Semiconductor Technology from A to Z

Devices

Characteristic Curves and Transistor Equations

Characteristic Curve Fields and Operating Point

A transistor has three terminals, and its behavior depends on two voltages and two currents. To present these dependencies clearly, they are resolved into characteristic curve fields: one curve shows the relationship between two quantities while a third is held fixed as a parameter. For the bipolar transistor these are the input characteristic $I_B(U_{BE})$, the transfer characteristic $I_C(U_{BE})$ or $I_C(I_B)$, and the output characteristic field $I_C(U_{CE})$ with I_B as parameter. For the MOSFET the input characteristic disappears, since the gate current is practically zero in static operation; what remains are the transfer characteristic $I_D(U_{GS})$ and the output characteristic field $I_D(U_{DS})$ with U_{GS} as parameter.

Characteristic curves describe the device alone. Only the external circuit determines where on the curve the transistor actually operates. For a load resistor R_L between supply voltage U_{DD} and drain, $U_{DS} = U_{DD} - I_D \cdot R_L$ holds – a straight line in the output characteristic field, the load line. Its intersection with the curve for the currently applied gate voltage is the operating point. If U_{GS} changes, the operating point moves along the load line; this movement is precisely the amplification mechanism of a transistor. The location of the operating point determines whether a transistor acts as a linear amplifier, a switch, or a current source.

Transistor equations are mathematical approximations that describe these curves with a few parameters. They allow circuits to be calculated by hand, but they are models with a limited range of validity. The models in circuit simulators such as SPICE are considerably more complex, yet build on these same underlying equations.

Characteristic Curves of the Bipolar Transistor

The basic structure of a bipolar transistor is covered in the chapter Bipolar Transistors; the following describes its electrical behavior in operation.

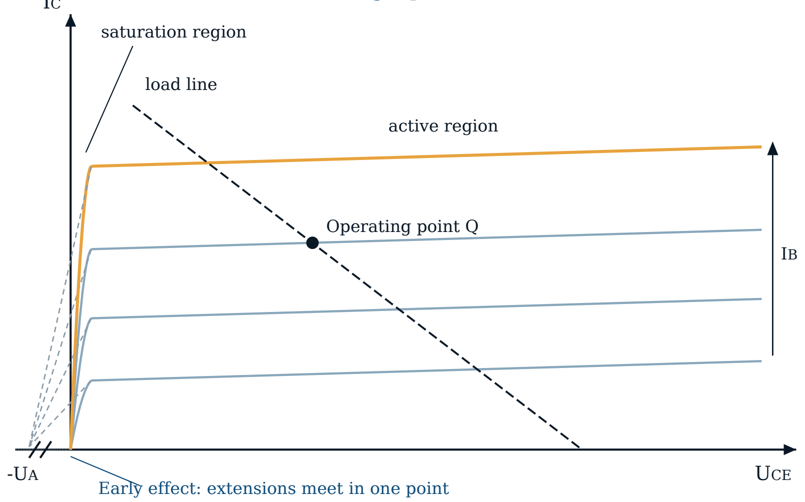
The input characteristic of an npn transistor is the diode characteristic of the forward-biased base-emitter junction: base current rises exponentially with U_{BE}, becoming significant only above about 0.6 V for silicon. Because collector current is coupled to base current through current gain B, the transfer characteristic $I_C(U_{BE})$ is likewise exponential. A change in U_{BE} of only 60 mV multiplies the collector current tenfold at room temperature – the bipolar transistor is an extremely steep, but also temperature-sensitive, converter of voltage into current.

The output characteristic field $I_C(U_{CE})$ shows, for each base current, a curve that rises steeply at small U_{CE} and then transitions into a nearly horizontal branch. The steep region below about 0.2 V is the saturation region: both pn junctions are forward-biased, and the transistor behaves like a closed switch with a small residual voltage U_{CEsat}. In the horizontal branch, the active region, the base-collector junction is reverse-biased, and collector current is determined almost entirely by base current – here the transistor operates as an amplifier. The slight upward slope of the curves in the active region is the Early effect: as U_{CE} increases, the space-charge region of the collector junction widens into the base, the effective base width shrinks, and collector current rises. Extending the curves to the left, they intersect the voltage axis approximately at a common point, the Early voltage $-U_A$ (typically 50 to 150 V).

With no base current, only a negligible residual current flows – the transistor is off. At sufficiently high collector-emitter voltage, avalanche breakdown of the collector junction eventually sets in; the breakdown voltage U_{CEO} marks the upper limit of the permissible operating range.

Output Characteristic Field of a Bipolar Transistor

Collector current vs. collector-emitter voltage, parameter: base current



Load line and characteristic curve intersect at the operating point; the slight upward slope in the active region shows the Early effect.

Equations of the Bipolar Transistor

In the active region, collector current follows the Shockley equation of the base-emitter junction:

$$I_C = I_S \cdot (\exp(U_{BE} / U_T) - 1) \approx I_S \cdot \exp(U_{BE} / U_T)$$

The saturation current I_S is a device constant in the range of 10^{-15} to 10^{-12} A and depends on emitter area, base doping, and base width. The thermal voltage $U_T = k \cdot T/q$ is about 26 mV at 300 K; it is the reason the characteristic is so steep and why bipolar transistors change so strongly with temperature. Base current follows from $I_B = I_C / B$ with the DC current gain B (usually h_{FE} in datasheets), which ranges from 50 to several hundred depending on type. Emitter current is the sum of both: $I_E = I_C + I_B = I_C \cdot (1 + 1/B)$.

The Early effect is captured by a correction factor:

$$I_C = I_S \cdot \exp(U_{BE} / U_T) \cdot (1 + U_{CE} / U_A)$$

This equation directly yields the output resistance in the active region, $r_{CE} \approx U_A / I_C$ – at 1 mA collector current and $U_A = 100$ V, roughly 100 kΩ. The boundary to saturation lies where U_{CE} falls below U_{BE} and the collector junction becomes conducting; in saturation the exponential equation no longer applies, and collector current is limited by the external circuit rather than by the transistor itself.

The complete equation system for all four operating modes (active, inverse, saturation, cutoff) is given by the Ebers-Moll model of 1954, which describes the transistor as two back-to-back diodes with controlled current sources. For hand calculation, the active-region approximation above is almost always sufficient.

Characteristic Curves of the MOSFET

The basic structure of a MOSFET is covered in the chapter Field-Effect Transistors; the following describes its electrical behavior in operation.

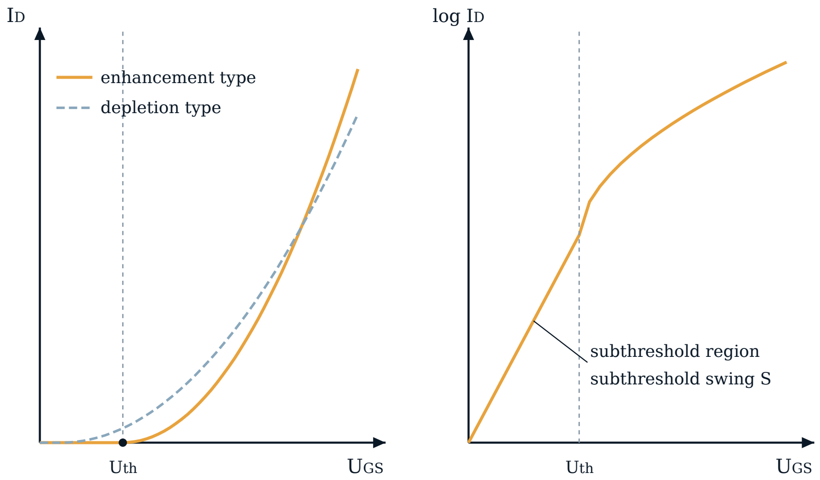
The transfer characteristic $I_D(U_{GS})$ of an n-channel enhancement MOSFET stays essentially at zero up to the threshold voltage U_{th} and then rises, initially quadratically, above it. The threshold voltage is the gate voltage at which an inversion layer forms at the oxide-semiconductor interface, creating a conductive channel between source and drain; it lies between 0.2 and 0.5 V for modern logic transistors and between 2 and 4 V for power MOSFETs. A depletion-mode MOSFET, by contrast, has a negative threshold voltage and already conducts at $U_{GS} = 0$; its transfer characteristic is simply shifted to the left – an overview of the different gate principles and operating modes is given in the chapter Field-Effect Transistor Families: An Overview. Plotting I_D logarithmically reveals that current below U_{th} does not end abruptly but decays exponentially – the subthreshold region, which is decisive for leakage currents and ultra-low-power circuits.

The output characteristic field $I_D(U_{DS})$ divides into two regions. For small drain-source voltages, current rises nearly linearly with U_{DS} – the transistor behaves like a resistor controlled by the gate, which is why this section is called the resistive or triode region. As U_{DS} increases, the voltage between gate and the drain end of the channel decreases, the channel narrows there, and at $U_{DS} = U_{GS} - U_{th}$ it pinches off completely at the drain. Beyond this point, current remains nearly constant: the saturation region. The name is historical and somewhat misleading, since for the MOSFET it denotes exactly the region called "active" in the bipolar transistor – for the bipolar transistor, saturation instead means the switch-on state at small U_{CE}.

The boundary between the two regions forms a parabola $I_D = (\beta/2) \cdot U_{DS}^2$ through the pinch-off points of all curves. As with the bipolar transistor, the curves rise slightly in the saturation region, here due to channel-length modulation: the pinch-off point moves toward the source as U_{DS} increases, the effective channel shortens, and current increases.

Transfer Characteristic of the MOSFET

Drain current versus gate-source voltage, linear and logarithmic



Below threshold, drain current does not drop to zero abruptly but decays exponentially – visible only on a logarithmic scale.

Equations of the MOSFET

The simplest analytical model of the long-channel MOSFET goes back to Shockley and is implemented in every simulator as the Level-1 or Shichman-Hodges model – the same square-law form also appears in the saturation-current equation of the junction FET (JFET). In the resistive region ($U_{DS} < U_{GS} - U_{th}$):

$$I_D = \beta \cdot [(U_{GS} - U_{th}) \cdot U_{DS} - U_{DS}^2 / 2]$$

In the saturation region ($U_{DS} \geq U_{GS} - U_{th}$):

$$I_D = (\beta/2) \cdot (U_{GS} - U_{th})^2 \cdot (1 + \lambda \cdot U_{DS})$$

The gain factor $\beta = \mu_n \cdot C_{ox} \cdot W/L$ combines all technology and geometry quantities: the electron mobility μ_n in the channel, the area-specific gate-oxide capacitance $C_{ox} = \varepsilon_{ox} / t_{ox}$, and the width-to-length ratio W/L. The circuit designer can only choose W/L; μ_n and C_{ox} are fixed by the process. The parameter λ (typically 0.01 to 0.1 V^{-1}) describes channel-length modulation and is the counterpart to the Early voltage: 1/λ corresponds to U_A. The difference $U_{GS} - U_{th}$ is called the overdrive voltage U_{ov} and is the central design quantity for the saturation region.

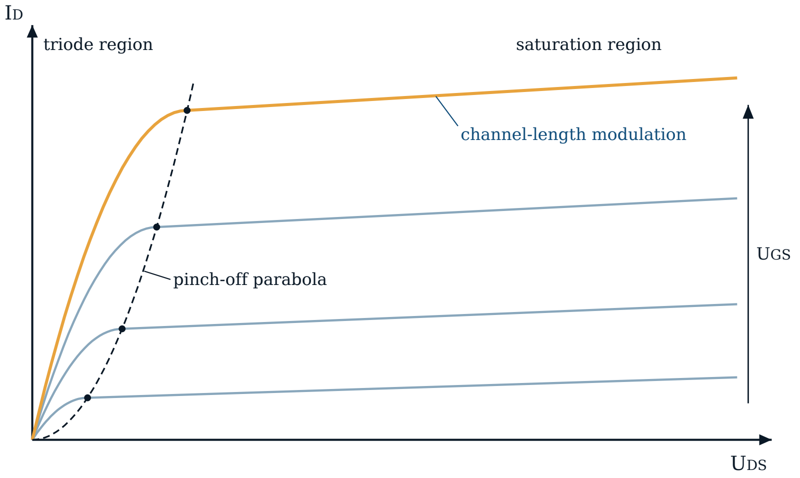
Below the threshold voltage, current follows an exponential law:

$I_D \approx I_0 \cdot \exp((U_{GS} - U_{th}) / (n \cdot U_T))$

with the subthreshold factor n between 1 and 1.5. The slope is expressed as the subthreshold swing $S = n \cdot U_T \cdot \ln 10$, at least 60 mV per decade at room temperature. This physical limit determines how far U_{th} can be lowered without unacceptably raising off-state current – one of the main reasons the supply voltage of digital circuits cannot be reduced indefinitely. If a reverse voltage U_{SB} exists between source and substrate, the threshold voltage increases via the body effect by $\gamma \cdot (\sqrt{(2\varphi_F + U_{SB})} - \sqrt{(2\varphi_F)})$, with the body-effect coefficient γ and the Fermi potential φ_F.

Output Characteristic Field of the MOSFET

Drain current vs. drain-source voltage, parameter: gate-source voltage



The pinch-off parabola separates the triode and saturation regions; the slight upward slope in saturation results from channel-length modulation (parameter λ).

Small-Signal Parameters and Limits of the Models

For amplifier design, what matters is not the full characteristic curve but its behavior in the immediate vicinity of the operating point. There the curve can be replaced by its tangent, and the small-signal parameters follow from the equations by differentiation. Transconductance g_m indicates how strongly the output current responds to a small change in input voltage. For the bipolar transistor it is proportional to collector current, $g_m = I_C / U_T$ – about 38 mS at 1 mA, independent of the specific device. For the MOSFET in saturation, $g_m = \beta \cdot U_{ov} = 2 \cdot I_D / U_{ov}$; it grows only with the square root of drain current and, at equal current, is typically an order of magnitude below that of the bipolar transistor. The output resistance $r_o = U_A / I_C$ or $1 / (\lambda \cdot I_D)$ limits the maximum achievable voltage gain of a stage to $g_m \cdot r_o$, the so-called intrinsic gain.

The simple equations hold for channel lengths above roughly one micrometer. At shorter channel lengths, charge carriers reach their saturation velocity even at moderate field strengths, and drain current no longer rises quadratically but only linearly with the overdrive voltage. In addition, drain-induced barrier lowering (DIBL) reduces U_{th} as U_{DS} rises, and mobility degradation occurs due to the high vertical field under the gate. Simulation models such as BSIM4 or BSIM-CMG for FinFETs capture these effects with several hundred parameters, extracted by semiconductor manufacturers from measurements on test structures and delivered to circuit designers as a process design kit. For understanding circuits, however, the equations in this chapter remain the foundation: they give the correct trends and orders of magnitude, even where the numerical values in the nanometer regime are only qualitatively accurate.