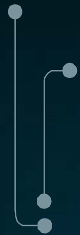


Interconnect Fabrics

Don Draper

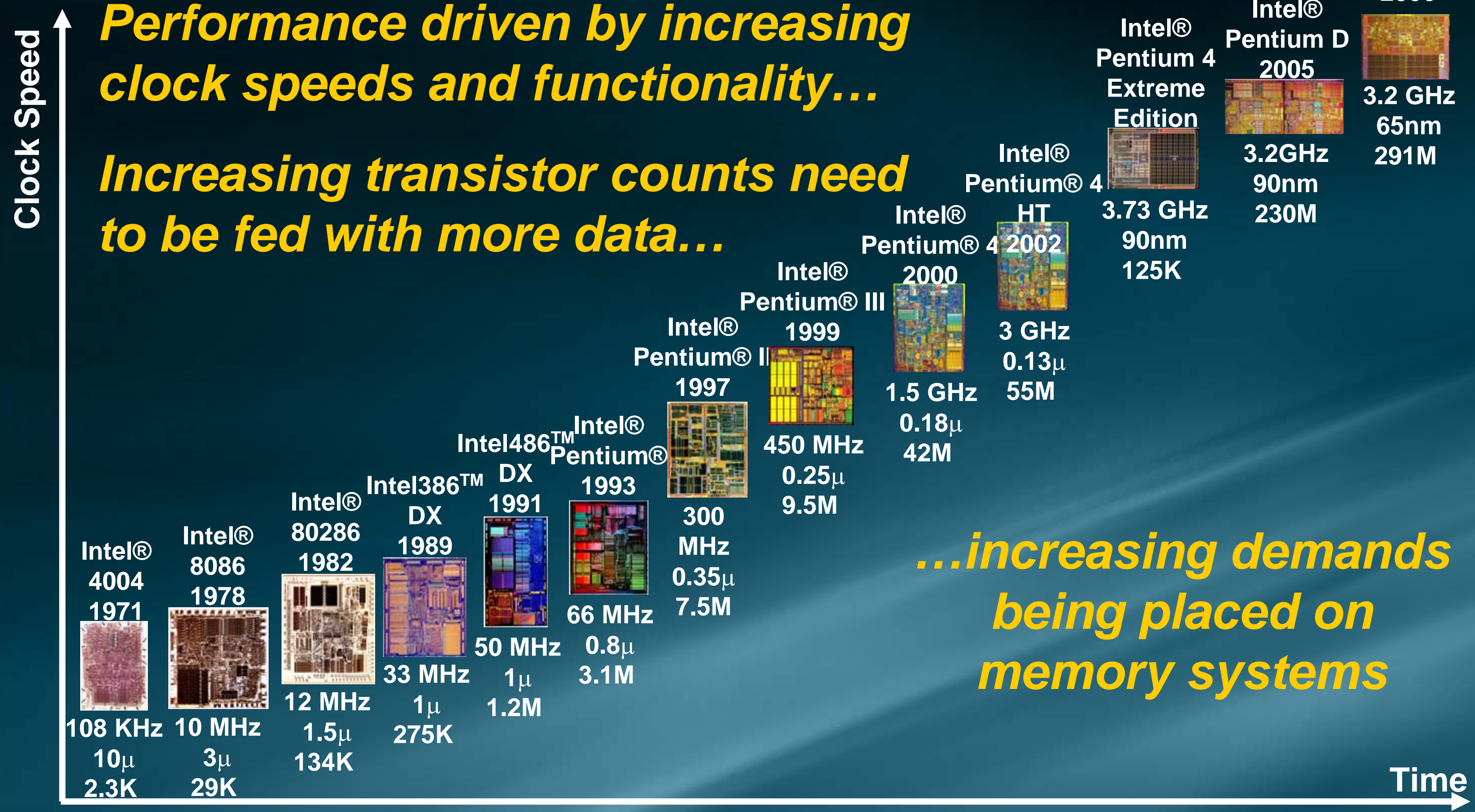
Rambus, Inc.

SRC/NSF/ITRS 2020 Virtual Immersion Architectures

Santa Cruz, CA

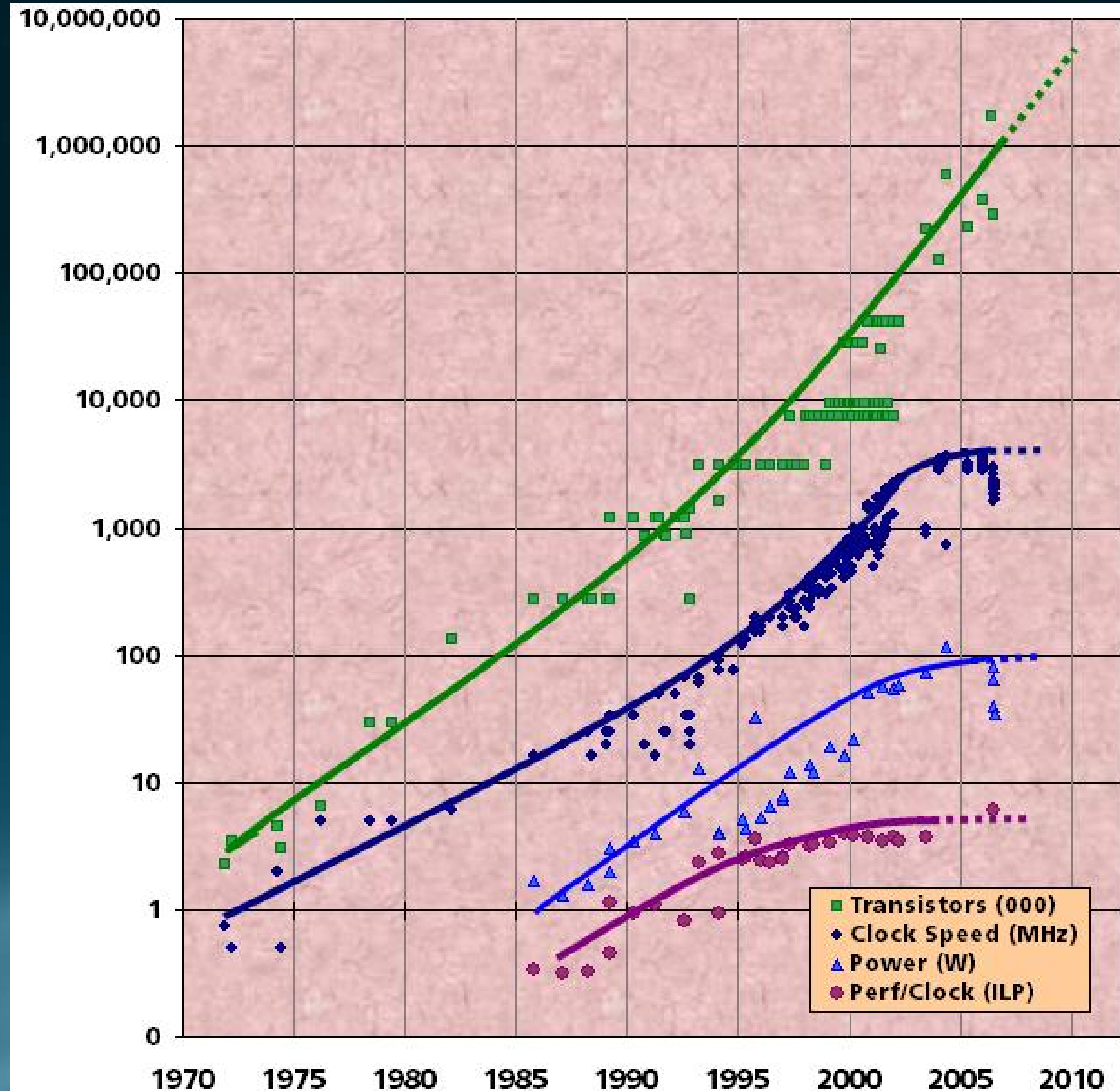
July 10-11, 2008

Moore's Law Driving Performance



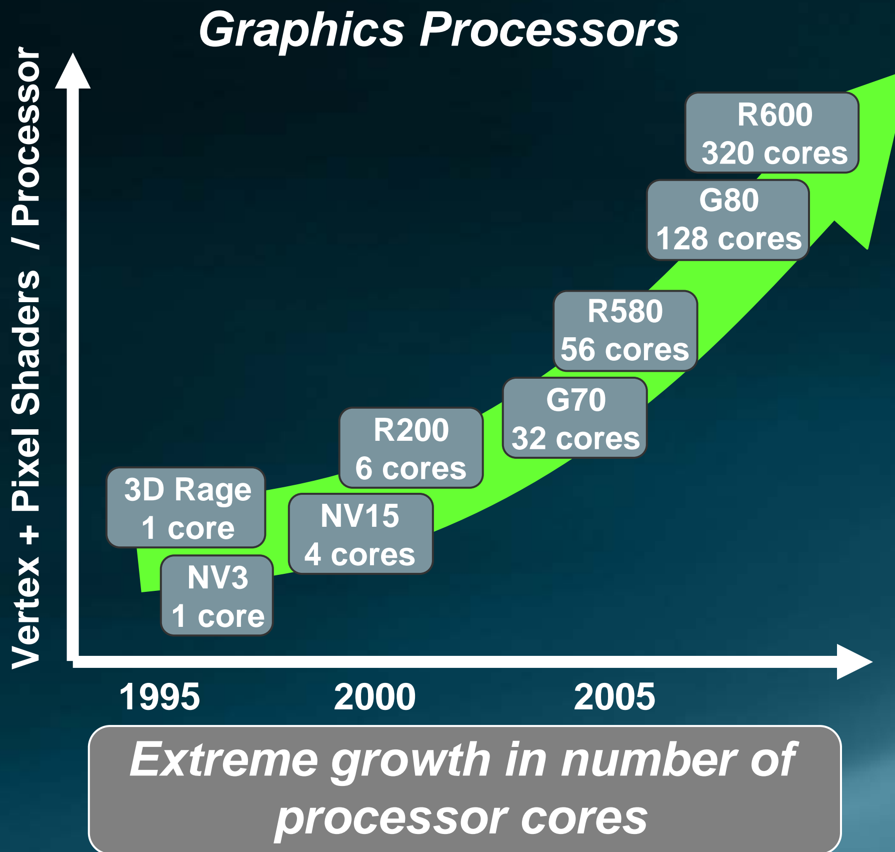
Clock Scaling Bonanza Has Ended

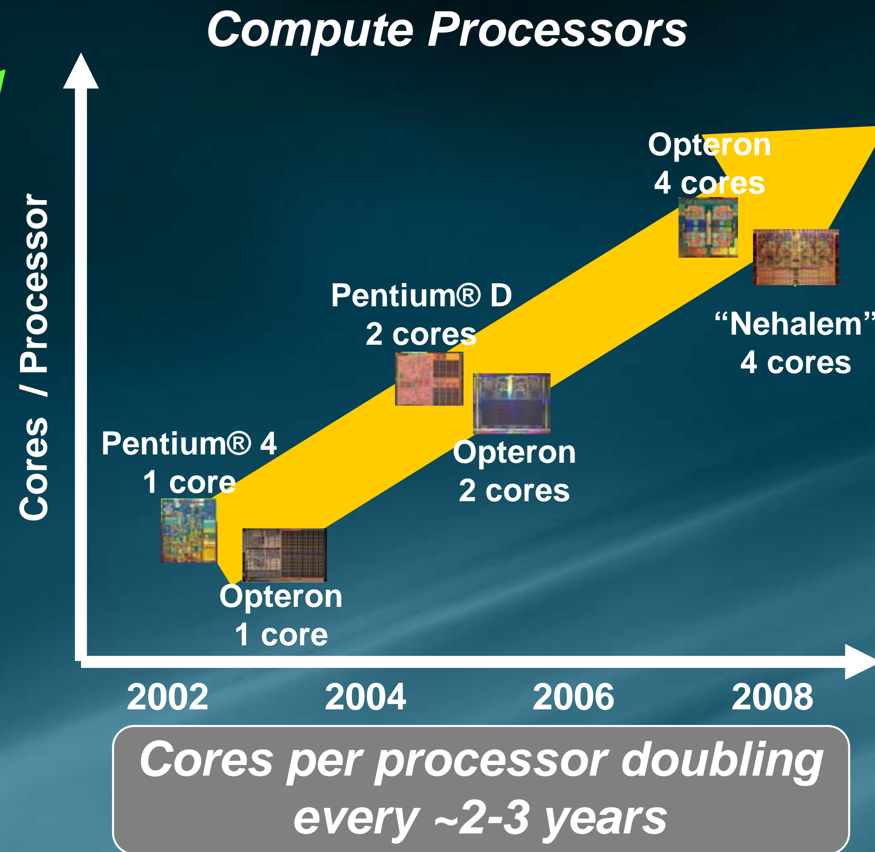
- Chip density is continuing to increase ~2x every 2 years
  - Clock speed is not.
  - Number of processor cores may double instead
- There is little or no hidden parallelism (ILP) to be found
- Parallelism must be exposed to and managed by software



Source: Intel, Microsoft (Sutter) and Stanford (Olukotun, Hammond)

Dramatic Growth in Number of Graphics and Compute Processor Cores





Extreme growth in number of processor cores

Cores per processor doubling every ~2-3 years

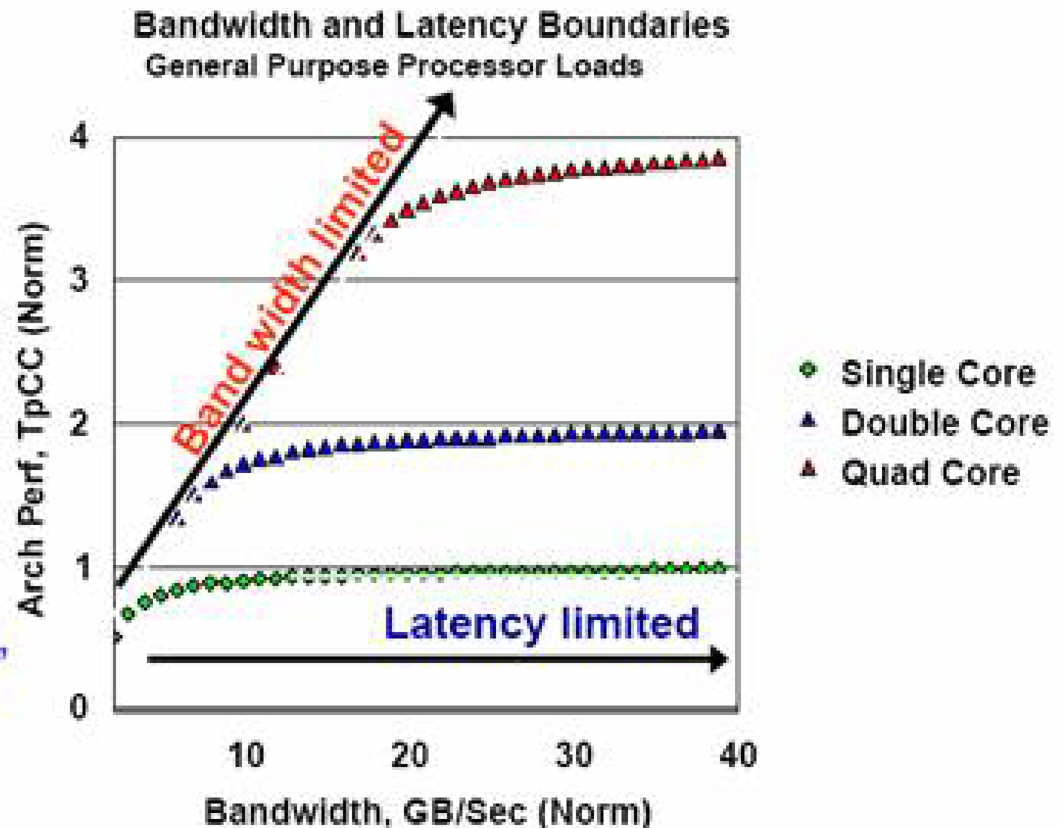
Rising core counts placing increasing demands on memory performance

Multi-core Processor Trend: More Cores Need Higher Bandwidth

Processor load trade-off between I/O Bandwidth, Bus Latency.

- For generic workloads, uni-processor perf saturates bandwidth benefit, becomes latency-limited.

- As core counts increase, I/O Bandwidth becomes increasingly important

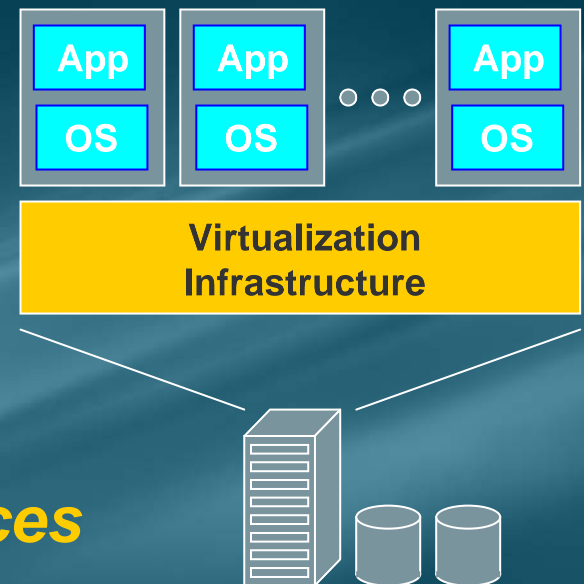


Ref: K. Bernstein, "New Dimensions in Microarchitecture," Micro-39, 12/06

TpCC – Transaction Processing Performance Council – C Benchmark
Industry standard benchmark for server online transaction processing (OLTP)

Virtualization Redefining Data Centers

- Difficult to write programs for multi-core CPUs
  - Run multiple programs in parallel instead
- Virtualization abstracts computer resources
  - Run multiple applications and operating systems at the same time
- Virtualization benefits
  - Increased resource utilization
  - Ease of management
  - Reduced operating costs



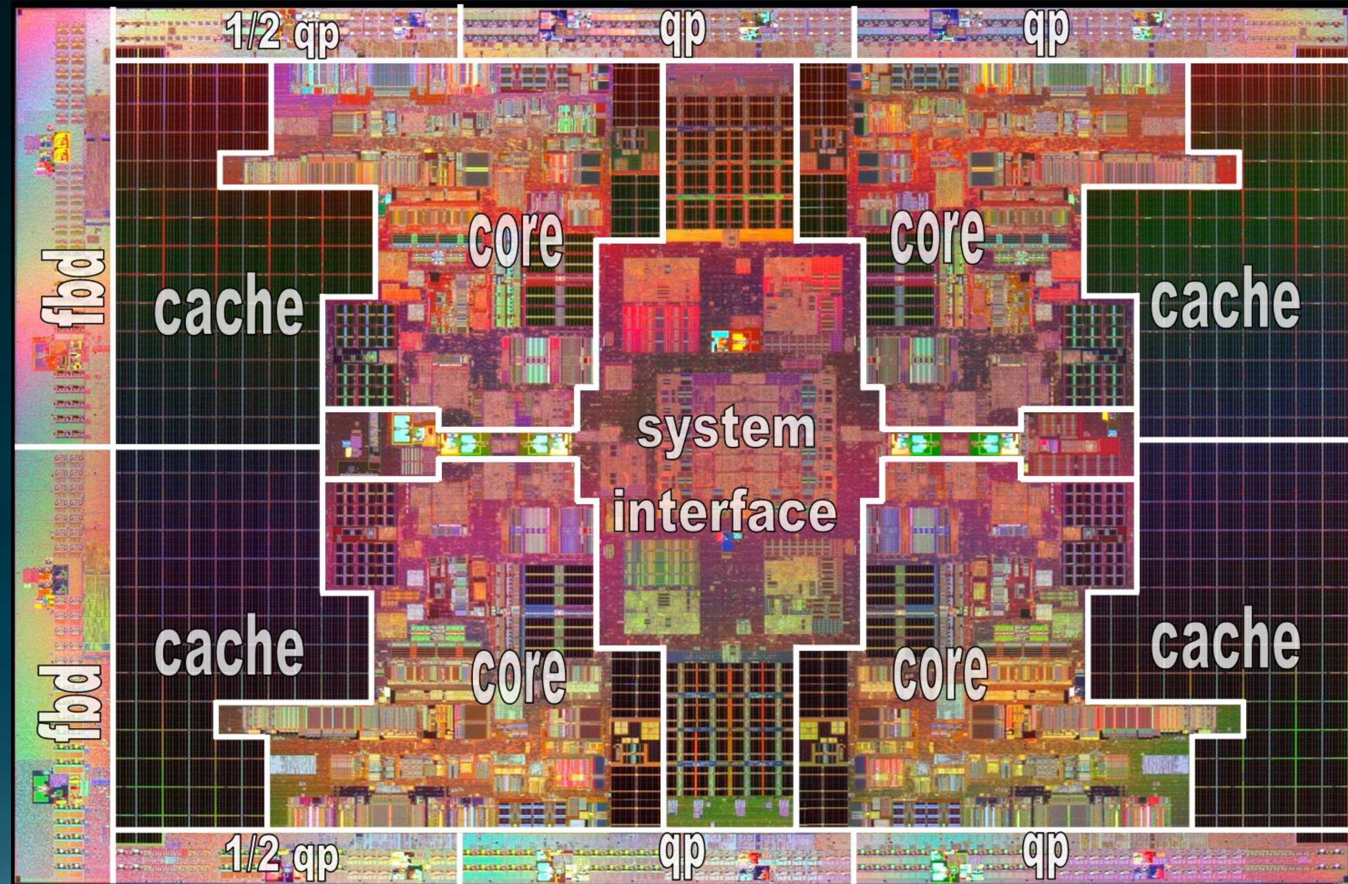
Virtualization increases memory bandwidth and capacity needs, reduces locality in memory system traffic

DDR DRAM Speed Ranges

Features/ Options	DDR2	DDR3	Projected DDR4	Comments
Speed (data pin)	400-, 533-, 667-, 800Mbps	800-, 1066-, 1333-, 1600Mbps	1600-, 2133-, 2667-, 3200Mbps	Migration to higher speed I/O
Prefetch (MINI READ burst)	4bit (2 clocks)	8bit (4 clocks)	8bit (4 clocks)	Reduced core speed dependency for better yield
Internal banks	4 (256Mbit, 512Mbit) 8 (1-, 2-, 4Gbit)	8 (512Mbit, 1Gbit, 2Gbit, 4Gbit, 8Gbit)	16	Larger density per monolithic package, 8 banks standard
Voltage	1.8V 1.8V I/O	1.5V 1.5V I/O	1.2V 1.2V I/O	Reduces memory system power demand
Densities	256 Mbit-4Gbit	512Mbit-8Gbit	up to 16Gbit	High-density components enable large capacity memory subsystems
Pinout/package	60-ball; x4, x8 84-ball; x16 FBGA only	78-ball; x4, x8 96-ball; x16 FBGA only	FBGA only	Independent pinout for x4/x8 and x16 (simplifies module design)
Data strobes	Single ended or differential	Differential only	TBD	Reduce data strobe crosstalk
R_{tt} values	50-, 75-, 150ohm	120-, 60-, 40-, 30-, 20ohm and dynamic ODT	TBD	Support higher data rates
DQ driver impedance	18ohm	34ohm, 40ohm	TBD	Optimized for 2-slot and point-to-point systems
Multipurpose register	None	Four registers - 2 defined, 2 RFU	TBD	Provides specialty readouts
RESET#	None	Dedicated input	Dedicated input	Disable outputs, resets DRAM

Interconnects on on a 65nm 2 Billion Transistor Quad-Core Itanium® Processor

- Interconnects have a transfer rate of 4.8GT/s.
- 4 full and 2 half QPI links give peak bandwidth of 96 GB/s
- 4 FBD channels give 34 GB/s



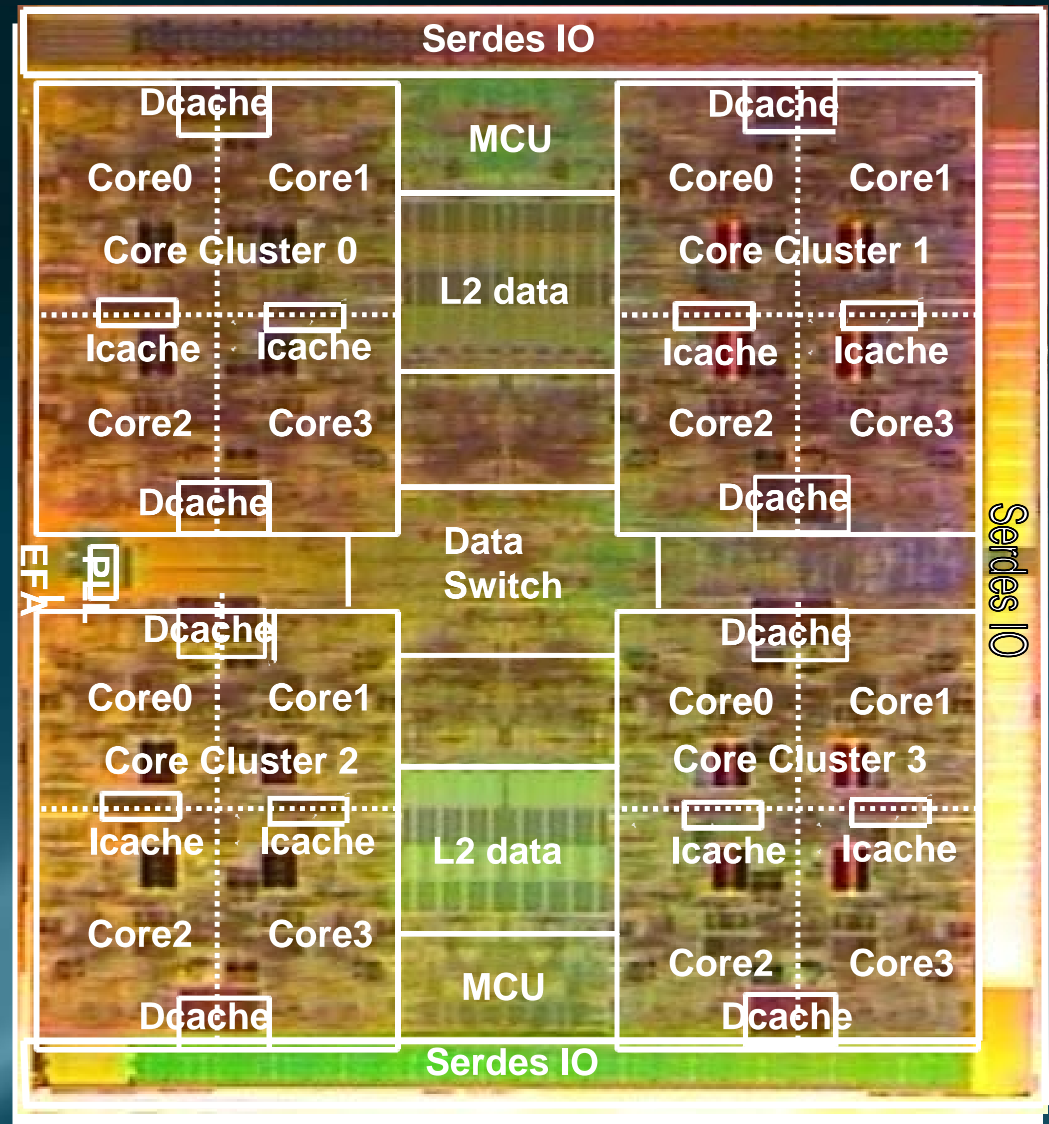
ISSCC 2008,
Blaine Stackhouse, et al
Intel

Quickpath Interconnect Implementation

- 4 tap transmit equalization
- Transmit clock duty cycle correction
- PVT tolerant, precision bias current generation
- Per lane tunable transmitter output swing
- Low jitter forwarded clock architecture
- Differential inputs employ voltage offset compensation
- AC coupled forwarded clock recovery scheme
- Receive clock duty cycle correction.
- 8 phase delay locked loop
- 3 interpolator receive data recovery scheme

Interconnects on a16-Core 32-Thread Plus 32-Scout-Thread CMT SPARC® Processor

- **Memory and system interfaces run at 2.67 Gb/s**
- **The memory interface consists of 96 transmit and 160 receive channels with a total throughput of 680 Gb/s**
- **The system interface consists of 15 transmit and receive channels with a total throughput of 80 Gb/s**



ISSCC2008 Marc Tremblay et al, Sun Microsystems

A 16Gb/s/link, 64GB/s Bidirectional Asymmetric Memory Interface Cell

Controller PHY Cell

7

DQ x8	DQ x8	C/A x2 / CTL	DQ x8	DQ x8

18 bumps

- 16 DRAMs @ 64GB/sec each
- Access granularity:
  - 64-byte: shown
  - 32-byte: + 32 diff pairs (6%)

16Gbps per differential pair

PHY PHY PHY PHY

2 C/A 32 DQ (×16)

Controller

17.2 x 12.9 = 222 mm^2 (150μ bump pitch)

ISSCC 2008, K. Chang et al, Rambus

Key Challenges for 1TB/sec+ Systems

- Maximize link bandwidth
  - 16 Gbps per differential pair and beyond
- Maximize link density at controller
  - Leverage packaging trends:

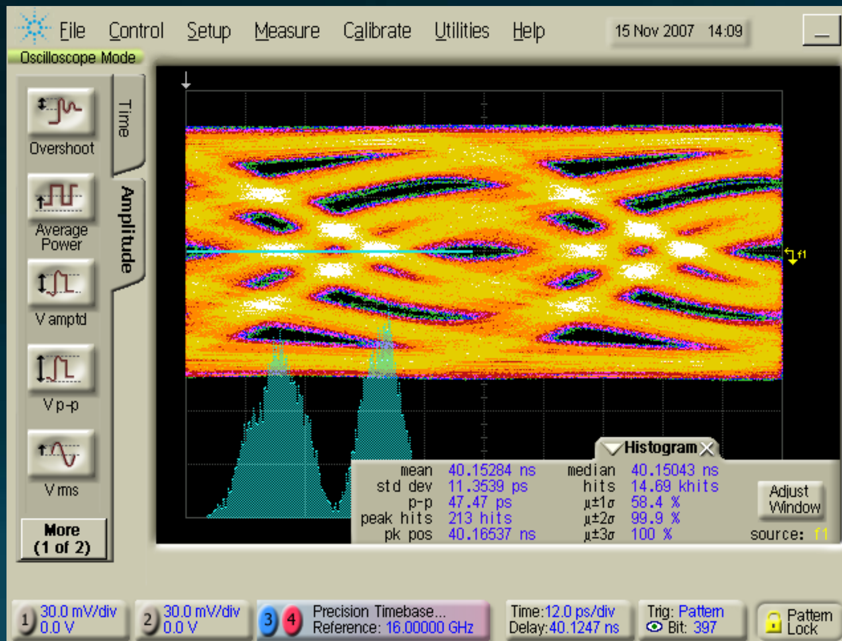
Parameter	Trend
Flip chip bump pitch	180μ going down to 150μ
BGA ball pitch	1.0mm going down to 0.8mm
Package body size	42.5mm and going up
Substrate thickness	400μ-800μ going down to 200μ

- Efficiently utilize available signals
  - Per-byte masks, strobes, clocks, DBI, etc. consume valuable pins
  - Use narrow command/address channels to reduce overhead
- Improve power efficiency

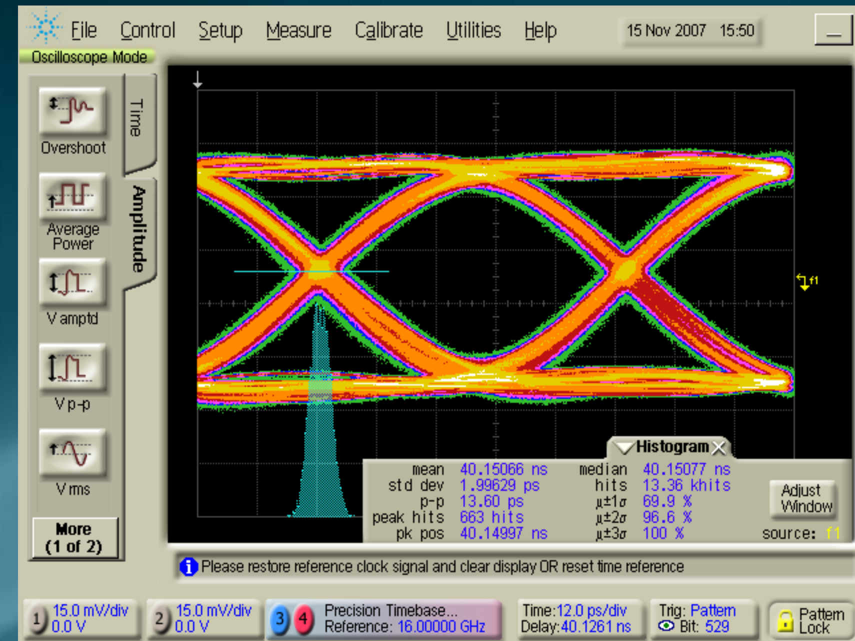
Equalization Enables Robust High Speed Signaling in Future Memory Systems

- Equalization required to reduce ISI at high data rates
- Asymmetric equalization reduces cost, complexity in DRAM

Controller unequalized 16Gbps TX eye



Controller equalized 16Gbps TX eye



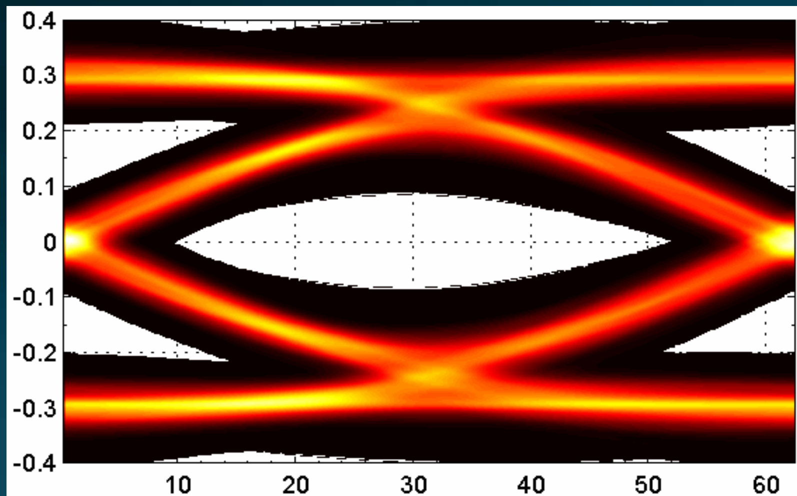
Pattern is PRBS 2^11-1

3" FR4 + 12" cable to scope

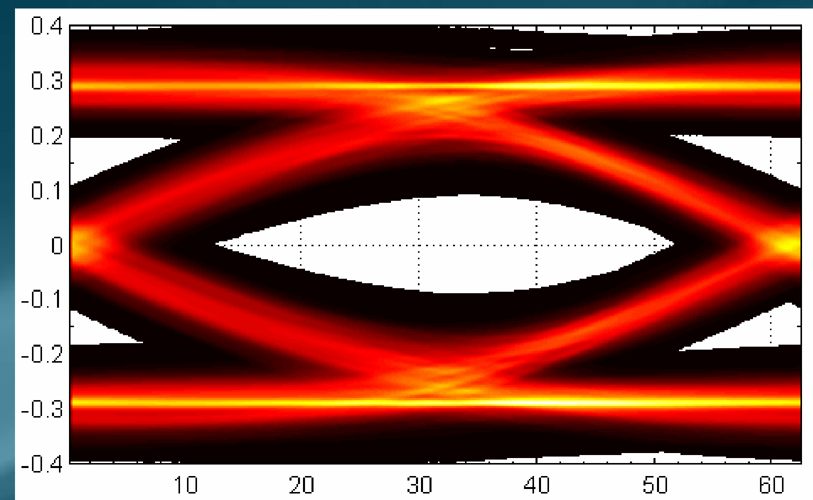
Asymmetric Equalization

- Equalization is required to reduce ISI at 16Gbps+
- Controller handles majority of equalization
  - Reduces cost and complexity of DRAM
  - Multi-tap TX FIR filter
  - RX linear equalizer
  - Adaptive algorithm to refine EQ coefficients in system

16Gbps Simulated Eye Diagrams



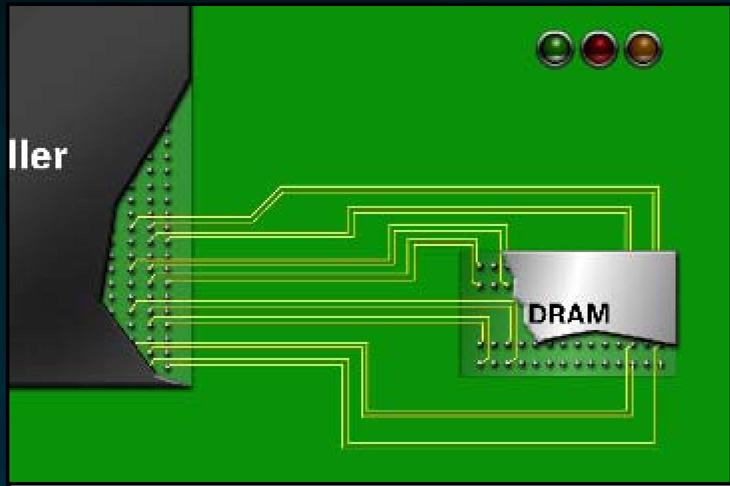
Write Eye
(@ DRAM RX input)



Read Eye
(after RX Linear EQ)

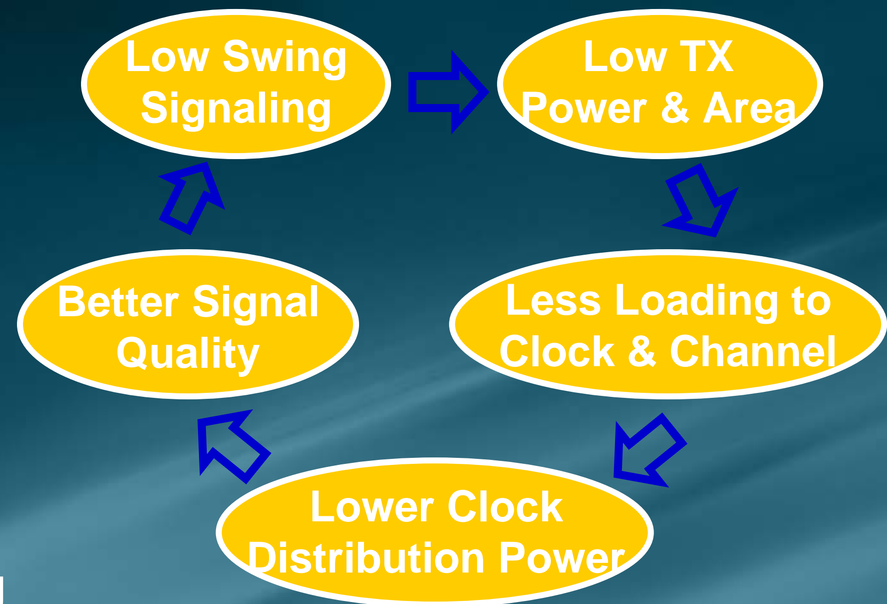
Robust Signaling and Reducing Power Becoming More Challenging in Future Memory Systems

- FlexPhase™ per-pin timing adjustment



- Benefits
  - Enables higher data rates
  - Compensates for manufacturing and environment variations
  - No trace length matching required, simplifies board design

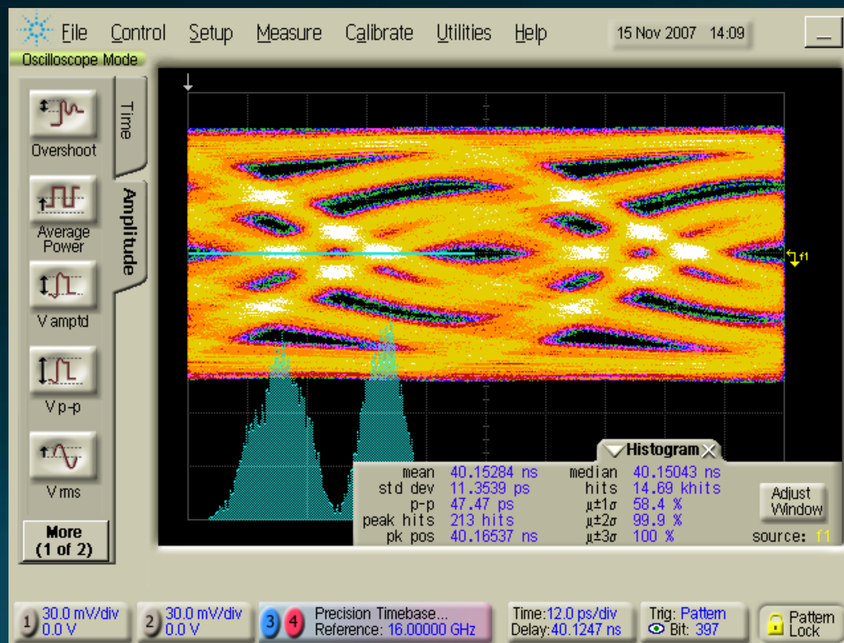
- Low swing signaling has <u>power</u> and <u>signal integrity</u> benefits



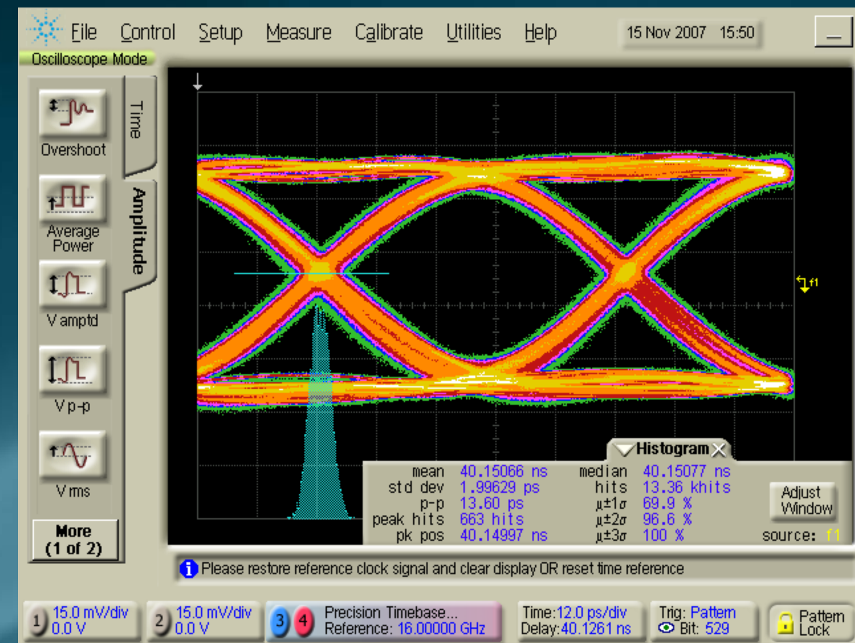
Equalization Enables Robust High Speed Signaling in Future Memory Systems

- **Equalization required to reduce ISI at high data rates**
- **Asymmetric equalization reduces cost, complexity in DRAM**

Controller unequalized 16Gbps TX eye



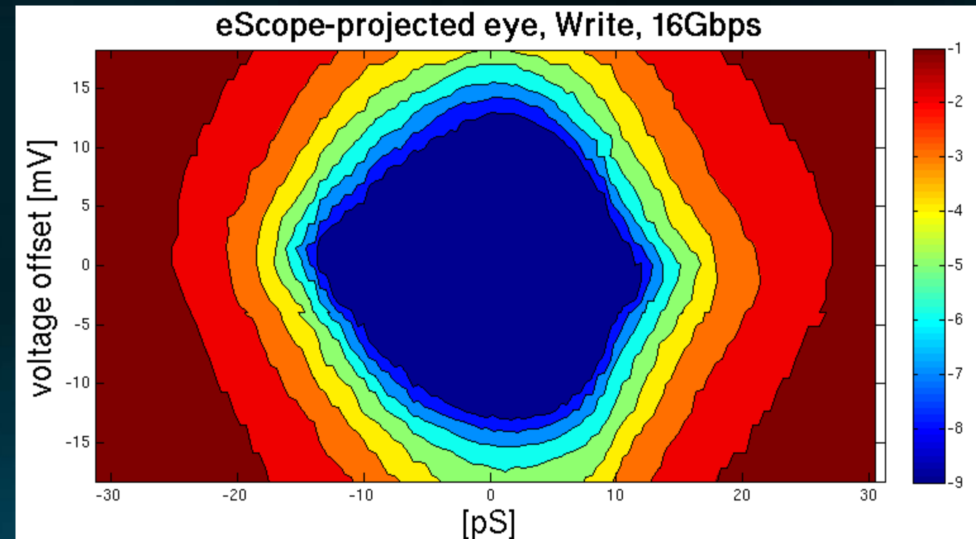
Controller equalized 16Gbps TX eye

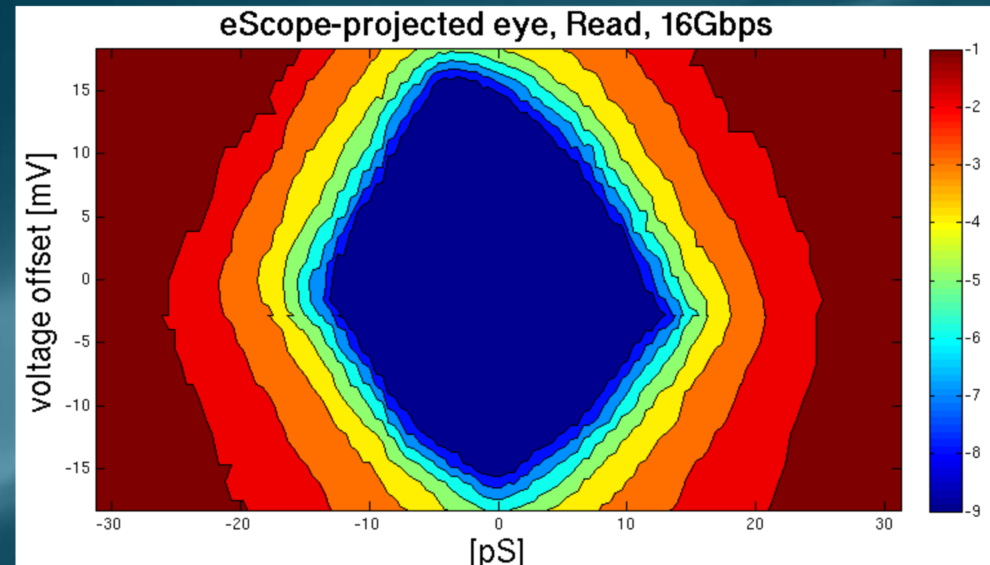


Pattern is PRBS 2^11-1
3" FR4 + 12" cable to scope

Receiver Eye Measurements at 16Gb/s

- **Measured** by using Tx twister to add offset on the channel for both read and write
- Half Tx swing Includes receiver effect
- Even with half swing, both directions show large timing margin (>0.3UI for BER = 10-9) and voltage margin (>30mV)

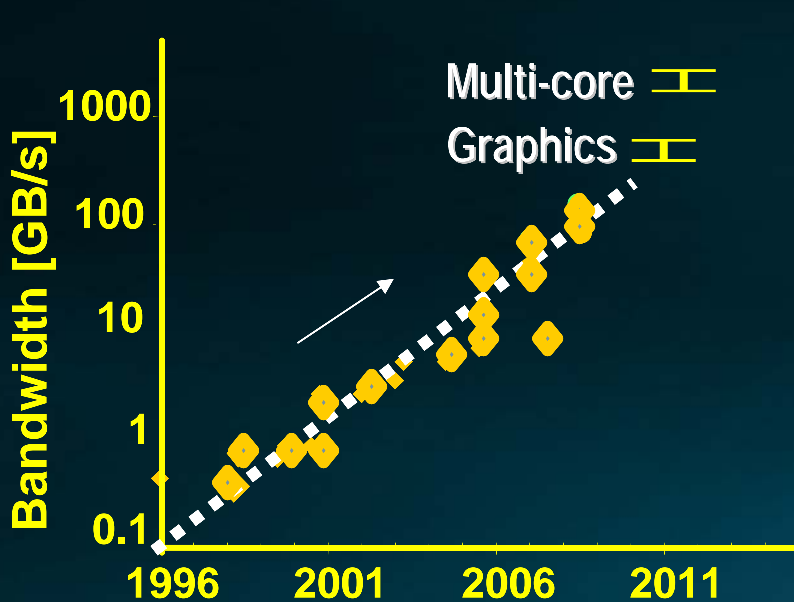


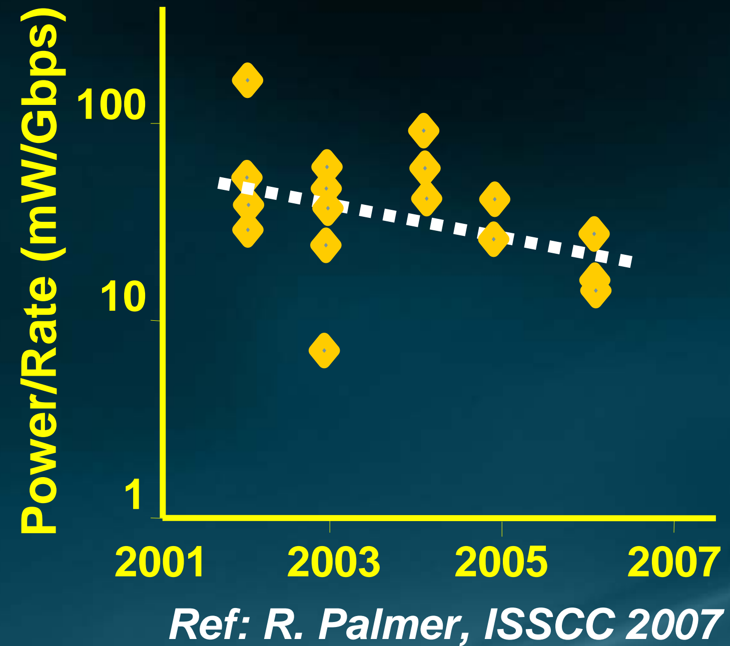


Performance Summary

Technology	TSMC 65nm G+
Supply Voltage (VddA/VddIO/Vdd)	1.1/1.2/1.1V nominal
Cell size (4 DQ bytes + RQ block)	3.6mm x 1.5mm
Bandwidth per DQ and RQ link	16Gb/s
Total data bandwidth per cell	64GB/s
Energy efficiency (nominal conditions)	13mW/Gb/s
LC PLL range	8GHz +/- 5%
LC clock jitter ($RJ_{,rms}/TJ_{,BER=10^{-12}}$)	318fs / 7.7ps
Tx jitter ($RJ_{rms}/TJ_{BER=10^{-12}}$), all links and bytes active	810fs / 24.4ps
Read/write BER @ 16Gb/s (from overnight runs)	$< 10^{-14}$
Write timing margin at BER = 10^{-12}	19% UI, 11.9ps
Read timing margin at BER = 10^{-12}	19% UI, 11.9ps

IO Power Considerations



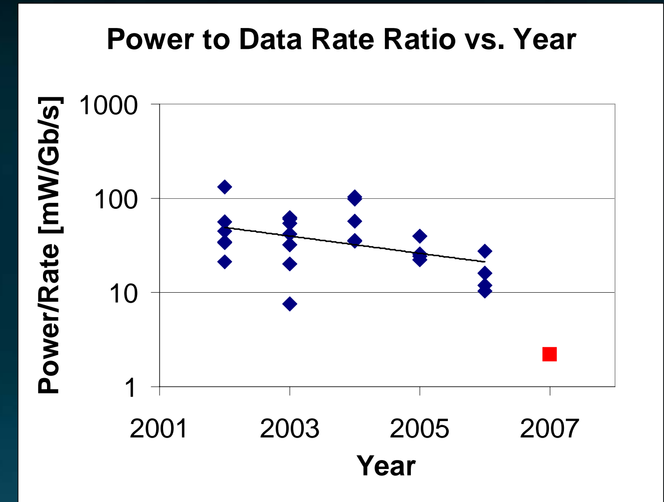


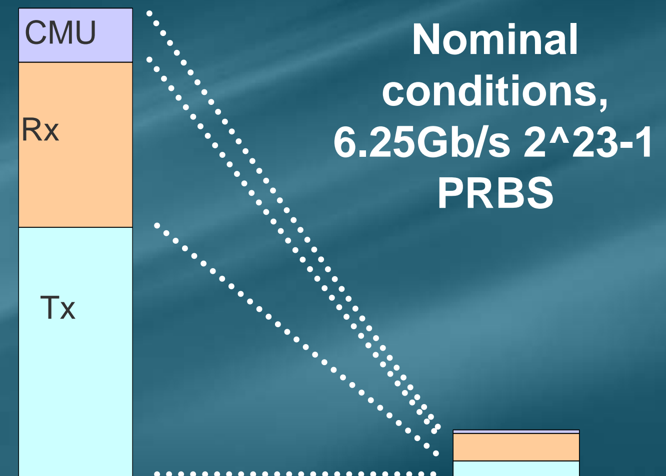
Ref: R. Palmer, ISSCC 2007

- Future off-chip bandwidth demand ⇒ >1TB/s
- Current I/O power efficiencies > 10mW/Gbps
- Significant improvement needed to enable future computing platforms

A 14mW 6.25Gb/s Transceiver in 90nm CMOS for Serial Chip-to-Chip Communications

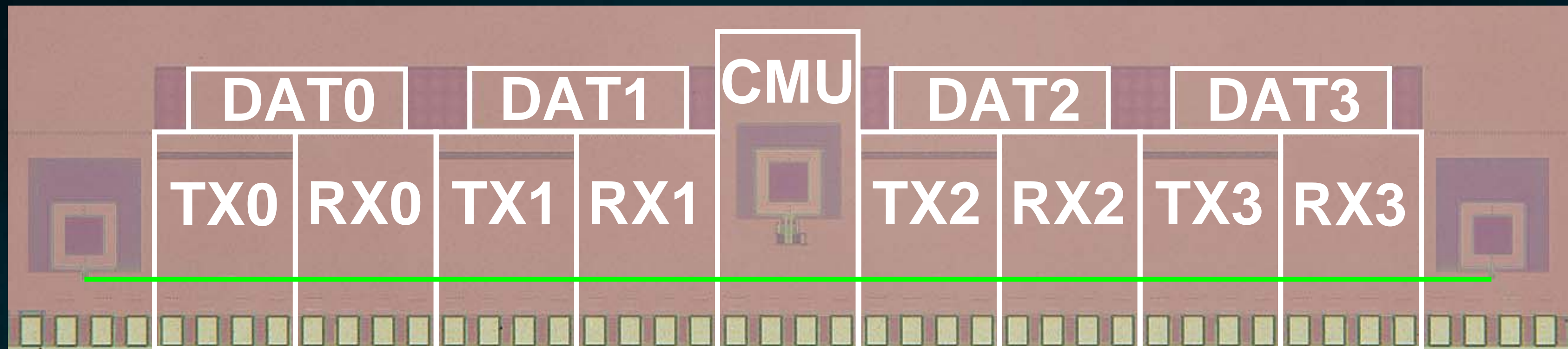
- 6.25 Gb/s in 90nm @ 2.2 mW/Gb/s
- Low-power techniques:
  - Low-swing, voltage-mode signaling
  - Shared CMU
  - Resonant clock distribution
  - Offset-trimmed receiver
  - Low-bandwidth CDR in software
  - Edge-based adaptive EQ in software
  - Low-power phase rotator





ISSCC 2007, R. Palmer et al, Rambus

Transciever Implementation (x4)

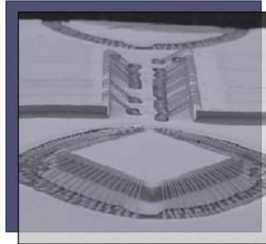
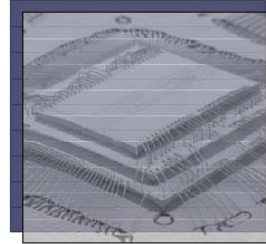


- TSMC 90-nm 1.0/2.5V "G" process
- 4 transceivers, each 640x480um
- Shared CMU, 320x712um
- Clock distribution inductors, each 140x140um
- 4 data generator/checker/loopback blocks
- Control register interface to FPGA (controller)
- Bond-wire BGA package

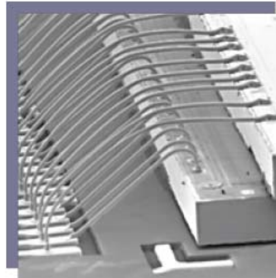
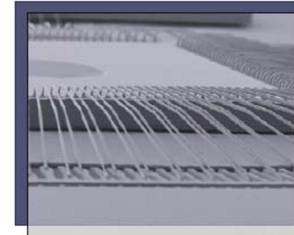
Other Memory Approaches – Chip Stacking

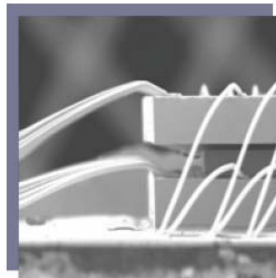
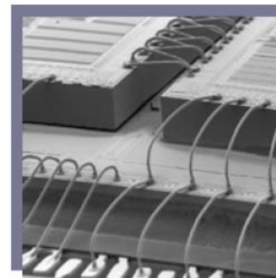
Wire Bond Stacks Work Today

- Complex
- Pitch Limited

- Large Form-Factor
- Performance Limited

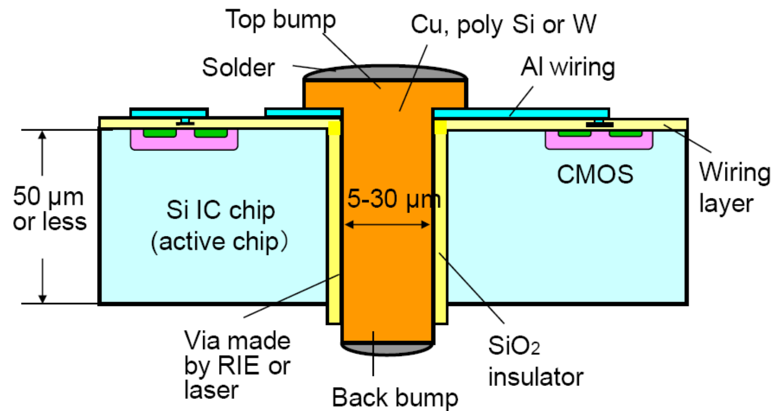



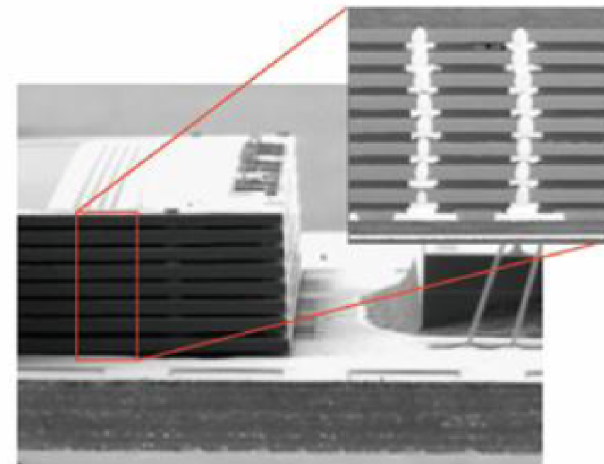
4

©2008 Micron Technology, Inc. All rights reserved.

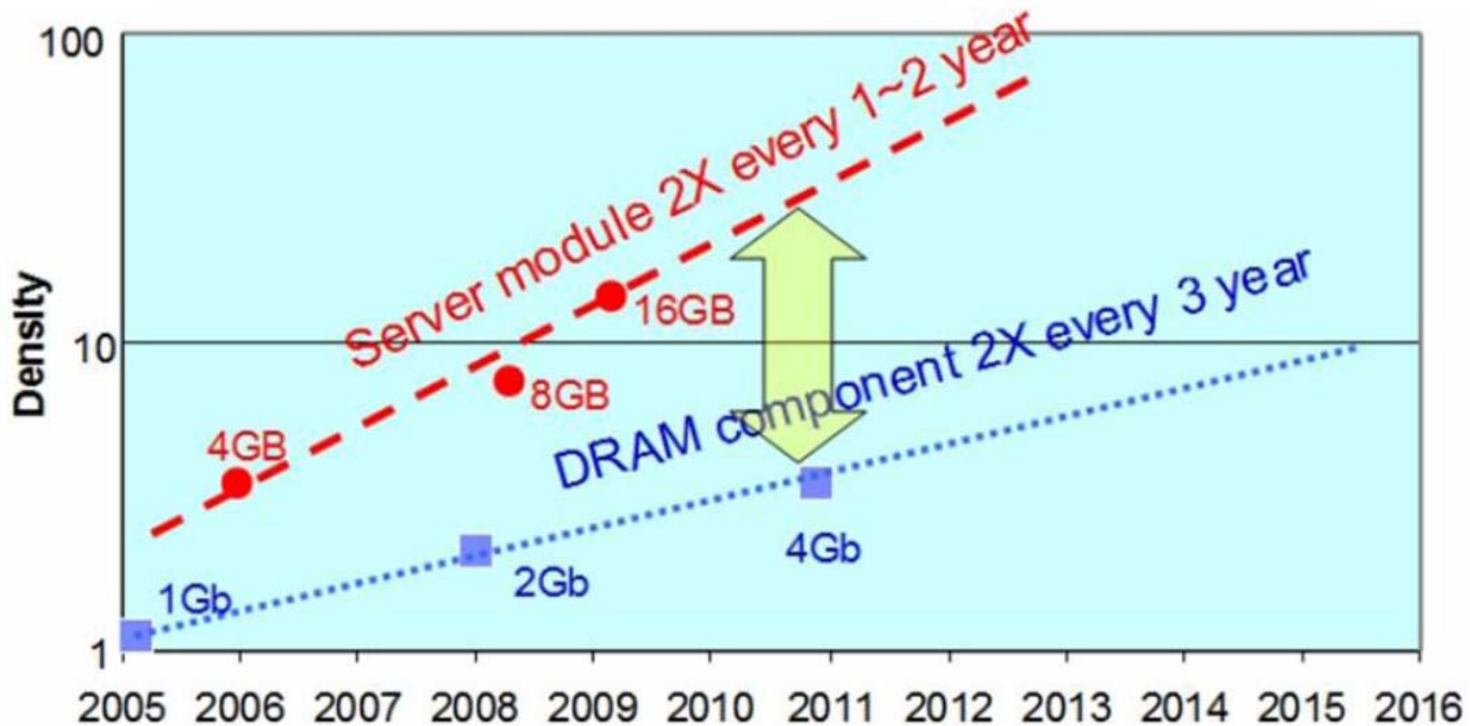
Micron

Die-to-Die Interconnect: Through-Silicon Vias





Driving Force for TSV in DRAM



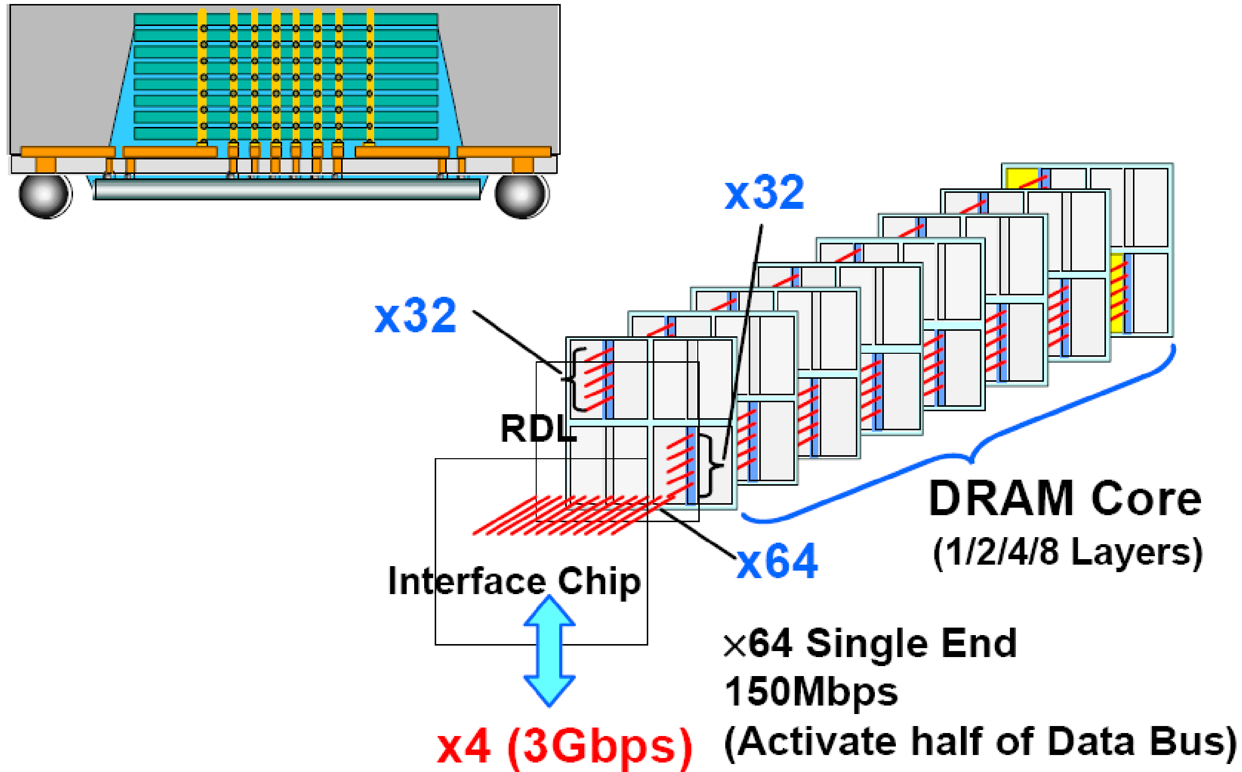
TSV Trends for Memory

	Today	Tomorrow	Future
Via Diameter	10–30µm	5–10µm	1–5µm
Via Depth	50–100µm	25–50µm	15–25µm
Wafer Thickness	50–100µm	25–50µm	15–25µm
Aspect Ratio	3:1–5:1	5:1–10:1	10:1 & greater
Pitch	75–100µm	50–75µm	20–50µm
TSV Density	~100 I/O	~500 I/O	~1000 I/O
Package Type	DDP	QDP	8DP

©2008 Micron Technology, Inc. All rights reserved.

Micron

DRAM Stacking



Fully Integrated VLSI CMOS and Photonics "CMOS Photonics"

Cary Gunn - Luxtera, Inc., Carlsbad, CA

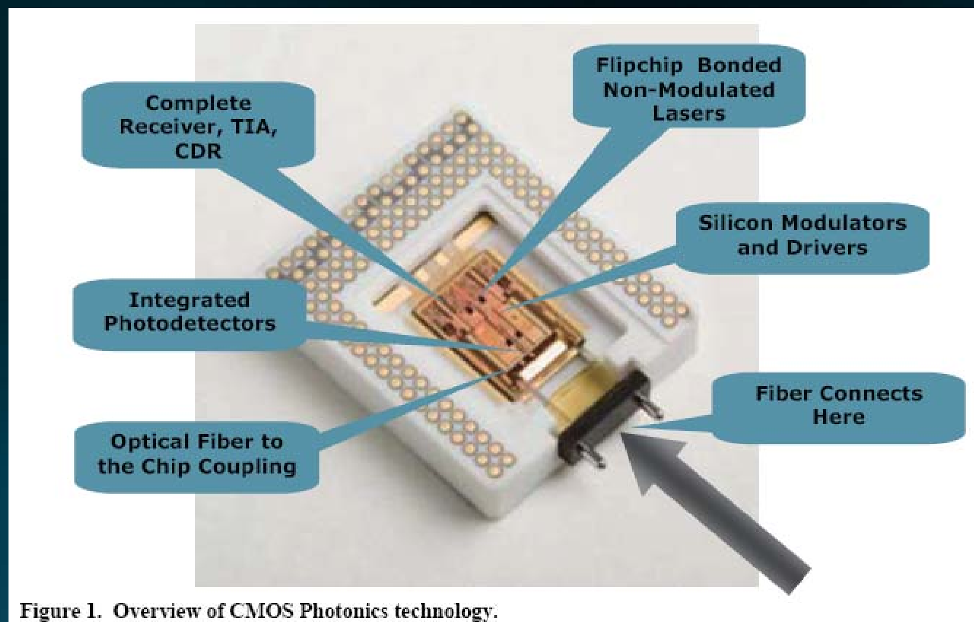


Figure 1. Overview of CMOS Photonics technology.

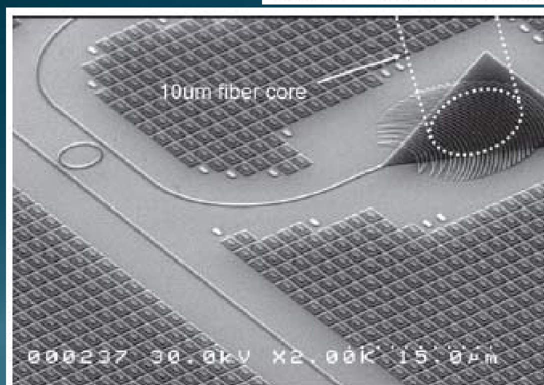


Figure 2. Oblique view of a holographic lens connected to silicon waveguides. The optical fiber illuminates the HL from a position normal to the surface. The core of the fiber is shown by the dashed lines.

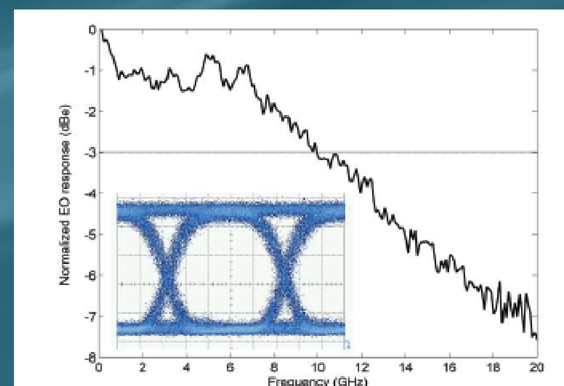
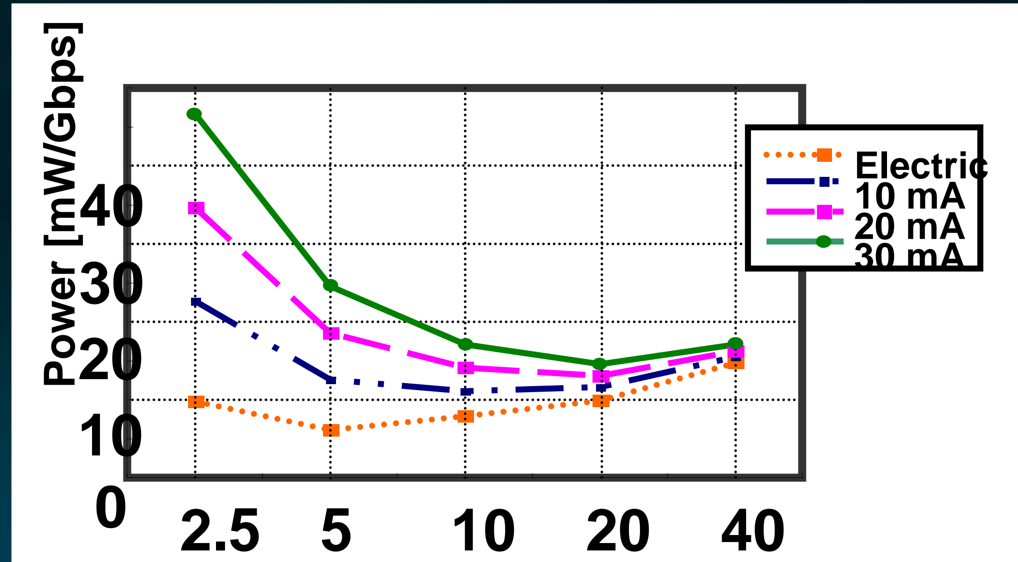


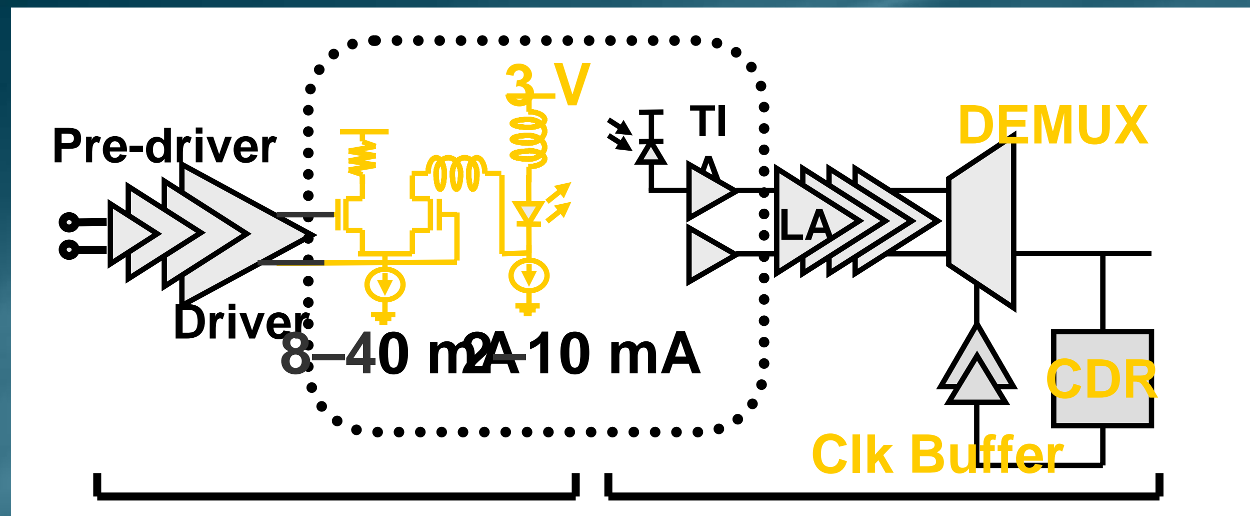
Figure 5. Silicon optical modulator 10Gb eye and frequency response.

IC Designs for Optical Communications

Biggest source of power dissipation is from laser



ISSCC2007 Wireline Forum,
Hirotaka Tamura, Fujitsu



Interconnect Fabric Conclusions

- Processor and system performance will be gated by chip-chip IO performance
- SI limitations will not scale with multi-core and multi-threaded processor performance
- Power dissipation will be traded off between processor core and IO interface to keep total power dissipation within bounds