High-speed Interconnects for Advanced Computing and Communication: Evolution, Challenges, and What lies ahead?

Rohit Sharma
Professor
Department of Electrical Engineering
Indian Institute of Technology Ropar

rohit@iitrpr.ac.in

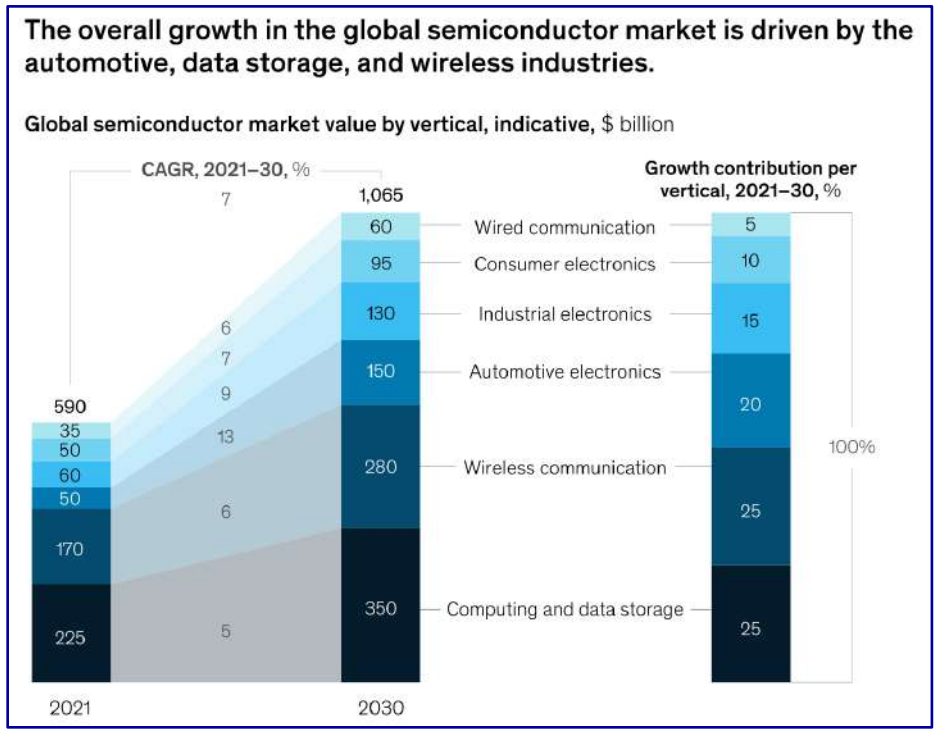
Outline

- Introduction
- The Interconnect Bottleneck
- Cu-Graphene Interconnects
- Co-design for Heterogeneous Integration
- Machine Learning based Design Approaches
- Summary

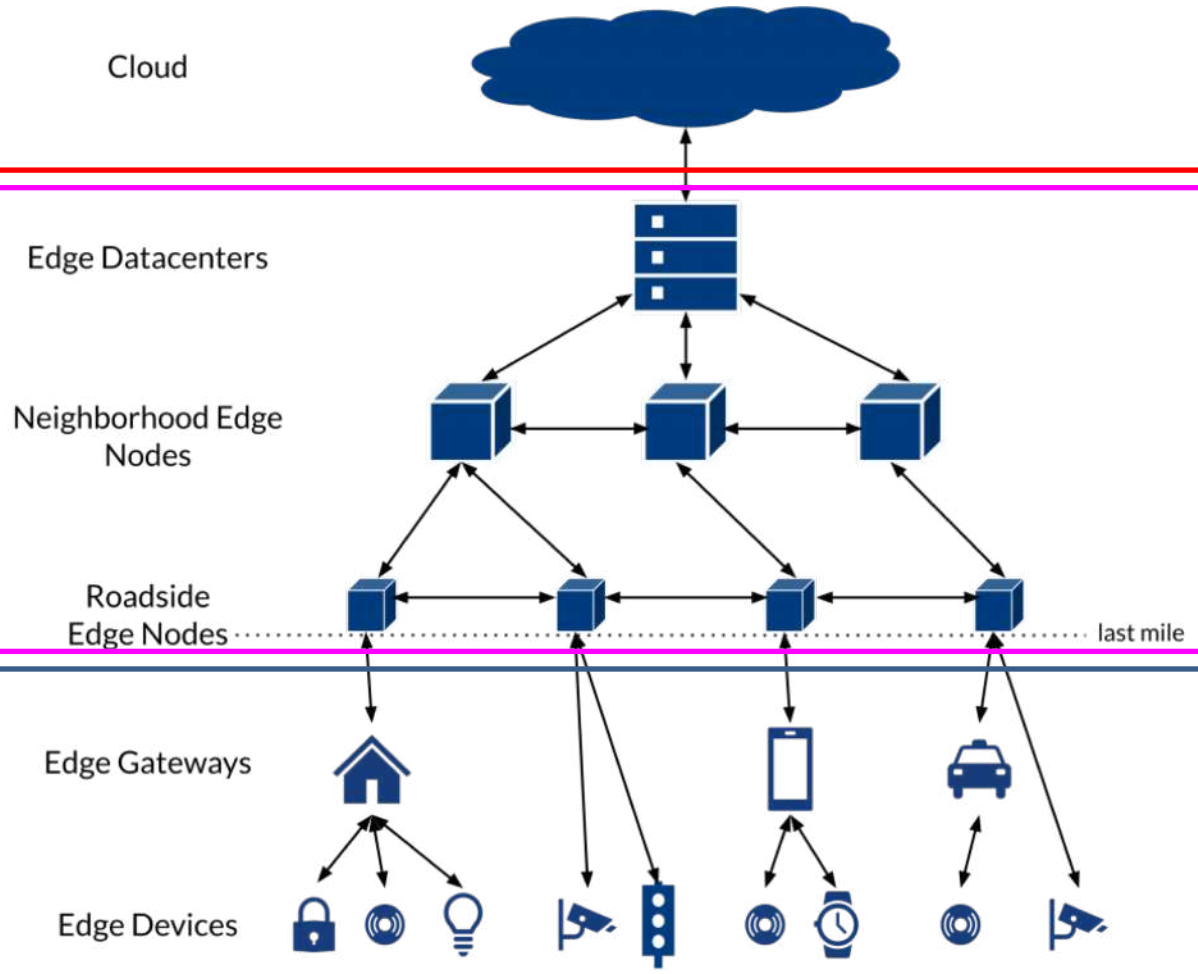
Semiconductor Industry Outlook

- Global semiconductor industry projected to become a trillion-dollar industry by 2030
  (Source: McKinsey & Company)
  - 55 years to become a 0.5T industry
  - Expected to double in the next 10 years
  - Drivers: Computing/Storage, Wireless, Automotive, ….

- 2022: Indian Semiconductor Mission
  - Establish Onshoring Capabilities and Indian Leadership
  - $11B Investment



Distributed Computing and Communications



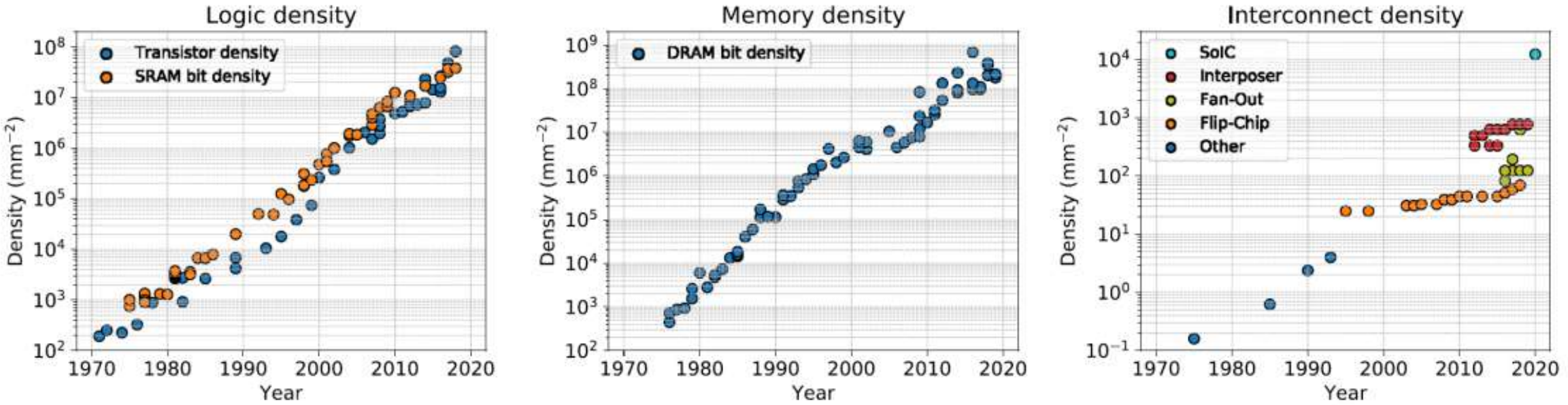
- Storage, Coordinator of lower levels
- Limitless Compute Power
- ~8000
- Power: 10s of KWs - 100s of KW

- Base stations, Routers, Switches
- High Compute Power
- 10-100 million devices
- Power: 10s of W – few KWs

- Sensors, Smart Phones, Automobiles
- Billions of devices
- Power: 10s of mW – 10s of W

Source: Jeff Burns, "Systems and Architectures for Distributed Compute", SRC Workshop, 2022.

Where are we & what is needed to move ahead?



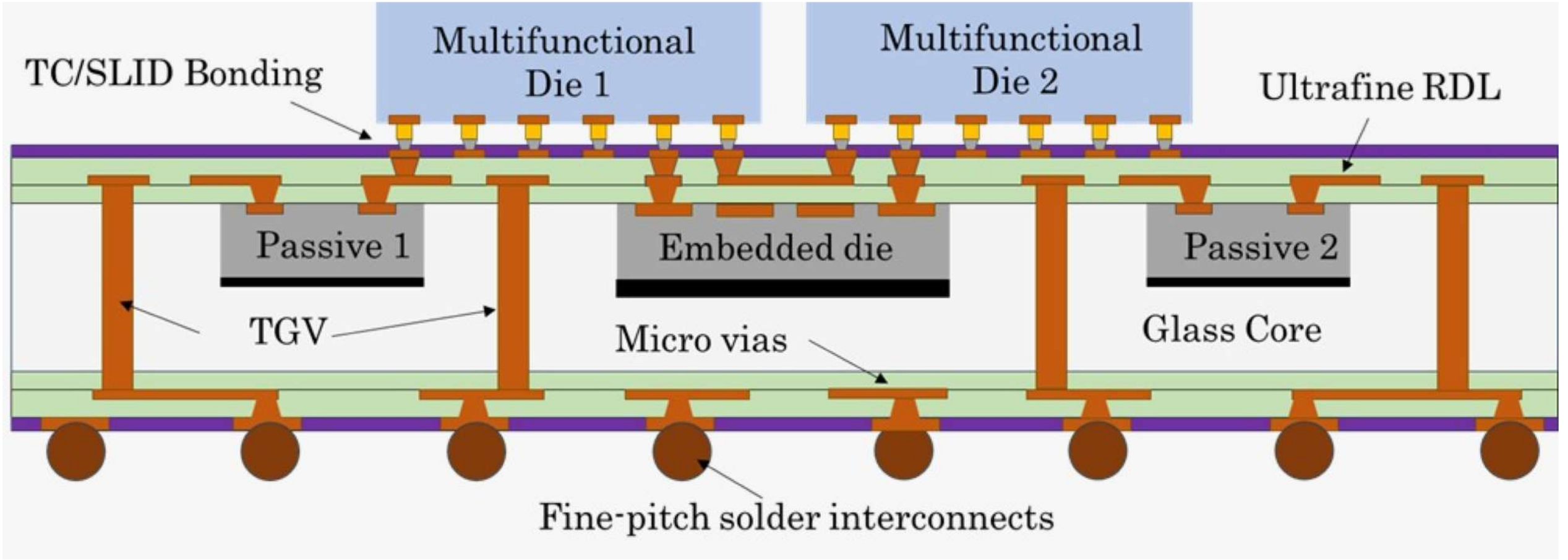
H.-S. Philip Wong, et al, "A Density Metric for Semiconductor Technology", Proceedings of the IEEE, April 2020

Current State of the Art

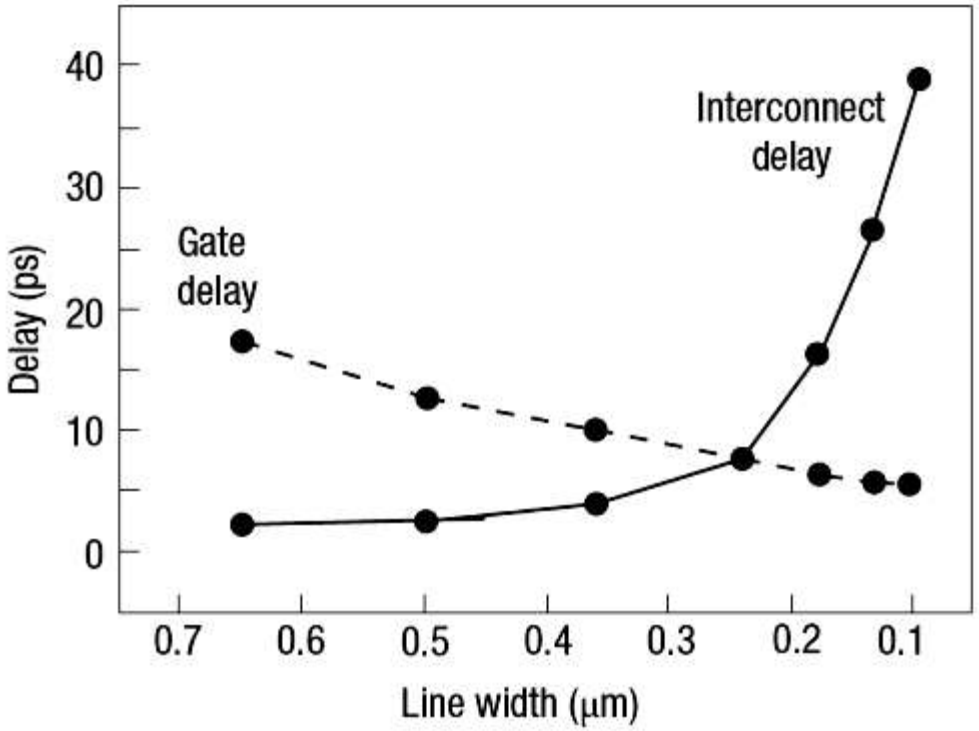
- Monolithic logic 10^8 transistors/mm^2
- DRAM 10^9 transistors/mm^2
- IO density 10^4 IO/mm^2
- SRAM Access 20-50 TBps

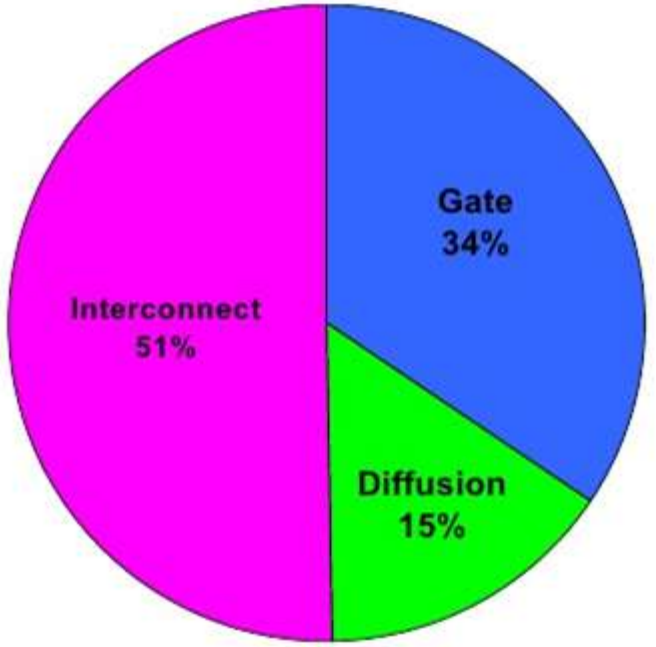
10-100X increase in transistor densities
Interconnect densities10^6 and higher (100X)
Energy per bit (EPB) reduced to femto-joules/bit
500TBps/mm^2 of bandwidth (10X increase)
Wireless communication at 1Tbps

Heterogeneously packaged embodiment



Where are we & what is needed to move ahead?





Total dynamic power

Source: Intel

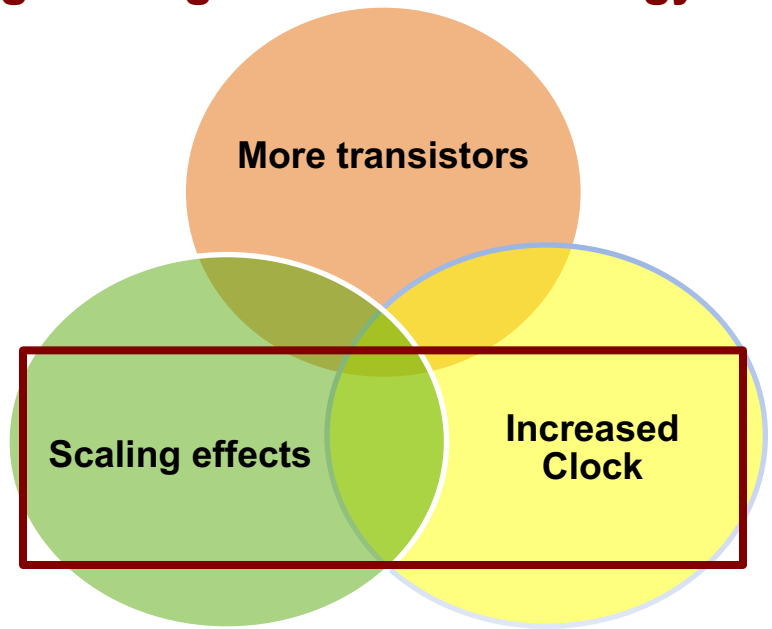
The big picture

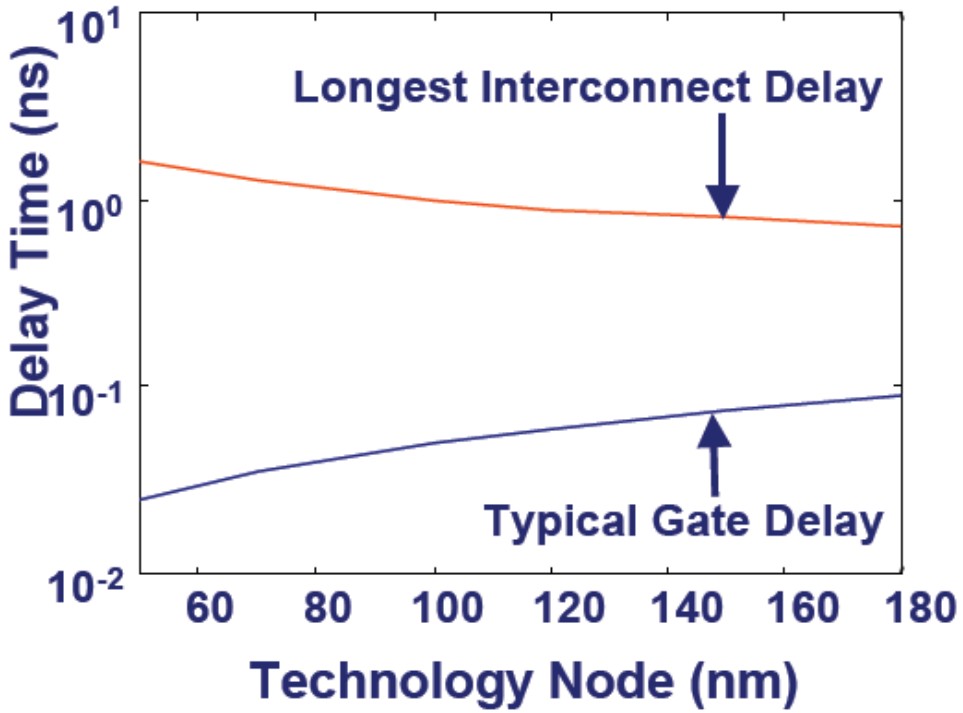
- With continuous scaling, opposite impact is seen on distributed resistance-capacitance (*rc*) product for transistors and interconnects.
- For current and future technology nodes, **interconnect delay** cannot be ignored and is considered to **dominate over transistor delay**.
- Majority **power** dissipation in current microprocessors is due to interconnects.

Source: R. Kirchain and L. Kimerling, "A roadmap for nanophotonics," *Nature Photonics*, vol. 1, pp. 303-305, 2007

The Interconnect Bottleneck

Design changes in new technology nodes





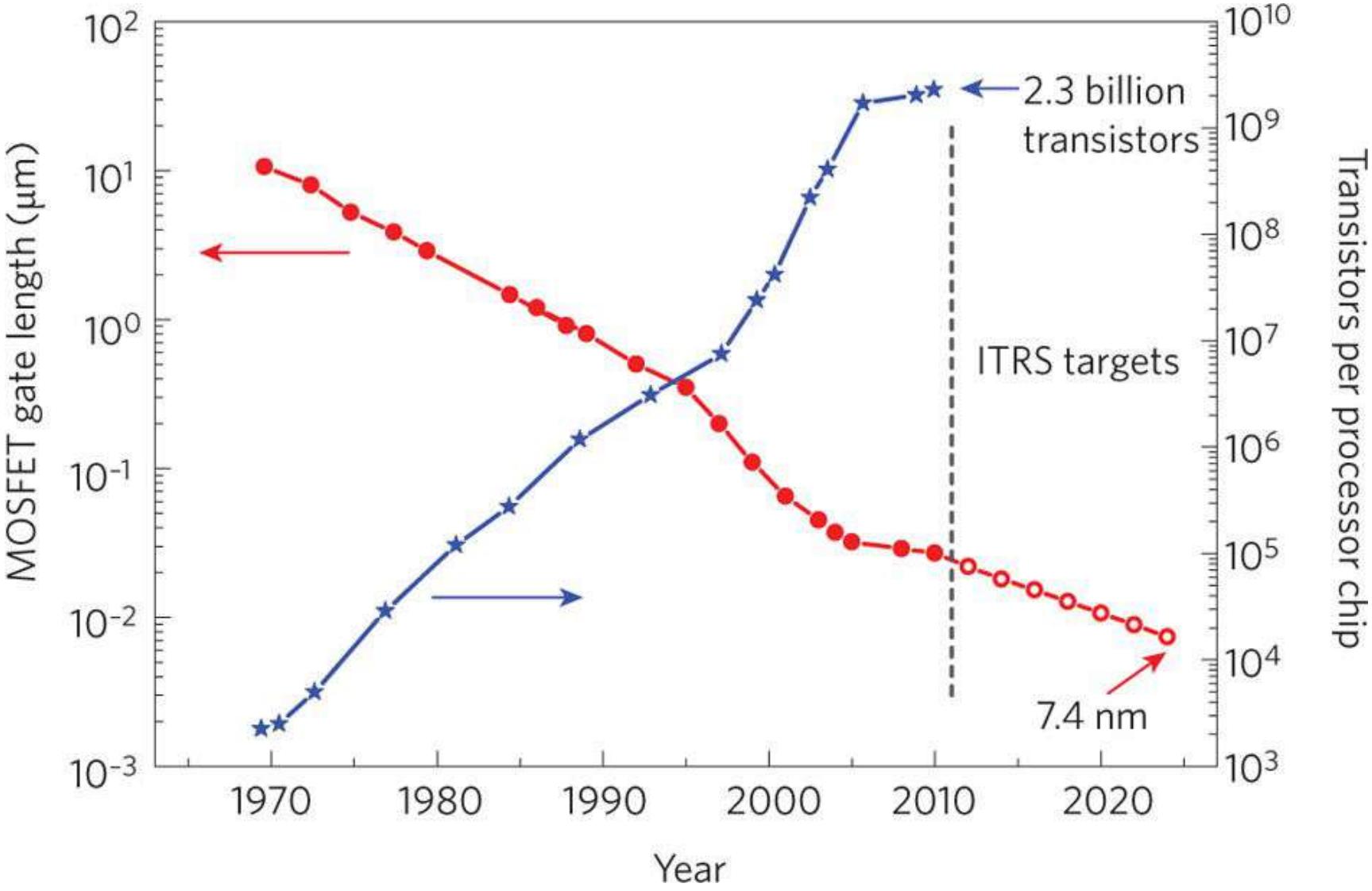
Key observations

Interconnects become **longer** and **thinner**

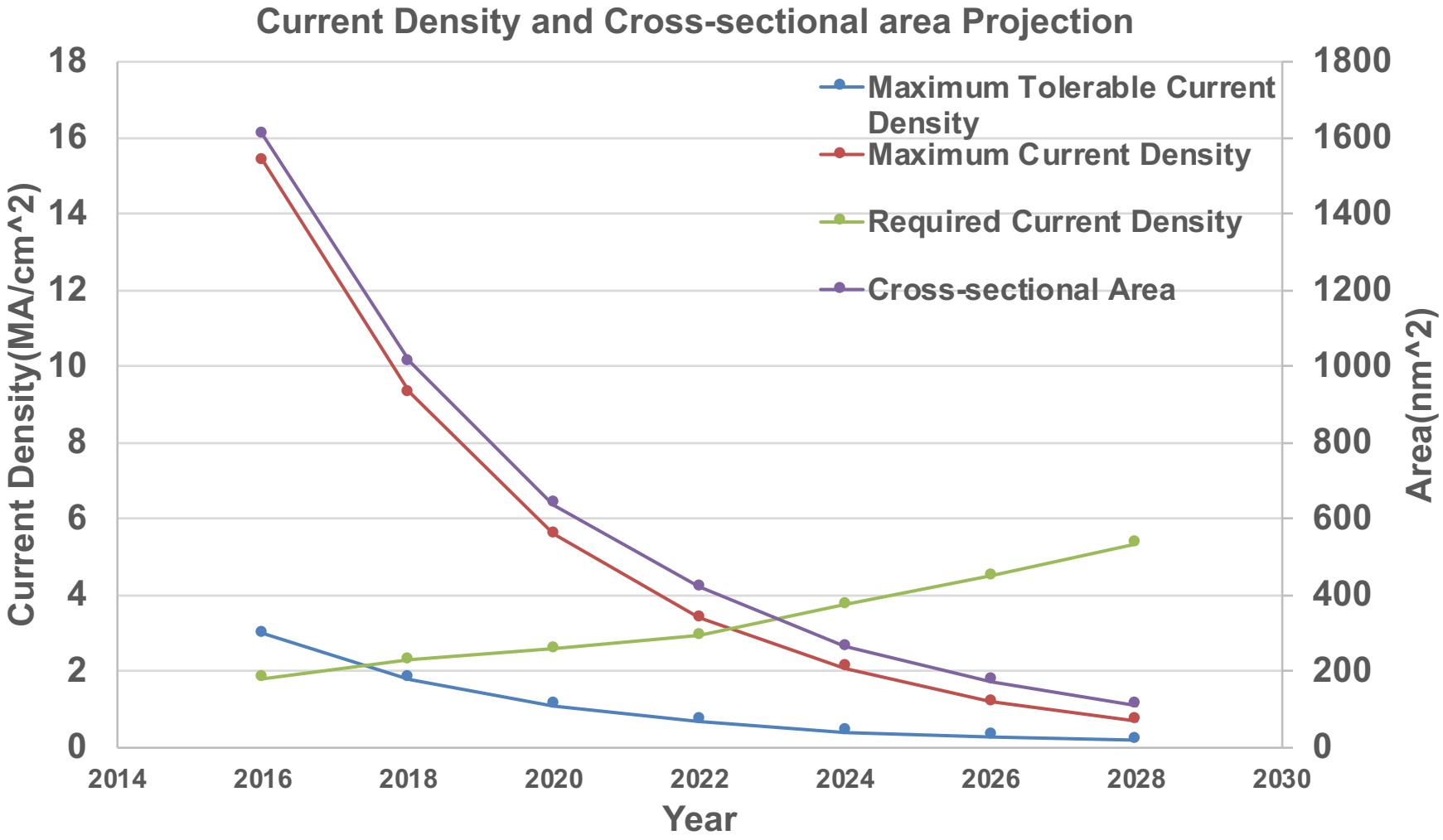
- Increased **delay**
- More **losses**

Source: International Semiconductor Technology Roadmap

The Interconnect Bottleneck

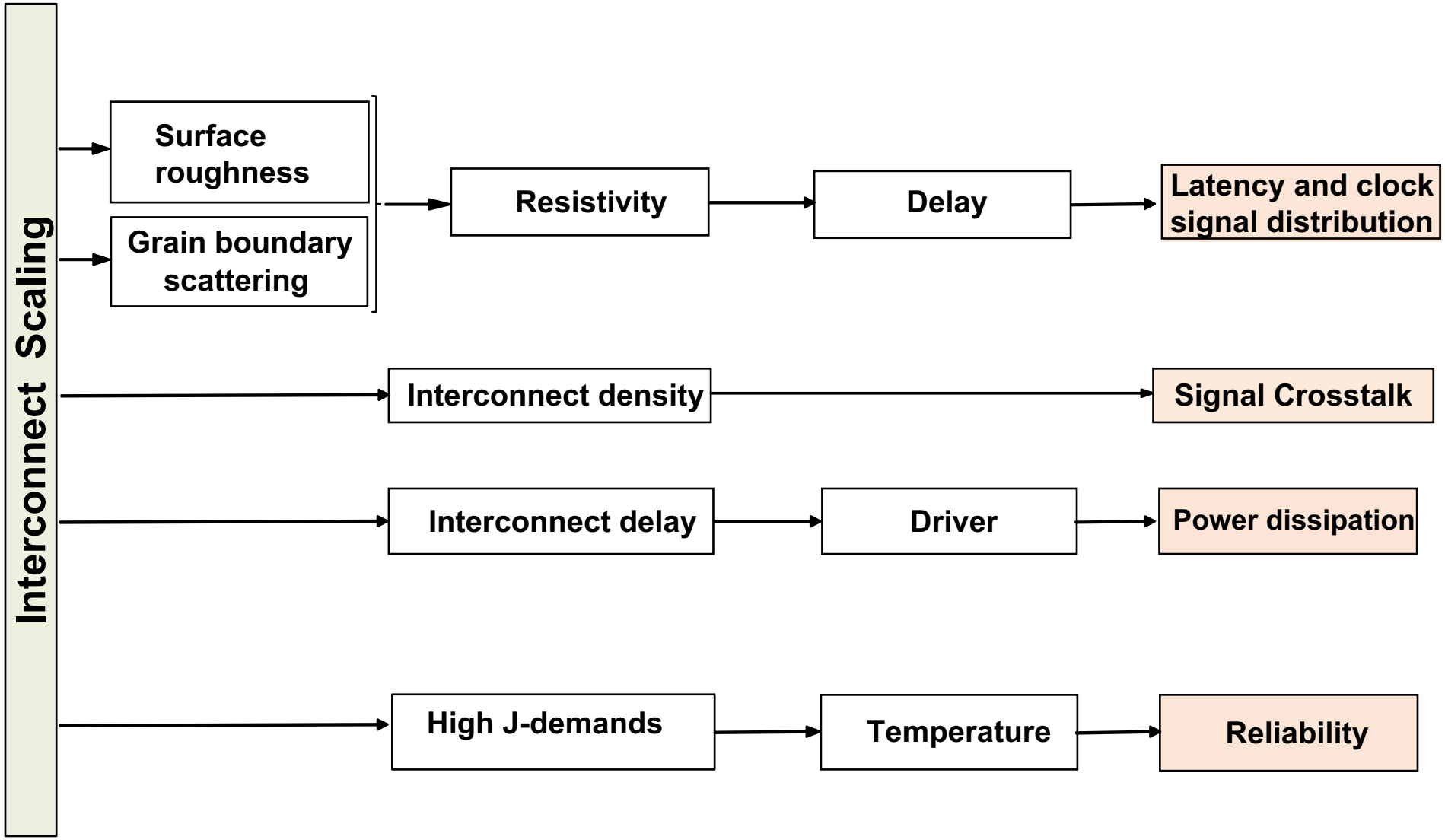


The Interconnect Bottleneck

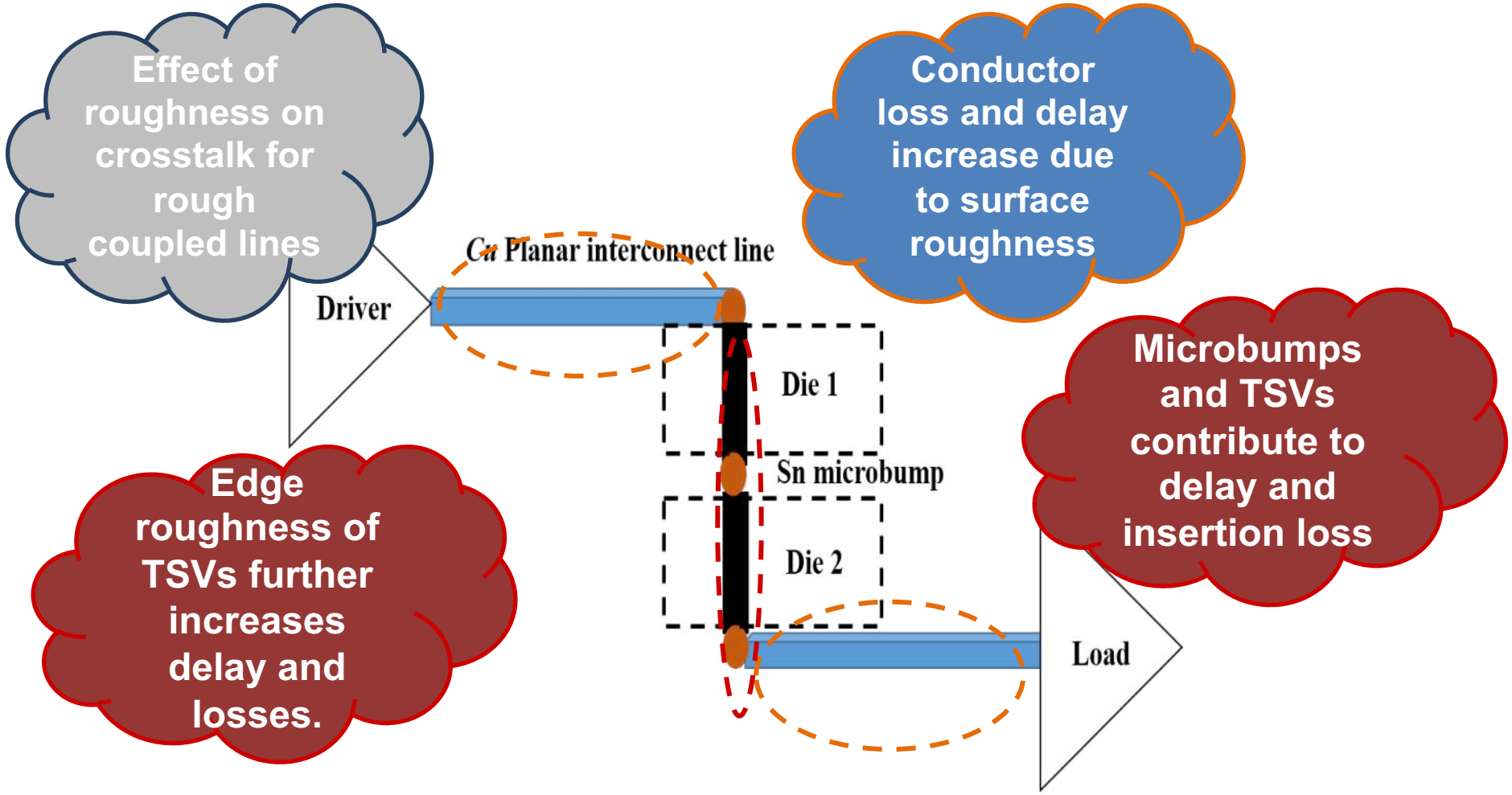


International Technology Roadmap for Semiconductors, 2015. [Online]. Available: http://www.itrs2.net/itrs-reports.html

Surface Roughness in Cu Interconnects

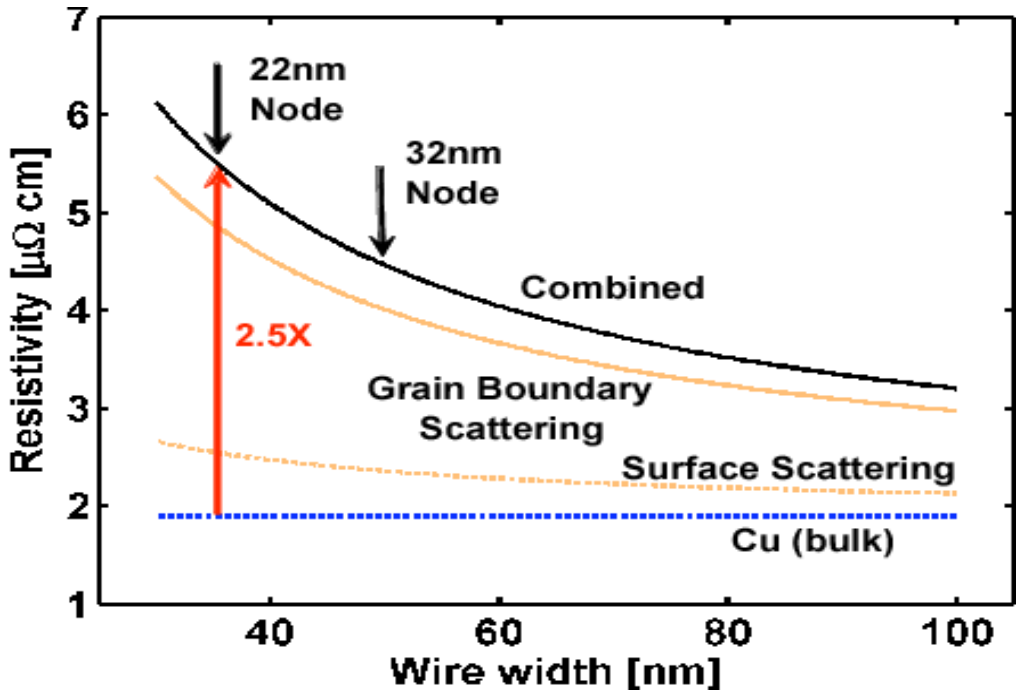


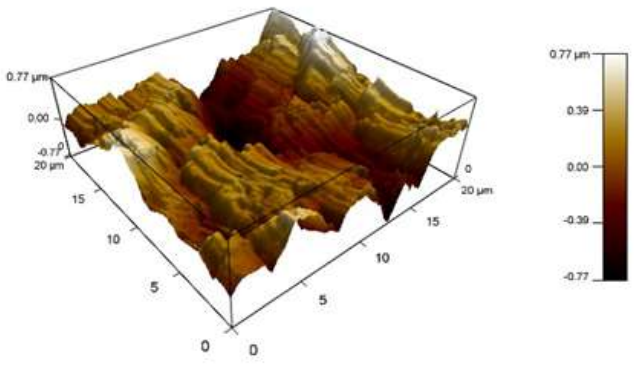
Surface Roughness in Cu Interconnects

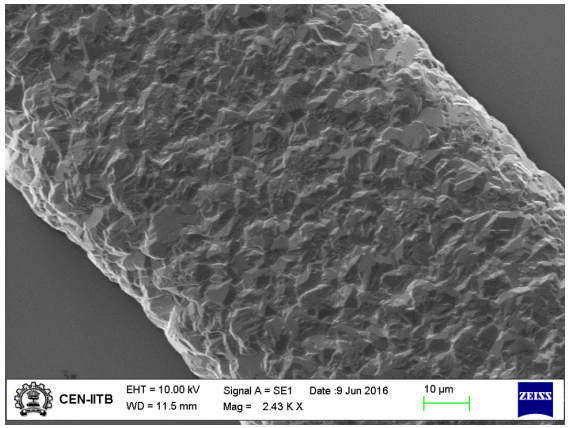


Source: S Kumar and R Sharma, "Analytical Model for Resistivity and Mean Free Path in On-Chip Interconnects with Rough Surfaces", *IEEE Transactions on Emerging Topics in Computing*, 2016.

Surface Roughness in Cu Interconnects



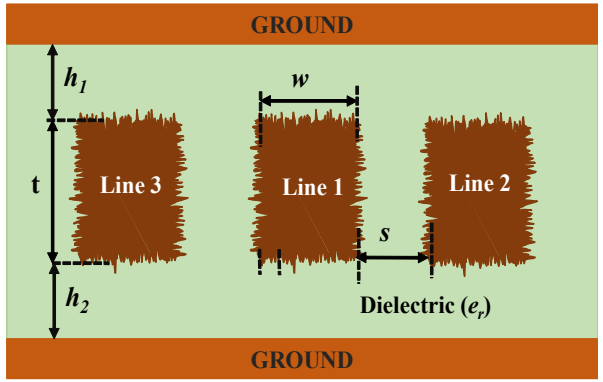


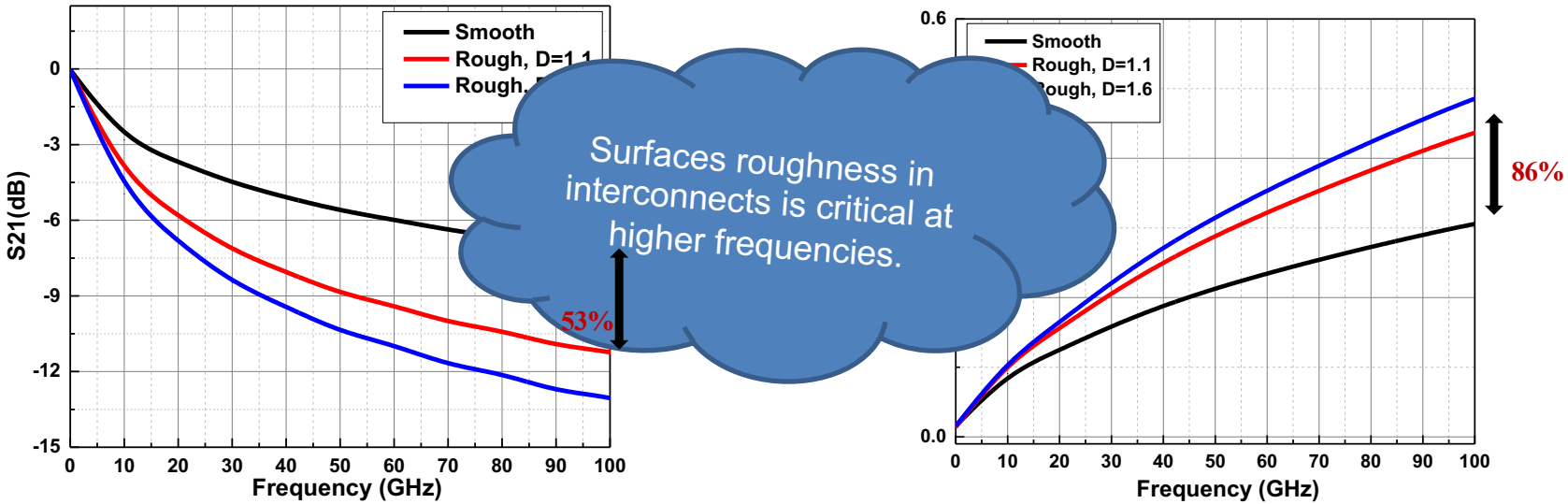


- Surface roughness is a random process characterized by *rms* height, slope, curvature etc.
- Statistical parameters strongly depend on the resolution and scan length of the instrument used to measure the roughness and hence are not unique for a particular surface.
- Fractal approach is used to define the natural rough surface because this approach is instrument independent and is scale invariant.

S. Kumar and R. Sharma, "Analytical Modeling and Performance Benchmarking of On-Chip Interconnects with Rough Surfaces," *IEEE Transactions on Multi-Scale Computing Systems,* 2018.

Surface Roughness is a Problem

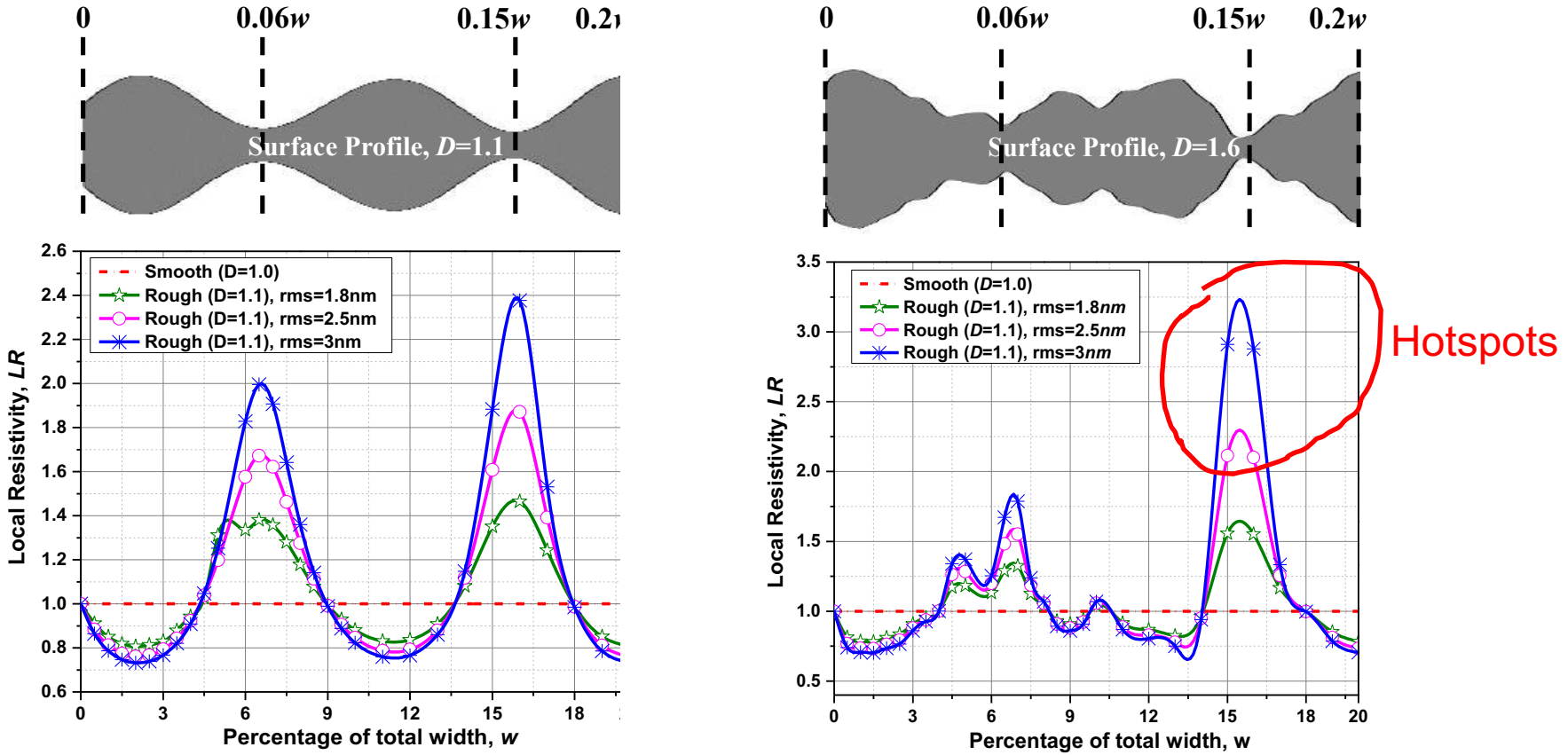




Insertion loss as a function of frequency for a 10*mm* long *Cu* chip to chip interconnect (w = 75μm, t = 8μm, h = 10μm).

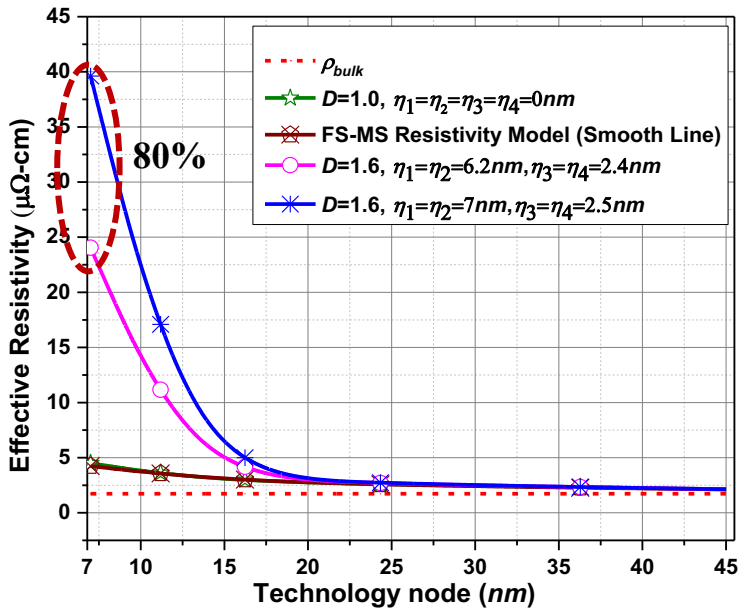
Attenuation as a function of frequency for a 10*mm* long *Cu* chip to chip interconnect (w = 75μm, t = 8μm, h = 10μm).

Variation in Local Resistivity Surface

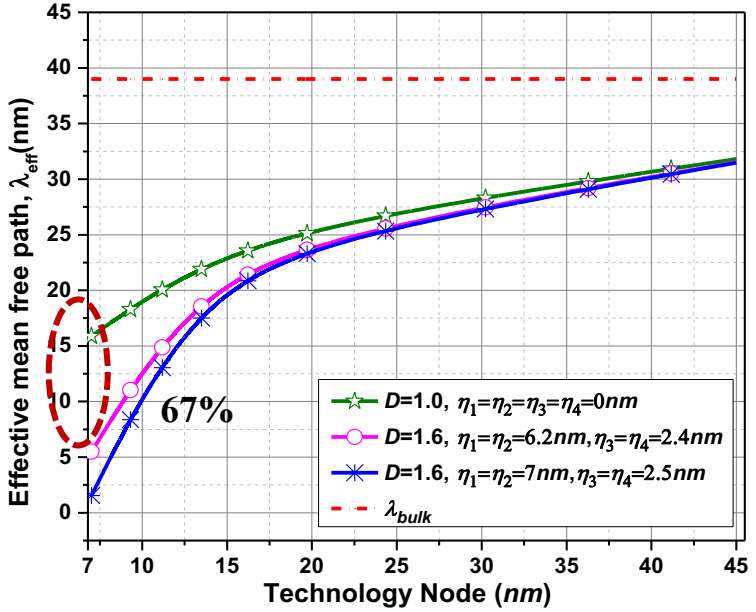


Local resistivity of 7 *nm* local/Intermediate line as a function of width for different RMS value of roughness (a) D = 1.1 and (b) D = 1.6. T = 300 K.

Effective Resistivity and Mean Free Path



(a)

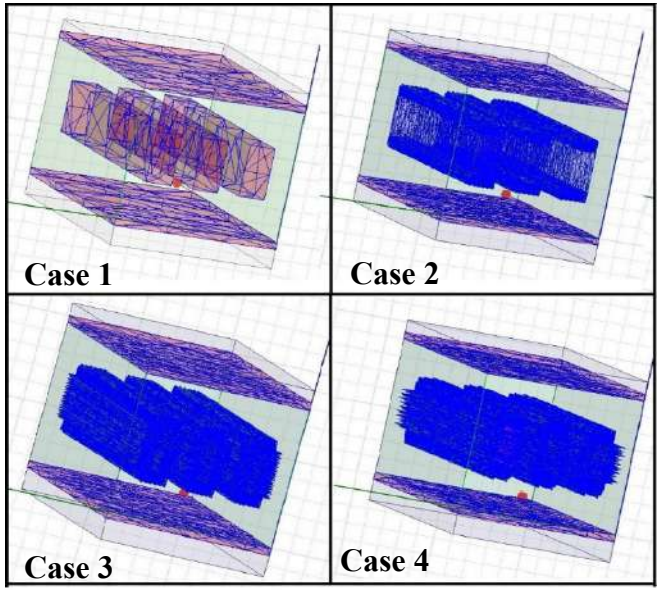


(b)

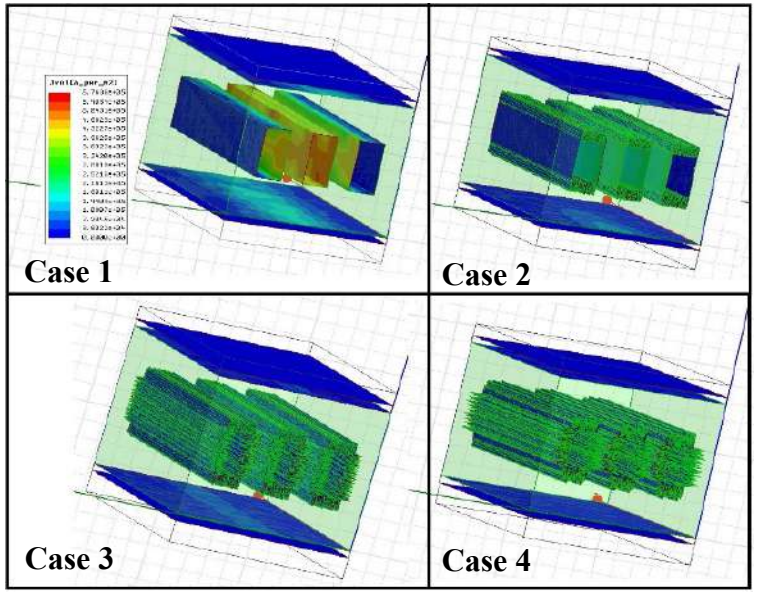
(a) Effective resistivity of Global Interconnects for different technology nodes and different values of *D* (b) Effective mean free path of Global Interconnects for different technology nodes and different values of *D.*

Source: Somesh Kumar and R Sharma, “Chip-to-Chip Copper Interconnects with Rough Surfaces: Analytical models for Parameter Extraction and Performance Evaluation”, IEEE Transactions on Components, Packaging and Manufacturing Technology, 2018.

Computational Complexity

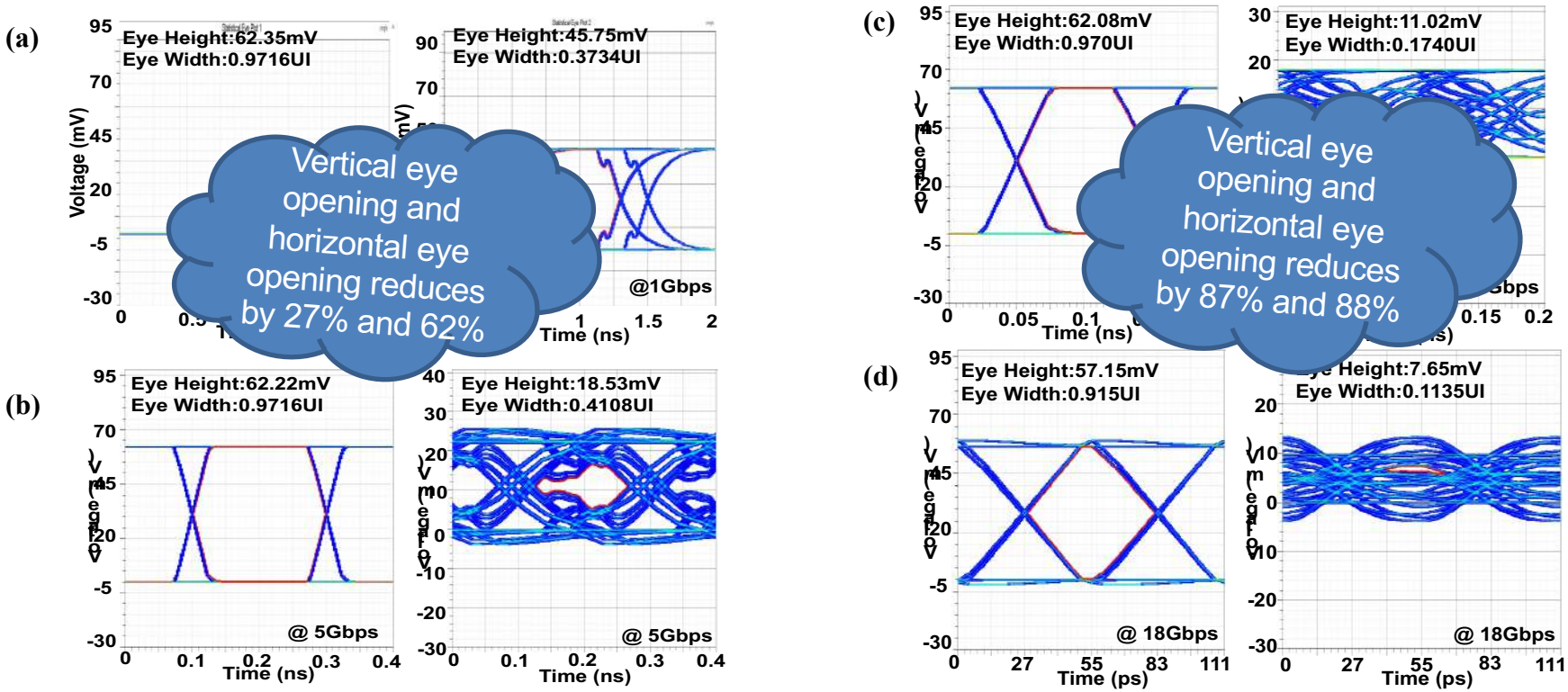


3D view of mesh generated in *HFSS*. (a) Case 1(Smooth line, D = 1.0) (b) Case 2 (Only top and bottom surface rough, D = 1.1) (c) Case 3 (all four surfaces rough, D = 1.1) (d) Case 4 (all four-surface rough, D = 1.6).



3D view of volume current density generated in *HFSS* (a) Case 1 (Smooth line, D = 1.0) (b) Case 2 (Only top and bottom surfaces are rough, D = 1.1) (c) Case 3 (all four surfaces are rough, D = 1.1) (d) Case 4 (all four surfaces are rough, D = 1.6).

Signal Integrity Analysis



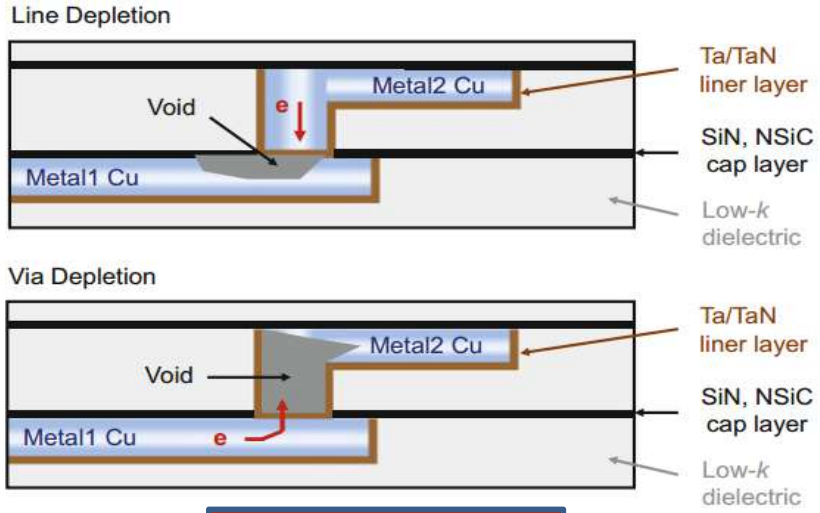
Comparison of eye diagram for smooth and rough 7*nm* technology node on-chip interconnect structure with (a) 1Gbps (b) 5Gbps (c) 10Gbps (d) 18Gbps random data of 2^{11}-1 bits travelling the channel.

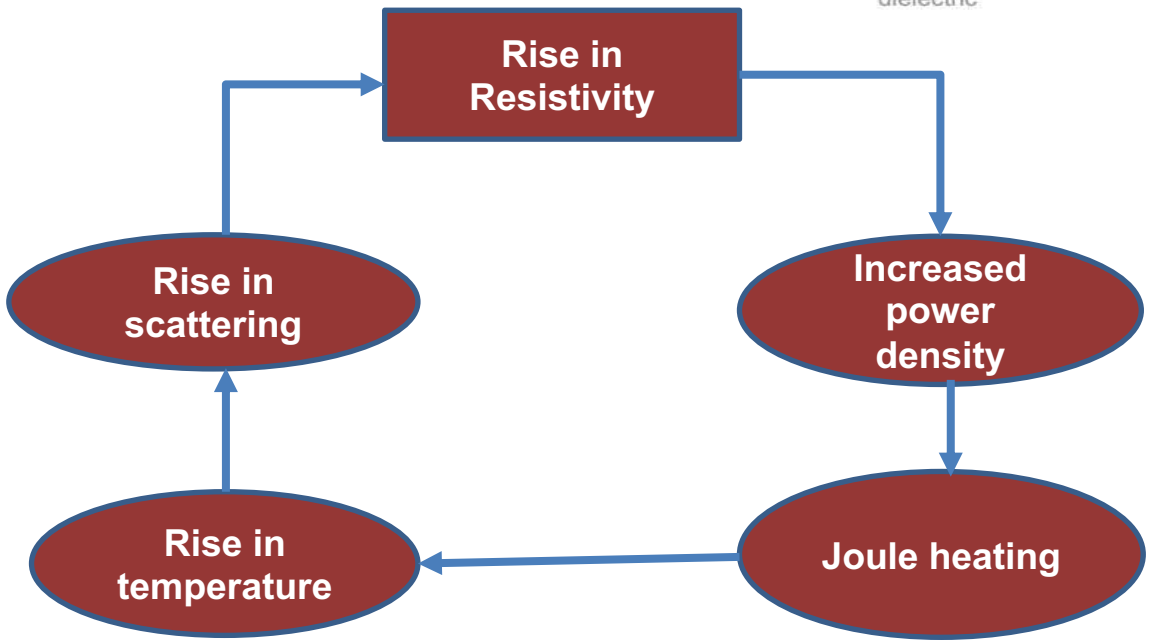
Thermal Impact on Interconnects

Aggressive interconnect scaling has resulted in increasing current densities and associated thermal effects

- Reduced feature sizes
- Inhomogeneities in feature size
- Higher current density
- Joule heating and varying thermal profiles

Thermal effects in interconnects have become a serious performance and reliability constraint



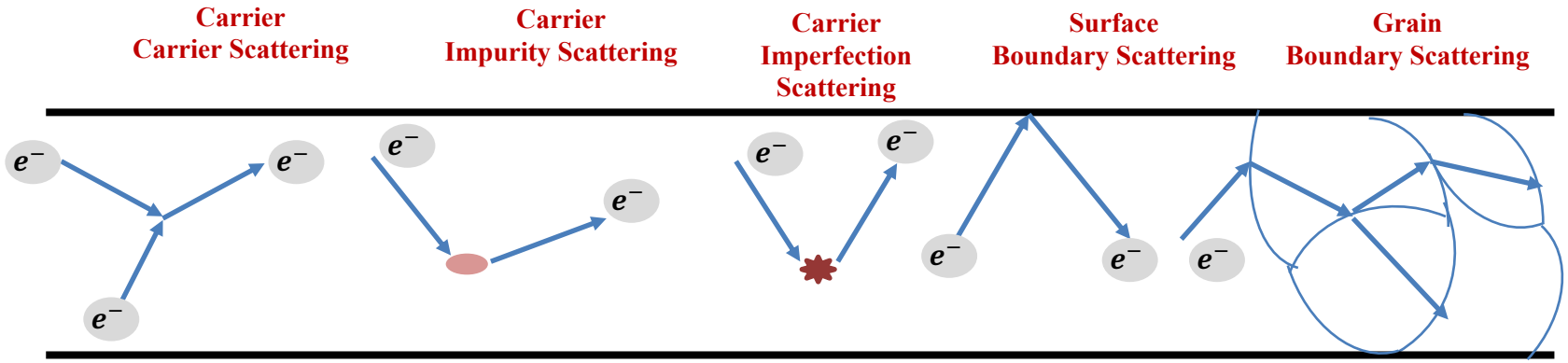


Source: T. Gupta, *Copper Interconnect Technology*, Springer

Causes of Thermal Issues

Electrical and thermal conductivity of thin films reduce due to

- Aggressive carrier scattering
- Carrier-carrier scattering, carrier impurity scattering, and carrier imperfection scattering are combinedly called bulk scattering.



Different scattering mechanism in Cu

With scaling thermal conductivity of materials reduces due to more impurity, disorder and grain boundaries.

With scaling thermal conductivity of materials reduces due to boundary scattering, phonon leakage, and related interactions.

- Generally affects the in-plane and cross plane thermal transport.

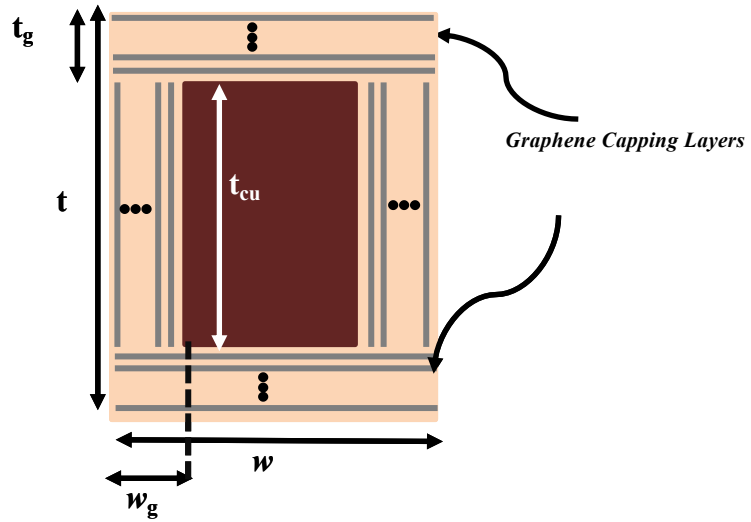
Copper Graphene Hybrid Interconnects

Conventional barrier layer possesses very high resistivity leading to creation of void and hillocks.

Graphene and Cu in some sort of hybrid heterogeneous structure can bring potential benefits of reducing Cu electromigration and diffusion.

Graphene as a barrier layer changes the diffusion path from surface to grain boundary.

Graphene as a barrier layer enhances the elastic surface scattering.



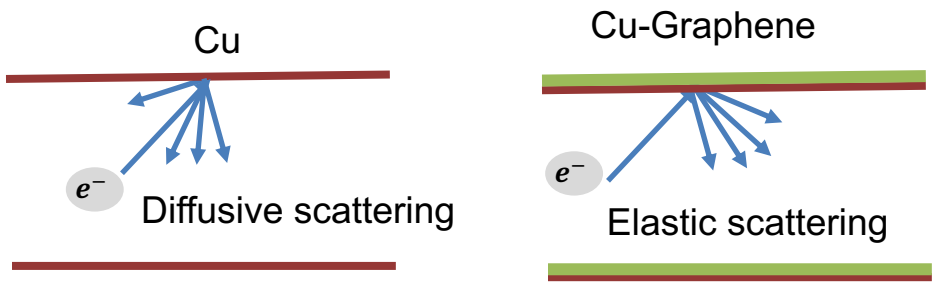
Cu line with diffusion barrier

Resistivity for different barrier layers at 22, 13, and 7 nm

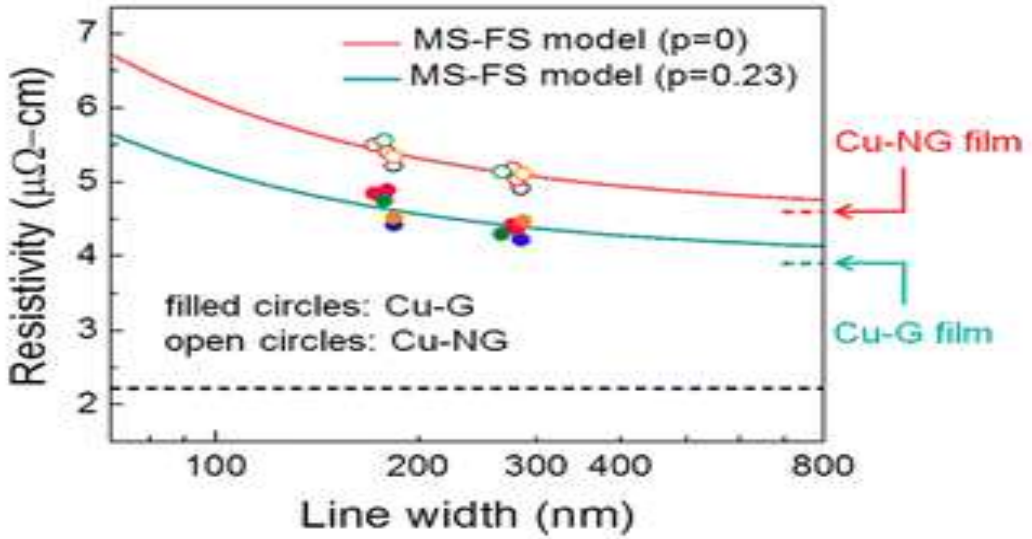
Thickness	2 nm	1.2 nm	0.6 nm
Ta Barrier Layer (ρ_{Ta})	278.6×10^{-8}	433.2×10^{-8}	852.4×10^{-8}
W Barrier Layer (ρ_{W})	65.1×10^{-8}	100×10^{-8}	194.2×10^{-8}
MLGNR Barrier Layer (ρ_{MLG})	**8×10^{-8}**	**10.22×10^{-8}**	**14×10^{-8}**

R Kumar and R Sharma, "A Temperature and Dielectric Roughness-Aware Matrix Rational Approximation Model for the Reliability Assessment of Copper– Graphene Hybrid On-Chip Interconnects," *IEEE Trans. on Components, Packaging and Manufacturing Technology*, 2020.

Copper Graphene Hybrid Interconnects



Surface scattering

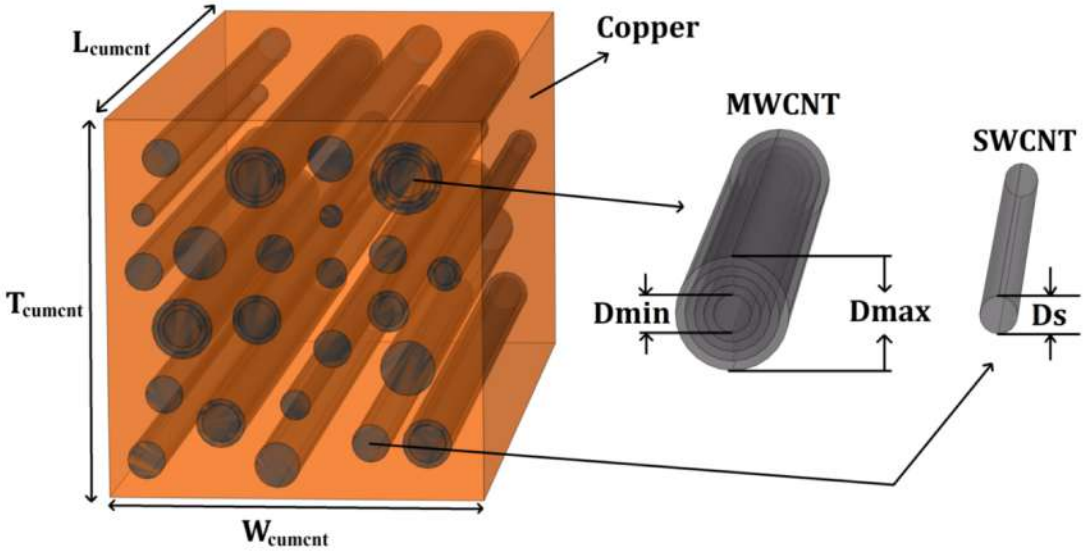


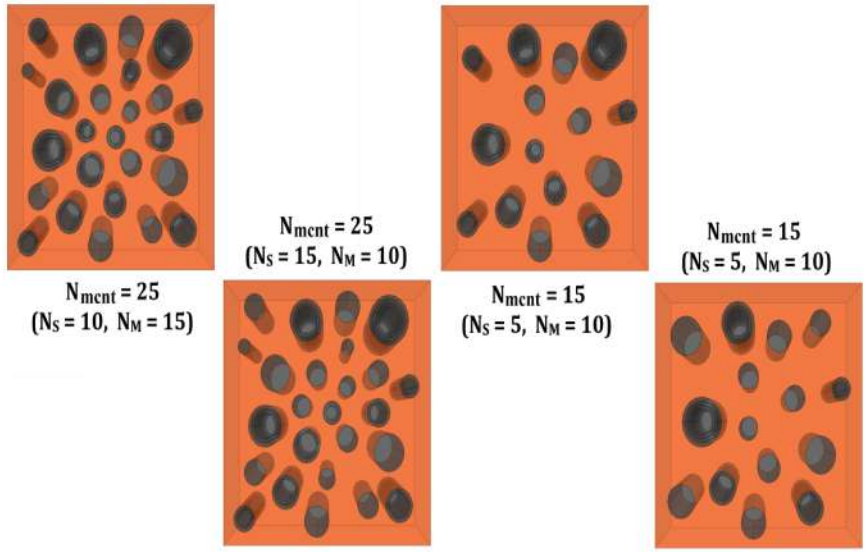
Resistivity w.r.t line width for different specularity constant

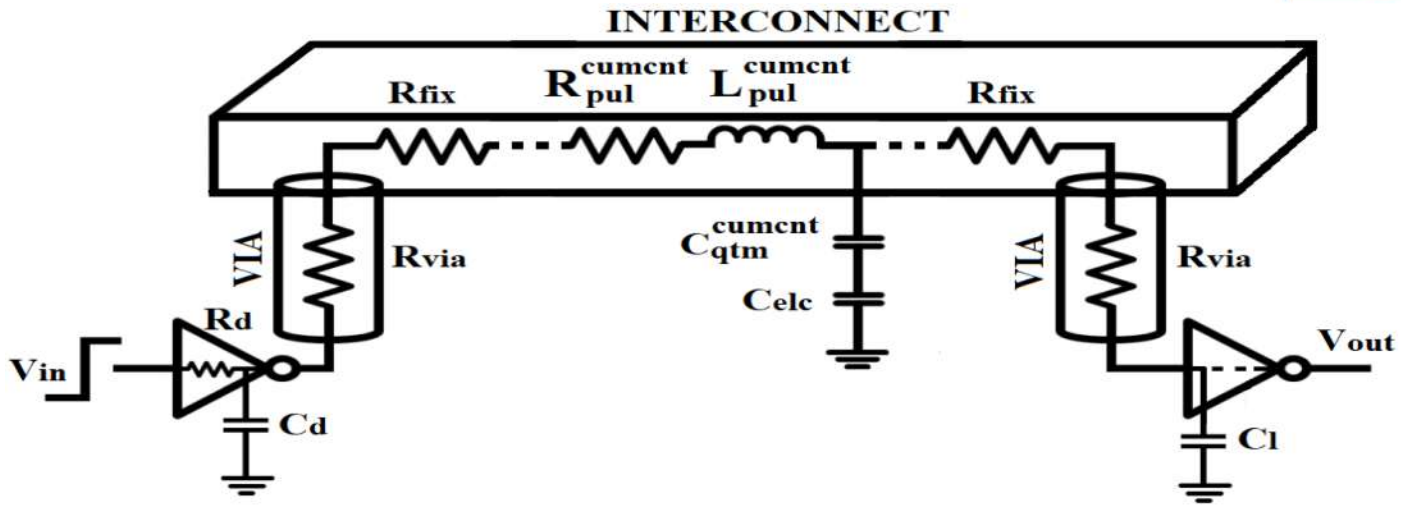
- The EM activation energies for pure Cu and graphene capped Cu are found to be 0.76 eV and 1.23 eV.
- This indicates that the Graphene barrier layer successfully suppresses the surface migration of Cu atoms.
- Dominant diffusion path of Cu atoms is shifted from the surface to the grain boundaries of the Cu line.
- Grown Graphene over Cu results in morphological change in Cu lines.
- As the Graphene is grown over Cu grain sizes increases substantially.
- Increment in grain size leads to reduction in grain boundary scattering.

R. Mehta, S. Chugh, and Z. Chen, "Enhanced Electrical and Thermal Conduction in Graphene-Encapsulated Copper Nanowires," *Nano Letters*, 2015.

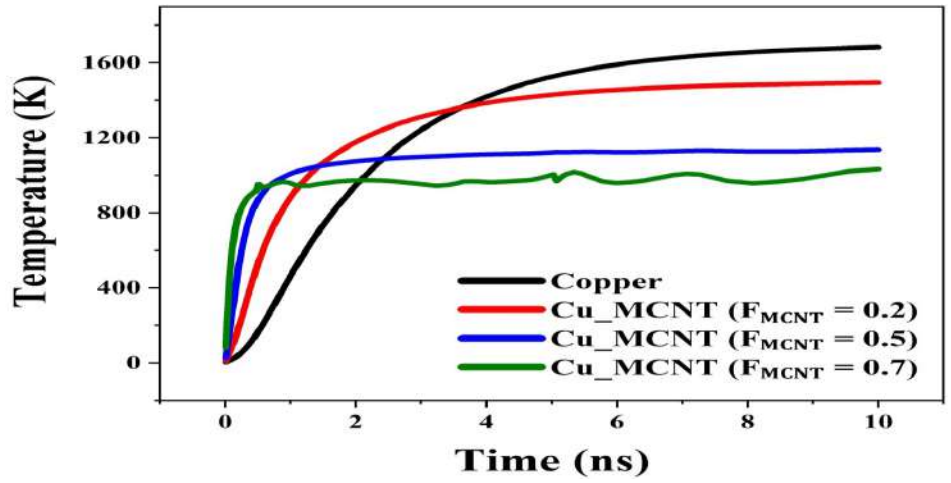
Copper CNT Hybrid Interconnects

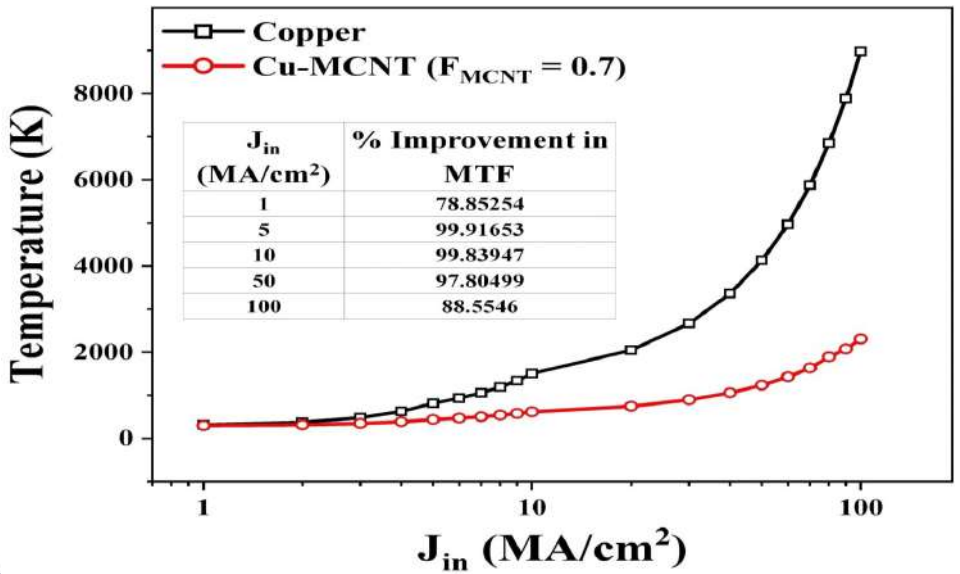




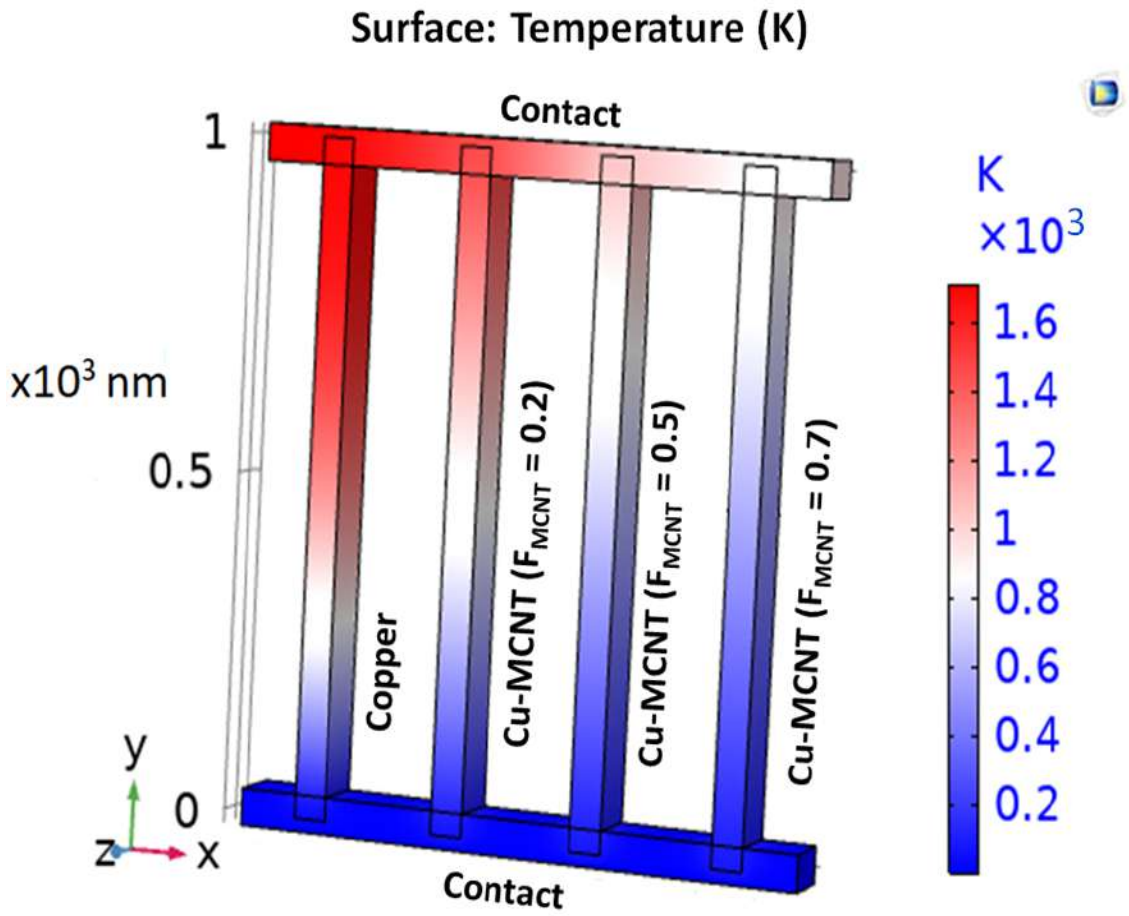


Copper CNT Hybrid Interconnects

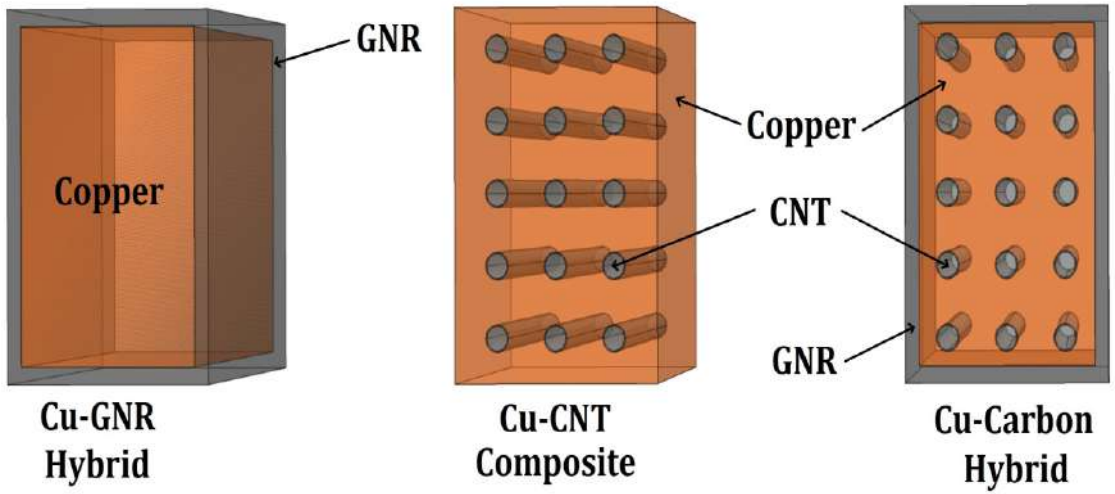


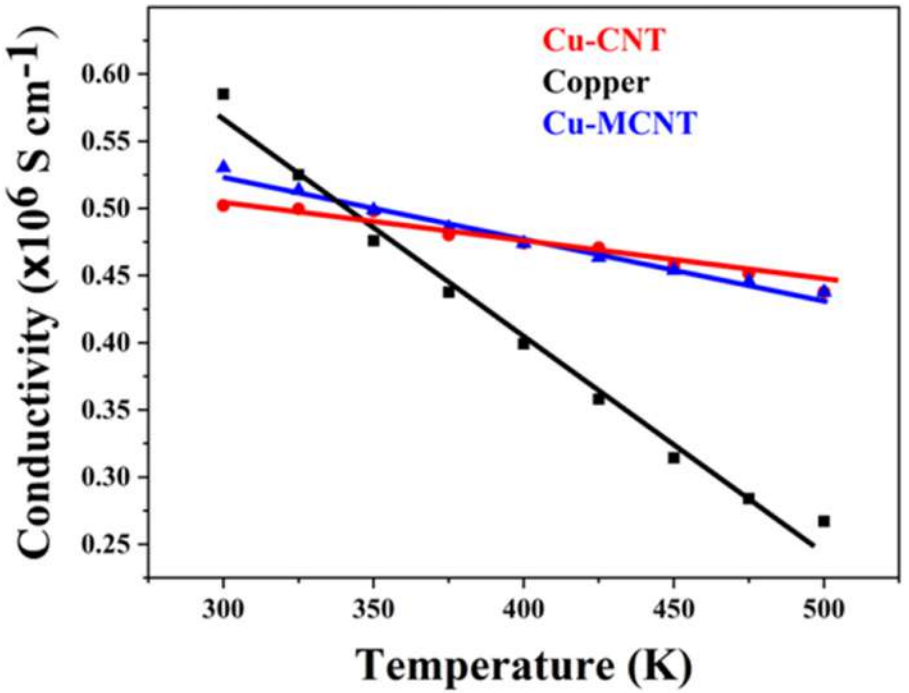


J_{in} (MA/cm²)	% Improvement in MTF
1	78.85254
5	99.91653
10	99.83947
50	97.80499
100	88.5546



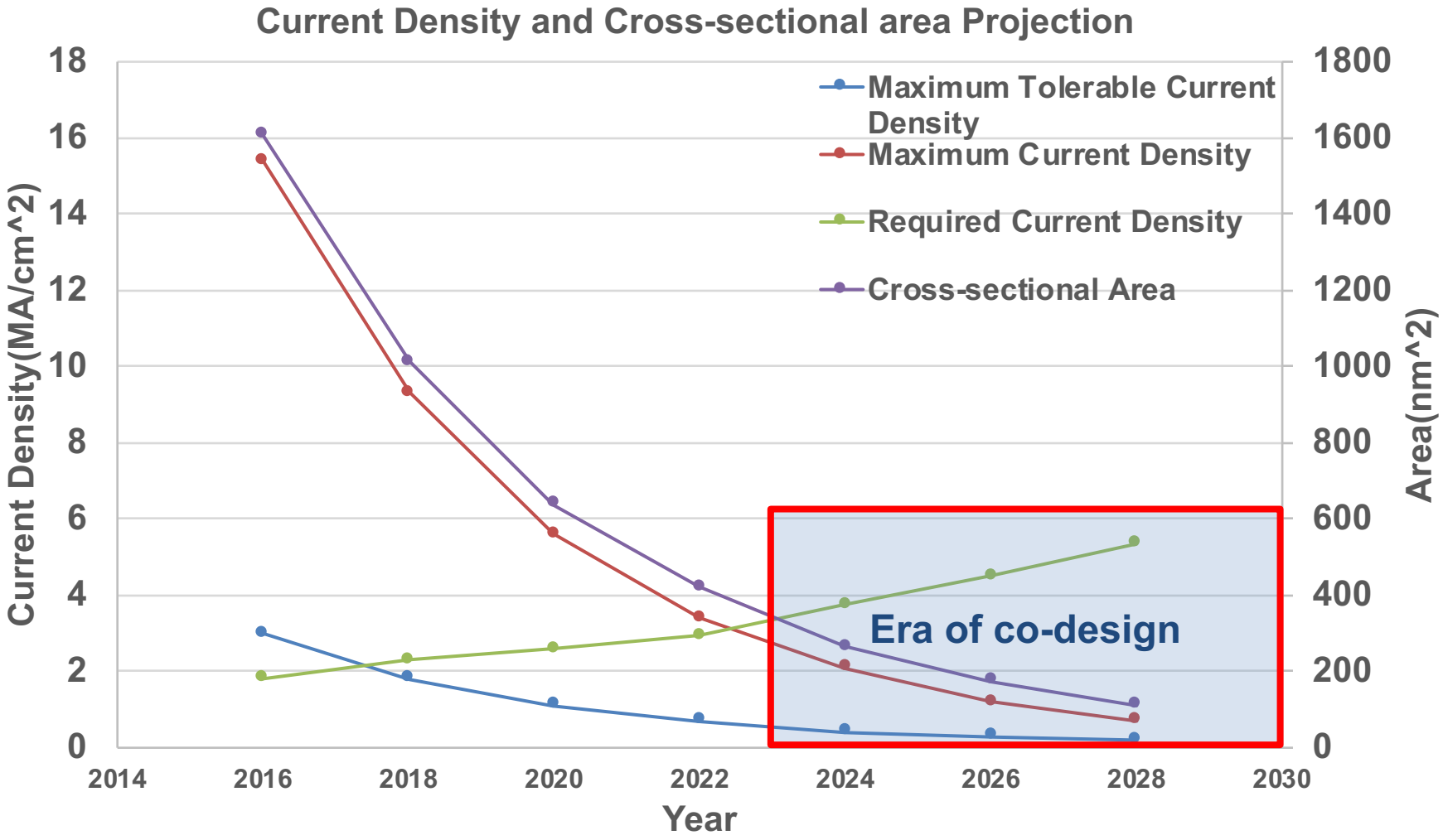
Copper CNT Hybrid Interconnects





- **Cu-GNR Hybrid Interconnect:** Graphene acts as the barrier layer over copper interconnects.
- **Cu-CNT Composite Interconnect:** Copper is electrodeposited over a bundle of CNTs
- **Cu-Carbon Hybrid Interconnect:** Graphene acts as the barrier layer over Cu-GNR interconnects.

Revisiting the Interconnect Bottleneck



International Technology Roadmap for Semiconductors, 2015. [Online]. Available: http://www.itrs2.net/itrs-reports.html

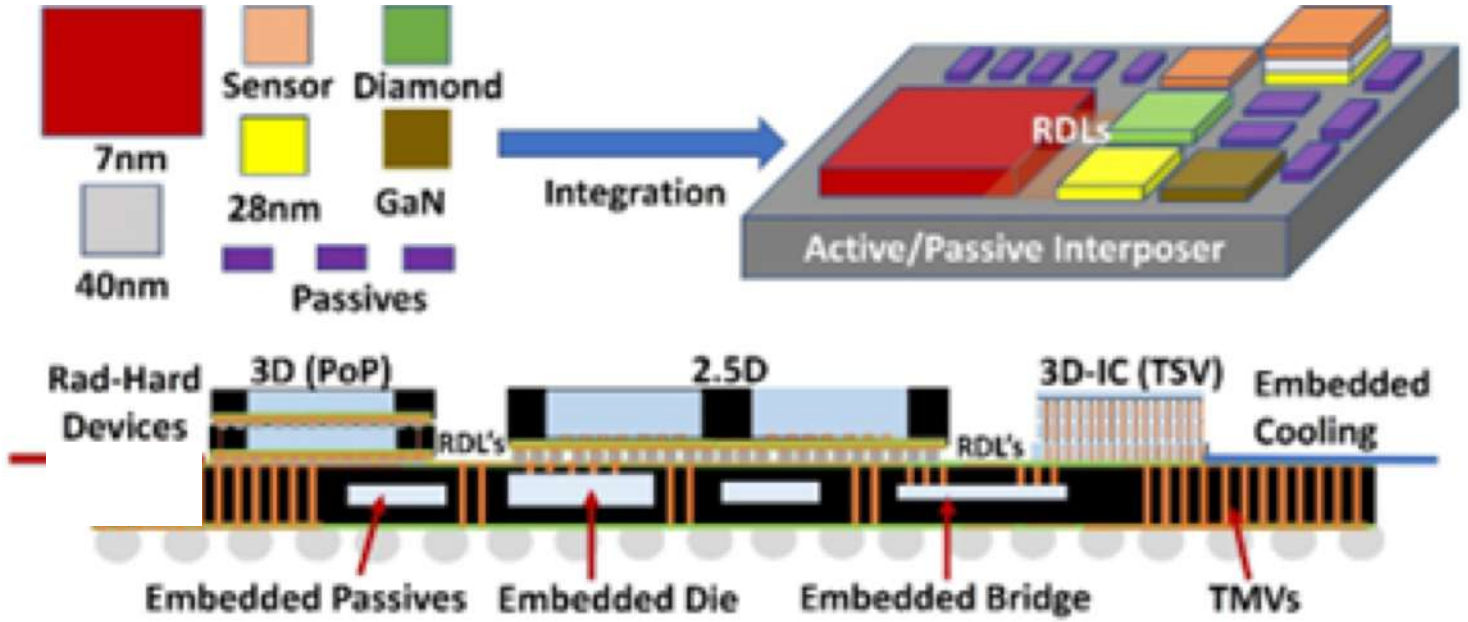
Co-design for Heterogeneous Integration

Design

Materials

Processes

Environments



Device Packaging	Bonding	Interposers	Interconnects
Wafer Level Packaging	Flip Chip	Organics	C4, Microbump's, Cu-Pillar
3D PoP, 3D-IC (Die Stack)	Thermo-Compression	Silicon	RDL's, TxVs, Optical
SiP, QFN, BGA, Hybrids…	Hybrid Direct Bonding	Ceramics	Cu-Cu (bumpless)

Electrical

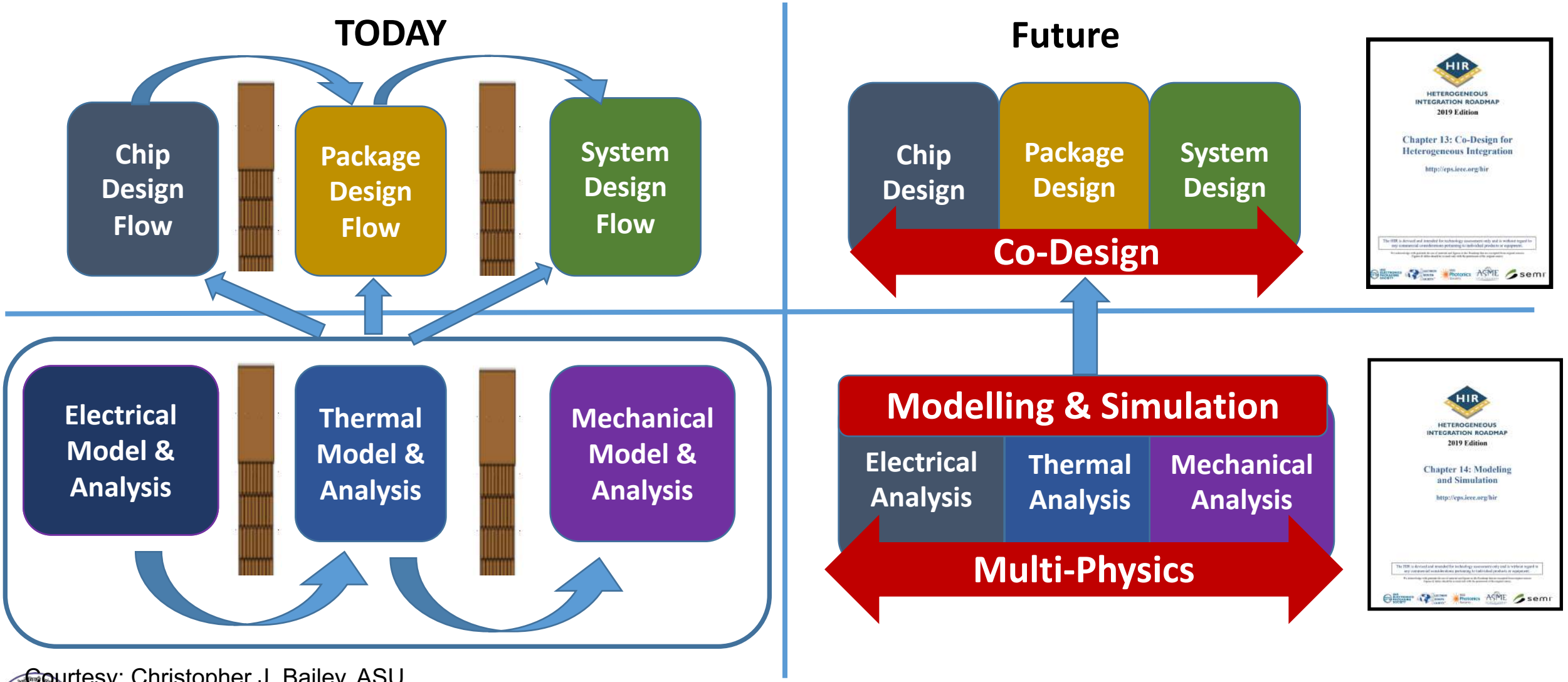
Thermal

Mechanical

Reliability

Courtesy: Christopher J. Bailey, ASU

Co-design for Heterogeneous Integration



Courtesy: Christopher J. Bailey, ASU

Uncertainty Quantification

Uncertainty propagation

Uncertain parameters (Temperature, geometrical dimensions etc.)

Output becomes a random variable!!

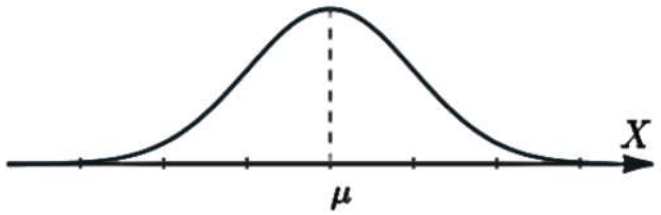
Variability in temperature

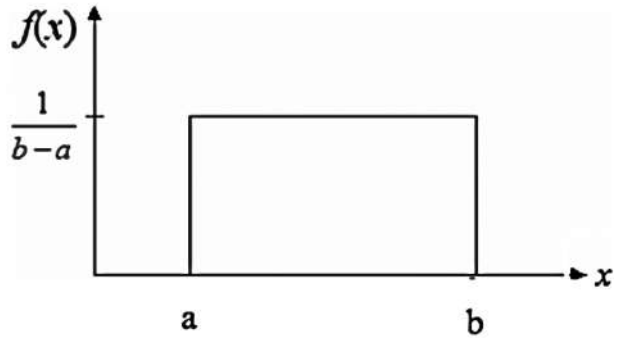
Systematic

Stochastic

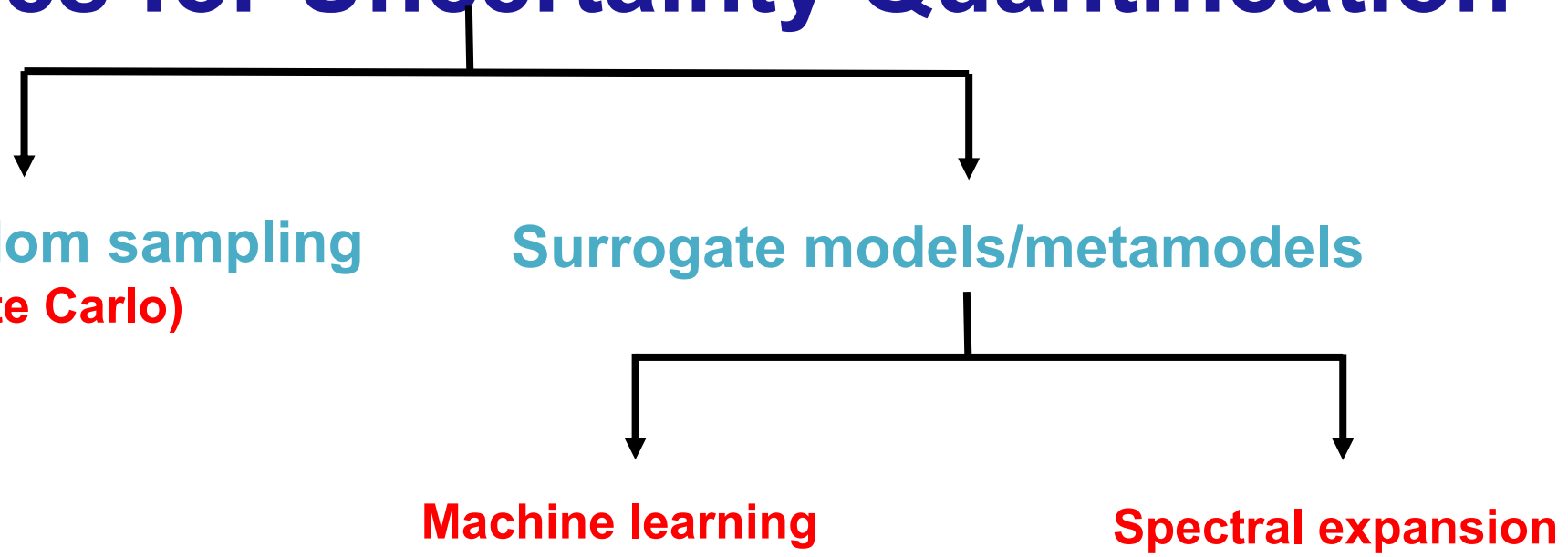
$[T_{min}, T_{max}]$

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$





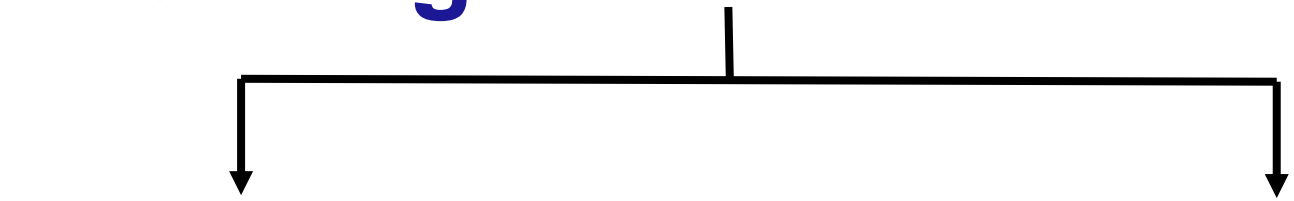
Methodologies for Uncertainty Quantification



Targets of uncertainty quantification:

- To evaluate mean and variance of the outputs.
- To evaluate the reliability of the outputs.
- To assess the complete probability distribution of the outputs.

Surrogate Models/Metamodels



Machine learning

- Artificial neural networks
- Support vector machines

Spectral expansion

- Karhunen-Loeve expansion
- Generalized polynomial chaos expansion

✓ Identification of trends and patterns

✓ Scope of improvement

× Data acquisition

× Algorithm selection

× High error susceptibility

✓ Optimal accuracy

✓ Time efficient

× Curse of dimensionality

× Uses polynomials – cannot capture high non-linearity

Summary

- The growth of semiconductor industry will be governed by aggressive energy aware design
  - Miniaturization will cease to pay dividends
- Functional 2D materials offer promising prospects; however, their acceptance by industry is thwarted by process constraints
- Surface roughness in *Cu* global interconnects needs to be address at higher frequencies and lower technology nodes
- Heterogeneous integration of packages and systems will require greater effort for co-design and co-analysis
  - Electrical, thermal and mechanical issues will coexist in next generation packages
  - Continue to pose serious challenges to overall system reliability

Thank You ☺

Q&A Session