CERN–2013–007
16 December 2013

ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE
CERN EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH

The CERN Accelerator School

Ion Sources

Senec, Slovakia
29 May – 8 June 2012

Proceedings

Editor: R. Bailey

GENEVA
2013

ISBN 978–92–9083–395–6
ISSN 0007–8328
Copyright © CERN, 2013
Creative Commons Attribution 3.0
Knowledge transfer is an integral part of CERN's mission.
CERN publishes this report Open Access under the Creative Commons Attribution 3.0 license (http://creativecommons.org/licenses/by/3.0/) in order to permit its wide dissemination and use.

This report should be cited as:

Proceedings of the CAS–CERN Accelerator School: Ion Sources,
Senec, Slovakia, 29 May – 8 June 2012, edited by R. Bailey, CERN-2013-007, (CERN, Geneva, 2013),
DOI: 10.5170/CERN–2013–007

A contribution in this report should be cited as:

[Author name(s)], in Proceedings of the CAS–CERN Accelerator School: Ion Sources,
Senec, Slovakia, 29 May – 8 June 2012, edited by R. Bailey, CERN-2013-007, pp. [first page] – [lastpage],
DOI: 10.5170/CERN–2013–007. [first page]

Abstract

These proceedings collate lectures given at the twenty-sixth specialized course organised by the CERN Accelerator School (CAS). The course was held in Senec, Slovakia from 29 May to 8 June 2012, in collaboration with the Slovak University of Technology, Bratislava.

Following lectures on the background accelerator physics and the fundamental processes underlying the production of particle beams, the course covered the different types of ion sources and associated technological issues, and different applications for which sources are used. The participants pursued one of six case studies in order to get "hands-on" experience of the issues connected with the design of an ion source.

Foreword

The aim of the CERN Accelerator School (CAS) is to collect, preserve and disseminate the knowledge accumulated in the world's accelerator laboratories over the years. This applies not only to general accelerator physics, but also to related sub-systems and associated technologies, and to how these are adapted to particular requirements. This wider aim is achieved by means of specialized courses currently held yearly. The topic of the 2012 specialized course was Ion Sources and was held at the Hotel Senec, Senec, Slovakia from 29 May to 8 June 2012.

The course was made possible by the productive collaboration with the Slovak University of Technology in Bratislava, Slovakia, in particular through the efforts of Marius Pavlovic. The collaboration with the Slovak University of Technology in Bratislava was one of activities of the KEGA 019STU-4/2012 project financed by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

The backing of the CERN management and the guidance of the CAS Advisory and Programme Committees enabled the course to take place, while the attention to detail of the Local Organising Committee and the management and staff of the Hotel Senec ensured that the school was held under optimum conditions.

The generous financial support offered by GSI Darmstadt, Germany and ZTS VVU Kosice a.s., Slovakia allowed scholarships to be offered to highly deserving young students who would otherwise not have been able to attend, while MedAustron kindly funded the banquet. An Endowment Fund of Slovenské elektrárne in the Pontis Foundation, Slovakia covered the participation fee of three local students from the Slovak University of Technology in Bratislava. Finally, the city of Senec provided maps and other useful material for visitors.

Special thanks must go to the lecturers for the preparation and presentation of the lectures, even more so to those who have written a manuscript for these proceedings.

For the production of the proceedings we are indebted to the efforts of Barbara Strasser and to the CERN Information and Publishing Services, especially Jerome Caffaro and Valeria Brancolini for their very positive collaboration.

Finally, the enthusiasm of the participants of 25 nationalities, from institutes in 19 countries, provides convincing proof of the usefulness and success of the course.

These Proceedings have been published in paper (black and white) and electronic form. The electronic version, with full color figures, can be found at http://cds.cern.ch/record/1445287?ln=en.

Roger Bailey
CERN Accelerator School

Programme - Ion Sources
29 May – 8 June, Senec, Slovakia 2012

Time	Tuesday 29 May	Wednesday 30 May	Thursday 31 May	Friday 1 June	Saturday 2 June	Sunday 3 June	Monday 4 June	Tuesday 5 June	Wednesday 6 June	Thursday 7 June	Friday 8 June
08:30 09:30		Requirements of Ion Sources R. Scrivens	Space Charge Effects N. Chauvin	Electron Cyclotron Resonance Ion Sources I T. Thuillier	Volume & Surface-Enhanced Negative Ion Sources M. Stockli		MEVVA Ion Sources I. Brown	Radioactive Ion Sources T. Stora	Plasma Diagnostics U. Fantz	Beam Extraction and Transport I T. Kalvas	
09:30 10:30		Definition of Ion Sources R. Scrivens	Fundamental Processes: Atomic Physics I M. Kowalska	Electron Cyclotron Resonance Ion Sources II T. Thuillier	Negative Ion Sources (Magnetron, Penning) D. Faircloth		Other Ion Sources P. Sortais	Charge Breeders F. Wenander	Beam Diagnostics U. Raich	Beam Extraction and Transport II T. Kalvas	
		COFFEE	**COFFEE**	**COFFEE**	**COFFEE**		**COFFEE**	**COFFEE**	**COFFEE**	**COFFEE**	
11:00 12:00		Theory of Relativity C. Prior	Fundamental Processes: Atomic Physics II M. Kowalska	Electron Beam Ion Sources G. Zschornack	Ion Sources for Fusion I W. Kraus		Fundamentals of Microwave Engineering & RF Coupling Issues I L. Celona	Technological Aspects: High Voltage D. Faircloth	Modelling and Simulations P. Spaedtke	Research Accelerator Applications R. Bailey	
12:00 13:00		Transverse Dynamics B. Holzer	Fundamental Processes: Plasma Physics I K. Wiesemann	Laser Ion Sources I B. Marsh	Ion Sources for Fusion II W. Kraus		Fundamentals of Microwave Engineering & RF Coupling Issues II L. Celona	Microwave Discharge Ion Sources L. Celona	Case Study Presentations	Industrial Applications P. Sortais	
		LUNCH	**LUNCH**	**LUNCH**	**LUNCH**		**LUNCH**	**LUNCH**	**LUNCH**	**LUNCH**	
14:30 15:30		Theory of Electro-magnetic Fields C. Prior	Fundamental Processes: Plasma Physics II K. Wiesemann	Laser Ion Sources II B. Marsh	Work on Case Study		Technological Aspects: Magnetic System Design T. Thuillier	Alternative Heating Methods L. Celona	Case Study Presentations	Medical Applications T. Haberer	
15:30 16:30		Longitudinal Dynamics B. Holzer	Production of High Intensity, Highly Charged Ions I S. Gammino	RF Ion Sources W. Kraus	Work on Case Study		Work on Case Study	Vacuum Requirements P. Chiggiato	Case Study Presentations	Summary of Case Study P. Spaedtke	
		TEA	**TEA**	**TEA**	**TEA**		**TEA**	**TEA**	**TEA**	**TEA**	
17:00 18:00		Introduction to Case Study P. Spaedtke	Production of High Intensity, Highly Charged Ions II S. Gammino	**Seminar** MedAustron Project A. Fabich	Discussion Session		Work on Case Study	**Seminar** Multi-Beam Mask Writer Develop-ment for the Semiconductor Industry H. Loeschner	Case Study Presentations	Directions for the Future A. Aanesland	
19:00 19:30	**Buffet Dinner**	**Welcome Drink/Dinner**	**Dinner**	**Dinner**	**Dinner**	**Special Dinner**	**Dinner**	**Dinner**	**Dinner**	**Dinner**	

A R R I V A L D A Y

E X C U R S I O N

D E P A R T U R E D A Y

Contents

Foreword
R. Bailey v

Requirements for ion sources
R. Scrivens 1

Classification of ion sources
R. Scrivens 9

Introduction to Transverse Beam Dynamics
B.J. Holzer 27

Introduction to Longitudinal Beam Dynamics
B.J. Holzer 47

Space-Charge effects
N. Chauvin 63

A Short Introduction to Plasma Physics
K. Wiesemann 85

Production of High-Intensity, Highly Charged Ions
S. Gammino 123

Electron Beam Ion Sources
G. Zschornack, M. Schmidt, A. Thorn 165

Resonance Ionization Laser Ion Sources
B. Marsh 203

Volume and Surface-Enhanced Volume Negative Ion Sources
M.P. Stockli 265

Negative Ion Sources: Magnetron and Penning
D.C. Faircloth 285

Vacuum Arc Ion Sources
I. Brown 311

Radioactive Ion Sources
T. Stora 331

Charge Breeding of Radioactive Ions
F.J.C. Wenander 351

Technological Aspects: High Voltage
D.C. Faircloth 381

Microwave Discharge Ion Sources
L. Celona 421

Microwave Coupling to ECR and Alternative Heating Methods
L. Celona 443

Vacuum Technology for Ion Sources
P. Chiggiato 463

Beam Diagnostics
U. Raich 503

Modelling
P. Spädtke 515
Beam Extraction and Transport
T. Kalvas 537
An Application for Research – the Large Hadron Collider
R. Bailey 565
Direction for the Future – Successive Acceleration of Positive and Negative Ions Applied to Space Propulsion
A. Aanesland, J. Bredin, L. Popelier and P. Chabert 575
List of Participants 585

Requirements for Ion Sources

R. Scrivens
CERN, Geneva, Switzerland

Abstract
Ion sources produce beams for a large variety of different physical experiments, industrial processes and medical applications. In order to characterize the beam delivered by them, a list of requirements is necessary. In this chapter the list of principal requirements is specified and definitions for them are given.

1 Uses for ion sources

Ion sources are one of the critical components for the production of high-energy beams of particles, which have many applications, not just in the field of scientific studies, but also for medical applications, material modification and surface treatments, and semiconductor doping. Even possibly, in the future, they may find applications in energy production through inertial confinement fusion, magnetic confinement fusion or energy amplification.

As the ion source is a driver for a further process or study, there are certain specifications on its performance, which must be quantified. In this chapter, the definitions of this characterization will be summarized.

2 The need for ion sources

One item that all particle accelerators have in common is that they have to start with a source of particles. For many applications and experiments, it is necessary to use heavier particles, such as protons, the nucleus of the hydrogen atom.

In order to impart a large amount of energy to an atom, it is necessary to apply a force to it. We are only able to create electric and magnetic fields in a well-defined manner, and in general the gravitational field is too weak to be able to attain a reasonable energy gain. (For example, a hydrogen atom allowed to free fall to the Earth's surface would reach a total energy of 0.65 eV as it reached sea level.) In order to use the electromagnetic forces, we must give atoms an overall charge. The simplest way to do this is by removing one or more electrons, or attaching additional electrons, and therefore creating ions. The device for doing this is an ion source.

Once transformed into a charged state, the ions can be manipulated with the Lorentz force – see Eq. (1) – which allows them to gain energy and therefore to be accelerated with an electric field, as well as to be deflected with a magnetic field:

$$\vec{F} = qe\left(\vec{E} + \vec{v} \times \vec{B}\right). \tag{1}$$

In a more general sense, we consider an *ion source* to be a device that ionizes atoms in order to produce ions, and provides a first stage of acceleration (ion extraction) in order to produce a beam of ions. We define an *ion beam* as an ensemble of ions that move with an average velocity and direction.

3 Characterization

The characterization of the ions, and the qualities of the beam into which they are formed, require several definitions. Seen from the viewpoint of the requirements of the user of the ion beam, they are necessary to allow the choice of the type of source to be made.

3.1 Particle type

The type of particle required for the application is the most fundamental parameter to be chosen, and is symbolically represented in the following way:

$$\left({}^{A}\mathrm{E}^{q+}\right)_{n} \tag{2}$$

The first principal choice is the type of atom from the periodic table (here given the general symbol E for 'element'). Further to this, the application may require a specific isotope to be provided (A), which could be one of a stable type, or, in the case of radioactive ion beams, an unstable isotope. The charge state of the ion (q) can be either a positive state, up to the fully stripped nucleus, or a negative ion, with the addition of an electron to the atom. Next, the requirements may not be for ionized atoms, but for ionized molecules (n), where the atom may be paired with the same atom type, or a molecule could be required of specific, dissimilar atoms (which may have their own isotope specifications).

Finally, the application may require different ion types to be available, which may only be possible with the use of several ion sources, and the time to switch between ion types may be a factor to consider.

3.2 Energy

It may be possible to produce ions of the energy required by the application directly from the source itself (and its extraction system), or the ions may need further acceleration in a particle accelerator (using DC or radiofrequency (RF) electric fields). In the case that a further accelerator is needed, the energy from the source must be chosen in collaboration with the subsequent acceleration stage.

The typical definition of the energy of the beam from an ion source is made purely in terms of the kinetic energy, i.e. the rest mass energy is ignored. The energy is then well defined by the voltage used to accelerate the ions, i.e.

$$E_{\text{kinetic, total}} = qeU \ , \tag{3}$$

where q is the charge state of the ion (1, 2, 3, …), e is the electron charge and U is the applied voltage (or equivalent) used for acceleration. The use of a negative ion requires that the voltage is also reversed. For the definition of the ion energy, the electron charge is usually contained in the units, so a ${}^{20}\mathrm{Ne}^{5+}$ ion accelerated across 100 kV would be defined as having an energy of 500 keV, whereas in SI units the energy would be 8×10^{-14} J.

Many types of RF particle accelerator require a fixed ion velocity at their input. In the case that different ion species are to be delivered by an ion source, and accelerated, it is very common to use the kinetic energy per nucleon as a definition:

$$E_{\text{kinetic, per nucleon}} = \frac{E_{\text{kinetic, total}}}{A} = \frac{qeU}{A} = \frac{1}{2}m_{\mathrm{u}}v^2 = (\gamma - 1)m_{\mathrm{u}}c^2 \ , \tag{4}$$

where A is the mass number of the ion, m_{u} is the rest nuclear mass unit (1.67×10^{-27} kg), c is the speed of light and γ is the relativistic gamma factor. The final two terms demonstrate that the kinetic energy per nucleon is only dependent on the velocity of the particle (either v or γ).

The advantages of using a higher particle energy are:

- space charge is reduced – higher beam currents are possible;
- reduces energy spread ratio ($\Delta E/E$);
- reduces the geometrical emittance ⇒ smaller apertures;
- higher velocity makes it easier to inject into an RF accelerator (e.g. drift tube linac).

And the disadvantages are:

- higher risk of sparking, implying a higher degree of complexity to avoid it;
- longer radio frequency quadrupole (as the input cells become longer);
- higher fields for beam devices;
- higher energy = higher beam power ⇒ consequences for beam intercepting devices.

3.2.1 Intensity and repetition rate

The beam intensity from a source is usually quoted in terms of an equivalent electrical current, i.e.

$$I_{\text{electric current}} = \frac{qeN_{\text{ions}}}{t} . \tag{5}$$

In the case of multiply charged ion beams, the particle current may be given, which is the equivalent current for a beam where all ions have a charged state of +1:

$$I_{\text{particle current}} = \frac{eN_{\text{ions}}}{t} = \frac{I_{\text{electric current}}}{q} . \tag{6}$$

It is normally accepted to quote the beam current during the beam pulse, although this is not always the case.

The repetition cycle must be defined (unless the beam is required in continuous mode). The pulse length and the repetition rate are sufficient, but care must be taken to understand the effects of beam outside this time window.

3.3 Space charge

Ions within our beam of particles will all have the same polarity of charge, which means they will all repel each other, an effect that is called *space charge*.

For beams of *electrons*, there is a relatively simple way to characterize how much the space-charge force will affect the beam. This is found by calculating the beam perveance (P), which is defined in terms of the beam electric current (I) and the beam energy (U) as

$$P = \frac{I}{U^{3/2}} . \tag{7}$$

If two electron beams of the same initial beam size and perveance are left to drift through an otherwise field-free region, they will expand at the same rate (relative to the longitudinal direction of travel).

However, this simple perveance equation for electrons is only valid because all electrons are alike. For ions, we have to take into account also the masses and charges of the ions in the beam. In this case, we develop the poissance (Π), which is defined as

$$\Pi = \frac{9}{4\varepsilon_0}\frac{I}{U^{3/2}}\sqrt{\frac{m}{2qe}}\frac{a}{b}, \tag{8}$$

with the usual definitions for all the symbols, and with a and b being the height and width of the beam. (This leads to a beam having a different poissance in each plane, with the smaller size having the large poissance value.) The poissance equation allows the comparison of different particle beams, of different energies and ion types. It shows that low-energy, low-charge-state, high-mass beams can have very strong space-charge characteristics; for example, a 40 mA, 50 keV proton beam has the same Π as a 0.2 mA, 5 keV Ar^+ beam.

An important effect that helps to reduce the space-charge forces of the beam at low energies is the trapping of ions or electrons produced by beam collisions with the residual gas. This space-charge compensation works when the beam is in a DC or quasi-DC mode (i.e. pulsed for some tens of microseconds or more).

3.4 Emittance

The emittance of a single particle or a beam can be considered from the viewpoint of the amplitude term of the solution to Hill's equation for the periodic structure of a circular particle accelerator [1]. However, when considering the beam from an ion source, there is no periodic structure with which to compare the ion position, and the emittance definition needs to be considered in a different (but consistent) way.

If we consider a particle in a beam travelling along the beam axis (which we will call the z-axis in this case), we can plot the particle's position (x) and angle (x') relative to z on a graph, which is referred to as a phase-space plot. If this is done for all the particles in a beam, we can draw an enclosing shape around all the particles (see Fig. 1). Then the total, geometric emittance is defined as the area of the phase space covered with particles, divided by π, i.e.

$$\varepsilon_{\text{total, geometric}} = \frac{\text{area}}{\pi}. \tag{9}$$

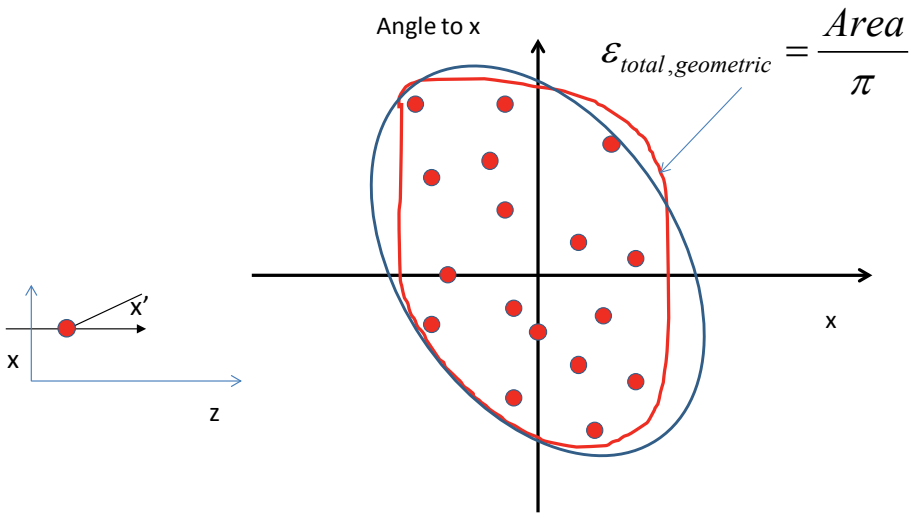


Fig. 1: Definition of axes for a beam particle, travelling along the z-axis, at a position x and angle x'. Many beam particle positions can be plotted in phase space, leading to the definition of a beam emittance.

The solution to Hill's equation describes that a particle orbiting a synchrotron, when measured as a fixed position along the synchrotron, will make the shape of an ellipse in phase space, Thus it is very common to consider the beam from a source also to be enclosed by an ellipse, and therefore the total emittance can be considered to be the area of this ellipse (divided by π).

Very often, the emittance will be quoted as $\varepsilon\,\pi$ mm mrad, i.e. there is the inclusion of a π in the units. Correctly, whether the units contain this *flag* of π or not, the value of ε is always that corresponding to Eq. (9).

Although the convention of considering an ellipse is retained, ion sources rarely produce a beam that fits well with the definition of an ellipse, and seldom is the beam distribution inside either uniform or Gaussian.

The root mean square (r.m.s.) emittance is a statistical definition of the amount of phase space covered by a beam, and is given by

$$\varepsilon_{\text{r.m.s.}} = \frac{1}{N^2}\left\{\left[N\sum x^2 - \left(\sum x\right)^2\right]\left[N\sum x'^2 - \left(\sum x'\right)^2\right] - \left[N\sum xx' - \sum x \sum x'\right]^2\right\}^{1/2}, \tag{10}$$

$$\varepsilon_{\text{r.m.s.}} = \frac{1}{N}\left(\sum x^2 \sum x'^2 - \left(\sum xx'\right)^2\right)^{1/2}, \qquad \overline{x} = \overline{x'} = 0, \tag{11}$$

where the second line is valid when the beam is centred in both position and angle. In this centred beam case, we can also calculate the Twiss parameter of the beam:

$$\gamma_x \varepsilon_{\text{r.m.s.}} = \left\langle x'^2 \right\rangle, \qquad \alpha_x \varepsilon_{\text{r.m.s.}} = \left\langle xx' \right\rangle, \qquad \beta_x \varepsilon_{\text{r.m.s.}} = \left\langle x^2 \right\rangle. \tag{12}$$

The value of the emittance defined above is reduced when the ions are accelerated. In order to express this requirement more clearly, the definition of an emittance that does not change with energy is needed, which is called the normalized emittance. This is normally symbolized by ε_{n} and sometimes the symbol ε^* is especially used to note that the emittance is normalized, 1 r.m.s.

We also consider the concept of the beam intensity normalized by the beam emittance, for which we give the name beam brightness (B_{b}):

$$B_{\text{b}} = \frac{I_{\text{b}}}{\pi^2 \varepsilon_x \varepsilon_y}. \tag{13}$$

Considering the various definitions of the emittance, it is not normally useful to quote a value for a beam brightness, but it is useful for comparing the relative performance of different sources.

3.5 Material efficiency

Ion sources take a raw material as an input to produce ions, and this raw material can sometimes be rare, difficult to produce or hazardous. Examples of these can be rare elements (for example, the element iridium) and radioactive isotopes. For ion sources that need these raw inputs, there is therefore a benefit to have a high efficiency of conversion from the input material to the output ion desired.

The simple definition of efficiency is therefore

$$\eta = \frac{N_{\text{ions}}}{N_{\text{atoms}}}. \tag{14}$$

3.6 Beam purity

Several applications will require a very pure beam – for example, for hadron therapy – and for many physics experiments – for example, where reaction rates are measured by knowing the incoming beam flux of the ion projectile being studied.

Typically, the ion transportation or acceleration system can play an important role in beam purification, with RF acceleration and spectrometer magnets (and Wien filters) usually allowing ions of the wrong charge-to-mass ratio to be filtered out.

In some cases, this will not be sufficient, or the application is placed close to the source, and therefore the source should already play a role in eliminating unwanted ions. Plasma-based sources are not very well suited to producing clean beams, as they have a large range of electron energies that can ionize any element. Resonant ionization laser ion sources are a good choice, as they are able to use internal excitation of an electron in an atom to an excited state, from which it can be ionized with a low-energy laser (which is not able to ionize other elements that are in their ground states).

Some selectivity is also available from surface ionization sources, where only atoms with low ionization potential will be ionized, but this can still lead to a few different ion types being produced.

3.7 Outlet pressure

Most (but not all) sources require a large amount of material to be injected in order to produce ions, and the un-ionized gas or vapour of the material will also exit the source. The gas ejected from the source can lead to breakdowns in both the high-voltage regions of the extraction system and in other high-field regions. The gas can also interact with the beam to cause beam recombination or stripping, which lowers the intensity and can lead to high powers of beam losses for high-intensity sources.

3.8 Reliability

As ion sources are the driver of other processes or experiments, it is clear that high reliability is desirable. In general, there will be a trade-off between the other performance factors of the source. For example, high-intensity sources (in particular, if they run at a high duty factor) can suffer from wear due to the ion bombardment of the surfaces, which can also coat other surfaces in the source, thereby limiting the lifetime.

In terms of reliability, any of the final beam parameters of the source being outside defined limits can lead to the source being effectively in fault, depending on the application. Lower than nominal beam intensities, for example, may be tolerated by increasing the length of an experiment, but the ion beam energy not being correct (e.g. due to voltage holding problems) may be equivalent to no beam at all.

Ion source faults can be due to issues with the ion source itself (e.g. failure of a cathode), or to sparking across an insulator, or to an external service to the source (e.g. a high-voltage power supply or failure of the cooling water supply to the source). Faults to the services to the source are often quickly repaired (assuming the spare parts are available), unless the source requires a thermal equilibrium to be reached (e.g. a cathode to be heated to a certain temperature), and the re-establishment of this configuration can take some time. Regarding problems to the source itself, the downtime can be much longer, owing to the requirement to open the vacuum system, and potentially to recondition the source after a change of component (or the complete change of the source).

Like any system, ion source reliability can be characterized by the following metrics:

- Availability, the fraction of the time that a process is working as required.
- Mean time between failures (MTBF), the average time between two faults that need intervention.

- Mean time to repair (MTTR), the average time it takes to repair a fault (weighted with the probability of the fault happening).

In order to increase the availability, there are some general rules that can help to increase the availability of the ion source:

- improve EMI sensitivity to avoid trips due to breakdowns;
- use uninterruptable power supplies to reduce sensitivity to power cuts;
- keep installations clean (inside the vacuum and outside) in order to avoid discharges;
- keep spares, and maintain and test them;
- cold-spare sources (fully built, tested and ready for installation);
- hot-spare sources (two installed sources, with a switch-yard into the beam line, to switch quickly between them).

All of these come at a cost, in both initial investment and operating costs.

3.9 Stability

The variation of any of the operating parameters of an ion source can be deemed to lead to an unstable situation for its operation. Although this is most obviously seen if the beam intensity is varying, other beam parameters are, of course, of concern to the system utilizing the beam.

4 High-charge-state ion sources

The energy of an ion beam can be increased more effectively by increasing the charge state of the ion. By being more effective, ions can therefore be accelerated in shorter distances, which reduces costs in the construction of the accelerator, and typically will reduce the electrical power required to produce the necessary acceleration.

Ion sources produce higher charge states through a process of stepwise ionization, that is, in a chain from $1+ \rightarrow 2+ \rightarrow \ldots \rightarrow n+$. This requires that two conditions be fulfilled so that the charge state $n+$ can be attained.

1. The ionizing process must have sufficient energy that excitation from $(n-1)+ \rightarrow n+$ is possible.
2. The ions are kept in the plasma long enough for them to attain the charge state required.

For the second of these points, and assuming that the ionization process is via electron impacts onto ions, a confinement time (t_{conf}) can be defined, which is the product of the electron density (n_{e}) and the lifetime of the ion in the source (τ), that is,

$$t_{\mathrm{conf}} = n_{\mathrm{e}}\tau \,. \tag{15}$$

The required charge state can be reached either by having a high-density ion ionizing electrons, or by confining the ions for a sufficiently long time.

5 Negative ions

The majority of elements in the periodic table have the ability to attach an additional electron, with the energy released by making the attachment being called the *electron affinity*. The values of the electron affinities for the elements of the first two rows of the periodic table are given in Table 1.

Table 1: Electron affinities for the elements of the first two rows of the periodic table.

Element	Electron affinity (eV)	Element	Electron affinity (eV)
Hydrogen (1)	0.75	Carbon (6)	1.26
Lithium (3)	0.62	Oxygen (8)	1.46
Boron (5)	0.28	Fluorine (9)	3.40

Negative ions are of special interest for acceleration systems, owing to the ease with which they can be converted into neutral or positive ions. There are several mechanisms for making this conversion: by stripping using material (in either the solid, gas or plasma form), by photo-detachment or by using a sufficiently strong magnetic field to pull the loosely bound electron off the ion.

The interest in changing the polarity of the ion is because this nonlinear process allows the fast reversal of the force produced by an applied electric and magnetic field (or to cancel it in the case of conversion to a neutral atom). The following are four well-known applications of negative ions:

1. *The tandem accelerator.* Negative ions are accelerated from ground potential to a high-voltage terminal, where they are stripped to positive ions, and accelerated by the opposite electric field to ground potential. The ion energy is two (or more) times the applied voltage.
2. *Ion extraction from cyclotrons.* Negative ions are accelerated in a cyclotron, and stripped into protons, allowing the magnetic field to extract them directly, avoiding complicated extraction geometries.
3. *Charge exchange injection into synchrotrons.* Negative ions are accelerated in a linear accelerator, and overlapped onto circulating positive ions in a synchrotron (due to the opposite bending in a magnetic field). Both beams are passed through a stripping system, emerging as positive ions in a higher-brightness beam.
4. *Magnetic confinement fusion – neutral beam injection.* Negative ions are accelerated, then stripped into a neutral, high-energy (and high-power) beam that can cross the magnetic fields into a magnetic confinement region.

A further application that is far beyond the scope of this chapter is the creation of polarized protons, where the orbital electrons are used to align the angular moment of the proton/nucleus.

Negative ions bring with them further challenges for source engineering and beam transport, including dealing with electron beams co-extracted with the ions, and unwanted gas stripping in the transport lines, but they do bring with them some interesting possibilities for beam instrumentation.

6 Summary

In this chapter, the requirements for ion sources and the ways to specify them have been given. In the following chapters, ion sources will be sorted into general types, and will be explained in more detail.

Reference

[1] CAS, Intermediate Accelerator Physics (Ed D. Brandt), DESY, Zeuthen, Germany, September 2003, CERN-2006-02, p. 135.

Classification of Ion Sources

R. Scrivens
CERN, Geneva, Switzerland

Abstract
In this chapter, the anatomy of an ion source is briefly described, as well as a few features of particle motion in electric and magnetic fields, and of particle dynamics and plasmas. Using this information, different types of ion sources are described, highlighting their main mode of operation.

1 Introduction

The generation of ions for an ion source is based fundamentally on three main processes, which are

i) electron impact ionization,

ii) photo-induced ionization and

iii) surface ionization.

In addition to these, some other atomic and molecular processes can be the process leading to ion extraction, for example, charge exchange between ions and atoms, and molecular dissociation into atoms and ions.

Ion sources have always been soundly rooted in the research arena, which has led to continual optimization of the design for each application type and results in an enormous array of ion source types (just a few of which are shown in Fig. 1). In this chapter we will journey through the main different types of ion source, dealing with the main physical processes required for each source to work. To do this, ion sources will be classified into the following general types:

- electron bombardment,
- d.c./pulsed plasma discharge,
- radio-frequency (RF) discharge,
- microwave and electron cyclotron resonance (ECR),
- laser driven,
- surface and
- charge exchange.

Several excellent books on ion sources exist, but, for the most part, Bernhard Wolf's book *Handbook of Ion Sources* [1] has been used as the main source of information for this chapter.

Overall, this chapter will give a very basic overview of each source type and the processes that make them work. Much more detailed information will be given for most of the source types in the other chapters of these proceedings specializing on each source configuration or process.

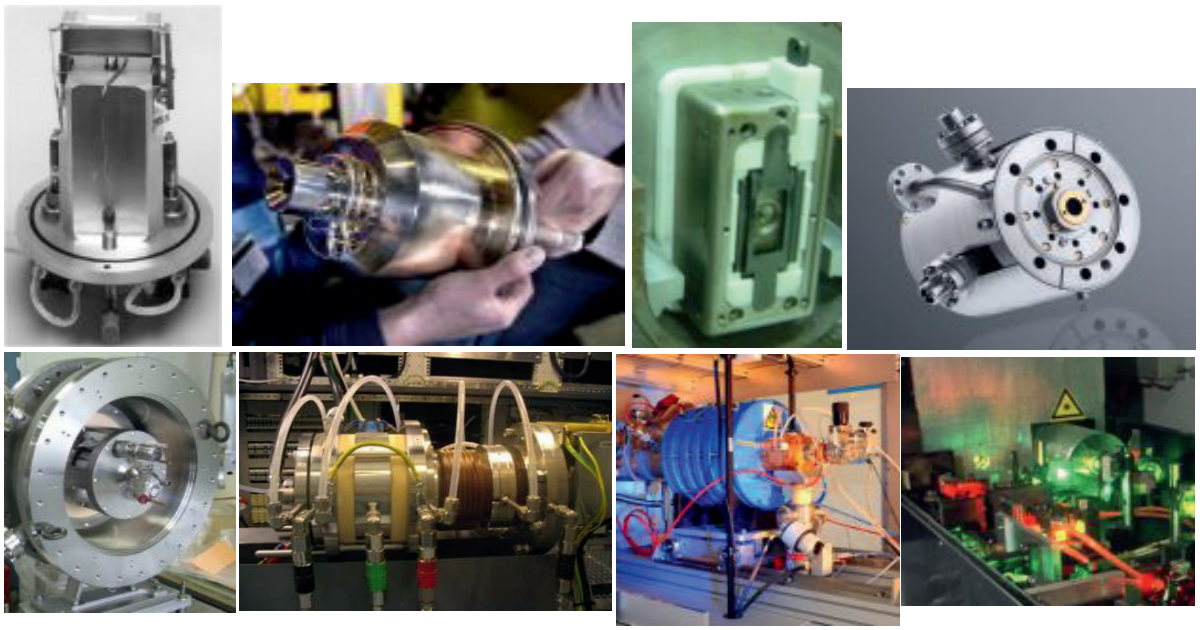


Fig. 1: An array of ion sources of various types. Top row: Freeman, Duoplasmatron, Magnetron, Penning. Bottom row: RF ion source, Chordis, ECR, a laser.

2 Basic components of an ion source

Before going any further with the description of ion source types, the basic components of an ion source need to be known. Ion sources are so diverse in type that a fully general scheme cannot be given, but for illustration Fig. 2 shows the main components inside a cathode-type ion source, and they can be summarized as follows.

- *Main chamber.* Somewhere in the source there needs to be a chamber in which the ionization processes will take place and through which the created ions will drift in order to arrive at the region for extraction. This chamber will have to be vacuum-tight (or placed in another outer vessel), but depending on the source type it could be metal or ceramic. Somewhere in this chamber there must be a hole through which the ions will exit the source.
- *Material.* For the ion type required, material must be supplied into the source. This can be in the form of gas or a compound gas containing the desired ion type, as well as solid (or even liquid) material that will be heated to produce a gas. Sometimes the material can be introduced into the electrodes, such that it will be sputtered into the main source chamber.
- *Ionization energy source.* The ionization of atoms requires some energy, and therefore power will have to be delivered in some form to the source (for example, via light, or electrical power).
- *Extraction system.* The hole from which the ions will emerge will need to have an electric field applied so that the ions can be extracted from the source, and accelerated to an initial energy.

From these fundamental components onwards, different source types start to diverge dramatically, which will be seen in the next sections.

In addition to the above components, many systems are needed to make the source operational, including vacuum pumping systems, power supplies and controls, to name but a few.

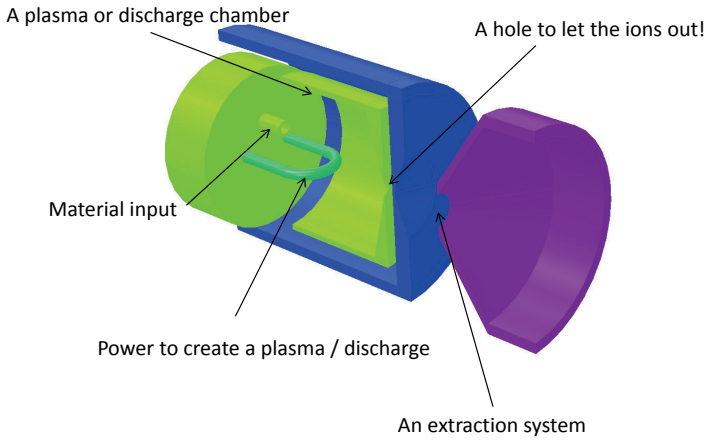


Fig. 2: The scheme of a typical ion source (for this example, using electron impact ionization)

3 Simple charged-particle dynamics

The charged particles in an ion source, that is, both the positive (and negative) ions and electrons, undergo forces due to the electric and magnetic fields. These fields are both externally applied to the source volume, and generated by the assembly of charged particles.

In order to have a preliminary understanding how the majority of sources are able to control the particles inside them, some simple single-particle dynamics is needed.

3.1 Charged particle in a magnetic field

In this simplest situation, a charged particle will make circular orbits in the plane perpendicular to the magnetic field, with a cyclotron radius (ρ_c) and a cyclotron frequency (ω_c) given by

$$\rho_c = \frac{\sqrt{2mE_\perp}}{qeB}, \tag{1}$$

$$\omega_c = \frac{qeB}{m}. \tag{2}$$

These equations show that the frequency of oscillation is constant for a given particle, and does not depend on the particle velocity around the magnetic field (as long as the energy is not high enough for relativistic corrections to be needed), but only on the magnetic field strength.

In the direction parallel to the magnetic field, the particle is able to drift without undergoing any force.

3.2 $E \times B$ drift

If an electric field is parallel to a magnetic field, a charged particle is accelerated along the direction of the electric field and is not subjected to any force due to the magnetic field.

However, if the electric field and magnetic field are perpendicular, a charged particle is accelerated by the electric field, and as it gains velocity and the magnetic field starts to curve its trajectory. Eventually the trajectory direction starts to oppose the electric field and the particle decelerates until it comes to rest, but displaced in a direction perpendicular to both the electric and magnetic field vectors.

This drift velocity is given vectorally by

$$v_{\text{drift}} = \frac{\boldsymbol{E} \times \boldsymbol{B}}{B^2}, \tag{3}$$

i.e., the drift velocity is independent of the particle charge and mass.

3.3 Collisional drift

The charged particles within a plasma undergo long-range elastic collisions that are able to change the particle's trajectory. The rate of these collisions can be characterized by the collisional frequency, given by

$$v_{\text{coll}} = 2 \times 10^{-6} n_{\text{e}} \ln \Lambda \; T_{\text{e}}^{-3/2} \; (m_{\text{e}}/m_{\text{particle}})^{1/2} \; \text{s}^{-1}, \tag{4}$$

where m_{particle} is the mass of the particle whose collisional frequency is being considered, and $\ln \Lambda$ is the Coulomb logarithm (a factor that describes the difference between long- and short-range collisions, and is typically taken to have a value of 10 for conventional plasmas).

As the charged particles in a plasma make orbits around the magnetic field, with a radius typically of the size of the cyclotron radius given by Eq. (1) (where the particle temperature can be used as an approximation for the energy), the collisions cause the trajectory to be changed. The rate at which the particles drift is a direction perpendicular to the magnetic field (D) due to these collisions can be approximated by

$$D \sim \rho_{\text{c}}^2 v_{\text{coll}} \sim \left(\frac{\sqrt{2 m_{\text{particle}} E_{\perp}}}{eB} \right)^2 \frac{1}{T^{3/2}} \left(\frac{m_{\text{e}}}{m_{\text{particle}}} \right)^{1/2} \sim \frac{m_{\text{particle}}^{1/2}}{T^{1/2}}. \tag{5}$$

From this result we see that, when a plasma is confined by a magnetic field, heavier particles drift perpendicularly to the field more quickly, as do particles in cooler plasmas, which is at odds with the conventional kinetic energy equation (i.e., $E = ½mv^2$).

4 Ion source types

4.1 Electron impact ionization sources

When an electron strikes a target atom (or molecule, or ion) there is a chance that during the impact one or more electrons are removed from the target particle, which turns it into an ion or increases its charge-state. Ion Sources using electron impact ionization work in practice by supplying a source of electrons (typically from a cathode), and a way to accelerate them to an energy sufficient to cause ionization of the material in a chamber. A schematic of an electron impact ionization source is shown in Fig. 3.

Electrons are made available from a cathode in the source, which in most cases will use thermionic emission, i.e., heating the material so that electrons gain some energy and are able to escape the work function of the cathode material (the energy required per electron to leave the material). The rate of electron emission from a surface (J, the electron current density) is given by the Richardson equation, Eq. (6), for a temperature T and material surface work function ϕ_{s}:

$$J = AT^2 \exp\left(\frac{-e\phi_{\text{s}}}{kT} \right) \tag{6}$$

where A is the Richardson–Dushman constant:

$$A = \frac{4\pi e m_e k^2}{h^3} \approx 1.2 \times 10^6 \ \mathrm{A\,m^{-2}\,K^{-2}}, \tag{7}$$

which, although a constant when the equation is derived from first principles, in practice turns out to be dependent on the material. Values for the work function and the constant A can be found in the literature for many cathode materials, and illustrative calculations of the current density that can be emitted as a function of temperature is shown in Fig. 4.

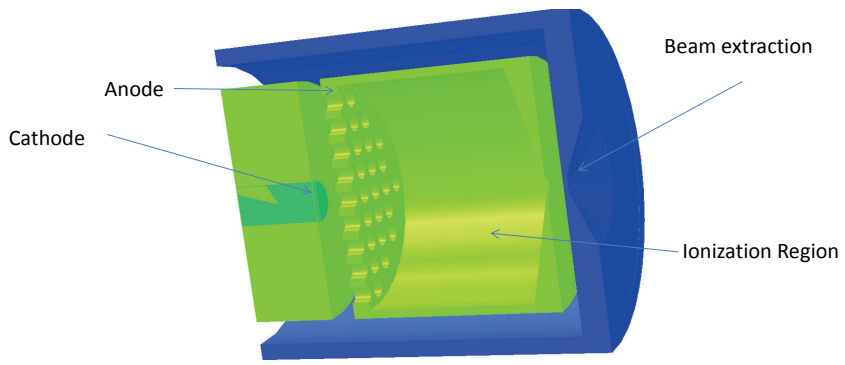


Fig. 3: Scheme of an electron impact ion source

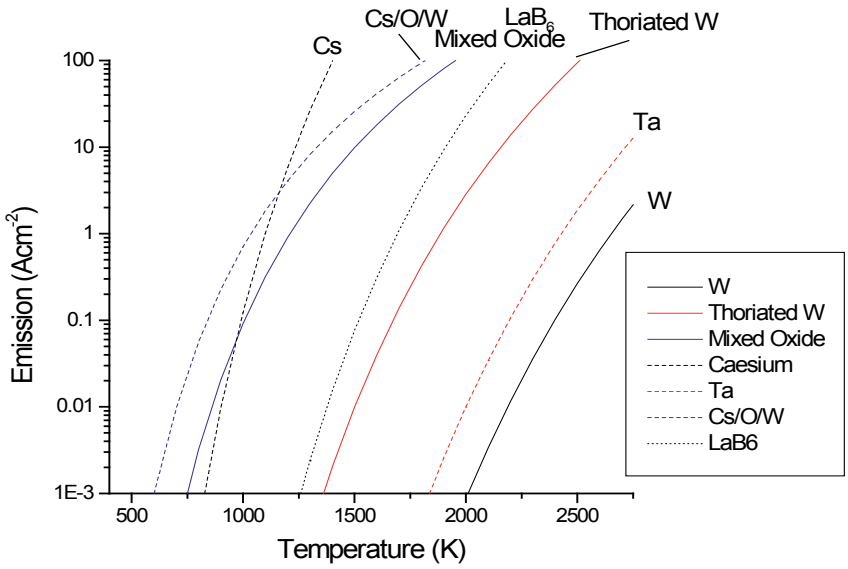


Fig. 4: Electron current density emitted from a material as a function of temperature for several cathode materials. Note that the temperature range shown for Cs is not valid, as it is well above the melting point for this element.

The target for the electrons will be neutral atoms or molecules inside the source ionization region. The rate at which ions will be created inside this source depends on the current of electrons available, and the cross-section for ionization of the target particle. We can understand the meaning of the cross-section (σ) as part of the formula:

$$\sigma n_{\text{atom}} = \frac{1}{L_{\text{collision}}}, \tag{8}$$

where n_{atom} is the number density of the target particles inside the source volume, and $L_{\text{collision}}$ is therefore the mean path the electron will travel between events causing ionization. If we generate an electron current in a source, and let each electron flow from the cathode to the anode with a distance L_{path}, then by comparing the total distance covered by all electrons to the distance needed to cause an ionization event, the number of ions created per unit time ($\mathrm{d}n_{\text{ion}}/\mathrm{d}t$) can be calculated as

$$\frac{\mathrm{d}n_{\text{ion}}}{\mathrm{d}t} = \frac{I\sigma n_{\text{atom}} L_{\text{path}}}{e}, \tag{9}$$

from which it can be seen that higher electron currents, cross-sections, gas density and path length for electrons to flow are all ways to increase the production of ions. In electron beam impact ionization sources, $L_{\text{collision}} >> L_{\text{path}}$, and therefore the majority of electrons flow from the cathode to the anode without causing an ionization event.

Data for the cross-section have been measured for most ions, and the values for a few of them are shown in Fig. 5. These data clearly show how the ionization cross-section can be very different depending on the ion type, and of course it is necessary that the electron energy is at least higher than the ionization energy.

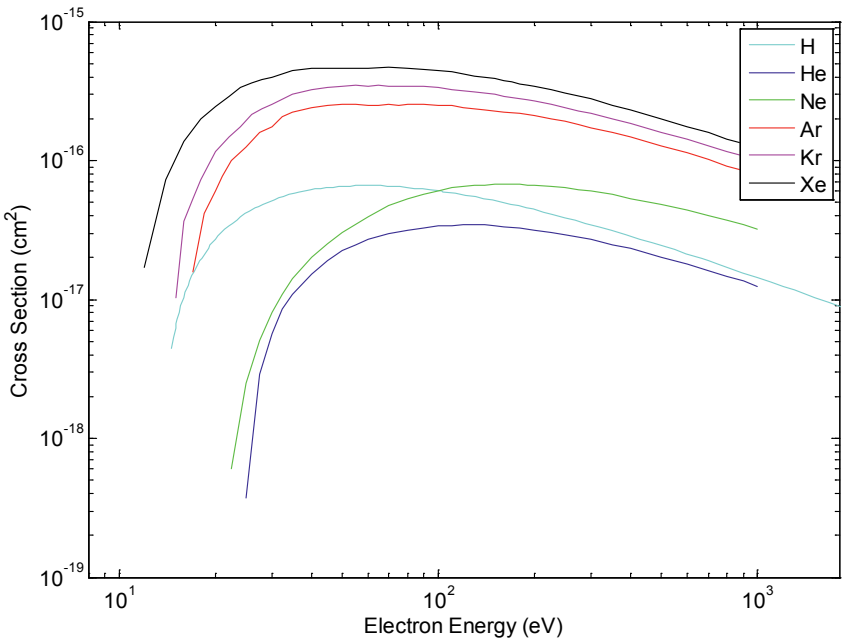


Fig. 5: Electron impact ionization cross-section as a function of electron energy, for selected noble gases [2] and hydrogen [3].

The operation mode of the source is therefore to generate electrons from a cathode, and accelerate them to an energy of a few times the ionization potential, often across a short distance to a grid. From there on, the electrons can drift in a field-free region, with the gas. The grid also prevents ions from the ionization region from being accelerated back towards the cathode, but instead have a chance to drift towards the exit hole of the source. A practical example of a FEBIAD (forced electron beam induced arc discharge) used for radioactive isotope ionization can be found in Ref. [4].

4.2 Plasma discharge ion sources

For the description of electron impact ionization sources above, the electron density and dynamics were dominated by thermionic emission from the cathode, and by the application of external fields. The charge density in the source was so far not high enough to affect the motion of electrons and ions.

As we move to high densities of neutral particles, electrons and then hopefully ions, several effects take place:

- Higher gas densities mean that $L_{\text{collision}} \sim L_{\text{path}}$ and therefore the majority of electrons will cause an ionization. This means that the electrons lose energy when they cause an ionization, and that new, low-energy electrons are created.
- Large numbers of ions are created, which can be accelerated to the cathode, where they also cause the emission of electrons from the cathode surface.

As the density of ions and electrons in the source increases, a plasma is considered to have been created, which leads to new effects, for example:

- quasi-neutrality (the density of positive and negative charges will generally balance),
- screening of externally applied electric fields (through sheath layers, which are not charge neutral).

The distance over which the screening of fields takes place in a plasma is characterized by the Debye length, which is given by

$$\lambda_{\text{D}} = \sqrt{\frac{\varepsilon_0 k T_{\text{e}}}{n_{\text{e}} e^2}}\,, \tag{10}$$

where T_{e} is the temperature of the electrons and n_{e} the electron density. If the Debye length is smaller than the source plasma, then bulk plasma effects will be important, and sheath and screening layers around electrodes will be present. The Debye length as a function of electron density and electron temperature is given in Fig. 6.

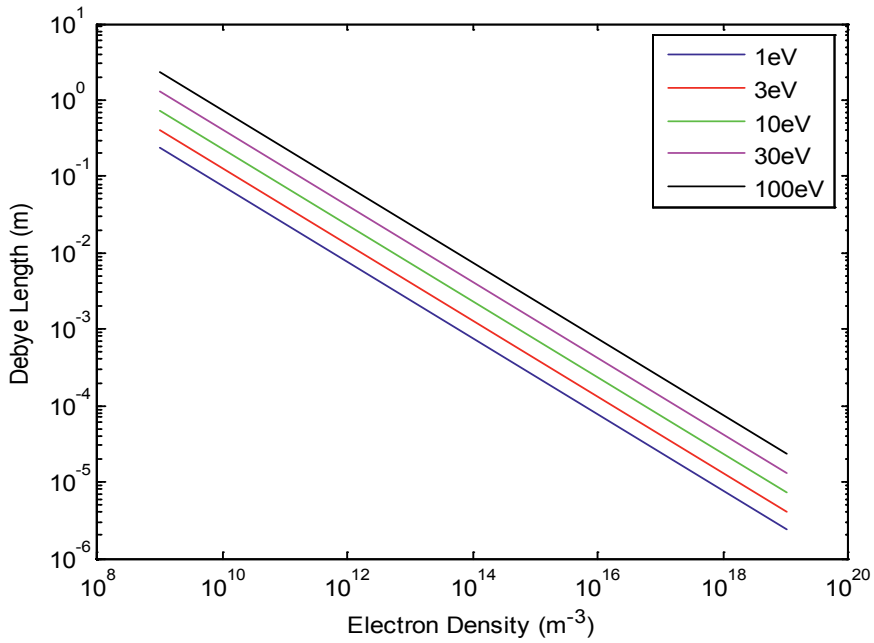


Fig. 6: Plasma Debye length as a function of electron density and electron temperature

The creation of significant numbers of ions and electrons from the initial gas will lead to the possibility of sustaining a discharge (i.e., the current flowing between cathode and anode electrodes is no longer dominated by the thermionic electrons from the cathode, but by new ions and electrons created from the gas, as well as secondary electrons from ion impact on the cathode surface).

Paschen investigated the breakdown in gases in the 1880s, and found that the voltage required to produce breakdown in a gas (V_b) was given by

$$V_b = \frac{aPd}{\ln(Pd) + b}, \tag{11}$$

where P is the pressure, d is the distance between the electrodes, and a and b are gas-dependent constants. For a given gas, the required voltage for a sustainable breakdown is a function of the product of the pressure and distance between the electrodes, and not just the electric field and pressure.

The plot of this function for a gas is called the Paschen curve (Fig. 7), for which we see that there is a minimum voltage needed to break down a gas (for example, in air this is 327 V). The plots also reveal some rather counter-intuitive phenomena. For example, considering a fixed pressure P, we can imagine applying a voltage V, and changing the distance between our electrodes (d). Starting with a very small distance (on the left side of the plot), we are below the Paschen curve, and so breakdown does not occur. As we increase the distance, we move above the curve and a breakdown can be made. Further increase d and we finally reach a distance where the discharge cannot be sustained.

It should be noted that the parameter b is also a function of the number of secondary electrons from the cathode per incident ion, and so Paschen curves are in fact dependent on the cathode characteristics.

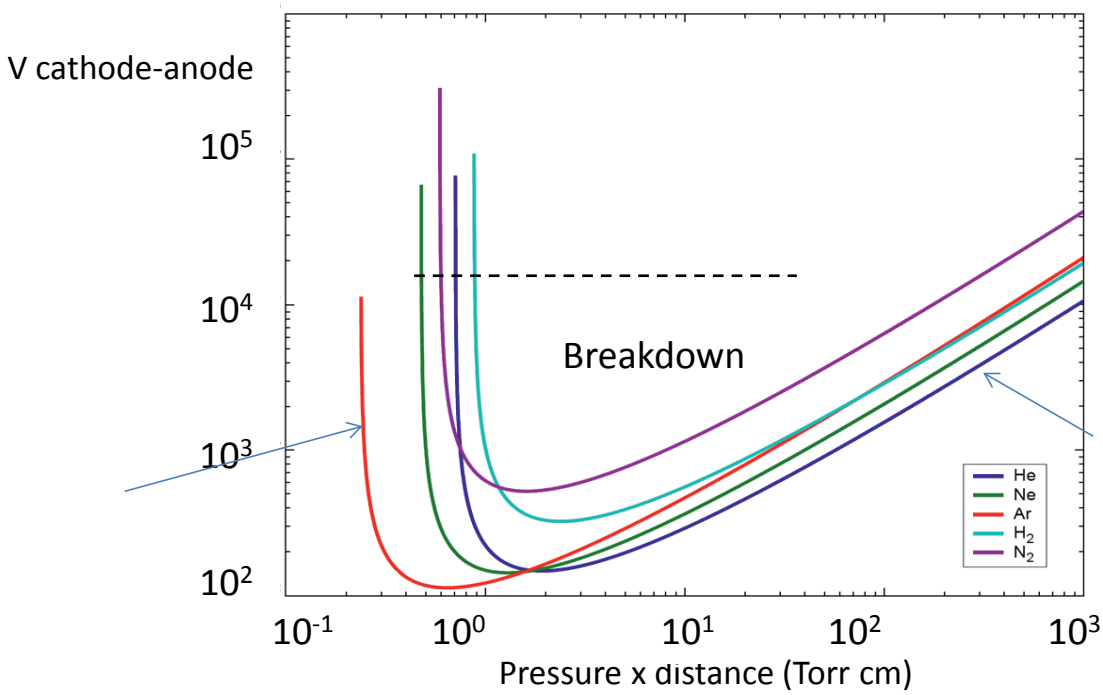


Fig. 7: Paschen curves, the lowest voltage for a sustained discharge as a function of the product of gas pressure and anode-to-cathode distance, for various gases.

The point of these electrical discharges is that they contain large quantities of ions that are of interest for ion sources. But to run a gas discharge at low pressures requires a large distance, as the

electrons must travel far enough to be able to have an ionizing collision. One way to increase the distance travelled by the electrons is to introduce a magnetic field to the source, and by changing the configuration of the electrodes and magnetic field we find a multitude of different source types, just a few of which are shown with a schematic of the discharge in Fig. 8.

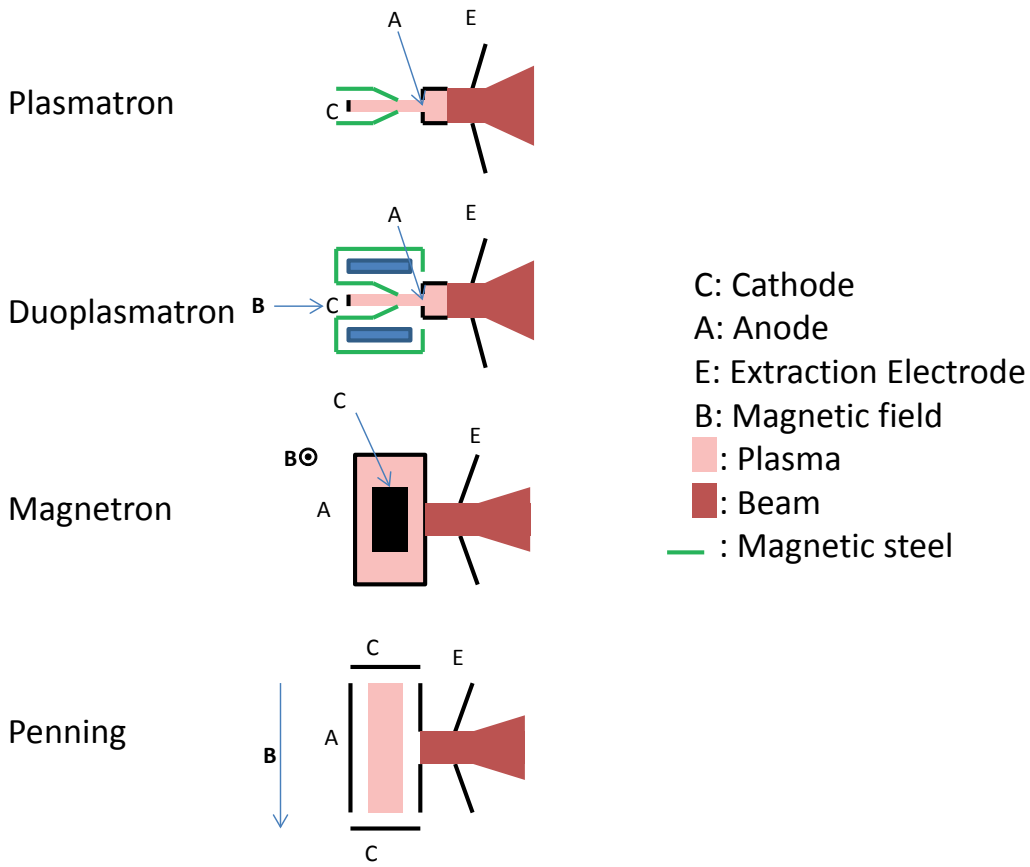


Fig. 8: Schematic configuration of a few discharge-type ion sources

Through all these source types there is the possibility to make almost any ion type, with a vast array of intensities and pulse structures, but all sources are limited in their lifetime by the problem that the cathode is constantly bombarded by ions, which sputter away the cathode material. This causes damage to the cathode (changing its surface shape, for example), can cause a change in the cathode composition (either to the crystal structure or to the ratio of elements in a compound cathode), and causes the cathode material to be deposited in places where it is not wanted (for example, across insulators).

The rate at which the cathode is sputtered depends on the cathode material, impacting ion type as well as the energy of impact. Measurements of the sputtering yield (atoms sputtered from a material per incident ion) can be found in the literature (e.g. Ref. [5]). A few plots are reproduced in Fig. 9 as examples, which show how the problem is particularly acute when heavier gases (or vapours) are used.

Part of the solution of the cathode sputtering problem is addressed with radio-frequency (RF), microwave, electron cyclotron resonance (ECR) and laser sources.

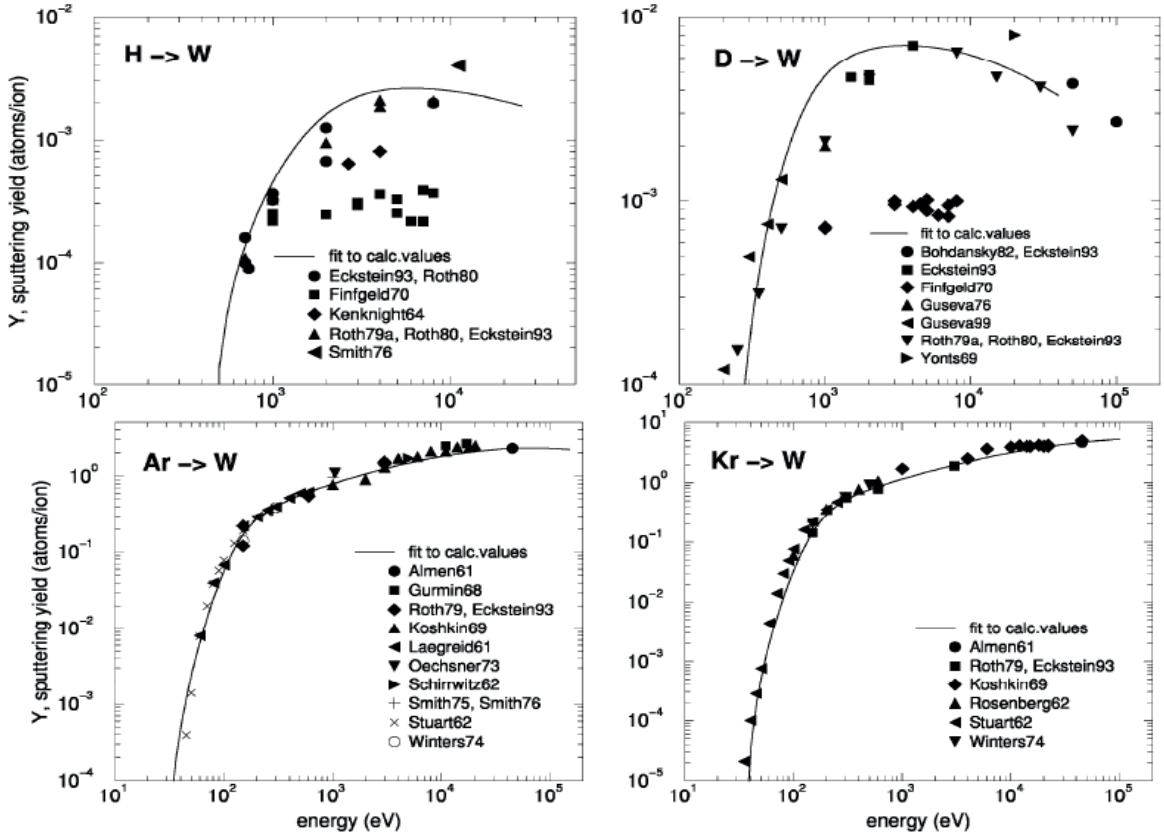


Fig. 9: Sputter yield (Y) for a few ion types onto tungsten, as a function of energy (from Ref. [5])

4.3 RF discharge

Instead of using a cathode–anode voltage to generate the electric field required to accelerate electrons to the energy required to cause ionization, a time-varying electric field can also be used. Two general configurations exist that are able to be applied around a plasma region, and these are to use a capacitive-type system with two electrodes, or a coil to give an inductive system. The two systems are shown schematically in Fig. 10.

Using a capacitively coupled system is conceptually fairly simple, but still has the disadvantage that the electric field lines terminate on the electrodes, which therefore take the place of the anode and cathode, which means that they share the plasma ion load that would normally be taken by the single cathode in a d.c. or pulsed system.

In an inductively coupled system, the electric field is generated by the time variation of the magnetic field, and in the scheme shown in Fig. 10 (right) it has electric field lines that do not terminate on the conductor. Therefore the electric field does not drive ions into the conductor surface and reduces the damage to them.

The electric field is strongest near the coil, and has a strength given by

$$E \approx \frac{\mu_0 \pi N_t I r f}{L}, \tag{12}$$

where N_t is the number of turns, I is the peak current flowing in the coil, r is the coil radius, f is the frequency and L is the coil length. The electric field strength can reach several kilovolts per metre in the right conditions.

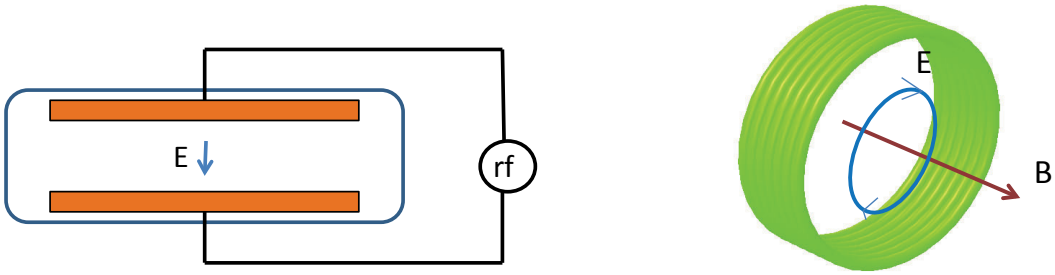


Fig. 10: Schematic of two excitation schemes using radio-frequency electromagnetic fields. Left: capacitive coupled system. Right: inductively coupled system.

RF sources typically use relatively low RF frequencies (from hundreds of kilohertz up to a few tens of megahertz) and typically they do not try to couple to any plasma resonance. Therefore, to a first approximation, they behave in a similar way to cathode–anode discharge sources but with a different electron acceleration mechanism.

An example of an RF discharge source using a solenoidal antenna is shown in Fig. 11.

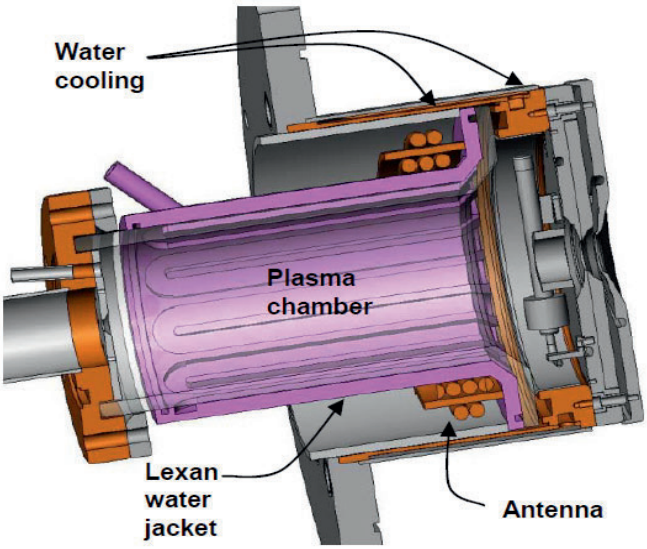


Fig. 11: An example of an RF discharge source, from the Spallation Neutron Source, for H^- production

4.4 Electron cyclotron resonance ion sources (ECRIS)

We have already learned that charged particles in a magnetic field make circular orbits in a direction perpendicular to the magnetic field, with a frequency (f_c) given by

$$f_c = \frac{1}{2\pi}\frac{qeB}{m}, \tag{13}$$

which is independent of the velocity of the particle (as long as relativistic energies are not reached). For electrons, this translates to an engineering formula of $f_c = 28$ GHz T^{-1}. Applying a strong magnetic field to a plasma (for example, 0.5 T with a pair of Helmholtz-type coils) leads to the electrons circulating with a frequency of 14 GHz.

The injection of microwaves of this same frequency into the plasma leads the electric field component of the wave to be in resonance with this electron orbit, and accelerates (or decelerates) the

electrons. This coupling of energy heats the electrons very efficiently, to the point where they are able to cause electron impact ionization of the injected gas, or further ionize ions in the plasma.

As they are heated, the process of collisional drift allows the heated electrons to be more efficiently contained in the plasma, which further benefits the creation of energetic electrons.

ECRIS have therefore become a very popular type of ion source, capable of delivering very different types of beams. Lower-frequency versions (typically 2.45 GHz frequency, often using permanent magnets to provide the required 87.5 mT resonance field) are well suited to producing beams of light ions, including high-intensity and high-duty-factor pulsed beams – one such source developed by CEA Saclay [6] is shown in Fig. 12. High-frequency versions are capable of producing beams of high-charge-state ions.

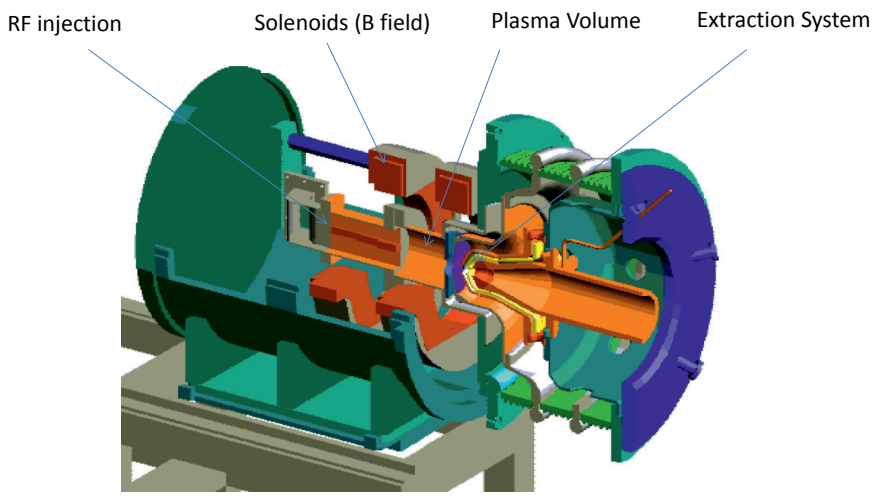


Fig. 12: Layout of the SILHI 2.45 GHz ECRIS from CEA

4.5 Laser ion sources

Lasers allow us to produce high-brightness, monochromatic beams of photons that can be used in two ways to produce ions, which are explained in the following sections.

4.5.1 Laser ionization ion source

Photons directed towards atoms, with the photon energy equal to or greater than the ionization potential, can directly ionize the atom. However, calculating the wavelength (λ) of the light required for this process,

$$\lambda = \frac{hc}{e\Phi_{\mathrm{i}}} = \frac{1.24}{\Phi_{\mathrm{i}}}\ \mu\mathrm{m}, \tag{14}$$

shows that, even for the element with the lowest ionization potential, francium at 3.83 eV, the longest suitable wavelength of 324 nm is already far into the ultraviolet (UV) part of the spectrum, and even this element requires a wavelength only available from a few laser types.

Therefore the ionization process is usually made using multiple photons of different wavelengths, where one beam of laser light raises an electron into an excited state, which then allows the ionization process to take place with a much lower-energy photon, shown schematically in Fig. 13. The use of these element specific energy levels allows a high degree of selectivity of the type of ion produced, as the excitation of atoms between two energy levels requires a very precise photon energy.

In this case, only one element will be put into the excited state, and only these excited atoms will be ionized.

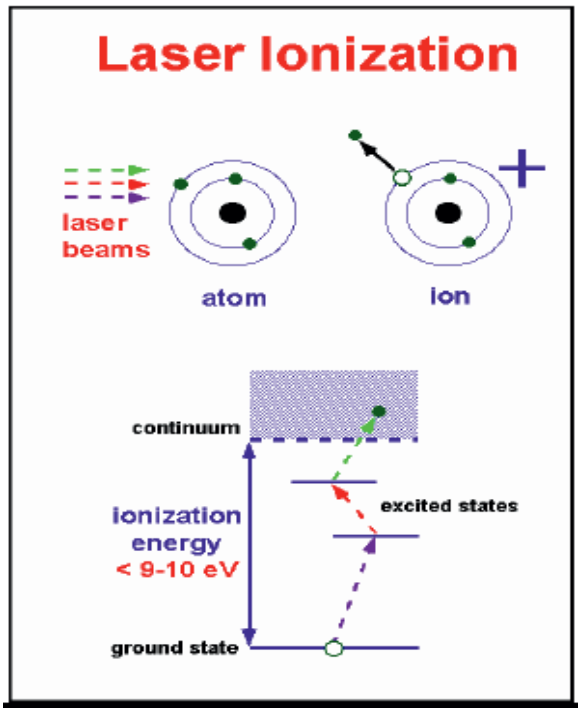


Fig. 13: Scheme for ionizing an atom using several laser wavelengths

This type of ion source has therefore lent itself very well to the radioactive ion beam community, where the selectivity based on element type is a complement to the mass spectrometry to allow a very high level of isotope selectivity. This has been taken even further by finding excitation levels whose energy is dependent on the mass of the radioactive nucleus, therefore allowing the isotope selectivity to be made even within the source.

4.5.2 Laser plasma ion sources

With a laser plasma ion source, the intense radiation from a laser is used to produce a dense, and sometimes hot, plasma, in which electron impact ionization is used to produce ions.

As many lasers can be used in a pulsed mode to deliver high instantaneous powers, and the beam of photons can be focused to a very small size, this means that very high power densities can be reached. When directed to a solid surface, these high power densities easily cause the surface to vaporize. In the transition from the density of the solid (approximately 10^{23} atoms/cm^3) to the low atomic density of vacuum, at some point the density will lead to a plasma electron oscillation frequency that is the equal to the laser frequency.

The relationship between the original laser light wavelength (λ) and the electron density at which this resonance occurs is given by

$$\lambda = \frac{2\pi c}{e}\sqrt{\frac{m_e \varepsilon_0}{n_e}}, \tag{15}$$

where m_e and n_e are the electron mass and electron number density. This resonant condition is ideal for strongly heating the plasma.

Atoms ablated from the solid surface pass through the plasma region and are ionized. In this case the ion type is changed easily through modification of the target material, allowing a large range of ions to be produced from this source type.

4.6 Surface ion sources

Within a solid material, electrons are present with a range of different energy states – Pauli's exclusion principle forces each electron to have a different wavefunction, and therefore a different energy level or spin. This means that electrons populate higher and higher energy levels within the material.

The temperature of the material causes the distribution of the populated energy levels by electrons to be modified, and some electrons to find higher energy levels, which may even allow the electrons to move to a high enough energy to escape the material (and become a thermionic electron emitted from a cathode).

If a layer of atoms coats this material surface, their outer electrons can populate the energy levels of the material; and if the atoms are desorbed from the surface (by heating or bombardment), a fraction will be liberated in a charged state. The Saha–Langmuir equation is used to calculate the ratio of the number of atoms desorbed as ions (n_i) and neutrals (n_0) as a function of the base material work function (ϕ_s) and the desorbed atom's first ionization potential (ϕ_i):

$$\frac{n_i}{n_0 + n_i} = \left(1 + \frac{g_0}{g_i} e^{(\phi_i - \phi_s)/kT}\right)^{-1}, \tag{16}$$

where T is the temperature and g_0/g_i is the ratio of the statistical weights (or degeneracy) of the neutrals to ions. This ratio is plotted in Fig. 14 for two temperatures, and the ratio of desorbed ions is circled on the plot for various elements. It can be seen that ions can still be desorbed even when the ionization potential is higher than the material work function, especially if the material is heated to higher temperatures.

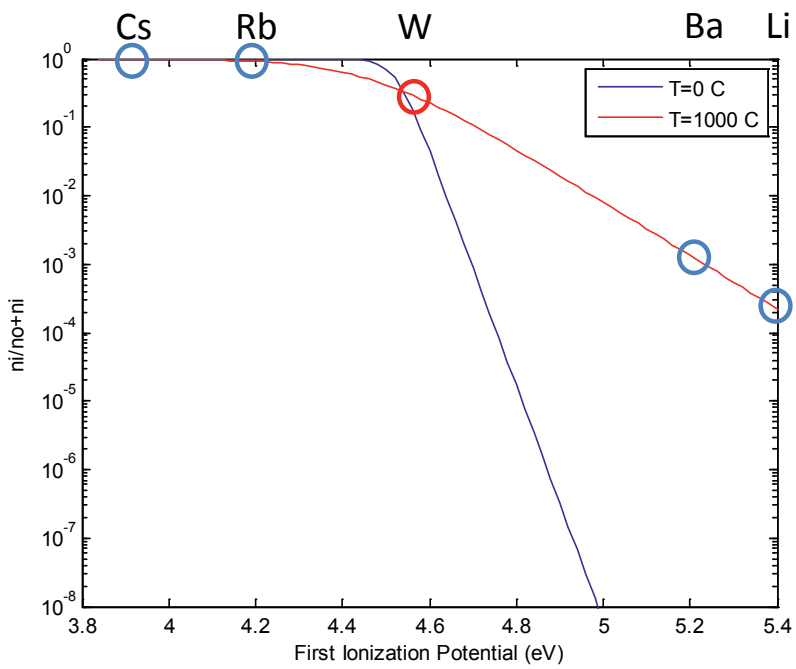


Fig. 14: Fraction of particles emitted from a tungsten surface as ions for two different temperatures. The ratios for several different elements are marked.

The design of the source is therefore relatively simple, with a tube of a high-work-function material heated to high temperatures, and injected with the material to be ionized, and is shown schematically in Fig. 15. In some applications, this type of source can be simply a hot filament.

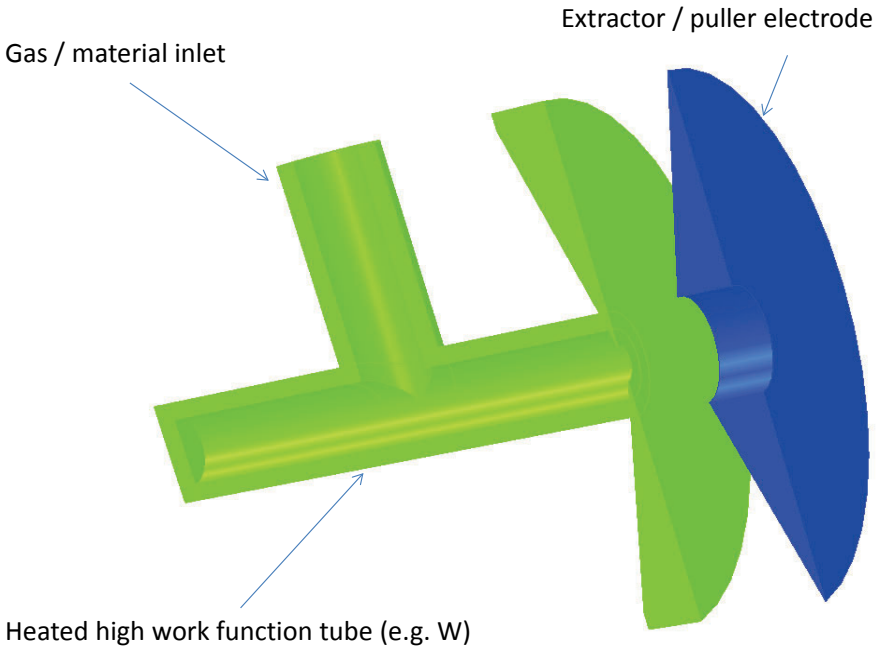


Fig. 15: Scheme of a surface ion source

4.7 Charge exchange ion sources

Negatively charged ions are possible for a range of elements, especially hydrogen and the halogens (fluorine, chlorine, bromine, etc.), which are said to have a positive electron affinity (i.e., the electron can be bound to the atom, and energy is required to release it). The binding energy is typically small (0.75 eV for H^-) and normally the cross-section for destruction is therefore high.

The charge exchange ion source uses the transfer of electrons directly between ions and neutral atoms and molecules. For example, Fig. 16 shows the cross-section for projectile protons (H^+) to capture directly two electrons from an H_2 molecule as a target, in order to produce H^- ions. This cross-section peaks around 20 keV energy for the projectile H^+ ion.

There is no need to have the same type of ion as the projectile and target, and alkali-metal vapours usually provide a much better ratio for electron capture towards the projectile, compared to the cross-section for stripping the electrons off the projectile. Therefore, it can be possible to convert a high fraction (10% or more of some elements, see Fig. 16) into negative ions.

Hence the charge exchange ion source (shown schematically in Fig. 17) firstly requires an ion source for positive ions, which can be of many different types (again depending on the application); it then passes the beam through a charge exchange cell, which should contain the gas or alkali vapour that will serve as the electron donor.

These source types also lend themselves to the production of polarized ions, where the metal vapour ions can be optically polarized, and during the charge transfer this polarization is transferred to the ion and then to the nucleus.

More information about polarized ion sources can be found in Ref. [10].

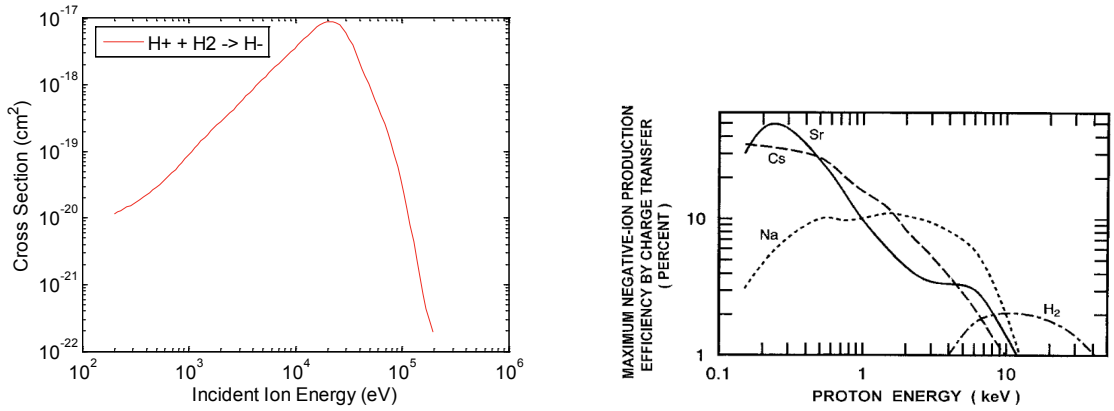


Fig. 16: Left: cross-section for double charge transfer of $H^+ + H_2$ to H^- ions, as a function of incident H^+ energy; data from Ref. [7]. Right: maximum conversion fraction of H^+ to H^- ions, as a function of proton energy and target type; reproduced from Ref. [8].

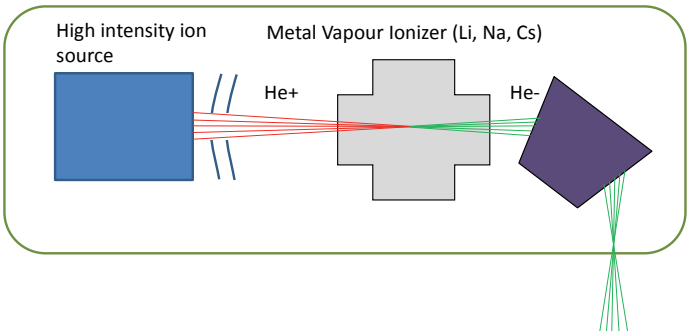


Fig. 17: Scheme of a charge exchange-type source; scheme derived from Ref. [9]

5 Methods for negative ion production

In section 4.7 it was seen how negative ions can be produced by the transfer of electrons from a gas or vapour to a beam of ions. For hydrogen, there is a range of sources able to deliver negative ions, even with high intensities. They rely on two main physical processes, namely, plasma volume production, and surface production.

The extraction of negative ions is also complicated by the fact that the electric field also extracts electrons, which, although they are accelerated quickly, can be present in sufficient numbers to cause a large increase in the space charge from the source.

5.1 Volume production

There are many ways for an additional electron to be attached to a hydrogen atom in a plasma. For example, a free plasma electron can be directly attached; and a charge exchange can occur between two hydrogen atoms, leading to one proton and one H^-. But the cross-sections for these processes are small, and, for a given plasma electron temperature, always have a higher cross-section for the destruction of the H^-.

There is a particular process where the rate of creation of H^- is higher than the destruction rate, which is by electron attachment of low-energy electrons to an excited H_2 molecule, which then disassociates into a neutral H and an H^-. In order to create the excited molecules, a higher electron

temperature is needed to create the excited H_2, but this zone can be separated from the H^- production area by a magnetic field, which as a result of collisional drift (see section 3.3) will allow lower-energy electrons to pass more quickly.

The plasma is produced and heated by several possible methods, including low-frequency RF as well as cathode and gas discharges (for example, in Penning sources).

5.2 Surface production

The production of positive ions from surfaces was discussed in section 4.7, where it was explained that an electron can be captured by a surface as an atom is desorbed from the surface. If the work function of the surface is very small, it is also possible that a desorbed atom has an additional electron attached.

For the best effect, the work function of the surface should be close to the binding energy of the electron to the atom, which for hydrogen would mean using a work function close to 0.75 eV. A coating of an alkali metal – the best being caesium – on a metal surface is the way to produce some of the lowest work functions, particularly if the layer is very thin (for the lowest possible work function, the layer of Cs should be less than one atom thick). A hydrogen plasma is then used to coat the caesium layer with hydrogen, and then bombard it with ions, which desorb these hydrogen atoms from the surface, with a high probability of them being negatively charged.

For consistent operating conditions, the source requires that the surface work function be kept reasonably constant, which is a challenge in the environment of an ion source. In addition, the heavy caesium atoms are themselves easily ionized if they are in the plasma, and they can cause significant sputtering of any cathode electrodes in the source.

Using caesium in the ion source usually leads to increased intensities, as well as decreasing the density of electrons at the extraction region, reducing the intensity of this co-extracted beam.

5.3 Highly charged ions

Highly charged ions are of interest for accelerators because they allow more energy gain for the same electric field. This is most interesting when the beam is of low intensity, where only a small fraction of the power put into an accelerator is transferred to the beam itself.

The production of highly charged ions cannot be normally made in a single step ($A^0 \rightarrow A^{q+}$), either by electron impact ionization or photoionization, as the electron or photon will need to have enough energy to liberate all the atomic binding energies. So logically the method is to continue to bombard ions with electrons, removing one electron at a time, called stepwise ionization. In this case the highest electron energy needs only be sufficient to cause the ionization of the highest charge state, and the ions need to be kept in the plasma sufficiently long for them to have enough collisions to allow them to reach the desired charge state.

The ionization potentials for all ions and charge states have been calculated by Carlson, Nestor, Wasserman and McDowell [11], and the cross-sections for ionization by electron impact from a charge state q to $q + 1$ ($\sigma_{q\rightarrow q+1}$) can be estimated for all charge states and ions from Lotz's formula [12].

As the electron energy requirements are therefore known to reach the highest charge state, and the cross-sections can then be estimated, we can use the following formula to link to the electron density (n_e) and the time required to reach the wanted charge state (τ):

$$n_e \tau = \frac{1}{v_e} \sum \frac{1}{\sigma_{q\rightarrow q+1}}, \tag{17}$$

where v_e is the average electron velocity. We see that it is the product of $n_e \tau$ that is important, i.e., we need high electron densities or a long time to reach the high charge states.

The three sources that generally fit these requirements for producing highly charged ions (i.e., ions where the last ionization potential is a few hundred volts) are the following:

- *Electron cyclotron resonance ion sources.* Electrons can be heated to high energies, and ions are trapped in the potential well of the electrons for many milliseconds.
- *Electron beam ion sources.* Electrons are produced by a cathode, the gun accelerates them to the energy required for the highest ionization, and this beam is focused. Ions are trapped by the potential of the electron beam, as well as electric fields applied by electrodes.
- *Laser plasma ion sources.* A very hot dense plasma is created by laser heating of a surface and then the plasma. Ions travel through the plasma quickly, but, owing to the high electron density in the plasma, they can undergo sufficient collisions to become highly ionized within a few nanoseconds.

It should not be forgotten that the counterpart of ionization is recombination, for which the cross-sections increase as the ion charge state increases. Of particular importance are recombination through charge exchange with other ions and atoms (avoided by keeping the background pressure low in the source) and three-body recombination, where one electron is recombined while another takes away the excess energy. This three-body recombination generally works against very dense plasma sources.

6 Conclusions

Within this chapter we discovered some of the fundamental processes that allow the production of ionized atoms, and saw how they are applied to ion sources. Depending on the main mechanism used for producing the ion, we find that there are different classifications of ion sources.

Ultimately the application should define the type of ion source used, where the source will be chosen that has the best compromise for plasma properties for the desired ion beam properties, as well as the intensity and pulse characteristics, simplicity of implementation and cost.

References

[1] B. Wolf (Ed.), *Handbook of Ion Sources* (CRC Press, Boca Raton, FL, 1995).

[2] R. Rejoub, B.G. Lindsay, and R.F. Stebbings, *Phys. Rev.* **A65** (2002) 042713.

[3] Y.-K. Kim and M.E. Rudd, *Phys. Rev.* **A50** (1994) 3954.

[4] R.L. Gill and A. Piotrowski, *Nucl. Instrum. Methods Phys. Res.* **A234**(2) (1985) 213.

[5] W. Eckstein, Sputtering yields, in *Sputtering by Particle Bombardment* (Topics in Applied Physics, Vol. 110, Springer, Berlin, 2007), pp. 33–187.

[6] M. Delaunay *et al.*, High current proton and deuteron ECR source developments at CEA, in *EPAC 96, Proc. Fifth European Particle Accelerator Conference, Barcelona, Spain* (Institute of Physics, Bristol, 1996), THP010L.

[7] C.F. Barnett (Ed.), *Atomic Data for Fusion*, Vol. 1, *Collisions of H, H_2, He, and Li Atoms and Ions with Atoms and Molecules*, ORNL-6086 (Controlled Fusion Atomic Data Center, Oak Ridge, TN, 1990).

[8] A.S. Schlachter, *AIP Conf. Proc.* **111** (1984) 300–332.

[9] N. Tanaka *et al.*, *Plasma Fusion Res.* **2** (2007) S1105.

[10] A. Zelenski, *Rev. Sci. Instrum.* **81** (2010) 02B308.

[11] T.A Carlson *et al.*, *At. Data Nucl. Data Tables* **2** (1970) 63–99.

[12] W. Lotz Z. Phys. 216, (1968) p. 24.

Introduction to Transverse Beam Dynamics

B.J. Holzer
CERN, Geneva, Switzerland

Abstract
In this chapter we give an introduction to the transverse dynamics of the particles in a synchrotron or storage ring. The emphasis is more on qualitative understanding rather than on mathematical correctness, and a number of simulations are used to demonstrate the physical behaviour of the particles. Starting from the basic principles of how to design the geometry of the ring, we review the transverse motion of the particles, motivate the equation of motion, and show the solutions for typical storage ring elements. Following the usual treatment in the literature, we present a second way to describe the particle beam, using the concept of the emittance of the particle ensemble and the beta function, which reflects the overall focusing properties of the ring. The adiabatic shrinking due to Liouville's theorem is discussed as well as dispersive effects in the most simple case.

1 Introduction

The transverse beam dynamics of charged particles in an accelerator describes the movement of single particles under the influence of external transverse bending and focusing fields. It includes the detailed arrangement of the accelerator magnets used (for example, their positions in the machine and their strength) to obtain well-defined and predictable parameters of the stored particle beam. It also describes methods to optimize the single-particle trajectories as well as the dimensions of the beam as an ensemble of many particles. A detailed treatment of this field in full mathematical lucidity and including sophisticated lattice optimizations, such as the right choice of the basic lattice cells, the design of dispersion suppressors or chromaticity compensation schemes, is beyond of the scope of this basic overview. For further reading and for more detailed descriptions, therefore, we would like to refer the reader to more detailed explanations [1–4, 9, 10].

2 Geometry of the ring

Magnetic fields are used in general in circular accelerators to provide the bending force and to focus the particle beam. In principle, the use of electrostatic fields would be possible as well, but at high momenta (i.e., if the particle velocity is close to the speed of light), magnetic fields are much more efficient. The force acting on the particles, the Lorentz force, is given by

$$\boldsymbol{F} = q \cdot (\boldsymbol{E} + (\boldsymbol{v} \times \boldsymbol{B}))$$

As an example let us consider a magnetic B field of 1 T, which is a conservative size. The resulting Lorentz force acting on a high-energy particle ($v \approx c \approx 3 \times 10^8$ m/s) is

$$F \approx q \cdot 3 \cdot 10^8 \frac{\mathrm{m}}{\mathrm{s}} \cdot 1 \frac{\mathrm{Vs}}{\mathrm{m}^2}$$

$$F \approx q \cdot 300 \frac{\mathrm{MV}}{\mathrm{m}}$$

This corresponds therefore to an equivalent electric field of 300 MV m^{-1}, which is far beyond the technical limits, which are set by field breakdown and discharges. In high-energy storage rings, or, more precisely, as soon as the speed of the particles is large enough, magnetic fields are much more efficient in accelerators. It has to be pointed out, however, that for low-energy heavy-ion rings situations might occur where the speed is low enough to make electrostatic fields more efficient.

2.1 The ideal circular orbit

We describe the transverse movement of the particles in a coordinate system rotating with the so-called ideal particle. As indicated in Fig. 1, the vertical coordinate, describing the displacement of a particle with respect to the ideal orbit, is called *y*, the corresponding horizontal one *x*, and the coordinate that is pointing into the direction of the longitudinal motion and that is moving with the particles around the ring is called *s*.

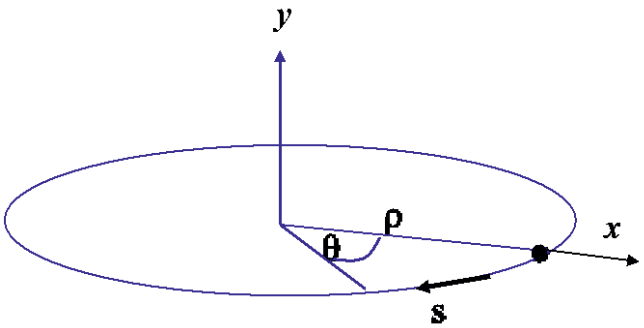


Fig. 1: Coordinate system used and orbit of an idealized particle

Assuming ideal conditions and a homogeneous dipole field to provide the bending force, the condition for a circular orbit is given by the equality between the Lorentz force and the centrifugal force. Neglecting any electrostatic field we get

$$q \cdot v \cdot B = \frac{mv^2}{\rho}$$

In a constant transverse magnetic field ***B***, a particle will see a constant deflecting force and the trajectory will be part of a circle, whose bending radius ρ is determined by the particle momentum $p = mv$ and the external B field:

$$B \cdot \rho = \frac{p}{q}$$

The term $B{\cdot}\rho$ is called the beam rigidity. Inside a dipole magnet, therefore, the bending angle (sketched in Fig. 2) is

$$\alpha = \frac{\int B ds}{B \cdot \rho}. \tag{1}$$

For the lattice designer, the integrated B field along the design orbit of the particles is one the most important parameters, as it is the value that enters Eq. (1) and defines the field strength and the number of such magnets that are installed. By requiring a bending angle of 2π for a full circle, we obtain a condition for the magnetic dipole fields in the ring. Figure 3 shows a photo of a small storage ring [5], where only eight dipole magnets are used to define the design orbit. The magnets are powered symmetrically, and therefore each magnet corresponds to a bending angle α of the beam of exactly 45°. The field strength B in this machine is of the order of 1 T.

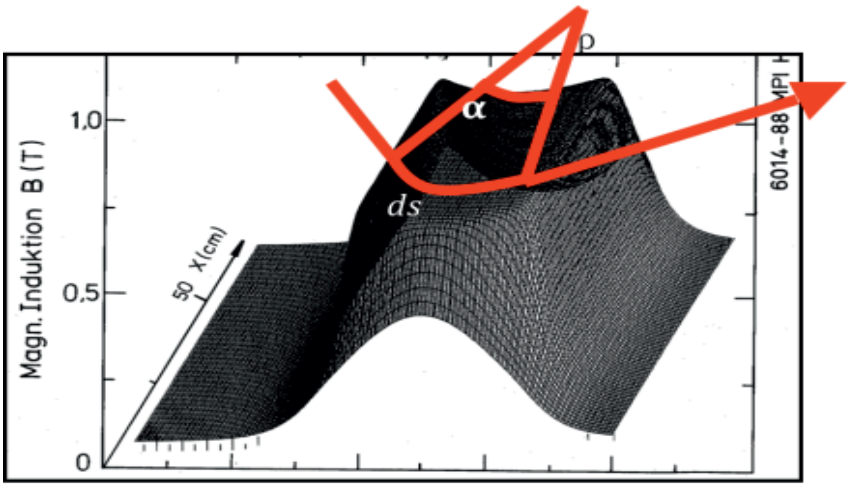


Fig. 2: Magnetic *B* field in a storage ring dipole and, schematically, the particle orbit

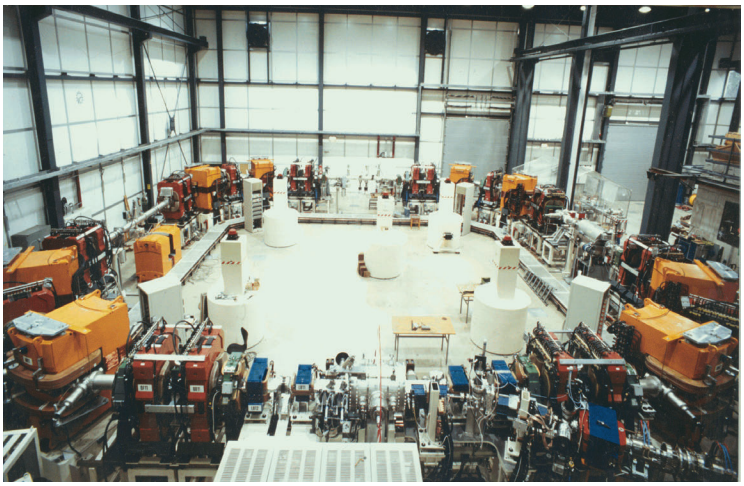


Fig. 3: The TSR heavy-ion storage ring at the Max-Planck-Institut in Heidelberg [5]

The lattice and, correspondingly, the beam optics in modern storage rings are usually split into several different characteristic parts. These include arc structures, which are used to guide the particle beam and to establish a regular pattern of focusing elements, leading to a regular, periodic beam dimension. These structures define the geometry of the ring and, as a function of the installed dipole magnets, the maximum energy of the stored particle beam. The arcs are connected by so-called insertions, long lattice sections where the optics is modified to establish the conditions needed for particle injection, to reduce the dispersion function, or to reduce the beam dimensions in order to increase the particle collision rate, for example, where the beam in a collider ring must be prepared for particle collisions.

In the case of the Large Hadron Collider (LHC) at CERN, for a momentum p = 7000 GeV/c, 1232 dipole magnets are needed in the arc to define the machine geometry, each having a length of 15 m and a B field of 8.3 T (Fig. 4). To express this absolutely exactly, the LHC, as any other storage ring, is not really a *ring* but rather a polygon – or, to be even more precise, a 1232-gon.

Fig. 4: The s.c twin-aperture dipole magnets of LHC

In general, we will try to cover a maximum part of the circumference by the main bending magnets, as they define the highest energy of the machine. For technical reasons, as a general rule about two-thirds of the circumference of the machine can be used to install the dipoles.

Requiring an overall bending angle of 2π for the complete dipole bending strength and approximating the integral of the B field by the sum over the magnets, we can calculate the required B field. In the case of the LHC we get

$$\oint B dl \approx N\, l\, B = 2\pi\, p/\mathrm{e}$$

$$B \approx \frac{2\pi \cdot 7000 \cdot 10^9\ \mathrm{eV}}{1232 \cdot 15\mathrm{m} \cdot 3 \cdot 10^8 \frac{\mathrm{m}}{\mathrm{s}} \cdot e} = 8.3\ \mathrm{T}$$

A part of the LHC tunnel is shown in Fig. 5, where the shape of the ring is hardly visible due to the large bending radius of $\rho = 2.5$ km.

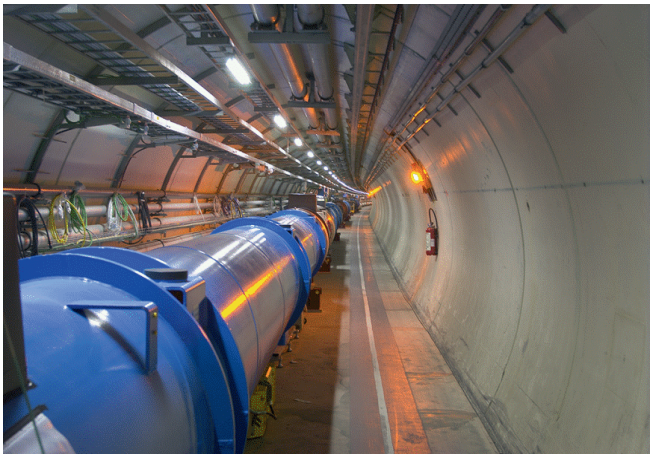


Fig. 5: The LHC tunnel with the dipole cryostats. The normalized bending force, which is represented by the radius of $\rho = 2.5$ km, is hardly visible.

3 Quadrupole magnets and equation of motion

Once the geometry and specification of the arc have been determined and the layout of the bending magnets has been done, the next step is to worry about the focusing properties of the machine. In general, we have to keep more than 10^{12} particles in the machine, distributed over a number of bunches, and these particles have to be focused to keep their trajectories close to the design orbit.

Gradient fields generated by quadrupole lenses are used to do this job. These lenses generate a magnetic field that increases linearly as a function of the distance from the magnet centre:

$$B_y = -g \cdot x\,, \;\; B_x = -g \cdot y$$

Here, x and y refer to the horizontal and vertical planes and the parameter g is called the gradient of the magnetic field. It is customary to normalize the magnetic fields to the momentum of the particles. In the case of dipole fields, we obtain from Eq. (1)

$$\alpha = \frac{\int B ds}{B \cdot \rho} = \frac{1}{\rho} \cdot L_{eff}$$

where L_{eff} is the so-called effective length of the magnet. The term $1/\rho$ is the normalized bending strength of the dipole. In the same way, the field of the quadrupole lenses is normalized to $B\rho$. The strength k is defined by,

$$k = \frac{g}{B \cdot \rho}$$

and the focal length of the quadrupole is given by

$$f = \frac{1}{k \cdot l}$$

As a rule of thumb, we get the normalized gradient k as the ratio between gradient g expressed in T m^{-1} and the particle momentum p in units of GeV/c and multiplying by 0.3 as an approximation of the speed of light:

$$k = 0.3 \frac{g(\frac{T}{m})}{p(\frac{GeV}{c})}$$

It is worth emphasizing that the gradient g will describe the focusing property of the quadrupole in both transverse planes. From Maxwell's equation,

$$\boldsymbol{\nabla} \times \boldsymbol{B} = \cancel{\boldsymbol{j}} + \cancel{\frac{\partial \boldsymbol{E}}{\partial t}} = 0$$

because at the position of the beam (i.e., inside the quadrupole bore) the density j of the magnet current is zero and no explicit changing electric field exists. We conclude that

$$\frac{\partial B_y}{\partial x} = \frac{\partial B_x}{\partial y}$$

To derive the equation of motion of the particles under the influence of the fields described above, we refer to the linear approximation. In the case of the fields this means that only constant terms or terms linearly dependent on x (or y) will be taken into account. Mathematically speaking, for a general Taylor expansion, all higher-order terms will be neglected:

$$\frac{\boldsymbol{B}(\boldsymbol{x})}{\boldsymbol{p}/\boldsymbol{e}} = \frac{1}{\rho} + \boldsymbol{k}\,\boldsymbol{x} + \cancel{\frac{1}{2!}\boldsymbol{m}\,\boldsymbol{x}^2} + \cancel{\frac{1}{3!}\boldsymbol{n}\,\boldsymbol{x}^3} + \ldots$$

This approximation is more serious than it might look like. Nonlinear terms would lead to an equation of motion that cannot be solved analytically any more. For the storage ring design, the consequence is to optimize the magnets and suppress higher multipoles in the dipole and quadrupole magnets as well as possible. Moreover, in general in modern accelerators, the magnets will be split and optimized according to their job: dipoles for bending, and quadrupoles for focusing. Multipoles like sextupole and octupole magnets will be introduced – if needed – carefully for higher-order corrections only.

In Fig. 3 we saw already an example of such a separate function storage ring [5] with eight dipole magnets and 4 × 5 quadrupole lenses for beam focusing.

3.1 Equation of motion

For the derivation of the equation of motion, we start with a general expression for the radial acceleration as known from classical mechanics:

$$a_r = \frac{d^2\rho}{dt^2} - \rho\left(\frac{d\theta}{dt}\right)^2$$

The first term refers to an explicit change of the bending radius, and the second one to the centrifugal acceleration. Referring to our coordinate system, and replacing for the general case the ideal radius ρ by $\rho + x$ (Fig. 6), we obtain for the balance between radial force and the counter-acting Lorentz force the relation

$$\boldsymbol{F = m\frac{d^2}{dt^2}(x+\rho) - \frac{mv^2}{x+\rho} = e\,B_y\,v}$$

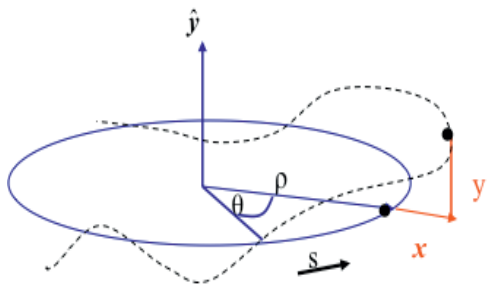


Fig. 6: Ideal circular orbit and real particle trajectory with its transverse coordinates x and y

On the right-hand side of the equation, we take only linear terms into account,

$$B_y = B_0 + x\frac{\partial B_y}{\partial x}$$

and for convenience we replace the independent variable t by the coordinate s,

$$x' = \frac{dx}{ds} = \frac{dx}{dt}\frac{dt}{ds}$$

to obtain an expression for the particle trajectories under the influence of the focusing properties of the quadrupole and dipole fields in the ring, described by a differential equation. This equation is derived in its full beauty elsewhere [6], so we shall just state it here:

$$x'' + Kx = 0. \tag{2}$$

Here, x describes the horizontal coordinate of the particle with respect to the design orbit; the derivative is taken with respect to the orbit coordinate s, as usual in linear beam optics; and the

parameter K combines the focusing strength k of the quadrupole and the weak focusing term $1/\rho^2$ of the dipole field. (Note that a negative value of k means a horizontal focusing magnet.) The value of K is given by

$$K = -k + 1/\rho^2.$$

In the vertical plane, in general, the term $1/\rho^2$ is missing, as in most (but not all) accelerators the design orbit is in the horizontal plane and no vertical bending strength is present. At the same time, the sign of the gradient changes due to the geometry of the quadrupole field lines, $k \to -k$. So, in the vertical plane, we have

$$K = k.$$

3.2 Single-particle trajectories

The differential equation (2) describes the transverse motion of a particle with respect to the design orbit. This equation can be solved in a linear approximation, and the solutions for the horizontal and vertical planes are independent of each other in the sense that the motions in the two transverse planes are uncoupled.

If the focusing parameter K is constant, which means that we are referring to a position inside a magnet where the field is constant along the orbit, the general solution for the position and angle of the trajectory can be derived as a function of the initial conditions x_0 and x_0'. In the case of a focusing lens, we obtain

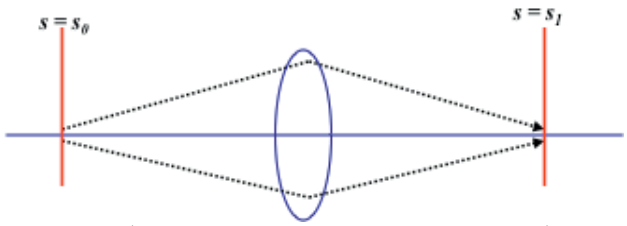
$$x(s) = x_0 * \cos(\sqrt{K} * s) + \frac{x_0'}{\sqrt{K}} * \sin(\sqrt{K} * s),$$
$$x'(s) = -x_0 * \sqrt{K} * \sin(\sqrt{K} * s) + x_0' * \cos(\sqrt{K} * s),$$

or, written in a more convenient matrix form,

$$\begin{pmatrix} x \\ x' \end{pmatrix}_s = M \cdot \begin{pmatrix} x \\ x' \end{pmatrix}_0 .$$

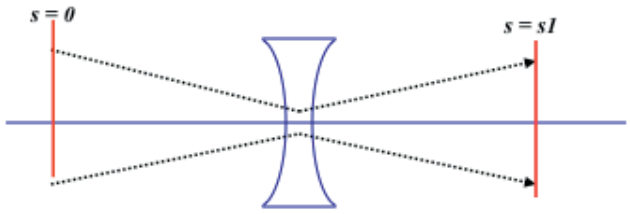
Given the particle amplitude and angle in front of the lattice element, x_0 and x_0', we obtain their values after the element by a simple matrix multiplication. The matrix M depends on the properties of the magnet, and we obtain the following expressions for three typical lattice elements. Also shown schematically are the situations for the three cases described.

Focusing quadrupole:



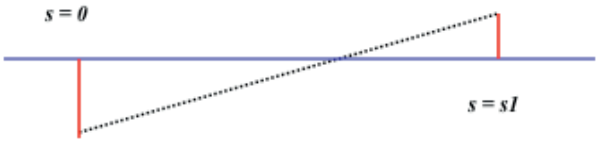
$$M_{QF} = \begin{pmatrix} \cos(\sqrt{|K|}l & \frac{1}{\sqrt{|K|}}\sin(\sqrt{|K|}l \\ -\sqrt{|K|}\sin(\sqrt{|K|}l & \cos(\sqrt{|K|}l \end{pmatrix} \tag{3a}$$

Defocusing quadrupole:



$$M_{QD} = \begin{pmatrix} \cosh(\sqrt{|K|}l & \frac{1}{\sqrt{|K|}}\sin\mathrm{h}(\sqrt{|K|}l \\ \sqrt{|K|}\sinh(\sqrt{|K|}l & \cos\mathrm{h}(\sqrt{|K|}l \end{pmatrix} \tag{3b}$$

Drift space:



$$M_{\text{drift}} = \begin{pmatrix} 1 & \ell \\ 0 & 1 \end{pmatrix} \tag{3c}$$

3.3 Transformation through a system of lattice elements

Given the matrix description for the lattice elements, it is evident that, to obtain the particle amplitude at a certain position in the storage ring, we can either start at an initial point and transform the vector (x, x') step-by-step through the machine, or – mathematically equivalent, but much more convenient – determine the product matrix between the two points of interest and apply it to the initial coordinates $(x, x')_0$. As an example, we refer to Fig. 7. For simplicity, we consider a storage ring built only out of focusing and defocusing quadrupoles and dipole magnets in between.

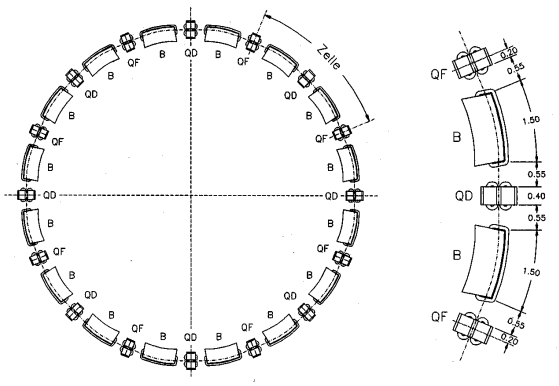


Fig. 7: Ideal storage ring consisting of focusing and defocusing quadrupoles and dipole magnets for bending (Courtesy of K. Wille [7]).

Starting in front of a focusing magnet, a typical part of this structure, expressed in matrix form, would be e.g.

$$M_{total} = M_{QF} * M_D * M_{QD} * M_{Bend} * M_{D*.....}$$

The particle coordinates at point s_2 can be obtained by repetitive multiplication of the corresponding single element matrices, or by applying the product matrix above:

$$\begin{pmatrix} x \\ x' \end{pmatrix}_{s2} = M(s_2, s_1) * \begin{pmatrix} x \\ x' \end{pmatrix}_{s1}$$

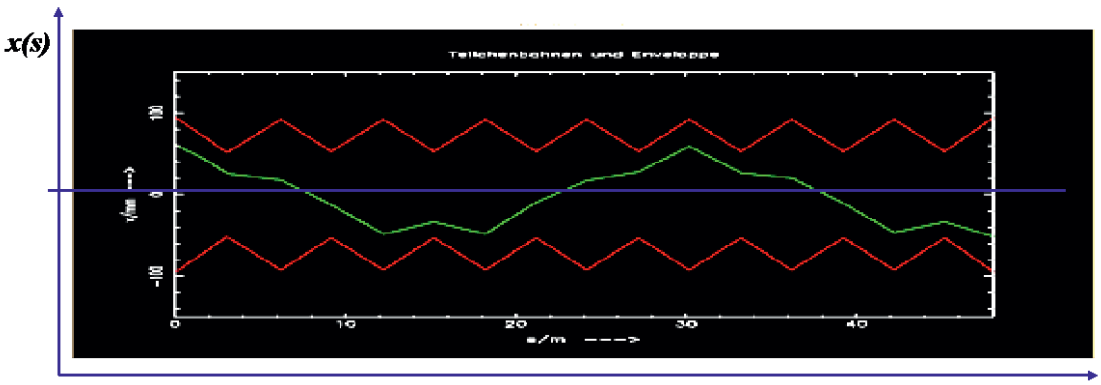


Fig. 8: Single-particle trajectory in a storage ring (green line). In red, the pattern of the beta function is shown, which can be interpreted as the maximum beam size or aperture needed in the example.

Using reasonable values for $1/\rho$ and k, we obtain the picture shown in Fig. 8 for the particle trajectory. The particle movement – looking zig-zag in the simplified model used here – represents in each single element the solution of a harmonic oscillator, each having eventually a different restoring force, still however leading to an overall focusing effect around the storage ring. The overall number of oscillations, or better the corresponding oscillation frequency of this transverse motion, is of major importance for beam stability reasons and represents the eigenfrequency of the motion, the so-called 'tune'. In the example of Fig. 8 we get $Q \approx 1.3$.

4 The Twiss parameters α, β and γ

Following the single-particle trajectory as shown in Fig. 8, the natural question will be how the trajectory of the second turn will look; and then the third one; and so on. Reproducing the calculations that lead to the orbit of the first turn will result in a large number of single-particle trajectories that will overlap somehow and form the beam envelope. Figure 9 shows the result for 50 turns. Clearly, as soon as we talk about many turns, or many particles, the use of the single-trajectory approach is quite limited and we need a description of the beam as an ensemble of many particles.

Fortunately, in the case of periodic conditions in the accelerator, there is another way to describe the particle trajectories that, in many cases, is more convenient than the above-mentioned formalism, which is valid within a single element. It is important to note that, in a circular accelerator, the focusing elements are necessarily periodic in the orbit coordinate s after one revolution. Furthermore, storage ring lattices have in most cases an inner periodicity: they are often constructed, at least partly, from sequences in which identical magnetic cells, the lattice cells, are repeated several times in the ring and lead to periodically repeated focusing properties.

In this case, the equation of motion has to be written now in a slightly different form:

$$x''(s) - k(s)x(s) = 0$$

Unlike the treatment above, the focusing parameters – or the restoring forces – are no longer constant, but are functions of the coordinate s. However, they are periodic in the sense that, at least after one full turn, they repeat each other, $k(s+L) = k(s)$, leading to the special so-called Hill differential equation. Following Floquet's theorem, the solution of this equation can be written in its general form as [2]

$$x(s) = \sqrt{\varepsilon} \cdot \sqrt{\beta} \cdot \cos(\psi(s) + \phi) \tag{4}$$

$$x'(s) = \frac{-\sqrt{\varepsilon}}{\sqrt{\beta(s)}} \cdot \sin(\psi(s) + \phi) + \alpha(s) \cdot \cos(\psi(s) + \phi)$$

The position and angle of the transverse oscillation of a particle at a point s are given by the value of a special β-function at that location, and ε and δ are constants of the particular trajectory. The β-function depends in a quite sophisticated manner on the overall focusing properties of the storage ring. It cannot be calculated directly in an analytical approach, but it has to be either determined numerically or deduced from properties of the single-element matrices described above (see e.g. [8]). In any case, like the lattice itself, it has to fulfil the periodicity condition

$$\beta(s+L) = \beta(s)$$

Inserting the solution (4) into the Hill equation and rearranging slightly, we get

$$\psi(s) = \int_0^s \frac{ds}{\beta(s)} \tag{5}$$

which describes the phase advance of the oscillation. It should be emphasized that ψ depends on the amplitude of the particle oscillation. At locations where β reaches large values, i.e., the beam has a large transverse dimension, the corresponding phase advance will be small; and vice versa, at locations where we create a small β in the lattice, we will obtain a large phase advance. In the context of Fig. 8 we introduced the tune as the number of oscillations per turn, which is nothing other than the overall phase advance of the transverse oscillation per revolution in units of 2π. So by integrating Eq. (5) around the ring, we get the expression

$$Q_y = \frac{1}{2\pi} \oint \frac{ds}{\beta(s)}$$

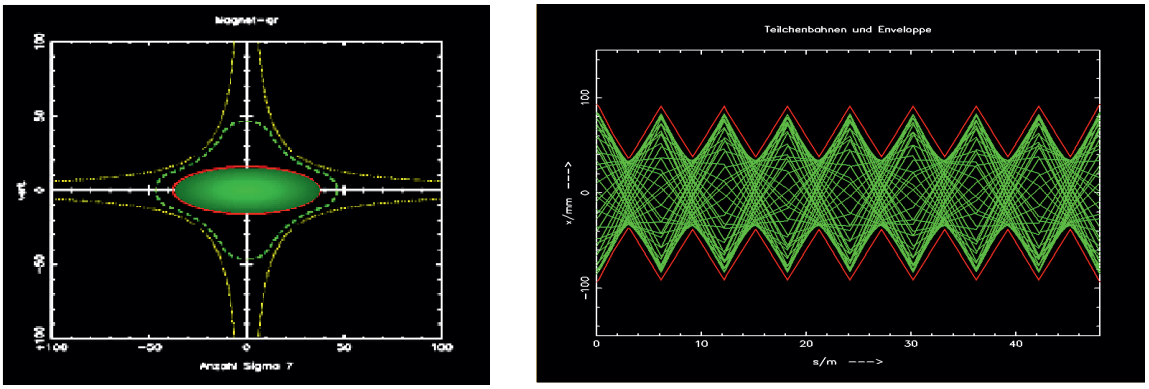


Fig. 9: Beam as an ensemble of many single-particle trajectories. Left: transverse shape of a typical particle beam. Right: overlapping of many particle trajectories.

The practical significance of the β-function is shown in Fig. 9. The left part shows schematically the transverse shape of the beam inside the vacuum chamber (red) and the hyperbolic pole shoe profile of the quadrupole lens. On the right-hand side the single-particle trajectories are shown. The envelope of the overlapping trajectories is given by $\hat{x} = \sqrt{\varepsilon\beta(s)}$ and is used to define the beam size in the sense of a Gaussian density distribution.

The integration constant ε has a clearly defined physical interpretation. Given the solution of Hill's equation

$$x(s) = \sqrt{\varepsilon}\,\sqrt{\beta(s)}\,\cos(\psi(s) + \phi) \tag{6}$$

and its derivative

$$x'(s) = -\frac{\sqrt{\varepsilon}}{\sqrt{\beta(s)}}\,\left\{\alpha(s)\cos(\psi(s) + \phi) + \sin(\psi(s) + \phi)\right\} \tag{7}$$

we can transform Eq. (6) to

$$\cos(\psi(s) + \phi) = \frac{x(s)}{\sqrt{\varepsilon}\,\sqrt{\beta(s)}}$$

and insert the expression into Eq. (7) to get an expression for the integration constant ε:

$$\varepsilon = \gamma(s)\,x^2(s) + 2\alpha(s)x(s)x'(s) + \beta(s)\,x'^2(s) \tag{8}$$

Here we follow the usual convention in the literature and introduce the two parameters

$$\alpha(s) = \frac{-1}{2}\beta'(s)$$
$$\gamma(s) = \frac{1+\alpha(s)^2}{\beta(s)} \tag{9}$$

We obtain for ε a parametric representation of an ellipse in x–x' 'phase' space, which according to Liouville's theorem is a constant of motion, as long as conservative forces are considered. The mathematical integration constant thus gains physical meaning.

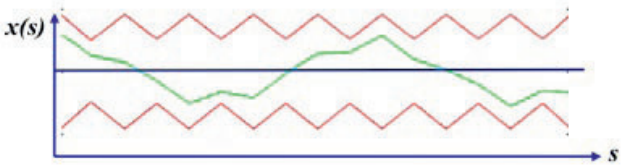


Fig. 10: A single-particle trajectory …

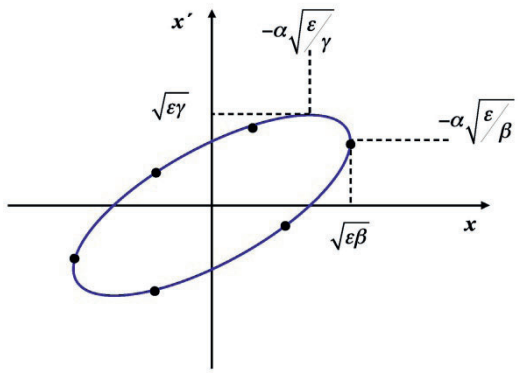


Fig. 11: … and its phase space coordinates at a position s turn by turn

Referring again to the single-particle trajectory, discussed above (Fig. 10), but now plotting at a given position s in the ring the coordinates x and x' turn-by-turn, we get the picture shown in Fig. 11. Plotted in phase space the particle will follow the shape of an ellipse whose shape and orientation are

defined by the optical parameters at the reference position s in the ring. Each point in Fig. 11 represents the coordinates for a certain turn at that position in the ring, and the particle will have performed from one turn to the next a number of revolutions in phase space that corresponds to its tune. It has to be emphasized that, as long as conservative forces are considered (i.e., no interaction between the particles in a bunch, no collisions with remaining gas molecules, no radiation effects, etc.), the size of the ellipse in x–x' space is constant and can be considered as a quality factor of the single particle. Large areas in x–x' space (where, to be exact, the area is given as $A = \pi\varepsilon$) mean large amplitudes and angles of the transverse particle motion, and we would consider it as a bad particle 'quality'.

To discuss the dependence of the shape of the ellipse on the beam optics in a bit more detail, we replace the two parameters α and γ by their corresponding expressions in Eq. (9),

$$\varepsilon = \frac{x^2}{\beta} + \frac{\alpha^2 x^2}{\beta} + 2\alpha \cdot xx' + \beta \cdot x'^2$$

solve for x'

$$x'_{1,2} = \frac{-\alpha \cdot x \pm \sqrt{\varepsilon\beta - x^2}}{\beta}$$

and determine the maximum values of the particle angle $\hat{x}'$ via

$$\frac{dx'}{dx} = 0$$

to obtain for the maximum angle x' and the position x at that point:

$$\hat{x}' = \sqrt{\varepsilon\gamma}\,, \qquad x = \pm\alpha\sqrt{\varepsilon/\gamma}$$

For completeness and to demonstrate the conservation of phase space area, we plot in Fig. 12 the emittance invariants of four particles in phase space. For the position s this time the centre of a focusing quadrupole has been chosen, where the orientation of the ellipses is flat, due to the fact that at this location the beam will have its largest size, i.e., $\alpha = 0$. Turn by turn the coordinates of the four particles will lie on the corresponding ellipses and never cross each other.

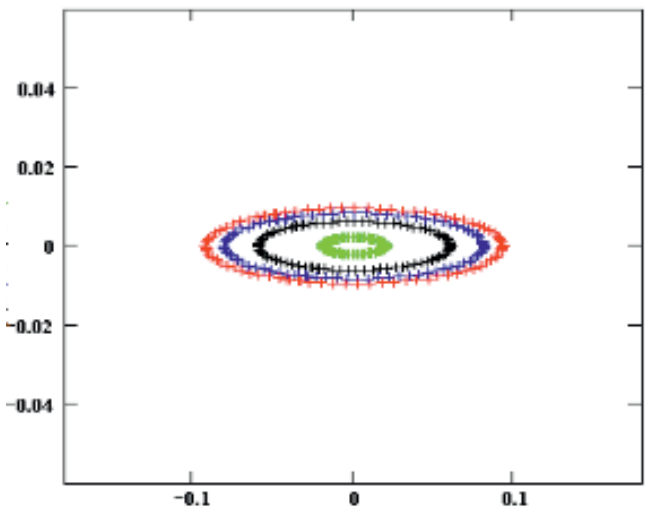


Fig. 12: Phase space ellipses at the location of a focusing quadrupole for four particles

4.1 The emittance of a particle ensemble

The ensemble of many single particles, as shown in Fig. 13, forms a pattern of overlapping trajectories that in the end we will observe as transverse intensity (or charge) distribution and that we will use to define the beam size. Along the storage ring coordinate s, its transverse size, in the sense of the maximum amplitude of a trajectory that will be observed at a given location, is defined by the β-function.

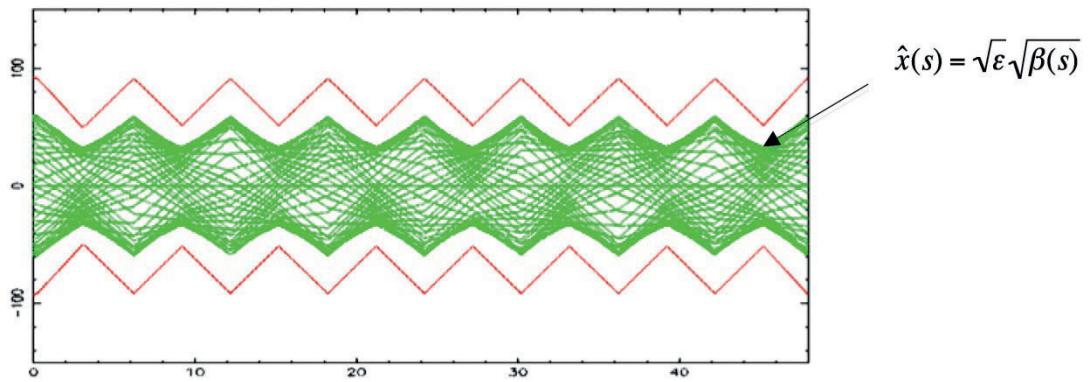


Fig. 13: The overlapping trajectories of many single particles define the beam cross-section as defined by ε and β.

As the general solution of the equation of motion is given by

$$x(s) = \sqrt{\varepsilon}\sqrt{\beta(s)} \cdot \cos(\Psi(s) + \phi)$$

the maximum amplitude, or beam size, is obtained from

$$\hat{x}(s) = \sqrt{\varepsilon}\sqrt{\beta(s)}$$

and depends via the β-function on the focusing properties of the lattice and via ε on the quality of the particle ensemble.

In many cases, the transverse particle density of the particles follows a Gaussian distribution. Referring to a particle within this distribution that is situated at one standard deviation σ, the ε parameter (sometimes called the Courant–Snyder invariant) of this particle can be used as representative of the complete beam. In this sense we talk about a beam emittance and thus about a general quality factor of the whole particle ensemble. Referring to the example of the LHC, for example, we have, in the arc of the storage ring at flat-top energy, the following values:

$$\beta = 180 \text{ m}$$

$$\varepsilon = 5 \cdot 10^{-10} \text{ m} \cdot \text{rad}$$

and the beam size (1σ) is

$$\sigma = \sqrt{\varepsilon \cdot \beta} \cdot \sqrt{180 \text{ m} \cdot 5 \cdot 10^{-10} \text{ m} \cdot \text{rad}} = 0.3 \text{ mm}$$

Figure 14 shows the result of a measurement of the transverse beam size. The points represent the measurement values, the curve a Gaussian fit, to obtain the sigma of the distribution as a number to qualify the beam dimension. In general, we will define the aperture dimensions of the vacuum chamber as a certain multiple of this beam sigma. In the case of the LHC, for example, we require a minimum aperture of $r_0 = 18\sigma$ inside the mini-beta quadrupoles.

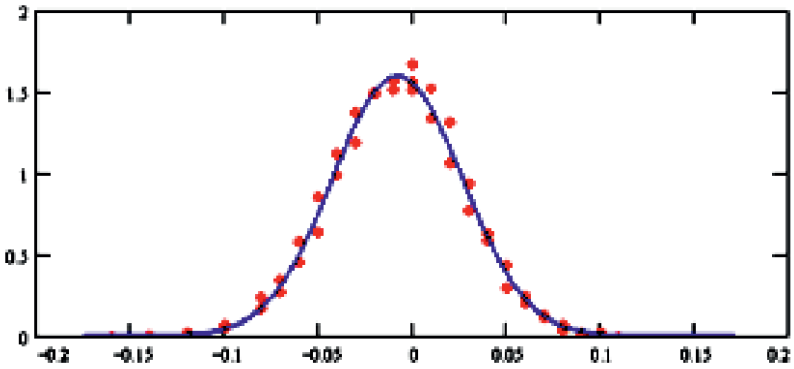


Fig. 14: Measured transverse particle density in a beam and a Gaussian fit to obtain the σ of the distribution

5 Liouville's theorem during acceleration

There is a special issue concerning the phase space area of a particle beam as soon as we start to accelerate. It has been explained that the expression of the beam emittances

$$\varepsilon = \gamma(s)\, x^2(s) + 2\alpha(s)x(s)x'(s) + \beta(s)\, x'^2(s)$$

represents an ellipse in the x–x' space, which we call in a rather sloppy manner 'phase space'.

The real phase space, however, as defined in classical mechanics, relates the position x with its canonical conjugate momentum p_x. It is the area in this real phase space that is conserved according to Liouville's theorem. So if

$$\int p_x dx = const$$

and rewriting for the angle

$$x' = \frac{dx}{ds} = \frac{dx}{dt}\frac{dt}{ds} = \frac{\beta_x}{\beta}$$

which is the ratio of the relativistic β-parameter between the transverse motion and the longitudinal one, we get from Liouville's theorem

$$\int pdq = mc\gamma\beta \underbrace{\int x'dx}_{\varepsilon} = const$$

with the expression in the integral being our 'phase space'. Still, it is true that the beam emittance is constant for a given energy, but as soon as we start to accelerate ε will shrink as

$$\varepsilon = \int x'dx \propto \frac{1}{\beta\gamma}$$

where γ and β are the relativistic parameters and should not be confused with the beam optics functions.

As a consequence, it turns out that a proton machine or an electron linac will need the highest aperture at injection energy. As soon as we start to accelerate, the beam size shrinks as $\gamma^{-1/2}$ in both transverse planes. At lowest energy the machine will have major aperture problems, and here we have to minimize the beta function to obtain a beam envelope that fits into the vacuum chamber. At high energy this aperture restriction is more relaxed and we can develop optics solutions that allow for higher values of β. For example, a mini-beta concept that, as we will see, leads to extreme beta values

in the focusing magnets will only be adequate at flat top. An example is shown in Fig. 15 for the case of the HERA proton storage ring. The particles were injected into this machine at an energy of 40 GeV and after acceleration reached a flat-top energy of 920 GeV. For the same beam optics, the figure compares the situation at injection energy and at flat top. Owing to the factor of 23 in energy increase, the beam size shrinks by nearly a factor of 5 in both planes.

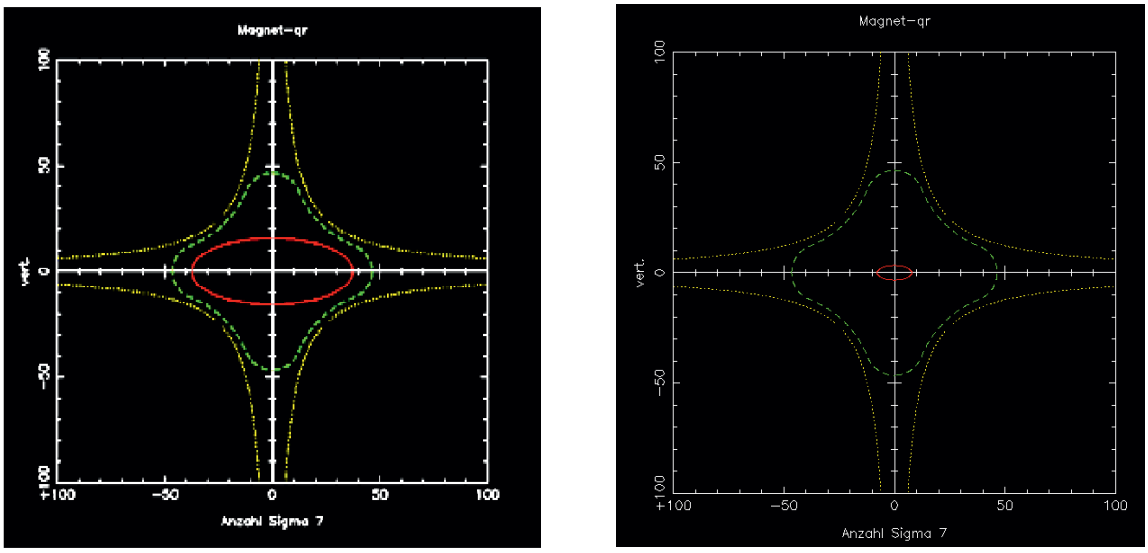


Fig. 15: Transverse beam envelope for design parameters of the HERA proton ring at injection and flat top. For the same optics, the beam size shrinks considerably during the acceleration.

In the case of the LHC, the two different beam optics optimized following these needs are shown in Fig. 16. While at injection energy of 450 GeV the maximum beta values have to be limited to 500 m, much higher values are possible at high energy, and close to the mini-beta insertions a maximum of $\beta = 4.5$ km is possible.

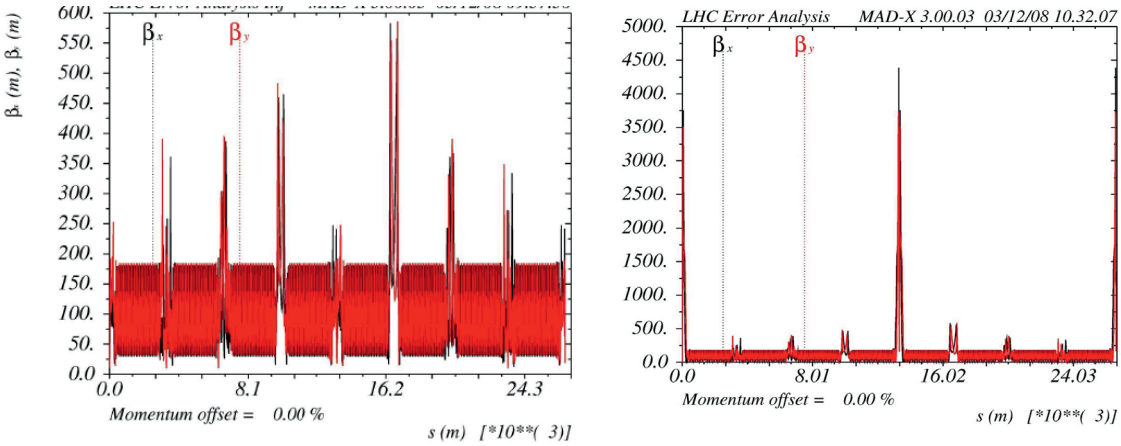


Fig. 16: LHC beam optics optimized for injection (left) and luminosity (right). The large beta functions needed for luminosity operation can only be applied at flat-top energy where the beam emittance is considerably reduced.

6 Dispersive effects

Until now we have treated the beam and the equation of motion as a mono-energetic problem. Unfortunately, in the case of a realistic beam, we have to deal with a considerable distribution of the particles in energy or momentum. Typical values are

$$\frac{\Delta \boldsymbol{p}}{\boldsymbol{p}} \approx 1.0\ 10^{-3}$$

This momentum spread will lead to several effects concerning the bending of the dipole magnets and the focusing strength of the quadrupoles. It turns out that the equation of motion, which was a homogeneous differential equation until now, will get a non-vanishing term on the right-hand side:

$$x'' + x(\frac{1}{\rho^2} - k) = \frac{\Delta p}{p} \cdot \frac{1}{\rho}$$

The general solution therefore is the sum of the solution of the homogenous equation of motion plus a special solution of the inhomogeneous one:

$$x(s) = x_h(s) + x_i(s)$$

Here x_h is the solution that we have discussed until now and x_i is an additional contribution that has yet to be determined. For convenience, we usually normalize this second term and define a function, the so-called dispersion:

$$D(s) = \frac{x_i(s)}{\Delta p / p}$$

This describes the dependence of the additional amplitude of the transverse oscillation on the momentum error of the particle. In other words it fulfils the condition

$$x_i''(s) + K(s) \cdot x_i(s) = \frac{1}{\rho} \cdot \frac{\Delta p}{p}$$

The dispersion function is usually calculated by optics programs in the context of the calculation of the usual optical parameters and is of equal importance. Analytically it can be determined for single elements via

$$D(s) = S(s) \int_{s0}^{s1} \frac{1}{\rho} C(\tilde{s}) d\tilde{s} - C(s) \int_{s0}^{s1} \frac{1}{\rho} S(\tilde{s}) d\tilde{s}$$

where $S(s)$ and $C(s)$ correspond to the sine-like and cosine-like elements of the single-element matrices in the lattice [6].

Typical values in the case of a high-energy storage ring are

$$x_\beta = 1\text{–}2 \text{ mm}, D(s) = 1\text{–}2 \text{ m}$$

and, for a typical momentum spread of $\Delta p / p \approx 10^{-3}$, we obtain an additional contribution to the beam size from the dispersion function that is of the same order as the one from the betatron oscillations x_β. An example of a high-energy beam optics including the dispersion function is shown in Fig. 17.

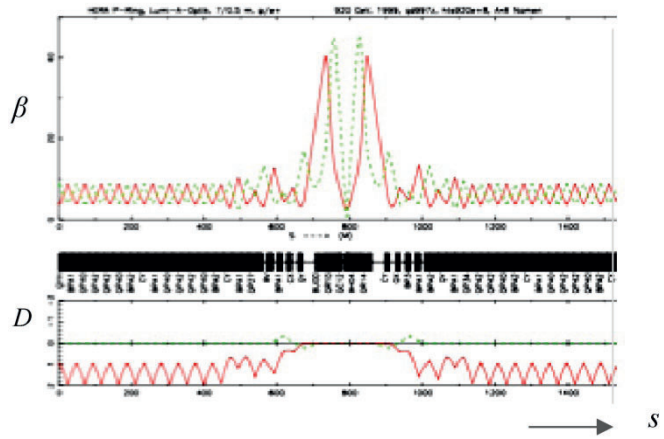


Fig. 17: Beam optics function β(s) (upper part) and dispersion function D(s) (lower part) for a part of a typical collider ring. The dispersion must vanish at the collision point in the middle of the horizontal axis.

It should be pointed out that the dispersion describes that special orbit that an ideal particle would have in the absence of no betatron oscillations ($x_\beta = x_\beta' = 0$) for a momentum deviation of $\Delta p/p = 1$. Still, it describes 'just another particle orbit' and so it is subject to the focusing forces of the lattice elements, as seen in Fig. 17.

7 Mini-beta insertions and luminosity

The straight sections of a storage ring are often designed for the collision of two counter-rotating beams. For a given overall cross-section of the particle collision (i.e., the probability for a physics process to occur during the interaction of the particles), the event rate of the collision process is determined by the so-called luminosity of the storage ring.

$$R = \sigma_R \cdot L$$

Schematically the situation is shown in Fig. 18.

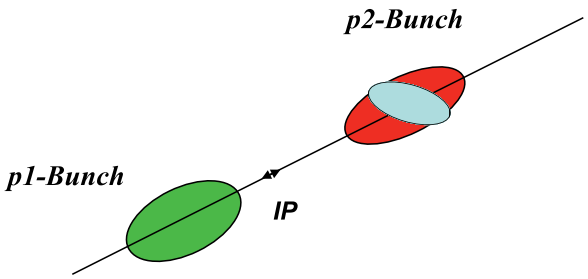


Fig. 18: Colliding bunches and luminosity

The luminosity is characterized by the storage ring parameters and depends on the stored beam current and the transverse size of the colliding bunches at the interaction point:

$$L = \frac{1}{4\pi e^2 f_0 n_b} * \frac{I_{p1} I_{p2}}{\sigma_x \sigma_y}$$

Here I_{p1} and I_{p2} are the values of the stored beam currents, f_0 is the revolution frequency of the machine, and n_b is the number of stored bunches. The quantities σ_x^* and σ_y^* in the denominator are the beam sizes in the horizontal and vertical planes at the interaction point. For a high-luminosity collider,

the stored beam currents have to be large and, at the same time, the beams have to be focused at the interaction point to very small dimensions. The β-functions at the collision points are therefore very small compared with their values in the arc cells. Typical values are more in the range of centimetres than of metres. At the same time, a large drift space is needed where focusing elements cannot be installed due to the particle detector that will be placed around the interaction point (Fig. 19).
As an example, the parameter list of the LHC including the design luminosity is presented:

$$\left.\begin{array}{ll} \beta^*_{x,y} = 0.55\ \mathrm{m} & f_0 = 11.245\ \mathrm{kHz} \\ \varepsilon_{x,y} = 5 \cdot 10^{-10}\ \mathrm{m \cdot rad} & n_b = 2808 \\ \sigma_{x,y} = 17\ \mu\mathrm{m} & I_p = 584\ \mathrm{mA} \end{array}\right\} L = 1.0 \cdot 10^{34}\ \mathrm{cm^{-2}s^{-1}}$$

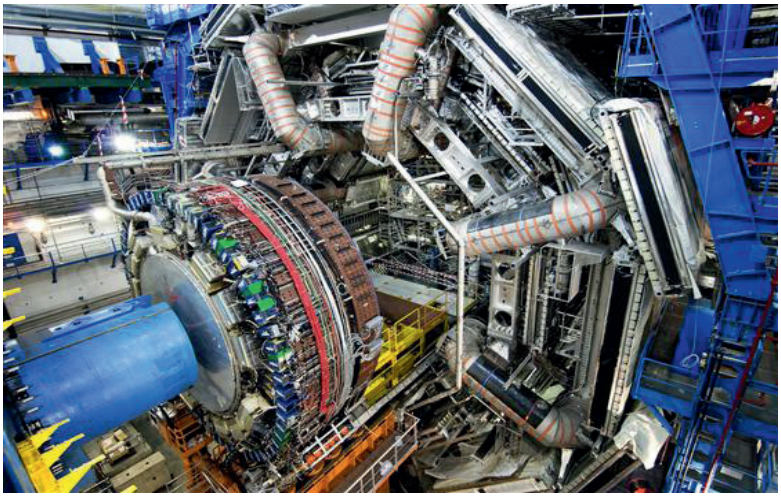


Fig. 19: Particle detector of the ATLAS collaboration at the LHC storage ring

Figure 20 shows the typical layout of such a mini-beta insertion. It consists in general of:

- a symmetric drift space that is large enough to house the particle detector and whose beam waist (where $\alpha_0 = 0$) is centred at the interaction point (IP) of the colliding beams;
- a quadrupole doublet (or triplet) installed on each side as close as possible to the IP;
- additional quadrupole lenses to match the optical parameters of the mini-beta insertion to the optical parameters of the lattice cell in the arc.

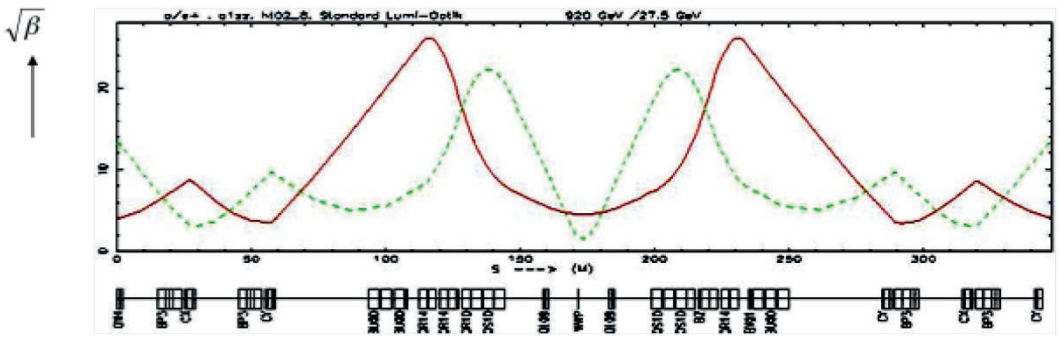


Fig. 20: Layout of a mini-beta insertion scheme

Following Liouville's theorem, the request of a small beam size immediately will lead to a corresponding increase of beam divergence. Figure 21 shows the corresponding situation in phase

space. It compares the x–x' ellipse in a typical location of the storage ring with the situation obtained at the IP.

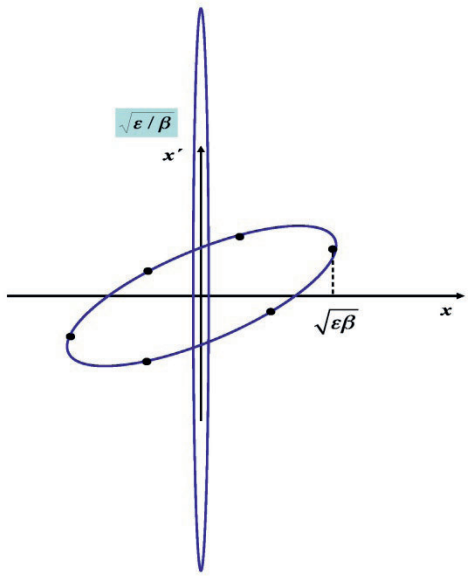


Fig. 21: Phase space ellipse in the arc of a storage ring and at the interaction point

While the beam size shrinks with reduced β as $\sigma = \sqrt{\varepsilon\beta}$, the corresponding divergence is increased as $\sigma' = \sqrt{\varepsilon/\beta}$. More quantitatively, the behaviour of the beta function in the vicinity the IP can be calculated and we obtain within the symmetric drift space of the mini-beta insertion

$$\beta(s) = \beta_0 + \frac{\ell_1^{\,2}}{\beta_0}\,.$$

Inevitably, the strong focusing that is needed to obtain the smallest spot size of the beam leads to a large increase of the beam inside the first quadrupoles after the IP. It is at this location that we will need the strongest and at the same time the largest aperture quadrupoles – this is the price that we have to pay for the luminosity that we will get out of the design.

References and further reading

[1] P. Bryant and K. Johnsen, *The Principles of Circular Accelerators and Storage Rings* (Cambridge University Press, Cambridge, 1993).

[2] F. Hinterberger, *Physik der Teilchenbeschleuniger* (Springer, Berlin, 1997).

[3] M. Sands, The Physics of e^+e^- Storage Rings, SLAC Report 121 (1970).

[4] D. Edwards and M. Syphers, An Introduction to the Physics of Particle Accelerators, Superconducting Super Collider Laboratory Report (1990).

[5] E. Jaeschke *et al.*, in European Particle Accelerator Conference, Rome (1988), p. 365.

[6] P. Schmüser, Basic course on accelerator optics, in CERN Accelerator School: 5th General Accelerator Physics Course, CERN 94-01 (1994).

[7] K. Wille, *Physics of Particle Accelerators* (Oxford University Press, Oxford, 2005).

[8] B. Holzer, Lattice design, in CERN Accelerator School: Intermediate Accelerator Physics Course, CERN 2006-002 (2006).

[9] A. Chao and M. Tigner, *Handbook of Accelerator Physics and Engineering* (World Scientific, Singapore, 1999).

[10] M. Reiser, *Theory and Design of Charged Particle Beams* (Wiley-VCH, Weinheim, 2008).

Introduction to Longitudinal Beam Dynamics

B.J. Holzer
CERN, Geneva, Switzerland

Abstract
This chapter gives an overview of the longitudinal dynamics of the particles in an accelerator and, closely related to that, the issue of synchronization between the particles and the accelerating field. Beginning with the trivial case of electrostatic accelerators, the synchronization condition is explained for a number of driven accelerators like Alvarez linacs, cyclotrons and finally synchrotrons and storage rings, where it plays a crucial role. In the case of the latter, the principle of phase focusing is motivated qualitatively as well as on a mathematically more correct level and the problem of operation below and above the transition energy is discussed. Throughout, the main emphasis is more on physical understanding rather than on a mathematically rigorous treatment.

1 Introduction

The longitudinal movement of the particles in a storage ring – and, closely related to this, the acceleration of the beam – is strongly related to the problem of synchronization between the particles and the accelerating system. This synchronization might be established via the basic hardware and design of the machine (like in a Wideroe structure), via the orbit (in a cyclotron), or in a more sophisticated way it can be a fundamental feature of the ring, which leads in the end to the name 'synchrotron'. We will treat these different aspects in more detail. But, before we do, we would like to start on a very fundamental ground and at the same time go back a bit in history.

2 Electrostatic machines

The most prominent example of these machines, besides the Cockcroft–Walton generator, is the Van de Graaff generator. A sketch of the principle is shown in Fig. 1.

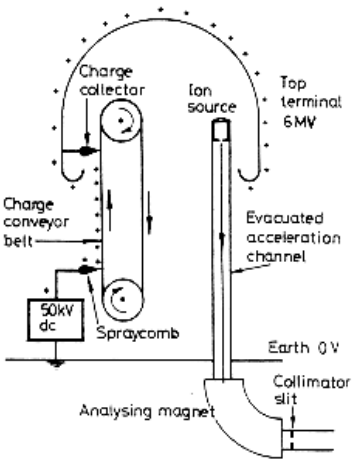


Fig. 1: Technical principle of a Van de Graaff accelerator

Using a moderate d.c. high voltage, charges are sprayed on a moving belt or chain with isolated links and transported to a kind of Faraday screen, where the charges accumulate, leading to considerable high voltages. Basically, it is the transformation of mechanical energy into electrostatic energy via the movement of the belt that defines the physical principle behind the generator. The high-voltage part is usually connected to a particle source (protons or heavy ions) and the potential difference between the high-voltage end and the ground potential will define the particle energy. Voltages in the range of some megavolts can be achieved, where usually a discharge-suppressing gas is needed to achieve the highest possible terminal voltages (up to ≈ 30 MV). By their design concept, these machines deliver an excellent energy resolution, as basically each and every particle sees the same accelerating potential (which is no longer the case as soon as we have to consider a.c. accelerating structures).

An example of such a machine is shown in Fig. 2. This 'tandem' Van de Graaff accelerator uses a stripper foil in the middle of the structure, thus, in the first half of the structure, accelerating negative ions that are produced in a Cs-loaded source. After stripping the electrons, the same voltage is applied in the second part of the structure to gain another step in energy. However, after leaving the accelerator at the down end, the particle energy is still limited by the high voltage that can be created, and that is finally always limited due to discharge effects of the electric field.

The kinetic energy of the particle beam is given by the integration over the electric field $\boldsymbol{E}$, in direction z of the particle motion and measured as usual in units of electronvolts (eV):

$$dW = eE_z ds \quad \rightarrow \quad W = e\int E_z ds \; = [\mathrm{eV}]$$

Fig. 2: Tandem Van de Graaff accelerator of the MPI Institute, Heidelberg. The ion source, delivering negative ions, is connected to the accelerator on the right-hand side of the picture. The acceleration structure is housed in a vessel filled with discharge suppressing gas (SF6). At the far end, the dipole spectrometer magnet is visible to separate different ion species.

It should be emphasized here that a special synchronization between accelerating voltage and particle beam is not needed as a d.c. voltage is used for acceleration.

3 The first radio-frequency accelerator: Wideroe linac

The basic limitation given by the maximum achievable voltage in electrostatic accelerators can be overcome by applying a.c. voltages. However, a more complicated design is needed to prevent the particles from being decelerated during the negative half-wave of the radio-frequency (RF) system. In 1928 Wideroe presented for the first time the layout of such a machine (and built it). Figure 3 shows the principle.

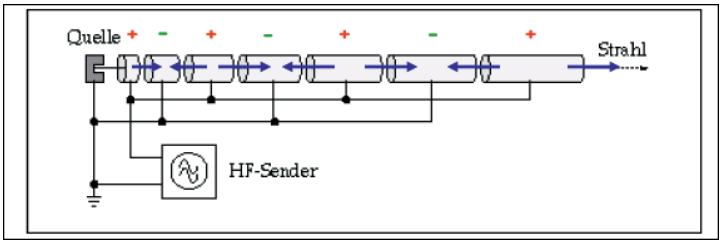


Fig. 3: Schematic layout of an RF linear accelerator (drift tube linac) as proposed by Wideroe in 1928

The arrows show the direction of the electric field for a given moment in time (i.e., the positive half-wave of the a.c. voltage) and the corresponding polarity applied to the electrodes indicated by the '+/–' sign in the figure. In principle, arbitrarily high beam energies can be achieved by applying the same voltage to a given number of electrodes, provided that the particle beam is shielded from the electric field during the negative RF half-wave. Accordingly the electrodes are designed as drift tubes ('drift tube linac') whose length is adopted according to the particle velocity and RF period. Schematically the problem is shown in Fig. 4.

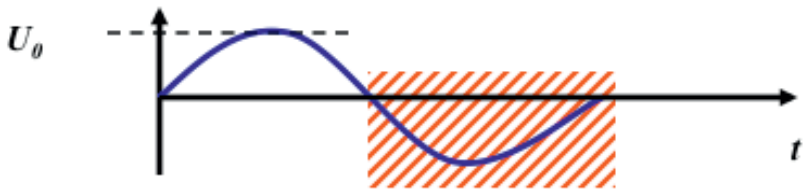


Fig 4: During the negative half-wave of the RF system the particles have to be shielded to avoid deceleration

The red area corresponds to the time during which the particle has to be shielded from the decelerating field direction and is defined by the RF period: $t_{\text{shield}} = \tau_{\text{RF}}/2$. Accordingly the drift tube length has to be

$$l_i = v_i \cdot \frac{\tau_{rf}}{2}$$

If the kinetic energy (in the classical regime) is given by

$$E_i = \frac{1}{2} m v^2$$

we obtain an equation for the drift tube length that, for a given accelerating voltage U_0 and charge q, depends on the RF frequency ν_{RF} and the accelerating step i:

$$l_i = \frac{1}{\nu_{rf}} \cdot \sqrt{\frac{i \cdot q \cdot U_0 \cdot sin\psi_s}{2m}} \tag{1}$$

The parameter ψ_s describes the so-called synchronous phase and can be chosen to be 0° in this case to obtain highest acceleration performance.

For completeness, it should be mentioned that in 1946 Alvarez improved the concept by embedding the structure inside a vessel to reduce RF losses. The drift tube linac is typically used as an accelerator for proton and heavy-ion beams, and beam energies of the order of some tens of Mega electronvolt (MeV) are obtained. At the time of writing, a new proton linac is installed at CERN to improve the beam performance for the LHC beams and – as you would expect – again an Alvarez drift tube linac is used to gain in three stages a kinetic energy of the protons of 150 MeV.

One of the best-known examples of such an accelerator is running at GSI in Darmstadt, used as a universal tool for the acceleration of (almost) any heavy-ion species. In Fig. 5 the inner structure is shown, with the drift tubes and the surrounding vessel.

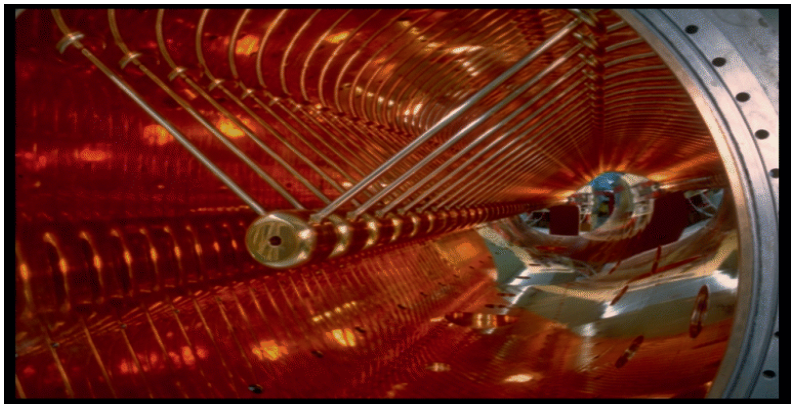


Fig. 5: 'Unilac' of GSI

Unlike the d.c. accelerators, synchronization has to be obtained between the particles and the accelerating RF field, which – as shown above – is represented by the drift tube length, and so in a certain sense built into the hardware of the system.

4 Cyclotrons

In principle, the drift tube linac concept can be applied to relatively high energies (doing the calculations in Eq. (1), relativistically correct even beyond the classical regime). The length of the drift tubes, however, makes the design very inefficient for high energies, and more sophisticated ideas are used to keep the structure compact and within reasonable financial and technical limits. The idea that is considered as a next natural step is the application of a magnetic dipole field to bend the particle orbit into a circle or, better (as we will see), into a spiral.

A constant dipole field applied perpendicular to the particle motion will lead to a transverse deflecting Lorentz force that is given by

$$\mathbf{F} = q \cdot (\boldsymbol{v} \times \boldsymbol{B}) = q \cdot v \cdot B$$

where we have replaced the cross-product by a simple multiplication.

The condition for a circular orbit is given by the equality of this Lorentz force and the centrifugal force, and leads to a relation that is of major importance for any circular machine:

$$q \cdot v \cdot B = \frac{m \cdot v^2}{R} \quad \rightarrow \quad B \cdot R = p/q \tag{2}$$

The expression $B{\cdot}R$ on the right-hand side is called the beam rigidity, as it describes the resistance of the particle beam to any external deflecting force and clearly depends on the particle momentum. At the same time it defines – for a given size of the machine (R) and for a given maximum magnetic field (B) of the dipoles installed – the highest momentum that can be carried by the (circular) accelerator.

The revolution frequency of the particles

$$\omega_z = \frac{v}{R} = \frac{q}{m} \cdot B_z \tag{3}$$

is constant as long as we can consider the classical approach, i.e., as long as the mass of the particles can be assumed to be constant. Accordingly the RF frequency applied for the acceleration can be kept constant and – in the ideal case – set equal to a multiple of the revolution frequency to obtain particle synchronization.

A schematic layout of a cyclotron is shown in Fig. 6. The accelerating RF voltage is applied between the two halves of the pill-box-like structure, leading to a step in energy (or better momentum) twice per revolution.

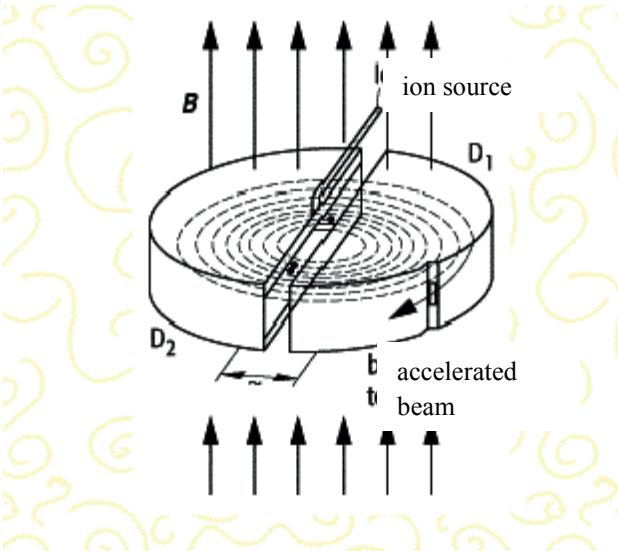


Fig. 6: Schematic layout of a cyclotron with the two half-pill-box-like electrodes that contain the spiralling orbit, and the external magnetic field line indicated.

According to Eq. (2), the radius of the particle orbit will increase due to the constant magnetic field, and a spiralling orbit is obtained. The maximum energy achievable in a cyclotron is determined by the strength and geometrical size of the dipole magnet. An example of such a classical cyclotron is shown in Fig. 7. The vacuum chamber of the cookie-like box, the magnet coils and the dipole magnet are clearly visible.

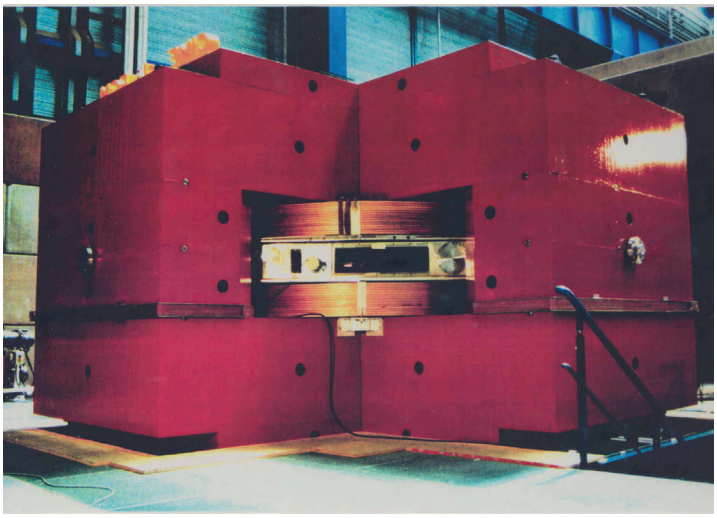


Fig. 7: Cyclotron SPIRAL at Ganil

Another limitation, however, should be mentioned. While the revolution frequency of the particles can be considered as constant in the classical treatment of Eq. (3), the situation changes as

soon as a considerable mass increase is obtained due to the relativistic particle energies. The correct expression for the revolution frequency ω_s, and thus as well for the RF frequency ω_{RF}, therefore is

$$\omega_s(t) = \omega_{rf}(t) = \frac{q}{\gamma(t) \cdot m_0} \cdot B$$

where γ describes the time-dependent relativistic parameter. During the acceleration process the particle frequency slows down, and to keep the particle motion synchronous with the applied RF frequency the external RF system has to be tuned to follow this effect. While higher particle energies can be obtained, the performance of the machine will degrade as a d.c.-like beam acceleration is no longer possible due to the RF cycling that is needed for each acceleration process.

As before, to refer to the problem of synchronization: the synchronous condition between particles and accelerating RF voltage in a cyclotron is defined by the length of the spiralling orbit.

5 Radio-frequency resonators and transit time factor

Synchrotrons are ring-like accelerators that use located RF systems for particle acceleration and, owing to their design principle, can achieve up to now the highest particle energies. The RF voltage – much like in the case of a drift tube linac – is usually created in a so-called resonator where a standing RF wave is created, with an electric field vector pointing in the direction of particle motion (Fig. 8).

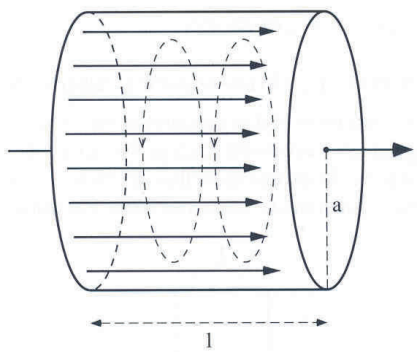


Fig. 8: Pill box cavity with longitudinal electric field

Before going into the detail of the particle dynamics in a synchrotron, a problem has to be mentioned that is present in all RF-like structures. The gain in kinetic energy (which is equal in this case to the gain in overall energy) in a d.c. accelerating system is given by

$$dW = dE = eE_z ds$$

with E_z describing the longitudinal component of the electric field, and the overall energy gain W is obtained via a simple integration over the complete path ds of the particle:

$$W = e\int E_z ds = eV$$

The situation changes quite a bit if RF voltage is applied. The electric field is changing as a function of the RF frequency used:

$$E_z = E_0 \cos\omega t = \frac{V}{g}\cos\omega t$$

and the total energy gain obtained is given by

$$\Delta W = \frac{eV}{g}\int_{-g/2}^{g/2} \cos\omega\frac{z}{v} dz$$

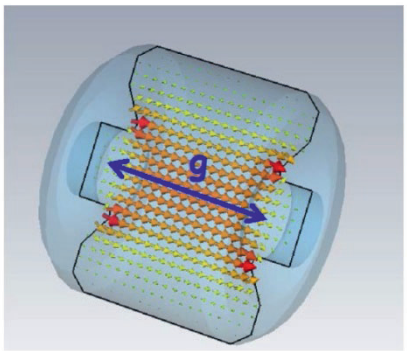


Fig. 9: Typical RF resonator with calculated field vector and gap size g

As usual z refers here to the longitudinal direction, V is the maximum amplitude of the applied voltage and v is the speed of the particle. The integration refers to the complete distance between the cavity walls (the *gap*), and we refer for symmetry reasons to the centre of the RF resonator. Solving the integral we obtain

$$\Delta W = eV \frac{\sin\theta/2}{\theta/2} = eVT$$

The parameter θ is called transit angle and is defined as

$$\theta = \frac{\omega g}{v}$$

and the transit time factor T, describing the effective available accelerating voltage in an RF system, is

$$T = \frac{\sin\theta/2}{\theta/2}$$

The ideal case of a transit time factor approaching 1 is obtained if the argument $\theta/2$ is as small as possible, ideally $\theta/2 \to 0$, which corresponds either to $\omega \to 0$ as in the case of d.c. acceleration or to $g \to 0$. This means that we obtain the most effective use of the RF system for the smallest gap distance of the cavity. The simple case shown in Fig. 8 therefore is not an ideal situation. An improved cavity design will rather look like the one in Fig. 10, where the gap is reduced to the limit of discharges that will occur if the resulting electric field is pushed too high.

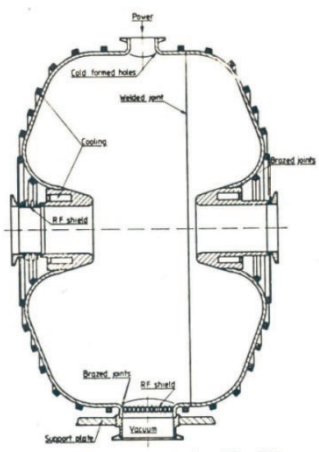


Fig. 10: Example of an optimised cavity design to increase the transit time factor

6 Synchrotrons

A synchrotron is a circular accelerator or storage ring with

- a design orbit of constant radius defined by the arrangement and strength of a number of dipole magnets that according to Eq. (2) define the particle rigidity (or momentum),
- an RF system, located at a distinct place in the ring and powered at an RF frequency that is equal to the revolution frequency of the particles or an integer multiple (so-called harmonic) of it.

For the description of the particle dynamics, we refer to a synchronous particle of ideal energy, phase and energy gain per turn. As we will see, the synchronization between the RF system and the particle beam is of major importance in this machine and has to be explicitly included in the design. To understand its principle, we have to refer briefly to the transverse dynamics of a particle with momentum error and the resulting dispersive effects (Fig. 11).

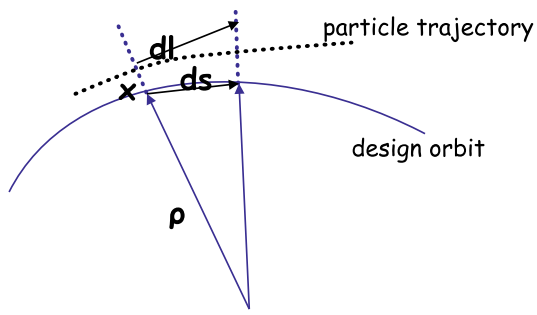


Fig. 11: Trajectories for particles of design energy and off momentum

While the ideal particle will run on the design orbit defined by the dipole magnets and will proceed a distance d*s*, a non-ideal particle running on a displaced orbit (displaced to the outer side of the ring in the example of Fig. 11) will pass the corresponding distance d*l*:

$$\frac{dl}{ds} = \frac{\rho + x}{\rho}$$

Solving for d*l* we obtain

$$dl = \left(1 + \frac{x}{\rho(s)}\right) ds$$

and integrating around the machine we get the orbit length of the non-ideal particle, which depends on the radial displacement *x*:

$$l_{\Delta E} = \int dl = \int \left(1 + \frac{x_{\Delta E}}{\rho(s)}\right) ds$$

where we assume that the radial displacement $x_{\Delta E}$ is caused by a momentum error and the dispersion function of the magnet lattice

$$x_{\Delta E}(s) = D(s)\frac{\Delta p}{p} \tag{4}$$

we obtain an expression for the difference in orbit length between the ideal and dispersive particle, which is determined by the amount of relative momentum error and the dispersion function in the storage ring:

$$\delta l_{\Delta E} = \frac{\Delta p}{p} \int \left(\frac{D(s)}{\rho(s)}\right) ds$$

The ratio between relative orbit difference and relative momentum error is called the momentum compaction factor α_p and using Eq. (4) it is determined by the integral of the dispersion function around the ring and the bending radius of the dipole magnets:

$$\frac{\delta l_\varepsilon}{L} = \alpha_p \frac{\Delta p}{p}$$

$$\alpha_p = \frac{1}{L} \oint \left(\frac{D(s)}{\rho(s)} \right) ds$$

For first estimates, let us assume equal bending radii in all dipoles, so $1/\rho$ = const and we replace the integral of the dispersion around the ring by a sum over the average dispersion in the dipole magnets (outside the dipoles the term $1/\rho = 0$, so this assumption is justified for a rough estimate):

$$\int_{dipoles} D(s)\, ds \approx l_{\Sigma(dipoles)} \cdot \langle D \rangle_{dipole}$$

So we get a nice and simple expression for the momentum compaction factor that depends only on the ratio of average dispersion and geometric radius R of the machine:

$$\alpha_p = \frac{1}{L}\, l_{\Sigma(dipoles)} \cdot \langle D \rangle \frac{1}{\rho} = \frac{1}{L}\, 2\pi\rho \cdot \langle D \rangle \frac{1}{\rho}$$

$$\alpha_p \approx \frac{2\pi}{L} \langle D \rangle \approx \frac{\langle D \rangle}{R}$$

Assuming finally that the particles are running at the speed of light, $v \approx c$ = const, the ratio between relative error in time is given by the relative change of the orbit length and thus by the momentum compaction factor and the relative momentum error:

$$\frac{\delta T}{T} = \frac{\delta l_\varepsilon}{L} = \alpha_p \frac{\Delta p}{p} \tag{5}$$

6.1 Dispersive effects in synchrotrons

In the case of non-relativistic particles, however, and the problem of synchronization, we need a more careful treatment, which will give us a relation between the revolution frequency and the momentum error of the particles. The parameter of interest is the ratio between the relative momentum error and the relative frequency deviation of a particle:

$$\frac{df_r}{f_r} = \eta \frac{dp}{p}$$

Given the revolution frequency as a function of machine circumference and speed,

$$f_r = \frac{\beta c}{2\pi R}$$

we obtain via logarithmic derivation

$$\frac{df_r}{f_r} = \frac{d\beta}{\beta} - \frac{dR}{R} \tag{6}$$

Remembering that the relative change in radius, i.e., the second term in the expression, is given by the momentum compaction factor α_p,

$$\frac{dR}{R} = \alpha \frac{dp}{p}$$

The momentum is related to the particle energy and its velocity

$$p = mv = \beta\gamma \frac{E_0}{c}$$

and its relative error is obtained as

$$\frac{dp}{p} = \frac{d\beta}{\beta} + \frac{d\left(1-\beta^2\right)^{-\frac{1}{2}}}{\left(1-\beta^2\right)^{-\frac{1}{2}}} = \left(1-\beta^2\right)^{-1} \frac{d\beta}{\beta}$$

Introducing the last two equations into Eq. (6), we finally obtain the required relation between frequency offset and momentum error

$$\frac{df_r}{f_r} = \left(\frac{1}{\gamma^2} - \alpha\right)\frac{dp}{p} \tag{7}$$

which for ultra-relativistic particles reduces to the simplified expression of Eq. (5). Accordingly the η parameter defined above is given as

$$\eta = \frac{1}{\gamma^2} - \alpha$$

As an important remark we state that the change of revolution frequency depends on the particle energy γ and possibly changes sign during acceleration. Particles get faster in the beginning and arrive earlier at the cavity location (classical regime), while particles that travel at $v \approx c$ will not get faster any more but rather get more massive and, being pushed to a dispersive orbit, will arrive later at the cavity (relativistic regime). The boundary between the two regimes is defined for the case where no frequency dependence on dp/p is obtained, namely, $\eta = 0$ and the corresponding energy is called the 'transition energy':

$$\gamma_{tr} = \frac{1}{\sqrt{\alpha}}$$

In general we will design machines in such a way as to avoid the crossing of this transition energy. As it involves changes of the RF phase unless, the particles lose their longitudinal focusing created by the sinusoidal RF function, the bunch profile will dilute and get lost. Qualitatively the longitudinal focusing effect and the problem of gamma transition is explained in Fig. 12.

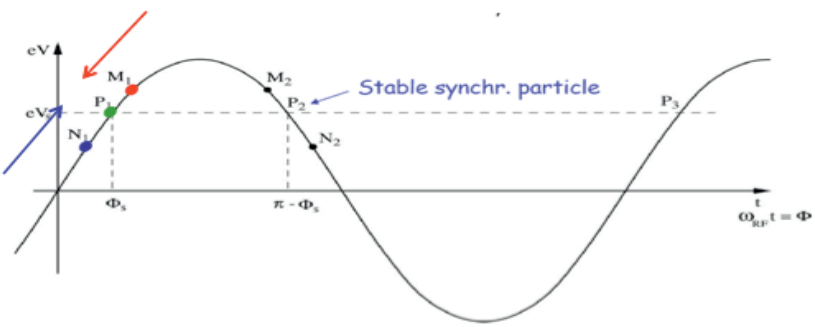


Fig. 12: Qualitative picture of the phase focusing principle below transition

6.2 The classical regime

Assume an ideal particle that is passing the cavity in time (or phase) the resonator at a certain position, as indicated by the green spot in Fig. 12. It will see a certain accelerating voltage and correspondingly receive an energy increase. We call this phase the synchronous phase, as indicated in the plot. A particle that has smaller energy than the ideal one will travel at lower speed and arrive after the next turn later, or at a larger phase and thus see a stronger accelerating voltage. It will therefore compensate the lack in energy and step-by-step come closer to the ideal particle. Just the opposite happens to a particle that has a positive energy offset. As it is faster than the synchronous particle, it will arrive earlier at the cavity, see a smaller voltage and again step-by-step will approach the ideal particle. In both cases a net focusing effect is obtained, which is due to the relation between momentum and speed and the right choice of the synchronous phase.

For highly relativistic particles the same effect exists but based now on the mass increase with energy. As visualized in Fig. 13 the high-energy particle (marked in blue) will, due to its higher mass, run on a longer orbit and – as its speed is constant $v \approx c$ – it will arrive later at the cavity location.

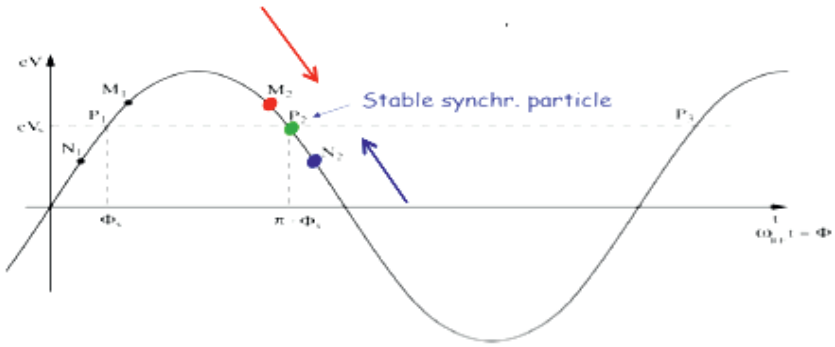


Fig. 13: Phase focusing situation above transition

As a consequence, the synchronous phase has to be chosen depending on whether we are running the machine below or above transition. Synchrotrons that have to pass through transition will have to apply a phase jump to keep the particles bunched.

In this context it is worth having a look at the acceleration mechanism itself. The particle momentum is defined via the beam rigidity by the dipole field B,

$$p = eB\rho$$

and as a consequence a change in particle momentum is reflected by an appropriate change of the B field:

$$\frac{dp}{dt} = e\rho\,\dot{B}$$

The momentum increase per turn is therefore given by

$$(\Delta p)_{turn} = e\rho\,\dot{B}T_r = \frac{2\pi\, e\rho\, R\dot{B}}{v}$$

and referring to the energy change rather than the change in momentum we get with

$$E^2 = E_0^2 + p^2c^2 \quad \Rightarrow \quad \Delta E = v\Delta p$$

the change in energy per turn, which is clearly related to the accelerating voltage and the synchronous phase ϕ_s of the particles:

$$\Delta E_{turn} = \Delta W_{turn} = 2\pi e\rho R\dot{B} = e\hat{V}\sin\phi_s$$

The following are some remarks that might be worth emphasizing:

- The energy gain depends on the rate of change of the dipole field.
- The number of stable synchronous particles is equal to the harmonic number h. They are equally spaced along the circumference.
- Each synchronous particle satisfies the relation $p = eB\rho$. They have the nominal energy and follow the nominal trajectory.
- As long as the particles are not fully relativistic yet, their revolution frequency will change and so does the RF frequency, which is a multiple of f_r to stay in synchronization during the complete acceleration process.

6.3 Frequency change during acceleration

The relation between the revolution frequency and RF frequency is defined by the harmonic number and depends on the size of the ring and the magnetic dipole field:

$$f_r = \frac{f_{RF}}{h} = f(B, R_s)$$

Hence, using the beam rigidity relation and the average dipole field to define the radius of the ideal particle, we get for an average magnetic field <B(t)>

$$\frac{f_{RF}(t)}{h} = \frac{v(t)}{2\pi R_s} = \frac{1}{2\pi}\frac{e}{m} < B(t) >$$

$$\frac{f_{RF}(t)}{h} = \frac{1}{2\pi}\frac{ec^2}{E_s(t)}\frac{r}{R_s}B(t)$$

Using the relativistic overall energy

$$E^2 = (m_0 c^2)^2 + p^2 c^2$$

we get finally an expression for the RF frequency as a function of the changing external dipole field:

$$\frac{f_{RF}(t)}{h} = \frac{c}{2\pi R_s}\left\{\frac{B(t)^2}{(m_0 c^2 / ecr)^2 + B(t)^2}\right\}^{1/2}$$

At high energies, or, more accurately, as soon as

$$B > \frac{m_0 c^2}{ecr} \tag{8}$$

the situation simplifies a lot and the equation above reduces to

$$\frac{f_{RF}(t)}{h} \approx \frac{c}{2\pi R_s} = const \tag{9}$$

so that we can keep the RF frequency constant. It is evident from Eq. (8) that for electron beams in general the condition (9) is (nearly) always fulfilled while proton or heavy-ion synchrotrons need a more sophisticated RF control.

7 Synchrotron motion

In the following we will again go through the longitudinal motion. However, unlike the previous section, we will try to put things on a mathematically more solid ground.

As in Fig. 12 we expect a longitudinal oscillation in phase and energy under the influence of the focusing mechanism explained above. The relation between relative frequency deviation and relative momentum error is given by Eq. (7) as

$$\frac{df_r}{f_r} = \left(\frac{1}{\gamma^2} - \alpha\right)\frac{dp}{p}$$

which translates into the difference in revolution time

$$\frac{dT}{T_0} = \left(\alpha - \frac{1}{\gamma^2}\right)\frac{dp}{p}$$

and leads to a difference in phase at the arrival at the cavity:

$$\begin{aligned}\Delta\psi &= 2\pi\frac{\Delta T}{T_{rf}} = \omega_{rf} * \Delta T \\ &= h * \omega_0 * \Delta T = h * 2\pi\frac{\Delta T}{T_0} \\ &= h * 2\pi\left(\alpha - \frac{1}{\gamma^2}\right)\frac{dp}{p} \\ &= \frac{h * 2\pi}{\beta^2}\left(\alpha - \frac{1}{\gamma^2}\right)\frac{dE}{E}\end{aligned}$$

As before, the revolution frequency ω_0 and the RF frequency ω_{RF} are related to each other via the harmonic number h. Hence the difference in energy and the offset in phase are connected to each other through the momentum compaction factor, or better the η parameter.

Differentiating with respect to time gives the rate of change of the phase offset per turn:

$$\Delta\dot{\psi} = \frac{\Delta\psi}{T_0} = \frac{h * 2\pi}{\beta^2 T_0}\left(\alpha - \frac{1}{\gamma^2}\right)\frac{dE}{E} \tag{10}$$

On the other hand, the difference in energy gain of an arbitrary particle that has a phase distance of $\Delta\psi$ to the ideal particle is given by the voltage and phase of the RF system:

$$\Delta E = e * U_0(\sin(\psi_s + \Delta\psi) - \sin\psi_s)$$

As before, we describe the phase of the ideal ('synchronous') particle by ψ_s and the phase difference by $\Delta\psi$. Referring to small amplitudes $\Delta\psi$ of the phase oscillations, we can simplify the treatment by assuming

$$\sin(\psi_s + \Delta\psi) - \sin\psi_s = \sin\psi_s \underbrace{\cos\Delta\psi}_{\approx 1} - \cos\psi_s \underbrace{\sin\Delta\psi}_{\Delta\psi} - \sin\psi_s$$

and get for the rate of energy change per turn:

$$\Delta\dot{E} = e * \frac{U_0}{T_0}\Delta\psi\cos\psi_s$$

A second differentiation with respect to time delivers

$$\Delta\ddot{E} = e * \frac{U_0}{T_0} \Delta\dot{\psi} \cos\psi_s \tag{11}$$

Combining Eqs. (10) and (11) we get finally a differential equation for the longitudinal motion under the influence of the phase focusing mechanism:

$$\Delta\ddot{E} = e * \frac{U_0}{T_0} \frac{2\pi h}{\beta^2 T_0} \left(\alpha - \frac{1}{\gamma^2} \right) \frac{dE}{E} \cos\psi_s$$

For a given energy the parameters in front of the right-hand side are constant and describe the longitudinal – or 'synchrotron' – oscillation frequency. Using therefore

$$\Omega = \omega_0 * \sqrt{\frac{-eU_0 h \cos\psi_s}{2\pi\beta^2 E} \left(\alpha - \frac{1}{\gamma^2} \right)} \tag{12}$$

we get the equation of motion in the approximation of small amplitudes:

$$\Delta\ddot{E} + \Omega^2 \Delta E = 0$$

It describes a harmonic oscillation in E–ψ phase space for the difference in energy of a particle to the ideal (i.e., synchronous) particle under the influence of the phase focusing effect of our sinusoidal RF function.

As already discussed qualitatively, the expression (12) leads to real solutions if the argument of the square root is a positive number. Two possible situations therefore have to be considered: below the gamma transition the η parameter is positive, and above it η is negative. The synchronous phase as the argument of the cos function therefore has to be chosen to get an overall positive value under the square root:

$$\gamma < \gamma_{tr} \qquad \eta > 0, \quad 0 < \phi_s < \pi/2$$
$$\gamma > \gamma_{tr} \qquad \eta < 0, \quad \pi/2 < \phi_s < \pi$$

As an example Fig. 14 shows the superconducting RF system of the LHC, and the basic parameters, including the synchrotron frequency, are listed in Table 1.

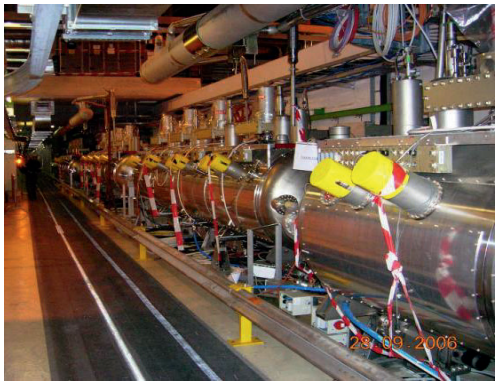


Fig. 14: LHC RF system

Table 1: Parameters of the LHC RF system.

Bunch length (4σ)	ns	1.06
Energy spread (2σ)	10^{-3}	0.22
Synchrotron rad. loss / turn	keV	7
RF frequency	MHz	400
Harmonic number h		35640
RF voltage / beam	MV	16
Energy gain / turn	keV	485
Synchrotron frequency	Hz	23

Figure 15 finally shows the longitudinal focusing effect that has been observed during the LHC commissioning phase. On the left-hand side, beam had been injected into the storage ring while the RF system was still switched off. The bunch, nicely formed by the RF voltage of the pre-accelerator, is visible only for a few turns and the bunch profile is decaying fast, as no longitudinal focusing is active. The right-hand side shows the situation with the RF system activated and the phase adjusted. The injected particles stay nicely bunched and the acceleration process can start.

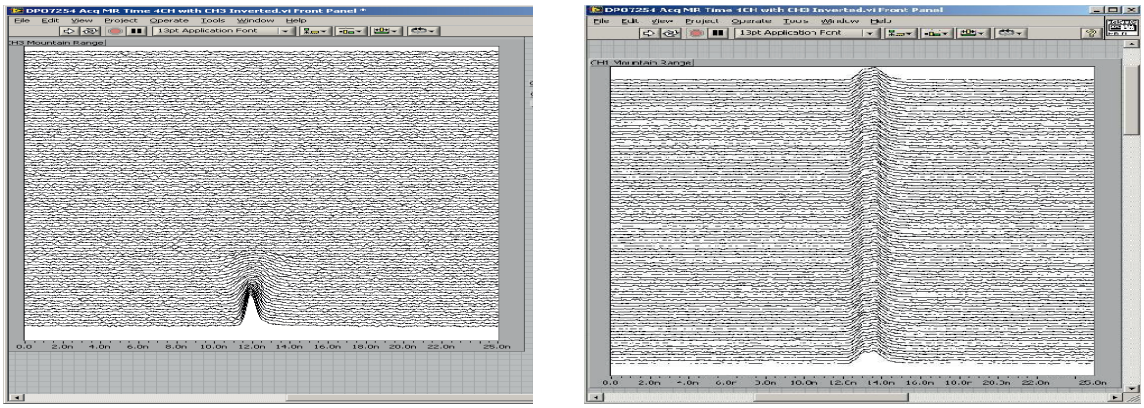


Fig. 15: Longitudinal beam profile measured during LHC injection. Left: the RF system is switched off, and the bunch profile is decaying fast. Right: the RF system is activated, and the particles stay bunched.

Bibliography

P. Bryant and K. Johnsen, *The Principles of Circular Accelerators and Storage Rings* (Cambridge University Press, Cambridge, 1993).

A. Chao and M. Tigner, *Handbook of Accelerator Physics and Engineering* (World Scientific, Singapore, 1999).

D. Edwards and M. Syphers, An Introduction to the Physics of Particle Accelerators, Superconducting Super Collider Laboratory Report (1990).

F. Hinterberger, *Physik der Teilchenbeschleuniger* (Springer, Berlin, 1997).

B. Holzer, Lattice design, in CERN Accelerator School: Intermediate Accelerator Physics Course, CERN 2006-002 (2006).

J. LeDuff, in CERN Accelerator School: 5th General Accelerator Physics Course, CERN 94-01 (1994).

M. Reiser, Theory and Design of Charged Particle Beams (Wiley-VCH, Weinheim, 2008).

M. Sands, The Physics of e^+e^- Storage Rings, SLAC Report 121 (1970).

P. Schmüser, Basic course on accelerator optics, in CERN Accelerator School: 5th General Accelerator Physics Course, CERN 94-01 (1994).

K. Wille, Physics of Particle Accelerators and Synchrotron Radiation Facilities (Teubner, Stuttgart, 1992).

Space-Charge Effects

N. Chauvin
Commissariat à l'Énergie Atomique et aux Énergies Alternatives, Gif-sur-Yvette, France

Abstract
First, this chapter introduces the expressions for the electric and magnetic space-charge internal fields and forces induced by high-intensity beams. Then, the root-mean-square equation with space charge is derived and discussed. In the third section, the one-dimensional Child–Langmuir law, which gives the maximum current density that can be extracted from an ion source, is exposed. Space-charge compensation can occur in the low-energy beam transport lines (located after the ion source). This phenomenon, which counteracts the space-charge defocusing effect, is explained and its main parameters are presented. The fifth section presents an overview of the principal methods to perform beam dynamics numerical simulations. An example of a particles-in-cells code, SolMaxP, which takes into account space-charge compensation, is given. Finally, beam dynamics simulation results obtained with this code in the case of the IFMIF injector are presented.

1 Space charge: beam self-generated fields and forces

1.1 Space charge

Consider two particles of identical charge q. If they are at rest, the Coulomb force exerts a repulsion (see Fig. 1(a). Now, if they are travelling with a velocity $v = \beta c$, they represent two parallel currents $I = qv$, which attract each other as a result of the effect of their magnetic fields (see Fig. 1(b)).

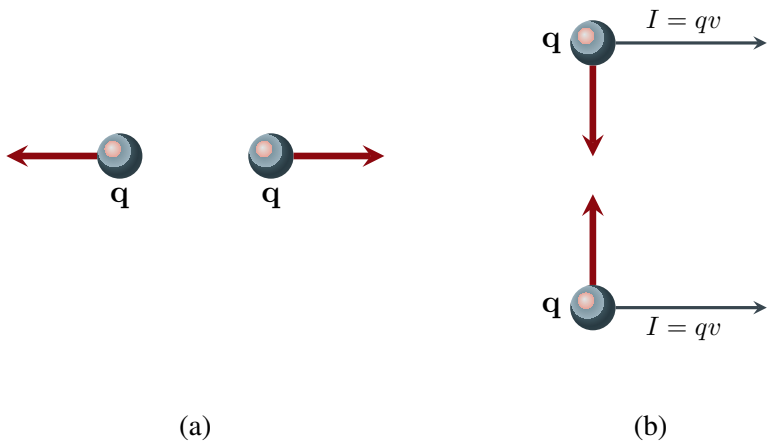


Fig. 1: (a) Coulomb force exerts a repulsion between stationary charges. (b) There is attraction between moving charges.

Now, consider a test particle of charge q in an unbunched beam of particles (charge q_i) with a circular cross-section (see Fig. 2). The Coulomb repulsion pushes the test particle outwards. The induced force is zero in the beam centre and increases towards the edge of the beam (Fig. 2(a)). The magnetic force is radial and attractive for the test particle in a travelling beam, represented as parallel currents (Fig. 2(b)).

Fig. 2: (a) Charges: repulsion. (b) Parallel currents: attraction

1.2 Space-charge fields

Consider a continuous beam with a cylindrical symmetry distribution that moves with a constant velocity $v = \beta c$. The beam charge density is

$$\rho(x, y, z) = \rho(r), \tag{1}$$

with $r = \sqrt{(x^2 + y^2)}$. For symmetry reasons, the electric field has only a radial component E_r. Using the integral form of Gauss's law over a cylinder centred on the beam axis gives

$$E_r(r) = \frac{1}{\epsilon_0 r} \int_0^r \rho(r') r' \, \mathrm{d}r' \tag{2}$$

The beam current density is

$$\mathbf{J}(x, y, z) = J(r)\mathbf{u}_z, \tag{3}$$

where $\mathbf{u}_z$ is the unitary vector of the beam propagation. If the particles of the beam have the same longitudinal speed, $\mathbf{v}_z = \beta_z c \, \mathbf{u}_z$, we have

$$\mathbf{J}(r) = \rho(r)\beta_z c \, \mathbf{u}_z. \tag{4}$$

For symmetry reasons, the magnetic field has only an azimuthal component B_θ. Using the integral form of Ampere's law over a cylinder centred on the beam axis gives

$$\begin{aligned} B_\theta(r) &= \frac{\mu_0}{r} \int_0^r J(r') r' \, \mathrm{d}r' \\ &= \frac{\mu_0 \beta_z c}{r} \int_0^r \rho(r') r' \, \mathrm{d}r'. \end{aligned} \tag{5}$$

From Eqs. (2) and (5), we obtain (as $c = 1/\sqrt{\epsilon_0 \mu_0}$)

$$B_\theta(r) = \frac{\beta_z}{c} E_r(r). \tag{6}$$

1.3 Space-charge forces

The space-charge fields exert a force $\mathbf{F}$ on a test particle at radius r that can be expressed by

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \tag{7}$$

which with our geometry can be written as

$$F_r = q(E_r + \beta_z c B_\theta). \tag{8}$$

If the particle trajectories obey the paraxial assumption, one can write

$$\beta^2 = \beta_x^2 + \beta_y^2 + \beta_z^2 \approx \beta_z^2. \tag{9}$$

From Eqs. (6) and (8), it follows finally that

$$F_r = qE_r(1 - \beta^2) = \frac{qE_r}{\gamma^2}. \tag{10}$$

The following should be noted about this derivation:

- In Eq. (10), the 1 represents the electric force and the β^2 the magnetic force.
- The electric force is defocusing for the beam; the magnetic force is focusing.
- The ratio of magnetic to electric forces, $-\beta^2$, is independent of the beam density distribution.
- For relativistic particles, the beam magnetic force almost balances the electric force.
- For non-relativistic particles (like low-energy ion beams), the space magnetic force is negligible: in the ion source extraction region and in the low-energy beam lines, the space charge has a defocusing effect.

1.3.1 Uniform beam density

A uniform beam density is expressed by

$$\rho(r) = \begin{cases} \rho_0 & \text{if } r \leq r_0, \\ 0 & \text{if } r > r_0. \end{cases} \tag{11}$$

The charge per unit length is

$$\lambda = \rho_0 \pi r_0^2. \tag{12}$$

The total beam current can be expressed by

$$I = \beta c \int_0^{r_0} 2\pi\rho(r')r' \, \mathrm{d}r' \tag{13}$$

So, we obtain

$$\rho_0 = \frac{I}{\beta c \pi r_0^2}. \tag{14}$$

From Eqs. (2) and (14):

$$E_r(r) = \frac{I}{2\pi\epsilon_0 \beta c r_0^2} r \quad \text{if } r \leq r_0, \tag{15a}$$

$$E_r(r) = \frac{I}{2\pi\epsilon_0 \beta c r} \quad \text{if } r > r_0. \tag{15b}$$

The following should be noted:

- The field is linear inside the beam.
- Outside of the beam, it varies according to $1/r$.

Similarly, from Eqs. (5) and (13):

$$B_\theta(r) = \mu_0 \frac{I}{2\pi r_0^2} r \quad \text{if } r \leq r_0, \tag{16a}$$

$$B_\theta(r) = \mu_0 \frac{I}{2\pi r} \quad \text{if } r > r_0. \tag{16b}$$

1.3.2 Space-charge forces – Gaussian beam density

A Gaussian beam density is expressed by

$$\rho(r) = \rho_{0\mathrm{g}} \exp\left(\frac{-r^2}{2\sigma_r^2}\right). \tag{17}$$

The charge per unit length is

$$\lambda = 2\rho_{0\mathrm{g}}\pi\sigma_r^2. \tag{18}$$

The space-charge electric field is

$$E_r(r) = \frac{\rho_{0\mathrm{g}}}{\epsilon_0}\frac{\sigma_r^2}{r}\left[1 - \exp\left(\frac{-r^2}{2\sigma_r^2}\right)\right]. \tag{19}$$

The following should be noted:

- The field is nonlinear inside the beam.
- Far from the beam (several σ_r), it varies according to $1/r$.

1.4 Space-charge expansion in a drift region

Consider a particle (charge q, mass m_0) beam of current I, propagating at speed $v = \beta c$ in a drift region, with the following hypotheses:

1. The beam has cylindrical symmetry and a radius r_0.
2. The beam is paraxial ($\beta_r \ll \beta_z$).
3. The beam has an emittance equal to 0.
4. The beam density is uniform.

Newton's second law for the transverse motion of the beam particles gives

$$\frac{\mathrm{d}(m_0\gamma\beta_r c)}{\mathrm{d}t} = m_0\gamma\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = qE_r(r) - q\beta c B_\theta(r). \tag{20}$$

Using Eqs. (15a) for E_r and (16a) for B_θ in Eq. (20) gives

$$m_0\gamma\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{qIr}{2\pi\epsilon_0 r_0^2\beta c}(1-\beta^2), \tag{21}$$

with

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \beta^2 c^2\frac{\mathrm{d}^2 r}{\mathrm{d}z^2}. \tag{22}$$

Equation (21) becomes

$$\frac{\mathrm{d}^2 r}{\mathrm{d}z^2} = \frac{qIr}{2\pi\epsilon_0 r_0^2 m_0 c^3\beta^3\gamma^3}. \tag{23}$$

We will now introduce several parameters used in the literature on beams with space charge. The characteristic current I_0 is defined by

$$I_0 = \frac{4\pi\epsilon_0 m_0 c^3}{q}. \tag{24}$$

The generalized perveance K, a dimensionless parameter, is defined by

$$K = \frac{qI}{2\pi\epsilon_0 m_0 c^3\beta^3\gamma^3}. \tag{25}$$

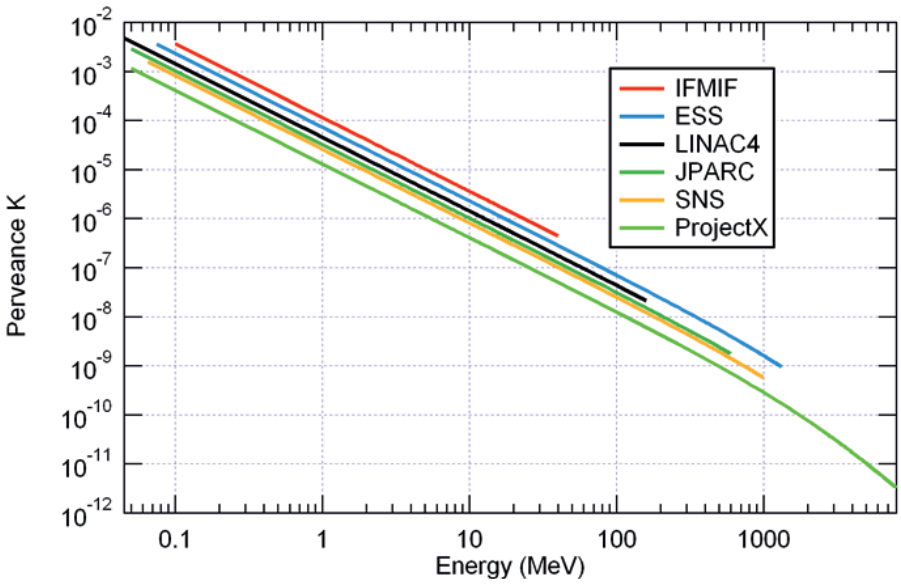


Fig. 3: Beam perveance for different hadron accelerators as a function of their beam energy

The perveance refers to the magnitude of space-charge effects in a beam. Figure 3 compares the generalized perveance, as a function of the beam energy, for various high-intensity accelerators.

Then, Eq. (23) for the particle trajectories can be reduced to the form

$$\frac{\mathrm{d}^2 r}{\mathrm{d}z^2} = \frac{K}{r_0^2} r. \tag{26}$$

In the case of a laminar beam, the trajectories of all particles are similar and, in particular, the particle at $r = r_0$ will always stay at the beam boundary. Considering $r = r_0 = r_{\mathrm{env}}$, the equation of the beam radius in a drift space can be written as

$$\frac{\mathrm{d}^2 r_{\mathrm{env}}}{\mathrm{d}z^2} = \frac{K}{r_{\mathrm{env}}}. \tag{27}$$

2 Root-mean-square envelope equation with space charge

2.1 Definition of root-mean-square quantities

In the following, $\langle A \rangle$ represents the mean of the quantity A over the beam particle distribution:

r.m.s. size $$\tilde{x} = \sqrt{\langle x^2 \rangle - \langle x \rangle^2}, \tag{28}$$

r.m.s. divergence $$\widetilde{x'} = \sqrt{\langle x'^2 \rangle - \langle x' \rangle^2}, \tag{29}$$

r.m.s. emittance $$\widetilde{\epsilon_x} = \sqrt{\tilde{x}^2 \widetilde{x'}^2 - \langle (x - \langle x \rangle)(x' - \langle x' \rangle) \rangle^2}. \tag{30}$$

The beam Twiss parameters can be expressed with the r.m.s. values as

$$\beta_x = \frac{\tilde{x}^2}{\widetilde{\epsilon_x}}, \tag{31}$$

$$\gamma_x = \frac{\widetilde{x'}^2}{\widetilde{\epsilon_x}}, \tag{32}$$

$$\alpha_x = -\frac{\langle (x - \langle x \rangle)(x' - \langle x' \rangle) \rangle}{\widetilde{\epsilon_x}}. \tag{33}$$

2.2 Root-mean-square emittance and nonlinear forces

Consider an ideal particle distribution in phase space (x, x') that lies on a line passing through the origin as represented on Fig. 4. Assume that, for any particles, the divergence x' is related to the position x by the expression

$$x' = Cx^n, \tag{34}$$

where n is positive and C is a constant. Considering the squared r.m.s. emittance (from Eq. (30))

$$\widetilde{\epsilon_x}^2 = \langle x^2 \rangle \langle x'^2 \rangle - \langle x x' \rangle^2, \tag{35}$$

and using Eq. (34) we obtain

$$\widetilde{\epsilon_x}^2 = C^2 (\langle x^2 \rangle \langle x^{2n} \rangle - \langle x^{n+1} \rangle^2). \tag{36}$$

When $n = 1$, the line is straight (Fig. 4(a)) and the r.m.s. emittance is $\widetilde{\epsilon_x}^2 = 0$. When $n \neq 1$, the relationship is nonlinear, the line of the phase space is curved (Fig. 4(b)), and the r.m.s. emittance is in general not zero. It is interesting to note that, even if the space phase area occupied by the beam distribution is zero, the r.m.s. emittance may be not zero if the distribution lies on a curved line.

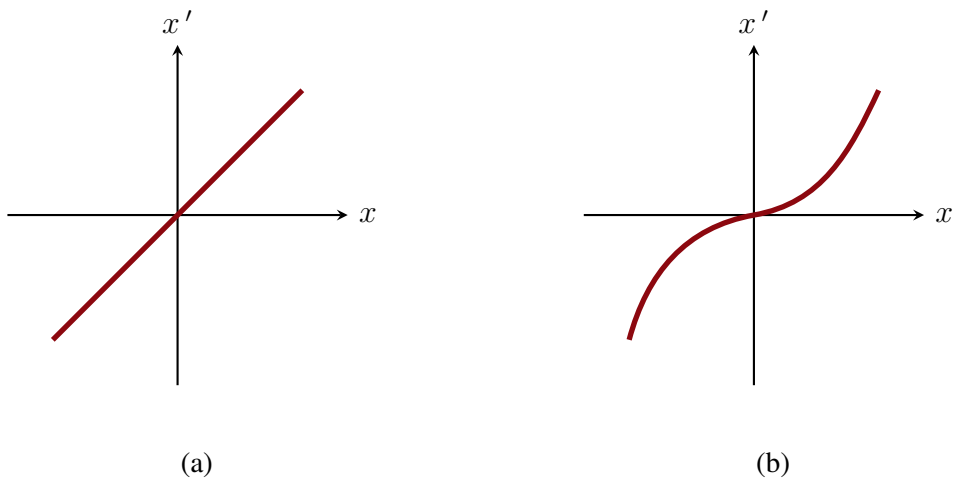


Fig. 4: Root-mean-square emittance for distributions with zero areas: (a) $\varepsilon = 0$; and (b) $\varepsilon \neq 0$

2.3 Root-mean-square envelope equation

Consider a beam moving in the direction s, where individual particles satisfy the equation of motion

$$x'' + \kappa(s)x - F_\mathrm{s} = 0, \tag{37}$$

where $\kappa(s)x$ represents a linear external focusing force (for instance quadrupole $\kappa = qB/\gamma m a \beta c$) and F_s is a *space-charge force term* (in general not linear).

To simplify the situation, assume that the beam is centred on the axis with no divergence, so that $\langle x \rangle = 0$ and $\langle x' \rangle = 0$ (i.e. $\tilde{x}^2 = \langle x^2 \rangle \equiv \overline{x^2}$). The equations of motion for the second moments of the distribution can then be written as

$$\frac{\mathrm{d}\overline{x^2}}{\mathrm{d}s} = 2\overline{xx'} \tag{38}$$

and

$$\begin{aligned} \frac{\mathrm{d}\overline{xx'}}{\mathrm{d}s} &= \overline{x'^2} + \overline{xx''} \\ &= \overline{x'^2} - \kappa(s)\overline{x^2} - \overline{xF_\mathrm{s}}. \end{aligned} \tag{39}$$

Differentiating Eq. (28) twice and using Eq. (38) gives

$$\tilde{x}'' = \frac{\overline{x''} + \overline{x^2}}{\tilde{x}} - \frac{\overline{xx'}}{\tilde{x}^3}. \tag{40}$$

Using Eqs. (39) and (30), we have finally that

$$\tilde{x}'' + \kappa(s)\tilde{x} - \frac{\widetilde{\epsilon_x}^2}{\tilde{x}^3} - \frac{\overline{xF_\mathrm{s}}}{\tilde{x}} = 0. \tag{41}$$

Equation (41) is the r.m.s. envelope equation and expresses the motion of the r.m.s. beam size. In three dimensions, one can write three envelope equations that are coupled through the space charge and can depend on the beam dimension in the three directions. These equations can be useful to find an analytic solution for simple problems. Furthermore, they are very important in the space-charge community.

2.4 Example: elliptical continuous beam of uniform density

An elliptical uniform beam density is expressed by

$$\rho(r) = \begin{cases} \rho_0 & \text{if } x^2/r_x^2 + y^2/r_y^2 < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

As the distribution is uniform, the semi-axes of the ellipse, r_x and r_y, are related to the r.m.s. beam sizes: $r_x = 2\tilde{x}$ and $r_y = 2\tilde{y}$. So, the r.m.s. envelope equations are given by

$$\tilde{x}'' + \kappa_x(s)\tilde{x} - \frac{\widetilde{\epsilon_x}^2}{\tilde{x}^3} - \frac{K}{2(\tilde{x}+\tilde{y})} = 0, \tag{43}$$

$$\tilde{y}'' + \kappa_y(s)\tilde{y} - \frac{\widetilde{\epsilon_y}^2}{\tilde{y}^3} - \frac{K}{2(\tilde{x}+\tilde{y})} = 0. \tag{44}$$

The second term of Eqs. (43) and (44) is a focusing term. The third term is the emittance term, which is defocusing (negative sign). The fourth term is the space-charge term, which is also defocusing. It can be seen that the two planes are coupled through the space-charge term of the r.m.s. envelope equation.

2.5 Linear space-charge force and equivalent beam

The space-charge force is linear only if the beam density is uniform, which is very unlikely in the case of practical beams. So the space-charge force is fundamentally nonlinear. Nevertheless, it was shown by Sacherer [16] that, for ellipsoidal bunches, where the r.m.s. emittance is either constant or specified in advance, the evolution of the r.m.s. beam projection is nearly independent of the beam density. This means that for calculation of the r.m.s. dynamics, the actual distribution can be replaced by an equivalent uniform beam that has the same intensity and the same r.m.s. second moments.

3 The Child–Langmuir law

The Child law states the maximum current density that can be carried by charged-particle flow across a one-dimensional extraction gap. The limit arises from the longitudinal electric fields of the beam space charge. It is a very important result in beam physics of collective effects and also in ion source fields where the extracted beam current is one of the most important parameters of an ion source.

In this section, the Child–Langmuir law for a one-dimensional extraction gap will be derived. It is enough to emphasize the fundamental physical phenomena that limit the beam current extracted from a particle source. However, for a practical extraction system, it is necessary to perform three-dimensional calculations by including form factors that depend on the geometry of the extraction electrodes.

3.1 Extraction gap

The extraction gap is the first stage of an accelerator: low-energy charged particles from a source are accelerated to moderate energy ($\sim$10 keV to $\sim$1 MeV) and formed into a beam. The Child law calculation applies to the one-dimensional gap of Fig. 5. A voltage $-V_0$ is applied across the vacuum gap of width d. Charged particles with low kinetic energy enter at the grounded boundary. The particles have rest mass m_0 and carry positive charge q. Particles leave the right-hand boundary with kinetic energy qV_0. We shall assume that the exit electrode is an ideal mesh that defines an equipotential surface while transmitting all particles.

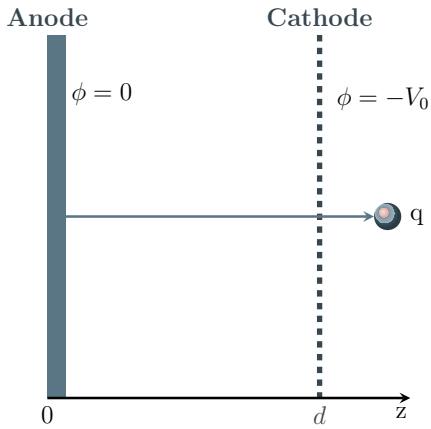


Fig. 5: Geometry of an infinite planar extraction gap

3.2 One-dimensional Child law for non-relativistic particles

To simplify the calculation, the following assumptions are made:

1. Particle motion is non-relativistic ($qV_0 \ll m_0c^2$).
2. The source on the left-hand boundary can supply an unlimited flux of particles. Flow restriction would result only from space-charge effects.
3. The transverse dimension of the gap is large compared with d.
4. The transverse magnetic force generated by particle current across the gap is negligible compared with the axial electric force. Consequently, the particle trajectories are straight across the gap. This assumption is valid for an ion beam, but can be violated in the case of a high-current relativistic electron gun.
5. Particles flow continuously; the electric field and particle density at all z are constant.

The steady-state condition means that the space-charge density, $\rho(z)$, is constant in time

$$\frac{\partial \rho(z)}{\partial t} = 0. \tag{45}$$

This equation implies that the current density, J_0, is the same at all positions in the gap. The charge density can be expressed by

$$\rho(z) = \frac{J_0}{v_z(z)}. \tag{46}$$

According to the above assumptions, the particle velocity in the gap is

$$v_z^2(z) = \frac{2q\phi(z)}{m_0}. \tag{47}$$

Using Eqs. (46) and (47) we obtain

$$\rho(z) = \frac{J_0}{\sqrt{2q\phi(z)/m_0}}. \tag{48}$$

Now, consider the one-dimensional Poisson equation:

$$\nabla^2\phi = \frac{\mathrm{d}^2\phi(z)}{\mathrm{d}z^2} = -\frac{\rho(z)}{\epsilon_0}. \tag{49}$$

Substituting Eq. (48) into Eq. (49) gives

$$\frac{\mathrm{d}^2\phi(z)}{\mathrm{d}z^2} = -\frac{J_0}{\epsilon_0\sqrt{2q/m_0}}\frac{1}{\phi^{1/2}}. \tag{50}$$

If we introduce the dimensionless variables $\zeta = z/d$ and $\Phi = -\phi/V_0$, Eq. (50) becomes

$$\frac{\mathrm{d}^2\Phi(z)}{\mathrm{d}\zeta^2} = -\frac{\alpha}{\phi^{1/2}}, \tag{51}$$

with

$$\alpha = \frac{J_0 d^2}{\epsilon_0 V_0\sqrt{2qV_0/m_0}}. \tag{52}$$

Three boundary conditions are needed to integrate Eq. (51): $\Phi(0) = 0$, $\Phi(1) = 1$ and $\mathrm{d}\Phi(0)/\mathrm{d}\zeta = 0$. Multiplying both sides of Eq. (51) by $\Phi' = \mathrm{d}\Phi/\mathrm{d}\zeta$, we can integrate and obtain

$$(\Phi')^2 = 4\alpha\sqrt{\Phi(\zeta)}. \tag{53}$$

A second integration gives

$$\Phi^{3/4} = \tfrac{3}{4}\sqrt{4\alpha}\zeta, \tag{54}$$

or, coming back to the initial variables,

$$\phi(z) = -V_0\left(\frac{z}{d}\right)^{4/3}. \tag{55}$$

In Eq. (54), the condition $\Phi(1) = 1$ implies $\alpha = \frac{4}{9}$. Substituting in Eq. (52) gives

$$J_0 = \frac{4}{9}\epsilon_0\left(\frac{2q}{m_0}\right)^{1/2}\frac{V_0^{3/2}}{d^2}. \tag{56}$$

This is the Child–Langmuir law, which represents the maximum current density that can be achieved in the diode by increasing the ion supply by the anode. This limitation is only due to space charge. The only way to increase the extracted current is to increase the electric field in the gap (i.e. increase the gap voltage or decrease the cathode–anode spacing). In a real ion source, the extracting electric field cannot be increased infinitely, as electrodes break down at fields exceeding a value around 10 MV m^{-1}.

For a given gap voltage and geometry, the current density is proportional to the square root of the charge-to-mass ratio of the particles, $\sqrt{q/m_0}$.

Let us compare the electrostatic potential for a space-charge-limited flow given by Eq. (55), with the potential in the same gap without space charge. In that case, the potential between the anode and the

cathode of the extraction gap (see Fig. 5) can be calculated from Laplace's equation (as there is no space charge, $\rho = 0$),

$$\nabla^2\phi = \frac{\mathrm{d}^2\phi(z)}{\mathrm{d}z^2} = 0, \tag{57}$$

with the solution

$$\phi(z) = -\frac{V_0}{d}\,z. \tag{58}$$

By comparing Eq. (55) with Eq. (58), it can be seen that the space charge of the extracted particle lowers the potential (in absolute value) at any given point between the two electrodes of the planar diode.

3.3 Child–Langmuir law, orders of magnitudes

The Child–Langmuir current for non-relativistic electrons is

$$J_0 = 2.33 \times 10^{-6}\,\frac{V_0^{3/2}}{d^2} \qquad (\mathrm{A\ m^{-2}}), \tag{59}$$

where V_0 is in volts and d in metres. For ions, Eq. (56) becomes

$$J_0 = 5.44 \times 10^{-8}\sqrt{\frac{Z}{A}}\,\frac{V_0^{3/2}}{d^2} \qquad (\mathrm{A\ m^{-2}}), \tag{60}$$

where Z is the charge state of the ion and A is its atomic number. For a given extraction voltage and geometry, the possible current density of electrons is around 43 times higher than that of protons.

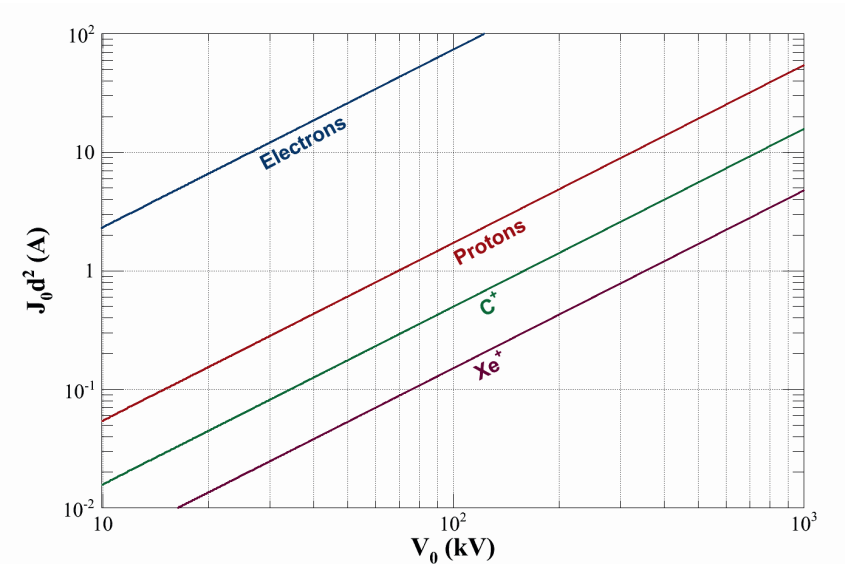


Fig. 6: Normalized space-charge-limited current densities for different charged species

Figure 6 shows a plot the variation of $J_0 d^2$ (in amperes) as a function of V_0 for electrons and some singly charged ions. The values for electrons are plotted only up to 100 kV, as relativistic corrections should be introduced for higher voltages.

4 Space-charge compensation and beam transport in low-energy beam transport lines

4.1 Space-charge compensation

4.1.1 Space-charge compensation principle

Space-charge compensation (SCC), or space-charge neutralization, occurs when a beam is propagating through the residual gas of the low-energy beam transport (LEBT) line (or some additional gas) and subsequently induces ionization of the molecules of this gas (Fig. 7). The secondary particles produced by

ionization (i.e. electrons or ions), which have an opposite polarity to the particles of the beam, are trapped by the beam potential until a steady state is reached. Thus, the low-energy beam can be considered as a plasma that creates a focusing effect that counteracts the space-charge effect.

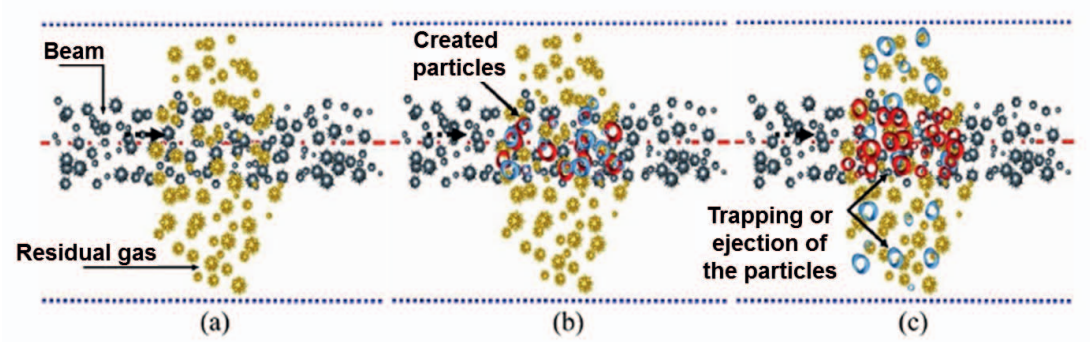


Fig. 7: Space-charge compensation process

As an example, consider a proton beam propagating through a residual gas of H_2. It induces the production of e^-/H_2^+ pairs by ionization according to the following reaction:

$$\mathrm{p} + \mathrm{H}_2 \rightarrow \mathrm{p} + \mathrm{e}^- + \mathrm{H}_2^+ \tag{61}$$

The created electron will be trapped by the beam potential and will contribute to reducing it. The created H_2^+ ion will be expelled from the proton beam.

In this section, basic expressions for the degree of space-charge compensation and characteristic time are given to obtain some orders of magnitude for design and experimental considerations. More elaborate descriptions of the SCC evolution can be found in detailed analytical (see [4,17]) and numerical (see [4,22]) works.

4.1.2 Degree of space-charge compensation

Now, consider a uniform cylindrical beam of radius r_B with an intensity I_B and a longitudinal speed $v = \beta_\mathrm{B} c$ propagating into a cylindrical beam pipe of radius r_P, which is assumed to be grounded (see Fig. 8).

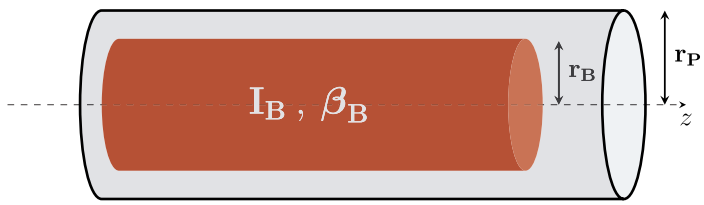


Fig. 8: Uniform cylindrical beam propagating through a beam pipe

If we assume that there is no space-charge compensation, the space-charge electric fields inside and outside the beam are given by Eqs. (15a) and (15b)

$$E_r(r) = \frac{I_\mathrm{B} r}{2\pi\epsilon_0 \beta c r_0^2} \quad \text{if } r \leq r_0, \tag{62a}$$

$$E_r(r) = \frac{I_\mathrm{B}}{2\pi\epsilon_0 \beta c r} \quad \text{if } r > r_0. \tag{62b}$$

By integrating these equations with the boundary condition, $\phi(r_\mathrm{P}) = 0$, we obtain

$$\phi(r) = \frac{I_\mathrm{B}}{4\pi\epsilon_0\beta_\mathrm{B}c}\left[1 + 2\ln\left(\frac{r_\mathrm{P}}{r_\mathrm{B}}\right) - \frac{r}{r_\mathrm{B}^2}\right] \quad \text{if } r \leq r_\mathrm{B}, \tag{63a}$$

$$\phi(r) = \frac{I_\mathrm{B}}{2\pi\epsilon_0\beta_\mathrm{B}c}\ln\frac{r_\mathrm{P}}{r} \quad \text{if } r_\mathrm{B} \leq r \leq r_\mathrm{P}. \tag{63b}$$

The potential on the beam axis (i.e. potential well) created by a uniform beam, without SCC, is given by Eq. (63a) for $r = 0$:

$$\phi_0 = \frac{I_B}{4\pi\varepsilon_0\beta_B c}\left[1 + 2\ln\left(\frac{r_P}{r_B}\right)\right] \tag{64}$$

During the SCC process, the neutralizing particles created by gas ionization are trapped by this potential well. Equation (64) shows that the potential well (i.e. the space-charge force) increases if the radius of the beam decreases. So, achieving a beam waist in a LEBT could be critical for the quality of the beam.

Now, consider a compensated beam at steady state. If we define by ϕ_c its potential on axis, the SCC degree is given by

$$\eta = 1 - \frac{\phi_\mathrm{c}}{\phi_0}. \tag{65}$$

The beam potential well of the compensated beam can be experimentally measured. The values found for the 75 keV, 130 mA proton beam of the LEDA range from 95% to 99% [9]. Along the LEBT, the SCC degree is not constant, as the neutralizing particle trajectories can be modified by external fields of the focusing element, for example. This phenomenon induces strongly non-uniform space-charge forces that can lead to beam emittance growth

4.1.3 Space-charge compensation time

The characteristic SCC transient time, τ_scc, can be determined by considering the time it takes for a particle of the beam to produce a neutralizing particle on the residual gas. It is given by

$$\tau_\mathrm{scc} = \frac{1}{\sigma_\mathrm{ioniz} n_\mathrm{g}\beta_\mathrm{B}c}, \tag{66}$$

where σ_ioniz is the ionization cross-section of the incoming particles on the residual gas and n_g is the gas density in the beam line. The space-charge compensation is expected to reach a steady state after two or three τ_scc times. As an example, the SCC transient time for a 95 keV proton beam propagating in H_2 gas of pressure 5×10^{-5} hPa is 15 µs.

4.2 Beam transport

Only the transport in the LEBT (Low Energy Beam Transport) will be considered here. Nevertheless, the source extraction system is a critical part, especially for a high-intensity injector, as the beam has to be properly formed under strong space-charge forces to be correctly transported in the LEBT. Thus, it seems mandatory to perform simultaneously both the design and simulations of the ion source extraction and the LEBT.

Once the beam is created and extracted from the ion source, it has to be transported and matched by the LEBT to the first accelerating structure like a radio frequency quadrupole (RFQ). The focus can be achieved with electrostatic or magnetic elements. After the ion source, because of the geometry of the extraction system, the beam usually presents a cylindrical symmetry. In order to preserve this symmetry and to simplify the beam tuning, magnetic solenoid lenses or electrostatic Einzel lenses are more commonly used than quadrupoles.

4.2.1 Electrostatic LEBT

In an electrostatic LEBT, the beam is propagating without any space-charge compensation because the neutralizing particles are attracted (or repelled) by the electric field induced by the focusing elements. This kind of beam line is compatible with beam chopping, as there is no transient time for the SCC. Furthermore, the design of electrostatic LEBTs is simplified by the fact that no repelling electrodes for trapping neutralizing particles are needed. So, the beam lines are very compact, which tends to minimize the beam losses by charge exchange. As an example, Fig. 9 shows the SNS ion source with the 12 cm long LEBT equipped with two Einzel lenses [19].

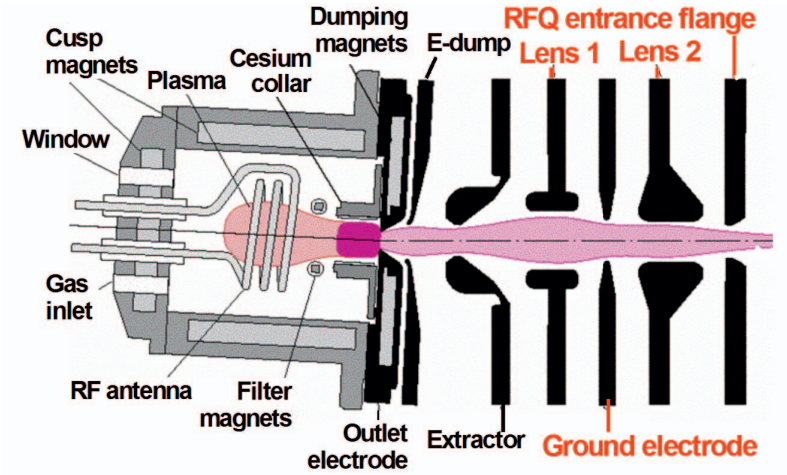


Fig. 9: The SNS ion source and electrostatic LEBT

On the down side, electrostatic LEBTs are vulnerable to beam losses that can lead to high-voltage breakdowns and beam trips. Besides, the Einzel lenses intrinsically induce optical aberrations that create a beam halo and emittance growth. To limit this effect, the beam radius should not exceed two-thirds of the lens aperture radius. Finally, the design of the electrostatic LEBTs is intensity-limited. As the beam is not compensated, its divergence and size will increase rapidly with its intensity (especially for currents of several tens of milliamps). So, it seems difficult to operate the LEBT with a higher current than the design current without expecting beam losses or dramatic emittance growth.

4.2.2 Magnetostatic LEBT

In this case, the beam is fully compensated by the ionization on the residual gas, as explained in the previous section. The gas in the LEBT comes mainly from the ion source, but it has been shown experimentally that the beam emittance can be improved with a higher pressure in the beam line [11]. Besides, the nature of the injected gas has an influence on this emittance improvement. An emittance reduction of a factor of 2 has been reported by replacing H_2 gas by the same partial pressure of Kr [3]. Nevertheless, the gas injection into the beam line has to be done carefully: the higher the pressure, the higher the beam losses by charge exchange. For example, a Kr partial pressure of 4×10^{-5} hPa in a 2 m LEBT leads to a H^+ (100 keV) loss rate due to electron capture of around 2.4%.

For a positive ion beam, another source of compensation particles should be mentioned, even if it is less significant: secondary electrons are produced when a beam hits the beam pipes.

At the end of the LEBT, the electric field of the RFQ tends to penetrate through the injection hole and have a significant effect on the SCC by attracting the neutralizing particles. Moreover, this region is critical from the space-charge point of view, because a beam waist is performed to match the beam for its injection into the RFQ. So, a polarized electrode is placed as close as possible to the RFQ entrance to repel the neutralizing particles in the LEBT and to minimize the uncompensated zone (see Fig. 10).

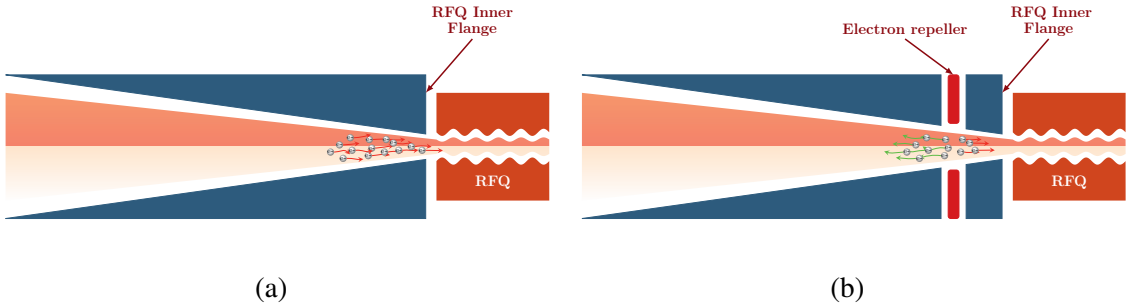


Fig. 10: (a) No electron repeller: neutralizing particles (electrons in this case) are attracted by the RFQ electric field. (b) Electrode repeller located before the RFQ: some neutralizing particles are repelled into the LEBT; the uncompensated zone is minimized.

In a magnetic LEBT, the rise time of the pulsed beams is dominated by the SCC transient time (i.e. several tens of microseconds). A fast chopping system has to be inserted to reach a rise time of the order of hundreds of nanoseconds. In the case of H^- ion beams, a phenomenon of overcompensation occurs during the SCC transient time [2]. When the beam is fully compensated, neutralizing particles (in that case H^+) are still created, but, as they are significantly slower than the electrons, the SCC degree can be greater than 1 during the time it takes for the excess H^+ to be expelled from the beam. During that time, the beam is over-focused and instabilities can be observed.

5 Beam dynamics simulation codes with space charge

5.1 Numerical codes for ion source extraction systems

Some 2D- or 3D-like codes AXCEL-INP [18], PBGUNS [20] and IBSimu [13] have been successfully used to design sophisticated ion source extraction systems as well as electrostatic LEBTs.

With these codes, one can shape the electrodes, compute the generated electric field and track the particle in the defined domain. Over the past few years, elaborate optimization of the geometry of the extraction system has been performed to increase the extracted beam intensity while minimizing the optical aberration and the beam divergence. As an example, the extraction system of the SILHI source, which has an intermediate and a repelling electrode, forming, together with the plasma and grounded electrodes, a pentode extraction system (see Fig. 11) [7], has been developed using AXCEL-INP.

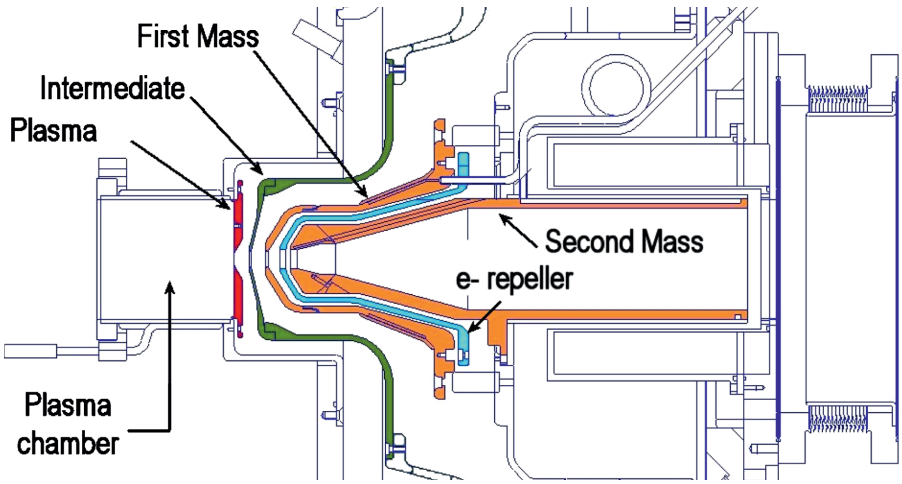


Fig. 11: Five-electrode beam extraction system of the SILHI source

5.2 Numerical codes for beam dynamics transport

In a classical numerical code, the beam is represented by N macro-particles (N is normally less than the actual particle number in the beam) that can be considered as a statistical sample of the beam with the same dynamics as the real particles. The macro-particles are transported through the accelerator step by step and at each time step $\mathrm{d}t$:

- external forces acting on each macro-particle are calculated,
- space-charge fields and the resulting forces are calculated, and
- tyhe equation of motion is solved for each macro-particle.

The space-charge electric field can be computed by a particle–particle interaction (PPI) method or a particles-in-cells (PIC) method, which are briefly described in the next subsections.

5.2.1 Particle–particle interaction method

For each macro-particle i of charge q, it is assumed that the applied space-charge electrostatic field, $\vec{E}_i$, is the sum of all the fields induced by all the other macro-particles:

$$\vec{E}_i = \frac{q}{4\pi\varepsilon_0} \sum_{i \neq j} \frac{\vec{r}_j - \vec{r}_i}{\|\vec{r}_j - \vec{r}_i\|^3}. \tag{67}$$

The advantages of this method are that it is easy to code and the electric field is directly computed on the macro-particles. The main drawbacks are that the method is time-consuming for calculation (proportional to the square of the number of macro-particles) and the obtained space-charge field map is not smooth (the lower the macro-particle number, the more granularity), which can lead to non-physical emittance growth.

5.2.2 Particles-in-cells method

In this case, the physical simulated space is meshed. The mesh geometry depends on the situation. The meshing can be in one, two or three dimensions, depending on the symmetry of the simulated geometry and the beam.

The average beam density at each node of the mesh is obtained by counting the number of particles that are located close to it (an interpolation can induce smoothing). Once the density function is obtained, the field at each node is computed by solving the Poisson equation. Several techniques can be used to solve this equation at each node of the mesh.

1. A direct method. The field is directly calculated at each node of the mesh. The calculation time is proportional to the square of the mesh number.

2. A fast Fourier transform (FFT) method. The field at one node is given by the convolution product of the density and a Green's function. This can be solved by using the fact that a Fourier transform of a convolution product is equal to the product of the Fourier transforms. If n is the mesh number, the calculation time is proportional to $n \log(n)$. One drawback is that the FFT method does not take into account the boundary conditions of the conductors.

3. A relaxation method [14], which is an iterative method. If n is the mesh number, the calculation time is proportional to $n \log(n)$. It can take into account the particular boundary conditions as conductive items, for instance.

Once the field at each node is known, the field at the macro-particle location is calculated by interpolation from the closest nodes. After the evaluation of the beam density and the field for each

particle, the computing time is proportional to the macro-particle number. The algorithm of a PIC code is presented in Fig. 12.

The PIC codes are the most commonly used for space-charge calculations, as they are the fastest and most efficient. A compromise has to be found between the mesh size and the particle number in order reach a sufficient resolution while avoiding some numerical noise that can lead to non-physical effects.

PIC codes that are commonly used for beam dynamics simulations with space charge are: TRACK [1], IMPACT [15] and TraceWin [21].

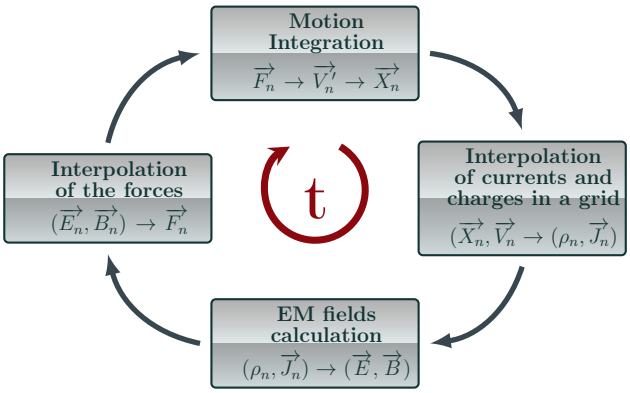


Fig. 12: Algorithm of a PIC code dedicated to particle transport with space charge

5.3 Beam dynamics numerical codes with space-charge compensation

In order to achieve realistic beam transport simulations of high-intensity ion beams at low energy ($\leq$100 keV), it is necessary to take into account the space-charge compensation of the beam by ionization of the residual gas. For that, it is necessary to use self-consistent codes, like WARP [12] or SolMaxP [5]. For example, SolMaxP, has recently been developed at CEA-Saclay and is now used to design and simulate high-intensity injectors.

SolMaxP is a PIC code with an additional module (Monte Carlo algorithm) to simulate the beam interactions with gas (ionization, neutralization, scattering) and beam line elements (secondary emission). The dynamics of the main beam is calculated as well as the dynamics of the secondary particles. Figure 13 shows the SolMaxP algorithm.

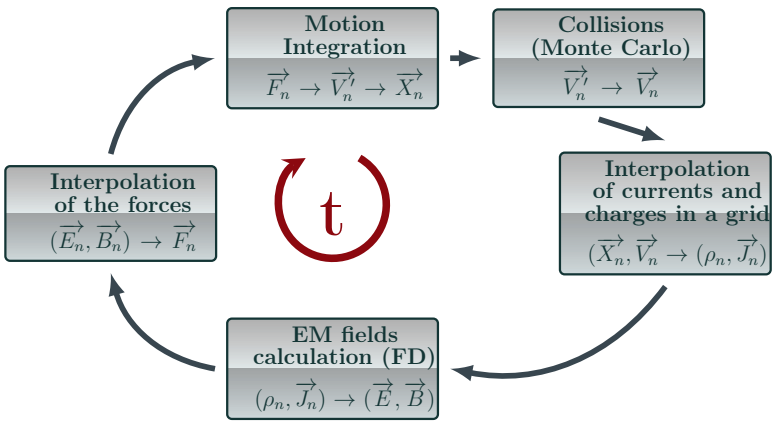


Fig. 13: Algorithm of SolMaxP, a self-consistent code for particle transport with space-charge compensation

The SolMaxP code inputs are the particle distribution of the beam, the applied external fields (e.g. focusing elements, electron repeller) and the beam line geometry and gas pressure. The outputs are the particle distributions (ions, electrons, neutral) all along the beam line and the electric field map derived from the potential map created by the space charge along the beam line.

5.4 Example of numerical simulations of a high-intensity injector: IFMIF

5.4.1 Simulation conditions

First, the modelling of the IFMIF/EVEDA ion source extraction system [8] has been done with AXCEL-INP. The D^+, D_2^+ and D_3^+ particle distributions coming from this model are the input of these simulations.

Then, the simulations presented have been done under the following conditions or hypotheses:

1. D^+, D_2^+ and D_3^+ beams are transported.
2. The electric field map of the source extraction system is included to get relevant boundary conditions.
3. The gas pressure is considered to be homogeneous in the beam line.
4. The gas ionization is produced by ion beam and electron impact.
5. No beam scattering on gas is considered.
6. No secondary electrons are created when the beam hits the beam pipe.

5.4.2 IFMIF injector parameters and layout

The IFMIF injector has to deliver a 140 mA, 100 keV CW D^+ beam of 0.25 π mm mrad emittance. It is composed by a 2.45 GHz electron cyclotron resonance (ECR) source based on the SILHI design and an LEBT with dual solenoid lens focusing system with integrated dipole correcting coils (see Fig. 14) [10].

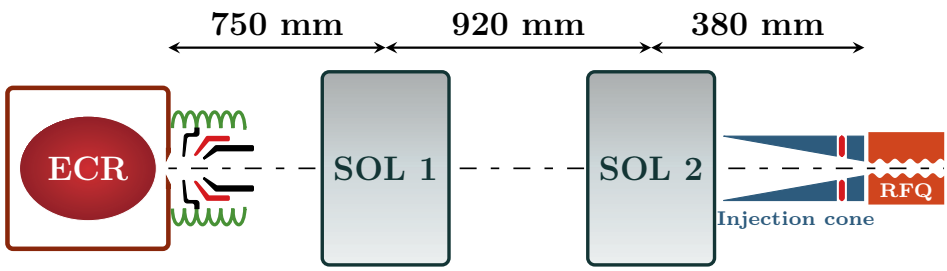


Fig. 14: IFMIF source and LEBT layout

5.4.3 Simulation results

5.4.3.1 SCC transient time

SolMaxP simulations make it possible to determine the SCC transient time. The simulation starts at the time $t = 0$ with no beam in the line but only a fixed gas pressure. Then, the beam starts to propagate in the beam line and its evolution can be followed. The evolution of the D^+ beam emittance measured between the two solenoids is represented on Fig. 15. In this plot, the emittance evolution is given for different pressure conditions in the beam line. Assuming that the D_2 gas contribution to the total pressure in the beam line is 10^{-5} hPa (coming from the ion source), two simulations were done by adding a partial pressure (2 or 4×10^{-5} hPa) of either D_2 gas or krypton, all the other parameters remaining constant.

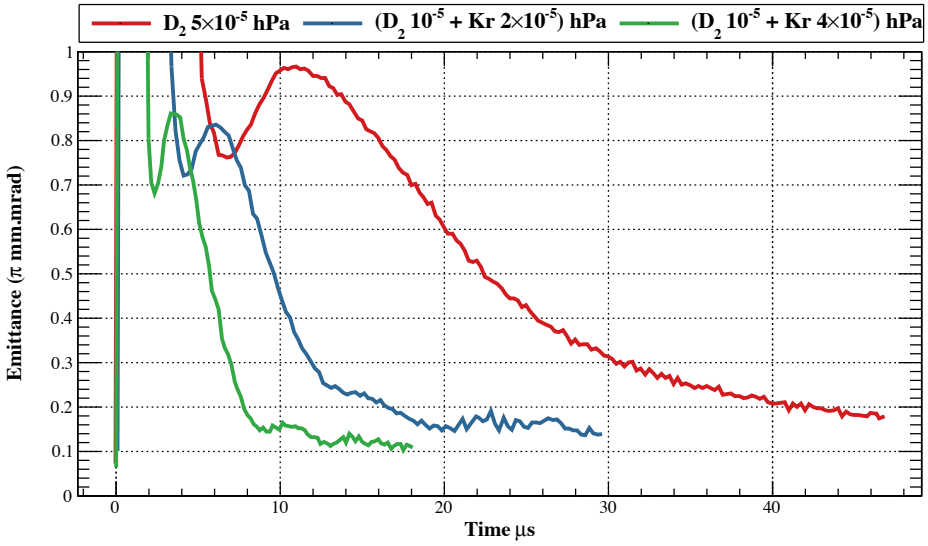


Fig. 15: Emittance evolution versus time for different gas pressure in the IFMIF LEBT between the two solenoids

As expected (see Eq. (66)), the SCC transient time decreases when the beam is transported through a gas with a higher ionization cross-section ($\sigma_{\text{ioniz Kr}} > \sigma_{\text{ioniz D}_2}$). Thus, the simulated space-charge compensation transient time is around 20 µs for a krypton pressure of 2×10^{-5} hPa.

5.4.3.2 SCC potential map

A cut in the $(z0y)$ plane of the space-charge potential in the LEBT, when the space-charge compensation has reached its steady state, is represented on Fig. 16. In this plot, the abscissa $z = 0$ represents the position of the repelling electrode of the source extraction system, while $z = 2.05$ m is the RFQ entrance. The solenoids are respectively located at $z = 0.75$ m and $z = 1.65$ m.

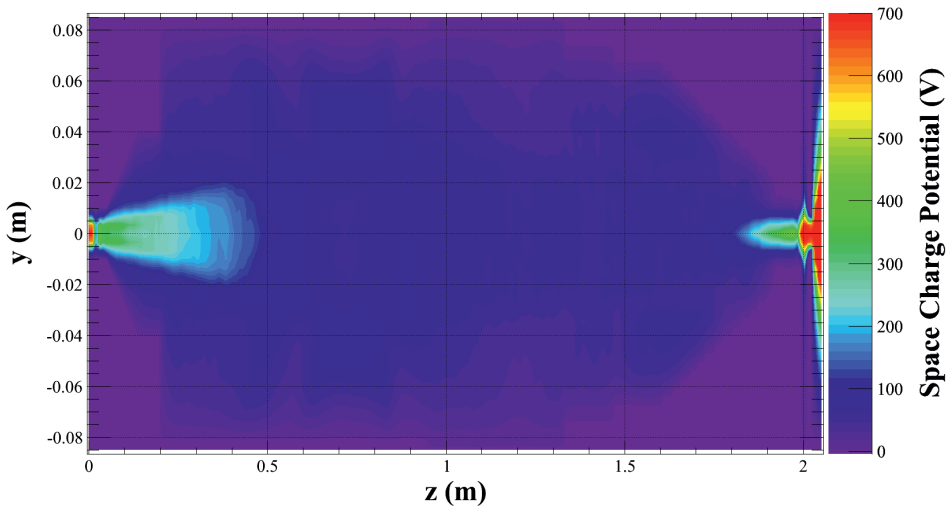


Fig. 16: Two-dimensional cut in the $(z0y)$ plane of a space-charge potential map in the IFMIF/EVEDA LEBT

From the space-charge potential map, the SCC degree can be calculated along the IFMIF/EVEDA LEBT. The potential on the uncompensated beam axis ϕ_0 is calculated with Eq. (65) and the SolMaxP simulations give the potential on the compensated beam axis (see Fig. 16 for $y = 0$). The result is shown

in Fig. 17. It can be observed that, in the ion source extraction region and after the repelling electrode at the RFQ injection, the SCC is poor because the electrons are attracted out of the beam. In the central part of the LEBT, where the solenoids and a drift are located, the SCC degree reaches around 95%, which is compatible with the experimentally measured values [9, 11].

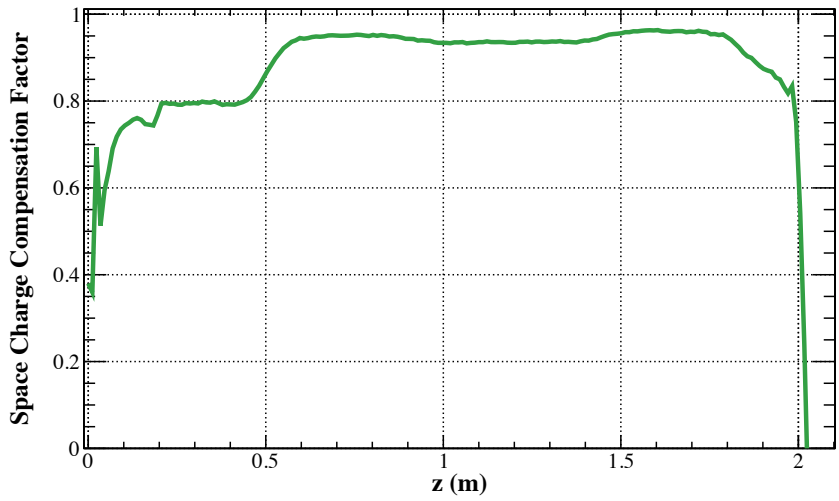


Fig. 17: Space-charge compensation degree along the IFMIF/EVEDA LEBT

5.4.3.3 IFMIF LEBT beam dynamics

The extensive calculations that have been performed for the IFMIF injector led to a very compact design of the LEBT and to an optimization of some parameters, like the position of electron repeller electrode in the injection cone [6].

The beam dynamics simulations showed that the IFMIF/EVEDA deuteron beam can be transported and injected into the RFQ with optimized emittance and Twiss parameters. Under these conditions, the RFQ transmission is above 95%. The transport of the D^+ beam in the injector and in the first section of the IFMIF RFQ is shown on Fig. 18.

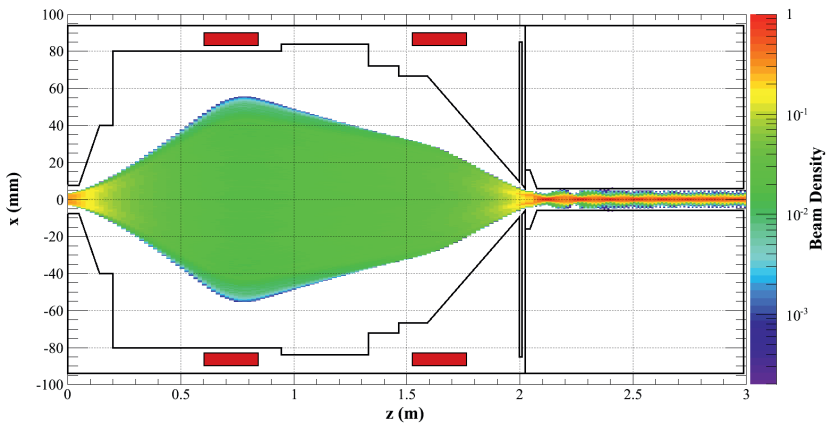


Fig. 18: Beam transport simulation in the IFMIF injector and in the first RFQ cells

The SolMaxP code is in qualitatively good agreement with experimental results, but some quantitative confrontations will be done with experimental results obtained with the beams of the IFMIF injector. Besides, the code will be improved in order to take into account more physical phenomena,

like a precise calculation of the gas pressure along the beam line (the pressure appears to be a critical parameter for the SCC) or the beam scattering on the gas.

References

[1] V.N. Aseev, P.N. Ostroumov, E.S. Lessner and B. Mustapha, TRACK: the new beam dynamics code, Proc. Particle Accelerator Conf., Knoxville, USA, 2005.

[2] R. Baartman and D.H. Yuan, Space charge neutralization studies of an H^- beam, Proc. European Particle Accelerator Conf., Rome, Italy, 1988.

[3] P.-Y. Beauvais, R. Ferdinand, R. Gobin, J.M. Lagniel, P.-A. Leroy, L. Celona, G. Ciavola, S. Gammino, B. Pottin and J. Sherman, Emittance improvement of the electron cyclotron resonance high intensity light ion source proton beam by gas injection in the low energy beam transport, *Rev. Sci. Instrum.* **71**(3) (2000) 1413–1416.

[4] A. BenIsmail, Etude des effets transitoires de la compensation de charge d'espace sur la dynamique d'un faisceau intense, Ph.D. thesis, Université Paris XI, Orsay, 2005.

[5] A. Chancé, N. Chauvin and R. Duperrier, The SolMaxP code, Proc. Int. Particle Accelerator Conf., New Orleans, USA, 2012.

[6] N. Chauvin, M. Comunian, O. Delferrière, R. Duperrier, R. Gobin, P.A.P. Nghiem and D. Uriot, Final design of the IFMIF-EVEDA low energy beam transport line, Proc. Particle Accelerator Conf., Vancouver, Canada, 2009.

[7] O. Delferriere and D. De Menezes, Extraction design and low energy beam transport optimization of space charge dominated multispecies ion beam sources, *Rev. Sci. Instrum.* **75**(5) (2004) 1659–1661.

[8] O. Delferriere, D. De Menezes, R. Gobin, F. Harrault and O. Tuske, Electron cyclotron resonance 140 mA D^+ beam extraction optimization for IFMIF-EVEDA accelerator, *Rev. Sci. Instrum.* **79**(2) (2008) 02B723.

[9] R. Ferdinand, J. Sherman, R.R. Stevens Jr. and T. Zaugg, Space-charge neutralization measurement of a 75-keV, 130-mA hydrogen-ion beam, Proc. Particle Accelerator Conf., Vancouver, Canada, 1997.

[10] R. Gobin, P.-Y. Beauvais, O. Delferriere, D. De Menezes, O. Tuske, G. Adroit, Y. Gauthier and F. Harrault, A 140 mA CW deuteron electron cyclotron resonance source for the IFMIF-EVEDA project, *Rev. Sci. Instrum.* **79**(2) (2008) 02B303.

[11] R. Gobin, P.-Y. Beauvais, R. Ferdinand, P.-A. Leroy, L. Celona, G. Ciavola and S. Gammino, Improvement of beam emittance of the CEA high intensity proton source SILHI, *Rev. Sci. Instrum.* **70**(6) (1999) 2652–2654.

[12] D.P. Grote, A. Friedman, J.-L. Vay and I. Haber, The WARP code: modeling high intensity ion beams, *AIP Conf. Proc.* **749**(1) (2005) 55–58.

[13] T. Kalvas, O. Tarvainen, T. Ropponen, O. Steczkiewicz, J. Ärje and H. Clark, Ibsimu: a three-dimensional simulation software for charged particle optics, *Rev. Sci. Instrum.* **81**(2) (2010) 02B703.

[14] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd edn. (Cambridge University Press, Cambridge, 2007).

[15] J. Qiang, R.D. Ryne, S. Habib and V. Decyk, An object-oriented parallel particle-in-cell code for beam dynamics simulation in linear accelerators, *J. Comput. Phys.* **163**(2) (2000) 434–451.

[16] F.J. Sacherer, RMS envelope equations with space charge, *IEEE Trans. Nucl. Sci.* **18**(3) (1971) 1105–1107.

[17] I.A. Soloshenko, Space charge compensation of technological ion beams, *IEEE Trans. Plasma Sci.* **27**(4) (1999) 1097–1100.

[18] Axcel INP code, http://www.aetassociates.com/software.php?inp.

[19] M.P. Stockli, B. Han, S.N. Murray, T.R. Pennisi, M. Santana and R.F. Welton, Ramping up the spallation neutron source beam power with the H^- source using 0 mg Cs/day, *Rev. Sci. Instrum.* **81**(2) (2010) 02A729.

[20] PBGINS code, http://far-tech.com/pbguns/index.html.

[21] TraceWin code, http://irfu.cea.fr/Sacm/en/logiciels/index3.php.

[22] J.L. Vay and C. Deutsch, Intense ion beam propagation in a reactor sized chamber, *Nucl. Instrum. Methods Phys. Res.* **A464**(1–3) (2001) 293–298.

Further readings

A. BenIsmail, R. Duperrier, D. Uriot and N. Pichoff, Space charge compensation studies of hydrogen ion beams in a drift section, *Phys. Rev. ST Accel. Beams* **10**(7) (2007) 070101.

M. Reiser, *Theory and Design of Charged Particle Beams* (John Wiley & Sons, New York, 2008).

F. Sacherer, RMS envelope equations with space charge, CERN Internal Report, SI/DL/70-12 (1970).

K. Schindl, Space charge, CERN Accelerator School 2003: Intermediate Course on Accelerators, CERN-2006-002 (2006).

T.P. Wangler, *RF Linear Accelerators* (Wiley-VCH, Weinheim, 2008).

A Short Introduction to Plasma Physics

K. Wiesemann
AEPT, Ruhr-Universität Bochum, Germany

Abstract
This chapter contains a short discussion of some fundamental plasma phenomena. In section 2 we introduce key plasma properties like quasi-neutrality, shielding, particle transport processes and sheath formation. In section 3 we describe the simplest plasma models: collective phenomena (drifts) deduced from single-particle trajectories and fundamentals of plasma fluid dynamics. The last section discusses wave phenomena in homogeneous, unbounded, cold plasma.

1 Introduction

Plasma exists in many forms in nature and has a widespread use in science and technology. It is a special kind of ionized gas and in general consists of:

- positively charged ions ('positive ions'),
- electrons, and
- neutrals (atoms, molecules, radicals).

(Under special conditions, plasma may also contain negative ions. But here we will not discuss this case further. Thus in what follows the term 'ion' always means 'positive ion'.) We call an ionized gas 'plasma' if it is *quasi-neutral* and its properties are dominated by electric and/or magnetic forces.

Owing to the presence of free ions, using plasma for ion sources is quite natural. For this special case, plasma is produced by a suitable form of low-pressure gas discharge. The resulting plasma is usually characterized as 'cold plasma', though the electrons may have temperatures of several tens of thousands of Kelvins (i.e. much hotter than the surface of the Sun), while ions and the neutral gas are more or less warm. However, owing to their extremely low mass, electrons cannot transfer much of their thermal energy as heat to the heavier plasma components or to the enclosing walls. Thus this type of cold plasma does not transfer much heat to its environment and it may be more exactly characterized as 'low-enthalpy plasma'.

2 Key plasma properties

2.1 Particle densities

Owing to the presence of free charge carriers, plasma reacts to electromagnetic fields, conducts electrical current, and possesses a well-defined space potential.

Positive ions may be singly charged or multiply charged. For a plasma containing only singly charged ions, the ion population is adequately described by the ion density n_i,

$$n_i = \frac{\text{number of particles}}{\text{volume}}, \quad [n_i] = \text{cm}^{-3} \text{ or } [n_i] = \text{m}^{-3}. \tag{1}$$

Besides the ion density, we characterize a plasma by its electron density n_e and the neutral density n_a.

2.2 Ionization degree, quasi-neutrality

Quasi-neutrality of a plasma means that the densities of negative and positive charges are (almost) equal. In the case of plasma containing only singly charged ions, this means that

$$n_i \approx n_e . \tag{2}$$

In the presence of multiply charged ions, we have to modify this relation. If z is the charge number of a positive ion and n_z is the density of z-times charged ions, the condition of quasi-neutrality reads

$$n_e \approx \sum_z z \cdot n_z . \tag{3}$$

The degree of ionization is defined with the particle densities, not with the charge densities. However, there are two different definitions in use:

$$\eta_i = \frac{\sum_z n_z}{n_a + \sum_z n_z} \quad \text{and} \quad \eta_i' = \frac{\sum_z n_z}{n_a} . \tag{4}$$

Strictly speaking, η_i' is an approximation of η_i for $\eta_i << 1$, which is the usual case. Typical values of η_i for plasmas in ion sources are in the range of 10^{-5} to 10^{-3}. Fully ionized plasma corresponds to $\eta_i = 1$ ($\eta_i' \to \infty$ in this case).

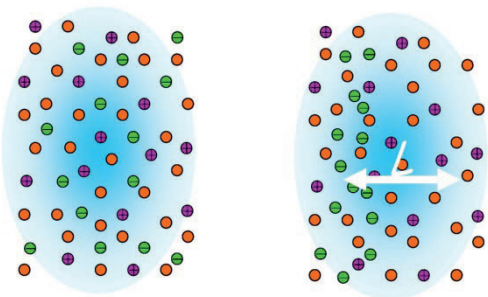


Fig. 1: Charge separation, schematic

To investigate quasi-neutrality further, we assume that a cloud of electrons in plasma has moved to a certain area, forming a negative space charge there. A similar ion cloud is left without electrons in a distance $L \approx \delta x$, forming a positive space charge (see Fig. 1). Thus one obtains between these space-charge clouds an electric field having its maximum value $\overrightarrow{E_{max}}$ at the mutual borders. We can estimate the value of $\overrightarrow{E_{max}}$ by using Poisson's equation:

$$E_{max} = \frac{e \cdot n_i \cdot \delta x}{\varepsilon_0} . \tag{5}$$

The direction of E is such that the electric force drives the two clouds back to overlap. Here e is the positive elementary charge and ε_0 is the permittivity of free space.

For further discussion, we calculate the gain in potential energy W_{pot} of a charged particle after moving by δx through the space-charge layer:

$$W_{\text{pot}} = \int_0^{\delta x} eE\,\mathrm{d}x = \frac{e^2 n_{\text{e}} (\delta x)^2}{2\varepsilon_0}\,. \tag{6}$$

The only energy available for this purpose is the thermal energy of electrons (and ions, but, owing to their usually low temperature compared to electrons, ion thermal effects can be neglected in cold plasma, i.e. in the plasma of interest for ion sources), at the average $\frac{1}{2}k_{\text{B}}T_{\text{e}}$ for a movement in one degree of freedom. Thus we may expect deviations from quasi-neutrality on a scale defined by

$$W_{\text{pot}} = \frac{1}{2}k_{\text{B}}T_{\text{e}}\,. \tag{7}$$

This corresponds to a charge separation over the so-called Debye–Hückel length λ_{D} [1] given by

$$\lambda_{\text{D}} = \left(\frac{\varepsilon_0 k_{\text{B}} T_{\text{e}}}{e^2 n_{\text{e}}}\right)^{1/2}\,. \tag{8}$$

A numerical value of this length is given by $\lambda_{\text{D}}/\text{m} = 7.434\times10^3 (k_{\text{B}}T_{\text{e}}/\text{eV})^{1/2}/(n_{\text{e}}/\text{m}^{-3})^{1/2}$ (see also Fig. 2).

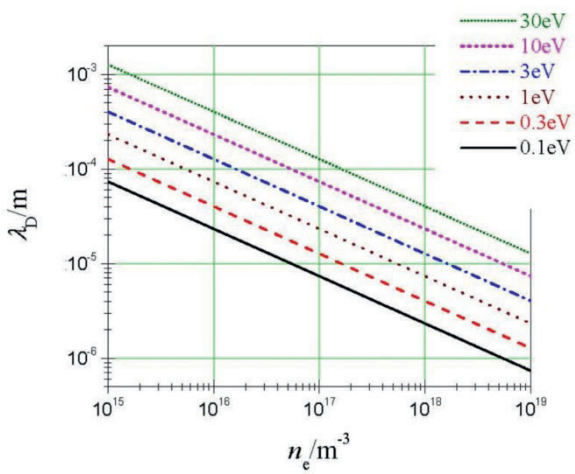


Fig. 2: Debye length versus plasma density and electron temperature

We may, on the other hand, ask what amount $\Delta n = |n_{\text{e}} - n_{\text{i}}|$ of deviation from quasi-neutrality is possible over a given length L. Again, we have only the thermal energy at our disposal. Thus

$$\frac{1}{2}k_{\text{B}}T_{\text{e}} \approx \frac{1}{2}\frac{e^2}{\varepsilon_0}\cdot\Delta n\cdot L\,. \tag{9}$$

When substituting $k_B T_e$ by λ_D we obtain as an estimate

$$\frac{\Delta n}{n} \approx \left(\frac{\lambda_D}{L}\right)^2 . \tag{10}$$

We may formulate the condition of quasi-neutrality as $\Delta n << n_e, n_i$. According to Eq. (10), this is equivalent to $L >> \lambda_D$. This means that the extension of an ionized gas must be large compared to the Debye–Hückel length, in order to fulfil the conditions of being plasma. Plasma quasi-neutrality is defined only on a large macroscopic scale. If we inspect plasma on a microscopic scale, we may find deviations from neutrality increasing with decreasing scale length.

In plasmas powered electrically (i.e. in any electric discharge), the electrons gain energy more easily from the external electric field than the inert ions, as the electrons are much lighter and thus electronic currents are much larger. Since in elastic collisions electrons can transfer kinetic energy only in small amounts – of the order of m_e/m_i – to the ions (m_e and m_i are the electronic and ionic masses, respectively), in steady state the electron temperature will be much higher than the ion (and neutral) temperatures, as discussed above. Thus, electrons are mainly responsible for local deviations from neutrality and the temperature in the formula of the Debye–Hückel length is the electron temperature T_e.

In a typical low-power ion source plasma, the electron temperature T_e is of the order 30 000–40 000 K, while the ion temperature T_i is around 500–1000 K. The electron density n_e amounts to about $10^{10}\,\mathrm{cm}^{-3} = 10^{16}\,\mathrm{m}^{-3}$ and higher. Under these conditions, the Debye–Hückel length is of the order of 0.12–0.16 mm and shorter.

(In many plasma physics texts, the symbol T stands for the product of temperature with the Boltzmann constant k_B, i.e. for the energy $k_B T$ measured in electronvolts (eV), instead of for the thermodynamic temperature. This characteristic energy is dubbed 'temperature measured in eV'. A temperature of 11 600 K corresponds to a characteristic energy of 1 eV.)

2.3 Plasma oscillations

The value of the electric field $\vec{E}$ created by charge separation is, as we have seen, proportional to the separation length, which we now call x:

$$E = \frac{e}{\varepsilon_0} n x . \tag{11}$$

Thus we obtain for the movement of, say, electrons, under the action of the restoring force $F = eE$,

$$F = eE = \frac{e^2}{\varepsilon_0} n x = m_e \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} . \tag{12}$$

This is the equation of a harmonic oscillator with the eigenfrequency

$$\omega_{pe} = \left(\frac{e^2 n}{\varepsilon_0 m_e}\right)^{1/2} , \tag{13}$$

the so-called (angular) electron plasma frequency. A numerical value of the (electron) plasma frequency is given by $\omega_{\mathrm{pe}} = 2\pi \times 8.9 \times \sqrt{n_{\mathrm{e}} / \mathrm{m}^{-3}}$. For the plasma data given above, this yields $\omega_{\mathrm{pe}} = 2\pi \times 8.9 \times 10^{8}\,\mathrm{s}^{-1}$.

What we describe here are oscillations of the electron charge cloud as a whole (see Figs. 1 and 3). The inert ions are considered to remain at rest. A more careful analysis will reveal that, instead of these oscillations, different types of acoustic waves can propagate in plasma. However, the electron and ion plasma frequencies will show up as important parameters for characterizing these different types of plasma waves.

By replacing the electron mass by the ion mass in Eq. (13), one obtains the (angular) ion plasma frequency. It is the natural frequency of ion space charge and may play a role in the ion sheaths in front of a wall or between plasma meniscus and extraction hole at the output of an ion source. In plasma, ion acoustic waves are strongly damped at this frequency – see the discussion below.

Fig. 3: Plasma oscillations

Our considerations show that neutrality is a dynamic equilibrium state of plasma from which deviations are possible on time scales defined by the (electron) plasma frequency and extending over spatial dimensions of the order of the Debye–Hückel length. These deviations are powered by the thermal energy of the charged plasma constituents and tend to decay into the neutral equilibrium state. Thus plasmas are always close to neutrality – they are quasi-neutral.

2.4 Plasma as a gas

A gas is described adequately by single-particle properties averaged over the particle distribution functions and parameters like pressure, temperature and density, which can be correlated to those averages, as we know from kinetic theory.

Plasma kinetic theory is classical Boltzmann statistics, if the distance between particles (electrons, ions, neutrals) is sufficiently large (*classical plasma*). For electrons, this is the case if their average distance,

$$\lambda_{\mathrm{n}} = 1/(n_{\mathrm{e}})^{1/3}\,, \tag{14}$$

is large compared to the average electron de Broglie wavelength λ_{B},

$$\lambda_{\mathrm{B}} = h/m_{\mathrm{e}} v_{\mathrm{th}}\,, \text{ with } \tfrac{1}{2} m_{\mathrm{e}} v_{\mathrm{th}}^{2} = k_{\mathrm{B}} T_{\mathrm{e}}\,. \tag{15}$$

Otherwise plasma is *degenerate*.

A plasma can be described as an *ideal gas* if the mutual potential energy of electrons and ions is small compared to the average kinetic energy $\frac{3}{2} k_{\mathrm{B}} T_{\mathrm{e}}$, that is, if

$$\frac{3}{2}k_{\mathrm{B}}T_{\mathrm{e}} >> \frac{e^2}{4\pi\varepsilon_0\lambda_{\mathrm{n}}}. \tag{16}$$

Substituting $k_{\mathrm{B}}T_{\mathrm{e}}$ by the Debye–Hückel length, we obtain as equivalent conditions:

$$\lambda_{\mathrm{D}} >> \lambda_{\mathrm{n}} = \frac{1}{n_{\mathrm{e}}^{1/3}} \tag{17}$$

or

$$g = \left(\frac{1}{n_{\mathrm{e}}}\frac{4\pi}{3}\lambda_D^3\right)^{-1} << 1. \tag{18}$$

The expression in the bracket is the number of electrons in a so-called Debye–Hückel sphere, that is, in a sphere with the volume $\frac{4\pi}{3}\lambda_{\mathrm{D}}^3$. Its reciprocal g is the 'plasma parameter', taken in plasma theory as a measure for degeneracy or absence of degeneracy in plasma.

Plasma with large plasma parameter is 'non-ideal' or 'strongly coupled classical plasma'. Now

$$g \propto \frac{n_{\mathrm{e}}^{1/2}}{(k_{\mathrm{B}}T_{\mathrm{e}})^{3/2}}. \tag{19}$$

Thus non-ideal classical plasma is very cold and very dense. In such a case, correlations between the plasma particles may become important. Under laboratory conditions, such correlations can be observed in dusty plasmas, where dust particles sometimes adjust themselves into regular structures. In ion source plasma we have $g << 1$ as a rule. Such plasma behaves classically, that is, obeys classical Boltzmann statistics, it is in general non-degenerate and the conditions in Eqs. (17) and (18) are fulfilled.

2.5 Particle transport in plasma

We restrict our discussion to drift and diffusion, transport processes important in ion source plasma. These processes are characterized by 'transport coefficients', of which we discuss mobility *b*, conductivity σ and diffusion coefficient *D*.

2.5.1 Mobility and conductivity

To understand the concept of mobility, we consider a simple method for measuring viscosity in a viscous fluid, the falling sphere method. Under the action of gravity, a small metallic sphere will, after a short distance, fall with constant speed, which can be taken as a measure of viscosity. In ion source plasma, viscosity is a higher-order effect and can be neglected. We will not consider it further. However, under the common action of a constant force due to, say, an electric field $\vec{E}$ and friction due to collisions with other particles, a charged particle may attain a constant drift velocity $\overrightarrow{v_{\mathrm{D}}}$, which under favourable conditions is proportional to the value of the force acting, that is, to qE. Here q is the particle charge. The proportionality constant is the *mobility* b:

$$\overrightarrow{v_{\mathrm{D}}} = bq\vec{E}. \tag{20}$$

Here we follow the mobility definition given by Allis [2]. However, a warning: many authors use a different mobility definition:

$$\overrightarrow{v_{\mathrm{D}}} = \tilde{b}\vec{E} \,. \tag{21}$$

The advantage of the definition according to Eq. (20) is that b is positive by definition, whereas $\tilde{b}$ carries the sign of the charge q, that is, $\tilde{b}$ is negative for electrons and positive for positive ions, thus causing some complications, especially when defining the conductivity.

The density of the electric current $\overrightarrow{j_{\mathrm{e}}}$ carried by drifting charged particles is given by

$$\overrightarrow{j_{\mathrm{e}}} = qn_{\mathrm{q}}\overrightarrow{v_{\mathrm{D}}} = q^2 n_{\mathrm{q}}\vec{E} = \sigma\vec{E} \,. \tag{22}$$

Here we have used Eq. (20) to eliminate $\overrightarrow{v_{\mathrm{D}}}$ and to define the *conductivity* σ. In our definition the conductivity also comes out to be positive by definition.

In weakly ionized plasma, friction is due to charged particle collisions with neutrals. For electrons, one obtains from kinetic theory

$$b_{\mathrm{e}} = \langle 1/m_{\mathrm{e}}\nu_{\mathrm{en}}\rangle = \langle 1/m_{\mathrm{e}}Q_{\mathrm{en}}v_{\mathrm{e}}\rangle \propto 1\big/\sqrt{m_{\mathrm{e}}} \,. \tag{23}$$

Here the angle brackets indicate an average over the electron distribution function, ν_{en} is the collision frequency for momentum transfer collisions between electrons and neutrals, Q_{en} is the respective cross-section for momentum transfer, and v_{e} is the velocity of a single electron. The averages are defined by

$$\langle 1/m_{\mathrm{e}}Q_{\mathrm{en}}v_{\mathrm{en}}\rangle = \iiint_{\overrightarrow{v_{\mathrm{e}}}} \left(1/m_{\mathrm{e}}Q_{\mathrm{en}}v_{\mathrm{en}}\right) f(\overrightarrow{v_{\mathrm{e}}})\,\mathrm{d}^3 v_{\mathrm{e}} \,, \tag{24}$$

where $f(\overrightarrow{v_{\mathrm{e}}})$ is the electron velocity distribution function. Our formulas are valid if the value of drift velocity $\overrightarrow{v_{\mathrm{D}}}$ is small compared to the average value of the electron velocity

$$\langle v_{\mathrm{e}}\rangle = \sqrt{8m_{\mathrm{e}}/\pi k_{\mathrm{B}}T_{\mathrm{e}}} \,. \tag{25}$$

Otherwise $f(\overrightarrow{v_{\mathrm{e}}})$ must be replaced by the distribution function of the drifting electrons, and b and σ will become functions of the electron drift velocity v_{eD}. Equation (20) describes a stationary equilibrium between electric force and friction. Under the condition of this equilibrium, the energy taken by the drifting particles is completely transformed into heat of those particles. In fusion research, this process constitutes an important mechanism for plasma (electron) heating and is called 'ohmic heating'. In the case of fully ionized fusion plasma, the necessary friction is produced by electron–ion collisions.

A detailed analysis reveals that such equilibrium is possible only if the decrease of the friction force with increasing absolute value of v_{eD} is sufficiently slow. The friction force decreases because of the decrease of the Coulomb collision cross-sections with increasing collision energy. In fully ionized fusion plasma, the condition for equilibrium is fulfilled for electrons as long as

$$\tfrac{1}{2} m_{\mathrm{e}} v_{\mathrm{eD}}^2 \le k_{\mathrm{b}} T_{\mathrm{e}} \,. \tag{26}$$

Otherwise electrons may be continuously accelerated to relativistic energies, an effect known as 'run-away' [3]. To avoid run-away, the drift velocity must be guided to rise sufficiently slowly that Eq. (26) is never violated. A similar effect may also occur in weakly ionized plasma because all collision cross-sections with neutrals decrease at higher energies [4]. In general, the equilibrium between collisional friction and an external force is a phenomenon at low energies.

2.5.2 Diffusion

Let us consider a virtual plane somewhere in homogeneous plasma. Owing to Brownian motion, there is a continuous flow of electrons, ions and neutrals from both sides through this plane. According to a famous formula of gas kinetics, its (particle) current density Γ is given by

$$\Gamma = n_{\mathrm{q}} \sqrt{m_{\mathrm{e}} / 2\pi k_{\mathrm{B}} T_{\mathrm{e}}} \,. \tag{27}$$

As the particle densities and temperatures in homogeneous plasma are equal on both sides, these flows will also be equal for any particle sort and compensate each other.

In the case of density (and temperature) gradients, however, the flows no longer equal each other, and a net flow results in the direction of decreasing particle density. This transport process is *diffusion*. It can be described by the equations known as Fick's laws. According to the first of Fick's laws, the diffusive particle flux $\overrightarrow{\Gamma_{\mathrm{diff}}}$ is given by

$$\overrightarrow{\Gamma_{\mathrm{diff}}} = -D \nabla n \,. \tag{28}$$

It follows from our discussion that the value of the current density of the diffusive flux is always smaller than the flux given in Eq. (27).

The diffusion constant D is related to the respective mobility by the famous Einstein relation

$$D_{\mathrm{e,i}} = b_{\mathrm{e,i}} k_{\mathrm{B}} T_{\mathrm{e,i}} = k_{\mathrm{B}} T_{\mathrm{ei}} / m_{\mathrm{ren,in}} \nu_{\mathrm{en,in}} \approx \langle v_{\mathrm{e,i}}^2 \rangle / \nu_{\mathrm{en,in}} \approx \nu_{\mathrm{en,in}} \langle \lambda_{\mathrm{en,in}} \rangle^2 \,, \tag{29}$$

if the mobility definition of Eq. (20) is used. Here the brackets $\langle \cdots \rangle$ stand for an average over the particle distribution function. The last two approximate expressions in Eq. (29) are very useful for estimating orders of magnitude. From relations (29) and Eq. (23) and the respective formula for the ion mobility it follows that

$$D_{\mathrm{e}} / D_{\mathrm{i}} = \sqrt{m_{\mathrm{rin}} / m_{\mathrm{e}}} \,, \tag{30}$$

that is, in weakly ionized (and non-magnetized) plasma, electrons diffuse much faster than ions. Here m_{rin} is the reduced mass of the ion–neutral collision system as used in collision theory. Also m_{e} stands for the reduced mass of the electron–neutral collision system here and in Eq. (23). However $m_{\mathrm{re}} \approx m_{\mathrm{e}}$ because m_{e} is so small. Equation (29) is in principle also valid in fully ionized plasma, if that is sufficiently close to equilibrium. In that case, the collision frequencies of electrons and ions with neutrals must be replaced by the electron–ion collision frequencies. 'Close to equilibrium' means that the distribution function of the diffusing particles is very close to a Maxwellian and any drift velocity is very small. In more complicated situations, D must be calculated from kinetic theory, which is not treated here. The interested reader may find a discussion in books on plasma kinetic theory – see for instance Ref. [5].

2.6 Sheath formation

Ion source plasma is enclosed by a vessel. Its walls are sinks for charged particles, causing continuous flows of electrons and ions to the walls.

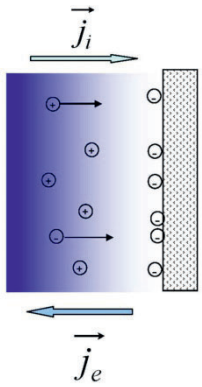


Fig. 4: Electron and ion flows from plasma to a wall (schematic)

Ions may stick on or in a wall or be neutralized and re-emitted as neutrals. Electrons may stick on a dielectric wall and bind in low-energy surface states, be absorbed by metallic walls, or be re-emitted and reflected. Further ion (and metastable) and electron bombardment of a wall may cause secondary electron emission. However, in our simplified discussion, we will neglect back-flows from the walls as a secondary effect. The electron and ion flows towards a wall correspond to electric currents with antiparallel current densities (see Fig. 4). Owing to their low mass, the transport of electrons is much faster than ion transport. This leads in the steady state to the formation of a tiny positive space charge in plasma and a negative charge on a wall. Because of these charges, the plasma potential is positive and the wall potential is negative. In the case of a dielectric wall, the electric current to the wall must be zero, the same holding true for the electric current density $\vec{j}$:

$$\vec{j} = \vec{j}_\mathrm{e} + \vec{j}_\mathrm{i} = 0 \,. \tag{31}$$

The charges in plasma and on the walls regulate themselves in such a way that Eq. (31) is fulfilled. Assuming the plasma potential to be zero and that the electrons obey a Maxwell distribution function, we can rewrite Eq. (31) as

$$\vec{j}_\mathrm{i} = -\vec{j}_\mathrm{e} = -en_\mathrm{e}\left(\frac{m_\mathrm{e}}{2\pi k_\mathrm{B}T_\mathrm{e}}\right)\exp\left(\frac{eU_\mathrm{wall}}{k_\mathrm{B}T_\mathrm{e}}\right). \tag{32}$$

Here e is the *positive* elementary charge. Using Eq. (32) we may easily calculate the wall potential U_wall. (It is also equal to the potential of a floating probe.) In the case of metallic walls, Eq. (31) may not necessarily be fulfilled, only the respective condition for the total electric currents. This is of special importance for the case of magnetized plasma. In this case diffusion across the magnetic field differs from diffusion along the magnetic field lines. Thus compensating currents may flow in a metallic vessel. These currents greatly influence plasma containment in such cases. We will discuss this problem further in section 3.2.

3 Plasma modelling

3.1 Movement of charged particles under the action of electric and magnetic fields

Consider a single singly charged particle (mass m, charge q) in a combination of a stationary force field $\vec{F}$ not depending on the particle velocity (like the force $q\vec{E}$ in an electric field $\vec{E}$), and a stationary magnetic field with induction $\vec{B}$. We denote by $\vec{v}$ the velocity of this particle and by $\dot{\vec{v}}$ the time derivative of velocity, its acceleration. The equation of motion in vector notation,

$$m\dot{\vec{v}} = q\vec{v} \times \vec{B} + \vec{F}\,, \tag{33}$$

constitutes a set of three differential equations. The first term on the right-hand side has only components perpendicular to $\vec{B}$. Thus the equation for the motion in the direction parallel to $\vec{B}$ is independent of $\vec{B}$ and describes a constant acceleration parallel to $\vec{B}$ by the respective component of $\vec{F}$, denoted by $\vec{F}_{\parallel}$. The components of $\vec{F}$ and $\vec{v}$ perpendicular to $\vec{B}$ and their time derivatives we will term $\vec{F}_{\perp}$, $\vec{v}_{\perp}$, and so on.

Thus, *perpendicular* to $\vec{B}$, we have a system of two coupled inhomogeneous differential equations:

$$m\dot{\vec{v}}_{\perp} = q\vec{v}_{\perp} \times \vec{B} + \vec{F}_{\perp}\,. \tag{34}$$

The solution is a combination of the general solution of the homogeneous part of these equations and a special solution of the inhomogeneous equation. The homogeneous part is given by

$$m\dot{\vec{v}}_{\perp} = q\vec{v}_{\perp} \times \vec{B}\,. \tag{35}$$

It describes the motion of our particle under the action of the magnetic field only. The solution is well known. The motion is a gyration in the plane perpendicular to $\vec{B}$, that is, a motion with constant velocity on a circle, the so-called cyclotron motion (see Fig. 5). The radius of the circle, the cyclotron radius, r_{B}, is given by

$$r_{\mathrm{B}} = \frac{mv_{\perp}}{qB}\,. \tag{36}$$

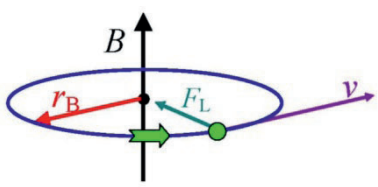


Fig. 5: Cyclotron motion of a charged particle

Here $v_{\perp}$ is the speed of the particle in the plane perpendicular to $\vec{B}$, thus $mv_{\perp} = p_{\perp}$ is the absolute value of the respective momentum. The respective angular frequency, the cyclotron frequency ω_{B} is given by

$$\omega_{\mathrm{B}} = \frac{v_{\perp}}{r_{\mathrm{B}}} = \frac{qB}{m} . \tag{37}$$

It does not depend on the particle velocity – at least in the non-relativistic case. Note that the charge q may be either positive or negative. In our definition, ω_{B} has the same sign – it defines the rotational direction.

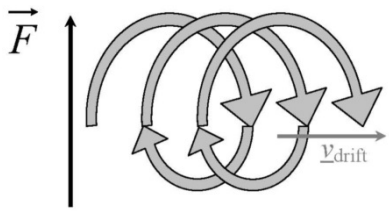


Fig. 6: Drift of a charged particle due to the combined action of a constant magnetic field with the induction *B* and a constant force *F*.

We obtain a special solution of the inhomogeneous equation by assuming the velocity $\vec{v}_{\perp}$ to be constant, which means zero acceleration. In this case we obtain

$$-q\vec{v}_{\perp} \times \vec{B} = \vec{F}_{\perp} . \tag{38}$$

To solve for the velocity we generate the cross-product of this equation with $\vec{B}$ from the right, yielding

$$qB^2 \vec{v}_{\perp} = \vec{F}_{\perp} \times \vec{B} \quad \text{or} \quad \vec{v}_{\perp} = \frac{\vec{F} \times \vec{B}}{qB^2} \equiv \vec{v}_{\mathrm{drift}} . \tag{39}$$

The complete particle motion in a plane perpendicular to $\vec{B}$ is a superposition of a gyration and a constant velocity $\vec{v}_{\mathrm{drift}}$ in the direction perpendicular to $\vec{B}$ and $\vec{F}$. The cyclotron motion depends on the initial velocities of the particles. These velocities are distributed at random. When we average over all particles, the mean value of the gyration velocity will be zero. In contrast to that, at least for all particles of equal charge $\vec{v}_{\mathrm{drift}}$ is the same. Its value and direction depend only on the direction and magnitude of the magnetic induction and the external force $\vec{F}$. The whole population of equal particles will drift in the same direction. In this way we have reduced the discussion of particle motion to the discussion of the motion of the guiding centre for gyration as a pseudo-particle. This strategy is known as the guiding centre approximation. Most important is the case of the force due to an electric field, $\vec{F} = q\vec{E}$. In this case the charge cancels out in Eq. (39). All charged particles drift with the same velocity in the same direction ($E \times B$ drift).

We have pictured this drift by considering the influence of the external force on the gyration (see Fig. 6): imagine a magnetic field with the field lines pointing perpendicular into the plane of the paper. Under this condition, the gyration is restricted to the plane of the paper (or a plane parallel to it). The case shown in Fig. 6 is that of a particle with positive charge under the influence of an electric field. When the particle moves upwards, it is accelerated and the radius of its gyration circle increases. Downwards it is decelerated and the radius of the gyration circle decreases. Thus its trajectory is a cycloid instead of a closed circle.

In the case of a negatively charged particle, the direction of the gyration is reversed. If the force is due to an electric field, the direction of the force *F* is also reversed and the drift, therefore, goes in

the same direction as before. If the direction of the force is independent of the particle charge (like, for example, in the case of inertial or gravitational forces), the direction of the drift will be reversed. In this latter case positively and negatively charged particles drift in opposite directions. The result may be an electric current or charge separation.

If only the magnetic field strength decreases in the direction indicated by the arrow for F, we get a similar effect because of the dependence of the gyration radius on B. This so-called gradient drift is also an example of current transport by a drift. Any effect changing the gyration radii in a similar way can thus cause a drift. A different approach to drift induced by a gradient of the magnetic induction B is by considering the gyrating particle as a pseudo-particle with a magnetic moment $\vec{M}$. In analogy to a current-carrying wire loop, we can treat the gyrating particle as a circular current $I = q / \tau B = q\omega B / 2\pi$. The area A within the loop is $A = r_B^2 \pi$. Then

$$M = I \cdot A . \tag{40}$$

This is a magnetic dipole experiencing a force in the presence of a magnetic field gradient. This force can be considered as the reason for the drift and can be treated in a similar way as we have treated the force due to an electric field. The direction of $\vec{M}$ is always opposite to the direction of $\vec{B}$ (see Fig. 7). The plasma is *diamagnetic*!

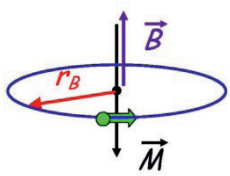


Fig. 7: Magnetic moment of a gyrating charged particle

If the direction of the magnetic field gradient is parallel to $\vec{B}$, we have a special effect. It turns out that the magnetic moment due to a gyrating particle moving along $\vec{B}$ is constant if the gradient of $\vec{B}$ is sufficiently small (i.e. $\vec{M}$ is an *adiabatic invariant*). Using the relations for the current I and the area of the gyration circle A given above, we obtain

$$M = I \cdot A = W_\perp / B . \tag{41}$$

Here $W_\perp = \frac{1}{2} m v_\perp^2$ is the kinetic energy due to the gyration of the considered charged particle, that is, due to the motion perpendicular to the direction of $\vec{B}$.

Further (assuming the potential energy to be zero), the total energy

$$W = W_\perp + W_\parallel \tag{42}$$

is also a constant of motion. If the particle moves in the direction of increasing magnetic field, $W_\perp$ must increase, because M is constant. Thus, $W_\parallel$, the energy of the motion parallel to $\vec{B}$, must decrease by the same amount to keep W constant. If $W_\parallel$ becomes zero, the particle cannot proceed further and must return. A configuration with B increasing along B is therefore called a magnetic mirror (Fig. 8).

In our discussion we have tacitly assumed that collisions between plasma particles are not important. If collisions are very frequent, cyclotron motion and drift will be disturbed. We can estimate the condition for drift by comparing the average frequency of gyration, $\omega_B = v_{th}/r_B = qB/m$, with the average collision frequency ν_c (note the difference in symbols: frequency ν and velocity v). If at least for one kind of charged particle $\omega_B >> \nu_c$, drift and gyration are fully developed for those particles. We call such plasma *magnetized*. Magnetization can, at least in principle, always be attained by a sufficiently strong external magnetic field. For a given magnetic field, magnetization depends on plasma and neutral density. It is strongest in dilute plasma at low pressure.

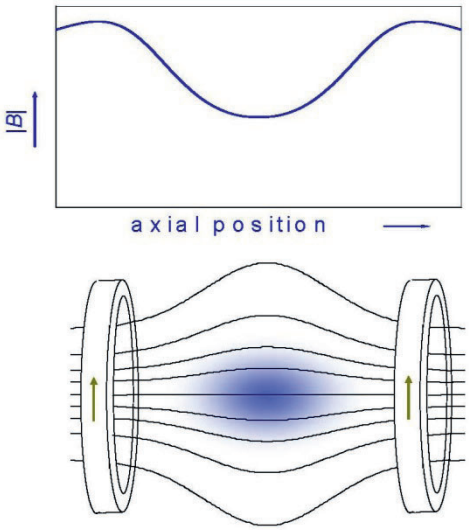


Fig. 8: Magnetic mirror configuration showing the shape of the magnetic field lines on a plane through the axis of the system (below) and the curve of the absolute value of the induction $\vec{B}$ on the axis of the system.

3.2 Diffusion in magnetized plasma

Under the action of an external magnetic field with induction $\vec{B}$, plasma becomes anisotropic. One consequence is that there is a difference between transport, such as diffusion, along and across the magnetic field. Along the magnetic field, diffusion resembles transport in non-magnetized plasma. Let $D_{\|e}$ and $D_{\|i}$ be the electron and ion diffusion coefficients for this case. They obey the same relation as in the absence of a magnetic field (see Eq. (30)):

$$\frac{D_{\|e}}{D_{\|i}} = \sqrt{\frac{m_{ri}}{m_e}}, \tag{43}$$

the same holding true for mobility and conductivity.

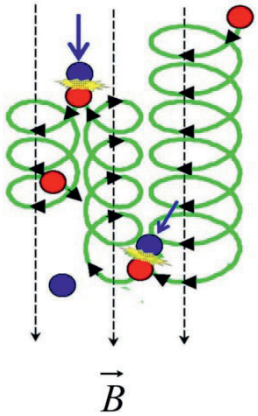


Fig. 9: Trajectories of the Brownian motion of a charged particle in a magnetic field

To understand diffusion across the magnetic field, we first consider the Brownian motion of, say, electrons in magnetized plasma (see Fig. 9). In Brownian motion, the trajectories of neutral particles will be straight between collisions and are bent by collisions. The resulting trajectory is a random zigzag path. In plasma, a strict distinction between free motion and collision is not possible: the trajectories of charged particles may be curved due the far-reaching Coulomb field.

In the presence of an external magnetic field (induction $\vec{B}$), free-moving charged particles spiral around a magnetic field line. If the cyclotron frequency of the spiralling particles is large compared to any collision frequency, one has the situation sketched in Fig. 9. Owing to collisions, particles hop from field line to field line. As discussed above, the picture of separate collisions and free flight in between must be modified in plasma. Here the particle motion is an $E \times B$ drift under the action of a randomly fluctuating electric field. However, calculation of diffusion coefficients leads to similar results in both models.

Taking the situation in Fig. 9 as given, we can conclude that the average distance travelled between successive collisions across the magnetic field is not the mean free path but the average cyclotron radius $\langle r_{\mathrm{B}} \rangle$ – compare Eq. (36). Thus we may estimate the diffusion coefficient for diffusion across the magnetic field by replacing in the last expression $\langle \lambda \rangle$ on the right-hand side of Eq. (29) by $\langle r_{\mathrm{B}} \rangle$:

$$D_{\perp \mathrm{e,i}} \approx \nu \langle r_{\mathrm{Be,i}} \rangle^2 \propto \sqrt{m_{\mathrm{e,i}}} \,. \tag{44}$$

Note that the masses in Eq. (44) are *not* the reduced masses! For the ratio of the diffusion coefficients, instead of Eq. (43) we obtain

$$\frac{D_{\perp \mathrm{e}}}{D_{\perp \mathrm{i}}} = \sqrt{\frac{m_{\mathrm{e}}}{m_{\mathrm{i}}}} \,. \tag{45}$$

Across the magnetic field, ion transport is much faster than electron transport. For plasma confined magnetically in a closed vessel, this has serious consequences. If the vessel has dielectric walls, regions hit by magnetic field lines will charge up negatively, other regions positively. The charges build up a potential, repelling the fast component so much that both kinds of particle hit the wall at the same rate. Thus transport (i.e. plasma losses) is ruled by the slowly transported species:

along the field lines this is the ions, and across the field lines it is the electrons. In the case of a metallic wall, compensating currents between the different wall regions inhibit the build-up of surface charges. Thus the plasma losses are ruled by the fast transported species: across the magnetic field lines this is the ions, and along the magnetic field lines it is the electrons. This effect is sometimes called the Simon short-circuit effect. [6]. It plays a decisive role in ion sources for multiply charged ions especially in the so-called ECRIS, a microwave discharge in a magnetic trap consisting of a superposition of a mirror trap as in Fig. 8 and a magnetic hexapole – see the discussion in Refs. [7], [8] and the literature cited therein. In these sources, it was found that biasing the metallic endplate of the discharge vessel greatly enhanced the production of high charge states. A further increase could be obtained by covering the inside of the discharge vessel with dielectric layers. Both measures intercept part of the compensating wall current and thus dramatically improve plasma containment. This in turn improves ionization into high charge states.

3.3 Fluid description of plasma

The discussion of single-particle motion neglects the strong coupling between positive and negative charges and thus cannot render all aspects of plasma physics. A model emphasizing more strongly the coupling of charges of opposite sign is the description of plasma as a conducting fluid. This description of plasma is called magneto fluid dynamics or plasma fluid dynamics.

In this model the fluid is considered as a continuum, which means that any partition of the fluid has the same properties – independently of its size. The granulation due to the atomic structure is neglected. This means that the discussion is only valid on macroscopic scales.

The kinematics of point masses describes the motion of a (rigid) body by its position vector $\overrightarrow{r(t)}$ and its time derivatives, velocity $\overrightarrow{\dot{r}(t)}$ and acceleration $\overrightarrow{\ddot{r}(t)}$. In fluid mechanics we have an extended medium, which we have to characterize by extended velocity and acceleration fields $\overrightarrow{\dot{r}(t,x)}$ and $\overrightarrow{\ddot{r}(t,x)}$. A position corresponds to a fluid particle, which should keep its identity while streaming. Thus we can define a trajectory for it. As a consequence, there exists an unambiguous transformation between the locations of fluid particles at a time t_0 and a later moment t. Let $\overrightarrow{r(0)} = \{x(0), y(0), z(0)\}$ be the coordinate of a particle at a time t_0, which we will in future denote by

$$\overrightarrow{r(0)} \equiv \vec{a} \equiv (\xi(0), \eta(0), \zeta(0)) . \tag{46}$$

At a later time t the position vector $\vec{r}$ of this particle is an unambiguous function of $\vec{a}$ and t:

$$\vec{r}(\vec{a},t) \equiv (x(\vec{a},t), y(\vec{a},t), z(\vec{a},t)) . \tag{47}$$

This latter relation can be reversed and solved for $\vec{a}$ (at least in principle), yielding

$$\vec{a}(\vec{r},t) = (\xi(\vec{r},t), \eta(\vec{r},t), \zeta(\vec{r},t)) . \tag{48}$$

Here $\vec{r}(\vec{a},t)$ is the position of the particle that at t_0 was at the position $\vec{a}$. The coordinate system of $\vec{r}(\vec{a},t)$ is called Euler coordinates. Their origin is fixed in space.

The coordinate $\vec{a}(\vec{r},t)$ identifies the fluid particle that at time t is found at the position $\vec{r}$. The respective coordinate system is named Lagrange coordinates. These coordinates identify fluid particles and are thus coupled to the flow. Other names for Lagrange coordinates are convective or material coordinates.

A streaming continuum represents a velocity field. By this phrase we denote that the velocity is a function of position and time, that is

$$\vec{v} = \vec{v}(\vec{r},t) \quad \text{or} \quad \vec{v} = \vec{v}(\vec{a},t) . \tag{49}$$

The two expressions can be unambiguously transformed into each other, but have different meanings: $\vec{v}(\vec{r},t)$ is the velocity of the fluid at the position $\vec{r}$ at the time *t*, while $\vec{v}(\vec{a},t)$ is the velocity of particle $\vec{a}$ at the time t. Besides velocity, a streaming fluid is also characterized by other quantities that are functions of position and space, like the mass density ρ, the pressure *p* and the temperature *T*. These quantities are called *intensive quantities*. By taking the volume integral over an intensive quantity, we end up with an *extensive quantity*. For example, the mass *m* of a certain area of the fluid is given by

$$m = \iiint_V \rho \, \mathrm{d}V . \tag{50}$$

Extensive quantities are additive, intensive quantities not: the mass of two regions is the sum of their masses. But for a homogeneous fluid, the mass density does not change, when adding material domains with equal densities together.

We now consider the space and time derivatives of intensive quantities. Consider such a quantity $\Phi(\vec{a}(\vec{r},t),t) = \Phi(\vec{r}(\vec{a},t),t)$. Here we understand that Φ is the *numerical value* of this function. In Euler coordinates we obtain for the differential $\mathrm{d}\Phi$

$$\mathrm{d}\Phi = \frac{\partial \Phi}{\partial t}\mathrm{d}t + \frac{\partial \Phi}{\partial x}\mathrm{d}x + \frac{\partial \Phi}{\partial y}\mathrm{d}y + \frac{\partial \Phi}{\partial z}\mathrm{d}z . \tag{51}$$

The convective time derivative is usually written $\mathrm{D}/\mathrm{D}t$. It is not a total derivative because $\vec{a}$ is kept constant. For the convective time derivative of Φ we obtain

$$\frac{\mathrm{D}\Phi}{\mathrm{D}t} = \left(\frac{\partial \Phi}{\partial t}\right)_{\vec{r}} + \frac{\partial x}{\partial t}\cdot\frac{\partial \Phi}{\partial x} + \frac{\partial y}{\partial t}\cdot\frac{\partial \Phi}{\partial y} + \frac{\partial z}{\partial t}\cdot\frac{\partial \Phi}{\partial z} = \frac{\partial \Phi}{\partial t} + (\vec{v}\cdot\vec{\nabla})\Phi . \tag{52}$$

This formula can be applied to the components of the vector $\vec{v}$, yielding

$$\frac{\mathrm{D}\vec{v}}{\mathrm{D}t} = \frac{\partial \vec{v}}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v} . \tag{53}$$

Note that the dot product on the right-hand side is a scalar, which multiplied with the vector $\vec{v}$ gives a vector. The term $\partial\vec{v}/\partial t$ is the so-called *local acceleration* of a non-stationary flow; the term $(\vec{v}\cdot\vec{\nabla})\vec{v}$ is named *convective acceleration*. To understand the difference, we consider the flow of an incompressible fluid in a tube with varying cross-section (e.g. a Venturi tube; see the sketch shown in Fig. 10). If the flow is stationary, the local acceleration is zero, $\partial\vec{v}/\partial t = 0$. However, a fluid particle passing from the left large cross-sectional area into the narrow one will be accelerated because the flow velocity is higher in the narrow section of the tube. This acceleration is described by the convective acceleration $(\vec{v}\cdot\vec{\nabla})\vec{v}$. If, however, the flow becomes non-stationary because the pressure at the left entrance changes, the flow velocity will change everywhere. Then $\partial\vec{v}/\partial t \neq 0$.

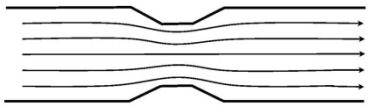


Fig. 10: Sketch of the stationary flow of an incompressible fluid in a tube with changing cross-sections

We now have the ingredients for formulating the equation of motion. We consider a fluid particle with mass m given by Eq. (50). Here ρ is the mass density of the fluid in the vicinity of the particle. Newton's equation for this particle reads

$$\frac{\mathrm{D}}{\mathrm{D}t}\iiint_{\hat{V}} \rho\vec{v}\,\mathrm{d}V = \sum \vec{F}\,. \tag{54}$$

Here $\sum \vec{F}$ is the summation over all forces acting on the particle. Our equation is in convective coordinates, the integration must be taken over $\hat{V}$, the volume of the fluid particle, which may change along the particle's trajectory. Thus the differentiation and the integration cannot simply be interchanged. In contrast to a volume V, fixed in space, we call $\hat{V}$ a *material domain*. However, for transforming Eq. (54) into Euler coordinates, we must replace the integration over $\hat{V}$ by integration over that fixed volume V, which just coincides with $\hat{V}$. The change of an extensive quantity of the particle inside this fixed volume is described by the *Reynolds transport theorem*, which states that the temporal change of an extensive quantity in a material domain is given by the temporal change of this quantity in the fixed volume just coinciding with the material domain and the total flow into and out of the fixed volume. Thus

$$\frac{\mathrm{D}}{\mathrm{D}t}\iiint_{\hat{V}} \rho\vec{v}\,\mathrm{d}V = \iiint_{V}\left(\frac{\partial\rho\vec{v}}{\partial t} + (\nabla\cdot\vec{v})\rho\vec{v}\right)\mathrm{d}V\,. \tag{55}$$

Here and in what follows, we omit indicating the symbol of the nabla operator ∇ by an arrow on top, as it is evident that it can be considered as a vector. Because V is fixed, we can interchange time derivation and volume integration. To proceed further, we must specify the forces on the right-hand side of Eq. (54). We distinguish between *surface forces* and *volume forces*. Surface forces can be represented by a surface integral of a force density. The only example we have to consider is the pressure force $\vec{F_\mathrm{p}}$ defined by

$$\vec{F_\mathrm{p}} = -\iint_{A} p\,\mathrm{d}\vec{A} = -\iiint_{V} (\nabla p)\,\mathrm{d}V\,. \tag{56}$$

For the latter transformation, we used the Gauss theorem.

Volume or external forces $\vec{F_V}$ will be specified below. They can be represented by a volume integral over a force density $\vec{f}$. Thus

$$\vec{F_V} = \iiint_{V} \vec{f}\,\mathrm{d}V\,. \tag{57}$$

By using these expressions we obtain for our momentum equation

$$\iiint_V \left(\rho \frac{\partial v}{\partial t} + \rho(\vec{v} \cdot \nabla)\vec{v} + \nabla p - \vec{f} \right) \mathrm{d}V = 0 . \tag{58}$$

This equation must be valid for any fixed volume. This is only possible if the integrand equals zero:

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \nabla)\vec{v} + \frac{1}{\rho}\nabla p - \frac{1}{\rho}\vec{f} = 0 \quad \text{(Euler equation)} . \tag{59}$$

Examples of external forces are the Lorentz force and the gravitational force. Further, we have an internal friction $\vec{f_{\mathrm{R}}}$ due to momentum exchange between the different plasma components (and due to viscosity, but viscosity can be neglected in ion source plasma).

The Lorentz force acting on a single ion or electron is given by

$$q_k \vec{E} + q_k \vec{v_k} \times \vec{B} \equiv \frac{1}{n_k} \vec{f_k} . \tag{60}$$

Here the index k characterizes the particle species (electron or ions of different charge and mass). The total Lorentz force density is the sum over these different contributions:

$$\vec{f_{\mathrm{L}}} = \sum \vec{f_k} = \left(\sum_k n_k q_k \right) \cdot \vec{E} + \left(\sum_k n_k q_k \vec{v_k} \right) \times \vec{B} = \rho_{\mathrm{el.}} \vec{E} + \vec{j} \times \vec{B} . \tag{61}$$

Here we have introduced for abbreviation the electrical space-charge density $\rho_{\mathrm{el.}} \equiv \sum_k n_k q_k$ and the electrical current density $\vec{j} \equiv \sum_k n_k q_k \vec{v_k}$. Analogously we obtain for the gravity force density ($\vec{g}$ is gravitational field strength)

$$\vec{f_{\mathrm{g}}} = \rho \vec{g} . \tag{62}$$

Thus we have finally

$$\rho \frac{\partial \vec{v}}{\partial t} + \rho(\vec{v} \cdot \nabla)\vec{v} + \nabla p - \rho_{\mathrm{el.}} \vec{E} - \vec{j} \times \vec{B} - \rho \vec{g} + \vec{f_{\mathrm{R}}} = 0 . \tag{63}$$

Without the Lorentz term, this equation is known as the *Navier–Stokes equation*. For some situations, it is sufficient to use this equation as a global plasma equation and consider ions and electrons to be strongly coupled. For instance, this may be the case for slow waves (model of plasma as a single fluid).

In the case of high-frequency waves, the ion inertia may be so large that ions cannot follow the fast oscillations, while electrons can. In this case it is better to formulate a separate fluid equation for every plasma component and consider the coupling by mutual friction and electric fields created by space charges. Neglecting gravitation, we have for a component k

$$\rho_k \frac{\partial \vec{v_k}}{\partial t} + \rho_k (\vec{v_k} \cdot \nabla)\vec{v_k} + \nabla p_k - \rho_{\mathrm{el.},k} \vec{E} - \vec{j_k} \times \vec{B} + \sum_l \vec{f_{k,l}} = 0 . \tag{64}$$

The last term contains the forces due to internal friction between the different plasma species, which is between electrons or ions and electrons, or between ions and neutrals. A *highly ionized plasma* is defined as plasma where electron–ion friction dominates. A *weakly ionized plasma* is dominated by friction between charged particles and neutrals.

For the force density $\overrightarrow{f_{k,l}}$ acting on particles of kind *k* due to collisions with particles of kind *l*, we have the general formula

$$\overrightarrow{f_{kl}} = n_k m_{kl} \nu_{kl} (\overrightarrow{v_l} - \overrightarrow{v_k}) , \tag{65}$$

where

$$m_{kl} = \frac{m_k m_l}{m_k + m_l} \tag{66}$$

is the reduced mass of the colliding particles and ν_{kl} is the respective collision frequency for momentum transfer.

In addition, we need the continuity equation describing the conservation of mass

$$\frac{\partial \rho}{\partial t} + (\vec{v} \cdot \nabla)\rho + \rho \nabla \cdot \vec{v} = 0 , \tag{67}$$

which can also be formulated for each of the different plasma components, and what is called Ohm's law, but is in reality the definition of the electrical conductivity σ. (Ohm's law states that σ depends neither on the electric field $\vec{E}$ nor on the current density $\vec{j}$; only in this we find the case $\vec{j} \propto \vec{E}$.) The current density in a flowing fluid is proportional to the electric field $\vec{\tilde{E}}$ in the frame of the moving fluid. Transformation to a system fixed in space yields $\vec{\tilde{E}} = \vec{E} + \vec{v} \times \vec{B}$. Thus the current density is proportional to $\vec{E} + \vec{v} \times \vec{B}$ in a coordinate system at rest. The proportionality constant is the conductivity σ:

$$\vec{j} = \sigma(\vec{E} + \vec{v} \times \vec{B}) . \tag{68}$$

In the presence of an external magnetic field, plasma is anisotropic. The density of electric currents flowing under the action of an electric field $\vec{E}$ is not necessarily parallel to $\vec{E}$. Thus σ must be considered as a tensor in the general case. We express this by doubly underlining the symbol. The more general definition of the conductivity thus reads

$$\vec{j} = \underline{\underline{\sigma}}(\vec{E} + \vec{v} \times \vec{B}) . \tag{69}$$

We discuss this in detail in the following section. Now

$$\vec{j} = e n_{\mathrm{i}} \overrightarrow{v_{\mathrm{i}}} - e n_{\mathrm{e}} \overrightarrow{v_{\mathrm{e}}} . \tag{70}$$

The ion and electron densities must be obtained from a two-fluid model.

3.4 AC conductivity of magnetized plasma

We now consider the case of particle motion under the action of an oscillating electric field $\vec{E} = \vec{E_0}\exp(-\mathrm{i}\omega t)$ in magnetized plasma. The vectors $\vec{E_0}$ and $\vec{B}$ define a plane. We introduce Cartesian coordinates (main directions labelled x, y, z) in such a way that this plane is the xz-plane and B is parallel to the z-direction (see Fig. 11).

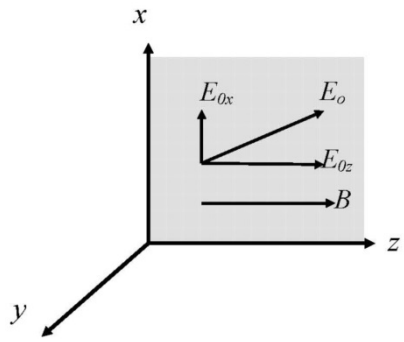


Fig. 11: Coordinates for calculating the AC conductivity of magnetized plasma

To solve the equation of motion of a charged particle, we start with an ansatz for the particle velocity v

$$\vec{v} = \left[a_1(t)\overrightarrow{E_{01}} + a_2(t)\overrightarrow{E_{01}} \times \vec{B} + a_3(t)\overrightarrow{E_{03}}\right]\exp(-i\omega t)\,. \tag{71}$$

(Note that $\vec{E_0}$ has by definition no component in the y-direction. We therefore constructed a respective coordinate vector by the cross product between $\overrightarrow{E_{01}}$ and $\vec{B}$). Introducing Eq. (71) into Eq. (33) gives, after some algebra, a set of three equations for the velocity components, respectively for the a_i. Of these equations, those for $a_{1,2}$ are coupled, while the equation for a_3 does not depend on the magnetic induction $\vec{B}$, and is similar to the equation for the case without a magnetic field, namely

$$\frac{\partial a_3}{\partial t} - \mathrm{i}\omega a_3 - \frac{q}{m} = 0 \quad \Rightarrow \quad a_3 = \frac{\mathrm{i}q}{\omega m} \quad \Rightarrow \quad v_z = \frac{\mathrm{i}qE_z}{\omega m}\,. \tag{72}$$

For $a_{1,2}$ we obtain

$$\frac{\partial a_1}{\partial t} - \mathrm{i}\omega a_1 - \frac{q}{m} + \frac{qB^2 a_2}{m} = 0\,,$$

$$\frac{\partial a_2}{\partial t} - \mathrm{i}\omega a_2 - \frac{qa_1}{m} = 0\,. \tag{73}$$

For separating and solving (73), we differentiate both equations and in the set of differentiated equations we replace the first derivatives of the a_i by the expressions of (73). By using (37) and with some algebra we obtain for a_2 the inhomogeneous differential equation

$$\frac{\partial^2 a_2}{\partial t^2} - 2\mathrm{i}\omega\frac{\partial a_2}{\partial t} + (\omega_\mathrm{B}^2 - \omega^2)a_2 = \frac{q^2}{m^2}\,. \tag{74}$$

The solution of the homogenous part of (74) is the gyration of the particle and will not be considered here. A special solution of the inhomogeneous equation yields for $\omega \neq \omega_{\mathrm{B}}$

$$a_2 = \frac{q^2}{m^2}\frac{1}{\omega_{\mathrm{B}}^2 - \omega^2} \quad \text{and} \quad a_1 = \frac{q}{m}\frac{\mathrm{i}\omega}{\omega_{\mathrm{B}}^2 - \omega^2}. \tag{75}$$

These solutions correspond to forced oscillations with a resonance at $\omega = \omega_{\mathrm{B}}$. In the case of this resonance, one obtains a different solution describing a gyration with the amplitude rising linearly in time. This is the well-known cyclotron resonance, where a charged particle gains energy continuously from the electric field.

A particle moving under the action of a constant force $\vec{F}$ in a viscous medium attains after some time a finite, constant velocity $\vec{v}$, which is proportional to $\vec{F}$. The proportionality constant is called the mobility *b*. Thus we have $\vec{v} = b\vec{F}$. The non-resonant oscillatory motion of a charged particle under the action of an oscillating force has constant amplitude proportional to the amplitude of the oscillating field even if there is no viscous medium. This is due to the inertia of the oscillating particles creating a phase shift of 90° between the force and the particle velocity. This limits the gain of energy in the oscillating electric field. Thus also in this case the proportionality constant between the amplitude of the force and the amplitude of the velocity is called the mobility, because of some analogy with the DC case. In complex notation, assuming an electric field of the form $\vec{E} = \vec{E_0}\exp(-\mathrm{i}\omega t)$, we obtain from the equation of motion

$$\vec{v} = \frac{\mathrm{i}q}{\omega m}\vec{E}. \tag{76}$$

Thus the mobility is given by $b = \mathrm{i}q/\omega m$, the imaginary unit i standing for a phase shift of $\pi/2$ between the velocity and the driving force $q\vec{E}$.

To generalize the case of magnetized plasma considered above, we assume the existence of an oscillating electric field $\vec{E} = \vec{E_0}\exp(-\mathrm{i}\omega t) = \vec{E_1} + \vec{E_2} + \vec{E_3}$. Here the $\vec{E_i}$ are the components of $\vec{E}$ in the coordinate directions. From the solution found above, we can construct the drift velocity $\overrightarrow{v_{\mathrm{drift}}}$ of a charged particle under the action of this field,

$$\overrightarrow{v_{\mathrm{drift}}} = \frac{\mathrm{i}\omega}{\omega_{\mathrm{B}}^2 - \omega^2}\frac{q}{m}(\vec{E_1} + \vec{E_2}) + \frac{q^2/m^2}{\omega_{\mathrm{B}}^2 - \omega^2}(\vec{E_2} \times \vec{B} + \vec{E_1} \times \vec{B}) + \frac{\mathrm{i}q}{\omega m}\vec{E_3}, \tag{77}$$

corresponding to the component equations

$$\begin{aligned} v_{\mathrm{drift},x} &= \frac{\mathrm{i}\omega}{\omega_{\mathrm{B}}^2 - \omega^2}\frac{q}{m}E_1 + \frac{q^2B/m^2}{\omega_{\mathrm{B}}^2 - \omega^2}E_2 + 0 \\ &\equiv b_{11}qE_1 + b_{12}qE_2 + b_{13}qE_3, \\ v_{\mathrm{drift},y} &= -\frac{q^2B/m^2}{\omega_{\mathrm{B}}^2 - \omega^2}E_1 + \frac{\mathrm{i}\omega}{\omega_{\mathrm{B}}^2 - \omega^2}\frac{q}{m}E_2 + 0 \\ &\equiv b_{21}qE_1 + b_{22}qE_2 + b_{23}qE_3, \end{aligned} \tag{78}$$

$$v_{\mathrm{drift},z} = 0 + 0 + \frac{\mathrm{i}q}{\omega m} E_3$$
$$\equiv b_{31} q E_1 + b_{32} q E_2 + b_{33} q E_3 .$$

In the respective second lines, we have formulated the component equations with the components b_{ij} of the mobility tensor $\underline{\underline{b}}$, which has to replace the scalar mobility b in the case of the anisotropic magnetized plasma. From Eqs. (78) we obtain for $\underline{\underline{b}}$

$$\underline{\underline{b}} = \frac{\mathrm{i}}{\omega m} \begin{Vmatrix} \dfrac{\omega^2}{\omega^2 - \omega_{\mathrm{B}}^2} & \dfrac{\mp \mathrm{i}\omega\omega_{\mathrm{B}}}{\omega^2 - \omega_{\mathrm{B}}^2} & 0 \\ \dfrac{\pm\omega\omega_{\mathrm{B}}}{\omega^2 - \omega_{\mathrm{B}}^2} & \dfrac{\omega^2}{\omega^2 - \omega_{\mathrm{B}}^2} & 0 \\ 0 & 0 & 1 \end{Vmatrix} \equiv \frac{\mathrm{i}}{\omega m} \underline{\underline{K}} . \tag{79}$$

Having found the mobility, we can define the conductivity of a medium. If we have drifting plasma, the density $\vec{j}$ of an electric current is given as the sum over the densities of the drift currents of the charged particles identified by the index k. In an isotropic medium (plasma without external magnetic field) we have

$$\vec{j} = \sum_k n_k q k \overrightarrow{v_{\mathrm{drift},k}} = \sum_k n_k q_k b_k q_k \vec{E} \equiv \sigma \vec{E} . \tag{80}$$

The direction of $\vec{j}$ depends only on the direction of $\vec{E}$, not on the sign of the charge q. From (80) we obtain

$$\sigma = \sum_k n_k b_k q_k^2 . \tag{81}$$

The sign of the conductivity does not depend on the sign of q. In the case of an anisotropic medium, we obtain analogously

$$\underline{\underline{\sigma}} = \sum_k n_k \underline{\underline{b}}_k q_k^2 \quad \Rightarrow \quad \sigma_{ij} = \sum_k n_k (b_{ij})_k q_k^2 . \tag{82}$$

Thus by using (79) we have

$$\underline{\underline{\sigma}}(\omega) = \frac{i}{\omega} \sum_k \frac{n_k q_k^2}{m_k} \underline{\underline{K}}_k . \tag{83}$$

Here $\underline{\underline{K}}_k$ is the tensor defined in Eq. (79) for the plasma component labelled k.

What we have considered here is collision-free plasma. In this case the plasma is a loss-free reactance and $\underline{\underline{\sigma}}$ is imaginary, that is, without ohmic contributions. In the case of collisions, we have to supplement Eq. (33) by the average momentum loss due to collisions

$$\delta \overrightarrow{p_{\mathrm{k}}} / \delta t = \sum_{\mathrm{j} \neq \mathrm{k}} m_{\mathrm{rjk}} \left(\overrightarrow{v_{\mathrm{j}}} - \overrightarrow{v_{\mathrm{k}}} \right) \nu_{\mathrm{m}} \left(\overrightarrow{v_{\mathrm{j}}} - \overrightarrow{v_{\mathrm{k}}} \right) . \tag{84}$$

Here the m_{rjk} are the reduced masses of the colliding particles, the $\overrightarrow{v_{\mathrm{j,k}}}$ are the particle velocities and $\nu_m(|\vec{v}_j - \vec{v}_k|)$ is the average momentum transfer collision frequency. By adding this term, the elements of the conductivity tensor will become complex, the real parts describing the ohmic losses in the plasma.

4 Plasma waves

4.1 Some general wave concepts

Let us recall the concept of electromagnetic waves in vacuum as described by Maxwell's equations, which read

$$\begin{aligned} \nabla \times \vec{H} &= \varepsilon_0 \frac{\partial \vec{E}}{\partial t} \quad (\vec{j} = 0, \text{ no electric current}), \\ \nabla \times \vec{E} &= -\mu \frac{\partial \vec{H}}{\partial t} \quad (\nabla \cdot \vec{E} = 0, \text{ no space charge}, \nabla \cdot \vec{B} = 0, \text{ no magnetic monopoles}). \end{aligned} \tag{85}$$

Multiplying the first equation by (the negative of) the vacuum permittivity ($-\mu_0$), and taking the time derivative yields

$$-\mu_0 \nabla \times \frac{\partial \vec{H}}{\partial t} = -\frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2}, \tag{86}$$

where we have used the relation $\varepsilon_0 \mu_0 = 1/c^2$, and c is the phase velocity of light in vacuum. Taking the curl ($\nabla \times$) of the second equation and eliminating $\vec{H}$ by using Eq. (85), we finally get (in similar ways)

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\vec{E} = 0 \quad \text{and} \quad \left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\vec{H} = 0. \tag{87}$$

These are the equations for electromagnetic waves in vacuum. We solve them by assuming plane waves

$$\vec{E} = \vec{E_0} \exp[\mathrm{i}(\vec{k} \cdot \vec{r} - \omega t)] + \vec{E_0^*} \exp[-\mathrm{i}(\vec{k} \cdot \vec{r} - \omega t)]. \tag{88}$$

Here $\vec{E_0}$ and $\vec{E_0^*}$ are the complex amplitudes, the wave vector $\vec{k}$ gives the direction of wave propagation, and ω is the angular frequency. The asterisk indicates the complex conjugate. When introducing this ansatz into the wave equation, we obtain equations in which ∇ is replaced by $\pm \mathrm{i}\vec{k}$ and the time derivative by $\mp \mathrm{i}\omega$. In principle, one needs to take only one of the two terms on the right-hand side of (88), because only the real part is physically important. It is the same for both terms. We will take for our discussion the first term (i.e. the upper sign), but you may find a different convention in some books.

In the case of vacuum we obtain

$$-k^2 + \omega^2 c^2 = 0 \quad \text{or} \quad \omega = \pm kc. \tag{89}$$

In general, the relations $\omega(k)$ or $k(\omega)$ are termed *dispersion relations*. For vacuum, the dispersion relation is linear.

We now consider the argument of the exponential functions, that is, the *phase*. By the condition

$$\frac{\mathrm{d}}{\mathrm{d}t}(\vec{k}\cdot\vec{r}-\omega t)=\vec{k}\cdot\dot{\vec{r}}-\omega=0 \tag{90}$$

we define a velocity for the propagation of the phase. It follows from this equation that the phase travels parallel to $\vec{k}$ and that the absolute value of the phase velocity is given by

$$v_{\mathrm{ph}}=\frac{\omega}{k}\,. \tag{91}$$

By

$$\mu=c/v_{\mathrm{ph}}=c\cdot k/\omega \tag{92}$$

we define the refractive index of a wave. In vacuum we have $c=v_{\mathrm{ph}}$, thus $\mu=1$.

Besides the phase velocity we define the group velocity v_{g} by

$$v_{\mathrm{g}}=\frac{\mathrm{d}\omega}{\mathrm{d}k}=\frac{\mathrm{d}(v_{\mathrm{ph}}k)}{\mathrm{d}k}=v_{\mathrm{ph}}+k\frac{\mathrm{d}v_{\mathrm{ph}}}{\mathrm{d}k}=v_{\mathrm{ph}}-\lambda\frac{\mathrm{d}v_{\mathrm{ph}}}{\mathrm{d}\lambda}\,. \tag{93}$$

4.2 Waves in plasma without external magnetic field

4.2.1 Electromagnetic waves

As a first step we consider waves in plasma without an external magnetic field, that is, for $\vec{B}=0$. The convective acceleration is *per se* nonlinear. By neglecting this term, we restrict our consideration to waves with sufficiently small amplitudes as a first step. In contrast to vacuum, we may have space charge and electric current in plasma. Thus $\nabla\cdot\vec{E}\neq 0$ and $\vec{j}=\rho_{\mathrm{el}}\vec{v}\neq 0$.

However, we neglect pressure effects and assume the plasma to contain only singly charged ions of one kind. The equation of motion thus becomes

$$\frac{\partial\vec{j}}{\partial t}=e^2n\left(\frac{1}{m_{\mathrm{e}}}+\frac{1}{m_{\mathrm{i}}}\right)\vec{E}=\frac{e^2n}{m_{\mathrm{ei}}}\vec{E}=\varepsilon_0\omega_{\mathrm{p}}^2\vec{E}\,, \tag{94}$$

where $n=n_{\mathrm{e}}=n_{\mathrm{i}}$ is the plasma density, and $m_{\mathrm{ei}}\approx m_{\mathrm{e}}$ is the reduced mass of electrons and ions. Assuming plane waves yields

$$\vec{j}=-\frac{\varepsilon_0\omega_{\mathrm{p}}^2}{\mathrm{i}\omega}\vec{E}\,. \tag{95}$$

Thus we obtain $\sigma=\mathrm{i}\varepsilon_0\omega_{\mathrm{p}}^2/\omega$. In this approximation, plasma is not an ohmic conductor but a reactance. There is a phase shift by 90° between the electric field and the current density. The imaginary conductivity is due to the inertia of the electrons. Plasma behaves like an inductance. The Maxwell equations for this case become

$$\nabla \times \vec{H} = -\frac{\varepsilon_0 \omega_\mathrm{p}^2}{\mathrm{i}\omega}\vec{E} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t},$$

$$\nabla \times \nabla \times \vec{E} = -\frac{\omega_\mathrm{p}^2}{\mathrm{i}c^2\omega}\frac{\partial \vec{E}}{\partial t} - \frac{\omega^2}{c^2}\frac{\partial^2 \vec{E}}{\partial t^2}.. \tag{96}$$

Transformations of these equations must allow for transverse as well as longitudinal waves as possible solutions. Multiplying the first of these equations by $(-\mu_0)$ and with $\vec{E} = \vec{E_0}\exp[\mathrm{i}(\vec{k}\cdot\vec{r} - \omega t)]$ we thus obtain finally

$$(\vec{k}\vec{k} - k^2)\vec{E} = \frac{\omega^2}{c^2}\left(1 - \frac{\omega_\mathrm{p}^2}{\omega^2}\right)\vec{E} = \varepsilon\frac{\omega^2}{c^2}\vec{E}, \tag{97}$$

where

$$\varepsilon = 1 - \omega_\mathrm{p}^2 / \omega^2 \tag{98}$$

is the permittivity of the plasma. Equation (98) is known as the *Eccles relation*.

For transverse waves, $\vec{k} \perp \vec{E}$, thus $\vec{k}\cdot\vec{E} = 0$ and we obtain as dispersion relation

$$k^2 = \frac{\omega^2}{c^2}\left(1 - \frac{\omega_\mathrm{p}^2}{\omega^2}\right) \quad \text{or} \quad \frac{\omega^2}{\omega_\mathrm{p}^2} = 1 + \frac{k^2c^2}{\omega_\mathrm{p}^2}. \tag{99}$$

The latter form we present in Fig. 12 (so-called *Brillouin diagram*). Here the inclinations of the dashed (green) lines give the phase velocity and the group velocity of the plasma wave at the point where these lines cross the (blue) plasma wave curve.

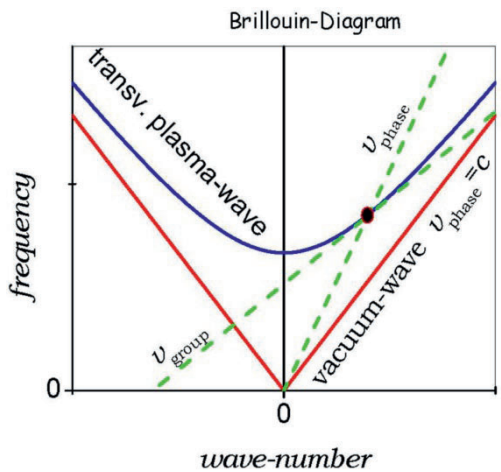


Fig. 12: Brillouin diagram (i.e. frequency versus wavenumber) for transverse electromagnetic waves in vacuum and in magnetic field-free plasma.

For longitudinal waves, $\vec{k} \parallel \vec{E}$ and thus $\vec{k}(\vec{k}\cdot\vec{E}) = k^2\vec{E}$. The left-hand side of Eq. (97) equals zero. Thus

$$1 - \frac{\omega_p^2}{\omega^2} = 0 \quad \Rightarrow \quad \omega = \omega_p . \tag{100}$$

There is no wave, only oscillations, with the plasma frequency as discussed in the introduction.

For high frequencies, the dispersion relation of transverse waves approaches asymptotically that of free space. The plasma behaves as a dielectric with a refractive index μ given by the so-called Maxwell relation

$$\mu = \sqrt{\varepsilon} = \sqrt{1 - \omega_p^2 / \omega^2} \le 1 . \tag{101}$$

The approximations used for deriving Eq. (101) become valid at sufficiently high frequencies of the waves. Even in the case of plasma with external magnetic field, our model is correct for $\omega >> \omega_B$. We expect deviations at lower frequencies.

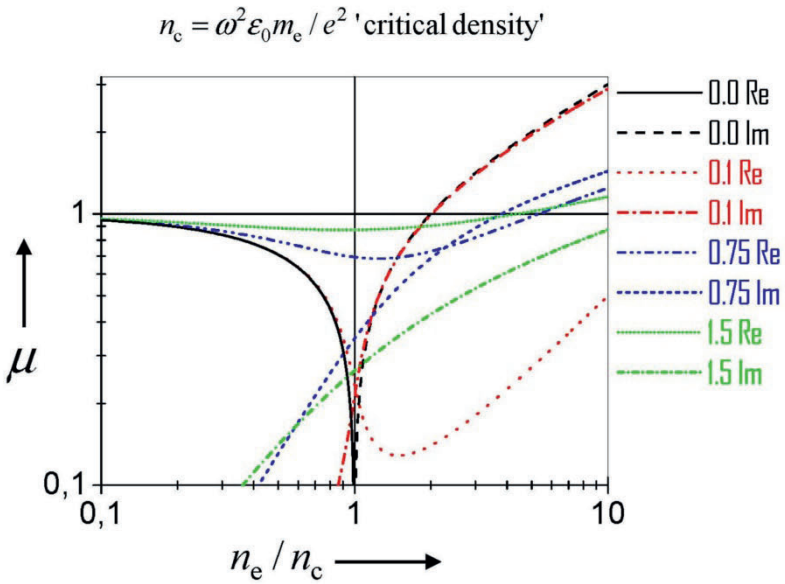


Fig. 13: Real (Re) and imaginary (Im) parts of the refractive index of transverse plasma waves versus plasma density over critical density for different ratios of momentum transfer collision frequency over wave frequency.

By using the refractive index with $\nu_m = 0$ we can distinguish different regions for transverse wave propagation in plasma (see Fig. 13):

- For $\omega > \omega_p$, the refractive index is real. Waves can propagate without damping. Plasma behaves as a lossless dielectric with a refractive index $\mu < 1$. Thus the phase velocity of waves exceeds the phase velocity of electromagnetic waves in vacuum, $v_{ph} > c$.
- For $\omega = \omega_p$, we have $\mu = 0$ and $k = 0$. There is no wave, only an oscillation.

– For $\omega < \omega_p$, both μ and k are imaginary. There is no wave propagation. Waves when penetrating a plasma will decay exponentially over the so-called skin depth L_{skin} given by

$$L_{skin} = \frac{1}{\operatorname{Im} k} = \frac{1}{(\omega / c)\sqrt{\omega_p^2 / \omega^2 - 1}} \quad \Rightarrow \quad \frac{c}{\omega_p} = \frac{\lambda_p}{2\pi} \quad \text{for} \quad \omega \to 0 . \tag{102}$$

For small frequencies the skin depth does not depend on frequency and is equal to the wavelength of an electromagnetic wave in vacuum having a frequency equal to the plasma frequency.

Collisions between charged particles and between charged particles and neutrals cause wave damping. Since electronic collisions are most frequent, they are most important. When considering these collisions, instead of Eq. (94) we obtain

$$\frac{\partial \vec{j}}{\partial t} + \vec{j} \cdot \nu_m = \varepsilon_0 \omega_p^2 \vec{E} . \tag{103}$$

Introducing plane waves yields

$$\vec{j} = -\varepsilon_0 \omega_p^2 \frac{\vec{E}}{i\omega - \nu_m} \equiv \sigma \vec{E} . \tag{104}$$

By a similar procedure as described above, for the dispersion relation with respect to the refractive index we obtain

$$\mu = \frac{kc}{\omega} = \sqrt{\left(1 - \frac{\omega_p^2}{\omega^2 + \nu_m^2}\left(1 - i\frac{\nu_m}{\omega}\right)\right)} . \tag{105}$$

This function is plotted in Fig. 13 for different values of the collision frequency ν_m. As a result of dissipation by collisions, the conductivity will become complex. The same holds true for the refractive index and for *k*. There are no longer frequency regions where μ is purely imaginary or real. Waves propagate even for $\omega < \omega_p$, though they are heavily damped. But damping is also observed for $\omega > \omega_p$.

4.2.2 Longitudinal waves

Longitudinal waves in neutral gases are (compressive) sound waves. In the equation of motion, the respective restoring force is described by the ∇p term. Acoustic waves are dispersion-free, that is, $\omega = kc_s$. The phase velocity of sound in neutral gas, c_s, is given by

$$c_s = \sqrt{\kappa p / \rho} . \tag{106}$$

Here κ is the ratio of the specific heat at constant pressure to that at constant volume. To obtain these relations, one considers small wave amplitudes and neglects all terms quadratic in wave amplitude or quantities proportional to wave amplitudes (linearization). The gas is compressed adiabatically by the wave, thus $\nabla p / p = \kappa (\nabla n / n)$.

In analogy to Eq. (106), for abbreviation, in plasmas we can define two further sound speeds:

$$c_{\mathrm{si}} = \sqrt{\kappa_{\mathrm{i}} p_{\mathrm{i}} / \rho_{\mathrm{i}}} \quad \text{ion sound speed} \tag{107}$$

and

$$c_{\mathrm{se}} = \sqrt{\kappa_{\mathrm{e}} p_{\mathrm{i}} / \rho_{\mathrm{e}}} \quad \text{electron sound speed}\,. \tag{108}$$

For simplicity we neglect collisional damping, and obtain the equations of motion for electrons and ions as

$$\begin{aligned} \rho_{\mathrm{e}} \frac{\partial \vec{v}_{\mathrm{e}}}{\partial t} &= -n_{\mathrm{e}} e \vec{E} - \nabla p_{\mathrm{e}}, \\ \rho_{\mathrm{i}} \frac{\partial \vec{v}_{\mathrm{i}}}{\partial t} &= n_{\mathrm{i}} e \vec{E} - \nabla p_{\mathrm{i}}. \end{aligned} \tag{109}$$

Here e is the *positive* elementary charge. By a linearization similar to that applied for ordinary sound waves, considering longitudinal waves with $\vec{k} \parallel \vec{E}$ (this is equivalent to reducing the Maxwell equations to the Poisson equation) and assuming plane waves, one ends up with two combined equations of motion for electrons and ions. These equations have a non-trivial solution only if their determinant is zero. This condition yields the dispersion relation, a bi-quadratic equation having two independent solutions corresponding to two different kinds of wave. At high frequencies we obtain approximately

$$\omega^2 = \omega_{\mathrm{p}}^2 + k^2 (c_{\mathrm{se}}^2 + c_{\mathrm{si}}^2) \approx \omega_{\mathrm{pe}}^2 + k^2 \kappa_{\mathrm{e}} \frac{k_{\mathrm{B}} T_{\mathrm{e}}}{m_{\mathrm{e}}} \quad \text{(Bohm–Gross dispersion relation).} \tag{110}$$

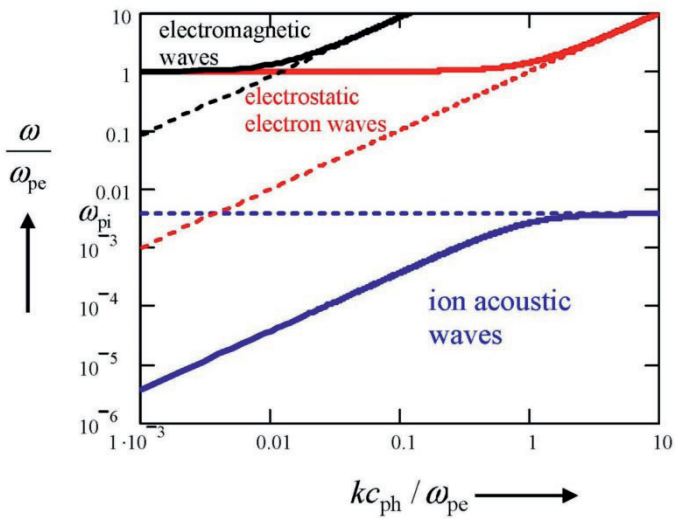


Fig. 14: Double-log Brillouin diagram for longitudinal plasma waves plotted from the exact formulas; and the same for electromagnetic (transverse) waves for comparison.

The respective waves are electrostatic electron waves. The dispersion relation is formally equivalent to that of electromagnetic waves. However, the phase velocity is much smaller in the case of acoustic waves: $c_{se} << c$. In Fig. 14 we show these waves for the case of argon plasma with cold

ions, $n_e = 10^{18}\,\mathrm{m}^{-3}$ and $k_B T_e = 10\,\mathrm{eV}$. The coefficient κ_e depends on the model used. Many authors set $\kappa_e = 3$ (see e.g. [9]).

The second branch represents ion acoustic waves. For small wavenumbers one obtains for the dispersion relation

$$\omega^2 \approx k^2 \left(c_{si}^2 + \frac{m_e}{m_i} c_{se}^2 \right) \approx k^2 \kappa_e \frac{k_B T_e}{m_i} . \tag{111}$$

The latter transformation is valid if (as mostly the case in low-pressure gas discharges) $T_i << T_e$. For larger values of k (i.e. at short wavelengths), the frequency is almost constant and corresponds to the ion plasma frequency ω_{pi} as sketched in Fig. 14 for an argon plasma with the values of density and electron temperature given above. We present the exact dispersion relation, not the approximations given above. To span the large differences in frequency of these types of waves, we have used a double-logarithmic plot (compare the Brillouin diagram, Fig. 12).

In reality electrostatic waves are strongly damped, especially at short wavelengths. The damping mechanism is not described by the fluid dynamic model used here. It relies on the concept of fluid particles. In a plasma any set of microscopic particles (electrons and ions) forming a (macroscopic) fluid particle will disintegrate by diffusion within a short time. Thus ions and electrons come out of phase with the wave motion and their wave energy is transformed into thermal energy. Further, the interaction of ions and electrons with the electric field of a wave depends strongly on the particle velocity and is most efficient for particles moving in the same direction as the wave with velocities close to the phase velocity. The waves are damped by transferring energy to those particles. These nonlinear processes (*Landau damping*, *particle trapping*) are adequately described only by kinetic theory.

4.3 Waves in plasma with external magnetic field

4.3.1 Cold plasma

When plasma is magnetized, not only does it become anisotropic, but also the force exerted by the magnetic field on charged particles will increase the number of possible wave modes. To describe all this would need more than the space available here. Therefore, we will discuss only a simple example as a model and then give some general remarks. To start, we neglect the ∇p force. This corresponds to considering only cold, homogeneous plasma. Our set of equations is the appropriate Maxwell equation,

$$\nabla \times \nabla \times \vec{E} = \nabla^2 \vec{E} - \nabla(\nabla \cdot \vec{E}) = \frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} + \frac{1}{\varepsilon_0 c^2} \frac{\partial \vec{j}}{\partial t} , \tag{112}$$

the momentum equation from a single-fluid model,

$$\rho \frac{\mathrm{D}\vec{v}}{\mathrm{D}t} = \rho_{el} \vec{E} + \vec{j} \times \vec{B} , \tag{113}$$

and (a simplified form of) the so-called Ohm's law obtained from a two-fluid model,

$$\frac{\partial \vec{j}}{\partial t} = \frac{e^2 n}{m_e} (\vec{E} + \vec{v} \times \vec{B}) - \frac{\omega_{Be}}{B} (\vec{j} \times \vec{B}) . \tag{114}$$

The decisive difference, that is, the anisotropy of the plasma, is represented by the $\vec{j}\times\vec{B}$ terms. As a consequence, $\vec{j}$ and $\vec{E}$ may not be parallel. For linearization we use the scheme

$$\vec{B}=\vec{B}_0+\vec{B}',\quad \vec{E}=0+\vec{E}',\quad \vec{j}=0+\vec{j}',\quad \vec{v}=0+\vec{v}',\quad \rho=\rho_0+\rho',\quad \rho_{\mathrm{el}}=0+\rho'_{\mathrm{el}}. \tag{115}$$

The primed quantities are considered as small perturbations of the equilibrium values and small to first order. When introducing (115) into Eqs. (112)–(114) and neglecting terms with products of small quantities, because they are small to second order, we obtain a new system of linear equations. By setting the determinant of this equation system to zero, we obtain a functional $F(\omega,k)$ as dispersion relation.

We simplify this procedure for two special cases, assuming transverse waves with $\vec{E}\perp\vec{k}$, propagating along the magnetic induction of the external field, that is, $\vec{k}\parallel\vec{B}_0$. By eliminating the electric field, we obtain an equation for $\vec{j}$,

$$\left\{\omega^2-\frac{\omega^2\omega_{\mathrm{p}}^2}{\omega^2-k^2c^2}-\omega_{\mathrm{Bi}}\omega_{\mathrm{Be}}\right\}\vec{j}=-\mathrm{i}\omega\omega_{\mathrm{Be}}\vec{j}\times\vec{B}_0. \tag{116}$$

Here $\omega_{\mathrm{Bi,Be}}$ are the gyro-frequencies of ions and electrons respectively (see Eq. (37)). For $\omega << \omega_{\mathrm{Bi}}$ we can neglect the term quadratic in ω and the term on the right-hand side of Eq. (116) and obtain for the refractive index of these waves

$$\mu^2=\left(\frac{kc}{\omega}\right)^2=1+\frac{nm_{\mathrm{i}}}{\varepsilon_0B_0^2}\approx\frac{\rho\mu_0c^2}{B^2}\equiv\frac{c^2}{v_{\mathrm{A}}^2}. \tag{117}$$

The phase velocity of these low-frequency waves is the so-called Alfvén velocity v_{A}. They were first described by H. Alfvén and are called Alfvén waves. These waves resemble waves on a string under tension. Here the oscillating strings are the plasma-filled magnetic flux tubes.

If we multiply Eq. (116) with its complex conjugate, we obtain an equation that can be reduced by erasing $|\underline{j}|^2$ on both sides. Taking the square root and after some algebra we obtain the refractive index

$$\mu^2=1-\frac{\omega_{\mathrm{p}}^2}{\omega^2-\omega_{\mathrm{Bi}}\omega_{\mathrm{Be}}\mp\omega\omega_{\mathrm{Be}}}. \tag{118}$$

We see that $\mu^2\to\infty$ and changes sign for

$$\omega^2-\omega_{\mathrm{Bi}}\omega_{\mathrm{Be}}\mp\omega\omega_{\mathrm{Be}}=0. \tag{119}$$

The two signs in the above correspond to two different types of waves, which show at two specific frequencies a resonance with the refractive index growing over all limits. For $\omega >> \omega_{\mathrm{Bi}}$ we can neglect the term $\omega_{\mathrm{Bi}}\omega_{\mathrm{Be}}$ and obtain as resonance frequency $\omega=\omega_{\mathrm{Be}}$. For $\omega << \omega_{\mathrm{Be}}$ we can neglect the ω^2 term and obtain as resonance frequency $\omega=\omega_{\mathrm{Bi}}$. (The values of the resonance frequencies are exact despite the many approximations made.) These are the electron cyclotron and ion cyclotron

waves. Both are circularly polarized: electron cyclotron waves right-hand, and ion cyclotron waves left-hand. The electric field vector changes its direction like the velocities of the gyrating electrons and ions, respectively. At resonance, electrons and ions respectively gyrate synchronously with the wave and can extract energy continuously.

Where the refractive index is imaginary, waves cannot propagate but are damped. The plasma behaves like a waveguide with propagation and cut-off regions (respectively stop bands). The frequencies limiting these regions are either the resonances ($|\mu^2| \to \infty$) or the cut-off frequencies where $\mu = 0$. The correlation between resonances, cut-offs and stop bands is shown schematically in Fig. 15. Stop bands are shaded. In non-magnetized plasma we find only a cut-off frequency, the plasma frequency, and no resonance (see Fig. 13). There we also demonstrated the effect of damping. As a consequence, μ^2 will be a complex number, its value remaining finite and non-zero, resonances and cut-offs becoming diffuse. But the cut-off regions remain regions of very strong damping, while damping in the propagation regions is much weaker. The simple damping-free description gives still the main features.

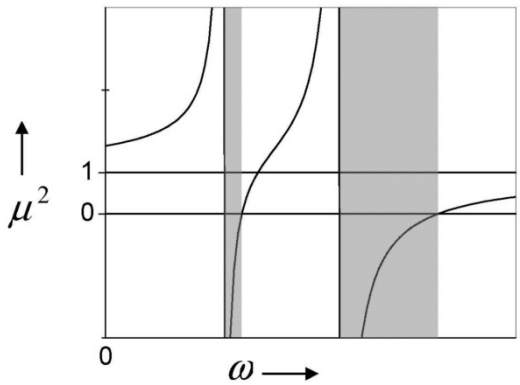


Fig. 15: The square of the refractive index μ^2 versus (angular) wave frequency ω. The shaded regions define stop bands with imaginary refractive index (schematic).

In magnetized plasma, waves may propagate in any direction; however, wave properties like refractive index (phase velocity) and resonance frequencies depend on the direction of propagation. An exception is the cut-off frequencies. They are the same for all directions. Usually one only considers wave propagation in the so-called principal directions: along and across the external magnetic field. Waves are classified according to their properties in the main directions. But these classifications may lose their sense when considering oblique wave propagation.

For cold infinite plasma, the dispersion relation again has two solutions, describing fast and slow wave types. These waves have different names depending on the direction of propagation, frequency range and also on the author. For waves where the electric vector $\vec{E}$ is parallel to the field lines of the external magnetic field (respectively the induction $\vec{B_0}$), the dispersion relation is the same as in the case of non-magnetized plasma. These waves are called ordinary waves. Thus this name is used in plasma physics in a manner different from in crystal optics (where 'ordinary' means waves propagating in any direction with the same phase speed). In plasma, ordinary waves exist only in the so-called main directions of wave propagation: $\vec{k} \parallel \vec{B}_0$ and $\vec{k} \perp \vec{B}_0$. For $\vec{k} \parallel \vec{B}_0$, the condition is fulfilled only for longitudinal waves; for $\vec{k} \perp \vec{B}_0$, only for transverse waves. Ordinary plasma waves propagating in other directions do not exist, that is, the dispersion relations for waves propagating in any other direction differ from that of the magnetic field-free case. In warm plasma, where pressure effects are important, at least two more electrostatic wave types can exist.

In bounded plasma there is a limitation by wavelength. Long-wavelength waves cannot propagate. On surfaces, special types of surface waves are possible. How does one find orientation? For the infinite plasma we will give a general treatment ending in the very helpful Clemmow–Mullaly–Allis (CMA) diagram.

4.3.2 General dispersion relation for cold magnetized plasma

For cold magnetized plasma, the Maxwell equation for plane waves reads

$$\mathrm{i}\vec{k}\times\vec{H}=\vec{j}-\mathrm{i}\varepsilon_0\omega\vec{E}=(\underline{\underline{\sigma}}-\mathrm{i}\varepsilon_0\omega)\vec{E}\equiv-\mathrm{i}\underline{\underline{\varepsilon}}\vec{E}\,. \tag{120}$$

The latter transformation corresponds to considering the currents induced by a wave as displacement current, which is considering plasma as a polarizable medium. Using (112) we obtain for the electric field

$$\left(k^2-\vec{k}\vec{k}-\frac{\omega^2}{c^2}\underline{\underline{\varepsilon}}\right)\cdot\vec{E}=0\,. \tag{121}$$

Here $\vec{k}\vec{k}$ is the dyadic product of the vector $\vec{k}$ with itself. We obtain the dispersion relation from this system of equations by setting the determinant to zero. Writing formally $\underline{\underline{\mu\mu}}\equiv\vec{k}\vec{k}c^2/\omega^2$, we can formulate this as

$$\left|\mu^2\underline{\underline{1}}-\underline{\underline{\mu\mu}}-\frac{\omega^2}{c^2}\underline{\underline{\varepsilon}}\right|=0\,. \tag{122}$$

We now have the problem of determining the dielectric tensor $\underline{\underline{\varepsilon}}$. For this purpose, we must determine $\underline{\underline{\sigma}}$. This is beyond the scope of this chapter, and we will give here only the (general) formula. For details see, for example, Refs. [10] and [11]. It is general use to abbreviate $\underline{\underline{\varepsilon}}$ in the form

$$\underline{\underline{\varepsilon}}\equiv\begin{Bmatrix} S & \mathrm{i}D & 0\\ -\mathrm{i}D & S & 0\\ 0 & 0 & P\end{Bmatrix}, \tag{123}$$

with the abbreviations

$$S\equiv1-\sum_k\frac{\omega_{\mathrm{p}k}^2}{\omega^2}\frac{1-\mathrm{i}\nu_k/\omega}{(1-\mathrm{i}\nu_k/\omega)^2-\omega_{\mathrm{B}k}^2/\omega^2}\,,$$

$$D\equiv-\sum_k\frac{\omega_{\mathrm{p}k}^2}{\omega^2}\frac{\omega_{\mathrm{B}k}}{\omega}\frac{1-\mathrm{i}\nu_k/\omega}{(1-\mathrm{i}\nu_k/\omega)^2-\omega_{\mathrm{B}k}^2/\omega^2}\,, \tag{124}$$

$$P\equiv\sum_k\frac{\omega_{\mathrm{p}k}^2}{\omega^2}\frac{1}{(1-\mathrm{i}\nu_k/\omega)}\,.$$

Here k indicates the different kinds of charged particle (electrons and different kinds of ions), and $\omega_{\mathrm{p}k}$ and $\omega_{\mathrm{B}k}$ are the plasma and gyro-frequencies respectively of species k.

To calculate the determinant, it is most convenient to introduce the coordinate system sketched in Fig. 16. In this system we have $\vec{B_0} = (0.0, B_0)$ and $\vec{k} = (k\sin\theta, 0, k\cos\theta)$.

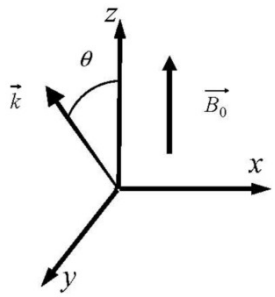


Fig. 16: Coordinate system used for Eq. (125)

With this notation our dispersion relation reads

$$\begin{vmatrix} \mu^2\cos^2\theta - S & -\mathrm{i}D & -\mu^2\cos\theta\sin\theta \\ +\mathrm{i}D & \mu^2 - S & 0 \\ -\mu^2\cos\theta\sin\theta & 0 & -\mu^2\sin^2\theta - P \end{vmatrix} = 0\,. \tag{125}$$

When explicitly calculating Eq. (125), the μ^6 terms vanish and we obtain a bi-quadratic expression for the refractive index,

$$A\mu^4 - B\mu^2 + C = 0\,, \tag{126}$$

with

$$\begin{aligned} &A = S\sin^2\theta + P\cos^2\theta, \\ &B = RL\sin^2\theta + SP(1+\cos^2\theta), \quad R \equiv S + D,\ L \equiv S - D, \\ &C = PRL. \end{aligned} \tag{127}$$

It has two solutions corresponding to *fast* and *slow* waves:

$$\mu_{\mathrm{f,s}}^2 = \frac{B}{2A}\left(1 \pm \sqrt{1 - 4AC/B^2}\right)\,. \tag{128}$$

It is also possible to solve for θ, yielding an expression known as the *Appleton–Lassen equation*

$$\tan\theta = \frac{-P(\mu^2 - R)(\mu^2 - L)}{(S\mu^2 - RL)(\mu^2 - P)}\,. \tag{129}$$

This equation is most convenient for discussing wave propagation in the 'principal' directions $\theta = 0$ and $\theta = \pi / 2$. Here $\theta = 0$ means propagation along $\vec{B_0}$, that is, $\vec{k} \parallel \vec{B}_0$. In this case also $\tan^2\theta = 0$, thus either

- $\mu_r \equiv \sqrt{R}$ (right circularly polarized waves) or
- $\mu_l \equiv \sqrt{L}$ (left circularly polarized waves).

For these waves we find in the approximation of high frequencies, $\omega >> \omega_{Bi}$, a cut-off frequency with $\mu_{r,l} = 0$ for

$$\frac{\omega_p^2}{\omega^2} = 1 \pm \frac{\omega_{Be}}{\omega}. \tag{130}$$

Right circularly polarized waves have a resonance $\mu_r \to \infty$ at the electron cyclotron frequency ω_{Be}, the 'electron cyclotron resonance', where the electric field vector rotates synchronously with the gyrating electrons. Left circularly polarized waves have no resonance at high frequencies. Instead, they become resonant with the gyrating ions at ω_{Bi}.

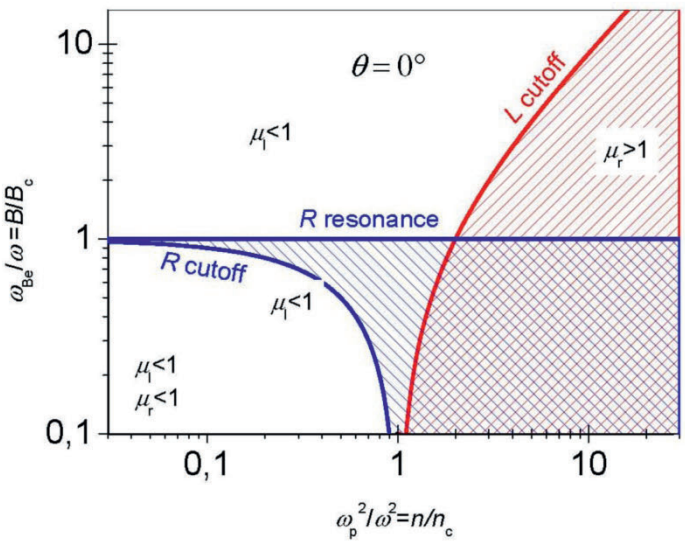


Fig. 17: Location curves for resonances and cut-offs at $\theta = 0°$ in a double-log ω_{Be} versus ω_p^2/ω^2 diagram (this is in principle a so-called CMA diagram).

For the discussion of wave propagation, one usually visualizes resonances and cut-offs by their localization curves in a diagram, where the abscissa displays $\omega_p^2 / \omega^2 \equiv n/n_c$ and the ordinate $\omega_{Be} / \omega \equiv B / B_c$. Here n_c and B_c are called critical density and critical magnetic induction, respectively. These are, respectively, the density for which a given frequency is equal to the plasma frequency, and an induction for which a given frequency is equal to the electron gyro-frequency. Thus this diagram can be read either, for a given frequency, as a diagram of magnetic induction versus plasmas density or, for a given plasma, as a frequency diagram (or as a mixture of both). In this diagram Eq. (130) describes two straight lines, the same holding true for the resonance condition $\omega = \omega_{Be}$. This is shown in Fig. 17, where shaded regions are regions of no wave propagation along the magnetic field lines.

The different patterns indicate stop bands for right and left circularly polarized waves. The sharp resonances and cut-offs appear only when collisions are absent. This figure represents the basic structure of the so-called CMA diagram in the approximation $\omega_{\mathrm{Bi}} = 0$.

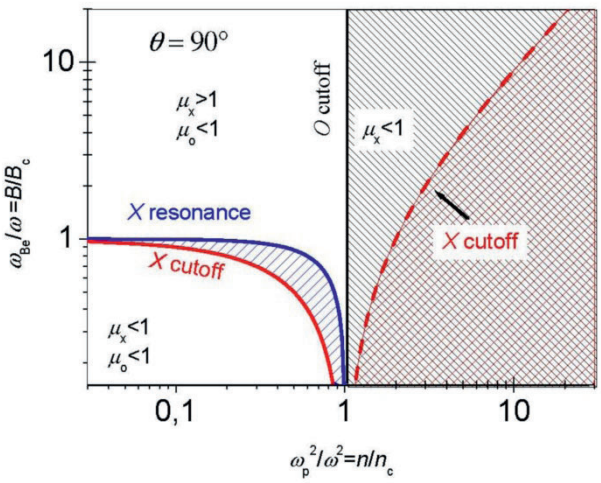


Fig. 18: Simplified CMA diagram for the case $\theta = 90°$ (see also Fig. 17)

In the case of waves propagating across the magnetic field, we have $\theta = 90°$, which implies $\tan\theta \to \infty$. This requires the denominator in Eq. (129) to become zero. For this case again we obtain two solutions: the waves with $\mu_{\mathrm{o}} \equiv \sqrt{P}$ are called *ordinary waves*, while those with $\mu_{\mathrm{x}} \equiv \sqrt{RL/S}$ are called *extraordinary waves*. The name *ordinary* is chosen to express that the refractive index for these waves does not depend on the external magnetic field. Ordinary waves have no resonance, only a cut-off at $\omega = \omega_{\mathrm{p}}$ like waves in magnetic field-free plasma. The cut-offs for the extraordinary waves are the same as for the left and right circularly polarized waves – the localization curves are the same. They are given by $R = 0$ and $L = 0$. The resonance condition for the extraordinary wave is $S = 0$. As resonance frequency we find in the approximation of high frequencies

$$\frac{\omega_{\mathrm{Be}}}{\omega} = \sqrt{1 - \frac{\omega_{\mathrm{p}}^2}{\omega^2}} \,. \tag{131}$$

The respective curves are shown in Fig. 18. There is no wave propagation in the shaded regions. The pattern is used to distinguish between ordinary (o) and extraordinary (x) waves. While ordinary waves do not propagate for any frequency below ω_{p}, extraordinary waves may propagate at lower frequencies if the magnetic field is sufficiently high.

To demonstrate the case of oblique direction of propagation, we show in Fig. 19 the case $\theta = 70°$. Here the designations X, O, R, L are extrapolations from the principal directions. Strictly speaking, these designations are defined only for the principal directions.

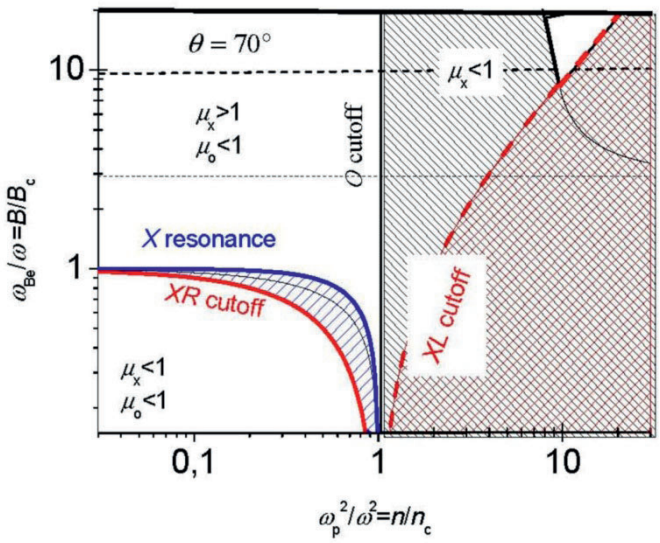


Fig. 19: Simplified CMA diagram for waves propagating at $\theta = 70°$ as an example for oblique propagation (see also Fig. 18).

What we have discussed so far were simplified versions of the so-called *Clemmow–Mullaly–Allis* (*CMA*) *diagram*, which helps in classifying plasma waves. In Fig. 20 we show a more complete version of this diagram. Here the ordinate represents $\omega_{\mathrm{Be}}\omega_{\mathrm{Bi}}/\omega^2$ instead of ω_{Be}/ω to involve the ion effects for hydrogen ions, for instance the resonance of ion cyclotron waves (L-resonance). The curves shown are those for resonances and cut-offs at the principal directions of wave propagation (called *principal resonances* and *principal cut-offs* in the literature). These curves divide the diagram into 13 regions, in each of which exist topological distinct wave-normal surfaces (polar plots of the phase velocity $\overrightarrow{v_{\mathrm{phase}}} = \omega\vec{k}/k^2$ versus the angle of propagation θ). These polar diagrams are shown, but not to identical scales. Instead, the dotted circles represent *c*, the vacuum phase velocity of light, as an aid for orientation. Lemniscate shaped wave-normal surfaces indicate resonances with $v_{\mathrm{phase}} = 0$. In any case these resonances occur in a distinct specific direction. The presentation is organized for the case that the *B* field vector points vertically upwards.

The CMA diagram is very helpful, for example, in discussing the accessibility of a plasma by a certain wave. Imagine the problem of launching a wave from an antenna outside plasma. 'Outside' means that the plasma density *n* is almost zero. If, also, the magnetic field is zero and rises inside the plasma, the conditions at the location of the antenna correspond to the left lower corner of the diagram. This means that the wave will propagate into a stop band limited by a cut-off and cannot penetrate further into the plasma. A possible solution is to install at the antenna a magnetic field with $B > B_{\mathrm{c}}$, thus creating conditions where waves can propagate.

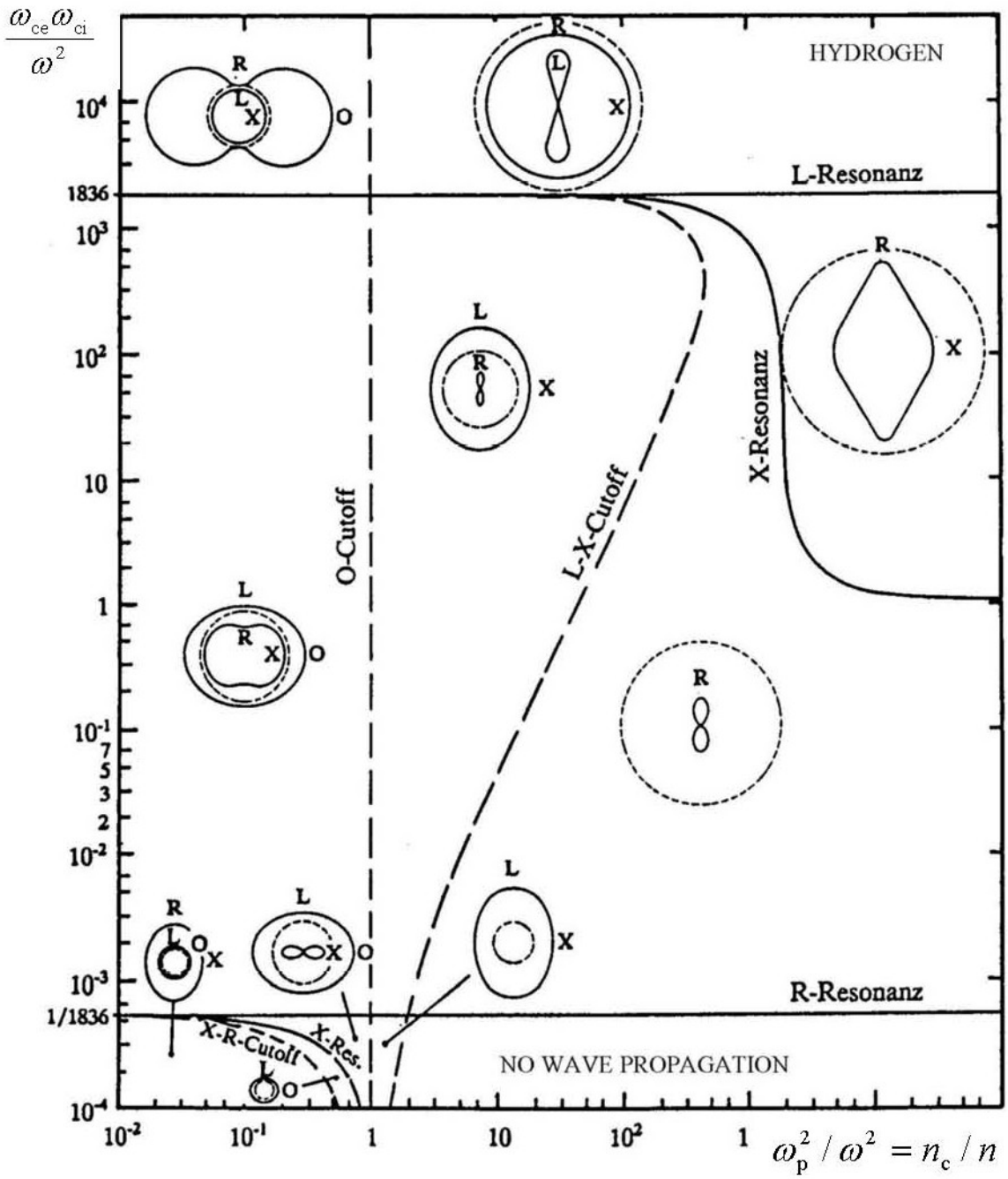


Fig. 20: Complete CMA diagram including ion effects for H^+ ions, such as, for example, the L-resonance for left circularly polarized ion cyclotron waves, and further phase velocity polar diagrams. (From Ref. [14], © G. Jansen, Wangen, Germany, reprinted with permission.)

5 Concluding remarks

Here ends our short excursion into the world of plasma. It was necessarily short and in no way complete. To learn more, one needs to study the literature. There are a number of introductory plasma physics books on the market. In my opinion Ref. [9] is one of the best, at least for a reader new to the field, because it is easy to read and gives a very good overview. However, I am not familiar with the very recently published books, and for those I cannot give a recommendation. In general, these books

emphasize the problems of fully ionized plasma as studied in fusion research and astrophysics. Plasma waves are discussed in much more detail in the book of Stix [10]. Our presentation further owes much to the treatment of wave phenomena in Refs. [10]–[13]. Jansen's book [14] gives many helpful diagrams on wave behaviour in plasma. The special problems of ion source plasma are discussed in a short review in Ref. [15]. Many ion source discharges have much in common with discharges for material processing. The book of Lieberman and Lichtenberg [16] gives a very concise and well-written treatment of the problems of those plasmas and may be found extremely useful when studying the physics of ion source discharges. Last, but not least, the book of Geller [8], the 'father of ECRIS' (electron cyclotron resonance ion source), is a very personal report of the author's involvement and struggle in developing the ECRIS principle and understanding the physics of ECRIS plasma.

References

[1] P. Debye and E. Hückel, *Phys. Z.* **24** (1923) 183 and 305.

[2] W.P. Allis, Motions of ions and electrons, in *Handbuch der Physik*, Ed. S. Flügge (Springer, Berlin, 1956), Vol. XXI, p. 383.

[3] H. Dreicer, *Phys. Rev.* **115** (1959) 238.

[4] L. Stenflo, *J. Nucl. Energy, C Plasma Phys*. **8** (1966) 665.

[5] I.P. Shkarovsky, T.W. Johnston and M.P. Bachynsky, *The Particle Kinetics of Plasma* (Addison-Wesley, London, 1966).

[6] A. Simon, *Phys. Rev.* **98** (1955) 317.

[7] A.G. Drentje, *Rev. Sci. Instrum.* **74** (2003) 2631.

[8] R. Geller, *Electron Cyclotron Resonance Ion Sources and ECR Plasmas* (IOP Publishing, Bristol, 1996).

[9] F.F. Chen, *Introduction to Plasma Physics and Controlled Fusion* (Plenum, New York, 1990).

[10] T.H. Stix, *The Theory of Plasma Waves* (American Institute of Physics, New York, 1992).

[11] W.P. Allis, S.J. Buchsbaum and A. Bers, *Waves in Anisotropic Plasmas* (MIT Press, Cambridge, MA, 1963).

[12] G. Bekefi, *Radiation Processes in Plasmas* (Wiley, New York, 1966).

[13] M.A. Heald and C.B. Wharton, *Plasma Diagnostics with Microwaves* (Wiley, New York, 1965).

[14] G. Jansen, *Plasmatechnik* (Hüthig, Heidelberg, 1992) [in German].

[15] I.G. Brown, in *The Physics and Technology of Ion Sources*, Ed. I.G. Brown (Wiley-Interscience, New York, 1989), p. 7.

[16] M.A. Lieberman and A.J. Lichtenberg, *Principles of Plasma Discharges and Material Processing* (Wiley-Interscience, New York, 1994).

Production of High-Intensity, Highly Charged Ions

S. Gammino
Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali del Sud, Catania, Italy

Abstract
In the past three decades, the development of nuclear physics facilities for fundamental and applied science purposes has required an increasing current of multicharged ion beams. Multiple ionization implies the formation of dense and energetic plasmas, which, in turn, requires specific plasma trapping configurations. Two types of ion source have been able to produce very high charge states in a reliable and reproducible way: electron beam ion sources (EBIS) and electron cyclotron resonance ion sources (ECRIS). Multiple ionization is also obtained in laser-generated plasmas (laser ion sources (LIS)), where the high-energy electrons and the extremely high electron density allow step-by-step ionization, but the reproducibility is poor. This chapter discusses the atomic physics background at the basis of the production of highly charged ions and describes the scientific and technological features of the most advanced ion sources. Particular attention is paid to ECRIS and the latest developments, since they now represent the most effective and reliable machines for modern accelerators.

1 Introduction

Research on ion sources is basically dictated by the growing need for intense currents of multicharged ions in modern nuclear physics facilities. Ion acceleration in a cyclotron obeys the scaling law $E \propto q/A^2$, where q is the charge state and A is the mass number, so the advantages of highly charged ion (HCI) sources is evident. Intense currents are required in order to reduce the acquisition time in low-cross-section nuclear physics experiments, or to increase the production rate of exotic beams in modern facilities such as SPIRAL2, FAIR and so on. To quote one example, the FRIB (Facility for Rare Isotope Beams) project at MSU, Michigan, USA, requires $I \geq 280$ μA for U^{34+}. Similarly, the FAIR facility at GSI, Darmstadt, Germany, will be based on a 1 mA beam of U^{28+}. Multicharged ion production can be obtained in the presence of a flow of electrons or in a plasma. The various ionization techniques differ from each other in terms of the electron temperature, electron density and plasma lifetime that can be achieved. The simplest way to create plasmas consists of electrical discharges in vacuum tubes; other ways use electron beams passing through neutral gases or electromagnetic waves interacting with gases or vapours in the presence of a well-shaped magnetic field. The latter method is preferable for stable and reproducible HCI formation, since the magnetic field lengthens the time spent by the ions in the volume where multiple ionizing collisions with energetic electrons occur. When the plasma heating is provided by means of microwaves and under the electron cyclotron resonance (ECR) condition, the plasmas are named ECR plasmas.

The characterization of the different types of ion sources can be done by taking into account some fundamental quantities (*quality factor*). In particular, as ionization to high charge states requires long times, it is clear that the ion confinement time will be a crucial parameter to determine the maximum achievable charge state. The number of available electrons is also important, because it is clearly related to the number of electrons able to ionize the atoms down to inner shells. Hence a qualitative criterion for evaluating the ability of a given plasma to provide HCIs is given by the quality parameter Q, defined as the product of the electron density n_e with the ion confinement time τ_i:

$$Q = n_\mathrm{e}\tau_i. \tag{1}$$

The following proportionality is valid:

$$\langle q \rangle \propto n_{\mathrm{e}} \tau_{\mathrm{i}}. \tag{2}$$

Motivations leading to such rules will become clearer in the following sections. Here, it is worth mentioning that each source, according to its peculiar characteristics, will have a specific capability to increase the electron density or the ion confinement time, and thus to maximize the Q parameter. Although Q can be considered as the main parameter to determine the quality of ion source performance, the electron temperature T_{e} (or more generally the average electron temperature in the plasma) also plays a key role: it fixes the maximum charge state that can be achieved depending on the given ionization potential. In terms of extracted currents, a confinement time that is too long would limit the amount of ions extracted from the system per unit time. Well-confined plasmas can produce extremely high charge states, but current intensity will be modest, i.e.

$$I \propto \frac{n_{\mathrm{e}}}{\tau_{\mathrm{i}}}. \tag{3}$$

Since Eq. (3) displays an inverse proportionality between current and confinement time, and a direct dependence on the electron density (as in the case of Q), the electron density n_{e} becomes the key parameter if one wants to maximize both currents and charge states.

In order to complete the set of quality parameters, we have to take into account the total background pressure in the chamber where the plasma is created. Fully ionized plasmas are rare, and in many cases the neutral atoms and the lowly charged ions may decrease the mean charge state because of charge exchange processes (whose cross-section is considerably higher than the cross-section of electron recombination). Hence another quality criterion can be inferred, and it states that

$$\frac{n_0}{n_{\mathrm{e}}} \leq f\left(T_{\mathrm{e}}, A, z\right), \tag{4}$$

where f is a function of the atomic mass A, the atomic number z and the electron temperature T_{e}. Finally, in order to perfectly match the ion sources with the accelerators, the beam quality must fit the accelerator constraints, in terms of brightness and emittance.

Additional characteristics of ion sources are envisaged, such as the reliability, the lifetime of the system, the maintenance and, last but not least, the cost (for building and for ordinary operations).

2 Atomic physics background for HCI production

The dynamics of the collisions among the plasma particles is of primary importance for ionization rate determination and for dealing with a great number of other phenomena such as electron and ion lifetime evaluation in magnetically confined plasmas. An atomic physics background is therefore needed to distinguish between the most significant plasma parameters influencing ion source performance. Basically, the collision dynamics is mainly determined by the electron temperature, the electron density and the background pressure.

It is convenient to introduce some physical quantities that are useful for the treatment of collisional dynamics. In particular, once we have defined the *cross-section*, we can introduce the mean free path, the mean time between a collision and the following one, and finally the collision frequency. The analysis of the various collisions in plasmas will be carried out principally in terms of the collision frequency.

If σ is the cross-section of a given collision, then the mean free path will be given by

$$\lambda_{\mathrm{m}} = \frac{1}{n\sigma}, \tag{5}$$

where n is the density.[1] Once the λ_{m} parameter has been defined, the time between two successive

[1] Here we denote the general density by n. The electron and ion densities are usually distinguished by using the notation n_{e} and n_i.

collisions will be

$$\tau = \frac{\lambda_{\mathrm{m}}}{v}, \tag{6}$$

where v is the velocity of the colliding particles. The collision frequency can be easily determined from (6), and is given by

$$\nu = \frac{1}{\tau} = \frac{v}{\lambda_{\mathrm{m}}} = n\sigma v. \tag{7}$$

Generally, in plasmas, all the particles (electrons, ions and neutrals) follow some energy distribution function,[2] and so we cannot consider a single velocity in (7). This expression can be still used to determine ν, but we have to average[3] over all the possible values of σv. Then we obtain

$$\nu = n\overline{\sigma v}. \tag{8}$$

The assumptions made so far are quite general and nothing specific has been said about the several types of plasma collisions. A schematic overview of the main plasma collision processes, and a classification, have been given according to the number of colliding particles: *multiple collisions* are characteristic of plasmas, because they occur thanks to the *long-range Coulomb interaction*; and *binary collisions* are more similar to the collisions in gaseous systems. However, in this case, non-elastic and ionizing collisions occur because of the high energy content of the plasma particles, especially of electrons.

2.1 Multiple-scattering collisions

Note that in plasmas any electric field can be felt by charged particles only within a distance that is named the *Debye length*,

$$\Lambda_{\mathrm{D}} = \left(\frac{\epsilon_0 K_{\mathrm{B}} T_{\mathrm{e}}}{ne^2}\right)^{1/2}. \tag{9}$$

This means that a proof particle feels the electromagnetic interaction due only to those plasma particles located inside the Debye sphere.[4] However, the number of particles included in the Debye sphere is still quite large (of the order of several thousands). A test particle, therefore, will be subjected to multiple interactions for which is more convenient to evaluate the cross section for a cumulative deflection of 90°. Then we can define a cross-section, σ_{90°, given by[5]

$$\sigma_{90^\circ} = 8\pi \left(\frac{z_1 z_2 e^2}{M v^2}\right)^2 \ln\left(\frac{\Lambda_{\mathrm{D}}}{b_{\mathrm{min}}}\right), \tag{10}$$

where z_1 and z_2 are the charges of the two particles, e is the electron charge, M is the mass of the colliding particle (if the other particles are much heavier) and the term $\Lambda_{\mathrm{D}}/b_{\mathrm{min}}$ is the so-called Coulomb logarithm. This equation, except for the logarithm, is practically the Rutherford formula for charged particle scattering (binary collision). The Rutherford cross-section is almost two orders of magnitude lower than (10). This means that a 90° scattering due to a single collision is much less probable than multiple deflections due to many collisions.[6]

[2] In many cases it is convenient to use a Maxwellian-like distribution, for which a temperature can be determined. However, for low-pressure and high-temperature plasmas, the energy distribution function does not follow a pure Maxwellian shape, but it can nevertheless be viewed as a superposition of several plasma populations with Maxwellian distributions at different temperatures.

[3] We must average not only over v, but also over σ because in general we have $\sigma \equiv \sigma(v)$.

[4] The Debye sphere is the sphere with radius Λ_{D}.

[5] The quantities in the formula are usually expressed in the c.g.s. system

[6] This result was formerly guessed by Chandrasekhar [1].

A more precise representation of the multiple collisions can be given in the centre-of-mass system, and the collision frequencies can be calculated by taking into account the collisions among the two different plasma species: electrons and ions. Omitting, for the sake of brevity, the mathematical description, we can use the final formulas to understand the role that each type of collision plays in the plasma:

$$\nu_{90^\circ}^{\mathrm{ee}} = \frac{1}{\tau_{\mathrm{sp}}} = 5\times 10^{-6} n \frac{\ln(\Lambda_{\mathrm{D}}/b)}{T_{\mathrm{e}}^{3/2}}, \tag{11}$$

$$\nu_{90^\circ}^{\mathrm{ei}} = \frac{1}{\tau_{\mathrm{sp}}} \sim 2\times 10^{-6} z n \frac{\ln(\Lambda_{\mathrm{D}}/b)}{T_{\mathrm{e}}^{3/2}}, \tag{12}$$

$$\nu_{90^\circ}^{\mathrm{ii}} = \frac{1}{\tau_{\mathrm{sp}}} \sim z^4 \left(\frac{m_{\mathrm{e}}}{m_{\mathrm{i}}}\right)^{1/2} \left(\frac{T_{\mathrm{e}}}{T_{\mathrm{i}}}\right)^{3/2} \nu_{90^\circ}^{\mathrm{ee}}, \tag{13}$$

where n_{e} is in cm^{-3} and T_{e} is in eV.

Each frequency is usually called the *characteristic Spitzer collision frequency*, while τ_{sp} is the characteristic Spitzer collision time elapsing on average between two subsequent collisions. Once the collision frequency[7] is given, we can easily determine the time between two successive events: $\nu_{90^\circ}^{\mathrm{ee}} \simeq \nu_{90^\circ}^{\mathrm{ei}}$ and $\tau_{90^\circ}^{\mathrm{ee}} \simeq \tau_{90^\circ}^{\mathrm{ei}}$. In contrast, if we consider ions colliding with electrons (ion–electron collisions), we have

$$\tau_{90^\circ}^{\mathrm{ie}} \simeq \left(\frac{m_{\mathrm{i}}}{m_{\mathrm{e}}}\right) \tau_{90^\circ}^{\mathrm{ei}}. \tag{14}$$

Thermalization among different plasma species is governed by collisions. The time needed for a 90° deflection and that needed for energy exchange are defined as

$$\tau_{\mathrm{m}}^{\mathrm{ee}} \sim \tau_{90^\circ}^{\mathrm{ee}} \sim \tau_{90^\circ}^{\mathrm{ei}}, \tag{15}$$

$$\tau_{\mathrm{m}}^{\mathrm{ii}} \sim \tau_{90^\circ}^{\mathrm{ii}} \sim \left(\frac{m_{\mathrm{i}}}{m_{\mathrm{e}}}\right)^{1/2} \tau_{90^\circ}^{\mathrm{ei}}, \tag{16}$$

$$\tau_{\mathrm{m}}^{\mathrm{ei}} \sim \tau_{m}^{\mathrm{ie}} \sim \frac{m_{\mathrm{i}}}{m_{\mathrm{e}}} \tau_{90^\circ}^{\mathrm{ei}}. \tag{17}$$

The subscript "m" indicates the average time needed for energy sharing collisions.

The Spitzer collisions are the dominant processes in highly ionized and low-pressure plasmas, and totally regulate the thermalization processes. From the above formulas, it can be seen that $\tau_{\mathrm{m}}^{\mathrm{ee}} \sim \tau_{90^\circ}^{\mathrm{ee}}$ and $\tau_{\mathrm{m}}^{\mathrm{ei}} \gg \tau_{90^\circ}^{\mathrm{ei}}$, implying a much easier electron–electron thermalization than electron–ion one. As an example, at typical electron energies ranging from a few electronvolts[8] to some kiloelectronvolts,[9] the value of ν_{ee} varies from 10^5–10^6 s^{-1} ($T_{\mathrm{e}} = 10$ eV) to 10^2–10^3 s^{-1} ($T_{\mathrm{e}} = 1$ keV), so that the electron thermalization requires a few microseconds, in the first case, or milliseconds in the second case. The ion thermalization proceeds on a $m_{\mathrm{i}}/m_{\mathrm{e}}$ (i.e. about 10^3–10^4) longer timescale. In an electron cyclotron resonance ion source (ECRIS), where the ion lifetime is shorter than the time required for the ion collisional heating, the ions remain cold. This result has an immediate consequence on the emittance, which can thereby be maintained sufficiently low.[10] Another consequence deriving from the set of equations (15) is that all the electrostatic collision frequencies are much smaller than some other characteristic frequency of the plasma, in particular the Larmor frequency (if the plasma is magnetized) and the microwave frequency (if the plasma is sustained by electromagnetic fields):

$$\omega_{\mathrm{RF}} \gg \nu_{\mathrm{ee}} \quad \text{and also} \quad \omega_{\mathrm{RF}} \gg \nu_{\mathrm{ei}}. \tag{18}$$

Under this condition the plasma is called *collisionless*.

[7] We talk about the collision frequency for the sake of simplicity, but strictly speaking the term 'collision' is incorrect, as we should speak about 90° scattering.

[8] This temperature is typical for the ion sources used to produce lowly charged ion beams at high currents (tens of mA): these sources are usually defined as microwave discharge ion sources.

[9] For ECRIS plasmas for highly charged ions.

[10] It strongly depends on the ion temperature.

2.2 Binary collisions

The binary collisions in plasmas can be divided into two groups: the *electron–neutral collisions* (elastic and inelastic) and the *ionizing collisions*. The former play an important role especially in the case of high-pressure plasmas with a low degree of ionization and a low electron temperature.[11] These collisions provide the thermalization in this kind of plasma, playing the same role as the Spitzer collisions in the case of highly ionized and low-pressure plasmas. As for the creation of multicharged ions in ECRIS, these collisions are not so important, and so we focus our attention on the ionizing collisions.

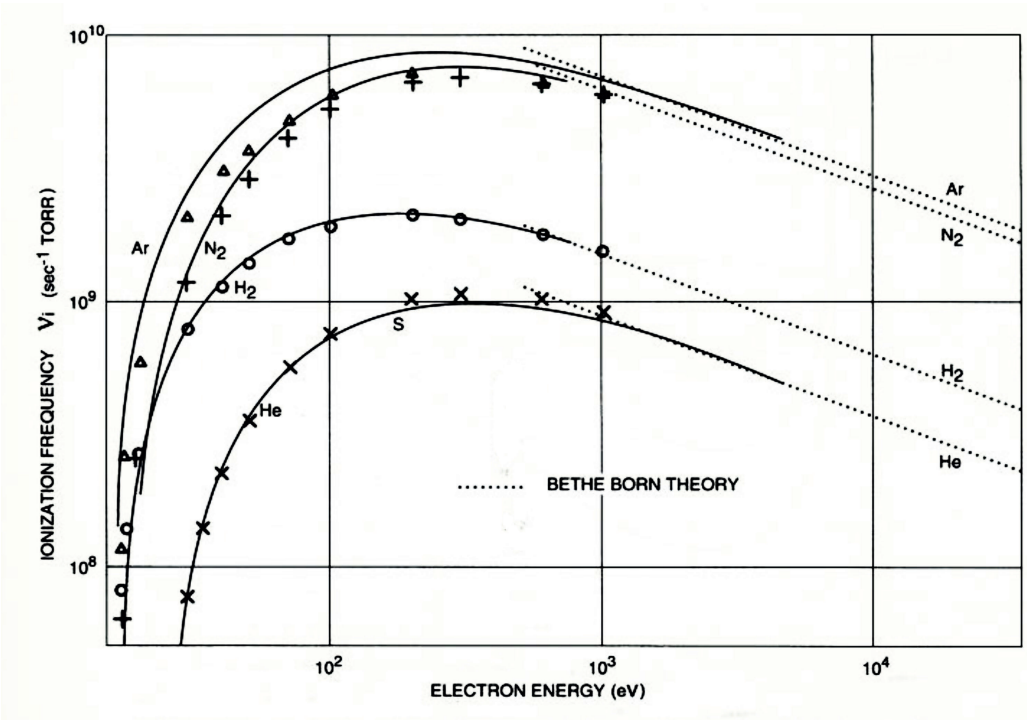


Fig. 1: Ionization frequencies versus the electron energy. For high electron energies the experimental data have been fitted with the Bethe–Born theoretical formula [2].

Figure 1 shows the trend of the ionization frequency versus the electron energy for different ion species. It can be seen that ν_i is parametrized by the background pressure. Hence, to determine the real ionization rate, we have to multiply the frequency shown in the figure by the operating pressure of our ion source. In addition, the trend shown in Fig. 1 considers all the possible ionizing events, i.e. the production rates for all the charge states that can be produced by a single collision. However, it can be demonstrated that generally a single ionizing collision expels just one electron from the atomic shells, thus requiring many collisions to achieve highly charged ions. For higher electron energies, ν_i can be approximated by the Bethe–Born formula:

$$q_\mathrm{i} \propto \frac{1}{E} \ln E \quad \text{or} \quad \nu_\mathrm{i} \propto \frac{1}{\sqrt{E}} \ln E. \tag{19}$$

It is clear from Fig. 1 that the ionizing frequency increases after a threshold value (corresponding to the ionization potential), and then slowly decreases for higher electron energies. If we take into account the usual Maxwellian distribution function of the electrons, we may replace the electron energy on the x axis of Fig. 1 with the electron temperature. For example, in a microwave discharge plasma the electron

[11] By high pressure, we mean, in this case, 10^{-3}–10^{-4} mbar, i.e. the operating pressures of a microwave discharge plasma (MDP). Low temperature here means $T_\mathrm{e} \leq 15$ eV, which also in this case is typical of an MDP.

temperature is usually $\sim$ 10 eV, but the atoms can be easily ionized because of the distribution tail electron component, whose energy is high enough to provide effective ionizing collisions.

In the case of ECRIS, where the production of highly charged ions is of primary importance, a particular relationship among electron density, electron temperature and ion confinement time has to be achieved. This discussion leads to the so-called *multicharged ion (MI) production criterion*. As mentioned above, the ionization down to inner shells of atoms requires a *step-by-step* process, because the cross-section for single ionization per collision is much higher than that for multi-ionization.

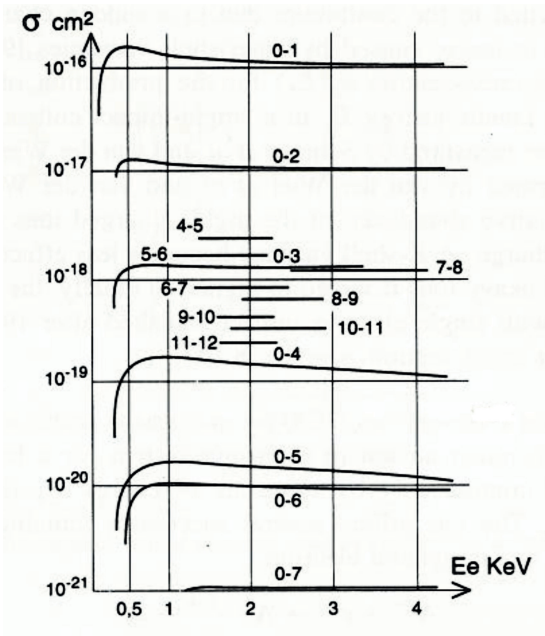


Fig. 2: Single-impact ionization cross-section σ_{0_z} compared to successive-impact ionization cross-sections versus electron energy (Ar).

This is confirmed by Fig. 2, which shows that the cross-section for single ionization due to a single collision is of the order of 10^{-16} cm^2, and then, for multi-ionizations, it rapidly decreases. It turns out that the probability to expel two electrons is one order of magnitude lower; to expel three electrons it is two orders of magnitude lower; and so on. The figure also shows the cross-section for single ionizations, but in the case of highly charged ions (transitions $z \to z+1$).

The time needed for the transition from charge state z_1 to charge state z_2 takes a time, on average, of

$$\tau_{z_1 z_2} = \frac{1}{[n_{\mathrm{e}} \sigma_{z_1 z_2}(v_{\mathrm{e}}) v_{\mathrm{e}}]}, \tag{20}$$

and averaging over a Maxwellian distribution[12] we have

$$\tau_z = \frac{1}{[n_{\mathrm{e}} S_z(T_{\mathrm{e}})]}, \tag{21}$$

$$S_z(T_{\mathrm{e}}) = \langle \sigma_z(v_{\mathrm{e}}) v_{\mathrm{e}} \rangle (T_{\mathrm{e}}). \tag{22}$$

[12]We usually average the collision parameters over a Maxwellian distribution. The energy distributions of ECRIS plasmas differ from a Maxwellian trend, but they can be considered as superpositions of three different distributions, each one having its own temperature. Then, for the sake of simplicity, we assume that the results obtained for a Maxwellian distribution are valid for each plasma population, with the further consideration that the warm component is the most effective to produce highly charged ions. From now on, we will use the parameters of the 'warm' ($\sim$ keV) population to determine the ionizing frequencies of ECRIS plasmas.

If the ion confinement time is longer than the time required for the given ionization (i.e. the time defined in Eq. (21)), the transition $z \to z+1$ takes place. Then the condition to produce highly charged ions in plasmas can be written as

$$n_{\mathrm{e}}\tau_{\mathrm{i}} \geq \frac{1}{[S_z(T_{\mathrm{e}})]}. \tag{23}$$

In plasmas with long confinement times, the step-by-step ionization provides highly charged ions. Equation (23) can be rewritten by substituting the $S_z(T_{\mathrm{e}})$ parameter, obtaining

$$\xi n_{\mathrm{e}}\tau_{\mathrm{i}} \geq 5 \times 10^4 (T_{\mathrm{e}}^{\mathrm{opt}})^{3/2}, \tag{24}$$

where $\xi \equiv \sum_{j=1}^{N} q_j$ is the number of subshells in the atomic outer shells.

The criterion that comes from Eq. (24) is of paramount importance, as it fixes the plasma operational parameters needed for good-performance HCI sources. Some numerical examples can be given for the required densities, confinement times and electron temperatures. If one aims to obtain completely stripped light ions, then $T_{\mathrm{e}} \simeq 5$ eV, with $n_{\mathrm{e}}\tau_{\mathrm{i}} \geq 10^{10}$ cm^{-3}, is needed. The required electron temperature is given by the energy threshold of the given charge state, to obtain:

$$T_{\mathrm{e}}^{\mathrm{opt}} \simeq 5W_{\mathrm{thr}}. \tag{25}$$

A more general picture of the criteria linked to the plasma parameters is given by Fig. 3, which features what combinations of temperatures, densities and confinement times are able to produce different ion species with different charge states. This is called a *Golovanivsky plot*, as it was presented for the first time by the late Russian scientist more than 20 years ago.

2.3 Recombination effects

The charge state of highly charged ions may decrease because of some electron–ion recombination process. The recombination mechanisms are of three different types:

- *Radiative recombination.* Free electrons are captured by ions with the subsequent emission of a photon.
- *Dielectronic recombination.* This is a three-step process, where an electron first excites an ion, then it is captured, and finally a photon is emitted.
- *Charge exchange recombination.* Highly charged ions, colliding with low charge species and/or with neutrals, capture some electrons from the electronic cloud of the other particle, thus decreasing their own charge state.

By analysing the cross-sections of each mechanism, it turns out that only charge exchange recombination plays a marked role in ion sources. Because of its nature, the recombination rates must depend on the background pressure: the higher the number of neutral or low-charge-state ions, the higher is the probability of charge exchange collisions. Again, in order to calculate the correct parameters for ECRIS, the time needed for a single charge exchange process must be evaluated. The charge exchange collision time must be longer than the ionization time for a given charge state (see Eq. (24)). Bypassing the mathematical passages, we can finally write

$$\frac{n_0}{n_{\mathrm{e}}} \leq 7 \times 10^3 \xi \frac{\sqrt{A}}{z[T_{\mathrm{e}}^{\mathrm{opt}}]^{3/2}}, \tag{26}$$

where n_0 is the density of neutrals, and A is the atomic mass number.

Then, in order to obtain fully stripped argon ions (Ar^{18+}) at $n_{\mathrm{e}} \simeq 10^{12}$ cm^{-3}, we need an operational pressure of $\sim 10^{-7}$ mbar, corresponding to a number of neutrals per cubic centimetre $\sim 10^9$ cm^{-3}.

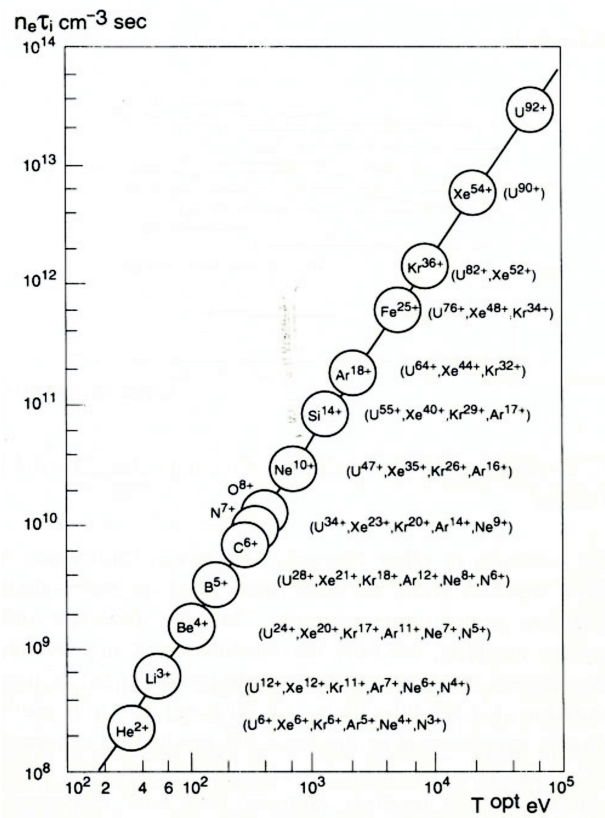


Fig. 3: Golovanivsky's plot of the $(n_e\tau_i)$ criteria for the production of highly charged ions. The ions enclosed by circles are completely stripped; then some combinations of electron temperature, electron density and ion confinement time allow one to produce completely stripped ions. Inside the brackets, incompletely stripped ions are shown: they can be produced with the corresponding plasma parameter of ions enclosed in circles.

These numbers are typical of many modern ECRIS, where the minimum operational pressure is between 10^{-8} and 5×10^{-7} mbar.

In summary, the analysis of collision dynamics in plasmas highlights the importance of low background pressure and of plasma confinement by means of magnetic fields. For low pressures ($p < 10^{-4}$ mbar), the mean free path for ionizing collisions becomes longer than the plasma itself:[13] $\lambda_i \geq 300$ cm. Only by means of a magnetic field is it possible to have some collisions in times shorter than the ion lifetime. Electrons spiral around the field lines and in addition they are reflected by a magnetic mirror, so that they go back and forth inside the magnetic trap and collide with neutral atoms or charged ions.

3 Ion sources for high charge states and high current intensity

The following are the principal ion sources able to fulfil the needs of modern accelerators in terms of high charge states and high current:

1. Electron beam ion sources (EBIS)
2. Laser ion sources (LIS)
3. Electron cyclotron ion sources (ECRIS)

[13] Plasma dimensions are comparable with ones of the metallic chambers usually employed in ion sources for sustaining the plasma, i.e. some tens of centimetres.

On the basis of the discussion in the previous section, a comparison of the data (both for currents and for charge states) often favours ECRIS, which in fact are the most used in laboratories all over the world. Otherwise, for specific applications, EBIS offer remarkable advantages (beam quality), whereas LIS are able to provide very high current with a short pulse.

3.1 Electron beam ion sources (EBIS)

The EBIS sources [3] follow a simple principle: an electron beam is focused by means of a strong magnetic field generated by a solenoid, thus creating a plasma with ions confined by the potential well due to the focused electron beam. A simplified picture of this mechanism is shown in Fig. 4. The atoms filling the plasma chamber are ionized by the electron beams. Once they became ions, they are confined by the potential well, and their confinement time increases as their charge state increases.

Cylindrical electrodes surrounding the device can be used to increase the potential well, thus increasing the ion confinement time and producing higher charge state. Figure 4 schematically describes an EBIS. The solenoids for plasma confinement are shown, and also the extraction system.

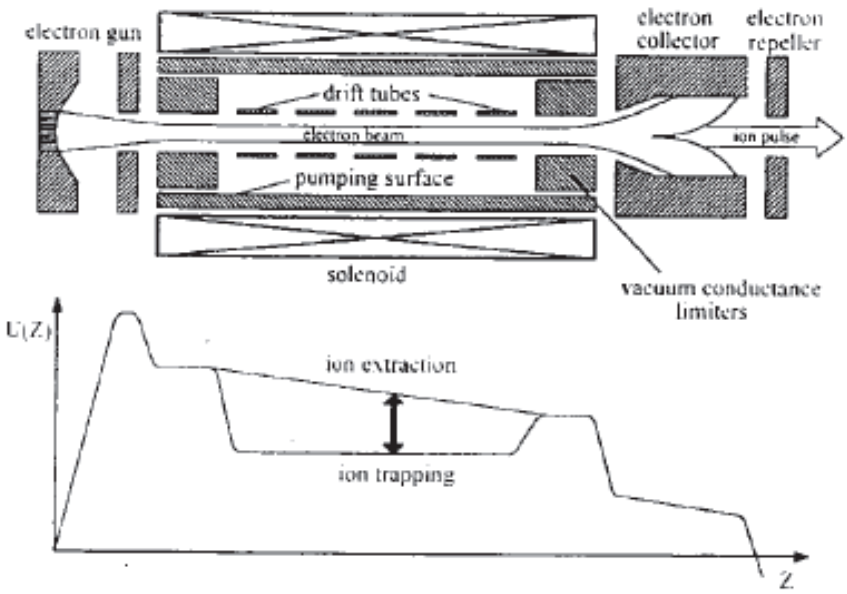


Fig. 4: EBIS scheme principles [3]

EBIS are widely used as the injectors for synchrotrons of completely stripped light ions (e.g. Dubna, Saclay or Stockholm). However, they can be also used as the injector of highly charged heavy ions (e.g. CERN or LNBL).

For very high charge states, the maximum currents are no higher than some nanoamps, but nevertheless they are able to produce charge states up to Xe^{53+} and Xe^{54+} (JINR Dubna, Russia). Some experimental tests carried out at the Los Alamos National Laboratories, USA, have demonstrated that the creation of totally stripped uranium ions is possible, even if the extracted current is of the order of a few tens of ions per second. Although these machines are completely useless for accelerators, they allow one to carry out very interesting experiments for atomic physics.

In these devices, to reach such high charge states, the second term of the expression (2) has been maximized. Because of this trapping optimization, these sources are usually named *electron beam ion traps* (EBIT). The expression (3) explains how only a few ions per second are produced in the Los Alamos

machine, as the considerable increase of the confinement time is not accompanied by a corresponding increase of the electron density.

Finally, we can summarize the main advantages and disadvantages of EBIS:

– *Advantages*
  High charge states can be easily produced
  Beam quality is quite good because of the low-temperature ions (emittance is negatively affected by high-temperature ions)
  High possibility of obtaining various pulse lengths or continuous-wave beams
– *Disadvantages*
  Low currents
  High realization costs
  Very complex device

3.2 Laser ion sources (LIS)

Laser ion sources [3] are based on laser beams focused onto a solid target surface (Fig. 5). If the power density of the laser beam is high enough, then a rapid vaporization of the target material occurs, and a *plasma plume* is generated, expanding at supersonic velocities (higher than 10^4 m s^{-1}). Figure 6 shows the expanding plume; this photo was taken during a recent experiment carried out at LNS [4] with a Langmuir probe, able to characterize the plasma plume in terms of temperature and density for different times from a laser pulse. It helps to understand the plasma dimensions: the probe was placed 1.5 cm from the tantalum target, and the plume front reaches the metallic tip after approximately 1 μs. The transverse plasma dimension is comparable with the probe tip length, about 1 cm.

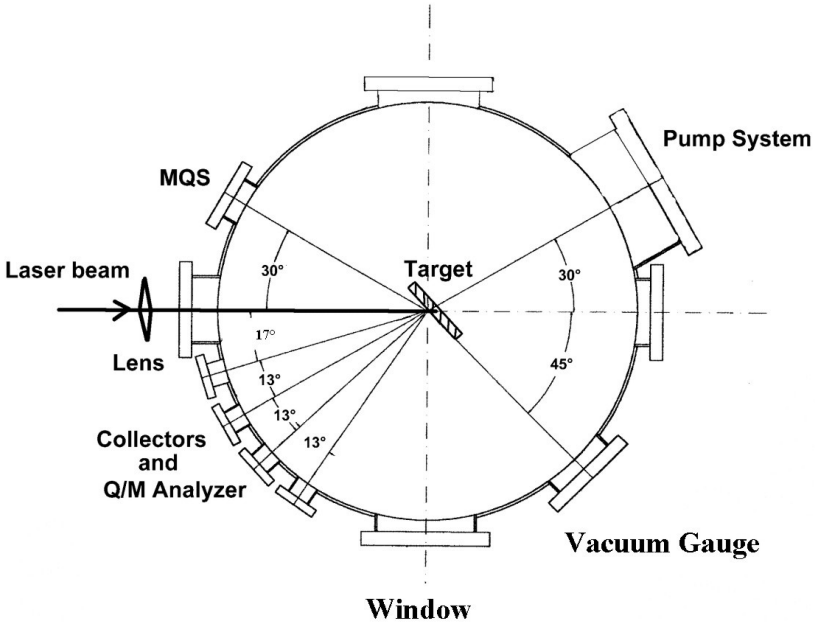


Fig. 5: Schematic view of a LIS source

Electrons generated during the target evaporation are heated by the remaining part of the laser radiation up to several tens of electronvolts when a Q-switched Nd:YAG laser, operating at 1064 nm wavelength, with 9 ns pulse duration and 1–900 mJ energy, is used. Such a laser operates at INFN-LNS, where a LIS is used instead of an ion source to study the physics of plasma produced by laser ablation.

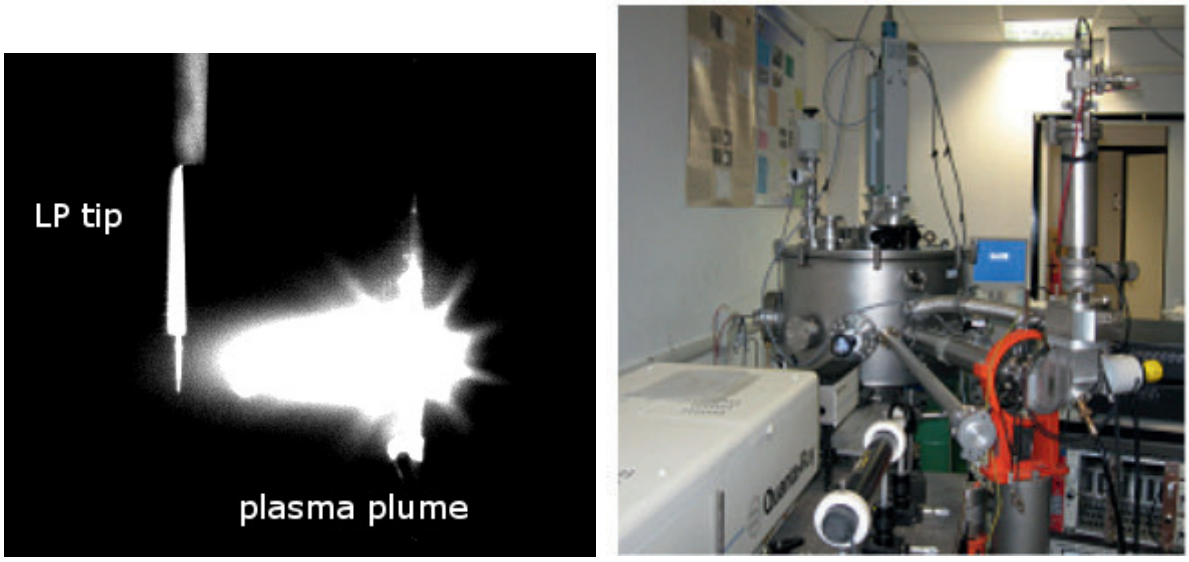


Fig. 6: Plasma plume in a LIS source and the experimental apparatus for Langmuir probe measurements of laser-generated plasma properties.

Usually the laser beam is focused through a convergent lens onto a metal target (Ta, Ag, Cu, etc.) placed inside a vacuum chamber at 10^{-6} mbar. The laser beam incidence angle is lower than 90° with respect to the target surface and the spot diameter is 1 mm. The plasma plume is emitted along the normal to the irradiated target whatever the direction of laser beam is with respect to the normal direction.

In general, the laser energy varies, from 1 J to 50 J, with *pulse lengths* up to 1 μs and *repetition rate* up to 1 Hz for CO_2 lasers. The electron temperature, T_e, and the charge state distribution depend on laser properties as

$$T_e \approx (P\bar{Z})^{2/7}, \tag{27}$$

where P is the laser power. Figure 5 shows in particular the scheme of the LNS's laser source. The system is composed of different parts:

- laser beam generator;
- optical system for laser focalization;
- vacuum chamber where the target is located;
- several diagnostics tools for analysis of ion properties (time-of-flight (TOF) measurements, ion energy analysers (IEA), etc.);
- pumping system for vacuum.

The main limitation to the use of LIS in accelerator facilities is linked to the low repetition rate. This is mainly due to the laser pumping, which requires a lot of time, especially in the case of high-power, high-intensity lasers. Anyway, as mentioned above, LIS are powerful methods to investigate non-equilibrium plasma physics, and they find applications in many fields of fundamental and applied physics (one of the most recent applications concerns cultural heritage).

The most important application of laser sources to accelerator physics consists of using the expanding plumes as accelerators themselves. It has been demonstrated that a strong electric field can be generated inside the plasma plume, and these fields can accelerate the ions up to MeV energies in the case of a laser with high power density. The possibility to install a small accelerator based on non-equilibrium expanding plasmas is under investigation in many laboratories, and also at LNS some interesting results have been obtained with low-power-density lasers, accelerating ions up to 10 keV [5].

If the plasmas produced by laser ablation are used as ion sources, the performance summarized in Table 1 can be achieved.

Table 1: Characteristic performance of laser ion sources.

Charge state	Power density (W cm^{-2})	Current (mA)	Pulse time (μs)
1–2	10^9	10–100	20
$z \geq 10^+$	$\geq 10^{12}$	0.5–10	5–10

4 Electron cyclotron resonance ion sources (ECRIS)

The ECR ion sources are characterized by high-level technology; and the use of superconducting magnets and high-frequency microwave generators has became mandatory for third-generation sources. The general layout of these sources can be summarized as in Fig. 7.

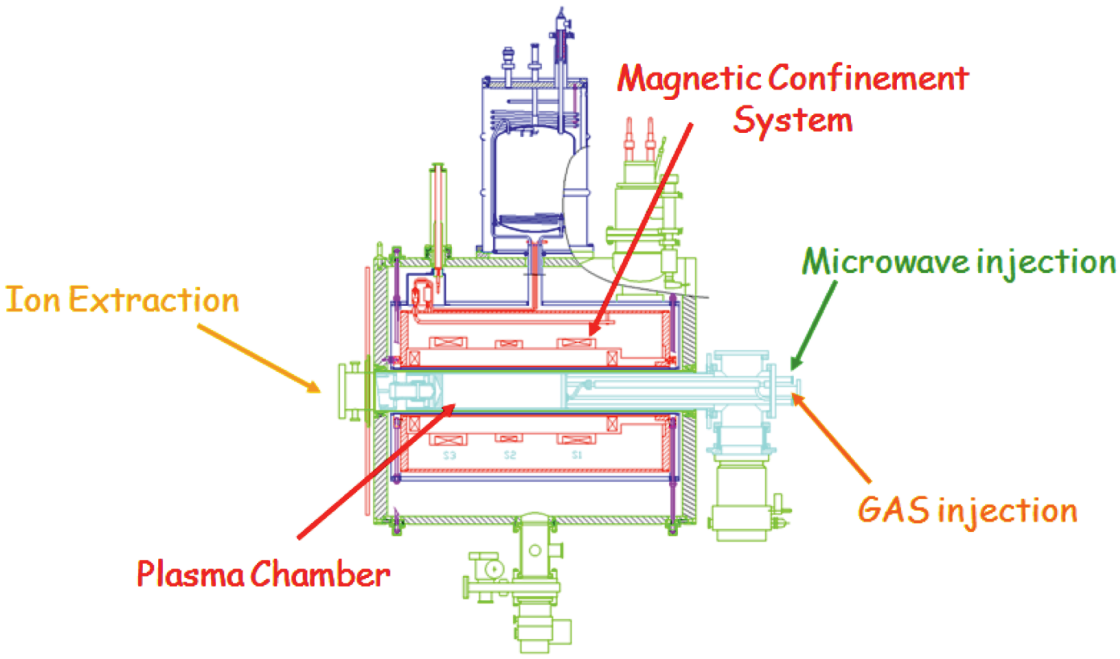


Fig. 7: Main subsystems of the ECRIS

ECR ion sources are based on the so-called electron cyclotron resonance, which ensures resonant absorption of electromagnetic energy by the plasma electrons. To this end, the plasma must be created in the presence of a magnetostatic field appositely shaped to guarantee effective plasma confinement. In brief, an ECRIS works with a gas or vapour injected into a vacuum chamber, with microwaves feeding the cavity and in the presence of an appositely shaped magnetic field. A few free electrons spiralling around the magnetic field lines are accelerated (and then heated) by the microwaves according to the process called electron cyclotron resonance [2]. The electron–atom collisions allow multiply charged ions to be obtained, and very high charge states can be reached because of the quite long confinement times.

Here we can summarize the advantages and disadvantages of such machines:

– *Advantages*
  Intense currents even for highly charged ions
  Long source lifetime
  High stability
  Capability to produce CW and pulsed beams
– *Disadvantages*
  High power consumption
  Expensive and complex technology both for magnets and for microwave components
  High ripple for highly charged ion beams (up to $\sim 10\%$)
  Long conditioning times
  Some difficulties for the production of metallic beams

4.1 ECRIS magnets technology

The different subsystems that compose the ECR ion sources are shown in Fig. 7. An efficient confinement system is mandatory to obtain high-density and stable plasmas. It consists of two or more solenoids for the axial confinement and a hexapole for the radial confinement. The dynamics of the plasma electrons along the plasma chamber axis can be described on the basis of simple mirror system characteristics. In order to obtain the minimum B (min-B) configuration, which ensures magnetohydrodynamic (MHD) stability, a sextupole encloses the plasma chamber, creating a field increasing monotonically along the chamber radius. Figure 8 illustrates the design of an ECRIS.

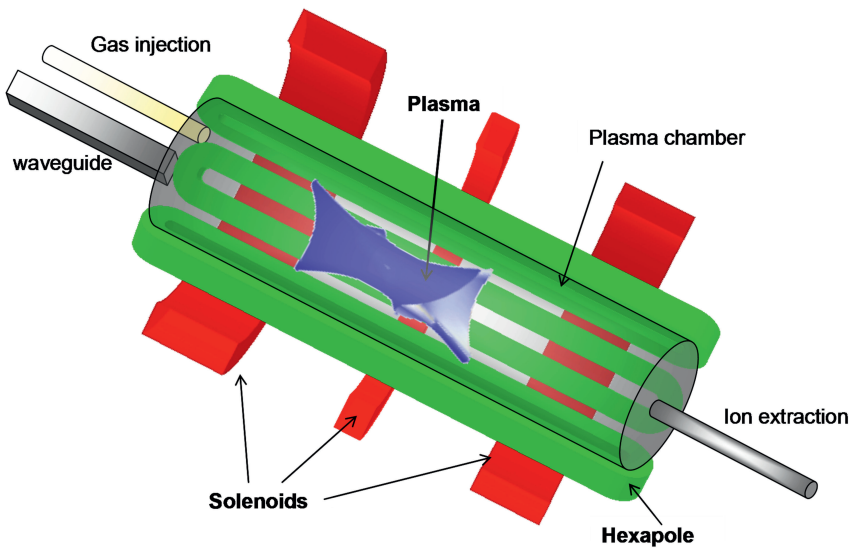


Fig. 8: Location of the main ECRIS devices (magnets, microwave injection, gas injection, ion extraction, plasma chamber) together with the six-cusp plasma generated in the central region of the plasma chamber.

The technological problems related to the magnetic systems used for ECRIS are not simple to solve, because of the very high values of the magnetic field needed for the source operations. In order to give an estimation of the magnetic field needed for operations at high frequencies, we report the expression giving the strength of the magnetic field at a fixed microwave frequency:

$$B_{\mathrm{ECR}} = 2\pi \frac{m}{q} f = 0.03577 f \quad (\mathrm{GHz}). \tag{28}$$

Hence, if $f = 18$ GHz, then $B_{\mathrm{ECR}} \simeq 0.64$ T. Obviously this value of the magnetic field concerns the resonance, but to ensure high stability and long confinement times the magnetic field must be considerably higher. In third-generation ECRIS (like the VENUS source operating at LNBL, California), magnetic fields up to 3.7 T are used. Furthermore, the possibility to operate with variable magnetic field is highly appreciated, as this allows the source performance to be optimized. In general, several types of multipoles may be used, and it is known that the higher the multipole order, the higher will be the loss area of the magnetic structure. Larger areas mean higher ion currents, but on the other hand it lowers the ion lifetime and thus the mean ion charge state (see Eq. (2)). In order to reach a compromise between the extraction current and the ion lifetime, the hexapole is employed in the most of ECRIS. The loss area of the hexapole is shown in Fig. 9.

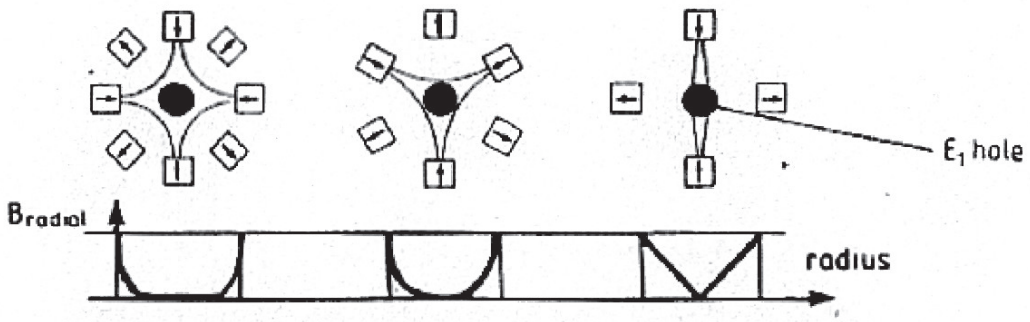


Fig. 9: Loss area and trend of the radial magnetic field in the case of multipolar magnets

Figure 9 shows the loss area shape for three different magnetic multipoles (octupole, hexapole and quadrupole). It also shows the radial trend of the magnetic field for each magnet: the field grows on going from the centre to the periphery. Let us define by B_{m} and r_{m} the magnetic field (on the multipole walls) and the radius of the multipole, and by r_{C} the plasma chamber radius. Then the magnetic field on the plasma chamber walls is given by

$$B_{\mathrm{C}} = B_{\mathrm{m}} \left(\frac{r_{\mathrm{C}}}{r_{\mathrm{m}}} \right)^{\frac{1}{2}n-1}, \tag{29}$$

where n corresponds to the multipole order. Note that the higher the multipole order, the lower will be the value of the magnetic field at the chamber walls, which is detrimental for the confinement. Hence the hexapole is the best compromise between the maximum value of the magnetic field and the loss area. Figure 10 shows the superconducting hexapole of the MS-ECRIS source. In this case the maximum hexapolar field on the chamber wall should be 2.7 T (the maximum available for ECRIS currently operating in the world).

4.2 The microwave injection system

The microwave injection system consists of one or more waveguides with the corresponding ports located in the injection flange of the plasma chamber. As an example, Fig. 11 shows the injection flange with the location of the microwave ports for the SERSE source.

The insulation of the chamber is ensured by a DC break placed in proximity to the injection flange. When two waveguide ports are present, the two injection ports are often perpendicular to each other. The *microwave generators* are usually of two types: *klystron generators* and *travelling wave tube (TWT) amplifiers*. For third-generation sources the gyrotron-based generators are used, with frequencies of 28 GHz or higher.

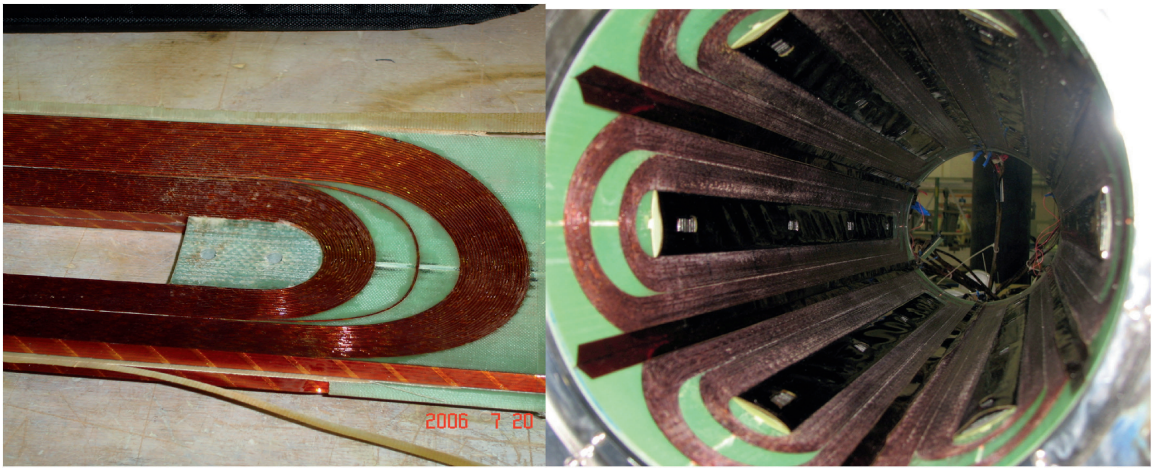


Fig. 10: Superconducting coils and hexapole assembly for the MS-ECRIS source

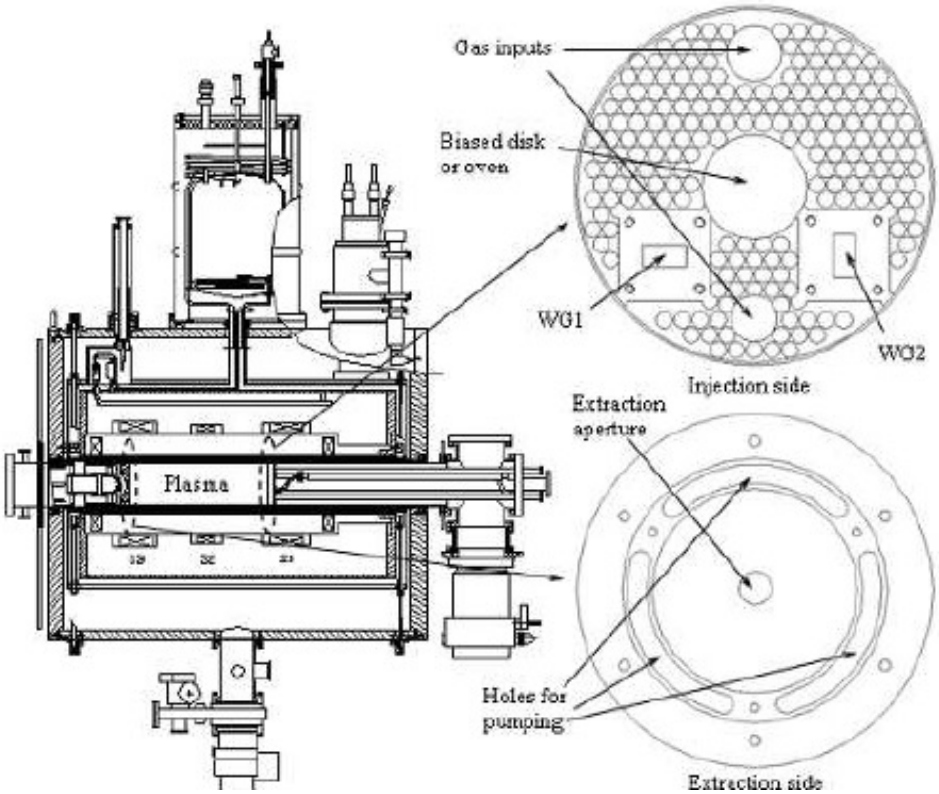


Fig. 11: Detail of the injection flange of the SERSE source with the two microwave ports and a sketch of the microwave generators–plasma chamber connections via waveguides.

4.3 The plasma chamber

The plasma chambers used for ECRIS usually consist of a cylindrical cavity with stainless steel or aluminium walls. As an example, the SERSE chamber is made of aluminium and has a radius of 6.5 cm and a length of 45 cm. The volume of the MS-ECRIS source, designed for the GSI Laboratory, Darmstadt, is slightly bigger, with a chamber radius of 7.5 cm and length of 50 cm. The chamber injection flange includes the *gas injection system* together with the holes for vacuum pumping and the microwave ports. On the opposite side of the plasma chamber there is the *extraction system*. The chamber has a hole, which permits the extraction of the ions. It is also used for the chamber vacuum pumping; it is usually biased up to high voltages (e.g. 25 kV for SERSE). The extraction of intense beams is ensured by a two- or three-electrode system. Details about the extraction system of ECRIS can be found elsewhere [2].

4.4 Additional devices

Additional devices are used for the ECRIS operations. A *solenoid* or other focusing elements are adopted after the extraction system. *Faraday cups* are used to measure the beam current; profilers and emittance measurement devices permit the beam quality to be defined. A magnetic dipole is used as charge-over-mass particle selector, and it permits the charge state distribution (CSD) to be determined. A pressure of the order of 1×10^{-7} mbar can be reached by means of rotative and turbomolecular pumps.

The ECRIS are also able to obtain metallic ion beams, but a gaseous medium needs to be ignited and transformed into plasma through the electron cyclotron resonance. The production of metallic beams requires the presence of an *oven* that can be installed into the plasma chamber. The oven can operate with different principles: resistive or inductively heated; and it is often able to reach a temperature higher than 2000°C. Figure 12 shows the inductively heated oven built for MS-ECRIS.

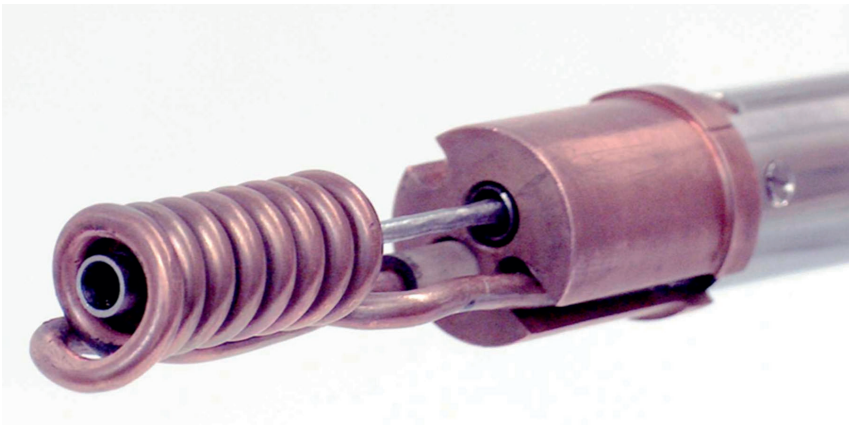


Fig. 12: The inductively heated oven constructed for the MS-ECRIS source

Some problems arise when metals from refractory elements have to be used (like tungsten, tantalum, etc.). For these cases a new method named ECLISSE was developed at LNS; it couples the advantages of the ECRIS in terms of HCI beams and high current, low energy spread and low emittance, with the possibility to obtain ions of several metals by means of laser ablation. Such a method may represent a solution for HCI production from metals, as it may give very high currents of extremely high charge states, but it requires that the ECRIS plasma has a high energy content $n_{\mathrm{e}}kT_{\mathrm{e}}$.

In order to increase the electron density, additional tools have been studied, designed and tested over the past years. In particular the *biased disc* is widely used and it consists of a metallic and cylindrically shaped electrode inserted into the plasma and biased at different voltages (from a few tens to hundreds of volts). Also its position inside the cavity can be varied, but usually it stays outside the plasma core, that is, out of the region of the plasma inside the resonance surface. This device permits

replacement of the electrons lost because of confinement leakages. Thus a biased electrode immersed into the plasma is able to supply additional electrons and to reduce the so-called 'electron starvation' that limits the performances of ECRIS, by reducing the number of ionizing collisions inside the plasma core.

Another ancillary component for modern ECRIS with superconducting magnets is the cryostat containing liquid helium. Figure 13 shows the design of the MS-ECRIS cryostat surrounding the plasma chamber. Owing to the complexity of ECRIS systems and to the parameters that need to be controlled (as well as for safety reasons), a remote control panel is usually installed, which in particular allows the microwave generators' power (and frequency where possible) to be varied and the magnetic field profile (in the case of variable magnetic field) to be changed.

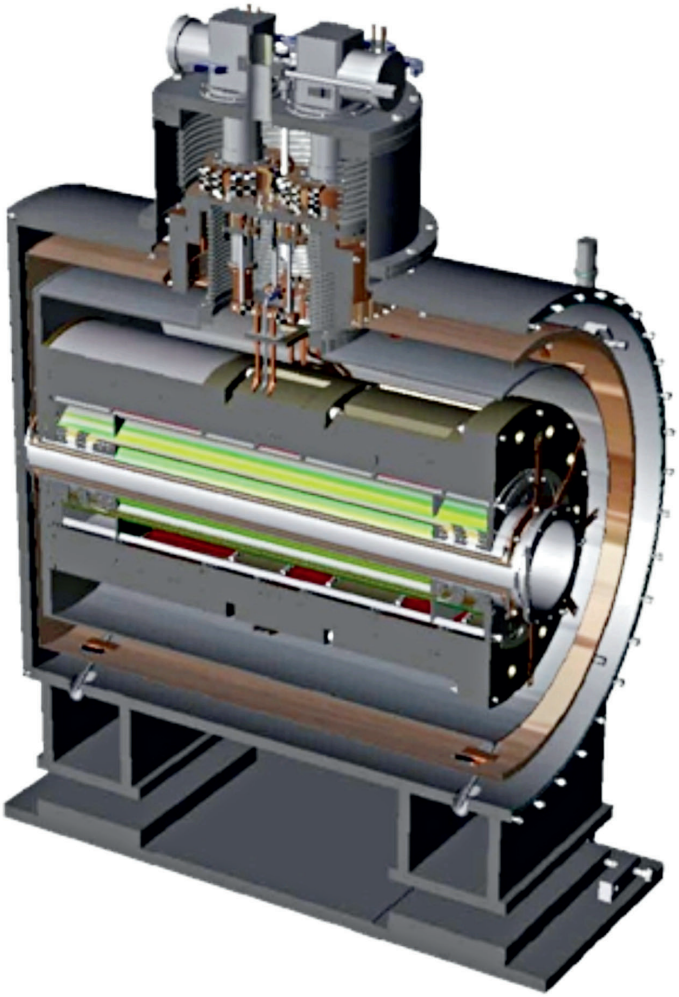


Fig. 13: The design of the MS-ECRIS cryostat and of the plasma chamber

5 Fundamentals about magnetic confinement of plasmas

Confinement of charged particles, and thus of plasmas, is of primary importance in many fields of physics. In the case of ion sources, the ions must live for milliseconds inside the plasma, so that they are subjected to a number of collisions. Confinement can be ensured by electric or magnetic fields. While the effect of an electric field on charged particles is quite trivial to discuss, the role of a non-uniform magnetic field is not. Magnetic confinement can be studied in either a single-particle or one-fluid approximation.

Let us start by recalling the main equations of fluid dynamics as applied to plasmas. The general equation that describes the momentum transport in a plasma subjected to electric and magnetic fields is

$$mn\left[\frac{\partial \boldsymbol{u}}{\partial t}+(\boldsymbol{u}\cdot\boldsymbol{\nabla})\boldsymbol{u}\right]=qn(\boldsymbol{E}+\boldsymbol{u}\times\boldsymbol{B})-\boldsymbol{\nabla}p. \tag{30}$$

In the one-fluid approximation, i.e. when the fluid equations are written for a mixture of ions and electrons and the average mass density and velocity are calculated, we can make some approximations that considerably simplify the model. We suppose that the particles move with velocities less than c (as in the case of the single-particle approximation), we take into account the plasma quasi-neutrality (i.e. $n_{\mathrm{e}} \simeq n_{\mathrm{i}}$), we neglect the ratio $m_{\mathrm{e}}/m_{\mathrm{i}}$ and minor terms, we consider only long-spatial-scale phenomena, with the further restriction that only low-frequency phenomena occur, and finally we consider that the system remains isotropic at all times. Hence the set of so-called *magnetohydrodynamic* equations becomes

$$\frac{\partial \rho_m}{\partial t}+\nabla\cdot\rho_m\boldsymbol{V} = 0, \tag{31}$$

$$\rho_m\frac{\partial \boldsymbol{V}}{\partial t} = \frac{\boldsymbol{J}\times\boldsymbol{B}}{c}\nabla p, \tag{32}$$

$$\boldsymbol{E}+\frac{\boldsymbol{V}\times\boldsymbol{B}}{c} = \eta\boldsymbol{J}, \tag{33}$$

$$\boldsymbol{\nabla}\times\boldsymbol{E} = -\frac{1}{c}\frac{\partial \boldsymbol{B}}{\partial t}, \tag{34}$$

$$\boldsymbol{\nabla}\times\boldsymbol{B} = \frac{4\pi\boldsymbol{J}}{c}. \tag{35}$$

where the subscript "m" in the density term indicates ions or electrons respectively, since the equations are valid for both species.

From Eq. (32), for plasmas in the stationary state, we obtain

$$\boldsymbol{\nabla}p=\boldsymbol{J}\times\boldsymbol{B}. \tag{36}$$

From Eq. (36) it follows that the currents and the magnetic field lines lie on isobaric surfaces, and the vectors $\boldsymbol{J}$ and $\boldsymbol{B}$ are orthogonal each other.

Furthermore, by making use of the equation

$$\boldsymbol{\nabla}\times\boldsymbol{B}=\mu_0\boldsymbol{j}, \tag{37}$$

Eq. (31) becomes

$$\boldsymbol{\nabla}\left(p+\frac{B^2}{2\mu_0}\right)=\frac{1}{\mu_0}(\boldsymbol{B}\cdot\boldsymbol{\nabla})\boldsymbol{B}. \tag{38}$$

Assuming that $\boldsymbol{B}$ varies weakly along its direction, we can write

$$p+\frac{B^2}{2\mu_0}=\text{constant}. \tag{39}$$

In Eq. (39) $B^2/(2\mu_0)$ is the magnetic pressure. The sum of the kinetic pressure and the magnetic pressure is constant, and thus, in a plasma with a density gradient, the magnetic pressure will be high where the particle pressure is low (i.e. in the plasma periphery), and it will be low where the plasma density is high. This diamagnetic effect is due to the so-called *diamagnetic current*, which arises as a consequence of the pressure gradient. The intensity of the diamagnetic effect can be evaluated by taking the ratio of the two terms in Eq. (39). Such a ratio is usually indicated as the β *parameter*:

$$\beta\equiv\frac{\sum nkT}{B^2/2\mu_0}. \tag{40}$$

Generally, β can be used to determine the extent of the plasma confinement:

(a) for $\beta \ll 1$, the magnetic pressure is much higher than the plasma kinetic pressure, which is the condition for an optimal confinement;
(b) for $\beta > 1$, the plasma pressure is higher than that due to the magnetic field, and thus the plasma cannot be magnetically confined; and
(c) for $\beta = 1$, the diamagnetic effect generates an internal magnetic field exactly equal to the external one, and thus two regions exist, one where only the magnetic field is present, and the other where only the plasma, without magnetic field, exists.

It can be demonstrated that only in case of $\beta \ll 1$ is the plasma well confined; otherwise the equilibrium is unstable.

The simplest device for plasma confinement was investigated long ago by Fermi and it is named a *simple mirror*.[14] In this magnetic configuration, the field is provided by two solenoids with coinciding axes, located at definite distances from other. Figure 14 features the shape of the field lines.

The principal constraints for charged particle confinement are summarized here:

1. a magnetic field gradient is needed $\nabla_{\parallel} B$, directed along the direction of the field itself;
2. even in simple cases, as for simple mirror configurations, the field may be axisymmetric but it must have a radial component; and
3. the velocity vector of the charged particle must have a particular angle with the field lines; a boundary exists for this angle, and for angles smaller than this, the particle is no longer confined.

In Fig. 14 only one of the two mirror elements is displayed, and the Larmor motions and the drift around the axis are shown. Along the mirror axis z, charged particles experience a force F_z, which is able to confine them provided that some specific conditions are satisfied.

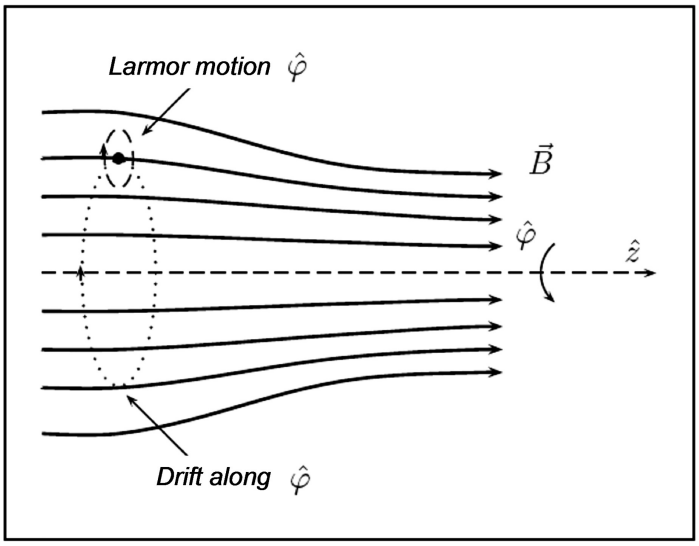


Fig. 14: Magnetic field lines produced in a simple mirror configuration. The density of the field lines increases from left to right, and this generates the gradient, which works as a mirror for charged particles.

The correct expression for F_z is to be averaged over a gyro-period. For the sake of simplicity we consider a particle moving along the mirror axis. Then v_φ is constant and depending on the sign of q it

[14]Fermi proposed the simple mirror configuration, with sources of the magnetic fields moving towards the centre of the system, as a possible mechanism involved in the generation of cosmic rays.

will be $v_\varphi = \mp v_\perp$. Furthermore $r = r_\mathrm{L}$ and thus

$$\bar{F}_z = \mp\frac{1}{2}qv_\perp r_\mathrm{L}\frac{\partial B}{\partial z} = \mp\frac{1}{2}q\frac{v_\perp^2}{\omega_\mathrm{c}}\frac{\partial B_z}{\partial z} = -\frac{1}{2}\frac{mv_\perp^2}{B}\frac{\partial B_z}{\partial z}. \tag{41}$$

By averaging F_z over a gyro-period, considering a particle moving along the mirror axis, we obtain [6]

$$\bar{F}_z = -\frac{mv_\perp^2}{B}B_1 z, \tag{42}$$

and thus the force is directed in the opposite direction with respect to the particle motion. We can also define the so-called *magnetic moment* of the charged particle as

$$\mu \equiv \frac{1}{2}\frac{mv_\perp^2}{B}. \tag{43}$$

Then Eq. (41) becomes

$$\bar{F}_z = -\mu\left(\frac{\partial B}{\partial z}\right). \tag{44}$$

Generalizing this for whatever gradient occurs along the particle motion, we obtain

$$\boldsymbol{F}_\parallel = -\mu\frac{\partial B}{\partial s} = -\mu\boldsymbol{\nabla}_\parallel B, \tag{45}$$

where $\mathrm{d}\boldsymbol{s}$ is a line element along the $\boldsymbol{B}$ direction.

The main property of μ is that it is an adiabatic invariant for the particle motion.[15]. The confinement of charged particles in mirror-like devices can be studied in terms of μ invariance. The breakdown of the adiabatic invariance of μ expels the particle from the confinement. The adiabatic invariance is valid as long as the B_z component slowly varies with z during a gyro-period. This condition can be written in the following manner:

$$\frac{r_\mathrm{L}}{L} \ll 1, \tag{46}$$

i.e. the Larmor radius r_L must be much smaller than L, the characteristic length for the B variation.[16]

The mirror effect in terms of the μ invariance can be explained as follows. If we consider a proof particle moving from the centre to the periphery of the magnetic bottle,[17] it sees a growing magnetic field. As μ must be kept constant, the perpendicular component of the velocity $v_\perp$ must increase because of the B increase. The magnetic field does not do any work on the system, and thus the kinetic energy remains constant. Hence when $v_\perp$ increases, $v_\parallel$ decreases. If B is sufficiently high at the device periphery, there will be a mirror point where $v_\parallel = 0$ and the particle is reflected.

6 Ion confinement in ECR plasmas

Magnetic confinement is effective only for particles that experience few collisions between the mirror points. This condition is often fulfilled for electrons but only rarely for ions. In ion sources, in fact, ions are cold, and their collisionality is a factor $m_\mathrm{i}/m_\mathrm{e}$ larger than for electrons. This makes them poorly magnetized, especially in the plasma core. The conservation of plasma quasi-neutrality thereby makes the ion confinement a direct consequence of the electron one. Presently, several models have been proposed to explain the ion confinement in magnetically confined plasmas, but only few experimental results are available to evaluate the best model. The most commonly accepted model assumes that well-confined electrons in the plasma core provide a sufficiently strong negative potential dip that incorporates

[15] To read more about the adiabatic invariants in plasmas, see Ref. [6]

[16] This property of the μ invariance is of primary importance also for ECRIS physics. If such a condition is not satisfied, then non-adiabatic effects begin to play a role, leading to the expulsion of the particle from the confinement.

[17] The simple mirror systems are often named 'magnetic bottles'.

multicharged ions. This theory applies only to the plasma core, while for the whole volume a more general approach must be taken into account. The quasi-neutrality must be preserved everywhere, and an ambipolar diffusion process arises. The slower and the faster species can be determined by analysing the collisions times when collisions are responsible for particle deconfinement. The slower species determine the sign of the plasma potential. Tonks and Langmuir formerly described the properties of plasmas located inside absorbing boundaries.[18] The preservation of the quasi-neutrality means that $n_{\mathrm{e}} \simeq Zn_{\mathrm{i}}$, with Z the mean ion charge state, and $j_{\mathrm{e}} \simeq j_{\mathrm{i}}$. Thus the electron and the ion densities must remain the same, along with the currents of particles leaving the plasma. Starting from this assumption in ECRIS plasma halo, the electrons,[19] having $\tau_{\mathrm{ei}}/\tau_{\mathrm{ii}} < 1$, leave the plasma more easily than ions, thus creating a positive potential at the plasma boundary, which accelerates ions and tends to retain the escaping electrons. On the contrary, in the plasma core, the electrons are less collisional, because their temperature is very high. Then they are well confined and generate a hot electron cloud, which retains the ions. The negative potential dip can be indicated as $\Delta\phi$:

$$\Delta\phi \propto -\frac{T_{\mathrm{i}}}{ze}. \tag{47}$$

Assuming that the bouncing time of ions inside the plasma trap is τ_{b} we obtain

$$\tau_{\mathrm{b}} \simeq \frac{B_{\max}}{B_{\min}} \left[\frac{\text{plasma length}}{\langle \text{ion velocity} \rangle}\right], \tag{48}$$

$$\tau_i \simeq \exp\left(\frac{Z\Delta\phi}{KT_{\mathrm{i}}}\right), \tag{49}$$

with $Z = ze$. Note that the plasma potential is globally positive, but the 'hollow shape', with the small dip in the plasma core, confines the ions.

According to this model, the electrons with high energies ($T_{\mathrm{e}} > 10\,\mathrm{keV}$) provide a stabilizing effect for the plasma because they practically confine the ions. In order to increase the ion lifetime, we have to increase the depth of the potential.

This model for ion confinement has been recently put in doubt by some results and needs further developments.

6.1 Onset of MHD instabilities and stabilized min-B traps

The simple mirror configuration allows confinement of those particles that obey the adiabatic invariance of the magnetic moment. Gibson, Jordan and Lauer proved experimentally that singly charged particles remain trapped in a magnetic mirror in a manner completely consistent with the prediction of the adiabatic theory. Thus this theory predicted (and the experiment confirmed) that single particles are satisfactorily trapped in a magnetic mirror field. However, the trapping of a single particle in some field configuration does not mean that a plasma has a stable equilibrium. Not only losses induced by collisions (which can change the velocity vector direction, putting it inside the loss cone) but also other instabilities may arise that increase particle losses. Because of the magnetic confinement, the simple mirror plasmas have a non-isotropic energy distribution function. This leads to some instabilities in the velocity space, the so-called *loss cone instability*. However, some hydrodynamic instabilities may be more dangerous. The most important instability involving simple mirror plasmas is the so-called *flute instability*. This instability for magnetically confined plasmas is equivalent to the well-known Rayleigh–Taylor instability, which arises in neutral fluids and also in plasmas subjected to the gravitational force. In the case of magnetically confined plasmas, the curvature of magnetic field lines triggers the instability. In Fig. 15 a schematic representation of the mechanism at the basis of the flute instability is reported.

[18] By absorbing boundaries we usually mean the walls of a metallic plasma chamber.

[19] Because of their higher mobility, due to the low plasma halo temperature and its high density, the electrons are more mobile than ions in the plasma periphery, and so they tend to leave the plasma at a higher rate than ions.

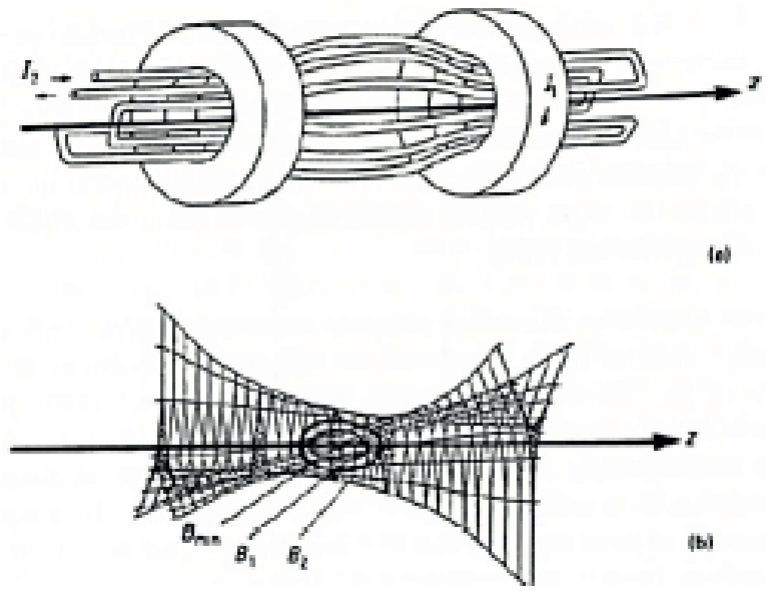


Fig. 15: (a) Magnetic system and (b) magnetic field originated by two solenoids surrounding a hexapole

If the magnetic field lines curve into the plasma everywhere, then the confinement will be stable. This result gives us some practical information to realize optimal plasma traps where plasmas are stably magnetically confined.

In 1962 Ioffe [7] first reported on plasma-confinement experiments in a magnetic field configuration that had the desirable feature that the magnetic field increased in every direction away from the plasma boundary, and that did not have the undesirable feature of a region where the magnetic field went to zero inside the plasma (as in the case of cusp configurations), producing an open wide loss cone. Such a configuration can be obtained as a superposition of two magnetic field structures: one created by two solenoids (simple mirror), and the other created by six conductors surrounding the plasma chamber (a hexapole). In this way the magnetic field increases in every direction that can be seen from the plasma centre. In fact, the field produced by the hexapole increases along the plasma radius. This provides a stabilizing effect also radially, the simple mirror already being able to confine the plasma axially. This configuration can also be thought of as a multi-mirror device: the particle is reflected at many points travelling along the field lines. Ioffe and his co-workers at the Kurchatov Institute in Moscow used a device known as the PR-5 (Probkotron-5) to demonstrate that the hydromagnetic instability could be suppressed with a min-B configuration. Figure 15 schematically shows the field lines of such a field.

Note that the geometrical loci where B is constant are ellipsoids that are usually defined as *egg-shaped surfaces*. As there are no axial symmetries, the μ adiabatic invariance cannot be invoked to study particle confinement. Another adiabatic invariant can be used for min-B configurations, i.e.

$$J = \oint v_{\parallel}\, \mathrm{d}l, \tag{50}$$

in which J essentially determines the length of field lines between two reflection points. The J invariance implies that, even after reflection, the particle will move along the same field line, or at least along a field line with the same J value. Field lines with the same J value define a surface over which particles with fixed values of W/μ will move.

The system is realized so that the constant-J surfaces do not intercept the plasma chamber walls. Up to now we have pointed out the ability of a min-B configuration in ensuring the MHD stability, but no

numbers have been given for the relative values of the confining magnetic field, i.e. for the mirror ratio at the extraction and injection sides, and also in the radial direction.

Gammino and Ciavola proposed, in 1990, the so-called *high-B mode* (HBM) concept [8], stating that 'only an appropriate optimization, in terms of mirror ratios, of the confining magnetic field, leads to the exploitation of the electron density as expected from the frequency scaling, thus leading to an effective increase of the sources performances'. Such a principle does not conflict with the Geller laws, but it limits their applications to well-confined plasmas. The main aspect of the HBM concerns the possibility to give some numbers for the configuration of the magnetic field. These results have been obtained by virtue of MHD considerations. Note that a stable MHD equilibrium can be reached only if the magnetic pressure is higher than the plasma pressure.[20] Then we estimate the plasma pressure in ECRIS, we compare it with the magnetic pressure (which depends on the maximum field that can be produced by the magnet), and finally we can state whether or not a given magnetic configuration is stable according to the MHD criteria for stability.[21]

The expression for the β parameter is

$$\beta = \frac{n_{\mathrm{e}} K T_{\mathrm{e}}}{B^2/2\mu_0}. \tag{51}$$

Well-confined plasmas are characterized by β parameters in the range

$$0.005 < \beta < 0.01. \tag{52}$$

The plasma electron density can be roughly estimated as

$$n_{\mathrm{e}} \simeq n_{\mathrm{cutoff}} = \frac{m_{\mathrm{e}}}{e^2}\epsilon_0 \omega_{\mathrm{RF}}^2 = \frac{\epsilon_0}{m_{\mathrm{e}}} B_{\mathrm{ECR}}^2, \tag{53}$$

with

$$\omega_{\mathrm{RF}} = \frac{q B_{\mathrm{ECR}}}{m}. \tag{54}$$

Hence, by substituting Eq. (53) into Eq. (51), and considering Eq. (52), we find

$$\frac{\epsilon_0}{m_{\mathrm{e}}} B_{\mathrm{ECR}}^2 k T_{\mathrm{e}} \quad < \quad 0.01 \frac{B^2}{2\mu_0}, \tag{55}$$

$$\left(\frac{B}{B_{\mathrm{ECR}}}\right)^2 \quad > \quad 100 k T_{\mathrm{e}} 2 \frac{\mu_0 \epsilon_0}{m_{\mathrm{e}}}. \tag{56}$$

We also have

$$\mu_0 \epsilon_0 = \frac{1}{c^2} \quad \text{and} \quad m_{\mathrm{e}} c^2 = 511 \text{ keV}, \tag{57}$$

so that we then obtain

$$\left(\frac{B}{B_{\mathrm{ECR}}}\right)^2 > \frac{200 k T_{\mathrm{e}}}{511 \text{ keV}}. \tag{58}$$

The above equation is of primary importance for the operations of modern ECRIS, as it gives some useful information on practical preparation of an experiment. For instance, if $kT_{\mathrm{e}} = 10$ keV, then

$$\frac{B}{B_{\mathrm{ECR}}} > 2. \tag{59}$$

These results have been confirmed by a great amount of experimental data. One example is shown in Fig. 16, showing the trend of the extracted current from the SERSE source of the INFN-LNS for different values of the radial mirror ratio B/B_{ECR}. Note that the extracted current rapidly increases, then tends to saturate when $B/B_{\mathrm{ECR}} > 2$, confirming that the electron density reaches its maximum value only

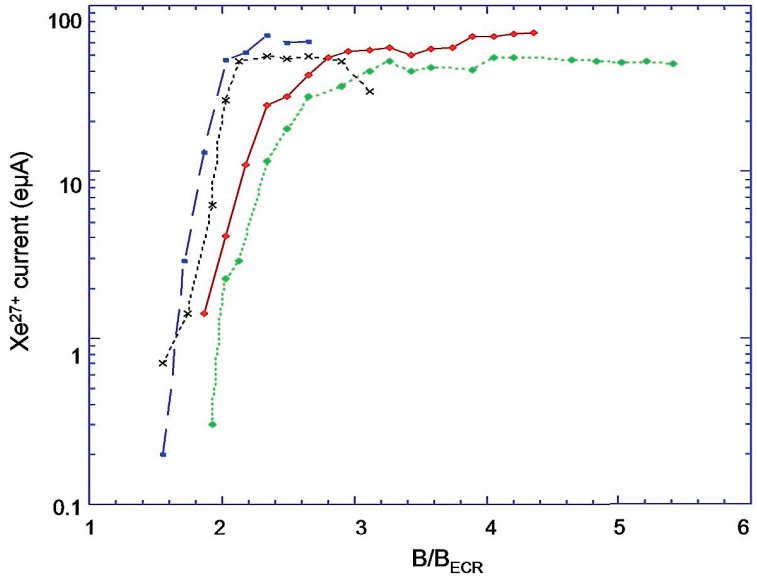


Fig. 16: Experimental confirmation of the HBM concept. The extracted current increases as the ratio B/B_{ECR} increases, but it saturates for $B/B_{\mathrm{ECR}} = 2$, in good agreement with Eq. (58) in the case of $T_{\mathrm{e}} \simeq 10$ keV.

when the plasma stability condition is established. More precise scaling laws can be determined by taking into account all the typical values of the magnetic field in the plasma trap: the injection, the extraction and the radial field, all of them compared with the field at ECR and with that at the bottom of the configuration (usually named B_{min} or B_0). Finally, note that the relation (58) strictly links the magnetic configuration to the used frequency (B_{ECR} is fixed by the microwave frequency). Hence, to effectively improve the ECRIS performance, one has to increase both the microwave frequency and the magnetic field, in particular the mirror ratios, of the plasma trap.

7 Simplified treatment of wave propagation in a magnetized plasma

In strongly magnetized plasmas the direction of the external magnetic field is a cause of anisotropy in the propagation of the electromagnetic waves. The electron motion is in fact not free in any direction, and waves are sustained in different manners according to their propagation direction. Skipping all the theoretical details, let us here introduce the propagation constant also in the case of a generic angle of propagation:

$$k'_\theta = \frac{\omega}{c}\left[1 - \frac{X(1+\mathrm{i}Z-X)}{(1+\mathrm{i}Z)(1+\mathrm{i}Z-X) - \frac{1}{2}Y_{\mathrm{T}}^2 + \sqrt{\frac{1}{4}Y_{\mathrm{T}}^4 + Y_{\mathrm{L}}^2(1+\mathrm{i}Z-X)^2}}\right]^{1/2} \tag{60}$$

and

$$k''_\theta = \frac{\omega}{c}\left[1 - \frac{X(1+\mathrm{i}Z-X)}{(1+\mathrm{i}Z)(1+\mathrm{i}Z-X) - \frac{1}{2}Y_{\mathrm{T}}^2 - \sqrt{\frac{1}{4}Y_{\mathrm{T}}^4 + Y_{\mathrm{L}}^2(1+\mathrm{i}Z-X)^2}}\right]^{1/2}. \tag{61}$$

Hence we have found two waves travelling along an arbitrary direction defined by θ with k'_θ and k''_θ. As a function of X, k'_θ is more similar to a wave propagating in an isotropic plasma, and thus the

[20] By plasma pressure we mean the usual kinetic pressure due to the particle temperature and density.

[21] Magnetically confined plasmas can be considered MHD-stable only if the β parameter is appreciably smaller than one.

associated wave will be called the *ordinary wave*; the wave with k''_θ is named the *extraordinary wave*. However, in plasma physics another classification prevails, according to the scheme shown in Figs. 17 and 18. Usually, the waves propagating along the magnetic field direction are called the R and L *waves*, according to their polarization.[22] The waves propagating along the perpendicular direction to $\boldsymbol{B}_0$ are named O and X, according to the orientation of the wave electric field with respect to the magnetostatic field direction. If $\boldsymbol{k} \perp \boldsymbol{B}_0$ and $\boldsymbol{E} \parallel \boldsymbol{B}_0$, then we have the O mode. Otherwise, if $\boldsymbol{k} \perp \boldsymbol{B}_0$ and $\boldsymbol{E} \perp \boldsymbol{B}_0$, whatever the direction of $\boldsymbol{E}$, the wave will be called the X mode.

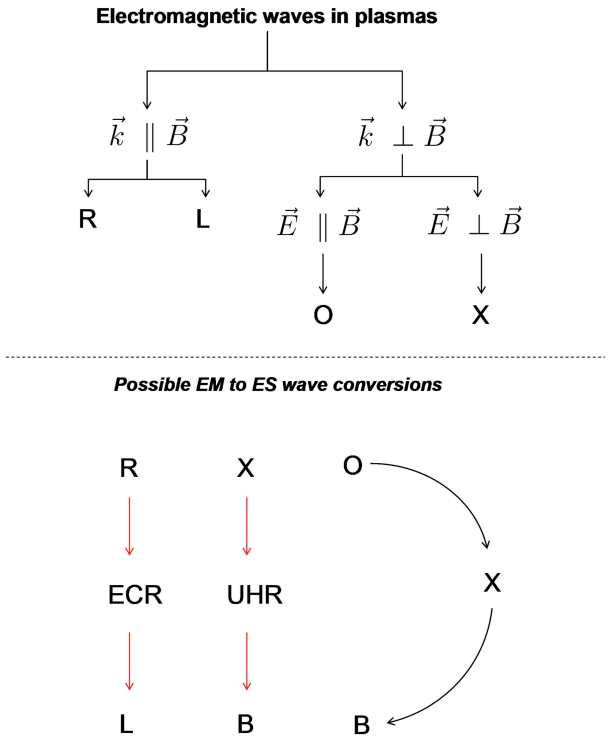


Fig. 17: Scheme of the wave propagation in anisotropic plasmas

Many other waves can be generated in plasmas, many of them involving ions, collective fluid motions, and so on. The various ways in which electromagnetic waves can be injected into a column of magnetized plasma are shown in Fig. 19. By applying the kinetic theory, one can also determine the wave propagation in a warm plasma.

We next consider the cutoffs and resonances of the R, L, O and X waves. For the R waves, k becomes infinite at $\omega = \omega_\mathrm{c}$. In this case the microwave frequency is equal to the Larmor frequency, hence the wave is in resonance with the cyclotron motion of the electrons. This resonance is called *electron cyclotron resonance* and is of primary importance for ECRIS. More generally, the cutoffs and the resonances of the various propagation modes can be summarized by reporting the different c^2k^2/ω^2 formulas obtained from Eqs. (60) and (61). A schematic summary of the main electron waves[23] is given in Table 2.

In Table 2, $\omega_\mathrm{h} = \sqrt{\omega_\mathrm{c}^2 + \omega_\mathrm{p}^2}$ is the so-called *upper hybrid frequency*. It corresponds practically to the frequency of the Langmuir oscillations in non-magnetized plasmas. Electrostatic electron waves

[22] As may be argued by the notation, the R wave is a right-hand polarized wave, and L is a left-hand polarized wave.

[23] The term 'electron waves' usually indicates the high-frequency waves, i.e. the electromagnetic waves. For these waves, the ions can be considered as an immobile medium, as the wave frequency is too high to have some ions responding to the electromagnetic field. It is usually assumed that ions ensure the quasi-neutrality, but they do not take part in the electromagnetic wave propagation.

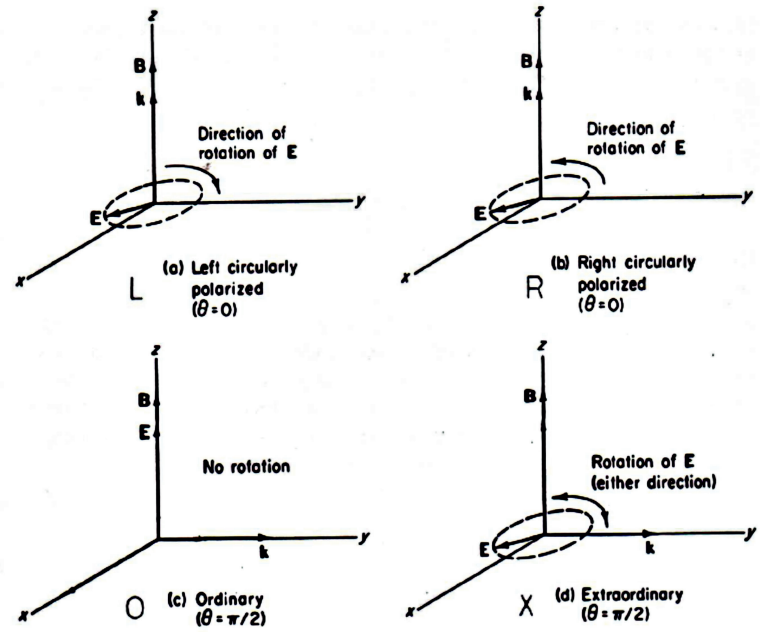


Fig. 18: Diagram showing the possible orientations of the electric field with respect to the magnetostatic field, and also the possible polarizations in waves propagating in magnetized plasmas.

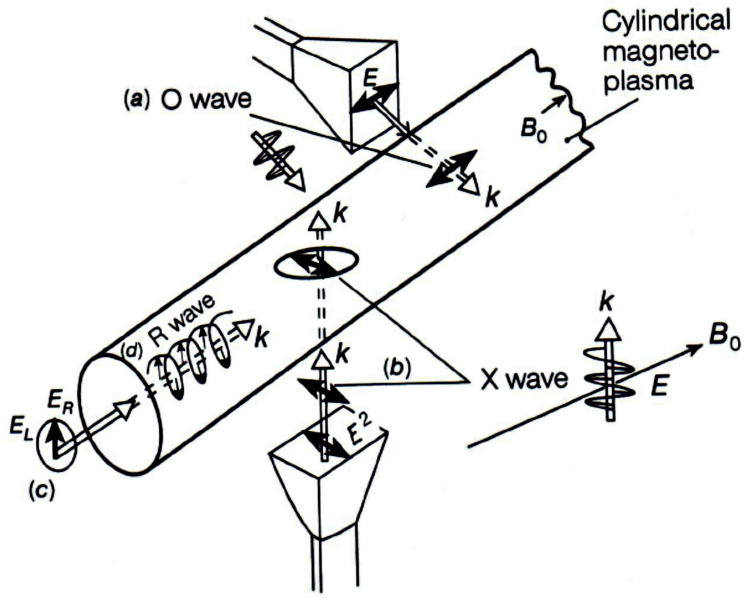


Fig. 19: Schematic representation of the possible ways to inject electromagnetic waves into a column of a magnetized plasma. The different modes can be excited, according to the location of the waveguide providing the electromagnetic field, and also according to the wave polarization inside the waveguide.

Table 2: Cutoffs and resonances of the waves propagating inside the plasma [6].

Vector orientation	Refractive index	Wave type
$\boldsymbol{B}_0 = 0$	$\omega^2 = \omega_\mathrm{p}^2 + k^2c^2$	Light waves
$\boldsymbol{k} \perp \boldsymbol{B}_0,\ \boldsymbol{E} \parallel \boldsymbol{B}_0$	$\dfrac{c^2k^2}{\omega^2} = 1 - \dfrac{\omega_\mathrm{p}^2}{\omega^2}$	O wave
$\boldsymbol{k} \perp \boldsymbol{B}_0,\ \boldsymbol{E} \perp \boldsymbol{B}_0$	$\dfrac{c^2k^2}{\omega^2} = 1 - \dfrac{\omega_\mathrm{p}^2}{\omega^2}\dfrac{\omega^2 - \omega_\mathrm{p}^2}{\omega^2 - \omega_\mathrm{h}^2}$	X wave
$\boldsymbol{k} \parallel \boldsymbol{B}_0$	$\dfrac{c^2k^2}{\omega^2} = 1 - \dfrac{\omega_\mathrm{p}^2/\omega^2}{1 - (\omega_\mathrm{c}/\omega)}$	R wave
$\boldsymbol{k} \parallel \boldsymbol{B}_0$	$\dfrac{c^2k^2}{\omega^2} = 1 - \dfrac{\omega_\mathrm{p}^2/\omega^2}{1 + (\omega_\mathrm{c}/\omega)}$	L wave

across $\boldsymbol{B}$ have this frequency, while those along $\boldsymbol{B}$ have the usual plasma oscillations with $\omega = \omega_\mathrm{p}$. Furthermore, from Table 2 it can be seen that the R wave suffers the resonance reported before (i.e. the ECR resonance), as its index of refraction goes to infinity when $\omega \to \omega_\mathrm{c}$. Also the X mode has a resonance, but at the upper hybrid frequency. Thus, inside a magnetized plasma, where an X mode is travelling, the so-called *upper hybrid resonance* (UHR) occurs if

$$\omega^2 = \omega_\mathrm{p}^2 + \omega_\mathrm{c}^2 = \omega_\mathrm{h}^2. \tag{62}$$

The L wave does not suffer any resonance, as well as the O mode. Instead, they suffer a cutoff that can be determined by the equations reported in Table 2 when the index of refraction, i.e. c^2k^2/ω^2, goes to zero. A powerful method to visualize the cutoffs and the resonances of the various modes is to plot them on a so-called Clemmow–Mullaly–Allis (CMA) diagram.[24] This diagram is reported in Fig. 20.

When the propagation occurs at a given angle with respect to the magnetic field, the phase velocities change with the angle. Some of the modes listed above with $\boldsymbol{k} \parallel \boldsymbol{B}_0$ or $\boldsymbol{k} \perp \boldsymbol{B}_0$ change continually into each other. Such a complicated picture is greatly simplified by the CMA diagram, which however works only in the case of the cold plasma approximation ($T_\mathrm{e} = T_\mathrm{i} = 0$). Any finite-temperature modification implies such a great number of complications that the diagram becomes completely useless.

The CMA diagram can be viewed as a plot of ω_c/ω versus $\omega_\mathrm{p}^2/\omega^2$, or equivalently a plot of the magnetic field versus the plasma density. For a given frequency ω, any experimental situation characterized by ω_c (i.e. the magnetic field) and by ω_p (i.e. the plasma density) is denoted by a point on the graph. The total plane is divided into several zones, and the boundaries of each zone are the cutoffs and the resonances mentioned above. For example, the upper hybrid resonance can be easily found in the graph. Considering an X wave propagating from a region with high magnetic field, inside a plasma with a fixed value of density (e.g. $\omega_\mathrm{p}^2/\omega^2 = 0.7$), the value of the magnetic field where the UHR occurs can be easily determined on the graph.

The vertical line at $\omega_\mathrm{p}^2/\omega^2 = 1$ is the so-called O cutoff, and it corresponds to the cutoff density in an unmagnetized plasma. Above this value, the plasma will be called *overdense*, as shown in Fig. 21. The small diagrams in each region delimited by the cutoffs and the resonances (i.e. the ellipsoidal small diagrams) indicate not only which wave propagates, but also the qualitative variation of the phase velocity with the angle (considering that the magnetic field is directed along the vertical direction).

[24] The CMA diagram takes its name from the three authors who proposed the graphical representation of the many cutoffs and resonances of electromagnetic waves propagating in a cold plasma.

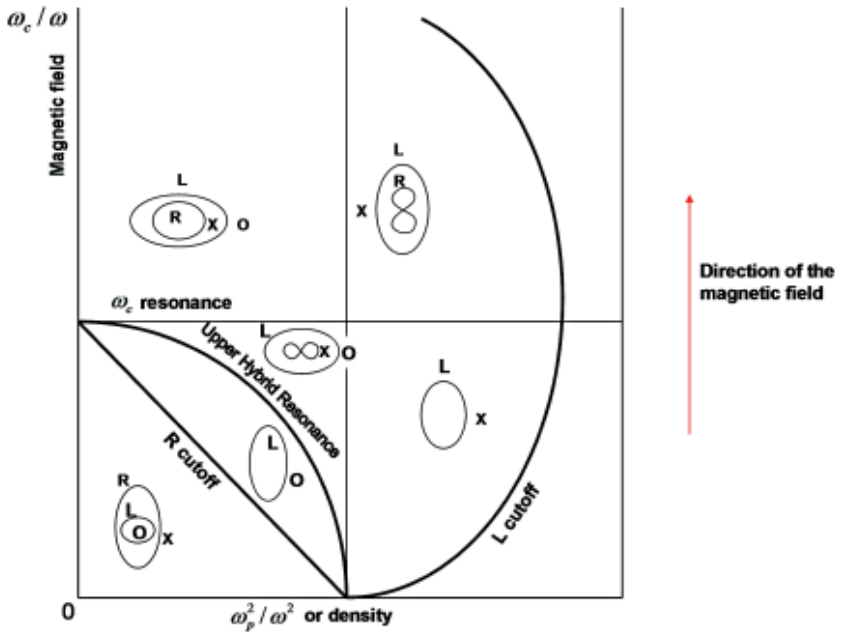


Fig. 20: The Clemmow–Mullaly–Allis diagram for the classification of waves in a cold plasma

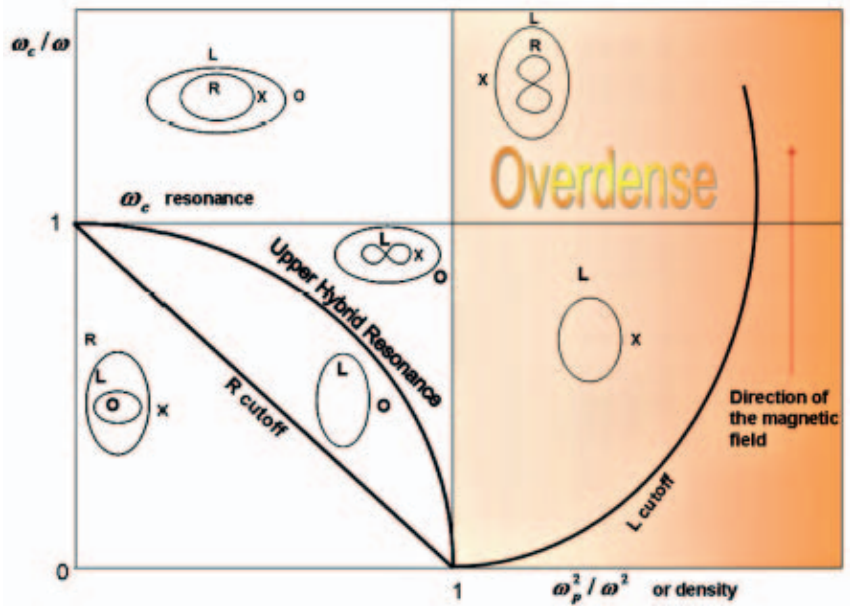


Fig. 21: A detailed view of the Clemmow–Mullaly–Allis diagram showing the region of the so-called overdense plasma, i.e. that region where the plasma density exceeds the cutoff of the ordinary (O) mode.

As mentioned above, the kinetic theory provides more information about the wave propagation than the macroscopic one-fluid approach and the single-particle investigation. The main assumption of both the one-fluid and single-particle approaches is that no velocity distributions of plasma particles are taken into account. This limits the applicability of such approaches to low-temperature plasmas only, or to those plasmas for which the particle velocities are near to the fluid element velocity. Landau damping, ion waves and all those waves at harmonics of the cyclotron frequency are outside of the fluid or single-particle schemes, and they all depend on the plasma temperature. For example, in the case of the O mode, whose dispersion relation is reported above and was based on a single-particle approach, if we apply the kinetic theory we find that other possible propagation modes exist in proximity of the cyclotron harmonics $n\omega_{\mathrm{c}}$.

8 Plasma heating by resonant stochastic wave–electron interaction

In ECRIS an effective heating is experimentally observed; on the other hand, we know that the electron lifetimes are of the order of milliseconds for well-confined plasmas. Thus, non-adiabatic effects, which provide the stochasticity, arise in times shorter than milliseconds. Reasonable estimations and numerical simulations give a time of the order of a few nanoseconds to observe non-adiabatic effects in such plasmas. Thus, collisionless plasma heating is completely determined by the stochasticity, as demonstrated also by Liebermann and Lichtenberg [9].

According to their theory, the stochasticity in ECR plasmas is provided by multiple passages through the resonance, as long as a 'forgetting' mechanism works well. In order to better understand how the heating develops in multi-passage interactions, it is better to introduce the *separatrix* concept. Particles that resonantly interact with waves can be considered as trapped by the wave potential. The resonant condition is achieved if the well-known equation $\omega = kv_{\mathrm{e}}$ is satisfied. Again, consider a frame moving with the wave velocity. Particles moving with a similar velocity can oscillate in the wave potential[25] $\varphi(z)$. Then the electron moves with a velocity v_{e}, while the wave potential, which can be indicated by $\varphi(z)$,

$$\varphi = \varphi_0 \sin(\omega t - kz), \tag{63}$$

is stationary and sinusoidal, as the frame moves with v_φ. Because of energy conservation, we can write

$$\tfrac{1}{2}mv_{\mathrm{e}}^2 + e\varphi(z) = \epsilon = \text{constant}. \tag{64}$$

Thus the velocity v of the electron is

$$v_{\mathrm{e}} = \pm\sqrt{\frac{2\epsilon - 2e\varphi(z)}{m}}. \tag{65}$$

This velocity can be plotted in the phase space z–v_{e}.

Particles at the bottom of the wave potential (i.e. perfectly resonant with the wave) stay exactly at the centre of each separatrix, and their trajectory in the phase space is just a point (their velocity with respect to the wave is zero, so they must lie on the z axis). If an electron is moved from the bottom of the potential, it will begin to oscillate around the equilibrium point, similarly to mechanical oscillators with small-amplitude oscillations around the equilibrium. As long as the perturbation is small enough, the motion around the minimum will be harmonic. In the phase space, these motions are represented by ellipsoidal trajectories, as shown in Fig. 22. The electron velocity may become higher and higher, and then the ellipsoidal trajectories become bigger and bigger. In fact, if the electron energy becomes appreciable with respect to $\varphi(z)$, the motion is no longer harmonic and nonlinear effects arise. If the electron stays on the top of the $\varphi(z)$ potential, then the electron velocity will be zero according to Eq. (65). In this case the electron has the same velocity as the wave, but, because it stays on top of

[25] As before, we consider a one-dimensional model to explain the particle trapping in the wave field at the resonance.

the wave potential, its equilibrium is not stable, and chaotic motions may arise. When this condition is satisfied, the phase space separates into different zones. Inside each zone the particle is trapped by the wave. Figure 22 features the separatrix of a sinusoidal wave interacting with a single particle. This theory, named the *pendulum model*, explains easily how the nonlinearity and the stochastic effects can play a fundamental role in the collisionless wave–particle interaction. Those electrons whose velocity differs considerably from the wave phase velocity are not trapped and their velocity is slightly perturbed by the wave. If the separatrix areas present some non-uniformities, for example, because of localized wave patterns, the heating becomes more effective, as such systems have other sources of stochasticity.

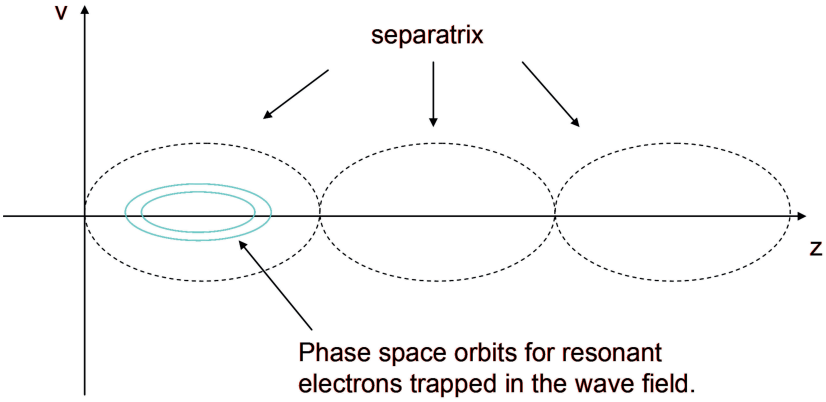


Fig. 22: Spread in the phase space of the electron velocity due to the activation of the stochastic effects

Let us discuss the heating in ECRIS through electron multi-passages across the ECR zones. An easier treatment is possible if we consider a simple mirror configuration, because in that case a one-dimensional model can be applied and the bouncing of the electrons inside the trap is subjected only to the potential μB. Then the equation for the z variables becomes

$$z = z_0 \cos(\omega_\mathrm{b} t + \psi_0), \tag{66}$$

where ω_b is the bouncing frequency. The perpendicular component of the wave electric field can be simply written as

$$E_\perp(t, z, r, \theta) = E_{\perp 0} \cos(\omega t - kz + \theta_\perp), \tag{67}$$

where $\theta_\perp$ is the initial perpendicular phase. By substituting Eq. (66) into Eq. (67), we obtain

$$\begin{aligned} E_\perp(t, z, r, \theta) &= E_{\perp 0} \cos[\omega t + \theta_\perp - k z_0 \cos(\omega_\mathrm{b} t + \psi_0)] \\ &= \sum_n A \exp\left[(\omega - \omega_\mathrm{c}) t + \theta_\perp + \theta_b\right], \end{aligned} \tag{68}$$

where

$$\theta_\mathrm{b} = \omega_\mathrm{b} t + \psi_0. \tag{69}$$

Equation (68) takes into account the cyclotron frequency, the frequency of the wave and the bouncing frequency. The Fourier expansion shows that the composition of the bounce and cyclotron frequencies leads to an effective multiwave interaction, i.e. we may imagine a single particle interacting with a pair of waves at the resonance. Each wave–particle interaction has a proper separatrix and many zones like those shown in Fig. 22 are present in the phase space.[26] Figure 23 shows the structure of several separatrices in the phase space in the proximity of the ECR zone.

[26]Here we consider the perpendicular phase space, i.e. that one formed by $p_\perp$–$\theta_\perp$.

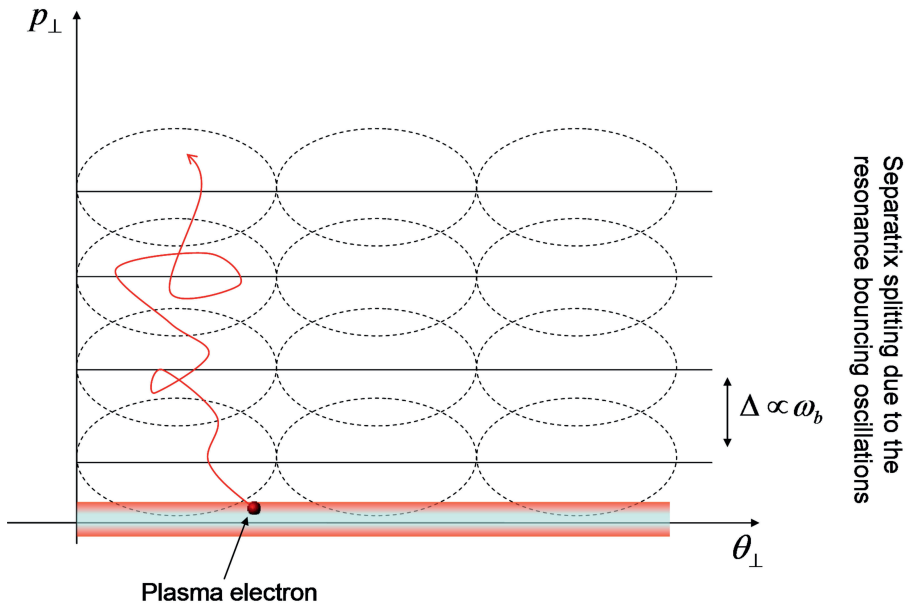


Fig. 23: The many separatrices occurring in the proximity of the ECR zones: they are overlapped

It is clear that all the separatrices are overlapped, the distance between them being proportional to ω_b. The lowest separatrix intercepts the plasma slab; this means that electrons of the plasma can be trapped into the lowest separatrix. According to its own kinetic energy, to the phase with respect to the wave, and so on, the motion will evolve into the several separatrices in a complicated manner, taking into account also the nonlinearity in the wave–particle interactions. As the separatrices are overlapped, the electron perpendicular momentum (thus its energy) will increase more and more on passing from one separatrix to the subsequent one, as schematically shown in Fig. 23. In the real situation the horizontal separatrix extension at the resonance is limited, and thus the electron will escape after a fixed time from the interaction zone. But because of its bouncing motion, it will pass again through the resonance. The former interaction was regulated by a stochastic motion in the phase space, so it will arrive again at the ECR with a completely random phase, thus being able to gain more energy. This process goes on for many passages through the resonance. However, the higher the electron energy, the higher is ω_b. When the bouncing frequency becomes comparable to the Larmor frequency, the separatrix splits and the electron cannot pass through them. This effect corresponds to the activation of some adiabatic invariants: the electron heating stops because the phase randomization process, strictly connected to the separatrix overlapping, no longer works.

A more quantitative treatment of the heating process starts from the Canobbio parameters. Liebermann and Lichtenberg formulated the theory to calculate the maximum value of the electron energy for particles bouncing between the mirror, then passing many times through the resonance. Let t_e be the time effectively spent by the electron in the ECR region:

$$t_\mathrm{e} = \frac{0.71}{\omega}\left(\frac{2\omega}{\delta v_{\perp i}}\right)^{2/3}, \tag{70}$$

where the perpendicular velocity $v_\perp$ is given by

$$v_\perp \propto \frac{e}{m} E t_\mathrm{e} \tag{71}$$

that is, in terms of δ,

$$v_\perp \propto \frac{E}{\delta^{2/3}} \tag{72}$$

and also

$$v_\perp \propto \frac{\sqrt{P}}{\delta^{2/3}}. \tag{73}$$

By virtue of Eq. (70), we have that t_e does not depend on $v_{\|R}$, and it is inversely proportional to $v_{\perp R}$, i.e. the higher the transverse velocity, the less effective will be the particle acceleration.

For multiple passages through the ECR region, the $v_\perp$ evolves according to

$$v_\perp^2 = v_{\perp 0}^2 + v^2 + 2(v v_{\perp 0}^2 \cos\theta), \tag{74}$$

where v is the maximum velocity gained for each passage, θ is a function of the phase and $v v_{\perp 0} \cos\theta$ is the stochastic term. After some passages, the mechanism explained above begins to play a role, the separatrix splits and the heating stops. This mechanism can be explained also in a different manner. As the velocity increases, some adiabatic invariants arise. Such invariants depend also on the wave electric field ($\boldsymbol{E}$). Defining by W_S the maximum value of the energy that can obtained through stochastic heating, we have

$$W_\mathrm{S} = 0.34 eEL\sqrt{\frac{\bar{t}}{\tau_\mathrm{S}}}, \tag{75}$$

with

$$\tau_\mathrm{S} \sim 4\pi \frac{m}{e}\frac{1}{B}\frac{L^2}{l^2}, \tag{76}$$

which is the time necessary to have a phase shift of 2π, L is the characteristic length of the mirror field, as can be seen by the $\boldsymbol{B}$ equation

$$B_z = B_0\left(1 + \frac{x^2}{L^2}\right), \tag{77}$$

and

$$\bar{t} = \sqrt{\left(\frac{mL}{eE}\right)}\left(1 + \frac{l^2}{L^2}\right)^{5/4}, \tag{78}$$

corresponding to the resonance position $x = l$. In these conditions the electron is mirrored just inside the ECR zone [9, 10].

The maximum energy achievable by the electrons will be

$$W_\mathrm{B} \sim 5 W_\mathrm{S}, \tag{79}$$

and W_B is usually called the *absolute stochastic barrier*. To give an estimate of such a barrier, we can substitute some realistic values into the equations (75)–(79), thus obtaining

$$W_\mathrm{S} \sim 2E^{3/4}\ (\mathrm{V\ m^{-1}}) \propto P_\mathrm{RF}^{3/8}, \tag{80}$$

with E and P_{RF} the wave's electric field and power. As for ECRIS we have

$$0.1 \leq E \leq 10\ (\mathrm{kV\ cm^{-1}}). \tag{81}$$

The absolute stochastic barrier will lie in the range

$$10\,\mathrm{keV} \leq W_\mathrm{B} \leq 270\,\mathrm{keV}. \tag{82}$$

The estimation of the maximum energy achievable by ECRIS plasma electrons is realistic if compared with experiments. This is a remarkable result, as the single-particle theory may be considered somewhat oversimplified. However, such calculations fail for modern ECRIS, like VENUS. Other possible heating mechanisms may be investigated thanks to numerical simulations in order to understand if the

adiabatic invariants, activated at high electron energy, can be broken by additional effects not considered in the Liebermann and Lichtenberg theory.

Strictly speaking, it can be noted, by the separatrix treatment, that the heating may continue if the separatrix splitting does not occur, i.e. if additional perturbations to the particle motion change the separatrix amplitudes or provide an additional separatrix, in order to have a separatrix overlap again. In this way the phase randomization should continue even above the stochastic barrier.

The so-called 'ECRIS standard model', which has traced the road for the development of electron cyclotron resonance ion sources in the past 20 years, is based on experimental evidence and mixes Geller's scaling laws and the high-B mode concept. This model has permitted an average increase of about one order of magnitude per decade in the performance of ECR ion sources since the time of the pioneering experiment of R. Geller at CEA-Grenoble [11]. However, nowadays such a trend is limited not only by the technological limits of the magnets, but also by the experimental evidence of hot electron population formation, which becomes more important for high plasma density (i.e. for an operational frequency of 28 GHz and higher) and for microwave power above 3 kW.

The simplest picture to determine a relationship between the microwave frequency and the maximum achievable plasma density considers the *O-mode cutoff frequency*, so that, writing it in terms of maximum density as a function of the field frequency, we obtain

$$n_{\mathrm{cutoff}} = \epsilon_0 m_{\mathrm{e}} \frac{\omega^2}{e^2}. \tag{83}$$

More generally, all the possible electromagnetic modes propagating in plasmas ($R,\ L,\ O,\ X$) suffer a cutoff. Then the maximum density of the plasma where these modes propagate cannot exceed some fixed values. Only electrostatic modes are able to propagate in plasmas of whatever density (e.g. Bernstein modes).

The main consequence of Eq. (83) is that, by increasing the electromagnetic field, the electron density can be increased. In magnetized plasmas the relationship between density and frequency may change, but the dependence of the extracted current on the square of the microwave frequency has been demonstrated in many experiments. On the basis of these considerations and of the experimental results, 25 years ago Geller proposed the following scaling laws [12], which for a long time have been the guideline for the ECRIS community:

$$q_{\mathrm{opt}} \propto \log B^{3/2}, \tag{84}$$

$$I^{q+} \propto f^2 M_{\mathrm{i}}^{-1}, \tag{85}$$

where q_{opt} is the optimum charge state, B is the peak field of the magnetic trap, f is the microwave frequency, I^{q+} is the intensity of the charge state q and M_{i} is the mass of ions.

These formulas have been verified in several ECRIS spread over the world. However, although the microwave frequency was continuously increased, in order to fulfil Eq. (84), the increasing trend of the source performance began to saturate, especially when it was clear that the MINIMAFIOS source [13] produced at 16.6 GHz a mean charge state lower than the SC-ECRIS operating at 6.4 GHz at MSU.[27] It was clear that the confinement, which for the SC-ECRIS was much better than for MINIMAFIOS, was the other key parameter to ensure the exploitation of the benefits provided by the increase of the microwave frequency increase.

9 Advanced research for ECRIS development

The best compromise in terms of intense currents and charge states is obtained by ECRIS. Up to now the 'ECRIS standard model' has traced the road for the development of ECRIS. Since its formulation

[27] The MINIMAFIOS source was able to extract a CSD with $\langle q \rangle = 9$, whereas the SC-ECRIS produced $\langle q \rangle = 12$.

(completed in 1990, when the HBM concept was proposed), this model has been proven by many experiments carried out with different sources, operating at different ECR frequencies. The main rules were confirmed by experiments performed at MSU-NSCL in 1993–94 and in 1995. At the powers, magnetic fields and microwave frequencies used up to now, the ECRIS obey the standard model: the extracted current increases strongly as the microwave frequency increases, but only the increase of the magnetic mirror ratio ensures that the increased plasma energy content can be efficiently exploited, by making more stable the confinement. Assuming the conclusions of the high-B mode concept, and linking them with the experimental results, it is known that the magnetic field distribution in the plasma chamber must obey the following rules:

(a) the radial magnetic field value at the plasma chamber wall must be $B_{\mathrm{rad}} \geq 2B_{\mathrm{ECR}}$;
(b) the axial magnetic field value at injection must be $B_{\mathrm{inj}} \simeq 3B_{\mathrm{ECR}}$ or more;
(c) the axial magnetic field value at extraction must be about $B_{\mathrm{ext}} \simeq B_{\mathrm{rad}}$; and
(d) the axial magnetic field value at minimum must be in the range $0.30 < B_{\mathrm{min}}/B_{\mathrm{rad}} < 0.45$.

Then, according to the standard model, the ECRIS development is strictly linked to the improvements of superconducting magnets and of the microwave generator technology. There is no evidence for scaling laws for the microwave power and the ECR heating process (ECRH), because the relation between the power and the magnetic field is not so simple. Some authors studied the RF coupling to the plasma in terms of the maximum power rate per unit volume and its relationship with the beam intensity produced by different ECR ion sources [14]. Only recently it was observed that the efficacy of the RF to plasma energy transfer depends also on the amount of wave energy coupled to the plasma chamber. Thus the cavity design, and in particular the microwave injection geometry, is of primary importance for a high RF energy transmission coefficient into the plasma chamber.

Note that the problem of the wave energy transmission into the plasma must be divided into two parts: the first is connected to the microwave generator–waveguide–plasma chamber coupling, while the second one relates to the wave–plasma interaction. The standard model does not allow one to argue for any prediction about the mechanism at the basis of electron and ion dynamics for different magnetic field profiles and for different microwave frequencies. Some preliminary experimental results obtained with a third-generation ECRIS (VENUS) showed the limitations of the standard model, in particular with regard to the magnetic field scaling. Although the standard model has permitted strong improvement in the ECRIS performance up to the third-generation ones (ECRIS that operate at 28 GHz and with magnetic field up to 4 T), we expect that the technological limits can be overcome by means of a better understanding of plasma physics, in particular of the ECR heating mechanism.

Since 1994 (first evidence for the two-frequency heating effect), several possible ways to improve ECRIS performance overcoming the 'brute force' scaling of the main operative parameters have been investigated worldwide.

9.1 Two-frequency heating

Historically, this can be considered the first non-conventional method of plasma heating outside the scheme traced by the standard model. Since 1994, the so-called two-frequency heating (TFH) has been used [15, 16] to improve the HCI production by feeding the plasma with two electromagnetic waves at different frequencies instead of one. In some cases, even three or more close frequencies or a white noise generator (WNG) [17] have been used. The TFH has been demonstrated to be a powerful method. For example, in the case for ^{238}U [15], it increased the production of higher charge states (from 35+ to 39+) by a factor of 2–4 and shifted the peak charge state from 33+ to 36+. Unfortunately, neither the relationship between the two frequencies nor the respective power was unequivocally determined. In fact, any source features a different set of parameters and the optimization is done empirically, just by looking to the maximization of beam current.

Several qualitative explanations have been given about this phenomenon, all related to the increase of the average electron temperature T_e and the ionization rate by assuming that the crossing of two resonance surfaces helps the electron to gain more energy. This simple picture does not explain the reason for the relevant changes in the charge state distribution (CSD) for different pairs of frequencies (even for the case of minor changes, let us say a few MHz over 14 or 18 GHz), which can be explained by the frequency tuning effect. It is important to underline that, even in the case of TFH applied to many existing sources, a TWT is often used, the other being a klystron-based generator. The choice of TWT allows experimentalists to vary the second frequency slightly. As an example, a strong variation of ECRIS performance has been observed in TFH operations [16]. It was observed that the production of O^{7+} with 60 W emitted by the TWT at the optimum frequency gives the same effect as 300 W from the fixed-frequency klystron. The maximum current is obtained by means of klystron at 427 W and TWT at 62 W ($I = 66$ e μA). In order to obtain the same current, they needed $P_{RF} = 800$ W by the klystron in SFH. Furthermore, in the case of TFH, the current increases almost 20% (from 57 to 66 e μA) when the TWT emitted frequency shifts from 11.06 to 10.85 GHz, the klystron and TWT emitted power remaining constant. Thus, the TFH is an effective method to increase the extracted current from ECRIS, but it can be fully exploited only by means of frequency tuning. Several measurements have been carried out with the SERSE ion source. The TFH has been used for operation at either 14 or 18 GHz, with a clear advantage with respect to SFH. It was not successful when a 28 GHz gyrotron and an 18 GHz klystron were used at the same time [18], and that is evidence that fixed-frequency generators are not adequate for TFH. The performance obtained by a two-fixed-frequency klystron was also not as good as that obtained by a fixed-frequency generator plus a TWT amplifier.

As an example, for tin production, a current of the order of 2–3 e μA was measured for the charge state 29^+ when a power of 1.4 kW from the 18 GHz klystron and a power of 1.0 kW from the 14.5 GHz klystron were used. By keeping all the parameters and the 18 GHz klystron power unchanged, a TWT amplifier operating in the range of 8–18 GHz replaced the 14.5 GHz klystron. Even if the maximum power from the TWT was only 200 W, it permitted one to obtain more than 3 e μA of the charge state 29^+ with a frequency of 17.600 GHz. Another interesting operating point was observed at the frequency around 17.000 GHz, which was less critical than any other set-up and permitted one to obtain the state 30^+ with the same beam current, by means of only minor changes of the gas pressure. It must be underlined that the vacuum inside the plasma chamber improved by decreasing the total power from 2.4 to 1.6 kW, which permitted one to move the CSD to higher charge states and to have an excellent reproducibility, even after some weeks. In order to compare these results with the performance in SFH mode, the Sn^{30+} current obtained with the same power from the 18 GHz klystron was only 0.7 e μA. All these data are explained by the description of the microwave–plasma interaction given here and they demonstrate the advantage of TFH operations, for highly charged ion production when the chosen frequency has the right electromagnetic field distribution in the plasma chamber. Moreover, the closer frequencies are the best choice.

To explain TFH on the basis of the fundamental processes occurring during the electron–wave interaction, numerical simulations based on a pure Monte Carlo approach were developed. To be adequately modelled, the TFH effects can be depicted in the following way:

1. We double the resonance zone width, so that electrons that do not gain energy during the first crossing may have another possibility to be heated by the second one. Figure 24 shows the locations of these two resonance regions inside the plasma chamber calculated for the initial set-up of our simulations.
2. Preliminary simulations have shown a significant role played by the ECR interaction in recovering electrons otherwise contained in the magnetic loss cones. This effect (a deterministic plug-in of cold electrons) is currently deemed to play the main role in confining most of the plasma inside the volume embedded by the resonance surface. Adding a second resonance, electrons that have not been plugged by the first wave can be recovered by means of the second one, with a further

increase of the plasma confinement.

A series of simulations has been carried out in order to verify the effect of the TFH. In the case in Fig. 24, the two resonance zones are far apart. This allows the electrons to randomize their gyro-phase with respect to the electromagnetic wave. The numerical study has dealt with the case of two electromagnetic waves at different frequencies supplying the mode $TE_{4,6,22}$ (1000 W) and the mode $TE_{4,4,23}$ (300 W), respectively. It is important to note in Table 3 that, by means of the TFH, the final energy is almost twice as high as in the case of SFH at 18 GHz, whereas the amount of recovered electrons (i.e. equal to 38.2%) is twice that obtained in single-frequency operation (18 GHz; 1000 W).

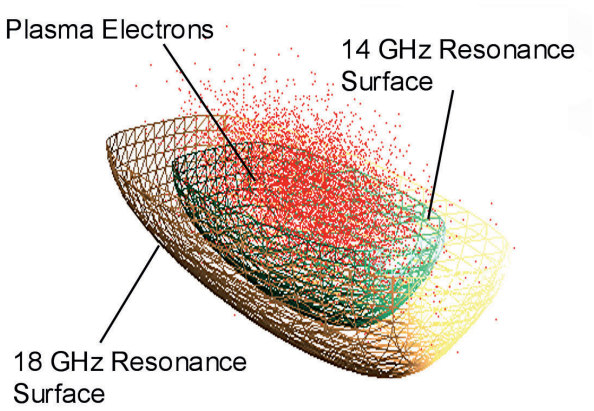


Fig. 24: Section of the two resonance surfaces showing the position of the plasma electrons

Table 3: Percentage of recovered electrons (i.e. electrons that are recovered by the electromagnetic field) and the electron energy after 50 ns for different heating configurations (SFH and TFH) and at different powers.

Frequency (GHz)	Power (W)	Recovered electrons (%)	Energy (keV)
14 + 18	1000 + 300	38.2	1.9589
18	2000	35.2	1.2612
18	1000	19.2	1.0141
14	300	12.9	1.5509

9.2 Frequency tuning effect

Many experiments over the past years have shown that significant improvements of ECRIS performances (in terms of both total extracted current and highly charged ion production) are obtainable by slightly varying the microwave frequency in the case of single frequency heating (SFH), defined as a 'frequency tuning' effect. It was known that the variation of the frequency increases the electron density, and thereby improves the ECRIS performance. However, this occurs only for variations of the order of GHz, and it is strictly connected with the increase of the cutoff density.

On the contrary, several experiments have demonstrated that even slight variations of the pumping wave frequency may lead to strong variation of the extracted current for SFH. Such variations are of the order of MHz to GHz, but they are sufficient to strongly change the wave propagation inside the plasma chamber and the mode pattern on the resonance surface.

The first evidence that significant improvements can be obtained by varying the frequency of the microwaves was given by the different performance observed for the SERSE and CAESAR ion sources

when fed by a klystron-based or a TWT-based generator [19–21] at either 14 or 18 GHz. Similar results came from experiments performed at ORNL and at JYFL [17].

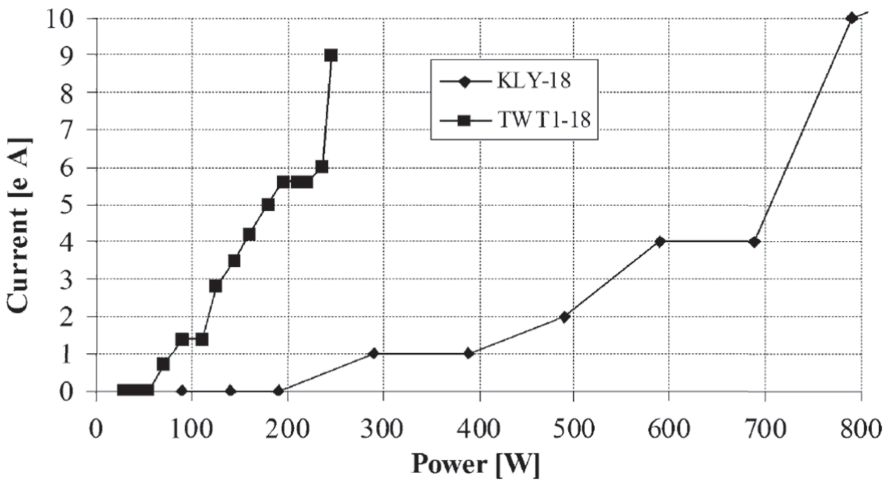


Fig. 25: Comparison between trends of O^{8+} at 18 GHz for a klystron (up to 800 W) and a TWT

Figure 25 shows that a remarkable increase in the production of O^{8+} can be obtained by using the TWT instead of klystrons. In particular, the current of O^{8+} obtained with the TWT at 240 W is obtained by the klystron at 800 W, i.e. a power three times higher. It is important to underline that the two generators operated at two different but close frequencies. In particular, the klystron was operating at 18.0 GHz, while the TWT was at 17.9 GHz (both the amplifiers were fed by dielectric resonator-type oscillators). Such difference of performance was initially explained as a greater frequency dispersion of the TWT, but a series of measurements carried out with the two microwave generators, by means of a spectrum analyser, pointed out that the spectrum of the emitted radiation of a TWT is similar to that of the klystron. So the only difference between the two generators is the output frequency, with the further possibility for the TWT to vary the emitted frequency, thus optimizing the source performance.

The experimental set-ups used for the confirmation of the frequency tuning effect were the SUPERNANOGAN ion source of CNAO, Pavia, and the CAPRICE source at GSI, Darmstadt, which demonstrated that the frequency tuning strongly affects also the ion beam formation. Figure 26 shows the current for C^{4+} obtained with the SUPERNANOGAN ion source in 2005 versus the microwave frequency, keeping the power and all the other source parameters unchanged. The figure features strong fluctuations of the extracted current in a frequency span of about 90 MHz.

The frequency was changed in the range 14.44 to 14.53 GHz with a step of 1 MHz and it was observed that changes of a few MHz changed the C^{4+} current even by 70%. Further experiments including the final validation of the frequency tuning effect were carried out with the SUPERNANOGAN source of CNAO. It permitted an increase by 30% in its performance for C^{4+} and by 50% for H^{3+} (these ions are particularly requested for medical applications), and an additional increase of reliability and availability figures was registered.

Interesting results come from the GSI experiment carried out in 2007. On that occasion, as stated before, the frequency tuning was demonstrated to strongly affect also the beam shape, along with the extracted current for each charge state.

As an example, Fig. 27 reports the shape evolution of a helium beam recorded on a viewer. The two helium charge states can be observed as well as the aberrations introduced by the hexapole. Furthermore, Fig. 27 evidently shows that the beam intensity distribution is inhomogeneous and that this distribution changes with the microwave frequency, while keeping all the other parameters constant.

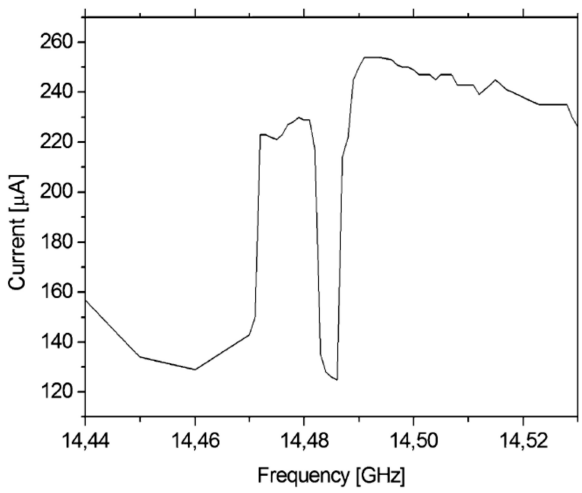


Fig. 26: Trend of the analysed C^{4+} current for the SUPERNANOGAN ECR ion source versus the RF frequency

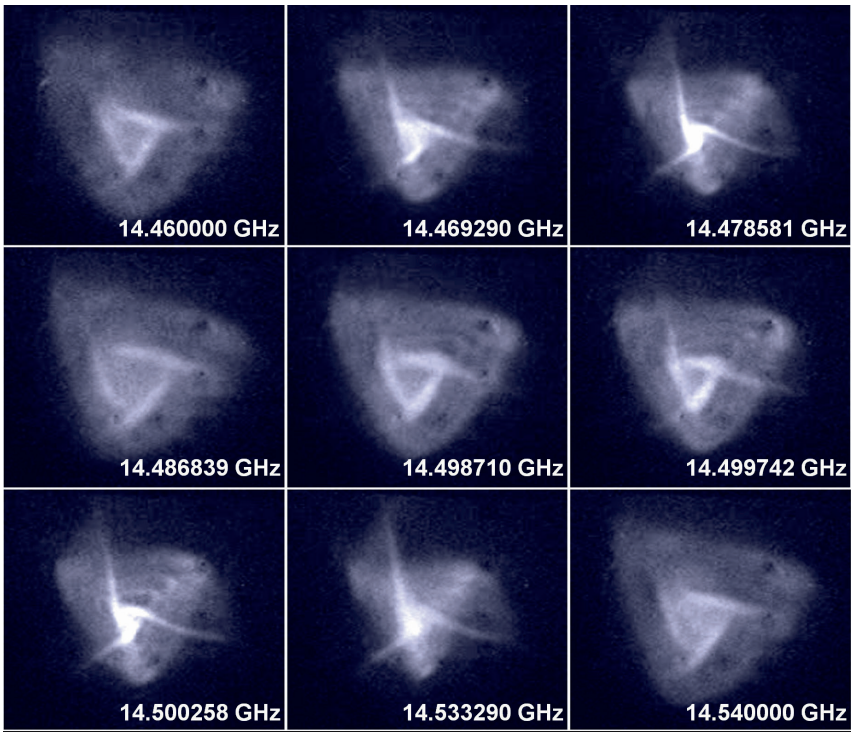


Fig. 27: Variation of the extracted beam shape for different microwave frequencies

Measurements of the S_{11} scattering parameter were carried out in parallel. The S_{11} parameter is connected with the reflection coefficient, and the measurements have shown that, for some frequencies, the amount of reflected power increases, thus demonstrating that some resonances are present inside the cavity with or without plasma. In addition, fluctuation in the extracted power occurs also in the case of quasi-constant reflection coefficient.

A series of simulations were carried out in order to verify the effect of different microwave frequencies on the electron dynamics. The mode pattern over the resonance surface was investigated.

The resonance surface can be calculated numerically once the magnetic field structure is known,

as it corresponds to that surface with constant field equal to $m\omega_{\mathrm{RF}}/q$. Figure 28 features the resonance surface of the SERSE source calculated numerically by means of MATLAB. Note that, by varying the frequency by about 50 MHz, the field distribution over the surface changes remarkably. In particular, if one consider the zones of maxima, the electric field is more than 10^2 higher in the case of Fig. 28(b) than in the case of Fig. 28(a).

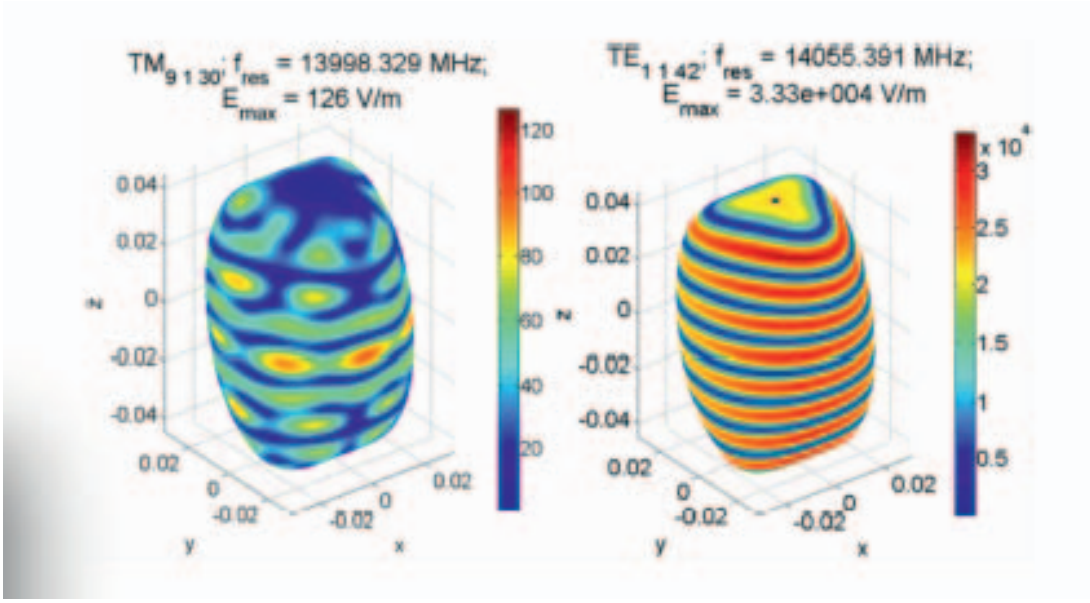


Fig. 28: Simulated resonance surface of the SERSE source with the distribution of the electromagnetic field (false colour representation).

Thus the excitation of near modes, even with distances of a few MHz, strongly changes not only the maximum field over the surface, but also the distribution of the zones of minima and maxima.

Simulations carried out considering different modes have demonstrated that some of them are able to heat the electrons up to energies much higher than others (even ten times higher). Figure 29 features the strong fluctuations of the energy gained by the electrons after 50 ns. The simulations have been done in the proximity of 14 GHz, and for the SERSE source parameters (in terms of magnetic field structure, plasma chamber dimensions and input RF power). The maximum energy after 50 ns has been reached when the $\mathrm{TE}_{1,1,42}$ mode has been used for plasma heating, whereas the minimum energy has been reached for the mode $\mathrm{TE}_{9,1,30}$. Note that these two extreme cases correspond to the field distribution shown in Fig. 28. In our study the net power flowing into the chamber was kept constant, and therefore the observed effects are only due to the different field patterns. In fact, it is clear from the figure that the $\mathrm{TE}_{1,1,42}$ has a higher electric field over a greater part of the resonance surface with respect to the $\mathrm{TE}_{9,1,30}$ mode.

In order to explain the simulation results shown in Fig. 29, we have to take into account that the greatest amount of energy transfer from waves to electrons occurs on the resonance surface because of the ECR condition. On the other hand, the electrons gyrate around the field lines with the Larmor frequency, with their guiding centre moving along the magnetic lines. Then the structure of the magnetic field determines the crossing points of the electron trajectories with the resonance surface.

Figure 30 points out that only in particular zones of the resonance surface does energy transfer occur, and the amount of energy gained by the electrons depends strongly on the mode structure, as only in the case of high-field zones in the proximity of the trajectory crossing points is it possible for the electron to gain very high energies.

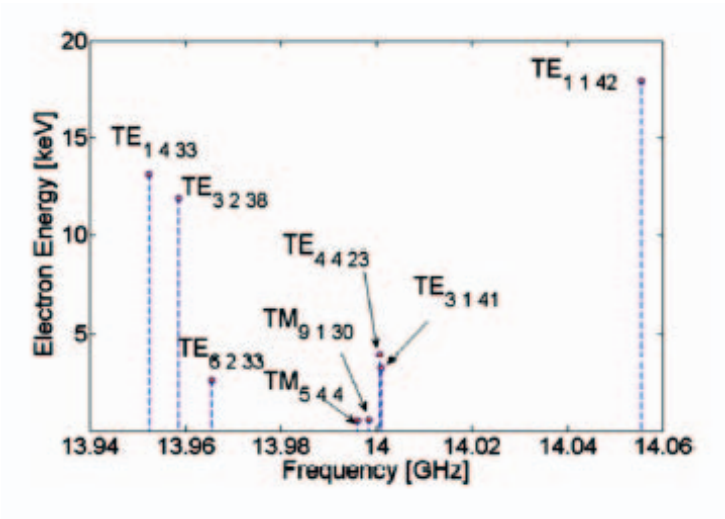


Fig. 29: Simulation of electron heating rapidity for different modes

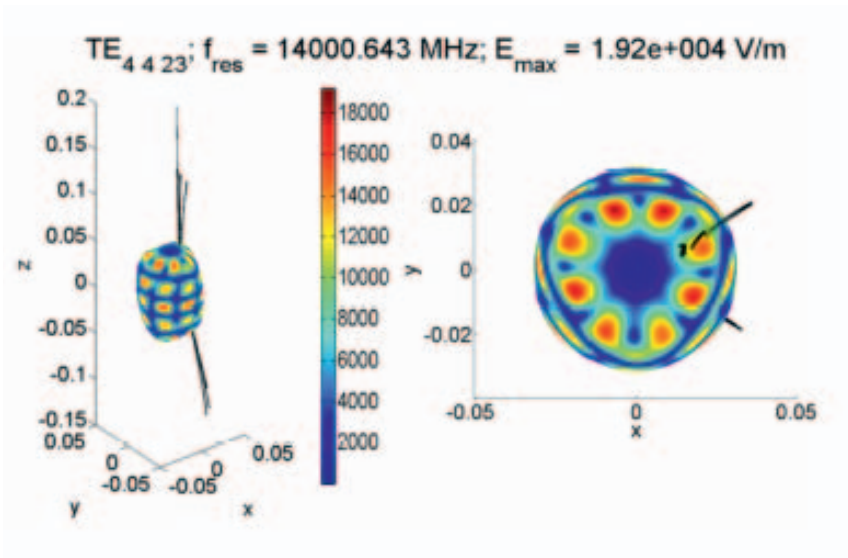


Fig. 30: The electron trajectory and the simulated resonance surface with the electromagnetic field. The crossing point between the electron trajectory and the resonance surface is of primary importance for the electron heating.

The situation here describes reasonably the CAPRICE results, and the simulation results demonstrate that variation of frequency of the order of tens of MHz (over a pumping frequency of 14 GHz) strongly affect the electron heating, and thus also the source performance.

The results obtained with SERSE fed with TWT or klystron microwave generators can also be explained thanks to the simulation results presented above. The TWT-based generators, in fact, are able to change the output frequency, and so by using them it is possible to inject into the plasma chamber the best frequency in order to maximize the heating rapidity v_{h}.

Note that the higher the heating rapidity, the higher is the replacement of the electrons lost from the confinement because of the Spitzer collisions. This effect is of primary importance because it permits one to obtain high densities of warm electrons.

The electromagnetic field patterns over the ECR surface for several frequencies are shown in Fig. 28. It is clear that on passing from 14 to 18 GHz the pattern of the resonant modes over the ECR surface changes considerably. This means that after the effects due to the scaling laws, we have to take into account the differences in the field patterns when the frequency changes.

10 Highly charged ions – perspectives

There is no doubt among experts that the incredible rate of increase of ion source performance cannot be sustained in the future, but some of the recent results have demonstrated that a better comprehension of plasma physics may permit additional increases of average charge states and currents. Further tests in *ad hoc* test benches are under way to improve the knowledge of the new plasma heating schemes, and the saturation of magnet performances may not be a limiting factor in the future. The proposal of fourth-generation ECRIS with operational frequency above 50 GHz is appealing but it is questionable if a similar device may run with stable and reproducible beams, to the extent that heating is not controlled.

Acknowledgements

The chapter presented here has been prepared in strict collaboration with David Mascali and Giuseppe Torrisi, whose active contributions permitted my schematic lesson notes to be translated into a readable form.

References

[1] S. Chandrasekhar, *Principles of Stellar Dynamics* (University of Chicago Press, Chicago, IL, 1942).

[2] R. Geller, *Electron Cyclotron Resonance Ion Sources and ECR Plasmas* (Institute of Physics, Philadelphia, PA, 1996).

[3] B. Wolf, *Handbook of Ion Sources* (CRC Press, Boca Raton, FL, 1995).

[4] D. Mascali, L. Torrisi, D. Margarone, S. Gammino, R. Miracoli, N. Gambino, L. Celona, G. Ciavola and F. Maimone, Langmuir probe characterization of laser generated plasmas. *Acta Technica CSAV (CeskoSlovensk Akademie Ved)* 53 (4) , pp. 333-345, accepted for publication.

[5] L. Torrisi, S. Gammino, L. Andò, L. Laska and J. Krasa, Tantalum ions production by 1064 nm pulsed laser irradiation. *J. Appl. Phys.* **91** (2002) 4685–4692.

[6] F.F. Chen, *Introduction to Plasma Physics and Controlled Fusion*, vol. 1, *Plasma Physics* (Kluwer Academic Plenum Plublisher).

[7] M.S. Ioffe,"Mirror traps", *Plasma Physics* (IAEA, Vienna, 1965),

[8] G. Ciavola and S. Gammino, *Rev. Sci. Instrum.* **63**(4) (1992) 2881.

[9] F. Jeager, A.J. Lichtenberg and M.A. Lieberman, Theory of electron cyclotron resonance heating II. Long time and stochastic effects. *Plasma Phys.* **15**(2) (1973) 125–150.

[10] F. Jeager, A.J. Lichtenberg and M.A. Lieberman, Theory of electron cyclotron resonance heating I. Short time and adiabatic effects. *Plasma Phys.* **14**(12) (1972) 1073–1100.

[11] S. Bliman, R. Geller, W. Hess, B. Jaquot and C. Jaquot, *IEEE Trans. Nucl. Sci.* **NS-23**(2) (1972) 200.

[12] R. Geller, F. Bourg, P. Briand, J. Debernardi, M. Delaunay, B. Jacquot, P. Ludwig, R. Pauthenet, M. Pontonnier and P. Sortais, The Grenoble ECRIS status 1987 and proposal for ECRIS scaling, Proc. 8th Int. Workshop on ECR Ion Sources, East Lansing, 1987, vol. 1, p. 22. NSCL Report MSUCP-47.

[13] C. Barue, M. Lamoureux, P. Briand, A. Girard and G. Mellin, *J. Appl. Phys.* **76** (1994) 2662.

[14] D. Hitz *et al.*, The comparison of 18 and 28 GHz behavior of SERSE: some conclusions, Proc. 15th Int. Workshop on ECR Ion Sources, Jyvaskyla, Finland, 2002, p. 100.

[15] Z.Q. Xie and C.M. Lyneis, Improvements on the LBL AECR source, Proc. 12th Int. Workshop on ECR Ion Sources, Tokyo, Japan, 1995.

[16] R.C. Vondrasek, R.H. Scott, R.C. Pardo and H. Koivisto, Operational improvements of the Argonne ECR sources, Proc. 15th Int. Workshop on ECR Ion Sources, Jyvaskyla, Finland, 2002, p. 174.

[17] Y. Kawai, G.D. Alton, O. Tarvainen, P. Suominen and H. Koivisto, Effect of broadband microwave radiation on the performance of a conventional B-minimum geometry electron cyclotron resonance ion source. *Rev. Sci. Instrum.* **77**(3) (2006) 03A331.

[18] S.Gammino, G. Ciavola, L. Celona, D. Hitz, A. Girard and G. Melin, Operation of the SERSE superconducting electron cyclotron resonance ion source at 28 GHz. *Rev. Sci. Instrum.* **72**(11) (2001) 4090–4097.

[19] L. Celona *et al.*, Analysis of the SERSE ion output by using klystron-based or TWT-based microwave generators. *AIP Conf. Proc.* **749** (2005) 99.

[20] S. Gammino, G. Ciavola and L. Celona, Difference in the performances of ECR ion sources when they are fed by klystron-based generator or travelling-wave-tube-based generator. *Nucl. Instrum. Methods Phys. Res.* **A491** (2002) 342–345.

[21] R. Vondrasek *et al.*, *AIP Conf. Proc.* **749** (2005) 31.

Electron Beam Ion Sources

G.Zschornack[a,b], *M.Schmidt*[b] *and A.Thorn*[b]
[a]University of Technology Dresden, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany
[b]Dreebit GmbH, Dresden, Germany

Abstract
Electron beam ion sources (EBISs) are ion sources that work based on the principle of electron impact ionization, allowing the production of very highly charged ions. The ions produced can be extracted as a DC ion beam as well as ion pulses of different time structures. In comparison to most of the other known ion sources, EBISs feature ion beams with very good beam emittances and a low energy spread. Furthermore, EBISs are excellent sources of photons (X-rays, ultraviolet, extreme ultraviolet, visible light) from highly charged ions. This chapter gives an overview of EBIS physics, the principle of operation, and the known technical solutions. Using examples, the performance of EBISs as well as their applications in various fields of basic research, technology and medicine are discussed.

1 Introduction

The idea to develop electron beam ion sources (EBISs) was constituted by the need for multiply charged ions for accelerator applications to derive high final particle kinetic energies and by a general scientific interest to study exotic states of matter as represented by highly charged ions (HCIs).

Highly charged ions possess properties that differ in many ways from atoms and low charged ions. The following are some of the distinguishing characteristics.

- Due to the ionization process, a large amount of potential energy is stored in HCIs. For example, a Xe^{44+} ion has a potential energy of about 51 keV.
- During ion–surface interactions, this potential energy is released within a few 10 fs over an area of about 100 nm^2, which leads to power densities of 10^{12} up to 10^{14} W cm^{-2}.
- Due to the high power density in the surface interaction area compared to low charged ions (LCIs), HCIs produce high amounts of secondary particles such as neutrals as well as secondary electrons and secondary ions.
- HCIs possess very strong static electric fields, reaching from 10^{14} up to 10^{16} V cm^{-1}.
- HCIs are characterized by a very effective stopping power (e.g. for Au^{69+}, 100 keV nm^{-1}).
- Due to their high ionic charge, HCIs can be accelerated very effectively ($\propto q$ for linear accelerators and $\propto q^2$ for circular accelerators; q denotes the ion charge state).
- Nowadays, HCIs can be produced in a compact laboratory set-up. Large accelerator structures are used if high HCI currents are necessary for a specific experiment but are not required for many applications.

These unique properties are the reason why the interest in HCIs is constantly growing and the number of applications increases. Figure 1 gives several examples for applications of HCIs in basic research as well as technology.

HCIs appear only in the plasma of fusion facilities, the corona of the Sun, at the border of black holes, and in similar exotic cosmic regions. The first publications on HCIs date back more than 80 years and the number of publications has increased continually. In the first classical papers Bowen and Millikan reported on the production of Se^{20+} ions in an arc discharge [1] in 1925 and Edlen explained the origin

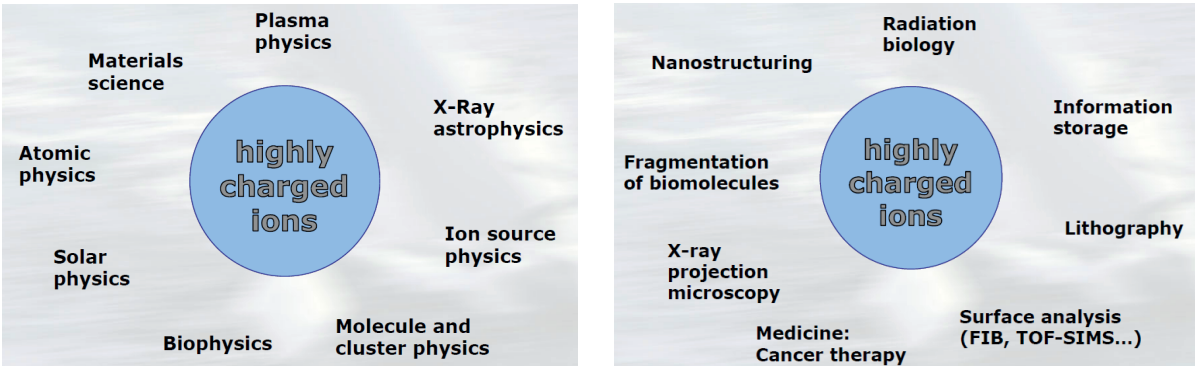


Fig. 1: Some applications of HCIs in basic research (left side) and possible applications in technology (right side)

of spectral lines in the corona of the Sun as emission lines from 10- to 15-fold ionized iron, calcium and nickel atoms in 1942 [2].

A classical approach to produce HCIs in the laboratory is the use of ion accelerators by stripping of energetic LCIs to all practically possible ion charge states. As a result of the stripping process, HCIs of high kinetic energy are produced, which must be decelerated in many cases for further use. Another way to access HCIs is the use of powerful ion sources such as the following:

- EBISs;
- electron cyclotron resonance (ECR) ion sources (ECRISs) (for basics see Ref. [3]); or
- laser ion sources (for basics see Ref. [4]).

This present chapter is about EBISs. For details on other kinds of ion sources, we refer to the citations above and the references therein. The aim of this work is to give a short up-to-date overview about the function and use of EBIS technology. More details can be found in the literature, e.g. in the book of Gillaspy [5] and the paper written by Currell [6]. More detailed information about the basic physics of HCIs can be found in the review by Gillaspy [7] and the book written by Beyer and Shevelko [8].

2 EBIS: the basic idea

2.1 The technical solution

EBIS/T ion sources have been known for more than 40 years and the technical design of these sources has been continually improved with increasing technical possibilities. The principle of operation of an EBIS is shown in Fig. 2.

The ion production in an EBIS is based on electron impact ionization in a high-density electron beam that is compressed by a strong magnetic field produced by magnetic coils. An electron gun with a cathode of high emissivity produces the electron beam. The ions that are produced by successive electron impact ionizations inside the electron beam are radially confined by the negative space charge of the electron beam, and the axial confinement of the ions is achieved by superposition of electrostatic potentials generated using a minimum of three collinear drift tubes. The higher potentials on the outer drift tubes create an axial electrostatic trap. Hence, the ionization time of the ions can be controlled by switching the axial trap potentials periodically on and off.

The magnetic field for the electron beam compression is generated by a special magnet system, which can consist of superconducting magnetic coils or permanent magnet rings. The electron beam is dumped in an electron beam collector where the electrons are separated from the extracted ions. Usually EBISs are equipped with one or more radial ports. This allows for spectrometry of electromagnetic radiation produced in the region of the ion trap during the ionization process.

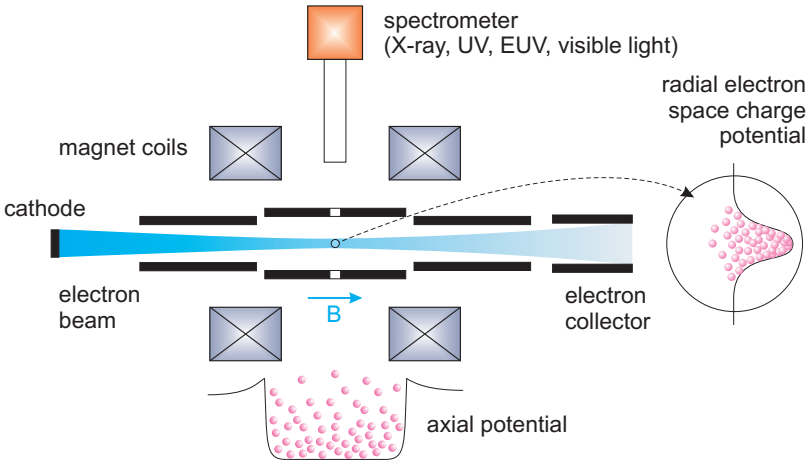


Fig. 2: Principle of operation of an EBIS

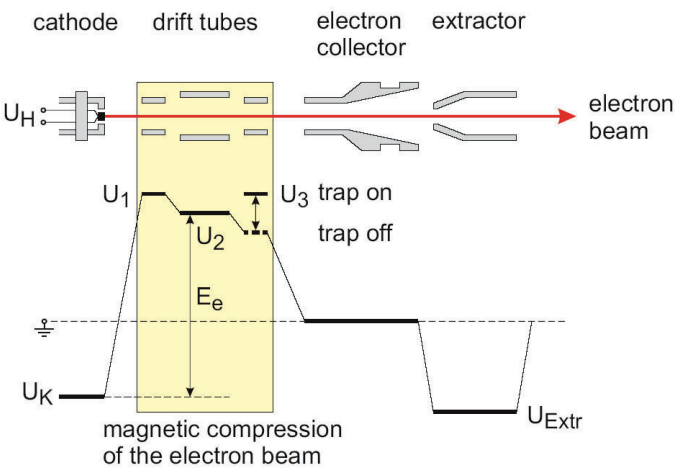


Fig. 3: Electrical scheme of EBIS operation: U_{K}, cathode potential; U_{Extr}, ion extraction potential; U_{H}, potential at the cathode heater; U_1, trap potential at the side of the electron gun; U_3, trap potential at the ion extraction side (switched for pulsed ion extraction); U_2, potential at the middle drift tube.

The electrical scheme of EBIS operation is shown in Fig. 3 for an EBIS with three drift tubes (solutions are also known with more drift tubes). The electron energy E_{e} in the centre of the drift tubes where the ions are produced and accumulated is

$$E_{\mathrm{e}} = e(U_2 - U_{\mathrm{K}} + U_{\mathrm{e}}), \tag{1}$$

where U_{e} is the space-charge potential of the electron beam, which slightly decreases the final electron beam energy. The axial ion trap is controlled by switching the potential U_3 in the indicated manner. The chosen dynamics and height of the potential wall formed by the voltage U_3 allows the following three principal modes of source operation.

– *Permanently opened trap – transmission mode.* The trap is permanently open and ions are produced in the electron beam without axial trapping. This mode delivers high currents of the lowest charged ions with ion beam currents up to microamps.

– *Partially closed trap – leaky mode.* Selecting a low axial potential wall, a certain number of ions with adequate kinetic energy can overcome the potential wall and are extracted continuously. This mode delivers ions with preferentially low up to intermediate ion charge states and a small fraction of higher ion charge states, with ion beam currents up to nanoamps.
– *Periodically opened and closed trap – pulsed mode.* In this case the voltage U_3 is high enough to trap all ions axially. Periodic opening of the trap releases pulses of ions extracted with typical pulse widths in the order of some microseconds and allows the production of the highest currents of HCIs (up to microamps during the pulse).

Table 1 compares important practically realized parameters of cryogenic and room-temperature EBISs.

Table 1: Parameters of cryogenic and room-temperature EBISs.

Parameter	Cryogenic EBIS	Room-temperature EBIS
Magnetic system	superconducting coils (3–8 T) on axis	permanent magnets (SmCo, NdFeB) (250–600 mT) on axis
Electron beam currents	up to amps	$\leq$200 mA
Electron beam energies	up to 200 keV	up to 30 keV
Electron beam densities	$j_e > 1000$ A cm^{-2}	$j_e < 500$ A cm^{-2}
Basic vacuum	up to 10^{-12} mbar	up to 10^{-10} mbar
Length of the ionization region	up to metres	$\leq$6 cm
Ion charge states	highest ion charge states Xe^{54+}, U^{92+} at maximum	bare ions up to $Z = 28$, Kr^{34+}, Xe^{48+}, Au^{60+}, etc.
Set-up time	$\geq$1 day	hours
Remarks	large devices liquid helium cooling latest developments: refrigerator cooling	compact transportable low initial and maintenance costs short set-up times

2.2 The history of EBISs: a short retrospect

EBIT/EBIS ion sources have now been available as reliable sources of HCIs for about 20 years. After the first demonstration of the working principle of an electron beam ion source (EBIS) in the pioneer works of Donets, Ovsyannikov and collaborators [9–11], the development of the first compact electron beam ion trap (EBIT) was reported by Levine *et al.* [12]. Starting from that date, many different EBIT/EBIS ion sources have been developed and put into operation at different places all over the world.

First of all, it seems to be necessary to clear up the terms 'EBIT' and 'EBIS'. Both terms can be found in the literature. The term EBIT stands for electron beam ion trap. Since all EBITs can in principle also work like an EBIS, it seems to be more common to speak about an EBIS unless the function as an ion trap is of interest, as is the case in spectroscopic studies on HCIs.

Making no claim to be complete, sources such as the LLNL EBIT, the NIST EBIT, the Tokyo EBIT, the Shanghai EBIT and the Heidelberg EBIT with their different modifications should be mentioned here. Furthermore, the first attempts to built a refrigerated (cryogenic closed cycle, no liquid helium) EBIT with a 2 cm ion trap using outer correction coils are known [13] and were realized with the Stockholm EBIT. Most of these EBITs are special constructions realized by several laboratories with the goal to perform specific investigations on highly charged ions.

Another line of development is the series of compact room-temperature EBIT/EBIS ion sources of the Dresden EBIT/EBIS family [14, 15]. They have a patented operation principle that allows the production of highly charged ions efficiently and with ion beam currents competitive with other known EBIT/EBIS solutions. Continuing this way a new refrigerated superconducting ion source, the Dresden EBIS-SC, was constructed and commissioned in 2010 [16]. In this context the basic idea was to build an ion source working with a liquid-helium-free refrigerator technique for cooling the superconductors that generate the strong magnetic field compressing the electron beam and to realize a very compact design of the source body. In spite of the compact ion source assembly, the system has proved successful for source operation with electron beams up to 700 mA.

An overview of EBIS installations in 2010 and earlier can be found in Ref. [17]. Nevertheless, the number of employed EBIS/EBIT is increasing permanently. In the last two years installations in Kielce and Krakow (both Poland), Huddersfield (Great Britain), Clemson (USA), Dresden-Rossendorf and Jena (both Germany) and Shanghai (China) have emerged.

Most of the EBISs of international reputation are special laboratory constructions, where each EBIS has its own particularities. Commercial versions of EBIT and EBIS devices are known only from two vendors:

- *Physics & Technology (Livermore, USA).* Offered is a so-called REBIT/S (Refrigerated Electron Beam Ion Trap/Source; for details see [13]).
- *Dreebit GmbH (Dresden, Germany).* Offered is a series of compact and efficient room-temperature EBIT and EBIS sources (Dresden EBIT, Dresden EBIS and Dresden EBIS-A) as well as a superconducting EBIS (Dresden EBIS-SC), using refrigerating technology for magnet cooling (for details see [15]).

3 EBISs: the basic physics

3.1 The ionization balance

The intended purpose of an EBIS is to produce highly charged ions. To produce HCIs a high-density electron beam acts as ionization medium. For neutrals and ions in the electron beam, the balance between

- charge generating processes and
- charge destructive processes

is of importance for reaching a certain ionization stage.

For ion production the dominant process is electron impact ionization. Therefore, only this process is considered at first. For the ion production in an EBIS we can write

$$\begin{array}{lcl} \text{to } q=1 & \dfrac{\mathrm{d}n_0}{\mathrm{d}t} & = & -\lambda_1 n_0, \\ \vdots & \vdots & & \vdots \\ \text{to } q & \dfrac{\mathrm{d}n_q}{\mathrm{d}t} & = & \underbrace{\lambda_q n_{q-1}}_{\text{ion production}} - \underbrace{\lambda_{q+1} n_q}_{\text{ion destruction}}, \\ \vdots & \vdots & & \vdots \\ \text{to } Z & \dfrac{\mathrm{d}n_Z}{\mathrm{d}t} & = & \lambda_Z n_{Z-1}, \end{array} \tag{2}$$

where n_0 is the neutral particle density. For λ_q we have that

$$\lambda_q = \sigma_q v_\mathrm{e} n_\mathrm{e} \qquad \text{(reaction rate)}, \tag{3}$$

where λ_q has the dimension of s^{-1} and σ_q is the ionization cross-section, v_e the electron velocity and n_e the electron density.

If the ionization process starts only with neutral particles it yields

$$t = 0, \qquad n_0(0) = n_0^0, \qquad n_q(0) = 0.$$

This leads to the solution

$$\begin{aligned}
n_0 &= n_0^0\,\mathrm{e}^{-\lambda_1 t}, \\
n_1 &= \frac{n_0^0\lambda_1}{\lambda_2-\lambda_1}(\mathrm{e}^{-\lambda_1 t} - \mathrm{e}^{-\lambda_2 t}), \\
n_2 &= n_0^0\lambda_1\lambda_2\left(\frac{\mathrm{e}^{-\lambda_1 t}}{(\lambda_2-\lambda_1)(\lambda_3-\lambda_1)} + \frac{\mathrm{e}^{-\lambda_2 t}}{(\lambda_3-\lambda_2)(\lambda_1-\lambda_2)} + \frac{\mathrm{e}^{-\lambda_3 t}}{(\lambda_1-\lambda_3)(\lambda_2-\lambda_3)}\right), \\
\vdots \quad & \quad \vdots \\
n_q &= n_0^0\prod_{l=1}^{q}\lambda_l\left(\sum_{j=1}^{q+1}\frac{\mathrm{e}^{-\lambda_j t}}{\prod_{k=1,\,k\neq q}^{q+1}(\lambda_k-\lambda_j)}\right).
\end{aligned} \tag{4}$$

In principle, Eq. (2) can also be integrated for other conditions, such as, for example, for $n_q(0) \neq 0$, for charge breeding processes or for external ion injection into an EBIS. With Eq. (3) the production rate of ions with individual ion charge states can be determined.

If we develop the exponents in Eq. (4) it yields

$$\begin{aligned}
n_0 &= n_0^0(1-\lambda_1 t), \\
n_1 &= n_0^0\lambda_1 t, \\
n_2 &= n_0^0\lambda_1\lambda_2\frac{t^2}{2}, \\
\vdots \quad & \quad \vdots \\
n_q &= n_0^0\frac{t^q}{q!}\prod_{j=1}^{q}\lambda_j.
\end{aligned} \tag{5}$$

The characteristic time τ for ion production (ionization time) follows using

$$t_q = \frac{1}{\lambda_q} = \frac{1}{\sigma_q v_\mathrm{e} n_\mathrm{e}} \qquad \left(\frac{1}{\text{reaction rate}}\right), \tag{6}$$

to give

$$\begin{aligned}
\tau_1 &= t_1, \\
\tau_2 &= t_1 + t_2, \\
\vdots \quad & \quad \vdots \\
\tau_q &= \sum_{k=1}^{q} t_q.
\end{aligned} \tag{7}$$

For estimating when a certain ionization stage reaches its maximum in the actual ion charge-state distribution, Eq. (7) can be used. Hence we write

$$j_\mathrm{e}\tau_q = \sum_{k=1}^{q}\frac{e}{\sigma_k} \qquad \text{(ionization factor)}, \tag{8}$$

with

$$j_e = n_e v_e$$

as the electron beam density (A cm^{-2}). The ionization factor is a fundamental quantity that determines the ionization stage in an ion source.

The production of a mean ion charge state $\bar{q}$ is only possible if a certain $j\tau$ value (ionization factor) is reached. The following must hold:

$$j_e \tau_q \geq \sum_{k=1}^{q} \frac{e}{\sigma_k}. \tag{9}$$

The precondition must be an electron energy E_e higher than the ionization potential I_q (ideally about (2–3)$\times I_q$).

The best values for the ionization factor and the electron beam energy under consideration of the excitation function of the electron-impact ionization cross-sections for selected elements are shown in Fig. 4. Thus producing high ionization stages requires

- a sufficient electron beam energy and
- a sufficiently high ionization factor.

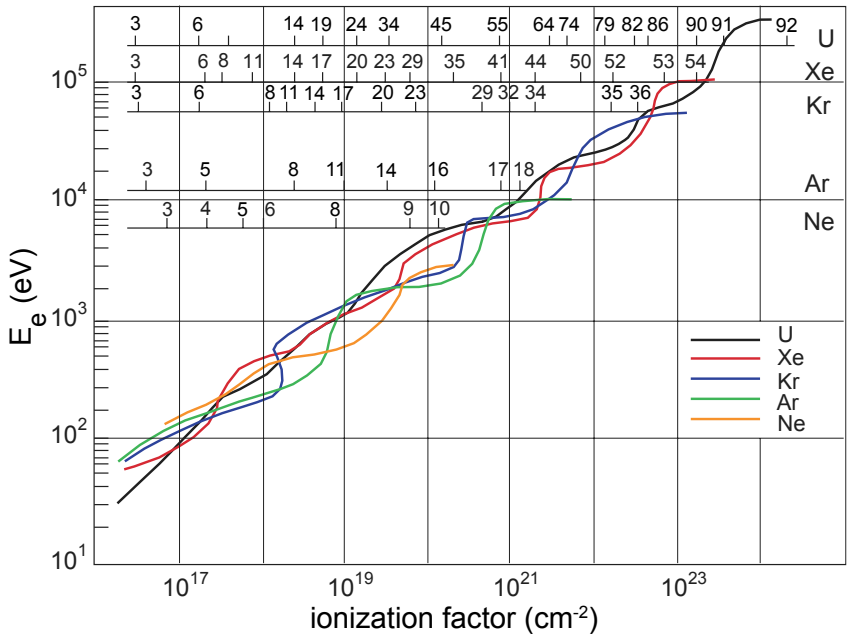


Fig. 4: Interrelation between ionization factor and electron energy for the effective production of ions in different ion charge states (according to an idea of Prof. E. D. Donets (Dubna)).

In order to explain the role of electron binding energies, in Fig. 5 we show in more detail the electron binding energies of ions for different isoelectronic sequences as a function of the atomic number. Thereby it is considered that the ionization cross-section has its maximum at about 2.7 times the electron binding energy of the electron to be ionized.

In practice, processes such as

- charge exchange and
- ion losses from the source

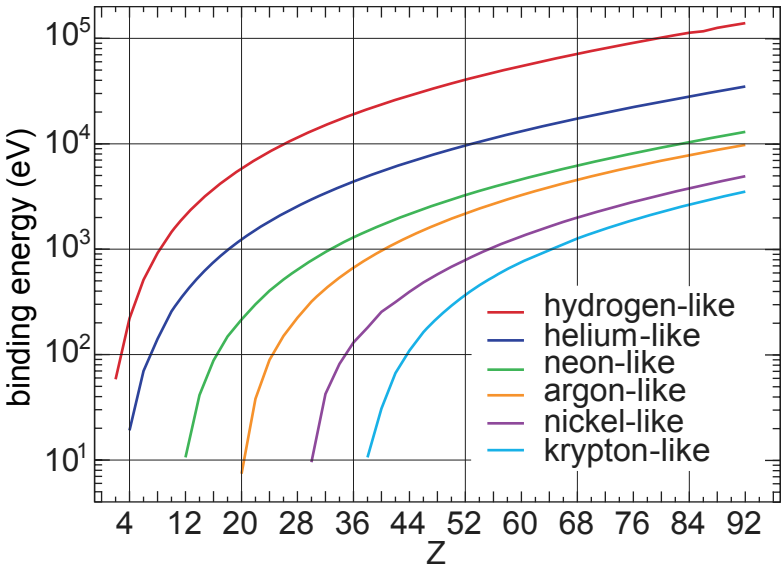


Fig. 5: Electron binding energies for the ion production of different isoelectronic sequences

can reduce the ionization power significantly. Of special importance in this case is the charge exchange with neutrals. The ionization factor as a function of the ionization time and the electron energy for selected isoelectronic sequences is given in Table 2 (according to Ref. [18]).

Figure 6 shows this behaviour for xenon ions. The ionization factor for the production of a certain ion charge state is minimal at 2.7 times the electron binding energy and increases again at higher electron energies. To produce the highest ionization charge states, ionization factors many orders of magnitude higher than those ones for low charged ions must be reached. It should be noted again that the issues discussed are only valid under the idealized condition that only electron impact ionization takes place.

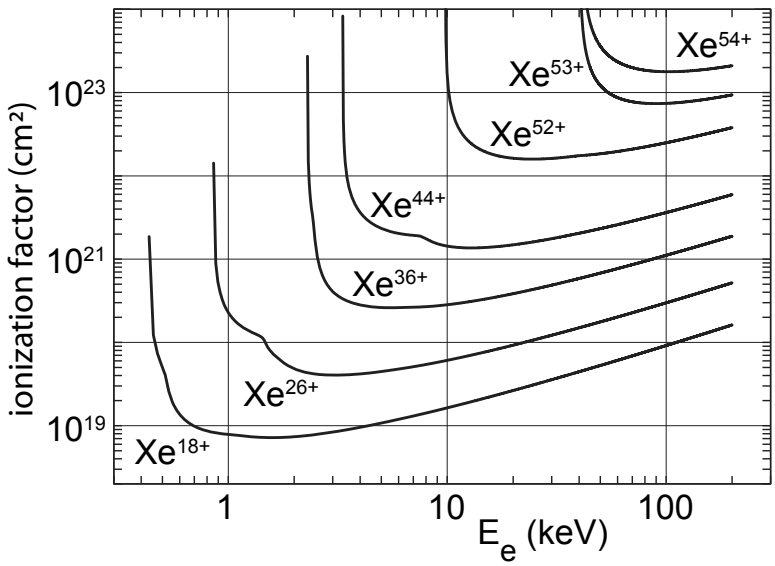


Fig. 6: Ionization factors for the production of individual xenon ionization charge states

Table 2: Ideal production conditions for ions of different isoelectronic sequences. Given are the ionization factor $j_e\tau$ (e^- cm^{-2}), the optimal electron beam energy (keV) and the required ionization time (ms or s) for an assumed ionization factor of $j_e\tau = 3 \times 10^{22}$ e^- cm^{-2}.

Sequence	Neon $Z = 10$	Argon $Z = 18$	Krypton $Z = 36$	Xenon $Z = 54$	Gold $Z = 79$	Uranium $Z = 92$
Atom fully ionized	Ne^{10+} 2×10^{21} 3 7 ms	Ar^{18+} 2×10^{21} 9 67 ms	Kr^{36+} 3×10^{22} 40 1 s	Xe^{54+} 2×10^{23} 80 7 s	Au^{79+} 6×10^{23} 180 20 s	U^{92+} $2 \times 10^{24+}$ 300 67 s
Helium- like	Ne^{8+} 8×10^{18} 0.6 0.3 ms	Ar^{16+} 1×10^{20} 2 3 ms	Kr^{34+} 2×10^{21} 7 67 ms	Xe^{52+} 2×10^{22} 20 0.7 s	Au^{77+} 6×10^{22} 45 2 s	U^{90+} 2×10^{23} 70 7 s
Neon- like		Ar^{8+} 3×10^{18} 0.3 0.1 ms	Kr^{28+} 3×10^{20} 4 10 ms	Xe^{44+} 2×10^{21} 8 67 ms	Au^{69+} 6×10^{21} 17 200 ms	U^{82+} 3×10^{22} 30 1 s
Argon- like			Kr^{18+} 1×10^{19} 1.5 0.3 ms	Xe^{36+} 2×10^{20} 5 7 ms	Au^{61+} 1×10^{21} 12 33 ms	U^{74+} 5×10^{21} 20 167 ms
Krypton- like				Xe^{18+} 6×10^{18} 1 0.4 ms	Au^{43+} 1×10^{20} 4 3 ms	U^{56+} 7×10^{20} 7 23 ms
Xenon- like					Au^{25+} 2×10^{19} 1.5 0.7 ms	U^{38+} 7×10^{19} 4 2 ms

3.2 Atomic physics for ion production

3.2.1 Overview

The following are the atomic processes contributing to ion production;

1. electron impact ionization,
2. charge exchange,
3. recombination,
4. three-body recombination,
5. dielectronic recombination,
6. photoionization,
7. vacancy cascades and
8. electron shake-off processes.

Later on we will discuss these further processes that are important for ion losses and the energy balance in EBISs:

- ion heating by elastic electron collisions,
- ion–ion energy exchange,
- ion confinement and
- ion losses from the trap.

In the following we will discuss the most important processes in more detail. For the basic atomic processes, some useful estimation formulas are given. Nevertheless, we note that these formulas do not completely substitute more precise quantum-mechanical calculations for individual processes.

3.2.2 Direct Coulomb ionization

In simple cases, the cross-sections for single electron impact ionization

$$\mathrm{X}^{q+} + \mathrm{e}^- \rightarrow \mathrm{X}^{(q+1)+} + 2\mathrm{e}^-$$

can be estimated for different subshells j by the Lotz formula [19]

$$\sigma_{qj} = a_{qj} g_{qj} \frac{\ln(E_\mathrm{e}/I_{qj})}{E_\mathrm{e} I_{qj}} \left\{ 1 - b_{qj} \exp\left[-c_{qj} \left(\frac{E_\mathrm{e}}{I_{qj}} - 1 \right) \right] \right\}. \tag{10}$$

For all ionization processes, the electron energy E_e must be greater than the related ionization potential I_{qj}, otherwise the ionization cross-section is equal to zero. The quantities a_{qj}, b_{qj} and c_{qj} are tabulated constants. For highly charged ions, they are given by

$$a_{qj} = 4.5 \times 10^{-14}\ \mathrm{cm}^2\ \mathrm{eV}^2$$

and

$$b_{qj} = c_{qj} = 0,$$

with g_{qj} as the occupation number of the subshell j of an ion in the charge state q. The quantity I_{qj} is the ionization potential for the ionization of the subshell j. The total ionization cross-section results as a sum over all occupied subshells:

$$\sigma_q(E_\mathrm{e}) = \sum_q \sigma_{qj}(E_\mathrm{e}). \tag{11}$$

As an example, the electron impact ionization cross-sections for argon ions are shown in Fig. 7 [20].

We will not go into the details for the double ionization process

$$\mathrm{X}^{q+} + \mathrm{e}^- \rightarrow \mathrm{X}^{(q+2)+} + 3\mathrm{e}^-$$

but we refer the reader to the estimation formula from Mueller and Frodel [21], which was derived from experimental data. As a rule double ionization cross-sections for most species are about one order of magnitude or more lower than cross-sections for single ionization cross-sections.

3.2.3 Charge exchange

The charge exchange between ions can be written as

$$\mathrm{X}^{q+} + \mathrm{X}^{p+} \rightarrow \mathrm{X}^{i+} + \mathrm{X}^{(q+p-i)+}$$

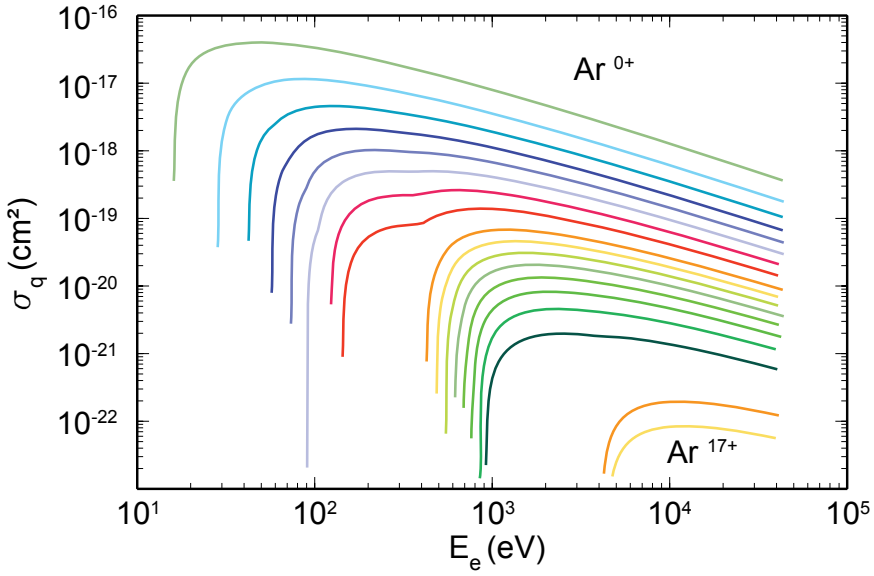


Fig. 7: Single electron impact ionization cross-sections as a function of the electron energy (according to Ref. [20])

and correspondingly for the charge exchange with a neutral atom

$$\mathrm{X}^{q+} + \mathrm{X} \rightarrow \mathrm{X}^{(q-1)+} + \mathrm{X}^{+}$$

Owing to the fact that

$$\mathrm{X}^{+} + \mathrm{X} \rightarrow \mathrm{X} + \mathrm{X}^{+}$$

charge exchange does not play a role for singly charged ions.

Of special importance in this case is the charge exchange of ions with neutrals due to the very high cross-sections. This circumstance is important for EBIS operation and can limit the reachable ionization stages in the electron beam. For collision energies below 25 keV/u, charge exchange can be estimated by the formula from Mueller and Salzborn [22]

$$\sigma_{q\rightarrow q-1}\,(\mathrm{cm}^2) \approx (1.43 \pm 0.76) \times 10^{-12}\, q^{1.17}\, [I_q\,(\mathrm{eV})]^{-2.76} \tag{12}$$

and is of the order of 10^{-14}–10^{-15} cm^{-2}. As a rule, charge exchange is the dominant loss mechanism for highly charged ions and is especially important if the gas pressure in the ion source is relatively high. As we can see from Eq. (12), the cross-section is independent of the energy of the projectiles. In Fig. 8 cross-sections for single and double charge exchange are given for xenon ions.

3.2.4 Radiative recombination

The scheme for radiative recombination (RR)

$$\mathrm{X}^{q+} + \mathrm{e}^{-} \rightarrow \mathrm{X}^{(q-1)+} + \hbar\omega$$

is shown in Fig. 9.

The cross-section for charge exchange processes can be estimated by the formula from Kim and Pratt [24]

$$\sigma_q^{\mathrm{RR}}(E_\mathrm{e}) = \frac{8\pi}{3\sqrt{3}}\alpha\lambda_\mathrm{e}^2\chi_q(E_\mathrm{e})\ln\left(1 + \frac{\chi_q(E_\mathrm{e})}{2\hat{n}}\right), \tag{13}$$

with

$$\chi_q(E_\mathrm{e}) = (Z+q)^2 \frac{I_\mathrm{H}}{4E_\mathrm{e}}$$

and

$$\hat{n} = n + (1 - W_n) - 0.3.$$

Here n is the main quantum number of the valence shell, W_n is ratio of the number of unoccupied states to the total number of states in the valence shell, $\lambda_\mathrm{e} = 3.861 \times 10^{-11}$ cm (electronic Compton wavelength) and I_H = 13.6 eV. As an example we show the cross-sections for radiative recombination on xenon ions as a function of the electron energy in Fig. 10.

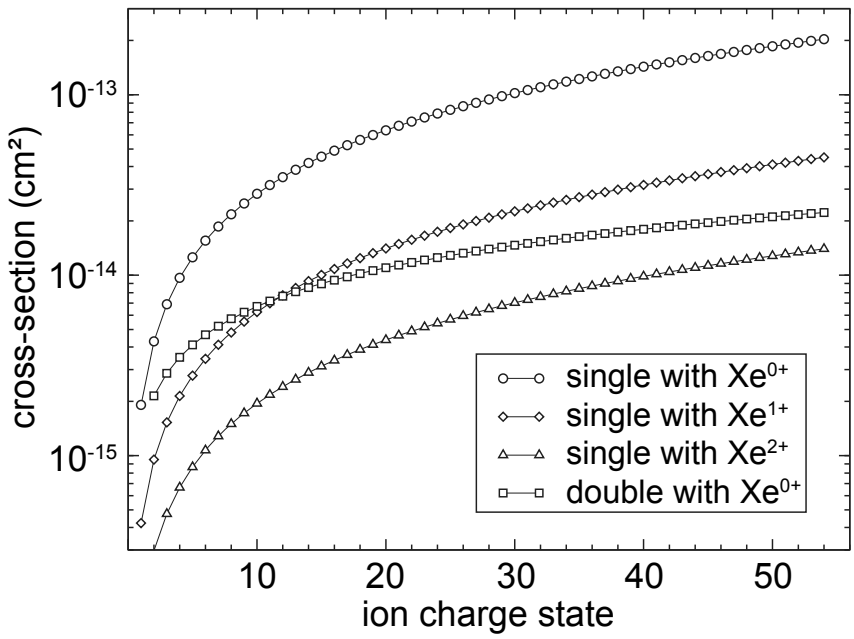


Fig. 8: Charge exchange cross-sections for single charge exchange with $\mathrm{Xe}^{(0,1,2)}$ ions as well as for the double charge exchange with xenon neutrals (after Ref. [23]).

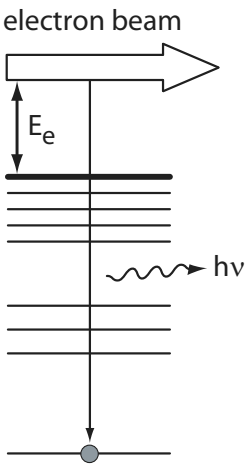


Fig. 9: Scheme of the radiative recombination process. The energy $\hbar\omega$ of the emitted photon is calculated as $\hbar\omega = E_\mathrm{e} - E_\mathrm{B}$, where E_B is the electron binding energy.

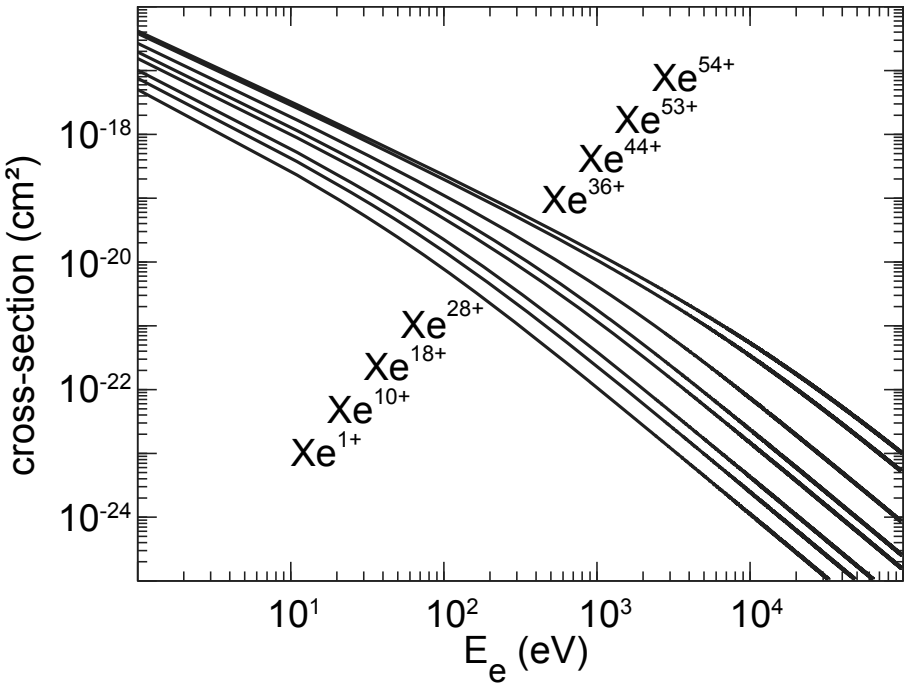


Fig. 10: Cross-sections for the radiative recombination on xenon ions (according to Ref. [23])

Since RR processes counteract the ionization for each element and ion charge state, there exists a characteristic electron energy at which the ionization rate is equal to the RR rate. This behaviour is discussed in more detail in Ref. [25].

3.2.5 Further processes

Besides the main processes discussed above, in special cases some further contributions from

- dielectronic recombination (DR),
- three-body recombination,
- vacancy cascades and
- photoionization

are of interest, but they will not be discussed here.

3.3 Balance equation for the ion charge-state distribution

The number of ions with charge q produced from ions with charge $(q-1)$ and ionized further to the charge state $(q+1)$ can be described by a balance equation for successive electron impact ionization and competing processes:

$$\begin{aligned}\frac{\mathrm{d}n_q}{\mathrm{d}t} &= R^{\mathrm{ion}}_{q-1\to q} - R^{\mathrm{ion}}_{q\to q+1} + R^{\mathrm{RR}}_{q+1\to q} - R^{\mathrm{RR}}_{q\to q-1} \\ &\quad + R^{\mathrm{exch}}_{q+1\to q} - R^{\mathrm{exch}}_{q\to q-1} - R^{\mathrm{ax,esc}}_{q} - R^{\mathrm{rad,esc}}_{q} + R^{\mathrm{source}},\end{aligned} \tag{14}$$

with ionization terms R^{ion}, radiative recombination terms R^{RR}, charge exchange terms R^{exch}, axial ion loss terms $R^{\mathrm{ax,esc}}$, radial ion loss terms $R^{\mathrm{rad,esc}}$ and particle source term R^{source}.

If the ionization process occurs in single-step processes and if we neglect the axial loss (which is a good approximation for appropriate field configurations) this yields

$$
\begin{aligned}
\frac{\mathrm{d}n_0}{\mathrm{d}t} &= \underbrace{-n_0\lambda_{0,1}}_{\text{ionization}} + \underbrace{n_1\lambda_{1,0}}_{\text{charge exchange}} \quad , \\
\frac{\mathrm{d}n_1}{\mathrm{d}t} &= n_0\lambda_{0,1} - n_1(\lambda_{1,2} + \lambda_{1,0}) + n_2\lambda_{2,1} - \left[\frac{\mathrm{d}n_1}{\mathrm{d}t}\right]^{\mathrm{rad,esc}} \quad , \\
&\vdots \\
\frac{\mathrm{d}n_q}{\mathrm{d}t} &= n_{q-1}\lambda_{q-1,q} - n_q(\lambda_{q,q+1} + \lambda_{q,q-1}) + n_{q+1}\lambda_{q+1,q} - \left[\frac{\mathrm{d}n_q}{\mathrm{d}t}\right]^{\mathrm{rad,esc}} \quad , \\
\frac{\mathrm{d}n_z}{\mathrm{d}t} &= n_{Z-1}\lambda_{Z-1,Z} - n_Z\lambda_{Z,Z-1} - \left[\frac{\mathrm{d}n_Z}{\mathrm{d}t}\right]^{\mathrm{rad,esc}} \quad .
\end{aligned}
\tag{15}
$$

Here $n_0, \ldots, n_Z$ are the densities for neutrals and ions of the charge state q. Further we have the ionization coefficients $\lambda_{q,q+1}$ and the recombination and charge exchange coefficients $\lambda_{q,q-1}$. Details about the corresponding terms, the influence on the development of the individual ion densities, the development of the energy densities and ion evaporation of ions from the ionization volume can be found, for example, in Refs. [26] and [27].

In Fig. 11, as a typical example we show the evolution of xenon ion charge states in the electron beam of a Dresden EBIT as it follows from equation system (14).

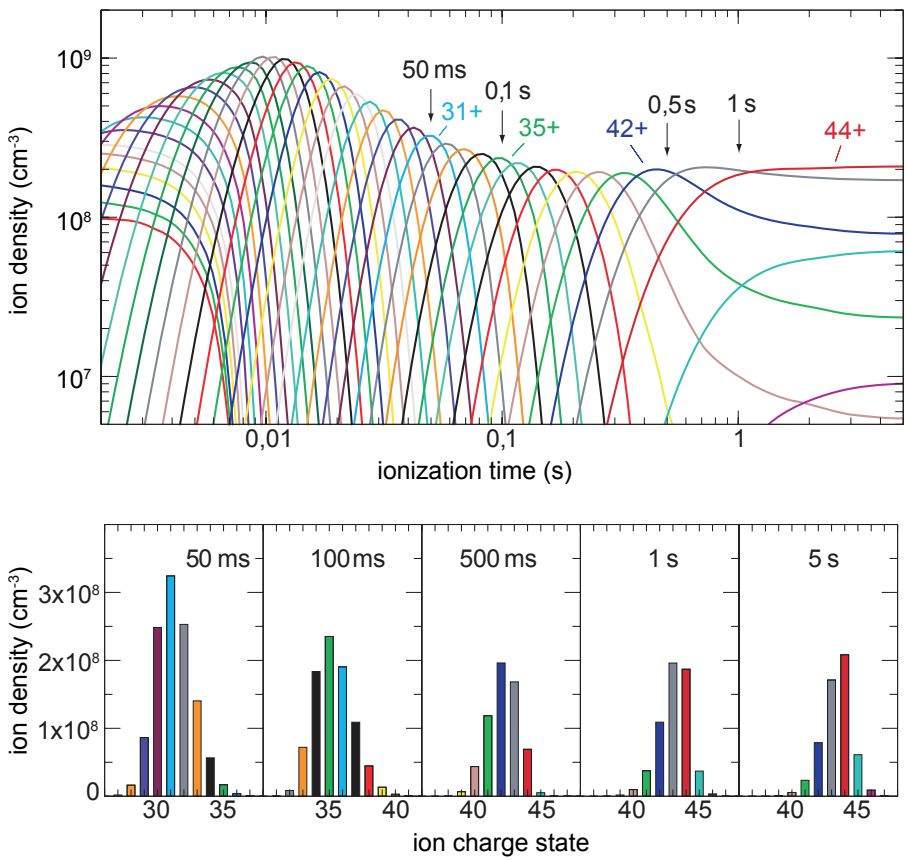


Fig. 11: Evolution of the ionization of xenon ions in a Dresden EBIT at $E_\mathrm{e} = 15\,\mathrm{keV}$, $I_\mathrm{e} = 40\,\mathrm{mA}$ and $p = 2 \times 10^{-9}$ mbar.

3.4 Ion source physics: selected topics

Owing to the restricted scope of this chapter, only some information about the important (but not complete) details of EBISs are given. For more detailed studies, please refer to the corresponding literature. For reasons of simplicity and copyright, in the following the author will give examples of results from EBISs of the Dresden EBIS type.

3.4.1 Electrical trap capacity

The ion trapping in an EBIS (see Fig. 12) occurs

- in the axial direction by electrostatic potentials and
- in the radial direction by the space-charge potential of the electron beam.

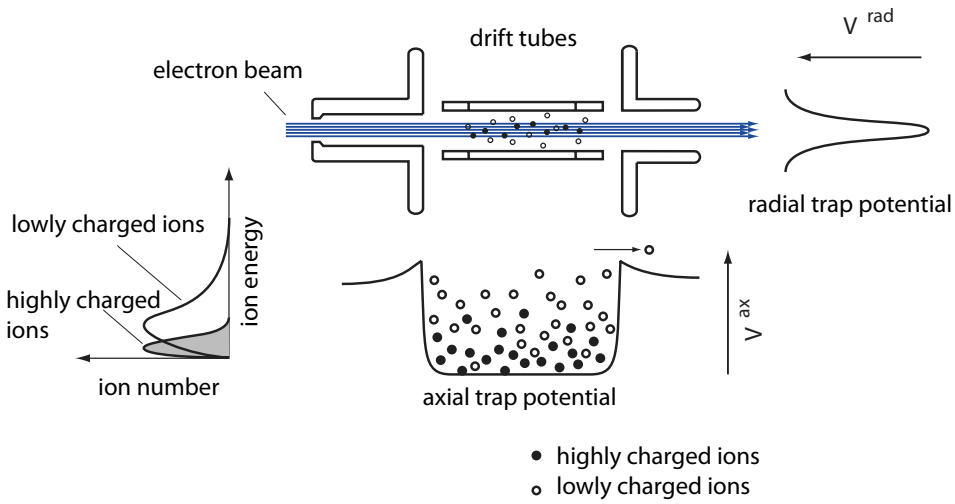


Fig. 12: Axial and radial potentials in an EBIT. Owing to the effective trap height according to $V_\mathrm{b} \times q$, highly charged ions are trapped more strongly than low charged ones. The energy exchange between the ions leads to an equilibrium distribution of the ion energy.

The total number of ions that can be stored in an EBIS is limited by the electrical trap capacity C_el. For an estimation of C_el, we assume a homogeneous electron beam that passes through an ion trap of length L with the electron beam current I_e and electron energy E_e. Then the number of negative charges in the beam volume is a measure for the trap capacity C_el. With

$$I_\mathrm{e} = \frac{\mathrm{d}Q}{\mathrm{d}t}, \qquad v_\mathrm{e} = \frac{\mathrm{d}x}{\mathrm{d}t}, \qquad v_\mathrm{e} = \sqrt{\frac{2E_\mathrm{e}}{m_\mathrm{e}}},$$

we have

$$\Delta Q = \frac{I_\mathrm{e}\Delta x}{v_\mathrm{e}} = \frac{I_\mathrm{e}L}{\sqrt{2E_\mathrm{e}/m_\mathrm{e}}}$$

and for the trap capacity

$$C_\mathrm{el} = 1.05 \times 10^{13}\,\frac{I_\mathrm{e}\,(\mathrm{A})\,L\,(\mathrm{m})}{\sqrt{E_\mathrm{e}\,(\mathrm{eV})}}. \tag{16}$$

In order to estimate the useable trap capacity, in practice we must supplement Eq. (16) to

$$C_\mathrm{el} = 1.05 \times 10^{13}\,\frac{I_\mathrm{e}\,(\mathrm{A})\,L\,(\mathrm{m})}{\sqrt{E_\mathrm{e}\,(\mathrm{eV})}}\,\alpha f, \tag{17}$$

with α the ratio of ions with a certain ion charge state in the ion charge-state distribution and f the charge compensation of the electron beam.

3.4.2 Boltzmann distribution of the ion energy

In the source volume ion–ion collisions lead very rapidly (in about milliseconds) to a Boltzmann distribution of the ion energy

$$f(E_\mathrm{i}) = \frac{2}{\sqrt{\pi}\,kT_\mathrm{i}}\sqrt{\frac{E_\mathrm{i}}{kT_\mathrm{i}}}\exp\left(-\frac{E_\mathrm{i}}{kT_\mathrm{i}}\right). \tag{18}$$

The formation of a Boltzmann energy distribution has consequences for the whole ion trap process:

– *Permanent ion losses.* At a certain mean ion energy, ions always exist with an energy greater than a critical energy

$$E_\mathrm{cr} = qU_\mathrm{b}, \tag{19}$$

and can leave the trap with the barrier U_b. This means that we have a constant ion loss from the trap. The ion loss ratio can be derived by integrating the distribution function. The behaviour of ions in the trap can be described by the non-stationary Boltzmann equation in a self-consistent model considering ionization and energy exchange processes.

– *Evaporative cooling* of multiply charged ions by light LCIs. Elastic collisions between ions with different charge states and masses lead to an equilibrium energy distribution for each ion sort. Thereby the critical energy for leaving the ion trap for LCIs is lower than for highly charged ions, i.e. the ion lifetime in the ion trap is lower for LCIs. LCIs leave the beam first, i.e. they evaporate and transport the energy transmitted from the highly charged ions to lower charged ones out of the ionization volume. As a result of the process, the mean temperature of the highly charged ions decreases and ion losses of these ions are reduced.

3.4.3 Radial trap potential

For simplicity, we consider an ion trap with a radial symmetric electron beam the electron density of which is homogeneously distributed over the radius r_e of the electron beam. Using Gauss's theorem the radial trap potential has the shape

$$V_\mathrm{e}(r) = \begin{cases} \dfrac{U_\mathrm{e} r^2}{r_\mathrm{e}^2} & \text{for } r < r_\mathrm{e}, \\ U_\mathrm{e}\left(2\ln\dfrac{r}{r_\mathrm{e}} + 1\right) & \text{for } r > r_\mathrm{e}. \end{cases} \tag{20}$$

Here we have

$$U_\mathrm{e} = \frac{I_\mathrm{e}}{4\pi\varepsilon_0 v_\mathrm{e}} = \frac{1}{4\pi\varepsilon_0}\sqrt{\frac{m_\mathrm{e}}{2}}\frac{I_\mathrm{e}}{\sqrt{E_\mathrm{e}}}. \tag{21}$$

We can estimate U_e with

$$U_\mathrm{e}\,(\mathrm{V}) = \frac{30I\,(\mathrm{A})}{\sqrt{1-\left(\dfrac{E_\mathrm{e}\,(\mathrm{keV})}{511}+1\right)^{-2}}}. \tag{22}$$

As an example the radial trap potential in Fig. 13 is shown for a typical operation regime of a Dresden EBIT.

The ion confinement is additionally influenced by the magnetic field enclosing the electron beam, so that we finally have an effective radial potential of

$$V_\mathrm{e}^\mathrm{eff}(r) = V_\mathrm{e}(r) + \frac{eq_\mathrm{i}B^2r^2}{8m_\mathrm{i}}, \tag{23}$$

with R the drift tube radius.

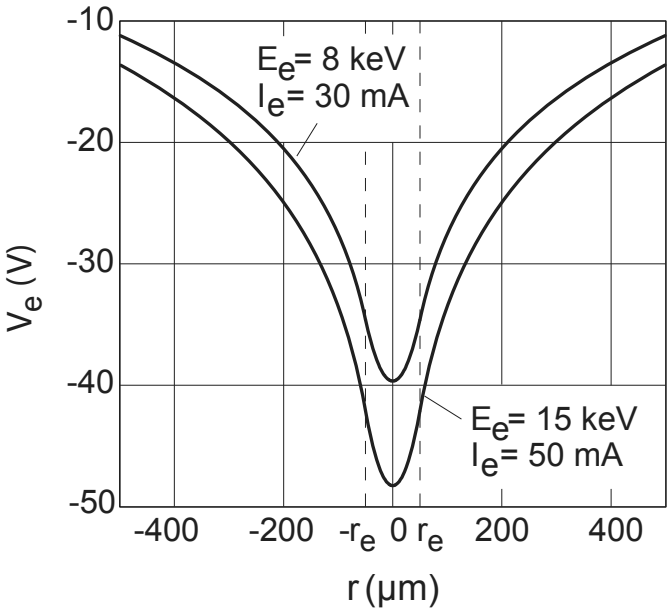


Fig. 13: Radial trap potential of a Dresden EBIT for two different electron beam currents

3.4.4 Ion heating and energy balance

Ions are heated in the electron beam by elastic Coulomb collisions and by microwave instabilities. If instabilities do not appear (huge Debye lengths), the main heating processes are elastic electron–ion collisions.

3.4.4.1 Elastic electron–ion collisions

Ion heating by elastic electron collisions can be described by

$$\left[\frac{\mathrm{d}E_\mathrm{i}}{\mathrm{d}t}\right]^{\mathrm{heat}} = (n_\mathrm{i}\sigma_\mathrm{i}v_\mathrm{e})\left(2\frac{m_\mathrm{e}}{m_\mathrm{i}}E_\mathrm{e}\right) = 2\frac{m_\mathrm{e}}{m_\mathrm{i}}E_\mathrm{e}f_\mathrm{ei}, \tag{24}$$

with

$$t_\mathrm{ei} = f_\mathrm{ei}^{-1} = \frac{1}{n_\mathrm{i}\langle v_\mathrm{e}\sigma_\mathrm{i}\rangle}$$

and σ_i the Coulomb collision cross-section

$$\sigma_\mathrm{i} = \frac{1}{(4\pi\varepsilon_0)^2}4\pi\left(\frac{q_\mathrm{i}e^2}{m_\mathrm{e}}\right)^2\frac{\ln\Lambda_\mathrm{ei}}{v_\mathrm{e}^4}.$$

The Coulomb logarithm for electron–ion collisions $\ln\Lambda_\mathrm{ei}$ has the shape

$$\ln\Lambda_\mathrm{ei} = \begin{cases} 23 - \ln\dfrac{q_\mathrm{i}\sqrt{n_\mathrm{e}}}{(kT_\mathrm{e})^{3/2}} & \text{for } q_\mathrm{i}^2\times 10\,\mathrm{eV} > kT_\mathrm{e} > \dfrac{m_\mathrm{e}}{m_\mathrm{i}}kT_\mathrm{i}, \\ 24 - \ln\dfrac{\sqrt{n_\mathrm{e}}}{(kT_\mathrm{e})} & \text{for } kT_\mathrm{e} > q_\mathrm{i}^2\times 10\,\mathrm{eV} > \dfrac{m_\mathrm{e}}{m_\mathrm{i}}kT_\mathrm{i}, \\ 30 - \ln\dfrac{q_\mathrm{i}^2\sqrt{n_\mathrm{e}}}{(kT_\mathrm{i})^{3/2}A_\mathrm{i}} & \text{for } kT_\mathrm{e} < \dfrac{m_\mathrm{e}}{m_\mathrm{i}}q_\mathrm{i}kT_\mathrm{i}, \end{cases} \tag{25}$$

with kT the energy (eV), n the density (cm^{-3}), A_i the atomic mass and k the Boltzmann constant. Energy transfer by Coulomb collisions then occurs with a rate

$$\left[\frac{\mathrm{d}(n_\mathrm{i}kT_\mathrm{i})}{\mathrm{d}t}\right]^{\mathrm{heat}} \approx \left(-\frac{\mathrm{d}E_\mathrm{e}}{\mathrm{d}t}\right)n_\mathrm{e}. \tag{26}$$

3.4.4.2 Elastic ion–ion collisions

Energy transfer between ions occurs by elastic ion–ion collisions. In these collisions energetic ('hot') ions are cooled by colder ions. If we assume a Boltzmann distribution of the ion energy, the cooling process can be characterized by

$$\left[\frac{\mathrm{d}E}{\mathrm{d}t}\right]^{\mathrm{exch}} = 2\nu_{ij} n_i \frac{m_i}{m_j} \frac{kT_j - kT_i}{\left(1 + \frac{m_i T_j}{m_j T_i}\right)^{3/2}}. \tag{27}$$

Here the ν_{ij} are Coulomb collision rates between the ions

$$\nu_{ij} = \frac{1}{(4\pi\varepsilon_0)^2} \frac{4\sqrt{2\pi}}{3} n_j \left(\frac{q_i q_j e^2}{m_i}\right)^2 \left(\frac{m_i}{kT_i}\right)^{3/2} \ln \Lambda_{ij}, \tag{28}$$

with Λ_{ij} as the Coulomb logarithm for ion–ion collisions

$$\ln \Lambda_{ij} = \ln \Lambda_{ji} = 23 - \ln \left[\frac{q_i q_j (m_i + m_j)}{m_i T_j + m_j T_i} \left(\frac{n_i q_i^2}{T_i} + \frac{n_j q_j^2}{T_j}\right)^{1/2}\right]. \tag{29}$$

The energy transfer by ion–ion collisions is applied for evaporative cooling where low charged ions cool down more highly charged ions. The the rest gas in the source acts as the source for the LCIs. Because ion–ion collisions occur very efficiently, the ions tend to thermal equilibrium. The energy that the ions have on leaving the ion trap is then proportional to the ion charge state. It is important to mention that there exists an upper limit for the neutral gas density because otherwise the number of highly charged ions will be reduced dramatically by charge exchange processes.

The energy that is transferred out of the ion source by the evaporation process can be described by the cooling rate

$$\frac{\mathrm{d}E_0}{\mathrm{d}t} \left(\frac{\mathrm{eV}}{\mathrm{s\,cm^3}}\right) = \frac{\mathrm{d}n_{\mathrm{LCI}}\,(\mathrm{cm^{-3}})}{\mathrm{d}t} U_{\mathrm{b}}\,(\mathrm{V})\, q_{\mathrm{LCI}}, \tag{30}$$

with n_{LCI} the equilibrium density of the LCIs and q_{LCI} the ion charge state of the LCIs.

3.4.5 Electron beam: equation of motion and beam radius

The equation of motion for the electron beam in the paraxial approximation can be written as

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{e I_{\mathrm{e}}}{2\pi\varepsilon_0 v_z r m_{\mathrm{e}}} + \frac{e^2}{4m_{\mathrm{e}}^2} \left(\frac{B_{\mathrm{c}}^2 r_{\mathrm{c}}^4}{r^3} - B_z^2 r\right), \tag{31}$$

with B_{c} the B field at the cathode, B_z the axial magnetic field, r_{c} the cathode radius, I_{e} the electron beam current and v_z the axial electron velocity.

If we assume $B_{\mathrm{c}} = 0$ then there exists a stationary solution of Eq. (31). The solution corresponds to an equilibrium flow of electrons with constant radius, the so-called Brillouin flow (index 'B'). For a Brillouin flow, all electrons have a constant distance to the beam centre. Thereby the Lorentz force caused by the magnetic field is compensated by the space charge and the centrifugal force of the rotating electrons.

If we consider more realistic assumptions, such as

- a magnetic field at the cathode,
- thermal effects at the cathode due to filament heating up to the temperature T_{c} and
- interactions between the electrons,

then this leads to a corrected electron beam radius (smaller than r_{B}) as described by Herrmann [28]. According to Herrmann's theory the electron beam radius enclosing 80% of the beam is calculated [29] to be

$$r_{\mathrm{e}} = r(0)\sqrt{\left(1-\frac{r_0}{r(0)}\right)^2 + \frac{2}{1+(B_{\mathrm{c}}^2 r_{\mathrm{c}}^4/B_z^2 r_0^4)}\left(\frac{v_{\mathrm{e}}\tan\gamma}{(e/m_{\mathrm{e}})B_z\gamma}\right)}, \tag{32}$$

and

$$r_0 = r_{\mathrm{B}}\left\{\frac{1}{2}+\frac{1}{2}\left[1+4\left(\frac{8kT_{\mathrm{c}}r_{\mathrm{c}}^2 m_{\mathrm{e}}}{e^2 B_z^2 r_{\mathrm{B}}^4}+\frac{B_{\mathrm{c}}^2 r_{\mathrm{c}}^4}{B_z^2 r_{\mathrm{B}}^4}\right)\right]^{1/2}\right\}^{1/2}, \tag{33}$$

with $r(0)$ the beam radius at the cathode and γ the angle deviation from the source axis.

Generally, the electron beam diameter of an EBIT depends on the actual technical solution. Typical diameters are of the order of 40 µm up to about 200 µm. A classical method to measure the electron beam diameter in an EBIS is shown in Fig. 14 and is discussed in more detail in Ref. [30]. The idea is to measure the X-rays emitted from ions in the electron beam that are ionized or excited by electron impact. Through an aperture the X-rays are detected on a charge-coupled device (CCD) chip and the evaluation of the measured intensity distributions allows the electron beam diameter to be determined.

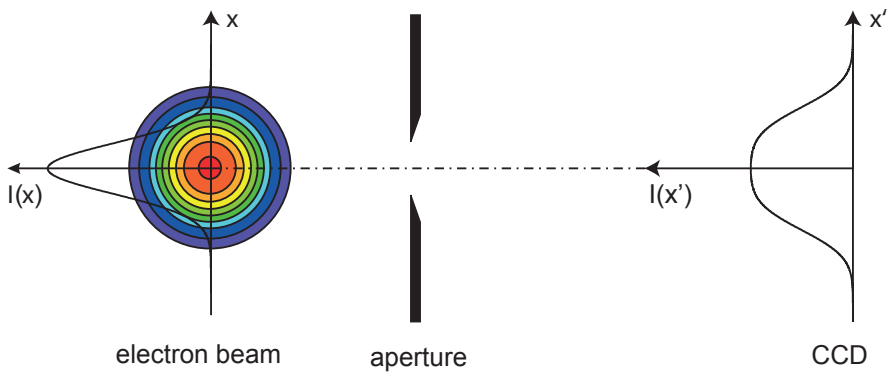


Fig. 14: CCD chip imaging of X-rays emitted from the electron beam (after Ref. [30]). Here $I(x)$ is the X-ray intensity distribution from the electron beam, and $I(x')$ is the intensity distribution on the CCD.

4 Ion extraction from EBISs

4.1 Basic ion extraction techniques

As mentioned before, EBISs allow for production of ion beams in DC as well as pulsed mode. In the following, we will discuss and characterize some basic features of EBIS ion extraction. If ions from an EBIS are applied in different experiments, a basic problem is the separation of an individual ion charge state from the extracted ions, which primarily contain a spectrum of individual ion charge states. Generally, the ion charge-state separation is realized by

- magnetic dipole analyzer or
- Wien filters.

4.2 Ion beam separation using an analysing magnet

A typical experimental set-up for experiments with extracted ions and magnetic analysis of the ion charge states is shown in Fig. 15. The facility shown was designed for experiments with highly charged ions for ion–solid interactions, atomic physics, biophysics, plasma physics, materials analysis, solid-state physics, etc. The ion source as well as the ion beamline can be insulated from ground potential. Thus ions with a wide range of kinetic energies (decelerated as well as accelerated) can be used for experiments in the target chamber of the facility, which can be subdivided into four functional units, the ion source (e.g. Dresden EBIS-A), the first and second beamline segments, and the target chamber.

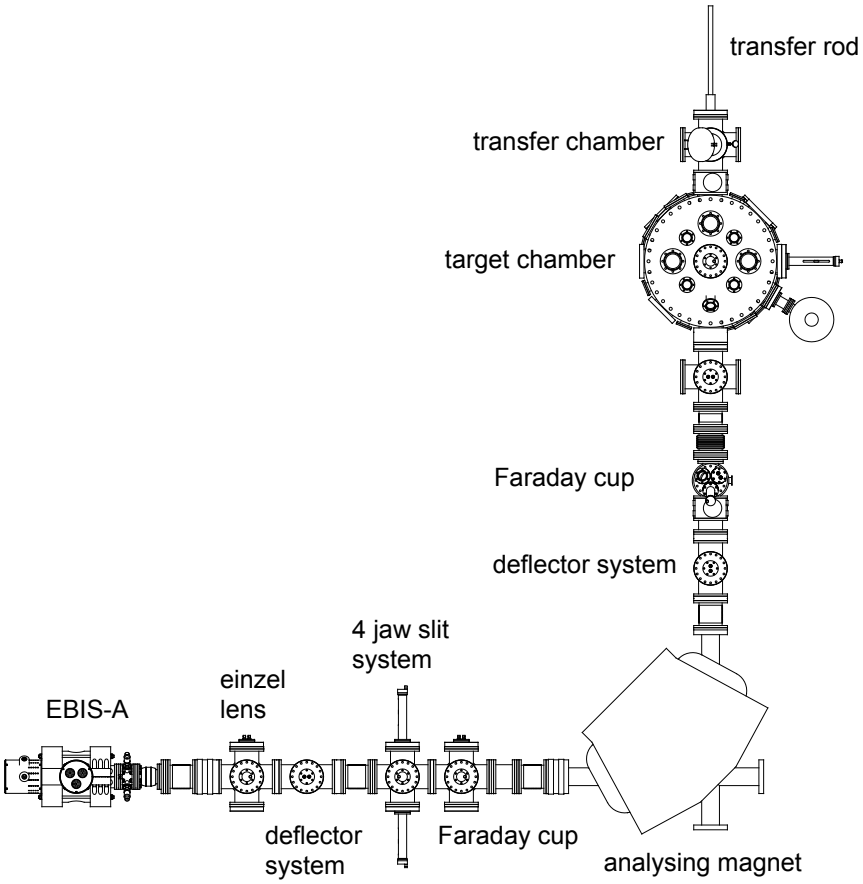


Fig. 15: Footprint of an ion beam facility featuring a Dresden EBIS-A including a 90° bending magnet for mass-to-charge ratio separation of the ion beam extracted from the EBIS.

An einzel lens and an x–y deflector as well as an aperture, a four-jaw slit system and a Faraday cup are integrated in the first beamline segment. The einzel lens and the x–y deflector are used to focus, deflect and align the ion beam. An aperture with variable diameters can be used, reducing the phase space of the ion beam and therefore improving the beam quality at the expense of beam intensity. A Faraday cup is used for detecting and measuring the ion beam current. It is mounted onto a translational feedthrough directly behind the four-jaw slit system, which is positioned in the focal point of the dipole magnet. Thus the determination of the beam position as well as the beam size is possible.

The ion beam extracted from the ion source can consist of ions of various elements and ion charge states. Thus a dipole magnet separates the ions according to their momentum in the second beamline segment. The momentum of the ion depends on the charge state and the mass of the ion. Hence ions with a certain charge-to-mass ratio can be selected.

The momentum-separated ions are aligned and deflected with a second x–y deflector positioned behind the dipole magnet. A second Faraday cup mounted in the focal point downstream of the dipole magnet ensures detection and current measurement of the ion beam using an electrometer.

The last functional unit of the facility consists of a deceleration lens and the target chamber. The ions are focused into the target chamber by the deceleration lens system. In principle, it is possible to decelerate the ions down to a few electronvolts as well as to accelerate them up to some tens of keV (40 keV) per charge state. In the example shown, the target chamber has two stages, an upper one for target irradiation experiments and a lower one for target preparation and target transfer. The upper stage is mounted onto the deceleration lens system. The target can be rotated (360°) and transferred via a sample holder mounted to a four-axes manipulator. The target can be heated by resistance heating as well as electron bombardment and cooled down via liquid nitrogen. With a 'wobble stick' a further target manipulation inside the target chamber is possible. For measuring the beam intensity in the target chamber, a third Faraday cup is mounted onto the four-axes manipulator below the sample holder. A large number of vacuum ports allow the installation of various devices for experiments.

To characterize ion extraction spectra with an analysing magnet, some examples for ion extraction spectra are given as they can be expected for EBISs. Figure 16 shows a xenon ion extraction spectrum as measured with a Dresden EBIS-A for an ionization time of 5 s. In Fig. 17 a carbon ion extraction spectrum is shown as measured with a superconductiong Dresden EBIS-SC for an ionization time of 200 ms (for details see Ref. [31]).

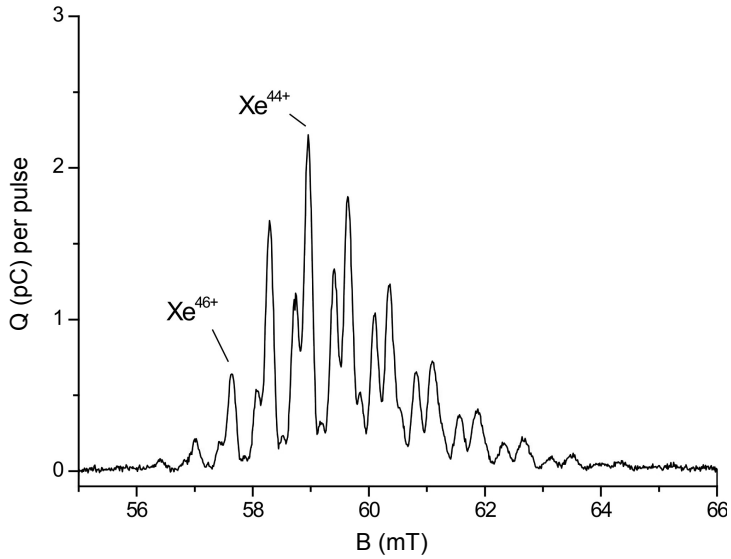


Fig. 16: A xenon ion extraction spectrum measured at a Dresden EBIS-A for a vacuum of 5×10^{-10} mbar, an electron beam energy of 12 keV and an electron beam current of 100 mA.

4.3 Ion beam separation using a Wien filter

A Wien filter is a particle separator with crossed magnetic and electric field configuration providing mass- and charge-state-separated beams of excellent quality and small spatial dimensions [32]. With the Wien filter as an ion source add-on, a very compact device is available substituting a complete standard beamline set-up with a space-consuming dipole magnet and other necessary equipment (see Fig. 18). The Wien filter can be used as a stand-alone solution for various beamline set-ups. Depending on the installed mass- and charge-state-separating aperture, a resolution of better than 80 is available. The geometrical dimensions of the Wien filter are small compared to a classical dipole magnet. The Wien filter fits easily on a beamline, making a change of the ion beam direction, as is necessary for a dipole magnet, obsolete.

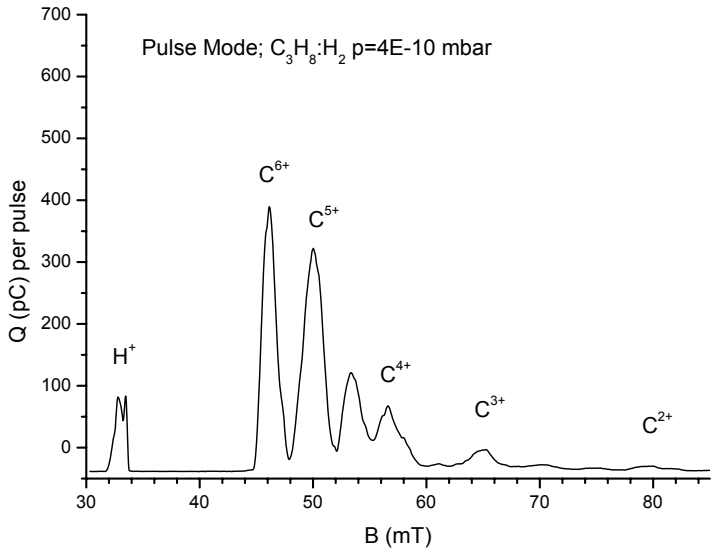


Fig. 17: A carbon ion extraction spectrum measured at a Dresden EBIS-SC for a vacuum of 4×10^{-10} mbar, an electron beam energy of 12 keV, an electron beam current of 400 mA and an ionization time of 200 ms.

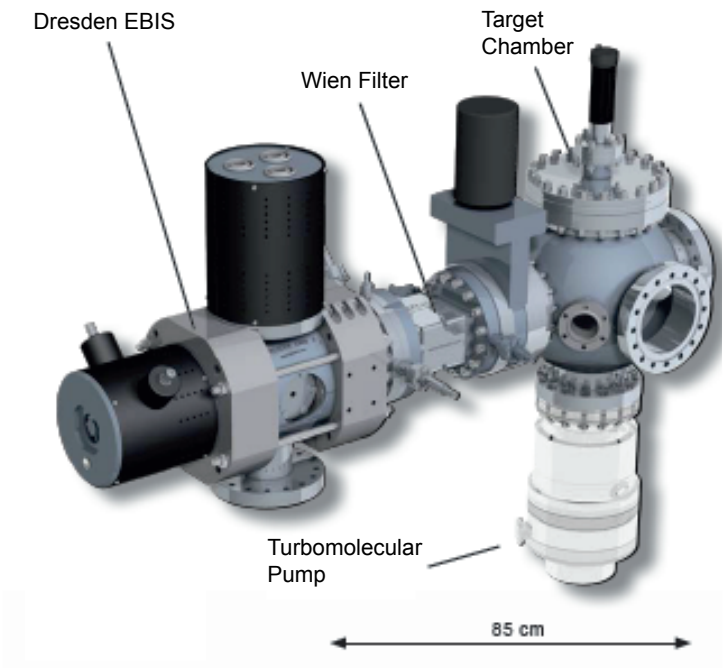


Fig. 18: A 3D presentation of a Wien filter-based beamline

The use of a Wien filter allows for separation of light and intermediate-Z ions of all ion charge states. For hydrogen and xenon, this is shown in Figs. 19 and 20. The most important difference to the use of analysing magnets is that a magnet has a gap of several centimetres, but for obtaining high resolution with a Wien filter we must work with aperture diameters in the millimetre or sub-millimetre region, i.e. beam losses through the Wien filter have to be taken into account.

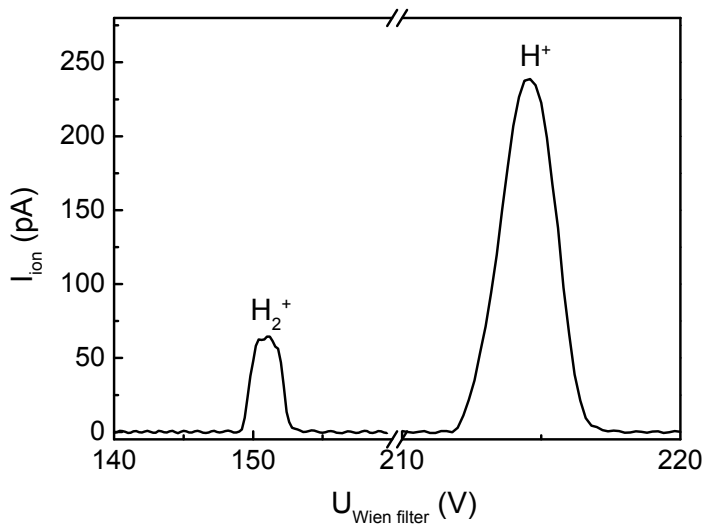


Fig. 19: A leaky mode hydrogen ion extraction spectrum measured at a Dresden EBIS for a vacuum of 3×10^{-9} mbar, an electron beam energy of 13.6 keV and an electron beam current of 30 mA.

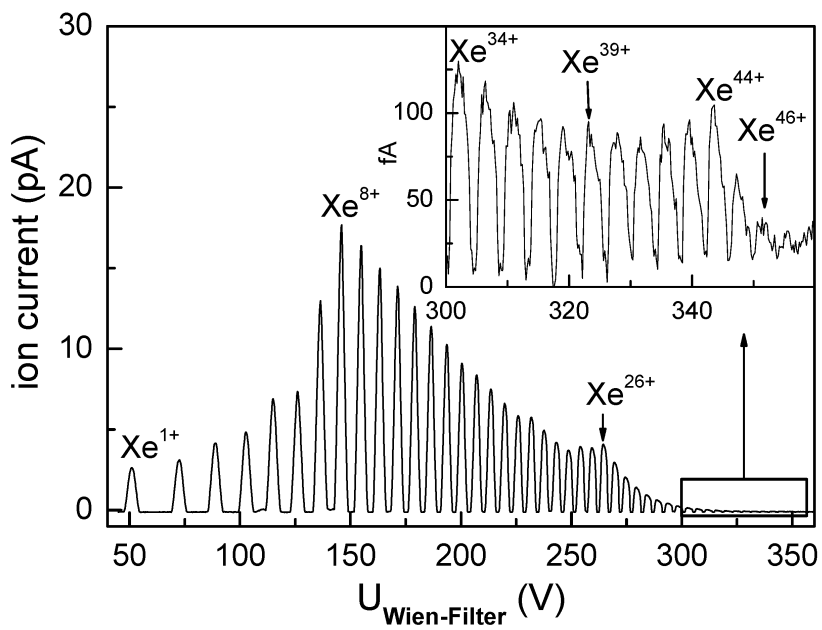


Fig. 20: A leaky mode xenon ion extraction spectrum measured at a Dresden EBIS for a vacuum of 3×10^{-9} mbar, an electron beam energy of 13.6 keV and an electron beam current of 30 mA.

4.4 Beams of singly charged molecular fragments and clusters

In leaky mode, EBISs can be used as sources for low charged ions from molecular fragments. The production of molecular fragments in an EBIS is of interest for various applications. This concerns issues such as studies on electronic transitions in molecules by spectroscopy of emitted electromagnetic radiation, protonation studies on complex molecules, as well as investigations on the interaction of molecular fragments with surfaces and on free molecules or gases in a corresponding gas jet. The use of a gas target may allow the production of neutral beams of different molecular fragments by charge exchange processes. Figure 21 shows an ion extraction from a Dresden EBIT after injection of propane. Thereby all possible stoichiometric ratios of propane fragments were detected. More details about molecule fragmentation in EBISs can be found in Ref. [33].

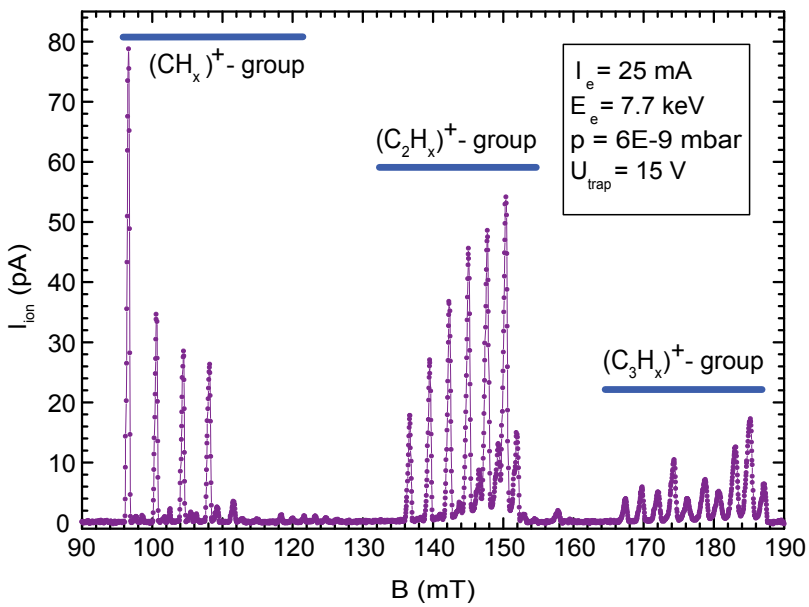


Fig. 21: Molecular fragment spectrum in leaky mode extracted from the Dresden EBIT after injection of propane

Additionally, one can also produce more complex structures such as metal clusters. In Fig. 22 this is shown for copper clusters as produced with a Dresden EBIT. The copper was introduced into the ion source by sputtering from a copper surface under gold ion irradiation. The capability of this technique is seen if we note that clusters with different isotope ratios can be clearly resolved, such as $[(^{63}\mathrm{Cu})_3]^{1+}$, $[(^{63}\mathrm{Cu})_2(^{65}\mathrm{Cu})_1]^{1+}$, $[(^{63}\mathrm{Cu})_1(^{65}\mathrm{Cu})_2]^{1+}$ and $[(^{65}\mathrm{Cu})_3]^{1+}$. Without a great deal of effort, cluster beam currents of up to 10^5 clusters per second were measured.

4.5 Absolute numbers of extracted ions

To characterize the potential of different EBIS types, we give some absolute ion numbers for pulsed ion extraction of ion sources of the Dresden EBIS type in Table 3. The summarized numbers characterize the expected order of magnitude for ion extraction but can deviate in their dependence on the individual ion source and the selected operation regime.

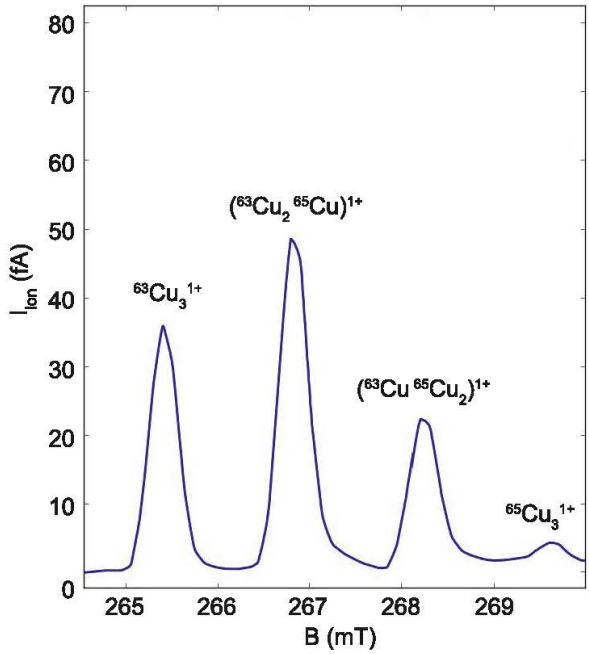


Fig. 22: Measured copper clusters from a Dresden EBIT at an electron beam energy of 6 keV and an electron beam current of 23 mA.

Table 3: Ions per pulse for pulsed ion extraction of ion sources of the Dresden EBIS type.

Ion	Dresden EBIT	Dresden EBIS	Dresden EBIS-A	EBIT : EBIS : EBIS-A
C^{6+}	10 000 000	30 000 000	400 000 000	1 : 3 : 40
Ar^{16+}	900 000	8000 000	250 000 000	1 : 9 : 280
Ar^{18+}	6000	90 000	1500 000	1 : 15 : 250
Xe^{44+}	10 000	700 000	10 000 000	1 : 70 : 1000

4.6 Pulse shape for typical ion extraction

Typical pulse widths at pulsed ion extraction from an EBIS are of the order of a few microseconds. Thereby the pulses are not completely symmetric but have a tail on the side of higher flight times. The reason is that, at the moment of trap opening, only those ions are extracted that have a flight direction towards the extraction side. All other ions will be reflected at the potential wall at the gun side and extracted after that. This situation is shown in Fig. 23. The absolute pulse width is determined by the speed of switching the trap potential at the extraction side ('slew rate'). We see that we can influence the pulse width from some microseconds up to about 100 µs, i.e. we come near to 'flat-top' pulses. As an example this is shown in Fig. 24 for ion pulses from a Dresden EBIS-A with a trap length L of 6 cm and different slew rates.

4.7 Short-time ion pulse extraction

For some applications the time structure of the extracted beams of HCIs is of fundamental interest. For example, for applications in time-of-flight secondary ion mass spectroscopy (TOF-SIMS) and in some solid-state physics experiments, the formation of short pulses up to the nanosecond range is of interest. Present studies prove that a short pulsed ion extraction from the Dresden EBIT is possible. The extracted

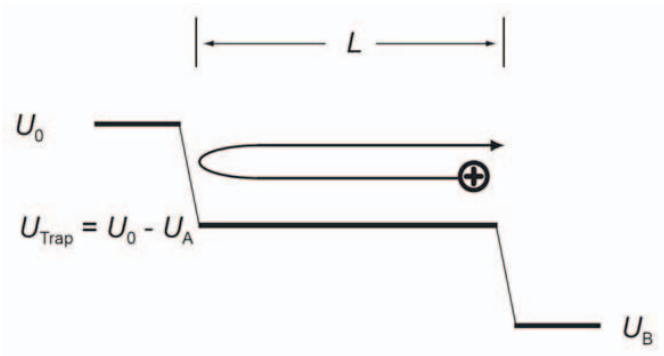


Fig. 23: Ion paths at the moment of ion extraction

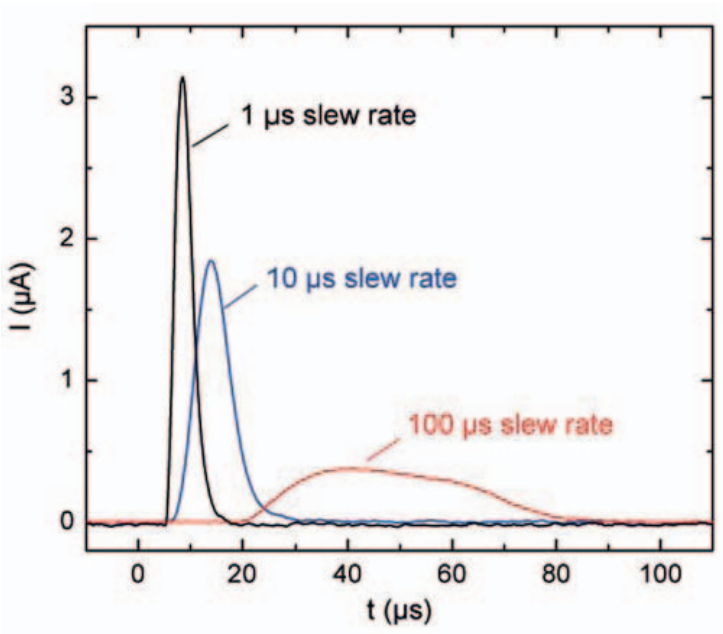


Fig. 24: Carbon ion pulses from a Dresden EBIS-A for different slew rates

ion charges per pulse are in the range of some femtocoulombs and it is possible to extract several pulses until the trap has to be reloaded. The maximum extracted ion charge depends significantly on the chosen extraction time. As an example we show in Fig. 25 a plot of extracted ion charges for an Ar^{16+} pulse as a function of the extraction time t_{extr}. From Fig. 25 it is seen that, after some hundreds of nanoseconds, the extracted ionic charge increases no further.

4.8 Ion beam emittance

For applications of ion beams, a quantitative measure of the quality of the beam is given by the root-mean-square (r.m.s.) emittance

$$\varepsilon_{x,\mathrm{rms}} = \sqrt{\overline{x^2}\,\overline{x'^2} - \overline{x'^2}} \tag{34}$$

in x–x' trace space and equivalently in the case of y–y'.

A common method to determine these ion beam parameters is the pepperpot technique, which can be traced back to works of the 1960s [35]. A scheme of a pepperpot emittance meter is shown in Fig. 26. The incoming ion beam passes the pepperpot mask and is separated into several beam spots. The particles hitting the microchannel plate (MCP) create an electron current, which is amplified passing

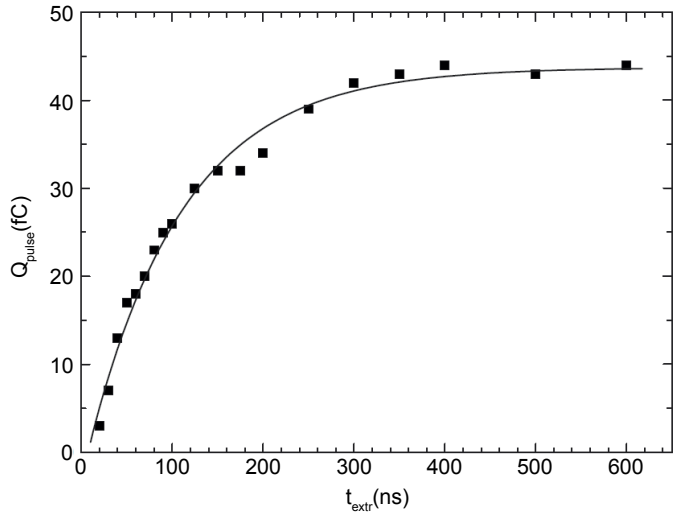


Fig. 25: Extracted ion charges of Ar^{16+} pulses from a Dresden EBIT as a function of the extraction time t_{extr} (after Ref. [34]).

the two MCPs. The electrons are then accelerated towards the phosphor screen. The visible light spots created at the phosphor screen are detected after 90° deflection by a CCD camera. The emittance of the ion beam can be determined from the positions, sizes and shapes of the light spots.

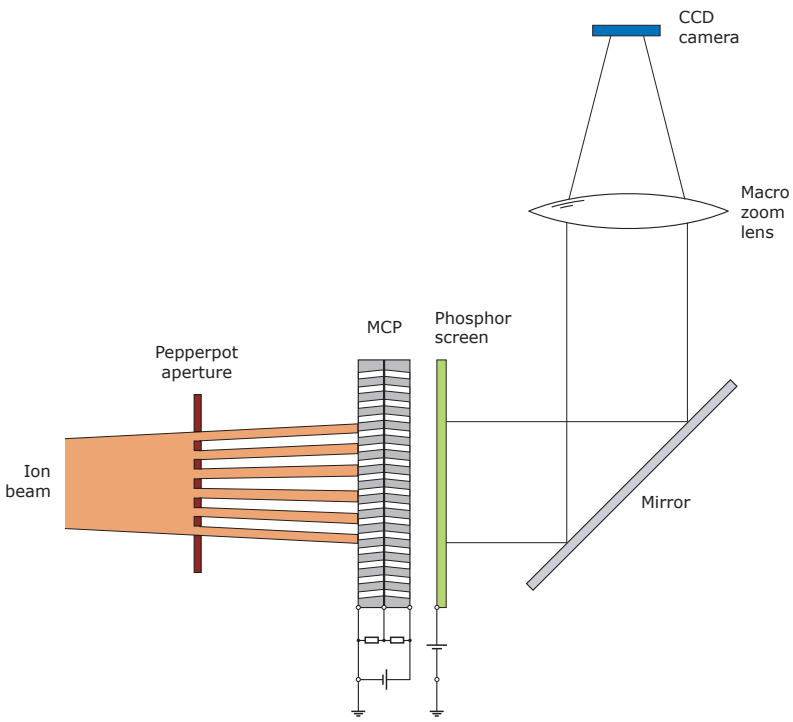


Fig. 26: Scheme of a pepperpot emittance meter

Because of the small ionization volume and the small electron beam diameter, EBISs produce ion beams with very low emittances. As an example in Fig. 27 the x and y r.m.s. emittance for C^{6+} ions

extracted from a Dresden EBIS-A is shown as a function of the electron beam current. Typical emittances are of the order of some mm mrad.

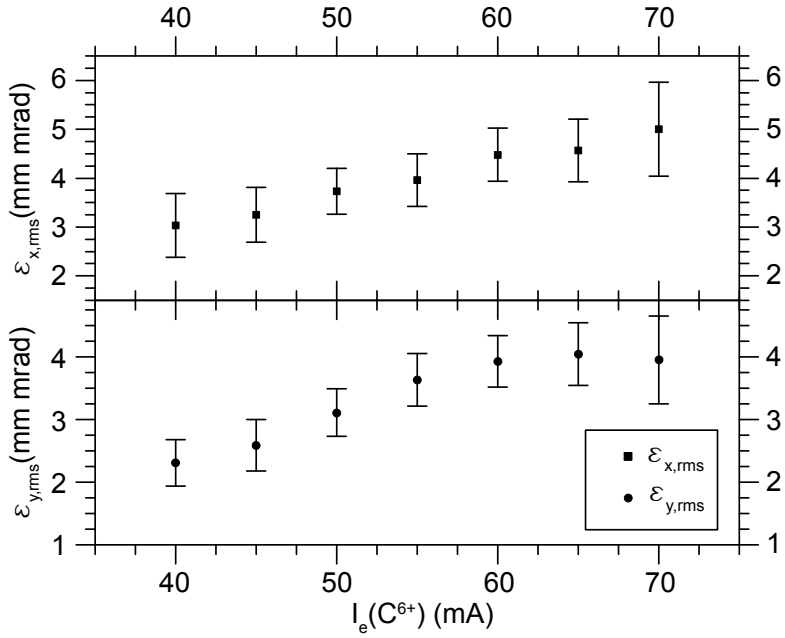


Fig. 27: R.m.s. emittance of C^{6+} ions extracted from a Dresden EBIS-A

For higher beam energies, the pulse components perpendicular to the beam direction are small compared with the pulse components in the beam direction, i.e. higher beam energies lead to lower emittances. This can be described by the normalized emittance and it yields

$$\varepsilon_{\text{norm}} = \beta'\gamma'\varepsilon \qquad \text{with} \quad \beta' = \frac{v}{c} \quad \text{and} \quad \gamma' = \frac{1}{\sqrt{1-\beta^2}}, \tag{35}$$

where v is the ion velocity and c is the speed of light. For typical operation regimes of EBISs, the normalized emittance is of the order of 10^{-3} mm mrad.

4.9 Energy spread and absolute energy of extracted ions

In order to measure the absolute energy and the energy spread of extracted ions, a retarding field analyser (RFA) can be applied. An RFA is an electrostatic energy analyser that allows the kinetic energy distribution of an ion beam to be measured. The analyser consists of a collimator with exchangeable apertures, three meshes, with the central mesh set on high potential, and a Faraday cup for current measurement.

The ion or electron current is measured in the Faraday cup of the RFA. In order to analyse the energy distribution of the incoming ion beam, the potential of the retarding mesh is increased stepwise while measuring the current in the cup. The measured dependence $I_{\text{ion}}(U_{\text{RFA}})$ is differentiated and fitted with a Gaussian to obtain the energy distribution. The mean energy of the charged particles E_{m} is defined as the centre position of the Gaussian. The energy spread of the charged particle beam ΔE is defined as the full width half-maximum (FWHM) of the Gaussian.

Figures 28 and 29 show examples of measured absolute energies and energy spread of ions extracted from a Dresden EBIS-A. The energy spread of the ion beam, which depends on the electron beam current, is generally quite small for Dresden EBIS/T systems. For the Dresden EBIS-A, it is of the order of several electronvolts. More details about the RFA technique can be found in Ref. [36].

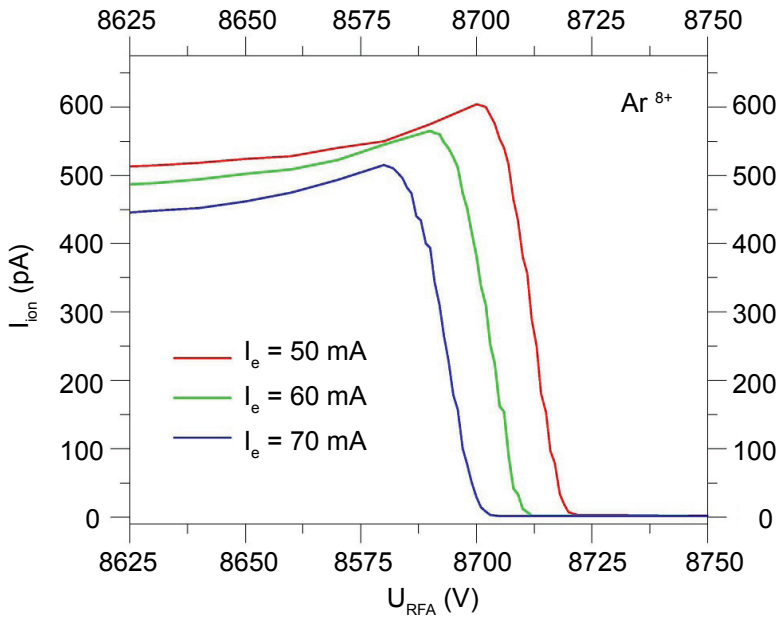


Fig. 28: Plot showing $I_{\mathrm{ion}}(U_{\mathrm{RFA}})$ curves for electron currents of 50, 60 and 70 mA measured with Ar^{8+} at a drift tube potential of 8.7 keV at a Dresden EBIS-A.

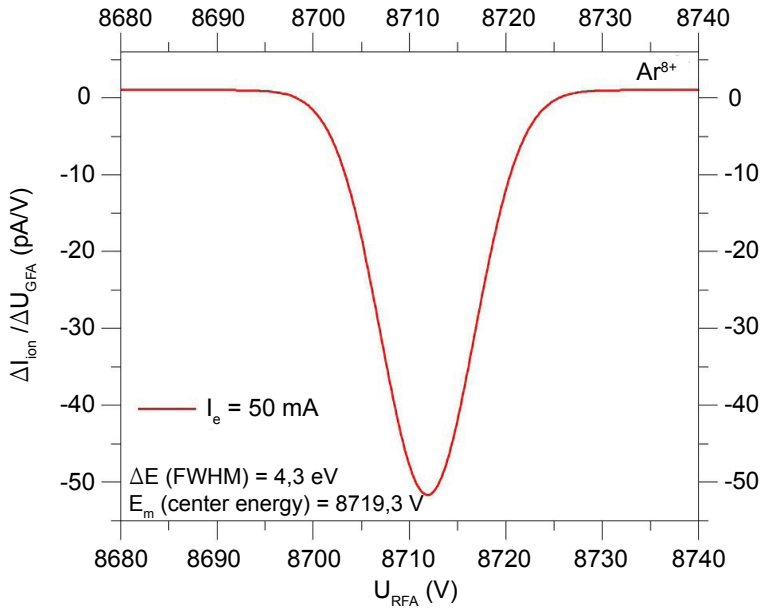


Fig. 29: The energy distribution of the extracted ion beam is $E = 4.3\,\mathrm{eV}$ at an electron beam of $I_{\mathrm{e}} = 50\,\mathrm{mA}$ measured at a Dresden EBIS-A.

5 EBISs as sources of X-rays from HCIs

5.1 X-ray emission from HCIs

EBISs are excellent tools for X-ray spectroscopy of highly charged ions. X-rays from HCIs are of fundamental interest for atomic physics, plasma diagnostics, astrophysics and other applications. In contrast to other sources of X-rays from HCIs such as hot plasmas or accelerator facilities, EBISs are comparatively small and low-cost devices. Additionally, they offer the possibility to control the electron beam energy very precisely, which creates superior conditions for advanced experiments studying X-rays from HCIs.

It is well known that vacancies in the atomic shell influence X-ray lines in terms of their transition energies, transition probabilities and the structure of the measured spectrum. Classical X-ray spectra have been known for more than 100 years. As long ago as 1917 Barkla received a Nobel Prize for the discovery of characteristic X-rays. Less than 50 years later, in 1962, the first X-ray source outside our Solar System was observed and from that date interest in X-rays from HCIs has grown more and more.

Compared to classical X-rays as parent diagram lines appearing after the production of an inner-shell vacancy and subsequent filling of this vacancy with an outer electron, X-rays from HCIs can be different from these classical transitions. Mainly, we can observe X-ray transitions from

- direct excitation (DE),
- radiative recombination (RR) and
- dielectronic recombination (DE).

Detailed information on the basic physics of X-rays from HCIs can be found, for example, in Refs. [37] and [38]. Measurements were performed by different groups working with EBISs – the pioneering work of the Livermore group of Dr. Beiersdorfer *et al.* should be pointed out especially (see e.g. Ref. [39]).

5.2 X-ray measurements

Measurements show that in the case of the Dresden EBIT individual dipole lines have a radiation power of the order of some nanowatts. For the Dresden EBIS-A this power increases by about one order of magnitude and for the superconducting Dresden EBIS-SC by a factor of about 200. As a rule, the number of X-rays emitted is high enough to perform energy-dispersive as well as wavelength-dispersive X-ray spectrometry in acceptable time frames of minutes up to some hours.

In Fig. 30 a measured wavelength-dispersive DE spectrum from carbon- to helium-like iron ions is shown. The measurement was realized with a flat crystal and for an electron beam current of 40 mA. Figure 31 shows an argon RR spectrum as measured with a Dresden EBIT. Here the electron capture into the highest possible argon charge state is shown with a clear signature.

DE as well as RR measurements can be realized with comparatively simple detection techniques. In both cases, without any synchronization with the EBIS, a detector can measure the emitted X-rays passing through a spectroscopic window. More sophisticated are measurements of DE processes where the measurement process must be synchronized with the electron beam energy because DE processes are sharp resonance processes. For example, in Fig. 32 a krypton DE KLL spectrum for Kr^{28+} up to Kr^{32+} ions is shown.

In addition to the classical energy- or wavelength-dispersive X-ray spectroscopy, time-resolved X-ray spectrometry is possible if we add to each measured count time information about the ionization time in the ion trap (time between closing the ion trap and registration of the X-ray quanta). Such a spectrum is shown in Fig. 33 for a Kr^{28+} DE KLL X-ray transition. The figure shows the evolution of the Kr^{28+} ionization charge state in the ion trap of a Dresden EBIT and demonstrates that such ions can be stored for 10 s or longer.

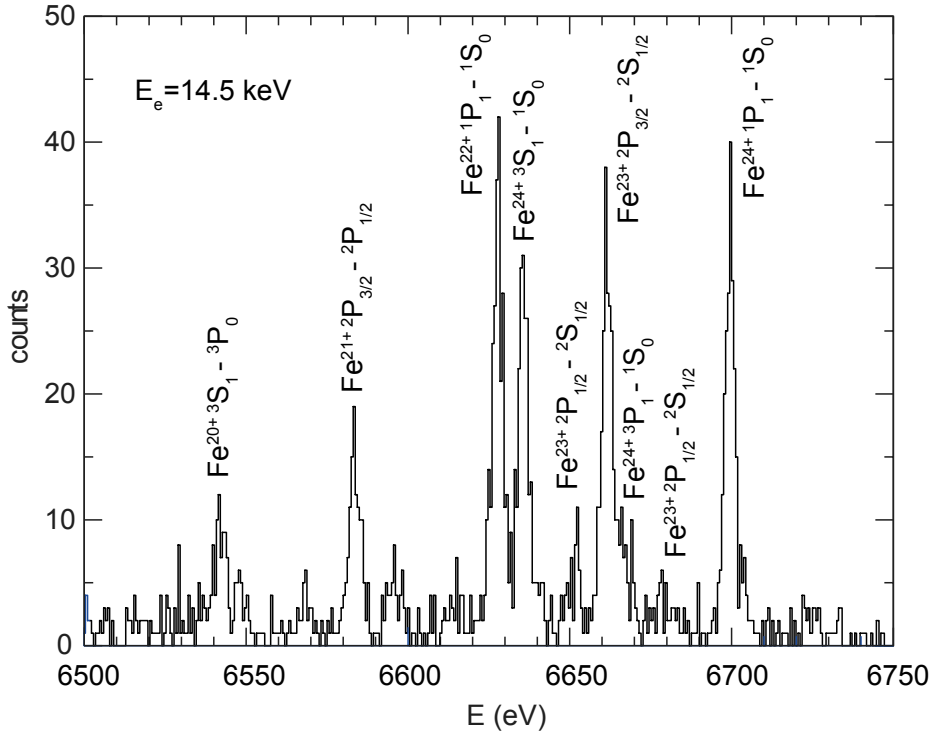


Fig. 30: Wavelength-dispersive iron X-ray spectrum from DE processes in a Dresden EBIT at 14.5 keV electron beam energy.

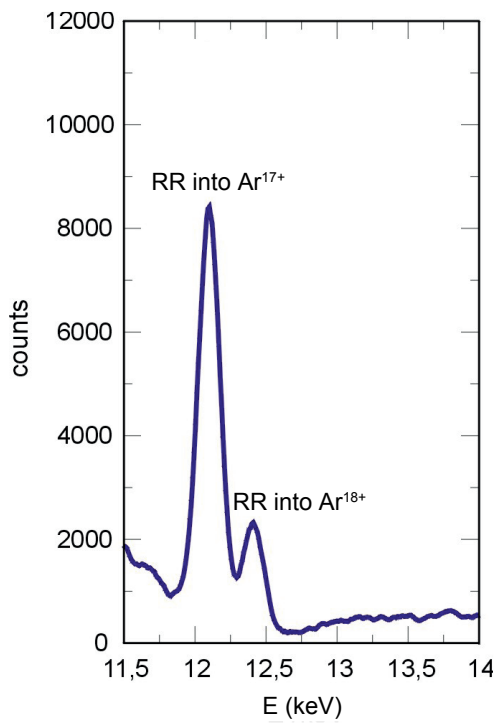


Fig. 31: Energy-dispersive measured argon RR spectrum from a Dresden EBIT for the electron capture into hydrogen-like and bare argon at an electron beam energy of 8 keV and an electron beam current of 15 mA.

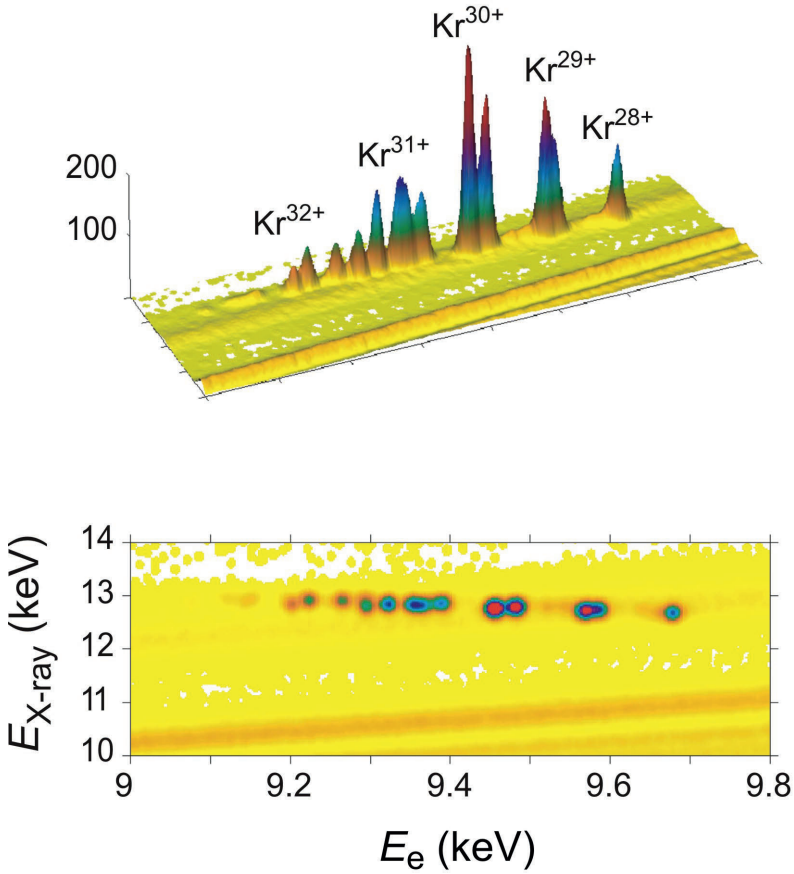


Fig. 32: Krypton KLL spectrum from DE processes for Kr^{28+} up to Kr^{32+} ions in a Dresden EBIT

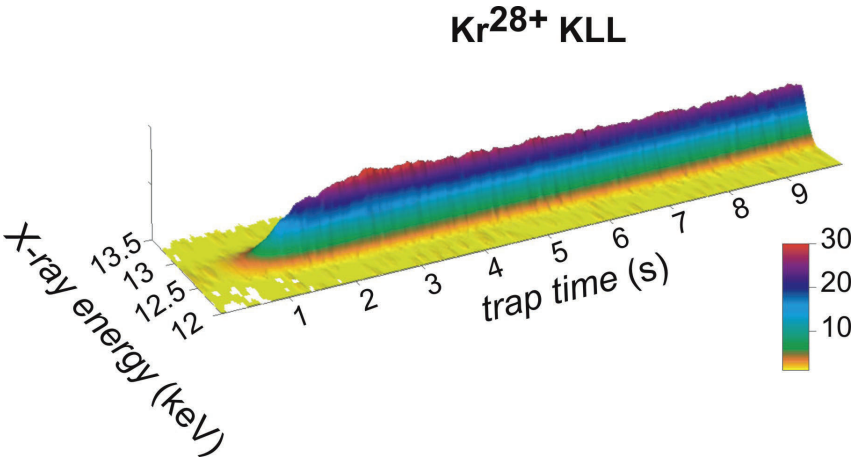


Fig. 33: Time-resolved krypton DE KLL spectrum for Kr^{28+} measured at a Dresden EBIT

6 Applications of EBISs (selected topics)

As mentioned at the beginning, EBISs are versatile tools for investigations on HCIs in basic and applied research as well as for applications with strong practical relevance. It is not the aim of the present chapter to discuss all these topics in detail, but in the following subsections we will give some examples of the versatile application potential of HCIs.

6.1 HCI-FIB for structuring and surface analysis

Standard focused ion beam (FIB) systems use liquid metal ion sources (LMISs) with their rather small standard spectrum of projectiles. Combining an EBIS, which provides a broader palette of ion species from different elements (from noble gases up to metals) ranging from high to low charge states, with an FIB column leads to new useable physical properties and new applications.

As already mentioned, the potential energy stored in the highly charged ions due to the ionization process leads to power densities of order 10^{12} up to $10^{14}\,\mathrm{W\,cm^{-2}}$ in single ion–surface interactions and therefore to higher yields of secondary ions and secondary electrons per incident ion compared to conventional ions. Preliminary experiments have been done at the CNRS in Marcoussis [40].

Furthermore, the ion implantation range can be varied by selecting a certain charge state, which is due to the final kinetic energy at a constant acceleration potential (see Fig. 34). This can be realized via the integration of a crossed-field ion beam separator, the already introduced Wien filter [32], which enables the user to separate ions according to their mass-to-charge ratio. Since the ion energy depends on the charge state as well as on the mass, a broad range of projectile energies is available. Also, the intrinsic potential energy of the projectile can be adjusted by the user by selecting the charge state of the ions.

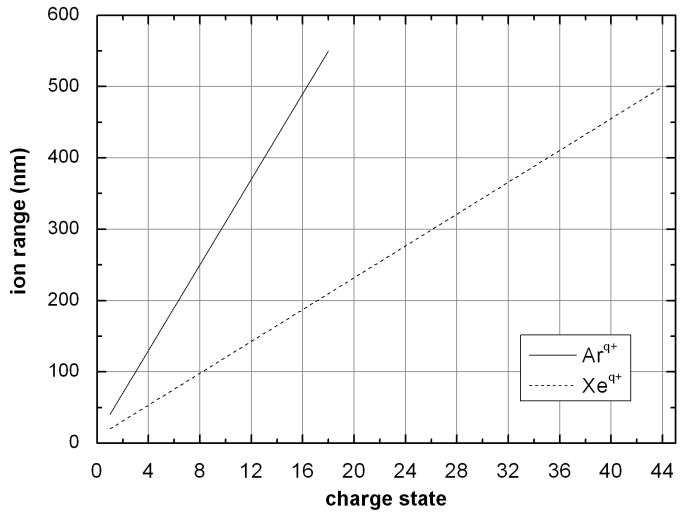


Fig. 34: Ion range of argon and xenon ions in silicon at different kinetic energies

The technology as described above complements the classical spectrum of projectiles for various applications beyond the latest developments using noble gas ions in FIB systems. Considering the ion charge state as a new physical parameter that can be varied offers suitable preconditions in sputtering processes with regard to the produced secondary particles being used for surface analysis (TOF-SIMS). This also takes into account the ability to vary the kinetic energy at a fixed acceleration potential of the ion source and FIB column assembly.

6.2 Particle therapy

Cancer is the second most common cause of death worldwide and, statistically, about 33% of all inhabitants of the European Union will be confronted with some kind of cancer in their lives. Currently about 45% of all cancer cases can be treated, mainly by surgery and/or by radiation therapy [41].

Within radiotherapy, radiation with hadrons (mainly protons and carbon ions) is the most promising treatment technique. In contrast to the classical irradiation with X-rays, treatment with ions has significant advantages. Energy deposition within the tissues obeys the Bragg interaction law, i.e. the maximum of the energy is deposited at the end of the particle trajectories, hence in the tumour cells. The ion range can be adjusted precisely by varying the kinetic energy of the particles. The intruding ions lose energy inside a dense medium according to the Bragg curve. Most of the kinetic energy is deposited at a certain depth along the ion path. This effect can be used to destroy a tumour while the healthy tissue around is spared (this is important for tumours in the vicinity of risky organs like the medulla).

Another important property of ions as opposed to photons is their electrical charge. Charged particles can be focused and deflected by electrical and magnetic fields, providing the ability to scan with the ion beam over the tumour very precisely. Owing to the variation of the kinetic energy of the ions, it is possible to vary the penetration depth. Hence the tumour can be scanned voxel by voxel (a voxel is a three-dimensional volume element). Besides the electrical properties, the relative biological efficiency (RBE) of the ions is higher by a factor of about 3. This means that the same dose of carbon ions affects the tumour tissue by a factor of 3 more compared to classical X-rays. A higher RBE allows a lower dose to produce the same effect on the tumour with less damage to the healthy tissue.

Nowadays, a substantial number of medical particle therapy facilities are based on synchrotron accelerators combined with ECR ion sources. Owing to the low purity of the ion beam, electron strippers are required. If using an EBIS ion source and its excellent beam purity, a subsequent electron stripper is no longer necessary. Also, the lower emittance of EBIS sources will lead to a higher transparency of the accelerator structure. For other accelerator strategies, like direct wall accelerators or direct driven accelerators, even the application of room-temperature EBIS ion sources might be possible.

6.3 Charge breeding

In nuclear mass measurements involving ions stored in Penning traps, the precision of the experimental method increases proportionally with the charge state of the investigated ions. Furthermore, for proposed nuclear fusion reaction studies, neutron-rich, relatively short-lived radioactive isotopes have to be accelerated to energies above the Coulomb barrier, which requires a small mass-to-charge ratio (A/q) to reduce the scale and cost of the accelerator. These and other endeavours rely on the availability of HCI from a broad spectrum of elements or even exotic radioactive species, respectively.

As described above, EBISs or EBITs are HCI sources which have many advantages. However, they are typically loaded via gas injection, which limits the range of available ion species from these machines. The injection of externally produced low charged ions into an EBIS or EBIT and re-extraction as HCIs, known as charge breeding, offers a way to increase the range of available HCI species. Various ion injection techniques have been used for charge breeding. Among these are experimental set-ups including metal vapour vacuum arc and liquid metal or liquid metal ion sources [42, 43] but also beams of short-lived radioactive isotopes produced using the isotope separation on-line (ISOL) technique [44].

Recently, a Dresden EBIS-A was fed by a liquid metal alloy ion source to produce HCI from low charged gold ions. For the charge-bred Au^{q+} ions, charge states up to $q = 60$ were achieved (see Fig. 35). This experimental technique represents a clean and elegant way of introducing metals or other elements into the Dresden EBIS-A that cannot be injected directly through a gas valve. Therefore, the range of applications for this type of ion source is extended.

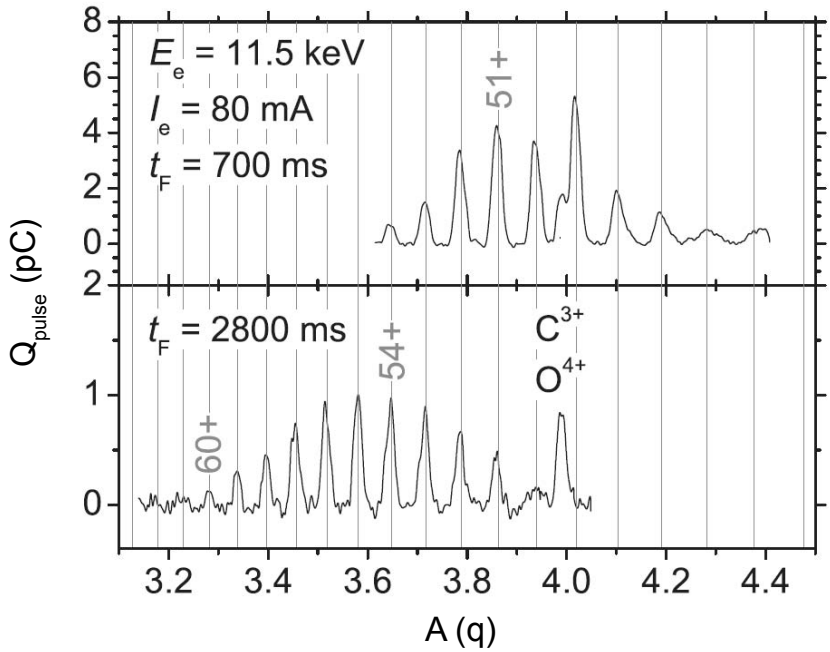


Fig. 35: Charge breeding spectra with gold ions injected into the trap of an EBIS-A with an external LMIS (from Ref. [45]).

7 Conclusion

Highly charged ions (HCIs) are becoming more and more important in basic research as well as technological applications (for example see Refs. [46] – [48]). This is because of their unique properties, which are listed briefly at the beginning of this chapter. In the past, large accelerator devices had to be used for the production of HCIs, while today's sources of HCIs will fit on a table top. This increases the availability of HCIs and opens up new possibilities for a larger community to realize experiments with HCIs.

This chapter has highlighted the electron beam ion source (EBIS) technology as an efficient and elegant way to produce HCIs. The basic physics and working principle of EBISs are explained in detail and compared to other existing HCI sources. Especially, aspects of HCI extraction but also the use of EBISs as sources of X-rays from HCIs trapped inside the machines are discussed.

We hope this work inspires researchers to continue the advancement of this extraordinary ion source technology and creates ideas for new fields of applications of HCIs produced by EBISs in the future.

References

[1] I. S. Bowen and R. A. Millikan, *Phys. Rev.* **25** (1925) 591.

[2] B. Edlen, *Z. Astrophys.* **22** (1942) 30.

[3] R. Geller, *Electron Cyclotron Resonance Ion Sources and ECR Plasmas* (IOP Publishing, Bristol 1996).

[4] B. Sharkov, Laser ion sources, in *The Physics and Technology of Ion Sources*, 2nd edn. (Wiley-VCH, Weinheim, 2004).

[5] J. Gillaspy, *Trapping Highly Charged Ions: Fundamentals and Applications* (Nova Science, Huntington, NY, 2001).

[6] F. Currell, Electron beam ion traps and their use in the study of highly charged ions, in *The Physics of Multiply and Highly Charged Ions*, Ed. F. Currell (Kluwer Academic, Dordrecht, 2003), Vol. I.

[7] J. Gillaspy, *J. Phys. B: At. Mol. Opt. Phys.* **34** (2001) R93.

[8] H. F. Beyer and V. P. Shevelko, *Introduction to the Physics of Highly Charged Ions* (IOP Publishing, Bristol, 2003).

[9] E. D. Donets, USSR Inventor's Certificate No. 248860, 16.03.1967; *Bull. OIPOTZ* **24** (1969) 65.

[10] E. D. Donets, V. I. Ilushenko and V. A. Alpert, Ultrahigh vacuum electron beam ion source of highly stripped ions, Proc. First Int. Conf. on Ion Sources, Saclay, France, 1969, p. 635.

[11] E. D. Donets and V. P. Ovsyannikov, *Sov. Phys. JETP* **53** (1981) 466.

[12] M. A. Levine, R. E. Marrs, J. R. Henderson, D. A. Knapp and M. B. Schneider, *Phys. Scr.* **T22** (1981) 157.

[13] J. W. McDonald and D. H. G. Schneider, *Nucl. Instrum. Methods Phys. Res.* **B241** (2005) 870.

[14] V. P. Ovsyannikov and G. Zschornack, *Rev. Sci. Instrum.* **70** (1999) 2646.

[15] For EBIT/EBIS ion sources of the Dresden EBIT/EBIS family see: `http://www.dreebit.com`.

[16] G. Zschornack, V. P. Ovsyannikov, F. Grossmann, A. Schwan and F. Ullmann, Proc. Int. Symp. on Electron Beam Ion Sources and Traps, Stockholm, Sweden, 7–10 April 2010; JINST 5 C08012.

[17] R. Becker and O. Kester, *Rev. Sci. Instrum.* **81** (2010) 02A513.

[18] V. P. Ovsyannikov, Technique and application of electron beam ion sources, Doctoral thesis, Joint Institute of Nuclear Research, Dubna, Russia, 1996.

[19] W. Lotz *et al.*, *Z. Phys.* **206** (1967) 205.

[20] T. Werner, Diploma thesis, Institut fuer Kern- und Teilchenphyik, TU Dresden, Dresden, Germany, 1997.

[21] A. Mueller and R. Frodel, *Phys. Rev.* **44** (1980) 29.

[22] A. Mueller and E. Salzborn, *Phys. Lett.* **A59** (1976) 19.

[23] F. Ullmann, Ph.D. thesis, Technische Universitaet Dresden, Fakultaet Mathematik und Naturwissenschaften, Dresden, Germany, 2005.

[24] Y. S. Kim and R. H. Pratt, *Phys. Rev.* **A27** (1983) 2913.

[25] R. Becker, Talk Int. Conf. on Ion Sources (ICIS 2009), Gatlinburg, TN, USA, September 2009.

[26] B. M. Penetrante, J. N. Bardsley, D. DeWitt, M. Clark and D. Schneider, *Phys. Rev.* **A43** (1991) 4861.

[27] I. V. Kalagin, D. Kuechler, V. P. Ovsyannikov and G. Zschornack, *Plasma Sources Sci. Technol.* **7** (1998) 441.

[28] G. Herrmann, *J. Appl. Phys.* **29** (1958) 127.

[29] H. Kuramoto, T. Kinnugawa, H. Watanabe, C. Yamada, S. Ohtani, I. Yamada and F. J. Curell, *Rev. Sci. Instrum.* **73** (2002) 42.

[30] A. Thorn, Ph.D. thesis, Technische Universitaet Dresden, Fakultaet fuer Mathematik und Naturwissenschaften, Dresden, Germany, 2012.

[31] G. Zschornack, A. Schwan, F. Ullmann, F. Grossmann, V. P. Ovsyannikov and E. Ritter, *Rev. Sci. Instrum.* **83** (2012) 02A507.

[32] M. Schmidt, H. Peng, G. Zschornack and S. Sykora, *Rev. Sci. Instrum.* **80** (2009) 063301.

[33] M. Kreller, G. Zschornack, U. Kentsch and R. Heller, *Rev. Sci. Instrum.* **79** (2008) 02A702.

[34] U. Kentsch, G. Zschornack, A. Schwan and F. Ullmann, *Rev. Sci. Instrum.* **81** (2010) 02A507.

[35] L. E. Collins and P. T. Stroud, *Nucl. Instrum. Methods* **26** (1964) 157.

[36] E. Ritter, Diploma thesis, Technische Universitaet Dresden, Department of Physics, Dresden, Germany, 2010.

[37] H. F. Beyer, H.-J. Kluge and V. P. Shevelko, *X-Ray Radiation of Highly Charged Ions* (Springer, Berlin, 1997).

[38] F. J. Currell (Ed.), *The Physics of Multiply and Highly Charged Ions* (Kluwer Academic, Dordrecht, 2003), Vol. I.

[39] P. Beiersdorfer, *AIP Conf. Proc.* **389** (1997) 121.

[40] F. Ullmann, F. Grossmann, V. P. Ovsyannikov, J. Gierak and G. Zschornack, *Appl. Phys. Lett.* **90** (2007) 083112.

[41] T. Auberger, E. Griesmayer, Das Projekt MedAustron - Designstudie, Fotec - Forschungs- und Technologietransfer GmbH, Wiener Neustadt, 2004 (chapter 1).

[42] N. Nakamura, T. Kinugawa, H. Shimizu, H. Watanabe, S. Ito, S. Ohtani, C. Yamada, K. Okazaki, M. Sakurai, M. R. Tarbutt and J. D. Silver, *Rev. Sci. Instrum.* **71** (2000) 684.

[43] A. Pikin, J. G. Alessi, E. N. Beebe, A. Kponou and K. Prelec, *Rev. Sci. Instrum.* **77** (2006) 03A909.

[44] F. Wenander, *J. Instrum.* **5** (2010) C10004.

[45] A. Thorn, E. Ritter, F. Ullmann, W. Pilz, L. Bischoff and G. Zschornack, *Rev. Sci. Instrum.* **83** (2012) 02A511.

[46] R. W. Schmieder and R. J. Bastasz, Proc. 6th Int. Conf. on the Physics of Highly Charged Ions, *AIP Conf. Proc.* **274**, Manhattan, KS, 1992.

[47] G. Zschornack, F. Grossmann, V. P. Ovsyannikov, R. Heller, U. Kentsch, M. Kreller, M. Schmidt, A. Schwan, A. Silze and F. Ullmann, *Materialwissenschaften und Werkstofftechnik* **40**(4) (2009) 1.

[48] V. Shevelko and H. Tawara (Eds.), *Atomic Processes in Basic and Applied Physics* (Springer, Heidelberg, 2012).

Resonance Ionization Laser Ion Sources

B. Marsh
CERN, Geneva, Switzerland

Abstract
The application of the technique of laser resonance ionization to the production of singly charged ions at radioactive ion beam facilities is discussed. The ability to combine high efficiency and element selectivity makes a resonance ionization laser ion source (RILIS) an important component of many radioactive ion beam facilities. At CERN, for example, the RILIS is the most commonly used ion source of the ISOLDE facility, with a yearly operating time of up to 3000 hours. For some isotopes the RILIS can also be used as a fast and sensitive laser spectroscopy tool, provided that the spectral resolution is sufficiently high to reveal the influence of nuclear structure on the atomic spectra. This enables the study of nuclear properties of isotopes with production rates even lower than one ion per second and, in some cases, enables isomer selective ionization. The solutions available for the implementation of resonance laser ionization at radioactive ion beam facilities are summarized. Aspects such as the laser requirements, ion source conditions, selectivity, efficiency and applicability are covered. The practical solutions implemented at the CERN ISOLDE facility for laser beam transport and control, reliability and ease of operation will also be described.

1 Introduction

In addition to providing a general overview of the history, operating principles, capabilities and status of existing resonance ionization laser ion sources (RILIS), this chapter is intended as a guide to the key aspects of the laser ion source through a description of laser solutions, technical considerations and atomic excitation schemes. The CERN ISOLDE RILIS is used as an example to illustrate how these considerations have been implemented at a well-established operating facility. Included in the appendix is a list of useful references and an updated table of the isotopes that have so far been ionized using laser ion sources at radioactive ion beam facilities worldwide.

1.1 Radioactive ion beam (RIB) facilities

Many RIB facilities are operating or are under development around the globe. An overview of the European facilities is given elsewhere [1]. Blumenfeld *et al.* [2] discuss the state-of-the-art ion beam production methods and summarize the plans for future facilities worldwide. In general, RIB facilities fall into one of two main categories: isotope separator on-line (ISOL) and in-flight fragment separators. These are illustrated schematically in Fig. 1. Common to all RIB facilities is the need for a primary (driver) projectile beam, a target and a magnetic isotope separator. The categories are defined by the method of ion beam preparation: radionuclide production, transport from the target, ionization, extraction and mass separation.

At ISOL facilities, a modular approach is adopted for RIB production [3], and the result is a low-energy (<60 keV), low-emittance ion beam that is easy to manipulate. There are two distinct types of ISOL facility, shown as **1** and **2** in Fig. 1. In-flight facilities (**3** in Fig. 1) rely on the fragmentation of a heavy projectile that is incident on a thin target:

1. *Thick target plus transfer line.* Reaction products are stopped within the thick target material and transported as an atomic vapour by diffusion and effusion processes to an ion source. The target–ion source assembly is on a high-voltage (e.g. 60 kV) platform. To minimize energy spread of the

resulting ion beam, a single charge state (usually +1) is preferred and the ionization takes place in a field-free region. Ions then drift towards the extraction region where they are accelerated to become an ion beam due to their proximity to a grounded extraction electrode.

2. *Thin target plus gas cell.* Reaction products are ejected from the thin target and stopped within a gas cell containing helium or argon. The gas pressure is optimized to favour the single positive charge state and the ions are ejected into an electric guiding and extraction field by the gas flow through the exit nozzle of the cell.
3. *Projectile fragment separator.* The projectile fragments continue on with a largely unperturbed forward velocity and enter a wide-acceptance mass separator directly. A thorough description of the projectile fragmentation technique can be found in Ref. [4] and references therein. The use of a thin target and the large diversity of reaction products (large energy spread and multiple charge states) mean that the ion flux for a chosen beam at such facilities is often much lower than can be achieved at thick-target ISOL facilities. On the other hand, because of the immediate extraction of reaction products, the yield of very exotic or short-lived species can be many orders of magnitude greater and the high beam energy gives immediate compatibility with some colliding-beam experiments or storage ring systems. The nature of the extraction process removes any need for an ion source and there is no chemical dependence of the release. Ion beams of even the most refractory elements can be produced.

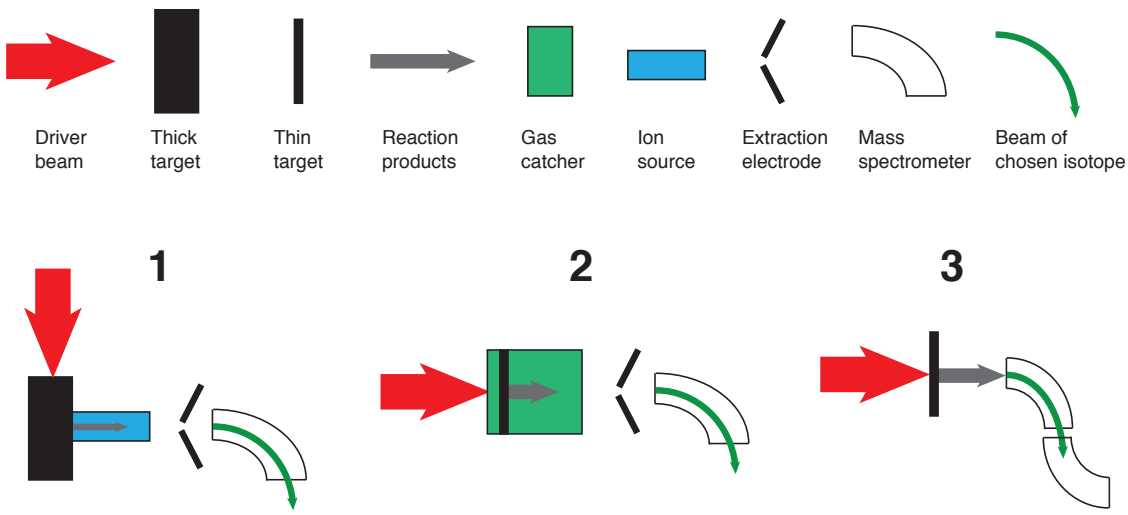


Fig. 1: Methods for radioactive ion beam production. Sketches **1** and **2** show the ISOL method. Sketch **3** shows the fragment separator method. In standard operation, only method **1** requires a separate ion source.

1.2 Ion source requirements

As is shown in Fig. 1, only thick-target ISOL facilities require a separate ion source as standard. When specifying an ion source, the following key requirements must be defined:

energy; intensity; selectivity; efficiency; universality; particle type; speed; charge state.

At an ISOL facility, each of these requirements can vary significantly from one experiment to another due to the diversity of the ion beams that may be produced. It is therefore necessary for several ion source types to be available. The ion sources of ISOL facilities [5] can be grouped into three categories:

Surface ion source. Ionization (X^+ or X^-) by electron exchange during the interaction of an atom with a heated metal surface. This can be a simple, robust heated metallic capillary. High efficiencies (e.g., >50% for Cs) are achieved for elements with a low ionization potential [6].

Electron impact ion source. Ionization (X^+ to X^n) by high-energy electrons within an electron beam or a plasma. This is a complex device and is sensitive to several carefully tuned parameters, such as buffer gas pressure, anode and cathode voltages, and magnetic field strength. Very high efficiencies are achieved for many elements and molecules. Any element that is released from the target can be ionized because the electron energy can exceed several hundred electronvolts. Electron impact ionization is often applied at ISOL facilities in a secondary ion source for high charge state production prior to post-acceleration of the ion beam [7].

Laser ion source. Ionization (X^+) by multi-step resonant (and non-resonant) photon absorption. This is an essential aspect of many ISOL facilities because it is an efficient ionization mechanism that is completely element-selective, without introducing any extra delay to the ion extraction process. Atoms enter a laser interaction region (usually the hot cavity of the surface ion source) into which multiple laser beams are transmitted. The laser wavelengths are precisely tuned so that the photon energies match successive electron transition energies of an ionization scheme [8].

1.3 Motivation for a resonance ionization laser ion source

The alkali metals with a low ionization potential can be efficiently and rather cleanly ionized with a surface ion source. On the other hand, the halogen elements, which have a very high ionization potential, are not surface ionized and their atomic structure, particularly the high energy of the first excited atomic state, renders them out of the practical reach of a laser ion source. In this case, electron impact ionization is the only option. Most of the elements, however, lie between these extremes of the periodic table. For many of these, the laser ion source can be the optimal solution due to the following additional advantages of this technique.

- Reduced isobaric contamination: better conditions for experiments and lower radioactive inventory at mass separator.
- Beam diagnostics: the ionization process for the element of interest can be switched on and off simply by blocking a laser beam, enabling easy identification of the isotope of interest within a background of other isotopes.
- The ionization region can be simple and robust, whilst the complexity of the device is safely located away from the hostile target environment.
- Isomer-selective ionization can be achieved for some isotopes [9, 10].

1.3.1 Mass separation and element selectivity

The General Purpose Separator (GPS) of ISOLDE is a typical example of a dipole mass separator at an ISOL facility. The mass resolution, $M/\Delta M$, is of the order of several thousand ($\approx 10^3$). This is sufficient for the selection of a particular atomic mass unit so long as the neighbouring mass components of the incident beam are not orders of magnitude more intense. Isobar separation, which typically requires a resolution exceeding 10^5, is outside the scope of standard dipole magnet mass separators. By adding an element-selective (Z) process to the ion beam preparation, the RILIS complements the mass (A) selection of the mass separator, to enable the selection of a single isotope of a single element (Fig. 2).

To fully appreciate the potential problem of isobaric contamination in a mass-separated beam, one can consider the production of A clean beam of ^{78}Cu at a thick-target ISOL facility (Fig. 3). Using a 1 GeV proton driver beam and a uranium target, the in-target production rates for the isobars, ^{78}Zn and ^{78}Ga, are several orders of magnitude higher than for ^{78}Cu.

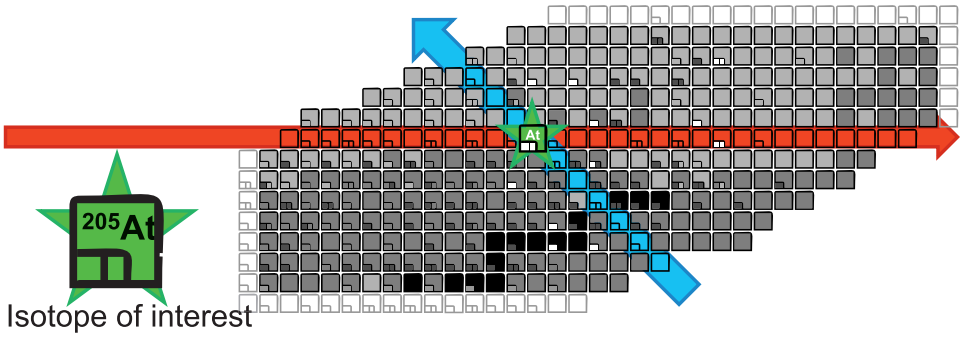


Fig. 2: The combination of element-selective (Z) ionization and mass (A/Q) selectivity results in an ion beam of one isotope of one element.

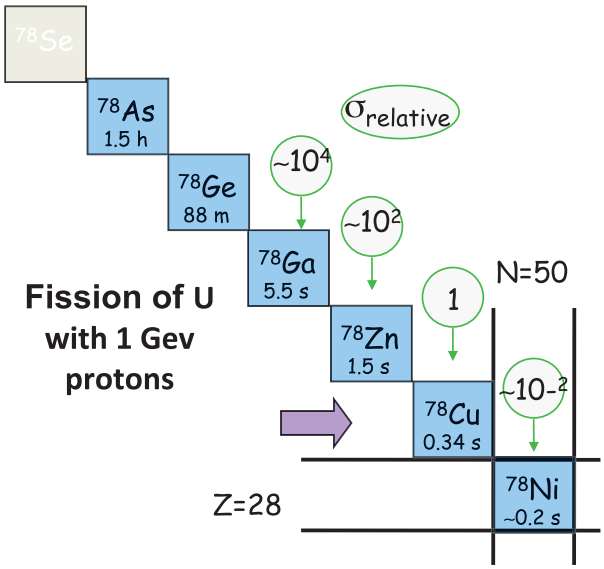


Fig. 3: The problem of isobaric contamination: high production rates of surface ionized isobars

The ionization potential (IP = 7.73 eV) of Cu is too high for efficient surface ionization and, neglecting any chemical dependence of release, unselective ionization with an electron beam ion source would result in an ion beam composition that corresponds to the production rate ratios (10 000 : 1 for ^{78}Ga : ^{78}Cu). A RILIS ionization efficiency for Cu of >7% has been measured for ionization inside the hot cavity of the surface ion source. At the normal operating temperature of ≈ 2000 K, surface ionization of Zn (IP = 9.39 eV) is negligible whereas the ionization efficiency of Ga is of the order of 0.1%. The resulting ion beam (surface plus laser ionization) therefore is improved to ≈ 100 : 1 ratio of ^{78}Ga : ^{78}Cu. This is a more manageable contamination level, and the effect of the remaining ^{78}Ga contamination can be limited by the use of methods such as ion beam gating (based on differences in release times after proton impact between Cu and Ga), selective detection using characteristic radiation decay. The ability to completely suppress the Cu ion component of the beam by blocking one of the RILIS lasers facilitates the optimization of these methods. It has been demonstrated that almost complete suppression of isobaric contamination is feasible but requires the active repulsion of the surface ions, and ionization must take place inside a separate laser–atom interaction region. This will be discussed in section 5.

2 Operating principle of a resonance ionization laser ion source

The resonance ionization laser ion source is based upon this principle of stepwise photo-ionization. This process was demonstrated by Letokhov and colleagues in 1972 [11]. The theory of resonance ionization is discussed in detail, with emphasis on the technique of resonance ionization spectroscopy, in Letokhov's book *Laser Photoionization Spectroscopy* [12].

Using tunable lasers, it is possible to match the photon energy of the laser light to an electronic transition of a particular element. If this light is incident on an atomic vapour containing many atomic elements, the photons will be resonantly absorbed by an electron occupying a compatible energy level: the transition energy must match the energy of the absorbed photon. If the photon flux is sufficient, and a series of laser frequencies are used, each precisely tuned so that the respective photon energies match transitions between successive electron energy levels of a particular atomic species, then the valence electron will be promoted through the energy level structure to reach an excitation energy close to the ionization potential of the element. An atom with a highly excited valence electron can be ionized by electric fields, atomic collisions or absorption of a photon with an energy greater than the difference between the ionization potential and the energy of the excited electron. These processes are illustrated in Fig. 4. Since, for each element, the energy level structure is discernibly different, the stepwise resonant nature of the process ensures a high degree of Z selectivity, and the atomic vapour of other elements will be unaffected by the laser light.

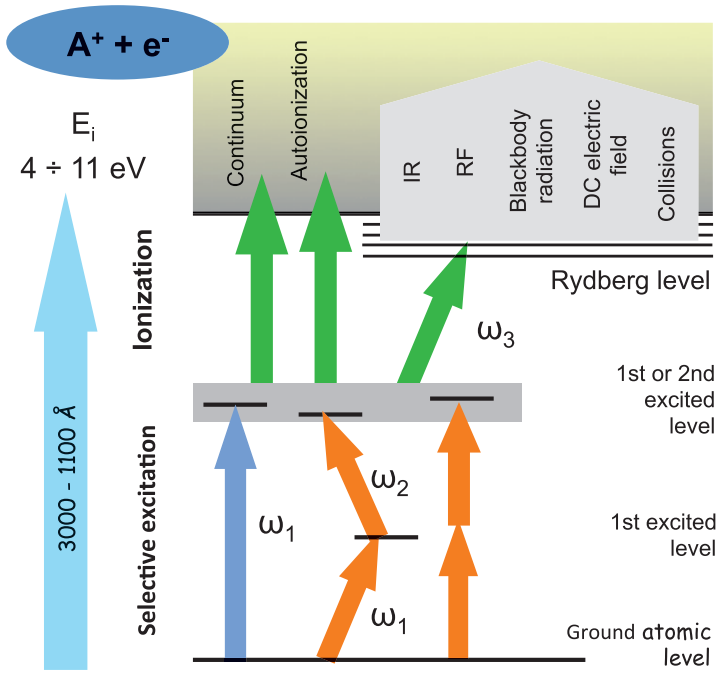


Fig. 4: The principle of the resonance ionization technique: a stepwise sequence of resonant photon absorption followed by resonant or non-resonant ionization.

2.1 Atomic structure

The atomic energy level structure is the basis of atomic physics and is therefore the subject of many textbooks [13,14]. The energy of an electron bound to an atom depends on its location within the discrete energy level structure and is a result of several electron–nucleus and electron–electron interactions.

The state of an electron is defined by the quantum numbers n, l, j, m_j, where n is the principal quantum number and l is the electron orbital angular momentum; n can be any positive integer and $l = 0, 1, 2, \ldots, n-1$. The total electronic angular momentum j, $j = l + s$, is given by coupling the

orbital angular momentum with the intrinsic electronic spin s to give $j = l \pm \frac{1}{2}$. The magnetic quantum number m_j is the projection of the total angular momentum j along its quantization axis. There are $2j + 1$ magnetic substates, m_j, which are degenerate in the absence of an external magnetic field. The quantum numbers for the valence electrons combine to describe the atomic energy level using the *term* symbol:

$$^{2S+1}L_J, \tag{1}$$

which contains the following information:

Total spin (multiplicity), S.

This is given by the spin–spin interaction and is the coupling of the intrinsic angular momentum of electrons (spin-$\frac{1}{2}$ particles): $S = 0$ (antiparallel), $S = 1$ (parallel).

Orbital angular momentum, L.

This is the coupling of the orbital angular momenta of the valence electrons:

$$L = \mathrm{S}(0),\ \mathrm{P}(1),\ \mathrm{D}(2),\ \mathrm{F}(3),\ \cdots$$

Total angular momentum, J.

This is the coupling of the total orbital angular momentum and the total spin (spin–orbit interaction, or *fine structure*): $J = |L - S|,\ \ldots,\ |L + S|$. The suitable angular momentum coupling scheme should be applied.

L–S coupling

For light multi-electron atoms ($Z < 30$), the electron spin coupling ($s \cdot s$) is larger with respect to the individual electron spin–orbit interaction ($l + s$) than it is for heavier atoms. In this case, the angular momentum coupling scheme used to describe the atomic state is Russell–Saunders or L–S coupling: the total electronic orbital angular momentum L is given by $\mathbf{L} = \sum_i \mathbf{l}_i$ and the total spin S is given by $\mathbf{S} = \sum_i \mathbf{s}_i$. The total angular momentum J of an atomic energy level is given by coupling these two terms, $\mathbf{J} = \mathbf{L} + \mathbf{S}$.

jj coupling

For heavier elements, spin–orbit interaction ($l \cdot s$) is dominant and the resulting values of j sum to give the total angular momentum $\mathbf{J} = \sum_i \mathbf{j}_i$.

Using the selection rules for atomic electric dipole transitions (which arise from momentum conservation and symmetry rules), comparing two term symbols for energy levels of an atom is a guide as to whether a transition between the levels is allowed. For electric dipole transitions between atomic energy levels, conservation of both angular momentum and parity demand that the condition $\Delta l = \pm 1$ is met. Where L–S coupling holds, further rules apply: $\Delta S = 0$ (not strict), $\Delta L = 0, \pm 1$, $\Delta J = 0, \pm 1$ except for $J = 0 \rightarrow J = 0$ transitions. Additionally, $\Delta m_J = 0, \pm 1$ and $\Delta m_J = 0$ is forbidden if $\Delta J = 0$.

2.2 Energy scales and commonly used units

Below is a description of the various quantities that are relevant to the following sections of this chapter, and to the field of atomic spectroscopy in general.

Wavelength. Wavelength λ has SI units of metres, but, for describing photons of suitable energy for excitation of atomic transitions (ultraviolet–visible–infrared range), nanometres ($1\ \mathrm{nm} = 10^{-9}\ \mathrm{m}$) or angstroms ($1\ \text{Å} = 10^{-10}\ \mathrm{m}$) is more convenient. The wavelength is dependent on the refractive index of the medium. In atomic spectroscopy literature, wavelength above 200 nm are often quoted as *air* wavelengths, λ_{air}. In units of angstroms, *vacuum* to *air* wavelength conversion uses this formula [15]:

$$\lambda_{\mathrm{air}} = \frac{\lambda_{\mathrm{vac}}}{1.0 + 2.735\,182 \times 10^{-4} + 131.4182/\lambda_{\mathrm{vac}}^2 + 2.762\,49 \times 10^8/\lambda_{\mathrm{vac}}^4}. \tag{2}$$

Energy. For the energy scales relevant to atomic physics, the electronvolt ($1\ \mathrm{eV} \approx 1.6022 \times 10^{19}$ J) is useful, since a typical electron transition energy is several electronvolts. The transition energy can be calculated from the wavelength using the relation $E = h/\nu = hc/\lambda$, but, because the value of h is not known to more than eight significant figures, a more precise description of transition energies is given by the *wavenumber*.

Wavenumber. This is the reciprocal of the *vacuum* wavelength $\bar{\nu} = 1/\lambda_{\mathrm{vac}}$ and is given in units of cm^{-1}. Conveniently, the *wavenumber* is directly proportional to the energy: $E = hc\bar{\nu}$. For this reason, transition energies are often given as *wavenumbers* (cm^{-1}).

Frequency. In atomic spectroscopy, units of frequency are used to describe both small-scale energy distributions, such as spectral linewidths, and the smallest energy intervals that are commonly observed in atomic physics: the isotope shifts and hyperfine structure. These are usually given in units of MHz or GHz ($1\ \mathrm{GHz} \approx 0.033\ \mathrm{cm}^{-1}$).

2.3 Resonance excitation of atoms

The excitation of an electron between a lower (E_1) and upper (E_2) level of a transition will occur if the atom is exposed to a photon flux Φ ($\mathrm{cm}^2\ \mathrm{s}^{-1}$) with a transition rate given by

$$W_{12} = \Phi\sigma_{12}(\lambda), \tag{3}$$

provided that the wavelength of the photon is equal to $\lambda = hc/E_{12}$. Here $\sigma_{12}(\lambda)$ is the absorption cross-section for atoms in state E_1. The decay rate by spontaneous emission of N_2 atoms in state E_2 is given by

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -A_{21}N_2, \tag{4}$$

where A_{21} (s^{-1}) is the Einstein A-coefficient [16], which is a characteristic constant for the transition.

For a two-level system, the transition is saturated when the absorption rate equals the rate of spontaneous emission and the populations N_1 and N_2 reach an equilibrium, with the relative occupancy of the levels given by their statistical weights, $2J + 1$. If the same principle is applied for successive excitation steps, it is possible for a high-lying energy level to become populated. If continuous excitation from the higher level results in a rate of ionization, a continuous net excitation and ionization sequence occurs. A thorough treatment of this process is given in Ref. [12].

Since the cross-sections $\sigma_{i,i+1}$ for resonant transitions between energy levels are often many orders of magnitude higher than for the ionizing step, the ionization rate is normally governed by the laser power available for the ionizing transition. This is discussed in section 2.5.

Assuming saturation of the resonant transitions, the ionization efficiency will be highest for a system that uses an upper excited state with a higher statistical weighting (highest J).

2.4 Spectral linewidth

The spectral shape of the absorption or emission lines of atomic transitions is Lorentzian. Demtröder [13] explains the atomic spectral line shape using the classical analogy of a damped harmonic oscillator, for

which the Fourier transform of the amplitude of oscillation gives a Lorentzian-shaped intensity profile. For a thorough quantum mechanical treatment of the atomic excitation and emission process, the density matrix formalism [17, 18] is required and the *optical Bloch equations* are used. For a two-level system, this approach yields the resonant saturation parameter, S_0:

$$S_0 = \frac{I}{I_{\mathrm{sat}}}, \qquad I_{\mathrm{sat}} = \frac{\pi h c}{3\lambda^3 \tau}. \tag{5}$$

For a $\Delta J = 0$ transition, as S_0 increases, the population of the excited state increases to a maximum value of 50%. An occupancy of the excited state that exceeds 50% is known as a *population inversion* and is only possible with a pumped system of more than two levels. This is prerequisite for laser operation [19].

For an unsaturated atomic transition, the Lorentzian linewidth is given by the natural lifetime, τ, of the excited state and can be most simply calculated using the *uncertainty principle* ($\Delta E \cdot \Delta t \geq \hbar/2$) to give

$$\Delta\nu_{\mathrm{nat}} = \frac{1}{4\pi\tau}. \tag{6}$$

For a two-level system the natural linewidth of the transition is simply related to the atomic transition probability A_{ki} using

$$\tau = 1/A_{ki}. \tag{7}$$

Example
$^3\mathrm{P}_{1/2} \rightarrow {}^3\mathrm{S}_{1/2}$ Na D line with $\tau = 16$ ns:
$\Delta\nu_{\mathrm{nat}} = 10^{-9}/16 \cdot 2\pi \approx 10\ \mathrm{MHz}$

Since typical lifetimes are in the range of several nanoseconds, the natural linewidth is in the range of 10–100 MHz. For the purposes of resonance ionization and optical spectroscopy, the mechanisms that result in a broadening of this linewidth or a distortion or shift of the line-shape must be considered because they are commonly far greater than $\Delta\nu_{\mathrm{nat}}$.

Power broadening. This effect can be observed if the atomic transition is excited with a laser intensity that approaches or exceeds the saturation intensity. The broadening, which is illustrated by the coloured curves in Fig. 5, occurs as follows: Below saturation (blue line), the linewidth is narrow but the tails of the resonance profile extend far from the central frequency. As the saturation intensity is approached (yellow line), the population of the excited state converges towards its maximum value for a two-level system of 50% at $\omega = \omega_0$. Increased laser intensity (red line) cannot significantly increase the population of the excited state if $\omega = \omega_0$, but if $\omega \neq \omega_0$ then saturation is not yet reached so the population can increase. A broadened Lorentzian profile is therefore observed. In this situation the power-broadened linewidth ($\Delta\nu_{\mathrm{p}}$) is determined by the degree of saturation,

$$\Delta\nu_{\mathrm{p}} = \sqrt{1 + S_0}. \tag{8}$$

Doppler broadening. This is the result of the velocity distribution of atoms in a vapour due to the vapour temperature. This velocity profile is given by the Maxwell–Boltzmann distribution, shown as black lines in Fig. 5. The influence of this velocity spread is described in detail, along with worked examples in Ref. [13] and other atomic physics textbooks. The resulting Doppler broadening contribution, $\Delta\nu_{\mathrm{D}}$, to the linewidth is

$$\Delta\nu_{\mathrm{D}} = \nu_0 \sqrt{\frac{8 k_{\mathrm{B}} T \ln 2}{m c^2}}. \tag{9}$$

Figure 5 shows the broadening of the Doppler profile due to a factor of 2 increase in temperature.

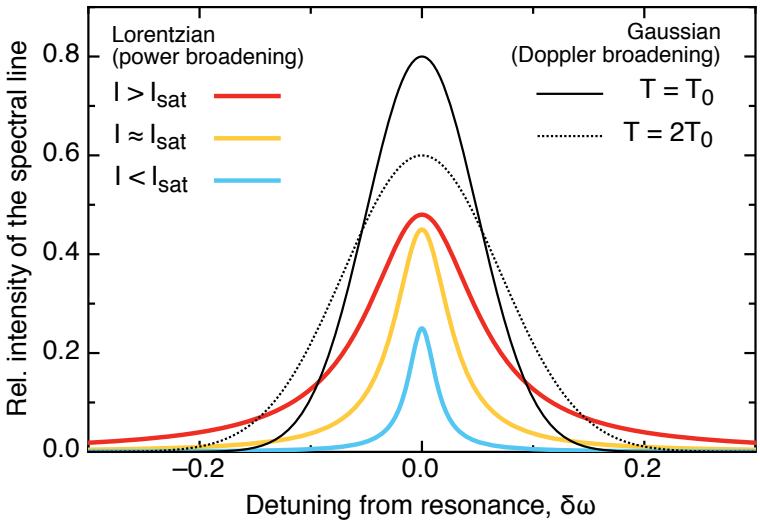


Fig. 5: Spectral line-shape broadening for atomic transitions. In the absence of Doppler broadening, a Lorentzian peak with a full width at half-maximum equal to the natural linewidth is observed if the spectral line is probed with a single-mode laser at an intensity below saturation (blue curve). Increasing the laser intensity broadens the line according to Eq. (8). In a high-temperature vapour, the thermal motion of the atoms gives rise to a Gaussian contribution to the line-shape (black curves) with a width that depends on the mass of the atom and the temperature.

Example

$^3\mathrm{P}_{1/2} \rightarrow {}^3\mathrm{S}_{1/2}$ Na ($m = 23$ g mol^{-1}) D line with $\lambda = 589.1$ nm, $T = 500$ K:

$$\Delta\nu_\mathrm{D} = 7.16 \times 10^{-7}(c/539.1) \times 10^{-9}\sqrt{500/23} = 1.7\ \mathrm{GHz}$$

Pressure broadening. This occurs if the atom is inside a buffer gas: collisions of atoms with other atoms or molecules (collisional broadening) results in reduction of the effective lifetime of the excited state and therefore a larger uncertainty for the emission energy. The result is a broadened Lorentzian linewidth that depends on the type, density and temperature of the gas. An additional effect occurs due to the presence of other atoms or molecules in the vicinity of the excited atom. This results in a shift, rather than a broadening, of the energy levels of the atom. This effect is rather insensitive to temperature changes. The extent of each of these effects can be seen in Fig. 6 [20].

2.5 Ionization of excited atoms

2.5.1 Non-resonant laser photo-ionization

Ionization by non-resonant photon absorption to induce a transition to the continuum from a suitably high-lying excited state is a commonly used method.

The cross-section for this type of transition is low ($\sigma_\mathrm{ni} = 10^{-19}$–$10^{-17}$ cm^2). The required photon flux, Φ (cm^{-2} s^{-1}), for this transition is therefore higher than for the stronger resonant transitions of earlier excitation steps in the ionization scheme. For saturation of the ionization step, the *flux* condition must be met:

$$\sigma_\mathrm{ni}\Phi_\mathrm{i} > \beta, \tag{10}$$

where β is the spontaneous emission rate (typically 10^6 s^{-1}). For a beam spot of 10 mm^2, the flux condition is $\Phi_\mathrm{i} = 10^{24}$ cm^{-2} s^{-1}. This corresponds to a continuous-wave (CW) laser power in the 10 kW range, unattainable or impracticable for use at an ISOL facility. Pulsed lasers are capable of attaining a peak power that exceeds this level, e.g., a 10 W, 10 ns pulsed laser with a repetition rate of

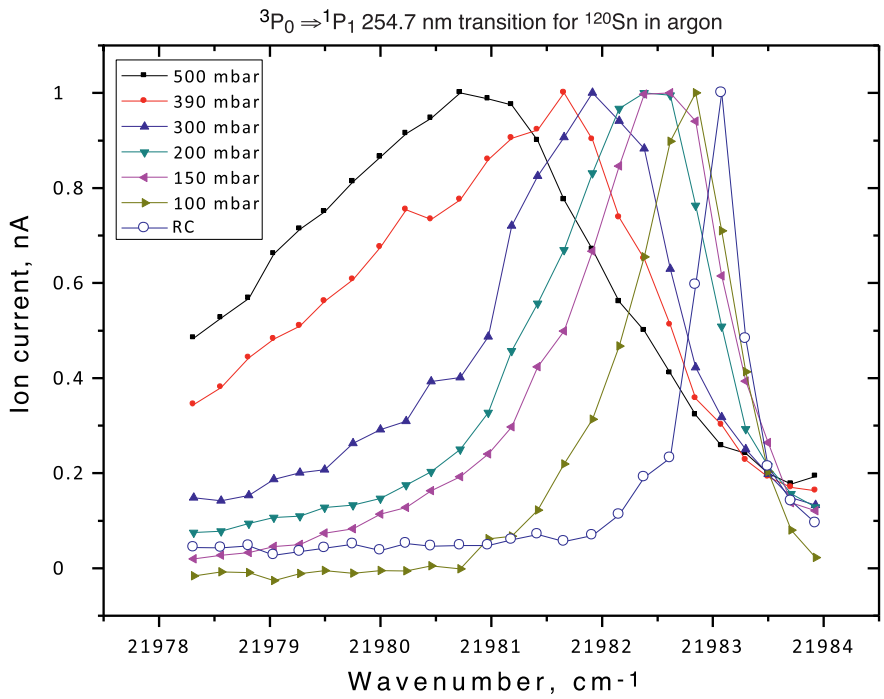


Fig. 6: Pressure broadening and shifts for ^{120}Sn in an argon gas cell compared to the spectral line observed in a gas-free reference cell (RC) [20].

10 kHz has a peak power of 100 kW. In this case, however, the ionization rate becomes limited by the number of photons that can be delivered within the effective laser interaction period. This leads to the *fluence* condition:

$$\sigma_{\mathrm{ni}}\Phi_{\mathrm{i}}\tau_{\ell} > 1, \tag{11}$$

where τ_{ℓ} is the laser pulse length. For a typical nanosecond pulsed laser system of a laser ion source, this corresponds to a pulse energy requirement of the order of 10 mJ (100 W at 10 kHz). A comparison of this value with the laser parameters of the ISOLDE RILIS system (Table 1) indicates that such transitions are usually below saturation, and the efficiency of ionization is directly proportional to the laser power available for the non-resonant ionization step.

For a laser ion source installation, the final step power is usually supplied using the laser with the highest power available at a wavelength that puts the ionization potential (IP) of the excited atom within reach. Depending on the laser configuration, this could be a Nd:YAG laser with second harmonic output at 532 nm. Such a laser would be suitable provided that the previous transitions of the ionization scheme excite the atom to within 18 797 cm^{-1} of the IP.

Generally the photon absorption cross-section for excitation above the ionization potential falls with increasing energy. However, the spectrum of this continuum region may contain interference profiles, broad structures and the strong resonance peaks of auto-ionizing states (AIS). For some cases, the choice of the previous resonant step may depend more on the structure of the continuum and the resulting non-resonant ionization cross-section than on the strength of the intermediate transition.

Neglecting any structure within the ionization continuum, the photo-ionization rate for an unsaturated non-resonant step is directly proportional to the number of photons available. One should therefore consider that the number of photons in a laser pulse of a given energy is directly proportional to the laser wavelength: a 1 mJ pulse of a Nd:YAG fundamental beam (1064 nm) contains twice as many photons as a 1 mJ pulse of the second harmonic (532 nm). This factor may influence the choice of laser for the final step of the ionization scheme.

For a pulsed laser system, the duty cycle becomes an important consideration for the design or required specification of the laser ion source. This is discussed in section 2.6.

2.5.2 Auto-ionizing states

As mentioned, for many elements, a complex energy level structure exists above the ionization potential. Excited states within this region are known as auto-ionizing states (AIS). They involve the simultaneous excitation of two electrons, each to an energy level below the IP of the atom. If the total excitation energy is higher than the IP, de-excitation may occur via a non-radiative process in which the excitation energy of one electron is transferred to the other, resulting in ionization. This decay channel is rapid and the typical lifetime of an AIS is 10^{-15} to 10^{-10} s. The natural linewidths are therefore often very large (several gigahertz).

It has been observed that these states are more abundant for transitions from excited states with a complex electronic configuration involving inner-shell electron excitations. A good example is the ISOLDE RILIS ionization scheme for manganese [21].

A transition to an auto-ionizing state can exhibit a cross-section several orders of magnitude higher than for non-resonant transitions. In practice, this benefit can be lost due to the lower achievable laser power of a tunable laser compared to the pump laser that may be used for non-resonant ionization. The importance of these states is therefore very much dependent on laser system or application. Examples of ionization scheme development involving the search for an AIS are shown in section 4.

2.5.3 Rydberg state

Rydberg levels are a series of high-lying excited states with a high, and successively increasing, principal quantum number, n. The energy spacing of the series members is described by the Rydberg formula,

$$E_n = E_{\mathrm{IP}} - \frac{R_M}{(n - \delta(n,l))^2}, \tag{12}$$

where R_M is the Rydberg constant, and $\delta(n,l)$ is the quantum defect, an element- and atomic-orbital-dependent correction to the hydrogen atom Rydberg formula that compensates for the fact that the inner electrons of a multi-electron atom do not fully screen their associated nuclear charge. The observation of many Rydberg series members and the analysis of the energy level spacing enables a precise determination of the ionization potential of the atom. Recently, this technique has enabled the first measurement of the ionization potential of astatine [22, 23].

The photon absorption cross-section for transitions to the high-lying and abundant Rydberg levels can be large ($\sigma_{n\mathrm{R}} \approx 10^{-14}$ cm^2). Since the atomic radius is approximately proportional to n^2, an electron occupying a high-n level has a binding energy of the order of tens of millielectronvolts. This is minute compared to a typical transition energy of $\sim$3 eV. The lifetimes of Rydberg levels can be long (several microseconds) and increase approximately as the cube of the effective quantum number [24]. Owing to their long lifetimes and their energy proximity to the ionization potential, Rydberg atoms are sensitive to external effects and can be ionized by the following mechanisms.

Field ionization. If an atom is subject to a sufficiently intense electric field, its high-lying discrete states within the Rydberg series will merge with the ionization continuum. An electron occupying one of these states has a cross-section for dissociation that is dependent on the electric field strength and its principal quantum number. If a transition to a high-n state is used, the ionization efficiency can be high in even a weak electric field (>2000 V cm^{-1}).

Collisional ionization. For an electron in a state of high n, the collisional ionization probability can be extremely large due to three contributing effects. The first is a simple geometric contribution; the n^2 dependence of the atomic radius translates to a collision probability proportional to n^4. Secondly, the state lifetimes are usually long (of the order of microseconds), so the probability of

an atom collision before the spontaneous decay of the state can be high. Finally, the states are characteristically weakly bound, so the probability of ionization in the event of a collision is also high.

Blackbody radiation. If the excited atomic vapour is confined within a high-temperature cavity, the intensity of blackbody radiation from the cavity walls may be sufficient to ionize the atom, particularly for the higher-n series members [25].

2.6 RILIS applied to radioactive ion beam facilities

Photo-ionization was demonstrated by Letokhov in 1971 by the two-step ionization of rubidium [26]. This led to the later application of the technique as a laser ion source for efficient and selective ionization at ISOL facilities. Several designs were tested, and a summary of the development stages, along with references for each, is given in Ref. [27]. The first operational laser ion source at an ISOL facility was at the IRIS laboratory in PNPI, Gatchina, Russia [28]. Their system of recently modernized copper vapour pumped dye lasers is still in regular operation and was the basis for the first RILIS installation at ISOLDE [29]. Today, there are many laser ion sources in operation or under development at a variety of facilities worldwide, as shown in Fig. 7.

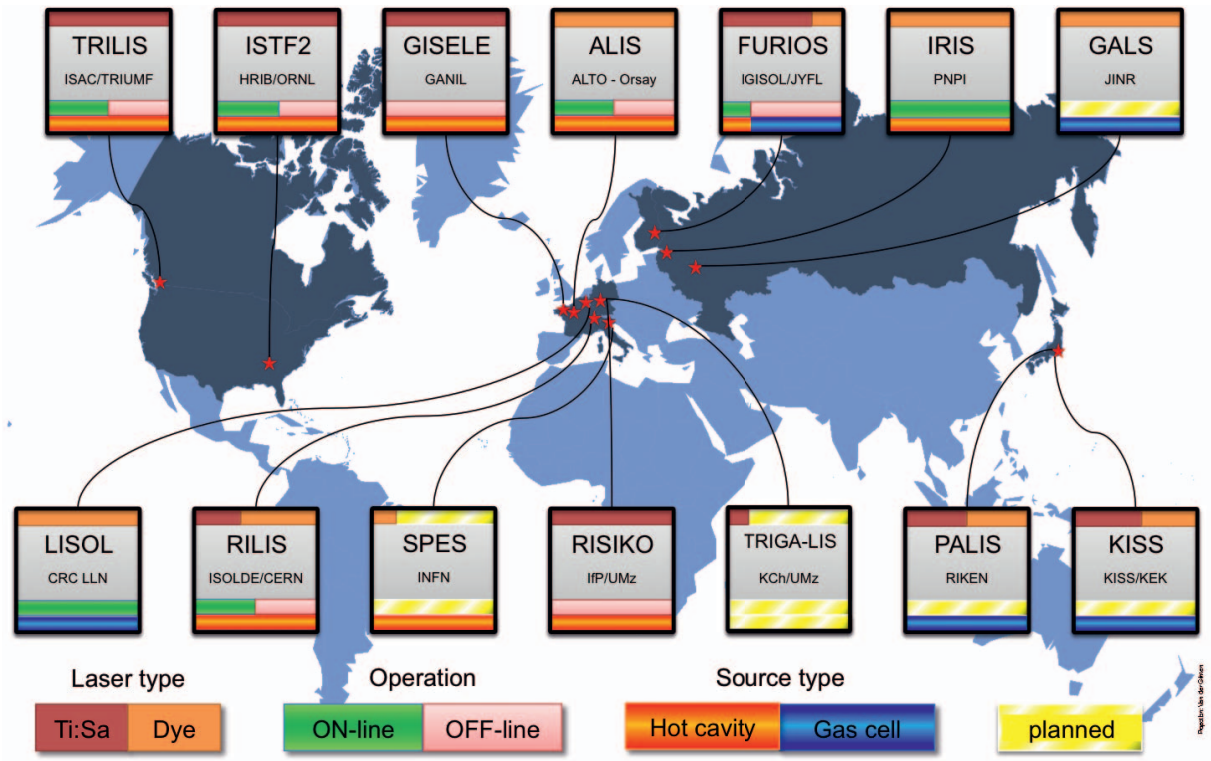


Fig. 7: Existing and planned laser ion sources at radioactive ion beam facilities around the world

As described in section 1.2, in the standard configuration, only thick-target ISOL facilities have a distinct ionization region that can easily accommodate the use of a laser ion source. The other types of RIB facilities directly extract ions that are created during the driver beam impact on the target. For these facilities, compatibility with a laser ion source requires the addition of a neutralization mechanism and a suitable laser interaction region.

This approach is illustrated in Fig. 8, which is an adapted version of Fig. 1 to show how the RILIS technique can be applied to all three types of RIB facility.

Common to all of these facilities is the prerequisite that there must be a clear line of sight to the

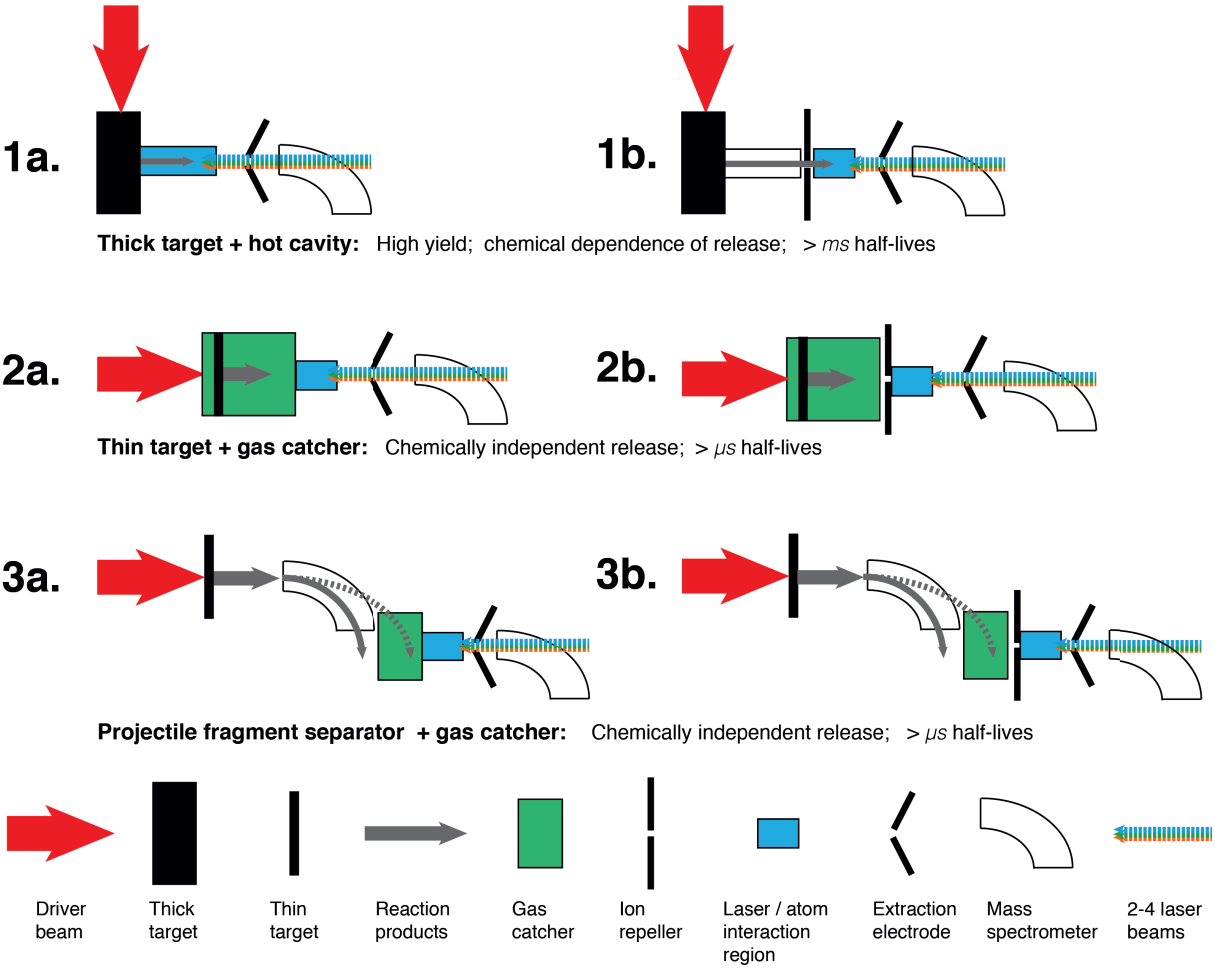


Fig. 8: Radioactive ion beam production methods compatible with resonance ionization laser ion sources

laser–atom interaction region. This means that an optical path must exist between the ion optics, isotope separator, beam diagnostics and radiation shielding constructions. The aperture sizes must be large enough to accommodate multiple laser beams, each with a diameter that avoids diffraction limitations. This is necessary for achieving a focal spot that enables a precise overlap with the atomic sample.

Any optical elements must be compatible with the full range of laser wavelengths that will be used. For broadband transmissive optics (prisms, vacuum windows, sampling plates) high-purity quartz such as Suprasil® should be used. Using narrower-band optics such as dielectric mirrors for beam transport can be more efficient but access to the full range of ionization schemes is likely to require changes of these elements. At some facilities this may be impracticable due to radiation safety considerations.

2.6.1 Overview of RILIS ion beams at isotope separator facilities

Figure 9 is an up-to-date compilation of all mass separated, laser ionized radioactive ion beams that have been produced at RIB facilities worldwide. Also shown are isotope beams that have been studied by resonance ionization spectroscopy as a secondary process, rather than during the initial radionuclide production and ionization stage. These techniques include thermal or pulsed laser ablated atomic beam studies [91], collinear resonance ionization spectroscopy [92] and resonance ionization mass spectrometry [93]. Typically these methods require an accumulation period of the radionuclide sample and/or an additional transport time after production to reach the experimental setup. This is reflected in the grouping of these isotopes closer to the line of stability than for the 'in-source' methods. The objectives

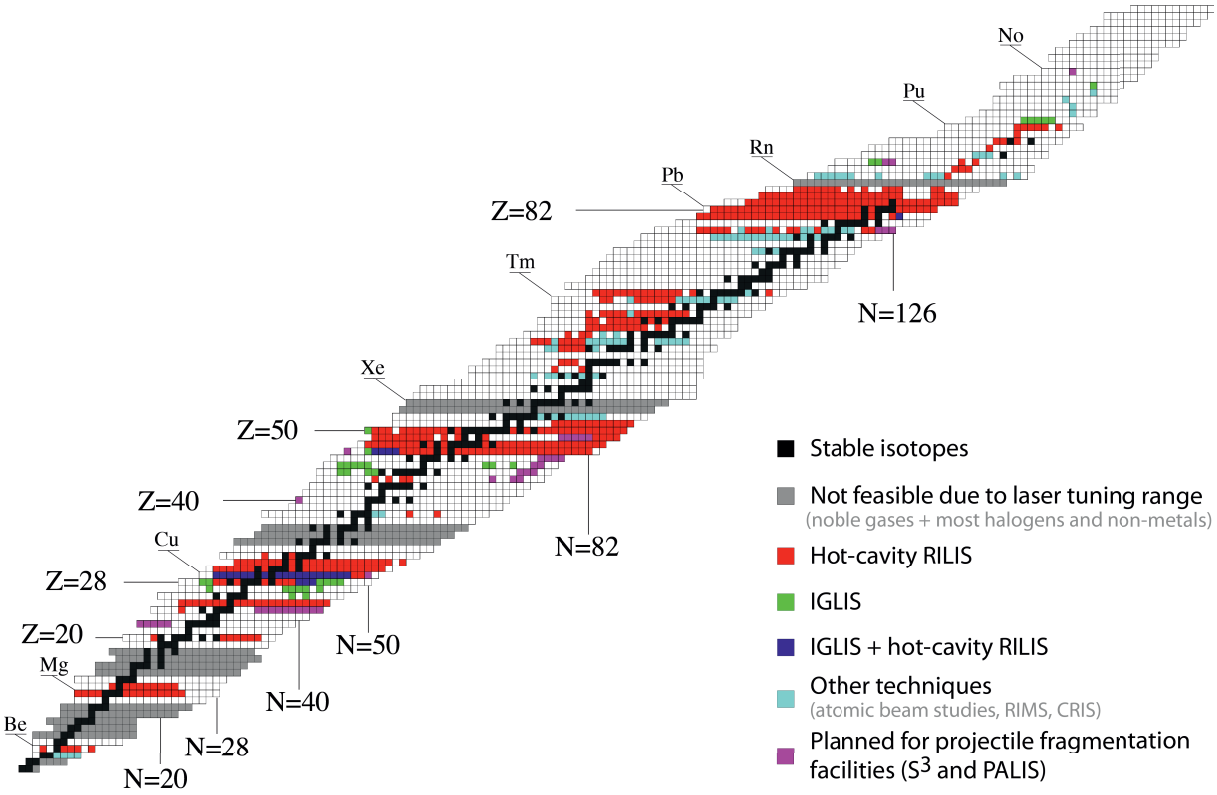


Fig. 9: The chart of nuclides showing the radioactive ion beams that have been resonantly ionized and mass separated at isotope separator facilities worldwide. These data are tabulated in [90] and Appendix A.

.

for the future IGLIS systems coupled to in-flight fragmentation facilities are also shown and highlight the complementary nature of this technique in providing access to beams of exotic isotopes of refractory metals as heavy isotopes (A >230) which are produced in projectile-target fusion reactions. Thick target ISOL facilities using the hot-cavity RILIS (HC RILIS) technique lay claim to the overwhelming majority of the laser ionized isotopes and so far radioactive ion beams of 39 elements have been produced. However, the restriction of this technique to the non-refractory metals and the inability to extract beams of isotopes with half lives <10 ms is evident. The GC RILIS technique has made an impact within these regions due to the rapid and chemistry independent nature of the radionuclide extraction process. The future GC RILIS projects (S^3 and PALIS) of the projectile fragment separators at GANIL and RIKEN promise to provide access to clean beams of many isotopes that have so far not been accessible with any other technique (green boxes in Fig. 9).

2.6.2 Thick-target, hot-cavity laser ion source

Atoms created inside a thick target reach the ion source by diffusion and effusion processes [30], as illustrated in Fig. 10. The rapidity of this transport is temperature- and material-dependent. The porosity or grain structure of the target and the geometry of the target–ion source assembly influences the path length required to reach the ion source. For short-lived species, finding the optimal balance between a high production rate and a short transport time is an ongoing field of research.

A requirement for high ionization efficiency is the irradiation of all atoms within the sample. If the atomic sample is provided as a continuous supply, the highest efficiency is achieved if the duty cycle losses are minimized; the importance of a high-repetition-rate laser system is easily calculable. The duty

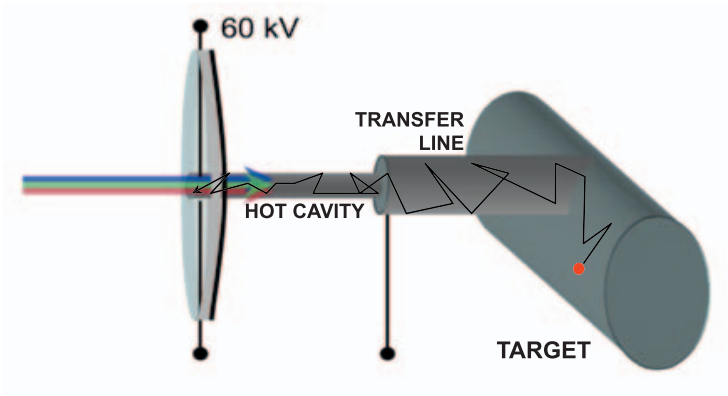


Fig. 10: An atom transported by diffusion and effusion to the laser interaction region of the hot-cavity RILIS: the surface ion source cavity.

cycle within the laser interaction region (ν_{L}) is given by

$$\nu_{\mathrm{L}} = f\tau, \tag{13}$$

where f is the laser repetition rate and τ is the transit time through the interaction region. To ensure at least one laser pulse–atom interaction, the duty cycle requirement is

$$f \geq 1/\tau. \tag{14}$$

The surface ionizer cavity of a thick-target ISOL facility can serve as the laser–atom interaction region and confines the atom sample within the 3 mm diameter, 30 mm long structure. The typical effusion time through the cavity is of the order of 100 µs. The laser repetition rate must therefore be at least 10 kHz to satisfy Eq. (14).

Atoms within the high temperature cavity environment have a broad thermal velocity distribution. This results in a significant Doppler broadening of the atomic spectral lines (>2 GHz for $A = 200$). Furthermore, for atoms with low-lying fine-structure levels close to the ground state, the high-temperature environment of the hot cavity will result in a relative population of these levels that follow Boltzmann statistics. Since the splitting is orders of magnitude larger than the laser linewidth, thermally excited atoms will not interact with a laser tuned to a first step transition from the ground state. Appendix C lists the populations of these levels for many elements at the typical hot-cavity temperature of 2200 °C.

These factors influence the choice of ionization scheme, set a laser line-width requirement (discussed later) and also limit the achievable resolution in resonance ionization spectroscopy experiments.

2.6.3 Thin-target, gas catcher laser ion source

The success of a laser ion source relies upon the availability of the sample in atomic form. Coupling laser ionization with a thin-target, gas-cell facility is therefore non-trivial. The first challenge is operating in 'atom guide' as opposed to 'ion guide' mode, requiring an adaptation of the gas-cell parameters to optimize the fraction of the reaction products in atomic form. Additionally, a good temporal and spatial overlap of the laser beams with the atomic sample must be achieved.

For laser access, the gas cell should be equipped with a quartz window to allow line of sight to the preferred laser–atom interaction region. Ideally, this region should be close to the exit nozzle of the gas cell for fast extraction of the laser ions and a reduced chance of neutralization. If laser access from the mass separator side of the gas cell along the ion beam axis is possible, then ionization outside the gas cell but within the gas jet, close to the exit nozzle, can be considered. These two options are shown as **2a** and **2b** respectively in Fig. 8. Figure 11 is an example showing both laser access routes that have been

used for the FURIOS laser ion source at the JYFL IGISOL facility in Finland [31]. The considerations for both laser–atom interaction options are discussed below.

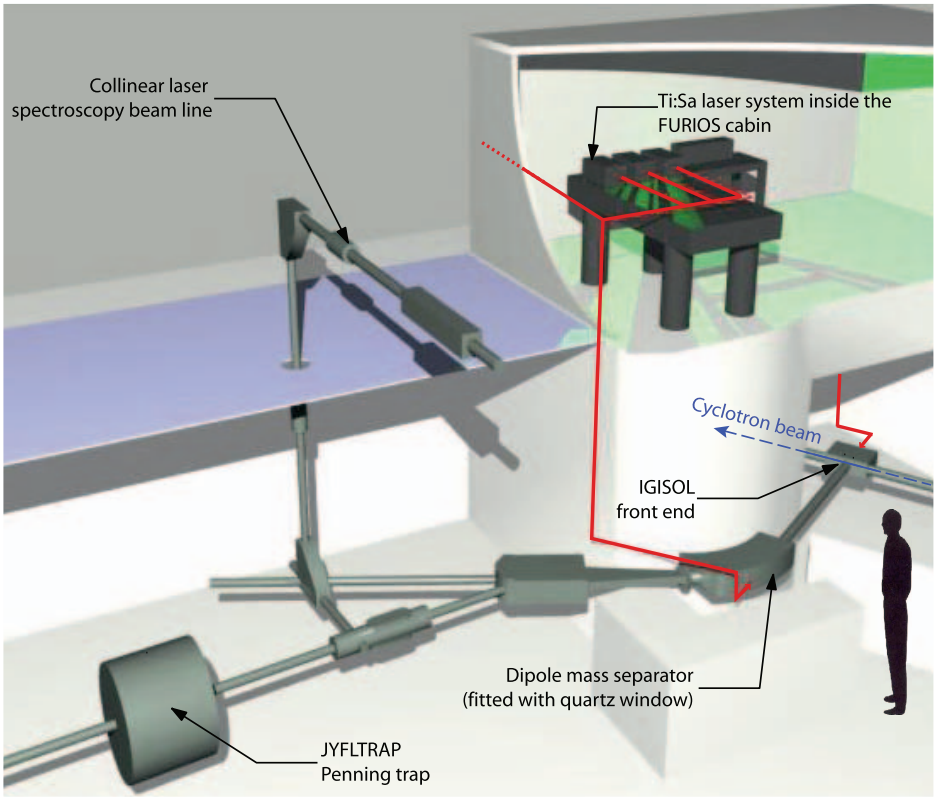


Fig. 11: The <2009 JYFL IGISOL beam-line layout showing the two laser beam access routes to the gas cell

2.6.3.1 Ionization inside the gas cell

Laser ionization in a gas cell (item **2a** of Fig. 8) has been extensively applied at the Leuven LISOL facility [32], where Ni, Co, Cu, Mn, Ti and Fe, and the refractory elements Rh and Re, have been ionized with efficiencies approaching 10% in the best cases. One of the striking findings from this development has been the high, and element-dependent, sensitivity to the presence of impurities in the buffer gas. Changes in gas purity, even on the parts per billion level, can have a dramatic effect on the photo-ion survival [33]. In the standard 'ion guide' mode of operation, the gas cell is optimized for the preferential release of reaction products in a 1+ charge state. In on-line operation, this process is somewhat dependent on the recoil energy and initial charge-state distribution of the reaction products, while the final charge-state distribution is a consequence of a series of charge-exchange collisions and recombination processes. The complexities are discussed in some detail in Ref. [34]. It is important to note that, for most elements, the second ionization potential is lower than the first ionization potential of helium or argon; therefore, the reduction to a 1+ charge state cannot be solely a result of collisions with these buffer gas atoms. In fact, the presence of molecular impurities such as O_2, H_2O, etc., greatly enhances the process of charge-state reduction.

The laser ion source efficiency is given by

$$\eta = \eta_{\text{neutralization}}\, \eta_{\text{ionization}}\, \eta_{\text{extraction}}. \tag{15}$$

This presents some difficulties when trying to optimize the gas cell for operation as a laser ion source. We have seen that the presence of impurities greatly increases the neutralization process, but, as demonstrated

at the LISOL facility, these impurities can also dramatically reduce the survival rate of the laser ions prior to extraction. To aid the neutralization process, the stopping power of the buffer gas can be improved by increasing the gas pressure. However, for spectroscopic applications of the laser ion source, where the goal is to precisely measure the positions of hyperfine components or the extent of the isotope shift, it is important to limit (or at least characterize) the pressure broadening of the spectral lines.

An additional complication is the effect of the plasma, generated by the driver beam impacting the thin target, on the neutral fraction or laser-ion survival within the gas cell. At the Leuven IGLIS, the 'dual chamber' gas cell was developed to screen the laser interaction volume from the production region and primary beam path. As a result, the plasma density at the location of the laser ionization is low enough for residual ions to be removed by applying electrical fields prior to the arrival of the laser pulses, thus greatly increasing the IGLIS selectivity; at LISOL a laser ON/OFF ratio of >2200 was achieved for ^{94}Rh, produced by fusion-evaporation using a ^{40}Ar beam incident on a ^{58}Ni target [35].

To establish the laser requirements for in-gas-cell ionization, the following differences with respect to hot-cavity ionization should be considered.

- Different spectral line-shape. Reduced Doppler broadening (room temperature) but instead pressure broadening and shifts dominate the distortion of the observed spectral line.
- Greater residence time. The atom evacuation time from the gas cell can be as large as several milliseconds, meaning that a lower-repetition-rate laser system can be used.
- Low temperature. The population of low-lying fine-structure levels is low. Only first step transitions from the ground state should be used for the ionization schemes.

2.6.3.2 Ionization outside the gas cell

With access through the dipole magnet and the lasers following a collinear and counter-propagating path with the ion beam (item **2b** of Fig. 8), the possibility exists for laser ionization outside the gas cell, within an interaction region close to the exit nozzle. Atoms are transported by gas flow from the gas cell into the laser interaction region immediately downstream of the gas-cell nozzle. Efficient ionization will therefore require a good geometrical overlap of the reaction products within the gas jet with the counter-propagating laser beams. This is not trivial, since the saturation of the atomic transitions requires a high photon flux, and therefore a small laser beam diameter. Furthermore, if the gas flow velocity is high (e.g., 1000 m s^{-1}), for an interaction region of 40 mm, a laser repetition rate of 25 kHz would be required for each atom to have at least one interaction with a laser pulse. It follows therefore that a long and well collimated gas jet should be created.

Several off-line studies of the gas flow characteristics outside gas cells have been performed and two methods of obtaining a suitably narrow gas-jet have been established [36].

A simple and universal solution is to use a capillary nozzle and a maintain a background pressure of several millibars in the expansion region. In this case the gas jet structure is a series of expansion and compression standing waves that encompass regions of gas in which the atoms have a very low energy spread. If there is simultaneous longitudinal and transverse laser access to the gas jet, a narrow linewidth excitation laser for the spectroscopic transition can interact in a high-resolution 'cross-beams' geometry with the low-energy spread portion of the supersonic gas jet whilst the broad linewidth laser beams for the subsequent excitation and ionization steps of the ionization scheme can overlap longitudinally with the gas jet. If a strong spectroscopic transition is used, an acceptable efficiency may even be achieved if a narrow linewidth tunable CW laser is used for the spectroscopic transition.

An alternative approach is to generate a collimated supersonic jet through the use of a Laval nozzle that is specifically tailored for a particular operating pressure. Test has shown that a jet of radius 3 mm and length 140 mm can be produced inside a region of background pressure of 0.3 mbar using a 250 mbar argon gas cell. This is a promising area of development for high-resolution laser resonance

ionization spectroscopy studies with gas cells [38]. For atoms transported inside the low divergence, low temperature ($\sim$ 5 K) and high velocity supersonic gas jet the Doppler and pressure broadening of the spectral lines is low. High sensitivity collinear resonance ionization spectroscopy studies can therefore be performed with unprecedented spectral resolution (<200 MHz).

In summary, ionization outside the gas cell offers the following advantages:

- pure isotope beams are possible through the complete suppression of the isobar background of gas-cell ions by using an ion collector or repeller;
- potential for high-resolution 'in-source' RIS on short-lived isotopes (sub-millisecond).
- with a high repetition rate laser system (>10 kHz), the efficiency should be similar to that of in-gas cell laser ionizaton.

A thorough description of state-of-the-art laser ionization techniques that have been applied for gas cells is provided elsewhere [37, 38].

2.6.4 Projectile fragment gas catcher laser ion source

The implementation of a laser ion source at a fragment separator facility is a novel idea under development at the SLOWRI facility [39] in Japan. This project, the Parasitic Laser Ion Source (PALIS) [40, 41] will require a gas cell positioned close to the second focal plane of the BigRIPS [42] fragment separator. The gas cell stops a portion of the ion beam that would otherwise be removed by a slit for beam purification. This would enable a low-energy gas-cell-based experiment to operate parasitically alongside a normal BigRIPS run.

A similar technique is proposed for coupling a gas cell to the S3 fragment separator facility at SPIRAL2, France [94].

Although the isotope production method is different, the laser ion source considerations for this type of facility, illustrated by items **3a** and **3b** of Fig. 8, are the same as for a standard gas-cell ISOL facility.

3 The CERN ISOLDE RILIS

In this section, the application of the RILIS technique to the ISOLDE facility is explained. It is intended as a guide to the RILIS installation and a source of information and references that should be relevant for aspects of other hot-cavity- or gas-cell-based laser ion source operation and development. ISOLDE [43] is a thick-target ISOL facility that uses the 1.4 GeV proton beam of the CERN Proton Synchrotron Booster (PSB) to produce radioactive isotopes in a chosen target material by spallation, fragmentation and fission reactions. There are two 'front ends' housing independent target/ion source assemblies. Figure 12 is a photograph of an ISOLDE target unit. Downstream of each target is a corresponding mass separator: the General Purpose Separator (GPS) or the High Resolution Separator (HRS). These are indicated in Fig. 13.

The RILIS laser installation is located in the experimental hall above and adjacent to the shielding for the separator areas. Laser access to the ion sources of both targets is possible through independent laser paths through the windows of each 90° separator magnet. The optical path lengths are $\approx$18 m (GPS) and $\approx$25 m (HRS).

The demand for RILIS ionized beams increases annually and is challenging to fulfil: the number of operating hours during the 2012 on-line period exceeded 3000, and ion beams of 13 of the 28 elements that have so far been ionized on-line were produced. To accommodate this demand, the capabilities of the RILIS have been increased through technical enhancements and ionization scheme developments. A stepwise upgrade of RILIS that began in 2007 was completed in 2011.

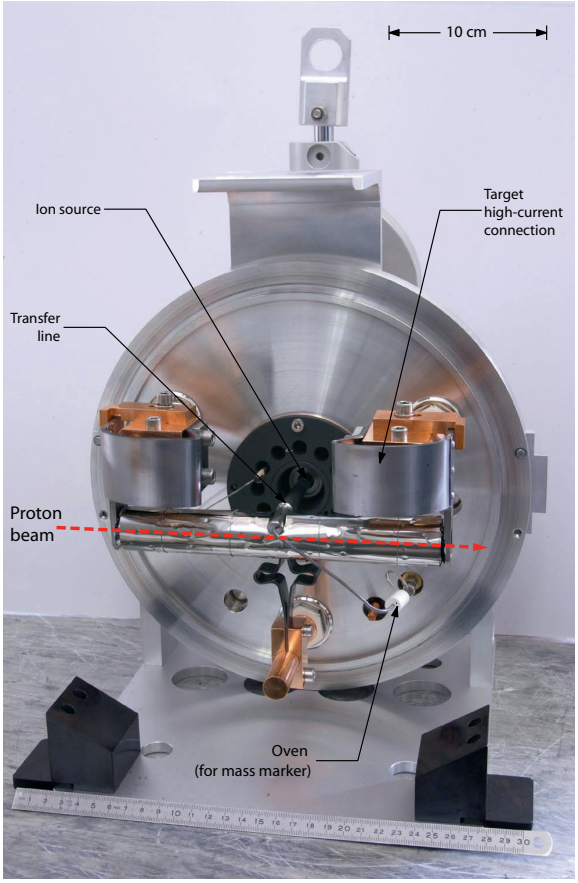


Fig. 12: The ISOLDE target and surface ion source assembly

The technical details of the ISOLDE RILIS are documented in several publications. Together these show the evolution of the laser installation over the various upgrade steps:

1. the original ISOLDE RILIS installation [29];
2. copper vapour laser (CVL) → Nd:YAG pump laser upgrade [21];
3. Nd:YAG pumped dye laser upgrade [44];
4. complementary Nd:YAG pumped Ti:Sa laser installation [45];
5. operation of the dual Ti:Sa and dye laser RILIS system [44].

3.1 The hot-cavity ionizer

The standard surface ion source cavity is a 30 mm long, 3 mm diameter, capillary made from a refractory metal (tungsten, tantalum or rhenium). A high cavity temperature is needed in order to confine the reaction products without losses due to wall sticking. Optimizing the cavity temperature for maximum laser-ion survival also enhances the standard surface ion source operation, and so ions can be created with a high degree of efficiency on the surface of the cavity. The RILIS can benefit greatly from the ionizer 'cavity effect', which is documented by Kirchner [6]. This is the transverse potential well that results from the electron emission from the cavity walls. The net positive charge on the walls gives a degree of transverse confinement for the laser ions, greatly improving their extraction efficiency.

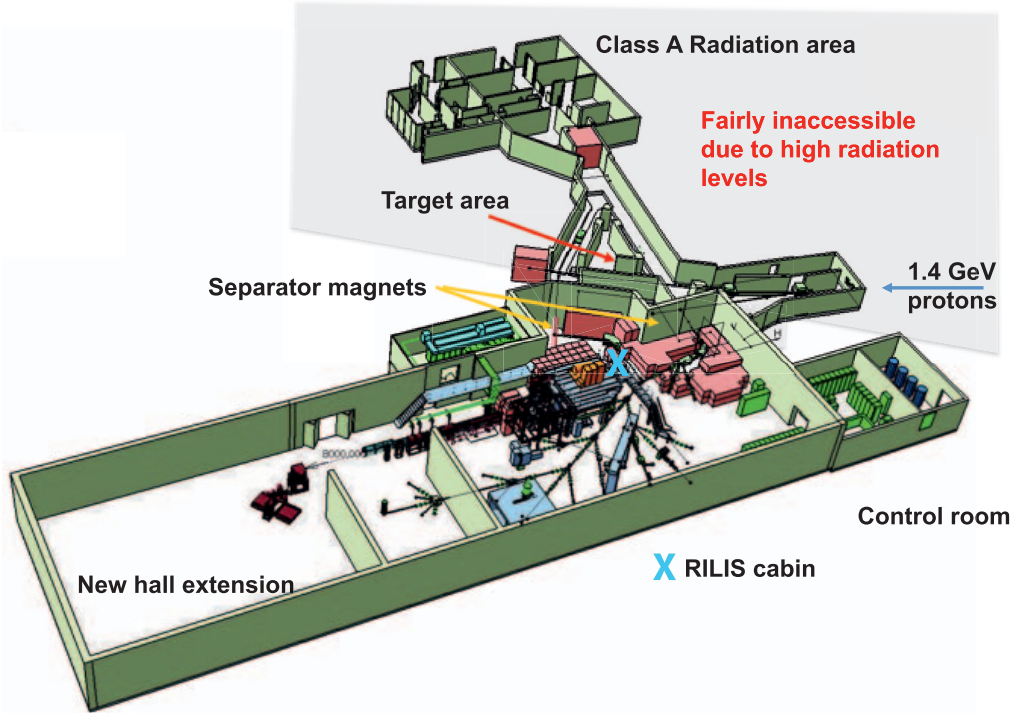


Fig. 13: Schematic of the ISOLDE laboratory. The area marked in grey is inaccessible during on-line operation due to the high levels of activity.

At the typical operating temperature of >2000 K, the Doppler-broadened linewidth exceeds 1 GHz, even for the heaviest elements that are produced. The thermal distribution of the atoms can also result in a significant population of the higher-energy fine-structure components of the atomic ground state. In this case a proportion of the atoms are unavailable for ionization if only one laser is available for a first excitation step of an ionization scheme. A table of thermal populations of low-lying atomic levels for many elements is given in Appendix C.

The hot-cavity environment sets the following laser requirements for efficient ionization.

Linewidth. This must exceed the effective spectral linewidth of the atomic sample (>1 GHz in the hot-cavity environment).

Repetition rate. The typical effusion time of 100 µs means that the laser system should operate at >10 kHz so that each atom 'sees' a laser pulse.

Beam quality[1] **and low divergence.** It must be possible to transmit the laser beam efficiently to the 3 mm cavity at a distance of up to 25 m from the focusing optics.

The influence of the cavity dimensions on the laser ionization efficiency is discussed in Ref. [29]. Neglecting any detrimental effect of increased cavity length (such as radioactive decay of the isotope of interest due to increased effusion time), an increased cavity length and narrow cavity diameter is favourable. In most cases, focusing the laser beams into a cavity of diameter less than 3 mm is impractical at ISOLDE. Furthermore, increasing the cavity length can cause problems related to the mechanical stability under thermal stress and may therefore lead to reliability issues.

[1]The M^2 factor of a laser beam is a wavelength-independent measure of beam quality that is given by the ratio of the beam parameter product (BPP) of the laser beam to that of an ideal Gaussian beam. The BPP is the product of the laser beam divergence and the radius of the beam waist. For a Gaussian beam, this is equal to λ/π. The minimum value of M^2 is unity, corresponding to a single-mode TEM_{00} Gaussian laser beam. For many applications, the M^2 is a convenient quantity to use since it is directly proportional to the area of the laser focal spot.

The influence of the inevitable surface ionization processes inside the hot-cavity ionizer tube on the RILIS selectivity is discussed in section 5.

3.2 The laser system

Figure 14 is a schematic of the current RILIS installation, with details of the optical path length and beam distribution to the HRS target. Six tunable lasers are installed: three dye lasers and three titanium–sapphire (Ti:Sa) lasers. Details of these systems are given in the following sections.

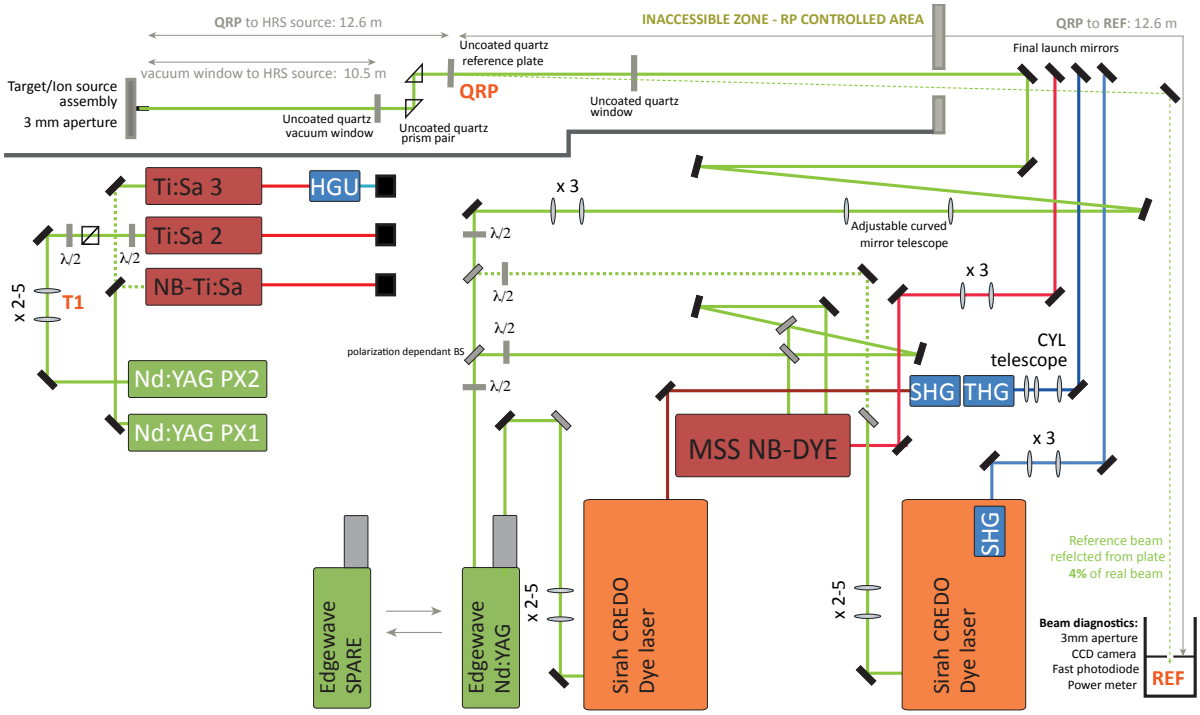


Fig. 14: The layout of the ISOLDE RILIS, including the optical elements of the HRS beam transport system

3.2.1 Pump lasers

Each tunable laser requires an energy (pump) source to reach and maintain a population inversion in the laser gain medium. In this case, the pumping energy is delivered by a non-tunable, high-power laser. Depending on the laser configuration and the ionization scheme, the pump laser may also be used for non-resonant ionization.

The fundamental requirements for the pump laser are as follows.

1. **Wavelength**
   (a) In the absence of nonlinear effects, for energy conservation, the pump wavelength must be shorter than the required laser emission wavelength.
   (b) For efficient energy transfer, the pump laser wavelength should be close to the peak of the absorption spectrum of the tunable laser gain medium.
2. **Power**
   The pump power should be high enough to achieve the required tunable laser output power (pumping efficiencies are typically in the range of 10–30%).
3. **Pulse length**
   The pulse length must suit the application. For dye laser pumping or non-resonant ionization, a

short pulse is required (<20 ns). This is because of the high gain and short energy storage time of the laser dye. The dye laser pulse length is approximately equal to the pumping pulse length. Short pulse length, and therefore high peak power, is preferable for saturating transitions or for non-resonant ionization. Ti:Sa crystals have a long 'build-up' time (>100 ns) and lower gain. In this case the pump energy can still be absorbed efficiently for longer pulses (>100 ns). The Ti:Sa pulse length is defined by the cavity dimensions and is not shortened by the use of a short pump pulse. In this case a longer pump pulse length is favourable because the lower peak power reduces the likelihood of damage to the optics or the Ti:Sa crystal.

4. **Repetition rate**
   As explained in section 2.6, a repetition rate of 10 kHz is required for compatibility with the 100 μs residence time of an atom in the hot cavity.
5. **Synchronization**
   (a) The delay between the pump laser and the tunable laser emission should be considered. If the tunable laser output is delayed by more than 10 ns, then the use of the same laser for pumping a different type of tunable laser, or for non-resonant ionization, would require long and impractical optical delay paths.
   (b) If several pump lasers are used, then they must be synchronized with low pulse timing jitter. The output pulse jitter with respect to an external trigger signal must be small compared to the pulse length.
6. **Beam quality,** M^2
   If the pump laser is to be used for non-resonant ionization, then the beam quality and low divergence requirements mentioned previously apply, so that efficient transmission to the 3 mm ionizer cavity can be achieved.

The ISOLDE RILIS pump laser specifications are given in Table 1. The lasers are all diode-pumped Nd:YAG lasers operating at a 10 kHz repetition rate. In principle, all four pump lasers can operate simultaneously, and it is possible for them to be synchronized with a low pulse timing jitter. For the Ti:Sa system, the use of two synchronized pump lasers simplifies the task of maintaining temporal overlap of the output pulses of two Ti:Sa lasers. In practice, however, this has been rarely applied and usually only one pump laser is used for each tunable laser system. The primary purpose for the pump laser redundancy for each system is to enable rapid replacement, minimizing downtime in the event of a laser malfunction.

Table 1: The ISOLDE RILIS pump laser specifications.

	Dye pump Nd:YAG			Ti:sapphire pump Nd:YAG	
	Beam A	Beam B	Beam C	Laser 1	Laser 2
Wavelength (nm)	532	532	355	532	532
Repetition rate (kHz)	10	10	10	10	10
Average power (W)	0–80	0–40	0–20	0–74	0–72
Pulse duration (ns)	8	9	11	128	170
Pulse jitter (ns)	3	3	3	9.3	8
Divergence (mrad)	0.2	0.2	0.3	7.6	6.4
Beam quality, M^2	<1.3	—	—	≈30	≈30

The 532 and 355 nm output beams are the second and third harmonics of the fundamental output wavelength of Nd:YAG lasers. These are generated by nonlinear frequency conversion processes (described in section 3.3). The dye pump lasers, supplied by Edgewave GmbH, each have three output beams (two at 532 nm and one at 355 nm). This is achieved by the use of three harmonic generation

crystals (SHG1, SHG2 and THG): the full-power fundamental beam enters SHG1 to produce beam A and a residual fundamental beam, which is sent to SHG2 to produce beam B. Beam B and the residual fundamental beam both pass through the THG crystal for the production of beam C, if required. The efficiency of each harmonic generation stage is controlled by temperature tuning of the corresponding crystal to the phase matching condition. This is a convenient way of distributing the power output for a particular application.

The Ti:Sa pump lasers are Photonics Industries DM-60 models, designed to provide >60 W at 532 nm each. The long pulse length and high M^2 makes these lasers unsuitable for non-resonant ionization or dye laser pumping, but such parameters are favourable for the pumping of the RILIS Ti:Sa lasers.

Owing to the limited space in the RILIS laboratory, an additional pump laser requirement for ISOLDE is a small laser head footprint. Both the Edgewave and the DM-60 lasers are particularly small compared to other commercially available ones that offer similar specifications.

3.2.2 Tunable lasers

Table 2 is a list of the main specifications of the ISOLDE RILIS tunable lasers. An overview of the tuning ranges and achievable power levels of the fundamental and multiple harmonic beams is given later in Fig. 22.

Table 2: Key parameters of the tunable lasers at the ISOLDE RILIS installation.

	Dye lasers		Ti:sapphire lasers
	Credo	DMK	
Tuning range (nm)	390–860	390–860	680–950
Linewidth (GHz)	9	15 or 0.8	5 or 0.8
Maximal power (W):			
Fundamental	20	10	6
Second harmonics	2.5	1	1
Third harmonics	0.2	0.2	0.15
Fourth harmonics			0.15
Pulse duration (ns)	7	10	35
Divergence (mrad)	<1	<1	<1

3.2.2.1 Dye lasers

The application of tunable dye lasers for atomic spectroscopy is discussed in Ref. [46]. References [47] and [48] explain the principle of several diffraction grating tuned pulsed dye laser configurations.

The RILIS dye laser system consists of two Credo broadband (≈9 GHz) lasers from Sirah Lasertechnik GmbH. These are in an oscillator and single-pass amplifier configuration. The oscillator is equipped with an angle-dependent resonator 'end mirror' in the form of a diffraction grating with groove spacing of 1800 lines/mm. Figure 15 shows the laser cavity design, grazing incidence grating–tuning mirror combination, and the frequency doubling crystal with a compensator prism [49].

In the standard configuration, the Credo laser linewidth is 0.08 cm^{-1} (≈2.4 GHz). For general RILIS use, a larger linewidth is preferred and the lasers were purchased with a non-standard intra-cavity beam expansion factor that increases the linewidth to ≈9 GHz by constricting the width of illumination of the diffraction grating. The increased linewidth offers several advantages: higher efficiency for the excitation of atoms with a broad hyperfine structure; reduced sensitivity to isotope shifts; and improved ion current stability in the event of small wavelength fluctuations or drifts.

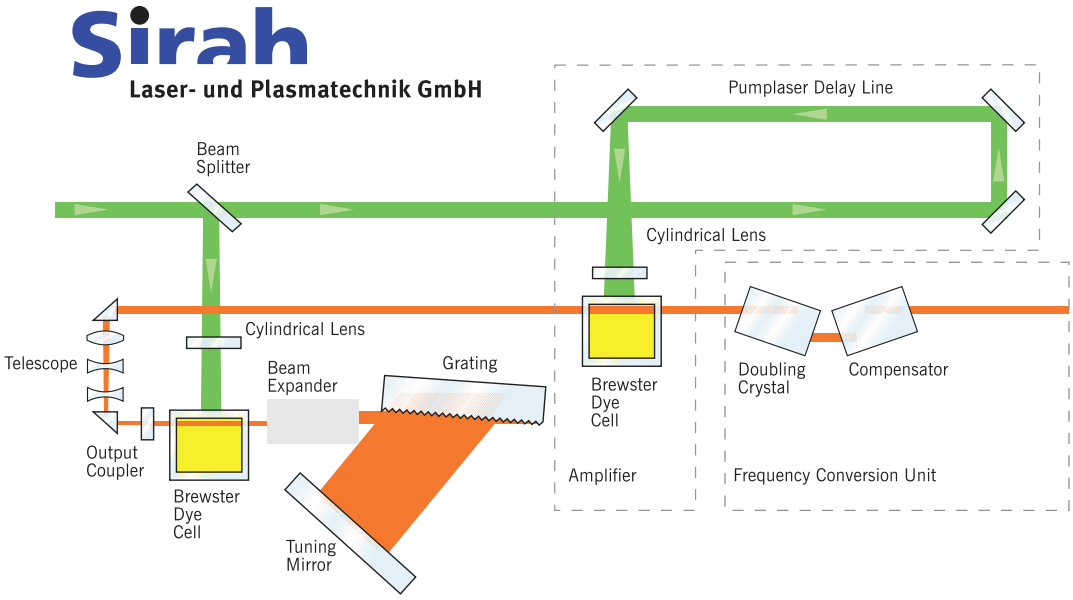


Fig. 15: The Sirah Credo dye laser optical layout [49]

The third dye laser is supplied by DMK Laser Microsystems. It also uses oscillator and single-pass amplification configuration with a similar specification to those of the Credo laser. The DMK laser also has reduced linewidth (0.7 GHz) option. The linewidth reduction is achieved by inserting a 6 mm thick glass Fabry–Perot etalon into the laser cavity.

All three dye lasers are equipped with pump optics that are coated for UV (355 nm) or green (532 nm) pumping. Coverage of the full range of the dye laser spectral range requires many different dyes. A rather comprehensive set of dye tuning curves can be found at the website of Exciton (http://www.exciton.com/laserdyescurves.html) [50]. Figure 16 shows examples of typical optical layouts for dye laser second and third harmonic generation. These techniques are necessary for producing dye laser radiation between 210 nm and 390 nm.

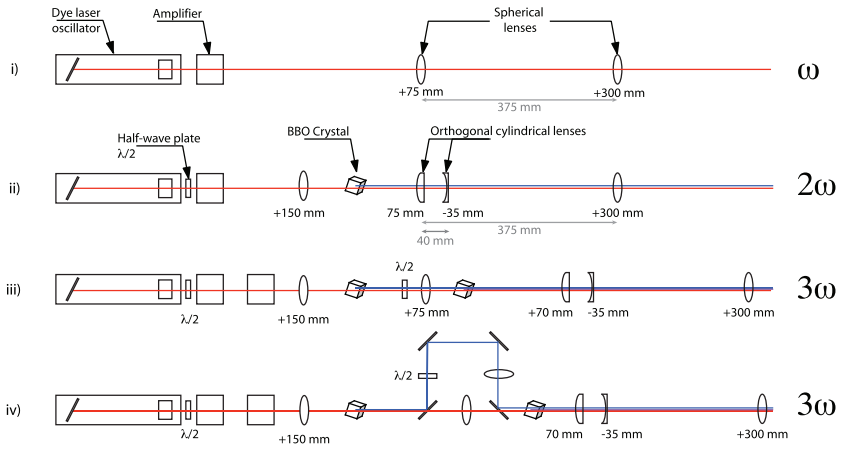


Fig. 16: RILIS dye laser beam focusing and harmonic generation options

3.2.2.2 *Titanium–sapphire (Ti:Sa) lasers*

Titanium-doped sapphire crystals exhibit a gain bandwidth that exceeds that of any individual laser dye. Because of this, the most popular application of Ti:Sa lasers [51] is for the generation of ultra-short (femtosecond) pulse lengths using mode-locking techniques. This broad bandwidth makes Ti:Sa crystals a suitable gain medium for a broadly tunable laser with pulse lengths in the nanosecond range, to give a spectral coverage of $\sim$ 690–950 nm. The Ti:Sa lasers of the ISOLDE RILIS are a recent adaptation of the University of Mainz design [52], with specific modifications to improve their use for intensive on-line operation at ISOLDE [45]. New features for improved ease of set-up, wavelength measurement and maintenance of the laser performance have been added. These modifications are apparent in the Ti:Sa photograph of Fig. 17, which is displayed alongside a labelled schematic of the laser that shows the following components:

MP high-reflectivity 532 nm pump mirrors;

L 1 inch, $f = 80$ mm pump beam focusing lens on an XYZ adjustable mount;

M2, M3 curved resonator mirrors ($R = 75$ mm) – the pump beam passes through M3;

M1 high-reflectivity cavity end mirror;

M4 output coupler;

BS beam sampler, to send a portion of the Ti:Sa output to the wavemeter (WM) fibre coupler – the BS is mounted on a magnetic pedestal and so can easily be exchanged for a He–Ne mirror to send a fibre-coupled He–Ne alignment beam into the resonator cavity;

MB broadband mirror to direct the sampled beam to the wavemeter;

BF birefringent filter, which is a broadband frequency selection device;

E etalon, to provide an additional degree of frequency selection.

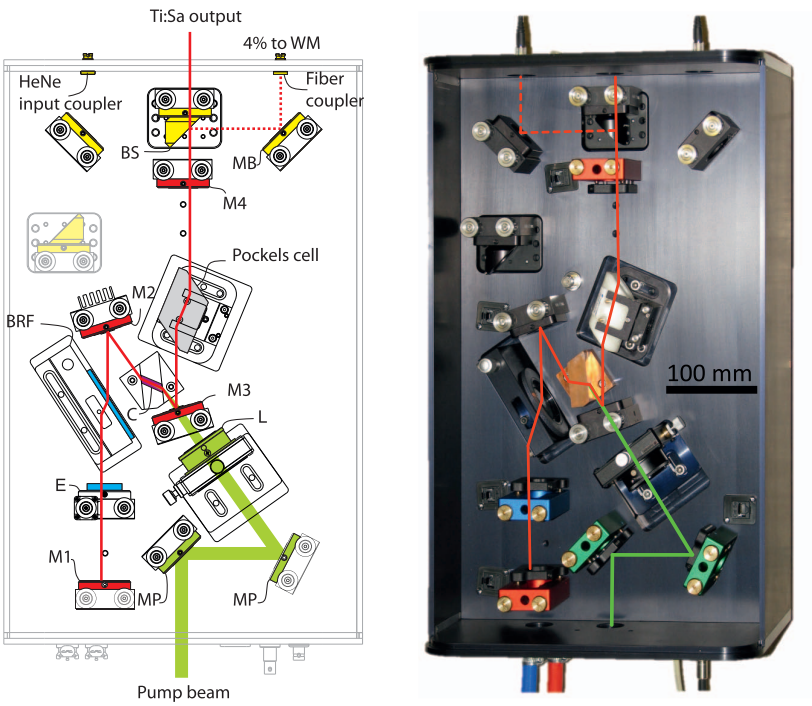


Fig. 17: A schematic and a photograph of the CERN Ti:Sa laser. New ease-of-use features such as the colour-coded mirrors, wavemeter pick-off, fibre-coupled helium–neon laser alignment input and pre-aligned optics inserts can be seen [45].

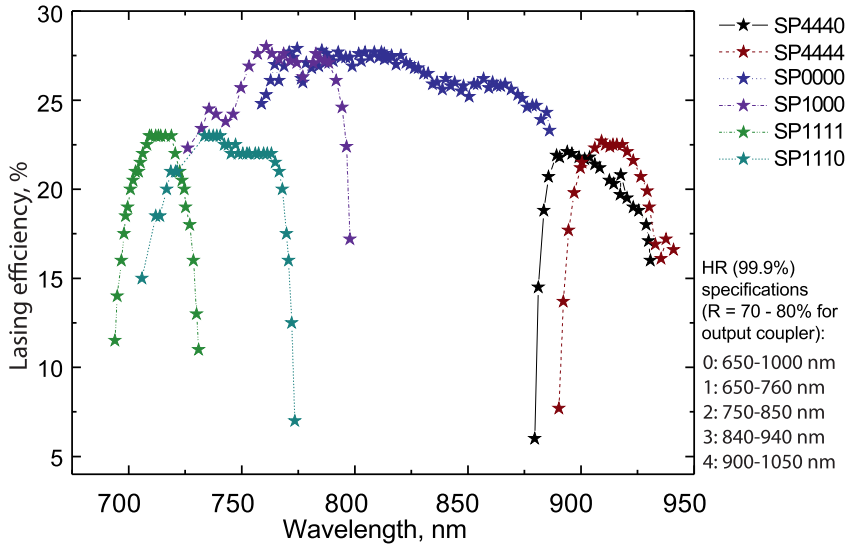


Fig. 18: Tuning curves for the different mirror sets for the CERN Ti:Sa laser. Pump power is 10 W. The Pockels cell and etalon have been removed.

Although it is possible to use a broadband mirror set that can cover the full Ti:Sa tuning range, for optimal and reliable performance, particularly at high pumping powers, several sets of cavity mirrors (**M1–4**) are used. The tuning ranges for each mirror set is shown in Fig. 18.

3.3 Harmonic generation

At the fundamental frequency, the spectral range of tunable laser systems is not sufficient to enable their application for efficient resonant ionization of a wide range of elements. In fact, for many elements, suitable first excited states lie at an energy that is far into the UV region, corresponding to a wavelength less than half of that achievable at the short-wavelength end of the tuning range of a dye-laser system.

These transitions are, however, accessible if the second or third harmonic of the fundamental emission frequency is generated. For laser light, the generation of multiple harmonics within certain optical materials is possible due to the nonlinear effects resulting from the characteristically high electromagnetic fields that can be generated.

3.3.1 Principle

When light passes through an optical material, a polarization response to the electric field occurs. At low laser intensity, the polarization of transparent materials can be considered to vary linearly with the applied field:

$$\mathbf{P} = \chi \varepsilon_0 \mathbf{E}, \tag{16}$$

where χ is the electric susceptibility and $\mathbf{P}$ is the polarization vector (the dipole moment per unit volume that is induced in a material due to an applied electric field).

For optically isotropic materials, χ is constant. In a nonlinear optical material (anisotropic with respect to the refractive index), χ becomes a tensor accounting for the multiple-order polarization responses of the material.

If the electric field is sufficiently high (e.g., a pulsed laser beam focused to a narrow beam waist), nonlinear interactions must be considered and Eq. (16) must be generalized to

$$\mathbf{P} = \varepsilon_0(\chi_1 \mathbf{E} + \chi_2 \mathbf{E}^2 + \chi_3 \mathbf{E}^3 + \cdots + \chi_n \mathbf{E}_n). \tag{17}$$

The polarization is a consequence of charge oscillations within the material induced by the intense oscillating field generated by the laser light. The radiation of these oscillations effectively adds to the electric field. The scalar quantities, χ_1, χ_2 and χ_3, are material properties of linear susceptibility and second- and third-order nonlinear susceptibilities, respectively. For some materials, the nonlinear susceptibilities are sufficiently large for their polarization effect to be comparable to that for the linear term. For laser light of angular frequency $\omega = 2\pi\nu_1$, the electric field E can be described by

$$E = E_0\,\mathrm{e}^{-\mathrm{i}\omega t} + E_0^*\,\mathrm{e}^{+\mathrm{i}\omega t}. \tag{18}$$

For an optical material with a non-zero value of χ_2, the second-order nonlinear polarization term given by $P_2 = \varepsilon_0\chi_2 E^2$ becomes

$$P_2 = 2\varepsilon_0\chi_2 E^* E + 2\varepsilon_0\chi_2 E^2\,\mathrm{e}^{-\mathrm{i}(2\omega)t} + 2\varepsilon_0\chi_2 (E^*)^2\,\mathrm{e}^{\mathrm{i}(2\omega)t} \tag{19}$$

for which the time-dependent terms have characteristic frequencies of precisely twice the frequency of the incident laser light. Under certain conditions, these terms can result in a significant conversion of the fundamental beam to a second beam with twice the original frequency. This was first demonstrated by Franken *et al.* [53]. The main difficulty associated with second harmonic generation (SHG) lies in the frequency dependence of the refractive index of the material. This results in incoherent production of the second harmonic beam at different positions within the material. This can severely affect the conversion efficiency due to destructive interference of the second harmonic beam. This can be avoided if the phase velocities of the fundamental and the second harmonic are the same. For a uniaxial nonlinear crystal, it is possible to select an angle of propagation for which the refractive index for the extraordinary ray (n_e) is equal to the refractive index of the ordinary ray (n_o). This technique, known as *phase matching*, is illustrated in Fig. 19 and is described in some detail in Ref. [19]. The efficiency of second harmonic generation depends on the intensity of the fundamental beam and the interaction length l:

$$I_{2\omega} \sim l^2 I_\omega^2 \epsilon_{2\omega} \propto PL. \tag{20}$$

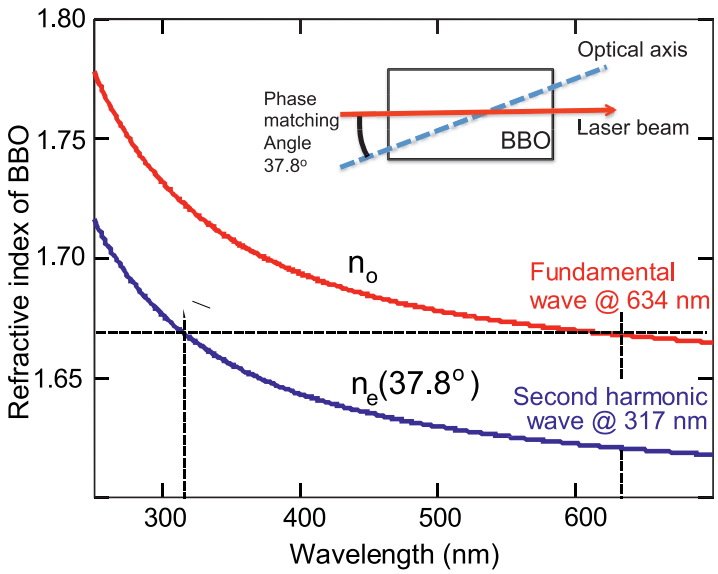


Fig. 19: An example of achieving the phase matching condition for the second harmonic generation of a 634 nm beam inside a BBO crystal. BBO (β-barium borate) is a negative uniaxial crystal ($n_\mathrm{e}(\omega) < n_\mathrm{o}(\omega)$). At an angle of propagation of 37.8° with respect to the optical axis, $n_\mathrm{e}(2\omega) = n_\mathrm{o}(\omega)$ and phase matching is achieved.

A consequence of the angle tuning that is required to achieve phase matching is that fundamental and second harmonic beams do not propagate collinearly. This is known as 'walk-off' and limits the

conversion efficiency by reducing the effective value of l^2 in Eq. (20). It also gives rise to an elliptical output beam. If the maximum value of l^2 is reached, an increase in conversion efficiency requires an increase in laser intensity (increased focusing of the input beam).

The principle of second harmonic generation can be generalized to account for two incident beams with different frequencies. If phase matching is achieved, it is possible to generate a beam at a frequency equal to the sum of the frequencies of the two incident beams. This technique is known as 'sum frequency' generation. A two-step sequence of 'second harmonic generation' followed by 'sum frequency generation' is commonly used for the generation of the third harmonic of the dye or Ti:Sa laser output. An early example of this method, applied to generate a 229 nm beam for the first step transition of cadmium, using the CVL pumped dye laser system at the ISOLDE RILIS, is described in Ref. [54].

3.3.2 Suitable nonlinear materials

The selection criteria for a suitable nonlinear optical medium for harmonic generation for the dye or Ti:Sa lasers of a laser ion source are the following:

- wide transparency range (low absorption from 200 to 1000 nm);
- relatively low temperature sensitivity, for reliable operation under fluctuating power levels or laboratory conditions;
- high damage threshold (in the range of GW cm^{-2}), due to the requirement for a high laser intensity (Eq. (20));
- high nonlinear coefficients, χ_2 and χ_3;
- broad phase matching range.

These criteria limit the choice of crystal types to a small number. The most commonly used crystals are BBO (β-barium borate, BaB_2O_4), BiBO (bismuth borate, BiB_3O_6) and LBO (lithium triborate, LiB_3O_5). BiBO crystals have the highest nonlinear coefficient of these (1.5 times higher than for BBO, 3.5 times greater than for LBO). BiBO is also insensitive to moisture, which gives a long-term reliability advantage over BBO. On the other hand, BBO and LBO are cheaper and have a higher damage threshold. At ISOLDE, BBO is the most commonly used crystal type for use with the tunable laser system. These crystals are all classed as 'negative uniaxial' ($n_{\mathrm{e}}(\omega) < n_{\mathrm{o}}(\omega)$). The polarization of the second harmonic beam is rotated by 90° with respect to the fundamental beam if type-I phase matching is used, and for third harmonic generation both input beams need to have the same polarization.

A free program called **SNLO** (http://www.as-photonics.com/snlo) is a useful non-linear crystal configuration tool which can help in the optimal selection of crystals for a particular application.

Because the fundamental tuning range of the Ti:Sa laser is in the red to infrared part of the spectrum, frequency doubling, tripling or quadrupling is essential for all ionization schemes if only Ti:Sa lasers are used. A custom-made harmonic generation (FCU) unit was developed at the University of Mainz to be used with their Ti:Sa lasers (Fig. 20). The ISOLDE RILIS has two updated versions of this device. The CERN FCU incorporates some changes to ergonomics and the beam shaping optics, which, based on $ABCD$ matrix calculations, results in reduced astigmatism of the third and fourth harmonic beams.

Set-up **A** in Fig. 21 is the standard scheme, which uses S_2 for compensating the divergence of the beams exiting the first BBO crystal (due to the focusing with S_1). The elliptical second harmonic (blue) beam is adjusted with C_1 to produce a beam aspect ratio close to 1 in the far field. S_3 focuses the blue beam at the location of the second crystal. S_4 is adjusted so that the foci of both beams coincide for THG in the second crystal. S_5 and C_2 perform the equivalent beam shaping of the UV beam as S_2 and C_1 did for the blue beam, to produce a suitable aspect ratio for focusing the UV beam the ion source using the spherical lens telescope ($S_6 + S_7$).

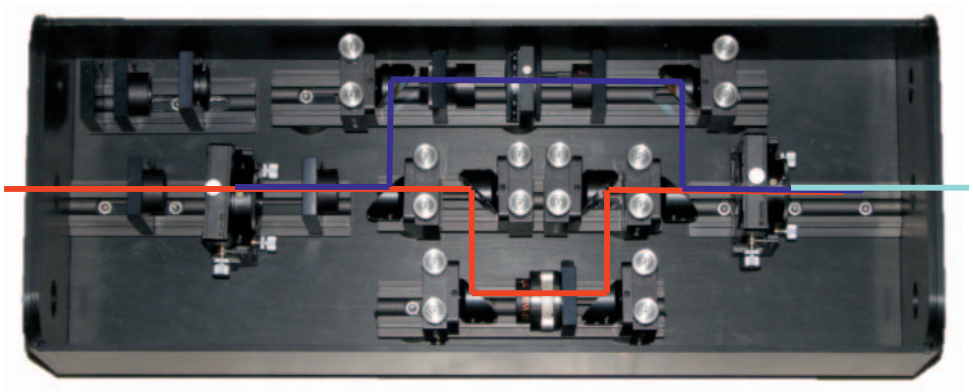


Fig. 20: The modified CERN version of the Mainz frequency conversion unit (FCU). The optical paths for the fundamental (red), second harmonic (dark blue) and third harmonic (light blue) beams are shown.

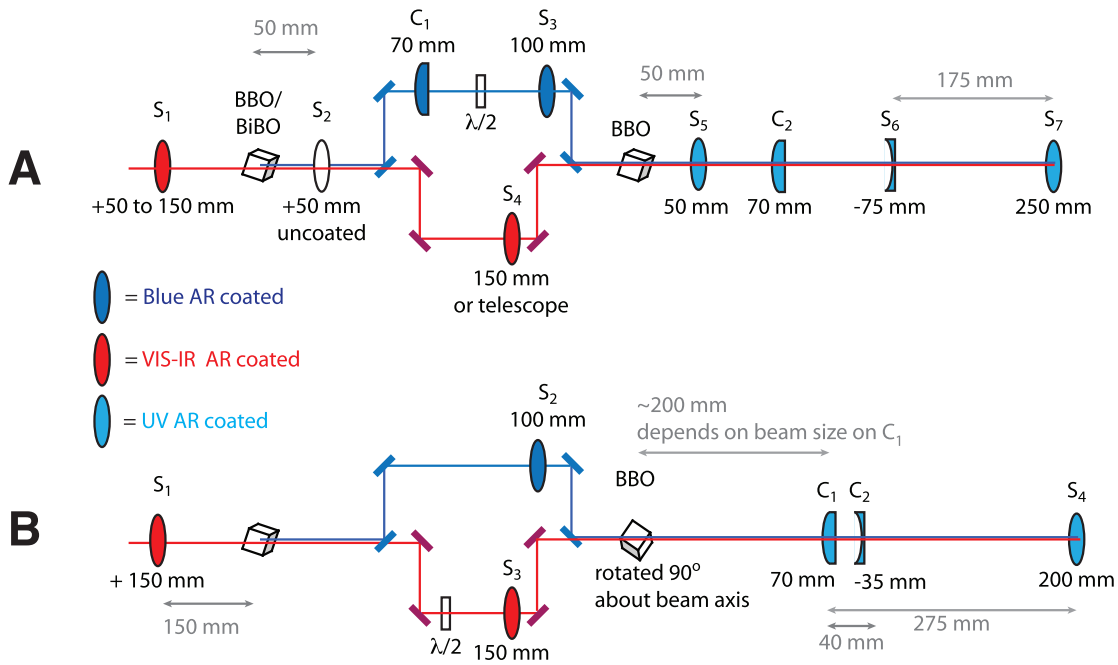


Fig. 21: Possible optical configurations for third or fourth harmonic generation using the Ti:Sa FCU. In this example the unit is configured for THG. For FHG the optical set-up is the same but the fundamental beam (red) is blocked after the first BBO crystal. Spherical lenses are labelled S_n and cylindrical lenses are labelled C_n.

Set-up **B** in Fig. 21 is an alternative, more simple optical set-up, similar to that used for the dye laser system, shown in Fig. 16. This configuration works particularly well if S_1 has a focal length of >100 mm. In this case the divergence of the uncompensated blue and red beams emerging from the first crystal does not result in a beam size that is larger than the surface of the lenses S_2 and S_3. Although this configuration does not enable correction of the elliptical blue beam, the larger beam diameters at S_2 and S_3 result in a smaller diffraction-limited spot size at the THG crystal. The half-wave plate is placed in the red beam to further reduce losses of the weaker blue beam and the tripling crystal is rotated accordingly. A further simplification is the configuration of the UV beam shaping and focusing optics. A cylindrical lens telescope ($C_1 + C_2 + S_4$) enables independent control of the X and Y axes of the elliptical UV beam. The telescope must be placed close to the output of the FCU box so that the UV beam spot is not larger

than C_1. Similarly, the diameter of S_4 must be at least 50 mm to avoid losses. In practice, using set-up **B**, a higher harmonic generation efficiency is usually achieved and the time required for set-up and optimization is greatly reduced. This is due the reduced number of degrees of freedom and fewer losses on optical surfaces. Whilst it is likely that the maximum achievable THG/FHG efficiency is greater using set-up **A**, the increased number of parameters to optimize makes this hard to achieve in practice.

3.4 The dual RILIS

The advantages of a dual system of complementary tunable lasers (dye plus Ti:Sa) are numerous [44], but the main benefit is the spectral tuning range of the complete system, as shown in Fig. 22. Complete coverage of the visible range is possible along with gap-free access to wavelengths down to almost 200 nm. To maximize the potential of the system, the ability to synchronize the output pulses of each laser with negligible jitter is essential. To achieve this, the pump lasers are triggered from a precise common master clock and delay generator. This device, a Quantum Composers® 9520, can provide output laser trigger pulses that can be delayed with 250 ps precision and <50 ps jitter. This has enabled the operation of the dual RILIS in the following four modes.

Dye-only mode. The dye laser system provides all the excitation steps of the ionization scheme. During this mode of RILIS operation, the Ti:Sa system can be prepared for the ionization scheme for the next element in the ISOLDE schedule, reducing the switching time between elements.

Ti:Sa-only mode. This is the Ti:Sa-only equivalent of the previous mode.

Mixed dye plus Ti:Sa mode. The ionization scheme uses the optimal combination of dye and Ti:Sa lasers, depending on the suitability of each laser type for a particular required wavelength.

Backup mode. The ionization scheme has one or more excitation steps that are provided by both the dye and the Ti:Sa lasers simultaneously. This is beneficial, as it increases the likelihood of saturating transitions for which the power output for a single laser is low (if third or fourth harmonics are required) or because of the improved reliability, and reduced downtime during laser maintenance tasks such as the replacement of used laser dyes.

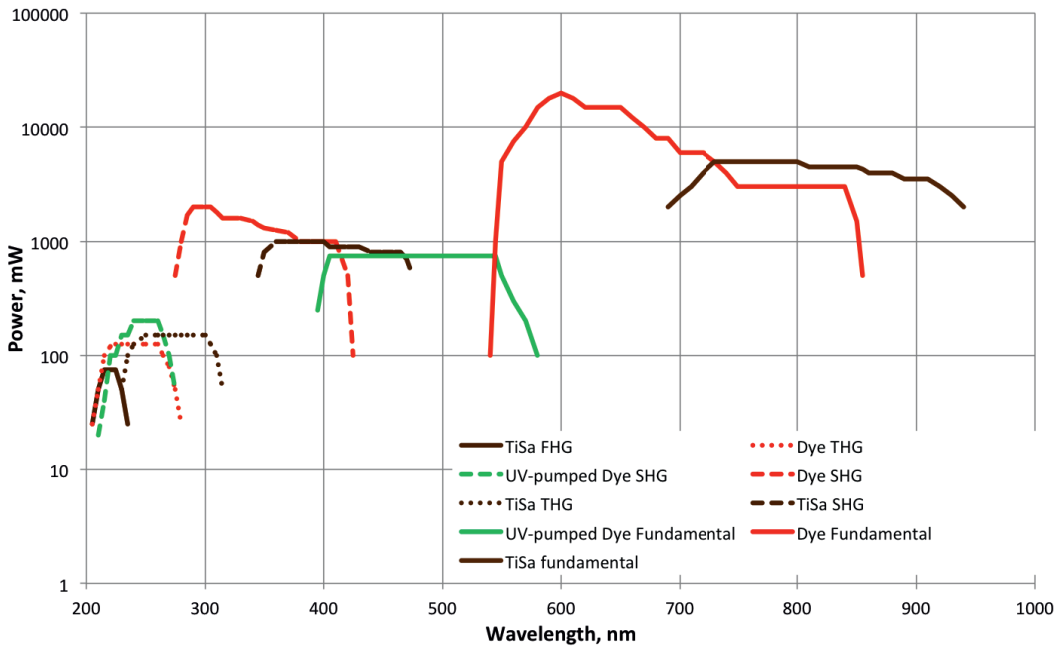


Fig. 22: The tuning range of the dual dye and Ti:Sa RILIS system, including second, third and fourth harmonics

3.5 Beam transport

The optical path for beam transport to the ion source should be as simple as possible whilst still enabling sufficient shielding from the high-radiation target and separator environments. For any new facility, such an optical path needs to be foreseen during the initial design phases of the laboratory. At ISOLDE, two laser paths are required so that access to each target is possible.

For the full range of ionization schemes, laser wavelengths in the range of 210–1000 nm are required, and for the RILIS ionization schemes, up to three laser beams are required simultaneously.

Because access to the optical elements outside of the laser laboratory, behind the radiation shielding, is not possible during on-line operation, a simple and universal system of beam steering optics is required. At ISOLDE, the solution is the use of high-quality uncoated fused silica prisms. These have to be large enough to accommodate all of the laser beams. Fused silica is chosen for all broadband transmissive optics (windows, lenses, prisms) due to its unique combination of the following properties: low absorption across the full wavelength range of the laser ion source; low thermal expansion coefficient; and relatively low refractive index. In addition, the internal absorption for all wavelengths must be low to avoid losses or thermal lensing effects.

At the TRIUMF TRILIS facility in Vancouver, Canada [55], an alternative approach is used to eliminate the ≈4% losses that occur at each uncoated surface–air interface.[2] By using separate sets of high-reflectivity broadband dielectric mirrors, the laser beams can be transported with very low loss for the specified wavelength range of the optics. The beams are then combined at the position of the mass separator using dichroic optics or polarizing beam splitter cubes. At facilities where the frequency of RILIS use and the range of elements to be produced is lower, or if the access conditions for changing optics are less restrictive, then this solution is advantageous due to the improved transport efficiency. Low-loss beam transport is particularly important for a Ti:Sa-only RILIS system because of the greater reliance on harmonic generation for reaching the required wavelength.

The fused silica window of the vacuum chamber through which the laser beams enter has a free aperture diameter of >20 mm, which is slightly larger than the space occupied by the multiple laser beams at this position. The beam size reduces and the spatial overlap improves as the beams converge upon the ionizer cavity after a further ≈8 m path length. At each separator window, a shutter is installed within the vacuum chamber. When the RILIS is not operating, this shutter blocks the window from potential ion beam deposits on the vacuum-side surface. Nevertheless, the window transmission is reduced over time due to excessive deposition of particles on the surface and is changed every year. This task involves work in a high-radiation-dose-rate environment, and opening the magnet vacuum chamber involves a contamination risk. The window therefore has to be easily accessible and a procedure for rapid exchange has to be established.

Figure 14 shows the optical elements in the path of the RILIS beams towards the HRS target, located ≈23 m from the exit of the final telescope (**T2**) on the RILIS laser table. The laser beams pass through a 2 inch diameter uncoated fused silica plate **QRP**, which is mounted 12.6 m from the ion source. The surface reflections from the uncoated quartz plate **QRP** produces a set of reference beams. The angle of **QRP** is adjusted so that they are directed to the reference area, **REF**, inside the RILIS room, via a 2 inch diameter UV-enhanced aluminium mirror that is mounted on the HRS launch mirror support. The location of **QRP** is chosen so that the optical path length for the reference beams from **QRP** to **REF** is equal to the distance travelled by the main beams from **QRP** to the ion source. Figure 23 shows the optical layout of the reference area. In this example the system is configured for the individual monitoring and stabilization of two laser beams. The laser beams converge upon a 3 mm diameter aperture at the reference point. This is representative of the hot-cavity entrance aperture. A power measurement taken after this point therefore enables the power transmission to the ion source to be calculated. Before

[2]The reflectivity of an uncoated optical surface is determined by the Fresnel equations and is dependent on the refractive index of the material. For a fused silica–air ($n_{\mathrm{FS}} = 1.461$) interface, there will be a 3.6% reflection of a 532 nm beam at 0° incidence.

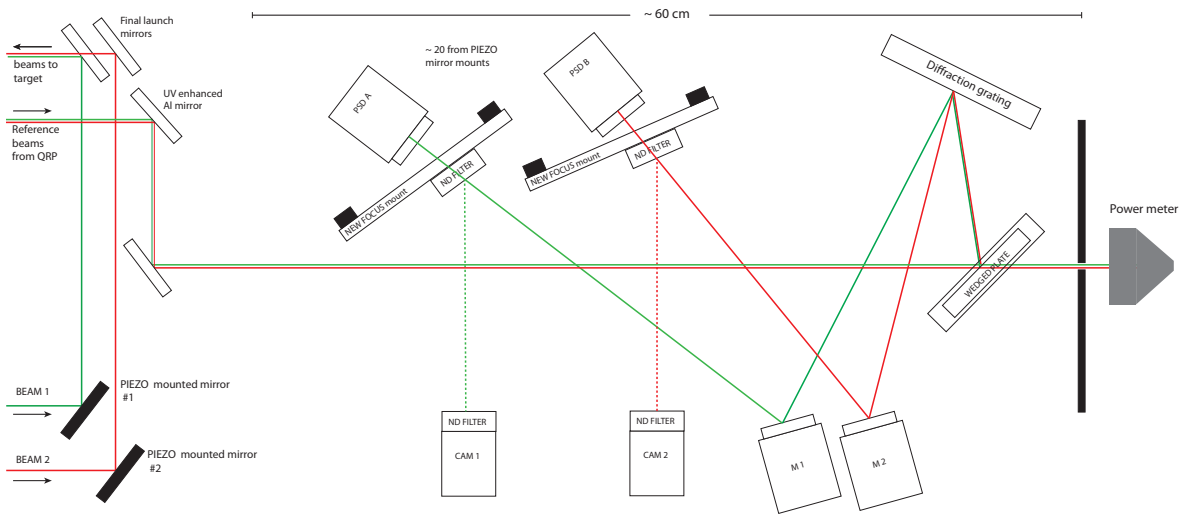


Fig. 23: The tuning range of the dual dye and Ti:Sa RILIS system, including second, third and fourth harmonics

Table 3: Edgewave and Blaze laser beam transport efficiency for the 25 m HRS optical path based on reference beam measurements.

	Laser power (W) (on laser table)	Reference beam power (mW) 3 mm aperture	Full beam	Transport efficiency (%) (correcting for 34% surface losses)
Edgewave	43	370 (21.5%)	470 (27.3%)	33
Blaze	15	350 (58%)	—	88
Blaze	40	800 (50%)	1000 (62.5%)	76

entering the aperture, the beams pass through a wedged beam sampler plate. The surface reflections from this plate are directed to a diffraction grating, which splits the overlapping beams due to its wavelength-dependent angle of reflection. This enables each beam to be directed to a separate monitoring and stabilization system. The beam size and position can be monitored using a charge-coupled device (CCD) camera and a position-sensitive detector. If used in combination with a launch mirror mounted on a piezo-actuator, rapid (10–50 Hz) laser beam position stabilization can be established. The stabilization system used at ISOLDE is a commercial device (Aligna® by TEM Messtechnik GmbH). Alternative solutions are available from several other companies (e.g., MRC Systems GmbH). Since the optical path for a laser ion source is long, the system required only two-dimensional (position) stabilization, but the large distance between the active mirror and the ion source necessitates the use of piezo- rather than pico-motor actuators.

The data in Table 3 show recently measured values of the beam transport efficiency for the Edgewave laser. Once the surface losses of 34% (10 uncoated surfaces in the optical path) are subtracted from the initial power value, the transmission efficiency through the 3 mm reference hole is 33% for the Edgewave beam. The poor efficiency for the Edgewave beam transport is due to losses and poor focusing of parts of the multi-component Edgewave beam. A beam diameter larger than the various aperture restrictions of the transport optics is required to achieve maximum transmission through the reference hole without diffraction limitations. The transmission data for a Blaze 532-40-HE diode-pumped Nd:YVO_4 laser, supplied as a loan by Lumera Laser GmbH, are also shown. The Blaze laser has a lower M^2 than the Edgewave laser ($\sim$1.1) and a circular beam shape. For this laser, 88% transmission through the reference hole could be achieved with an initial beam diameter of <10 mm, significantly smaller than any aperture restrictions in the optical path. Further details of a test of the Blaze 532-40-HE laser, conducted at the ISOLDE, can be found in Ref. [56].

3.6 Ion beam detection

Typically, the RILIS parameters are optimized under off-line (no protons on target) conditions using a heated sample of the element of interest. The sample is referred to as a 'mass marker' and is a specially prepared known quantity (usually several thousand nanoamp hours) of the element of interest that is placed in a resistively heated capillary 'oven' container, which is connected to the transfer line of the target unit. The use of an independently controlled oven ensures that a reliable and adjustable supply of the element is available during an experiment, facilitating the initial and regular optimization of the laser parameters. The oven is heated to provide a steady atom supply that results in an ion current that exceeds the measurement threshold of an ISOLDE Faraday cup (FC) (>0.2 pA)[3] [57]. The RILIS parameters (beam positions, pulse timing, focusing and wavelengths) are optimized whilst monitoring the ion current of the mass-separated beam. For exclusively radioactive elements (such as At and Po), a 'mass marker' is not available and the supply of atoms has to be provided by in-target production. Depending on the release and decay properties, this can be done either during proton taking or shortly after an accumulation of the radiogenic material. If the ion rate is below the FC detection limit, then the laser optimization has to be determined using a more sensitive device such as a microchannel plate (MCP) detector, or by characteristic radiation detection.

3.6.1 Beta and gamma decaying isotopes

A tape station (Fig. 24) semi-permanently attached to the ISOLDE beam-line, can be used to measure β- or γ-active nuclei. A short bunch of the ion beam is implanted into a tape, which is then moved into a 4π β-detector or a germanium scintillator detector for γ detection. The yield is obtained by integrating the measured release curve. This method is feasible for nuclei with a half-life that is greater than the minimum tape transport time ($\sim$100 ms) but not so long (several hours) as to cause long-term contamination of the tape. Depending on the decay chain, for shorter-lived nuclei it may be possible to determine the yield by measuring the activity associated with the decay of the daughter nuclei. The realization of a reliable fast tape station for sub-100 ms measurements is an ongoing project at ISOLDE.

3.6.2 Alpha decaying isotopes

For α-active isotopes, the alpha decay spectroscopy system of the IKS Leuven nuclear spectroscopy group (Fig. 25) is often used. It is referred to as the *windmill* because it contains a rotating wheel of up to ten carbon catcher discs. These are made at GSI [59] and have a diameter of 6–10 mm and a thickness of 20 μg cm^{-2}. At two of the ten disc locations, the wheel is surrounded by a pair of Si α detectors. Three of these are circular detectors with an active area of 300 mm^2, placed as close as possible to the catcher disc. The fourth, placed 4 mm in front of the implantation site, is a larger diameter (24 mm) annular detector that allows passage of the ion beam, whilst maximizing the α detection solid angle ($\approx$66%). A germanium detector can be placed behind the implantation site, at the position of the narrow-walled end flange, for gamma detection.

The alpha count rate can be shared with the RILIS control PCs through the use of a laser scan program that is part of the LabVIEW-based RILIS control infrastructure [95] This enables synchronized exchange of RILIS and windmill variables (laser wavelength, power, gated α-count rate), the generation of a RILIS-ready transistor-transistor logic (TTL) signal and the plotting of live wavelength versus α rate spectra. For radiogenic isotopes, any measurement to determine the ion rate should be made with knowledge or control of the incident proton beam characteristics so that the influence of proton beam fluctuations during a laser scan can be accounted for.

Example of two alpha spectra acquired using the 'windmill' system during an in-source spectroscopy study of neutron-deficient Bi isotopes are provided in Fig. 32 of Section 6. The alpha energy resolution is sufficient to allow clean gating on the isomer or the ground-state alpha decay channels.

[3] With an average proton beam of 2.5 μA, 0.2 pA corresponds to a yield of about 5×10^6 per microcoulomb of protons.

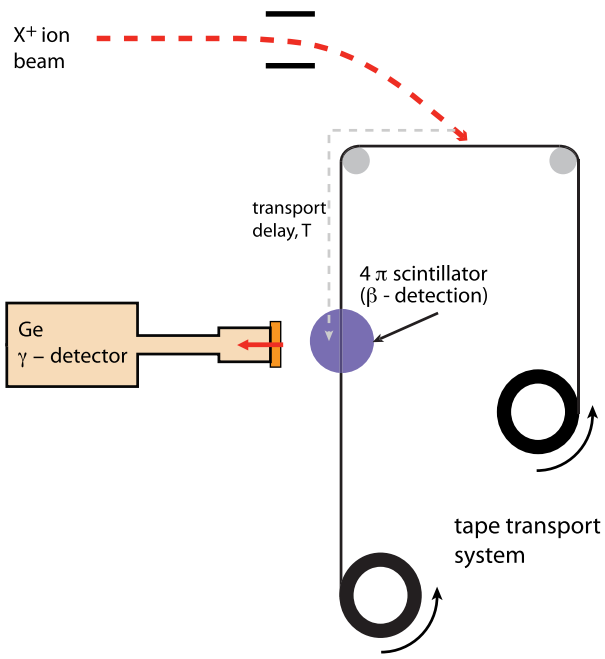


Fig. 24: The ISOLDE tape station used for yield and lifetime measurements of β and γ decaying isotopes

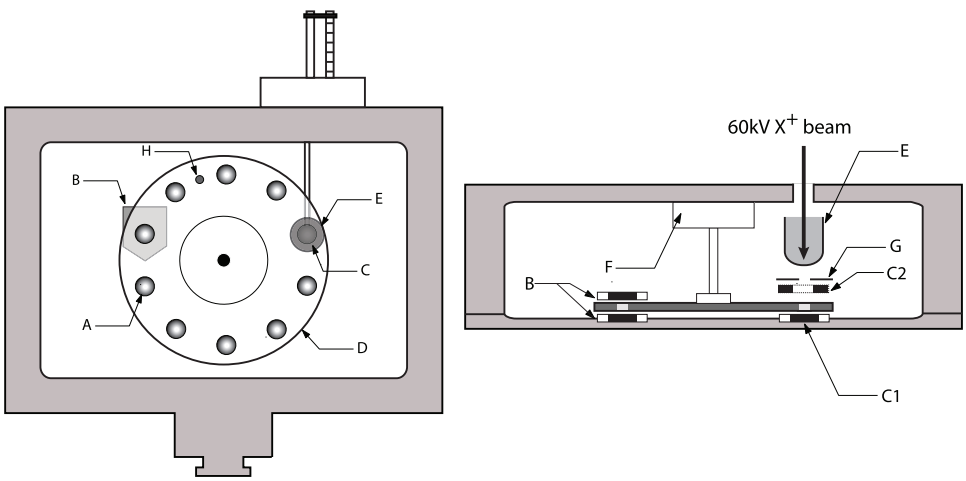


Fig. 25: The KU Leuven windmill alpha detection device. The ten carbon discs (A) are mounted on a wheel (D) that can be rotated using a stepping motor (F). The incident beam current can either be measured with a movable Faraday cup (E) or be sent through a collimator (G) onto the foil. A pair of Si detectors (C1 and C2) measure the alpha spectrum during implantation on the foil. A second detector pair (B) is used to measure a previously implanted foil to verify its cleanliness before re-use. A 30–50 Bq ^{241}Am source (H) can be moved to the detector locations for calibration purposes.

3.7 Typical annual operation

As the most commonly used ion source of the ISOLDE facility, the RILIS is in a virtually uninterrupted cycle of set-up and operation during the on-line period. In 2012, for example, RILIS-ionized beams of 13 elements were produced for 24 on-line runs. This required approximately 3000 hours of laser operation, for which 344 RILIS operator shifts were scheduled. This intensity of use is unique among laser ion sources worldwide and requires at least four laser operators per experiment, a highly reliable laser system and careful scheduling of ISOLDE experiments. Figure 26 shows the yearly operating hours of the RILIS. Although the demand for RILIS has increased annually, a maximum feasible operating time of approximately 2000 hours was reached for the dye-laser-based system. The influence of the 2011 upgrade to the dual (dye plus Ti:Sa) system on the maximum operating time is clearly evident. A future increase in intensity of RILIS use will require a change from a 'shift-based' to an 'on-call-based' working mode. A project is under way to equip the RILIS installation with a suitable machine protection and safety interlock system that will enable safe running under these conditions. Reliable 'on-call' operation will also require more advanced and comprehensive RILIS parameter monitoring tools and an upgrade of the automated laser beam position stabilization system. Details of this system can be found in Ref. [95].

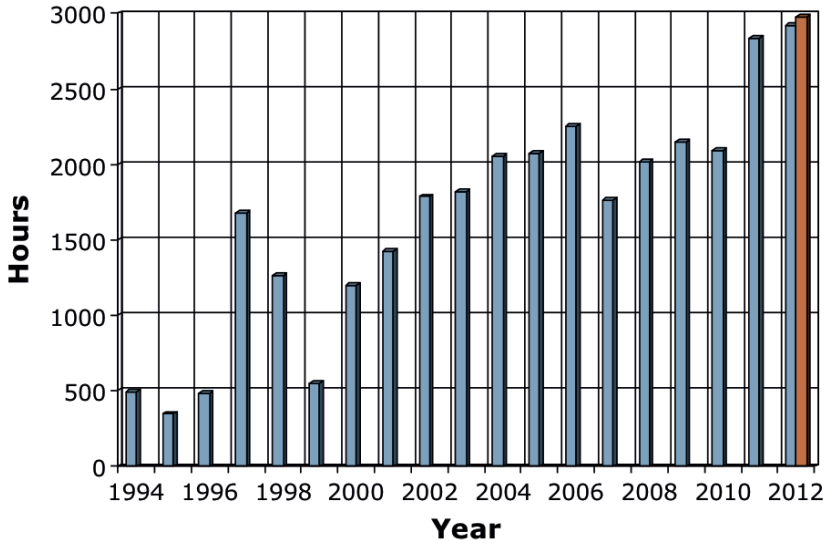


Fig. 26: Annual operating statistics for the ISOLDE RILIS. The jump in operating hours from 2010 to 2012 was only possible due to the availability of the second (Ti:Sa) laser system.

3.8 Ionization schemes

So far, ion beams of 31 elements have been produced using the ISOLDE RILIS, and 28 of these have been used for on-line production of radioactive isotopes. The details of these schemes are provided in Fig. 27 and Table 4.

4 Ionization scheme development

The periodic table of Fig. 28 is colour-coded to indicate the availability of laser ionization schemes for the elements that have been developed at various different facilities using a range of different laser systems. For the coloured elements, an ionization scheme (dye, or Ti:Sa, or mixed dye plus Ti:Sa) has been applied, either off-line or on-line under standard operating conditions for a hot-cavity or gas-cell laser ion source.

Element	Setting time*	Step 1		Step 2		Step 3		Efficiency	Step 1		Step 2		Step 3		Efficiency
	days	λ_1, nm	Dye	λ_2, nm	Dye/YAG	λ_3, nm	Dye/YAG	off-line	λ_1, nm	Fundamental	λ_2, nm	Fundamental	λ_3, nm	TiSa/YAG	off-line
$_{4}$Be	3	234.9	Pyridin 1	297.3	Rhod B or Pyrr 597	–		>7%	234.861	704.583	297.3	Dye 594.6			
$_{12}$Mg	2	285.2	Rhod 6G	552.8	Fluorescein 27	532.1	YAG	10%	285.213	855.639	880.917	880.9	291.63	874.89	
$_{13}$Al	2	309.3 308.2	Rhod B Rhod B	532.1	YAG	–		20%	308.305	924.915	462	924	–		
$_{20}$Ca	2	272.2	Fluorescein 27	532.1	YAG	–		0.50%	422.792	845.584	731.816		868.52		
$_{21}$Sc	2	327.4	Phenox 9	719.8	Pyridin 2	532.1	YAG	15%	402.152	804.304	627.5	Dye	809.162		
$_{25}$Mn	3	279.8	Rhod 6G	628.3	DCM	635.8	DCM	19%	279.827	839.481	628.3	Dye	635.8	Dye	
$_{27}$Co	2	304.4	Rhod B	544.5	Fluorescein 27	532.1	YAG	>3.8%	240.799	722.397	827.16	827.16	816.618	TiSA	
$_{28}$Ni	3	305.1	Rhod B	611.1	Rhod B	748.2	Styr 8	>6%	305.1	Dye	611.1	Dye	748.219	TiSA	
$_{29}$Cu	2	327.4	Phenox 9	287.9	Rhod 6G	–		>7%	244.238	732.714	441.68	883.36	–		
$_{30}$Zn	3	213.9	Phenox 9	636.2	DCM	532.1	YAG	4.90%	213.857	855.428	636.2	Dye	532.1	YAG	
$_{31}$Ga	2	287.4 294.4	Rhod 6G Rhod 6G	532.1	YAG	–		21%	287.423 294.36	862.269 883.08	532.1	YAG	–		
$_{39}$Y	2	414.3	Styr 9	662.4	Phenox 9	510.6	YAG		414.284	828.568	662.4	Dye	510.6	YAG	
$_{47}$Ag	2	328.1	Phenox 9	546.6	Fluorescein 27	532.1	YAG	14%	328.162	984.486	421.214	842.428	532.1	YAG	
$_{48}$Cd	3	228.8	Pyridin 1	643.8	DCM	532.1	YAG	10.40%	228.802	915.208	643.8	Dye	532.1	YAG	
$_{49}$In	2	303.9 325.6	Rhod B Phenox 9	532.1	YAG	–			303.94	911.82	532.1	YAG	–		
$_{50}$Sn	4	286.3	Rhod 6G	811.4	Styr 9	823.5	Styr 9	9%	286.3	858.9	811.399	811.399	823.49	823.49	
$_{51}$Sb	3	217.6	Phenox 9	560.2	Rhod 6G	532.1	YAG	2.70%	217.581	870.324	560.2	Dye	532.1	YAG	
$_{60}$Nd	2	588.8	Rhod 6G	596.9	Rhod B or Pyrr 597	596.9	Pyrr 597								
$_{62}$Sm	3	600.4	Rhod B	675.2	Phenox 9	676.18	Phenox 9								
$_{65}$Tb	3	579.6	Rhod 6G	551.7	Fluorescein 27	618.3	Rhod B								
$_{66}$Dy	2	625.9	DCM	607.5	Rhod B	532.1	YAG	20%							
$_{70}$Yb	2	555.6	Pyrr 567	581.0	Rhod 6G	581.0	Rhod 6G	15%	267.275	801.825	736.67	736.7			
$_{79}$Au	5	267.6	Coum 540A	306.5	DCM	673.9	Phenox 9	>3%	267.59	802.77	406.621	813.242	532.1	YAG	
$_{80}$Hg	3	253.7	Styr 8	313.2	DCM	626	DCM	0.10%	253.7	761.1	313.184	939.552	626	Dye	
$_{81}$Tl	2	276.8	Rhod 110	532.1	YAG	–		27%	276.787	830.361	532.1	YAG	–		
$_{82}$Pb	2	283.3	Rhod 6G	600.2	Rhod B	532.1	YAG	>3%							
$_{83}$Bi	2	306.8	Rhod B	555.2	Fluorescein 27	532.1	YAG	6%	306.77	920.31	555.2	Dye	532.1	YAG	
$_{84}$Po	4	255.8	Styr 8	843.4	Styr 9	532.1	YAG		255.801	767.403	843.38	843.38	532.1	YAG	
$_{85}$At	4	216.9 224.4	Phenox 9 Phenox 9	795.4 793.2 ?	Styr 9 Styr 9	532.1 532.1	YAG YAG		216.9 216.9	867.6 867.6	795.4 915	795.4 915	RYD 532.1	Dye YAG	

Fig. 27: Laser settings and typical set-up time for the ISOLDE RILIS ionization scheme. Several elements can be ionized with either a Ti:Sa, a dye or a combined dye plus Ti:Sa configuration. The values for the transition wavelengths, and the corresponding fundamental beam output from the lasers, are given as vacuum wavelengths for ease of conversion to the transition energy.

Table 4: Excitation schemes used at the ISOLDE RILIS: IP, ionization potential; λ_1, λ_2, λ_3, wavelengths (air) of the first, second and third excitation transitions; P_1, P_2, P_3, laser power transmitted to the RILIS cavity; η_{ion}, ionization efficiency.

Element	IP (eV)	λ_1 (nm)	P_1 (mW)	λ_2 (nm)	P_2 (mW)	λ_3 (nm)	P_3 (mW)	η_{ion} (%)
$_{4}$Be	9.32	234.86	20	297.32	600	—	—	>7
$_{12}$Mg	7.65	285.21	40	552.84	80	532	13 500	>10
$_{13}$Al	5.99	309.27	50	511, 578	5000	—	—	>20
$_{21}$Sc	6.56	327.36	20	719.83	60	511, 578		15
$_{20}$Ca	6.11	422.67	300	586.63	2000	655.29	2000	>2
		272.16	100	532	15 000	—	—	0.45
$_{25}$Mn	7.44	279.83	50	628.27	5000	647.34	4000	>15
$_{27}$Co	7.88	304.40	120	544.46	150	511, 578	5000	>4
$_{28}$Ni	7.64	305.08	60	611.11	275	748.22	1200	>6
$_{29}$Cu	7.73	327.40	50	287.89	250	—	—	>7
$_{30}$Zn	9.39	213.86	20	636.23	500	532	7000	>5
$_{31}$Ga	6.00	287.4	50	532	10 000			>21
		294.45	50					
$_{39}$Y	6.22	408.37	100	581.91	1900	581.91	1900	—
$_{47}$Ag	7.58	328.07	70	546.55	100	511, 578	5000	14
$_{48}$Cd	8.99	228.80	20	643.85	300	511	3000	10
$_{49}$In	6.00	303.94	50	511, 578	5000	—	—	—
$_{50}$Sn	7.34	300.91	50	811.40	150	823.49	1100	≈9
$_{51}$Sb	8.61	217.58	12	560.21	150	511	6000	2.7
$_{60}$Nd	5.53	588.79	700	596.94	4500	596.94	4500	
$_{65}$Tb	5.86	579.56	300	551.65	200	618.25	1200	—
$_{66}$Dy	5.94	625.91	170	607.50	90	511	3500	20
$_{70}$Yb	6.25	555.65	200	581.03	1200	581.03	1200	—
$_{79}$Au	9.23	267.59	80	306.54	40	673.9	370	>3
$_{80}$Hg	10.44	253.65	8	313.18	100	626.3	1200	0.1
$_{81}$Tl	6.11	276.79	100	532	15 000	—	—	27
$_{82}$Pb	7.42	283.30	50	600.19	625	532	8500	3
$_{83}$Bi	7.29	306.77	50	555.20	500	511, 578		6
$_{84}$Po	8.42	255.80	20	843.38	150	532	7500	>0.4
$_{85}$At	9.32	216.22	20	915.21	1000	532	15 000	>4
		216.22	20	795.21	1000			

For the ISOLDE RILIS, the feasibility of ionizing many of the elements is hindered by their inability to escape the thick-target matrix in atomic form (the vapour pressure is too low). These elements are indicated in Fig. 28 by the use of an outlined font. Table 6 of Appendix B is a list of vapour pressures at 2000°C, sorted into ascending order.

For several other elements, laser ionization is not feasible because of the atomic energy level structure. The halogens, for example, exist only in vapour form at high temperatures, but the excitation of their first excited atomic state requires a photon of less than 200 nm, which is very difficult to produce and cannot be easily transported to the ion source. Fortunately, these can be ionized with a plasma ion source and a good degree of selectivity can be achieved by condensing isobaric contaminants in a cooled transfer line.

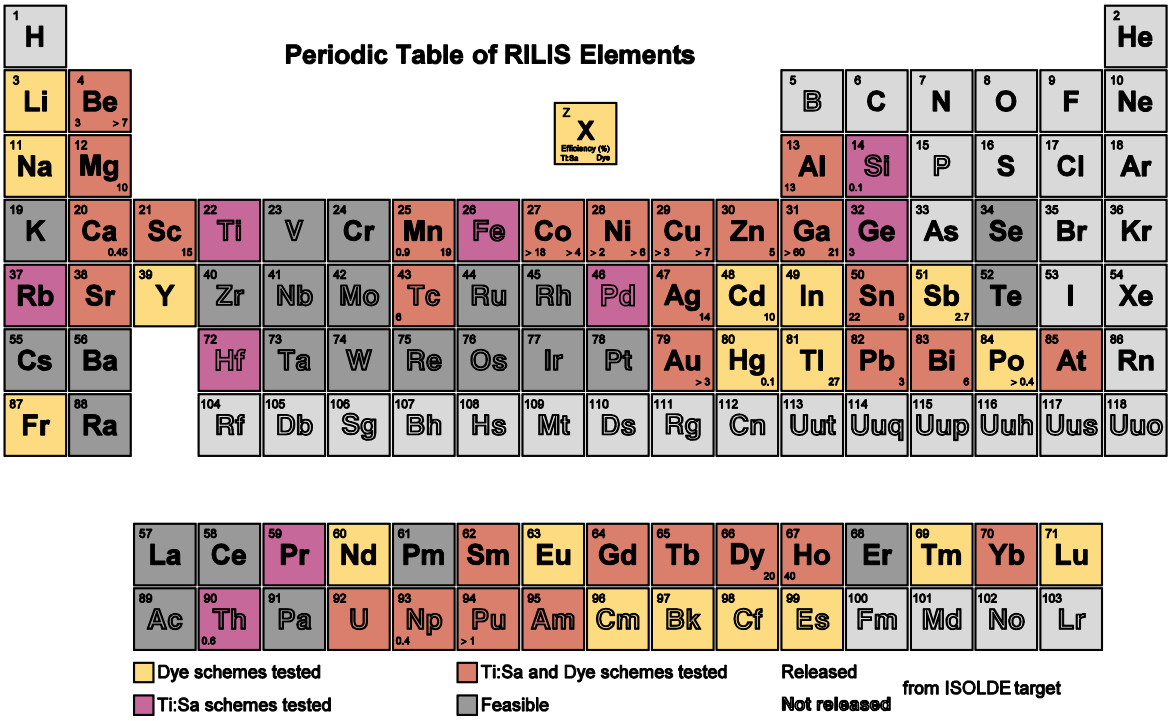


Fig. 28: The elements accessible by resonant ionization laser ion sources

For the elements that can be ionized with the ISOLDE RILIS a principal limitation to ionization efficiency is the laser power available for any unsaturated transitions within the ionization scheme. Currently 16 (Mg, Sc, Co, Zn, Ga, Y, Ag, Cd, In, Sb, Tl, Pb, Bi, Po, At, Dy) of the 31 RILIS ionization schemes rely on a non-resonant ionization step using the second harmonic (532 nm) output of the Nd:YAG pump laser. For these elements, the ionization step is not saturated and the efficiency is proportional to the laser power that is delivered to the ion source. For these elements, an improvement in the ionization efficiency is likely if the final step instead populates an auto-ionizing state (AIS). The search for AISs for the RILIS elements is an ongoing but time-consuming process and, for many elements, a suitable AIS does not exist. For a non-resonant ionization step the residual pump laser power must be transported to the 3 mm ionizer cavity of the target unit, which is at a distance of ≈23 m (for the ISOLDE HRS target) from the final focusing optics inside the RILIS cabin. Typically 45 W of the 90 W Edgewave laser output is available for this application. The beam quality of this laser limits the transport efficiency to the ion source and the power, measured at the reference point (which corresponds to 4% of the transmitted power to the source), is typically 500 mW (implying that only 12.5 W is delivered to the source). If a dye or Ti:Sa laser is used for the final step, the power transmitted to the source is typically approximately 20% of this value. The use of an additional tunable laser for a transition to an AIS can only be justified if the transition cross-section is roughly an order of magnitude higher than for non-resonant ionization with a 532 nm photon. Figure 29 uses the example of scheme development for manganese using a dedicated off-line ionization spectroscopy facility [21] to illustrate the search for an auto-ionizing state by scanning several tunable laser ranges over an energy region above the ionization potential.

The scope for the development of the ionization scheme for the ISOLDE RILIS falls into the following five categories.

1. RILIS ionization is feasible in standard conditions but a suitable ionization scheme has not yet been developed: Cr, Se, Te, Ba, Ra, La, Ce, Pm, Er.

2. RILIS ‘enhancement’ or ion rate is low due to efficient surface ionization in standard operating conditions. RILIS will only be applied with the use of a particularly efficient ionization scheme and/or surface ionization suppression techniques (section 5): Li, Na, K, Cs,(Ba), (Ra).
3. The existing scheme uses non-resonant ionization for the final step. The efficiency is therefore limited by the power available for ionization. Localizing a strong AIS would be beneficial: Mg, Sc, Co, Zn, Ga, Y, Ag, Cd, In, Sb, Tl, Pb, Bi, Po, At, Dy.
4. The laser ionization efficiency is low due to the use of a sub-optimal ionization scheme: Hg.
5. Before attempting ionization scheme development for the refractory metals, further target–ion source research and development is required to produce an atomic sample within a suitable laser interaction region: V, Zr, Nb, Mo, Ru, Rh, Ta, W, Re, Os, Pt, Ac, Pa.

More generally, the scope for scheme development is clearly dependent on the facility type and the technical specifications of the laser system that is to be used. For example, due to the chemical independence of isotope release from the target, the development of an ionization scheme for a GC RILIS facility can be applied immediately without the need for accompanying target/ion source development.

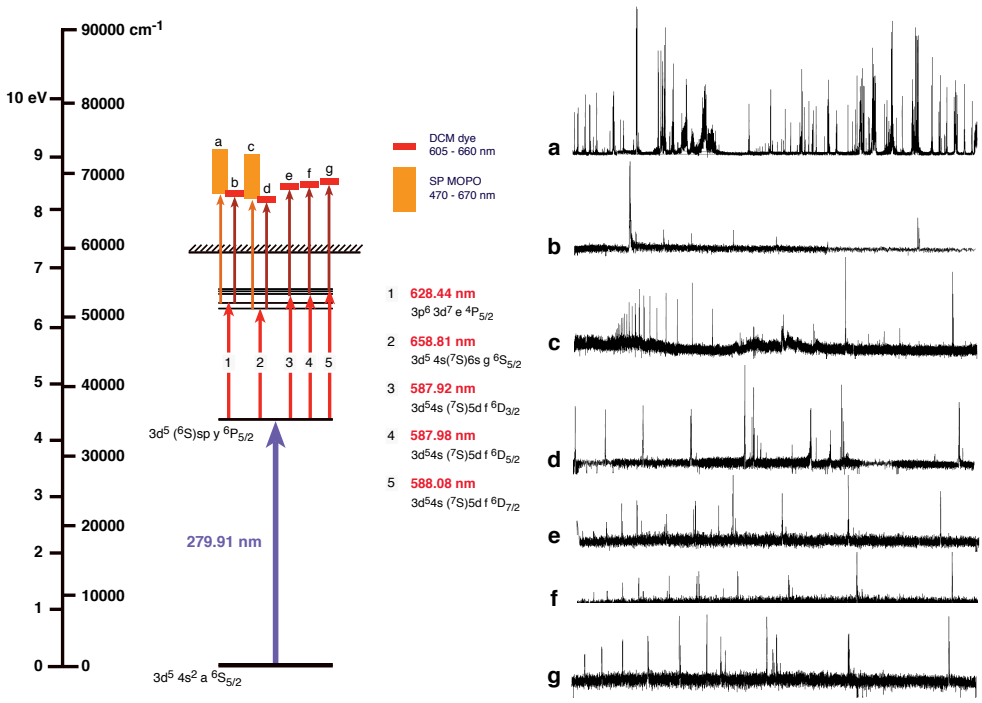


Fig. 29: The search for autoionizing states for manganese by off-line resonance ionization spectroscopy of an atomic beam. The plots a–g show the photo-ion rates versus laser wavelength for the 7 energy regions that were studied.

5 The problem of surface-ionized isobars for hot-cavity laser ion sources

For hot-cavity laser ion sources, the origin of isobaric impurities is the surface ionization process that occurs alongside the laser ionization. The surface ionization process inside the hot cavity is well understood. This is confirmed by the agreement between the experimental data and numerical models of the ionization process [60] (shown in Fig. 30).

For many isotopes that can be laser ionized using the RILIS, the production of a pure ion beam is hindered by in-target production rates of low-ionization-potential atomic or molecular isobars that are efficiently surface-ionized. In these cases, the production of a sufficiently pure laser ionized beam

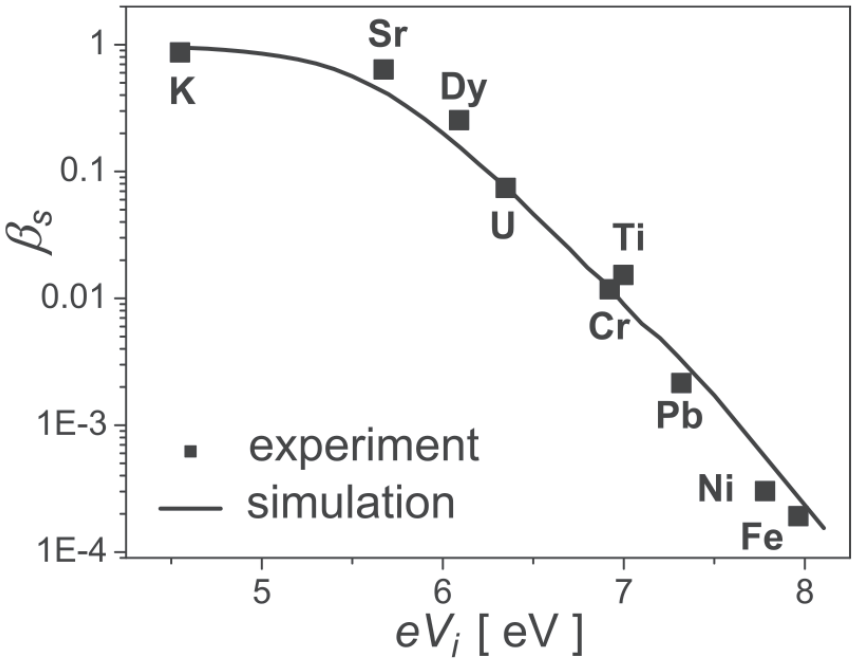


Fig. 30: Experimental data [6] (squares) and simulation results [60] (line) showing the dependence of the hot tungsten cavity ionization efficiency, β_s, on the ionization potential of the element.

requires a means of improving the RILIS selectivity. This problem can be tackled using the following three approaches:

(i) Reduce the surface ionization efficiency. For this approach, the challenge is to selectively decrease the surface ionization efficiency whilst maintaining a high laser ionization efficiency. A starting point is the understanding of the factors that influence the surface ionization process. The Saha–Langmuir equation shows the exponential dependence of the probability of ionization of an atom with ionization potential W_i interacts with a metal surface of work function Ψ at a temperature T:

$$\frac{\mathrm{X}^+}{\mathrm{X}^0} = \frac{g^+}{g^0}\,\mathrm{e}^{-(\Psi - W_i)/kT}. \tag{21}$$

Here g^0 and g^+ are the statistical weights $(2J+1)$ of the atomic and first ionic ground states. This formula suggests that a reduction of either (a) the operating temperature or (b) the material work function can have a significant influence on the probability of surface ionization. The feasibility of these options depends on the effect of these changes on the production of an ion beam of the isotope of interest.

(a) The importance of a high temperature to achieve a low effusion time and to avoid condensation of metals with low vapour pressure on the metal surface has already been discussed. The effect in the resulting ion beam current is dependent on both the element and the isotope (half-life). A reduction in cavity temperature is therefore not a universally feasible solution. Furthermore, it has been shown that the extraction efficiency, or survival rate, of laser ions relies on the temperature-dependent thermionic emission of electrons from the cavity walls [28].

(b) The typically used tungsten or tantalum ionizer materials exhibit material properties, other than the work function, that make them a suitable choice for a surface ion source. They operate reliably at a high temperature, are mechanically robust and easy to produce, and are

easily resistively heated. The use of a low-work-function material inevitably compromises some of these properties. Finding an ideal cavity material is therefore challenging and requires a considerable research and development effort. The performance for radioactive ion beam production can only be reliably assessed in on-line tests. These are costly and time-consuming. An off-line study of candidates for suitable low-work-function cavity materials (indirectly heated GdB_6 and resistively heated W with a BaOSrO coating) has been performed at CERN [61]. A significant reduction of surface ionization has been demonstrated, with only a moderate loss of laser ionization efficiency [64]. Under standard on-line conditions, the same selectivity was not achieved due to surface ions that originate from the hot target and transfer line, and the increased effusion time for some isotopes limits the use of these cavities to specific applications.

(ii) Spatial separation of the laser and surface ions. If the laser ionization process takes place outside of the hot cavity, as illustrated by sketch **1b** of Fig. 8, it is possible to reject the surface ions using a repeller electrode, creating a surface-ion-free laser ionization region downstream of the hot cavity. Clearly this presents some technical challenges. The most crucial of these is obtaining a geometrical overlap of the expanding atom beam from the hot cavity with RILIS laser beams. Clearly, the best overlap is achieved with the laser beams directed towards the ionizer cavity. Whilst this is also the optimal geometry for creating laser ions inside the cavity, these ions will be rejected along with the repelled surface ions. Since the laser ionization efficiency inside the cavity is typically <10%, this loss factor is not the primary concern for this technique. Greater challenges are designing the laser interaction region itself and producing a device that is suitable for on-line operation in a hostile radiation and high-temperature environment.

In recent years there has been considerable progress in this area, and a device known as the laser ion source trap (LIST) has been coupled to ISOLDE Ti and UC_x target units and used on-line for isobar-free laser ionization of Mg and Po isotopes.

The LIST is the result of many years of collaborative development between the Mainz University LARISSA group and ISOLDE at CERN. The initially foreseen device was a miniaturized radio-frequency cooler–buncher [62]: a segmented gas-filled radio-frequency quadropole (RFQ) with a longitudinal trapping potential and a fast-switching end electrode. This would enable the cooling and trapping of ions and their release as an intense ion bunch with a low energy spread and high purity. A prototype was built and tested off-line, but its unsuitability for on-line use at a facility such as ISOLDE (gas load, space charge issues for bunching, number of electrodes, etc.) led to a series of simplifications that resulted in the current ISOLDE LIST unit. This simplified LIST comprises a surface ion repeller electrode, an RFQ ion guide and an end electrode. Its design is based upon the following key requirements for the laser interaction region:

(a) The laser interaction region should be as close to the cavity exit as possible for good laser–atom overlap.
(b) Laser ionization should take place in a longitudinally 'field-free' region so that the energy spread of the extracted ions is low.
(c) The ions must be guided towards the extraction field using a transverse potential.
(d) The device should be easily switchable to 'ion guide', whereby ions originating from the hot cavity are transmitted.

In parallel, an equivalent device was developed at the TRIUMF TRILIS facility, initially referred to as the Ion-Guide Laser Ion Source (IG-LIS) [96], but later renamed Surface Ion Repelling Laser Ion Source (SIRLIS) to avoid confusion with the IGLIS (In-Gas Laser Ionization and Spectroscopy) method which encompasses the GC RILIS and GC LIST techniques of **2** and **3** in Fig. 8). The ISOLDE and TRIUMF LIST devices have both been fully tested and characterized under standard on-line operating conditions [97, 98]. At ISOLDE the LIST was recently used for in-source

resonance ionization spectroscopy of ^{217,219}Po isotopes, made possible due to the suppression of francium isobars by over three orders-of-magnitude. The LIST performance is characterized by two key values: the suppression factor ($\sim 10^3$-10^4) and the efficiency loss factor ($\sim$20-50). The loss factor is due to the expansion of the atomic beam downstream of the ionizer cavity. This would therefore be minimized with a higher repetition rate laser system (with the same pulse energy), a reduced cavity-to-LIST gap, improved atom beam collimation, or a larger laser beam diameter (provided that all transitions of the ionization scheme are saturated).

(iii) Temporal separation of laser and surface ions. This can be achieved by exploiting the time structure of the pulsed laser ionization process which occurs on top of the d.c. surface ion background through the use of a fast beam-gate, synchronized with the laser pulse repetition rate. In a standard tungsten cavity, the resistive heating of the cavity results in a voltage drop along the cavity of several volts. The resulting drift velocity of the ions towards the extraction region reduces the laser-ion bunch width. If cavity voltage is increased (through the use of a higher resistance material or a cavity design), the bunch width can be reduced further and a higher selectivity can be achieved [58]. In this case the challenge is to maximize the cavity voltage drop whilst maintaining a universal, robust and reliable laser interaction region. The optimization of this approach through the testing of different cavity materials, pulsed heating techniques and the use of improved gating methods is the subject of ongoing research and development. The maximum selectivity would be achieved if the beam-gate is applied at the time focus of the ion bunch, this approach, known as ToFLIS, has been investigated off-line by Mishin et al [99]. The implementation of this system for on-line use is the subject of ongoing research and development.

6 In-source resonance ionization spectroscopy

The technique of in-source resonance ionization spectroscopy is a sensitive method of studying the structure of atoms that are released from the target of a radioactive ion beam facility. For rare, exclusively radioactive, elements, this method enables the study of the atomic properties such as the transition energies, atomic spins and the ionization potential.

Laser spectroscopy can be precise enough to resolve the influence of nuclear structure differences (spins, moments, shapes and volumes), between different isotopes, on the electronic energy levels. A broad overview of the application of lasers in the field of nuclear physics can be found in Ref. [68]. The study of nuclear ground-state and isomer properties by laser spectroscopy of atomic transitions has existed for over 30 years [69,70]. There are two recently published reviews of measurements in this field, covering the many different approaches at various facilities worldwide [65,71]. These articles include an introduction to the measurement observables – the *hyperfine structure* (HFS) and *isotope shift* (IS) – and formulas for the extraction of nuclear properties such as spins, moments and changes in mean-squared charge radii. Together, they form a useful repository of references for work that has been carried out in the field. Most of the work reported is sub-Doppler collinear laser fluorescence spectroscopy of fast atom or ion beams [72]. Typically single-mode CW laser light is used, thereby achieving a spectral resolution that approaches the natural linewidth of the electron transition.

For in-source resonance ionization spectroscopy, the measurement is performed during the ion production process of a laser ion source: a laser wavelength scan across the resonance structure of a suitable excitation step is made whilst recording the resulting ion rate. The first application of this method is reported in Ref. [73]. The efficiency of laser ionization, coupled with the selectivity and efficiency of modern particle detection techniques based on characteristic radiation decay, makes this is an extremely sensitive method. It is therefore suitable for the measurement of exotic isotopes with very low production rates (less than one per second).

At ISOLDE, this technique has been successfully applied for the study of the nuclear spins and electromagnetic moments of Cu isotopes [74, 75], for the separation of Ag isomers, for Be isotope shift

measurements [66] and for the ongoing campaign to measure the systematics of the elements in the lead ($Z = 82$) region (Pb [76–78], Bi [79], Tl, Po [80, 100].

The same technique has also been successfully applied to gas-cell facilities [81], and there are plans to perform in-source resonance ionization spectroscopy studies using the proposed laser ion sources of the fragment separator facilities at GANIL, France, and RIKEN, Japan.

Hot-cavity-based in-source resonance ionization spectroscopy is limited in spectral resolution due to Doppler broadening. This is larger for the lighter elements but still exceeds 1 GHz for the heaviest isotopes that can be produced at ISOLDE. Useful information can therefore only be obtained from atomic transitions that are particularly sensitive to changes of nuclear properties. The choice of the atomic transition for spectroscopy is therefore important. For transitions involving electronic s orbitals, the IS or HFS is likely to be larger due to the higher probability density of the electron at the position of the nucleus. An additional consideration is that the extraction of useful information may rely on atomic calculations, and these are more reliable for transitions from the ground state of the atom [82].

The total isotope shift is the sum of two components, known as the *field shift* (FS) and the *mass shift* (MS). Only the FS contains information about the nuclear charge radius. Deducing radii from the observed IS therefore relies on an accurate extraction of the MS contribution, or restricting the measurements to the study of heavy elements for which FS is dominant.

Only isotopes with a non-zero nuclear spin have atomic transitions with a hyperfine structure. The hyperfine structure is the lifting of the degeneracy of an electronic level due to the coupling of the total electronic angular momentum J with the nuclear spin I. This results in a new quantum number to describe the electronic level, F ($F = |I - J|, \ldots, I + J$). Transitions between the F states of different energy levels are possible provided that $\delta F < 2$. Fig. 31 shows the ionization scheme for bismuth and the F levels that result in the hyperfine structure of the first step transition. Fig. 32 is an example of how this looks when observed by scanning the RILIS dye laser in narrow-band mode whilst recording the ion rate via alpha detection using the windmill detector set-up. There is a pair of unresolved triplet peaks. The separation of each multiplet is ∼25 GHz, due to the splitting of the upper level of the transition. The substructure of each multiplet corresponds to the F level spacing of the lower level of the transition.

The ^{191}Bi optical spectrum shows a significant difference between the HFS for the ground state and for the isomer. This is clearer if one looks at the accompanying alpha spectra, where the left-hand spectrum was obtained at a scan position with the highest $\frac{1}{2}^+ : \frac{9}{2}^-$ spin-state ratio and the right-hand spectrum shows the highest obtainable $\frac{9}{2}^- : \frac{1}{2}^+$ ratio. This is an example of the RILIS operating as an isomer-selective ion source. In this case, the isomer selectivity was just a consequence of the hyperfine-structure scans, but this example demonstrates that, for nuclear isomers that result in a discernibly different hyperfine structure, it is possible to selectively ionize atoms that have a particular nuclear spin state. This is particularly desirable for nuclear physics experiments such as Coulomb excitation [83] or precision mass measurements [84], where a pure isomer beam can greatly simplify the interpretation of the data. For ^{68}Cu, the isomer selectivity ($^{68\mathrm{m}}\mathrm{Cu}/(^{68\mathrm{m}}\mathrm{Cu} + {}^{68\mathrm{g}}\mathrm{Cu})$) was ≈400 [10].

6.1 Narrow linewidth lasers

For standard RILIS operation or spectroscopy studies for ionization scheme development, the primary concern is ionization efficiency. In this case the large bandwidth (∼10 GHz) of the RILIS lasers is advantageous, as one can be confident that, in most cases, the HFS components of the atomic levels are excited simultaneously and the sensitivity to the IS is low. In-source spectroscopy attempts to resolve these features at the expense of overall ionization efficiency.

A reduction in the laser bandwidth and the spectral line broadening for the spectroscopic transition is required to achieve a measurement precision approaching the limit of ∼1 GHz set by the broadened spectral linewidth in the hot cavity.

For the DMK dye laser and the CERN Ti:Sa lasers, a linewidth reduction to less than 1 GHz has

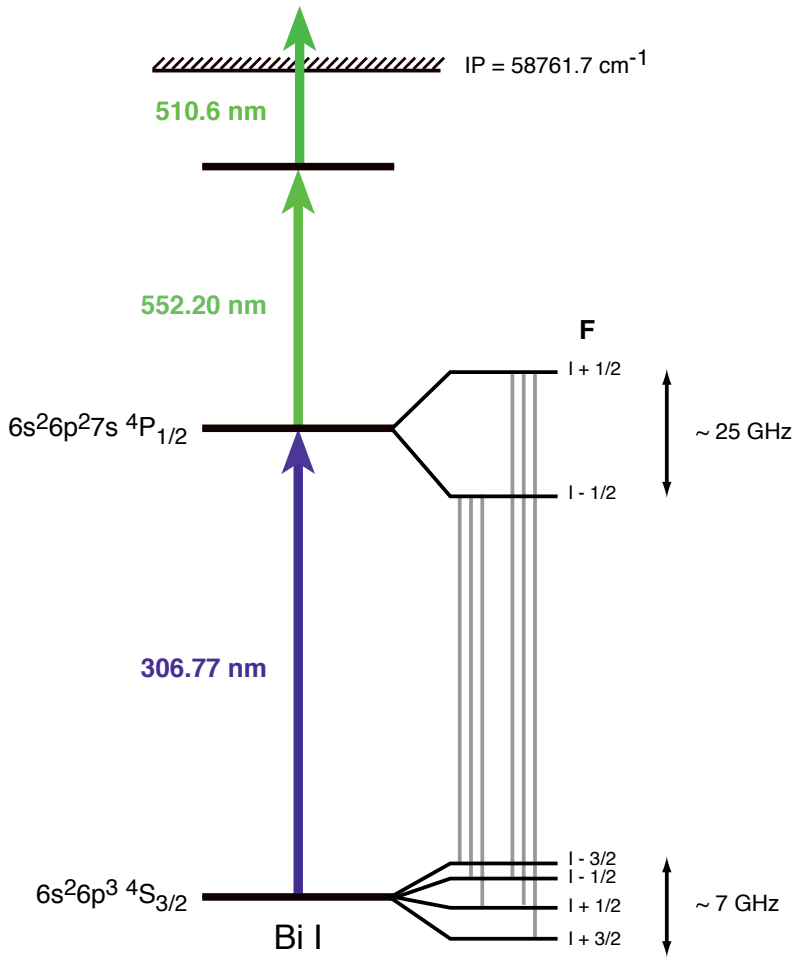


Fig. 31: The RILIS ionization scheme for bismuth. The hyperfine splitting of the ground state and the first excited state is shown (not to scale). The F quantum numbers and allowed transitions are shown.

been achieved by inserting a thick (≈6 mm) Fabry–Perot etalon in the resonator cavities [101]. If the atomic transition is saturated, the power (saturation) broadening can contribute to the spectral linewidth. By attenuating the output power (using neutral density filters or a half-wave plate/polarizing beam-splitter cube combination) to below saturation, power broadening is avoided, and the ionization efficiency is only slightly reduced. The influence of the linewidth reduction on the spectral resolution for hyperfine-structure measurement is demonstrated by the scans of the hyperfine structure of ^{197}Au, shown in Fig. 33.

Recent work at the Leuven LISOL facility has demonstrated the use of pulsed dye amplification of a CW tunable diode laser to obtain pulsed single-mode CW output with a Fourier-limited linewidth (≈90 MHz for a 5 ns pulse). At the LISOL facility, this linewidth reduction will enable high-resolution in-gas jet laser spectroscopy. At ISOLDE, if a laser with such a narrow linewidth was used for hot-cavity in-source spectroscopy, the Doppler broadening of the atomic lines would mean that little enhancement of the spectral resolution would be achieved but the efficiency would be reduced by a factor equal to the ratio of the laser Doppler-broadened atomic linewidth to the laser linewidth.

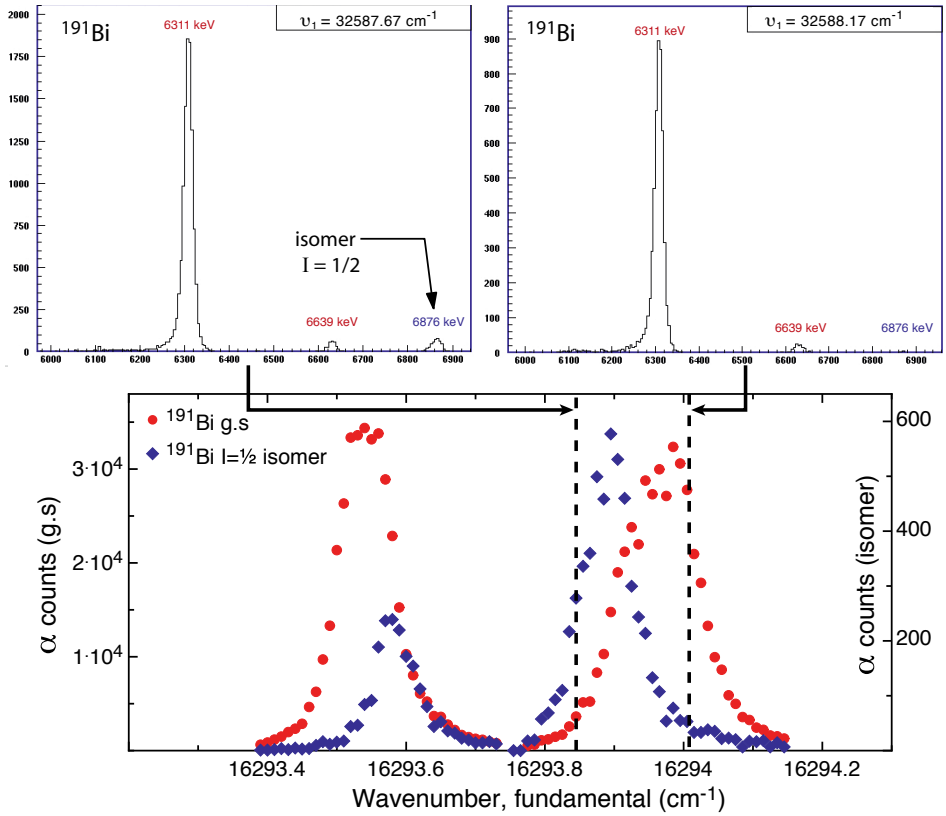


Fig. 32: Examples of alpha spectra acquired during scans of ^{189}Bi at ISOLDE using the in-source resonance ionization spectroscopy technique.

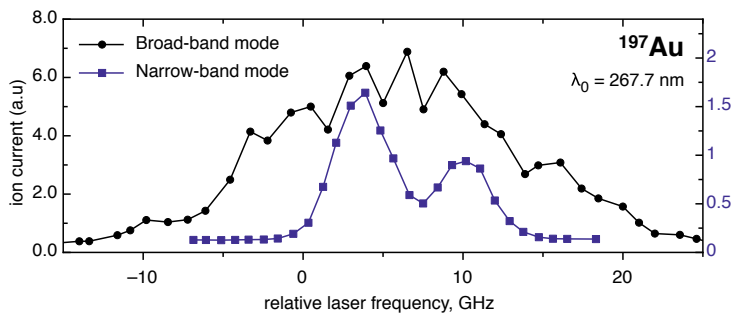


Fig. 33: A comparison of broad- and narrow-band hyperfine-structure scans of the ^{197}Au first step transition using the frequency-tripled output of the Ti:Sa laser.

Appendices

A Radioactive isotopes ionized with the RILIS technique

Table A.1: Radionuclides and radioactive ion beams delivered for atomic and nuclear physics experiments applying methods of resonance laser ionization. The information on the ionization scheme used includes the number of excitation steps and the ion source type (HC RILIS, hot-cavity RILIS; GC RILIS, gas-cell RILIS; ABT, atomic beam with thermal evaporation; ABPL, atomic beam with pulsed laser evaporation; CRIS, collinear resonance ionization spectroscopy; RIMS, resonance ionization mass spectroscopy; HC LIST, laser ion source trap coupled with a hot cavity), and the type of ionizing transition (C, continuum; A, auto-ionization; R, Rydberg state). This is an update of the equivalent table in Ref. [27].

	Z	Scheme	A	Technique	Facility	Reference
Li	3	Four-step C	8, 9	ABT	UNILAC/GSI	[Ewa04]
			8–11			[San06]
Be	4	Two-step A	7, 10–12, 14	HC RILIS	ISOLDE/CERN	[Kos98]
		Three-step C	10–12			[Pri06]
		Two-step A	9–12			[Las11]
Mg	12	Three-step C	23, 27–34	HC RILIS	ISOLDE/CERN	[Kos03b]
		Three-step C	22, 23, 27	HC LIST	ISOLDE/CERN	[Fin13]
		Three-step C	20–28		ISAC/TRIUMF	[Rae13b]
			22			[Muk04]
			21			[Kra09]
		Three-step C	21, 23, 27, 28		ISAC/TRIUMF	[Las09]
		Three-step A	20, 21, 23–35			[Las11]
Al	13	Two-step C	26, 28–34	HC RILIS	ISOLDE/CERN	[Kos03b]
		Two-step C	26		ISAC/TRIUMF	[Pri06]
			26, 28, 29			[Las09]
			30–31			[Las11]
			24–29		ISAC/TRIUMF	[Rae13b]
Ca	20	Three-step C	41	ABT	Mainz University	[Mue01]
		Three-step C	49–52	HC RILIS	ISAC/TRIUMF	[Las11]
		Three-step A	49–54	HC RILIS	ISOLDE/CERN	[Mar13]
Sc	21	Three-step A	47	HC RILIS	ISAC/TRIUMF	[Rae13b]
Mn	25	Three-step A	53, 54, 56–65	HC RILIS	ISOLDE/CERN	[Fed97]
			61–69			[Han99]
			48–54, 56–69			[Oin00]
Fe	26	Two-step A	65, 67	GC RILIS	LISOL/LLN	[Pau09]
Co	27	Two-step A	66, 68, 70	GC RILIS	LISOL/LLN	[Mue00]
			54			[Kud96]
			67			[Wei99]
			65, 67			[Pau09]
Ni	28	Three-step A	56, 57, 59, 63, 65–70	HC RILIS	ISOLDE/CERN	[Jok97]
		Two-step A	54, 55	GC RILIS	LISOL/LLN	[Kud96]
			68–74			[Fra01]
Cu	29	Two-step A	57–62, 64, 66–78	HC RILIS	ISOLDE/CERN	[Kos00b]
		Two-step A	57–59	GC RILIS	LISOL/LLN	[Coc10]
			70–76			[Kru02]
Zn	30	Three-step C	58	HC RILIS	ISOLDE/CERN	[Jok98]
			58–61, 63			[Oin00]
			58–63, 69, 71–74			[Kos03]
			61–63, 65, 69, 71–81			[Kos05]
			71–82			[Kos08]
Ga	31	Two-step C	61	HC RILIS	ISOLDE/CERN	[Wei02b
			74–86			[Kos02a]
			61–68, 70, 72–75			[Kos03b]

Continued on next page

Table A.1 – continued from previous page.

	Z	Scheme	A	Technique	Facility	Reference
		Two-step R	62		ISAC/TRIUMF	[Pri06]
		Two-step C	61–68, 70, 72–75			[Las09]
			84		ALTO/INPO	[Kol12]
		Three-step C	83, 85, 86		HRIBF/ORNL	[Liu13]
Ge	32	Three-step A	67, 71, 76	HC RILIS	ISAC/TRIUMF	[Las11]
Sr	38	One-step R	89, 90	CRIS	Mainz University	[Mon93]
Tc	43	Three-step A	99	HC RILIS	Mainz University	[Ame90]
	43	Three-step A	94, 96, 98, 99		ISAC/TRIUMF	[Las11]
	43	Three-step R	99		Mainz University	[Mat10]
Ru	44	Two-step A	90, 91	GC RILIS	LISOL/LLN	[Dea04]
	44		92, 94, 95			[Fac04]
Rh	45	Two-step A	91–93	GC RILIS	LISOL/LLN	[Dea04]
			113			[Kud96]
			94, 112			[Kud09]
			116			[Kud08]
Ag	47	Three-step C	112, 121–127	HC RILIS	ISOLDE/CERN	[Fed95b]
			107m, 122–129			[Kra98]
			101–108, 110–129	HC RILIS	ISOLDE/CERN	[Fed00]
		Three-step C	129–130			[Kra05]
			98–107, 109–117		ISAC/TRIUMF	[Las09]
		Two-step C	97–101	GC RILIS	LISOL/LLN	[Fer13]
Cd	48	Three-step C	131, 132	HC RILIS	ISOLDE/CERN	[Han00]
			98–105, 107, 109, 111, 115, 117–132			[Kos03b]
			129–133			[Kra05]
In	49	Two-step C	100–108	HC RILIS	ISOLDE/CERN	[Kos02a]
			132–135			[Dil02]
		Two-step C	108, 110, 112, 116		ISAC/TRIUMF	[Las11]
Sn	50	Three-step A	101–103, 108	HC RILIS	UNILAC/GSI	[Fed95a]
			109–111, 113, 117, 119, 121, 123, 125–137		ISOLDE/CERN	[Fed00]
			136–138			[Wal05]
			105–110, 113, 117, 119, 121, 123, 125, 128–138			[Kos08]
		Two-step A	125–132	ABPL	ISOLDE/CERN	[Bla02]
		Three-step A	107–111, 113, 121	HC RILIS	ISAC/TRIUMF	[Las11]
		Two-step A	97–101	GC RILIS	LISOL/LLN	[Fer13a]
Sb	51	Three-step C	128–138	HC RILIS	ISOLDE/CERN	[Fed08]
			137–139			[Arn11]
Te	52	Three-step C	120, 122–136	ABPL	ISOLDE/CERN	[Sif06]
Pr	59	Three-step C	136, 140	HC RILIS	ISOLDE/CERN	[Got11]
Nd	60	Three-step C	132, 134–141	ABT	IRIS/PNPI	[Let92]
			138, 139, 140	HC RILIS	ISOLDE/CERN	[Got11]
Sm	62	Three-step A	138–143, 145	ABT	IRIS/PNPI	[Mis87b]
			140–143	HC RILIS	ISOLDE/CERN	[Got11]
Eu	63	Three-step A	145–150	ABT	IRIS/PNPI	[Alk83]
			141–144			[Fed84]
			155–159			[Alk90a]
			138–145			[Let92]
			137–139, 141–144	HC RILIS		[Bar04]
Gd	64	Three-step A	146, 148, 150	ABT	IRIS/PNPI	[Alk88]
			143, 145, 146	HC RILIS		[Bar05]

Continued on next page

Table A.1 – continued from previous page.

	Z	Scheme	A	Technique	Facility	Reference
Tb	65	Three-step A	147–155, 157, 159		IRIS/PNPI	[Alk90b]
			149	HC RILIS	ISOLDE/CERN	[Kos03b]
Dy	66	Three-step C	151–155, 157, 159	HC RILIS	ISOLDE/CERN	[Got11]
Ho	67	Three-step A	152–163	ABT	IRIS/PNPI	[Alk89a]
			152	HC RILIS		[Alk91]
Tm	69	Three-step A	157–168, 170–172	ABT	IRIS/PNPI	[Mis87a]
			153, 154, 158–163	HC RILIS		[Bar00]
Yb	70	Two-step R	157, 159, 175	CRIS	ISOLDE/CERN	[Sch91]
		Three-step A	154–156, 160–166	HC RILIS	IRIS/PNPI	[Alk92]
			155			[Bar00]
			153–156, 158–164			[Bar02]
			157–167		ISOLDE/CERN	[Mis93]
			178			[Got11]
Ir	77	Three-step C	182–189	ABPL	ISOLDE/CERN	[Ver00]
Pt	78	Three-step A	186, 188	ABPL	ISOCELE/IPN	[Lee88]
			185–189, 191			[Duo89]
			183–189, 191, 193		ISOLDE/CERN	[Hil92]
			178–185			[Bla99]
Au	79	Three-step C	185–189	ABT	ISOLDE/CERN	[Wal87]
		Three-step A	195	ABPL		[Kro87]
		Two-step C	194–196, 198		McGill University	[Lee87a]
		Three-step C	186, 187, 190, 192, 194–196		ISOCELE/IPN	[Kro91]
			198, 199			[Sav90]
		Three-step A	183, 184		ISOLDE/CERN	[Lee87b]
		Three-step C	183–186, 191		ISOCELE/IPN	[Bla92]
		Two-step C	184, 188, 190, 191, 193, 195		ISOLDE/CERN	[Sau00]
		Three-step A	194	HC RILIS		[Eli10]
			184, 186, 190, 194, 201, 202			[Pod10]
			177–182, 185, 191			[And13]
Tl	81	Two-step R	208	GC RILIS	Mainz University	[Lau92]
		Two-step C	179–200	HC RILIS	ISOLDE/CERN	[Kos03b]
			183–191, 193–195, 197, 207		IRIS/PNPI	[Bar13]
Pb	82	Three-step C	185	HC RILIS	ISOLDE/CERN	[And02]
			184–203, 205, 209–215			[Kos02b]
			183–203, 215			[Kos03b]
			182–190			[Sel06]
Bi	83	Three-step C	188–208, 210–218	HC RILIS	ISOLDE/CERN	[Kos03b]
Po	84	Three-step C	193–198, 200, 202, 204	HC RILIS	ISOLDE/CERN	[Coc08]
			192–210, 216, 218 (even)			[Coc11]
			191–203, 209, 211 (odd)			[Sel13]
			208, 216–219	HC LIST		[Fin13]
At	85	Three-step C	198, 199, 212, 217, 218	HC RILIS	ISAC/TRIUMF	[Las11]
		Two-step C	199		ISOLDE/CERN	[Rot13]
		Three-step R	205			
		Three-step C	194, 196–212, 217–219			[And13]
Fr	87	Two-step C	221	HC RILIS	ISAS, Troitsk	[And86b]
		Two-step R	221			[And87]
		Two-step C	202–207, 211, 218, 219–221, 229, 231	CRIS	ISOLDE/CERN	[Coc13]
Ac	89	Two-step A	212, 213	GC RILIS	LISOL/LLN	[Fer13a]
			225	HC RILIS	ISAC/TRIUMF	[Las11]
			227		Mainz University	[Ros12]
Th	90	Two-step C	230	RIMS	LANL, USA	[Joh93]
		Three-step A	228–230		Mainz University	[Rae11b]

Continued on next page

Table A.1 – continued from previous page.

	Z	Scheme	A	Technique	Facility	Reference
		Two-step A	229			[Son12b]
Pa	91	Three-step A	231	HC RILIS	Mainz University	[Got11]
U	92	Three-step A	233	HC RILIS	Mainz University	[Rae10]
Np	93	Two-step A	237	RIMS	Mainz University	[Rie93]
		Three-step A	237	HC RILIS		[Rae11a]
Pu	94	Three-step A	239–242, 244	RIMS	Mainz University	[Rus89]
Pu			233	HC RILIS	Mainz University	[Rae12]
Am	95	Three-step R	243	RIMS	Mainz University	[Erd98]
		Two-step C	241, 243, 240f, 242f	GC RILIS	MPIK, Heidelberg	[Bac98]
			244f			[Bac00]
Cm	96	Three-step R	248	RIMS	Mainz University	[Erd98]
Bk	97	Three-step R	249	RIMS	Mainz University	[Erd98]
Cf	98	Three-step R	249	RIMS	Mainz University	[Erd98]
Es	99	Three-step R	254	RIMS	Mainz University	[Erd98]
Fm	100	Two-step C	255	GC RILIS	Mainz University	[Sew03]

References for table A.1

Alk83 G.D. Alkhazov *et al.*, Measurement of isotopic variations in the charge radii of europium nuclei by the method of three-stepped laser photoionization of atoms. *JETP Lett.* **37** (1983) 274–8 (Engl. transl.)

Alk88 G.D. Alkhazov, A.E. Barzakh, V.P. Denisov, V.S. Ivanov, I.Ya. Chubukov, N.B. Buyanov, V.S. Letokhov, V.I. Mishin, S.K. Sekatskii, and V.N. Fedoseev, Mean-square charge radius of the magic nucleus ^{146}Gd. *JETP Lett.* **48** (1988) 413–15 (Engl. transl.).

Alk89a G.D. Alkhazov, A.E. Barzakh, I.Ya. Chubukov, V.P. Denisov, V.S. Ivanov, N.B. Buyanov, V.N. Fedoseyev, V.S. Letokhov, V.I. Mishin, and S.K. Sekatskii, Nuclear deformation of holmium isotopes *Nucl. Phys.* **A504** (1989) 549–61

Alk90a G.D. Alkhazov *et al.*, Odd–even staggering in nuclear charge radii of neutron-rich europium isotopes. *Z. Phys.* **A337** (1990) 257–9

Alk90b G.D. Alkhazov, A.E. Barzakh, V.P. Denisov, V.S. Ivanov, I.Ya. Chubukov, V.S. Letokhov, V.I. Mishin, S.K. Sekatsky, and V.N. Fedoseyev, Electromagnetic moments and nuclear charge radii for neutron-deficient Tb isotopes and the deformation jump near $Z = 64$, $N = 90$. *Z. Phys.* **A337** (1990) 367–70

Alk91 G.D. Alkhazov, L.K. Batist, A.A. Bykov, V.D. Vitman, V.S. Letokhov, V.I. Mishin, V.N. Panteleyev, S.K. Sekatsky, and V.N. Fedoseyev, Application of a high efficiency selective laser ion source at the IRIS facility. *Nucl. Instrum. Methods Phys. Res.* **A306** (1991) 400–2

Alk92 G.D. Alkhazov *et al.*, A new highly efficient method of atomic spectroscopy for nuclides far from stability. *Nucl. Instrum. Methods Phys. Res.* **B69** (1992) 517–20.

Alk92 S.V. Andreev, V.S. Letokhov, and V.I. Mishin, Laser resonant photoionization detection of traces of the radioactive isotope ^{221}Fr in a sample. *JETP Lett.* **43** (1986) 736–40 (Engl. transl.).

Ame90 Ames F, Brumm T, Jager K, Kluge H-J, Suri B M, Rimke H, Trautmann N and Kirchner R 1990, A high-temperature laser ion source for trace analysis and other applications *Appl. Phys.* **B51** 200-6

And87 S.V. Andreev, V.S. Letokhov, and V.I. Mishin, Laser resonance photoionization spectroscopy of Rydberg levels in Fr. *Phys. Rev. Lett.* **59** (1987) 1274–6

And02 A.N. Andreyev *et al.*, Nuclear spins, magnetic moments and α-decay spectroscopy of long-lived isomeric states in ^{185}Pb. *Eur. Phys. J.* **A14** (2002) 63–75

And13 A.N. Andreyev *et al.*, 2013 Addendum to the IS534 Proposal CERN-INTC-2013-002; INTC-P-319-ADD-1.

Arn11 O. Arndt *et al.*, Decay of the r-process nuclides ^{137,138,139}Sb, and the $A = 130$ solar r-process abundance peak. *Phys. Rev.* **C84** (2011)

Bac98 H. Backe *et al.*, Isotope shift measurements for superdeformed fission isomeric states. *Phys. Rev. Lett.* **80** (1998) 920–3

Bac00 H. Backe, A. Dretzke, M. Hies, H. Kunz, W. Lauth, M. Sewtz, N. Trautmann, R. Repnow, and H.J. Maier, Isotope shift measurement at ^{244f}Am. *Hyperfine Interact.* **127** (2000) 35–9

Bar00 A.E. Barzakh, I.Ya. Chubukov, D.V. Fedorov, V.N. Panteleev, M.D. Seliverstov, and Yu.M. Volkov, Mean square charge radii of the neutron-deficient rare-isotopes in the region of the nuclear shell $N = 82$ by the laser ion source spectroscopy technique. *Phys. Rev.* **C61** (2000) 034304

Bar02 A.E. Barzakh, D.V. Fedorov, V.N. Panteleev, M.D. Seliverstov, and Yu.M. Volkov, Measurements of charge radii and electromagnetic moments of nuclei far from stability by photoionization spectroscopy in a laser ion source. *AIP Conf. Proc.* **610** (2002) 915–19

Bar04 A.E. Barzakh, D.V. Fedorov, A.M. Ionan, V.S. Ivanov, F.V. Moroz, K.A. Mezilev, S.Yu. Orlov, V.N. Panteleev, and Yu.M. Volkov, Changes in the mean square charge radii of neutron-deficient europium isotopes measured by the laser ion source resonance ionization spectroscopy. *Eur. Phys. J.* **A22** (2004) 69–74

Bar05 A.E. Barzakh, D.V. Fedorov, A.M. Ionan, V.S. Ivanov, F.V. Moroz, K.A. Mezilev, S.Yu. Orlov, V.N. Panteleev, and Yu.M. Volkov, Laser spectroscopic studies of ^{145}Gd, ^{145}Gd^m, and ^{143}Gd^m. *Phys. Rev.* **C72** (2005) 017301

Bar13 A.E. Barzakh, D.V. Fedorov, V.S. Ivanov, P.L. Molkanov, V.N. Panteleev, and Yu.M. Volkov, New laser setup at the IRIS facility. Magnetic moments and mean squared charge radii of neutron deficient Tl isotopes. *Hyperfine Interact.* **216** (2013) p27-31.

Bla92 F. Le Blanc *et al.*, PILIS: Post-ISOCELE Laser Isobar Separation – a high efficiency apparatus for laser spectroscopy. *Nucl. Instrum. Methods Phys. Res.* **B72** (1992) 111–18.

Bla99 F. Le Blanc *et al.*, Large odd–even radius staggering in the very light platinum isotopes from laser spectroscopy. *Phys. Rev.* **C60** (1999) 054310.

Bla02 F. Le Blanc *et al.*, Charge radius change in the heavy tin isotopes until $A = 132$ from laser spectroscopy. *Eur. Phys. J.* **15** (2002) 49–51.

Coc08 T.E. Cocolios *et al.*, Resonant laser ionization of polonium at RILIS-ISOLDE for the study of ground- and isomer-state properties. *Nucl. Instrum. Methods Phys. Res.* **B266** (2008) 4403–6.

Coc10a T.E. Cocolios *et al.*, Magnetic dipole moments of ^{57,58,59}Cu *Phys. Rev.* **C81** (2010) 014314.

Coc11 T.E. Cocolios *et al.*, Early onset of ground state deformation in neutron deficient polonium isotopes. *Phys. Rev. Lett.* **106** (2011) 052503.

Coc13 T.E. Cocolios *et al.*, The Collinear Resonance Ionization Spectroscopy (CRIS) experimental setup at CERN-ISOLDE. *Nucl. Instrum. Methods Phys. Res. B* In Press http://dx.doi.org/10.1016/j.nimb.2013.05.088

Dar12 I.G. Darby *et al.*, In-gas-cell laser resonance ionization spectroscopy in the vicinity of ^{100}Sn: magnetic moments of $N = {}^{50-54}$Ag. (in preparation).

Dea04 S. Dean *et al.*, The beta decay of neutron-deficient rhodium and ruthenium. *Eur. Phys. J.* **A21** (2004) 243–55.

Dil02 I. Dillmann *et al.*, Selective laser ionisation of $N \approx 82$ indium isotopes: the new r-process nuclide ^{135}In. *Eur. Phys. J.* **A13** (2002) 281–4.

Duo89 H.T. Duong *et al.*, Shape transition in neutron deficient Pt isotopes. *Phys. Lett.* **B217** (1989) 401–5.

Eli10 S. Eliseev *et al.*, Direct mass measurements of ^{194}Hg and ^{194}Au: a new route to the neutrino mass determination? *Phys. Lett.* **B693** (2010) 426–9.

Erd98 N. Erdmann, Determination of the first ionization potential of nine actinide elements by resonance ionization mass spectrometry (RIMS). *J. Alloys Compounds* **271–273** (1998) 837–40.

Ewa04 G. Ewald *et al.*, Nuclear charge radii of ^{8,9}Li determined by laser spectroscopy. *Phys. Rev. Lett.* **93** (2004) 113002.

Fac04 M. Facina , B. Bruyneel, S.Dean, J. Gentens, M. Huyse, Yu. Kudryavtsev, P. Van den Bergh, P. Van Duppen, A gas cell for thermalizing, storing and transporting radioactive ions and atoms. Part II: On-line studies with a laser ion source, *Nucl. Instrum. Meth.* **B226** (2004) 401-418.

Fed84 V.N. Fedoseyev, V.S. Letokhov, V.I. Mishin, G.D. Alkhazov, A.E. Barzakh, V.P. Denisov, A.G. Dernyatin, and V.S. Ivanov, Atomic lines isotope shifts of short-lived radioactive Eu studied by high-sensitive laser resonance photoionization method in 'on-line' experiments with proton beams. *Opt. Commun.* **52** (1984) 24–8.

Fed95a V.N. Fedoseyev *et al.*, Study of short-lived tin isotopes with a laser ion source. *AIP Conf. Proc.* **329** (1995) 465–7.

Fed95b V.N. Fedoseyev *et al.*, Study of short-lived silver isotopes with a laser ion source. *Z. Phys.* **A353** (1995) 9–10.

Fed97 V.N. Fedoseyev, Chemically selective laser ion source of manganese. *Nucl. Instrum. Methods Phys. Res.* **B126** (1997) 88–91.

Fed00 V.N. Fedoseyev, G. Huber, U. Köster, J. Lettry, V.I. Mishin, H.L. Ravn, and V. Sebastian, The ISOLDE laser ion source for exotic nuclei. *Hyperfine Interact.* **127** (2000) 409–16.

Fed08 V.N. Fedoseev *et al.*, ISOLDE RILIS: new beams, new facilities. *Nucl. Instrum. Methods Phys. Res.* **B266** (2008) 4378–82.

Fer12 R. Ferrer *et al.*, Performance of a high repetition pulse rate laser system for in-gas-jet laser ionization studies with the Leuven laser ion source at LISOL. *Nucl. Instrum. Methods Phys. Res.* **B291** (2012) 29–37.

Fer13 R. Ferrer *et al.*. Submitted to Phys. Lett. B, In-gas-cell laser ionization spectroscopy in the vicinity of 100Sn:Magnetic moments and mean-square charge radii of N=50-54 Ag.

Fer13a R. Ferrer *et al.*, In Gas Laser Ionization and Spectroscopy experiments at the Superconducting Separator Spectrometer (S3): Conceptual studies and preliminary design, *Nucl. Instrum. Methods Phys. Res.* **B** DOI:10.1016/j.nimb.2013.07.028

Fin13 D. Fink *et al.*First application of the Laser Ion Source and Trap (LIST) for on-line experiments at ISOLDE *Nucl. Instrum. Methods Phys. Res.* In Press http://dx.doi.org/10.1016/j.nimb.2013.06.039

Fra01 S. Franchoo *et al.*, Monopole migration in ^{69,71,73}Cu observed from β decay of laser-ionized $^{68-74}$Ni. *Phys. Rev. C* **64** (2001) 054308.

Got11 Gottwald, Tina. Studium hochkomplexer atomarer Spektren mittels Methoden der Laserresonanzionisation. Diss. PhD thesis, Johannes Gutenberg-Universität, 2011. German.

Han99 M. Hannawald *et al.*, Decay of neutron-rich Mn nuclides and deformation of heavy Fe isotopes. *Phys. Rev. Lett.* **82** (1999) 1391–4.

Han00 M. Hannawald *et al.*, Selective laser ionization of very neutron-rich cadmium isotopes: decay properties of ^{131}Cd^{83} and ^{132}Cd^{84}. *Phys. Rev.* **C62** (2000) 054301.

Hil92 T. Hilberath, S. Becker, G. Bollen, H.-J. Kluge, U. Krönert, G. Passler, and R. Wyss, Ground-state properties of neutron-deficient platinum isotopes. *Z. Phys.* **A342** (1992) 1–15.

Joh93 S.G. Johnson and B.L. Fearey, Spectroscopic study of thorium using continuous-wave resonance ionization mass spectrometry with ultraviolet ionization. *Spectrochim. Acta* **B48** (1993) 1065–77.

Jok97 A. Jokinen *et al.*, Selective laser ionization of radioactive Ni-isotopes. *Nucl. Instrum. Methods Phys. Res.* **B126** (1997) 95–9.

Jok98 A. Jokinen *et al.*, Beta decay of the $MT = -1$ nucleus ^{58}Zn studied by selective laser ionization. *Eur. Phys. J.* **A3** (1998) 271–6.

Kol12 K. Kolos, Probing the nuclear structure in the vicinity of 78Ni via beta decay spectroscopy of 84Ga, PhD Thesis Universite Paris Sud - Paris XI (2012).

Kos02a U. Köster *et al.*, Isomer separation and measurement of nuclear moments with the ISOLDE RILIS. *Hyperfine Interact.* **127** (2000) 417–20.

Kos02b U. Köster, Intense radioactive-ion beams produced with the ISOL method. *Eur. Phys. J.* **A15** (2002) 255–63.

Kos00b U. Köster *et al.*, Isomer separation of ^{70g}Cu and ^{70m}Cu with a resonance ionization laser ion source. *Nucl. Instrum. Methods Phys. Res.* **B160** (2000) 528–35.

Kos03a U. Köster *et al.*, Oxide fiber targets at ISOLDE. *Nucl. Instrum. Methods Phys. Res.* **B204** (2003) 303–13.

Kos03b U. Köster *et al.*, On-line yields obtained with the ISOLDE RILIS. *Nucl. Instrum. Methods Phys. Res.* **B204** (2003) 347–52.

Kos05 U. Köster, ISOLDE beams of neutron-rich zinc isotopes: yields, release, decay spectroscopy. *AIP Conf. Proc.* **798** (2005) 315–26.

Kos08 U. Köster, O. Arndt, E. Bouquerel, V.N. Fedoseyev, H. Franberg, A. Joinet, C. Jost, I.S.K. Kerkines, and R. Kirchner, Progress in ISOL target-ion-source systems. *Nucl. Instrum. Methods Phys. Res.* **B266** (2008) 4229–39.

Kra09 J. Krämer *et al.*, Nuclear ground-state spin and magnetic moment of ^{21}Mg. *Phys. Lett.* **B678** (2009) 465–9.

Kra98 K.-L. Kratz, Laser isotope and isomer separation of heavy Ag nuclides: half-life of the r-process waiting point nuclide ^{129}Ag and structure of neutron-rich Cd isotopes. Proc. Int. Conf. on Fission and Properties of Neutron-Rich Nuclei, Sanibel Island, USA, 10–15 November 1997, Eds. J.H. Hamilton and A.V. Ramayya (World Scientific, Singapore, 1998), p. 586.

Kra05 K.-L. Kratz, B. Pfeiffer, O. Arndt, S. Hennrich, and A. Wöhr, r-process isotopes in the ^{132}Sn region. *Eur. Phys. J.* **A25** (2005) 633–8.

Kro87 U. Krönert, S. Becker, T. Hilberath, H.-J. Kluge, and C. Schulz, Resonance ionization mass spectroscopy with a pulsed atomic beam. *Appl. Phys.* **A44** (1987) 339–45.

Kro91 U. Krönert, S. Becker, G. Bollen, M. Gerber, T. Hilberath, H.-J. Kluge, and G. Passler, On-line laser spectroscopy by resonance ionization of laser-desorbed, refractory elements. *Nucl. Instrum. Methods Phys. Res.* **A300** (1991) 522–37.

Kru02 K. Kruglov *et al.*, Production of neutron-rich copper isotopes in 30-MeV proton-induced fission of ^{238}U. *Nucl. Phys.* **A701** (2002) 145–9.

Kud96 Yu. Kudryavtsev, J. Andrzejewski, N. Bijnens, S. Franchoo, J. Gentens, M. Huyse, A. Piechaczek, J. Szerypo, I. Reusen, P. Van Duppen, P. Van Den Bergh, L. Vermeeren, J. Wauters, A. WÃűhr. Beams of Short Lived Nuclei Produced by Selective Laser Ionization in a Gas Cell. *Nucl. Instrum.Meth.* **B114**, 350-365, (1996).

Kud08 Yu. Kudryavtsev, T.E. Cocolios, J. Gentens, O. Ivanov, M. Huyse, D. Pauwels, M. Sawicka, T. Sonoda, P. Van den Bergh, P. Van Duppen, Characterization of the ISOL laser ion source using spontaneous fission of 252Cf , *Nucl. Instrum.Meth.* **B266** 4368-4372 (2008).

Kud09 Yu. Kudryavtsev, T.E. Cocolios, J. Gentens, M. Huyse, O. Ivanov, D. Pauwels, T. Sonoda, P. Van den Bergh, and P. Van Duppen, Dual chamber laser ion source at LISOL. *Nucl. Instrum. Methods Phys. Res.* **B267** (2009) 2908–17.

Las11 J. Lassen (ISAC/TRIUMF), private communication (2011).

Las09 J. Lassen, P. Bricault, M. Dombsky, J.P. Lavoie, M. Gillner, T. Gottwald, A. Hellbusch, A. Voss, and K.D.A. Wendt, Laser ion source operation at the TRIUMF radioactive ion beam facility. *AIP Conf. Proc.* **1104** (2009) 9–15.

Lau92 W. Lauth, H. Backe, M. Dahlinger, I. Klaft, P. Schwamb, and G. Schwickert, Resonance ionization spectroscopy in a buffer gas cell with radioactive decay detection, demonstrated using Tl. *Phys. Rev. Lett.* **68** (1992) 1675–8.

Lee87a L.K.P. Lee, J.E. Crawford, V. Raut, G. Savard, G. Thekkadath, T.H. Duong, and J. Pinard, Resonant ionization using synchronized laser pulses. *Nucl. Instrum. Methods Phys. Res.* **B26** (1987) 444–7.

Lee87b J.K.P. Lee, Resonant ionization spectroscopy of laser-desorbed gold isotopes. *AIP Conf. Proc.* **164** (1987) 207–8.

Lee88 J.K.P. Lee *et al.*, Charge–radius changes in even-A platinum nuclei. *Phys. Rev.* **C38** (1988) 2985–8.

Let92 V.S. Letokhov, V.I. Mishin, S.K. Sekatsky, V.N. Fedoseyev, G.D. Alkhazov, A.E. Barzakh, V.P. Denisov and V.E. Starodubsky, Laser spectroscopic studies of nuclei with neutron number $N < 82$ (Eu, Sm and Nd isotopes). *J. Phys. G: Nucl. Part. Phys.* **18** (1992) 1177–93.

Let98 J. Lettry *et al.*, Recent development of the ISOLDE laser ion source. *Rev. Sci. Instrum.* **69** (1998) 761–3.

Liu13 Y. Lui *et al.*, On-line commissioning of the HRIBF resonant ionization laser ion source. *Nucl. Instrum. Methods Phys. Res.* **B298** (2013) 5–12.

Mat10 Mattolat C, Gottwald T, Raeder S, Rothe S, Schwellnus F, Wendt K, ThÃűrle-Pospiech P and Trautmann N 2010, Determination of the first ionization potential of technetium *Phys. Rev.* **A81** (2010) 052513.

Mis87a V.I. Mishin, S.K. Sekatskii, V.N. Fedoseev, N.B. Buyanov, V.S. Letokhov, G.D. Alkhazov, A.E. Barzakh, V.P. Denisov, V.S. Ivanov, and I.Ya. Chubukov, Ultrasensitive resonance laser photoionization spectroscopy of the radioactive chain $^{157-172}$Tm produced by a proton accelerator. *Sov. Phys.–JETP* **66** (1987) 235–42 (Engl. transl.).

Mis87b V.I. Mishin *et al.*, Resonance photoionization spectroscopy and laser separation of ^{141}Sm and ^{164}Tm nuclear isomers. *Opt. Commun.* **51** (1987) 383–6.

Mis93 V.I. Mishin, V.N. Fedoseyev, H.-J. Kluge, V.S. Letokhov, H.L. Ravn, F. Scheerer, Y. Shirakabe, S. Sundell, and O. Tengblad, Chemically selective laser ion-source for the CERN-ISOLDE on-line mass separator facility. *Nucl. Instrum. Methods Phys. Res.* **B73** (1993) 550–60.

Mon93 L. Monz *et al.*, Fast, low-level detection of strontium-90 and strontium-89 in environmental samples by collinear ionization spectroscopy. *Spectrochim. Acta* **B48** (1993) 1655–71.

Mue00 W.F. Mueller *et al.*, Beta decay of ^{66}Co, ^{68}Co, and ^{70}Co. *Phys. Rev.* **C61** (2000) 054308.

Mue01 Müller P, Bushaw P A, Blaum K, Diel S, Geppert Ch, Nähler A, Trautmann N, Nörtershäuser W, Wendt K, 2001 ^{41}Ca ultratrace determination with isotopic selectivity >1012 by diode-laser-based RIMS, *Fresenius journal of analytical chemistry* **370** (01)508-12, doi: 10.1007/s002160100815

Muk04 M. Mukherjee *et al.*, The mass of ^{22}Mg. *Phys. Rev. Lett.* **93** (2004) 150801.

Oin00 M. Oinonen *et al.*, Laser ionization in β-decay studies of Zn and Mn nuclei. *Hyperfine Interact.* **127** (2000) 431–6.

Pau09 D. Pauwels *et al.*, Structure of ^{65}Co, ^{67}Co studied through the beta decay of ^{65}Fe, ^{67}Fe and a deep-inelastic reaction. *Phys. Rev.* **C79** (2009) 044309.

Pod10 Z. Podolyak, private communication (2010).

Pri06 E.J. Prime *et al.*, TRIUMF resonant ionization laser ion source. *Hyperfine Interact.* **171** (2006) 127–34.

Rae10 S. Raeder, Spurenanalyse von Aktiniden in der Umwelt mittels Resonanzionisations Resonanzionisations-massenspektrometrie Diss. PhD Thesis, Germany, 2010.

Rae11b S. Raeder, V. Sonnenschein, T. Gottwald, I.D. Moore, M. Reponen, S. Rothe, N. Trautmann, and K. Wendt, Resonance ionization

spectroscopy of thorium isotopes towards a laser spectroscopic identification of the low-lying 7.6 eV isomer of ^{229}Th. *J. Phys. B: At. Mol. Opt. Phys.* **44** (2011) 165005.

Rae12 Raeder S, Hakimi A, StÃűbener N, Trautmann N and Wendt K, Detection of plutonium isotopes at lowest quantities using in-sources resonance ionization mass spectrometry. *Anal. Bioanal. Chem.* **404** 2163-72 (2012).

Rae13a Raeder S, Dombski M, Hedden H, Lassen J, Quenzel T, SjÃűdin M, Teigelhöfer A, Wendt K, In-source laser spectroscopy developments at TRILIS - towards spectroscopy on actinium and scandium *Hyperfine Interac.* **216** 33-9 (2013).

Rae13b S. Raeder 2013, Private communication.

Reu99 I. Reusen *et al.*, β-decay study of ^{54,55}Ni produced by an element-selective laser ion source. *Phys. Rev.* **C59** (1999) 2416–21.

Rie93 J. Riegel *et al.*, Resonance ionization mass spectroscopy for trace analysis of neptunium. *Appl. Phys.* **B56** (1993) 275–80.

Ros12 Rossnagel J, Raeder S, Hakimi A, Ferrer R, Trautmann N and Wendt K, Determination of the first ionization potential of actinium. *Physical Review* **A85** (2012) 01252.

Rot13 S. Rothe *et al.*, Measurement of the first ionization potential of astatine by laser ionization spectroscopy, *Nature Commun.* **4** 1835 (2013) DOI: 10.1038/ncomms2819.

Rus89 W. Ruster *et al.*, A resonance ionization mass spectrometer as an analytical instrument for trace analysis. *Nucl. Instrum. Methods Phys. Res.* **A281** (1989) 547–58.

San06 R. Sànchez *et al.*, Nuclear charge radii of ^{9,11}Li: the influence of halo neutrons. *Phys. Rev. Lett.* **96** (2006) 033002.

Sau00 J. Sauvage *et al.*, COMPLIS experiments: collaboration for spectroscopy measurements using a pulsed laser ion source. *Hyperfine Interact.* **129** (2000) 303–17.

Sav90 G. Savard *et al.*, Laser spectroscopy of laser-desorbed gold isotopes. *Nucl. Phys.* **A512** (1990) 241–52.

Sch91 C. Schulz *et al.*, Resonance ionization spectroscopy on a fast atomic ytterbium beam. *J. Phys. B: At. Mol. Opt. Phys.* **24** (1991) 4831–44.

Sel06 M. Seliverstov *et al.*, Study of the neutron deficient $^{182-190}$Pb isotopes by simultaneous atomic- and nuclear-spectroscopy. *Hyperfine Interact.* **171** (2006) 225–31.

Sel13 M. Seliverstov *et al.*, Charge radii of odd-A 191-211Po isotopes. *Physics Letters* **B719** 362-366 (2013).

Sew03 M. Sewtz *et al.*, First observation of atomic levels for the element fermium ($Z = 100$). *Phys. Rev. Lett.* **90** (2003) 163002.

Sif06 R. Sifi *et al.*, Laser spectroscopy measurements of neutron-rich tellurium isotopes by COMPLIS. *Hyperfine Interact.* **171** (2006) 173–9.

Son12a V. Sonnenschein *et al.*, The search for the existence of ^{229m}Th at IGISOL. *Eur. Phys. J.* **A48** (2012) 52.

Son12b V. Sonnenschein *et al.*, Determination of the ground-state hyperfine structure in neutral ^{229}Th. *J. Phys. B: At. Mol. Opt. Phys.* **45** (2012) 165005.

Pro12 T.J. Procter *et al.*, Development of the CRIS (Collinear Resonant Ionisation Spectroscopy) beam line. *J. Phys.: Conf. Ser.* **381** (2012) 012070.

Ver00 D. Verney *et al.*, Nuclear moments of neutron-deficient iridium isotopes from laser spectroscopy. *Hyperfine Interact.* **127** (2000) 79–82.

Wal87 K. Wallmeroth *et al.*, Sudden change in the nuclear distribution of very light gold isotopes, *Phys. Rev. Lett.* **58** (1987) 1516–19.

Wal05 W.B. Walters, Structure and decay of neutron-rich nuclides in the 115–138 mass range and r-process nucleosynthesis. *AIP Conf. Proc.* **764** (2005) 335–9.

Wei99 L. Weissman *et al.*, Beta-decay of ^{67}Co. *Phys. Rev.* **C59** (1999) 2004–8.

Wei02b L. Weissman *et al.*, β+ decay of ^{61}Ga. *Phys. Rev.* **C65** (2002) 044321.

B Vapour pressures of the elements

Table B.1: Approximate values of vapour pressures for the elements at 2000°C, taken from the vapour pressure curves in Ref. [85]. The elements are sorted in order of decreasing vapour pressure.

Element	Vapour pressure (mbar at 2000°C)	Element	Vapour pressure (mbar at 2000°C)
At	>100	Au	11
Hg	>100	Ge	9.3
S	>100	Er	9.3
P	>100	Ho	9.3
Fr	>100	Sc	8.0
K	>100	Fe	4.0
Rb	>100	Ni	2.7
Se	>100	Gd	2.7
As	>100	Co	2.0
Cs	>100	Pd	2.0
Na	>100	Nd	1.3
Po	>100	Y	0.9
Cd	>100	Si	0.9
Zn	>100	Tb	0.4
Te	>100	Pr	0.4
Sr	>100	Ti	0.3
Mg	>100	La	0.1
Tl	>100	Pu	0.1
Sb	>100	V	9E−02
Pb	>100	Lu	8E−02
Li	>100	Ac	4E−02
Eu	>100	Ce	2E−02
Yb	>100	Rh	1E−02
Ba	>100	B	9E−03
Sm	>100	Pt	4E−03
Ca	>100	U	1E−03
Bi	>100	Ru	1E−04
Mn	>100	Th	7E−05
Tm	>100	Mo	3E−05
In	>100	Ir	2E−05
Ag	>100	C	1E−05
Ga	>100	Nb	1E−05
Al	80	Zr	2E−06
Be	67	Hf	2E−06
Cu	27	Tc	3E−07
Sn	27	Os	8E−08
Cr	20	Re	3E−08
Dy	15	Ta	3E−08
Am	12	W	1E−09

C Thermal populations

Table C.1: Thermal population of levels of selected elements that are populated inside a typical hot-cavity environment. The values are calculated from the Boltzmann distribution at 2200°C and take into account the statistical weights ($2J + 1$) of each level. For each element, the level with the highest population appears in **bold** type.

Element	J	Energy (cm^{-1})	Population at 2200°C (%)	Element	J	Energy (cm^{-1})	Population at 2200°C (%)
Al	0.5	0	35	Pm	**2.5**	**0**	**28**
	1.5	**112**	**65**		3.5	804	24
Sc	1.5	0	42		4.5	1749	18
	2.5	**168**	**58**		5.5	2797	12
Ti	2	0	27	Sm	0	0	6
	3	170	34		1	293	16
	4	**387**	**39**		**2**	**812**	**20**
Co	**4.5**	**0**	**44**		**3**	**1490**	**20**
	3.5	816	23		4	2273	16
	2.5	1407	12		5	3126	12
Ni	**4**	**0**	**36**	Gd	2	0	16
	3	205	25		3	215	20
	2	880	12		**4**	**533**	**21**
	3	1332	13		5	999	20
Ga	0.5	0	45		6	1719	16
	1.5	**826**	**55**	Tb	**7.5**	**0**	**21**
Ge	0	0	18		6.5	286	16
	1	557	39		7.5	462	16
	2	**1410**	**41**		5.5	510	12
Se	**2**	**0**	**81**	Dy	**8**	**0**	**89**
	1	1989	15		7	4134	8
Y	1.5	0	47	Lu	**1.5**	**0**	**64**
	2.5	**530**	**53**		2.5	1994	32
Tc	**2.5**	**0**	**54**	Ac	**1.5**	**0**	**70**
	4.5	2573	22		2.5	2231	30
	3.5	3251	12	Th	**2**	**0**	**55**
Pd	**0**	**0**	**80**		3	2869	16
	3	6564	15	U	**6**	**0**	**46**
In	**0.5**	**0**	**63**		5	620	27
	1.5	2213	37	Np	**5.5**	**0**	**52**
Sn	0	0	34		4.5	2034	14
	1	**1692**	**40**	Pu	**0**	**0**	**33**
	2	3428	25		1	2204	29
La	**1.5**	**0**	**33**		2	4300	15
	2.5	1053	28	Cm	2	0	16
	3.5	3495	10		3	302	19
Ce	**4**	**0**	**18**		4	816	19
	2	229	9		**6**	**1214**	**22**
Pr	**4.5**	**0**	**40**		5	1764	14
	5.5	1377	22	Ti	2	0	27
	6.5	2847	12		3	170	34
Nd	**4**	**0**	**40**		**4**	**387**	**39**
	5	1128	26				
	6	2367	16				

D Ionization scheme data sources

- **R. L. Kurucz database**: http://www.cfa.harvard.edu/amp/ampdata/kurucz23/sekur.html
- **NIST database**: http://physics.nist.gov/PhysRefData/ASD/lines_form.html
- The datasheets of E.B. Saloman:
  As, B, Cd, C, Ge, Au, Fe, Pb, Si and Zn [86]
  Ga, Mn, Sc and Tl [87]
  Al, Ca, Cs, Cr, Co, Cu, Kr, Mg, Hg and Ni [88]
  Be, In, Li, K, Rb, Ag, Ti, V and Ni update [89]

D.1 The RILIS ionization scheme database

A convenient and open database of RILIS ionization schemes has been written by Martin Klein and Sebastian Rothe. Details are reported in Ref. [22]. It is a Web-based resource that can be accessed and developed by the laser ion source community:

RILIS elements database: http://riliselements.web.cern.ch/riliselements/

Users of the database are asked to contribute by entering data concerning ionization schemes that have been applied. Additional information such as laser details, AIS search energy regions and saturation requirements can be added. Using this as a central resource for sharing and communicating ionization scheme developments will be of considerable benefit to all users of laser ion source facilities.

Acknowledgements

I would like to acknowledge the numerous contributions from colleagues for supplying material for the lecture slides and many of the figures included in this chapter. The main contributors were Jonathan Billowes, Klaus Blaum, Valentin Fedosseev, Kieran Flanagan, Yuri Kudrayastev, Jans Lassen, Viacheslav Mishin, Iain Moore, Wilfried Nörtershäuser, Sebastian Rothe, Piet Van Duppen and Klaus Wendt.

References

[1] Y. Blumenfeld, Radioactive ion beam facilities in Europe, *Nucl. Instrum. Methods Phys. Res.* **B266**(19–20) (2008) 4074–4079.

[2] Y. Blumenfeld, T. Nilsson, and P. Van Duppen, Facilities and methods for radioactive ion beam production, *Phys. Scr.* **152** (2013) 014023.

[3] U. Köster, How to produce intense and pure ISOL beams, *Prog. Part. Nucl. Phys.* **46**(1) (2001) 411–412.

[4] G. Münzenberg, The separation techniques for secondary beams, *Nucl. Instrum. Methods Phys. Res.* **B70**(1–4) (1992) 265–275.

[5] R. Kirchner, Ion sources for radioactive beams and related problems, *Rev. Sci. Instrum.* **67**(3) (1996) 928.

[6] R. Kirchner, On the thermoionization in hot cavities, *Nucl. Instrum. Methods Phys. Res.* **A292**(2) (1990) 203–208.

[7] R. Kirchner and E. Roeckl, Investigation of gaseous discharge ion sources for isotope separation on-line, *Nucl. Instrum. Methods* **133**(2) (1976) 187–204.

[8] U. Köster, Resonance ionization laser ion sources, *Nucl. Phys.* **A701**(1–4) (2002) 441–451.

[9] V.N. Fedoseyev, V.I. Mishin, L. Weissman, M. Huyse, K. Kruglov, W.F. Mueller, P. Van Duppen, J. Van Roosbroeck, P. Thirolf, H.G. Thomas, D. Weisshaar, W. Schulze, R. Borcea, M. La Commara, G. Huber, V. Sebastian, K.L. Kratz, H. Schatz, K. Schmidt, R. Catherall, U. Georg, J. Lettry,

M. Oinonen, H.L. Ravn, and H. Simon, Isomer separation of Cu and ^{70m}Cu with a resonance ionization laser ion source, *Nucl. Instrum. Methods Phys. Res.* **B160** (2000) 528–535.

[10] U. Köster, R. Catherall, V.N. Fedoseyev *et al.*, Isomer separation and measurement of nuclear moments with the ISOLDE RILIS, *Hyperfine Interact.* **127** (2000) 417–420.

[11] R.V. Ambartzumian and V.S. Letokhov, Selective two-step (STS) photoionization of atoms and photodissociation of molecules by laser radiation, *Appl. Opt.* **11**(2) (1972) 354.

[12] V.S. Letokhov, *Laser Photoionization Spectroscopy* (Academic Press, Orlando, FL, 1987).

[13] W. Demtröder, *Atoms, Molecules and Photons* (Graduate Texts in Physics, Springer, Berlin, 2010).

[14] C.J. Foot, *Atomic Physics* (Oxford Master Series in Physics, Oxford University Press, New York, 2005).

[15] D.C. Morton, Atomic data for resonance absorption lines. I – Wavelengths longward of the Lyman limit, *Astrophys. J. Suppl. Ser.* **77** (1991) 119.

[16] A. Einstein, Strahlungs-Emission und -Absorption nach der Quantentheorie, *Deutsche Physikalische Gesellschaft* **18** (1916) 318–323.

[17] K. Blum, *Density Matrix Theory and Applications* (Springer Series on Atomic, Optical, and Plasma Physics, vol. 64, Springer, Berlin, 2012).

[18] B.W. Shore and P.R. Berman, Theory of coherent atomic excitation, *Phys. Today* **44**(12) (1991) 71.

[19] A.E. Siegman, *Lasers* (University Science Books, Mill Valley, CA, 1986).

[20] R. Ferrer, Low-energy radioactive isotope beam production by in-gas laser ionization for decay spectroscopy, 2012 Gas Cell Workshop, RIKEN-Wako.

[21] B.A. Marsh, L.-E. Berg, D.V. Fedorov, V.N. Fedosseev, O.J. Launila, M. Lindroos, R. Losito, F.K. Österdahl, T. Pauchard, I.T. Pohjalainen, U. Sassenberg, M.D. Seliverstov, A.M. Sjödin, and G. Tranströmer, The ISOLDE RILIS pump laser upgrade and the LARIS Laboratory, *Hyperfine Interact.* **196**(1–3) (2010) 129–141.

[22] S. Rothe, An all-solid state laser system for the laser ion source RILIS and in-source laser spectroscopy of astatine at ISOLDE, CERN, PhD thesis, Johannes Gutenberg-Universitat, 2012.

[23] S. Rothe, A.N. Andreyev, S. Antalic, A. Borschevsky, L. Capponi, T.E. Cocolios, H. De Witte, E. Eliav, D. V Fedorov, V.N. Fedosseev, D.A. Fink, S. Fritzsche, L. Ghys, M. Huyse, N. Imai, U. Kaldor, Y. Kudryavtsev, U. KÃűster, J.F.W. Lane, J. Lassen, V. Liberati, K.M. Lynch, B.A. Marsh, K. Nishio, D. Pauwels, V. Pershina, L. Popescu, T.J. Procter, D. Radulov, S. Raeder, M.M. Rajabali, E. Rapisarda, R.E. Rossel, K. Sandhu, M.D. Seliverstov, A.M. SjÃűdin, P. Van den Bergh, P. Van Duppen, M. Venhart, Y. Wakabayashi, and K.D.A. Wendt, Measurement of the first ionization potential of astatine by laser ionization spectroscopy, *Nature Communications* **4** 1835 (2013).

[24] T. Gallagher, S. Edelstein, and R. Hill, Radiative lifetimes of the S and D Rydberg levels of Na, *Phys. Rev.*, **A11**(5) (1975) 1504–1506.

[25] I.I. Beterov, D.B. Tretyakov, I.I. Ryabtsev, V.M. Entin, A. Ekers, and N.N. Bezuglov, Ionization of Rydberg atoms by blackbody radiation, *New J. Phys.* **11**(1) (2009) 013052.

[26] R.V. Ambartsumian, V.P. Kalinin, and V.S. Letokhov, Selective two-step photoionization of rubidium atoms by laser radiation, *JETP Lett.* **13** (1971) 217–219 (Engl. transl.).

[27] V.N. Fedosseev, Yu. Kudryavtsev, and V.I. Mishin, Resonance laser ionization of atoms for nuclear physics, *Phys. Scr.* **85**(5) (2012) 58104.

[28] G.D. Alkhazov, L.Kh. Batist, A.A. Bykov, V.D. Vitman, V.S. Letokhov, V.I. Mishin, V.N. Panteleyev, S.K. Sekatsky, and V.N. Fedoseyev, Application of a high efficiency selective laser ion source at the IRIS facility, *Nucl. Instrum. Methods Phys. Res.* **A306**(1–2) (1991) 400–402.

[29] V. Mishin, V. Fedoseyev, H. Kluge, V. Letokhov, H. Ravn, F. Scheerer, Y. Shirakabe, S. Sundell, and O. Tengblad, Chemically selective laser ion-source for the CERN-ISOLDE on-line mass separator facility, *Nucl. Instrum. Methods Phys. Res.* **B73**(4) (1993) 550–560.

[30] J.R.J. Bennett, Delays in thick targets, *Nucl. Phys.* **A701**(1–4) (2002) 296–302.

[31] I.D. Moore, A. Nieminen, J. Billowes, P. Campbell, C. Geppert, A. Jokinen, T. Kessler, B. Marsh, H. Penttilä, S. Rinta-Antila, B. Tordoff, K.D.A. Wendt, and J. Äystö, Development of a laser ion source at IGISOL, *J. Phys. G: Nucl. Part. Phys.* **G31**(10) (2005) S1499–1502.

[32] Yu.A. Kudryavtsev, J. Andrzejewski, N. Bijnens, S. Franchoo, M. Huyse, A. Piechaczek, J. Szerypo, I. Reusen, P. Van Duppen, L. Vermeeren, J. Wauters, and A. WoỲĹh, Laser ion source for the Leuven Isotope Separator On-Line, *Rev. Sci. Instrum.* **67**(3) (1996) 938.

[33] Yu. Kudryavtsev, M. Facina, M. Huyse, J. Gentens, P. Van den Bergh, and P. Van Duppen, Beams of isotopes produced at LISOL by laser ionization after thermalization of energetic ions in a gas cell, *Nucl. Instrum. Methods Phys. Res.* **B204** (2003) 336–342.

[34] I.D. Moore, T. Kessler, T. Sonoda, Y. Kudryavstev, K. Peräjärvi, A. Popov, K.D.A. Wendt, and J. Äystö, A study of on-line gas cell processes at IGISOL, *Nucl. Instrum. Methods Phys. Res.* **B268**(6) (2010) 657–670.

[35] Yu. Kudryavtsev, T.E. Cocolios, J. Gentens, M. Huyse, O. Ivanov, D. Pauwels, T. Sonoda, P. Van den Bergh, and P. Van Duppen, Dual chamber laser ion source at LISOL, *Nucl. Instrum. Methods Phys. Res.* **B267**(17) (2009) 2908–2917.

[36] M. Reponen, I.D. Moore, I. Pohjalainen, T. Kessler, P. Karvonen, J. Kurpeta, B. Marsh, S. Piszczek, V. Sonnenschein, and J. Äystö, Gas jet studies towards an optimization of the IGISOL LIST method, *Nucl. Instrum. Methods Phys. Res.* **A635**(1) (2011) 24–34.

[37] M. Reponen, I.D. Moore, T. Kessler, I. Pohjalainen, S. Rothe, and V. Sonnenschein, Laser developments and resonance ionization spectroscopy at IGISOL, *Eur. Phys. J.* **A48**(4) (2012) 45.

[38] Yu. Kudryavtsev, R. Ferrer, M. Huyse, P. Van den Bergh, and P. Van Duppen, The in-gas-jet laser ion source: resonance ionization spectroscopy of radioactive atoms in supersonic gas jets, *Nucl. Instrum. Methods Phys. Res.* **B297** (2013) 7–22.

[39] M. Wada, A. Takamine, T. Sonoda, K. Okada, and P. Schury, Developments at the SLOWRI facility at RIKEN: precision optical spectroscopy of $^{7,9,10,11}Be^{+}$ ions, *Hyperfine Interact.* **199**(1–3) (2011) 269–277.

[40] T Sonoda, The development of the gas cell-based laser ion source for RIKEN PALIS, *AIP Conf. Proc* **1104** (2009) 132–137.

[41] T. Sonoda, M. Wada, H. Tomita, C. Sakamoto, T. Takatsuka, T. Furukawa, H. Iimura, Y. Ito, T. Kubo, Y. Matsuo, H. Mita, S. Naimi, S. Nakamura, T. Noto, P. Schury, T. Shinozuka, T. Wakui, H. Miyatake, S. Jeong, H. Ishiyama, Y.X. Watanabe, Y. Hirayama, K. Okada, and A. Takamine, Development of a resonant laser ionization gas cell for high-energy, short-lived nuclei, *Nucl. Instrum. Methods Phys. Res.* **B295** (2013) 1–10.

[42] T. Kubo. In-flight RI beam separator BigRIPS at RIKEN and elsewhere in Japan. *Nucl. Instrum. Methods Phys. Res.* **B204** (2003) 97–113.

[43] E. Kugler, ISOLDE Laboratory portrait, *Hyperfine Interact.* **129**(1–4) (2000) 12.

[44] V.N. Fedosseev, L.-E. Berg, D.V. Fedorov, D. Fink, O.J. Launila, R. Losito, B.A. Marsh, R.E. Rossel, S. Rothe, M.D. Seliverstov, A.M. Sjödin, and K.D.A. Wendt, Upgrade of the resonance ionization laser ion source at ISOLDE on-line isotope separation facility: new lasers and new ion beams, *Rev. Sci. Instrum.* **83**(2) (2012) 02A903.

[45] S. Rothe, B.A. Marsh, C. Mattolat, V.N. Fedosseev, and K. Wendt, A complementary laser system for ISOLDE RILIS, *J. Phys.: Conf. Ser.* **312**(5) (2011) 52020.

[46] H. Walther, Application of tunable dye lasers to atomic spectroscopy, *Phys. Scr.* **9**(5) (1974) 297–305.

[47] F.P. Schäfer, *Dye Lasers* (Topics in Applied Physics, vol. 1, Springer, Berlin, 1973).

[48] T.W. Hänsch, Repetitively pulsed tunable dye laser for high resolution spectroscopy, *Appl. Opt.*

11(4) (1972) 895–898.

[49] Sirah, Sirah webpage, *http://www.sirah.com/2013*.

[50] Exciton, Dye tuning ranges, *http://www.exciton.com/laserdyescurves.html*.

[51] K.F. Wall, Titanium sapphire lasers, *Lincoln Lab. J.* **3**(3) (1990) 447–462.

[52] C. Mattolat, S. Rothe, F. Schwellnus, T. Gottwald, S. Raeder, K. Wendt, T. Iguchi, and K. Watanabe, An all-solid-state high repetition rate titanium:sapphire laser system for resonance ionization laser ion sources, *AIP Conf. Proc.* **1104** (2009) 114–119.

[53] P.A. Franken, A.E. Hill, C.W. Peters, and G. Weinreich, Generation of optical harmonics, *Phys. Rev. Lett.* **7**(1) (1961) 118–120.

[54] N. Erdmann, V. Sebastian, V.N. Fedoseyev, M. Hannawald, G. Huber, T. Kautzsch, K.-L. Kratz, V.I. Mishin, M. Nunnemann, G. Passler, and N. Trautmann, Frequency tripling for resonance ionization spectroscopy (RIS) of Cd, *Appl. Phys. B: Lasers Opt.* **B66**(4) (1998) 431–433.

[55] J. Lassen, P. Bricault, M. Dombsky, F. Izdebski, J.P. Lavoie, M. Gillner, T. Gottwald, F. Hellbusch, A. TeigelhoÌĹfer, A. Voss, K.D.A. Wendt, F.D. McDaniel, and B.L. Doyle, Solid-state laser, resonant ionization laser ion source (RILIS) and laser beam transport at radioactive ion beam facilities, *AIP Conf. Proc.* **1099** (2009) 769–773.

[56] B. Marsh, Suitability test of a high beam quality Nd:YVO_4 industrial laser for the ISOLDE RILIS installation, Technical Report, 2013.

[57] G.J. Focker, PicoAmpereMeter (PAM) *http://isolde.web.cern.ch/isolde/group/PG-minutes-2011/20110504_Mini-PicoAmpMeter-talk_Physisists.pdf*, (2011).

[58] J. Lettry, R. Catherall, P. Drumm, P. Van Duppen, A. H. M. Evensen, G. J. Focker, A. Jokinen, O. C. Jonsson, E. Kugler, and H. Ravn, Pulse shape of the ISOLDE radioactive ion beams *Nucl. Instrum. Methods Phys. Res.*, **126**, no. 1-4, pp. 130-134, (1997).

[59] B. Lommel, W. Hartmann, B. Kindler, J. Klemm, and J. Steiner, Preparation of self-supporting carbon thin films, *Nucl. Instrum. Methods Phys. Res.* **A480**(1) (2002) 199–203.

[60] M. Turek, A. Droździel, K. Pyszniak, D. Maczka, and B. Slowinski, Simulations of ionization in a hot cavity surface ion source, *Rev. Sci. Instrum.* **83**(2) (2012) 023303.

[61] M. Menna, R. Catherall, J. Lettry, E. Noah, and T. Stora, R&D for the development of negative ion beams of halogens, *Nucl. Instrum. Methods Phys. Res.*, **266** 19-20, pp. 4391-4393 (2008).

[62] K. Wies, C. Geppert, K. Blaum, K. Brück, H.-J. Kluge, S. Schwarz, and K. Wendt, Development towards a laser ion source trap for the production of exotic species, *Hyperfine Interact.* **162**(1–4) (2006) 29–38.

[63] F. Schwellnus, K. Blaum, C. Geppert, T. Gottwald, H.-J. Kluge, C. Mattolat, W. Nörtershäuser, K. Wies, and K. Wendt, The laser ion source and trap (LIST) – a highly selective ion source, *Nucl. Instrum. Methods Phys. Res.* **B266**(19–20) (2008) 4383–4386.

[64] F. Schwellnus, K. Blaum, C. Geppert, T. Gottwald, H.-J. Kluge, C. Mattolat, W. Nörtershäuser, K. Wies, and K. Wendt, The laser ion source and trap (LIST) - A highly selective ion source, *Nucl. Instrum. Methods Phys. Res.* **266**(19-20) (2008) 4383-4386.

[65] K. Blaum, J. Dilling, and W. Nörtershäuser, Precision atomic physics techniques for nuclear physics with radioactive beams, *Physica Scripta* **T152** (2013) 014017.

[66] V. Sebastian, R. Catherall, V. N. Fedoseyev, U. Georg, G. Huber, Y. Jading, O. Jonsson, U. Köster, M. Koizumi, K.-L. Kratz, E. Kugler, J. Lettry, V. I. Mishin, H. L. Ravn, C. Tamburella, and A. Wöhr, Spectroscopic applications of the ISOLDE laser ion source, in *Proceedings of Exotic nuclei and atomic masses* **ENAM 98** (1998) 126-129.

[67] F. Schwellnus, K. Blaum, R. Catherall, B. Crepieux, V. Fedosseev, T. Gottwald, H.-J. Kluge, B. Marsh, C. Mattolat, S. Rothe, T. Stora, and K. Wendt, The laser ion source trap for highest isobaric selectivity in online exotic isotope production, *Rev. Sci. Instrum.* **81**(2) (2010) 02A515.

[68] H.-J. Kluge and W. Nörtershäuser, Lasers for nuclear physics, *Spectrochim. Acta B: At. Spectrosc.* **B58**(6) (2003) 1031–1045.

[69] K. Heilig and A. Steudel, Changes in mean-square nuclear charge radii from optical isotope shifts, *At. Data Nucl. Data Tables* **14**(5–6) (1974) 613–638.

[70] W.H. King, *Isotope Shifts in Atomic Spectra*, 1st edn (Springer, Berlin, 1984).

[71] B. Cheal and K.T. Flanagan, Progress in laser spectroscopy at radioactive ion beam facilities, *J. Phys. G: Nucl. Part. Phys.* **G37**(11) (2010) 113101.

[72] K. Anton, S. Kaufman, W. Klempt, G. Moruzzi, R. Neugart, E. Otten, and B. Schinzler, Collinear laser spectroscopy on fast atomic beams, *Phys. Rev. Lett.* **40**(10) (1978) 642–645.

[73] G.D. Alkhazov, A.E. Barzakh, V.P. Denisov, K.A. Mezilev, Yu.N. Novikov, V.N. Panteleyev, A.V. Popov, E.P. Sudentas, V.S. Letokhov, V.I. Mishin, V.N. Fedoseyev, S.V. Andreyev, D.S. Vedeneyev, and A.D. Zyuzikov, A new highly efficient method of atomic spectroscopy for nuclides far from stability, *Nucl. Instrum. Methods Phys. Res.* **B69**(4) (1992) 517–520.

[74] L. Weissman, U. Köster, R. Catherall, S. Franchoo, U. Georg, O. Jonsson, V. Fedoseyev, V. Mishin, M. Seliverstov, J. Van Roosbroeck, S. Gheysen, M. Huyse, K. Kruglov, G. Neyens, and P. Van Duppen, Magnetic moments of ^{68}Cu^{g,m} and ^{70}Cu^{g,m1,m2} nuclei measured by in-source laser spectroscopy, *Phys. Rev.* **C65**(2) (2002) 024315.

[75] K. Flanagan, P. Vingerhoets, M. Avgoulea, J. Billowes, M. Bissell, K. Blaum, B. Cheal, M. De Rydt, V. Fedosseev, D. Forest, C. Geppert, U. Köster, M. Kowalska, J. Krämer, K. Kratz, A. Krieger, E. Mané, B. Marsh, T. Materna, L. Mathieu, P. Molkanov, R. Neugart, G. Neyens, W. Nörtershäuser, M. Seliverstov, O. Serot, M. Schug, M. Sjoedin, J. Stone, N. Stone, H. Stroke, G. Tungate, D. Yordanov, and Yu. Volkov, Nuclear spins and magnetic moments of Cu^{71,73,75}: inversion of π^2p$^{3/2}$ and π^1f$^{5/2}$ levels in Cu^{75}, *Phys. Rev. Lett.* **103**(14) (2009) 142501.

[76] M. Seliverstov, A. Andreyev, N. Barré, H. Witte, D. Fedorov, V. Fedoseyev, S. Franchoo, J. Genevey, G. Huber, M. Huyse, U. Köster, P. Kunz, S. Lesher, B. Marsh, B. Roussière, J. Sauvage, P. Duppen, and Yu. Volkov, Study of the neutron deficient $^{182-190}$Pb isotopes by simultaneous atomic- and nuclear-spectroscopy, *Hyperfine Interact.* **171**(1–3) (2007) 225–231.

[77] H. De Witte, A. Andreyev, N. Barré, M. Bender, T. Cocolios, S. Dean, D. Fedorov, V. Fedoseyev, L. Fraile, S. Franchoo, V. Hellemans, P. Heenen, K. Heyde, G. Huber, M. Huyse, H. Jeppessen, U. Köster, P. Kunz, S. Lesher, B. Marsh, I. Mukha, B. Roussière, J. Sauvage, M. Seliverstov, I. Stefanescu, E. Tengborn, K. Van de Vel, J. Van de Walle, P. Van Duppen, and Yu. Volkov, Nuclear charge radii of neutron-deficient lead isotopes beyond $N = 104$ midshell investigated by in-source laser spectroscopy, *Phys. Rev. Lett.* **98**(11) (2007) 112502.

[78] M.D. Seliverstov, A.N. Andreyev, N. Barré, A.E. Barzakh, S. Dean, H. De Witte, D.V. Fedorov, V.N. Fedoseyev, L.M. Fraile, S. Franchoo, J. Genevey, G. Huber, M. Huyse, U. Köster, P. Kunz, S.R. Lesher, B.A. Marsh, I. Mukha, B. Roussière, J. Sauvage, I. Stefanescu, K. Van de Vel, P. Van Duppen, and Yu.M. Volkov, Charge radii and magnetic moments of odd-A $^{183-189}$Pb isotopes, *Eur. Phys. J.* **A41**(3) (2009) 315–321.

[79] B. Marsh, In-source laser resonance ionization at ISOL facilities, PhD thesis, University of Manchester, February 2007.

[80] T.E. Cocolios, A.N. Andreyev, S. Antalic, A.E. Barzakh, B. Bastin, J. Büscher, I.G. Darby, W. Dexters, D.V. Fedorov, V.N. Fedosseev, K.T. Flanagan, S. Franchoo, S. Fritzsche, G. Huber, M. Huyse, M. Keupers, U. Köster, Yu. Kudryavtsev, E. Mané, B.A. Marsh, P.L. Molkanov, R.D. Page, M.D. Seliverstov, A.M. Sjödin, I. Stefan, J. Van de Walle, P. Van Duppen, M. Venhart, and S.G. Zemlyanoy, Early onset of deformation in the neutron-deficient polonium isotopes, *J. Phys.: Conf. Ser.* **381** (2012) 012072.

[81] T. Cocolios, A. Andreyev, B. Bastin, N. Bree, J. Büscher, J. Elseviers, J. Gentens, M. Huyse, Yu. Kudryavtsev, D. Pauwels, T. Sonoda, P. Van den Bergh, and P Van Duppen, Magnetic dipole

moment of $Cu^{57,59}$ measured by in-gas-cell laser spectroscopy, *Phys. Rev. Lett.* **103**(10) (2009) 5–8.

[82] S. Fritzsche, Ratip – a toolbox for studying the properties of open-shell atoms and ions, *J. Electron Spectrosc. Rel. Phenom.* **114–116** (2001) 1155–1164.

[83] I. Stefanescu, G. Georgiev, F. Ames, J. Äystö, D. Balabanski, G. Bollen, P. Butler, J. Cederkäll, N. Champault, T. Davinson, A. Maesschalck, P. Delahaye, J. Eberth, D. Fedorov, V. Fedosseev, L. Fraile, S. Franchoo, K. Gladnishki, D. Habs, K. Heyde, M. Huyse, O. Ivanov, J. Iwanicki, J. Jolie, B. Jonson, Th. Kröll, R. Krücken, O. Kester, U. Köster, A. Lagoyannis, L. Liljeby, G. Bianco, B. Marsh, O. Niedermaier, T. Nilsson, M. Oinonen, G. Pascovici, P. Reiter, A. Saltarelli, H. Scheit, D. Schwalm, T. Sieber, N. Smirnova, J. Walle, P. Duppen, S. Zemlyanoi, N. Warr, D. Weisshaar, and F. Wenander, Coulomb excitation of ^{68,70}Cu: first use of postaccelerated isomeric beams, *Phys. Rev. Lett.* **98**(12) (2007) 68–71.

[84] J. Van Roosbroeck, C. Guénaut, G. Audi, D. Beck, K. Blaum, G. Bollen, J. Cederkall, P. Delahaye, A. De Maesschalck, H. De Witte, D. Fedorov, V. Fedoseyev, S. Franchoo, H. Fynbo, M. Górska, F. Herfurth, K. Heyde, M. Huyse, A. Kellerbauer, H.-J. Kluge, U. Köster, K. Kruglov, D. Lunney, V. Mishin, W. Mueller, Sz. Nagy, S. Schwarz, L. Schweikhard, N. Smirnova, K. Van de Vel, P. Van Duppen, A. Van Dyck, W. Walters, L. Weissman, and C. Yazidjian, Unambiguous identification of three β-decaying isomers in Cu^{70}, *Phys. Rev. Lett.* **92**(11) (2004) 112501.

[85] K. Marton, *Vacuum Physics and Technology Methods of Experimental Physics* **14**, Elsevier Science, Amsterdam, (1980).

[86] E.B. Saloman, A resonance ionization spectroscopy/resonance ionization mass spectrometry data service. I – Data sheets for As, B, Cd, C, Ge, Au, Fe, Pb, Si and Zn, *Spectrochim. Acta B: At. Spectrosc.* **B45**(1–2) (1990) 37–83.

[87] E.B. Saloman, A resonance ionization spectroscopy/resonance ionization mass spectrometry data service: V – Datasheets for Ga, Mn, Sc and Tl, *Spectrochim. Acta B: At. Spectrosc.* **B49**(3) (1994) 251–281.

[88] E.B. Saloman, A resonance ionization spectroscopy/resonance ionization mass spectrometry data service. II – Datasheets for Al, Ca, Cs, Cr, Co, Cu, Kr, Mg, Hg and Ni, *Spectrochim. Acta B: At. Spectrosc.* **B46**(3) (1991) 319–378.

[89] E.B. Saloman, A resonance ionization spectroscopy/resonance ionization mass spectrometry data service: IV – Datasheets for Be, In, Li, K, Rb, Ag, Ti and V and an update of the datasheet for Ni, *Spectrochim. Acta B: At. Spectrosc.* **B48**(9) (1993) 1139–1203.

[90] V. N. Fedosseev, Y. Kudryavtsev, and V. I. Mishin, Resonance laser ionization of atoms for nuclear physics, *Physica Scripta* **85**(5) (2012) 058104.

[91] F. Le Blanc, L. Cabaret, J. E. Crawford, S. Essabaa, V. Fedoseyev, W. Geithner, J. Genevey, M. Girod, R. Horn, G. Huber, S. Kappertz, J. Lassen, J. K. P. Lee, G. Le Scornet, V. Mishin, R. Neugart, J. Obert, J. Oms, a. Ouchrif, S. Peru, J. Pinard, H. Ravn, B. Roussière, J. Sauvage, and D. Verney, Charge radius change in the heavy tin isotopes until A = 132 from laser spectroscopy, *The European Physical Journal A - Hadrons and Nuclei* **15**(1-2) (2002) 49-51.

[92] T. E. Cocolios, H. H. Al Suradi, J. Billowes, I. Budincevic, R. P. de Groote, S. De Schepper, V. N. Fedosseev, K. T. Flanagan, S. Franchoo, R. F. Garcia Ruiz, H. Heylen, F. Le Blanc, K. M. Lynch, B. A. Marsh, P. J. R. Mason, G. Neyens, J. Papuga, T. J. Procter, M. M. Rajabali, R. E. Rossel, S. Rothe, G. S. Simpson, A. J. Smith, I. Strashnov, H. H. Stroke, D. Verney, P. M. Walker, K. D. A. Wendt, and R. T. Wood, The Collinear Resonance Ionization Spectroscopy (CRIS) experimental setup at CERN-ISOLDE, *Nucl. Instrum. Methods Phys. Res.*, In press (2013).

[93] P. Kunz, G. Huber, G. Passler, N. Trautmann, and K. Wendt, Resonance Ionization Mass Spectrometry (RIMS) with Pulsed and CW-Lasers on Plutonium, *Hyperfine Interactions* **162**(1-4) (2006) 159-166.

[94] R. Ferrer, B. Bastin, D. Boilley, P. Creemers, P. Delahaye, E. Lienard, X. Flechard, S. Franchoo, L.

Ghys, M. Huyse, Y. Kudryavtsev, N. Lecesne, H. Lu, F. Lutton, E. Mogilevskiy, D. Pauwels, J. Piot, D. Radulov, L. Rens, H. Savajols, J. C. Thomas, E. Traykov, C. Van Beveren, P. Van den Bergh, and P. Van Duppen, In gas laser ionization and spectroscopy experiments at the Superconducting Separator Spectrometer (S3): Conceptual studies and preliminary design, *Nucl. Instrum. Methods Phys. Res.*, In press (2013).

[95] R. E. Rossel, V. N. Fedosseev, B. A. Marsh, D. Richter, S. Rothe, and K. D. A. Wendt, Data acquisition, remote control and equipment monitoring for ISOLDE RILIS, *Nucl. Instrum. Methods Phys. Res.*, In press (2013).

[96] J. P. Lavoie, P. Bricault, J. Lassen, and M. R. Pearson, Segmented linear radiofrequency quadrupole/laser ion source project at TRIUMF, *Hyperfine Interactions* **174**(1-3) (2007) 33-39.

[97] D. A. Fink, S. D. Richter, B. Bastin, K. Blaum, R. Catherall, T. E. Cocolios, D. V. Fedorov, V. N. Fedosseev, K. T. Flanagan, L. Ghys, A. Gottberg, N. Imai, T. Kron, N. Lecesne, K. M. Lynch, B. A. Marsh, T. M. Mendonca, D. Pauwels, E. Rapisarda, J. P. Ramos, R. E. Rossel, S. Rothe, M. D. Seliverstov, M. Sjodin, T. Stora, C. Van Beveren, and K. D. A. Wendt, First application of the Laser Ion Source and Trap (LIST) for on-line experiments at ISOLDE, *Nucl. Instrum. Methods Phys. Res.*, In press (2013).

[98] S. Raeder, Laser-2013 Workshop proceedings, *Hyperfine Interactions*, In preparation (2013).

[99] V. I. Mishin, A. L. Malinovsky, D. V. Mishin, T. Iguchi, and K. Watanabe, Resonant Ionization Laser Ion Source (RILIS) With Improved Selectivity Achieved By Ion Pulse Compression Using In-Source Time-of-flight Technique, in *AIP Conference Proceedings* **1104** (2009) 207-212.

[100] M. D. Seliverstov, T. E. Cocolios, W. Dexters, A. N. Andreyev, S. Antalic, A. E. Barzakh, B. Bastin, J. Buscher, I. G. Darby, D. V. Fedorov, V. N. Fedoseyev, K. T. Flanagan, S. Franchoo, S. Fritzsche, G. Huber, M. Huyse, M. Keupers, U. Koster, Y. Kudryavtsev, B. A. Marsh, P. L. Molkanov, R. D. Page, A. M. Sjodin, I. Stefan, J. Van de Walle, P. Van Duppen, M. Venhart, and S. G. Zemlyanoy, Charge radii of odd-A 191-211Po isotopes, *Physics Letters* **B719**(4-5) (2013) 362-366.

[101] S. Rothe, V. N. Fedosseev, T. Kron, B. A. Marsh, R. E. Rossel, and K. D. A. Wendt, Narrow linewidth operation of the RILIS titanium: Sapphire laser at ISOLDE/CERN, *Nucl. Instrum. Methods Phys. Res.* In Press (2013).

Volume and Surface-Enhanced Volume Negative Ion Sources

M.P. Stockli
Spallation Neutron Source, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

Abstract
H^- volume sources and, especially, caesiated H^- volume sources are important ion sources for generating high-intensity proton beams, which then in turn generate large quantities of other particles. This chapter discusses the physics and technology of the volume production and the caesium-enhanced (surface) production of H^- ions. Starting with Bacal's discovery of the H^- volume production, the chapter briefly recounts the development of some H^- sources, which capitalized on this process to significantly increase the production of H^- beams. Another significant increase was achieved in the 1990s by adding caesiated surfaces to supplement the volume-produced ions with surface-produced ions, as illustrated with other H^- sources. Finally, the focus turns to some of the experience gained when such a source was successfully ramped up in H^- output and in duty factor to support the generation of 1 MW proton beams for the Spallation Neutron Source.

1 Introduction

This chapter discusses the production of large quantities of negative hydrogen ions (H^-) to form small-diameter H^- beams, which then are accelerated to high energies to create powerful proton beams. Colliding with other beams or impacting on specially engineered targets, these proton beams will then produce high yields of the desired secondary particles, such as neutrons in the 1 MW Spallation Neutron Source (SNS) [1]. The discussion focuses on the successful high-yield H^- production with volume and caesiated-volume sources, such as the SNS H^- source [2], the DESY H^- source at the Deutsches Electronen Synchrotron [3], the J-PARC H^- source at the Japan Proton Accelerator Research Complex [4], and a few others that were important for those developments. It excludes the compact surface plasma sources, which are the topic of a chapter by D. Faircloth.

The SNS H^- source was specified with requirements that exceeded the capabilities of all existing H^- sources [5]. Although there were H^- sources successfully producing more than the required 50 mA H^- beam current, those sources were operated at duty factors that were orders of magnitude smaller. Although there were H^- sources successfully operating with duty factors exceeding the 6–7% required for SNS, those sources were producing only a fraction of the required 50 mA. At that time, it was commonly assumed that increasing the duty factor would inversely decrease the lifetime. Accordingly, the specified 6–7% duty factor made the requirement for three-week-long, maintenance-free service cycles [6] a leap of faith.

In 1994 Lawrence Berkeley National Laboratory (LBNL) hosted a workshop to evaluate options and issues for an H^- source, a low energy beam transport (LEBT) system, and beam chopping for SNS [7]. Although a broad range of required R&D was identified, the radio-frequency (RF)-driven, Cs-enhanced, multicusp H^- source, which was developed at LBNL and tested at SSC, looked very promising [8]. The Cs cartridges, containing only a few milligrams of Cs, were very successful in tests at LBNL, and later at the Superconducting Super Collider (SSC) [9], although the duty factor was 60 times smaller than the SNS requirement. Also, to intercept the co-extracted electrons before they gain

prohibitively high energies, a strong dipole magnet was integrated into the outlet aperture [10], an elegant and compact, but untested, solution that remains to be fully evaluated.

Since 2009 SNS has been operating near 1 MW, with the H^- ions supplied with slightly modified LBNL H^- sources. These deliver daily ~230 C of H^- ions, which is unprecedented for pulsed H^- sources for accelerators [2].

In six-week maintenance-free service cycles the sources deliver ~10 kC or ~2.7 A h of H^- ions, which is also unprecedented for pulsed H^- sources for accelerators [2]. Ramping up the duty factor 25-fold over a three-year period revealed many unprecedented issues, which had to be understood and mitigated to continue the ramp-up with an acceptable availability.

Accordingly, the CERN Accelerator School (CAS) programme committee requested the inclusion of some of the SNS lessons learned, including the Cs_2CrO_4 system, and scaling rules for H^- sources.

2 Historical background

Negative ion sources for accelerators were initially developed to allow the sources to be placed on easily accessible, service- and control-friendly, isolated platforms that are negatively charged to a few tens of kilovolts. So called 'tandems' accelerate the negative ions to a highly charged positive high-voltage (HV) terminal, strip at least two electrons, and then accelerate the resulting positive ions back to ground, which doubles the energy of singly charged, or multiplies the energy of multiply charged, heavy ions.

Double-stripping H^- ions produces protons, which travel with opposite curvature in magnetic fields. This change of direction is used in cyclotrons to extract protons without the need of an extraction channel, which is prone to arcing and strong activation.

Accumulator rings can easily accumulate identical particles as long as the particle bunches remain separated in space so that kickers can be activated between bunches. This limit can be overcome by stripping negative ions inside an injection magnet, after which the now positive ions join the trajectories of the already stored positive ions. For example, SNS accelerates ~1000 beamlets of ~40 mA H^- ions, which are (mostly) double stripped in the injection magnet. Here they join the proton beam in the accumulator ring, which growths to several tens of amperes as a result. This allows the delivery of all protons to the Hg spallation target in less than 1 μs, although it takes ~1 ms to produce those protons.

The production of negative ions was a specialty limited to a few accelerator laboratories. However, that changed when it became clear that magnetically confined fusion needed neutral beam heating to reach the temperatures required for fusion reactors. Many researchers and laboratories started programmes to model, study, or develop large negative ion sources and the required neutralizers for a new class of high-voltage, high-current accelerators.

The extensive research on H^- production and the development of H^- sources are primarily documented in the proceedings of the 'International Symposia on the Production and Neutralization of Negative Ions and Beams' (PNNIB), which were mostly held at Brookhaven National Laboratory. Except for the first two, all proceedings are available as conference proceedings from the American Institute of Physics (AIP). Reduced funding in the 1990s reduced the interest and the symposia stopped in 1997. The symposium was restarted in 2002 in part to support the unprecedented challenge SNS was facing. In 2008 the biennial symposium was renamed as the 'International Symposium on Negative Ions, Beams and Sources' (NIBS), while the proceedings continue to be published by AIP.

3 The volume production of H^- ions

The second electron on an H^- ion is bound with 0.75 eV, about 20 times less than the ~15 eV ionization energy of hydrogen atoms or molecules. This mismatch makes the direct formation of H^- ions in plasma rare events, such as the radiative electron capture by hydrogen atoms, which peaks with $\sim 10^{-18}$ cm^2 near 1 eV [11]. Even less likely is the dissociative attachment of fast electrons (>7 eV) to ground-state hydrogen molecules, yielding H^- ions with a cross-section of $\sim 10^{-20}$ cm^2 [12].

Accordingly, it was rather surprising when in 1977 M. Bacal found signals of large negative ion populations in hydrogen plasma. Years of research have shown that the dominant production occurs when highly rovibrationally excited hydrogen molecules ($4 \leq \nu \leq 9$) disintegrate after colliding with slow electrons (~1 eV), having cross-sections of up to $\sim 10^{-15}$ cm^2 [11, 12].

Highly rovibrationally excited molecules are easily produced with fast electrons (>20 eV), with cross-sections of up to 5×10^{-18} cm^2 [12, 13]. Unfortunately, such fast electrons (>5 eV) destroy H^- ions rather rapidly, with cross-sections of up to 4×10^{-15} cm^2 [11, 12], severely limiting the H^- population in hot plasma. H^- ions have much longer lifetimes in cold plasma because of their ionization threshold at 0.75 eV.

This problem of production and destruction was overcome with tandem sources [14, 15], which contain a magnetic filter between the plasma-producing region and the ion outlet, as shown in Fig. 1. The magnetic dipole field generates the Lorentz force, which is perpendicular and proportional to the velocity of charged particles. This force turns fast electrons around, returning them towards the filament. The Lorentz force has fewer effects on slow, collisional electrons and even slower protons, which diffuse through the filter field and generate much colder plasma near the ion outlet, so extending the lifetime of the locally generated H^- ions.

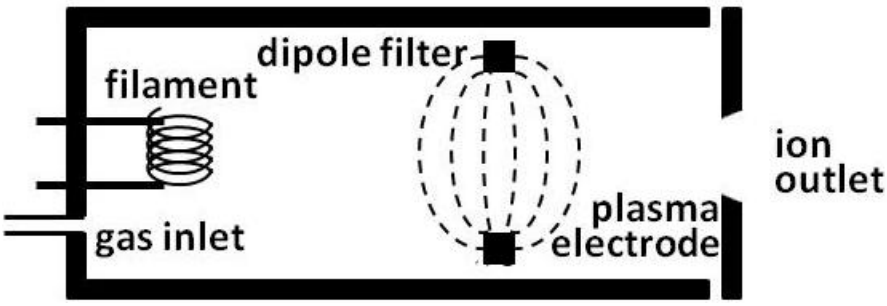


Fig. 1: Schematic of a tandem source with a magnetic filter

Filter magnets keep the destructive hot electrons away from the extraction region. However, the loosely bound second electron can also be detached by photons ($\sim 10^{-17}$ cm^2) [12] or in collisions with H atoms or molecules ($\sim 10^{-15}$ cm^2) [11, 12]. The dominant loss is probably the mutual neutralization with protons, especially with slow protons (e.g., $\sim 7 \times 10^{-14}$ cm^2 for 0.5 eV H^+) [12]. Accordingly, H^- ions formed near the source outlet have a much better chance to be extracted, and therefore the filter fields are now located close to the outlet.

4 Volume negative ion sources

Many so-called 'volume' negative ion sources have been developed that deliver predominantly H^- ions produced in the volume near the outlet. Foremost is 'Camembert' [16] at the Ecole Polytechnique in Palaiseau, France, which is extensively used to study the volume production of H^- and D^- ions [17]. A filament discharge generates plasma, which is confined by multicusp magnets and features a plasma electrode [14].

In the 1980s LBNL started to develop filament-driven H^- sources. In the late 1980s a small source with a 75 mm inner diameter was developed using all permanent magnets as shown in Fig. 2(a) [18]. Fourteen rows of water-cooled $SmCo_5$ magnets provide radial confinement, and another four rows in the back flange provide rear confinement. The filter field was generated with a pair of water-cooled permanent magnets mounted on the outlet flange. Placing a pair of $SmCo_5$ magnets 40 mm apart produced the highest H^- current when ~1 V was applied the plasma electrode (PE), as the outlet electrode is normally called. Up to 2 mA H^- peak current was extracted through a 1 mm diameter outlet when 80 kW of discharge power was applied to generate the 1 ms long pulses [18].

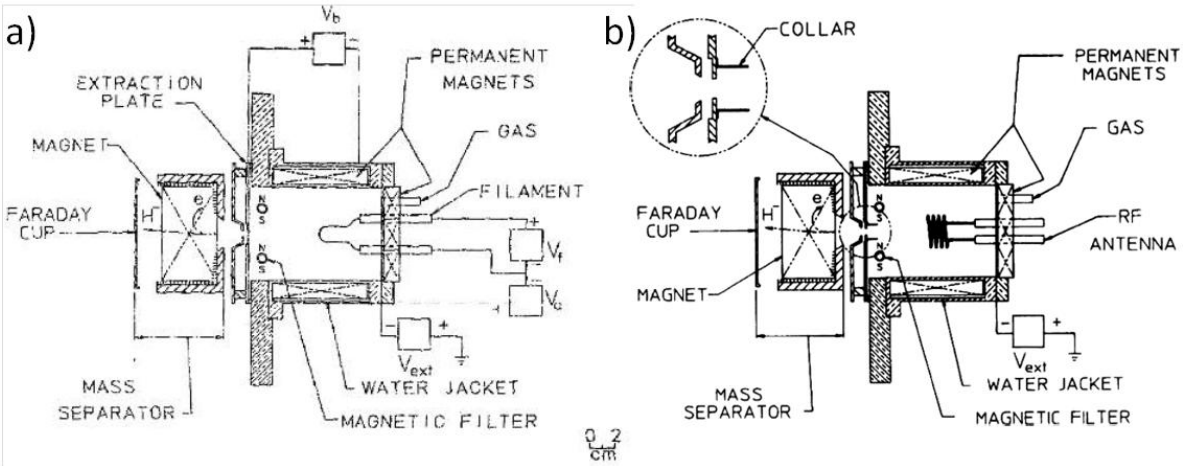


Fig. 2: Schematics of the small (a) filament-driven and (b) RF-driven volume H− source developed at LBNL

In 1990 LBNL started to develop an H^- source for the SSC. Replacing the filament in their small multicusp source by a three-turn inductive coil, shown in Fig. 2(b), increased the H^- current from 4.2 mA to 6 mA [19]. The outlet had 2 mm diameter and the plasma was driven with 25 kW discharges. A collar surrounded the outlet, which was shown to reduce the number of co-extracted electrons.

In 1992, the LBNL-built SSC H^- source met its requirement by producing 30 mA with a two-turn antenna driven by 35–45 kW RF and ignited with a W filament [20]. Studying the emittance of the H^- beam, it was shown that that a ~45% downstream taper of the 6.3 mm outlet yield a favourable emittance and divergence [21].

After acquiring a copy of the LBNL RF-driven H^- source in 1994, DESY found the antenna lifetime unacceptable [22]. Despite testing other configurations, DESY was unable to produce robust antennas, which would yield adequate H^- beam current with an adequate lifetime [23]. Finally, in the late 1990s DESY decided to develop an external antenna H^- source where the antenna is wound around an Al_2O_3 cylinder, which forms the wall of the plasma chamber shown in Fig. 3(a) (adapted from [23]). In addition, it was decided to reduce operational risks [24] by not using Cs.

The resulting source was very successful, producing a 40 mA H^- beam for 0.15 ms at 6 Hz for over 100 weeks [25]. A pulse length of 3 ms [26] was demonstrated and 60 mA peak currents were also achieved [25]. Figure 3(b) (from [25]) shows the segmented collar that was studied to elucidate the function of the collar. Briefly, the collar allows the electron density to be lowered, especially at the entrance of the collar. This draws positive ions to the collar wall, where they neutralize to form hyper-thermal neutrals, which in turn can form additional rovibrationally exited molecules, enhancing the production of H^- ions [25].

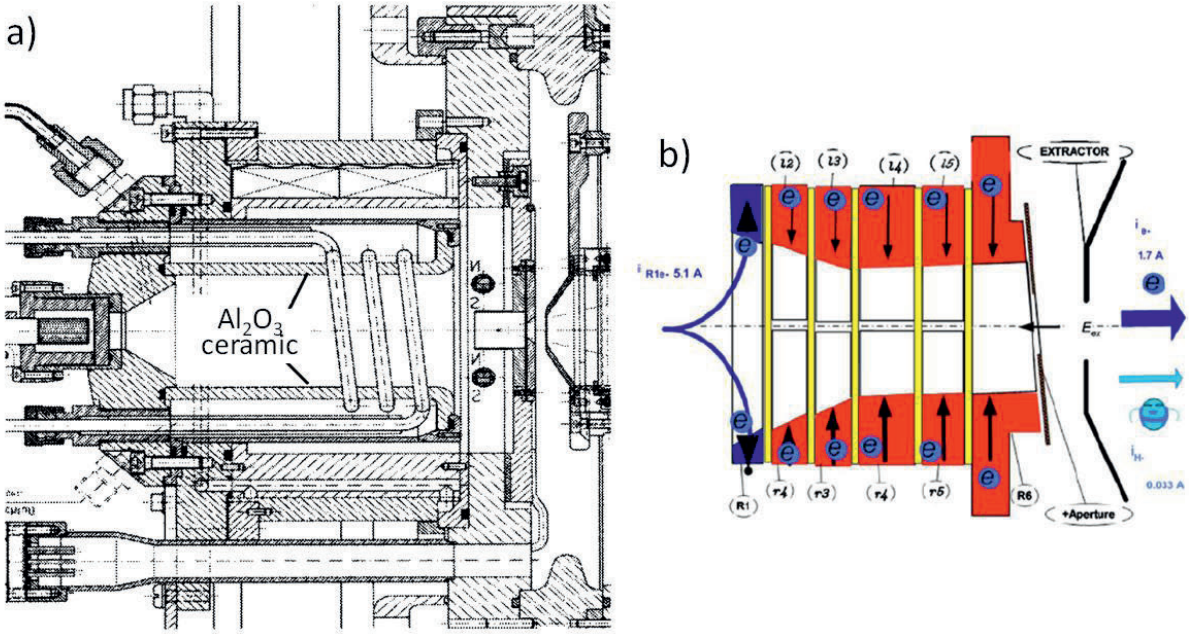


Fig.3: (a) Schematic and (b) detailed outlet collar of the external antenna H− source developed at DESY ((a) adapted from [23]; (b) from [25]).

In the late 1980s TRIUMF started to develop a 10 cm inner diameter, filament-driven H^- source to inject a continuous H^- beam into their cyclotron [27]. Figure 4(a) shows a cross-section through the source with two of the ten rows of $SmCo_5$ magnets that produce the plasma-confining multicusp field. At the outlet end, two diametrically opposed magnets are reversed to produce the filter field [27]. This causes the electron temperature to drop from ~2 eV in the centre of the source to <0.3 eV at the 13 mm diameter outlet, which is shown in detail in Fig. 4(b) [28, 29]. The source yields up to 15 mA of H^- at 20–30 kV, producing ~3 mA per kilowatt of required power [26], which is very efficient for H^- sources. The design has been licensed and the source is commercially available [30]. Operation near the peak current requires the filament to be replaced about every two weeks [30].

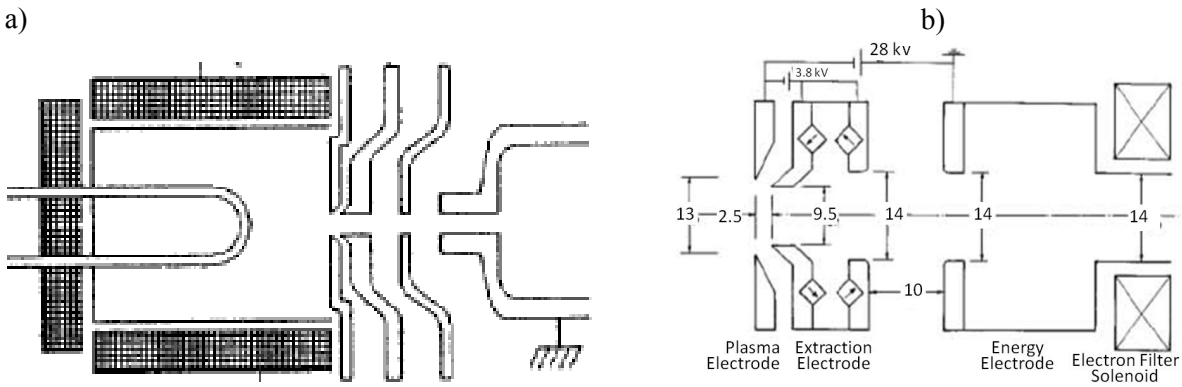


Fig. 4: (a) Schematic and (b) detailed extraction region of the 15 mA continuous H^- source developed by TRIUMF ((a) adapted from [27]; (b) adapted from [26]).

To optimize the extraction of the 'volume'-produced H^- ions, production 'volume' H^- sources have large outlets of the order of 1 cm diameter, in contrast to the small source outlets featured by the compact surface plasma sources discussed by D. Faircloth. The large outlet causes significant leakage

of neutral hydrogen gas, typically 20–40 sccm (standard cubic centimetres per minute) , which needs to be differentially pumped to avoid large stripping losses in the LEBT (~10^{-15} cm^2 for $E_{H^-} > 100\,\text{eV}$) [12].

5 The surface production of H⁻ ions

Compared to the ~15 eV ionization energy of hydrogen atoms and molecules, it is easier to remove electrons from metal surfaces, which have work functions between 4 and 5.6 eV, as the electron affinities of surfaces are called [12]. This is relevant because the walls of the plasma container are continuously exposed to a flux of thermal hydrogen molecules, which tend to stick for a while and later thermally desorb. In addition the plasma generates significant fluxes of hyper-thermal atoms and molecules, which also impact on the surfaces. There are also fast positive hydrogen ions (H^+, H_2^+ and H_3^+), which are accelerated in the sheath by the plasma potential, and then Auger-neutralize when they reach the surface. Owing to their kinetic energy, the hyper-thermal neutrals and neutralized ions are likely to bounce back or alternatively sputter an adsorbed hydrogen atom or molecule with a hyper-thermal velocity.

When a hydrogen atom leaves the surface, a conduction electron near the surface can get trapped in the field of the atom, forming an H^- ion. However, the surface work function is always more attractive than the 0.75 eV electron affinity of H atoms. Accordingly the dominant fraction of electrons return to the surface, especially for atoms leaving the surface slowly, thus giving the electrons time to choose.

Rasser gives a low-velocity approximation for the probability β^- for an atom to form a negative ion:

$$\beta^-(v_\perp) \approx (2/\pi)\exp[-\pi(\Phi - S)/(2av_\perp)], \tag{1}$$

where $v_\perp$ is the emitted atom's velocity normal to the surface, Φ is the work function of the surface, S is the electron affinity of the atom, and a is a decay constant [31]. The approximation shows that the probability depends exponentially on the inverse of the normal velocity, yielding very small probabilities for small velocities. Furthermore, it depends exponentially on the difference between the work function and the electron affinity, which for hydrogen is dominated by the larger work function.

Rasser's full model yields a probability that saturates for large velocities (>100 eV H) at ~4% for H on clean W(110). However, the probability drops to ~2% for ~10 eV H atoms [31]. The work function of W(110) is 5.25 eV, slightly higher than the 4.95 eV for Mo(110) [32], and accordingly Mo should yield similar percentages.

It is possible to enhance the probabilities by lowering the work function through the adsorption of alkali atoms [32]. Figure 5 shows the approximate work function of a Mo surface as a function of coverage by adsorbed Cs atoms. The coverage is measured in units of a monolayer, a one-atom thick layer of densely packed Cs atoms. The work function starts near 4.6 eV, the value of a clean, polycrystalline Mo substrate, free of Cs. As Cs is gradually added, the work function decreases and reaches a minimum of ~1.6 eV near 0.6 monolayer [32]. Adding more Cs increases the work function until an equilibrium value is reached close to 2.1 eV, the work function of bulk Cs.

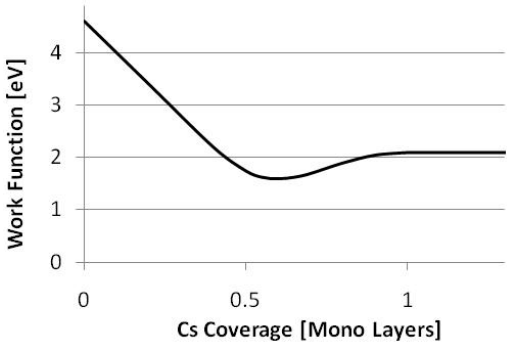


Fig. 5: Work function for Mo surfaces as a function of adsorbed Cs

For a W(110) surface covered with 0.5 monolayer of Cs, Rasser calculated ~40% probability for fast H atoms to form negative ions, and ~30% for 10 eV H atoms [31]. Rasser's calculations have been confirmed by measurements of up to 34% H^- ions when a proton beam was reflected at small grazing angle from a caesiated W(110) surface with a 1.45 eV work function [33].

While these are impressive fractions, they apply only to H^+ ions that bounce back elastically after hitting a Cs atom or a heavy atom of the substrate. Molecular ions have smaller velocities, and many of the ions lose a significant fraction of their energy when scattering with the light hydrogen atoms and molecules adsorbed on the surface, some of which are sputtered.

Much smaller percentages can be expected for atoms ejected due to impacting hyper-thermal atoms and molecules. Okuyama scattered 0.11–0.14 eV hydrogen atoms from a Cs-covered Mo substrate, and found ~0.02% H^- production probabilities, which are in reasonable agreement with Eq. (1) [34]. Model calculations find large fluxes of hyper-thermal atoms, and their contributions to H^- production can outnumber the contributions from the ions [35].

It is important to understand that many processes contribute to the extracted H^- beam in many interconnected ways. Optimizing the performance of an H^- source must always be judged from the carefully measured, extracted H^- beam and its emittance.

6 Filament-driven surface-enhanced volume H^- ion sources

In the 1990s the Japan Atomic Energy Research Institute (JAERI) started to develop H^- ion sources for a multipurpose high-energy proton accelerator project. The development produced a 150 mm inner diameter, multicusp H^- source driven by two filaments, featuring a Mo plasma electrode with an 8 mm outlet. Adding Cs increased the extracted H^- threefold, yielding up to 72 mA with 56 kW discharge power [36]. Unfortunately the lifetime was limited by the filament, which fell short of the three-week project requirement.

In the 2000s the Japan Atomic Energy Agency (JAEA) and the High Energy Accelerator Research Organization (KEK) constructed J-PARC jointly. Concerns that Cs may compromise the performance of the radio-frequency quadrupole (RFQ) accelerator drove the development of a new H^- source that could operate without Cs. After developing long-lived, highly emissive LaB_6 filaments, optimizing the geometry and the local fields, shown in Fig. 6(a), and optimizing the Mo plasma electrode with a 9 mm outlet and the extraction system, shown in Fig. 6(b), up to 38 mA current was obtained for ~50 kW discharge power [37]. The 2.5 eV work function of LaB_6, the careful shaping of the plasma electrode, its coating of mostly B and some La after use [38], as well as its high operational temperature suggest that a significant contribution of surface-produced H^- ions [37] can be obtained without using Cs. This source has served J-PARC with high availability, easily meeting the 1200 h lifetime requirement for the 17 mA, 0.5 ms long pulses at 25 Hz [4].

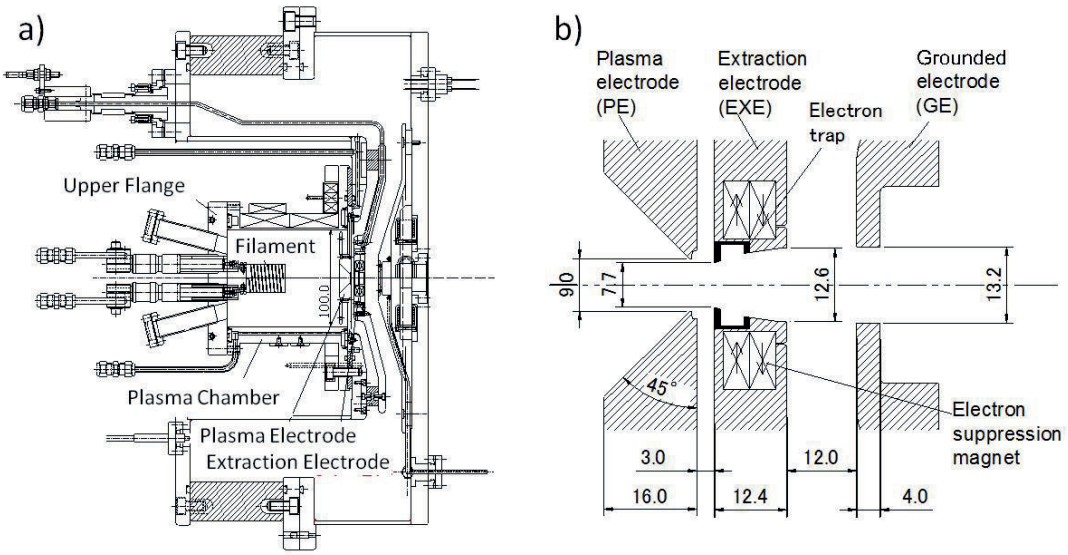


Fig. 6: (a) Schematic and (b) detailed extraction system of the H^- source developed and used at J-PARC [39]

However, 60 mA will be required to upgrade J-PARC to 1 MW with 0.5 ms pulses at 25 Hz and a three-week lifetime. Renewed filament optimizations yielded up to 45 mA [38], missing the 60 mA goal. Adding Cs yielded only a modest enhancement for the LaB_6 filaments [37]. Using a W filament yielded 18 mA before and 72 mA after caesiation, with a discharge power of only 16 kW. However, this level was maintained only for a limited time and there was consumption of the Cs [37].

Arc discharges sputter the filaments, which limits their lifetime. Some of the sputtered material coats the surfaces near the outlet, which in some cases can enhance the surface production of H^- ions. However, it can also sputter the Cs or cover the Cs layer, and therefore filament-driven H^- sources require a steady supply of Cs. The J-PARC experiments suggest that using Cs in filament-driven sources can only yield long lifetimes at low duty factors.

7 RF-driven surface-enhanced volume H^- ion sources

In 1992 the electron-suppressing collar of the RF-driven LBNL H^- source was replaced with a collar holding two Cs cartridges, shown in Fig. 7(a). This increased the H^- current output by a factor of three [40], converting this source to a surface-enhanced volume H^- source.

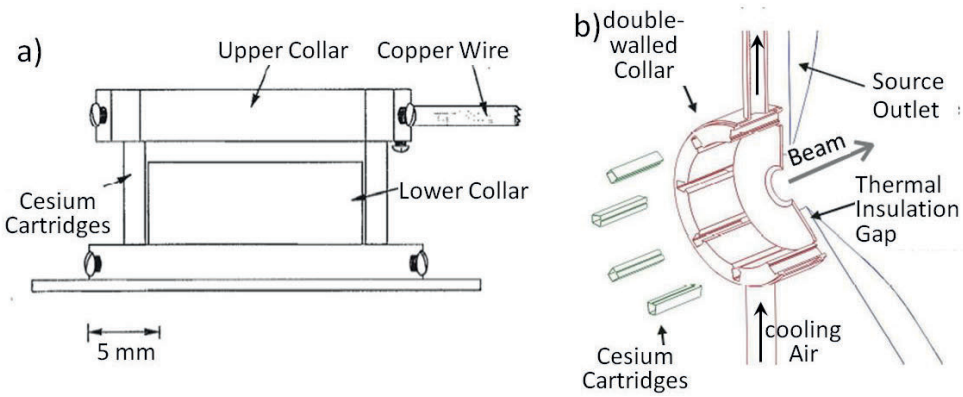


Fig. 7: Schematics of the (a) prototype and (b) SNS Cs collar developed at LBNL

In 1995, the same collar was tested in the SSC H^- source, which produced up to 100 mA and continued for three days to deliver over 80 mA for 0.1 ms at 10 Hz [9]. Again, a volume source was converted to a surface-enhanced volume H^- source.

In the late 1990s, LBNL started the detailed design of the SNS source, building on previous success and incorporating the advances achieved with the SSC source. Several changes had to be made to address the drastically higher duty factor of the SNS source. While the use of Cs cartridges was adopted from the SSC source, they were integrated into an air-cooled collar to control their temperature, as shown in Fig. 7(b). Increasing the collar temperature to 400°C increased the H^- current extracted through the 7 mm outlet from 18 mA to 56 mA for a 30 kW RF discharge, while drastically reducing the co-extracted electron current [41].

In addition, a dipole magnet was integrated into the outlet flange to drive the co-extracted electrons sideways onto the e-dump electrode, shown in Fig. 8(a) [10]. Several ignition schemes were tested, but these were abandoned in favour of a continuous 13 MHz discharge [42]. To limit the emittance growth, a very compact LEBT was designed. It comprised two electrostatic einzel lenses, with the second lens split into four quadrants to steer and chop the beam [43]. The source and LEBT were successfully commissioned at LBNL, producing 50 mA beam pulses [44]. In 2002 it was shipped to, installed and recommissioned at Oak Ridge National Laboratory (ORNL). After initial problems with high duty factors, the commissioning of the SNS accelerator was performed with short pulses at low repetition rates, which yielded ~99% availability for the ion source and LEBT system [45].

Challenges arose in 2007 after stepping up to 15 Hz and trying to extend the pulse length to 0.3 ms. However, these problems were analysed and mitigated, normally by the start of the next ~20 week run [45, 46], yielding availabilities mostly between 95 and 99%.

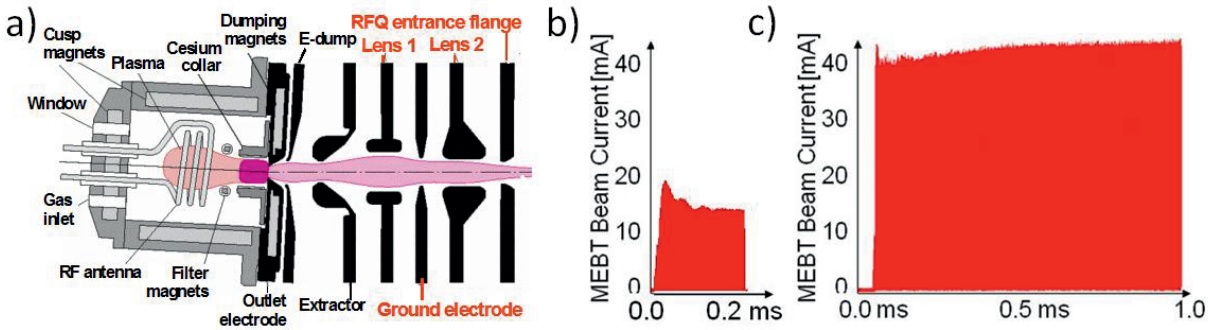


Fig. 8: (a) Schematic of the SNS H^- source and LEBT; (b) RFQ output in 2007; and (c) RFQ output in 2009

Ramping up the duty factor from ~0.5% to ~5% over a two-year period required the modification of configurations and procedures [46]. The result was an increase in the H^- beam current from ~13 mA in summer of 2007, as shown in Fig. 8(b), to ~40 mA in autumn of 2009, as shown in Fig. 8(c), which exceeded the 38 mA requirement for ~1 MW of beam power.

A series of interesting lessons was learned regarding the radio frequency: ~300 W of 13 MHz sustains continuous plasma, which eases [42] but does not guarantee the breakdown of the pulsed high-power plasma. The high-power plasma pulses are generated by superimposing high power at 2 MHz for the desired pulse length and repetition rate.

In 2007, when the high-power RF pulse length was extended beyond 0.1 ms, the beam current initially dropped off, as seen in Fig. 8(b). However, increasing the matching capacity increased the latter part of the current while lowering the initial overshoot, just as desired.

Ramping up the H^- beam current required increasing the RF power and the match accordingly, which frequently led to plasma outages. Trying to match the high-power plasma, which includes plasma inductance, increasingly mismatched the lower-inductance initial state until the induced electric fields were too weak to break down the plasma [47]. Using a compromise tune and increasing the matching network inductance reduced the problem, but did not eliminate it.

In 2009 an increasing number of plasma outages were noted towards the end of the four-week source service cycles, especially towards the end of the approximately 20-week run [47]. This was caused by the decreasing plasma impurities, which increased the breakdown voltage of the hydrogen gas. Using a compromise tune and increasing the inductance reduced the problem, but did not eliminate it [47].

Measuring the capacitive shift of the RF resonance with and without plasma allowed the estimation of the plasma inductance at ~0.15 μH, a significant fraction of the coil inductance of ~0.5 μH, but only ~4% of the total inductance. The problem was finally mitigated by starting each pulse with 1.96 MHz for ~5 μs before switching to 2 MHz, where the source is matched for maximum H^- beam current [48]. This reduced the ~30 μs H^- beam rise, seen in Fig. 9(a), to ~10 μs, seen in Fig. 9(b).

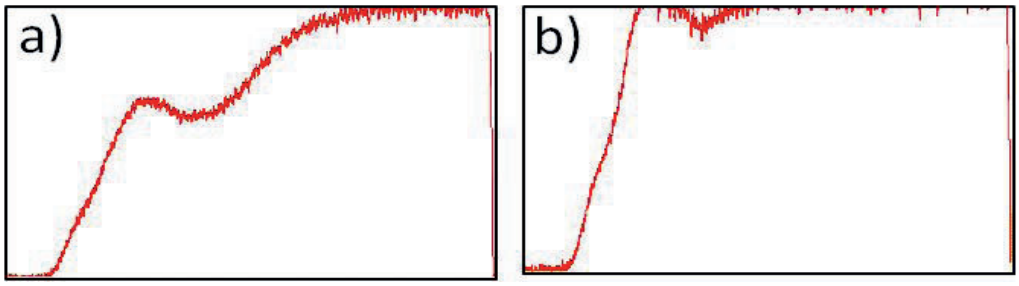


Fig. 9: The H^- beam rise time (a) for 2 MHz and (b) for a 1.96 MHz start followed by 2 MHz

Developing a 70 kV, 2 MHz transformer [49] allowed the 2 MHz amplifier to be moved from the 65 kV high-voltage platform enclosure to ground, which significantly reduced the RF amplifier problems [47].

The source is kept at a potential of −65 kV, which generates the extraction field. An 8 kV supply located on the −65 kV platform generates the −6.2 kV required to intercept most of the co-extracted electrons [46], which are driven to the side by the dipole dumping magnet.

There was also a significant learning curve for the LEBT [45], but its availability is very high since the summer of 2010. While models predict no significant losses, recently implemented thermocouples in the centre ground of the LEBT show significant heat generated by the beam as well as by occasional discharges.

8 RF antennas for high-current, high-duty-factor negative ion sources

High H^- beam currents require high radio-frequency (RF) power, with several hundreds of amperes flowing through the antenna. Hundreds of amperes induce oscillating voltages of the order of a kilovolt, which can cause sputtering of the antenna surface and large fluctuations in the plasma potential. Increasing the RF power increases the plasma density. This can cause an increasing fraction of the current to pass through the plasma, bypassing the ~0.5 μH antenna and leading to a saturation of the plasma density. Electrical insulation and water cooling are required for the high plasma densities needed to produce large H^- beam currents. This was initially achieved with a glass/enamel-coated Cu antenna [19].

In 1994 DESY acquired a copy of the LBNL RF source and spare antennas to test its suitability as an injector for the Hadron–Electron Ring Accelerator (HERA). All antennas were coated by the Porcelain Patch and Glaze Corporation (P&G) with a single layer of porcelain [50]. Unfortunately, each antenna developed one or several holes in apparently random locations, compromising the H^- beam output. Glassy deposits needed to be cleaned from the source before operation could restart with a new antenna. Of the 21 antennas, four failed during the first day of operation and another six failed

over the next 14 days. One antenna survived for 167 days, as seen in Fig. 10(a) (from [22]). Converted to daily failure probabilities, after 19% failed on the first day, the probability dropped to ~3.3% for the next 20 days, and then to ~1.2% for the next 146 days. Most remarkable is the lack of old-age failures. This suggests that the sputtering of the bulk antenna coating was not the root cause of the failures.

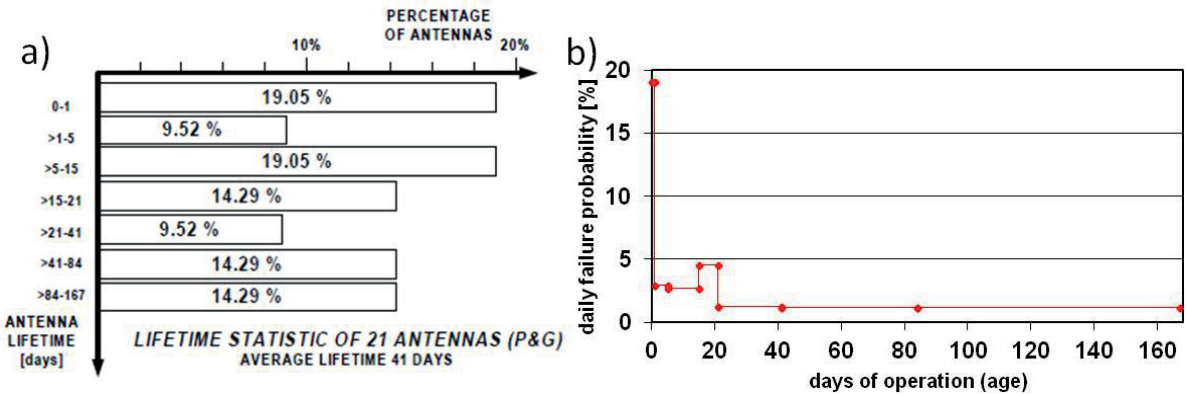


Fig. 10: Antenna lifetimes (a) as reported by DESY and (b) plotted as daily failure probability ((a) from [22])

When the construction of SNS started in 1999, multicusp ion sources with internal antennas were in use or being tested at several laboratories throughout the world. Reports of antenna failures raised concern over whether the H^- source developed for SNS could meet the SNS lifetime requirement. Naturally, each laboratory tested the antennas with their specific requirements, which raised the question of how such lifetimes should be scaled to the SNS requirements.

- If the limiting factor was the heating of certain parts, the average RF power would be the appropriate scaling factor. This meant that the 50 h observed at the Paul Scherrer Institute (PSI), Switzerland, with continuous 6–8 kW, 2 MHz discharges disqualified the P&G antennas [51]. However, the 250 h achieved at PSI [51] with a quartz antenna [52] was promising. The lifetime of more than 500 h of a Ti tube inside a quartz tube driven by a continuous 13 MHz, 2 kW discharge [53] could meet the SNS requirements if the required 50 mA H^- current could be produced with 30 kW or less.
- However, if the limiting factor was related to turn-on issues of the high-power RF, the lifetime had to be scaled with the repetition rate. Only DESY tested the antennas with pulsed high-power RF, and if 45 kW could produce the required 50 mA, scaling with the repetition rate reduces the observed 41-day average lifetime to 0.7 days at 60 Hz.
- Finally, if the limiting factor was related to the high-power RF, the lifetime should be scaled with the duty factor, which would reduce the observed DESY lifetime to 0.2 days at 60 Hz and 1 ms.

The DESY results clearly questioned the chances of the P&G antennas meeting the SNS 1 MW requirements. It was found that the single layer of porcelain was only ~0.1–0.2 mm thick, which can only withstand 1–2 kV. This is marginal compared to the voltages generated in the coil at high power. A microscopic cross-section revealed highly variable porosity, which is common in air-sprayed, air-fired porcelain. A chemical analysis showed a significant amount of TiO_2, which is a dielectric and reduces the voltage drop inside the porcelain [54].

The antenna lifetimes were extended by Cherokee Porcelain [55], coating Cu antennas with four or five layers of TiO_2-free porcelain, which yielded a ~0.7 mm thick, low-dielectric insulation. Operational experience showed practically no failures at low power and low duty factor. Antenna failures started above 3% duty factor and yielded approximately one failure in about seven source service cycles of each ~20-week run. After the duty factor during conditioning was increased to 7%, three antennas failed within hours, after which the conditioning duty factor was reduced to the production duty factor. When the production duty factor was raised to 5.4% and the RF power was

raised to ~60 kW, the antenna failures increased to two or three failures in about seven or eight source service cycles of each ~20-week run [56].

One antenna failed after 22 days; all other failures occurred in the first 11 days, and half of the failures occurred in the first six days, despite increasing the source service cycle up to six weeks [56]. This is consistent with infant mortality and the absence of old-age failure. This is also consistent with plasma emission spectra showing that Na sputtered from the antenna only during the first day when the plasma contains heavy impurities, initially water and later Cs [2]. It is also consistent with the removed antennas being blackened by carbon, most likely emitted from the stainless heat shield. The carbon becomes a sputter-resistant antenna coating after the impurities disappear from the plasma.

A more recent analysis revealed that most failures occur at the end of the coil where the antenna tube is exposed to the plasma. An improved geometry is being developed [48]. More recently, antenna failures have been significantly reduced by excluding all antennas with visible or tangible surface imperfection from production runs. In addition, efforts are under way to improve the cleanliness of the coating process [57].

Why not simply adopt DESY's external antenna scheme? The reason is because SNS's 67 times larger duty factor puts a significant heat load on the plasma chamber. SNS started to develop H^- sources with external antennas in 2003 [58]. Several designs, some of which included a Faraday shield, did not bring the anticipated results. After two Al_2O_3 chambers failed, as expected from model calculations, the effort adopted aluminium nitride (AlN) because of its much higher thermal conductivity. Early in 2009 the AlN H^- source was approved for neutron production. However, within about five weeks, five AlN source failures forced source changes, and the internal antenna source had to be reinstated to restore acceptable availability. Another serious issue is continuous beam loss during the entire one- to two-week service cycles, which is now suspected to be caused by emissions from AlN [57].

9 Caesium and the management of Cs_2CrO_4 cartridges

Its 260 pm atomic radius makes caesium the largest naturally occurring atom, as seen in Fig. 11 [59]. When adsorbed on the surface of another metal, its outermost electrons mix with the conduction electrons, forming strong, ionic-like bonds with the surface. Accordingly, as an adsorbate, the ionic radius of 181 pm becomes more relevant and is shown as a dashed line in Fig. 11. The line shows that there is a significant mismatch between the Cs ionic radius and the atomic radius of metals typically used in ion sources.

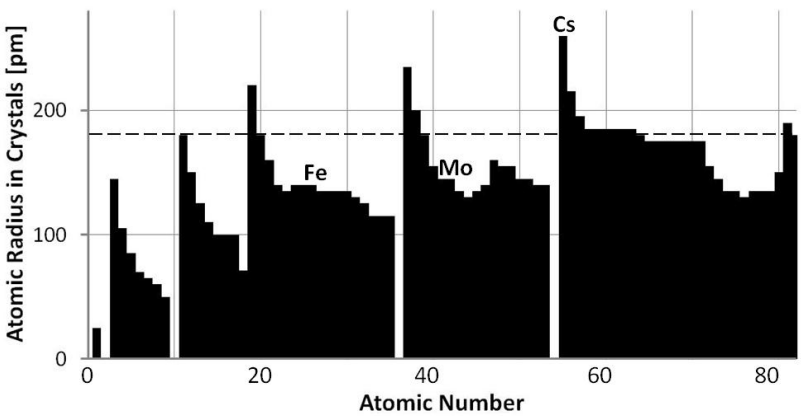


Fig. 11: Atomic radii versus atomic number. The dashed line shows the ionic radius of Cs

At low doses Cs adsorbs at the most attractive locations of the metal substrate, but the radius mismatch forces additional Cs to adsorb at less and less attractive locations. This causes the surface bond energy to decrease with increasing Cs coverage. This can be seen in Fig. 12, which shows functions fitted to data for Cs on a polycrystalline Mo surface [60] and on a 110 Mo surface [61].

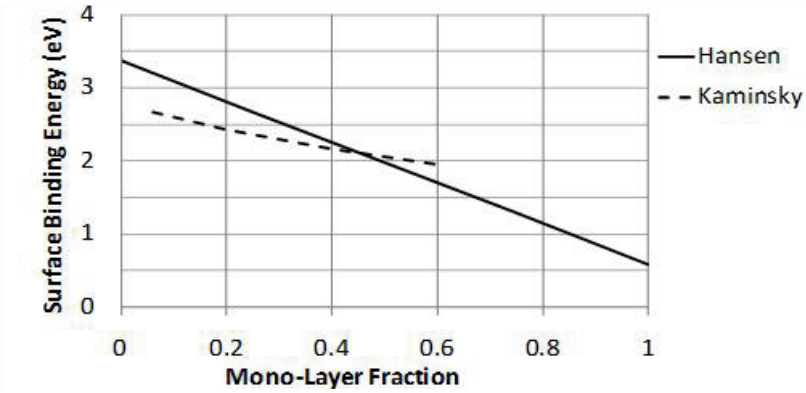


Fig. 12: Surface binding energy for Cs on polycrystalline Mo (Kaminsky [60]) and 110 Mo (Hansen [61]) versus the Cs coverage.

Cs atoms that adsorb on top of a Cs layer have small binding energies, and therefore are very rapidly emitted at room temperature. The decreasing binding energy of the first monolayer has an interesting effect on its thermal emission. For constant bond energies, thermal emission emits a certain fraction of the remaining adsorbates, causing an exponential decay of the adsorbate population. However, because the bond energy increases with a decreasing population, the bond energy practically stabilizes the population at a certain fraction, which depends on the temperature. This is not quite obvious from Fig. 13, which shows the expected fractions for different temperatures versus the time on a logarithmic scale calculated with the Hansen approximation [62].

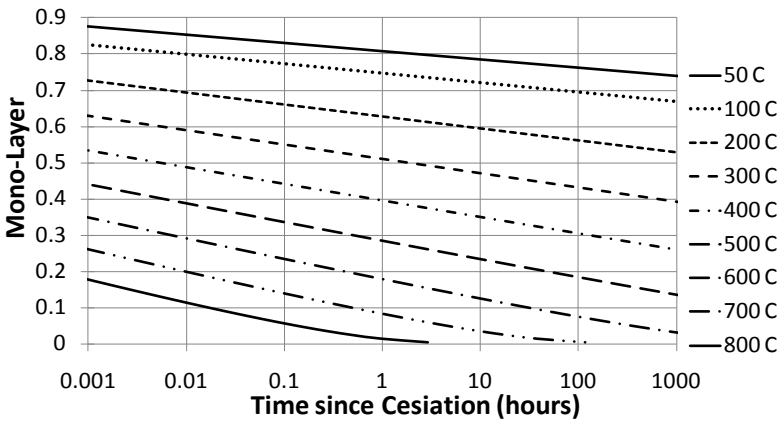


Fig. 13: The Cs surface coverage versus time for different surface temperatures

However, using linear time scales, Fig. 14 [56] shows that significant changes only happen at the beginning and then become too small to be noticed: looking at a scale of 5 h in Fig. 14(a), the significant change happens in the first half hour, and looking on a scale of 35 days in Fig. 14(b), the significant change occurs in the first few days. This means that one can use a constant temperature to approximately control the Cs population on a metal surface. This explains why successful caesium-enhanced volume H^- sources control the temperature of the surfaces near the outlet. The temperature is adjusted to obtain the fractional monolayer that yields the most H^- beam.

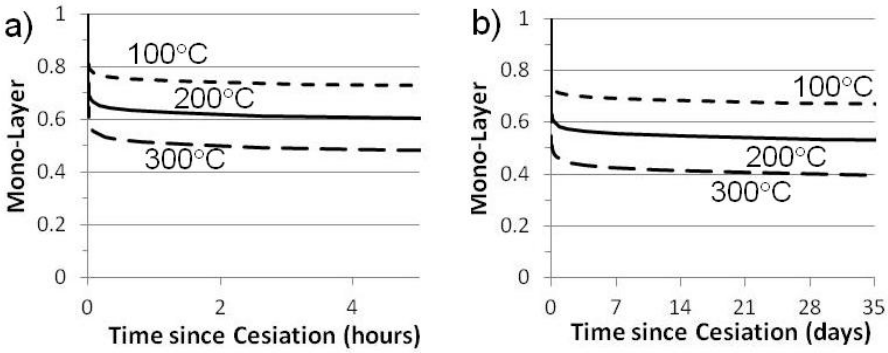


Fig. 14: Same as Fig. 13, but linear time scales of (a) 5 h and (b) 35 days

Cs has an ionization energy of 3.9 eV and an electron affinity of 0.47 eV, and therefore remains mostly neutral in the cold plasma near the outlet. Accordingly, some of the Cs escapes the source as neutral atoms and adsorbs on surfaces that are in line of sight with the outlet. For the SNS H^- source, these surfaces include the extractor, the first lens and the ground electrode. The adsorbed Cs lowers the work function, which leads to electron emission and discharges from the negatively charged first lens, which can make the LEBT temporarily inoperable. Lowering the use of Cs has significantly improved the performance of the first lens [62].

Figure 15(a) shows the cross-section of a Cs_2CrO_4 cartridge [63]. Eight such cartridges, containing slightly less than 30 mg of Cs, are inserted into the Cs collar, shown in Fig. 15(b). The figure also shows the water-cooled filter magnets, the 7 mm diameter source outlet, and the Mo converter, which was introduced in 2007 to boost the H^- current output [45].

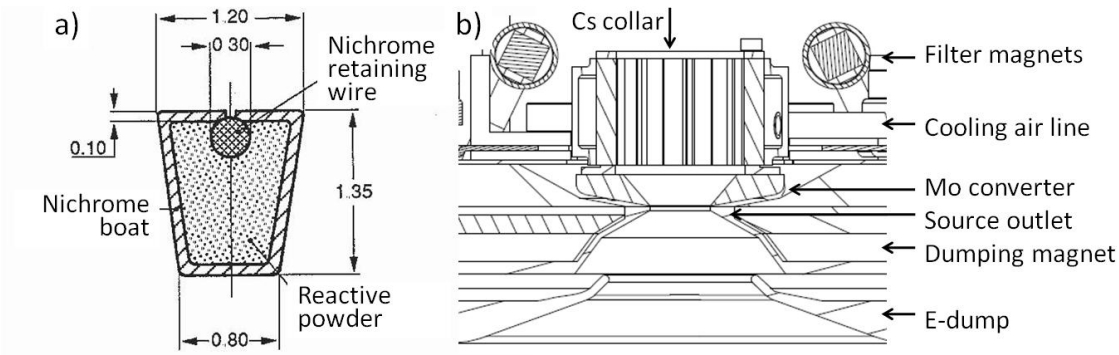


Fig. 15: (a) Cs_2CrO_4 cartridge cross-section and (b) the cartridge holding Cs collar with the SNS converter

The mixture of Cs_2CrO_4 and St101 (a getter made of 16% Al and 84% Zr) is a reactive, fine powder having a very large surface area covered with adsorbates. SNS found that the cartridges can be fully degassed without emitting large amounts of Cs by heating them for 3 h to ~250°C. This is confirmed by the residual gas mass analyser (RGA), which shows the same partial pressures before and after caesiation.

Without sufficient degassing, the St101 will getter the surface sorbates before starting to reduce the Cs_2CrO_4, and so release less, and sometimes insufficient, Cs. Insufficient degassing is confirmed by the lower partial pressures of masses 18, 28 and 44 after caesiation, because the caesiation stopped the outgassing from the cartridges.

For Cs to form strong, ionic bonds with the converter, the Mo converter must also be cleaned of sorbates. SNS found that 3 h of 50 kW plasma at a 5.3% duty factor sputter-cleans the Mo converter, and accordingly it coincides with the conditioning of the cartridges and the high-voltage conditioning of the two electrostatic lenses.

In summary, the SNS source is installed and evacuated while being leak-checked. When complete, the protective cages are closed, the collar is heated to ~250°C, and the 13 MHz plasma is started, followed by the 2 MHz. After 3 h of conditioning and ramping up the lens voltages, the collar temperature is raised to 550°C. After 12 min the temperature is lowered to ~180°C to retain a near-optimal fractional Cs layer. The extracted H^- beam typically grows for the first few days as the Cs fraction decreases from above optimal towards the optimal value.

10 Persistent H^- beams and the plasma potential

Despite short pulses, low repetition rates and a cold air-cooled Cs collar, the H^- beam decayed in the first neutron production runs, with the first run shown in Fig. 16. However, the beam could be restored by raising the temperature of the Cs collar for a certain time period.

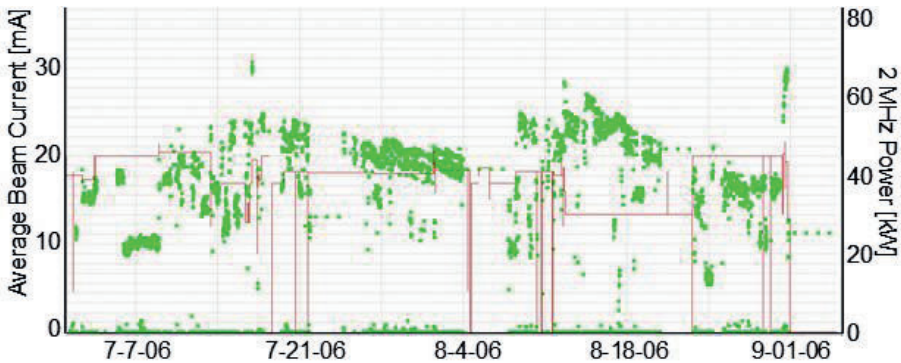


Fig. 16: Decaying H^- beam during the first neutron production run

Caesiations yielded unpredictable results until the need for degassing the cartridges was recognized in 2007. Even then, the H^- beam decayed during the first night, but became persistent after recaesiating the following morning [56]. Only in 2008, after the plasma conditioning was extended to 3 h, did the H^- beams become persistent after the first caesiation. In 2009, to meet the growing H^- current requirement, the Cs collar temperature was raised without noticing beam loss during the three-week and later four-week source service cycles. This lead to the calculations presented in the last section [62]. After a contamination event in 2011, the source service cycle of the uncontaminated source was extended to six weeks without observing decay in the H^- beam [2].

However, a thermal air leak during the summer of 2011 caused the H^- beam to decay by ~1% per hour. The beam was restored with recaesiations, after which it decayed again with the same rate [56]. This is consistent with the sputtering of the Cs by the heavy air ions and its subsequent loss from the plasma. Fig. 17 shows the ratio of the Cs sputter threshold to the Cs surface bond energy versus the square root of the mass of the sputtering ion, using the Bohdansky approximation [64]. The figure indicates that water and CO ions have the most effective masses for sputtering Cs, while N^+ and NH_3^+ need just a little more energy to sputter Cs. On the other hand, hydrogen ions are much too light to sputter Cs effectively.

Assuming surface bond energies of 2 eV for Cs on Mo [60], the fact that pure hydrogen plasma fails to sputter Cs suggests that the plasma potential is less than 25 V, whereas the fact that nitrogen ions sputter Cs suggests that the plasma potential exceeds 8 eV.

RGA data from the leaky source suggested roughly a 10^{-6} cm^3 s^{-1} air leak. This is consistent with a plasma impurity of ~1 ppm, which would suggest that the SNS hydrogen plasma has normally less than 10^{-8} impurities when it provides persistent beams for six weeks [2]. At this time it is unclear how that can be reconciled with the use of ultra-high-purity hydrogen (<10 ppm impurities) without further purification.

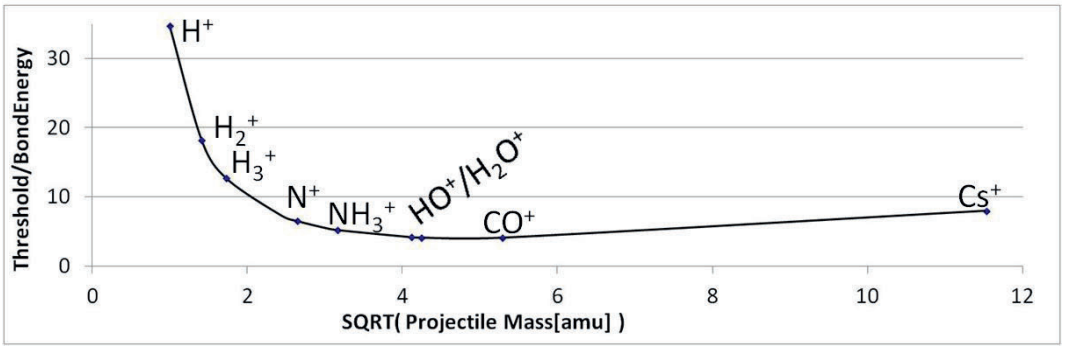


Fig. 17: Cs sputter threshold to surface bond energy ratio for various ions

11 The caesium consumption

Many papers have pointed out that H is too light for sputtering Cs. This is not exactly true because sputtering depends on the energy of the hydrogen particles. On the other hand, it is well known that Cs-enhanced H^- sources require a small flux of Cs to maintain a constant output of the H^- beam. SNS appears to have the first caesiated H^- source that is able to run at high duty factor for long time periods without a continuous supply of Cs [46].

The Cs consumption is difficult to determine because it depends on the design and even more on the operation of the Cs system. Starting up a new system, Cs should be generously dispensed to be sure to reach the optimal Cs distribution within a reasonable time period, normally judged by the H^- beam current output. However, once the H^- output peaks, the Cs dispense rate should be drastically reduced for sustainable operations. The details depend on the temperatures of certain surfaces. There is a long learning curve of adjustments to maintain that maximum H^- output current by replacing the Cs that is lost and not more. After sustainability is established, the learning curve begins on how to reach the optimal caesiation in a minimum of time with a minimum of Cs.

An excellent example is the HERA magnetron source, which used 6 mg of Cs when it was initially started in 1983. In its last run of 2008, that amount was reduced to 0.7 mg per day [65].

Another example is the multicusp, filament-driven LANSCE H^- source, where the outlet-facing Mo converter features −300 V to boost the extraction of the converter-surface-produced H^- ions. This voltage is sufficient to empower hydrogen ions to sputter Cs, which could contribute to the consumption of 0.7 g of Cs per day operating at a 5% duty factor [66]. This exceeds the HERA consumption 1000-fold. As discussed, thermal emission initially contributes to Cs losses, but if properly adjusted, those losses are marginalized by self-stabilization within a few days. Eventually Cs losses are dominated by sputter losses, and therefore only occur when plasma is present. Accordingly, the Cs consumption should be normalized to the plasma duty factor (5%), which yields for the LANSCE source ~14 g per plasma-day (daily consumption divided by the plasma duty factor), where a plasma-day is the length of time needed to accumulate a full day of plasma operation.

Normalizing the HERA source with its 0.0075 % duty factor yields ~10 g per plasma-day [65], not very different from the LANSCE source.

After heating the SNS cartridges three times for 30 min to 550°C, an analysis showed the Cs to be fully depleted. That suggests that ~4 mg Cs is released during the first 12 min. Producing H^- beam for up to six weeks suggests a daily consumption of <0.1 mg per day. Normalizing with the 5.3% duty factor yields <1.8 mg per plasma-day, almost four orders of magnitude less than the other sources.

12 The co-extracted electrons, a real challenge for high-power H^- sources

Negative ion sources are placed on negative high-voltage platforms, which generate the field to accelerate the escaping negative ions towards the near-ground extractor. Naturally this field also accelerates all electrons that venture to the meniscus, forming an electron beam. While this is a small inconvenience for low-current and/or low-duty-factor H^- sources [22], it becomes a challenging problem at high power and high duty factor [67] because there are normally many more electrons than H^-. Co-extracted electron currents exceeding the H^- current by a factor between 100 and 200 have been reported [18], although the introduction of the outlet collar and the use of NdFe filter magnets reduced this ratio to ~50 [19]. Further optimizations can lower this value by another factor of 2 [21, 68, 69].

While Cs normally increases the H^- beam current, it also decreases the electron current by a factor of two [40], which brings the electron to H^- ratio typically into the single-digit range [68]. When extracting ~50 mA H^- beam, the SNS source co-delivers only about 20 mA of electrons, which is surprisingly small.

Accelerating amps of electron currents from uncaesiated sources or fractions of amps from caesiated H^- sources would generate destructive electron beams with kilowatts of power, while posing a destabilizing excessive burden on the extraction supply. The common solution is to stop the electron beam at an intermediate electrode, powered by a high-current, reduced-voltage supply located on the high-voltage platform. This decouples the energy of the H^- beam from fluctuations of the electron beam, which can be a serious problem for pulsed H^- sources.

Separating the electron beam from the H^- beam requires a magnetic field, which will also deflect the H^- beam by a small angle of 2 to 3°. In the SNS source, this angle is compensated by tilting the source by a few degrees. As shown in Figs. 4(b) and 6(b), other sources typically use two opposing dipole fields to replace the angular deflection with small doglegs in the ion trajectories.

For smaller source voltages, the e-dump can be placed beyond the grounded extractor [26], although energy-reducing voltages will focus the ion beam. The SNS source (see Fig. 8(a)), the J-PARC source (see Fig. 6(b)) and the TRIUMF source (see Fig. 4(b)) all place the e-dump between the source and the extractor, which minimizes focusing. However, this couples the energy of the electron beam and the extraction field near the outlet, which is not ideal.

For high-current, high-duty-factor H^- sources, it is essential to properly model a robust dumping of the electron beam, which will determine the voltage of the e-dump. Placing an extraction or 'puller' electrode between the source outlet and the e-dump allows the control of the extraction field while dumping the electrons at a safe voltage [70].

Acknowledgements

This manuscript was prepared at Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC, under contract DE-AC05-00OR22725 for the US Department of Energy. The author appreciates the invaluable comments from M. Bacal, J. Galambos, J. Holmes and H. Oguri.

References

[1] S. Henderson, Spallation Neutron Source operation at 1 MW and beyond, in Proc. 25th Linear Accelerator Conf., LINAC10, Tsukuba, Japan (2010), MO103, pp. 11–15.

[2] M.P. Stockli, B.X. Han, S.N. Murray, T.R. Pennisi, M. Santana, and R.F. Welton, *AIP Conf. Proc.* **1515** (2013) 292–301.

[3] J. Peters, The HERA volume H source, in Proc. Particle Accelerator Conf. 2005, PAC 2005 (IEEE, New York, 2005), pp. 788–790.

[4] H. Oguri, K. Ikegami, K. Ohkoshi, I. Koizumi, Y. Namekawa, A. Ueno, A. Takagi, and S. Yamazaki, *AIP Conf. Proc.* **1515** (2013) 379–385.

[5] R.F. Welton, Overview of high-brightness H^- ion sources, in Proc. 21st Linear Accelerator Conf., LINAC 2002, Gyeongju, Korea (2002), TH101, pp. 559–563.

[6] M.P. Stockli, R.F. Welton, M. Janney, R. Lauf, R. Gough, R. Keller, K.N. Leung, J. Reijonen, T. Schenkel, J. Staples, and R. Thomae, *Rev. Sci. Instrum.* **73** (2002) 1007.

[7] L. Schroeder, K.N. Leung, and J. Alonso (Eds.), Proc. Workshop on Ion Source Issues Relevant to a Pulsed Neutron Source, Berkeley, CA, 24–26 October 1994, LBNL Report 36347.

[8] J.R. Alonso, *Rev. Sci. Instrum.* **67** (1996) 1308–1313.

[9] K. Saadatmand, G. Arbique, J. Hebert, R. Valicenti, and K.N. Leung, *Rev. Sci. Instrum.* **67** (1996) 1318–1320.

[10] M.A. Leitner, R.A. Gough, K.N. Leung, M.L. Rickard, P.K. Scott, A.B. Wengrow, M.D. Williams, and D.C. Wutte, *Rev. Sci. Instrum.* **69** (1998) 962–964.

[11] R.K. Janev, D. Reiter, and U. Samm, Collision Processes in Low-Temperature Hydrogen Plasmas, Jülich Report 4105 (2003).

[12] H. Zhang, *Ion Sources* (Springer, Berlin / Science Press, Beijing, 1999).

[13] J.R. Hiskes, *J. Appl. Phys.* **51** (1980) 4592–4594.

[14] M. Bacal, A. Hatayama, and J. Peters, *IEEE Trans. Plasma Sci.* **33** (2005) 1845–1871.

[15] M. Bacal, *Nucl. Fusion* **46** (2006) S250–S259.

[16] M. Bacal, A.M. Bruneteau, H.J. Doucet, W.G. Graham, and G.W. Hamilton, in Proc. 2nd Int. Symp. on the Production and Neutralization of Negative Ions and Beams, Brookhaven National Laboratory, New York, 1980, pp. 126–133.

[17] M. Bacal and G.W. Hamilton, *Phys. Rev. Lett.* **42** (1979) 1538–1540.

[18] K.N. Leung, K.W. Ehlers, C.A. Hauck, W.B. Kunkel, and A.F. Lietzke, *Rev. Sci. Instrum.* **59** (1988) 453–456.

[19] K.N. Leung, G.J. DeVries, W.F. DiVergilio, R.W. Hamm, C.A. Hauck, W.B. Kunkel, D.S. McDonald, and M.D. Williams, *Rev. Sci. Instrum.* **62** (1991) 100–104.

[20] K. Saadatmand, J.E. Hebert, and N.C. Okay, *AIP Conf. Proc.* **287** (1992) 448–457.

[21] K. Saadatmand, J.E. Hebert, and N.C. Okay, *Rev. Sci. Instrum.* **65** (1994) 1173–1175.

[22] J. Peters, *Rev. Sci. Instrum.* **71** (2000) 1069–1074.

[23] J. Peters, *Rev. Sci. Instrum.* **69** (1998) 992–994.

[24] J. Peters, *Rev. Sci. Instrum.* **65** (1994) 1237–1239.

[25] J. Peters, *Rev. Sci. Instrum.* **79** (2008) 02A515.

[26] J. Peters, The HERA RF-driven multicusp H^- ion source, in Proc. Linear Accelerator Conf., LINAC 2006, Knoxville, TN (2006), TUP061, pp. 388–390.

[27] K.R. Kendall, M. McDonald, D.R. Mosscrop, P.W. Schmor, D. Yuan, G. Dammertz, B. Piosczyk, and M. Olivo, *Rev. Sci. Instrum.* **57** (1986) 1277–1281.

[28] T. Kuo, D. Yuan, K. Jayamanna, M. McDonald, R. Baartman, P. Schmor, and G. Dutto, *Rev. Sci. Instrum.* **67** (1996) 1314–1316.

[29] K. Jayamanna, M. McDonald, D.H. Yuan, and P.W. Schmor, The TRIUMF compact DC H^-/D^- ion source, in Proc. European Particle Accelerator Conf., EPAC 90, Nice (1990), pp. 647–649.

[30] D-Pace at http://www.d-pace.com/products_hion.html.

[31] B. Rasser, J.N.M. Van Wunnik, and J. Los, *Surf. Sci.* **118** (1982) 697–710.

[32] W.G. Graham, in Proc. 2nd Int. Symp. on the Production and Neutralization of Negative Ions and Beams, Brookhaven National Laboratory, New York, 1980, pp. 126–133.

[33] J.N.M. Wunnik, B. Rasser, and J. Los, *Phys. Lett.* **87A** (1982) 288–290.

[34] T. Okuyama and Y. Mori, *Rev. Sci. Instrum.* **63** (1992) 2711–2713.

[35] M. Ogasawara, R. Sakurai, T. Koishimine, S. Mitsuhashi, and A. Hatayama, *Fusion Eng. Design* **26** (1995) 507–521.

[36] H. Oguri, Y. Okumura, K. Hasegawa, Y. Namekawa, and T. Shimooka, *Rev. Sci. Instrum.* **73** (2002) 1021–1023.

[37] A. Ueno, H. Oguri, K. Ikegami, Y. Namekawa, and K. Ohkoshi, *Rev. Sci. Instrum.* **81** (2010) 02A720.

[38] H. Oguri, A. Ueno, K. Ikegami, Y. Namekawa, and K. Ohkoshi, *Rev. Sci. Instrum.* **81** (2010) 02A715.

[39] H. Oguri, K. Ikegami, K. Ohkoshi, Y. Namekawa, and A. Ueno, *AIP Conf. Proc.* **1390** (2011) 235–244.

[40] K.N. Leung, D.A. Bachman, and D.S. McDonald, *Rev. Sci. Instrum.* **64** (1993) 970–973.

[41] M.A. Leitner, D.W. Cheng, R.A. Gough, R. Keller, K.N. Leung, S.K. Mukherjee, P.K. Scott, and M.D. Williams, High-current, high-duty-factor experiments with the H^- ion source for the Spallation Neutron Source, in Proc. 1999 Particle Accelerator Conf., New York (1999), WEA13, pp. 1911–1913.

[42] T. Schenkel, J.W. Staples, R.W. Thomae, J. Reijonen, R.A. Gough, K.N. Leung, and R. Keller, *Rev. Sci. Instrum.* **73** (2002) 1017–1019.

[43] J.W. Staples, The NSNS front end accelerator system, in Proc. 1997 Particle Accelerator Conf., Vancouver, Canada (1997), pp. 1174–1176.

[44] R. Keller, D. Cheng, R. DiGennaro, R.A. Gough, J. Greer, K.N. Leung, A. Ratti, J. Reijonen, R.W. Thomae, T. Schenkel, J.W. Staples, R. Yourd, A. Aleksandrov, M.P. Stockli, and R.W. Welton, *Rev. Sci. Instrum.* **73** (2002) 914–916.

[45] M.P. Stockli, B.X. Han, S.N. Murray, D. Newland, T.R. Pennisi, M. Santana, and R.F. Welton, *AIP Conf. Proc.* **CP1097** (2009) 223–235.

[46] M.P. Stockli, B. Han, S.N. Murray, T.R. Pennisi, M. Santana, and R.F. Welton, *Rev. Sci. Instrum.* **81** (2010) 02A729.

[47] B. Han, T. Hardek, Y. Kang, S.N. Murray Jr., T.R. Pennisi, C. Piller, M. Santana, R.F. Welton, and M.P. Stockli, *AIP Conf. Proc.* **CP1390** (2011) 216–225.

[48] M.P. Stockli, B.X. Han, T.W. Hardek, Y.W. Kang, S.N. Murray, T.R. Pennisi, C. Piller, M. Santana, and R. Welton, Highly-persistent SNS H^- source fueling 1-MW beams with 7–9 kC service cycles, in Proc. 2011 Particle Accelerator Conf., New York (2011), pp. 1993–1995.

[49] Y.W. Kang, R. Fuja, T. Hardek, S.W. Lee, M.P. McCarthy, M.C. Piller, K. Shin, M.P. Stockli, A. Vassioutchenko, and R.W. Welton, Developments for performance improvement of SNS H^- ion source RF systems, in Proc. 25th Linear Accelerator Conf., LINAC10, Tsukuba, Japan (2010), pp. 1019–1021.

[50] Porcelain Patch and Glaze Corporation, Oakland, CA, USA.

[51] H. Einenkel, Paul Scherrer Institute, Switzerland, private communication (2001).

[52] Y. Lee, R.A. Gough, K.N. Leung, L.T. Perkins, D.S. Pickard, J. Vujic, L.K. Wu, M. Olivo, and H. Einenkel, *Rev. Sci. Instrum.* **69** (1998) 1023–1025.

[53] J. Reijonen, M. Eardley, R. Gough, R. Keller, K. Leung, R. Thomae, D. Pickard, and M.D. Williams, *Rev. Sci. Instrum.* **71** (2000) 1134–1136.

[54] R.F. Welton, M.P. Stockli, Y. Kang, M. Janney, R. Keller, R.W. Thomae, T. Schenkel, and S. Shukla, *Rev. Sci. Instrum.* **73** (2002) 1008–1012.

[55] Cherokee Porcelain Enamel Corporation, 2717 Independence Lane, Knoxville, TN37914, USA.

[56] M.P. Stockli, B. Han, T.W. Hardek, Y.W. Kang, S.N. Murray, T.R. Pennisi, C. Piller, M. Santana, and R.F. Welton, *Rev. Sci. Instrum.* **83** (2012) 02A732.

[57] R.F. Welton, V.G. Dudnikov, K.R. Gawne, B.X. Han, and S.N. Murray, *Rev. Sci. Instrum.* **83** (2012) 02A725.

[58] R.F. Welton, M.P. Stockli, R.T. Roseberry, Y. Kang, and R. Keller, *Rev. Sci. Instrum.* **75** (2004) 1789–1792.

[59] J.C. Slater, *J. Chem. Phys.* **41** (1964) 3199.

[60] M. Kaminsky, *Atomic and Ionic Impact Phenomena on Metal Surfaces* (Springer, New York, 1965).

[61] L.K. Hansen, Thermionic Converters and Low Temperature Plasma, Information Center/US-DOE-tr-1 (1978).

[62] M.P. Stockli, B.X. Han, S.N. Murray, T.R. Pennisi, M. Santana, and R.F. Welton, *AIP Conf. Proc.* **CP1390** (2011) 123–133.

[63] CS/NF/3.6/11 from SAES Getters S.p.A., Via Gallarate 215, 20151 Milano, Italy.

[64] J. Bohdansky and J. Roth, *J. Appl. Phys.* **51** (1980) 2861–2865.

[65] J. Peters, *AIP Conf. Proc.* **CP1097** (2009) 236–242.

[66] R. Keller, O. Tarvainen, E. Chacon-Golcher, E.G. Geros, K.F. Johnson, G. Rouleau, J.E. Stelzer, and T.J. Zaugg, *AIP Conf. Proc.* **CP1097** (2009) 161–170.

[67] O. Midttun, T. Kalvas, M. Kronberger, J. Lettry, H. Pereira, C. Schmitzer, and R. Scrivens, *Rev. Sci. Instrum.* **83** (2012) 02B710.

[68] D.P. Moehs, J. Peters, and J. Sherman, *IEEE Trans. Plasma Sci.* **33** (2005) 1786–1798.

[69] M. Kronberger, D. Kuchler, J. Lettry, O. Midttun, and M.O. Neil, *Rev. Sci. Instrum.* **81** (2010) 02A708.

[70] O. Midttun, T. Kalvas, M. Kronberger, J. Lettry, H. Pereira, and R. Scrivens, *AIP Conf. Proc.* **CP1515** (2013) 481–490.

Negative Ion Sources: Magnetron and Penning

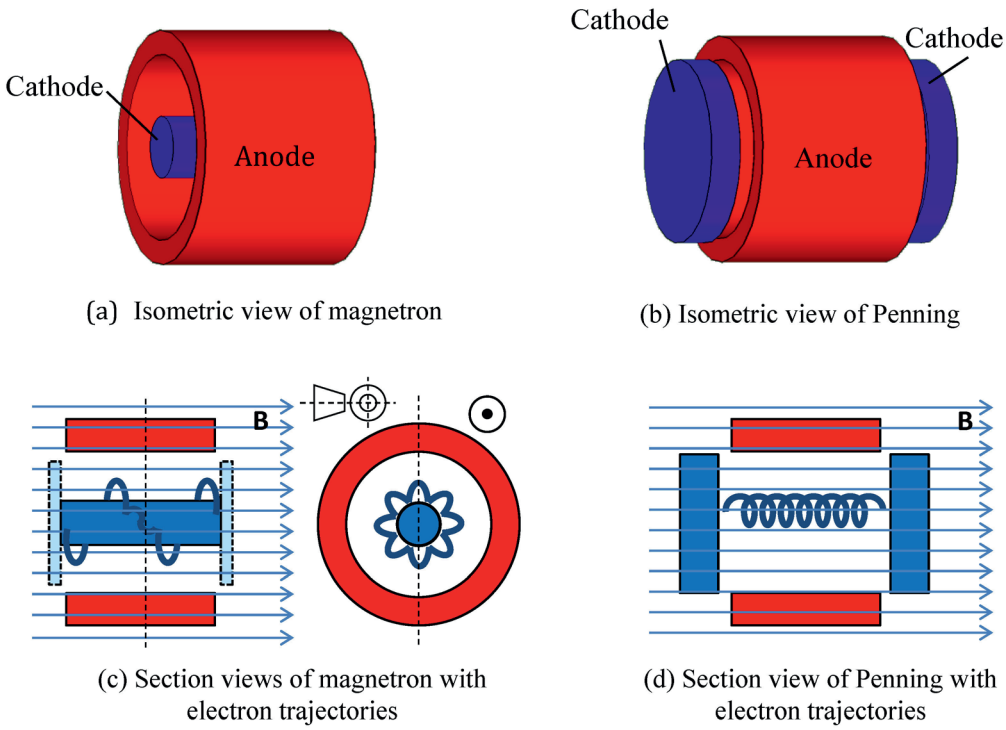
D.C. Faircloth
Rutherford Appleton Laboratory, Chilton, Oxfordshire, UK

Abstract
The history of the magnetron and Penning electrode geometry is briefly outlined. Plasma generation by electrical discharge-driven electron impact ionization is described and the basic physics of plasma and electrodes relevant to magnetron and Penning discharges are explained. Negative ions and their applications are introduced, along with their production mechanisms. Caesium and surface production of negative ions are detailed. Technical details of how to build magnetron and Penning surface plasma sources are given, along with examples of specific sources from around the world. Failure modes are listed and lifetimes compared.

1 Introduction

1.1 Overview

Magnetron and Penning ion sources have a long history going back to the early part of the 20th century. Their use for negative ion production really took off in 1970s Soviet Russia with the introduction of caesium. Caesiated magnetron and Penning negative ion sources are referred to as surface plasma sources. They are the brightest of all negative ion sources, with H^- beam current densities at extraction exceeding 1 A cm^{-2}.



(a) Isometric view of magnetron

(b) Isometric view of Penning

(c) Section views of magnetron with electron trajectories

(d) Section view of Penning with electron trajectories

Fig. 1: Different views of the basic magnetron and Penning electrode geometries

1.2 Basic electrode topology

The fundamental magnetron and Penning topologies are shown in Fig. 1. Topologically, they both consist of a tube-shaped anode with a magnetic field running along the axis of the tube. The magnetron has a cathode in the centre of the tube as shown in Fig. 1(a), whereas the Penning has two cathodes at either end of the tube as shown in Fig. 1(b).

Although Fig. 1 shows the fundamental topologies, actual geometries used in practice are often quite different from those shown in Fig. 1. Anode tubes can be very short, so, instead of a tube, a window frame could be a better description of the anode shape. Electrodes can be square rather than round, they can be stretched to be rectangular, and often include a range of extra features such as lips, ridges, notches, grooves and dimples. Magnetron cathodes are often spool-shaped as it is shown in Fig. 1(c).

Whatever the geometry details, the basic topology of a magnetic field running perpendicularly through a hole in an anode is fundamental to both magnetron and Penning. It is the cathode topology and the resultant motion of the electrons (see section 2.4) that differentiates between them.

1.3 Early history

1.3.1 Magnetron

The magnetron electrode topology was first invented, not as an ion source, but as an electron tube by Albert Hull working at the General Electric Research Laboratory, Schenectady, New York in 1916 [1]. He was trying to use magnetism to find alternatives to the patented electrostatic control of valve amplifiers. The tubes he invented had various names, such as comet valves, boomerang valves and ballistic valves, because of the electron trajectories in the tube (see Fig. 1(c)) – however, the name magnetron valve eventually stuck [2].

To reach higher and higher powers, Hull started to add various gases to his tubes. By 1921 an experimental 1000 kW water-cooled magnetron with an anode voltage of 20 kV had been developed [2]. At the time Hull was working for the influential Irvin Langmuir. Langmuir discussed Hull's results with his fellow New England researchers and suggested that the gaseous discharge magnetron could make an excellent source of ions for the early particle accelerators being developed to study nuclear processes. The magnetron ion source was initially developed by Lamar and Luhr in the early 1930s at the Massachusetts Institute of Technology [3]. In 1934, Stanley Van Voorhis and his team in Princeton were the first to report using a magnetron as a proton source for an accelerator [4].

The possibility of using the magnetron electrode geometry to generate high-power microwaves was first practically demonstrated in 1940 by two British physicists, Henry Boot and John Randall working at Birmingham University, UK. By adding circular caves or cavities into the inner walls of the anode tube, electrons can be made to circulate resonantly in the cavities at microwave frequencies. With the addition of an output coupler positioned in one of the anode cavities, it is possible to produce high-power microwaves. These types of devices are known as cavity magnetrons and are not the subject of this paper. Cavity magnetrons are in common usage in microwave ovens.

1.3.2 Penning

The Penning geometry was first reported as a positive ion source by Louis Maxwell working at the Franklin Institute in Philadelphia in 1930 [5]. However, the Penning electrode geometry gets its name from Frans Penning, a researcher working at Philips Physics Laboratory in Eindhoven, The Netherlands. In 1937 he developed the Penning ionization gauge or Philips ionization gauge (PIG) [6]. By measuring the discharge current in the electrode geometry, he was able to measure accurately and reliably very low pressures in gases, a technique that is still widely used today.

1.3.3 Variations

The success of the magnetron and Penning electrode geometries as ion sources spawned a series of variations: the magnetron source was modified to give the Freeman source; the Penning source developed into the Calutron, Bernas and Nielson sources. Within a decade it was possible to produce multi-milliamp beams of positive ions from almost all the elements, no matter if they were solids, liquids or gases.

2 Plasma generation by electrical discharge

2.1 Introduction

An electrical discharge in the magnetron and Penning electrode configurations generates a plasma from which the required ions can be extracted. A low-pressure (< 1 mbar) gas is fed into the space between the electrodes. The type of gas used depends on the type of ions required. The electrical discharge is created by applying an electric potential between the anode and cathode. The resultant electric field accelerates free electrons, which can ionize the gas molecules. This process is called electron impact ionization.

2.2 Electrical discharges

2.2.1 Townsend breakdown

Electron impact ionization occurs when the applied electric field is high enough to accelerate the electrons (within their mean free path) to the ionization energy of the gas. At this point the current between the electrodes rapidly increases. The electrons ionize the neutral atoms and molecules, producing more electrons. These additional electrons are accelerated to ionize even more atoms, producing even more free electrons in an avalanche breakdown process known as Townsend breakdown. This runaway process means that the voltage needed to sustain the discharge drops significantly as the discharge enters the glow regime.

2.2.2 Glow discharge

A glow discharge is so called because it emits a significant amount of light. Most of the photons that make up this light are produced when atoms that have had their orbital electrons excited by electron bombardment relax back to their ground states. Photons are produced in any event that needs to release energy, for example when ions recombine with the free electrons or when vibrationally excited molecules relax.

A glow discharge is self-sustaining because positive ions that are accelerated to the cathode impact on it, producing more electrons in a process called secondary emission. This is why the voltage at which a discharge goes out is lower than the voltage at which it initiates.

The current in a glow discharge increases with very little increase in discharge voltage. The plasma distributes itself around the cathode surface as the current increases. Eventually the current reaches a point where the cathode surface is completely covered with plasma and the only way to increase the current further is to increase the current density at the cathode. This causes the plasma voltage near the cathode to rise.

2.2.3 Arc discharge

The increased current density leads to cathode heating, and eventually the cathode surface reaches a temperature where it starts to thermionically emit electrons (see section 2.3.9) and the discharge moves into the arc regime with a negative current versus voltage characteristic.

Magnetron and Penning ion sources operate with discharges in the glow arc regime. Technically the discharge is not an arc because the cathode is kept cool enough so it does not thermionically emit many electrons. However, the addition of caesium (see section 4) causes significant electron emission, which also yields a negative current versus voltage characteristic. The discharge current in magnetron and Penning ion sources is often called the 'arc current'.

2.3 Plasma properties

2.3.1 Introduction

The properties of the plasma produced in both the magnetron and Penning discharge share a lot of similarities. This section describes some basic plasma and electrode properties that will be referred to in later sections of this chapter.

2.3.2 Density, n

The most basic parameter is the density of each of the constituents in the plasma. It is usually written as n with a subscript to represent the type of particle, and is expressed in number of particles per cubic metre. Some older papers give density in particles per cubic centimetre. Thus we have

n_e = density of electrons
n_i = density of ions
n_n = density of neutrals

2.3.3 Temperature, T

The temperature of the plasma is a measurement of how fast each of the particles is going, otherwise known as the particle's kinetic energy. Particle temperatures are usually expressed in electronvolts. The Boltzmann constant k_B makes 1 eV = $1.602 \times 10^{-19}/1.381 \times 10^{-23}$ = 11 600 K. Similarly we have

T_e = temperature of electrons
T_i = temperature of ions
T_n = temperature of neutrals

The temperatures of the electrons, ions and neutrals can be different.

2.3.4 Temperature distributions

Obviously not all electrons in plasma will have the same temperature, and the same is true for the ions of the same species. The temperatures T_e, T_i and T_n are merely averages (of absolute temperatures). If the plasma is in thermal equilibrium, then the distribution will be Maxwellian and obey Maxwell–Boltzmann statistics.

Often the plasma is in a magnetic field. The ions and electrons will spiral around the magnetic field lines and slowly move along them. Hence, the particle velocities (temperatures) will not be the same in all directions. The particle temperatures in different directions are defined as follows: $T_{i\parallel}$ is the ion temperature parallel to the magnetic field, and $T_{i\perp}$ is the ion temperature perpendicular to the magnetic field.

2.3.5 Quasi-neutrality

Plasma is generally charge neutral, so all of the charge states of all the ions add up to the same number as the number of electrons:

$$\sum q_i n_i = n_e \tag{1}$$

2.3.6 Percentage ionization

The percentage ionization is a measure of how ionized the gas is, i.e. what proportion of the atoms have actually been ionized:

$$\text{percentage ionization} = 100 \times \frac{n_\mathrm{i}}{n_\mathrm{i}+n_\mathrm{n}} \tag{2}$$

When the percentage is above 10%, the plasma is said to be highly ionized and the interactions that take place within are dominated by plasma physics. Less than 1% ionization and interactions with neutrals must be considered.

2.3.7 Collisions

Collisions between charged particles in a plasma are fundamentally different from collisions in a neutral gas. The ions in a plasma interact by the Coulomb force: they can be attracted or repelled from a great distance. As an ion moves in a plasma, its direction is gradually changed as it passes the electric fields of neighbouring particles; whereas, in a neutral gas, the particles only interact when they get so close to each other that they literally bounce off each other's outer electron orbitals. In a neutral gas, the average distance the particles travel in a straight line before bouncing off another particle is referred to as the mean free path. In a plasma, the mean free path concept does not work because the ions and electrons are always interacting with each other by their electric fields. Instead, a concept called 'relaxation time' is invoked: this is the time it takes for an ion to change direction by 90°. The relaxation time τ_θ can also be described as 'the average 90° deflection time'. In a plasma there are different relaxation times between each of the different particle species.

2.3.8 Work function, φ

In any solid metal, there are one or two electrons per atom that are free to move from atom to atom. This is sometimes collectively referred to as a 'sea of electrons'. Their velocities follow a statistical distribution, rather than being uniform. Occasionally an electron will have enough velocity to exit the metal without being pulled back in. The minimum amount of energy needed for an electron to leave a surface is called the work function. Specifically, the work function, φ, shown in Fig. 2, is the energy needed to move an electron from the Fermi level E_f into vacuum E_0. 'Fermi level' is the term used to describe the top of the collection of electron energy levels in an atom at absolute zero temperature. The work function is characteristic of the material surface and for most metals is of the order of several electronvolts.

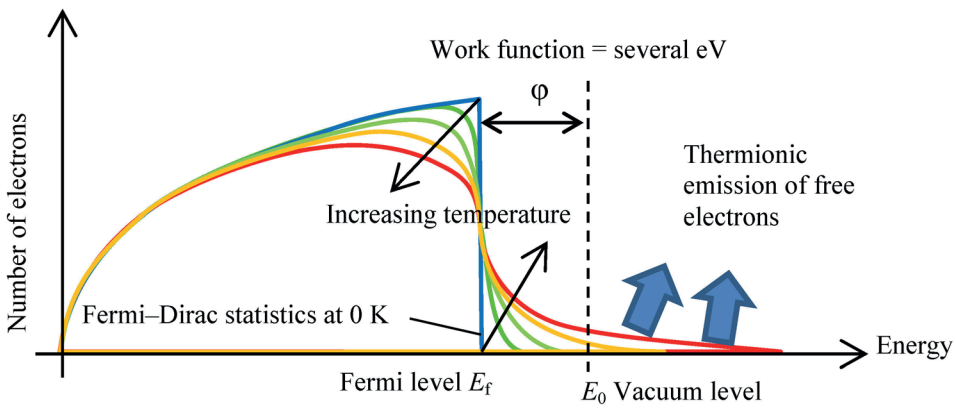


Fig. 2: Electron energy distribution in metals at different temperatures and the work function

2.3.9 Thermionic emission

Thermionic emission is the heat-induced flow of charge carriers from a surface or over a potential energy barrier. This occurs because the thermal energy given to the carrier overcomes the work

function of the metal. Heating a material causes the energy distribution of the electrons to spread out, as shown in Fig. 2. Thermionic currents can be increased by decreasing the work function. This often-desired goal can be achieved by applying various oxide coatings to the surface.

In 1901 Owen Richardson found that the current from a heated wire varied exponentially with temperature. He later proposed this equation:

$$J = A_{\mathrm{G}} T^2 \mathrm{e}^{\frac{-\varphi}{kT}} \tag{3}$$

where J is the electron current density on the surface of the cathode, φ is the cathode work function and T is the temperature of the cathode. Here A_{G} is given by

$$A_{\mathrm{G}} = \lambda_{\mathrm{R}} A_0 \tag{4}$$

where λ_{R} is a material-specific correction factor that is typically of order 0.5 and A_0 is a universal constant given by

$$A_0 = \frac{4\pi m k^2 e}{h^3} = 1.20173 \times 10^6 \text{ A m}^{-2} \text{ K}^{-2} \tag{5}$$

where m is the electron mass, e is the elementary charge and h is Planck's constant.

2.3.10 Debye length

Named after the Dutch scientist Peter Debye, the Debye length, λ_{D}, is the distance over which the free electrons redistribute themselves to screen out electric fields in plasma. This screening process occurs because the light mobile electrons are repelled from each other while being pulled by neighbouring heavy, low-mobility positive ions. Thus the electrons will always distribute themselves between the ions. Their electric fields counteract the fields of the ions, creating a screening effect. The Debye length not only limits the influential range that particles' electric fields have on each other but also limits how far electric fields produced by voltages applied to electrodes can penetrate into the plasma. The Debye length effect is what makes the plasma quasi-neutral over long distances.

The higher the electron density, the more effective the screening, and thus the shorter this screening (Debye) length will be. The Debye length is given by:

$$\lambda_{\mathrm{D}} = \sqrt{\frac{\epsilon_0 k T_{\mathrm{e}}}{n_{\mathrm{e}} q_{\mathrm{e}}^2}} \tag{6}$$

where

λ_{D} is the Debye length (of order 0.1–1 mm for ion source plasmas),

ϵ_0 is the permittivity of free space,

k is the Boltzmann constant,

q_{e} is the elemental charge,

T_{e} is the temperature of the electrons,

n_{e} is the density of electrons.

2.3.11 Plasma sheath

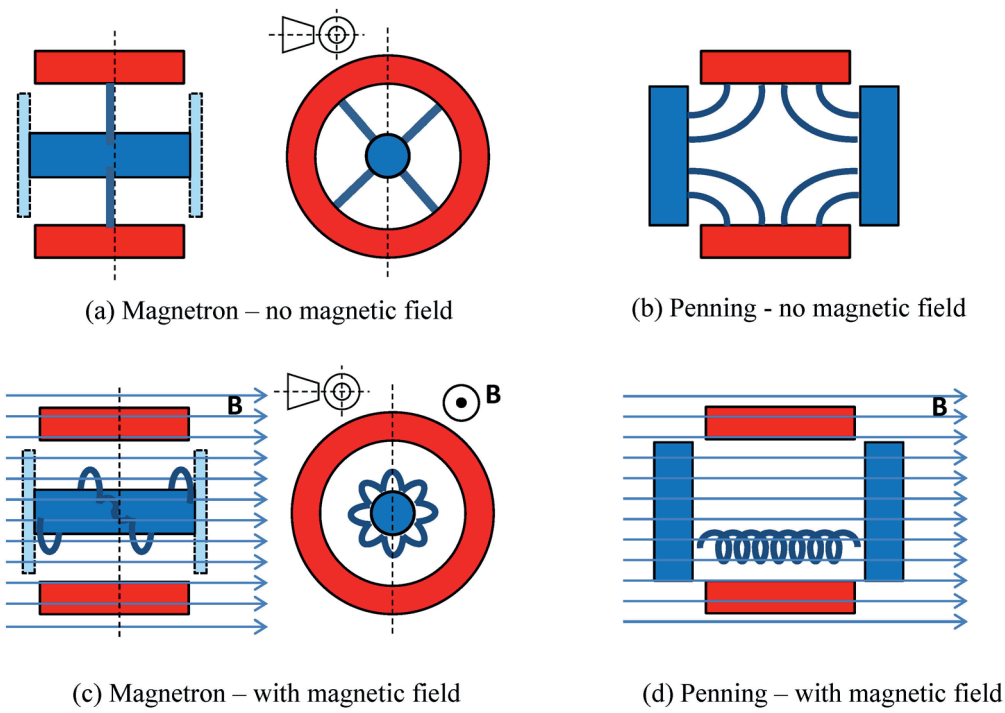
The screening effect of the plasma creates a phenomenon called the plasma sheath around the cathode electrode(s). The plasma sheath is also called the Debye sheath. The sheath has a greater density of positive ions, and hence an overall excess positive charge. It balances an opposite negative charge on the cathode with which it is in contact. The plasma sheath is several Debye lengths thick.

The quasi-uniform plasma potential is closest to the anode voltage, so the largest potential drop in a plasma is across the plasma sheath near the cathode. Positive ions that drift into the cathode sheath region are accelerated towards the cathode by this potential drop. Similarly, negative ions and electrons produced on the cathode surface are accelerated away.

2.4 Magnetic field and electron trajectories

Although the magnetron and Penning discharge geometries share a lot of the basic plasma properties listed in the previous section, they have fundamentally different electron trajectories. This is caused by the different way their electric and magnetic fields are oriented.

Figure 3(c) shows the direction of the dipole magnetic field in the magnetron geometry. Without the magnetic field, the electrons would simply flow axially from the cathode to the anode, as shown in Fig. 3(a). The magnetic field causes the electrons emitted from the cathode to rotate and end up back near the cathode. They are repelled from the cathode only to be rotated back towards the cathode again. This process repeats, causing the electrons to propagate around the cathode in a looping circular motion. As well as propagating around the cathode, the electrons also drift along its length in a helix. The magnetron cathode usually has end discs that make the cathode resemble a spool. The purpose of the end discs is to repel the electrons when they reach the end of the cathode. In this way the electrons spiral up and down the spool-shaped cathode.



(a) Magnetron – no magnetic field

(b) Penning - no magnetic field

(c) Magnetron – with magnetic field

(d) Penning – with magnetic field

Fig. 3: Electron trajectories in magnetron and Penning electrode geometries with and without magnetic fields

Figure 3(d) shows the direction of the dipole magnetic field in the Penning geometry. Without the magnetic field, the electrons would follow the quadrupole-shaped electric field lines, as shown in Fig. 3(b). With the magnetic field, the electrons are emitted from a cathode, and then they drift along the magnetic field lines by spiralling around them. They eventually get to the opposite cathode, where they are reflected back.

In reality, the electron trajectories are much more complex than the idealized ones that are shown in Fig. 3. They have a range of energies (section 2.3.4) and they interact with the electric fields of the ions and other electrons (section 2.3.7), leading to effects such as screening (section 2.3.10) and

the cathode sheath (section 2.3.11). However, the overall effect is that the magnetic field confines the electrons: to circulate around the cathode in the case of the magnetron, and to bounce between the two cathodes in the Penning.

3 Negative ions

3.1 The negative ion

Negative ions are neutral atoms with an additional electron attached. The binding energy of the additional electron to the atom is termed the electron affinity. Some elements (such as beryllium, nitrogen and the noble elements) have a negative electron affinity, which means that they cannot form stable negative ions. Both magnetron and Penning ion sources have been used to produce H^- and D^- ions as well as other heavy negative ions, such as O^-, B^-, C^-, etc.

H^- ions are the most commonly produced negative ion. Hydrogen has an electron affinity of 0.7542 eV. Considering that the binding energy of an electron to a proton is 13.6 eV, the extra electron on an H^- ion is very loosely held on.

It is also important to note that H^- has a much larger ionization cross-section than H^0: 30 times larger for electron collisions and 100 times for H^+ collisions. H^- ions are very fragile.

3.2 Uses

Negative ion sources were first developed to allow electrostatic accelerators to effectively double their output beam energy. In a tandem generator, H^- ions are first accelerated from ground to terminal volts; they are then stripped of their two electrons when they pass through a thin foil. The resulting protons are then accelerated from terminal volts back to ground, at which point they have an energy of twice the terminal volts.

Cyclotrons use negative ions and stripping foils to extract the beam from the cyclotron. The stripping foil is positioned near the perimeter of the cyclotron poles. As the negative ion beam is accelerated, it circulates on larger and larger radii until it passes through the stripping foil, which converts the beam from being negative to positive. The Lorenz force on the beam is reversed and, instead of the force pointing into the centre of the cyclotron, it points outwards and the beam is cleanly extracted.

In high-power proton accelerators, H^- ions are used to allow charge accumulation via multiturn injection. An H^- beam from a linear accelerator is fed through a stripping foil into a circular ring (a storage, accumulator or synchrotron ring), leaving protons circulating in the ring. The H^- beam from the linear accelerator continues to enter the ring while the circulating beam repeatedly passes through the stripping foil unaffected. The incoming beam curves in one way through a dipole as an H^- beam, then curves out of the dipole in the opposite direction as a proton beam on top of the circulating beam. This allows accelerator designers to beat Liouville's theorem and build up charges in phase space. Without this negative ion stripping trick, only one turn on the same orbit could be accumulated in the ring.

3.3 Early negative ion sources

3.3.1 Charge exchange

The first H^- ion sources were charge exchange devices. There are two ways of doing this: with foils or gases. With foils, a proton beam, with an energy of about 10 keV, is passed through a negatively biased foil and by electron capture an H^- beam is produced. For gases, the proton beam is passed through a region filled with a gas. The H^- beam is produced by sequential electron capture; protons are

converted first to neutral H^0, then to H^-. The gas acts as an electron donor. Only about 2% of the protons are converted into H^- ions. Until the 1960s this was the main technique used to make H^- beams. Beams of up to 200 μA were produced using this method. In 1967 Bailey Donnally [7] discovered that the yield of He^- ions can be increased by using caesium vapour as an electron donor. This led to the development of a series of negative ion sources using alkali vapour.

3.3.2 Extraction from the edge of the plasma

For several decades, numerous researchers [8, 9] had been experimenting with sources originally designed to produce positive ions, but by reversing the polarity of the extraction they were able to extract negative ions. Nevertheless, the co-extracted electron current was always at least an order of magnitude higher than the negative ion current.

In the early 1960s George Lawrence and his team at Los Alamos [10] were using a duoplasmatron to produce H^- ions when they first noticed that substantially higher beam currents and lower electron currents could be extracted when the extraction hole was moved away from the centre of the discharge (Fig. 4). They concluded that the extracted H^- ions must be produced near the edge of the plasma. (This was also discovered independently by a team at the UK Atomic Weapons Establishment [11].)

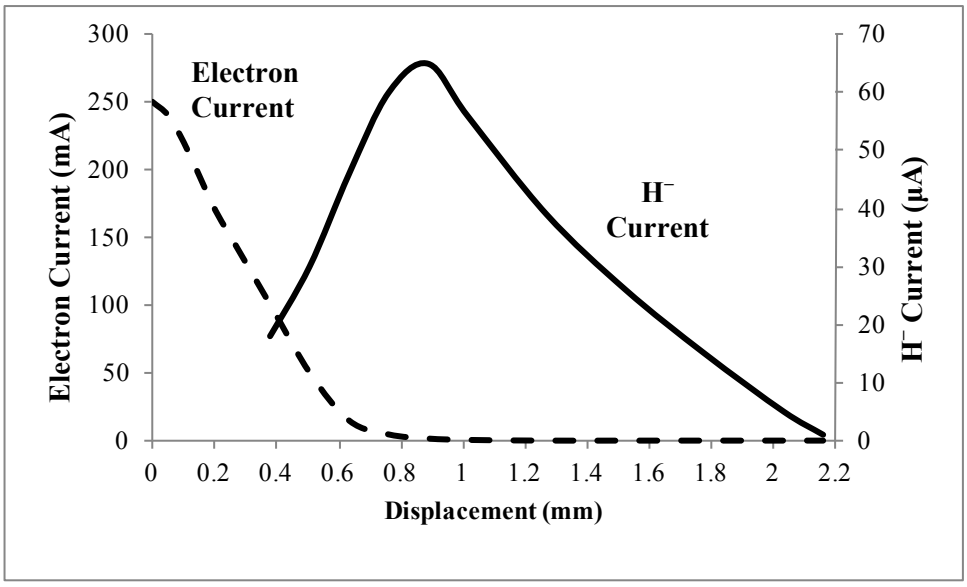


Fig. 4: H^- and electron currents as a function of extraction offset in a duoplasmatron measured at Los Alamos

During the 1960s various sources originally designed to produce positive ions were adapted and modified to produce H^- ions, and beam currents up to a few milliamps were produced.

3.4 Production mechanisms

3.4.1 Overview

The physical processes involved with the production of negative ions are still not fully understood, but they can be generally described as: charge exchange, surface and volume production processes. In different types of source, one production process may dominate; however, all three processes might contribute to the overall extracted negative ion current.

3.4.2 Charge exchange

Already described in section 3.3.1, charge exchange was the first method used to produce H^- ions. Resonant charge exchange processes also occur in the Penning surface plasma source. H^- ions produced on the cathode surface are accelerated by the cathode sheath (see section 2.3.11) to the plasma potential (≈ 60 V). If these fast H^- were to be directly extracted they would create a beam with a large energy spread. The fast H^- undergo resonant charge exchange with slow (thermal) neutral hydrogen atoms (H^0) in the plasma, leaving slow H^- ions ready for extraction:

$$H^- (\approx 60 \text{ eV}) + H^0 (< 1 \text{ eV}) \rightarrow H^- (< 1 \text{ eV}) + H^0 (\approx 60 \text{ eV}) \quad (7)$$

3.4.3 Surface production

In magnetron and Penning negative ion sources, the primary means of H^- ion production is via caesium-enhanced surface production. This was discovered by Soviet scientists in the 1970s. They called ion sources using this production mechanism 'surface plasma sources' (SPS). Caesium-enhanced surface production of negative ions is discussed in more detail in section 4.

3.4.4 Volume production

Also in the 1970s Marthe Bacal working at Ecole Polytechnique discovered a completely new production mechanism that relied on H^- production in the plasma volume itself. Initially people were sceptical because H^- ions are so fragile (as discussed in section 3.1). The plasma in the discharge was thought to be too energetic for any H^- ions produced in the volume to survive long enough to make it to the extraction region. The breakthrough that made volume production possible was separation by a magnetic filter field of the plasma production region from the extraction region. In the 1980s, Leung, Ehlers and Bacal used a filament-driven multicusp ion source with a magnetic dipole filter field positioned near the extraction region. The filter field blocked high-energy electrons from entering the extraction region, whereas ions and cold electrons could diffuse across the filter field. This effectively separated the discharge into two regions: a high-temperature driver plasma on the filament side of the filter field, and a low-temperature H^- production plasma on the extraction region side. Magnetically filtered multicusp ion sources are sometimes called 'tandem' sources because of these two regions of different plasma temperatures (not to be confused with tandem accelerators).

The volume production process relies on the dissociative attachment of low-energy electrons to rovibrationally excited H_2 molecules:

$$H_2^* + e (\leq 1 \text{ eV}) \rightarrow H^- + H^0 \quad (8)$$

If the H_2 molecule is vibrationally cold, the dissociative attachment cross-section is extremely low (10^{-21} cm^2). When the H_2 molecule is rovibrationally excited, however, the cross-section increases by five orders of magnitude. Thus, low-energy electrons can be very effective in generating H^- ions by dissociative attachment to highly vibrationally excited molecules. The rovibrationally excited molecules are produced not only in the plasma but also on the walls of the chamber and electrode surfaces.

It is possible that some of the extracted H^- ions in the magnetron and Penning sources are created by volume production near the extraction aperture, with the magnetic dipole field acting as the fast electron filter.

3.5 Negative ion extraction

One of the challenges with negative ion source design is how to deal with the co-extracted electrons. Extracting electrons with the negative ions is obviously unavoidable because they both have the same charge. The ion source engineer must first try to minimize the amount of co-extracted electrons, and

then find a way to separate and dump the unwanted electrons from the negative ion beam. The electron current can be 10 times greater than the negative ion current itself.

4 Caesium and surface production

4.1 History

4.1.1 Caesium and surfaces

In the early 1960s Victor Krohn, Jr and his team [12] at Space Technology Laboratories, Inc., California were experimenting with surface sputter ion sources. Surface sputter ion sources are mainly used to produce beams of heavier ions (such as metals) for coating and etching applications. Krohn noticed that, when Cs^+ ions were used to sputter a metal target, the yield of sputtered negative ions increased by an order of magnitude.

4.1.2 Soviet breakthrough

In the early 1970s Gennadii Dimov, Yuri Belchenko and Vadim Dudnikov at the Budker Institute of Nuclear Physics started experimenting with caesium in ion sources. Using a magnetron ion source (which the Soviets called a planotron), Dudnikov added Cs vapour to the discharge for the first time. A dramatic increase in H^- current was observed, along with a decrease in co-extracted electrons. The Dimov team went on to extract a colossal 880 mA pulsed H^- beam from an experimental magnetron ion source [13].

This success led them to develop a Penning-type ion source that could produce 150 mA of pulsed H− beam current with only 250 mA of extracted electrons [14]. The H− currents produced were orders of magnitude higher than anything seen previously. When these revolutionary results were published, interest in caesiated ion sources took off. Researchers all over the world started using caesium in their ion sources, and a large number of new ion source designs were developed.

The Soviets went on to develop a semi-planotron (essentially half a magnetron geometry where the electrons only circulate round half of a cathode spool), which they claimed was capable of delivering absolutely enormous 11 A H^- beam currents.

4.1.3 American developments

The most intense development occurred in the USA. Krsto Prelec and his team at Brookhaven National Laboratory (BNL) improved the geometry with the addition of a cathode indent (see section 5.3). They also developed a 2 A H^- beam current magnetron for neutral beam injectors for fusion projects. Chuck Schmidt *et al.* took the BNL design to Fermilab and developed a magnetron for accelerator applications that still runs today.

At Los Alamos National Laboratory, Paul Allison *et al.* developed the Penning source design, which, combined with some aspects of the Fermilab design, was then used on the ISIS spallation neutron source in the UK.

A very different type of H^- source that also relies on surface production of H^- ions is the surface converter source. Developed in the 1980s by Ehlers and Leung at the Lawrence Berkeley Laboratory, it also relies on a caesiated surface [15]. A caesiated surface sits in the middle of a large filament-driven multicusp confined plasma. The surface is curved, with a radius centred on the extraction region. H^- ions produced on this surface are geometrically 'focused' towards the extraction hole because they are repelled by a negative potential applied to the converter surface. This type of source is said to be 'self-extracting'.

4.2 Surface production of negative ions

4.2.1 Surface physics processes

Surfaces are very interesting places. Many different types of particles arrive at a surface: protons, ionized hydrogen molecules, ionized caesium, neutral atoms or molecules. They could be travelling at thermal velocities of a fraction of an electronvolt, or they could be travelling at up to 50–80 eV after being accelerated by the plasma sheath potential (section 2.3.11).

When a particle interacts with a surface, many complex and competing processes can occur:

- reflection,
- adsorption,
- sputtering,
- desorption,
- recombination,
- dissociation,
- ionization,
- neutralization,
- secondary electron emission,
- photoemission,
- excitation.

Particles ejected from the surface could be of a different charge state from the incoming particle, be excited, be in a molecule, or be some combination of all three. The surface may also be altered. Complex interactions can take place at the surfaces.

4.2.2 Low-work-function surface production

To make H^- ions, the surface must provide electrons. Surfaces tend to hold on to electrons – they have an electron affinity, or work function, φ. So to make H^- ions, a low-work-function surface is essential.

The work function of a surface obviously depends on what it is made of. If atoms of a different element are adsorbed on that surface (such as caesium), then the work function can be altered. The 'thickness' of the adsorbed layer will also have an effect on the surface's work function. The thickness of the adsorbed layer is usually defined in terms of the number of 'monolayers' of the adsorbed atoms:

$$\text{thickness (number of monolayers)} = \frac{\text{number of adsorbate atoms per unit area}}{\text{number of adsorbate atoms for a monolayer per unit area}} \tag{9}$$

When talking about negative ion production, the surface is usually the cathode and is typically made of a high-melting-point metal such as tungsten, φ = 4.55 eV, or molybdenum, φ = 4.6 eV. Caesium has the lowest work function of all elements: φ = 2.14 eV. The work function of a caesium-coated molybdenum surface is actually lower than that of bulk caesium. As caesium covers the molybdenum surface, the work function decreases to 1.5 eV at 0.6 of a monolayer and then rises to about 2 eV for one monolayer or greater of caesium, as shown in Fig. 5. This minimum at 0.6 monolayers is caused by atomic interactions increasing the Fermi level at the surface, thus decreasing the amount of energy required to liberate the electrons.

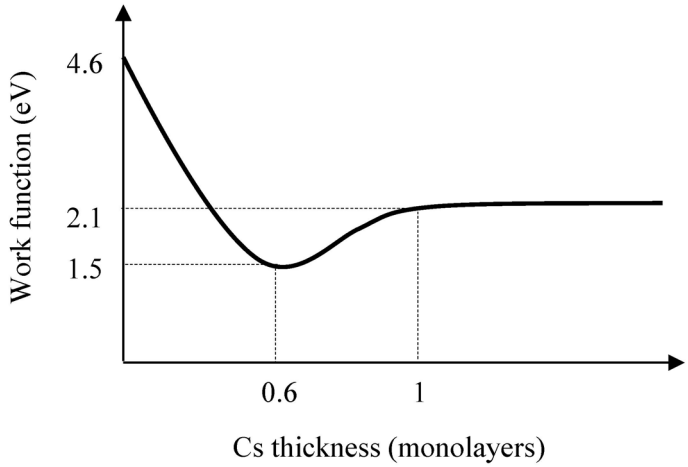


Fig. 5: Surface work function versus caesium thickness on a molybdenum surface

Although this sounds like an impossibly precise surface coverage to maintain, it is relatively simple to engineer. The thickness of the caesium coating on electrode surfaces will rarely ever come close to one full monolayer because thermal desorption and plasma sputtering remove excess caesium. This is because the Cs–Cs bond is weaker than the Cs–Mo, Cs–Ta or Cs–W bonds, so Cs atoms adsorbed on Cs atoms are rapidly sputtered away by the plasma or thermally emitted from hot surfaces. (It is possible to build up multiple layers but only on cold surfaces that are shielded from the plasma.)

4.2.3 Maintaining caesium coverage

To minimize the work function and maximize the H^- production, an optimum layer of caesium must be maintained on the surface. The surface is a dynamic place – caesium atoms are constantly being desorbed by plasma bombardment. To maintain optimum caesium coverage in a surface plasma source, a constant flux of caesium is required. This is often provided by an oven containing pure elemental caesium, but it can also be provided using caesium chromate cartridges that release Cs when heated.

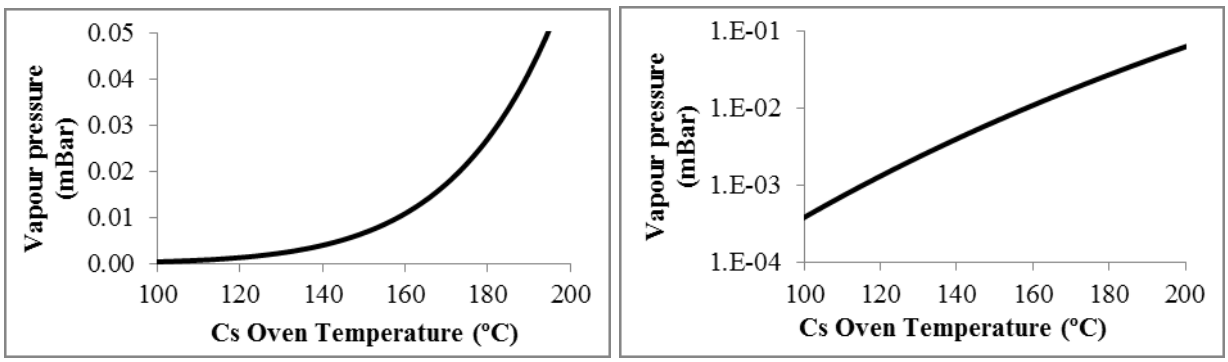


Fig. 6: Caesium vapour pressure versus caesium oven temperature

The flux of caesium into the plasma can be precisely controlled by setting the temperature of the caesium oven. Fig. 6 shows how the vapour pressure of caesium varies with temperature [16]. Ion sources that use elemental caesium in an oven usually operate somewhere between 70°C and 190°C. This covers a very large range of vapour pressures suitable for most applications.

4.3 H^- destruction

There are many processes that can destroy H^- ions. The most common are:

$H^- + H^+ \rightarrow H^0 + H^0$	mutual neutralization
$H^- + e \rightarrow H^0 + 2e$	electron detachment
$H^- + H^0 \rightarrow H_2^* + e$	associative detachment

As already mentioned, the H^- ion cross-section is 30 times larger than the neutral H^0 atom for collisions with electrons and 100 times larger for collisions with H^+ ions. As well as being fragile and easily destroyed, it is much more likely to be hit.

The aim of the source designer is to minimize the H^- destruction processes by controlling the geometry, temperature, pressure and fields in the source.

4.4 The additional benefits of caesium

In addition to aiding H^- surface production, caesium also helps to stabilize the plasma by readily ionizing to produce additional electrons for the discharge. This reduces the amount of noise in the discharge and extracted beam current.

5 Magnetron surface plasma source

5.1 Introduction

Thus far this chapter has discussed features common to all magnetron and Penning electrode geometries, along with negative ions and surface production on caesiated surfaces. This section gives more details specifically about magnetron surface plasma negative ion sources. Examples from different operational particle accelerators around the world are given.

5.2 Construction

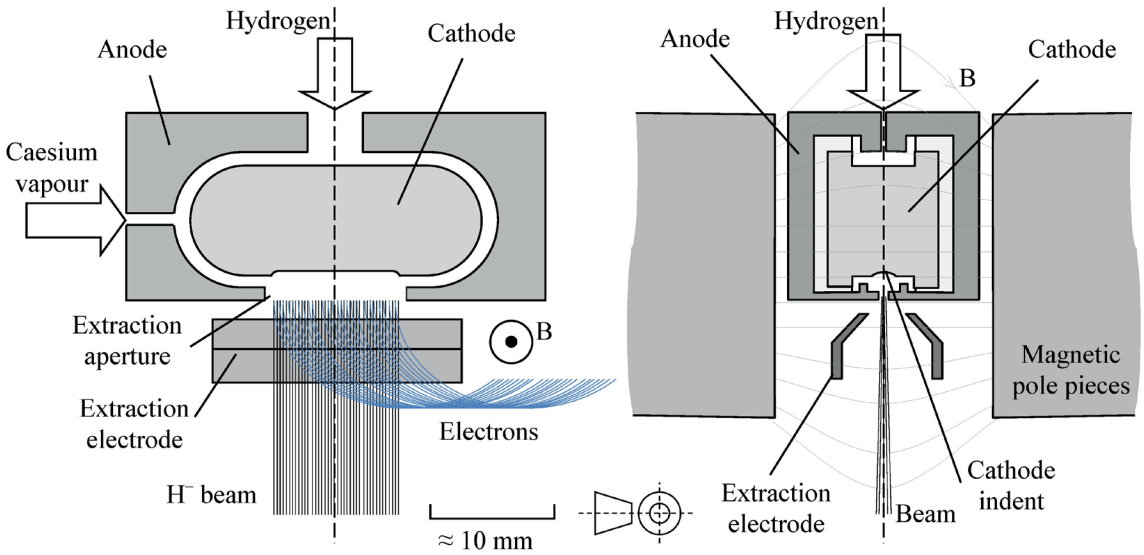


Fig. 7: Sectional schematic of a magnetron with slit extraction

The magnetron surface plasma source has a racetrack-shaped discharge bounded on the inside by the spool-shaped cathode and on the outside by the anode, as shown in Fig. 7. The anode and cathode are

only about 1 mm apart, so the discharge is quite thin in the shape of a ribbon wrapped around the cathode. The anode and cathode are separated by a ceramic spacer. A magnetic field of between 0.1 T and 0.2 T is applied perpendicular to the plane of the racetrack, which causes the electrons to circulate around the racetrack. On one of the long sides of the racetrack discharge the anode has a hole through which the beam is extracted. Hydrogen is pulsed with a piezoelectric valve into the discharge on the opposite side to the extraction hole. Caesium vapour is introduced via an inlet on one side from a heated oven containing elemental caesium.

5.3 H^- production

There is an indent in the cathode opposite the extraction aperture. The indent increases the extracted H^- beam current by geometrically focusing the H^- ions produced on the cathode surface. The shape of the indent depends on the shape of the extraction geometry. A spherical dimple indent is used for a circular aperture and extraction electrode. A cylindrical groove indent is used for a slit aperture and extraction electrode, as shown in Fig. 6.

H^- ions are produced on the cathode surface and are accelerated away by the cathode sheath potential. Some of the cathode sheath-accelerated H^- ions undergo resonant charge exchange with slow thermal H^0 on the way to extraction, resulting in a beam energy distribution with two peaks.

5.4 Extraction

The amount of beam current extracted increases with extraction voltage. Voltages up to 35 kV have been used for ion sources on operational machines that produce 100 mA [17]. It is possible that the co-extracted electron current will be over 10 times the H^- beam current depending on how the magnetron is operated. There must be provision for dumping electron beam currents of this magnitude.

The ion source designer must consider the co-extracted electron current when specifying extraction power supplies. Caesium will also cause the extraction system to flash over in normal operation and during start-up. The high-voltage extraction power supply must be tough enough to withstand these breakdowns.

5.5 Temperature control

Technically speaking, the type of magnetron under discussion is a cold-cathode surface plasma magnetron. The term ‘cold-cathode’ is used because the cathode is not in the form of a heated filament. However, the name can be misleading because the cathode must still operate at elevated temperatures to maintain the Cs balance allowing stable and reliable operation. The electrodes generally run at a few hundred degrees Celsius; they are heated by the power of the discharge, and sometimes additional heating elements are used if the source operated at very low duty cycles. The temperature can be controlled by some form of cooling system if the source runs at higher duty cycles.

5.6 Caesium trapping

Caesium is exceptionally good at producing lots of electrons, so it can also promote high-voltage breakdowns if it gets onto electrode surfaces. The first accelerator applications of the negative magnetron used extra high voltages (EHV) generated by a Cockcroft–Walton voltage multiplier as the primary post-extraction acceleration technique. It was therefore essential to prevent caesium entering the high-voltage acceleration column by using some form of trap. The most common trap design is a 90° dipole inside a cold box, as shown in Fig. 8. The beam is bent through 90°, which also serves to separate out the co-extracted electrons; the beam exits through a hole in the cold box after the 90° bend. The cold box is held between −40°C and 0°C using refrigerant; this causes caesium vapour escaping from the source to condense on the walls of the cold box. The dipole field is also used in the

ion source itself to confine the electrons. The refrigerated cold box is usually biased at the extraction electrode voltage so it can also be used to cool the extraction electrode by a direct thermal contact.

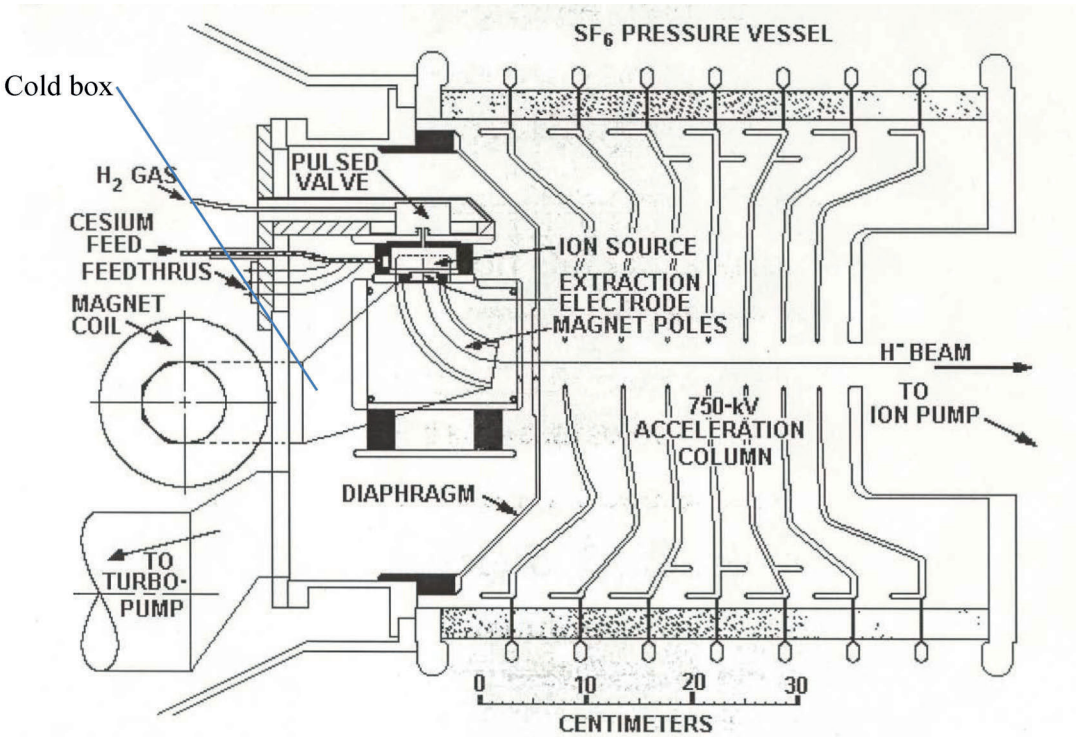


Fig. 8: The FNAL magnetron, cold box and 750 kV acceleration column

Radio frequency quadrupole (RFQ) accelerators have now largely replaced EHV. It is now common to post-extraction accelerate the beam up to somewhere between 30 and 70 keV, then pass the beam though a magnetic low-energy beam transport (LEBT) system into an RFQ. The LEBT is usually long enough (1–2 m) to prevent any caesium from entering the RFQ without a cold box trap. The caesium is deposited on the beam pipe walls in the LEBT, but this is less of a problem because only magnetic fields are present.

5.7 Duty cycle and noise limitations

The magnetron can deliver high H^- currents of over 100 mA and can have very long lifetimes of over six months, but only at very low duty factors less than 0.5%. This because it is not possible to maintain optimum caesium coverage on the cathode surface opposite the extraction aperture for higher duty factors.

The beam from a magnetron is also noisier than that of a Penning ion source.

5.8 Examples of negative ion source magnetrons around the world

5.8.1 Budker Institute of Nuclear Physics (BINP), Novosibirsk, Russia

The magnetron (also called a planotron) was the first ion source where the H^- current was significantly increased by adding caesium vapour. This work was done by Gennadii Dimov, Yuri Belchenko and Vadim Dudnikov at the Budker Institute of Nuclear Physics in the early 1970s [13]. Many different magnetron designs were developed at BINP.

5.8.2 Fermi National Accelerator Laboratory (FNAL), USA

In the late 1970s Chuck Schmidt developed the FNAL version of the magnetron ion source shown in Fig. 9. The ion source is mounted face down on a flange inserted from the rear, as shown in Fig. 8. The negative ion beam is then extracted and bent through a 90° bend in a cold box caesium trap.

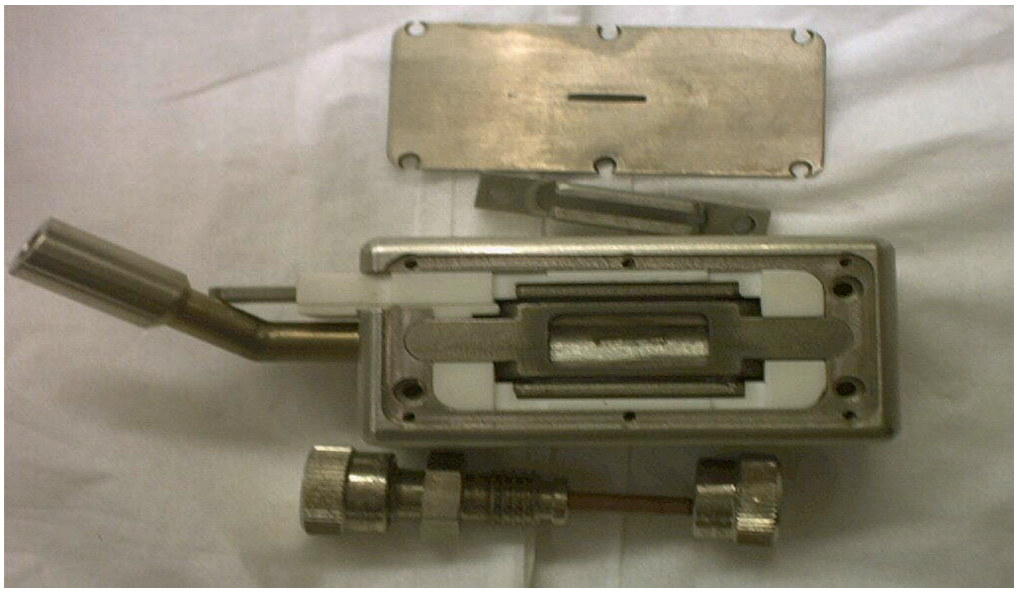


Fig. 9: The FNAL surface plasma magnetron source

5.8.3 Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany

The FNAL magnetron design was adopted by DESY and further developed by Jens Peters [18]. The DESY magnetron ran on the HERA accelerator. The design is slightly different from the FNAL design in that the ion source is top loaded into the 90° magnet and caesium trap cold box, as shown in Fig. 10.

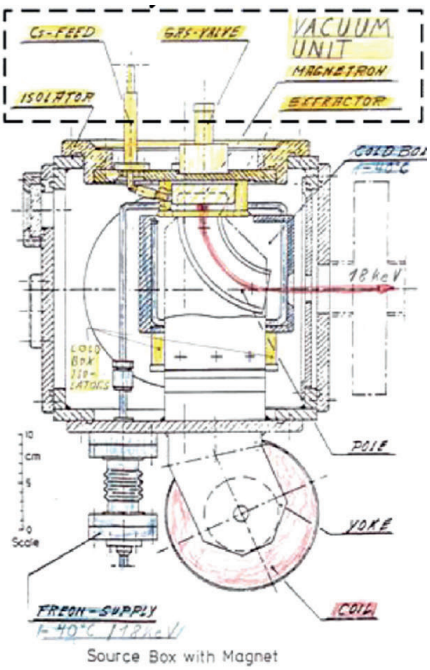


Fig. 10: The HERA top-loading magnetron and caesium trap

The different components of the HERA magnetron are shown in Fig. 11.

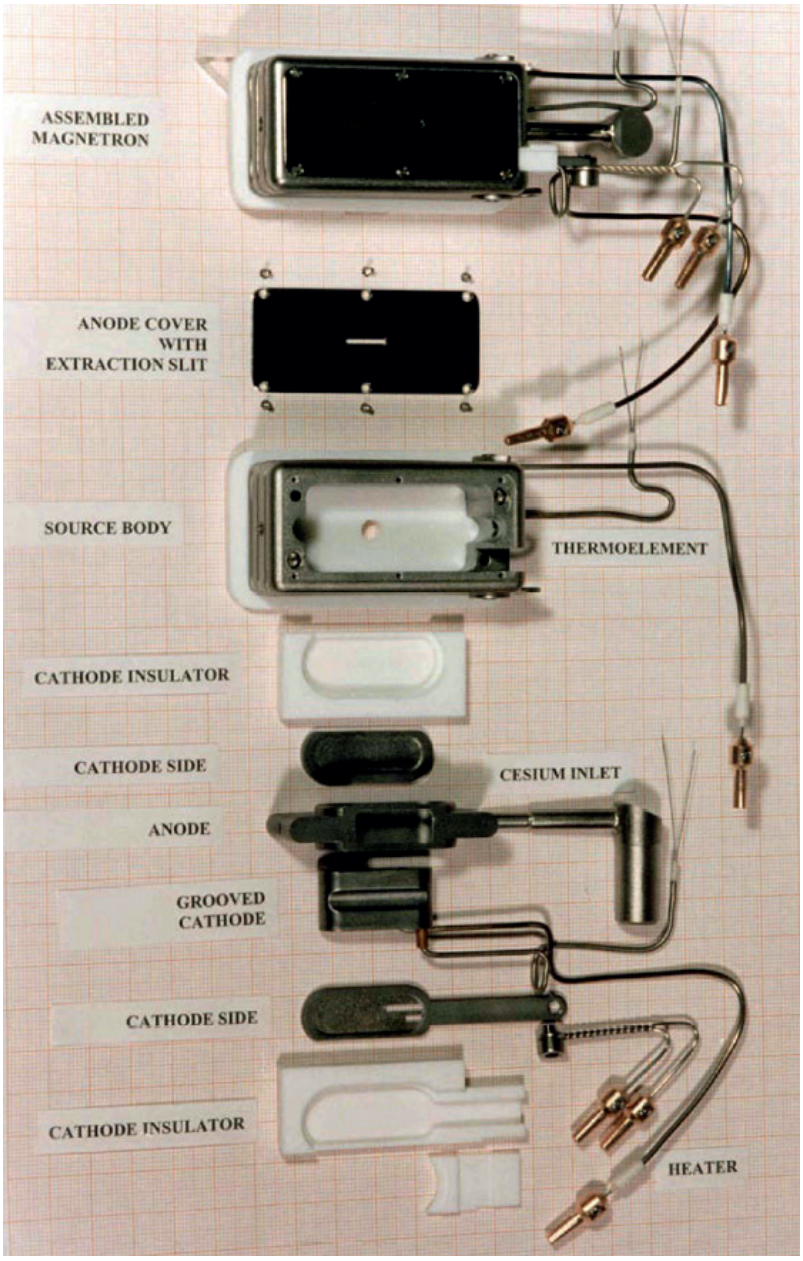


Fig. 11: The HERA magnetron components

5.8.4 Brookhaven National Laboratory (BNL), USA

Jim Alessi at Brookhaven further optimized the FNAL design [17]. He was able to reduce the discharge current, increase the extraction voltage and use permanent magnets. This highly reliable, very efficient magnetron, shown in Fig. 12, is still in use today. The operational parameters for this source are shown in Table 1.

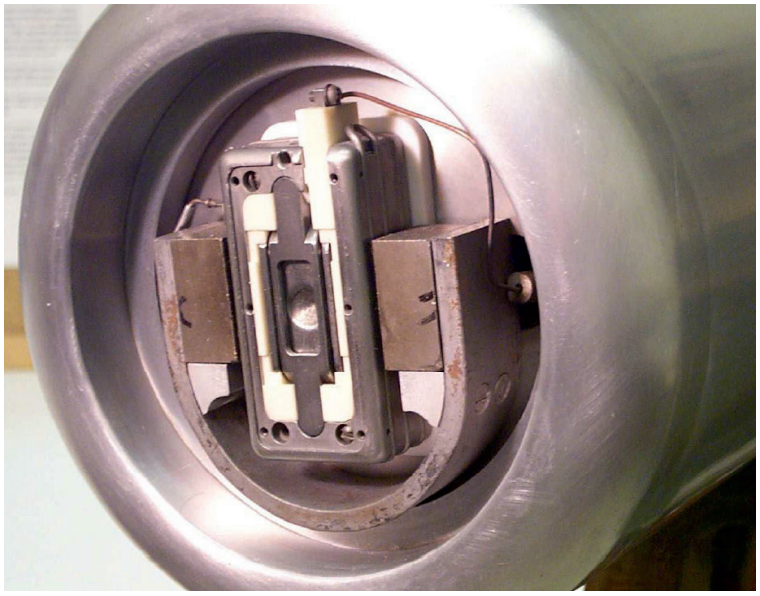


Fig. 12: The BNL magnetron with the anode cover plate and extract electrode removed

Table 1: The operational parameters of the BNL magnetron.

H^- beam current	90–100 mA
Extraction voltage	35 kV
Discharge voltage	140–160 V
Discharge current	8–18 A
Rep. rate	7.5 Hz
Beam pulse width	700 μs
Duty factor	0.5%
Cs consumption	0.5 mg h^{-1}
H_2 gas flow	3 mL min^{-1}
R.m.s. emittance	0.4 π mm mrad (normalized)
Lifetime	270 days

6 Penning surface plasma source

6.1 Introduction

This section gives more details specifically about Penning surface plasma negative ion sources. Examples from different operational particle accelerators around the world are given.

6.2 Construction

The Penning negative ion source (Fig. 13) has a small (approximately 10 mm × 5 mm × 5 mm) rectangular discharge region with a transverse magnetic field. The long sides of discharge are bounded by two cathodes, with the other four walls at anode potential, creating a 'quadrupole-like' electric field arrangement. The two cathodes are parallel jaws machined from a single piece of molybdenum to give good thermal conduction to the mounting flange. The magnetic field is oriented orthogonally to the cathode surfaces so that electrons emitted from the cathode are confined by the magnetic field lines and reflect back and forth between the parallel cathode surfaces. The primary anode is hollow and has holes through which hydrogen and caesium vapour are fed into the discharge. The extraction aperture plate is connected to the primary anode, which completes the window frame anode. The beam is extracted from the plasma through a slit in the extraction aperture plate by a high voltage applied to an extraction electrode. All electrodes are made of molybdenum.

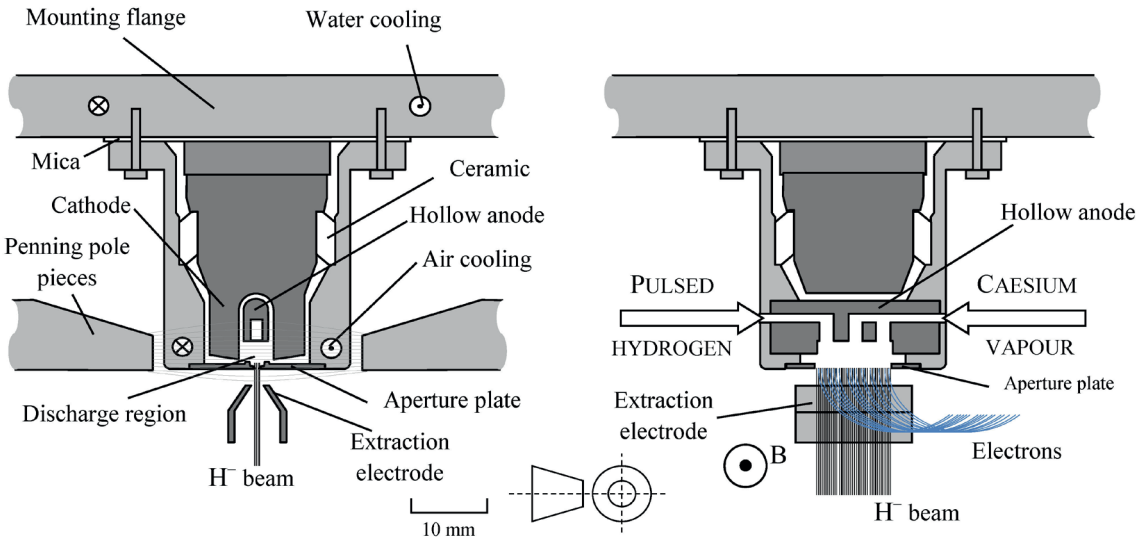


Fig. 13: Sectional schematic of a Penning H^- ion source

6.3 H^- production

Like with the magnetron, H^- ions are produced on the cathode surfaces and accelerated by the plasma sheath potential that exists next to the cathode. The plasma sheath potential is about 60 V. Unlike the magnetron, however, the cathode surface is not directly opposite the extraction aperture. The addition of ribs on the inside of the extraction aperture plate means that there is no direct line of sight between the cathode surface and the extraction aperture. Thus, it is impossible for the fast cathode-produced H^- to be extracted directly. Only H^- ions that have undergone resonant charge exchange with slow H^- ions (see section 3.4.2) are extracted, resulting in a beam with a lower energy spread than from the magnetron.

6.4 Temperature control

Like the magnetron discussed in the previous section, technically speaking the Penning negative ion source is a cold-cathode surface plasma Penning. Also, like the magnetron, 'cold-cathode' refers to the

fact that the cathode is not a hot filament, but the cathode must still operate at elevated temperatures to maintain the Cs balance required for stable operation. The Cs-coated electrodes must also run at an elevated temperature to emit significant numbers of electrons. The electrodes generally run at a few hundred degrees Celsius, heated by the power of the discharge and temperature-controlled by some form of cooling system. From Eq. (3) the thermionic emission for Cs on Mo increases by 13 orders of magnitude between 300 K and 600 K. This is why Penning surface plasma sources will not run at room temperature.

6.5 Extraction

The Penning is the brightest H^- ion source, with current densities at extraction above 1 A cm^{-2} possible. The same points as mentioned in section 5.4 about the magnetron also apply to the Penning.

6.6 Caesium trapping

The same points as mentioned in section 5.6 about the magnetron also apply to the Penning.

6.7 Examples of negative Penning sources around the world

6.7.1 Budker Institute of Nuclear Physics (BINP), Novosibirsk, Russia

The H^- Penning ion source was invented by Vadim Dudnikov, and developed with Gennadii Dimov and Yuri Belchenko at the Budker Institute of Nuclear Physics in the early 1970s at the same time as the magnetron source. Dudnikov reported pulsed H^- currents up to 150 mA and that the source could run with duty cycles all of the way up to d.c. [14]. New versions of the Penning source continue to be developed at BINP by Belchenko and his team. In 2012 they reported H^- beam currents of 25 mA d.c. [19].

6.7.2 Institute for Nuclear Research (INR), Moscow, Russia

An almost unmodified version of the original 1970s BINP source is still routinely used at INR. Two views of the INR Penning are shown in Fig. 14. The first shows the anode and cathode with the anode cover plate and extract electrode removed, and the second shows the ion source assembly.

Fig. 14: The INR Penning H^- ion source

6.7.3 Los Alamos National Laboratory (LANL), USA

The Americans took the BINP Penning design and developed it further. In the 1990s Vernon Smith, Paul Allison and Joe Sherman at Los Alamos developed a scaled-up Penning source [20] shown in Fig. 15 that produced 120 mA H^- beam currents in 500 µs pulses at 60 Hz.

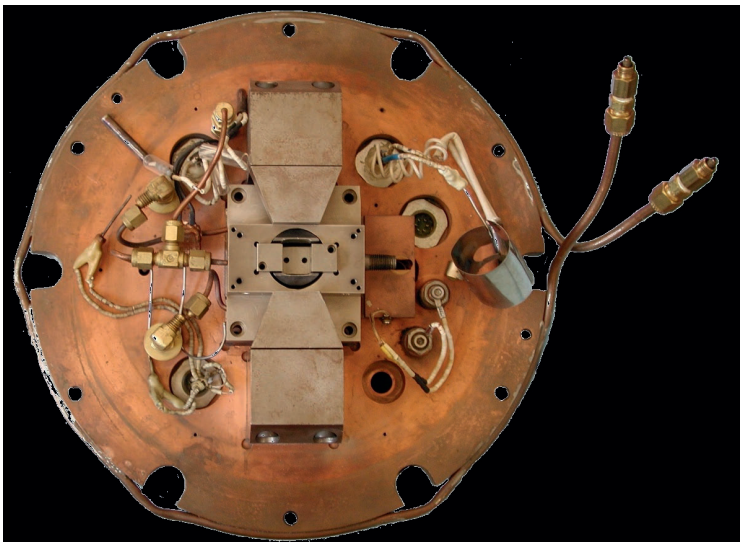


Fig. 15: The LANL scaled Penning H^- ion source with the anode cover plate and extract electrode removed

6.7.4 Rutherford Appleton Laboratory (RAL), UK

In the early 1980s Peter Gear and Roger Bennett at RAL combined an early LANL Penning design with some of the aspects of the FNAL magnetron design. They used the FNAL 90° magnet, cold box and rear-loading design for use on the ISIS accelerator. The design has slowly evolved over the years. The present design used for ISIS operations is shown in Fig. 16.

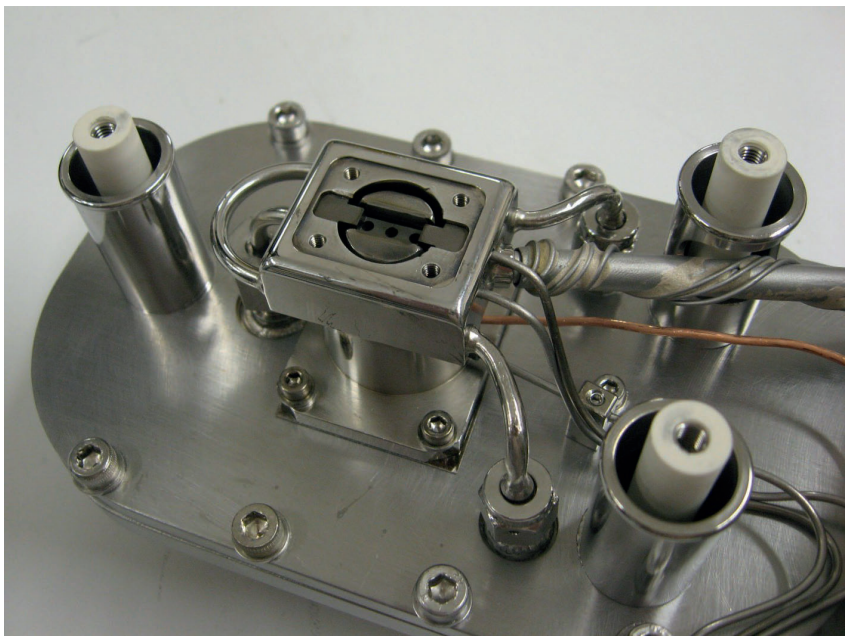


Fig. 16: The RAL Penning H^- ion source with the anode cover plate and extract electrode removed

The operational parameters for the ISIS source are shown in Table 2.

This source is currently under development at the Rutherford Appleton Laboratory by the author and his team, and 60 mA, 1 ms, 50 Hz pulses are now routinely produced [21].

Table 2: The operational parameters of the ISIS operational Penning source.

H^- beam current	35 mA
Extraction voltage	17 kV
Discharge voltage	60–80 V
Discharge current	55 A
Rep. rate	50 Hz
Beam pulse length	250 μs
Duty factor	0.5%
Cs consumption	3 mg h^{-1}
H_2 gas flow	20 mL min^{-1}
R.m.s. emittance	0.35 π mm mrad (normalized)
Lifetime	25 days

7 Discussion

7.1 Operating and tuning

All ion sources require a certain amount of tuning of the operating parameters for optimum performance. Surface plasma sources are no exception.

The main operating parameters are: hydrogen pressure, caesium vapour pressure, discharge current, magnetic field strength and the temperature of the electrodes. Secondary parameters such as timing and repetition rate have an effect on the transient hydrogen pressure and the degree of caesium coverage on the electrodes. Electrode cooling controls the temperature of the electrodes, which also controls the degree of caesium coverage.

Surface plasma ion sources are difficult to work with because their operation depends on optimum electrode surface temperature and optimum caesium coverage. Unfortunately, these optima are in the range of several hundred degrees Celsius and less than a monolayer, respectively, so it is impossible to just set all the parameters and turn the source on. The optima must be arrived at over time.

The optimum surface temperature must be arrived at by heating the electrodes (either with heaters or by the discharge itself). The optimum caesium coverage must be arrived at by allowing the flux of caesium onto the electrode surfaces to stabilize with the caesium sputtering/desorption rate from the surfaces.

When the optimum electrode temperature and caesium coverage have been reached, the source must be stabilized. This involves applying the correct amount of cooling to maintain the average electrode temperature, and the correct caesium oven temperature to maintain the flux onto the surface.

The caesiated Penning geometry operates in two distinct types of discharge mode: high-impedance (low-current), and low-impedance (high-current) modes. There is a sharp transition between the two modes that occurs when the electrodes are hot enough to maintain the correct caesium coverage. It is the low-impedance (high-current) discharge mode that is required for high-current H^- ion production.

The two operating modes have very different operating conditions: the high-impedance mode has a discharge voltage of several hundred volts, whereas the low-impedance mode has a discharge voltage of about 60–70 V. When the source is cold and first switched on, it runs in high-impedance mode. It is impossible to produce a high-current plasma in high-impedance mode.

In normal running conditions, the source operates in low-impedance (high-current) mode. If the source drifts from its optimum conditions, it can slip back into high-impedance mode. The discharge current drops, the electrodes cool, the source becomes unstable and the discharge goes out. It can only be restarted by applying a high voltage and recovering the optimum surface temperatures and caesium coverage. This can make experimenting with different operating parameters quite time-consuming.

The best way to run surface plasma sources for operational accelerators (once they have started up) is to try and keep everything steady. Any changes to the operating parameters should be made one at a time and in very small steps with long pauses (hours or days) in between. This process is called tuning, and requires skill and patience.

7.2 Sputtering

In both magnetron and Penning ion sources, the discharge is in direct contact with the anode and cathode, so sputtering processes will eventually erode the electrode surfaces. This puts a fundamental limit on the lifetime of magnetron and Penning surface plasma sources. The degree of sputtering increases with the size of the projectile atom. Caesium atoms are therefore a major contributor to the total amount of sputtering that occurs.

Sputtering not only removes material from electrode surfaces but also redeposits it. This can cause build-up of sputtered material in the discharge volume. Figure 17 shows the inside of the plasma aperture plate electrode from the RAL Penning H^- ion source after 26 days running. Significant erosion of the ribs either side of the aperture slit are visible. Deposited material is also visible all over the plate. The deposited material is molybdenum that has been sputtered from the cathode surfaces by bombardment with Cs^+ ions accelerated by the cathode sheath.

The sputtered material can flake off and bounce around inside the discharge volume, causing the discharge to appear unstable. The high-current discharge is a violent place and the flake debris can be broken up and vaporized; however, sometimes a flake can partially block the aperture slit, as shown in Fig. 17. This causes a step change in the beam current. If the blockage is small, it might be possible to increase the output current by tuning. It is also possible that a flake might short out the anode and cathode. If this happens, all beam is lost, and the source starts to cool. It is very rare, but possible, that a flake large enough to short out the extraction electrode is spat out of the aperture slit.

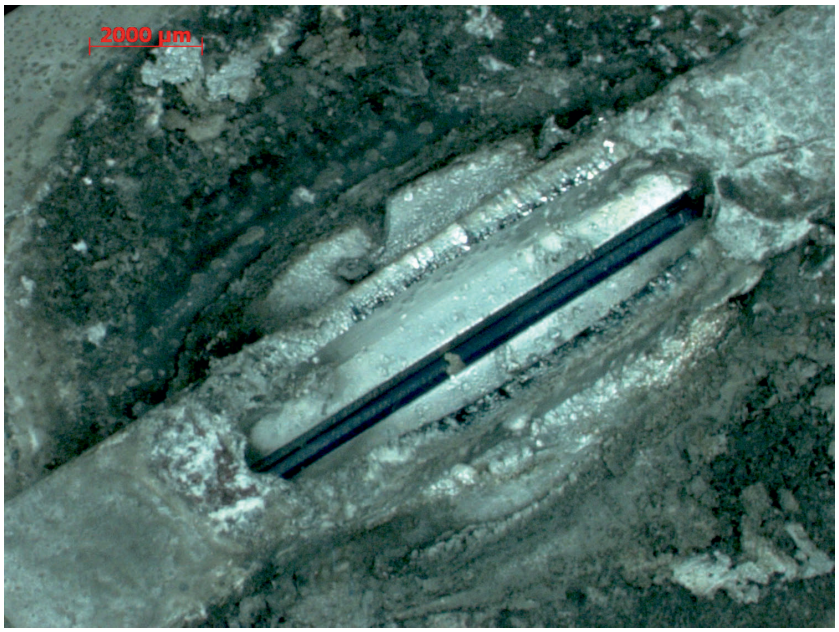


Fig. 17: The inside of the plasma aperture plate electrode from the RAL Penning H^- ion source after 26 days running.

7.3 Other failure modes

The other source failures are caused by failures of the ancillary equipment used to make the source work. These could be failures in power supplies, temperature controllers, heating, cooling, hydrogen delivery, caesium delivery, vacuum pumps, timing and control systems etc.

7.4 Lifetime comparison

Magnetrons are generally used for lower duty cycles and Penning sources for higher duty cycles. Table 3 shows the lifetimes of some of the different sources that have good operational data. On first inspection, the sources have very different lifetimes; however, when the plasma duty factors are taken into account, the lifetimes in integrated plasma days are similar.

Table 3: Comparison of the lifetimes of the different surface plasma sources.

	DESY magnetron	FNAL magnetron	BNL magnetron	RAL Penning
Discharge current (A)	47	50	18	55
Plasma pulse length (µs)	75	80	700	800
Rep. rate (Hz)	6.25	15	7.5	50
Plasma duty factor (%)	0.047	0.12	0.525	4
Lifetime (days)	900	200	270	25
Lifetime (plasma days)	**0.42**	**0.24**	**1.42**	**1.00**

Acknowledgements

Many thanks are due to Alan Letchford and David Findlay, who diligently proof-read this paper and gave excellent suggestions for improvement. Thanks also go to Scott Lawrie for reading before publication. Thanks also to Jim Alessi, Dan Bollinger and Jens Peters for pictures of their magnetron sources, and to Viktor Klenov for INR Penning source photographs.

References

[1] A.W. Hull, *Phys. Rev.* **18** (1921) 31–57.
[2] J.E. Brittain, *Proc. IEEE* **98** (2010) 635–637.
[3] E.S. Lamar and O. Luhr, *Phys. Rev.* **44** (1933) L947–948.
[4] S.N. Van Voorhis *et al.*, *Phys. Rev.* **45** (1934) L492–493.
[5] L.R. Maxwell, *Rev. Sci. Instrum.* **2** (1931) 129.
[6] F.M. Penning, *Physica* **IV**(2) (1937) 71–75.
[7] B.L. Donnally, *Phys. Rev.* **159** (1967) 87.
[8] K.W. Ehlers *et al.*, *Nucl. Instrum. Methods* **22** (1963) 87–92.
[9] K.W. Ehlers, *Nucl. Instrum. Methods* **32** (1965) 309–316.
[10] G.P. Lawrence *et al.*, *Nucl. Instrum. Methods* **32** (1965) 357–359.
[11] L.E. Collins and R.H. Gobbett, *Nucl. Instrum. Methods* **35** (1965) 277–282.
[12] V.E. Krohn, Jr, *J. Appl. Phys.* **33** (1962) 3523–3525.
[13] Yu.I. Belchenko, G.I. Dimov and V.G. Dudnikov, *Nucl. Fusion* **14** (1974) 113.
[14] V.G. Dudnikov, Proc. 4th All-Union Conf. on Charged Particle Accelerators, Moscow, 1974 (Nauka, Moscow, 1975), vol. 1, pp. 323–325.
[15] K.N. Leung and K.W. Ehlers, *Rev. Sci. Instrum.* **53** (1982) 803.
[16] J.B. Taylor and I. Langmuir, *Phys. Rev.* **51** (1937) 753–760.
[17] J.G. Alessi, in *High Intensity and High Brightness Hadron Beams*, 20th ICFA Int. Beam Dynamics Workshop, AIP Conf. Proc. 642 (American Institute of Physics, Melville, NY, 2002), pp. 279–281.
[18] J. Peters, in *Negative Ions, Beams and Sources*, Proc. 1st Int. Symp., AIP Conf. Proc. 1097 (American Institute of Physics, Melville, NY, 2009), pp. 236–242.
[19] Yu.I. Belchenko *et al.*, in *Negative Ions, Beams and Sources*, Proc. 3rd Int. Symp., AIP Conf. Proc. 1515 (American Institute of Physics, Melville, NY, 2013), pp. 448–455.
[20] H.V. Smith *et al.*, *Rev. Sci. Instrum.* **65** (1994) 123.
[21] D.C. Faircloth *et al.*, in *Negative Ions, Beams and Sources*, Proc. 3rd Int. Symp., AIP Conf. Proc. 1515 (American Institute of Physics, Melville, NY, 2013), pp. 359–368.

Vacuum Arc Ion Sources

I. Brown
Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Abstract
The vacuum arc ion source has evolved into a more or less standard laboratory tool for the production of high-current beams of metal ions, and is now used in a number of different embodiments at many laboratories around the world. Applications include primarily ion implantation for material surface modification research, and good performance has been obtained for the injection of high-current beams of heavy-metal ions, in particular uranium, into particle accelerators. As the use of the source has grown, so also have the operational characteristics been improved in a variety of different ways. Here we review the principles, design, and performance of vacuum arc ion sources.

1 Introduction

The defining characteristic of the vacuum arc ion source is its production of *high-current* beams of *metal* ions. The metal plasma from which the ion beam is formed is produced by a vacuum arc plasma discharge, and hence the name. Vacuum arc ion sources have been used primarily for ion implantation in material surface modification research, and also for particle accelerator injection, as well as for other fundamental and applied purposes. Beams have been produced from over 50 of the solid metals of the Periodic Table (i.e., almost all of them) and carbon, with extracted ion energy up to several hundred kiloelectronvolts and beam current up to several amperes. The source is usually operated in a repetitively pulsed mode, with pulse length typically a few hundred microseconds and repetition rate up to a few tens of pulses per second; operation of a simplified d.c. embodiment has been demonstrated. In general, the ions have low but multiply ionized charge states. The mean charge state lies between 1+ and 3+ depending on the particular metal, and the charge states can be increased by up to a factor of two in a number of different ways.

The history of vacuum arc ion sources extends back to the 1940s, when attempts to use this approach for isotope separation were first made in the USA as part of the Manhattan Project during World War II [1]; the effort was abandoned, however, for reasons having to do with arc instability. Research in the general area of vacuum arcs has a long and rich history in the Former Soviet Union (FSU) [2], where the development of vacuum arc-based ion sources began in the late 1950s and early 1960s, led primarily by Plyutto and co-workers at the Sukhumy Physical Technical Institute in Georgia [3]. This early work was largely a demonstration of the basic concept of metal ion extraction from vacuum arc plasma. In the mid-1960s a kind of vacuum arc ion source was developed by workers in Ukraine for the production of beryllium ion beams with current up to 170 mA [4]. In 1979 Prewett and Holmes developed a carbon vacuum arc ion source at the University of Liverpool, UK, to produce a low-energy C^+ ion beam current of up to 0.5 A [5]. In the early 1980s Humphries and co-workers at the University of New Mexico, USA, embarked on an extensive investigation of the vacuum arc for production of ion beams for heavy-ion fusion research application [6–8]. A vacuum arc ion source development programme was initiated at the Lawrence Berkeley National Laboratory (LBNL) in 1982 for the production of high-current uranium ion beams for injection into the LBNL heavy-ion synchrotron (the Bevalac) for fundamental heavy-ion nuclear physics research, and later for ion implantation application. The LBNL sources, called *Mevva* (metal vapour vacuum arc) ion

sources, were developed in a number of different directions, including embodiments with multiple-cathode assemblies, very large diameter extractors, miniature sources, and a test d.c. version [9, 10]. At virtually the same time, development of vacuum arc ion sources was initiated at the High Current Electronics Institute (HCEI) of the Russian Academy of Sciences, Tomsk. These sources (called *Diana* [11] and *Titan* [12]) played a significant role in the rapid growth of activity in the field throughout the FSU. A series of vacuum arc ion sources called *Raduga* were developed at the Nuclear Physics Institute (NPI) of the Tomsk Polytechnic University [13, 14], and the *Tamek* sources were developed at the Tomsk Institute of Automatic Control Systems and then at the Applied Physics Institute, Sumy, Ukraine [15]. Programmes were subsequently established in other world universities, institutes, and laboratories. A historical review of the early development of vacuum arc ion sources has been given elsewhere [16].

Vacuum arc ion source performance has been improved in a number of important ways over the years. One critical beam parameter, for example, is the ion charge state spectrum; for many purposes there is a need for higher charge states than normally formed in the vacuum arc plasma. Following this need, several different approaches to increasing the ion charge state have been developed. Beam noise and pulse-to-pulse reproducibility are also very important for particle accelerator application, and these characteristics have been improved vastly in work carried out at the Gesellschaft für Schwerionenforschung (GSI), Darmstadt, Germany. The vacuum arc ion source is now routinely used for high-current metal ion injection into the GSI heavy-ion accelerators. Reliable and long-lifetime arc triggering is another important source feature, and good progress has been made in this direction also.

Here the source performance and beam characteristics are summarized, typical design features outlined, and some of the facilities that have been established at several laboratories around the world are described.

2 Plasma physics of the vacuum arc discharge

Reviews of vacuum arc plasma discharges have been given by a number of authors [17–19]. The vacuum arc is a high-current discharge between two electrodes in vacuum. At the cathode the current is concentrated at a small number of tiny, discrete sites, called cathode spots. The formation of cathode spots is a fundamental characteristic of the vacuum arc discharge. The spots are where the metal plasma is produced, and it is this plasma that provides the current path between cathode and anode that keeps the arc alive. Thus some of the plasma that is generated at the cathode must necessarily deposit on the anode so as to form the cathode-to-anode current path, and some of the metal plasma can be taken away, using suitable geometry, and used for another purpose – for example, to form the ion beam of an ion source. At a cathode spot the cathode material is heated, vaporized, and ionized into the plasma state. The spot has a diameter of the order of a micrometre, and the current density is extremely high, of the order of 10^6–10^8 A cm^{-2}. The arc current is constricted to a small number of such spots. Most of the parameters of the vacuum arc plasma are determined by the plasma physics within the cathode spot. Individual spots move around on the cathode surface, and the lifetime of a particular spot may be only microseconds or less; on the other hand, small surface irregularities like edges or protuberances tend to anchor the spots to these sites. The plasma pressure within a cathode spot is high, and the strong pressure gradient causes the plasma generated there to plume away from the surface in a manner quite similar to the plasma plume generated by an intense focused laser beam at a solid surface. The current carried by a cathode spot is typically about a few to a few tens of amperes, depending on the metal, and if the arc is caused to conduct a higher total current, then more cathode spots are formed. Thus a typical metal vapour arc discharge of several hundred amperes current might involve the participation of several tens of cathode spots. The assemblage of cathode spots gives rise to a dense plasma of cathode material that streams away from the cathode as a jet. The ion streaming velocity, or drift speed, is typically 1–3 cm μs^{-1}.

Ambient gas is not essential to the discharge. Much vacuum arc research has been carried out in the 10^{-6} Torr range, and ~10^{-4} Torr might be considered a rough upper limit. Residual gas can have a significant influence on the plasma parameters, especially on the metal ion charge state distribution [20].

The arc current can be anywhere in the range from several tens of amperes up to many kiloamperes, and for most vacuum arc ion sources the current is typically one hundred to several hundred amperes. The arc voltage (when the arc is alight, i.e., the *burning voltage*) lies in the range 10–30 V, say typically about 20 V, and varies with the cathode metal used. Note that the arc current cannot be decreased to arbitrarily low value; the cathode spot requires a certain minimum current to stay alight, typically a few to a few tens of amperes depending on the metal species, below which the spot extinguishes. This has significance for the minimum power level at which the vacuum arc can be operated and thus on the concern of a d.c. (continuous, or steady-state) vacuum arc ion source. Essentially all vacuum arc ion sources are operated in a repetitively pulsed mode, although a d.c. version has been demonstrated.

Conduction of the arc current is supported by the metal plasma that is evolved from the solid electrode (cathode) material itself. In most vacuum arc ion sources, triggering of the vacuum arc has been accomplished by initiation of a surface flashover across a thin insulating (ceramic) surface between the cathode and an annular trigger electrode, and other methods have also been used [21].

A well-established basic feature of the vacuum arc is the relationship between the metal ion flux that is generated and the current that drives the arc. The arc current is composed of both an electron component and an ion component. The plasma ion current is the plasma flux generated at the cathode. Over a wide range of conditions the plasma ion current is a constant fraction of the arc current,

$$I_{\text{ion}} = \varepsilon I_{\text{arc}}, \tag{1}$$

where ε is in the range 0.06–0.12, say typically about 10%. Thus the electrical efficiency of the vacuum arc discharge – the ratio of total generated metal ion plasma current to arc current – is high, about 10%. The electrical efficiency of the vacuum arc ion source can also be high, although the steps (total metal plasma generated by the arc) to (plasma transported to the beam formation electrodes) to (formation of ion beam from plasma presented to the extractor) to (ion beam delivered to a downstream target or beam-line) conspire together to yield an overall electrical efficiency that is considerably less than the fundamental vacuum arc discharge efficiency of 10%.

The relationship between the mass evolved from the cathode and the arc parameters has also been investigated. Generalizing, one can say that the mass of the plasma generated by the vacuum arc is of the order of several tens of micrograms per coulomb of arc current, where the precise value depends primarily on the metal used. At the cathode spots, cathode material is converted into metal plasma and also into solid *macroparticles* (so called because they are macroscopic compared to plasma particles, the ions). Macroparticles are metallic globules that are ejected from the cathode in the molten state and then rapidly solidify; they typically have a diameter in the range 0.1–10 μm. For many cases (e.g., for high-melting-point materials), the macroparticle content of the ion beam is small and is not a concern.

3 The basic vacuum arc ion source

The basic operating principles of the vacuum arc metal ion source can be described using the example of the LBNL Mevva II ion source [10], a simplified schematic of which is shown in Fig. 1 and its associated electrical schematic in Fig. 2. As for any plasma-based ion source, there are two basic parts – the plasma generator, and the beam extraction system. The cathode of the vacuum arc is a simple cylindrical rod (typically 5–10 mm diameter) of the material of interest, and the discharge is triggered by a high-voltage pulse applied to a trigger electrode that surrounds the cathode coaxially, separated

from it by a thin (about 0.5 mm) alumina insulator. A part of the metal plasma created at the cathode flows through the anode hole, of diameter about 1 cm, and then through a drift space of several centimetres to the extractor grids. That part of the plasma that strikes the anode carries the current that keeps the plasma alight, while the plasma plume that streams through the central hole in the anode is used to form the ion beam.

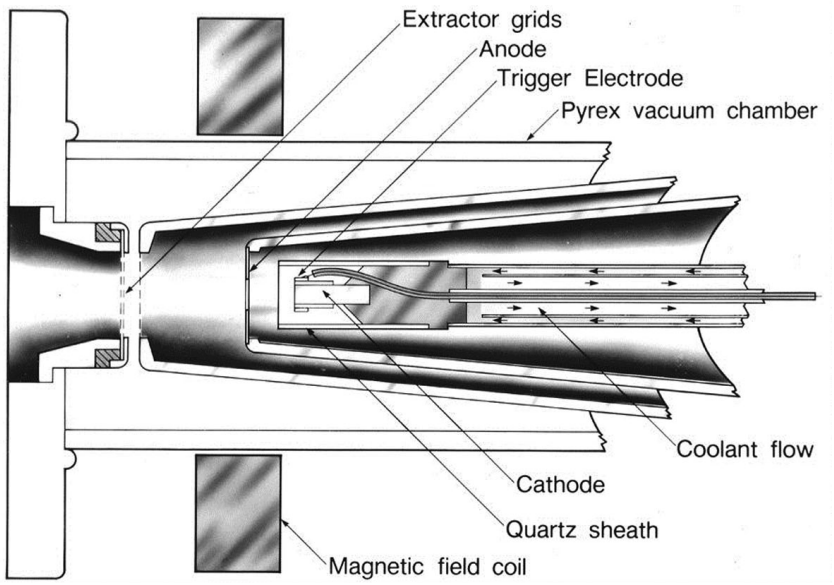


Fig. 1: Simplified schematic of the LBNL Mevva II ion source

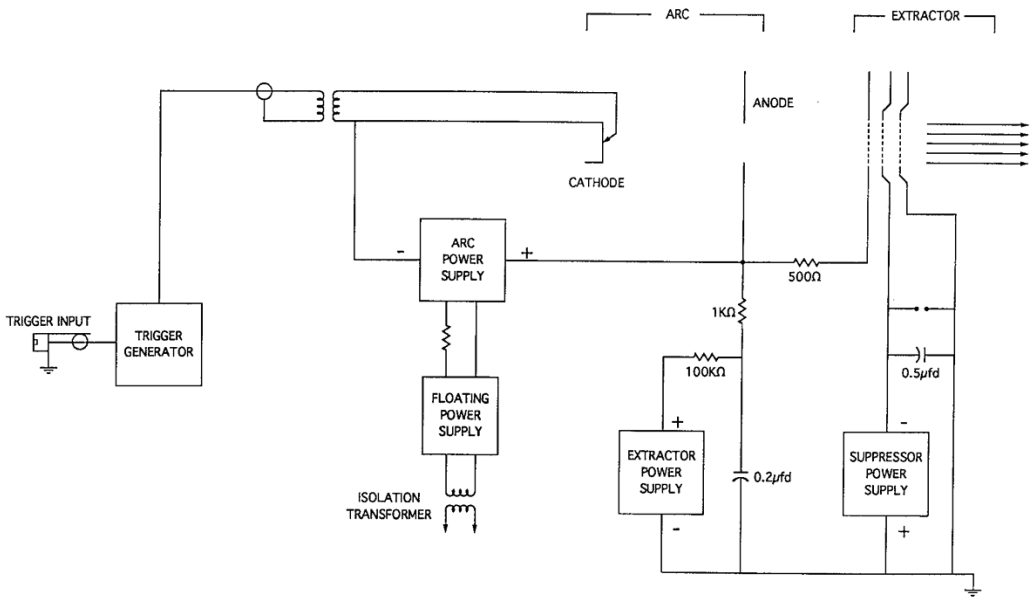


Fig. 2: Simplified electrical schematic of a vacuum arc ion source set-up

The kind of beam formation electrodes (*extractor grids*) used can be chosen to suit the particular application, and a conventional multi-aperture, accel–decel configuration is simple and effective. A set of three circular grids is located appropriately with respect to the expanding plasma plume, with the first grid (nearest the plasma) (*plasma grid*) at or near plasma potential, the middle grid (*suppressor grid*) typically several kilovolts negative with respect to ground so as to suppress the backflow of low-energy electrons, and the third grid (*ground grid*) at ground potential. The first grid and the entire plasma generation part of the source are biased to high positive potential. The magnetic field coil shown in Fig. 1 is optional and is often omitted. A modest longitudinal magnetic field in the arc region can be used to control the amount of plasma that is transported from the cathode to the extractor. In the absence of magnetic field, much of the plasma is lost to the walls. With an applied field of just a few hundred gauss, the radial plasma loss can be much reduced, and a greater fraction of the plasma that is formed can be presented to the extractor and converted into ion beam.

Usually the source is operated in a repetitively pulsed mode, with pulse width 100–500 μs and repetition rate up to several tens of pulses per second. The arc current is typically in the range of 50–500 A, often conveniently supplied by an *LC* (inductance-capacitance) pulse line. The trigger pulse is usually of O/C amplitude ~10 kV and duration ~10 μs or more, supplying a peak current of a few tens of amperes trigger-to-cathode by a step-up transformer, which also serves for high-voltage isolation. The ion beam extraction voltage in typical operation is several tens of kilovolts up to a maximum of about 100 kV and might be either d.c. or pulsed (d.c. in the schematic shown).

The extracted metal ion beam current can easily be as high as several amperes, and for very broad beam sources up to several tens of amperes [10], but for usual arc parameters and extractor sizes (initial beam diameter about 1–10 cm) a typical beam current might be 100–500 mA.

Beam divergence is determined primarily by the extraction optics, and if the extractor grids have been well designed and fabricated, the extraction optics can be empirically matched to the plasma by variation of plasma density via the arc current. For optimum extraction conditions (the *perveance match* condition), the beam divergence is typically about 3° half-angle for uranium.

The vacuum arc ion source produces ions that are multiply ionized, and the charge state spectrum of the ion beam is important for most applications. The ions generated in the vacuum arc plasma are in general multiply stripped, with a mean charge state of 1+ to 3+, depending on the particular metal species, and the charge state distribution can have components from Q = 1+ to 6+; thus the ion energy is greater than the extraction voltage by this same factor.

4 Source performance

The source performance is determined in part by the specific source embodiment used, but one can nevertheless generalize, referring to specific source embodiments for specific experimental results. The operation and performance of the LBNL Mevva IV source have been described in detail [10, 22–24]. Summarized here are beam parameters that are typical of vacuum arc ion sources. Note that in all cases the source is designed and operated repetitively pulsed, not d.c. Usually the beam pulse width is a few hundred microseconds and the repetition rate, say, 10 pps. Thus this kind of ion source is suited to synchrotron injection but not cyclotron.

4.1 Beam current

Depending on the extractor size and thus the fraction of plasma presented for beam formation, the ion beam current can easily be as high as several amperes, and for very broad beam sources up to several tens of amperes [10]. The fraction of this beam that is usable, however – for example, for injection into an accelerator low-energy beam-line or for implantation of a downstream target – depends on the particular set-up. The beam transported downstream depends on the beam emittance and the target (or

beam-line) acceptance as well as losses in the beam transport system. For example, a 10 cm extractor source version (Mevva V) has produced a titanium ion beam current into a nearby large-area Faraday cup of up to 3.5 A at 90 kV extraction voltage [10]. In another series of measurements, a source embodiment (Mevva IV) with a 2 cm diameter extractor was used to characterize the beam extraction, as shown in Figs. 3 and 4. Figure 3 shows the Ti beam current as a function of arc current for a range of extraction voltages, and Fig. 4 shows the Ta beam current as a function of extraction voltage for a range of arc currents. Measurements of this kind have been made for a number of different cathode materials (ion beam species), and good agreement was found with the well-known Child–Langmuir prediction for extracted ion current under space-charge-limited conditions,

$$I = \tfrac{4}{9}\varepsilon_o S\left(\frac{2eQ_i}{M_i}\right)^{\frac{1}{2}}\frac{V^{\frac{3}{2}}}{d^2} = 1.72\,S\left(\frac{Q_i}{A}\right)^{\frac{1}{2}}\frac{V^{\frac{3}{2}}}{d^2} \qquad (2)$$

where S is the extractor open area, Q_i the ion charge state (ion charge $q = eQ$), $M_i = Am_u$ the ion mass (where A is the atomic weight in amu and m_u the mass of 1 amu), V the extractor voltage, and d the extractor gap. In the second expression, I is in mA, V in kV, S in cm^2, and d in cm. The implication is that the ion beam current that can be expected from any embodiment of vacuum arc ion source can be predicted with reasonable accuracy from the Child–Langmuir equation above.

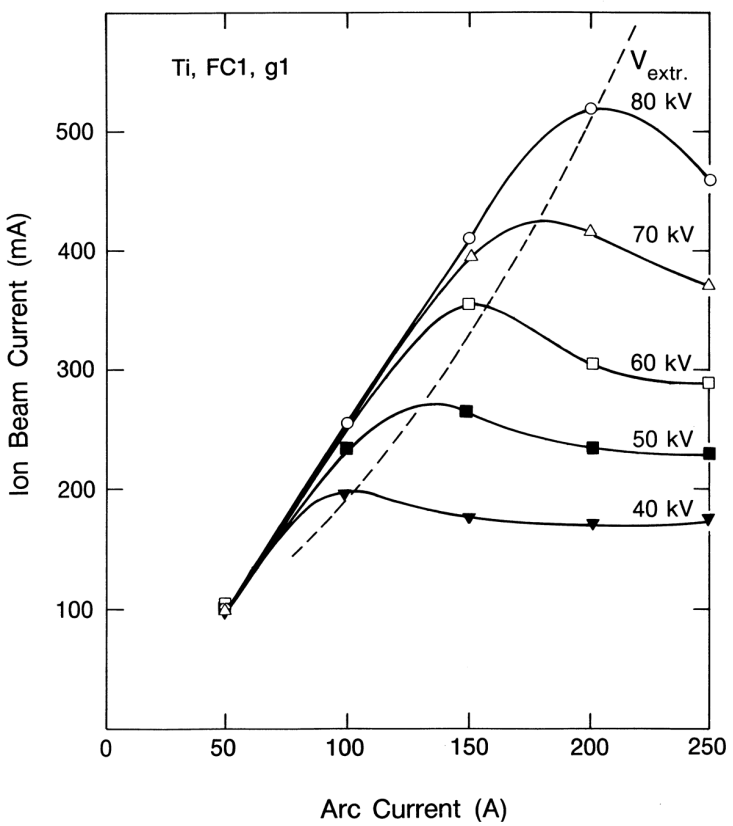


Fig. 3: Ion beam current as a function of arc current; titanium beam (LBNL Mevva IV)

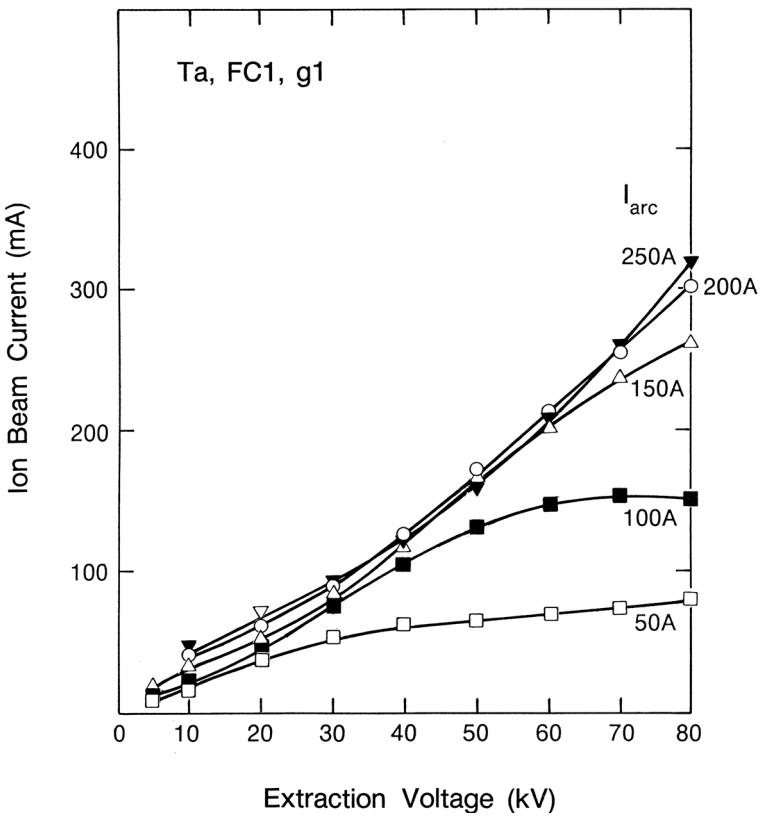


Fig. 4: Ion beam current as a function of extractor voltage; tantalum beam (LBNL Mevva IV)

A source embodiment has been demonstrated [10] incorporating a 50 cm diameter extractor. This source has formed Ti ion beams at 50 kV extraction voltage, i.e., about 100 keV mean ion energy (the mean charge state for Ti is 2.1+), and with peak total extracted ion current over 20 A.

4.2 Beam profile, divergence, and emittance

Beam divergence is determined primarily by the extraction optics, and, assuming that the extractor grids have been designed and fabricated reasonably, the extraction optics can be empirically matched by variation of the plasma density via the arc current. At optimum (the *perveance match* condition), the beam divergence is typically about 3° half-angle. The normalized emittance, ε_N, is about 0.3 π mm mrad (normalized) for uranium at perveance match. The ion temperature in the vacuum arc plasma is quite generally about 1 eV, and the plasma-temperature-determined beam divergence half-angle $\theta = v_\perp/v_\parallel = \sqrt{(T_{i\perp}/E_i)} = 0.2°$ for E_i = 100 keV. Thus the measured beam divergence, or emittance, is indeed determined mostly by the extraction and not by the transverse plasma temperature $T_{i\perp}$.

The initial beam shape, i.e., the beam radial profile immediately after extraction, is determined by the radial plasma density distribution at the extractor location together with the extraction optics. If the plasma density is not uniform across the extractor, then the extraction optics cannot be matched everywhere. An open-ended multipole magnetic bucket configured of rare-earth magnets and located in the ion source drift region can be used to flatten the plasma profile at the extractor [25], but, even so, the beam profile loses memory of its shape at the extractor fairly rapidly, and after some tens of centimetres downstream propagation the beam profile reverts to the usual Gaussian [26]. Experimentally, control of the arc current provides a convenient means of optimizing the extraction, via the plasma density. Alternatively, the extraction optics can be tailored by the extractor voltage for a fixed arc current (plasma density).

4.3 Beam composition

The vacuum arc ion source produces ions that are multiply ionized, and the charge state spectrum of the ion beam is important for most vacuum arc ion source applications. Ion charge state distributions (CSD) have been studied in detail experimentally [27], using time-of-flight (TOF) charge state diagnostics [28], and theoretically [29]. Almost all of the solid metals of the Periodic Table have been successfully used in vacuum arc ion sources, as well as cathodes made from metallic alloys, compounds, and pressed mixtures. Compound cathodes produce ions of the cathode constituents, and it is interesting to note that beams containing non-metallic elements, such as B and S, can be made by using conducting compound electrodes of which the non-metal is a constituent such as LaB_6 or PbS. An example of the charge state spectra obtained is shown in Fig. 5, where an oscillogram of an iridium TOF spectrum is shown (current measured by a magnetically suppressed Faraday cup). Note that for multiply charged ions, electrical current i_{elec} is not the same as particle current i_{part}, since each particle can carry multiples, Q, of the electronic charge, e: $i_{elec} = Qi_{part}$. This can be important, as, for example, in the case of the beam current measured by a Faraday cup being electrical current, while the current needed for estimating ion implantation dose is particle current.

The mean charge state of the ion charge state distribution in general lies in the range from 1+ to 3+, depending on the particular metal species, and the distribution can have components from Q = 1+ to 6+; thus the ion energy is greater than the extraction voltage by this same factor. The measured charge state distributions and mean charge states for a wide range of elemental species, under typical vacuum arc operational parameters and in the absence of techniques to enhance the ion charge state, are given in Table 1. More detail is given elsewhere [27].

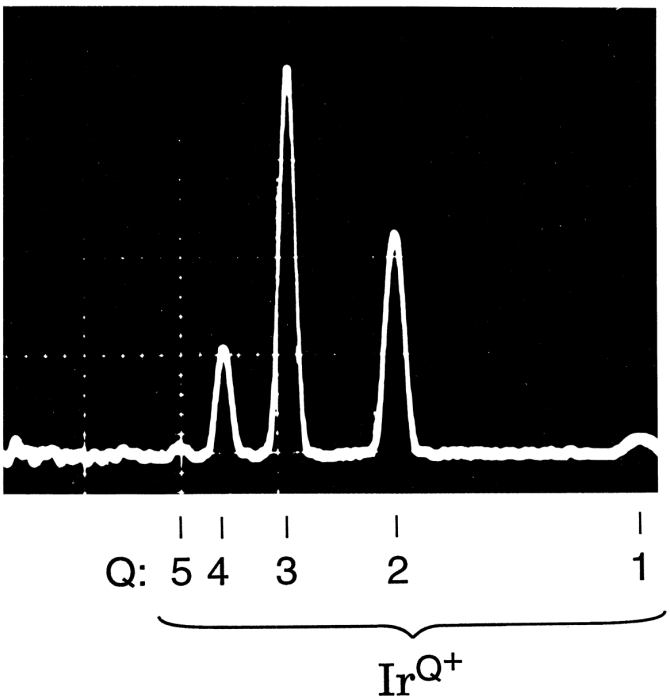


Fig. 5: Time-of-flight charge state distribution for an iridium ion beam. The amplitudes are electrical current measured by a Faraday cup.

4.4 Beam noise and pulse shape reproducibility

The basic mechanism of vacuum arc plasma formation is via explosive emission from microscopic sites on the cathode surface [30], leading to a relatively high fluctuation level in plasma density and

thus also in extracted ion beam current. That is, the vacuum arc ion source beam is inherently a relatively noisy beam, and the pulse-to-pulse reproducibility is not high. While for some applications, such as broad-beam ion implantation for material surface modification, this is not a concern, for other applications, such as for particle accelerator injection, it can be a severe drawback, and for this purpose it is highly desirable to develop methods for forming a quieter beam and improved pulse shape reproducibility.

Table 1: Ion charge state fractions and mean charge states, expressed in terms of particle current. Asterisks (*) indicate a trace (under 1%).

Element	Z	Charge state						$\langle Q \rangle$
		1+	2+	3+	4+	5+	6+	
Li	3	100						1.0
C	6	100						1.0
Mg	12	46	54					1.5
Al	13	38	51	11				1.7
Si	14	63	35	2				1.4
Ca	20	8	91	1				1.9
Sc	21	27	67	6				1.8
Ti	22	11	75	14				2.1
V	23	8	71	20	1			2.1
Cr	24	10	68	21	1			2.1
Mn	25	49	50	1				1.5
Fe	26	25	68	7				1.8
Co	27	34	59	7				1.7
Ni	28	30	64	6				1.8
Cu	29	16	63	20	1			2.0
Zn	30	80	20					1.2
Ge	32	60	40	*				1.4
Sr	38	2	98					2.0
Y	39	5	62	33				2.3
Zr	40	1	47	45	7			2.6
Nb	41	1	24	51	22	2		3.0
Mo	42	2	21	49	25	3		3.1
Rh	45	46	43	10	1			1.7
Pd	46	23	67	9	1			1.9
Ag	47	13	61	25	1			2.1
Cd	48	68	32					1.3
In	49	66	34	*				1.4
Sn	50	47	53					1.5
Sb	51	1	*					1.0
Ba	56		100					2.0
La	57	1	76	23				2.2
Ce	58	3	83	14				2.1
Pr	59	3	69	28				2.2
Nd	60		83	17				2.2
Sm	62	2	83	15				2.1
Gd	64	2	76	22				2.2
Dy	66	2	66	32				2.3
Ho	67	2	66	32	*			2.3
Er	68	1	63	35	1			2.4
Tm	69	13	78	9				2.0
Yb	70	3	88	8				2.1
Hf	72	3	24	51	21	1		2.9
Ta	73	2	33	38	24	3		2.9
W	74	2	23	43	26	5	1	3.1
Ir	77	5	37	46	11	1		2.7
Pt	78	12	69	18	1			2.1
Au	79	14	75	11				2.0
Pb	82	36	64					1.6
Bi	83	83	17					1.2
Th	90		24	64	12			2.9
U	92	20	40	32	8			2.3

Minimum beam noise is found empirically [31, 32] to occur when the extraction voltage and plasma density are optimally matched – the perveance match condition for optimum beam formation.

(By beam noise we mean the fractional fluctuation level of ion beam current about the mean beam current – $\delta i_b/i_b$.) Then the r.m.s. noise is a minimum of about 7%. Increased gas pressure in the discharge gap and in the extraction region improve the beam quality, but this method shifts the CSD to lower charge states. There is also a significant influence of magnetic field on beam noise and pulse stability.

Humphries and co-workers developed a technique for reduction in ion beam noise level by the use of wire meshes to reflect plasma electrons and to limit ion flow by its own space charge [33]. In the case of space-charge-limited flow, the extracted ion beam current is independent of the plasma fluctuation level and random plasma generation at the cathode spots. The method has been further developed specifically for accelerator injection application [34, 35].

Work carried out at GSI Darmstadt, Germany, over a period of several years has resulted in a vast improvement in both the beam noise and the beam pulse shape reproducibility [36]. By combining a number of techniques, including the use of meshes, addition of an optimized magnetic field, and geometric optimization, the performance of the GSI vacuum arc ion sources has been brought to the point where they are fully acceptable for routine accelerator injection application; the source is described below in section 5.2. The good performance of this source with respect to beam noise is evident from Fig. 6. Figure 6(a) shows the pulse shape of a uranium ion beam, pre-analysis, with all charge states, as measured by a Faraday cup 30 cm downstream: current is 156 mA and beam noise is ±4%. Figure 6(b) shows the post-analysis U^{4+} beam as measured by a beam transformer at the entrance to the pre-linac radio-frequency quadrupole (RFQ) accelerator, a point about 12 m downstream from the ion source: U^{4+} beam current is about 25 mA and beam noise is ±5%. The pulse shape, beam noise, and reproducibility are adequate for the source to be used routinely at GSI for high-current injection of metal ions, especially U^{4+}, into the UNILAC and SIS accelerators.

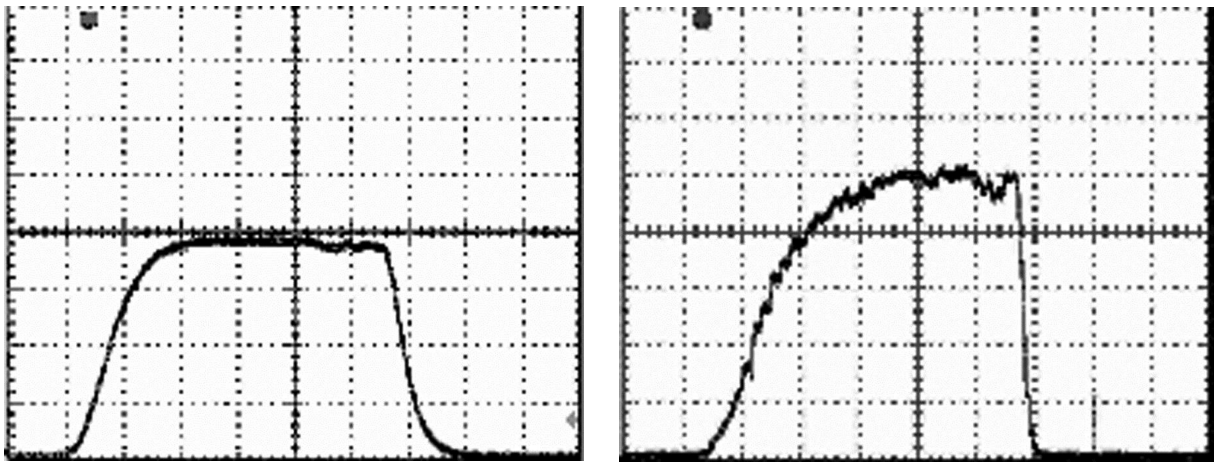


Fig. 6: (a) Uranium ion beam (all charge states) current pulse from the GSI vacuum arc source measured 30 cm from the source extractor; 40 mA/div. (b) Post-analysis U^{4+} beam at the RFQ entrance, a distance of 12 m from the source; 5 mA/div. Sweep speed is 100 μs/div.

4.5 Gaseous operation

The vacuum arc ion source is above all a *metal* ion source. However, it can be operated so as to form beams that are controllable mixtures of gas and metal ions. Gas is fed into the source near the arc region where the plasma density is high, and a modest magnetic field (B ~ 200–400 G) is applied. The gas is ionized and gaseous ions are mixed with the metal plasma and extracted ion beam. The fraction of gaseous ions is regulated by the neutral gas feed, and the gas-to-metal ion ratio can be controlled

from zero up to as much as about 99% gaseous. An example of this effect is shown in Fig. 7 for a titanium ion beam with added nitrogen. In this case a magnetic field was added to the arc region, and thus the low-pressure Ti charge states are somewhat elevated over the charge state distribution that is usual for the case of zero *B*. As the pressure is increased, there is a slow downshift of Ti states, and a simultaneous increase in the fraction of N ions in the beam. Two conclusions follow from these kinds of observations: (i) high fractions of highly charged ions call for the lowest possible background gas pressure; and (ii) ambient gas pressure provides a means for controlling the ion charge state distribution towards lower states. These effects can be used to advantage. For example, hybrid Ti–N ion beams have application for implantation of subsurface TiN layers, and similar buried ceramic layers can be formed from other species such as Al–O, Zr–O, etc. [37]. In work related to accelerator injection, a vacuum arc ion source was used to provide Mg^{+} ions for injection into the GSI heavy-ion accelerator, having increased the ion source gas pressure so as to maximize the fraction of singly charged ions with respect to the doubly charged Mg^{2+} fraction [38].

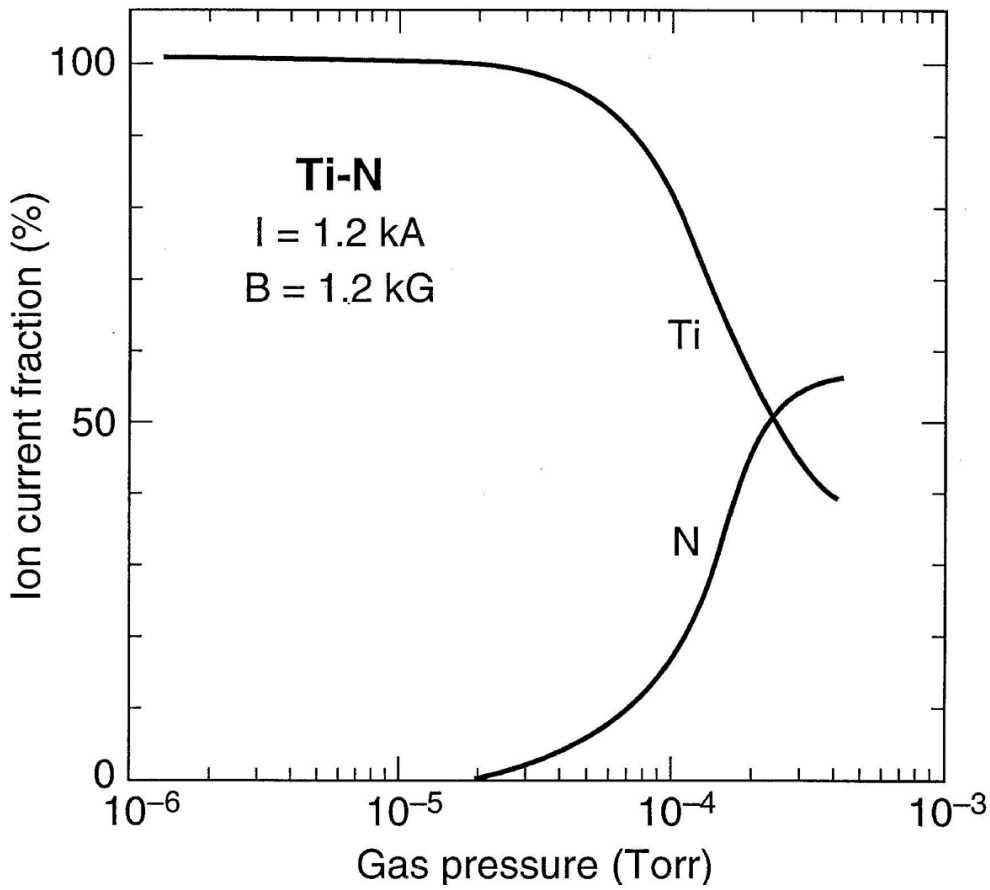


Fig. 7: Metal and gas ion beam composition fraction as a function of nitrogen gas pressure, for a hybrid Ti–N ion beam formed using the LBNL Mevva V ion source.

A way of operating a vacuum arc source in a 100% gaseous mode has been reported [39]. In this case, the vacuum arc feature is not used at all, but instead a hollow-cathode glow discharge is formed in the same ion source geometry.

There is a small gaseous contamination of the metal ion beam even at the lowest pressures at which the vacuum arc ion source is normally operated (~10^{-6} Torr). The primary origin of this is condensation of gas (oxygen, nitrogen, water vapour) on the fresh front surface of the cathode from

the residual ambient between arc pulses. This also affects, to some extent, the metal ion charge state distribution. The magnitude of the effect depends on the time between pulses, and can be minimized by operating at high repetition rate, greater than about 10 pps [40].

5 Source embodiments

We briefly describe here some of the vacuum arc ion sources that have been made and used at various laboratories. A comprehensive survey of all such sources is not attempted or implied. We distinguish the sources by the main application to which the work was directed.

5.1 Ion implantation application

The vacuum arc ion source has been adopted widely for metal ion implantation application, since the source can readily provide the high-current beams necessary for high-dose implantation for material surface modification. Vacuum arc ion implantation facilities have been set up at LBNL, Berkeley, USA [9, 10, 22, 23, 41–43], HCEI, Tomsk, Russia [11, 12, 44], NPI, Tomsk, Russia [14, 45, 46], Izmir, Turkey [47], the Australian Nuclear Science and Technology Organization (ANSTO), Sydney, Australia [48], Nagasaki, Japan [49], and most impressively at Beijing Normal University, China [50–52], and elsewhere.

The LBNL Mevva V ion source has operated steadily for many years with infrequent need for maintenance or repair. It remains a reliable workhorse and has served as a basic model for vacuum arc ion sources elsewhere. A schematic of the source is shown in Fig. 8 and a photograph in Fig. 9. It uses a 10 cm diameter, multi-aperture, accel–decel extractor configuration, incorporates a rotatable 18-cathode assembly, and can operate at up to about 65 kV extraction voltage or more (record high was 110 kV).

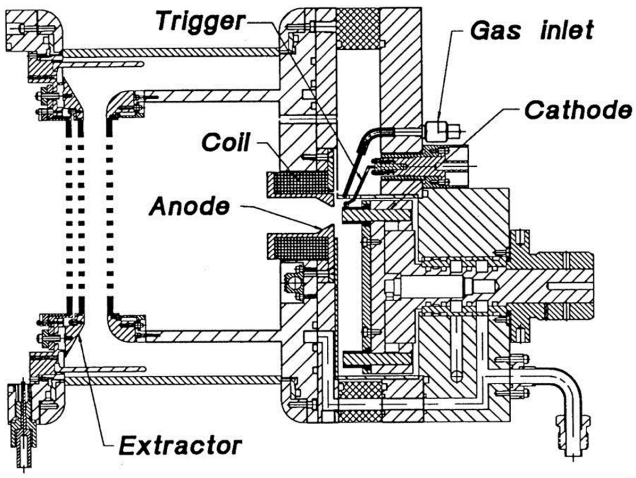


Fig. 8: LBNL Mevva V source, fitted with pulsed solenoid surrounding the arc region and with gas inlet feed

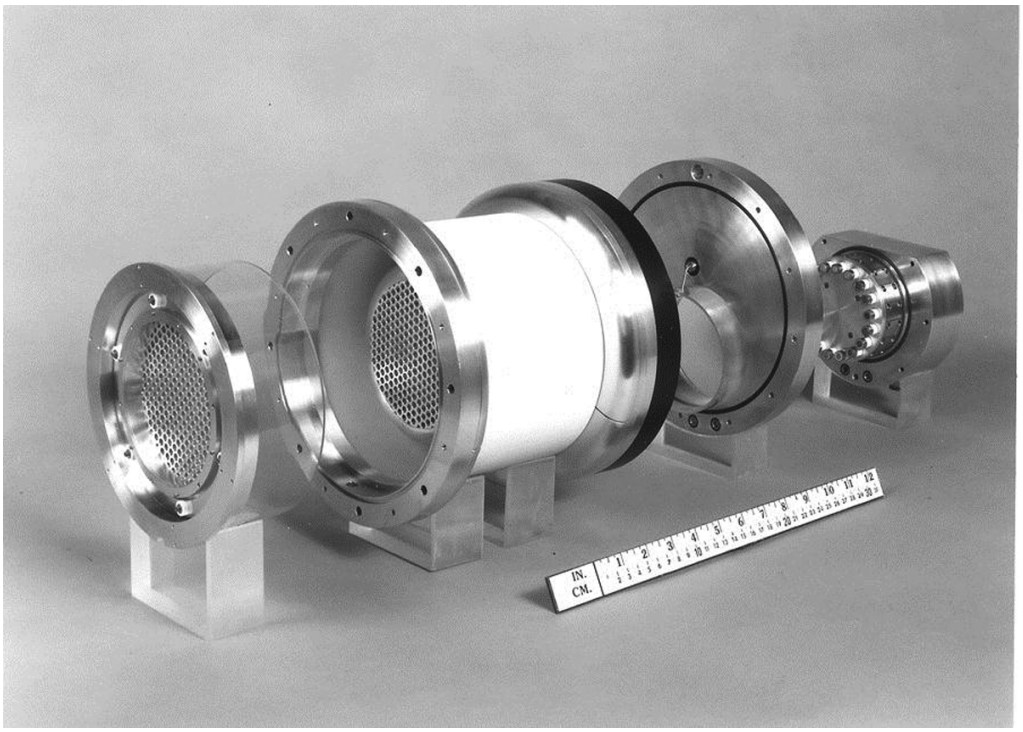


Fig. 9: LBNL Mevva V source, partly disassembled showing the 10 cm extractor grids on the left and the multiple cathode assembly on the right.

Beijing Normal University has a substantial vacuum arc ion source programme, initiated in 1987 with support from the National High Technology Research and Development Programme of China. Since then, more than 10 different embodiments of vacuum arc ion source-based implanters have been developed and deployed in factories and companies in China, Hong Kong, and Taiwan, with increasingly impressive ion source and implantation parameters. Among these, the largest presently in use is the Mevva 50 implanter, with a time-averaged ion beam current of 50 mA, installed at the Advanced Materials Center in the southern China city of Shenzhen. An even larger implanter has been made; this is the Mevva 100 implanter, with time-averaged ion beam current of 100 mA. This metal ion implantation facility is the largest vacuum arc metal ion implanter for industrial surface processing in the world.

5.2 Accelerator injection application

Vacuum arc ion sources have been used successfully at the Institute for Theoretical and Experimental Physics (ITEP), Moscow [53, 54] and at the GSI heavy-ion accelerator research centre at Darmstadt, Germany, for injecting metal ions into a linear accelerator and synchrotron, including uranium. At ITEP, beams of a range of metals were formed, including Be, C, Al, Fe, Cu, Zr, and W at beam current of up to 400 mA. For a pulse length of 100 μs and pulse repetition frequency 0.3 pps, operation times of up to 300–400 hours were obtained between necessary down-times for source maintenance (cathode replacement and other).

Much of the GSI work was performed collaboratively with the group at HCEI, Tomsk. For accelerator application, the quality of each individual ion beam pulse is important, as opposed to the case for ion implantation application. Beam noise and/or pulse shape variation from pulse to pulse can lead to poor accelerator performance. Thus the GSI source has incorporated a number of developments to enhance the high ion charge state fraction, to reduce the beam noise, and to improve the pulse-to-pulse beam shape reproducibility [34–36].

A schematic of the GSI source is shown in Fig. 10 and a photograph in Fig. 11. The cathode assembly holds 17 individual cathodes, each 17 mm long and 5.7 mm in diameter. The stainless-steel anode is located a distance of 15 mm from the cathode and has a central aperture 15 mm in diameter. Magnetic field coils are located external to the vacuum chamber. Two stainless-steel grids are installed to reduce the beam fluctuation level and to reduce plasma density. A multi-aperture accel–decel extraction system (13 holes, each 3 mm in diameter, aspect ratio 0.5) is used to form the ion beam. As an example of source performance, 25 mA of U^{4+} ion current with an energy of 2.2 keV/u (156 mA full beam, or 170 mA cm^{-2}, electrical current) was measured at the entrance of the RFQ, for a typical extraction voltage of 32 kV, with a high fraction (67% electrical current fraction) of U^{4+} ions, a post-analysis U^{4+} beam noise level of about ±5%, and good beam pulse shape reproducibility. Examples of the total extracted uranium beam and the separated U^{4+} beam pulse shapes are shown above in Fig. 6, and the main source and beam parameters are summarized in Table 2.

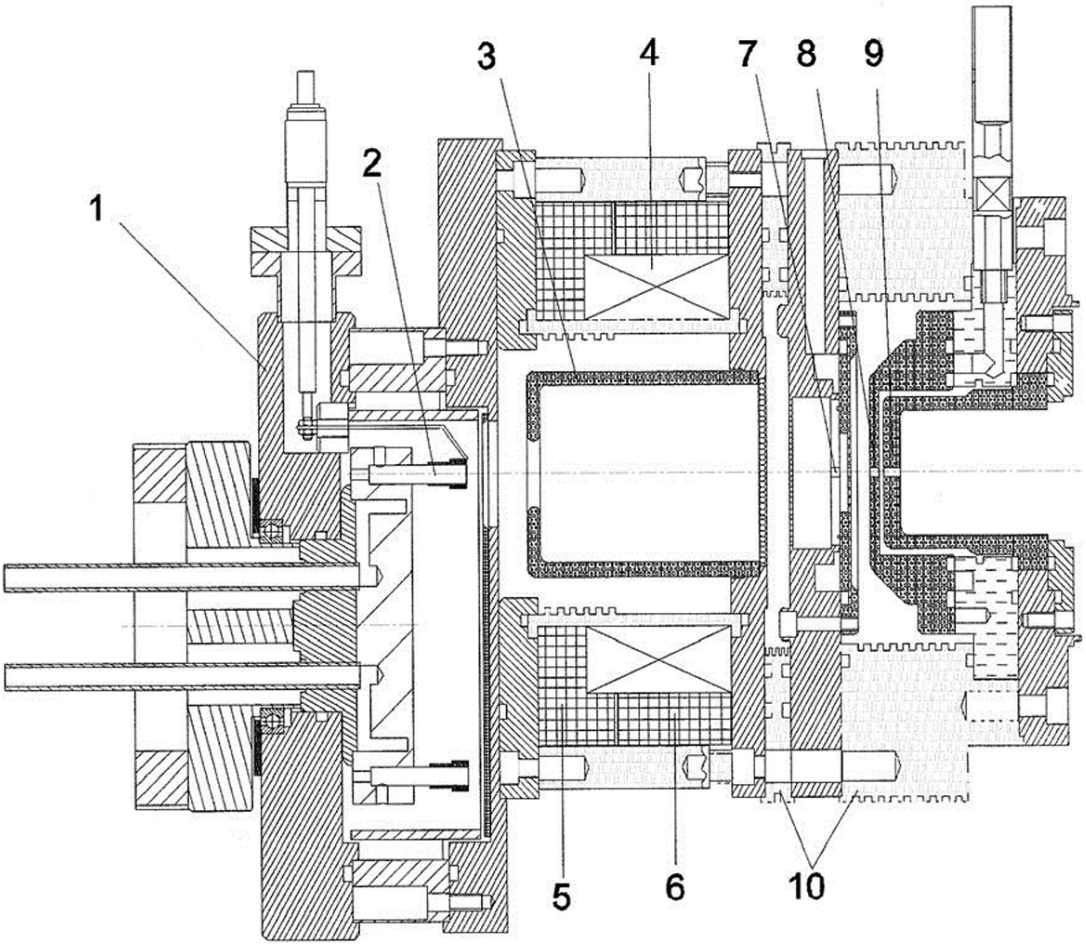


Fig. 10: GSI vacuum arc ion source: 1, cathode flange; 2, cathode; 3, anode (stainless steel); 4, SmCo cusp magnets (10 in total); 5, coil I; 6, coil II; 7, plasma electrode and grid; 8, screening electrode; 9, ground electrode; 10, insulators.

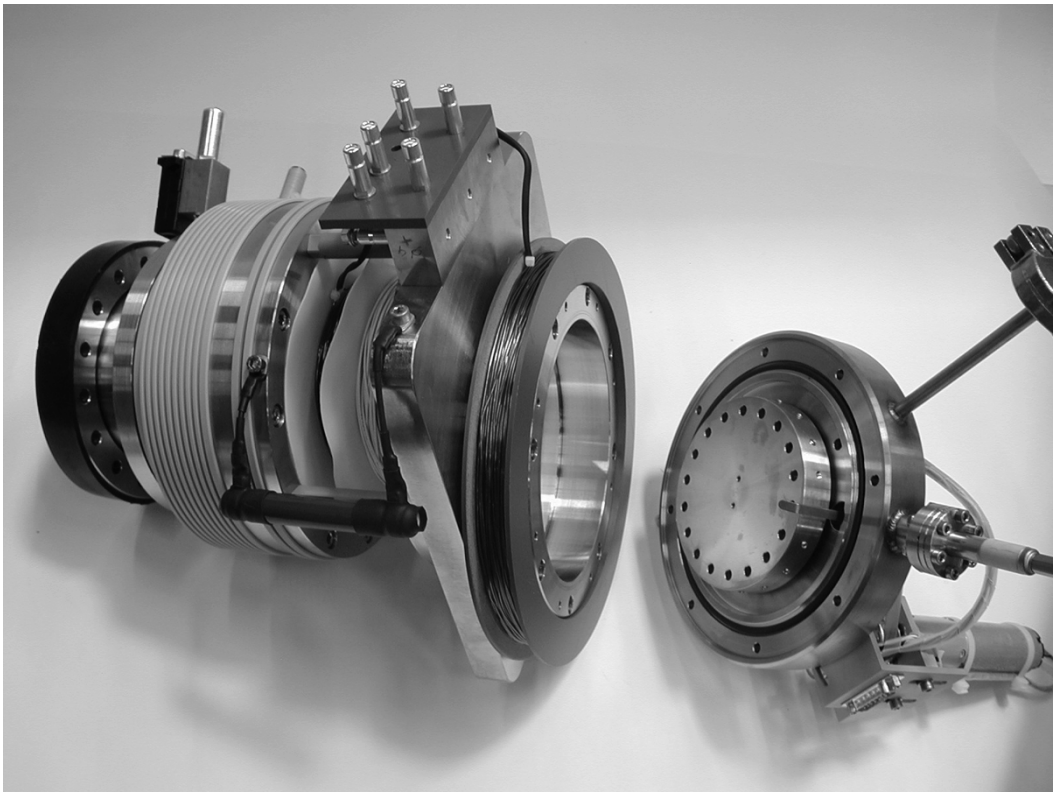


Fig. 11: Photograph of the partially disassembled GSI vacuum arc ion source

Table 2: GSI vacuum arc ion source and ion beam parameters.

Ion species	Uranium
Ion charge state distribution	U^{3+} = 16%, U^{4+} = 67%, U^{5+} = 14%, U^{6+} = 3%
Extracted ion beam current	156 mA at 35 kV
Accelerated ion beam current	55 mA at 131 kV
Analysed U^{4+} ion beam current	25 mA
Pulse length / repetition rate	0.6 ms / 1 pps
Extraction system	13 × 3 mm, multi-aperture
$\varepsilon_{x,y}$ (156 mA at 35 kV)	200 π mm rad
$\varepsilon_{x,y}$ (55 mA at 131 kV)	350 π mm rad
$\varepsilon_{x,y}$ (after separation, 15 mA)	100 π mm rad
Noise, full beam / analysed (r.m.s.)	< ±4% / ±5%
Pulse-to-pulse stability	Better than 80%
Voltage breakdowns (extractor grids)	two per day
Cathode lifetime	12 hours at 0.6‰ duty
Lifetime of the ion source (between maintenance)	7 days for SIS injection (0.2‰ duty)

This ion source has proven its capability in extended tests at the GSI high-current injector and has been put into regular operation for the GSI accelerator facility, providing uranium ion beams with current about an order of magnitude greater than previously possible. We note that, because of the fixed input velocity required to match into the RFQ (2.2 keV/u), the ion source extraction voltage and the post-acceleration must be changed for the various different ion species used; thus each new species needs to be independently optimized. The GSI sources have been described in detail elsewhere [34–36, 55–57].

5.3 Heavy-ion fusion application

Heavy-ion fusion (HIF) is one approach to the problem of controlled thermonuclear power production, in which a small DT (deuterium-tritium) target is bombarded by an intense flux of heavy ions and compressed to fusion temperatures. There is a need in present HIF research and development for a reliable ion source for the production of heavy-ion beams with low emittance, low beam noise, ion charge states Q = 1+ to 3+, beam current 0.5 A, pulse width 5–20 μs, and repetition rate 10 pulses per second. The suitability of a vacuum arc ion source for this application has been explored [58–61]. Energetic, high-current, gadolinium ion beams were produced, using the LBNL Mevva V source, with parameters as required or close to those required, and it is probable that the performance parameters can all be improved yet further in an optimized source design. An example of the beam pulse obtained is shown in Fig. 12, which shows an oscillogram of a Gd beam (85% in the Gd^{2+} charge state) formed at an ion energy of 120 keV (60 kV extraction voltage), a collected beam current of 120 emA (electrical milliamperes), for a 20 μs pulse width and a pulse rise time of less than 1 μs [59]. The beam emittance was about 0.3 π mm mrad (normalized). These preliminary results indicate that a vacuum arc-based metal ion source of this kind is a good potential candidate for an HIF ion source.

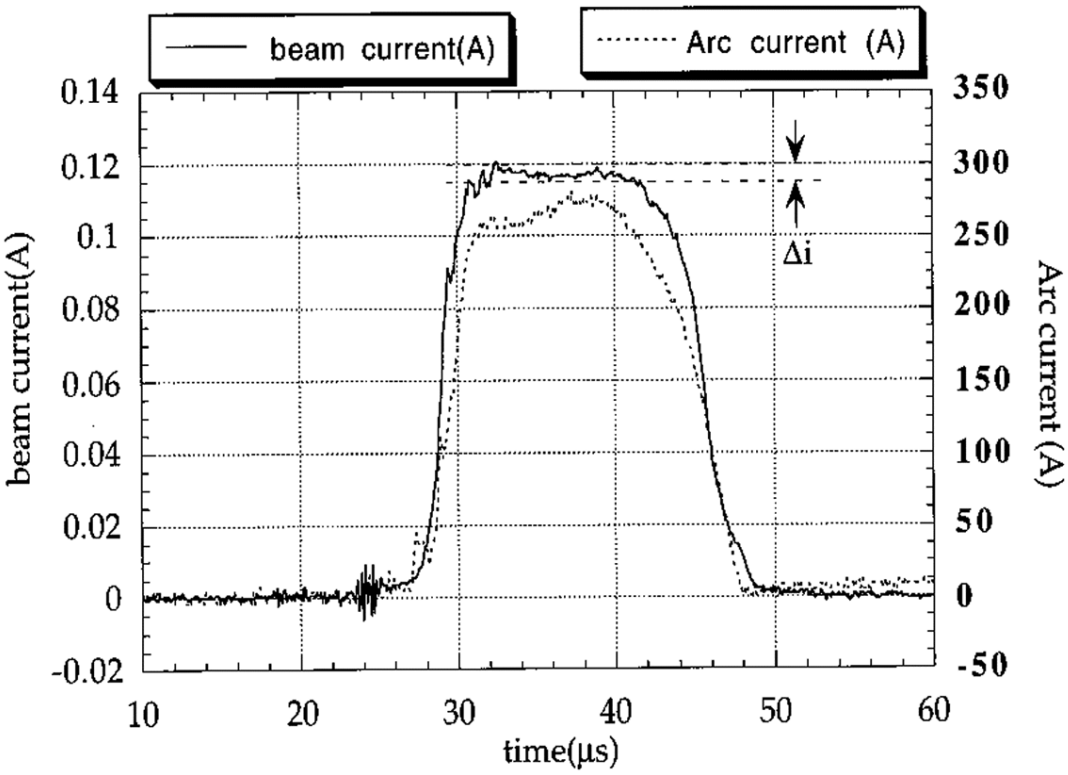


Fig. 12: Gd ion beam current (solid line), and arc current (dotted line), when the extraction is optimized (beam-formation perveance match condition).

6 Conclusion

High-current beams of metal ions can be provided by the vacuum arc ion source. Virtually all the metals of the Periodic Table, and carbon, have been used, and beams containing mixtures of metal ions and non-metallic species can also be made. The ions are in general multiply ionized but of low charge state, the beam typically containing a distribution of charge states spanning the rough range from 1+ to 5+, and techniques for increasing the charge states by as much as a factor of 2 have been demonstrated. The mean ion energy of the extracted beam is typically around 50–100 keV, with the

charge state components of energy up to as much as several hundred kiloelectronvolts. The ion current that can be extracted is large, and the record high demonstrated, albeit with a 50 cm diameter extractor, is about 20 A; from smaller diameter extractors and thus for beams with a tighter emittance, beam current of 100–200 emA (all charge states) is typical. Although beam noise and shot-to-shot irreproducibility were initially a daunting concern, these features have been vastly improved, in particular by the GSI/HCEI collaborative work and as manifested by the highly successful use of the GSI vacuum arc ion source for routine production and accelerator injection, particularly of high-current U^{4+} beams.

References

[1] R.K. Wakerling and A. Guthrie (Eds.), *Electromagnetic Separation of Isotopes in Commercial Quantities* (National Nuclear Energy Series, USAEC, 1951), p. 324.

[2] I.G. Kesaev, *Cathode Processes in Electric Arcs* (Nauka, Moscow, 1968) (in Russian).

[3] A.A. Plyutto, *JETP* **39** (1960) 1589 (in Russian).

[4] E.I. Revutskii, G.M. Skoromnyi, Yu.F. Kulygin, and I.I. Goncharenko, Proc. Soviet Conf. on Charged-Particle Accelerators, Moscow, 1968, Ed. A.A. Vasiliev (transl. published by the USAEC), p. 447.

[5] P.D. Prewett and R. Holmes, *J. Phys. E: Sci. Instrum.* **12** (1979) 179.

[6] S. Humphries, Jr., M. Savage, and D.M. Woodall, *Appl. Phys. Lett.* **47** (1985) 468.

[7] C. Burkhart , S. Coffey, G. Cooper, S. Humphries, Jr., L.K Len, A.D. Logan, M. Savage, and D.M. Woodall, *Nucl. Instrum. Meth. Phys. Res. B* **10–11** (1985) 792.

[8] S. Humphries, Jr., C. Burkhart, and L.K. Len, in *The Physics and Technology of Ion Sources*, Ed. I.G. Brown (Wiley, New York, 1989), p. 397.

[9] I.G. Brown, J.E. Galvin, and R.A. MacGill, *Appl. Phys. Lett.* **47** (1985) 358.

[10] I. G. Brown, *Rev. Sci. Instrum.* **65** (1994) 3061.

[11] G.P. Badgenov, S.P. Bugaev, G.P. Erokhin, V.N. Kiselev, A.E. Ligatchev, S.M. Chesnokov, and A.V. Ianchiuck, Proc. 5th All-Union Symp. on High Current Electronics, Tomsk, Russia, 1984, Part II, p. 93 (in Russian).

[12] S.P. Bugaev, A.G. Nikolaev, E.M. Oks, P.M. Schanin, and G.Yu. Yushkov, *Rev. Sci. Instrum.* **65** (1994) 3119.

[13] A.I. Ryabchikov, *Rev. Sci. Instrum.* **63** (1992) 2425.

[14] A.I. Ryabchikov, S.V. Dektjarev, and I.B. Stepanov, *Rev. Sci. Instrum.* **65** (1994) 3126.

[15] A.M. Tolopa, *Rev. Sci. Instrum.* **65** (1994) 3134.

[16] I.G. Brown and E.M. Oks, *IEEE Trans. Plasma Sci.* **25** (1997) 1222.

[17] G.A. Mesyats, *Cathode Phenomena in a Vacuum Discharge* (Nauka, Moscow, 2000).

[18] R.L. Boxman, P.J. Martin, and D.M. Sanders (Eds.), *Handbook of Vacuum Arc Science and Technology* (Noyes, New York, 1995).

[19] J.M. Lafferty (Ed.), *Vacuum Arcs – Theory and Application* (Wiley, New York, 1980).

[20] P. Spädtke, H. Emig, B.H. Wolf, and E.M. Oks, *Rev. Sci. Instrum.* **65** (1994) 3113.

[21] A. Anders, I.G. Brown, R.A. MacGill, and M.R. Dickinson, *J. Phys. D: Appl. Phys.* **31** (1998) 584.

[22] I.G. Brown, in *The Physics and Technology of Ion Sources*, Ed. I.G. Brown (Wiley, New York, 1989), p. 331.

[23] E. Oks and I. Brown, in *The Physics and Technology of Ion Sources*, 2nd edn., Ed. I.G. Brown (Wiley-VCH, Berlin, 2004), p. 257.

[24] E.M. Oks, *Rev. Sci. Instrum.* **69** (1998) 776.

[25] S. Anders, S. Raoux, K. Krishnan, R.A. MacGill, and I.G. Brown, *J. Appl. Phys.* **79** (1996) 6785.

[26] R.A. MacGill, A. Vizir, and I.G. Brown, *Rev. Sci. Instrum.* **71** (2000) 672.

[27] I.G. Brown and X. Godechot, *IEEE Trans. Plasma. Sci.* **19** (1991) 713.

[28] I.G. Brown, J.E. Galvin, R.A. MacGill, and R.T. Wright, *Rev. Sci. Instrum.* **58** (1987) 1589.

[29] A. Anders, *Phys. Rev. E* **55** (1997) 969.

[30] G.A. Mesyats, *IEEE Trans. Plasma Sci.* **23** (1995) 879.

[31] I.G. Brown, X. Godechot, P. Spädtke, H. Emig, D.M. Rück, and B.H. Wolf, Proc. IEEE Particle Accelerator Conf. (IEEE Conf. Proc. Series. 91CH3038-7, 1991), p. 1943.

[32] I.G. Brown, *IEEE Trans. Plasma Sci.* **21** (1993) 537.

[33] S. Humphries, Jr., C. Burkhart, S. Coffey, G. Cooper, L.K. Len, M. Savage, D.M. Woodall, H. Rutkowski, H. Oona, and R. Shurter, *J. Appl. Phys.* **59** (1986) 1790.

[34] E. Oks, P. Spädtke, H. Emig, and B.H. Wolf, *Rev. Sci. Instrum.* **65** (1994) 3109.

[35] E. Oks, G. Yushkov, I. Litovko, A. Anders, and I. Brown, *Rev. Sci. Instrum.* **73** (2002) 735.

[36] R. Hollinger, M. Galonska, and P. Spädtke, *Rev. Sci. Instrum*. **75** (2004) 1595.

[37] E.M. Oks, G.Yu. Yushkov, P.J. Evans, A. Oztarhan, I.G. Brown, M.R. Dickinson, F. Liu, R.A. MacGill, O.R. Monteiro, and Z. Wang, *Nucl. Instrum. Meth. Phys. Res. B* **127/128** (1997) 782.

[38] P. Spädtke, H. Emig, B.H. Wolf, and E.M. Oks, *Rev. Sci. Instrum.* **65** (1994) 3113.

[39] G.Y. Yushkov, R.A. MacGill, and I.G. Brown, *Rev. Sci. Instrum.* **75** (2004) 1582.

[40] G.Y. Yushkov and A. Anders, *IEEE Trans. Plasma Sci.* **26** (1998) 220.

[41] I.G. Brown, *IEEE Trans. Plasma Sci.* **21** (1993) 537.

[42] H. Shiraishi and I.G. Brown, *Rev. Sci. Instrum.* **61** (1990) 3775.

[43] H. Shiraishi and I.G. Brown, *Rev. Sci. Instrum.* **61** (1990) 589.

[44] A.S. Bugaev, V.I. Gushenets, A.G. Nikolaev, E.M. Oks, K.P. Savkin, P.M. Schanin, G.Yu. Yushkov, and I.G. Brown, in *Emerging Applications of Vacuum-Arc-Produced Plasma, Ion and Electron Beams*, Eds. E. Oks and I. Brown (Kluwer Academic, Dordrecht, 2002), p. 79.

[45] A.I. Ryabchikov, I.B. Stepanov, S.V. Dektjarev, E.I. Lukonin, and I.A. Shulepov, *Rev. Sci. Instrum.* **71** (2000) 704.

[46] A.I. Ryabchikov, I.B. Stepanov, S.V. Dekjarev, and O.V. Sergeev, *Rev. Sci. Instrum.* **69** (1998) 893.

[47] A. Oztarhan, I. Brown, C. Bakkaloglu, G. Watt, P. Evans, E. Oks, A. Nikolaev, and Z. Tek, *Surf. Coat. Technol.* **196** (2005) 327.

[48] P.J. Evans, G.C. Watt, and J.T. Noorman, *Rev. Sci. Instrum.* **65** (1994) 3082.

[49] K. Baba, S. Nagata, R. Hatada, T. Daikoku, and M. Hasaka, *Nucl. Instrum. Meth. Phys. Res. B* **80/81** (1993) 297.

[50] A.D. Liu, H.X. Zhang, and T.H. Hang, *Surf. Coat. Technol*. **193** (2005) 65.

[51] T. Zhang, H. Zhang, C. Ji, X. Zhang, Y. Wu, F. Ma, H. Liang, H. Shou, and J. Shi, *Surf. Coat. Technol.* **128–129** (2000) 1.

[52] H. Zhang, X. Zhang, F. Zhou, S. Zhang, Q. Li, and Z. Han, *Rev. Sci. Instrum.* **65** (1994) 3088.

[53] V.A. Batalin, J.N. Volkov, T.V. Kulevoy, and S.V. Petrenko, *Rev. Sci. Instrum.* **65** (1994) 3104.

[54] M.M. Katz, L.N. Kondratiev, N.N. Pomelov, and A.D. Rogal, *Rev. Sci. Instrum.* **65** (1994) 3101.

[55] H. Reich, P. Spädtke, and E.M. Oks, *Rev. Sci. Instrum*. **71** (2000) 707.

[56] F. Heymach, M. Galonska, R. Hollinger, K. Leible, P. Spädtke, and M. Stork, in *Emerging Applications of Vacuum-Arc-Produced Plasma, Ion and Electron Beams*, Eds. E. Oks and I. Brown (Kluwer Academic, Dordrecht, 2002), p. 59.

[57] R. Hollinger, M. Galonska, and P. Spädtke, Proc. 20th Int. Symp. on Discharges and Electrical Insulation in Vacuum (Yalta, Ukraine, 2004), IEEE Cat. No. 04CH37565, vol. 2, p. 547.

[58] I.G. Brown, *AIP Conf. Proc.* **152** (1986) 207.

[59] F. Liu, N. Qi, S. Gensler, R.R. Prasad, M. Krishnan, and I.G. Brown, *Rev. Sci. Instrum.* **69** (1998) 819.

[60] N. Qi, J. Schein, R.R. Prasad, M. Krishnan, A. Anders, J. Kwan, and I.G. Brown, *Nucl. Instrum. Meth. Phys. Res. A* **464** (2001) 576.

[61] A. Anders and J.W. Kwan, *Nucl. Instrum. Methods Phys. Res. A* **464** (2001) 569.

Radioactive Ion Sources

T. Stora
CERN, Geneva, Switzerland

Abstract
This chapter provides an overview of the basic requirements for ion sources designed and operated in radioactive ion beam facilities. The facilities where these sources are operated exploit the isotope separation online (ISOL) technique, in which a target is combined with an ion source to maximize the secondary beam intensity and chemical element selectivity. Three main classes of sources are operated, namely surface-type ion sources, arc discharge-type ion sources, and finally radio-frequency-heated plasma-type ion sources.

1 Introduction

Radioactive ion sources are used in facilities that deliver radioactive ion beams. These beams are exploited by a large community of physicists active in nuclear physics, astrophysics, atomic physics, material science and life science. The production of the radioactive ions is based on two generic principles: a heavy-ion beam fragmented by a thin target and mass-separated 'in flight' in a series of large-acceptance dipole magnets; and online mass separation of nuclear products diffused out of thick targets impinged by light ion beams, this being referred to as the 'ISOL technique' [1]. In the in-flight technique, the fragments are charged and thus do not require the exploitation of ion sources; while in ISOL production, the fragments diffuse out of a production target or catcher as neutral species.

While ion sources are intimately related to the history and developments of stable particle accelerators, radioactive ion sources are likewise linked to the development of ISOL facilities. In such facilities, ion sources are one of the important links in the full production chain of the secondary radioactive ion beams. The first generation of radioactive ion sources were those developped for stable ions in accelerators such as cyclotrons and mass separators. At a later stage they were adapted to match the specific needs for the production and acceleration of radioisotopes, which are produced with several orders of magnitude less intensities than their stable counterparts, and have, by definition, a finite lifetime before radioactive decay.

The lecture given during the school could not cover all the required aspects in an exhaustive manner. This chapter will introduce some additional aspects such as some historical elements concerned with the development of the sources used in radioactive beam facilities, and some references will be provided to those which need information at an expert level. The chapter is organized in the following manner. In a first part, a brief technical description of a radioactive ion beam facility and the main requirements for a radioactive ion source are introduced. The different classes of ion sources are then reviewed. The sources are split into three main categories; surface-type ion sources, the forced electron beam sources, and finally the RF-heated plasma sources. For each of these categories, historical developments are reported first, followed by the theoretical considerations underlying their operation, and finally the main operational parameters. An outlook for the on-going and future needs in our facilities will finally be reported.

2 Radioactive ion beam facilities

The first report of the online production of a radioactive ion beam dates back to the 1950s, when exotic krypton fission fragments (of a few tens of seconds half-life) were produced from large UO_2 targets heated and connected to an isotope mass separator [1]. The ion source was of a glow discharge Nielsen type (Fig. 1) [2]. About 15 years after, a number of so-called first-generation facilities came into operation in different laboratories across the world, amongst which was ISOLDE at CERN, still in operation today after a series of major upgrades and evolutions. The family of ion sources used throughout the facilities was enlarged over the years with new members, namely forced electron beam-induced arc discharge (FEBIAD) ion sources, surface ion sources, and later on RF-driven ion sources, more particularly electron cyclotron resonance (ECR) ion sources [3]. A schematic layout of an ISOL-type facility is shown in Fig. 2. The main components present in such facilities are:

- a primary particle driver;
- a target station with standard interfaces (vacuum, electrical power, cooling circuit) and the first optical elements of the secondary isotope beam line;
- a target and ion source unit, which is mechanically tight to the station and regularly exchanged, often by remote handling exploiting robot and piloted cranes;
- a magnetic dipole where the different components of the isotope secondary beam are selected according to mass over charge figures;
- additional electrostatic optical elements to steer the beam along the beam lines.

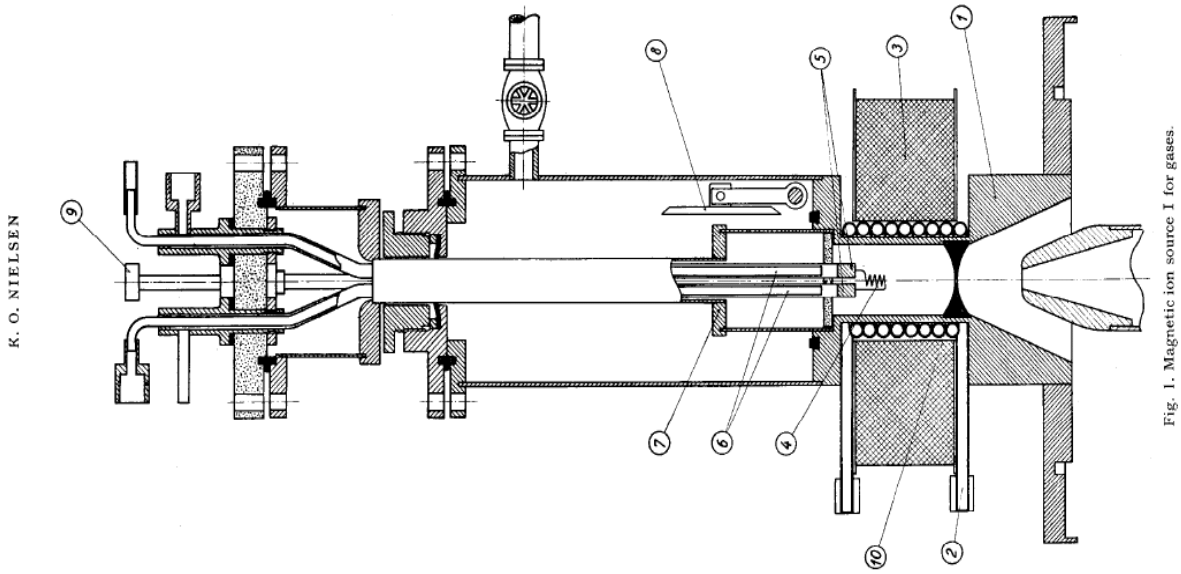


Fig. 1: Ion source used in the first online separation of a radioactive ion beam at the Niels Bohr Institute in 1951 [1] (reprinted from Ref. [2]). The ion source was said to be of 'magnetic type' by Nielsen, and is known today to be of the Nielsen type. The temperature of the arc discharge chamber could be held up to 1100°C. The 0.4 mm diameter tungsten filament lifetime was 50–100 h, operated at 0.1 mbar gas pressure. Components: 1, brass anode block; 2, Cu water cooling tube; 3, magnet coil; 4, filament; 5, Cu rods; 6, water-cooled brass tubes; 8, 'vacuum lock'; 9, needle valve for controlling the gas inlet; 10, stainless-steel bottom plate. Operating parameters (5% efficiency on Kr): anode voltage, 100–200 V; arc current, 1–2 A; magnetic field, 0–400 G; diameter of outlet opening, 2–5 mm; beam current for krypton, 50–150 μA.

The figure of merit of a given radioactive ion beam facility comes from the folding of the following parameters:

- number of available radioactive ion beams;
- intensity of the beam;
- beam quality, for instance its purity and emittance;
- non-degradation of the beam performance over time, like intensity, purity, energy;
- facility up-time;
- (post)-accelerated beam characteristics.

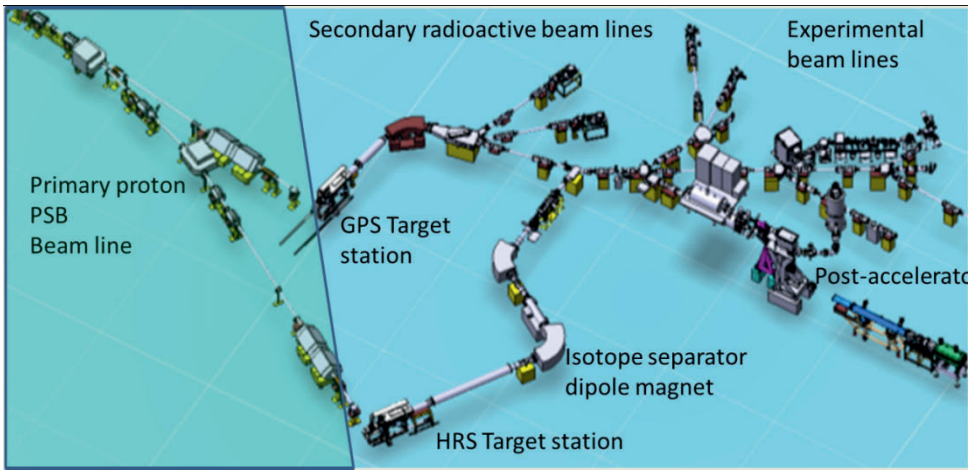


Fig. 2: General view of an isotope online facility such as ISOLDE at CERN. The buildings and shielding structure are not represented.

In a given facility, the characteristics of the primary beam, often a large particle accelerator, are defined in the design and construction phases. The rest of the infrastructures is also determined at the conception of the facility, such as the buildings or services. The parts that deal with the design of the target units, and to a lesser extent the beam lines, on the other hand, can still be and mostly are developed over time, to meet the evolving needs of the experiments. These developments more particularly focus on the different types of target materials to generate the isotopes, on the engineering aspects of the target production unit design, and – in the focus here – on the development of the ion sources. For a given facility, several combinations of target materials and ion sources are integrated into compact units, and are regularly exchanged in order to deliver the required isotope beams during a year of operation. The integration of the source into a compact unit dates back to the 1970s at ISOLDE (Fig. 3). A contemporary target and ion source unit is shown in Fig. 4.

3 Requirements for a radioactive ion source

An ideal ion source for ISOL radioactive ion beam facilities must fulfil a set of basic criteria. The development of any existing source or the design of a new one will therefore consist in satisfying the maximum number of these requirements. Often compromises have to be made, i.e., increasing the heating power of a source to achieve high ionization efficiency could affect the source stability over time.

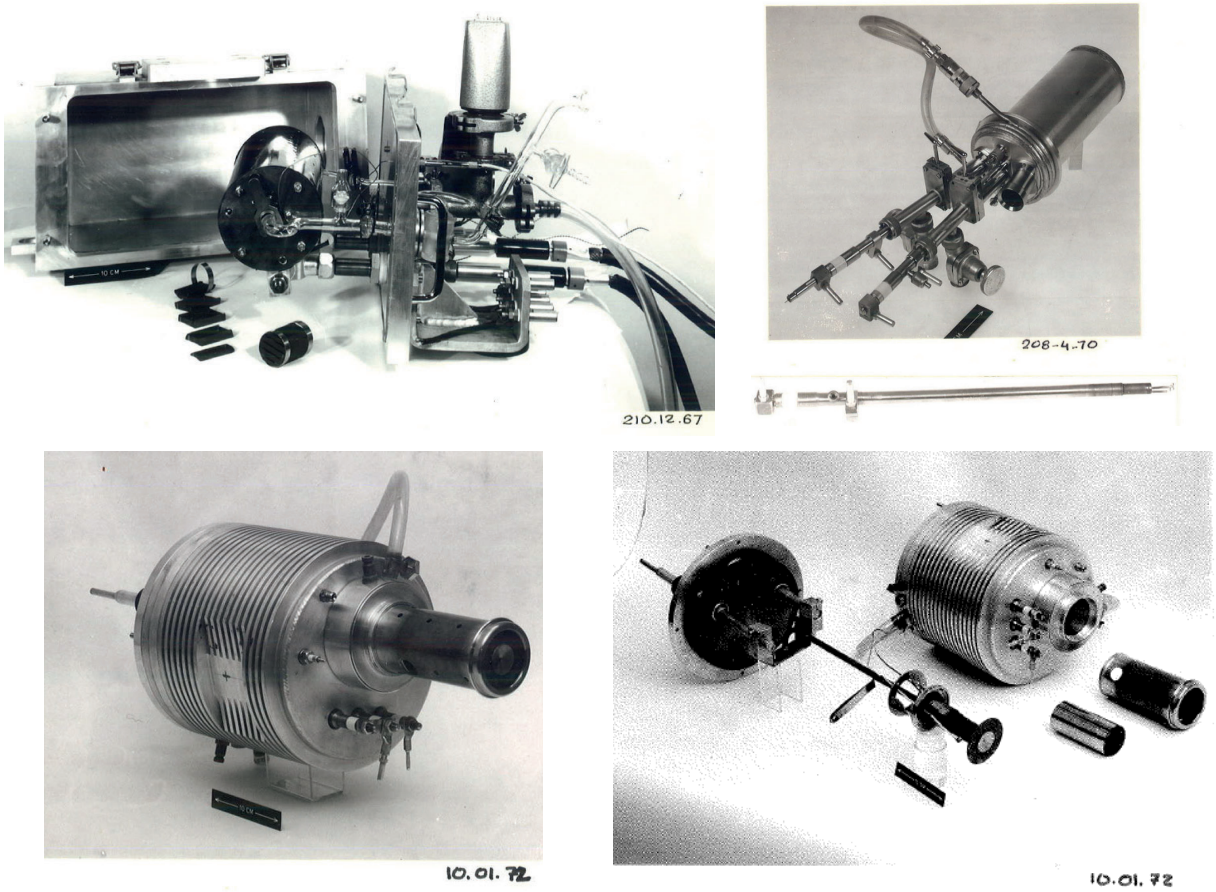


Fig. 3: Top: target and glow discharge ion source used at ISOLDE until the 1970s. Bottom: integrated MK-IV target and ion source unit in 1972.

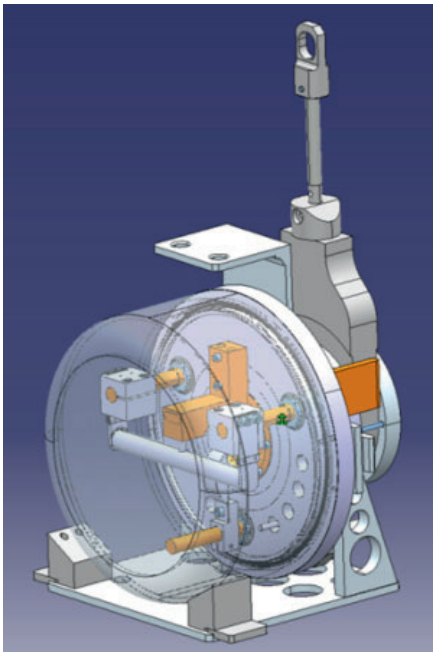


Fig. 4: Modern target and ion source unit at ISOLDE. The diameter of the confinement box (seen in transparency) is 30 cm.

An ideal source must:

- be compact;
- have radiation-hard components to withstand a predefined dose (e.g., 1 MGy for an installation like ISOLDE);
- be compatible with the predefined interfaces and first beam optical elements, i.e., the electrostatic extraction puller;
- provide a maximum ionization efficiency, if possible approaching 100%;
- operate in a stable manner with varying gas loads coming from the target heated at high temperatures and impacted by the primary beam (i.e., there should be no drift in performance during operation which requires a retuning of its parameters);
- be element-selective, i.e., ionize only isotopes from one element amongst a gas phase made up of other isotopes, stable impurities and molecules with several orders higher concentrations;
- produce a total extracted beam intensity compatible with the low-energy beam-line elements, i.e., typically of about 100 μA or less;
- have a perfect beam quality, i.e., no energy spread and a minimal transverse emittance;
- have small residence time with respect to the isotope of interest, i.e., in the millisecond range for the most exotic isotopes produced by the ISOL method.

It is clear that no source is universal, and since a facility aims to deliver as many different elements and isotopes as possible, sources need to be developed over the years, and, during operation, be exchanged from one production unit to the next one, typically once a week or once a month, to meet the requirements for the production of the element under consideration.

It is easy to understand that ion sources need very different characteristics to ionize noble gases such as argon, metallic elements such as nickel, or chemically reactive atoms such as boron. To help provide a more rational view, some key properties are reported in Fig. 5. A periodic table of the chemical elements is displayed, showing their ionization potential, their electron affinity and their melting and boiling points.

In the following section, the three important categories of radioactive ion sources for ISOL facilities are introduced, namely those based on surface ionization processes, on electron beam impact, and on RF-heated plasma.

4 Surface ion sources

4.1 Historical elements

Surface ion sources are used in most present-day ISOL-type facilities. They were developed in the early days of the isotope separation technique, being simple, robust, fast and selective ion sources. A layout of the target and ion source unit first developed and used in Orsay before being transferred to the Proton Synchrotron (PS) beam at CERN is shown in Fig. 6 [4]. In this configuration, the target, with direct interaction of the primary particle beam, for the production and recoil of the reaction products, is located inside an oven. It is made of graphite or refractory metal and brought up to a high temperature of 1500°C. The surfaces of the oven act as the ion source itself, exploiting the well-known surface ionization phenomenon first identified and described by Saha and Langmuir for heated metallic filaments [5]. The ions are then emitted through a vertical slit before acceleration and injection in the magnetic dipole mass spectrometer. The ionization efficiency for alkali isotope beams was reported to be small, but beams of ^{9}Li of 180 ms half-life could already be extracted [6]. Ten years later, the concept had been improved, and ^{11}Li and ^{32}Na beams with half-lives in the 10 ms range could be produced with a rhenium foil inserted in the oven as an ionizing surface [7]. Efficiencies in the few per cent range were reported.

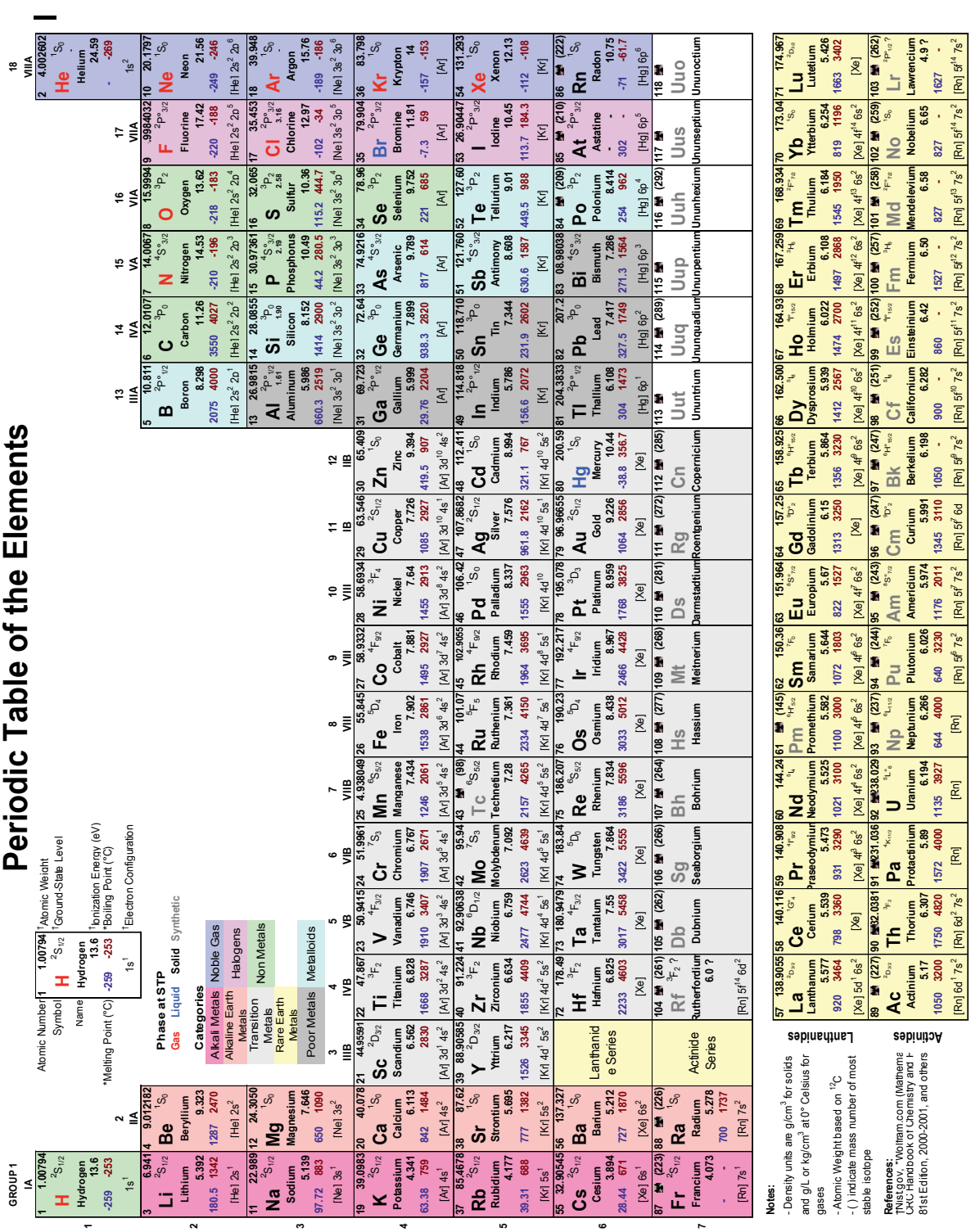


Fig. 5: Periodic table of the chemical elements

Surface ion sources made of hollow tubes were introduced in the 1970s [8]. They were heated by electron bombardment. A version adapted for online operation at ISOLDE was available soon after [9] (Fig. 7). In this last version, the source is connected to a large target container via a transfer line, in such a way that the beam irradiation and target heating are spatially decoupled from the operating source. Heating was provided either by electron bombardment or by ohmic current. Efficiencies in the several tens of per cent for lanthanides were reported in this configuration. Later on, other developments took place, for instance providing negative surface sources and investigating different materials [10]. They are based on similar processes as their positive counterparts for singly negative ions. Some examples of such sources are shown in Fig. 8.

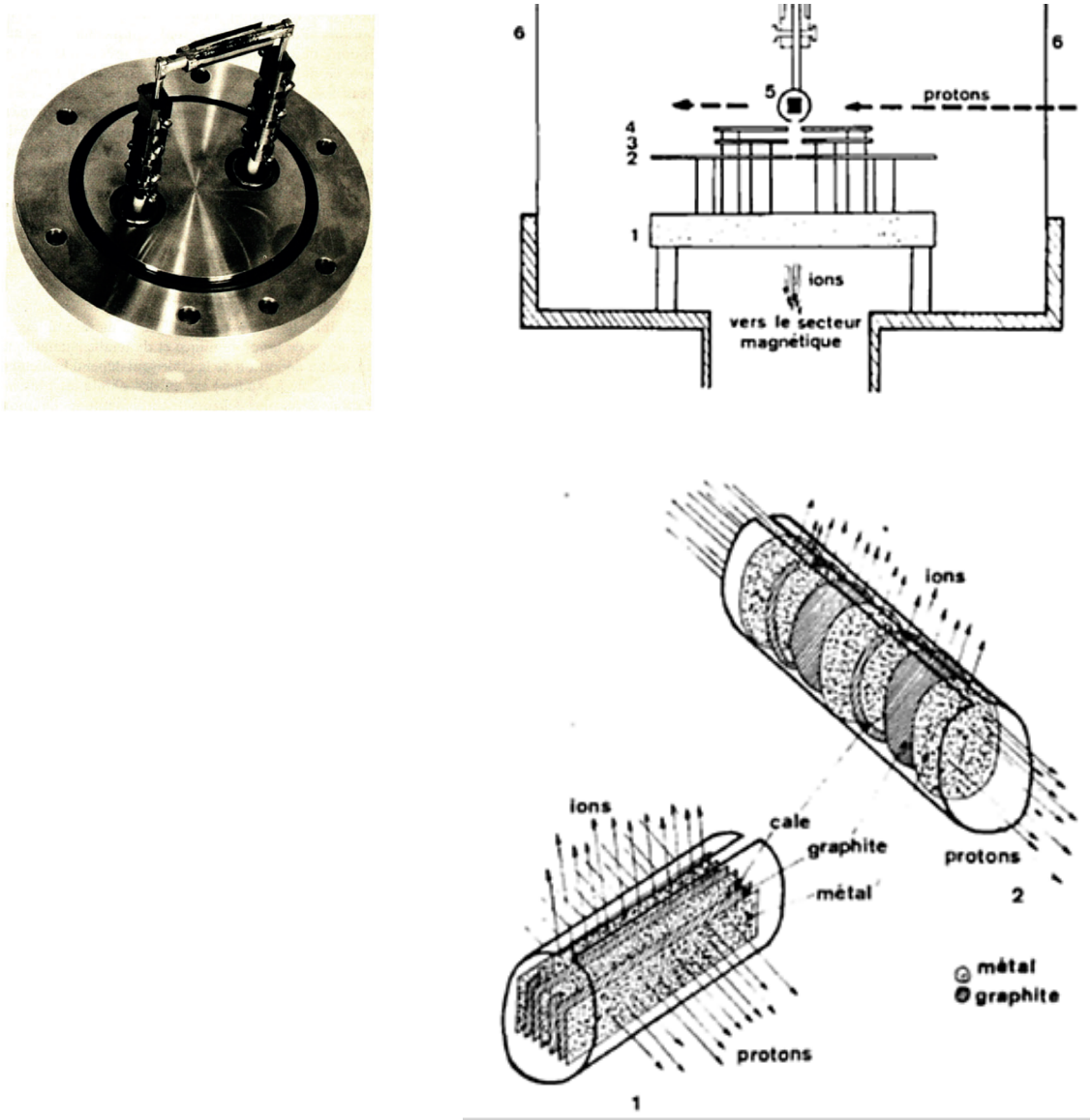


Fig. 6: Surface ion source developed at Orsay to ionize alkali metal beams and used at CERN PS in the 1960s (from Ref. [4]).

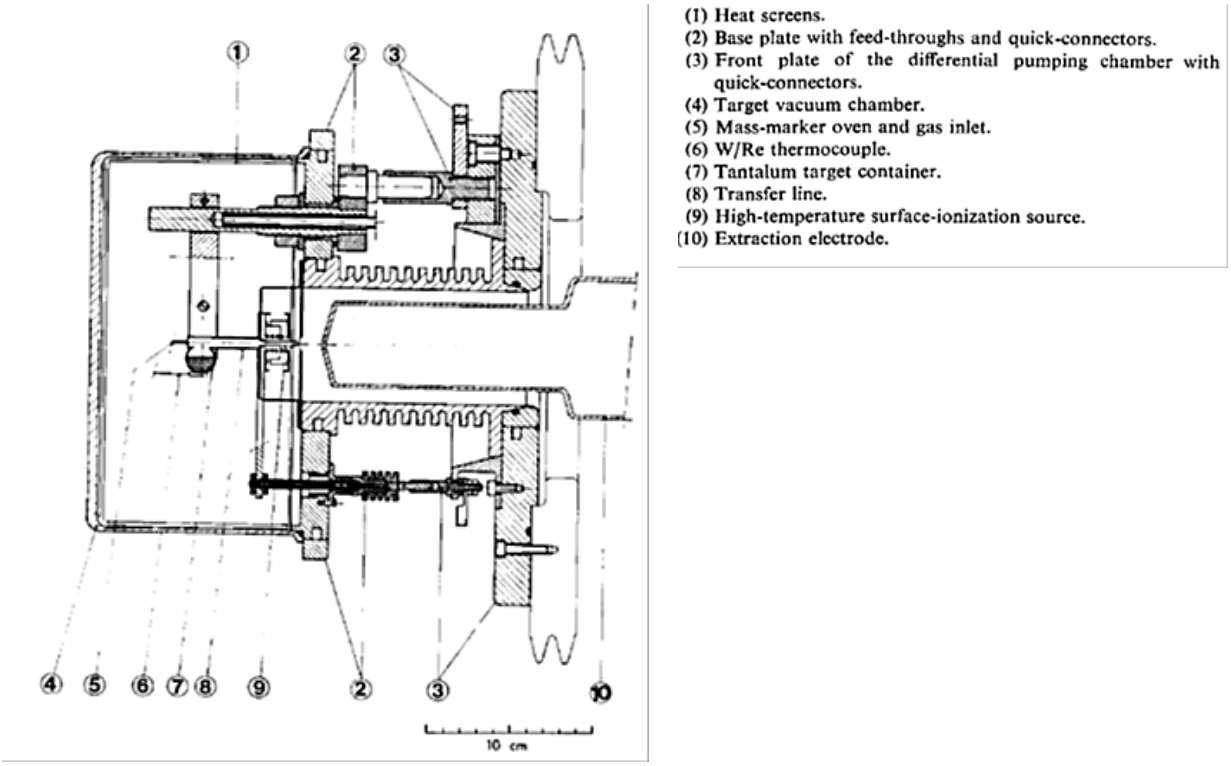


Fig. 7: Surface ion source used at CERN-ISOLDE (from Ref. [9])

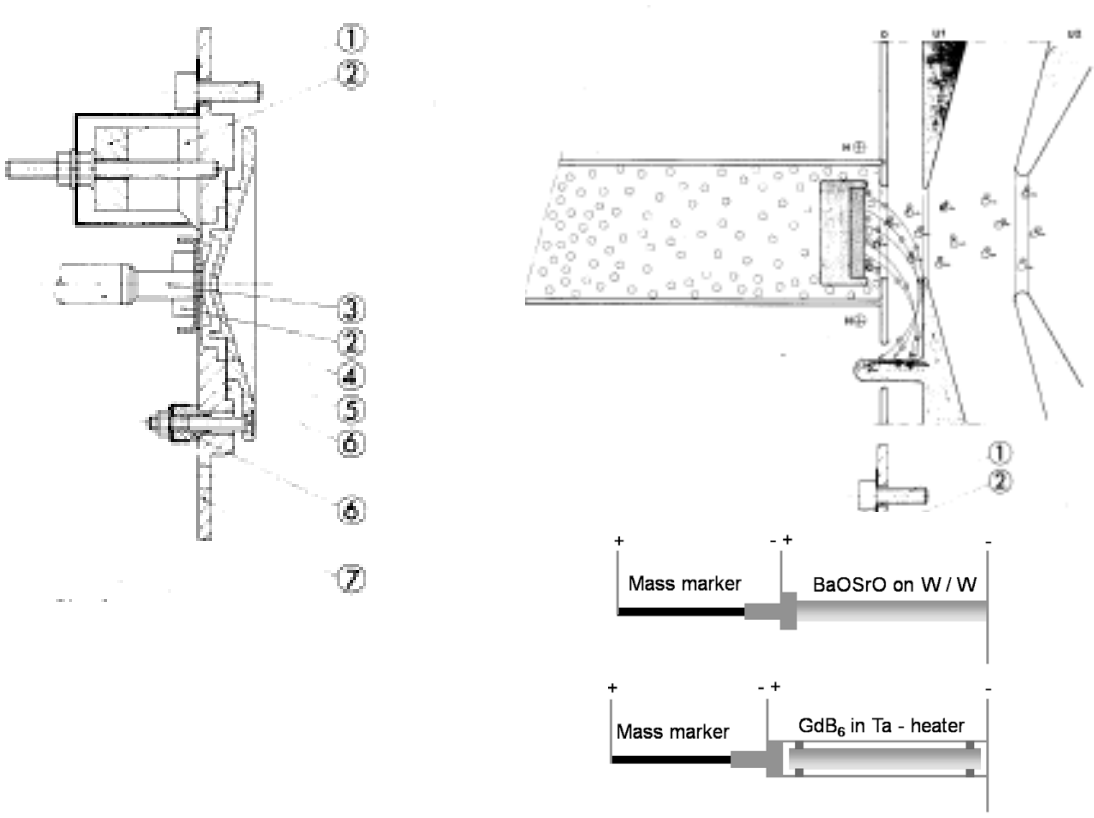


Fig. 8: MK4 negative surface ion source developed at ISOLDE and some prototype tubular geometries (from Ref. [10]).

4.2 Mode of operation and underlying physics principles

The principle of operation is rather simple. It is based on the following well-known Saha–Langmuir equation, which describes the ionization degree α and efficiency ε for weak plasmas [5] and for heated surfaces, where g_+ and g_0 are the quantum degeneracy from the charged and neutral states, W is the work function of the ionizing material, and ϕ_i is the ionization potential of the atoms under consideration:

$$\alpha_{\text{surface}} = (g_+/g_0)\exp[(W-\phi_i)/kT]\,, \tag{1}$$

$$\alpha_{\text{surface}} = \exp[(W-\phi_i^*)/kT] \qquad \text{with} \qquad \phi_i^* = \phi_i - kT\ln(g_+/g_0)\,, \tag{2}$$

$$\varepsilon_{\text{surface}} = \frac{1}{1+1/\alpha_{\text{surface}}}\,. \tag{3}$$

Eqs. (1) and (3) were established in 1925 and were used to describe the measured efficiencies of ion sources of simple geometries, such as heated surfaces and filaments. Eq. (2) is a convenient form in which the pre-factor originating from multiple quantum states of the neutral atom and the charged ion are already included in a normalized ionization potential parameter ϕ_i^*.

However, this model was found to be inaccurate in properly describing the behaviour of the source in a hollow-tube configuration. Corrections to account for the proper plasma parameters in such heated tubes or capillaries had to be included.

It took ten years to develop a satisfactory model and to test it against experimental data. More details can be found in two review articles written by R. Kirchner [11]. In a hollow-tube configuration, correction factors are required to properly describe the observed ionization efficiencies. In this case, a very low-density plasma is created in the tube; this introduces a first correction factor κ, which accounts for the average number of collisions that the isotope experiences with the hot surface of the tube before being extracted, and a second one ω, the trapping efficiency of the emitted ion inside the plasma potential well. Some arguments based on the geometry and plasma properties can be given to estimate these two factors; however, in practice, the product $\omega\kappa$ is used as a free parameter to fit a set of experimental data obtained with a series of chemical elements that span a relevant range of ionization potentials and efficiencies. The model in Eq. (4) is therefore now accepted to describe ionization efficiencies for hollow-tube ionizers as found in most of the present ISOL-type facilities:

$$\varepsilon_{\text{ionizer}} = \frac{1}{1+1/\omega\kappa\alpha_{\text{surface}}}\,. \tag{4}$$

A graph plotting $\varepsilon_{\text{ionizer}}$ versus ϕ_i^* for a source made of tungsten with two different partial pressures of neutral xenon gas is shown in Fig. 9. A similar graph has been obtained very recently at JAEA-ARSC with a hot tantalum cavity as shown in Fig. 10 [12].

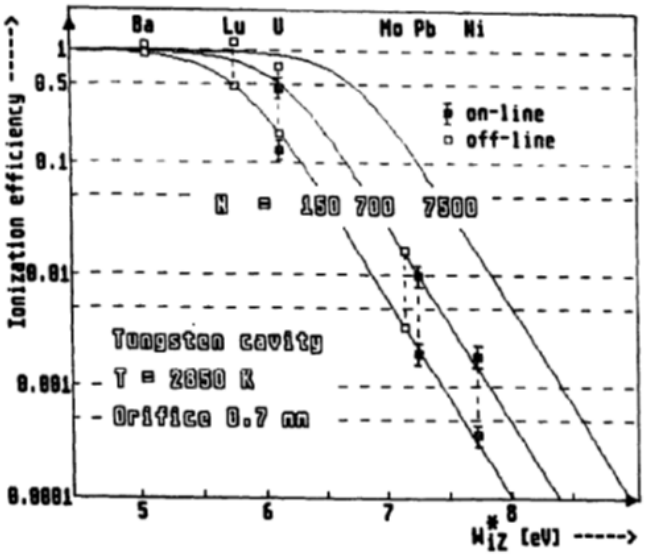


Fig. 5. Ionization efficiency of various trace elements in a cavity operated without xenon (lower values) and with a xenon input of around 10^{-3} cm^3(STP)/s (upper values), respectively. The latter corresponds to a neutral density of $\approx 5\times10^{14}$/cm^3 in the cavity. Curves are calculated according to eqs. (7) or (9), $N = 700$ implying that all surface-ionized particles leave the cavity as ions.

Fig. 9: Efficiency plotted versus ionization potential of different condensable chemical elements. Two sets of data recorded at different partial pressure in the hollow cavities are reported. The fit of the data using Eq. (4) and different values for $\omega\kappa$ (referred to as N in the figure) are also shown. From Ref. [11].

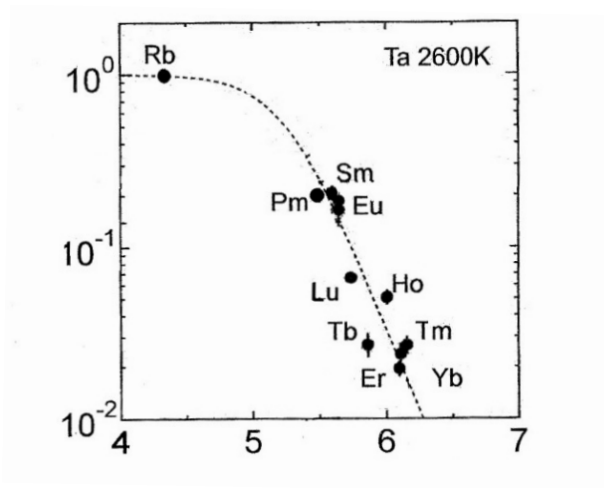


Fig. 10: Tantalum surface ion source developed at ASRC-JAEA heated by electron bombardment with two filaments. The ionization efficiencies are reported for a set of exotic radio-lanthanides for a cavity operated at 2600 K [12].

4.3 Specific applications, future developments

Surface ion sources are used in most ISOL-type facilities in operation or in construction. They are exploited to produce the most intense alkali and alkaline-earth isotope beams, and provide in addition some chemical selectivity. Besides their utilization as selective ionizers, they are nowadays combined with RILIS lasers to ionize metallic elements. This last mode of ionization is described elsewhere in this present volume by B. Marsh [13]. In this case, choices of materials with lower work functions are generally preferred, with a successful demonstration of using ceramics of GdB_6 to suppress isobaric contaminants [10]. Surface ion sources have been used to produce molecular beams with suitable ionization potentials. This is mainly lanthanide oxide and fluorine compounds, and alkaline-earth fluoride molecules [9, 14]. In that case the theoretical treatment to account for the operational behaviour of the source has not been investigated in great detail. The main elements required to account for evolving intensities and efficiencies of atomic versus molecular ions have been reported in the case of Ba and O atoms for Ba^+ and BaO^+ ion production [15]. The interested reader should refer to that publication for more details.

5 FEBIAD ion sources

5.1 Historical elements

Forced electron beam-induced arc discharge (FEBIAD) ion sources were developed to overcome some of the drawbacks observed while operating the previous generation of sources equipped with a heated cathode emitting electrons in a high-pressure plasma chamber. In that previous generation, the configuration of the filament, the magnetic field, the polarization of the plasma chamber walls, i.e., anode, and the layout of the ion extraction led to different versions referred to as the Nielsen, Bernas–Nier or hollow-cathode ion sources [11].

One version of the Nielsen source was shown in Fig. 1 and another one is shown in Fig. 11. Reported drawbacks are a finite lifetime of the cathodes, which could be dependent on the chemical reactivity of the gas phases, operation under high pressures of 10^{-2} mbar, high total extracted beam intensities, and difficulties in reaching and maintaining stable modes of operation.

The principal technical modifications of the previous source designs were the change of the cathode, from coiled wire geometry to a flat disc, and the insertion of a polarized grid in front to stabilize the electron emission current. The first development and application of these sources were reported by R. Kirchner at GSI and S. Sundell at Isolde in the 1970s [16]. Different versions of the FEBIAD sources were developed, exploiting various structural materials for the cathode, the accelerating grid, the anode chamber or the heat screens. Figure 12 shows the three FEBIAD sources MK3, MK5 and MK7 out of a larger series developed at ISOLDE and used until 2009 [17]. The different models were combined with different types of targets and connected via transfer lines kept at different temperatures.

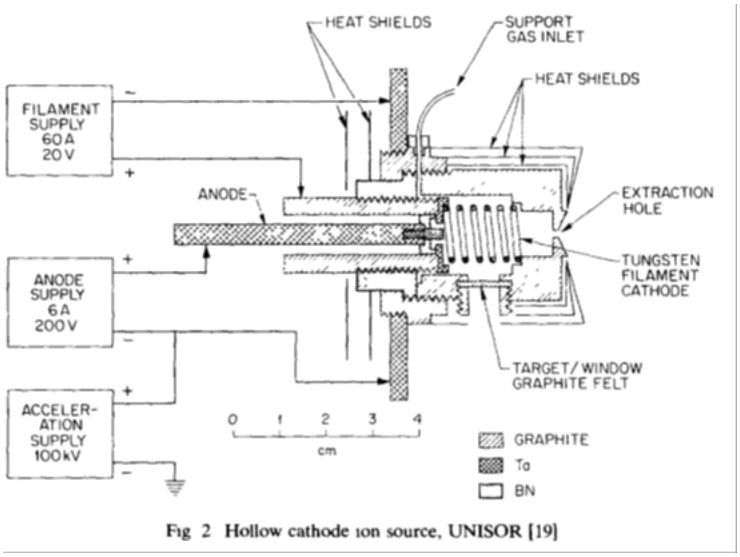


Fig. 11: Hollow-cathode ion sources applied at UNISOR at GSI (from Ref. [11])

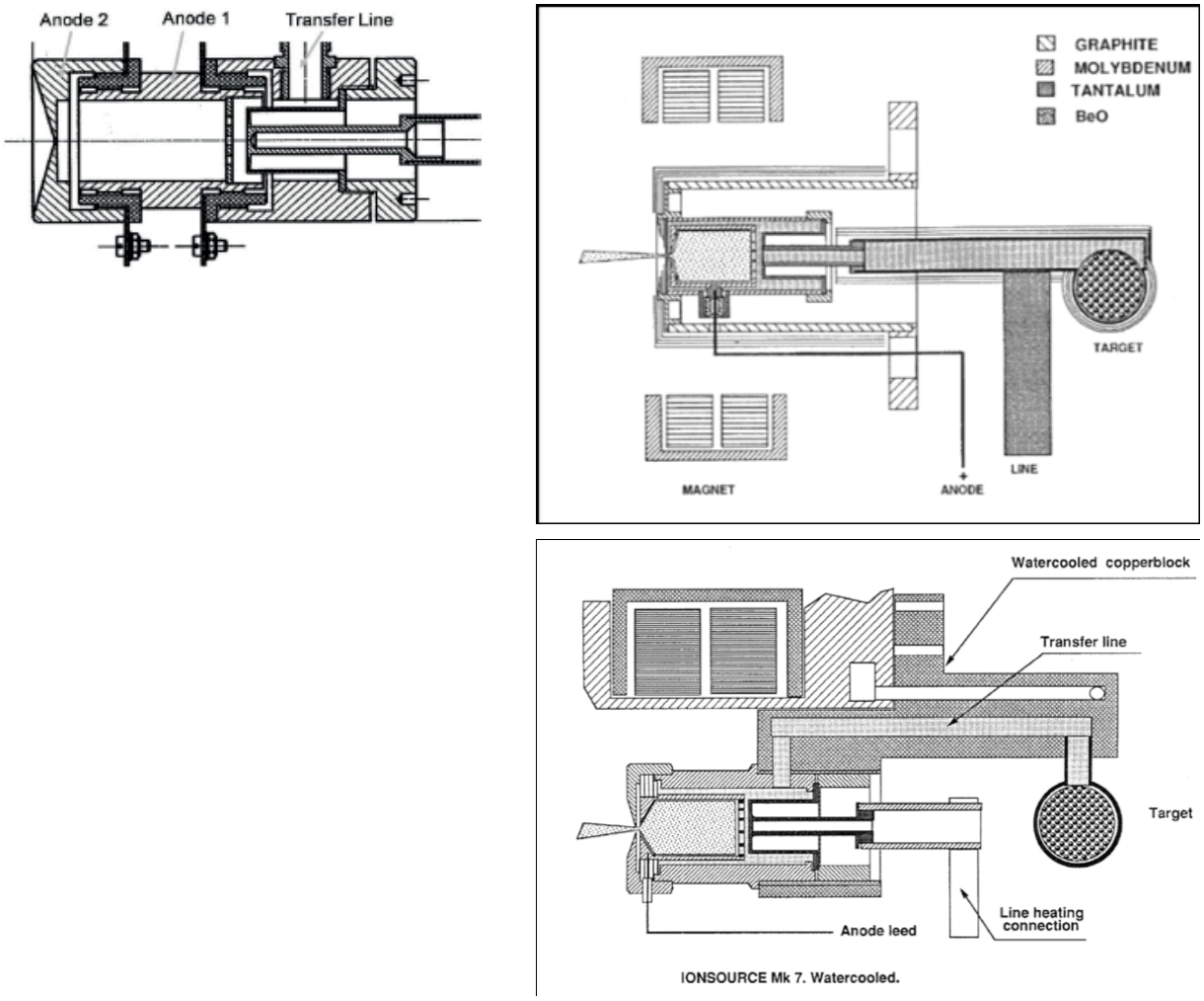


Fig. 12: MK3, MK5 and MK7 FEBIAD sources developed at ISOLDE (from Ref. [17])

5.2 Mode of operation and underlying physics principles

None of the theoretical models put forward could well describe the reported experimental efficiencies of the FEBIAD sources over extended ranges of operational parameters. In addition, some of the used parameters were not physically sound. During his PhD, L. Penescu proposed a new model that seemed to provide a reasonably good description of the observed ion beam currents using simple physical parameters [18]. This model and detailed experimental investigation of the three FEBIAD sources in operation at ISOLDE led to a new generation of such sources, which we named VADIS for versatile arc discharge ion source.

The proposed ionization efficiency model for FEBIAD sources is the following:

$$\varepsilon = \frac{R_{\text{ion}} V_{\text{source}} f_{\text{extr}}}{I_{\text{n_in}}} = \frac{(n_{\text{e}} n_{\text{n}} \sigma_{\text{ion}} v_{\text{rel}}) V_{\text{source}} f_{\text{extr}}}{I_{\text{n_in}}} . \tag{5}$$

The model described by Eq. (5) is a linear combination of a set of parameters that can be calculated from an experimentally accessible set of data, such as the emitted electron current from the cathode or the injected neutral buffer gas from the calibrated leak rate. Indeed, the different terms in Eq. (5) can be explicitly computed:

$$v_{\text{rel}} \approx v_{\text{e}} = \sqrt{\frac{2eU}{m_{\text{e}}}} \tag{6}$$

is the average relative speed between the neutral isotopes and the electrons emitted from the cathode and accelerated by the anode grid at a potential U, typically at 30–250 V;

$$n_{\text{e}} = \varepsilon_{\text{e}} \frac{j_{\text{e}}}{v_{\text{e}}} , \tag{7}$$

$$j_{\text{e}}\ [\text{mA mm}^{-2}] = \frac{4}{9} \varepsilon_0 \left(\frac{2e}{m_{\text{e}}} \right)^{1/2} \frac{U^{2/3}}{d^2} , \tag{8}$$

with n_{e} the electron density in the source, ε_{e} a scaling factor that accounts for the grid transparency and electron reaching the drift plasma chamber, and j_{e} the electron current emitted from the cathode, described by the Child–Langmuir equation (8) when space charged limited, with U the acceleration potential of the grid and d the distance between the cathode and the grid. This is for sources operated typically at temperatures above 2000°C in the configuration of a VADIS source. For low temperatures, the electron emission is then limited by the Richardson–Dushmann electron emission from a cathode:

$$j_{\text{e}}\ [\text{mA mm}^{-2}] = AT^2 \exp(-W/kT) \qquad \text{with} \qquad A \cong 1200\ \text{mA mm}^{-2}\ \text{K}^{-2} . \tag{9}$$

Here the work function W of the cathode is in eV and A is the Dushman constant, which can adopt different values for different materials. In Fig. 13 the current measured at the anode power supply is recorded, and can as a first approximation be compared to Eqs. (8) and (9), using the recorded temperature of the cathode and the anode potential. The neutral atom density is expressed by Eq. (10), in which the gas leak rate Q_{in} has been measured with a tracer gas, M is the mass of the isotope, the temperature T is assumed homogeneous and is that of the cathode, and S_{out} is the surface area of the extraction hole of the source:

$$n_{\mathrm{n}} = 2\times 10^{18} Q_{\mathrm{in}} \frac{\sqrt{M}}{T^{3/2} S_{\mathrm{out}}} . \tag{10}$$

The 1+ ionization cross-section is provided by the Lotz formula:

$$\sigma_{+} = 4.5\times 10^{-14} \sum_{nl} \frac{\ln(E_{\mathrm{e}}/E_{1+,nl})}{E_{\mathrm{e}} E_{1+,nl}} \ [\mathrm{cm}^2], \tag{11}$$

And finally V_{source} is the volume of the source, $I_{\mathrm{n_in}}$ is the isotope current entering the source, and f_{extr} is the extraction efficiency once an ion is produced.

By folding together Eqs. (5)–(11), the ionization efficiency can be expressed as

$$\varepsilon = 2.33\times 10^{4} f V_{\mathrm{source}} A \exp(-W/kT) l \frac{\ln(U/\phi_{\mathrm{i}})}{U\phi_{\mathrm{i}}} \frac{\sqrt{M}\sqrt{T}}{S_{\mathrm{out}}} . \tag{12}$$

Equation (12) can be tested for different types of FEBIAD ion sources, and different parameters of operation. Values of V_{source} and S_{out} are directly deduced from the mechanical design. W is a known property of the cathode (classically made of graphite or refractory metals), provided the surface is free of contaminants. The temperature of the source T is deduced from a calibration curve reporting T measured at the surface of the cathode, with respect to the applied electrical heating power. M and ϕ_{i} are the mass and ionization potential of the isotope under consideration. U is the anode potential which sets the energy of the electrons. The last parameter f cannot easily be independently computed. However, it directly reflects if the design of the source and its mode of operation are chosen to optimize the electron–ion interactions and ion extraction before losses on the walls of the chamber. It is typically affected by the grid transparencies for the emitted electrons, by the internal pressure, by the configuration of the extraction geometry, by the penetration of the electrostatic field of the beam extraction electrode and by the magnetic field deflecting the electron trajectories. The efficiencies are plotted in Figs. 13 and 14 for a VADIS source, with the characteristic exponential increase in a first part, the saturation in a second region, and an eventual decrease in a last part.

As can be seen in Fig. 13, above about 2000°C, the electron emission of the cathode becomes space-charged-limited and follows the Child–Langmuir Eq. (8). In this case, Eq. (12) can be re-expressed accounting for the proper formulation of the electron current j_{e}:

$$\varepsilon = 5.42\times 10^{-2} f V_{\mathrm{source}} \frac{U^{3/2}}{d^2} l \frac{\ln(U/\phi_{\mathrm{i}})}{U\phi_{\mathrm{i}}} \frac{\sqrt{M}}{S_{\mathrm{out}} T^{3/2}} . \tag{13}$$

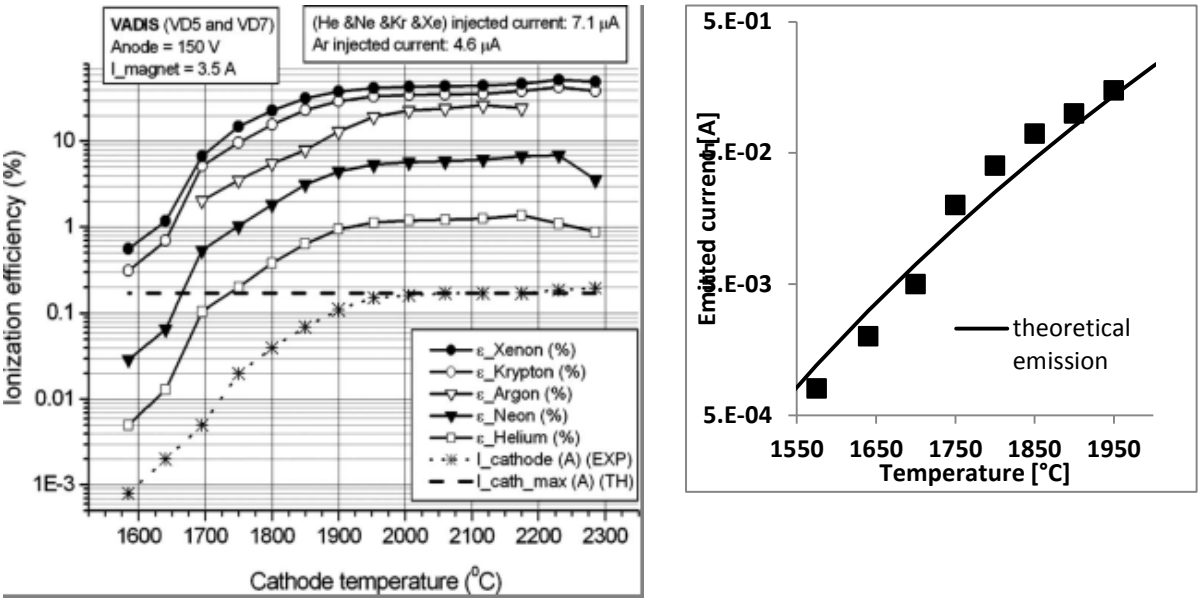


Fig. 13: Left: efficiencies measured for a series of noble gases in a VADIS source, electron emission from the cathode, and theoretical limitation predicted from Child–Langmuir space-charged emission (from Ref. [18]). Right: electron emission superimposed upon the theoretical prediction of Eq. (9), with A = 600 mA mm^{-2} K^{-2}, W = 4.12 eV and S_{cathode} = 1.13 cm^2.

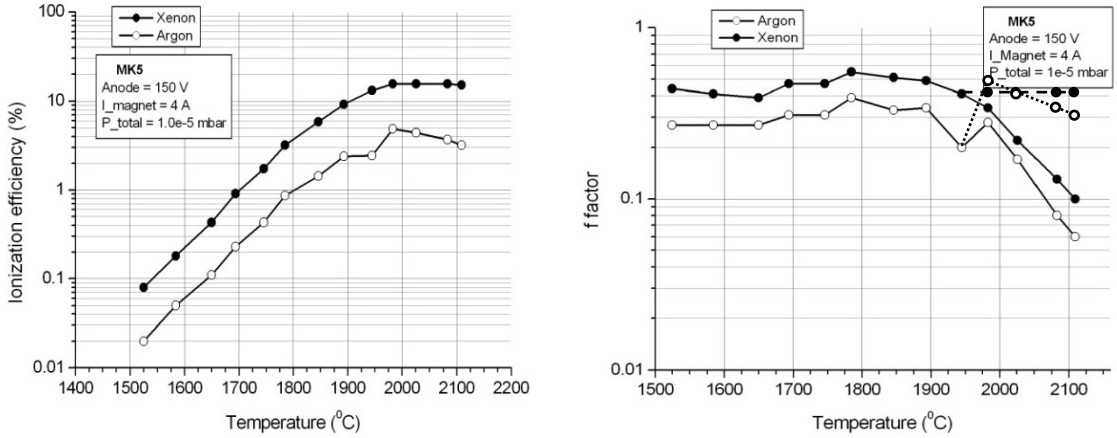


Fig. 14: Efficiencies measured in a MK5 source, and f factor deduced from Eq. (12). The dashed lines are obtained when the cathode emission is space-charge-limited, Eq. (13) [18].

The deduced f factors are also shown in Fig. 14. When the electron emission is properly measured or computed with appropriate models, this factor gives important insights into the quality of the design, and an optimal range of operational parameters. In particular, it shows if the design of the source can be improved and what are the elements that need modification.

5.3 Specific applications, future developments

The FEBIAD sources are versatile sources, compact, fast, and provide high ionization efficiencies for condensable elements. They have also been used to produce molecular beams of condensable elements such as SnS^+, $SeCO^+$ and AlF^+ [19]. They can be adapted to match specific requirements such as an improved beam quality from a reduced extraction hole surface area, a reduction of CO_2^+ beam impurities replacing graphite parts, and so forth. Interesting developments are the on-going coupling of these cavities with laser sources, and should provide exciting developments in the near future.

6 RF heated sources

6.1 Historical elements

This last section covers radio-frequency-heated sources applied for radioactive ion beams. This type of source was applied for radioactive ion beam production much more recently than the other two preceding classes. For instance, ECR sources, extensively described in these series of lectures for stable beams, have been proposed and developed for operation in different facilities. Such sources were developed for stable beams in the 1970s. They were applied for radioactive beam generation about 20 years after, at GANIL in France, in Louvain-La-Neuve in Belgium and at the TISOL facility in Canada [20, 21]. These sources are mostly used to ionize molecular beams that are volatile at room temperature. The source operated in Louvain-La-Neuve to produce intense beams of oxygen and nitrogen isotopes is shown in Fig. 15. This type of source, however, presents a number of challenges for operation in the harsh environments found at target stations of the ISOL facilities. As can be seen in Fig. 15, this was avoided by positioning the source in a remote position with respect to the target position. However, this is at the expenses of the speed of transport of the radioactive isotopes from the target to the ion source. The combined graphite target and Nanogan 3 ion source is shown in Fig. 16.

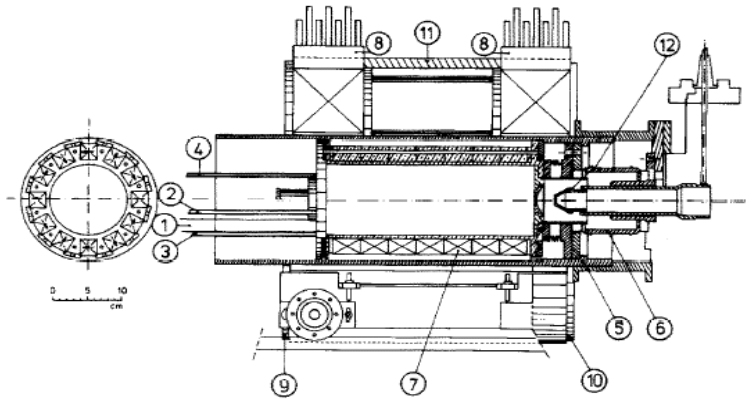


Diagram of the ECR ion source: micro wave guide (1), cooling water in/outlet (2, 3), gas input (4), HV insulator (5, 6), permanent magnetic bars (7), solenoids (8), iron yoke (9, 10, 11), extraction electrode (12).

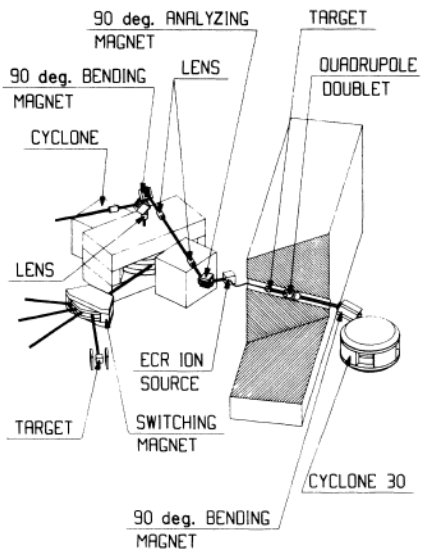


Fig. 15: Top: ECR source operated at CRC, Louvain-La-Neuve (from Ref. [20]). Bottom: layout of the facility [21].

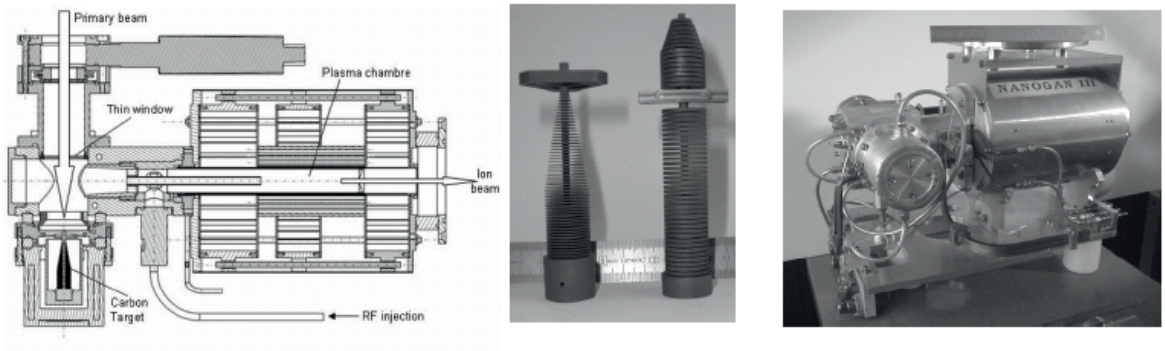


Fig. 16: Target and Nanogan III ECR source operated at SPIRAL I, GANIL facility

While these sources have demonstrated the production of noble gases, molecular oxygen, carbon and nitrogen isotope beams, the main factors that affect the ionization efficiencies are the distribution along different charge states, the residence time in a larger plasma volume, and possible losses of the isotopes on the chamber walls. Two different lines of development have since been followed. One is the development of compact 1+, 2.45 GHz ECR sources, such as those reported in Ref. [22]. The second line of development was the implementation of a so-called charge breeding scheme, in which a 1+ beam is injected in an ECRIS or EBIS charge booster or charge breeder device. This second line of development is specifically described in the course given by F. Wenander; the interested reader should therefore refer to his contribution on charge breeding.

6.2 Mode of operation and underlying physics principles

These sources are operated following the same principles as their counterpart operated for stable beams. Here also the interested reader is advised to look in publications describing ECR ion sources, one of those is R. Geller Electron Cyclotron Resonance ion Source and ECR Plasmas, (IOP Bristol and Philadelphia, 1996). The main point of interest when applied to the production of radioactive ion beams is the residence time of the isotopes, coupled to the ionization efficiency. Studies have been performed in an attempt to link plasma chamber volume, residence time and ionization efficiency dependent on isotope half-life, as shown in Fig. 17 for Ne isotopes. Obviously more systematic studies are required, especially for molecular beams of short-lived isotopes.

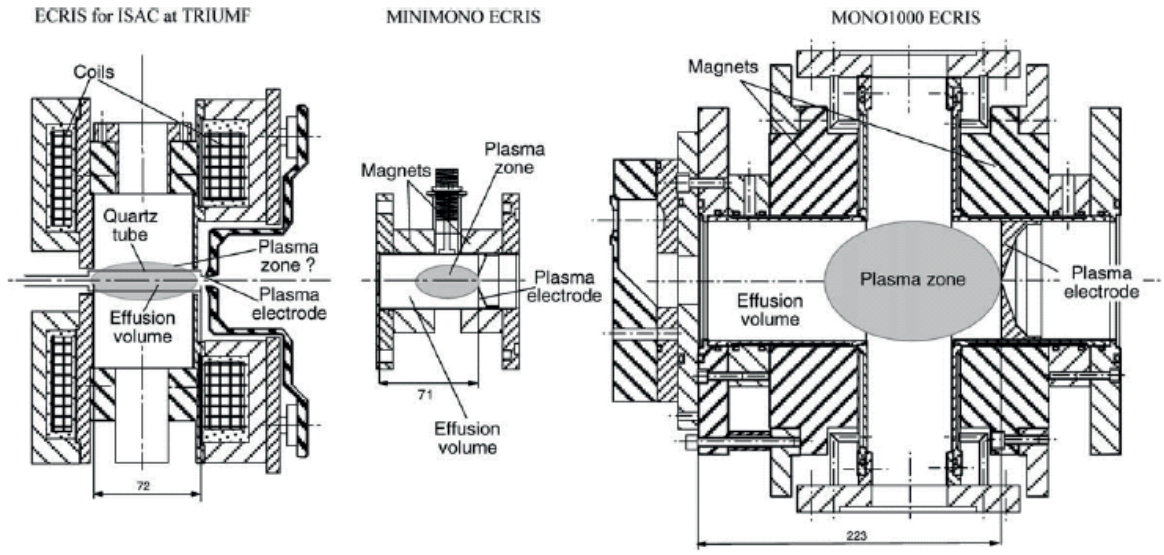
6.3 Specific applications, future developments

Development of RF-heated plasma sources for radioactive ion beams is a very active field in our community. Production of radiation-hard ECR sources is progressing well. Furthermore, new concepts of plasma sources have appeared during the past few years, for instance, low-frequency inductively coupled RF sources or Helicon-type RF sources. These compact and robust sources are expected to provide interesting results for the production of different molecular beams, in an attempt to produce plasmas that are more appropriate for the ionization of the molecules without destruction of the molecular bonds [24].

7 Conclusion

I hope I have been able to share some of the exciting recent developments that our field has witnessed over the past few years. Many new facilities are under conception or construction around the world. At the same time, new types of radioactive beams are produced each year by the ISOL method, reflecting the progress in the various fields linked to beam production. The family of robust, compact, simple, rapid and efficient sources suited for these facilities will almost certainly continue its expansion.

Progress in the understanding of the basic physical phenomena and the development of new concepts will probably lead to increases in quality and intensity of the beams.



Comparisons between expected and measured ionisation efficiencies with four 2.45-GHz ECR ion sources, for different isotopes of Ne

	ISAC TRIUMF	MINIMONO GANIL	MONO1000 GANIL	MONOBOB GANIL
Volume (cm^3)	18	**117**	**1500**	**687**
Exit diameter (mm)	**5**	**4**	7	**5**
Transient time (ms)	**<40**	34	15	81
Ion. efficiency (%)				
$^{20}Ne^+$ (stable)	11.5	**25**	**92**	**53**
$^{22}Ne^+$ (9.25 s)	**12**	25	92	53
$^{18}Ne^+$ (1.67 s)	11	25	91	51
$^{17}Ne^+$ (0.109 s)	10.5	19	81	30

Fig. 17: Comparison of the different 2.45 GHz ECR sources at GANIL and TRIUMF (from Ref. [23])

References

[1] O. Kofoed-Hansen and K.O. Nielsen, *Dan. Mat. Fys. Medd.* **26** (1951) 7.

[2] K.O. Nielsen, *Nucl. Instrum.* **1**(6) (1957) 289.

[3] J. Lettry, Review of ion-source developments for radioactive ion-beam facilities, in Proc. 1999 Particle Accelerator Conference, New York (1999), pp. 92–96.

[4] R. Klapisch *et al.*, *Nucl. Instrum. Meth.* **53** (1967) 216.

[5] I. Langmuir and K.H. Kingdom, *Proc. R. Soc. Lond.* **A107** (1925) 61.

[6] R. Klapish *et al.*, *Nucl. Instrum. Meth.* **38** (1965) 291.

[7] R. Klapish *et al.*, *Phys. Rev.* **C12** (1975) 644; *Phys. Rev. Lett.* **29** (1972) 1254.

[8] G.J. Beyer *et al.*, *Nucl. Instrum. Meth.* **96** (1971) 437; M. Remy and R. Haug, *Rev. Sci. Instrum.* **41** (1970) 650; P.G. Johnson, *Nucl. Instrum. Meth.* **106** (1973) 83.

[9] H.L. Ravn *et al.*, *Nucl. Instrum. Meth.* **123** (1975) 131.

[10] B. Vosicki *et al.*, *Nucl. Instrum. Meth.* **186** (1981) 307; M. Menna *et al.*, *Nucl. Instrum. Meth.* **B266** (2008) 4391; F. Schwellnus *et al.*, *Nucl. Instrum. Meth.* **B267** (2009) 1856.

[11] R. Kirchner, *Nucl. Instrum. Meth.* **186** (1981) 275; R. Kirchner, *Nucl. Instrum. Meth.* **A292** (1990) 203.

[12] N. Sato *et al.*, *Rev. Sci. Instrum.* **84** (2013) 023304.

[13] B. Marsh, these proceedings.

[14] H.L. Ravn *et al.*, *J. Inorg. Nucl. Chem*. **123** (1975) 131.

[15] R. Mueller and H. Wagner, in Proc. Int. Conf. on Electromagnetic Isotope Separators and the Techniques of Their Applications, Marburg, 1970, p. 461.

[16] R. Kirchner and E. Roeckl, *Nucl. Instrum. Meth.* **139** (1976) 276.

[17] S. Sundell and H. Ravn, *Nucl. Instrum. Meth.* **B70** (1992) 160.

[18] L. Penescu, PhD thesis, Universitatea Politehnica, Bucharest (2009); L. Penescu *et al.*, *Rev. Sci. Instrum.* **81** (2010) 02A906.

[19] E. Hagebo *et al.*, *Nucl. Instrum. Meth*. **B70** (1992) 165.

[20] P. Decrock *et al.*, *Nucl. Instrum. Meth*. **B58** (1991) 252; P. Bricault *et al.*, *Rev. Sci. Instrum.* **63** (1992) 2494; K. Jayamanna *et al.*, *Rev. Sci. Instrum.* **69** (1998) 756.

[21] D. Darquennes *et al.*, *Phys. Rev.* **C42** (1990) R804.

[22] P. Jardin *et al.*, *Nucl. Instrum. Meth*. **B204** (2003) 377.

[23] C. Huet-Equilbec *et al.*, *Nucl. Instrum. Meth*. **B240** (2005) 752.

[24] P. Suominen and T. Stora, Target and ion source development, *Isolde Newsletter* (2011), p. 20, http://isolde.web.cern.ch/sites/isolde.web.cern.ch/files/April%202011.pdf; M. Kronberger *et al.*, *Nucl. Instrum. Meth*. **B** (2013), available online.

Charge Breeding of Radioactive Ions

F.J.C. Wenander
CERN, Geneva, Switzerland

Abstract
Charge breeding is a technique to increase the charge state of ions, in many cases radioactive ions. The singly charged radioactive ions, produced in an isotope separator on-line facility, and extracted with a low kinetic energy of some tens of keV, are injected into a charge breeder, where the charge state is increased to Q. The transformed ions are either directed towards a dedicated experiment requiring highly charged ions, or post-accelerated to higher beam energies. In this paper the physics processes involved in the production of highly charged ions will be introduced, and the injection and extraction beam parameters of the charge breeder defined. A description of the three main charge-breeding methods is given, namely: electron stripping in gas jet or foil; external ion injection into an electron-beam ion source/trap (EBIS/T); and external ion injection into an electron cyclotron resonance ion source (ECRIS). In addition, some preparatory devices for charge breeding and practical beam delivery aspects will be covered.

1 Introduction and motivation

It is believed that around 6000 nuclei exist in the nuclear landscape, and presently some 3000 have been observed experimentally (of these, less than 10% are stable). The radioactive isotopes are currently the focus of nuclear physics research. In general, there are two complementary ways to generate good-quality beams of exotic nuclei, the in-flight (IF) separation method [1] and the isotope separation on-line (ISOL) technique [2]. The two methods inherently produce radioactive ions with beam energies at the opposite ends of the spectrum, IF in the GeV and ISOL in the keV region, as shown in Fig. 1. Nevertheless, the in-between region, from 0.1 to 10 MeV/A,[1] holds several opportunities for interesting physics [3], such as:

- Coulomb excitation;
- few-particle transfer, e.g. (d,p), (^{9}Be,2α), (p,γ); and
- fusion reactions at the Coulomb barrier.

This has pushed the technical development, and, with the introduction of post-acceleration (i.e. consecutive acceleration of ISOL-produced low-energy elements to higher energies), the energy gap could be bridged. To boost the ISOL beam energy, it is not sufficient to raise the electrostatic potential of the primary ion source (limited to a few hundred kilovolts). Therefore, conventional radiofrequency (RF) acceleration techniques such as cyclotrons or linacs have to be used. In both cases, the final beam energy per nucleon (MeV/A) is governed by the ion charge state: for the cyclotron with bending limit K, it is $K(Q/A)^2$; and for a linac with length L_{linac} and average acceleration field E_{acc}, it is $E_{\mathrm{acc}}L_{\mathrm{linac}}Q/A$. The following discussion will focus on linacs, as this kind is predominantly used for post-acceleration at present-day facilities.

[1] Note that the term MeV/A, energy per nucleon, is commonly used and the unity is related to the particle velocity. The total acceleration energy is obtained by multiplying with A.

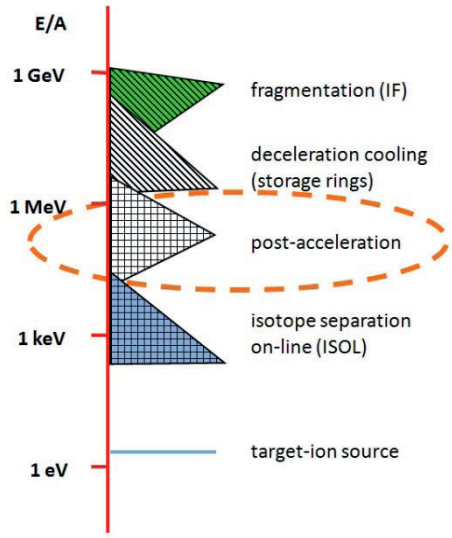


Fig. 1: Energy regions for nuclei produced with IF and ISOL techniques. The beam energy of IF ions can be reduced in storage rings, and the energy of ISOL beams increased in post-accelerators.

The ion charge is furthermore important for a linac as a high A/Q ratio imposes a low f_{RF} (a few tens of megahertz for very heavy ions with charge 1^+) to achieve adequate

- transverse focusing (focal strength ~ $(Q/A)/f_{RF}^2$) and
- period length (L_{period}) of the initial RF structure part as the extraction velocity from the primary source is limited.[2]

The transverse dimension r_{linac} of the cavities generally increases for lower frequencies, thus the technical design challenges and the cost increase with ion A/Q. In conclusion, the cost of the post-accelerator depends very much on the initial charge state, as a low A/Q leads to a short linac length with small transverse dimensions, and the cost scales approximately as $L_{linac}(r_{linac})^p$, where $1 < p < 2$.

As will be seen in the following section, most ISOL systems produce 1^+ ions. The traditional method to increase the charge of ions – the foil or gas-jet stripping technique – was challenged some 15 years ago by novel schemes for charge breeding. The novelty was the transformation from 1^+ to Q charged ions inside an electron-beam ion source/trap (EBIS/T) or electron cyclotron resonance ion source (ECRIS) by electron–ion collisions. These charge breeders are located in the low-energy part of the machine before the accelerating structures. Because of the capability of these charge-breeding devices to produce highly charged ions, A/Q ratios between 3 and 9 are easily obtained. The layout of a combined ISOL and post-accelerator system is shown in Fig. 2, including an A/Q analyser after the breeder, which selects the correct A/Q value and separates the radioactive ions from stable ion beams formed from the residual gas.

2 ISOL beam parameters and breeder criteria

The charge breeder acts a link between the ISOL production stage and the post-accelerator. It has to accept the different beams effectively and deliver charge-bred beams with the right characteristics to the post-accelerator. The properties of the radioactive ion beams from an ISOL facility may vary greatly depending on the element, type of ion source and if preparatory devices are being used or not.

[2] For example: $A = 220$, $Q = 1$, $U_{extr} = 100$ kV $\Rightarrow v_{extr} = 3 \times 10^5$ m s^{-1}; $L_{period} = 2$ cm $\Rightarrow f_{RF} \sim v_{extr}/L_{period} = 15$ MHz. Unless otherwise indicated SI units are used throughout the paper.

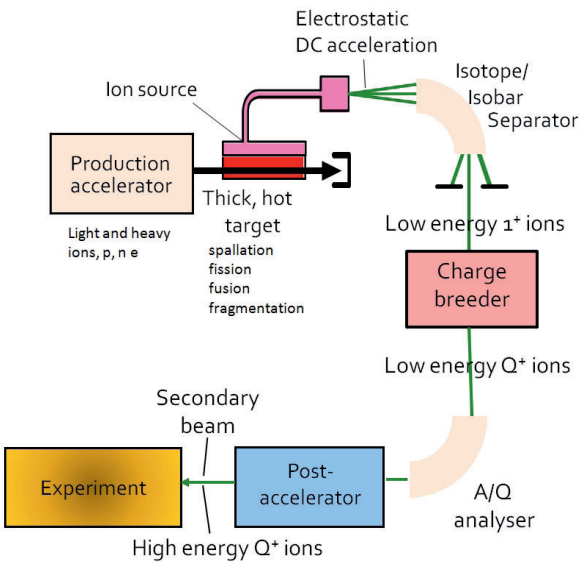


Fig. 2: Schematic layout of an ISOL production system with subsequent charge breeder and post-accelerator, followed by an experimental set-up.

2.1 What goes in …

The ion mass stretches from light $A = 6$ (even ^{4}He is used for test purposes) to very heavy elements with $A \sim 238$. As the neutron drip-line is further away from stability than the proton drip-line, many of the ions have a large neutron excess. This means that, to reach the required A/Q ratio, charge breeding to a relatively high charge state is imposed. The ion current spans an extensive region from just a few ions per second to $>10^{11}$ ions per second, where the upper value may increase with next-generation ISOL facilities. The ion charge is predominantly 1^+ but sometimes 2^+ or even higher charges can be produced by the initial ion source on the target and transferred to the charge breeder. The beam energy is a few tens of keV (total energy) with an energy spread of a few eV. If the driver beam is continuous, the radioactive ions are released continuously from the primary target; although, with a pulsed driver, a semi-continuous, but intensity-modulated, time structure is introduced. The release time, ranging from some tens of milliseconds to minutes or longer, is mainly governed by the diffusion of the radioactive atoms in the target, the transport process from the target to the ion source, and the half-life of the ion itself. An example of a time release curve for ^{8}Li is given in Fig. 3.

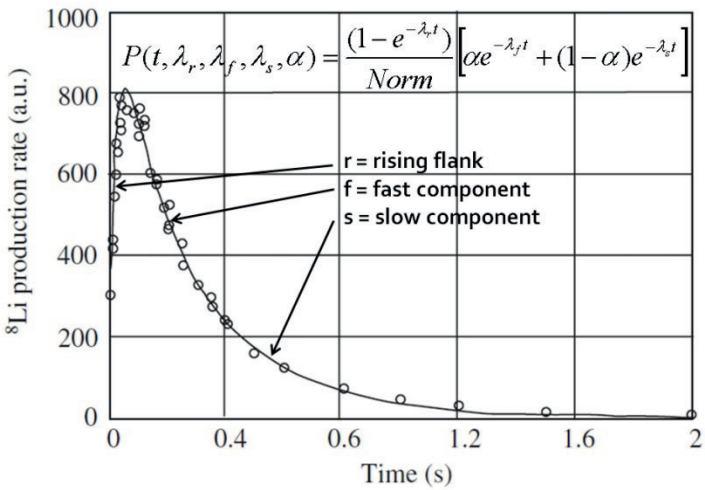


Fig. 3: Release curve of ^{8}Li ions from a tantalum foil target. A proton pulse impinges on the target at t = 0 s. An empirical fitting formula is included. Based on [4].

When one moves away from the valley of stability, the half-life of the nuclide to be investigated is typically reduced. Close to the drip-lines it can be as short as a few milliseconds or even shorter. Nevertheless, the production and ionization processes in the ISOL system set a lower limit of around 10 ms. Finally, the extracted beam might not be isobarically clean but might have a superposition of isobaric contaminants. They can be abundant, even dominant, particularly far from stability. Several techniques to suppress them within the ISOL system exist, such as using resonant laser ionization [5], chemical suppression and molecular sideband beams (see section 7.2), to mention but a few. The transverse emittance of the 1^+ radioactive beam is strongly linked to the type of ion source (e.g. plasma or surface ionizer) and its design. The beam properties of the ISOL system are summarized in Table 1.

Table 1: Summary of ISOL beam parameters.

Entity	Value	Comments
Ion mass	4 to ~238	He to U
Intensity	Few to >10^{11} ions/s	Large dynamic range
Charge	1^+	Sometimes 2^+, 3^+, …
Energy	Several tens of keV	Total beam energy
Energy spread	Few eV	
Temporal structure	CW or quasi-CW	Driver beam – CW or pulsed
Transverse emittance	10–50 mm mrad	90% at 50 keV
Half-life	>10 ms	Limited by ISOL system
Selection	Not necessarily isobarically clean	Make use of resonant laser ionization, for example

2.2 … and what comes out

The charge breeder has to produce multiply charged ions. Extremely high charge states are not required in most cases, as long as the upper A/Q limit of the accelerator is fulfilled (typically between 4.5 and 9), although exceptions exist, as will be pointed out below. The breeding efficiency η_{breed}, defined as the number of extracted divided by the number of injected particles, or expressed in terms of electrical currents as

$$\eta_{\text{breed}} = \frac{I_{\text{extr}}(Q)}{QI_{\text{inj}}(1^+)}, \tag{1}$$

where I_{inj} and I_{extr} are the injected and extracted electrical currents, is of utmost importance. At least 10% is expected, as many isotopes are rare and difficult to produce. Moreover, a high efficiency leads to reduced machine contamination. To further reduce the losses, the breeding time should be short (ideally down to 10 ms) for short-lived ions. Apart from these criteria, a sufficient throughput capacity should be provided for high-intensity beams (up to microamps), some serving as secondary driver beams for the production of even more neutron-rich nuclei or superheavy elements. Another important factor is the purity of the extracted beam, as stable beam contaminants from the breeder may disturb the experiment, as will be described later. Ideally, the breeder system should suppress isobaric and molecular contaminants coming from the incoming ISOL system. To trap the injected ions efficiently, the transverse acceptance should be large, covering the emittance of the ISOL beam. The extracted beam emittance, on the other hand, should be as small as possible to achieve a sufficient resolution in the subsequent A/Q separator and to fit into the acceptance space of the accelerator. It is possible that the output emittance of the beam from the charge breeder is smaller than the incoming beam emittance, owing to the cooling effect of the ions in the charge breeder plasma. Depending on the

post-accelerator and experiment, a continuous wave (CW) or pulsed extracted beam is requested. For low intensities, a pulsed beam at the experiment might be preferable to increase the signal-to-noise ratio, while high-intensity beams should be delivered as CW in order to avoid detector saturation.

3 Atomic physics processes for multiply charged ions

The removal of an electron from an atom or molecule requires an electric field in excess of 10^{10} V m^{-1} for field ionization, only achievable within atomic distances typically reached in collisions with charged particles (or intense lasers). The conservation of energy and momentum favours electrons as the most efficient ionizing particles compared to protons and photons, for example. The electron-impact ionization cross-sections σ are of the order of 10^{-16} cm^2 for neutral atoms, approximately matching the geometrical size of the atoms. There are two different ways of producing multiply charged ions:

- in a single collision, where many electrons are removed from the ion (double ionization, triple ionization, etc.); and
- by multistep ionization, where only one electron is removed per collision and high charge states are produced in successive different collisions.

In single-step ionization, the incident electron must have an energy of at least the sum of all the binding energies of the removed electrons, whereas in multistep ionization they only have to exceed the energy of each electron removed. In practice, multistep ionization, expressed as

$$\mathrm{e}^- + \mathrm{A}^{i+} \rightarrow \mathrm{A}^{(i+1)+} + 2\mathrm{e}^- \tag{2}$$

is the only feasible route to high-charge-state ions, but the process takes time. The time depends on the plasma density and the ionization cross-section, and must be shorter than the ion lifetime in the plasma, limited by the confining field(s) and recombination processes. The mean time $\bar{\tau}_Q$ required to reach a charge state Q can be calculated by adding up the individual ionization times, assuming no multiple ionization:

$$\bar{\tau}_Q = \sum_{i=1}^{Q-1} \bar{\tau}_{i\rightarrow i+1} = \frac{e}{j_\mathrm{e}} \sum_{i=1}^{Q-1} \frac{1}{\sigma_{i\rightarrow i+1}} . \tag{3}$$

According to the expression, valid for a monoenergetic electron beam, the breeding time can be adjusted by varying the electron current density j_e (A cm^{-2}).

The ionization cross-section can be estimated using Lotz's semi-empirical formula [6]. The simplified version given in Eq. (4) is valid for electron-impact beam energies E_kin much larger than the electron binding energies $E_{i,nl}$,

$$\sigma_{Q\rightarrow Q+1} = 4.5\times 10^{-14} \sum_{nl} \frac{\ln(E_\mathrm{kin} / E_{i,nl})}{E_\mathrm{kin} E_{i,nl}} \quad (\mathrm{cm}^2), \tag{4}$$

where $E_{i,nl}$ denotes the binding energy for electron i in subshell nl, the energies are given in eV and the sum is over all removable electrons in all orbitals nl. This typically leads to a maximum of the ionization cross-section when the electron energy is approximately three times the ionization potential. The Lotz formula gives approximate values with uncertainty of +40%/−30% for most species [7]. The formula is based on rather low-energy experimental data and should be applied with care for highly charged ions of high Z. For these cases, one should consider use of atomic numerical codes such as Flexible Atomic Code [8], paying attention to relativistic treatment for the highest ionization energies.

Ionization is a statistical process leading to a charge-state development as seen in the upper part of Fig. 4. The width of the distribution given in the lower part of the figure depends mainly on the Z of

the element, and usually 15–35% of the ions are found in the most abundant charge state. If the electron distribution is monoenergetic, its energy can be chosen just below an ionization level, thereby pushing a very high fraction of the ions into more or less a single charge state.

Electron-impact ionization is not the only process on-going inside a plasma. Several competing processes reduce the ion charge, for instance radiative and dielectronic recombination with electrons, and charge exchange with other atom/ion species. In addition, the ions may escape from the confinement. A differential equation describing the evolution of one charge state is given in Eq. (5). The model is suitable to be used for ions in an EBIS/T, where the electron-beam energy is known, partial neutral gas pressure can be estimated and the trapping potential easily calculated:

$$\begin{aligned}\frac{\mathrm{d}N_i}{\mathrm{d}t} &= n_\mathrm{e} v_\mathrm{e}\left[\sigma^{\mathrm{EI}}_{i-1\to i}N_{i-1} - \left(\sigma^{\mathrm{EI}}_{i\to i+1} + \sigma^{\mathrm{RR}}_{i\to i-1} + \sigma^{\mathrm{DR}}_{i\to i-1}\right)N_i + \left(\sigma^{\mathrm{RR}}_{i+1\to i} + \sigma^{\mathrm{DR}}_{i+1\to i}\right)N_{i+1}\right] \\ &\quad - n_0 v_\mathrm{ion}\left[\sigma^{\mathrm{CX}}_{i\to i-1}N_i - \sigma^{\mathrm{CX}}_{i+1\to i}N_{i+1}\right] - N_i R_i^{\mathrm{ESC}}.\end{aligned} \tag{5}$$

Here N_i denotes the number of ions with charge i, n_e and v_e the electron density and velocity, n_0 the neutral particle density of other atoms and v_ion the average ion velocity, defined by the ion injection energy, ionization and electron scattering heating [10]. The index EI denotes electron-impact ionization, RR radiative recombination, DR dielectronic recombination, CX charge exchange and ESC escape rate. CX between different ion charge states is suppressed due to Coulomb repulsion of ions, and therefore only charge exchange with neutrals is considered.

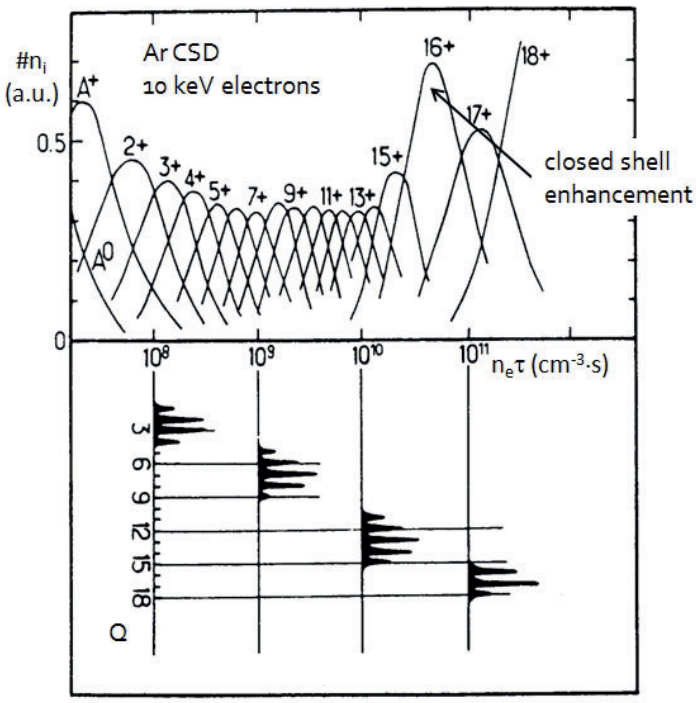


Fig. 4: Calculated relative atom and ion density of argon under bombardment of monoenergetic 10 keV electrons as a function of electron density times breeding time. The calculation assumes only single-step ionization. The He-like properties of Ar^{16+} reduce the cross-section for ionization from 16^+ to 17^+ compared to 15^+ to 16^+ and therefore enhances the abundance of 16^+. Based on [9].

The charge exchange process can be calculated using the Müller–Salzborn empirical formula [11] or the Selberg approximation [12]. The significance of charge exchange is defined by the ratio of EI and CX probabilities at given electron and neutral densities, electron and ion velocities and corresponding cross-sections, and is of little importance for low ion charge states.

For instance, for Pb^{20+} immersed in an electron-beam current density of 100 A cm^{-2}, the rate of electron-impact ionization equals the charge exchange rate with neutrals at a relatively high pressure of a few 10^{-7} mbar. For Pb^{70+}, on the other hand, the rates are similar at a significantly lower pressure of 10^{-11} mbar. Likewise, during the breeding to high charges, the ions gain significant momentum from elastic scattering by the highly energetic electrons, making them more likely to be lost from the confinement potential, and as a result of the higher ion velocity the CX rate is increased. The increase in ion charge state, however, augments the confinement force of the electric fields, and ion cooling actions can be present, both processes reducing the ion loss rate. Finally, to minimize the radiative recombination, the energy of the electrons should be considerably higher than the ionization potential, although that lowers the electron-impact ionization cross-section. For highly charged ions, the final population of the highest charge states is defined by the equilibrium between RR and EI at the given electron density and velocity. This is illustrated for the uranium case in Fig. 5, where the two graphs plotted indicate the residual gas pressure at which the cross sections for EI and CX, and RR and CX processes, are equal. Compared to the lead example above, a higher residual gas pressure can be tolerated as the electron current density is significantly higher.

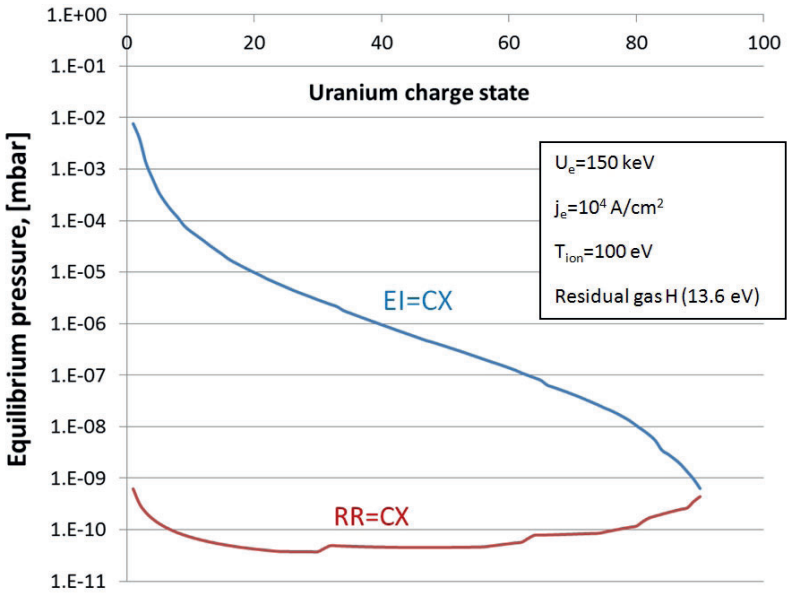


Fig. 5: The two graphs plot the residual gas pressure at which the cross sections for EI and CX, and RR and CX equal for charge breeding of uranium in a very high performance EBIS/T.

4 The gas-jet/foil stripper alternative

Although strictly not considered as a charge breeder, charge stripping in a dense, cold electron cloud is the classical way of increasing the charge state. This method is used at numerous heavy-ion injectors, for example, in the Pb injector chain at CERN. The idea of shooting ions with low charge through either a thin foil or a layer of gas in order to strip away further electrons has been used more or less since the invention of ion accelerators, and will therefore be introduced briefly as a reference for the other breeder types.

The ions need to be pre-accelerated to around 10 keV/*A* for efficient stripping in a gas jet, and up to several 100 keV/*A* for stripping in a solid foil. The stripping process may be repeated at different ion velocities in order to reach even higher charge states, as will be explained below. Usually He is used as gas-jet stripper medium, and Be, C, Al, Al_2O_3 or Mylar are used as foil material. Carbon foils have the advantage of being stable in vacuum at high temperatures, combined with good electrical and thermal conductivity. Furthermore carbon has the advantage of being the material with the lowest *Z* that can be fabricated into very thin foils to minimize multiple scattering and energy straggling [13].

The foil contains in its crystalline structure cold electrons with a density n_e of around 10^{24} cm^{-3}. The relative interaction velocity equals approximately the transit velocity v_{proj} of the accelerated ions (10^8 to a few 10^9 cm s^{-1}). The interaction time $t_{transit}$, of the order of 10^{-14} s, is given by the transit time, thus $n_e t_{transit} \sim 10^{10}$ s cm^{-3} and $n_e v_{proj} t_{transit} > 10^{19}$ cm^{-2}. During the interaction, two types of collisions are in competition: step-by-step ionization, and recombination through electron capture. At high speed, the former dominates. According to Bohr's criterion, an ion penetrating through matter retains only those electrons whose orbital velocity is greater than the ion velocity v_{proj}. Baron's empirical formula [14] can be used to calculate the mean equilibrium charge state $\bar{Q}$ for carbon foil stripping:

$$\bar{Q} = Z_{proj} C_1 \left(1 - C_2 \exp\left(\frac{-83.28\beta}{Z_{proj}^{0.447}} \right) \right), \tag{6}$$

where relativistic $\beta = v_{proj}/c$, $C_1 = 1$ for $Z_{proj} < 54$, $C_1 = 1 - \exp(-12.9 + 0.21 Z_{proj} - 0.001 Z_{proj}^2)$ for $Z_{proj} \geq 54$, and $C_2 = 1$ for ion energies $W_{proj} \geq 1.3$ MeV/A, $C_2 = 0.9 + 0.08 W_{proj}$ for $W_{proj} < 1.3$ MeV/A. Note that the average charge state is only dependent on v_{proj} and Z_{proj}. In Fig. 6 the attainable charge states as a function of the ion number for different particle velocities are plotted.

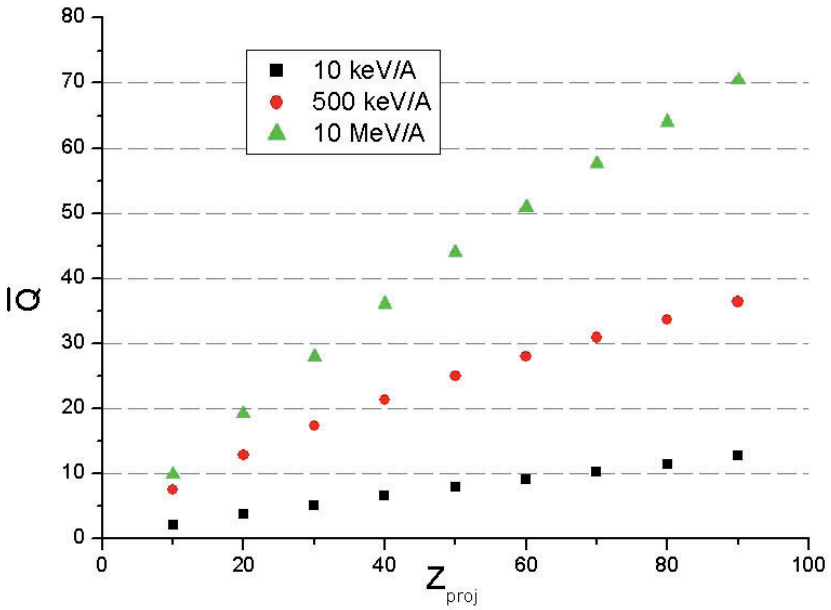


Fig. 6: Mean equilibrium charge state as a function of ion number for three different projectile energies after stripping in a carbon foil.

The charge-state distribution can be assumed to be Gaussian if no significant atomic shell effects are present and $\bar{Q}$ is not too close to Z_{proj}. The width σ_w of the distribution is given by

$$\sigma_w = 0.5 \sqrt{\bar{Q}\left[1 - \left(\frac{\bar{Q}}{Z_{proj}} \right)^{1.67} \right]} \quad \text{for } Z_{proj} < 54, \tag{7a}$$

$$\sigma_w = \sqrt{\bar{Q}\left[0.075 + 0.19\left(\frac{\bar{Q}}{Z_{proj}} \right) - 0.27\left(\frac{\bar{Q}}{Z_{proj}} \right)^2 \right]} \quad \text{for } Z_{proj} \geq 54. \tag{7b}$$

Light elements will have a narrow distribution, leading to a relatively high fraction in a single charge state, while heavy elements will have the ions distributed over more charges, as demonstrated in Fig. 7. Revised formulas for the stripping processes in foils and gases are given in Ref. [15].

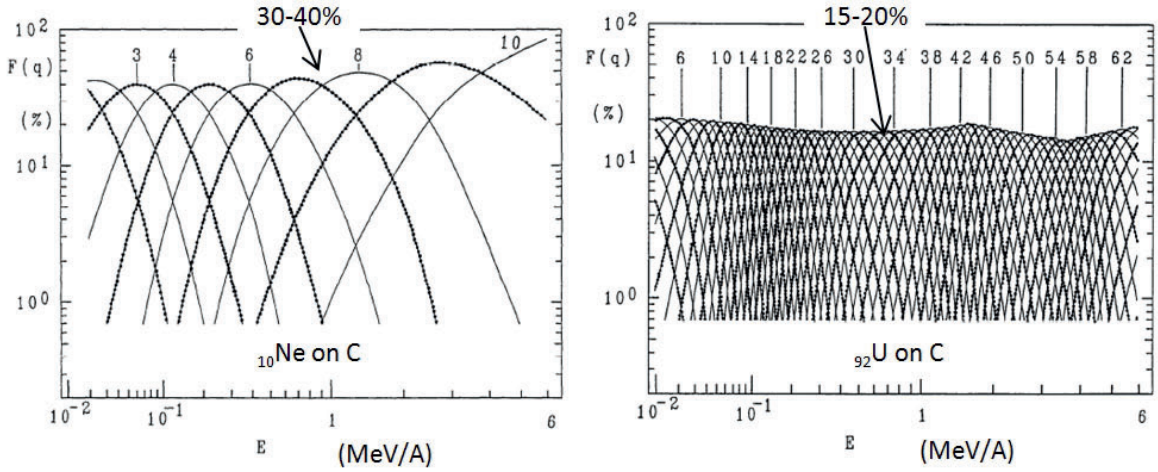


Fig. 7: Equilibrium charge-state distributions for Ne (left) and U (right) after stripping in a carbon foil. The plots show the fraction in different charge states as a function of the projectile energy. Based on [16].

The discussion above assumes that an equilibrium charge-state distribution has been reached, which means that the distribution at the stripper exit does not depend on the initial ion charge-state distribution, and nor does it change if the thickness of the target is further increased. It is important to know the equilibrium thickness, as the desired charge state is not attained if the foil is too thin, and excessive thickness will introduce further energy straggling and loss, and angular straggling.[3] From the approximation in Fig. 8, we learn that the carbon foil equilibrium thickness for a 500 keV/*A* beam is < 5 μg cm^{-2} (corresponding to < 25 nm). Manufacturing and handling of thinner foils is practically very difficult, and therefore gas targets are used for lower particle velocities.

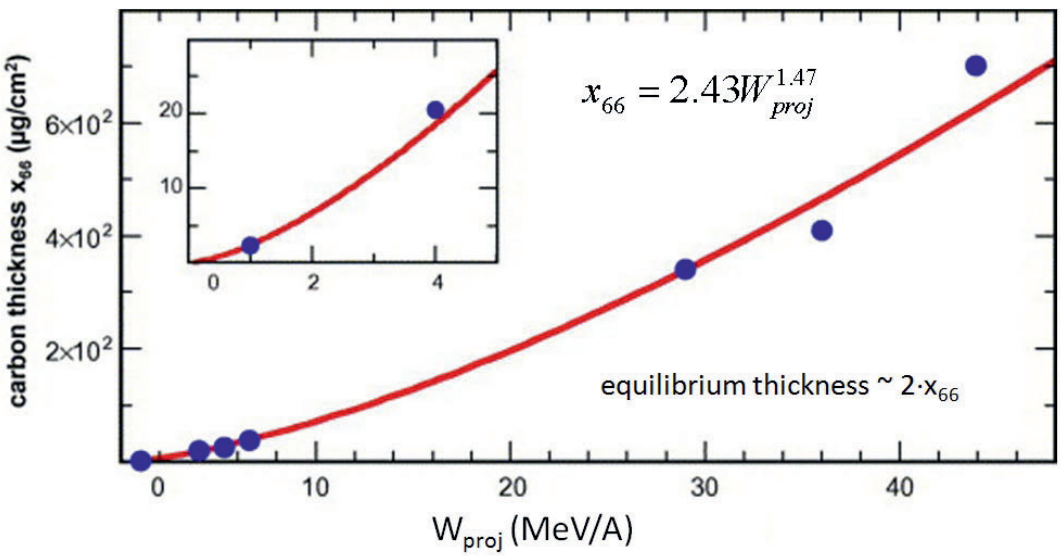


Fig. 8: Approximate equilibrium thickness for carbon stripper foil as a function of projectile energy. Based on [17].

The integrated thickness for a gas-jet stripper is a fraction of a microgram per square centimetre, suitable for ion velocities between 5 and 25 keV/*A*. Even though a relatively high fraction can be attained in a single charge state (35–55%), the charge is modest, ranging from 2^+ to 4^+ when going from Xe to U. It is interesting to note that in a solid stripper the collision frequency is larger than in a dilute gas-jet stripper, leading to less pronounced recombination effects due to Auger and radiative decays, and therefore a higher average charge for the same thickness.

[3] Calculated using, for example, the SRIM/TRIM package [18].

As the attainable charge in a stripping process is velocity-dependent, one would ideally like to introduce a stripper stage as soon as the increased particle velocity enables a higher charge state in order to make best use of the acceleration voltage. This would, however, lead to poor transmission due to the selection of a single charge state after each stripping stage. Instead, a typical layout makes use of one gas stripper followed by one or two stripper foils. See Table 2 for typical numbers on post-acceleration of ^{132}Sn using two stripper stages.

Table 2: Charge stripping of ^{132}Sn.

Entity	Transmission	Comments
Bunching efficiency[a]	65%	
Gas stripping to 2^+	55%	At 8 keV/A
Stripping foil to 23^+	20%	At 500 keV/A
Total transmission	7%	In a single charge state

[a] A bunch-rotating RF cavity is mandatory in order to generate a time focus at the stripper to minimize the longitudinal emittance growth due to energy straggling.

To increase the efficiency for heavy-ion beams, where multiple stripping stages are required, multi-charge-state acceleration is an option. Instead of selecting a single charge state after the stripping, a broader band of charges is accelerated, leading to a significant increase (a factor 2–3) in particle intensity for high-Z ions [19]. A charge variation acceptance $\Delta Q/Q \sim 10\%$ can be achieved in a linac. The disadvantages are complicated beam transport and an increased transverse and longitudinal emittance of at least a factor of 3 compared with single-charge-state acceleration. Precise longitudinal beam manipulations are also required, with an additional cavity for phase synchronization after each stripping station, and another one for beam re-collection before each additional stripping station [20].

To summarize, the foil (gas-jet) stripping method has several attractive features:

- It is simple, as it only uses passive elements and the foil arrangement itself is inexpensive.
- The ionization process equals the travel time through the foil, which is sub-picosecond.
- It is capable of handling very high beam intensities.
- It introduces no extra beam contamination to the radioactive beam.
- It is a very efficient method for the production of bare light ions.

The method has, however, a few drawbacks, summarized below:

- It needs pre-acceleration in order to attain sufficient velocity for stripping in either gas-jet or thin solid foil. The cost of the pre-accelerator adds to the overall project cost.
- The quasi-CW macrostructure[4] of the beam, determined by the release properties of the ISOL system, calls for a CW post-accelerator in order not to introduce losses.
- The 6D phase-space beam emittance increases due to multiple angular scattering and energy-loss straggling of the ions in the stripper material.
- The lifetime of the carbon foils is limited. Nevertheless, for most radioactive beams, the foil lifetime should be of little concern, as the intensity is low. It is not sufficient to reach sublimation temperatures (>150 W cm^{-2}), and a long operation time before radiation damage occurs is expected.

[4] A microstructure (several MHz) is imposed on the beam by the pre-buncher and the RF accelerating cavities.

5 ECRIS-based charge breeders

The concept of using an ECRIS as charge breeder is straightforward, as illustrated in Fig. 9. Singly charged ions are injected slowly into an ECRIS, and subsequently captured by the plasma, where hot electrons perform electron-impact step-by-step ionization of the ions. The highly charged ions migrate out of the plasma and are extracted by an extraction electrode before being further post-accelerated.

The ECRIS has to be of minimum *B* (min-B) field type, which means that the magnetic field increases in all directions from the centre of the magnet configuration, and the plasma is confined longitudinally by Helmholtz coils and radially by a strong permanent magnet multipole.

The electrons are confined by the magnetic field and the ions in addition through the charge-neutrality condition of the plasma. In a min-B field device, the ions can be confined for a longer time (some 100 ms) and high-energy electrons (exceeding 100 keV) are present.

In Fig. 10 the required electron energy and electron density multiplied by confinement time are indicated for different ion charges. As the confinement time is around 0.1 s, an electron density $n_e \sim 10^{12}$ cm^{-3} is necessary for charge breeding. The plasma cut-off frequency, that is, the lowest frequency that can penetrate a plasma with a certain electron density, is around 10 GHz, given by Eq. (8) – up to now ECRIS breeders have usually operated at a frequency of 14–18 GHz:

$$f_p = \frac{1}{2\pi}\sqrt{\frac{e^2 n_e}{\varepsilon_0 m_e}}. \tag{8}$$

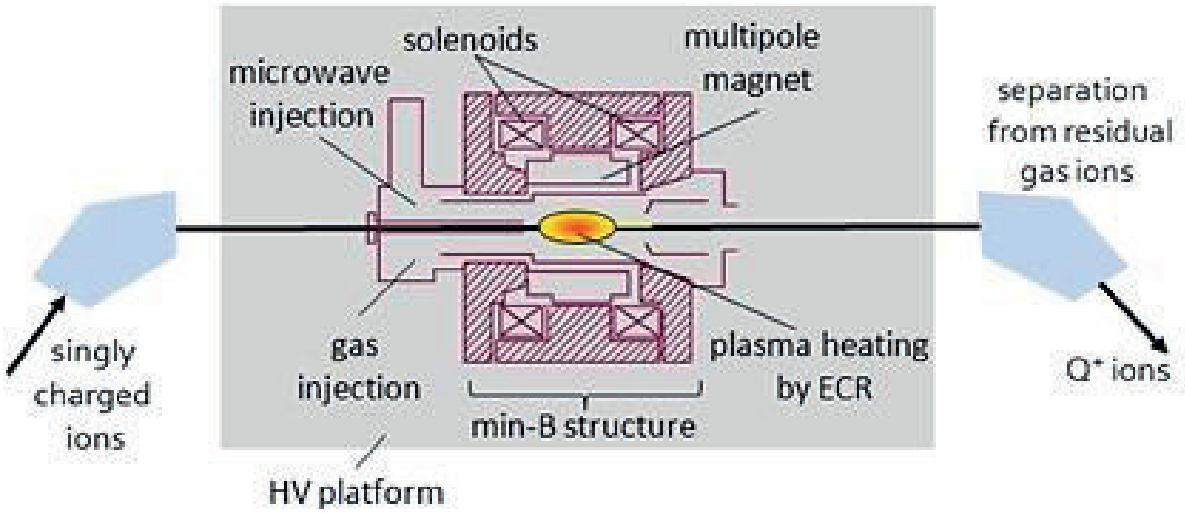


Fig. 9: Schematic concept making use of an ECRIS for charge breeding

Here m_e and e represent the mass and charge of the electron, respectively. As the electron density and confinement time are given, the breeding capacity and ion throughput can be estimated. For this we assume a plasma volume of 100 cm^3 and charge breeding to 10^+. Approximately 20% of the ions are found in charge state 10^+, with the rest in the neighbouring charge states.

Finally we assume that the radioactive ions occupy 10% of the total space charge, the rest being support gas ions. A total of 2×10^{12} radioactive ions per second would then be extracted, corresponding to 400 particle nanoamps. The estimation has been confirmed with tests where beam intensities of a few microamps have been injected and successfully charge-bred [21]. Such a large capacity is an important strength of the ECRIS charge breeder concept.

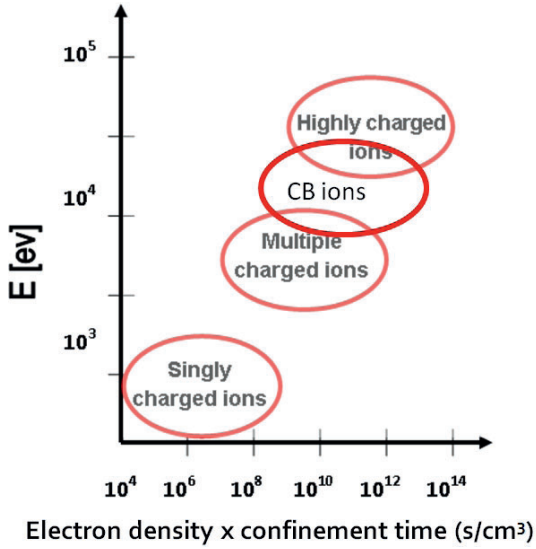


Fig. 10: Required electron energy and density multiplied by confinement time for production of different ion charge states.

The ions entering an ECRIS charge breeder are not trapped in a strong electrostatic potential from an electron beam, as is the case inside an EBIS/T. Instead, the singly charged ions are first slowed down electrostatically by positioning the ECRIS on a positive high-voltage potential, then they are thermalized inside the plasma region, and finally ionized to a higher charge state and thereby trapped by a weak electrostatic potential dip in the centre of the plasma. The general stopping procedure, the so-called plasma capture, is depicted in Fig. 11.

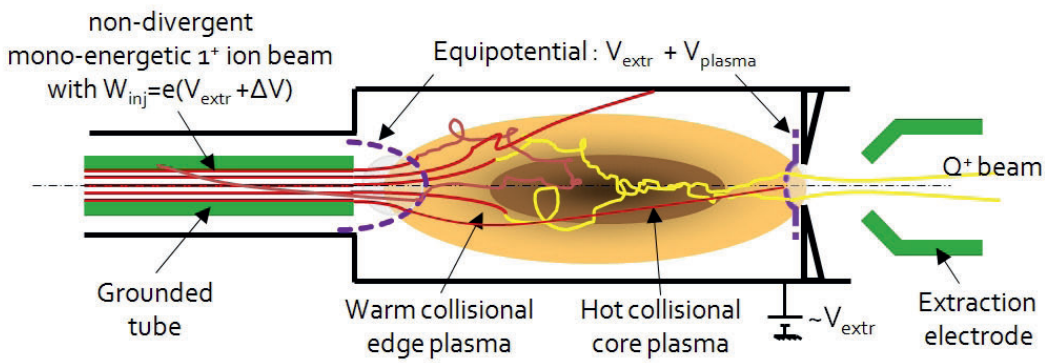


Fig. 11: The stopping process, or plasma capure, for injection of 1^+ ions into an ECRIS plasma. After trapping and stepwise ionization, the multiply charged ions subsequently migrate towards the extraction side and are accelerated by the extraction electrode.

An ion entering a plasma is deflected by the Coulomb force from the ions and electrons. In a strongly ionized, quiescent plasma, as found in a min-B structure ECRIS, the cumulative deflection due to small-angle, long-range scattering with many particles is larger than single two-body large-angle scattering. Each collision with a plasma electron or ion only gives a small-angle deviation to the injected particle, but the cumulative effect of many small-angle collisions eventually yields large-angle scattering of 90°. During the collision process, energy is transfered from the injected particle to the plasma charges.

In the complete collision damping theory proposed by Spitzer [22], the incident particles with velocity v_{inj} transfer momentum via collisions with the thermal electrons and ions of the magnetoplasma with velocities v_{therm-} and v_{therm+}, respecticely. For $v_{inj}/v_{therm+} < 10$ the ion–ion collision is the guiding mechanism, while for higher ratios the damping efficiency drops rapidly until the ratio

v_{inj}/v_{therm-} reaches values of the order of unity. The collisions between injected ions and the thermal electrons then become frequent. However, the mean free path length for this type of collision is too long to play a role in an ECRIS plasma. Thus, the governing slowing-down process is ion–ion collisions, and the slowing-down coefficient can be calculated with [23]

$$\frac{\langle \Delta v_{inj} \rangle}{\Delta t} = \frac{n_{therm+}}{2\pi\varepsilon_0}\left[\frac{Q_{inj}Q_{therm+}e^2}{m_{inj}\overline{v}_{therm+}}\right]^2\left(1+\frac{m_{inj}}{m_{therm+}}\right)R\left(\frac{v_{inj}}{\overline{v}_{therm+}}\right)\ln\Lambda\,. \tag{9}$$

Here n_{therm+} is the density of plasma ions, Q_{inj} and Q_{therm+} the charge states of the injected particle and the mean charge state of the plasma particles, respectively, m_{inj} and m_{therm+} their masses, $\ln\Lambda$ the Coulomb logarithm and $R(v_{inj}/\overline{v}_{therm+})$ a function of the injected particle and plasma ion velocities. We see that high plasma density and high ion charge states lead to a rapid slowing down. This has also been proven, as doubly charged Pb ions are easier to trap than singly charged. The function $R(v_{inj}/\overline{v}_{therm+})$ peaks when the velocities are similar, which has some consequences for the ion injection [24]. For example, if $^{85}Rb^+$ ions enter a plasma mainly constituted by ^{16}O ions from the support gas at a temperature of approximately 2 eV, their energy should ideally be 2 eV times 85/16, which equals ~10 eV. Ions with much higher injection energy, for instance 40 eV, would not be trapped. Following this argument, the injection energy of light ions should be low, around 2 eV for $^{11}Li^+$, which is challenging from a beam optics and magnetic stray-field point of view.

It should be noted that a few assumptions are essential for the capturing model to be valid, namely, the following:

- The intensity of the injected particles is low, meaning that they do not interact with each other. In reality, this means that the injected ion intensity should only be a fraction of the support gas ions inside the plasma.
- The injected particles interact with the plasma particles only via long-range cumulative plasma collisions.
- The plasma particles have a Maxwellian velocity distribution.
- The distance between 90° deviations is smaller than the plasma size.

Is the critical condition for 90° deflection inside an ECRIS fulfilled? The mean free path for a 90° deviation, $\lambda_{90°}$, can be estimated with the following expression [25]:

$$\lambda_{90°} \approx \frac{W_{inj}}{4\pi n_e Q_{inj}Q_{therm+}e^2\ln\Lambda}\,, \tag{10}$$

where W_{inj} is the energy of the injected particle inside the plasma. If we assume W_{inj} = 10 eV (as shown above to be reasonable), Q_{inj} = 1, Q_{therm+} = 10, $n_e = 10^{12}$ cm^{-3} and approximate $\ln\Lambda$ with 10, then we find that $\lambda_{90°}$ ~ 5 cm, thus comparable with the dimensions of the ECRIS plasma. The thermalization process of the injected ions takes a few ion–ion plasma collisions, and within some 10^{-4} s they have obtained a Maxwellian temperature distribution with an energy of some eV and are then confined like the support gas ions. Subsequently, the energetic electrons can strip the 1^+ ions through step-by-step ionizing collisions and transform them into higher-charge-state Q ions.

Based on the theory for the ion capture inside the plasma, one can conclude that the high-voltage adjustment of the ISOL ion source relative to the charge breeder is crucial. The layout of the electrical potentials is shown in Fig. 11 (left) and due to the often unknown plasma potential (i.e. the potential difference between the plasma and the plasma chamber; typically some tens of eV) inside the breeder, the voltage setting needs to be tuned. The normalized extracted charge-bred current is plotted as a function of the charge breeder potential for different primary ion sources in Fig. 12 (right). If the charge breeder potential is too low, meaning that the ions enter the plasma with a high velocity, only

noble gases are charge-bred. These ions can impinge at the plasma chamber wall without sticking and thereafter re-enter into the plasma region. Therefore the energy range of non-sticking elements is large (some 100 eV). Condensable, or metallic, ions, on the other hand, need to have the exact entrance energy (within some eV): if the energy is too low, they will not enter the plasma region; and if too high, they will hit the chamber wall and be stuck. The mean stay time (sojourn time) for atoms on the wall, given by Frenkel's law, is for Ar only some 10 ps, while for Ni it is around 100 years at room temperature.

The ability to adjust the A/Q value of the radioactive ion is essential in order to avoid superimposed stable beam contaminants. In an ECRIS the charge state is mainly dependent on the electron density n_e, the ion confinement time τ_{ion} and the electron energy distribution. These properties cannot be adjusted directly, so instead the accessible parameters are tuned. In practice, the injected RF power, and thus the power density, is varied as $Q \propto P_{RF}^{1/3}$. Secondly, the support gas pressure, which influences the number of cold electrons, the ion–ion cooling effect and charge exchange probability, can be changed. Finally, the magnetic field at the extraction side, B_{ext}, is varied as $Q \propto \ln(B_{ext})$. The latter effect can be explained by the magnetic-mirror effect, consisting in reflecting the charged particles back towards the weak-field region (illustrated in Fig. 13) with a force

$$F = -\frac{1}{2}\frac{mv_\perp^2}{B}\frac{\mathrm{d}B_z}{\mathrm{d}z} = -\mu\frac{\mathrm{d}B_z}{\mathrm{d}z},$$

where $v_\perp$ is the ion velocity perpendicular to the axial B field and μ the magnetic moment of the ion. The axial confinement is due to the conservation of total kinetic energy and the magnetic moment. A higher magnetic field reduces the loss rate of ions through the loss cone of the magnetic bottle.

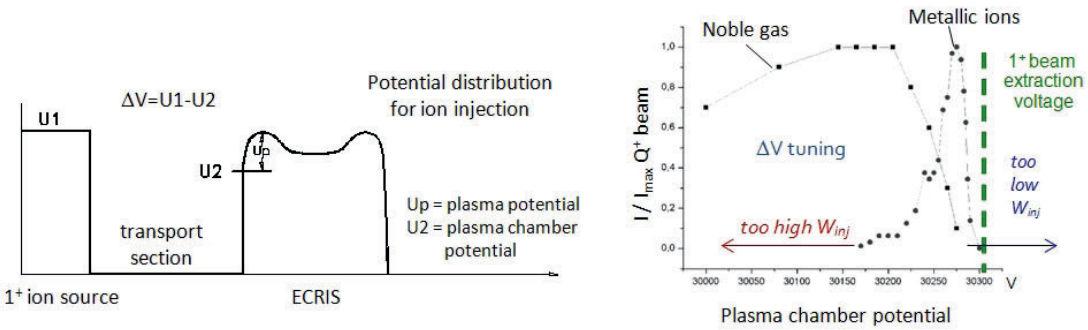


Fig. 12: (left) Potential distribution for the primary 1^+ ion source and the charge-breeding ECRIS. (right) Normalized extracted charge-bred beam as function of the high-voltage potential of the charge breeder for two different primary ion sources. The noble gases were produced in a plasma source with a certain plasma potential, while the metallic ions were produced in a surface ionizer (0 V plasma potential); thus the different optimal potential of the breeder.

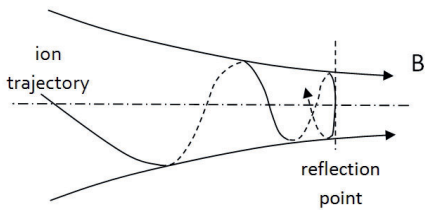


Fig. 13: The magnetic-mirror effect showing how a charged particle is reflected in a high B field region. To conserve the magnetic moment μ, the cyclotron frequency is increased while the cyclotron radius is decreased. In turn, the longitudinal momentum is reduced in order to conserve the total kinetic energy.

An ECRIS charge breeder delivers beams with A/Q values between 3 and 9 and an energy spread of a few eV. Breeding efficiences from a few up to 20% have been reported [26]. Apart from the radioactive beam, many other stable elements originating from the charge breeder are extracted,

with a total electric current of around 100 μA. These stable ions originate either from the support gas (often He, O, Ne or Ar) needed to supply the plasma with a permanent source of cold electrons, or from the plasma chamber. Electrons and ions escaping the multi-pole confinement can sputter chamber material into the plasma or desorb ions that have been implanted from previous runs (the so-called memory effect). To limit the A/Q contamination, several techniques related to the plasma chamber cleaning exist [27], and the use of isotopically enriched support gas is recommended.

The normal operation for an ECRIS is CW mode, meaning that the highly charged ions leak out from the plasma towards the extraction electrode. Likewise, the ion injection is also continuous. An ECRIS can, however, be operated in a different mode, the so-called afterglow mode, where the ions are extracted in a pulse coinciding with the switching off of the RF heating. The hot plasma electrons, strongly bound to the magnetic field lines, leave the plasma at the same moment as the RF heating is switched off. The ions, partly confined by the negative space charge of the electrons, are therefore expulsed in a short pulse by Coulomb repulsion. The pulse is a few milliseconds long and shown for Pb^{27+} in Fig. 13 (left). An ECRIS charge breeder operated in afterglow mode can consequently collect ions from a CW ion source, charge breed them and confine them for some 100 ms, and thereafter extract them in a compressed pulse. This has been demonstrated with Rb ions and the general result is illustrated in Fig. 14 (right). The drawback of this mode, also known as ECRIT mode [28], is a reduced transmission efficiency compared with CW operation.

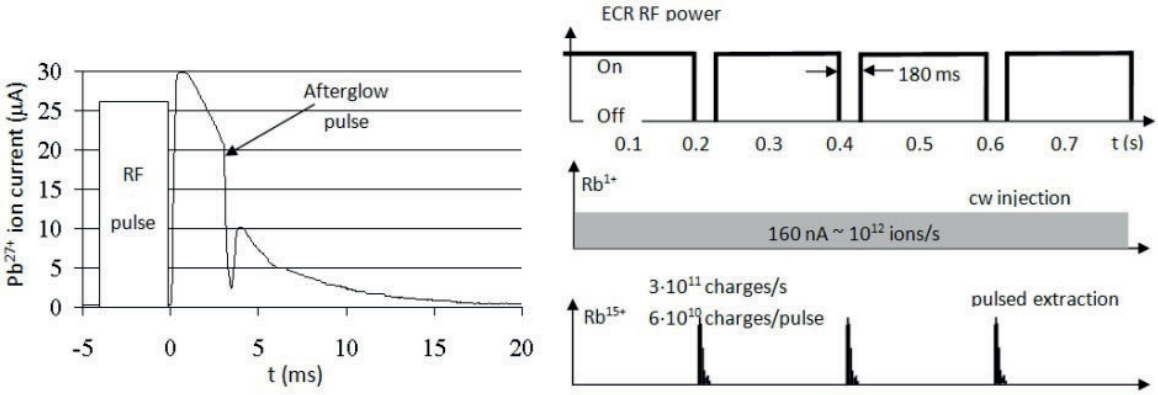


Fig. 14: (left) Pb^{27+} afterglow pulse occurring after the RF heating is switched off. Based on [21]. (right) Charge breeding of Rb^{1+} to 15^{+} using the afterglow mode. The pulsed extraction is suited for linacs operated in pulsed mode. Based on [28].

As demonstrated, a charge-breeding ECRIS has to operate with frequency > 10 GHz, and therefore with a magnetic field strength of a couple of teslas. The charge breeder follows the same magnetic field relations as for a high-Q, high-current ECRIS, that is, $B_{inj}/B_{ecr} \sim 4$, $B_{ext}/B_{ecr} \sim 2$, $B_{min}/B_{ecr} \sim 0.8$, $B_{rad}/B_{ecr} > 2$ and $B_{ext}/B_{rad} < 0.9$, where B_{ecr} is the magnetic field at ECR resonance, B_{inj} (B_{ext}) the field maximum at injection (extraction) side, B_{rad} the radial B field of the sextupole at the plasma chamber wall and B_{min} the minimum B field between the magnetic mirrors. Characteristic for an ECRIS charge breeder is the open end at the injection side, at which the gas and RF injections are usually located for a normal ECRIS. Instead, at this place, a cylindrical tube, connected to a low electrical potential relative to the plasma chamber, guides the ions into the breeder (see Fig. 10). The highest breeding performance is presently achieved by breeders with radial RF injection in the centre of the chamber as a symmetric B field is maintained at the injection region, therefore avoiding beam trajectory deflection of the injected ions, which can be important for light ions [29]. A good pumping conductance is essential to reduce excessive amounts of residual gas ions from the breeder.

The layout of an operational ECRIS charge breeder for radioactive ions is shown in Fig. 15. The use of a combined magnetic and electrostatic analyser will be discussed in section 7.2. The separation of the breeding stage from the production of the 1^{+} ions in the primary ion source means that the breeder is accessible with limited radio-protection restrictions.

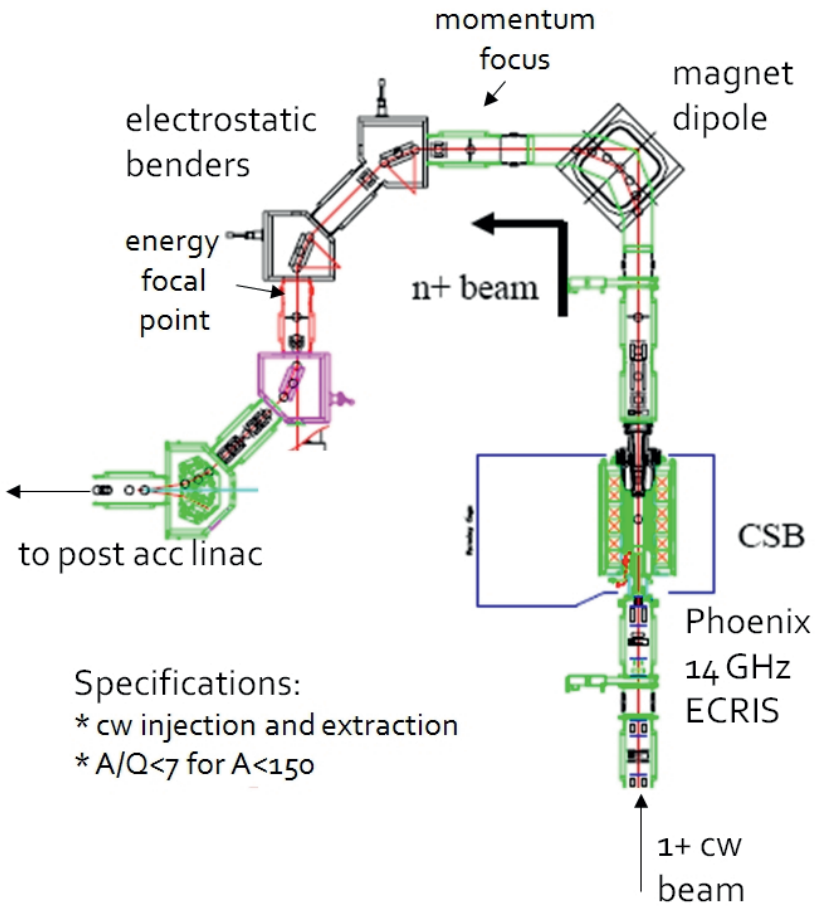


Fig. 15: The charge breeder for the ISAC post-accelerator facility at TRIUMF [30]. The 1^+ beam is charge-bred in a 14.5 GHz ECRIS. The cw extracted beam is separated in a combined magnetic and electrostatic analyser before being injected into a radiofrequency quadrupole (RFQ) cavity and the consecutive accelerating cavities.

6 EBIS/T-based charge breeders

By tradition, an EBIS has a long trapping region (>50 cm) with a limited j_e (<1000 A cm^{-2}) compressed by an extended solenoid, in most cases superconducting. An EBIT, on the other hand, is based around a pair of Helmholtz coils, has a very small electron-beam diameter r_{ebeam} and therefore has high j_e (>1000 A cm^{-2}) owing to the high compression, and a short trapping region (~10 cm), from which usually no ions are extracted. These differences have consequences for the charge-breeding performance, as will be pointed out. For some time now, the distinction between EBIS and EBIT has not been that clear-cut; in fact, the machine types are approaching each other, and EBIT can also provide ions to external set-ups.

The main elements of an EBIS/T, such as electron gun, trapping tubes, electron collector, extraction electrode and magnetic solenoid, are represented in Fig. 16. The ion injection is performed from the collector side, as well as the ion extraction. Once injected, the ions are trapped in a magneto-electrostatic trap: radially by the electron-beam space-charge potential and the magnetic field, and axially by electrostatic potential barriers applied on the trapping drift tubes. The ionization to higher charge states is obtained by bombardment from the fast, dense monoenergetic electron beam generated by the electron gun.

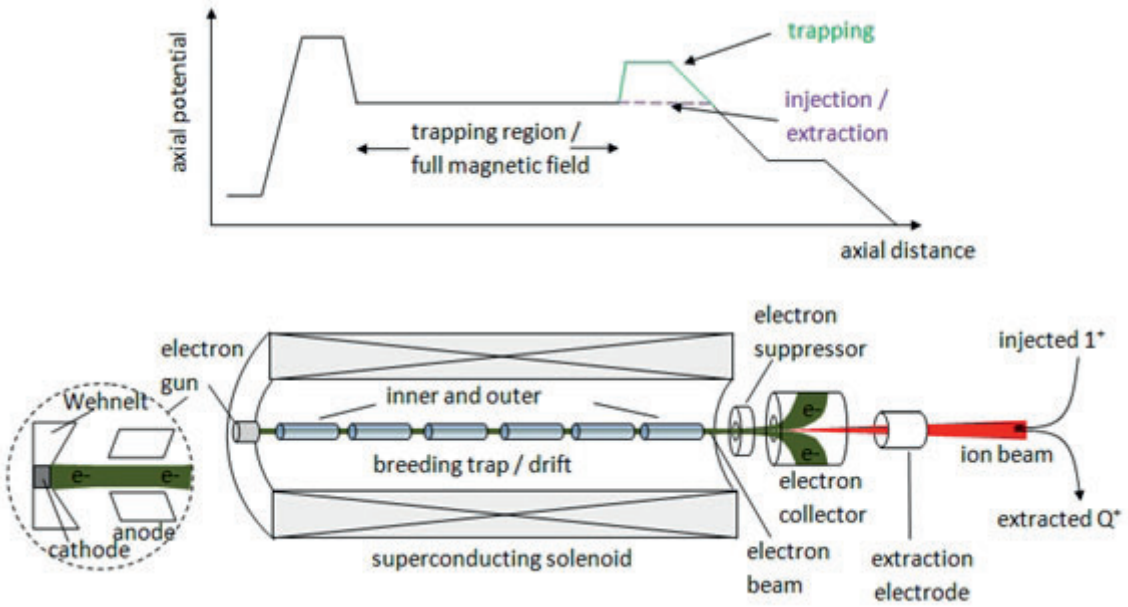


Fig. 16: Schematic picture of an EBIS/T. The ion injection and extraction are performed from the collector side using an electrostatic kicker to direct the injected ions from the ISOL system into the breeder and the extracted ions to the succeeding mass separator.

To reach a certain charge state, either the electron current density or the charge-breeding time has be adjusted, as given by Eq. (3). The electron current density, typically varying between 50 and 5000 A cm^{-2}, is less convenient to change compared to the breeding time. It is important to have the possibility to select a specific charge state in order to avoid superimposed stable beam contaminants from the breeder, as will be discussed section 7.2. In Fig. 17 two extracted charge spectra of Cs with different breeding times are shown. For a breeding time of 78 ms, the charge-state distribution peaks at 29^{+}; while for 158 ms, the maximum number of ions is found at 32^{+}. In general, one would like to keep the breeding time as short as possible, and therefore one profits from a high total electron current I_e as $j_e = I_e/(\pi r_{\mathrm{ebeam}}^2)$.

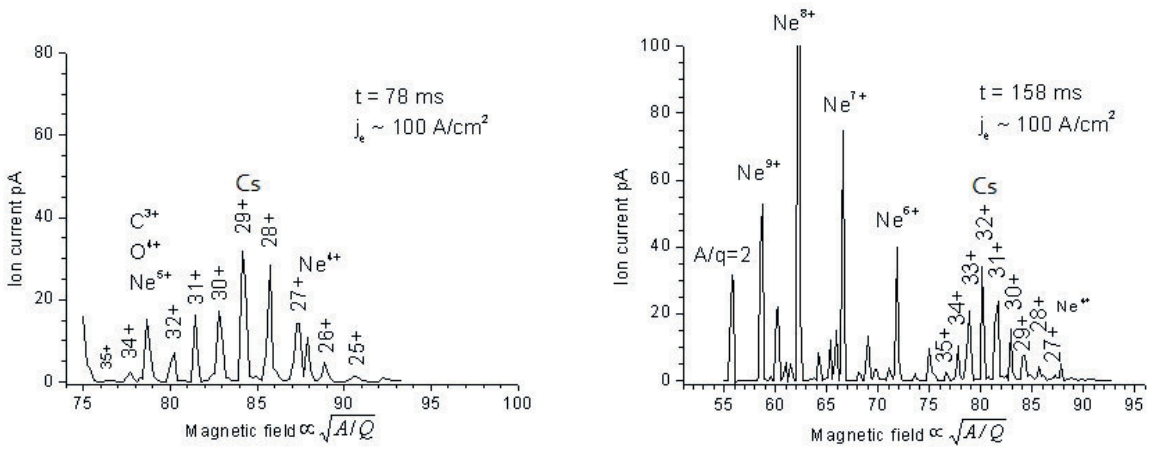


Fig 17: Extracted charge-state distributions of Cs for two different breeding times. In both cases, the electron current density was approximately 100 A cm^{-2}. The right spectrum is performed with a wider scan so several residual gas peaks are present; most dominant are the Ne peaks originating from buffer gas diffusing from the preceding Penning trap (see section 7.1).

When the ions are trapped inside the breeder, it is desired that they reside within the electron beam in order to undergo stepwise ionization. If they spend part of their time outside the electron beam, the effective electron current density will be lower ($j_e(\text{effective}) = j_e T_{\text{inside}}/T_{\text{total}}$) and the charge-breeding time subsequently increases. The ion injection into the EBIS/T is therefore critical, and a maximum overlap between the ion beam and the electron beam is essential. Figure 18 illustrates different ion injection cases, the middle one being ideal, and the right one being completely flawed. The ions will never cross the electron beam in the latter situation and will therefore remain as 1^+. The left situation, although not ideal but acceptable if a prolonged breeding time is tolerated, will yield charge-bred ions as they traverse the electron beam intermittently and will therefore be further ionized. For each ionization step, the potential depth from electron-beam space charge seen by the ion will increase and, owing to conservation of energy, the ions will be attracted to the beam axis and the overlap therefore increase.

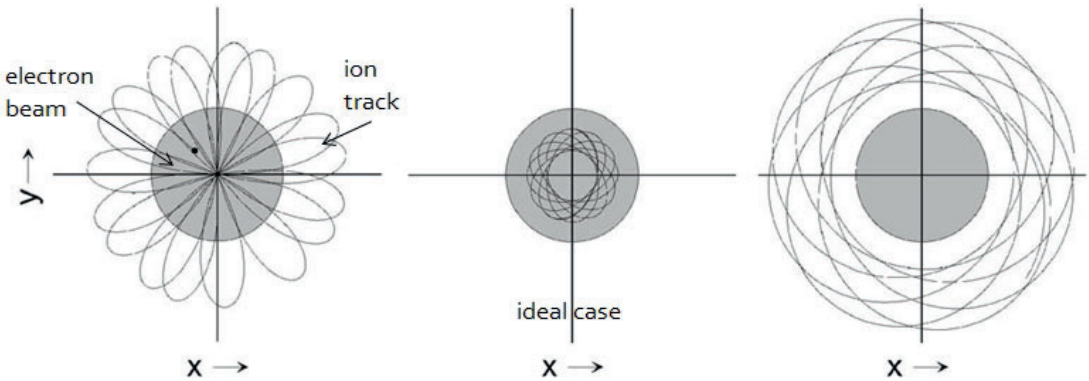


Fig 18: Ion tracks for different injection scenarios. To the left the ions occasionally traverse the electron beam and will at some point become ionized to higher charge states, while this never happens in the case to the right. The ideal situation is a complete overlap between the 1^+ ion trajectory and the electron beam, as shown in the middle. Figure modified from [32].

The transverse phase-space acceptance of the electron beam can be estimated using Eq. (11) [31]. This formula gives the normalized phase-space area within which ions have to fit to be fully trapped within the electron beam when entering a non-space-charge-neutralized EBIS/T:

$$\alpha = \pi\gamma \frac{r_{\text{ebeam}}}{c}\sqrt{\frac{Q_{\text{inj}}e}{m_{\text{ion}}}}\left(Br_{\text{ebeam}}\sqrt{\frac{Q_{\text{inj}}e}{m_{\text{ion}}}} + \sqrt{\frac{Q_{\text{inj}}eB^2r_{\text{ebeam}}^2}{4m_{\text{ion}}} + \frac{\rho_l}{2\pi\varepsilon_0}}\right) \tag{11}$$

Here the charge per metre of the electron beam is denoted by ρ_l and γ represents the relativistic gamma factor. Examining the formula, one finds that two terms originate from the magnetic field, while the second term under the square root is related to the electron-beam space charge. A high electron-beam current results in a large acceptance. Furthermore, an electron beam being space-charge-neutralized by ions has a smaller acceptance compared to an empty beam, and its acceptance is governed by the solenoid magnetic field.[5] As expected, the acceptance is strongly dependent on the electron-beam radius, meaning that traditional EBIT devices with very small beam radii may have difficulties trapping the ions.

There are two schemes to introduce the external ions into the EBIS/T: continuous injection, also called 'accu-mode'; and pulsed injection. In the former, the ions are continuously injected over the outer barrier, and during the round trip inside the trapping region they are ionized to a higher charge state and subsequently trapped (see Fig. 19). Hence, in order to succeed with this injection mode, the ions have first of all to be within the acceptance of the electron beam, as discussed above. In addition,

[5] The partial neutralization of the electron beam by ions is not only harmful. In fact, the compensating ions act as a well-defined Coulomb target for the injected ion and can improve the slowing-down and trapping process, particularly for continuous over-the-barrier injection [33].

since there are no dissipative forces present (a trapping region with only a few ions confined is assumed), a successive ionization is mandatory, as 1^+ ions can escape again from the trapping region over the outer barrier. Singly charged ions ionized to 2^+, on the other hand, will see a doubled outer barrier potential.

Equation (3) gives the average time $\tau_{1\to2}$ it takes for a singly charged ion to lose one electron, that is $\tau_{1\to2} = e/(j_e\sigma_{1\to2})$. The probability P for the ionization process to have taken place within a time τ_{ebeam} spent inside the electron beam is given by $P = 1 - \exp(-\tau_{\text{ebeam}}/\tau_{1\to2})$. Using the example of $^{55}Fe^+$ with $\sigma_{1\to2} \sim 10^{-17}$ cm^2 being injected into an electron beam with density $j_e = 200$ A cm^{-2}, and requiring that $P > 0.5$, a time within the electron beam exceeding 55 μs is required. To increase the probability, the spontaneous solution is to reduce the ion injection energy to increase the round-trip flight time. The magnetic-mirror effect, however, sets a limitation to this, as the ions will be reflected while entering over the outer barrier if the energy is too low, as illustrated in Fig. 12.

To obtain a high trapping efficiency for a continuously injected beam, the trapping region should be long, the electron-beam radius large and the electron current density high. Unfortunately, the two last requirements are in contradiction, as $I_e = j_e\pi r_{\text{ebeam}}^2$. In general, a careful tuning of the injection energy is required (within ~10 eV), as well as a small energy spread and transverse emittance of the injected beam in order to obtain an acceptable injection and trapping efficiency. A more efficient injection method is pulsed injection.

The pulsed injection scheme requires a fast pulsing of the outer barrier and a pre-bunching of the beam from the ISOL system. The pulse length has to be shorter than the round-trip time inside the trapping region. The pre-bunching will be dealt with in section 7.1. Figure 20 (top) demonstrates the pulsed injection process with the quick ramping of the outer barrier. To assure an immersed injection into the electron beam, the ion injection energy should ideally be less than the potential depth of the electron beam, given by

$$U_{\text{well}} = \frac{I_e}{4\pi\varepsilon_0}\sqrt{\frac{m_e}{2eU_e}} \tag{12}$$

and illustrated in Fig. 20 (bottom). The notation U_e represents the electron-beam voltage in the trapping region, and a typical value for the electron-beam well is 100 V. For an element with $A = 55$ injected into a 0.5 m long EBIS/T, this corresponds to a pulse length shorter than 50 μs. In reality, somewhat higher injection energy is often used to reduce the magnetic mirror reflection at the entrance at the cost of a longer breeding time (due to the reduced effective electron current density). The pulsed injection method is relatively flexible and has a higher transverse and energy-spread acceptance compared to the continuous injection. The major drawback is the need for pulse formation prior to the breeder.

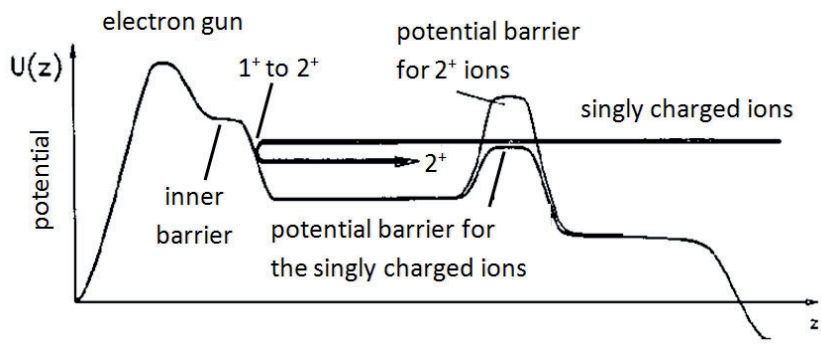


Fig. 19: Continuous ion injection into an EBIS/T. The ions are injected with a low velocity over the outer barrier, and inside the trapping region they must undergo ionization to 2^+ or higher within one round trip to get trapped.

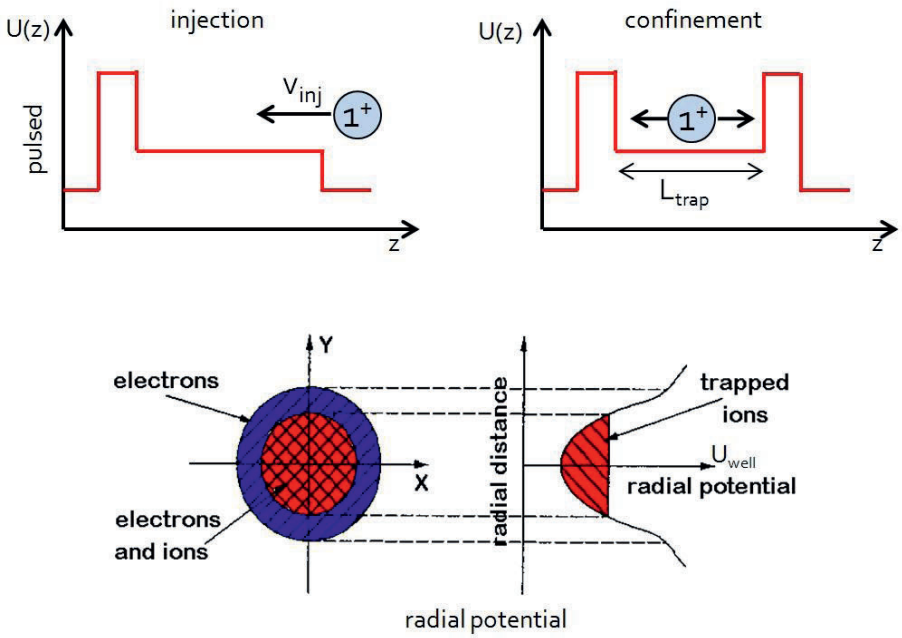


Fig. 20: (top) Pulsed injection scheme. The pulse length and the ramping time of the outer barrier are shorter than the round trip within the trapping region. (bottom) The radial electron-beam well and ions injected at the bottom of the well.

The electron-beam energy in a charge breeder is dictated either by the perveance limitation or the desired ion charge states. The perveance, a measure of the space-charge effect in the electron beam, relates the electron current to the electron-beam energy as $P = I_e/U_e^{3/2}$. This is of particular importance at the electron gun, but applies as well for the drift tube and collector regions. As a guideline, the perveance should be below 5×10^{-6} perv, which for an electron beam of 1 A corresponds to a minimum electron energy of 3500 V. To reach the desired A/Q value, the electron-beam energy has to be higher than the ionization potential for charge state Q of the element. Particularly difficult are elements along the neutron-rich drip-line due to the excess of neutrons, therefore requiring a high charge state. In Table 3 the ionization potential for different elements along the neutron drip-line are listed assuming that an $A/Q \sim 4$ is requested. Even for the heaviest elements, the ionization potential is modest. Using Lotz's formula, Eq. (4), one can show that the optimal electron-beam energy for ionization is ~2.7 times the ionization potential; thus a maximum electron-beam energy of 10 keV is necessary to cover the full nuclear landscape. Nonetheless, certain dedicated experiments may require much lower A/Q ratios, even fully stripped ions, and then the electron-beam energy should reach 100 keV or higher.

Table 3: Ionization potentials for charge state Q for different neutron-rich elements, where the charge state is chosen such that $A/Q \sim 4$.

Z	A (neutron-rich)	Q	Ionization potential (eV)
20	60	15	900
40	110	27	1500
60	161	40	2800
80	210	52	3100

The extraction of the charge-bred ions from the EBIS/T can be done in three different ways: with passive, fast or slow extraction, shown in Fig. 21. The passive extraction is achieved by abruptly lowering the outer barrier and letting the ions leave the trap with their inherent energy. The total ion

energy at extraction is the sum of the injection energy, the energy gained during each ionization step where the potential energy of the ion is increased due to the change in charge, and electron–ion heating and ion–ion cooling processes. Owing to Coulomb scattering among the ions, the momentum space is assumed to be almost isotropic. Heavier ions with higher charge state gain more energy than the lighter ones during the ionization process (even disregarding the electron–ion heating), but the large mass of the ions makes them leave the trap slower. A characteristic passive extraction pulse from a 0.8 m long trapping region ranges from 25 μs (full width at half-maximum, FWHM) for light ions to around 80 μs (FWHM) for very heavy ions ($A/Q = 4$). If an axial electric field is applied during the extraction phase, effectively attained by putting different voltages on the trapping tubes, the ions can be extracted faster (within less than 10 μs) but the energy spread will increase accordingly. Fast extraction is advantageous for injection into synchrotrons, for example.

Finally, by slowly lowering the other barrier, or by slowly increasing the potential of the trapping tubes, the ions can be leaked over the extraction barrier for several 100 μs, or even milliseconds. This is of interest for experiments requiring a low instantaneous particle rate.

The ion throughput rate is determined by the storage capacity of positive charges and the rate at which the injection, breeding and extraction cycle can be repeated. The ion storage capacity is determined by the trap length L_{trap}, the electron-beam current I_{e}, the electron-beam voltage U_{e} and the electron-beam compensation degree f, and the number of charges N^{-} is given by the expression

$$N^{-} = \frac{f I_{\mathrm{e}} L_{\mathrm{trap}}}{e} \sqrt{\frac{m_{\mathrm{e}}}{2eU_{\mathrm{e}}}} \,. \tag{13}$$

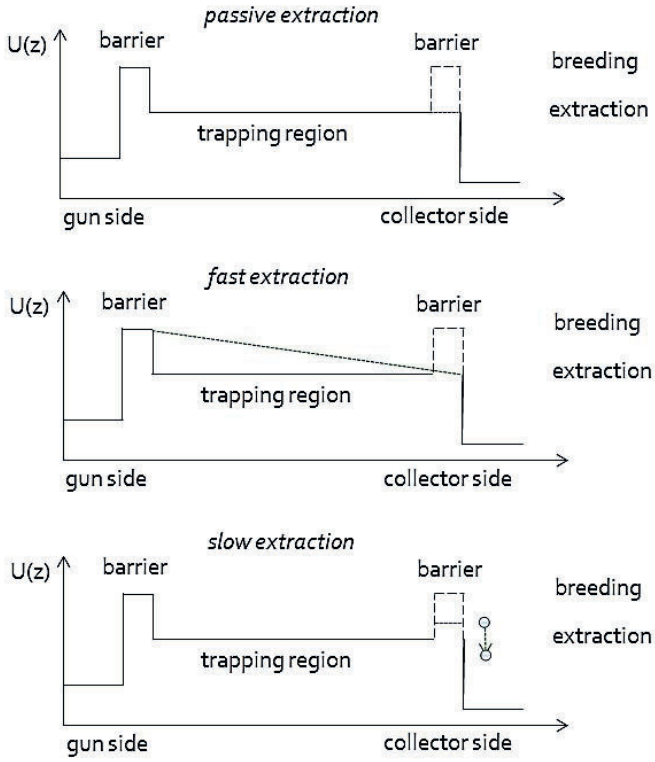


Fig. 21: Different ion extraction schemes from an EBIS/T. The fastest extraction is obtained with the second mode, although with a high energy spread. The lowest energy spread, down to a few eV total spread, is reached with slow extraction mode, which also produces long beam pulses.

A long trapping region and high electron current are the main contributors to a high storage capacity. As an example, an electron beam with 0.5 A and 5 keV in a 0.8 m long trap with a 50% compensation degree has a capacity of 3×10^{10} charges. With approximately 20% of the beam in the desired charge state, that means that about 10^9 ions with charge 5^+ can be extracted from one pulse, and 10^8 ions with charge 50^+. This assumes that no evaporative ion cooling [34] using lighter ions is employed, as the light ions will occupy part of the electron space-charge. The ion throughput per second equals the storage capacity divided by the repetition rate, mainly governed by the breeding time. Additional capacity limitation due to a preparatory Penning trap or similar is treated later on in section 7.1.

An example of an RIB facility using an EBIT as breeder is the ReA3 at NSCL/MSU, which is a 3 MeV/*A* re-accelerator of thermalized projectile fragmentation and fission beams [35]. Its layout is presented in Fig. 22.

The set-up uses a very high-compression electron gun aiming for $j_e > 10^4$ A cm^{-2}, meaning that the charge breeding of ions with $Z < 35$ into Ne-like or higher should take less than 10 ms, and the ionization from 1^+ to 2^+ occurs within 1 μs. The rapid ionization, in combination with a long trapping region with a variable magnetic field strength in the axial direction, ensures a satisfactory accumulation efficiency for the continuous injection of 1^+ ions. High ion beam rates (>10^9 ions per second) should also be attainable, as the electron current is large and the breeding time very short. The elements in the drift-tube region are at cryogenic temperatures, leading to a good vacuum and low degree of residual gas ions in the extracted beam.

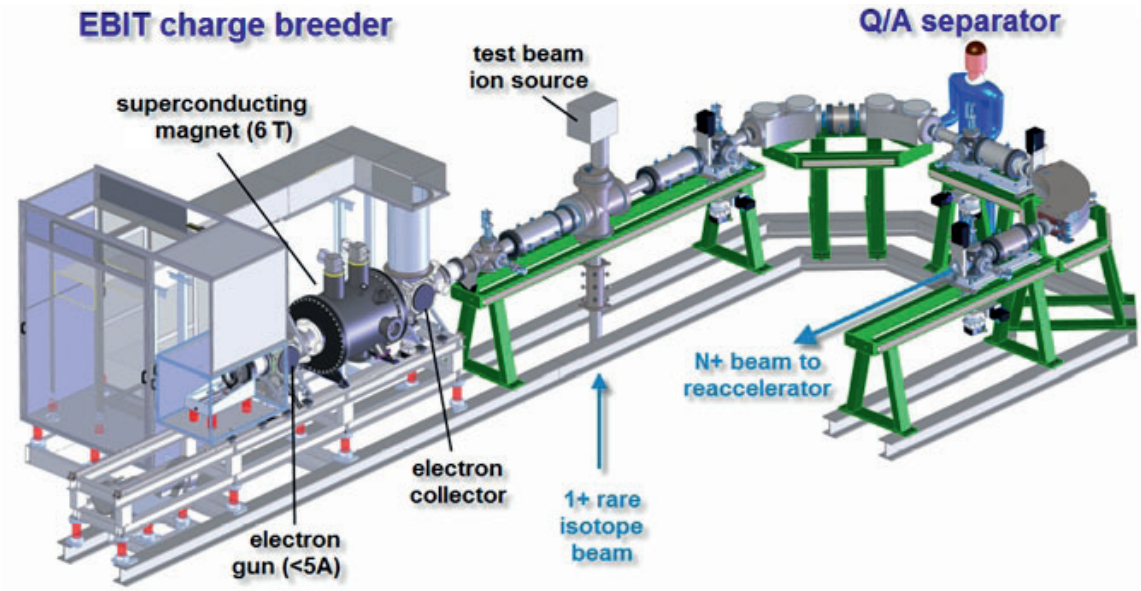


Fig 22: The EBIT-based charge breeder for the ReA3 facility. The radioactive 1^+ ions come from a gas stopper in which they have been cooled to a low emittance and energy spread. The pulsed extracted beam is analysed in an *A/Q* separator before being injected into a superconducting linac.

7 Preparatory devices and tricks

7.1 Beam preparation in a Penning trap

The performance of the charge breeder can be improved if a beam preparatory device such as buffer-gas-filled RFQ cooler [36] or Penning trap [37] is introduced prior to the breeder. Apart from reducing the transverse beam emittance and the energy spread to a few mm mrad (50 keV, 90%) and to less than 1 eV, respectively, such devices can change the temporal beam structure, that is, bunch the beam.

In addition, the Penning trap approach offers, in the case of low-intensity beams, the extra opportunity of mass purification, as will be explained below.

In a Penning trap, ions are confined in the longitudinal direction by a quadratic electrostatic potential and in the radial direction by a strong magnetic field parallel to the symmetry axis of the electric field (see Fig. 23). The trapped ion performs three independent eigenmotions: one harmonic oscillation in the axial direction (ν_z) and two oscillations in the radial direction, the reduced cyclotron motion (ν_+) and the magnetron motion (ν_-). These frequencies are related as

$$\nu_{\pm} = \frac{\nu_c}{2}\left(1 \pm \sqrt{1 - \frac{2\nu_z^2}{\nu_c^2}}\right), \tag{14a}$$

$$\nu_+ + \nu_- = \nu_c = \frac{QeB}{2\pi m_{ion}}. \tag{14b}$$

The cyclotron frequency (ν_c) depends on the ion mass and the magnetic field strength, while the magnetron motion is almost mass-independent. Typical values for an ion with A = 30 in a magnetic field of 3 T, an electrical potential depth of 10 V and a characteristic trap dimension of 15 mm are ν_z = 60 kHz, ν_- = 1.2 kHz and $\nu_c \approx \nu_+$ = 1.5 MHz.

In a buffer-gas-filled trap, cooling occurs due to the frictional force provided by the gas. The time constant for the cooling is inversely proportional to the frequency of the oscillation. Thus, the axial and reduced cyclotron motions are cooled fast and their amplitudes are consequently reduced. The magnetron motion, on the other hand, is cooled much more slowly and has negative energy radius dependence, that is, the radius increases when the ion is cooled. To prevent ions from hitting the electrode walls, the so-called sideband cooling technique is used [38]. A quadrupolar RF electrical field is applied in the transverse plane at the cyclotron frequency, which couples the magnetron and reduced cyclotron motions. In a trap without a gas, an oscillation between the two motions would occur for a sideband excitation at ν_c. When gas is added, the cyclotron motion decreases fast, while the magnetron radius increases slowly, but as a result of the coupling both radii are decreased. Since ν_c is mass-dependent, the sideband excitation can be used to centre specific ion species as shown below.

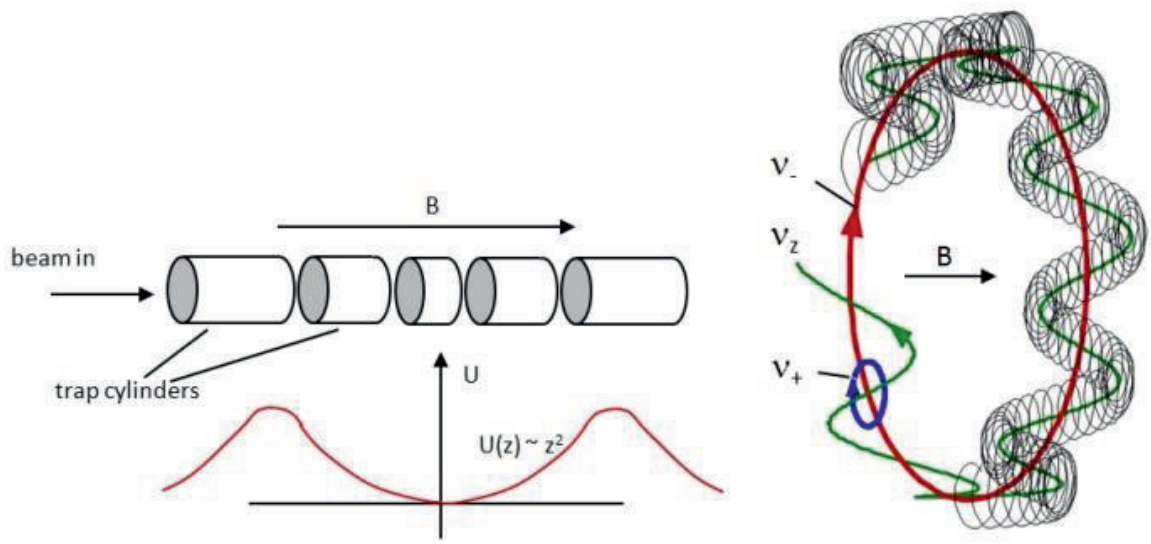


Fig. 23: (left) Penning trap electrode stack, with magnetic field and quadratic electrostatic potential. (right) The three ion motions inside a Penning trap: the large-radius magnetron motion with a low frequency ν_-; the axial motion with a frequency ν_z; and the fast spinning reduced cyclotron motion with a frequency ν_+.

To capture the continuous beam from the ISOL system, the ions are slowed down so they have just enough energy to overcome the electrostatic potential wall on the injection side. Because of the presence of the buffer gas, a dissipative force causes the ions to lose energy.

The viscous force can be approximated with $F = -q_{\text{ion}}v_{\text{ion}}/K$, where K is the ion mobility, which is inversely proportional to the buffer-gas pressure. For example, ^{23}Na with an energy of 50 eV will lose 25 eV m^{-1} in Ne gas with a pressure of 10^{-3} mbar. After a cooling period of some tens of milliseconds, the ions are ejected from the trap as a short bunch (a few microseconds long) when the outer barrier is lowered. The bunch length Δt_{extr} and the energy spread ΔE_{extr} are related by $\Delta E_{\text{extr}}\Delta t_{\text{extr}} \approx 5$ eV μs.

An example of a preparatory Penning trap is REXTRAP used in front of the REXEBIS charge breeder in the REX-ISOLDE post-accelerator [39]. It accumulates and cools the beam for as long as the charge breeding is on-going inside the EBIS, and, once the EBIS is ready to accept new ions, a bunch is released from the Penning trap. As the extracted beam has a transverse emittance of about 10 mm mrad (30 keV, 90%), and the pulse length is less than 20 μs, the injection into REXEBIS becomes efficient. Normally the Penning trap is operated with a low mass resolution ($m/\Delta m < 300$). The Brillouin limit [40]

$$n_{\text{ion}} = \frac{\varepsilon_0 B^2}{2m_{\text{ion}}} \tag{15}$$

yields the maximum ion density that can be stored within the trap, and, for a field of a few teslas, around 10^8 ions with $A = 100$ can be stored per cubic centimetre. In REXTRAP space-charge effects start occurring for more than a several 10^7 ions per pulse, with an emittance increase and efficiency decrease as a result. This means that, at present, the Penning trap–EBIS concept can provide approximately $10^8 \cdot \eta_{\text{trap-EBIS}}/T_{\text{breeding}}$ number of accelerated particles per second, where $\eta_{\text{trap-EBIS}}$ is the combined transmission efficiency for the Penning trap and EBIS, and T_{breeding} the breeding time.

7.2 Beam purity

In contrast to other application fields for EBIS/T and ECRIS, a high ion current out of a charge breeder is rarely required. Instead, the intensity of the radioactive beams can be very low, for instance, 10^3 ions per second corresponding to ~0.2 fA is not exceptional. At such low currents, contaminating beams can easily dominate. In decay studies performed at standard ISOL energies, the stable beam component is of little concern, but for post-accelerated beams it causes problems at the experiment. For instance, the excitation from the contaminations may dominate the direct emission from the experimental target, causing detector overload (pile-up, dead-time and random coincidences) and cross-section normalization problems.

The origin of contamination is either from the ISOL system (radioactive or stable contamination) or from the breeder itself (stable contamination). In the former case, it is almost always a question of isobaric contamination, while from the charge breeder a non-isobaric contamination with a similar A/Q value as the radioactive beam can be present. In the breeder, residual gases are ionized, producing beams of, for instance, C, N and O.

The extracted current is proportional to the partial pressure of the contributing gases, so the good vacuum inside an EBIS/T is advantageous. An A/Q spectrum of an extracted EBIS/T beam is shown in Fig. 24. In an ECRIS, elements sputtered or desorbed from the internal wall surfaces (Al, Fe, etc.) and support-gas ions are abundantly extracted.

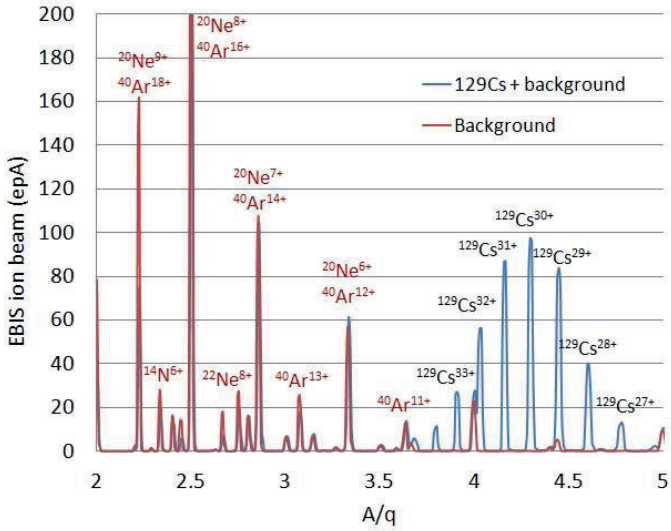


Fig. 24: Extracted beams from REXEBIS as a function of A/Q showing residual gas peaks and charge-bred ^{129}Cs. The blue trace is with and the red trace without ^{129}Cs being injected. The dominating Ne peaks are due to buffer gas migrating from the Penning trap, while Ar had previously been used as venting gas inside the EBIS.

To suppress the breeder contaminations, a combined electrostatic deflector and magnetic bender is often used in order to attain an achromatic selection system with large energy-spread acceptance [41]. If only a separator magnet, with rigidity $B\rho$, is employed, the velocity spread out of the breeder placed at a potential U_{extr} can cause an ambiguity in the A/Q selection as $B\rho = m_{ion}v_{extr}/q_{ion}$. Nevertheless, with an electrostatic deflector included having an electrical bending field E_{defl} and radius r_{defl}, which selects ions according to $E_{defl}r_{def} = 2U_{extr}$, the combined selection becomes $m_{ion}/q_{ion} = (B\rho)^2/(E_{defl}r_{defl})$, and only a single A/Q is transmitted. A typical A/Q resolution of a breeder separator is of the order of a few hundred. Therefore, for example, $^{14}N^{6+}$ could possibly be separated from $^{7}Be^{3+}$ (required resolution 450), while $^{12}C^{6+}$ is difficult to separate from $^{18}F^{9+}$ (required resolution 19 200). Owing to the limited resolution, it is imperative to have background-free A/Q regions, as for $^{129}Cs^{30+}$ in Fig. 24, and the possibility to tune the ion charge to these regions.

The beam leaving the ISOL system is occasionally superimposed by isobaric contaminants. For instance, the double-magic ^{132}Sn is overwhelmed by ^{132}Cs by several orders of magnitude. In the target ion source, several tricks can be applied in order to suppress the contamination, but no universal solution exists. Isobaric selection in the separator is difficult to attain, as a resolution of 5000 to 50 000 is typically needed (see also Table 4).

To combat the isobaric contaminations from the ISOL system, one can use so-called molecular sideband beams. The method makes use of the fact that elements along an isobar have different chemical properties and in most cases do not form similar molecular compounds. Thereby the desired element can be transferred into a clean molecular sideband with a high chemical selectivity. The method is thoroughly treated in [42]. For example, instead of selecting ^{132}Sn directly from the target and primary ion source, with a substantial ^{132}Cs contamination superimposed, the molecule $^{132}Sn^{34}S$ is mass-selected (produced by injection of $^{34}S_2$ gas). Cs does not preferably bind as CsS, and is therefore suppressed. The SnS molecule is transferred to the breeder, where it is dissociated and Sn charge-bred in a standard way. The procedure is illustrated in Fig. 25. If some other element of mass 166 leaves the ISOL separator, it can be suppressed in the breeder separator by an appropriate selection of A/Q. This two-stage suppression is in general very effective. The transmission efficiency depends on the electronegativity of the desired ion in the molecule, disfavouring radioactive ions with highly electronegative values (particularly group VII elements), since they tend to become neutral (or even negative) when the molecule is broken up and therefore escape the confining potential of the breeder.

Table 4: Typical required mass resolving powers.

	Required resolution
Neighbouring masses	250
Molecular ions (e.g. CO from N_2)	500–1000
Isobars (e.g. ^{96}Sr from ^{96}Rb)	5000–50 000
Isomers	10^5–10^6

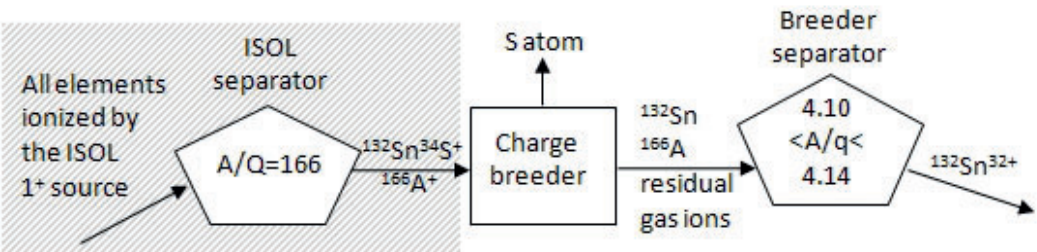


Fig. 25: By extracting ^{132}Sn^{34}S from the ISOL system, the contaminating ^{132}Cs is avoided. The SnS molecule is dissociated in the breeder and Sn is charge-bred and extracted. The ^{166}A contamination is suppressed in the second separator stage.

Another technique to suppress low-intensity isobaric contaminants from the ISOL system is to make use of the intrinsic mass resolution of a Penning trap. By ion-motion manipulation inside the trap (see Fig. 26), a mass-resolving power of the order of 10^4 to 10^5 is achievable in principle. The complete trapped ion cloud is first compressed by a quadrupolar excitation at ν_c with a large amplitude (the resonance thereby covers all injected elements, including the contaminants) inside the Penning trap for approximately 50 ms. Thereafter the cloud is displaced radially with a dipolar mass-independent excitation at ν_- for 20 ms. Once offset, a quadrupolar excitation at ν_c is again applied for 100–200 ms, though this time with a low amplitude and therefore with a narrow sideband resonance, which selectively re-centres the desired ions according to $\nu_c = QeB/(2\pi m_{\mathrm{ion}})$. These can then be extracted from the trap and injected into the breeder in the normal way, while the still displaced contaminants are stopped by a diaphragm. The complete cycle takes up to 400 ms. Owing to space-charge effects, the method is only applicable for less than 10^6 ions per bunch, including the contaminants [43].

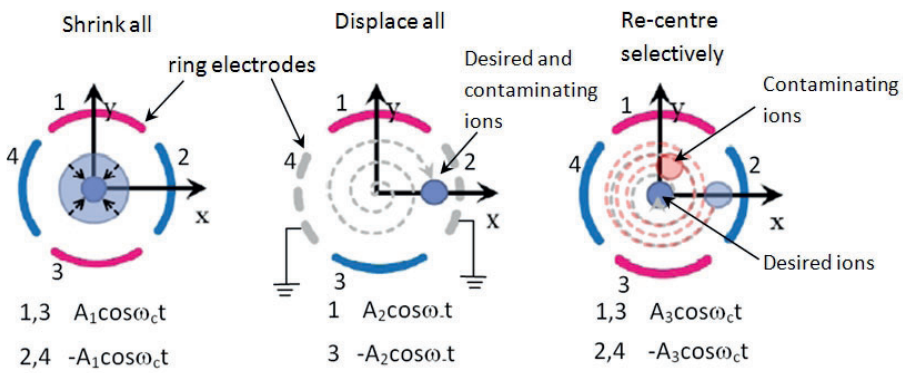


Fig. 26: Sequence of RF excitations to achieve isobaric separation in a Penning trap. Axial view of the trap centre. (left) High-amplitude quadrupolar excitation at the true cyclotron frequency to compress the complete ion cloud. (middle) Dipolar excitation at the magnetron frequency to displace the ion cloud outside the acceptance of the extraction diaphragm. (right) Low-amplitude mass-selective quadrupolar excitation at the cyclotron frequency to re-centre the desired ions. Amplitudes $A_1 \gg A_2 > A_3$. Picture adjusted from [44].

8 Facilities and future

Charge breeders are not exclusively used as a preparatory stage for injection into a post-accelerator. A few set-ups exist that use EBIT as charge breeder to feed mass-measurement systems, for example TRIUMF's Ion Trap for Atomic and Nuclear science (TITAN) [45]. High-precision mass measurements using traps benefit from the higher charge state as the resolution is given as

$$\frac{m_{\text{ion}}}{\Delta m_{\text{ion}}} \propto \frac{QeB}{m_{\text{ion}}} T_{\text{RF}} \sqrt{N}\,, \tag{16}$$

where T_{RF} is the duration of the radiofrequency excitation and N the number of measurement cycles. For short-lived radioactive ions, the advantage becomes evident, as the excitation time may be restricted by the half-life, so instead of several seconds excitation time, a 10^+ charged ion can be measured with similar accuracy within a fraction of a second. In contrast to the post-accelerator context, these breeders emphasize a high charge state and the amount of ions to breed is limited, of the order of a few hundred, as ideally only a single ion should be injected into the mass-measuring trapping system at each cycle. The efficiency should nonetheless be high, as many of the ions are exotic and produced in small numbers. Additionally, the extracted ions must not be contaminated with stable ions originating from the breeder, as that will impair the mass-measurement process inside the trap.

The use of laser ion sources for the production of highly charged ions has been explored at different laboratories; for example, the production of Pb^{27+} was reported in Ref. [46]. The ionization process is non-resonant and requires high-intensity lasers ($P > 10^{10}$ W cm^{-2}) if high charge states are to be attained. Their applicability for charge breeding is, however, expected to be limited, as the efficiency is low for non-solid elements, the transverse beam emittance large (for a Pb beam, 140 mm mrad for 100 kV, 90%) and the energy spread several keV Q. The high power requirement also implies that the lasers are pulsed with a limited repetition rate, which negatively affects the efficiency.

The number of charge breeders for RIB facilities have rapidly grown since the year 2000. The present situation, listing operational breeders and breeders in the design or commissioning stage, is illustrated in Fig. 27. All new ISOL facilities involving a post-accelerator, or an advanced Penning trap system for mass measurements, have a breeder of either EBIS/T or ECRIS type.

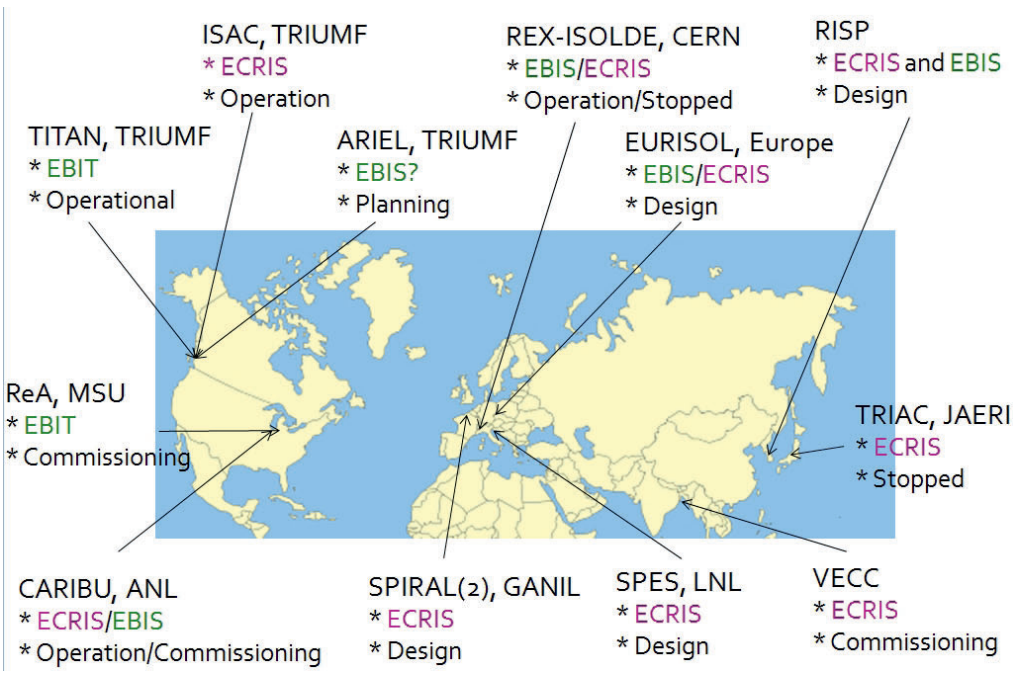


Fig. 27: Charge breeders for radioactive ion beams as of spring 2012

From the next-generation RIB facilities, first of all higher beam intensities are to be expected. Accelerated beam intensities exceeding 10^{12} particles per second could be expected in an ultimate EURISOL facility [47]. The higher beam intensities implicitly lead to access further out towards the proton and neutron drip-lines, with short nuclear lifetimes as a result. Existing breeding techniques, with parallel operation of EBIS/T for traditional low-intensity experiments and ECRIS breeders for high-current beams, as shown in Fig. 28, will cover these upcoming requirements.

More challenging will be the charge breeding of ions to very high charge states, even fully stripped, for consecutive injection into storage rings [48]. The collection and charge breeding of extreme amounts of ^{6}He and ^{18}Ne to be used in beta-beam facilities [49] are also very challenging. The state-of-the-art breeding techniques need to be pushed even further to cover these cases.

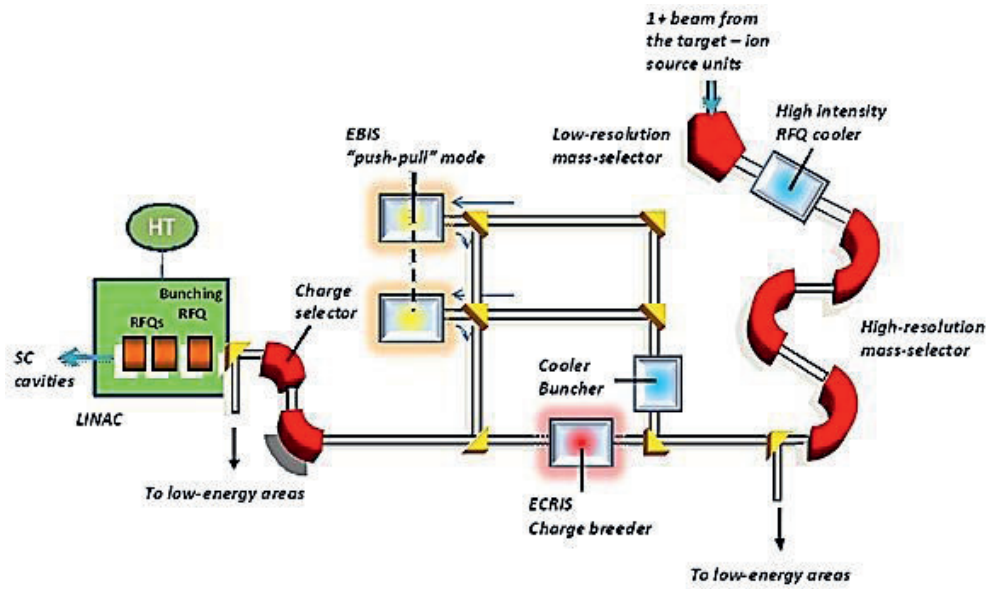


Fig. 28: Two main charge-breeding paths in a future high-intensity RIB facility. For exotic low-intensity beams ($<10^{8}$ ions per second) used for 'standard' nuclear spectroscopy experiments, two EBIS/T operating in push–pull mode are foreseen. For high-intensity RIBs, for example, intended to generate even more neutron-rich nuclei, an ECRIS is used as a charge breeder as the beam purity is of less importance.

9 Conclusions

Charge breeding of exotic isotopes produced in ISOL facilities has during the past decade opened up unique possibilities for nuclear and atomic physic experiments at both low and high energies. The higher charge states of the radioactive ions give access to precise mass measurements of very short-lived ions, but, above all, it allows for more compact and cost-effective post-accelerators achieving some MeV/A thanks to the lowered mass-to-charge ratio.

The breeding time of some tens to a few hundreds of milliseconds is compatible with the hold-up time in an ISOL target-ion source assembly. As pointed out, the characteristics of the breeding device have a decisive influence on the layout of the post-accelerator, as the time structure of the beam (CW or pulsed extraction) and the A/Q are determined by the breeder. Furthermore, the level of beam contamination in the extracted charge-bred beam, originating from mainly residual gases in the breeder, is an important aspect.

References

[1] D.J. Morrissey and B.M. Sherill, In-flight separation of projectile fragments, *Lect. Notes Phys.* **651** (2004) 113.

[2] H.L. Ravn and B.W. Allardyce, On-line mass separators, in *Treatise on Heavy-Ion Science*, Ed. D.A. Bromley (Plenum, New York, 1989), Vol. 8, p. 363.

[3] Focus section: Physics with reaccelerated ISOL beams, *J. Phys. A: Nucl. Part. Phys.* **38** (2011), February.

[4] J.R.J. Bennett, *Nucl. Instrum. Methods* **B204** (2003) 215.

[5] B. Marsh, these Proceedings.

[6] W. Lotz, *Z. Phys.* **206** (1967) 205.

[7] W. Lotz, *Z. Phys*. **232** (1970) 101.

[8] M.F. Gu, *Can. J. Phys*. **86** (2008) 675.

[9] J. Arianer, A. Cabrespine and C. Goldstein, *Nucl. Instrum. Methods* **193** (2003) 401.

[10] F. Currell and G. Fussmann, *IEEE Trans. Plasma Sci.* **33** (2005) 1.

[11] A. Müller and E. Salzborn, *Phys. Lett.* **A62** (1977) 391.

[12] N. Selberg, C. Biedermann and H. Cederquist, *Phys. Rev.* **A54** (1996) 4127.

[13] V. Liechtenstein, Carbon stripper foils – preparation and quality, in *Electrostatic Accelerators Particle Acceleration and Detection*, Ed. R. Hellborg (Springer, Berlin, 2005), p. 187.

[14] E. Baron, M. Bajard and Ch. Ricaud, *Nucl. Instrum. Methods* **A328** (1993) 177.

[15] G. Schiwietz and P.L. Grande, *Nucl. Instrum. Methods* **B175-177** (2001) 125.

[16] K. Shima, N. Kuno, M. Yamanouchi and H. Tawara, *At. Data Nucl. Data Tables* **51** (1992) 173.

[17] M. Toulemonde, *Nucl. Instrum. Methods* **B250** (2006) 263.

[18] J. F. Ziegler, *Nucl. Instrum. Methods* B219-220 (2004) 1027. http://www.srim.org.

[19] P.N. Ostroumov, Proc. Particle Accelerator Conf., Knoxville, USA, 2005, p. 3301.

[20] D. Berkovitz and A. Facco, Preliminary design and beam dynamics simulation of the EURISOL linac post-accelerator, LNL/INFN 197/2003, November 2002.

[21] T. Lamy *et al.*, Proc. European Particle Accelerator Conf., Paris, France, 2002, p. 1724.

[22] L. Spitzer Jr., *Physics of Fully Ionized Gases* (Interscience, New York, 1956).

[23] J.L. Delcroix and A. Bers, *Physique des Plasmas* (EDP Sciences, Les Ulis, France, 1994), vol. 2, p. 357.

[24] R. Geller, T. Lamy and P. Sortais, *Rev. Sci. Instrum.* **77** (2006) 03B107.

[25] R. Geller, *Electron Cyclotron Resonance Ion Source and ECR Plasmas* (IOP Publishing, Bristol, 1996), sec. 1.3.6.2.

[26] R. Vondrasek, *Rev. Sci. Instrum*. **83** (2012) 02A913.

[27] N. Imai *et al.*, *Rev. Sci. Instrum.* **79** (2008) 02A906.

[28] R. Geller, *Rev. Sci. Instrum.* **71** (2000) 612.

[29] R.C. Pardo *et al.*, *Chin. Phys. Soc. (C)* **31**(S1) (2007) 219.

[30] F. Ames *et al.*, *Rev. Sci. Instrum.* **79** (2008) 02A902.

[31] F. Wenander, REXEBIS the Electron Beam Ion Source for the REX-ISOLDE project, CERN-OPEN-2000-320 (1998).

[32] S. Schwarz *et al.*, *Rev. Sci. Instrum*. **79** (2008) 02A706.

[33] R. Becker and O. Kester, *J. Phys.: Conf. Ser.* **2** (2004) 20.

[34] B.M. Penetrante *et al.*, *Phys. Rev*. **A43** (1991) 4861.

[35] S. Schwarz *et al.*, *J. Instrum.* **5** (2010) C10002.

[36] R.B. Moore and G. Rouleau, *J. Mod. Opt.* **39** (1992) 361.

[37] G. Bollen, *Nucl. Phys.* **A616** (1997) 457c.

[38] G. Savard *et al.*, *Phys. Lett.* **A158** (1991) 247.

[39] D. Habs *et al.*, *Hyperf. Interact.* **129** (2000) 43.

[40] L. Brillouin, *Phys. Rev.* **67** (1945) 260.

[41] R. Rao *et al.*, *Nucl. Instrum. Methods* **A427** (1999) 170.

[42] R. Kirchner, *Nucl. Instrum. Methods* **B204** (2003) 179.

[43] A. Herlert *et al.*, *Hyperf. Interact.* **199** (2011) 211.

[44] S. Sturm, Implementation of a high resolution mass resolving technique for REXTRAP and investigation of space charge related phenomena, Master's Thesis, Universitat Heidelberg, 2007.

[45] A Lapierre *et al.*, *J. Instrum.* **5** (2010) c07001.

[46] S. Kondrashev *et al.*, Proc. European Particle Accelerator Conf., Lucerne, Switzerland, 2004, p. 1402.

[47] EURISOL Design Study Final Report, Ed. J. Cornell, 2009, http://www.eurisol.org/eurisol_design_study_final_report-1156.html.

[48] M. Grieser *et al.*, *Eur. Phys. J. – Spec. Topics* **207** (1) (2012) 1.

[49] M. Lindroos and M. Mezzetto, *Annu. Rev. Nucl. Part. Sci.* **60** (2010) 299.

Bibliography

I.G. Brown, The Physics and Technology of Ion Sources (Wiley, New York, 1989).

F.F. Chen, Introduction to Plasma Physics and Controlled Fusion, Vol. 1, Plasma Physics (Plenum Press, New York, 1974).

P. Delahaye, O. Kester, C. Barton, T. Lamy, M. Marie-Jeanne and F. Wenander, Evaluation of charge-breeding options for EURISOL, Eur. Phys. J. **A46** (2010) 421.

R. Geller, Electron Cyclotron Resonance ion Source and ECR Plasmas (IOP, Bristol, 1996).

O. Kester, Charge breeding application of EBIS/T devices, AIP Conf. Proc. **1099** (2009) 7.

T. Lamy et al., Status of charge breeding with electron cyclotron resonance ion sources, Rev. Sci. Instrum. **77** (2006) 03B101.

Z.H. Shun, Ion Sources (Springer, Berlin, 1999).

F. Wenander, Charge breeding of radioactive ions with EBIS and EBIT, J. Instrum. **5** (2010) C10004.

F. Wenander, Charge state breeders: on-line results, Nucl. Instrum. Methods **B266** (2008) 4346.

F. Wenander, Charge breeding techniques, CERN-AB-2004-035.

B. Wolf, Handbook of Ion Sources (CRC Press, Boca Raton, FL, 1995).

Relevant conference Proceedings

Proceedings of the International Workshop on ECR Ion Sources

Proceedings of the International Symposium on EBIS/T

Proceedings of the International Conference on Radioactive Nuclear Beams (now discontinued)

Proceedings of the International Conference on Electromagnetic Isotope Separators and Techniques Related to Their Applications

Technological Aspects: High Voltage

D.C. Faircloth
Rutherford Appleton Laboratory, Chilton, UK

Abstract
This paper covers the theory and technological aspects of high-voltage design for ion sources. Electric field strengths are critical to understanding high-voltage breakdown. The equations governing electric fields and the techniques to solve them are discussed. The fundamental physics of high-voltage breakdown and electrical discharges are outlined. Different types of electrical discharges are catalogued and their behaviour in environments ranging from air to vacuum are detailed. The importance of surfaces is discussed. The principles of designing electrodes and insulators are introduced. The use of high-voltage platforms and their relation to system design are discussed. The use of commercially available high-voltage technology such as connectors, feedthroughs and cables are considered. Different power supply technologies and their procurement are briefly outlined. High-voltage safety, electric shocks and system design rules are covered.

1 Introduction

1.1 Uses

High voltages (HV) are a fundamental component of ion sources, they are used for:

- extracting beams from plasmas (up to 50 kV),
- accelerating beams (up to 28 000 kV),
- initiating discharges and pre-ionizing gases (up to 20 kV),
- focusing and deflecting beams (up to 50 kV),
- suppressing unwanted particles (up to 5 kV).

1.2 Challenges

Ion sources are particularly challenging for HV design because they often operate at high temperatures in magnetic fields with large numbers of free charge carriers and stray beams. In negative ion sources, caesium often coats surfaces and encourages breakdowns. To keep efficiencies high and to minimize emittance and beam size, ion sources are designed to be compact. Unfortunately, this only makes the high-voltage aspect of the design even more difficult. There are very rarely any experimental data available for the specific combination of conditions present to aid design calculation. Therefore new source designs often have to undergo several prototype iterations. Modelling can help point to problems, but test-stands and a ‘suck it and see’ approach are essential because problems often emerge only after extended periods of operation.

High voltages are an obvious electrocution safety hazard and must be treated with great care. Safety systems must be put in place to prevent accidental electric shock. Also, ion sources often run with explosive gases such as hydrogen, and high-voltage sparks are an ignition source. Compliance with European ATEX directives [1] is essential.

1.3 Aims

The main aim of high-voltage design for ion sources is as follows: to produce reliable breakdown where it is wanted and to prevent breakdown where it is unwanted.

Reliable breakdown must occur in the discharge volume and any pre-ignition system. Breakdown must be prevented (or its occurrence minimized to acceptable levels) in the extraction, acceleration, focusing, deflection and suppression systems. Particular care must be given to the design of the extraction system because it is always right next to the discharge volume.

1.4 Factors affecting HV breakdown

High-voltage breakdown occurs when the electric field in a system becomes high enough to cause electron avalanches (see section 3.4). Electric field is the potential gradient, or the rate at which the voltage changes per unit length. It has many names: electric field strength, electric field intensity, stress, or just E. It is expressed in units of V m^{-1}, kV m^{-1}, kV mm^{-1}, kV cm^{-1} or MV m^{-1}.

In general, high-voltage breakdown is most likely to occur where the electric field is the highest, but this depends on various factors. Different insulating materials under different conditions will break down at different electric field strengths. The electric field where a material breaks down is called the material's dielectric strength and is expressed in the same units as electric field.

Different gases have different dielectric strengths. Increasing a gas's temperature will decrease its dielectric strength, as will decreasing its pressure. However, at very low pressures the dielectric strength sharply increases again, as we enter the realm of vacuum insulation (see section 3.7 on the Paschen curve).

Solid insulators generally have a much greater dielectric strength than gaseous ones. Gaseous and liquid insulators generally recover after breakdown, whereas breakdown of solid insulators usually causes irreversible damage.

The geometry of the electrodes plays a critical role in the type of discharge produced and the overall breakdown voltage. Electrode surface roughness can also reduce the voltage at which a system breaks down by causing localized field enhancements on a microscopic level. This is why high-voltage electrodes are often highly polished.

Another very important factor determining the voltage at which a system will break down is the presence of insulating surfaces. Surfaces are inevitable in high-voltage systems because the electrodes must be supported somehow. Solid insulators may have a very high dielectric strength in their bulk volume (i.e., their resistance to being punctured by the high field.) However, their surfaces can be very weak. Once initiated, discharges can propagate along insulator surfaces in very low field strengths. This often causes permanent damage to the insulator surface.

The point where the insulator surface meets the electrode can lead to localized field enhancements that can initiate discharges, dramatically lowering the breakdown voltage (see section 4.3.2 on triple junction effects).

Insulators and insulator surfaces can charge up, affecting the electric field. This can either increase or decrease the voltage at which the system breaks down depending on polarity. The actual mechanism of high-voltage breakdown is statistical by nature. No two sparks are ever the same. This combined with the complexities of charging mean that it is impossible to predict the exact voltage at which a system will break down.

2 Calculating electric fields

2.1 Introduction

Knowing the electric field strengths present in a high-voltage system is essential to predicting its performance. Electric fields can be calculated using equations, measured in experimental analogues, or solved with numerical techniques.

2.2 Maxwell's equations

Maxwell's equations are a set of four equations that completely describe the relationships between electric and magnetic fields. They define not only the shape and strength of the electromagnetic field, but also how it varies with time. The equations can be expressed in different ways, but the most compact form is in differential notation:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{1}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{2}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{3}$$

$$\nabla \cdot \mathbf{D} = \boldsymbol{\rho} \tag{4}$$

where

E is the electric field strength,

B is the magnetic flux density,

H is the magnetic field strength,

J is the current density,

D is the electric flux density,

$\boldsymbol{\rho}$ is the charge density.

Equation (1) describes how a time-varying magnetic field generates an electric field. It explains how electrical generators work. It is the differential form of Faraday's law of induction.

Equation (2) describes how electrical currents produce magnetic fields. It explains how electromagnets and motors work.

Equation (3) describes the shape of the magnetic field. It implies that magnetic monopoles do not exist.

Equation (4) describes the shape of the electric field in the presence of electrical charges. Both point and diffuse charges can exist. It is the differential form of Gauss's law.

2.3 Poisson and Laplace equations

2.3.1 Derivation

If no magnetic fields are present Eq. (1) becomes

$$\nabla \times \mathbf{E} = 0 \tag{5}$$

The definition of electric field is the rate at which the scalar potential ϕ varies with distance:

$$\mathbf{E} = -\text{grad}\,\phi \tag{6}$$

The electric field and electric displacement field (electric flux density) are related by the electrical permittivity ε:

$$\mathbf{D} = \varepsilon \mathbf{E} \tag{7}$$

Substituting Eq. (6) into Eq. (7) gives

$$\mathbf{D} = -\varepsilon\,\text{grad}\,\phi \tag{8}$$

Substituting Eq. (8) into Eq. (4) gives Poisson's equation

$$\nabla^2 \phi = \frac{\rho}{\varepsilon} \tag{9}$$

which can also be written as

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = \frac{\rho}{\varepsilon} \tag{10}$$

If no charges are present, Eq. (10) becomes Laplace's equation:

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = 0 \tag{11}$$

2.4 Solving Laplace's equation

2.4.1 Infinite parallel-plate capacitor

Laplace's equation can be used to calculate the electric fields present in different geometries. Consider an infinite parallel-plate capacitor, as shown in Fig. 1. One plate is grounded, and the other is held at a potential V. The plates are a distance d apart.

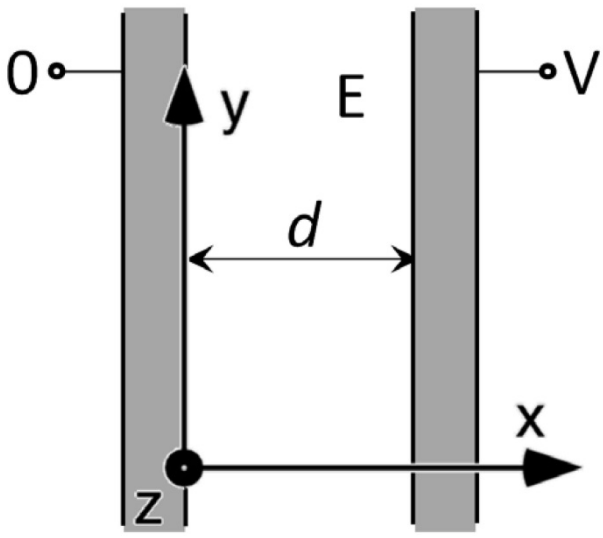


Fig. 1: Infinite parallel-plate capacitor

If the axis are defined as shown in Fig. 1, then the potential ϕ between the plates does not vary in the y or z directions:

$$\frac{\partial \phi}{\partial y} = \frac{\partial \phi}{\partial z} = 0 \tag{12}$$

Therefore the second derivative is also zero:

$$\frac{\partial^2 \phi}{\partial y^2} = \frac{\partial^2 \phi}{\partial z^2} = 0 \tag{13}$$

Substituting into the Laplace equation (11) gives

$$\frac{\partial^2 \phi}{\partial x^2} = 0 \tag{14}$$

Integrating this gives

$$\frac{\partial \phi}{\partial x} = c_1 \tag{15}$$

and integrating again gives

$$\phi(x) = c_1 x + c_2 \tag{16}$$

At $x = 0$, $\phi = 0$, and at $x = d$, $\phi = V$, so

$$c_1 = \frac{V}{d} \tag{17}$$

and

$$c_2 = 0 \tag{18}$$

So Eq. (16) becomes

$$\phi(x) = \frac{V}{d} x \tag{19}$$

From Eq. (6) we get

$$\mathbf{E} = -\frac{V}{d} \tag{20}$$

or

$$|\mathbf{E}| = \frac{V}{d} \tag{21}$$

2.4.2 Infinite coaxial line

Consider and infinite coaxial line as shown in Fig. 2. The inner cylindrical conductor is held at a potential ϕ_1 and the outer cylindrical conductor is held at a potential ϕ_2. The radii of the inner and outer conductors are r_1 and r_2, respectively.

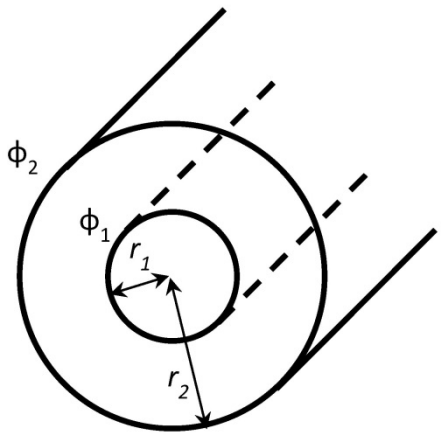


Fig. 2: Infinite coaxial line

The Laplace equation (11) can be expressed in terms of cylindrical symmetric coordinates:

$$\frac{1}{r}\cdot\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2\phi}{\partial\varphi^2}+\frac{\partial^2\phi}{\partial z^2}=0 \tag{22}$$

If the axes in Fig. 2 are defined so that the z-axis runs along the centre of the coaxial line, then the voltage ϕ does not vary with angle φ or distance z. So

$$\frac{\partial^2\phi}{\partial\varphi^2}=\frac{\partial^2\phi}{\partial z^2}=0 \tag{23}$$

Substituting into the cylindrical systematic form of Laplace's equation (22) gives

$$\frac{1}{r}\cdot\frac{\partial}{\partial r}\left(r\frac{\partial\phi}{\partial r}\right)=0 \tag{24}$$

By integration and substitution we get

$$\phi(r)=\phi_1+\frac{\phi_1-\phi_2}{\ln\left(\frac{r_1}{r_2}\right)}\left(\ln r-\ln r_1\right) \tag{25}$$

From Eq. (6),

$$E(r)=\frac{\phi_1-\phi_2}{r\ln\left(\frac{r_2}{r_1}\right)} \tag{26}$$

The field is greatest on the surface of the inner conductor where $r = r_1$:

$$E_{\max} = \frac{\phi_1 - \phi_2}{r_1 \ln\left(\frac{r_2}{r_1}\right)} \tag{27}$$

2.5 Electric fields measured in experimental analogues

2.5.1 Introduction

Algebraically solving the electric field distribution is relatively easy for the idealized geometries shown in the previous section. It is possible to solve the fields in real-world geometries as long as the electrode surfaces are made up of basic shapes, although as the electrode geometries become more complex the equations rapidly become very difficult to solve. When the electrode surface geometry cannot accurately be described by an equation, then algebraic solution of the fields is no longer possible.

Although rarely used these days, it is possible to set up an experimental analogue to directly measure electric potentials using a model of the electrode geometry.

2.5.2 Teledeltos paper

Teledeltos paper is an electrically conductive paper. It is formed by coating carbon on one side of a sheet of paper. This provides a sheet of uniform resistance, with isotropic resistivity in each direction.

Teledeltos paper can be used to create a two-dimensional model of any scalar field, such as the electric field. Conductive paint is used to paint the shape of the electrodes on the paper. Power supplies are used to apply voltages to the painted-on electrodes. A probe connected to a voltmeter is then used to measure the voltage at various points between the painted-on electrodes. In this way a map of the scalar electric potential ϕ can be constructed. Then equipotential lines can be drawn and electric field strengths calculated.

The resistances involved are in the kilohm range. This is low enough that the Teledeltos paper may be used with safe low voltages, yet high enough that the currents remain low, avoiding problems with contact resistance.

2.5.3 Electrolytic tank

A similar technique to the Teledeltos paper method involves the use of a tank of conductive fluid [2]. Scaled or actual size copies of the electrodes are placed in a tank of fluid, which could be a solution of $NiSO_4$, $CuSO_4$ or just pure distilled water. Voltages are applied to the electrodes and a probe is used to measure the voltage at various positions to generate a map of scalar electric potential in the system.

Using a fluid instead of a conductive surface has two main advantages: three-dimensional field maps can be generated; and the actual electrodes can be tested rather than conductive paint drawings.

In reality it is quite difficult to accurately measure three-dimensional field maps, so electrolytic tanks were often used with a layer of fluid to generate two-dimensional field maps instead. By tilting the tank, it is possible to generate two-dimensional field maps at different 'slices' through the electrode geometry under test.

Figure 3 shows a sophisticated electrolytic tank system capable of automatically plotting equipotential lines. A simple electrode geometry consisting of two circular electrodes is being tested. The electrodes sit in a tank of electrolytic fluid and a voltage is applied between them. The voltage measurement probe is attached to the lid of the tank and can be moved in the x and y directions using

two motor drives. When the lid is closed, the probe is immersed in the electrolytic solution and, with the help of control electronics, starts to follow an equipotential line. A pen is mounted on the probe, which plots the trajectory of the probe on a sheet of paper mounted on the lid, thus drawing the equipotential line. This process is repeated for different probe voltages until an adequate map of the field between the electrodes is obtained.

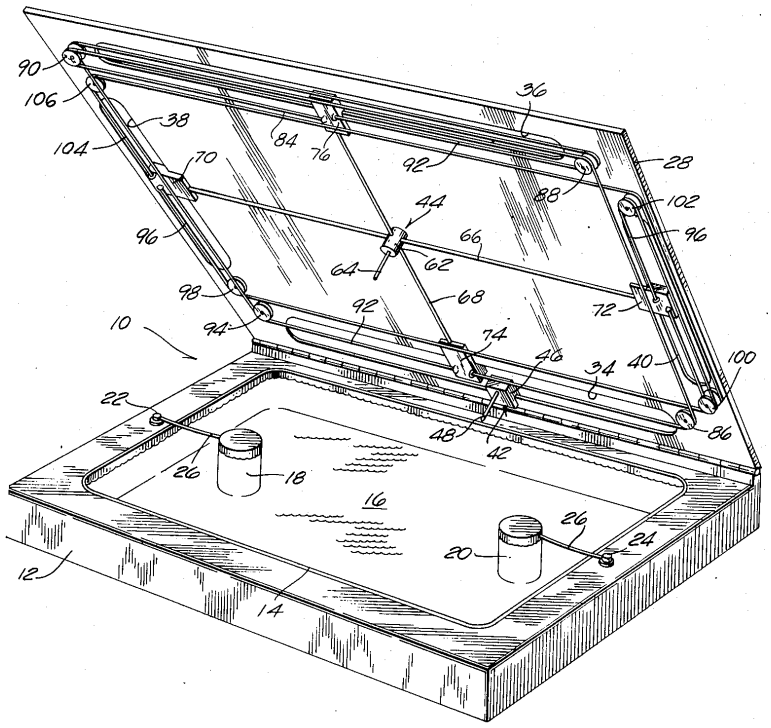


Fig. 3: An automated electrolytic tank measuring the equipotentials between two circular electrodes (from United States Patent No. 3764900).

2.6 Solving electric field distributions with numerical techniques

2.6.1 Introduction

Thankfully, plotting fields with experimental analogues as described in the previous section is now a thing of the past. Today, experimental analogues are only used in student laboratories as demonstrations. The increased availability of computing power from the 1970s onwards has made solving electric fields using numerical methods commonplace.

2.6.2 Finite elements

The finite element technique involves breaking the electrode geometry and the space between the electrodes into discrete regions called elements. This process is called meshing because the resultant pattern resembles the mesh in a fisherman's net, or the delicate mesh pattern in a net curtain. A typical two-dimensional finite element mesh is shown in Fig. 4.

The corners of each element are called nodes. Using the discrete form of Poisson's equation, it is possible to define the potential of each node in terms of its neighbour's potentials and their distance away (the length of the element's side). In this way a set of simultaneous equations defining the potential of every node can be defined. It is then just a case of solving all the simultaneous equations – a problem ideally suited to computers.

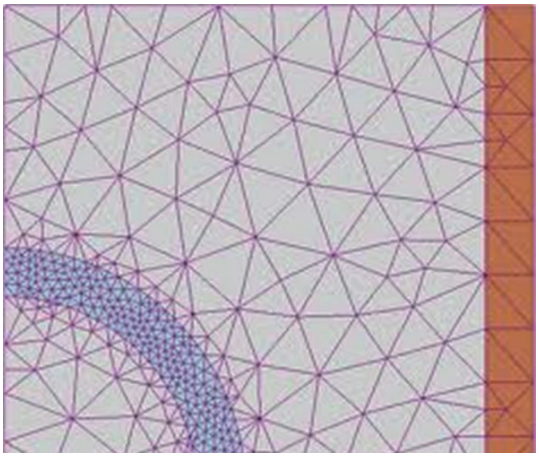


Fig. 4: Finite elements in a two-dimensional mesh

2.6.3 Element shape and size

The example shown in Fig. 4 is two-dimensional and uses triangular elements, but it is equally possible to use square or rectangular elements. Three-dimensional problems can also be solved using tetrahedral or brick-shaped elements.

The accuracy of the field distribution obtained depends on the size of the elements: smaller elements and more dense elements will yield higher-accuracy results. However, smaller elements means more elements. More elements means more simultaneous equations to solve. It takes longer to solve more simultaneous equations, so more computing resources are required. Therefore, when meshing a geometry, it is always important to be as efficient with element size as possible: do not make the mesh high density everywhere, only where it is needed. This is most important in three dimensions, where very large problems can take hours to solve.

2.6.4 Boundaries

To solve simultaneous equations, there must be some known values. In electric field problems, these are nodes that lie on the conducting electrode surfaces. Electrodes that have known voltages applied to them will have their nodes set to that specific voltage. Electrodes that are left to float in potential can have a condition imposed to make all the nodes have an equal value.

An important technique used to reduce the total number of elements in a model is the use of symmetry planes. If the electrode geometry has reflectional symmetry, then only half the geometry needs to be modelled. The equipotential field lines pass through the line of reflectional symmetry at right angles. This means there is no electric field component normal to the line of symmetry. In the model, the line of symmetry becomes a boundary where the electric field normal to the boundary must be zero. This imposes a condition on the nodes on that boundary.

Rotational symmetries can also be exploited to reduce the number of elements. This is especially true for three-dimensional geometries with an axis of rotational symmetry. The geometry can be represented as a two-dimensional plane, which will be much quicker to solve and give an identical result to solving the full three-dimensional geometry.

It may also be possible to reduce a three-dimensional geometry to a two-dimensional geometry by taking a slice through it. However, unless the objects being modelled are very long in the direction orthogonal to the slice, this will only be an approximation to the true field.

2.6.5 Solution methods

There are many ways to solve the simultaneous equations produced by the finite element technique. Here just are a few of the more commonly used techniques:

Direct methods:
- Gaussian elimination
- LU decomposition method

Iterative methods:
- Mesh relaxation methods
  - Jacobi
  - Gauss–Seidel
  - Successive over-relaxation (SOR) method
  - Alternating directions implicit (ADI) method
- Matrix methods
  - Thomas tridiagonal form
  - Sparse matrix methods
  - Conjugate gradient (CG) methods
  - Multi-grid (MG) method

It is perfectly possible to write your own solver, and this is the subject of many undergraduate and Masters degree projects; however, you will be reinventing the wheel. There are plenty of commercial and open-source codes available to do this written in every language for every platform.

2.6.6 Professional modelling software

There are many advantages to using existing codes other than time saved in not having to develop your own code. Efficiently meshing a geometry, especially a three-dimensional one, is a difficult mathematical problem in itself. Even the expensive commercially written codes sometimes have problems meshing geometries containing a lot of fine details.

There are many codes available. A particularly good free two-dimensional code is Poisson Superfish, developed by the code group at Los Alamos National Laboratory in the USA [3]. Here is a non-exhaustive list of some commercially available codes:

- ANSYS
- Cobham Opera
- COMSOL
- CST Studio Suite
- SimIon
- INP

The commercially available codes can be expensive, especially the three-dimensional ones that are capable of solving magnetic and other transient problems. The advantages of professional codes are:

- Quick and easy geometry importing directly from CAD files.
- Advanced meshing tools.
- Multicore solvers.
- Proven benchmarked code.
- User support helpdesks and large user communities.

3 Electrical discharges

3.1 Introduction

There are many different types of electrical discharges. Some are found in nature, such as lightning, aurora and Saint Elmo's fire. This section will cover the different types of electrical discharges found in engineering. They can be classified in different ways: Are they localized or do they completely bridge the gap between the electrodes? Are they transient or continuous? Are they wanted or unwanted?

'Wanted' electrical discharges are used to create the plasma in ion sources, and these can be glow or arc discharges (see sections 3.13 and 3.14). These types of discharges must be reliably and repeatably produced. Starting a discharge is referred to as striking the discharge.

'Unwanted' discharges in high-voltage systems can lead to catastrophic failure, but not always: it depends if the discharge remains localized in a medium that self-heals. A good example is a corona discharge from a sharp point in air (see section 3.11). If the system geometry is large enough, then the corona can fizz away without much problem. The corona will cause a few extra microamps to be drawn from the high-voltage power supply, create electromagnetic noise, and a small amount of ozone, but it will not lead to an overall failure of the high-voltage system.

However, if the localized discharge is in a non-self-healing medium, then problems can occur. A good example is a void (or hole) in solid insulation (see section 3.10). Initially the discharge remains localized, just causing electromagnetic noise, but slowly it will damage the insulator material and potentially lead to a catastrophic failure.

An 'unwanted' discharge that completely bridges the gap between the electrodes is referred to as a 'breakdown' or a 'flashover' and is often visible as a spark accompanied by an audible crack. If a solid insulating medium is between the electrodes and the discharge breaks through it, the insulator is said to have been 'punctured' or 'ruptured'. The weakest point of an insulator is its surface (see section 3.9.1). A discharge that propagates along the surface is referred to as 'tracking' because it often leaves carbonized tree- or fern-like tracks on the insulator surface. A similar term in solid insulation is 'treeing' (see Fig. 5). The paths left by high-voltage discharges can be quite beautiful.

Fig. 5: Treeing in a block of Plexiglas. © Bert Hickman

Electrical discharges are plasmas. Plasmas are chemically reactive, so they can cause chemical damage to insulators. Plasmas contain fast-moving ions that can damage electrodes by sputtering. This must be considered when designing the plasma chamber for ion sources. Discharges also produce a significant amount of ultraviolet (UV) light that can degrade insulator materials.

The route to high-voltage breakdown depends on the degree of field uniformity in the system. Discharges can be initiated in localized regions of high field strength, then continue to propagate into regions where the field is not strong enough to start a discharge, but it is strong enough to allow the discharge to continue to propagate. A well-designed high-voltage system removes regions of high field by rounding electrodes and screening sharp points (see section 4.2).

Whether the discharge is transient or continuous can also depend on how the discharge gets its power (see section 3.12).

3.2 Ionization

3.2.1 Background ionization

Electrical discharges are fundamentally about ionizing materials (gases, liquids or solids) to produce charge carriers (electrons and ions). There is always a low level of background ionization occurring due to background radiation. Background radiation comes mainly from natural sources such as radon gas and cosmic rays but also includes a small contribution from manmade nuclear tests and accidents.

3.2.2 Electron impact ionization

The most important ionization mechanism in electrical discharges is electron impact ionization. Free electrons impact on the outer electron orbitals of atoms and molecules all the time. If the free electrons have enough energy, they can knock the outer electron out of its orbital, leaving a positive ion and another free electron. The free electrons get their energy from externally applied fields. The field applied to create a discharge is usually the electric field; however, a magnetic field can also be applied in the case of inductively coupled discharges.

Inductively coupled discharges can only be produced with a time-varying magnetic field. The free electrons are actually accelerated by the spatially varying, non-conservative electric field produced by the time-varying magnetic field. This is the Maxwell–Faraday law of electromagnetic induction shown in Eq. (1).

Electric-field-driven discharges can also be produced by a.c. electric fields, but they extinguish and reignite each time the electric field reverses polarity.

3.2.3 Photo-ionization

Photons with enough energy can also cause ionization. These are usually photons in the UV part of the spectrum. The discharge itself can create UV light when excited atoms relax or ions recombine with electrons. This plays an important role in helping the discharge to propagate.

An excited (or metastable) state is where the electrons are still attached to the atom; they occupy higher ‘excited’ orbitals. An exited atom can go on to be completely ionized by absorbing more energy, either from another photon or by being hit by a free electron.

If the excited atom does not receive any more energy, it will relax back into a lower state by emitting a photon, which could go on to excite the electrons in another atom.

3.3 Discharge current/voltage landscape

There are many names for the basic phenomenon that comprise electrical discharges, such as avalanches, streamers, corona, leaders, glows and arcs. These will all be defined in the rest of this

section. They cover a huge range of different currents. To provide a conceptual landscape in which these discharges exist, it is useful to consider a simple set-up with two electrodes shown in Fig. 6(a). A d.c. voltage can be applied between the electrodes and the resulting current measured.

The general current–voltage characteristic of such a discharge is summarized in Fig. 6(b). The exact shape of the curve depends on the type of gas, pressure, electrode geometry, electrode temperatures, electrode materials and any magnetic fields present.

At low voltages, the current between two electrodes is extremely small, but it slowly increases as the voltage between the electrodes increases (as shown in the bottom left corner of the graph in Fig. 6(b)). This tiny current comes from charge carriers produced by background ionization. They are swept out of the gap by the electric field between the electrodes that is created by the applied voltage. There are only enough charge carriers produced by background radiation for a few nanoamps of current, so the current quickly saturates. The voltage can then be increased with no increase in current. The ions and electrons are pulled towards the electrodes through the gas molecules interacting with them as they go.

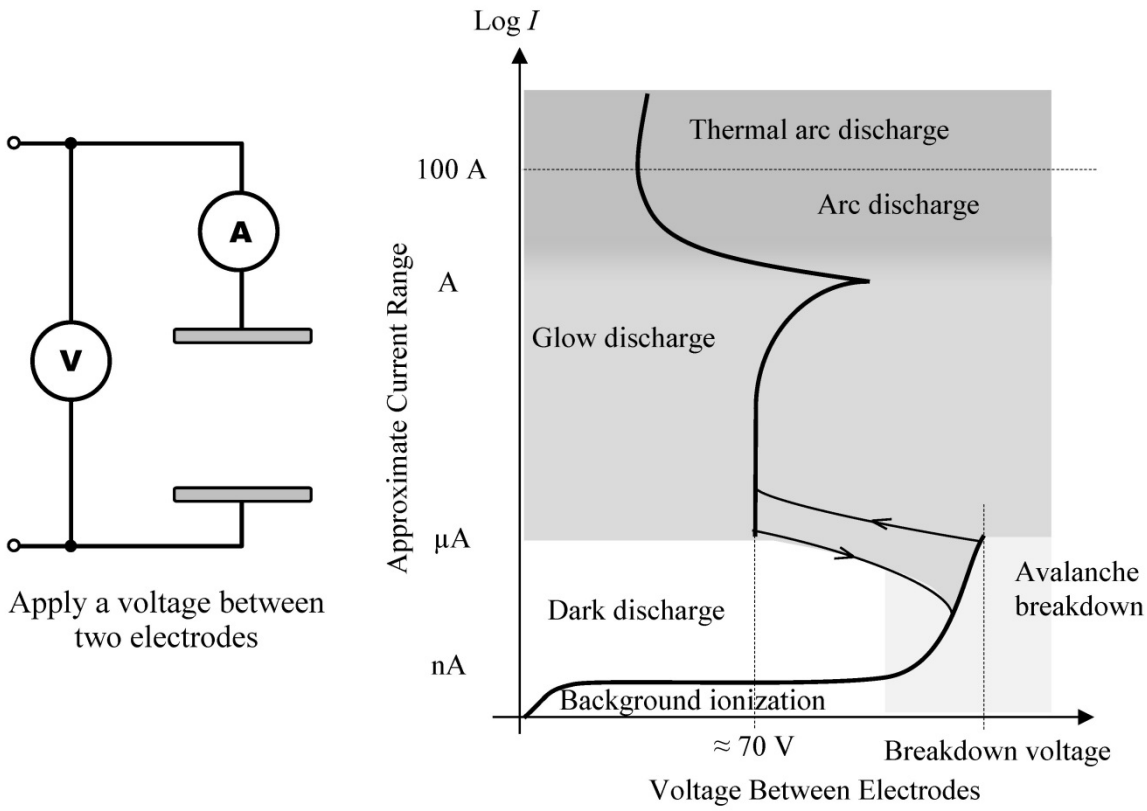


Fig. 6: The current–voltage characteristic of a typical electrical discharge between two electrodes

3.4 Avalanche breakdown

3.4.1 Introduction

The voltage between the electrodes shown in Fig. 6(a) is increased until the applied electric field is high enough to accelerate the electrons to the ionization energy of the gas. At this point the current rapidly increases, as shown in the bottom right corner of the graph in Fig. 6(b). The electrons ionize the neutral atoms and molecules, producing more electrons. These additional electrons are accelerated to ionize even more atoms, producing even more free electrons in an avalanche build-up process, as shown in Fig. 7.

This is the moment of inception of a high-voltage breakdown, and if the conditions are right, it can cause complete flashover of the electrodes.

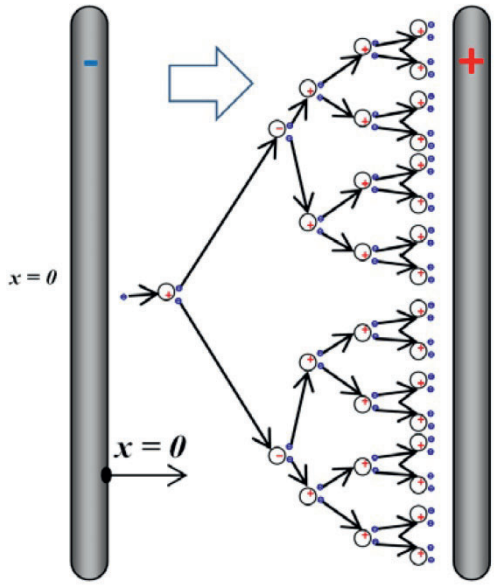


Fig. 7: The avalanche breakdown mechanism in electrical discharges

3.4.2 Townsend's primary ionization coefficient, α

The avalanche process was first mathematically described by John Townsend in 1897. Consider an avalanche discharge between two electrodes. The position between the electrodes in Fig. 7 is given by the variable x. The number of additional free electrons $\mathrm{d}n_x$ produced in a distance $\mathrm{d}x$ depends on the number of electrons at that point n_x and the primary ionization coefficient α:

$$\mathrm{d}n_x = n_x \alpha \, \mathrm{d}x \tag{28}$$

The primary ionization coefficient α is the number of additional electrons produced per unit length. By integration and the fact that $n_x = n_0$ at $x = 0$, this gives

$$n_x = n_0 \, \mathrm{e}^{\alpha x} \tag{29}$$

The number of free electrons (and ions) increases exponentially in the avalanche.

3.5 Streamers

3.5.1 Electric field in an avalanche

Figure 8(a) shows the distribution of particles in a single avalanche. The discharge has a negative head, comprising the avalanche of free electrons, and a positive tail, comprising the positive ions left behind after the ionization avalanche has passed. The ions are at least 1800 times heavier than the electrons so they take much longer to accelerate than the electrons. Compared to the fast free electrons, the ions only move a tiny distance from where they were born.

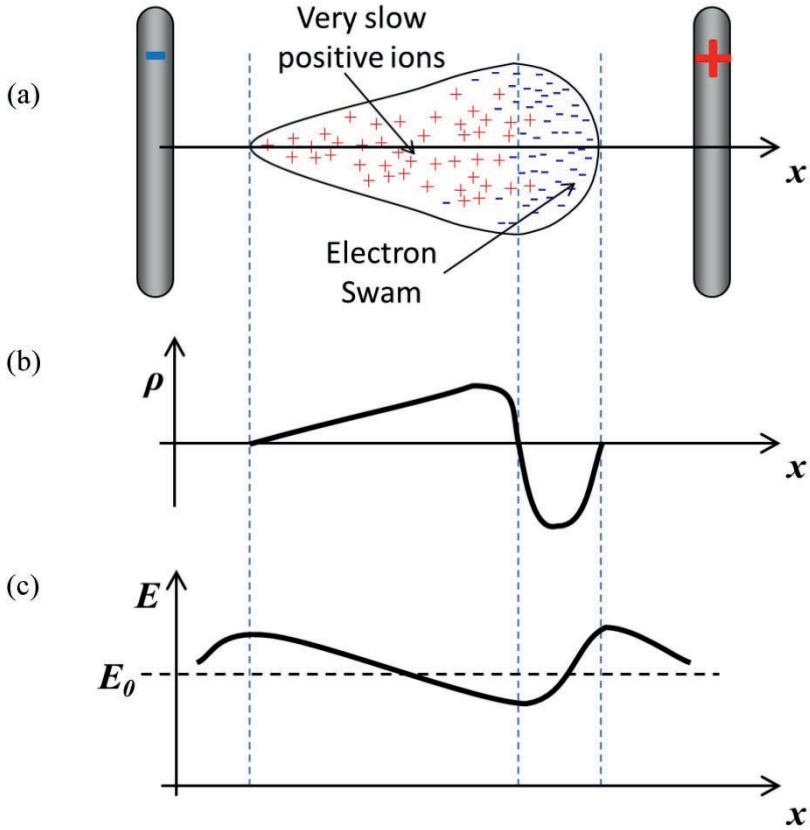


Fig. 8: Distribution of (a) particles, (b) charge density and (c) electric field distribution in a single avalanche discharge.

The difference in mobility of the ions and electrons creates the charge distribution shown in Fig. 8(b), which in turn produces the electric field distribution shown in Fig. 8(c). If the discharge was not present, the electric field between the electrodes would be E_0. Even though the total amount of positive and negative particles is approximately equal, the distribution of charges in the avalanche causes an increase in the field in front of and behind the discharge. This is caused by the negative electron avalanche being closer to the positive electrode and the positive ions being closer to the negative electrode.

3.5.2 Streamer formation

The avalanche emits photons that can ionize nearby atoms creating free electrons as described in section 3.2.3. Any free electrons created in the higher-field region in front of or behind the initial avalanche will go on to produce additional avalanches, as shown in Fig. 9.

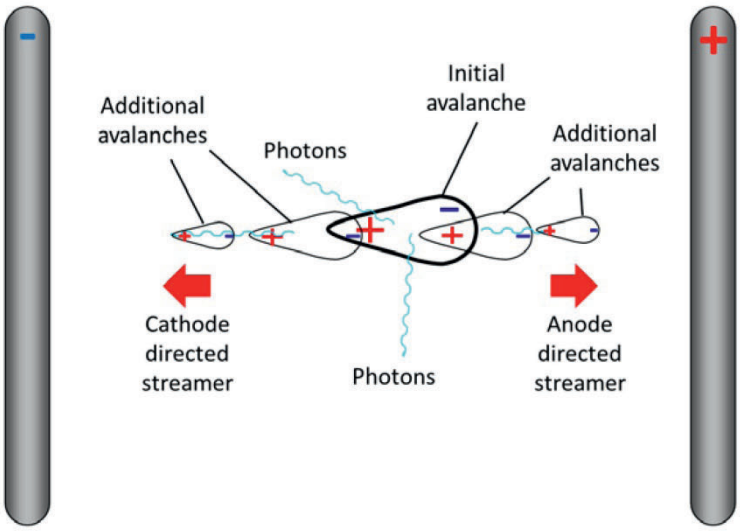


Fig. 9: Streamer formation made up of a series of avalanches propagating in both directions from the initial avalanche.

Each additional avalanche creates more photons and further enhances the electric field. In this way a chain of avalanches propagates from the head and tail of the initial avalanche. The name for this chain of avalanches is a streamer. Streamers propagate in both directions, as shown in Fig. 9. Eventually the chain of avalanches bridges the gap between the electrodes, creating an ionized conductive channel. This is the moment the gap actually flashes over. As much current flows from one electrode to the other as the power supply can deliver.

3.5.3 Townsend's secondary ionization coefficient, γ

Townsend also considered secondary ionization processes. In addition to photoionization and metastable effects, ion impact ionization should also be considered. This is where the positive ions impact on neutral atoms or the cathode electrode surface. Positive ions impacting on the cathode significantly enhance the growth of the discharge at the cathode.

The secondary ionization coefficient γ is the number of secondary electrons produced per electron in the primary avalanche. It includes all secondary processes:

$$\gamma = \gamma_{\mathrm{p}} + \gamma_{\mathrm{m}} + \gamma_{\mathrm{ion}} \tag{30}$$

Take a pair of electrodes a distance d apart and assume all electrons that leave the cathode and are produced by primary and secondary processes reach the anode. Then using Eq. (29) the total current in the discharge can be shown to be

$$I = \frac{I_0 \, \mathrm{e}^{\alpha d}}{1 - \lambda(\mathrm{e}^{\alpha d} - 1)} \tag{31}$$

When breakdown occurs, the current increases very rapidly. This allows a criterion for breakdown to be obtained from Eq. (31):

$$\gamma(\mathrm{e}^{\alpha d} - 1) = 1 \tag{32}$$

The term $\mathrm{e}^{\alpha d}$ is usually very large, so Eq. (32) becomes

$$\gamma \, \mathrm{e}^{\alpha d} = 1 \tag{33}$$

This is the Townsend criterion for breakdown and gives the conditions required for a self-sustaining discharge. Eq. (32) can be used when modelling breakdowns.

3.6 Leaders

Leaders only occur in very long breakdowns (more than a metre) so they only happen when using extra-high voltages (EHV). Pre-injectors and tandems use EHV d.c. to accelerate beams after extraction from the ion source.

A leader is a hot plasma channel that starts from either the positive or negative electrode. The current from many streamers forms a highly ionized region at the electrode surface, which then propagates into the gap. The highly conductive leader channel is effectively at the same potential as the electrode. Leader growth is significantly slower than the fast streamers, and it propagates by creating more streamers at its head.

Leader formation is a vital component of lighting propagation, but its presence in high-voltage engineering is a rare and unwanted phenomenon.

3.7 Paschen curve

The breakdown voltage of a gas between two flat electrodes depends only on the electron mean free path and the distance between the electrodes. The electron mean free path is the average distance the electrons travel before hitting atoms, so it is the key factor defining the growth of an avalanche discharge. It is directly related to pressure. Fig. 10 shows how the breakdown voltage of hydrogen varies with the product of pressure p and distance d between electrodes. This was first stated in 1889 by Friedrich Paschen [4].

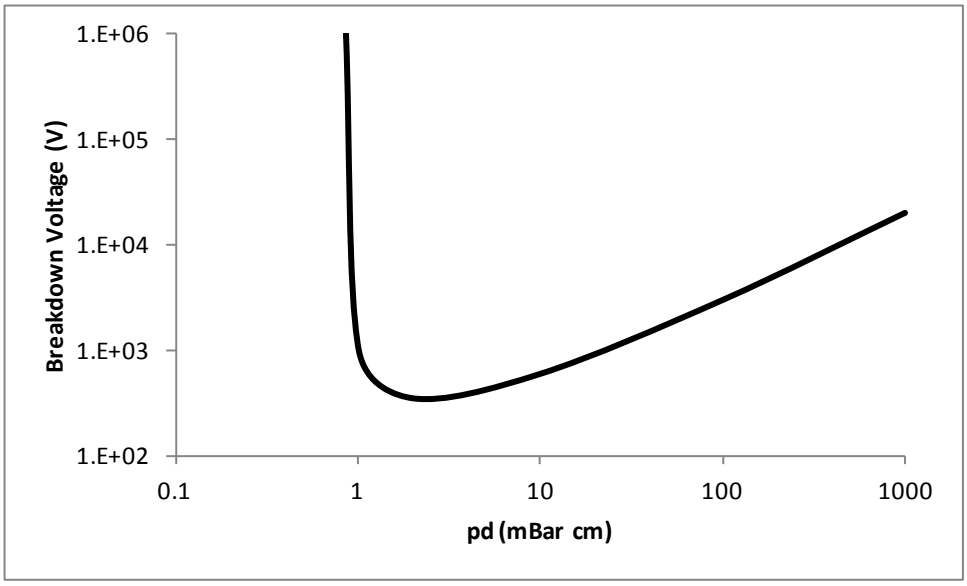


Fig. 10: The Paschen curve for hydrogen

At very high pressures the mean free path is very short. This means that the electrons never have enough time to be accelerated to the ionization energy before hitting an atom or molecule. This means that the breakdown voltage is high at very high pressures.

At very low pressures, the mean free path between collisions is longer than the distance between the electrodes. So, although the electrons can be accelerated to ionizing energies, they are unlikely to hit anything other than the anode. This means that the breakdown voltage is very high at very low pressures. This is why vacuum is such a good insulating medium.

Between these two extremes is a minimum whose position depends on the type of gas and the electrode material. This 'Paschen minimum' leads to a counterintuitive phenomenon: operating just below this minimum, electrodes further apart will have a lower breakdown voltage than those closer together. This is because, in a longer gap, there is more space for the electron avalanches to develop.

Ion sources often operate within an order of magnitude of the Paschen minimum because gas is added to a vacuum to create plasma for particle creation. Ion source engineers should always keep this in mind when trying to diagnose unwanted breakdowns. Figure 10 shows the Paschen curve for hydrogen, the Paschen curves for all other gases are very similar, with slightly different Paschen minima, ranging from 0.5 to 2 mbar cm and 100 to 800 V.

3.8 Vacuum breakdown

3.8.1 Introduction

Avalanche and streamer breakdown only occur at pressures and gap lengths to the right of the Paschen minimum shown in Fig. 10. To the left of the Paschen minimum is vacuum breakdown territory. This side of the Paschen minimum, the voltage required to break down a gap rapidly increases by several orders of magnitude.

Vacuum breakdown is a very different type of breakdown: a type in which the electrode surface plays a critical role. For breakdown to occur, a conductive plasma channel must bridge the gap between the electrodes. There are not enough atoms left between the electrodes to produce enough charge carriers, so they must instead come from the electrodes themselves.

3.8.2 Mechanisms

A surface can never be completely flat. Small or microscopic protuberances always exist at some scale. These cause localized field enhancements. When the voltage applied between the electrodes is high enough, small points on the cathode surface begin to emit electrons through quantum tunnelling processes. This highly localized current flow causes rapid localized overheating of the first cathode protrusion to start emitting. The protrusion explodes, creating a plume of vapour that is ionized and can cause breakdown.

The electrode material and surface finish are very important for vacuum breakdown. Highly polished surfaces are the best, but even they have a limit. The shiny smooth surface might still have microscopic oxide deposits in inclusions on the surface. The oxide deposits will generally be non-conductive, so they will have a permittivity different from vacuum. This step change in permittivity at the inclusion/oxide perimeter causes localized field enhancements. These localized field enhancements cause electrons to be emitted from the electrode. The resultant localized high current densities lead to a micro-explosion of material that is ionized, potentially causing breakdown.

3.8.3 High-voltage conditioning

High-voltage conditioning is essential to operating vacuum systems at very high field strengths. The conditioning process is very simple, but requires patience. The voltage is very slowly increased until the first sign of a discharge. The voltage is then held steady, or reduced slightly. When the discharge activity dies away, the voltage is very slowly increased again, until the discharge reappears. The voltage is then held steady, or reduced slightly. This process is repeated until the desired voltage is reached, or until a voltage is reached where the discharge activity does not die away.

Discharge activity might be detected through a sudden increase, or spike, in the supply current. If there is a window in the vessel, light might be visible, or if there is an antenna, electromagnetic noise could be picked up. Electrons emitted from the discharge are accelerated by the field and, if the voltage is high enough, could produce X-rays by bremsstrahlung. Depending on the configuration,

there might not be any indication of a discharge until a complete flashover and resulting voltage collapse.

Whichever way the discharge is detected, the conditioning process works by slowly destroying each surface emitter in a controlled manner. If the voltage is increased too quickly, the discharge activity is too violent: instead of removing the protuberance, the localized electrode explosion creates more surface irregularities. These additional protuberances caused by the explosion might actually make the situation worse. The key to high-voltage conditioning is patience: the voltage must be increased very slowly, often being held at a voltage for some time before attempting to increase it further.

High-voltage conditioning is sometimes referred to as 'spot knocking' for obvious reasons.

Depending on the maximum field strength required, the conditioning process could take minutes or days. Below 1 MV m^{-1} d.c., the conditioning process is relatively fast. Between 1 and 10 MV m^{-1} d.c., the conditioning process takes longer. The best way to monitor how conditioning is going is to record the number of transient discharges (or sparks) per hour. At very high fields the spark rate might never get better than a few sparks per hour. The tolerable spark rate depends on the application. If no sparks can be tolerated, then the system must first be conditioned to a higher field, then when the voltage is reduced to the operating level the spark rate drops almost to zero. For very high field strengths above 10 MV m^{-1} d.c., it is very difficult to condition the electrodes to give a spark rate of zero. The electrode material becomes very important at these sorts of field levels (see section 4.2.4).

3.9 Insulator breakdown

3.9.1 Surface breakdown

Something non-conducting has to hold the electrodes in position. This introduces an insulating surface between the electrodes. Insulator surfaces are the weakest point of any high-voltage system and care must be taken in their design (see section 4.3).

Discharges tend to propagate preferentially along surfaces because the surface is a good source of material to ionize and produce electrons. This allows surface avalanches to propagate along insulator surfaces in relatively low fields.

Triple junction effects (see section 4.3.2) where the insulator and the electrode meet can cause field enhancements, which can initiate discharges. Once started, the surface allows the discharge to propagate further than they would without the surface present.

The insulating surface can hold charge from previous discharge activity. Dependent on polarity, this can cause field enhancements, which can initiate surface discharges at lower voltages.

3.9.2 Tracking

Solid insulators are not self-healing like air and liquids. If there is a high-voltage surface breakdown, it can (but not always) cause permanent damage to the surface of the insulator. The high current that flows causes localized heating, leading to melting, carbonization or chemical changes of the insulator material and surrounding gases. This process is known as tracking: a conducting channel grows from one of the electrodes. It can be quite slow, building up a conducting track over a period of years, or it can be very quick, tracking over in one discharge event. The rate of growth depends on the voltages applied, the insulator material and its cleanliness. Once an insulator surface has completely tracked over, the electrodes are shorted together with a relatively low resistance, making application of a high voltage impossible. Some materials are track-resistant (in that they do not carbonize or form conductive compounds); these tend to be materials that do not contain hydrocarbons.

3.9.3 Bulk breakdown

Unlike its surface, the body of an insulator has a very high dielectric strength. The dielectric strength of a material is the field at which it fails or breaks down through its bulk. Gases and liquids also have dielectric strengths. In a solid, breakdown is always catastrophic; the material becomes punctured or ruptured with a track. Depending on the material, the track might be carbonized – it depends on what chemistry occurs when the material is turned into plasma. The shape of the track is defined by the shape of the electric field and the nature of the voltage applied. A uniform field will cause an almost straight track. A divergent field will cause a beautiful tree-like track as shown in Fig. 5.

The bulk of an insulator can also fail if it contains any voids. The effect of a void is twofold: firstly, the reduced permittivity in the void increases the field in the void; and secondly, the void is likely to contain a gas with a much lower dielectric strength than the insulator. Partial discharges in voids will eventually cause failure by slowly eroding their way through the insulation.

3.10 Partial discharges

Partial discharges are small high-voltage breakdowns that do not completely bridge the gap between the electrodes. They are caused by field enhancements created by changes in permittivity. This can either be voids in solid insulation (as discussed in the previous section) or mixtures of different insulator types.

The electric field in a void is higher because the permittivity of the void is lower than that of the surrounding bulk insulator. This causes the equipotential field lines to be concentrated in the void. Another way of looking at this is shown in Fig. 11(a). Consider that the void has a capacitance C_v, and the capacitance of the void is C_x to one electrode and C_y to the other. This makes a capacitive circuit between the electrodes, which can be thought of as a capacitive voltage divider. The void has a lower permittivity, and so a smaller capacitance. In a capacitive voltage divider the largest voltage drop is across the smallest capacitance, i.e., the void.

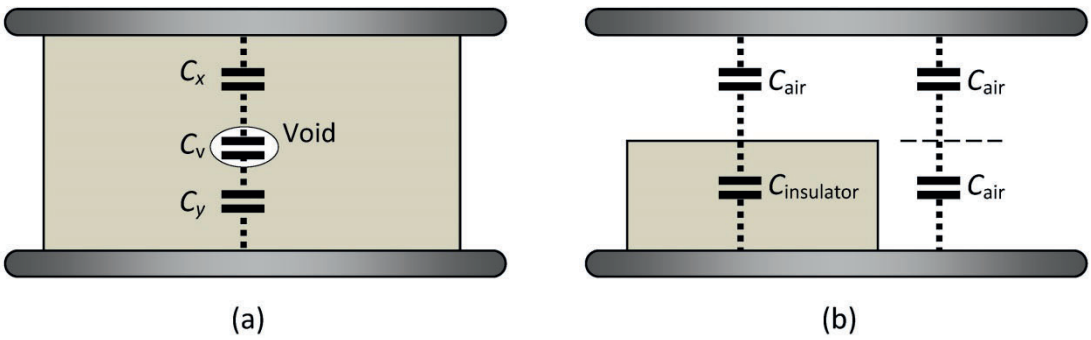


Fig. 11: Capacitive voltage divider explanation for how voids can cause field enhancement

The capacitive voltage divider analogy can also be used to explain why it is never a good idea to partially fill the gap between two electrodes with insulation. To increase the breakdown voltage between two electrodes, it would seem like a good idea to put some insulation between the electrodes. However, partially filling the gap, as shown in Fig. 11(b), actually increases the field in the region not filled with insulator. The capacitive voltage divider set up by C_{air} and $C_{insulator}$ will make the field in the air greater than it would be without the insulator present! The top electrode will break down to the insulator surface. This could be a partial discharge, or it could go on to puncture the solid insulation.

To increase the breakdown voltage between electrodes, the only option is to completely fill the gap between them with a material of higher dielectric strength, making sure that no voids are present.

This is why vacuum injection moulding is often used to make solid high-voltage insulators because this avoids the presence of bubbles or voids.

Composite insulators can be built, but special care must be taken to understand the voltage distribution in them caused by different material permittivities.

3.11 Corona discharges

3.11.1 Introduction

A corona discharge is another type of partial discharge. They occur when there is a rapid change in field strength near a sharp point. Although this type of discharge usually falls into the 'unwanted' discharge territory, they can be tolerated. However, like partial discharges in voids, corona discharges generate electrical noise, which may be a problem in some situations. If they occur in air, they generate ozone, which can be harmful to health if there is inadequate ventilation. Corona discharges are used for some engineering applications, such as printing and filtration.

3.11.2 Mechanism

Figure 12 shows an electrode arrangement that will create a corona discharge before the electrodes flash over. Geometric field enhancement occurs near the tip of the triangle. The point causes the equipotential lines to bunch up, making the very localized high-field region shown in Fig. 12(b).

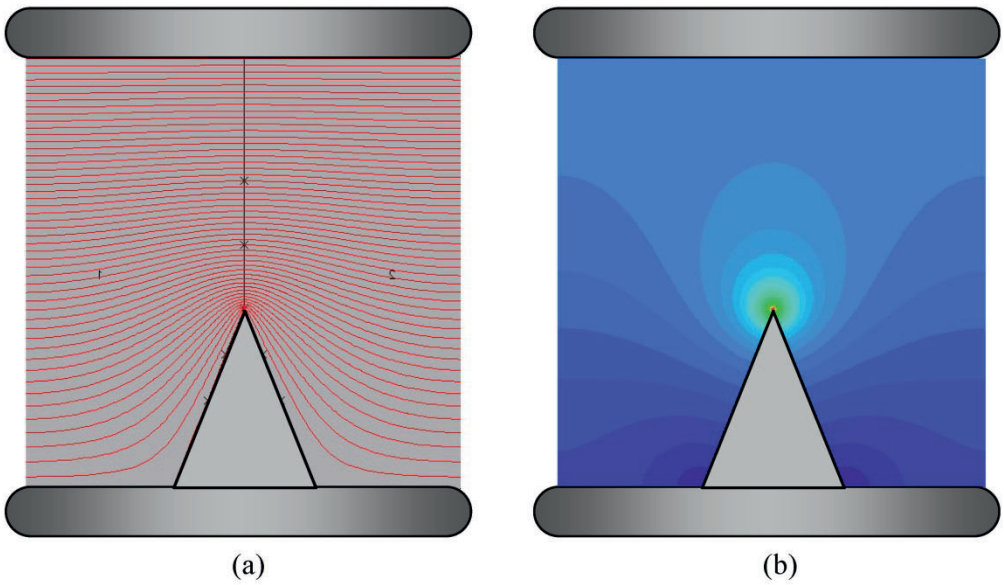


Fig. 12: (a) Equipotential lines and (b) electric field strength around a point

When the applied voltage is high enough, discharges start in the high-field region at the tip and propagate into a decreasing field. The field strength drops off so quickly that the discharge can only propagate a short distance before the field is not strong enough to accelerate electrons to the ionization energy of the gas. The field in the rest of the gap is not strong enough for the discharges to develop either. The electron avalanches stop a certain distance from the tip.

3.11.3 Importance of polarity

In highly divergent fields like that shown in Fig. 12, the polarity of the applied field plays a critical role in the properties of the discharge. This is obvious when looking at the direction of the avalanches shown in Fig. 13.

For a positive point, shown in Fig. 13(a), the electron avalanches develop into an increasing field. Any avalanche that is initiated is guaranteed to develop until it reaches the positive point.

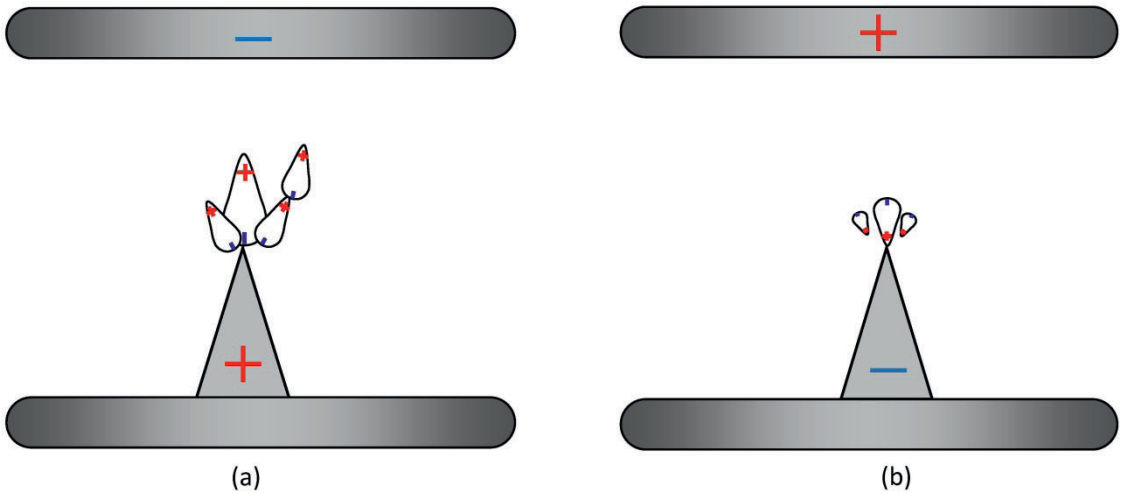


Fig. 13: Electron avalanches for (a) positive and (b) negative point–plane discharges for the same applied voltage.

For a negative point, shown in Fig. 13(b), the electron avalanches develop into a decreasing field. Avalanches will stop propagating when the field strength drops below the value required to accelerate the electrons to the ionization energy of the gas.

The asymmetry of the field means that discharges produced from a positive point propagate much further into the gap than those from a negative point at the same voltage. The field enhancement caused by the charge separation in the initial avalanches causes secondary avalanches. Positive point discharges shown in Fig. 13(a) are more likely to cause secondary avalanches because the size of the initial avalanches is much larger.

3.11.4 Corona inception and extinction voltage

The voltage at which the corona activity starts is referred to as the inception voltage. The field asymmetry in a point–plane gap also means that the voltage at which the corona discharge starts is different for each polarity. Once started, the corona discharge will continue until the applied voltage is reduced. Avalanches form on a nanosecond timescale, and new avalanches are continuously developing near the point.

If the applied voltage is reduced, the corona discharge eventually stops. The voltage at which this happens is referred to as the corona extinction voltage. It is always lower than the inception voltage because, once started, there are always lots of free electrons present to produce more avalanches. Extra free electrons are also created for a negative point corona from electron emission caused by ion bombardment.

3.12 Importance of the power supply

The power source attached to the electrodes will have a large effect on the type of discharge produced. If it is an a.c. supply, the voltage applied to the electrodes will change polarity each cycle, which will

cause the discharge to start and stop twice each cycle. If it is a pulsed supply, there may be time for only a few avalanches before the voltage is switched off. The amount of energy stored in the supply will limit how much current flows before the voltage collapses and the discharge stops. If the power source has a large capacitance (or large inductance), then there could be enough stored energy to cause catastrophic damage to the electrodes and insulators. If the power supply is capable of sustaining an 'unwanted' discharge, it is vital that discharge detection systems can effectively inhibit the power supply output before serious damage occurs by arcing (section 3.14).

All the discharges discussed so far fall into the category of 'unwanted' avalanche discharges. Looking back to the landscape of different discharges shown in Fig. 6, the regimes for 'wanted' discharges occur after the point of avalanche breakdown.

The current–voltage characteristics of discharges are not ohmic. In fact, the arc discharge regime shown in Fig. 5 (discussed in section 3.14) actually has a negative resistance: the volts drop as the current increases.

For 'wanted' discharges the supply must be able to deliver power reliably to the discharge and provide stable control of the current flowing. The current and voltage of a discharge will be where the power supply load curve intersects the characteristic shown in Fig. 6. The gradient of the discharge characteristic at the intersection point determines whether the discharge is stable or not. A resistor is often put in series with the discharge to allow stable control of the current.

3.13 Glow discharge

3.13.1 Introduction

Glow discharges are used in many applications: fluorescent tubes, neon tubes, plasma televisions, for surface treatment and, of course, in ion sources. The glow discharge is so called because it emits a significant amount of light. Most of the photons that make up this light are produced when atoms that have had their orbital electrons excited by electron bombardment relax back to their ground states. Photons are produced in any event that needs to release energy, for example when ions recombine with the free electrons and when vibrationally excited molecules relax.

3.13.2 Striking a glow

The process of starting a glow discharge is referred to as striking. The gas between the electrodes must be ionized. This is usually achieved by simply applying a voltage high enough to cause breakdown. The runaway avalanche breakdown process causes the voltage needed to sustain a discharge to collapse to some tens of volts, as shown in Fig. 6. This collapse in voltage signifies that the discharge has entered the glow discharge regime. The gas between the electrodes has become highly conductive plasma.

When struck, a glow discharge is self-sustaining because positive ions are accelerated to the cathode, and when they impact on it, they produce more electrons. This is one of the secondary emission processes, γ_{ion}, discussed in section 3.5.3. The self-sustaining nature of glow discharges is why there is a hysteresis in the current versus voltage curve at the glow-to-dark discharge transition shown in Fig. 6.

3.13.3 Debye length

Named after the Dutch scientist Peter Debye, the Debye length λ_D is the distance over which the free electrons redistribute themselves to screen out electric fields in plasma. This screening process occurs because the light mobile electrons are repelled from each other while being pulled by neighbouring heavy low-mobility positive ions. Thus the electrons will always distribute themselves between the ions. The electron electric fields counteract the fields of the ions, creating a screening effect. The

Debye length not only limits the influential range that particles' electric fields have on each other but also limits how far electric fields produced by voltages applied to electrodes can penetrate into the plasma. The Debye length effect is what makes the plasma quasi-neutral over long distances.

The higher the electron density, the more effective the screening, and thus the shorter this screening (Debye) length will be. The Debye length is given by:

$$\lambda_{\mathrm{D}} = \sqrt{\frac{\epsilon_0 k T_{\mathrm{e}}}{n_{\mathrm{e}} q_{\mathrm{e}}{}^2}} \tag{34}$$

where

λ_{D} is the Debye length (of order 0.1–1 mm for ion source plasmas),

ϵ_0 is the permittivity of free space,

k is the Boltzmann constant,

q_{e} is the charge of an electron,

T_{e} is the temperatures of the electrons,

n_{e} is the density of electrons.

3.13.4 Cathode plasma sheath

The screening effect of the plasma creates a phenomenon called the cathode plasma sheath around the cathode electrode. The plasma sheath is also called the Debye sheath. The sheath has a greater density of positive ions, and hence an overall excess positive charge. It balances an opposite negative charge on the cathode with which it is in contact. The plasma sheath is several Debye lengths thick.

The majority of the voltage between the anode and the cathode is dropped across this thin cathode plasma sheath. The bulk of plasma is almost at the anode potential.

3.13.5 Towards arc transition

The current in a glow discharge can be increased over several orders of magnitude with very little increase in discharge voltage. The plasma distributes itself around the cathode surface as the current increases. Eventually the current reaches a point where the cathode surface is completely covered with plasma and the only way to increase the current further is to increase the current density at the cathode. This causes a larger voltage drop near the cathode and the plasma voltage rises as shown in Fig. 6.

3.14 Arc discharge

3.14.1 Introduction

Arcs are used in welding, cutting, machining, furnaces and ion sources. Unwanted arc discharges can be very destructive, causing significant damage to electrodes and insulators by sputtering and melting from the large currents that flow. They can also destroy power supplies because of their negative resistance characteristic shown in Fig. 6. More current flows as the power supply volts drop. Without adequate protection, the components in the power supply output stage will fail due to overcurrent. The current in an arc is only limited by what the power supply can deliver.

3.14.2 Glow to arc transition

If the power supply is capable of driving an arc, then arc discharges can occur directly after high-voltage breakdown. The current increases so rapidly that there is no time for a glow discharge to form.

The current density at the cathode causes heating and eventually the cathode surface reaches a temperature where it starts to emit electrons thermionically. This is technically when the discharge moves into the arc regime with a negative current–voltage characteristic as shown in Fig. 6.

The transition between glow and arc can be unstable because of the rapid increase in emitted electrons with temperature and the dynamic surface processes.

Magnetron and Penning ion sources that operate in the arc discharge regime enhance the electron emission from the cathode with the addition of work-function-lowering caesium. Vacuum arc ion sources rely on the heating and vaporizing of the cathode material to produce ions.

3.14.3 Thermal arcs

As the current in an arc increases, the plasma becomes almost completely ionized (there are few neutral particles left). Eventually the current density in the plasma reaches a point where the ions have the same average velocity as the electrons: they have reached thermal equilibrium. The discharge enters the thermal arc regime where the discharge voltage rises as the current increases, as shown in Fig. 6.

3.15 Factors affecting properties of discharges

3.15.1 Statistical variability

Even with identical conditions, the same electrode gap will break down at different voltages each time the voltage is applied. This is because of the statistical nature of high-voltage breakdown: no two sparks are ever the same. The initial location of the free electrons will be different every time.

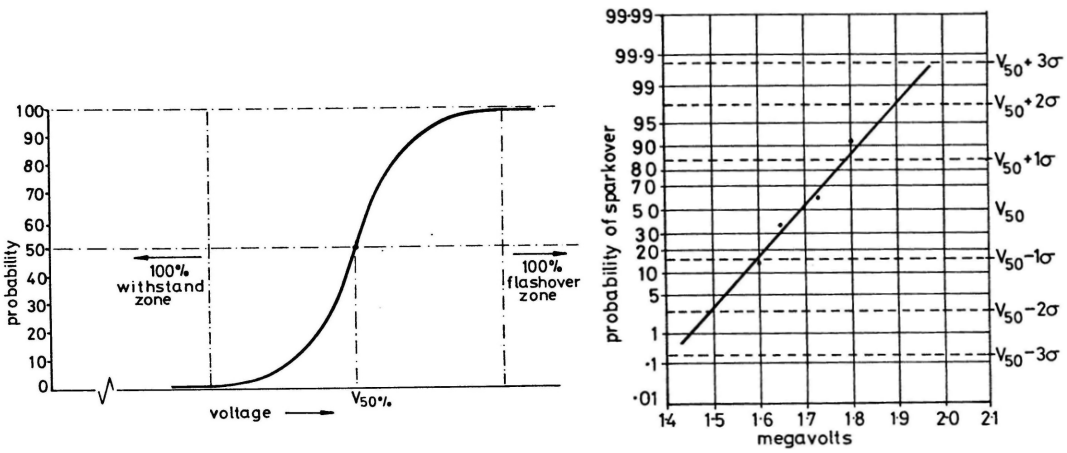


Fig. 14: The statistical variability of high-voltage breakdown

To help to quantify this, the concept of $V_{50\%}$ is introduced. This is the voltage at which the system will break down 50% of the time the voltage is applied. Figure 14 illustrates this point. It also demonstrates that operating at a voltage just below $V_{50\%}$ does not guarantee that breakdowns will not occur: they are just less likely. For applications where high-voltage breakdowns cannot be tolerated, significant safety margins must be introduced.

The value of $V_{50\%}$ can only be obtained by repeated experiment, but this may not be possible in certain circumstances. For example, an insulator might be permanently damaged after a single breakdown. High-voltage conditioning of vacuum systems is essentially a process of carefully finding the $V_{50\%}$ then slowly increasing the voltage until $V_{50\%}$ can no longer be increased.

3.15.2 Environmental conditions

Higher temperatures and lower pressures lead to lower flashover voltages. A correction factor for $V_{50\%}$ in air can be found from

$$V_{50\%} \text{ correction factor} = \frac{0.386 \times P}{273 + t} \tag{35}$$

where P is in mmHg and t is in degrees Celsius.

The humidity of air can also affect the breakdown voltage. Higher humidity leads to lower breakdown and corona inception voltages.

3.15.3 Polarity and geometry

The influence that voltage polarity has on discharges has already been discussed in section 3.11. If there is perfect mirror symmetry between the anode and cathode, the polarity of the applied voltage does not matter. As soon as the mirror symmetry is broken, the positive and negative breakdown voltages of the geometry become different. The difference in breakdown voltage is largest when the difference between the anode and cathode geometry is largest. The most extreme case is the point–plane gap shown in Fig. 12. In this case a negative point breaks down at a much higher voltage than a positive point.

3.15.4 Type of applied voltage

The rise-time and duration of the applied voltage will affect at what voltage the gap breaks down. Shorter applied pulses, or higher-frequency a.c., have a higher breakdown voltage than d.c. or slow pulses. If the applied voltage pulse is short enough, the discharge does not have time to build up before the voltage drops.

3.15.5 Magnetic fields

Magnetic fields significantly affect electrical discharges. Charged particles will rotate around magnetic field lines: this means that they tend to travel along magnetic field lines, by spiralling along them. This can be exploited to confine plasma in an ion source.

Electrons are much lighter than ions, so magnetic fields have a much greater effect on electrons than they do on ions. Stray magnetic fields can cause unwanted breakdowns by channelling electrons along the field lines.

3.15.6 Charges

Discharges by their very nature contain large numbers of charge carriers. The space charge of the free charge carriers created by a discharge can have a significant effect on the overall electric field. Depending on the polarity and timeframes of the applied voltages, this can either enhance or reduce the applied field. The same effect can happen when surface charge is deposited on insulator surfaces. Residual charge left by a previous discharge can alter the behaviour of the following discharge. The overall effect that charges have on discharges is complicated and situation-dependent. Residual charges are a cause of some of the statistical variability of discharges.

3.15.7 Contamination and lost beams

The cleanliness of the electrode and insulator surfaces can reduce the breakdown voltage of a system. Pollution, dirt, grease and water can all cause tracking (section 3.9.2) to occur. Figure 15 shows a 35 kV insulator that has tracked by water flowing down its surface. Stray particle beams impacting on

insulators or electrodes can trigger breakdowns. In vacuum, X-rays from bremsstrahlung created when electrons hit electrode surfaces can cause insulators to break down.

Fig. 15: Tracking caused by water leak across a 35 kV insulator

4 High-voltage design and technology

4.1 Introduction

This section covers some of the design principles required to successfully design and build high-voltage systems. It also discusses the commercial high-voltage technology required to bring complete ion source systems together.

4.2 Electrodes

4.2.1 Electrode design

Electrodes are used to shape the electric field or as a source/sink for current. Field shaping electrodes are used in accelerating, deflecting and focusing beams. Current source/sink electrodes are used in plasma generation and beam dumps. Some electrodes such as extraction electrodes can be both field shaping and current sinking.

Electrode design obviously depends on application. For field shaping electrodes, the shape of the electrode is critical. For current source/sink electrodes, the sputtering resistance and current-carrying capability is critical.

If the electrode's aim is to shape fields while avoiding unwanted discharges, the key to high-voltage electrode design is smoothness, both microscopically and macroscopically. Electrodes should be physically smooth and have gently curving features with large radii.

Using modern finite element techniques, the electric fields produced by electrodes can be accurately calculated (see section 2.6). This allows designs to be tested and optimized before manufacture. Extraction and beam acceleration electrodes can be modelled and simulated beams tracked through them.

4.2.2 Corona shields

Corona shields are a type of field shaping electrode used to shield sharp points that would otherwise suffer from corona. The sharp points are usually unavoidable components like bolts or other fixtures and fittings. A corona shield increases the radius of these points by literally covering them up with rounded electrodes that are electrically connected so they are effectively at the same potential. Other names for corona shields are guard rings and stress rings. Figure 16 shows two examples of how to shield a point with two different types of corona shield.

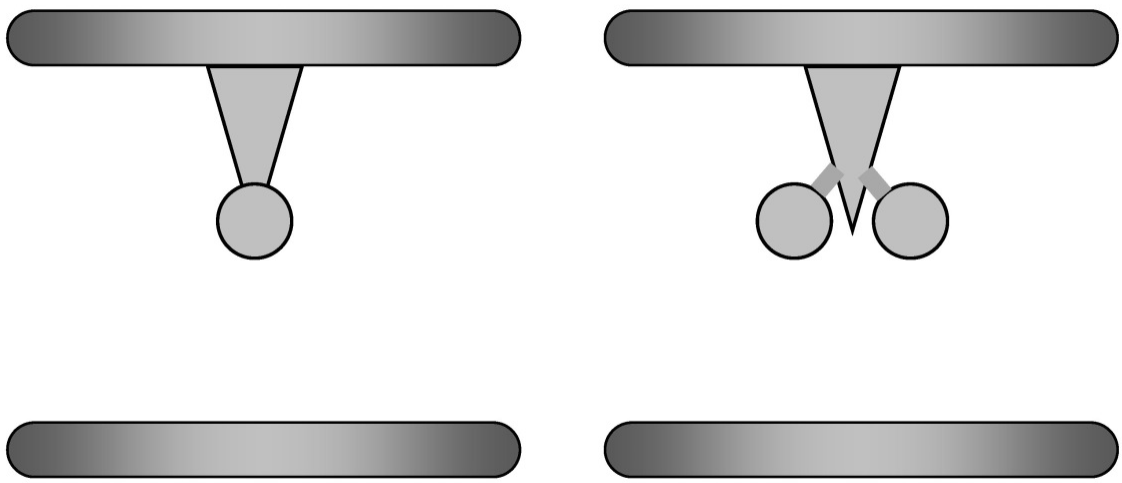


Fig. 16: Two examples of shielding a point with corona shields

4.2.3 Sputtering

Sputtering is where ions bombard electrode surfaces, knocking atoms off the surface. This process causes erosion of the electrode surface. Electrodes in plasmas suffer the most from sputtering. Sputtering eventually leads to failure of the plasma electrode system. The process also slowly deforms extraction electrodes.

4.2.4 Electrode materials

For cathodes in high-sputter environments, molybdenum or tungsten are often used. For field shaping electrodes, any conductive material can be used, so material choice is application-dependent.

For higher fields, electrodes must be smooth and highly polished to minimize surface irregularities. Polished 316L stainless steel is commonly used for most applications.

For operation at very high field strengths in vacuum >10 MV m^{-1} d.c., titanium or its alloys give lower spark rates, especially when used as the cathode electrode material. Titanium–aluminium–molybdenum alloy and pure titanium produce approximately equal spark rates. Titanium–vanadium or titanium–manganese alloys have the lowest spark rates of all [5].

4.3 Insulators

4.3.1 System design

Insulator design is based on the fact that an insulator's surface is the weakest part of a high-voltage system. Good design must: keep the electric field strength along its surface as low as possible; avoid field enhancement caused by triple junctions; and keep the surface clean.

Breakdowns on insulator surfaces cause permanent damage, so good HV system design must ensure that breakdowns are unlikely to occur across an insulator surface. The highest field strengths should be across electrode-to-electrode gaps. The field strengths along insulator surfaces should be kept as low as possible by making the insulators as long as is practically possible. Triple junctions that could initiate a discharge should be screened.

4.3.2 Triple junction effect

The triple junction effect is a localized geometric field enhancement caused when insulators of two different permittivities meet at an electrode. Triple junctions always exist at some scale because it is impossible to have a completely flat surface (see Fig. 17).

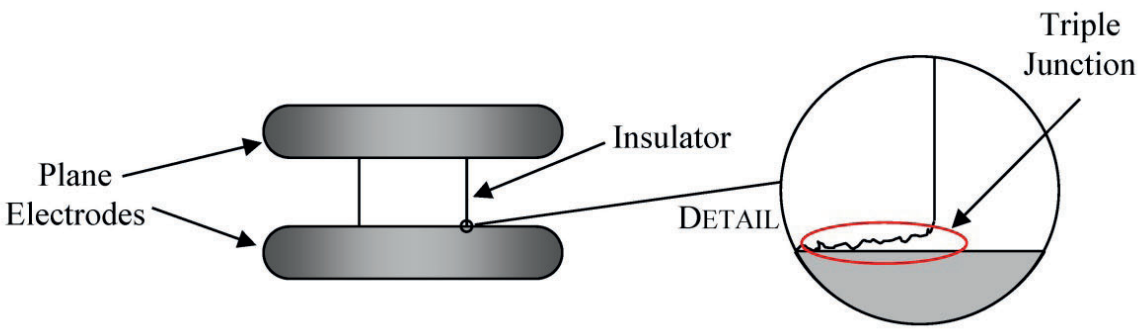


Fig. 17: Triple junctions occur where insulators and electrodes meet

The triple junctions cause field enhancement because free space (air or vacuum) has a lower permittivity than the insulator. The insulator pushes the equipotential field lines into the free space, increasing the field strength in that region. An alternative capacitive voltage divider explanation is given in section 3.10.

Triple junctions are a problem because the field enhanced region is in the perfect position to initiate discharges, which can then propagate across the entire insulator surface. The field enhancement effect actually gets worse the smaller the triple junction becomes. This is illustrated in Fig. 18, which shows how the field in the free space region increases as the triple junction gap is reduced from 4 mm to 0.5 mm in a 500 kV m^{-1} ambient field strength with an insulator relative permittivity of 2.2. At 0.5 mm the field in the free space region has almost doubled to 1000 kV m^{-1}.

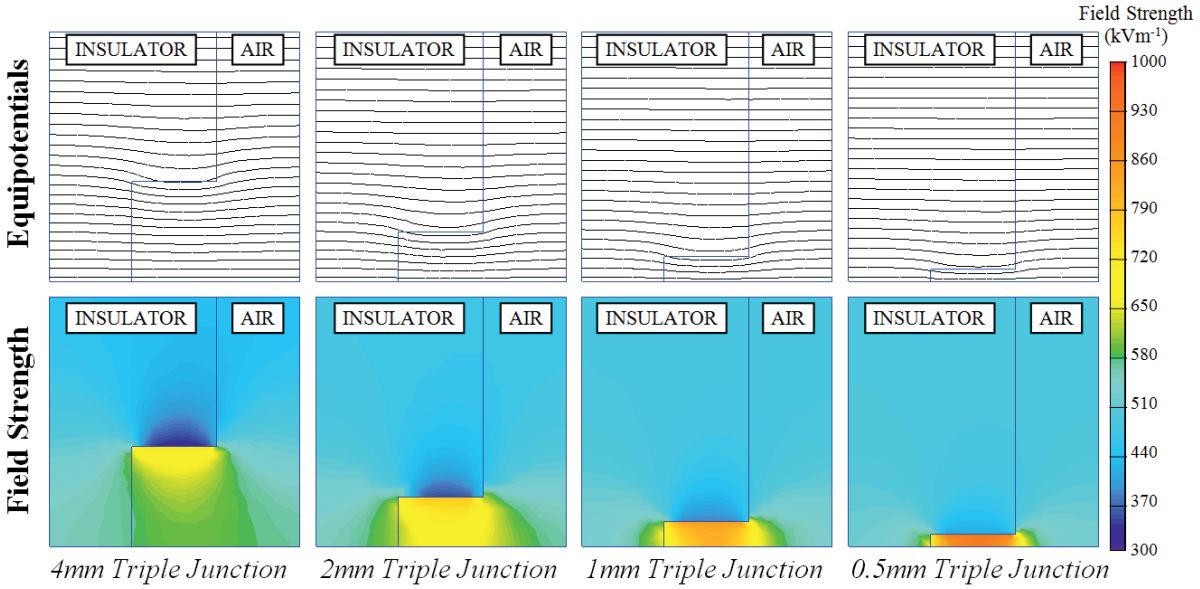


Fig. 18: Triple junction field enhancement increases as the gap become smaller

The permittivity of the insulator amplifies the triple junction effect because the increased relative permittivity pushes more field lines into the free space.

Figure 19 shows how the field in a 1 mm triple gap increases as the relative permittivity of the insulator increases in a 500 kV m^{-1} ambient field.

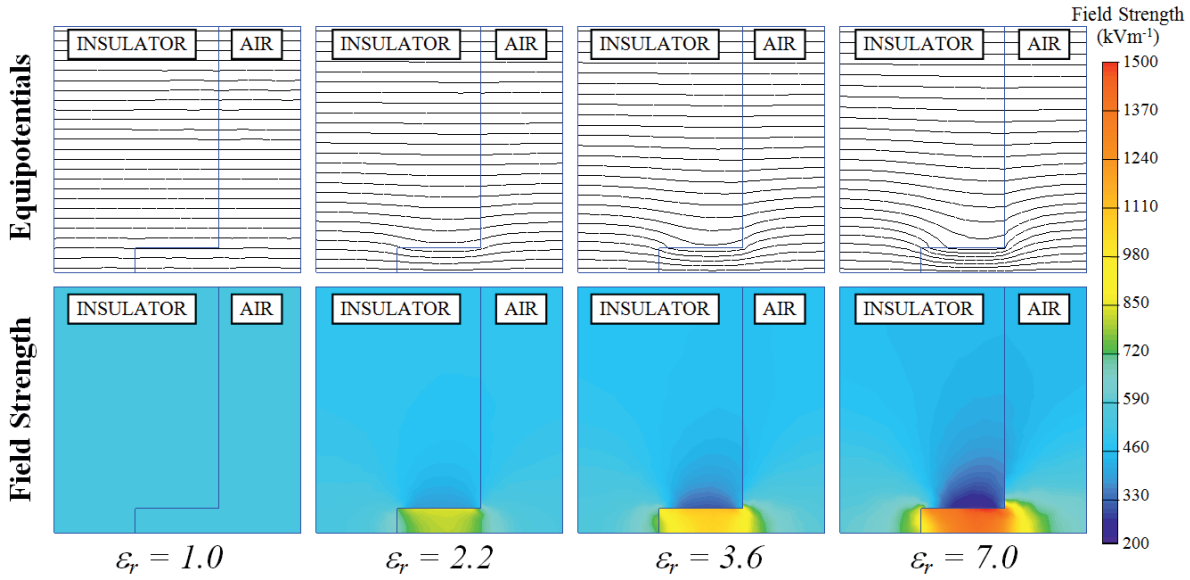


Fig. 19: Triple junction field enhancement increases as the permittivity of the insulator increases

4.3.3 Triple junction shielding

By shaping the electrodes, it is possible to shield any triple junction. Figure 20 shows some of the different possible techniques, which all work by reducing the field in the triple junction region.

Figure 20(a) shows the field in an insulator with a relative permeability of 2.2 and a 1 mm triple junction in an ambient field of 500 kV m^{-1}.

If the insulator cannot be modified, then the best option is to recess the electrode as shown in Fig. 20(b). If neither insulator nor the electrode can be modified, then the best option is to add a guard ring onto the electrode, as shown in Fig. 20(c). Both the guard rings and the recesses need to be a short distance away from the insulator: close enough to shield the triple junction, yet far enough away so the field on the insulator surface is not increased. The precise shape of the guard ring cross-section can be optimised to minimise the field on the insulator surface.

If the insulator can be modified, the best solution is to add an internal guard ring, as shown in Fig. 20(d). This reduces the field in free space and increases it in the insulator material. This is the best engineering solution because the bulk of the insulator has a very high dielectric strength. Also the conditions inside the insulator are much more stable and controllable than outside. Surface contamination and lost beams can cause the surface of an insulator to fail (section 3.15.7). It is also possible to reduce the surface field even further by also adding an external guard ring or recess, as shown in Fig. 20(e).

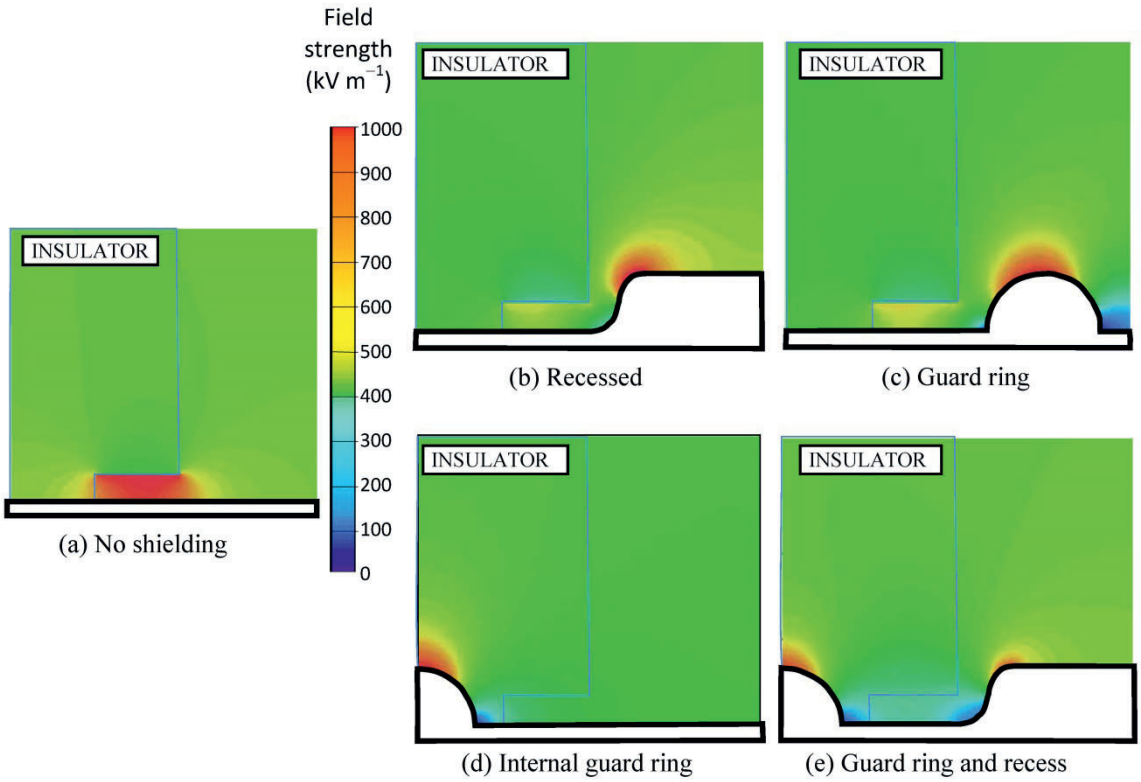


Fig. 20: Different insulator triple junction shielding techniques

4.3.4 Insulator material

There are many types of insulation material. Which insulator to use obviously depends on the application. For example, Al_2O_3 is commonly used in vacuum. AlN is used when high thermal conductivity is required. Macor is used when small complex shapes need to be machined. Porcelain is used in compression. Epoxy resin is used to impregnate and pot. Mica is used for thin high-voltage withstand. Glass is used when visible transparency is required.

Table 1 shows the relative permittivity and dielectric strength of some commonly used insulation materials.

Table 1: The relative permittivity and dielectric strength of some commonly used insulation materials.

Material	Relative permittivity, ε_r	Dielectric strength (kV mm^{-1})
Air	1	3
PTFE	2.2	19.7
Al_2O_3	8–10	13.4
Mica	6	118
Epoxy resin	3.6	20
SF_6	1.002	7.5
Oil	2.5–4	10–15

4.3.5 Insulator surface profile

If the electric fields along a surface are low ($\ll$ 5 kV cm^{-1}), then simple, straight, smooth insulators can be used in many applications. However, for high-field applications, the shape of an insulator's surface profile is a very important part of its high-voltage design.

The ridged or contoured profile that is commonly seen on high-voltage insulators is there to increase the surface path length along which the discharge must propagate to reach the other electrode. The profile should also make the electric field component parallel to the surface drop to zero along parts of the surface, which will inhibit surface propagation.

Another name for the insulator ridges is 'sheds'. This name is more common in insulators designed to work outdoors and in wet conditions. Water flowing down insulator surfaces can cause tracking (see section 3.9.2) in fields much lower than 5 kV cm^{-1}. The field strength that caused the track shown in Fig. 15 was only 1.2 kV cm^{-1}. Sheds work by protecting some of the insulator's surface from the water, keeping it dry like a garden shed keeps its contents dry. In extreme environments, the underside of the shed has additional undulations. The sheds are designed so that water cannot collect on their surfaces. The water flowing off the surfaces has the additional benefit of cleaning off any pollution.

4.3.6 Insulator protection

In ion sources that use caesium or sources that sputter large amounts of conductive metals, it is sometimes necessary to implement insulator shields around the insulator. These are metal shields surrounding, but not touching, the insulator. They are mounted on one electrode, and stop short of the other electrode. They leave a small section of insulator visible, with the majority of it covered. If the environment is particularly harsh, it is sometimes necessary to add a second shield of larger diameter connected to the other electrode, which covers both the insulator and first insulator shield.

When an insulator flashes over, it can be irreparably damaged. To prevent this, extra electrodes can be added that act as a protective spark gap. Called 'arcing horns', they limit the voltage that can be applied to the insulator, ensuring that it does not become damaged.

4.3.7 Insulation coordination

When looking at the overall system design, it is important to make sure that each part of the high-voltage system is capable of working at the voltages required. Starting at the high-voltage power supply, the cable connecting it to the vacuum vessel should be able to withstand the maximum output voltage. The vacuum feedthrough should be able to withstand the maximum output voltage. Any

insulators inside the vessel should be able to withstand the maximum output voltage. It should only be the electrodes themselves that might have a lower flashover voltage.

4.4 Gaseous and liquid insulation

4.4.1 Air

Air is obviously the most common form of gaseous insulation. As a rule of thumb, air at normal ambient conditions requires about 30 kV cm^{-1} to break down, for uniform fields. This number slowly drops for longer gaps and is significantly lower for asymmetrical fields. In fact, a positive point–plane 1 m gap only requires 500 kV to break down.

The operating voltage for a high-voltage system should always be significantly lower (see Fig. 13) than its breakdown voltage if the system is to operate reliably.

4.4.2 Sulphur hexafluoride (SF_6)

SF_6 is the most commonly used insulating gas. It has a dielectric strength twice that of air at one atmosphere pressure because it is a very electronegative gas. This makes it very good at absorbing the free electrons produced in avalanches. It thus inhibits electrical breakdown.

It is non-toxic and non-flammable, making it very safe to work with. It allows lightweight, compact, high-voltage systems to be developed without the need for vacuum pumps. Often used in vessels filled to only 0.5 bar over atmospheric pressure, it can easily deliver dielectric strengths three times that of air. For specialist applications, even higher pressures are used.

The major drawback of SF_6 is that it is the most potent greenhouse gas: 1 kg of SF_6 is equivalent to 23 900 kg of CO_2. It requires gas handling systems and must not be released into the atmosphere. If significant amount of sparking occurs, by-products can be produced that are hazardous to health and can be corrosive. The most toxic gaseous arcing by-product is S_2F_{10}, which has a permitted exposure limit of only 0.01 ppm.

4.4.3 Oil

Oil is the most common liquid insulator. It has a similar dielectric strength to that of SF_6 at 0.5 bar overpressure (i.e., three times that of air). Oil is commonly used in transformers and high-current applications because it can also be used as a coolant.

Oil is much less harmful to the environment, but it still must be disposed of properly, and a back-up containment vessel must be provided to prevent leakage. This is in the form of either double-skinned tanks or bunds.

4.5 Commercially available HV components

4.5.1 Bushings

Bushings are high-voltage feedthroughs designed to pass a high voltage through a barrier. For ion sources, this barrier is often a vacuum vessel, so the bushing must also withstand the pressure difference. The length of the bushing is much more compact on the vacuum side of the bushing than it is on the air side. Bushings are used for semi-permanent high-voltage connections.

4.5.2 Connectors

Connectors are used to connect high-voltage cables between the high-voltage systems and the ion source. Like bushings, they can also be mounted on a vacuum vessel. Connectors are used when the high-voltage connection needs to be regularly removed.

4.5.3 Cables

High-voltage cable is a specialized system in itself and it will be designed to operate at a certain voltage. High-voltage cable is usually coaxial, with a centre high-voltage conductor surrounded by an earthed outer conductor. The key to successful cable use is proper cable termination. High-voltage cables cannot just be stripped back like normal wires. The insulation must be carefully removed, and often other components such a stress cones or stress shields are required to screen the sharp edge created when the outer conductor was cut. Sometimes self-amalgamating high-permittivity tape is used to spread the electric field away from the sharp-edged outer conductor.

High-voltage connectors are designed for specific cables, so they include cable terminating features in their design.

4.5.4 Insulators

Ion sources contain lots of custom-built insulators designed to fit in specific places, but wherever possible commercially available insulators should be used. Commercially available insulators are relatively cheap and, if used according to their specifications, incredibly reliable. A well-designed insulation system is one you should never have to worry about.

4.5.5 Voltage dividers

It is necessary to measure and monitor the high voltages applied in a system. This is usually achieved with voltage dividers. These can be either capacitive or resistive. Some dividers that are designed to work with d.c. and give a flat frequency response are a combination of both.

4.6 High-voltage platforms

4.6.1 Introduction

Ion source technology nearly always requires the biasing of power supplies at high voltages. This is because electric fields are used to extract beams from plasma. To do this, the source plasma must be held at a high voltage, and the extracted beam accelerated to ground potential. To achieve this set-up, the power supply required to generate the plasma must also be biased at a high voltage. Usually the power supply needs some sort of control system, which also must be held at a high voltage, also called ‘at volts’. In addition the ion source usually needs some sort of monitoring ‘at volts’.

All this is achieved by electrically floating all the equipment on a high-voltage platform, powering them with isolated power supplies, and controlling them with isolated signals.

4.6.2 System design

There are two possible system design philosophies for high-voltage platforms.

High-voltage ‘room’

This involves physically putting all the equipment on a metal platform supported by insulators. The connections to the ion source are made with wires capable of carrying the required currents and suitable vacuum feedthroughs. The ion source is mounted on a high-voltage insulator. The platform, connecting wires and ion source are all surrounded by a cage, room or partition walls to prevent people from touching equipment biased at volts.

This design philosophy takes up a lot of room. If electrical noise or X-rays are produced, then the entire ‘room’ must be adequately screened and shielded. The main advantage of a high-voltage ‘room’ is that it allows plenty of space for expansion. Temporary diagnostics and experimental

systems can easily be accommodated and simply connected to the ion source. Access for maintenance is easy. For a source under development, this is the best design philosophy.

High-voltage 'chassis'

The equipment that needs to run at volts is mounted in a chassis. The chassis is mounted on insulators in an enclosure to prevent people from touching the high-voltage chassis. High-voltage cable is used to take the power and monitoring signals to the ion source. High-voltage feedthroughs then take the electrical connections into the vacuum vessel and onto the ion source.

This design philosophy is very inflexible. There may be no room in the HV chassis for any more equipment. Any new equipment will need new high-voltage cables and feedthroughs to the ion source. Equipment maintenance requires opening two enclosures, and access is often restricted. The main advantage of an HV chassis is compactness. All high voltages are safely enclosed, electrical screening is easy, and X-ray shielding can be implemented without difficulty if required. Personnel can have full access to the area while the source is running. For a mature, fully developed system, this is the best design philosophy.

Hybrid design philosophies are possible. For example, an enclosed HV chassis could be implemented with a caged cable tray connecting to the ion source. This removes the need for HV cabling to carry low-voltage power and monitoring signals. The main question the system designer needs to ask is this: How flexible does the set-up need to be?

4.6.3 Isolated power supplies

Isolated power supplies are used to get power to the equipment on the HV platform. The simplest solution is a 1 : 1 mains isolation transformer. An isolation transformer has a large amount of insulation between its secondary winding and the yoke. This allows the primary winding and the magnetic steel yoke to remain at ground potential, while the electrically isolated secondary can be floated to platform volts. Isolation transformers can be air- or oil-filled, and they can be single- or three-phase.

There is a limit to how much insulation can be wrapped around a transformer core. For most cases this is a few hundred kilovolts. Beyond this voltage, the only way to get power to the high-voltage platform is with a motor alternator set. A motor is mounted at ground potential and mechanically connected via an insulating rod of material to a generator mounted on the high-voltage platform.

The number of moving parts makes motor alternator sets much less reliable than isolation transformers. Whichever system is used, the isolated neutral must be connected to the high-voltage platform for safety. Otherwise, when the platform is earthed down, the platform power supply would not be referenced to anything.

4.6.4 Isolated control signals

There must be a way to control and monitor the equipment on the high-voltage platform. An old-fashioned technique is to use long insulating rods. This has now been almost entirely replaced by signals sent over fibre optics. There are many commercially available control systems that can communicate over fibre.

4.6.5 Other isolated systems

It is sometimes necessary to connect a water chiller at ground potential to the ion source on the high-voltage platform. To do this, long insulating hoses are used. The water circuit must be kept deionized

with an ion exchange column. For safety it is a good idea also to have a conductivity meter, which provides an interlock signal to the high-voltage power supply.

Gases, including hydrogen can be delivered to the high-voltage platform with simple insulating tubes.

Radio-frequency (RF) power can be fed to the platform through specially designed waveguides with insulating breaks.

4.7 High-voltage power supplies

4.7.1 Power supply technologies

There are many different ways to generate high voltages:

- Semiconductors – thyristor, insulated gate bipolar transistor (IGBT), gate turn-off (GTO) thyristor
- Tube – tetrode
- Pulse-forming network
- Cascade rectifier (Greinacher, Cockcroft–Walton multiplier)
- Van de Graaff generator, pelletron
- Linear (usually front end only)
- Switched mode transformer – HV diode and capacitor

4.7.2 High-voltage power supply manufacturers

There are many manufacturers of high-voltage power supplies. Standard designs are available or custom-built power supplies can be procured. Tight specification is essential. Ion sources are an incredibly difficult load for a power supply to deliver power to. The power supply must be able to cope with breakdowns – this must be specified.

If the resource, ability and skills are available, it is possible build high-voltage power supplies ‘in house’.

4.7.3 Insulation test equipment

High-voltage systems can be tested with special high-voltage power supplies that only allow very small currents to flow in the event of a breakdown. This allows systems to be tested with a reduced risk of permanently damaging any solid insulation.

5 Safety

5.1 Introduction

There are many hazards that must be considered when working with ion sources: high temperatures, strong magnetic fields, dangerous materials (e.g. caesium) and heavy equipment. Before any work starts, thorough risk assessments should be made. This section covers the safety hazards involved when working with high voltages.

5.2 Electric shocks

The main hazard from high voltage is electric shock. Small electric shocks can startle people, potentially causing them to fall or injure themselves by hitting or dropping something. We can sense small electric currents as low as 5 mA (either a.c. or d.c.) as a mild tingling sensation. Higher currents can cause the muscles to tense: 60 mA (either a.c. or d.c.) flowing across the chest cavity is enough to fibrillate the heart, potentially causing heart attack and death. Therefore, the most dangerous situation is holding an earthed handrail and touching a live conductor. A saying goes: 'An old electrician always has one hand in his pocket.'

Dry hand-to-hand resistance can be as high as 100 kΩ depending on how thick the skin is, but this can drop to 1 kΩ for wet or broken skin. This means that voltages up to 60 V are safe. The problem with high voltages is the breakdown voltage of the skin itself. The outer layer of the skin, the stratum corneum, breaks down at about 450–600 V, leaving only about 500 Ω hand-to-hand resistance.

Another problem with high voltages is the amount of energy stored by capacitances at high voltages. Energy stored increases with the square of the voltage. Consider a 1 μF capacitor charged to 30 kV: the energy stored is $\frac{1}{2}CV^2 = 0.5 \times 1\ \mu\text{F} \times 30\ \text{kV} = 450\ \text{J}$. This is a large amount of energy. Medical defibrillators only use a few hundred joules, so this is more than enough to stop the heart.

5.3 Earthing systems

High-voltage systems are made safe by connecting all live electrodes to ground. This is achieved using multiple earthing systems: interlocked earthing arms that are automatically applied; earth switches that can be locked in place with an interlock key; and/or earth sticks that are applied by the operator.

5.4 Interlock systems

Interlock systems are in place to protect both personnel and the equipment. They are a combination of mechanical locks, switches and relays. Interlocks force personnel to operate the system in a specific order. Operation is prevented until all necessary actions have been carried out. Access is prevented until the system has been made safe.

5.5 Other hazards related to high voltage

X-rays can be produced by bremsstrahlung. Any electrons produced in a vacuum vessel can be accelerated by the applied high voltage and generate X-rays when they hit the anode. X-rays can be produced with voltages as low as 20 kV. In fact, bremsstrahlung does not have a low-voltage cut off – it is just that, below 20 kV, the configuration is usually shielded by the vacuum vessel walls. At higher voltages, shielding may become necessary, so X-ray detectors should be used during system commissioning above 20 kV.

Electric sparks can ignite hydrogen and cause an explosion. European ATEX directives [1] must be complied with.

High-voltage capacitors that are not being used should have their terminals shorted together. This is because the dielectric inside can have some 'memory' of the previous energization. Previously used high-voltage capacitors just sitting on a shelf can charge up enough to give an electric shock. The same is true for long lengths of high-voltage cable, which can have a significant capacitance.

5.6 HV safety design rules

The most fallible part of any system is the person operating it. High-voltage safety systems must be designed to make them idiot-proof. For regular use it is not acceptable to have safety reliant on an operator correctly following a procedure. Complacency can be a problem; the operator might have

started and shut down the system thousands of times. It is important that the system is designed in such a way that it makes it impossible for an absent-minded operator to hurt themselves or others. These design rules must be followed:

5.6.1 Impossible to accidentally lock someone in the HV area

If the HV area is very small or if the high-voltage chassis (section 4.6.2) design approach is used, then this rule takes care of itself. However, if the HV area is large enough that someone could be hidden behind equipment, then a 'Search' button must be implemented. The interlock system makes it impossible to shut the door to the HV area until the 'Search' button has been pressed. The 'Search' button should be positioned at the far end of the HV area. This forces the operator to go to the far end of the HV area where they will see if anybody is in the area. Simply calling out 'Is there anybody in here?' is not acceptable because someone could be unconscious.

It must be possible to open the door from inside the HV area, even if it has been locked from the outside. This is usually implemented using the same technology as for fire escapes: crash-bars or break-glass latches. Exit signs must be provided.

5.6.2 Ability to shut down the power inside and outside the HV area

In case of emergency it should be possible to shut down the power to the high-voltage system with an emergency shut-off switch. These are usually red push buttons that lock when pushed, requiring a key to reset them. They should be positioned inside and outside the HV area and be clearly labelled.

5.6.3 Impossible to power on the HV without locking the area

The HV power supplies should have some kind of interlock that prevents them being energized until the doors to the HV area have been secured. This can be done with interlock switches or key exchanges; often it is best to use both. It should be impossible to remove the earth connection to the HV platform until the HV area has been secured.

5.6.4 Impossible to enter the HV area without making it safe

The interlock switches and key exchanges should make it impossible to open the door with the high-voltage power supplies on. However, the equipment could still be charged to a high voltage. When the door to the HV area opens, an automatic earthing arm should short the high-voltage platform to ground. If there are no capacitances or stored energy in the system, then an automatic earthing arm may not be necessary. This should be thoroughly risk-assessed.

5.7 Reasonably practicable

Of course, like any health and safety system, the safety design rules covered in the previous section can only be implemented as far as is reasonably practicable. Safety systems should be designed to prevent people unfamiliar with the equipment or absent-minded people from hurting themselves. You can never stop wilful people from bypassing interlocks or climbing over fences. Make sure there are signs and warning messages on all equipment, doors and entrances. That way, if anybody does bypass the safety systems, they will have only themselves to blame.

Acknowledgements

I would like to thank David Findlay and Alan Letchford for their excellent proof-reading skills and suggestions for improvement; and Mike Perkins and Scott Lawrie for their supportive comments.

References

[1] http://en.wikipedia.org/wiki/ATEX_directive.

[2] P.A. Kennedy and G. Kent, *Rev. Sci. Instrum.* **27** (1956) 916.

[3] Poisson Superfish, http://laacg1.lanl.gov/laacg/services/download_sf.phtml.

[4] F. Paschen Wiedemann, *Ann. Phys. Chem.* **37** (1889) 69–96.

[5] J. Huguenin and R. Dubois, Measurements on a high-gradient accelerating tube model: investigation of the properties of titanium electrodes, CERN 65-23 (1965).

Bibliography

R. Arora and W. Mosch, *High Voltage and Electrical Insulation Engineering* (Wiley, Hoboken, NJ, 2011).

T.J. Gallagher and A.J. Pearmain, *High Voltage: Measurement, Testing and Design* (Wiley, Chichester, UK, 1984).

Microwave Discharge Ion Sources

L. Celona
Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali del Sud, Catania, Italy

Abstract
This chapter describes the basic principles, design features and characteristics of microwave discharge ion sources. A suitable source for the production of intense beams for high-power accelerators must satisfy the requirements of high brightness, stability and reliability. The 2.45 GHz off-resonance microwave discharge sources are ideal devices to generate the required beams, as they produce multimilliampere beams of protons, deuterons and singly charged ions. A description of different technical designs will be given, analysing their performance, with particular attention being paid to the quality of the beam, especially in terms of its emittance.

1 Introduction

The production of high-current beams is a key requirement for various applications, and this is expected to increase in coming years, either for industrial applications or for research projects. High-current and high-brightness H^+ beams can be provided by microwave discharge ion sources (MDISs), which present many advantages in terms of compactness, high reliability, ability to operate in continuous-wave (CW) or pulsed mode, reproducibility and low maintenance. Some applications based on intense proton beams [1] are as follows:

- accelerator-driven systems (ADSs) for nuclear waste transmutation and energy production,
- radioactive ion beams,
- intense neutron spallation sources,
- radiation processing, and
- neutrino factory.

The main parameters of the related proton drivers are listed in Table 1, while Table 2 shows a list of projects (operating and under construction) using high-current proton beams or intense H^- sources with low transverse emittance.

Table 1: Proton driver requirements.

Proton driver	Energy (GeV)	Beam power (MW)
ADS: XADS	~ 0.6	~ 5
Ind. burner	~ 1	~ 50
Spall. neutron source (ESS)	1.33	5
Irradiation facility	1	> 10
Neutrino factory (CERN)	2.2	4
RIB: 'one stage'	~ 0.2	~ 0.1
'two stage'	~ 1	$\sim$ 5–10

The optimization of beam formation and transport through the low-energy beam transport (LEBT) plays a fundamental role in the provision of a high-quality beam to the accelerator. This is a common requirement of the projects reported in Table 2, where emittances at the entrance to the radio frequency

Table 2: High-power accelerator requirements.

	p/H$^-$	mA	ms	Hz	Duty factor (%)	π mm mrad
LEDA	p	100	CW	CW	100	0.25
IPHI	p	100	CW	CW	100	0.25
TRASCO	p	30	CW	CW	100	0.2
ESS	p	60/90	2.84	14	4	0.3
SPL	H$^-$	50	1.5	50	7.5	0.2
SNS	H$^-$	50	1	60	6	0.25
JKJ	H$^-$	30	0.5	50	2.5	0.25

quadrupole (RFQ) of the order of 0.20–0.30 π mm mrad are needed, making it essential to design and test the ion source and LEBT as a whole. The major challenge of the accelerator front-end is therefore the preparation of a high-quality beam, with a pulse that is well defined in time and has a small transverse emittance. In the following, a review of the major experiences in the production of intense proton beams are reported (Table 2) together with future perspectives.

2 Microwave ion source for high-intensity proton production

2.1 Historical notes

The history of 2.45 GHz high-current source (HCSs) started about 35 years ago with different source designs proposed by Sakudo [2] and by Ishikawa *et al.* [3], especially for industrial applications. The sources produced remarkable results not only for protons, but also for deuterons and singly charged light ions. A simple concept of the microwave discharge source was based on a non-confining magnetic field higher than the resonance field (i.e. 87.5 mT). Sakudo and his collaborators at the Central Research Laboratory of Hitachi Limited pioneered the development of high-current microwave ion sources for ion implantation [4]. The first Hitachi ion source (see Fig. 1) is composed of a plasma generator, that is, essentially, a section of coaxial waveguide with an axial magnetic field supplied by three solenoids. The 2.45 GHz microwaves are introduced via a water-cooled antenna connected to the inner conductor of a coaxial-to-rectangular waveguide transition. The magnetic induction is varied along the length of the antenna to match the impedance of the plasma-filled chamber to the impedance of the microwave line [5]. The extraction system is a multiaperture triode with 124 apertures 3 mm in diameter distributed over a 50 mm diameter circle. The sources built by Sakudo's group were able to supply 2 mA of As^+ and 15 mA of B^+, and they were successfully adapted to industrial application setups.

The second microwave ion source developed by Sakudo *et al.* was especially designed to generate a slit-shaped ion beam. The plasma chamber illustrated in Fig. 2 is a tapered ridged waveguide with all but the volume between the ridges filled with boron nitride. The 2.45 GHz microwaves are introduced through a dielectric window from a rectangular waveguide. The electric field between the ridges is relatively uniform, ensuring a reasonably constant plasma density over the entire length.

A step forward was made by Ishikawa [6], whose design was very compact (chamber diameter was 50 mm) and the source was able to produce milliampere beams of any species, finding applications not only in ion implantation devices but also for ion beam deposition. The absence of antennas made this equipment more reliable for long-time operations.

2.2 The CRNL ion source

In 1991 a simple and robust design was proposed by Taylor and Mouris at Chalk River National Laboratory (CRNL; see Fig. 3). This source can be considered as the basis of all the different designs proposed

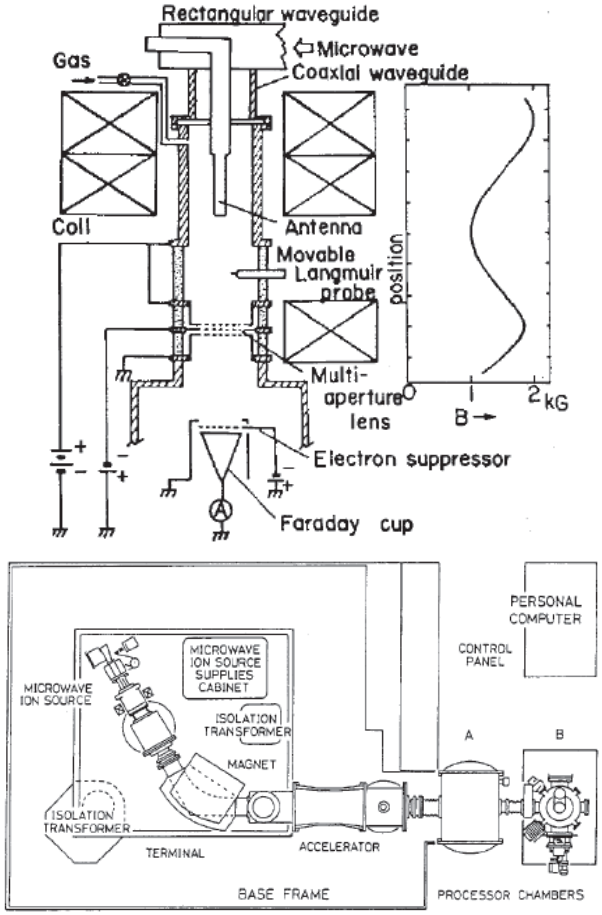


Fig. 1: Hitachi microwave ion source with multiaperture extraction system

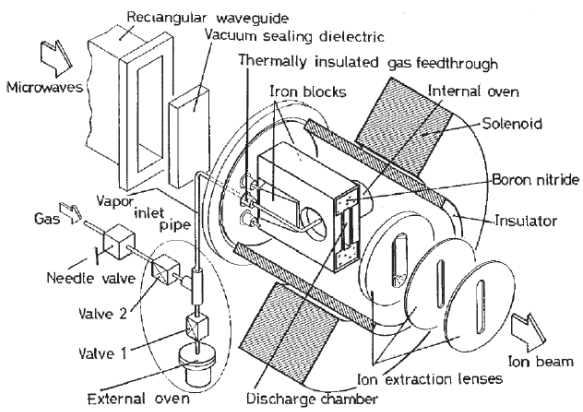


Fig. 2: Hitachi microwave ion source with slit extraction system

in the past 25 years. The main innovation consisted in the use of a matching unit to adapt the waveguide to plasma impedance, which enhanced the plasma density and finally the current density of the extracted beam. Moreover, two separately fed solenoids, approximately placed at the two extremes of the plasma chamber, allowed the magnetic field profile to be adapted. The extraction system was not sophisticated but adequate for high-current beam formation, using the three electrodes in the accelerate–decelerate operation mode.

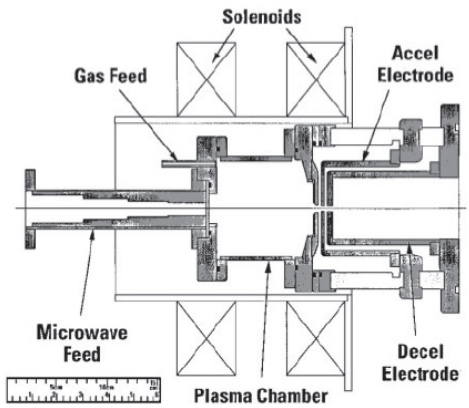


Fig. 3: The CRNL microwave source

The configuration of the Chalk River electron cyclotron resonance (ECR) proton source is illustrated in Fig. 4. The plasma generator is simply a hydrogen-filled chamber, with a ceramic rectangular waveguide window, encircled by two solenoids. The extraction system is a 50 kV multiaperture triode.

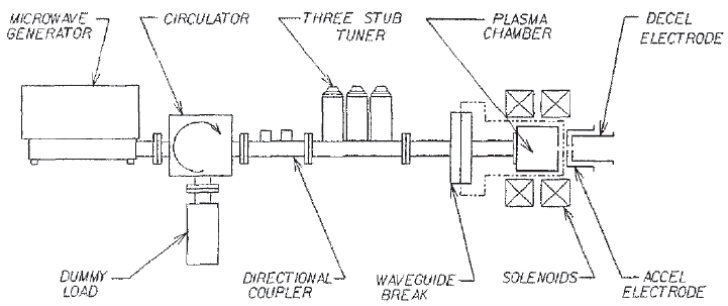


Fig. 4: The CRNL microwave source

The microwave line is isolated from the plasma chamber by a d.c. waveguide break consisting of a Teflon sheet sandwiched between a choke flange and a standard flange. Since the solenoids are also electrically isolated from the plasma chamber, only the plasma chamber itself remains at high voltage. All of the power supplies, d.c. as well as microwave, are at ground potential. The ion source generates beam current densities of up to 350 mA cm^{-2} at a microwave power of 1000 W. The three-stub tuner is adjusted to optimize the impedance match. Figure 5 illustrates the evolution of the beam current with the mass flow, microwave power and magnetic induction on the plasma chamber axis.

2.3 The LEDA injector

This design was further improved by J. Sherman and co-workers at Los Alamos National Laboratory (LANL), who improved the extraction system and the LEBT in order to optimize the beam coupling to the

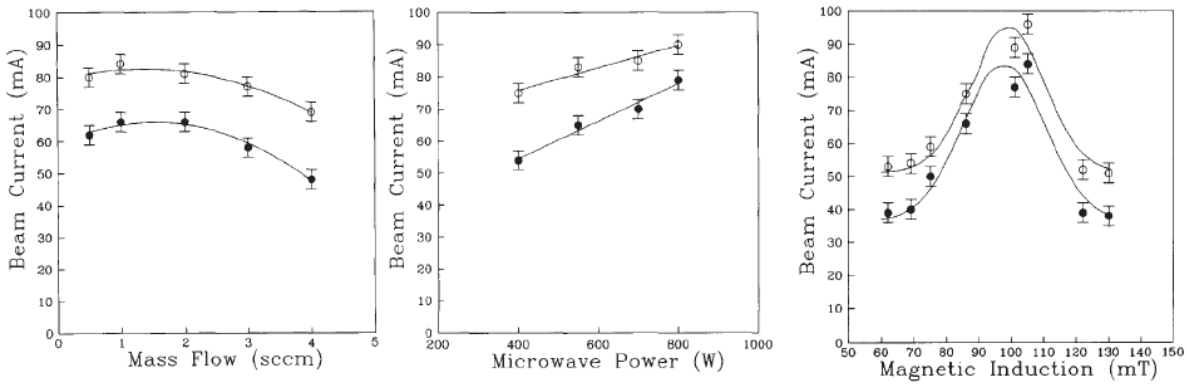


Fig. 5: Beam current as a function of mass flow, microwave power and magnetic induction

RFQ of the LEDA project. Then the CRNL plasma generator was integrated with a 75 kV acceleration structure at Los Alamos (Fig. 6) [7]. Plasma is generated by the interaction of 2.45 GHz microwaves with H_2 gas in the presence of an approximate 875 G axial magnetic field.

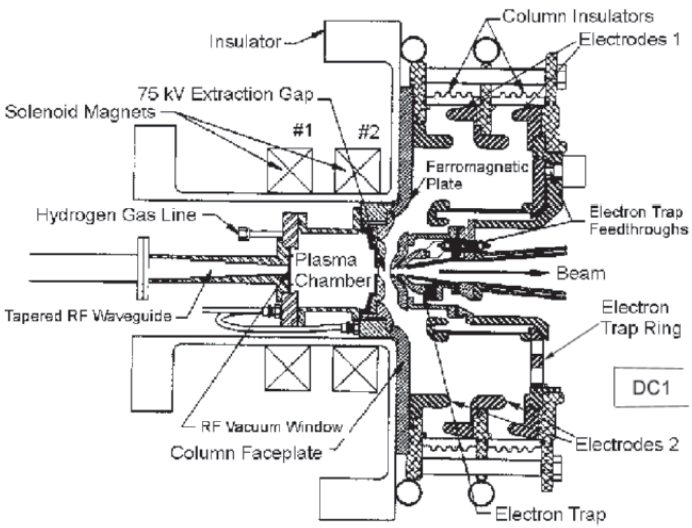


Fig. 6: Line drawing of the 75 keV microwave proton source

Beam measurements were made on a prototype injector shown in Fig. 7. The proton source is mounted on a beam diagnostics and pump box, which is followed by two solenoids. The beam current is characterized by two d.c. current monitors (DC1 and DC2), an a.c. beam current transformer for beam noise measurements, an x and y video profile system for beam position and width measurements, and an emittance measuring unit (EMU), which also serves as the beam dump. The typical operating parameters of the injector are listed in Table 3.

From Fig. 8 we can see that beam current is measured at DC1, DC2 and in the RFQ entrance collimator for 6.7 MeV RFQ; beam focusing accomplished with LEBT is achieved thanks to solenoid magnets S1 and S2; and beam centroid is controlled with steering magnets SM1 and SM2.

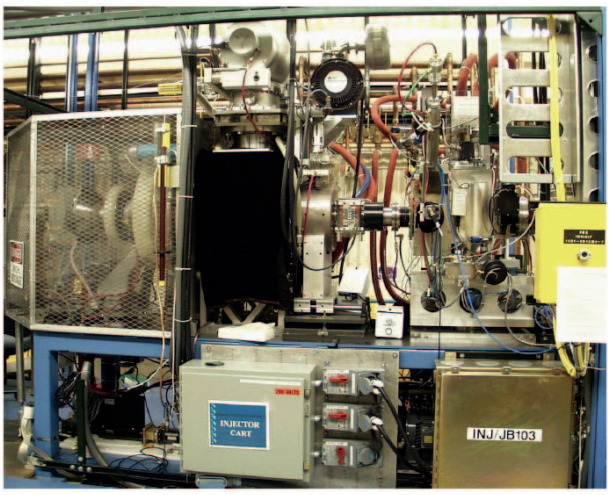


Fig. 7: Injector used for the beam measurements

Table 3: LEDA injector parameters.

Injector parameter	Value
H_2 gas flow	4.1 sccm*
Ion source pressure	2 mTorr
Ion source gas efficiency	24%
Discharge power, at 2.45 GHz	1.2 kW
Beam energy	75 keV
High-voltage power supply current	165 mA
DC1 current	154 mA
DC2 current	120 mA
Proton fraction	90%
Injector emittance, 1 r.m.s. norm.	0.18 π mm mrad

*sccm = standard cubic centimetres per minute.

2.4 The SILHI source

Following the same track, at CEA-Saclay (Commissariat à l'Énergie Atomique, Saclay), the SILHI source obtained large brightness and high reliability in the second half of the 1990s. All the parts of the source, which is shown in Figs. 9 and 10, are robustly engineered, including the beam line elements and diagnostics, adapted to beams exceeding 10 kW, typically 95 kV, 140 mA. A proton/deuteron fraction above 80% and an emittance below 0.2 π mm mrad have been obtained, while the reliability has increased over the years, even reaching 99.9% in a one-week test. The SILHI source (see Fig. 10) has been developed at CEA-Saclay, in the framework of the IPHI project, devoted to the production and acceleration of intense proton beams (currents up to 100 mA). The source design and results are described in detail elsewhere [8–10], and in Table 4 we report the main features.

A collaboration arranged between CEA and the Istituto Nazionale di Fisica Nucleare (INFN) has allowed in recent years a significant increase of beam quality produced by SILHI, reaching the outstanding value of 0.1 π mm mrad for 0.75 mA as described in detail in the next section.

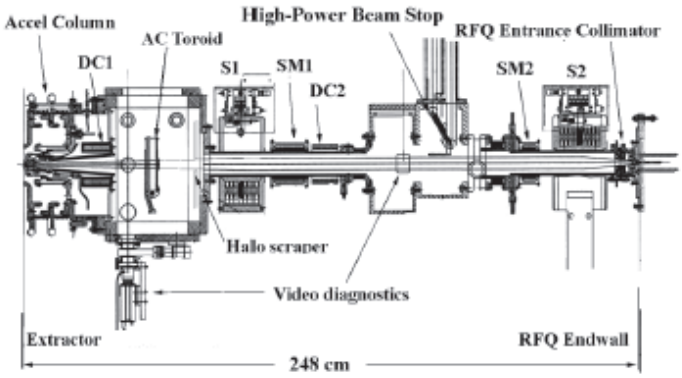


Fig. 8: Injector used for the beam measurements

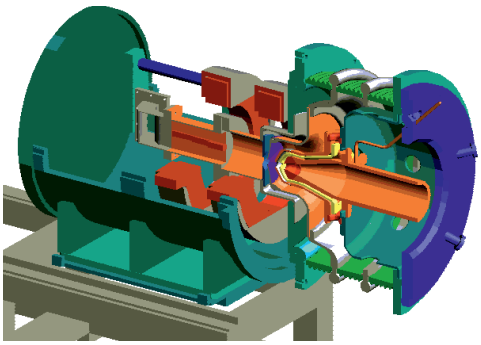


Fig. 9: Render view of the SILHI source operating at CEA

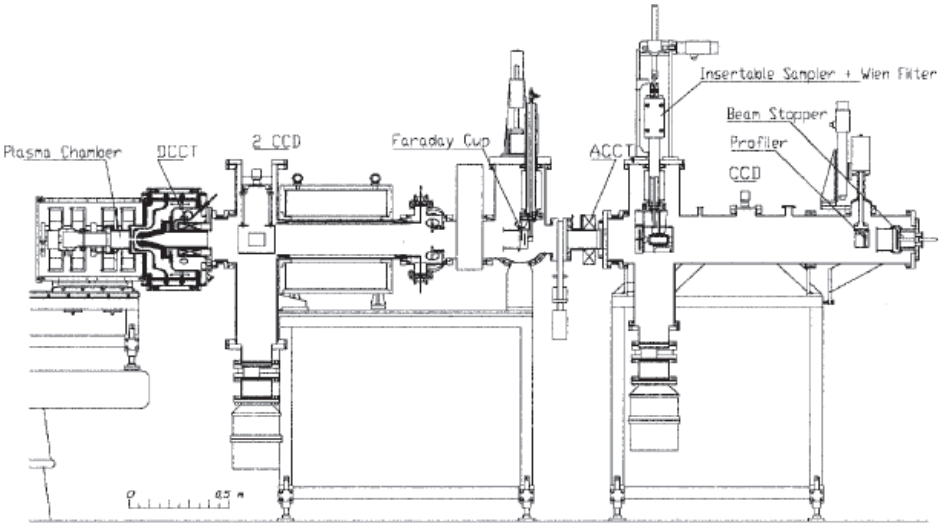


Fig. 10: Layout of the SILHI ion source and the beam line

Table 4: SILHI source typical operating parameters.

Parameter	Value
Total beam current	140 mA
Proton fraction	$\sim 85\%$
Beam density	220 mA cm^{-2}
Beam energy	95 keV
Discharge RF power, at 2.45 GHz	1.2 kW
Beam emittance, at 75 mA	0.1 π mm mrad
Hydrogen mass flow	$\sim$ 2 sccm

2.5 TRASCO Intense Proton Source (TRIPS) at INFN-LNS

The TRIPS ion source [11] has been designed, realized and commissioned at INFN-LNS, within the framework of the TRASCO Project (an R&D programme, the goal of which was the design of an accelerator driving system (ADS) for nuclear waste transmutation). With reference to the three experiments outlined above, a series of innovations were implemented on the TRASCO project. The goal of TRIPS was less demanding in terms of currents than it was for the other projects in Table 2, but there was a very strong requirement for source reliability, with an r.m.s. normalized emittance below 0.2 π mm mrad for an operating voltage of 80 kV (Table 5).

Table 5: TRIPS typical operating parameters.

Parameter	Value
Total beam current	60 mA
Proton fraction	90%
Beam energy	80 keV
Microwave power, at 2.45 GHz	0.3–1 kW
Beam emittance, at 35 mA	0.07 π mm mrad
Gas flow	0.4–0.6 sccm
Duty factor, at 35 mA	99.8%
Reliability, at 35 mA	99.8%

The major innovation consisted in the use of two movable coils permitting the magnetic field profile to be varied. The plasma chamber dimension and the four-step matching transformer were defined to get a uniformly dense plasma at the extraction hole (6 or 8 mm diameter). The microwave coupling with the matching transformer and the automatic tuning unit allowed operation with low values of reflected power (below 5%) and a high electric field on the axis, thus increasing proton fraction and current density, up to 200 mA cm^{-2}. The extraction system was designed in collaboration with Saclay, exploiting the experience already gained at INFN-LNS for the design of extraction geometries from plasma sources. Measurements with the 8 mm hole plasma electrode have given results largely exceeding the TRASCO design current: up to 61 mA were extracted at 80 kV and about 90% of the beam has been transported to the beam stop.

Figure 11 shows the TRIPS source and Fig. 12 shows the experimental set-up. The first section of the low-energy beam transfer line (LEBT) devoted to beam analysis consists of a current transformer (DCCT1), a focusing solenoid, a four-sector ring to measure beam misalignments and inhomogeneities, a second current transformer (DCCT2) and an insulated 10 kW beam stop (BS), which measures the beam current.

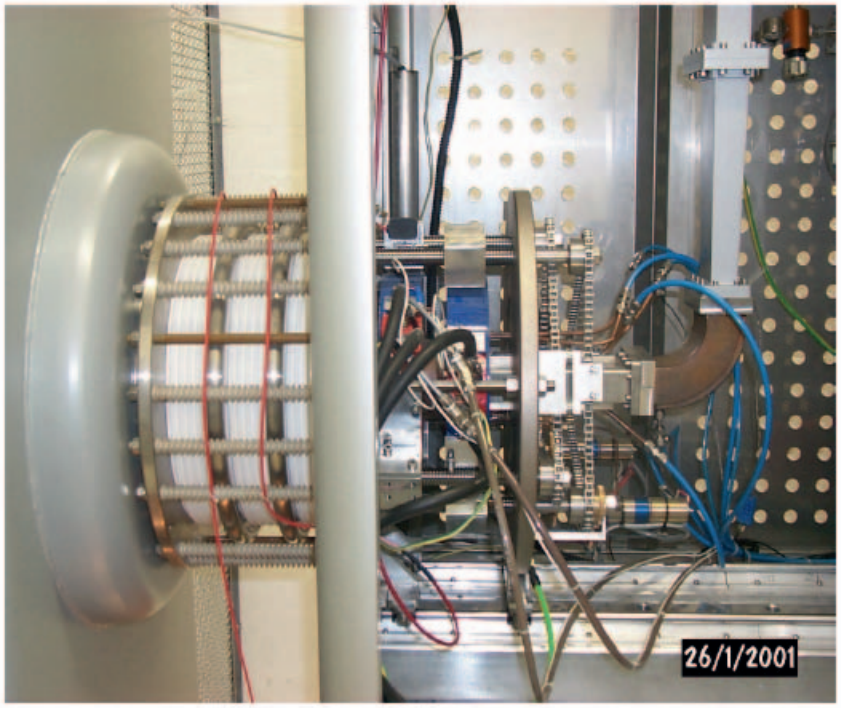


Fig. 11: The TRIPS ion source on the 100 kV platform

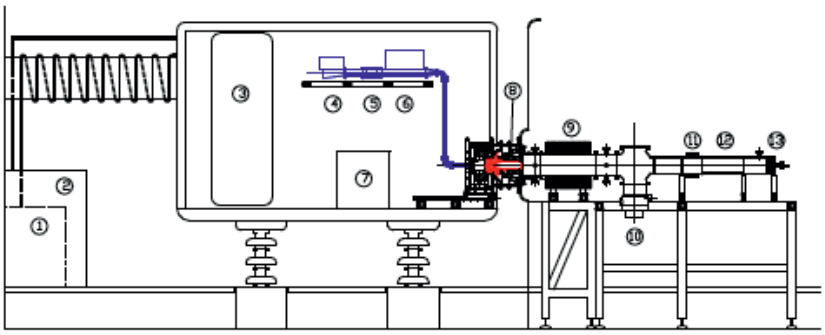


Fig. 12: The experimental set-up: (1) demineralizer; (2) 120 kV insulating transformer; (3) 19 inch rack for the power supplies and for the remote control system; (4) magnetron and circulator; (5) directional coupler; (6) automatic tuning unit; (7) gas box; (8) DCCT1; (9) solenoid; (10) turbomolecular pump; (11) DCCT2; (12) quartz tube; (13) 10 kW beam stop.

Systematic measurements with movable coils have been carried out for different positions and currents, by leaving the other parameters unchanged. From these measurements it can be determined that the source is more sensitive to variations in the extraction coil than in the injection one. The best performance is clearly obtained when the two ECR zones are located exactly on the boron nitride discs at the two ends of the plasma chamber. Figure 13 shows the best obtained profile (the straight line represents the value of the magnetic field corresponding to the resonance at 2.45 GHz).

Systematic emittance measurements have been made with an emittance measurement unit provided by the CEA-Saclay SILHI group, showing the effect of each parameter on the beam emittance; more details are reported in the next section.

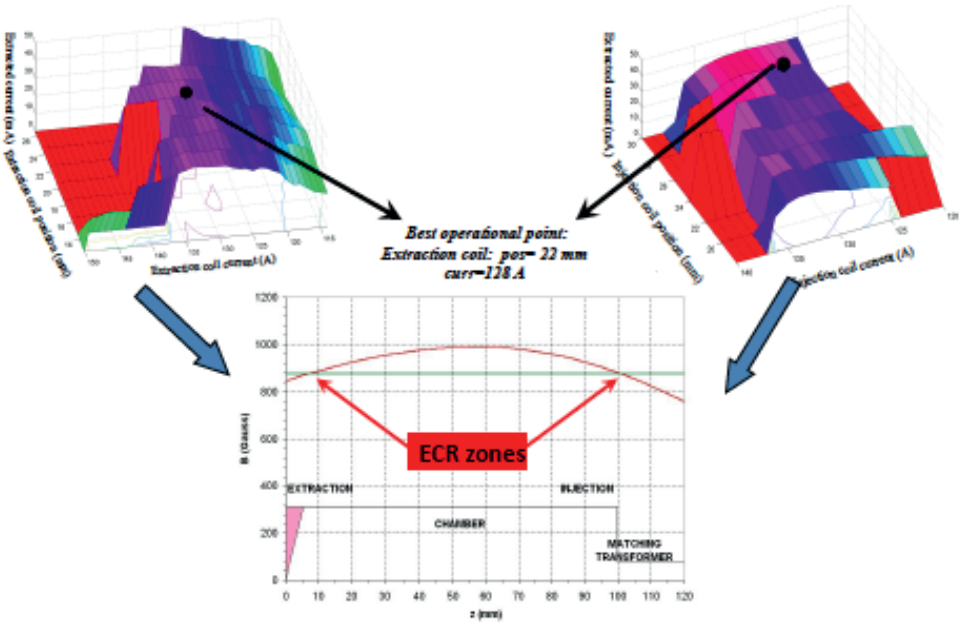


Fig. 13: Magnetic field optimization through changes in the coil current and position (above) and ideal profile (below).

2.6 Versatile Ion Source (VIS)

The TRIPS source exceeded all the requirements of the TRASCO project, but in order to simplify it, the VIS source was built in 2006 [12]. In particular, in the VIS source, the movable coils have been replaced with permanent magnets, and the extraction geometry and extraction column have been simplified. With these modifications it has been possible to avoid the high-voltage platform and the insulating transformer, by insulating the gas pipe and the waveguide line. All these changes decreased the high-voltage sparks and increased the source reliability. All the devices for the remote control were placed at ground potential, thus leaving only the plasma chamber and the permanent magnets at high voltage. The compact dimensions have also resulted in better and easier maintenance. The magnetic system consists of a set of three Vacodym 745 HR permanent magnets; they are packed together with two soft iron spacers and they are supported by a stainless-steel tube. The layout of the VIS source is reported in Fig. 14 [13].

The microwave line, in Fig. 15, is the result of an optimization study carried out with tools for high-frequency structure simulation in order to reduce microwave losses, simultaneously with an adequate matching of the waves to the plasma chamber. A plasma is generated by means of the microwaves provided by a 2.45 GHz magnetron through a WR 340 (86.4 mm $\times$ 43.2 mm) waveguide excited in the TE_{10} dominant mode. An automatic tuning unit adjusts the modulus and phase of the incoming wave in order to match the plasma chamber impedance with and without the plasma, and a dual-arm directional coupler is used to measure the forward and the reflected power.

In TRIPS both the microwave line and the microwave generator are placed on the high-voltage platform, while in VIS the microwave generator is placed at ground potential. Then, in order to separate the high-voltage region from the grounded one, a waveguide d.c. break has been designed and realized with the support of the HFSS electromagnetic simulator. It is made of 31 aluminium discs of a WR 340 waveguide insulated from one another by means of fibreglass. The conductive parts will be fixed to voltages gradually decreasing from 80 kV to ground voltage, still keeping the insertion loss low. The high-pressure quartz window is placed before the WR 284 water-cooled copper bend in order to avoid any damage due to back-streaming plasma electrons. Finally, a maximally flat matching transformer has been inserted before the plasma chamber, shown in Fig. 16, which is an optimized version of a similar device

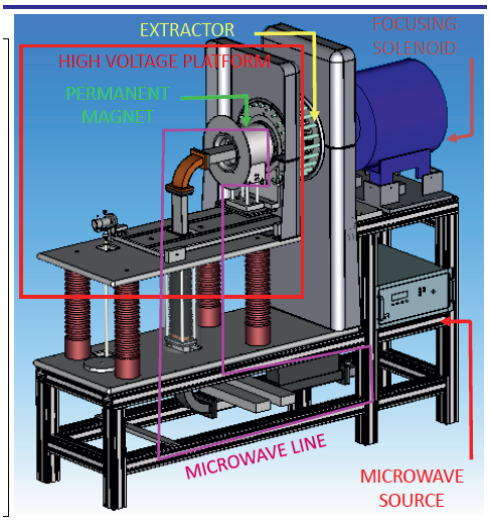


Fig. 14: A render view of the VIS source together with the focusing solenoid

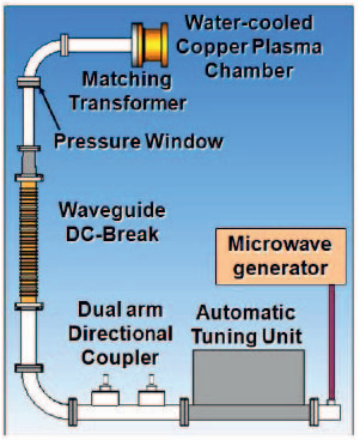


Fig. 15: VIS microwave line

used in MIDAS and TRIPS ion source. This realizes an impedance matching with the plasma chamber and concentrates the electromagnetic field near the axis. The enhancement of the electromagnetic field in the plasma chamber cavity increases the plasma density locally and finally the ionization process.

The ionic component of the plasma produced in the plasma chamber is then extracted by means of a four-electrode extraction system. The low-energy beam transport line (LEBT) allows beam analysis, and it consists of a focusing solenoid, a four-sector diaphragm to measure the beam misalignments, a d.c. current transformer, a 30° bending magnet (Fig. 17) and an insulated beam stop to measure the beam current, shown in Fig. 18. The extraction system consists of a plasma electrode made of molybdenum at 65 kV voltage, two water-cooled grounded electrodes and a 3.5 kV negatively biased screening electrode inserted between them, in order to avoid secondary electrons due to residual gas ionization, back-streaming to the extraction area. The VIS extraction has been optimized to work around 40 mA, fulfilling the requirement of high brightness. The beam emittance measurements are described in the next section. The beam availability has been further increased and damage to the electronics because of high-voltage sparks is negligible.

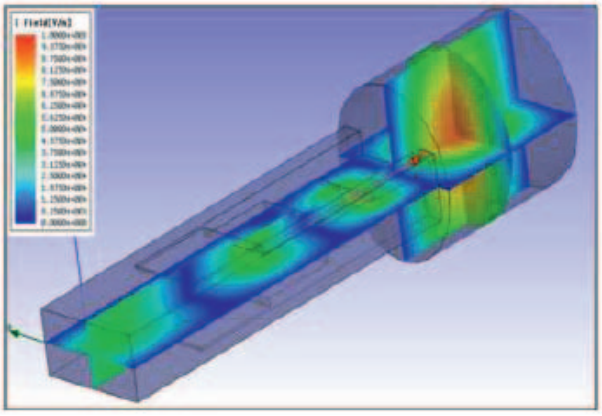


Fig. 16: Electric field amplitude at 2.45 GHz

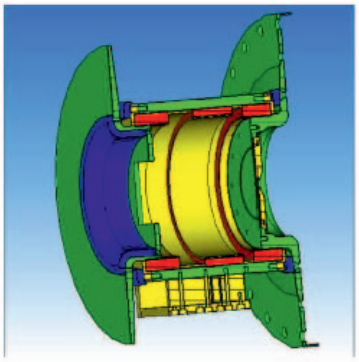


Fig. 17: Magnet

Fig. 18: VIS beam line

3 Beam emittance measurements

The production of a high-quality beam has been, since the first measurements made at CRNL by Taylor and Wills, one of the major challenges for such sources, together with beam reliability. Over the years, many developments have been carried out to improve these aspects, and the most important results have been obtained through a deep optimization of the extraction geometry together with an appropriate design of the low-energy beam transfer line.

3.1 SILHI beam emittance improvements

A significant step forward in this field, namely the production of a high-quality beam, was made within the framework of the collaboration between INFN and CEA, in which an innovative method based on the controlled injection of a gas into the line was developed [14]. The idea is based on the fact that the ions obtained from residual gas ionization are expelled from the centre of the beam line, where the potential is positive, towards the wall. Electrons from the wall are attracted towards the beam, so that the beam is compensated, provided that the pressure is high enough to have an adequate number of electrons (a compromise between beam losses and space-charge compensation is to be found experimentally). According to that approach, the most effective gases are the heaviest, which easily release a large number of electrons. If N hydrogen atoms per unit volume are required for optimum compensation (i.e. if N electrons neutralize the beam space charge), for a species that gives Z_{eff} electrons when the atoms interact with the 95 keV proton beam, the optimum number of atoms per unit volume is N/Z_{eff} (this clue neglects the dissociation process, which does not require a large amount of energy).

Experiments have been carried out by injecting different gases (H_2, N_2, ^{84}Kr and Ar) into the SILHI beam line through a leak valve placed after the LEBT solenoid and comparing the emittance measurements at different pressures with an extracted beam in range of 75–80 μA. Two gauges (measuring p_1 and p_2) are placed, respectively, at the extraction column exit and between the sampler and the profiler in the diagnostic box (Fig. 10). In all the cases considered here, a decrease of beam emittance has been observed with the beam line pressure increase. These results have been explained by a higher degree of space-charge compensation as confirmed by a series of measurements carried out in collaboration with the Los Alamos National Laboratory (LANL).

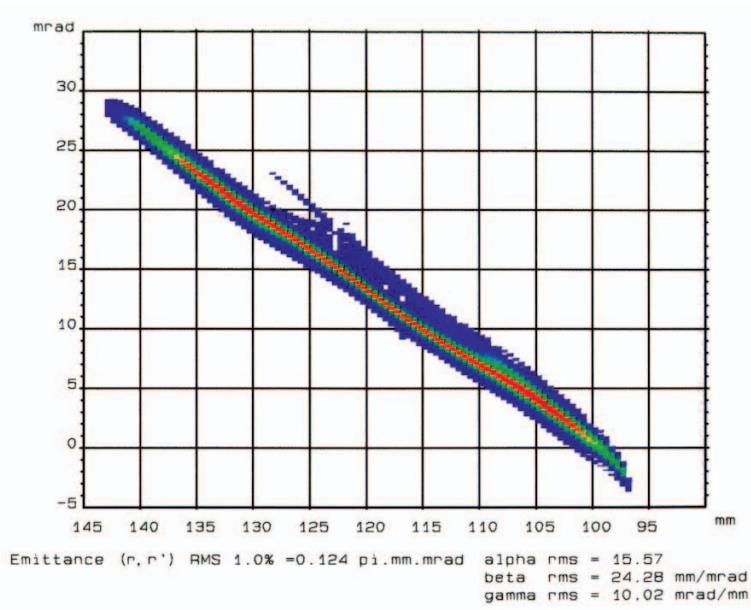


Fig. 19: Emittance plot after injecting Ar into the beam line: $p_1 = 4.5 \times 10^{-5}$ Torr, $p_2 = 4.4 \times 10^{-5}$ Torr, $\epsilon_{\text{rms}} = 0.124\,\pi$ mm mrad.

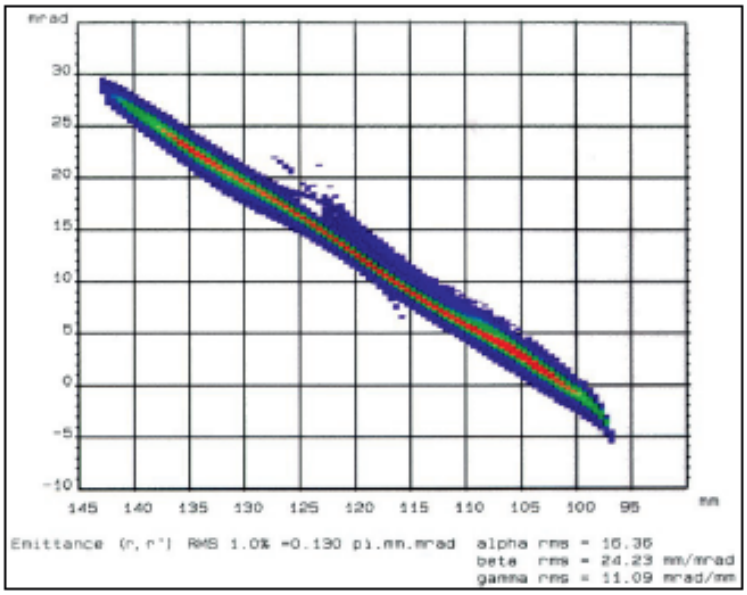


Fig. 20: Emittance plot after injecting N_2 into the beam line: $p_1 = 4.5 \times 10^{-5}$ Torr, $p_2 = 4.5 \times 10^{-5}$ Torr, $\epsilon_{\rm rms} = 0.13\,\pi$ mm mrad.

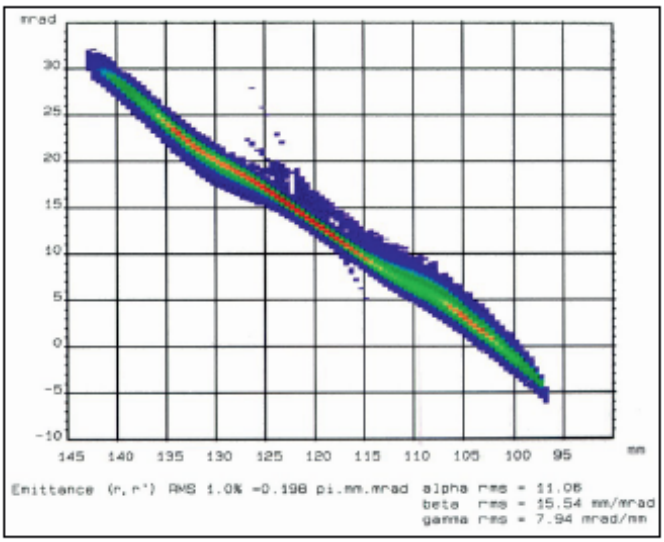


Fig. 21: Emittance plot after injecting H_2 into the beam line: $p_1 = 5 \times 10^{-5}$ Torr, $p_2 = 4.9 \times 10^{-5}$ Torr, $\epsilon_{\rm rms} = 0.198\,\pi$ mm mrad.

As observed, the behaviour depends upon the atomic mass of the species injected. In particular, an r.m.s. normalized emittance of 0.1 π mm mrad has been obtained with ^{84}Kr injected into the beam line ($p_1 = 3 \times 10^{-5}$ Torr). The emittance is reduced by a factor of 3 with respect to the value 0.33 π mm mrad measured without gas ($p_1 = 1.8 \times 10^{-5}$ Torr). Emittance values below 0.15 π mm mrad have been easily obtained also using Ar with pressure values of the first gauge around 2.5×10^{-5} Torr. For this gas, the minimum value of r.m.s. normalized emittance of 0.125 π mm mrad has been measured (see Fig. 19). A lower efficacy is obtained with N_2 and H_2 injection: in particular, with N_2 injection we have measured 0.13 π mm mrad at relatively high pressure ($p_1 = 4.5 \times 10^{-5}$ Torr; see Fig. 20), while for H_2 the minimum value of emittance obtained is 0.198 π mm mrad, as shown in Fig. 21.

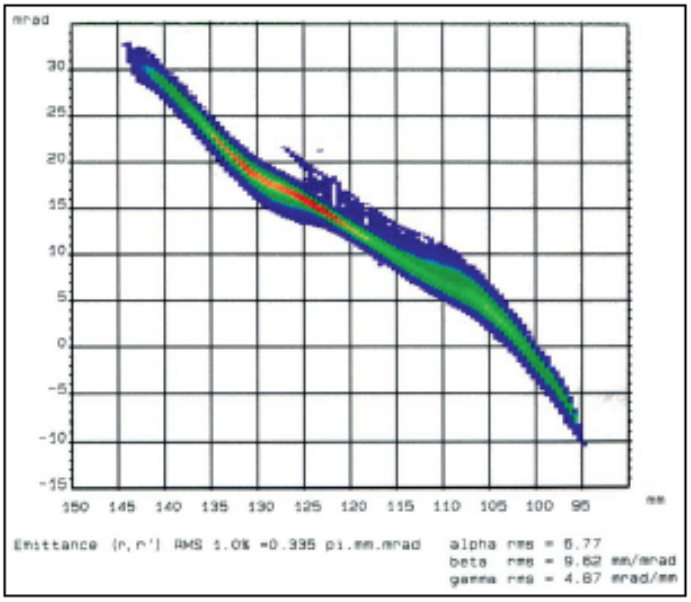


Fig. 22: Emittance picture without injecting ^{84}Kr into the beam line: $\epsilon_{\rm rms} = 0.33\,\pi$ mm mrad, $p_1 = 8\times10^{-5}$ Torr, $p_2 = 1.2 \times 10^{-5}$ Torr.

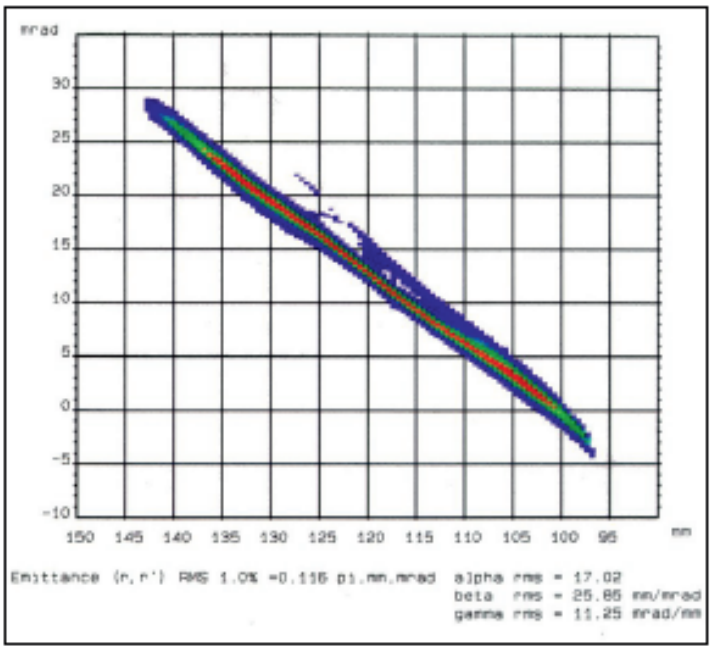


Fig. 23: Emittance picture after injecting ^{84}Kr into the beam line: $\epsilon_{\rm rms} = 0.11\,\pi$ mm mrad, $p_1 = 3.5 \times 10^{-5}$ Torr, $p_2 = 2.7 \times 10^{-5}$ Torr.

Figures 22 and 23 show the emittance measurement obtained with ^{84}Kr. It must be pointed out that the emittance pictures become straight with the gas injection and present a lower beam size and lower aberrations. This is also evident during the measurement because each beamlet selected by the sampler has a lower width and a higher intensity.

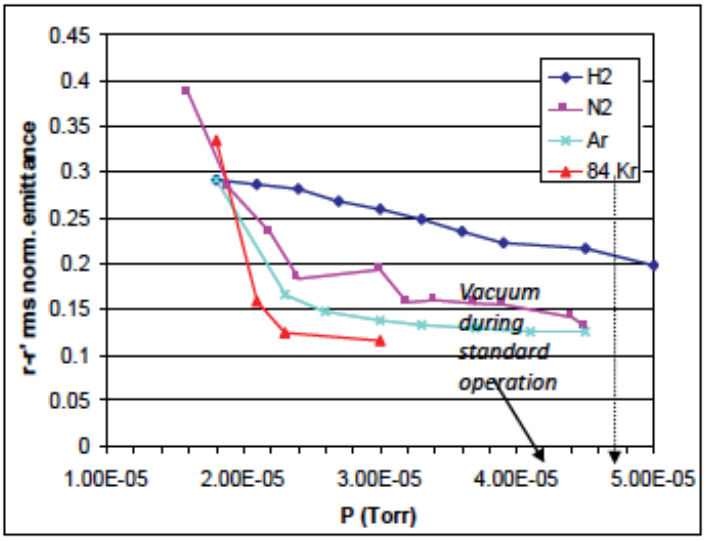


Fig. 24: Space-charge compensation with H_2, N_2, Ar and Kr

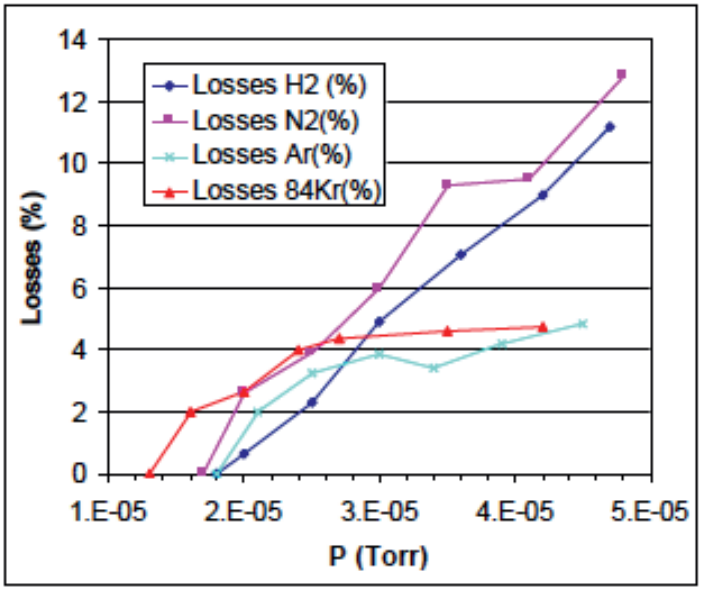


Fig. 25: Losses of the beam current

Figure 24 summarizes the results obtained and in all cases a decrease of beam emittance has been observed with the increase of beam line pressure. Figure 25 summarizes also the losses at the end of the beam line at the different pressures, losses that are less than 5% for masses heavier than Ar.

3.2 TRIPS beam emittance improvements

The requirement in terms of proton current for the TRIPS source (Table 2) was significantly lower than that for the SILHI source, and therefore (at the nominal beam current) gas injection has been not necessary to achieve the required goal. However, also in this case this method has been important to decrease the emittance at higher currents.

In this case the measured values were coherent with the calculated ones and close to those predicted by the equation

$$\epsilon_{\mathrm{normal,rms}} = \frac{1}{4} \times 0.0653r \left(\frac{kT}{A}\right)^{1/2} \qquad (\pi\,\mathrm{mm\,mrad}), \tag{1}$$

where r is the radius of the extraction aperture in millimetres, A is the rest mass of the ion in atomic units and kT is the ion temperature in electronvolts. Assuming an ion temperature of 1 eV and with an extraction hole of 3 mm radius, the expected r.m.s. normalized emittance is circa 0.05 π mm mrad. The emittance values measured are close to the theoretical one for a beam current close to the nominal value, while at higher currents the space-charge effects play an important role in the beam blowup, and a gas must be added into the beam line to decrease the beam emittance.

Measurements [15] have been carried out at different current levels (between 30 and 50 mA); for a fixed current value, we have explored the effect of some parameters such as puller voltage and solenoid current on the beam emittance. A first observation was that the emittance slightly changes with the puller voltage for a fixed current. Measurements performed on a 42 mA beam, with the solenoid fixed at 290 A, a hydrogen flux of 0.49 sccm and a discharge power of 650 W, show a decrease of emittance from 0.208 to 0.172 π mm mrad just by increasing the voltage drop between the extraction and puller electrode from 38 to 43 kV. The optimal operating value ranges between 40 and 42 kV; unfortunately, for higher voltage HV discharges occur and it was not possible to operate the source safely. Another important observation was obtained by looking at the evolution of the beam emittance by changing the solenoid strength for fixed source conditions.

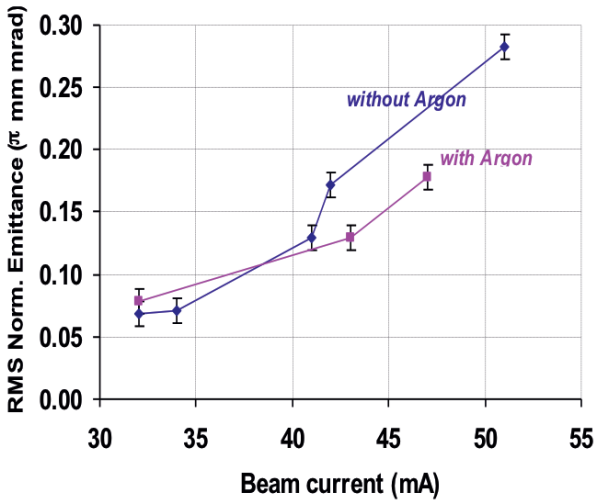


Fig. 26: The r.m.s. normalized emittance versus beam current

Measurements were carried out on a 32 mA beam with the puller fixed at 40 kV, the discharge power at 550 W and the hydrogen flux at 0.45 sccm; the emittance changes between 0.069 and 0.137 π mm mrad by decreasing the solenoid current. This in turn means that the position of the crossover along the beam line is extremely important for the optimum coupling with the RFQ, as at the crossover

the space-charge forces are more important, as confirmed by the measurements with a four-grid analyser. An optimal value of solenoid current for the operation with the EMU around 280 A has been found to avoid an excessive power density over the EMU sampler. Finally Fig. 26 shows the evolution of the emittance with the beam current increase; this confirms the influence of beam line pressure on the space-charge compensation (and thus on the emittance) by means of the injection of a controlled amount of argon into the beam line, as already done for the SILHI source for different gases.

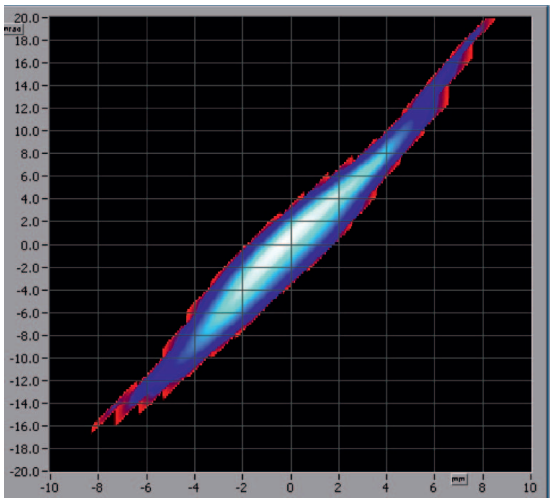


Fig. 27: Proton beam emittance measured without gas ($P_{\text{source}} = 1.5 \times 10^{-5}$ mbar, $P_{\text{line}} = 8.8 \times 10^{-6}$ mbar, $I_{\text{extr}} = 32$ mA, $\epsilon_{\text{norm}} = 0.069\,\pi$ mm mrad).

The measures have been carried out for fixed solenoid current (280 A) and puller voltage (40 kV) and by changing the discharge power from 450 to 650 W and the hydrogen flux from 0.45 to 0.59 sccm. It has been observed that the space-charge forces play a role for currents greater than 40 mA; the background pressure in that condition was around 1.5×10^{-5} mbar in the extraction column and around 8.5×10^{-6} mbar in the beam line. By adding an argon leak, the pressure increased to 2.1×10^{-5} mbar in the whole line, and a 30% beam emittance decrease was obtained without significant beam losses.

Figure 27 shows the emittance measured for the nominal current, while Figs. 28 and 29 show the emittance patterns for higher currents with and without gas injection.

In these cases the emittance growth is justified because of the lower space-charge compensation; then the benefits of argon injection are visible. In fact such gas injection improves the compensation and leads to lower value of emittance.

4 Future perspectives

In the off-resonance microwave ion sources, the ECR is not a predominant condition for plasma generation; in fact, in this case higher electron densities can be obtained by means of a microwave discharge at higher magnetic field value and higher pressures. The plasma–cavity system is a distributed parameter resonant circuit, since the length of the plasma is of the same order of magnitude as the free-space electromagnetic wavelength. Then, differently from the ECR ion sources, the cavity diameter and excitation frequency are chosen to allow only one cavity mode to be excited for a given cavity length. In this case the power coupling into a given mode is usually accomplished by means of tuning stubs in order to achieve the necessary impedance match. The steady-state microwave discharge is characterized by the equality between the power absorbed by the plasma and the lost power, mainly due to inelastic ionization, excitation collisions and energy transmission out of the active discharge region. The power absorbed by the plasma is given by one-half of the real part of the complex Poynting vector, and therefore it depends

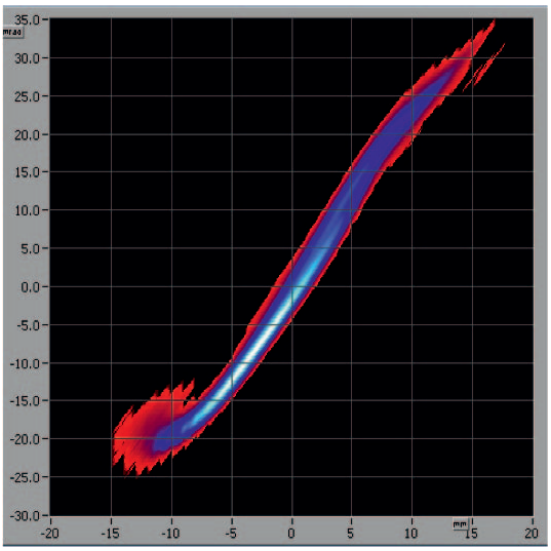


Fig. 28: Proton beam emittance measured without Ar ($P_{\text{source}} = 1.9 \times 10^{-5}$ mbar, $P_{\text{line}} = 1.1 \times 10^{-6}$ mbar, $I_{\text{extr}} = 51$ mA, $\epsilon_{\text{norm}} = 0.283\,\pi$ mm mrad).

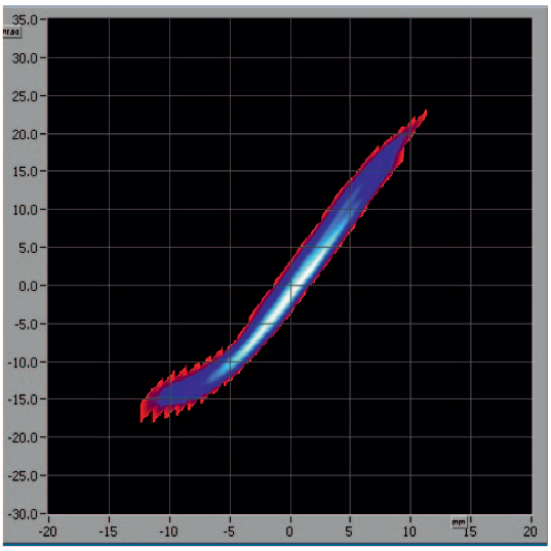


Fig. 29: Proton beam emittance measured with Ar ($P_{\text{source}} = 2.45 \times 10^{-5}$ mbar, $P_{\text{line}} = 2.2 \times 10^{-5}$ mbar, $I_{\text{extr}} = 47$ mA, $\epsilon_{norm} = 0.178\,\pi$ mm mrad).

on the electric field strength in the plasma. In order to optimize the coupling for a given mode, the maximization of that electric field is therefore important. A waveguide transformer (usually maximally flat design) is widely used for such purposes for most microwave sources in operation nowadays. Such a device realizes a progressive matching between the waveguide, normally operating in the dominant mode, and the equivalent impedance of the plasma-filled chamber, also concentrating the electric field at its centre. Figure 30 shows a comparison of the electric field on the plasma chamber axis in the case that no transformer is used or by employing the TRASCO Intense Proton Source (TRIPS) or Versatile Ion Source (VIS) [16]. The excitation frequency is 2.45 GHz in all cases and it can be observed that an increase by a factor of 2 can be obtained, in the frequency range of circa 400 MHz, by appropriately shaping the waveguide ridges as discussed in detail elsewhere [17].

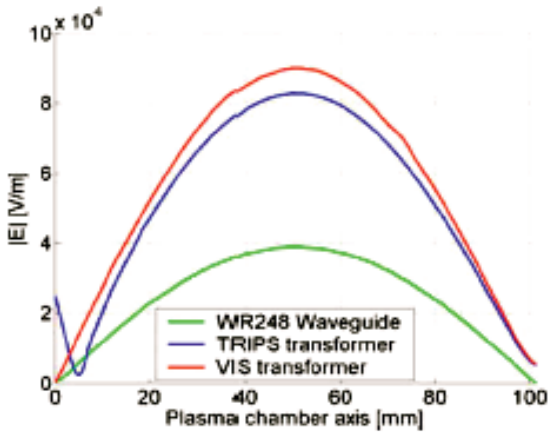


Fig. 30: Comparison of the electric field on the plasma chamber axis in different cases

Therefore, both matching transformers concentrate the electric field around its axis in a smaller region than the original WR 284 cross-section. This feature has been observed in the traces on the boron nitride disc at the injection side of the TRIPS and VIS plasma chambers, and it is particularly remarkable for the production of ion beams because the extraction system is centred on the axis of the plasma chamber and any enhancement of the plasma density at the centre of the cavity leads to a similar increase in the ion beam current. Even higher enhancements are made possible with such a device with different designs or by extending the ridge also within the span of the plasma chamber. Nowadays most of the high-intensity ion sources use automatic tuning units to optimize the power coupling into the operational mode and waveguide transformers similar to those previously described to enhance the plasma density inside the source. The increase in the latter parameter is mandatory for further increase in the produced currents.

Usually in microwave discharge ion sources used for intense beam production, the microwaves are provided by means of a waveguide located longitudinally with respect to the chamber axis. In these devices, the axis of symmetry of the magnetic field coincides with the chamber axis; this means that the injected wave is mainly an O or R wave when it propagates inside the plasma. However, because of the complex structure of the magnetic field lines, and because of possible reflections at the chamber walls, these modes may convert each other, and also X modes can be generated somewhere inside the cavity, through an O–X conversion. At that time, the X waves can be directly converted into a Bernstein electrostatic wave (BW) at the upper hybrid resonance (UHR) layer. BWs are a great advantage, as they travel into the plasma without any cutoff and provide heating even in the case of overdense plasmas. However, if the injection angle of the O mode is not optimal, the BW creation process has a low efficiency. The variations of the power and the background pressure change the plasma properties, and in turn affect the conversion efficiency. As BWs are known to be absorbed at cyclotron harmonics, the electrostatic wave generation in an MDIS is possible and can be used as an alternative plasma heating method. In order to increase the BW creation efficiency, the proper injection angle is needed: this can be achieved by using single cut antennas (waveguides) launching O waves in the right direction with respect to the magnetic field lines. In this way O–X–B conversion is possible, as observed in Ref. [18]. More details are discussed in Ref. [19]. A detailed investigation of this new approach to plasma heating in an MDIS may make it possible to take a large step towards higher extracted beam currents: in fact the key parameter in MDISs is the electron density, more so than the temperature (low charged ions are usually required), and by means of electrostatic wave heating we should be able to overcome the cutoff density at 2.45 GHz of a factor of 5 or 6.

References

[1] B. Wolf, *Handbook of Ion Sources* (CRC Press, Boca Raton, 1995).

[2] N. Sakudo, *Rev. Sci. Instrum.* **49** (1978) 940.

[3] J. Ishikawa, Y. Takeiri and T. Takagi, *Rev. Sci. Instrum.* **55** (1984) 449.

[4] I.G. Brown, *The Physics and Technology of Ion Sources* (Wiley-VCH, Weinheim, 2004).

[5] T. Taylor, *Rev. Sci. Instrum.* **63** (4), April (1992), 2507

[6] J. Ishikawa, Y. Takeiri and T. Takagi, *Rev. Sci. Instrum.* **55** (1994) 449.

[7] J. Sherman, A. Arvin, L. Hansborough, D. Hodgkins, E. Meyer, J.D. Schneider, H.V. Smith Jr., M. Stettler, R.R. Stevens Jr., M. Thuot, T. Zaugg and R. Ferdinand, Status report on a dc 130 mA, 75 keV proton injector (invited). *Rev. Sci. Instrum.* **69** (2) (1998) 1003.

[8] P.-Y. Beauvais *et al.*, *Rev. Sci. Instrum.* **71**(3) (2000) 1413–1416.

[9] J.-M. Lagniel *et al. Rev. Sci. Instrum.* **71**(2) (2000) 830–835.

[10] R. Gobin *et al. Rev. Sci. Instrum.* **73**, (2002), 922

[11] L. Celona, G. Ciavola, S. Gammino, R. Gobin and R. Ferdinand, TRIPS: the high intensity proton source for the TRASCO project. *Rev. Sci. Instrum.* **71** (2000) 771–773.

[12] G. Ciavola, L. Celona, S. Gammino, M. Presti, L. Ando, S. Passarello, X.Zh. Zhang, F. Consoli, F. Chines, C. Percolla, V. Calzona and M. Winkler, A version of the Trasco Intense Proton Source optimized for accelerator driven system purposes. *Rev. Sci. Instrum.* **75** (2004) 1453–1456,.

[13] F. Maimone, G. Ciavola, L. Celona, S. Gammino, D. Mascali, N. Gambino, R. Miracoli, F. Chines, S. Passarello, G. Gallo and E. Zappala, Status of the Versatile Ion Source VIS, Proc. 11th European Particle Accelerator Conf. (EPAC08), Genoa, Italy, 2008, MOPC151, 430.

[14] R. Gobin, P.-Y. Beauvais, R. Ferdinand, P.-A. Leroy, L. Celona, G. Ciavola and S. Gammino, Improvement of beam emittance of the CEA high intensity proton source SILHI. *Rev. Sci. Instrum.* **70** (6) (1999) 2652.

[15] L. Celona, G. Ciavola, S. Gammino, R. Gobin and R. Ferdinand, Status of the TRASCO proton source and emittance measurements. *Rev. Sci. Instrum.* **75** (2003) 5.

[16] L. Celona, G. Ciavola, S. Gammino, F. Chines, M. Presti, L. Ando, X.H. Guo, R. Gobin and R. Ferdinand, *Rev. Sci. Instrum.* **75** (2004) 1423.

[17] L. Celona et al., *Rev. Sci. Instrum.* **81** (2010) 02A333.

[18] H.P. Laqua, W7-AS Team and ECRH Group, *Plasma Phys. Controlled Fusion* **41** (1999) A273.

[19] H. Laqua, M. Otte, Y. Podoba, D. Mascali, S. Gammino, L. Celona, G. Ciavola, N. Gambino, R. Miracoli and F. Maimone, Study of the Bernstein wave heating in the WEGA stellarator plasma and possible application to ECRIS, INFN, Report No. INFN/AE-09/1 (2009).

Microwave Coupling to ECR and Alternative Heating Methods

L. Celona
Istituto Nazionale di Fisica Nucleare, Laboratori Nazionali del Sud, Catania, Italy

Abstract
The Electron Cyclotron Resonance Ion Source (ECRIS) is nowadays the most effective device that can feed particle accelerators in a continuous and reliable way, providing high-current beams of low- and medium-charge-state ions and relatively intense currents for highly charged ions. The ECRIS is an important tool for research with ion beams (in surface, atomic, and nuclear science) while, on the other hand, it implies plasma under extreme conditions and thus constitutes an object of scientific interest in itself. The fundamental aspect of the coupling between the electromagnetic wave and the plasma is hereinafter treated together with some variations to the classical ECR heating mechanism, with particular attention being paid to the frequency tuning effect and two-frequency heating. Considerations of electron and ion dynamics will be presented together with some recent observations connecting the beam shape with the frequency of the electromagnetic wave feeding the cavity. The future challenges of higher-charge states, high-charge breeding efficiency, and high absolute ionization efficiency also call for the exploration of new heating schemes and synergy between experiments and modelling. Some results concerning the investigation of innovative mechanisms of plasma ignition based on upper hybrid resonance will be described.

1 Introduction

The Electron Cyclotron Resonance Ion Source (ECRIS) is used to deliver beams of singly or multiply charged ions for a wide range of applications in many laboratories. In particular, such devices are well suited for the production of Highly Charged Ions (HCIs) – a key consideration for the new acceleration facilities (e.g. FAIR, RIA, HRIBF, etc.), which will require milliampere levels of HCIs. In order to get such high currents, the recent progress in the performance of the ECRIS has been mainly linked to improvement of the plasma magnetic confinement within the source chamber and to an increase of the frequency for the feeding microwaves which, under proper conditions, develop to higher plasma densities [1]. The trend followed until now has been to increase the frequency and the magnetic field, leading to rising costs for the technology and safety problems for the magnet's cryostat because of the growth of hot electrons: the frequency and magnetic field scaling are close to saturation. Apart from this main route, various techniques are also employed to enhance the production of HCIs, such as the use of secondary emission materials, the wall coatings, the installation of a bias disc, or gas mixing, to mention the most important ones [2].

Moreover, in the past decade, some experiments involving new approaches to feeding have allowed an increase in the performance of the conventional ECR ion sources with a minimum-B magnetic field structure by feeding them with electromagnetic waves that have a large spectral content, or that are obtained by the superimposition of a discrete set of microwaves at different frequencies [3–6].

Even if these experiments have provided interesting results, they have not given us an explanation or a methodology to convey a better understanding of the coupling mechanism between a feeding waveguide and a cavity filled with plasma, and the energy transfer between the electromagnetic field in the source plasma chamber and the plasma confined therein. An increase of knowledge in terms of microwave coupling to ECRIS plasma, and therefore optimization of the ECR power transfer processes, may allow us to design ion sources with higher performance. Further improvements of ECRIS output

currents and the average charge state require a deep understanding of electron and ion dynamics in the plasma, and of the impact of the electromagnetic structure on the electron density distribution: theoretical investigations, in order to find a reasonable and adequate model for these mechanisms and to predict their possible improvement, may lead to the design of a future innovative ECR ion source.

Over recent years, dedicated experiments at INFN-LNS have aimed to investigate these topics and, at the same time, some theoretical studies have been undertaken to describe them in detail [7].

The microwave coupling between the electromagnetic wave and the plasma determines the efficient transfer of the energy from the microwaves to the plasma electrons inside the ECRIS. The plasma chamber, on the other hand, can be considered as a resonant cavity for the electromagnetic waves and, when the magnetized plasma is created, the electrical permittivity of the medium filling the chamber is no longer homogeneous and no longer electrically isotropous. A detailed investigation has been carried out to demonstrate that the performance of ECR ion sources depends on the electromagnetic field excited inside the plasma chamber and on the coupling mechanism used to provide the microwaves to the plasma. In the following, the various approaches used over the years are presented, together with the possible alternative heating schemes.

2 The ECR Standard Model

Up to now, the so-called 'ECRIS Standard Model' has, for the greater part of a decade, been the road map followed by ECRIS designers. The main rules were confirmed by experiments performed at MSU-NSCL in 1993–4 and 1995, and can be summarized as follows [8]:

- the radial magnetic field value at the plasma chamber wall must be $B_{\mathrm{rad}} \geq 2B_{\mathrm{ECR}}$;
- the axial magnetic field value at injection must be $B_{\mathrm{inj}} \simeq 3B_{\mathrm{ECR}}$ or more;
- the axial magnetic field value at extraction must be about $B_{\mathrm{ext}} \simeq B_{\mathrm{rad}}$;
- the minimum value of the axial magnetic field must be in the range $0.30 < B_{\mathrm{min}}/B_{\mathrm{rad}} < 0.45$; and
- the optimum power must increase with the volume of the plasma and with the square of the frequency.

Up to now, almost every operating ECRIS has complied with the Standard Model: the extracted current increases as the microwave frequency increases, but only an increase in the mirror ratio can exploit the optimal performance, making the increase of the electron density with frequency effective. Also, according to the Standard Model, the development of the ECRIS is strictly linked to improvements in superconducting magnets and in the technology of microwave generation. Various authors have studied the RF coupling to the plasma in terms of the maximum power rate per unit volume and its relationship with the beam intensity produced by different ECR ion sources [9], but this description is not satisfactory. The cavity design, and the microwave injection geometry, are of primary importance for high-RF-energy transmission to the plasma chamber. Note that the problem of wave energy transmission into the plasma must be divided into two parts: the first is related to the microwave generator–waveguide–plasma chamber coupling, while the second concerns the wave–plasma interaction. Both of these two aspects play notable roles for the future improvement of ECRIS performance. Many experiments have been undertaken outside the framework traced by the Standard Model, to verify the possibility of improving the plasma heating. Principally, they follow three different road maps. A first series of experiments has been devoted to the study of the variation of ECRIS performance with slight variations of the microwave frequency. The second route regards the possibility of operating with more than one frequency for plasma heating (usually two) or of using broad microwave spectra. The third method is actually a mixture of the previous ones: two or more frequencies are used for plasma heating, but at least one is provided by a broadband microwave generator (such as a Travelling-Wave Tube, or TWT), which allows the effects of frequency tuning and of multi-frequency heating to be combined.

2.1 The ECRIS plasma chamber as a resonant cavity

In order to characterize the electromagnetic field that is present inside the plasma chamber of an ECR ion source and where the particle motion occurs, it is fundamental to make the following assumption: the plasma chamber is a resonating cavity for the electromagnetic wave feeding the plasma. The coupling between the electromagnetic wave and the plasma-filled chamber and the electromagnetic field patterns that can be excited inside the resonating cavity are of primary importance for the characterization and for the properties of the plasma. Furthermore, the characterization of the plasma chamber in terms of the excitable electromagnetic field allows us to make some assumptions about the charged particle motion and energy, as described in the following sections.

2.2 Resonant modes inside a cylindrical plasma chamber

A plasma chamber can be represented, in a first-order approximation, by a cylinder of radius a and length l, filled by a medium with a certain electric permittivity ϵ and a certain magnetic permeability μ. Then, a discrete number of electromagnetic field patterns can exist inside the plasma chamber: the so-called resonant modes. They can be defined by the following equations, defined using a system of cylindrical coordinates (ρ,ϕ,z):

$$\begin{cases} E_\rho = -\dfrac{x_{n\nu}}{a}\dfrac{r}{l}\dfrac{\pi}{h^2}C_n J_n'\left(\dfrac{x_{n\nu}\rho}{a}\right)\cos n\phi \sin\left(\dfrac{r\pi z}{l}\right), \\ E_\phi = \dfrac{n}{\rho}\dfrac{r}{l}\dfrac{\pi}{h^2}C_n J_n\left(\dfrac{x_{n\nu}\rho}{a}\right)\sin n\phi \sin\left(\dfrac{r\pi z}{l}\right), \\ E_z = C_n J_n\left(\dfrac{x_{n\nu}\rho}{a}\right)\cos n\phi \cos\left(\dfrac{r\pi z}{l}\right), \\ H_\rho = -i\dfrac{\epsilon\omega}{h^2}\dfrac{n}{\rho}C_n J_n\left(\dfrac{x_{n\nu}\rho}{a}\right)\sin n\phi \cos\left(\dfrac{r\pi z}{l}\right), \\ H_\phi = -i\dfrac{\epsilon\omega}{h^2}\dfrac{x_{n\nu}}{a}C_n J_n'\left(\dfrac{x_{n\nu}\rho}{a}\right)\cos n\phi \cos\left(\dfrac{r\pi z}{l}\right), \\ H_z = 0. \end{cases} \tag{1}$$

These are the equations describing the Transverse Magnetic (TM) modes, with only the magnetic field components on the transverse plane. The Transverse Electric (TE) modes, with only the electric field components on the transverse plane, can be calculated using the following equations:

$$\begin{cases} E_\rho = \dfrac{\mu\omega}{h^2}\dfrac{n}{p}C_n J_n\left(\dfrac{x_{n\nu}'\rho}{a}\right)\sin n\phi \sin\left(\dfrac{r\pi z}{l}\right), \\ E_\phi = \dfrac{\mu\omega}{h^2}\dfrac{x_{n\nu}}{a}C_n J_n'\left(\dfrac{x_{n\nu}'\rho}{a}\right)\cos n\phi \sin\left(\dfrac{r\pi z}{l}\right), \\ E_z = 0 \\ H_\rho = -i\dfrac{x_{n\nu}'}{a}\dfrac{r}{l}\dfrac{\pi}{h^2}C_n J_n'\left(\dfrac{x_{n\nu}\rho}{a}\right)\cos n\phi \cos\left(\dfrac{r\pi z}{l}\right), \\ H_\phi = i\dfrac{n}{\rho}\dfrac{r}{l}\dfrac{\pi}{h^2}C_n J_n\left(\dfrac{x_{n\nu}'\rho}{a}\right)\sin n\phi \cos\left(\dfrac{r\pi z}{l}\right) \\ H_z = -iC_n J_n\left(\dfrac{x_{n\nu}'\rho}{a}\right)\cos n\phi \sin\left(\dfrac{r\pi z}{l}\right). \end{cases} \tag{2}$$

Here, the time dependence $\mathrm{e}^{i\omega t}$ is omitted; J_n and J_n' are, respectively, the Bessel functions and the derivatives of the Bessel functions of order n; and $x_{n\nu}$ and $x_{n\nu}'$ are their related ν roots. A discrete set of frequencies $\omega/2\pi$ can exist inside the resonance cavity and they are defined as follows:

$$\omega_{n\nu r} = c\sqrt{\frac{r^2\pi^2}{l^2} + h^2}, \tag{3}$$

where the value h is as follows:

$$h = \frac{x_{n\nu}'}{a} \qquad \text{(TE modes)}, \tag{4}$$

$$= \frac{x_{n\nu}}{a} \qquad \text{(TM modes)}. \tag{5}$$

Then, the resonant modes $\mathrm{TE}_{n,\nu,\rho}$ and $\mathrm{TM}_{n,\nu,\rho}$ that exist inside the plasma chamber are identified by the mode parameters and their strength is related to the amplitude associated with the field strength of the wave. By supposing that the chamber is filled with a homogeneous and unmagnetized medium, the corresponding electrical permittivity is given by the relation

$$\epsilon_r = 1 - \left(\frac{\omega_p}{\omega}\right)^2, \tag{6}$$

where ω is the pulsation generating the plasma and ω_p is the plasma pulsation:

$$\omega_p = \sqrt{\frac{n_\mathrm{e} e^2}{m_\mathrm{e} \epsilon_o}}, \tag{7}$$

in which m_e and e are, respectively, the electron mass and the electron charge, ϵ_o is the electrical permittivity in vacuum, and n_e is the electron density. Therefore, under these hypotheses, the plasma build-up can be seen as a change in the electrical permittivity and therefore as a change in the resonant frequencies, because of the change of ϵ in Eq. (3). This model is certainly oversimplified if we consider the plasma of an ECR ion source, but has been applied successfully, giving information on the plasma parameters generated in a small reactor at INFN-LNS.

3 Two- and multiple-frequency heating

Since 1994, the so-called Two-Frequency Heating (TFH) has been used [4, 5] to improve the HCI production by feeding the plasma with two electromagnetic waves at different frequencies instead of one. In some cases, even three or more close frequencies have been used.

The performance of the LBL AECR source has been improved by simultaneously heating the plasma with microwaves of 10 and 14 GHz. The plasma stability was improved and the ion charge-state distribution was shifted to a higher-charge state. The production of high-charge-state ions was increased by a factor of between 2 and 5, or higher, for the very heavy ions such as bismuth and uranium, as compared to the case of single-frequency (14 GHz) heating [4].

In an ECR source, electron cyclotron resonance heating couples microwave power into the plasma electrons. This occurs when the microwave frequency ω_{rf} matches the cyclotron frequency, $\omega_c = B/m_\mathrm{e}$, of the electrons. In high-charge-state ECR sources with one frequency, the geometry of the minimum-B field results in a closed, approximately ellipsoidal ECR surface. The electrons are heated in a thin resonance zone at the surface, as they spiral back and forth between the magnetic mirrors. When two frequencies are used, it is possible to produce two concentric surfaces, the physical separation of which depends on the frequency difference and the strength and gradient of the magnetic field, leading to a higher density of the energetic electrons. With TFH, the ECR plasma is more quiescent than with single-frequency heating. Both the short-term and the long-term plasma stability are improved, and more microwave power could be launched into the plasma. TFH has been demonstrated to be a powerful method: in the case in Ref. [4] for ^{238}U, it increased the production of higher-charge states (from 35+ to 39+) by a factor ranging from 2 to 4 and shifted the peak charge state from 33+ to 36+. Two-frequency heating using 10 and 14 GHz in the AECR provided significantly better performance and indicated that still higher performance with multiple-frequency heating may be possible. While multiple-frequency heating would increase the complexity and cost of an ECR source, it could provide significant gains in performance. In [5], the performance of the Argonne ECR ion source was improved through the use of TFH, with the primary frequency of 14 GHz from a klystron and the second frequency from a Travelling-Wave Tube Amplifier (TWTA) with a tunable range of 11.0–13.0 GHz. Source output as well as stability were improved, with a shift to higher-charge states observed. The use of a second frequency increased the

intensity of the medium charge states by 50–100% and the higher-charge states by a factor of between 2 and 5. In a modern second-generation ECRIS, TFH is usually implemented by using frequencies between 14 and 18 GHz, while the best-performing ECRIS use a combination of 24 or 28 GHz with 18 GHz.

3.1 Multiple-frequency plasma heating

Microwaves of various frequencies can be simultaneously launched into and absorbed by a high-charge-state ECR plasma. The minimum-B magnetic field configuration in an ECRIS can provide many closed and nested ECR heating surfaces, as graphically shown in Fig. 1. If two or more significantly different frequencies are used, two or more well-separated and nested ECR surfaces will exist in the ECR plasma. With the multiple ECR surfaces, electrons can be heated four times or more for one pass from one mirror end to the other, whereas they are only heated twice in the case of single-frequency heating. Multiple-frequency heating can couple more microwave power with better efficiency into the plasma and it leads to a higher density of hot electrons, which is essential for the production of highly charged ions. To demonstrate that the frequency-domain technique can be used to enhance the performance of a traditional minimum-B ECR ion source, comparative studies have been undertaken to assess the relative performance of the Oak Ridge National Laboratory CAPRICE ECR ion source [10], in terms of its multiply charged ion-beam generation capability, when excited with high-power, single-frequency, or multiple-discrete-frequency microwave radiation, derived from standard klystron and/or TWT technologies. These studies demonstrate that the charge-state populations for Ar^{q+} and Xe^{q+}move towards higher values when excited with two- and three-discrete-frequency microwave power compared to those observed when single-frequency microwave power is used. For example, the most probable charge state for Xe is increased by one charge-state unit, while the beam intensities for charge states higher than the most probable one are increased by factors of ~ 3 compared to those observed for single-frequency plasma excitation. In Ref. [11], particle-in-cell codes have been used to simulate the magnetic field distributions, to demonstrate the advantages of using multiple, discrete frequencies over single frequencies to power conventional ECR ion sources.

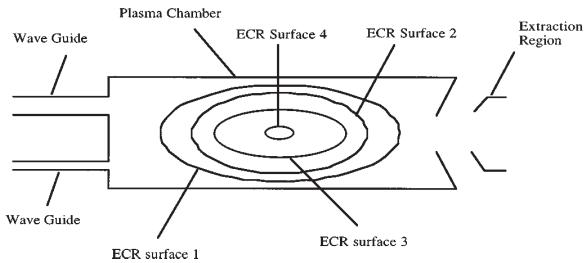


Fig. 1: A schematic view of four nested ECR surfaces in a high-charge-state ECR ion source for four well-separated frequency waves [12].

One, two and three-frequency heating experiments have been conducted using Xe feed gas. With the addition of the second and third frequencies, the most probable Xe^{q+} charge state moves towards higher values by one unit; this clearly illustrates that the performance of conventional-geometry ECR ion sources can be significantly improved by the use of multiple-discrete-frequency plasma heating.

4 Electron heating with broadband microwave radiation and the volume effect

The ionization process in Electron Cyclotron Resonance (ECR) ion sources is based on the sequential removal of electrons in collisions between ions or neutral atoms with electrons that have been heated stochastically to high energies by the adsorption of microwave power under ECR conditions. The ECR

zones in a conventional minimum-B geometry ECRIS, powered by narrow-bandwidth microwave radiation, are thin volumes that are substantially smaller in relation to their total plasma volumes. Consequently, the probability of an acceleration resulting in a substantial energy gain for the electrons that arrive in the ECR zone in phase with the electric field vector of the electromagnetic wave is lower than is possible in extended-volume ECR zones. Therefore, it is reasonable to believe that the performance of conventional minimum-B ECR ion sources can be improved by increasing their respective resonance volumes. The volume effect has been demonstrated by tailoring the central magnetic field so that it forms a large resonant volume and by increasing the number of discrete operational frequencies of the ion source. Broadband sources of RF power provide a simple and potentially more effective alternative for increasing the physical sizes of the resonance zones in conventional B_{minimum} ECR ion sources.

4.1 Experimental procedures

Some experiments have been performed with the JYFL 6.4 GHz conventional minimum-B geometry ECR ion source at the University of Jyväskylä, using Ar as the feed gas. The performance of the source was compared when operated with either narrow or with broadband microwave radiation, under the same neutral gas pressure and input power conditions. For the narrow-bandwidth experiments, the carrier signal from the local oscillator (LO, bandwidth 1.5 MHz) was fed directly into the TWTA. For the broadband experiments, the signal from the LO was fed into a White-Noise Generator (WNG), producing an output signal with a bandwidth of 200 MHz, FWHM, equally distributed about the central frequency of the LO (6.4 GHz). The measurements were carried out at the same absorbed microwave power level (200 W), determined by subtracting the reflected power from the forward power of the TWTA. The input power was limited due to impedance mismatch of certain frequencies in the amplified broadband signal, resulting in 20% reflected power of the injected power. This problem cannot be avoided and it constitutes the main drawback of such a technique, since frequencies are differently coupled into the ion source depending on the intrinsic electromagnetic structure of the source itself. In the case of narrow- bandwidth operation, the reflected power was typically 6%. For the WNG mode of operation at optimum pressure, the ion plasma density, as evaluated from the measured drain current and the charge-state distribution, using Bohm's criterion, was within 15% of the neutral density. Although the beam intensities were higher by factors >2 with the WNG at low pressures, there was no difference between the broadband and narrow-bandwidth modes of operation in the high neutral pressure regime due to increased rates of charge exchange [13,14]. Enlarged ECR zones improve the production of high-charge-state ions due to enhanced bombardment with higher populations of energetic electrons and, especially, due to the enhanced ionization rates of neutrals in the ECR zone, thereby lowering the probability of charge-exchange collisions. The tremendous potential of the broadband technique for enhancing high-charge-state beam intensities over conventional means at equivalent input power levels is illustrated in Fig. 2, where the ratio of WN/LO-generated high-charge-state beams is seen to increase continually with the charge state. Clearly, the potential of the broadband method can only be fully evaluated by comparing the high charge states produced with the two modes of operation at the same power density, if this is achievable in practice in terms of the operational limitations imposed by outgassing effects and the required RF power.

The performance of ECR ion sources can also be significantly improved by tailoring the central region of the magnetic field so that it is resonant with single-frequency microwave radiation (spatial domain) [15–17]. The spatial-domain technique employs a magnetic field configuration with an extended central flat region, tuned to be in resonance with single-frequency microwave radiation. Because of the large resonant plasma volume, significantly more RF power can be coupled into the plasma, resulting in heating of electrons over a much larger volume than is possible in conventional ECR ion sources. The ability to ionize a larger fraction of the particles in the plasma volume effectively reduces the probability of resonant and non-resonant charge exchange, thereby increasing the residence time of an ion in a given charge state and for subsequent and further ionization. All other parameters being equal, the 'volume'

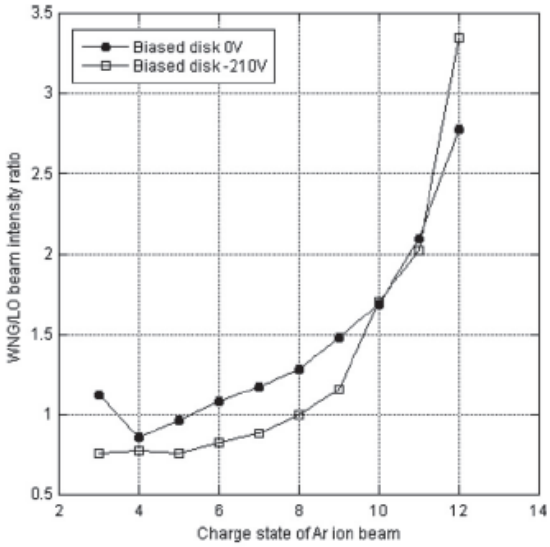


Fig. 2: Intensity ratios of argon ion beams generated with signals from the WNG and LO versus the charge state

ECR source should result in higher-charge-state distributions, higher beam intensities, and improved operational stability. The axial magnetic field is shown in Fig. 3: its profile is flat (constant mod-B) in the centre, and extends over the length of the central field region, along the axis of symmetry, and radially outwards to form a uniformly distributed ECR plasma 'volume'. This magnetic field design strongly contrasts with those used in conventional ECR ion sources, where the central field regions are approximately parabolic and the consequent ECR zones are 'surfaces'.

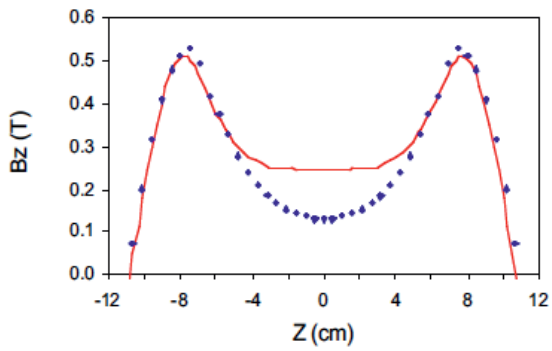


Fig. 3: Axial magnetic field profiles of the 'volume' (solid line) and conventional minimum-B configurations in 'surface' (dotted line) ECR ion sources [18].

According to computational studies, the new configuration will result in dramatic increases in the absorption of RF power, thus enabling the heating of electrons over a much larger volume, thereby increasing the electron temperature and the 'hot' electron population in the plasma.

5 Frequency tuning

Many experiments over recent years have shown that significant improvements of ECRIS performance (both in terms of total extracted current and the production of highly charged ions) can be obtained by slightly varying the microwave frequency in the case of Single-Frequency Heating (SFH), this being

defined as a 'frequency tuning' effect. It was already known that a large increase in the frequency increases (∼GHz) the electron density and improves the ECRIS performance because of the increase of the cut-off density, but since 2001 several experiments have demonstrated that even slight variations of the pumping wave frequency (∼MHz) may lead to strong variations in the extracted current. The first evidence was provided by the different performances observed for the SERSE and CAESAR ion sources when fed by a klystron-based or a TWT-based generator [19–21] at either 14 or 18 GHz. Other interesting results came from experiments performed at GSI and at JYFL. In frequency tuning, the microwave frequency is varied in order to select the most efficient heating mode inside the plasma chamber of the ECRIS. In the experiments, the input frequency for the klystron was swept from 14.05 to 14.13 GHz in 100 s, using a Rohde & Schwartz signal generator. This bandwidth was found adequate to maintain a constant output power over the whole frequency sweep, using the automatic level control feature of the klystron.

5.1 Effect on beam structure and emittance

The beam emittance and the beam structure were studied as a function of the frequency with the aid of an Allison-type emittance scanner and a KBr beam viewer. Figure 4 shows three beam viewer pictures: the beam structure varies strongly with the microwave frequency, as an indication of the change in the plasma–wave coupling during the scan. However, no unequivocal explanation concerning the origin of the beam structure variations can be given: they could originate from the changes in the plasma, in the electron ion dynamics, due to electromagnetic field variations and/or in the beam line due to changes in ion-beam intensity and space charge. The explanation of the data presented above is based on the assumption that the frequency tuning changes the electromagnetic field distribution inside the resonator in terms of its distribution over the resonance surface; that is, where the wave–electron energy transfer takes place. This assumption requires that even in case of plasma filling the cavity, the resonant modes persist; that is, the formation of standing waves is still possible.

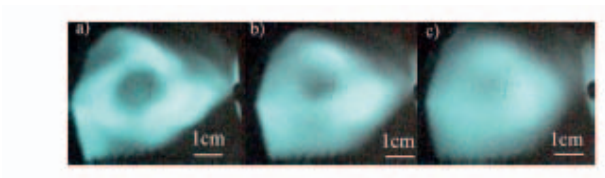


Fig. 4: (Colour online) The structure of the Ar^{9+} ion beam with different plasma heating frequencies is shown: (a) 14.050 GHz, (b) 14.090 GHz, and (c) 14.108 GHz.

5.2 The microwave frequency dependence of the properties of the ion beam extracted from a CAPRICE-type ECRIS

Measurements have been performed with the CAPRICE-type ECRIS installed at the ECR Injector Setup (EIS) of GSI. The experimental set-up uses a microwave sweep generator that feeds a TWTA, covering a wide frequency range from 8 to 18 GHz. This arrangement provides a precise determination of the frequencies and of the reflection coefficient along with the beam properties. A sequence of viewing targets positioned inside the beam line monitors the evolution of the beam shape. In the present experiment, the frequency tuning effect has been analysed in the 12.5–16.5 GHz frequency range. The availability of a TWTA driven by a signal generator made it possible to change the source operating frequency in steps of a few hundred kilohertz. This experiment allows us to analyse the beam properties when the ECRIS operative frequency sweeps over a wide range of 4 GHz, and hence for increasing ECR surfaces. The influence on lower- and higher-charge states has been analysed for different source conditions concerning the magnetic field configuration, the gas pressure, and the power setting.

5.3 Description of experimental set-up

The CAPRICE-type [22] ECR ion source used for this experiment is equipped with a 1.2 T maximum radial magnetic field. The plasma chamber was 179 mm long and 64 mm in diameter. The RF power was provided by a TWTA working in the 8–18 GHz frequency range and able to provide an output power higher than 650 W in the frequency range of 12–18 GHz. The input of the amplifier was driven by a signal generator able to sweep from 1 to 20 GHz. According to the maximum manageable power reflected to the amplifier, it has been restricted to working at a power of 100 W and in the frequency range of 12.5–16.5 GHz. The use of a waveguide microwave isolator covering this frequency range and handling up to 650 W could allow us to work with higher powers.

The frequency steps were set to 200 kHz, with a dwell time of 20 ms for each step. The duration of one measurement was then around 400 s. Two directional couplers with high directivity were inserted in the waveguide line in order to measure the forward power and the reflected power by means of two microwave power probes. The experiment was carried out with argon and helium as support gases, at gas pressures of $(3.9–5.0)\times10^{-6}$ mbar. The ion currents of the Ar^{7+}, Ar^{8+}, and Ar^{9+} charge states were measured using a Faraday cup; the drain current of the high-voltage power supply of the extraction was also recorded. The extraction voltage was set to 15 kV; a voltage of -2 kV was applied to the screening electrode. Viewing targets could be remotely inserted at three positions along the beam line in order to monitor the evolution of the beam shape right after the extraction, the focused beam, and the analysed beam [23]. KBr was used as the target coating material for this experiment.

5.4 Results and discussion

Different measurements were carried out by sweeping the frequency and setting different ion source parameters; that is, the injection and extraction magnetic field values, the gas pressure, and the microwave power. The source parameters were set to operate using a charge-state distribution with a maximum on the Ar^{8+} current (by feeding the plasma with 100 W microwave power at 14.5 GHz). An Ar^{8+} current of 85 μA and a drain current of 2.36 mA were obtained. From these source conditions, the frequency sweep was started by ramping the signal generator from 12.5 GHz up to 16.5 GHz, while the reflection coefficient, the Ar^{7+}, Ar^{8+}, and Ar^{9+} currents, and the drain current were recorded simultaneously. The evolution of the reflection coefficient with the frequency is shown in Fig. 5. As expected by comparison with previous experiments, the matching impedance between the cavity filled by the plasma and the electromagnetic wave is strongly dependent on the frequency [6]. It is remarkable that the plasma properties also change considerably with varying frequency. The strong correlation between the peaks of the reflection coefficient and the current amplitude are clearly visible around the frequencies at which the reflection coefficient is higher than -9.54 dB (matching condition). The relationship between the resonance frequencies and the heating efficiency has been analysed theoretically and particle-in-cell codes have been used to correlate the electromagnetic field patterns and the electron cyclotron resonance surface [24]. However, it is not possible with our analysis to determine the electromagnetic field patterns (modes) related to these peaks. The comparison of the current evolution in the frequency range indicated above is presented in Fig. 5. It is clear how the current amplitude is affected by the choice of the operative frequency. Looking at the Ar^{8+} current, it ranges from a few microamperes up to 200 μA. The experimental results were clearly reproducible in several runs, thus confirming the reliability of the measurements. The evolution of the Ar^{7+} and Ar^{8+} currents is similar, but their amplitudes differ. In fact, at the frequencies at which both currents present a peak, the Ar^{8+} current is quite a bit higher than the Ar^{7+} current (e.g. at 14.119 GHz, the difference between the two currents is 80 μA). The opposite behaviour is visible for the minima of the current amplitudes, where the Ar^{7+} current is higher than the Ar^{8+} current. In the range 15.64–16.5 GHz, the currents of the higher-charge states – that is, Ar^{9+} – tend to lower values even if the reflected power is less than 10%. This seems to be due to the confining magnetic field, which restricts the source operation to lower frequencies. It has been also observed that the evolution of the drain current follows the trend of the three charge states presented in Fig. 5. In order to have a

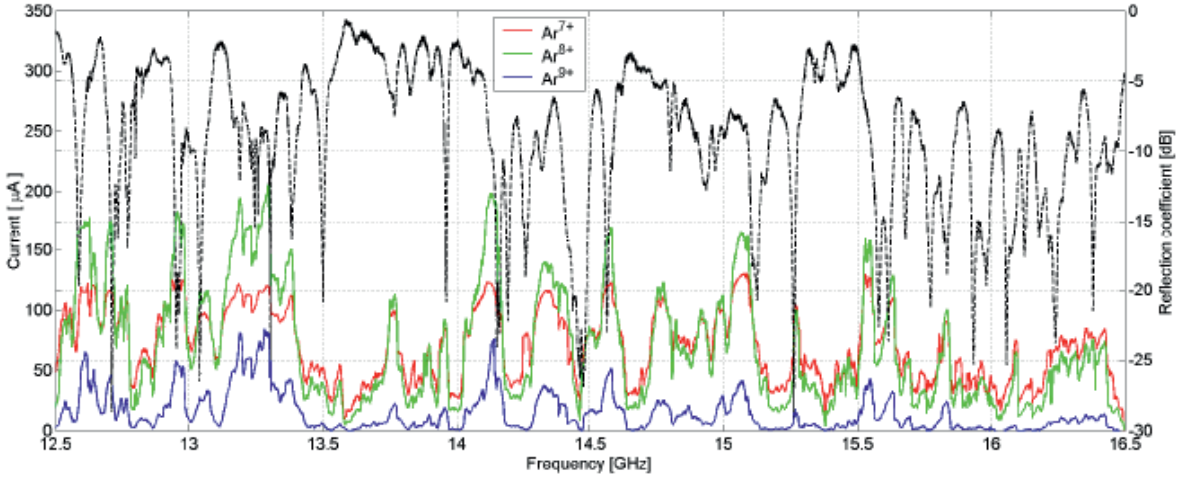


Fig. 5: (Colour online) The reflection coefficient and current evolution versus microwave frequency (the coloured solid lines refer to the left-hand scale, while the dashed black line is the reflection coefficient, referring to the right-hand scale).

complete understanding of the sweep effects on the ionization process, the charge-state distribution has been analysed for different frequencies. It was decided to restrict the analysis to the frequency range of 14–15 GHz. Several frequencies have been considered at which peaks and minimum amplitudes occur. In Fig. 6, the charge-state distributions are presented for four different frequencies. The 14.5 GHz value

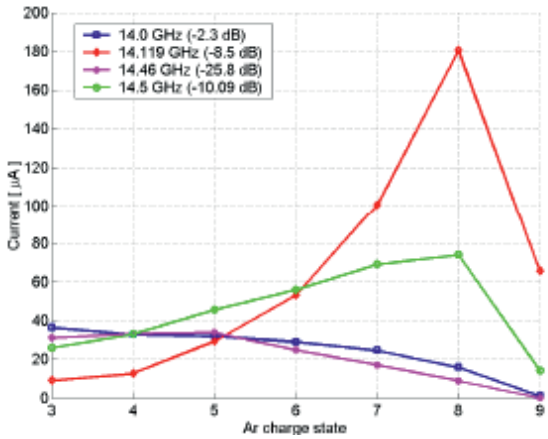


Fig. 6: The Ar charge-state distributions for four significant operating frequencies

is the normal operational frequency of the CAPRICE ion source; 14.46 GHz and 14.119 GHz are the frequencies at which the minimum and the maximum Ar^{8+} currents, respectively, were measured in the 14–15 GHz frequency range. The charge-state distribution related to 14.0 GHz operation is also reported in order to emphasize the importance of the electromagnetic field pattern and the choice of the frequency. In fact, the charge-state distributions at 14.0 GHz and at 14.46 GHz are quite similar, even though the amount of power feeding the plasma was more than doubled. In the first case, the reflection coefficient (indicated in the legend of Fig. 2) was -2.3 dB and more than half of the power was reflected; and in the second case it was -25.8 dB, the impedance matching condition was fulfilled, and hardly any power was reflected. It is also interesting that for the frequencies at which the higher-charge states are favoured,

the current of the lower charge state is decreasing and vice versa. The analysed charge-state distributions confirm that frequency tuning affects the higher-charge states to a greater extent (at 14.119 GHz, the current enhancement with respect to the 14.5 GHz operational frequency is 244% for the Ar^{8+} current and 456% for the Ar^{9+} current). The enhancement of the current is not the only effect of frequency sweeping; in fact, the quality, the shape, and the emittance of the ion beam also vary. The use of beam viewing targets has proven to be a favourable technique for monitoring the beam shape and a promising beam diagnostic tool. The images recorded beyond the extracted beam-focusing solenoid (position VT2) and in the diagnostic box after the mass/charge selection dipole (position VT3) are shown in Fig. 7.

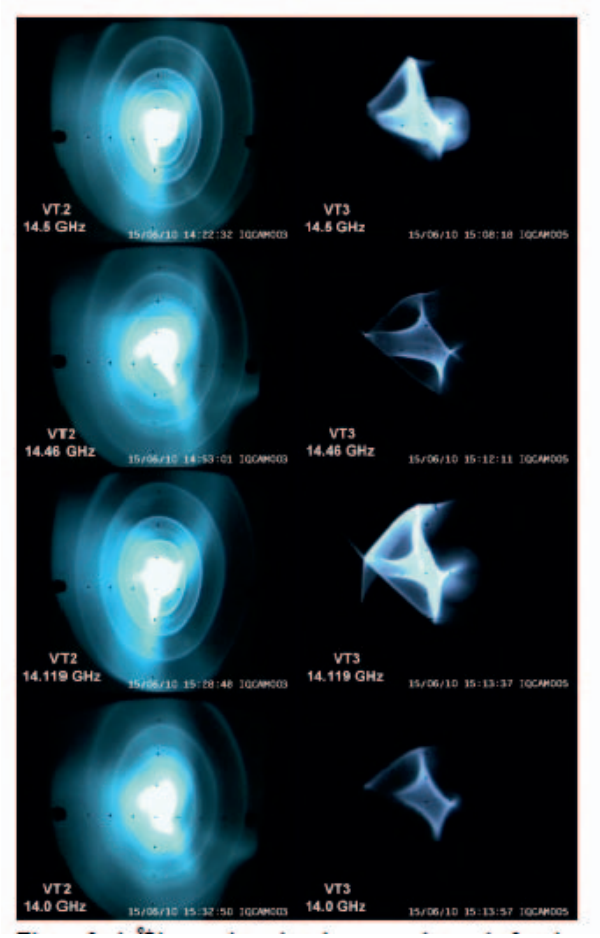


Fig. 7: The Ar^{8+} beam viewed at the targets located after the focusing solenoid (left-hand column) and after the dipole (right-hand column).

Here, the same frequencies are chosen as in Fig. 6. The Ar^{8+} beam is shown in the right column and the focused beam images also refer to the Ar^{8+} magnetic field setting. At 14.0 GHz and 14.46 GHz, which are the frequencies at which the Ar^{8+} current presents a minimum, the focused beam shape seems to remain unchanged in the orientation of the arms (at 14.0 GHz it is a little bit brighter, also according to the higher intensity measured at the Faraday cup). At 14.119 GHz, the focused beam is a little bit bigger and brighter than the one at 14.5 GHz, and in both cases the orientation of the arms is turned by more then 40° clockwise with respect to the 14–14.6 GHz beam shapes.

5.5 Frequency tuning combined with the two-frequency heating effect

The relationship between the two frequencies and the respective power was not univocally determined. In fact, any source features a different set of parameters and the optimization is done empirically, just by looking at the maximization of the beam current. Several qualitative explanations have been offered for this phenomenon, all related to the increase of the average electron temperature, T_e, and to the ionization rate, by assuming that the crossing of two resonance surfaces helps the electrons to gain more energy. This simple picture does not explain the reason for the relevant changes in the charge-state distribution for different pairs of frequencies (even for the case of minor changes – such as a few megahertz over 14 or 18 GHz), which can be explained nowadays in terms of the frequency tuning effect. It is important to underline that, even in the case of TFH applied to many existing sources, a TWT is often used, the other option being a klystron-based generator. The choice of a TWT allows the experimentalists to vary the second frequency slightly. It has been observed – for example, for the production of O^{7+} – that 60 W emitted by the TWT at the optimum frequency gives the same effect as 300 W emitted from the fixed-frequency klystron. The maximum current is obtained by means of a klystron at 427 W and a TWT at 62 W (I = 66 μA), operating simultaneously; in order to obtain the same current, the experimenters needed P_{RF} = 800 W from the klystron in SFH. Furthermore, in the case of TFH, the current increases almost 20% (from 57 to 66 μeA) when the TWT-emitted frequency shifts from 11.06 to 10.85 GHz, the klystron and TWT emitted power both being held constant. Then, the TFH is an effective method to increase the extracted current from an ECRIS, but it can only be fully exploited by means of frequency tuning. Several measurements have been carried out with the SERSE ion source. The TFH has been used for operation at either 14 or 18 GHz, with a clear advantage with respect to SFH.

In section 5.2, the evolution of the ion current intensities of Ar^{7+}, Ar^{8+}, and Ar^{9+} over the 12.5–16.5 GHz frequency range has been described. The argon charge-state distributions were analysed for different frequencies, and for some frequencies an enhanced intensity of the higher-charge states was observed. We decided to perform an experiment to investigate the double-frequency heating effect using a single fixed frequency (14.5 GHz in one case and 14.119 GHz in another) and a second frequency swept over the 12.5–16.5 GHz range. First, the power feeding the plasma was kept at 100 W, in order to compare the results of double-frequency heating with single-frequency sweeping. Then the power distribution between the fixed frequency and the sweeping frequency was unbalanced by doubling the power of the fixed frequency in one case and the power of the sweeping frequency in the other. The ion source parameters – including the magnetic field, the gas pressure, the extraction voltage, and so on – were set to and maintained at the same values as used for single-frequency sweep analysis, with the charge-state distribution optimized for the Ar^{8+} intensity. The Ar^{7+}, Ar^{8+}, and Ar^{9+} currents were recorded together with the extractor drain current. The forward power and the reflected power were measured simultaneously. However, in the case of double-frequency heating, the power probes cannot distinguish the measured values of the forward and reflected power related to each of the two waves. The results of these measurements are shown in Fig. 8. A comparison between the two techniques concerning the behaviour of the ion source is possible. As expected, in the case of the two frequencies, the minima of the reflection coefficient are higher because of the higher level of reflected power due to the superposition of the two waves. It is interesting to note that these minima occur at almost the same frequencies for both techniques. Since the positions of the reflection coefficient minima do not vary, it can be expected that the peaks of the ion current are located at the same frequencies for the two heating methods. This comparison is shown in Fig. 9 for the drain current of the high-voltage extractor power supply. In this case, 50 W is provided at one fixed frequency and the effect of sweeping a second frequency with a power of 50 W produces a different current behaviour from that with a single sweeping frequency (at the same power). This can be seen in Fig. 9 and Figs. 10–11. In fact, during the single-frequency sweep, the drain current varies from around 1.4 mA to 2.8 mA, while with double-frequency heating the range of variation is 0.6 mA. The drain current appears more stable with frequency when the double frequency is applied instead of single-frequency tuning. This behaviour is also observed in

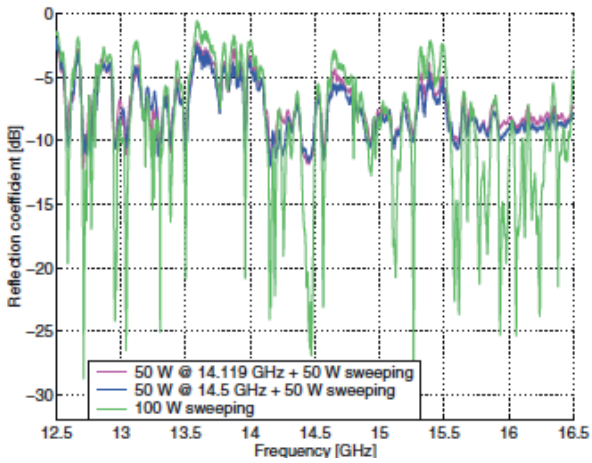


Fig. 8: (Colour online) The reflection coefficient as a function of frequency: a comparison between the frequency tuning effect and double-frequency heating.

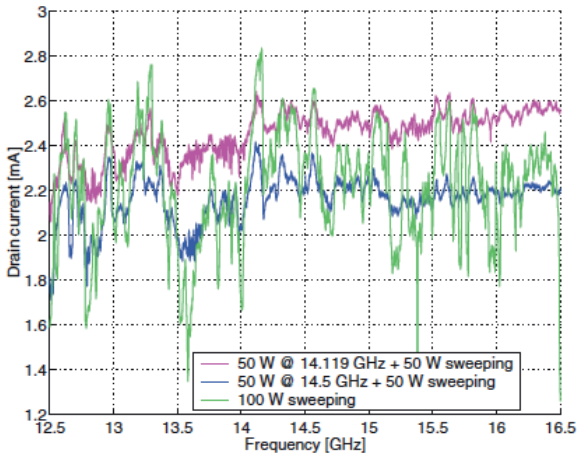


Fig. 9: (Colour online) The drain current of the high-voltage extractor power supply as a function of frequency: a comparison between the frequency tuning effect and double-frequency heating.

the Ar^{7+} current variation shown in Fig. 10. A possible explanation might be found in the microwave power provided by the two methods. In the case of single-frequency tuning, 100 W is provided to the source and during the frequency sweep, when the impedance is not matched, most of the forward power is not coupled to the source but is reflected (see Fig. 8). But in the case of double-frequency heating, just 50 W of the total forward power of 100 W is associated with the sweeping frequency, and at the frequencies at which the power is not coupled to the plasma, one half of the total power is provided at a fixed frequency, which is well coupled to the source (i.e. at 14.5 GHz or at an optimized frequency such as 14.119 GHz). The choice of the fixed frequency allows enhancement of the ion currents of the higher-charge states. With respect to single-frequency operation, with two frequencies the current peaks are broadened and reduced in amplitude, the frequencies at which they occur remain the same, and the average current level remains high (see Fig. 10 for Ar^{7+}), while the higher-charge states tend to higher ion currents when the frequency is increased (see Figs. 11 and 12). The experimental results indicate that there can be strong differences when the operating frequency is optimized by frequency tuning. In fact, when operating in double-frequency mode and optimizing one of the two frequencies (14.119 GHz

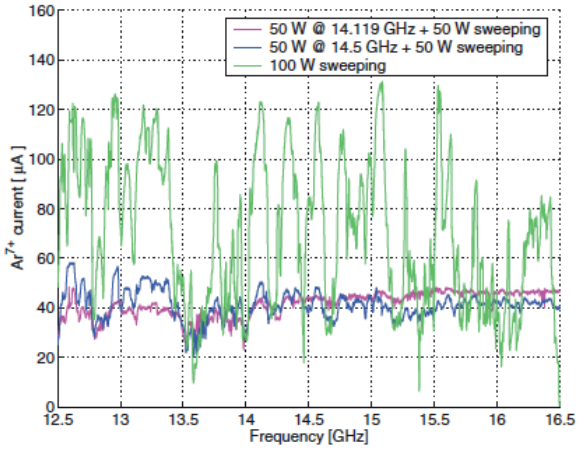


Fig. 10: (Colour online) The Ar^{7+} current as a function of frequency: a comparison between the frequency tuning effect and double-frequency heating.

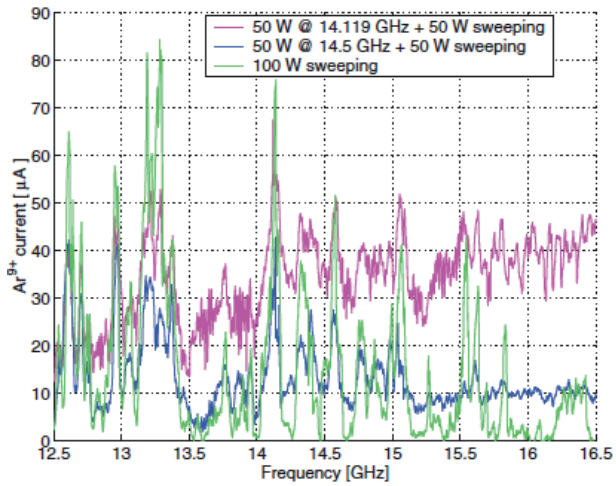


Fig. 11: (Colour online) The Ar^{9+} current as a function of frequency: a comparison between the frequency tuning effect and double-frequency heating.

for the present case, as described above), the current increases linearly with frequency (with respect to the 14.5 GHz fixed frequency).

In comparing the two techniques while keeping the ion source settings and the microwave power constant, note that the behaviour of the ion currents and the reflection coefficient remain almost the same with frequency. This result is useful to optimize the performance of an ECRIS further when double-frequency heating is used. In fact, single-frequency tuning can be used to find the optimum frequencies to be used for operating in the multiple heating mode. Note also that when operating with two generators set at the same frequency, the current level measured is different from the single-generator case. This is because when the two frequencies become equal, the current of each charge state and the power provided to the ion source are not stable. A possible explanation for this phenomenon might be overlapping of the two signals and the phase difference between the two waves. We plan to explore the use of two phase-locked generators in forthcoming experiments. The behaviour of argon ion beams has been explored for different power levels of the fixed frequency and the sweep frequency. In both cases, two fixed

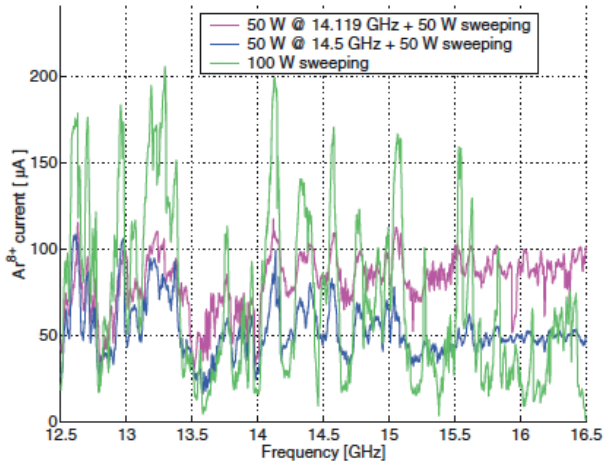


Fig. 12: (Colour online) The Ar^{8+} current as a function of frequency: a comparison between the frequency tuning effect and double-frequency heating.

frequencies were chosen and in this case the possibility of setting the fixed frequency to an optimized value, obtained by fine-tuning the frequency, allowed us to increase the beam intensity (Fig. 13).

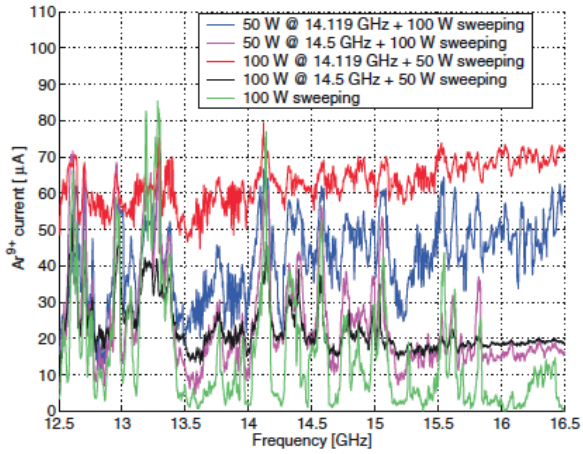


Fig. 13: (Colour online) The Ar^{9+} current as a function of frequency: differing power distributions between the two waves.

Figure 13 shows the evolution of the Ar^{9+} current when 150 W of total power is distributed between the two microwave feeds in different proportions, one of fixed frequency and the second swept from 12.5 GHz to 16.5 GHz. Again, the single-frequency sweep carried out at 100 W is included in the figure for comparison. For the optimum frequency mentioned previously (14.119 GHz), the current intensity is still comparable between the two methods even though the total power is 50% greater.

When the power associated with the fixed frequency is higher, all the analysed charge states are less sensitive to variation of the second frequency. This follows from the reduced dynamic range of the current, indicating a more stable plasma; in particular, when the fixed frequency is chosen to be at the optimum value (14.119 GHz). Furthermore, when one of the two frequencies is optimized, the average extracted current remains high with respect to an unoptimized frequency (such as 14.5 GHz), increasing with the second frequency (quite evident above 15 GHz). All these observations are more evident for

higher-charge states. With the ion source running at higher microwave power, it was possible to perform an experiment with an Ar^{11+} ion beam. The results are shown in Fig. 14.

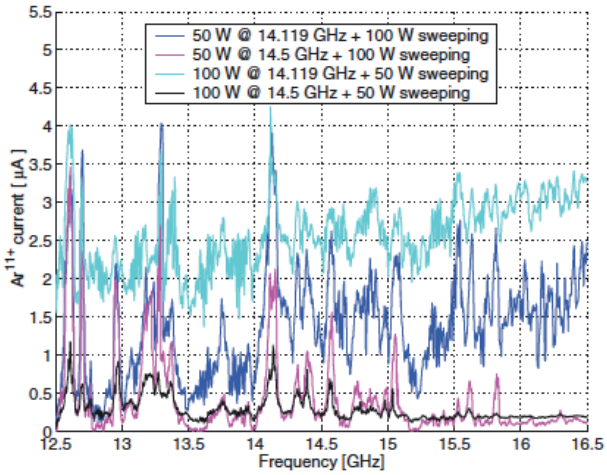


Fig. 14: (Colour online) The Ar^{11+} current as a function of frequency, showing different power distributions between the two waves.

The power distribution between the two waves is also important. Obviously, it is more effective to increase the power at the fixed frequency. Figure 14 shows once more that the choice of the optimum frequency, as obtained in the previous experiment on frequency tuning, is very important for double-frequency heating at higher power. This result is interesting with respect to the operation of new-generation ECRIS machines working at 28 GHz, using gyrotron microwave generators in combination with a second microwave generator of lower frequency, working in double-frequency mode.

6 Alternative heating method: Bernstein waves

ECR devices are density limited, because the electromagnetic waves cannot propagate beyond a certain density, called the cut-off density. An alternative to the classical ECR interaction is electrostatic wave heating, driven by Bernstein waves. Bernstein waves (BW) are very interesting for nuclear fusion devices [25], because the plasma heating occurs in the absence of any density cut-off. Electromagnetic Ordinary (O) or Extraordinary (X) modes can be externally launched into the plasma. O-mode waves may be coupled to an X mode in the O-mode cut-off layer, and then the X mode can be coupled to a BW mode at the upper hybrid resonance (UHR). This process is called O-X-B mode conversion and was described in 1973 by Preinhaelter and Kopecky [25].

It was shown that O-X-B mode conversion may be optimized by changing the O-mode insertion angle with respect to the external magnetic field direction: in this case, the O mode is completely converted into a slow X mode, which under certain conditions is in turn converted to BWs in the Upper Hybrid Resonance (UHR) layer [26]. The generated Bernstein waves travel inside the plasma until they are absorbed at the ECR or at the higher-order cyclotron harmonics. A possible design for a microwave injection system producing the O-X-B conversion is shown in Fig. 15: first, an O wave is launched from the outside, with an oblique angle of incidence obtained with a proper orientation of the parabolic mirrors. For an optimal launch angle, it is possible to obtain a correlated optimal parallel refractive index and then a coincidence of the O and X modes at the critical plasma density (cut-off). This means that both modes have the same phase and group velocities and the power is transferred without reflections. Once the X waves are generated, they propagate towards the UHR. Here, the X mode coincides with the electron Bernstein mode. In the linear description, the X waves are completely converted into EBWs.

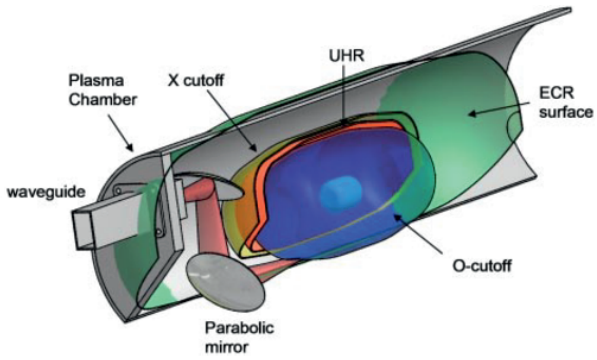


Fig. 15: A possible design for a microwave injection system producing O-X-B conversion: two parabolic mirrors are used to bring a focused microwave beam into the optimal region of the O cut-off layer.

This process is called X–B conversion. It should be noted that the O-X-B process can only take place if the plasma density is above the O-wave cut-off density.

O-X-B plasma heating and current driving with BW in an overdense plasma were demonstrated in the Stellarator WEGA, operating at the Max Planck Institute for Plasma Physics in Greifswald, Germany [27]. The heating of a plasma by means of EBW at particular frequencies enabled us to reach densities much larger than the cut-off ones. Evidence of EBW generation and absorption together with X-ray emissions due to high-energy electrons is shown in ray tracing simulations and CCD photographs in Ref. [29].

A plasma reactor operating at the Laboratori Nazionali del Sud of INFN, Catania, has been used as a test-bench for the investigation of innovative mechanisms of plasma ignition based on electrostatic waves (ES-W), obtained via the inner plasma EM-to-ES wave conversion. Evidence of Bernstein wave (BW) generation is shown in Ref. [28].

In particular, during the experiment, a microwave discharge ion source has been used: a plasma reactor consisting of a stainless-steel cylinder that is 24 cm long and 14 cm in diameter. A NdFeB permanent magnet system generates an off-resonance magnetic field along the plasma chamber axis (with a maximum of 0.1 T on axis).

Microwaves have been generated by using a TWT, which is able to generate microwaves from 3.2 to 4.9 GHz. The typical working frequency when using the TWT was 3.7478 GHz. The temperature and plasma density measurements have been carried out by using a movable Langmuir Probe (LP). The LP can host a small wire used as a local electromagnetic antenna, which can be connected to a spectrum analyser for the plasma spectral emission analysis. An Si-Pin and a HPGe X-ray detector have been used for the measurement of X-ray spectra in different plasma conditions. Both detectors are able to detect X rays with energies greater than about 1 keV. A CCD camera has been used to visualize the plasma structure within the chamber at different working frequencies, and at different microwave powers and pressures. A series of LP measurements has been carried out with the plasma reactor at the frequency of 2.45 GHz, when both under-resonance and off-resonance regions are present. In Fig. 16, it is evident that the electron density is drastically enhanced in regions where the condition $B < B_{\mathrm{ECR}}$ is satisfied.

This effect is observed for all of the different power values that we used; in particular, at 200 W, an electron density of about 1.5×10^{12} cm^3 has been measured, a value 20 times greater than the cut-off density. Note that the electron density is everywhere comparable or larger than the cut-off density ($n_c = 7.5 \times 10^{10}$ cm^3). In the same magnetic configuration, X-ray measurements have been carried out at 2.45 GHz and 3.7478 GHz. By increasing the pumping frequency, it is possible totally to remove the ECR, so that at 3.7478 GHz the entire plasma chamber is at under-resonance and EBW heating becomes the unique heating mechanism. Spectral temperatures measured in the two cases are shown in Fig. 17.

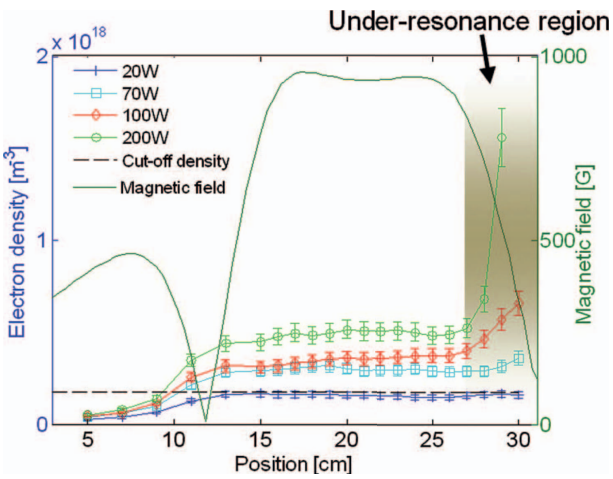


Fig. 16: (Colour online) Electron density and magnetic field profile at 1.5×10^{-4} mbar, and frequency 2.45 GHz, at different microwave powers. Microwave injection occurs at the right-hand side of the figure.

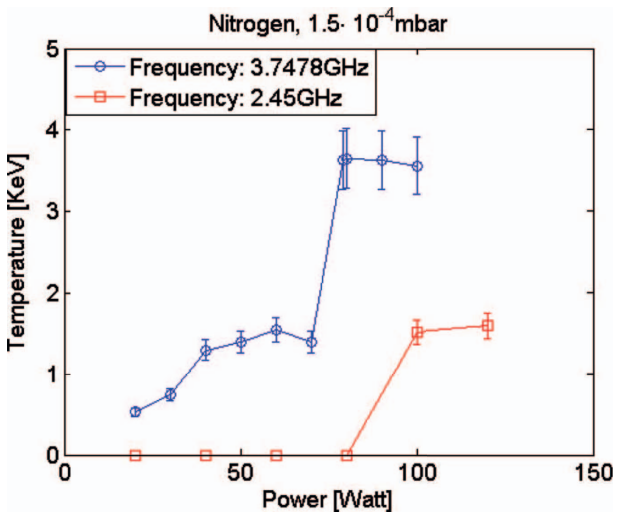


Fig. 17: (Colour online) Spectral temperatures. At a frequency of 2.45 GHz, ECR heating and EBW heating coexist; at 3.7478 GHz, EBW heating is dominant.

At the frequency of 3.7478 GHz, the spectral temperature is significantly larger (up to 4 keV) than at 2.45 GHz. Around 80 W, it is possible to identify a threshold for which the temperature slope increases suddenly and becomes steeper for both configurations. The end-point energy is about 10 times the value of the spectral temperature. The spectral temperature decreases slightly with the atomic mass of the gas used and when the pressure is increased. Further results and discussion can be found in Ref. [29]. The results are interpreted through the Bernstein wave heating theory, and are very promising for future high-intensity multicharged ion sources that can, therefore, be based on the results described here, by employing a simplified magnetic configuration with respect to typical minimum-B ECR ion sources.

7 Conclusion

Microwave coupling to ECR ion sources plays a fundamental role in enhancing their performance in terms of current extracted and average charge state produced. Over the past decade, new heating methods

have been studied to overcome the technological limitations due to the magnetic field and frequency scaling, to achieve the production of milliampere levels of HCI, as requested for the new accelerating facilities all over the world.

References

[1] S. Gammino, *High Energy Phys., Nucl. Phys.* **31** (2007) 137.

[2] A.G. Drentje, *Rev. Sci. Instrum.* **74** (2003) 2631.

[3] Y. Kawai, G.D. Alton, O. Tarvainen, P. Suominen, and H. Koivisto, *Rev. Sci. Instrum.* **77** (2006) 03A331.

[4] Z.Q. Xie and C.M. Lyneis, *Rev. Sci. Instrum.* **66** (1995) 4218.

[5] R.C. Vondrasek, R.H. Scott, R.C. Pardo, and H. Koivisto, Operational improvements of the Argonne ECR sources, Proc. 15th Int. Workshop ECR Ion Sources, Jyvaskyla, Finland, 12–14 June 2002.

[6] G.D. Alton *et al.*, *Rev. Sci. Instrum.* **69** (1998) 2305.

[7] L. Celona, G. Ciavola, F. Consoli, S. Gammino, F. Maimone *et al.*, Observations of the frequency tuning effect in the 14 GHz CAPRICE ion source, *Rev. Sci. Instrum.* **79** (2008) 023305.

[8] S. Gammino and G. Ciavola, ECR ion sources and scaling laws, Proc. 14th Int. Conf. Cycl. Capetown, South Africa, 1995, p. 377.

[9] D. Hitz *et al.*, The comparison of 18 and 28 GHz behavior of SERSE: some conclusions, Proc. 15th Int. Workshop ECR Ion Sources. Jyvaskyla, Finland, 2002, p. 100.

[10] G.D. Alton, F.W. Meyer, Y. Liu, J.R. Beene, and D. Tucker. Enhancing the performances of traditional electron cyclotron resonance ion sources with multiple-discrete-frequency microwave radiation, *Rev. Sci. Instrum.* **69** (1998) 2305.

[11] G.D. Alton, *Computational studies for a multiple-frequency electron cyclotron resonance ion source.Rev. Sci. Instrum.* **67** (1996) 989.

[12] Z.Q. Xie, Production of highly charged ion beams from electron cyclotron resonance ion sources (invited), *Rev. Sci. Instrum.* **69** (1998) 625.

[13] O. Tarvainen, Y. Kawai, G.D. Alton, P. Suominen, and H. Koivisto, Electron heating with broadband microwave radiation: a new method for improving the performances of conventional B-minimum electron cyclotron resonance ion sources, *Nucl. Instrum. Methods Phys. Res. B* **261** (2007) 1044.

[14] Y. Kawai, G.D. Alton, O. Tarvainen, P. Suominen, and H. Koivisto, Effect of broadband microwave radiation on the performance of a conventional B-minimum geometry electron cyclotron resonance ion source, *Rev. Sci. Instrum.* **77** (2006) 03A331.

[15] G.D. Alton and D.N. Smithe, *Rev. Sci. Instrum.* **65** (1994) 775.

[16] G.D. Alton and D.N. Smithe, *Physica Scripta* **T71** (1996) 66.

[17] G.D. Alton, Proc. 14th Int. Conference on Cyclotrons and their Applications (Capetown, South Africa, 8–13 October 1995), Ed. J.C. Cornell, World Scientific, Singapore, p. 362.

[18] G.D. Alton and Y. Liu, Enhancing the performances of ECR ion source, Proc. 1999 Particle Accelerator Conference, New York (1999).

[19] L. Celona *et al.*, Analysis of the SERSE ion output by using klystron-based or TWT-based microwave generators, Proc. AIP Conf., vol. 749 (2005), p. 99.

[20] S. Gammino, G. Ciavola, and L. Celona. *Nucl. Instrum. Methods Phys. Res. A*, **491** (2002) 342.

[21] R. Vondrasek *et al.*, AIP Conf. Proc., vol. 749 (2005), p. 31.

[22] D. Hitz, A. Girard, K. Serebrennikov, G. Melin, D. Cormier, J.M. Mathonnet, J. Chartier, L. Sun, J.P. Briand, and M. Benhachoum, *Rev. Sci. Instrum.* **75** (2004) 1403.

[23] J. Mäder *et al.*, *Rev. Sci. Instrum.* **81** (2010) 174.

[24] D. Mascali *et al.*, *Rev. Sci. Instrum.* **81** (2010) 02A334.

[25] H.P. Laqua *et al.*, *Plasma Phys. Control. Fusion* **41** (1999) A273.

[26] N.A. Krall and A.W. Trivelpiece, *Principles of Plasma Physics* (San Francisco Press, San Francisco, CA, 1986).

[27] Y.Y. Podoba, H.P. Laqua, G.B. Warr, M. Schubert, M. Otte, S. Marsen, F. Wagner, and E. Holzhauer, *Phys. Rev. Lett.* **98** (2007) 255003.

[28] D. Mascali, L. Celona, S. Gammino, R. Miracoli, G. Castro, N. Gambino, and G. Ciavola, Electrostatic wave heating and possible formation of self-generated high electric fields in a magnetized plasma, *Nucl. Instrum. Methods Phys. Res. A.* **653** (2011) 11.

[29] G. Castro, D. Mascali, F. P. Romano, L. Celona, S. Gammino *et al.*, Comparison between off-resonance and electron Bernstein waves heating regime in a microwave discharge ion source, *Rev. Sci. Instrum.* **83** (2012) 02B501.

Vacuum Technology for Ion Sources

P. Chiggiato
CERN, Geneva, Switzerland

Abstract
The basic notions of vacuum technology for ion sources are presented, with emphasis on pressure profile calculation and choice of pumping technique. A Monte Carlo code (Molflow+) for the evaluation of conductances and the vacuum–electrical analogy for the calculation of time-dependent pressure variations are introduced. The specific case of the Linac4 H^- source is reviewed.

1 Introduction

Ion sources are among the very few places in accelerators where gas is intentionally injected during operation. Therefore, the first goal of the vacuum system is to remove the injected gas before it reaches the rest of the acceleration chain, where a lower gas density is required. The injected and then evacuated gas flow is in general very high. Such a requirement, together with the long uninterrupted operational time, imposes a drastic constraint on the selection of the vacuum pumps and the way they are operated.

Mechanical pumps are the obvious choice close to the injection points, where the gas throughput is the highest. Such pumps evacuate the gas molecules outside the ion source. However, as soon as most of the injected gas is evacuated and lower pressures must be reached, the vacuum system should rely on capture pumps because they are in general more reliable than mechanical pumps.

The evaluation of the dynamic and static pressure profiles should be undertaken in parallel with the mechanical design. Basic calculations may be carried out for simple geometries. However, the complexity of the mechanical design requires advanced computation. The molecular trajectories in the vacuum system are simulated by Monte Carlo codes, and pressures are evaluated by counting the molecular collisions with virtual surfaces. The analogy between vacuum systems and electrical networks, in conjunction with the Monte Carlo simulations, may be used to obtain time-dependent pressure profiles.

In ion sources, pressure is measured by applying different principles. In the highest pressure range, near the injection points, mechanical and thermal conductivity effects are still significant in the gas phase. In this case, thermal and capacitance pressure gauges are employed. For a more rarefied gas, pressure readings are obtained by measuring gas-ion currents intentionally produced by electron impact ionization.

This chapter focuses mainly on high-flow gas pumping and the calculation of pressure profiles. The first pages are dedicated to the presentation of the basic knowledge and terminology of vacuum technology, especially conductance and pumping speed. Then the main pumping mechanisms are explained, with the focus on operational details that are relevant to ion sources. Pressure measurement is not included because, in general, it is not considered as a critical issue in ion sources. A comprehensive introduction to vacuum technology may be found in the books listed in the references.

2 Basic notions of vacuum technology

In the framework of vacuum technology for particle accelerators, a rarefied gas in equilibrium is always described by the ideal-gas equation of state:

$$PV = N_{\text{moles}}RT, \tag{1}$$

where P, T and V are the gas pressure, temperature and volume, respectively, and R the ideal-gas constant (8.314 J K^{-1} mol^{-1} in SI units). From statistical physics considerations, Eq. (1) may be rewritten in terms of the total number of molecules N in the gas:

$$PV = Nk_{\text{B}}T, \tag{2}$$

where k_{B} is the Boltzmann constant (1.38×10^{-23} J K^{-1} in SI units).

In the International System of Units (SI), the pressure is reported in pascals (Pa): 1 Pa is equivalent to the pressure exerted by one newton on a square metre (1 N m^{-2}). Other units are regularly used in vacuum technology, in particular bar and its submultiple the millibar (mbar). The Torr is still occasionally used, mostly in the USA; it is equivalent to the pressure exerted by a 1 mm high column of Hg. The conversion values between the common pressure units are shown in Table 1.

Table 1: Conversion values for the most common pressure units of vacuum technology.

	Pa	**bar**	**atm**	**Torr**
1 Pa	1	10^{-5}	9.87×10^{-6}	7.5×10^{-3}
1 bar	10^{5}	1	0.987	750.06
1 mbar	10^{2}	10^{-3}	0.967×10^{-3}	0.75
1 atm	1.013×10^{5}	1.013	1	760
1 Torr	133.32	1.33×10^{-3}	1.32×10^{-3}	1

The number density ($n = N/V$) of gas molecules is easily calculated by Eq. (2). Some typical examples are collected in Table 2.

Table 2: Typical number density at room temperature and helium boiling point.

	Pressure (Pa)	**293 K (molecules cm^{-3})**	**4.3 K (molecules cm^{-3})**
Atmospheric pressure at sea level	1.013×10^{5}	2.5×10^{19}	1.7×10^{21}
Typical plasma chambers	1	2.5×10^{14}	1.7×10^{16}
Linac pressure upper limit	10^{-5}	2.5×10^{9}	1.7×10^{11}
Lowest pressure ever measured at room temperature [1]	10^{-12}	250	1.7×10^{4}

In addition to the number of molecules or moles, gas quantities are expressed as pressure–volume (PV) values at a given temperature. Pressure–volume quantities are converted to number of molecules by dividing them by $k_B T$ as given in the equation of state. For example, 1 Pa m^3 at 293 K is equivalent to

$$N = \frac{1\ [\mathrm{Pa\ m^3}]}{1.38 \times 10^{-23}\ [\mathrm{J\ K^{-1}}]\ \times 293\ [\mathrm{K}]} = 2.47\ \times 10^{20}\ \text{molecules.}$$

In the same way, at 293 K, 1 Torr litre = 3.3 × 10^{19} molecules and 1 mbar litre = 2.47 × 10^{19} molecules. In vacuum technology, pressure–volume units are used most of the time to report gas quantities and flow rates. For example, the gas flow rate is generally reported in mbar litre s^{-1} or Torr litre s^{-1}.

In vacuum systems, pressures span several orders of magnitude (Table 3). Degrees of vacuum are defined by upper and lower pressure boundaries. Different degrees of vacuum are characterized by different pumping technologies, pressure measurement, materials and surface treatments. Ion sources operate in the degrees of vacuum that are usually called medium and high vacuum.

Table 3: Degrees of vacuum and their pressure boundaries.

	Pressure boundaries (mbar)	**Pressure boundaries (Pa)**
Low vacuum (LV)	1000–1	10^5–10^2
Medium vacuum (MV)	1–10^{-3}	10^2–10^{-1}
High vacuum (HV)	10^{-3}–10^{-9}	10^{-1}–10^{-7}
Ultra-high vacuum (UHV)	10^{-9}–10^{-12}	10^{-7}–10^{-10}
Extreme vacuum (XHV)	<10^{-12}	<10^{-10}

2.1 Gas kinetics

The kinetics of ideal-gas molecules is described by the Maxwell–Boltzmann theory [2]. For an isotropic gas, the model provides the probabilistic distribution of the molecular speed magnitudes. The mean speed of molecules $\langle v \rangle$, i.e. the mathematical average of the speed distribution, is given by

$$\langle v \rangle = \sqrt{\frac{8 k_B T}{\pi m}} = \sqrt{\frac{8RT}{\pi M}}, \tag{3}$$

where m is the mass of the molecule and M is the molar mass. The unit of both masses is kg in SI. Typical mean speed values are shown in Table 4.

Table 4: Mean speed of gas molecules of different mass at room temperature and helium boiling point.

	H_2	**He**	**CH_4**	**N_2**	**Ar**
$\langle v \rangle$ at 293 K (m s^{-1})	1761	1244	622	470	394
$\langle v \rangle$ at 4.3 K (m s^{-1})	213	151	75	57	48

Another important result of the Maxwell–Boltzmann theory is the calculation of the molecular impingement rate φ on a surface, i.e. the rate at which gas molecules collide with a unit surface area exposed to the gas. Assuming that the density of molecules all over the volume is uniform, it can be shown [3] that

$$\varphi = \frac{1}{4} n \langle v \rangle, \tag{4}$$

and using Eq. (3) for the mean speed as obtained by the Maxwell–Boltzmann theory,

$$\varphi = \frac{1}{4} n \sqrt{\frac{8 k_{\mathrm{B}} T}{\pi m}}. \tag{5}$$

Numerical values in terms of P, T and molar mass are obtained by Eq. (6); some selected values are reported in Table 5:

$$\varphi\ [\mathrm{cm^{-2}\ s^{-1}}] = 2.635 \times 10^{22} \frac{P\ [\mathrm{mbar}]}{\sqrt{M[\mathrm{g}]\ T[\mathrm{K}]}}. \tag{6}$$

Table 5: Molecular impingement rates at room temperature for H_2, N_2 and Ar at some selected pressures.

Gas	Pressure (mbar)	Impingement rate ($\mathrm{cm^{-2}\ s^{-1}}$)
H_2	10^{-3}	1.1×10^{18}
	10^{-8}	1.1×10^{13}
	10^{-14}	1.1×10^{7}
N_2	10^{-3}	2.9×10^{17}
	10^{-8}	2.9×10^{12}
Ar	10^{-3}	2.4×10^{17}
	10^{-8}	2.4×10^{12}

2.2 Mean free path and Knudsen number

In any physically limited vacuum system, molecules collide between each other and with the walls of the vacuum envelope/container. In the first case, the average length of the molecular path between two points of consecutive collisions, i.e. the mean free path $\bar{\lambda}$, is inversely proportional to the number density $n = P/k_{\mathrm{B}}T$ and the collision cross-section σ_{c} [4], as given by

$$\bar{\lambda} = \frac{1}{\sqrt{2} n \sigma_{\mathrm{c}}}. \tag{7}$$

For elastic collisions between hard spheres, Eq. (7) can be written in terms of the molecular diameter δ as

$$\bar{\lambda} = \frac{1}{\sqrt{2} \pi n \delta^2} = \frac{k_{\mathrm{B}} T}{\sqrt{2} \pi P \delta^2}. \tag{8}$$

The collision cross-sections for common gas species in vacuum systems are listed in Table 6.

Table 6: Elastic collision cross-sections for five different molecules.

Gas	H_2	He	N_2	O_2	CO_2
σ_c (nm^2)	0.27	0.27	0.43	0.40	0.52

For numerical purposes, Eq. (8) can be rewritten for a specific gas as a function of temperature and pressure. For example, for H_2, we obtain

$$\bar{\lambda}_{H_2}[m] = 4.3 \times 10^{-5} \frac{T[K]}{P[Pa]} \tag{9}$$

The values of the mean free path for H_2 at room temperature are shown in Fig. 1 as a function of the gas pressure.

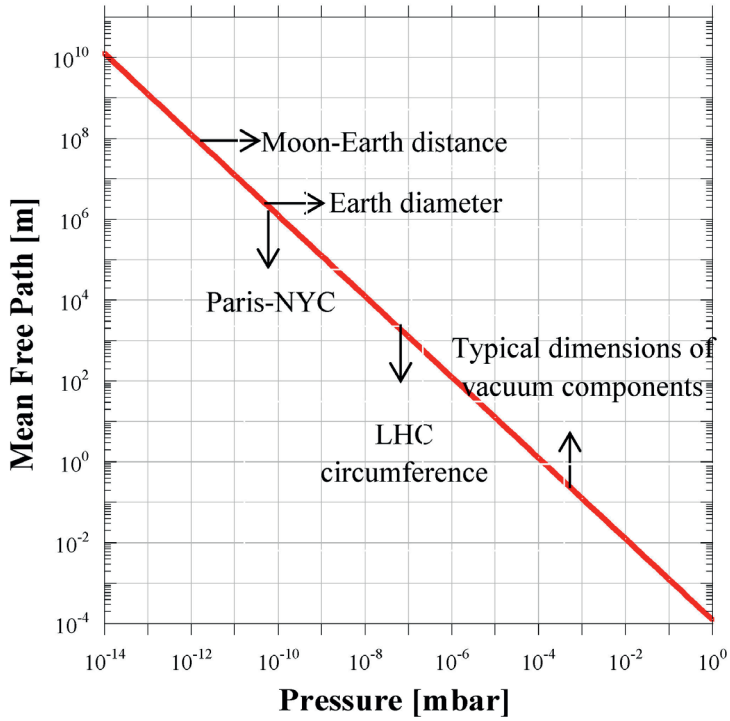


Fig. 1: Mean free path of H_2 molecules at 293 K

When the mean free path is of the order of typical dimensions of the vacuum vessel, for example, the diameter of cylindrical beam pipes, molecular collisions with the wall of the vacuum envelope become preponderant. For even longer $\bar{\lambda}$, the gas dynamics is dominated by molecule–wall collisions; intermolecular interactions lose any effect on the gas displacement.

The adimensional Knudsen number K_n translates into numerical values the considerations expressed above. It is defined as the ratio between the mean free path and a characteristic dimension of a vacuum system (D):

$$K_n = \frac{\bar{\lambda}}{D} \tag{10}$$

The values of K_n delimit three gas dynamic regimes as reported in Table 7.

Table 7: Gas dynamic regimes defined by the Knudsen number.

K_n range	Regime	Description
K_n >0.5	Free molecular flow	Molecule–wall collisions dominate
K_n <0.01	Continuous (viscous) flow	Gas dynamics dominated by intermolecular collisions
0.5 < K_n < 0.01	Transitional flow	Transition between molecular and viscous flow

Typical beam pipe diameters are of the order of 10 cm. Therefore, the free molecular regime is obtained for pressures in the low 10^{-3} mbar range or lower. Except for the ion source's plasma chambers, the vacuum systems of accelerators operate in the free molecular regime; only this vacuum regime is considered in this chapter. Excellent introductions to vacuum systems in viscous and transitional regimes can be found in Refs. [5] and [6].

The free molecular flow regime characterizes and determines the pumping and pressure reading mechanisms that can be used in particle accelerators. Pumps and instruments must act on single molecules since there is no interaction between molecules. Collective phenomena such as pressure waves and suction do not influence gas dynamics in free molecular flow.

2.3 Conductance in free molecular flow

In the free molecular regime, the net gas flow Q between two points of a vacuum system is proportional to the pressure difference ($P_1 - P_2$) at the same points:

$$Q = C(P_1 - P_2), \tag{11}$$

where C is called the gas conductance of the vacuum system between the two points. In the free molecular regime, the conductance does not depend on pressure. It depends only on the mean molecular speed and the vacuum system geometry. If the gas flow units are expressed in terms of pressure–volume (e.g. mbar litre s^{-1} or Pa m^3 s^{-1}), the conductance is reported as volume per unit time (i.e. l s^{-1} or m^3 s^{-1}, from here onward litre are reported with the letter l).

The conductance is easily calculated for the simplest geometry, i.e. a small wall slot of surface area A and infinitesimal thickness dividing two volumes of the same vacuum system (see Fig. 2).

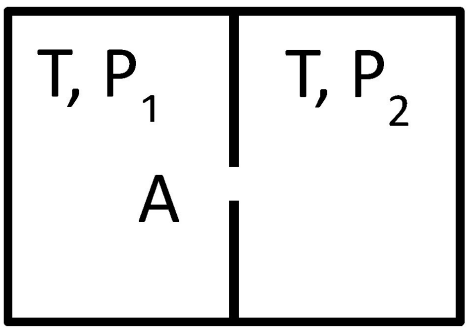


Fig. 2: Schematic drawing of two volumes communicating through a thin and small wall slot

The net flow of molecules from one volume to the other may be calculated by the molecular impingement rate given by Eq. (4). The number of molecules of volume 1 that go into volume 2 ($\varphi_{1\to2}$) is

$$\varphi_{1\to2} = \frac{1}{4} A n_1 \langle v \rangle$$

while that from volume 2 to volume 1 is

$$\varphi_{2\to1} = \frac{1}{4} A n_2 \langle v \rangle.$$

The net molecular flow is given by the difference between the two contributions,

$$\varphi_{1\to2} - \varphi_{2\to1} = \frac{1}{4} A (n_1 - n_2) \langle v \rangle,$$

and from Eqs. (2) and (3)

$$\varphi_{1\to2} - \varphi_{2\to1} = \frac{1}{4} A \frac{\langle v \rangle}{k_B T} (P_1 - P_2).$$

Multiplying both terms of the equality by $k_B T$ and applying Eq. (2), the gas flow in pressure–volume units is obtained as

$$Q = \frac{1}{4} A \langle v \rangle (P_1 - P_2). \tag{12}$$

Comparing Eqs. (11) and (12), it comes out that the conductance of the wall slot is proportional to the surface area of the slot and the mean speed of the molecules:

$$C = \frac{1}{4} A \langle v \rangle \propto \sqrt{\frac{T}{m}}. \tag{13}$$

From Eqs. (3) and (13), it can be shown that the conductance of the wall slots is inversely proportional to the square root of the molecular mass. Therefore, for equal pressure drop, the gas flow of H_2 is the highest. Finally, for gas molecules of different masses, the conductance scales as the square root of the inverse mass ratio:

$$\frac{C_1}{C_2} = \sqrt{\frac{m_2}{m_1}}. \tag{14}$$

As an example, the conductance for N_2 is $\sqrt{2/28} = 0.27$ times that for H_2, namely 3.7 times lower. Table 8 collects conductance values, for an orifice, per unit surface area (C') at room temperature for common gas species.

Table 8: Unit surface area conductances for common gas species in two different units.

Gas	H_2	He	CH_4	H_2O	N_2	Ar
C' at 293 K ($m^3\ s^{-1}\ m^{-2}$)	440.25	311	155.5	146.7	117.5	98.5
C' at 293 K ($l\ s^{-1}\ cm^{-2}$)	44	31.1	15.5	14.7	11.75	9.85

For more complex geometries than wall slots, the transmission probability τ is introduced. If two vessels, at the same temperature, are connected by a duct (see Fig. 3 for symbols), the gas flow from vessel 1 to vessel 2 ($\varphi_{1\to2}$) is calculated multiplying the number of molecules impinging on the entrance section of the duct by the probability $\tau_{1\to2}$ for a molecule to be transmitted into vessel 2 without coming back to vessel 1:

$$\varphi_{1\to2} = \frac{1}{4} A_1 n_1 \langle v \rangle \tau_{1\to2}. \tag{15}$$

Fig. 3: Schematic drawing of two vessels connected by a duct

Similarly, the gas flow from vessel 2 to vessel 1 is written as

$$\varphi_{2\to1} = \frac{1}{4} A_2 n_2 \langle v \rangle \, \tau_{2\to1}. \tag{16}$$

In the absence of net flow, $\varphi_{1\to2} = \varphi_{2\to1}$ and $n_1 = n_2$, then

$$A_1 \tau_{1\to2} = A_2 \tau_{2\to1}. \tag{17}$$

When $n_1 \neq n_2$, a net flow is set up. It can be calculated by taking into account Eqs. (15)–(17) and (2):

$$\varphi_{1\to2} - \varphi_{2\to1} = \frac{1}{4} A_1 \langle v \rangle \tau_{1\to2} \frac{(P_1 - P_2)}{k_\mathrm{B} T}. \tag{18}$$

In pressure–volume units, Eq. (18) becomes

$$Q = \frac{1}{4} A_1 \langle v \rangle \tau_{1\to2} (P_1 - P_2) = C' A_1 \tau_{1\to2} (P_1 - P_2), \tag{19}$$

where, as already mentioned, C' is the conductance of the unit surface area wall slot. Comparing Eqs. (11) and (19), it can be seen that the conductance of the connecting duct is equal to the conductance of the duct entrance in vessel 1, considered as a wall slot, multiplied by the molecular transmission probability from vessel 1 to vessel 2:

$$C = C' A_1 \tau_{1\to2}. \tag{20}$$

2.3.1 Evaluation of the transmission probability

The transmission probabilities depend only on vacuum component geometry. They may be calculated analytically for simple geometry by means of relatively complex integral equations (Clausing equations, see Ref. [7]). Approximate formulas are reported in many vacuum-technology books, for example in Ref. [8].

For the very common case of tubes of uniform circular cross-section of length L and radius R, the Santeler equation [9] gives the transmission probability with less than 0.7% error:

$$\tau = \tau_{1\to2} = \tau_{2\to1} = \frac{1}{1 + \frac{3L}{8R}\left(1 + \frac{1}{3\left(1 + \frac{L}{7R}\right)}\right)}. \tag{21}$$

For long tubes, i.e. $L/R \gg 1$, Eq. (21) can be simplified to

$$\tau \approx \frac{1}{1 + \frac{3L}{8R}} \approx \frac{8}{3}\frac{R}{L}. \tag{22}$$

Equation (22) combined with Eq. (20) gives the conductance of long circular pipes; it is one of the most used equations in vacuum technology. For N_2 it may be written as

$$C \approx 11.75 \times \frac{\pi D^2}{4} \times \frac{4D}{3L} = 12.3 \frac{D^3}{L} \ [\mathrm{l\,s^{-1}}] \ \ ([D] \text{ and } [L] = \mathrm{cm}). \tag{23}$$

As a result, the conductance of a tube is strongly dependent on its diameter.

The Santeler equation and its approximation for long tubes are plotted in Fig. 4. The latter can be applied with errors less than 10% for $L/R \gg 20$. As shown in the figure, the transmission probability is about 0.5 for circular tubes for which the diameter is equal to their length. In other words, the conductance of such tubes is half that of their entrance surface.

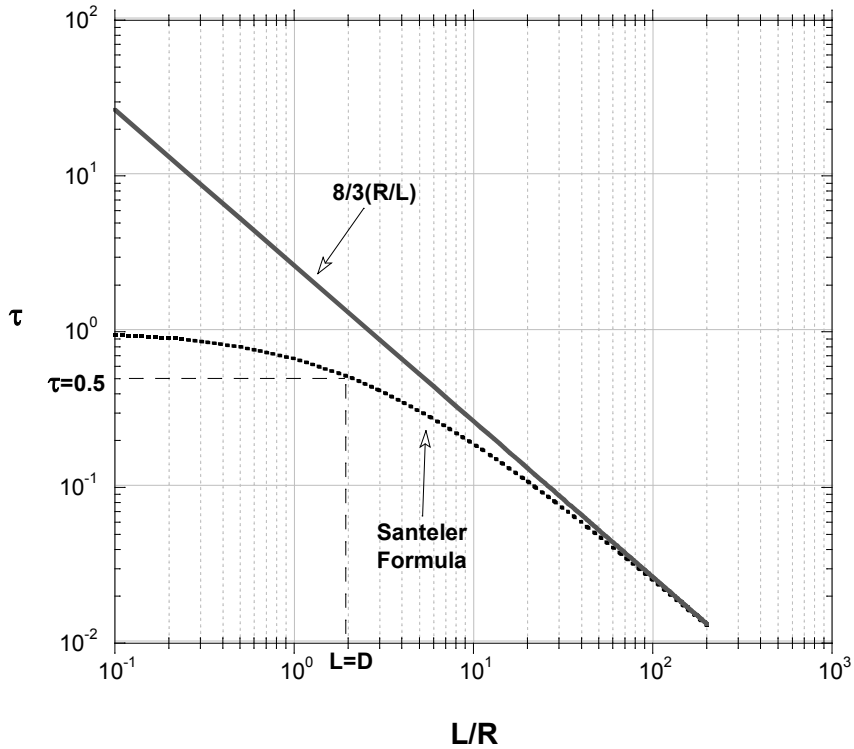


Fig. 4: Transmission probability of tubes of uniform circular cross-section calculated by the Santeler equation and its approximation for high *L/R*.

Conductances of more complicated components are calculated by test-particle Monte Carlo methods (TPMC). The components are modelled in three dimensions; the TPMC codes generate molecules at the entrance of the component pointing in 'random' directions according to the cosine distribution. When the molecules impinge on the internal wall of the component, they are re-emitted again randomly. The program follows the molecular traces until they reach the exit of the component. The transmission probability is given by the ratio between the number of 'escaped' and 'injected' molecules [10]. Many simulated molecules are needed to reduce the statistical scattering.

The reference TPMC software at CERN is MolFlow+ [11]. This powerful tool for high-vacuum applications imports the 3D drawing of the vacuum components and generates 'random' molecules on any surface of interest. Figure 5 shows the model and the molecular tracks for the extraction chamber of the Linac4's ion source.

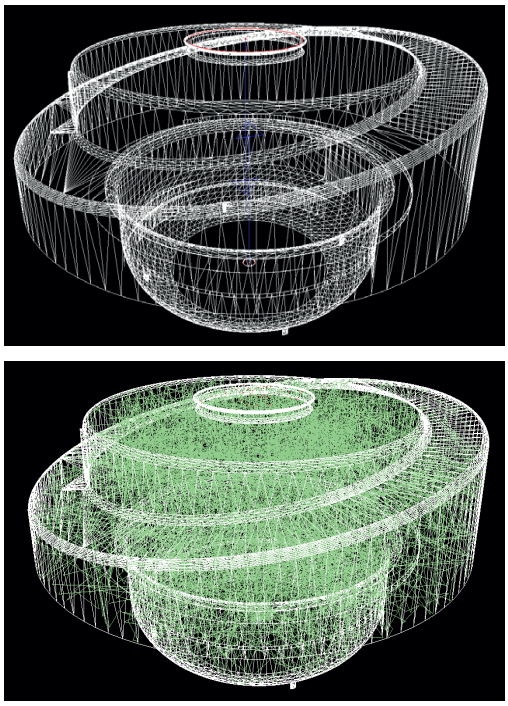


Fig. 5: Meshes used for the Monte Carlo simulation of the extraction chamber of Linac4. The second picture shows molecular tracks (in green) generated by the MolFlow+ code.

2.3.2 Combination of conductances

Elementary vacuum components are installed either in series, i.e. traversed by the same net gas flow, or in parallel, i.e. equal pressures at the extremities (Fig. 6).

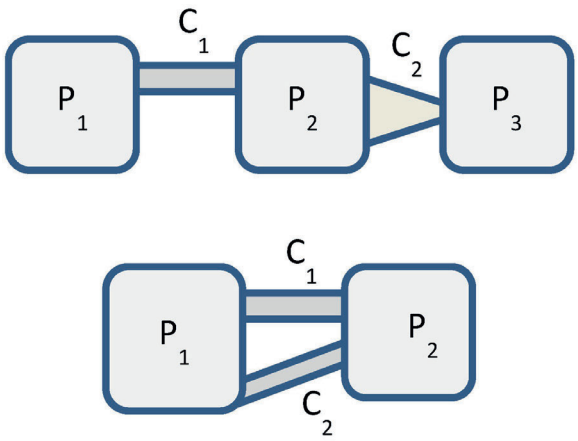


Fig. 6: Schematic drawings of components installed in series (top) and in parallel (bottom)

With reference to Fig. 6, the net gas flow in the two components connected *in series* is given by

$$Q = C_1(P_1 - P_2), \quad Q = C_2(P_2 - P_3). \tag{24}$$

A total conductance C_{TOT} equivalent to C_1 and C_2 is introduced in such a way that

$$Q = C_{\mathrm{TOT}}(P_1 - P_3). \tag{25}$$

Combining Eqs. (24) and (25), C_{TOT} is easily calculated as

$$\frac{1}{C_{\mathrm{TOT}}} = \frac{1}{C_1} + \frac{1}{C_2}. \tag{26}$$

In general, for N components in series:

$$\frac{1}{C_{\mathrm{TOT}}} = \sum_{1}^{N} \frac{1}{C_i}. \tag{27}$$

In the same way, it can be shown that for N components installed *in parallel*, the total conductance is the sum of the conductances of all components:

$$C_{\mathrm{TOT}} = \sum_{1}^{N} C_i. \tag{28}$$

2.4 Pumping speed

In vacuum technology, a pump is any 'object' that removes gas molecules from the gas phase. A vacuum pump is characterized by its pumping speed S, which is defined as the ratio between the pumped gas flow Q_{P} (pump throughput) and the pump inlet pressure P:

$$S = \frac{Q_{\mathrm{P}}}{P}. \tag{29}$$

The pumping speed unit is volume over time, thus the same unit as conductance. In high vacuum, S is in general expressed in l s^{-1} or m^3 s^{-1}; in low and medium vacuum, m^3 min^{-1} is used. In a more general way, S can be defined as the derivative of the pump throughput with respect to the pump inlet pressure:

$$S = \frac{\partial Q_{\mathrm{P}}}{\partial P}. \tag{30}$$

The pump throughput can be written as the gas flow through the cross-section of the pump inlet (surface area A_{P}) multiplied by the capture probability σ, i.e. the probability for a molecule that enters the pump to be definitely removed and never more to reappear in the gas phase of the vacuum system (see Eq. (31)). In the literature, σ is also called the Ho coefficient:

$$Q_{\mathrm{P}} = \varphi A_{\mathrm{P}} \sigma = \frac{1}{4} A_P n \langle v \rangle \sigma. \tag{31}$$

Considering Eqs. (13) and (2), it turns out that

$$Q_{\mathrm{P}} = A_{\mathrm{P}} C' n \sigma = A_{\mathrm{P}} C' \sigma \frac{P}{k_{\mathrm{B}} T}.$$

From the definition of the pumping speed and converting the throughput into pressure–volume units:

$$S = A_{\mathrm{P}} C' \sigma. \tag{32}$$

Therefore, the pumping speed is equal to the conductance of the pump inlet cross-section multiplied by the capture probability. The maximum theoretical pumping speed of any pump is obtained for σ= 1 and it is equal to the conductance of the pump inlet cross-section. Table 9 reports some values of the

maximum pumping speed of lump pumps for typical diameters of the pump inlet. Because S depends on C', and so on the inverse of the square root of the molecular mass, the maximum theoretical pumping speed is that for H_2.

The pumping speed given by the suppliers is called the *nominal pumping speed*; it refers to the pump inlet. The *effective pumping speed* S_{eff} is that acting directly in the vacuum vessel of interest. The effective pumping speed is lower than the nominal pumping speed owing to gas flow restrictions interposed between the pump and the vessel.

Table 9: Maximum pumping speeds in l s^{-1} for different diameters of circular pump inlets.

ID (mm)	H_2	N_2	Ar
36	448	120	100
63	1371	367	307
100	3456	924	773
150	7775	2079	1739

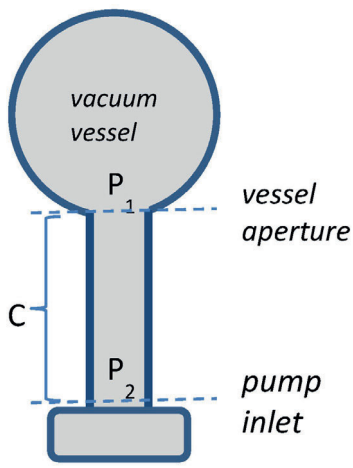


Fig. 7: Schematic drawing of a gas flow restriction of conductance C interposed between a pump of pumping speed S and a vacuum vessel.

The effective pumping speed is calculated considering the gas flow from the vessel and the pump. Taking into account Eqs. (11) and (29), with reference to Fig. 7, one obtains

$$Q = C_1(P_1 - P_2) = SP_2 = S_{\text{eff}}P_1$$

and so

$$\frac{1}{S_{\text{eff}}} = \frac{1}{S} + \frac{1}{C}. \tag{33}$$

As a result, for $C \ll S$, one finds $S_{\mathrm{eff}} \approx C$. In other words, the effective pumping speed does not depend on the installed pump if the conductance of the interposed connection is very low. This conclusion is of primary importance in the design of efficient vacuum systems.

2.5 Outgassing

Several gas sources contribute to the total gas load in a vacuum system. In addition to intentional gas injections and air leaks, outgassing of materials plays a crucial role in the global gas balance. Materials release gas molecules that are both adsorbed onto their surfaces and dissolved in their bulk [12]. The outgassing is distributed uniformly all over the vacuum exposed surfaces; its rate is in general reported for unit surface area (q).

The outgassing rate is reduced by dedicated surface treatments. For metals, in general, several surface treatments are available, each aiming at removing a specific set of contaminants. Gross contamination and the sorption layer (hydrocarbons, Cl compounds, silicone greases, etc.) are removed by solvent and detergent cleaning. Solvent molecules interact with contaminants and transport them away from the surface by diffusion. They are in general quite selective: perchloroethylene (C_2Cl_4) has a wide spectrum of applications, while hydrofluorocarbon (HFC) has a more restricted action. Detergent molecules are dissolved in water; they are based on an aggregate of surfactant (surface-acting agent) molecules, called a micelle, with hydrophilic and lipophilic extremities; they allow organic molecules and water to combine. Detergents are less selective than solvents. Thick oxide layers (1–10 nm thick) are removed by chemical pickling. The damaged skin of metals – full of dislocations, voids and impurities – is removed by acid etching or electropolishing.

For all materials, after thorough surface treatment, water vapour dominates the outgassing process. Its outgassing rate is inversely proportional to the pumping time (t) for smooth metals. The following empirical relationship is in general applied:

$$q_{\mathrm{H_2O}} \approx \frac{3 \times 10^{-9}}{t\,[\mathrm{h}]} \left[\frac{\mathrm{mbar\,l}}{\mathrm{s\,cm^2}}\right]. \tag{34}$$

For organic materials, in particular polymers, in the first phase of pumping, the water vapour outgassing rate is inversely proportional to the square root of the pumping time $q_{\mathrm{H_2O}} \propto 1/\sqrt{t}$, the absolute value being strongly dependent on the nature of the material. The square root dependence indicates that water molecule diffusion in the bulk is the leading process.

The water vapour outgassing is reduced by *in situ* heating in vacuum (bake-out). Such a thermal treatment is very effective for metals if it is carried out for at least 12 h at temperatures higher than 120°C. After bake-out, H_2 becomes the main outgassed species for metals; the outgassing rate can be regarded as constant at room temperature. It can be reduced by heat treatment at higher temperature either *in situ* or *ex situ* in a vacuum furnace [12].

As shown in Table 10, the value of outgassing rate spans several orders of magnitude. The right choice of materials and treatments is essential in the design of vacuum systems to limit the gas charge. The outgassing features should be taken into account at the very beginning of the design process when materials are selected.

Table 10: Values of outgassing rates for selected materials used in vacuum technology. The reported heat treatments are carried out both *in situ* and in vacuum.

Material	q (mbar l s^{-1} cm^{-2})	Main gas species
Neoprene, not baked, after 10 h of pumping [13]	order of 10^{-5}	H_2O
Viton, not baked, after 10 h of pumping [13]	order of 10^{-7}	H_2O
Austenitic stainless steel, not baked, after 10 h of pumping	3×10^{-10}	H_2O
Austenitic stainless steel, baked at 150°C for 24 h	3×10^{-12}	H_2
OFS copper, baked at 200°C for 24 h	order of 10^{-14}	H_2

3 Calculation of pressure profiles

The calculation of the pressure profile along vacuum systems is an essential task of vacuum experts and should be tackled at the design phase. In general, the contributions to the total pressure of localized and distributed gas sources are considered separately and finally added. This is possible because in most cases the equations that describe pressure profiles are linear. This may not be true if the pumping speed is pressure-dependent.

3.1 Pressure profiles generated by localized gas sources

The pressure in a vacuum vessel is obtained by taking into account Eq. (29) and the intrinsic pressure limitation P_0 of the installed pumping system:

$$P = \frac{Q}{S} + P_0. \tag{35}$$

The pumping speed S is either given by the supplier or preliminarily measured; P_0 is the pressure attained in the system without any gas load. When a restriction of conductance C is interposed between the pump and the vessel, the effective pumping speed S_{eff} is considered and Eq. (35) becomes

$$P = \frac{Q}{S_{\text{eff}}} + P_0 = \frac{Q(C+S)}{CS} + P_0. \tag{36}$$

When many vessels are interconnected, the flow balance is written in each vessel (node analysis). This analysis leads to a system of linear equations, from which the pressure values in each vessel are calculated. As an example, with reference to Fig. 8, in the first vessel, the injected gas flow (Q) is either pumped (P_1S_1) or transmitted to the second vessel ($C_1(P_1 - P_2)$). This latter flow is pumped in the second vessel or transmitted to the third vessel, and so on. Thus

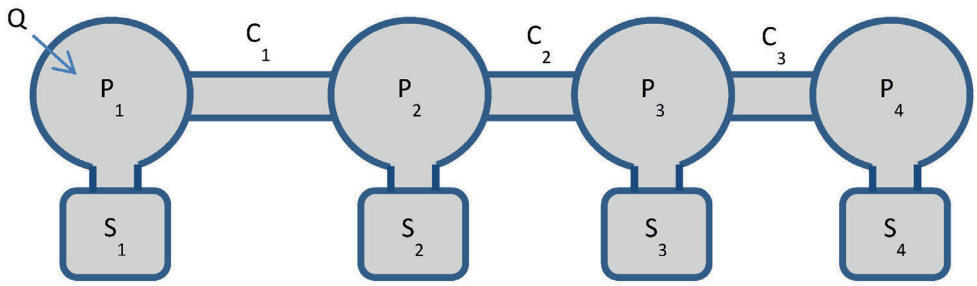


Fig. 8: Schematic drawing of four interconnected vacuum vessels. In each vessel, the gas flow balance is written (node analysis).

$$\begin{gathered} Q = P_1 S_1 + C_1(P_1 - P_2), \\ C_1(P_1 - P_2) = C_2(P_2 - P_3) + P_2 S_2, \\ C_2(P_2 - P_3) = C_3(P_3 - P_4) + P_3 S_3, \\ C_3(P_3 - P_4) = P_4 S_4. \end{gathered} \tag{37}$$

When a second localized gas flow is settled, the node analysis is repeated. The contributions of each localized gas flow to the pressure values are then added. If the cross-section of the interconnection ducts is constant, it can be shown that the pressure varies linearly between two connected vessels.

3.2 Pressure profiles generated by distributed gas sources

The pressure profiles generated by distributed outgassing and lump pumps is calculated analytically for simple geometries. Let us consider the case of a cylindrical beam pipe (radius R, length L) connected to a pump at one of its extremities. The calculation starts by imposing a mass balance equation in a generic small fraction of the beam pipe (see Fig. 9):

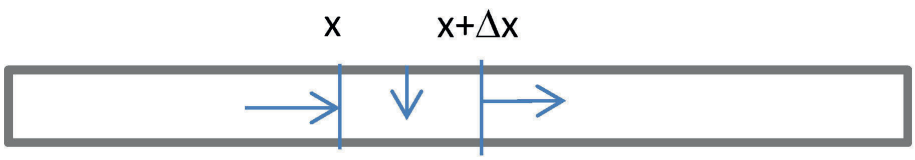


Fig. 9: Schematic geometrical model used to write the mass balance equation

$$\begin{gathered} Q(x+\Delta x) - Q(x) = 2\pi R \Delta x\, q \quad \Rightarrow \quad \frac{\mathrm{d}Q}{\mathrm{d}x} = 2\pi R q, \\ Q(x+\Delta x) = -C\frac{L}{\Delta x}\big(P(x+\Delta x) - P(x)\big) = -CL\frac{\Delta P}{\Delta x} \quad \Rightarrow \quad Q(x) = -CL\frac{\mathrm{d}P}{\mathrm{d}x}, \\ \Rightarrow \quad CL\frac{\mathrm{d}^2 P}{\mathrm{d}^2 x} = -2\pi R q, \end{gathered} \tag{38}$$

where q is the outgassing rate of the unit surface area, and C, CL and $CL/\Delta x$ are the conductances of the beam pipe, the unit length and the section Δx of the beam pipe, respectively.

Equation (38) indicates that the pressure profile is parabolic. The analytical expression is obtained by imposing the pressure at the pump location $P(0)$ and the absence of pressure gradient at the opposite beam pipe extremity (see Fig. 10).

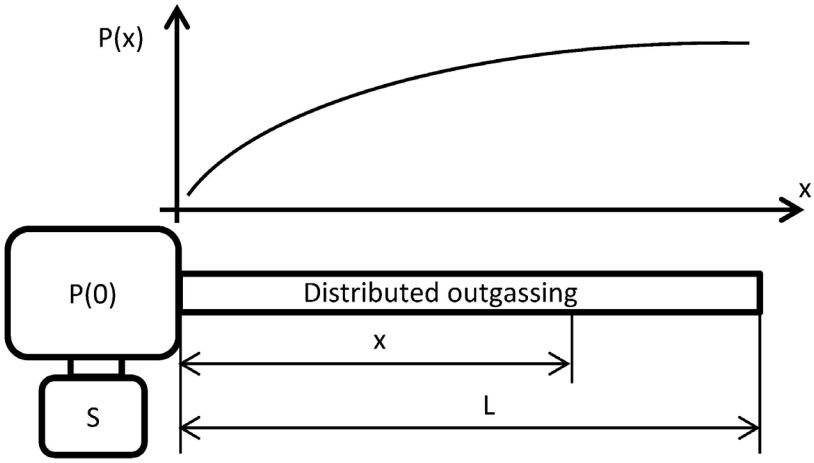


Fig. 10: Pressure profile in a tube pumped at one extremity with uniformly distributed outgassing rate

$$P(0)=\frac{Q_{TOT}}{S}=\frac{2\pi RLq}{S},$$
$$\left(\frac{dP}{dx}\right)_{x=L}=0.$$

Therefore,

$$P(x)-P(0)=-\frac{Q_{TOT}}{C}\left[\left(\frac{x}{L}\right)-\frac{1}{2}\left(\frac{x}{L}\right)^2\right],$$
$$P(L)-P(0)=-\frac{Q_{TOT}}{2C}. \tag{39}$$

where Q_{TOT} is the outgassing rate of the whole beam pipe.

If a second identical pump is installed on the other extremity of the beam pipe (see Fig. 11), the pressure profile is again parabolic, with a maximum at the centre of the vacuum chamber ($x = L/2$):

$$P(0)=\frac{2\pi RL\,q}{2S}=\frac{Q_{\mathrm{TOT}}}{2S},$$
$$P(x)-P(0)=-\frac{Q_{\mathrm{TOT}}}{2C}\left[\left(\frac{x}{L}\right)-\left(\frac{x}{L}\right)^2\right], \tag{40}$$
$$P\left(\frac{L}{2}\right)-P(0)=\frac{Q_{\mathrm{TOT}}}{8C}.$$

Numerical codes and Monte Carlo simulations are available for geometries that are more complex. Dedicated codes have been extensively used [14].

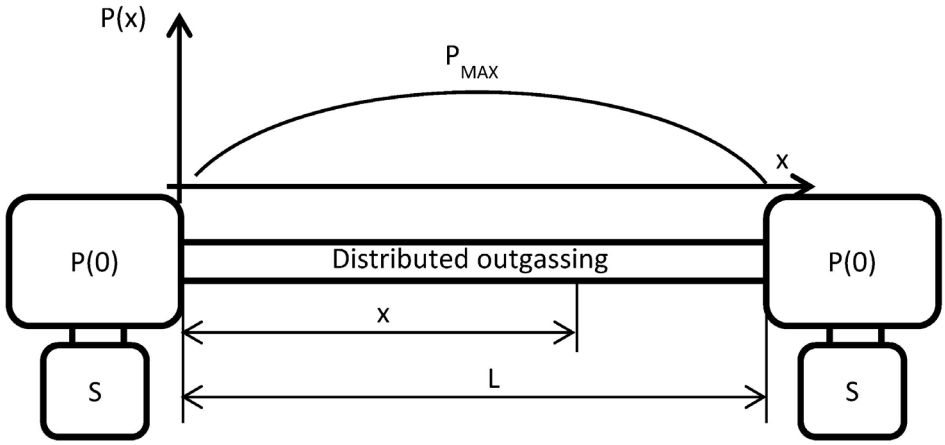


Fig. 11: Pressure profile in a tube pumped at both extremities with uniformly distributed outgassing rate

3.3 Time-dependent pressure profiles

The calculations shown in the previous subsection are valid only for time-independent conditions. In time-dependent conditions, the variation of the quantity of gas in the vacuum system has to be added to the balance equations such as Eq. (38). For the simplest vacuum system (vessel and pump), the gas balance equation becomes

$$\frac{\mathrm{d}N}{\mathrm{d}t} = Q - SP,$$

where Q is the gas load, SP is the gas removed by the pump, and N is the number of molecules in the vacuum system. Differentiating Eq. (2) and converting to pressure–volume units, the following equation is obtained:

$$V\frac{\mathrm{d}P}{\mathrm{d}t} = Q - SP. \tag{41}$$

Equation (41) is easily solved:

$$P(t) = Ae^{-t/\tau_\mathrm{p}} + \frac{Q}{S}.$$

The characteristic time of pumping $\tau_\mathrm{p} = V/S$ characterizes all pressure transients in a vacuum system. The integration constant A is calculated by imposing the pressure at $t = 0$. An exponential pressure variation is obtained, attaining the stationary value asymptotically. The pressure variations for $P(0) = P_0$ and $P(0) = 0$ are

$$P(0) = P_0 \quad \Rightarrow \quad P(t) = \left(P_0 - \frac{Q}{S}\right)\mathrm{e}^{-t/\tau_\mathrm{p}} + \frac{Q}{S}, \tag{42}$$

$$P(0) = 0 \quad \Rightarrow \quad P(t) = \frac{Q}{S}\left(1 - \mathrm{e}^{-t/\tau_\mathrm{p}}\right). \tag{43}$$

In the case of time-dependent gas sources, as in pulsed ion sources, the solution of Eq. (41) is given by

$$\begin{aligned} P(t) &= \frac{\int \mathrm{e}^{t/\tau_\mathrm{p}} \frac{Q_\mathrm{in}(t)}{V}\mathrm{d}t + A}{\mathrm{e}^{t/\tau_\mathrm{p}}} \\ &= \frac{\int Q_\mathrm{in}(t)\mathrm{d}t}{V} - \frac{\mathrm{e}^{-t/\tau_\mathrm{p}}}{V\tau_\mathrm{p}}\int \mathrm{e}^{t/\tau_\mathrm{p}}\left[\int Q_\mathrm{in}(t)\mathrm{d}t\right]\mathrm{d}t + A\mathrm{e}^{-t/\tau_\mathrm{p}}. \end{aligned} \tag{44}$$

For complex vacuum systems where several components are interconnected, the balance equation (41) is written for each vessel. This results in a system of coupled differential equations whose solution could be time-consuming. However, the analogy between vacuum systems and electrical networks may be used to accelerate the calculation.

3.4 Calculation of pressure profiles by the electrical analogy

Equations (11), (26) and (27) show that there is an analogy between vacuum systems and electrical networks. In Table 11 vacuum components and variables are correlated with electrical elements and characteristics.

Table 11: Electrical analogy of vacuum components and variable.

Vacuum element	Electrical elements	Electrical symbol
Conductance C	Conductance $1/R$	
Gas flow Q	Current I	
Pressure P	Voltage V	
Volume V	Capacitance C	
Pump	Conductance to ground	
Gas source	Current generator	
Constant pressure source	Voltage supply	

The ground potential is equivalent to zero pressure. A vacuum chamber of a given conductance and volume corresponds to two resistances and a capacitance. For symmetry, the capacitance is placed in the middle of the two resistances. If a local gas source and a pump are added, a current generator and a resistance to ground are connected to the circuit (see Fig. 12).

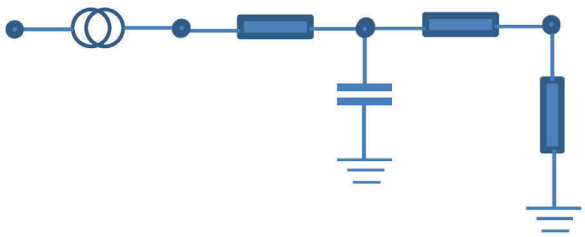


Fig. 12: Electrical analogy of a vacuum chamber with gas injection and pumping

Another typical example is given in Fig. 13; it represents a differential pumping. This configuration is used when a high gas flow must be prevented from reaching the low-pressure side of a vacuum system. It consists of a small diaphragm with pumps on both sides.

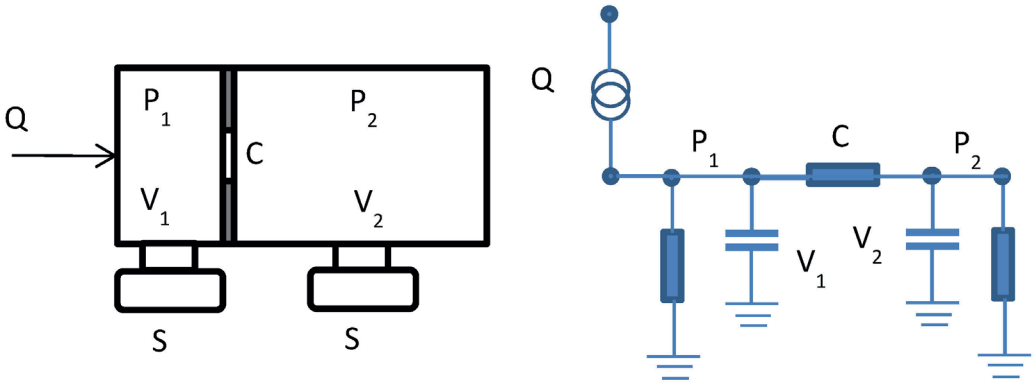


Fig. 13: Electrical analogy of a vacuum system with differential pumping

Long beam pipes are subdivided into small units to calculate the axial pressure distribution; the small units are considered as single vacuum chambers (volume and conductance) in series. The conductance of a single small unit is equal to the conductance of the entire vacuum chamber times the number of units. Distributed outgassing is taken into account by inserting a current generator on each unit.

The electrical network is solved by dedicated software, for example LTSpice. The time evolution and pulsed sources are easily included in the calculation. Nonlinear electrical components are used to simulate pressure- and time-dependent conductances and pumping speeds.

The MedAustron ring [15] and the Linac4 ion source [16] have been fully simulated by means of the electrical analogy. The effects of modifications in the shape of the components, position and size of the pumps have been checked and optimized.

4 Gas pumping

Vacuum pumps in the molecular regime are classified into two families: momentum transfer pumps and capture pumps. Both act on each molecule singularly since no momentum and energy transfer is possible between molecules in this pressure range. In the first family, molecules receive a momentum component pointing towards the pump outlet (foreline) where the gas is compressed and evacuated by pumps working in the viscous regime (e.g. rotary vane, diaphragm and scroll pumps). The second family removes gas molecules by fixing them on a surface exposed to the vacuum.

4.1 Momentum transfer pumps

In this family of pumps, gas molecules are steered by collision with moving surfaces (molecular pumps) or supersonic jets (diffusion pumps). The latter mechanism is no longer used in particle accelerators, though it still has extensive applications in industrial vacuum systems.

4.1.1 Molecular pumps

In molecular pumps, gas molecules impinge and adsorb on the moving surface; on desorption, the molecular emission is no longer isotropic due to the superposition of the wall velocity (see Fig. 14). As a result, molecules are preferentially redirected towards the direction of the wall movement, i.e. the density of molecules increases in the same direction.

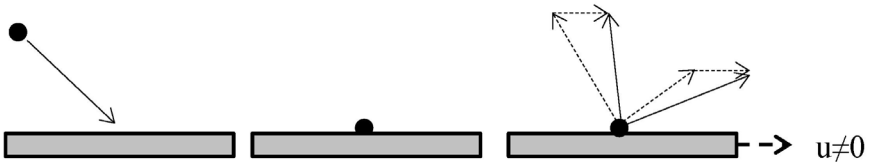


Fig. 14: Schematic drawing showing the mechanism of momentum transfer by collision onto a moving surface

The first molecular pump was invented by W. Gaede. In the original design, the moving surface is a rotor revolving at high frequency (see Fig. 15). To prevent back-streaming, the inlet and the outlet were separated by a very thin slot (locking slot). The slot height was of the order of 10^{-2} mm.

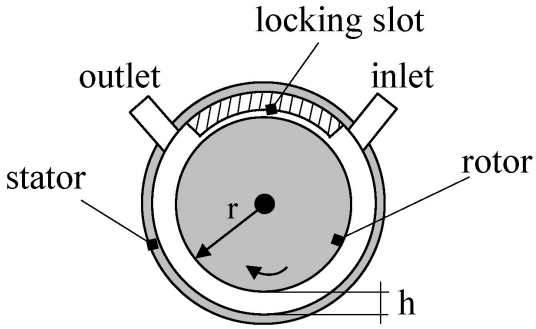


Fig. 15: Schematic drawing of a molecular pump

The most important characteristics of a molecular pump are the pumping speed S and the maximum compression pressure ratio that can be achieved between the pump inlet P_{IN} and outlet P_{OUT}:

$$K_0 = \left(\frac{P_{\mathrm{OUT}}}{P_{\mathrm{IN}}}\right)_{\mathrm{MAX}}. \tag{45}$$

The crucial parameters affecting S and K_0 are highlighted by a simplified model (see Fig. 16).

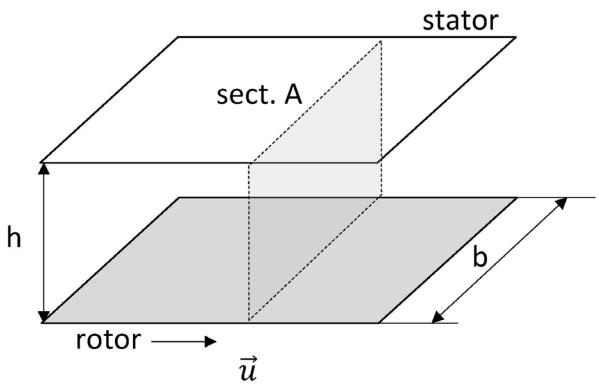


Fig. 16: Plane section of a molecular pump

At any time, half of the molecules have just collided with the moving surface (rotor) and drift in the direction of the velocity $\vec{u}$. The other half hit the stator where the drift component is lost. As a result, the drifted molecular flow Q_{p} towards an imaginary cross-sectional area A (see Fig. 16) is

$$Q_{\mathrm{p}} = \frac{1}{2} nubh = \frac{1}{2} \frac{P}{k_{\mathrm{B}} T} ubh,$$

where 'bh' is the area A of the cross-section and 'ubh' is the gas volume drifting through area A in one second. Converting into pressure–volume units and dividing by P, the pumping speed is obtained as

$$S = \frac{Q_p}{P} = \frac{1}{2} ubh. \tag{46}$$

Therefore, the pumping speed of molecular pumps is proportional to the speed of the moving wall. It does not depend on the nature of the gas; thus, molecular pumps are not selective. Finally, a large cross-section 'bh' is needed to receive a large quantity of gas.

The moving wall produces a higher gas density near the pump outlet. In turn, the pressure gradient generates a gas backflow Q_{bf} that can be written as

$$Q_{\mathrm{bf}} = -\bar{c}\, \nabla P,$$

where $\bar{c}$ is the conductance of a unit-length duct of cross-sectional area bh. For $h \ll b$, one has

$$\bar{c} = \frac{2}{3} \frac{(bh)^2}{b+h} \langle v \rangle.$$

The net flow is therefore given by

$$Q = Q_{\mathrm{p}} - Q_{\mathrm{bf}} = \frac{1}{2} Pubh - \bar{c}\, \nabla P.$$

When the net flux is zero, the maximum compression ratio is reached, in which case

$$\frac{\mathrm{d}P}{P} = \frac{1}{2} \frac{ubh}{\bar{c}} \mathrm{d}x.$$

Integrating between inlet and outlet (distance L) and using Eqs. (13) and (20) gives

$$K_0 = \left(\frac{P_{\mathrm{OUT}}}{P_{\mathrm{IN}}}\right)_{\mathrm{MAX}} = \exp\left(\frac{ubhL}{\bar{c}}\right) \propto \exp\left(\frac{u}{\langle v \rangle} \frac{L}{h}\right) \propto \exp\left(u\sqrt{m_i}\, \frac{L}{h}\right). \tag{47}$$

To obtain high K_0:

- the velocity of the moving surface must be of the order of the mean molecular speed (high $u/\langle v \rangle$ ratio), and
- the duct where molecules drift must be narrow and long (high L/h),

In addition, K_0 depends strongly on the molecular mass, the lowest being for H_2. As a result, the ultimate pressure of molecular pumps is dominated by H_2. More generally, the low K_0 for H_2 is an intrinsic limitation of momentum transfer pumps.

4.1.2 Turbomolecular pumps

For many years, the industrialization of molecular pumps was hindered by the machining of the locking slot. The narrow gap between rotor and stator created mechanical issues due to tolerances and thermal expansion. The problem was overcome in 1957 when W. Becker invented the turbomolecular pump (TMP). In the TMP, the momentum transfer is produced by rapidly rotating blades rather than parallel surfaces.

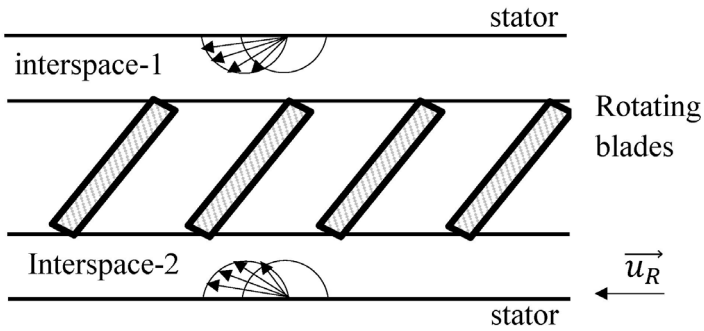


Fig. 17: A single rotor–stator stage in a turbomolecular pump, showing the molecular speed distributions

The rotating blades are tilted with respect to the rotational axis. As a result, oblique channels between successive blades are formed. The rotor is inserted between two static surfaces defining two interspaces (see Fig. 17). When the molecule comes from interspace-1, the angular distribution of the molecular velocity seen from the blades is preferentially oriented towards the blades' channels. Conversely, for those coming from interspace-2, the velocity vectors point preferentially towards the blades' wall, therefore increasing backscattering. A gas flow is consequently generated from interspace-1 to interspace-2. This mechanism works only if the angular distribution of the molecular speed as seen from the blade is significantly deformed, i.e., only if the blades' speed is at least of the order of the mean molecular speeds (hundreds of metres per second).

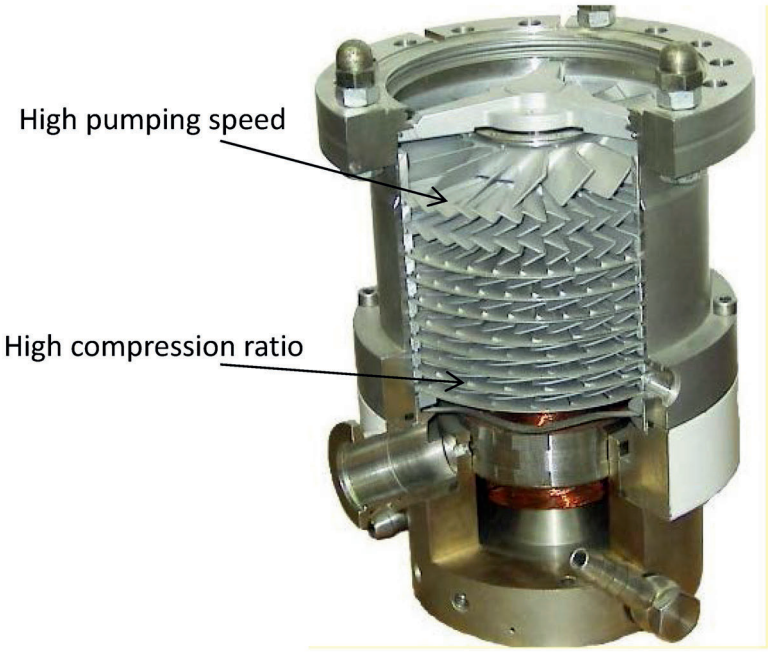


Fig. 18: Cut-out view of a turbomolecular pump.
Courtesy of Wikipedia (http://en.wikipedia.org/wiki/Turbomolecular_pump).

In a real turbomolecular pump, the gas is compressed by several series of rotating blades (see Fig. 18). Every series of rotating blades is followed by a series of static blades. Molecules transmitted through the rotating blades' channels hit the static blades; as a result, the angular distribution of velocity is randomized and the molecules are ready for the next compression stage. The momentum transfer is effective only if the molecules do not experience intermolecular collisions after hitting the

blades; this is equivalent to saying that the mean free path has to be larger than the blade distances. As a result, this type of pump works at full pumping speed only in molecular regimes ($P < 10^{-3}$ mbar).

The conclusions drawn for molecular pumps hold also for turbomolecular pumps. The dimension of the blades and the rotor–stator distance are larger for the first series of blades in view of maximizing the pumping speed (see Eq. (46)). The last compression stages are tighter to increase the maximum compression ratio (see Eq. (47)) and compensate for the increased gas density and thus for the resulting lower molecular mean free path. The pumping speed of a TMP (Fig. 19) is constant in the molecular regime ($P < 10^{-3}$ mbar) and ranges between 10 and 3000 l s^{-1} depending on pump inlet diameter and mechanical design. As expected, the maximum compression ratio is the lowest for H_2; in classical designs it is about 10^3 (see Fig. 20). Nowadays, values up to 10^6 are achieved by integrating a molecular drag pump (Gaede or Holweck drag stage) to the rear of the blade sets.

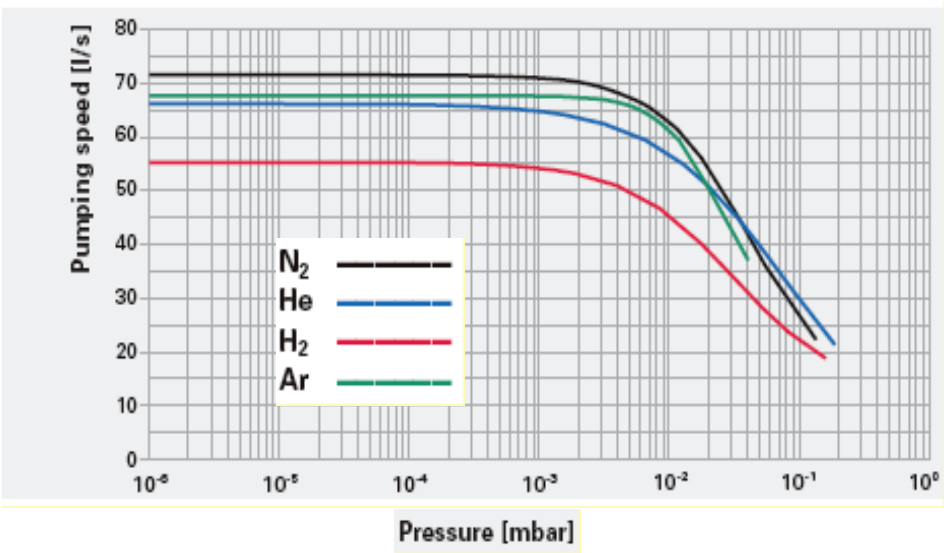


Fig. 19: Pumping speeds of a turbomolecular pump equipped with a DN63CF flange. For TMP, the pumping speeds are constant in the free molecular regime and the nature of the pumped gas has a limited effect, if compared with other type of pumps. Courtesy of Pfeiffer Vacuum.

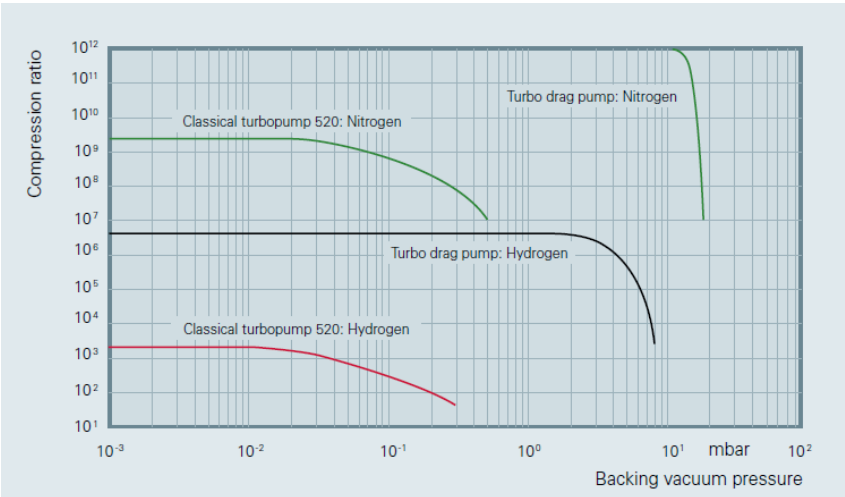


Fig. 20: Maximum compression ratio of two different turbomolecular pumps for N_2 and H_2. Courtesy of Pfeiffer Vacuum.

The TMP ultimate pressure is of the order of 10^{-10}–10^{-11} mbar for baked and all-metal vacuum systems. Lower pressures may be obtained if another type of pump removes the TMP H_2 back-streaming.

At the pump outlet, the compressed gas is evacuated by a mechanical pump (often called backing pump) operating in the viscous regime. The turbomolecular pump and its backing pump are in general assembled in a single unit that includes power supplies, controls and instrumentation.

Each set of rotor blades is machined from a single block of high-strength aluminium alloys. Typical pumps equipped with a DN100 flange rotate at a frequency of 1 kHz, therefore approaching circumferential speeds up to 500 m s^{-1}.

The TMP is the best available pump for the first stages of ion sources. It can evacuate high gas flow at relatively high pressure without selectivity and memory effects. For example, a DN100 TMP can withstand a continuous gas flow up to 2×10^{-1} mbar l s^{-1} without damage.

The use of dry pumps for TMP backing has surmounted the risk of oil back-streaming from rotary vane pumps. The higher ultimate pressure of dry pumps is compensated by the integration of molecular drag pumps (higher compression ratio). The complete removal of all lubricated mechanical bearings has been obtained by magnetic rotor suspension.

The main drawback of a TMP is related to possible mechanical failures leading to definitive damage of the high-speed rotor. In addition, in case of unwanted rotor deceleration caused by power cut or rotor seizing, the vacuum system has to be protected by safety valves and dedicated pressure sensors against air back-stream. TMPs have limited application in radioactive environments due to possible damage to their electronics and power supplies. Some radiation-resistant TMPs are now available; with a significant cost, the whole electronics is moved beyond the radiation shielding by means of long cables.

4.2 Capture pumps

Capture pumps remove gas molecules by fixing them onto an internal wall. To be efficient, capture pumps must block the gas molecules for an average time (sojourn time t_s) much longer than the typical running time of the accelerator. An estimation of the sojourn time is given by the Frenkel law [17]:

$$t_\mathrm{s} = t_0\, \mathrm{e}^{E_\mathrm{a}/k_\mathrm{B}T}, \tag{48}$$

where E_a is the adsorption energy and $t_0 \approx h/k_\mathrm{B}T \approx 10^{-13}$ s. Very long sojourn times are obtained either for high adsorption energies

$$E_\mathrm{a} \gg k_\mathrm{B}T,$$

or for very low temperatures

$$T \ll \frac{E_\mathrm{a}}{k_\mathrm{B}}.$$

In the former case, capture pumps are called *chemical pumps* or *getter pumps*; in the latter case, *cryogenic pumps*.

In chemical pumps, the binding force involves electron exchange between gas molecules and surfaces (*chemisorption*). Typical binding energies for getter pumps are higher than 1 eV/molecule. In cryogenic pumps, Van der Waals' interactions might be sufficiently strong to fix molecules to cold surfaces (*physisorption*). In this case, the binding energy can be much lower than 0.5 eV/molecule.

Chemisorption, implantation and physical burial of gas molecules by reactive metal atoms are combined in sputter ion pumps. They represent the most important and widely used pumping technology in particle accelerators.

4.2.1 Sputter ion pumps

In a sputter ion pump (SIP) the residual gas is ionized in a Penning cell. The ions are then accelerated towards a cathode made of a reactive metal (Fig. 21). The cathode–ion collisions provoke sputtering of reactive-metal atoms that are deposited on the nearby surfaces (Fig. 22). The Penning cells are assembled in honeycomb structures to form the pump's elements.

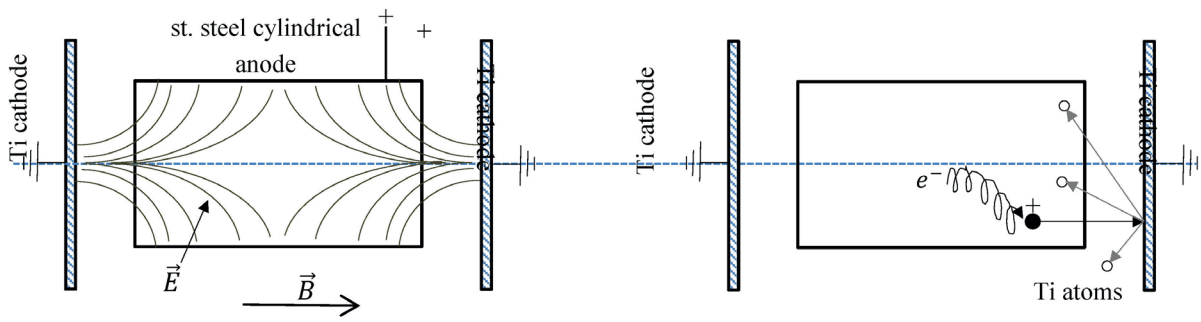


Fig. 21: Schematic drawing depicting the pumping mechanism of sputter ion pumps in the diode configuration

The pumping action is given by three distinct mechanisms:

- Chemical adsorption of gas molecules onto the reactive metal layer and subsequent burial by additional sputtered atoms; this works for all gas species except for rare gases, which cannot form chemical compounds.
- Implantation both of gas ions in the cathode and of energetic neutrals bounced back from the cathode into the deposited film; this is the only pumping mechanism for rare gases.
- Diffusion into the cathode material and the deposited film; this is possible only for H_2 because of its very large diffusivity in metals in its atomic form.

In the diode SIP (Fig. 21), the anode is composed of several open cylinders made of stainless steel at a positive electrical potential (from 3 to 7 kV). The cathodes are plates of Ti – at ground potential – placed a few millimetres from both anode extremities. A magnetic field parallel to the cells' axis is generated by external permanent magnets (about 0.1 T). In this configuration, the crossed electrical and magnetic fields trap electrons in long helical trajectories, resulting in an increased probability of gas ionization.

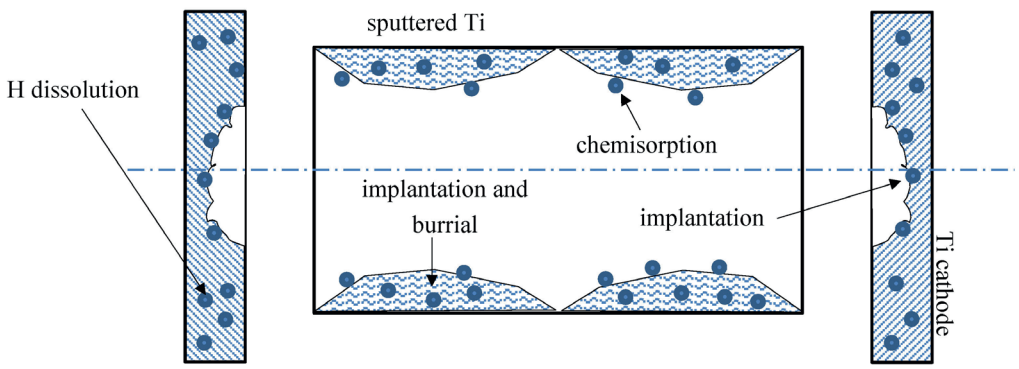


Fig. 22: Schematic drawing of a SIP cell after prolonged operation

Molecular implantation in the cathodes is not a permanent pumping; in fact, the progressive erosion due to gas ion sputtering sooner or later frees the implanted gas (Fig. 22). Once released from the cathodes, reactive gas species are chemisorbed while rare gases remain in the gas phase until the next implantation. Pressure instabilities have been reported after excessive pumping of rare gas: the continual erosion liberates gas that increases the pressure, which in turn increases the erosion. Consequently, a pressure rise is generated; this is stopped when most of the rare gas is implanted in a deeper zone of the cathode. A new pressure spike appears when the erosion reaches the new rare gas front. This type of instability is typically observed when large quantities of air are pumped (Ar represents 1% in air) and is known as 'argon disease' [18].

The pumping efficiency for rare gas is improved by reducing the number of ions implanted into the cathodes, while increasing the energetic neutrals bounced back from the cathode and their chance to be buried by Ti atoms onto the anode. Two different approaches have been commercially exploited:

- cathodes made of metals with higher atomic mass, and
- different geometry than the diode configuration.

In the first case, Ta is used instead of Ti (181 amu against 48 amu). Gas ions, once neutralized on the cathode surface, have a higher probability to escape from the cathode and bounce back with higher energy when colliding on Ta rather than Ti. Consequently, the implantation probability on the anode is higher and the quantity of gas implanted into the cathode is lower. Pumps that use Ta cathodes are called 'noble diode' or 'differential ion' pumps.

In the second case, the diode configuration is modified by introducing a third electrode (triode sputter ion pumps, see Fig. 23). The cathode consists of a series of small Ti plates aligned along the cell axis and polarized negatively with respect to the ground. The anode is at the ground potential.

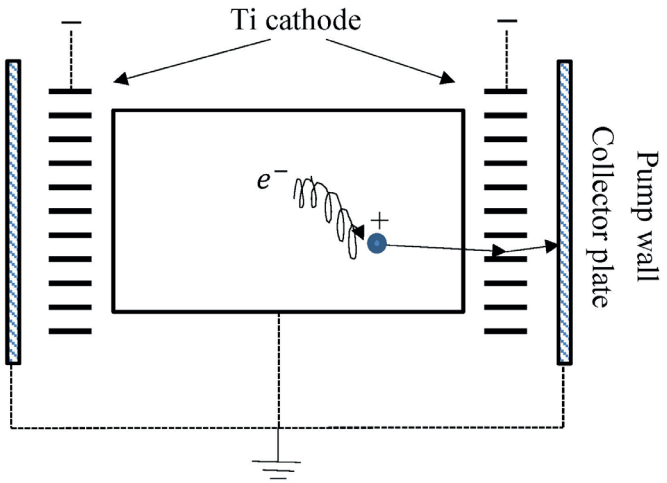


Fig. 23: Schematic drawing of a triode SIP cell. The Ti cathodes are negatively biased

In this configuration, the collisions between ions and cathode are at glancing angle: the sputtering rate, the neutralization and the bouncing probability are higher. As a result, more Ti atoms are sputtered and more gas molecules are implanted into the collector plate and anode; the gas implanted into the cathode is minimized.

An improved triode pump is the StarCell® produced by Agilent. Two sheets of Ti that are cut and bent to form radial blades (see Fig. 24) replace the series of small plates. This increases the rigidity of the cathode and reduces the risk of short-circuits.

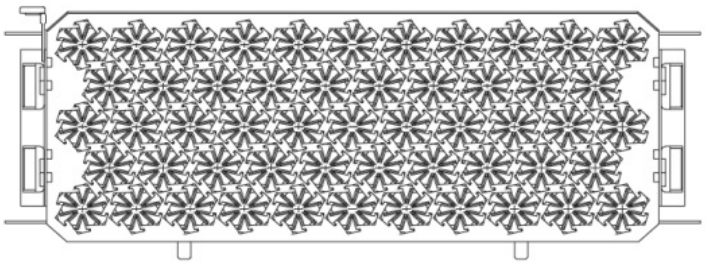


Fig. 24: Drawing of Agilent's StarCell Ti cathode. Courtesy of Agilent Vacuum

4.2.1.1 Hydrogen pumping by SIP

In SIP, H_2 is mainly pumped by dissolution into the Ti cathodes. To be absorbed, H_2 must be dissociated. Because in the Penning discharge only a small fraction of the hydrogen ions are H^+, the dissociation is possible only on the Ti cathodes after removal of the oxide layer by sputtering. This surface cleaning process is very slow if H_2 is the leading gas in the vacuum system due to its low sputtering yield (0.01 at 7 keV on Ti). Therefore, at the beginning of the operation, the pumping speed for H_2 is lower than the nominal value and increases gradually with time.

The simultaneous pumping of another gas has a strong effect on H_2 pumping efficiency. If the additional gas has a higher sputtering yield, the cathode cleaning is faster and the nominal pumping speed is reached in a shorter time. Conversely, it can contaminate the cathode surface and reduce the H_2 adsorption. Finally, dissolved H atoms can be sputtered away, reducing the pumping efficiency.

When the concentration of H_2 in the cathodes is higher than the solubility limit in Ti, hydride precipitates are formed. The cathodes expand and become brittle, generating anomalous field emission and short-circuits. A typical concentration limit is 10 000 mbar l for a commercial 500 l s^{-1} SIP. Water vapour pumping also contributes to the total quantity of H_2 charged in the cathodes.

During the pumping of H_2 or after the dissolution of a high quantity of the same gas, the operation of SIP at high pressures (higher than 10^{-5} mbar) can lead to thermal run-away. The Penning discharge heats the cathode and provokes gas desorption, which reinforces the discharge. This positive feedback mechanism can cause local melting of the cathode. To avoid this problem, the electrical power provided to the pump is limited at high pressure and during the ignition.

4.2.1.2 Pressure measurement by SIP

The ion current collected by the SIP's cathodes is used as an indication of the pressure. In fact, a defined relationship is measured between the two variables down to pressures of the order of 10^{-9} mbar. For lower pressures, the pump current is limited by field emission (leakage current). The low-pressure reading can be extended by one decade if the applied voltage is reduced in the lower-pressure range (e.g. from 7 to 3 kV). The pressure reading by SIP is extensively used in particle accelerators.

4.2.1.2 Pumping speed of SIP

The pumping speed of SIP depends on the gas pressure at the pump inlet, the nature of the gas, and the quantity of gas already pumped. The latter dependence is correlated with the implantation into the cathode. For new pumps and gases other than H_2, the rate of implantation is larger than the rate of release by sputtering. Conversely, for 'saturated' pumps, the two rates equate and the pumping in the cathode is negligible. The nominal pumping speed of SIP is in general given for 'saturated' pumps.

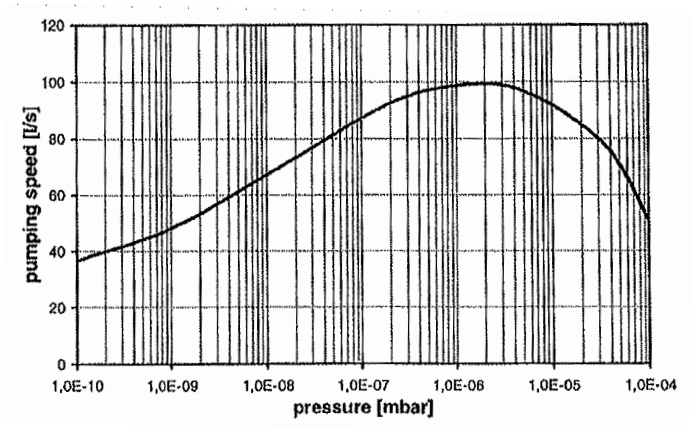


Fig. 25: Typical inlet pressure dependence of SIP pumping speed; the maximum pumping speed is reached at about 10^{-6} mbar. Courtesy of Agilent Vacuum.

The pumping speed attains a maximum for inlet pressures of about 10^{-6} mbar. This is the value given by the suppliers for the nominal pumping speed. For lower pressures, the pumping speed decreases, half of the maximum being obtained at about 10^{-9} mbar (see Fig. 25). Typical values of S are collected in Tables 12 and 13. A cut-away view of the StarCell pump is shown in Fig. 26.

Table 12: N_2 nominal pumping speed of commercial StarCell SIP ('saturated') for different standard pump inlet diameters.

Pump inlet diameter, DN (mm)	S_{N_2} **(l s^{-1})**
63	50
100	70 and 125
150	240 and 500

Table 13: Nominal pumping speed normalized to that of air for diode and triode SIP.

Gas	**Air**	**N_2**	**O_2**	**H_2**	**CO**	**CO_2**	**H_2O**	**CH_4**	**Ar**	**He**
Diode	1	1	1	1.5-2	0.9	0.9	0.8	0.6-1	0.03	0.1
Triode	1	1	1	1.5-2	0.9	0.9	0.8	0.6-1	0.25	0.3

Fig. 26: Cut-away view of a triode StarCell pump. The inlet flange is DN63CF. Courtesy of Agilent Vacuum

4.2.2 Getter pumps

Getter materials fix gas molecules on their surface by forming stable chemical compounds. The chemical reaction is possible only if the getter atoms are not already combined with surface contaminants, i.e. oxygen, carbon, etc. A contamination-free surface may be produced in two ways:

- sublimating the getter material *in situ* – evaporable getters or sublimation pumps; and
- dissolving the surface contamination into the bulk of the getter material by heating – non-evaporable getters (NEGs); the dissolution process is called activation.

A getter surface is characterized by its sticking probability α, namely the capture probability per molecular impingement. From Eq. (32), the pumping speed of getter material is written as

$$S = \alpha A_{\text{getter}} C', \tag{49}$$

where A_{getter} is the geometric surface area of the getter material. The α value depends on many variables, first of all the nature of the gas. Getter materials do not pump rare gases and methane at room temperature. In that respect, they always need auxiliary pumping to keep a stable pressure. The sticking probabilities of homonuclear diatomic molecules are in general lower than those of molecules composed of dissimilar atoms. The values of α for common gas species in vacuum systems may be placed in increasing order as follows:

$$N_2 < H_2 < O_2 < H_2O < CO < CO_2.$$

For N_2 typical α values are in the high 10^{-3} to low 10^{-2} range; for CO and CO_2, α values close to 1 may be measured. In addition, the sticking probability is strongly dependent on the gas surface coverage, i.e. the quantity of gas previously pumped on the getter material. As the gas surface coverage increases, the sticking probability decreases. The pumping becomes negligible when the surface is fully covered by gas molecules. As an example, complete saturation of a smooth NEG surface is obtained when about 10^{15} CO molecules cm^{-2} are adsorbed. Quantities about 5–10 times higher are reported for H_2O and O_2. H_2 dissociates on the getter surface and diffuses into the bulk of the NEG material. As a result, its pumping does not block the adsorption of other molecules. The pumping capacity of hydrogen is orders of magnitude higher than that for other gas species.

The surface roughness also has a strong impact on the sticking probability. A single molecular arrival may result in several collisions with a rough getter surface, therefore enhancing the capture probability. The pumping capacity per unit geometric area is also increased.

Owing to the high mobility of H atoms and the relatively low binding energy, the pumping of this gas is reversible: H_2 may be released by heating. The opposite effects of pumping and desorption set up an equilibrium pressure P_{H_2} that depends on heating temperature T and hydrogen concentration c_H. The three variables are correlated by Sieverts' law, i.e. the law of mass action for the H_2–metal system:

$$\log P_{H_2} = A + 2\log c_H - \frac{B}{T} \tag{50}$$

where A and B are typical values for a given getter material.

4.2.2.1 Sublimation pumps

In particle accelerators, Ti is the only getter material used as evaporable getter. It is sublimated on the inner walls of a dedicated vessel from Ti–Mo filaments heated up to 1500°C (Fig. 27). The maximum sticking probability varies in the range (1–5) × 10−2 for H2 and 0.5–1 for CO (Fig. 28). In some cases, the wall of the sublimation pumps may be cooled to liquid-nitrogen temperature. In this case, the sublimated film is rougher and consequently the sticking probabilities are higher. When the deposited film is saturated, a new sublimation is needed to recover the initial pumping speed. A single filament withstands tens of sublimation cycles.

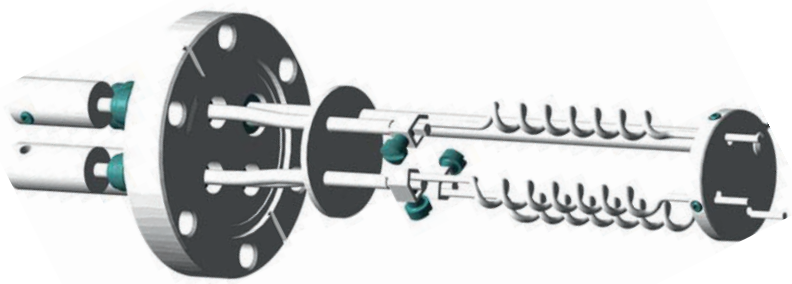


Fig. 27: Ti wires of a sublimation pump. Courtesy of Kurt J. Lesker Company

Sublimation pumps provide very high pumping speeds for a relatively low cost and have a limited maintenance. They are in general used to achieve pressures in the UHV range. Operation with continuous sublimation in higher pressure ranges may be conceived; in this case, cooling of the pump envelope must be added.

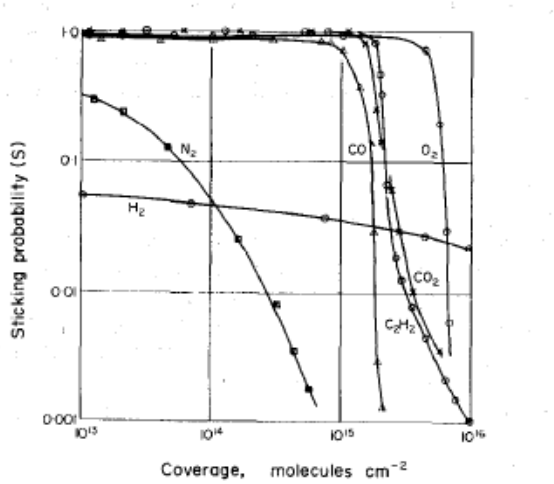


Fig. 28: Typical single-gas coverage dependence of the pumping speed of sublimation pumps for selected gas species. This plot and other interesting measurements with sublimation pumps may be found in the very instructive paper of Ref [19].

4.2.2.2 Non-evaporable getter pumps

During the activation process, the native oxide layer is dissolved in the NEG material. This relocation is possible only if it is energetically feasible and the bulk of the material has a large solubility limit for oxygen. Only few elements have these peculiar characteristics. Among them, those of the fourth group (Ti, Zr and Hf) are the most relevant. These metals are able to dissolve oxygen at room temperature to more than 20% in atomic concentration. Other metallic elements are added to increase oxygen diffusivity and so reduce the activation temperature and time; this is typically the case for V. For specific purposes, other elements may be added in the alloy, for example to reduce pyrophoricity in air.

NEG materials are usually produced in powder form to increase the surface exposed and the pumping capacity. The powdered materials are either sintered to form discs and pellets or pressed on metallic substrates shaped as filaments and ribbons. The latter have found an extensive application in particle accelerators as linearly distributed pumps [20] (see Fig. 29). Thin-film NEG coatings produced by magnetron sputtering are also used in modern accelerators in the UHV pressure range [21].

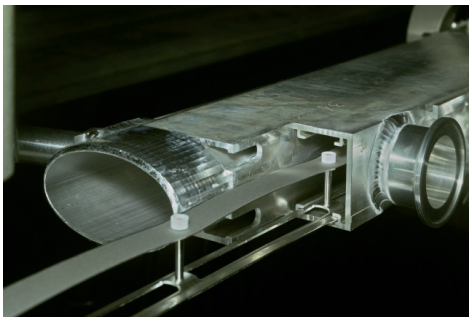


Fig. 29: Photograph of the dipole magnet vacuum chamber of the Large Electron–Positron collider. The NEG pump is made of Zr–Al powder fixed on both sides of a constantan ribbon.

A typical NEG alloy commercialized by SAES Getters (Milan, Italy) is the St707 made of Zr–V–Fe. It is activated at 400°C for 1 h or at 300°C for about 24 h; such characteristics allow passive activation during the bake-out of the stainless-steel vacuum chamber.

Modern lump NEG pumps provide gas surface capacities about 200 times higher than smooth surfaces. The embrittlement of the NEG elements limits the maximum quantity of H_2 that may be dissolved in the pumps. A safe limit of 20 Torr l g^{-1} is given by SAES Getters for the St707. For the same alloy, the H_2 equilibrium pressure is shown in Fig. 30 for various temperatures. The constants A and B of Eq. (50) are 4.8 and 6116, respectively, when the pressure is expressed in Torr and the concentration in Torr l g^{-1}.

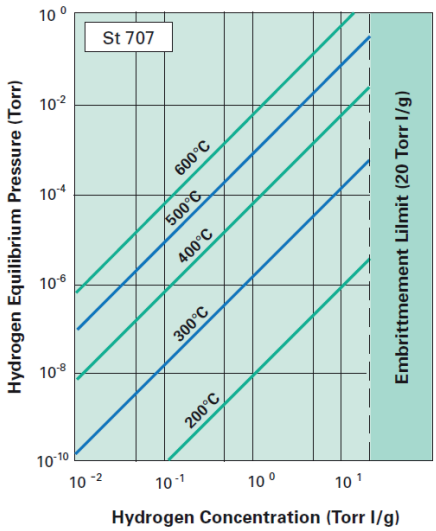


Fig. 30: Hydrogen equilibrium pressure for the St707 alloy. Courtesy of SAES Getters

After the pumping of a large quantity of H_2, a regeneration of the NEG pump may be necessary to avoid embrittlement. To do so, the NEG material is heated and the released gas is removed by a turbomolecular pump. The duration of the regeneration (t_R) is given [22] by

$$t_R = \frac{M_{NEG}}{S_{TMP}}\left(\frac{1}{q_f} - \frac{1}{q_i}\right) \times 10^{-A-(B/T)}, \tag{51}$$

where M_{NEG} (g) is the mass of the NEG material, S_{TMP} (l s^{-1}) is the pumping speed of the turbomolecular pump, q_i and q_f (Torr l g^{-1}) are the initial and the final concentrations of hydrogen, and T (K) is the absolute temperature.

At room temperature, the pumping speed decreases as a function of the amount of gas pumped, except for H_2, which diffuses into the alloy. The drop of pumping speed is less pronounced than for the sublimation pumps, due to the high porosity and related surface area of commercial NEG materials (see Fig. 31). The SAES Getters CapaciTorr pumps (see Fig. 32) have a sorption capacity for CO at room temperature up to about 5 Torr l (~1.5 × 10^{20} molecules). NEG pumps may operate at higher temperatures; the diffusion of C and O atoms is enhanced and the surface capacities largely increased. The practical result is a less pronounced pumping speed drop as a function of the quantity of gas pumped. However, the H_2 release might be an obstacle to achieve the required pressure.

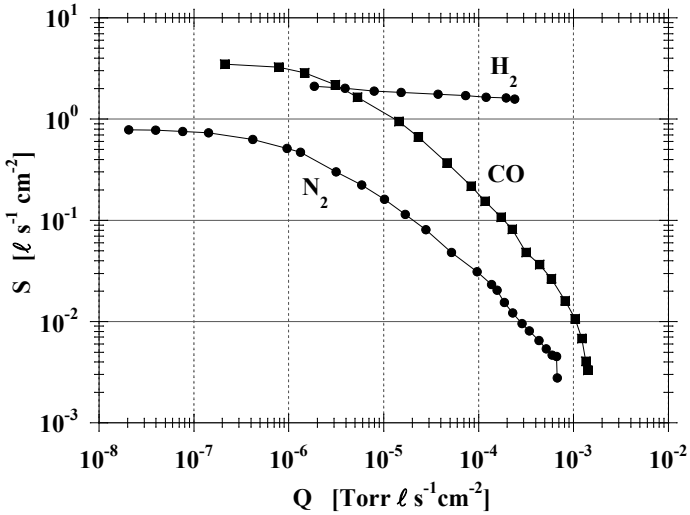


Fig. 31: Typical single-gas coverage dependence of the pumping speed of a St707 ribbon for selected gas species. This plot and other interesting measurements with NEG ribbons may be found in [23].

CO and CO_2 pumping hinder the adsorption of all other gases, while N_2 has only a limited effect and H_2 leaves the NEG surface unaffected.

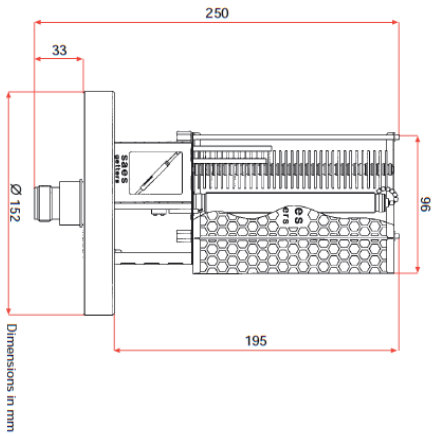


Fig. 32: Drawing and dimensions of a CapaciTorr pump. Courtesy of SAES Getters

4.2.3 Cryopumps

Cryopumps rely on three different pumping mechanisms:

- *Cryocondensation.* This mechanism is based on the mutual attraction of similar molecules at low temperature (Fig. 33). The key property is the saturated vapour pressure P_v, i.e. the pressure of the gas phase in equilibrium with the condensate at a given temperature [24]. The lowest pressure attainable by cryocondensation pumps is limited by the saturated vapour pressure. Among all gas species, only Ne, H_2 and He have P_v higher than 10^{-11} Torr at 20 K. The P_v of H_2 at the liquid He boiling temperature is in the 10^{-7} Torr range, and is 10^{-12} Torr at 1.9 K. The quantity of gas that may be cryocondensed is very large and limited only by the thermal conductivity of the condensate.
- *Cryosorption.* This is based on the attraction between gas molecules and substrates (Fig. 33). The interaction forces with the substrate are much stronger than those between similar molecules. As a result, providing the adsorbed quantity is lower than one monolayer, the

sojourn time is much longer and gas molecules are pumped at pressure much lower than the saturated vapour pressure. A significant quantity of gas may be pumped below one monolayer if porous materials are used; for example, in one gram of standard charcoal for cryogenic application, about 1000 m^2 of surface are available for adsorption. The important consequence is that significant quantities of H_2 and He may be pumped at 20 K and 4.3 K, respectively. In general, submonolayer quantities of all gas species are effectively cryosorbed at their own boiling temperature.

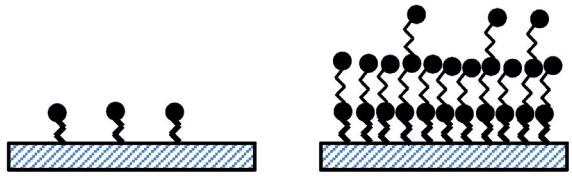


Fig. 33: Schematic drawing depicting (left) submonolayer cryosorption, where the mechanism is molecule–substrate interaction; and (right) cryocondensation, where the leading mechanism is intermolecular interaction.

– *Cryotrapping.* In this mechanism, the molecules of a low-boiling-temperature gas are trapped in the condensation layer of another gas. This is possible because the interaction energy between dissimilar molecules may be much higher than that between similar molecules. The trapped gas has a saturated vapour pressure several orders of magnitude lower than in its pure condensate. Typical examples are Ar trapped in CO_2 at 77 K and H_2 in N_2 at 20 K.

Modern cryopumps exploit the first two mechanisms. Cryocondensation takes place on a cold surface, in general at 80 K for water vapour and at 10–20 K for the other gas species. The cryosorption of He, H_2 and Ne is localized on a hidden surface coated with a porous material. This part of the pump is kept out of the reach of the other type of molecules, i.e. they have a probability close to one to be intercepted and adsorbed by another surface before reaching the cryosorber (see Fig. 34). The cooling is obtained by He gas cooled by a Gifford–McMahon cryocooler.

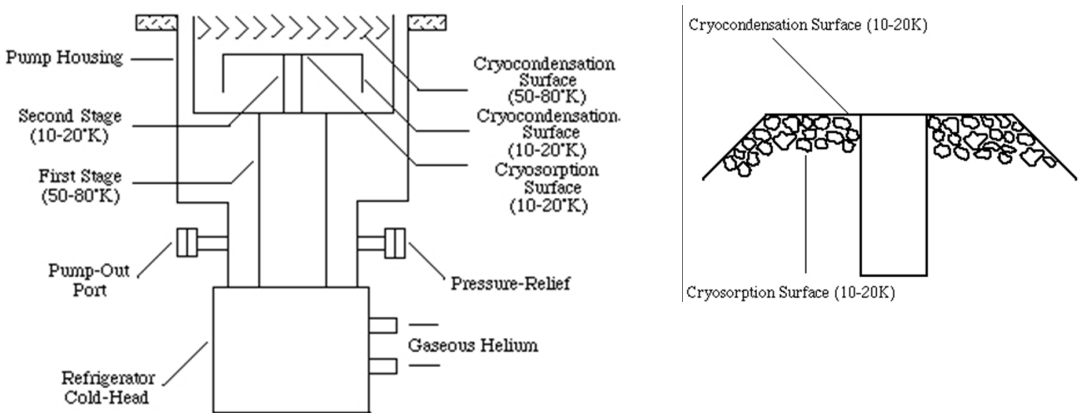


Fig. 34: Drawing of a generic cryopump with a closer view to the cryosorption surface where the porous material is fixed.

Cryopumps having pumping speeds in the range 800–60000 l s−1 are commercially available (Table 14). For condensable gas molecules, the capture probability is close to 1 (e.g. for water vapour). The maximum gas capacity (also called maximum gas intake) for the condensable gas is limited only by the thermal conductivity of the condensate. To avoid a thick condensate layer and excessive thermal load, cryopumps should be started in the molecular regime ($P < 10^{-3}$ mbar). The

quantity and properties of the porous material determine the maximum gas intake of a cryosorbed gas. In general, it is orders of magnitude lower than that for a condensable gas.

Table 14: Pumping speeds and maximum gas capacities of a commercial cryopump (Oerlikon-Leybold 800 BL UHV); the pump inlet diameter is 160 mm. Courtesy of Oerlikon-Leybold.

	H_2O	N_2	Ar	H_2	He
S [l s^{-1}]	2600	800	640	1000	300
Capacity [Torr l]		225 000	225 000	3225	375

Cryopumps require periodic regeneration to evacuate the gas adsorbed or condensed; in this way, the initial pumping speed is fully recovered. To do so, the cryopumps are warmed up to room temperature and the released gas is removed by mechanical pumps (mobile TMP in particle accelerators). During regeneration, the pump is separated from the rest of the system by a valve.

Excessive gas adsorption on the cryosorber leads to performance deterioration. A partial and much faster regeneration (1 h against more than 10 h) may be carried out at temperatures lower than 140°C in such a way as to remove the sorbed gas without releasing water vapour from the pump stage at higher temperature.

4.3 Comparison of pumps for ion sources

All pumps presented in the previous sections have been used (or are going to be used) in the vacuum systems of ion sources and in their adjacent vacuum sectors. The choice is based not only on the gas load and required pressure, but also on the costs of the pumps, their maintenance, the duty time of the source and safety aspects. The advantages and disadvantages of the four types of pumps presented in this chapter are summarized in Table 15.

Table 15: Advantages and disadvantages of the presented types of pumps.

Type of pump	Advantages	Disadvantages
TMP	– No memory effects – Constant pumping speed for pressures lower than 10^{-3} mbar – Pumping speed independent of total gas load – Starts working at high pressures (molecular regime)	– Mechanical fragility – Risk of contamination from the backing pump – Need of venting any time the pump is stopped – Need of valve on the inlet flange – Intrinsic limitation in ultimate pressure of H_2 – Possible vibrations – Regular maintenance
SIP	– Clean pumping – No maintenance – No vibrations – Installation in any orientation – Relatively long lifetime – Relatively low cost – Limited but high H_2 capacity – The pump current gives a pressure reading	– Low capture probability – Gas selectivity and limited capacity – Memory effects (in particular for rare gases) – Ignition not recommended in the 10^{-3}–10^{-4} mbar range – Heavy due to permanent magnets – Difficult starting for old pumps – Production of charged particles in particular at start-up – Field emission problems for old pumps – Fringing magnetic field – Safety issue: high voltage
Subli-mation	– Clean vacuum – High pumping speed for reactive gases – With SIP, extremely low vacuum can be achieved – Low cost per l s^{-1} – Electrical power only for sublimation; it works in case of power cut – Limited maintenance (filament change) – No vibration	– Very limited surface gas capacity – Need frequent sublimations at high pressure – Ti film peel-off for high sublimation rates – Selective pumping (no pumping of rare gases and methane) – Risk of leakage current in high-voltage insulators
NEG	– Clean vacuum – High pumping speed for reactive gases – With SIP, extremely low vacuum can be achieved – High gas capacity for porous NEG – Low cost per l s^{-1} – Electrical power needed only for activation; it works in case of power cut – No maintenance – No vibration	– Selective pumping (no pumping of rare gases and methane) – H_2 embrittlement if regeneration cannot be applied – Formation of dust particles is not excluded – Safety issue: pyrophoric, it burns when heated in air at high temperature
Cryo	– Very large pumping speed for all gases – Clean vacuum – High pumping capacity – Limited selectivity	– Cost and maintenance – Relatively large volume needed (including refrigerator) – Gas release in case of power cut – Reduced pumping efficiency for H_2 for high quantity of gas adsorbed: regeneration needed – Need of valve on the inlet flange

5 The vacuum layout of the Linac4 H^- source

Linac4 is a linear accelerator, under construction at CERN, conceived in the framework of the high-intensity and high-luminosity upgrade of the LHC. It will be connected to the LHC injection chain at the latest during the LHC long shutdown 2 (LS2) [25].

In the first sections of Linac4, the main gas loads are the pulsed H_2 injection in the source and the continuous injection of the same gas in the Low Energy Beam Transport (LEBT) for beam charge compensation. The vacuum system must evacuate the injected gas and ensure a pressure that is lower than 5×10^{-7} mbar in the Radio Frequency Quadrupole (RFQ).

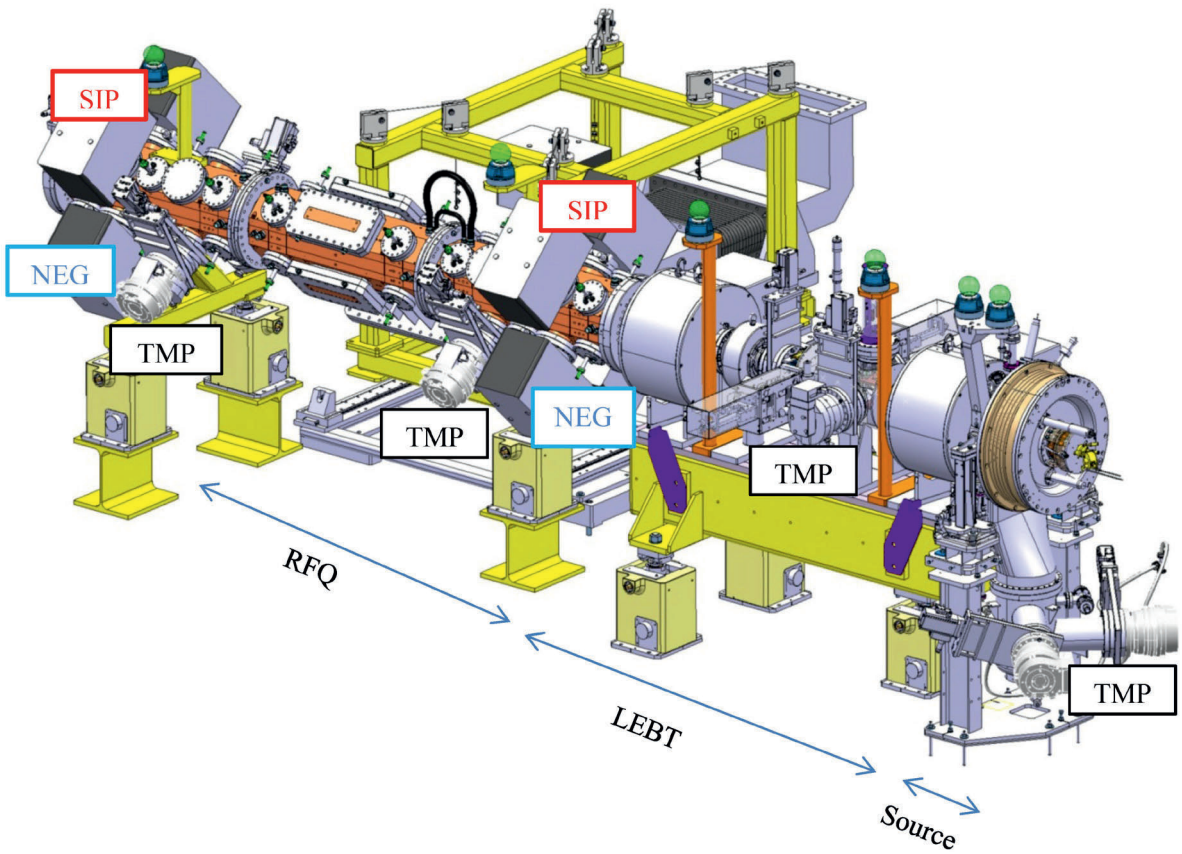


Fig. 35: Linac4 layout from the source to the RFQ. Courtesy of D. Steyaert, CERN EN-MME

Figure 35 shows the layout of the source up to the end of the RFQ. The extraction region of the source is equipped with two TMPs that evacuate most of the pulsed flux from the source. An additional TMP is connected to the LEBT diagnostic tank to pump the local continuous injection of H2. The RFQ is divided into three tanks. SIP, NEG pumps and TMP are installed on the first and the third tanks. The NEG pumps provide high pumping speed and capacity for hydrogen. The SIPs contribute to the H2 pumping and ensure the removal of gas species that are not pumped by getters. The TMP are backup pumps in case a problem occurs.

The electrical–network vacuum analogy was applied to evaluate the time-dependent pressure profiles from the source to the end of the RFQ. Figures 36 and 37 show the H2 partial pressure profiles as a function of time and position. Two situations are illustrated, i.e. the pressure profiles with and without continuous injection in the LEBT. A detailed analysis of the vacuum system of the Linac4 source can be found in Ref. [16].

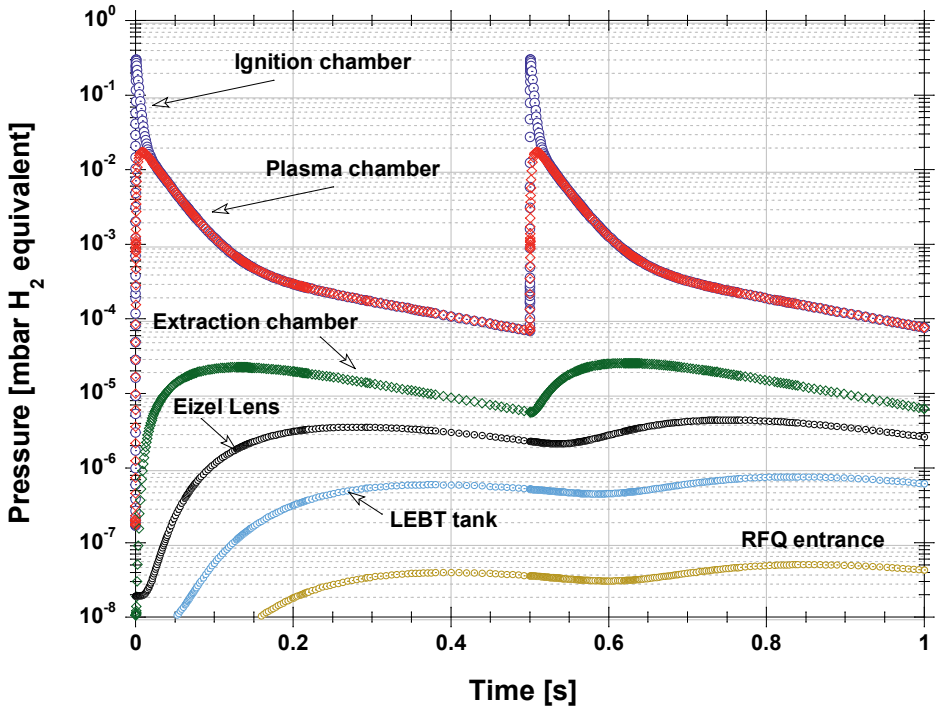


Fig. 36: Calculated H_2 pressure variations in the Linac4 H^- source. The H_2 gas burst from the piezo valve is about 5×10^{-3} mbar l per pulse. Each pulse lasts 5×10^{-4} s at a repetition frequency of 2 Hz. In this example, the injection in the LEBT is not taken into account.

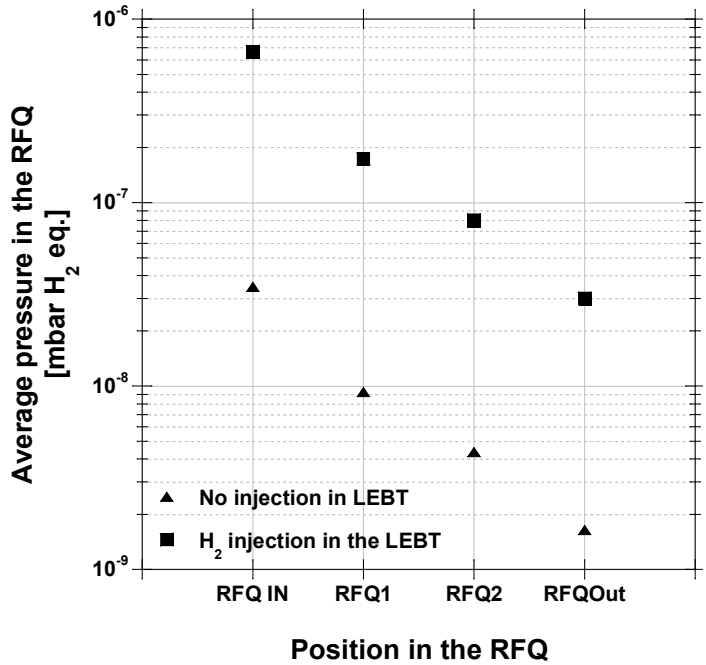


Fig. 37: Pressure profiles in the RFQ of Linac4 with and without H_2 injection in the LEBT. For the calculation, the injection in the LEBT is equivalent to fixing the pressure in the LEBT to 10^{-5} mbar (H_2 equivalent).

Acknowledgement

I wish to thank all my colleagues of the Vacuum, Surfaces and Coatings group of the Technology Department at CERN, including students and visitors. The VSC group has become a centre of excellence in Europe for vacuum technology encompassing all aspects, from surface treatments to computation, operation and control of unique facilities. A significant part of my experience originates from years of work I have been sharing with them. Chiara Pasquino prepared Figs. 36 and 37 and wrote part of section 5.

References

[1] C. Benvenuti and M. Hauer, Proc. Eighth Int. Vacuum Congress, Cannes, 1980, vol. II, pp. 199–202; available on CERN Document Server at http://cdsweb.cern.ch/record/314288/files/CM-P00064854.pdf.

[2] Sir James Jeans, *An Introduction to the Kinetic Theory of Gases* (Cambridge University Press, Cambridge, 1967).

[3] J.B. Hudson, *Surface Science: an Introduction* (John Wiley & Sons, New York, 1998), p. 125.

[4] H. Kuhn, H.-D. Försterling and D.H. Waldeck, *Principles of Physical Chemistry*, 2nd edn (John Wiley & Sons, Hoboken, NJ, 2009), p. 405.

[5] K. Jousten (Ed.), *Handbook of Vacuum Technology* (Wiley-VCH, Weinheim, 2008).

[6] J.M. Lafferty (Ed.), *Foundations of Vacuum Science and Technology* (Wiley-Interscience, New York, 1998).

[7] P. Clausing, *Ann. Phys.* **12** (1932) 961; reprinted in *J. Vac. Sci. Technol.* **8** (1971) 636.

[8] A. Berman, *Vacuum Engineering Calculations, Formulas, and Solved Exercises* (Academic Press, San Diego, 1992).

[9] D.J. Santeler, *J. Vac. Sci. Technol. A* **4** (1986) 348.

[10] J. Cole, *Prog. Astronaut. Aeronaut.* **51** (1976) 261.
D.H. Davis, *J. Appl. Phys.* **31** (1960) 1169.

[11] R. Kersevan and J.-L. Pons, *J. Vac. Sci. Technol. A* **27** (2009) 1017.
Molflow+ are available at http://test-molflow.web.cern.ch/.

[12] P. Chiggiato, CERN-ATS-Note-2012-048 TECH (in French).
Lecture given at CAS on Vacuum in Particle Accelerators, 2006; available at http://cas.web.cern.ch/cas/Spain-2006/Spain-lectures.htm.

[13] R.N. Peacock, *J. Vac. Sci. Technol.* **17** (1980) 330.

[14] A. Rossi, LHC Project Note 341, 2004; available in the CERN library at http://cdsweb.cern.ch/record/728512/files/project-note-341.pdf?version=1.

[15] C. Yin Vallgren *et al.*, CERN ATS-Note-2012-041 TECH; available in the CERN library at http://cdsweb.cern.ch/record/1445914/files/ATSnote_final.pdf.

[16] C. Pasquino *et al.*, CERN-ATS-Note-2012-043 TECH; available in the CERN library at http://cdsweb.cern.ch/record/1447099/files/ATS_Note%20-LINAC4%20H-%20source.pdf.

[17] J. Frenkel, *Z. Phys.* **26** (1924) 117.

[18] K.M. Welch, *Capture Pumping Technology* (North-Holland Elsevier, Amsterdam, 2001), p. 106.

[19] A.K. Gupta and J.H. Leck, *Vacuum* **25** (1975) 362.

[20] C. Benvenuti, *Nucl. Instrum. Methods* **205** (1983) 391.

[21] C. Benvenuti, Proc. Sixth European Particle Accelerator Conf., 1998, Stockholm; available at http://accelconf.web.cern.ch/AccelConf/e98/PAPERS/THZ02A.PDF.

[22] B. Ferrario, Proc. 9th Symp. on Fusion Technology, Garmisch Partenkirchen, 1976.
C. Boffito, B. Ferrario, and D. Martelli, *J. Vac. Sci. Technol.* **A1** (1983) 1279.

[23] C. Benvenuti and P. Chiggiato, *J. Vac. Sci. Technol.* **A14** (1996) 3278.

[24] R. A. Haefer, *Cryopumping Theory and Practise* (Clarendon Press, Oxford, 1989), p. 85.

[25] L. Arnaudon *et al.*, CERN-ATS-2011-041; available in the CERN library at http://cdsweb.cern.ch/record/1378473/files/CERN-ATS-2011-041.pdf.

Beam Diagnostics

U. Raich
CERN, Geneva, Switzerland

Abstract
As soon as the first particles emerge from an ion source, the source characteristics need to be determined. The total beam intensity, the transverse particle distributions, the beam divergence and emittance as well as the longitudinal parameters of the beam must be measured. This chapter provides an overview of typical measurement methods and the instruments used, and shows the results obtained.

1 Introduction

The performance of an ion source determines to a large extent the performance of a complete accelerator chain. For example, the beam current, pulse length as well as the transverse beam parameters at CERN's proton source limit the luminosity and therefore the number of collisions observed in the Large Hadron Collider (LHC) experiments. Even though the cost of the source and the Low Energy Beam Transport (LEBT) is low compared to the cost of the LHC, careful design is needed to reach the performance and reliability required by such a big accelerator chain. In order to make sure the source fulfils the requirements, its beam must be carefully and precisely measured. This chapter describes typical examples of such beam measurements, explains the instruments used and shows a number of results obtained on ion sources at different laboratories. Some figures were taken from References [1–6] which contain additional useful information.

2 CERN's Linac4 H^- source and Low Energy Beam Transport

In order to reach the *ultimate* performance of the LHC and to avoid the risk of breakdown of the aging 50 MeV proton linac, a new 160 MeV H^- linac will be built at CERN. The H^- ions will be injected into the PS Booster before being accelerated and transferred to the LHC via the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS).

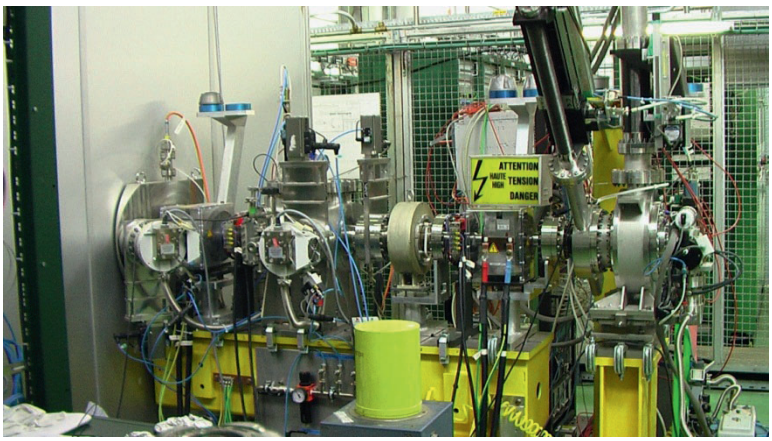


Fig. 1: Linac4 test stand, source and LEBT

In order to test the source, the Low Energy Beam Transport (LEBT) line and the Radio Frequency Quadrupole (RFQ) and chopper line, a test facility (Fig. 1) has been set up, allowing beam characterization at different stages in the acceleration chain. Commissioning the LEBT was accomplished in several stages (Fig. 2):

- The total current coming from the source was measured with a Faraday cup.
- The transverse beam parameters were determined using a slit/grid emittance meter.
- A spectrometer was used to measure the energy spread.
- A diagnostics box, a current transformer and the emittance meter were used to verify the beam optics, to maximize transmission through the line and to match the beam parameters to the RFQ.

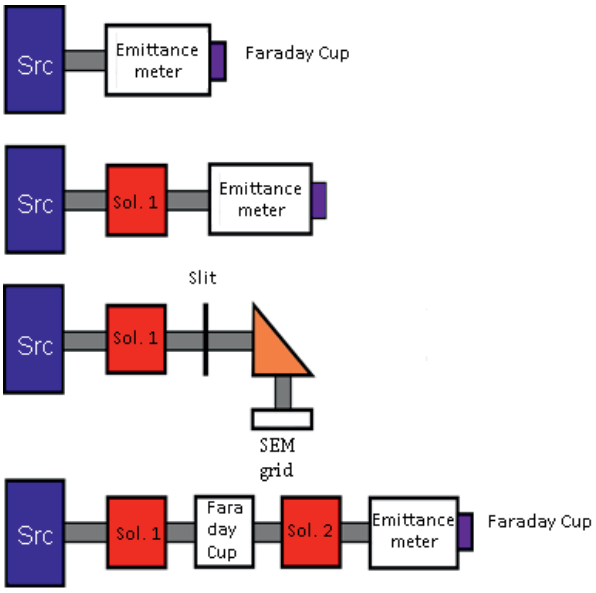


Fig. 2: LEBT layout for source parameter measurements

In the following, the beam parameters to be measured on an ion source will be described, together with the instruments used to perform the measurements. These are typically:

- Beam intensity
  - Faraday cup (destructive measurement)
  - Current transformer (non-destructive)
- Transverse profile
  - Wire harps
  - Residual gas monitors
- Transverse phase space
  - Slit/grid device
  - Allison scanner
  - Pepperpot
- Energy and energy spread
  - Spectrometer

3 Beam intensity measurements

Beam intensities are measured either destructively by collection of all beam charges in a Faraday cup or non-destructively with a current transformer, making use of the magnetic field associated with a charged particle beam.

3.1 The Faraday cup

A Faraday cup (Fig. 3) consists of an active electrode collecting the totality of beam charges, which usually is retractable. Since all beam charges are used for such a measurement, this is the most sensitive measurement possible, and currents down to a few picoamps can be detected. Typically, pneumatic in/out mechanisms are used to insert the device into the beam or to retract it.

When beam particles impinge on the electrode surface, secondary electrons are emitted. These may escape the cup and, by doing so, falsify the measurement. When measuring a proton beam, the beam current is thus overestimated, while for electron beams it is underestimated. Typically, the Faraday cup electrode has the form of a cone, such that the electrons created are recaptured in the cup. In addition, a guard ring with negative polarization voltage pushes the electrons trying to escape back into the cup. These secondary electrons typically have energies of few tens of electronvolts, such that rather low polarization voltages are usually sufficient to repel them.

In order to determine the polarization voltage needed, intensity measurements are taken while increasing the polarization voltage. The current measured decreases until all electrons are pushed back into the cup. Increasing the voltage further then has no effect on the signal.

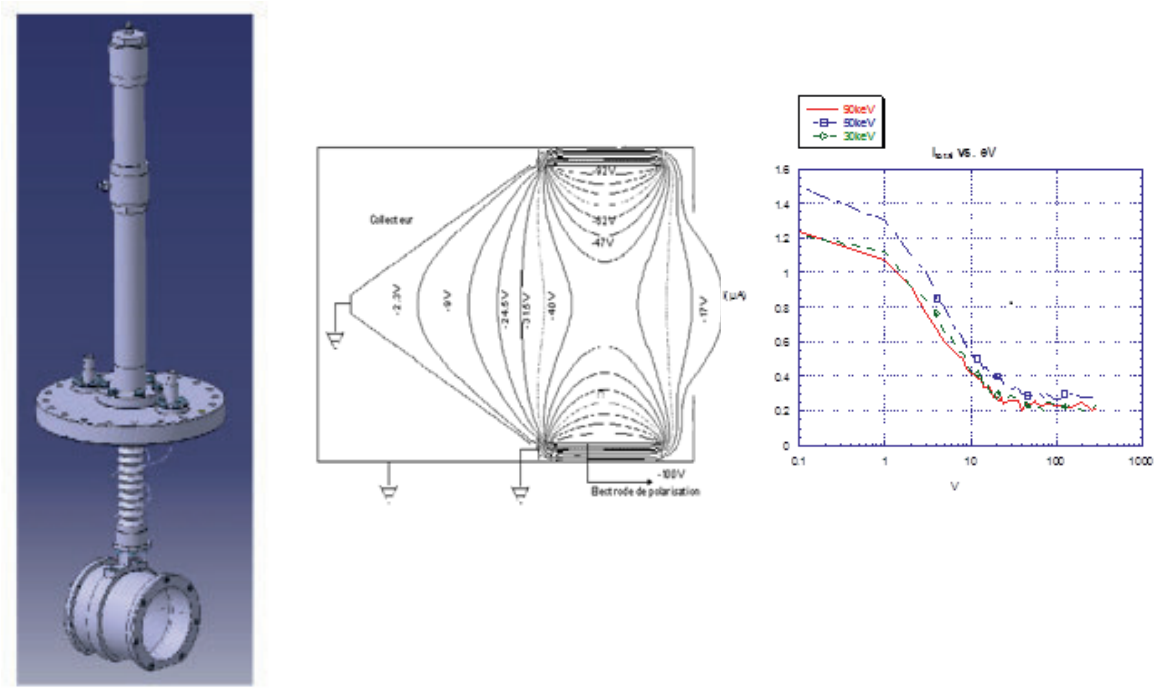


Fig. 3: Faraday cup, field map and typical measurements.to determine the polarization voltage

3.2 Fast current transformers

When measuring higher beam intensities, the magnetic field associated with the charged particle beam current may be used. These fields are extremely low and the field lines are therefore captured by a torus of magnetic material with very high permeability (inner torus in Fig. 4). The beam constitutes

the primary winding, while the signal is extracted from a secondary winding supplied by a wire wound around the magnetic material. A third winding is also usually used for calibration.

The signal voltage seen at the secondary winding is given by

$$U(t) = L\frac{I_{\mathrm{beam}}}{\mathrm{d}t}, \tag{1}$$

where the inductance L is given by

$$L = \frac{\mu_0\mu_{\mathrm{r}}}{2\pi} l N^2 ln\frac{r_{\mathrm{o}}}{r_{\mathrm{i}}}, \tag{2}$$

l is the length of the torus while r_{i} and r_{o} are the inner and outer radius, respectively. The typical relative permeability used, for example, that for the CoFe-based amorphous alloy Vitrovac, is $\mu_{\mathrm{r}} = 10^5$.

As is easily seen from Eq. (1), the signal does not depend on the current but on the current change. For a rectangular beam signal, only a positive and negative spike would therefore be observed. In order to better reconstruct the actual beam signal, a suitable capacitance and resistance network must be added, often coupled with an active amplification circuit.

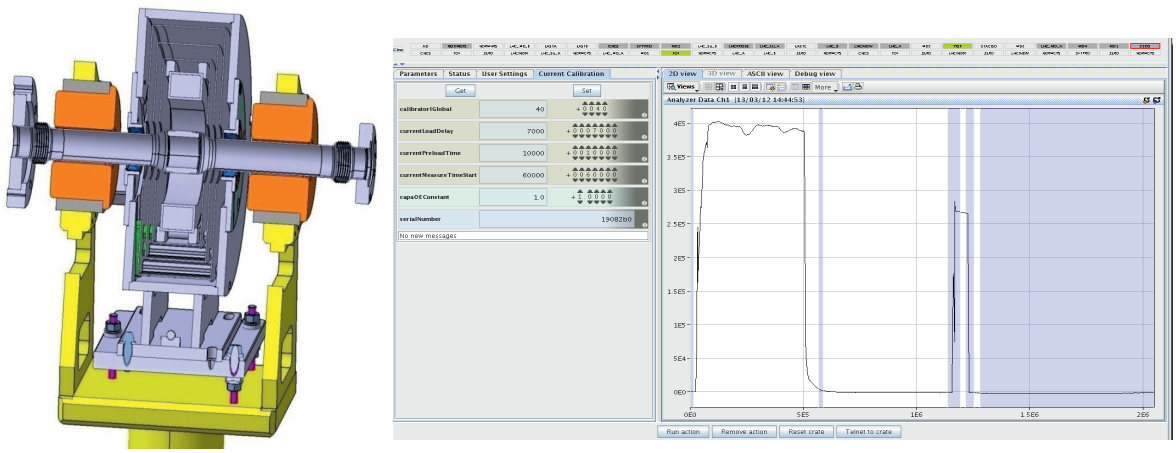


Fig. 4: Fast current transformer and typical signal including calibration pulse

The magnetic field generated by the beam is extremely weak, and it is important to shield the transformer against any external magnetic fields, for example those generated by nearby pulsed magnets. Multi-layer shielding enclosures of metals of increasing permeability from the outside towards the torus are therefore necessary to minimize the influence of perturbing fields. In addition, background measurement and subtraction may be used to further decrease the measurement errors.

When the beam travels in a conducting vacuum chamber, an image current is induced in the metallic walls, having the same amplitude but travelling in the opposite direction. If both these currents were to traverse the magnetic core, then the overall magnetic field generated by each would cancel. For such transformer measurements, the vacuum chamber is therefore interrupted by a ceramic insert and the image current is guided around the torus by a metallic housing (Fig. 5).

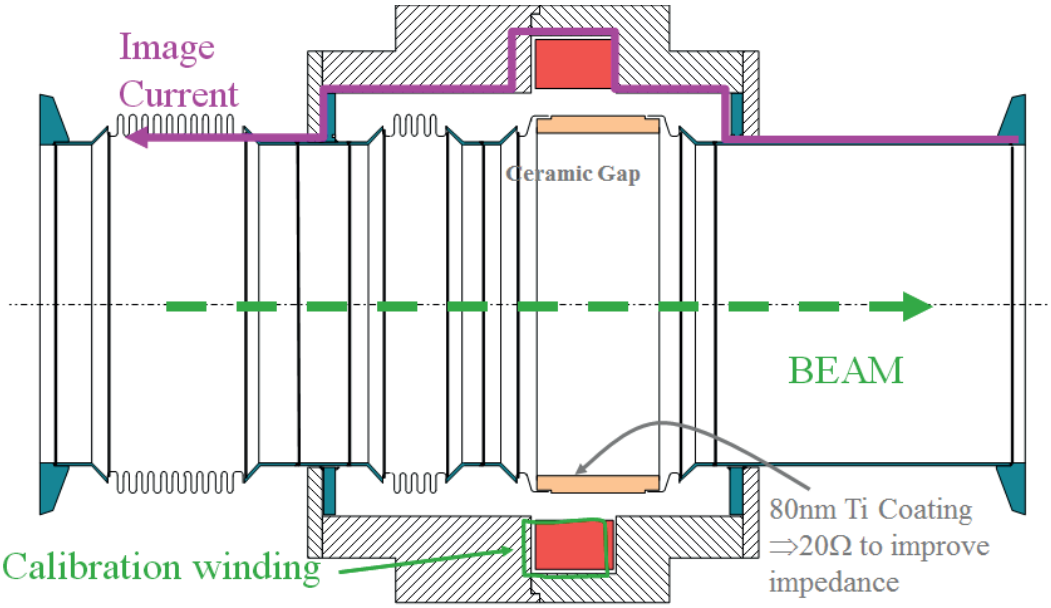


Fig. 5: Image current and ceramic gap in fast current transformer

While it is possible to minimize the signal droop to allow measurement of beam pulses up to a few seconds using an active fast current transformer, a different measurement principle is needed if DC currents are to be measured.

3.3 The DC current transformer

The DC current transformer (Fig. 6) consists of two identical tori with primary windings in opposite directions. The tori must be selected such that their magnetic differences are as small as possible. The windings are powered by a single modulator current, high enough to force the tori into saturation. The excitation currents in the windings and therefore the magnetic fields produced are of exactly the same strength in both tori such that the difference voltage seen at the output of the differential amplifier comparing the two is zero. When the beam passes through the transformer, an additional magnetic field is created that shifts the working point on the hysteresis curve for both tori. However, as the modulation current is opposite in each, the effect is asymmetric, resulting in a signal with twice the modulation frequency at the output of the differential amplifier. A feedback is created by means of a compensation current in the opposite direction with respect to the beam current, until the signal $v_a - v_b$ becomes zero again. Once this is the case, the feedback current is equal to the DC beam current and can be easily measured.

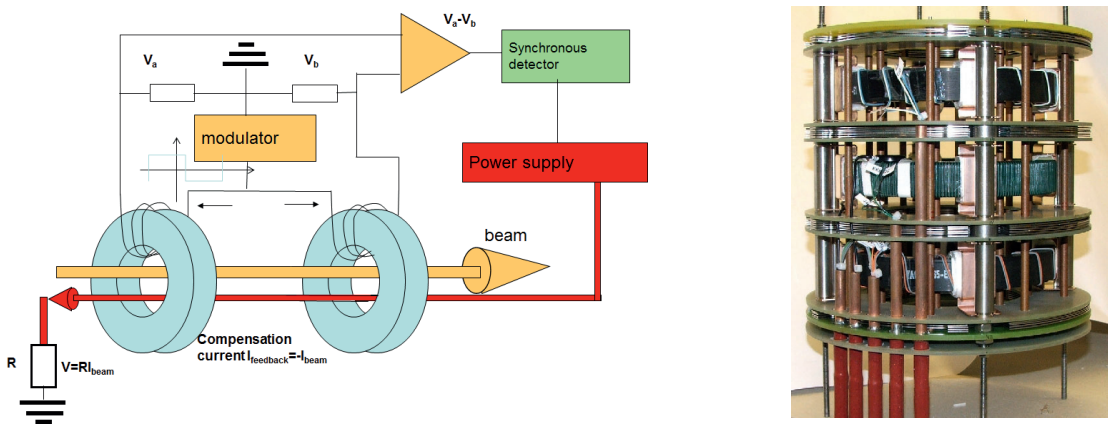


Fig. 6: DC current transformer

4 Profile measurements

There are different types of devices measuring the transverse distribution of particle beams. Wire grids or wire scanners are intercepting devices, which can have an effect on the beam, while the ionization profile monitor (IPM) or the luminescence monitor are non-intercepting devices, which rely on detecting the ionization or excitation of the small amounts of rest gas remaining in the vacuum chamber, respectively.

4.1 Wire grids

The wire grid (Fig. 7) is often used in transfer lines at high energies where the beam traverses the wire and creates secondary electrons when entering and exiting the wire material. This secondary emission current is proportional to the beam current at the position of the wire. Reading the current on each wire results in a transverse profile where the resolution depends on the number of wires covering the beam distribution. The minimum wire spacing achievable is typically around 300 μm.

At the very low energies that we are dealing with at the ion source, the particles cannot traverse the wire but deposit their charge in it. The wire grid therefore works like a multi-channel Faraday cup. Polarization can be provided with additional wires mounted in front of and/or at the back of the wire plane used for measurement or by polarizing the wires themselves. The cost of such a system is largely dominated by the number of electronic channels needed. In order to limit this cost, grids with non-constant wire spacing are often built (Fig. 7), where a fine spacing in the centre allows for a high resolution, while a wider spacing at the edges still allows measuring the tails of the distribution, albeit with less resolution. Since the interceptive nature of the device disturbs the beam in this case, the grids are usually mounted on an in/out mechanism driven by a motor or pneumatic system.

With wire grids, a complete beam profile is obtained with a single beam pulse, which is a big advantage on machines with a low duty cycle.

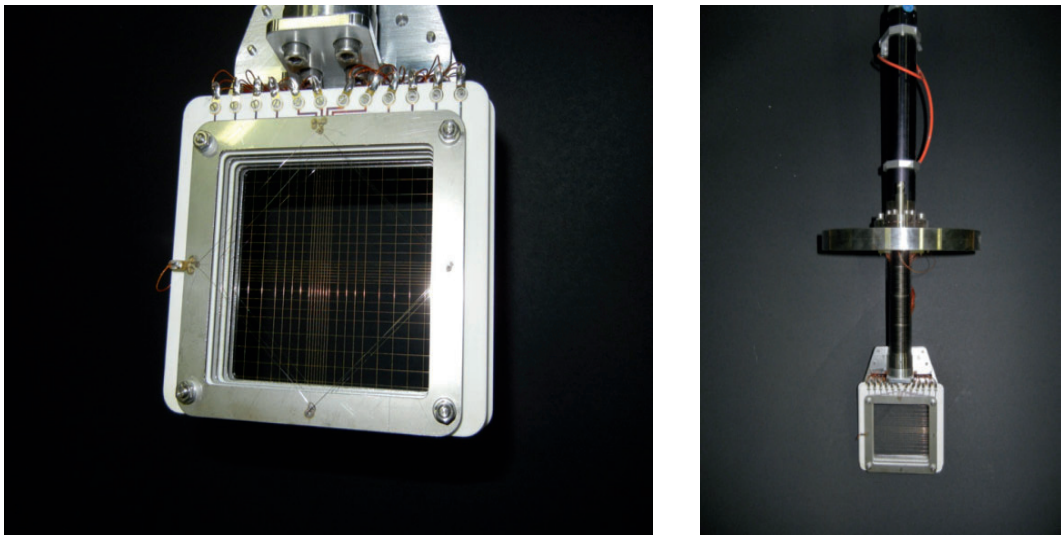


Fig. 7: Profile grids

4.2 Wire scanners

Instead of inserting a grid of wires, a single wire can instead be driven through the beam. Measuring the position of the wire and the signal induced on the wire at that position, a profile can be built up. The spatial resolution achievable is now determined by the minimum displacement of the wire, which is typically in the range below 100 μm. Wire scanners often use two wires assembled as a cross or L-

shape and are mounted at a 45° angle with respect to the beam (Fig. 8). In this way it is possible to measure horizontal and vertical profiles with a single movement through the beam.

Wire scanners are interesting because of their high resolution and rather low cost, since only a single electronic chain per measurement plane is needed. The long measurement times required in machines with low duty cycles can, however, be a problem. In a machine pulsing at 1 Hz, a profile with 60 measurement points will take a minute to be acquired, during which time the beam profile and position need to remain constant.

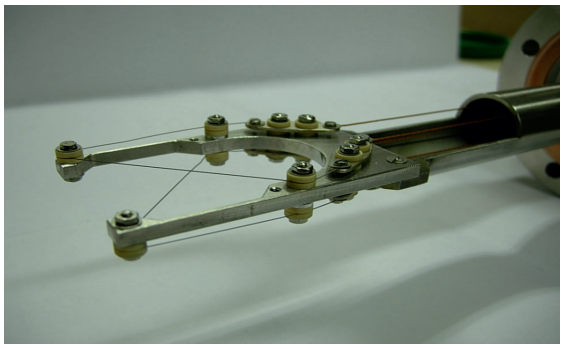


Fig. 8: Profile measurements with a wire scanner

4.3 Non-interceptive profile monitors

Some of the modern high-intensity proton or H^- machines provide currents and duty cycles so high that interceptive devices will be destroyed within a single beam pulse. Continuous observation of the beam profile in such a case is only possible with non-interceptive devices, which are based on the ionization of the rest gas or on the excitation of the rest-gas molecules resulting in luminescence. Using a laser to mimic a solid wire scanner is also a possibility, but is limited to H^- and electron machines. In the former case, the laser is used to strip the electron from the H^- ion, with the number of electrons detected at each position used to build up a profile; while in the latter case, Compton-scattered photons are detected.

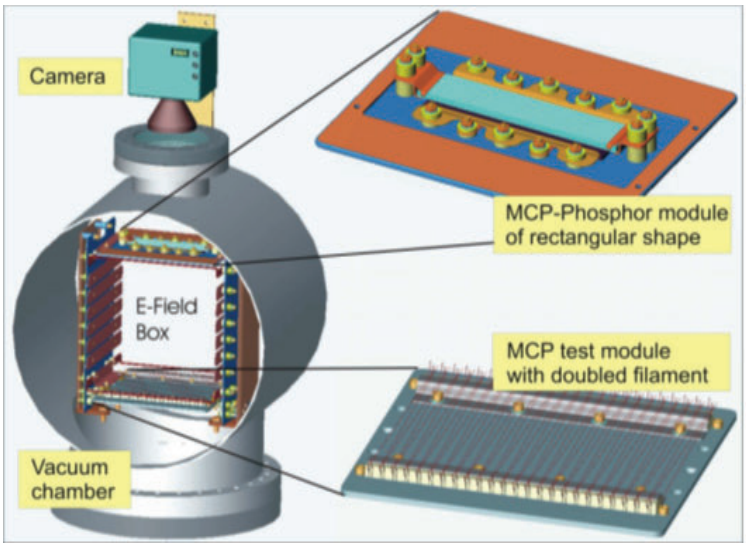


Fig. 9: Ionization profile monitor

In an ionization profile monitor (Fig. 9), the ions or electrons created by the interaction of the primary beam with the rest gas are guided by a homogeneous electric field to a multichannel plate (MCP), where the signal is amplified. The electrons created in the MCP are then either sent to a luminescent screen observed by a camera or directly detected using strip detectors. If the vacuum level is very good, the signal may be low and gas injection may be needed. At very high intensity, space-charge effects may blow up the secondary electron beam while it is guided to the MCP, resulting in an over-estimation of the primary beam size.

The luminescence monitor (Fig. 10) makes use of the light emitted through de-excitation of rest-gas molecules that have been excited through the passage of the primary beam. Here the setup is very simple: only a viewport and a camera are needed. However the cross-section for excitation is much lower than for ionization, and gas injection is almost always a necessity. In addition, the initial excited state can be relatively long-lived, such that the molecule can drift a significant distance before emitting the photon, which can result in distortion of the profile.

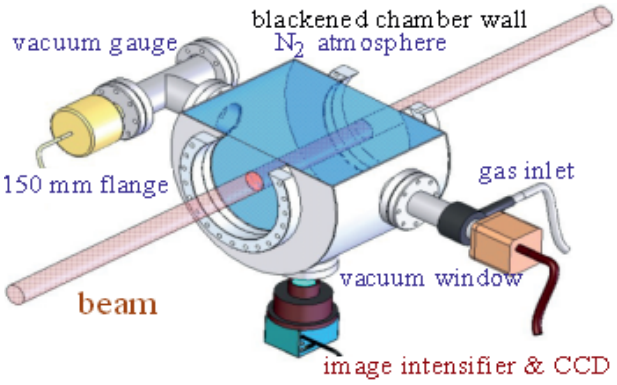


Fig. 10: Luminescence monitor

5 Emittance measurements

5.1 Slit/grid device

To measure the phase space distribution of the beam requires not only the position distribution (transverse profile) of the particles but also their angular distribution. A typical method consists of selecting a portion of particles at a specific position within the beam through insertion of a slit and measuring the angular distribution of these particles using a wire grid after they have traversed a suitably long drift space. The slit is slowly moved across the beam and for each slit position the angular distribution is measured. Each of these angular distributions forms a vertical slice in the x/x' (or y/y') phase space. For this reason the method is also called a *phase space scan*.

In the instrument that is shown in Fig. 11, both the L-shaped slit, mounted at 45°, and the two wire planes (horizontal and vertical) are controlled by stepping motors with a positioning resolution of 50 µm. The analogue signals from the wire grids are sampled with a time resolution of 6 µs, which permits one to observe phase space variations along the beam pulse, which in this case was several hundred microseconds in length. Owing to the mechanical complexity, the number of readout channels and the positioning precision required, such a device is rather expensive but allows very fine scanning of the phase space.

The method is only practicable at very low energy, because otherwise the conversion of angular to spatial distribution would require very long drift spaces.

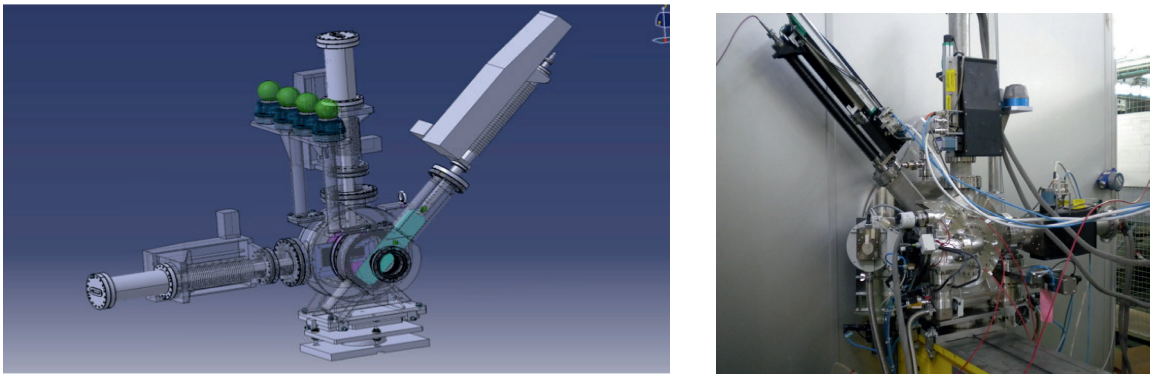


Fig. 11: Slit/grid emittance meter

When using a proton source, several ion species (proton, H^0, H_2^+, H_3^+) are created and are separated in phase space in the LEBT by a solenoid magnet (see Fig. 2). These ion species can then be clearly identified in the phase space plot measured after the first solenoid (Fig. 12).

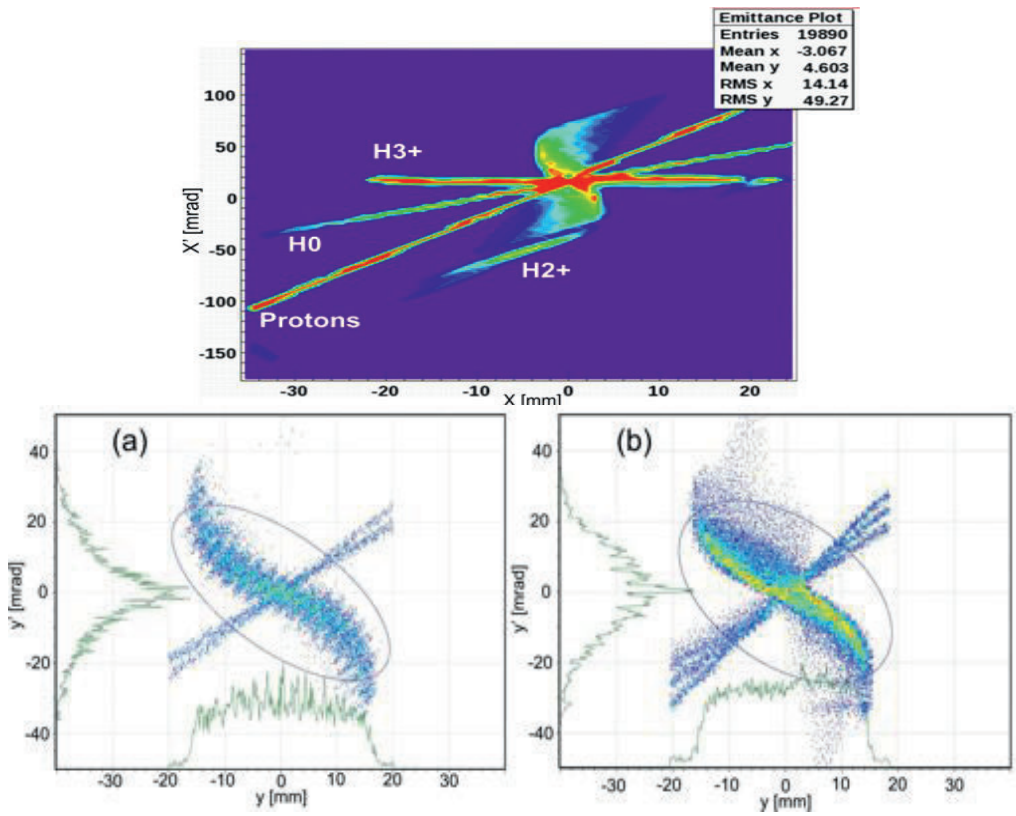


Fig. 12: Emittance meter results and simulations

Using the measured phase space parameters of the proton beam, the Twiss parameters (phase space distribution) at the exit of the source can be calculated using optics simulation programs. Mixing into the proton distribution a small amount of other ion species and following the mixed beam back to the position of the wire grid allows the simulated beam to be compared with the measurement, showing an impressive agreement.

5.2 The Allison scanner

The same principle of phase space scanning is used by the Allison scanner (Fig. 13), where the angular distribution is measured by an electrostatic deflector and a Faraday cup. The whole detector is passed through the beam, and for each position the angles are scanned by sequentially increasing the voltage on the deflection plate.

Again the Faraday cup signal is time-resolved with a resolution similar to the slit/grid device. In Fig. 13 it can clearly be seen that the phase space distributions at the beginning and the end of the beam pulse differ significantly from the distribution in the centre of the pulse. On the other hand, the plots are fairly stable within the rest of the pulse.

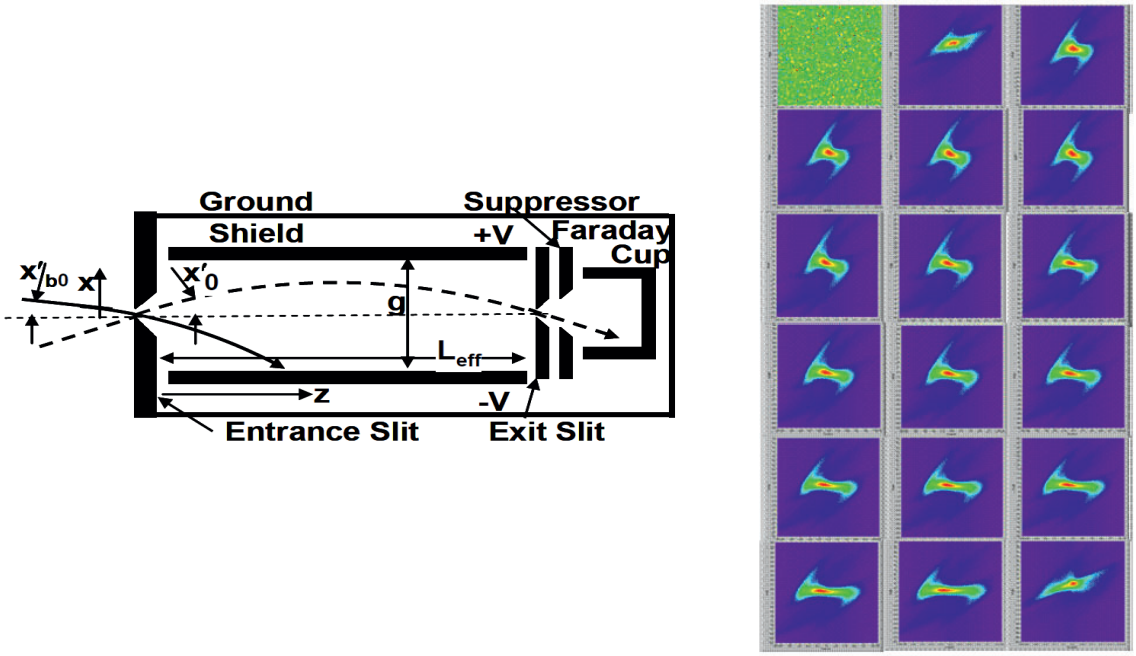


Fig. 13: Allison scanner

5.3 The pepperpot

The phase space scanning methods described above suffer from the long measurement times required to achieve a fine resolution.

The pepperpot eliminates this problem, allowing single-shot measurements to be performed. Here a plate with a rectangular arrangement of holes is inserted into the beam and the image of each hole is observed on a luminescent screen after a drift space using a camera (Fig. 14).

Projecting the contributions of each row/column of holes onto the *x*/*y* axis respectively allows the angular distribution at each row/column to be determined. Knowing the distance between the holes allows the pixel readout to be calibrated (ratio of pixels to millimetres), after which the emittance ellipse can be reconstructed.

Of course we must make sure that the individual hole images do not overlap, as this would make disentangling the contributions from each hole impossible. On the other hand, the distance between the holes should not be too big, otherwise a camera with a very large number of pixels would be needed. A good knowledge of the beam optics expected at the pepperpot location is therefore required before designing a pepperpot.

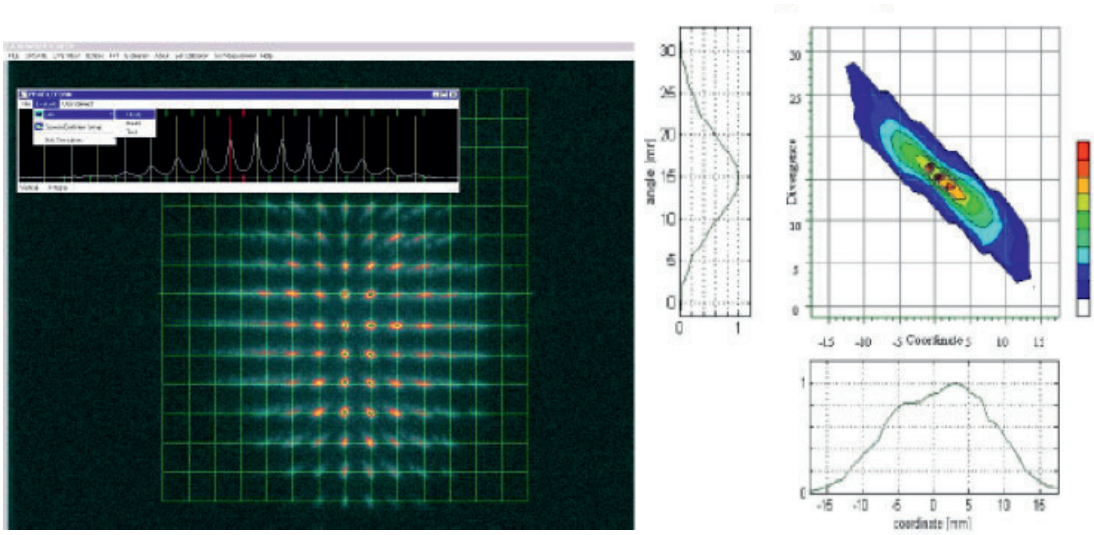


Fig. 14: Pepperpot measurements

The pepperpot also allows the observation of coupling between horizontal and vertical phase space by looking at a third projection, at 45° for example.

6 Spectrometer measurements

In order to determine the absolute energy as well as the energy spread of the beam, a spectrometer (bending) magnet whose magnetic field is precisely known is often used (Fig. 15). A vertical slit is placed in the beam, selecting only a small fraction of particles with the same horizontal position, which are bent in the spectrometer field and whose position after the bend is then measured. Depending on the particle energy, the bending angle will change, which results in a transverse profile in the horizontal plane that is proportional to the spread in particle energies.

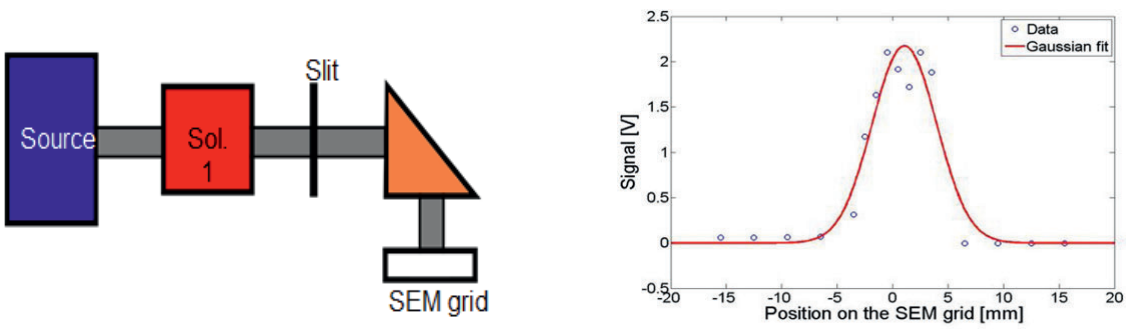


Fig. 15: Energy spread measured with a spectrometer

One way to calibrate the system is to modify the source extraction voltage and therefore the mean energy of the particles and observe the corresponding shift in the peak of the profile.

Another application of a spectrometer is for the selection of a particular charge state on a heavy-ion source (Fig. 16). The source is not able to fully ionize the atoms, which results in a charge state distribution of the extracted ions. Normally, only an ion with a precisely defined charge over mass ratio q/m can be accelerated in the following accelerator chain, and it is therefore advisable to select this charge state at the source in order to avoid radiation problems at higher energy. To measure the charge state distribution of the source, the ions are passed through a spectrometer magnet whose

magnetic field is ramped. At the same time, the ion current is measured in a Faraday cup located behind a slit situated after the spectrometer. The charge state spectrum of the ion source can therefore be measured (Fig. 16) and the ion with the desired charge state selected.

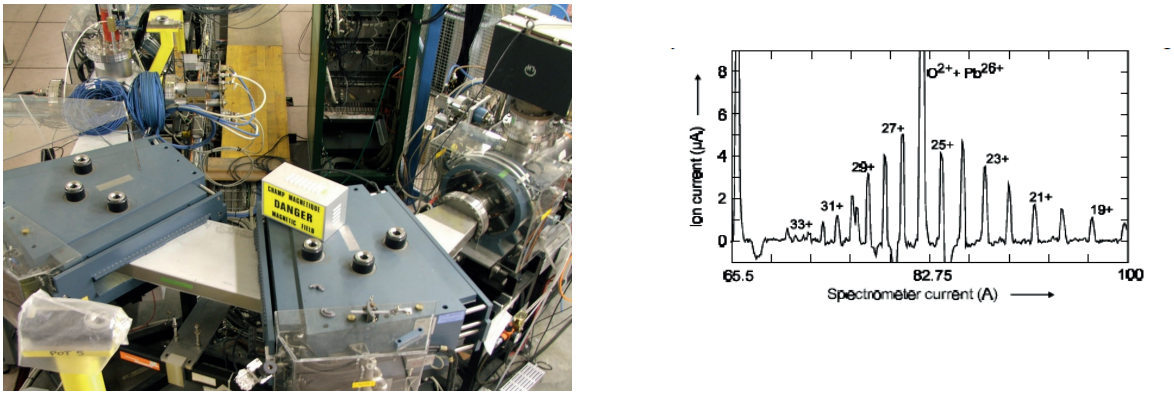


Fig. 16: Charge state spectrum measured with a spectrometer

7 Summary

This chapter gives an overview of the typical measurements needed at ion sources:

- intensity measurements;
- transverse profile, emittance measurements and determination of phase space parameters;
- energy and energy spread measurements.

The instruments used for each type of measurement are described, with their strengths and weaknesses discussed. The type of instrument selected depends on the characteristics of the beam to be measured in terms of beam current or brightness, duty cycle, etc. Interceptive devices, for example, may not be suitable because of the high energy deposition in the device, while non-interceptive methods may not work if the beam intensity is too weak or the vacuum requirements are too stringent to produce enough signal from the remaining rest gas. Finally Refs. [1–6] contain some further useful information.

References

[1] P. Forck, Lecture notes on beam diagnostics, Joint University Accelerator School (JUAS), March 2011; available at http://www-bd.gsi.de/conf/juas/juas_script.pdf.

[2] B. Cheymol *et al.*, First results from beam measurements at the 3 MeV test stand for CERN Linac4, Proc. DIPAC 2011, Hamburg, Germany.

[3] C.E. Hill and K. Langbein, Pulsed ECR source in afterglow operation at CERN, *Rev. Sci. Instrum.* **67** (1996) 1328 (doi:10.1063/1.1146709).

[4] M.P Stöckli *et al.*, Low-energy emittance studies with the new SNS Allison emittance scanner, Proc. PAC09, Vancouver, BC, Canada.

[5] T. Giacomini *et al.*, Ionization profile monitors – IPM at GSI, Proc. DIPAC 2011, Hamburg, Germany.

[6] C.A. Andre *et al.*, Beam induced fluorescence (BIF) monitors as standard operating tool, Proc. DIPAC 2011, Hamburg, Germany.

Modelling

P. Spädtke
GSI Darmstadt, Germany

Abstract
Modeling of technical machines became a standard technique since computer became powerful enough to handle the amount of data relevant to the specific system. Simulation of an existing physical device requires the knowledge of all relevant quantities. Electric fields given by the surrounding boundary as well as magnetic fields caused by coils or permanent magnets have to be known. Internal sources for both fields are sometimes taken into account, such as space charge forces or the internal magnetic field of a moving bunch of charged particles. Used solver routines are briefly described and some bench-marking is shown to estimate necessary computing times for different problems. Different types of charged particle sources will be shown together with a suitable model to describe the physical model. Electron guns are covered as well as different ion sources (volume ion sources, laser ion sources, Penning ion sources, electron resonance ion sources, and H^--sources) together with some remarks on beam transport.

1 Introduction

Modelling is the task of defining a correct picture of a physical device described by rules that can be handled by a computer. Because in these proceedings we are dealing with ion sources, we restrict ourselves here to ion sources, ion beams and related topics. Charged particles are influenced by electric and magnetic fields, and they themselves produce such fields. To solve the equation of motion, all the participating fields have to be known: the electric field $\boldsymbol{E}$ and the magnetic flux density $\boldsymbol{B}$. These fields might be either static or time-dependent; in general, they are not symmetric but are real three-dimensional fields. Depending on the number of charged particles, they will create their own fields: $\boldsymbol{E} = \boldsymbol{E}_{\mathrm{ext}} + \boldsymbol{E}_{\mathrm{int}}$, and $\boldsymbol{H} = \boldsymbol{H}_{\mathrm{ext}} + \boldsymbol{H}_{\mathrm{int}}$.

Before computers existed, analytical models were required to estimate unknown field distributions, or an experimental method like the electrolytic trough had to be used. Today, differential equations can be solved with numerical methods, and therefore it is possible to predict the behaviour of physical systems using digital computers. However, the applied models have to describe the real behaviour correctly. Let us assume a specific case to demonstrate the problem: a drifting ion beam. No external fields are present and no time dependence is assumed in this case.

2 Drifting ion beam

If all external fields are known, the solution of the initial value problem needs to be solved. The easiest case is when all the external fields disappear: $\boldsymbol{E} = 0$ and $\boldsymbol{H} = 0$. Let us assume a beam tube of 80 mm diameter and 1 m length (see Fig. 1). Discretization is achieved by a grid of 100 nodes in each Cartesian direction. This gives a longitudinal resolution of 1 cm, and a transverse resolution of 1 mm.

Because no external electric fields (grounded beam tube) and no internal magnetic fields (electric current is negligible) are present, the final coordinates x_f, y_f, z_f of each particle are easy to calculate, and are given by

$$x_f = x_0 + v_x t, \tag{1}$$

$$y_f = y_0 + v_y t, \tag{2}$$

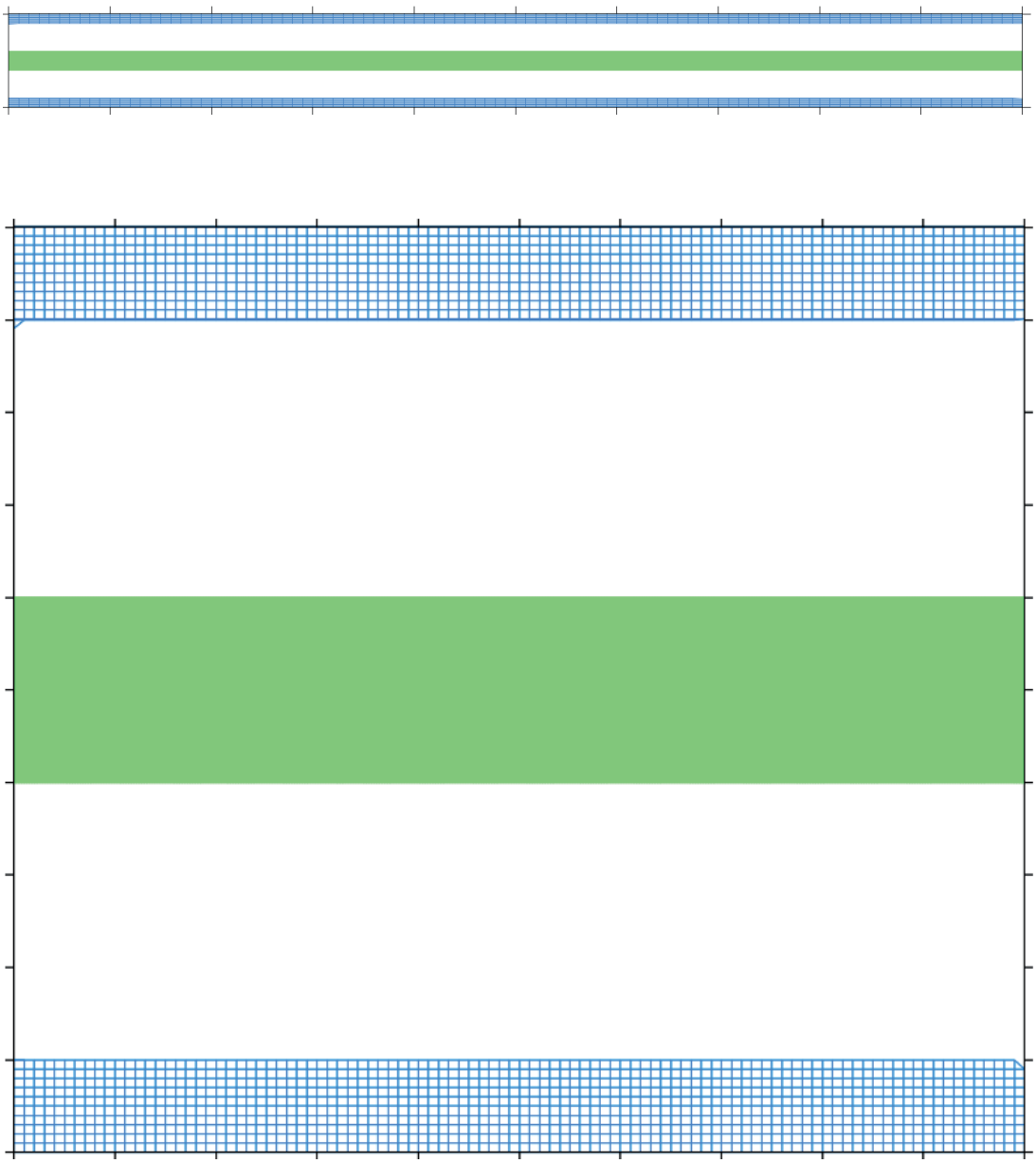


Fig. 1: Drifting ion beam in a beam tube of 1 m length and 80 mm diameter. The beam tube is indicated by the blue mesh. The ion beam drift energy of singly charged argon ions is assumed to be 35 keV. The trajectories are shown in green. Top: correct geometric scaling. Bottom: scaling differently in horizontal and vertical directions, to show more detail. Beam radius is 10 mm.

$$z_f = z_0 + v_z t. \tag{3}$$

If the initial conditions $(x_0, y_0, z_0, v_x, v_y, v_z)$ are known, the final coordinates can be calculated exactly, and the problem has been solved.

However, it is well known that charged particles will repel each other if they are of the same sign. These fields are generated by all participating particles, described by their space charge

$$\rho = \boldsymbol{I}/(\boldsymbol{\varepsilon}_0 \cdot \boldsymbol{v}), \tag{4}$$

where I is the current (A), v is the velocity of the charged particle (m s^{-1}), ε_0 the permittivity in vacuum, and ρ is the resulting space charge. The space charge of such a beam is shown in Fig. 2.

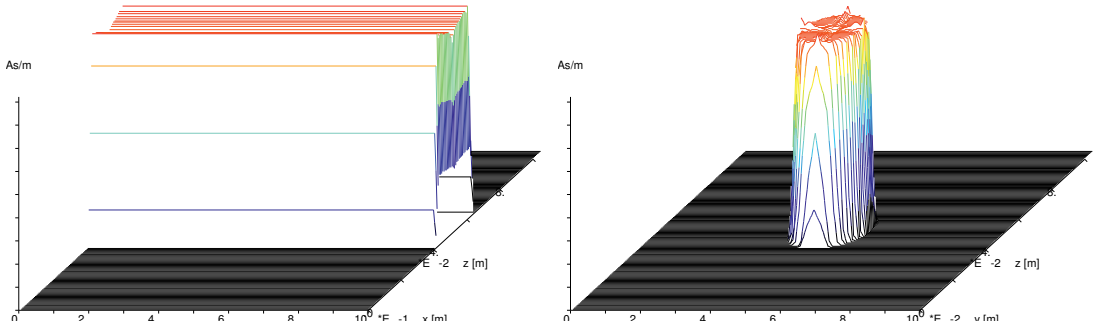


Fig. 2: Charge density distribution of the drifting ion beam: left, in the beam direction; right, perpendicular to the beam direction.

This space charge produces an electric potential Φ as shown in Fig. 3, described by Poisson's equation, for which the Cartesian formulation is shown in Eq. (5):

$$\partial^2\Phi/\partial x^2 + \partial^2\Phi/\partial y^2 + \partial^2\Phi/\partial z^2 + \rho/\varepsilon_0 = 0. \tag{5}$$

The resulting electric fields will influence the particles themselves. Only a few problems can be solved analytically, for example, an infinitely long ion beam, homogeneously distributed space charge, drifting with a constant velocity.

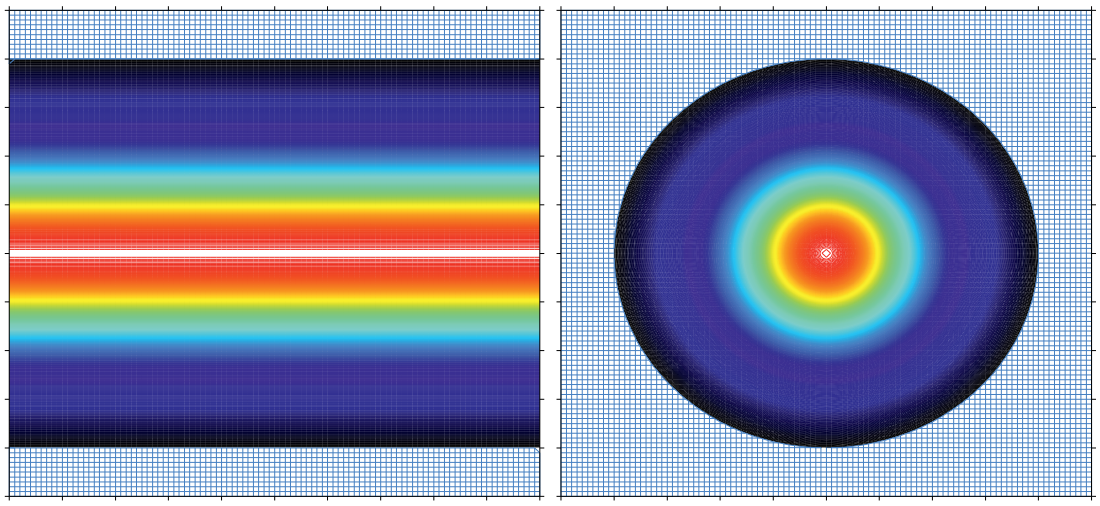


Fig. 3: Electric potential caused by the charge density, as shown in Fig. 2, $I = 5$ mA

If this is not the case, the differential equation (5) is to be replaced by a difference equation for each node point. Boundary conditions have to be defined at all six surfaces of the surrounding box. Each difference equation can then be solved consecutively. With a successive over-relaxation method, the set of equations can then be solved much faster. The result of such a simulation is shown in Fig. 3. Differences in computing time are demonstrated later in Fig. 12.

This simulated potential distribution can be compared with the analytic expression. For an infinitely long, cylindrical ion beam with zero emittance, carrying current I with constant drift velocity, the potential Φ can be calculated analytically as a function of the radius r:

$$\Phi(r) = \frac{I}{4\pi\varepsilon_0 v}\left(1 + 2\ln\frac{r_\mathrm{b}}{r} - \frac{r^2}{r_\mathrm{b}^2}\right) \qquad \text{if } r \le r_\mathrm{b}, \tag{6}$$

with velocity v, permittivity ε_0 and beam pipe radius r_b. Outside the beam the potential is

$$\Phi(r) = \frac{I}{2\pi\varepsilon_0 v}\left(\frac{1}{2} + \ln\frac{r_\mathrm{b}}{r}\right) \qquad \text{if } r \geq r_\mathrm{b}. \tag{7}$$

For a singly charged argon-ion beam of 5 mA and an energy of 35 keV, this simulation method predicts 412 V on the axis for the given ratio of beam tube diameter to beam diameter.

The magnetic field of the drifting ion beam (an azimuthal component $B_\Theta(r)$ of about 1×10^{-6} T at the beam edge) can be neglected at velocities of the order of $\beta \approx 0.1\%$:

$$B_\Theta(r) = -\frac{\mu_0 I}{2\pi r}. \tag{8}$$

Using Eqs. (6) and (7), one can check that the simulation yields the correct space-charge behaviour, but, of course, this solution is not a self-consistent solution. The interaction of the electric field with the particles is neglected. The self-consistent solution (Fig. 4) can be obtained by the successive repetition of solving Poisson's equation and the equation of motion, which creates the space-charge map. This procedure converges in most cases to the self-consistent solution.

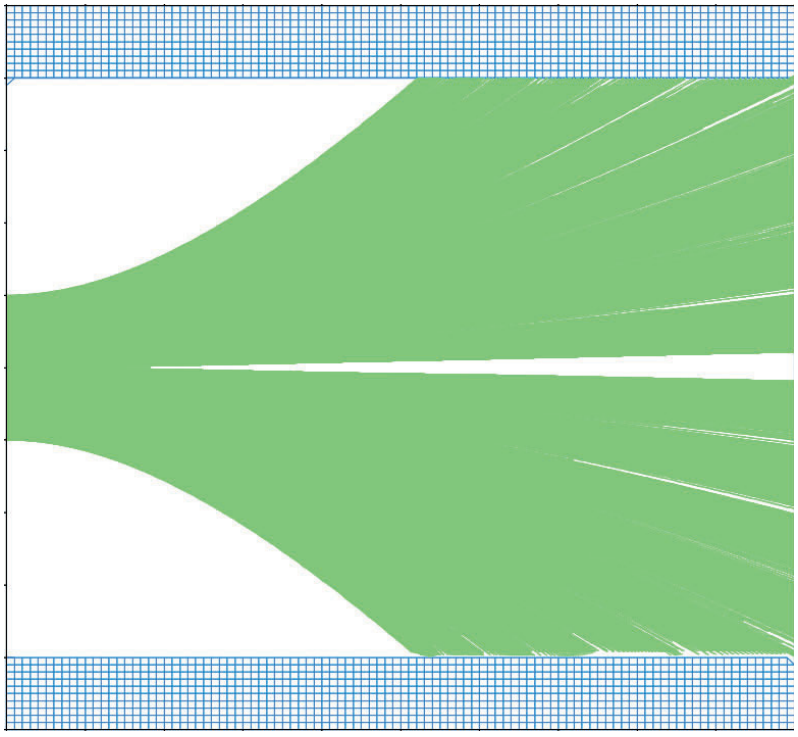


Fig. 4: Self-consistent solution of the drifting ion beam of 5 mA without any space-charge compensation

Besides the repulsive force of particles with the same sign of charge, the force is contractive for charges of different sign. A positive ion beam tries to attract electrons, and if the space charge of the ions cancels the space charge of the electrons, no repulsive force is acting at all. This process is called space-charge compensation [1]. The effect can be and has been measured [2], but how does one model it? The simplest method is to assume a net current.

2.1 Linear compensation

Instead of taking the full current into consideration (5 mA), only the uncompensated fraction is used for the simulation. Assuming a space-charge compensation of 90%, the current used for the simulation

is only 0.5 mA. This assumption, however, excludes nonlinear interaction of both particle groups. The resulting trajectory plot is shown in Fig. 5. Different projections of the 6D phase space into 2D subspaces are shown in Fig. 6

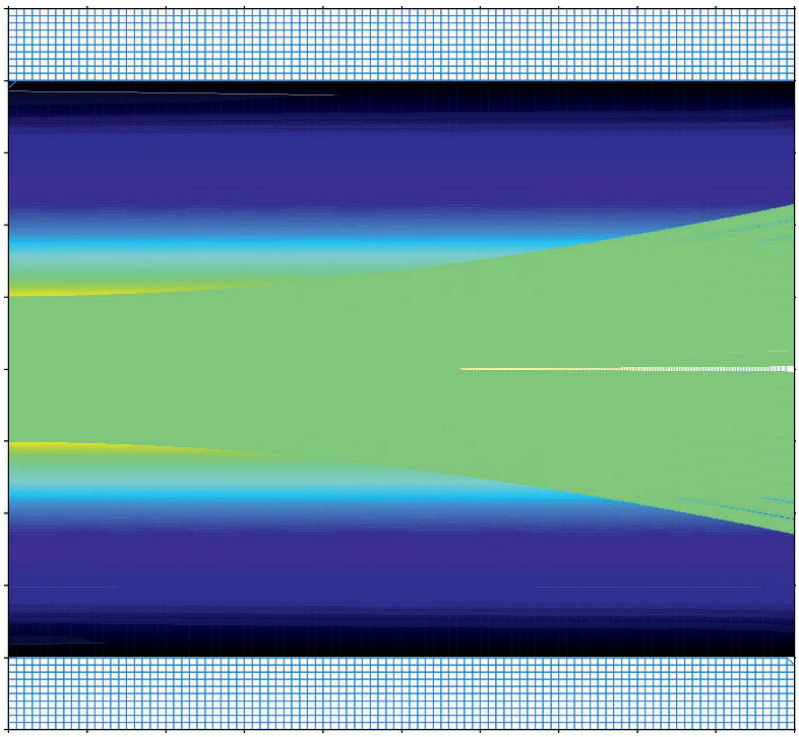


Fig. 5: Self-consistent solution of the drifting, partially compensated ion beam, plotted together with the potential distribution, assuming a linear space-charge compensation. Active current 0.5 mA, corresponding to 90% space-charge compensation.

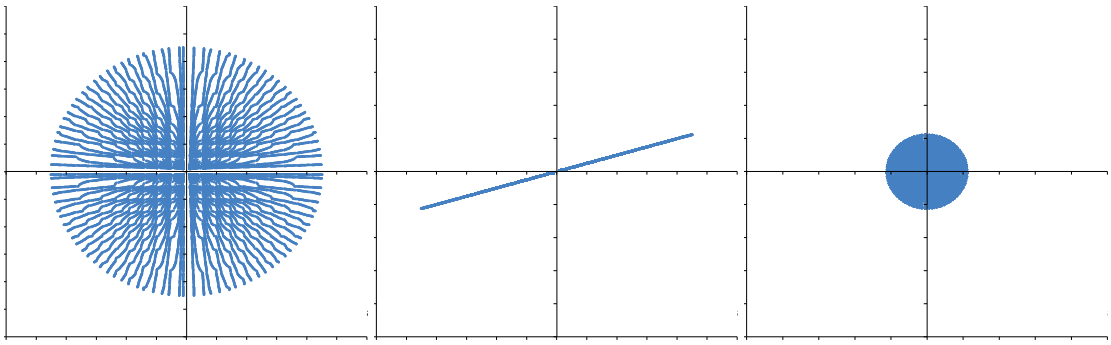


Fig. 6: Beam profile, emittance and momentum space at the end of the drift section, assuming 90% linear space-charge compensation.

2.2 Ion beam plasma – analytic solution

A better approach is to assume that the remaining uncompensated charges lead to a beam plasma potential. This assumption creates a field-free region within the beam, but with an electric gradient at the beam boundaries. The trajectory plot is shown in Fig. 7, the resulting potential distribution is shown in Fig. 8, and different projections of the 6D phase space onto 2D subspaces are shown in Fig. 9. Instead of

defining the degree of space-charge compensation, the value of the resulting space-charge potential has to be given. In our case, a beam plasma potential of 50 V would roughly correspond to 90% space-charge compensation.

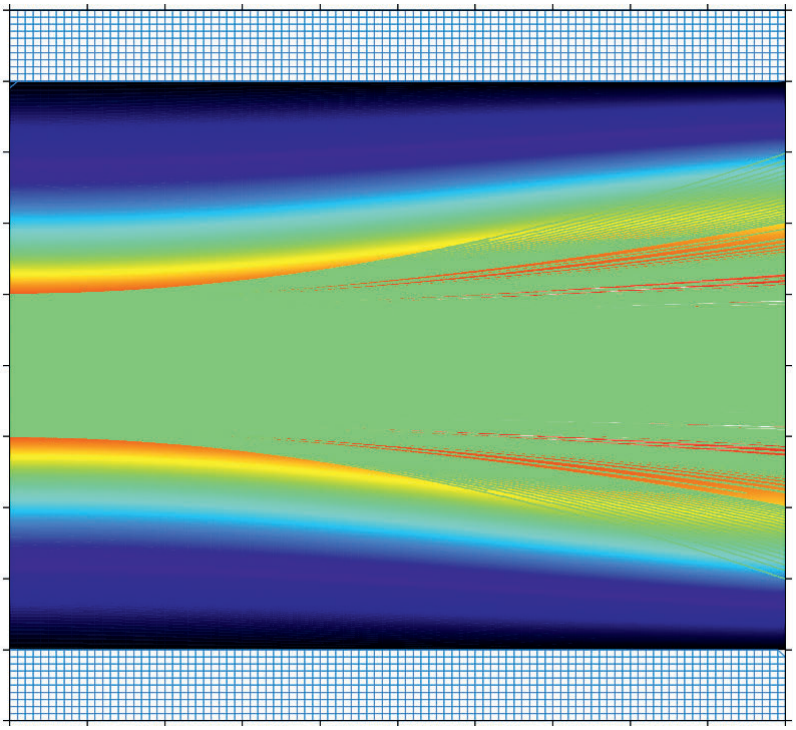


Fig. 7: Self-consistent solution of the drifting ion beam, plotted together with the potential distribution, assuming a beam plasma of 50 V.

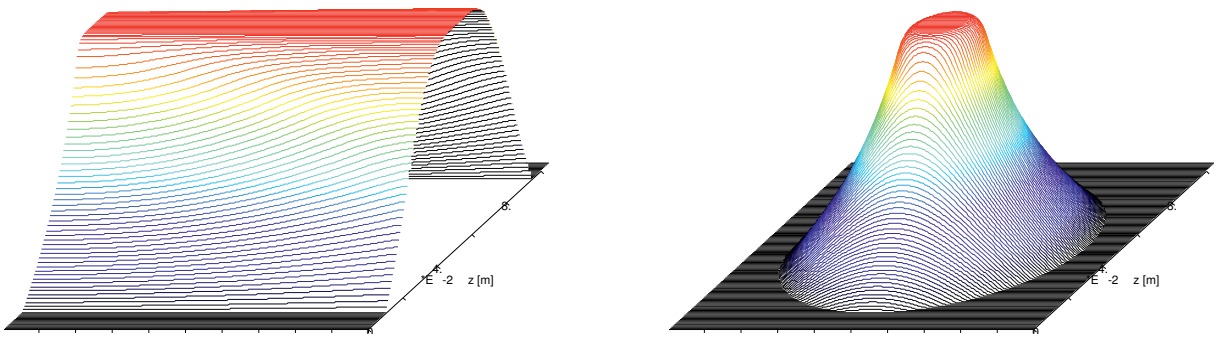


Fig. 8: Self-consistent potential distribution generated by the drifting ion beam in beam direction (left), and perpendicular to the beam direction (right).

2.3 Creation of secondaries

A third approach is to model the generation of secondaries. In that case the neutral gas pressure in the beam line has to be known, as well as the cross-sections for the different assumed processes. Ion–ion collisions might take place, as well as electron–ion collisions and ion–neutral collisions, to mention only a few of the possible interactions. In Fig. 10 a realistic neutral gas density distribution is shown. Together with correct cross-section and the primary ion intensity, the rate of secondaries can be calculated. The

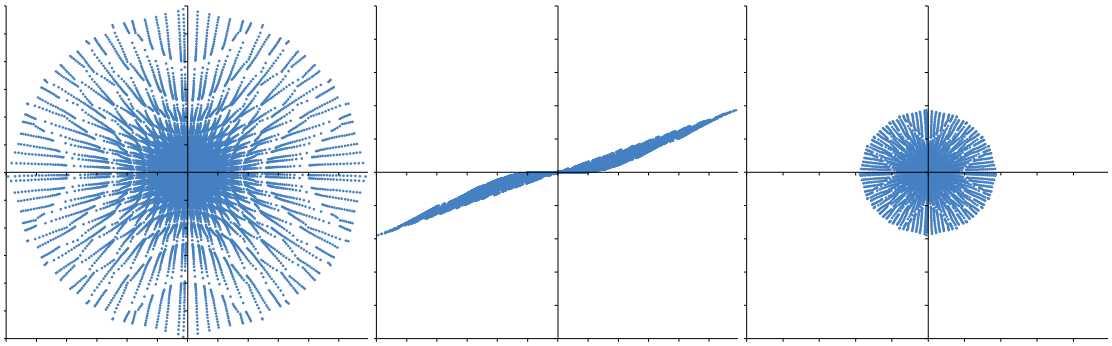


Fig. 9: Beam profile, emittance and momentum space at the end of the drift section for the beam plasma compensation model.

resulting secondaries can be used to distribute their space charge in addition to the space charge of the primary ions.

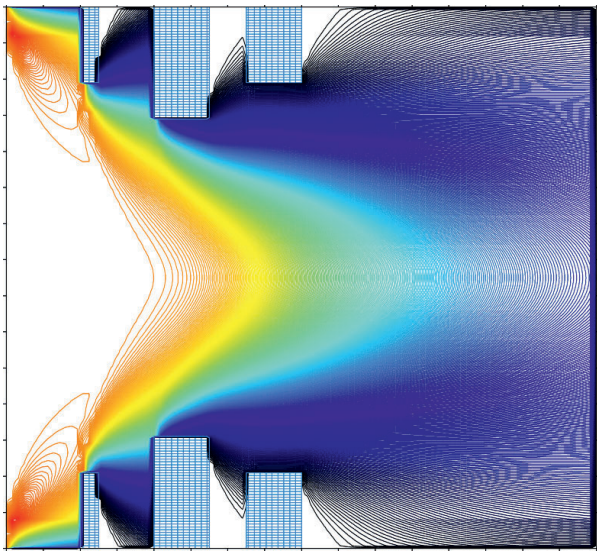


Fig. 10: Neutral gas pressure caused by the ion source, which can be used for the generation of secondaries

Because the number of secondaries for each primary trajectory should be large to minimize the discretization error, the total number of trajectories could become too high for a reasonable solution time.

When the development with time is to be simulated, another dimension has to be added, and a particle-in-cell (PIC) code is required. An example is given in Ref. [3].

Up to now, we have assumed that the beam is infinitely long, and its generation is located in an infinite distance. In most cases this is not true, and any modelling needs to assume something like a source for the simulated ion beam. Therefore the question arises how to model this. This will be covered in section 4.

3 Solver

Some general remarks should be made about the method of solution. Different numerical methods are available to solve a huge set of equations. Each method may have its own advantages and disadvantages. For the kind of modelling we are talking about, a point-to-point solver has the big advantage that analytic

expressions can be used on each point, for example, to calculate the electron density term as defined in Eq. (10) or the secondary ion density term in Eq. (12). In the case of a direct solver, the iterative determination of electron density would not be possible and the correct simulation of a plasma boundary would fail.

The rate of convergence depends on several aspects: the total number of nodes to be calculated (see Fig. 11), the question whether different mesh distances are used or not, and whether an over-relaxation method is applied (see Fig. 12).

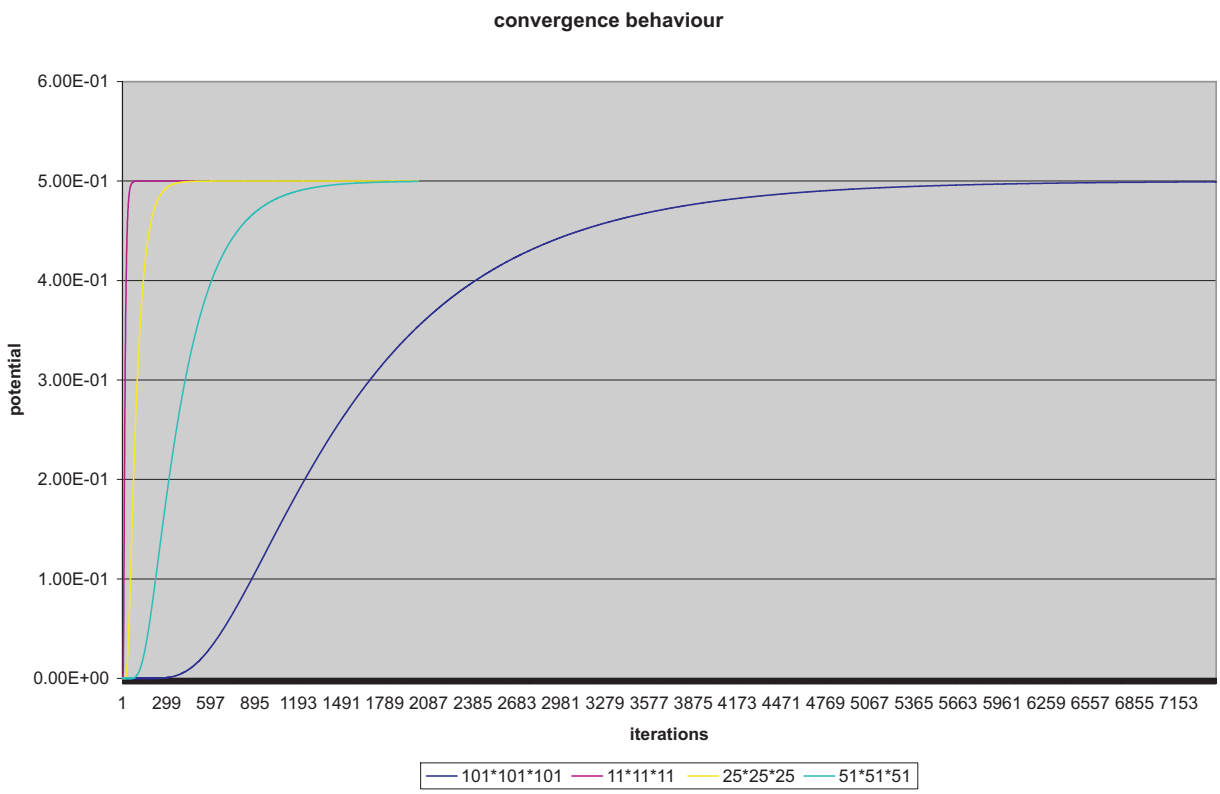


Fig. 11: Convergence behaviour: number of nodes. The more nodes that are to be calculated, the more iterations are required to achieve sufficient precision [4].

The necessary CPU time to run such a benchmark is given in Table 1: a Laplace solver on a 2D grid with 101×101 nodes and defined boundary condition.

4 Particle sources

Even though the general method to extract a charged particle from a particle source seems to be similar (biasing the particle source positively, to extract positive charges from the source), mostly any particle source requires its own model.

4.1 Fixed cathode

The first model that was implemented in computer codes was the model of cathode emission of electrons from a solid cathode surface (Fig. 13). This model is also applicable to surface ionization sources. The Child–Langmuir law (9) describes the particle density or here the electron current as a function of the applied extraction voltage:

$$j = \frac{4\varepsilon_0}{9}\sqrt{\frac{2q}{m}}\,\frac{\Phi^{3/2}}{d^2}. \tag{9}$$

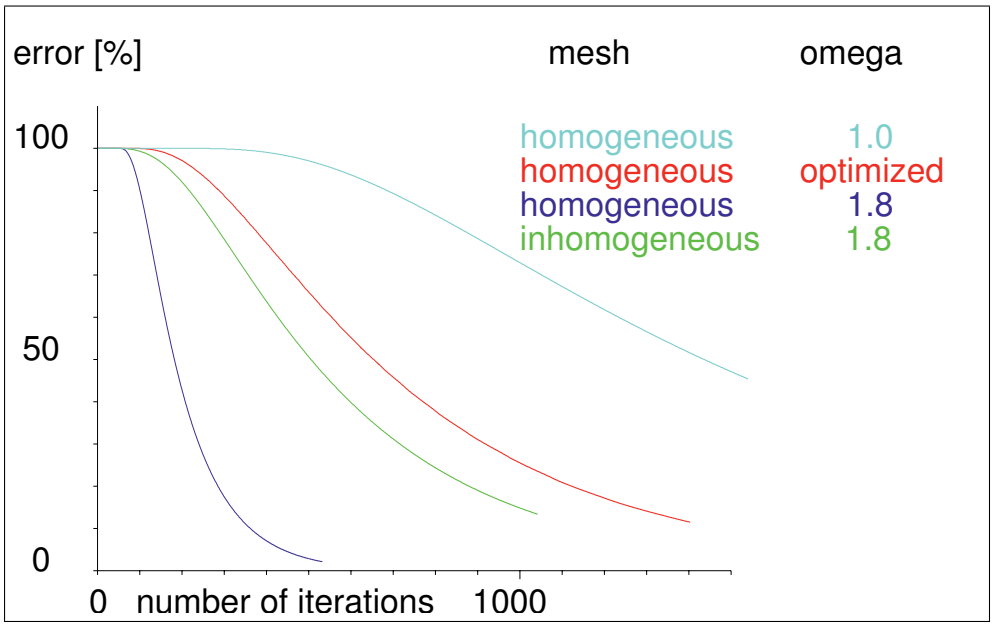


Fig. 12: Convergence behaviour: deviation of the analytical correct value in the centre of a simulation as a function of the iteration count. The influence of a mesh with variable mesh distances or constant mesh distances and the influence of the over-relaxation parameter is shown.

Table 1: Necessary CPU time for different hardware. The asterisk indicates that an interpreter instead of a compiler has been used. All rows beside 'today's laptop' have been compiled in 1987 already [5].

Computer	Language	Precision	
		single	double
McIntosh	Pascal*	60 000	—
McIntosh	Basic*	54 000	—
Commodore C64	Assembler	18 000	—
Atari ST	GFA Basic*	7720	—
Atari ST	Assembler	1800	—
HP9816	Basic*	7710	—
IBM PC	Fortran	7055	—
IBM XT/8087	Fortran	1723	—
VAX 8700	Fortran	10.5	15.8
IBM 3090-200	Fortran	2.61	—
Cray XM-P	Fortran	2.2	—
Today's laptop	Fortran	0.031	0.031

A successive iteration scheme (solving the Poisson equation, followed by solving the Lorentz equation) will modify Φ and therefore the current density. The modelling requires a slow increase of the current density, so as not to run into convergence problems. It should be mentioned that this assumption is valid only if the emission is space-charge-limited, and not emission-limited.

In the second case, the achievable current density is of course below the current density described in Eq. (9). With this restriction, the model can also be applied for the simulation of surface ionization sources.

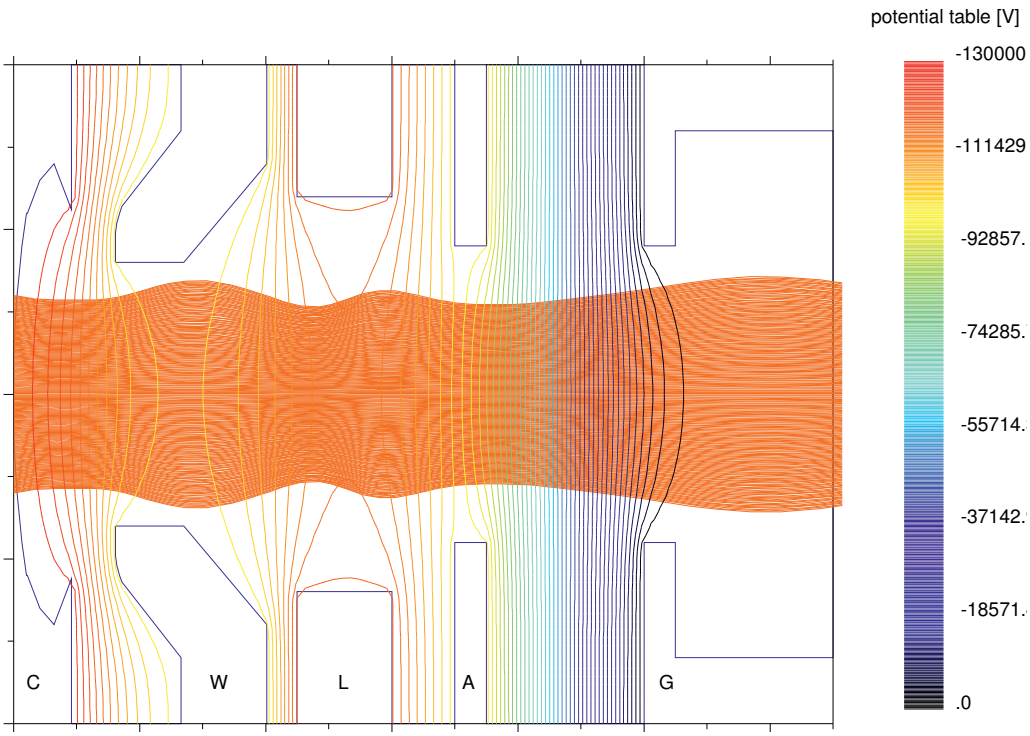


Fig. 13: Electron beam coming from a thermionic cathode: C, cathode; W, Wehnelt cap; L, lens; A, anode; G, ground.

4.2 Plasma sources – no magnetic field

4.2.1 Volume ion sources

Volume ion sources have a radial magnetic confinement, but the magnitude of $\boldsymbol{B}$ is negligible for extracted ions close to the axis. Without extraction voltage, the plasma would penetrate into the extraction system, but with positive biasing of the ion source, ions are retracted from the ion source, whereas electrons are reflected back into the source plasma. This can be modeled using Self's [6] theory, describing the electron density n_e as a function of the potential Φ:

$$n_\mathrm{e} = n_{\mathrm{e},0} \exp\left[-\frac{q(\Phi_\mathrm{pl} - \Phi)}{kT_\mathrm{e}}\right], \tag{10}$$

where $n_{\mathrm{e},0}$ is the electron density within the undisturbed plasma with plasma potential Φ_pl, k is Boltzmann's constant and T_e is the electron temperature. Using Eq. (10) an electron term can be added to the space charge ρ in Eq. (5). The influence of space charge and space-charge compensation can be seen in Fig. 14, where the current density is increased from zero to 1000 A m^{-2} (0, 16.8, 75, 100, 500, 750, 1000 A m^{-2}). For discretization, 100 nodes are used in the longitudinal direction and 80 nodes in each

transverse direction. The resolution is then in the range of 200 μm. The potential at the three electrodes is 20, −20 and 0 kV.

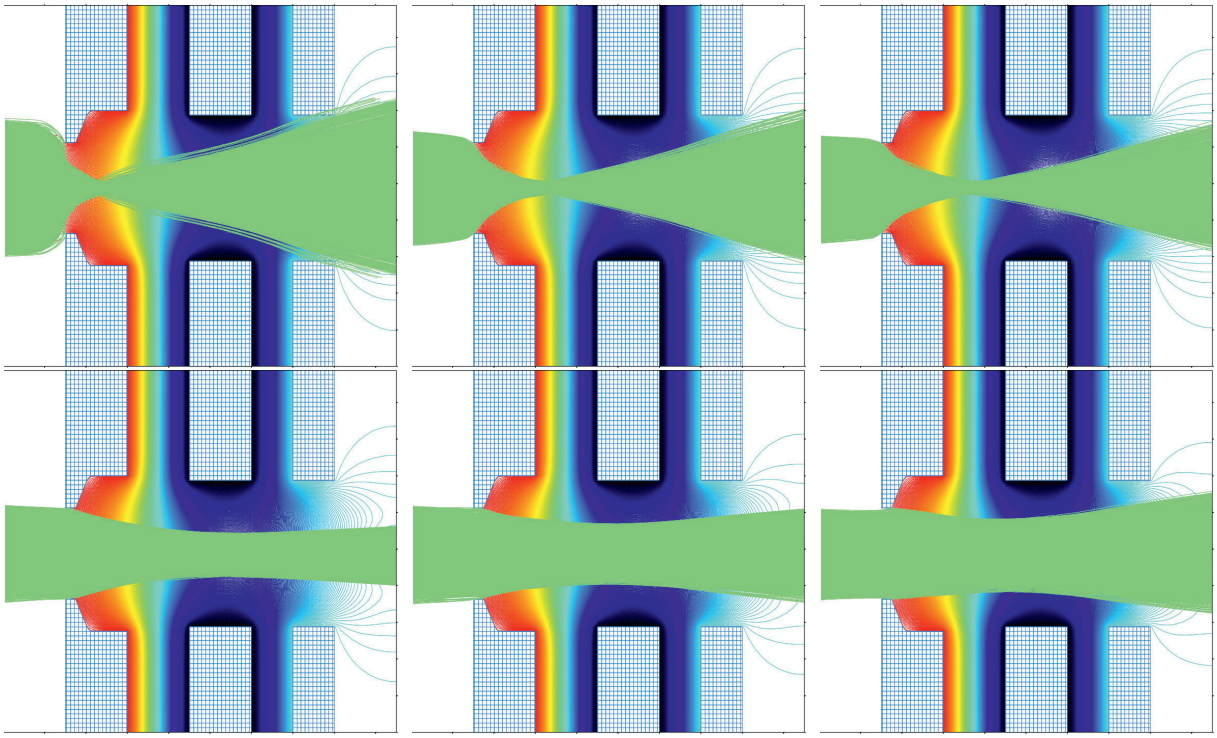


Fig. 14: Beam profiles with increasing particle density (from top left to bottom right)

The plasma boundary develops as shown in Fig. 15. The plasma boundary is visualized by plotting a few potential lines close to the plasma potential.

4.2.2 Laser ion sources

Laser ion sources creates a plasma by focusing a powerful laser beam onto a solid target (Fig. 16). The temperature of the target close to the focal point increases drastically until the target material evaporates and becomes ionized. The starting velocity perpendicular to the target surface is not negligible.

Furthermore, the time of laser irradiation is very short, of the order of picoseconds. If a drift section inside the plasma chamber (before extraction) is used to increase the pulse length, the extraction voltage has to be ramped to avoid a different energy along the pulse. A large momentum spread is typical for laser ion sources. Longitudinal effects become important, and a particle-in-cell (PIC) code might become necessary. As long as these effects can be neglected, a trajectory code can be used when the additional starting velocity is taken into account.

In case of a much more powerful laser (Fig. 17), not only can higher particle densities be achieved, but also the first section of the linear accelerator can be replaced. High energies, up to the MeV range, have been measured for protons created by laser radiation [8]. For such a simulation not only are protons with a large momentum distribution taken into account, but also co-moving electrons that compensate the space charge of the ions.

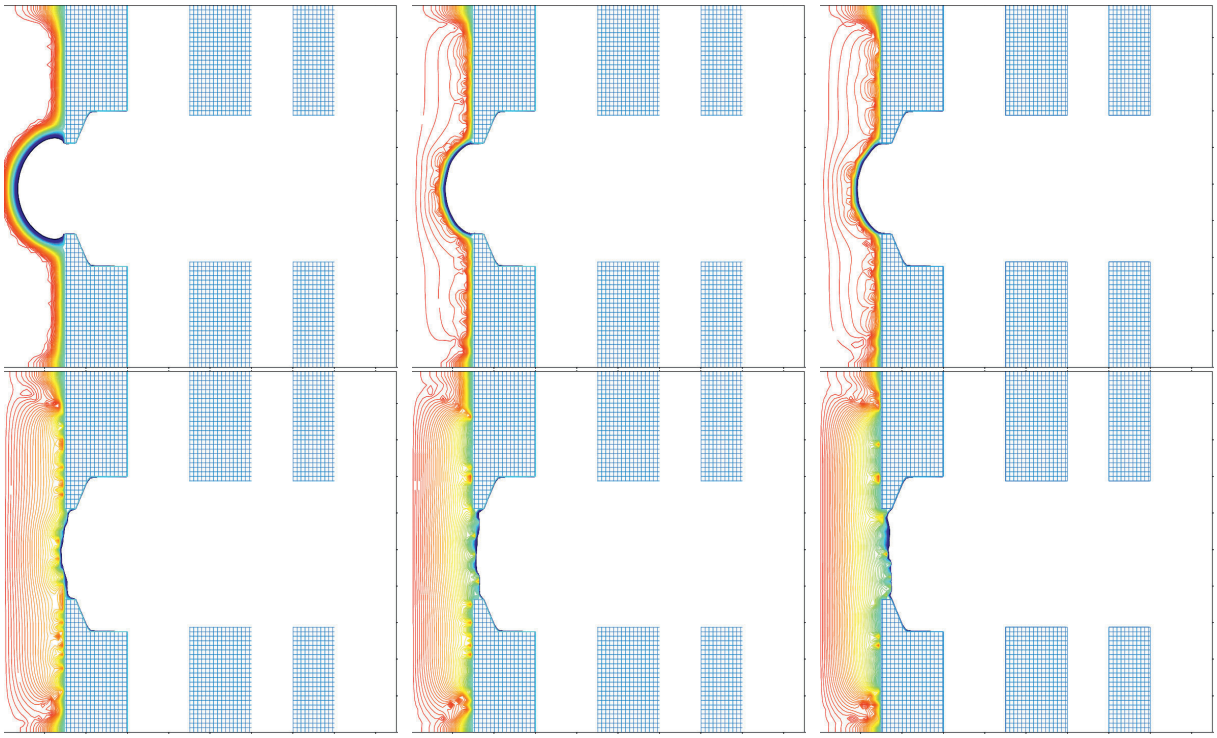


Fig. 15: Plasma boundary with increasing particle density (from top left to bottom right)

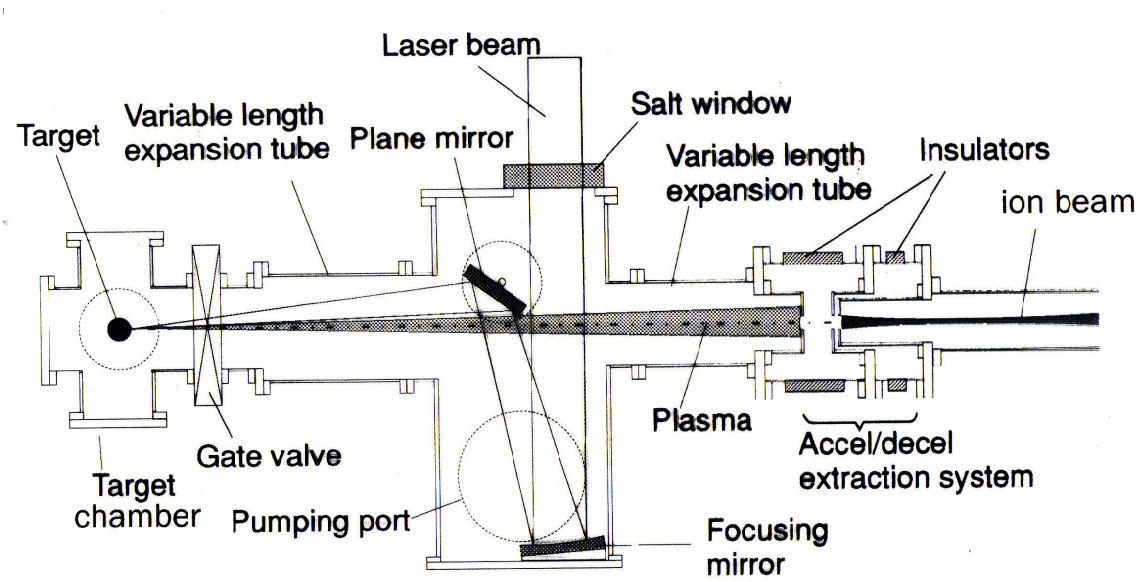


Fig. 16: A CO_2-laser ion source [7]. The long drift space from the laser target to the extraction system allows the particles to de-bunch. As a result, the high current density decreases, but the pulse length increases.

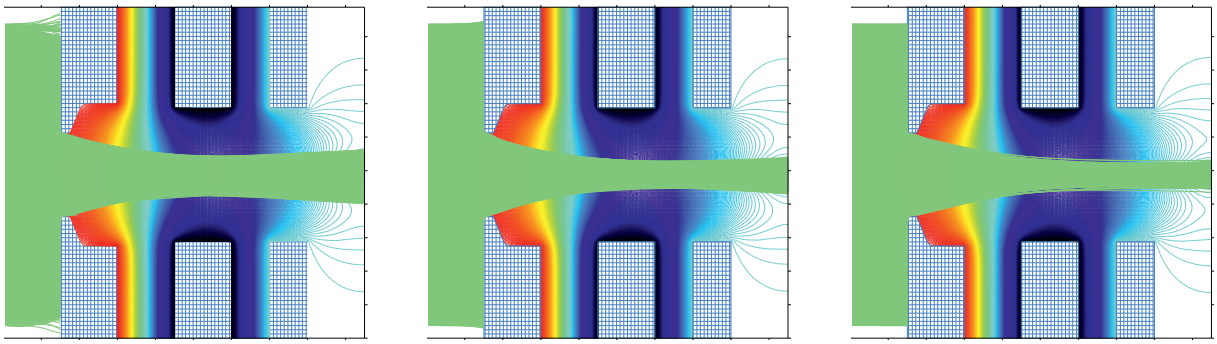


Fig. 17: With increasing laser power, the ions have higher velocity

4.3 Plasma sources – with magnetic field

4.3.1 PIG ion sources

There are different technical solutions available for a Penning ionization gauge (PIG) ion source (Fig. 18). One main classification might be the direction of extraction, which is in either the radial or the axial direction (with respect to the anode). Here only the radial extraction system is covered [9]. If the source is operated in sputter mode, an additional electrode is introduced into the plasma chamber, electrically insulated against the annular anode tube. Typically this sputter electrode, made from the desired material, is biased negatively with respect to the anode. A certain fraction of the ions that are present from the main gaseous discharge inside the anode chamber will be accelerated towards the sputter electrode. Depending on the material, the sputter voltage and the mass of the primary ions, neutral atoms can reach the discharge chamber, where they eventually become ionized.

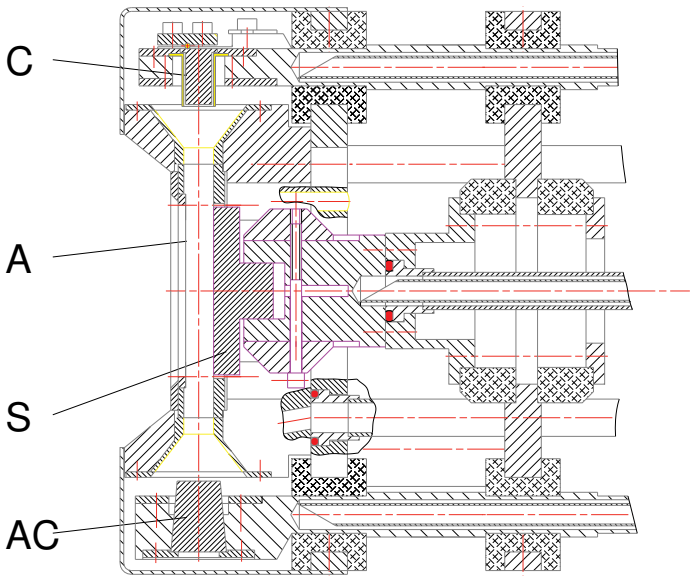


Fig. 18: Radial extraction from a Penning ion source: C, cathode; AC, anti-cathode; S, sputter electrode; A, anode

Simulations have been made for gaseous operation (see Figs. 19, 20 and 21) and for the sputter mode (see Figs. 22 and 23). The magnetic field direction is within the plane of plotting in the direction from the cathode to the anti-cathode.

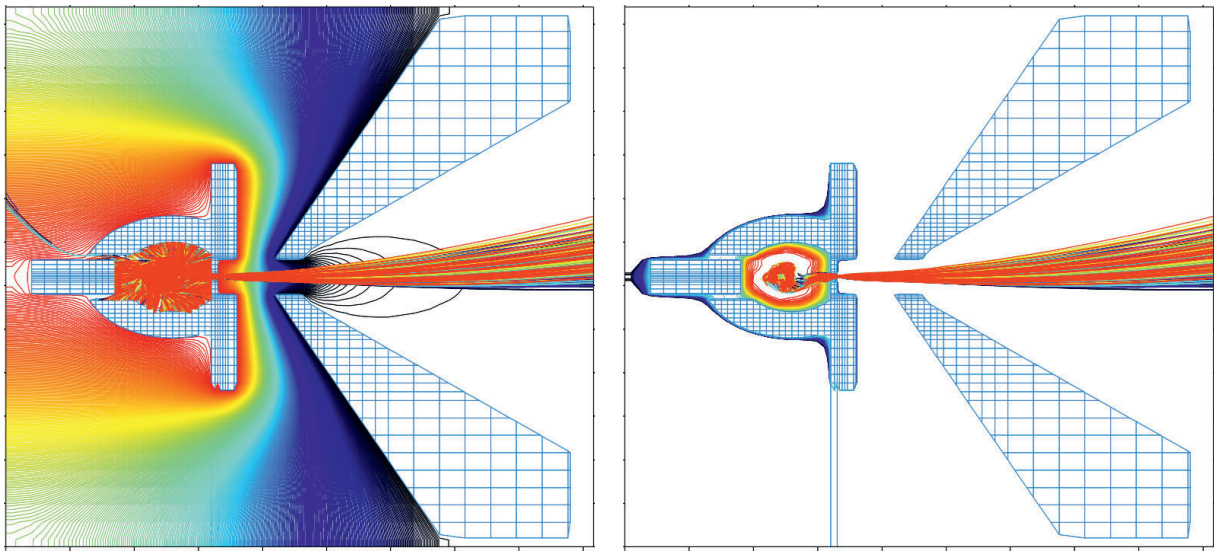


Fig. 19: Radial extraction from a Penning ion source (no sputtering). Left: all trajectories plotted; right: only ion trajectories that are extracted are shown. Low-current mode (≈ total 50 μA).

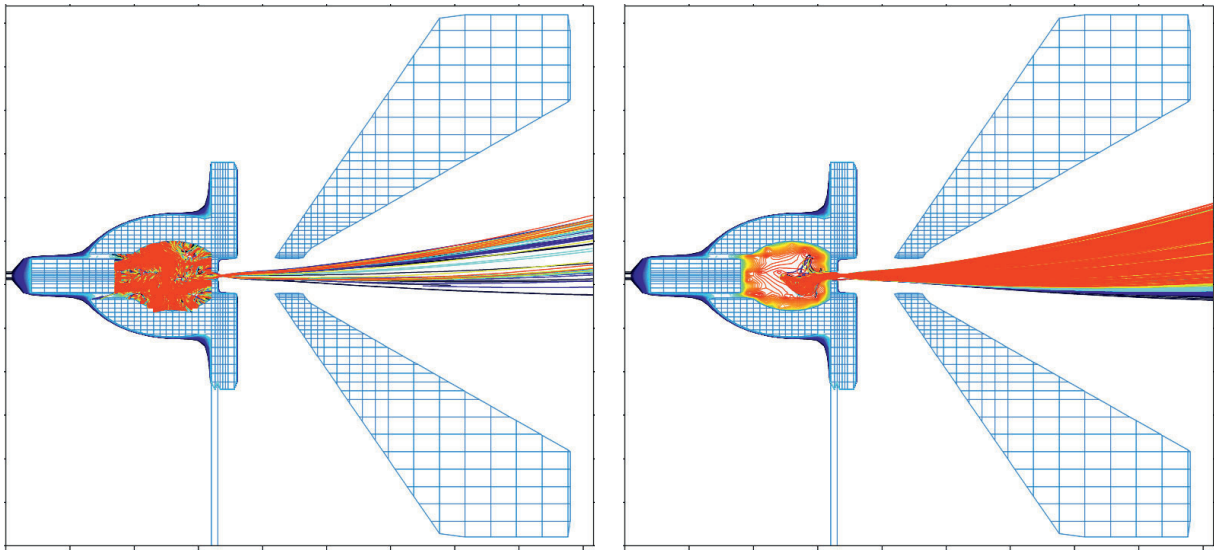


Fig. 20: Radial extraction from a Penning ion source (no sputtering), higher current density (≈ total 1 mA). Left: all trajectories plotted; right: only ion trajectories that are extracted are shown.

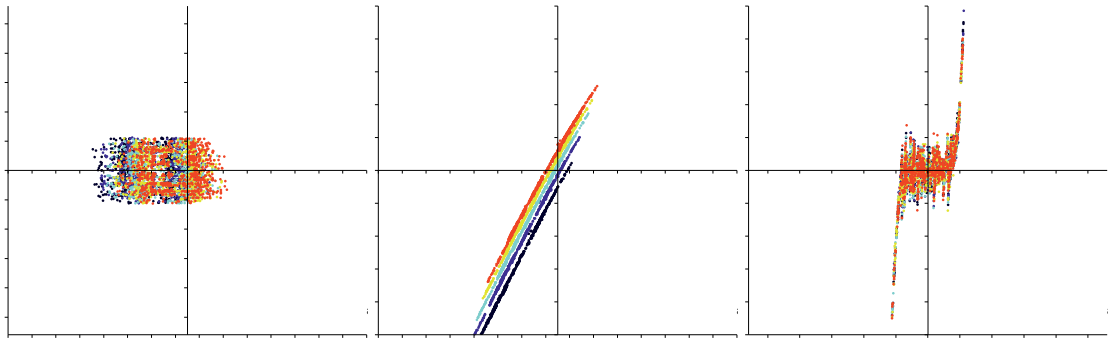


Fig. 21: Radial extraction from a Penning ion source (no sputtering). Left: real space; middle: horizontal emittance (perpendicular to the slit direction); right: vertical emittance (slit direction). Higher current mode (≈ total 1 mA).

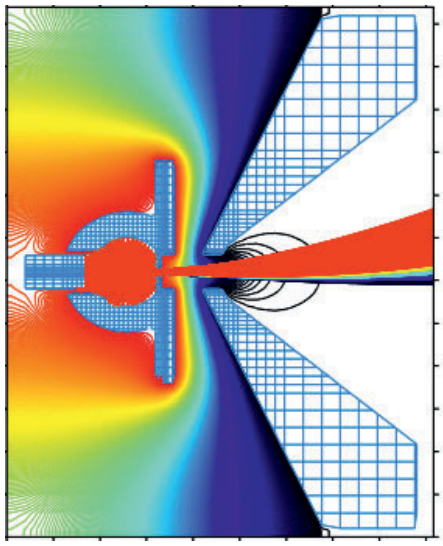


Fig. 22: Radial extraction from a Penning ion source in sputter mode

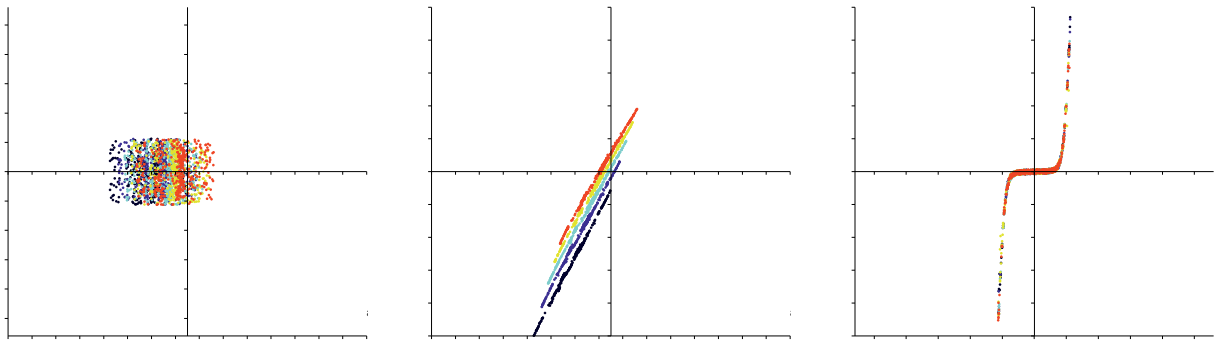


Fig. 23: Radial extraction from a Penning ion source in sputter mode. Low-current mode (≈ total 50 μA)

To increase the efficiency of the discharge, the complete ion source is embedded in a strong magnetic field (several kilogauss). The electron density can be increased by this method. Simultaneously the extracted ions are affected by this magnetic field, shown in Fig. 24. The path length of the ions within the magnetic field is in the range of several 10 cm. Space-charge-compensating electrons would not be able to move freely in that region. Nevertheless, experimental experience shows that the extracted ion beam has to be space-charge-compensated, otherwise the obtained electrical currents would not be possible. Obviously, electrons are present within the magnetic field, where they can move (oscillate) along the magnetic field lines.

4.3.2 ECR ion sources

Electron cyclotron resonance ion sources (ECRISs) have a strong axial magnetic mirror field. A radial multipole component (in most cases a hexapole) is added to increase the radial confinement and to stabilize the plasma confinement, especially for higher particle densities. Electrons are magnetized under the influence of the strong magnetic flux densities of the order of a few teslas. Because of the low temperature of the ions, they are magnetized as well. The property 'magnetized' means that the Larmor radius described in Eq. (11) is small compared to other dimensions:

$$r_{\mathrm{L}} = \frac{m v_{\perp}}{q B}. \tag{11}$$

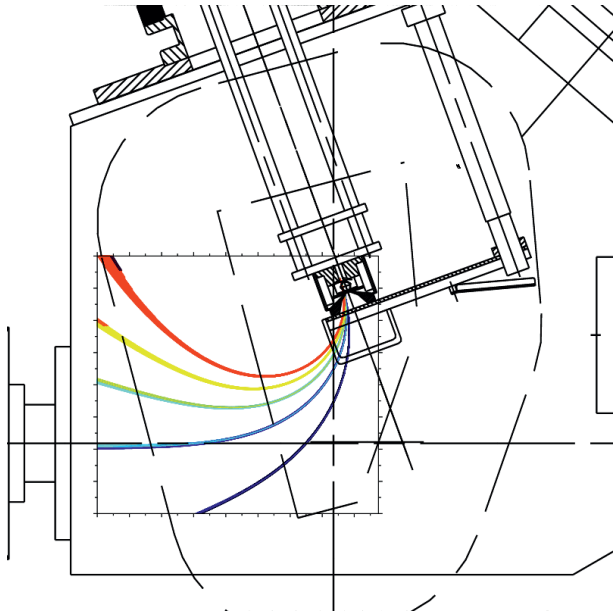


Fig. 24: Top view of a PIG source with an extraction system inside a 110° source dipole magnet, and the vacuum beam line. The rectangular insert shows the region of computation together with an extracted argon beam with charge states 1 to 5. Charge state 2 has been selected for the accelerator. A fully space-charge-compensated beam is assumed. As starting coordinates for the ions, the output coordinates from the extraction simulation have been used.

Table 2: Larmor radii r_L for different particles with different m/q at 1 T magnetic flux density and typical initial starting energy.

	Mass	Charge	Velocity (m s^{-1})	$r_{\mathrm{L}}(1\ \mathrm{T})$ (m)
Electron	e	−1	1 327 075	7.50×10^{-6}
Ion	p	1	30 971	0.32×10^{-3}
Ion	Ar	10	9 793	0.41×10^{-3}
Ion	Xe	10	8 525	1.17×10^{-3}
Ion	U	1	2 000	4.97×10^{-3}
Ion	U	28	10 620	0.94×10^{-3}

Some examples, with the assumption of a certain voltage drop (here 5 V) and a given magnetic flux density (here 1 T) are given in Table 2.

Because of the magnetization, the trajectories of ions inside the plasma are not affected, or at least not strongly affected, by ion–ion collisions. Ions as well as electrons can move freely along a magnetic field line, and, depending on the gradient of the magnetic flux density along the magnetic field line, energy in the direction of the magnetic field line is transferred into the plane perpendicular to the magnetic field line and vice versa. Ions can be extracted from the ion source, even if they are generated at any place within the plasma chamber far away from the extraction aperture. Especially in the configuration of an ECRIS, this is equivalent to the fact that ions are generated at different levels of magnetic flux density [10]. This means for this specific ion source that not only $\int \boldsymbol{B}\,\mathrm{d}s \neq 0$, but furthermore $\int \boldsymbol{B}\,\mathrm{d}s$ is different for ions coming from different starting points. These different initial conditions have to be taken into account when modelling ECRISs.

Of course, the magnetic field distribution is not a free parameter to optimize the conditions for a good extraction. It has to be chosen according to physical boundary conditions and to provide a good plasma confinement.

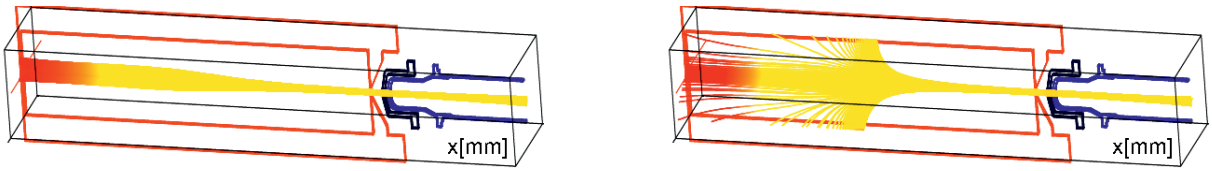


Fig. 25: Magnetic field lines going through the extraction aperture on a circle of 5 mm for two different magnetic settings.

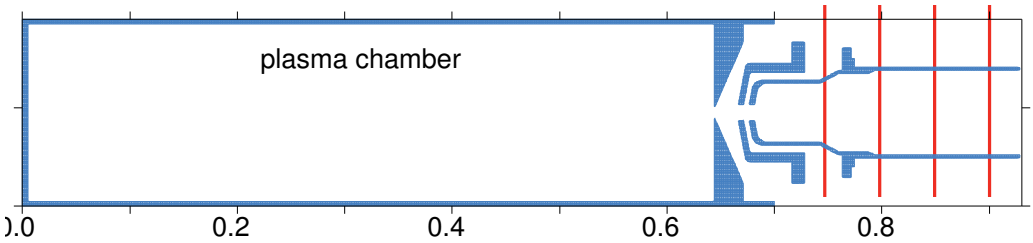


Fig. 26: Horizontal projection of the extraction system of the MS-ECRIS including plasma chamber. The red lines indicate the four locations of the given profiles, respectively emittances.

The magnetic field lines can now be used to localize the starting locations of the ions (see Fig. 25). Depending on the specific magnetic setting, ions from different locations are extracted. The colour along the field line indicates the relative magnetic flux density compared to the flux density at the extraction aperture. In this example, the magnetic flux density is so high that the ion beam will be focused after extraction already in the fringing field of the solenoid. The following profiles and emittance figures are taken at different locations of the beam line: 0.75 m 0.80 m, 0.85 m and 0.90 m (see Fig. 26). The different colours for different particles in Figs. 27–32 indicate different initial magnetic flux density at the origin of particles.

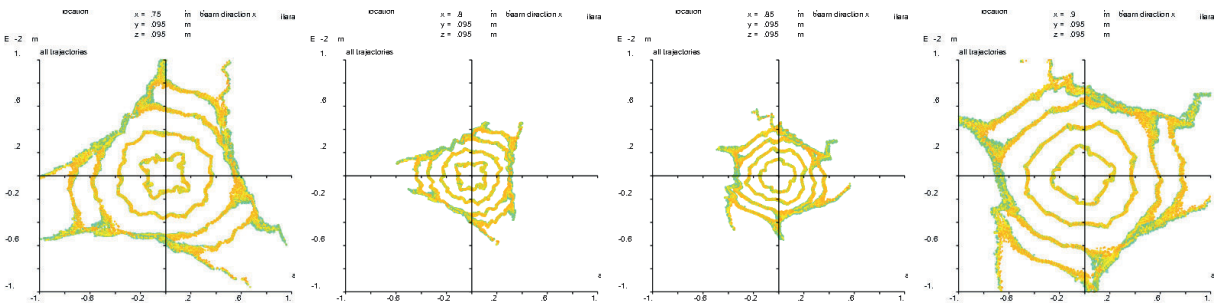


Fig. 27: Real space (profile) at 0.75 m, 0.80 m, 0.85 m and 0.90 m. Scale is ± 1 cm

These kinds of structures, as can be seen in the profiles of Figs. 27–32, have been observed experimentally. This confirms that the assumptions are realistic, but the model is of course not complete. The distribution needs to be smoothed; here only magnetic field lines going through the extraction hole on 1 mm, 2 mm, 3 mm and 4 mm radius have been used. A charge-state distribution (such as, for example, that shown in Figs. 33 and 34) has to be used instead of only one charge state.

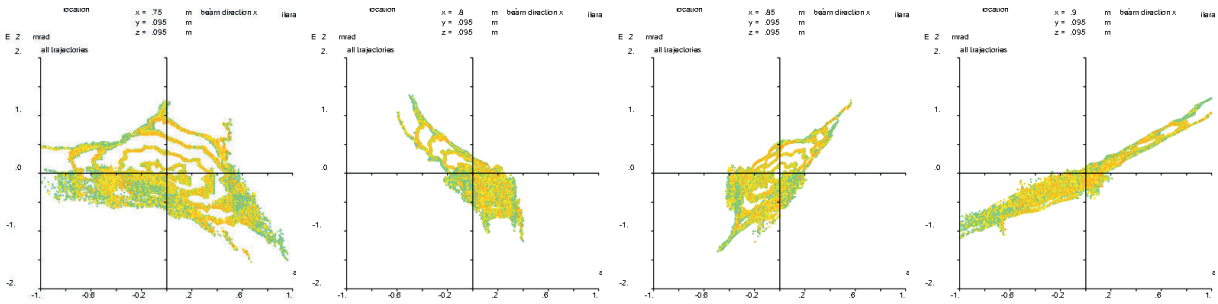


Fig. 28: Horizontal emittance at the same positions. Scale is ±1 cm, respectively ±200 mrad

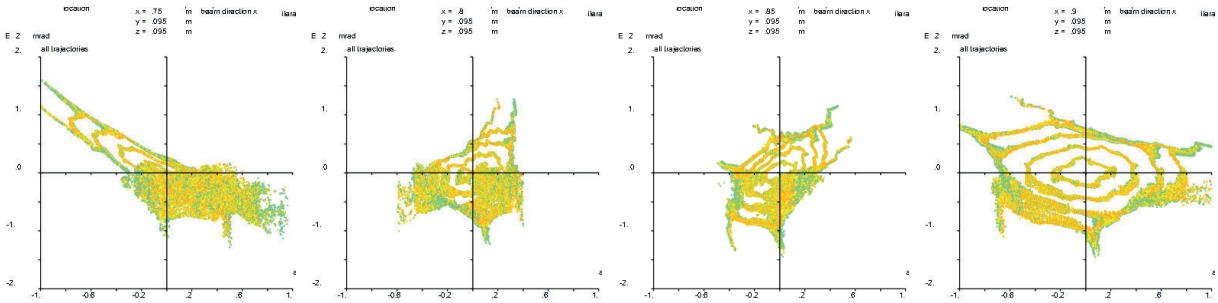


Fig. 29: Vertical angles as a function of horizontal location (mixed phase space). Scale is ±1 cm, respectively ±200 mrad.

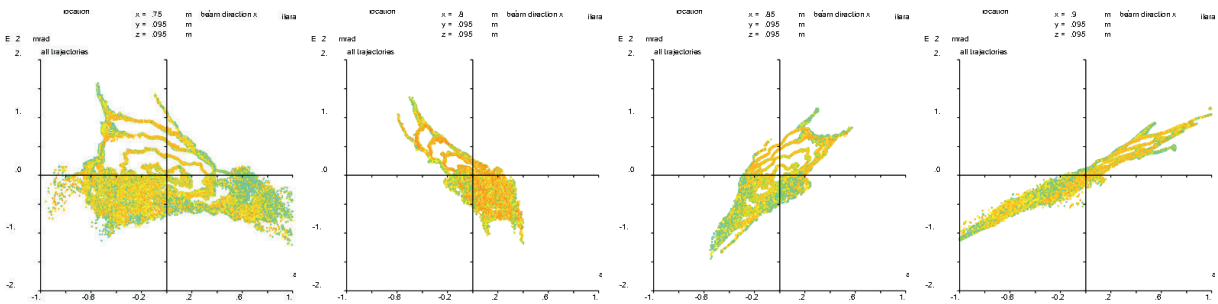


Fig. 30: Vertical emittance. Scale is ±1 cm, respectively ±200 mrad

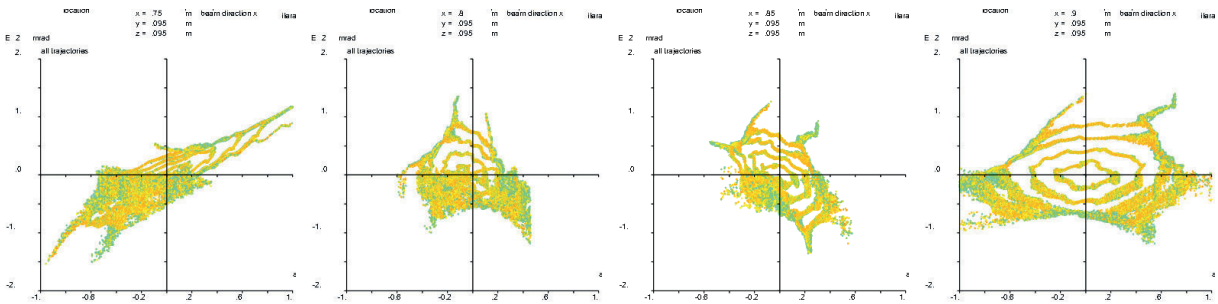


Fig. 31: Horizontal angles as a function of vertical location (mixed phase space). Scale is ±1 cm, respectively ±200 mrad.

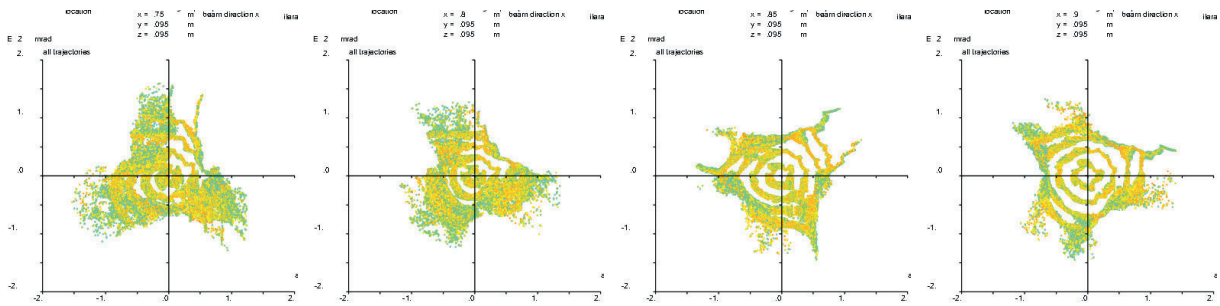


Fig. 32: Momentum space; ±200 mrad in both directions

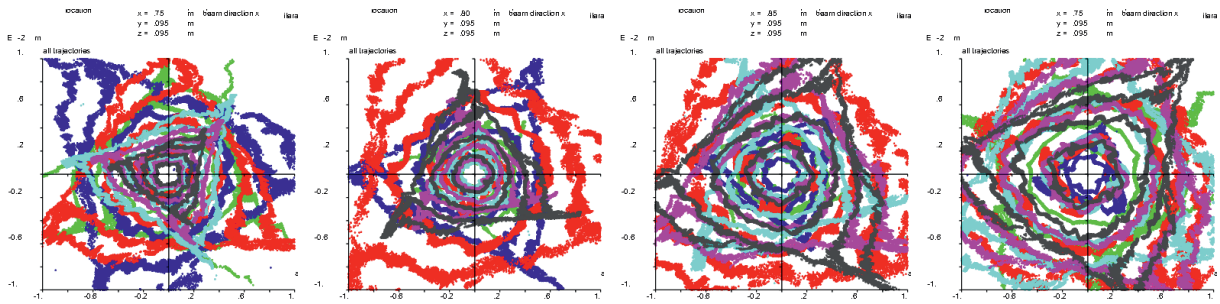


Fig. 33: Behaviour of different charge states in real space. Green, Ar^{+}; blue, Ar^{2+}; red, Ar^{3+}; light blue, Ar^{4+}; cyan, Ar^{5+}; grey, Ar^{6+}. The different focal lengths of different charge states are clearly visible.

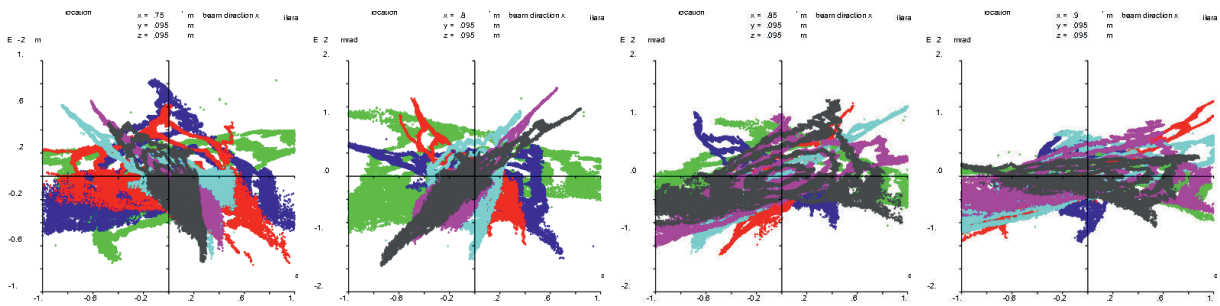


Fig. 34: Behaviour of different charge states for the horizontal emittance. Colours as defined in Fig. 33

Furthermore, the particle density at different places inside the plasma chamber (at the beginning of the trajectory) has to be known. Experiments with variable frequency have shown a strong influence on the extracted beam intensity [11], even with small frequency changes. This could be explained by the assumption that the existing modes of the RF determine the local available plasma density.

For this simulation 450 nodes in longitudinal direction, and 201 nodes in each transverse direction, have been used, in total 28 350 450 nodes. For one charge state, about 100 000 trajectories are necessary. Computing time, including space charge, for a full spectrum is of the order of several hours.

4.3.3 Negative ion sources

For negative ion beam extraction the neutralization of electrons and negative ions inside the plasma is performed by positive ions:

$$n_{\mathrm{i}+} = n_{\mathrm{i},0} \exp\left[-\frac{q(\Phi - \Phi_{\mathrm{pl}})}{kT_{\mathrm{i}}}\right], \tag{12}$$

where $n_{\mathrm{i}+}$ is the density of positive ions as a function of the potential Φ. The number of positive ions within the undisturbed plasma is $n_{\mathrm{i},0}$ at plasma potential Φ_{pl} with an ion temperature of T_{i}. This equation is similar to the case of the positive ion sources (Eq. (10)), but positive ions play the role of electrons within the plasma. For extraction of H^- the source has to be biased negatively, but this will extract electrons as well. Because the electron density can be high compared to negative ions, the load of the undesired electrons for the power supply is high. Typically, a magnetic filter is used at a position close to the extraction electrode to prevent the extraction of electrons. The magnetic field can be produced by permanent magnets, or by a coil inside the plasma chamber. The coil is indicated by the black conductor shown in Figs. 35 and 37. However, the magnetic field influences not only the electrons, but also the extracted negative ions. The extracted negative ions have an angle that might have to be compensated.

In the first example, the location of the magnetic filter is too close to the extraction system. A lot of electrons will be extracted (see Figs. 35 and 36).

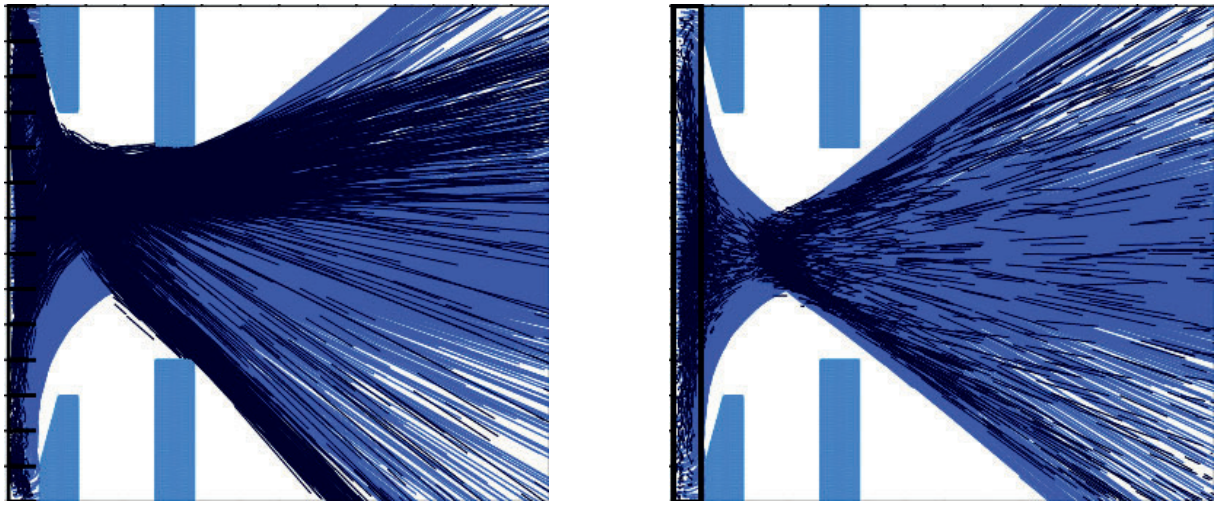


Fig. 35: Horizontal and vertical projection of extraction geometry together with the ion and electron trajectories. The magnetic flux density in this case is not strong enough to retain the electrons from being extracted.

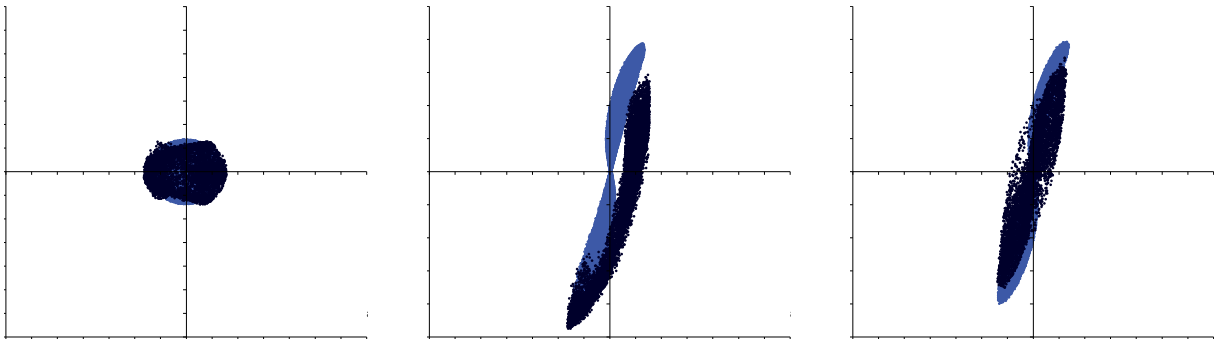


Fig. 36: Real space, horizontal and vertical emittances. H^- in blue, electrons in black

To improve the action of the filter it has been moved 4 mm back in the plasma direction. In Figs. 37 and 38 it is shown that this modification was successful; all electrons remain within the plasma.

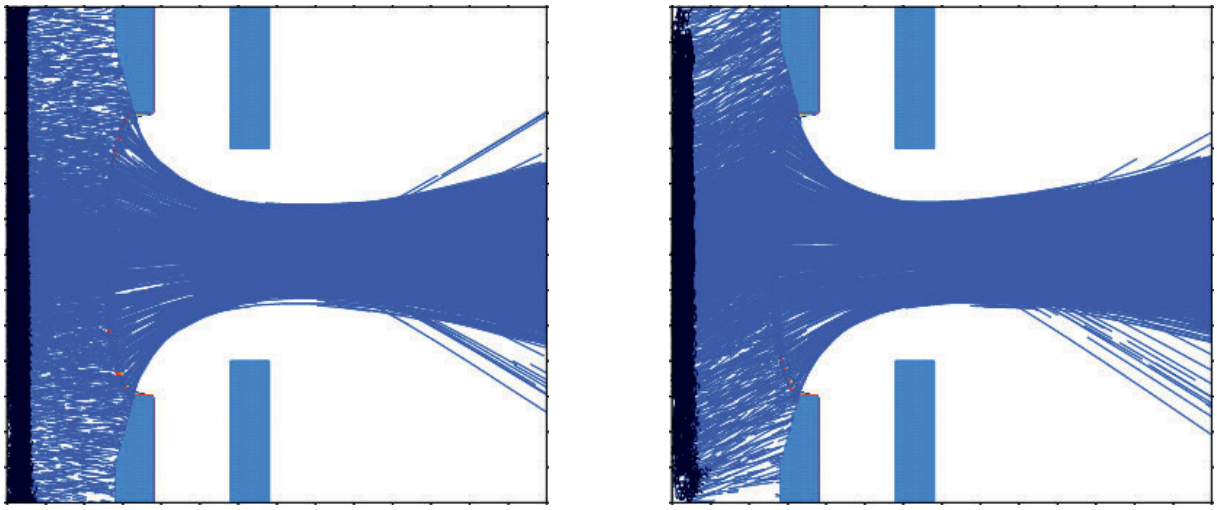


Fig. 37: Plasma boundary and extracted H^- beam in blue. Electrons are trapped within the plasma

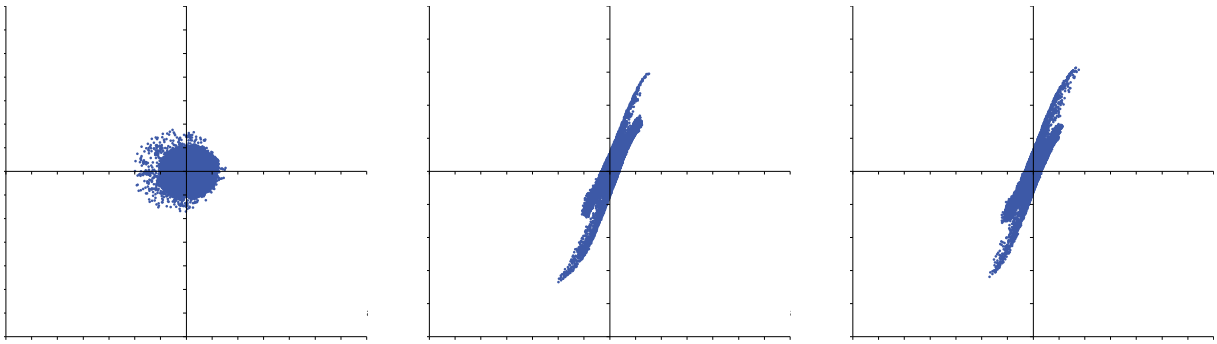


Fig. 38: Real space, horizontal and vertical emittances. H^- in blue, electrons are trapped within the plasma

It should be noted that all other parameters beside the relative position of the magnetic filter are unchanged. In both cases, a H^- current density of 2000 A m^{-2} and an electron current density of 8000 A m^{-2} have been assumed.

For this simulation 401 nodes in the longitudinal direction and 201 nodes in each transverse direction have been used, giving in total 16 200 801 nodes. For the simulation of H^- 50 000 trajectories and 10 000 trajectories for electrons have been used. Computing time including space charge is of the order of several hours.

References

[1] M. V. Nezlin, *Plasma Phys.* **10** (1968) 337.

[2] R. Dölling, J. Pozimski, P. Groß, J. Klabunde and P. Spädtke, Investigation of space-charge compensation with residual-gas-ion energy analyser, Proc. 18th Int. Linear Accelerator Conf., 1996.

[3] N. Chauvin, O. Delferrière, R. Duperrier, R. Gobin, P. A. P. Nghiem and D. Uriot, Transport of intense ion beams and space charge compensation issues in low energy beam lines, Proc. 14th Int. Conf. on Ion Sources, *Rev. Sci. Instrum.* **83**(2) (2012) 02B320.

[4] INP, KOBRA3-INP user manual, INP, Junkernstr. 99, 65205 Wiesbaden, Germany.

[5] I. G. Brown (Ed.), *The Physics and Technology of Ion Sources* (John Wiley, New York, 1989).

[6] S. A. Self, *Phys. Fluids* **6** (1963) 1762.

[7] V. Dubenkov, B. Sharkov, A. Golubev, A. Shumshurov, O. Shamaev, I. Rudskoy, A. Streltsov, Y.

Satov, K. Makarov, Y. Smakovsky, D. Hoffmann, W. Laux, R. W. Müller, P. Spädtke, C. Stöckl, B. Wolf and J. Jacoby, Acceleration of Ta^{10+} ions produced by laser ion source in RFQ MAXILAC, *Laser Part. Beams* **14** (1996) 385–392.

[8] K. Harres, I. Alber, A. Tauschwitz, V. Bagnoud, H. Daido, M. Günther, F. Nürnberg, A. Otten, M. Schollmeier, J. Schüttrumpf, M. Tampo and M. Roth, Beam collimation and transport of quasineutral laser-accelerated protons by a solenoid field, *Phys. Plasmas* **17** (2010) 023107.

[9] P. Spädtke and C. Mühle, Simulation of ion extraction and beam transport, *Rev. Sci. Instrum.* **71**(2) (2000) 820.

[10] P. Spädtke, R. Lang, J. Mäder, F. Maimone, J. Roßbach and K. Tinschert, Investigations on the structure of the extracted ion beam from an electron cyclotron resonance ion source, Proc. 14th Int. Conf. on Ion Sources, *Rev. Sci. Instrum.* **83**(2) (2012) 02B720.

[11] F. Maimone, K. Tinschert, L. Celona, R. Lang, J. Mäder, J. Roßbach and P. Spädtke, Operation of the CAPRICE electron cyclotron resonance ion source applying frequency tuning and double frequency heating, *Rev. Sci. Instrum.* **83**(2) (2012) 02A304.

Beam Extraction and Transport

T. Kalvas
Department of Physics, University of Jyväskylä, Finland

Abstract
This chapter gives an introduction to low-energy beam transport systems, and discusses the typically used magnetostatic elements (solenoid, dipoles and quadrupoles) and electrostatic elements (einzel lens, dipoles and quadrupoles). The ion beam emittance, beam space-charge effects and the physics of ion source extraction are introduced. Typical computer codes for analysing and designing ion optical systems are mentioned, and the trajectory tracking method most often used for extraction simulations is described in more detail.

1 Introduction

In principle, the task of beam extraction and the following low-energy beam transport (LEBT) system are quite simple. The ion source extraction consists of the front plate of the ion source, which is known as the plasma electrode, and at least one other electrode, the puller (or extractor) electrode, which provides the electric field for accelerating the charged particles from the ion source to form an ion beam. Whether or not the extraction contains any other electrodes, the beam leaves the extraction at energy

$$E = q(V_{\text{source}} - V_{\text{beamline}}), \tag{1}$$

defined by the charge q of the particles and the potential difference between the ion source, V_{source}, and the following beamline, V_{beamline}, which is typically the laboratory ground, as shown in Fig. 1. The ion source voltage is therefore set according to the requirements of the subsequent application. The intensity of the particle beam depends, as a first approximation, on the flux of charged particles hitting the plasma electrode aperture. The extracted ion beam current can therefore be estimated as

$$I = \tfrac{1}{4} A q n \bar{v}, \tag{2}$$

where A is the plasma electrode aperture, q is the charge of the particles, n is the ion density in the plasma and $\bar{v}$ is the mean velocity of extracted particles in the ion source plasma. Assuming a Maxwell–Boltzmann distribution for the extracted plasma particles, the mean velocity $\bar{v} = \sqrt{8kT/\pi m}$. From the point of view of the extraction, the plasma electrode aperture can be adjusted to tune the beam intensity.

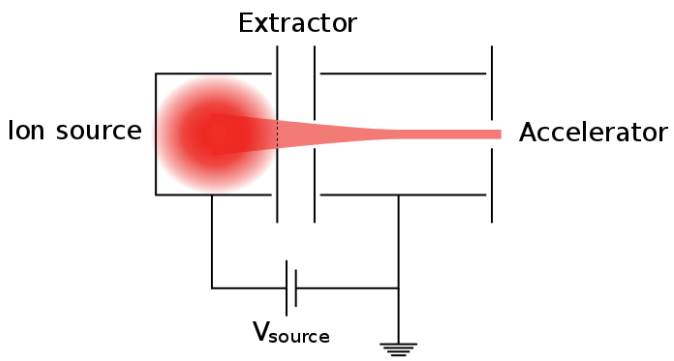


Fig. 1: The most basic electrostatic extraction system possible

The practical solutions are unfortunately much more complicated in most cases. The applications following the LEBT, which typically are accelerators to bring the beam to higher energies, often pose

strict requirements for the ion beam parameters. Not only do the beam intensity, energy and species spectrum need to meet the requirements, but also the beam spatial and temporal structure are specified. For an example of what kind of parameters the specifications might contain, please see Table 1.

Table 1: Specifications of the H^- ion beam at the end of the LEBT, entering the 2.5 MeV Radio Frequency Quadrupole (RFQ) accelerator of the Spallation Neutron Source (SNS) [1].

Parameter	**Value**
Beam current (H^-)	50 mA
Beam energy	65 keV
Emittance (normalized r.m.s.)	0.20 mm mrad
Twiss α	1.5
Twiss β	0.06 mm mrad^{-1}
Macro-pulse length	1 ms
Macro-pulse duty factor	6%
Mini-pulse length	645 ns
Mini-pulse duty factor	68%

The spatial requirements for the beam mean that focusing is necessary in the LEBT. Similarly, the temporal requirements necessitate beam chopping. Without careful design of the focusing elements, the space-charge force of the beam blows up the beam to the walls of the vacuum chamber, and only a part of the generated beam gets transported to the following accelerator. The extraction focusing systems must also provide some adjustability because, in most cases, the plasma conditions might not be constant in day-to-day operations. The LEBT has to be able to adapt to days of lower and higher performance, while maximizing the throughput to the following accelerator. Designing such systems is not easy. From the ion optics point of view, a system that is as short as possible would be preferred, but at the same time the system design also has to take into account the practical engineering constraints. For example, the beamline needs to have space for diagnostics and vacuum pumps in addition to the focusing elements.

This chapter concentrates mainly on the topic of ion optics in low-beam-energy systems: the focusing elements, beam space-charge blow-up, beam–plasma interface and how to model these systems with computer codes.

2 Low-energy beam transport

The ion beam travels in the beam transport line from one ion optical element to another along a curved path, which is usually defined as the longitudinal direction z. The transverse directions x and y are defined relative to the centre of the transport line, the optical axis, where $x = 0$ and $y = 0$. The transport line is usually designed in such a way that a so-called reference particle travels along the optical axis with nominal design parameters. The ion beam (bunch) is an ensemble of charged particles around the reference particle, with each individual particle at any given time described by spatial coordinates (x, y, z) and momentum coordinates (p_x, p_y, p_z). This six-dimensional space is known as the particle *phase space*. In addition to these coordinates, often inclination angles α and β or the corresponding tangents x' and y' are used. These are defined by

$$x' = \tan\alpha = \frac{p_x}{p_z} \quad \text{and} \quad y' = \tan\beta = \frac{p_y}{p_z}. \tag{3}$$

The motion of a charged particle in electromagnetic fields E and B is described by the Lorentz force F and Newton's second law, giving

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t} = \boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}), \tag{4}$$

where p is the momentum and v is the velocity of the particle with charge q. In general, the trajectory of a charged particle can be calculated by integrating the equation of motion if the fields are known. In the case of beam transport, the fields have two origins: (i) external fields, which are mainly generated by the ion optical elements; and (2) beam-generated fields. Generally, in LEBT ion optics, we assume that the particle density in a beam is low enough that single particle–particle interactions are negligible. It is therefore sufficient to take into account the collective beam-generated fields.

2.1 Beamline elements

The ion optical elements of the beam transport line come in two varieties: magnetic and electric. In the case of high-energy beams, where $v \approx c$, magnetic elements are used because the force, which is created with an easily produced magnetic field of 1 T, equals the force from an electric field of 300 MV m^{-1}, which is impossible to produce in a practical device. In LEBT systems, where the beam velocity is low, and the practical limit for electric fields is about 5 MV m^{-1}, the achievable forces are comparable, and other factors, such as size, cost, power consumption and the effects of beam space-charge compensation, come into play. An important factor in the selection of the type of beamline elements is also the fact that electrostatic fields do not separate ion species. In electrostatic systems the particles follow trajectories defined only by the system voltages. The magnetic elements, on the other hand, have a dependence on mass-to-charge ratio m/q. This allows separation of different particle species from each other.

The common beamline elements that are used to build LEBT systems include immersion lens, einzel lens, solenoid, dipole and quadrupole lenses. A short introduction to each of these elements is given below. For more detailed analyses, the reader is referred to literature (Refs. [2] and [3], for example).

2.1.1 Immersion lens

The immersion lens (or gap lens) is simply a system of two electrodes with a potential difference of $\Delta V = V_2 - V_1$. The lens can be either accelerating or decelerating, and, in addition to changing the particle energy by $q\Delta V$, the element also has a focusing action. The focal length of the immersion lens is given by [2]

$$\frac{f}{L} = \frac{4(\sqrt{V_2/V_1} + 1)}{V_1/V_2 + V_2/V_1 - 2}, \tag{5}$$

where L is the distance between the electrodes. The electrostatic extraction systems always have gap lenses, which accelerate the beam to the required energy. The first acceleration gap (plasma electrode to puller electrode) is a special case of the immersion lens because of the effect of the plasma on the electric field. It will be addressed later in this chapter.

2.1.2 Einzel lens

The einzel lens is made by combining two gap lenses into one three-electrode system with first and last electrodes at the beamline potential V_0 and the centre electrode at a different potential V_{einzel}. The einzel lens, which is typically cylindrically symmetric for round beams, is the main tool for beam focusing in many electrostatic extraction systems. The einzel focusing power is dependent on the geometry and the voltage ratio $R = (V_{\text{einzel}} - V_0)/V_0$, assuming that zero potential is where the beam kinetic energy is zero. The einzel lens may have the first gap accelerating and the second gap decelerating (known as accelerating einzel lens, $R > 0$) or vice versa (known as decelerating einzel lens, $R < 0$). Both configurations are focusing, but the refractive power of the einzel in decelerating mode is much higher than in accelerating mode with the same lens voltage. For example, for the geometry shown in Fig. 2, the focal length $f \approx 10D$ for a decelerating mode, $V_{\text{einzel}} - V_0 = -0.5V_0$. To achieve the same focal length in accelerating mode, a voltage $V_{\text{einzel}} - V_0 = 1.1V_0$ is needed. On the other hand, accelerating einzel lenses should be preferred if the required higher voltage (and electric fields) can be handled, because they

have lower spherical aberrations than decelerating einzel lenses, especially when the required refractive power is high.

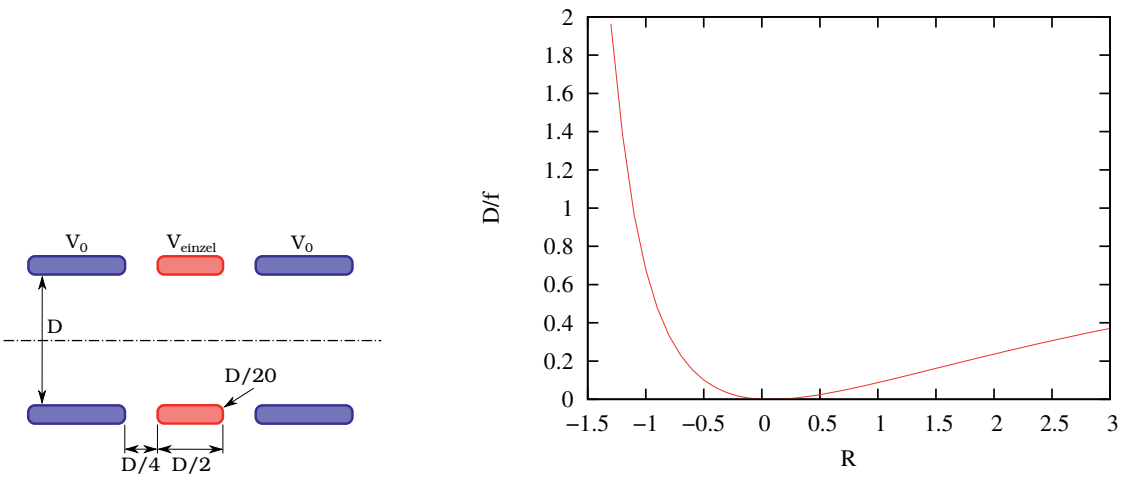


Fig. 2: An example geometry for an einzel lens and its refractive power scaled with the einzel internal diameter D as a function of the voltage ratio R. The lens is much stronger in decelerating mode compared to accelerating mode.

A special case of the einzel lens, where the first electrode and the third electrode are at different potentials, is also possible. This kind of set-up is known as a three-aperture immersion lens or zoom lens. It provides adjustable focusing in a system that otherwise acts as an immersions lens [2].

2.1.3 Solenoid lens

A solenoid lens is the magnetic equivalent of the electrostatic einzel lens. It consists of rotationally symmetric coils wound around the beam tube, creating a longitudinal magnetic field peaking at the centre of the solenoid. The focusing action of the solenoid is somewhat difficult to derive, but the idea can be described as follows, assuming a thin lens [4]. The radial magnetic field at the entrance of the solenoid gives the particle entering the field with $v_r = 0$ at radius r_0 an azimuthal thrust

$$v_\theta = \frac{qBr_0}{2m}, \tag{6}$$

which makes the trajectories helical inside the solenoid. At the exit of the solenoid, the particle receives a thrust cancelling the azimuthal velocity, but leaving the particle with a radial velocity

$$v_r = -\frac{r_0 q^2}{4m^2 v_z} \int B^2 \, \mathrm{d}z. \tag{7}$$

This radial velocity causes the beam to converge towards the optical axis. The refractive power of the lens is given by

$$\frac{1}{f} = \frac{q^2}{8mE} \int B^2 \, \mathrm{d}z. \tag{8}$$

2.1.4 Electrostatic and magnetic dipoles

The electrostatic dipole and magnetic dipole are elements that are primarily used to deflect charged-particle beams. The magnetic dipole is constructed from coil windings, creating a constant magnetic field in the transverse direction. The particles in the magnetic field follow circular trajectories as usual,

with radius

$$\rho = \frac{p}{qB} \approx \frac{mv_z}{qB} = \frac{1}{B}\sqrt{\frac{2mV_0}{q}}, \tag{9}$$

where V_0 is the voltage used to accelerate the particles from zero to v_z. Similarly, an electrostatic dipole may be constructed from cylindrical electrodes of radii r_1 and r_2 with voltages V_1 and V_2. The radius of curvature of the particle between the plates becomes

$$\rho = \frac{2V_0}{E}, \tag{10}$$

where E is the electric field and V_0 is the potential at the orbit (again assuming that zero potential is where the beam kinetic energy is zero). The voltage and electric field between the plates are

$$V = V_1 + (V_2 - V_1)\frac{\log(r/r_1)}{\log(r_2/r_1)}, \tag{11}$$

$$E = -\frac{V_2 - V_1}{\log(r_2/r_1)}\,\frac{1}{r}. \tag{12}$$

By choosing the plate voltages symmetrically as $V_1 = V_0 + V_{\text{plate}}$ and $V_2 = V_0 - V_{\text{plate}}$, the required plate voltage can be found as

$$V_{\text{plate}} = V_0 \log(r_2/r_1). \tag{13}$$

The optical axis of such a system is at radius $\rho = \sqrt{r_1 r_2}$. This is not the only possibility for the cylindrical dipole. The optical axis can also be chosen to be in the middle of the plates, which leads to asymmetric voltages.

The dipole elements also have focusing/defocusing properties. For example, the magnetic dipole with edges perpendicular to the optical axis (edge angle 0°) focuses the beam in the bending plane (x) as shown in Fig. 3(a). Directly from geometry, a so-called Barber's rule can be derived: the centre of curvature of the optical axis and the two focal points are on a straight line. For a symmetric set-up this means that $A = B = R/\tan(\phi/2)$. There is no focusing action in the y direction.

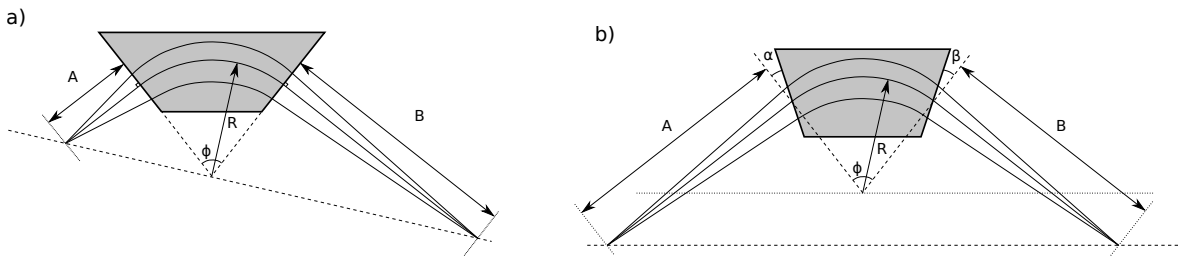


Fig. 3: Focusing of a magnetic dipole in the bending plane. (a) The case where the magnet has 0° edge angles can be described by Barber's rule: the center of curvature of optical axis and the two focal points are on a straight line. (b) If the edge angles are positive, as shown, the focusing power is decreased.

If the magnet edge angles deviate from 90°, the focusing power in the x direction can be adjusted. If the edge angle is made positive (as shown in Fig. 3(b)), there is weaker focusing in the x direction. If the angle is negative, there is stronger focusing in the x direction. Changing the edge angle also has an important effect in the y direction: if the angles are positive, the fringing field of the magnet will focus the beam in the y direction, as shown in Fig. 4. Overall, this means that the focusing in the x direction can be traded for y focusing. The focal length from the edge focusing is given by

$$f_y = \frac{R}{\tan\alpha}. \tag{14}$$

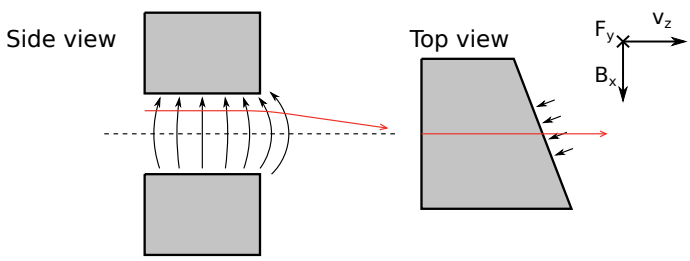


Fig. 4: In a dipole magnet with positive edge angle, the fringing field has a B_x component at non-zero y coordinates, focusing the beam in the y direction.

In a symmetric double-focusing dipole (where there is the same focal length in x and y directions), the angles and distances are given by

$$2\tan\alpha = 2\tan\beta = \tan(\phi/2), \tag{15}$$

$$A = B = \frac{2R}{\tan(\phi/2)}. \tag{16}$$

For a $\phi = 90°$ bending magnet, the edge angles become $\alpha = \beta = 26.6°$ and the focal distances $A = B = 2R$.

A dipole magnet focusing in the y direction can also be made by making the pole faces conical, as shown in Fig. 5. This kind of magnet is known as a radially inhomogeneous sector magnet. The magnetic field of such a magnet can be approximated as

$$\begin{aligned} B_y(x,y) &= B_0\left(1 - n\frac{x}{R} + \cdots\right), \\ B_x(x,y) &= B_0\left(n\frac{y}{R} + \cdots\right), \end{aligned}$$

where B_0 is the magnetic field on the optical axis with radius R and n is the so-called field index, which depends on the angle of the magnet pole. The first-order approximation shows that the magnet is focusing in the x direction if $n < 1$ and focusing in the y direction if $n > 0$. The focusing forces inside the magnet are symmetric at $n = \frac{1}{2}$.

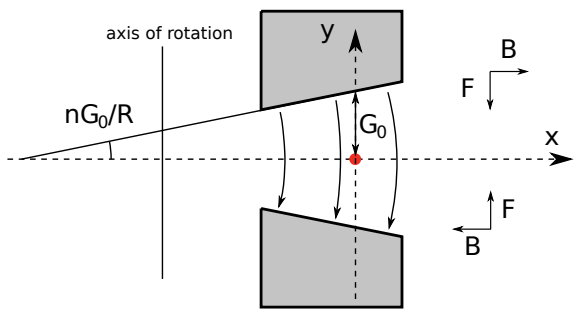


Fig. 5: In the inhomogeneous sector magnets, the field strength decreases with increasing radius x due to increasing gap, which also leads to B_x increasing with y.

The regular cylindrical electrostatic dipole only has focusing in the x direction similar to the regular straight edge magnetic dipole. The y focusing can be introduced by adjusting the ends of the cylindrical plates for edge focusing or by using spherical or toroidal plates.

For small-angle deflection, typically electrostatic parallel plates or so-called XY magnets are used. The parallel plates with $\pm V_{\text{plate}}$ voltages bend the beam by an angle

$$\theta = \frac{V_{\text{plate}} L}{V_0 d}, \tag{17}$$

where L is the length of the plates in the z direction and d is the distance between the plates. This kind of system is typically used for small corrections in beamlines and for beam chopping. For example, in Fig. 6, a simulation of a fast chopping system in an LBNL-built neutron generator is shown. The XY magnets are the magnetic equivalent of parallel plates, with typically two pairs of windings in a single instrument for correction in both transverse directions. The beam deflection is given by

$$\theta = LB\sqrt{\frac{q}{2mV_0}}, \tag{18}$$

where L is the field length and B is the field strength inside the device.

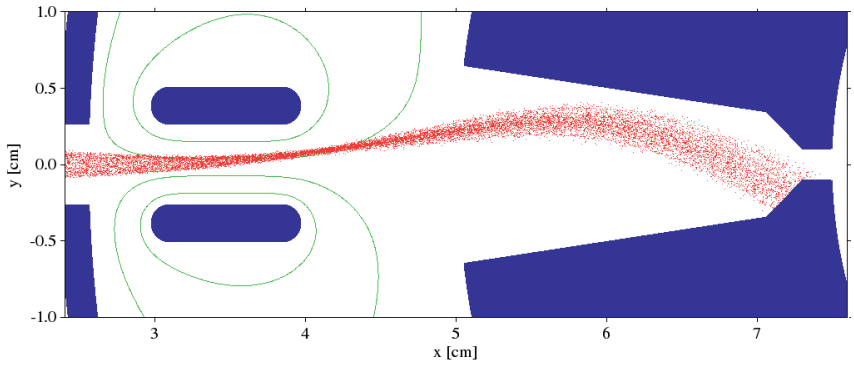


Fig. 6: Fast beam chopping can be done with parallel plates. A particle-in-cell (PIC) simulation of an LBNL-built neutron generator using 15 ns risetime ± 1500 V switches for generating 5 ns beam pulses is shown.

2.1.5 Quadrupole lenses

Electrostatic and magnetic quadrupoles are often used as focusing elements in LEBT systems in addition to einzel lenses and solenoids. The electrostatic quadrupole consists of four hyperbolic electrodes placed symmetrically around the beam axis with positive potential V_{quad} on the electrodes in the $+x$ and $-x$ directions and negative potential $-V_{\text{quad}}$ on the electrodes in the $+y$ and $-y$ directions, as shown in Fig. 7. The potential in such a configuration is given by

$$V = \frac{x^2 - y^2}{a^2} V_{\text{quad}}, \tag{19}$$

where a is the radius of the quadrupole. This leads to an electrostatic field

$$\boldsymbol{E} = -\frac{2V_{\text{quad}}}{a^2} x\hat{\boldsymbol{x}} + \frac{2V_{\text{quad}}}{a^2} y\hat{\boldsymbol{y}}, \tag{20}$$

from which we can see that such a quadrupole focuses a positive ion beam in the x direction and defocuses in the y direction. By analysing the particle trajectories in such fields, it can be shown that the refractive powers of the system are

$$1/f_x = k\sin(kL), \tag{21}$$
$$1/f_y = -k\sinh(kL), \tag{22}$$

where $k^2 = V_{\text{quad}}/aV_0$ and L is the effective length of the quadrupole.

The magnetic quadrupole has a similar construction to the electrostatic one: the magnet poles are made to be hyperbolic and windings are coiled in such a way that every other pole has magnetic flux in to the beam and every other out of the beam. The magnetic field in such a system is

$$\boldsymbol{B} = \frac{B_{\text{T}}}{a} y\hat{\boldsymbol{x}} + \frac{B_{\text{T}}}{a} x\hat{\boldsymbol{y}}, \tag{23}$$

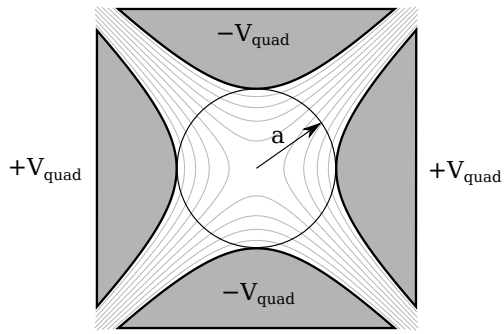


Fig. 7: Electrostatic quadrupole, with a being the distance from the optical axis to the electrode tip

where B_T is the magnetic field density at the pole tip. Positively charged particles having velocity $\boldsymbol{v} = v_z\hat{\boldsymbol{z}}$ feel a force $\boldsymbol{F} = qB_T v_z(-x\hat{\boldsymbol{x}} + y\hat{\boldsymbol{y}})/a$, which is focusing in the x direction and defocusing in the y direction. The magnetic force leads to the same refractive powers as presented by Eqs. (21) and (22), but with

$$k_B^2 = \frac{q}{p}\frac{B_\mathrm{T}}{a}.$$

Quadrupole lenses are typically used as doublets or triplets for solutions that are focusing in both transverse directions. Quadrupoles can also be used for transforming asymmetric beams such as slit beams from a Penning ion source, for example, into a round beam.

2.2 Beam emittance

Traditionally the emittance is defined as the six-dimensional volume limited by a contour of constant particle density in the (x, p_x, y, p_y, z, p_z) phase space. This volume obeys the Liouville theorem and is constant in conservative fields. With practical accelerators, a more important beam quality measure is the volume of the *envelope* surrounding the beam bunch. This is not conserved generally – only in the case where the forces acting on the particles are linear (see Fig. 8). Typically in the case of continuous (or long pulse) beams, where the longitudinal direction of the beam is not of interest, transverse distributions (x, x') and (y, y') are used instead of the full phase-space distribution for simplicity. Also for these distributions the envelope surrounding the distribution changes when nonlinear forces (non-idealities of beamline elements, for example) act on the particles. The size and shape of the transverse distribution envelope are important quality measures for beams because most complex ion optical devices such as accelerators have an acceptance window in the phase space within which they can operate.

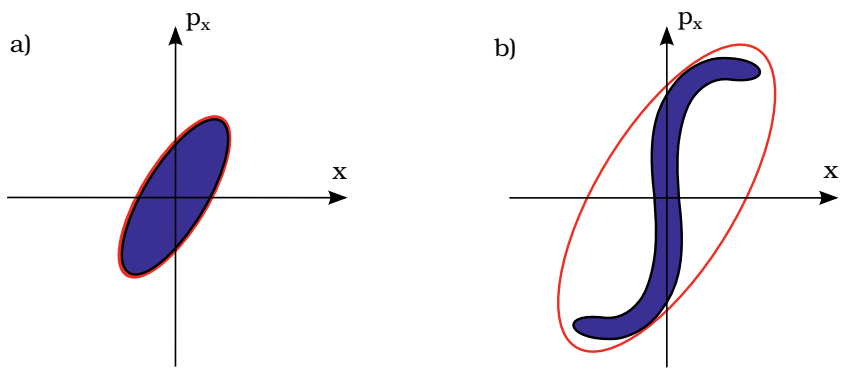


Fig. 8: A two-dimensional projection of an ensemble of particles (a) before going through a nonlinear optical system and (b) after it. The area of the particle distribution (shown in blue) is conserved but the area of the elliptical envelope (shown in red) increases.

2.2.1 Emittance ellipse

For calculation and modelling purposes, a simple shape is needed to model the ion beam envelope in (x, x') phase space. Real well-behaved ion beams usually have Gaussian distributions in both x and x' directions. Because the contours of two-dimensional (2D) Gaussian distributions are ellipses, it is an obvious solution to use the ellipse as the model in 2D phase spaces (and ellipsoids in higher dimensions). The equation for an origin-centred ellipse is

$$\gamma x^2 + 2\alpha x x' + \beta x'^2 = \epsilon, \tag{24}$$

where the scaling

$$\beta\gamma - \alpha^2 = 1 \tag{25}$$

is chosen. Here ϵ is the *two-dimensional transverse emittance*, and α, β and γ are known as the *Twiss parameters* defining the ellipse orientation and aspect ratio. The area of the ellipse is

$$A = \pi\epsilon = \pi R_1 R_2, \tag{26}$$

where R_1 and R_2 are the major and minor radii of the ellipse. Because of the connection between the area of the ellipse and ϵ, there is sometimes confusion about whether to include π in the above formula for quoted emittance values. The unit of emittance is often written as π mm mrad. This is done to emphasize that the quoted emittance number is the product of the radii and *not* the area of the ellipse. Always, when communicating about emittance numbers, it should be clearly indicated what the number is to avoid confusion.

From Eq. (24) the dimensions of the ellipse can be calculated. Some of the most important dimensions needed in calculations are shown in Fig. 9.

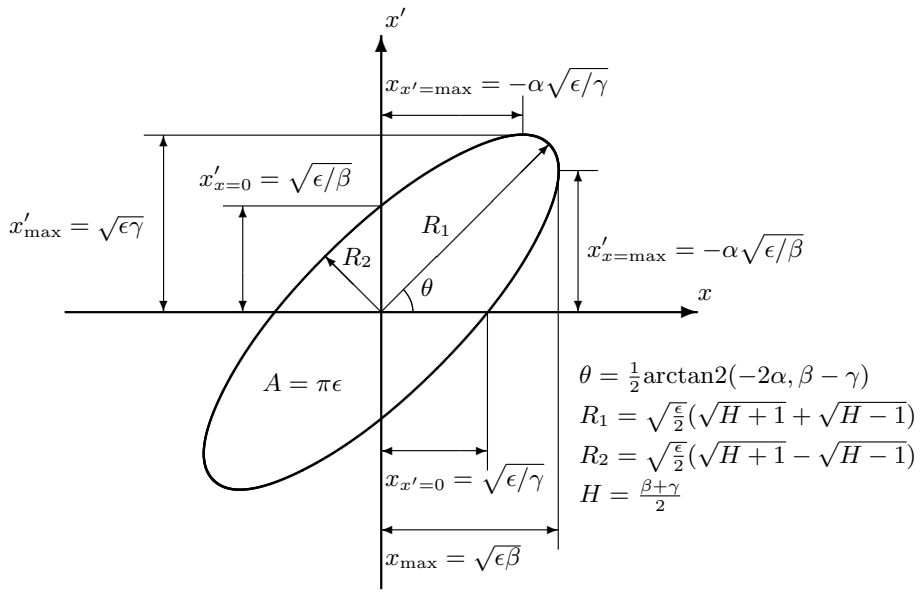


Fig. 9: Emittance ellipse geometry with the most important dimensions

2.2.2 Calculating r.m.s. emittance

How do ϵ and the Twiss parameters relate to phase-space distributions? How is the envelope defined? There are numerous ways to fit an ellipse to particle data. Often, a minimum-area ellipse containing some fraction of the beam is wanted (e.g. $\epsilon_{90\%}$), but unfortunately this is difficult to produce in a robust

way. A well-defined way to produce the ellipse is by using a statistical definition known as the *r.m.s. emittance*,

$$\epsilon_{\text{rms}} = \sqrt{\langle x'^2 \rangle \langle x^2 \rangle - \langle x x' \rangle^2}, \tag{27}$$

with the expectation values defined as

$$\left\langle x^2 \right\rangle = \frac{\iint x^2 I(x,x')\,\mathrm{d}x\,\mathrm{d}x'}{\iint I(x,x')\,\mathrm{d}x\,\mathrm{d}x'}, \tag{28}$$

$$\left\langle x'^2 \right\rangle = \frac{\iint x'^2 I(x,x')\,\mathrm{d}x\,\mathrm{d}x'}{\iint I(x,x')\,\mathrm{d}x\,\mathrm{d}x'}, \tag{29}$$

$$\left\langle x x' \right\rangle = \frac{\iint x x' I(x,x')\,\mathrm{d}x\,\mathrm{d}x'}{\iint I(x,x')\,\mathrm{d}x\,\mathrm{d}x'}, \tag{30}$$

where $I(x,x')\,\mathrm{d}x\,\mathrm{d}x'$ is the magnitude of the beam current at the differential area $\mathrm{d}x\,\mathrm{d}x'$ of phase space at (x,x'). Similarly, the Twiss parameters can be calculated from the particle distribution with

$$\alpha = -\frac{\left\langle x x' \right\rangle}{\epsilon}, \quad \beta = \frac{\left\langle x^2 \right\rangle}{\epsilon} \quad \text{and} \quad \gamma = \frac{\left\langle x'^2 \right\rangle}{\epsilon}. \tag{31}$$

For these formulas, it is assumed that the emittance distribution is centred at the origin, so that $\langle x \rangle = 0$ and $\langle x' \rangle = 0$. With measured emittances, additional difficulties arise from background noise and amplifier offsets in $I(x,x')$ data. Filtering methods for processing experimental data exist, from simple thresholding to more refined algorithms such as SCUBEEx [5].

2.2.3 Amount of beam inside the emittance ellipse

The meaning of the r.m.s. emittance can be more easily understood by looking at the amount of beam that is enclosed by the ellipse. This, of course, depends on the particle distribution shape. For real measured distributions, there is no direct rule. For theoretical known distributions, this can be calculated. The two most used model distributions used for beams are the bi-Gaussian and the Kapchinskij–Vladimirskij (KV) distribution.

The bi-Gaussian distribution oriented along the axes is given by

$$I(x,x') = \frac{1}{2\pi\sigma_x\sigma_{x'}} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{x'^2}{\sigma_{x'}^2}\right)\right], \tag{32}$$

where σ_x and $\sigma_{x'}$ are the standard deviations of the distribution in the x and x' directions. In practice, the distribution can be additionally rotated by angle θ.

The KV distribution has a uniform beam density inside an elliptical phase space given by

$$I(x,x') = \begin{cases} \dfrac{1}{\pi\epsilon} & \text{if } \gamma x^2 + 2\alpha x x' + \beta x'^2 \le \epsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{33}$$

In Fig. 10 the beam fraction inside the ellipse is shown as a function of the area of the ellipse for hard-edged KV and bi-Gaussian distributions. Because four times the r.m.s. ellipse fully encloses the KV distribution, the so-called 4-r.m.s. emittance is often used as the quoted number instead of the 1-r.m.s. emittance.

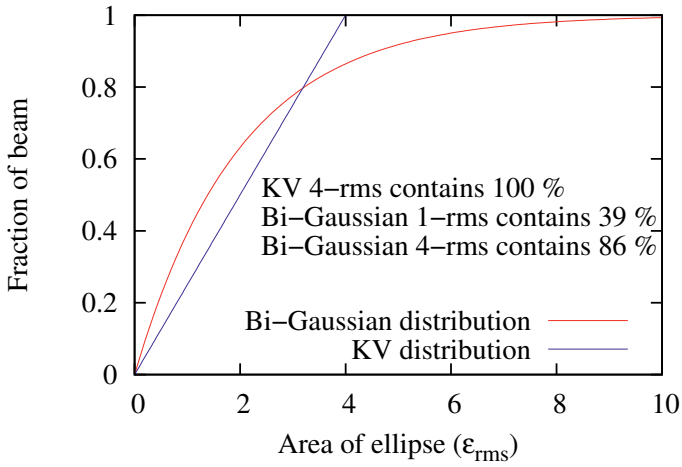


Fig. 10: Fraction of beam inside ellipse with $\epsilon = q\epsilon_{\text{rms}}$ for hard-edged KV distribution (blue) and for Gaussian distribution (red).

2.2.4 Normalization of emittance

The transverse emittance defined in (x, x') space has the property that it is also dependent on the longitudinal beam velocity. If the beam is accelerated and p_z increases, then $x' = p_x/p_z$ decreases. This effect is eliminated by normalizing the velocity to the speed of light c, which gives

$$x'_{\text{n}} = \frac{p_x}{p_{z1}}\frac{v_{z1}}{c} = \frac{v_x}{c} = \frac{p_x}{p_{z2}}\frac{v_{z2}}{c} \tag{34}$$

at non-relativistic velocities. The normalized emittance can therefore be calculated from the unnormalized emittance using

$$\epsilon_{\text{n}} = \epsilon\frac{v_z}{c}. \tag{35}$$

2.2.5 Emittance from plasma temperature

An ion beam formed by letting charged particles from a plasma be emitted from a round aperture has an emittance defined by the plasma ion temperature T and the aperture radius r, assuming that the acceleration to velocity v_z does not add aberrations. This minimum emittance can be calculated by using Eqs. (27)–(30) and using a particle distribution defined by a circular extraction hole and Gaussian transverse ion distribution, i.e.

$$I(x, x') = \frac{2}{\pi r^2}\sqrt{r^2 - x^2}\ \sqrt{\frac{m}{2\pi kT}}\exp\left(\frac{-m(x'v_z)^2}{2kT}\right). \tag{36}$$

After normalization, the resulting r.m.s. emittance becomes

$$\epsilon_{\text{rms,n}} = \frac{1}{2}\sqrt{\frac{kT}{m}}\frac{r}{c}. \tag{37}$$

The calculation can be performed similarly for a slit-beam extraction, which gives

$$\epsilon_{\text{rms,n}} = \frac{1}{2}\sqrt{\frac{kT}{3m}}\frac{w}{c}. \tag{38}$$

In the round aperture case, the emittance of the beam is linearly proportional to the plasma aperture radius. On the other hand, the beam current is roughly proportional to the area of the plasma aperture.

Scaling of the aperture size does not therefore change the *beam brightness*,

$$B = \frac{I}{\epsilon_{\mathrm{n},x}\,\epsilon_{\mathrm{n},y}} \tag{39}$$

in a first approximation.

2.2.6 *Emittance from solenoidal B field*

In electron cyclotron resonance (ECR) and microwave ion sources, there is a strong solenoidal magnetic field at the plasma electrode location, where the beam formation happens. This has a strong influence on the beam quality. As the particles exit the solenoidal magnetic field, they receive an azimuthal thrust described by Eq. (6). The emittance of the beam can be calculated outside the solenoid by considering the particle coordinates far away, where the azimuthal particle motion has completely changed to radial motion,

$$r' = \frac{v_r}{v_z} = \frac{v_\theta}{v_z} = \frac{qBr_0}{2mv_z}. \tag{40}$$

The r.m.s. emittance of the beam can be calculated from the radius of the constant-current-density beam at extraction and the asymptotic radial angle,

$$\epsilon_{\mathrm{rms}} = \frac{1}{4} r_0 r' = \frac{qBr_0^2}{8mv_z}. \tag{41}$$

The normalized emittance is given by

$$\epsilon_{\mathrm{rms,n}} = \frac{qBr_0^2}{8mc}. \tag{42}$$

For ECR ion sources, the effect of the magnetic field dominates the emittance compared to the effect of the ion temperature, as a result of the high magnetic fields in these devices. Unfortunately, the formula given here is not able to predict the emittance values, as measurements indicate that, for a given element, the higher-charge-state ions have lower emittances than lower-charge-state ions (see Fig. 11). This trend contradicts the prediction of Eq. (42). The only possible interpretation is that the ions are not being extracted from a uniform plasma. The higher charge states are confined closer to the axis at the extraction aperture, and therefore their emittance in the beam is lower. This example shows the additional difficulty in analysing ECR extractions, as there are no simple self-consistent plasma models describing the starting conditions for the ions [6].

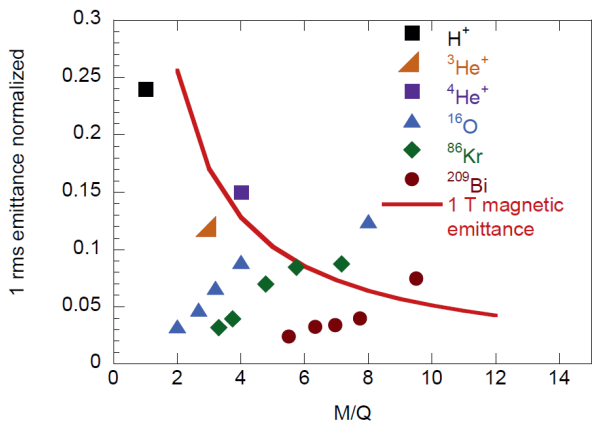


Fig. 11: Emittance measurements on the AECR-U ion source for various species in comparison with emittances calculated from Eq. (42) for a magnetic field of 1 T. Reproduced from Ref. [6].

2.3 Space charge

The ion beam charge density

$$\rho = \frac{J}{v} = \frac{I}{Av} \tag{43}$$

plays a major role in beam extraction systems, where current densities are high and velocities are low compared to other parts of accelerator systems. The space charge induces forces, which increase the divergence and emittance, 'blowing up' the beam. At higher-energy parts of the accelerator, the magnetic force generated by the beam particles starts to compensate the blow-up, but it is insignificant at $v \ll c$.

2.3.1 Space-charge effects on beam

Assuming a cylindrical constant-current-density beam with radius r propagating with constant velocity v_z, the beam-generated electric field is given by Gauss's law,

$$E = \begin{cases} \dfrac{I}{2\pi\epsilon_0 v_z}\dfrac{r}{r_{\text{beam}}^2} & \text{if } r \leq r_{\text{beam}}, \\ \dfrac{I}{2\pi\epsilon_0 v_z}\dfrac{1}{r} & \text{otherwise.} \end{cases} \tag{44}$$

The potential inside a beam tube with radius r_{tube} is therefore

$$\phi = \begin{cases} \dfrac{I}{2\pi\epsilon_0 v}\left[\dfrac{r^2}{2r_{\text{beam}}^2} + \log\left(\dfrac{r_{\text{beam}}}{r_{\text{tube}}}\right) - \dfrac{1}{2}\right] & \text{if } r \leq r_{\text{beam}}, \\ \dfrac{I}{2\pi\epsilon_0 v}\log\left(\dfrac{r}{r_{\text{tube}}}\right) & \text{otherwise.} \end{cases} \tag{45}$$

The potential distribution is plotted in Fig. 12 for a 10 mA, 10 keV proton beam inside a 100 mm beam tube.

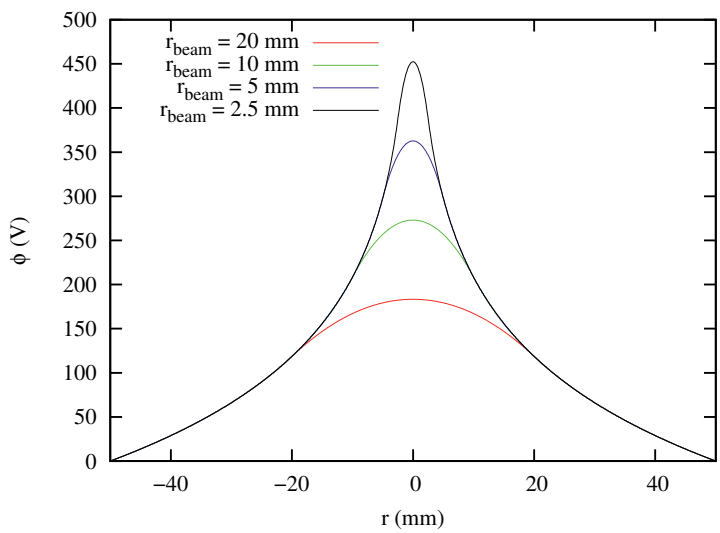


Fig. 12: The potential distribution inside a cylindrical 100 mm tube with 10 mA, 10 keV proton beam

The electric field in the constant-current-density case, given by Eq. (44), is linear with radius and therefore does not cause emittance growth, but it does cause increasing divergence of the beam. A particle at the beam boundary experiences a repulsive force

$$F_r = qE_r = ma_r = \frac{qI}{2\pi\epsilon_0 r v_z}. \tag{46}$$

Therefore, the particle acceleration is

$$a_r = \frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{\mathrm{d}^2 r}{\mathrm{d}z^2}\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = v_z^2 \frac{\mathrm{d}^2 r}{\mathrm{d}z^2}. \tag{47}$$

The particle trajectory is given by the differential equation

$$\frac{\mathrm{d}^2 r}{\mathrm{d}z^2} = \frac{1}{v_z^2} a_r = K\frac{1}{r}, \tag{48}$$

where

$$K = \frac{qI}{2\pi\epsilon_0 m v_z^3}, \tag{49}$$

assuming that the beam divergence is small (i.e. Eq. (46) is valid). The differential equation can be integrated after a change of variable $\lambda = \mathrm{d}r/\mathrm{d}z$, which gives

$$\frac{\mathrm{d}r}{\mathrm{d}z} = \sqrt{2K\log(r/r_0)}, \tag{50}$$

assuming $\mathrm{d}r/\mathrm{d}z = 0$ at $z = 0$. The solution is separable and can be integrated again to achieve the final solution [7]

$$z = \frac{r_0}{\sqrt{2K}} F\left(\frac{r}{r_0}\right) \tag{51}$$

with

$$F\left(\frac{r}{r_0}\right) = \int_{y=1}^{r/r_0} \frac{\mathrm{d}y}{\sqrt{\log y}}. \tag{52}$$

The last integral is not analytic, but can be numerically integrated for estimates of divergence. As an example, a parallel zero-emittance beam of $^{181}\mathrm{Ta}^{20+}$ accelerated with 60 kV has initial radius of $r_0 =$ 15 mm. The size of a 120 mA beam after a drift of 100 mm can be solved from $F(r/r_0) = 1.189$, which gives $r = 20$ mm.

With practical drifting low-energy beams, a more realistic model for the beam distribution is bi-Gaussian, for example. This kind of distribution leads to nonlinear space-charge forces, which cause emittance growth in addition to increase of beam divergence. Computer simulations are required to estimate these effects.

2.3.2 Space-charge compensation

The potential well of the beam formed by the accelerated charged particles acts as a trap for oppositely charged particles in areas where there are no external electric fields to drain the created charges. The trapped particles compensate the charge density of the beam, decreasing the depth of the potential well and therefore also decreasing the magnitude of the beam space-charge effects described above. This process is called *space-charge compensation*. The most abundant process for the production of compensating particles is the ionization of the background gas within the beam. In the case of positive ion beams, the electrons produced in the background gas ionization are trapped in the beam, while slow positive ions are repelled to the beamline walls. In the case of negative ion beams, the compensating particles are the positive ions created in the gas. The creation rate of the compensating particles can be estimated with

$$\frac{\mathrm{d}n_\mathrm{c}}{\mathrm{d}t} = \Phi n_\mathrm{gas}\sigma_\mathrm{i}, \tag{53}$$

where Φ is the flux of beam particles, n_gas is the gas density and σ_i is the ionization cross-section. If the creation rate is high enough, the space-charge compensation is finally limited by the leakage of compensating particles from the potential well as the compensation factor approaches 100%. The

compensation factor achieved in a real system is difficult to estimate accurately because it depends on the lifetime of the compensating particles in the potential well. The most important processes affecting the lifetime are (i) leakage of particles at the beamline ends, which can be limited with accelerating einzel lenses or magnetic fields, for example, (ii) recombinative processes and (iii) scattering processes leading to ejection of particles from the potential well. Assuming that the creation rate of compensating particles is high, the time-scale for achieving full compensation is

$$\tau = \frac{\rho_{\text{beam}}}{e\,\mathrm{d}n_{\text{c}}/\mathrm{d}t} = \frac{Q}{v n_{\text{gas}} \sigma_{\text{i}}}, \tag{54}$$

where Q is the charge state of the beam and v is the velocity of the beam. This equation can be used, for example, to estimate if compensation is possible in pulsed beams.

In high-beam-intensity LEBT systems, a controlled amount of background gas is often added to the vacuum chamber to increase the amount of compensation. A 1–2% beam loss in the increased ionization processes can lead to 10% increase in total beam transmission due to decreased divergence. The magnitude of the compensation can be estimated, for example, by measurement of the energy distribution of the ions ejected from the beam as a function of background gas pressure, as shown in Fig. 13. There are also particle-in-cell (PIC) computer codes, such as WARP [8] and SOLMAXP for modelling the relevant processes affecting beam compensation. These programs can be used for analysing beam transport with self-consistent compensation. With other beam transport programs, a typical solution is to scale the beam current locally with the space-charge compensation factor estimated by the user. For more information on space-charge compensation, please see the chapter by N. Chauvin in these proceedings dedicated to this topic [9].

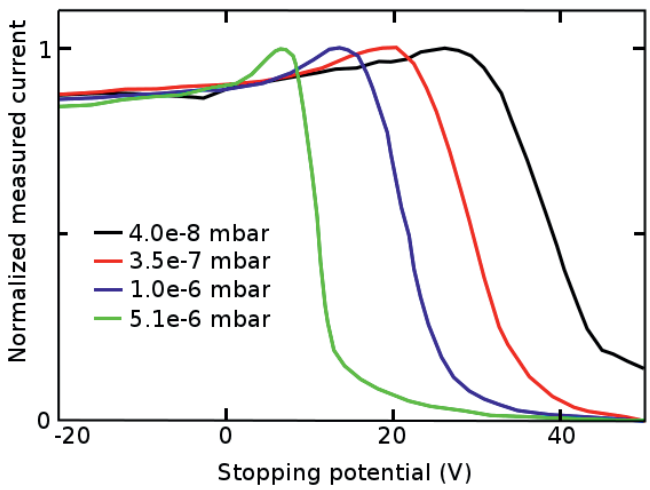


Fig. 13: Measurement of ion energy distribution of the ejected ions created in the positive ion beam potential well with different gas pressures. The potential well depth changes from ∼40 V to ∼10 V when the pressure is increased from 4.0×10^{-8} mbar to 5.1×10^{-6} mbar. Reproduced from Ref. [10].

3 Beam formation

In the introduction of this chapter it was assumed that the ion beam is simply formed by accelerating the plasma particles hitting the plasma electrode aperture, which separates the quasi-neutral plasma and the unneutralized beam. In this section, the physics of beam formation is analysed in more detail.

3.1 Space-charge-limited emission

In the first acceleration gap, where the beam is formed, the space-charge forces acting on the beam are largest. The situation can be evaluated in one dimension by assuming a beam starting with zero velocity with Poisson equation

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} = -\frac{\rho}{\epsilon_0} = -\frac{J}{\epsilon_0}\sqrt{\frac{m}{2q\phi}}, \tag{55}$$

where z is the location, ϕ is the gap potential, J is the beam current density and ϵ_0 is the vacuum permittivity. The emission surface is at $\phi(z=0)=0$ and the extractor surface is at $\phi(z=d)=V$. For $J=0$, the potential distribution between the surfaces is linear. As the emission current density increases, the electric field at the emission surface decreases until it becomes zero, as shown in Fig. 14(a). At that point the emission current is at the maximum level, for which Eq. (55) can be solved with the boundary condition $\frac{\mathrm{d}\phi}{\mathrm{d}z}(z=0)=0$. This condition is known as space-charge-limited emission, and the resulting limit for the maximum emission current density can be calculated using the following equation, which is known as the Child–Langmuir law [11]:

$$J_{\mathrm{max}} = \frac{4}{9}\epsilon_0\sqrt{\frac{2q}{m}}\,\frac{V^{3/2}}{d^2}. \tag{56}$$

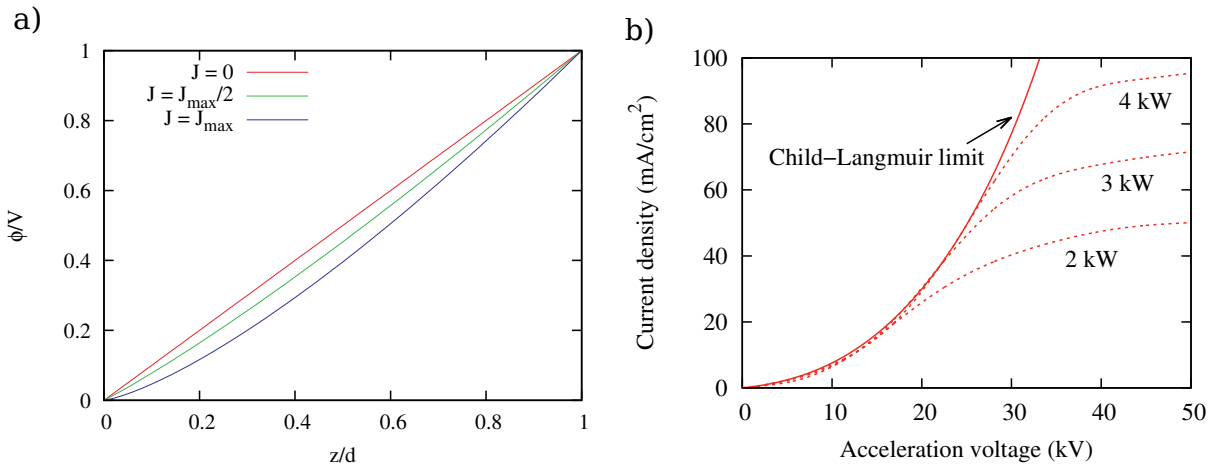


Fig. 14: (a) Potential distribution between emission and extractor surfaces with different beam current densities in the system. (b) Typical current–voltage characteristic of plasma extraction. With low acceleration voltages, the emission is operating in the Child–Langmuir limit. At higher voltages the extracted current density saturates due to the emission limit of the plasma (which is varied by adjusting the driver power from 2 kW to 4 kW).

Thermionic electron guns are often operated in the space-charge-limited mode for stability and uniformity reasons – even if the local emission conditions change, the emitted current stays constant as long as the emission is limited by the Child–Langmuir law.

The plasma ion sources are typically operated in emission-limited mode, i.e. the potential difference between the plasma electrode and puller electrode is made sufficiently large to handle the beam space charge. The law in the form shown here is not strictly valid for ion source plasma extraction because of the effects of plasma neutralization and higher 'starting' velocity of particles in plasma extraction. The physics of the space-charge limit is still valid and the Child–Langmuir law (56) can be used to estimate it.

In any system, the maximum extractable current is dependent on the geometry, the emission current density and the voltage via the space-charge limit. In the space-charge-limited region, the current is

proportional to $V^{3/2}$. This leads to the definition of the beam *perveance* as

$$P = \frac{I}{V^{3/2}}, \tag{57}$$

which is the proportionality constant describing the system. As long as the emission is space-charge-limited, the beam perveance is roughly constant. When the voltage is further increased and the beam emission is no longer space-charge-limited, the beam perveance decreases. See Fig. 14(b) for an example of the current–voltage characteristic of a plasma extraction.

3.2 Electrode geometry

The space-charge forces try to blow up the beam, as was shown above. This happens especially in the first acceleration gap because of the low velocity of the beam. To counteract the space-charge forces in the transverse direction, the electrodes can be shaped in such a way that the electric field in the first gap is not only accelerating but also focusing. In the case of space-charge-limited surface-emitted electrons, there is a perfect solution providing a parallel electron beam accelerated from the cathode [12]. The solution is to have a field shaping electrode around the cathode (at cathode potential) in a $67.5°$ angle with respect to the emitting surface normal, as shown in Fig. 15. This geometry is known as Pierce geometry. For plasma ion sources, there is no such magic geometry because the ions do not start from a fixed surface, but from plasma with varying starting conditions.

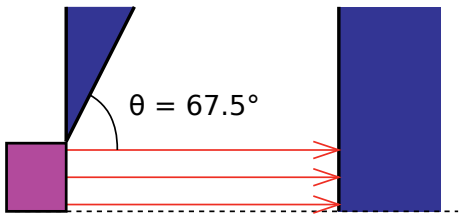


Fig. 15: Perfectly parallel extraction of space-charge limited surface emission electrons using the Pierce geometry

3.3 Positive ion plasma extraction

In the case of ion plasma extraction, the beam formation is more complicated than in the case of surface-emitted electrons described above. The ions are born in the quasi-neutral plasma and get extracted into the unneutralized beam. It is obvious that the extraction cannot be modelled without considering the neutralizing effect of the plasma. The simplest description of the necessary transition layer or *plasma sheath* was given by Bohm for an ion–electron plasma [13]. The ions are assumed to arrive from the bulk plasma into the sheath with velocity v_0. The charge density of ions can be calculated by assuming a quasi-neutral situation $\rho_0 = \rho_\mathrm{i} = \rho_\mathrm{e}$ at the bulk plasma in the plasma potential $\phi = \phi_\mathrm{P}$, where $\phi = \phi_\mathrm{wall} = 0$ is the plasma electrode potential. Using ion continuity $\rho_0 v_0 = \rho_\mathrm{i} v_\mathrm{i}$ and energy conservation $m_\mathrm{i} v_\mathrm{i}^2/2 = m_\mathrm{i} v_0^2/2 - q_\mathrm{i}(\phi - \phi_\mathrm{P})$, the ion density becomes

$$\rho_\mathrm{i} = \rho_0 \sqrt{1 - \frac{2q_\mathrm{i}(\phi - \phi_\mathrm{P})}{m_\mathrm{i} v_0^2}}. \tag{58}$$

The electrons are assumed to be in thermal equilibrium and therefore they follow the Boltzmann distribution

$$\rho_\mathrm{e} = \rho_0 \exp\left(\frac{e(\phi - \phi_\mathrm{P})}{kT_\mathrm{e}}\right). \tag{59}$$

The potential in the sheath is described by the Poisson equation

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} = -\frac{\rho_0}{\epsilon_0}\left[\sqrt{1 - \frac{2q_\mathrm{i}(\phi - \phi_\mathrm{P})}{m_\mathrm{i} v_0^2}} - \exp\left(\frac{e\phi}{kT_\mathrm{e}}\right)\right]. \tag{60}$$

An important feature can be observed from the equation: the shielding condition, $\mathrm{d}\phi/\mathrm{d}x(x=0)=0$ is only fulfilled when the space charge is non-negative, i.e. $\rho_\mathrm{i} \geq \rho_\mathrm{e}$ for all $\phi \leq \phi_\mathrm{P}$. The necessary condition

$$v_0 \geq v_\mathrm{B} = \sqrt{\frac{kT_\mathrm{e}}{m_\mathrm{i}}} \tag{61}$$

is known as the *Bohm sheath criterion* and v_B as the *Bohm velocity*. The criterion sets a low-velocity limit for ions arriving at the sheath edge and in most cases the equation holds with equality [14]. The Poisson equation (60) is impossible to solve analytically, and often a numerical approach or approximations are used even in the presented one-dimensional case. The computational approach that is used in plasma extraction simulations in higher dimensions is presented in section 4.

The situation at the plasma extraction is simple according to the model (see Fig. 16(a)). Positive ions flow from quasi-neutral bulk plasma into the extraction sheath with velocity v_B. The compensating electron density, which is defined by the potential, is equal to the ion density in bulk plasma and decays exponentially towards the extraction. Far enough in the extraction, the compensation becomes (essentially) zero. From the model, it is obvious that there is no well-defined boundary between neutralized plasma and the unneutralized extraction. Often, such a boundary would be useful for judging the focusing action of the electric field close to the plasma electrode and for communicating about the plasma sheath shape. Therefore, an equipotential surface at ϕ_wall in the case of positive ion extraction is often chosen as an artificial 'boundary', known as the *plasma meniscus*. This choice works as a thought model even though in reality there is no such boundary. See Fig. 16(b) for a two-dimensional example of the plasma sheath.

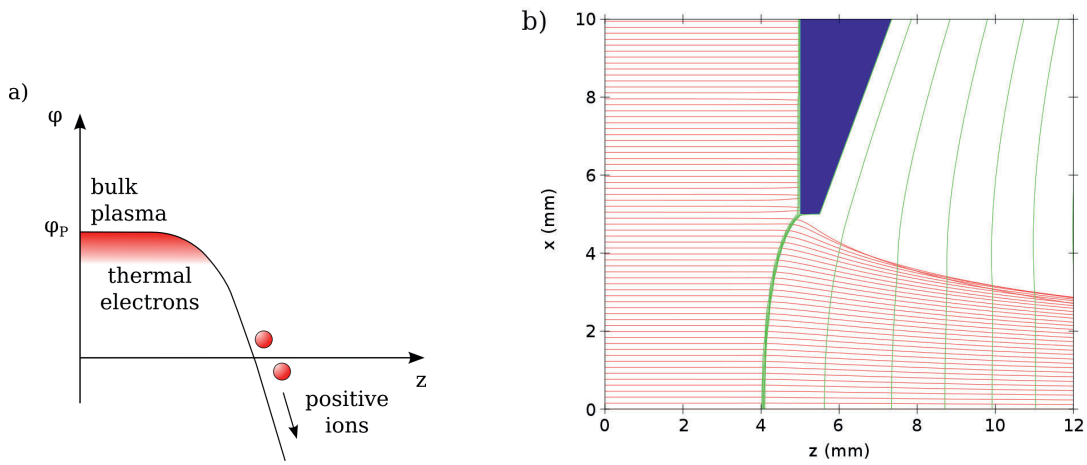


Fig. 16: Positive ion extraction. (a) Schematic representation of the model with thermal electrons populating the bulk plasma and positive ions accelerated by the extraction field through the plasma sheath. (b) Plot from a 2D simulation of a proton extraction with several equipotential lines drawn close to 0 V to visualize the location of the plasma meniscus. The ion temperature in the simulation is set to zero to show how the particle trajectories are accelerated perpendicular to the meniscus after leaving the plasma.

The process of beam formation varies not only with the extraction electric field strength and shape, but also with the properties of the plasma, i.e. plasma density, electron and ion temperatures. In Fig. 17, three simulated cases are shown with differing plasma densities. All other parameters are unchanged. In case (a) the plasma density is low and 25 mA of protons are extracted. The strong electric field in the extraction makes the plasma meniscus concave and the extracted beam is over-focused, causing increase in beam emittance. In case (b) the plasma density is higher and 60 mA of beam is extracted. The meniscus shape is almost flat, which provides the lowest beam emittance. In case (c), the plasma density is even higher and 95 mA of beam is extracted. The plasma meniscus is convex, the beam is divergent and the emittance is higher than in the optimal case. Because of this effect, it is important that

the electric field strength of the extraction system can be somehow adjusted if changing plasma densities are expected. Possible adjustments are changing the plasma electrode to puller electrode gap or changing the puller electrode voltage.

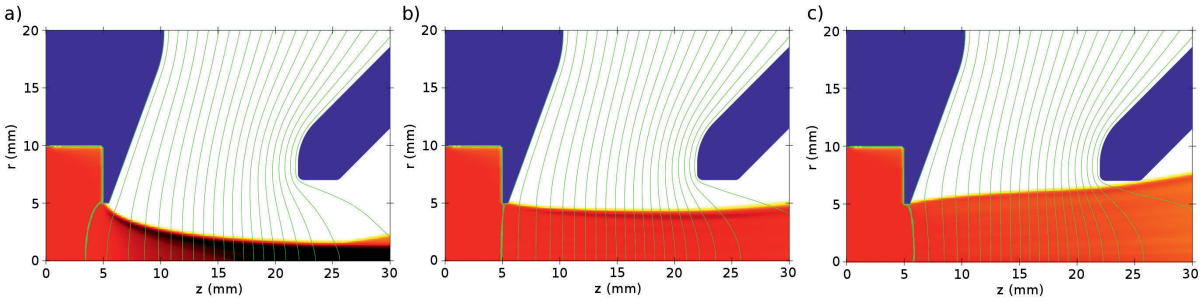


Fig. 17: Three simulations of 30 keV proton extraction with varying plasma densities. In (a) 25 mA of ions is extracted with 0.16 mm mrad r.m.s. emittance; in (b) 60 mA with 0.09 mm mrad; and in (c) 95 mA with 0.13 mm mrad r.m.s. emittance. The highest-quality beam is achieved with the flat plasma meniscus.

3.4 Negative ion plasma extraction

The negative ion plasma extraction model is similar to the positive ion extraction model presented above. The bulk plasma is at positive plasma potential ϕ_{P} and it is separated from the plasma electrode at $\phi = \phi_{\mathrm{wall}} = 0$ V by a plasma sheath. It is assumed that the extractable negative ions, which are either volume- or surface-produced, are born close to the wall potential and extracted from a uniform plasma volume. These charges form a potential well and counteract the formation of a saddle point at the extraction aperture. The non-existence of the saddle point is supported by the observed good emittance from H^- ion sources [15, 16]. The potential deviates from zero going into the bulk plasma due to the plasma potential and towards the extraction due to the acceleration voltage. This potential structure causes positive ions from the bulk plasma to be accelerated towards the extraction, having energy $e\phi_{\mathrm{P}}$ at the zero potential. These ions propagate until they are reflected back into the plasma by the increasing potential in the extraction. The potential well acts as a trap for thermal positive ions. The negative ions and electrons are accelerated from the wall potential towards the bulk plasma and more importantly towards the extraction. A schematic view of the negative ion extraction model is shown in Fig. 18 [17].

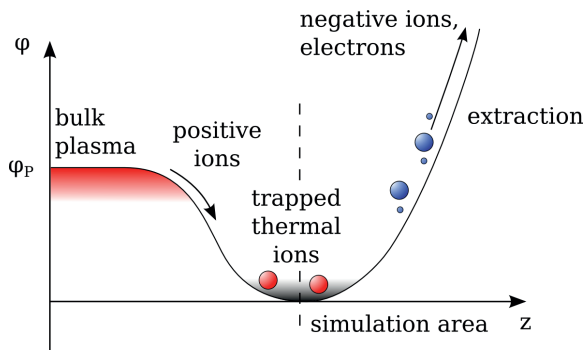


Fig. 18: Schematic representation of negative ion plasma extraction model with fast positive ions flowing from bulk plasma towards the extraction, thermal positive ions populating the potential well at the plasma electrode potential, and negative charged particles accelerated by the extraction field.

The negative ion plasma sheath from the zero potential towards the extraction can be described by the Poisson equation

$$\nabla^2\phi = -\frac{\rho}{\epsilon_0}, \tag{62}$$

where $\rho = \rho_{\text{neg}} + \rho_{\text{f}} + \rho_{\text{th}}$. Here ρ_{neg} is the space-charge density of negative particles, ρ_{f} is the space-charge density of fast positive ions and ρ_{th} is the space-charge density of trapped positive thermal ions.

The model allows several different negative ion species to be extracted from the ion source and also many positive ion species to be used as compensating plasma particles. Each of the thermal ion species has a separate Maxwellian velocity distribution with the associated space-charge distribution

$$\rho_{\text{th}} = \rho_{\text{th},0} \exp\left(\frac{-e\phi}{kT_{\text{i}}}\right), \tag{63}$$

where $\rho_{\text{th},0}$ is the space-charge density of the thermal ion species at the wall potential and T_{i} is the corresponding thermal ion temperature. The fast ions are decelerated and turned back into the plasma by the extraction voltage. The space-charge distribution of the fast ions is defined by the virtual cathode formation and it is

$$\rho_{\text{f}} = \rho_{\text{f},0}\left(1 - \frac{e\phi}{E_{\text{i}}}\right), \tag{64}$$

at $e\phi < E_{\text{i}}$ and zero otherwise. Here $\rho_{\text{f},0}$ is the space-charge density of fast ions at the wall potential and E_{i} is the corresponding kinetic energy, which should be around $e\phi_{\text{P}}$ as the particles are flowing from the bulk plasma. The quasi-neutrality of the plasma requires $\rho_{\text{neg}} + \rho_{\text{f}} + \rho_{\text{th}} = 0$ at $\phi = 0$ V.

In the negative ion extraction, the plasma sheath acts similarly to the positive ion extraction case. The smallest beam emittance is achieved with extraction field optimized for the plasma density of the ion source. The biggest difference from the positive ion case is that, with a positive puller electrode voltage, also electrons will be extracted from the plasma in addition to the negative ions. Depending on the ion source, the amount of co-extracted electrons may be as high as 100–200 times the amount of negative ions extracted or as low as 1 as is the case in caesiated surface production H^- sources. Especially in the cases where the amount of electrons is high, the electrons need to be dumped in a controlled manner as soon as possible to avoid unnecessary emittance growth due to the additional space charge. Often the electron beam current is so high that the dumping cannot be done at the full beam energy required by the application. In other words, the electron beam has to be dumped on an intermediate electrode at lower potential than ground.

Typically the electron dumping is done by utilizing a magnetic dipole field generated with permanent magnets. As the magnetic field also deflects the negative ion beam, the negative ion source may be mounted to the rest of the beam transport line at an angle to compensate for the deflection. Another solution is to use another dipole field in the opposite direction to correct the angle of the negative ion beam back perpendicular to the original axis. The resulting offset in the beam centre has to be corrected using deflector plates, XY magnets or mechanical offset.

There are three locations, which are generally used for dumping the electron beam in negative ion sources: 1. Puller electrode with low voltage with respect to the ion source can be used as an electron dump [17, 18], 2. a separate intermediate electrode before the puller electrode can be used if a higher voltage is needed on the puller electrode [19] or 3. the electrons are dumped on an electrode after the puller electrode, as shown in Fig. 19. The third method has the advantage that the voltage of the puller electrode can be optimized for plasma density matching without affecting the electron dumping [20, 21].

4 Computer codes for beam extraction and transport problems

The modelling and analysis of modern beam extraction and transport systems is so complicated that most of the work is done using specialized computer simulation codes. In this section, a brief overview of the wide spectrum of codes is presented.

4.1 Transfer matrix codes

Traditionally, modelling of ion optical transport lines has been done using the transfer matrix formalism, which describes the effect of ion optical elements (and drifts between them) on trajectories using transfer

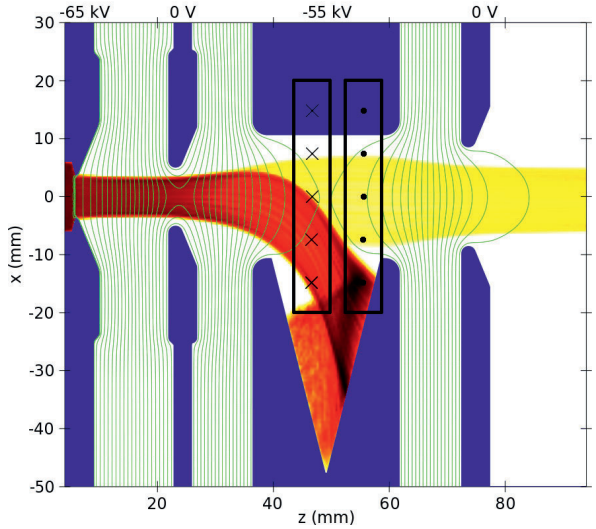


Fig. 19: An extraction system designed for the SNS for transporting up to 100 mA of H^- and for dumping up to 1 A of electrons into a V-shaped electron dump. The puller electrode can be adjusted to cope with varying current levels. The electron dump (einzel) electrode has permanent magnets embedded for a dipole–antidipole field structure [21].

matrices. In the computer program TRANSPORT [22], for example, the properties of a single-particle trajectory at any given point of the transport line are described by a vector $\boldsymbol{X} = (x, x', y, y', l, \delta)$, where x and y are the transverse displacements of the trajectory with respect to the reference trajectory, x' and y' are the tangents of angles of the trajectory with respect to the reference trajectory, l is the path-length difference between the trajectory and the reference trajectory, and $\delta = \Delta p/p$ is the fractional momentum deviation of the trajectory from the reference trajectory. The trajectory component X_i after propagation through an ion optical element can be calculated from

$$X_i = \sum_j Y_j \left\{ (X_i \mid Y_j) + \sum_k \frac{Y_k}{2} \left\{ (X_i \mid Y_j Y_k) + \sum_l \frac{Y_l}{3} \left\{ (X_i \mid Y_j Y_k Y_l) + \cdots \right\} \right\} \right\}, \tag{65}$$

where Y_i are the components of the trajectory before the ion optical element, and $(X_i \mid Y_j)$, $(X_i \mid Y_j Y_k)$, $(X_i \mid Y_j Y_k Y_l)$, ... are the first-order, second-order, third-order, ... transfer coefficients. This can be described as matrix–vector multiplication, with a 6×6 matrix in first order, a 6×6^2 matrix in second order, a 6×6^3 matrix in third order, etc., using an extended trajectory vector containing also the higher-order terms, for example, $(x, x', y, y', l, \delta, x^2, xx', xy, xy', xl, x\delta, x'^2, x'y, \ldots, \delta^2)$ in the second order. If the element matrices are expanded as square matrices, the propagation of the trajectory through the whole beamline can be calculated from the product of the element matrices $\boldsymbol{R}$ instead of using Eq. (65) for each element $\boldsymbol{R}(1)$, $\boldsymbol{R}(2)$, ... The propagation of trajectory $\boldsymbol{Y}$ through the whole system is given by

$$\boldsymbol{X} = \boldsymbol{R}\,\boldsymbol{Y} = \boldsymbol{R}(n) \cdots \boldsymbol{R}(3)\,\boldsymbol{R}(2)\,\boldsymbol{R}(1)\,\boldsymbol{Y}. \tag{66}$$

The cumulative matrix $\boldsymbol{R}$ can be used to calculate several trajectories through the transport line much faster than just by using the projection from element to element one after another.

The first-order approximation matrices can be easily derived analytically by assuming that the displacements and angles relative to the reference trajectory are small enough to justify the truncation of

higher-order terms. For example, the first-order matrix for a magnetic quadrupole lens is

$$\boldsymbol{R} = \begin{pmatrix} \cos kL & (1/k)\sin kL & 0 & 0 & 0 & 0 \\ -k\sin kL & \cos kL & 0 & 0 & 0 & 0 \\ 0 & 0 & \cosh kL & (1/k)\sinh kL & 0 & 0 \\ 0 & 0 & k\sinh kL & \cosh kL & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & L/\gamma^2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{67}$$

where L is the effective length of the quadrupole,

$$k_B^2 = \frac{B_\mathrm{T}}{a}\frac{q}{p}$$

and γ is the relativistic factor. Some ion optical effects can also be derived to higher-order approximations analytically. Usually the highest-order matrices are constructed using numerical integration of known fields.

In addition to transporting single trajectories or distributions, the same matrix formalism can be used to transport elliptical beam envelopes through the transport line in the first order. In this case, the beam envelope (in one dimension) is described by the beam matrix

$$\boldsymbol{\sigma} = \epsilon \begin{pmatrix} \beta & -\alpha \\ \alpha & \gamma \end{pmatrix}, \tag{68}$$

where α, β and γ are the Twiss parameters and ϵ is the beam emittance. The transformation of the envelope $\boldsymbol{\sigma}$ through a system described by matrix $\boldsymbol{R}$ can be calculated with

$$\boldsymbol{\sigma}' = \boldsymbol{R}\,\boldsymbol{\sigma}\,\boldsymbol{R}^\mathrm{T}. \tag{69}$$

Programs using the transfer matrix formalism are typically used for modelling long beam transport lines at relatively high energies, where the beam size and angles are small. The calculation of trajectories is fast, which allows automatic optimization of ion optics, etc. Many of the programs even have models for space-charge-induced divergence growth (similar to what was presented in section 2.3.1) and/or r.m.s. emittance growth modelling using the paraxial approximation for particle distributions, which makes them more suitable for modelling the low-energy and high-current systems such as LEBT and accelerator injection systems. The use of the space-charge models, of course, makes the calculation slower, but the codes are still much faster than the ray-tracing codes discussed in section 4.2. Some of the most commonly used codes of this type include the following:

- COSY INFINITY [23], up to infinite order, no space charge modelling
- DIMAD [24], up to third-order calculation, space charge of KV beam
- GIOS [25], up to third-order calculation, space charge of KV beam
- PATH MANAGER (TRAVEL) [26], up to second order, more advanced space-charge modelling for particle distributions (mesh or Coulomb model)
- TRACE-3D [27], mainly linear with space charge of KV beam
- TRANSPORT [22], up to third-order calculation, no space-charge modelling

The transfer matrix codes are unfortunately not usable for electrostatic extraction systems, because no general matrices exist for describing the optics of the electrode systems.

4.2 Ray-tracing and extraction codes

For systems where the approximations made in the transfer matrix codes are not valid, another more fundamental method has to be used. An approach taken by so-called *ray-tracing codes* is to directly integrate the particle equation of motion (Eq. (4)) using mesh-based maps for $\boldsymbol{E}$ and $\boldsymbol{B}$ fields. In the extraction system case, it is typically assumed that the $\boldsymbol{B}$ field is defined only by external sources, i.e. the beam-generated magnetic field is negligible. On the other hand, the space charge of the beam plays a major role, and the self-consistent solution of the beam transport can be found using the iterative approach shown in Fig. 20.

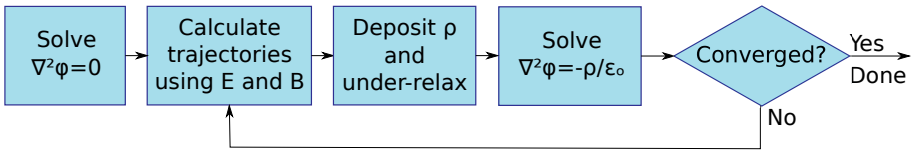


Fig. 20: Flow diagram for the iterative solution of high-space-charge beam transport with ray tracing

The iteration starts with computation of the electric potential on the system using the Laplace equation. The electric field is then calculated from the potential, and particles are traced through the system. The current carried by the particles is deposited onto the underlying mesh for the following solution of the Poisson equation for a new estimate of the electric potential. The iteration is continued like this until convergence is reached. A code of this type may also have a nonlinear Poisson solver for taking into account the compensating particles of the plasma using analytical formulas, as described in sections 3.3 and 3.4. Pioneering work in the field using this method was done by S. A. Self [28] and J. H. Whealton [29, 30] in the 1960s and 1970s. This type of code is often called an *extraction code* or *ion gun code*. The most used codes of this type include the following:

- IGUN [15], a code with plasma modelling for negative and positive ions, only 2D and cylindrically symmetric geometries
- PBGUNS [31], plasma modelling for positive and negative ions, only 2D and cylindrically symmetric geometries
- KOBRA [32], a more advanced 3D E field solver, positive ion plasma modelling, simple particle-in-cell capability
- IBSIMU [33, 34], 1D, 2D or 3D and cylindrically symmetric E field solver, plasma modelling for positive and negative ions

In the following, the typical methods used in extraction codes are described in more detail. The particular choices presented here are those that are used in the IBSIMU code, which was written by the author, but many codes of this type use the same or similar methods.

4.2.1 Electric potential and field

The geometry of the simulation domain is discretized using a regular grid, where the coordinates of calculation nodes can be calculated from integer indices i, j, k, with $x_i = x_0 + ih$, $y_j = y_0 + jh$ and $z_k = z_0 + kh$. The Poisson equation

$$\nabla^2\phi = \frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} + \frac{\mathrm{d}^2\phi}{\mathrm{d}y^2} + \frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} = -\frac{\rho}{\epsilon_0} \tag{70}$$

in three dimensions is discretized using the finite difference method (FDM), i.e. by replacing the derivatives with finite differences. The Poisson equation becomes

$$\frac{\phi_{i-1,j,k} + \phi_{i+1,j,k} + \phi_{i,j-1,k} + \phi_{i,j+1,k} + \phi_{i,j,k-1} + \phi_{i,j,k+1} - 6\phi_{i,j,k}}{h^2} = -\frac{\rho_{i,j,k}}{\epsilon_0} \tag{71}$$

for nodes that do not have electrodes as close neighbours. The nodes that are close to electrodes use a modified form of the Poisson equation with uneven node distances to take into account the real distance of the calculation node from the surface for more accurate, smooth solution near the surface. The finite difference for the partial derivative in the x direction becomes

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} = \frac{\beta\phi(x_0 - \alpha h) - (\alpha+\beta)\phi(x_0) + \alpha\phi(x_0 + \beta h)}{\frac{1}{2}(\alpha+\beta)\alpha\beta h^2}, \tag{72}$$

where αh is the distance from x_i to the location where the potential is known in the negative x direction and βh is the distance from x_i to the location where the potential is known in the positive x direction. The near-solid distances (α and β coefficients) are stored in a table for all calculation nodes that are neighbours to surfaces. The nodes that are on the boundaries of the simulation domain have to be constrained by a boundary condition, either a Dirichlet boundary condition (fixed potential)

$$\phi_{i,j,k} = \phi_{\text{const.}} \tag{73}$$

or a Neumann boundary condition (fixed derivative with respect to the normal of the boundary)

$$\frac{-3\phi_{i,j,k} + 4\phi_{i+1,j,k} - \phi_{i+2,j,k}}{2h} = \left(\frac{\mathrm{d}\phi}{\mathrm{d}x}\right)_{\text{const.}}, \tag{74}$$

in the case of a boundary with normal in the positive x direction. The numerical formulation of the Poisson equation then becomes a system of N simultaneous equations, where N is the number of free (non-constant) nodes. The problem is often described as a matrix equation

$$\boldsymbol{A}\,\boldsymbol{\phi} = \boldsymbol{B}, \tag{75}$$

where $\boldsymbol{A}$ is an $N \times N$ matrix of coefficients, $\boldsymbol{B}$ is a vector of coefficients from Eqs. (71)–(74) and $\boldsymbol{\phi}$ is the solution electric potential vector. In the case of plasma modelling, the Poisson equation contains an analytical term for the space-charge density of the compensating plasma particles. This leads to a nonlinear Poisson problem, which is typically formulated as

$$\boldsymbol{A}(\boldsymbol{\phi}) = 0. \tag{76}$$

In the typical scale of systems being investigated, N is 10^6–10^8, which makes the solution of Eq. (76) computationally intensive. The problem is typically solved using N-dimensional Newton–Raphson methods with iterative linear solvers.

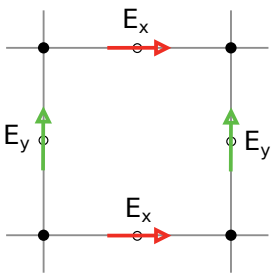


Fig. 21: The electric field is evaluated on a different mesh with $h/2$ offset from the electric potential mesh to avoid further approximation.

The electric field is calculated using meshes with $h/2$ location offset from the electric potential map nodes to avoid making further approximations. For E_x field the node location is $x_i = x_0 + h/2 + ih$, while y_j and z_k coordinates of the nodes are as before. The electric field is evaluated with

$$E_{x,i,j,k} = \frac{\phi_{i,j,k} - \phi_{i+1,j,k}}{h}, \tag{77}$$

and similarly for E_y and E_z as shown in Fig. 21. The near-solid distances (α and β) described before are also used here to modify the electric field calculation near the solids.

4.2.2 Trajectory calculation

The particles in this type of simulation represent several physical particles, and therefore they are typically called *macro-particles*. Each particle is given charge q and mass m, which are parameters of the physical particle, and current I, which is a parameter of the macro-particle. The particle also has location $\boldsymbol{x}$ and velocity $\boldsymbol{v}$, which are set according to the definition given by the user. In a typical case, where particles are starting inside the plasma, the location is sampled from a quasi-random distribution to fill a cylindrical surface on the boundary of the simulation domain. The longitudinal z component of the velocity is defined to be larger than the Bohm velocity, while the transverse components are sampled from a Gaussian distribution with standard deviation $\sigma_v = \sqrt{T_\mathrm{i} e/m}$, where T_i is the ion transverse temperature. The code also contains many other possibilities for beam definition, such as rectangular beams, beams with KV- or Gaussian-distributed emittance pattern or definition of beam particle-by-particle.

The propagation of particles in the simulation domain is calculated by integrating the particle equation of motion in the form of a set of ordinary differential equations. Using the non-relativistic approximation in 3D, the set is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v_x, \tag{78}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = v_y, \tag{79}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = v_z, \tag{80}$$

$$\frac{\mathrm{d}v_x}{\mathrm{d}t} = a_x = \frac{q}{m}(E_x + v_y B_z - v_z B_y), \tag{81}$$

$$\frac{\mathrm{d}v_y}{\mathrm{d}t} = a_y = \frac{q}{m}(E_y + v_z B_x - v_x B_z), \tag{82}$$

$$\frac{\mathrm{d}v_z}{\mathrm{d}t} = a_z = \frac{q}{m}(E_z + v_x B_y - v_y B_x). \tag{83}$$

The integration of the system of equations is done with a Runge–Kutta Cash–Karp adaptive algorithm and, at each step, the full set of particle coordinates $(t, x, v_x, y, v_y, z, v_z)$ is stored.

4.2.3 Space-charge deposition

The particle trajectories are used to calculate the space-charge density in the simulation domain. This is done by depositing the charge carried by the trajectories into the nodes of the mesh in which the Poisson equation is solved. Each particle trajectory carries current I. The simulated macro-particles must have a finite size so that space charge can be defined. The simplest scheme for space-charge deposition would be to assume a box-like particle shape with

$$\rho(x_1, x_2) = \begin{cases} \dfrac{I}{h^2 v} & \text{if } -\frac{1}{2}h < x_1 < \frac{1}{2}h \text{ and } -\frac{1}{2}h < x_2 < \frac{1}{2}h, \\ 0 & \text{elsewhere,} \end{cases} \tag{84}$$

where x_1 and x_2 are the coordinates transverse to the trajectory. This is known as the closest-node weighting and it is prone to numerical noise. A better solution is to distribute the space charge to closest nodes with bilinear weighting. The particle space-charge distribution

$$\rho(x_1, x_2) = \begin{cases} \dfrac{I}{h^2 v}(1 - |x_1/h|)(1 - |x_2/h|) & \text{if } -h < x_1 < h \text{ and } -h < x_2 < h, \\ 0 & \text{elsewhere,} \end{cases} \tag{85}$$

leads to much smoother space-charge densities. The charge deposition to the nodes is done based on the closest distance from the trajectory, as shown in Fig. 22.

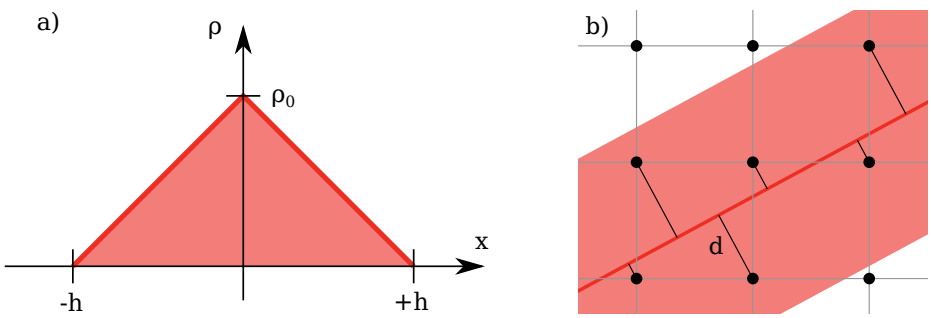


Fig. 22: (a) In 2D the particle shape is a triangle with base width of $2h$ and height $\rho_0 = \lambda/hv$, where λ is line charge density carried by the 2D trajectory. (b) As the particle passes through the domain, it covers mesh nodes with space-charge density according to the particle shape. The space charge deposited to mesh nodes is calculated according to the closest distance d of the node from the trajectory.

Even with the smooth space charge deposition of Eq. (85), the overall iteration may not converge in all cases. In a typical case a region of space gets a high ρ, which leads to electric fields directing the particles away from the region in the next iteration round. This will lead to a low ρ, which then leads to electric fields directing the particles into the area. This phenomenon happens easily especially where the particle velocities are low and trajectories get deflected by low electric fields. A simple method for suppressing the effect is to use under-relaxation of the space-charge map to avoid the over-shooting of ρ.

4.3 Other useful computer codes for ion optics

Other computer codes that may be useful for ion optics include POISSON SUPERFISH [35] and FEMM [36], which can be used to calculate magnetostatics in 2D and cylindrically symmetric geometries using the finite element method, and RADIA-3D [37], which uses boundary integral methods to solve magnetostatic problems in 3D. Commercial computational suites such as COBHAM VECTOR FIELDS [38], COMSOL MULTIPHYSICS [39] and IES LORENTZ [40] provide advanced finite element- and boundary element-based field solvers. These packages provide nicely packaged, easily used tools with graphical user interfaces for typical problems such as electrostatic and magnetostatic field solution, heat transfer and even charged-particle tracers. Unfortunately, serious plasma modelling capabilities are still missing from the commercial packages. All of the above-mentioned codes can be used together with IBSIMU, for example, for providing B field maps of magnetic elements.

There also exist many specialized programs for modelling space-charge compensation, beam bunching, cyclotron injection, radio frequency quadrupole (RFQ) acceleration, collisional ion source plasmas, etc., with many different methods. The reader is encouraged to seek more information about such software from the literature and the World Wide Web.

References

[1] R. Thomae, R. Gough, R. Keller, K. N. Leung, T. Schenkel, A. Aleksandrov, M. Stockli and R. Welton, Beam measurements on the H^- source and low energy beam transport system for the Spallation Neutron Source. *Rev. Sci. Instrum.* **73** (2002) 2016.

[2] H. Liebl, *Applied Charged Particle Optics* (Springer, Berlin, 2008).

[3] H. Wollnik, *Optics of Charged Particles* (Academic Press, Orlando, FL, 1987).

[4] V. Kumar, Understanding the focusing of charged particle beams in a solenoid magnetic field. *Am. J. Phys.* **77** (2009) 737.

[5] M. P. Stockli, R. F. Welton and R. Keller, Self-consistent, unbiased root-mean-square emittance analysis. *Rev. Sci. Instrum.* **75** (2004) 1646.

[6] D. Leitner, Ion beam properties and their diagnostics for ECR ion source injector systems, Proc. 14th Beam Instrumentation Workshop, Santa Fe, NM, USA, May 2010.

[7] S. Humphries, *Charged Particle Beams* (John Wiley & Sons, New York, 1990).

[8] J.-L. Vay, P. Colella, J. W. Kwan, P. McCorquodale, D. B. Serafini, A. Friedman, D. P. Grote, G. Westenskow, J.-C. Adam, A. Héron and I. Haber, Application of adaptive mesh refinement to particle-in-cell simulations of plasmas and beams. *Phys. Plasmas* **11** (2004) 2928.

[9] N. Chauvin, Space-Charge Effects, in these proceedings.

[10] D. S. Todd, D. Leitner and M. Strohmeier, Low energy beam diagnostics at the VENUS ECR ion source, Proc. 13th Beam Instrumentation Workshop, Tahoe City, CA, May 2008.

[11] C. D. Child, Discharge from hot CaO. *Phys. Rev.* **32** (1911) 492.

[12] J. R. Pierce, Rectilinear electron flow in beams. *J. Appl. Phys.* **11** (1940) 548.

[13] D. Bohm, Minimum ionic kinetic theory for a stable sheath, in *The Characteristics of Electrical Discharges in Magnetic Fields*, Eds. A. Guthrie and R. K. Wakerling (McGraw-Hill, New York, 1947).

[14] B. N. Chapman, *Glow Discharge Processes* (John Wiley & Sons, New York, 1980).

[15] R. Becker, NIGUN: a two-dimensional simulation program for the extraction of H^- ions. *Rev. Sci. Instrum.* **75** (2004) 1723.

[16] R. Becker, Mathematical formulation and numerical modelling of the extraction of H^- ions, Proc. 10th Int. Symp. on the Production and Neutralization of Negative Ions and Beams. *AIP Conf. Proc.* **763** (2005) 194.

[17] T. Kalvas, O. Tarvainen, H. Clark, J. Brinkley and J. Ärje, Application of 3D code IBSimu for designing an H^-/D^- extraction system for the Texas A&M facility upgrade, Proc. 2nd Int. Symp. on Negative Ions, Beams and Sources, Takayama, Japan. *AIP Conf. Proc.* **1390** (2011) 439.

[18] T. Kuo, D. Yuan, K. Jayamanna, M. McDonald, R. Baartman, P. Schmor and G. Dutto, On the development of a 15 mA direct current H^- multicusp source. *Rev. Sci. Instrum.* **67** (1996) 1316.

[19] R. Keller, D. Cheng, R. DiGennaro, R. A. Gough, J. Greer, K. N. Leung, A. Ratti, J. Reijonen, R. W. Thomae, T. Schenkel, J. W. Staples, R. Yourd, A. Aleksandrov, M. P. Stockli and R. W. Welton, Ion-source and low-energy beam-transport issues with the front-end systems for the Spallation Neutron Source. *Rev. Sci. Instrum.* **73** (2002) 914.

[20] Ø. Midttun, T. Kalvas, M. Kronberger, J. Lettry, H. Pereira, C. Schmitzer and R. Scrivens, A new extraction system for the Linac4 H^- ion source. *Rev. Sci. Instrum.* **83** (2012) 02B710.

[21] T. Kalvas, R. F. Welton, O. Tarvainen, B. X. Han and M. P. Stockli, Simulation of H^- ion source extraction systems for the Spallation Neutron Source with IBSimu. *Rev. Sci. Instrum.* **83** (2012) 02A705.

[22] D. C. Carey, K. L. Brown and F. Rothacker, Third-order TRANSPORT with MAD input – a computer program for designing charged particle beam transport systems, FERMILAB-Pub-98/310, Fermi National Accelerator Laboratory, October 1998.

[23] K. Makino and M. Berz, COSY INFINITY Version 9. *Nucl. Instrum Meth.* **A558** (2005) 346.

[24] R. V. Servranckx, Users' guide to the program DIMAD, TRI-DN-93-K233, Triumf Design Note, July 1993.

[25] H. Wollnik, B. Hartmann and M. Berz, Principles of GIOS and COSY. *AIP Conf. Proc.* **177** (1988) 74.

[26] A. Perrin, J.-F. Amand, T. Mütze, J.-B. Lallement and S. Lanzone, Travel user manual, CERN, April 2007.

[27] K. R. Crandall and D. P. Rusthoi, TRACE 3-D documentation, LA-UR-97-886, Los Alamos National Laboratory Report, May 1997.

[28] S. A. Self, Exact solution of the collisionless plasma-sheath equation. *Phys. Fluids* **6** (1963) 1762.

[29] J. H. Whealton, Optics of single-stage accelerated ion beams extracted from a plasma. *Rev. Sci. Instrum.* **48** (1977) 829.

[30] J. H. Whealton, E. F. Jaeger and J. C. Whitson, Optics of ion beams of arbitrary perveance extracted from a plasma. *J. Comput. Phys.* **27** (1978) 32.

[31] J. E. Boers, PBGUNS: a digital computer program for the simulation of electron and ion beams on a PC, Proc. Int. Conf. on Plasma Sciences, Vancouver, BC, 7–9 June 1993.

[32] P. Spädtke, KOBRA3-INP user manual, 2000.

[33] T. Kalvas, O. Tarvainen, T. Ropponen, O. Steczkiewicz, J. Ärje and H. Clark, IBSimu: a three-dimensional simulation software for charged particle optics. *Rev. Sci. Instrum.* **81** (2010) 02B703.

[34] T. Kalvas, Ion beam simulator, the distribution website of IBSimu code, http://ibsimu.sourceforge.net.

[35] K. Halbach and R. F. Holsinger, SUPERFISH – a computer program for evaluation of RF cavities with cylindrical symmetry. *Part. Accel.* **7** (1976) 213.

[36] D. C. Meeker, Finite element method magnetics, Version 4.0.1 (3 December 2006 build), http://www.femm.info.

[37] O. Chubar, P. Elleaume and J. Chavanne, Radia3D – a computer program for calculating static magnetic fields, http://www.esrf.eu/Accelerators/Groups/InsertionDevices/Software/Radia.

[38] Cobham Vector Fields, Opera, http://www.cobham.com.

[39] Comsol Multiphysics, http://www.comsol.com.

[40] Integrated Engineering Software, http://www.integratedsoft.com.

An Application for Research – the Large Hadron Collider

R. Bailey
CERN, Geneva, Switzerland

Abstract
The Large Hadron Collider (LHC) machine at CERN was designed and built primarily to find or exclude the existence of the Higgs boson, for which a large amount of data is needed by the LHC experiments. This requires operation at high luminosity, which in turn requires running with thousands of high-intensity proton bunches in the machine. After quantifying the data required by the experiments and elucidating the LHC parameters needed to achieve this, this paper explains how the LHC beams are fabricated from the pulse(s) coming from the CERN Duoplasmatron source.

1 Introduction

At the start of 2012, data from Large Electron–Positron collider (LEP), the Tevatron and the Large Hadron Collider (LHC) excluded the existence of the Higgs boson for the mass regions $m_H < 114$ GeV and $m_H > 150$ GeV. The expected cross-section for the production of the Higgs boson in the remaining mass range at LHC energies is measured in tens of picobarns (pb), to be compared with the total inelastic proton–proton cross-section, measured in tens of millibarns (mb) (1 barn = 10^{-28} m^2). When combined with the expected decay branching ratios to the experimentally most sensitive channels at LHC, values of the quantity cross-section × branching ratio are measured in femtobarns (fb). Hence large amounts of data are needed to produce and detect merely a handful of Higgs boson events.

The quantity used to define the instantaneous performance level of a collider is luminosity, L, measured in cm^{-2} s^{-1}. An average luminosity $\langle L \rangle$ delivered over a period of time t in collision yields a value of the delivered luminosity $L_{DEL} = \langle L \rangle t$, measured in barn^{-1}. Running for 10^6 s at an average luminosity of 10^{33} cm^{-2} s^{-1} yields an integrated luminosity of 1 fb^{-1}.

LHC therefore has to operate routinely at high luminosity. The design value is 10^{34} cm^{-2} s^{-1}, which requires a number of parameters to be achieved, not least the huge number of bunches in each counter-rotating beam. This is realized by a complex series of manipulations in the LHC injector chain.

2 Cross-sections and branching ratios

The expected cross-sections for production of the Higgs boson at $\sqrt{s}$ of 7 TeV (LHC in 2011) are shown in Fig. 1 [1]. For $m_H < 1$ TeV, the dominant process is gluon fusion and the expected cross-section is of the order of 10 pb in the Higgs mass range not already excluded.

The expected branching ratios for the decay of the Higgs, as a function of the Higgs mass, are shown in Fig. 2 [1]. For example, at a Higgs mass ~125 GeV, the branching ratio Higgs → ZZ is around 2×10^{-2}. To detect the Z through its decay into any of the three lepton families, a further branching ratio of 10^{-1} has to be applied to the decay of each Z. Therefore, the resulting branching ratio of Higgs → ZZ → four leptons is around 2×10^{-4}.

For a Higgs mass of 125 GeV, combining the production cross-section (σ) at $\sqrt{s}$ of 7 TeV with the branching ratio (BR) of Higgs → ZZ → four leptons, we obtain a value of σBR of 2 fb.

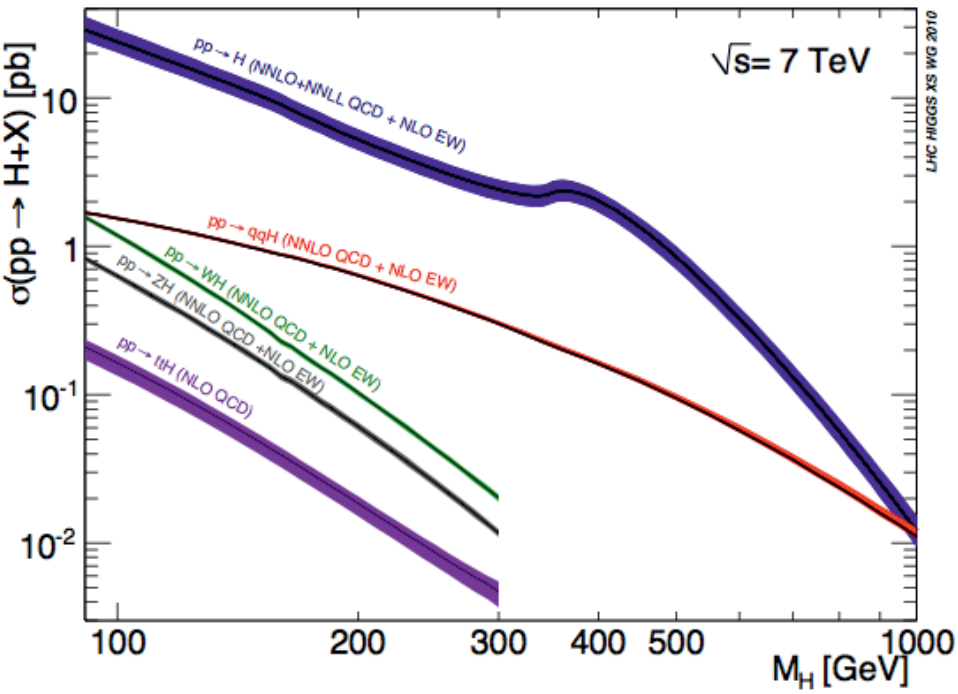


Fig. 1: Expected cross-sections for production of the Higgs boson at $\sqrt{s}$ = 7 TeV

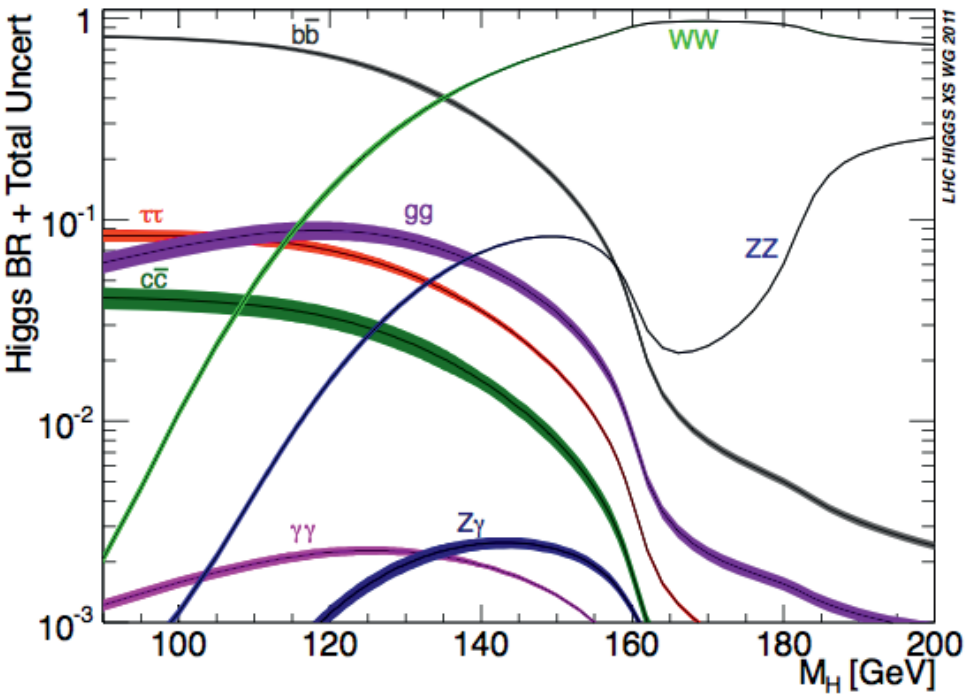


Fig. 2: Expected Higgs decay branching ratios

3 Luminosity and associated parameters

The performance of a collider is characterized by the luminosity delivered, L_{DEL}, over a period of time. If the average instantaneous luminosity over the time period t is $\langle L \rangle$, then

$$L_{\mathrm{DEL}} = \langle L \rangle t \text{ barn}^{-1}. \tag{1}$$

The number of events N_{event} with production cross-section σ and branching ratio BR detected in these data is given by

$$N_{\mathrm{event}} = \sigma \mathrm{BR}\, L_{\mathrm{DEL}}. \tag{2}$$

For the Higgs → ZZ → four leptons process mentioned above, a delivered luminosity of 5 fb^{-1} would yield just 10 events. In 2011, the LHC provided colliding beams for approximately 5×10^6 s in order to deliver 5 fb^{-1} of integrated luminosity, which from Eq. (1) corresponds to an average instantaneous luminosity of 10^{33} cm^{-2} s^{-1}. When one considers that the luminosity decays during a physics fill of several hours' duration, it is clear that instantaneous luminosities well in excess of 10^{33} are required to deliver even tens of these kinds of events per year at the LHC.

The luminosity at any moment can be written as

$$L = \frac{N^2 k_{\mathrm{b}} f \gamma}{4\pi \varepsilon_{\mathrm{n}} \beta^*} F \tag{3}$$

where γ, the relativistic factor, is given by the beam energy E divided by the proton rest mass, f is the revolution frequency, and F is a reduction factor resulting from operating with a crossing angle, θ_{c}.

The explanation of the other variables, together with LHC parameters for nominal, 2011 values and aims for 2012, is given in Table 1.

Table 1: LHC parameters in 2011 and 2012 compared to nominal.

Symbol	Meaning	Units	Nominal	2011	2012
E	Beam energy	TeV	7.0	3.5	4.0
N	Number of particles per bunch		1.15×10^{11}	1.5×10^{11}	1.5×10^{11}
k_{b}	Number of bunches per beam†		2808	1380	1380
ε_{n}	Normalized emittance (= $\varepsilon\gamma$)	μm rad	3.75	2.5	2.5
β^*	Beat function at IP 1 and 5	m	0.55	1.0	0.6
θ_{c}	Crossing angle through IP 1 and 5	μrad	285	240	290
L	**Peak instantaneous luminosity**	**cm^{-2} s^{-1}**	**10^{34}**	**3.5×10^{33}**	**6×10^{33}**

†Owing to the details of the filling scheme, a few per cent of bunches do not contribute to luminosity.

While the energy, the beta function at the interaction point (IP) and the crossing angle are all very much in the domain of the LHC itself, the collider is reliant on the injector chain to provide beams with thousands of high-intensity bunches of small transverse emittance.

4 The LHC injector chain

The LHC is supplied with protons from an injector chain comprising accelerators that have been in use for many years, namely Linac2, the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), as shown in Fig. 3. These accelerators were all upgraded in the 1990s to meet the very stringent needs of the LHC.

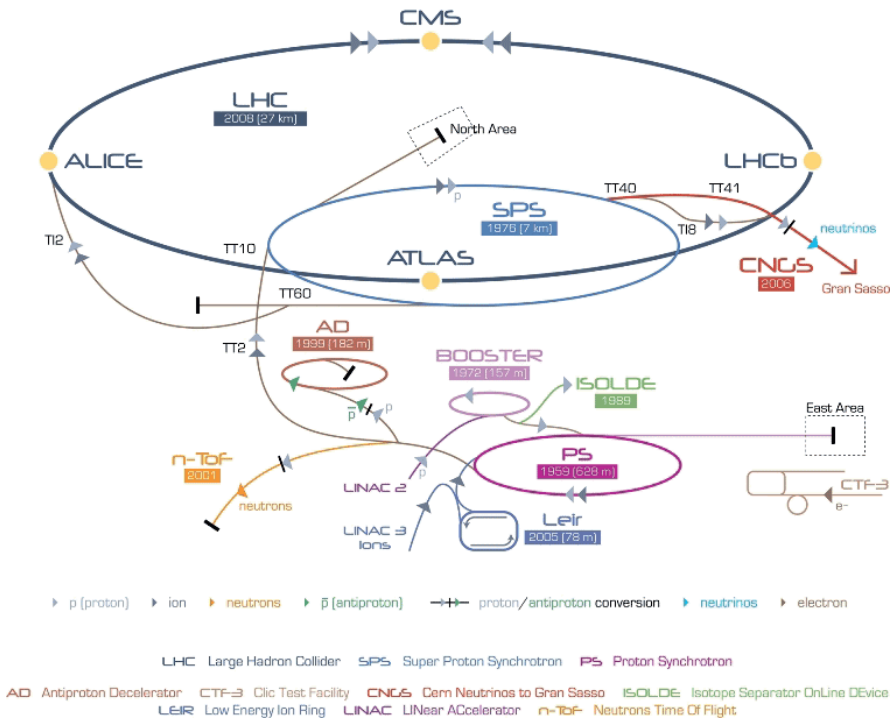


Fig. 3: The CERN accelerator complex

Linac2 has been the primary source of protons for the CERN accelerator complex for the past 30 years and has been modified to meet the demands of the LHC. The ion source is a Duoplasmatron giving up to 300 mA of beam current, followed by a four-vane Radio Frequency Quadrupole (RFQ) with output energy of 750 keV. A three-tank drift-tube linac with quadrupole focusing in the drift tubes follows. An 80 m long transfer line then carries the beam at 50 MeV to the entrance to the PSB.

The PSB is actually four rings stacked on top of each other. Each has one-quarter of the PS circumference, allowing beam to be extracted sequentially from each ring towards the PS, and thus in principle filling the circumference of the PS machine. The extraction energy of the PSB was increased from 1 GeV to 1.4 GeV in order to decrease the space-charge problems at injection in the PS machine.

The PS accelerates the beam to 14 GeV (for SPS physics beams) or to 26 GeV (for LHC beams) before transfer to the SPS, which in turn can accelerate the beam up to an energy of 450 GeV. The circumference ratios of PS/SPS and SPS/LHC are 1/11 and 7/27, respectively.

5 Production of LHC bunch trains

The LHC radiofrequency (RF) system operates on h = 35 640 at a frequency of 400 MHz, which with one in ten buckets used for beam yields 3564 bunch places spaced by 25 ns to completely fill the circumference. In practice, not all of the bunch places can be used in order to leave gaps for the

various kickers used in the injection and extraction processes. The optimum scheme developed [2] results in 2808 bunches in each ring of the LHC for nominal performance.

5.1 From PS extraction to LHC

The 2808 bunches in each LHC ring are provided by multiple injections from the SPS, which in turn is fed by multiple injections from the PS. For each extraction or injection, a fast kicker magnet is used, and the characteristics of these kickers determine the minimum gaps needed in the trains of bunches at various stages of beam production.

Figure 4 summarizes the bunch trains from PS extraction to LHC injection. A PS batch consists of 72 bunches on $h = 84$ at extraction. Either three or four of these batches are sequentially transferred to the SPS, thereby partially filling 3/11 or 4/11 of the SPS circumference. For each LHC ring, 12 of these 216 or 288 bunch trains are transferred from SPS to LHC. With 9×216 + 3×288 injections, the LHC is filled with 2808 bunches.

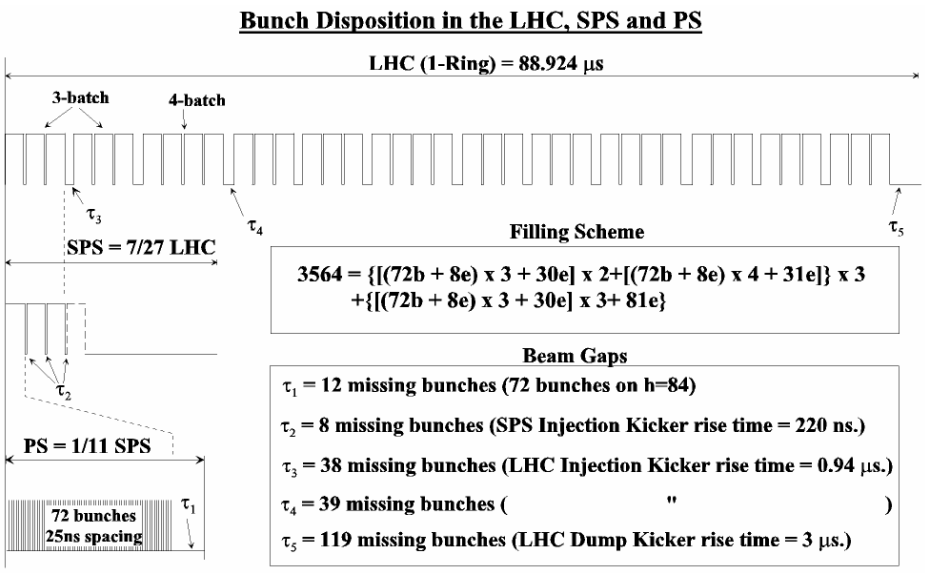


Fig. 4: Bunches in the LHC, SPS and PS machines

The SPS cycle needed to allow up to four injections from the PS, followed by the acceleration from 25 GeV to 450 GeV, followed by extraction to the LHC, has a duration of 21.6 s. Therefore, in an ideal situation, where every transfer works perfectly, the time needed to fill each LHC ring is a little under 5 min. In practice, this takes significantly longer.

5.2 From PSB to PS injection

With the PSB operating on $h = 1$, one bunch per ring is transferred to the PS, operating on $h = 7$. It is possible to fill the 6/7 of the PS ring with six bunches using a double-batch filling scheme, with either four bunches in the first batch and two in the second, or with three bunches per batch as depicted in Fig. 5. Each PSB cycle is 1.2 s long.

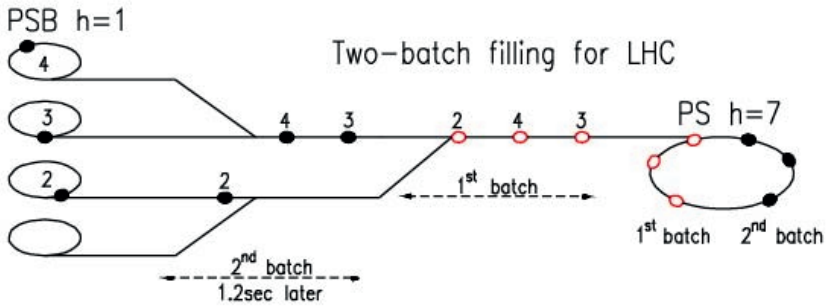


Fig. 5: PS double-batch filling scheme for the LHC

5.3 Bunch splitting in the PS

We have seen that the standard LHC filling scheme is based on a PS batch of 72 bunches being sent to the SPS, while the PSB fills the PS circumference with only six bunches. The six become 72 through a multiple splitting scheme in the PS machine [3].

Bunches from the PSB arrive in the PS and are captured by one of the PS RF systems. This is illustrated in Fig. 6, where the use of a sinusoidal RF waveform provides a stationary bucket in longitudinal phase space, into which the injected bunch arrives (provided that the phase is correct).

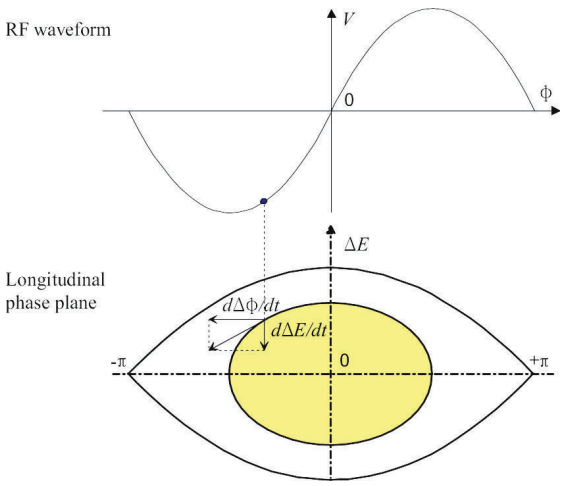


Fig. 6: Capture into a stationary bucket

The principle of using more than one RF waveform is shown in Fig. 7. By the simultaneous application of two RF waveforms, on h and $2h$, and an adiabatic change of the voltages as shown, the single stationary bucket can be transformed into two stationary buckets in the higher-frequency RF system. The bunch that was captured in the lower-frequency system is split into two bunches in the higher-frequency system.

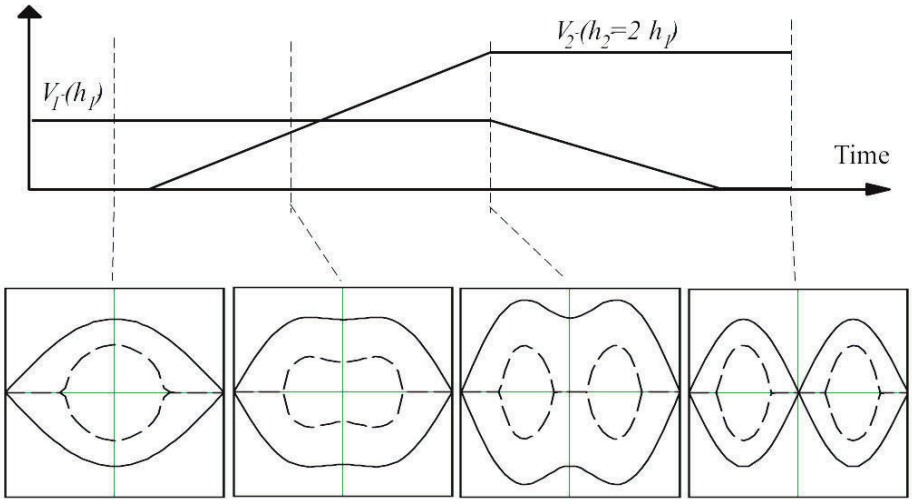
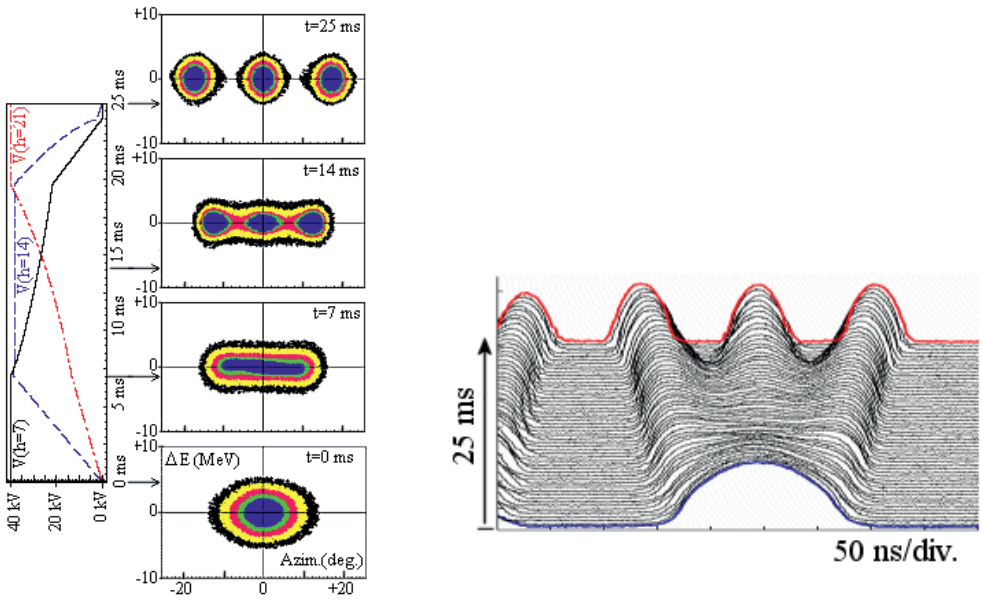


Fig. 7: Bunch splitting principle

Adding a third RF waveform, with appropriate amplitude and phase parameters, allows a bunch to be split into three. This is shown in simulation in Fig. 8(a), with real data shown in Fig. 8(b).



(a) (b)

Fig. 8: Triple bunch splitting in the PS

These techniques were developed and perfected in the 1990s and 2000s in the PS for the production of the LHC beam [4]. The complete process presently used is shown in Fig. 9.

Six PSB bunches are captured on harmonic 7 in the PS. The bunches are then split in three on the 1.4 GeV injection plateau using three groups of RF cavities operating on harmonics 7, 14 and 21. When bunched on h = 21, the beam is accelerated to 25 GeV, where each bunch is split twice (quadruple splitting) by consecutive application of RF systems operating on harmonics 42 and 84, at 20 MHz and 40 MHz, respectively. Finally, an 80 MHz system is used to perform bunch rotation, in

order to shorten each bunch to a length of 4 ns before extraction, in order to fit into the buckets of the SPS 200 MHz RF system.

Note the development of the empty bucket (only six bunches are captured by the $h = 7$ system) to leave a beam gap allowing the clean voltage rise of the PS extraction kicker.

Thus six bunches injected become 72 bunches extracted, with bunch length 4 ns, through the careful manipulation of the voltage waveforms of six RF systems in the PS machine.

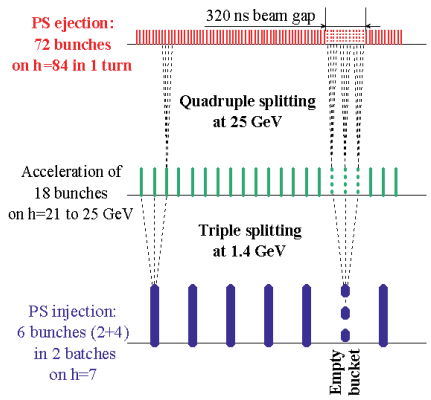


Fig. 9: Bunch evolution in the PS

5.4 Intensities

For nominal LHC operation, the single bunch intensity is 1.15×10^{11} protons. With the scheme described, the implications for the different machines are summarized in Table 2, assuming zero losses throughout the entire chain. This is too idealistic. Each injection process, each acceleration cycle and each extraction process are never loss-free, and with four machines involved the cumulative losses become significant. Taking a combined loss of 10% for each extraction or injection process, and a 10% loss through the acceleration cycle, yields the intensities shown in Table 3. Comparing the booster intensity in the two cases shows approximately a factor of two arising from the cumulative losses.

Table 2: Intensities in the various machines for nominal LHC beams, no losses.

Machine	Bunch intensity	Number of bunches	Total intensity	Scheme
LHC	1.15×10^{11}	2808	3.23×10^{14}	12 SPS cycles, 234 334 334 334
SPS	1.15×10^{11}	288	3.31×10^{13}	Four PS batches of 72
PS	1.15×10^{11}	72	8.28×10^{12}	Six injections, each split ×12
PSB	1.38×10^{12}	1	1.38×10^{12}	Two-turn injection into PSB

Table 3: Intensities in the various machines for nominal LHC beams, 10% losses.

Machine	Bunch intensity	Number of bunches	Total intensity	Scheme
LHC	1.15×10^{11}	2808	3.55×10^{14}	12 SPS cycles, 234 334 334 334
SPS	1.39×10^{11}	288	4.41×10^{13}	Four PS batches of 72
PS	1.68×10^{11}	72	1.33×10^{13}	Six injections, each split ×12
PSB	2.22×10^{12}	1	2.44×10^{12}	Two-turn injection into PSB

6 Requirements on the Linac2 complex

In order to fill each ring of the PSB, Linac2 has to deliver in a single pulse, in the most demanding case, four times the PSB intensity, or around 10^{13} protons. This is the intensity needed at the entrance to the PSB. Losses in the various parts of the Linac2 complex (50 MeV transfer, RFQ, source) push the required intensity from the source to somewhere closer to 2×10^{13} protons.

The PSB revolution time at the proton injection energy of 50 MeV is 1.67 μs, and since each PSB ring is filled using a two-turn low-emittance injection scheme, the pulse length of the beam coming from the linac/source has to be at least 13 μs.

Typical operating conditions therefore required for nominal LHC beams are for a current during the pulse of 165 mA in a 20 μs pulse, which corresponds to around 2×10^{13} protons. In practice, these requirements have been surpassed, enabling the Linac2 complex to provide higher bunch intensities than nominal.

7 Transverse emittance

For nominal performance of the LHC, the normalized emittance ε_n of the colliding bunches should be kept below 3.75 μm rad (Table 1).

Each bunch of protons emanating from the source is:

(i) accelerated in Linac2 and transferred to the PSB;

(ii) injected into (one ring of) the PSB, accelerated, extracted and transferred to the PS;

(iii) injected into the PS, split, accelerated, split, extracted and transferred to the SPS;

(iv) injected into the SPS, accelerated, extracted and transferred to the LHC; and

(v) injected into the LHC, accelerated, squeezed and brought into collision.

Each of these many operations (and more not mentioned) could lead to blow-up of the transverse emittance. In order to meet the LHC requirement, the emittance budget in the LHC Design Report specifies emittances for nominal bunches at extraction from the PBS, PS and SPS machines to be 2.5 μm rad, 3 μm rad and 3.5 μm rad, respectively.

In practice, the performance of the injectors has far exceeded these specifications. In 2011 the average values for normalized emittances in the three machines, corresponding to bunch populations in the SPS of 1.5×10^{11}, were 1.2 μm rad, 1.6 μm rad and 1.8 μm rad [5]. This has allowed colliding beam emittances in the LHC of under 2.5 μm rad, making a considerable contribution to the luminosity delivered. It should be noted that these figures correspond to operation with 50 ns bunch spacing, where one less bunch splitting is needed in the PS. This could contribute to the lower emittance.

8 Summary

The LHC injector chain plays a crucial role in the performance of the LHC. The fabrication of the thousands of bunches needed for high-luminosity performance involves a complex series of manipulations, each of which has to be made without significant detrimental impact on the transverse emittance of the beam being provided to the LHC. This not only has to be realized, but also has to be maintained throughout several months of operation per year.

The fact that LHC delivered luminosity levels that allowed the discovery of the Higgs boson in 2012 is testimony to the ingenuity of the physicists and engineers who have developed the injection scheme, and to the dedicated professionalism of the operations personnel who ensured that high performance was regularly delivered.

References

[1] The LHC Higgs Cross Section Working Group, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections.

[2] R. Bailey, and P. Collier, Standard filling schemes for the various LHC operation modes, LHC-Project Note 323 (2004).

[3] LHC Design Report, vol. III, chapter 7, CERN-2004-003 (2004).

[4] R. Garoby, Proceedings of the LHC Workshop, Chamonix, XI, p. 32, CERN-SL-2001-003 DI (2001).

[5] V. Kain, Proceedings of the LHC Performance Workshop, Chamonix, 2012, p. 81, CERN 2012-006 (2012).

Direction for the Future – Successive Acceleration of Positive and Negative Ions Applied to Space Propulsion

A. Aanesland, J. Bredin, L. Popelier and P. Chabert
Laboratoire de Physique des Plasmas, CNRS – Ecole Polytechnique, France

Abstract
Electrical space thrusters show important advantages for applications in outer space compared to chemical thrusters, as they allow a longer mission lifetime with lower weight and propellant consumption. Mature technologies on the market today accelerate positive ions to generate thrust. The ion beam is neutralized by electrons downstream, and this need for an additional neutralization system has some drawbacks related to stability, lifetime and total weight and power consumption. Many new concepts, to get rid of the neutralizer, have been proposed, and the PEGASES ion–ion thruster is one of them. This new thruster concept aims at accelerating both positive and negative ions to generate thrust, such that additional neutralization is redundant. This chapter gives an overview of the concept of electric propulsion and the state of the development of this new ion–ion thruster.

1 Introduction

In the late 1950s, the golden years of space exploration began. The technology improvements in rocketry after World War II allowed us to meet the challenge of creating thrust that could overcome Earth's gravity. Human exploration to the Moon became a reality; and also the huge technology breakthrough led to increasing numbers of satellites in orbit around the Earth and unmanned missions to our neighbouring planets, comets, asteroids, etc. The early *Pioneer* and *Voyager* missions have now reached interstellar space and still transmit signals back to Earth.

Electrical space thrusters show important advantages for applications in outer space compared to chemical thrusters, as they allow a longer mission lifetime with lower weight and propellant consumption. The first mission that used electric propulsion as a unique propulsion system was launched in 1989 with the *Deep Space I* programme. Since then, this technology has become more and more popular within the space sector, but still this makes up only a few per cent of the launched missions. The mature technologies on the market today rely on accelerating ions either via the Hall current (i.e., Hall thrusters or closed drift thrusters) or via electrostatic grids (i.e., ion thrusters) [1, 2]. Hall thrusters and gridded thrusters have various advantages and drawbacks, and it is mainly political issues that determine which one is used on a mission. The advantages of gridded thrusters (over Hall thrusters) are lower beam divergence, slightly longer lifetime and the fact that they can deliver higher thrust and ion beam velocity. The common point is that positive ions are accelerated from the plasma via electric fields to generate thrust. As the beam is positively charged, electrons are injected into the downstream space to ensure current and charge neutralization. This need for downstream neutralization is a drawback in all existing systems, as it takes up space and adds to the weight of the system. It also increases the complexity of the thrusters and consequently the risk of failure. The *HAYABUSA* mission is one example illustrating this problem [3].

The PEGASES thruster concept has been proposed to improve electric propulsion and to remove the need for additional neutralization systems [4]. PEGASES is an acronym for 'plasma propulsion with electronegative gases'. It belongs to the electrostatic gridded thruster family, but, contrary to classical systems, it accelerates alternately positive and negative ions to provide thrust. In this way additional electron neutralization is redundant. It is also thought that, since the recombination rate of oppositely charged ions is much higher than that for electron–ion recombination, the downstream beam or plume will mainly consist of fast neutrals and a low density of charged particles.

This chapter will outline the various stages of the PEGASES thruster, from plasma generation to the formation of an electronegative plasma and the acceleration of both positive and negative ions to provide thrust.

2 Thrust and specific impulse in space propulsion

The thrust T acting on a spacecraft is given by the change in momentum such that

$$T = -\frac{\mathrm{d}(mv)}{\mathrm{d}t} = -v_{\mathrm{ex}}\frac{\mathrm{d}m}{\mathrm{d}t}, \tag{1}$$

where v_{ex} is the exhaust velocity and $\mathrm{d}m/\mathrm{d}t$ is the rate of change of the total mass ejected. The thrust required in space applications can vary from meganewtons to micronewtons. To escape the Earth's gravitational field, the thrust needs to overcome the weight of the spacecraft. This requires very high thrust, which today can only be reached by chemical rockets that expel a large amount of mass very quickly (i.e., a large $\mathrm{d}m/\mathrm{d}t$). However, when in space, the thrust can be much lower. Sometimes, only micro-thrust is needed to precisely position space vehicles with respect to one another, and at other times the goal is to reach a target with as little propellant as possible but with no strict requirement on the mission duration. With low thrust, the spacecraft will not travel as fast, but will eventually reach the target.

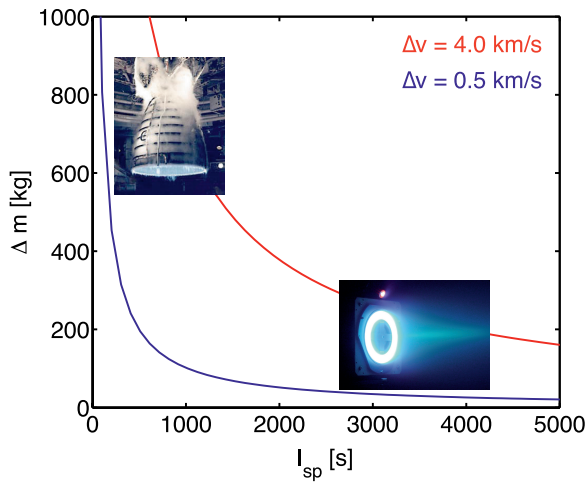


Fig. 1: Propellant mass required as a function of the specific impulse for two different types of missions

Integrating Eq. (1) gives the well-known rocket equation,

$$\Delta v = v_{\mathrm{ex}} \ln\left(\frac{m_0}{m_{\mathrm{f}}}\right), \tag{2}$$

where Δv is the change in velocity of the space vessel, and m_0 and m_{f} are its initial and final mass, respectively. Hence, $\Delta m = m_0 - m_{\mathrm{f}}$ is the required propellant mass. This equation was first derived by Konstantine E. Tsiolkovsky (1857–1935) and is sometimes referred to as the Tsiolkovsky rocket equation. The specific impulse is defined as

$$I_{\mathrm{sp}} = \frac{v_{\mathrm{ex}}}{g_0}, \tag{3}$$

and is given in the unit of seconds (g_0 is the gravitational constant at sea level). Eq. (2) shows that the I_{sp} of a thruster is indirectly a measure of the propellant consumption: the higher I_{sp}, the less propellant is

needed to reach the same change in velocity. Figure 1 is an example given for an interplanetary mission that will need a velocity change Δv of 0.5 km s^{-1}, or for a positioning in low Earth orbit with a velocity change of 4 km s^{-1}. In the first example, to achieve the same manoeuvre one could go from a propellant mass of around 400–800 kg to only 25 kg by choosing electric rather than chemical propulsion. Yet, the duration to achieve the manoeuvre is very different due to the thrust level. To bring 1 kg up to low Earth orbit with an Ariane V rocket costs around 20 000 euros [5]. Hence reducing the propellant consumption translates into real money for the industry. In the example above, a mission to the Moon saves 7–15 million euros.

2.1 Thrust from gridded ion thrusters

For an ion beam, v_{ex} is equivalent to the ion beam velocity v_{b} and $\mathrm{d}m/\mathrm{d}t$ can be expressed as the ion flux Γ_{i} out of the thruster with the given mass M_{i} through an effective grid area for ions A_{i}. Assuming that $v_{\mathrm{b}} = \sqrt{2eV_0/M_{\mathrm{i}}}$ is obtained by accelerating the ions across a voltage difference V_0, we can express the thrust provided by an ion beam as

$$T_{\mathrm{i}} = A_{\mathrm{i}}\Gamma_{\mathrm{i}}M_{\mathrm{i}}v_{\mathrm{b}} = A_{\mathrm{i}}en_{\mathrm{s}}\sqrt{2T_{\mathrm{e}}V_0}. \tag{4}$$

Hence, the thrust depends on the plasma density at the sheath edge n_{s}, the electron temperature T_{e}, the acceleration voltage V_0, and the area of active extraction/acceleration A_{i}. When the thruster is operated at its maximum performance, the current is limited by the space-charge-limited current through the grids, given by the Child–Langmuir relation [6]

$$J_{\mathrm{CL}} = \frac{4}{9}\varepsilon_0\left(\frac{2e}{M_{\mathrm{i}}}\right)^{1/2}\frac{V_0^{3/2}}{d^{*\,2}}, \tag{5}$$

where d^* is the effective space-charge-limited distance between the grids. Assuming that the plasma density and hence the flux from the plasma is balanced with the space-charge-limited current, such that $\Gamma_{\mathrm{i}} = J_{\mathrm{CL}}/e$, the maximum thrust can be expressed as

$$T_{\max} = \frac{8}{9}\varepsilon_0\frac{A_{\mathrm{i}}}{d^{*\,2}}V_0^2. \tag{6}$$

Here we neglect beam divergence, which would reduce the thrust by $\cos^2\theta$, where θ is the beam divergence angle, and we neglect any losses in the grid system. Hence, provided that the plasma is of sufficient density, the thrust depends only on the grid dimensions and the acceleration voltage. Higher thrust is achieved with larger surfaces and smaller grid distance (or, to be precise, a smaller space-charge-limited distance). As a short remark, note that the mass of the ions does not effect the thrust under these conditions. However, the mass plays a significant role in the specific impulse, and hence the propellant consumption.

A global (volume-averaged) model of a gridded ion thruster has recently been developed [7]. The neutral propellant (xenon gas) is injected into the thruster chamber at a fixed rate and a plasma is generated by circulating a radio-frequency (RF) current in an inductive coil. The ions generated in this plasma are accelerated out of the thruster by a pair of d.c. biased grids. The neutralization downstream is not treated. Xenon atoms also flow out of the thruster across the grids. The model, based on particle and energy balance equations, solves for four global variables in the thruster chamber: the plasma density, the electron temperature, the neutral gas (atom) density, and the neutral gas temperature. Figure 2 shows (a) the densities and (b) the thrust as functions of the input RF power, for a gridded system where the grids are separated by 1 mm. Intuitively, the thrust increases with increasing plasma density, as the current or flux from the plasma increases. What is less intuitive is that the maximum thrust, limited by the space charge, is reached for rather low RF power and plasma density. The ionization degree in gridded thrusters needs therefore only to be around 5–10%.

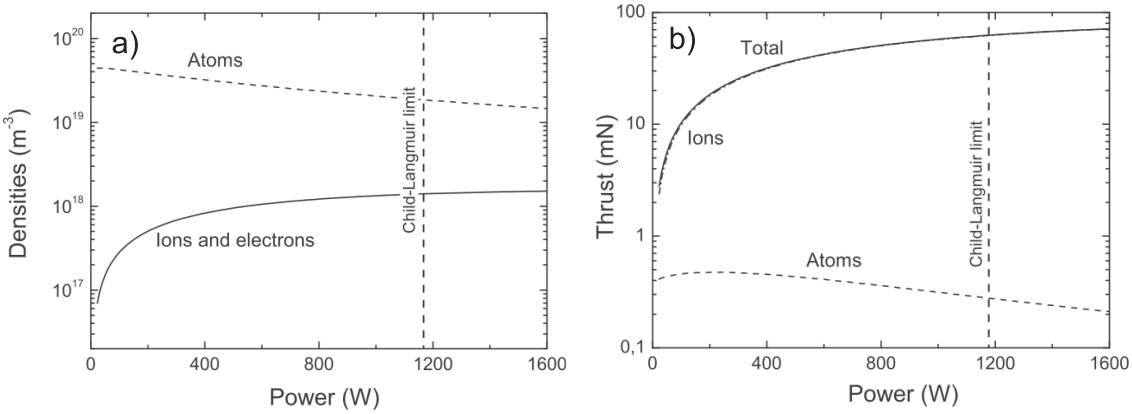


Fig. 2: (a) Atom and ion densities and (b) corresponding thrust as a function of power. The dotted line indicates the Child–Langmuir space-charge limit for the grids used in this calculation ($d = 1$ mm).

3 The physics of the PEGASES thruster

Our team at Laboratoire de Physique des Plasmas (LPP) is developing a new thruster for space propulsion. This thruster is called PEGASES (for 'plasma propulsion with electronegative gases') and belongs to the electric thrusters family. As for classical thrusters, the aim is to generate thrust and high specific impulse by the acceleration of charged particles. The innovative idea behind the PEGASES thruster is to generate thrust by both positive and negative ions, which has many advantages over existing technologies, such as eliminating the additional neutralization system and decreasing the weight and size of the thruster.

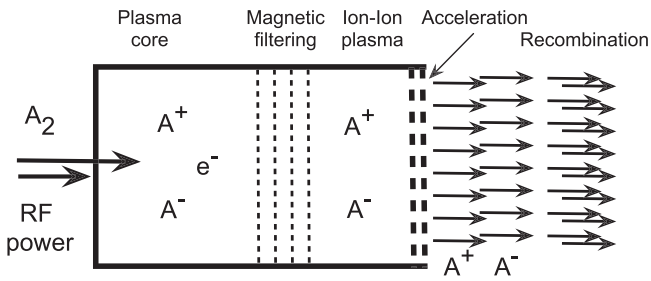


Fig. 3: Illustration of the PEGASES thruster, where positive and negative ions are accelerated alternately from an ion–ion plasma to generate thrust.

The PEGASES concept is illustrated in Fig. 3 and can be described by three almost independent stages:

Stage 1 is the plasma generation stage, where RF power is coupled to the plasma such that electrons are heated and cause ionization. This stage is important for the power efficiency of the thruster.

Stage 2 is the formation of an ion–ion plasma a certain distance from the ionization in Stage 1. For this we use a magnetic filter to confine and cool down the electrons. The hot electrons upstream of the barrier lead to high ionization and production of positive ions, while the cold electrons downstream of the barrier lead to efficient electron attachment and production of negative ions (for this to occur we need to use electronegative gases such as halogen-containing gases, e.g., O_2, SF_6 or I_2). As a result, positive and negative ions exist downstream with negligible contribution of electrons. This downstream region is described as an ion–ion plasma.

Stage 3 is the acceleration stage, where an alternate acceleration of positive and negative ions (from the

ion–ion region in Stage 2) will provide the thrust for the spacecraft. The acceleration in PEGASES is based on classical gridded thrusters, where the acceleration is obtained by creating an electric field between two or more grids. In PEGASES the acceleration field is changed in time by applying a square waveform to the plasma grid; this allows consecutive bursts of positive and negative ion beams.

3.1 Stage 1: plasma generation

RF plasmas are very often used in the semiconductor industry for etching and deposition, and a detailed description of the physics of these plasma discharges can be found elsewhere [8]. Briefly, the electromagnetic field from the RF power supply will couple to the electrons in the plasma and transfer energy to them. As illustrated in Fig. 4, there are a variety of methods to couple this energy to the electrons. For example, in Fig. 4(a) the RF voltage (from an electrode or a coil) couples capacitively to the electrons within an oscillating sheath. In Fig. 4(b) the RF current flowing in a coil induces an RF current in the plasma. The induced electric field transfers the energy to the electrons and consequently the field decays over a distance of a few centimetres in the plasma, called the skin depth. In Fig. 4(c) an additional static magnetic field allows an electromagnetic wave to propagate in the plasma and in this case transfers the energy to the electrons in the volume.

The plasma source in the PEGASES thruster is a purely inductively coupled plasma (ICP) without capacitive coupling, symmetrically driven at 4 MHz. The planar inductor is separated from the plasma by a thin (3 mm) ceramic window and encapsulated in a ferrite to reduce losses and to enhance the ICP coupling to the plasma [9, 10]. The RF power is fed to the inductor via an impedance matching network using a low-loss transmission-line transformer and air variable capacitors in symmetrical (push–pull) configuration. The symmetrical drive of the ICP inductor practically eliminates capacitive coupling to the plasma, resulting in negligible plasma RF potential. The power efficiency in such systems can reach up to 90% or more [9] and is very important in space applications where every gram and every unit of electric power counts.

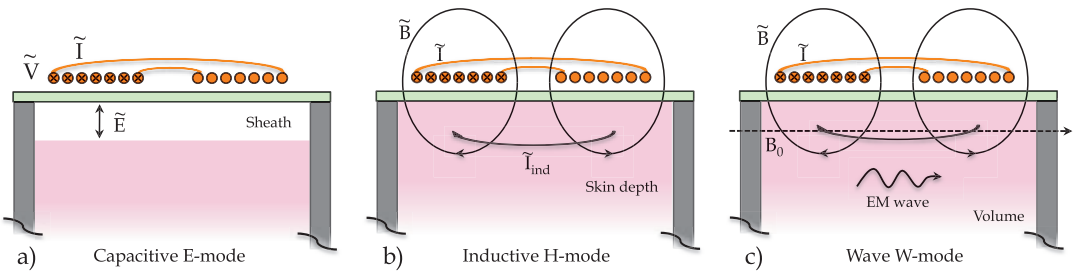


Fig. 4: Illustration of RF power discharges operated in the (a) capacitive, (b) inductive and (c) wave modes

3.1.1 Ionization and attachment in electronegative plasmas

The plasma in the PEGASES thruster contains both positively and negatively charged ions. To create negative ions, the gas has to be electronegative, allowing electrons to attach to neutral species. This process can occur in the volume or on surfaces with low work function. In the negative ion sources for fusion, negative hydrogen ions are for example produced on a caesium surface [11, 12].

In most other experiments with negative ions, negative ion production occurs in volume by what we call dissociative attachment collisions. Electrons are typically too energetic to attach directly to a neutral, so in the process either the excess energy leads to the neutral molecule breaking up and/or it dissipates into vibrational and rotational energies. Electronegative gases are therefore molecular gases and typically formed by halogens (group 17 in the periodic table) such as Cl_2 and I_2 or halogen-containing molecules such as SF_6 and CF_4. Although not halogens, O_2 and H_2 are also electronegative gases.

The ionization and attachment rate coefficients are typically increasing and decreasing functions of the electron temperature, where the typical reactions are given in the form [13]

$$\mathrm{AB}_x + \mathrm{e} \rightarrow \mathrm{AB}_x^+ + 2\mathrm{e} \tag{7}$$

$$\mathrm{AB}_x + \mathrm{e} \rightarrow \mathrm{AB}_y^- + \mathrm{B}_z \tag{8}$$

for ionization and dissociative attachment. Hence, for most electronegative gases, ionization dominates for high electron temperatures whereas attachment occurs for low temperatures.

3.2 Stage 2: ion–ion plasma formation

The aim for this stage is to segregate the plasma into two regions: (i) the plasma core, where electrons have a rather high temperature for efficient ionization; and (ii) a downstream ion–ion plasma, where the electron density can be neglected and the plasma dynamics is controlled by ions. As seen above, to acquire this segregation, we need to control the electron temperature in the plasma. Commonly, in low-temperature plasmas, the electron temperature T_e is governed by the ionization balance (electron creation and loss processes) and is a function of the kind of gas and the product pL, where p is the gas pressure and L is the characteristic size of the plasma [14]. The specific mechanism of electron heating seen in Fig. 4 and the value of discharge power have a minor influence on the electron temperature. In negative ion sources, the electron cooling is usually achieved with magnetic filters (a localized transverse magnetic field) placed in front of the ion extracting aperture. This technique is also used in the PEGASES thruster, where a localized magnetic barrier is generated by a set of permanent neodymium magnets forming a localized Gaussian magnetic field perpendicular to the extraction axis. The illustration of the PEGASES prototype with the magnetic field lines is shown in Fig. 5(a). We have recently shown that the gradient and strength of the magnetic field strongly affect the electron cooling, and the position of the minimum electron temperature is achieved at the maximum magnetic field strength [10]. We have also shown that the ion–ion plasma is also generated around the maximum field region [15]. As an example, the electron temperature measured in argon, and the measured densities in SF_6 are shown in Fig. 5. Close to the maximum magnetic field, the electron density is three orders of magnitude lower than the ion densities. The ion densities in this region remain high, showing that ion transport across the magnetic field is not affected by the magnetic field. For an RF power of only 120 W, the ion density reaches 3×10^{17} m^{-3} in the ion–ion region.

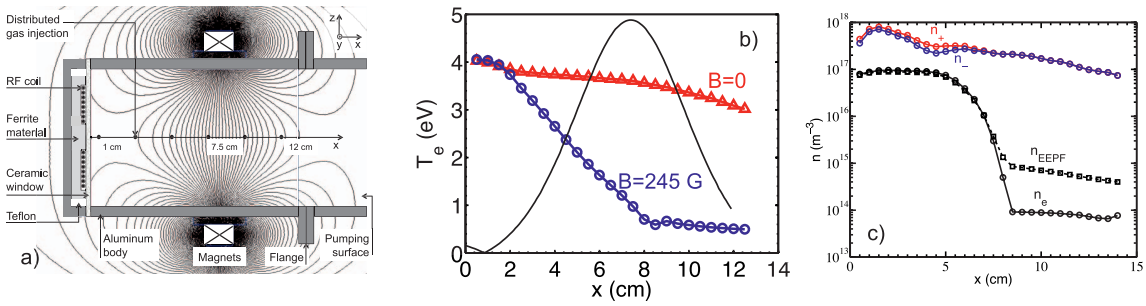


Fig. 5: Illustration of RF power discharges operated in the (a) capacitive, (b) inductive and (c) wave modes

3.3 Stage 3: alternate acceleration of ions to generate thrust

The PEGASES thruster is a gridded thruster based on the same principles as classical gridded acceleration systems. However, the originality of this thruster is that positive and negative ions are alternately accelerated from an ion–ion plasma using the same grids. To achieve this, the first grid in contact with

the plasma is biased with alternate voltage waveforms and the second grid is grounded [16]. The walls in the thruster are floating such that the potential of the plasma follows the plasma grid potential, and hence the electric field between the two grids will change direction during one bias period. A simplified illustration of this concept is shown in Fig. 6, where (a) shows one hole in the grid system with idealized ion trajectories, and (b) shows the potential distribution in the plasma and across the grids for the positive and negative bias period applied to the first grid. Note that possible variations of the plasma potential within the bulk plasma are not shown here.

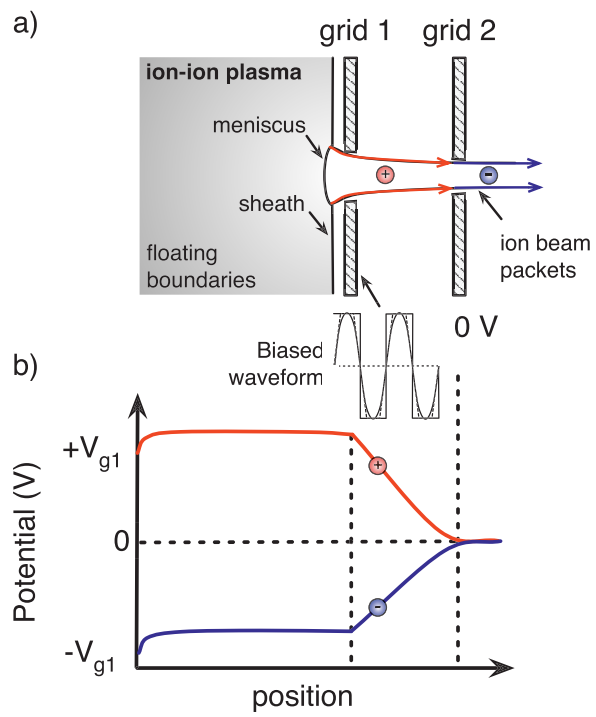


Fig. 6: Illustration of RF power discharges operated in the (a) capacitive, (b) inductive and (c) wave modes

The acceleration grids in any ion source are designed and adapted to the acceleration voltage and the available plasma parameters (density and temperature), to achieve the desired beam properties with respect to thrust, beam current, divergence, etc. Square waveforms are therefore best suited. Particle-in-cell simulations have shown that the ratio of extracted positive and negative ion fluxes from the plasma depends strongly on the electronegativity, particularly when the negative ion density $n_{\mathrm{i}-}$ is much higher than the electron density n_{e} [17]. When negative ions are extracted from the plasma, we should therefore expect co-extracted electrons. In this case, the period over which the negative charges are accelerated should be shorter than the corresponding period for positive ions in order to compensate both the charge density and the current (i.e., the duty cycle can be optimized). This asymmetry could even be pushed towards a system of accelerating positive ions and electrons. For an optimized system, the acceleration voltages for the positive and negative charges should also be adapted independently to compensate for the difference in their effective temperatures. The most suitable waveform is therefore a square waveform with adjustable duty cycle and voltage offset.

The upper and lower frequency requirements for the applied voltage should take into account: (i) both the ion plasma frequency (so the ions can react to the variations) and the transit time of the ions through the grids; and (ii) the potential barrier created in the downstream space from the space charge of the single beam packet or envelope, respectively. Analytical models developed in one dimension, which completely neglect any secondary electron emission or other external sources of space-charge neutralization, predict an operating frequency in the lower megahertz region [18].

3.4 First laboratory tests in PEGASES II

Successive positive and negative ion beams have been measured downstream of the alternately biased grids in the PEGASES II thruster. The grids are placed in the ion–ion plasma region and the bias frequency is only 1 kHz biased. Figure 7(a) shows a typical result of time-resolved ion energy distribution functions (IEDFs) measured downstream of the acceleration stage. The red and blue curves are obtained during the positive and negative bias periods, respectively. Both the positive and negative ions are measured with a constant amplitude and energy during their respective bias period. However, their mean energies are not the same. Figure 7(b) shows the IEDFs measured for various positions downstream of the grounded grid. These measurements are continuous while the grid is alternately biased. The peak ion energies remain constant downstream of the grid, but interestingly the positive peak amplitude decreases while the negative ion peak increases as a function of position.

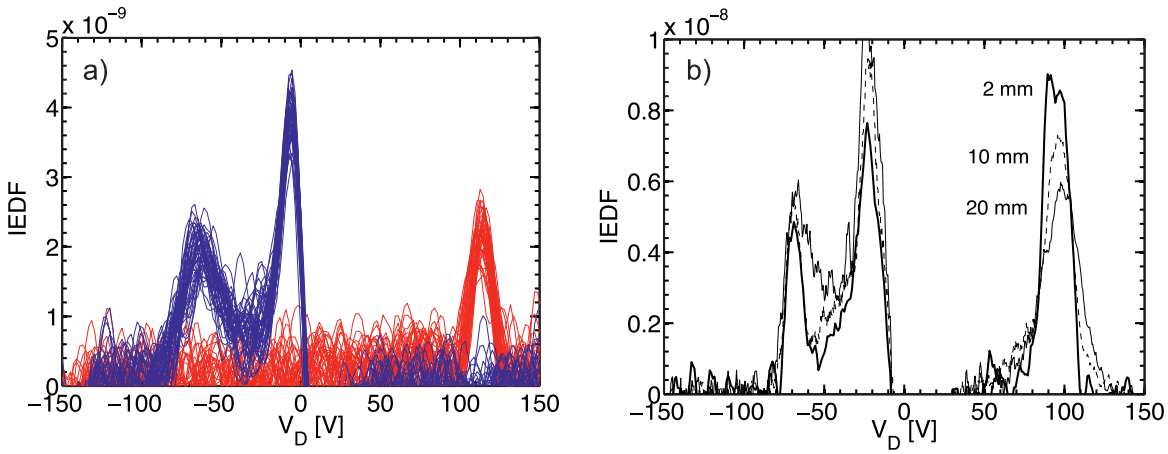


Fig. 7: (a) IEDFs measured during the bias period, where the red and blue curves are measured during the positive and negative bias period, respectively. (b) IEDF measured continuously for various positions downstream of the grounded grid. The acceleration voltage is here $V_0 = \pm 100$ V at a bias frequency of 1 kHz.

Figure 8(a) shows the ion mean energies as a function of the applied acceleration voltage for continuous and alternate acceleration operated in a mixture of argon and SF_6 (about 20% SF_6). The positive ions have energies equivalent to the acceleration voltage V_0. Some variations are obtained between each dataset, possibly due to slight differences in the ion–ion plasma conditions and therefore a change in the sheath voltage in front of the grids. The negative ions are, on the contrary, measured with lower energies than expected from the applied acceleration voltage. Similar results are obtained with planar probes. Figure 7(b) shows that the beam energy does not depend on the downstream position, indicating that the bipolar beam is space-charge-neutralized downstream.

The difference in the positive and negative ion energies can be understood as follows. In plasma immersed ion implantation (PIII) sudden negative bias pulses are applied to a surface such that positive ions are implanted in the material [19]. Several authors have seen that implanting into a dielectric substrate results in a significant voltage build-up in the wafer, reducing the effective implant energy [20,21]. In our case, a film composed of fluorine and sulfur can deposit on the acceleration grids. A dielectric film is observed by eye on the grids after some hours of operation in mixed Ar/SF_6 and after some tens of minutes in pure SF_6. This film is also detectable by measuring a decrease in the current on the grids. In the presence of a dielectric layer on the plasma grid, a certain voltage will be dropped in the dielectric such that the effective acceleration voltage is [21]

$$V_{\text{eff}} = V_0 - \frac{e\delta Q}{\varepsilon_0 \varepsilon_{\text{r}}}, \tag{9}$$

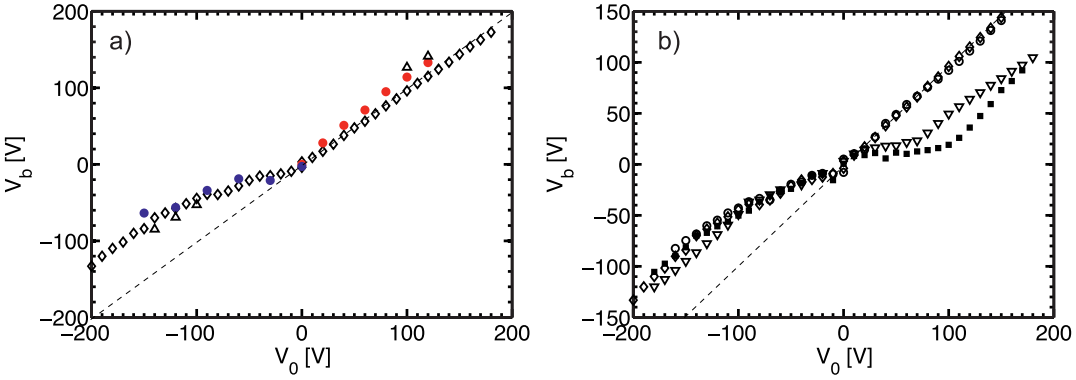


Fig. 8: Ion mean energy as a function of the acceleration voltage. (a) Circles and diamonds are obtained for alternate and continuous acceleration, respectively, and measured with the Retarding Field Energy Analyser (RFEA). Triangles are measured with a planar probe in continuous mode. (b) Increasing deposition of a fluorine–sulfur film on the acceleration grid. Diamonds are the initial conditions, triangles after about 5 min and diamonds after 30 min of operation, and circles after cleaning the grids.

where δ is the thickness of the dielectric film and Q is the surface charge. For a quick estimate using $\delta \sim 1$ µm, $Q \sim 1 \times 10^{16}$ and $\varepsilon_r = 3.7$ as for pure sulfur, the voltage drop in the dielectric is about 50 V, which is of the order of the measured energy loss for negative ions. Figure 8(b) shows the beam energies as a function of V_0 in pure SF_6 after about 5 min and 30 min of operation, i.e., the deposit on the grids became thicker and thicker. The negative ions are not affected much with increasing film thickness. However, the positive ion energies are similar to the negative ion energies after some time of operation. It seems likely that it is the dielectric film that is responsible for the lower beam energies. However, it is still uncertain why the negative ions are affected more than the positive ions in the initial conditions. This might be due to a change in the dielectric layer formed as a function of the bias voltage or a change in how this dielectric is charged when applying a positive or negative bias. Further experiments are needed before any definite conclusion can be drawn.

4 Conclusion

The PEGASES thruster is a new and promising gridded ion thruster where, in contrast to classical gridded thrusters, both positive and negative ions are accelerated alternately from the same source. This concept makes the additional neutralization system redundant and provides therefore many advantages over existing thrusters. The summary given here provides a brief overview of the state of the art in the development of this thruster. Although there is still a long way to go before this thruster is flying in space, we show that an ion–ion plasma is efficiently produced downstream of a localized magnetic filter, such that the degree of ionization required in a gridded thruster is met. The ion–ion plasma source can therefore provide high enough densities for gridded ion thrusters operating with an acceleration of 500 V or so, and can therefore provide similar thrust and I_{sp} as classical gridded thrusters. We have also shown the first laboratory evidence that alternate acceleration of positive and negative ions can be achieved.

5 Acknowledgements

This work was funded by EADS Astrium and by Agence Nationale de la Recherche (ANR) under contract ANR-11-BS09-040. The authors would like to thank V. Godyak, J.-P. Booth and G. Hagelaar for useful discussions and collaboration.

References

[1] H.R. Kaufman, Technology of closed-drift thrusters. *AIAA J.* **23** (1985) 78–87.

[2] D.M. Goebel and I. Katz, *Fundamentals of Electric Propulsion: Ion and Hall Thrusters* (JPL Space Science and Technology Series, Wiley, Hoboken, NJ, 2008).

[3] N. Kazutaka and H. Kuninaka, *Trans. Jpn. Soc. Aeronaut. Space Sci., Aerosp. Technol. Jpn.* **10** (2012) Tb-1.

[4] A. Aanesland, A. Meige, and P. Chabert, Electric propulsion using ion–ion plasmas. *J. Phys.: Conf. Ser.* **162** (2009) 012009.

[5] R. Janovsky and D.E. Koelle, Development and transportation costs of space launch systems. Presented at the DGLR/CEAS European Air and Space Conference (2007).

[6] H.R. Kaufman, Technology and applications of broad-beam ion sources used in sputtering. Part I. Ion source technology. *J. Vac. Sci. Technol.* **21** (1982) 725.

[7] P. Chabert, J. Arancibia Monreal, J. Bredin, L. Popelier, and A. Aanesland, Global model of a gridded-ion thruster powered by a radiofrequency inductive coil. *Phys. Plasmas* **19**(7) (2012) 073512.

[8] P. Chabert and N.S.J. Braithwaite, *Physics of Radio-Frequency Plasmas* (Cambridge University Press, Cambridge, 2011).

[9] V.A. Godyak, Electrical and plasma parameters of ICP with high coupling efficiency. *Plasma Sources Sci. Technol.* **20** (2011) 025004.

[10] A. Aanesland, J. Bredin, P. Chabert, and V.A. Godyak, Electron energy distribution function and plasma parameters across magnetic filters. *Appl. Phys. Lett.* **100** (2012) 044102.

[11] Yu.I. Belchenko, G.I. Dimov, and V.G. Dudnikov, A powerful injector of neutrals with a surface-plasma source of negative ions. *Nucl. Fusion* **14**(1) (1974) 113.

[12] M. Bacal, Physics aspects of negative ion sources. *Nucl. Fusion* **46** (2006) S250–S259.

[13] D. Rapp and D.D. Briglia, Total cross sections for ionization and attachment in gases by electron impact. II. Negative-ion formation. *J. Chem. Phys.* **43** (1965) 1480.

[14] Lieberman, M. A., and Lichtenberg, A. J., Principles of Plasma Discharges and Materials Processing. (2nd ed.). Wiley-Interscience (2005).

[15] J. Bredin, P. Chabert, and A. Aanesland, Langmuir probe analysis of highly electronegative plasmas. *Appl. Phys. Lett.* **102** (2013) 154107.

[16] A. Aanesland, P. Chabert, M. Irzyk, and S. Mazouffre, Method and device for forming a plasma beam. Patent WO 2012/042143 A1 (2012).

[17] N. Oudini, J.-L. Raimbault, P. Chabert, A. Meige, and A. Aanesland, Particle-in-cell simulation of an electronegative plasma under direct current bias studied in a large range of electronegativity. *Phys. Plasmas* **20** (2013) 043501.

[18] A. Aanesland, Low temperature plasmas exposed to external magnetic and electric fields. Habilitation à Diriger des Recherches, Université Pierre et Marie Curie, 2013.

[19] M.A. Lieberman, Model of plasma immersion ion implantation. *J. Appl. Phys.* **66** (1989) 2926.

[20] N. Cheung, Plasma immersion ion implantation for semiconductor processing. *Mater. Chem. Phys.* **46** (1996) 132–139.

[21] R.K.Y. Fu, Influence of thickness and dielectric properties on implantation efficacy in plasma immersion ion implantation of insulators. *J. Appl. Phys.* **95**(7) (2004) 3319.

Participants

ARAZ, A.	Siemens AG, Erlangen, DE
AUGUSTIN, M.	CERN, Geneva, CH
BABCOCK, C.	CERN, Geneva, CH
BARBISAN, M.	Consorzio RFX, Padova, IT
BRUNNER, B.	Slovak University of Technology, Bratislava, SK
CANET, C.	GANIL, Caen cedex 5, FR
CARRASCO DOMINGUEZ, R.	University of Huelva, Seville, ES
CASTRO, G.	INFN, Catania, IT
CETINKAYA, H.	Dumlupinar University, Kutahya, TR
COLCIAGO, M.	CERN, Geneva, CH
DI BARTOLO, F.	INFN, Catania, IT
DOGAR, A.	Pakistan Institute of Nuclear Science & Technology, Islamabad, PK
DONZEL, X.	Pantechnik, Bayeux, FR
DUBOIS, M.	GANIL, Caen cedex 5, FR
FINK, D.	CERN, Geneva, CH
GALATA, A.	INFN, Legnaro, IT
GAMBINO, N.	ETH-Zurich, Zurich, CH
GARBAYO, A.	AVS - Added Value Solutions, Chilton, Didcot, UK
GIL FLORES, J.	CERN, Geneva, CH
GILES, T.	CERN, Geneva, CH
GIOVANNINI, A.	ETH-Zurich, Zurich, CH
GRISHNYAEV, E.	BINP, Novosibirsk, RU
HENROTIN, S.	IBA, Louvain-la-Neuve, BE
HODGKINSON, A.	LBNL, Berkeley, US
IZAOLA, Z.	ESS-Bilbao, Leioa, ES
JENSEN, H.	PET and Cyclotron Unit, Copenhagen, DK
KAISER, M.	Slovak University of Technology, Bratislava, SK
KANG, B.	Institute for Basic Science, Daejeon, KR
KARPUS, S.	NSC KIPT, Kharkov, UA
KASNAR, F.	Slovak University of Technology, Bratislava, SK
KOIVISTO, T.	University of Jyvaskyla, Jyvaskyla, FI
KOMPPULA, J.	University of Jyvaskyla, Jyvaskyla, FI
KRONBERGER, M.	CERN, Geneva, CH
KUCHLER, D.	CERN, Geneva, CH
LIN, S.	Institute of Modern Physics, Lanzhou, CN
LU, W.	Institute of Modern Physics, Lanzhou, CN
MANZOLARO, M.	INFN - LNL, Legnaro, IT
MATLOCHA, T.	Czech Academy of Science, Rez, CZ
MATTEI, S.	CERN, Geneva, CH
MIDTTUN, O.	CERN, Geneva, CH
MIRACOLI, R.	ESS-Bilbao, Zamudio, ES
MONTANO, J.	CERN, Geneva, CH
NOCENTINI, R.	Max-Planck-Institut, Garching, DE
PEAUCELLE, C.	IPNL, Villeurbanne cedex, FR
PEREIRA, H.	CERN, Geneva, CH
PFISTER, J.	GSI, Darmstadt, DE
PISELLI, E.	CERN, Geneva, CH
POPESCU, S.	CERN, Geneva, CH
POTTIN, B.	CEA-SACLAY, Gif-sur-Yvette, FR

RAMSDORF, A.	JINR, Dubna, RU
RECCHIA, M.	Consorzio RFX, Padova, IT
ROEHRICH, J.	Helmholtz-Zentrum Berlin, Berlin, DE
ROTHE, S.	CERN, Geneva, CH
RUF, B.	Max-Planck-Institut, Garching, DE
RUIZ POMARES, M.	University of Huelva, Seville, ES
SCARPA, D.	INFN-LNL, Legnaro, IT
SEIFFERT, C.	CERN, Geneva, CH
SHA, S.	Institute of Modern Physics, Lanzhou, CN
SIESLING, E.	CERN, Geneva, CH
STAFFORD-HAWORTH, J.	CERN, Geneva, CH
STEYAERT, D.	CERN, Geneva, CH
TKACHEV, A.	BINP, Novosibirsk, RU
TOIVANEN, V.	University of Jyvaskyla, Jyvaskyla, FI
VALERIO LIZARRAGA, C.A.	CERN, Geneva, CH
WALLNER, J.	EBG MedAustron, Wiener Neustadt, AT
WIMMER, C.	Max-Planck-Institut, Garching, DE
WU, Q.	Institute of Modern Physics, Lanzhou, CN
ZAMENGO, A.	Consorzio RFX, Padova, IT
ZHANG, Z.	Institute of Modern Physics, Lanzhou, CN