Lectures on Detector Techniques
Stanford Linear Accelerator Center
September 1998 – February, 1999

Electronics I - Devices and Noise

Helmuth Spieler

Physics Division
Lawrence Berkeley National Laboratory

copies of transparencies in pdf format at
http://www-physics.lbl.gov/~spieler

for more details see UC Berkeley Physics 198 course notes
at http://www-physics.lbl.gov/~spieler/physics_198_notes

Electronic Noise

Two basic noise mechanisms: input noise current i_n
input noise voltage v_n

Equivalent Noise Charge:

$$Q_n^2 = i_n^2 \, T_s \, F_i \; + \; C_i^2 \, v_n^2 \, \frac{F_v}{T_s}$$

where T_s Characteristic shaping time (*e.g.* peaking time)

F_i , F_v "Form Factors" that are determined by the shape of the pulse.

C_i Total capacitance at the input node (detector capacitance + input capacitance of preamplifier + stray capacitance + …)

Typical values of F_i , F_v

CR-RC shaper	F_i = 0.924	F_v = 0.924
CR-$(RC)^4$ shaper	F_i = 0.45	F_v = 1.02
CR-$(RC)^7$ shaper	F_i = 0.34	F_v = 1.27
CAFE chip	F_i = 0.4	F_v = 1.2

In previous lecture noise was analyzed by determining frequency dependence of noise contributions and integrating over the frequency response of the shaper.

Not possible for all shapers !
(notably those involving switching – time variant shapers)

Noise Analysis in the Time Domain

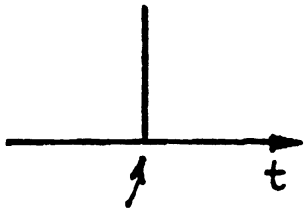
What pulse shapes have a frequency spectrum corresponding to typical noise sources?

1. Voltage Noise

   The frequency spectrum at the input of the detector system is "white", i.e.

$$\frac{dA}{df} = \text{const.}$$

   This is the spectrum (Fourier transform) of a δ impulse:



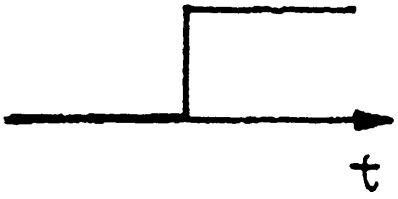
inifinitesimally narrow, but area = 1.

2. Current Noise

   The spectral density is inversely proportional to frequency, i.e.

$$v_{nb} = i_{nb} \frac{1}{\omega C_D} \qquad \Rightarrow \qquad \frac{dA}{df} \propto \frac{1}{f}$$

   This is the spectrum of a step impulse:



- Input noise can be considered as a sequence of δ and step pulses whose rate determines the noise level.

- The shape of the primary noise pulses is modified by the pulse shaper:

  δ pulses become longer,

  step pulses are shortened.

- The noise level at a given measurement time T_m is determined by the cumulative effect (superposition) of all noise pulses occurring prior to T_m.

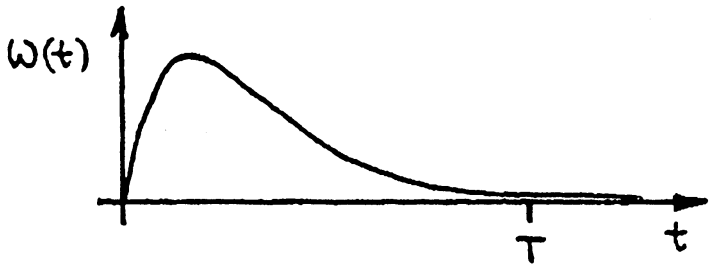
- Their individual contributions at $t = T_m$ are described by the shaper's "weighting function" $W(t)$.

References:

V. Radeka, Nucl. Instr. and Meth. **99** (1972) 525
V. Radeka, IEEE Trans. Nucl. Sci. **NS-21** (1974) 51
F.S. Goulding, Nucl. Instr. and Meth. **100** (1972) 493
F.S. Goulding, IEEE Trans. Nucl. Sci. **NS-29** (1982) 1125

Consider a single noise pulse occurring in a short time interval dt at a time T prior to the measurement. The amplitude at $t = T$ is

$$a_n = W(T)$$



If, on the average, $n_n\,dt$ noise pulses occur within dt, the fluctuation of their cumulative signal level at $t = T$ is proportional to

$$\sqrt{n_n dt}$$

The magnitude of the baseline fluctuation is

$$\sigma_n^2(T) \propto n_n \left[W(t)\right]^2 dt$$

For all noise pulses occurring prior to the measurement

$$\sigma_n^2 \propto n_n \int_0^\infty \left[W(t)\right]^2 dt$$

where

n_n determines the magnitude of the noise

and

$\int_0^\infty \left[W(t)\right]^2 dt$ describes the noise characteristics of the shaper – the “noise index”

The Weighting Function

a) current noise — $W_i(t)$ is the shaper response to a step pulse, i.e. the "normal" output waveform.

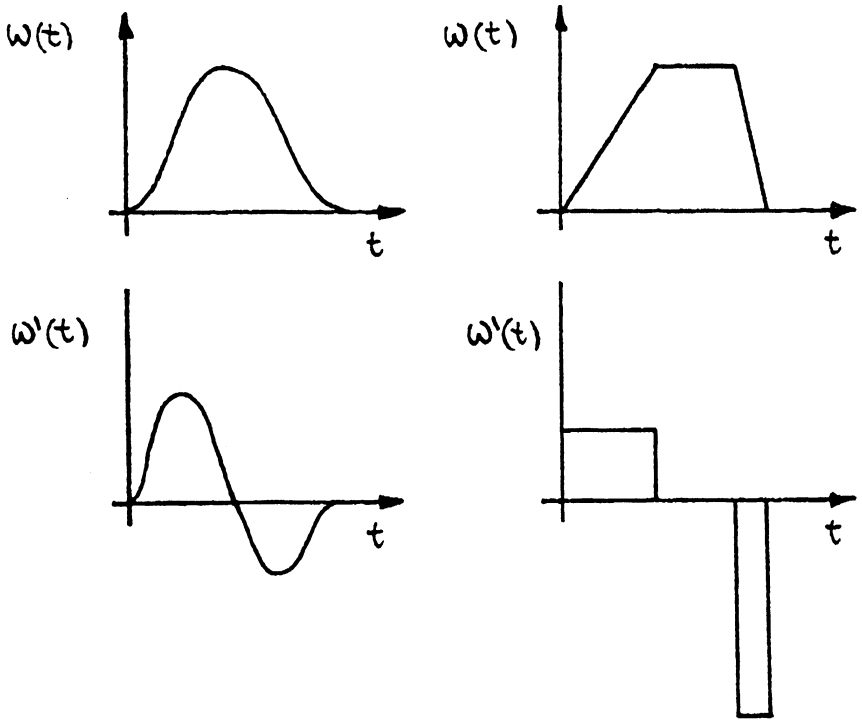
b) voltage noise — $$W_v(t) = \frac{d}{dt} W_i(t) \equiv W'(t)$$

(Consider a δ pulse as the superposition of two step pulses of opposite polarity and spaced inifinitesimally in time)

Examples: 1. Gaussian 2. Trapezoid

current ("step") noise

voltage ("delta") noise



Goal: Minimize overall area

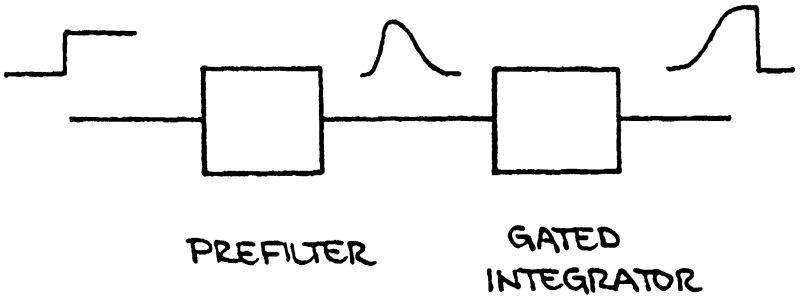
⇒ For a given pulse duration a symmetrical pulse provides the best noise performance.

Time-Variant Shapers

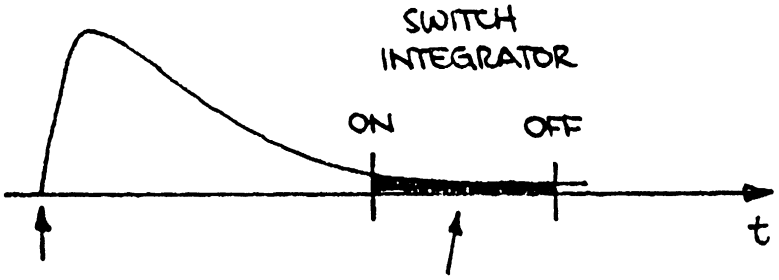
Example: gated integrator with prefilter

The gated integrator integrates the input signal during a selectable time interval (the “gate”).

In this example, the integrator is switched on prior to the signal pulse and switched off after a fixed time interval, selected to allow the output signal to reach its maximum.



Consider a noise pulse occurring prior to the “on time” of the integrator.



occurrence of the noise pulse

contribution of noise pulse to integrator output

For W_1 = weighting function of the time-invariant prefilter

W_2 = weighting function of the time-variant stage

the overall weighting function is obtained by convolution

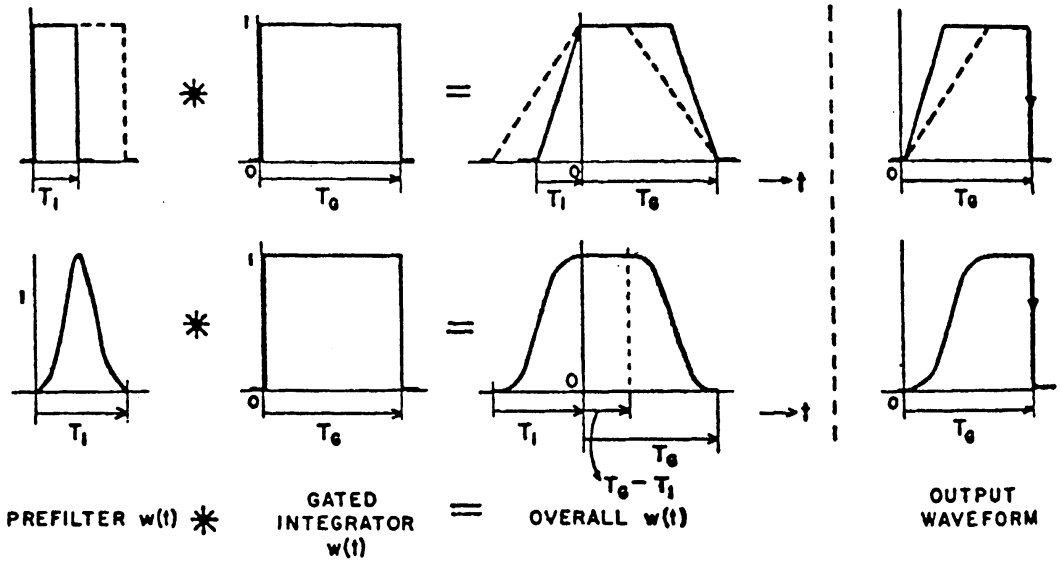
$$W(t) = \int_{-\infty}^{\infty} W_2(t') \cdot W_1(t - t')\, dt'$$

Weighting function for current ("step") noise: $W(t)$

Weighting function for voltage ("delta") noise: $W'(t)$

Example

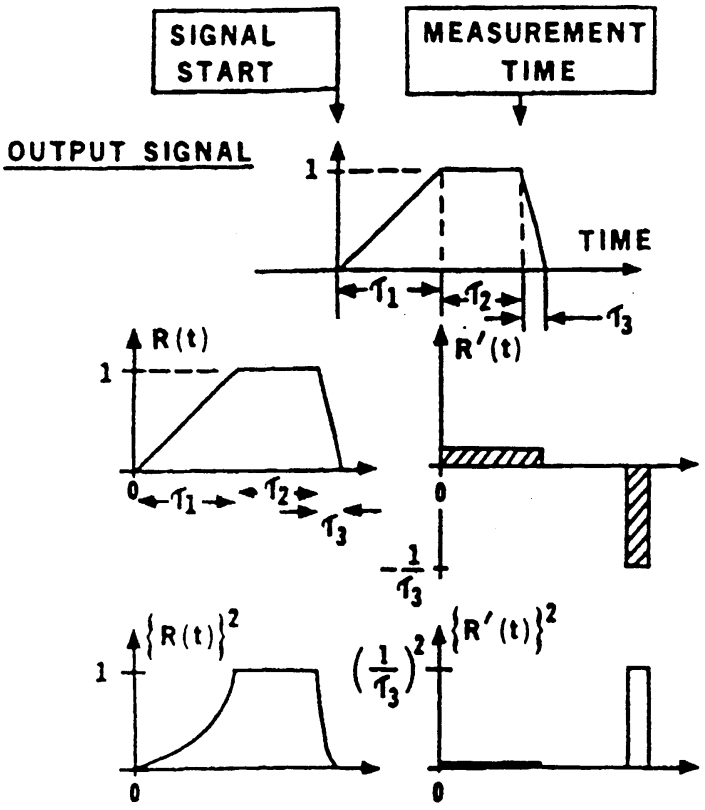
Time-invariant prefilter feeding a gated integrator
(from Radeka, IEEE Trans. Nucl. Sci. **NS-19** (1972) 412)



Comparison between a time-invariant and time-variant shaper
(from Goulding, NIM **100** (1972) 397)

Example: trapezoidal shaper Duration= 2 μs
Flat top= 0.2 μs

1. Time-Invariant Trapezoid



Current noise

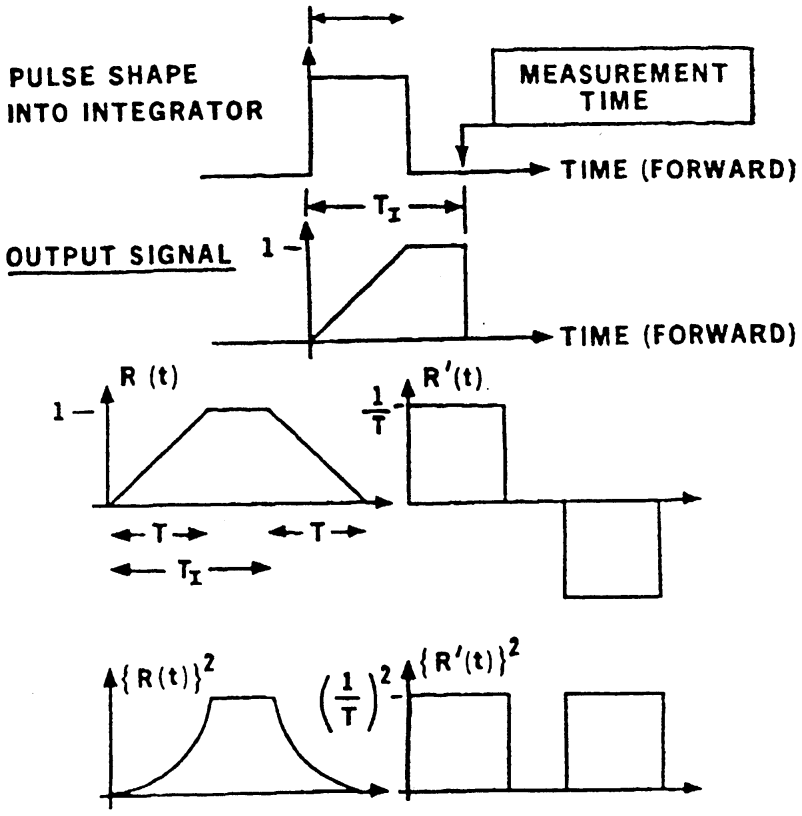
$$N_i^2 = \int_0^{\infty} [W(t)]^2 dt = \int_0^{\tau_1} \left(\frac{t}{\tau_1}\right)^2 dt + \int_{\tau_1}^{\tau_2} (1)^2 dt + \int_{\tau_2}^{\tau_3} \left(\frac{t}{\tau_3}\right)^2 dt = \tau_2 + \frac{\tau_1 + \tau_3}{3}$$

Voltage noise

$$N_v^2 = \int_0^{\infty} [W'(t)]^2 dt = \int_0^{\tau_1} \left(\frac{1}{\tau_1}\right)^2 dt + \int_{\tau_2}^{\tau_3} \left(\frac{1}{\tau_3}\right)^2 dt + = = \frac{1}{\tau_1} + \frac{1}{\tau_3}$$

Minimum for $\tau_1 = \tau_3$ (symmetry!) $\Rightarrow$ $N_i^2 = 0.8$, $N_v^2 = 2.2$

Gated Integrator Trapezoidal Shaper



Current Noise

$$N_i^2 = 2\int_0^T \left(\frac{t}{T}\right)^2 dt + \int_T^{T_I - T} (1)^2 dt = T_I - \frac{T}{3}$$

Voltage Noise

$$N_v^2 = 2\int_0^T \left(\frac{1}{T}\right)^2 dt = \frac{2}{T}$$

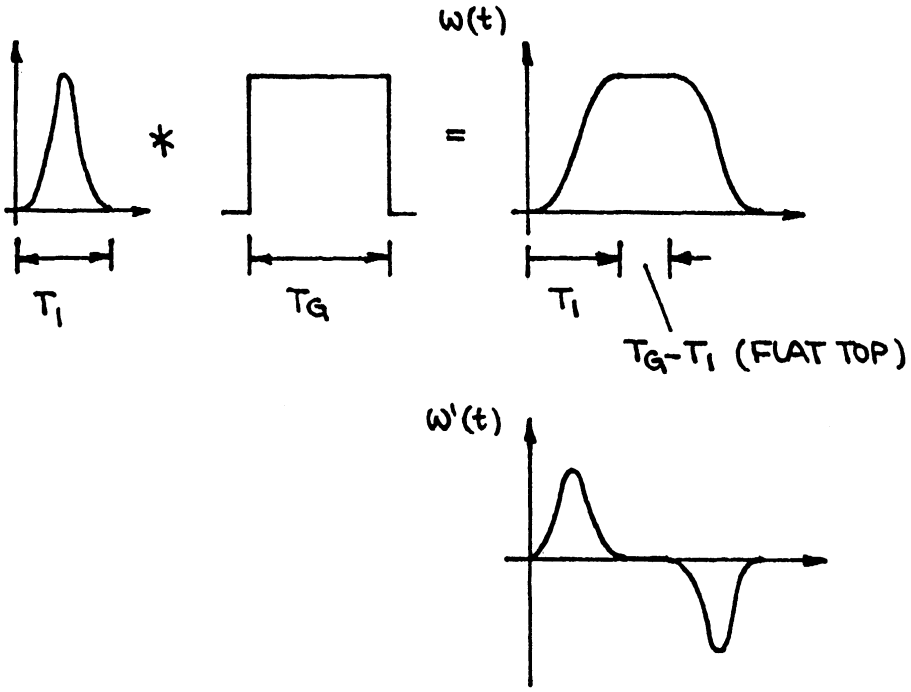
$\Rightarrow$ time-variant shaper N_i^2= 1.4, N_v^2= 1.1

time-invariant shaper N_i^2= 0.8, N_v^2= 2.2

time-variant trapezoid has more current noise, less voltage noise

Interpretation of Results

Example: gated integrator



Current Noise

$$Q_{ni}^2 \propto \int [W(t)]^2 \, dt$$

Increases with T_1 and T_G (i.e. width of $W(t)$)

(more noise pulses accumulate within width of $W(t)$)

Voltage Noise

$$Q_{nv}^2 \propto \int [W'(t)]^2 \, dt$$

Increases with the magnitude of the derivative of $W(t)$

(steep slopes $\rightarrow$ large bandwidth — *determined by prefilter*)

Width of flat top irrelevant

(δ response of prefilter is bipolar: net= 0)

Quantitative Assessment of Noise in the Time Domain

(see Radeka, IEEE Trans. Nucl. Sci. **NS-21** (1974) 51)

$$Q_n^2 = \frac{1}{2}\, i_n^2 \int_{-\infty}^{\infty} [W(t)]^2 dt + \frac{1}{2}\, C_i^2\, v_n^2 \int_{-\infty}^{\infty} [W'(t)]^2 dt$$

current noise — voltage noise

Q_n= equivalent noise charge [C]

i_n= input current noise spectral density [A/√Hz]

v_n= input voltage noise spectral density [V/√Hz]

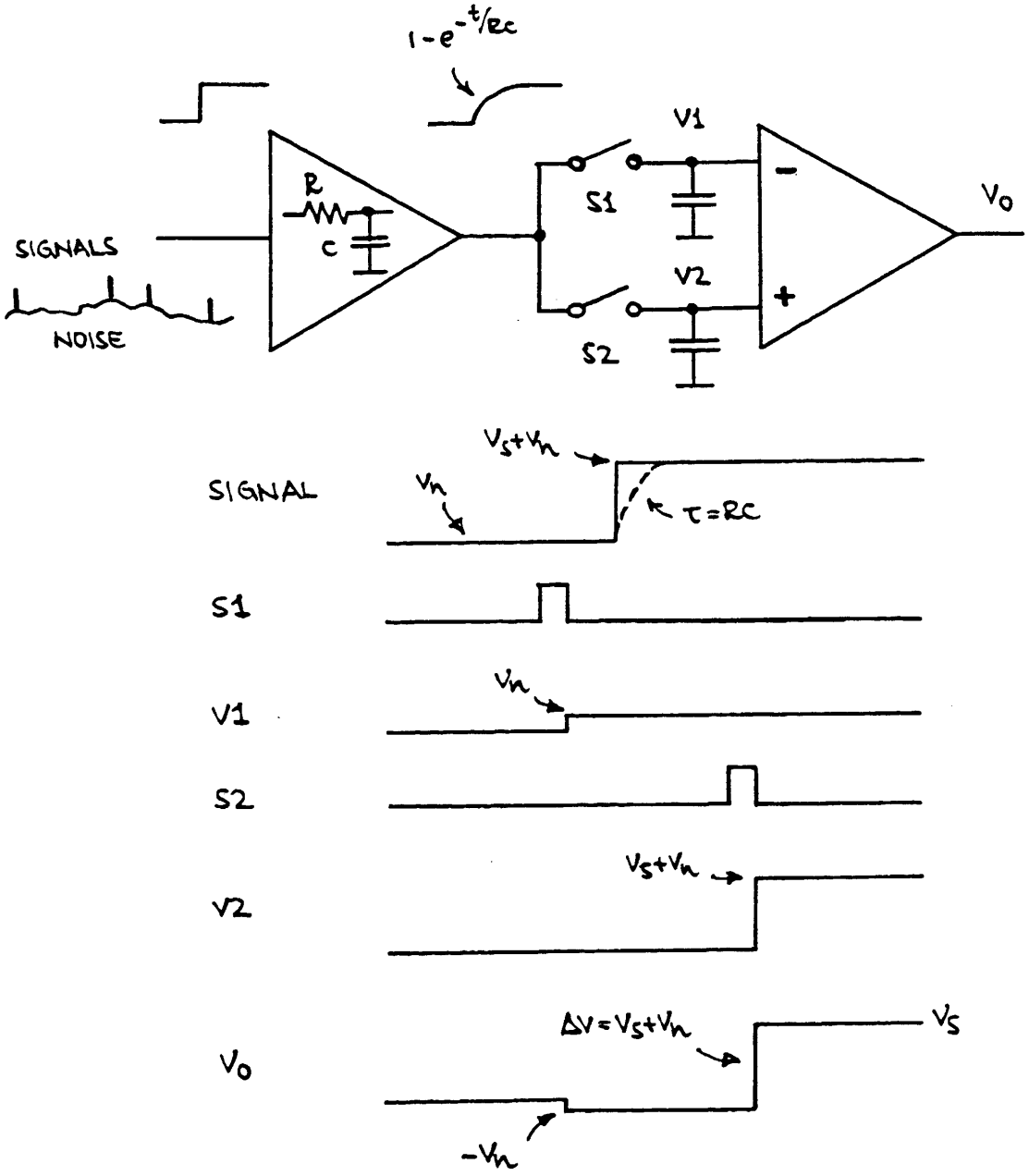
C_i= total capacitance at input

$W(t)$ normalized to unit input step response

or rewritten in terms of a characteristic time $t \rightarrow T/t$

$$Q_n^2 = \frac{1}{2}\, i_n^2\, T \int_{-\infty}^{\infty} [W(t)]^2 dt + \frac{1}{2}\, C_i^2\, v_n^2\, \frac{1}{T} \int_{-\infty}^{\infty} [W'(t)]^2 dt$$

Correlated Double Sampling



1. Signals are superimposed on a (slowly) fluctuating baseline
2. To remove baseline fluctuations the baseline is sampled prior to the arrival of a signal.
3. Next, the signal + baseline is sampled and the previous baseline sample subtracted to obtain the signal

1. Current Noise

Current (shot) noise contribution:

$$Q_{ni}^2 = \frac{1}{2} i_n^2 \int_{-\infty}^{\infty} [W(t)]^2 \, dt$$

Weighting function (T= time between samples):

$$\begin{aligned} t < 0: &\qquad W(t) = 0 \\ 0 \le t \le T: &\quad W(t) = 1 - e^{-t/\tau} \\ t > T: &\qquad W(t) = e^{-(t-T)/\tau} \end{aligned}$$

Current noise coefficient

$$F_i = \int_{-\infty}^{\infty} [W(t)]^2 \, dt$$

$$F_i = \int_0^T \left(1 - e^{-t/\tau}\right)^2 dt + \int_T^{\infty} e^{-2(t-T)/\tau} \, dt$$

$$F_i = \left(T + \frac{\tau}{2} e^{-T/\tau} - \frac{\tau}{2} e^{-2T/\tau} \right) + \frac{\tau}{2}$$

so that the equivalent noise charge

$$Q_{ni}^2 = \frac{1}{2} i_n^2 \left[T + \frac{\tau}{2} \left(e^{-T/\tau} - e^{-2T/\tau} + 1 \right) \right]$$

$$Q_{ni}^2 = i_n^2 \; \tau \;\; \frac{1}{4} \left(\frac{2T}{\tau} + e^{-T/\tau} - e^{-2T/\tau} + 1 \right)$$

Reality Check 1:

Assume that the current noise is pure shot noise

$$i_n^2 = 2q_e I$$

so that

$$Q_{ni}^2 = q_e I \tau \frac{1}{2}\left(\frac{2T}{\tau} + e^{-T/\tau} - e^{-2T/\tau} + 1\right)$$

Consider the limit Sampling Interval >> Rise Time, $T >> \tau$:

$$Q_{ni}^2 \approx q_e I \cdot T$$

or expressed in electrons

$$Q_{ni}^2 \approx \frac{q_e I \cdot T}{q_e^2} = \frac{I \cdot T}{q_e}$$

$$Q_{ni} \approx \sqrt{N_i}$$

where N_i is the number of electrons “counted” during the sampling interval T.

2. Voltage Noise

Voltage Noise Contribution

$$Q_{nv}^2 = \frac{1}{2} C_i^2 v_n^2 \int_{-\infty}^{\infty} [W'(t)]^2 \, dt$$

Voltage Noise Coefficient

$$F_v = \int_{-\infty}^{\infty} [W'(t)]^2 \, dt$$

$$F_v = \int_0^T \left(\frac{1}{\tau} e^{-t/\tau} \right)^2 dt + \int_T^{\infty} \left(\frac{1}{\tau} e^{-2(t-T)/\tau} \right)^2 dt$$

$$F_v = \frac{1}{2\tau} \left(1 - e^{-2T/\tau} \right) + \frac{1}{2\tau}$$

$$F_v = \frac{1}{2\tau} \left(2 - e^{-2T/\tau} \right)$$

so that the equivalent noise charge

$$Q_{nv}^2 = C_i^2 \; v_n^2 \; \frac{1}{\tau} \; \frac{1}{4} \left(2 - e^{-2T/\tau} \right)$$

Reality Check 2:

In the limit $T >> \tau$:

$$Q_{nv}^2 = C_i^2 \cdot v_n^2 \cdot \frac{1}{2\tau}$$

Compare this with the noise on an RC low-pass filter alone (i.e. the voltage noise at the output of the pre-filter):

$$Q_n^2(RC) = C_i^2 \cdot v_n^2 \cdot \frac{1}{4\tau}$$

(see the discussion on noise bandwidth)

so that

$$\frac{Q_n(\text{double sample})}{Q_n(RC)} = \sqrt{2}$$

If the sample time is sufficiently large, the noise samples taken at the two sample times are uncorrelated, so the two samples simply add in quadrature.

3. Signal Response

The preceding calculations are only valid for a signal response of unity, which is valid at $T >> \tau$.

For sampling times T of order τ or smaller one must correct for the reduction in signal amplitude at the output of the prefilter

$$V_s / V_i = 1 - e^{-T/\tau}$$

so that the equivalent noise charge due to the current noise becomes

$$Q_{ni}^2 = i_n^2 \ \tau \ \frac{\frac{2T}{\tau} + e^{-T/\tau} - e^{-2T/\tau} + 1}{4\left(1 - e^{-T/\tau}\right)^2}$$

The voltage noise contribution is

$$Q_{nv}^2 = C_i^2 \ v_n^2 \ \frac{1}{\tau} \ \frac{2 - e^{-2T/\tau}}{4\,(1 - e^{-T/\tau})^2}$$

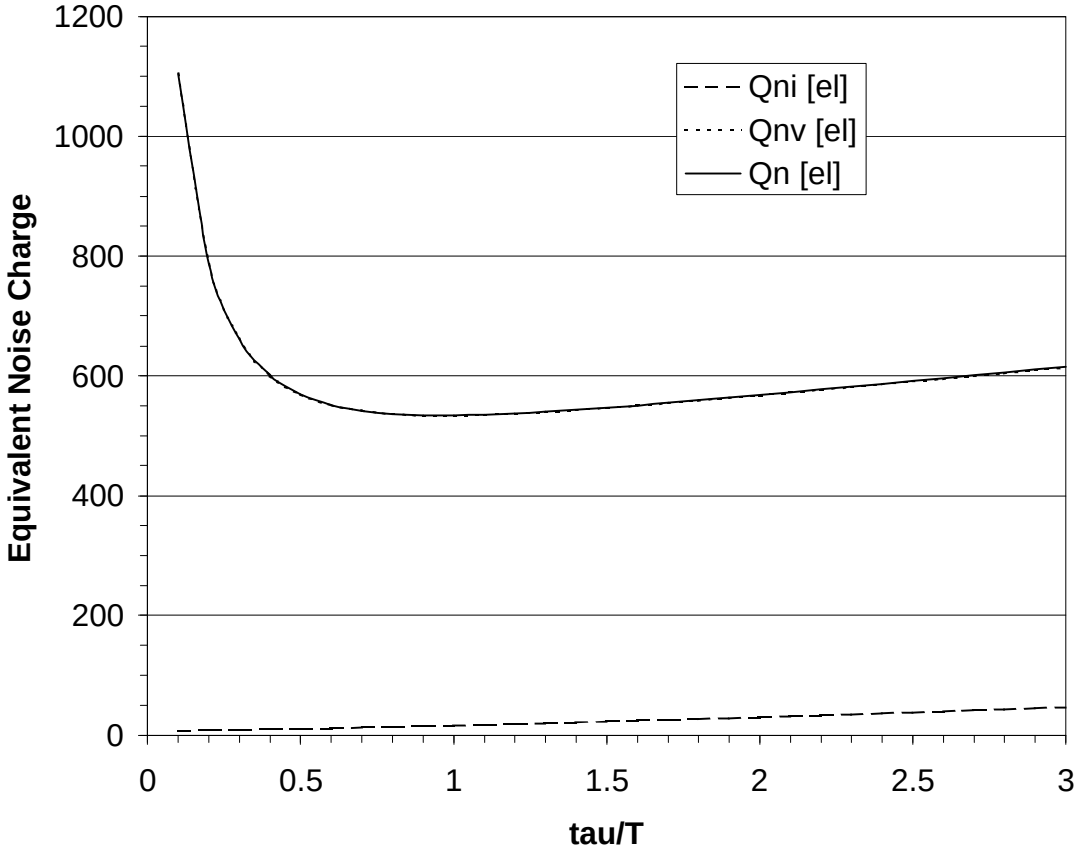
and the total equivalent noise charge

$$Q_n = \sqrt{Q_{ni}^2 + Q_{nv}^2}$$

Optimization

1. Noise current negligible

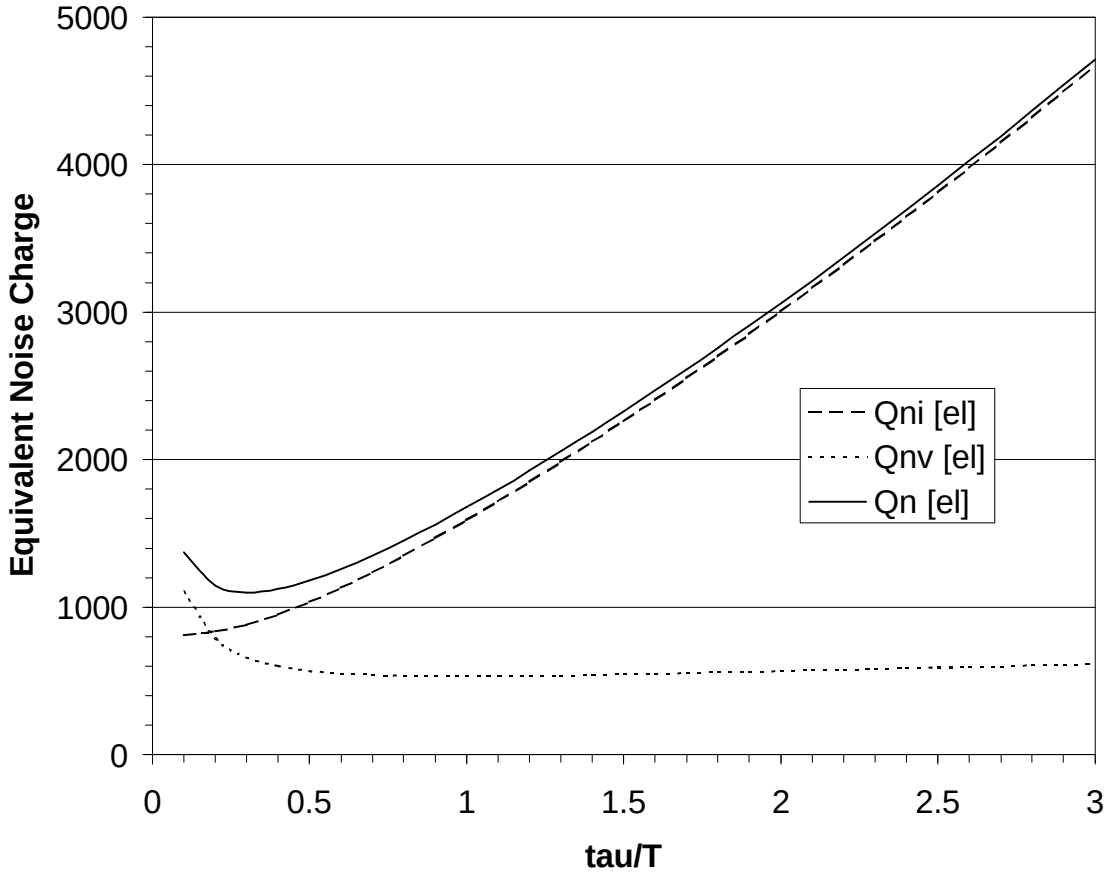
Parameters: T= 100 ns
C_d= 10 pF
v_n= 2.5 nV/√Hz
→ i_n= 6 fA/√Hz (I_b= 0.1 nA)



Noise attains shallow minimum for τ = T .

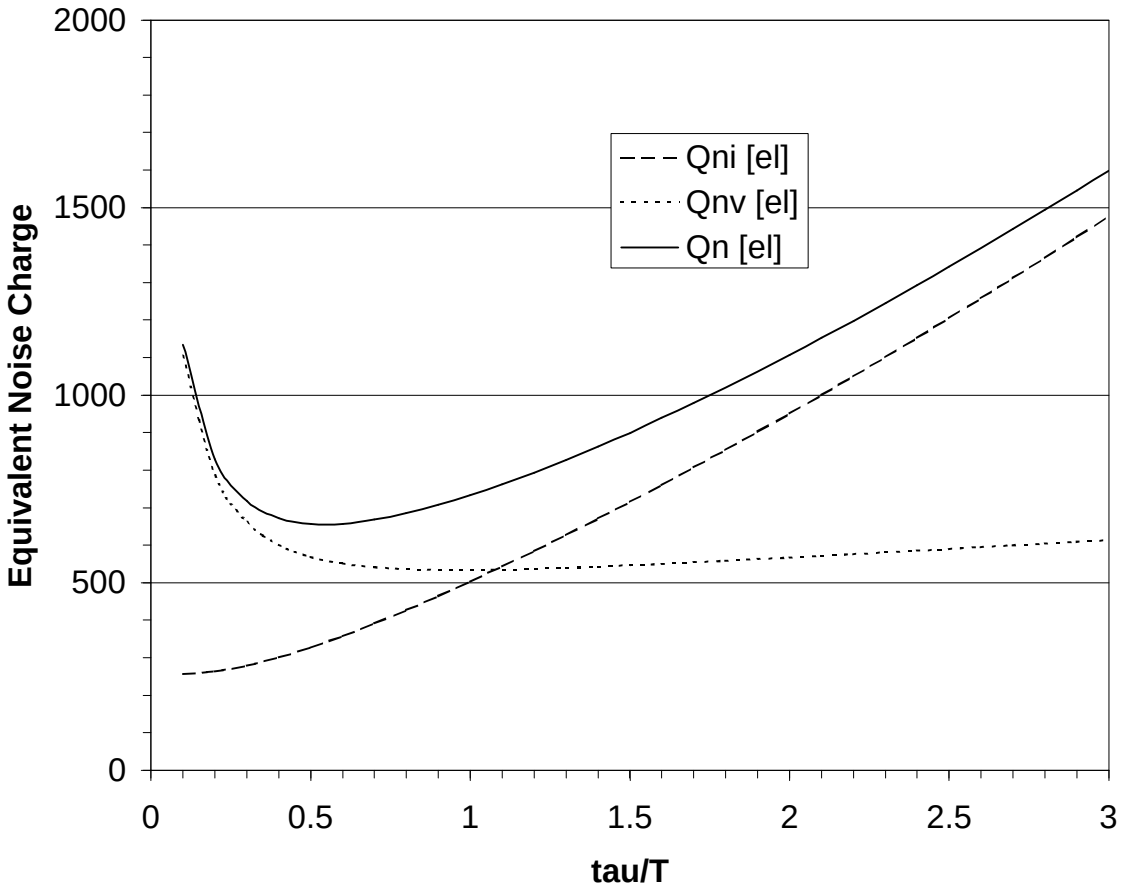
2. Significant current noise contribution

Parameters: T= 100 ns
C_d= 10 pF
v_n= 2.5 nV/√Hz
→ i_n= 0.6 pA/√Hz (I_b= 1 μA)



Noise attains minimum for τ = 0.3 T .

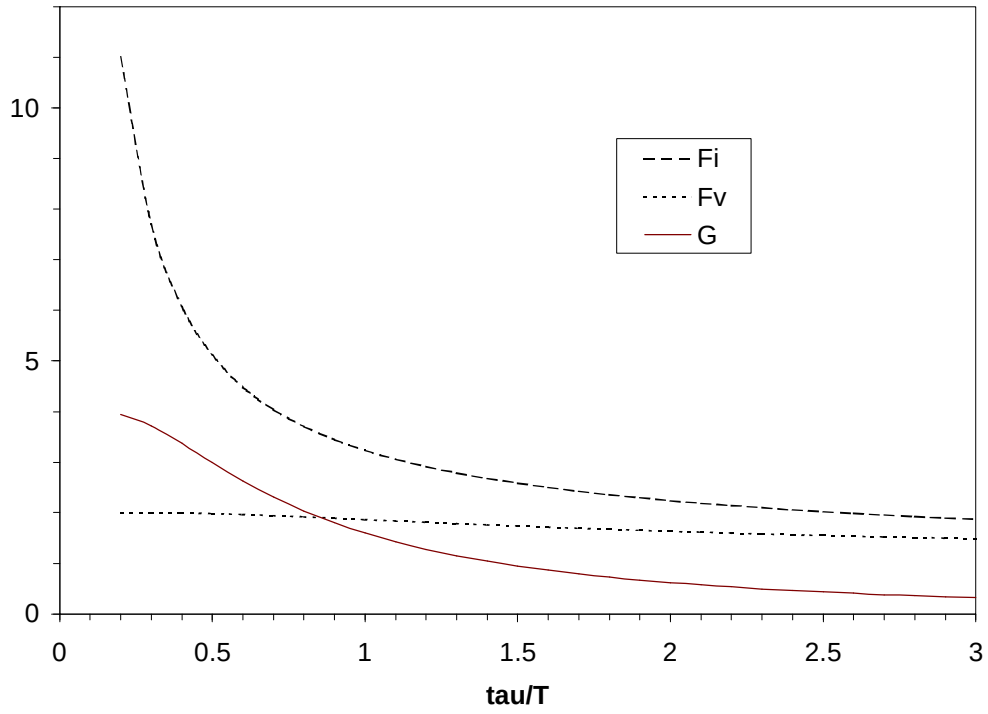
Parameters: T= 100 ns
C_d= 10 pF
v_n= 2.5 nV/√Hz
→ i_n= 0.2 pA/√Hz (I_b= 100 nA)



Noise attains minimum for τ = 0.5 T .

3. Form Factors F_i, F_v and Signal Gain G vs. τ / T

Note: In this plot the form factors F_i, F_v are not yet corrected by the gain G.



The voltage noise coefficient is practically independent of τ / T .

Voltage contribution to noise charge dominated by C_i/τ .

The current noise coefficient increases rapidly at small τ / T .

At large τ / T the contribution to the noise charge increases because of τ dependence.

The gain dependence increases the equivalent noise charge with increasing τ / T (as the gain is in the denominator).

Summary

Two basic noise mechanisms: input noise current i_n
input noise voltage v_n

Equivalent Noise Charge:

$$Q_n^2 = i_n^2 \; T_s \, F_i \; + \; C_i^2 \, v_n^2 \; \frac{F_v}{T_s}$$

front end, shaper, front end, shaper

where T_s Characteristic shaping time (*e.g.* peaking time)

F_i, F_v "Form Factors" that are determined by the shape of the pulse.

They can be calculated in the frequency or time domain.

C_i Total capacitance at the input node (detector capacitance + input capacitance of preamplifier + stray capacitance + …)

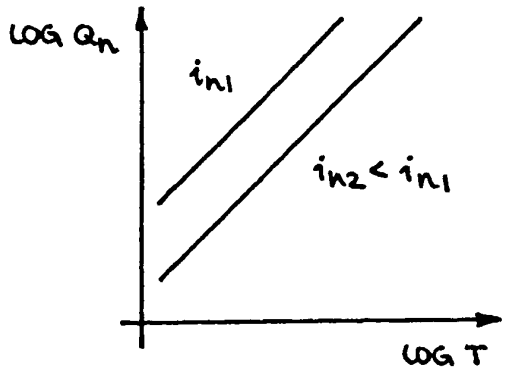
- Current noise contribution increases with T
- Voltage noise contribution decreases with increasing T

only for “white” voltage & current noise sources + capacitive load

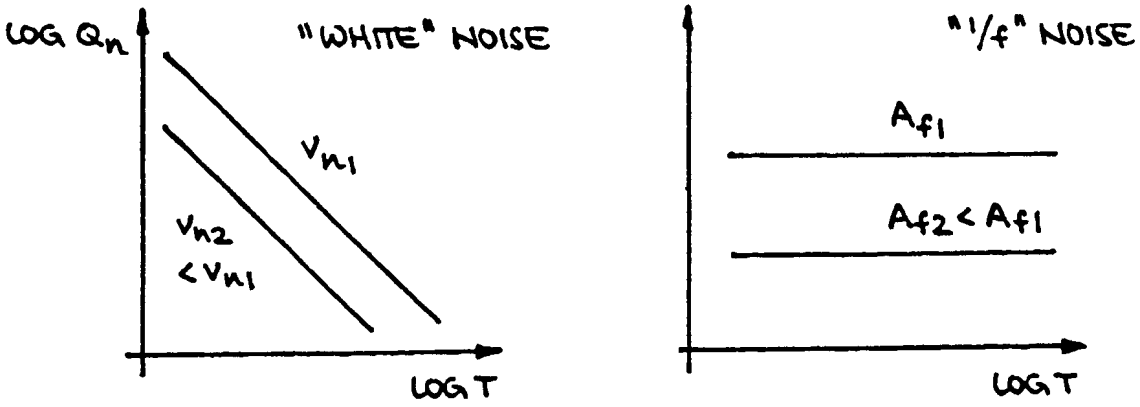
“$1/f$” voltage noise contribution constant in T

1. Equivalent Noise Charge vs. Pulse Width

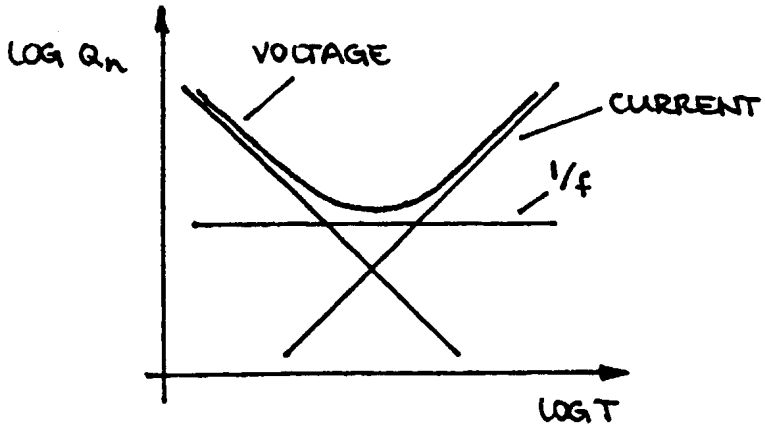
Current Noise vs. T



Voltage Noise vs. T



Total Equivalent Noise Charge



2. Equivalent Noise Charge vs. Detector Capacitance ($C_i = C_d + C_a$)

$$Q_n = \sqrt{i_n^2 F_i T + (C_d + C_a)^2 v_n^2 F_v \frac{1}{T}}$$

$$\frac{dQ_n}{dC_d} = \frac{2C_d v_n^2 F_v \frac{1}{T}}{\sqrt{i_n^2 F_i T + (C_d + C_a)^2 v_n^2 F_v \frac{1}{T}}}$$

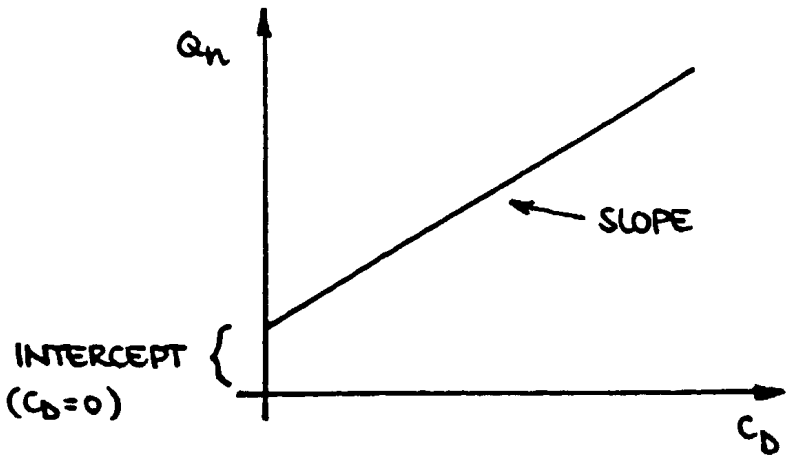
If current noise $i_n^2 F_i T$ is negligible

$$\frac{dQ_n}{dC_d} \approx 2v_n \cdot \sqrt{\frac{F_v}{T}}$$

↑ input stage ↑ shaper

Zero intercept

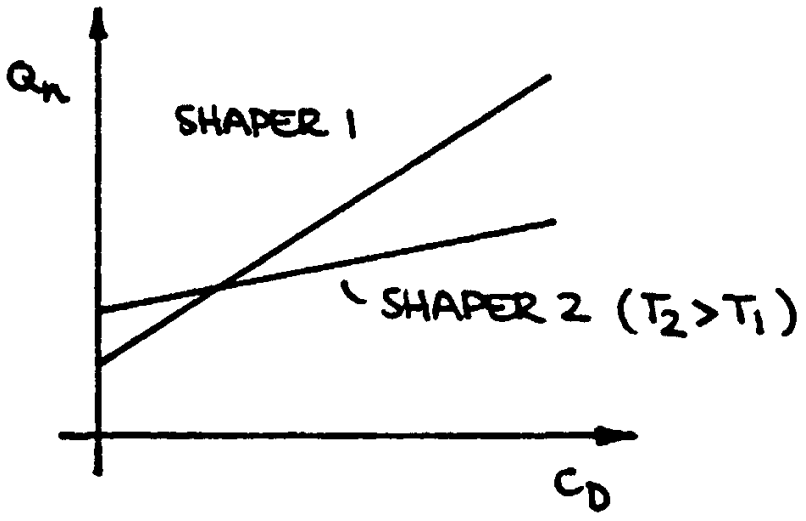
$$Q_n\big|_{C_d=0} = C_a v_n \sqrt{F_v / T}$$



Noise slope is a convenient measure to compare preamplifiers and predict noise over a range of capacitance.

Caution: both noise slope and zero intercept depend on *both the preamplifier and the shaper*

Same preamplifier, but different shapers:

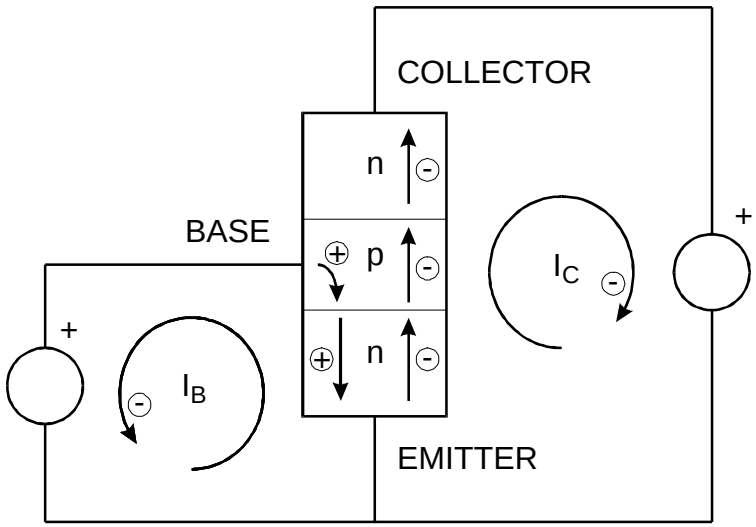


Caution: Current noise contribution may be negligible at high detector capacitance, but not for $C_d=0$.

$$Q_n\big|_{C_d=0} = \sqrt{i_n^2 F_i T + C_a^2 v_n^2 F_v / T}$$

Bipolar Transistors

Consider the *npn* structure shown below.



The base and emitter form a diode, which is forward biased so that a base current I_B flows.

The base current injects holes into the base-emitter junction.

As in a simple diode, this gives rise to a corresponding electron current through the base-emitter junction.

If the potential applied to the collector is sufficiently positive so that the electrons passing from the emitter to the base are driven towards the collector, an external current I_C will flow in the collector circuit.

The ratio of collector to base current is equal to the ratio of electron to hole currents traversing the base-emitter junction. In an ideal diode

$$\frac{I_C}{I_B} = \frac{I_{nBE}}{I_{pBE}} = \frac{D_n / N_A L_n}{D_p / N_D L_p} = \frac{N_D}{N_A}\frac{D_n L_p}{D_p L_n}$$

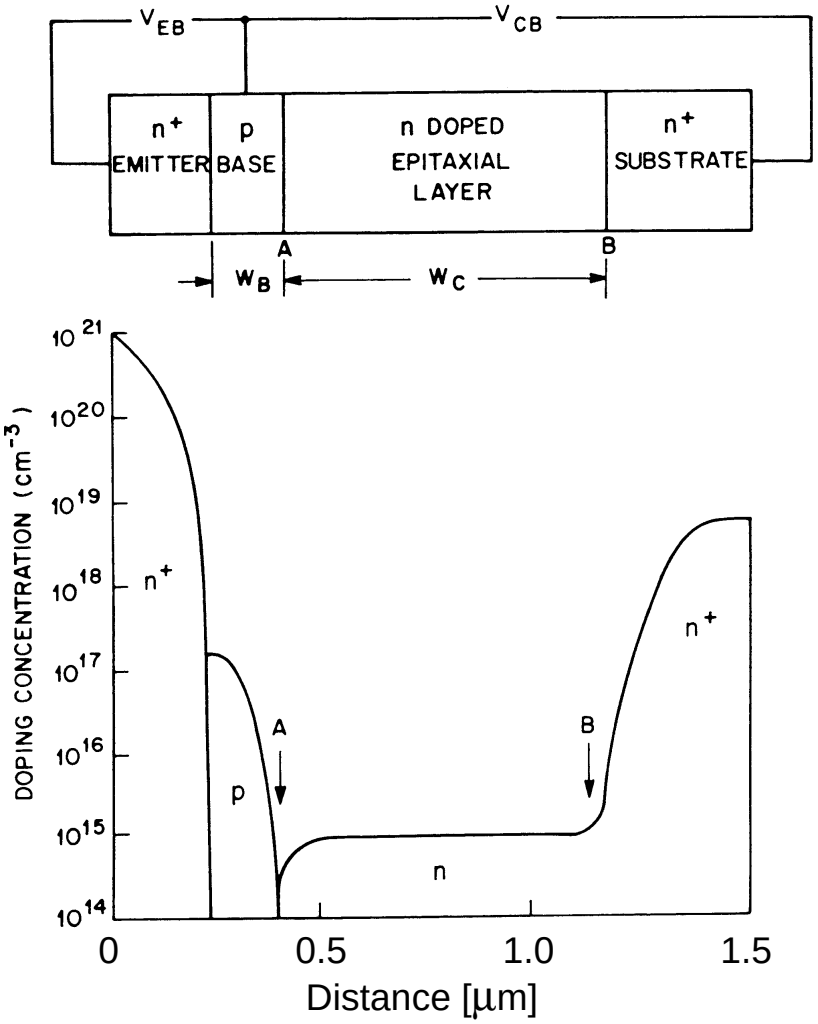
If the ratio of doping concentrations in the emitter and base regions N_D/N_A is sufficiently large, the collector current will be greater than the base current.

$\Rightarrow$ DC current gain

Furthermore, we expect the collector current to saturate when the collector voltage becomes large enough to capture all of the minority carrier electrons injected into the base.

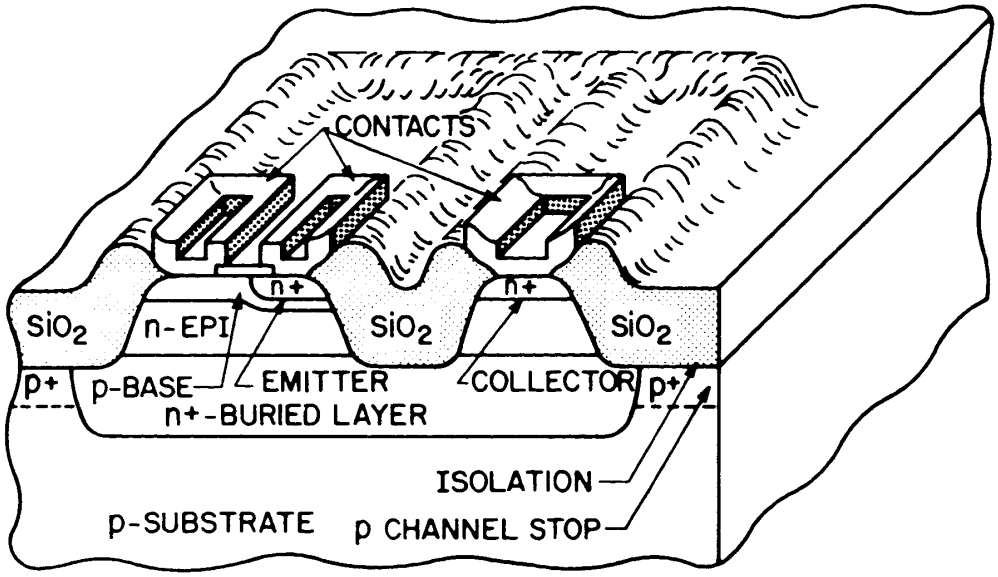
Since the current inside the transistor comprises both electrons and holes, the device is called a bipolar transistor.

Dimensions and doping levels of a modern high-frequency transistor (5 – 10 GHz bandwidth)



(adapted from Sze)

High-speed bipolar transistors are implemented as vertical structures.



(from Sze)

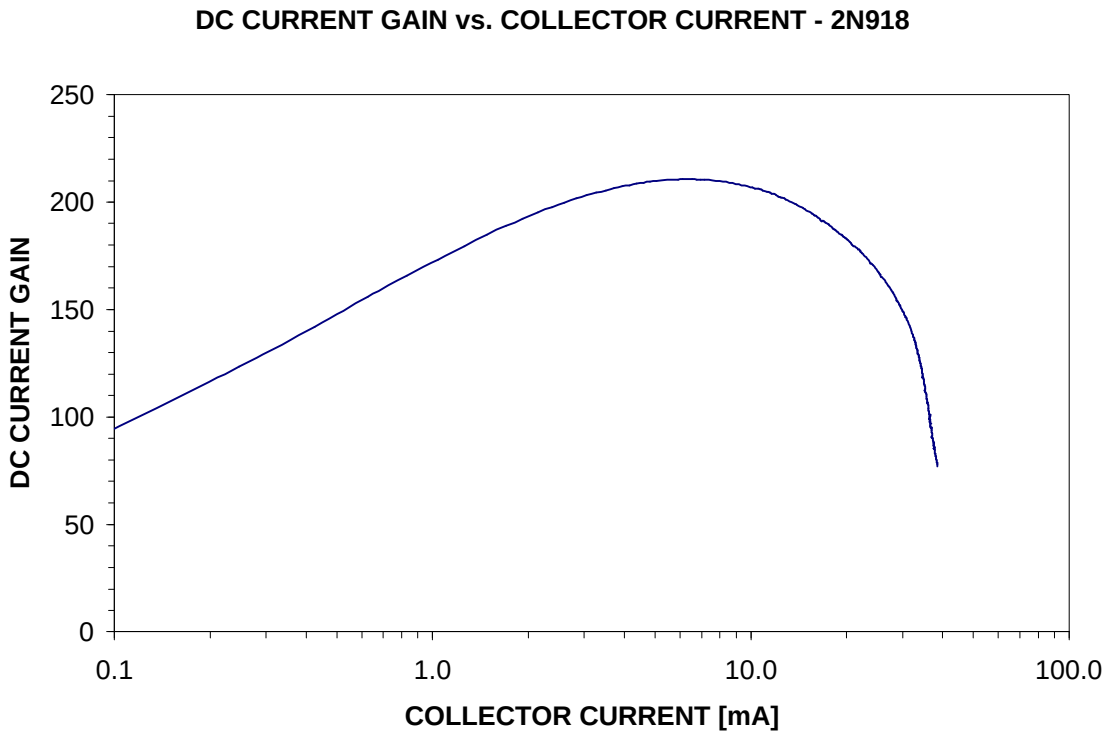
The base width, typically 0.2 μm or less in modern high-speed transistors, is determined by the difference in diffusion depths of the emitter and base regions.

The thin base geometry and high doping levels make the base-emitter junction sensitive to large reverse voltages.

Typically, base-emitter breakdown voltages for high-frequency transistors are but a few volts.

As shown in the preceding figure, the collector region is usually implemented as two regions: one with low doping (denoted “epitaxial layer in the figure) and the other closest to the collector contact with a high doping level. This structure improves the collector voltage breakdown characteristics.

The result of this simple analysis implies that for a given device the DC current gain should be independent of current. In reality this is not the case.



The decrease at low currents is due to recombination in the base-emitter depletion region.

At high currents the current gain drops because of resistive voltage losses, shift of the high field region at the collector (increased base width), loss of injection efficiency as the carrier density approaches the doping concentration, etc.

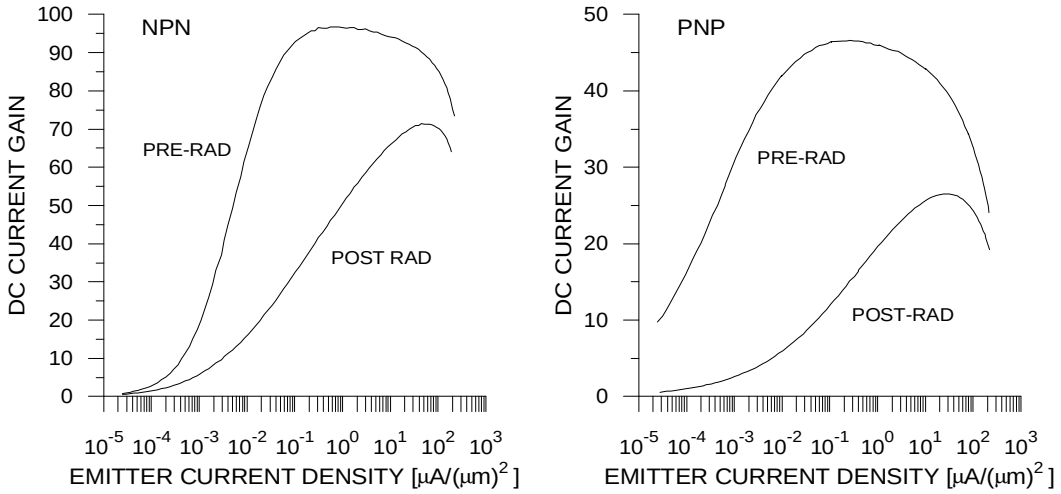
The “ideal” DC current gain depends only on device and material constants, whereas the recombination depends on the local density of injected electrons and holes with respect to the concentration of recombination centers.

Thus, the relative degradation of DC current gain due to recombination depends on the current density.

> Within a given fabrication process, a large transistor will exhibit more recombination than a small transistor at the same current. Stated differently, for a given current, the large transistor will offer more recombination centers for the same number of carriers.

As shown in the plots below, modern devices exhibit DC current gain that is quite uniform over orders of magnitude of emitter current.

The figures also show the degradation of current gain after irradiation, here after exposure to the equivalent of 10^{14} minimum ionizing protons/cm^2. The decrease of DC current gain at low current densities due to increased recombination is apparent.



In a radiation-damaged transistor the reduction in DC current gain for a given DC current will be less for smaller devices.

Although the bipolar transistor is a current driven device, it is often convenient to consider its response to input voltage.

Since the dependence of base current on base-emitter voltage is given by the diode equation

$$I_B = I_R(e^{q_e V_{BE}/k_B T} - 1) \approx I_R e^{q_e V_{BE}/k_B T}$$

The resulting collector current is

$$I_C = \beta_{DC} I_B = \beta_{DC} I_R e^{q_e V_{BE}/k_B T}$$

The transconductance, i.e. the change in collector current vs. base-emitter voltage

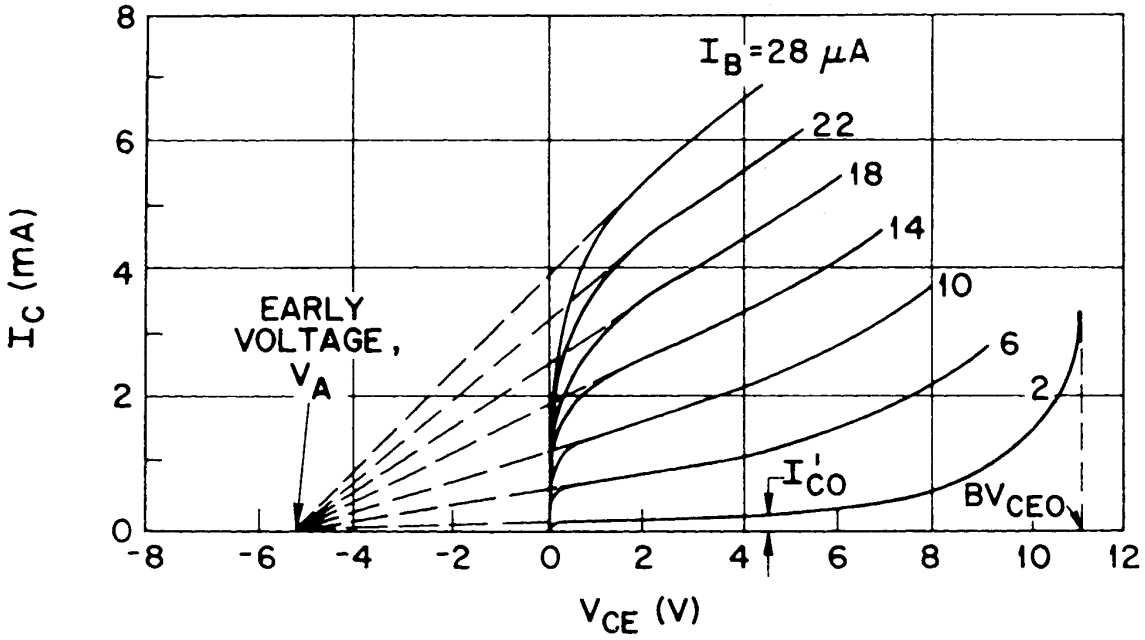
$$g_m \equiv \frac{dI_C}{dV_{BE}} = \beta_{DC} I_R \frac{q_e}{k_B T} e^{q_e V_{BE}/k_B T} = \frac{q_e}{k_B T} I_C$$

The transconductance depends only on collector current, so for any bipolar transistor – regardless of its internal design – setting the collector current determines the transconductance.

Since at room temperature $k_B T/q_e$ = 26 mV

$$g_m = \frac{I_C}{0.026} \approx 40 I_C$$

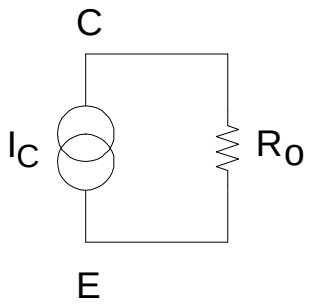
The obtainable voltage gain of an amplifier depends on the output characteristics of the transistor.



At low collector voltages the field in the collector-base region is not sufficient to transport all injected carriers to the collector without recombination. At higher voltages the output current increases gradually with voltage (saturation region), due to the change in effective base width.

An interesting feature is that the extrapolated slopes in the saturation region intersect at the same voltage V_A for I_c= 0 (“Early voltage”).

The finite slope of the output curves is equivalent to a current generator with a shunt resistance



where

$$R_o = K\frac{V_A}{I_C}$$

K is a device-specific constant of order 1.

The maximum obtainable voltage gain is

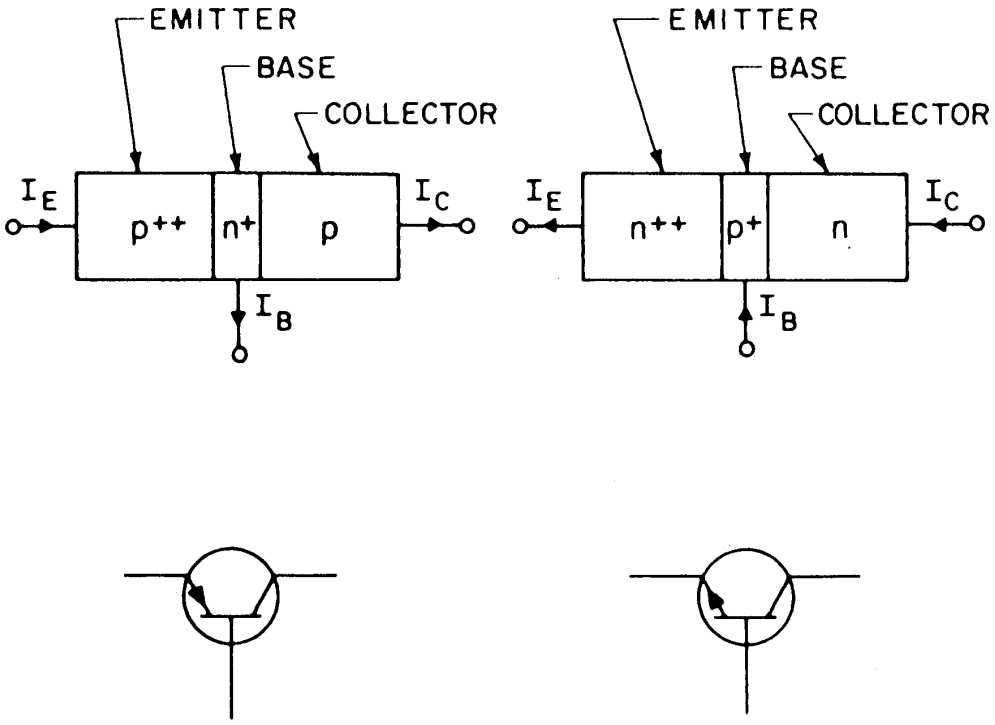
$$A_{v,\max} = \frac{dI_C}{dV_{BE}}R_o = g_m R_o = \frac{I_C}{k_B T / q_e}K\frac{V_A}{I_C} = K\frac{V_A}{k_B T / q_e}$$

which at room temperature is about ($40 V_A K$).

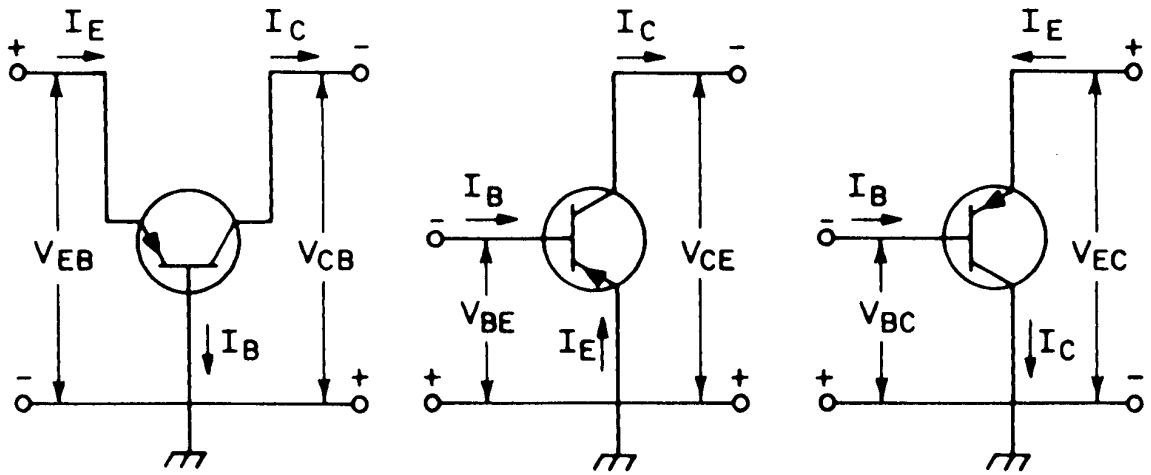
Note that to first order the maximum obtainable voltage gain is independent of current.

Transistors with large Early voltages will allow higher voltage gain.

Bipolar transistors can be implemented as *pnp* or *npn* structures.



In amplifier circuits transistors can be connected in three different circuit configurations:



(from Sze)

a) Common base configuration

The input signal is applied to the emitter, the output taken from the collector.

This configuration is used where a low input impedance is required.

$$R_i = \frac{dV_{EB}}{dI_E} \approx \frac{dV_{EB}}{dI_C} = \frac{1}{g_m} = \frac{k_B T}{q_e}\frac{1}{I_C}$$

Since at room temperature $k_B T/q_e$ = 26 mV

$$R_i = \frac{0.026}{I_C}$$

i.e. R_i = 26 Ω at I_c = 1 mA.

b) Common emitter configuration

The input signal is applied to the base, the output taken from the collector.

The input resistance is higher than that of the common base stage.

$$R_i = \frac{dV_{BE}}{dI_B} \approx \beta_{DC} \frac{dV_{BE}}{dI_C} = \frac{\beta_{DC}}{g_m} = \frac{k_B T}{q_e} \frac{\beta_{DC}}{I_C}$$

Since the base current is about β_{DC} times smaller than the emitter current, the input resistance of the common emitter stage is β_{DC} times larger than that of the common base stage.

For β_{DC} = 100 and I_c = 1 mA, R_i = 2600 Ω.

c) Common collector configuration

The signal is applied to the base and the output taken from the emitter ("emitter folower").

The voltage gain of this configuration cannot exceed 1, but the input resistance can be very high.

Since the load resistance R_L introduces local negative feedback, the input resistance depends on the load

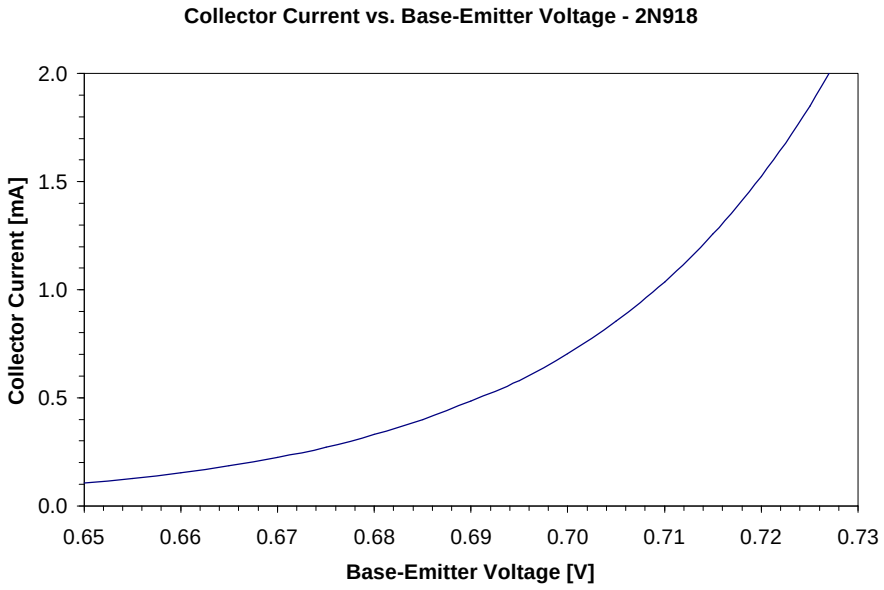
$$R_i \approx \beta R_L$$

The output resistance of the emitter follower is

$$R_o = \frac{dV_{out}}{dI_{out}} = \frac{d(V_{in} - V_{BE})}{dI_E} \approx \frac{dV_{BE}}{dI_E} \approx \frac{1}{g_m}$$

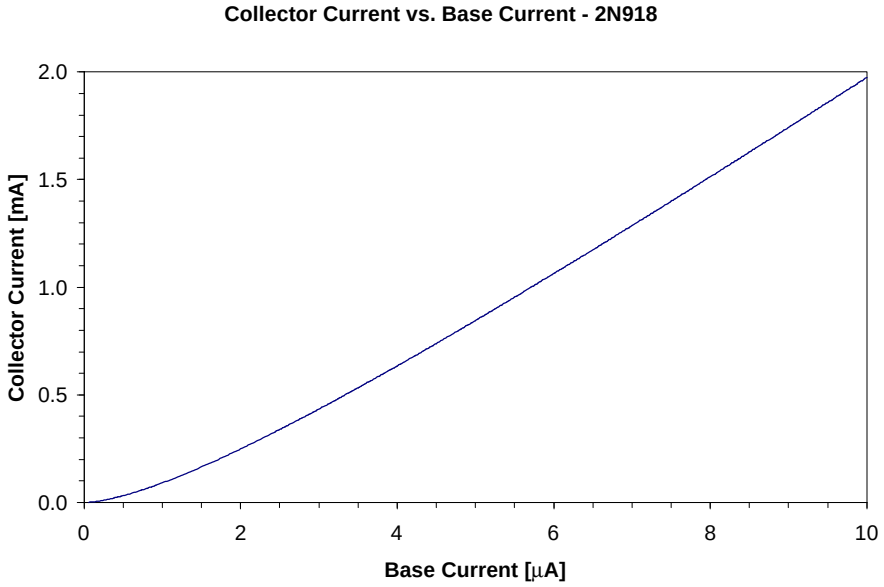
so at 1 mA collector current R_o = 26 Ω, i.e. although the stage only has unity voltage gain, it does have current gain.

Although the bipolar transistor is often treated as a voltage-driven device, the exponential dependence of current on input voltage means that as an amplifier the device is very non-linear.



In audio amplifiers, for example, this causes distortion.
Distortion may be limited by restricting the voltage swing, which to some degree is feasible because of the high transconductance.

With current drive the linearity is much better.



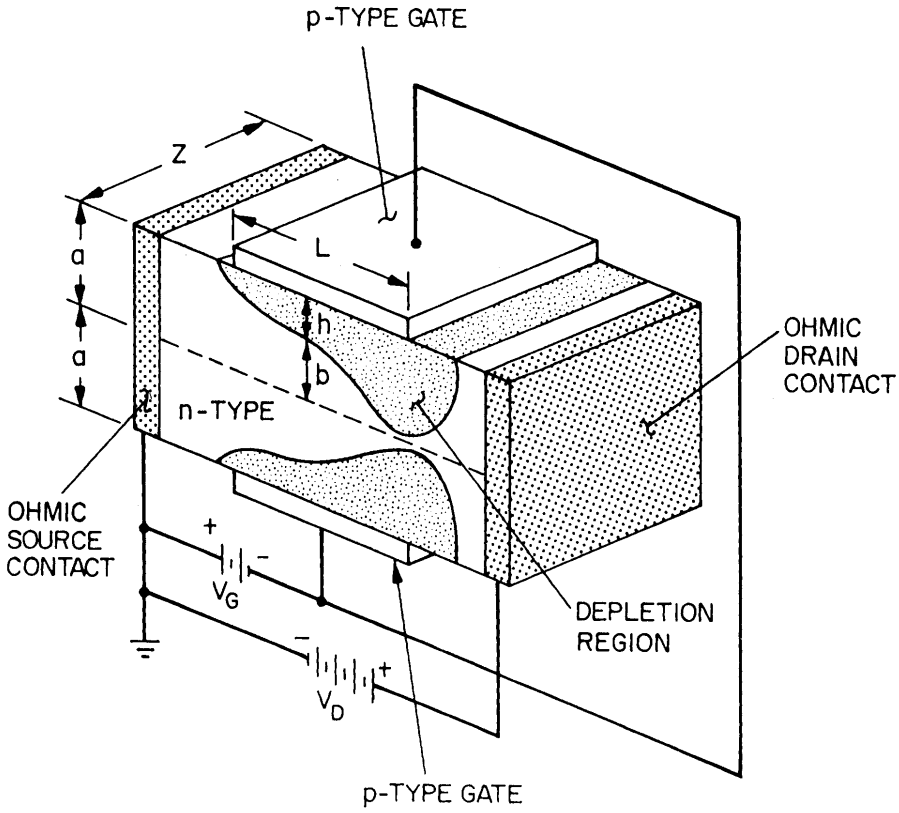
VIII.4. Field Effect Transistors

Field Effect Transistors (FETs) utilize a conductive channel whose resistance is controlled by an applied potential.

1. Junction Field Effect Transistor (JFET)

In JFETs a conducting channel is formed of *n* or *p*-type semiconductor (GaAs, Ge or Si).
Connections are made to each end of the channel, the Drain and Source.

In the implementation shown below a pair of gate electrodes of opposite doping with respect to the channel are placed at opposite sides of the channel. Applying a reverse bias forms a depletion region that reduces the cross section of the conducting channel.



(from Sze)

Changing the magnitude of the reverse bias on the gate modulates the cross section of the channel.

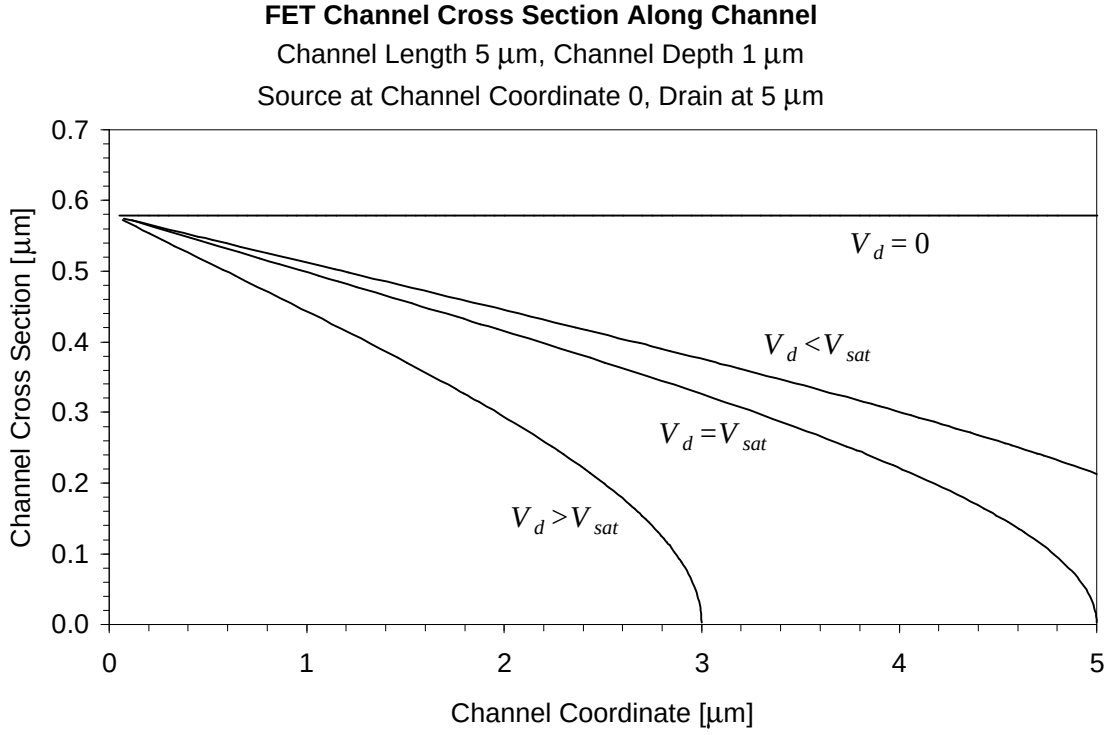
First assume that the drain voltage is 0.
Increasing the reverse gate potential will increase the depletion width, i.e. reduce the cross section of the conducting channel, until the channel is completely depleted. The gate voltage where this obtains is the “pinch-off voltage” V_P.

Now set both the gate and drain voltages to 0. The channel will be partially depleted due to the “built-in” junction voltage.

Now apply a drain voltage. Since the drain is at a higher potential than the source, the effective depletion voltage increases in proximity to the drain, so the width of the depletion region will increase as it approaches the drain.

If the sum of the gate and drain voltage is sufficient to fully deplete the channel, the device is said to be “pinched off”.

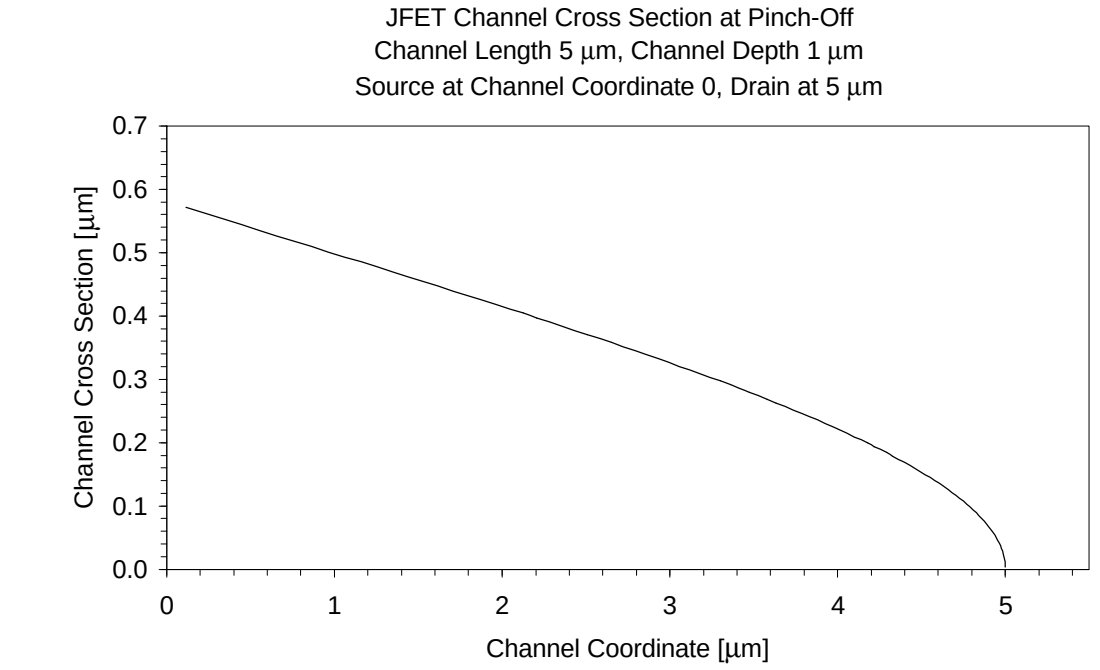
Increasing the drain voltage beyond this point moves the pinch-off point towards to the source.

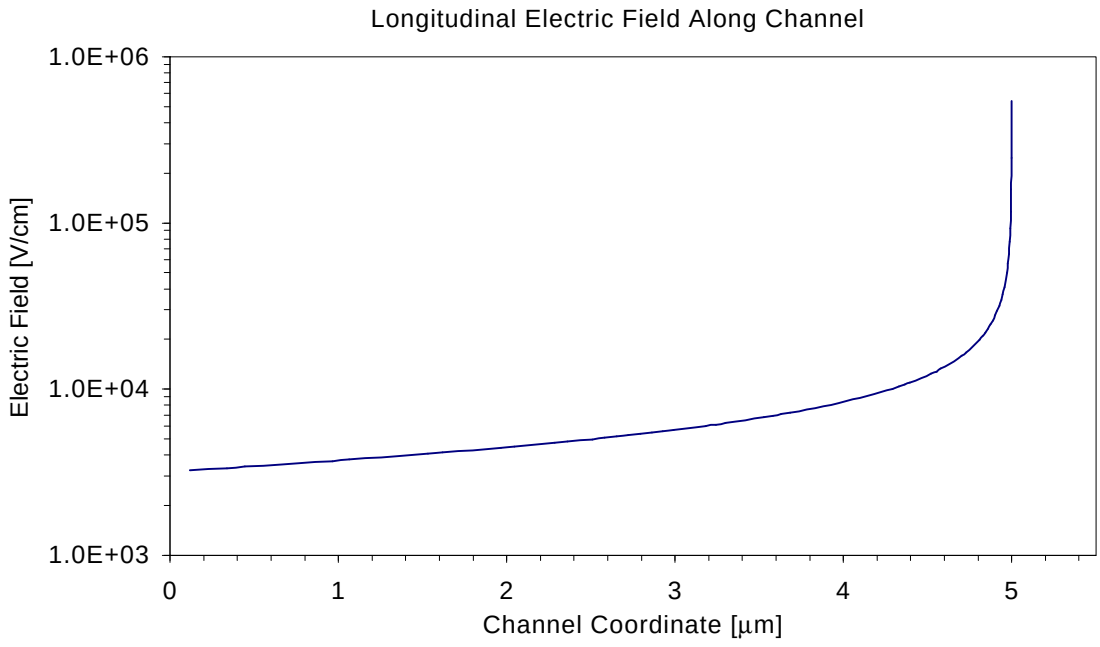


Pinching off the channel does not interrupt current flow. All thermally excited carriers have been removed from the depleted region, but carriers from the channel can still move through the potential drop to the drain.

The profile of the depletion region is not determined by the static potentials alone.

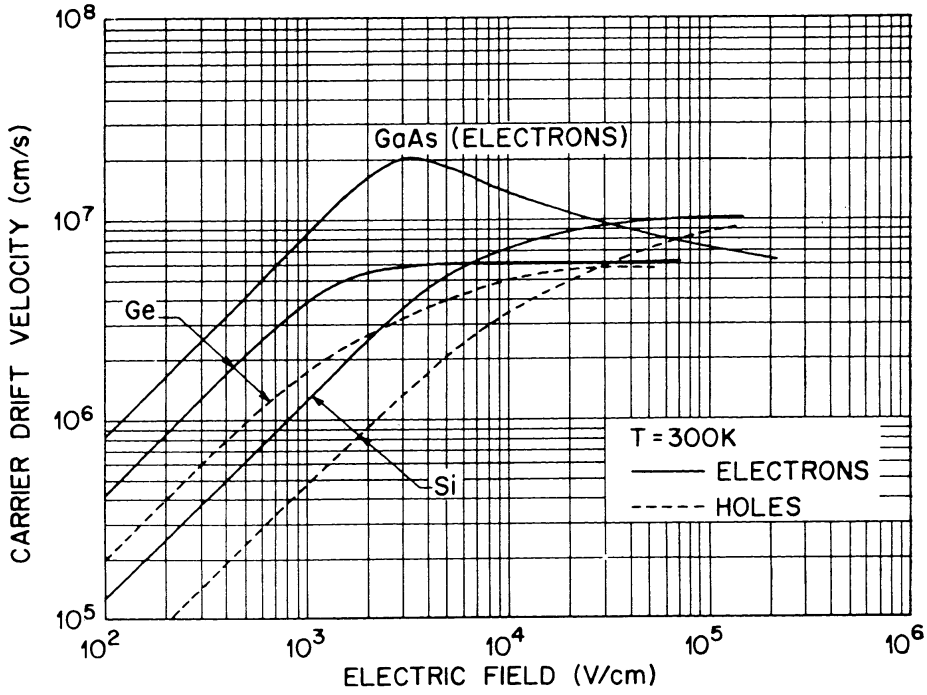
Current flow along the channel changes the local potential.
As the channel cross section decreases, the incremental voltage drop increases, i.e. the longitudinal drift field that determines the carrier velocity increases.





At high electric fields the mobility decreases.

This comes about because as the carriers' velocity increases they begin to excite optical phonons. At fields above 10^5 V/cm practically all of the energy imparted by an increased field goes into phonon emission.



(from Sze)

Since the velocity saturates at high fields, the current

$$I = N_C q_e v$$

also saturates, since the number of carriers N_C remains constant, i.e. at high fields silicon acts as an incremental insulator ($dI/dV = 0$).

As the drain voltage is increased beyond pinch-off, the additional voltage decreases the length of the resistive channel, but also increases the potential drop in the drain depletion region.

$\Rightarrow$ The current increases only gradually with drain voltage.

Current Voltage Characteristics

Low drain and gate voltages:

The resistive channel extends from the source to the drain.

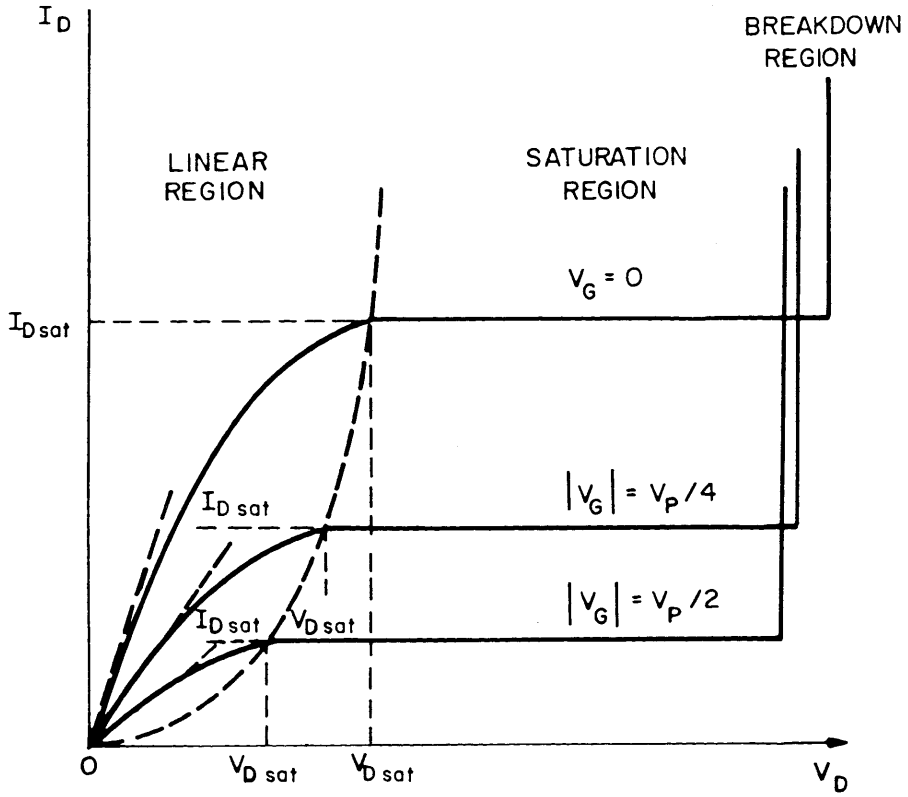
The drain current increases linearly with drain voltage.

"Linear Region"

Gate and drain voltages sufficiently high to pinch off the resistive channel:

The drain current remains constant with increasing drain voltage.

"Saturation Region"



(from Sze)

The drain saturation voltage V_{Dsat} increases as the gate voltage is changed from the static pinch off voltage V_p towards 0.

For use in amplifiers the characteristics in the saturation region are of the most interest.

To a good approximation

$$I_D = I_{DSS}\left(1-\left(\frac{V_G + V_{bi}}{V_P}\right)\right)^2$$

where V_p is the "pinch-off" voltage, i.e. the gate voltage at which the channel is fully depleted. The drain saturation current

$$I_{DSS} = \frac{1}{6\varepsilon}\mu(q_e N_{ch})^2 d^3 \frac{W}{L}$$

is determined by the carrier mobility μ, the doping level in the channel N_{ch} and the channel depth d, width W and length L. ε is the dielectric constant.

The transconductance

$$g_m = \left|\frac{dI_D}{dV_G}\right| = \frac{2I_{DSS}}{V_P}\left(1-\frac{V_G + V_{bi}}{V_P}\right)$$

is maximum for V_G= 0, i.e. maximum drain current, and for small pinch off voltages. Then

$$g_m\big|_{V_G=0} \approx \frac{2I_{DSS}}{V_P}$$

The transconductance depends primarily on current.

$$g_m = \frac{2I_{DSS}}{V_P}\left(1-\frac{V_G + V_{bi}}{V_P}\right) = \frac{2\sqrt{I_{DSS}}}{V_P}\sqrt{I_D}$$

The applied voltages only provide the boundary conditions to set up the required current. To maintain performance it is important to control the current (rather than the voltages).

To see how device parameters affect the transconductance, we'll ignore the built-in voltage since it varies only weakly with doping $V_{bi} = (k_B T/q_e) \log(N_{ch}/n_i)$.

With this approximation

$$g_m\big|_{V_G=0} \approx \frac{2I_{DSS}}{V_P} \approx \frac{W}{L}\frac{\mu(q_e N_{ch})^2 d^3}{3q_e N_{ch} d^2} \propto \frac{W}{L}\mu N_{ch} d$$

Obviously, a high carrier mobility will increase the transconductance, since for a given carrier concentration this will increase the magnitude of the current.

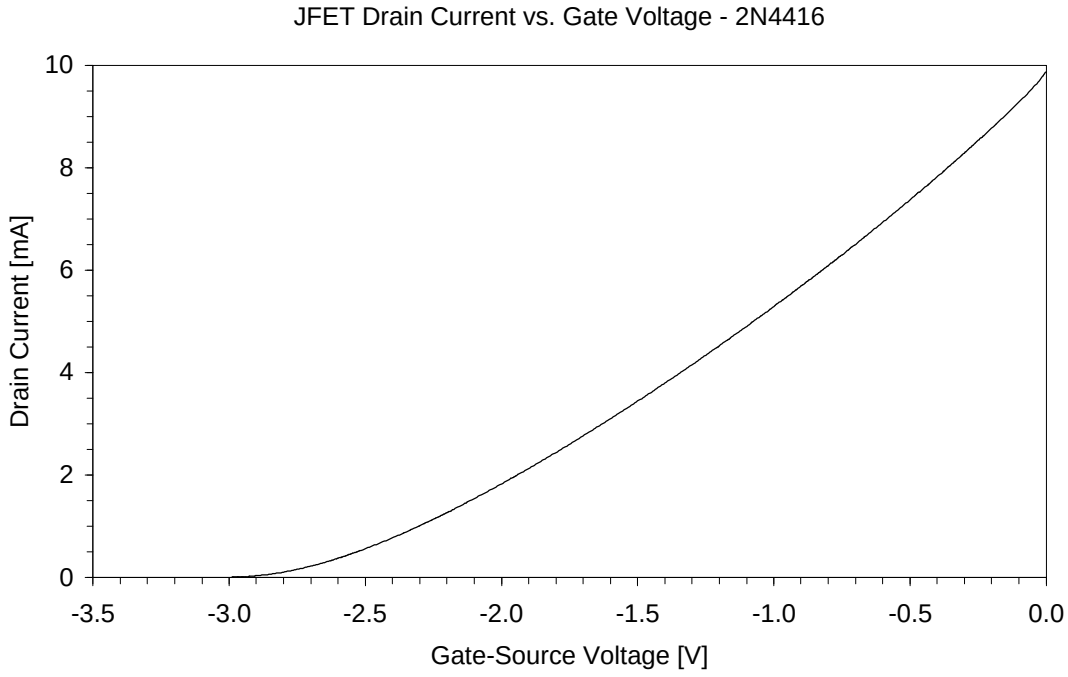
1. The proportionality of transconductance to width W is trivial, since it is equivalent to merely connecting device in parallel, so the normalized transconductance g_m/W is used to compare technologies.
2. The normalized transconductance increases with the number of carriers per unit length $N_{ch}d$ and decreasing channel length L.
3. Transconductance increases with drain current

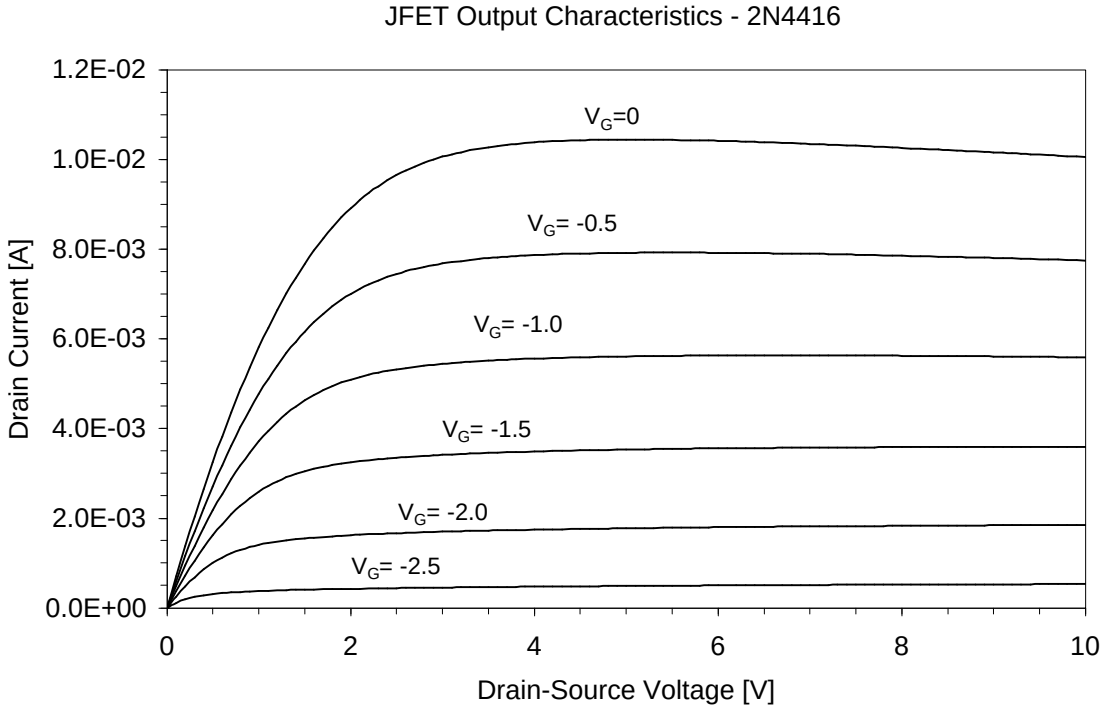
$$g_m = \left|\frac{dI_D}{dV_G}\right| = \frac{2I_{DSS}}{V_P}\left(1 - \frac{V_G + V_{bi}}{V_P}\right) = \frac{2\sqrt{I_{DSS}}}{V_P}\sqrt{I_D}$$

   i.e. drain current is the primary parameter; the applied voltages are only the means to establish I_D.

All of these optimizations also increase the power dissipation. For low power systems optimization is more involved.

Measured JFET Characteristics



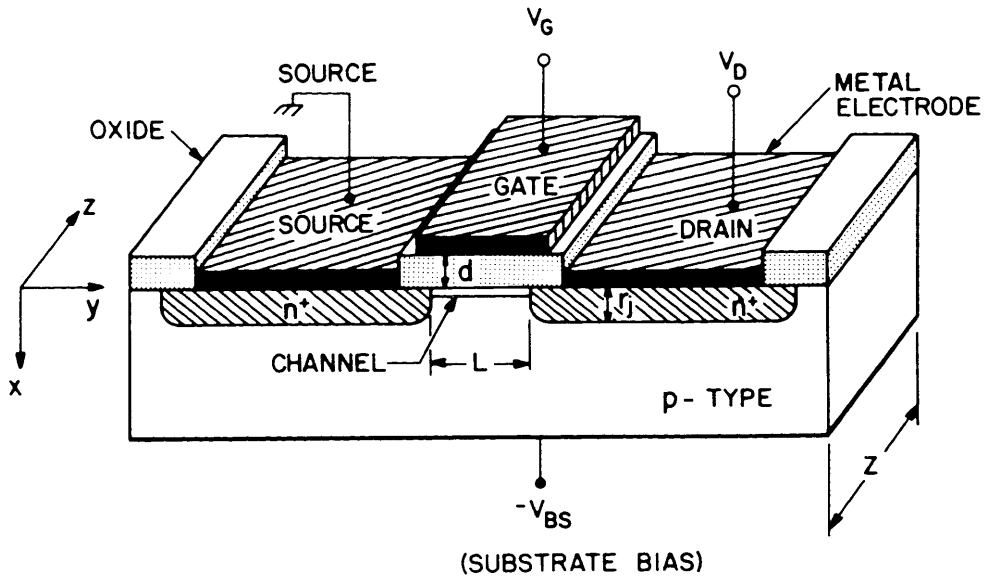


The slight drop in drain current for V_G=0 is due to self-heating.

Metal Oxide Field Effect Transistors (MOSFETs)

Both JFETs and MOSFETs are conductivity modulated devices, utilizing only one type of charge carrier. Thus they are called unipolar devices, unlike bipolar transistors, for which both electrons and holes are crucial.

Unlike a JFET, where a conducting channel is formed by doping and its geometry modulated by the applied voltages, the MOSFET changes the carrier concentration in the channel.



(from Sze)

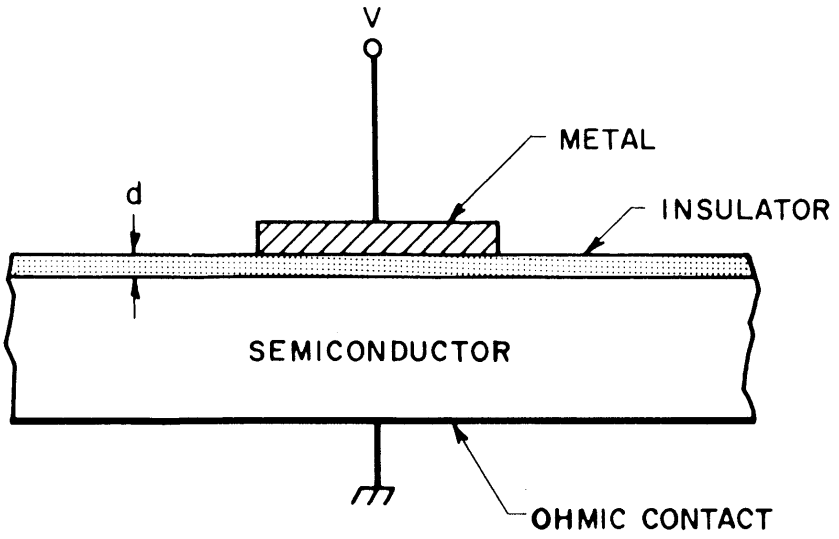
The source and drain are n^+ regions in a p-substrate.

The gate is capacitively coupled to the channel region through an insulating layer, typically SiO_2.

Applying a positive voltage to the gate increases the electron concentration at the silicon surface beneath the gate.

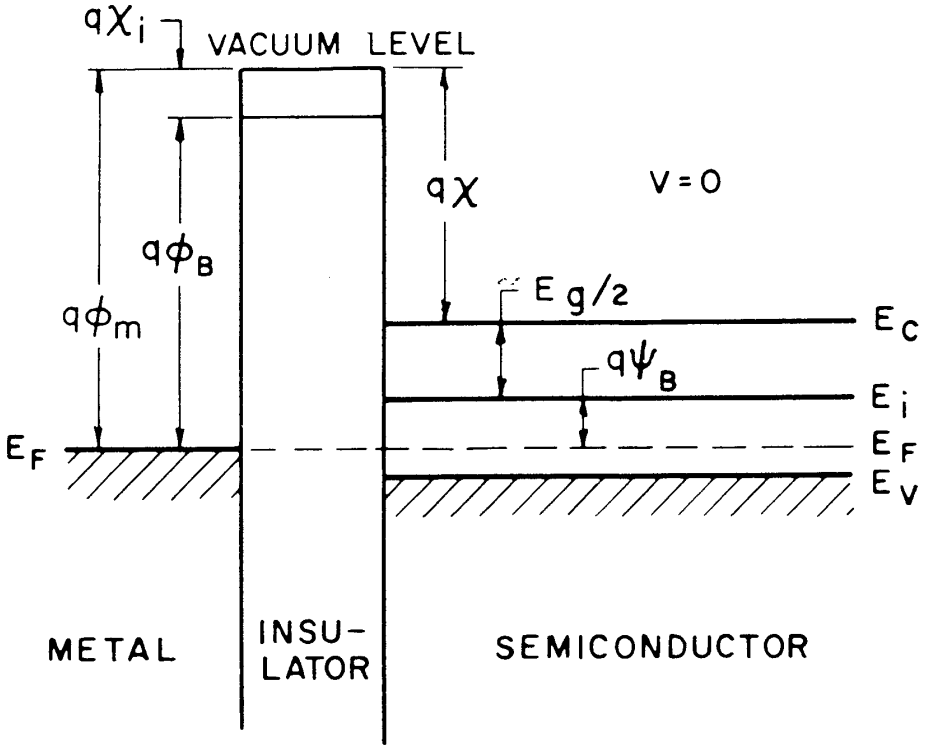
As in a JFET the combination of gate and drain voltages control the conductivity of the channel.

Formation of the Channel - The MOS Capacitor



(from Sze)

Band structure in an ideal MOS capacitor on a *p*-Substrate

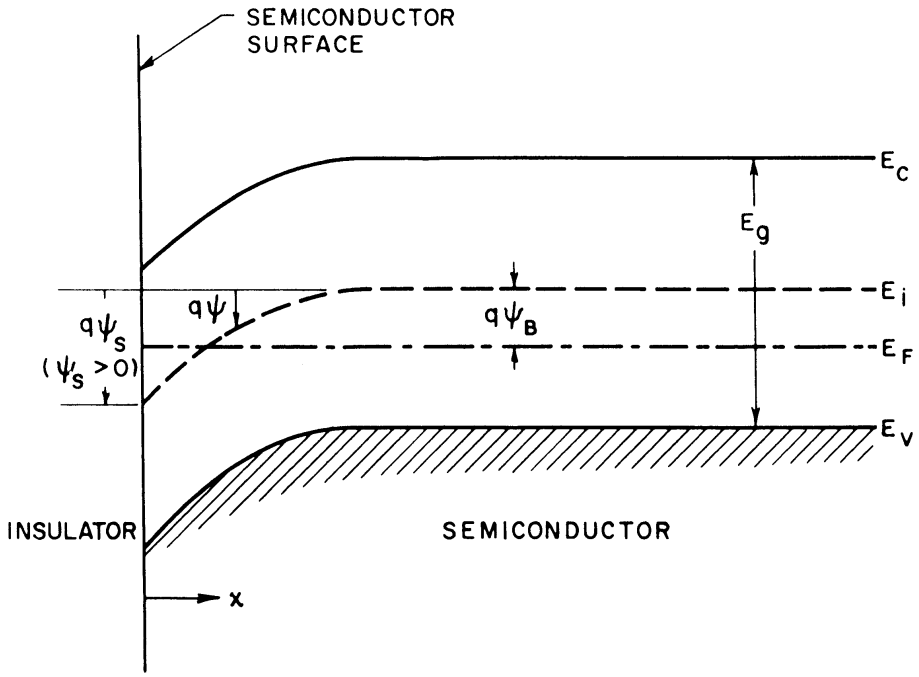


(from Sze)

In its natural state, however, the band structure is not flat as just shown.

The discontinuity in the crystal structure and charge trapped at the surface change the potential at the surface, so the bands bend.

Energy band diagram of a *p*-type semiconductor



(from Sze)

As shown above the surface potential Ψ_s is positive.

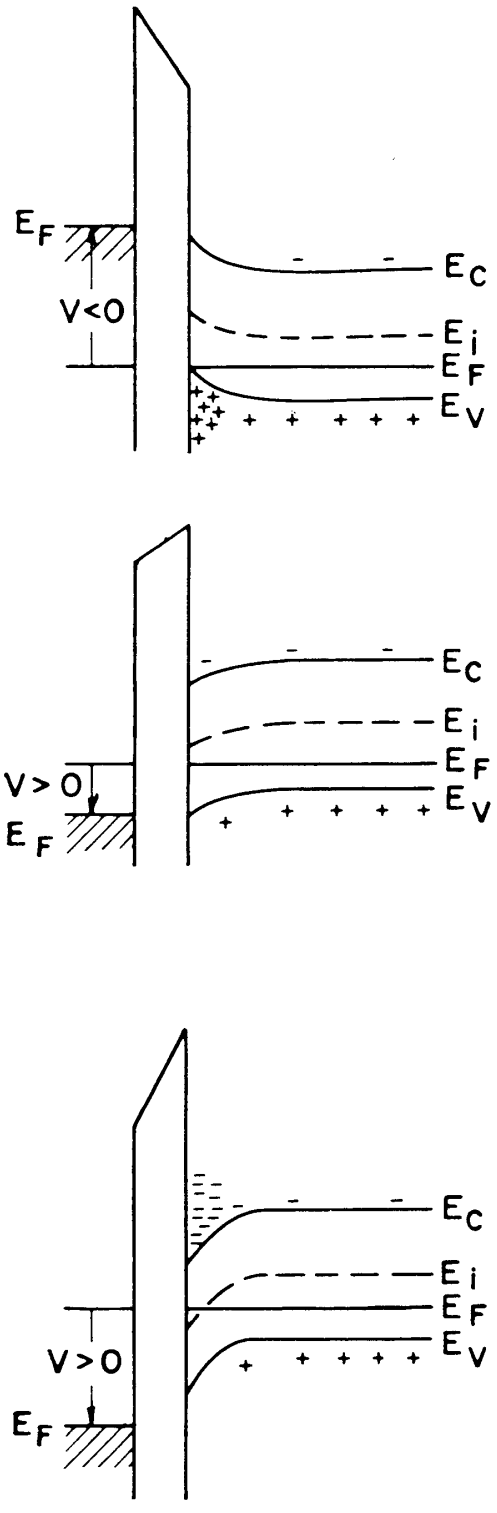
$\Psi_S < 0$ the bands bend upwards, increasing the hole concentration at the surface.

$\Psi_B > \Psi_S > 0$ the bands bend slightly downwards, reducing the concentration of holes at the surface (depletion)

$\Psi_S > \Psi_B$ the conduction band edge dips below the Fermi level, leading to an accumulation of electrons at the surface (inversion)

In the absence of any special surface preparation the surface of silicon is *n*-type, i.e. *p*-type silicon inverts at the surface.

Surface concentration vs. band-bending in *p*-type material



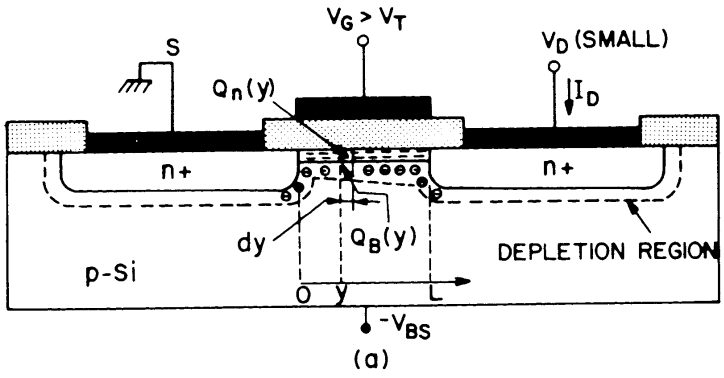
(from Sze)

An *n*-channel MOSFET utilizes an *n*-channel in a *p*-substrate, so application of a positive potential to the gate forms the inversion layer needed for the channel.

As in the JFET, the combination of current flow in the channel and the applied potentials forms a depletion region that is greatest near the drain. At a sufficiently large drain potential the channel “pinches off”.

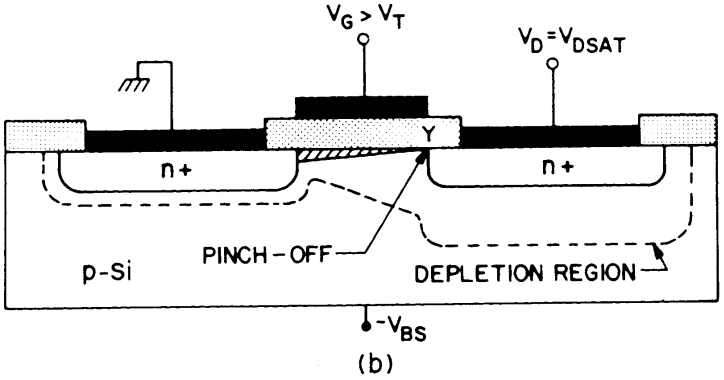
Low drain voltage
$V_D < V_{sat}$

Resistive channel



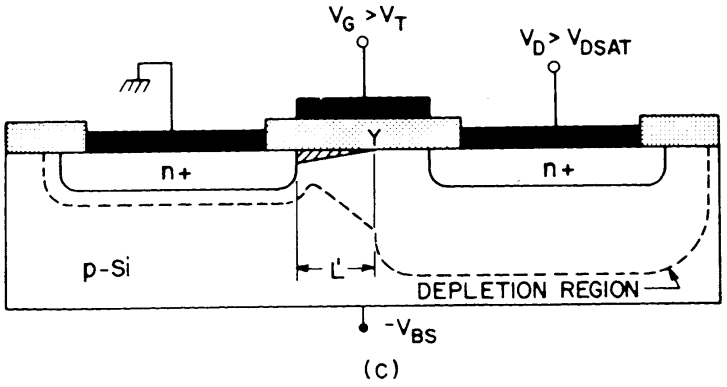
$V_D = V_{sat}$

Onset of current saturation



$V_D > V_{sat}$

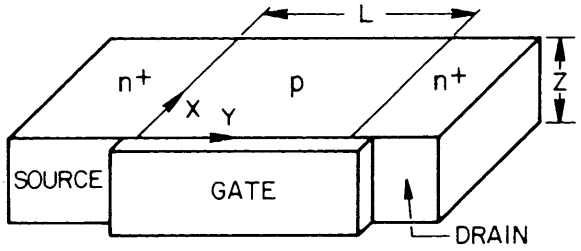
Output current saturated



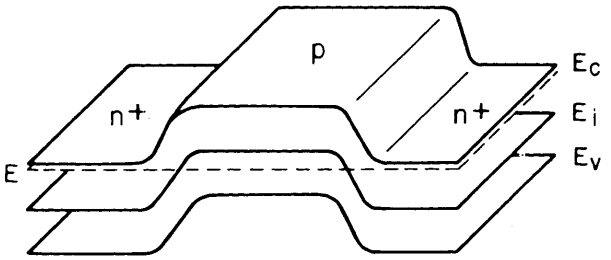
(from Sze)

Three-Dimensional Potential Diagram of an *n*-Channel MOSFET

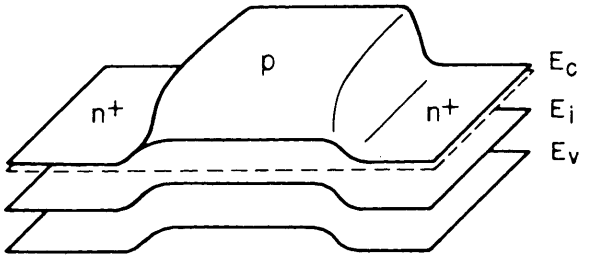
Device orientation



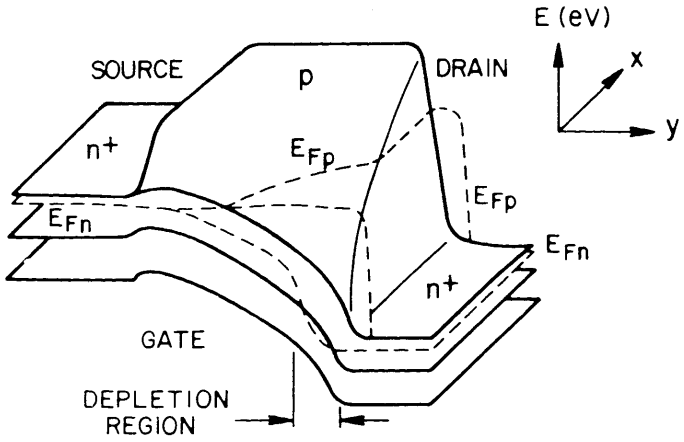
Flat-band condition



Formation of inversion layer



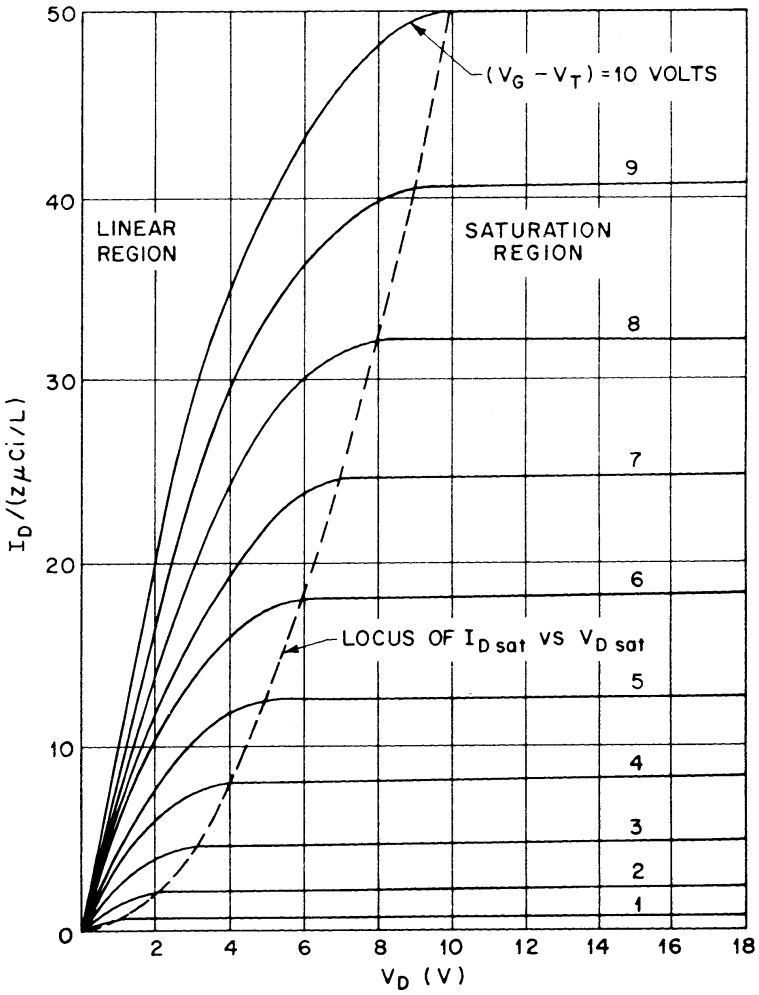
Drain voltage beyond saturation



(from Sze)

The output curves of a MOSFET are similar to a JFET.

The drain voltage required to attain saturation increases with the operating current.



(from Sze)

In saturation

$$I_D = \frac{W}{L}\frac{\mu\, C_i}{2}(V_G - V_T)^2$$

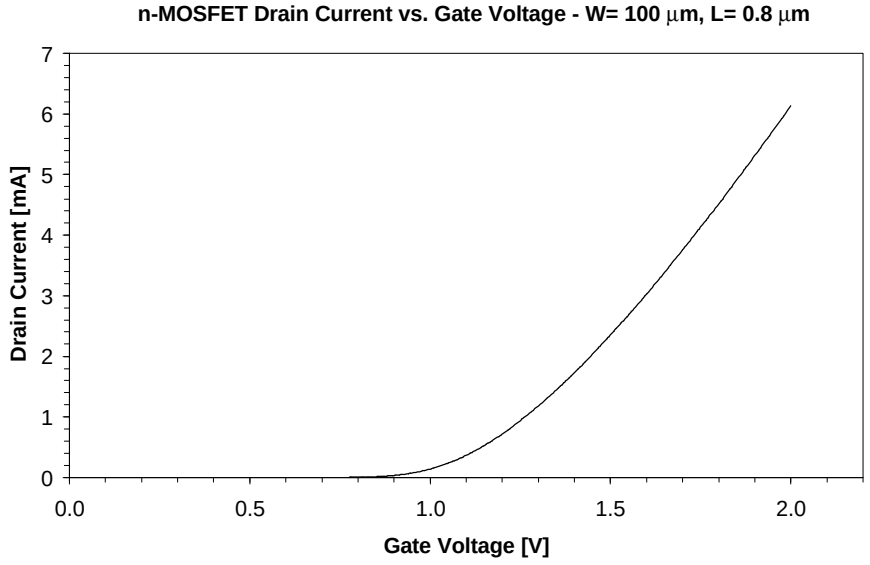
where C_i is the gate capacitance per unit area ε_{ox}/d_{ox} and V_T is the gate voltage corresponding to the onset of strong inversion ("threshold voltage")

From this the transconductance is

$$g_m = \frac{W}{L} C_i \mu\, (V_G - V_T) = \frac{W}{L}\frac{\varepsilon_{ox}}{d_{ox}}\mu\, (V_G - V_T) = \sqrt{\frac{W}{L}\cdot\frac{\varepsilon_{ox}}{d_{ox}}\mu \cdot I_D}$$

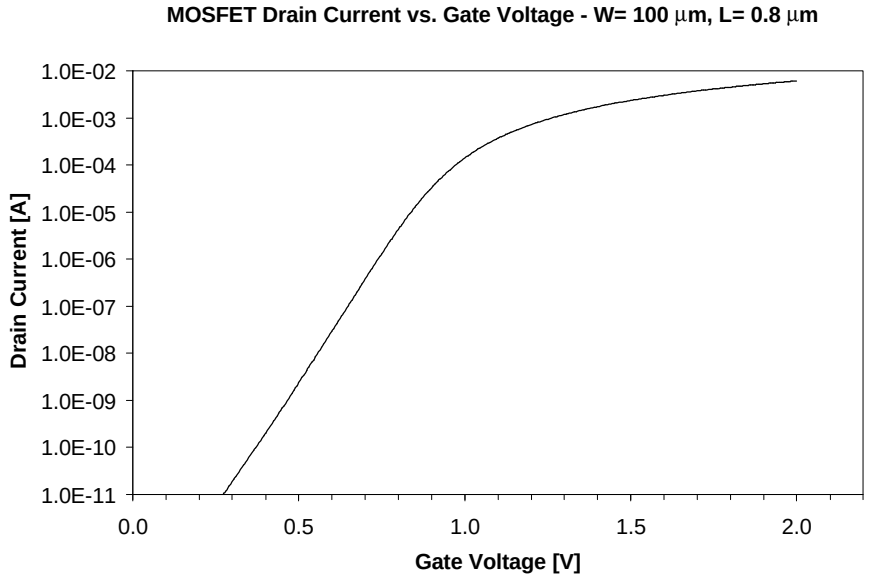
For a given width W and drain current I_D the transconductance is increased by decreasing the channel length L and the thickness of the gate oxide d_{ox}. As for the JFET, the transconductance depends primarily on current.

Measured characteristics of an *n*-channel MOSFET with 0.8 μm channel length and 20 nm gate oxide thickness



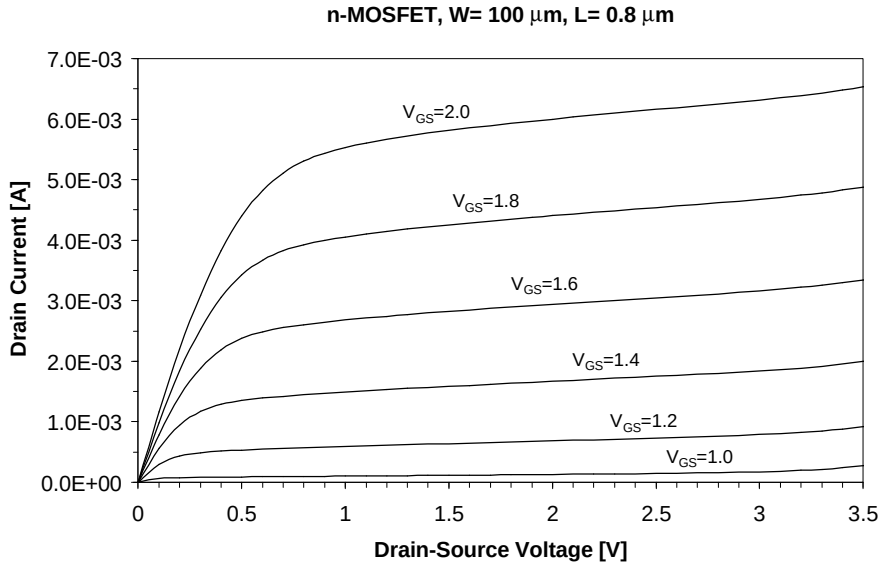
For this device the threshold voltage V_T is about 1.2 V.

The transition from weak to strong inversion becomes more apparent in a logarithmic plot.



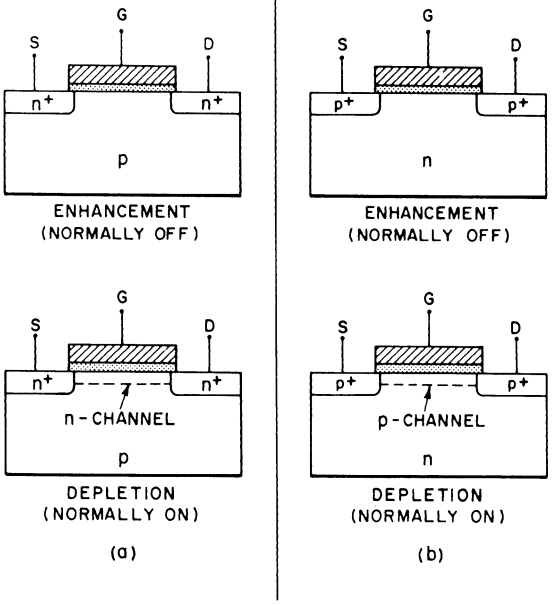
In the subthreshold regime the drain current is proportional to the inversion carrier concentration, i.e. it increases exponentially with gate voltage.

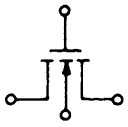
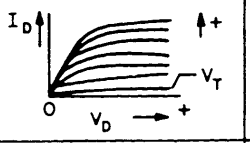
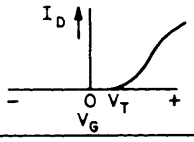
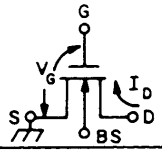
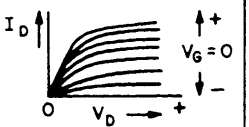
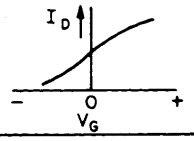
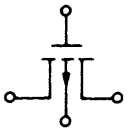
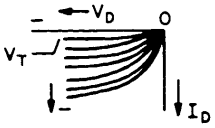
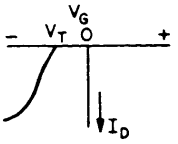
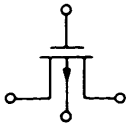
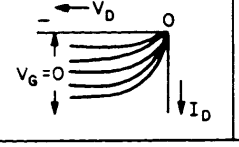
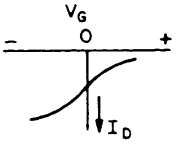
The increase in output saturation voltage with gate voltage can be seen clearly in the output curves.



Depending on the substrate doping MOSFETs can be implemented with either *n* or *p*-channels.

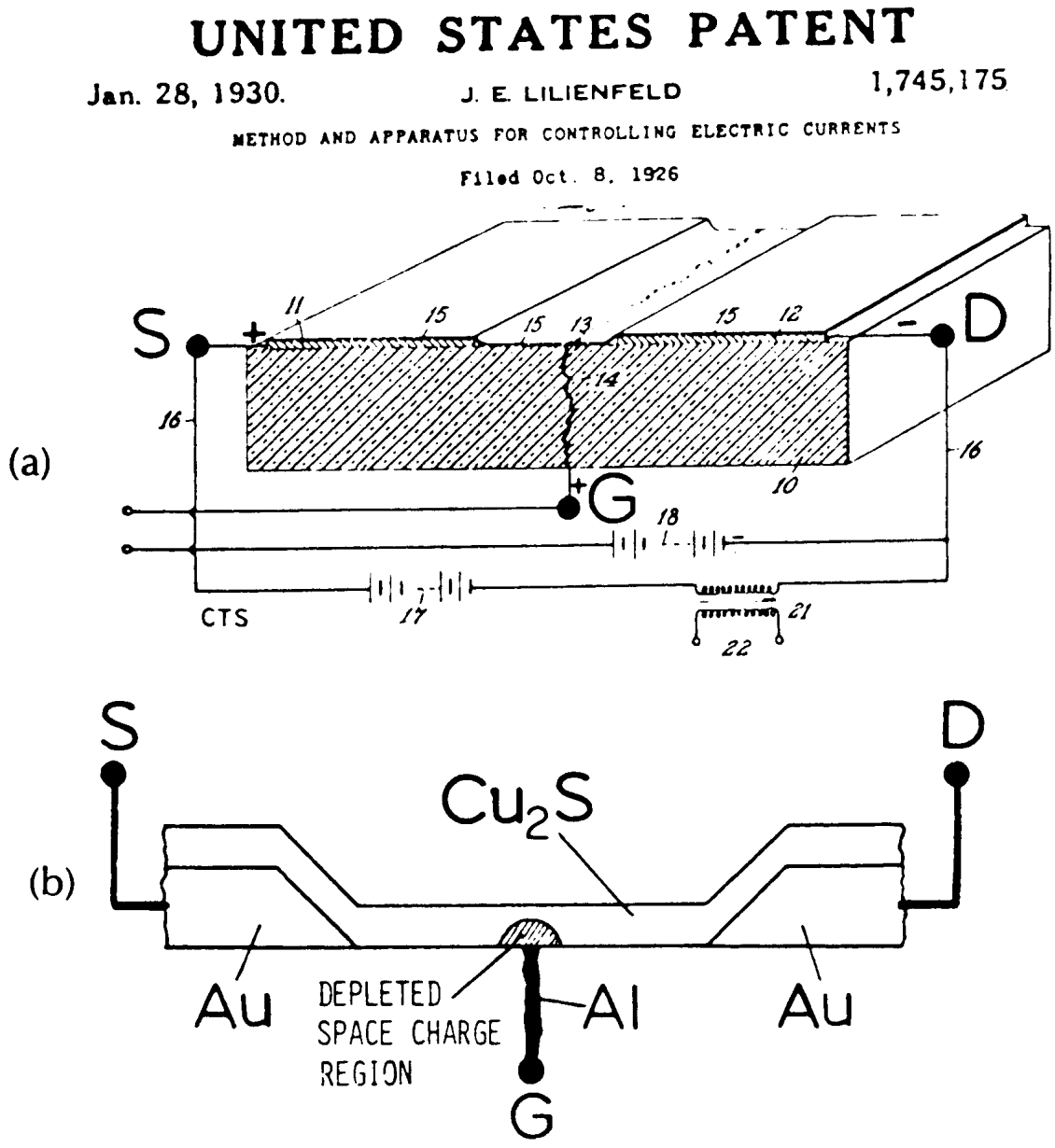
A thin surface layer can be implanted to adjust the threshold voltage. With this devices can be normally on at zero gate voltage (depletion mode) or normally off, i.e. require additional voltage to form the inversion layer (enhancement mode), as illustrated below



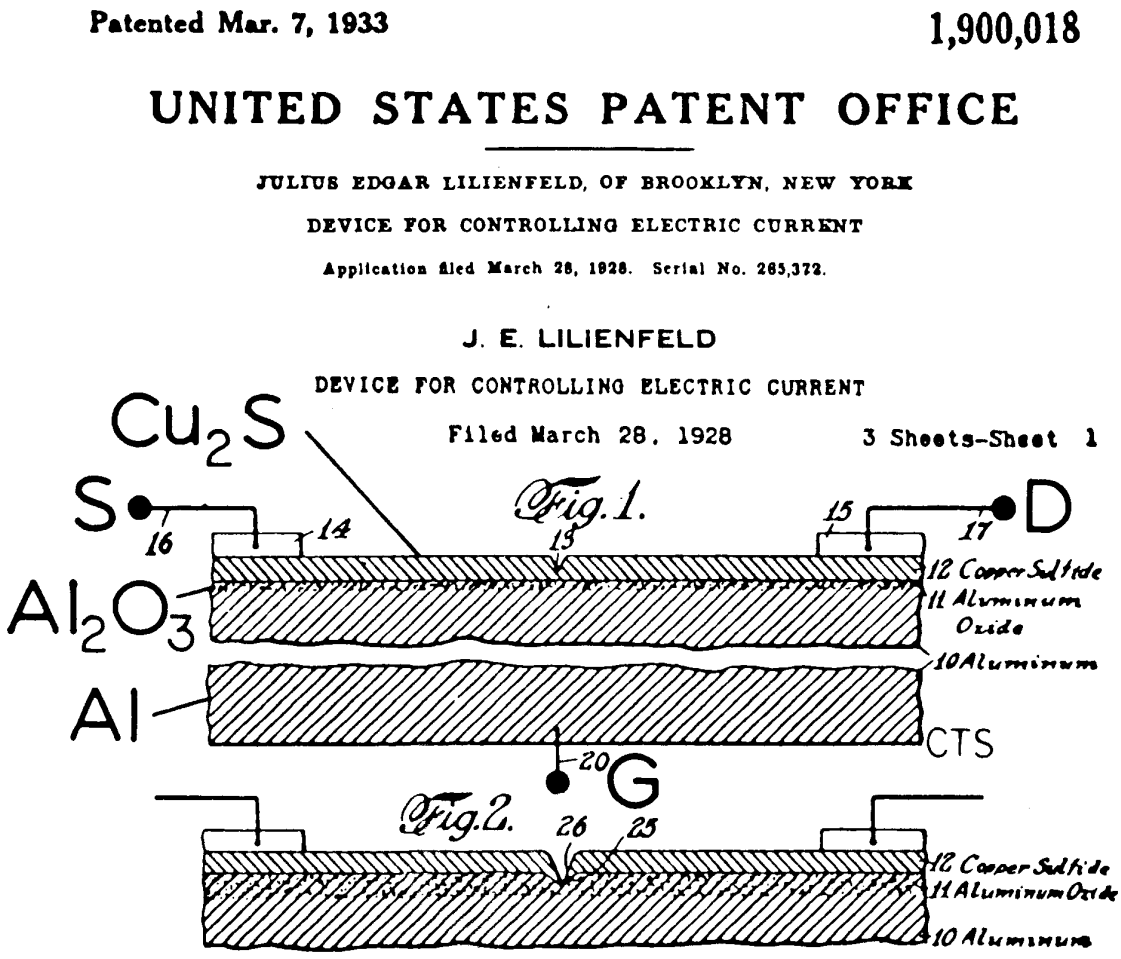
TYPE	ELECTRICAL SYMBOL	OUTPUT CHARACTERISTIC	TRANSFER CHARACTERISTIC
N-CHANNEL ENHANCEMENT (NORMALLY OFF)			
N-CHANNEL DEPLETION (NORMALLY ON)			
P-CHANNEL ENHANCEMENT (NORMALLY OFF)			
P-CHANNEL DEPLETION (NORMALLY ON)			

… So what else is new?

The first patent awarded for a junction field effect transistor was submitted in 1926



In 1928 Lilienfeld submitted a patent application for a MOSFET



Patented Mar. 7, 1933 1,900,018

UNITED STATES PATENT OFFICE

JULIUS EDGAR LILIENFELD, OF BROOKLYN, NEW YORK

DEVICE FOR CONTROLLING ELECTRIC CURRENT

Application filed March 28, 1928. Serial No. 265,372.

J. E. LILIENFELD

DEVICE FOR CONTROLLING ELECTRIC CURRENT

Filed March 28, 1928 3 Sheets-Sheet 1

Although Lilienfeld appears to have fabricated prototypes, the results were not reproducible, because surface states and impurity levels could not be controlled.

Furthermore, unknown to everyone at the time, the dynamics of electrons and holes and practically all of semiconductor physics had yet to be understood.

Nevertheless, these concepts provided the impetus for the research that lead to the first

bipolar transistor in 1947,
JFET in 1953 and
practical Si MOSFETs in 1960.

VIII.5. Noise in Transistors

a) Noise in Field Effect Transistors

The primary noise sources in field effect transistors are

a) thermal noise in the channel

b) gate current in JFETs

Since the area of the gate is small, this contribution to the noise is very small and usually can be neglected.

Thermal velocity fluctuations of the charge carriers in the channel superimpose a noise current on the output current.

The spectral density of the noise current at the drain is

$$i_{nd}^2 = \frac{N_{C,tot} q_e}{L^2} \mu_0 \, 4k_B T_e$$

The current fluctuations depend on the number of charge carriers in the channel $N_{C,tot}$ and their thermal velocity, which in turn depends on their temperature T_e and low field mobility μ_0. Finally, the induced current scales with 1/L because of Ramo's theorem.

To make practical use of the above expression it is necessary to express it in terms of directly measureable device parameters. Since the transconductance in the saturation region

$$g_m \propto \frac{W}{L} \mu N_{ch} d$$

one can express the noise current as

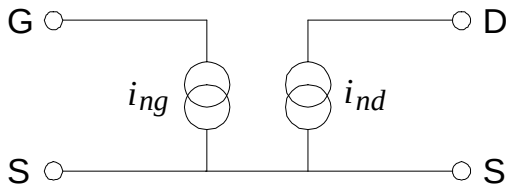
$$i_{nd}^2 = \gamma_n g_m \, 4k_B T_0$$

where T_0= 300K and γ_n is a semi-empirical constant that depends on the carrier concentration in the channel and the device geometry.

In a JFET the gate noise current is the shot noise associated with the reverse bias current of the gate-channel diode

$$i_{ng} = 2q_e I_G$$

The noise model of the FET



The gate and drain noise currents are independent of one another.

However, if an impedance Z is connected between the gate and the source, the gate noise current will flow through this impedance and generate a voltage at the gate

$$v_{ng} = Z\, i_{ng}$$

leading to an additional noise current at the output $g_m v_{ng}$, so that the total noise current at the output becomes

$$i_{no}^2 = i_{nd}^2 + (g_m Z i_{ng})^2$$

To allow a direct comparison with the input signal this cumulative noise will be referred back to the input to yield the equivalent input noise voltage

$$v_{ni}^2 = \frac{i_{no}^2}{g_m^2} = \frac{i_{nd}^2}{g_m^2} + Z i_{ng}^2 \equiv v_n^2 + Z i_n^2$$

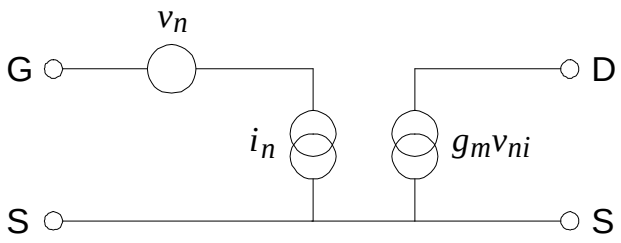
i.e. referred to the input, the drain noise current i_{nd} translates into a noise voltage source

$$v_n^2 = 4k_B T_0 \frac{\gamma_n}{g_m}$$

The noise coefficient γ_n is usually given as 2/3, but is typically in the range 0.5 to 1 (exp. data will shown later).

This expression describes the noise of both JFETs and MOSFETs.

In this parameterization the noise model becomes



where v_n and i_n are the input voltage and current noise. A shown above these contribute to the input noise voltage v_{ni}, which in turn translates to the output through the transconductance g_m to yield a noise current at the output $g_m v_{ni}$.

The equivalent noise charge

$$Q_n^2 = i_n^2 F_i T + v_n^2 C_i^2 \frac{F_v}{T}$$

For a typical JFET g_m= 0.02, C_i= 10 pF and I_G < 150 pA. If F_i=F_v=1

$$Q_n^2 = 1.9 \cdot 10^9 T + \frac{3.25 \cdot 10^{-3}}{T}$$

As the shaping time T decreases, the current noise contribution decreases and the voltage noise contribution increases. For T= 1 μs the current contribution is 43 el and the voltage contribution 3250 el, so the current contribution is negligible, except in very low frequency applications.

Optimization of Device Geometry

For a given device technology and normalized operating current I_D/W both the transconductance and the input capacitance are proportional to device width W

$$g_m \propto W \quad \text{and} \quad C_i \propto W$$

so that the ratio

$$\frac{g_m}{C_i} = const$$

Then the signal-to-noise ratio can be written as

$$\left(\frac{S}{N}\right)^2 = \frac{(Q_s/C)^2}{v_n^2} = \frac{Q_s^2}{(C_{\det} + C_i)^2}\frac{g_m}{4k_B T_0 \Delta f}$$

$$\left(\frac{S}{N}\right)^2 = \frac{Q_s^2}{\Delta f}\frac{1}{4k_B T_0}\left(\frac{g_m}{C_i}\right)\frac{1}{C_i\left(1+\frac{C_{\det}}{C_i}\right)^2}$$

S/N is maximized for $C_i = C_{det}$ (capacitive matching).

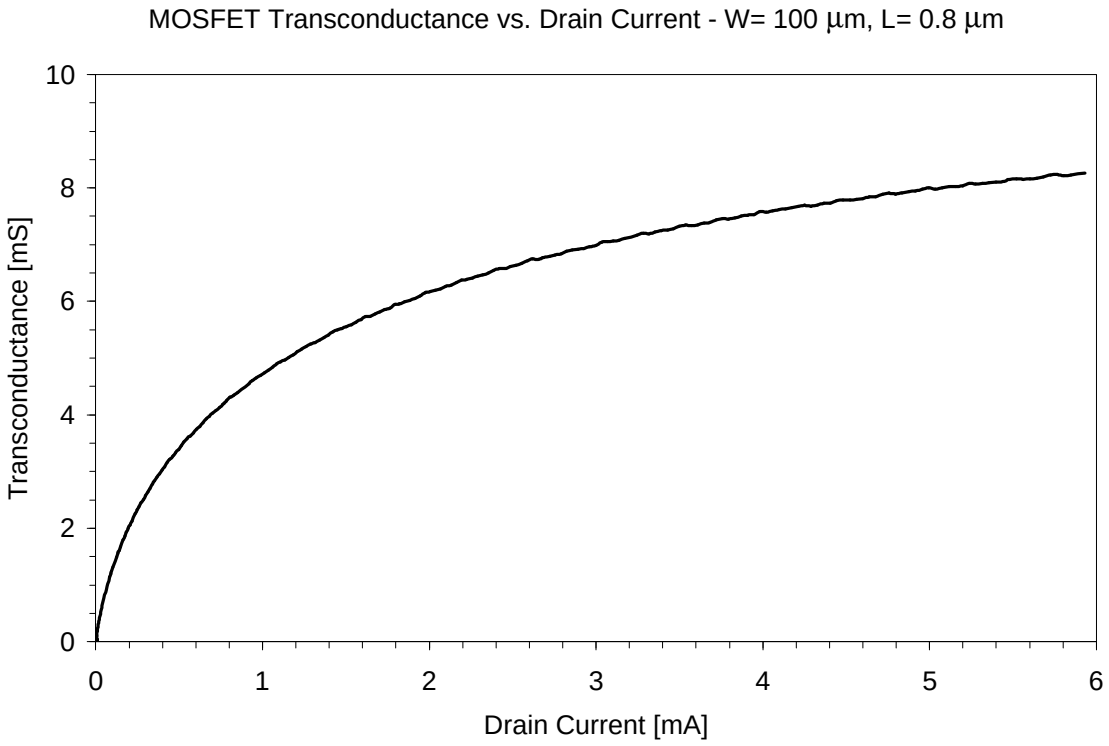
$C_i << C_{det}$: The detector capacitance dominates, so the effect of increased transistor capacitance is negligible.
As the device width is increased the transconductance increases and the equivalent noise voltage decreases, so S/N improves.

$C_i > C_{det}$: The equivalent input noise voltage decreases as the device width is increased, but only with $1/\sqrt{W}$, so the increase in capacitance overrides, decreasing S/N.

Minimum Obtainable Noise Charge

Device scaling can be used to determine the minimum obtainable noise charge for a given device technology.

The transconductance of an FET increases with drain current as shown below.



However, noise only decreases up to a certain current. The reason is that the noise from parasitic source and gate resistances becomes significant.

Assume that a transistor of width W assumes its minimum noise at a current I_d with an associated transconductance g_m.

Since the parasitic gate and source resistances are both inversely proportional to device width, the current density I_d/W will be the same for all widths of transistors using the same technology (and device length).

Thus, to obtain minimum noise one can tailor the FET to a given detector by scaling the device width and keeping the current density I_d/W constant.

Within this framework one can characterize the device technology by the normalized transconductance and input capacitance

$$g_m' = \frac{g_m}{W} \quad \text{and} \quad C_i' = \frac{C_i}{W}$$

and use these quantities to scale to any other device width. Since the equivalent input noise voltage

$$v_n^2 \propto \frac{1}{g_m}$$

the normalized input noise voltage is

$$v_n' = v_n \sqrt{W}$$

Using these quantities the equivalent noise charge can be written as

$$Q_n^2 = 4k_B T_0 \frac{\gamma_n}{W g_m'} \frac{F_v}{T} (C_d + C_s + W C_i')^2$$

where C_s is any stray capacitance present at the input in addition to the detector capacitance C_d and the FET capacitance WC_i'.

For $WC_i' = C_d + C_s$ the noise attains its minimum value

$$Q_{n,\min} = \sqrt{\frac{16 k_B T_0}{\kappa_n} \frac{F_v}{T} (C_d + C_s)}$$

where

$$\kappa_n \equiv \frac{g_m}{\gamma_n C_i}$$

is a figure of merit for the noise performance of the FET.

Example:

CMOS transistor with 1.2 μm channel length

At I_d/W = 0.3 A/m $g_m/C_i = 3 \cdot 10^{-9}$ s^{-1} and $\gamma_n = 1$.

For a CR-RC shaper with a 20 ns shaping time and an external capacitance

$C_d + C_s$ = 7.5 pF

$Q_{n,min}$ = 88 aC = 546 electrons,

achieved at a device width W = 5 mm, and a drain current of 1.5 mA.

The obtainable noise improves with the inverse square root of the shaping time, up to the point where $1/f$ noise becomes significant. For example, at T = 1 μs

$Q_{n,min}$ = 1.8 aC = 11 electrons,

although in practice additional noise contributions will increase the obtainable noise beyond this value.

Measured values of the noise coefficient γ_n for various *n*- and *p*-MOSFETs of various geometries for three normalized drain currents I_d/W.

Type	**NMOS**	**PMOS**	**NMOS**	**PMOS**	**NMOS**	**PMOS**	**NMOS**	**PMOS**
Width	75	75	1332	1332	888	888	1332	1332
Length	1.2	1.2	1.2	1.2	2.2	2.2	3.2	3.2
***I_d /W*= 0.03**								
0 Mrad			0.81	0.61	0.64	0.59	0.66	0.50
5 Mrad			2.17	0.84	1.00	0.58	1.50	0.69
***I_d /W*= 0.1**								
0 Mrad	1.10	0.70	1.20	1.10	0.80	0.80	0.80	0.60
5 Mrad	3.80	1.10	3.40	1.60	1.30	0.90	1.70	0.70
***I_d /W*= 0.3**								
0 Mrad	1.60	1.30	2.00	1.70	1.10	1.00	1.10	0.77
5 Mrad	5.00	2.90	4.80	2.70	1.60	1.40	1.20	0.81

Short channel *n*-MOSFETs tend to have higher noise coefficients at short channel lengths, probably due to increased electron temperature at high fields.

The table also shows the noise degradation after irradiation.

Since they are majority carrier devices, MOSFETs are insensitive to displacement damage, but they are affected by ionization damage, which leads to charge buildup in the oxide and the formation of interface states.

See H. Spieler, Introduction to Radiation-Resistant Semiconductor Devices and Circuits, tutorial, in A.H. Lumpkin, C.E. Eyberger (eds.) *Beam Instrumentation, Proceedings of the Seventh Workshop*, AIP Conference Proceedings 390, American Institute of Physics, Woodbury, NY, 1997, ISBN 1-56396-612-3

and http://www-atlas.lbl.gov/strips/strips.html

The preceding discussion has neglected $1/f$ noise, which adds a constant contribution independent of shaping time

$$Q_{nf}^2 \propto A_f C_{tot}^2$$

Although excess low frequency noise is determined primarily by the concentration of unwanted impurities and other defects, their effect in a specific technology is also affected by device size. For some forms of $1/f$ noise

$$A_f = \frac{K_f}{WL\, C_g^2}$$

where C_g is the gate-channel capacitance per unit area, and K_f is an empirical constant that is device and process dependent.

Typical values of the noise constant for various device types:

p-MOSFET	$K_f \approx 10^{-32}$ C^2/cm^2
n-MOSFET	$K_f \approx 4{\cdot}10^{-31}$ C^2/cm^2
JFET	$K_f \approx 10^{-33}$ C^2/cm^2

Specific implementations can improve on these values.

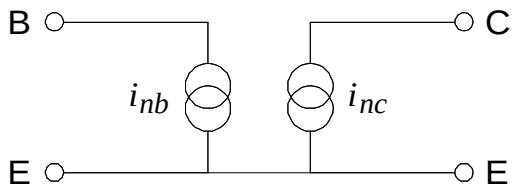
One should note that this model is not universally applicable, since excess noise usually does not exhibit a pure $1/f$ dependence; especially in PMOS devices one often finds several slopes. In practice, one must test the applicability of this parameterization by comparing it with data before applying it to scaled amplifiers.

Nevertheless, as a general rule, devices with larger gate area $W{\cdot}L$ tend to exhibit better "$1/f$" noise characteristics.

b) Noise in Bipolar Transistors

In bipolar transistors the shot noise from the base current is important.

The basic noise model is the same as shown before, but the magnitude of the input noise current is much greater, as the base current will be 1 – 100 μA rather than <100 pA.



The base current noise is shot noise associated with the component of the emitter current provided by the base.

$$i_{nb}^2 = 2q_e I_B$$

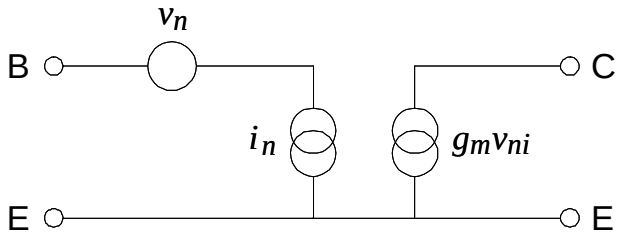
The noise current in the collector is the shot noise originating in the base-emitter junction associated with the collector component of the emitter current.

$$i_{nc}^2 = 2q_e I_C$$

Following the same argument as in the analysis of the FET, the output noise current is equivalent to an equivalent noise voltage

$$v_n^2 = \frac{i_{nc}^2}{g_m^2} = \frac{2q_e I_C}{(q_e I_C / k_B T)^2} = \frac{2(k_B T)^2}{q_e I_C}$$

yielding the noise equivalent circuit



where i_n is the base current shot noise i_{nb}.

The equivalent noise charge

$$Q_n^2 = i_n^2 F_i T + v_n^2 C_i^2 \frac{F_v}{T} = 2q_e I_B F_i T + \frac{2(k_B T)^2}{q_e I_C} C_{tot}^2 \frac{F_v}{T}$$

Since $I_B = I_C / \beta_{DC}$

$$Q_n^2 = 2q_e \frac{I_C}{\beta_{DC}} F_i T + \frac{2(k_B T)^2}{q_e I_C} C_{tot}^2 \frac{F_v}{T}$$

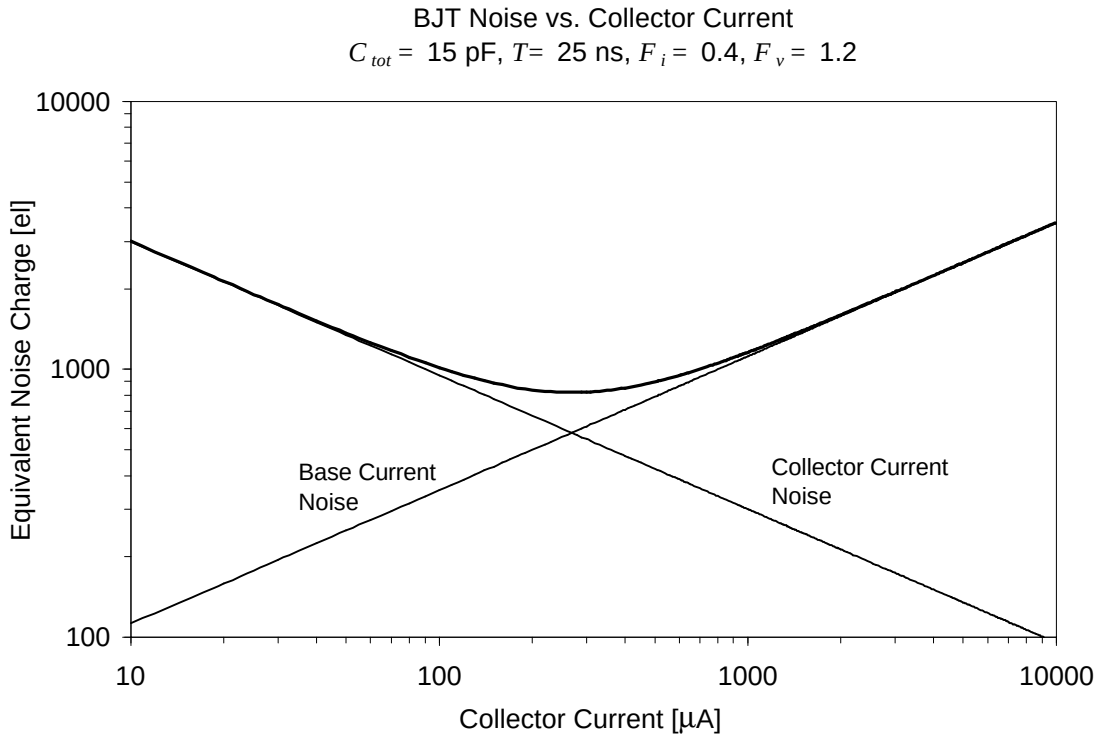
The current noise term increases with I_C, whereas the second (voltage) noise term decreases with I_C.

Thus the noise attains a minimum

$$Q_{n,\min}^2 = 4k_B T \frac{C_{tot}}{\sqrt{\beta_{DC}}} \sqrt{F_i F_v}$$

at a collector current

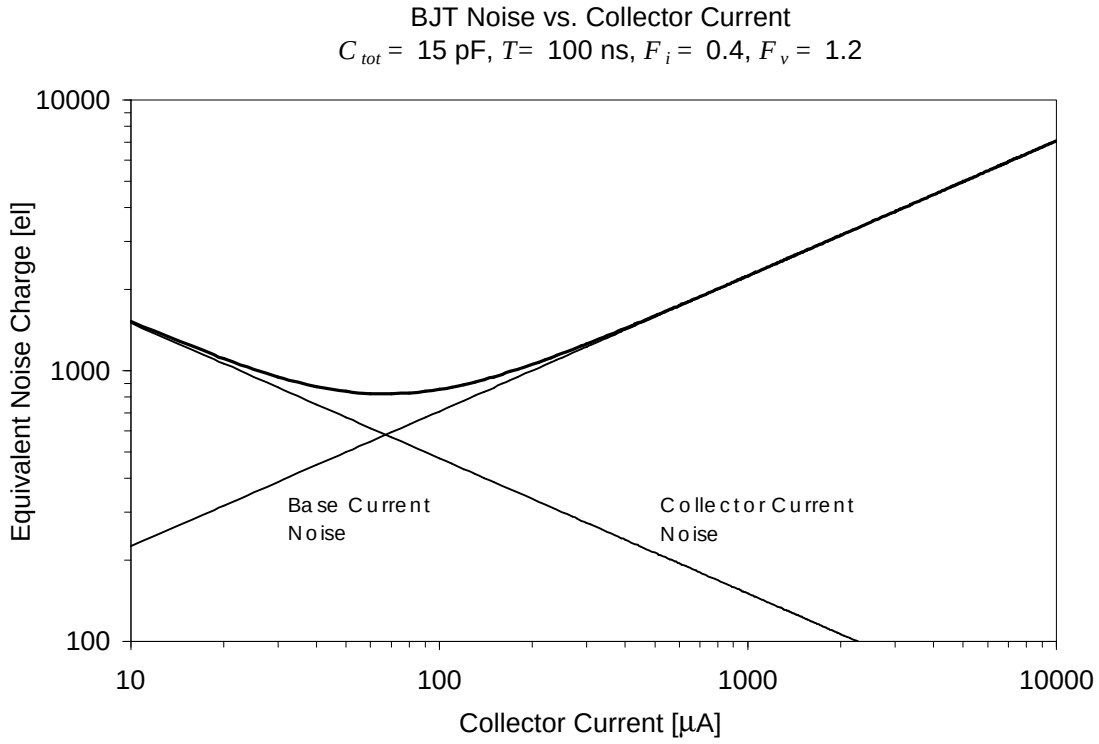
$$I_C = \frac{k_B T}{q_e} C_{tot} \sqrt{\beta_{DC}} \sqrt{\frac{F_v}{F_i}} \frac{1}{T}$$



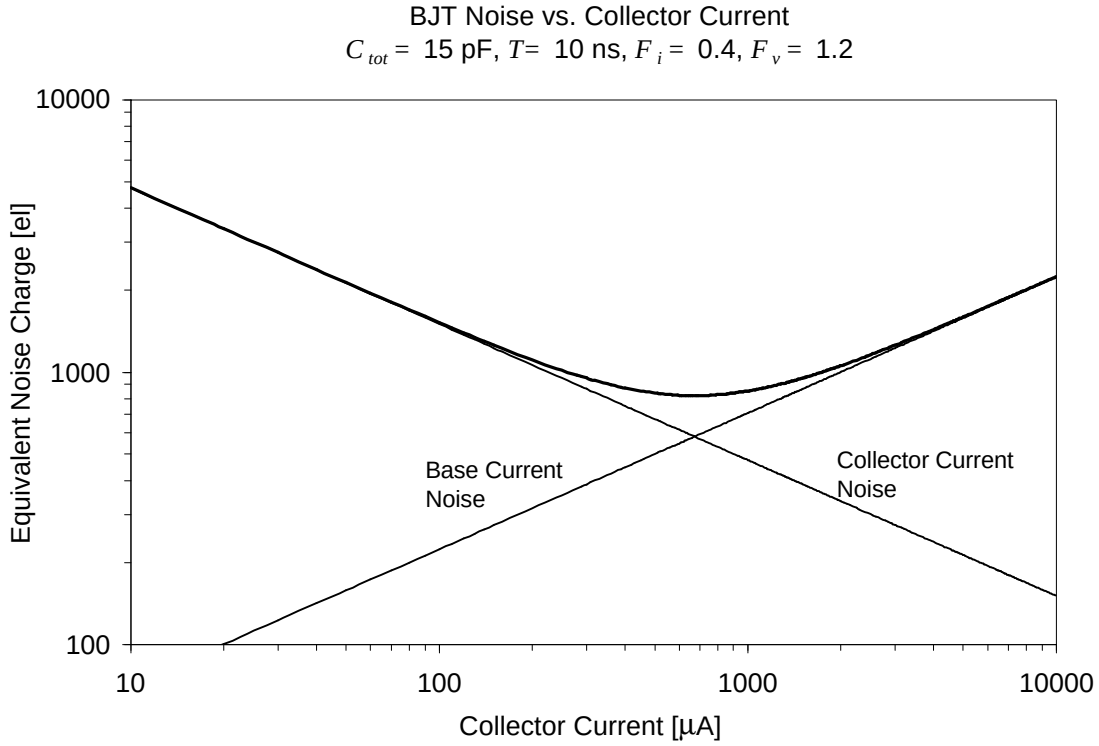
For a given shaper, the minimum obtainable noise is determined only by the total capacitance at the input and the DC current gain of the transistor, *not by the shaping time*.

The shaping time only determines the current at which this minimum noise is obtained

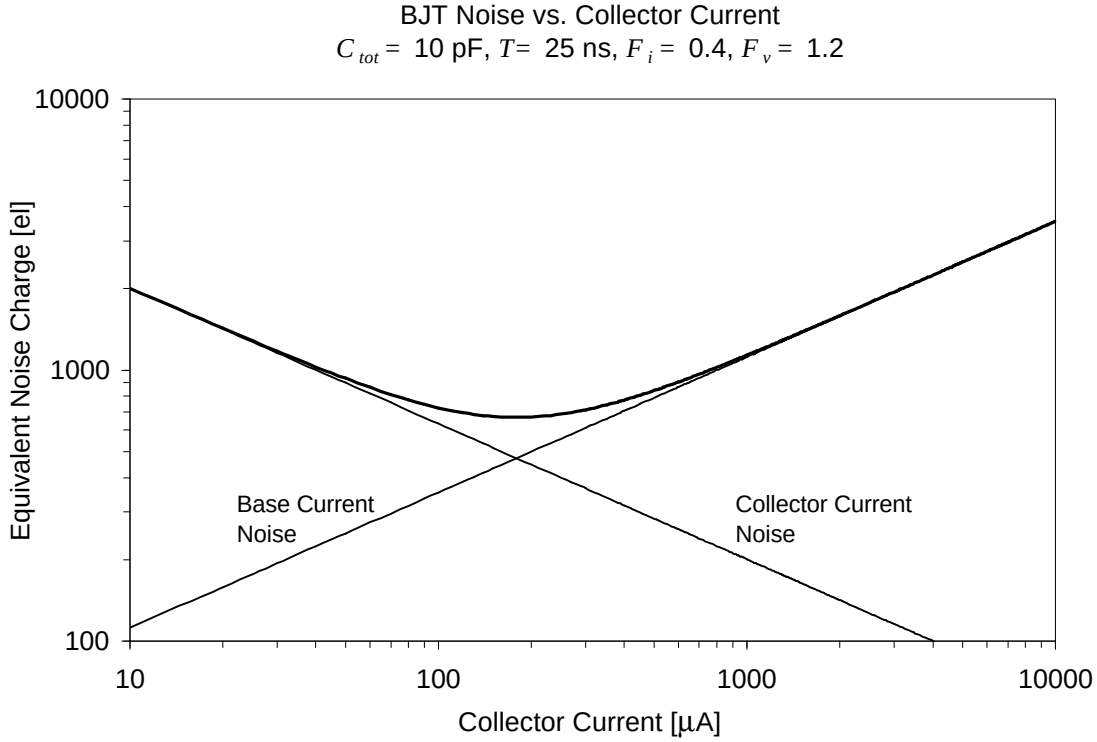
T= 100 ns

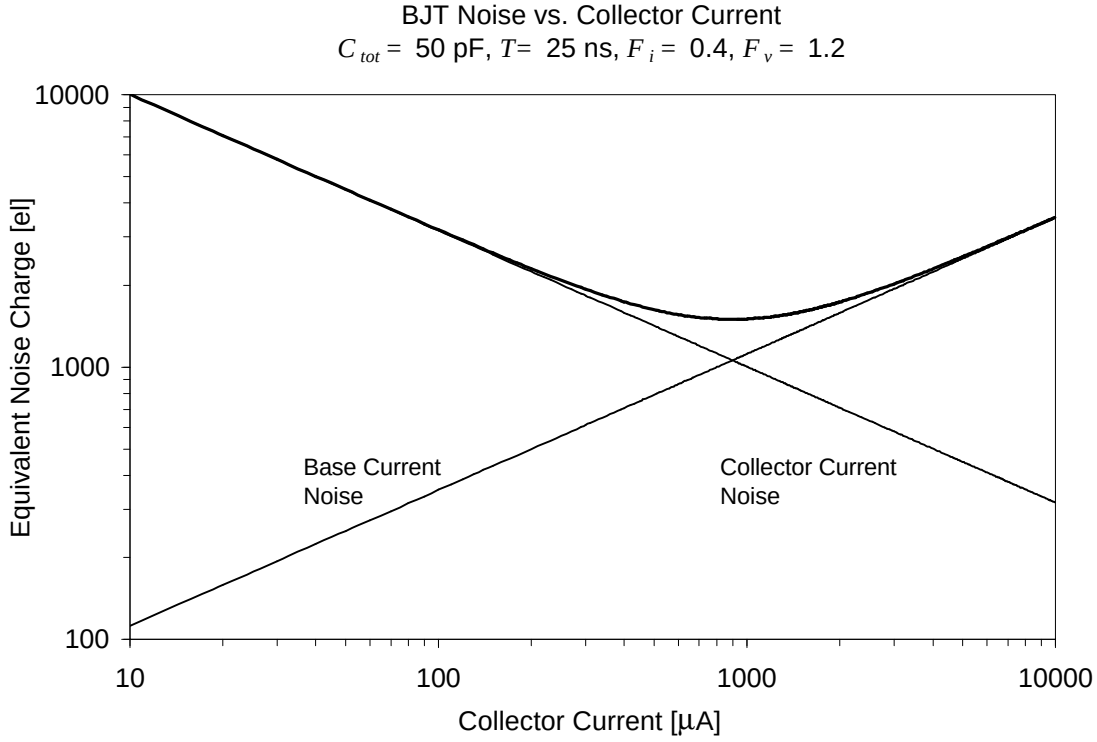


T= 10 ns



Increasing the capacitance at the input shifts the collector current noise curve upwards, so the noise increases and the minimum shifts to higher current.





Simple Estimate of obtainable BJT noise

For a CR-RC shaper

$$Q_{n,\min} = 772\left[\frac{el}{\sqrt{pF}}\right]\cdot\frac{\sqrt{C_{tot}}}{\sqrt[4]{\beta_{DC}}}$$

obtained at $$I_c = 26\left[\frac{\mu A\cdot ns}{pF}\right]\cdot\frac{C_{tot}}{\tau}\sqrt{\beta_{DC}}$$

Since typically $\beta_{DC} \approx 100$, these expressions allow a quick and simple estimate of the noise obtainable with a bipolar transistor.

Note that specific shapers can be optimized to minimize either the current or the voltage noise contribution, so both the minimum obtainable noise and the optimum current will be change with respect to the above estimates.

The noise characteristics of bipolar transistors differ from field effect transistors in four important aspects:

1. The equivalent input noise current cannot be neglected, due to base current flow.
2. The total noise does not decrease with increasing device current.
3. The minimum obtainable noise does not depend on the shaping time.
4. The input capacitance is usually negligible.

The last statement requires some explanation.

The input capacitance of a bipolar transistor is dominated by two components,

1. the geometrical junction capacitance, or transition capacitance C_{TE}, and
2. the diffusion capacitance C_{DE}.

The transition capacitance in small devices is typically about 0.5 pF.

The diffusion capacitance depends on the current flow I_E through the base-emitter junction and on the base width W, which sets the diffusion profile.

$$C_{DE} = \frac{\partial q_B}{\partial V_{be}} = \frac{q_e I_E}{k_B T}\left(\frac{W}{2D_B}\right) \equiv \frac{q_e I_E}{k_B T} \cdot \frac{1}{\omega_{Ti}}$$

where D_B is the diffusion constant in the base and ω_{Ti} is a frequency that characterizes carrier transport in the base. ω_{Ti} is roughly equal to the frequency where the current gain of the transistor is unity.

Inserting some typical values, I_E=100 μA and ω_{Ti} =10 GHz, yields C_{DE}= 0.4 pF. The transistor input capacitance C_{TE}+C_{DE}= 0.9 pF, whereas FETs providing similar noise values at comparable currents have input capacitances in the range 5 – 10 pF.

Except for low capacitance detectors, the current dependent part of the BJT input capacitance is negligible, so it will be neglected in the following discussion. For practical purposes the amplifier input capacitance can be considered constant at 1 ... 1.5 pF.

This leads to another important conclusion.

Since the primary noise parameters do not depend on device size and there is no significant linkage between noise parameters and input capacitance

- Capacitive matching does not apply to bipolar transistors.

Indeed, capacitive matching is a misguided concept for bipolar transistors. Consider two transistors with the same DC current gain but different input capacitances. Since the minimum obtainable noise

$$Q_{n,\min}^2 = 4k_B T \frac{C_{tot}}{\sqrt{\beta_{DC}}} \sqrt{F_i F_v} \ ,$$

increasing the transistor input capacitance merely increases the total input capacitance C_{tot} and the obtainable noise.

When to use FETs and when to use BJTs?

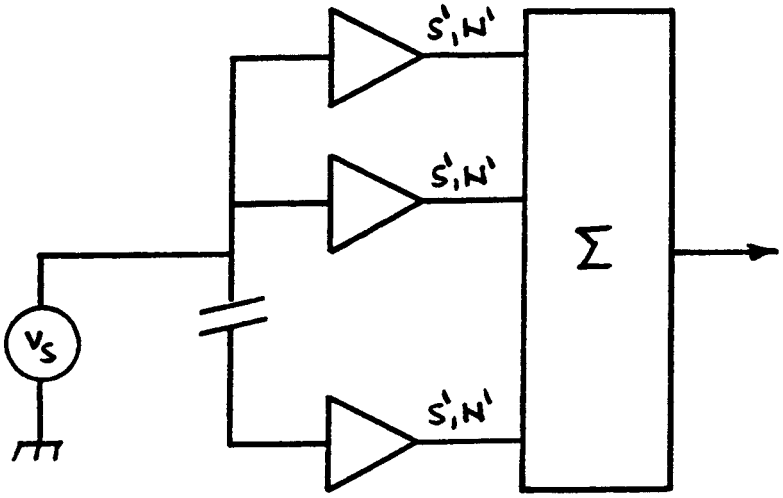
Since the base current noise increases with shaping time, bipolar transistors are only advantageous at short shaping times.

With current technologies FETs are best at shaping times greater than 50 to 100 ns, but decreasing feature size of MOSFETs will improve their performance.

Noise Optimization - Capacitive Matching Revisited

"Capacitive Matching" is often presented as a universal criterion for noise optimization. The results derived for bipolar transistors already show that capacitive matching does not apply in all amplifiers. This discussion is supposed to clarify where capacitive matching is useful and where it isn't.

Consider the array of amplifiers discussed in chapter VIII, but now both current and voltage noise will be considered. As in previous derivations of the equivalent noise charge, the amplifiers are assumed to have voltage-sensitive inputs. Furthermore, to simplify the analysis, the amplifiers do not utilize feedback.



Of course, in considering the current and voltage noise contributions, one can follow a formal argument based on the noise charge

$$Q_n^2 = i_n^2 F_i T + v_n^2 C_i^2 \frac{F_v}{T}$$

Since the current noise contribution does not depend on capacitance, matching the amplifier input capacitance to the detector capacitance should be irrelevant. On the other hand, since the voltage contribution does depend on capacitance, a correlation between v_n and C_i can yield an optimization condition.

Nevertheless, reviewing the formation of signal and noise in detail is useful to clarify the limits of capacitive matching.

1. Current Noise

For the noise currents originating in the individual amplifiers, the common connection to the signal source is a summing node, so if i_n' is the equivalent noise current of a single amplifier, for n amplifiers the total input noise current flowing through the signal source impedance is

$$i_n = \sqrt{n} \cdot i_n'$$

The flow of this current through the input impedance Z_i formed by the parallel connection of the detector capacitance and amplifier capacitances gives rise to a noise voltage

$$v_n = i_n Z_i$$

This voltage is applied in parallel to all amplifier inputs, so at the output of an individual amplifier (assuming gain A) the noise level is

$$N'(n) = Av_n = A\sqrt{n}\, i_n' Z_i$$

At the output of the summing circuit, the cumulative noise from all amplifier outputs becomes

$$N(n) = \sqrt{n}N'(n) = An\, i_n' Z_i$$

Since the amplifiers respond to voltage, the magnitude of the signal applied to all amplifiers is the same, which for a signal current i_s is

$$v_s = i_s Z_i$$

In the summed output the signals add coherently, so that

$$S(n) = nAi_s Z_i$$

and the signal-to-noise ratio

$$\frac{S(n)}{N(n)} = \frac{nAi_s Z_i}{Ani_n' Z_i} = \frac{i_s}{i_n'} = \frac{S(1)}{N(1)} \ ,$$

the same as for a single amplifier

Paralleling amplifiers does not affect the signal-to-noise ratio if only current noise is present.

Varying the amplifier input capacitance is irrelevant. As the total input capacitance increases, the noise voltage developed at the input decreases with Z_i, but so does the signal voltage, so the signal-to-noise ratio is unaffected.

2. Voltage Noise

The voltage noise contribution differs from the current noise in an important aspect:

- Voltage noise is not additive at the input.

This statement can be justified with two arguments, the first more physical and the second more formal.

1. Voltage noise tends to originate within a device (e.g. thermal noise of an FET channel or collector shot noise in a BJT) and appears as a noise current at the output, which is mathematically transformed to the input. This noise voltage is not physically present at the input and is not affected by any connections or components in the input circuit.

2. The noise voltage sources that represent all voltage noise contributions of a given amplifier are in series with the individual inputs. Since the input impedance of the amplifier is postulated to be infinite, the source impedance is by definition negligible in comparison, so the noise voltage associated with a given amplifier only develops across the input of that amplifier.

Assume that each amplifier has an input referred noise v_n'and an input capacitance C_i'.

Then the input signal voltage

$$v_s = \frac{Q_s}{C}$$

where C is the total input capacitance including the detector

$$C = C_{\text{det}} + nC_i'$$

The signal at each amplifier output is

$$S' = Av_s = A\frac{Q_s}{C_{\text{det}} + nC_i'}$$

The noise at each amplifier output is

$$N' = Av_n'$$

After summing the n outputs the signal-to-noise ratio

$$\frac{S(n)}{N(n)} = \frac{nS'}{\sqrt{n}N'} = \sqrt{n}\frac{S'}{N'} = \sqrt{n}\frac{A\dfrac{Q_s}{C_{\text{det}} + nC_i'}}{Av_n'} = \frac{Q_s}{v_n'/\sqrt{n}} \cdot \frac{1}{C_{\text{det}} + nC_i'}$$

which assumes a maximum when $C_{\text{det}} = nC_i$'.

Under this "capacitive matching" condition $\Sigma C_i' = C_{\text{det}}$ the signal-to-noise ratio

$$\frac{S}{N} = \frac{Q_s}{v_n'}\sqrt{\frac{C_{\text{det}}}{C_i'}}\frac{1}{C_{\text{det}} + nC_i'} = \frac{Q_s}{v_n'}\sqrt{\frac{C_{\text{det}}}{C_i'}}\frac{1}{2C_{\text{det}}}$$

or

$$\frac{S}{N} = \frac{1}{2}\frac{Q_s}{v_n'\sqrt{C_i'}} \cdot \frac{1}{\sqrt{C_{\det}}}$$

Since v_n' and C_i' are properties of the individual amplifier, i.e. constants, the signal-to-noise ratio decreases with the square root of detector capacitance

This relationship only holds if

1. the noise of the input amplifier/device decreases with increasing input capacitance.

2. the input capacitance is scaled with the detector capacitance ("capacitive matching")

- The first point is critical; if the noise voltage of a device and its input capacitance are not correlated, capacitive matching is deleterious.

Example: A MOSFET operated in weak inversion has a transconductance that depends only on current, independent of geometry. If power consumption is to be kept constant, increasing the size of the device at the same operating current will not increase the transconductance, but it will increase the input capacitance and, as a result, the equivalent noise charge.

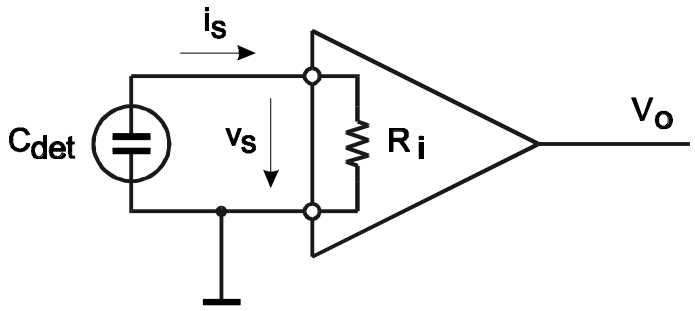
For both BJTs and FETs, the minimum obtainable noise increases with the square root of detector capacitance, although the physical origins for this behavior are quite different in the two types of devices.

S/N vs. Capacitance in a Current-Sensing Amplifier

The dependence of signal-to-noise ratio on capacitance was shown for voltage and charge-sensing amplifiers. The same result holds for current sensing amplifiers, but for different reasons.

Assume that the current induced on the detector electrodes is a delta pulse with charge Q_s.

The measured current pulse depends on the input time constant $\tau = C_{det}R_i$ that discharges the detector capacitance.



The signal current is flowing into the amplifier is

$$i_s = i_0 e^{-t/\tau}$$

Since $Q_s = \int_0^\infty i_s dt \Rightarrow \quad i_0 = Q_s/\tau$ and $\quad i_s = \frac{Q_s}{\tau} e^{-t/\tau}$

Assume that only the peak signal at $t = 0$ will be sensed, so $i_s = Q_s/\tau$.

If the equivalent input noise current of the amplifier is i_n and the noise bandwidth Δf, the signal-to-noise ratio is

$$\frac{S}{N} = \frac{Q_s/\tau}{i_n\sqrt{\Delta f}} = \frac{Q_s}{C_{\text{det}}} \frac{1}{R_i i_n \sqrt{\Delta f}}$$

$\Rightarrow$ *For a given amplifier system* (parameters R_i, i_n, Δf), the signal-to-noise ratio of the current signal is inversely proportional to the total input capacitance.

Decreasing the input resistance will increase the peak current.

Will this improve the signal-to-noise ratio?

Consider the simplest form of a low-impedance input, a common base bipolar transistor or a common gate FET.

For both transistor types the input resistance is equal to the reciprocal transconductance

$$R_i = \frac{1}{g_m}$$

A bipolar transistor provides the most favorable condition for a given current

$$g_m \equiv \frac{dI_C}{dV_{BE}} = \frac{q_e}{k_B T} I_C$$

$$R_i = \frac{dV_{EB}}{dI_E} \approx \frac{dV_{EB}}{dI_C} = \frac{1}{g_m} = \frac{k_B T}{q_e} \frac{1}{I_C}$$

$\Rightarrow$ Decreasing the input resistance requires an increase of the collector current.

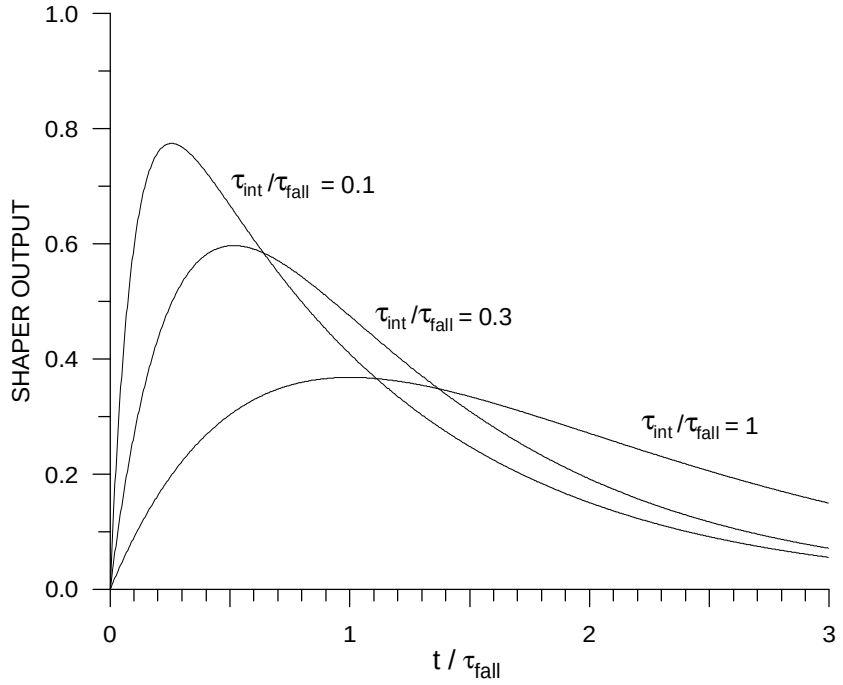
Increasing the collector current also increases the input noise current

$$i_n^2 = 2q_e(I_B + I_C) \approx 2q_e I_C$$

Decreasing the pulse width requires a faster shaper, which increases the noise bandwidth Δf.

Too large a bandwidth will increase the noise without increasing the signal.

Increasing the amplifier time constant decreases the peak signal.

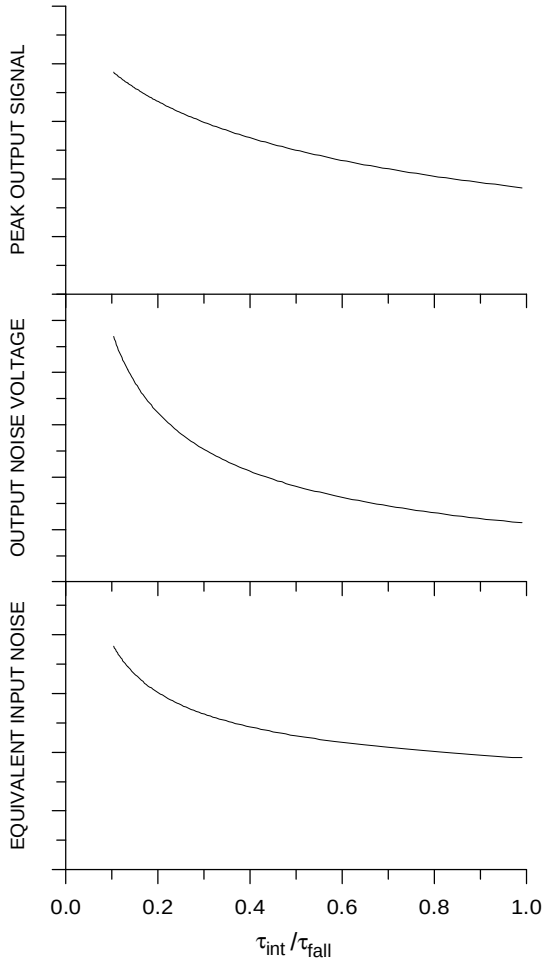


However …

as the signal decreases,

the noise drops even more,

so the equivalent input noise improves.



Assume the optimum shaping time $\tau = R_i C_{det}$ with a noise bandwidth $\Delta f = 1/4\tau = 1/4R_i C_{det}$. In any case the optimum noise bandwidth will scale with the input time constant τ.

Including the signal loss by $1/e$ this yields the signal-to-noise ratio

$$\frac{S}{N} = \frac{Q_s}{C_{\det}} \frac{1/e}{R_i i_n \sqrt{\Delta f}} = \frac{Q_s / e}{i_n R_i C_{\det} \sqrt{1/4R_i C_{\det}}} = \frac{2Q_s / e}{i_n \sqrt{R_i C_{\det}}}$$

$$\frac{S}{N} = \frac{Q_s}{\sqrt{C_{\det}}} \frac{2}{e k_B T} \propto \frac{Q_s}{\sqrt{C_{\det}}}$$

$\Rightarrow$ S/N scales inversely with $\sqrt{C_{\det}}$, as shown earlier for an optimized BJT or FET amplifier.

This analysis used some simplifications to better illustrate the various contributions to the equivalent noise charge.

An exact analysis yields the same result as obtained previously for FETs and BJTs.

Optimization for Low-Power

Optimizing the readout electronics in large vertex or tracking detector systems is not optimizing one characteristic, e.g. noise, alone, but finding an optimum compromise between noise, speed, and power consumption.

The minimum obtainable noise values obtained from the equations for both FETs and BJTs should be viewed as limits, not necessarily as desirable goals, since they are less efficient than other operating points.

First, consider two input transistors, which

provide the same overall noise with a given detector,

but differ in input capacitance.

Since the sum of detector and input capacitance determines the voltage noise contribution, the device with the higher input capacitance must have a lower equivalent noise voltage v_n, i.e. operate at higher current.

In general,

- low capacitance input transistors are preferable, and
- systems where the total capacitance at the input is dominated by the detector capacitance are more efficient than systems that are capacitively matched.

Capacitive matching should be viewed as a limit, not as a virtue.

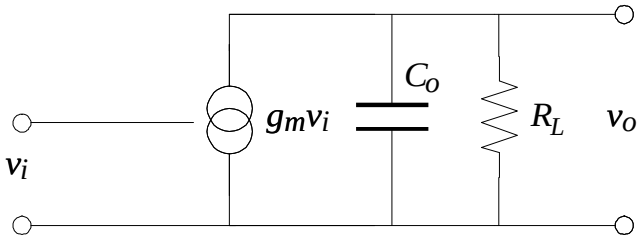
What is the optimum operating current for a given device?

Both response time, i.e. bandwidth, and noise depend on a common parameter, transconductance.

The relationship between noise and transconductance was shown above.

The dependence of bandwidth on transconductance is easy to derive.

Consider an amplifying device with transconductance g_m and a load resistance R_L.



The total capacitance at the output node is C_o.

The low frequency voltage gain is

$$A_v = \frac{v_o}{v_i} = g_m R_L$$

The bandwidth of the amplifier is determined by the output time constant

$$\tau_o = R_L C_o = \frac{1}{\omega_o}$$

Hence the gain-bandwidth product

$$A_v \omega_o = g_m R_L \cdot \frac{1}{R_L C_o} = \frac{g_m}{C_o}$$

is independent of the load resistance R_L, i.e. the voltage gain, but depends only on the device transconductance g_m and the capacitance at the output C_o.

The capacitance at the output node C_o depends on circuit topology and basic characteristics of the IC technology used. Often, the bandwidth is determined less by the inherent device speed, than by the stray capacitance to the substrate.

Since increasing transconductance yields both improved bandwidth and noise, a useful figure of merit for low power operation is the ratio of transconductance to device current g_m / I.

In a bipolar transistor

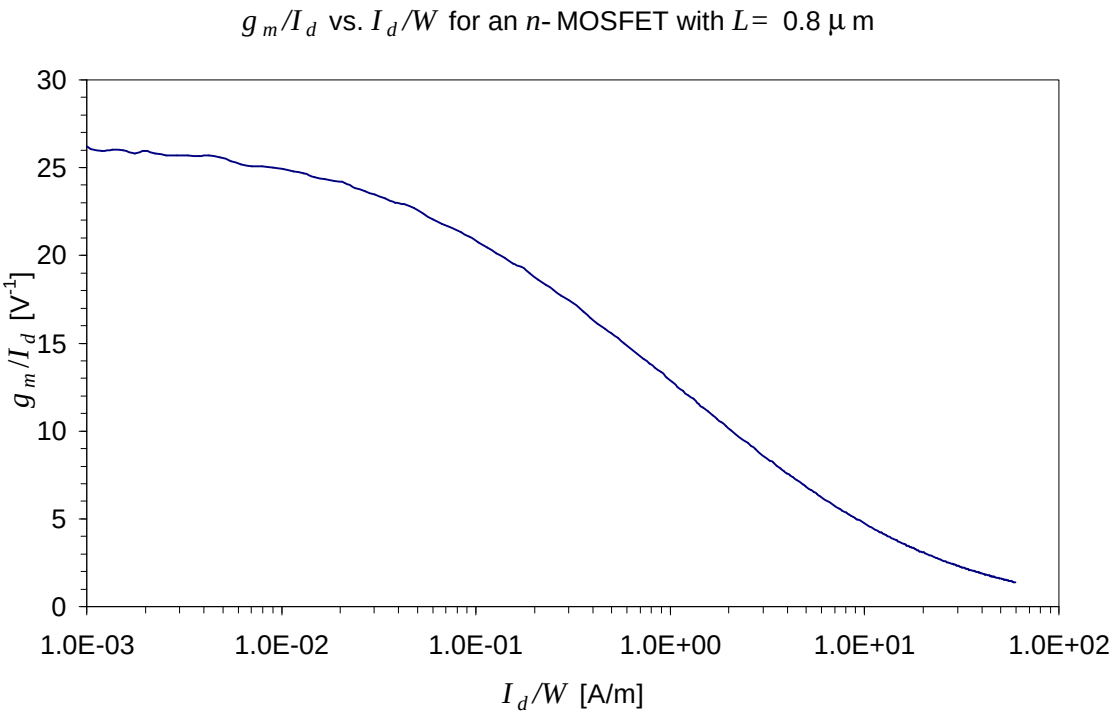
$$g_m = \frac{q_e}{k_B T} I_C$$

so g_m / I_c is constant

$$\frac{g_m}{I_C} = \frac{q_e}{k_B T}$$

In an FET the dependence of transconductance on drain current is more complicated.

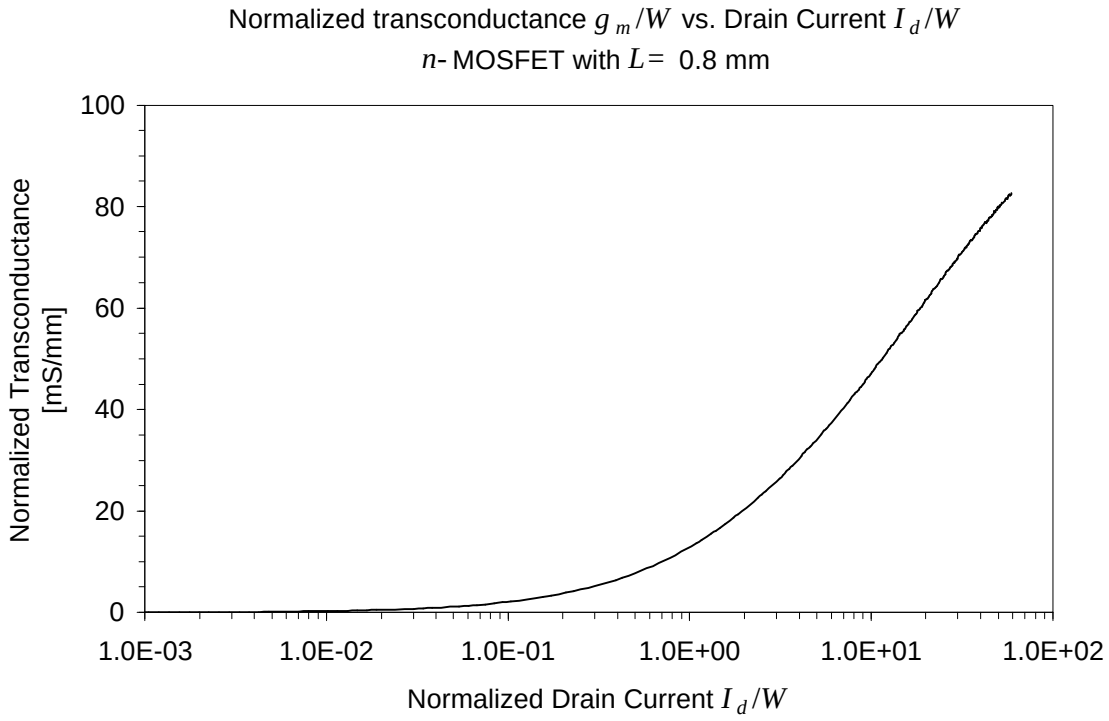
The figure below shows g_m/I_d as a function of normalized drain current I_d/W for a MOSFET with 0.8 µm channel length. This is a universal curve for all transistors using the same technology and channel length.



At low currents the MOSFET starts out with constant g_m/I_d, equal to a bipolar transistor. This is the weak inversion regime.

One then sees a rapid decrease in the regime 0.1 < I_d < 10 A/m (moderate inversion) and finally a gradual decrease at I_d > 10 A/m (strong inversion).

Note that although g_m/I_d is decreasing with current, the transconductance itself is increasing, but at a substantial penalty in current.



The strong inversion regime is most commonly used, especially when minimum noise is required, since it yields the highest transconductance.

Note, however, that the abscissa is logarithmic, and that the high transconductance in strong inversion comes at the expense of substantial power.

In systems where both speed and noise must be obtained at low power, for example HEP tracking detectors, the moderate inversion regime is advantageous, as it still provides 20 to 50% of the transconductance at 1/10 the power.

Having chosen a value of normalized drain current I_d/W that provides an adequate gain-bandwidth product, the required noise can be achieved adjusting the width of the FET, limited by capacitive matching.

Since in this scaling the current density remains constant, both the current and the transconductance increase proportionally with width. Thus the noise

$$v_n^2 \propto \frac{1}{g_m} \propto \frac{1}{I_D}$$

$$Q_n^2 \propto v_n^2 C_{tot}^2 \propto \frac{C_{tot}^2}{I_D}$$

so for a given noise level the required power

$$P_D \propto I_D \propto C_{tot}^2$$

A similar result obtains for bipolar transistors. The most efficient operating regime with respect to power is below the current for minimum noise

$$I_C = \frac{k_B T}{q_e} C_{tot} \sqrt{\beta_{DC}} \sqrt{\frac{F_v}{F_i}} \frac{1}{T}$$

In this regime the noise is dominated by voltage noise, so

$$Q_n^2 \approx \frac{2(k_B T)^2}{q_e I_C} C_{tot}^2 \frac{F_v}{T}$$

and as above for a given noise level

$$P \propto I_C \propto C_{tot}^2$$

The capacitance of the detector element strongly affects the required current in the input transistor, so reducing the capacitance not only improves noise performance, but also reduces power consumption.

Comparison: Power Dissipation of a Random Access Pixel Array vs. Strip Readout

If a strip readout for the LHC requires 2 mW per strip on an 80 μm pitch, i.e. 250 mW/cm width, is it practical to read out 15000 pixels per cm^2?

strip detector: n strips
pixel detector: n x n pixels

The capacitance is dominated by the strip-strip or pixel-pixel fringing capacitance.

$\Rightarrow$ capacitance proportional to periphery (pitch p and length l)

$$C \propto 2(l+p) \;\Rightarrow\; C_{pixel} \approx \frac{2}{n} C_{strip}$$

In the most efficient operating regime the power dissipation of the readout amplifier for a given noise level is proportional to the square of capacitance (discussed in VIII.5)

$$P \propto C^2$$

$\Rightarrow$

$$P_{pixel} \approx \frac{4}{n^2} P_{strip}$$

n times as many pixels as strips

$\Rightarrow$

$$P_{pixel,tot} \approx \frac{4}{n} P_{strip}$$

$\Rightarrow$ Increasing the number of readout channels can reduce the total power dissipation!

The circuitry per cell does not consist of the amplifier alone, so a fixed power P_0 per cell must be added, bringing up the total power by $n^2 P_0$, so these savings are only realized in special cases.

Nevertheless, random addressable pixel arrays can be implemented with overall power densities comparable to strips.