

Exploiting Capacitance in High-Performance Computer Systems

Alex Chow, David Hopkins, Robert Drost, Ron Ho
Sun Microsystems Laboratories
Menlo Park, CA 94025
USA

ABSTRACT

Aggressive scaling of transistor feature sizes has enabled unprecedented levels of integration and corresponding performance improvements in VLSI systems. However, fabrication costs present barriers to continued growth in transistor density. Proximity Communication breaks these barriers by providing high-density, high-bandwidth, low-power, and scalable off-chip I/O, allowing designers to partition their designs into separate chips with significantly reduced performance penalties. This partitioning greatly improves chip and package yield, and enables modular assemblies of heterogeneous systems with customizable mixes of functional units tailored for specific end-user applications.

INTRODUCTION

By any measure, microprocessor performance has dramatically increased over the past forty years. Normalizing the performance of computer systems to the SPEC 2000 integer benchmark shows a cumulative annual growth rate over that period of 35%, or computer performance doubling roughly every two years [1]. This growth curve is virtually unmatched by that of any other industry.

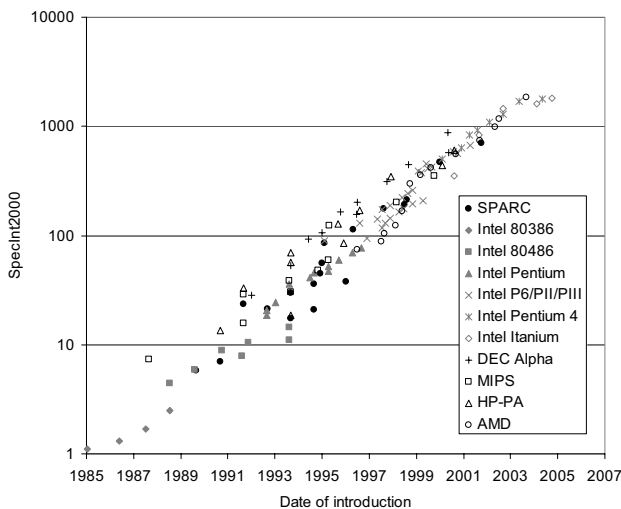


Figure 1. SpecInt2000 performance of microprocessors.

These and similar performance trends are often misattributed to Moore's Law, or what Carver Mead dubbed Gordon Moore's now-famous prediction of transistor density scaling. In a landmark paper, Moore, who was later to found Intel, predicted that the most cost-efficient number of integrated components per die will double every twelve months [2]. A decade later, Moore revised his prediction's doubling time to every two years [3], and in later years the doubling time was closer to every eighteen months. However, the idea was clear: when accounting for semiconductor yield, the number of integrated components that minimize a system's cost will grow geometrically over time.

While Moore's Law has acted as an important self-fulfilling prophecy for semiconductor manufacturing, Moore's publications made no predictions about computer performance. However, as

shown above, computer system performance **has** followed transistor density scaling. As a result, when papers, pundits, and panels ask when Moore's Law will end, they often mean to ask instead, "when will the historically geometric improvement in microprocessor performance end?"

In this paper, we consider this question from an economic die-area perspective, and describe a technology that exploits capacitance in order to dramatically improve chip-to-chip interconnections. Our coupled data technique offers a chance to extend computer system performance beyond the die size limits imposed by the economics of Moore's Law.

SYSTEM PERFORMANCE AND MOORE'S LAW

From a hardware perspective, the time a computer takes to execute an instruction is simply the product of the clock period and the number of clock cycles per instruction (CPI). While processor clock periods have historically fallen with process improvements, further reductions will be constrained by power limitations, clock distribution costs, and transistor variability.

Improving performance therefore increasingly calls for reducing CPI, largely through architectural improvements. Super-scalar, speculative, multi-threaded, or multi-processor architectures all reduce CPI by exploiting parallelism at instruction, thread, or program levels. Such techniques, however, require additional chip area for more complex logic, larger memories, and multiple cores. System architects therefore turn transistors into performance, and hence need increased transistors per die to continue to provide increased performance.

Merely assembling multiple dice together does not easily improve performance, because the multiple dice must still communicate. As technologies scale, on-chip features shrink much faster than off-chip features. Today, on-chip wires have pitches under 1 μm , while off-chip wires typically have pitches in excess of 100 μm . This "impedance mismatch" between on-chip and off-chip wires gives rise to a bandwidth bottleneck when sending data between chips. Serializer-deserializer (SerDes) circuits overclock off-chip wires to approach on-chip wire bandwidth, but at a high area and power cost, and they still fail to match on-chip bandwidth per area. As a result, increasing transistor count by simply bonding together multiple discrete chips is not efficient.

Continued transistor scaling appears fundamentally possible. Novel transistor structures, such as tri-gate or FinFET devices, help circumvent physical constraints such as transistor leakage. Innovative system architectures and circuit techniques help keep the anti-scaling trends of on-chip wires manageable [4]. New lithographic technologies, such as immersion bars, electron projection, and extreme UV imaging overcome the diffraction limits of traditional optical processes.

Unfortunately, the cost of transistor scaling also keeps growing. Million-dollar mask costs in modern technologies are prohibitive for low-volume and prototyping efforts. Design effort and time to market grow with transistor count; the complexity of billion-transistor chips requires almost an unmanageable level of engineering effort, reducing performance and challenging the feasibility of physical design [5]. Device reliability and yield degradation of large dice further exacerbate exponential trends in

manufacturing costs. Combined, these economic constraints will limit continued advances in transistor scaling [6].

That Moore’s Law, fundamentally an economic argument, may ultimately be limited by spiraling financial costs, is ironic. But it also invites us to break through these constraints by revisiting Moore’s original analysis of the most cost-effective number of transistors on a single die. We do this by redefining what we mean by a “single die.”

PROXIMITY COMMUNICATION

Proximity Communication extends the cost effectiveness of transistors to multiple chips by providing low-cost chip-to-chip communication: it delivers off-chip I/O at a power, bandwidth, and latency much closer to on-chip wires than SerDes links. With Proximity Communication, chips exchange data without wires [7]. They are placed face-to-face and in close proximity, with a gap of under 10 μm between their top surfaces. The chips communicate through capacitive coupling between top layer metal plates; the transmitter on one chip drives a metal plate that couples to a corresponding metal plate on the receiving chip (Figure 2). On-chip position sensors [8] and electronic alignment circuits [9] keep the plates of one chip well aligned with the plates of the other to ensure proper communication and minimize crosstalk.

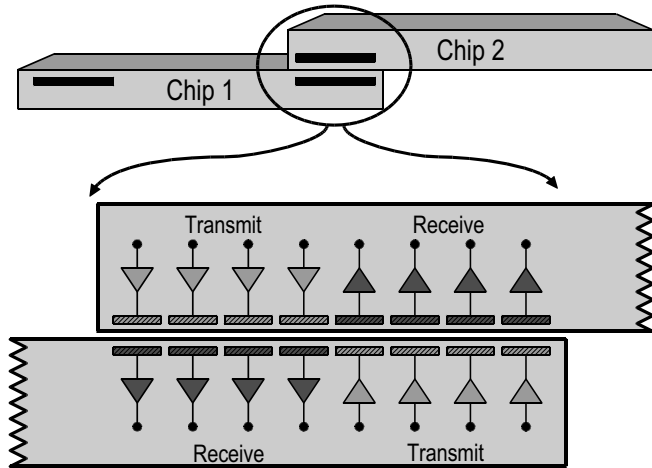


Figure 2. Two face-to-face chips.

By replacing off-chip wires with on-chip metal plates, Proximity Communication provides 60 \times greater channel density over conventional solder bumps, and this improvement increases under technology scaling. Proximity Communication also lowers I/O power requirements, not only because the on-chip transmitting and receiving structures are small and present low parasitic capacitance, but also because they are passivated and do not need ESD protection. In addition, the absence of long wires obviates the need for channel equalization, and the large number of available channels averts the use of SerDes circuits, further reducing power and latency.

Recent results in 0.18 μm CMOS showed reliable communication over 144 channels, with an aggregate bandwidth density of 430 Gb/s/mm² and a BER < 10⁻¹⁵ [10]. Measured energy consumption per channel is 3 pJ/bit. In terms of bandwidth, power, and reliability, this level of performance matches that of on-chip wires.

As a scalable, low-cost, high-bandwidth I/O technology, Proximity Communication provides two major benefits to designers. First, it allows a collection of small chips to function as one large chip without any degradation in performance. This enables high-performance systems without the yield-fallout cost of large dice. Second, Proximity Communication uses neither wires nor soldered connections. This absence of permanent attachment enables the

removal and replacement of individual chips. This solves the known-good-die problem for multi-chip packages, and enables the application-specific assembly of modular systems that leverage functional or physical heterogeneity.

LOWERING INTEGRATION COSTS

To improve performance, designers are integrating an increasing number of functional units on a single chip. This has led to system-on-chip (SoC) designs, which integrate onto a single die many different functional units, including processing cores, memory, data conversion circuits, oscillators, and radio-frequency interfaces. However, more functional units need more die area, and large die sizes prohibit building a cost-effective SoC. A simple formulation [11] of overall chip yield Y , given chip area A , is

$$Y = \int_0^{\infty} e^{-DA} f(D) dD \quad [1]$$

where D is the spot defect density, and $f(D)$ its probability distribution. This well-known yield formula gives an exponential increase in die cost with area. More refined models [12,13] provide better agreement with modern fabrication processes, but with similar overall trends.

The high yield fallout of large dice has motivated multi-chip modules (MCMs) and systems-in-package (SiP). However, these technologies present two main limitations. First, they are not well-matched to high-bandwidth applications: the chips still have an off-chip bandwidth bottleneck that requires the high area and power overhead of SerDes interfaces. For example, even with processor MCM modules, first- and second-level cache memories are still integrated into processing cores and not across chips in the MCM. Second, these technologies suffer from the known-good-die problem, as most components cannot be fully tested until they are packaged into a system. Therefore, unless the MCM or SiP can be reworked, the final package yield can still be low, especially for systems with large numbers of components.

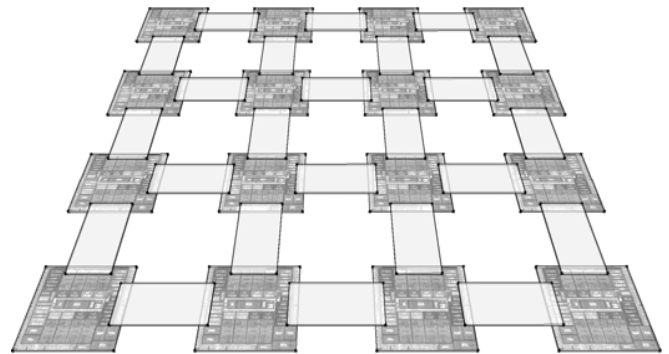


Figure 3. A Proximity Communication multi-chip system.

Proximity Communication lowers these costs of integration. By eliminating off-chip I/O bottlenecks, it enables high-performance systems using smaller, high-yield chips. By eliminating permanent soldered connections, it solves the known-good-die problem present in MCM and SiP approaches. Packages become reworkable because defective chips can be easily replaced or added. This improves overall package yield and lowers production costs. While many previous studies [14 - 17] have presented yield analyses of MCM and SiP systems, a similar analysis can extend those models to packages of chips using Proximity Communication. Figure 4 shows the total cost of a packaged system as a function of chip area. We compare three implementations: a single chip module (SCM), a 4-chip multi-chip module (MCM), and a 4-chip Proximity Communication module (PxC-MCM). To compute chip yield, we approximate

Equation 1 above with the simplified yield formula which uses a commonly-observed defect distribution [12],

$$Y = \frac{1}{(1 + AD/\alpha)^\alpha} \quad [2]$$

We assume $\alpha = 4$ and a defect density of $D = 0.014$ per cm^2 [18]. We add a cost of \$10,000 per wafer, and a chip area overhead of 15% for MCM and PxC-MCM. We also include a package cost of 40% of the chip cost for the SCM, and 80% of the chip cost for MCM [17] and PxC-MCM. The high channel density of Proximity Communication should lower PxC-MCM's chip area overhead, but for simplicity we assume the same area overhead as MCMs. Also, we expect the cost of PxC-MCM packages to eventually approach that of SCM packages.

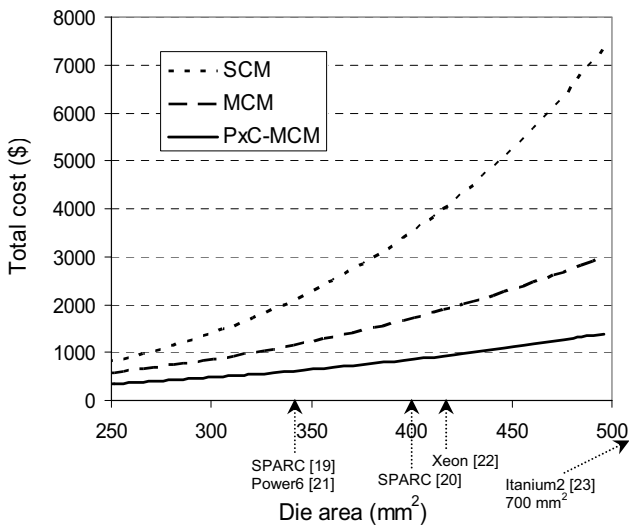


Figure 4. Total system cost as a function of die area. CPUs sizes [19 - 23] are shown for illustrative purposes only and do not indicate actual cost.

For the traditional MCM, we account for known-good-die by applying a wafer test yield—the percentage of unpackaged, defective chips that can be identified before assembly—of 75%. We assume that the MCM cannot be reworked. While rework is sometimes possible for packages with low chip count [16], it is costly because it often involves detachment of connected chips, and the operation may degrade the reliability of the module itself. For the PxC-MCM, we apply the same wafer test yield of 75%, and also add a rework cost of 10% of the package cost. Simple rework is now possible because the replacement of a defective chip does not require a complex detachment procedure.

Figure 4 shows the yield improvement that the PxC-MCM offers over the SCM and traditional MCM. Applying the above assumptions, the cost improvement offered by the PxC-MCM for a 400 mm^2 chip is 4× over the SCM and 2× over the traditional 4-chip MCM. These benefits improve for larger chips and higher counts of integrated chips per module, respectively. While this figure implies that MCMs are always more cost-effective than SCMs, it does not consider the bandwidth reduction and potential performance loss that results from breaking a single chip into many small chips in an MCM. Note that chips on PxC-MCMs are not limited in this way due to their high bandwidth (and low-power) chip-to-chip I/O.

ENABLING MODULAR HETEROGENEOUS SYSTEMS

Functional Modularity

Increasing performance along with application flexibility has motivated the integration of many disparate specialized functional units on a chip. This is especially true in recent multimedia processors for the consumer electronics market [24 - 28] which typically integrate functional units with potentially very different activity and power requirements. For example, a recent processor [25] integrates onto a single chip three CPUs, each supporting a different operating system, media engines for MPEG4 and H.264 video, baseband units for WCDMA and GSM, a GPS unit, a 3D graphics accelerator, a mobile video unit, and a music unit.

Supporting different media standards (JPEG, MPEG, H.264) [27] and quality profiles [28] may also have very different hardware and bandwidth requirements. The H.264 video standard, for example, defines 15 levels of profiles with bit rates that vary from 64 kb/s to 240 Mb/s [29]. H.261 bit rates vary between 40 kb/s to 2Mb/s [30].

Satisfying a large application space with a single chip or MCM necessitates these heterogeneous systems with many different functional units. These implementations, however, are costly not only due to the large die area required to support the functional units, but also from the design challenges in supporting IP blocks from different vendors, managing clocking and voltage interfaces between IP blocks, and debugging chip errors across multiple IP boundaries. In addition, power dissipation constraints are forcing designers to partition designs at higher levels [24], separate functional units into many power domains [25], and develop advanced power management techniques [26].

A modular approach to building such heterogeneous systems enables application-specific configurability, and hence an optimal mix of functional units that target specific applications. By assembling the appropriate mix of base processors, media processors, baseband units, and memory, designers can configure systems for a particular media standard, wireless interface, or quality profile, without having to incorporate all hardware functions into a chip and disable the unneeded ones in order to satisfy power budgets. By providing high-bandwidth I/O and a packaging approach that encourages the addition and replacement of individual chips, Proximity Communication allows this level of modularity in building heterogeneous systems.

Physical Modularity

The flexibility to break functional units into separate monolithic chips is also useful because it allows different process optimizations. For example, SRAM circuits can benefit from better transistor matching, and different threshold or supply voltages, while not needing as many metal layers as what a standard digital process provides. DRAM circuits require specialized structures like discrete trench or stacked capacitors. Radio-frequency circuits benefit from optimized substrate resistivity and processing steps that provide high-Q inductors, and may be better implemented in SiGe bipolar technologies. Optoelectronic interfaces may require direct bandgap materials unlike silicon. Sensor units may employ MEMS techniques that require materials and chemical etching processes that are incompatible with traditional CMOS chips. Chips with lower performance requirements may reduce the overall system cost when built in an older, less expensive fabrication technology.

Proximity Communication allows designers to easily assemble a physically heterogeneous system because there are no direct electrical connections between chips. It eliminates the requirement to match signaling levels or perform level conversion between chips that use different I/O voltages. The freedom to combine chips of very different processes and materials that best serve their respective functions alleviates some of the challenges that designers face in trying to adapt a standard digital CMOS process for more foreign applications.

SUMMARY

Proximity Communication eliminates traditional chip-to-chip I/O bottlenecks, allowing multiple chips to function together as one large chip without penalties in performance or cost due to yield fallout. It enables reworkable packages, providing not only performance but also yield benefits over traditional MCM and SiP approaches. By enabling the modular assembly of many separate chips, Proximity Communication simplifies and lowers the cost of building scalable, high-performance, heterogeneous systems.

REFERENCES

- [1] M. Horowitz, "Scaling, Power, and the Future of CMOS," in *IEEE International Electron Devices Meeting*, 2005.
- [2] G. E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics*, Vol. 38, No. 8, pp. 114-117, 19 Apr. 1965.
- [3] G. E. Moore, "Progress in Digital Integrated Electronics," in *IEEE International Electron Devices Meeting*, 1975.
- [4] R. Ho, *et al.*, "High-speed and Low-Energy Capacitively-Driven On-Chip Wires," in *2007 IEEE International Solid-State Circuits Conference*, 11-15 Feb. 2007, pp. 412-612.
- [5] P. R. Groeneveld, "Physical Design Challenges for Billion Transistor Chips," in *Proceedings of the 2002 IEEE International Conference on Computer Design: VLSI in Computers and Processors (ICCD'02)*, 16-18 Sept. 2002, pp. 78-83.
- [6] G. E. Moore, "Lithography and the Future of Moore's Law," in *Proceedings of SPIE*, Vol. 2438, Jun. 1995, pp. 2-17.
- [7] R. J. Drost, R. D. Hopkins, R. Ho, and I. Sutherland, "Proximity Communication," in *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 9, Sept. 2004, pp. 1529-1535.
- [8] A. Chow, D. Hopkins, R. Ho, and R. Drost, "Measuring 6D Chip Alignment in Multi-Chip Packages," in *Proceedings of the 2007 IEEE Sensors Conference*, 28-31 Oct. 2007, pp. 1307-1310.
- [9] R. Drost, R. Ho, D. Hopkins, and I. Sutherland, "Electronic alignment for proximity communication," in *2004 IEEE International Solid-State Circuits Conference*, 15-19 Feb. 2004, pp. 144-518.
- [10] D. Hopkins, *et al.*, "Circuit Techniques to Enable 430Gb/s/mm² Proximity Communication," in *2007 IEEE International Solid-State Circuits Conference*, 11-15 Feb. 2007, pp. 368-609.
- [11] B. T. Murphy, "Cost-Size Optima of Monolithic Integrated Circuits," in *Proceedings of the IEEE*, Vol. 52, No. 12, Dec. 1964, pp. 1537-1545.
- [12] C. H. Stapper, "On Murphy's Yield Integral," in *IEEE Transactions on Semiconductor Manufacturing*, Vol. 4, No. 4, Nov. 1991, pp. 294-297.
- [13] B. Meister, "On Murphy's Yield Formula," in *IBM Journal of Research and Development*, Vol. 27, No. 6, Nov. 1983, pp. 545-548.
- [14] S. Kim and F. Lombardi, "Yield Analysis of Fault-Tolerant Multichip Module Systems for Massively Parallel Computing," in *7th Annual IEEE International Conference on Wafer Scale Integration*, 18-20 Jan. 1995, pp. 308-317.
- [15] L. Zheng, M. Shen, and H. Tenhunen, "System-on-Chip or System-on-Package: Can we make an accurate decision on system implementation in an early design phase?" in *Southwest Symposium on Mixed-Signal Design*, 23-25 Feb. 2003, pp. 1-4.
- [16] J. K. Hage and R. J. Wagner, "High-Yield Assembly of Multichip Modules Through Known-Good IC's and Effective Test Strategies," in *Proceedings of the IEEE*, Vol. 80, No. 12, Dec. 1992, pp. 1965-1994.
- [17] E. E. Davidson, "Large Chip vs. MCM for a High-Performance System," in *IEEE Micro*, Vol. 18, No. 4, Jul-Aug. 1998, pp. 33-41.
- [18] *International Technology Roadmap for Semiconductors*, 2007 Edition, 31 Dec. 2007.
- [19] U. Nawathe, *et al.*, "An 8-Core 64-Thread 64b Power-Efficient SPARC SoC," in *2007 IEEE International Solid-State Circuits Conference*, 11-15 Feb. 2008, pp. 108-590.
- [20] M. Tremblay and S. Chaudhry, "A Third-Generation 65nm 16-Core 32-Thread Plus 32-Scout-Thread CMT SPARC[®] Processor," in *2008 IEEE International Solid-State Circuits Conference*, 3-7 Feb. 2008.
- [21] J. Friedrich, *et al.*, "Design of the Power6[™] Microprocessor," in *2007 International Solid-State Circuits Conference*, 11-15 Feb. 2007, pp. 96-97.
- [22] S. Rusu, *et al.*, "A Dual-Core Multi-Threaded Xeon[®] Processor with 16MB L3 Cache," in *2006 International Solid-State Circuits Conference*, 5-9 Feb. 2006, pp. 102-641.
- [23] B. Stackhouse, "A 65nm 2-Billion-Transistor Quad-Core Itanium[®] Processor," in *2008 International Solid-State Circuits Conference*, 3-7 Feb. 2008.
- [24] A. Abbo, *et al.*, "XETAL-II: A 107 GOPS, 600mW Massively-Parallel Processor for Video Scene Analysis," in *2007 IEEE International Solid-State Circuits Conference*, 11-15 Feb. 2007, pp. 270-602.
- [25] M. Ito, *et al.*, "A 390MHz Single-Chip Application and Dual-Mode Baseband Processor in 90nm Triple-Vt CMOS," in *2007 IEEE International Solid-State Circuits Conference*, 11-15 Feb. 2007, pp. 274-602.
- [26] B. Nam, *et al.*, "A 52.4mW 3D Graphics Processor with 141Mvertices/s Vertex Shader and 3 Power Domains of Dynamic Voltage and Frequency Scaling," in *2007 IEEE International Solid-State Circuits Conference*, 11-15 Feb. 2007, pp. 278-603.
- [27] C. Chien, *et al.*, "A 252kgate/71mW Multi-Standard Multi-Channel Video Decoder for High Definition Video Applications," in *2007 IEEE International Solid-State Circuits Conference*, 11-15 Feb. 2007, pp. 282-603.
- [28] H. Chang, *et al.*, "A 7mW-to-183mW Dynamic Quality-Scalable H.264 Video Encoder Chip," in *2007 IEEE International Solid-State Circuits Conference*, 11-15 Feb. 2007, pp. 280-603.
- [29] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, Jul. 2003, pp. 560-576.
- [30] "Line Transmission of Non-Telephone Signals: Video CODEC for Audiovisual services at $p \times 64$ kbits," *ITU-T Recommendation H.261*, March 1993.