ECE 546
Introduction

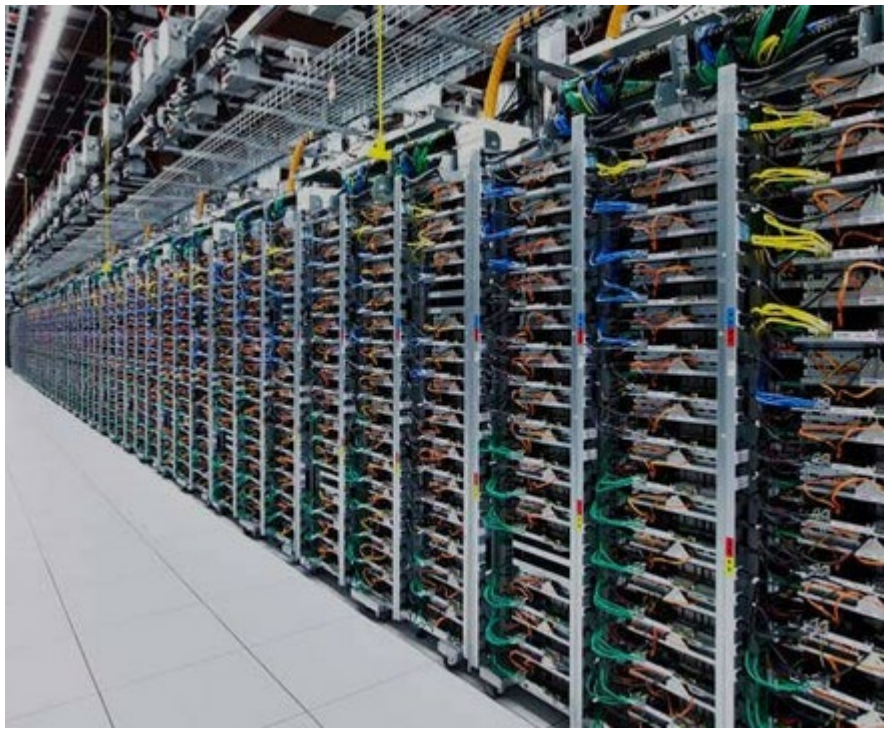
Spring 2026

Jose E. Schutt-Aine
Electrical & Computer Engineering
University of Illinois
jesa@illinois.edu

AI Energy Requirements

- **Data centers used 2.5% of US electricity in 2022**
- **Projected to increase to 20% by 2030**

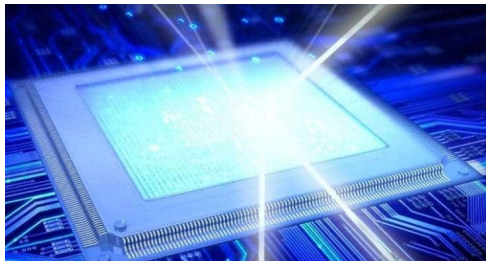
Training a large language model like **GPT-3** is estimated to use **1,300 megawatt hours (MWh)** of electricity.



Energy Solutions

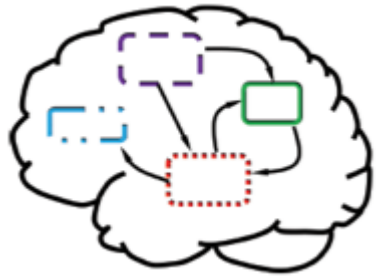
Renewable

Co-Packaged Optics



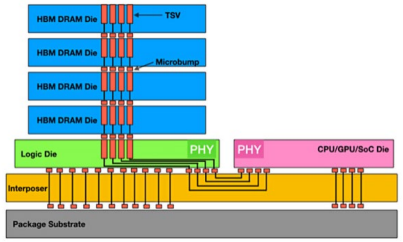
Algorithms

```
//Package level
for p3 = [0 : P3) :
  for q3 = [0 : Q3) :
    parallel_for k3 = [0 : K3) :
      parallel_for c3 = [0 : C3) :
// Chiplet level
for p2 = [0 : P2) :
  for q2 = [0 : Q2) :
    parallel_for k2 = [0 : K2) :
      parallel_for c2 = [0 : C2) :
// PE level
for r = [0 : R) :
  for s = [0 : S) :
    for k1 = [0 : K1) :
      for c1 = [0 : C1) :
        for p1 = [0 : P1) :
          for q1 = [0 : Q1) :
// Vector-MAC level
parallel_for k0 = [0 : K0) :
  parallel_for c0 = [0 : C0) :
    p = (p3 * P2 + p2) * P1 + p1;
    q = (q3 * Q2 + q2) * Q1 + q1;
    k = ((k3 * K2 + k2) * K1 + k1) * K0 + k0;
    c = ((c3 * C2 + c2) * C1 + c1) * C0 + c0;
    OA[p,q,k] += IA[p-1+r,q-1+s,c] * W[r,s,c,k];
```



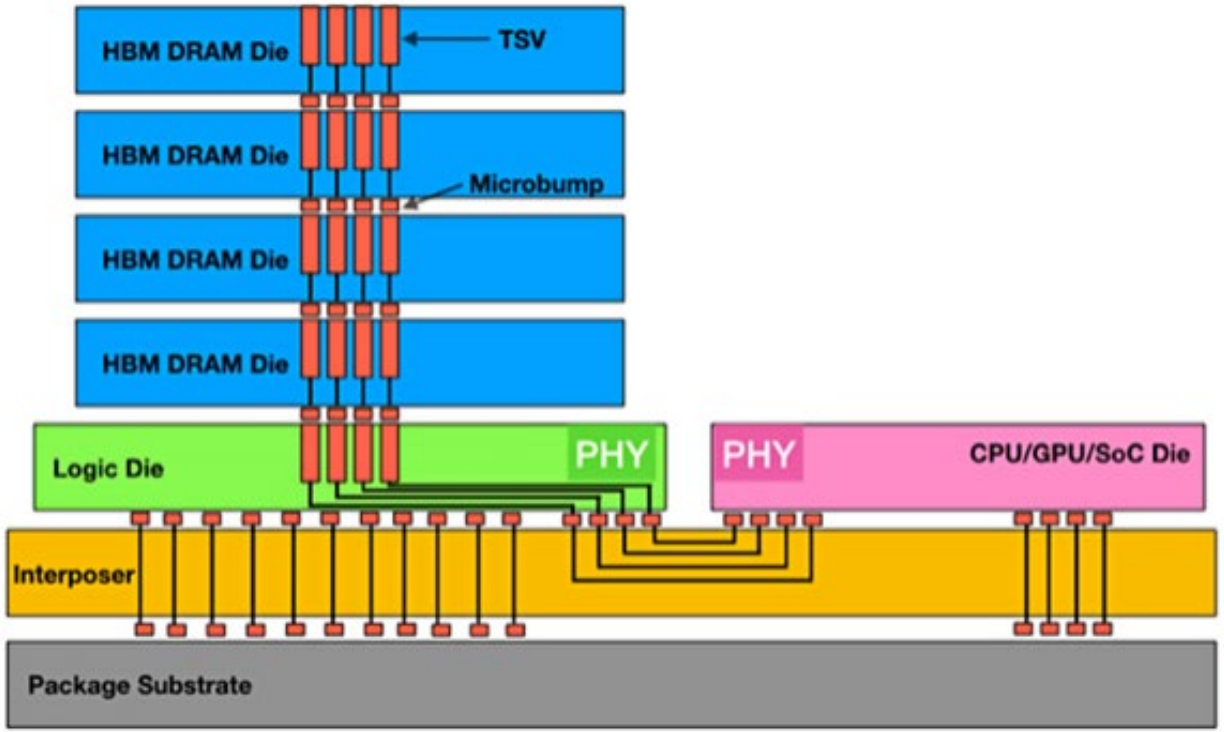
Neuromorphic

HI/Advanced packaging



HBM stack for maximum data throughput. Source: Rambus

Solution via Advanced Packaging

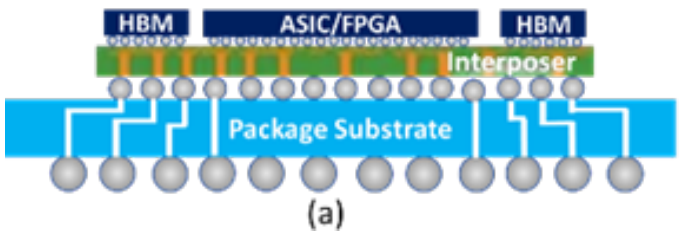


HBM stack for maximum data throughput. Source: Rambus

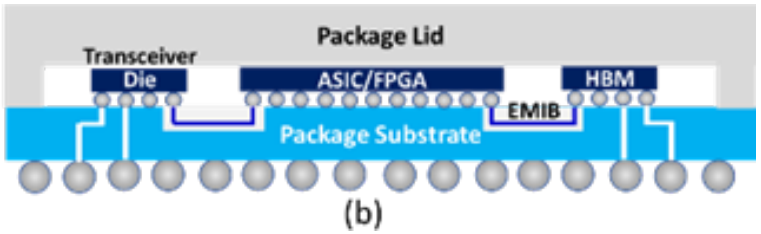
80% of power consumption is due to data movement through interconnects

The Interconnect Challenge

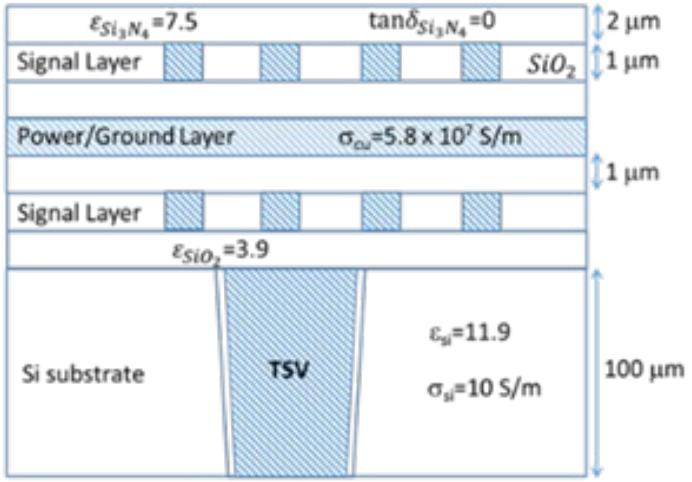
"There are many solutions and techniques for improving transistors; options for improving interconnects are very few…"



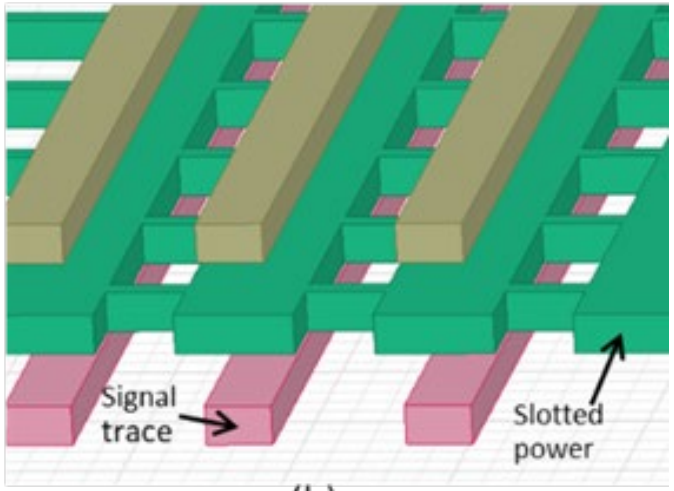
(a)



(b)



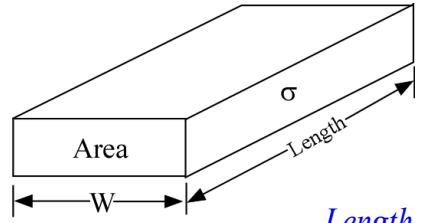
(a)



(b)

Challenges

- Signal/power integrity
- Thermal effects
- IR Drop
- Power dissipation
- High I/O count
- Reliability



Resistance: R $R = \frac{Length}{\sigma \cdot Area}$

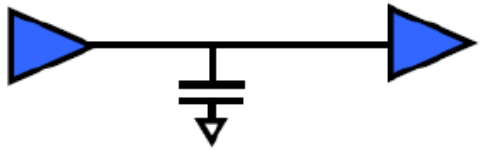
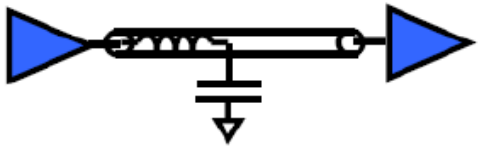
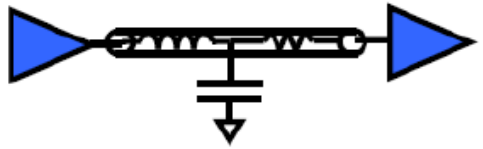
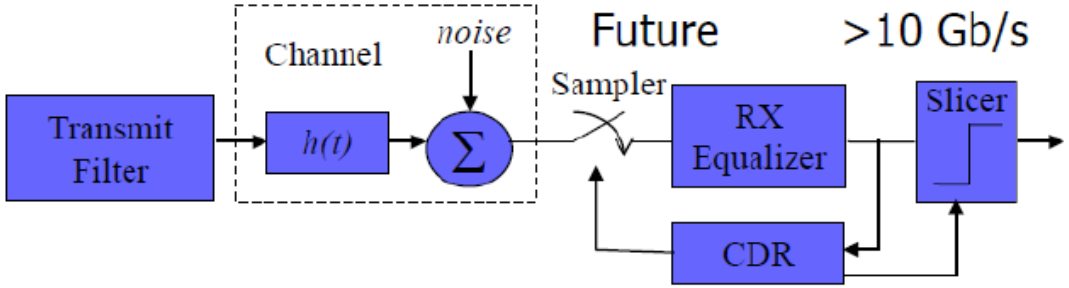
Resistance per unit length

Package level
W = 3 mils
R = 0.0045 Ω/mm

Chip-level
W = 0.25 microns
R = 422 Ω/mm

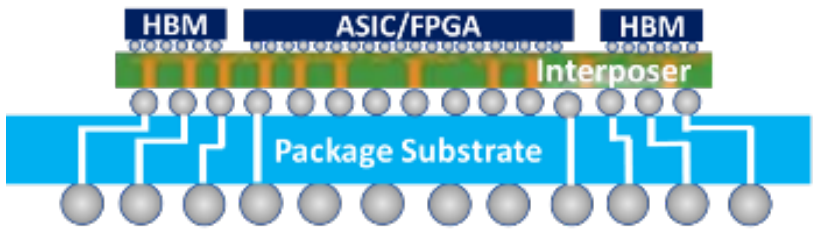
$Power = RI^2$

Interconnect Evolution

	Decade	Speeds	Transceiver Features
Lumped capacitance 	1980's	>10Mb/s	Inverter out, inverter in
Transmission line 	1990's	>100Mb/s	Termination Source-synchronous clk.
Lossy transmission line 	2000's	>1 Gb/s	Pt-to-pt serial streams Pre-emphasis equalization
	Future	>10 Gb/s	Adaptive Equalization, Advanced low power clk. Alternate channel materials

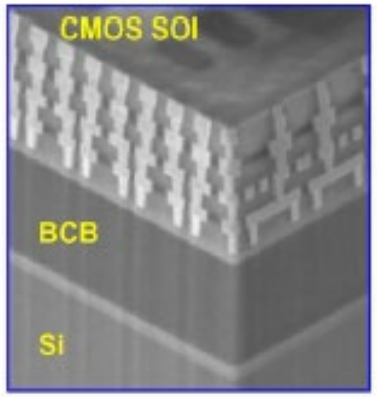
Slide Courtesy of Frank O'Mahony & Brian Casper, Intel

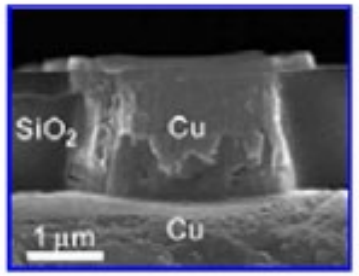
Advanced Packaging

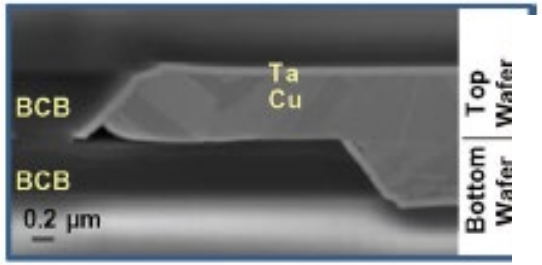


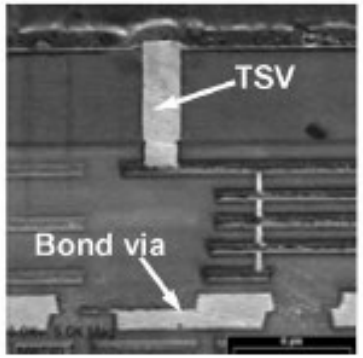
(a)

(b)

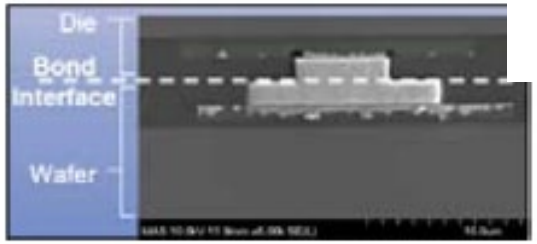


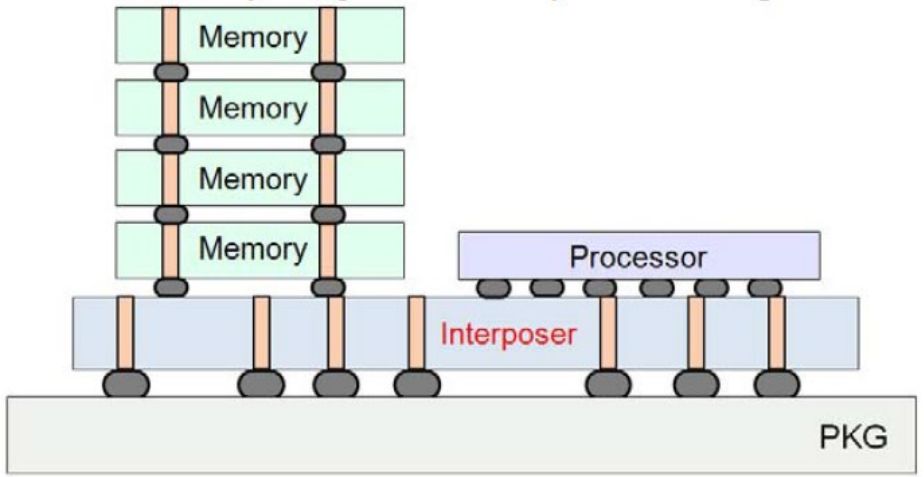






(a) (c) (e)
(b) (d) (f)

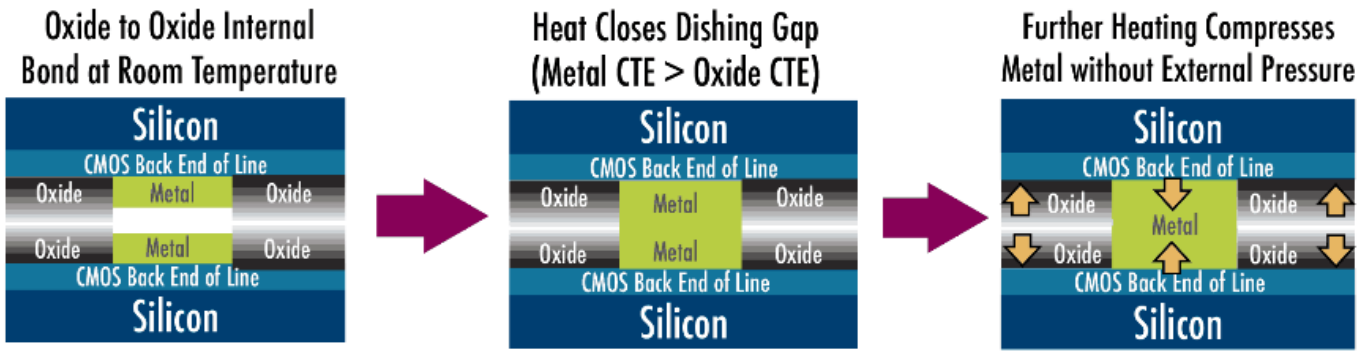




In order to meet the demands of AI, HPC, and next-generation data centers, advanced packaging will be key to performance, power efficiency, and scalability.

Hybrid Bonding

simultaneous bonding of dielectric and metal bond pads in one bonding step



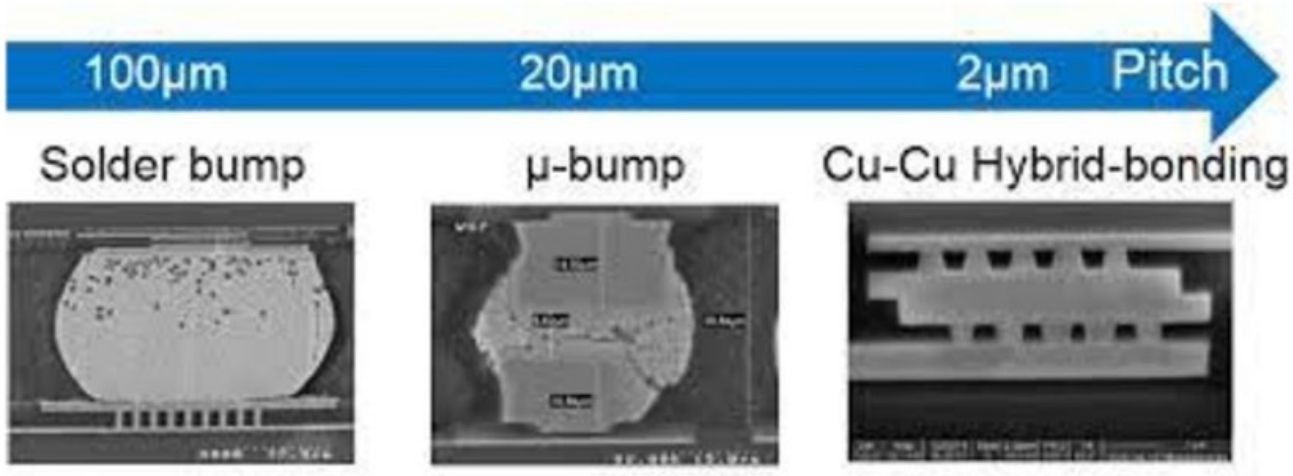
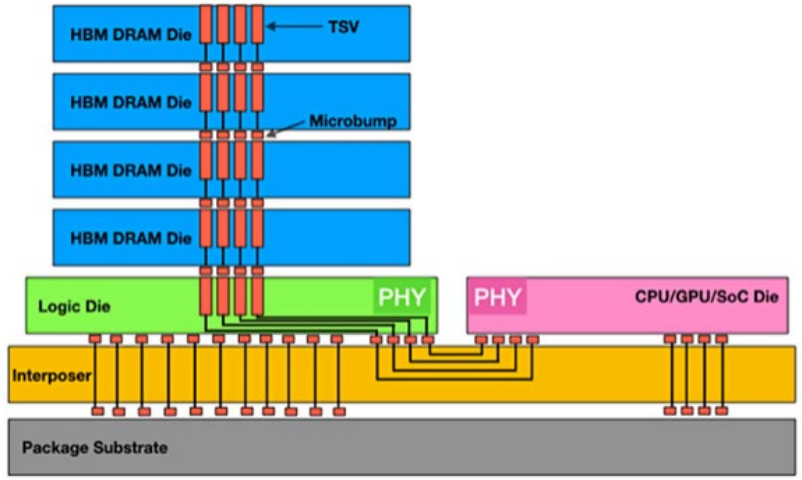


Image Credit: Imed Jani. Test and characterization of 3D high-density interconnects. Micro and Nanotechnolo- gies/Microelectronics. Université Grenoble Alpes, 2019. English. NNT : 2019GREAT094 . tel- 02634259



HBM stack for maximum data throughput. Source: Rambus

- Allows advanced 3D device stacking
- Highest I/O
- Enables sub-10-µm bonding pitch
- Higher memory density
- Expanded bandwidth
- Increased power
- Improved speed efficiency
- Eliminates the need for bumps, improving performance with no power or signal penalties

Heterogeneous Integration

- Heterogeneous Integration is defined as the integration of separately manufactured components into a higher-level assembly (Chiplets, SiPs, Modules) that, in the aggregate, provides enhanced functionality and improved operating characteristics

- The size, complexity and stochasticity associated with future advanced packaging platforms will render their design intractable using the traditional approaches. In addition, such design will need to be performed with consideration for thermal management, reliability, architecture and floorplanning/placement and routing constraints

Need for: Heterogeneous Integration

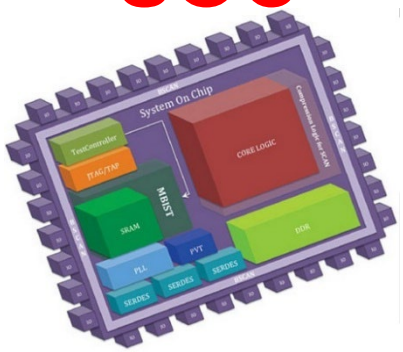
- Motivation: Ubiquitous materials and technologies
  - Die, interposer, substrate
  - Scaling is with wavelength rather than technology
- Opportunity: New interconnect technologies
  - Increased density of I/Os, finer pitches
  - 2.5D/3D integration
- Performance objective: minimize energy and delay
- Approach: Use chiplets

Need for: Chiplets

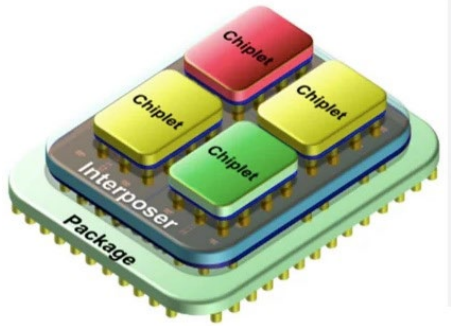
- Advantages
  - Better cost per die, higher yield, mix and match, reusability
- Challenges
  - Disaggregation is ad-hoc
  - Multi-level, multi-scale, increased complexity

Solution: Co-Design

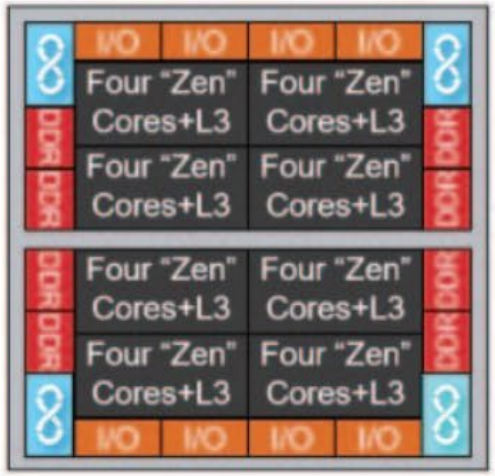
SOC



SiP

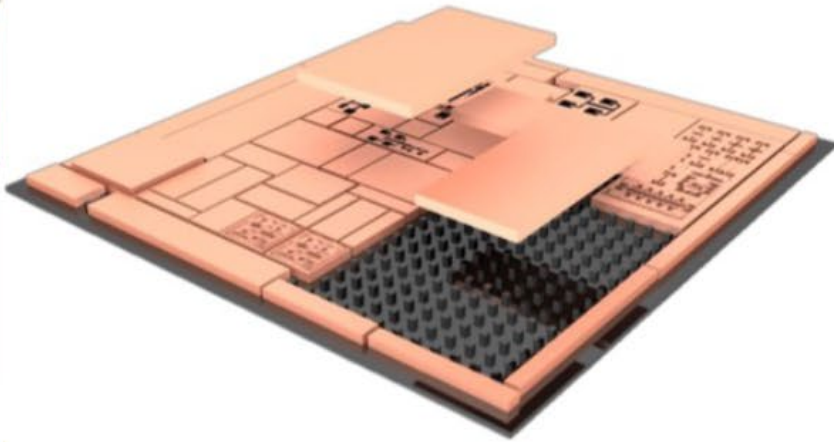


Chiplet-Based Design: AMD EPYC



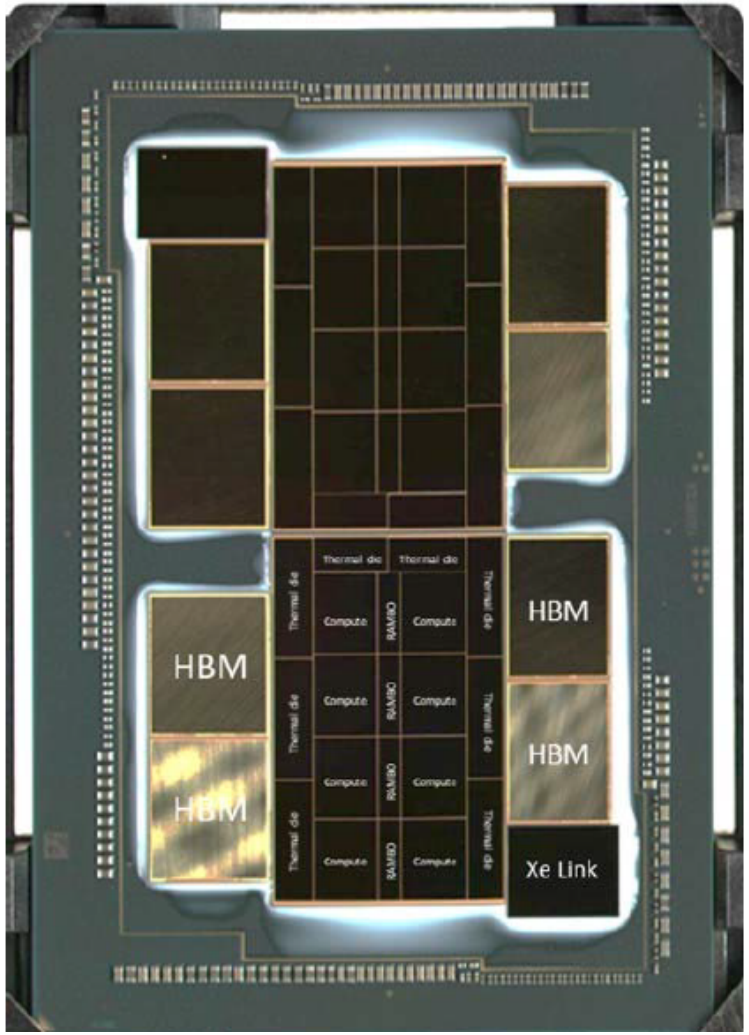
Monolithic 32-core Chip
777mm² total area
1.0x Cost

4 x 8-core Chiplet, 213mm² per chiplet
852mm² total area (+9.7%)
0.59x Cost



Naffziger, S., Noah Beck, T. Burd, K. Lepak, Gabriel H. Loh, M. Subramony and Sean White. "Pioneering Chiplet Technology and Design for the AMD EPYC™ and Ryzen™ Processor Families : Industrial Product." 2021 ACM/IEEE 48th Annual International Symposium on Computer Architecture (ISCA) (2021): 57-70.

Chiplet-Based Design: Intel's Ponte Vecchio

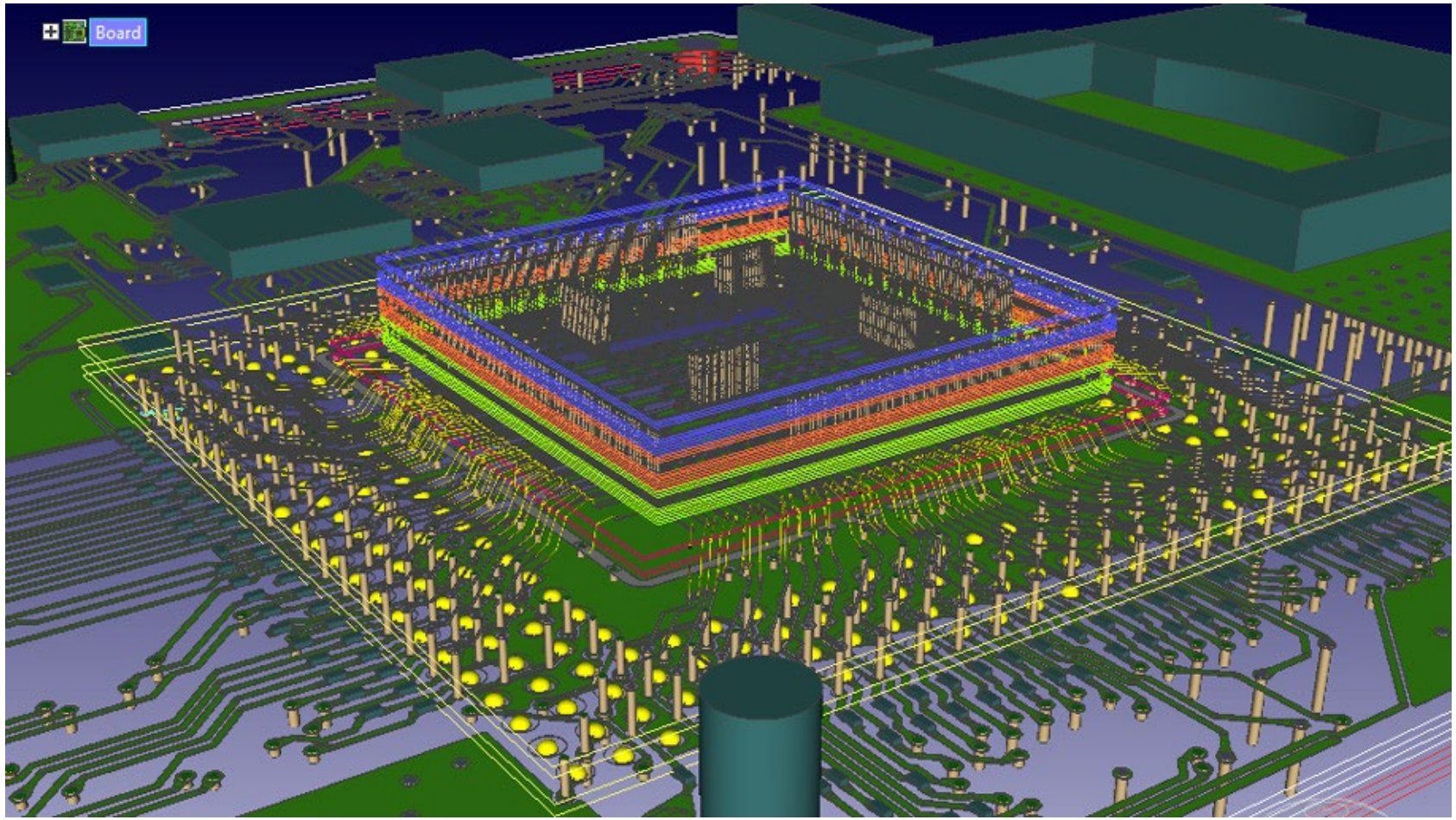


Integration	Foveros + EMIB
Power Envelope	600W
Transistor count	> 100B
Total Tiles	63 (**47 functional** + 16 thermal Tiles)
HBM count	8
Package Form factor	77.5 x 62.5 mm (4844 mm^2)
Platforms	3 platforms
IO	4x16 90G SERDES, 1x16 PCIe Gen5
Total Silicon	3100 mm^2 Si
Silicon footprint	2330 mm^2 Si footprint
Package layers	11-2-11 (24 layers)
2.5D Count	11 2.5D connections
Resistance	0.15 mΩ R_{path}/tile
Package pins	4468 pins
Package Cavity	186 mm^2 x4 cavities

Wilfred Gomes, Altug Koker, Pat Stover, Doug Ingerly, Scott Siers, Srikrishnan Venkataraman, Chris Pelto, Tejas Shah, Amreesh Rao, Frank O'Mahony, Eric Karl, Lance Cheney, Iqbal Rajwani, Hemant Jain, Ryan Cortez, Arun Chandrasekhar, Basavaraj Kanthi, Raja Koduri, "*Ponte Vecchio: A Multi-Tile 3D Stacked Processor for Exascale Computing*", ISSC 2022

Co-Design Requirements

- Tradeoffs in advance
- Translation and domains
- Propagate information
- Manage connectivity
- Database formats



Courtesy of Zuken

Future System Needs and Functions

Automotive

Sensors/MEMS

Data Center

High bandwidth

Mobile Wireless

Analog, RF

Computer

High-speed Digital

Trends in Electronic Design

- Low Power
- High Bandwidth
- Reduced Size and Weight
- Increased Functionality
- Reduced Cost

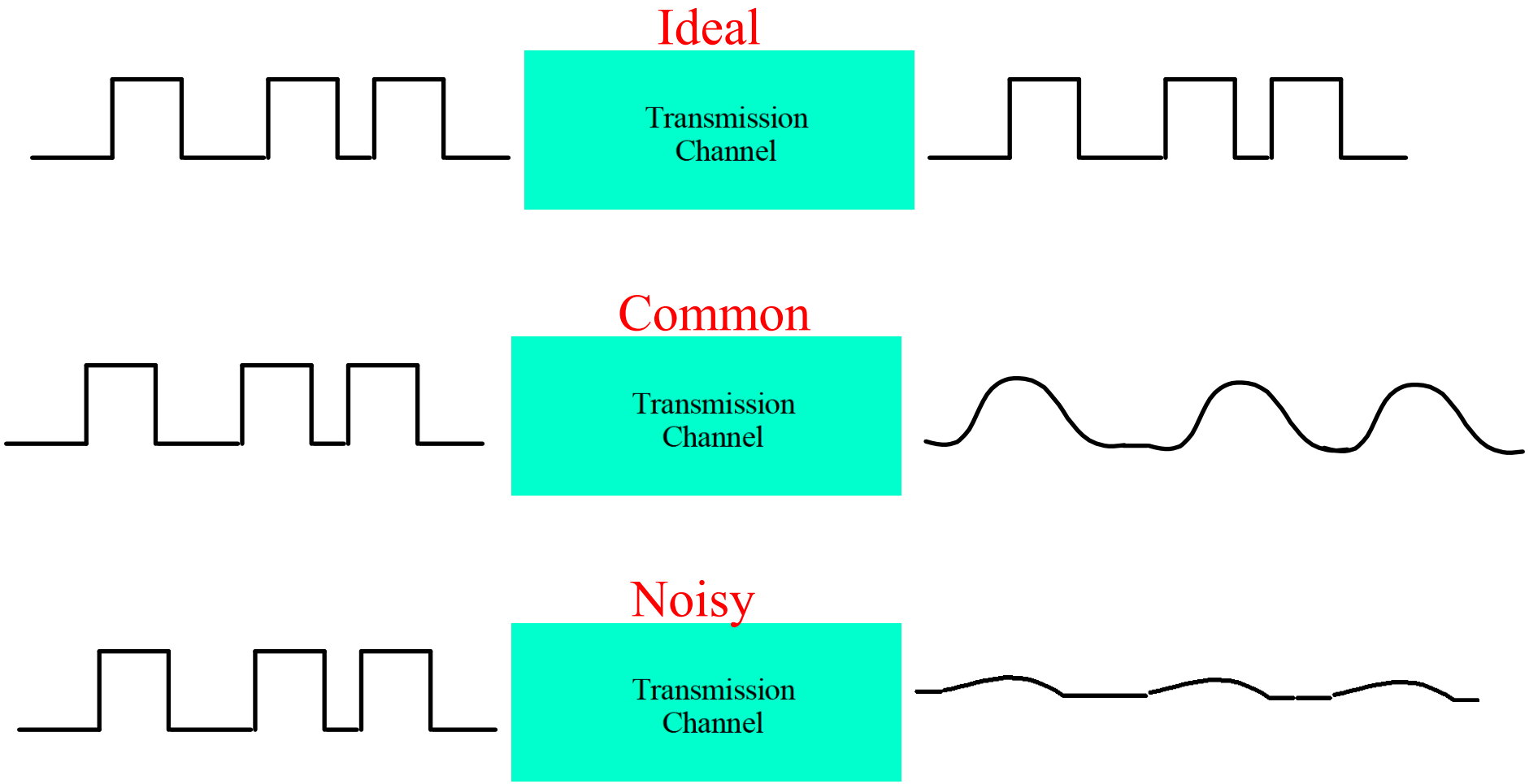
Modern systems are limited by data movement bandwidth, power, and signal integrity

Advanced Packaging

- Parallelism preferred over raw speed
- PAM4 required beyond 25 Gbps
- Nyquist frequency limits scaling
- Channel loss dominates design
- Equalization is mandatory

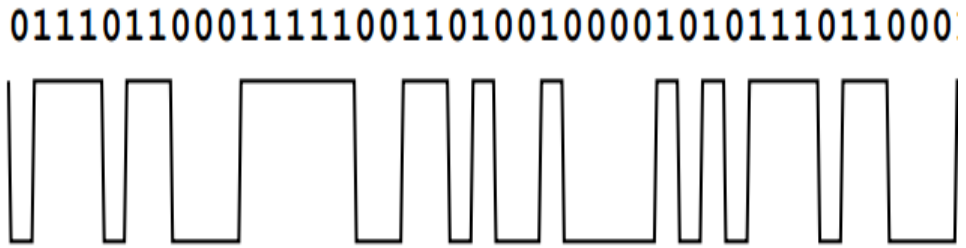
Packaging today is a signal integrity problem

Signal Integrity



Signal Integrity

- Serial data transmission sends binary bits of information as a series of optical or electrical pulses

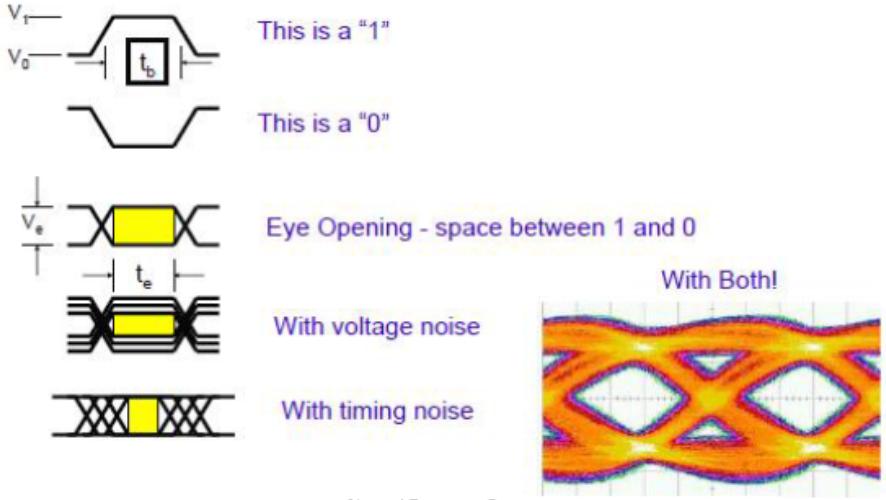


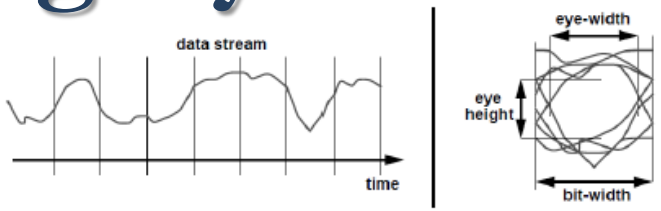
- The transmission channel (coax, radio, fiber) generally distorts the signal in various ways

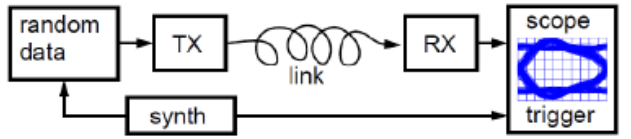
- From this signal we must recover both clock and data

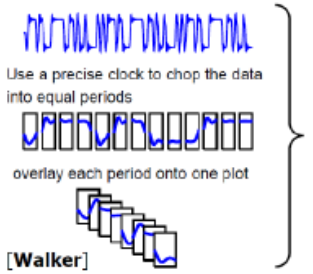
Signal Integrity

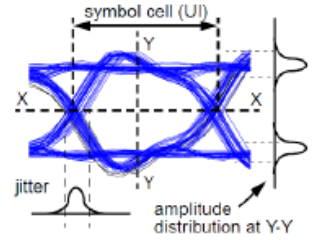
Eye-Diagrams

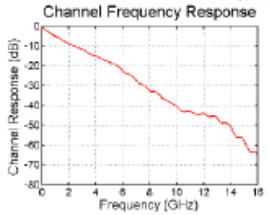


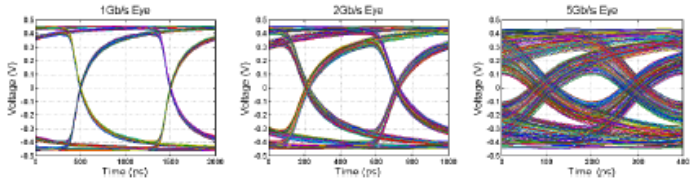


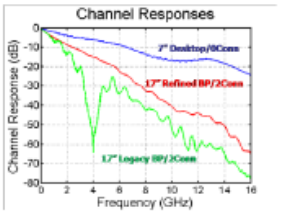


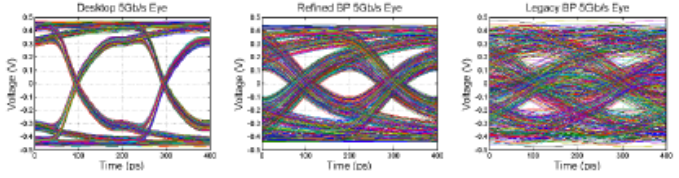




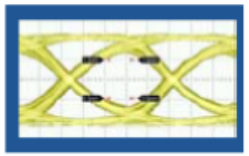




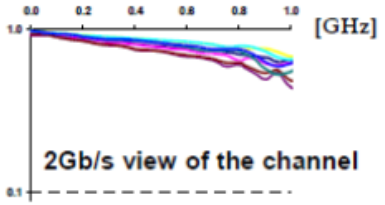


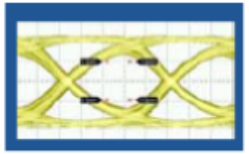


Channel

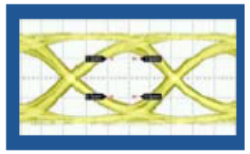


Signal at Tx

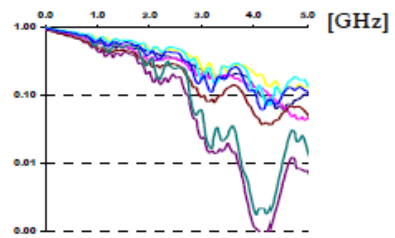


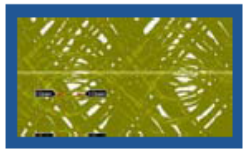


Signal at Rx

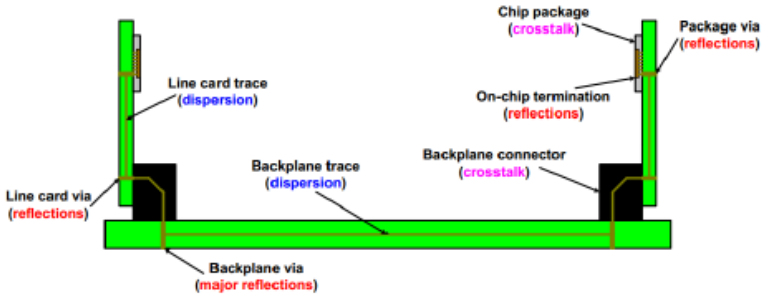


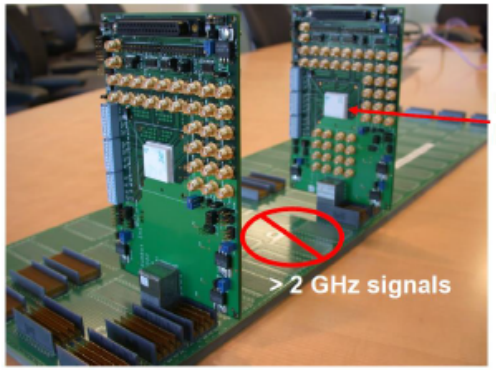
Signal at Tx

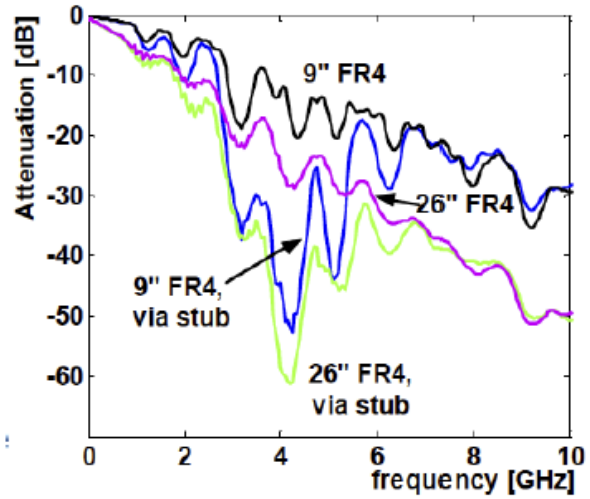




Signal at Rx

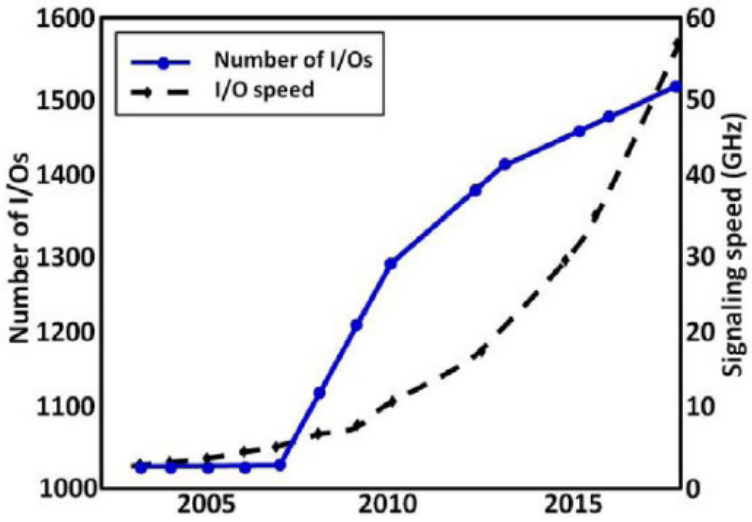




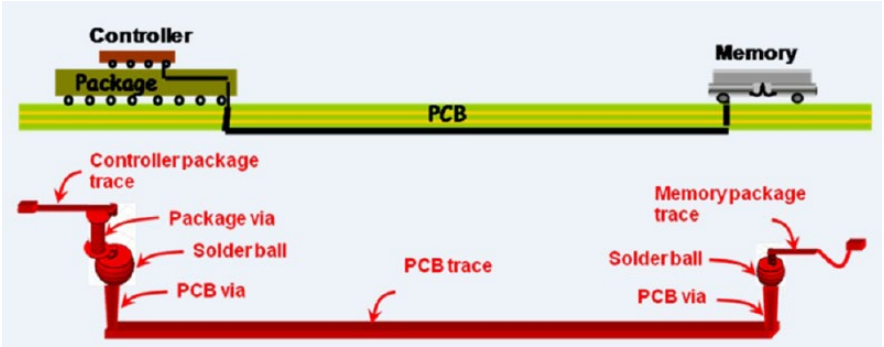


Design Challenges for High-Speed Links

- Modern computer systems require Tb/s aggregate off-chip signaling throughput
  - Interconnect resources are limited
    - Parallel buses with fast edge rates must be used
  - Package size and pin count cannot keep up with speed
  - Stringent power and BER requirements to be met
  - Channel attenuation increases with the data rate
  - High-performance signaling requires high-cost channels
  - Crosstalk-induced jitter



Available number and required speed of I/Os (ITRS roadmap)

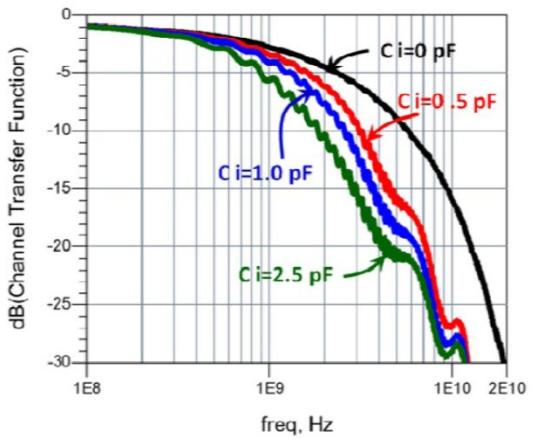


A typical controller-memory interface

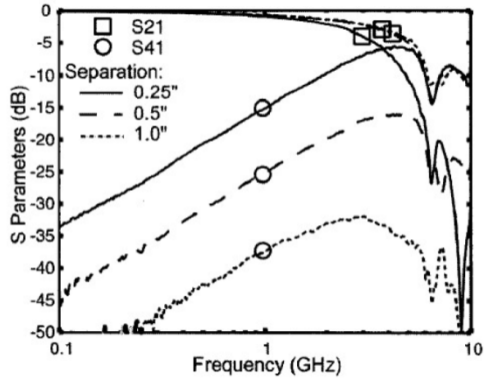
Signal Integrity Impairments In High-Speed Buses

- SI issues limit system performance to well below channel Shannon capacity
- Inter-Symbol Interference (ISI) is an issue for long backplane buses
- For short, low-cost parallel links, dominant noise source is crosstalk
  - Far-end crosstalk (FEXT) induces timing jitter (CIJ), impacts timing budget
- Other SI impairments:
  - Simultaneous-switching (SSO) noise
  - Thermal noise
  - Jitter from PLL/DLL

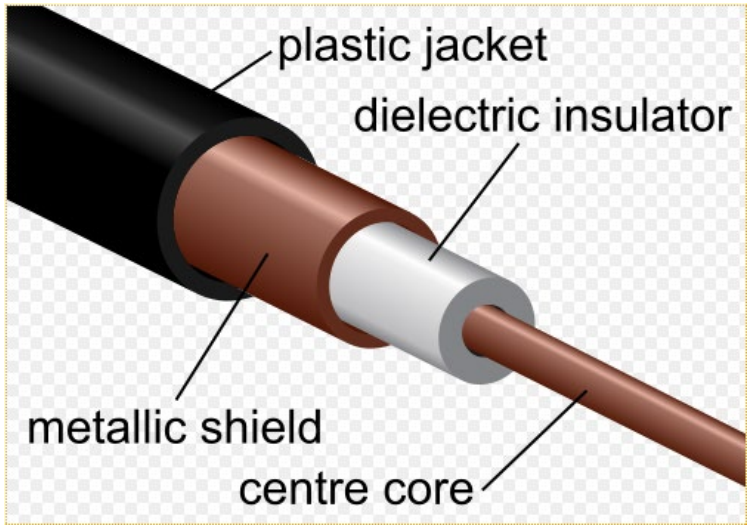
Insertion loss of a single DDR channel



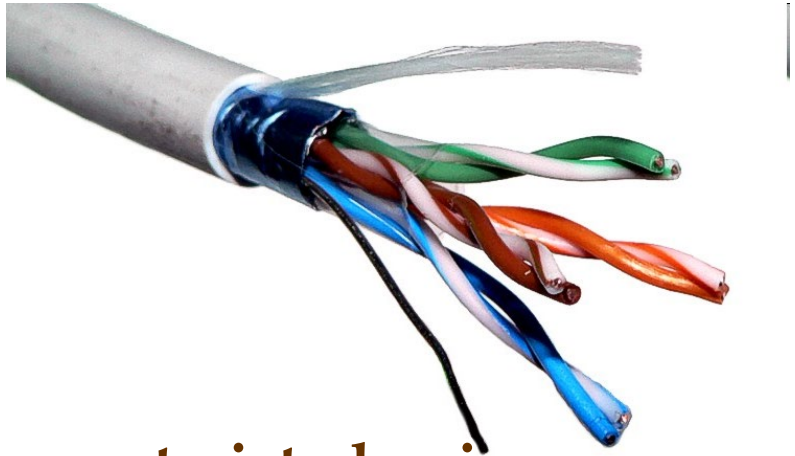
FEXT increases with routing density



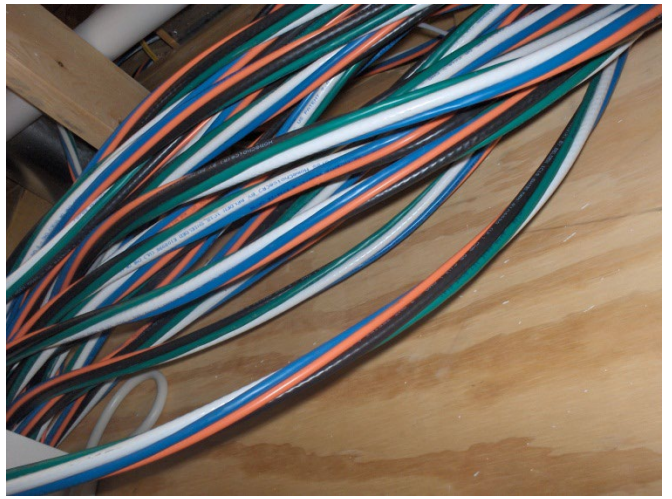
Cables and Transmission Lines



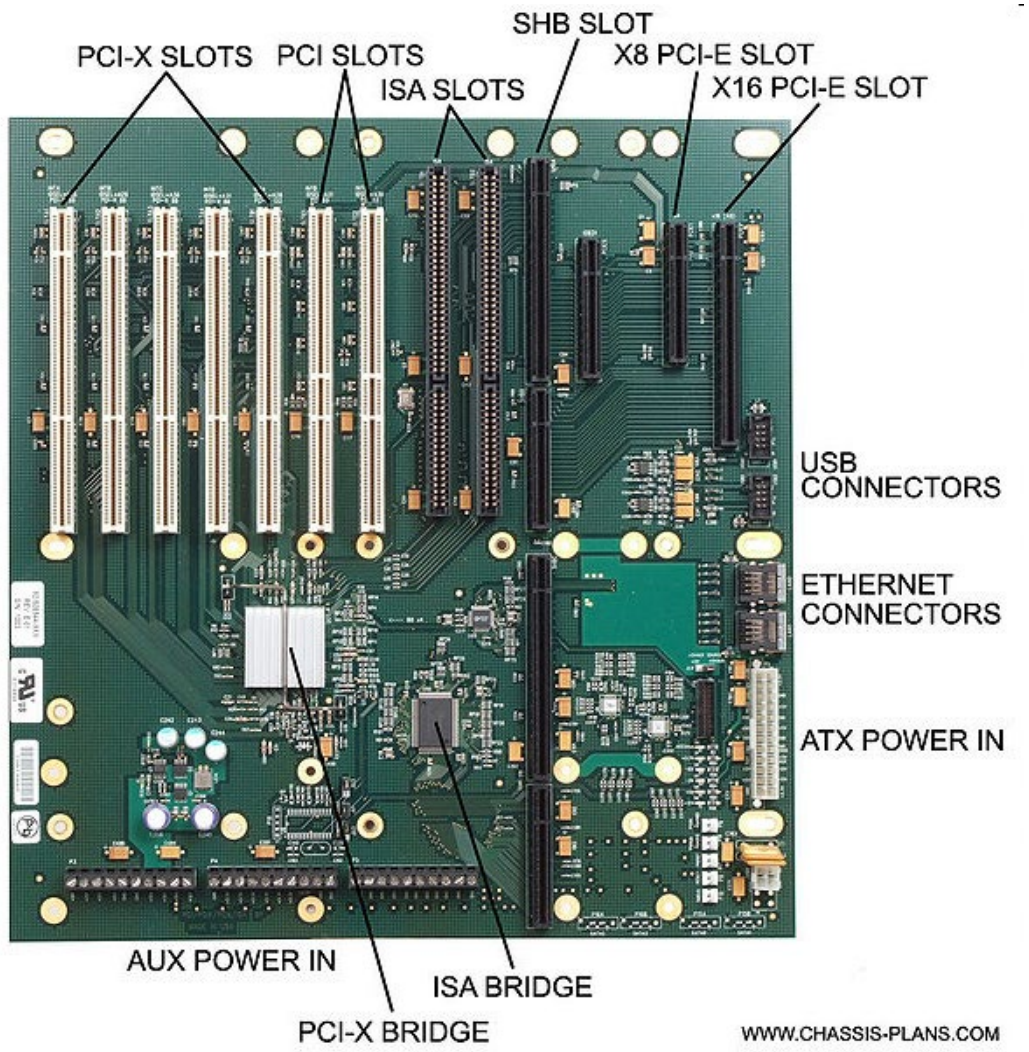
coaxial



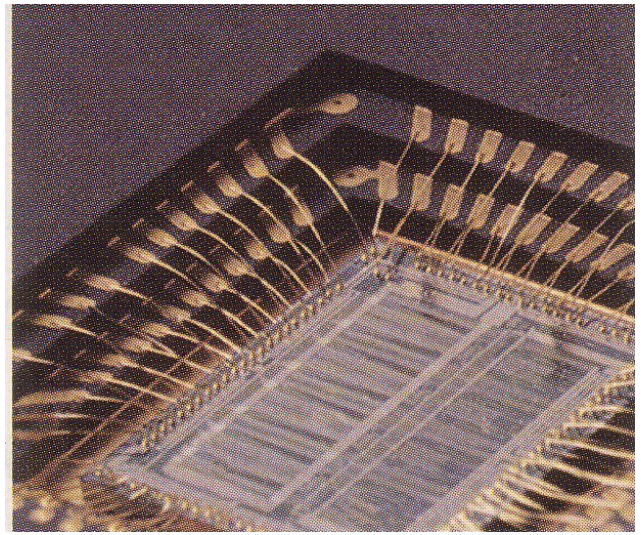
twisted pairs

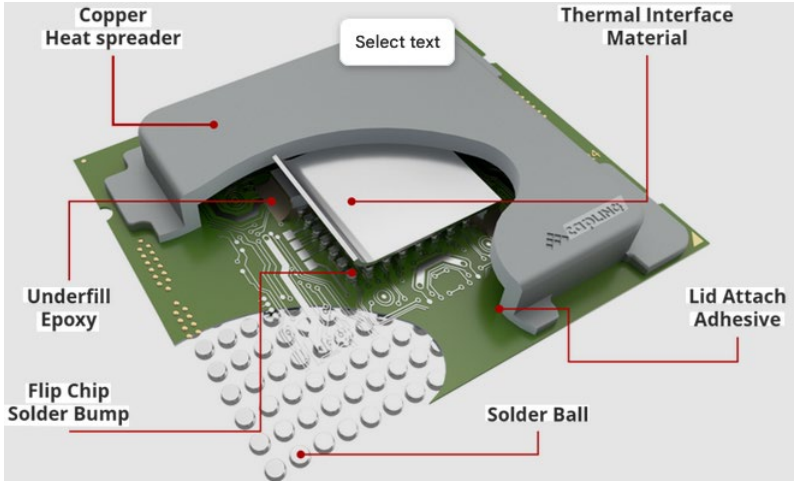


Motherboards and Backplanes



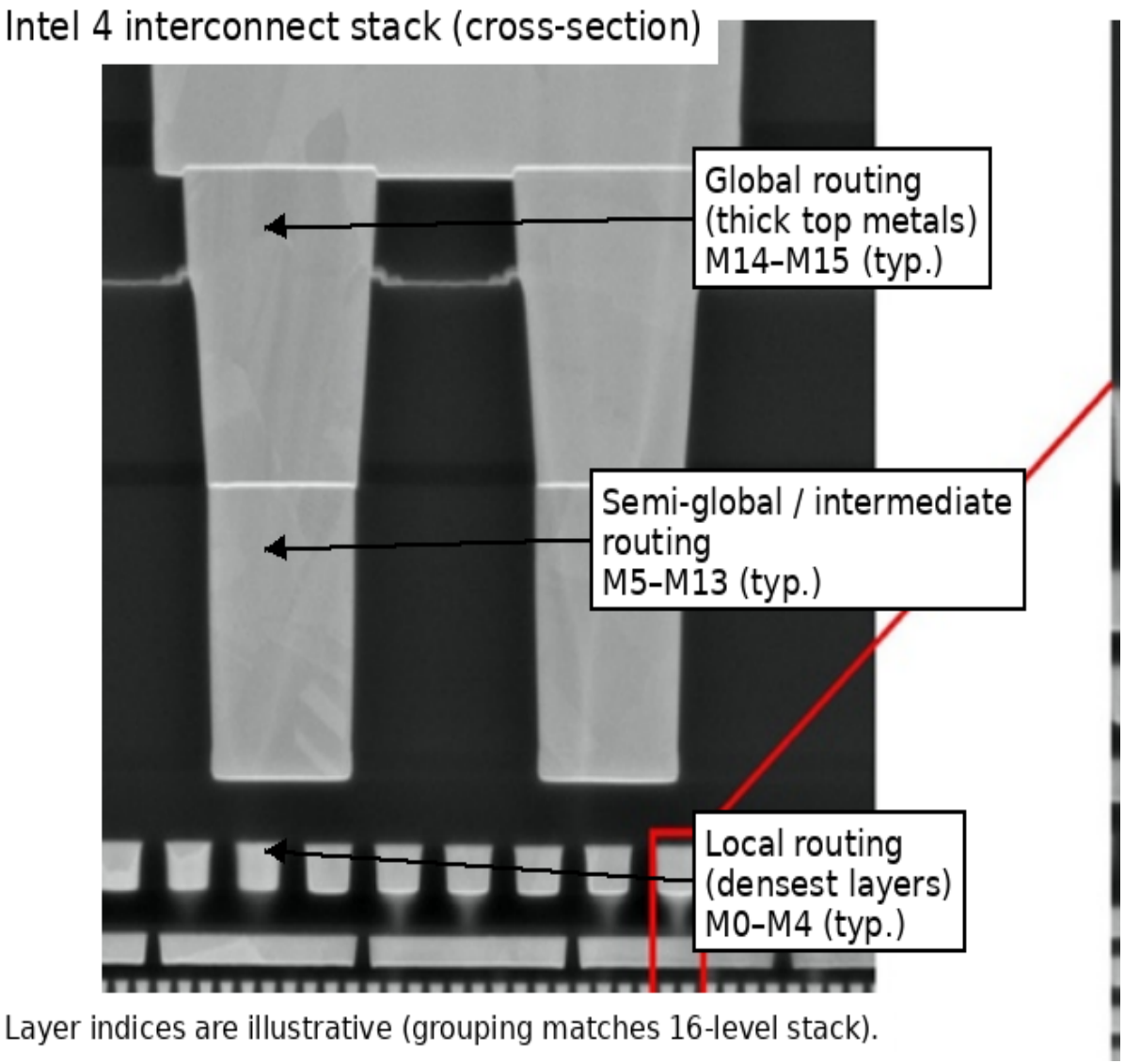
Package-Level Complexity





- Silicon interposers: 40–70+ metal layers
- TSV pitch: 10–50 µm, depth up to 100 µm
- RDL linewidth/spacing: 1–2 µm (fan-out)
- Hybrid bonding pitch: ≤10 µm
- 20–40 metal layers typical
- Core + build-up architecture
- Linewidth/spacing: 5–15 µm
- Skew and impedance variation across substrate

Modern BEOL Metal Stack (Intel 4-class) and Capacitance Scaling



Why capacitance becomes fringing-dominated

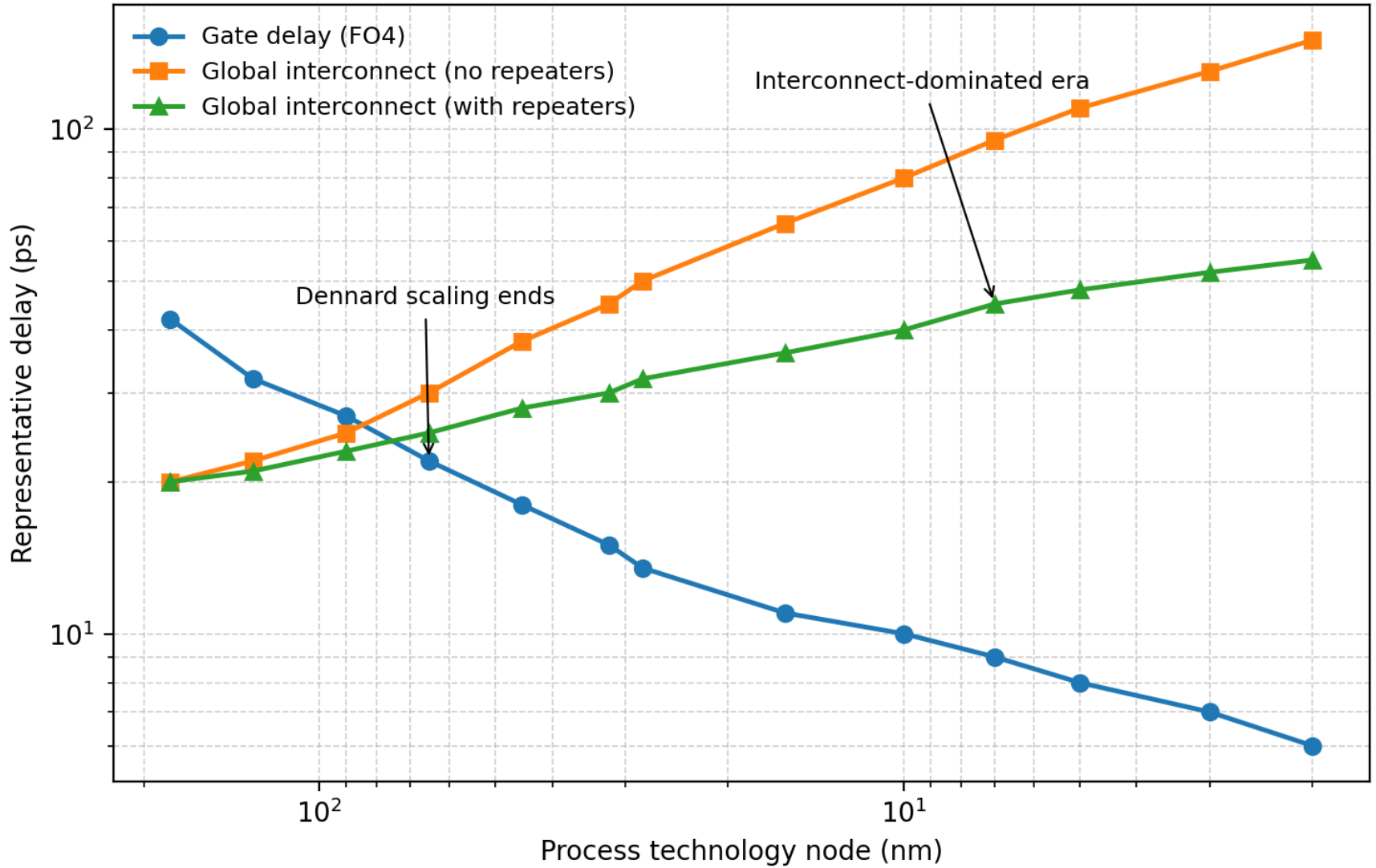
- When pitch shrinks, spacing becomes comparable to thickness → strong lateral coupling.
- Parallel-plate $C \approx \varepsilon \cdot W \cdot L/t$ stops scaling when t cannot shrink proportionally.
- Fringing / sidewall C grows in importance ($C_{edge} \approx \varepsilon \cdot L \cdot f(W,s,t)$), often dominating local layers.
- Modern low-k + air-gaps reduce ε, but coupling paths remain (especially M0–M4).
- Result: coupling noise, delay, and crosstalk become layout- and environment-dependent.
- Design response: shielding, spacing rules, RC extraction, and hierarchical routing (local→global).

Image source: VLSI Symposium 2022 Technical Tip Sheet, Intel (Paper T1-1), Figures 6 & 9. © Intel / VLSI Symposium. Used for educational purposes.

High-Speed Bus and Networks

- Memory Interfaces
  - DDR5 (6.4-8.8 Gbps/pin)
  - LPDDR5X (up to 8.5 Gbps)
  - GDDR6X (21-24 Gbps)
  - HBM3E (> 1 Tbps aggregate BW)
- Networking & Fabrics
  - Ethernet (up to 800 G (PAM4))
  - InfiniBand XDR (800 Gbps)
- Peripheral & Consumer Interfaces
  - USB4 v2 (80-120 Gbps)
  - HDMI 2.1 (48 Gbps)
  - Thunderbolt 4 (40 Gbps)
- Storage & Expansion
  - SATA III (6 Gbps)
  - UFS 4.0 (23.2 Gbps)
  - PCIe 4.0 (16 GT/s)
  - PCIe 5.0 (32 GT/s)
  - PCIe 6.0 (64 GT/s PAM4)

Signal Delay Trend

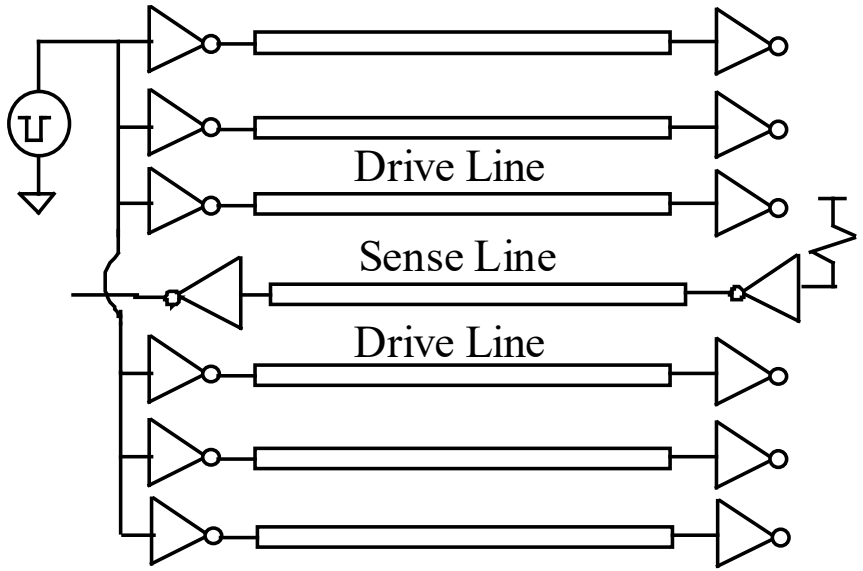


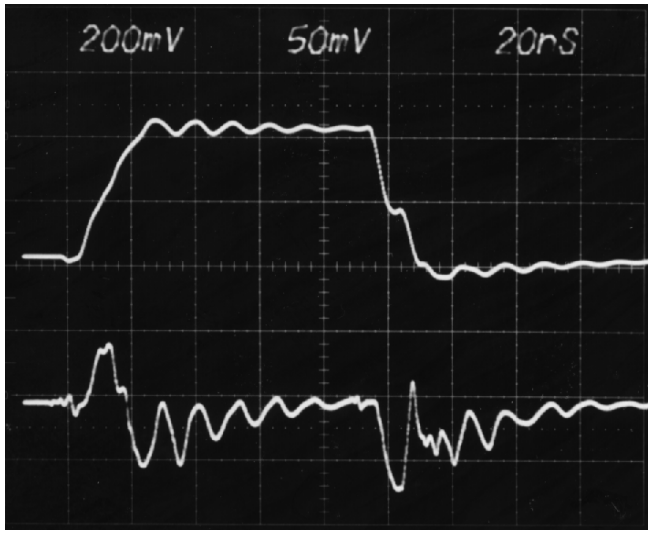
Interconnects

- Total interconnect length (m/cm^2) – active wiring only, excluding global levels will increase
- Interconnect power dissipation is more than 50% of the total dynamic power consumption in 130nm and will become dominant in future technology nodes
- Interconnect centric design flows have been adopted to reduce the length of the critical signal path

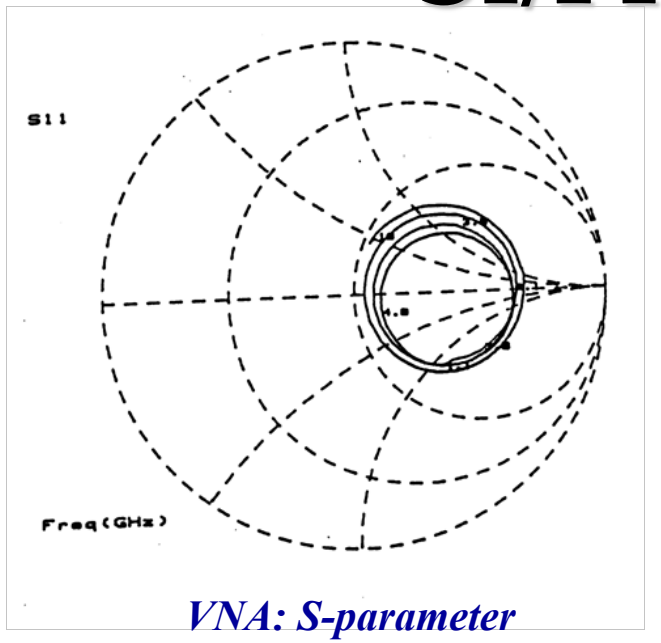
Signal Integrity Impairments

Crosstalk	Dispersion	Attenuation
Reflection	Distortion	Loss
Delta I Noise	Ground Bounce	Radiation

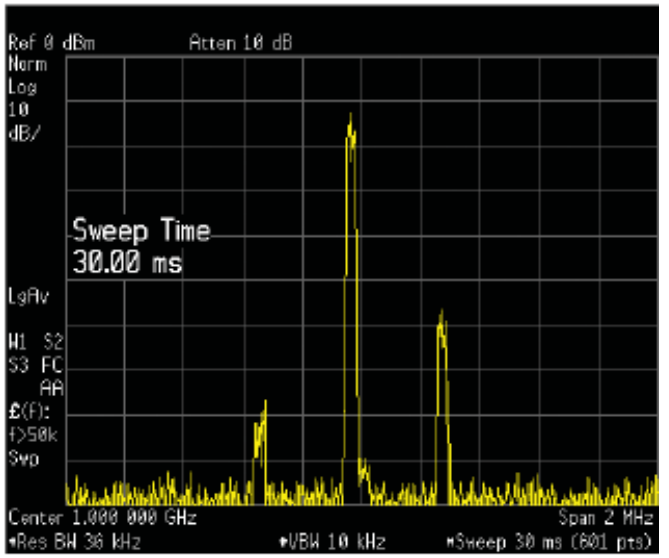




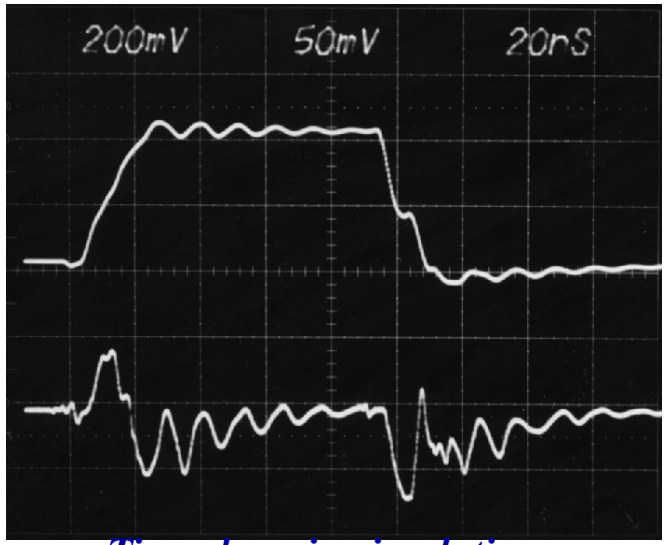
SI/PI Metrics



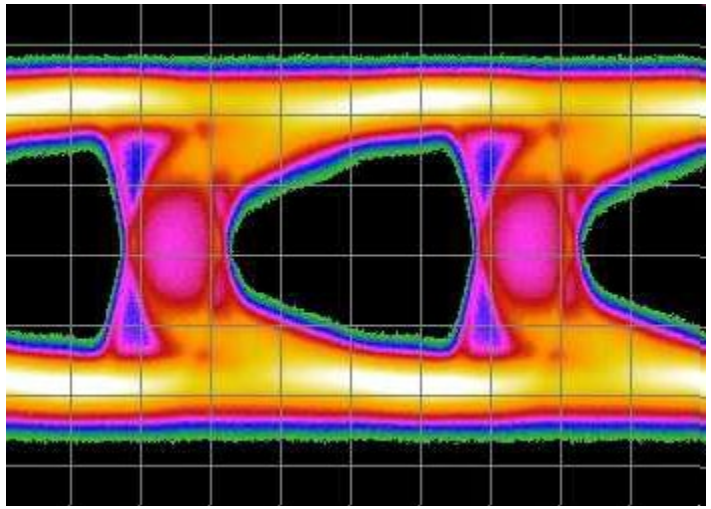
VNA: S-parameter



Spectrum Analyzer



Time-domain simulation

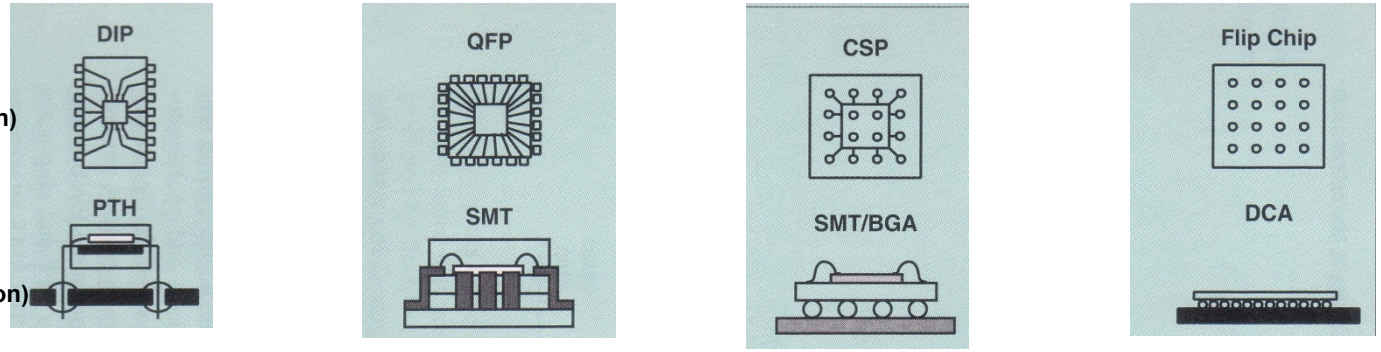


Eye diagram

Package Technologies (1995-2005)

Top View
(showing chip topackage connection)

Plane View
(showing package to board connection)



	DIP	QFP	CSP	Flip Chip
Chip Size (mm × mm)	5 ×5	16 × 16	25 ×25	36 × 36
Chip Perimeter (mm)	20	64	100	144
Number of I/Os	64	500	1600	3600
Chip Pad Pitch (μm)	312	128	625	600
Package Size (in × in)	3.3 × 1.0	2.0 × 2.0	1.0 × 1.0	1.4 × 1.4
Lead Pitch (mils)	100	16	25	24
Chip Area (mm^2)	25	256	625	1296
Feature Size (μm)	2.0	0.5	0.25	0.125
Gates/Chip	30K	300K	2M	10M
Max Frequency (MHz)	5	80	320	1280
Power Dissipation (W)	0.5	7.5	30	120
Chip Pow Dens (W/cm^2)	2.9	4.8	9.3	2.0
Pack Pow Dens (W/cm^2)	0.024	0.3	4.8	9.8
Supply Voltage (V)	5	3.3	2.2	1.5
Supply Current (A)	0.1	2.3	13.6	80

Package Technologies (2025)

Package Technology	Typical Die Size	I/O Count	Pitch	Bandwidth / Data Rate	Power Density	Primary Use Cases
Wire-Bond QFP / BGA	5–15 mm	64–400	400–800 µm	≤1 Gb/s	<5 W/cm²	Legacy MCUs, PMICs
Flip-Chip BGA (FC-BGA)	10–35 mm	1k–10k	150–180 µm	10–32 Gb/s	10–50 W/cm²	CPUs, GPUs
Fine-Pitch FC-BGA	20–50 mm	10k–40k	90–130 µm	32–56 Gb/s	50–150 W/cm²	HPC, AI accelerators
2.5D Interposer (Si/Organic)	Chiplets	20k–80k	40–55 µm	56–112 Gb/s	150–300 W/cm²	GPUs + HBM
Fan-Out Wafer-Level (FOWLP)	≤20 mm	500–5k	40–80 µm	10–32 Gb/s	20–80 W/cm²	Mobile SoCs
3D-IC (TSV stacked)	Stacked dies	10k–100k	<40 µm	112+ Gb/s	300–1000 W/cm²	AI, HBM, logic-on-logic

Package Electrical Parameter Evolution

Metric	2000-Era	State-of-the-Art (2025)
Feature Size	2 µm → 0.125 µm	5 nm → 3 nm
Max Clock / Data Rate	5–1280 MHz	5–10 GHz clocks, 112 Gb/s links
Supply Voltage	5 V → 1.5 V	0.6–0.9 V (core)
I/O Pitch	600–300 µm	40–55 µm (micro-bumps)
Chip Power	≤120 W	500–1000 W (module-level)
Power Density	≤10 W/cm²	100–1000 W/cm²

Tools for Signal Integrity

- **Electromagnetics**
  - Fundamentals
  - Modeling and extraction
  - Numerical techniques
- **Circuits, Devices & Networks**
  - Compact models
  - Reduced-order models
  - Behavioral models
- **Signal Processing**
  - Time/frequency domain
- **Measurements**
  - Scattering parameters
  - Time-domain reflectometry
  - Eye diagrams