International Symposium
on Physical Design

CHASE: A CHiplet Architecture Simulation and Exploration Framework with Decoupled Multi-Fidelity Optimization

Shixin Chen[1], Hengyuan Zhang[2], Jianwang Zhai[2], Bei Yu[1]
[1]The Chinese University of Hong Kong
[2]Beijing University of Posts and Telecommunications
sxchen22@cse.cuhk.edu.hk

February 27, 2026

Outline

1. Background & Motivation
2. CHASE Framework Overview
3. Core Technology Details
4. Experimental Results
5. Conclusion

Background & Motivation

Chiplet Architecture: Opportunity & Challenges

Moore's Law Saturation: Transistor scaling faces physical limits and rising costs (complex lithography, reduced yields). Chiplet-based architecture emerges as a promising alternative.

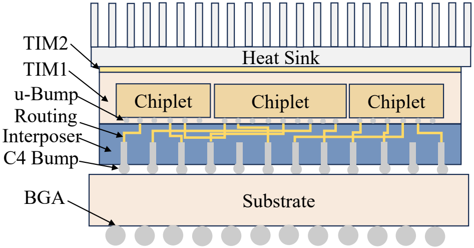


Figure 1: Typical 2.5D Chiplet Architecture (CoWoS)

Core Advantages of Chiplets:

- Heterogeneous integration (mix tech nodes for different modules)
- Higher yield & lower manufacturing cost
- Faster time-to-market (reuse existing IPs)
- Flexible system scaling

Critical Challenges in Chiplet Design

Existing solutions fail to address the comprehensive needs of chiplet-based architecture:

1. **Lack of Comprehensive Simulation:**
   - Most tools focus on NoC-based performance (coarse-grained)
   - Ignore advanced packaging overhead (interface, routing, area)
2. **Single-Dimension Optimization:**
   - Separate performance (latency/power) and manufacturing (area/thermal/cost) metrics
   - Metrics are interdependent (e.g., power $\leftrightarrow$ thermal, area $\leftrightarrow$ cost)
3. **Exploration in Large Design Space:**
   - Design space size: 1.07×10^9 combinations (Table 1)
   - Time-consuming full evaluations waste resources

Critical Challenges in Chiplet Design

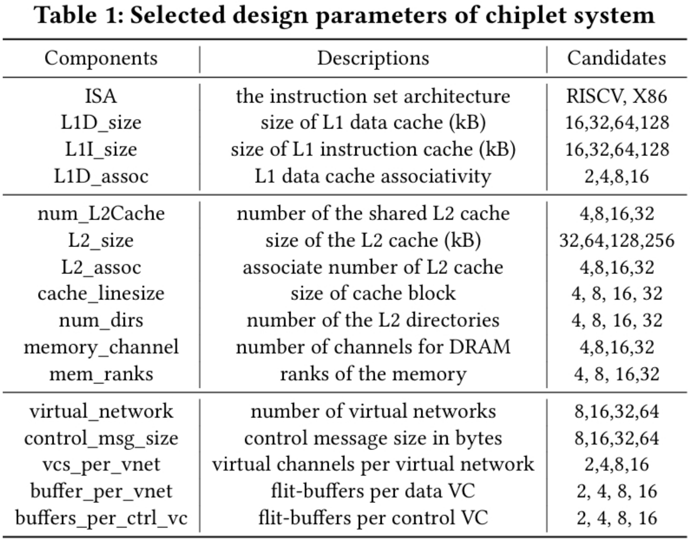


Table 1: Selected design parameters of chiplet system

Components	Descriptions	Candidates
ISA	the instruction set architecture	RISCV, X86
L1D_size	size of L1 data cache (kB)	16,32,64,128
L1I_size	size of L1 instruction cache (kB)	16,32,64,128
L1D_assoc	L1 data cache associativity	2,4,8,16
num_L2Cache	number of the shared L2 cache	4,8,16,32
L2_size	size of the L2 cache (kB)	32,64,128,256
L2_assoc	associate number of L2 cache	4,8,16,32
cache_linesize	size of cache block	4, 8, 16, 32
num_dirs	number of the L2 directories	4, 8, 16, 32
memory_channel	number of channels for DRAM	4,8,16,32
mem_ranks	ranks of the memory	4, 8, 16,32
virtual_network	number of virtual networks	8,16,32,64
control_msg_size	control message size in bytes	8,16,32,64
vcs_per_vnet	virtual channels per virtual network	2,4,8,16
buffer_per_vnet	flit-buffers per data VC	2, 4, 8, 16
buffers_per_ctrl_vc	flit-buffers per control VC	2, 4, 8, 16

Figure 2: Chiplet Design Space Complexity

CHASE Framework Overview

CHASE: Core Definition & Architecture

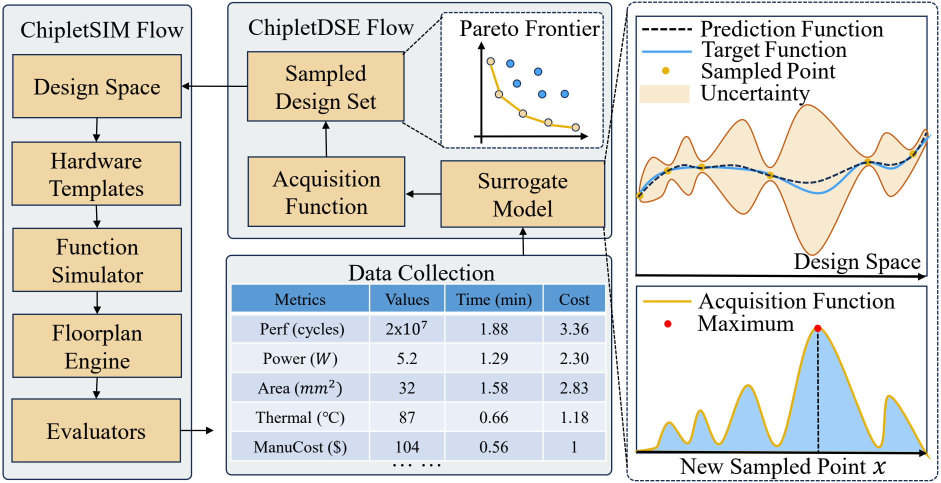
Core Definition:

A unified framework for **Chiplet-based Architecture Simulation and Exploration** that jointly optimizes performance and manufacturing metrics via decoupled multi-fidelity optimization.

Two Core Components:

- **ChipletSIM:** Comprehensive simulation flow
  - Performance-level metrics (latency, power)
  - Manufacturing-level metrics (area, thermal, cost)
- **ChipletDSE:** Efficient design space exploration
  - Decoupled multi-fidelity evaluation
  - Pareto frontier-driven optimization

CHASE: Core Definition & Architecture



Metrics	Values	Time (min)	Cost
Perf (cycles)	$2\text{x}10^7$	1.88	3.36
Power (W)	5.2	1.29	2.30
Area (mm^2)	32	1.58	2.83
Thermal (°C)	87	0.66	1.18
ManuCost ($)	104	0.56	1

… …

Figure 3: CHASE Framework Overview

Key Advantages Over Prior Work

Feature	Prior Work	CHASE
Simulation Scope	Single-level (performance/manufacturing)	Comprehensive (both levels)
Design Metrics	Partial (2-3 metrics)	Full (latency/power/area/thermal/cost)
Exploration Strategy	Full evaluation / Single-fidelity	Decoupled multi-fidelity
Design Space Coverage	Limited by budget	Maximized via early pruning
Pareto Optimization	No explicit support	Core objective (HV maximization)

Table: Comparison with Prior Chiplet Design Tools

Core Innovations:

- First framework to unify performance and manufacturing metrics for chiplets
- Decoupled evaluation aligns with staged simulation (saves 3.17x evaluation time)
- Hypervolume Knowledge Gradient (HVKG) for efficient Pareto exploration

Core Technology Details

ChipletSIM: Performance Level Evaluation

Performance Metrics: Latency & Power

Customized simulation flow based on gem5 and McPAT, supporting chiplet-specific characteristics.

1. Function Simulator (gem5-based):

- Flexible latency modeling for inter/intra-chiplet communication:

$$l_{L2} = l_{inter-chip} + l_{mem}$$

$$l_{inter-chip} = 2(l_{protocol} + l_{phy} + l_{adaptor}) + l_{interposer}$$

- Supports RISCV/X86 ISAs and parallel benchmarks (NPB, SPEC06, MatMul)
- Captures protocol overhead (UCIe, BoW)

ChipletSIM: Performance Level Evaluation

2. Power Evaluator (McPAT-based):

- Heterogeneous tech node support (22nm-90nm)
- Parallel evaluation for multi-chiplet systems
- Reports dynamic/peak power per chiplet

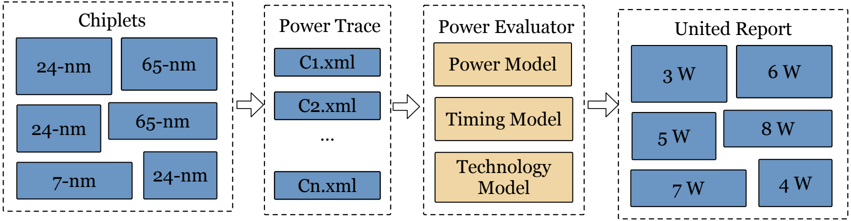


Figure 4: Power Evaluation for Heterogeneous Chiplets

ChipletSIM: Manufacturing Level Evaluation

Manufacturing Metrics: Area, Thermal, Cost

Integrated evaluation engine to model packaging-specific constraints and trade-offs.

1. Floorplan Engine:

- Constraints: No overlap, aspect ratio limits
- Objectives: Balance area, thermal, HPWL

$$\lambda_A \cdot A + \lambda_T \cdot T + \lambda_W \cdot W$$

- Algorithms: MP, Simulated Annealing

2. Thermal Evaluator (HotSpot-based):

- Power-driven thermal mapping
- Captures inter-chiplet heat transfer
- Reduces peak temp by 1.7 C (Fig. 9)

ChipletSIM: Manufacturing Level Evaluation

3. Cost Evaluator:

- Full 2.5D package cost model:

$$C_{2.5D} = \frac{\frac{C_{pack}}{Y_{pack}} + \sum_{i \in n} \left(\frac{C_{chip_i}}{Y_{chip_i}} + C_{bond_i} \right)}{Y_{bond}^{n}}$$

- Includes yield, bonding, interposer costs

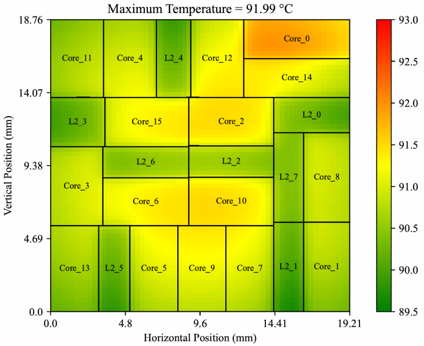


Figure 5: Floorplan Optimization (Area -19.4%, Temp -1.7C)

ChipletDSE: Design Space & Optimization Strategy

Decoupled Multi-Fidelity Optimization:

Aligns with staged simulation (performance → power → thermal/cost) to prune bad designs early.

Design Space Construction (Table 1): 3 key parameter groups (total 1.07×10^9 combinations):

- Core & L1 Cache (ISA, cache size/associativity)
- L2 Cache & Memory (number of channels, ranks, cache size)
- Communication Protocols (virtual networks, buffer sizes, msg size)

ChipletDSE: Design Space & Optimization Strategy

Key Concepts:

- **Pareto Frontier:** Non-dominated design set
- **Hypervolume (HV):** Measures Pareto quality
- **HVKG Acquisition:**

$$\alpha_{D-HVKG}(x, m) = \frac{\mathbb{E}[HV_{(x)} - HV^*]}{\lambda_D(x, m)}$$

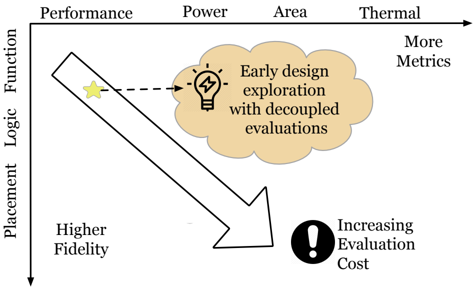


Figure 6: Decoupled Evaluation Workflow

ChipletDSE: Exploration Flow

Iterative Exploration Process:

1. **Initial Sampling:** Randomly sample design points in the large design space
2. **Surrogate Modeling:** Train GP models for each metric (latency/power/area/thermal/cost)
3. **Acquisition Maximization:** Select next design via α_{D-HVKG} (balances HV gain and cost)
4. **Staged Evaluation:** Prune designs with poor low-fidelity metrics (no full evaluation)
5. **Model Update:** Incorporate new data to refine GP models
6. **Pareto Construction:** Generate final non-dominated design set

ChipletDSE: Exploration Flow

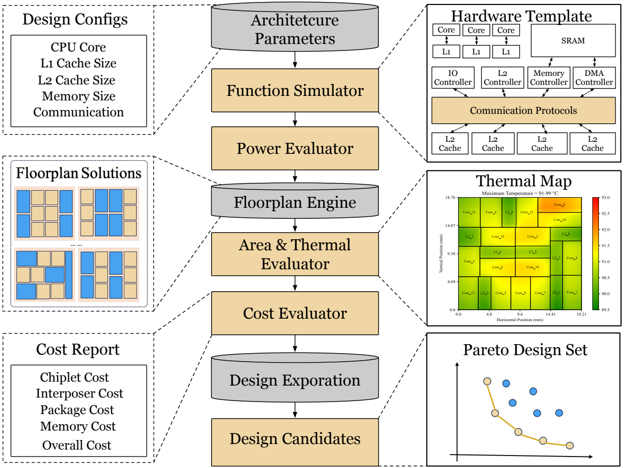


Figure 7: CHASE Exploration Flowchart

Key Benefit: Evaluates 3.17x more designs under the same budget (vs. full evaluation methods)

Experimental Results

Experimental Setup

Hardware & Software:

- Hardware: Intel Xeon Silver 4210R (40 cores, 128GB RAM), 2x RTX 3090
- Software: gem5 v22.1, McPAT v1.3, HotSpot v7.0, PyTorch 2.3, BoTorch 0.11.1
- Tech Node: 22nm-90nm (heterogeneous integration)
- Benchmarks: NPB, SPEC06, MatMul (128x128 to 512x512)
- Baselines: XGBoost, Random Forest, SVR, SoC-Tuner, BOOM-Explorer, Regression

Evaluation Metrics:

- Simulation: Metric accuracy, evaluation time breakdown
- Exploration: Pareto frontier quality (HV), design coverage
- Practical: DSE efficiency, resource savings

Result 1: ChipletSIM Comprehensive Evaluation

Power & Area Evaluation:

- Supports heterogeneous tech nodes (22nm-90nm)
- Accurate power modeling for mixed-node systems

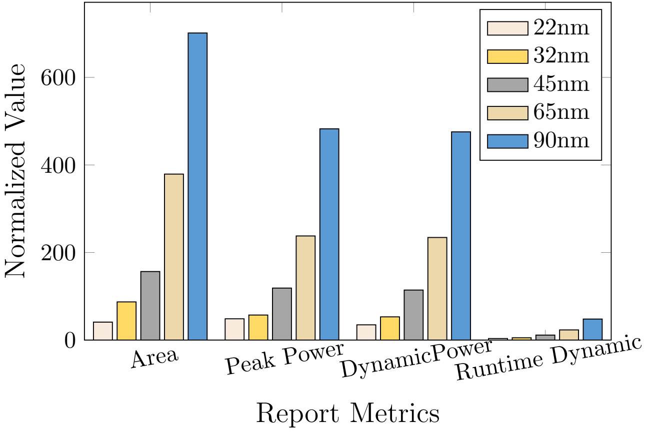


Figure 9: Heterogeneous Chiplet Power/Area

Result 1: ChipletSIM Comprehensive Evaluation

Simulation Time Breakdown (16-core, 8-cache chiplet system):

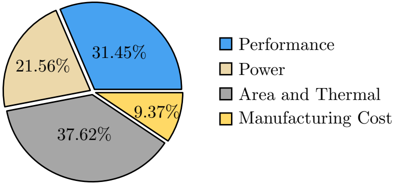


Figure 8: Evaluation Time Distribution

Cost Composition:

- CPU Core: 27.74%, L2 Cache: 32.87%
- Interposer: 11.8%, Defects/Bonding: 27.58%

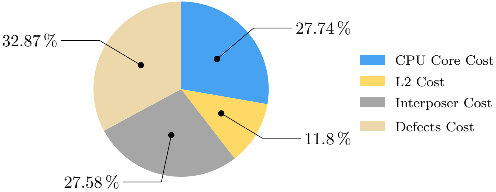


Figure 10: Chiplet System Cost Breakdown

Result 2: ChipletDSE Exploration Quality

Normalized Hypervolume (Higher = Better):

Method	RF	XGBoost	SVR	Regression	BOOM-Explorer	SoC-Tuner	**CHASE**
HV	0.62	0.68	0.59	0.71	0.75	0.78	**0.92**

Table: Hypervolume Comparison Under Same Budget

Pareto Frontier Comparison (Budget = 2000):

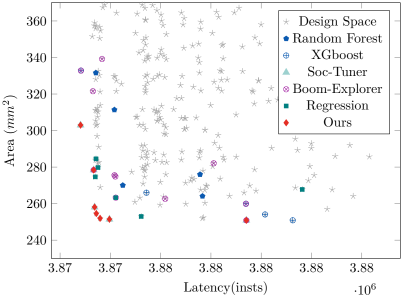


Figure 11: Latency vs. Area Pareto Sets

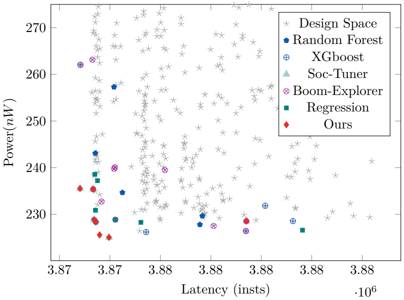


Figure 12: Latency vs. Power Pareto Sets

Conclusion

Summary of CHASE Framework

Core Contributions:

- **Comprehensive Simulation:** First framework unifying performance (latency/power) and manufacturing (area/thermal/cost) metrics for chiplets.
- **Efficient Exploration:** Decoupled multi-fidelity optimization maximizes design coverage under fixed budget (3.17x more designs evaluated).
- **Practical Impact:** Pareto frontier-driven design provides actionable trade-offs for real-world chiplet systems (19.4% area reduction, 1.7 C thermal improvement).

Key Takeaways:

- CHASE bridges the gap between chiplet architecture design and manufacturing constraints.
- Decoupled evaluation is critical for large design spaces (avoids wasting resources on bad designs).
- Future chiplet design will rely on unified frameworks like CHASE to balance performance, cost, and reliability.

THANK YOU!