ECE 546
Lecture -27
Co-Design for HI

Spring 2026

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

AI Energy Requirements

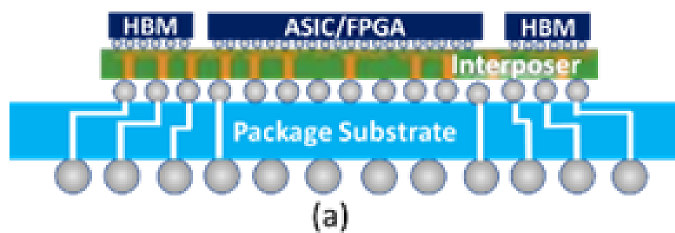
- **Data centers used 2.5% of US electricity in 2022**
- **Projected to increase to 20% by 2030**

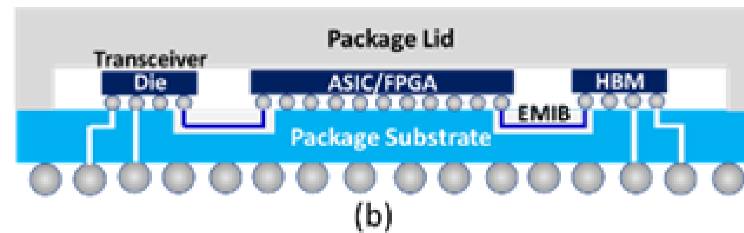
Training a large language model like **GPT-3** is estimated to use **1,300 megawatt hours (MWh)** of electricity.

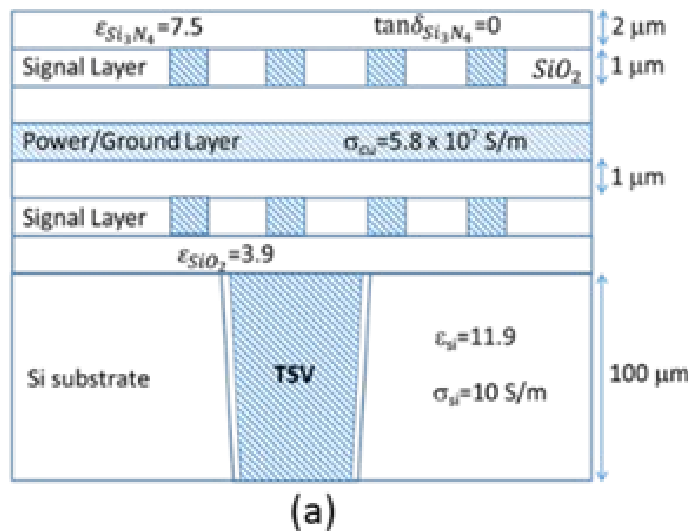
80% of power consumption is due to data movement through interconnects

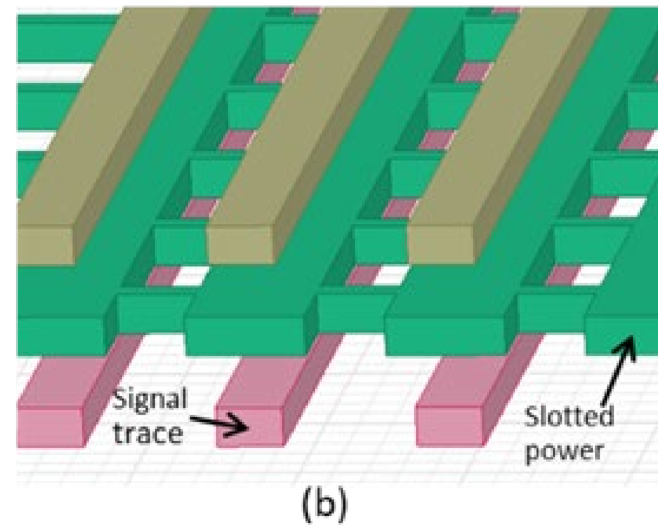
HI – The Interconnect Challenge

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





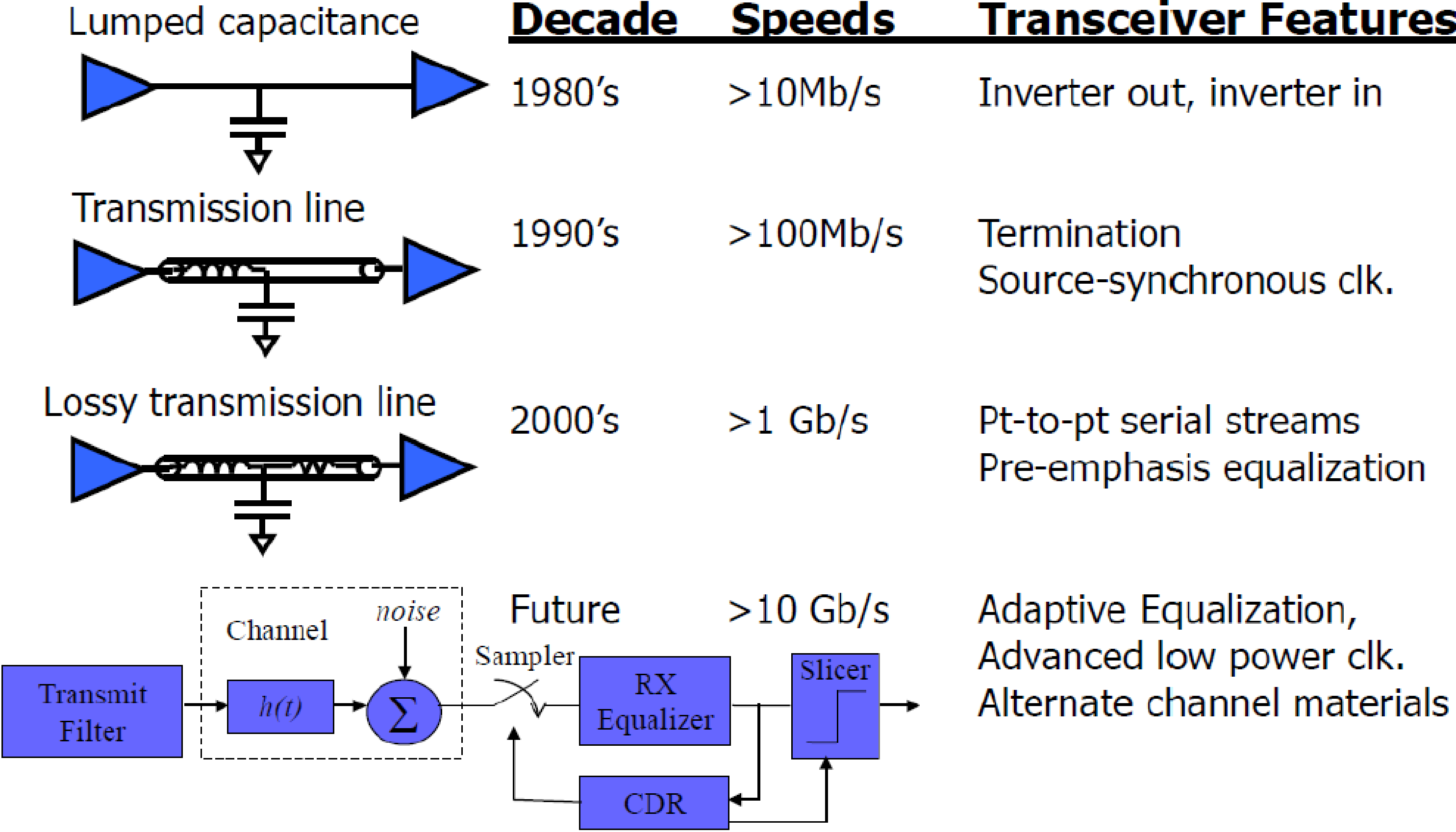




Challenges

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

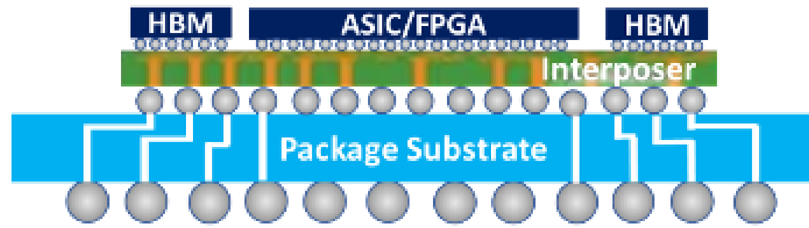
Example - Interconnect Evolution

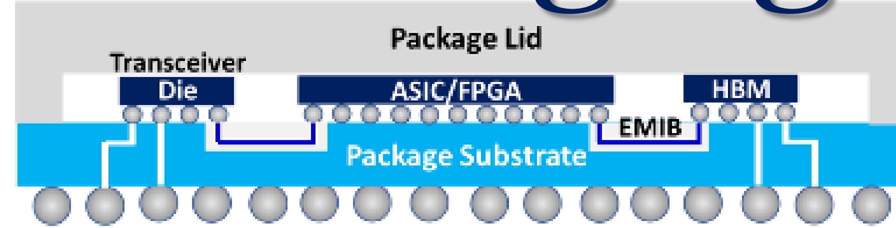


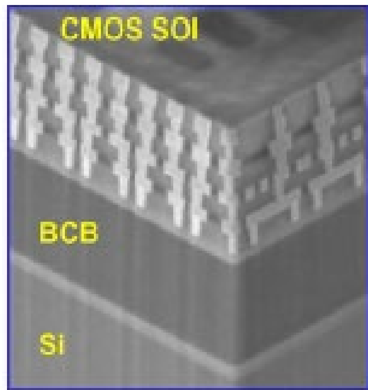
Decade	Speeds	Transceiver Features
1980's	>10Mb/s	Inverter out, inverter in
1990's	>100Mb/s	Termination Source-synchronous clk.
2000's	>1 Gb/s	Pt-to-pt serial streams Pre-emphasis equalization
Future	>10 Gb/s	Adaptive Equalization, Advanced low power clk. Alternate channel materials

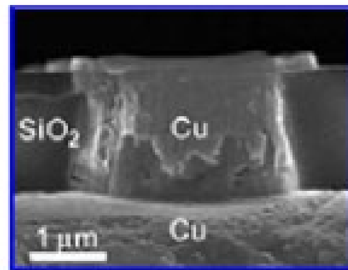
Slide Courtesy of Frank O'Mahony & Brian Casper, Intel

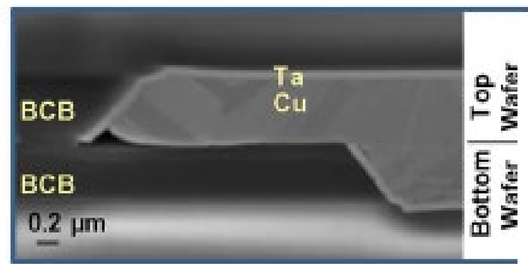
Advanced Packaging

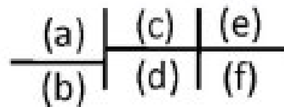


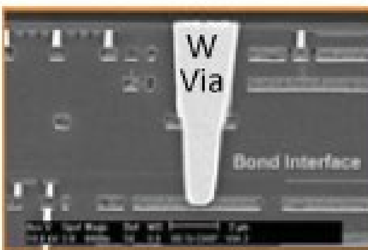


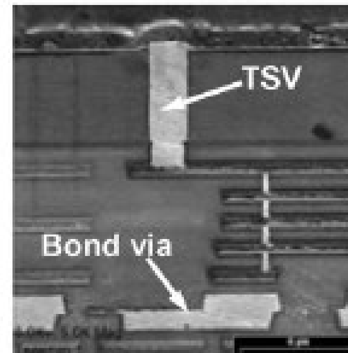


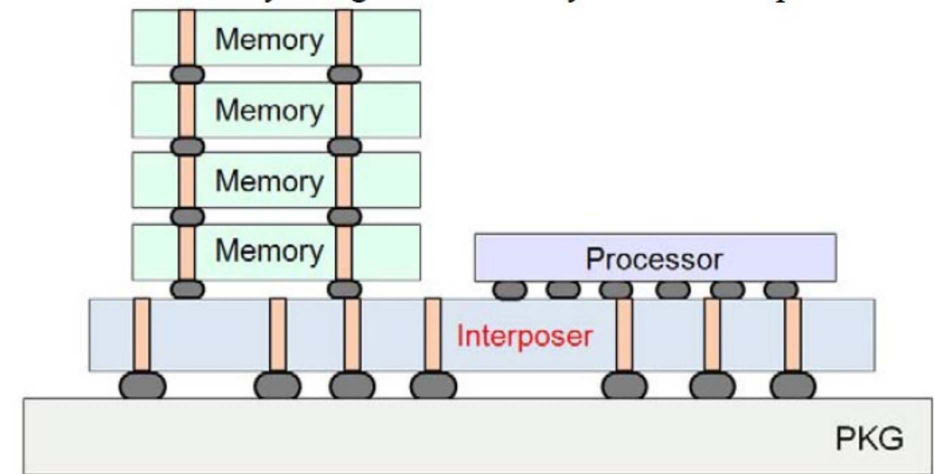








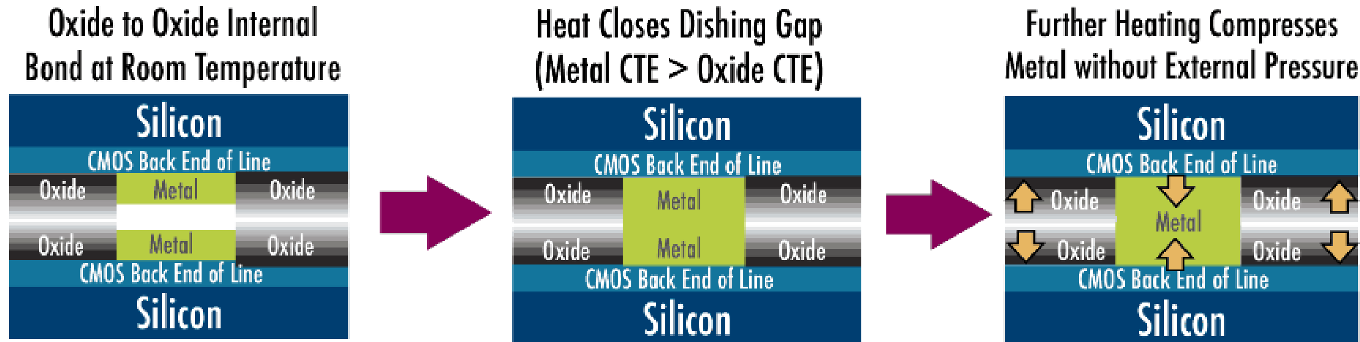




In order to meet the demands of AI, HPC, and next-generation data centers, advanced packaging will be key to performance, power efficiency, and scalability.

Hybrid Bonding

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



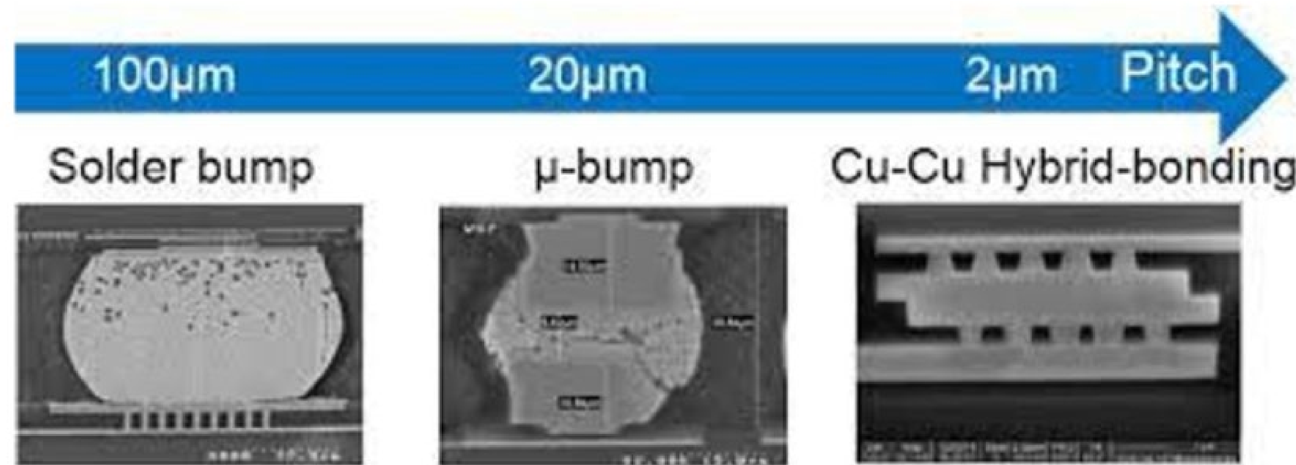
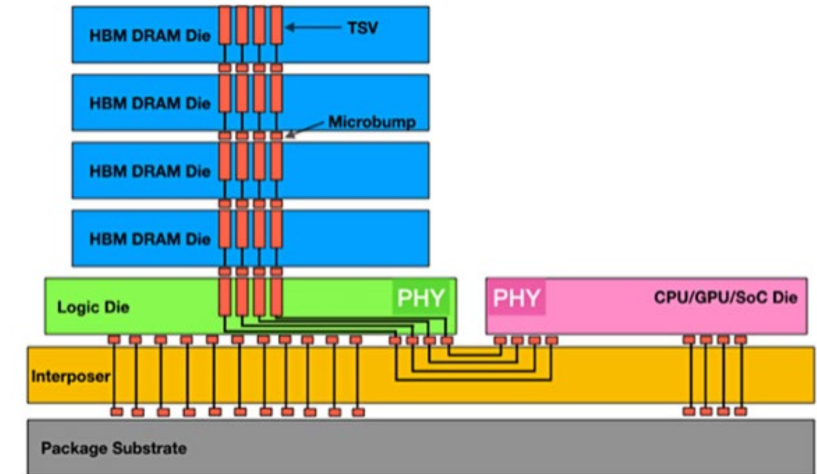


Image Credit: Imed Jani. Test and characterization of 3D high-density interconnects. Micro and Nanotechnolo- gies/Microelectronics. Université Grenoble Alpes, 2019. English. NNT : 2019GREAT094 . tel- 02634259



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

Objectives for HI

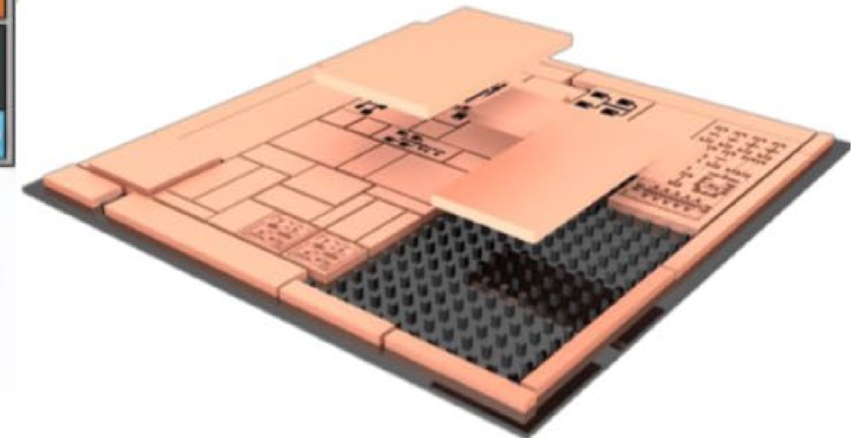
- Focus on minimizing energy and delay
- Move away from Von Neumann computing model
- Identify and address conflicting requirements,
- Make use of new physics and emerging technologies
- Future applications are drivers
- End of Moore's Law necessitates new paradigm

Universal co-design implies collaboration between many disciplines with the common goal of improved performance at lower cost

Need for: Heterogeneous Integration

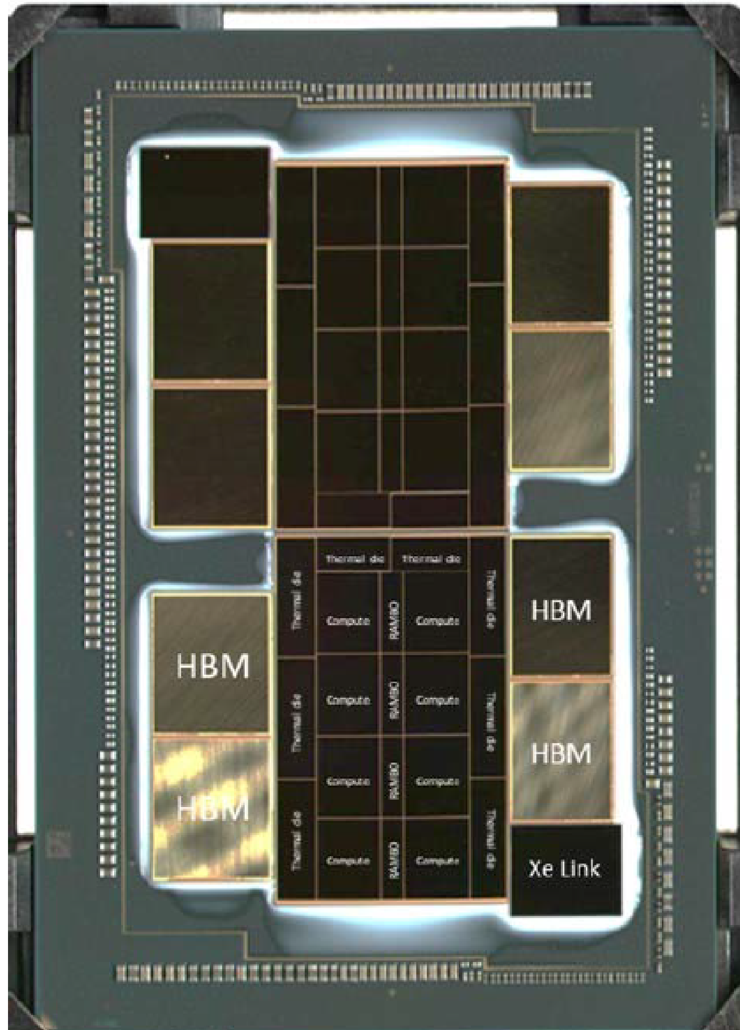
- Motivation: Ubiquitous materials and technologies
  - Die, interposer, substrate
  - Scaling is with wavelength rather than technology
- Opportunity: New interconnect technologies
  - Increased density of I/Os, finer pitches
  - 2.5D/3D integration
- Performance objective: minimize energy and delay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

Chiplet-Based Design: AMD EPYC



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



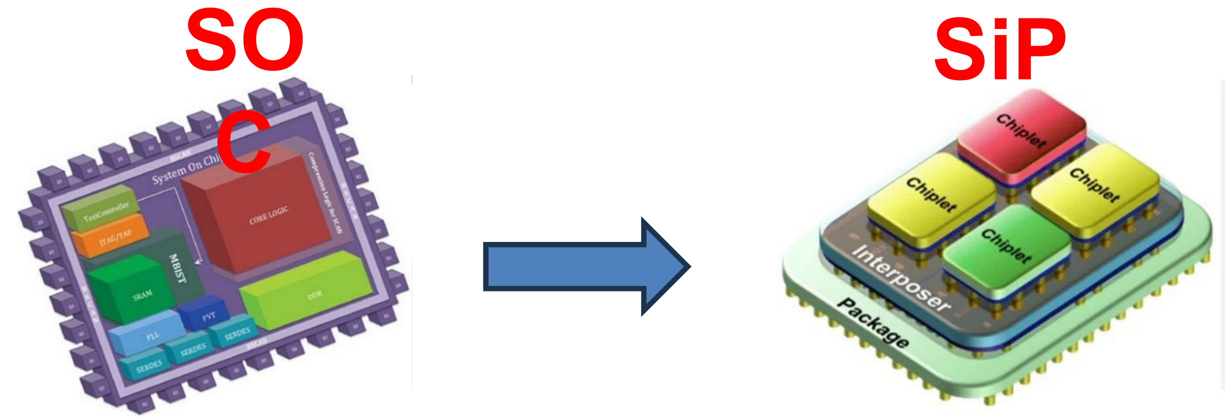
Integration	Foveros + EMIB
Power Envelope	600W
Transistor count	> 100B
Total Tiles	63 (**47 functional** + 16 thermal Tiles)
HBM count	8
Package Form factor	77.5 x 62.5 mm (4844 mm^2)
Platforms	3 platforms
IO	4x16 90G SERDES, 1x16 PCIe Gen5
Total Silicon	3100 mm^2 Si
Silicon footprint	2330 mm^2 Si footprint
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

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



What is Co-Design?

- Level Co-design
  - Chip
  - Package
  - Board
- Physics-based Co-design
  - Thermal
  - Electrical
  - Mechanical
  - Optical
- Function-based Co-design
  - Thermal aware
  - Signal integrity aware
  - Testability
  - Security aware
- Domain-based Co-design
  - Hardware
  - Software
  - Architecture

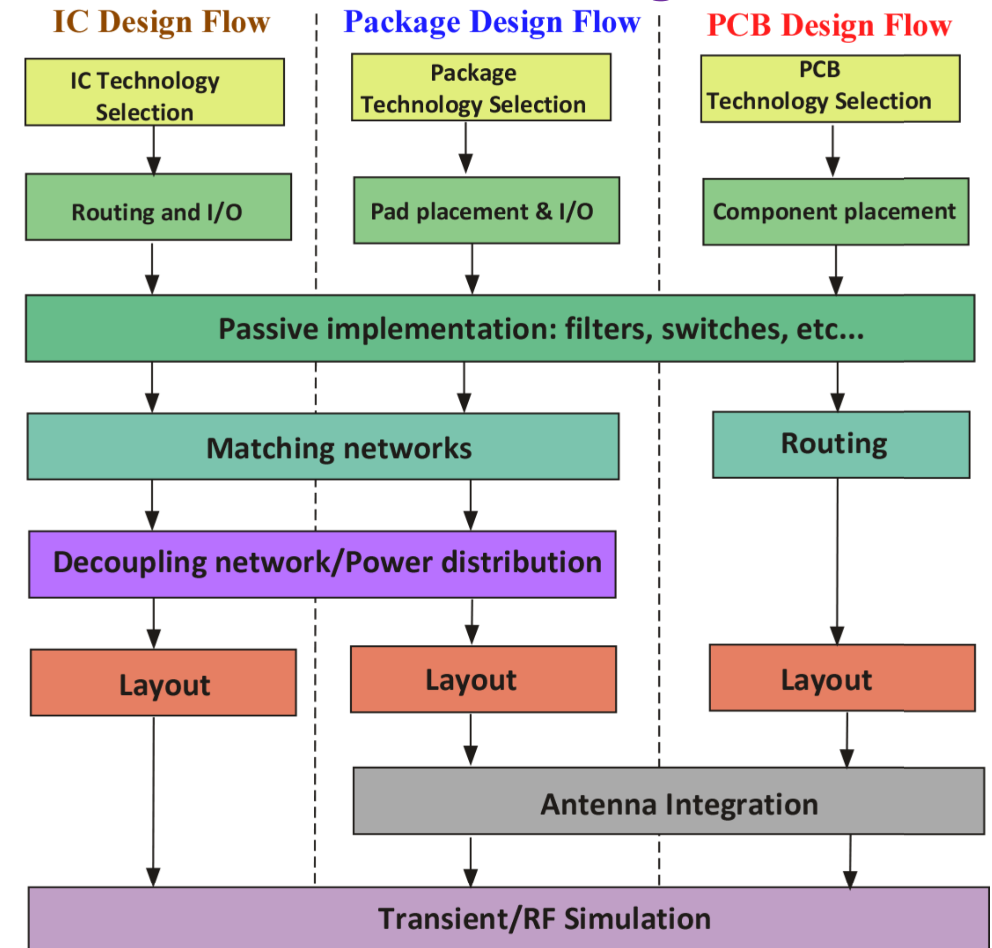
Design Flow

VERTICAL

- Bridge IC, Package, Board
- Bridge System, Architecture, Layout
- Bridge Synthesis, Analysis, Verification
- Bridge Hardware, Software, Firmware

HORIZONTAL

- Energy aware
- Signal/power integrity aware
- Stress/thermal aware
- Security aware
- Testing aware



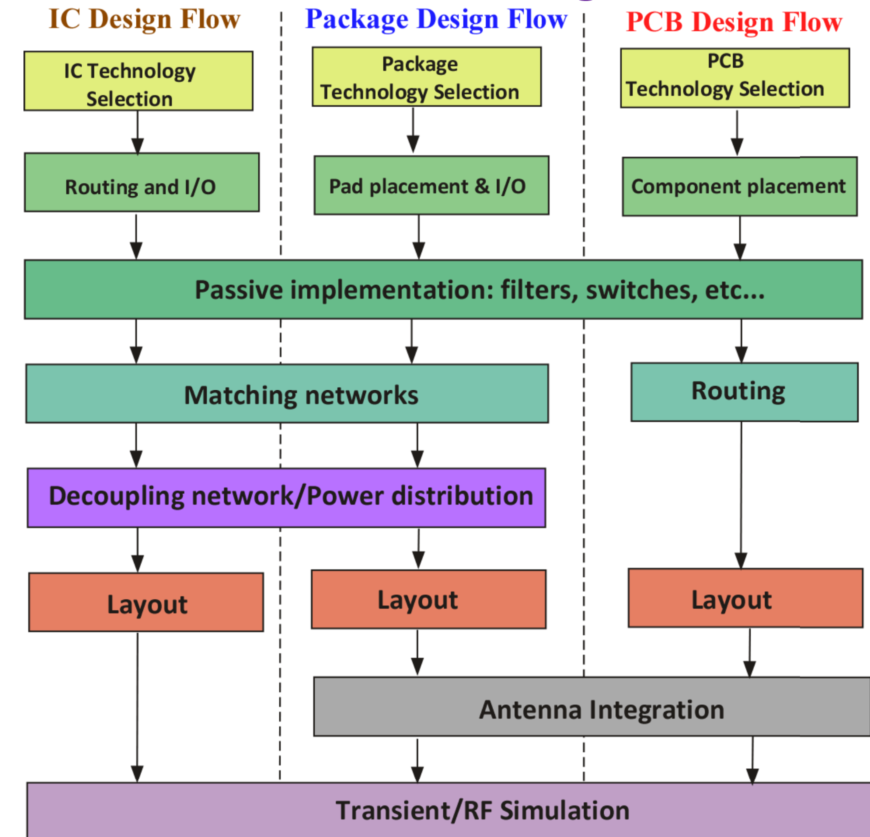
Co-Design

VERTICAL

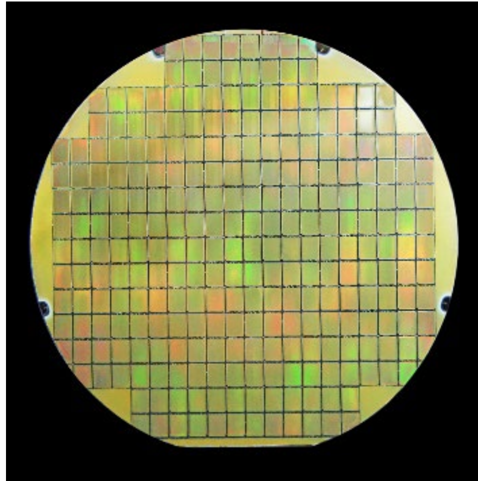
- Bridge IC, Package, Board
- Bridge System, Architecture, Layout
- Bridge Synthesis, Analysis, Verification
- Bridge Hardware, Software, Firmware

HORIZONTAL

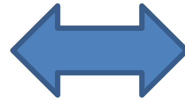
- Energy aware
- Signal/power integrity aware
- Stress/thermal aware
- Security aware
- Testing aware

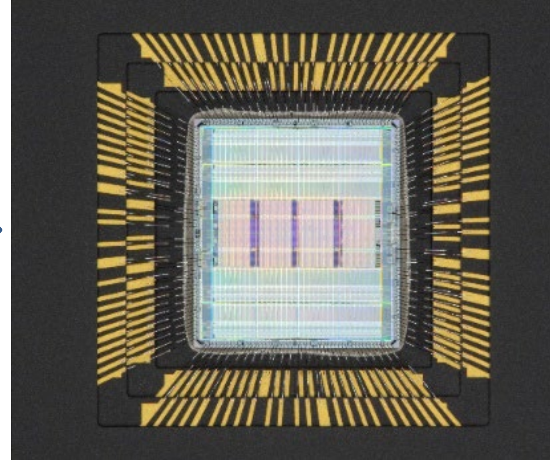


Traditional Co-Design

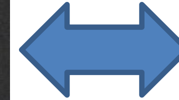


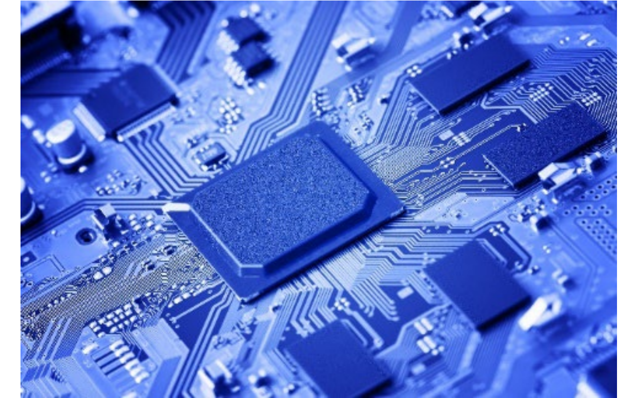
CHIP





PACKAGE





BOARD

- Transistors
- Nonlinear
- SPICE
- Scaling with tech

- Interconnects
- Linear
- EM Tools
- Scaling with λ

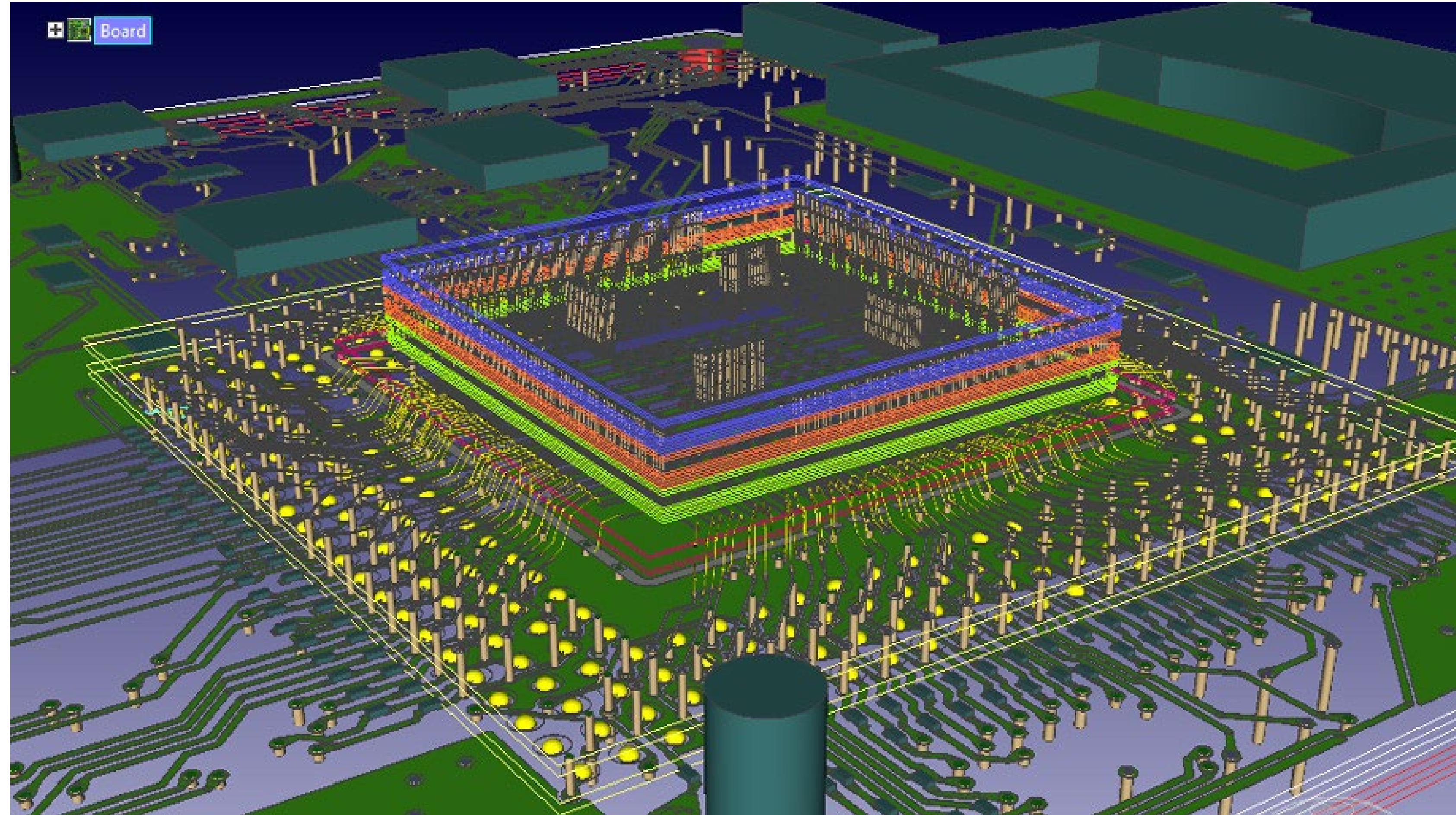
- Transmission lines, sensors
- Linear+Nonlinear
- EM Extraction, SPICE, IBIS,…
- Scaling with λ

Key Questions

- What is the state of the art in co-design?
- What system products will drive creation of new co-design tools/methods?
- What are the key challenges that need to be overcome?
- What added value will new co-design tools/methods bring to those products?
- What needs to happen for these challenges to be overcome?
- When will answers to the above become apparent?

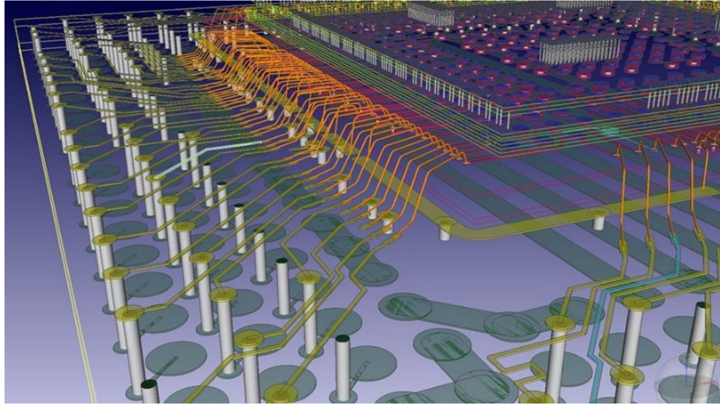
Co-Design Requirements

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

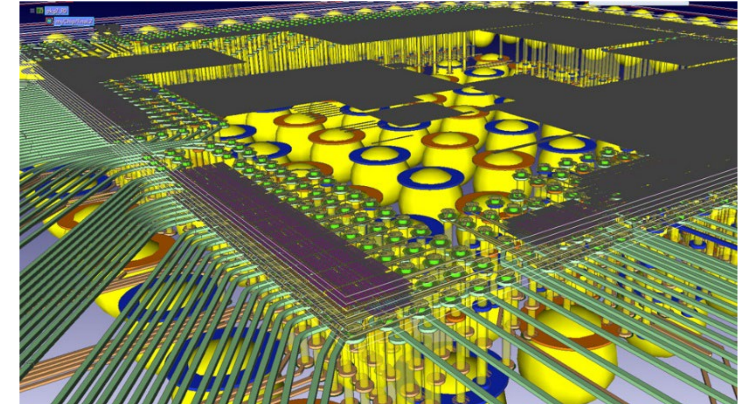


Courtesy of Zuken

Pathfinding Methodologies



Courtesy of Zuken



Courtesy of Zuken

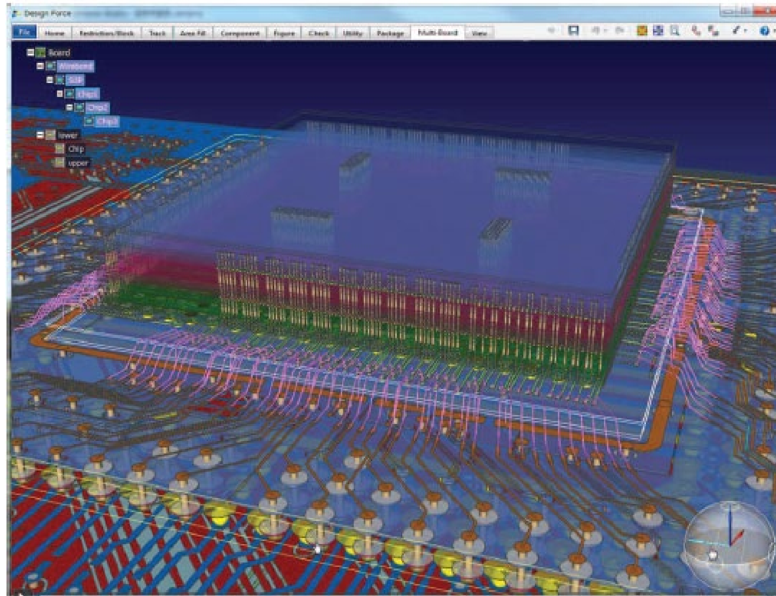
- Unified workflow, including partitioning, floor-planning, design of system-level interconnects, route pathway exploration and feasibility analysis. Capability to create abstract package models and virtual die models from multiple sources

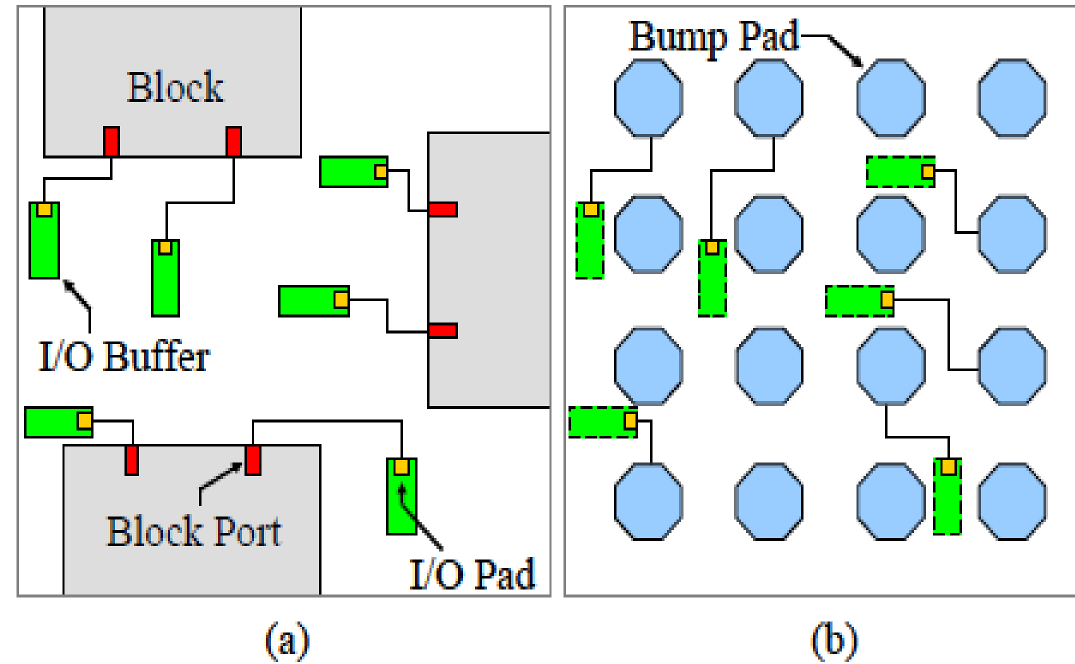
- Ability to visualize and modify component placement scenarios and make connectivity changes in a preliminary floorplan. Provision of dynamic manipulation of pin arrays within the abstract models

- Preserving signal assignments and rules while making adjustments to the physical pin array. Support for multiple package variables and PCB form factors to verify and compare different system configurations

- Enabling the interaction of design tools from different EDA vendors

Challenges for Placement and Routing

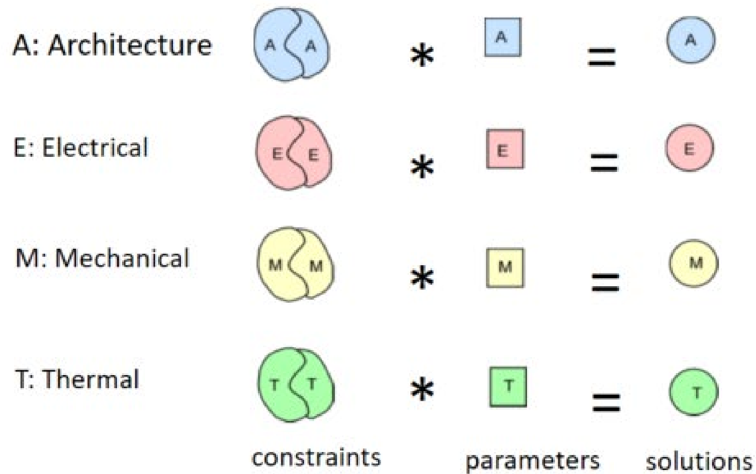


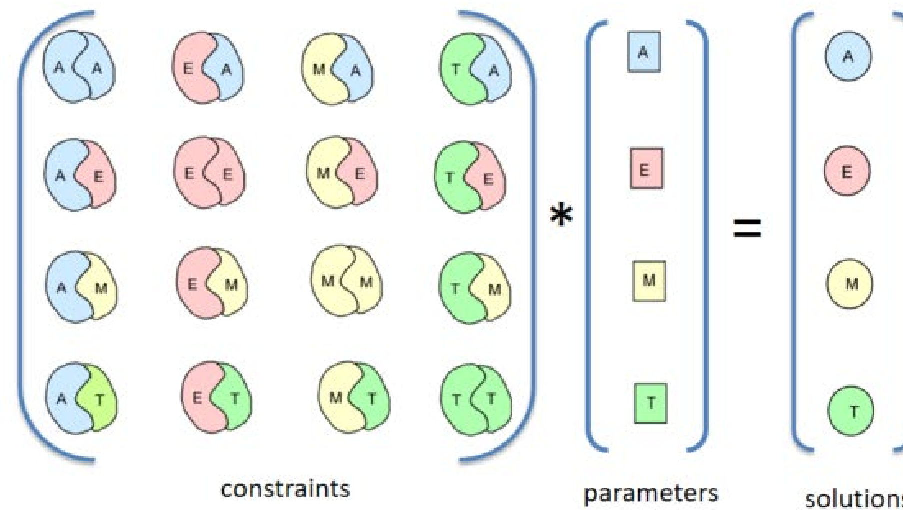


Need to concurrently manage connectivity between intra- and inter-levels

Co-Design Challenges

Current co-design methods are simply a cascade or combination of independent solutions. Real co-design requires concurrent solution from a formulation that accounts for all multi-physics interactions while embodying conflicting requirements

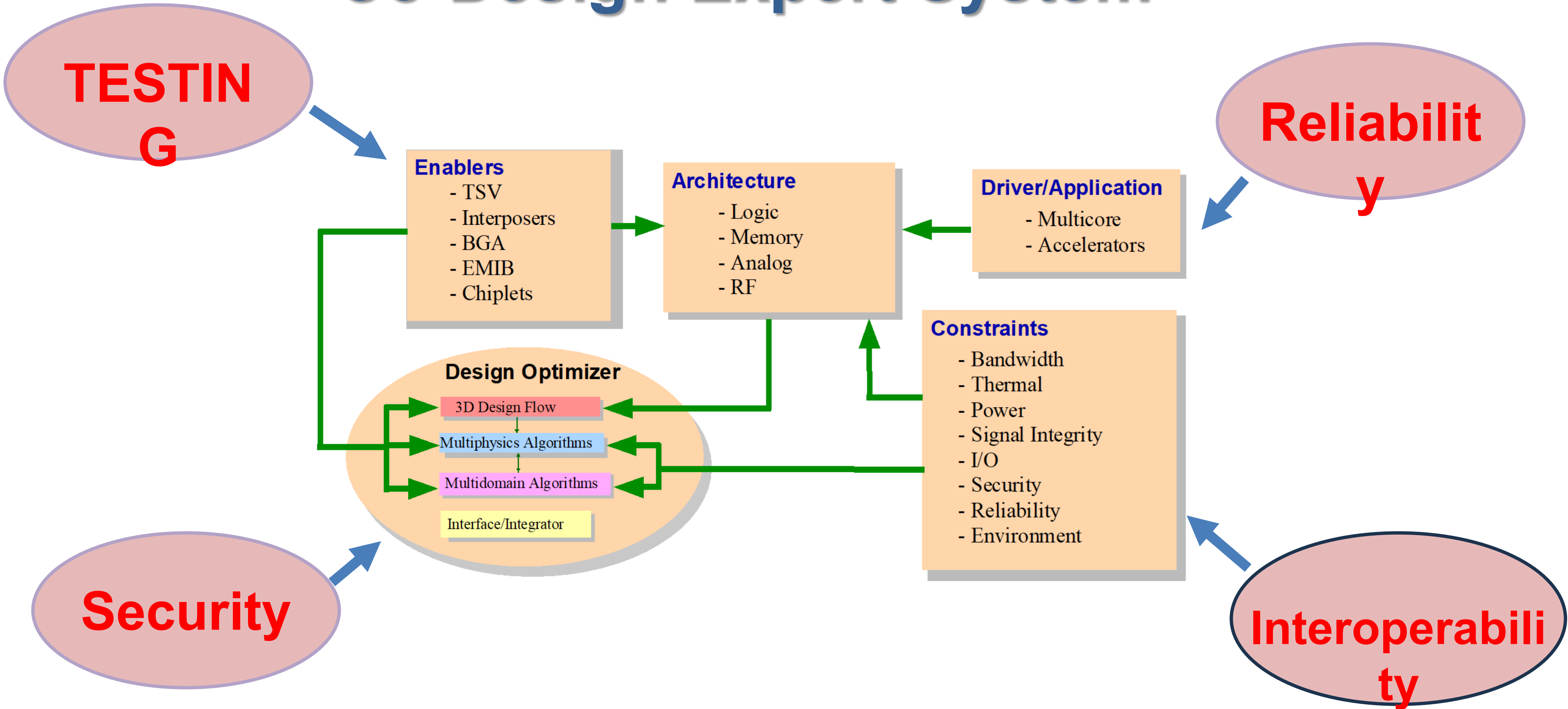




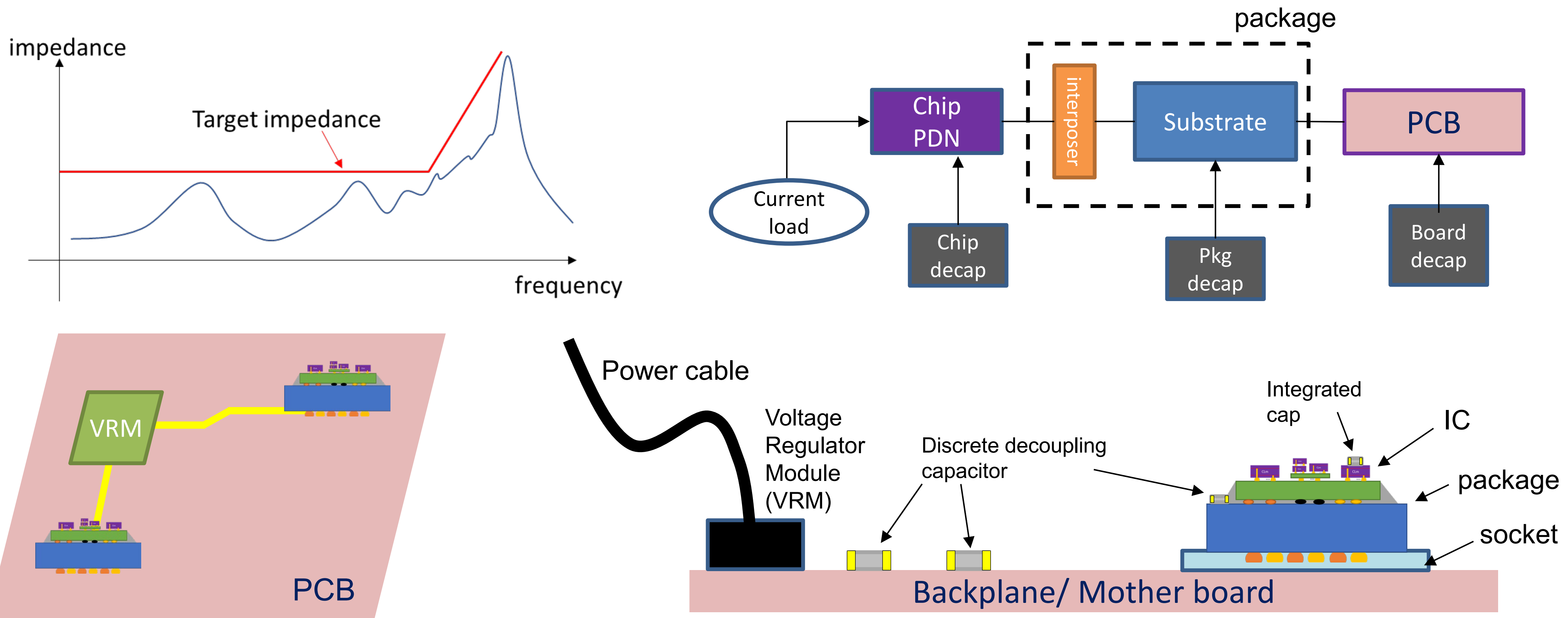
Constraints

- Bandwidth
- Thermal
- Power
- Signal Integrity
- I/O
- Security
- Reliability
- Environment
- Size
- Cost

Co-Design Expert System



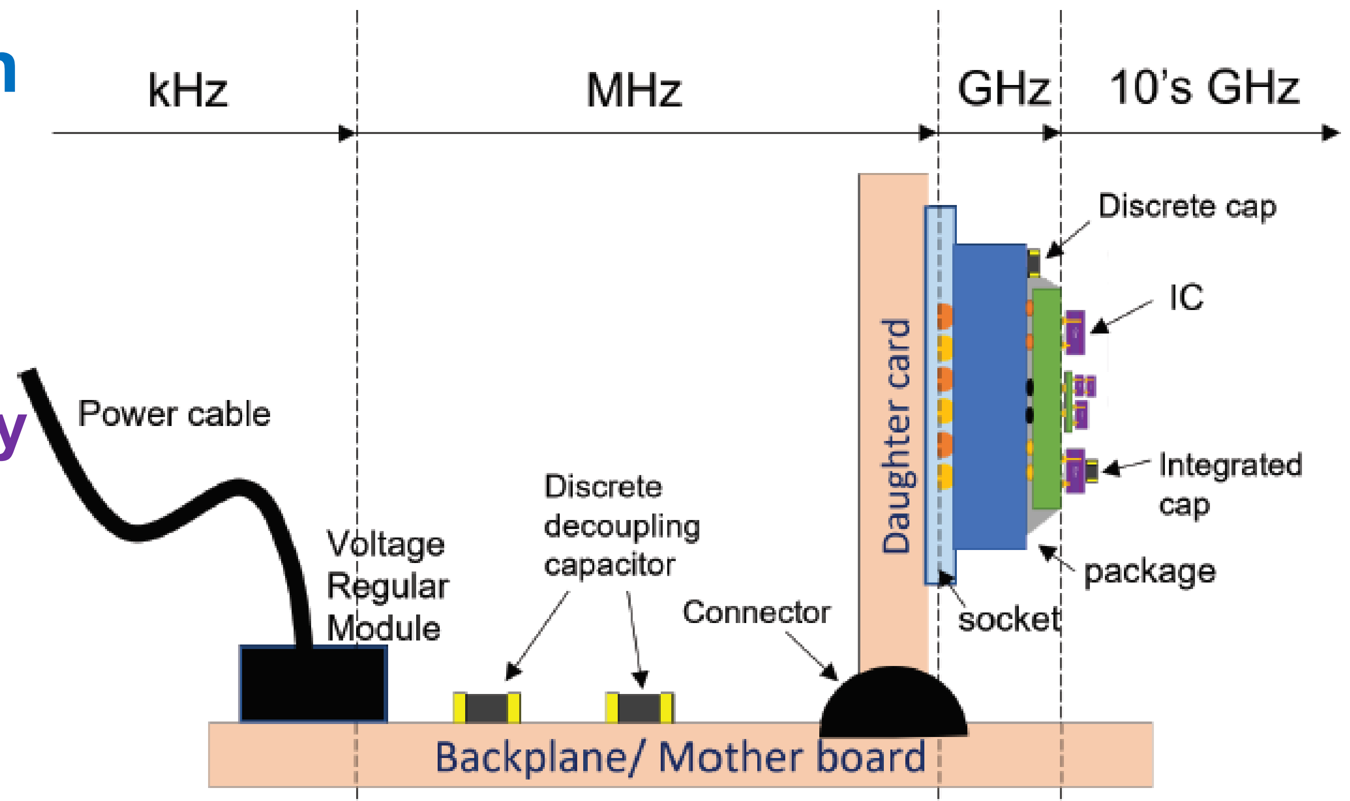
Power Distribution Network (PDN)



Goal: Achieve target impedance through use of decoupling capacitors at chip, package and board levels

Chiplets and Heterogeneous Integration

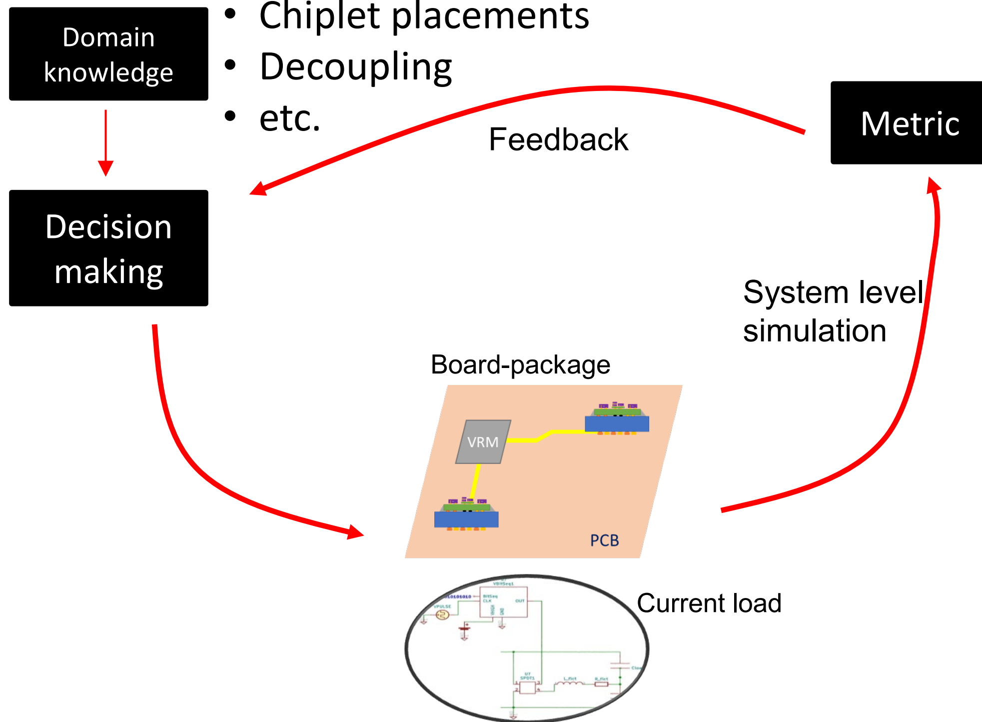
- **Multilevel Power Distribution**
  - Chiplets advantageous
  - Boundary between chip and package is blurred
  - Concurrent design is necessity
  - SI/PI, multilevel frequencies
  - System-level verification challenging

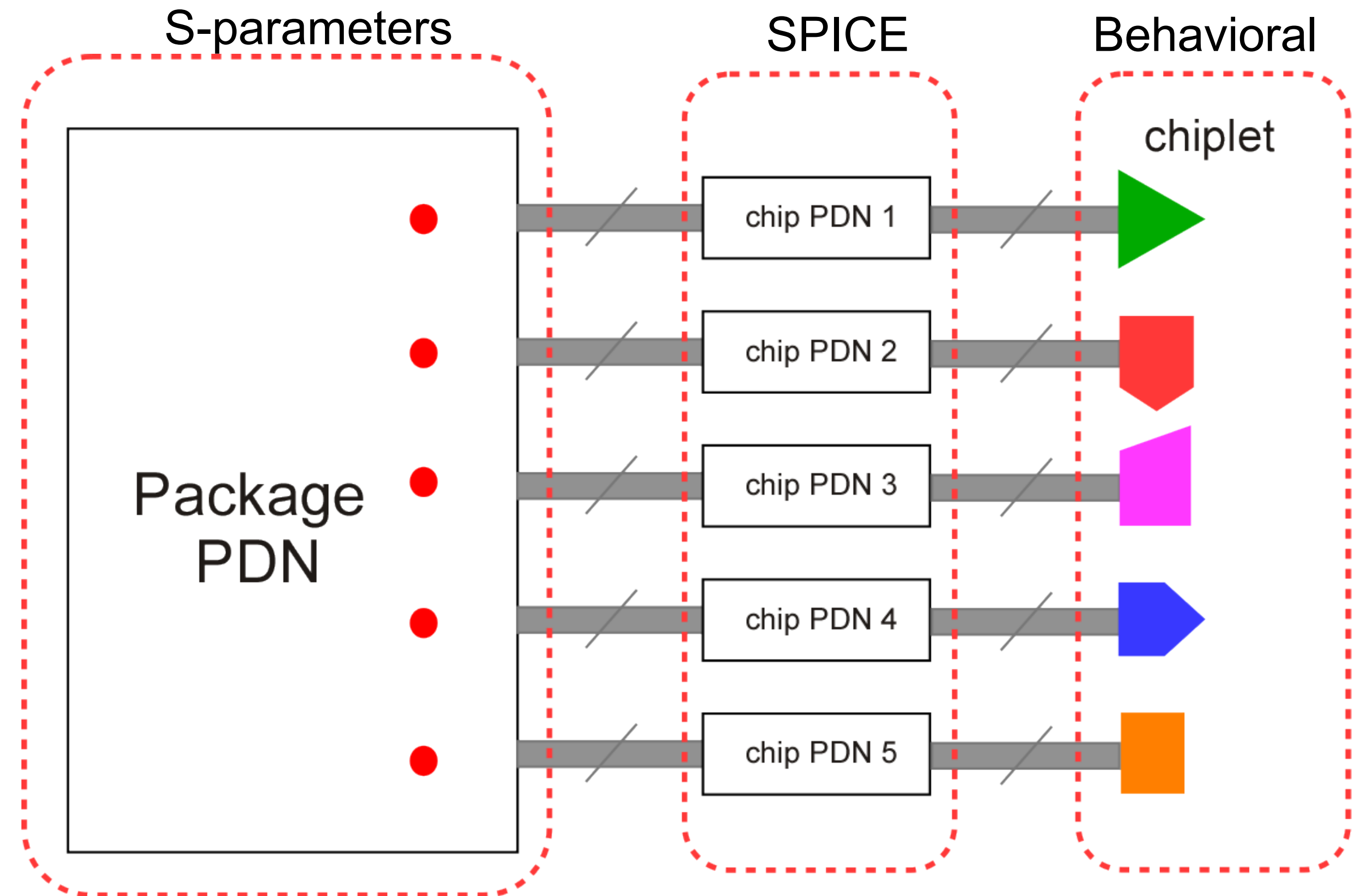


Chiplet PDN

Update configuration:

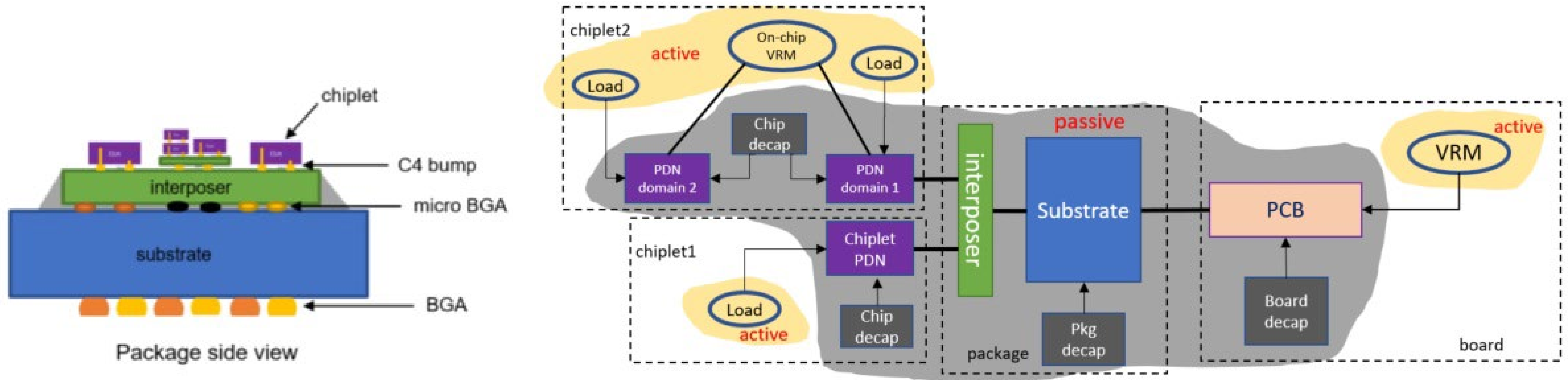
- Package design (geometry, bumps placement etc.)
- Power domain allocation
- Chiplet placements
- Decoupling
- etc.





Power Delivery for Chiplet-Based SiP

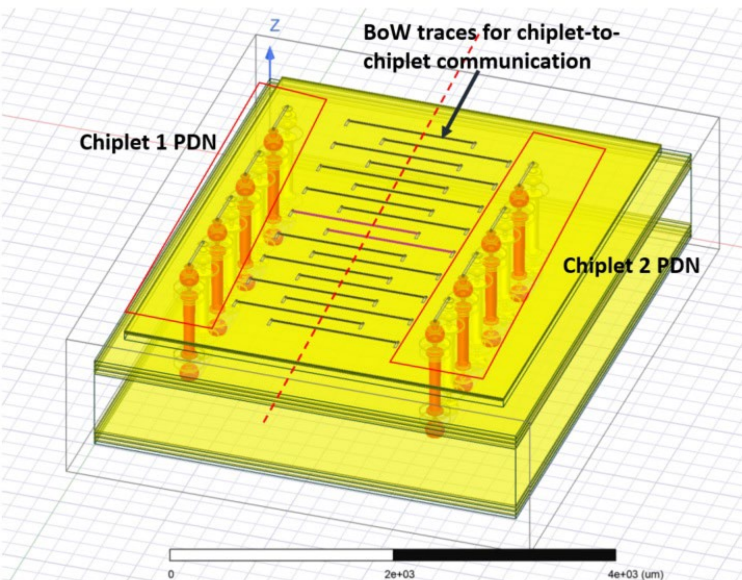
Chiplets must be positioned in their respective locations and provided with paths through the power distribution network. There can be multiple power domains. Package builder tools are used to automate the placement of vias, ball grid arrays and traces to the PDN. To mitigate supply voltage fluctuations decoupling capacitors ("decaps") are placed on the PCB, package substrate, package interposer, and the silicon dies.



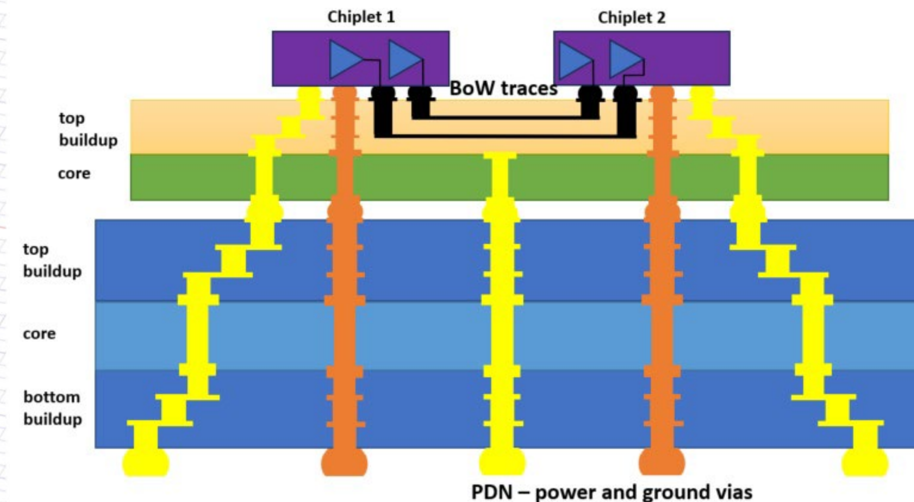
Bump Mapping for Chiplets

Many problems take too long to simulate and thus make it to difficult to predict the "what if" scenarios. This slows down the design process. This gets exacerbated when handling multiphysics situations (e.g. electrical + thermal) One approach is to first start with simplest possible models that will capture some physics and produce meaningful outputs

Top View



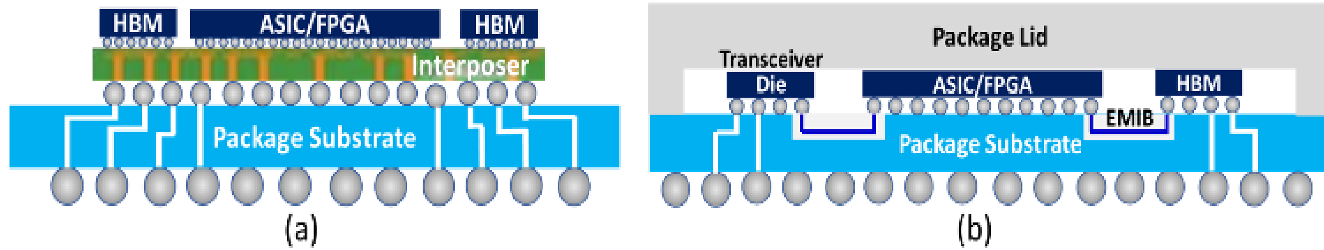
Cross-Section

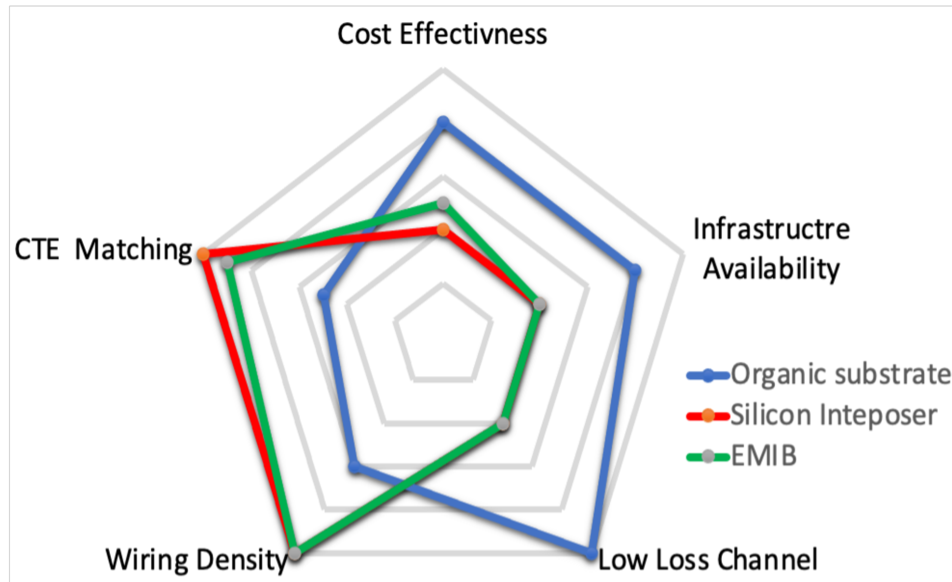


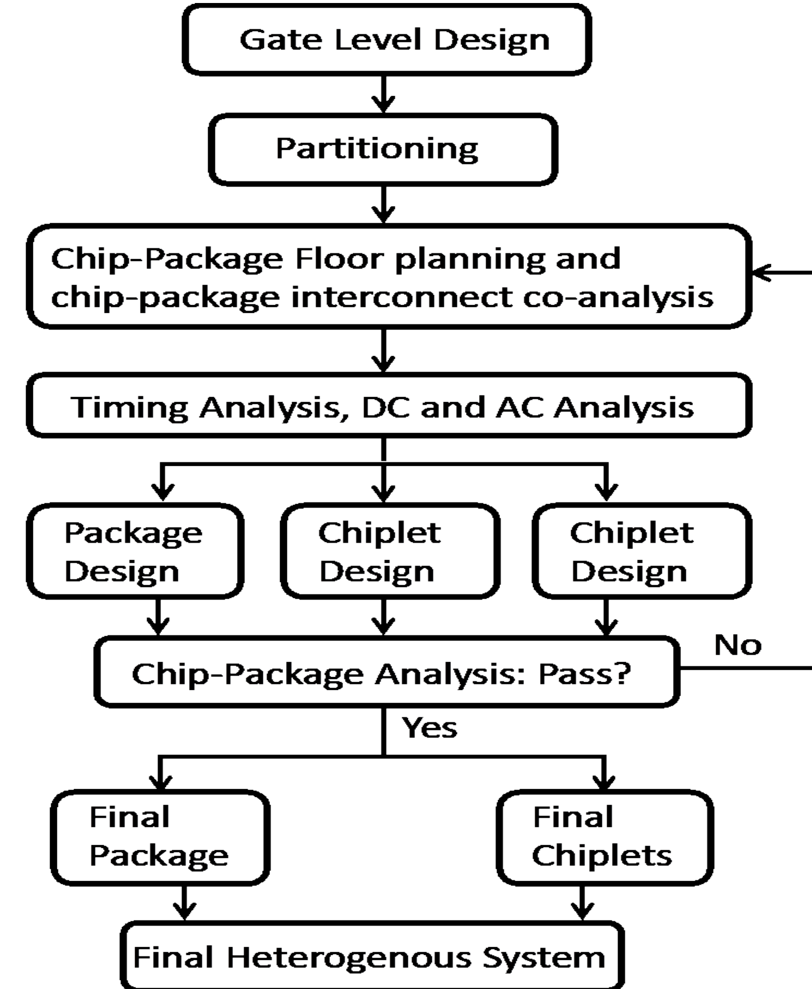
Chiplet-based design with PDN and D2D network

For 50 X 50 μBGA structure, extraction of network parameters takes more than a week!

Codesign for Chiplets



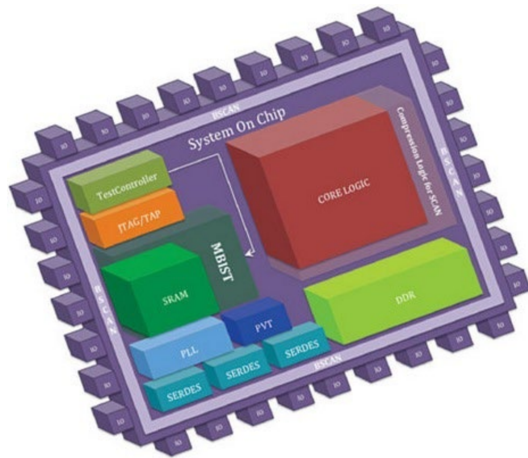




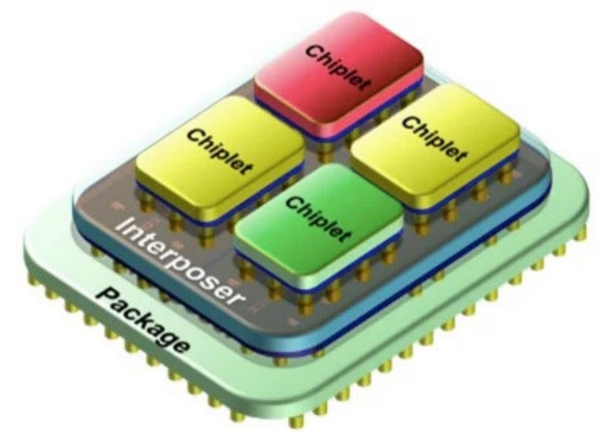
The Problem of Disaggregation

Architecting an IC as a chiplet-based SiP rather than a SoC is referred to as disaggregation of function. Today, it is performed *ad hoc*; there is no established methodology to optimize the disaggregation, i.e., to determine how many separate chiplets should be used in order to meet specifications. Once chiplet design is democratized, there will be more choices from different vendors which could make the process more chaotic.

SOC

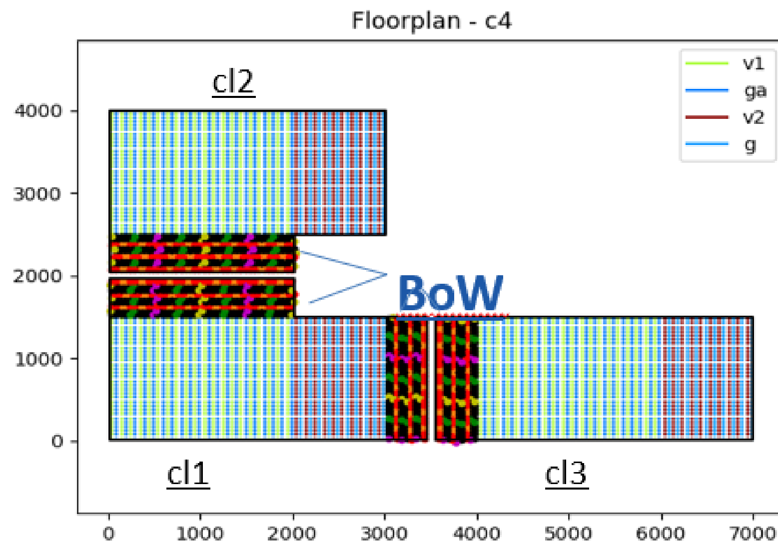


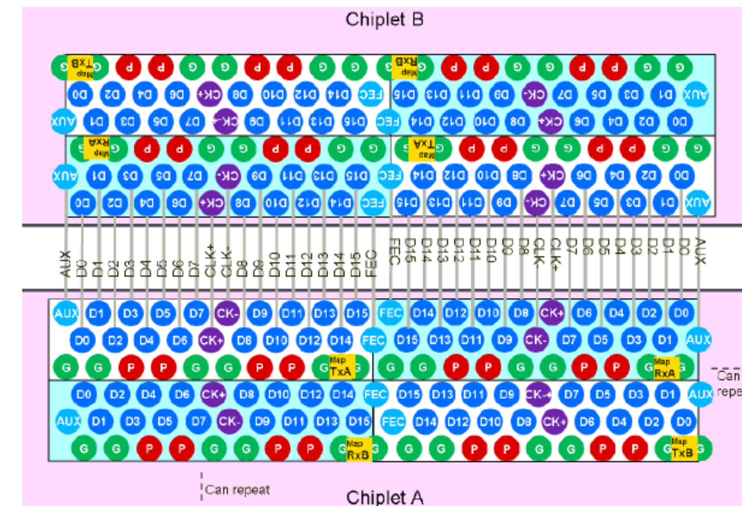
SiP



Automating Disaggregation

- Implement co-design into EDA tools
- Assist with AI/ML
- Integrate interconnection standards (e.g. UciE, BoW)





[1] Shahab Ardalan, Ramin Farjadrad,Mark Kuemerle,Ken Poulton,Suresh Subramaniam, Bapiraju Vinnakota, "Chiplet Communication Link: Bunch of Wires (BoW)", IEEE Micro, Jan/Feb 2021

Open-Source Opportunities

Packaging Benchmarks

UCIe

ODSA/BoW

IBIS

OpenROAD

Packaging Benchmark Suite

Welcome to the IBIS Open Forum

NEW Touchstone 2.1 Specification is approved and available for download

NEW IBIS 7.2.1 Parser IBISCHK7.2.1 is now available: IBISCHK7

NEW IBIS 7.2 Specification approved and available for download.

NEW IBIS Quality Specification 3.0 approved and available for download.

Barriers to HI Co-Design

- Technical Challenges
  - ➢problem size and complexity → computational bottleneck
  - ➢design tools need more intelligent interfaces
- Socio-Academic Challenges
  - ➢- lack of information sharing within entities (data)
  - ➢- lack of common standardized formats
  - ➢- lack of access to low-volume manufacturing for validation
  - ➢- lack of access to design tools for exploration purposes
  - ➢- lack of multidisciplinary curriculum in advanced packaging

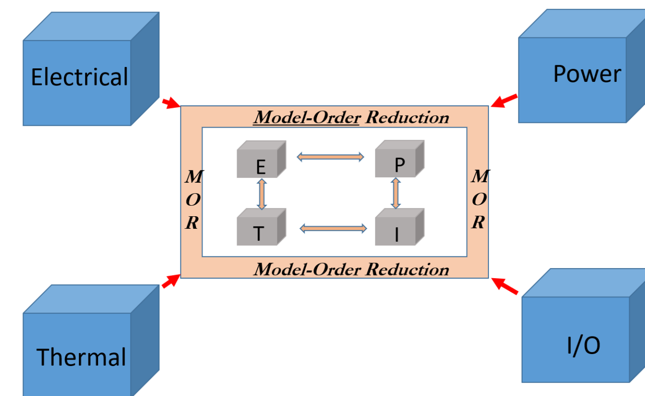
Possible Solutions for Co-Design

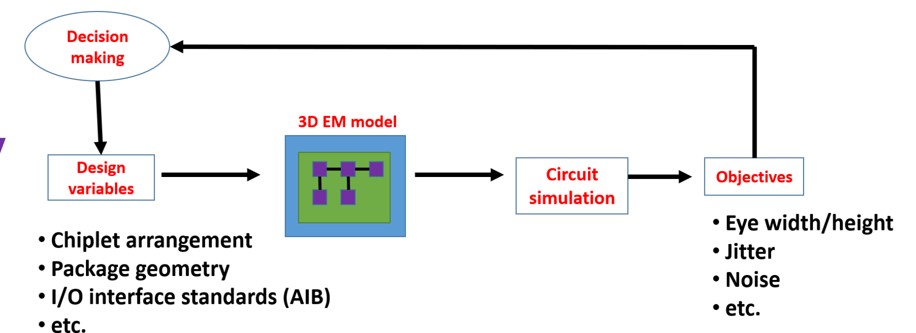
- Quantum Computing
  - Several decades away
  - Solutions exist for specialized problems
- Artificial Intelligence (AI) or Machine Learning (ML)
  - Data must be available
  - Training can be very tedious

➔ ***Need FASTER Electromagnetic Solvers*** **(>10X)**

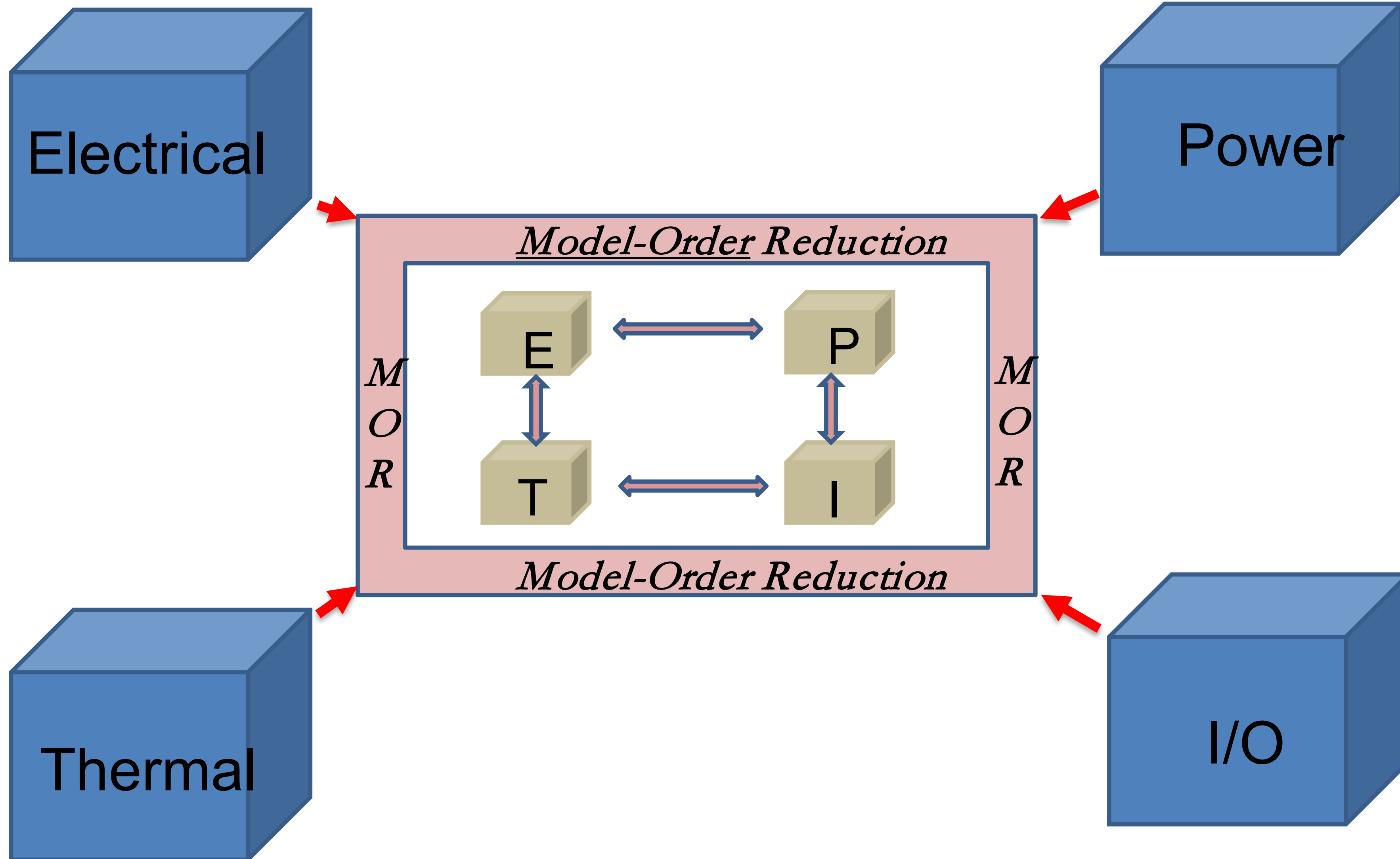
Research Needs

- Methodical abstractization
  - Compact models
  - Reduced-order models
  - Behavioral models
- Faster verification platforms (>10X
  - Multi-physics solvers (EM , thermal management, materials)
  - Transistor-level circuit simulation
- AI/ML assisted solutions
  - Digital twins
  - Optimization, mitigation of uncertainty

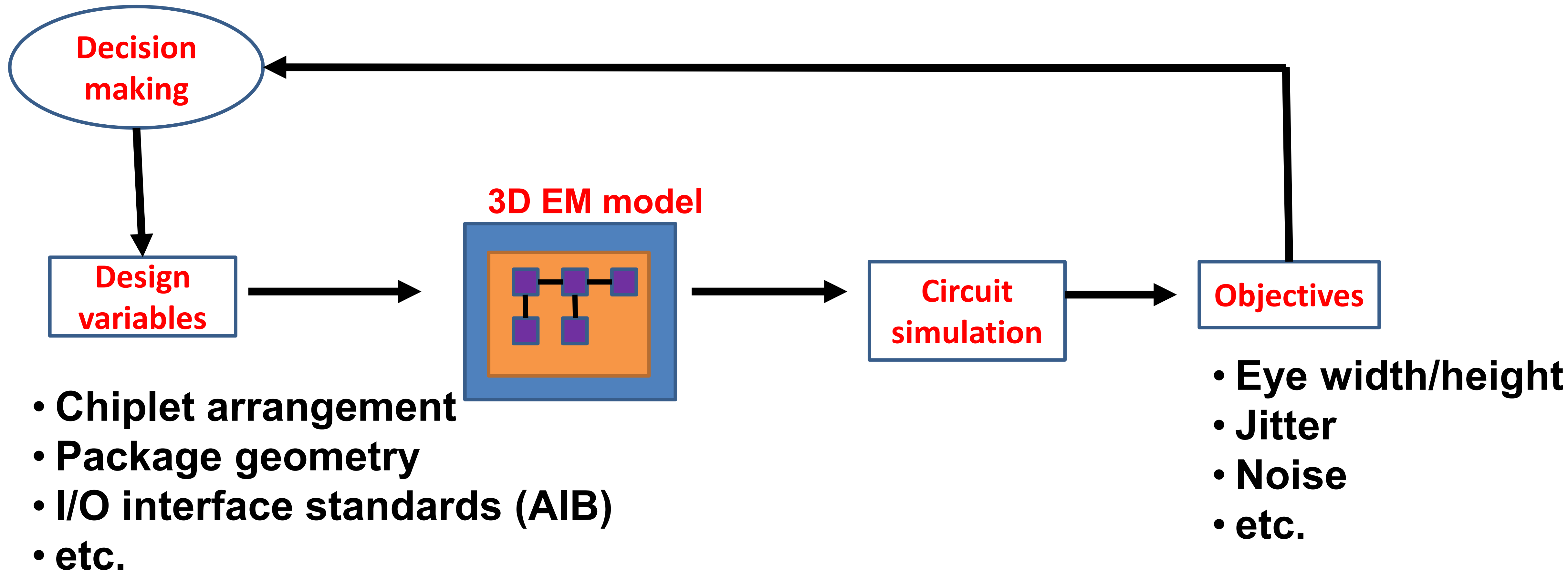




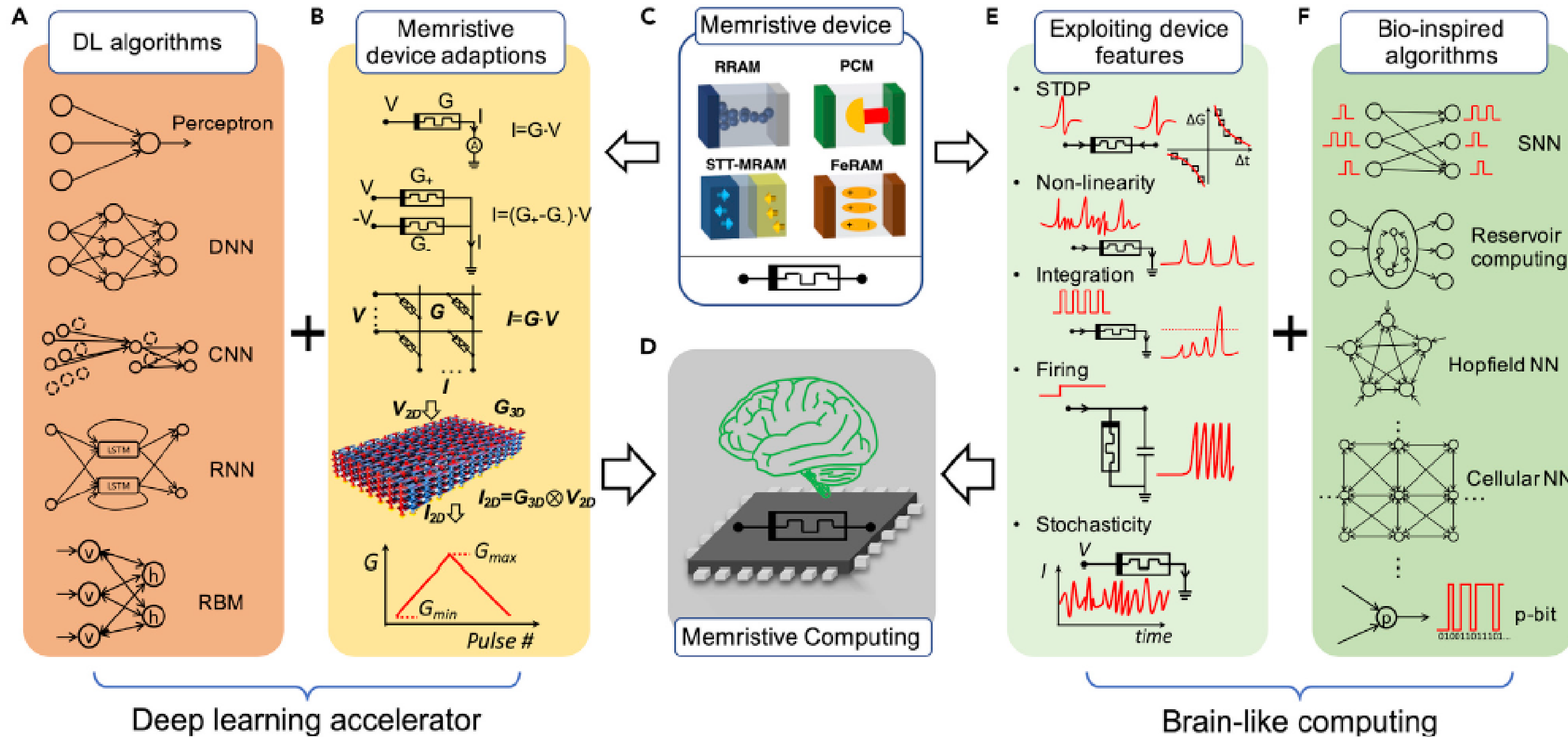
Multiphysics & MOR



AI/ML for System-Level



Driver: Brain-Inspired AI Chiplets



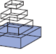
SOURCE: Wei Wang, Wenhao Song, Peng Yao, Yang Li, Joseph Van Nostrand, Qinru Qiu, Daniele Ielmini and J. Joshua Yang, "Integration and Co-design of Memristive Devices and Algorithms for Artificial Intelligence", *iScience 23, 101809*, December 18, 2020

Neurobiology-Driven Algorithms

- Similarities with biological synapses
  - Stochastic
  - pulse-pair facilitation
  - short-term plasticity
- Approach
  - Modify devices and materials to emulate algorithms
  - Modify algorithms to exploit special properties of memristive devices

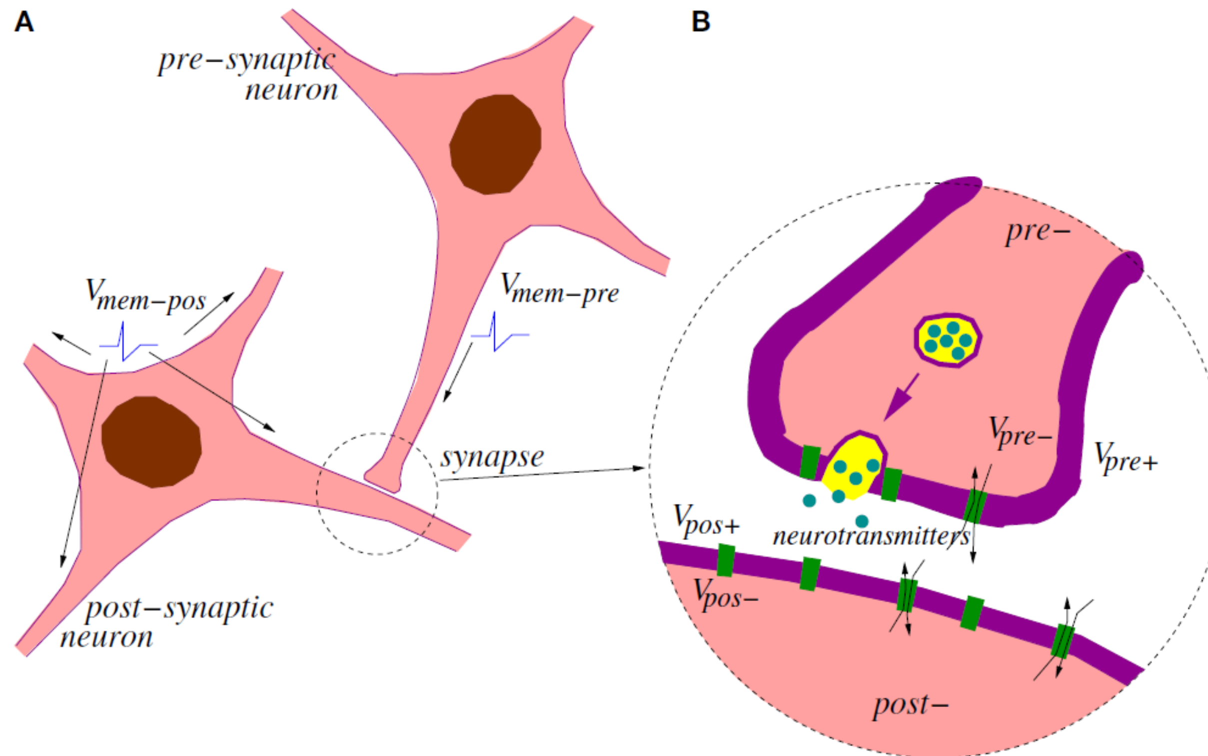
STDP-Based Memristor

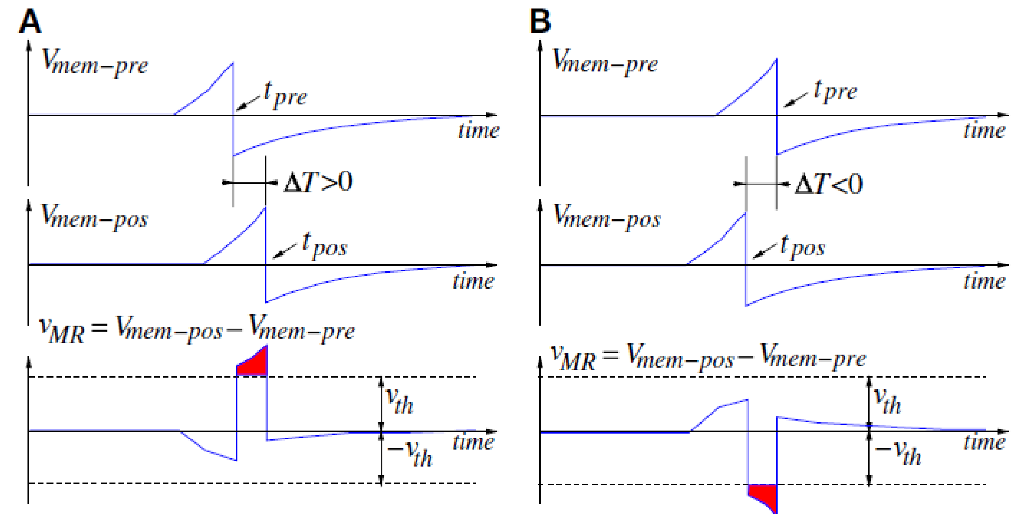
frontiers in
NEUROSCIENCE
ORIGINAL RESEARCH ARTICLE
published: 17 March 2011
doi: 10.3389/fnins.2011.00026



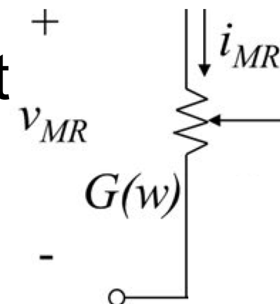
On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex

*Carlos Zamarreño-Ramos[1], Luis A. Camuñas-Mesa[1], Jose A. Pérez-Carrasco[1], Timothée Masquelier[2], Teresa Serrano-Gotarredona[1] and Bernabé Linares-Barranco[1]**





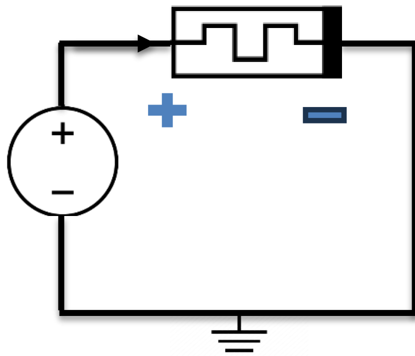
Equivalent Circuit



$$\Delta G\left(v_{\mathrm{MR}}\right)=\begin{cases}G_0\left[e^{v_{\mathrm{MR}}/v_o^+}-e^{v_{\mathrm{th}}^+/v_o^+}\right], & v_{\mathrm{MR}}>v_{\mathrm{th}}^+\\ -G_0\left[e^{-v_{\mathrm{MR}}/v_o^-}-e^{-v_{\mathrm{th}}^-/v_o^-}\right], & v_{\mathrm{MR}}<v_{\mathrm{th}}^-\\ 0, & otherwise\end{cases}$$

Memristor Model: Bipolar memristor with threshold

Reliable SPICE Simulations of Memristors, Memcapacitors and Meminductors, 2013



a. Behavioral Model is used
b. Old and New Parameters:

i. $V_{th} = 4.6V \rightarrow 1.6V$
ii. $R_{on} = 1k$
iii. $R_{off} = 10k$
iv. $R_{init} = 5k \rightarrow 3K$
v. $\beta = gain = 5e12$

Ix(mem1:PLUS)

Current Vs Voltage

1000µA 800µA 600µA 400µA 200µA 0µA -200µA -400µA -600µA -800µA -1000µA

-2.5V -2.0V -1.5V -1.0V -0.5V 0.0V 0.5V 1.0V 1.5V 2.0V 2.5V

G_{on} G_{off} G_{init}

Impact of β

V(1)

V(1) V_{in}

2.4V 1.6V 0.8V 0.0V -0.8V -1.6V -2.4V

Ix(mem1:PLUS)/V(1) G

600µΩ-1 540µΩ-1 480µΩ-1 420µΩ-1 360µΩ-1 300µΩ-1 240µΩ-1 180µΩ-1 120µΩ-1 60µΩ-1 0µΩ-1

0ns 6ns 12ns 18ns 24ns 30ns 36ns 42ns 48ns 54ns 60ns

$G_{on} = \frac{1}{R_{on}}$

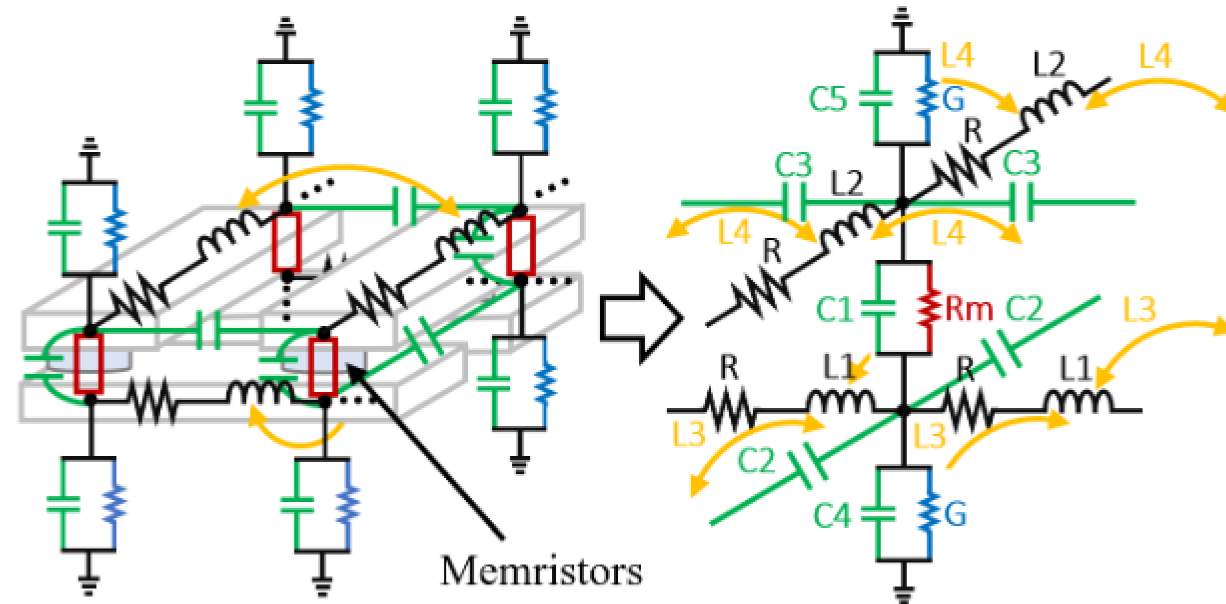
$G_{init} = \frac{1}{R_{init}}$

$G_{off} = \frac{1}{R_{off}}$

$g = \frac{I_{xmem}}{V_{in}}$

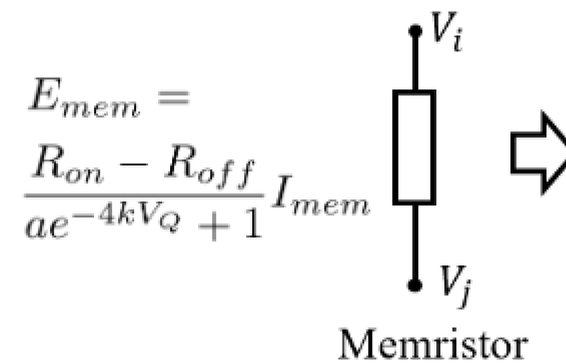
Memristor

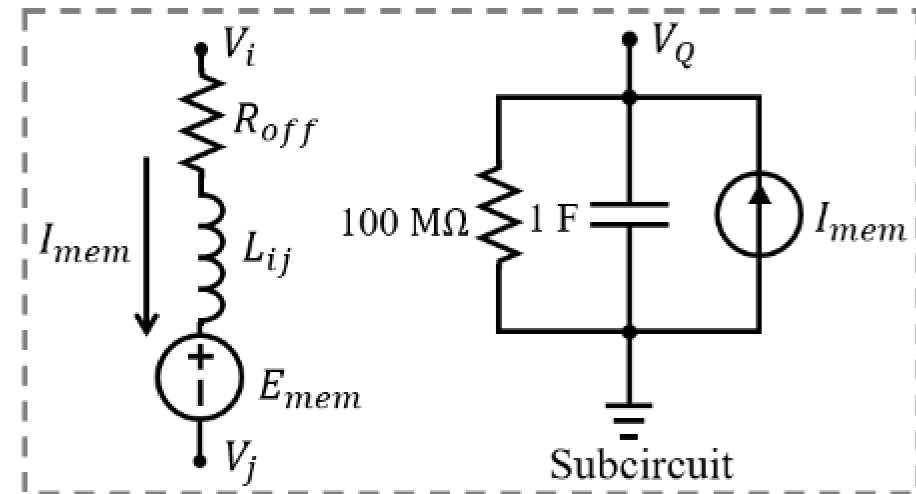
Crossbar array:



Memristors:

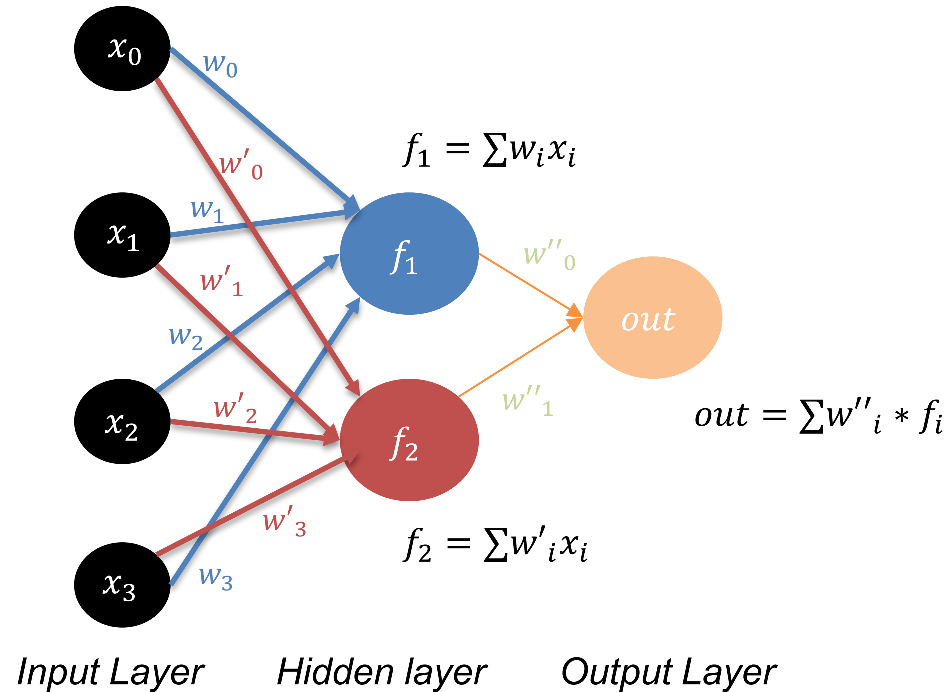
$$E_{mem} = \frac{R_{on} - R_{off}}{ae^{-4kV_Q} + 1} I_{mem}$$



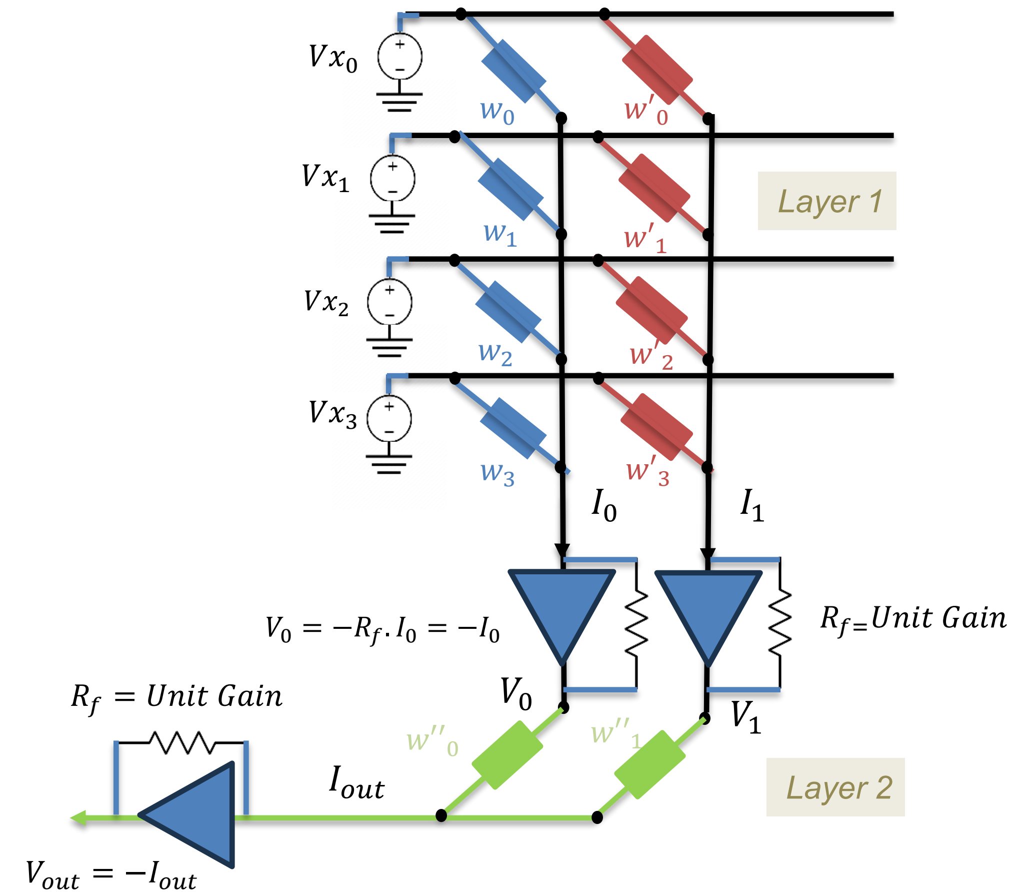


Fully Connected Layer for NN

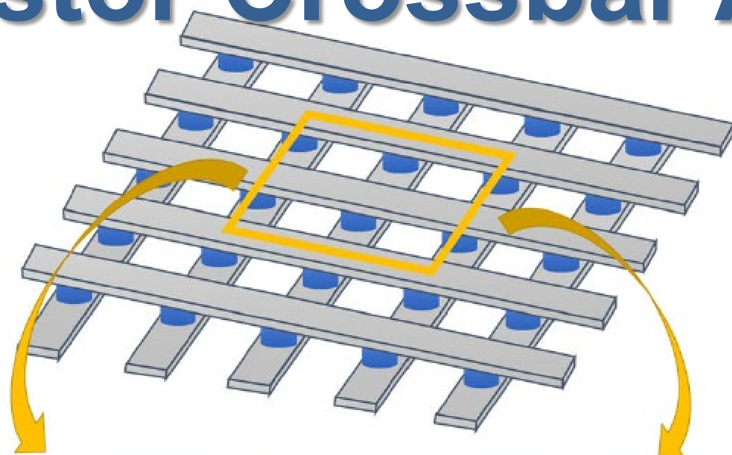
Linear Neural Network (Without any activation Function)

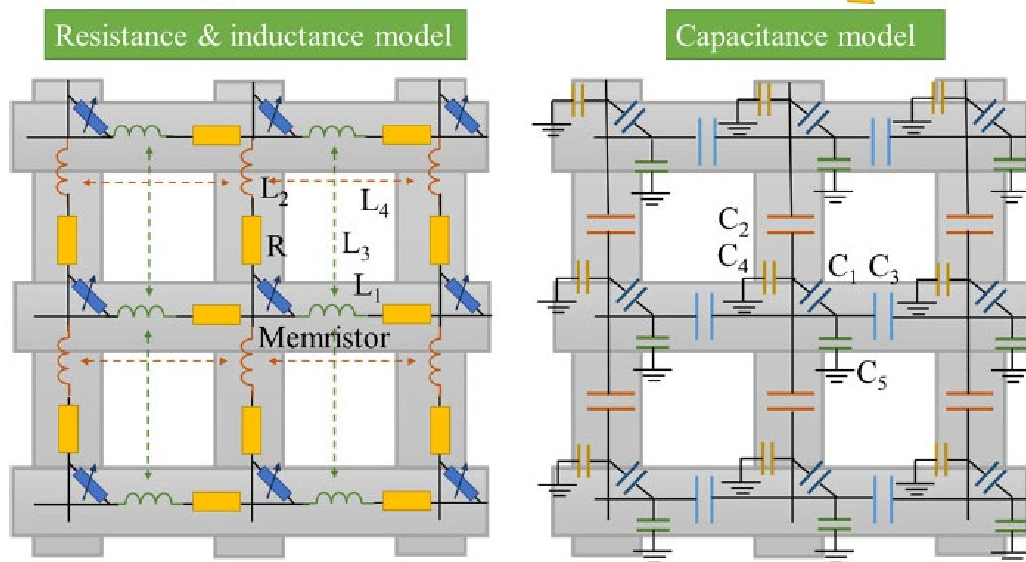


Conductance Mapping of NN in memristor array

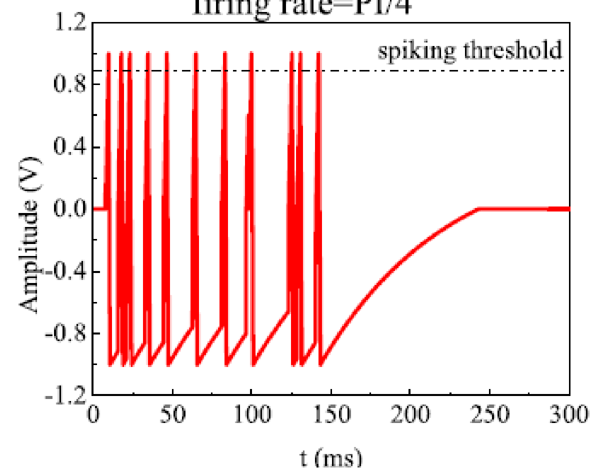


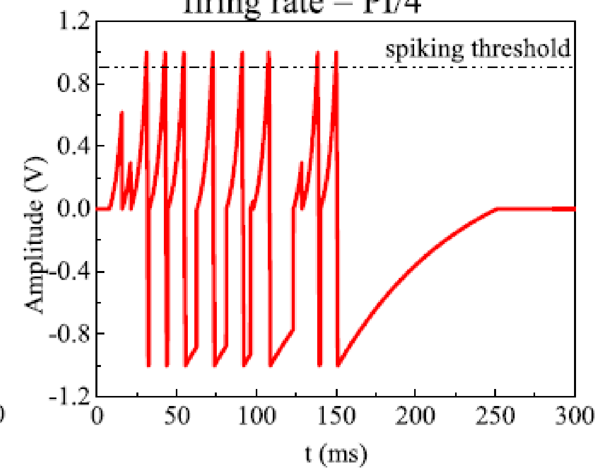
Memristor Crossbar Array





Will perform VMMA in analog domain





Conclusion

- Chiplets: HI solution for future SiP
- Can only be addressed through co-design
- Ecosystem still in nascent stage
- Size, complexity and stochasticity are increasing
- Brain-inspired computing will be a driver
- Great opportunities for MOR