

Fiber to the x (FTTX) is a generic term for any broadband network architecture using optical fiber to provide all or part of the local loop used for last mile telecommunications. As fiber optic cables are able to carry much more data than copper cables, especially over long distances, copper telephone networks built in the 20th Century are being replaced by fiber. **FTTX** is a generalization for several configurations of fiber deployment, arranged into two groups: **FTTP/FTTH/FTTB** (Fiber laid all the way to the premises/home/building) and **FTTC/N** (fiber laid to the cabinet/node, with copper wires completing the connection).

Telecommunications Optical Interconnect Requirements

Year	2015	2017	2019	2021	2023	2025	2027	2029
Data rate/wavelength single transceiver (Gb/s)	200	400	400	400	1000	1000	1000	1000
Bandwidth efficiency, Bits/ symbol	8	12	16	16	20	20	20	20
Modulation Method	DP-QPSKx 2	DP-16QAM	DP-16QAM	DP-16QAM	DP-64QAM	DP-64QAM	DP-64QAM	DP-64QAM
Max data rate/fiber	50Tb/s	50Tb/s	100Tb/s	100Tb/s	250 Tb/s	250 Tb/s	250 Tb/s	250 Tb/s
Distance between repeaters	80km to 1000km							
Optical wavelength	1310, 1550/1610	1310, 1550/1610	1310, 1550/1610	1310, 1550/1610	1310, 1550/1610	1310, 1550/1610	1310, 1550/1610	1310, 1550/1610
Single channel optical power	16dbm							
Total optical power	20dbm							
# wavelengths/waveguide	<168	<168	<168	<168	<168	<168	<168	<168
Wavelength spacing, GHz	50/100	50/100	50/100	50/100	50/100	50/100	50/100	50/100
Physical Modulation Method (direct or secondary)	Secondary modulator, PIC							
Optical mode; multi/single	single							

8. MORE MOORE

8.1. OVERALL TRANSISTOR TRENDS

Transistor cost and performance will continue to be strongly correlated to dimensional and functional scaling of CMOS as information processing technology is driving the semiconductor industry into a broadening spectrum of new applications according to 2015 ITRS 2.0.

Strained silicon, high- κ /metal-gate and Multigate transistors are now widely used in IC manufacturing. A significant part of the research to further improve device performance is presently concentrated on III-V materials and Ge. These materials promise higher mobilities than Si devices.

In order to take advantage of the well-established Si platform, it is anticipated that the new high-mobility materials will be epitaxially grown on Si substrate. Beyond implementation of these new materials, the Emerging Research Device (ERD) section reports completely new transistors, operating on new principles like tunneling (e.g. TFET) or spin that offer the possibility of operating at very low power. A lot of progress has been reported on these devices and their introduction into manufacturing appears consistent with the advent of the next decade.

Furthermore, a large variety (like never before) of new memory devices operating on completely new principles are extensively reported and some of them are approaching the pre-manufacturing phase.

32 Executive Summary

FIN FET devices can be simply characterized by observing that these transistors are nothing more than planar transistors rotated by 90 degrees along the source-gate-drain side. It has been already demonstrated that by means of a combination of horizontal scaling and by reducing the number of FINs it is possible to scale linear dimensions of transistors by more than historical 70%. Elimination of one FIN does not have any deleterious effect on performance since making the remaining FINs proportionally taller can compensate the lack of current carrying capability of the missing FIN. On the other hand, FIN FET transistors were primarily introduced to reduce leakage current and for being capable of operating at lower voltages than classical planar CMOS. Low power operation is now the most sought after attribute. Despite this enhanced scaling capability 2D scaling will eventually approach fundamental limits towards the end of this decade and therefore both logic and memory devices are exploring the use of the vertical dimension (3D) as a means of continuing packing more transistors in each cm^2 of silicon.

The combination of 3D device architecture and low power device will usher the (Third) Era of Scaling, identified in short as “3D Power Scaling”. Increase in the number of transistors per unit area will eventually be accomplished by stacking multiple layers of transistors

No new breakthroughs are reported for classical interconnections since no viable materials with resistivity below copper exist. However, progress in manipulation of edgeless wrapped materials (e.g., carbon nanotubes, graphene combinations etc.) offer the promise of “ballistic conductors,” which may emerge in the next decade.

3D integration of multiple dice offers possible avenues towards reducing interconnect resistance by increasing the conductor cross-section (vertical) and by reducing the length of each interconnect path. For instance, integrating memory device (die) immediately above logic device (die) and connecting them by means of wide through silicon vias (TSV) can accomplish this result.

8.2. MEMORY TRENDS.

Memory technologies will continue to drive the most aggressive pitch scaling and the highest transistor count. ***Memory technologies have been and will always be the drivers of Moore’s Law.***

DRAM products are approaching fundamental limitations as scaling DRAM capacitors is becoming very difficult in 2D structures. It is expected that these limits will be reached by 2024 and after this year DRAM technology will saturate at the 32Gbit level unless some major breakthrough will occur.

Flash memory on the other hand will lead the semiconductor industry towards the next revolution in transistor density. By 2020 the 2D Flash topological method will reach a practical limit with respect to cost effective realization of pitch dimensions. Due to the small volume of the floating gate charge amount and charge retention will become fundamental problems as fewer and fewer electrons will be available for the memory function. However, Flash producers are turning the transistors completely vertical in order to increase density while relaxing the dimensions of the storage gate to higher values. By so doing it will be possible to store more charge in the floating gate. Already prototypes of 32 and 48 layers of Flash transistor planes have been announced. This integration technology will become dominant in the next few years and will spread to logic circuits as well. However, the concept of “increased functionality” in the era of IoE is not longer exclusively equivalent to either increased transistor density or performance. New ways of integrating a system have become essential to producing system of higher functionality. The first method consists in extending the functionality of the CMOS platform via heterogeneous integration of new technologies, and the second method consists in stimulating inventions of devices that support new information-processing paradigms

DRAM TECHNOLOGY							
YEAR OF PRODUCTION	2015	2017	2019	2021	2024	2027	2030
<i>Half Pitch (Calculated Half pitch) (nm)</i>	24	20	17	14	11	8.4	7.7
<i>DRAM cell size (μm^2)</i>	0.00346	0.00240	0.00116	0.00078	0.00048	0.00028	0.00024
<i>DRAM cell FET structure</i>	RCAT+Fin	RCAT+Fin	VCT	VCT	VCT	VCT	VCT
<i>Cell Size Factor: a</i>	6	6	4	4	4	4	4
<i>Array Area Efficiency</i>	0.55	0.55	0.5	0.5	0.5	0.5	0.5
<i>V_{int} (support FET voltage) [V]</i>	1.1	1.1	1.1	1.1	0.95	0.95	0.95
<i>Support min. V_{in} (25C, $G_{m,max}$ $V_d=55mV$)</i>	0.40	0.40	0.40	0.40	0.37	0.37	0.37
<i>Minimum DRAM retention time (ms)</i>	64	64	64	64	64	64	64
<i>DRAM soft error rate (fits)</i>	1000	1000	1000	1000	1000	1000	1000
<i>Gb/1chip target</i>	8G	8G	16G	16G	32G	32G	32G

NAND Flash							
Year of Production	2015	2016	2020	2022	2024	2028	2030
<i>2D NAND Flash uncontacted poly 1/2 pitch - F (nm)</i>	15	14	12	12	12	12	12
<i>3D NAND minimum array 1/2 pitch -F (nm)</i>	80nm						
<i>Number of word lines in one 3D NAND string</i>	32	32-48	64-96	96-128	128-192	256-384	384-512
<i>Dominant Cell type (FG, CT, 3D, etc.)</i>	FG/CT/3D						
<i>Product highest density (2D or 3D)</i>	256G	384G	768G	1T	1.5T	3T	4T
<i>3D NAND number of memory layers</i>	32	32-48	64-96	96-128	128-192	256-384	384-512
<i>Maximum number of bits per cell for 2D NAND</i>	3	3	3	3	3	3	3
<i>Maximum number of bits per cell for 3D NAND</i>	3	3	3	3	3	3	3

8.3. LOGIC TRENDS

The advent of ubiquitous access to the Internet by means of Wi-Fi communication technologies has led to a vast deployment of mobile products. The need for extended time of operation by means of energy provided by a battery has changed the driver of transistor performance from maximum speed to reduced power consumption. This is accomplished by reducing power supply voltage and by designing transistors with steeper sub threshold slope. However, the need for improved performance is reflected in the drive for V_t reduction of more than 100% in the time horizon for Low (Power) Performance transistors to partially compensate for the reduction of power supply voltage.

Electrical Properties of Transistors

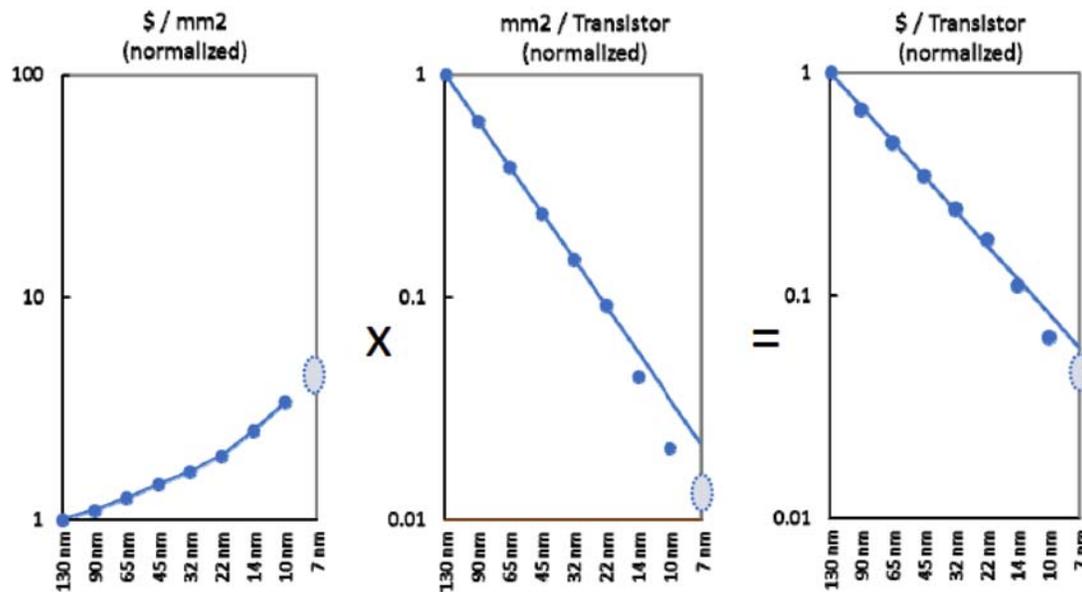
YEAR OF PRODUCTION	2015	2017	2019	2021	2024	2027	2030
Logic device technology naming	P70M56	P48M36	P42M24	P32M20	P24M12G1	P24M12G2	P24M12G3
Logic industry "Node Range" Labeling (nm)	"16/14"	"11/10"	"8/7"	"6/5"	"4/3"	"3/2.5"	"2/1.5"
Logic device structure options	FinFET FDSOI	FinFET FDSOI	FinFET LGAA	FinFET LGAA VGAA	VGAA, M3D	VGAA, M3D	VGAA, M3D

DEVICE ELECTRICAL SPECS							
Power Supply Voltage - V_{dd} (V)	0.80	0.75	0.70	0.65	0.55	0.45	0.40
Subthreshold slope - [mV/dec]	75	70	68	65	40	25	25
Inversion layer thickness - [nm]	1.10	1.00	0.90	0.85	0.80	0.80	0.80
V_t sat (mV) at $I_{off}=100nA/\mu m$ - HP Logic	129	129	133	136	84	52	52
V_t sat (mV) at $I_{off}=100pA/\mu m$ - LP Logic	351	336	333	326	201	125	125
Effective mobility ($cm^2/V.s$)	200	150	120	100	100	100	100
R_{ext} (Ohms. μm) - HP Logic	280	238	202	172	146	124	106
Ballisticity. Injection velocity (cm/s)	1.20E-07	1.32E-07	1.45E-07	1.60E-07	1.76E-07	1.93E-07	2.13E-07
V_{dsat} (V) - HP Logic	0.115	0.127	0.136	0.128	0.141	0.155	0.170
V_{dsat} (V) - LP Logic	0.125	0.141	0.155	0.153	0.169	0.186	0.204
Ion ($\mu A/\mu m$) at $I_{off}=100nA/\mu m$ - HP logic w/ $R_{ext}=0$	2311	2541	2782	2917	3001	2670	2408
Ion ($\mu A/\mu m$) at $I_{off}=100nA/\mu m$ - HP logic, after R_{ext}	1177	1287	1397	1476	1546	1456	1391
Ion ($\mu A/\mu m$) at $I_{off}=100pA/\mu m$ - LP logic w/ $R_{ext}=0$	1455	1567	1614	1603	2008	1933	1582
Ion ($\mu A/\mu m$) at $I_{off}=100pA/\mu m$ - LP logic, after R_{ext}	596	637	637	629	890	956	821
C_{ch} , total (fF/ μm^2) - HP/LP Logic	31.38	34.52	38.35	40.61	43.14	43.14	43.14
C_{gate} , total (fF/ μm) - HP Logic	1.81	1.49	1.29	0.97	1.04	1.04	1.04
C_{gate} , total (fF/ μm) - LP Logic	1.96	1.66	1.47	1.17	1.24	1.24	1.24
CV/I (ps) - FO3 load, HP Logic	3.69	2.61	1.94	1.29	1.11	0.96	0.89
$I/(CV)$ (1/ps) - FO3 load, HP Logic	0.27	0.38	0.52	0.78	0.90	1.04	1.12
Energy per switching [CV ²] (fj/switching) - FO3 load, HP Logic	3.47	2.52	1.89	1.24	0.94	0.63	0.50

Requirements for power reduction apply also to the High Performance transistors. Similarly, despite these reductions in operating voltage it is expected that Ion for both types of transistors will remain almost constant across the time horizon and this is not an easy result to accomplish. Total channel capacitance will increase by about 30% even though gate capacitance will continue to decrease. Finally transistor speed will continue to increase by a 4X factor for HP transistors enabling 1THz intrinsic functionality. Overall energy per switching will decrease to about 15% from nowadays values.

The introduction of FinFET transistors into manufacturing in 2011 revolutionized the way transistors are built and it resulted essential not only to extending Moore's Law but to also to accelerate the pace of reduction of transistor and pitch dimensions in this decade [Fig.8.1].

Offsetting Wafer Cost with Density



Source: Intel
© 2016 IEEE
International Solid-State Circuits Conference

1.1: Moore's Law: A Path Forward

17 of 30

From: 2016 ISSCC. "Moore's Law: A Path Forward". Bill Holt, Intel

Fig. 8.1 Increased transistor density enabled by FinFET

However, it is expected that new types of transistors will be required in the next decade to accomplish the indicated reduction in Subthreshold slope. Beginning with 2005 an extensive search for a new "switch" was initiated across the world under very broad guidelines. Multiple new types of switches were identified by 2010 and substantial experimental results were reported in 2015. Early on the new desirable properties of the new "switches" were classified in three categories:

1. *The new devices were required to function in at least two separate logic states but the more the better*
2. *It was desirable that the devices could operate with minimal power consumption (i.e., consuming less power than CMOS devices) and with no power consumption at all in their stand-by mode*
3. *It was also desirable that the devices should have the ability to retain memory information with essentially no power consumption (i.e., no power supply)*

Based on past experience few fundamental modes of physical operation were identified. In one case the devices relied on flow of electric charge, in another case the devices relied on magnetic properties and finally some devices would operate in a completely new way. While the first mode of operation could still be associated with charges flowing from one location to another the second mode of operation was associated with stationary magnetic dipoles that did not consume any power in their stand by condition. It was quickly realized that even though it was well known that electrons carried a negative charge it was also known that they also carried a magnetic dipole (spin) associated with them even though this latter property had never been used for construction of commercial Integrated Circuits.

In 2010 an extensive review of progress on novel devices and possible applications was published in the Proceeding of IEEE (Vol.98, No.12, December 2010). Among these devices the *tunnel transistor* appeared as a leading candidate with respect to operation at very low voltage. The operation can be explained in very simple terms. In a typical NMOS

36 Executive Summary

transistor the charges in the conduction band of the source region are prevented from flowing to the drain region by the potential barrier generated by the NP junction existing between the source and the body of the transistor. Application of positive voltage to the gate lowers this barrier and lets electrons flow to the drain. However, the electrons in the source region have an energy distribution that spreads to values exceeding the potential barrier even when there is no voltage applied to the gate. Under these conditions some level of leakage cannot be eliminated. Conceptually, it is possible to completely eliminate the effect of the tail of the electron energy distribution by deciding to operate via the electrons located in the valence band of the source region. Under these conditions charges would not be able to overcome the combination of the band gap and junction barrier voltage when operating in a stand-by mode! While it is true that any leakage from source to drain could be eliminated under these conditions it is also true that in order to allow any current at all to flow from the source to the drain region it would be necessary to somehow apply a voltage that would overcome the combination of band-gap and PN junction potentials. So, no leakage but also no current could flow under normal voltage operations.

However, Prof. Leo Esaki had demonstrated that when two energy bands are brought into very close proximity, charges could flow from one band to the other by tunneling through this band-to-band potential barrier. He received a Nobel Prize in 1973 for this invention.

Further improvements in the practical implementation of this concept to transistor fabrication has led to very promising results in recent years. It has been demonstrated that, under proper conditions, flow of current can occur at very low gate voltage and leakage current is practically negligible in the off state. Subthreshold slopes below 30mV/dec have been demonstrated. The research effort is now concentrated on selecting the appropriate materials. However, while essentially steep subthreshold transition from off to on state have been observed when TFETs operate at low temperature (i.e., 28mV/Dec at 77K) at present leakage increases as operation approaches the 300K temperatures. This is due to unwanted electronic states that allow current to flow between the two bands. The problem to be solved is not different from the problem of eliminating surface states in early MOS that scientists had to deal with for 20 years. This time the “states” are buried somewhere between the two bands across which the charges have to tunnel. Among the many proposed solution the use of 2D material seems very promising since their very own structure eliminates one set of undesirable “dangling bonds”.

Several devices utilizing the spin property of electrons have also shown promising results.

In particular, the concept of operating a device with current (dynamic mode) but then storing the result by means of a magnetic state (static mode) has been demonstrated. Spin-transfer torque (STTM) is an effect in which the orientation of a magnetic layer in a magnetic tunnel junction or spin valve can be modified using a spin-polarized current. Once the magnet is polarized no current is required to keep the magnet in this state. In essence this device embodies both the current carrying properties of electrons as well as the ability of transferring magnetic information to create a permanent state in a magnetic layer. This effect represents the ability to electrically program a magnetic memory that can store information without using any energy.

These few are just few examples indicating that by the year 2020 several new devices will be available to work in conjunction or better than CMOS on some specific applications.

“Moore’s Law is dead, long live Moore’s Law”

The question of how long will Moore’s Law last has been posed an infinite number of times since the 80s and every 5-10 years publications claiming the end of Moore’s Law have appeared from the most unthinkable and yet “reputedly qualified” sources. Despite these alarmist publications the trend predicted by Moore’s Law has continued unabated for the past 50 years by morphing from one scaling method to another, where one method ended the next one took over. This concept has completely eluded the comprehension of casual observers that have mistakenly interpreted the end of one scaling method as the end of Moore’s Law. As stated before, bipolar transistors were replaced by PMOS that were replaced by NMOS that were also replaced by CMOS. Equivalent Scaling succeeded Geometrical Scaling when this could no longer operate and now 3D Power Scaling is taking off.

By 2020-25 device features will be reduced to a few nanometers and it will become practically impossible to reduce device dimensions any further. At first sight this consideration seems to prelude to the unavoidable end of the integrated circuit era but once again the creativity of scientists and engineers has devised a method

“To snatch victory from the jaws of defeat”

The basic concept of this solution is actually rather obvious if we only observe places like Manhattan, Tokyo, Seoul or Hong Kong; once real estate space was fully utilized people discovered the unexplored space of the vertical dimension and began building skyscrapers.

Nowadays, Flash memory producers are facing a similar problem as they are running out of space to further scale Flash transistors. Furthermore, cost of producing integrated memory circuits of small dimensions keeps on rising while the number of stored electrons in the floating gate keeps on decreasing. To eliminate this problem Flash memory producers have already demonstrated and announced several new products that stack multiple layers of memory transistors on top of each other in a single integrated circuit. As many as 32 and 48 layers of Flash memory have been reported. Flash memory devices constituted by more than 100 layers have been predicted. The basic approach consists in building transistors that are 100% vertically oriented. Arrays of columns of 32 or 48 transistors have been built to increase transistor density as measured by the number of transistors/cm² and reduce cost since this approach requires fewer mask layers than traditional 2D manufacturing methods.

This vertical trend is led by Flash memories but it is expected that it will become an industry wide trend in the next decade. FinFET structures were built to control as much as possible the potential in the transistor channel. It was however anticipated by the ITRS in the 1998-2000 timeframe that eventually transistors would be fabricated with gates completely surrounding the semiconductor. Orienting the transistor substrate vertically and then completely surrounding it with a sequence of dielectric and metal layers deposited by means of deposition to fabricate the composite gate structure can more easily be accomplished if the transistor is vertically oriented. It is clear that this method reduces the transistor footprint and in conjunction with creating multiple layers of transistors one on top of the other will accelerate the level of transistor density beyond Moore's Law traditional trends.