ECE 546
Lecture -26
Packaging Technologies

Spring 2026

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

System-Level Integration (Microelectronic Packaging)

Semiconductor

* Unprecedented Innovations in CMOS, Si-Ge,Copper Wiring
* Fundamental technical Limits

Electronic Systems

* Computers, telecom & Consumer Products Merge
* Portable, Wireless, & Internet Accessible
* Very Low Cost & Very High Performance

Microelectronic Packaging

* High Cost, Low Performance, Low Reliability
* Lack of Skilled Human Resources

Packaging Challenges

- Package is bottleneck to system performance

- Package cost is increasing percentage of system cost

- Package limits IC technology

- On-chip system can outperform package capability

Advantages of SOC

* Fewer Levels of Interconnections
* Reduced Size and Weight
* Merging of Voice, Video, Data,...

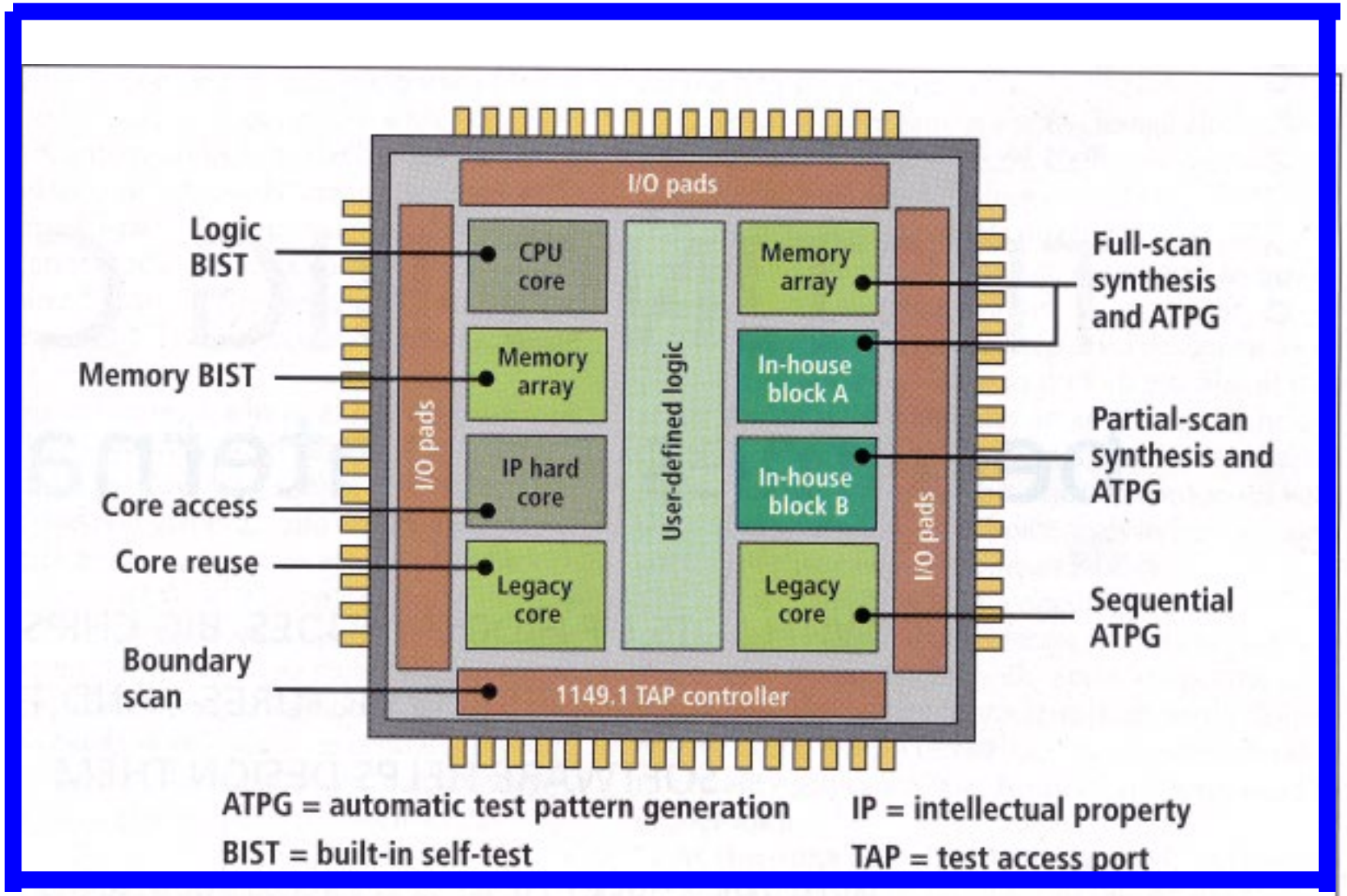
Arguments against SOC

* Challenges too Big
* Legal issues

Challenges for SOC

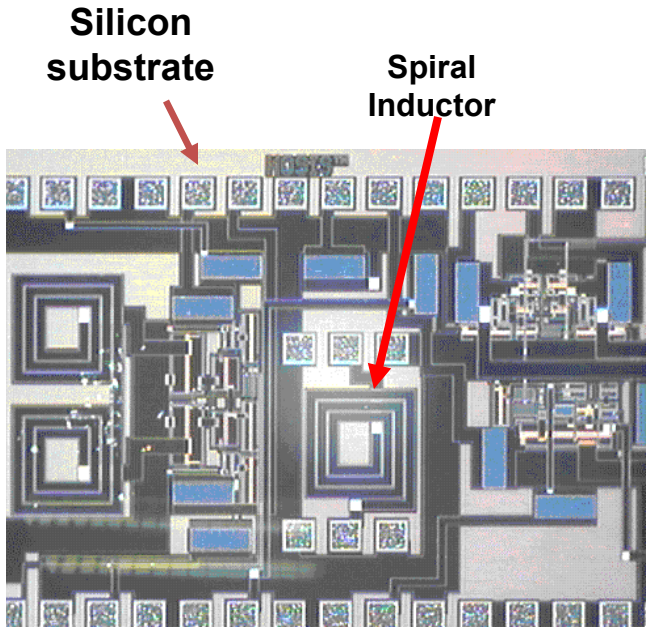
* Different Types of Devices
* Single CMOS Process for RF and Digital
* Design Methodology not available
* EDA Tools cannot handle level of complexity
* Intellectual Property
* Signal Integrity
* High-Power Requirements of PA

System on a Chip (SOC)



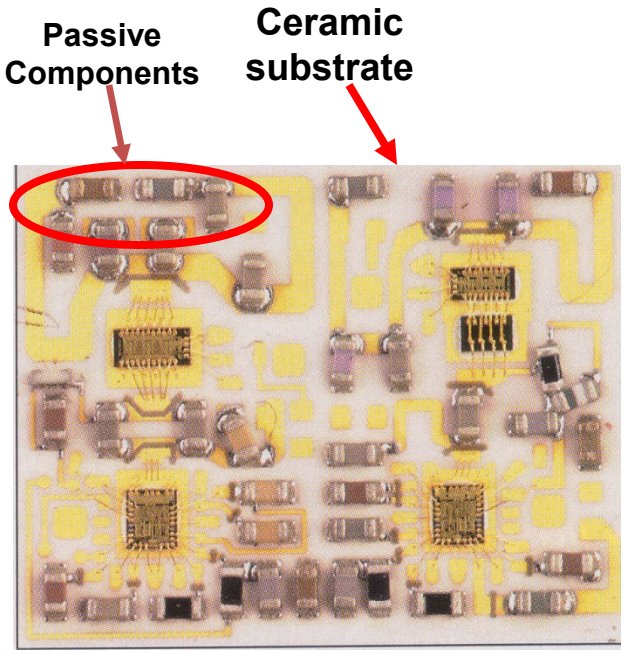
SOC vs SOP

System on Chip



Voltage Controlled Oscillator
(UIUC-CAD group – 1999)

System on Package

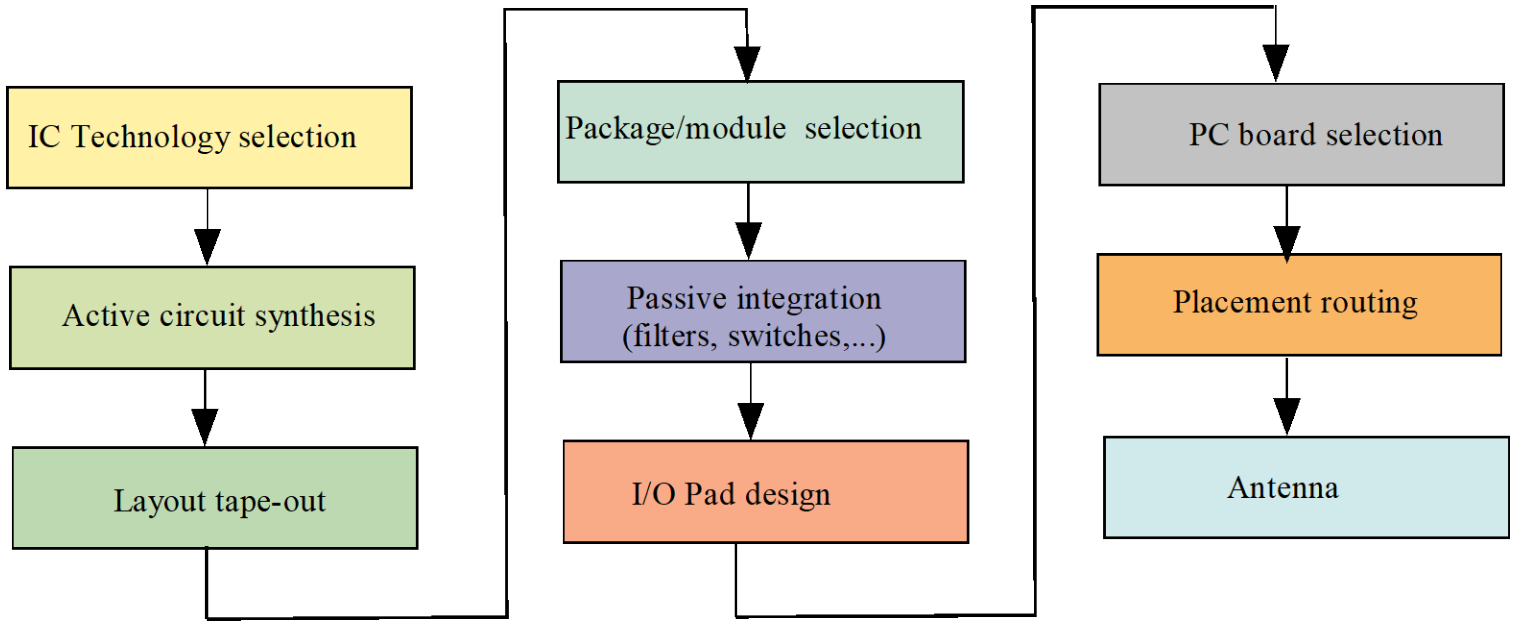


Triple-band GSM/EDGE Power Amp Module
(RF Design Magazine – 4/02)

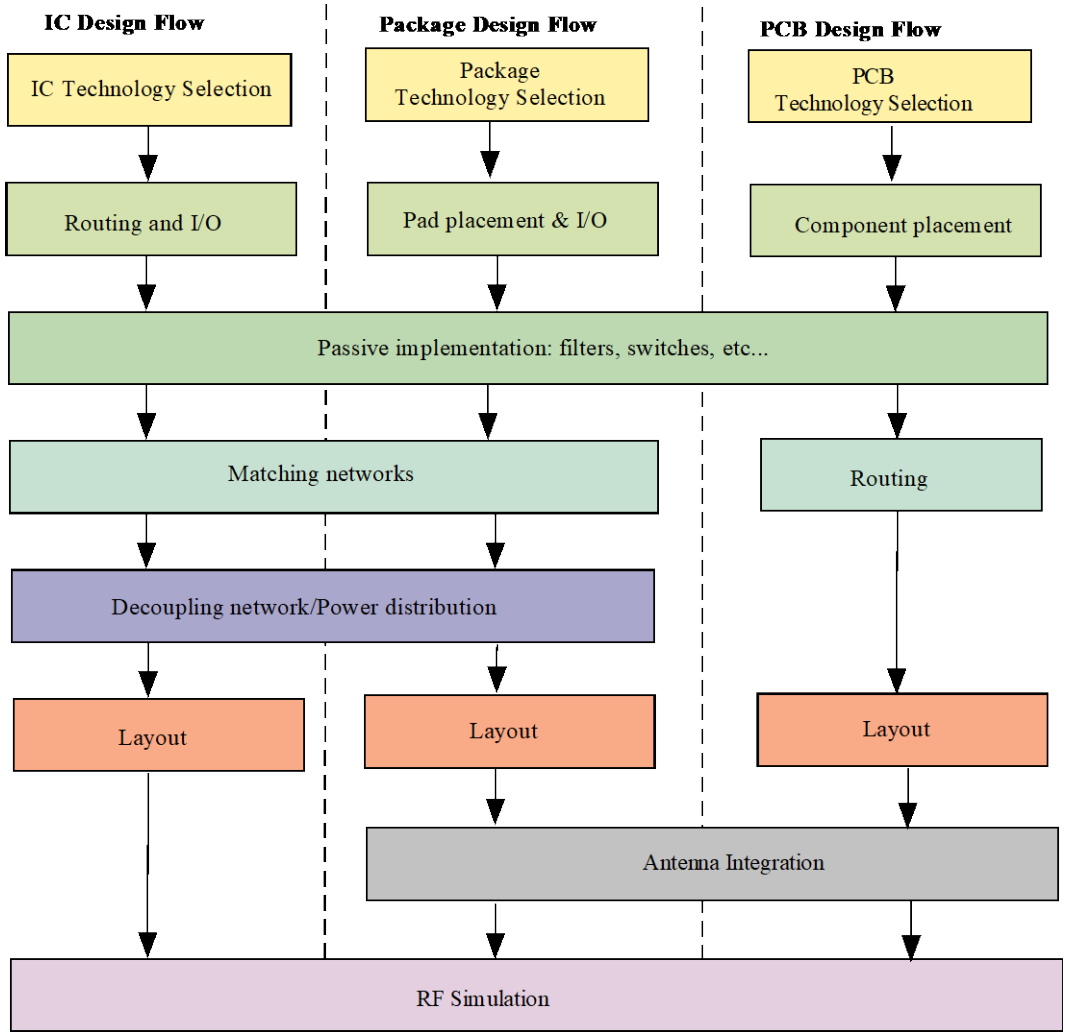
SOP vs SOC

	SOP	SOC
Low cost consumer products (<$200)	YES	YES
Portable products ($200-$2000)	YES	NO
Single processor products ($1-$5K)	YES	NO
High Performance Products (>5K)	YES	NO
Automotive and Space Applications	YES	NO

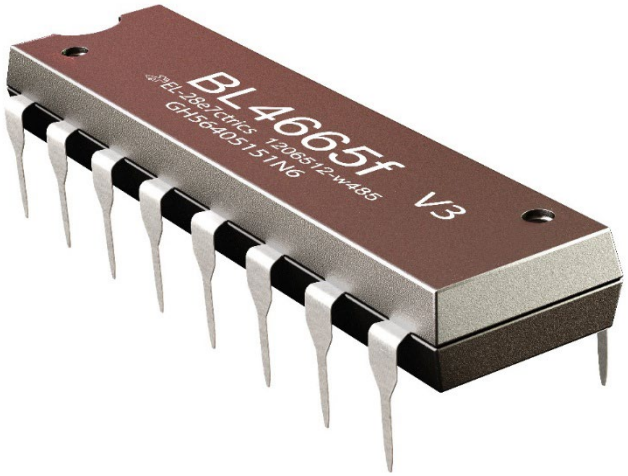
Traditional Design Flow

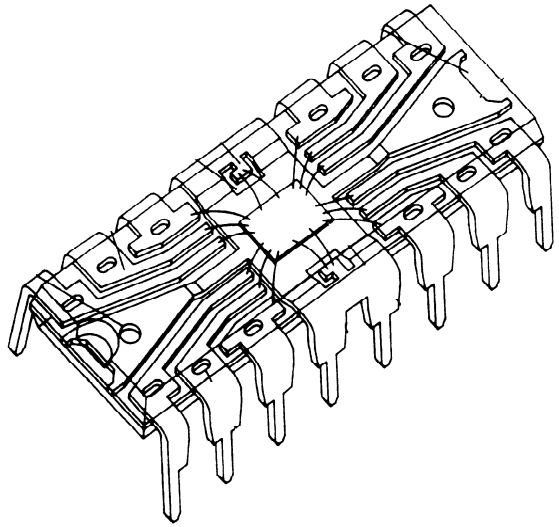


Co-Design Flow



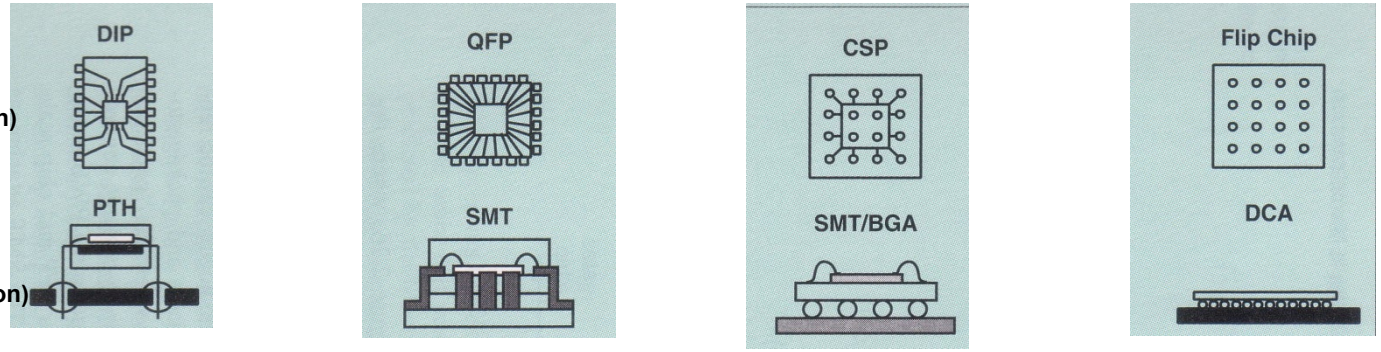
Dual-in-Line (DIP) Package





- Mounted on PWB in pin-through-hole (PTH) configuration
- Chip occupies less than 20% of total space
- Lead frame with large inductance

Package Types (1995-2005)



	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

Substrate Materials

Material	Surface roughness (μm)	$10^4 \tan\delta$ at 10 GHz	ε_r	Thermal conductivity K ($W/cm^2/^oC$)	Dielectric strength (kV/cm)
Air (dry)	N/A	~0	1	0.00024	30
Alumina: 99.5% 96% 85%	 0.05-0.25 5-20 30-50	 1-2 6 15	 10.1 9.6 15	 0.37 0.28 0.2	 4×10^3 4×10^3 4×10^3
Sapphire	0.005-0.025	0.4-0.7	9.4,11.6	0.4	4×10^3
Glass, typical	0.025	20	5	0.01	-
Polyimide	-	50	3.2	0.002	4.3

Substrate Materials

Material	Surface roughness (μm)	$10^4 \tan\delta$ at 10 GHz	ε_r	Thermal conductivity K ($W/cm^2/^oC$)	Dielectric strength (kV/cm)
Irradiated polyolefin	1		2.3	0.001	~300
Quartz (fused) i.e. SiO2	0.006-0.025	1	3.8	0.01	10×10^3
Beryllia	0.05-1.25	1	6.6	2.5	-
Rutile	0.25-2.5	4	100		-
Ferrite/garnet	0.25	2	13-16	0.03	4×10^3

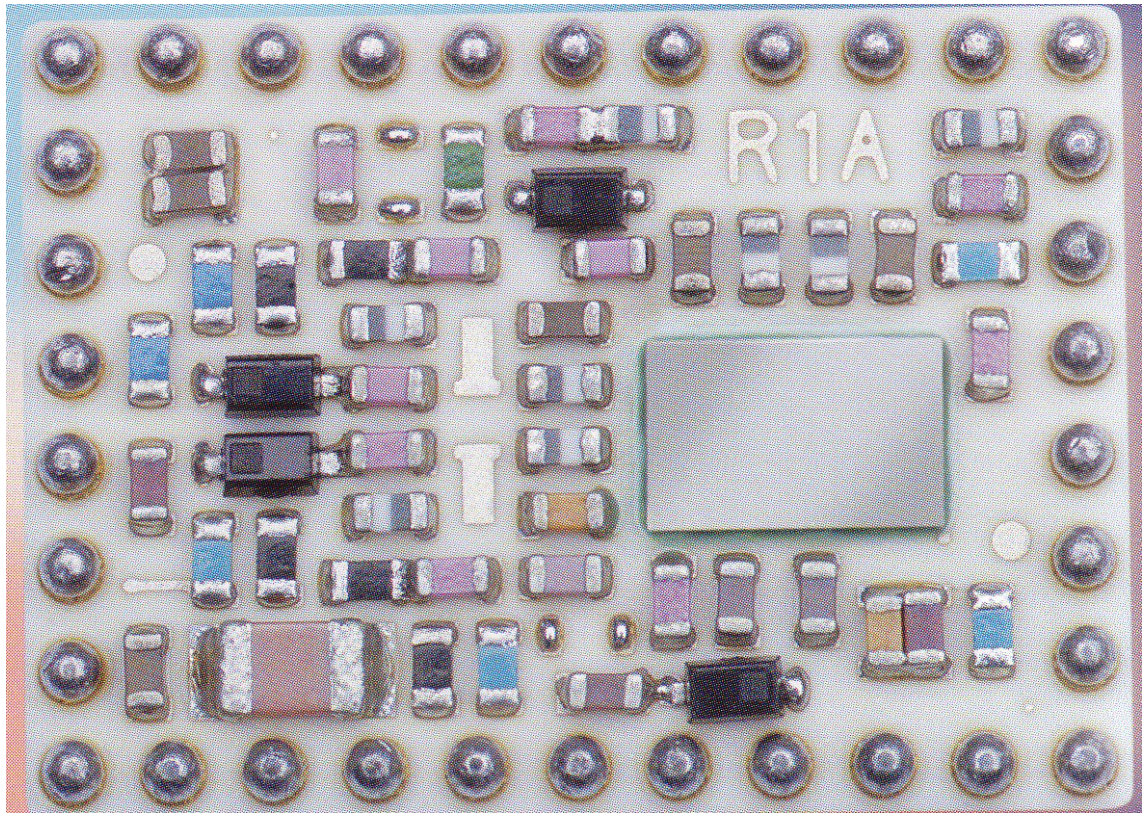
Substrate Materials

Material	Surface roughness (μm)	$10^4 \tan\delta$ at 10 GHz	ε_r	Thermal conductivity K ($W/cm^2/^oC$)	Dielectric strength (kV/cm)
FR4 circuit board	~6	100	4.3-4.5	0.005	-
RT-duroid 5880	0.75-1 4.25-8.75	5-15	2.16-2.24	0.0026	-
RT-duroid 6010	0.75-1 4.25-8.75	10-60	10.2-10.7	0.0041	-
AT-1000	-	20	10.0-13.0	0.0037	-
Cu-flon	-	4.5	2.1	-	-

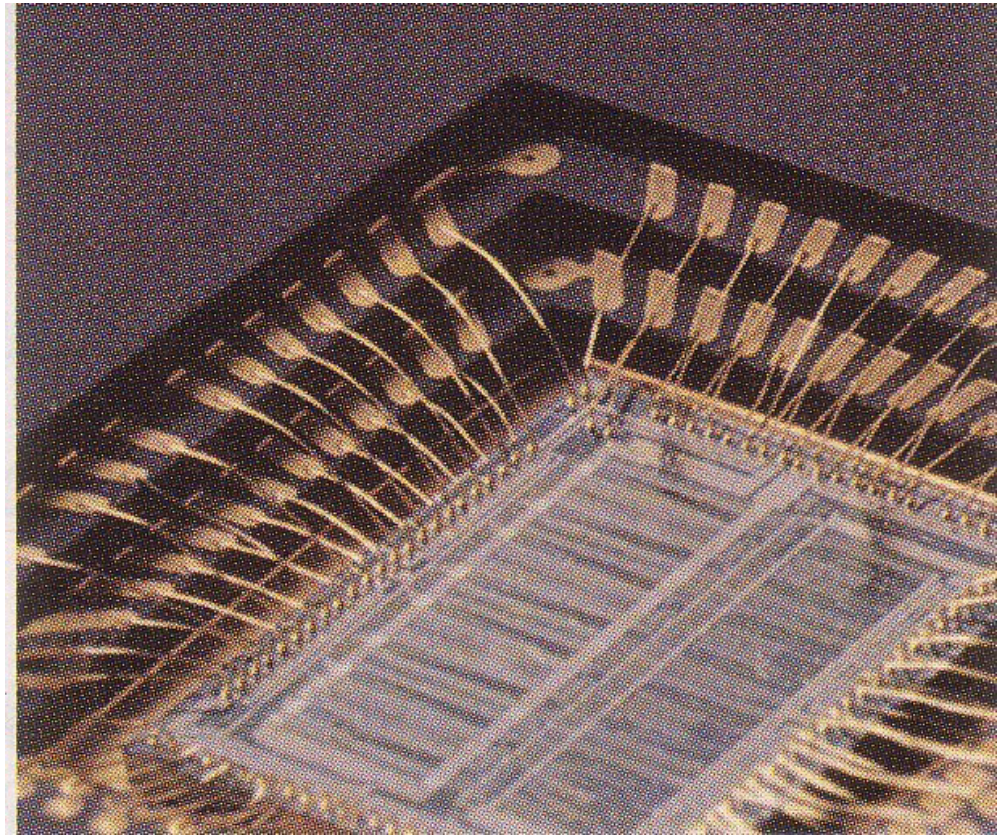
Substrate Materials

Material	Surface roughness (μm)	$10^4 \tan\delta$ at 10 GHz	ε_r	Thermal conductivity K ($W/cm^2/^oC$)	Dielectric strength (kV/cm)
Si (high resistivity)	0.025	10-100	11.9	0.9	300
GaAs	0.025	6	12.85	0.3	350
InP	0.025	10	12.4	0.4	350
SiO2 (on chip)	-	-	4.0-4.2	-	-
LTCC (typical green tape 951)	0.22	15	7.8	3	400

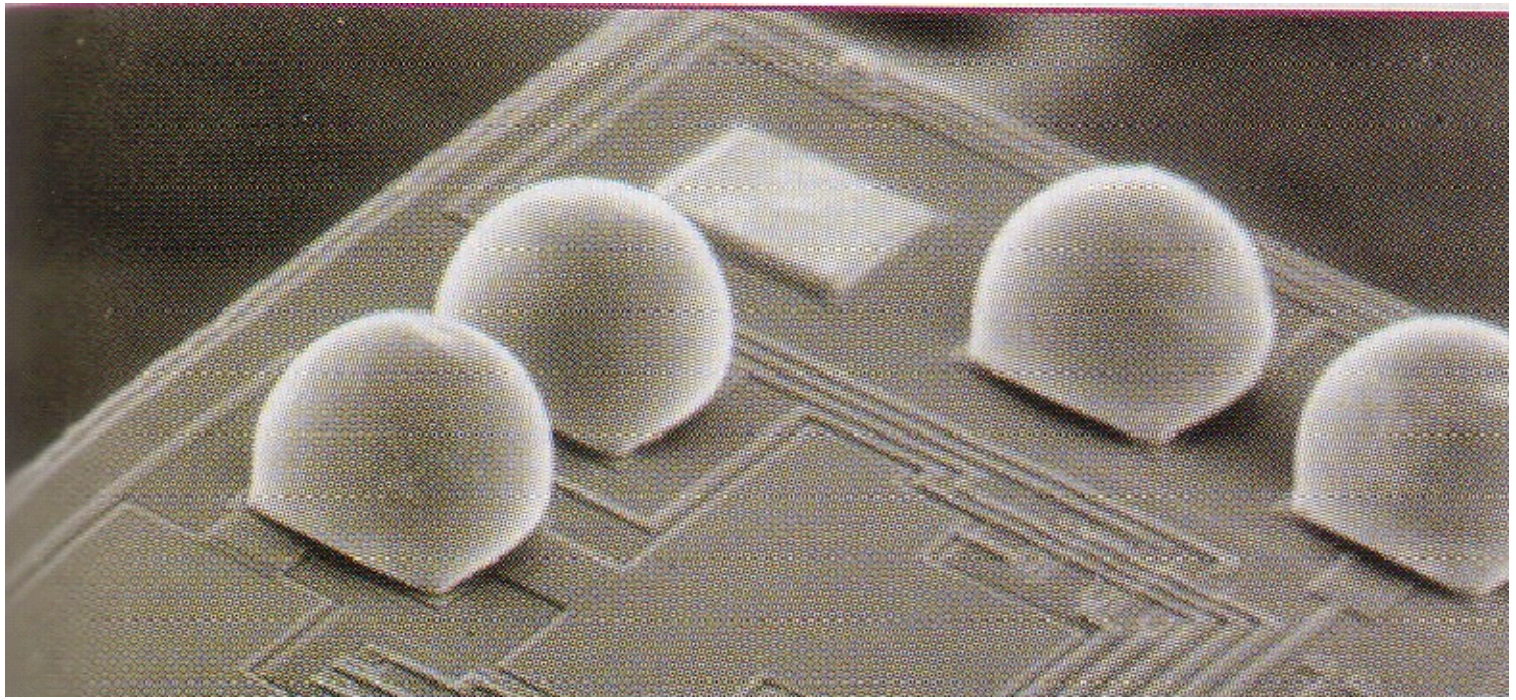
Ceramic Substrate



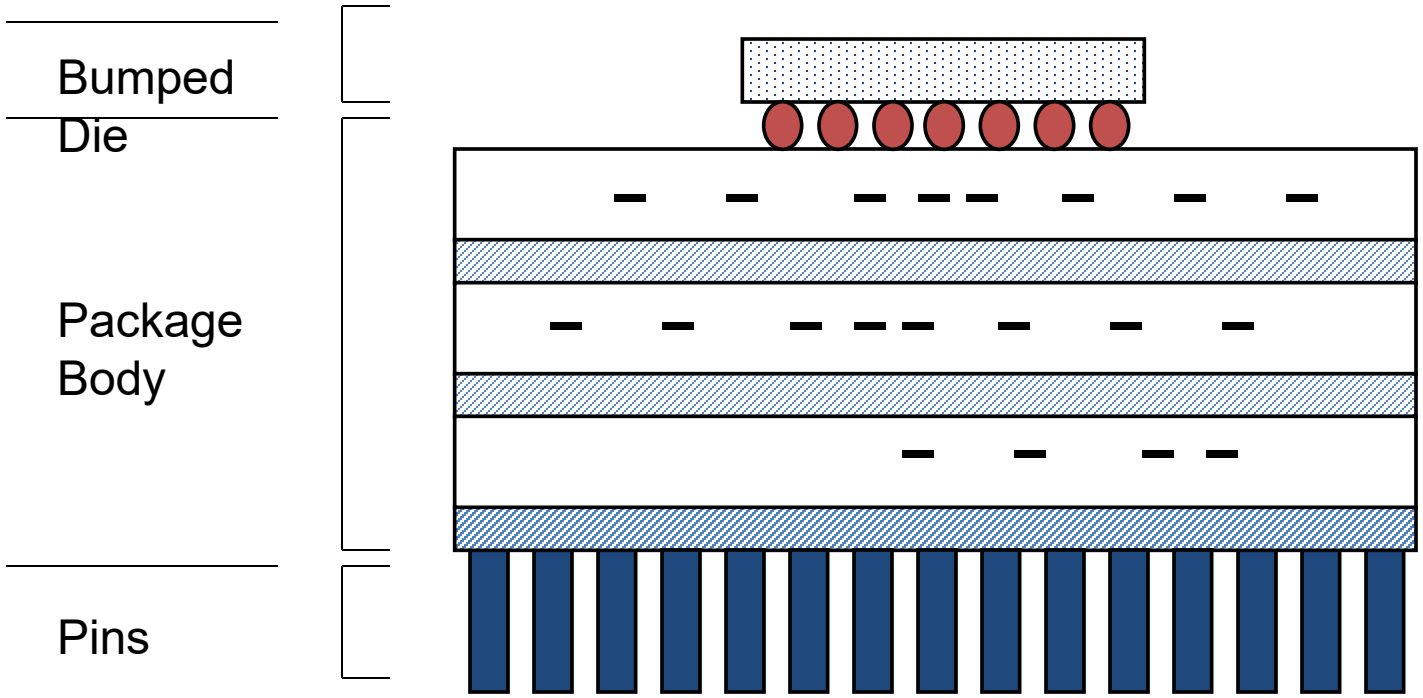
Stacked Wire Bonds



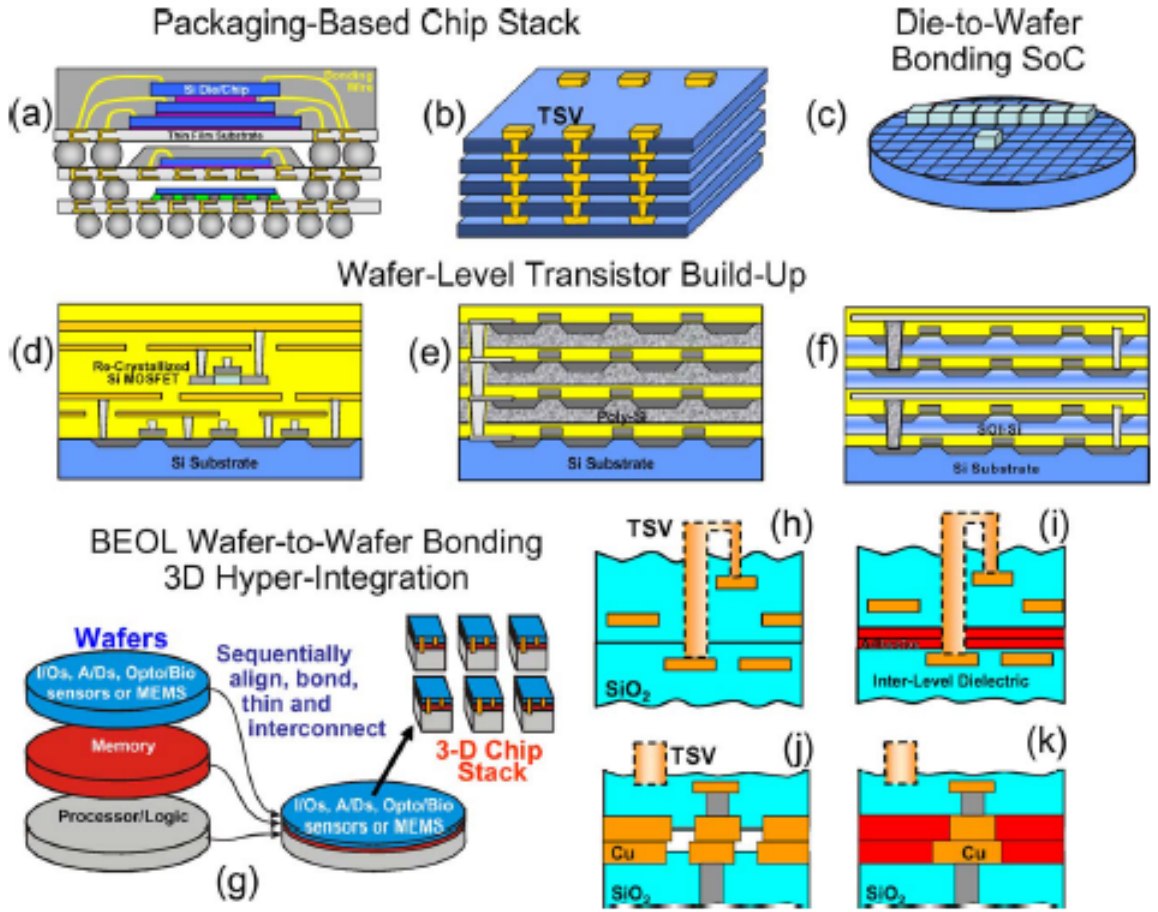
Ball Bonding for Flip Chip



Flip Chip Pin Grid Array (FC-PGA)

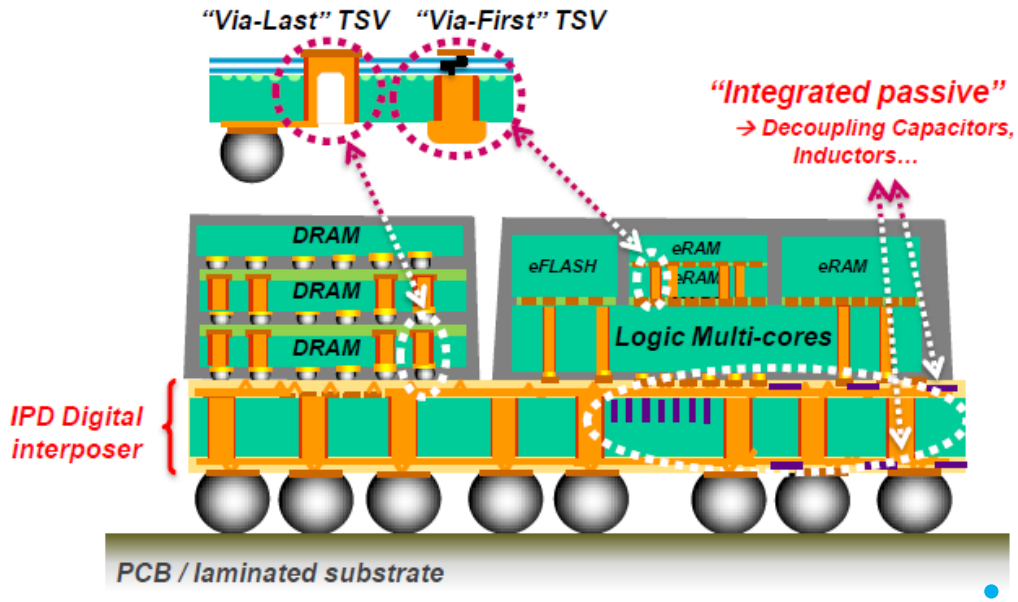


3D Packaging



Source: Jian-Qiang Lu, "3-D Hyperintegration and Packaging Technologies for Micro-Nano Systems", Proceedings of the IEEE, pp 18-30, Vol. 97, No. 1, January 2009.

3D Packaging



Source: Yole Report 2009.

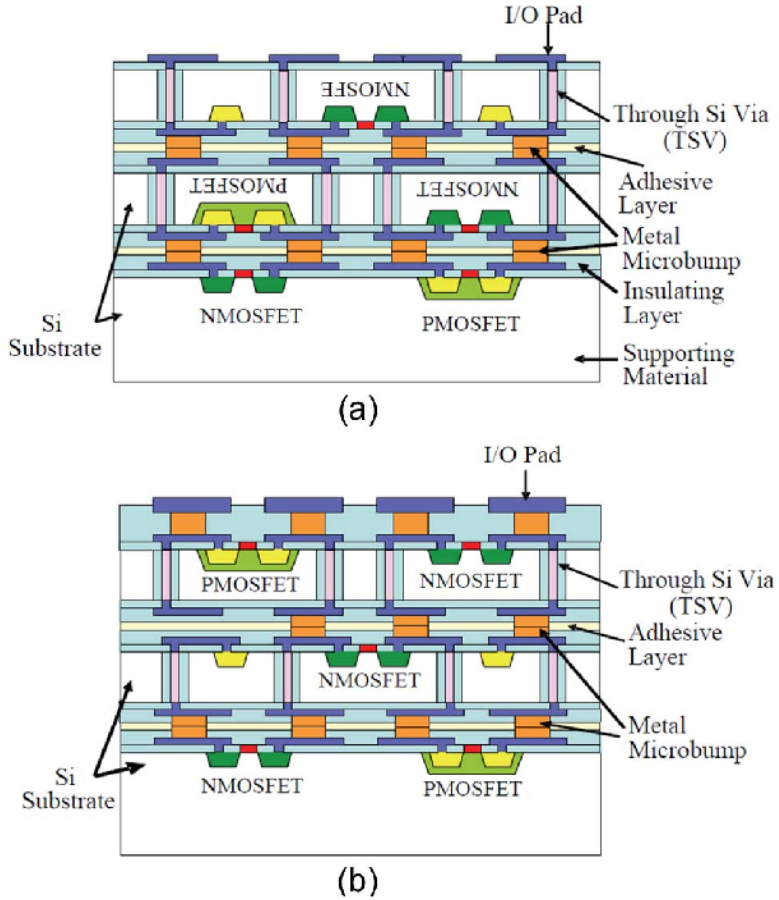
Key concepts

- Wires
  - shorter
  - lots of it
- Heterogeneous integration
  - Analog and digital
  - Technologies (GaAs and Si?)

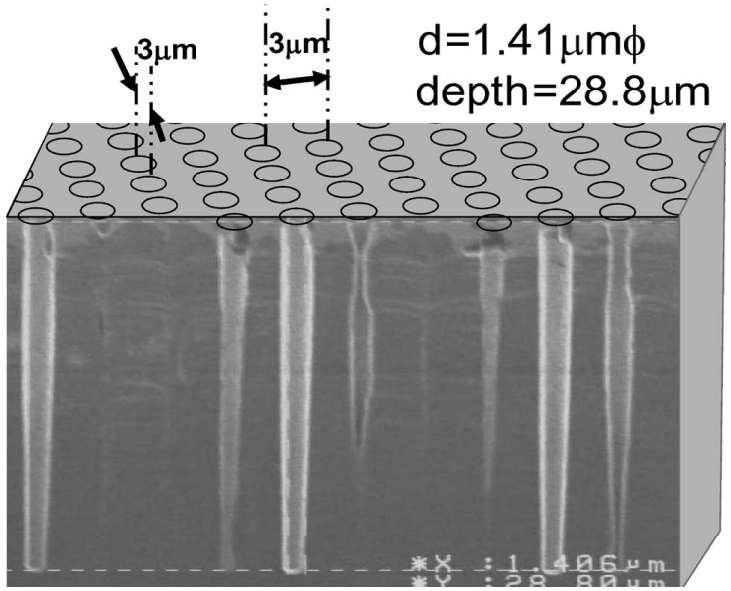
3D Industry

- Samsung
  - 16Gb NAND flash (2Gx8 chips) Wide Bus DRAM
- Micron
  - Wide Bus DRAM
- Intel
  - CPU + Memory
- OKI
  - CMOS Sensor
- Xilinx
  - 4 die 65 nm interposer
- Raytheon/Ziptronix
  - PIN Detector Device
- IBM
  - RF Silicon Circuit Board/ TSV Logic & Analog
- Toshiba
  - 3D NAND

Through-Silicon Vias



Mitsumasa Koyanagi," High-Density Through Silicon Vias for 3-D LSIs"
Proceedings of the IEEE, Vol. 97, No. 1, January 2009



From: M. Motoshi, "Through-Silicon Via, Proc. of IEEE Vol. 97, No. 1, January 2009.

TSV Density: $10/cm^2$ - $10^8/cm^2$

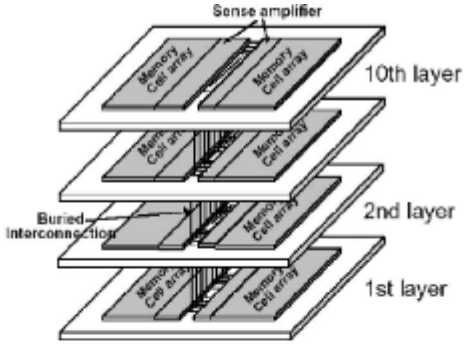
Through-Silicon Vias (TSV)

Advantages

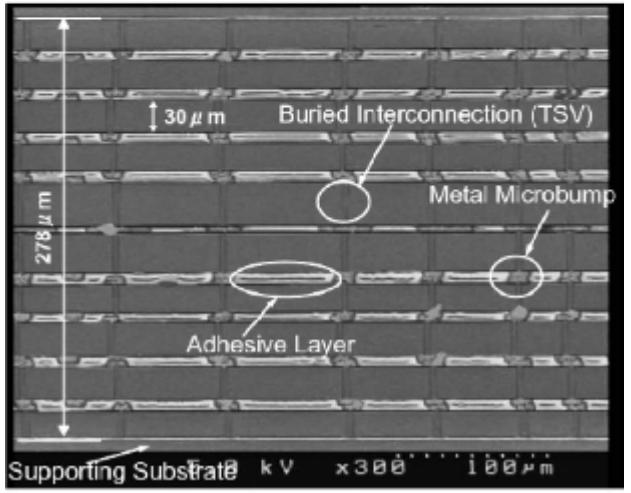
- Make use of third dimension
- several orders of magnitude ($10/cm^2$ to $10^8/cm^2$)
- Minimize interconnection length
- More design flexibility

Issues

- **3D Infrastructure & supply chain**
- **I/O Standardization**
- **EMI**
- **Thermal management and reliability**



(a)



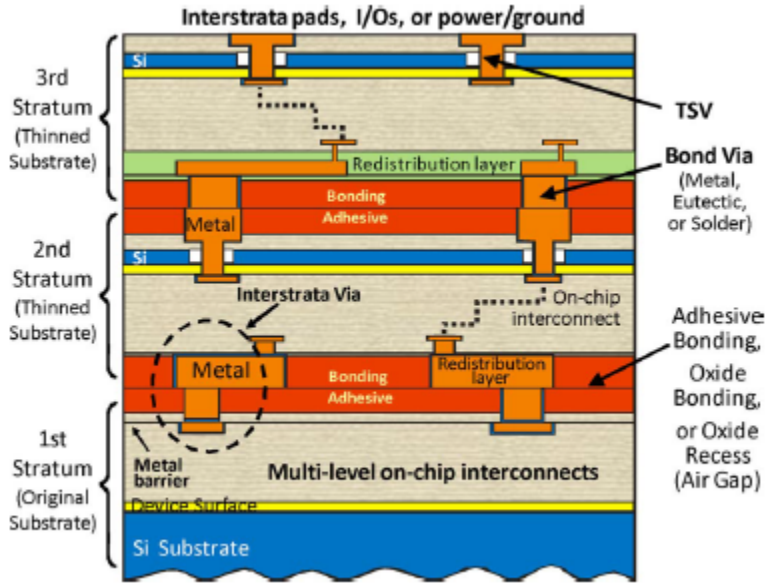
(b)

From Koyanagi et al., IEEE Proceedings, Feb 2009

TSV Pitch

TSV Pitch ≠ Area / Number of TSVs

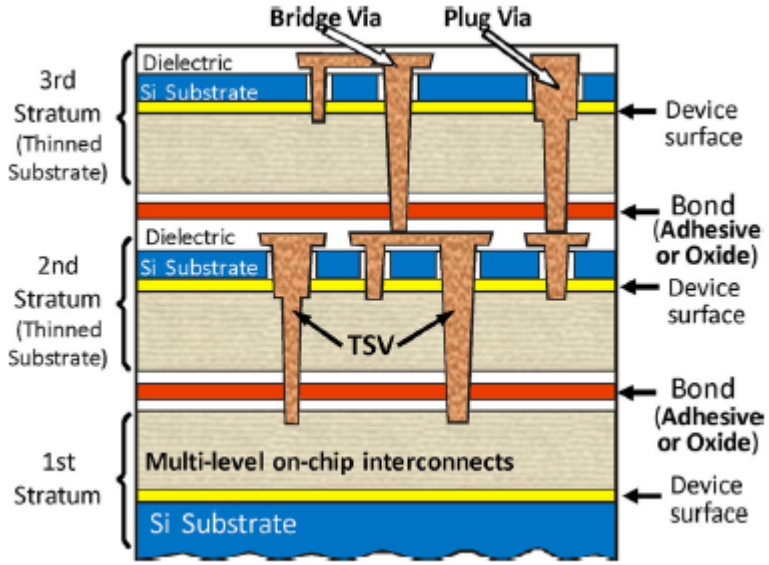
- TSV pitch example
  - 1024 bit busses require a lot of space with larger TSVs
  - They connect to the heart and most dense area of processing elements
  - The 45nm bus pitch is ~100 nm; TSV pitch is > 100x greater



Source: Jian-Qiang Lu, "3-D Hyperintegration and Packaging Technologies for Micro-Nano Systems", Proceedings of the IEEE, pp 18-30, Vol. 97, No. 1, January 2009.

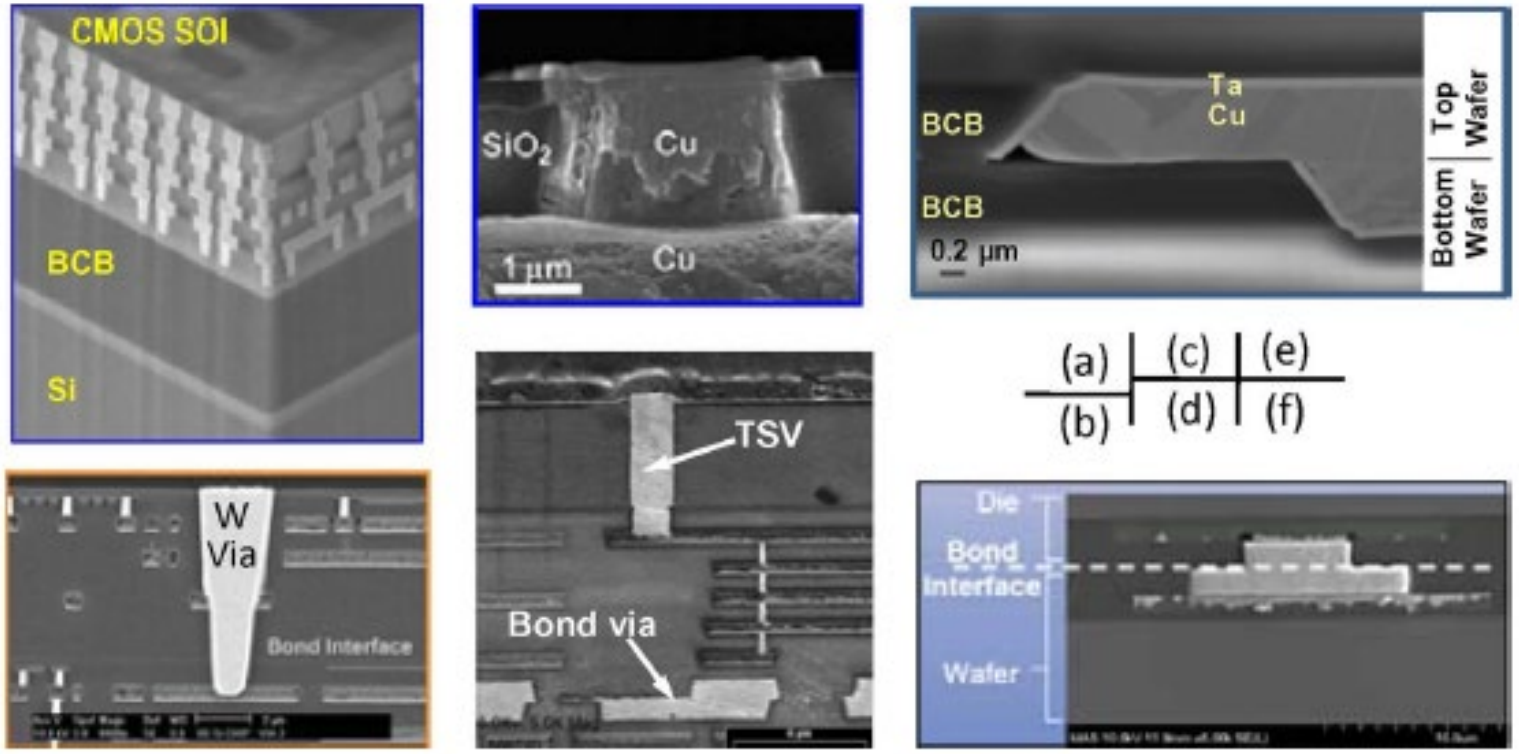
Through-Silicon Vias (TSV)

- Via First
- Via Last
- Via at Front End (FEOL)
- Via at Mid line
- Via at Back end (BEOL)



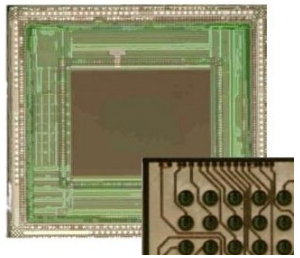
Source: Jian-Qiang Lu, "3-D Hyperintegration and Packaging Technologies for Micro-Nano Systems", Proceedings of the IEEE, pp 18-30, Vol. 97, No. 1, January 2009.

Through-Silicon Vias (TSV)



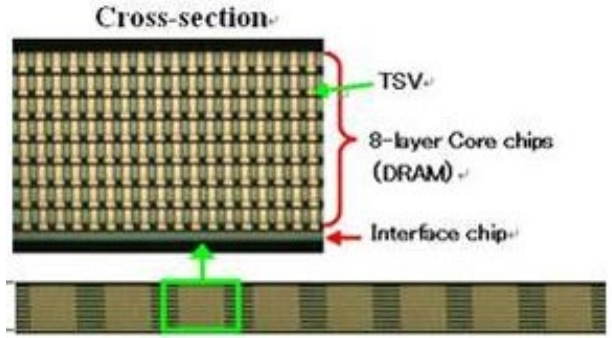
Source: Jian-Qiang Lu, "3-D Hyperintegration and Packaging Technologies for Micro-Nano Systems", Proceedings of the IEEE, pp 18-30, Vol. 97, No. 1, January 2009.

TSV-Based Products



STMicro CMOS image sensor in WLP/TSV package

Sony Video / DSC camera with BSI CMOS image sensors



Elpida's 3D TSV stacked DRAM memory

There are currently about 15 different 3D-IC pilot lines worldwide

3D-IC and TSV

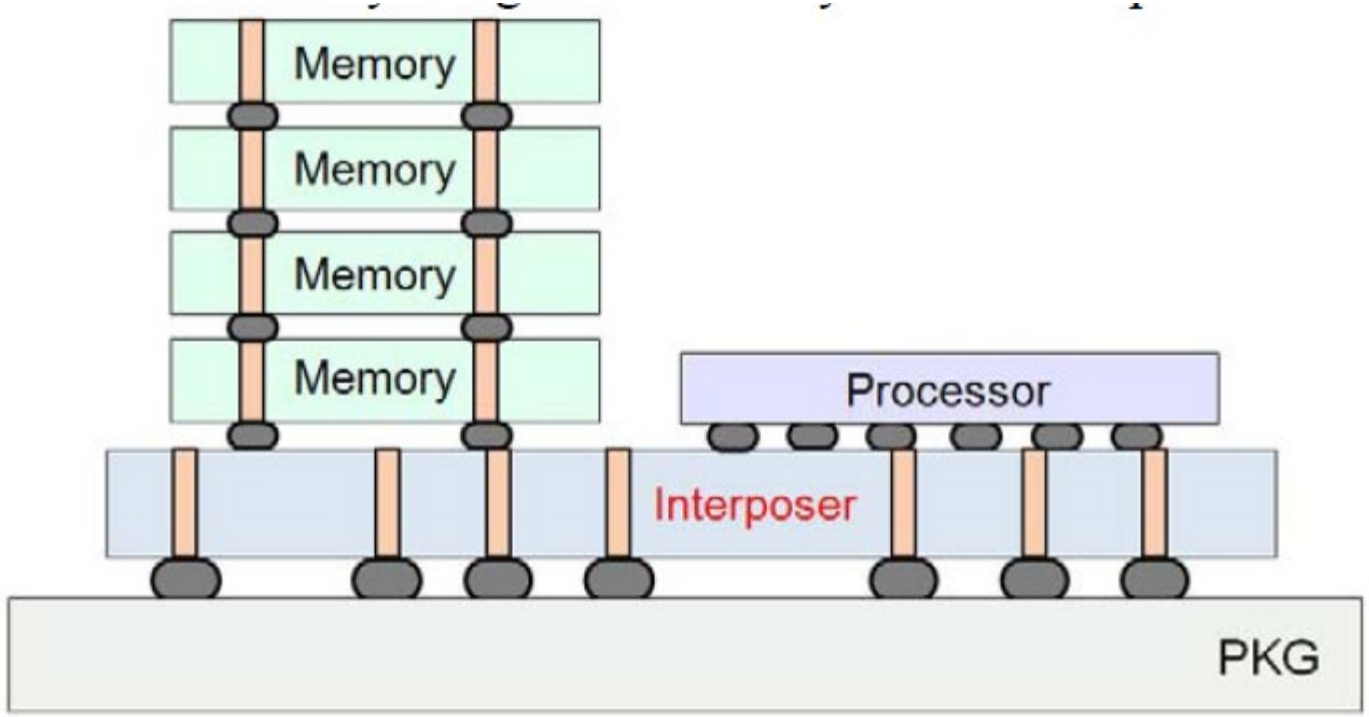
- Stacking of chips makes heat transfer through the z-direction difficult.
- Lossy silicon substrate makes coupling between adjacent TSVs strong.
- TSV noise can be easily coupled to the adjacent TSV through conductive silicon substrate
- 3D IC yields are much lower than 2D-IC
- Difficult to detect TSV and MOS failures

Solution: Use 2.5D integration

2.5D Integration

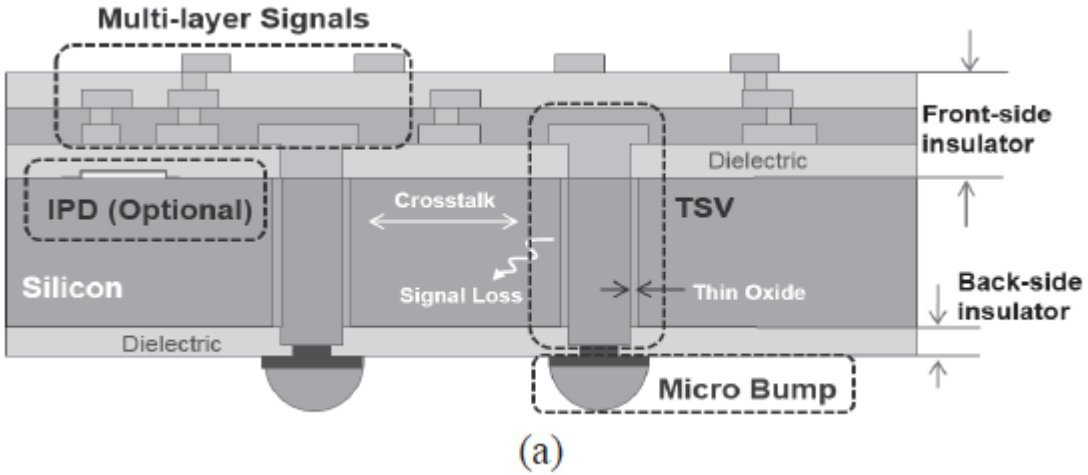
- 2.5D-IC emerges as a temporary solution
- In 2.5D-IC, several chips are stacked on interposer only homogeneous chip stacking is used.
- fine-pitch metal routing is necessary because it increase I/O counts
- For this purpose, an interposer is used where small width and small space metal routing is possible.
- Silicon substrate is usually used for an interposer because on-silicon metallization process is mature and fine-pitch metal routing is possible

Silicon Interposers



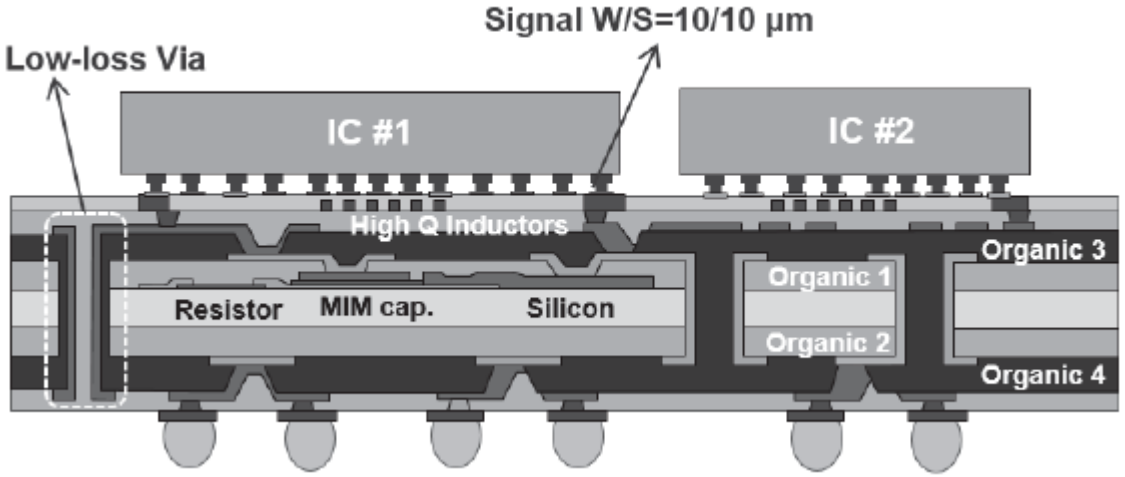
Source: J. Kim et al – DesignCon 2013.

Silicon Interposers



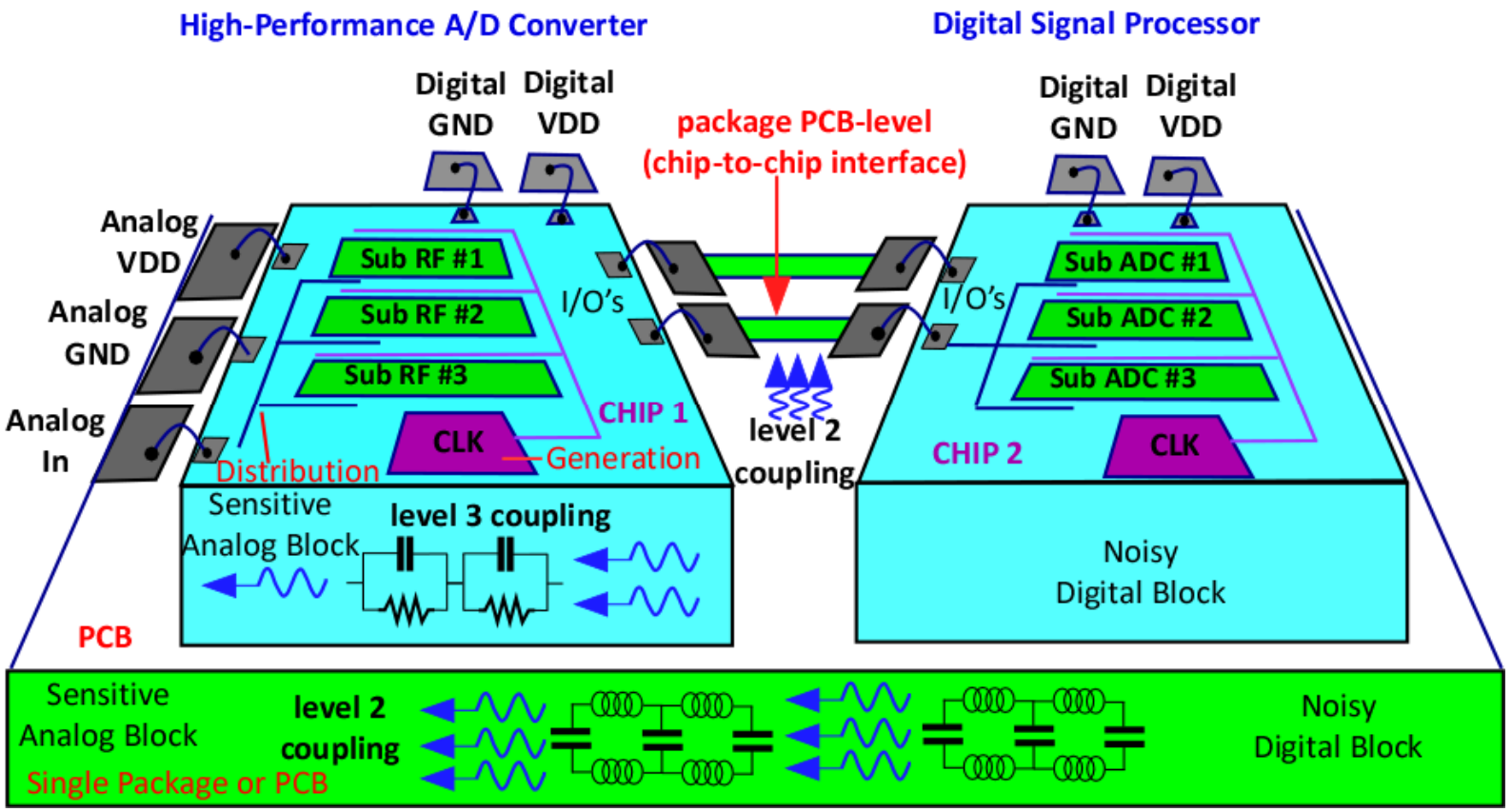
Source: Jong-Min Yook, Dong-Su Kim, and Jun-Chul Kim, "Double-sided Si-Interposer with Embedded Thin Film Devices", 2013 IEEE 15th Electronics Packaging Technology Conference (EPTC 2013), pp 757-760.

Silicon Interposers



Source: Jong-Min Yook, Dong-Su Kim, and Jun-Chul Kim, "Double-sided Si-Interposer with Embedded Thin Film Devices", 2013 IEEE 15th Electronics Packaging Technology Conference (EPTC 2013), pp 757-760.

Coupling Noise in Mixed Signal Systems

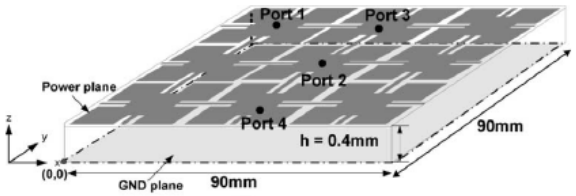


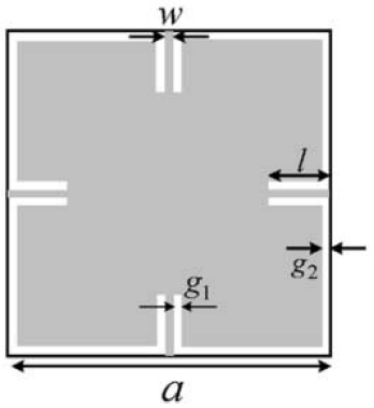
Proposed Solution

- Electromagnetic Bandgap (EBG) Structures
  - Definition: One-, Two- or Three-Dimensional Periodic Metallic/Dielectric System which Exhibits Band Rejection Behavior
  - Bandstop filter characteristics due to shunt capacitances and series inductance
  - Design can be optimized for PDN applications

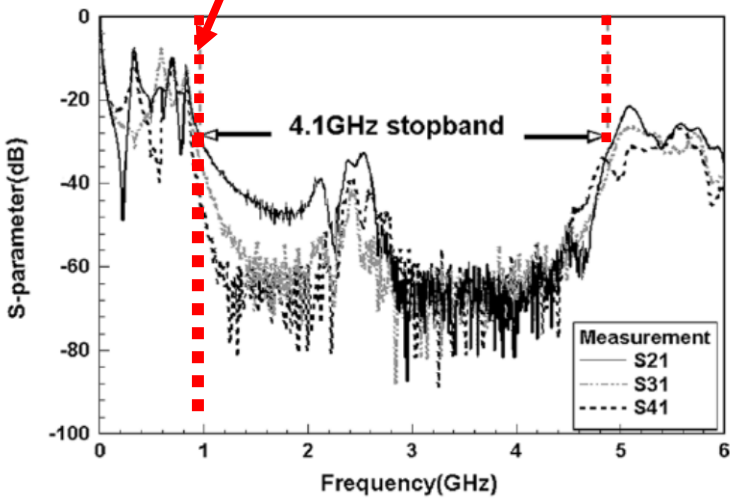
Electromagnetic Bandgap Structures (EBG)

IEEE MWCL July 2004

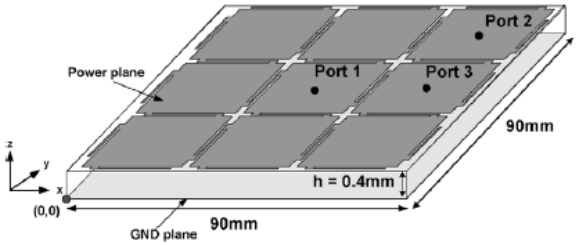


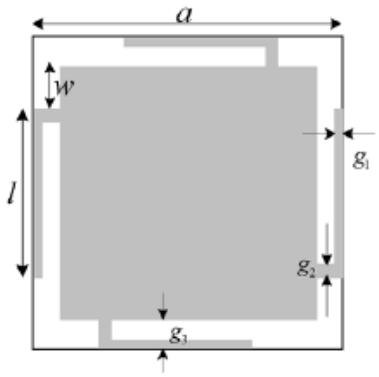


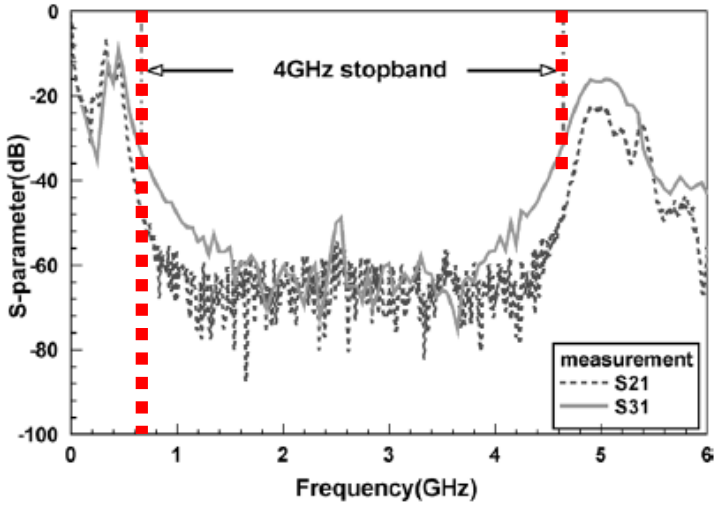
LPF: C_P and L_{bridge}



IEEE MWCL Mar. 2005

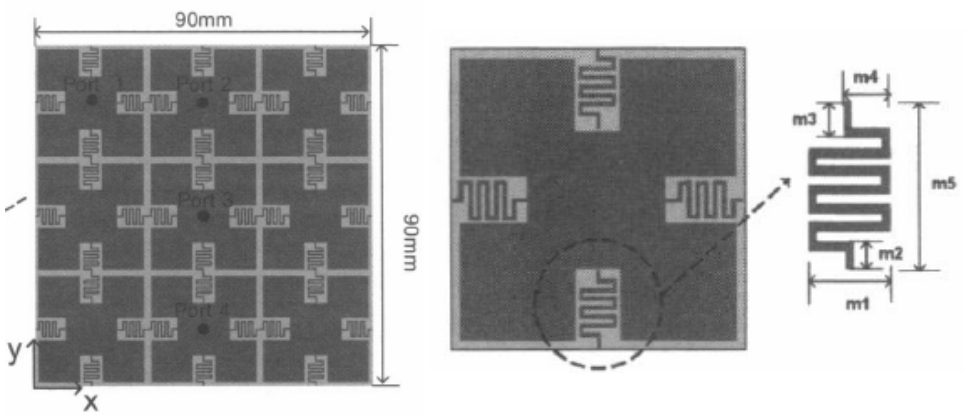


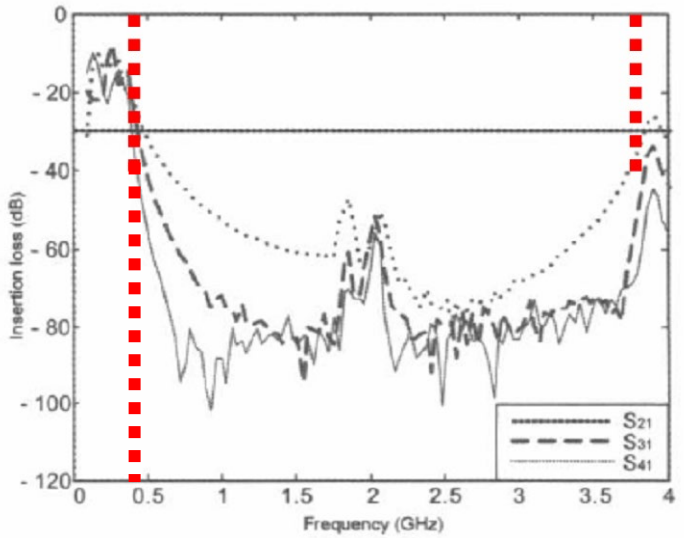




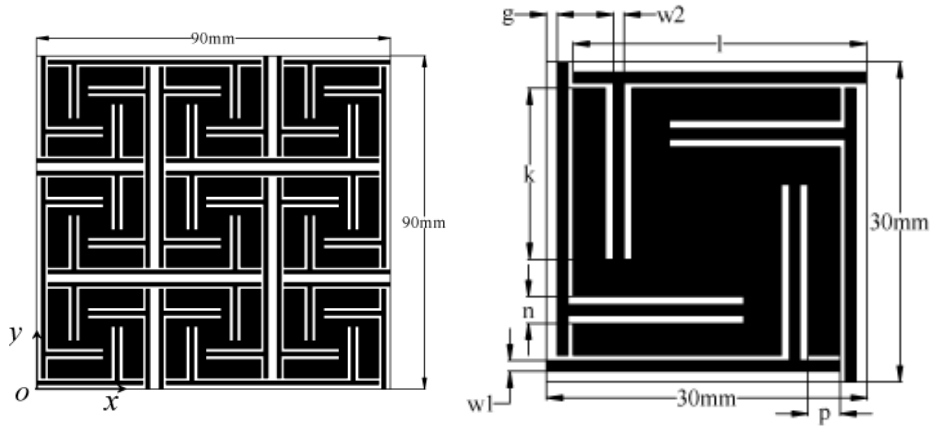
Planar-type EBG Structures

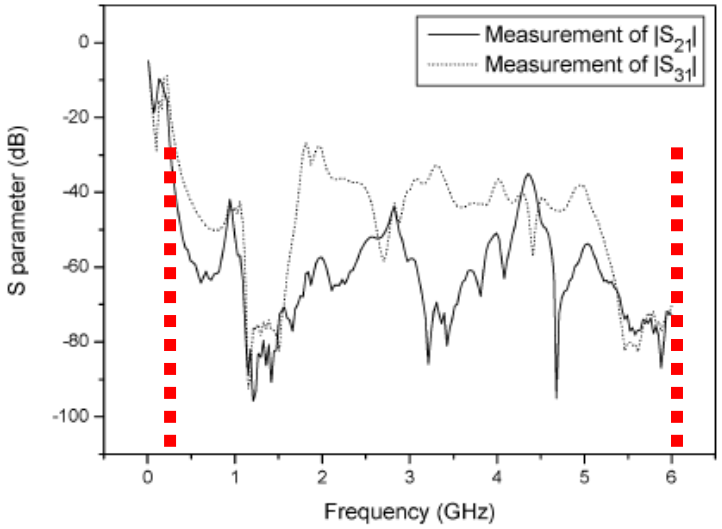
IEEE APS July 2005





IEEE MWCL Mar. 2006

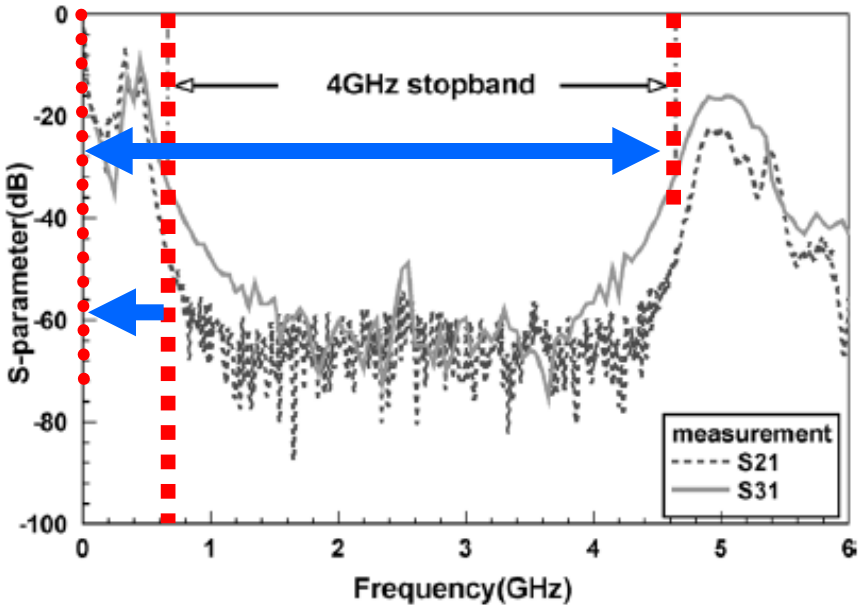


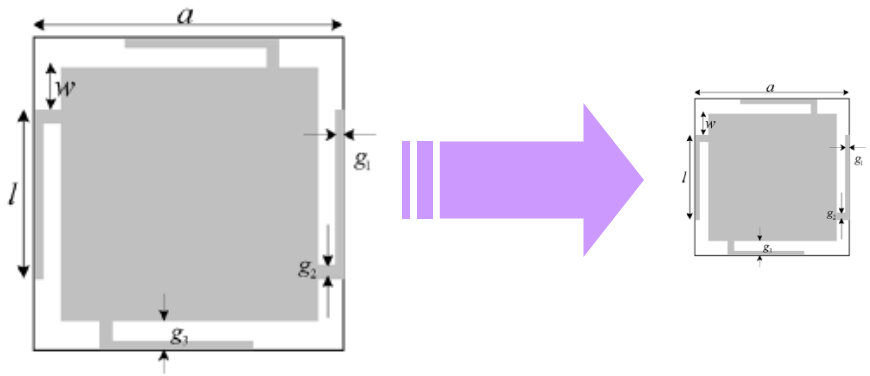


EBG Structure

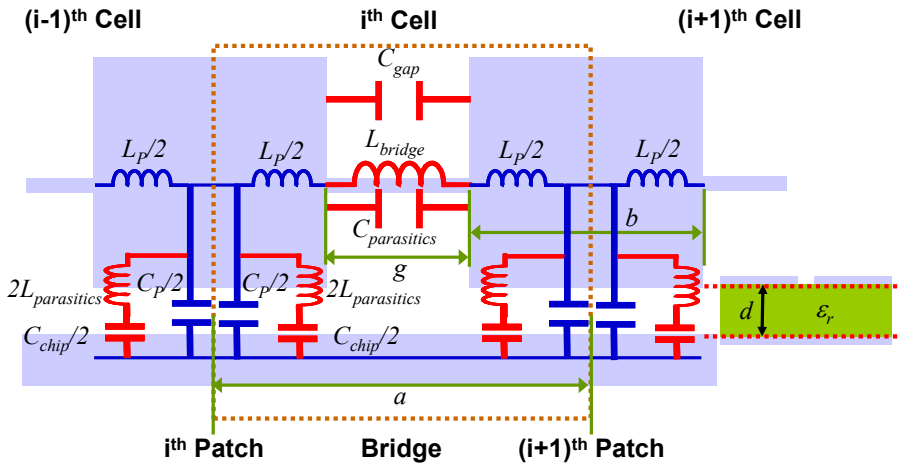
1. Reducing Cut-off Frequency of Planar-type EBG Structure
   - (In Consequence) Enhancing Noise Suppression Bandwidth

2. Miniaturizing Unit Cell of Planar-type EBG Structure
   - Without Degradation in Stopband Bandwidth



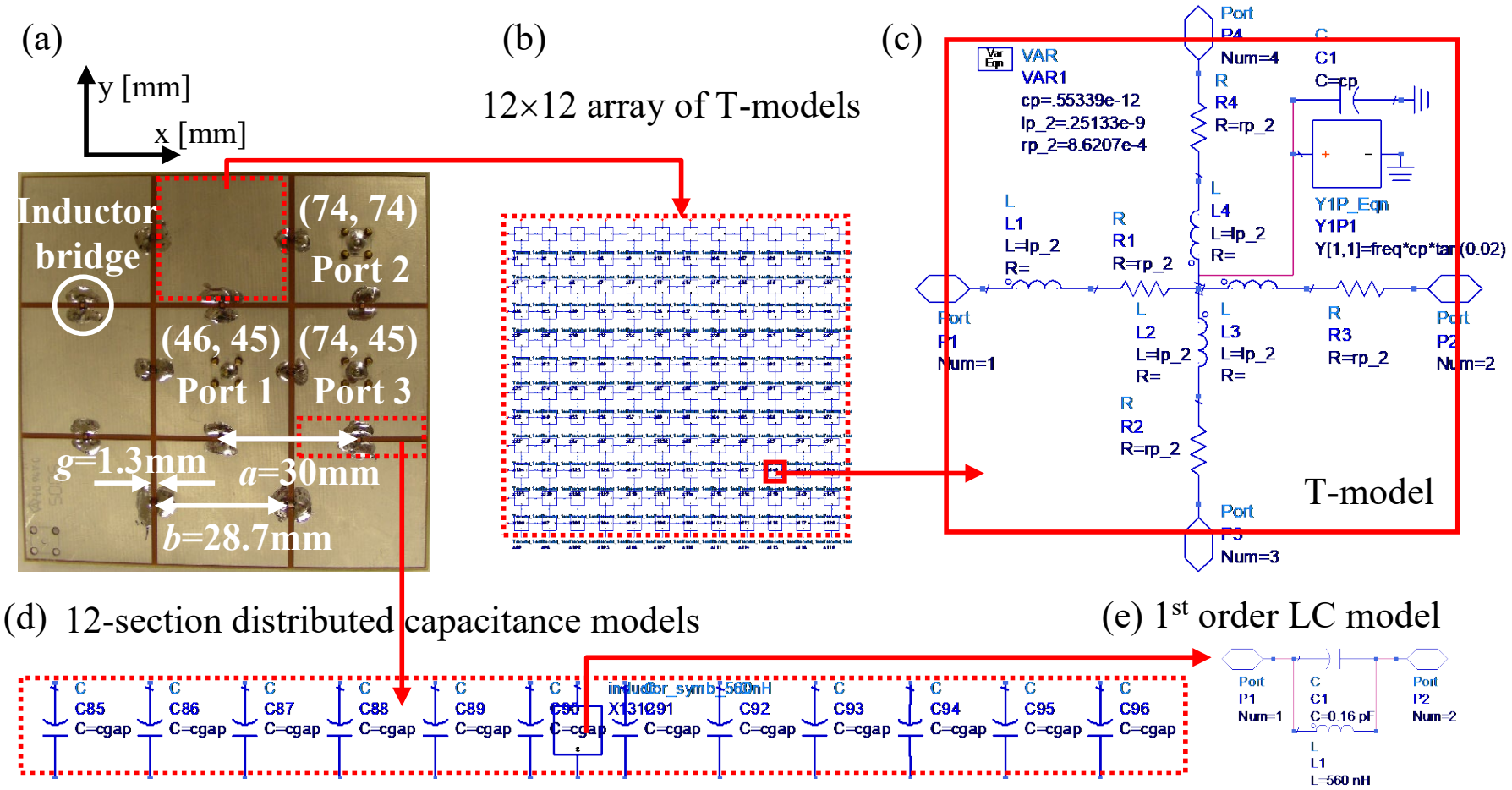


EBG Structure



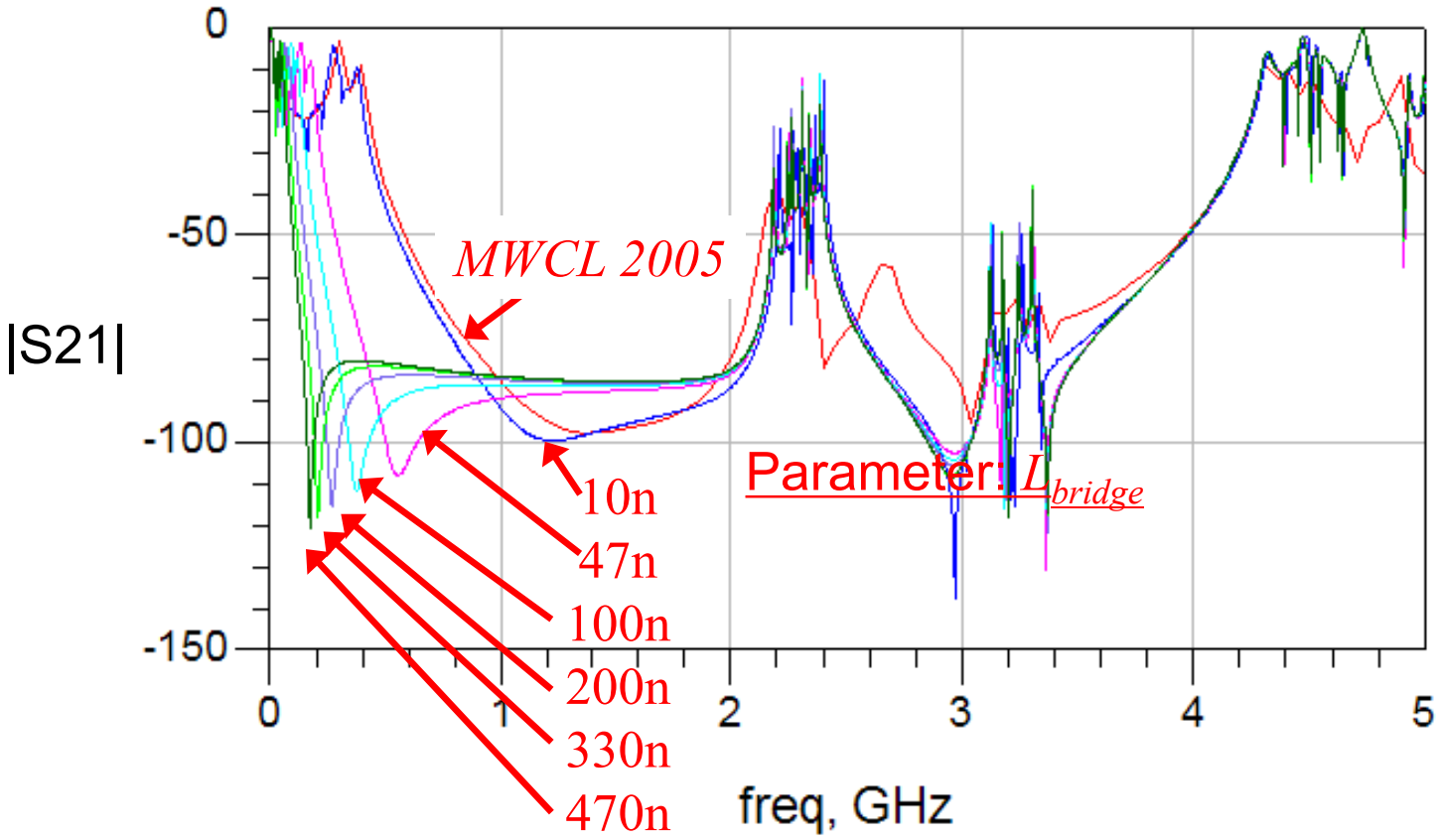
Methods of Cut-off Frequency Enhancements	Dominant Circuit Level Components in Fig.	Cut-off Frequency ($f_{lowpass\_cutoff}$) [MHz]	High Frequency Limitation of EBG Structure
Method 1: Conventional Planar-type EBG Structure	L_P, L_{bridge} (=L_{MSL}), and C_P	$\left[\pi\sqrt{C_P\left(L_P + L_{MSL}\right)}\right]^{-1}$	1st Resonant Frequency of Patch at $c/\left(2b\sqrt{\varepsilon_r}\right)$
Method 2: Increasing Bridge Inductance using Series Lumped Chip Inductors	L_P, L_{bridge} (=L_{chip}), and C_P	$\left[\pi\sqrt{C_P\left(L_P + L_{chip}\right)}\right]^{-1}$	1st Resonant Frequency of Patch at $c/\left(2b\sqrt{\varepsilon_r}\right)$
Method 3: Increasing Patch Capacitance using Shunt Lumped Chip Capacitors	L_P, L_{bridge}, $L_{parasitics}$, C_P, and C_{chip}	$\left[\pi\sqrt{C_P'\left(L_P + L_{bridge}\right)}\right]^{-1}$	Parallel Resonant Frequency of C_P and $L_{parasitics}$ at $1/\left(2\pi\sqrt{C_P L_{parasitics}}\right)$

Verification: ADS Simulation



Simulation Results

– EBG Structure with 90x90 mm^2 Ground Plane Area



Electrical-Thermal AC Analysis

Electrical Analysis:

$$\nabla \times \left(\frac{1}{\mu_r} \nabla \times \mathbf{E} \right) - k_0{}^2 \epsilon_r \mathbf{E} = -j k_0 Z_0 \mathbf{J}$$

- Waveguide port boundary condition
- Absorbing boundary condition

Thermal Analysis:

$$\nabla \cdot k \nabla T = -P$$

$$T = T_c \qquad \text{on } \Gamma_{\text{tc}}$$

$$k \frac{\partial T}{\partial n} = -h(T - T_a) \qquad \text{on } \Gamma_{\text{conv}}$$

Electrical-Thermal DC Analysis

Electrical Analysis:

$$\nabla \cdot \sigma \nabla \phi = 0$$

$$\phi = \phi_c \quad \text{on } \Gamma_{\text{vc}}$$
$$\sigma \frac{\partial \phi}{\partial n} = \frac{\phi}{RS} \quad \text{on } \Gamma_{\text{load}}$$

Thermal Analysis:

$$\nabla \cdot k \nabla T = -P$$

$$T = T_c \quad \text{on } \Gamma_{\text{tc}}$$
$$k \frac{\partial T}{\partial n} = -h(T - T_a) \quad \text{on } \Gamma_{\text{conv}}$$

Electrical Analysis:

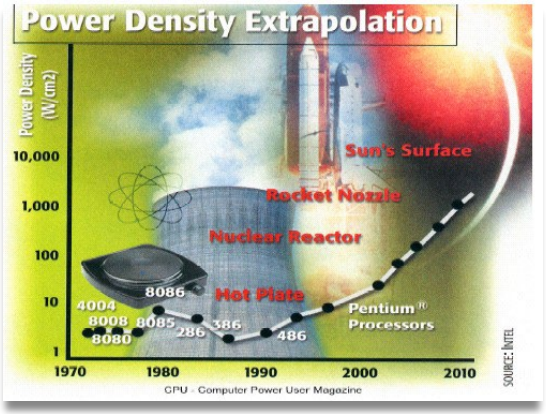
- DC
- AC
- Transient

Thermal Analysis:

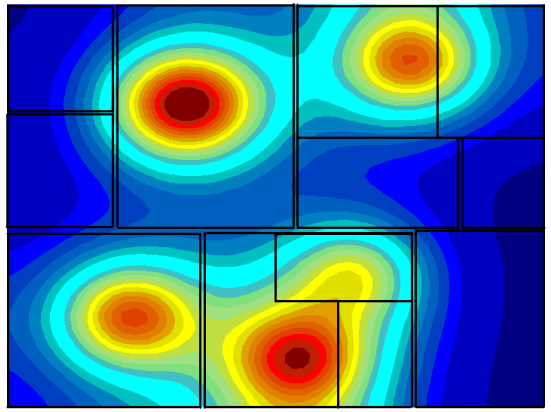
- steady-state
- Transient

Electro-Thermal Analysis. Motivation

- 3D integration technologies
  - 3D stacked IC designs
  - Increased power density
  - Heat removal difficulties



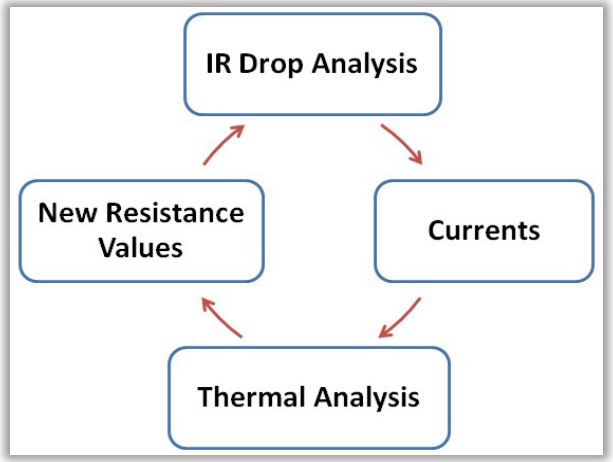
- Design challenges due to thermal issues
  - Electrical reliability (electro-migration)
  - Power delivery (IR drop)
  - Signal propagation (RC delay)
  - Memory retention time (Leakage)



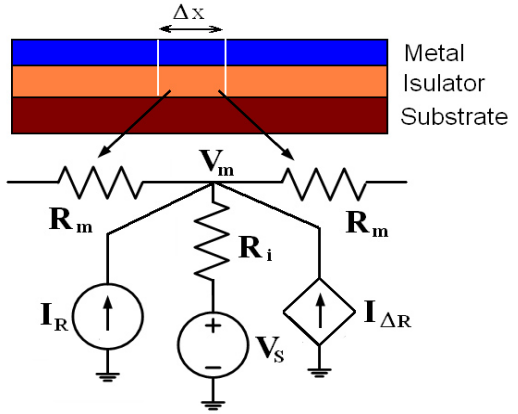
- Lack of suitable CAD tools
  - Thermal-aware design at the earliest stages
  - Using the floor plan and early power distribution analysis (know the current distribution – want to use that information)

Temperature-Dependent IR Drop

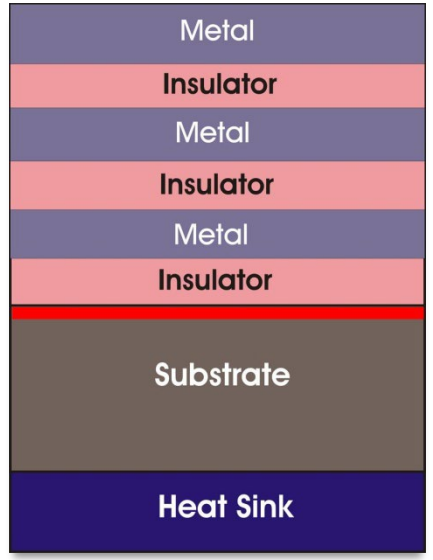
- Interconnect resistance depends on temperature → Current depends on resistance → Temperature depends on current → …
- Temperature-dependent phenomena must be accounted for
  - **Sources of heat**
    - **Self heating (Joule)**
    - **Heating from the substrate (from active devices)**



The flow of thermal-aware [8] IR drop analysis



Lumped model of the interconnect thermal system [9]

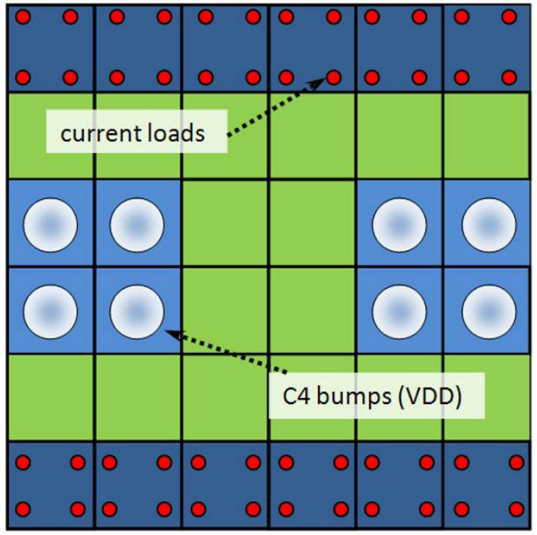


Multilayered structure of an IC

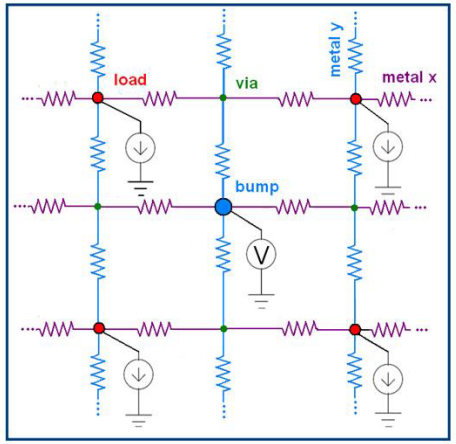
[8] Y. Zhong and M. D. F. Wong, "Thermal-aware IR drop analysis in large power grid," *IEEE ISQED*, 2008, pp. 194-199

[9] C. C. Teng, Y. K. Cheng, E. Rosenbaum, and S. M. Kang, "iTEM: A temperature-dependent electromigration reliability diagnosis tool," *IEEE Trans. Computer-Aided Design*, vol. 16, pp. 882–893, Aug. 1997.

Pre-Layout IR Drop Analysis



Floor plan and pad out

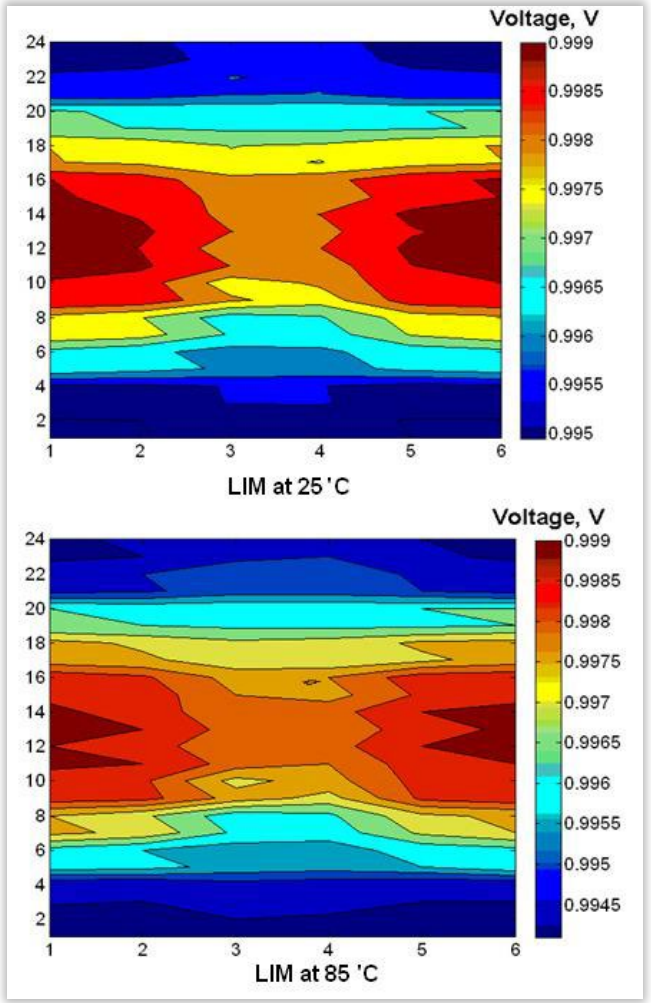


Equivalent circuit for a single segment

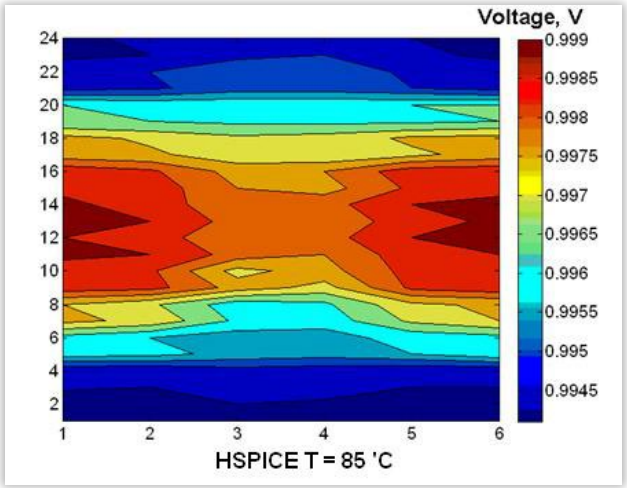
Example based on a segment of the actual Rambus test-chip

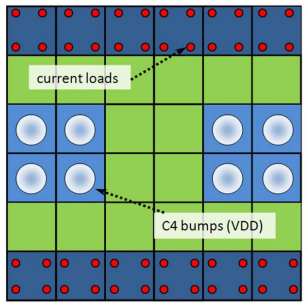
- PDN parameters are in the form of
  - process parameters
  - floor plan
  - pad out
  - current load
  - voltage budget
- Floor plan is divided into segments
- Each segment is modeled with a resistive grid
- A script is used to generate the equivalent circuit model
- Once the circuit representation is available a circuit solver can be used to perform static IR drop analysis

Pre-Layout IR Drop Analysis

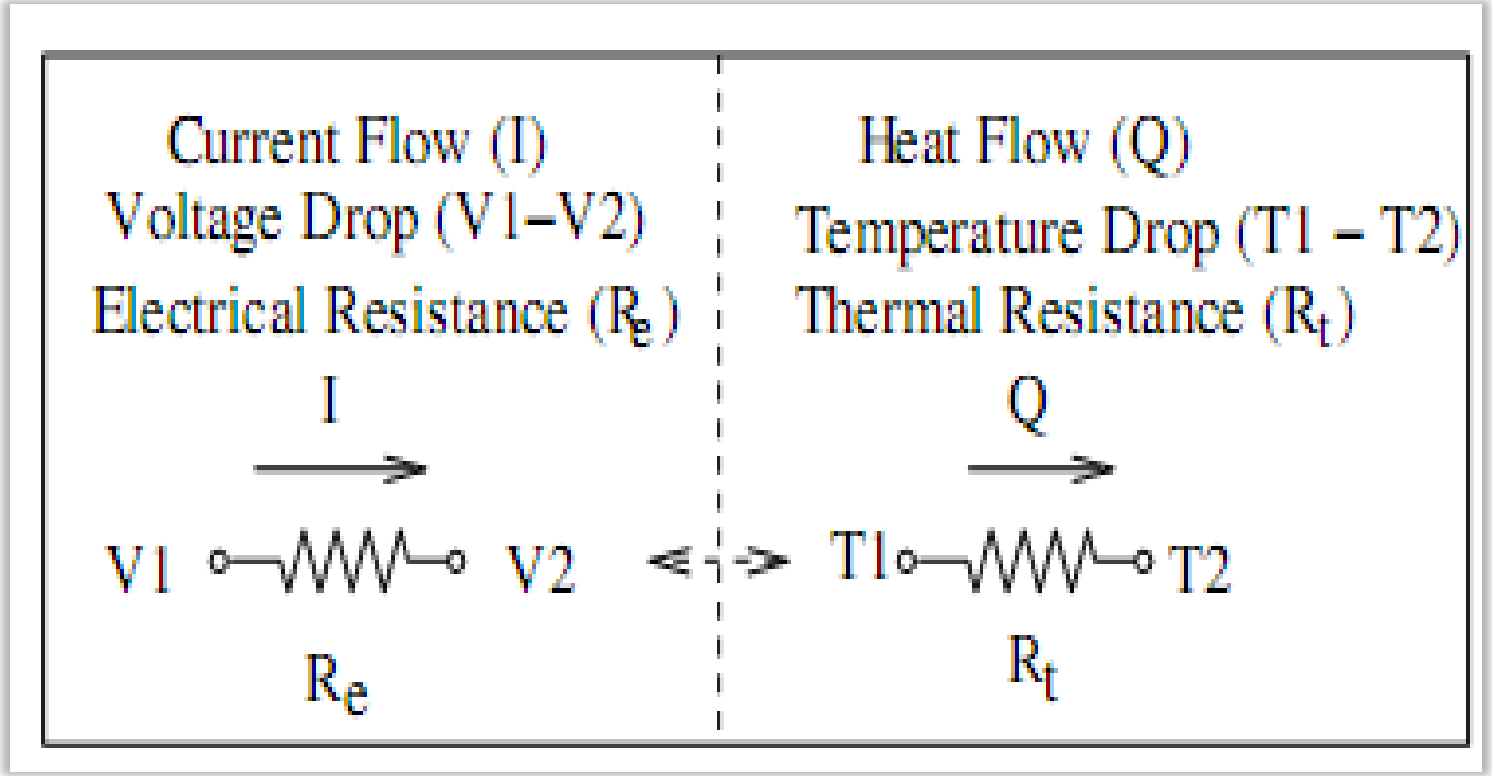


- Nominal VDD voltage of 1V is used
- Voltage drop is smaller at the locations of the bumps and becomes higher as the distance from the VDD pads increases
- Two simulations are performed at 25 °C and 85 °C
- LIM simulation at 85 °C is verified with HSPICE
- For this particular example temperature has noticeable but minor effect on the IR drop
- No thermal-electrical iterations are performed



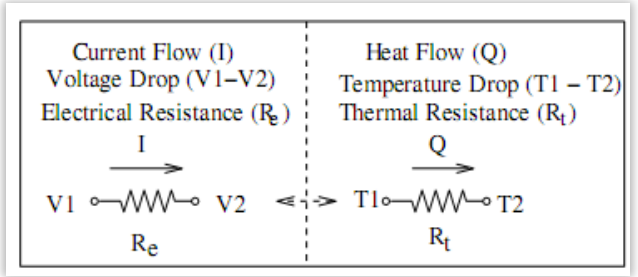


Electrical-Thermal Isomorphism



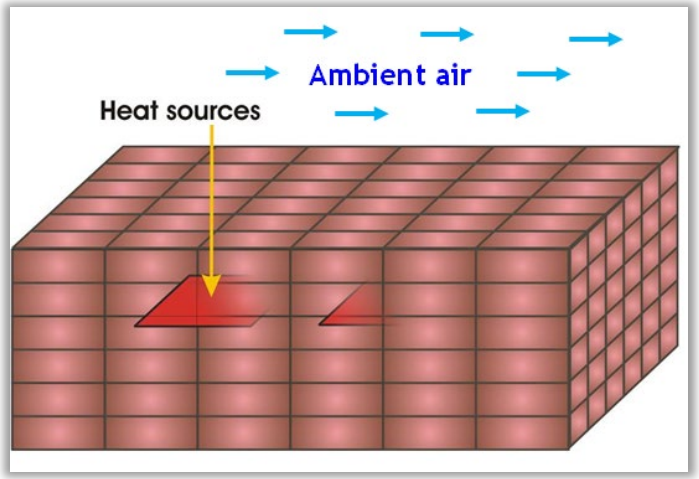
Temperature distribution in IC structure

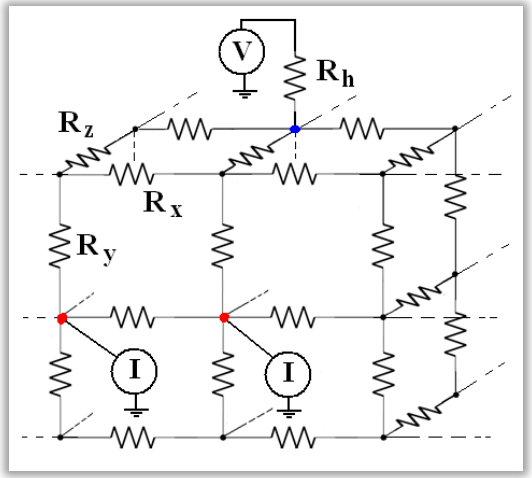
- **Modeling methodology**
  - Use thermal – electrical analogy
  - Thermal problem → electrical circuit
  - Bulk of the material → 3D Resistive network
  - Heat sources → Constant current sources
  - Convective boundaries → Effective resistances
  - Ambient temperature → Constant voltage sources



$$R_x = \frac{\Delta x}{kA}\left[\frac{^oC}{W}\right]$$

$$R_h = \frac{1}{h_e A}\left[\frac{^oC}{W}\right]$$



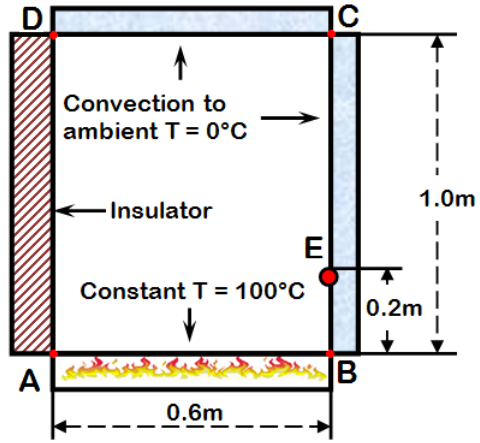


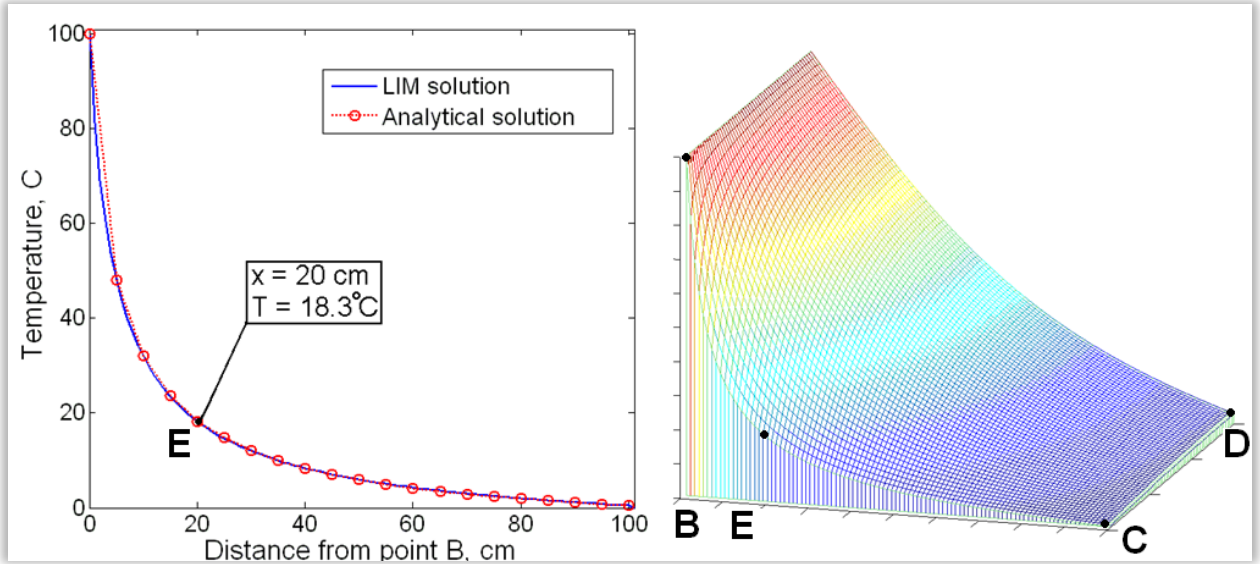
Apply circuit solver

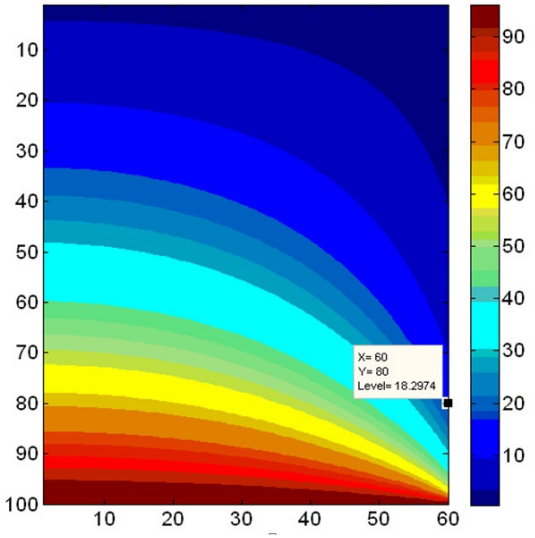
- **Solve the resulting network for node voltages**
  - A major issue – the SIZE of the model

Benchmark Thermal Problem

- 2D benchmark problem (NAFEMS)
  - Simple geometry
  - Has all typical components
  - There is analytical solution
  - Target temperature at E is 18.3 °C

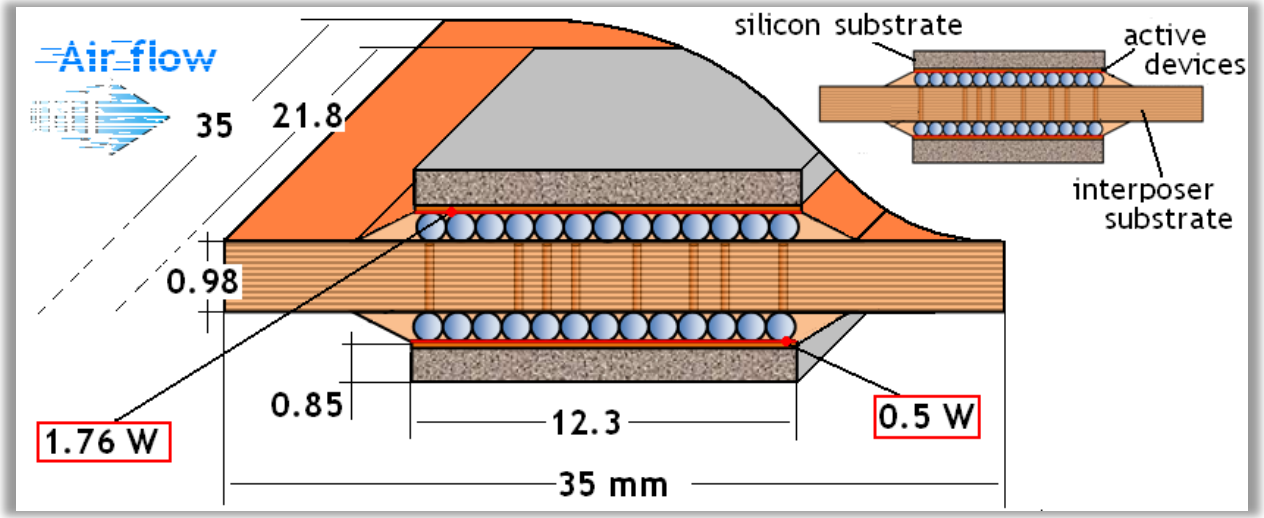


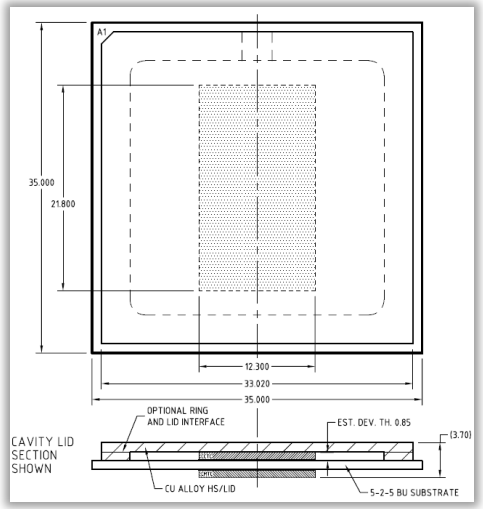




[10] Davies, G.A.O. and Fenner, R. T. and Lewis, R. W., *Background to benchmarks*, NAFEMS, 1993

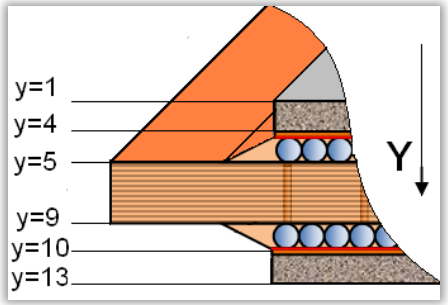
3D Structure. Chip-Interposer-Chip





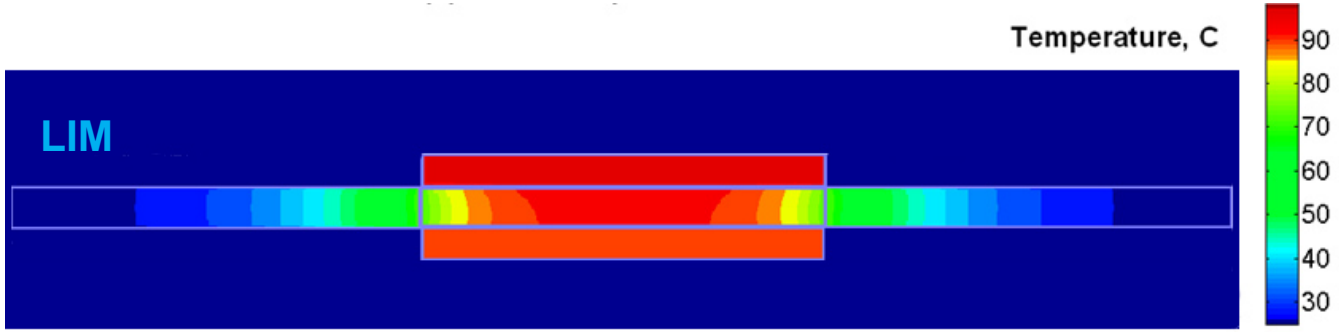
- How hot does the system get ?
- How much heat is transferred from the top chip (controller) to the bottom (memory) ?
- If the via density or distribution is changed, how does that affect the temperature distribution ?
- How much heat can be dissipated through the interposer substrate ?

Model parameters	
Unit cell size (cube)	Δx = 0.2833 mm
Number of nodes	135,089
Number of branches	326,168
Total number of elements	**461,257**

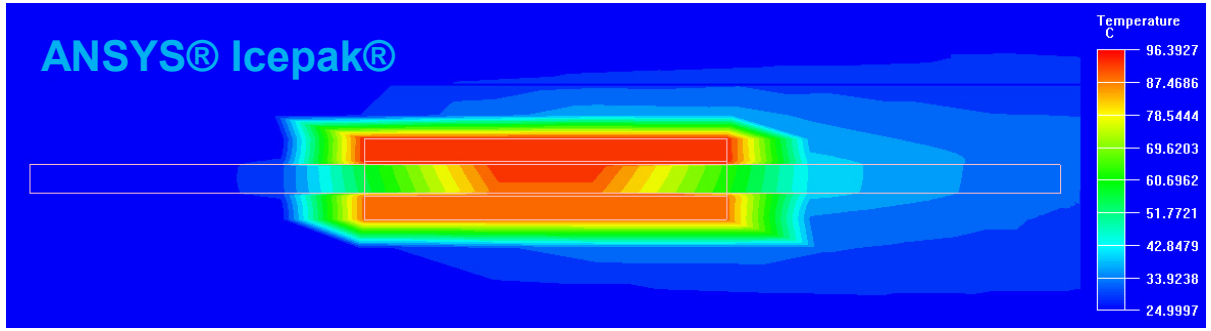


Results of the simulation

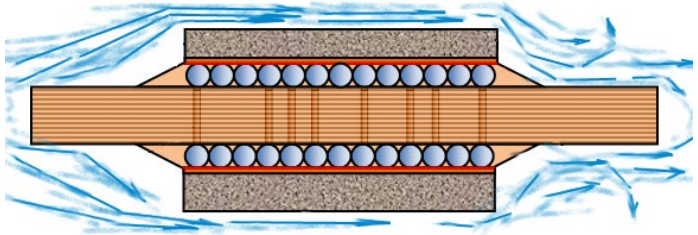
- LIM simulation results in a symmetric temperature profile



- Icepak result is not symmetric

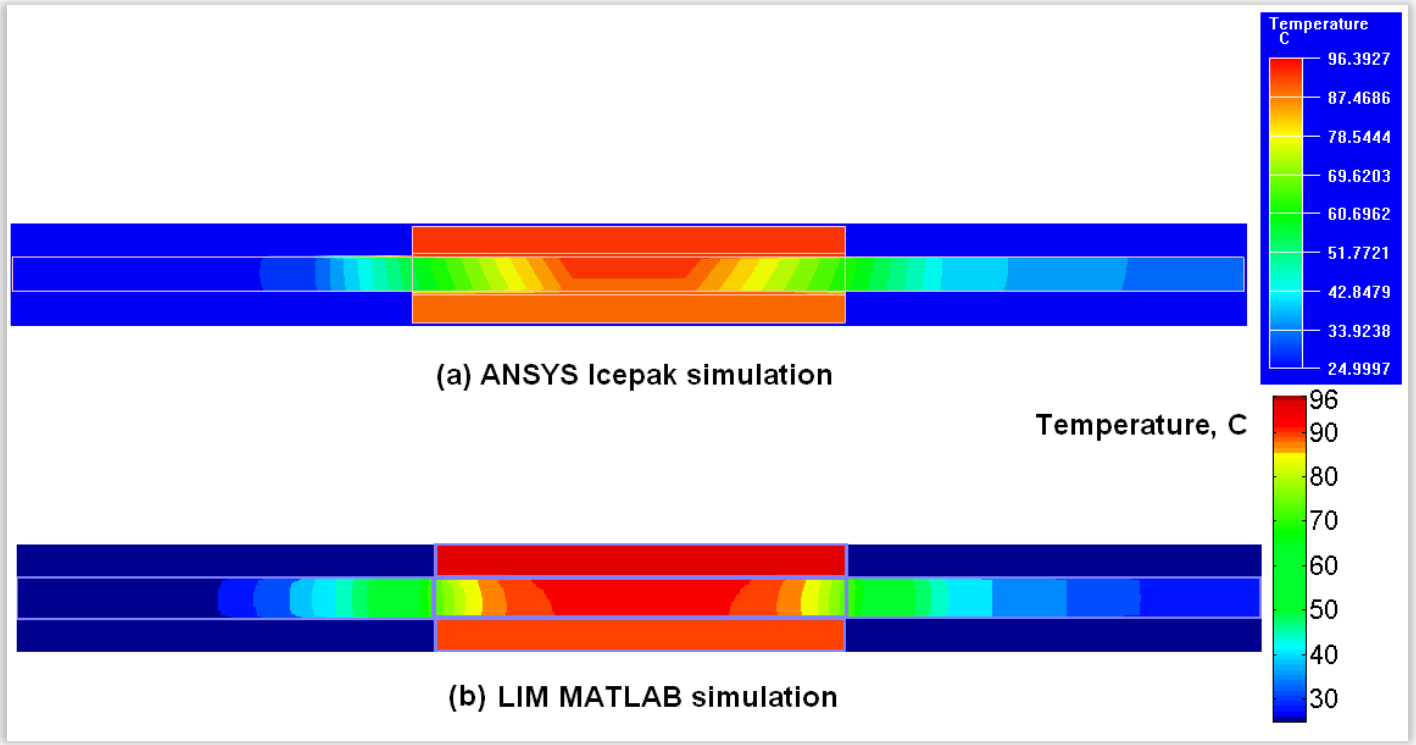


- We need to account for non-uniform cooling



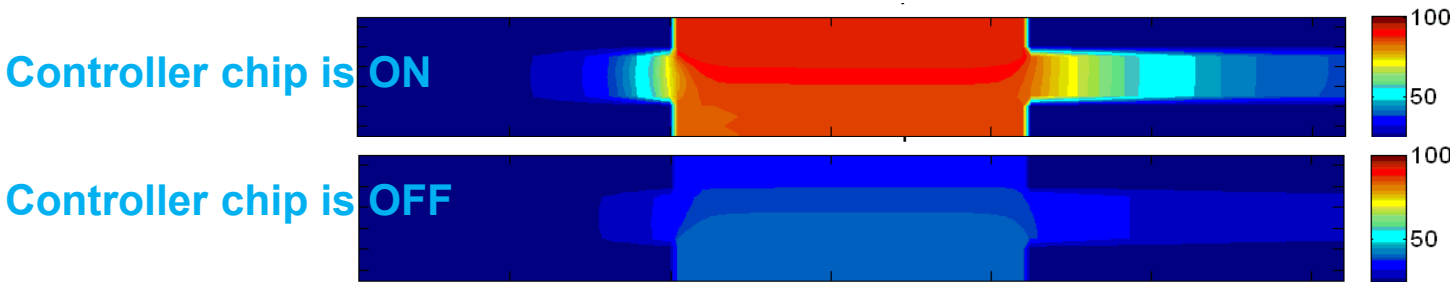
Results with non-uniform cooling

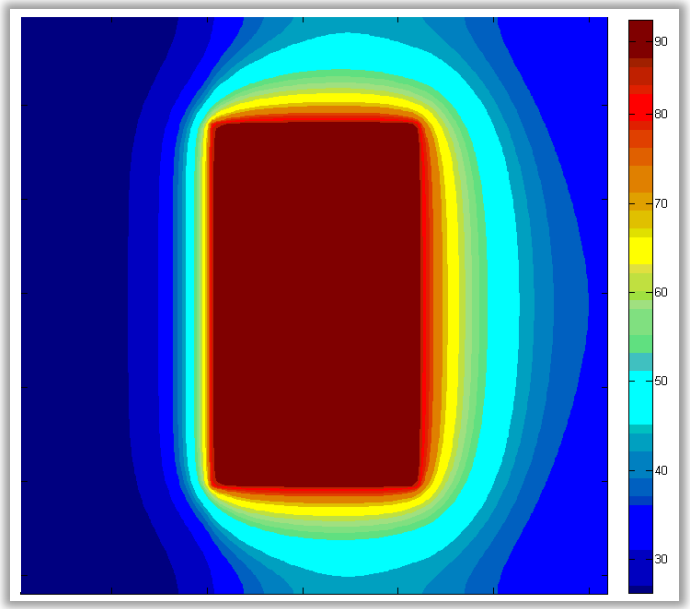
- Cross section of the 3D structure (center cut)
  - Comparison between two pictures from different tools
  - Looking for correct temperature range and general distribution (color maps used by the tools are not exactly the same)
  - In general, very good correlation is observed



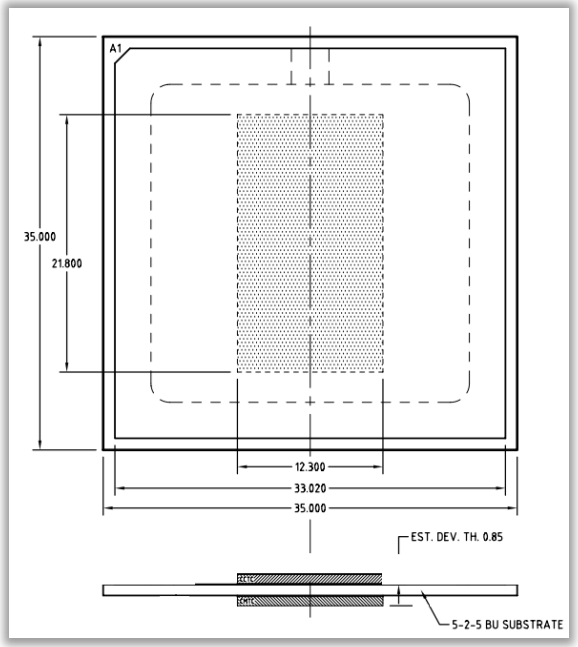
More Results

- Can look at two scenarios





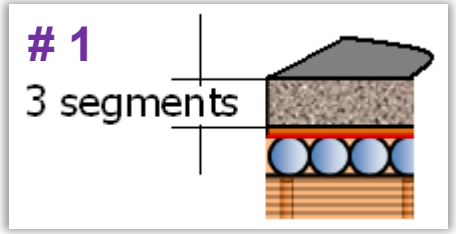
Steady-state temperature profile.
Top view of the structure

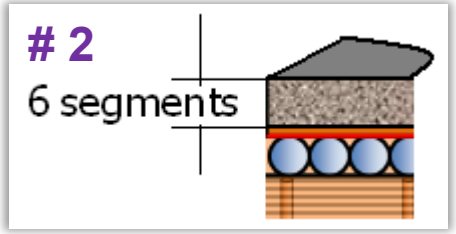


- In our 3D model we can observe any cross-section of the structure

Model Size Issues

- Mesh density considerations
  - coarse mesh results in errors in heat flux calculation
  - geometry of the structure
  - structure of the underlying PDN
  - sizes of elements of interest (TSVs, solder balls, etc.)





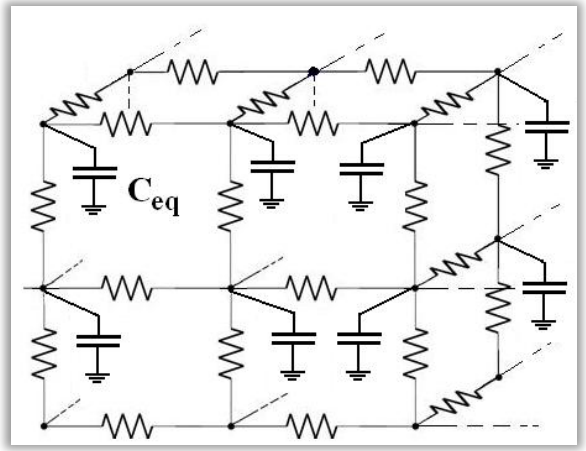
#	Elements	Run time LIM (C++)	Run time LIM (MATLAB)	Run time HSPICE 2010
1	461,257	7	16	27 (total)
2	3,514,400	65	154	949 (total)

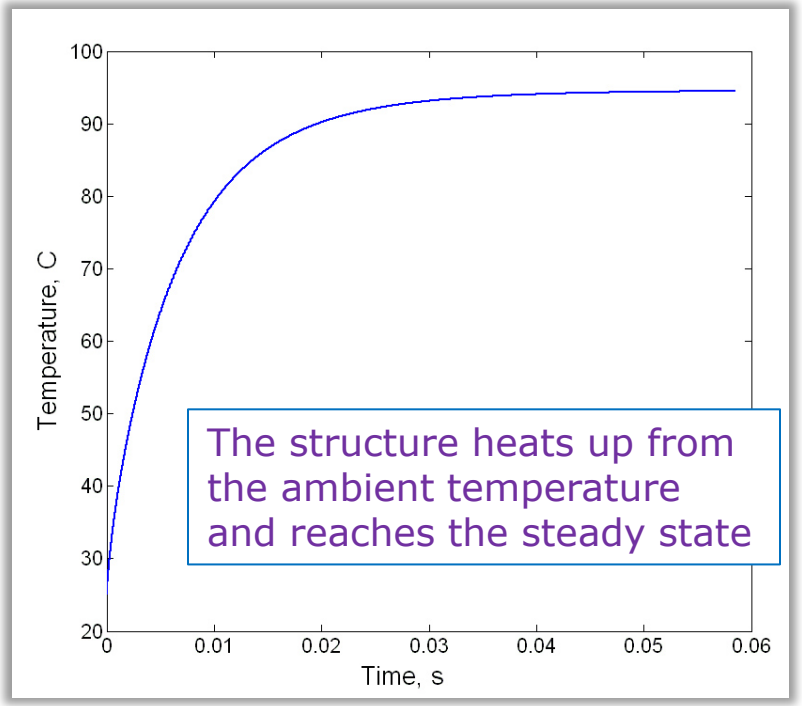
- Typically the size of the equivalent circuit is very large
- Traditional solvers (SPICE) do not scale well with the size of the model

Transient results

- Transient analysis is naturally performed by the LIM
- Insert the actual capacitance instead of fictitious

$$C_{eq} = c_p \cdot \rho \Delta x^3 \left[\frac{W \cdot s}{^oC} \right]$$





- Dynamic heat management through workload distribution
- Cooling management

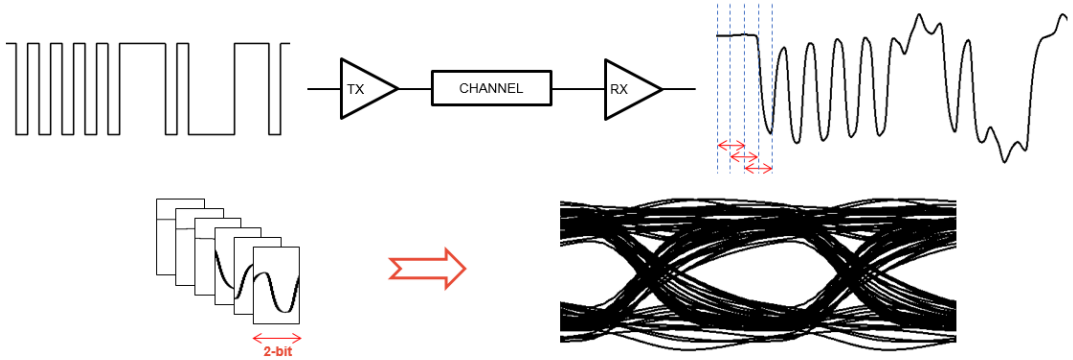
Novel CAD Technologies

- High-Speed Link Simulation
  - Surrogate modeling with polynomial chaos
  - Statistical eye diagram
- Machine Learning Techniques
  - Electromagnetic Modeling
  - Macromodeling
  - Equalization and

Eye Diagram Simulation Method

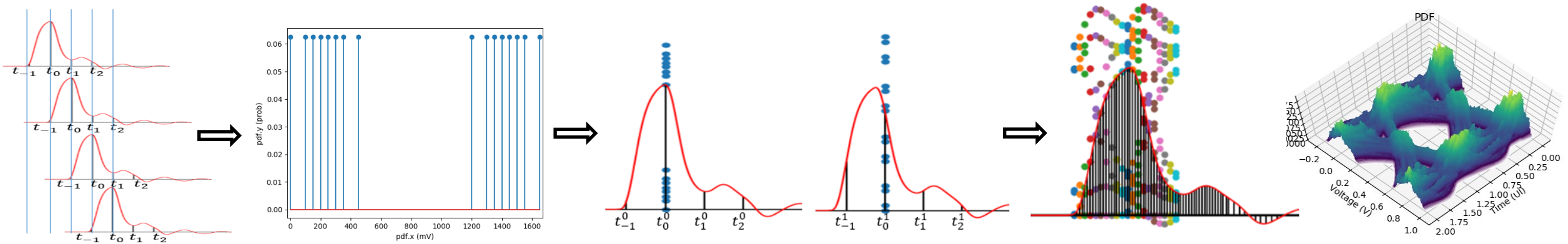
Transient Simulation

- The conventional Simulation Program with Integrated Circuit Emphasis (SPICE) transient simulation can handle nonlinear features.
- The time required for simulation increases in proportion to the length of the input sequence. For BER=10^{-12}, 10^{12} bits are expected to run. The runtime for 10^4 bits is 1 sec, so 10^{12} bits will take about 3 years to complete.



Statistical Simulation

- The statistical eye diagram is directly estimated from single bit response and its statistical information without running the full waveform transient simulation in LTI system.
- $$ISI_{k+1}(v,t) = \frac{\delta(v - V_0(t)) + \delta(v - V_1(t))}{2} * ISI_k(v,t)$$



Superposition of cursors | Probability of a given ISI | ISI at t^0 | ISI at t^1 | Multiple ISI at $t \in [0, 2T]$ | 3D view of eye diagram

Eye Diagram Simulation Method

Limits of Current Methods:

- **Transient** simulation: to completely capture nonlinearity, the full bit sequence is fed into the HSL and the full waveform is obtained, then deriving to eye diagram, which is computationally time-demanding.
- **Statistical** simulation: the fast and efficient simulation sacrifices the nonlinearity property and leads to a less accuracy as the statistical analysis bases on LTI and superposition.

Proposed Solutions:

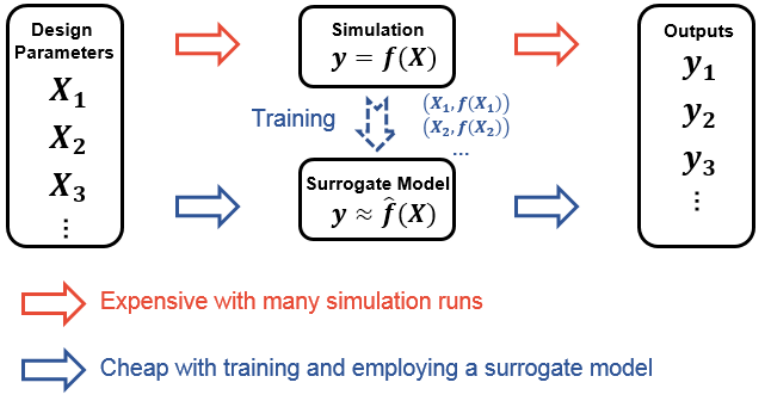
1. **Surrogate model through Polynomial Chaos Expansion (PCE) technique.**
   a) HSL Waveform and Eye Diagram Estimation via PCE
   b) **HSL Eye Width and Eye Height Estimation**
   c) **Decision-Feedback Equalizer (DFE) Taps Estimation with Surrogate Modeling Methods**
2. **Statistical analysis of the eye diagram in nonlinear systems via Wiener model.**
   a) Theoretical Wiener System
   b) **Differential FinFET Buffer**
   c) **High-speed Link System with CTLE and DFE**

Bobi Shi,Yi Zhou,Haofeng Sun,Thong Nguyen,Jose. Schutt-Aine, "Statistical Method for Eye Diagram Simulation in High-Speed Link Nonlinear System Applications", IEEE Transactions on Components, Packaging and Manufacturing Technology, 2024.

Surrogate Model – Polynomial Chaos Expansion

What is Surrogate Model?

- A surrogate model is a simplified model that is used to approximate the behavior of a complex or computationally expensive model that is difficult to work directly.



What is Polynomial Chaos Expansion?

- A Polynomial Chaos Expansion (PCE) is a way of representing an arbitrary random variable of interest as a function of another random variable with a given distribution, and of representing that function as a polynomial expansion.

$$y(\lambda) = \sum_{i=0}^{P} c_i \phi_i(\lambda)$$

$$P + 1 = \frac{(m+n)!}{m!n!}$$

$$\lambda = [\lambda_1, \lambda_2, .., \lambda_n]^T$$

$m =$ polynomial order

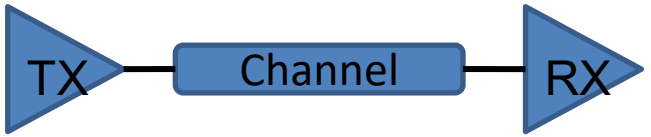
	Orthonormal Hermite Polynomial
$\phi_0(\lambda)$	1
$\phi_1(\lambda)$	λ
$\phi_2(\lambda)$	$(\lambda^2 - 1)/\sqrt{2}$
$\phi_3(\lambda)$	$(\lambda^3 - 3\lambda)/\sqrt{6}$
$\phi_4(\lambda)$	$(\lambda^4 - 6\lambda^2 + 3)/(2\sqrt{6})$

$$\mathrm{A} = \begin{bmatrix} \Phi_0(\lambda^1) & \dots & \Phi_P(\lambda^1) \\ \vdots & \ddots & \vdots \\ \Phi_0(\lambda^N) & \dots & \Phi_P(\lambda^N) \end{bmatrix}, \Gamma = \begin{bmatrix} c_0 \\ \vdots \\ c_P \end{bmatrix}, \mathrm{Y} = \begin{bmatrix} y(\lambda^1) \\ \vdots \\ y(\lambda^N) \end{bmatrix}$$

$$\Gamma = (A^T A)^{-1} A^T Y$$

Surrogate Model

Waveform and Eye Diagram Estimation



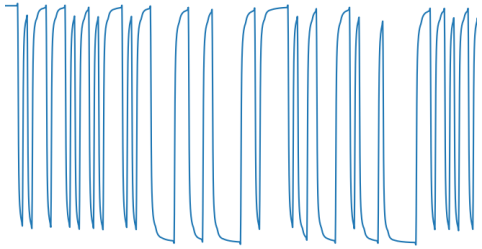
- Original input-output data set: bit sequence – output waveform
- Issue: a small bit sequence can result in a large matrix to be solved in PCE. 2000 bits leads to matrix column size in the order of billion.

$$P+1=\frac{(m+n)!}{m!\,n!}=\frac{2003!}{3!\,2000!}\approx 1.3E9$$

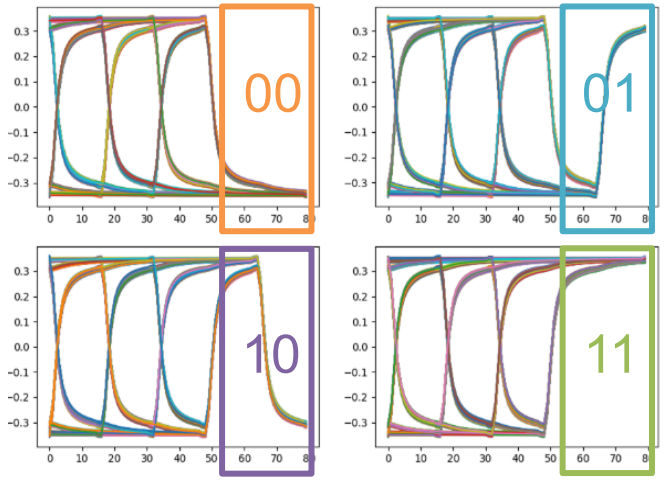
$$\mathrm{A}=\begin{bmatrix}\Phi_0(\lambda^1) & \dots & \Phi_P(\lambda^1)\\ \vdots & \ddots & \vdots\\ \Phi_0(\lambda^N) & \dots & \Phi_P(\lambda^N)\end{bmatrix} \qquad \Gamma=(A^TA)^{-1}A^TY$$

- Updated input-output data sets:
  - 4 groups of data set based on the last 2 bit transition states
  - Inputs: N-bit length sequence ending with 00/01/10/11
  - Outputs: Corresponding last 2-bit output waveform
  - Testing bits: 20,000 unseen bits

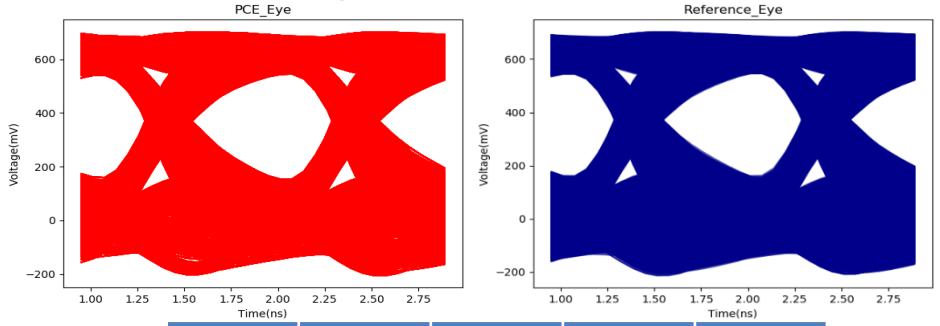
101110110000… $\hat{f}(X)$

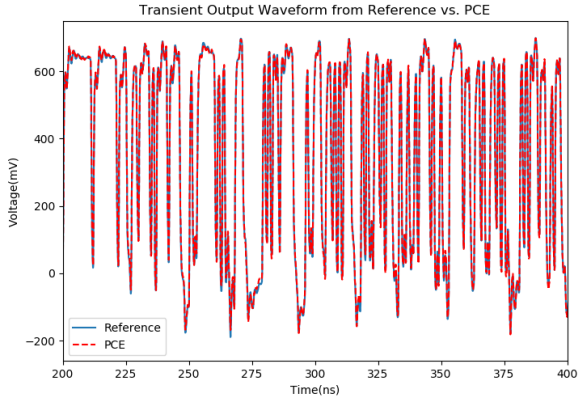


Decompose into 4 groups



Group00: 10100, 01000, 01100…
Group01: 01101, 10101, 10001…
Group10: 00110, 11010, 01010…
Group11: 10011, 00011, 01011…





	EH (mV)	EW (ns)	One Level	Zero Level
PCE	295	755	623	17
REF	281	740	622	18
%Error	4.98	2.02	0.16	5.56

Nonlinearity Modeling

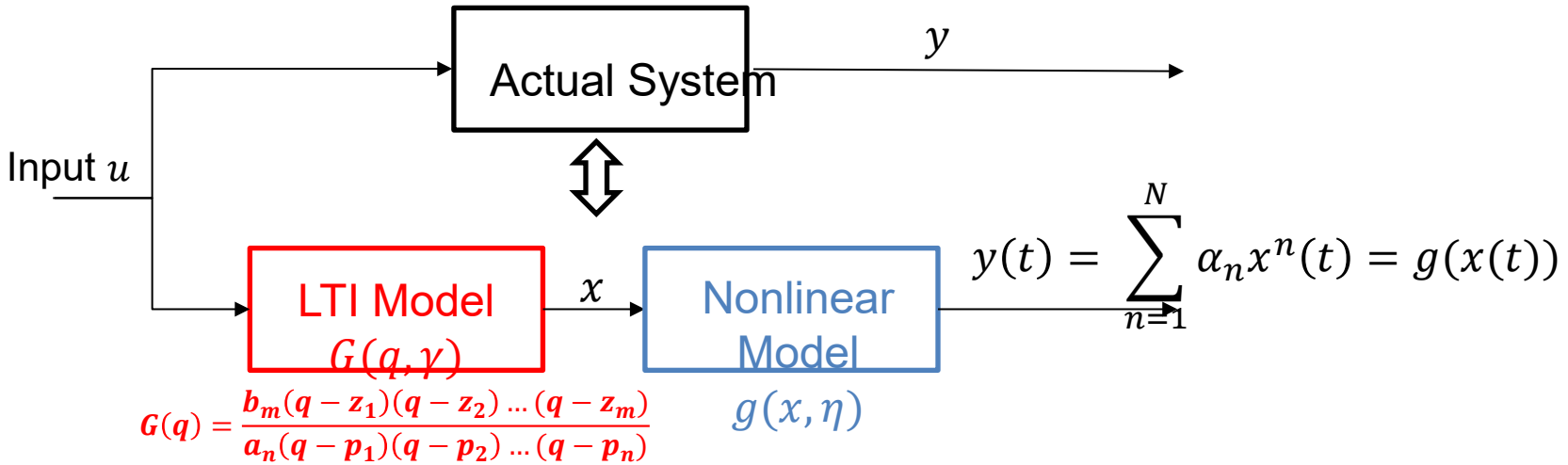
Volterra-Wiener Model System Identification

- Given a nonlinear time-invariant (NLTI) system along with the memory effects, the Volterra series expansion is widely used to represent such system as

$$y(t) = \sum_{n=1}^{N} \frac{1}{n!} \int h_n(\tau_1, \cdots, \tau_n) \prod_{i=1}^{n} u(t - \tau_i) d\tau_i$$

where $h_n(\tau_1, \tau_2, \ldots, \tau_n)$ is n^{th} order Volterra kernel.

- **Wiener model**: a special case of the Volterra series model, describes a static polynomial nonlinear model cascading with an LTI network.

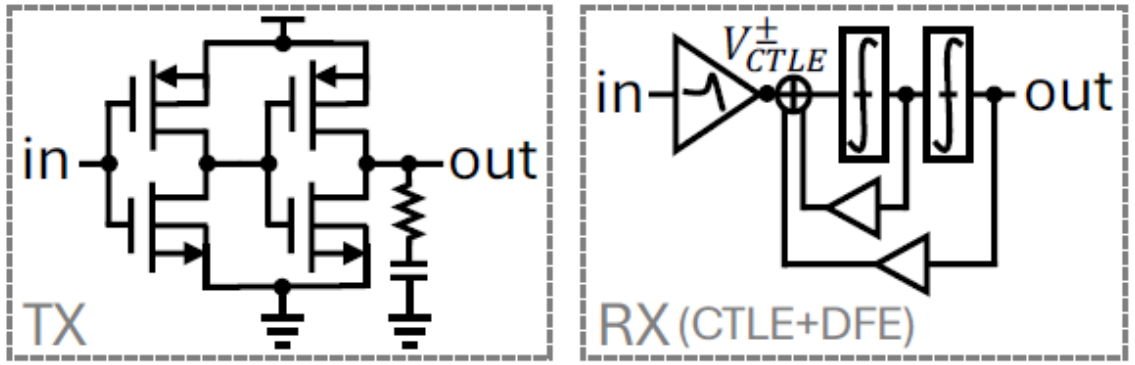


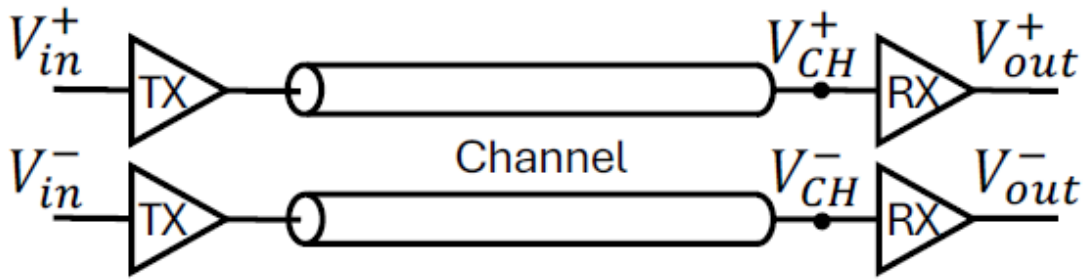
- A joint parameter vector $\theta = [\gamma^T, \eta^T]^T$ to describe the unknown intended to be found.
- γ represents the poles and zeros in the LTI system and η represents the polynomial coefficients in the nonlinear model.

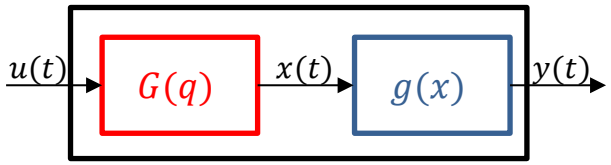
Statistical Analysis in Nonlinear System

High-speed Link System with Rx DFE and CTLE (1)

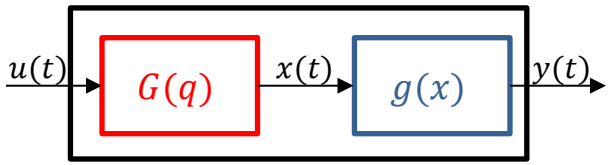
- A single-bit input pulse $u(t)$ at Vin is sent to the system to acquire the corresponding output response $y(t)$ at Vout.
- Modeling the $u(t) \Rightarrow y(t)$ system as Wiener model, which is consist of LTI block and polynomial block.



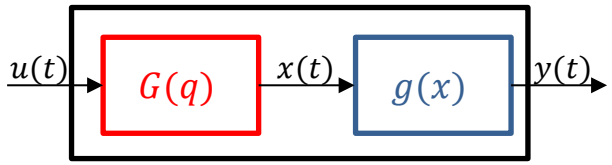




$V_{in} \Rightarrow V_{CH}$



$V_{in} \Rightarrow V_{CTLE}$

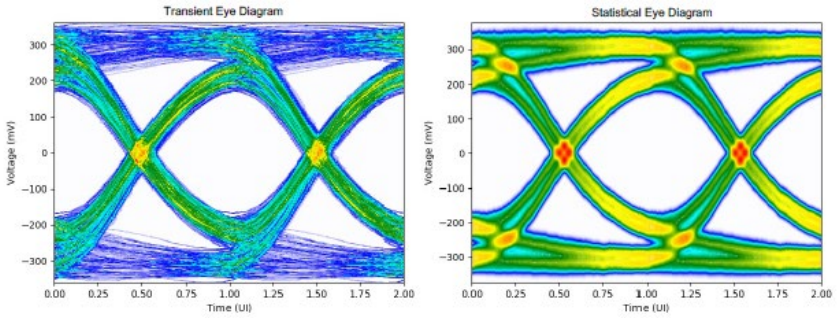


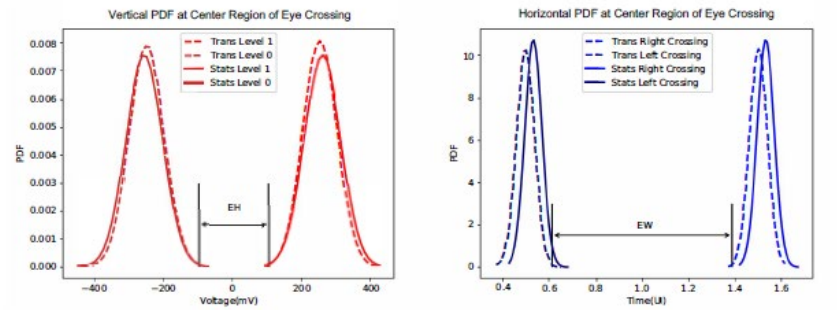
$V_{in} \Rightarrow V_{OUT}$

Statistical Analysis in Nonlinear System

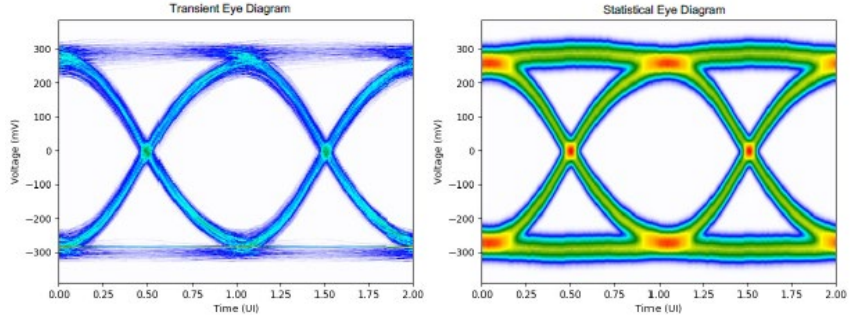
Example 3 – High-speed Link System with Rx DFE and CTLE (2)

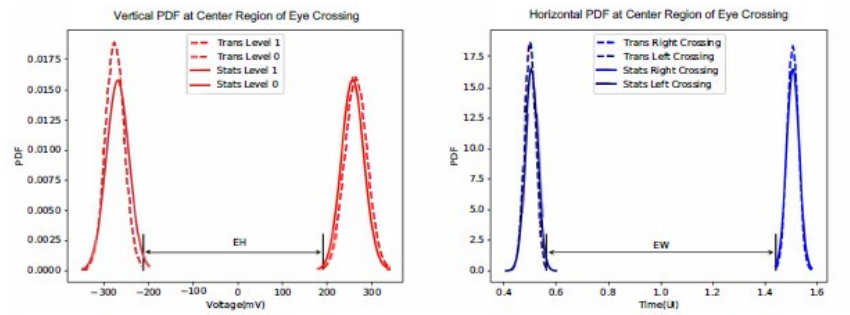
Eye Diagram Comparison at V_{CH}



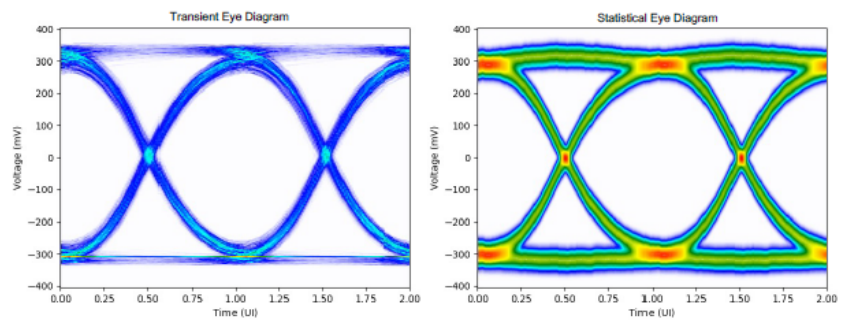


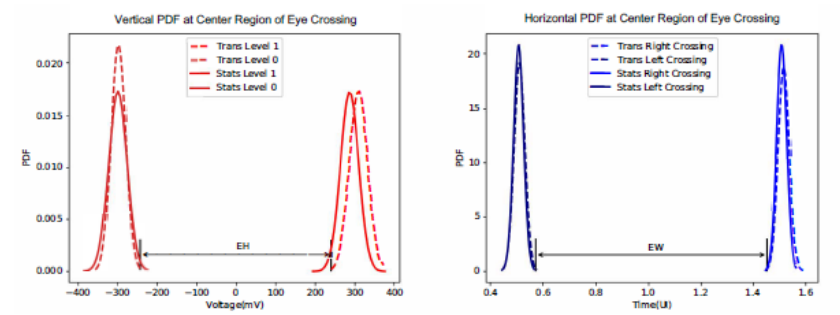
Eye Diagram Comparison at V_{CTLE}





Eye Diagram Comparison at V_{OUT}





Transient 1: 10^4bits PRE
Transient 2: 10^5bits PRE

	V_{CH}		V_{CTLE}		V_{OUT}	
	EH (mV)	EW (UI)	EH (mV)	EW (UI)	EH (mV)	EW (UI)
Transient 1	483.1616	0.8851	165.6665	0.6005	394.2349	0.5966
Transient 2	454.6452	0.8898	143.6442	0.5999	351.4231	0.5912
Statistical	447.9806	0.8782	137.3159	0.5922	344.7546	0.5928
Error 1	-7.28%	-0.78%	-17.11%	-1.38%	-12.55%	-0.64%
Error 2	-1.47%	-1.30%	-4.41%	-1.28%	-1.90%	-0.27%

	V_{CH}	V_{CTLE}	V_{OUT}
Transient 1	144 s		
Transient 2	43 m 16.9 s		
Statistical	16.58 s	5.53 s	5.96 s

Channel Modeling using ML*

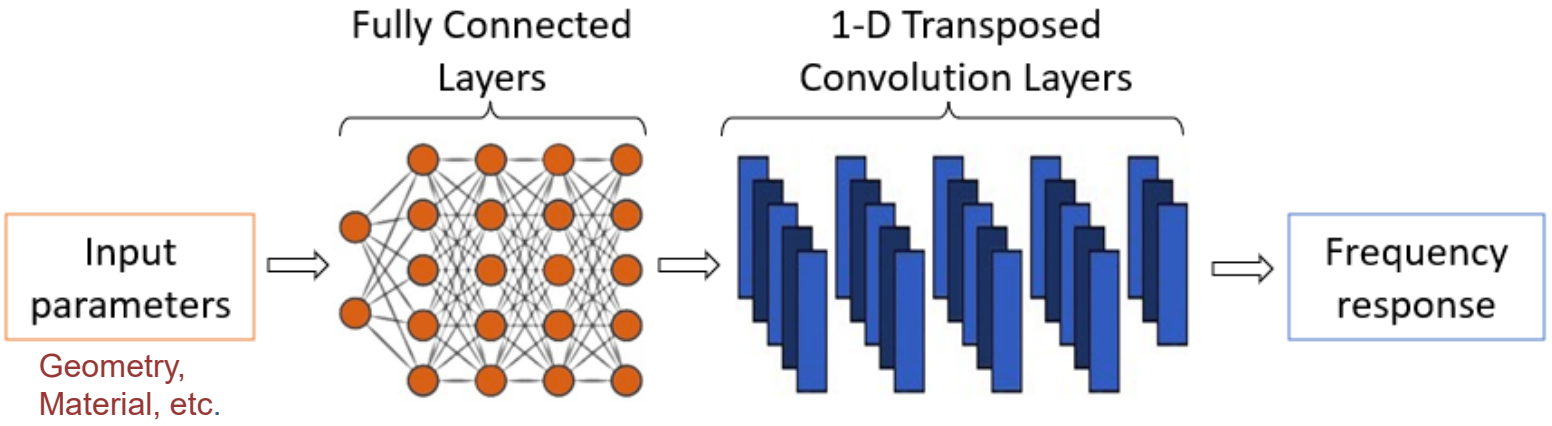
Goal: Create a fast channel model using Neural networks that can accurately predict S-parameters from various geometry information

Importance of using ML

- ❑To reduce computational complexity
- ❑For better understanding of the physical systems
- ❑Simplifies Design
- ❑Efficient Solutions

*Juhitha Konduru, Oleg Mikulchenko, Loke Yip Foo, Jose E. Schutt-Aine, "Signal Integrity Analysis and Design Optimization using Neural Networks", 2024 IEEE 74th Electronic Components and Technology Conference (ECTC), May 2024.

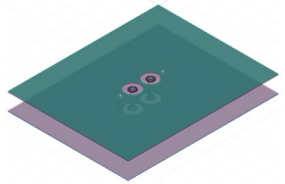
Neural Network Architecture



S-TCNN Architecture

Neural Network Architecture

Plated Through Hole layout Modeling



Physical Input parameters

Parameter	Unit	Min	Max
µVia Diameter	µm	30	70
µVia Pad Diameter	µm	31	140
µVia Antipad Radius	µm	100	500
PTH Pitch	µm	300	1200
Core Thickness	µm	100	1200
PTH Diameter	µm	100	250
PTH Pad Diameter	µm	110	500
PTH Antipad Radius	µm	50	500
S-G Via Pitch	µm	300	1200

*Juhitha Konduru, Oleg Mikulchenko, Loke Yip Foo, Jose E. Schutt-Aine, "Signal Integrity Analysis and Design Optimization using Neural Networks", 2024 IEEE 74th Electronic Components and Technology Conference (ECTC), May 2024.

- PTH structure is partially symmetric and reciprocal
  - the output frequency response of the model consists of the real and imaginary parts of the frequency responses of S_{11}, S_{12}, S_{13}, S_{14}, S_{33} and S_{34}
  - Broadband frequency response between 0.1 – 100GHz with 100MHz steps
  - Output dimensionality of the NN: 12000

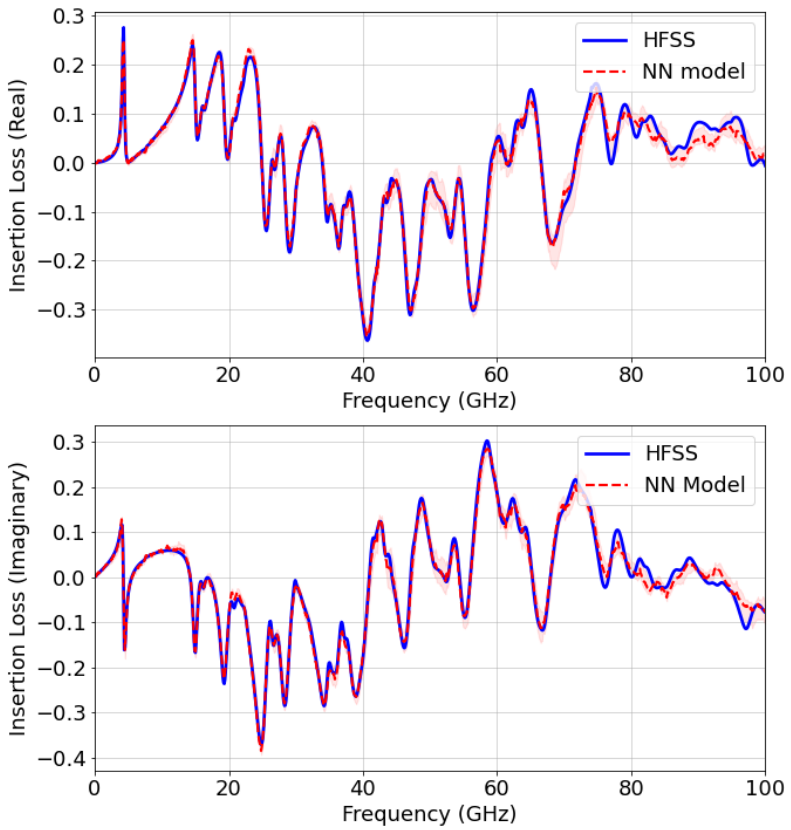
- 393 S-parameter samples of the PTH were generated by varying the 9 parameters
- Sobolset – Quasi random generator was used to determine the 393 samples.
- 390 sample are used for training and 3 samples are used for validation

Results - Training

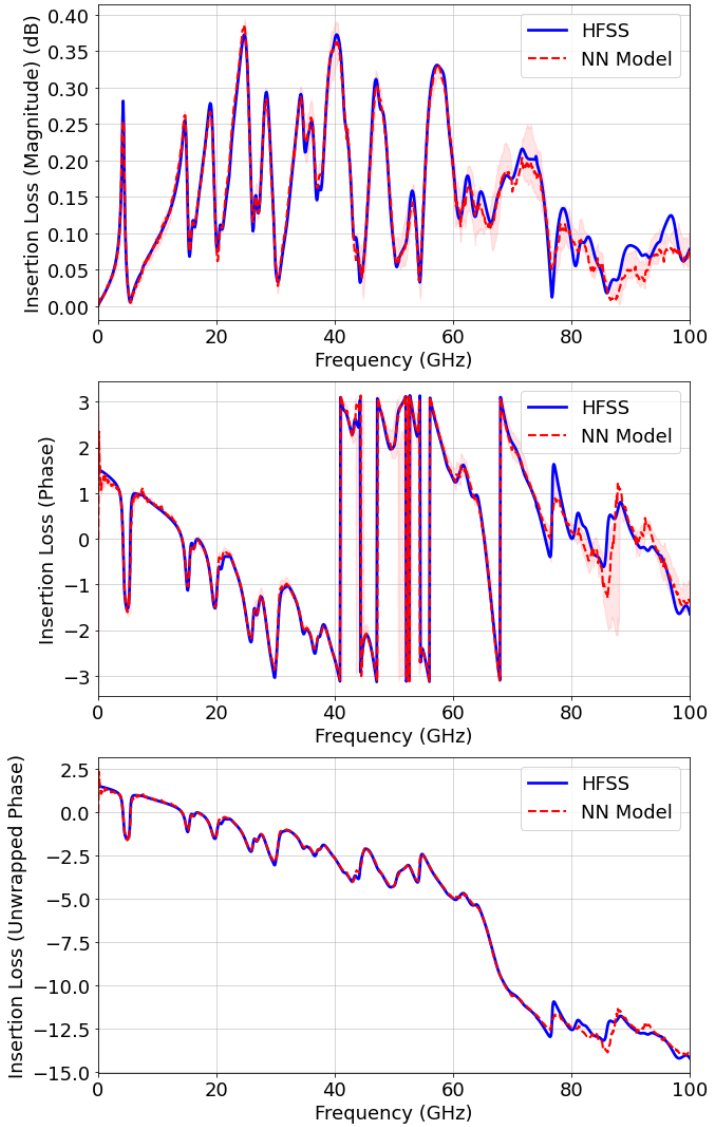
- STCNN model can predict S-parameter frequency responses very accurately
- Validation NMSE: 8%

Run-time comparison to generate 1000 different S-parameters

Mode	Run-time
Each HFSS simulation	40mins = 2400s
Dataset generation	393x2400s = 943200s
Training STCNN model	250s



(a)



(b)

S-parameters comparison between HFSS and S-TCNN Model a) Real-Imaginary plots, b) Magnitude-Phase-Unwrapped phase plots.

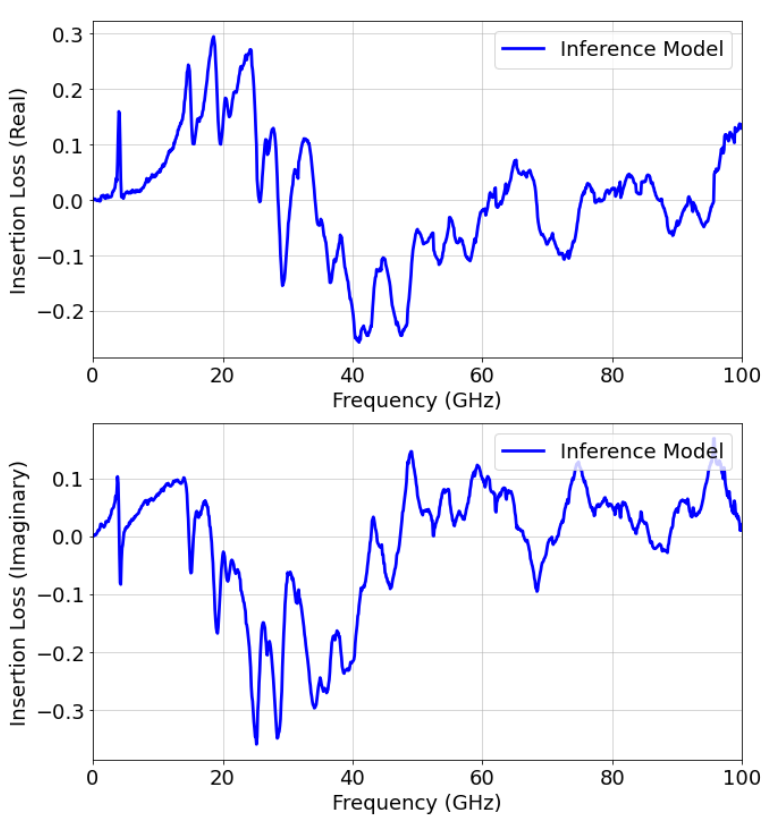
Inference Results

Input parameter values

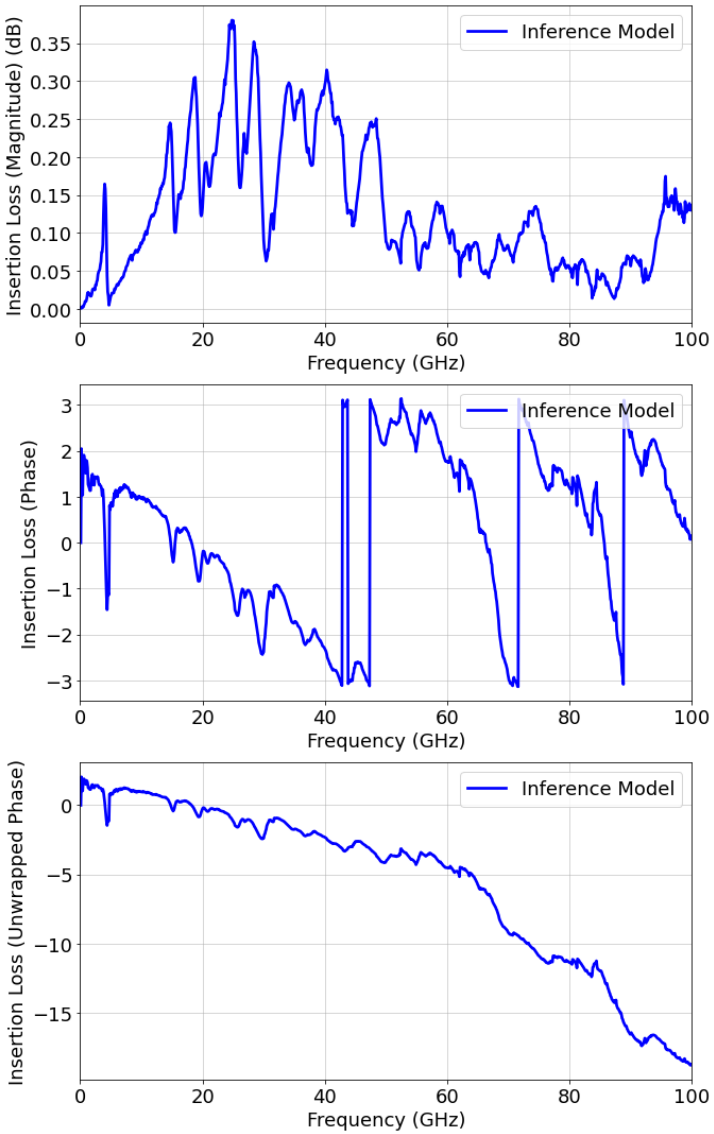
Parameter	Value	Min	Max
µVia Diameter	45	30	70
µVia Pad Diameter	120	31	140
µVia Antipad Radius	120	100	500
PTH Pitch	600	300	1200
Core Thickness	600	100	1200
PTH Diameter	120	100	250
PTH Pad Diameter	150	110	500
PTH Antipad Radius	300	50	500
S-G Via Pitch	600	300	1200

Run-time comparison to generate 1000 different S-parameters

Mode	Run-time
HFSS simulation	40mins = 2400s
Fast NN Model	0.2s (12000X speedup)



(a)



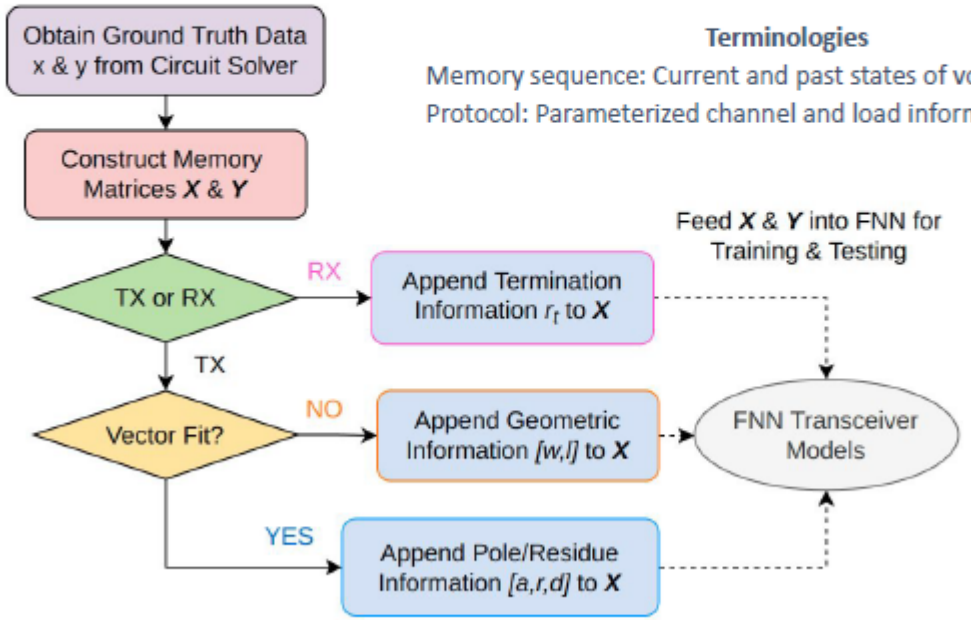
(b)

S-parameters generated from Inference Model a) Real-Imaginary plots, b) Magnitude-Phase-Unwrapped phase plots.

FFN for Cascade-able Transceiver

We developed a **feed-forward neural network (FNN) based modeling technique** that generates ***cascade-able*** transceiver blocks for HSL transient analysis. We verified the method by accurately modeling nonlinear CMOS and NAND buffers.

To train/generate the FNN TX/RX models:



Terminologies

Memory sequence: Current and past states of voltage inputs.

Protocol: Parameterized channel and load information.

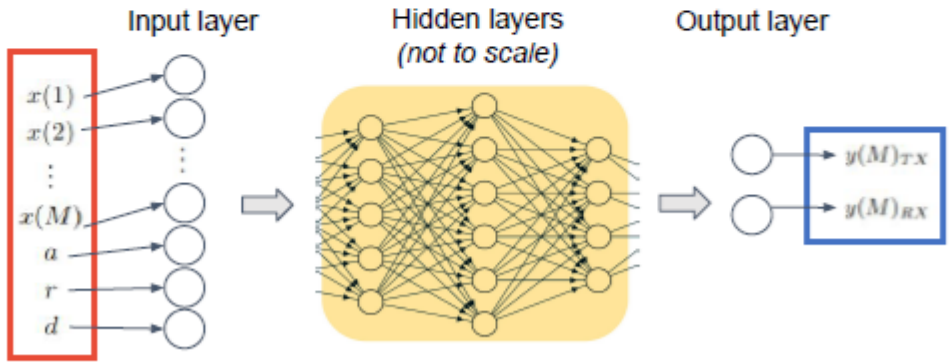
FNN modeling workflow.

Input matrix to FNN

$$\begin{bmatrix} x(1) & x(2) & \cdots & x(M) & a & r & d \\ x(2) & x(3) & \cdots & x(M+1) & a & r & d \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x(n-M) & x(n-M+1) & \cdots & x(n) & a & r & d \end{bmatrix}$$

Output matrix of FNN

$$\begin{bmatrix} y(M)_{TX} & y(M)_{RX} \\ y(M+1)_{TX} & y(M+1)_{RX} \\ \vdots & \vdots \\ y(n)_{TX} & y(n)_{RX} \end{bmatrix}$$



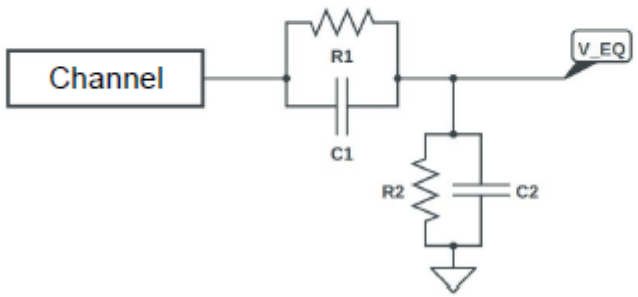
FNN structure for training TX with VF protocol.

Yixuan Zhao, Thong Nguyen, Hanzhi Ma, Er-Ping Li, Andreas C. Cangellaris, Jose E. Schutt-Aine, "Modular Neural Network-Based Models of High-Speed Link Transceivers", IEEE Transactions on Components, Packaging and Manufacturing Technology, Vol. 13, Issue 10, 2023

Application to CTLE

What is CTLE:

CTLE stands for *continuous-time linear equalizer*. A passive CTLE of one tap can be achieved by two sets of RC circuit.

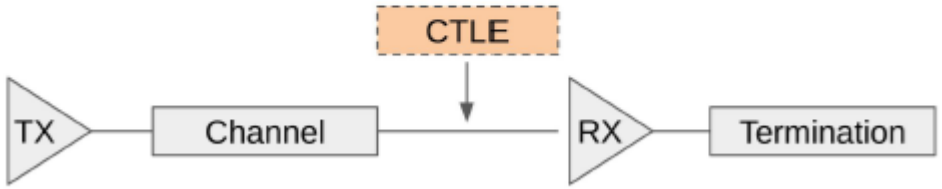


$$\omega_z = \frac{1}{R_1 C_1}$$

$$\omega_p = \frac{1}{\frac{R_1 R_2}{R_1 + R_2}(C_1 + C_2)}$$

$$A_{DC} = \frac{R_2}{R_1 + R_2}$$

Example circuit of one tap passive CTLE.



CTLE located in the interconnect channel.

Why we wish to include CTLE:

- In HSL, high frequency signals are attenuated more than the lower ones → inter-symbol interference (ISI).
- CTLE attenuates the low frequency signal and boost the high frequency signal → larger opening in the eye diagram.
- Many modern RX designs contains CTLE option by specifying the pole/zero frequency (ω_p/ω_z) and DC gain (A_{DC}).

Yixuan Zhao, Thong Nguyen, Hanzhi Ma, Er-Ping Li, Andreas C. Cangellaris, Jose E. Schutt-Aine, "Modular Neural Network-Based Models of High-Speed Link Transceivers", IEEE Transactions on Components, Packaging and Manufacturing Technology, Vol. 13, Issue 10, 2023

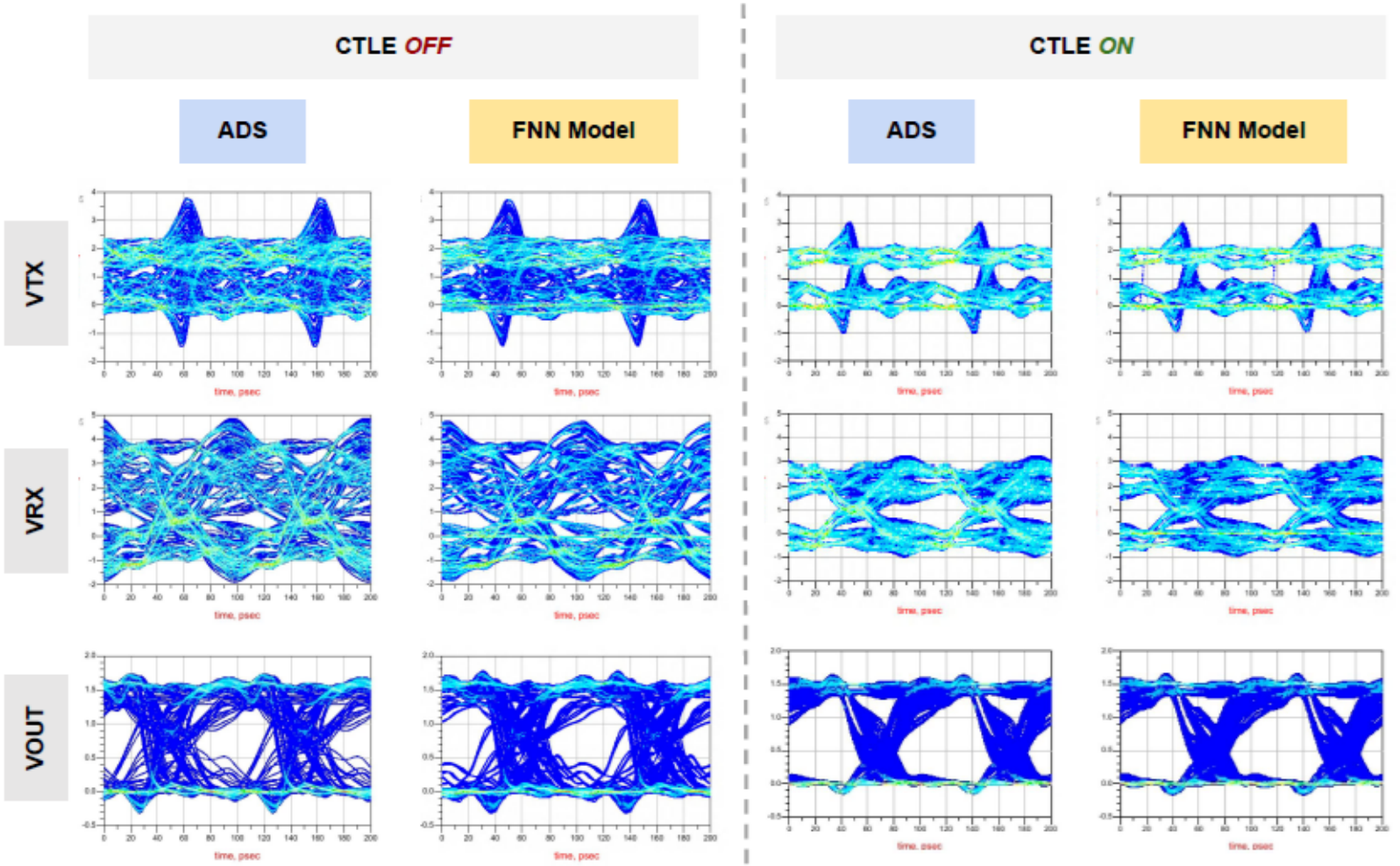
CTLE Results

Error in Eye Measurement

Percentage Error (%)		Eye Width	Eye Height	Zero Level	One Level
CTLE OFF	VTX	N/A			
	VRX	5.33	0	1.18	1.36
	VOUT	3.71	2.15	0	0.79
CTLE ON	VTX	0.40	2.12	1.06	0.17
	VRX	0.57	1.33	2.70	0.31
	VOUT	2.60	0	0	0.07

Good correlation: Predicted eye errors < 6 %
Time efficiency: Speed up eye analysis x 25

FNN models always **underestimates** the eyes:
→ **Tighter timing budget (preferred)**.
→ Might result in minor over-design.
→ Never result in false-design.



FNN eye predictions of V_{TX} , V_{RX} and V_{OUT} without and with CTLE.

Yixuan Zhao, Thong Nguyen, Hanzhi Ma, Er-Ping Li, Andreas C. Cangellaris, Jose E. Schutt-Aine, "Modular Neural Network-Based Models of High-Speed Link Transceivers", IEEE Transactions on Components, Packaging and Manufacturing Technology, Vol. 13, Issue 10, 2023