Why VLSI?
- Integration improves the design:
  - lower parasitics = higher speed;
  - lower power;
  - physically smaller.
- Integration reduces manufacturing cost—almost no manual assembly.

Moore’s Law
- Predicted that number of transistors per chip would grow exponentially (double every 18 months).
- Exponential improvement in technology is a natural trend: steam engines, dynamos, automobiles.

The Cost of Fabrication
- Current cost $2 - 3 billion
- Most profitable period is first 18 months to 2 years
- For large volume IC’s packaging and testing is largest cost
- For low volume IC’s, design costs may swamp manufacturing costs

The VLSI design process
- May be part of larger product design.
- Major levels of abstraction:
  - specification
Challenges in VLSI design
- Multiple levels of abstraction: transistors to CPUs.
- Multiple and conflicting constraints: low cost and high performance are often at odds.
- Short design time: Late products are often irrelevant.

Dealing with complexity
- Divide-and-conquer: limit the number of components you deal with at any one time.
- Group several components into larger components:
  - transistors form gates;
  - gates form functional units;
  - functional units form processing elements;

VLSI Applications
- VLSI is an implementation technology for electronic circuitry - analogue or digital
- It is concerned with forming a pattern of interconnected switches and gates on the surface of a crystal of semiconductor
- Microprocessors
  - personal computers
  - microcontrollers
- Memory - DRAM / SRAM
- Special Purpose Processors - ASICS (CD players, DSP applications)
- Optical Switches
- Has made highly sophisticated control systems mass-producible and therefore cheap
Metal-oxide-semiconductor (MOS) and related VLSI technology

- pMOS
- nMOS
- CMOS
- BiCMOS
- GaAs

Technology Background
What is a Silicon Chip?

- A pattern of interconnected switches and gates on the surface of a crystal of semiconductor (typically Si)
- These switches and gates are made of:
  - areas of n-type silicon
  - areas of p-type silicon
  - areas of insulator
  - lines of conductor (interconnects) joining areas together
    - Aluminium, Copper, Titanium, Molybdenum, polysilicon, tungsten
- The geometry of these areas is known as the layout of the chip
- Connections from the chip to the outside world are made around the edge of the chip to facilitate connections to other devices

MOS TRANSISTOR - INTRODUCTION

- We have treated transistors as ideal switches
- An ON transistor passes a finite amount of current
  - Depends on terminal voltages
  - Derive current-voltage (I-V) relationships
- Transistor gate, source, drain all have capacitance
  - \( I = C \frac{\Delta V}{\Delta t} \Rightarrow \Delta t = \frac{(C/I)}{\Delta V} \)
  - Capacitance and current determine speed
MOS Transistor Theory

- Modulated by voltage applied to the gate voltage controlled device
- nMOS transistor: majority carriers are electrons (greater mobility), p-substrate doped (positively doped)
- pMOS transistor: majority carriers are holes (less mobility), n-substrate (negatively doped)

Terminal Voltages

- Mode of operation depends on $V_g$, $V_d$, $V_s$

\[
\begin{align*}
V_{gs} &= V_g - V_s \\
V_{gd} &= V_g - V_d \\
V_{ds} &= V_d - V_s = V_{gs} - V_{gd}
\end{align*}
\]

Source and drain are symmetric diffusion terminals
- By convention, source is terminal at lower voltage
- Hence $V_{ds} \geq 0$

nMOS body is grounded. First assume source is 0 too.
- Three regions of operation
  - Cutoff
  - Linear
  - Saturation

Gate Biasing
**nMOS Device Behavior**

- **Accumulation mode**: $V_{gs} < V_t$
- **Depletion mode**: $V_{gs} = V_t$
- **Inversion mode**: $V_{gs} > V_t$

- **Enhancement-mode transistor**: Conducts when gate bias $V_{gs} > V_t$
- **Depletion-mode transistor**: Conducts when gate bias is zero

- $V_{gs} = 0$: no current flows from source to drain (insulated by two reverse biased pn junctions)
- $V_{gs} > 0$: electric field created across substrate

- Electrons accumulate under gate: region changes from p-type to n-type
- Conduction path between source and drain
Basic MOS Transistors
- Minimum line width
- Transistor cross section
- Charge inversion channel
- Source connected to substrate
- Enhancement vs Depletion mode devices
- pMOS are 2.5 time slower than nMOS due to electron and hole mobilities

Types of Fabrication
- nMOS Fabrication
- CMOS Fabrication
  - p-well process
  - n-well process
  - twin-tub process

Fabrication Technology
- Silicon of extremely high purity
  - chemically purified then grown into large crystals
- Wafers
  - crystals are sliced into wafers
  - wafer diameter is currently 150mm, 200mm, 300mm
  - wafer thickness <1mm
  - surface is polished to optical smoothness
- Wafer is then ready for processing
- Each wafer will yield many chips
  - chip die size varies from about 5mmx5mm to 15mmx15mm
  - A whole wafer is processed at a time
- Different parts of each die will be made P-type or N-type (small amount of other atoms intentionally introduced - doping - implant)
Interconnections are made with metal.
Insulation used is typically SiO2. SiN is also used. New materials being investigated (low-k dielectrics).
All the devices on the wafer are made at the same time.
After the circuitry has been placed on the chip:
- the chip is overglassed (with a passivation layer) to protect it.
- only those areas which connect to the outside world will be left uncovered (the pads).
The wafer finally passes to a test station:
- test probes send test signal patterns to the chip and monitor the output of the chip.
The *yield* of a process is the percentage of die which pass this testing.
The wafer is then scribed and separated up into the individual chips. These are then packaged.
Chips are ‘binned’ according to their performance.

Fabrication Process – Lithography

**Basic sequence**
- The surface to be patterned is:
  - spin-coated with photoresist
  - the photoresist is dehydrated in an oven (photo resist: light-sensitive organic polymer)
- The photoresist is exposed to ultra violet light:
  - For a positive photoresist exposed areas become soluble and non exposed areas remain hard
- The soluble photoresist is chemically removed (development).
  - The patterned photoresist will now serve as an etching mask for the SiO2.
• The SiO2 is etched away leaving the substrate exposed:
  – the patterned resist is used as the etching mask
• Ion Implantation:
  – the substrate is subjected to highly energized donor or acceptor atoms
  – The atoms impinge on the surface and travel below it
  – The patterned silicon SiO2 serves as an implantation mask
• The doping is further driven into the bulk by a thermal cycle
• The lithographic sequence is repeated for each physical layer used to construct the IC. The sequence is always the same:
  – Photoresist application
  – Printing (exposure)
  – Development
  – Etching
1. **Photoresist coating**

   ![Photoresist coating diagram](image)

2. **Exposure**

   ![Exposure diagram](image)

3. **Development**

   ![Development diagram](image)

4. **Etching**

   ![Etching diagram](image)

5. **Ion implant**

   ![Ion implant diagram](image)

6. **After doping**

   ![After doping diagram](image)

Patterning a layer above the silicon surface
• Etching:
  – Process of removing unprotected material
  – Etching occurs in all directions
  – Horizontal etching causes an under cut
  – “preferential” etching can be used to minimize the undercut
• Etching techniques:
  – Wet etching: uses chemicals to remove the unprotected materials
  – Dry or plasma etching: uses ionized gases rendered chemically active by an rf-generated plasma
Physical structure – NMOS
NMOS physical structure:
- p-substrate
- n+ source/drain
- gate oxide (SiO₂)
- polysilicon gate
- CVD oxide
- metal 1
- \(L_{\text{eff}} < L_{\text{drawn}}\) (lateral doping)

NMOS layout representation:
- Implicit layers:
  - oxide layers
  - substrate (bulk)
- Drawn layers:
  - n+ regions
  - polysilicon gate
  - oxide contact cuts
  - metal layers

Physical structure - PMOS
PMOS physical structure:
- p-substrate
- n-well (bulk)
- p+ source/drain
- gate oxide (SiO₂)
- polysilicon gate
- CVD oxide
- metal 1

PMOS layout representation:
- Implicit layers:
  - oxide layers
- Drawn layers:
  - n-well (bulk)
  - n+ regions
  - polysilicon gate
  - oxide contact cuts
  - metal layers

**Structure of MOS transistors**

We will discuss the structure of two MOS Field-Effect-Transistors (FETs) that are building blocks for all digital devices. The nMOS transistor shown in Figure 2.1 (n-type, n-channel, enhancement mode field-effect transistor) is built on the p-type semiconductor substrate, which is usually acceptor-doped silicon.
• Two \textbf{n+ diffusion} regions (’+’ indicates the high degree of doping) form the source and drain of the transistor. The area in between forms a conducting channel. Potentially, electrons, negative carriers, will form the current in the channel.
• The gate, which is formed from a conductor, typically polysilicon, is insulated from the source-channel-drain structure (and from the substrate) by the layer of silicon dioxide.
• The voltage between the gate and the substrate induces the electric field which controls the flow of the carriers in the channel.
This gives the rise to the name: field-effect transistor (FET).
• Transistor structure is completely \textbf{symmetrical} with respect to the source and drain. The role is defined by terminal voltages which establish the direction of the current (carriers) flow.
The pMOS transistor (p-type, p-channel) is a complementary structure to the nMOS transistor as depicted in Figure 2.2.

- The pMOS transistor is built on the n-type substrate which is donor-doped silicon.
- The source and drain of a pMOS transistor are now p+ diffusion regions.
- The carriers in the channel are now positive holes. As previously, their flow is controlled by the gate-substrate voltage.

MOS transistors described above are referred to as enhancement mode transistors. There are also depletion mode transistors used mainly in the analog
circuitry.

**Operation of MOS transistors**

We will describe operation of an enhancement-mode n-channel MOS field effect transistor (nMOS) as illustrated in Figure 2.1. The pMOS operates in the dual way. The basic principle of operation can be stated as follows. The flow of the current between the source and the drain is controlled by the electric field generated by the gate-substrate voltage. In order for the drain-source current to exist there must be carriers existing in the area between the source and drain referred to as the conducting channel. We will examine first how the channel is created and then how the drain current depends on relevant voltage.

Note that a MOS transistor is a four terminal device. In most cases, the substrate and the source of an nMOS are connected to the ground potential (GND) as in Figure 2.3.

**Step 1:** (Figure 2.3) $V_{GS} < V_T \Rightarrow I_D = 0$
We start with the case when the gate voltage, $V_{GS}$, is smaller than the threshold voltage, $V_T$. The drain voltage, $V_{DS}$, is irrelevant and can be zero. The electric field induced by the gate voltage points down from the gate through the channel. This field repels the majority carriers for the p-type substrate, that is, positive holes, from the channel hence forming a region depleted of carriers as shown in Figure 2.3. As a result, due to the lack of free carriers, no current flows between the source and the drain at this stage, that is, $I_D = 0$.

**Step 2:** (Figure 2.4) $V_{GS} > V_T$ and $V_{DS} = 0 =) I_D = 0$

In this case (Figure 2.4), when the gate voltage $V_{GS}$ increases above
the threshold voltage $V_T$, then the electric field repels more holes from
the channel area leaving an excess of electrons.
The field also pulls out electrons from the source and drain
area which,
by virtue of being the n+ regions, have excess of electrons.
As a result in the area between source and drain an inversion
layer is
created in which there is an excess of the negative carriers,
that is, electrons. In other words a conducting channel has been
formed
between the source and drain.
Due to the fact that in this case we assume that the drain-
source
voltage $V_{DS} = 0$, thermal equilibrium exists in the channel
region and
the drain current, $I_D = 0$.
The threshold voltage $V_T$ depends on a specific transistor
configuration, that is, on a specific technology of fabrication
of MOS
transistors and usually is in a range of 0.5V.
**Step 3: Linear region** (Figure 2.5)
In this case, in the presence of free electrons in the conducting channel, when the drain-source voltage increases above zero, $V_{DS} > 0$, the drain-source current, $I_D$, starts to flow. When the $V_{DS}$ voltage is relatively small, the transistor operates in the so-called linear region. In this region of operation the drain current $I_D$ is a quadratic function of the source-drain voltage, $V_{DS}$. Descriptively it means that the increase of the drain current slows down when the source-drain voltage increases. The channel depth at the drain end decreases with the increase of the source-drain voltage as illustrated in Figure 2.5. Equivalently we can...
say that the channel region acts as a voltage controlled resistor: the resistance increases when the source-drain voltage increases. **Step 4: pinch-off point** (Figure 2.6) \( V_{GS} > V_T, \ V_{DS} = V_{sat} \Rightarrow I_D > I_{sat} \)

![Diagram](image)

Figure 2.6: The pinch-off point for an nMOS transistor
When the source-drain voltage, \( V_{DS} \), reaches a certain value, \( V_{sat} \), the channel depth at the drain end is reduced to zero. This is called the pinch-off point. In other words, at the pinch-off point, \( V_{DS} = V_{sat} \).
From now on, the further increase of the source-drain voltage does not result in an increase of the source-drain current. The transistor now operates in the saturation mode. **Step 5: Saturation mode** (Figure 2.7) \( V_G > V_T, \ V_D > V_{DSat} \Rightarrow I_D = I_{DSat} \)
In the saturation mode, the depletion region adjacent to the drain is enlarged. Note that in the depletion region there are no free electric carriers and the area acts as a dielectric. The source-drain current, $I_D$, is now independent of the source-drain voltage, $V_{DS}$. Electrons arriving from the source to the channel are injected into the depleted part of the channel and are accelerated towards the drain by the high electric field induced by the source-drain voltage. Finally, it is important to remember that under no conditions there is a constant current flowing between the gate and other transistor terminals because the gate is insulated by a layer of SiO2.

**The MOS Transistor**

Placing an insulating layer between the gate and the channel allows for a wider range of control (gate)
voltages and further decreases the gate current (and thus increases the device input resistance). The insulator is typically made of an oxide (such as silicon dioxide, SiO$_2$). This type of device is called a metal-oxide-semiconductor FET (MOSFET) or insulated-gate FET (IGFET). The substrate is often connected to the source internally. The insulated gate is on the opposite side of the channel from the substrate (see Fig ). The bias voltage on the gate terminal either attracts or repels the majority carriers of the substrate across the PN junction with the channel. This narrows (depletes) or widens (enhances) the channel, respectively, as $V_{GS}$ changes polarity. For N-channel MOSFETs, positive gate voltages with respect to the substrate and the source ($V_{GS} > 0$) repel holes from the channel into the substrate, thereby widening the channel and decreasing channel resistance. Conversely, $V_{GS} < 0$ causes holes to be attracted from the substrate, narrowing the channel and increasing the channel resistance. Once again, the polarities discussed in this example are reversed for P-channel devices. The common abbreviation for an N-channel MOSFET is NMOS, and for a P-channel MOSFET, PMOS.

Because of the insulating layer next to the gate, input resistance of a MOSFET is usually greater than $10^{12}$ Ohms (a million megohms). Since MOSFETs can both deplete the channel, like the JFET, and also enhance it, the construction of MOSFET devices differs based on the channel size in the resting state, $V_{GS} = 0$. A depletion mode, device (also called a normally on MOSFET) has a channel in resting state that gets smaller as a reverse bias s applied, this device conducts current with no bias applied (see Fig ). An enhancement mode device (also
called a *normally off* MOSFET) is built without a channel and does not conduct current when $V_{GS} = 0$; increasing forward bias forms a channel that conducts current (see Fig. ).

Figure 9.4 illustrates how an N-channel enhancement MOSFET works. The transistor differs from the J-FET in two ways. Firstly, the gate electrode is placed on top of a very thin insulating layer (of oxide — hence the ‘oxide’ in the name) which means it isn't in direct electrical contact with the rest of the transistor. Secondly, the device is made without manufacturing an N-type doped channel between the source and drain. Instead, there are a pair of small N-type regions just under the drain & source electrodes. As a result, when the gate source voltage $V_{gs} = 0$, then $I_{ds} = 0$ no matter what drain-source voltage we apply.

If we apply a positive voltage to the gate we'll set up an electrostatic field between it and the rest of the transistor. The positive gate voltage will push away the ‘holes’ inside the p-type substrate and attracts the
moveable electrons in the n-type regions under the source & drain electrodes. This produces a layer just under the gate's insulator through which electrons can get into and move along from source to drain. The positive gate voltage therefore ‘creates’ a channel in the top layer of material. Increasing the value of the positive gate voltage pushes the p-type holes further away and enlarges the thickness of the created channel. As a result we find that the size of the channel we've made increases with the size of the gate voltage and enhances or increases the amount of current which can go from source to drain — this is why this kind of transistor is called an enhancement mode device.

The above explanation is based on using n-type patches at source & drain in a p-type substrate. The resulting transistor behaves as if the gate voltage creates a channel of n-type material, hence it's called an ‘n-channel’ device. It's possible to build devices the ‘other way around’ where p-type patches are used in an n-type substrate. These behave in a similar way, but they pass current when a negative gate voltage creates an effective p-type channel layer under the insulator. By swapping around p-type for n-type we can make pairs of transistors whose behaviour is similar except that all the signs of the voltages and currents are reversed. Pairs of devices like this care called complimentary pairs. Figure 9.5 illustrates the behaviour of a typical complimentary pair of power MOSFETs made by Hitachi for use in hi-fi amplifiers.
MOS Transistors - Types and Symbols

Symbols for N-channel MOSFET

Symbols for P-channel MOSFET

2SK135 (Hitachi)

$V_{gs}$ Volts

$I_{ds}$ Amps

2SJ50 (Hitachi)

Volts $V_{gs}$

$I_{ds}$ Amps

Polysilicon

Aluminum

Gate Oxide

Field Oxide

P-Type

Source / Drain Regions

NMOS Enhancement

NMOS Depletion

PMOS Enhancement

NMOS with Bulk Contact
NMOS and PMOS

**NMOS Transistor**

\[ V_{GS} > 0 \]

**PMOS Transistor**

\[ V_{GS} < 0 \]

---

**Electrical Properties**

**The Drain Current**

Charge in the channel is controlled by the gate voltage:

\[ Q_i(x) = -C_{ox} \left[ V_{GS} - V(x) - V_T \right] \]

\[ C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \]

Drain current is proportional to charge and velocity:

\[ I_D = -\nu_n(x) Q_i(x) W \]

\[ \nu_n = -\mu_n \xi_n(x) = \mu_n \frac{dV}{dx} \]

Combining velocity and charge:

\[ I_D dx = \mu_n C_{ox} W (V_{GS} - V - V_T) dV \]

Integrating over the channel:

\[ I_D = k_n W \left[ \left( V_{GS} - V_T \right) V_{DS} - \frac{V_{DS}^2}{2} \right] = k_n \left[ \left( V_{GS} - V_T \right) V_{DS} - \frac{V_{DS}^2}{2} \right] \]

**Transconductance:**

\[ k_n = \mu_n C_{ox} = \frac{\mu_n \varepsilon_{ox}}{t_{ox}} \]
Transistor in Linear
Linear (Resistive) mode

MOS transistor and its bias conditions

Transistor in Saturation

Saturation
For $V_{GD} < V_T$, the drain current saturates

$$I_D = \frac{k'n}{2} \frac{W}{L} (V_{GS} - V_T)^2$$

Including channel-length modulation

$$I_D = \frac{k'n}{2} \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS})$$

Modes of Operation

Cutoff:

$$V_{GS} < V_T \quad \quad I_D = 0$$

Resistive:

$$V_T < V_{GS}; \quad V_{GS} - V_T > V_{DS} \quad \quad I_D = \frac{k'n}{2} \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2}\right]$$

Saturation:

$$V_T < V_{GS}; \quad V_{GS} - V_T < V_{DS} \quad \quad I_D = \frac{k'n}{2} \frac{W}{L} (V_{GS} - V_T)^2$$

Current-Voltage Relations

Threshold Voltage: Concept
The Threshold Voltage

Threshold \[ V_T = V_{T0} + \gamma(\sqrt{2\phi_F + V_{SB}}) - \sqrt{2\phi_F} \]

Fermi potential \[ \phi_F = \phi_T \ln\left(\frac{N_A}{n_i}\right) \]

- \(2\phi_F\) is approximately - 0.6V for p-type substrates
- \(\gamma\) - the body factor
- \(V_{T0}\) is approximately 0.45V for our process

The Body Effect
Carrier Mobility: Velocity Saturation
The mobility of the carriers reduces at higher electric fields normally encountered in small channel length devices due to velocity saturation effects.

Velocity as a function of electric field

As the channel length, L, is reduced while the supply voltage is not, the tangential
electric field will increase, and the carrier velocity may saturate. \( \Sigma c \approx 104 \text{ V/cm} \) for electrons. Hence for N-channel MOSFET with \( L < 1 \ \mu\text{m} \), velocity saturation causes the channel current to reach saturation before \( V_D = V_G - V_T \). Instead of \( I_{DSAT} \) being proportional to \( (V_G - V_T)^2 \) it is linearly proportional to \( (V_G - V_T) \) and is approximately given by

\[
I_{DSat} = WC_{ox} (V_G - V_T) v_{sat}
\]

\( ec \approx 5 \cdot 104 \text{ V/cm} \) for holes, hence velocity saturation for P-channel MOSFET will not become important until \( L < 0.25 \ \mu\text{m} \).

Current-Voltage Relations The Deep-Submicron Era
ID versus VGS

- \( \nu_{sat} = 10^5 \)
  - Constant velocity

- Constant mobility (slope = \( \mu \))

- \( \xi_c = 1.5 \)

- Long-channel device
  - \( V_{GS} = V_{DD} \)

- Short-channel device

\( I_D \) versus \( V_{DSAT} \), \( V_{GS} - V_T \), \( V_{DS} \)
ID versus VDS

Including Velocity Saturation
Approximate velocity:
\[ v = \frac{\mu_n \xi}{1 + \xi / \xi_c} \quad \text{for} \quad \xi \leq \xi_c \]
\[ = v_{sat} \quad \text{for} \quad \xi \geq \xi_c \]

And integrate current again:
\[ I_D = \frac{\mu_n C_{ox}}{1 + \left( \frac{V_{DS}}{\xi_c L} \right)} \left( \frac{W}{L} \right) \left[ (V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \]

In deep submicron, there are four regions of operation:
- (1) cutoff,
- (2) resistive, 
- (3) saturation and 
- (4) velocity saturation

**Subthreshold Conduction**
When the surface is in weak inversion (i.e., \( \xi < \xi_c < \xi_p \), \( V_G < V_T \)), a conducting channel starts to form and a low level of current flows between source and drain.

In MOS subthreshold slope \( S \) is limited to \( kT/q \) (60mV/dec)
- ID leakage ↑
- Static power ↑
- Circuit instability ↑
VDD is scaled for low power, delay, VT must scale to maintain ID (ON)
With subthreshold slope limited to 60mv/decade the dynamic range becomes limited.

**Hot Carrier Effects**
From our p-n junction discussion we remember that the maximum electric field intensity is near the junction itself and it increases with the reverse bias.

\[
\xi_{\text{max}} = \sqrt{\frac{2qN_a(\phi_i - V_D)}{\varepsilon_{\text{ox}}}}
\]

In the case of MOS transistor the maximum electric field is near the drain-substrate
junction. The drain reverse bias has to be dropped from drain to source. As the channel length is reduced the electric field intensity in the channel near the drain increases more rapidly in comparison to the long channel case as \( i \) does not scale.

The free carriers passing through the high-field can gain sufficient energy to cause several *hot-carrier* effects. This can cause many serious problems for the device operation. Hot carriers can have sufficient energy to overcome the oxide-Si barrier. They are injected from channel to the gate oxide (process 1) and cause gate current to flow. Trapping of some of this charge can change VT permanently. Avalanching can take place producing electron-hole pairs (process 2). The holes produced by avalanching
drift into the substrate and are collected by the substrate contact (process 3) causing I_{sub} IR drop due to I_{sub} (process 4) can cause substrate-source junction to be forward biased causing electrons to be injected from source into substrate (process 5). Some of the injected electrons are collected by the reversed biased drain and cause a parasitic bipolar action (process 5).

**Effect of Reducing Channel Width on VT**

There are no diffusions on the side of the channel. Hence the depletion region extends sideways in areas lying outside the gate controlled region increasing the apparent channel width. As a result the VT is increased. Note that the effect here is opposite to that of reducing channel length.
MOS Device Scaling

Side View of the MOS transistor

V_T vs W

MOS Device Scaling

Scaled MOS Transistor
Why do we scale MOS transistors?

1. Increase device packing density
2. Improve frequency response (transit time) $\propto \frac{1}{L}$
3. Improve current drive (transconductance $g_m$)

$$g_m = \frac{\partial I_D}{\partial V_G} \bigg|_{V_D = \text{const}}$$

$$= \frac{W}{L} \mu_n \frac{K_{ox}}{t_{ox}} V_D \quad \text{for } V_D < V_{D_{sat}}, \quad \text{linear region}$$

$$= \frac{W}{L} \mu_n \frac{K_{ox}}{t_{ox}} (V_G - V_T) \quad \text{for } V_D > V_{D_{sat}}, \quad \text{saturation region}$$

Decreasing the channel length and gate oxide thickness increases $g_m$, i.e., the current drive of the transistor. Much of the scaling is therefore driven by decrease in $L$ and $t_{ox}$.

However if only these two parameters are scaled many problems are encountered, e.g., increased electric field.

In reality constant field scaling has not been observed strictly. Since the transistor current is proportional to the gate overdrive $(V_G - V_T)$, high performance demands have dictated the use of higher supply voltage. However, higher supply voltage implies increased power dissipation $(CV^2f)$. In the recent past low power applications have become important and have required a scaling scenario with lower supply voltage.

**Limitations of Scaled MOSFET**

**Effect of Reducing Channel Length: Drain Induced Barrier Lowering (DIBL)**
In devices with long channel lengths, the gate is completely responsible for depleting the semiconductor (Q_B). In very short channel devices, part of the depletion accomplished by the drain and source bias.

Since less gate voltage is required to deplete Q_B, V_T ↓ as L ↓. Similarly, as V_D ↑, more Q_B is depleted by the drain bias, and hence V_T ↓. These effects are particularly pronounced in lightly doped substrates.

If the channel length becomes too short, the depletion region from the drain can reach the source side and reduces the barrier for electron injection. This is known as punch through.

In devices with long channel lengths, the gate is completely responsible for depleting the semiconductor (Q_B). In very short channel devices, part of the depletion is accomplished by the drain and source bias. Since less gate voltage is required to deplete Q_B, the barrier for electron injection from source to drain decreases. This is known as drain induced barrier lowering (DIBL).
Potential variation along the channel illustrating drain induced barrier lowering (DIBL).
DIBL results in an increase in drain current at a given $V_G$. Therefore $V_T \downarrow$ as $L \downarrow$. Similarly, as $V_D \uparrow$, more $Q_B$ is depleted by the drain bias, and hence $I_D \uparrow$ and $V_T \downarrow$.

**A model for manual analysis**

**Design Rules**
- Designs are represented by geometries on specific layers
- Each layer in the layout represents a process during manufacture
- Design rules set the limitations on the layout geometries
- They are a consequence of the methods used to manufacture the devices

*Design rules can be absolute measurements (e.g. in nm) or scaled to an abstract unit, the lambda.*

Lambda-based designs are scaled to the appropriate absolute units depending on the manufacturing process finally used.
One lambda (\(\lambda\)) = one half of the “minimum” mask dimension.

Typically the length of a transistor channel is 2\(\lambda\). Usually all edges must be “on grid”, e.g., in the MOSIS scalable rules, all edges must be on a lambda grid.

Layer Representation
- The layers may be represented as:
  - A suitable colour scheme
  - Varying shading or stipple patterns
  - Varying line styles
  - A combination of the above

Wire Layers
Transistors
<table>
<thead>
<tr>
<th>LAYER</th>
<th>COLOUR</th>
<th>SYMBOLIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-well</td>
<td>Brown</td>
<td></td>
</tr>
<tr>
<td>Thin-oxide</td>
<td>Green</td>
<td>N-transistor</td>
</tr>
<tr>
<td>Poly</td>
<td>Red</td>
<td>Polysilicon</td>
</tr>
<tr>
<td>P+</td>
<td>Yellow</td>
<td>P-transistor</td>
</tr>
<tr>
<td>Metal1</td>
<td>Light Blue</td>
<td>Metal1</td>
</tr>
<tr>
<td>Metal2</td>
<td>Tan</td>
<td>Metal2</td>
</tr>
<tr>
<td>Contact-cut, Via</td>
<td>Black</td>
<td>Contact</td>
</tr>
<tr>
<td>Metal3</td>
<td>Grey</td>
<td>Metal3</td>
</tr>
</tbody>
</table>

**Layout - Line of Diffusion**

- Very common layout method
  - Start with a "line of diffusion" for each type
  - Cross with poly to make transistors
  - This is the "type 2" NOR gate

**Line of Diffusion in General**

- Start with lines of diffusion for each transistor type
A "stick" diagram is a simplified layout form which does contain the information related to each of the process steps, but does not contain the actual size of the individual features. Instead the features are represented by simple lines hence the name "stick" diagram.
Transistors

Are formed when poly (red) crosses diffusion (green or yellow). (lots of fab steps to make it seem that simple)

Representation of MOS Circuitry

Figure 2.12: Three representations of MOS circuitry: schematics, stick diagrams and circuit layouts.