THRESHOLD VOLTAGE

The **threshold voltage** of a MOSFET is usually defined as the gate voltage where an inversion layer forms at the interface between the insulating layer (oxide) and the substrate (body) of the transistor. The purpose of the inversion layer's forming is to allow the flow of electrons through the gate-source junction. The creation of this layer is described next. In an n-MOSFET the substrate of the transistor is composed of p-type silicon (see doping (semiconductor)), which has positively charged mobile holes as carriers. When a positive voltage is applied on the gate, an electric field causes the holes to be repelled from the interface, creating a depletion region containing immobile negatively charged acceptor ions. A further increase in the gate voltage eventually causes electrons to appear at the interface, in what is called an inversion layer, or channel. Historically the gate voltage at which the electron density at the interface is the same as the hole density in the neutral bulk material is called the threshold voltage. Practically speaking the threshold voltage is the voltage at which there are sufficient electrons in the inversion layer to make a low resistance conducting path between the MOSFET source and drain.

In the figures, the source (left side) and drain (right side) are labeled $n+$ to indicate heavily doped (blue) n-regions. The depletion layer dopant is labeled N_A^- to indicate that the ions in the (pink) depletion layer are negatively charged and there are very few holes. In the (red) bulk the number of holes $p = N_A$ making the bulk charge neutral.

If the gate voltage is below the threshold voltage (top figure), the transistor is turned off and ideally there is no current from the drain to the source of the transistor. In fact, there is a current even for gate biases below threshold (subthreshold leakage) current, although it is small and varies exponentially with gate bias.

If the gate voltage is above the threshold voltage (lower figure), the transistor is turned on, due to there being many electrons in the channel at the oxide-silicon interface, creating a low-resistance channel where charge can flow from drain to source. For voltages significantly above threshold, this situation is called strong inversion. The channel is tapered when $V_D > 0$ because the voltage drop due to the current in the resistive channel reduces the oxide field supporting the channel as the drain is approached.

- In modern devices the threshold voltage is a much less clear-cut parameter subject to variation with the biases applied to the device; see drain induced barrier lowering.

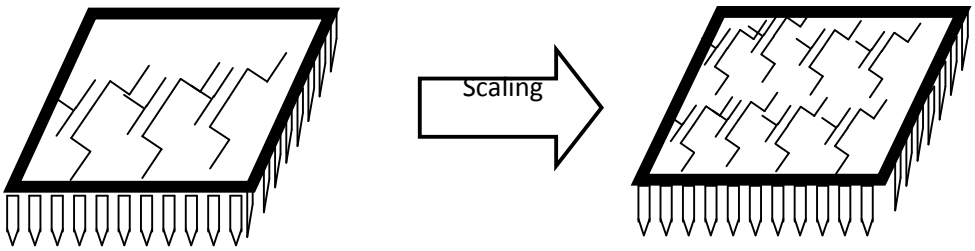

Fig. 1.14.MOSFET Scaling and Small Geometry Effects

To increase the number of devices per IC, the device dimensions had to be shrunk from one generation to another (i.e. scaled down)

- In theory, there are two methods of scaling:

    1. Full-Scaling (also called Constant-Field scaling): In this method the device dimensions (both horizontal and vertical) are scaled down by 1/S, where S is the scaling factor. In order to keep the electric field constant within the device, the voltages have to be scaled also by 1/S such that the ratio between voltage and distance (which represents the electric field) remain constant. The threshold voltage is also scaled down by the same factor as the voltage to preserve the functionality of the circuits and the noise margins relative to one another. As a result of this type of scaling the currents will be reduced and hence the total power per transistor (P=IxV) will also be reduced, however the power density will remain constant since the number of transistors per unit area will increase. This means that the total chip power will remain constant if the chip size remains the same (this usually the case).

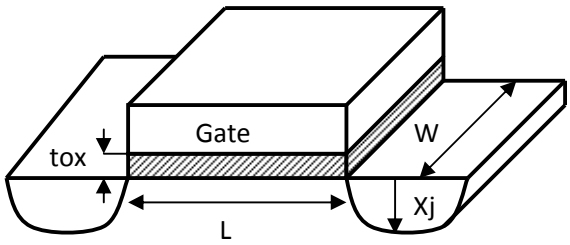

Fig. 1.15The table below summarizes how each device parameter scales with S (S>1)

Parameter	Before scaling	After scaling
Channel length	L	L/S
Channel width	W	W/S
Oxide thickness	tox	tox/S
S/D junction depth	Xj	Xj/S
Power Supply	VDD	VDD/S
Threshold voltage	V_{TO}	V_{TO} /S
Doping Density	N_A & N_D	N_A *S and N_D *S
Oxide Capacitance	Cox	S*Cox
Drain Current	I_{DS}	I_{DS} /S
Power/Transistor	P	P/S^1
Power Density/cm^1	*p*	*p*

2. Constant-Voltage scaling (CVS): In this method the device dimensions (both horizontal and vertical are scaled by S, however, the operating voltages remain

constant. This means that the electric fields within the device will increase (filed =Voltage/distance). The threshold voltages remain constant while the power per transistor will increase by S. The power density per unit area will increase by S^3! This means that for the same chip area, the power chip power will increase by S^3. This makes constant-voltage-scaling (CVS) very impractical. Also, the device doping has to be increased more aggressively (by S^1) than the constant-field scaling to prevent channel punch-through. Channel punch-through occurs when the Source and Drain Depletion regions touches one another. By increasing the doping by S^1, the depletion region thickness is reduced by S (the same ratio as the channel length). However, there is a limit for how much the doping can be increased (the solid solubility limit of the dopant in Silicon). Again, this makes the CVS impractical in most cases. The following table summarizes the changes in key device parameters under constant-voltage scaling:

Parameter	Before scaling	After scaling
Channel length	L	L/S
Channel width	W	W/S
Oxide thickness	tox	tox/S
S/D junction depth	Xj	Xj/S
Power Supply	VDD	VDD
Threshold voltage	V_{TO}	V_{TO}
Doping Density	N_A & N_D	$N_A * S^1$ and $N_D * S^1$

Oxide Capacitance	Cox	S*Cox
Drain Current	I_{DS}	I_{DS} * S
Power/Transistor	P	P*S
Power Desity/cm^1	*p*	*p* * S^3

In almost all cases, the scaling is a combination of constant-field scaling and constant-voltage scaling, such that the number of devices is increased and the total power/chip does not increase much.

Source : http://msk1986.files.wordpress.com/2013/09/7ec5-vlsi-design-unit-1-notes.pdf