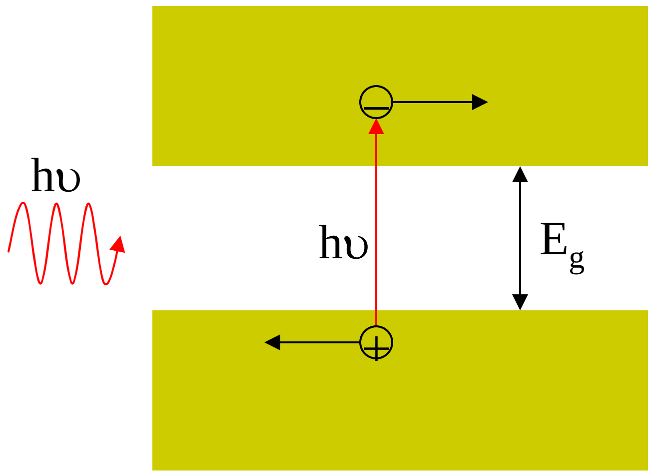
Lecture 12: Photodiode detectors

- Background concepts
- p-n photodiodes
- Photoconductive/photovoltaic modes
- p-i-n photodiodes
- Responsivity and bandwidth
- Noise in photodetectors

References: This lecture partially follows the materials from Photonic Devices, Jia-Ming Liu, Chapter 14. Also from Fundamentals of Photonics, 2nd ed., Saleh & Teich, Chapters 18.

Electron-hole photogeneration

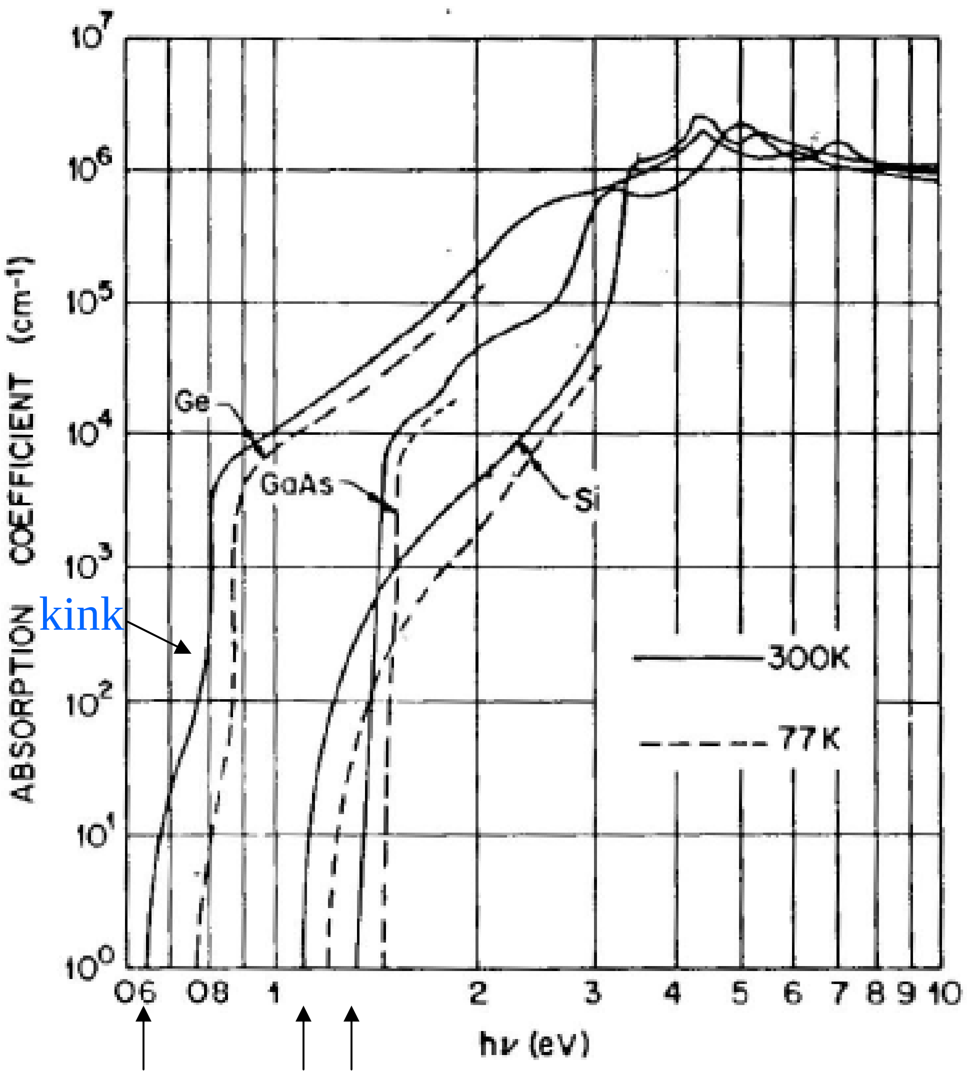
- Most modern photodetectors operate on the basis of the *internal photoelectric effect* – the photoexcited electrons and holes remain within the material, increasing the electrical conductivity of the material
- *Electron-hole photogeneration* in a semiconductor



- absorbed photons *generate* free electron-hole pairs

- *Transport* of the free electrons and holes upon an electric field results in a *current*

Absorption coefficient



Bandgaps for some *semiconductor photodiode materials* at 300 K

	Bandgap (eV) at 300 K	
	Indirect	Direct
Si	1.14	4.10
Ge	0.67	0.81
GaAs	-	1.43
InAs	-	0.35
InP	-	1.35
GaSb	-	0.73
$In_{0.53}Ga_{0.47}As$	-	0.75
$In_{0.14}Ga_{0.86}As$	-	1.15
$GaAs_{0.88}Sb_{0.12}$	-	1.15

Absorption coefficient

- E.g. absorption coefficient $\alpha = 10^3\ cm^{-1}$
- Means an 1/e optical power absorption length of

$$1/\alpha = 10^{-3}\ cm = 10\ \mu m$$

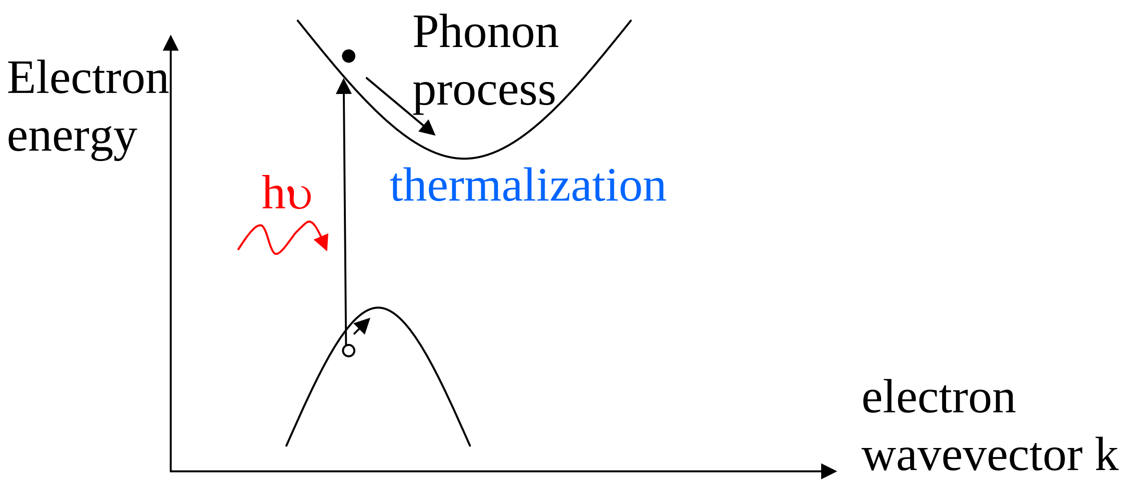
- Likewise, $\alpha = 10^4\ cm^{-1}$ => 1/e optical power absorption length of 1 μm.

$\alpha = 10^5\ cm^{-1}$ => 1/e optical power absorption length of 100 nm.

$\alpha = 10^6\ cm^{-1}$ => 1/e optical power absorption length of 10 nm.

Indirect absorption

- *Silicon* and *germanium* absorb light by both <u>indirect</u> and <u>direct</u> optical transitions.
- ***Indirect*** absorption requires the assistance of a <u>*phonon*</u> so that momentum and energy are conserved.
- *Unlike the emission process, the absorption process can be sequential, with the excited electron-hole pair thermalize within their respective energy bands by releasing energy/momentum via phonons.*
- This makes the *indirect absorption <u>less efficient</u> than direct absorption* where no phonon is involved.



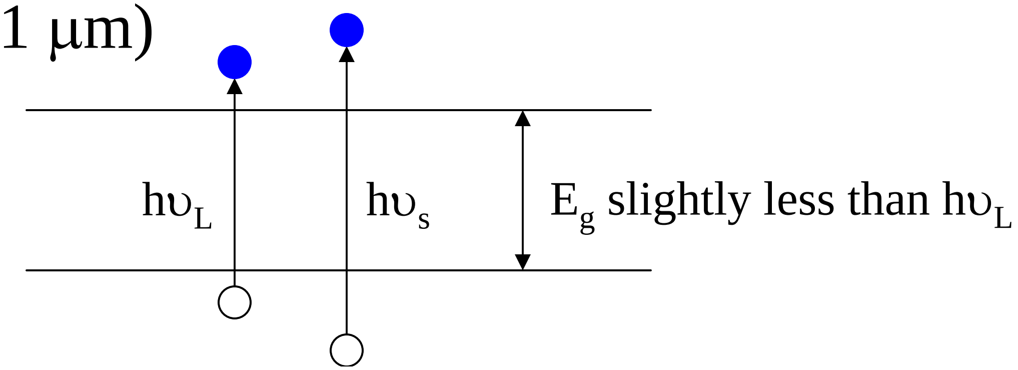
Indirect vs. direct absorption in silicon and germanium

- ***Silicon*** is *only weakly* absorbing over the wavelength band 0.8 – 0.9 μm. This is because transitions over this wavelength band in silicon are due only to the indirect absorption mechanism. The *threshold* for indirect absorption (*long wavelength cutoff*) occurs at 1.09 μm.
- The bandgap for direct absorption in silicon is 4.10 eV, corresponding to a threshold of 0.3 μm.

- ***Germanium*** is another semiconductor material for which the lowest energy absorption takes place by indirect optical transitions. Indirect absorption will occur up to a threshold of 1.85 μm.
- However, the *threshold for direct absorption* occurs at 1.53 μm, for shorter wavelengths germanium becomes strongly absorbing (*see the kink in the absorption coefficient curve*).

Choice of photodiode materials

- A photodiode material should be chosen with a *bandgap energy slightly less than the photon energy* corresponding to the *longest* operating wavelength of the system.
- This gives a *sufficiently high absorption coefficient* to ensure a good response, and yet limits the number of *thermally generated* carriers in order to attain a low "*dark current*" (i.e. current generated with no incident light).
- *Germanium photodiodes have relatively large dark currents* due to their *narrow bandgaps* in comparison to other semiconductor materials. This is a major shortcoming with the use of germanium photodiodes, *especially at shorter wavelengths* (below 1.1 μm)



III-V compound semiconductors

- Direct-bandgap III-V compound semiconductors can be better material choices than germanium for the longer wavelength region.

- Their *bandgaps can be tailored* to the desired wavelength by changing the relative concentrations of their constituents (*resulting in lower dark currents*).

- They may also be fabricated in *heterojunction* structures (which *enhances their high-speed operations*).

e.g. $In_{0.53}Ga_{0.47}As$ lattice matched to InP substrates responds to wavelengths up to around 1.7 μm. (*most important for 1.3 and 1.55 μm*)

Junction photodiodes

- The *semiconductor photodiode detector* is a p-n junction structure that is based on the internal photoeffect.

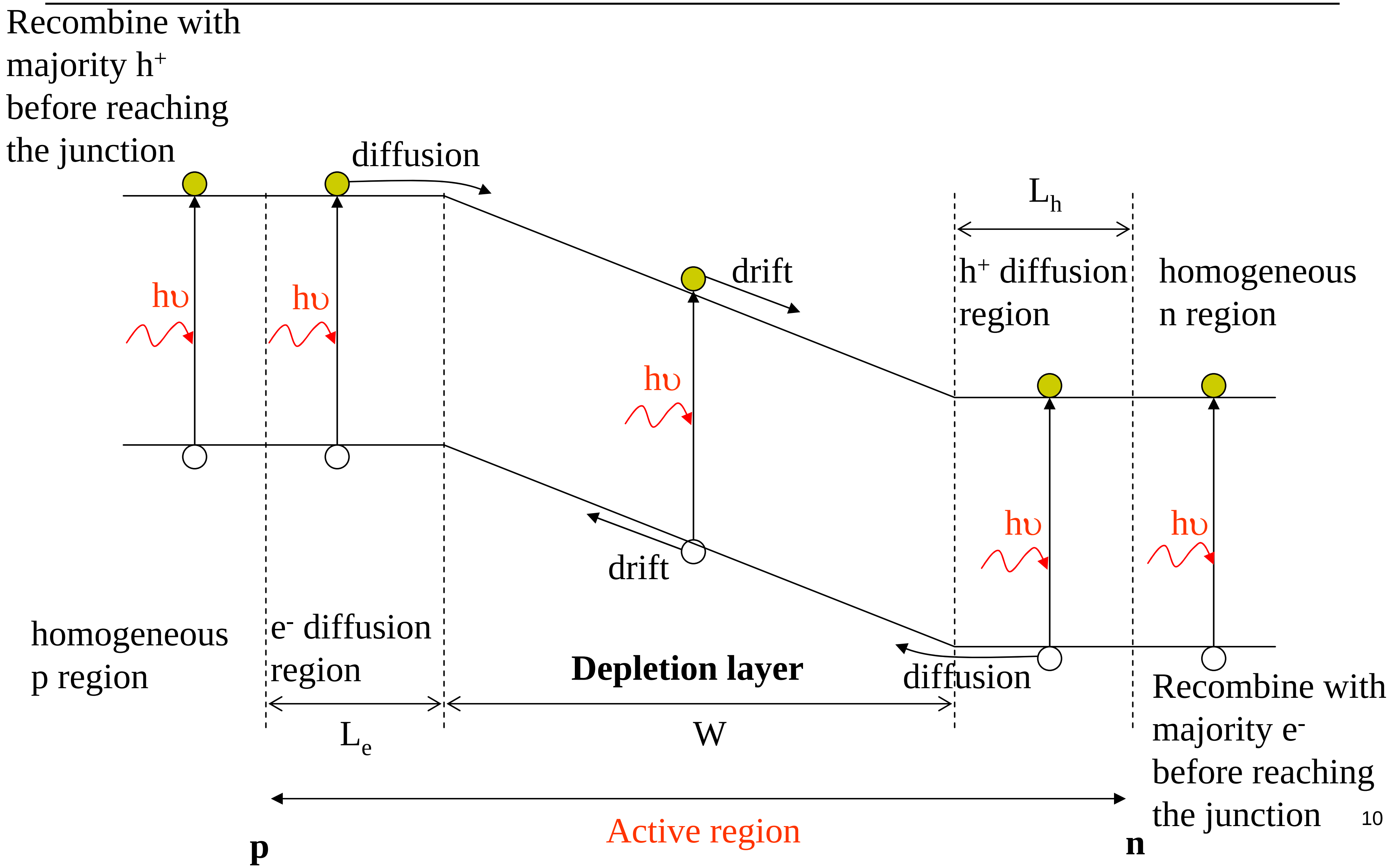
- The photoresponse of a photodiode results from the *photogeneration of electron-hole pairs through band-to-band optical absorption*.

  => The *threshold* photon energy of a semiconductor photodiode is the bandgap energy E_g of its active region.

- The *photogenerated* electrons and holes in the *depletion layer* are subject to the local electric field within that layer. The electron/hole carriers *drift* in opposite directions. This *transport* process induces an electric current in the external circuit.

- Here, we will focus on semiconductor *homojunctions*.

Photoexcitation and energy-band diagram of a p-n photodiode



- ***In the depletion layer***, *the internal electric field sweeps the photogenerated electron to the n side and the photogenerated hole to the p side*.

  => a *drift current* that flows in the *reverse* direction from the n side (cathode) to the p side (anode).

- Within one of the ***diffusion regions*** at the edges of the depletion layer, the photogenerated *minority* carrier (*hole in the n side and electron in the p side*) can reach the depletion layer by *diffusion* and then be *swept to the other side by the internal field*.

  => a *diffusion current* that also flows in the *reverse* direction.

- In the p or n ***homogeneous region***, *essentially no current is generated* because there is essentially no internal field to separate the charges and a minority carrier generated in a homogeneous region *cannot diffuse to the depletion layer before recombining with a majority carrier*.

Photocurrent in an illuminated junction

- If a junction of *cross-sectional area A* is uniformly illuminated by photons with $h\upsilon > E_g$, a *photogeneration rate* G (EHP/cm^3-s) gives rise to a photocurrent.
- *The number of holes* created per second within a diffusion length L_h of the depletion region on the n side is AL_hG.
- *The number of electrons* created per second within a diffusion length L_e of the depletion region on the p side is AL_eG.
- Similarly, AWG carriers are generated *within the depletion region* of width W.
- The resulting *junction photocurrent from n to p*:

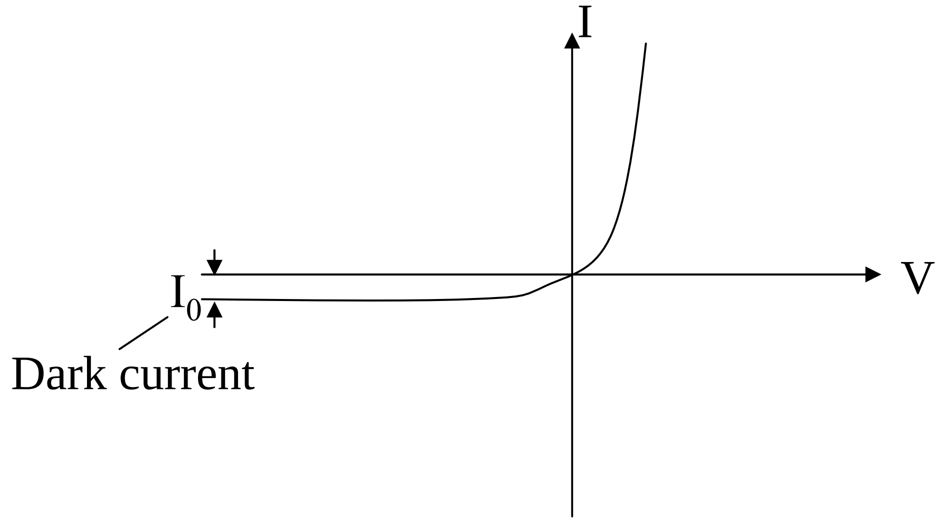
$$I_p = eA\,(L_h + L_e + W)\,G$$

Diode equation

- Recall the current-voltage (I-V) characteristic of the junction is given by the diode equation:

$$I = I_0(\exp(eV/k_BT) - 1)$$

- The current I is the injection current under a *forward* bias V.
- I_0 is the "saturation current" representing *thermal-generated* free carriers which flow through the junction (*dark current*).

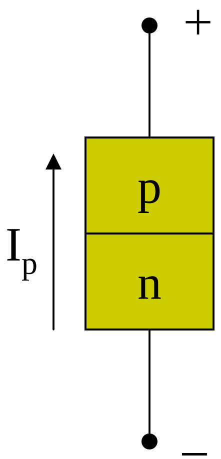


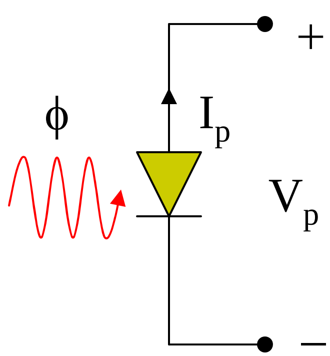
I-V characteristics of an illuminated junction

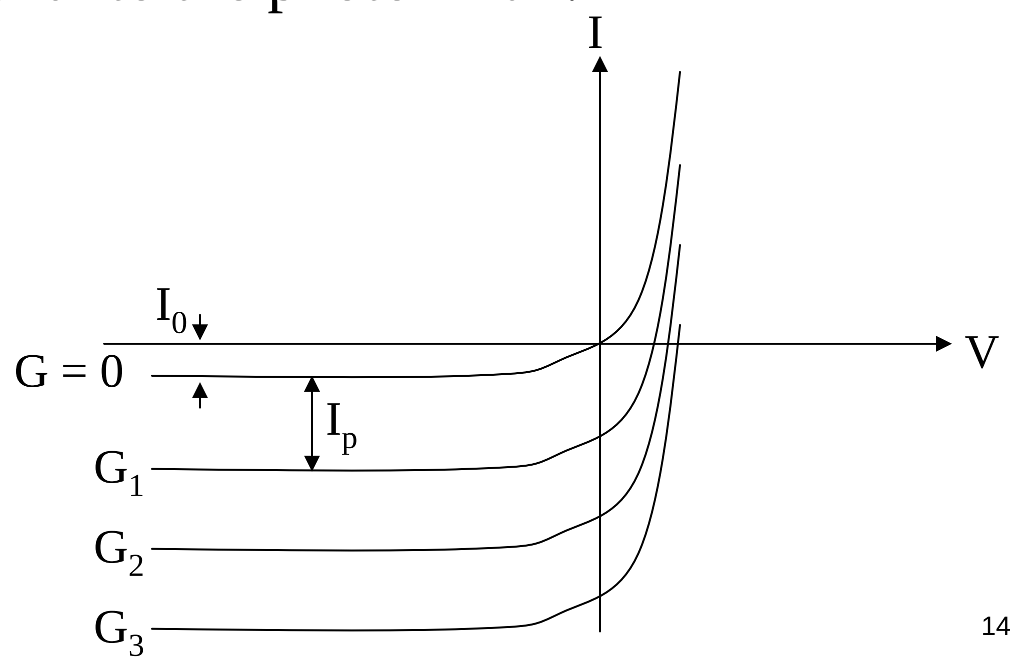
- The photodiode therefore has an I-V characteristic:

$$I = I_0(\exp(eV/k_BT) - 1) - I_p$$

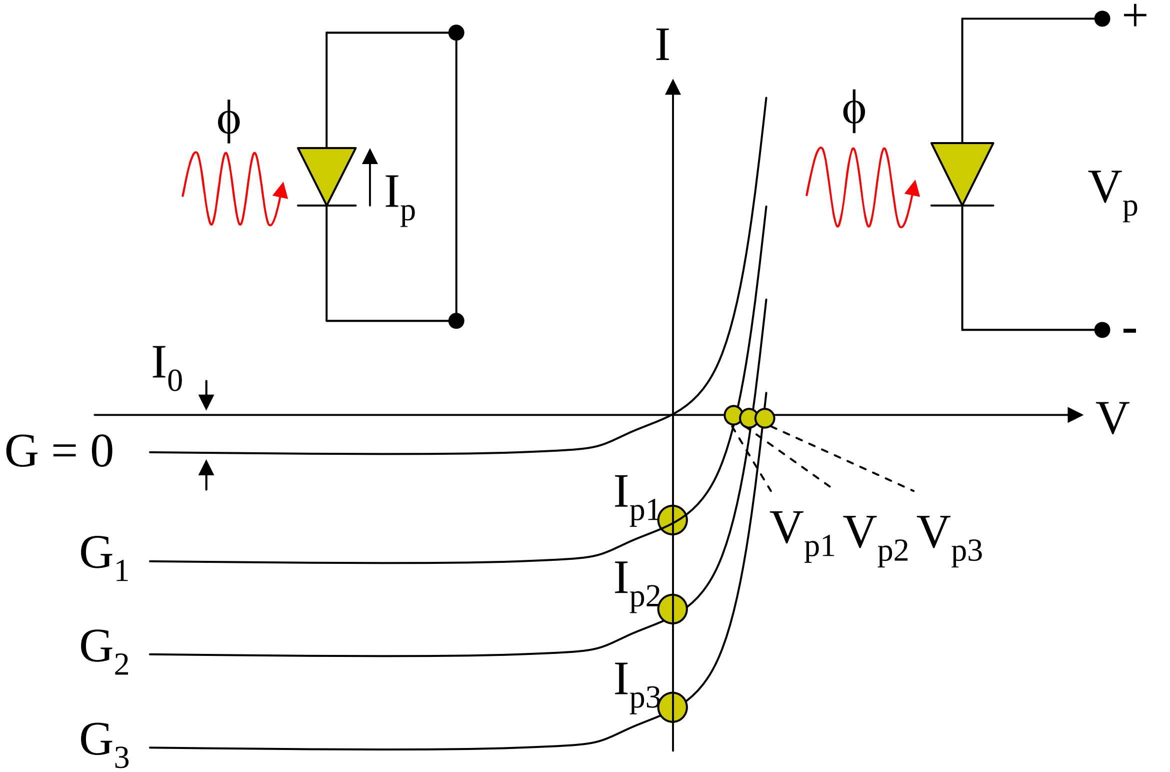
- This is the usual I-V curve of a p-n junction with an added photocurrent $-I_p$ proportional to the photon flux.







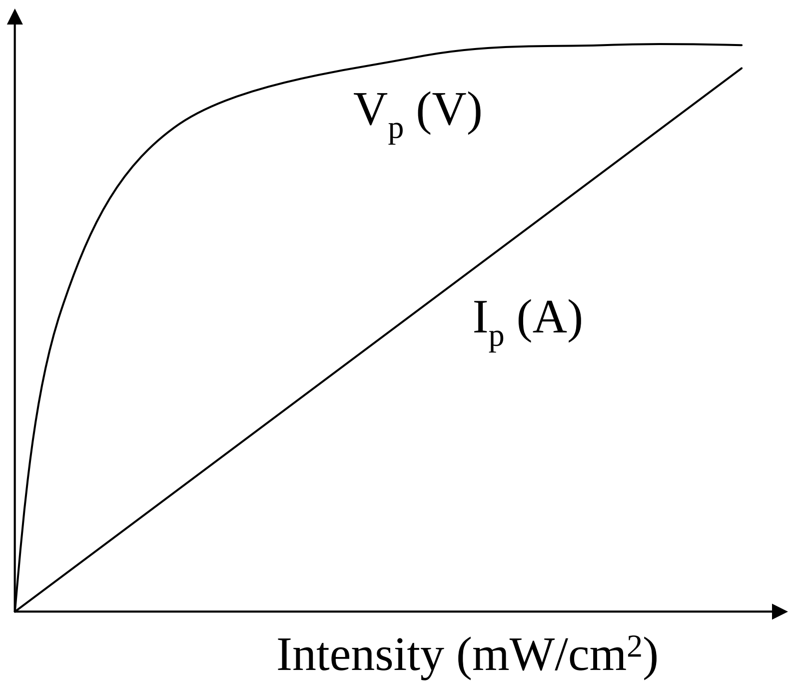
Short-circuit current and open-circuit voltage



- The *short-circuit current* (V = 0) is the *photocurrent* I_p.
- The *open-circuit voltage* (I = 0) is the *photovoltage* V_p.

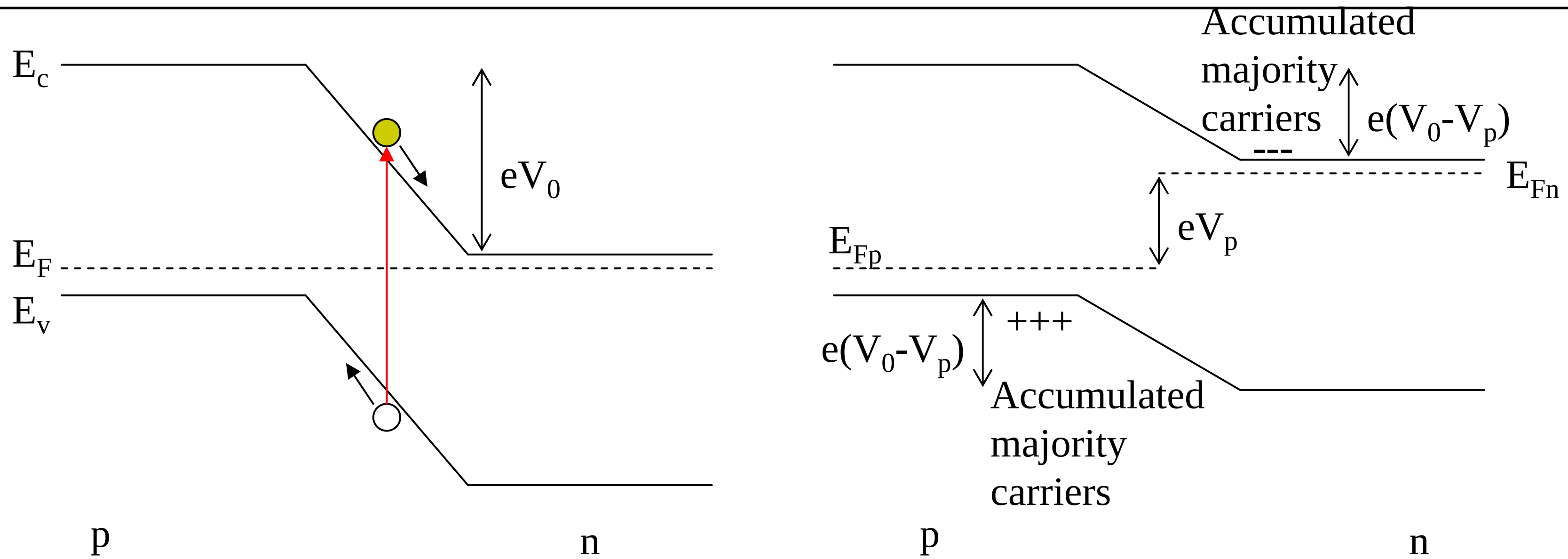
$$(I = 0) \Rightarrow \quad V_p = (k_B T/e)\ln(I_p/I_0 + 1)$$

Photocurrent and photovoltage



- As the light intensity increases, the short-circuit current increases linearly ($I_p \propto G$);
- The open-circuit voltage increases *only logarithmically* ($V_p \propto \ln (I_p/I_0)$) and limits by the *equilibrium contact potential*.

Open-circuit voltage



- The photogenerated, field-separated, majority carriers (+ve charge on the p-side, -ve charge on the n-side) forward-bias the junction.
- The appearance of a forward voltage across an illuminated junction (photovoltage) is known as the photovoltaic effect.
- The limit on V_p is the equilibrium contact potential V_0 as the contact potential is the maximum forward bias that can appear across a junction. (drift current vanishes with $V_p = V_0$)

Photoconductive and photovoltaic modes

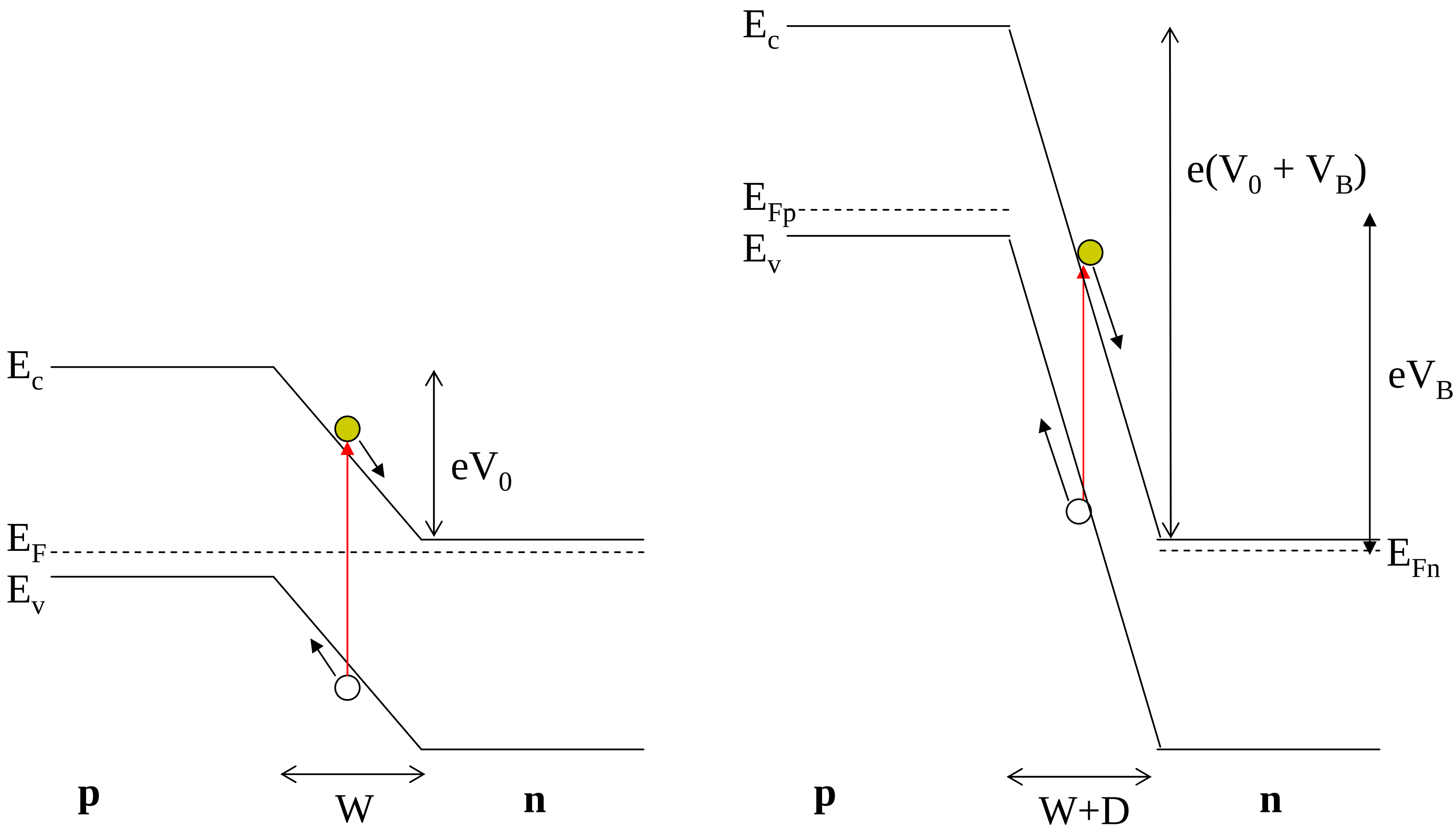
- There are *two* modes of operation for a junction photodiode: *photoconductive* and *photovoltaic*

- The device functions in *photoconductive* mode in the *third* quadrant of its current-voltage characteristics, including the *short-circuit condition* on the vertical axis for V = 0. (*acting as a current source*)

- It functions in *photovoltaic* mode in the *fourth* quadrant, including the *open-circuit condition* on the horizontal axis for I = 0. (*acting as a voltage source with output voltage limited by the equilibrium contact potential*)

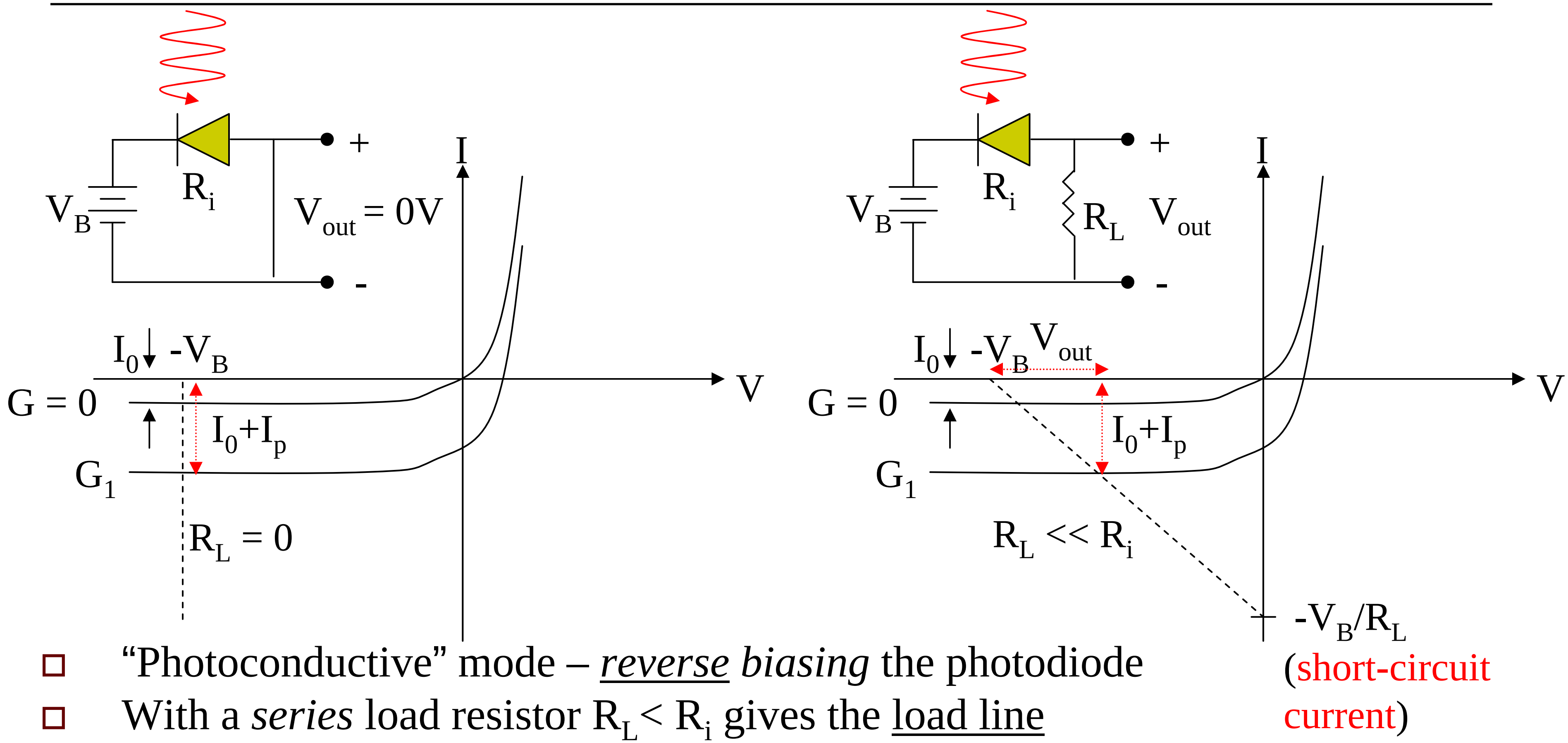
- The mode of operation is determined by the *bias condition* and the *external circuitry*.

Photoconductive mode under reverse bias



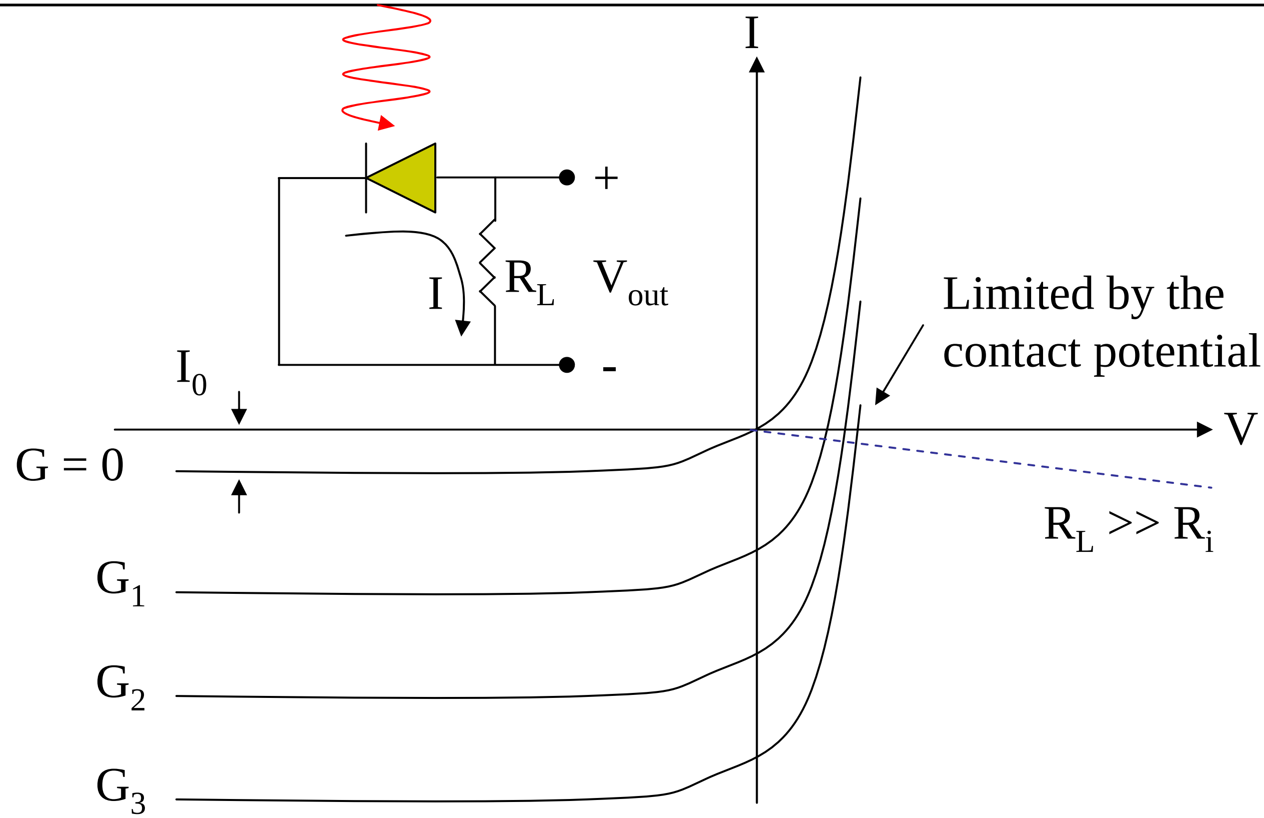
(For silicon photodiodes, $V_0 \approx 0.7$ V, V_B can be up to -5 – -10 V)

Basic circuitry and load line for the photoconductive mode



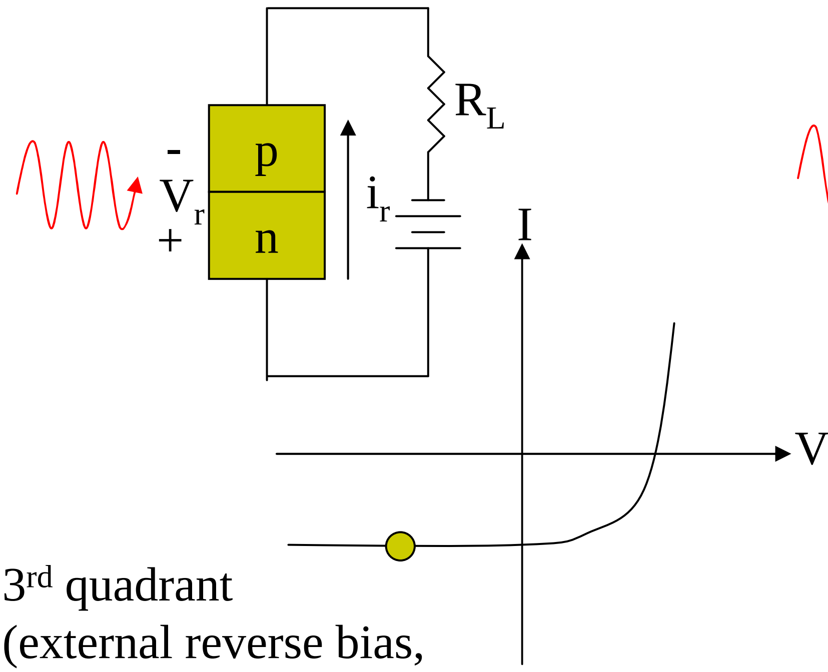
- "Photoconductive" mode – *reverse biasing* the photodiode
- With a *series* load resistor $R_L < R_i$ gives the load line
- Keep $V_{out} < V_B$ so that the photodiode is *reverse* biased (V_B *is sufficiently large*)
- Under these conditions and before it saturates, a photodiode has the following *linear response*: $V_{out} = (I_0 + I_p) R_L$

Basic circuitry and load line for the photovoltaic mode

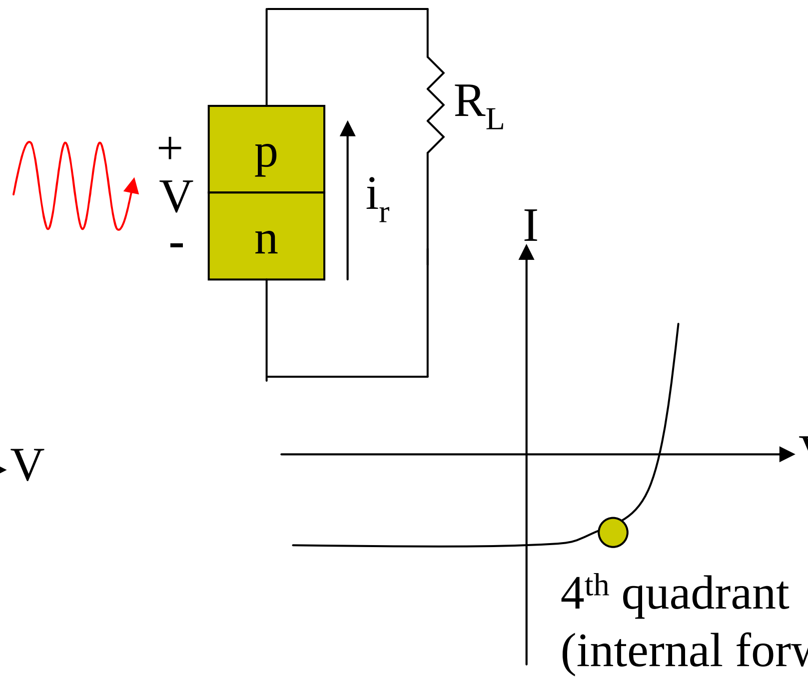


- *Does not require a bias voltage* but requires a ***large load resistance***.
- $R_L >> R_i$, so that the current I flowing through the diode and the load resistance is negligibly small.

Operation regimes of an illuminated junction



3rd quadrant
(external reverse bias,
reverse current)

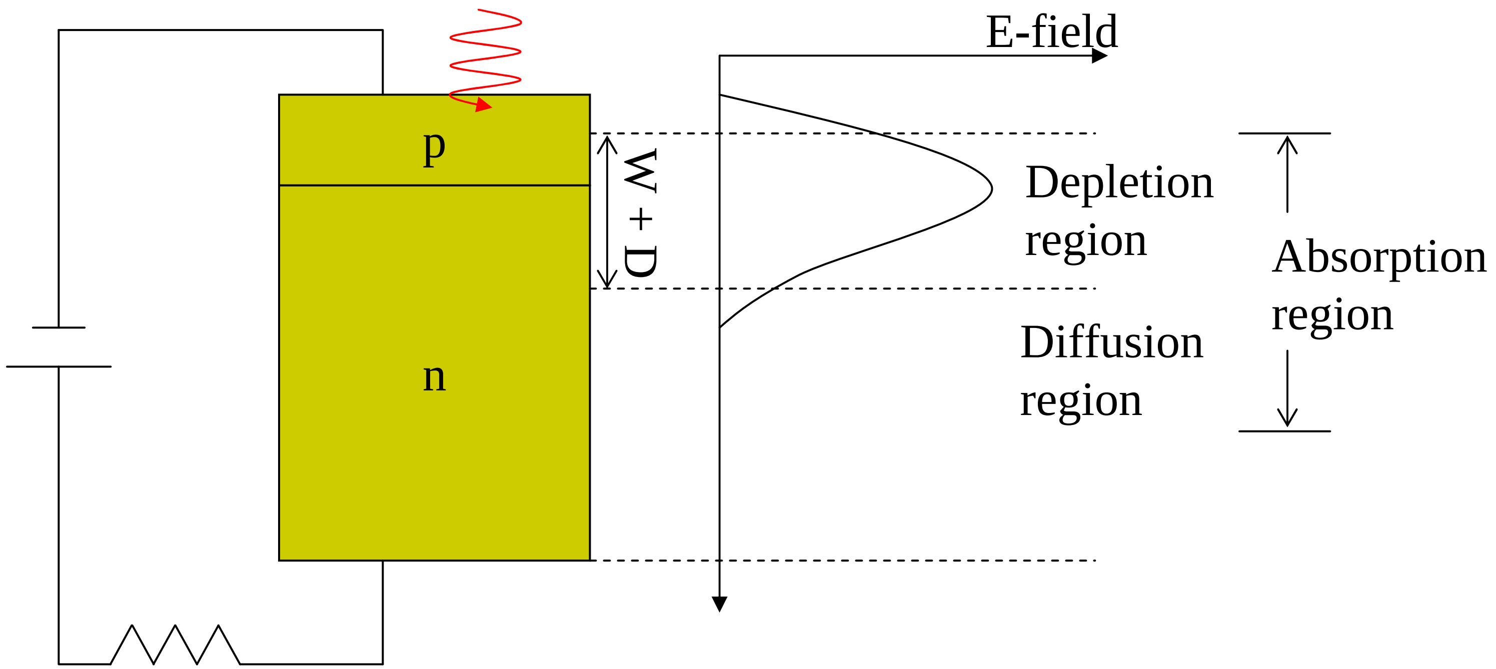


4th quadrant
(internal forward bias,
reverse current)

Photoconductive:
Power (+ve) is delivered *to the device by the external circuit* (photodetector)

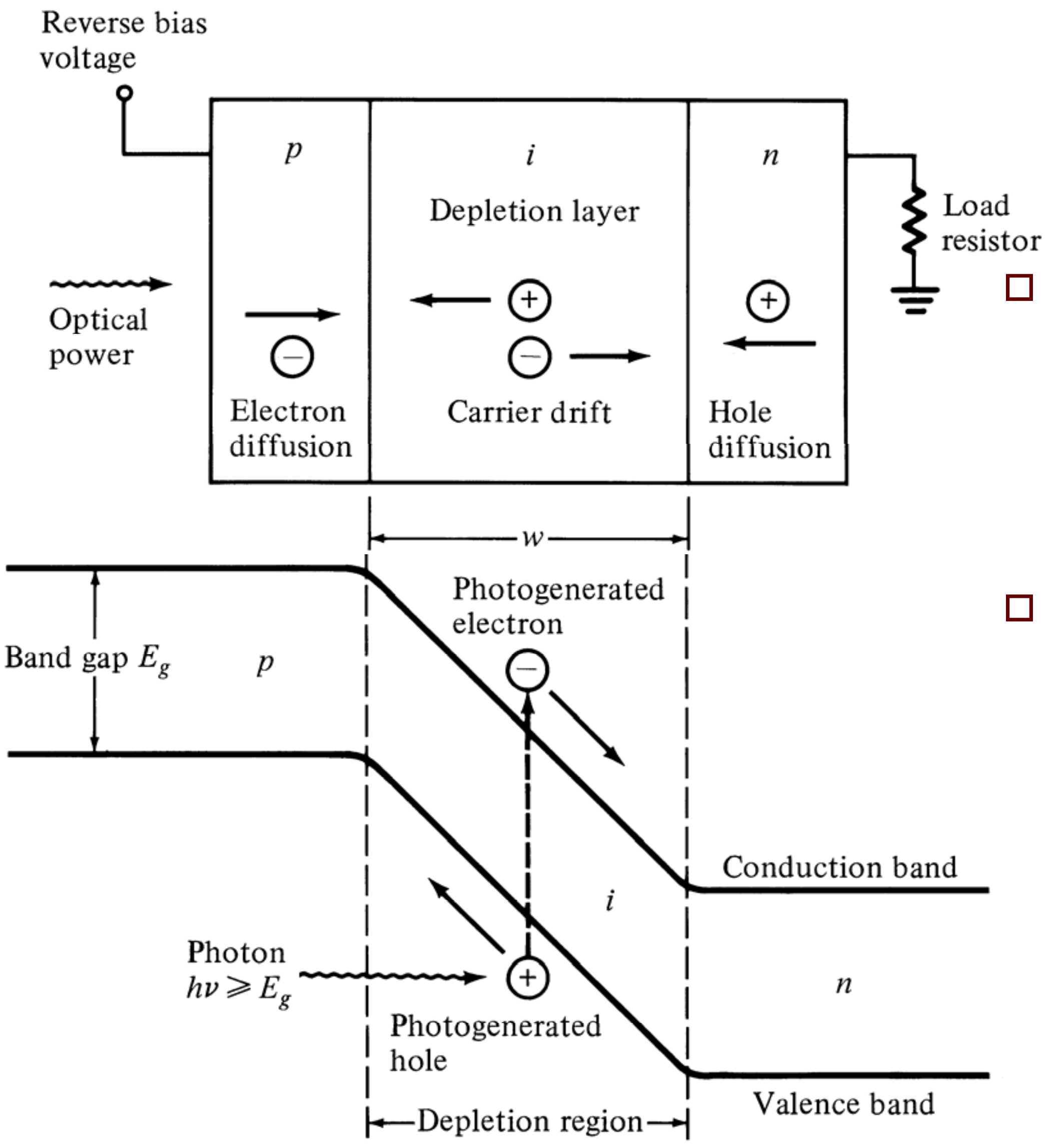
Photovoltaic:
Power (-ve) is delivered *to the load by the device* (solar cell/ energy harvesting)

A reverse-biased p-n photodiode



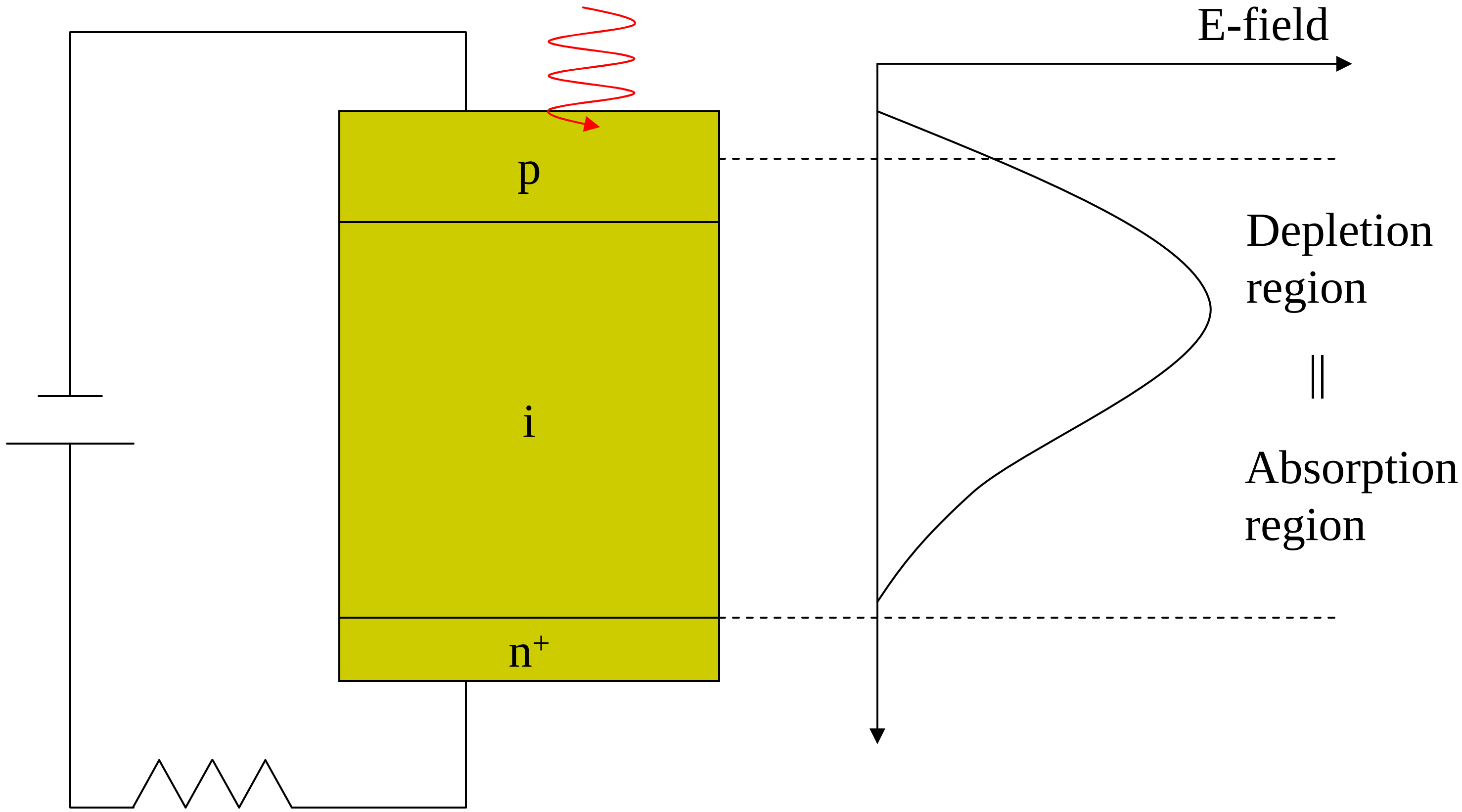
- It is important that the photons are absorbed in the depletion region. Thus, it is made as long as possible (say by decreasing the doping in the n type material). The depletion region width in a p-n photodiode is normally 1 – 3 μm.
- The *depletion-layer width widens* and the *junction capacitance drops* with reverse voltage across the junction.

p-i-n photodiodes



- A p-i-n photodiode consists of an *intrinsic* region sandwiched between heavily doped p^+ and n^+ regions. The *depletion layer* is almost completely defined by the intrinsic region.
- In practice, the intrinsic region does not have to be truly intrinsic but only has to be highly resistive (lightly doped p or n region).

A reverse-biased p-i-n photodiode

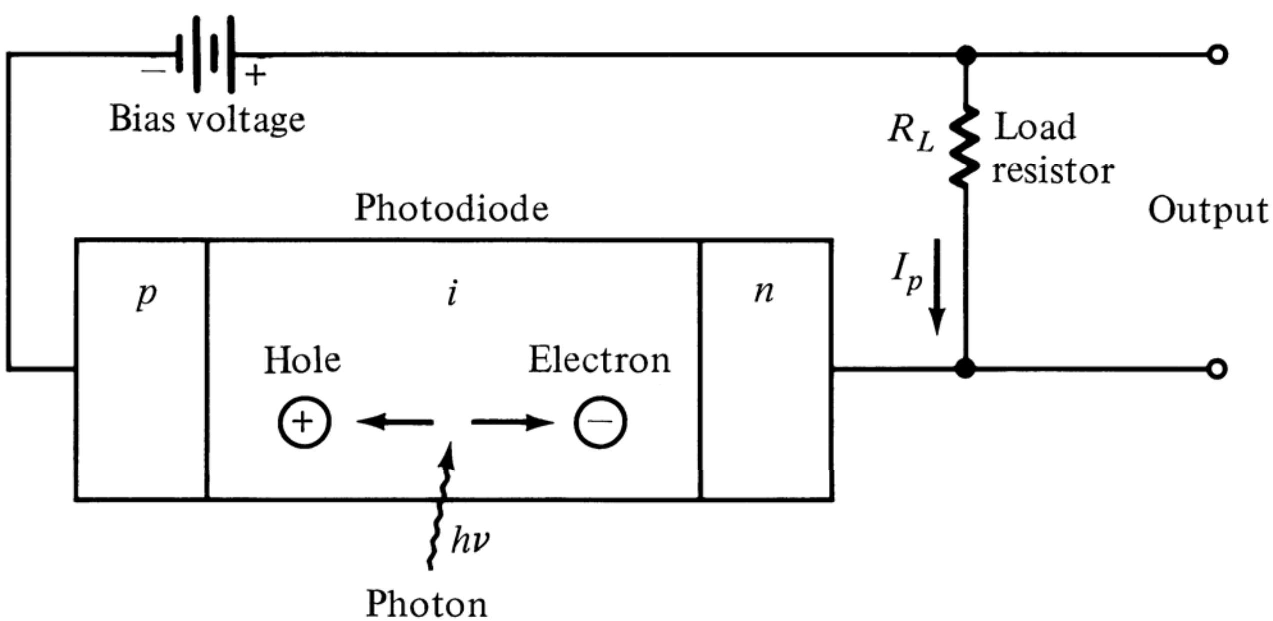


- All the absorption takes place in the depletion region. The intrinsic region can be an n-type material that is lightly doped, and to make a low-resistance contact a highly doped n-type (n^+) layer is added.

- The depletion-layer width W in a p-i-n diode does *not* vary significantly with bias voltage but is essentially fixed by the thickness, d_i, of the intrinsic region so that $W \approx d_i$.
- The internal capacitance of a p-i-n diode can be designed:

$$C_i = C_j = \varepsilon A/W \approx \varepsilon A/d_i$$

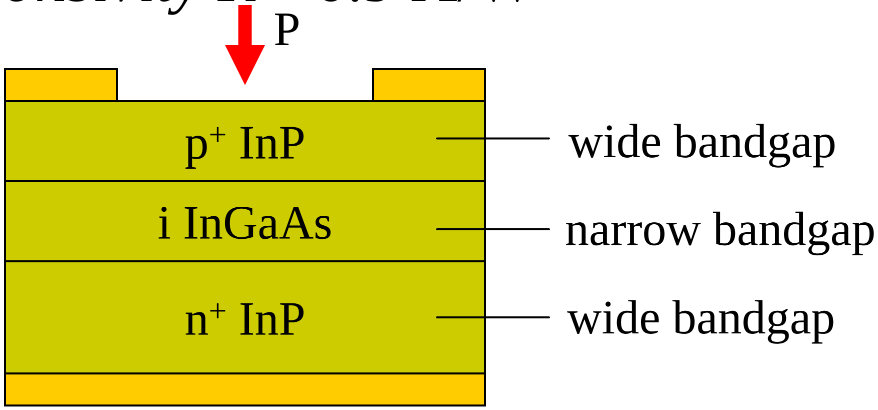
This capacitance is essentially independent of the bias voltage, remaining constant in operation.



- p-i-n photodiodes offer the following advantages:
  - *Increasing the width of the depletion layer* (where the generated carriers can be transported by drift) *increases the area available for capturing light*
  - Increasing the width of the depletion layer *reduces the junction capacitance* and *thereby the RC time constant*. Yet, the transit time increases with the width of the depletion layer.
  - Reducing the ratio between the diffusion length and the drift length of the device results in *a greater proportion of the generated current being carried by the faster drift process*.

Heterojunction photodiodes

- ***Many III-V p-i-n photodiodes have heterojunction structures***.
- Examples: p^+-AlGaAs/GaAs/n^+-AlGaAs, p^+-InP/InGaAs/n^+-InP, or p^+-AlGaAs/GaAs/n^+-GaAs, p^+-InGaAs/InGaAs/n^+-InP.
- AlGaAs/GaAs (0.7 – 0.87 μm)
- InGaAs/InP (1300 – 1600 nm). A typical InGaAs p-i-n photodetector operating at 1550 nm has a *quantum efficiency* $\eta \approx 0.75$ and a *responsivity* $R \approx 0.9$ A/W



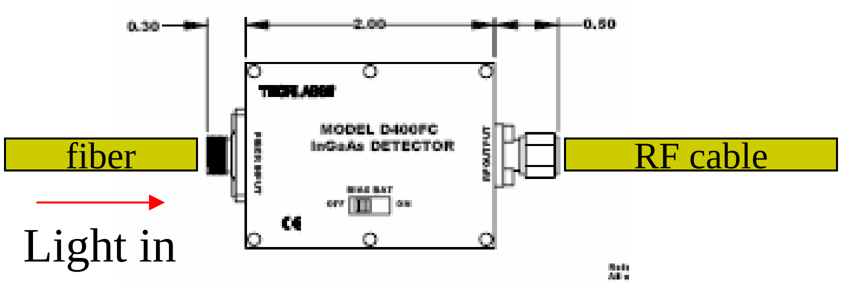
Heterojunction photodiodes

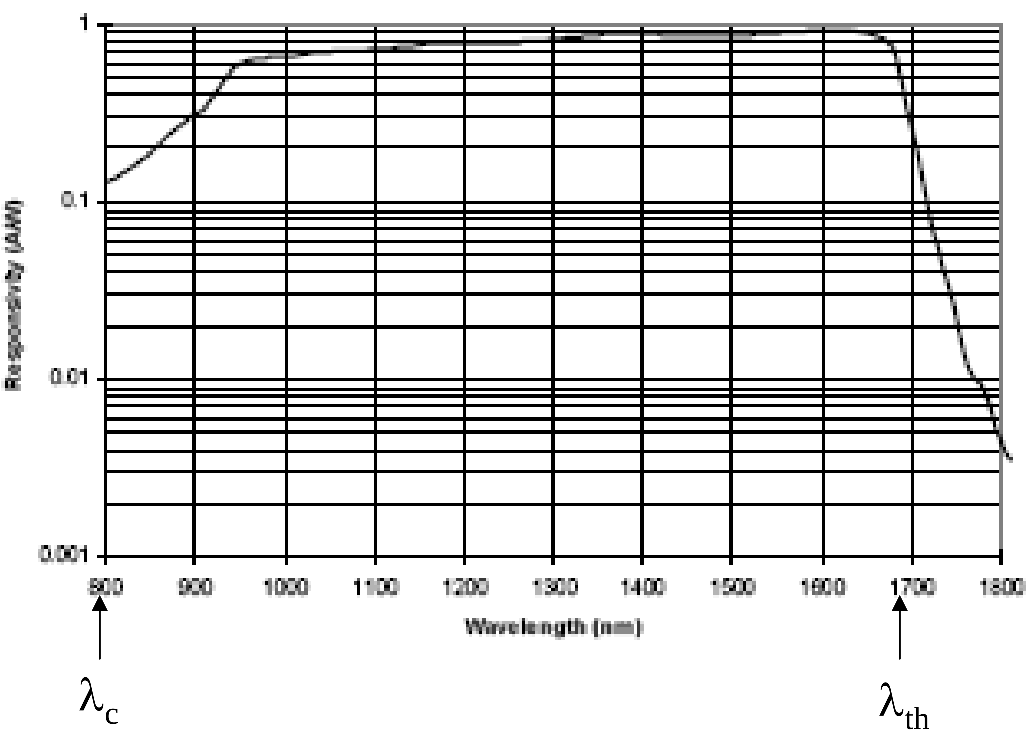
- Heterojunction structures offer additional flexibility in optimizing the performance of a photodiode.
- In a heterojunction photodiode, the active region normally has a bandgap that is *smaller* than one or both of the homogeneous regions.
- A *wide-bandgap homogeneous region*, which can be either the top p^+ region or the substrate n region, serves as a *window* for the optical signal to enter.
- The *small bandgap of the active region* determines the *long-wavelength cutoff* of the photoresponse, λ_{th}.
- The *large bandgap of the homogeneous window region* sets the short-wavelength cutoff of the photoresponse, λ_c.

=> For an optical signal that has a wavelength λ_s in the range $\lambda_{th} > \lambda_s > \lambda_c$, the *quantum efficiency* and the *responsivity* can be optimized.

InGaAs fiber-optic pin photodetector

(Thorlabs D400FC)





Spectral response	800 – 1700 nm
Peak response	0.95 A/W @ 1550 nm
Rise/fall time	0.1 ns
Diode capacitance	0.7 pF (typ)
NEP @ 1550 nm	1.0×10^{-15} W/√Hz
Dark current	0.7nA (typ), 1.0nA (max)
PD Active diameter	0.1 mm
Bandwidth	1 GHz (min)
Damage threshold	100 mW CW
Bias (reverse)	12V battery
Coupling lens	0.8” dia. Ball lens
Coupling efficiency	92% (typ) from both single- and multi-mode fibers over full spectral response

Application notes – output voltage

- The RF output signal (suitable for both pulsed and CW light sources) is the direct photocurrent out of the photodiode anode and is a function of the incident light power and wavelength.
- The *responsivity* $R(\lambda)$ can be used to estimate the amount of photocurrent.
- *To convert this photocurrent to a voltage* (say for viewing on an oscilloscope), add an external *load resistance*, R_L.
- The *output voltage* is given as:

$$V_0 = P\, R(\lambda)\, R_L$$

Responsivity

- The *responsivity* of a photodetector relates the electric current I_p flowing in the device circuit to the optical power P incident on it.

$$I_p = \eta \, e\Phi = \eta \, eP/h\upsilon \equiv R \, P \qquad \eta\text{: quantum efficiency}$$

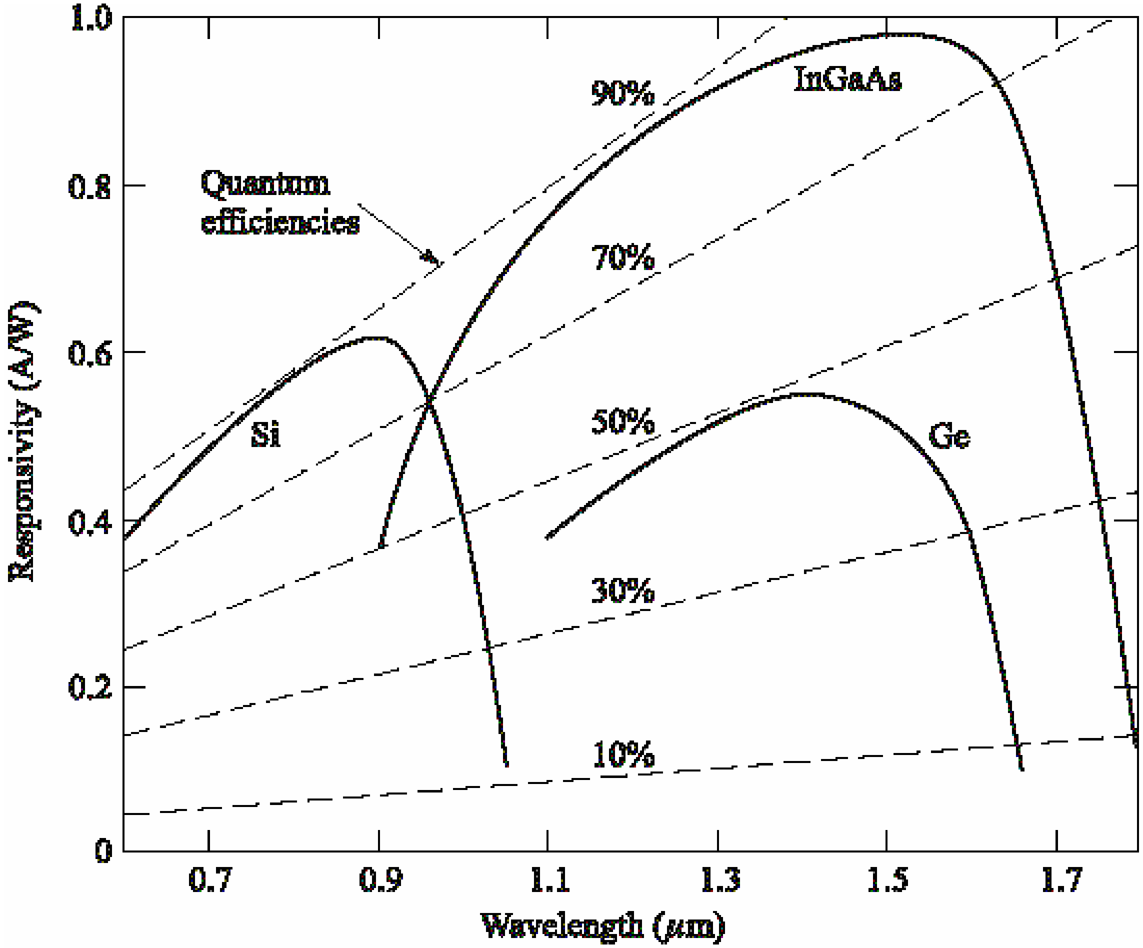
$$\text{Responsivity } R = I_p/P = \eta e/h\upsilon = \eta \, \lambda/1.24 \text{ [A/W]}$$

(Recall the LED responsivity [W/A])

- The responsivity is *linearly proportional* to both the *quantum efficiency* η and the free-space wavelength λ.

  (e.g. for $\eta = 1$, $\lambda = 1.24$ μm, R = 1 A/W)

Responsivity vs. wavelength



- Responsivity R (A/W) vs. wavelength with the quantum efficiency η shown on various dashed lines

Quantum efficiency

- The ***quantum efficiency*** (***external quantum efficiency***) η of a photodetector is the probability that a single photon incident on the device generates a photocarrier pair that contributes to the detector current.

$$\eta(\lambda) = \zeta\,(1\text{-}R)\,[1 - \exp(-\alpha(\lambda)d)]$$

R is the optical power reflectance at the surface, ζ is the fraction of electron-hole pairs that contribute to the detector current, $\alpha(\lambda)$ the absorption coefficient of the material, and d the photodetector depth.

ζ is the fraction of electron-hole pairs that *avoid recombination (often dominated at the material surface)* and contribute to the useful photocurrent. *Surface recombination* can be reduced by careful material growth and device design/fabrication.

$[1 - \exp(-\alpha(\lambda)d)]$ represents the fraction of the photon flux absorbed in the bulk of the material. The device should have a value of d that is *sufficiently large*. ($d > 1/\alpha$, $\alpha = 10^4\ cm^{-1}$, $d > 1\ \mu m$)

Dependence of quantum efficiency on wavelengths

- The characteristics of the semiconductor material determines the spectral window for large η.
- The bandgap wavelength $\lambda_g = hc/E_g$ is the *long-wavelength limit* of the semiconductor material.
- *For sufficiently short λ*, η also *decreases* because most photons are absorbed *near the surface* of the device (e.g. for $\alpha = 10^4$ cm^{-1}, most of the light is absorbed within a distance $1/\alpha = 1$ µm; for $\alpha = 10^5 - 10^6$ cm^{-1}, most of the light is absorbed within a distance $1/\alpha = 0.1 - 0.01$ µm).
- *The recombination lifetime is quite short near the surface*, so that the *photocarriers recombine before being collected*. (short-*wavelength limit*)
- In the near-infrared region, silicon photodiodes with *antireflection coating* can reach 100% quantum efficiency near 0.8 – 0.9 µm.
- In the 1.0 – 1.6 µm region, Ge photodiodes, InGaAs photodiodes, and InGaAsP photodiodes have shown high quantum efficiencies.

Application notes - bandwidth

- The *bandwidth*, f_{3dB}, and the *10 - 90% rise-time response*, t_r, are determined from the *diode capacitance* C_j, and the *load resistance* R_L:

$$f_{3dB} = 1/(2\pi R_L C_j)$$

$$t_r = 0.35/f_{3dB}$$

- *For maximum bandwidth*, use a direct connection to the measurement device having a 50 Ω input impedance. An SMA-SMA RF cable with a 50 Ω terminating resistor at the end can also be used. This will *minimize ringing by matching the coax with its characteristic impedance*.
- If bandwidth is not important, such as for continuous wave (CW) measurement, one can increase the amount of voltage for a given input light by increasing the R_L up to a maximum value (say 10 kΩ).

Speed-limiting factors of a photodiode

- ***High-speed photodiodes*** are by far the most widely used photodetectors in applications requiring high-speed or broadband photodetection.

- The speed of a photodiode is determined by *two* factors:
  - The *response time of the photocurrent*
  - The RC *time constant of its equivalent circuit*

- Because a photodiode operating in *photovoltaic* mode has a large RC time constant due to the large internal *diffusion capacitance upon internal forward bias* in this mode of operation

  => *only photodiodes operating in a photoconductive mode are suitable for high-speed or broadband applications*.

Response time of the photocurrent (photoconductive mode)

- The response time is determined by two factors:
  - *Drift* of the electrons and holes that are photogenerated in the *depletion layer*
  - *Diffusion* of the electrons and holes that are photogenerated in the *diffusion regions*
- ***Drift*** of the carriers across the depletion layer is a *fast* process - given by the *transit times* of the photogenerated electrons and holes across the depletion layer.
- ***Diffusion*** of the carriers is a *slow* process – caused by the optical absorption in the diffusion regions outside of the high-field depletion region.

  (*diffusion current can last as long as the carrier lifetime*)

  => *a long tail in the impulse response of the photodiode*

  => a *low-frequency falloff* in the device frequency response

Drift velocity and carrier mobility

- A constant electric field **E** presented to a semiconductor (or metal) causes its free charge carriers to *accelerate*.
- The accelerated free carriers then encounter frequent *collisions with lattice ions moving about their equilibrium positions* via thermal motion and imperfections in the crystal lattice (e.g. associated with impurity ions).
- These collisions cause the carriers to suffer *random decelerations* (*like frictional force*!)

  => the result is motion at an *average velocity* rather than at a constant acceleration.
- The *mean* ***drift velocity*** of a carrier

$$v_d = (eE/m)\,\tau_{col} = \mu E$$

where m is the effective mass, τ_{col} is the *mean time between collisions*, $\mu = e\tau_{col}/m$ is the ***carrier mobility***.

Drift time upon saturated carrier velocities

- When the field in the depletion region exceeds a *saturation* value then the carriers travel at a *maximum* drift velocity v_d.
- The *longest* transit time τ_{tr} is for carriers which must traverse the full depletion layer width W:

$$\tau_{tr} = W/v_d$$

A field strength above 2 x 10^4 Vcm^{-1} (say 2 V across 1 μm distance) in silicon gives maximum (*saturated*) carrier velocities of approximately 10^7 cms^{-1}. (max. v_d)

=> The transit time through a depletion layer width of 1 μm is around 10 ps.

Diffusion time

- ***Diffusion time of carriers generated outside the depletion region*** – carrier diffusion is a relatively *slow* process. The diffusion time, τ_{diff}, for carriers to diffuse a distance d is

$$\tau_{diff} = d^2/2D$$

where D is the *minority carrier diffusion coefficient*.

e.g. The hole diffusion time through 10 μm of silicon is 40 ns. The electron diffusion time over a similar distance is around 8 ns.

=> *for a high-speed photodiode, this diffusion mechanism has to be eliminated* (by reducing the photogeneration of carriers outside the depletion layer through design of the device structure, say using heterojunction pin diode).

Photodiode capacitance

- ***Time constant incurred by the capacitance of the photodiode with its <u>load</u>*** – the *junction capacitance*

$$C_j = \varepsilon A/W$$

where ε is the permittivity of the semiconductor material and A is the diode junction area.

⇒ A small depletion layer width W increases the junction capacitance.

(The *capacitance of the photodiode* C_{pd} is that of the <u>junction</u> *together with the capacitance of the <u>leads</u> and <u>packaging</u>*. This capacitance must be *minimized* in order to reduce the RC time constant. In ideal cases, $C_{pd} \approx C_j$.)

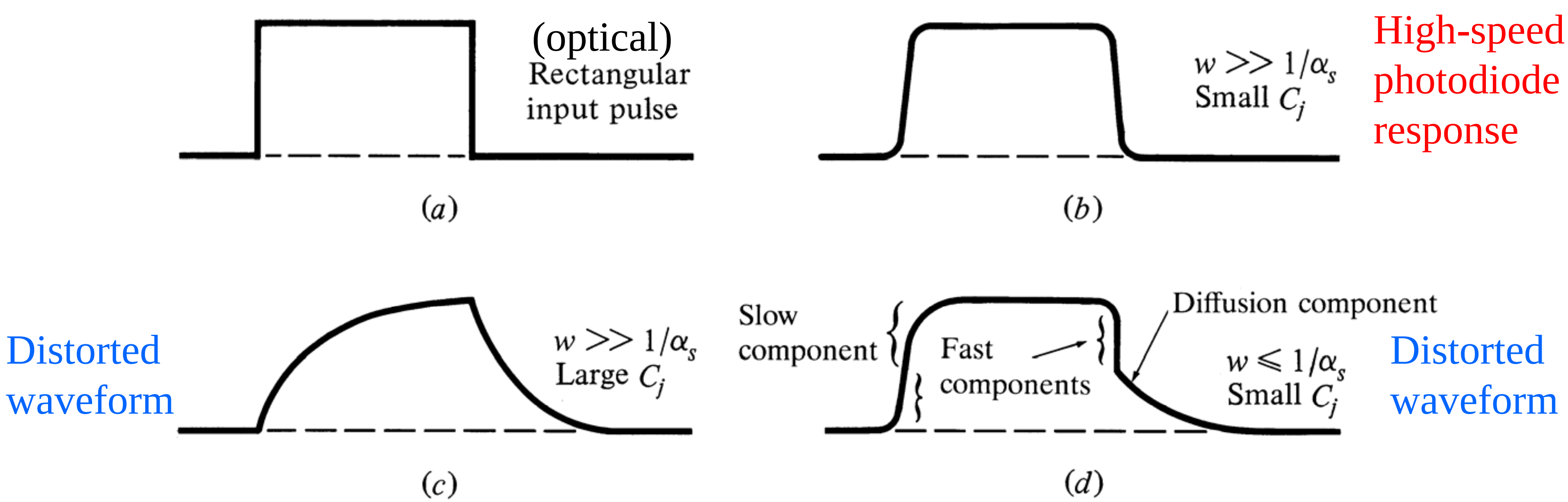
Remarks on junction capacitance

- ***For pn junctions***, because the width of the depletion layer decreases with forward bias but increases with reverse bias, the junction capacitance increases when the junction is subject to a forward bias voltage but *decreases when it is subject to a reverse bias voltage*.
- ***For p-i-n diodes***, the width of the depletion (*intrinsic*) layer is fixed => the junction capacitance is *not* affected by biasing conditions.
- **e.g.** A GaAs p-n homojunction has a 100 μm x 100 μm cross section and a width of the depletion layer W = 440 nm. Consider the junction in thermal equilibrium *without* bias at 300 K. Find the *junction* capacitance.

$$\varepsilon = 13.18\varepsilon_0 \text{ for GaAs, } \varepsilon_0 = 8.854 \times 10^{-12}\ \text{Fm}^{-1}$$

$\Rightarrow$ $C_j = 13.18 \times 8.854 \times 10^{-12} \times 1 \times 10^{-8}/(440 \times 10^{-9}) = 2.65$ pF

Photodiode response to rectangular optical input pulses for various detector parameters



(b) W >> 1/α (all photons are absorbed in the depletion layer) and small C_j.

(c) W >> 1/α, *large photodiode capacitance*, RC time limited

(d) W ≤ 1/α, (some photons are absorbed in the diffusion region) *diffusion component limited*

Transit-time-limited

- Thus, *for a high-speed photodiode, diffusion mechanism has to be eliminated* (by reducing the photogeneration of carriers outside the depletion layer through design of the device structure).
- *When the diffusion mechanism is eliminated*, the frequency response of the photocurrent is *only limited by the transit times of electrons and holes*.

- In a semiconductor, electrons normally have a higher mobility (*smaller electron effective mass*), thus a smaller transit time, than holes.

- *For a good estimate* of the detector frequency response, we use the *average* of electron and hole transit times:

$$\tau_{tr} = \tfrac{1}{2}(\tau_{tr}^{e} + \tau_{tr}^{h})$$

Approximated transit-time-limited power spectrum

- In the simple case when the process of carrier drift is dominated by a *constant transit time* of τ_{tr}

  => the temporal response of the photocurrent is ideally a *rectangular* function of duration τ_{tr} in the *time domain*

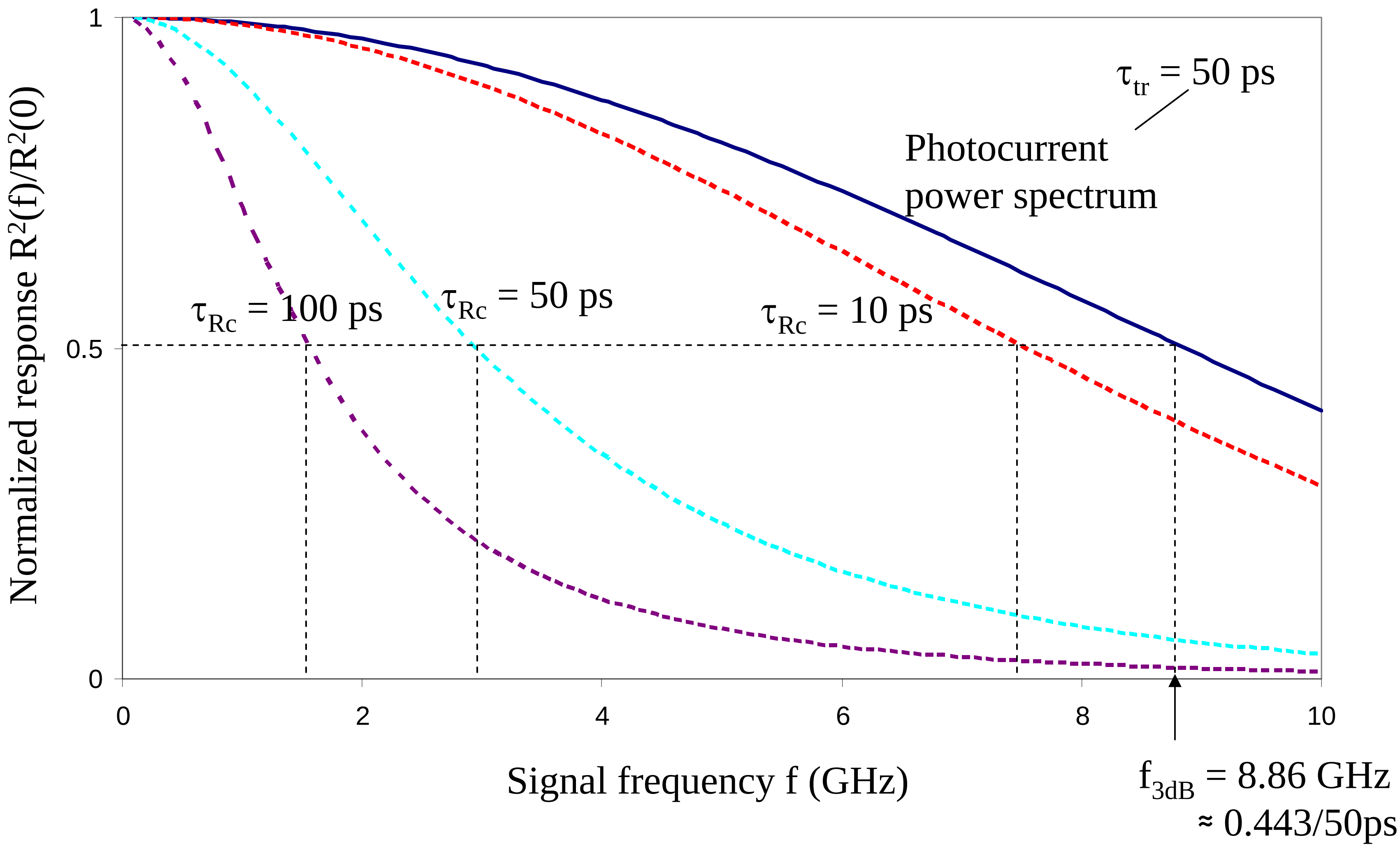
  => the *power spectrum* of the photocurrent frequency response can be approximately given as a sinc function in the *frequency domain*:

$$\mathcal{R}_{ph}^2(f) = |i_{ph}(f)/P(f)|^2 \approx \mathcal{R}_{ph}^2(0)\,(\sin(\pi f \tau_{tr})/\pi f \tau_{tr})^2$$

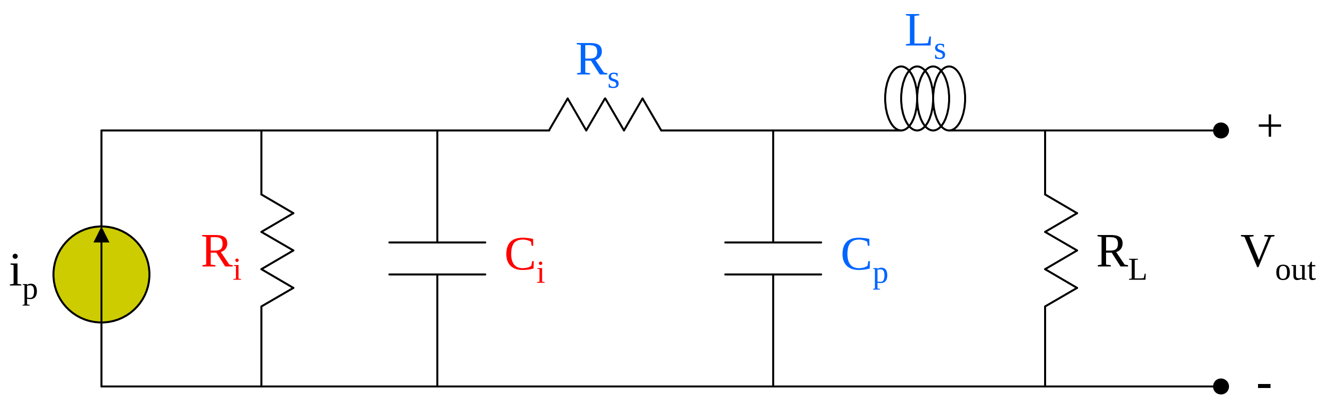
  =>a *transit-time-limited 3-dB frequency*:

$$f_{ph,3dB} \approx 0.443/\tau_{tr}$$

Total frequency response



Small-signal equivalent circuits



- A photodiode has an *internal resistance* R_i and an *internal capacitance* C_i *across its junction*.
- The *series resistance* R_s takes into account both resistance in the *homogeneous regions* of the diode and *parasitic resistance* from the contacts.
- The *external parallel capacitance* C_p is the *parasitic capacitance* from the contacts and the package.
- The series inductance L_s is the parasitic inductance from the wire or transmission-line connections.
- The values of R_s, C_p, and L_s can be minimized with careful design, processing, and packaging of the device.

- Both R_i and C_i depend on the *size* and the *structure* of the photodiode and *vary with the voltage across the junction*.

- In ***photoconductive*** mode under a *reverse* voltage, the diode has a *large* R_i normally on the order of 1 – 100 MΩ for a typical photodiode, and a *small* C_i dominated by the junction capacitance C_j.
- As the reverse voltage increases in magnitude, R_i increases but C_i decreases *because the depletion-layer width increases with reverse voltage*.

- In ***photovoltaic*** mode with a *forward* voltage across the junction, the diode has a *large* C_i dominated by the diffusion capacitance C_d.
- It still has a large R_i, *though smaller than that in the photoconductive mode*.

Remark on diffusion capacitance

- Because the ***diffusion capacitance*** is associated with the *storage of majority carrier charges in the diffusion region (photogenerated electrons and holes swept from the depletion region stored in the n side and the p side)*, it exists *only when a junction is under forward bias*.

- When a junction is under forward bias, *C_d can be significantly larger than C_j at high injection currents*.

- When a junction is *under reverse bias, C_j is the only capacitance of significance*.

=> *the capacitance of a junction under reverse bias can be substantially smaller* than when it is under forward bias.

Frequency response of the equivalent circuit

- The frequency response of the equivalent circuit is determined by
  - The *internal* resistance R_i and capacitance C_i of the photodiode
  - The *parasitic* effects characterized by R_s, C_p, and L_S
  - The *load* resistance R_L
- The *parasitic effects must be eliminated* as much as possible.

- A high-speed photodiode normally operates under the condition that $R_i >> R_L, R_s$.
  => *equivalent resistance* ≈ R_L

- *In the simple case, when the parasitic inductance/capacitance are negligible*, the speed of the circuit is dictated by the RC time constant $\tau_{RC} = R_L C_i$.

Approximated power spectrum

- The *equivalent circuit frequency response*:

$$R_c^2(f) \approx R_c^2(0)/(1 + 4\pi^2 f^2 \tau_{RC}^2)$$

- An *<u>RC-time-limited</u> 3-dB frequency*

$$f_{c,3dB} \approx 1/2\pi\tau_{RC} = 1/2\pi R_L C_i$$

- Combining the *photocurrent response* and the *circuit response*, the *total output power spectrum* of an optimized photodiode operating in *photoconductive* mode

$$R^2(f) = R_c^2(f)\, R_{ph}^2(f) \approx [R_c^2(0)/(1+4\pi^2 f^2 \tau_{RC}^2)]\, (\sin(\pi f \tau_{tr})/\pi f \tau_{tr})^2$$

RC-time-limited bandwidth

e.g. In a silicon photodiode with W = 1 μm driven at *saturation drift velocity*,

$$\tau_{tr} \approx 10^{-4}\ \text{cm}/10^{7}\ \text{cms}^{-1} \approx 10\ \text{ps}$$

suppose the diode capacitance = 1 pF and a load resistance of 50 Ω,

$$\tau_{RC} \approx 50\ \text{ps}$$

$$\Rightarrow\ f_{3dB} \approx 1/2\pi\tau_{RC} \approx 3.2\ \text{GHz}$$

Rise and fall times upon a square-pulse signal

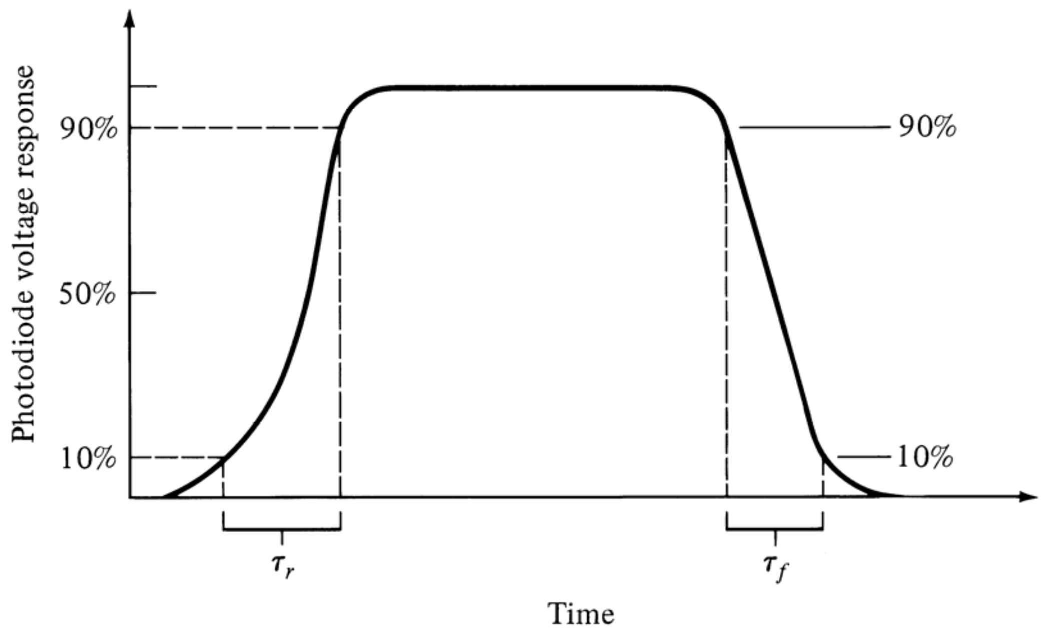
- In the time domain, the speed of a photodetector is characterized by the *risetime*, τ_r, and the *falltime*, τ_f, of its response to a square-pulse signal.

- The *risetime* - the time interval for the response to rise from 10 to 90% of its peak value.

- The *falltime* - the time interval for the response to decay from 90 to 10% of its peak value.

- The risetime of the *square-pulse response* is determined by the *RC circuit-limited bandwidth* of the photodetector.

Rise time and the circuit 3-dB bandwidth



- Typical response of a photodetector to a square-pulse signal
  => the *3-dB bandwidth (for the RC circuit)* is

$$\tau_r = 0.35/f_{3dB}$$

- For a voltage step input of amplitude V, the output voltage waveform $V_{out}(t)$ as a function of time t is:

$$V_{out}(t) = V[1 - \exp(-t/RC)]$$

$\Rightarrow$ the 10 to 90% rise time τ_r for the circuit is given by:

$$\tau_r = 2.2\ RC$$

- The *transfer function* for this circuit is given by

$$|H(\omega)| = 1/[1 + \omega^2(RC)^2]^{1/2}$$

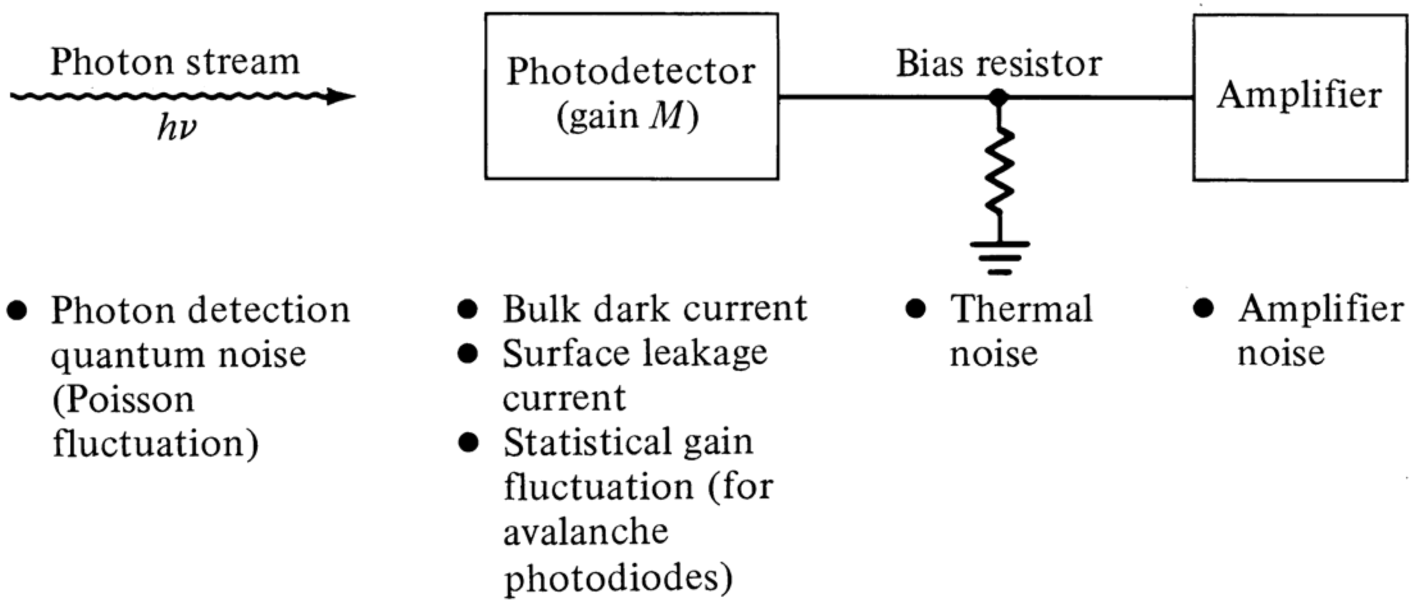
- The *3-dB bandwidth for the circuit* is

$$f_{3dB} = 1/2\pi RC$$

$$\Rightarrow\ \tau_r = 2.2/2\pi f_{3dB} = 0.35/f_{3dB}$$

Noise in photodetectors

Noise sources and disturbances



Photon noise – the most fundamental source of noise is associated with the *random* arrivals of the photons (usually described by *Poisson statistics*)

Photoelectron noise – a single photon generates an electron-hole pair with probability η. The photocarrier-generation process is random.

Gain noise – the amplification process that provides internal gain in certain photodetectors is stochastic.

Receiver circuit noise – various components in the electrical circuitry of an optical receiver, such as *thermal noise* in resistors.

Performance measures

- The ***signal-to-noise ratio*** (SNR) of a random variable - the ratio of its *square-mean to its variance*. Thus, the SNR of the current i is SNR = $\langle i \rangle^2/\sigma_i^2$, while the SNR of the photon number is SNR = $\langle n \rangle^2/\sigma_n^2$

- The ***minimum-detectable signal*** – the *mean* signal that yields SNR = 1

- The ***bit error rate*** (BER) – the probability of error per bit in a digital optical receiver.

- The ***receiver sensitivity*** – the signal that corresponds to a prescribed value of the SNR. While the *minimum-detectable signal* corresponds to a *receiver sensitivity* that provides SNR = 1, a higher value of SNR is often specified to ensure a given value of accuracy

  (e.g. SNR = 10 – 10^3 corresponding to 10 – 30 dB).

  For a digital system, the receiver sensitivity is defined as the *minimum optical energy or corresponding mean number of photons per bit* required to attain a prescribed BER (e.g. BER = 10^{-9} or better).

Photon noise

- The photon flux associated with a fixed optical power P is inherently uncertain (statistical).

- The mean photon flux is $\Phi = P/h\upsilon$, but this quantity fluctuates randomly in accordance with a probability law that depends on the nature of the light source.

- The number of photons n counted in a time interval T is thus random with *mean* $\langle n \rangle = \Phi T$.

- For *monochromatic coherent* radiation, the photon number statistics obeys the *Poisson probability distribution* $\sigma_n^2 = \langle n \rangle$ (i.e. variance equals mean)

=> the fluctuations associated with an *average of 100* photons result in an actual number of photons that lies approximately within the range 100 ± 10.

Poisson distribution

- The statistics arriving at a detector follows a *discrete probability distribution* which is *independent of the number of photons previously detected*.
- The probability P(z) of detecting z photons in time period τ when it is expected on *average* to detect z_m photons obeys the *Poisson distribution*

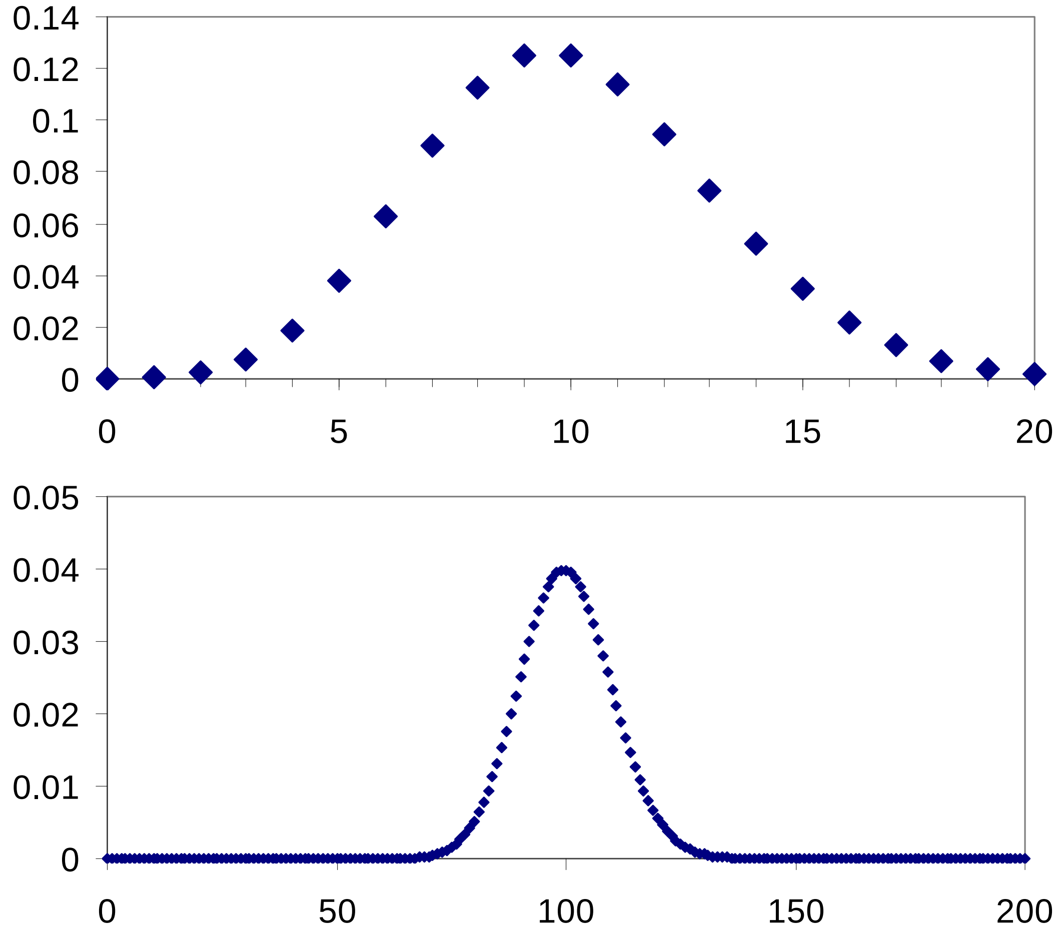
$$P(z) = z_m^{\,z} \exp(-z_m)/z!$$

where z_m *the mean is equal to the variance* of the probability distribution.

- The number of electrons generated in time τ is equal to the average number of photons detected over this time period

$$z_m = \eta P\tau/h\upsilon$$

Poisson distributions for $z_m = 10$ and $z_m = 100$

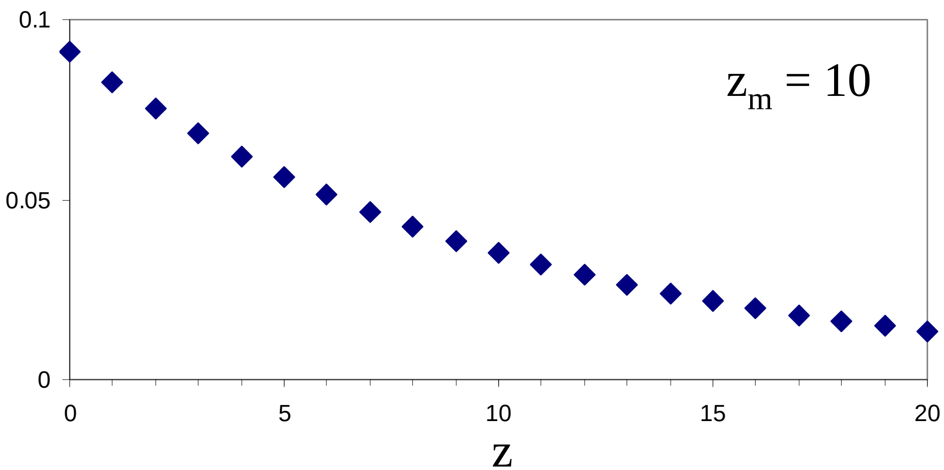


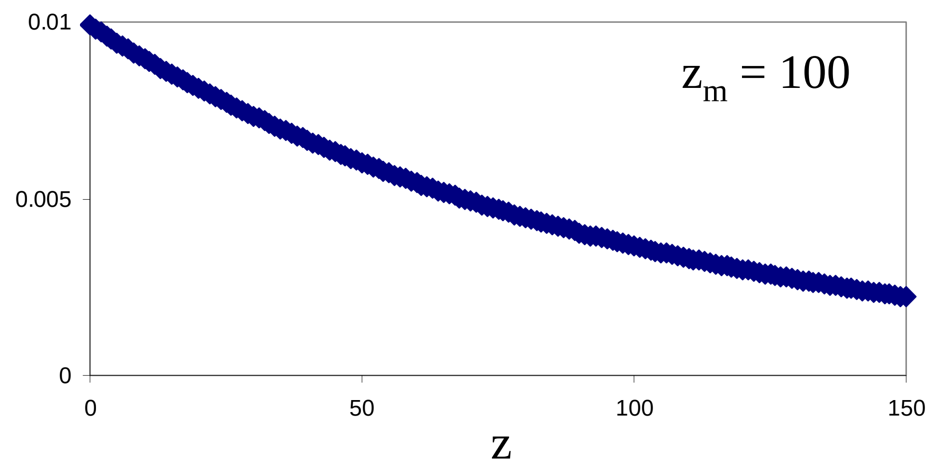
- Represent the detection process for monochromatic coherent light

- *Incoherent* light is emitted by independent atoms and therefore there is *no phase relationship among the emitted photons*. This property dictates *exponential distribution* for incoherent light (*if averaged over the Poisson distribution*)

$$P(z) = z_m^{\,z}/(1 + z_m)^{z+1}$$

- This is identical to the *Bose-Einstein distribution* which is used to describe the random statistics of light emitted in black body radiation (thermal light).





Photon-number signal-to-noise ratio

- The *photon-number signal-to-noise ratio*

$$SNR = \langle n \rangle^2/\sigma_n^2 = \langle n \rangle$$

  and the *minimum-detectable photon number*

$$\langle n \rangle = 1 \text{ photon}$$

- If the observation time T = 1 μs and the wavelength λ = 1.24 μm, this is equivalent to a *minimum detectable power* of <u>0.16 pW</u>. (e = 1.6×10^{-19} C)

- The *receiver sensitivity* for SNR = 10^3 (30 dB) is 1000 photons. If the time interval T = 10 ns, this is equivalent to a sensitivity of 10^{11} photons/s or an optical power sensitivity of <u>16 nW</u> at λ = 1.24 μm.

Photoelectron noise

- A photon incident on a photodetector of quantum efficiency η generates an electron-hole pair or liberates a photoelectron with probability η.
- An incident mean photon flux Φ (photons/s) therefore results in a *mean photoelectron flux* $\eta\Phi$.
- The number of photoelectrons m detected in the time interval T is a random variable with mean

$$\langle m \rangle = \eta\, \Phi T = \eta\, \langle n \rangle$$

- If the photon number follows the *Poisson probability distribution*, so is the photoelectron number.

- => the photoelectron-number variance $\sigma_m^2 = \langle m \rangle = \eta\, \langle n \rangle$

- ***Photoelectron-number signal-to-noise ratio***

$$\text{SNR} = \langle m \rangle = \eta \langle n \rangle$$

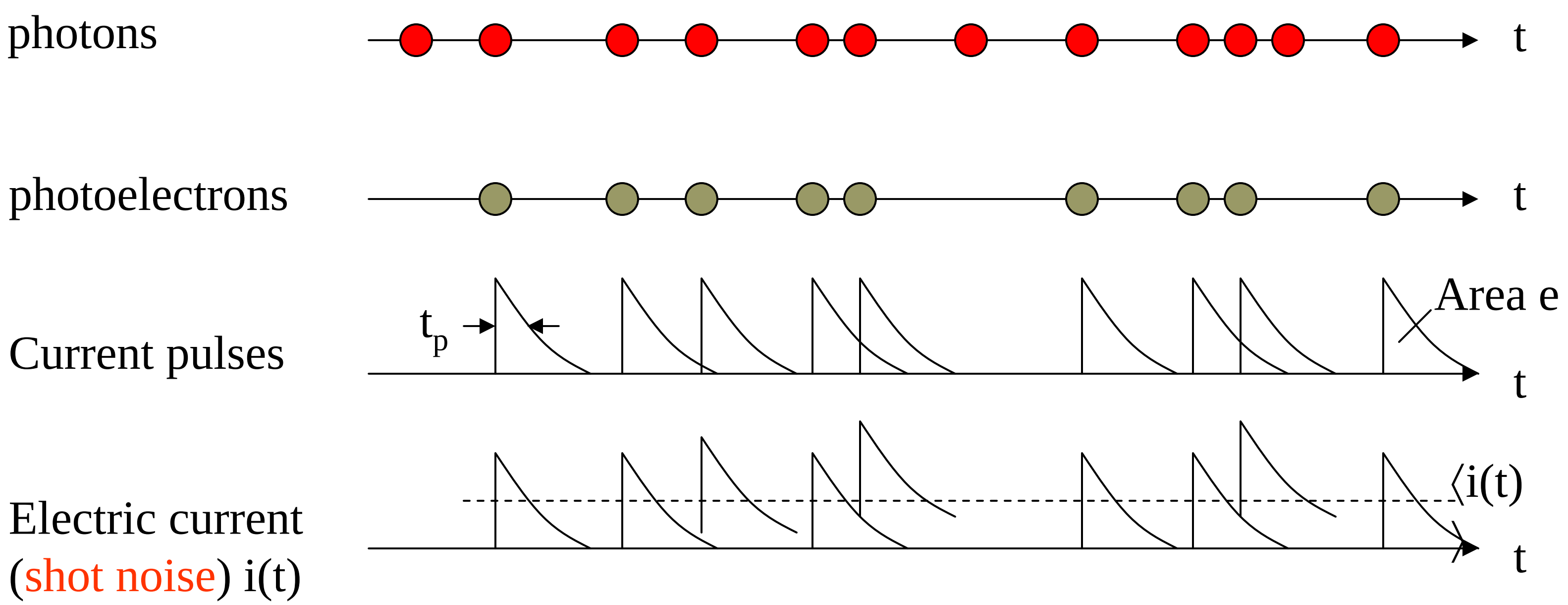
- The *minimum-detectable photoelectron number* is $\langle m \rangle = \eta \langle n \rangle =$ 1 photoelectron, corresponding to $1/\eta$ photons (i.e. > 1 photons).
- The *receiver sensitivity* for SNR = 10^3 is 1000 photoelectrons or $1000/\eta$ photons.

Photocurrent noise

- Here we examine the properties of the electric current i(t) induced in a circuit by a random photoelectron flux with mean $\eta\Phi$.
- We include the effects of *photon noise*, photoelectron *noise*, and the *characteristic time response of the detector and circuitry* (*filtering*).
- Assume every photoelectron-hole pair generates a pulse of electric current with charge (*area*) e and time duration τ_p in the external circuit of the photodetector.
- A photon stream incident on a photodetector therefore results in a stream of current pulses which add together to constitute the photocurrent i(t).

  => The randomness of the photon stream is transformed into a fluctuating electric current. *If the incident photons are Poisson distributed*, these fluctuations are known as ***shot noise***.

Shot noise



- The photocurrent induced in a photodetector circuit comprises a *superposition of current pulses*, each associated with a detected photon. The individual pulses illustrated are exponentially decaying step functions but they can assume an arbitrary shape.

- Consider a photon flux Φ incident on a photoelectric detector of quantum efficiency η.
- Let the random number m of photoelectrons counted within a *characteristic time interval* $T = 1/2B$ (the resolution time of the circuit) generate a photocurrent i(t), where t is the instant of time immediately following the interval T. (*The parameter B represents the bandwidth of the device/circuit system*.)
- For *rectangular* current pulses of duration T, the current and photoelectron-number random variables are related by $i = (e/T)\, m$.
- The *photocurrent mean* and *variance* are

$$\langle i \rangle = (e/T)\, \langle m \rangle$$

$$\sigma_i^2 = (e/T)^2\, \sigma_m^2$$

where $\langle m \rangle = \eta\Phi T = \eta\Phi/2B$ is the *mean number of photoelectrons collected* in the time interval $T = 1/2B$.

- Substituting $\sigma_m^2 = \langle m \rangle$ for the *Poisson* law yields the *photocurrent mean* and *variance*

$$\langle i \rangle = e\eta\Phi$$
$$\sigma_i^2 = 2e\,B\,\langle i \rangle$$

- => the *photocurrent SNR*

$$\mathrm{SNR} = \langle i \rangle^2/\sigma_i^2 = \langle i \rangle/2eB = \eta\Phi/2B = \langle m \rangle$$

- The current SNR is *directly proportional to the photon flux Φ* and *inversely proportional to the electrical bandwidth of the circuit B.*
- *The result is identical to that of the photoelectron-number SNR ratio* $\langle m \rangle$ as expected as the circuit introduces no added randomness.

- **e.g.** ***SNR and receiver sensitivity***. For $\langle i \rangle$ = 10 nA and the electrical bandwidth of the circuit B = 100 MHz, $\sigma_i \approx 0.57$ nA, corresponding to a SNR = 310 or 25 dB.

  => An *average* of 310 *photoelectrons* are detected in every time interval T = 1/2B = 5 ns.

  => The *minimum-detectable photon flux* for SNR = 1 is

  $\Phi = 2B/\eta$

  => The *receiver sensitivity* for SNR = 10^3 is

  $\Phi = 1000\ (2B/\eta) = 2 \times 10^{11}/\eta$ photons/s

Dark current noise

- *When there is no optical power* incident on the photodetector a *small reverse leakage current* still flows from the device terminals.
- Dark-current noise results from random electron-hole pairs generated *thermally* (or by tunneling).
- This dark current contributes to the *total system noise* and *gives random fluctuations about the average photocurrent*.

  => *It therefore manifests itself as shot noise* on the photocurrent.

- The ***dark current noise*** is

$$\sigma_d^2 = 2\, eB \langle I_d \rangle$$

Thermal noise

- Thermal noise (also called *Johnson noise* or *Nyquist noise*) results from *random thermal motions of the electrons in a conductor*. It is associated with the *blackbody radiation of a conductor* at the radio or microwave frequency range of the signal.
- Because *only materials that can absorb and dissipate energy* can emit blackbody radiation, thermal noise is generated *only by the resistive components* of the detector and its circuit. (*Capacitive and inductive components do not generate thermal noise because they neither dissipate nor emit energy*.)
- These motions give rise to a *random electric current even in the absence of an external electrical power source*. The thermal electric current in a resistance R is a random function i(t) whose mean value $\langle i(t) \rangle = 0$.

=> the variance of the current $\sigma_i^2 = \langle I_{th}^2 \rangle$

- In normal operation of most photodetectors, $f << k_B T/h$ = 6.24 THz at room temperature, the *frequency dependence of the thermal noise power is negligible*
- The *total thermal noise power* for a detection system of a bandwidth B is

$$P_{n,th} = 4k_B TB$$

- *For a resistor* that has a resistance R, the thermal noise can be treated as either *current noise* or *voltage noise*

$$P_{n,th} = \langle I_{th}^2 \rangle R = \langle v_{th}^2 \rangle / R$$

$$\Rightarrow \ \langle I_{th}^2 \rangle = 4k_B TB/R \quad \text{and} \quad \langle v_{th}^2 \rangle = 4k_B TBR$$

e.g. A 1-kΩ resistor at $T = 300^o$ K in a circuit bandwidth B = 100 MHz exhibits an RMS thermal noise current $\langle I_{th}^2 \rangle^{1/2} \approx 41$ nA.

- *For an optical detection system*, the resistance R is the *total equivalent resistance*, including the internal resistance of the detector and the load resistance R_L from the circuit, at the output of the detector.

- ***For a detector that has a current signal***, the thermal noise is determined by the *lowest shunt resistance to the detector*, which is often the load resistance of the detector.

  => *The thermal current noise* $\langle I_{th}^2 \rangle = 4k_B TB/R_L$ *can be reduced by increasing this resistance at the expense of reducing the response speed of the system*.

- ***For a detector that has a voltage signal***, the *thermal voltage noise* $\langle v_{th}^2 \rangle = 4k_B TBR_L$ can be reduced by *decreasing* this resistance, but *at the expense of reducing the output voltage signal*.

Signal-to-noise ratio

- The *total noise of a photodetector* is basically the sum of its *shot* noise and *thermal* noise:

$$\sigma_n^2 = \sigma_{sh}^2 + \sigma_{th}^2$$

- A photodetector is said to function in the *quantum regime* if $\sigma_{sh}^2 > \sigma_{th}^2$ (*shot-noise limited*)
- A photodetector is in the *thermal regime* if $\sigma_{th}^2 > \sigma_{sh}^2$ (*thermal-noise limited*)

$$\mathrm{SNR} = \langle I_p \rangle^2/\sigma_n^2 = \langle I_p \rangle^2/[2eB(\langle I_p \rangle + \langle I_d \rangle) + 4k_B TB/R_L]$$

$$= P^2 \mathcal{R}^2/[2eB(P\mathcal{R} + \langle I_d \rangle) + 4k_B TB/R_L]$$

where $\mathcal{R} = \eta e/h\upsilon$ is the *responsivity* of a photodetector.

Noise equivalent power

- The NEP of a photodetector is defined as the input power required of the optical signal for the signal-to-noise ratio to be unity, SNR = 1, at the detector output.
- The NEP for a photodetector that has an output current signal can be defined as

$$\text{NEP} = \langle i_n^2 \rangle^{1/2} / R$$

where $\langle i_n^2 \rangle$ is the *mean square noise current* at an input optical power level for SNR = 1 and *R* is the responsivity.

- For most detection systems at the *small input signal* level for SNR = 1, the *shot noise contributed by the input optical signal is negligible* compared to both the shot noise from other sources and the thermal noise of the detector.

- In this situation, the ***NEP of a photodetector***:

$$\mathrm{NEP} = (2e\langle i_d\rangle + 4k_BT/R_L)^{1/2}\, B^{1/2} / \mathcal{R}$$

- The NEP of a photodetector is often specified in terms of the NEP for a bandwidth of 1 Hz as $\mathrm{NEP}/B^{1/2}$, in the unit of W $\mathrm{Hz}^{-1/2}$.

- ***In order to reduce the RC time constant***, a high-speed photodetector that has a current signal normally has a *small area, thus a small dark current*, but requires a *small load resistance, thus a large thermal noise*.
- *=> the NEP of a high-speed photodetector is usually limited by the thermal noise from its external load resistance ($4k_BTB/R_L$) rather than by the shot noise from its internal dark current ($2eB\langle i_d\rangle$).*

E.g. A Si photodetector has an active area of $A = 5\ \mathrm{mm}^2$, a bandwidth of $B =$ 100 MHz, and a dark current of $i_d = 10$ nA. Find its shot-noise limited NEP, its thermal-noise-limited NEP, and its total NEP, all for a bandwidth of 1 Hz. Calculate the total NEP for its entire bandwidth.

- The shot noise from the dark current

$\langle i_{sh}^2 \rangle = 2eB \langle i_d \rangle = 2 \times 1.6 \times 10^{-19} \times 10 \times 10^{-9} \times B\ \mathrm{A}^2\ \mathrm{Hz}^{-1}$
$= 3.2 \times 10^{-27}\ B\ \mathrm{A}^2\ \mathrm{Hz}^{-1}$

- The thermal noise

$\langle i_{th}^2 \rangle = 4k_BTB/R_L = (4 \times 25.9 \times 10^{-3} \times 1.6 \times 10^{-19} / 50) \times B\ \mathrm{A}^2\ \mathrm{Hz}^{-1}$
$= 3.32 \times 10^{-22}\ B\ \mathrm{A}^2\ \mathrm{Hz}^{-1}$

- The total noise $\langle i_n^2 \rangle = \langle i_{sh}^2 \rangle + \langle i_{th}^2 \rangle = 3.32 \times 10^{-22}\ B\ \mathrm{A}^2\ \mathrm{Hz}^{-1}$, which is completely dominated by thermal noise.

- Suppose $R = 0.5\ \mathrm{AW}^{-1}$. The shot-noise-limited NEP for a bandwidth of 1 Hz:

$$(\mathrm{NEP})_{sh}/B^{1/2} = \langle i_{sh}^2 \rangle^{1/2}/(B^{1/2}R) = 113\ \mathrm{fW\ Hz}^{-1/2}$$

- The thermal-noise-limited NEP for a bandwidth of 1 Hz is

$$(\mathrm{NEP})_{th}/B^{1/2} = \langle i_{th}^{2}\rangle^{1/2}/(B^{1/2}R) = 36.4\ \mathrm{pW\ Hz}^{-1/2}$$

- The total NEP for a bandwidth of 1 Hz is

$$\mathrm{NEP}/B^{1/2} = 36.4\ \mathrm{pW\ Hz}^{-1/2}$$

- For B = 100 MHz, the total NEP for the entire bandwidth is

$$\mathrm{NEP} = 36.4 \times 10^{-12} \times (100 \times 10^{6})^{1/2}\ \mathrm{W} = 364\ \mathrm{nW}$$

This detector is completely limited by the thermal noise of its load resistance.