Thermal and Performance Modeling and Optimization for Chiplet-Based Systems

Lukas Pfromm

Preliminary Examination
20 Apr2026

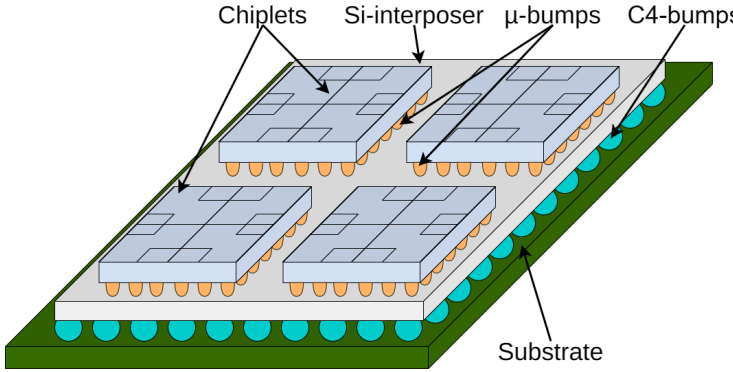
Committee Members: Joshua San Miguel, Janardhan Rao Doppa,
Eric Tervo (Ph.D. Co-Advisor) & Umit Y. Ogras (Ph.D. Advisor)

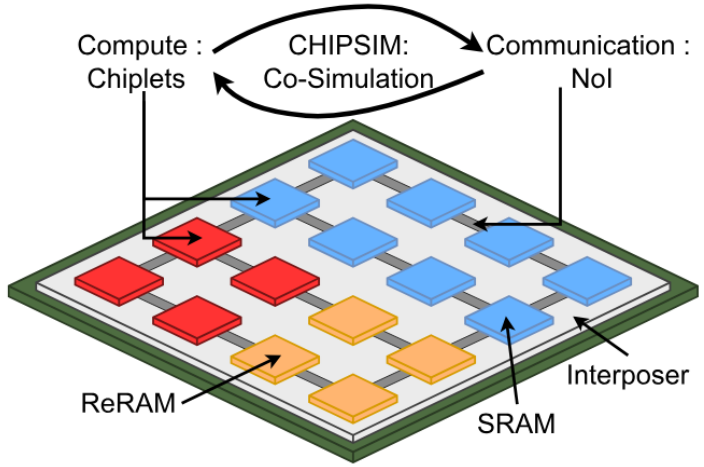
e lab

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Outline

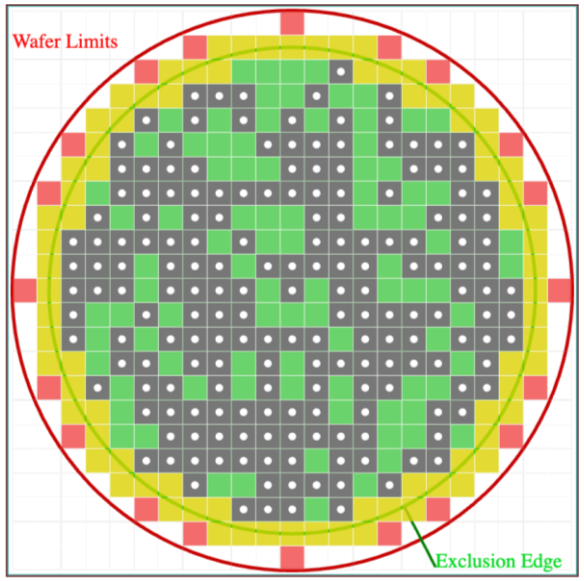
- **Motivation: Chiplet-based platforms**
- **Preliminary Work-1:**
  - CHIPSIM: A Co-Simulation Framework for Deep Learning on Chiplet-Based Systems
- **Preliminary Work-2:**
  - MFIT : Multi-Fidelity Thermal Modeling for 2.5D and 3D Multi-Chiplet Architectures
- **Ongoing and Proposed Work:**
  - Exploration of Thermoelectric Coolers (TECs) for Chiplet-Based Systems
  - TEC Control for Chiplet-Based Systems
  - Tutorial on the State of Thermal Research for Chiplet-Based Systems
- **Timeline**
- **Conclusions**

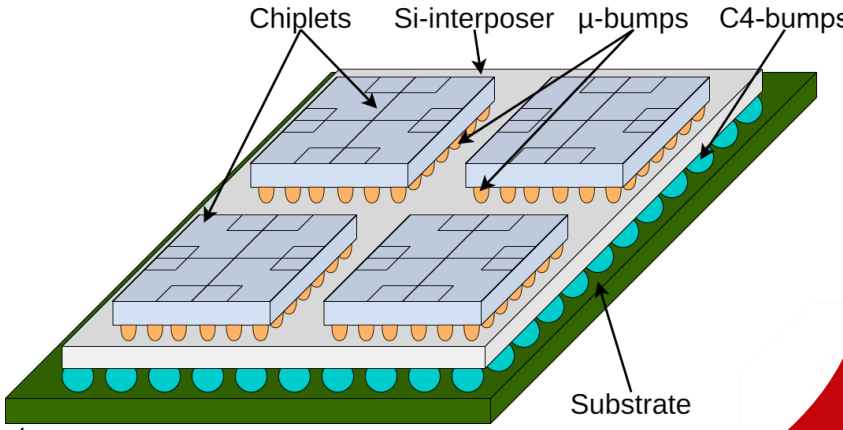




How did Chiplets Become Crucial?

- **Monolithic 2D fabrication technologies reach physical limits**
  - Manufacturing of a large die is not economical anymore
- **However, the need for more compute power keeps growing**
  - Ever-increasing processing and memory requirements of AI workloads
- **A promising avenue: heterogeneous integration and chiplet-based manycore systems**
  - Integrating many small dies on interposer to sustain performance growth
  - Higher yield and lower fabrication cost





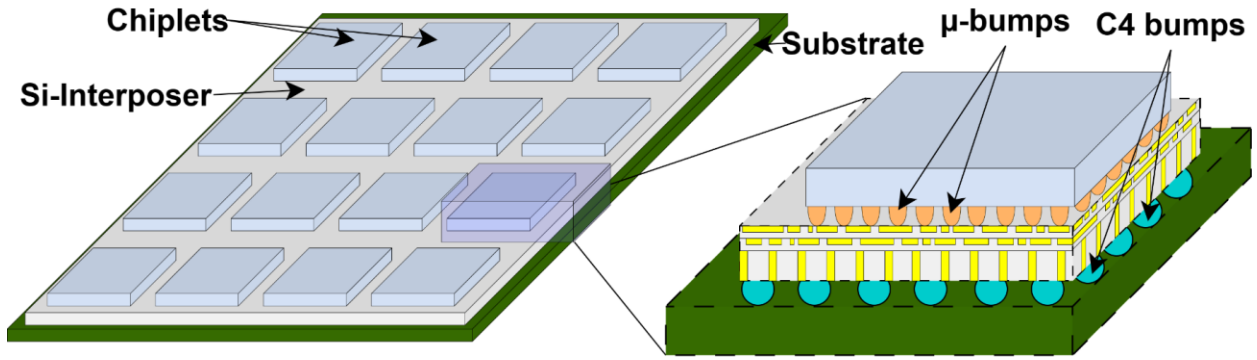
J. A. Cunningham et al., "ITRS2.0:Toward a reframing of the semiconductor technology roadmap.," in *IEEE 32nd Intl. Conf. Computer Design*, Oct 2014

Chiplet-Based Platforms

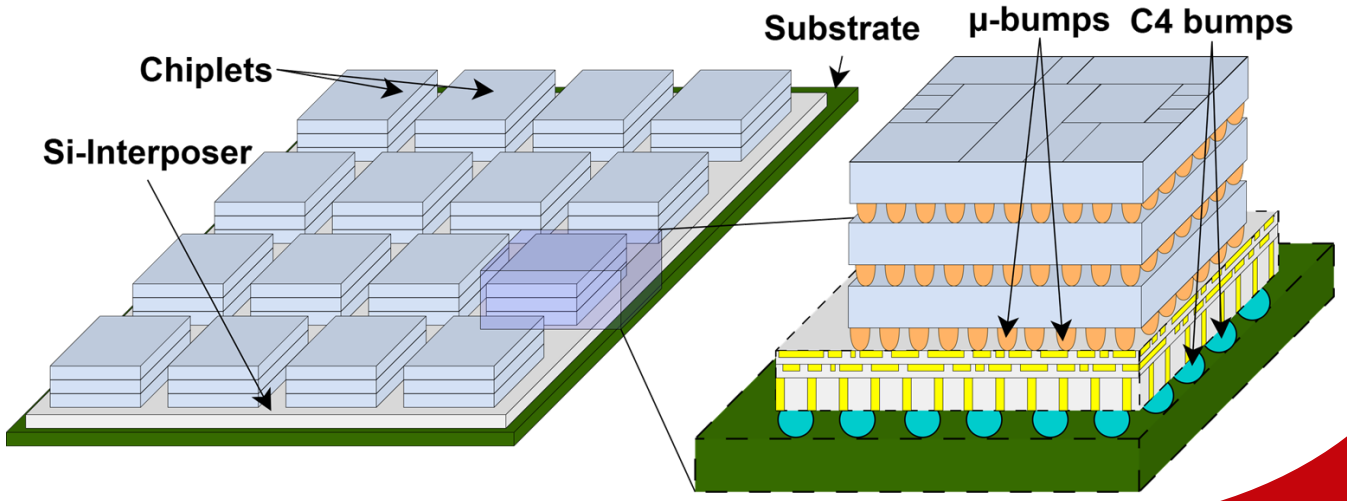
- Small dies fabricated on a single wafer
- Connected through Network-on-Interposer (NoI)
- Compact scale-out implementations
- Economical manufacturing

- Unsolved challenges we address:
  - ➢ *Performance Modeling*
  - ➢ *Thermal Issues*

2.5D Geometry

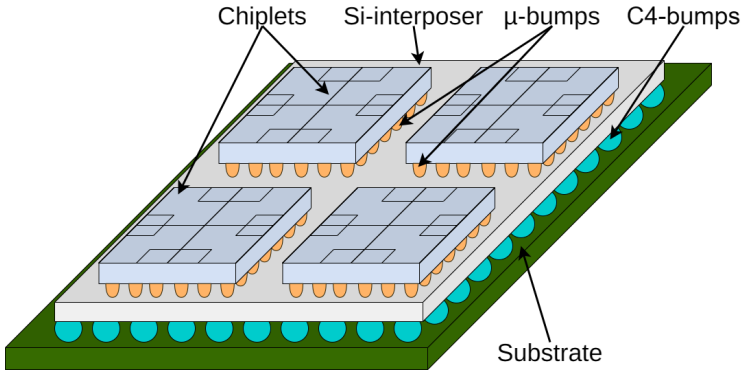


3D Geometry

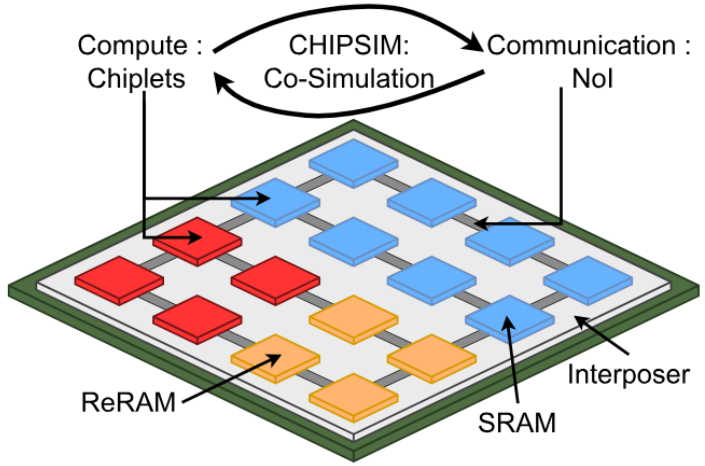


Outline

- Motivation: Chiplet-based platforms
- **Preliminary Work-1:**
  - CHIPSIM: A Co-Simulation Framework for Deep Learning on Chiplet-Based Systems

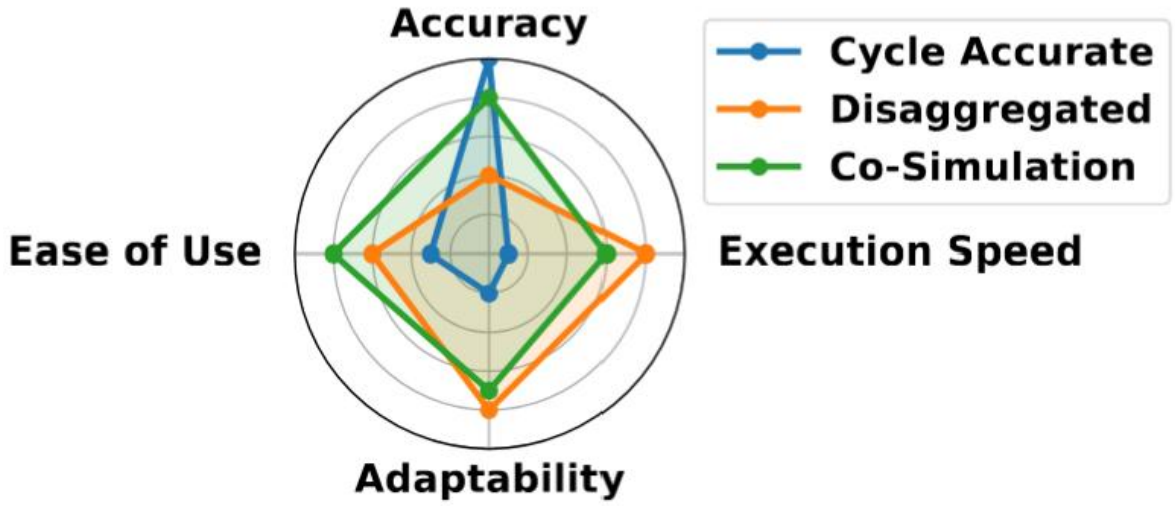


Pfromm, Lukas, Alish Kanani, Harsh Sharma, Janardhan Rao Doppa, Partha Pratim Pande, and Umit Y. Ogras. "CHIPSIM: A Co-Simulation Framework for Deep Learning on Chiplet-Based Systems."
IEEE Open Journal of the Solid-State Circuits Society (2025).



Chiplet-Based System Evaluation

- **Chiplet-Based systems required fast and accurate evaluation methods**
  - Highlighted by growing popularity of chiplet-based systems and recent federal programs (NAPMP)
- **Current methods**
  - Accurate but slow (infeasible for exploration)
  - Or fast and inaccurate (decouple processing from communication and do not support modeling of multiple models executing in parallel in the same system)
  - Incapable of modeling time-based power and thermal behavior
- **We propose a co-simulation framework which is fast, flexible, and accurate.**



Simulation Requirements

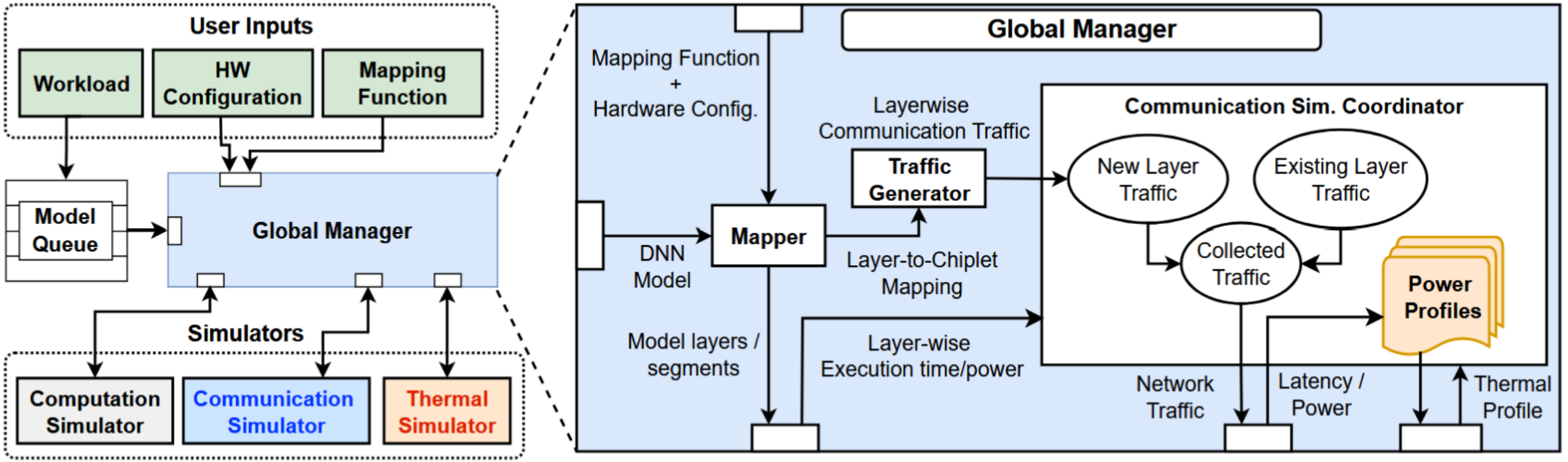
1. Target chiplet-based system executing multiple DNN models in parallel
2. The simulator must be able to model
   - Custom interconnect topologies,
   - Custom system sizes,
   - Integration techniques such as 3D stacking,
   - Heterogeneous chiplet configurations
3. The simulator must report
   - Execution time of each model
   - Power consumption over time
   - Thermal behavior of the system over time

Key Contributions

- **Novelty:**
  - Co-simulation of compute and communication for chiplet-based systems executing DNN models.
  - Explicit modeling of contention and pipelining for each individual model.
  - Capturing the notion of time, allowing for transient power profiles.
- **Challenges addressed:**
  - Existing approaches are either too slow or too inaccurate.
  - Decoupled compute and communication simulation, as is common in existing simulation approaches, is fundamentally inaccurate.

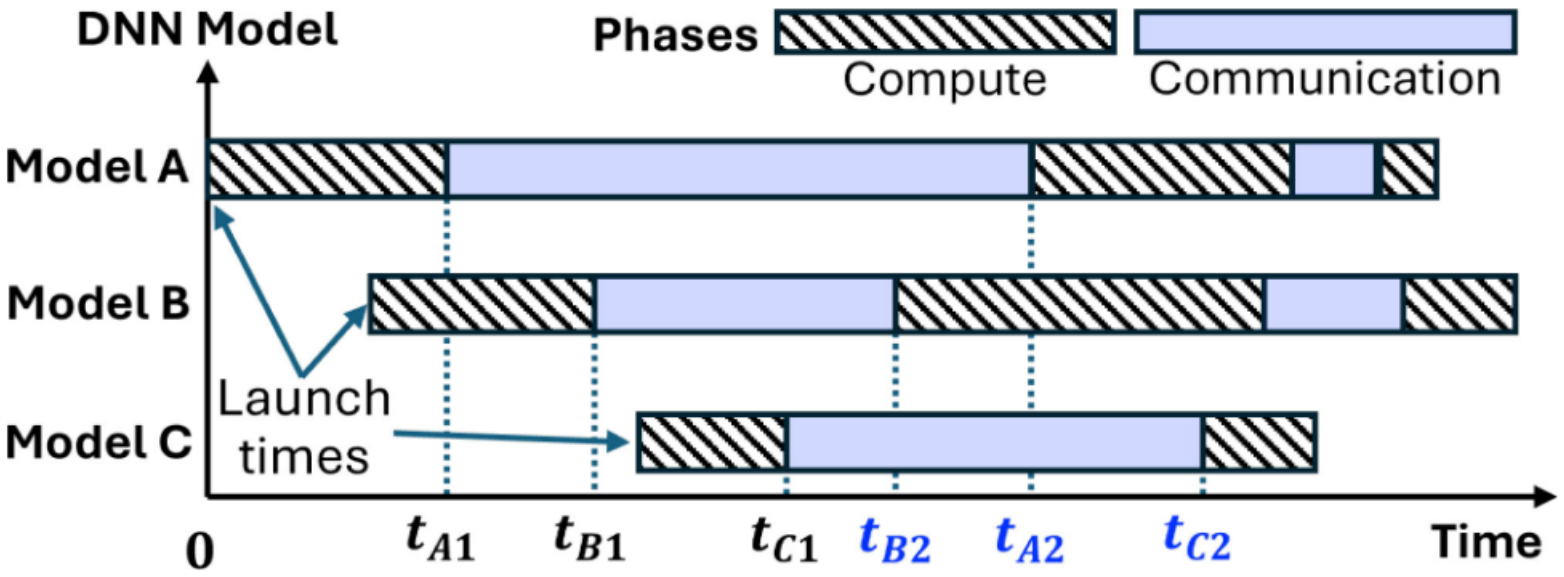
Chiplet System Co-Simulation

- Our approach coordinates the execution of processing and communication simulation
- Critically, our approach considers all active traffic in the system, not just the traffic from a single network. **Existing ad-hoc approaches do not meet this need.**



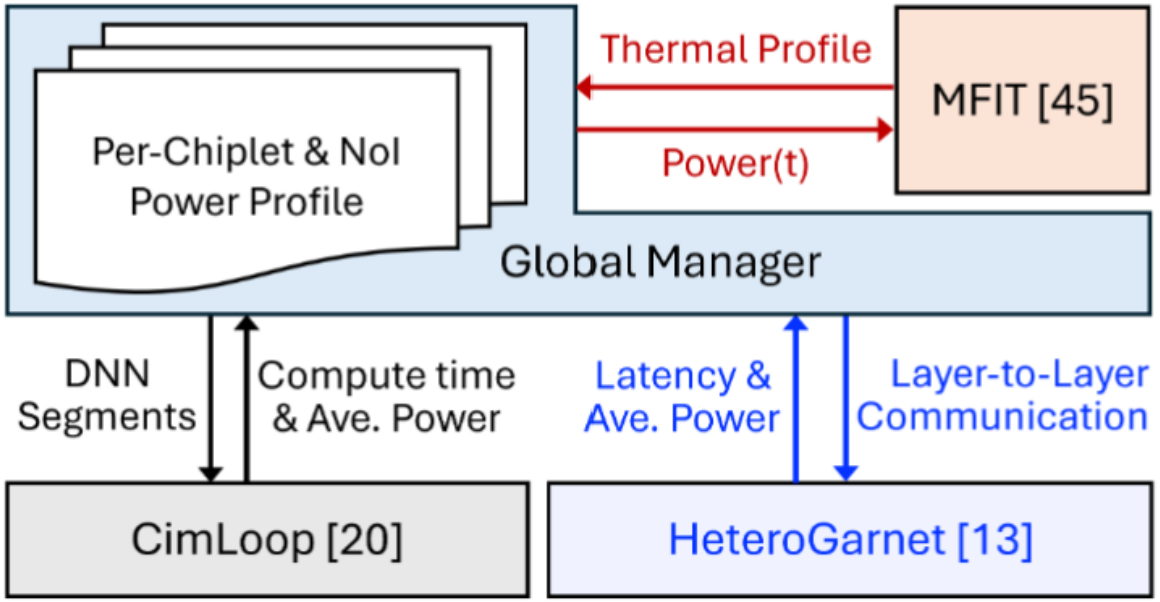
Model Interleaving

- The most critical functionality of CHIPSIM is tracking the state of the system over time, and performing simulation operations based on that state
- Below is a simple example with 3 DNN models executing in parallel:



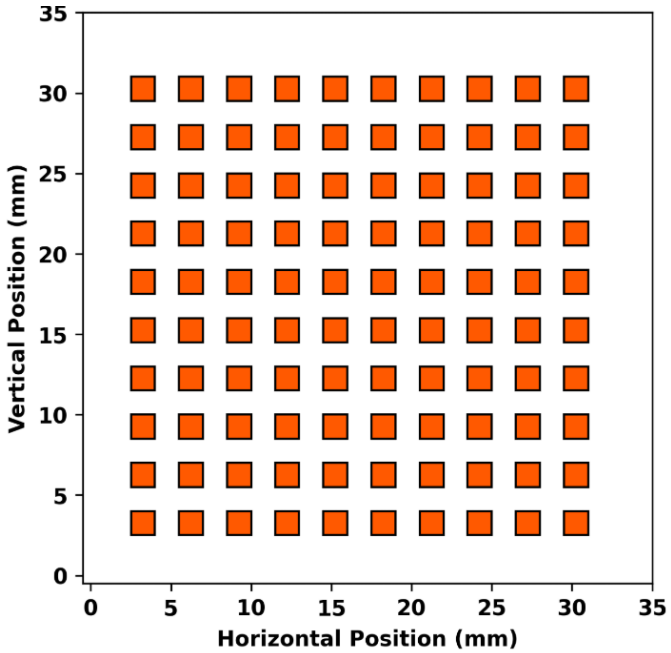
Our implementation of the framework

- To demonstrate the proposed framework, we create a concrete implementation
- Existing simulators are implanted to model individual processes (compute, communication, thermal behavior)
- CHIPSIM framework is general, any desired simulators can be implemented as required



Approach Performance Comparison

- We compare the results of running co-simulation with the existing baseline.
- 100 chiplet system executing 4 DNN models simulated (ResNet 18, 34, 50 + AlexNet)
- Baseline methods use same setup, but simulate each DNN model individually and do not co-simulate compute and communication (no parallel execution with other models)
- For pipelined experiments, number of inferences per model instance equates to how many layers of a DNN model can be active in parallel.



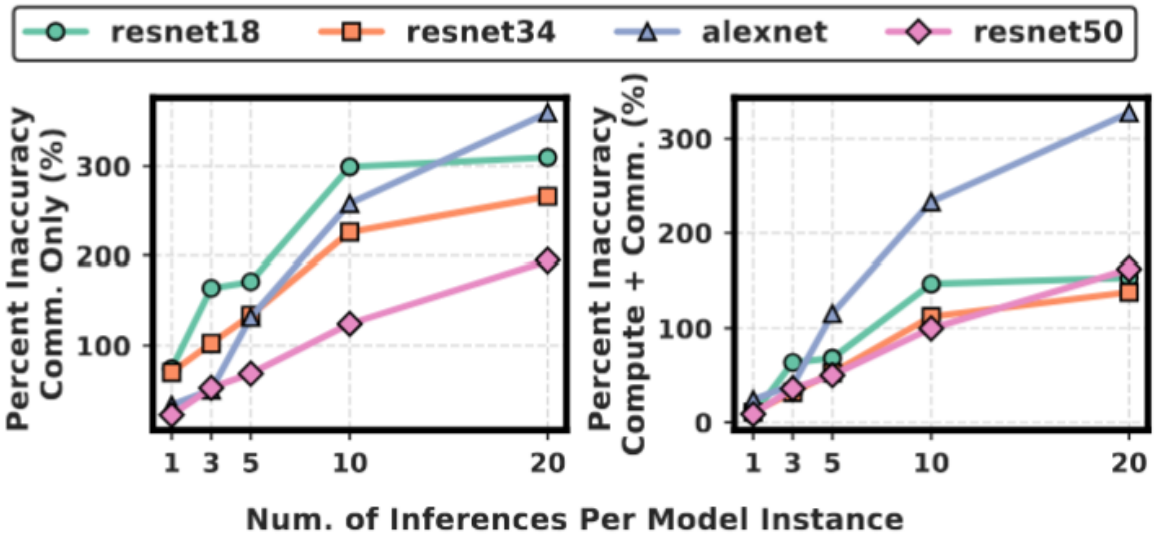
Results (CNN Models)

- Comm. Only baseline assumes no compute. This approach sometimes used for pure NoI evaluation.
- Comm. + Compute assumes both compute and communication modeled.
- While error is significant for non-pipelined execution, error increases dramatically when pipelined execution is implemented.

Non-Pipelined Execution

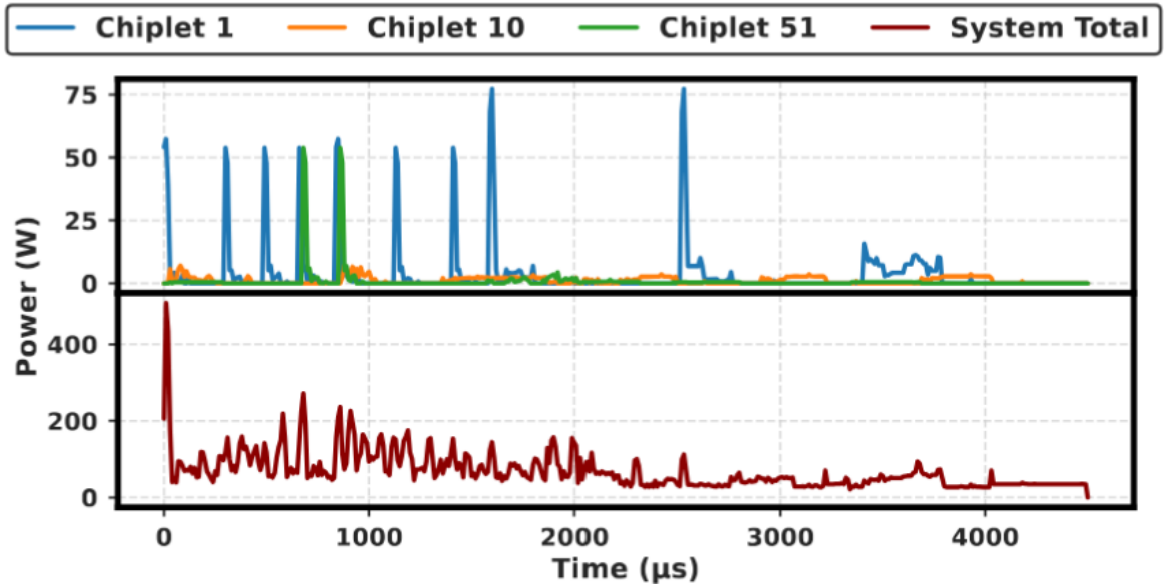
DNN Model	Percent Inaccuracy	
	Comm. Only	*Comm. + Compute*
ResNet18	74%	8%
ResNet34	70%	11%
ResNet50	22%	9%
AlexNet	33%	24%

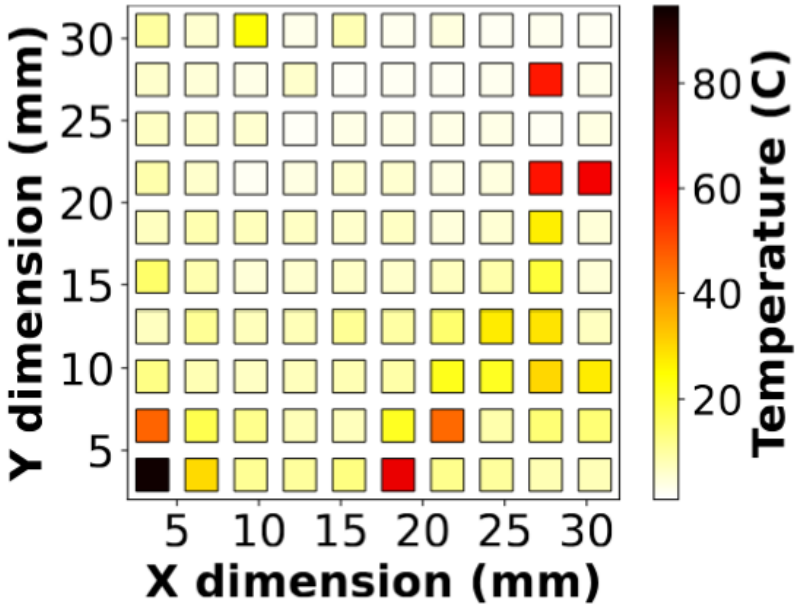
Pipelined Execution



Power and Thermal Analysis

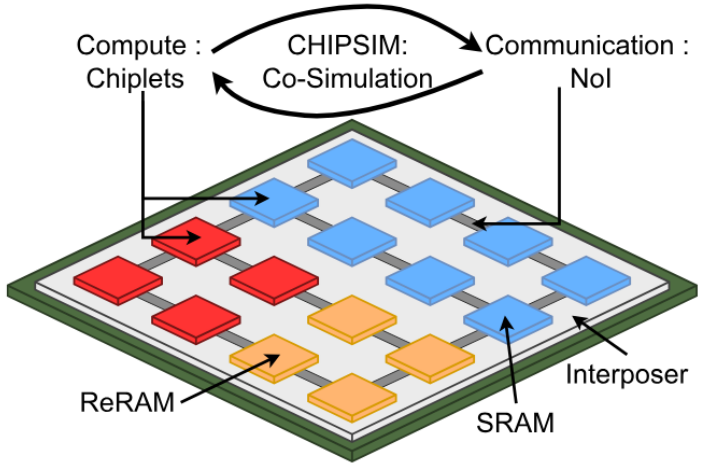
- Existing simulation approaches have no concept of time, so are unable to create power over time profiles.
- Our approach accurately models the power consumption of each chiplet over time on the order of microseconds
- Power then applied to thermal model to model temperature over time of each chiplet





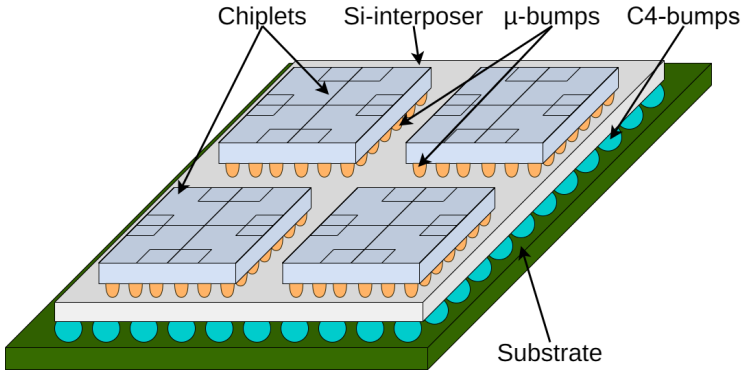
CHIPSIM Conclusions

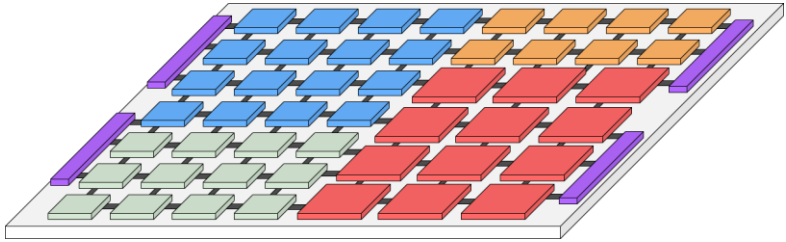
- Contention between models executing in parallel on chiplet-based systems is a major source of overhead in execution time.
- This overhead is not captured in existing simulation methods, compromising their accuracy.
- Additionally, by co-simulating compute and communication, CHIPSIM captures the notion of time absent in other simulators, enabling transient power profile generation.
- Power outputs of the simulator become inputs for the next work: MFIT



Outline

- **Motivation: Chiplet-based platforms**
- **Preliminary Work-1:**
  - CHIPSIM: A Co-Simulation Framework for Deep Learning on Chiplet-Based Systems
- **Preliminary Work-2:**
  - MFIT : Multi-Fidelity Thermal Modeling for 2.5D and 3D Multi-Chiplet Architectures





Heterogeneous chiplet system

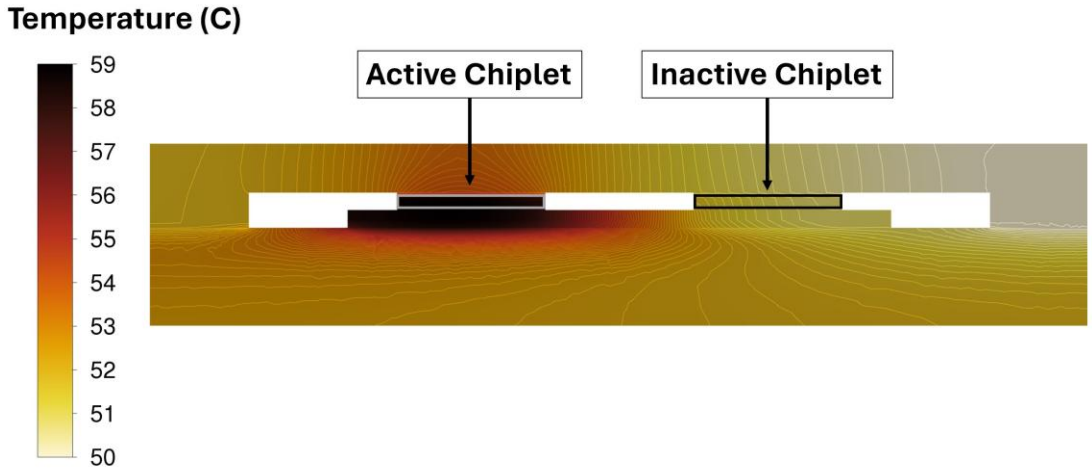
Pfromm, Lukas, Alish Kanani, et al. "MFIT: Multi-fidelity thermal modeling for 2.5 D and 3D multi-chiplet architectures."
ACM Transactions on Design Automation of Electronic Systems (2025).

Open source: Thermal tool in GitHub: https://github.com/AlishKanani/MFIT

19 citations in about one year

Unique Thermal Paths in 2.5D Platforms

- **Thermal bottlenecks have long been a significant barrier to performance**
- **Chiplet-based systems aggravate this barrier due to their dense integration and add unique challenges**
  - **Monolithic chip:** Heat is spread directly across the die
  - **Chiplet-based platforms:** Heat also flows through the interposer and heat spreader
- **Dense large-scale integration and new thermal conductance paths**
  - Lead to thermal hotpot even at moderate power consumption

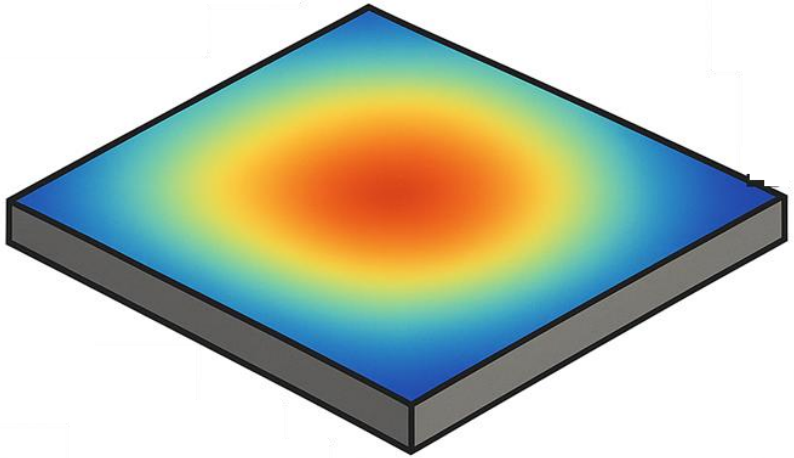
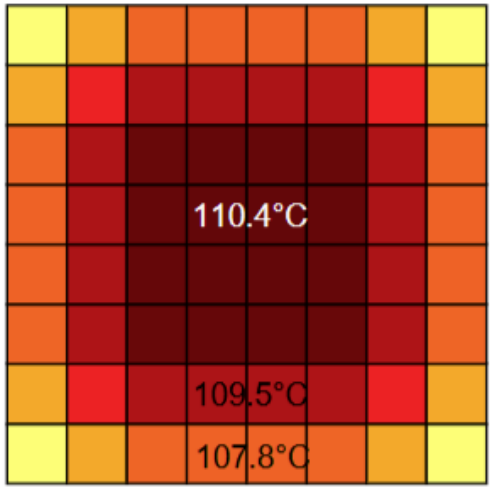
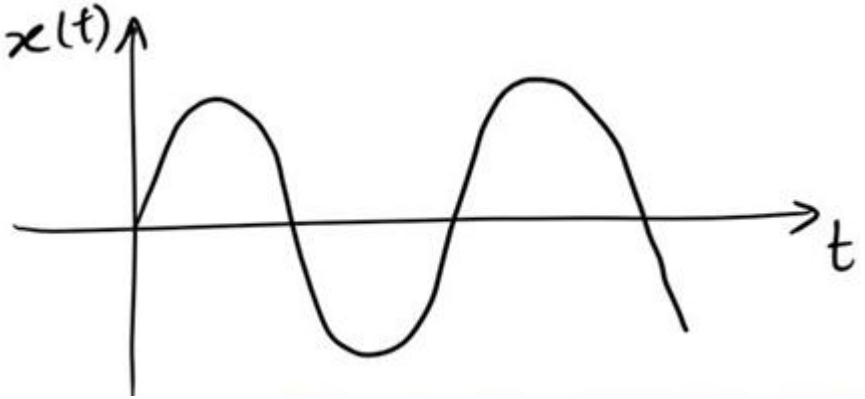
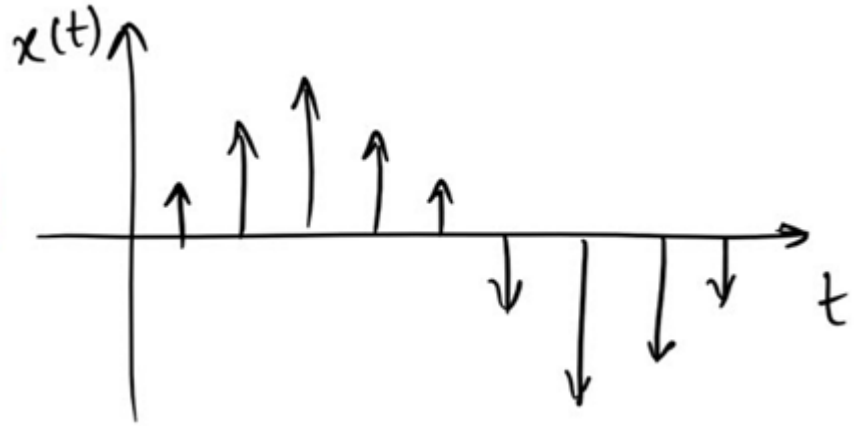


Need new approaches for this unique structure!

Overview of SOTA

- **The most accurate and direct thermal evaluation: hardware measurements**
  - Thermal imaging or temperature sensors
  - But requires hardware availability (infeasible for future systems)
- **This limitation motivated FEM-based modeling for temperature and analyzing heat flow**
  - Tools, such as ANSYS Fluent and COMSOL, commonly used for FEM simulations
  - Due to computational cost, they are suitable only for validation
- **Computational overhead and impractical execution time of FEM solvers motivate analytical models that enable rapid thermal evaluation in early design phases**
  - Most common method is constructing thermal RC networks
  - Recently proposed PACT framework employs a similar methodology by utilizing SPICE tools as solvers, focusing on standard-cell-level thermal analysis for 2.5D systems
- **However, these tools are not fast enough for large-scale, thermal-aware DSE and dynamic resource management of multi-chiplet systems**

Space & Time : Continuous vs Discrete

	Continuous	Discrete
Space		
Time		

Multi-Fidelity Thermal Modeling

Increasing speed →

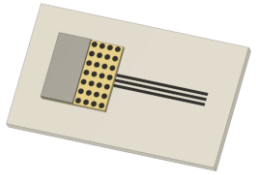
← Increasing accuracy

	Finite Element Method (FEM)		Analytical Models	
	1. Fine-grained	**2. Abstracted**	**3. Thermal RC Model**	**4. Discrete State-Space (DSS)**
Features	Accurate (e.g., real µbumps, links)	Replace micro-structures with equivalent material blocks	Independent of specific geometry, continuous time	Derived a specific architecture, discrete time
Error	Golden reference	< 0.5 °C	< 1.7 °C	Same as thermal RC*
Runtime	Not possible to model entire package	Days	Seconds	Milliseconds
Use case	Validate the abstracted FEM models	Ground truth to tune C values in Thermal RC model	Thermally-aware design space exploration, reference for DSS model	Large-scale optimization, thermal management

$$\boldsymbol{T[k+1] = A \times T[k] + B \times P[k]}$$

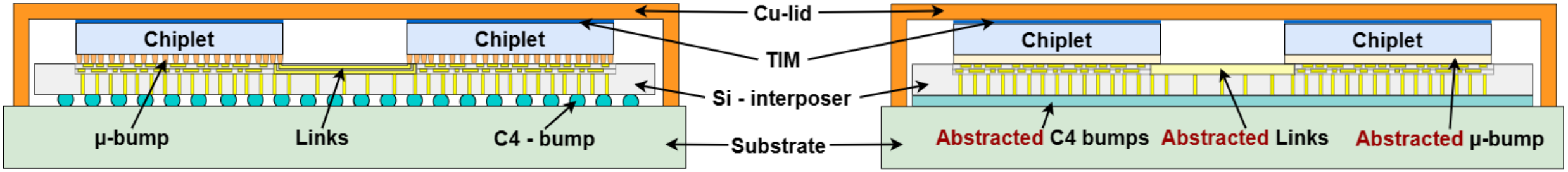
*Loses the generality.
Since the model is derived for a specific configuration, it needs to be regenerated if the architecture changes.

A Novel Multi-Fidelity Thermal Modeling Approach



1. Develop fine-grained FEM Models

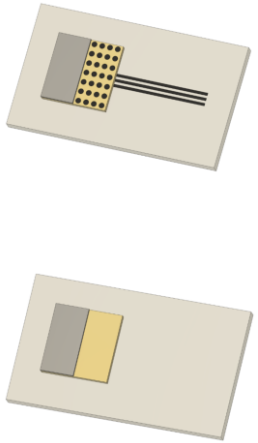
- **Model all the architecture, geometry, and parameters in as much detail possible**
- **Examples:**
  - All links
  - Microbumps
  - C4 (controlled-collapse chip connection) bumps



(a) Detailed FEM Model

(b) Abstracted FEM Model

A Novel Multi-Fidelity Thermal Modeling Approach

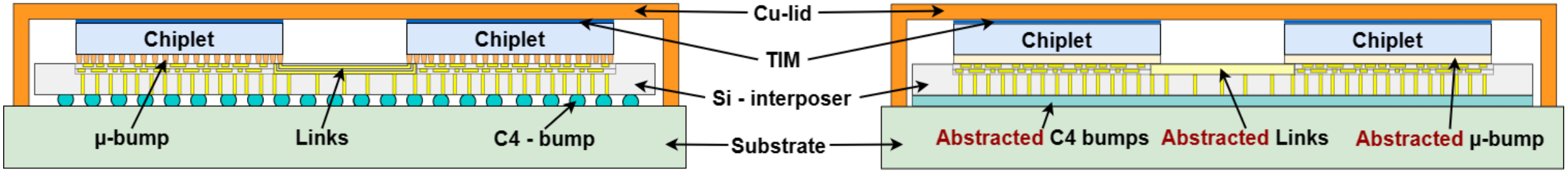


1. Develop fine-grained FEM Models
2. Abstract complex and detailed components

Iterate until desired fidelity achieved

Objective: < 0.5° temperature difference

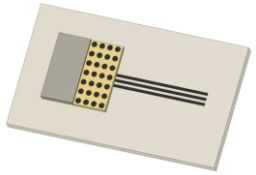
- **Abstract micro-structures that increase the FEM simulation time without significant accuracy impact**



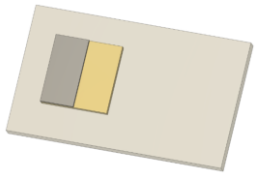
(a) Detailed FEM Model

(b) Abstracted FEM Model

A Novel Multi-Fidelity Thermal Modeling Approach

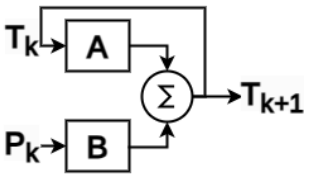


1. Develop fine-grained FEM Models



2. Abstract complex and detailed components

3. Develop thermal RC models



4. Derive design-specific DSS model

Iterate until desired fidelity achieved

Objective: < 0.5° temperature difference

Construct a thermal RC network.

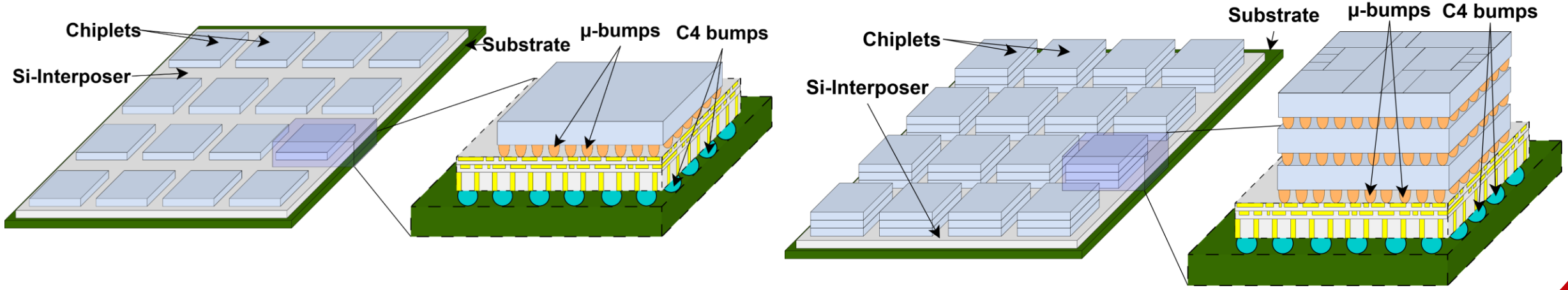
Objective: Minimize the temperature difference with the FEM result

Derive linear discrete state-space models

Objective: Approximate the thermal RC network for the desired fidelity

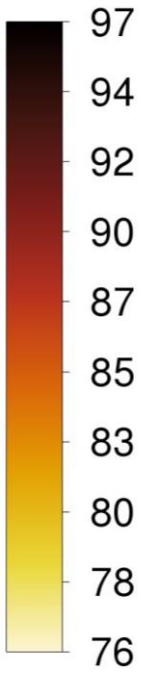
Experimental Setup

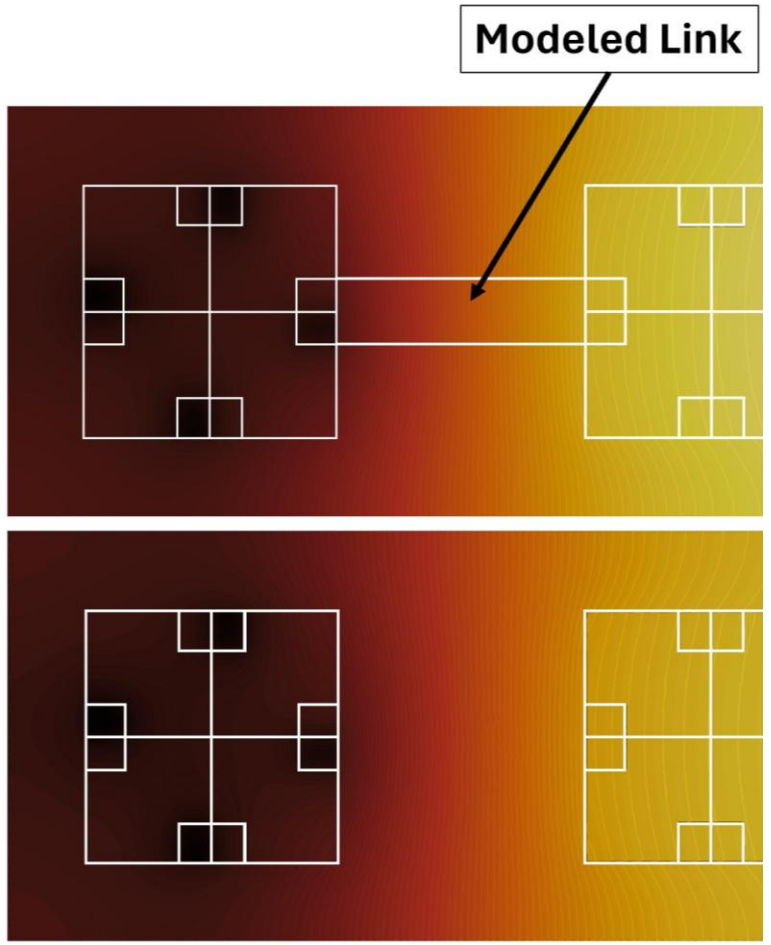
- **FEM simulation is assumed as the golden reference**
  - Evaluate abstracted FEM
  - Thermal RC network
  - Discrete state-space models
- **4x4, 6x6, 8x8 and 4x4x3 chiplet-based systems**
- **Tested on five different neural network workloads**
  - Workloads are a mix of deep neural networks



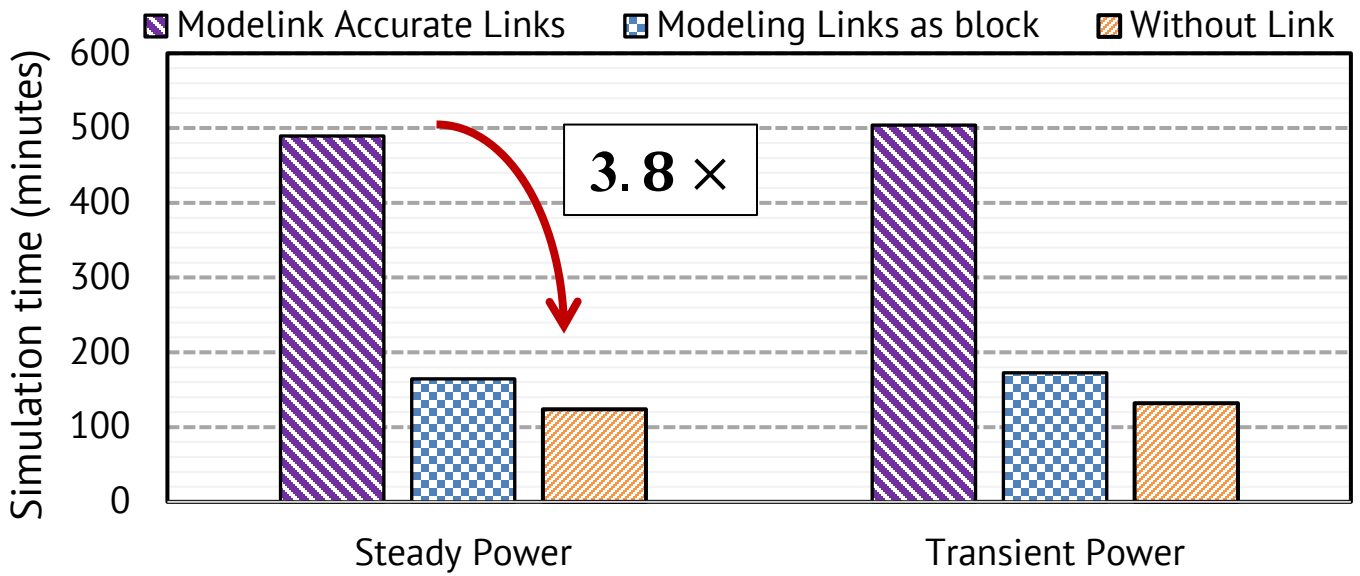
FEM: Accurate vs Abstracted Links

- **One chiplet is active – generating heat**
- **Inactive chiplet heated by *Thermal Crosstalk***
  - ➢ Interposer, heat spreader
- **Three alternative interposer models:**
  1. All links modeled accurately
  2. Links are abstracted as a block
     Material is assumed a mix of Cu and Si
  3. Links are removed



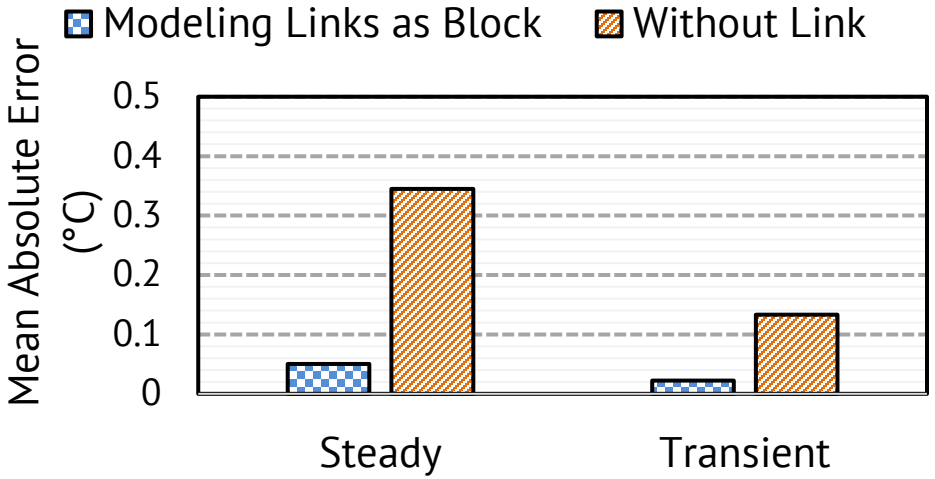


FEM: Accurate vs Abstracted Links



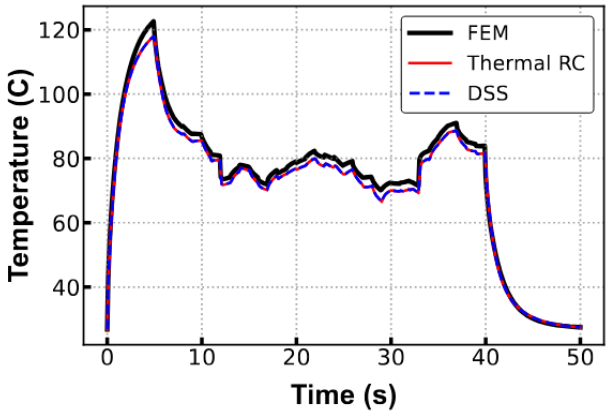
- **Abstracting the links as a single block**
  - Has negligible (<0.05°C) impact
- **Removing links degrades accuracy**

- **Abstracting the links as a single block**
  - Speed up from 503 min to 132 min
- **Removing links has diminishing returns**

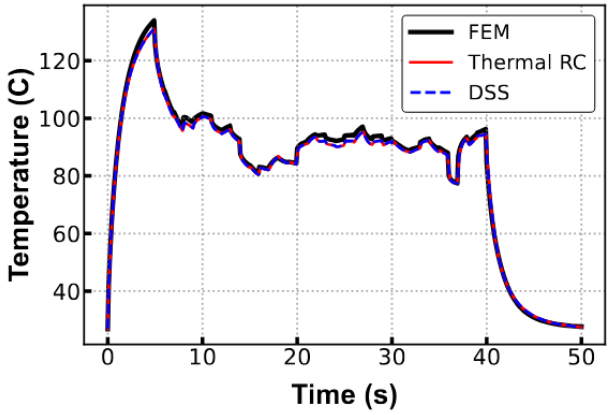


Experimental Results: Accuracy

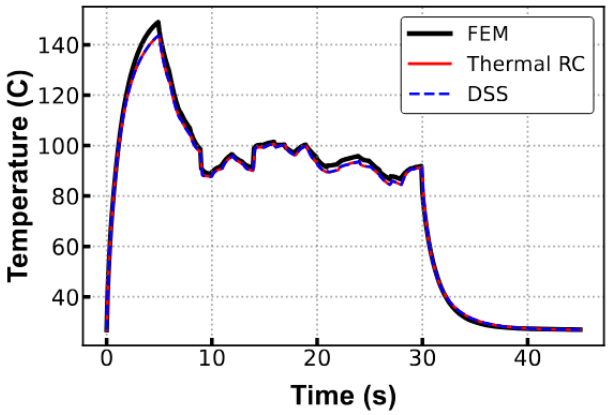
- **Temperature as a function of time**
  - Constant and pseudo-random power inputs to capture transient and steady-state behavior



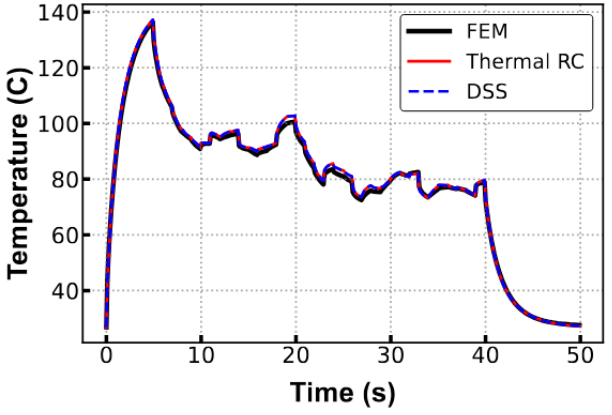
(a) 2.5D - 16 chiplet system



(b) 2.5D - 36 chiplet system



(c) 2.5D - 64 chiplet system

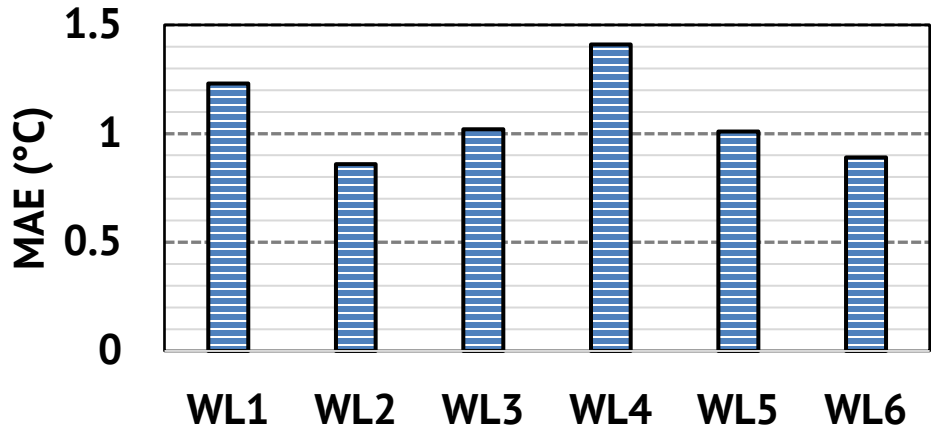


(d) 3D - 16x3 chiplet system

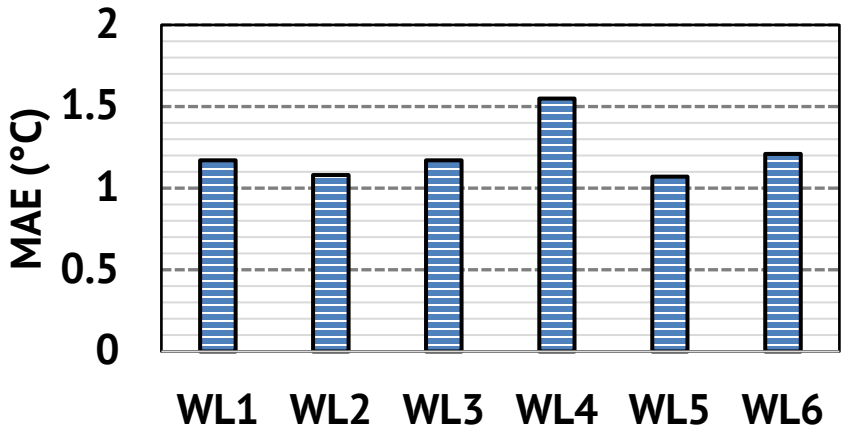
Negligible difference in transient temperature (even smaller difference in steady-state)

Experimental Results: Error

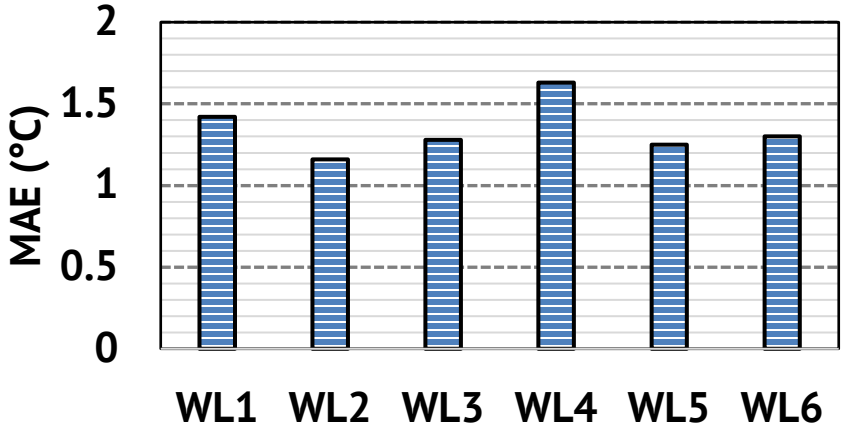
- **<1.7°C Mean Absolute Error wrt FEM**



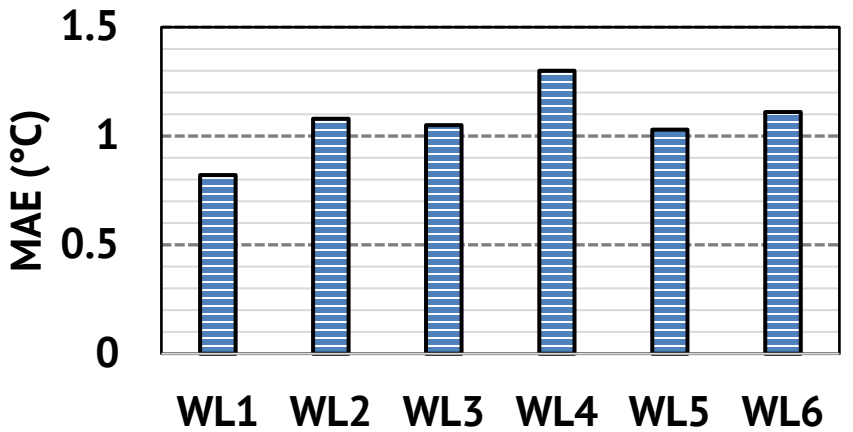
4x4 Chiplet system. Errors wrt FEM model



8x8 Chiplet system. Errors wrt FEM model



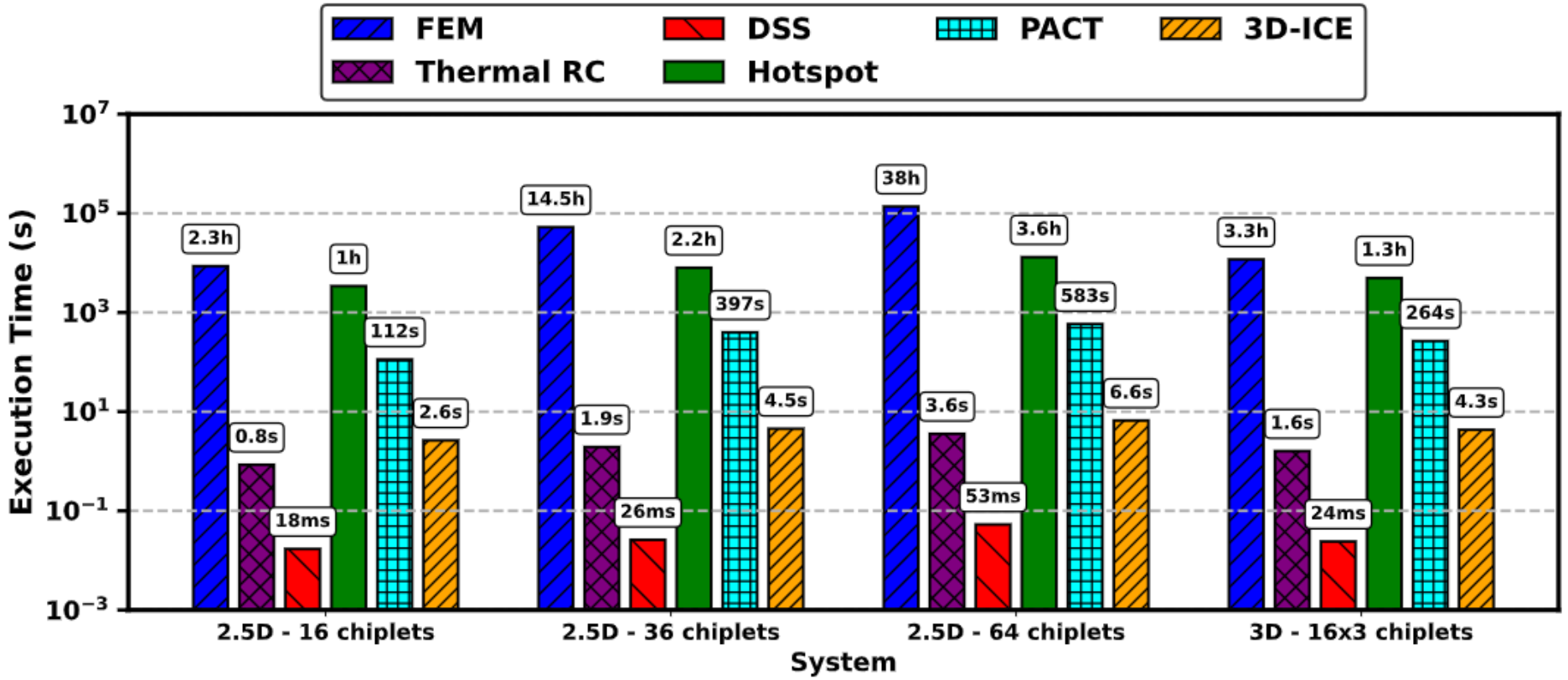
6x6 Chiplet system. Errors wrt FEM model



3D-16x3 Chiplet system. Errors wrt FEM model

Experimental Results: Speedup

- **Thermal RC model is ~10,000x faster than FEM**
- **DSS model is >50x faster compared to our thermal RC**



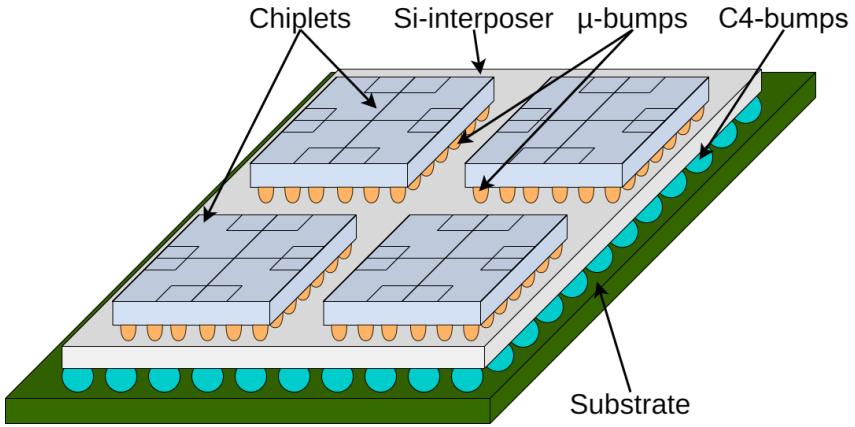
[1] Z. Yuan, et al, "PACT: An Extensible Parallel Thermal Simulator for Emerging Integration and Cooling Technologies," in IEEE TCAD, 2021

[2] A. Sridhar, et al, "3D-ICE: Fast compact transient thermal modeling for 3D ICs with inter-tier liquid cooling" ICCAD 2010

[3] Wei Huang, et al, "HotSpot: a compact thermal modeling methodology for early-stage VLSI design," in *IEEE Transactions on VLSI* 2006

Qualitative analysis

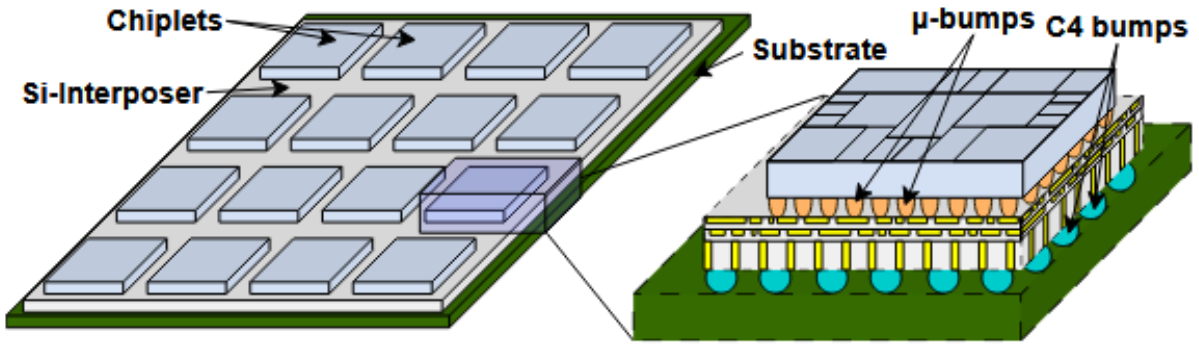
	Non-uniform grid	Anisotropic Materials	Non-homogeneous layers	Heat dissipation from both boundaries	Flexible with Architecture changes
Hotspot	✗	✗	✓	✓	✓
PACT	✗	✗	✓	✗	✓
3D-ICE	✓	✗	✗	✗	✓
Thermal RC(ours)	✓	✓	✓	✓	✓
DSS (ours)	✓	✓	✓	✓	✗

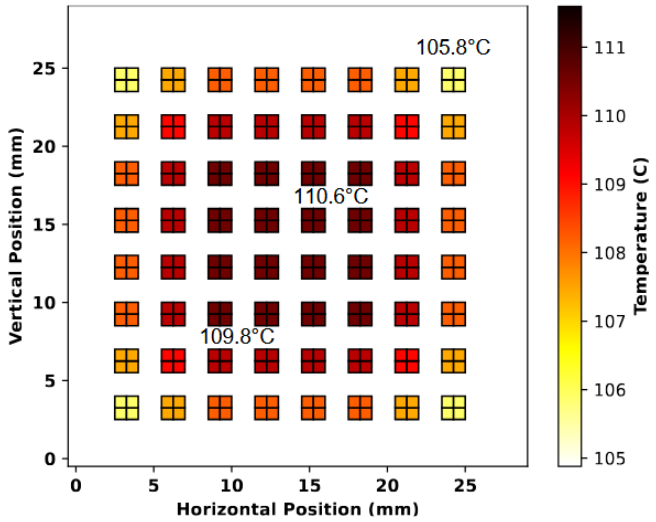


[1] Z. Yuan, et al, "PACT: An Extensible Parallel Thermal Simulator for Emerging Integration and Cooling Technologies," in IEEE TCAD, 2021
[2] A. Sridhar, et al, "3D-ICE: Fast compact transient thermal modeling for 3D ICs with inter-tier liquid cooling" ICCAD 2010
[3] Wei Huang, et al, "HotSpot: a compact thermal modeling methodology for early-stage VLSI design," in *IEEE Transactions on VLSI* 2006

MFIT Conclusions

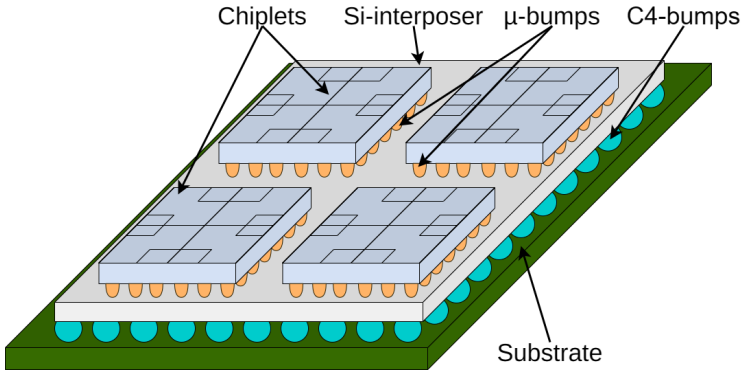
- We proposed MFIT, a multi-fidelity thermal modeling tool for every stage of the chiplet system design process
- MFIT balances speed and accuracy, matching the requirements of the current design stage, while supporting a broad range of chiplet system configurations
- Published at ACM Transactions on Design Automation of Electronic Systems, 2025
- This tool enables the next work: an exploration of a novel cooling approach for chiplet-based systems.

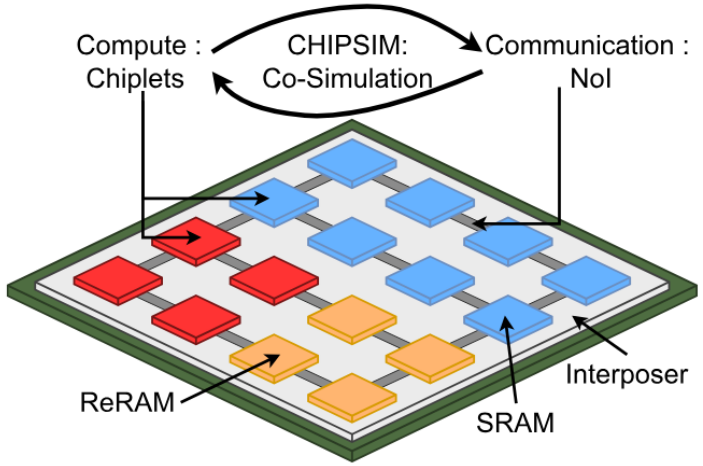




Outline

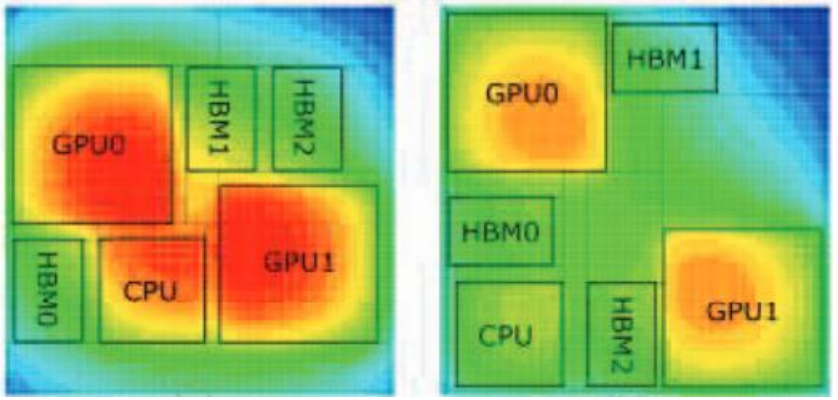
- **Motivation: Chiplet-based platforms**
- **Preliminary Work-1:**
  - CHIPSIM: A Co-Simulation Framework for Deep Learning on Chiplet-Based Systems
- **Preliminary Work-2:**
  - MFIT : Multi-Fidelity Thermal Modeling for 2.5D and 3D Multi-Chiplet Architectures
- **Ongoing and Proposed Work:**
  - Exploration of Thermoelectric Coolers (TECS) for Chiplet-Based Systems





Existing Cooling Approaches

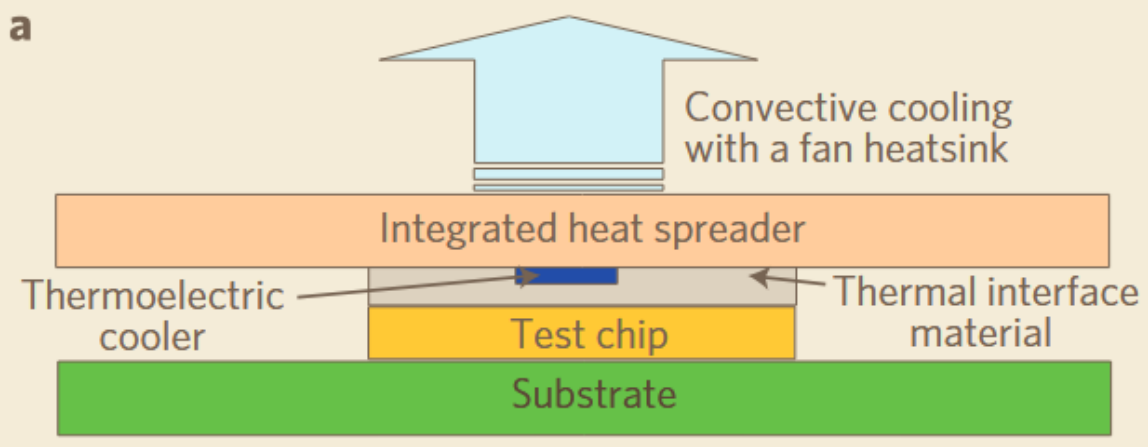
- **Many approaches exit to manage the temperature of chiplet-based systems. For example:**
  - Air, liquid, and micro-channel cooling
  - Thermal-aware floorplanning
  - DVFS and Scheduling optimizations
- **In practice, several of these techniques are combined in a single system**
- **However, in chiplet-based systems, Thermoelectric Coolers (TECs) have not been explored as a possible addition.**
- **In this work, we explore the efficacy of TECs for cooling chiplet-based systems.**

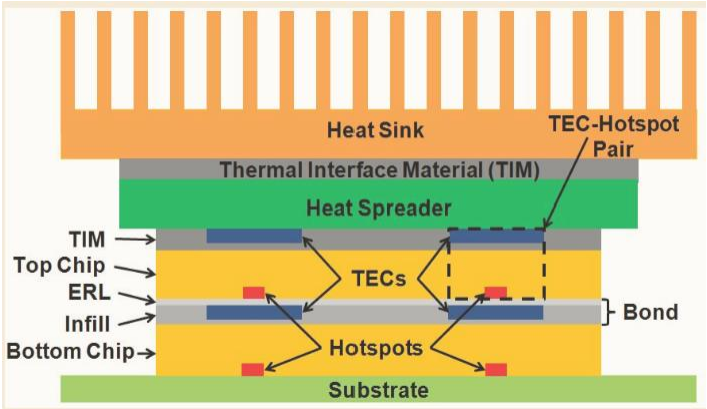


[1] EDA Association, IEEE Computer Society. Technical Council on Test Technology, and IEEE Solid-State Circuits Society, *TAP-2.5D: A Thermally-Aware Chiplet PlacementMethodology for 2.5D Systems*.

Prior Work

- **Thermoelectric devices have been previously applied to cool hotspots in monolithic chips (left fig.)**
  - Artificial hotspot is created by heater and managed by the thermoelectric cooler.
- **Also explored in 3D integration (simulation only, right fig.)**
  - TECs placed above hotspots in 3D stacked chips, impact of parameters and use examined.
- **No prior work specifically for chiplet-based systems.**
- **Our contribution: a simulation framework specifically for TEC-to-chiplet application, and evaluation of the effectiveness of TECs when applied to chiplets.**



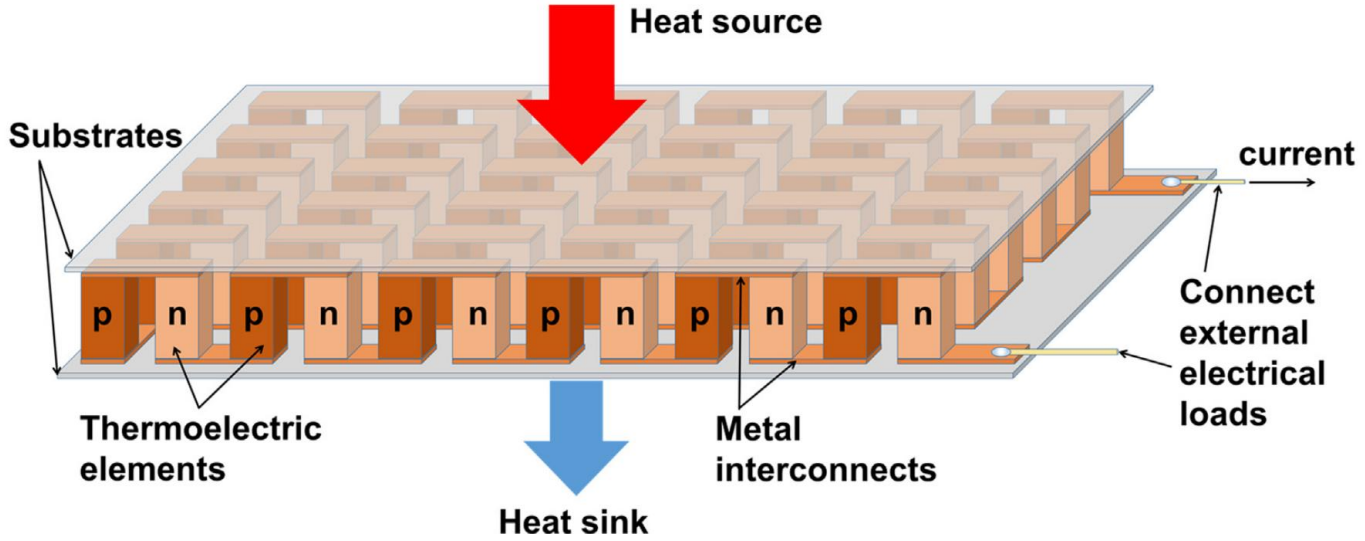


[1] I. Chowdhury *et al.*, "On-chip cooling by superlattice-based thin-film thermoelectrics," *Nature Nanotech*, vol. 4, no. 4, pp. 235–238, Apr. 2009, doi: 10.1038/nnano.2008.417.

[2] M. Redmond, K. Manickaraj, O. Sullivan, S. Mukhopadhyay, and S. Kumar, "Hotspot Cooling in Stacked Chips Using Thermoelectric Coolers," *IEEE Trans. Compon., Packag. Manufact. Technol.*, vol. 3, no. 5, pp. 759–767, May 2013, doi: 10.1109/TCPMT.2012.2226721.

Background: Thermoelectric Coolers

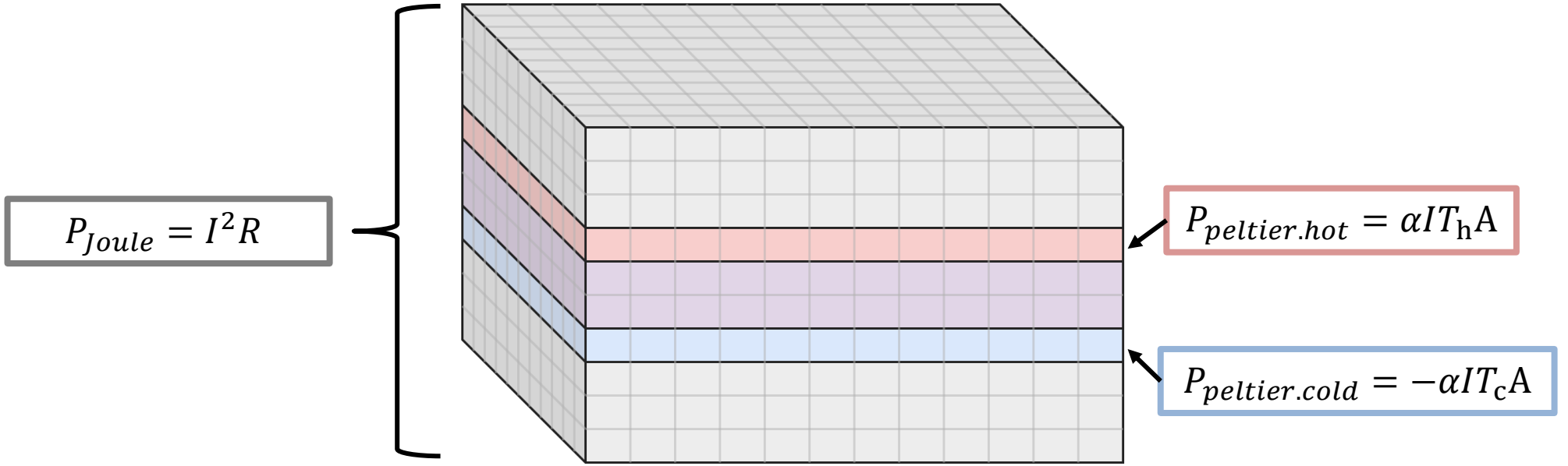
- **TECs are small electric heat pumps. Heat is absorbed on one side and rejected on the other.**
- **As a cost, they consume power, which adds heat to the system.**
- **Physically, they consist of many silicon legs which are thermally in parallel and electrically in series.**
- **Heat absorbed and rejected by the device is modeled by:**
  - $Q_c = \alpha I T_c - \left(\frac{1}{2}\right) I^2 R - K(T_h - T_c)$
  - $Q_h = \alpha I T_h + \left(\frac{1}{2}\right) I^2 R - K(T_h - T_c)$
  - Where α is a material parameter, I is current, R is electrical resistance, K is thermal conductivity, and T is hot/cold side temperature.



[1] P. Wang, B. L. Wang, and J. E. Li, "Temperature and performance modeling of thermoelectric generators," *International Journal of Heat and Mass Transfer*, vol. 143, p. 118509, Nov. 2019, doi: 10.1016/j.ijheatmasstransfer.2019.118509.

Modeling of TECs

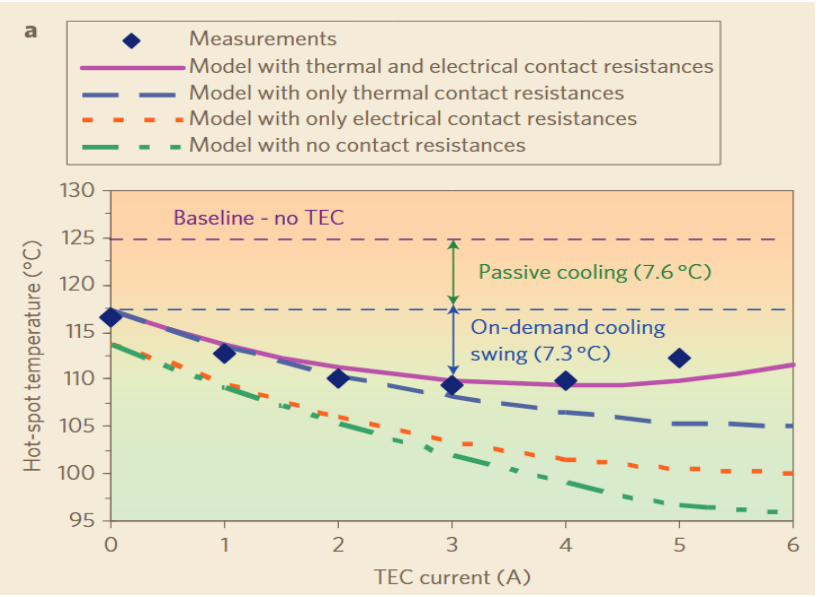
- **Instead of modeling TECs at leg-level details, layers are homogenized, and the simplified device is integrated to our thermal model, MFIT.**
  - This integration leverages the benefits of MFIT for quick and broad simulation
- **Each parameter of TEC equation is added to the existing Thermal RC network.**

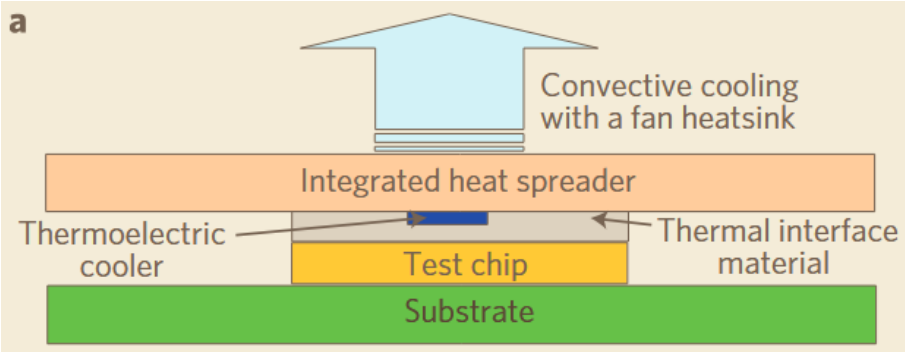


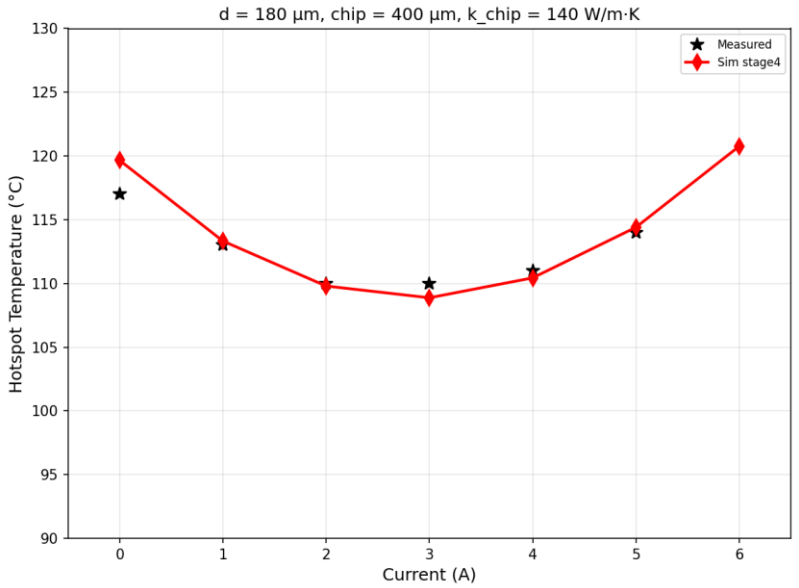
* Layers not to scale

Model Validation

- We validate our implemented model against the results of prior work. Prior work reports the temperature of the chip hotspot as a function of TEC current.
- Upper plot is prior work results, lower is our fitted model results.
- Our TEC model performance is nearly identical to the physically measured device





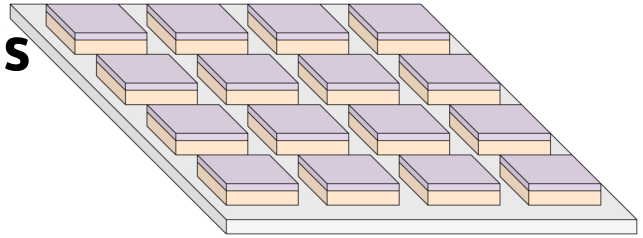


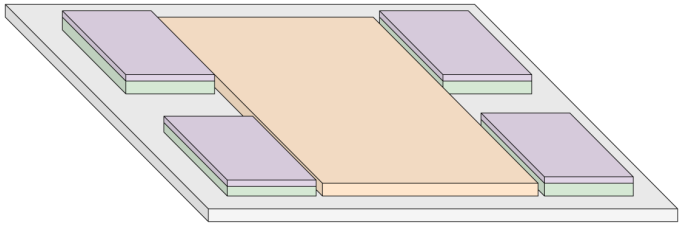
[1] I. Chowdhury *et al.*, "On-chip cooling by superlattice-based thin-film thermoelectrics," *Nature Nanotech*, vol. 4, no. 4, pp. 235–238, Apr. 2009, doi: 10.1038/nnano.2008.417.

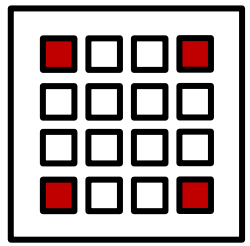
System Exploration

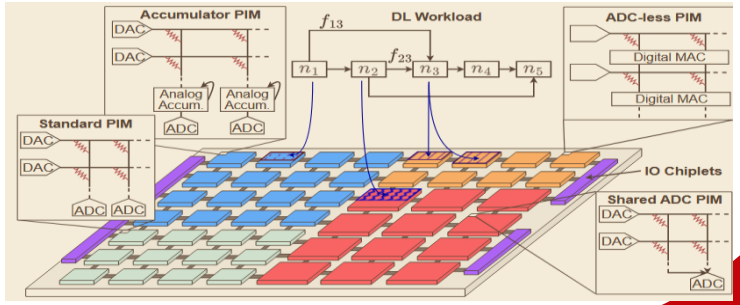
- We examine 4 cases in total. Two are highlighted in this presentation:

1. Simba-like: homogeneous power consumption, temperature sensitivity, chiplet size, spacing
2. V100: Commercial system with temperature-sensitive memory chiplets
3. Simba-like + Memory Chiplets: Same scenario as (1), but with added sensitive, lower power memory chiplets
4. THERMOS-like: replicate thermos system with clusters of more sensitive chiplets



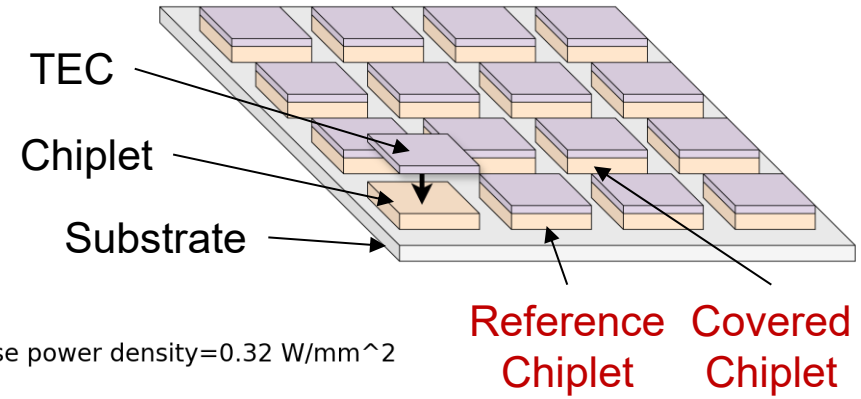


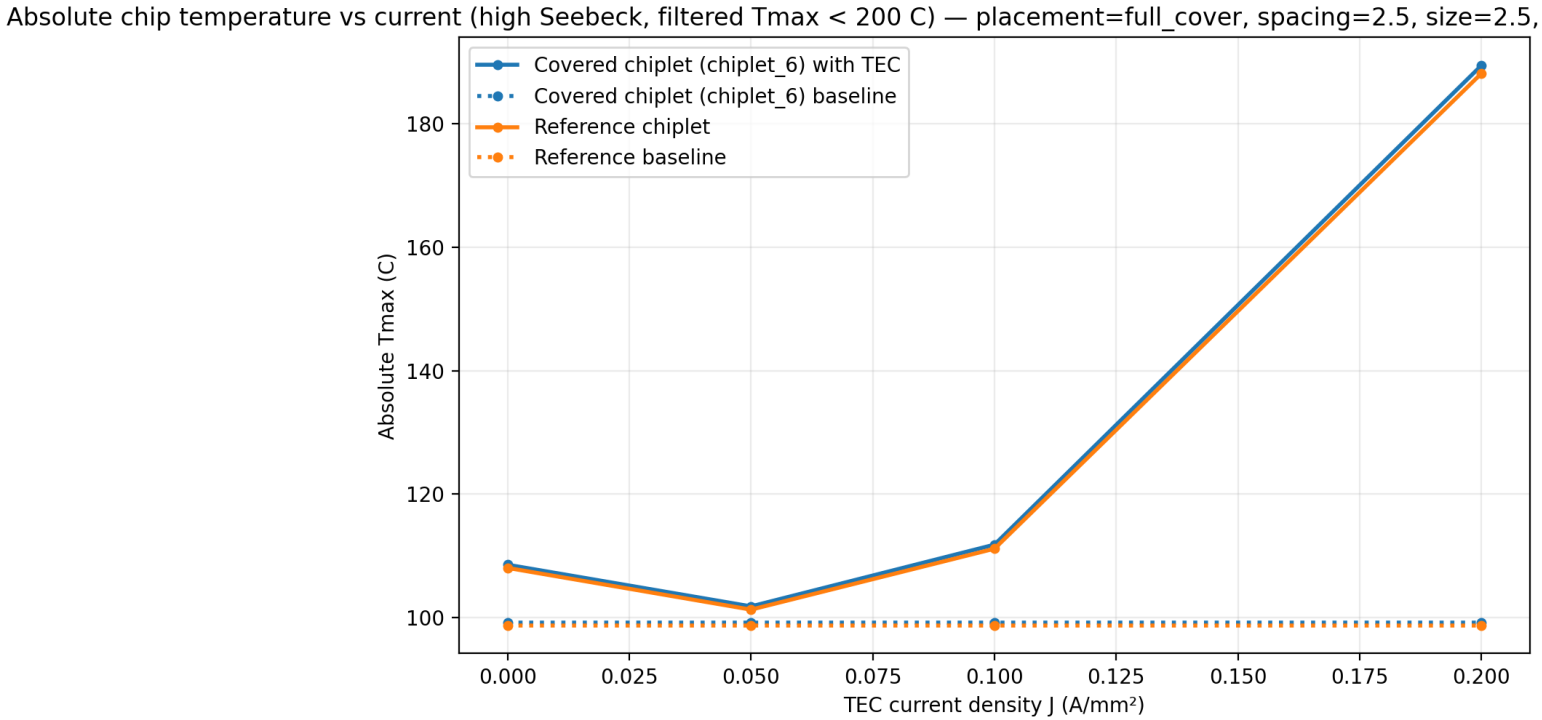




System 1: Cover All

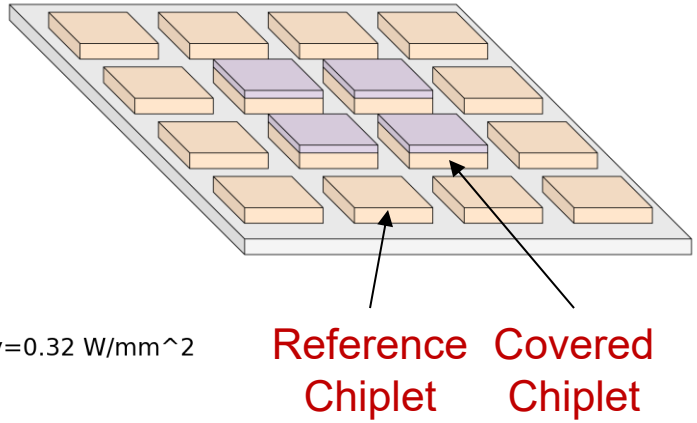
- **Rationale: We care about all chiplet temperatures equally, so we should cover all equally.**

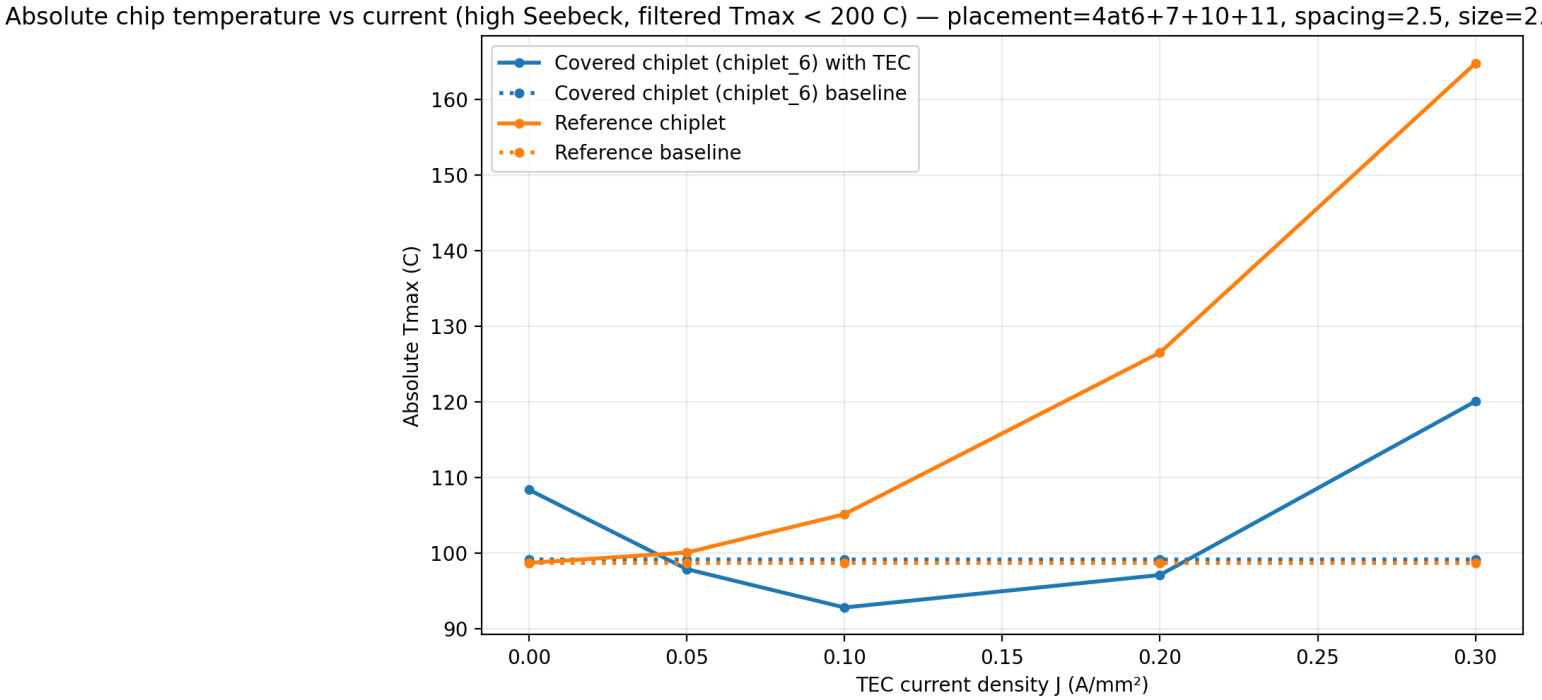




System 1: Cover Center Chiplets Only

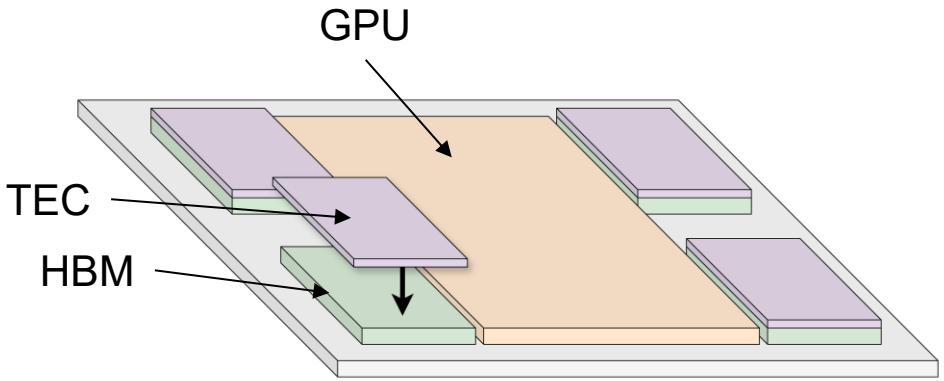
- **Rationale: center chiplets should generally be hotter, so they should get extra cooling**
- **Base system is identical to prior case.**



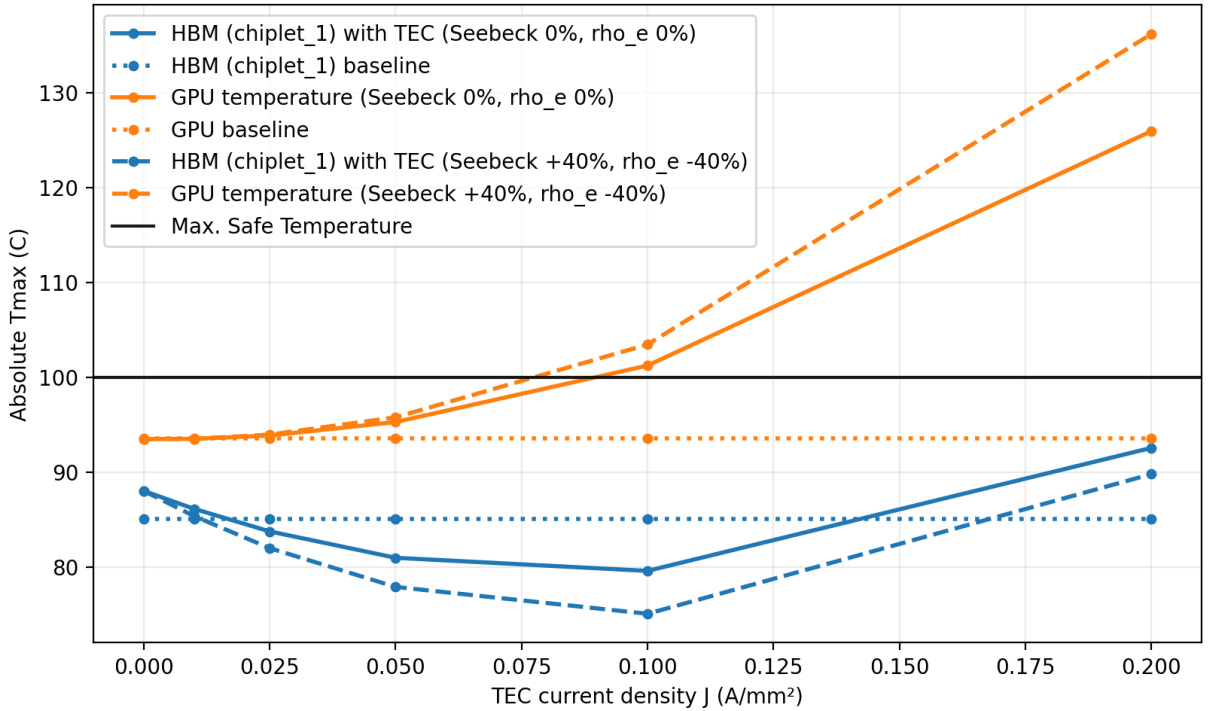


System 2: GPU System Results

- **Below plot shows temperature of GPU and HBM chiplet as TEC current is swept.**



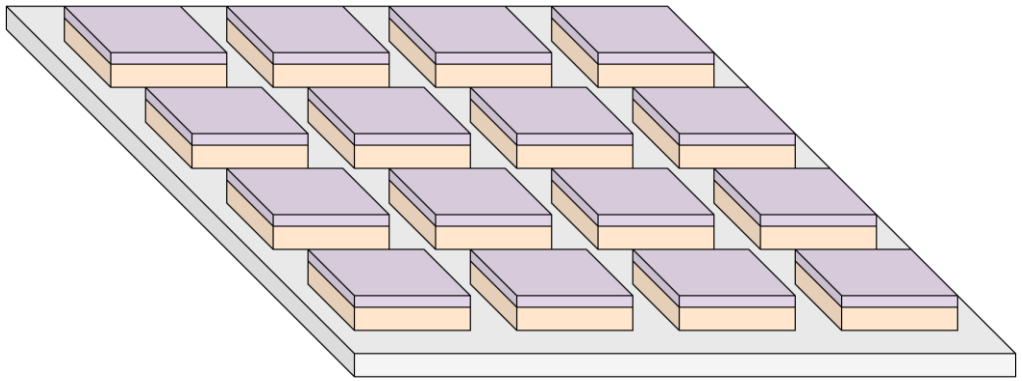
Absolute chip temperature vs current (filtered Tmax < 200 C; overlay Seebeck +40%, rho_e -40%) — placement=hbm_only, HBM power=12.0 W



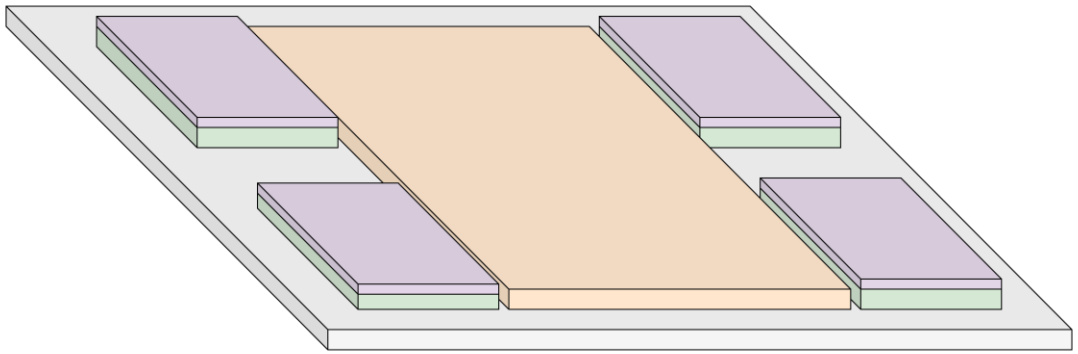
Early Conclusions

- When system is made up of homogeneous chiplets, in terms of both power consumption and temperature sensitivity (System 1), then TEC integration is harmful.
- When system is heterogeneous in power consumption or temperature sensitivity (System 2), TECs have the potential to reduce temperatures and improve performance.

No Good

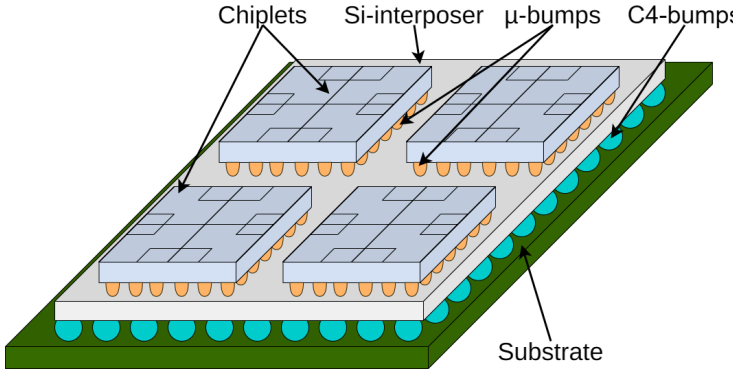


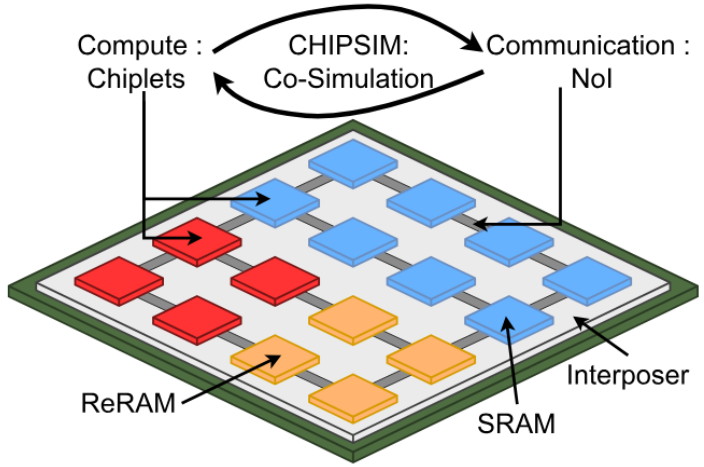
Has Potential



Outline

- **Motivation: Chiplet-based platforms**
- **Preliminary Work-1:**
  - CHIPSIM: A Co-Simulation Framework for Deep Learning on Chiplet-Based Systems
- **Preliminary Work-2:**
  - MFIT : Multi-Fidelity Thermal Modeling for 2.5D and 3D Multi-Chiplet Architectures
- **Ongoing and Proposed Work:**
  - Exploration of Thermoelectric Coolers (TECs) for Chiplet-Based Systems
  - TEC Control for Chiplet-Based Systems
  - Tutorial on the State of Thermal Research on Chiplet-Based Systems

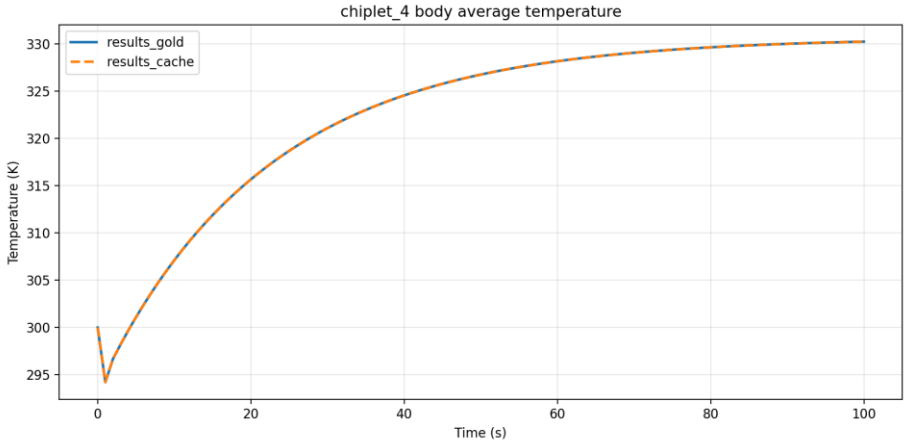




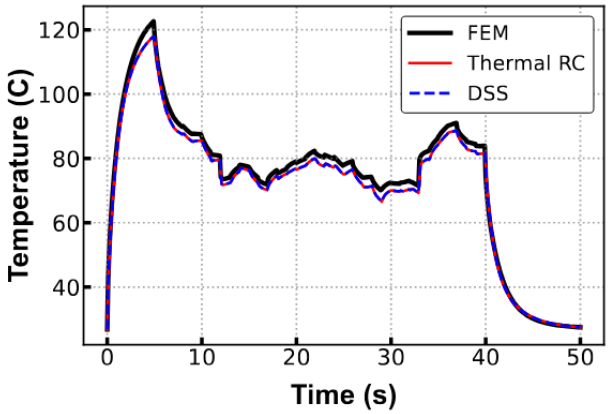
Chiplet-Thermoelectric Control

- **Our current exploration of TECs on chiplet-based system is limited to steady-state evaluation.**
  - All chiplet power consumption values and TEC input current are constant with respect to time. Temperatures are measured at steady-state.
- **In real systems, chiplet power consumption is transient, and TEC input current can be controlled at runtime.**
- **Our follow-up to our Chiplet-TEC exploration paper aims to do the following:**
- **Formulate control methods for TECs applied to chiplets which minimize system temperature and TEC power consumption.**

Steady-State



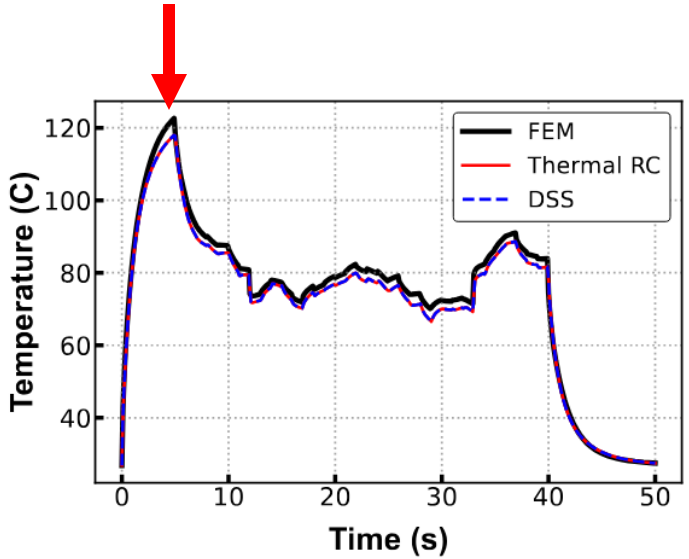
Transient



(a) 2.5D - 16 chiplet system

Approach: Minimize Peak Temp. and Power

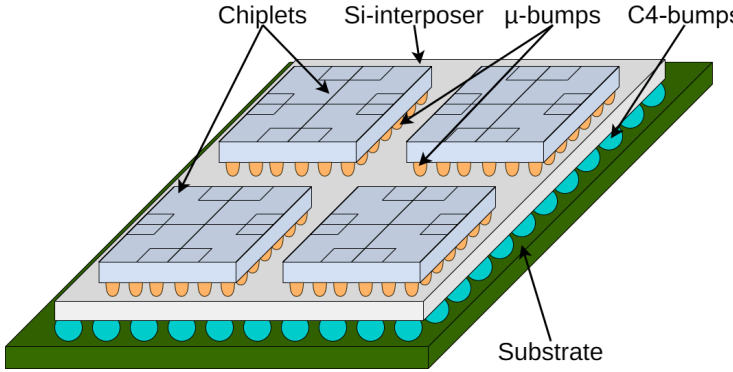
- When a TEC is active, it consumes power. Therefore, it should only be turned on when needed
- Most pressing need is to mitigate temperature spikes and reduce thermal throttling, increasing performance.
- Develop control method to optimize this temperature spike mitgation

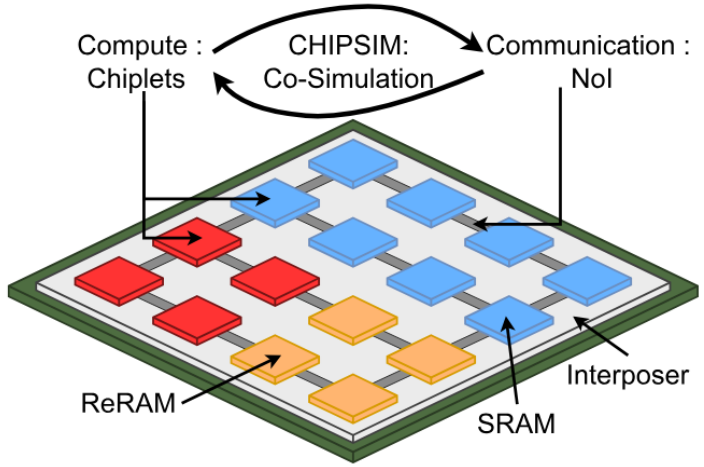


(a) 2.5D - 16 chiplet system

Outline

- **Motivation: Chiplet-based platforms**
- **Preliminary Work-1:**
  - CHIPSIM: A Co-Simulation Framework for Deep Learning on Chiplet-Based Systems
- **Preliminary Work-2:**
  - MFIT : Multi-Fidelity Thermal Modeling for 2.5D and 3D Multi-Chiplet Architectures
- **Ongoing and Proposed Work:**
  - Exploration of Thermoelectric Coolers (TECs) for Chiplet-Based Systems
  - TEC Control for Chiplet-Based Systems
  - Tutorial on the State of Thermal Research on Chiplet-Based Systems





Need for the Tutorial

- Thousands of papers exist on the thermal aspects of chiplet-based systems.
- Topics span modeling approaches, management strategies, new design techniques, and so on.
- However, no recent survey papers perform an exhaustive review of this area.
- Therefore, we propose to fill this gap.
- Additionally, for my own progress, this survey will provide deeper insight into the state of the field and potential new research.

Tutorial Approach

- **We apply a filtering methodology to identify the corpus of papers to consider.**
- **First, obtain dataset of all papers (sources include all major journals and conferences)**

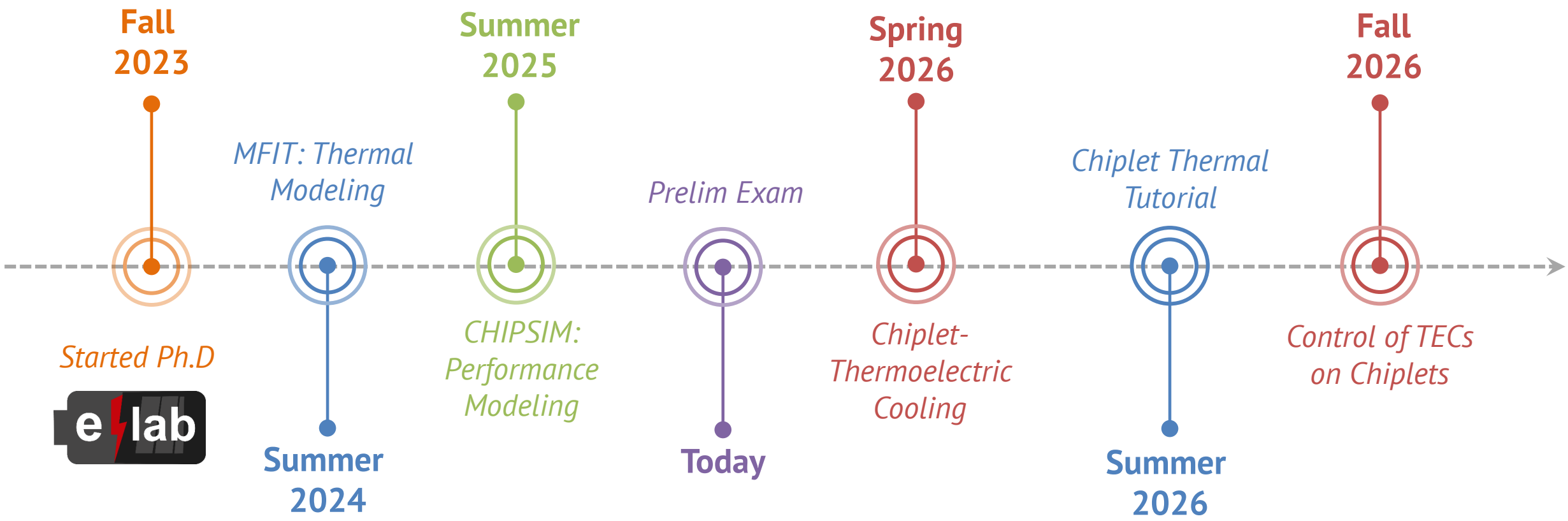
1. **Filter by domain (engineering, CS, materials, etc.)**
2. **Filter by Chiplet keywords**
3. **Filter by Thermal keywords**
4. **Semantic filter**
   - Apply AI model trained specifically to identify similar papers, trained on citation relationships

Early Results

- **Our filtering pipeline identifies around 2000 papers which are related to both the domains of Chiplets and Heat/Thermal.**
- **Next step is to identify clusters within these candidate papers.**
- **From paper clusters, identify potential key sections in the final paper.**

Filter Stage	Filter Description	# Papers Before	# Papers After
1	Domain Filter (Engineering, CS, etc.	230,000,000	130,000,000
2	Chiplet Keywords	130,000,000	35,000
3	Thermal Keywords	35,000	2,700
4	Semantic Filter	2,700	1,800

Tentative Timeline



List of Publications

[1] **Pfromm, L.*,** Kanani, A.*, Sharma, H., Solanki, P., Tervo, E., Park, J., Doppa, J., Pande, P.P. and Ogras, U., 2025. *"MFIT: Multi-fidelity thermal modeling for 2.5 D and 3D multi-chiplet architectures,"* ACM TODAES.

[2] Park, J., Kanani, A., **Pfromm, L.,** Sharma, H., Solanki, P., Tervo, E., Doppa, J.R., Pande, P.P. and Ogras, U.Y., 2024. *"Thermal modeling and management challenges in heterogenous integration: 2.5 D chiplet platforms and beyond,"* IEEE VTS. (INVITED)

[3] Kanani, A., **Pfromm, L.,** Sharma, H., Doppa, J., Pande, P. and Ogras, U., 2025. *"THERMOS: Thermally-Aware Multi-Objective Scheduling of AI Workloads on Heterogeneous Multi-Chiplet PIM Architectures,"* ACM TECS.

[4] **Pfromm, L.,** Kanani, A., Sharma, H., Doppa, J., Pande, P. and Ogras, U., 2025. *"CHIPSIM: A Co-Simulation Framework for Deep Learning on Chiplet-Based Systems,"* IEEE OJ-SSCS.

*Equal contribution

Thank you!

e lab

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON