http://www.realworldtech.com/page.cfm?AID=RWT050802020022&p=1 Go
1 capture
11 May 2003
APR MAY JUN
2002 11 2003 2004
About this capture

The Stuff Dreams Are Made Of [Part 1]

By: Paul DeMone **Updated:** 05-08-2002

This is the first installment of a two part article that examines CMOS, the nearly perfect technology for implementing the digital logic that powers much of our modern age. The n-type and p-type field effect transistor are introduced along with how these are combined to form basic logic gates. In the second part I'll examine performance issues and the methods microprocessor circuit designers use to reach today's GHz clock rates.

The Straight Dope on CMOS

A single technology underlies the modern microprocessor, and indeed, just about every integrated circuit you will find inside computers, telecommunications equipment, and digital consumer electronic devices. That technology is called CMOS, or Complementary Metal Oxide Semiconductor. CMOS is a nearly perfect silicon based technology with which to implement digital logic. It is fast, flexible, dissipates low power, has excellent noise immunity, and scales nicely with shrinking feature size and reduced supply voltage. Constant improvement to photolithographic techniques allows the number of transistors that can be incorporated into a mass-produced silicon chip to grow exponentially in time, a trend commonly known as Moore's Law. What is less well known is that without CMOS technology the degree to which this trend can be harnessed for practical applications would be severely constrained and many areas of microelectronics would be at least a decade behind where they are today.

Silicon is called a semiconductor because its electrical properties lie in between a good insulator like glass, and a good conductor, like copper. Pure silicon is a poor conductor at room temperature because it has very few free charge carriers. Nearly all electrons in silicon reside in a low energy state called the *valence band* where they are tied to individual atoms within the crystal lattice, and are thus unable to move and carry electrical current. It takes a relatively large amount of energy to kick an electron from the valence band to the *conduction band* where it is free to move and thus create electrical current. An electron that moves into the conduction band in pure silicon leaves behind a so-called hole, a missing covalent bond between a silicon atom in the crystal lattice and its neighbors. Remarkably enough, a hole acts as a kind of positive charge carrier. In response to an electric field, an electron from the neighboring atom in the direction of the electric field will shift into the position of the missing covalent bond leaving a gap in its place. Thus a hole effectively moves in the direction of an applied electric field.

In pure silicon every electron in the conduction band has a corresponding hole in the valence band. These are called an electron-hole pair. There is a very strong tendency for a free electron to fall back into the valence band and eliminate the hole, a process called electron hole recombination. Because of this there are relatively few electrons or holes to act as charge carriers in pure silicon, and thus it is a good insulator. By introducing an atom of an element with 3 valence electrons (one less than silicon) into pure silicon a hole will be created in the valence band. Because there is no corresponding electron in the conduction band looking to recombine with this hole it will persist over time. An example of a 3 valence electron (or *acceptor*) atom is boron. Because a hole acts like a virtual positive charge carrier, silicon that has been deliberately contaminated (or *doped*) with acceptor atoms is called p-type or p-doped silicon. The greater the doping level (i.e. the concentration of acceptor atoms in the silicon) the more holes there are per unit volume and the greater the conductivity of the doped silicon. In a similar fashion silicon can be doped using an element with 5 valence electrons, one more than silicon. The extra valence electron is free to move in the conduction band because it has no corresponding hole for it to recombine with. A valence 5 element, or *donor*, is called an n-type dopant because an electron is a negative charge carrier. An example of a donor element is phosphorous.

When a region of n-type silicon is formed next to a region of p-type silicon the interface is known as a pn junction, which is the basis of the diode, an electronic component that only conducts current in one direction. Because current is carried by holes in p-type silicon and electrons in n-type silicon, current normally can only flow in the direction p to n (so called forward bias). This current is composed of a stream of holes in the p-type silicon that flows to the pn junction and recombine with a corresponding stream of electrons in the n-type silicon that also flows to the junction. Current can't flow in the opposite direction (reverse bias) because that would require electron hole pair generation at the junction rather than electron hole pair recombination. Except in special cases (like a photodiode exposed to light) there is no energy source available to create electron hole pairs at the junction and thus current can't flow.

Next Page >>

Copyright © 2002 - All Rights Reserved

http://www.realworldtech.com/page.cfm?AID=RWT050802020022&p=2 Go
5 captures
6 Jul 2002 - 28 Apr 20
MAR APR MAY
2002 28 2003 2004
About this capture

The Stuff Dreams Are Made Of [Part 1]

By: Paul DeMone Updated: 05-08-2002

Transistor Yin and Yang

There are two basic types of field effect transistor (FET) that can be constructed, the n-channel FET (NFET) and the p-channel FET (PFET). The physical structure of the NFET and PFET are shown in Figure 1.

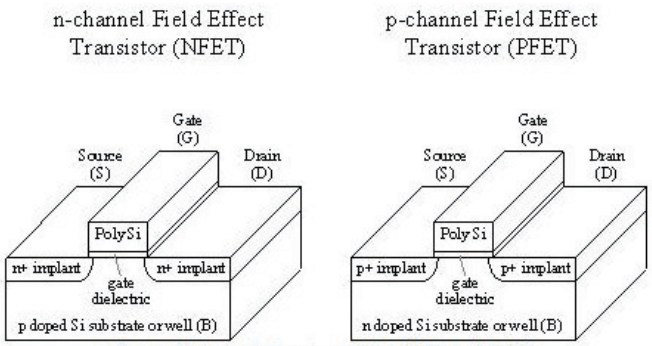


Figure 1 Physical Structure of the NFET and PFET

The NFET starts with an underlying body or substrate of p-type silicon. Two side by side regions in the top of the substrate are changed to n-type silicon by introducing (or implanting) a higher concentration of n-type dopant atoms than the concentration of p-type dopant atoms in the substrate. This is also known as an n+ implant, the "+" indicates a relatively high doping density. The two n+ regions are known as the source and drain. The designation is essentially arbitrary because a FET is electrically symmetrical and will function identically if source and drain are swapped. Because the source and drain regions are separated by p-doped silicon (which is generally kept at the lowest voltage potential on the chip) no current can flow between them. A field effect transistor is formed by depositing a polycrystalline silicon (also known as polysilicon or poly) structure called the gate over the substrate and between the source and drain regions. The gate is electrically isolated from the source, drain, and substrate region by an extremely thin layer of insulator called the gate dielectric.

The gate dielectric is typically silicon dioxide (SiO_2) and in the early days of the semiconductor industry the gate was composed of a metal like aluminum rather than polysilicon. This metal, oxide, semiconductor sandwich used historically to form FETs gave this technology the name MOS and why FETs are sometimes called MOSFETs. In modern MOS processes the polysilicon gate is actually formed before the source and drain regions are implanted and helps to precisely define the inner, opposing edges of the n+ implant. This technique is called *self aligned gate* and allows much faster transistor operation due to virtual elimination of the overlap between the gate and source and drain regions that creates parasitic capacitance. (The minute amount of overlap shown in Figure 1 is due to the fact that the dopant atoms in the source and drain regions tend to diffuse laterally a little bit during manufacturing).

The NFET operates when the voltage of the gate is raised relative to that of the source or drain region. The electric potential on the gate generates an electric field that tends to pull electrons from the source or drain region into the substrate immediately underneath the gate to balance the positive charge on the gate. If the gate voltage reaches a high enough level, the *threshold voltage*, the concentration of electrons in the substrate under the gate close to the dielectric (the channel region) will exceed the hole density in the p-type substrate. That charge carrier ratio *inversion* creates an effectively n-type region in the channel that allows current to flow from source to drain or vice versa. The higher the amount by which the voltage on the gate exceeds the threshold voltage, the greater the excess concentration of electrons in the channel and more current can flow.

The PFET is a mirror image (or complement) of the NFET. It starts with an n-type substrate (or well i.e. a substrate within a substrate) into which p-type source and drain regions are implanted on either side of a gate structure. The 'n' substrate is generally tied to the highest voltage potential on the chip. The PFET operates when the voltage of the gate is lowered relative to that of the source or drain region. When the gate voltage is lowered below a specific level, again called the threshold voltage, the concentration of holes induced in the channel region under the gate exceeds the background electron density in the n-substrate and the channel region is inverted to effectively p-type doping, which allows current to flow between source and drain and vice versa. The more the gate voltage is lowered below the threshold voltage the higher the excess hole concentration in the channel and more current can flow.

Copyright © 2002 - All Rights Reserved

http://www.realworldtech.com/page.cfm?AID=RWT050802020022&p=3 Go
3 captures
16 May 2002 - 11 May
SEP MAY JUN
2002 11 2003 2004
About this capture

The Stuff Dreams Are Made Of [Part 1]

By: Paul DeMone **Updated:** 05-08-2002

Paying the Ultimate Complement

The rise of the monolithic microprocessor and high density semiconductor memory began in the 1970's with the 4 bit microprocessor and the 1 Kb DRAM, both introduced by Intel and both based on PMOS technology. PMOS devices incorporate digital logic composed entirely of p-channel field effect transistors (PFETs). This was soon superseded by faster NMOS technology, which uses n-channel field effect transistors (NFETs) to form logic circuits. CMOS has both NFETs and PFETs (its name comes from the fact that it has the two complementary types of FETs) and went mainstream in the early 1980's with early 32-bit microprocessors like the MC68020 and i80386, as well as 256Kb DRAMs. Although CMOS integrated circuits require more processing steps to manufacture than PMOS or NMOS based devices and have slightly lower logic density, its desirable qualities far outweigh those considerations. CMOS has dominated the semiconductor industry for the past two decades, and this domination will continue for years to come as there is no credible successor in sight that can match all its strengths as well as exceed it in either performance or cost effectiveness.

The reason why having both NFETs and PFETs available is such a tremendous advantage arises from the process of constructing digital logic, the mathematical building block of computers and modern telecommunications. As most people know, digital logic relies on two discrete signal states commonly represented as the binary digits 0 and 1. These are typically represented in circuitry by low and high voltage levels respectively. When only one type of transistor is available, like NFETs in an NMOS process, logic circuitry is often constructed using a weak transistor that is always on and one or more strong transistors that are switched on and off. In NMOS the weak transistor is used to generate a high output voltage level or 1 when the strong transistors are turned off. When a strong transistor turns on it overwhelms the weak transistor and forces the output voltage low, which represents a 0. The process by which NMOS logic circuitry generates a 0 output consumes a steady stream of DC current, which is turned directly into heat. In addition, because of the disparity in strength between the weak and strong transistors, a NMOS logic circuit has a pronounced asymmetry in switching speed. It may take four times longer, or more, for a circuit output to switch from a 0 to 1 than from a 1 to a 0, yet one more headache for design engineers.

With a CMOS process, logic circuits can be constructed with both NFETs and PFETs in such a way that transistors never fight against each other except momentarily when switching states, and thus consume virtually no DC current when left quiescent in either output state. A side effect of this property is the ability to control power consumption over a tremendous range by adjusting the amount of switching activity per unit time as the application permits. This can be done by reducing an MPU's clock frequency, for example, or automatically disabling the clock to unused function units like a FPU. This attribute is a tremendous advantage for CMOS in being able to exploit Moore's Law. One of the last high-performance microprocessors constructed using NMOS logic was an early PA-RISC processor designed and manufactured by HP. This 1.5 um MPU used 71k transistors and consumed 10 Watts at 15 MHz. The next generation PA-RISC chip was implemented in 1.0 um CMOS. It contained 196k transistors, ran at 32 MHz and consumed 7.5 Watts. The amazing progress made in the microprocessor field over the last twenty years would have been a lot less impressive if CMOS didn't exist.

In addition to much lower effective power consumption, CMOS logic circuits can be balanced to have nearly symmetrical switching speeds. That is, the output can change from a 0 to a 1 as fast as from a 1 to a 0. In very rare and specific cases it may be desirable to build logic using an NMOS style methodology with constant current draw. The beauty of CMOS is that it is flexible enough to accommodate a wide variety of circuit design styles, including what is called pseudo-NMOS that can accommodate such needs. Another advantage of CMOS is that its logic circuits are very resistant to electrical design errors, like an error in transistor sizing. Using a wrong sized transistor in an NMOS circuit can render it completely non-functional, while a conventional CMOS circuit will still operate robustly, although perhaps at somewhat reduced speed. This makes it much simpler to automate the design of chips, or portions thereof, where the ultimate in performance is not required.

<< Prev Page Next Page >>

Pages: Paying the Ultimate Complement Go

Copyright © 2002 - All Rights Reserved

http://www.realworldtech.com/page.cfm?AID=RWT050802020022&p=4 Go
2 captures
8 Sep 2002 - 17 M
SEP MAY JUN
17
2002 2003 2004
About this capture

NEWS
Tech News
World Headlines
Special Reports
COLUMNS
Benchmark Examiner
Industry Update
Silicon Insider
RWT LABS
System Health
Technologies
Evaluations
MEMBERS
Login
Join RWT
FORUMS
SEARCH
LINKS
ABOUT US

The Stuff Dreams Are Made Of [Part 1]

By: Paul DeMone Updated: 05-08-2002

Print EMail Author

Pages: Gates Within Gates Go

Gates within Gates

Because CMOS logic circuits do not depend on detailed transistor electrical performance it is possible to understand their function using a simplified switch model for NFET and PFET transistors. The schematic symbol and simplified switch model for the NFET and PFET are shown in Figure 2 (the connection to the transistor substrate or body is sometimes denoted as a fourth terminal, B, on the FET symbol but for purely digital design it is often ignored through use of a 3 terminal symbol because it is invariably tied off to the appropriate power supply).

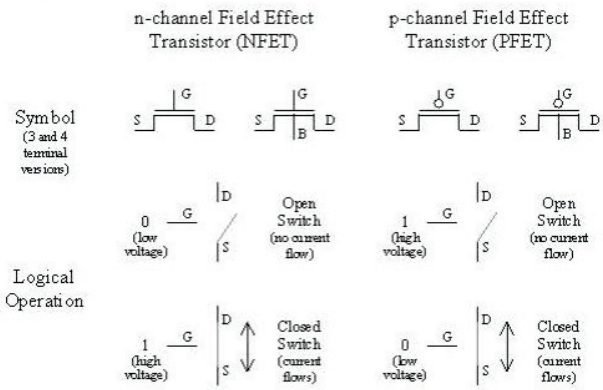


Figure 2. Symbol and Conceptual Switch Model of the NFET and PFET

In conceptual terms, the NFET and PFET both act like switches that can be closed (turned on) and thus conduct electrical current between the source and drain terminals. These switches turn on and off under control of the voltage applied to the gate terminal. The NFET model is that of a switch that is open when a low voltage (0) is applied to the gate and on when the gate is raised to a high voltage level (1). The PFET operates in a complementary fashion - it conducts current when the gate voltage is low (0) and turns off when the gate voltage is raised to a high level (1).

The simplest CMOS logic gate (the word gate is confusingly used to denote both the controlling structure in a FET and as a general term for a basic logic circuit, e.g. an AND gate) is called the inverter and its symbol, schematic and operation are shown in Figure 3.

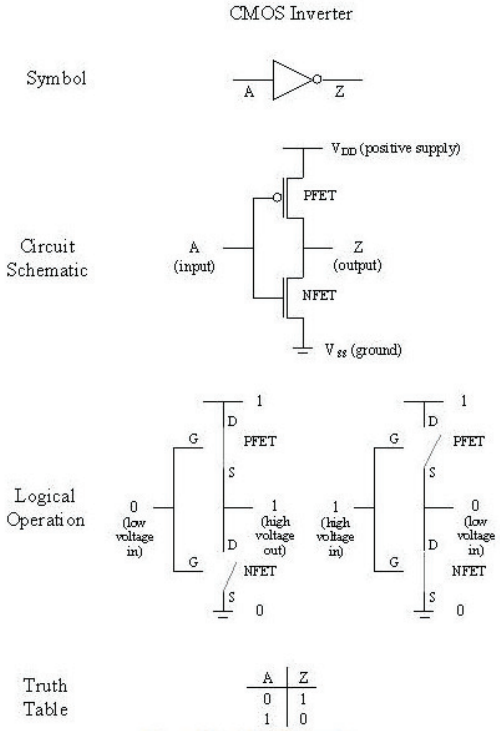
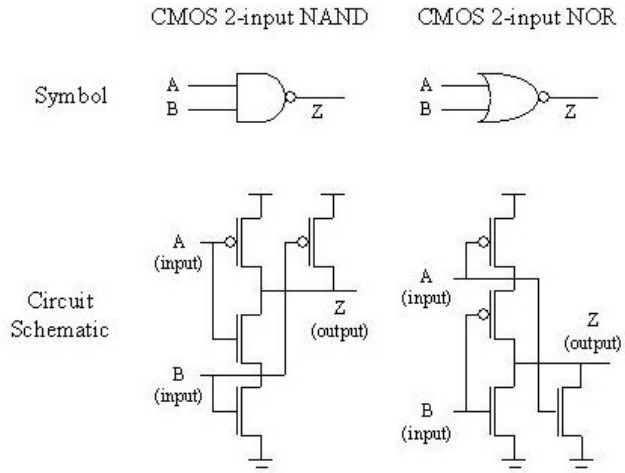


Figure 3 Basic CMOS Inverter

The inverter has terminals, an input and an output and it simply converts a logic 0 to a logic 1 and a logic 1 to a logic 0. Besides the obvious role as a means to flip the sense of a logic value the inverter is also used to restore and/or amplify weak or noisy logic signals. The basic CMOS inverter is composed of one NFET and one PFET. The input of the inverter connects to the gate of both transistors. The source of the NFET is tied to Vss (the low voltage power supply, equivalent to a logic 0) while its drain connects to the inverter output. The drain of the PFET is connected to Vdd (the high voltage power supply, equivalent to a logic 1) while its source connects to the inverter output.

Since the inverter input controls both the NFET and PFET simultaneously, one transistor is always on while the other is always off. If the input is a logic 0 then the PFET is turned on while the NFET is turned off. The PFET will source current from the Vdd supply to the output until its voltage level is raised to that of Vdd, which is a logic 1. If the input is a logic 1 then the PFET is turned off while the NFET is turned on. The NFET will sink current from the output to the Vss supply until the output voltage level falls to that of Vss, which is a logic 0.

The inverter is an essential logic element but by itself cannot do anything particularly useful. For that you need logic gates with two or more inputs. The simplest multiple input logic gates in CMOS are the two input negative-AND (NAND) gate and the two input negative-OR (NOR) gate. The NAND gate outputs a 0 when *all* its inputs are 1, otherwise its output is 1. The NOR gate outputs 0 when *any* of its inputs are 1, otherwise its output is 1. All possible Boolean logic function can be built using just NAND or NOR gates. Thus theoretically speaking it is possible to build an entire CPU entirely out of NAND or NOR gates although in practice this is often far from optimal. The symbol, schematic, and truth table for the basic CMOS two input NAND and NOR gates are shown in Figure 4.



Truth Table

A	B	Z
0	0	1
0	1	1
1	0	1
1	1	0

A	B	Z
0	0	1
0	1	0
1	0	0
1	1	0

Figure 4 Basic CMOS two input NAND and NOR gates

The NAND and NOR gates are specific examples of the general form of a static CMOS logic gate. That is, the output is connected to Vdd through a network of PFETs and to Vss through a complementary network of NFETs. In this context, complementary networks of PFETs and NFETs means that for every combination of inputs the output is always connected either just to Vdd through turned on PFET(s) or just to Vss through turned on NFET(s). A properly designed CMOS logic gate will never float the output (i.e. leave it not connected electrically to either Vdd or Vss) or create a short circuit path for current to flow from Vdd to Vss.

The two input NAND gate has two NFETs and two PFETs. The NFETs are in series while the PFETs are connected in parallel. That means that for the output to be low (0) both NFETs must be turned on (i.e. both inputs high or 1). The topology of the NOR gate is the same as that of the NAND gate but turned upside down. The output of the NOR gate is only high or 1 when both PFETs are turned on (i.e. both inputs are low or 0). Static CMOS NAND and NOR gates can be extended to any arbitrary number of inputs by adding more series and parallel FETs to the circuit, although in practice when NAND or NOR functions with fan-in (i.e. the number of inputs) greater than 3 or 4 are required these are usually implemented in a different fashion for performance reasons.

In the second installment of this two part article I will describe the operation of field effect tranistors in more detail using the example of a state of the art 0.13 um process and examine how FET characteristics affect logic performance. This will lead into an examination of the more advanced logic circuit design techniques used in modern MPUs to achieve the highest possible clock frequencies and performance.

Further Reading

[1] Maly, W., "Atlas of IC Technologies", Benjamin/Cummings Publishing Company Inc., ISBN 0-8053-8650-7, 1987.

[2] Weste, N. and Eshraghian, K., "Principles of CMOS VLSI Design, A Systems Perspective", Addison-Wesley Publishing Co, ISBN 0-201-08222-5, 1985.

<< Prev Page

Pages: Gates Within Gates Go

Copyright © 2002 - All Rights Reserved