

Tb/s Chip I/O - how close are we to practical reality?

Rick Walker

Hewlett-Packard Company

Palo Alto, California

walker@opus.hpl.hp.com

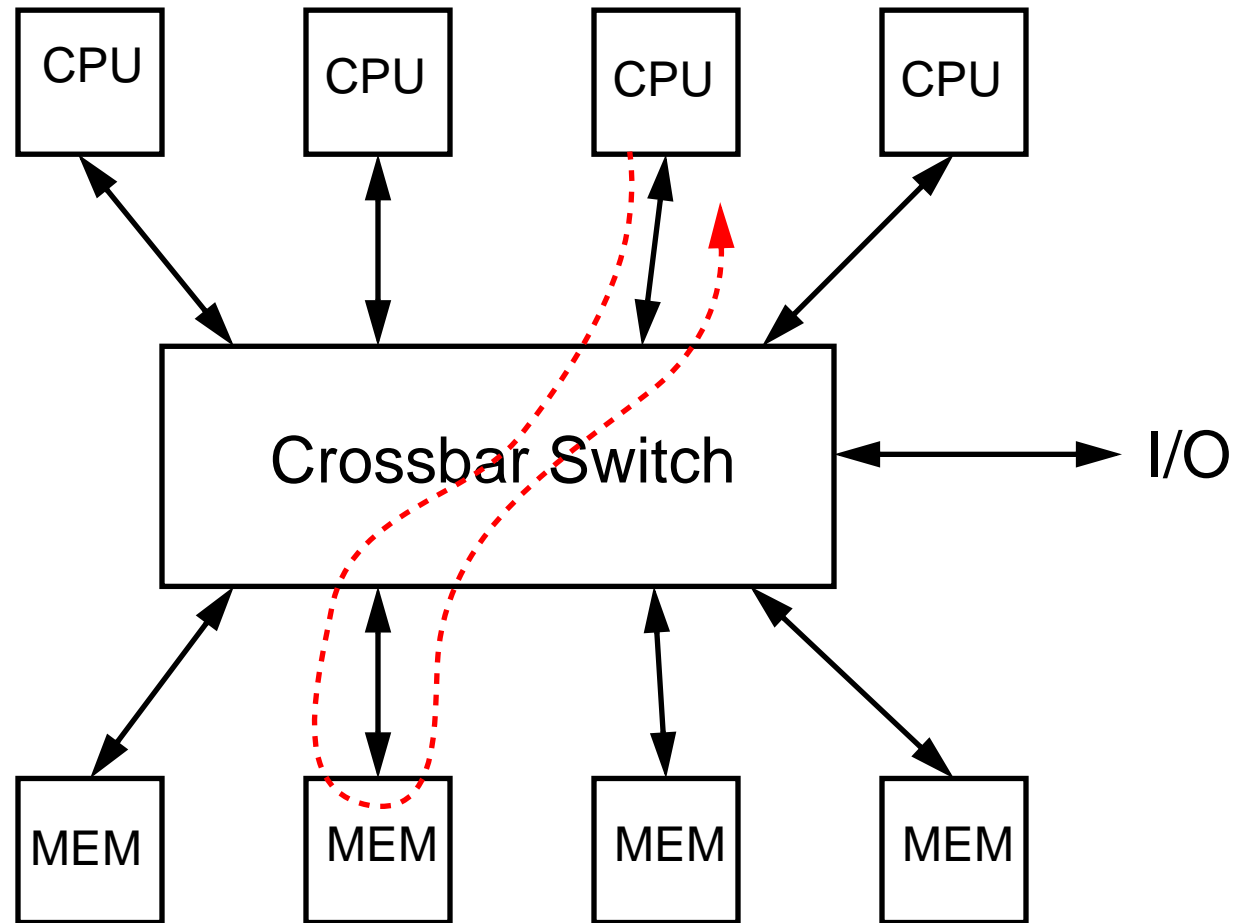
Agenda

- Applications and Key Specifications
- General Architecture for inter-chip communication
- Limitations
 - Skin-Loss
 - Delay Matching for Multi-phase sampling
 - CMOS Scaling
- Industry Trends
- Conclusions

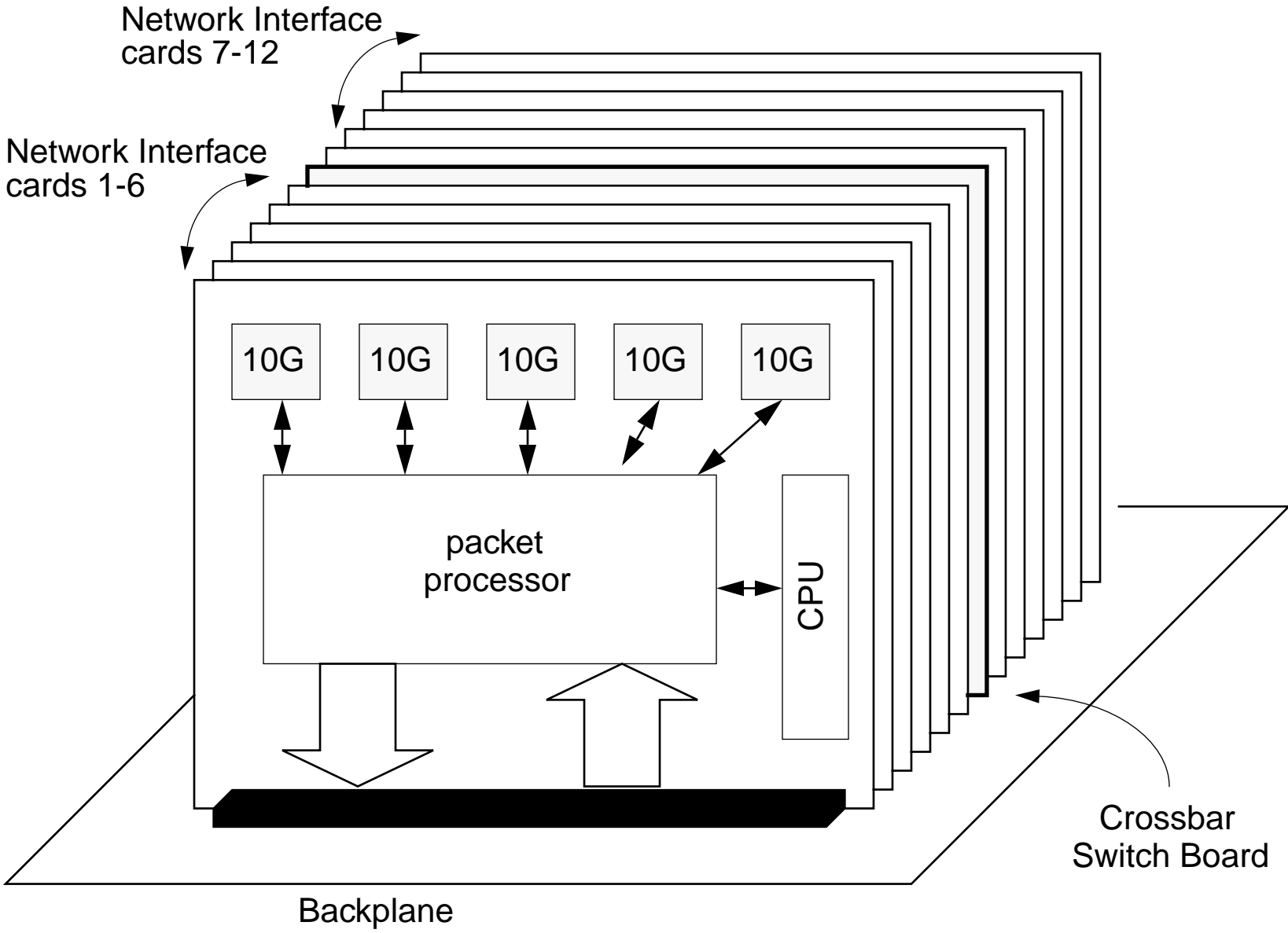
Current Practice

- Current high-performance systems are skew limited using parallel data clocked at 250-500Mb/s.
- Using clock and data recovery on Gb/s links eliminates the skew problem and improves system BW by factor of 8-16X.
- What are the limits for advanced systems?

CPU-CPU/Memory Application



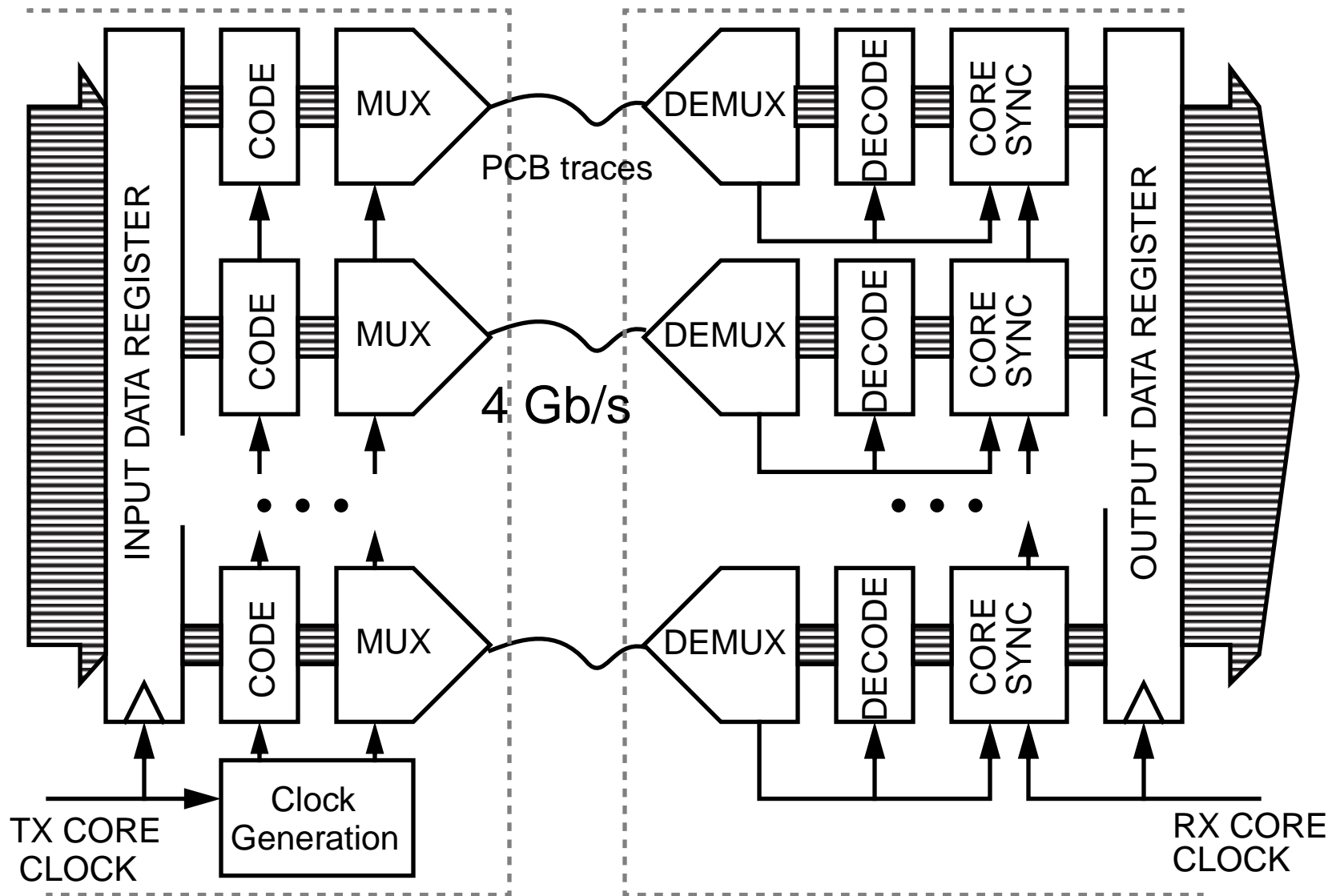
Router Application



Key Specifications

- Speed: As high as possible - at least 1Tb/s I/O per chip
- Latency: critical - less than 10ns plus time of flight
- BW/link: limited to 4-5 Gb/s by PCB loss
- Power: for a 100W chip, all 250 links should dissipate less than 40W -> 160 mW per link
- Size: a typical processor may be 9cm², if links use 20% of the total area, then each 4Gb/s link cell should be less than 720000um² in size.

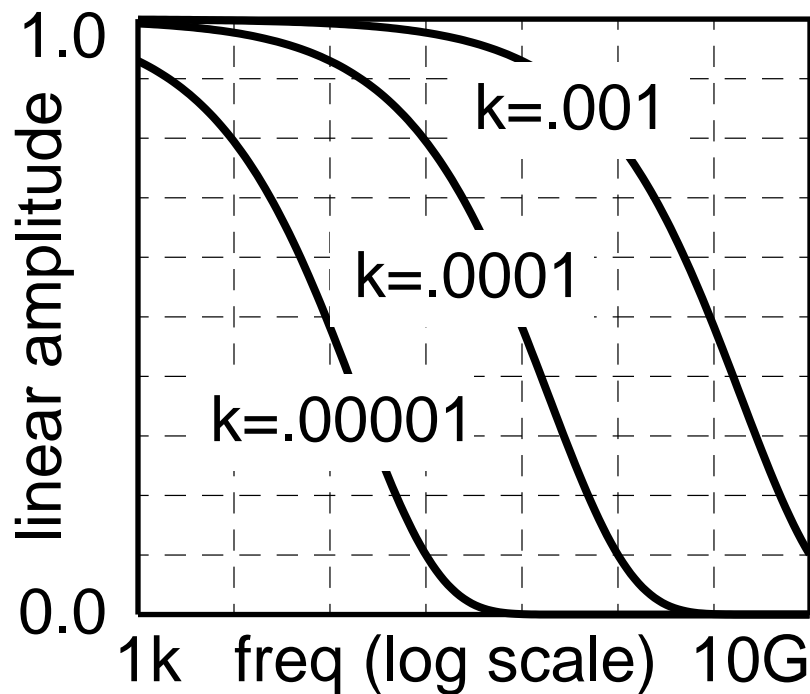
General Architecture



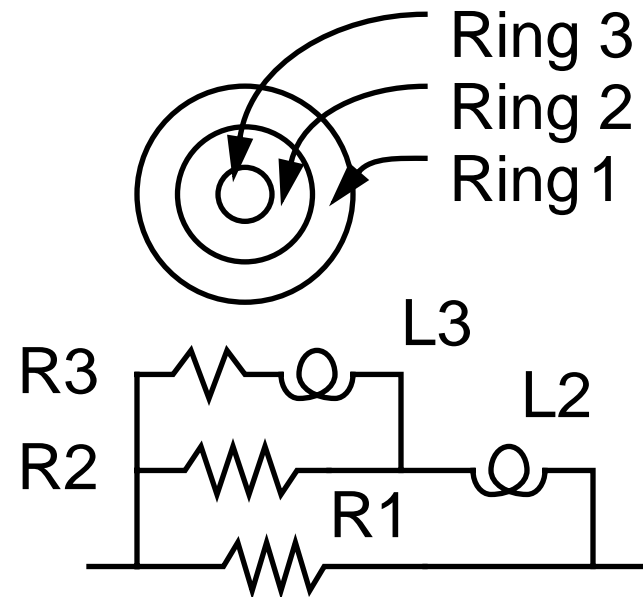
Skin Loss and Dielectric Loss

Nearly all cables are well modeled by a product of Skin Loss

$S(f) = e^{-k_s(1+j)l\sqrt{f}}$, and Dielectric Loss $D(f) = e^{-k_d lf}$ with appropriate k_s, k_d factors. Dielectric Loss dominates in the multi-GHz range. Both plot as straight lines on log(dB) vs log(f) graph.



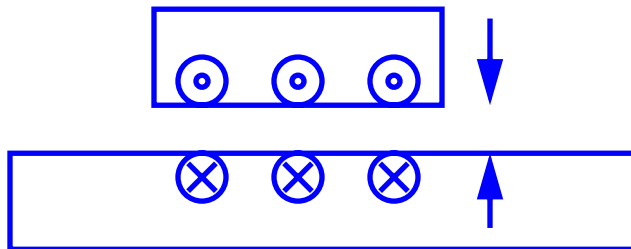
[YFW82]



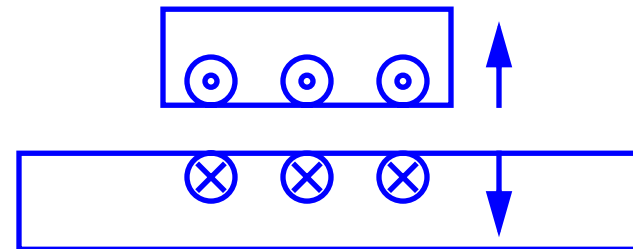
Three-element equivalent circuit of a conductor with skin loss

Skin Loss and Dielectric Loss

Two effects are operating in balance:



current filaments in conductor and ground attract each other

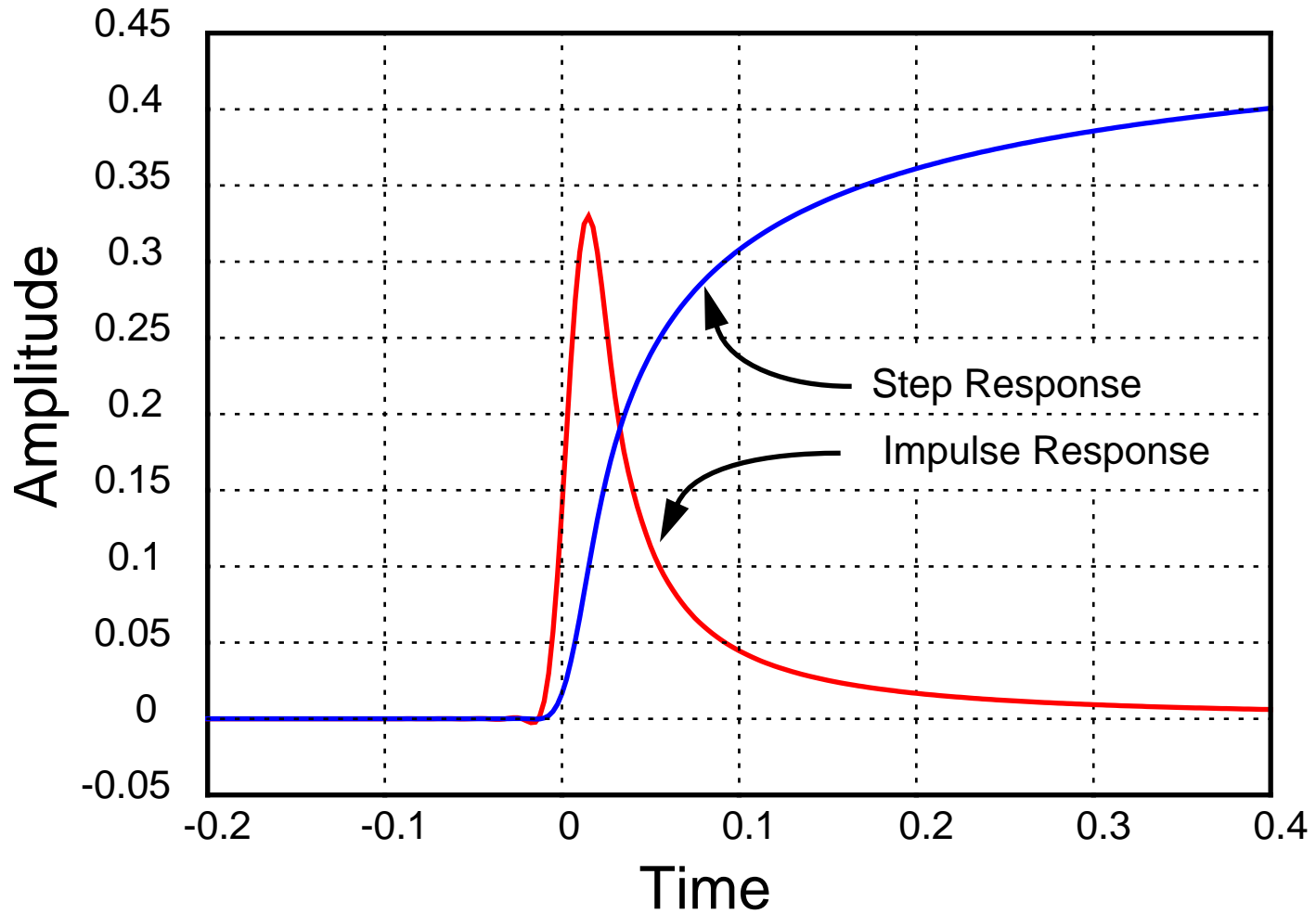


resistive drops tend to force current to spread out and flow evenly

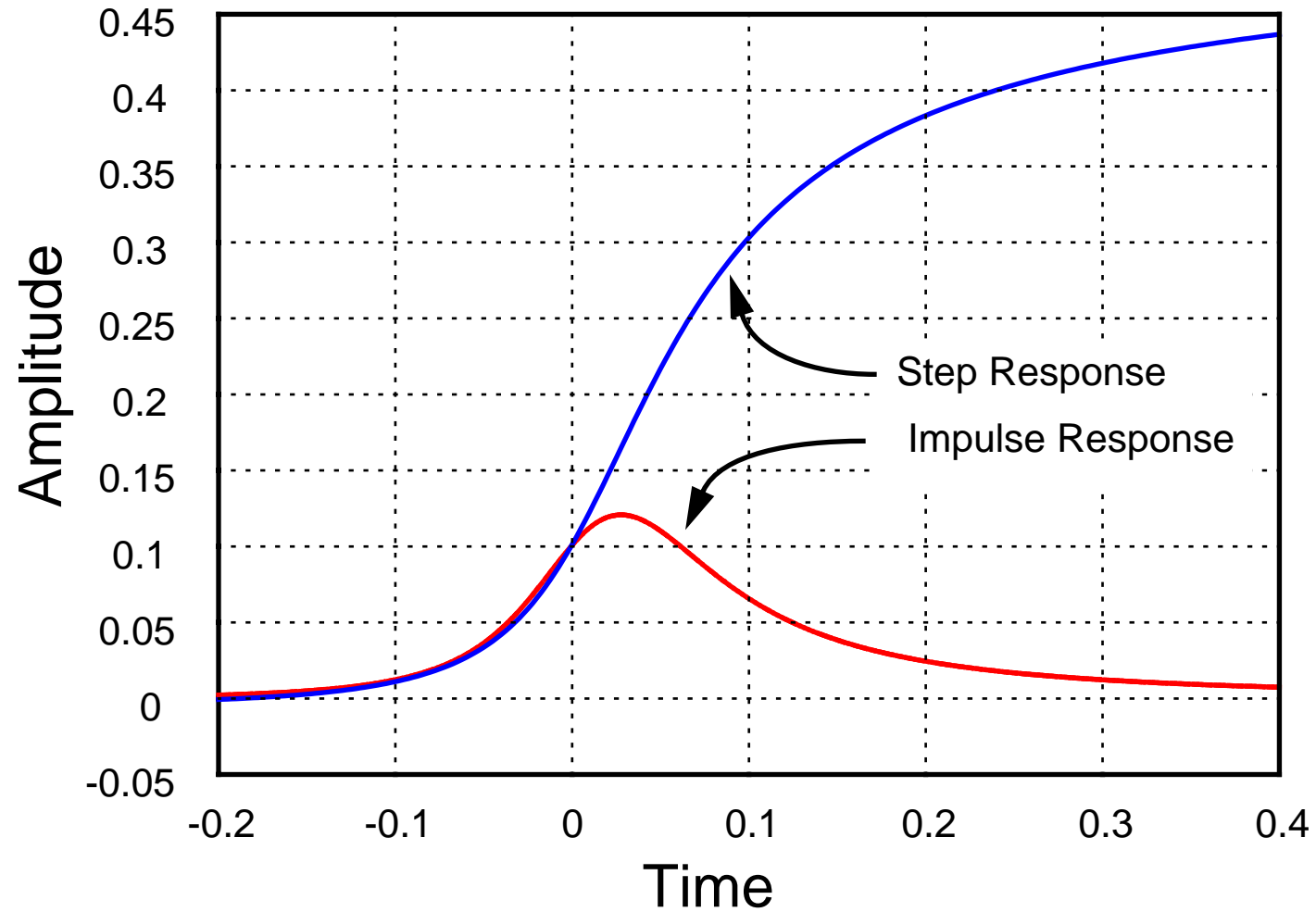
The net result is that current flows in a “skin” near the surface of the conductor towards the ground plane.

Current density falls off exponentially with depth and the depth is proportional to \sqrt{f} .

Skin Loss in the time domain

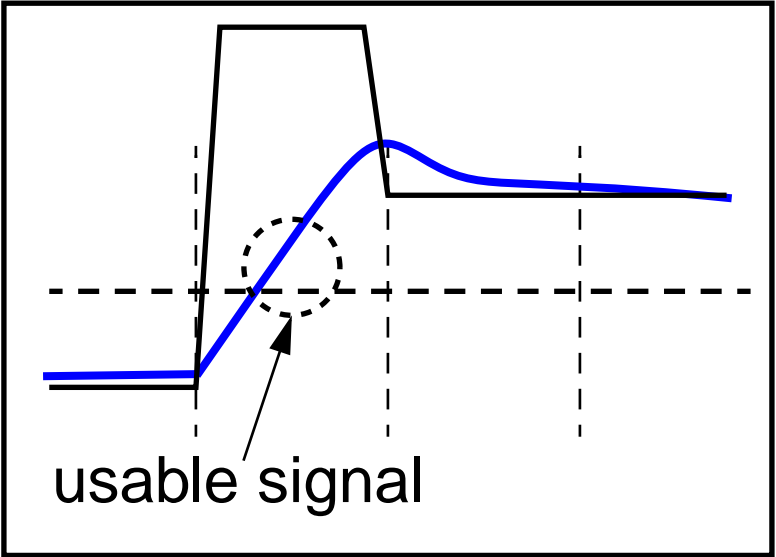
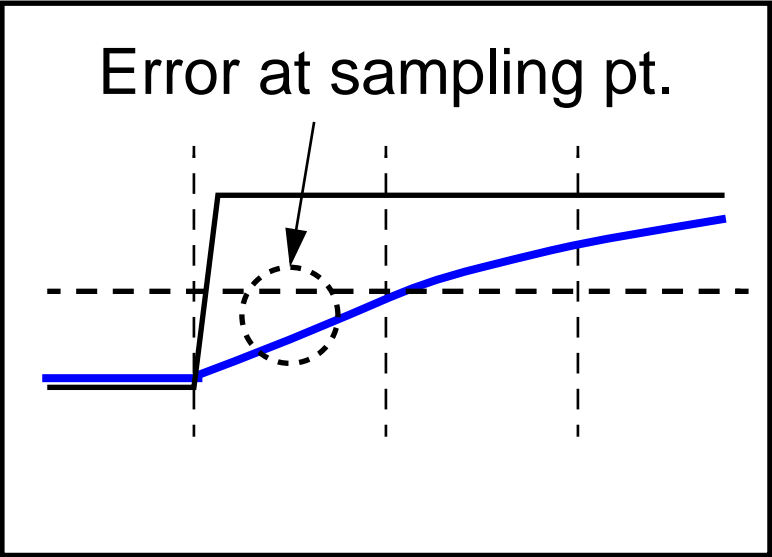


Dielectric Loss in the time domain

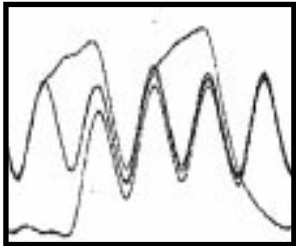


Skin Loss Equalization at Transmitter

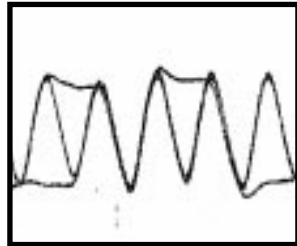
 boost the first pulse after every transition



[FMW97]

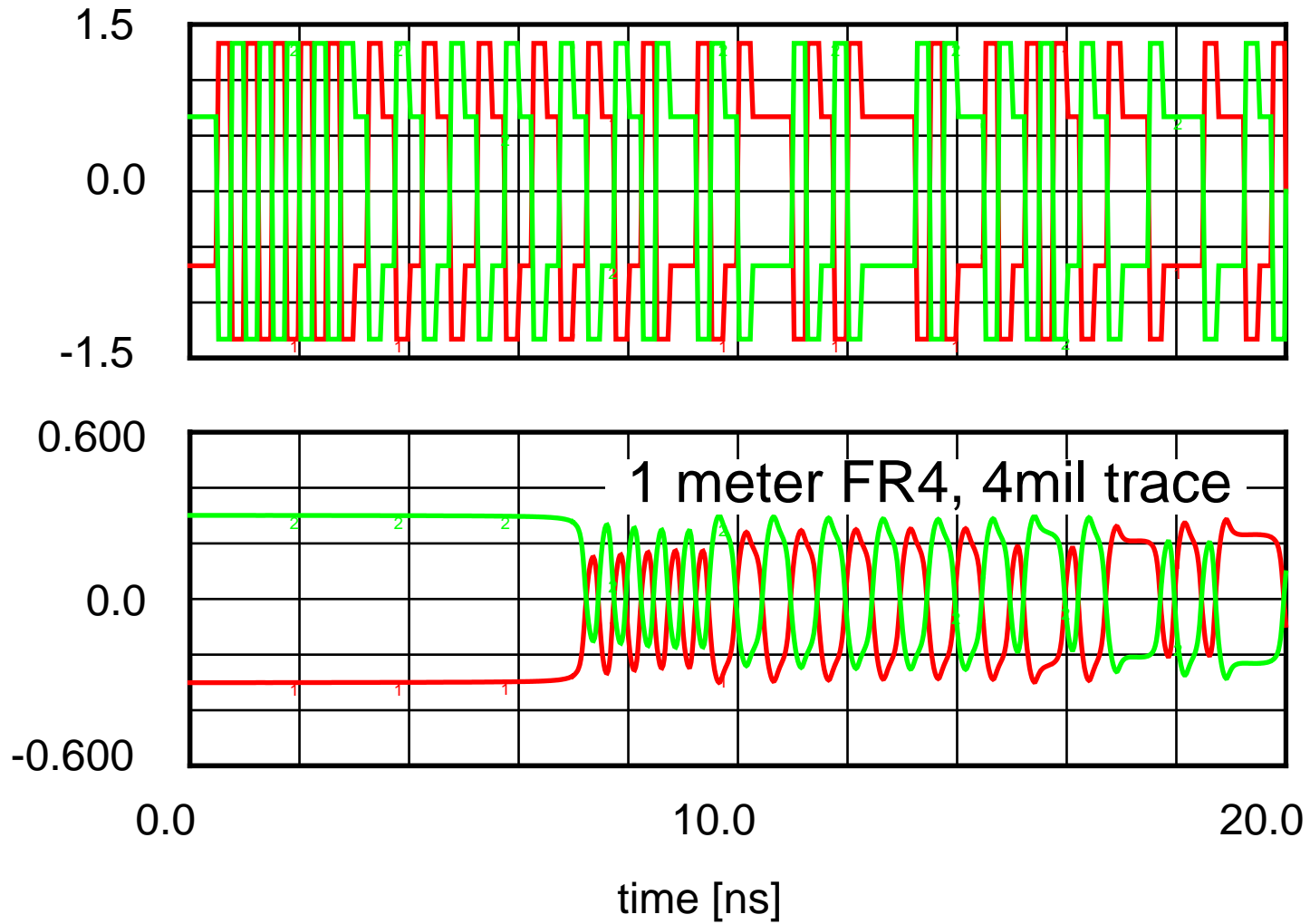


before

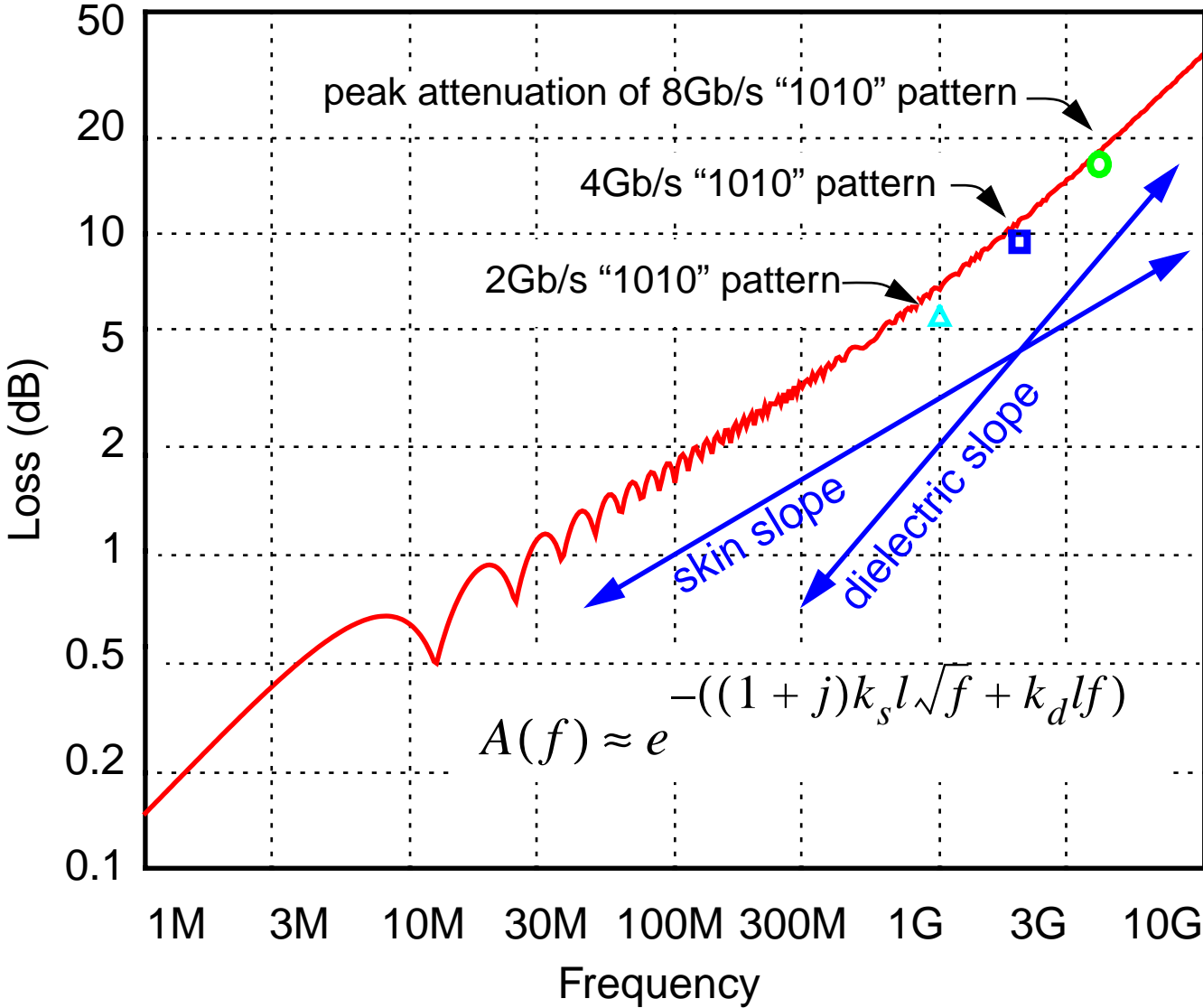


after

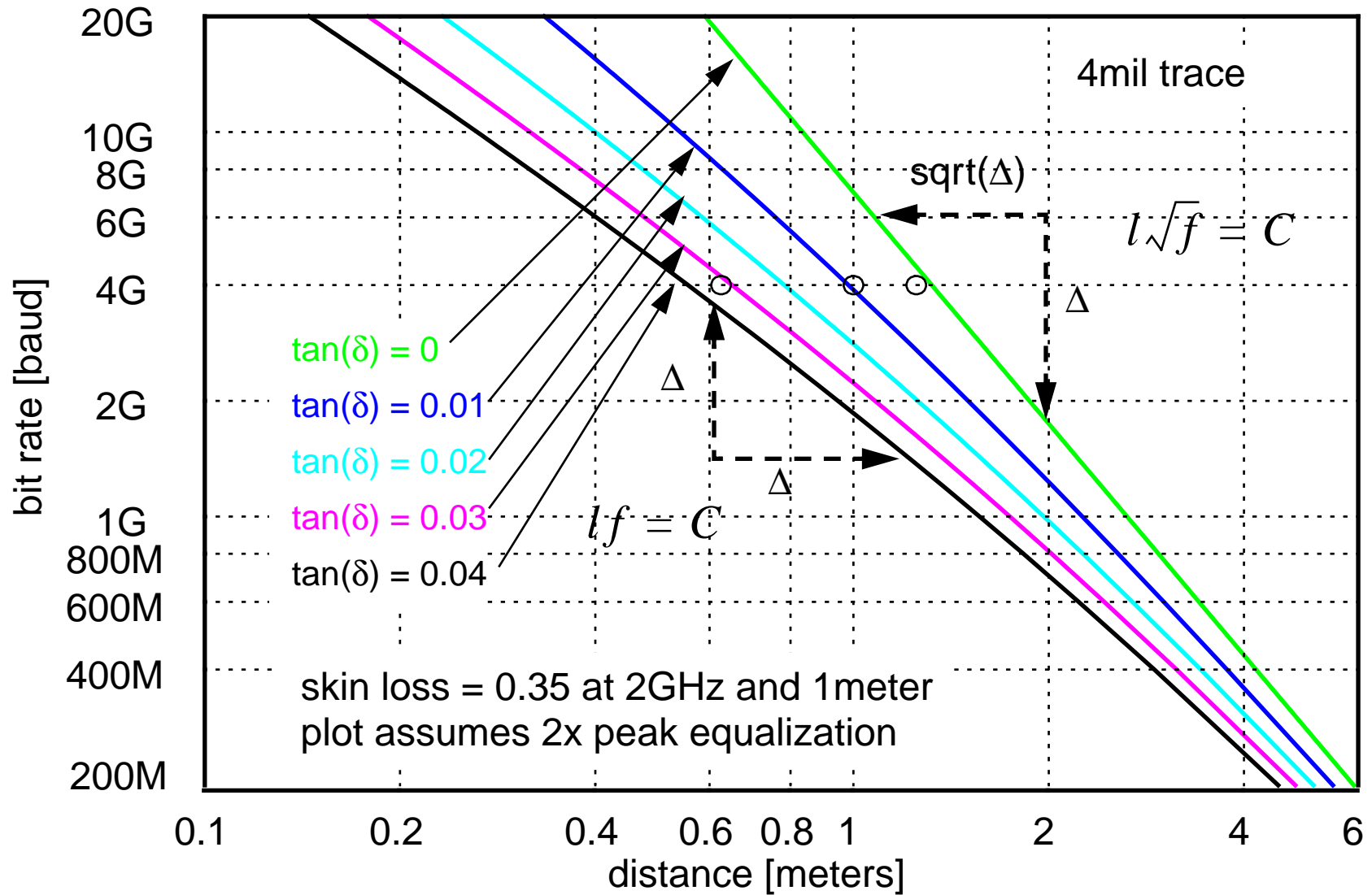
6dB Equalized Data



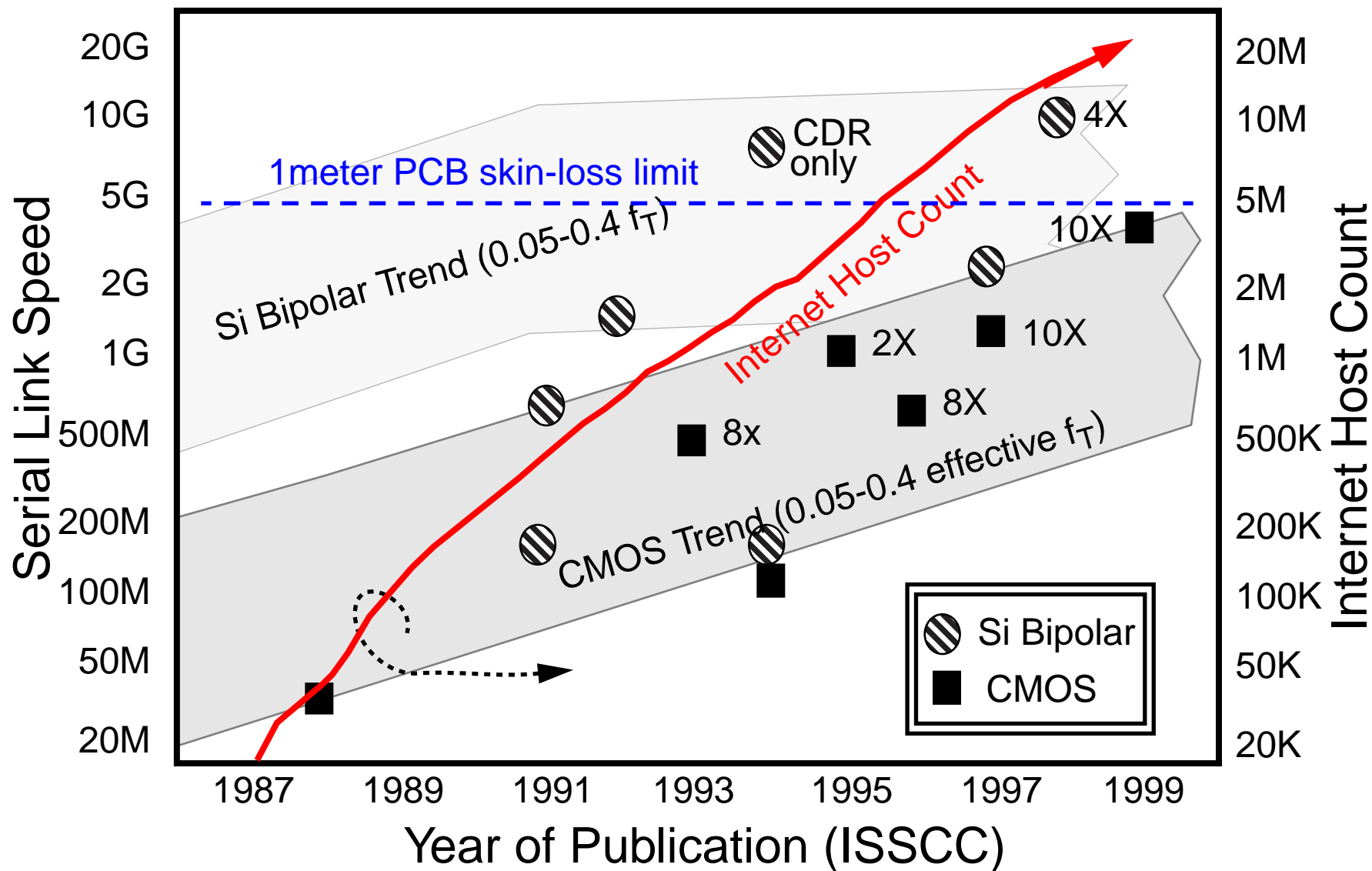
Skin Loss



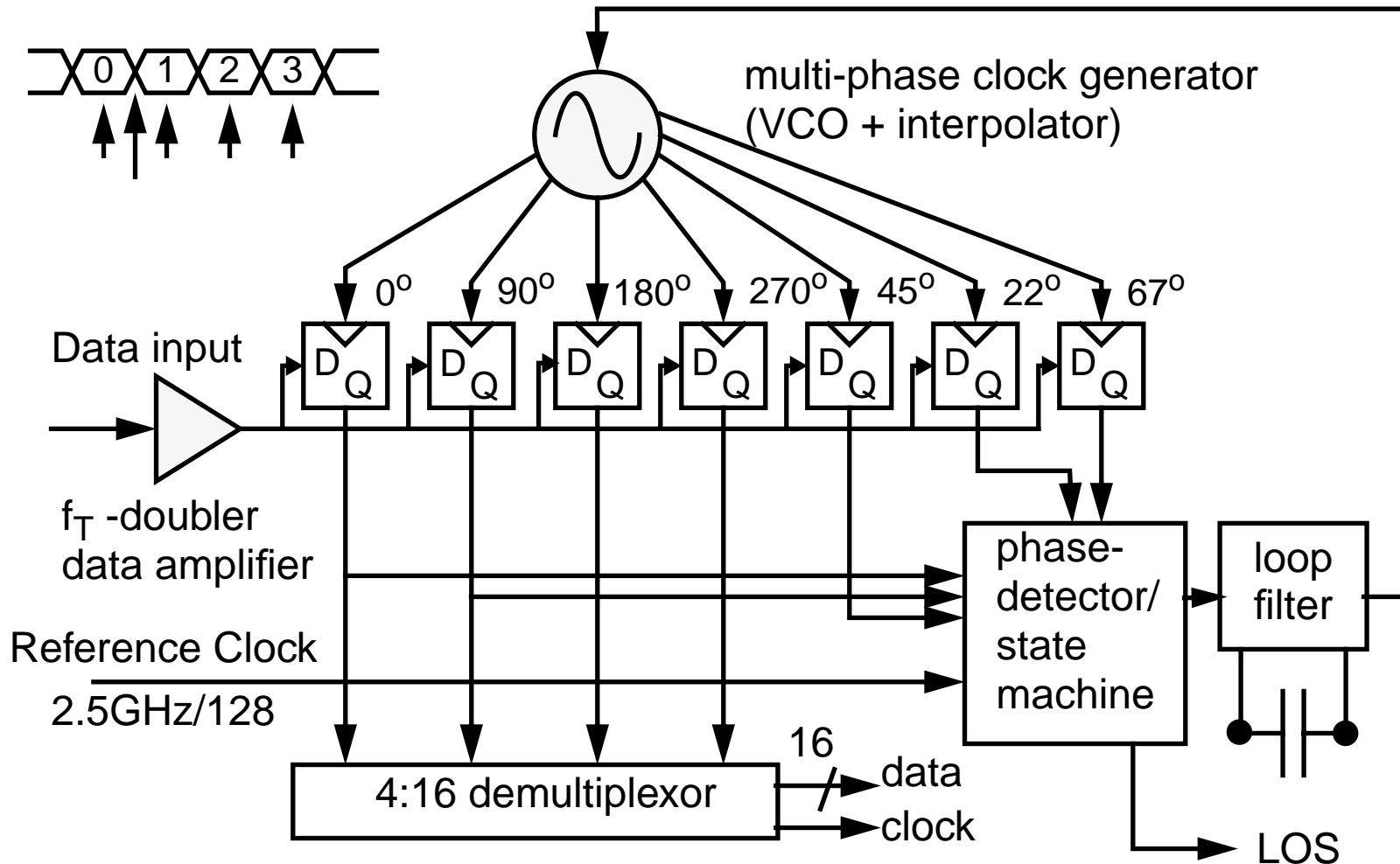
Data rate vs distance and $\tan(\delta)$



Communication Trends

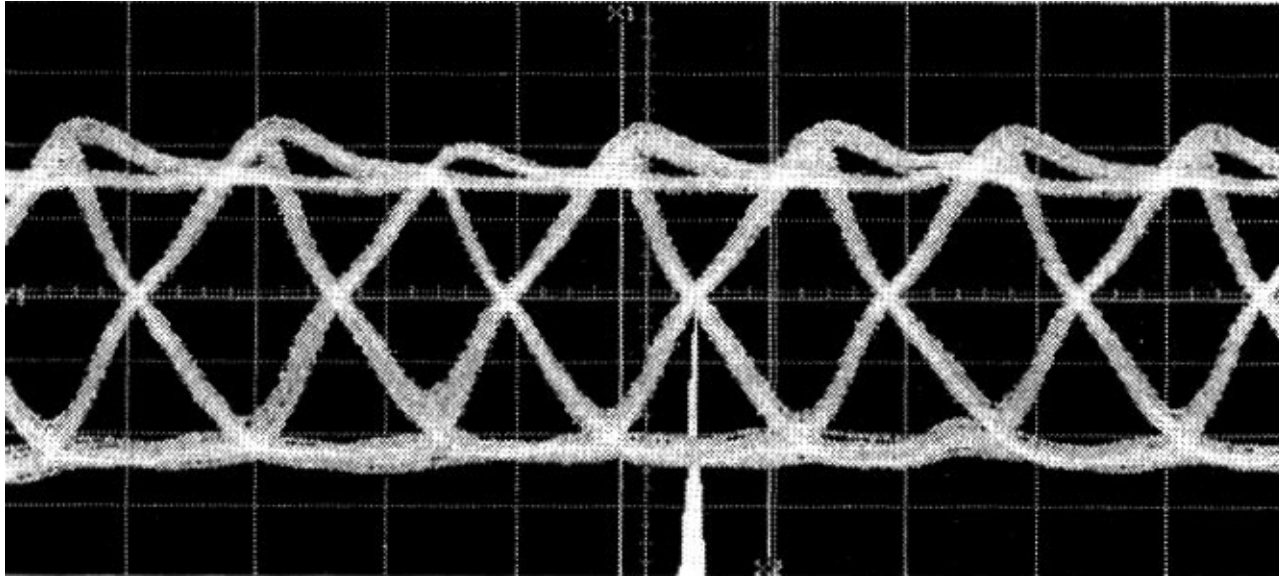


Example Multiphase RX Block Diagram



[WHK98]

Measurement of a Multi-phase System

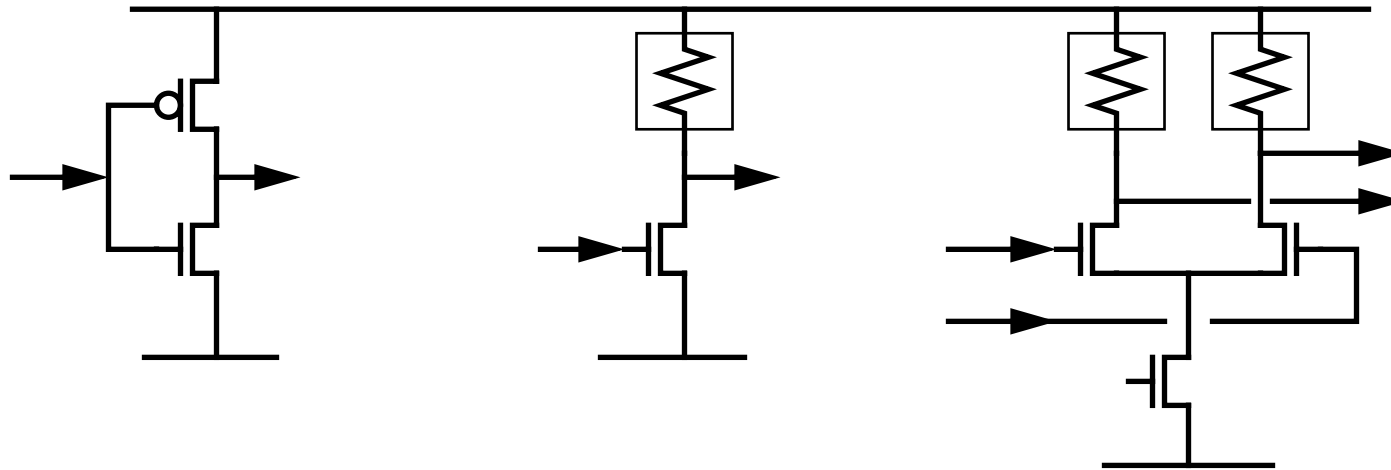


Reported Jitter: 8ps rms, 44ps pk-pk at 3.5Gb/s.

Measurement of photo shows 26ps difference between widest and narrowest eye, so true eye margin for end-end system is $44ps\sqrt{2} + 2 \cdot 26ps = 118ps$, or a total eye closure of 41%.

Attention to delay *matching* is critical!

Techniques to Improve Delay Matching and Power Supply Noise Immunity



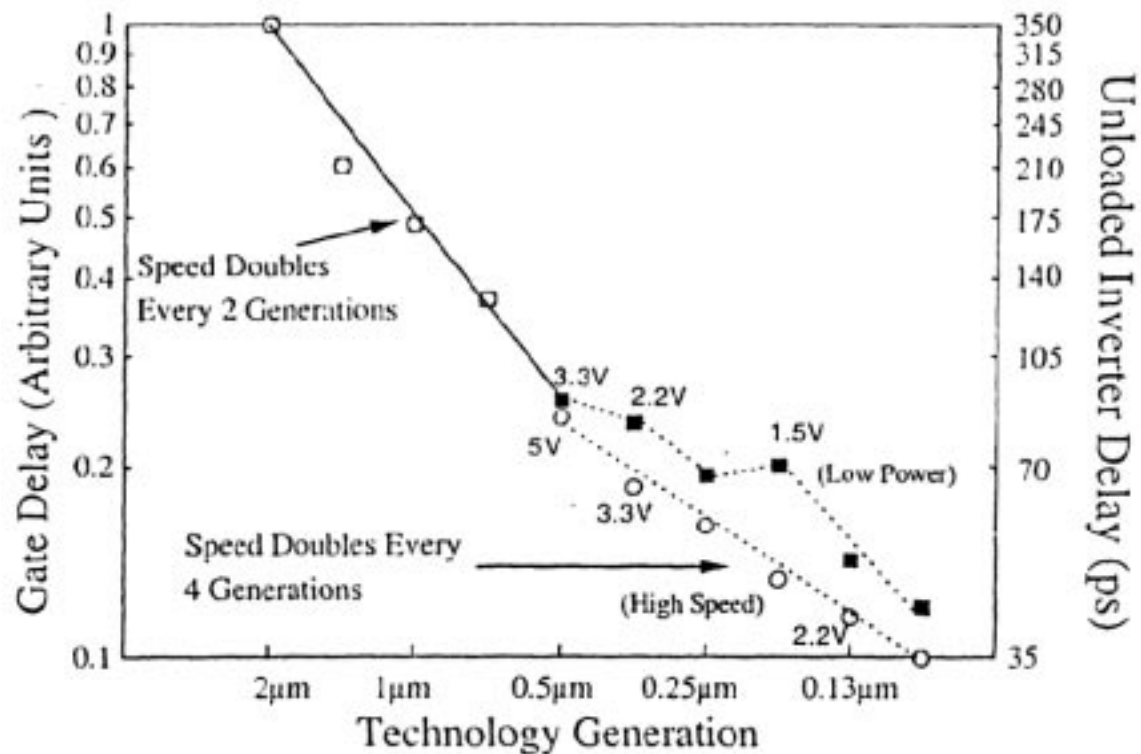
- Rising edge delay dependent on *absolute* V_T
- RC time-constant changes with supply voltage

- Rising edge delay dependent on *absolute* V_T

- Delay depends on *VT matching*
- Current source absorbs supply voltage changes

CMOS Scaling Issues

- Gate delay no longer scales with process

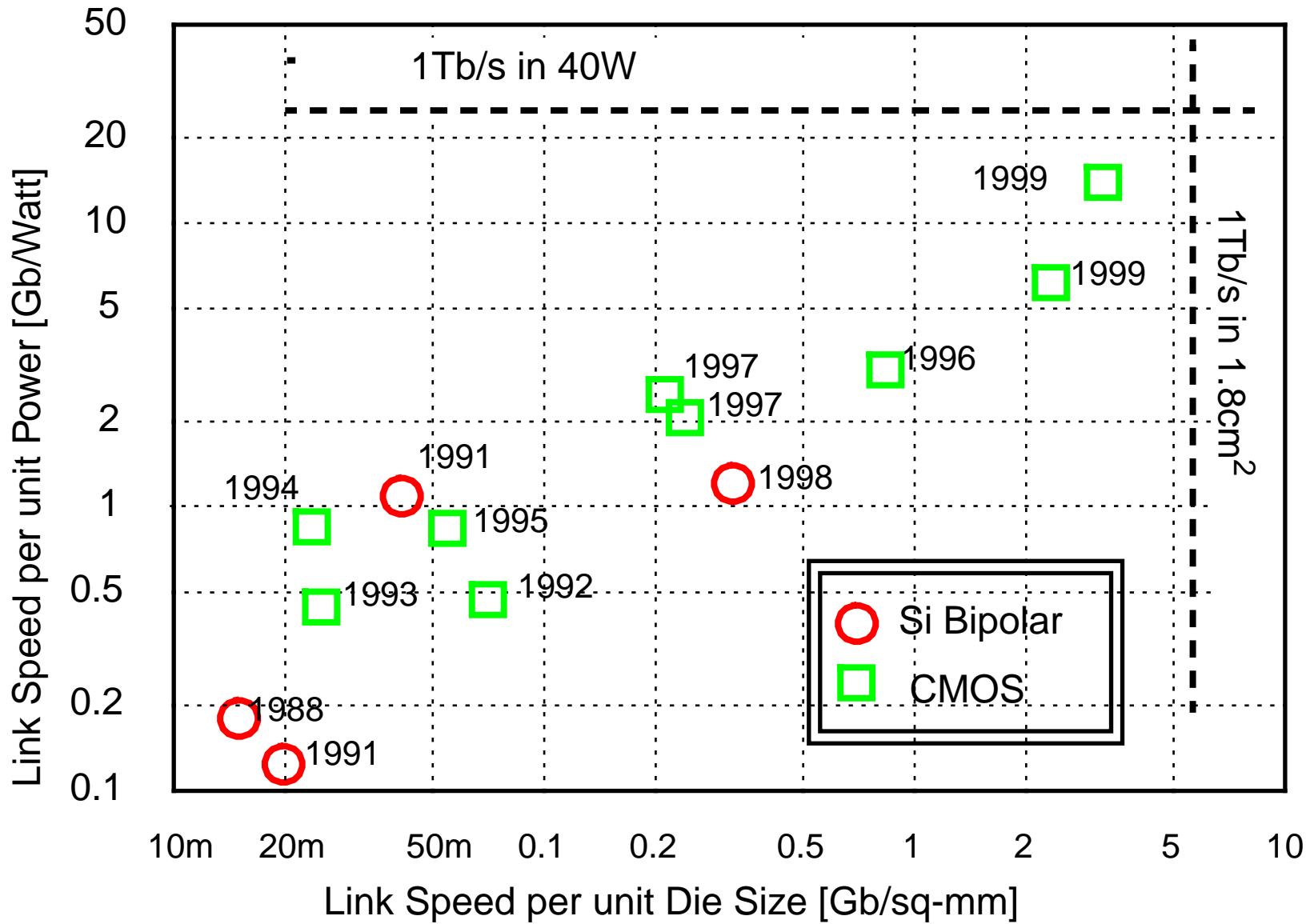


See: Chenming Hu, "Low-Voltage CMOS Device Scaling" 1994 ISSCC Digest, pp 86-87.

CMOS Scaling Issues (continued)

- V_t doesn't track with power supply - so we gradually lose ability to make ECL-like differential circuits.
- Full-swing circuits show worse delay matching than ECL-like topologies.
- Full-swing circuits show worse power-supply delay modulation than differential circuits.
- V_t matching gets worse due to statistical dopant variations in channel.
 - All of these trends make power supply noise rejection and multi-phase alignment more difficult with each process scaling.

Power and die size vs target



Industry Trends

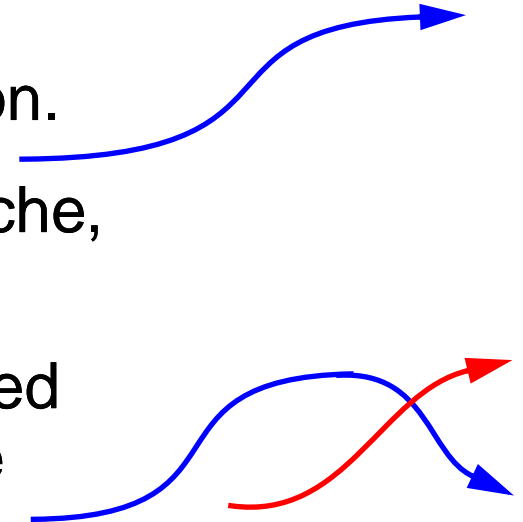
- 50% of all U.S. Families now have home computers
- Computer performance has surpassed needs of most users: witness the drop of P.C. prices in the last 3 years from a stable \$2K down to \$500 levels.
- Internet host count was doubling every 6 months in 1988, is now doubling every 24 months - we are clearly past the 50% adoption point in the growth curve.
- What applications will continue to drive expensive and exotic improvements in interconnect technology?
 - Without a new “killer app” to drive development, we may be stuck with the limitations of FR4/CMOS for quite some time.

Viability of “exotic” technologies

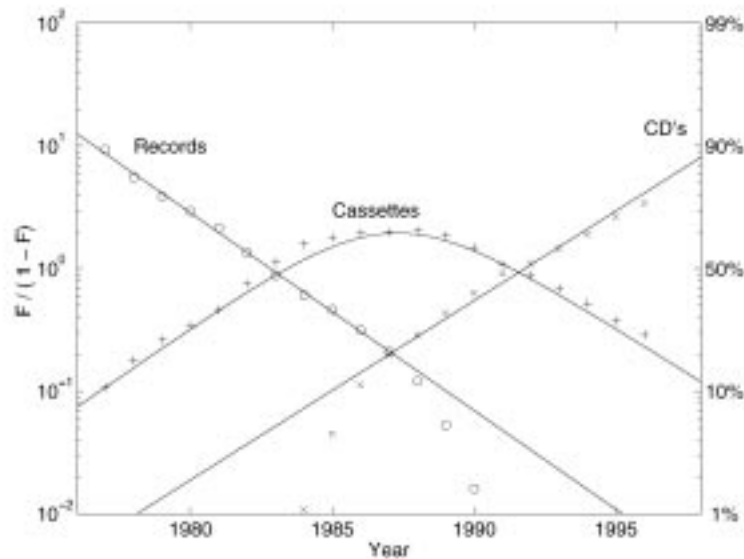
- Yielded CMOS parts come in at \$10/cm²
- Tb/s chip-chip links are probably feasible in the next few years.
- This performance can be achieved with existing BGA packages across commodity FR-4 PC Backplanes.
- The incremental cost of a Tb/s link in these applications will be about \$18 + connector cost.
 - For optical solutions to take hold in these applications, they must provide either significantly higher performance (>10Tb/s) or cheaper system cost (not likely).

Logistic growth law

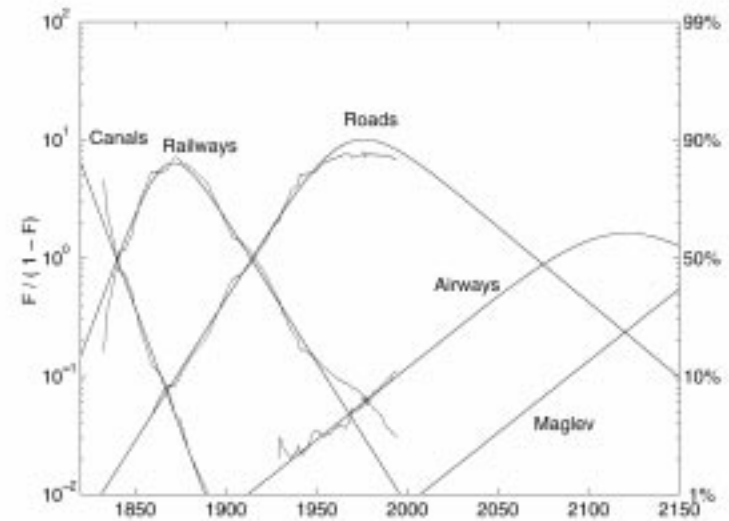
- Natural systems grow in a sigmoidal fashion.
- Early in the colonization of an ecological niche, the growth looks exponential.
- In the mature phase, steady state is reached until a competitor arises to outcompete the incumbent.



Meyer, Yung, & Assubel, *Technological Forecasting & Social Change* 61(3):247-271, 1999



Meyer, Yung, & Assubel, *Technological Forecasting & Social Change* 61(3):247-271, 1999



Conclusions

- Still much work to be done, but 1 Tb/s chip I/O seems an attainable target.
- 5Gb/s on 1meter PCB is the fastest that can be feasibly supported for the foreseeable future with *low latency*.
- Fiber seems to be progressing along either a 1-10-100-1000-10,000MHz or a 622-2488-10,000MHz evolutionary path. There may be an economically important need for 5Gb/s links.
- 10 Tb/s chip I/O is probably out of the question for current high-volume technologies (CMOS, FR-4 PCB). Computer designs and programs may have to give up cache coherency, and move towards cooperative computing architectures to break out of this limitation.