Migration from Electronics to Photonics in Multicore Processor

by

Zhoujia Xu

B.Eng, Electrical Engineering (2007)
National University of Singapore

Submitted to the Department of Materials Science and Engineering
in Partial Fulfilment of the Requirements for the Degree of

Master of Engineering in Materials Science and Engineering

at the

Massachusetts Institute of Technology

September 2008

Signature of Author: ______________________________
Department of Materials Science and Engineering
August 8, 2008

Certified by: ______________________________
Lionel C. Kimerling
Thomas Load Professor of Material Science and Engineering
Thesis Supervisor

Accepted by: ______________________________
Samuel M. Allen
POSCO Professor of Physical Metallurgy
Chair, Departmental Committee for Graduate Students

Migration from Electronics to Photonics

by

Zhoujia Xu

Submitted to the Department of Materials Science and Engineering
on August 8, 2008 in Partial Fulfillment of
the Requirement for the Degree of
Master of Engineering in Materials Science and Engineering

Abstract:

Twenty – first opportunities for Gigascale Integration will be governed in part by a hierarchy of physical limits on interconnect. Microprocessor performance is now limited by the poor delay and bandwidth performance of the on – chip global wiring layer. This thesis is envisioned as a critical showstopper of electronic industry in the near future. The physical reason behind the interconnect bottleneck is the resistive nature of metals. The introduction of copper in place of aluminum has temporarily improved the interconnect performance, but a more disruptive solution will be required in order to keep the current pace of progress, optical interconnect is an intriguing alternative to metallic wires. Many – core microprocessors will push performance per chip from the 10 gigaflop to the 10 teraflop range in the coming decade. Pin limitations, the energy cost of electrical signaling, and the non – scalability of chip – length global wires are significant bandwidth impediments. Silicon nanophotonic based many core architecture are introduced in order to meet the bandwidth requirements at acceptable power levels.

Thesis Supervisor: Lionel C. Kimerling

Title: Thomas Lord Professor of Materials Science and Engineering

Acknowledgement

There are a lot of people who have made it possible for me to be at this stage, and have helped me in my academic and personal growth. I would like to thank sincerely all of them.

I would like to thanks Professor Kimerling for his guidance and encouragement throughout this thesis, and his incredible insights that I benefited so much from.

Thanks to the whole EMAT group for being very supportive in this endeavor and my year at MIT.

I wish to express my sincere gratitude to my SMA friends, especially, Kwan Wee, Naga, Yu Yan, Luo Jia, they really helped me a lot in my thesis work, and they are my dictionary.

Last but no least, I would like to thank my wonderful parents, girl friend RTT, for their love and teaching that has brought me to where I am. Without them I would never have made it here in the first place.

Table of Contents:

Abstract: 3

Table of Contents: 7

Chapter One: Introduction 9

1.1 Electrical Interconnect Showstoppers 9

Chapter Two: Electrical Interconnect 13

2.1 Global and Local Interconnects 13

2.2 The Electronic Interconnect Speed Bottleneck 15

2.3 The Power Consumption, Signal Integrity and Trades – off 18

Chapter Three: Interconnect Limits 20

3.2 Device Limits 22

Chapter Four: Optics in Microprocessor 24

4.1 Introduction 24

4.2 Future Trend: Thousand Core Chips 26

4.3 Multicore bottleneck 30

4.4 Photonic Technology 32

4.5 Current multicore chips topology 34

4.5.1 Full Mesh Interconnect 34

4.5.2 Shared bus 34

4.5.3 Pipeline 35

4.5.4 Switched 36

Chapter Five: Suggested Architecture Overview 37

5.1 ATAC Architecture 38

5.1.2 Architecture Overview 38

5.2 The Corona Architecture 41

5.2.1 Cluster Architecture 41

5.2.2 Optical Crossbar 42

5.3 Optics vs Electronics 45

Chapter Six: Cost Analysis 46

6.1 Optical Cost Issue 46

6.2 Target bandwidth – cost 47

Chapter Seven: Multi-Core Market Analysis 48

7.1 Multi-core 48

7.2 Intel vs AMD 50

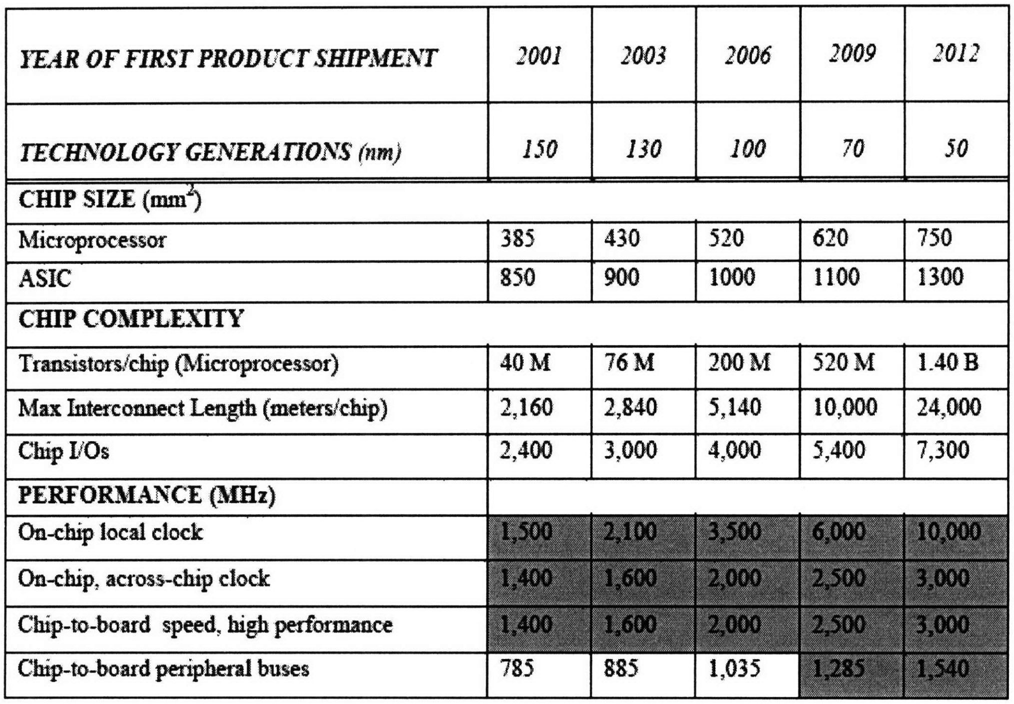
7.3 Start-up Company: 51

Chapter Eight: Conclusion 53

Chapter One: Introduction

1.1 Electrical Interconnect Showstoppers

The movement of data in a computer is almost the converse of the movement of traffic in a city. Downtown, in the congested core of the microprocessor, the bits fly at an extraordinary rate. But further out, on the broad avenues of copper that link one processor to another and one circuit board to the next, things slow down to a comparative crawl. A Pentium 4 introduced this spring operates at 2.4 GHz, but the data travels on bus operating at only 400GHz[1]. Current trends indicate that future chip multiprocessors (CMPs) may comprise hundred or even thousands of processing elements. Feeding data to so many on – chip cores, however, will be possible only if architecture and technology developments provide sufficient chip to chip and on – chip communication performance.



YEAR OF FIRST PRODUCT SHIPMENT	*2001*	*2003*	*2006*	*2009*	*2012*
TECHNOLOGY GENERATIONS (nm)	*150*	*130*	*100*	*70*	*50*
CHIP SIZE (mm^2)					
Microprocessor	385	430	520	620	750
ASIC	850	900	1000	1100	1300
CHIP COMPLEXITY					
Transistors/chip (Microprocessor)	40 M	76 M	200 M	520 M	1.40 B
Max Interconnect Length (meters/chip)	2,160	2,840	5,140	10,000	24,000
Chip I/Os	2,400	3,000	4,000	5,400	7,300
PERFORMANCE (MHz)					
On-chip local clock	1,500	2,100	3,500	6,000	10,000
On-chip, across-chip clock	1,400	1,600	2,000	2,500	3,000
Chip-to-board speed, high performance	1,400	1,600	2,000	2,500	3,000
Chip-to-board peripheral buses	785	885	1,035	1,285	1,540

Table 1: Technology Trends, Complexity, and Performance Requirements. [2]

Even though the design challenges have its roots in on – chip interconnects complexity and speed bottleneck, it extends to the package and system level, as clearly illustrated by the data in Table 1. This table illustrates trends in chip, package and system complexity and performance. From Table 1 it is apparent that circuit designers are faced with an unprecedented technical challenge, namely, the design and integration of giant microwave circuits, in which massive arrays of transistors are combined into compact, multi-functional systems, and subsequently packaged into low-cost products. The challenges are many. However, the interconnect bottleneck surfaces as the major showstopper at this point.

Optical technology and 3D integration are two potential solutions to chip – to – chip communication performance limitations. Still, on – chip communication faces considerable technological and architectural challenges. For example, global electrical interconnects do not scale well with technology[3]. Although delay – optimized repeater insertion and proper wire sizing can keep the delay nearly constant, this comes at the expense of power and active area, as well as a reduction in bandwidth. Researchers have developed techniques for optimizing the power – delay product, but these techniques yield neither optimal power nor optical latency[4]. This and other technological issue – such as manufacturability, conductivity, crosstalk, and so on – constitute important roadblock. As more cores are integrated, we expect the on – chip interconnect to take an increasingly large fraction of chip area and power budgets.

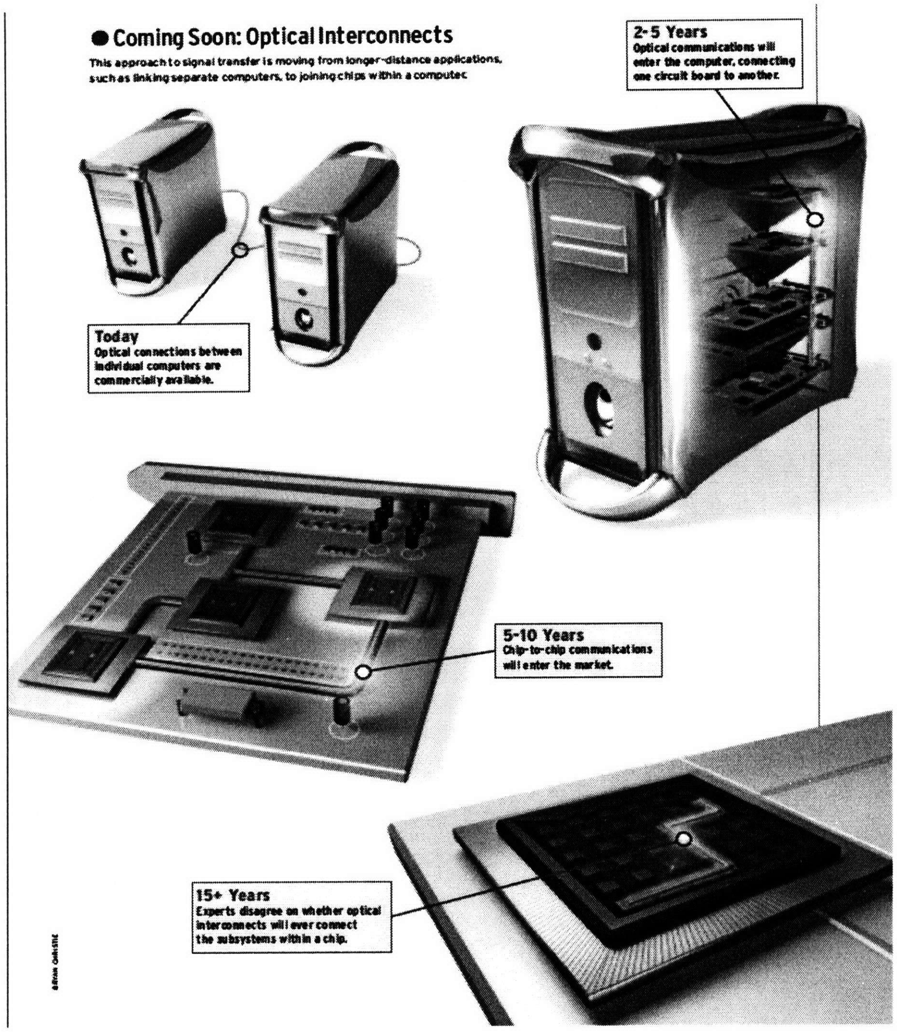

● Coming Soon: Optical Interconnects
This approach to signal transfer is moving from longer-distance applications, such as linking separate computers, to joining chips within a computec
2-5 Years
Optical communications will enter the computer, connecting one circuit board to another.
Today
Optical connections between individual computers are commercially available.
5-10 Years
Chip-to-chip communications will enter the market.
15+ Years
Experts disagree on whether optical interconnects will ever connect the subsystems within a chip.
©Bryan Christie

Figure 1: Future optical interconnects [1]

Whereas 10 years ago electrical – to – optical translation costs and CMOS incompatibility appeared to be insurmountable barriers to the use of optics in on – chip communication, today the outlook is dramatically brighter. Because of rapid progress in the past five years in CMOS – compatible detectors, modulators, and even light sources, the latest ITRS considers on – chip optical interconnects as a potential replacement for global wires by 2013. in global signaling applications, optical interconnects have the potential to fare favorably compared to their electrical counterparts, owing to their high speed, high bandwidth, low on – chip power, good electrical isolation, low electromagnetic interference, and other characteristics[5].

Chapter Two: Electrical Interconnect

2.1 Global and Local Interconnects

Technologically, on – chip interconnect is identified as electrical wiring. The definition of interconnect from International Technology Roadmap for Semiconductors (ITRS) is an "electrical wiring system which distributes clock and other signals, and provides power / ground to and among the various circuits / systems functions on a chip" [6]

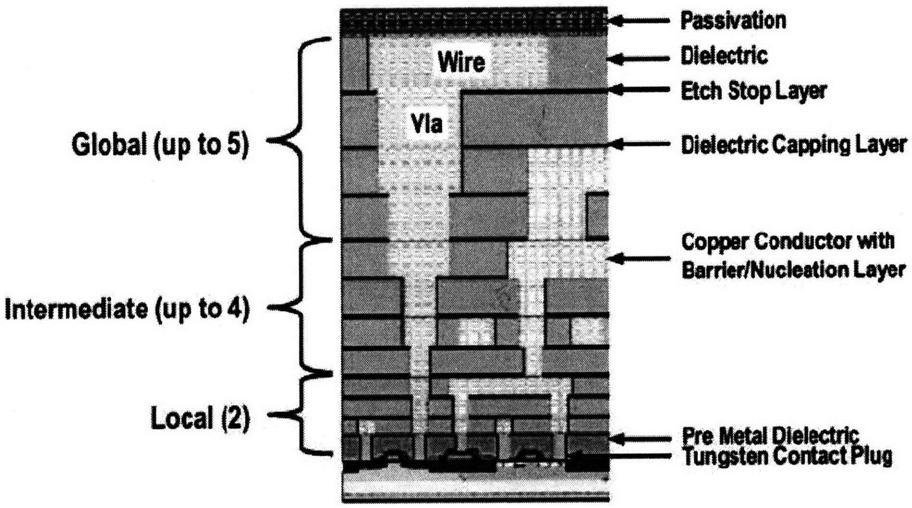


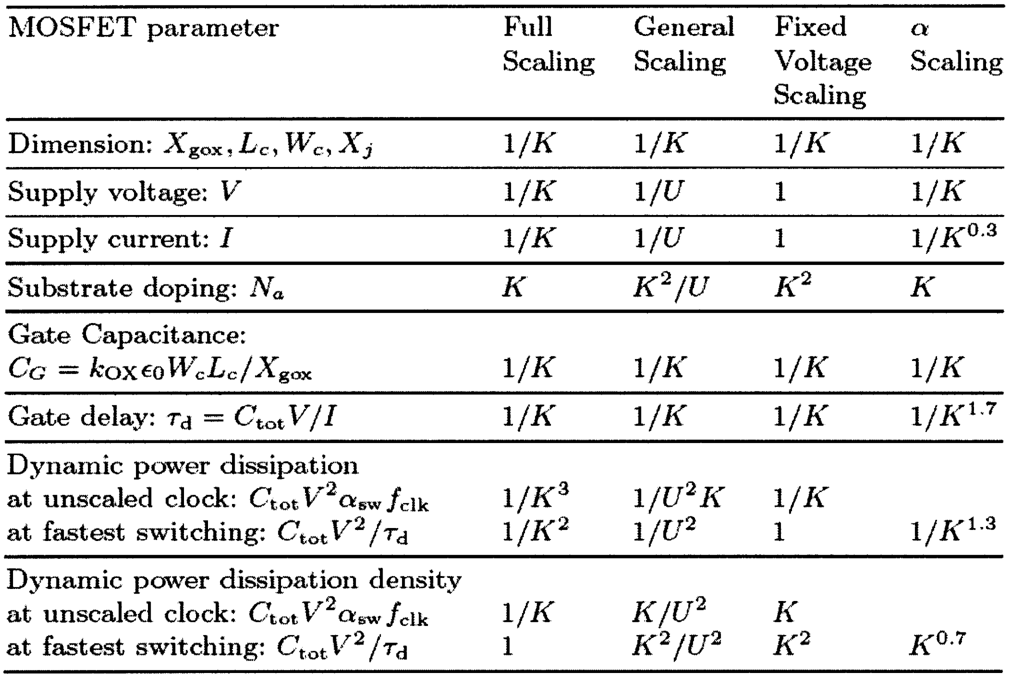
Figure 2: Schematic View of the back – end structure of electrical interconnect[6]

A typical interconnect structure is shown in Fig. 2. It is important to differentiate between local and global intrachip interconnects. Local interconnects have a delay of less than one clock cycle, while global interconnects typically take longer than one or two clock cycles, they are used to distribute clock signals and power among them. They can be much longer than gate size, and span over the whole chip size, reaching lengths of the order of cm. Thus, the length of global wires scales down with chip rather gate size. [4]

Local interconnects are used for short – distance communication and comprise the majority of on – chip wires. While there are fewer global interconnect, these links are no less important. Improving the performance of a small number of critical global links can significantly enhance the total system performance.

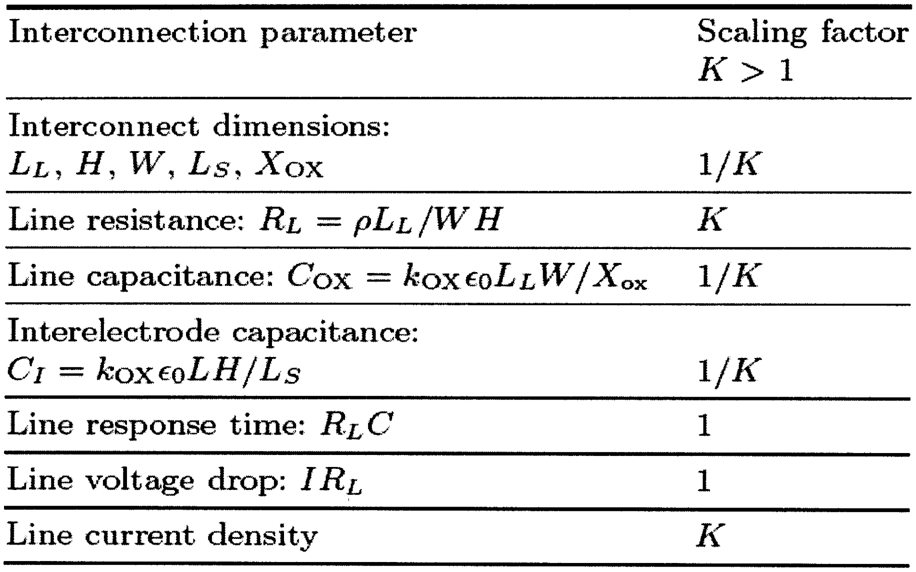
As you can see in the Fig. 2, local wires occupy the lower levels of interconnect and often can afford somewhat significant resistivity if they are very short, but must withstand higher process temperature than global interconnect. Material used for local interconnect include polysilicon, silicided gates, TiN and W. It is paramount that the resistivity of global wires is as low as possible, thus only metals (Al or Cu) are employed for global interconnect. [7]

2.2 The Electronic Interconnect Speed Bottleneck



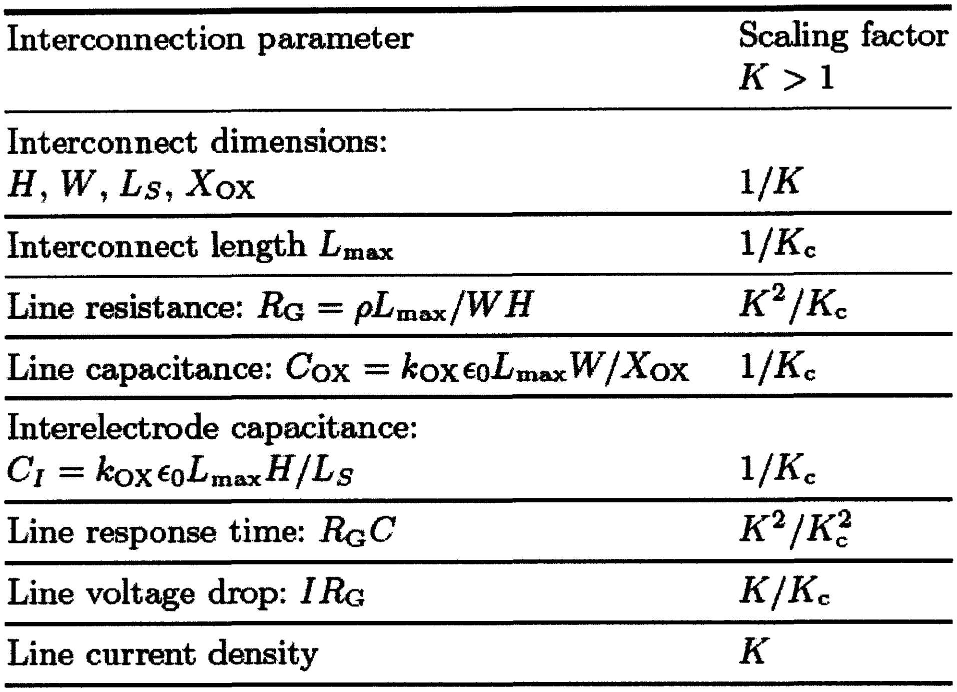
MOSFET parameter	Full Scaling	General Scaling	Fixed Voltage Scaling	α Scaling
Dimension: $X_{\text{gox}}, L_c, W_c, X_j$	$1/K$	$1/K$	$1/K$	$1/K$
Supply voltage: V	$1/K$	$1/U$	1	$1/K$
Supply current: I	$1/K$	$1/U$	1	$1/K^{0.3}$
Substrate doping: N_a	K	K^2/U	K^2	K
Gate Capacitance: $C_G = k_{\text{OX}}\epsilon_0 W_c L_c / X_{\text{gox}}$	$1/K$	$1/K$	$1/K$	$1/K$
Gate delay: $\tau_{\text{d}} = C_{\text{tot}} V/I$	$1/K$	$1/K$	$1/K$	$1/K^{1.7}$
Dynamic power dissipation				
at unscaled clock: $C_{\text{tot}} V^2 \alpha_{\text{sw}} f_{\text{clk}}$	$1/K^3$	$1/U^2 K$	$1/K$	
at fastest switching: $C_{\text{tot}} V^2 / \tau_{\text{d}}$	$1/K^2$	$1/U^2$	1	$1/K^{1.3}$
Dynamic power dissipation density				
at unscaled clock: $C_{\text{tot}} V^2 \alpha_{\text{sw}} f_{\text{clk}}$	$1/K$	K/U^2	K	
at fastest switching: $C_{\text{tot}} V^2 / \tau_{\text{d}}$	1	K^2/U^2	K^2	$K^{0.7}$

Table 2: MOSFET scaling. K is scaling factor (K > 1 implies that devices are shrinking)[7]



Interconnection parameter	Scaling factor $K > 1$
Interconnect dimensions: L_L, H, W, L_S, X_{OX}	$1/K$
Line resistance: $R_L = \rho L_L / WH$	K
Line capacitance: $C_{\text{OX}} = k_{\text{OX}}\epsilon_0 L_L W / X_{\text{ox}}$	$1/K$
Interelectrode capacitance: $C_I = k_{\text{OX}}\epsilon_0 LH/L_S$	$1/K$
Line response time: $R_L C$	1
Line voltage drop: IR_L	1
Line current density	K

Table 3: Local interconnect scaling[7]



Interconnection parameter	Scaling factor $K > 1$
Interconnect dimensions: H, W, L_S, X_{OX}	$1/K$
Interconnect length $L_{\mathbf{max}}$	$1/K_c$
Line resistance: $R_G = \rho L_{\mathbf{max}}/WH$	K^2/K_c
Line capacitance: $C_{OX} = k_{OX}\epsilon_0 L_{\mathbf{max}} W/X_{OX}$	$1/K_c$
Interelectrode capacitance: $C_I = k_{OX}\epsilon_0 L_{\mathbf{max}} H/L_S$	$1/K_c$
Line response time: $R_G C$	K^2/K_c^2
Line voltage drop: IR_G	K/K_c
Line current density	K

Table 4: Global interconnect scaling. Kc is the chip-scaling factor[7]

In contract to transistors, downscaling of interconnect does not enhance speed performance. Scaling effect for transistors, local and global wires are summarized above, respectively. In 1980, a MOSFET's delay was about 20ps, whereas the latency associated to a typical aluminum electrical interconnect 1.0 mm long surrounded by SiO_2, was 6.6ps. At that time, the interconnect delay was propagation limited, i.e. the delay bottleneck was the time – of – flight of electromagnetic waves associated to SiO_2, rather than the RC time constant, which is around 1 ps. According to prediction for the 35 nm technology generation (expected to appear in production in 2014), the estimated interconnect response time of a 1.0 mm copper line with a low – k dielectric (k ≈2) is τ = 250 ps[8]. Such value is mainly contributed by RC time constant, which has grown by more than 2 orders o magnitude since the 80s. The global RC delay grows because of scaling, according to Table 4. In comparison, the switching delay of a minimum geometry 35 nm generation MOSFET is τ = 2.5 ps, 100 times faster than interconnect. This simple numerical example emphasizes how the speed performance bottleneck has shifted over the years from gates to interconnect. For 10 nm technology, a 1000 ratio between interconnect latency and gate delay is expected[8].

2.3 The Power Consumption, Signal Integrity and Trades – off

The consumed power density in ICs is rising exponentially along scaling[6]. This can be quantitatively understood as follows.

A scaling factor K > 1 means that linear dimensions of gates are multiplied by 1/K. the capacitance at gate output C scales as 1/K ($C \propto 1/K$). This applied to all capacitive elements because capacitor's area scales as $1/K^2$ and capacitor's thickness scales as 1/K (for a ideal parallel plate capacitor, $C = \varepsilon A/d$).

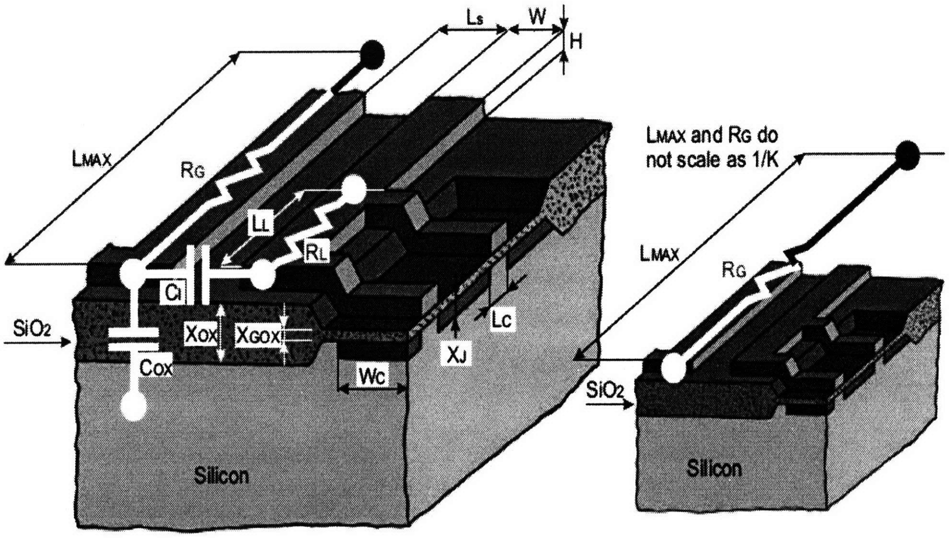


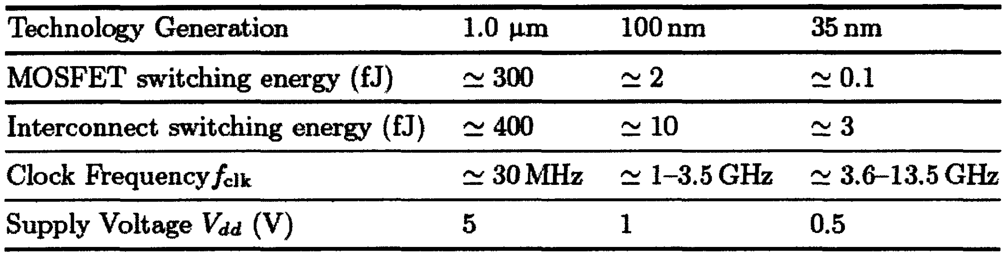
Figure 3: Global lines do not scale their length $L_{max,}$ whereas local lines L_L and MOSFET do[7]

Local wires connecting one gate's output to a following gate are the dominant contribution to this capacitive load C. If DC supply voltage V_{dd} does not change, the dynamically moving charge per gate is $Q = CV \propto 1/K$, and energy per cycle per gate is $E = CV_{dd}^2 \propto 1/K$.

The energy gets all dissipated. The dissipated power per device P is E times the gate switching frequency f, which is proportional to clock frequency f_{clk}. The proportionality factor α_{sw} is called gate switching activity ($f = \alpha_{sw} \times f_{clk}$). Reasonable assumptions can be that $\alpha_{sw} \approx 1/4$. Therefore, given the number of gates per unit area ($N = K^2$), the power density (power per unit area) $\wp$ is given by:

$$\wp = NCV_{dd}^2\alpha_{sw}\ f_{clk} \propto Kf_{clk}$$

From this equation, it is apparent that dynamic power density is inversely proportional to gate length, and that increasing the clock frequency f_{clk} exacerbates the power density problem.



Technology Generation	1.0 μm	100 nm	35 nm
MOSFET switching energy (fJ)	≃ 300	≃ 2	≃ 0.1
Interconnect switching energy (fJ)	≃ 400	≃ 10	≃ 3
Clock Frequency f_{clk}	≃ 30 MHz	≃ 1–3.5 GHz	≃ 3.6–13.5 GHz
Supply Voltage V_{dd} (V)	5	1	0.5

Table 5: Projected comparison between the dissipated energy[8]

Above table 5 have compared the energy dissipated by a MOSFET and by a 1 mm long interconnect for a single switch, for three technology generations. Table 4 emphasize that in earlier technologies, the power dissipated by the benchmark interconnect was comparable to MOSFET's dissipation, whereas the projection for 35 nm generation reveals a ratio of 30 between interconnect and MOSFET dissipation.

Chapter Three: Interconnect Limits

3.1 Material Limits

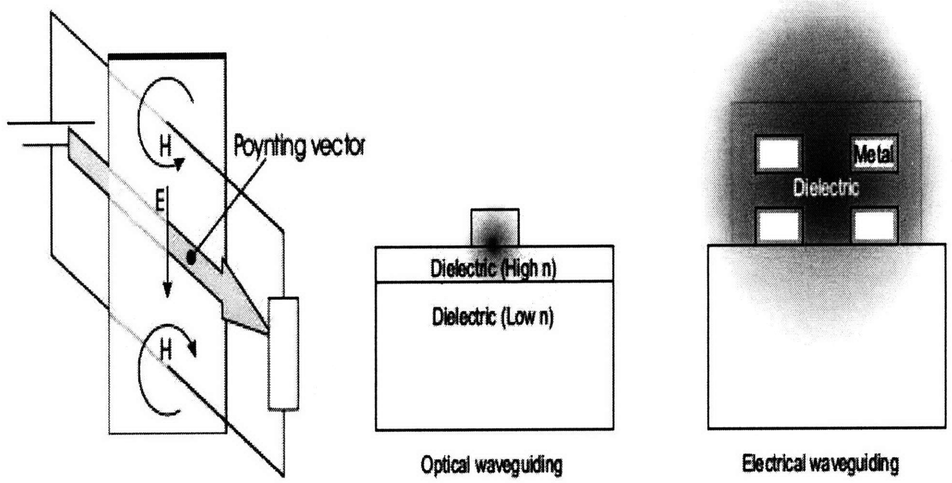


Figure 4: Direction of Poynting vector in an electrical communication and waveguiding.

An ideal conductor – i.e., an material in which charges can move under no electrical field, the Poynting vector steers towards the receiver, as shown in Fig. 4. The plot also emphasizes that the wave propagates inside the dielectric between the ideal metals, rather than in the metals themselves. Dielectric properties control the propagation characteristics, including the speed.

This kind of ideal conductors should be imagined as containing free charge carriers with zero mass. Charge carriers would act as “slaves” of electromagnetic fields in the dielectric, adjusting themselves, at the metal surface, as fast as required to make the fields transversal. In this sense, metallic “waveguiding” is achieved, by letting the charge carriers inside the metal (the cladding) to screen the field, confining it into the dielectric (the core). Unfortunately, microstrip and coplanar waveguides circuits are open waveguides, which means that the dielectric is in contact with basically all the IC’s metallic wires. This is the physical origin the **cross – talk** problems in electrical connections. In contrast, small optical wavelength permit wave guiding using total internal reflection in dielectric materials. The wave is effectively confined and smart cladding will allow dense packing with no major cross – talk problem.

However, metals are no ideal conductors. The **first reason** is that charge carriers have a nonzero mass. Free carriers are resonantly excited only at zero frequency, and their ability to follow electromagnetic fields drops with frequency. Beyond plasma frequency:

$$\omega_p = \sqrt{\frac{4\pi nq^2}{\varepsilon_0 m_e}}$$

Electrons loose any capability of following the radiation, which freely enters the metal without being absorbed.

The **second non - ideality** is damping, a mechanism which transfer mechanical energy from carriers to the solid lattice. Damping allows penetration of waves at frequencies below plasma frequency, where absorption is still significant. This is mechanism is the origin of the "skin effect", hinting at a thin sheet at the metallic inner surface, of thickness δ, in which carriers move for a given frequency.

$$\delta = \sqrt{\frac{1}{2\pi f \sigma \mu}}$$

For ideal metals, since conductivity σ is infinity, hence the thickness would be 0 at all frequencies. In real metals, however, the skin effect is normally associated with the idea of a reduction of the metal cross – section.

The **third non - ideality** is non – zero, finite resistivity which leads to power dissipation (Joule effect) and to speed limiting effects, the most known being the RC delay. Hence it is advantageous to employ low resistivity materials as conductors.

3.2 Device Limits

No matter the device is RC effect dominated, or inductive effect dominated, bandwidth both are limited by the wire length and cross area[9].

Numerically, it is $f_{\max} \propto A / L^2$. Therefore, the urge of large bandwidth in long wires forces a large wire cross section. *Svensson*[10] has proposed a graph plotted below which allows evaluation of the required cross section. The graph of Fig. 5 has two regions. The top left region is limited by line RC, whereas the bottom right region is skin effect limited.

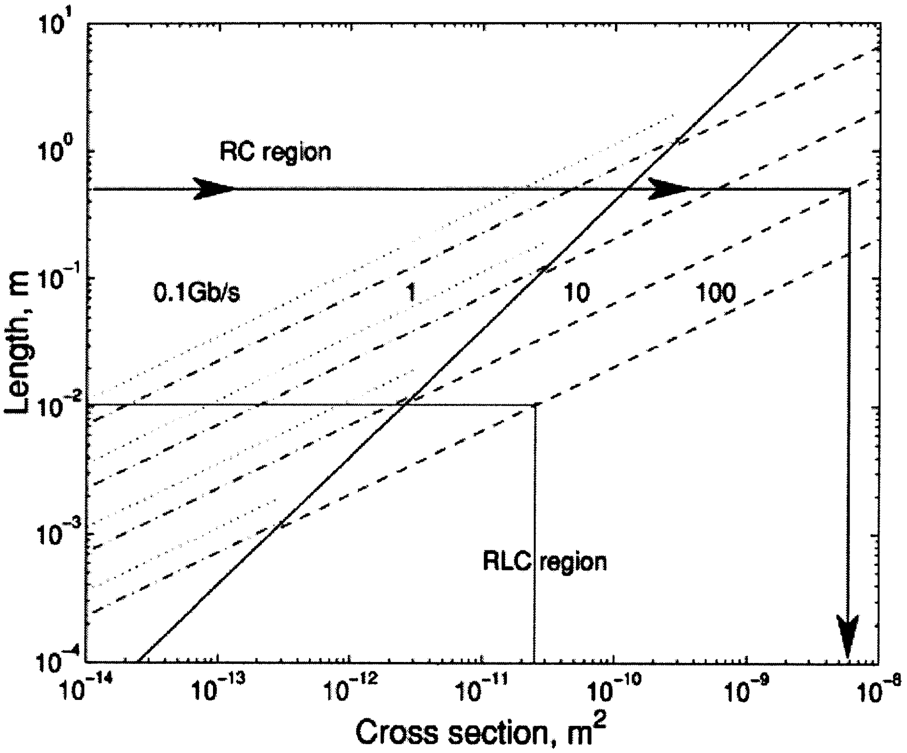


Figure 5: Wire length versus cross – section for different data rates (copper conductor with square cross section assumed)

As you can spot that, when the length is around 1 cm, taking bandwidth 100 GHz, the required minimum cross section is around 26 μm^2.

In optical interconnects, on the other hand, bandwidth limitations are much less important.

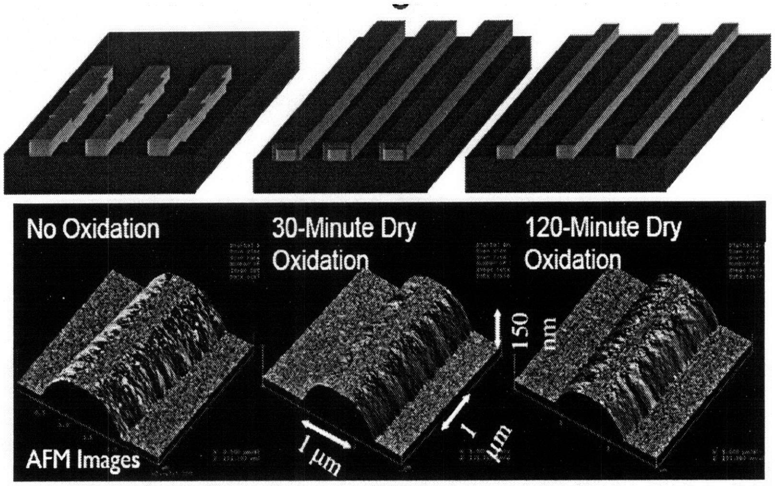


Figure 6: AFM images of silicon waveguides[11]

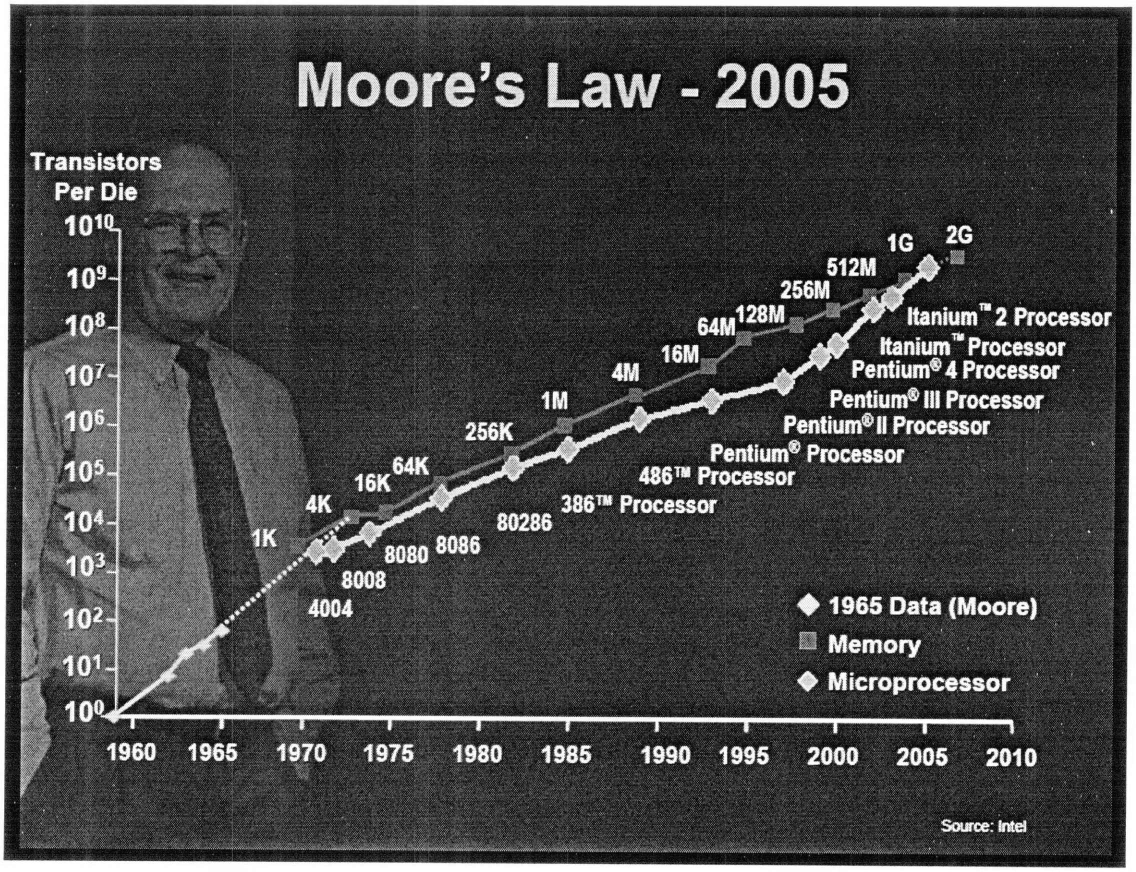
As you can spot of Fig. 6, the area of waveguide is:

$A = 0.5^2 \times \pi / 2 \approx 0.4 \mu m^2$ which is very smaller compared with electronics waveguide requirements.

Chapter Four: Optics in Microprocessor

4.1 Introduction

Moore's law describes a trend in the history of computer hardware: that the number of transistors that can be inexpensively placed on an integrated circuit is increasing exponentially, doubling approximately every two years.



Source: Intel

Figure 7: Moore's Law

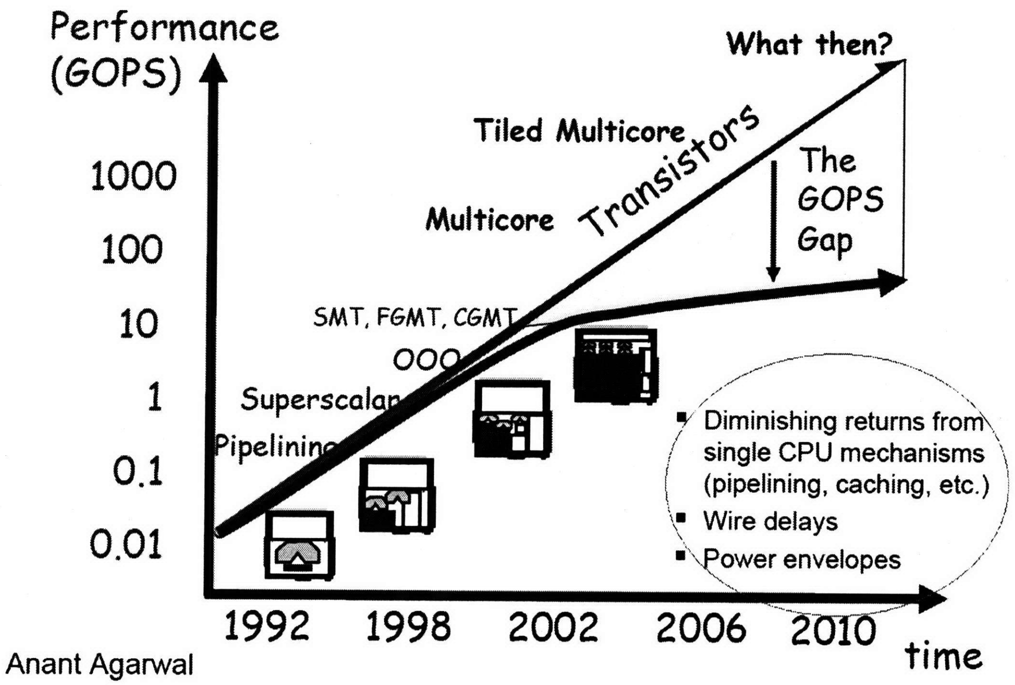


Figure 8: Single core bottleneck

In the past 30 years, as Moore's law predicted, processors kept getting more and more transistors and became more and more complex. The clock frequency kept increasing. After 2002, you simply could not take it any further. All the single processor mechanisms we knew about then were giving diminishing returns. Everything had gotten so big, the pipeline was full and there were no more benefits.

In addition, wire delays became a big issue, in that, the speed of transistors themselves was not the big issue, rather the time it took for a signal to propagate on a wire became the issue. What that meant is, building bigger and bigger structures was not possible anymore. The bigger the structure, the more power it would require and the more latency placed on the wires.

Thus, the power, latency, performance three issues were all interrelated, and there was no way to progress further in a single core design.

4.2 Future Trend: Thousand Core Chips

Integration capacity of billions of transistors exists today, and will double every two years. This trend is shown in Figure 9; starting from 2001 with 130nm technology generation, with a 300mm^2 die capable of integrating one billion transistors.

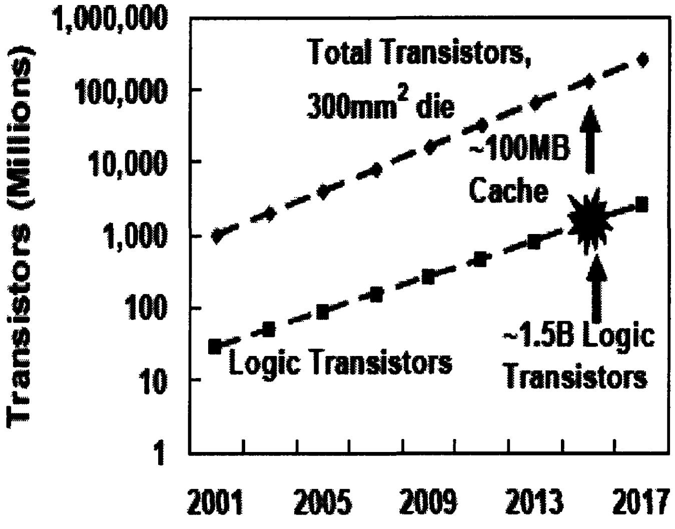


Figure 9: Transistor integration capacity[12]

Assuming about half the die area being allocated for logic, and the other half for large memory arrays such as caches, the trend shows that by 2015 you will have 100B transistors on a 300mm^2 die, with almost 1.5B transistors available for logic. The logic transistors tend to be larger than transistors in the memory, take larger space, and consume more power.[12]

How will you employ these logic transistors to deliver performance? The evolutionary approach is to continue today's trend with a few large processor cores, each employing 20 to 100 million logic transistors, and a large shared cache.

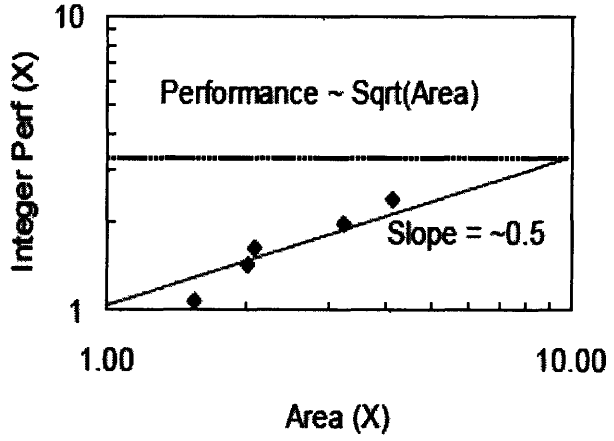


Figure 10: Pollack's Rule

Performance increase by microarchitecture alone is governed by Pollack's Rule, which states that performance increase is roughly proportional to square root of increase in complexity. In other words, if you double the logic in a processor core, then it delivers only 40% more performance—as evidenced by all the leading processors in the past as shown in Figure 10. It plots integer performance increase of new microarchitectures against area (power) increase from the previous generation microarchitecture, in the same process technology.

A multi-core microarchitecture, on the other hand, has potential to provide near linear performance improvement with complexity and power. Two smaller processor cores, instead of a large monolithic processor core, can potentially provide 70-80% more performance, as compared to only 40% from a large monolithic core. Multiprocessors have several other benefits as well: (1) each processor core can be individually turned on or off, thereby saving power; (2) each processor core can be run at its own optimized supply voltage and frequency; (3) easier to load balance among processor cores to distribute heat across the die; and (4) can potentially produce lower die temperatures improving reliability and leakage.

As technology scales further, transistor performance will not increase at the historic rates, due to excessive sub-threshold leakage current, and supply voltage scaling slowing down. Taking these effects into consideration, Figure 8 estimates power consumption of a 300mm^2 processor die.

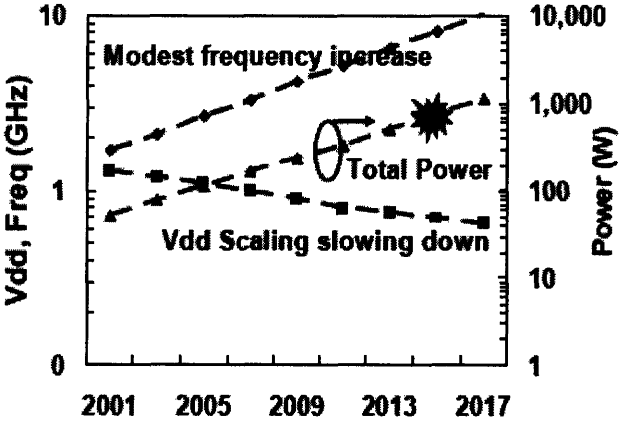


Figure 11: Frequency and Power Consumption [12]

Notice that such a multi-core die will consume almost 1,000 watts of power, which is unreasonable. Therefore, we need to go beyond multi-core, and apply Pollack's rule to the extreme to deliver compute performance in a reasonable power envelope.

Therefore, business as usual is not an option. You cannot simply follow the path of multi-core evolution, integrating multiple complex cores on a die. Instead, we propose that you integrate lots of smaller cores. Each small core delivers lower performance than a large complex core; however, the total compute throughput of the system is much higher as follows.

If there is 1B logic transistor budget, instead of integrating 10 large 100M transistors cores, we propose to integrate 100 medium 10M transistors cores, or even 1000 small 1M transistors cores. Applying the Pollack's rule inversely, after we reduce the complexity of each core, performance of a small core reduces as square – root of the size, but power reduction is linear, resulting in smaller performance degradation with much larger power reduction. Overall, the compute throughput of the system, on the other hand, increases linearly with the larger number of small cores.

A many-core system on a die does not necessarily have to be symmetric or homogenous. An asymmetric system may have a few large cores to deliver higher single-thread performance, but will predominantly have large number of small cores. A heterogeneous system may even integrate diverse special purpose cores for hardware acceleration, e.g. graphics engines.

Figure 12 illustrates such a heterogeneous many-core system with general purpose cores (GP), and special purpose cores (SP), each core having local cache memory, and all cores connected together with an on-die interconnection network.

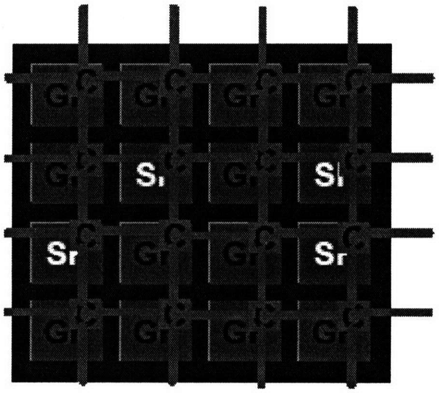
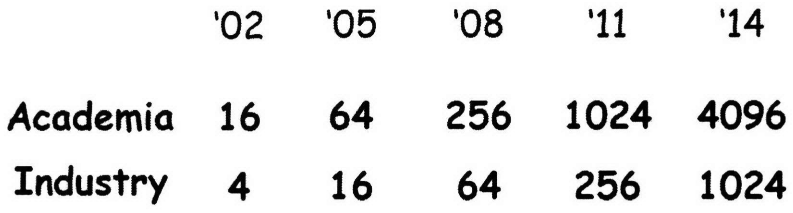


Figure 12: Illustration of a Many Core System[12]

4.3 Multicore bottleneck



	'02	'05	'08	'11	'14
Academia	16	64	256	1024	4096
Industry	4	16	64	256	1024

Figure 13: Roadmap for Multi-cores

Fig. 13 indicates the roadmap for multicore processor, compare with the real products, it is reasonable. Currently, 64 multi-core processor has been commercialized by Tilera Company.

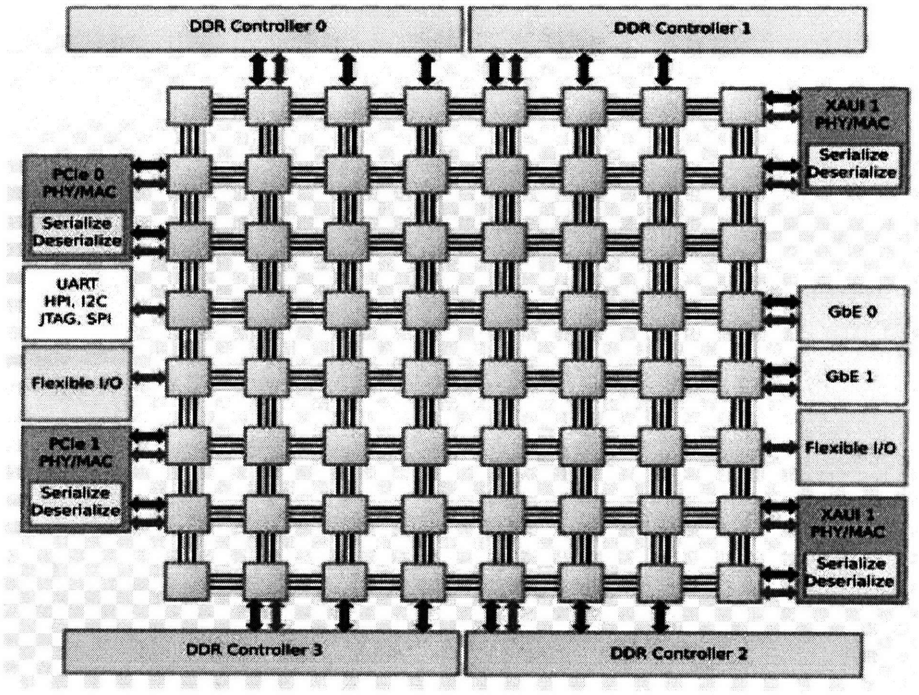


Figure 14: Tile64 from Tilera Company

However, as more and more cores integrated together, there are three big challenges with a multicore design; performance, power efficiency, and programmability. For multicore processors to take off, we had to start with a clean slate. There was nowhere to incrementally improve existing designs. It was a huge upheaval; we had to rethink the architecture, software, and processor design.

Next few sections, some basic knowledge will be introduced which sets the basic knowledge for suggested architecture.

4.4 Photonic Technology

Although Moore's Law enables larger numbers of computational elements to be placed on a single chip, the extent to which they can be used to improve performance is limited by the cost of communication between those elements. As core count grows into the hundreds, the main memory bandwidth required to support concurrent computation on all cores will increase by orders of magnitude. This creates a significant bandwidth bottleneck. Evidence suggests that many – core systems using electrical interconnects may not be able to meet these high bandwidth demands while maintaining acceptable performance, power, and area.

Nanophotonics offers an opportunity to reduce the power and area while meeting future system bandwidth demands. This section we will generally introduce a complete nanophotonic network.

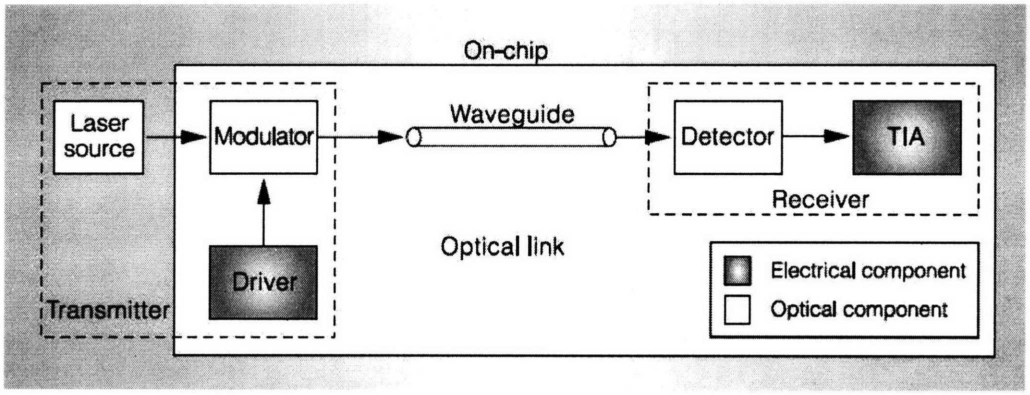


Figure 15: Simplified diagram showing the main components of on – chip optical transmission[13]

Transmitter

Optical transmission requires a laser source, a modulator, and a modulator driver circuit. The laser course provides light to the modulator, which transduces electrical information (supplied by the modulator driver), into a modulated optical signal.

Waveguide

Waveguides are the paths through which light is routed. For on – chip applications, silicon and polymer are the most promising waveguide material.

Receiver

An optical receiver performs the optical to electrical conversion of the light signal. It consists of a Photodetector and a transimpedance amplifier (TIA) stage.

The TIA stage converts Photodetector current to a voltage that subsequent stages threshold to digital levels.

4.5 Current multicore chips topology

4.5.1 Full Mesh Interconnect

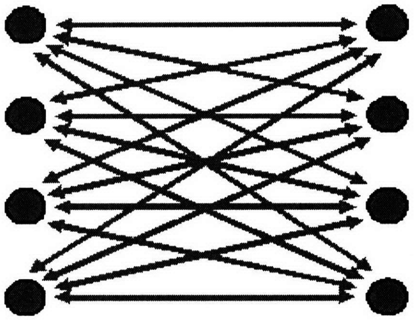


Figure 16: Point to Point

In this simplest of logical structures, every source has a point – to – point interconnect link with every destination with which it may have to communicate.

4.5.2 Shared bus

For most of the early Symmetric Multi – Processor (SMP) server designs, it was common to interconnect multiple processors, memory, and I/O ports using a single shared multidropped bus. This is a "many – to – many" configuration, with multiple senders and multiple receivers on each common electrical line.

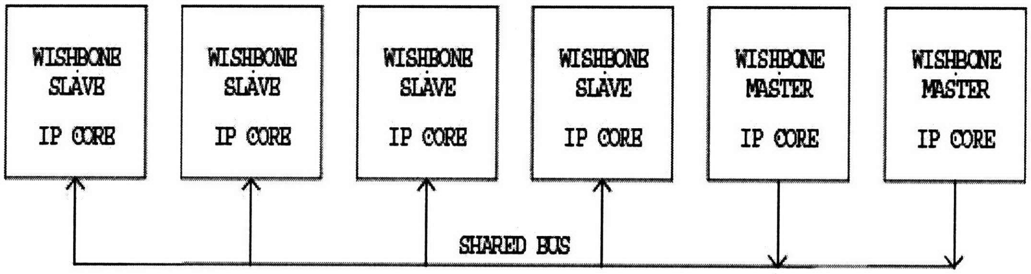


Figure 17: Shared bus[14]

An arbitration mechanism is used to select one sender to transmit on the bus at a time. This approach has the advantages of simplicity of design and ease of system expansion. However, the electrical characteristics of a multidropped bus limit its useful frequency to the 200 – 400 MHz range, and limit its length to 10 – 20 cm. In addition, as the number of cores on a bus increases, the length of the bus wires increase, forcing the use of a slower clock. Since all cores share the same bus, contention increases very quickly.

4.5.3 Pipeline

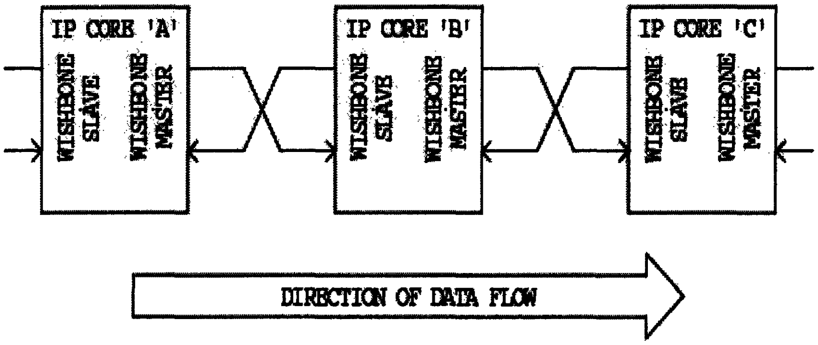


Figure 18: Pipeline[14]

The Raw microprocessor [7] consists of a 2 – D mesh of cores where each core can communicate directly with his neighbors. In this case, it avoids long global wires but communication between distant cores requires multiple hops. This design allows for much less contention as long as communication patterns do not physically overlap in the mesh.

4.5.4 Switched

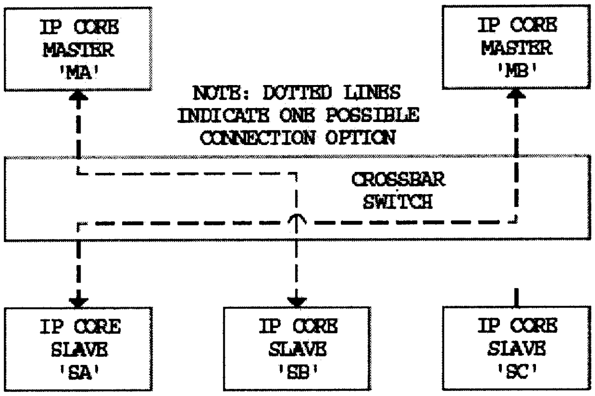


Figure 19: Cross bar switch[14]

The most generally useful networks currently are multistage packet – switched networks, in which modules are connected with point – to – point links, and switching elements within the modules route data and control packages to their proper destination on the basis of explicit routing headers or address – based routing.

Chapter Five: Suggested Architecture Overview

Current multicore architectures discussed above will not allow performance to scale with Moore's law. For applications with irregular, broadcast communication patterns, contention is unavoidable and may become unacceptable as processors are scaled to thousands of cores. Meanwhile, the scalability of today's multicore architecture is also threatened by the challenge of programming them. It is necessary to optimize the balance between computation and communication of multicore programmers. Further, the techniques that are currently used to program multicores will not scale to thousands of cores. Neither bus – based nor point – to – point multiprocessors are able to scale to thousands of cores since contention would overwhelm the network and coordinating thousands of processors is very difficult.

Two architectures are discussed below in order to solve the issue of contention and programming difficulty.

5.1 ATAC Architecture [15]

Fundamentally, the ATAC processor architecture addresses contention and programmability issues using on – chip optical communications technologies to replace or augment electrical communication channels.

ATAC leverages these advances to eliminate communication contention using Wavelength Division Multiplexing (WDM). In addition, optical waveguide can also transmit data at higher speeds than electrical wires because photons propagate faster in a waveguide than electrons do in a wire. Optical signaling can use less power than electrical signaling, especially for long wires due to low loss of optical waveguides and absence of periodic repeaters the way long electrical wires do.

In summary, an ATAC processor will provide programming transparency, power efficiency, high bandwidth, and performance scalability.

5.1.2 Architecture Overview

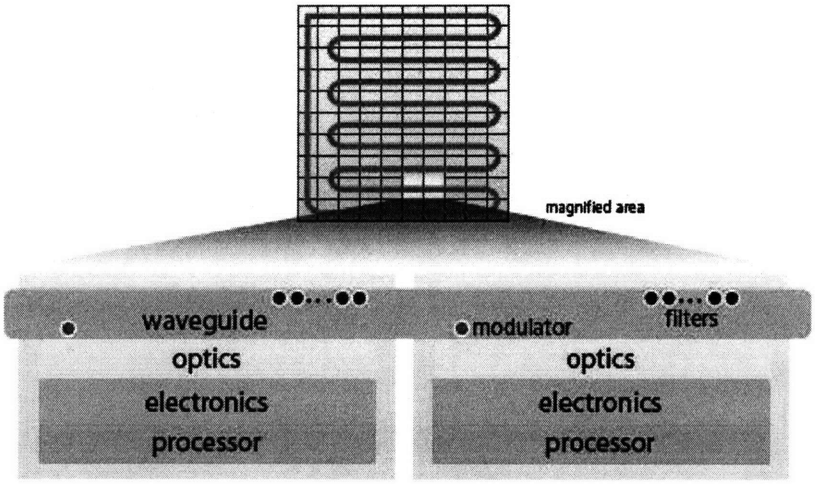


Figure 20: ATAC chip showing two adjacent cores[15]

In Fig. 20, ATAC consists of a 2 – D mesh of cores, each containing a processor pipeline and electrical and optical network resource.

ATAC cores are interconnected via an optical waveguide ring bus that is similar to a fully – connected, bi – directional point – to – point network. The waveguide bus passes through every core on the chip and incorporates WDM to overlap simultaneous communications without contention. To contrast, the use of WDM is not feasible with traditional electrical interconnect, and the wire delay and power issue limit the scalability of electrical buses, forcing alternatives with distance – dependent communication latencies. In addition, for the programmability issue, ATAC's architecture simplifies programming considerations because programmers need only specify the recipient of messages without having to deal with the complicated routing, non – uniform latencies, and bottlenecks inherent to electrical interconnect.

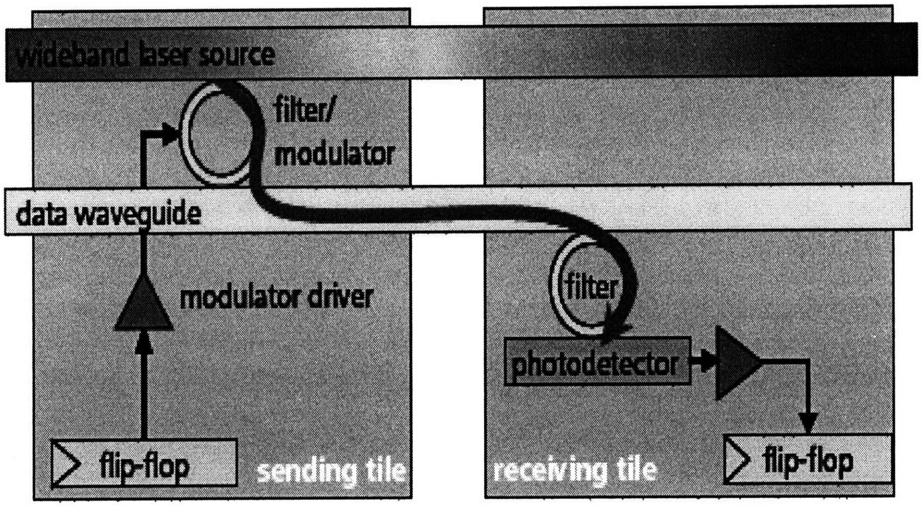


Figure 21: Optical transmission of one bit between two cores

Fig. 21 elaborates the process of sending one bit from one core to another. A modulator is an optical component that writes 1s and 0s into the network. Each core has one modulator per bit statically tuned to a frequency that is dedicated to that core. The modulator is turned on and off by a driver that is controlled by the processor electronics. The optical signal is transmitted over a Si waveguide. The signals are received using optical filter detectors, which are each statically tuned to receive a particular wavelength. For an N – core ATAC chip, each core contains N × M filters, where M is the bit – width (dictate how many information can flow in each clock cycle) of the network, which enable each core to receive from any other core. The filter channels the light onto a Photodetector which converts the incoming optical signal into an electrical signal that is buffered and stored in a flip – flop. The data is fed into a FIFO (refer to a way data stored in a queue is processed) and then into a CAM (Content – addressable memory), at last receiving core's processor can read the value from CAM.

5.2 The Corona Architecture [16]

Corona is tightly coupled, highly parallel NUMA system. As NUMA (Non – Uniform Memory Access) systems and applications scale, it becomes more difficult for the programmer to manage the placement and migration of programs and data. It is suggested to use homogeneous cores and caches, a crossbar interconnect that has near – uniform latency, a fair interconnect arbitration protocol, and high bandwidth between cores and from caches to memory to lessen the burden.

5.2.1 Cluster Architecture

Corona is a nanophotonically connected 3D many – core NUMA system that meets the future bandwidth demands of data – intensive applications at acceptable power levels. Corona is targeted for a 16 nm process in 2017. Corona comprises 256 general purpose cores, organized in 64 four core clusters, and is interconnected by an all – optical, high – bandwidth DWDM crossbar.

Inside Cluster, each core has a private L1 instruction and data cache, and all four cores share a unified L2 cache. A hub routes message traffic between the L1, directory, memory controller, network interface, optical bus, and optical crossbar.

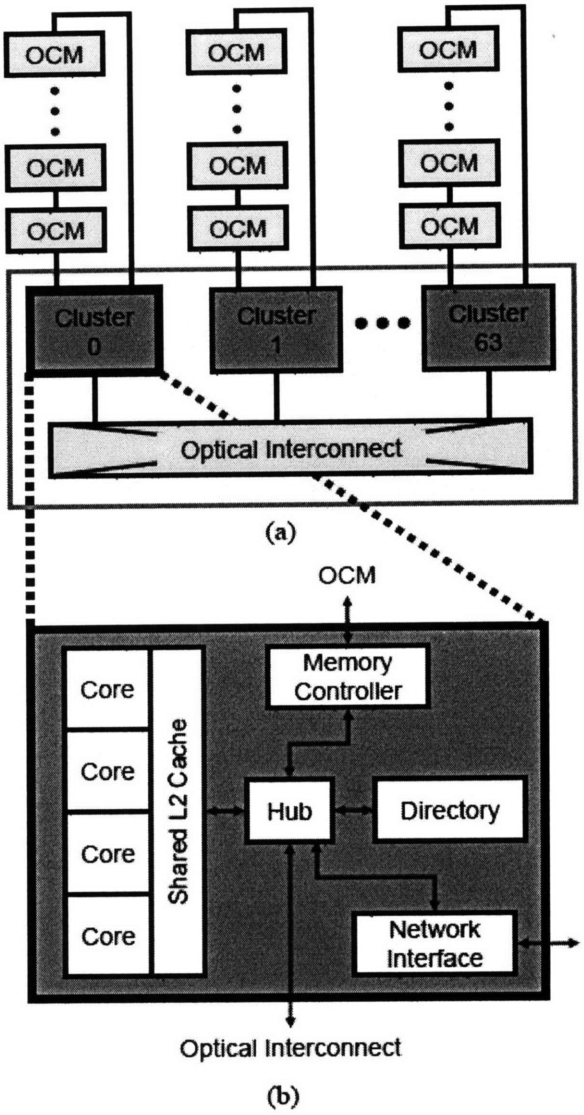


Figure 22: Architecture Overview[16]

5.2.2 Optical Crossbar

Each cluster has a designated channel that address, data, and accord to messages share.

Any cluster may write to a given channel, but only a single fixed cluster may read from the channel. A fully – connected 64 × 64 crossbar can be realized by replicating this mechanism 64 times.

The channels are 256 wavelengths (correspond to 256 cores), or 4 bundled waveguides, wide. When laid up, the waveguide bundle forms a broken ring that originates at the destination cluster (home cluster), is routed past every other cluster, and eventually terminates back at its origin. Light is sourced at a channel's home by a splitter that provides all wavelengths of light from a power waveguide. Communication is unidirectional, in cyclically (crossbar) increasing order of cluster number.

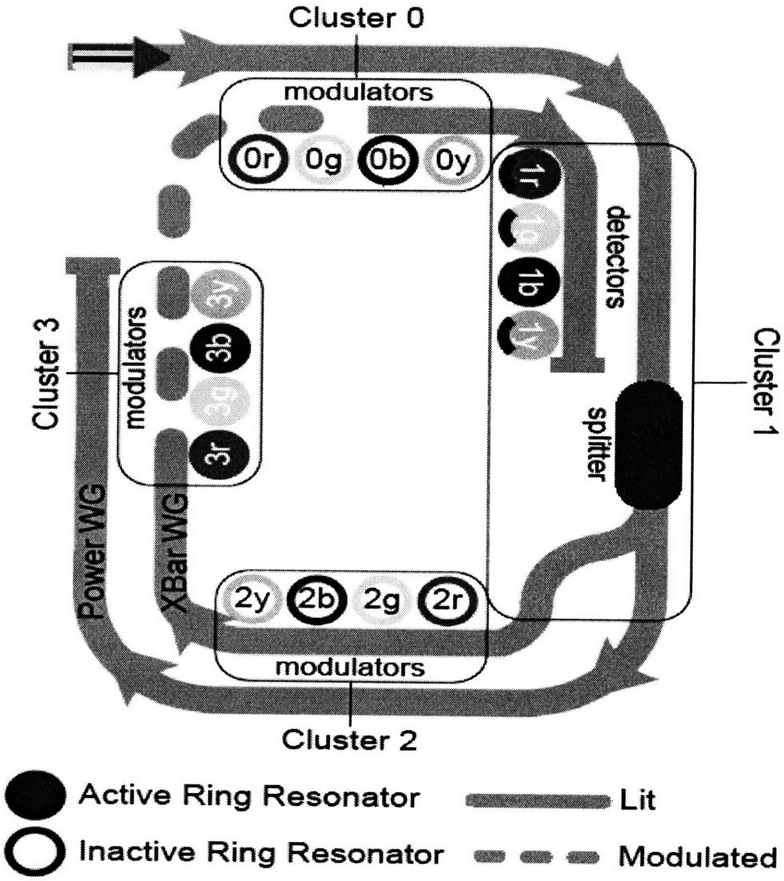


Figure 23: A Four Wavelength Data Channel Example[16]

A cluster sends to another cluster by modulating the light on the destination cluster's channel.

Fig. 23 illustrates the conceptual operation of a four – wavelength channel: A home cluster (cluster 1) sources all wavelengths of light (r, g, b, y). The light travels clockwise around the crossbar waveguides. It passes untouched by cluster 2's inactive (off – resonance) modulators. As it passes by cluster 3's active modulators, all wavelengths are modulated to encode data. Eventually, cluster 1's detectors sense the modulation, at which point the waveguide terminate.

5.3 Optics vs Electronics

Both two processor architecture addresses these contention and programmability issues using on – chip communications technologies to replace electrical communication channels. Current research in optical communications technologies is making strides at integrating optoelectronic components with standard CMOS fabrication processes. ATAC and Corona leverages these advances to eliminate communication contention using Wavelength Division Multiplexing (WDM). WDM allows a single optical waveguide to simultaneously carry multiple independent signals on different wavelengths. In addition, optical waveguides, however, can also transmit data at higher speeds than electrical wires because photons propagate faster in a waveguide than electrons do in a wire. This virtually eliminates the heterogeneous distance – based cost function for communication that complicates multicore programmability. In addition to speed, optical signaling can use less power than electrical signaling, especially for long wires. This is because optical waveguides have very low loss and do not require periodic repeaters the way long electrical wires do.

In conclusion, using any of these two architectures, it will provide programming transparency, power efficiency, high bandwidth, and performance scalability.

Chapter Six: Cost Analysis

A designer of a communication system should always understand what the physical fundamental limit of the communication system that is being designed is and try to approach that limit as closely as possible. If the design is far away from fundamental limits, there is a chance of significant pay – offs on looking how to improve system performance.

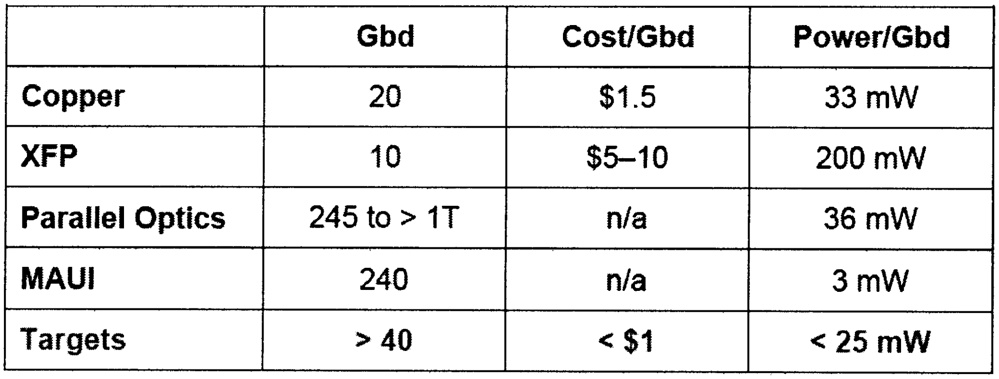
Nowadays, electronic communication links are impeded by fundamental physical loss mechanisms which we discussed in chapter 3. These include dielectric losses and skin effect losses which both are a function of distance and bandwidth. Shannon's law predicts that for any given signal strength there will be maximum communication rate determined by losses and noise level. As industries move to ever higher bandwidths, they have already approached Shannon's limitations. At this time, it is possible that bandwidth continue increasing by utilizing parallel channels, changing to lower loss interconnect material system (reduce the noise power over bandwidth, extend the Shannon's limit), or using repeater. However, each of these solutions will make the whole system becoming much more expensive. In this situation, photonics is generally chosen in many industries in order to ease the cost pressure and also extend the high performance request.

6.1 Optical Cost Issue

Although integrated optical components had their beginnings over thirty years ago, they have not evolved with the same success, neither in complexity nor in functionality as integrated electronics. Most of reasons for the technological success of integrated electronics can be attributed to the advancements in silicon processing.

Factors that effect cost of integrated optics are the lack of convergence to one material and the absence of device standards, as in the case of volume manufacturing of electronics. The successful development of additional optical components in silicon technology, the standardization of market specifications, and a focus on a common material platform may help alleviate some of the costs associated with integrated optics. Monolithic integration will need to manage production costs, which rise as the complexity and number of processing operations increase. [17]

6.2 Target bandwidth – cost



	Gbd	Cost/Gbd	Power/Gbd
Copper	20	$1.5	33 mW
XFP	10	$5–10	200 mW
Parallel Optics	245 to > 1T	n/a	36 mW
MAUI	240	n/a	3 mW
Targets	**> 40**	**< $1**	**< 25 mW**

Figure 24: Key metrics for silicon microphotonic solution to drivers

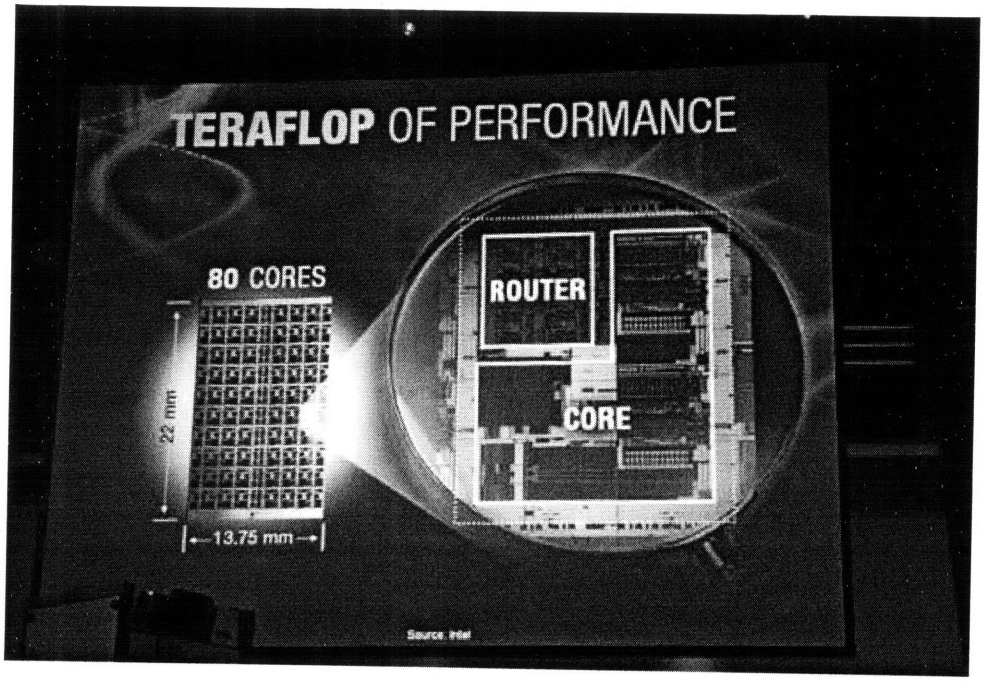
At present, the bandwidth – cost target is $1 per GHz for approximately 10 – meter distances; however, this cost structure is not apparent in today's start of the art commercial optical links such as XFP. But as electronic – photonic integration is considered for broader applications, this cost target could drop to $0.50 per GHz. [18]

Chapter Seven: Multi-Core Market Analysis

In this section, we will mainly consider multi-core consumer PC market.

7.1 Multi-core

There has been the paradigm change during the last 18 month to move from higher frequency single processors to multiple lower frequency processors. The chip production lines are setup to produce CPUs with many cores on the chip. Prototype of 80 cores on a chip has already been presented by Intel:



Source: Intel

Figure 25: Prototype of 80 cores multi-core

But it is not clear what the market directions really are and the paradigm change in hardware requires also a paradigm change in software. There are currently several independent developments ongoing in industry to exploit the different possibilities, which also show the uncertainties in this area.

A special case is the trend to combine much closer the CPU and the graphics processor (GPU). On the one side we have products like the FireStream 'coprocessor' from AMD, a product developed during the 2006 after AMD bought the graphics specialist ATI.

And we have the market leader in graphics processors, NVIDIA, who launched lately their Tesla product, a standalone 'supercomputer' based on several linked graphics cards. Intel on the other side is trying to establish a model, where one of the cores is doing the graphics work. They have bought in September 2007 the company HAVOC which produces the widely used physics engine for PC games which would run specially optimized on another core on a multi-core chip. Intel is also pushing a 'new' paradigm in the graphics processing model. Today the rendering principles in games are based on so called shader units, which require the special graphic cards from ATI and NVIDIA. Ray–tracing is a much better approach to produce realistic environment, but his requires much more computing capacity, which multi-core chip could offer in the future.

7.2 Intel vs AMD

While in 2005 and 2006 the processors from AMD had clear advantages over the corresponding processor generations from INTEL, the picture has changed completely in 2007. After internal restructuring and a change in their strategies, Intel was able to push their new multi-core processor generations with much success into the market. AMD was able to keep their market shares in the desktop, server and mobile area only because of a very aggressive pricing strategy.

Till now, market shares of AMD is 23.5% overall, 76.3% for Intel, as you can spot that the total market share for these two company reaches 99.8%.

This strategy caused heavy losses for AMD during the last 4 quarters (more than 2 billion $) and AMD is from the technology point of view still about 12 month behind Intel. In addition they are late in introducing their newest processor lines (Barcelona, Phenom).

These are all reflected in the revenue and profit for Intel and AMD:

	Q1/05	Q2/05	Q3/05	Q4/05	Q1/06	Q2/06	Q3/06	Q4/06	Q1/07	Q2/07	Q3/07
INTEL Revenue	9.4	9.2	10.0	10.2	8.9	8.0	8.7	9.7	8.9	8.7	10.1
INTEL Profit	2.2	2.0	2.0	2.5	1.4	0.9	1.3	1.5	1.6	1.3	1.9
AMD Revenue	1.2	1.3	1.5	1.8	1.3	1.2	1.3	1.8	1.2	1.4	1.6
AMD Profit	-0.02	0.01	0.08	0.96	0.19	0.09	0.14	-0.57	-0.61	-0.60	-0.40

Figure 26: Revenue and profits for AMD and Intel

AMD has to scale down and becomes less of a competitor for Intel with the corresponding consequences for the pricing strategies.

7.3 Start-up Company:

In consumer PC market; there are several hurdles for new company to enter.

First of all, consumer PC manufacturer is acted as the buyer for Intel, AMD, Tlera or other companies. Through several collaborations, buyers have built Brand Identity, in order to avoid risk; they will prefer the Intel and AMD instead of new company. However, if the performance of the start-up company microprocessor is much superior to historic company such as Intel and AMD, gradually, the monopoly situation (if Intel and AMD is a whole entity, it is acted as monopoly to start-up company) is going to be changed.

Secondly, because of high fixed cost for Semiconductor Company, Start-up Company would only choose fabrication outsourcing, obviously, from the technology point of view, it is behind the Intel and AMD, which means that in single cores, it is difficult to put same amount of the transistor into cores, and the performance is not as good as them.

Last but not least, Intel and AMD can be considered monopoly, and they will set the general price, and Start-up Company acts as price taker. As a result, it contains huge risk for Start-up Company when the price set by Intel and AMD variate.

Solution:

Instead of focusing on mainly computer PC market, Start-up Company can put its energy to the market where many-core processors have a significant impact.

For example, the largest multi-core chip that is actually shipping today is Tilera's 64-core TILE64^{TM} family of CMP unveiled last year. Tilera is not going to focus on the Consumer PC market, instead, it will attack the intelligent networking and Video Server markets via embedded processors. In both of these markets, they have the words "highest performance".

Chapter Eight: Conclusion

The appealing feature of optical interconnect is the very large bandwidth potentiality. In order to take the optical leap, however, the ability of efficient handling of optical signal at low cast is required.

At present, such an objective seems quite ambitious for Si technology. Optical functionality, however, is sneaking in, the situation is reminiscent of early modern Si electronic industry, when the devices were first born as bulky stand – alone elements. Integration started up then, and within few years Moore's law was stated.

Over the coming decade, memory and inter – core bandwidths must scale by orders of magnitude to support the expected growth in core performance resulting from increased transistor counts and device performance. It is believe that nanophotonics can be crucial in providing required bandwidths at acceptable power levels.

To investigate the potential benefits of nanophotonics on computer systems, two architectures are introduced in order to satisfy the future bandwidth requirements at acceptable power levels.

References:

1. Savage, N. Linking with light [high-speed optical interconnects]. *Spectrum, IEEE* **39**, 32-36(2002).
2. Cangellaris, A. The interconnect bottleneck in multi-GHz processors; new opportunities for hybrid electrical/optical solutions. *Massively Parallel Processing, 1998. Proceedings. Fifth International Conference on* 96-103(1998).doi:10.1109/MPPOI.1998.682132
3. Ho, R., Mai, K. & Horowitz, M. The future of wires. *Proceedings of the IEEE* **89**, 490-504(2001).
4. Haurylau, M. et al. On-Chip Optical Interconnect Roadmap: Challenges and Critical Directions. *Selected Topics in Quantum Electronics, IEEE Journal of* **12**, 1699-1705(2006).
5. Miller, D. Rationale and challenges for optical interconnects to electronic chips. *Proceedings of the IEEE* **88**, 728-749(2000).
6. *INTECONNECT*. (International Technology Roadmap for semiconductor.:).at <http://www.itrs.net/>
7. Gaburro, Z. Optical Interconnect. *Silicon Photonics* 1999(2004).at <http://www.springerlink.com/content/gr59t1m5qnlt8j9j>
8. Meindl, J. Interconnect opportunities for gigascale integration. *Micro, IEEE* **23**, 28-35(2003).
9. Davis, J. et al. Interconnect limits on gigascale integration (GSI) in the 21st century. *Proceedings of the IEEE* **89**, 305-324(2001).
10. Svensson, C. Electrical interconnects revitalized. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* **10**, 777-788(2002).
11. Sparacin, D.K., Spector, S.J. & Kimerling, L.C. Silicon Waveguide Sidewall Smoothing by Wet Chemical Oxidation. *J. Lightwave Technol.* **23**, 2455(2005).
12. Borkar, S. Thousand Core ChipsA Technology Perspective. *Design Automation Conference, 2007. DAC '07. 44th ACM/IEEE* 746-749(2007).
13. Nevin Kırman On-Chip Optical Technology in Future Bus-Based Multicore Designs. (2007).at <http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/mags/mi/&toc=comp/mags/mi/2007/01/m1toc.xml&DOI=10.1109/MM.2007.18>
14. *Wishbone*. (Wikipedia:).at <http://en.wikipedia.org/wiki/Wishbone_%28computer_bus%29>
15. MichaelWatts, J.P.J.E.J.M.^.K. *ATAC: On-Chip Optical Networks forMulticore Processors*. (Massachusetts Institute of Technology:).
16. Dana Vantrease Corona: System Implications of Emerging Nanophotonic Technology. (2008).at <http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath=/dl/proceedings/&toc=comp/proceedings/isca/2008/3174/00/3174toc.xml&DOI=10.1109/ISCA.2008.35>
17. Paniccia, M., Morse, M. & Salib, M. Integrated Photonics. *Silicon Photonics* 1999(2004).at <http://www.springerlink.com/content/05dmjbj8n0cj9m2y>
18. Kirchain, R. & Kimerling, L. A roadmap for nanophotonics. *Nature Photonics, Volume 1, Issue 6, pp. 303-305 (2007).* **1**, 303-305(2007).