Unit 5

FPGA – Introduction:

- The full form of **FPGA** is "**Field Programmable Gate Array**". It contains ten thousand to more than a million logic gates with programmable interconnection.
- Programmable interconnections are available for users or designers to perform given functions easily.
- There are I/O blocks, which are designed and numbered according to function. For each module of logic level composition, there are **CLB's (Configurable Logic Blocks)**.
- CLB performs the logic operation given to the module. The inter connection between CLB and I/O blocks are made with the help of horizontal routing channels, vertical routing channels and PSM (Programmable Multiplexers).
- The number of CLB it contains only decides the complexity of FPGA.

What is an FPGA?

- Field Programmable Gate Arrays (FPGAs) are semiconductor devices that are based around a matrix of configurable logic blocks (CLBs) connected via programmable interconnects.
- FPGAs can be reprogrammed to desired application or functionality requirements after manufacturing.

Difference between an ASIC and an FPGA :

S NO	ASIC	FPGA
1	ASIC's can't be reconfigured with different design. They contain only one design in them for the whole life time.	FPGA's can be reconfigured with different designs
2	It is suited for bulk production	It is not suited for bulk production
3	It is energy efficient and consumes less power than FPGA	It consumes more power than ASIC
4	Higher frequencies can be achieved on the same node.	Limited in operating frequency on a particular node
5	It is built only after validating the design for permanent use of that design on silicon.	It is used to prototype and validate a design

FPGA Applications :

Consumer Electronics - Cost-effective solutions enabling next generation, full-featured consumer applications, such as converged handsets, digital flat panel displays, information appliances, home networking, and residential set top boxes

- **Medical** - For diagnostic, monitoring, and therapy applications, the Virtex FPGA and Spartan® FPGA families can be used to meet a range of processing, display, and I/O interface requirements
- **Security** - Xilinx offers solutions that meet the evolving needs of security applications, from access control to surveillance and safety systems.
- **Video & Image Processing** - Xilinx FPGAs and targeted design platforms enable higher degrees of flexibility, faster time-to-market, and lower overall non-recurring engineering costs (NRE) for a wide range of video and imaging applications.
- **Wired Communications** - End-to-end solutions for the Reprogrammable Networking Linecard Packet Processing, Framer/MAC, serial backplanes, and more
- **Wireless Communications** - RF, base band, connectivity, transport and networking solutions for wireless equipment, addressing standards such as WCDMA, HSDPA, WiMAX and others.

FPGA Architectures :

- An FPGA comprises of an array of programmable logic blocks that are connected to each other through programmable
- interconnect network Programmability in FPGAs is achieved through an underlying programming technology.

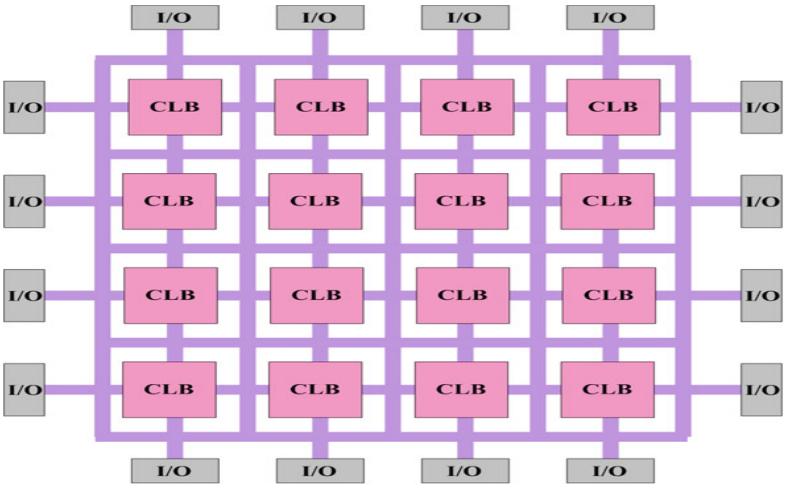


Fig: Overview of FPGA architecture

Programming Technologies :

- There are a number of programming technologies that have been used for reconfigurable architectures. Each of these technologies have different characteristics which in turn have significant effect on the programmable architecture.
- Some of the well known technologies include static memory , flash , and anti-fuse .

SRAM-Based Programming Technology :

- Static memory cells are the basic cells used for SRAM-based FPGAs use static memory (SRAM) based programming technology in their devices. These devices use static memory cells which are divided throughout the FPGA to provide configurability.
- In an SRAM-based FPGA, SRAM cells are mainly used for following purposes:

1. To program the routing interconnect of FPGAs which are generally steered by small multiplexors.
2. To program Configurable Logic Blocks (CLBs) that are used to implement logic functions.

Flash Programming Technology :

- One alternative to the SRAM-based programming technology is the use of flash or EEPROM based programming technology. Flash-based programming technology offers several advantages.

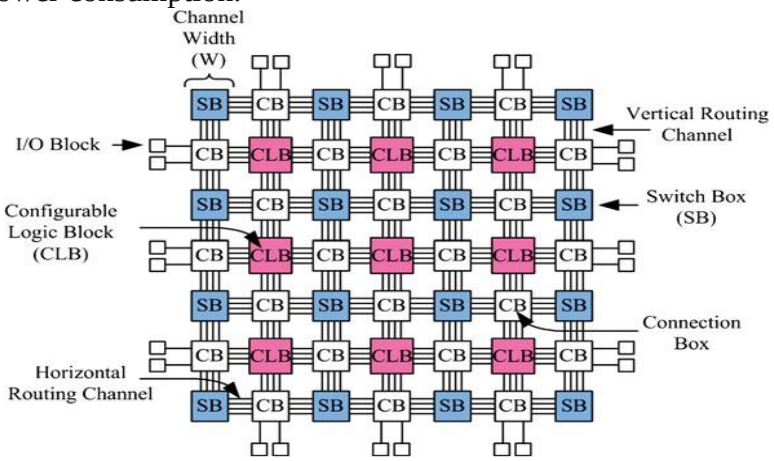
- Flash-based programming technology is also more area efficient than SRAM-based programming technology. Flash-based programming technology has its own disadvantages also.
- Unlike SRAM-based programming technology, flash based devices can not be reconfigured/reprogrammed an infinite number of times. Also, flash-based technology uses non-standard CMOS process.

Anti-fuse Programming Technology :

- An alternative to SRAM and flash-based technologies is anti-fuse programming technology.
- The primary advantage of anti-fuse programming technology is its low area.
- Also this technology has lower on resistance and parasitic capacitance than other two programming technologies.

Configurable Logic Block (CLB) :

- A configurable logic block (CLB) is a basic component of an FPGA that provides the basic logic and storage functionality for a target application design.
- In order to provide the basic logic and storage capability, the basic component can be either a transistor or an entire processor.
- However, these are the two extremes where at one end the basic component is very fine-grained (in case of transistors) and requires large amount of programmable interconnect which eventually results in an FPGA that suffers from area-inefficiency, low performance and high power consumption.



Overview of mesh-based FPGA architecture

FPGA Routing Architectures :

- The computing functionality is provided by its programmable logic blocks and these blocks connect to each other through programmable routing network. This programmable routing network provides routing connections among logic blocks and I/O blocks to implement any user-defined circuit.
- The routing interconnect of an FPGA consists of wires and programmable switches that form the required connection. These programmable switches are configured using the programmable technology.

FPGA Design Flow :

- One of the major benefits of using FPGAs is that they are reprogrammable, meaning they can be modified to serve a function that is completely different from the one they were performing before every time the designer uploads a new code to the FPGA.

FPGA Design Flow Diagram :

- FPGA flow starts with the design specifications.
- The functional description of the system is written in a Hardware Description Language (VHDL / Verilog) in the behavioural modelling style.
- The functionality is checked by performing behavioural simulation using a set of vectors the next step is performing a synthesis
- The synthesis step translates the behavioural netlist into a gate level netlist
- Synthesis step requires the behavioural netlist ,the selected device family(Spartan,virtex)
- Gate level netlist is again checked for functionality and next step is implementation step is performed.
- In the implementation step , the mapping of logic gates are done to the functional blocks in the FPGA, and the placement and routing are done to complete the implementation.
- Next, Bit stream file is generated which contains the programming data.
- Bit stream file is downloaded through the JTAG cable into the FPGA device.
- The final step is to test the FPGA device in the system and debug for any problems in functionality

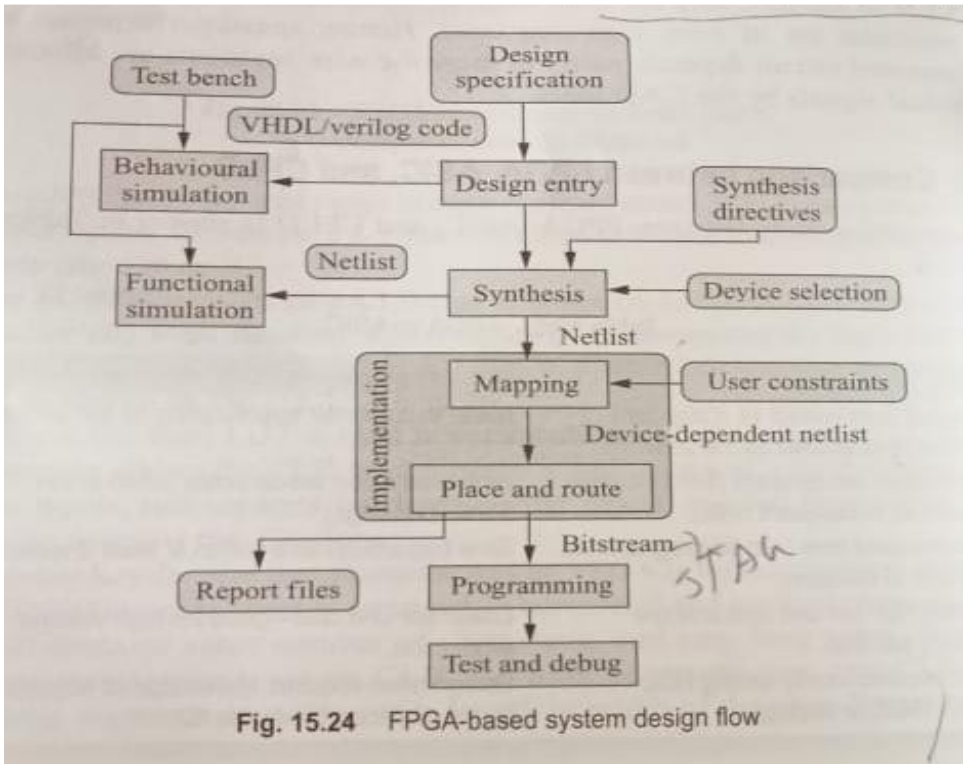


Fig. 15.24 FPGA-based system design flow

Xilinx FPGA Families

- Old families
  - XC3000, XC4000, XC5200
  - Old 0.5µm, 0.35µm and 0.25µm technology. Not recommended for modern designs.
- High-performance families
  - Virtex (220 nm)
  - Virtex-E, Virtex-EM (180 nm)
  - Virtex-II (130 nm)
  - Virtex-II PRO (130 nm)
  - Virtex-4 (90 nm)
  - Virtex-5 (65 nm)
  - Virtex-6 (40 nm) coming in 2009
- Low Cost Family
  - Spartan/XL – derived from XC4000
  - Spartan-II – derived from Virtex
  - Spartan-IIE – derived from Virtex-E
  - Spartan-3 (90 nm)
  - Spartan-3E (90 nm) – logic optimized
  - Spartan-3A (90 nm) – I/O optimized
  - Spartan-3AN (90 nm) – non-volatile,
  - Spartan-3A DSP (90 nm) – DSP optimized
  - Spartan-6 (45 nm) – coming in 2009

7 Series FPGA Families

Maximum Capability	ARTIX-7 Lowest Power and Cost	KINTEX-7 Industry's Best Price/Performance	VIRTEX-7 Industry's Highest System Performance
Logic Cells	20K – 355K	70K – 480K	285K – 2,000K
Block RAM	12 Mb	34 Mb	65 Mb
DSP Slices	40 – 700	240 – 1,920	700 – 3,960
Peak DSP Perf.	504 GMACS	2,450 GMACs	5,053 GMACS
Transceivers	4	32	88
Transceiver Performance	3.75Gbps	6.6Gbps and 12.5Gbps	12.5Gbps, 13.1Gbps and 28Gbps
Memory Performance	1066Mbps	1866Mbps	1866Mbps
I/O Pins	450	500	1,200
I/O Voltages	3.3V and below	3.3V and below 1.8V and below	3.3V and below 1.8V and below

Short-Channel Devices :

- A MOSFET device is considered to be short when the channel length is the same order ofmagnitude as the depletion-layer widths (x_{dD}, x_{dS}) of the source and drain junction.
- As the channel length L is reduced to increase both the operation speed and the number ofcomponents per chip, the so-called short-channel effects arise.

Short-Channel Effects :

The short-channel effects are attributed to two physical phenomena:

1. the limitation imposed on electron drift characteristics in the channel,
2. the modification of the threshold voltage due to the shortening channel

length.In particular five different short-channel effects can be distinguished:

1. Drain-induced barrier lowering and punch through
2. Surface scattering
3. Velocity saturation
4. Impact ionization
5. Hot electrons

1. Drain-induced barrier lowering and punch through :

The expressions for the drain and source junction widths are:

$$x_{dD} = \sqrt{\left(\frac{2e_{Si}}{qN_A}\right)\left(V_{DS} + f_{Si} + V_{SB}\right)}$$

and

$$x_{dS} = \sqrt{\left(\frac{2e_{Si}}{qN_A}\right)\left(f_{Si} + V_{DB}\right)}$$

where V_{SB} and V_{DB} are source-to-body and drain-to-body voltages.

- When the depletion regions surrounding the drain extends to the source, so that the two depletion layer merge (i.e., when $x_{dS} + x_{dD} = L$), punch trough occurs.
- Punch through can be minimized with thinner oxides, larger substrate doping, shallower junctions, and obviously with longer channels.
- The current flow in the channel depends on creating and sustaining an inversion layer on thesurface.
- If the gate bias voltage is not sufficient to invert the surface ($V_{GS}<V_{T0}$), the carriers (electrons) in the channel face a potential barrier that blocks the flow.
- Increasing the gate voltage reduces this potential barrier and, eventually, allows the flow of carriers under the influence of the channel electric field.
- In small-geometry MOSFETs, the potential barrier is controlled by both the gate-to-source voltage V_{GS} and the drain-to-source voltage V_{DS}.

- If the drain voltage is increased, thepotential barrier in the channel decreases, leading to drain-induced barrier lowering (DIBL).
- The reduction of the potential barrier eventually allows electron flow between the source and the drain, even if the gate-to-source voltage is lower than the threshold voltage.
- The channel current that flows under this conditions ($V_{GS}<V_{T0}$) is called the sub-threshold current.

- The potential barrier is controlled by both the gate-to-source voltage V_{GS} and the drain-to-source voltage V_{DS}.

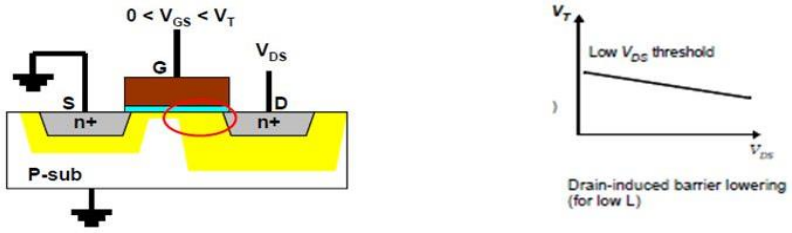


Fig : Drain-induced barrier lowering and punch through

2. Surface scattering :

- As the channel length becomes smaller due to the lateral extension of the depletion layer into the channel region, the longitudinal electric field component e_y increases, and the surface mobility becomes field-dependent.

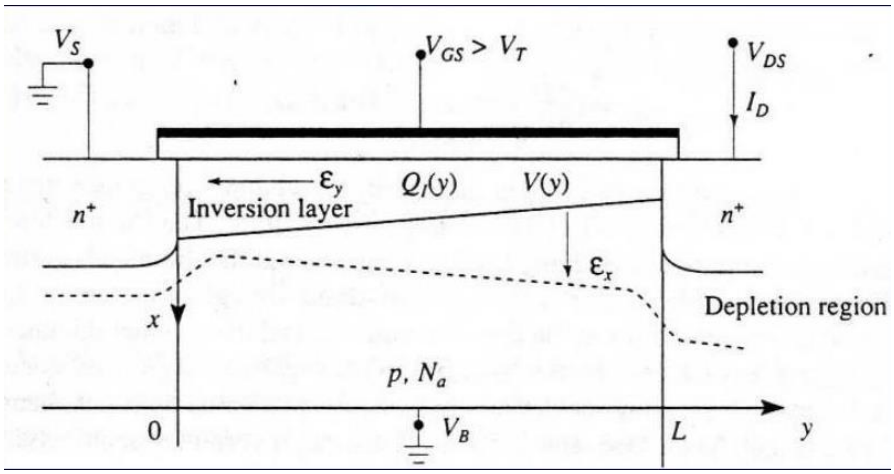


Fig: Surface scattering

- Since the carrier transport in a MOSFET is confined within the narrow inversion layer, and the surface scattering (that is the collisions suffered by the electrons that are accelerated toward the interface by e_x) causes reduction of the mobility, the electrons move with great difficulty parallel to the interface, so that the average surface mobility, even for small valuesof e_y, is about half as much as that of the bulk mobility.

3. Velocity saturation :

- The performance short-channel devices is also affected by velocity saturation, which reduces the transconductance in the saturation mode.
- At low e_y, the electron drift velocity v_{de} in the channel varies linearly with the electric field intensity.

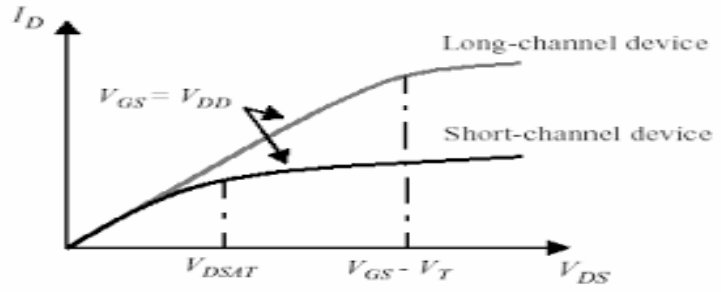


Fig : Velocity saturation

4. Impact ionization :

- Another undesirable short-channel effect, especially in NMOS, occurs due to the high velocity of electrons in presence of high longitudinal fields that can generate electron-hole (e-h) pairs byimpact ionization, that is, by impacting on silicon atoms and ionizing them.

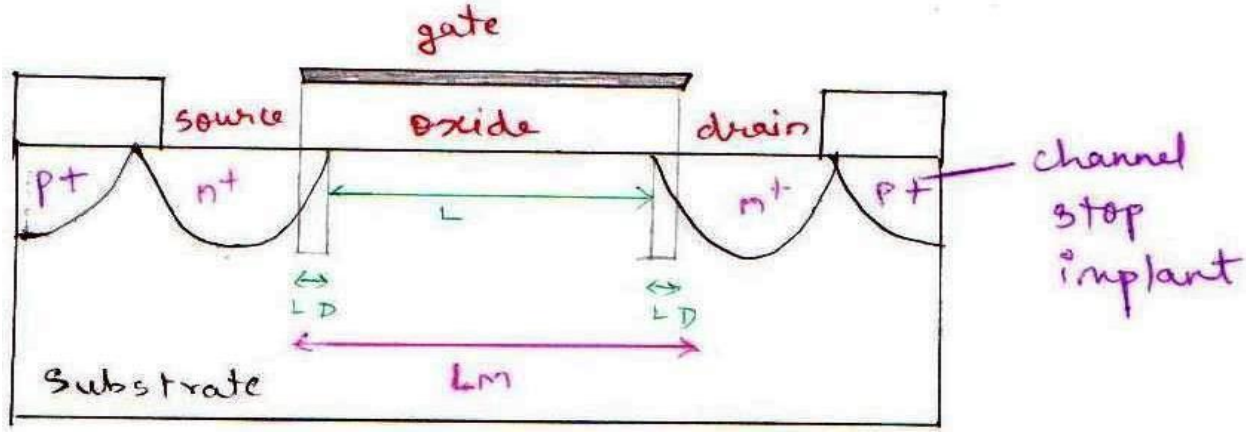


Fig: Impact ionization

5. Hot electrons :

- Another problem, related to high electric fields, is caused by so-called hot electrons.
- This high- energy electrons can enter the oxide, where they can be trapped, giving rise to oxide charging that can accumulate with time and degrade the device performance by increasing V_T and affect adverselythe gate's control on the drain current.

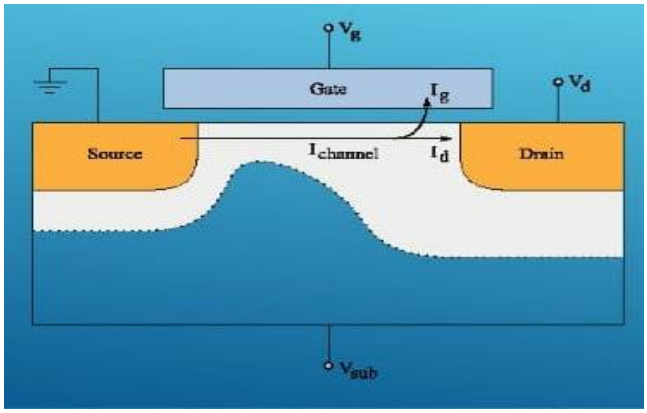


Fig : Hot electrons

- The hot electron (or short channel) effect is described in as occurring when a high voltage is applied across the source and drain of a device, the electric field is high, and the electrons are accelerated in the channel.
- The fastest electrons may damage the oxide and the interface near the drain, thus inducing transistor threshold shift and mobility change over the life of the part.

High-κ dielectric :

- The term **high-κ dielectric** refers to a material with a high dielectric constant (κ, kappa), as compared to silicon dioxide.
- High-κ dielectrics are used in semiconductor manufacturing processes where they are usually used to replace a silicon dioxide gate dielectric or another dielectric layer of a device.
- The implementation of high-κ gate dielectrics is one of several strategies developed to allow further miniaturization of microelectronic components, colloquially referred to as extending Moore's Law.
- Sometimes these materials are called "high-k" (pronounced "high kay"), instead of "high-κ" (high kappa).

Need for high-κ materials :

- Silicon dioxide (SiO_2) has been used as a gate oxide material for decades. As metal-oxide-semiconductor field-effect transistors (MOSFETs) have decreased in size, the thickness of the silicon dioxide gate dielectric has steadily decreased to increase the gate capacitance (per unit area) and thereby drive current (per device width), raising device performance.
- As the thickness scales below 2 nm, leakage currents due to tunneling increase drastically, leading to high power consumption and reduced device reliability. Replacing the silicon dioxide gate dielectric with a high-κ material allows increased gate capacitance without the associated leakage effects

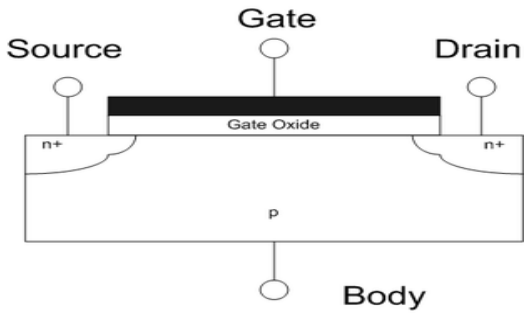
- The gate oxide in a MOSFET can be modeled as a parallel plate capacitor. Ignoring quantum mechanical and depletion effects from the Si substrate and gate, the capacitance C of this parallel plate capacitor is given by

$$C = \kappa \varepsilon_0 A / t$$

where

- A is the capacitor area
- κ is the relative dielectric constant of the material (3.9 for silicon dioxide)
- ε_0 is the permittivity of free space
- t is the thickness of the capacitor oxide insulator

- Since leakage limitation constrains further reduction of t, an alternative method to increase gate capacitance is alter κ by replacing silicon dioxide with a high-κ material.
- In such a scenario, a thicker gate oxide layer might be used which can reduce the leakage current flowing through the structure as well as improving the gate dielectric reliability.



The drain current I_D for a MOSFET can be written (using the gradual channel approximation) as

$$I_D(\text{sat}) = \{W/L\} \mu C_{inv} \{ V_G - V_{th} \}^2/2$$

where

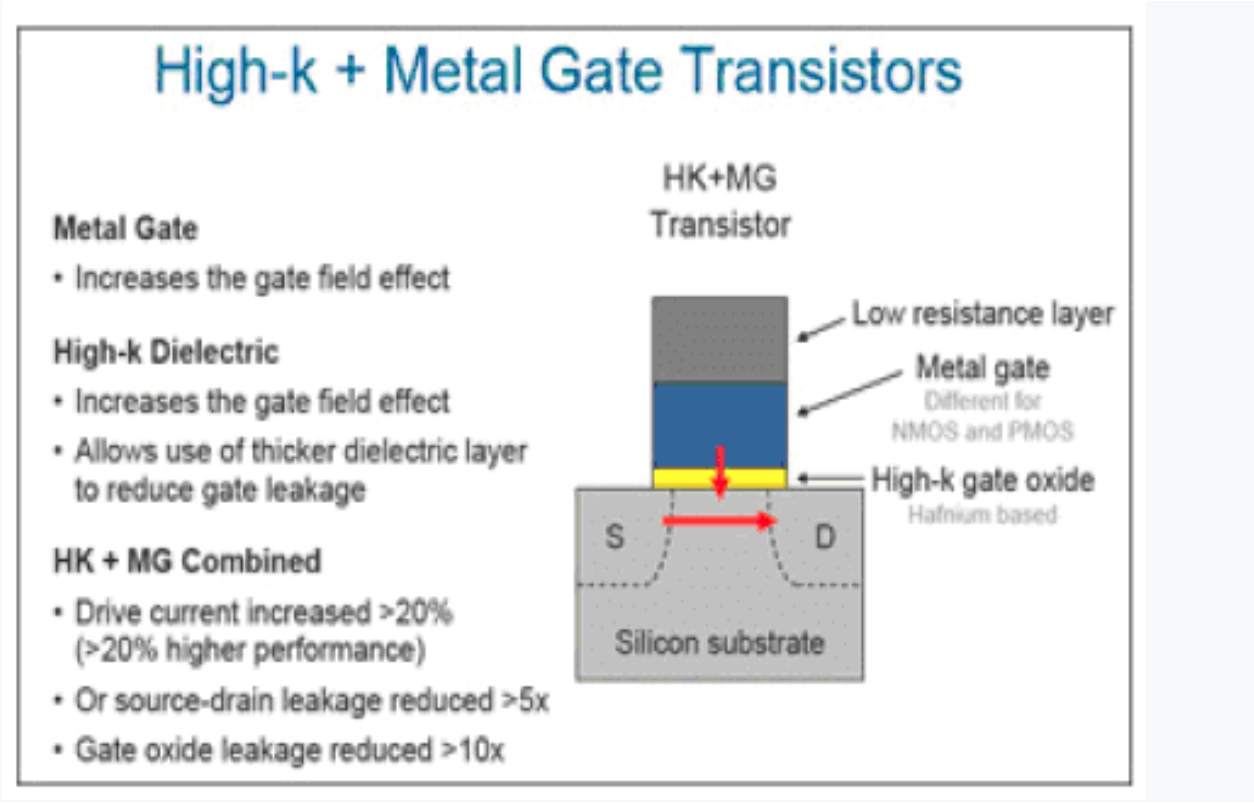
- W is the width of the transistor channel
- L is the channel length
- μ is the channel carrier mobility (assumed constant here)
- C_{inv} is the capacitance density associated with the gate dielectric when the underlying channel is in the inverted state
- V_G is the voltage applied to the transistor gate
- V_{th} is the threshold voltage

- The term $V_G - V_{th}$ is limited in range due to reliability and room temperature operation constraints, since a too large V_G would create an undesirable, high electric field across the oxide. Furthermore, V_{th} cannot easily be reduced below about 200 mV, because leakage currents due to increased oxide leakage and subthreshold conduction raise stand-by power consumption to unacceptable levels.

- Thus, according to this simplified list of factors, an increased $I_{D,sat}$ requires a reduction in the channel length or an increase in the gate dielectric capacitance.

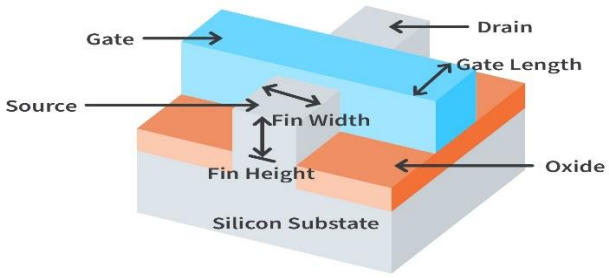
Metal gate :

- A metal gate, in the context of a lateral metal–oxide–semiconductor (MOS) stack, is **the gate electrode separated by an oxide from the transistor's channel** – the gate material is made from a metal.The "M" for metal has been replaced by a non-metal gate material.



FinFET (fin field-effect transistor) :

- A FinFET (fin field-effect transistor) is a MOSFET (metal-oxide semiconductor field-effect transistor) built on a silicon substrate where the gate is placed on multiple sides of the channel (fin) or wrapped around the channel, forming a multigate structure.
- It is called FinFET because the source/drain region on the silicon surface resembles the back fin of a fish. FinFET replaces the conventional MOSFET to allow transistors to be scaled down to the nanoscale region. It is a type of nonplanar or “3D” transistor used in modern processor designs due to its significant advantages.



- FinFET is built on a SOI (Silicon-on-insulator) or bulk silicon wafer that can be fabricated using an existing CMOS compatible technology. Like the conventional MOSFET transistor, FinFET also has a source, drain, and gate terminal to control the flow of current. What makes the FinFET differ from a MOSFET is the channel between the source and drain of FinFET. The channel on top of the silicon substrate is a three-dimensional bar, which is called a “fin”.
- The three-dimensional 'fins' form the source and drain, enabling more volume than a planar transistor for the same area. The gate orientation is at a right angle to the vertical fin, traversing one side of the fin to the other; it wraps over the fin, enabling it to interface with three sides of the fin. This forms several gate electrodes on each side of the fin or channel which reduce current leakage effects and improve drive current.
- Based on the design structure, the fin height of a single-fin FinFET must be half of the effective channel width (electrical width), W_{eff}. The drive current of the FinFET can be increased by increasing the height of the fin, thereby increasing the width of the channel.

$$W_{eff} = NFIN\,(\,TFIN + 2\,HFIN\,)$$

where,

- *NFIN* is the number of fins aligned in parallel
- *TFIN* is the thickness (width), and
- *HFIN* is the height of each fin.

Fin FET Advantages :

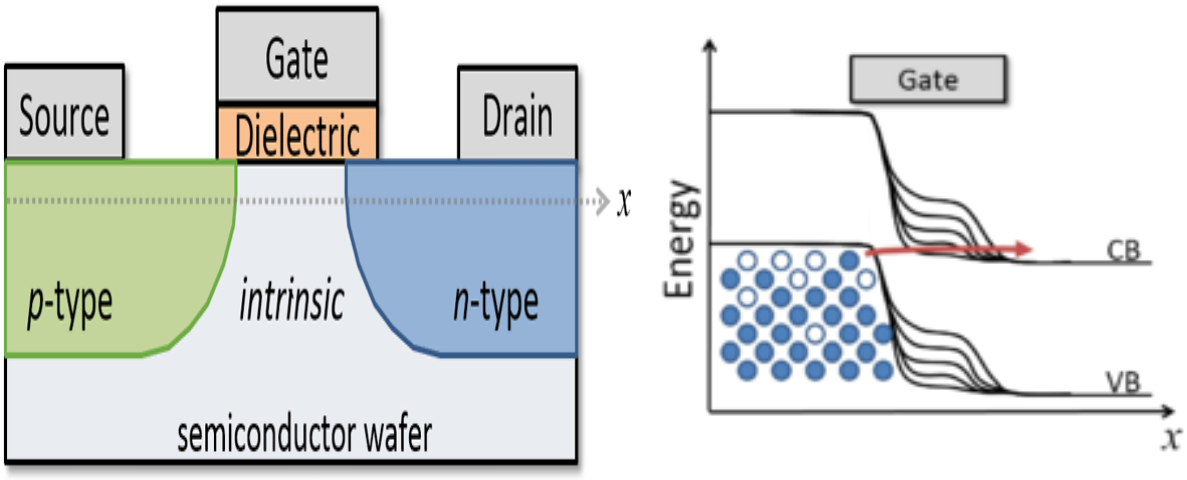
- Low power switching voltage
- Low power consumption
- Faster switching speed
- High drain current

Fin FET Disadvantages :

- Difficult to control dynamic Vth
- Very high capacitances
- High Fabrication cost

Tunnel field-effect transistor :

- The ***tunnel field-effect transistor*** **(*tunnel FET*** or **TFET)** is a transistor that operate by tunneling through the source/drain barrier rather than diffusion over the barrier.
- Tunnel FETs belongs to the family of so-called steep-slope devices which can switch on/off at lower voltages than metal oxide semiconductor FETs (MOSFETs) and are being investigated for ultra-low-power electronic applications
- A common TFET device structure consists of a P-I-N (p-type, intrinsic, n-type) junction, in which the electrostatic potential of the intrinsic region is controlled by a gate terminal.



- The device is operated by applying gate bias so that electron accumulation occurs in the intrinsic region for an n-type TFET.
- At sufficient gate bias, band-to-band tunneling (BTBT) occurs when the conduction band of the intrinsic region aligns with the valence band of the P region.
- Electrons from the valence band of the p-type region tunnel into the conduction band of the intrinsic region and current can flow across the device.
- As the gate bias is reduced, the bands becomes misaligned and current can no longer flow.