# Bayesian Optimization for Chiplet Placement on 2.5D System in Package Platforms: Tackling Rotation and Constraint Challenges

**Furen Zhuang [a], Partha Pratim Kundu [a], Sezin Kircali Ata[a], Yubo Hou [a], Mihai Dragos Rotaru [b], Rahul Dutta[b], Ashish James[a]**

[a] *Institute for Infocomm Research (I2R), A*STAR, Singapore, zhuangfr@i2r.a-star.edu.sg*

[b] *Institute for Microelectronics (IME), A*STAR, Singapore*

## 1. Abstract

In System-in-Package (SiP) design, current Bayesian optimization (BO)-based floor-planning methods fail to address chiplet rotation and infeasibly large placements, leading to inefficiencies in performance and optimization. Our work introduces a novel rotation- and constraint-aware BO approach for multi-objective chiplet placement, and effectively overcomes these limitations for enhanced performance.

## 2. Introduction

In SiP design, floor-planning is critical to optimizing performance, power consumption, and thermal management of 2.5D and 3D packaging systems. The primary goal is to minimize total wirelength between interconnects while adhering to temperature and physical layout constraints. Reducing wirelength lowers system latency and power consumption, enhancing overall chip performance. Additionally, placements must ensure peak temperatures remain below a designed threshold, typically 85˚C, to maintain proper chiplet functionality. The layout must also be feasible - chiplets cannot overlap nor exceed the interposer's boundaries.

### *2.1 Related work*

State-of-the-art placement methods often rely on Simulated Annealing (SA), which requires numerous iterations and costly thermal simulations, making it impractical for SiP design. Bayesian Optimization (BO) has emerged as a promising alternative for SiP floor-planning, efficiently balancing exploration and exploitation to achieve optimal placement with fewer simulation steps.

Current BO-based placement methods optimize a sequence pair -- a pair of sequences denoting the relative positions of each chiplet in two oblique directions, which specify their final positions in a placement. A position or Mallows kernel is typically used to assess the similarity between two sequence pairs and then predict the expected improvement (EI) for a new candidate sequence pair.

While some works address variable aspect ratios of macros [1], this approach is not always practical in SiP design and allowing chiplet rotation presents a more realistic problem. However, the discrete nature of rotations complicates the optimization process. In summary, current approaches suffer from two key limitations: 1) they cannot account for rotation, drastically reducing placement performance, and 2) they lack infeasibility prediction, leading to high computational costs as each candidate sequence pair must be realized into actual placements during optimization.

## 3. Proposed Method

Our method addresses these issues by first, deriving a comprehensive kernel that integrates the sequence-aware position kernel and the rotation-aware Hamming kernel, enabling our surrogate model to efficiently search both the sequence pair and rotation spaces:

$$\begin{aligned} K([x,y,r],[x',y',r']) &= \gamma k_{pos}(x,x')k_{pos}(y,y')k_{ham}(r,r') \\ &+ (1-\gamma)[k_{pos}(x,x')k_{pos}(y,y') \\ &+ k_{ham}(r,r')] \end{aligned}$$

$$k_{pos}(\pi,\pi') = \exp\left(-\frac{1}{N}\sum_n w_n\left|\pi^{-1}(n) - \pi'^{-1}(n)\right|\right)$$

$$k_{ham}(r,r') = \exp\left(-\frac{1}{N}\sum_n r(n) \neq r'(n)\right)$$

Second, we introduced a model that predicts the longest placement dimension to estimate the feasibility probability, ensuring the placement stays within canvas limits. This probability is used to weigh the raw acquisition function and derive the actual EI [5].

$$EI_C(x,y,r) = P(\max dim < intp\ size|x,y,r)EI_{UC}(x,y,r)$$

The subscripts C and UC refer to constrained and unconstrained respectively. Consequently, infeasible sequence pairs result in low EI and are not realized into placements, leading to more efficient optimizations.

We ran experiments on two designs from [2], with network topologies [4] shown in Figs 1 and 2. Comparison results are shown in Figs 3 and 4, and Tables 1 and 2. The objective loss shown is from [2] which balances the tradeoff between temperature and wirelength:

$$Loss = \alpha\frac{T - T_{min}}{T_{max} - T_{min}} + (1-\alpha)\frac{WL - WL_{min}}{WL_{max} - WL_{min}}$$

$$\alpha = \begin{cases} \min\{0.01T - 0.35, 0.9\}, if\ T > 85 \\ 0, otherwise. \end{cases}$$

Our method outperforms state-of-the-art benchmarks SA [2] and BO [3], achieving the best objective value across all methods run for 150 steps. For the multi-GPU environment, our method achieves a loss of 0.216, which is 62.6% and 26.0% lower than SA and BO respectively. For the Ascend910 environment, our method achieves a loss of 0.039, which is 92.5% and 79.1% lower than SA and BO respectively.

# Bayesian Optimization for Chiplet Placement on 2.5D System in Package Platforms: Tackling Rotation and Constraint Challenges

**Furen Zhuang** [a], **Partha Pratim Kundu** [a], **Sezin Kircali Ata**[a], **Yubo Hou** [a], **Mihai Dragos Rotaru** [b], **Rahul Dutta**[b], **Ashish James**[a]

[a] *Institute for Infocomm Research (I2R), A*STAR, Singapore, zhuangfr@i2r.a-star.edu.sg*

[b] *Institute for Microelectronics (IME), A*STAR, Singapore*

Although our method requires slightly more steps to converge in the multi-GPU environment compared to BO, this is expected due to the larger search space that includes rotation. Additionally, we observed that BO suffers from plateauing, where the objective stagnates for many steps. This occurs because BO selects infeasible candidate sequence pairs, which do not result in improvements. Its raw acquisition function struggles to identify the best feasible candidates, whereas our method, using an enhanced acquisition function, efficiently focuses on the solution space. This enables more effective optimization and faster convergence.

## 4. Conclusion

In summary, the proposed rotation-aware approach enhances the thermal and wirelength performance of chiplet placements. Additionally, it leverages a constraint-aware optimization method that enables rapid discovery of optimal placements, even within a larger search space. Overall, the method efficiently yields placements with higher performance, lower power consumption, and superior thermal management.

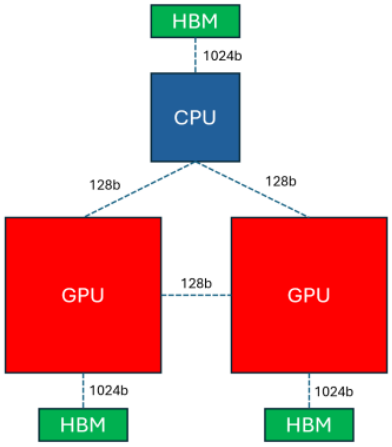


Fig. 1: Network topology of multi-GPU system [4].

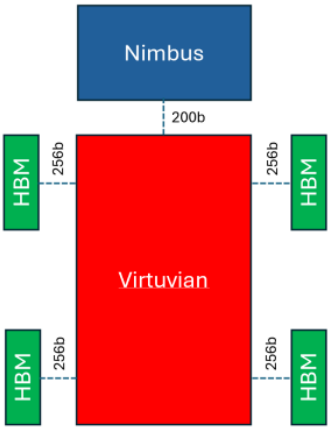


Fig. 2: Network topology of Huawei Ascend910 system [4].

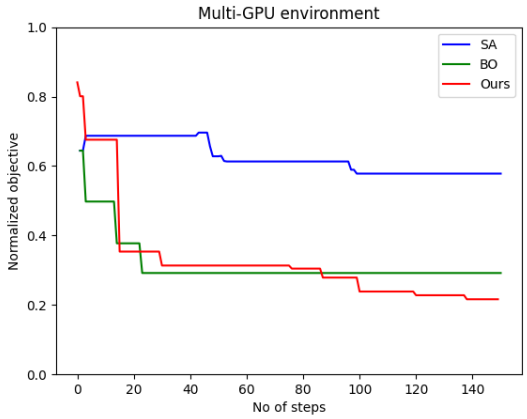


Fig. 3: Comparison of convergence in Multi-GPU environment on normalized objective loss.

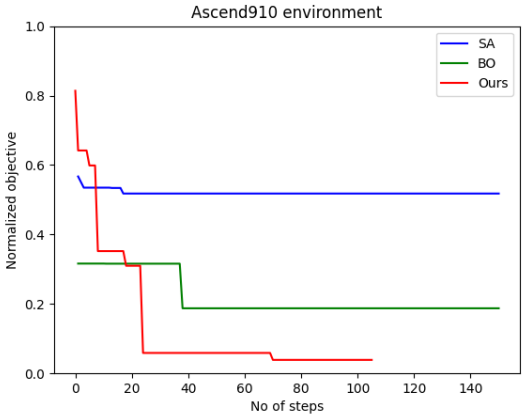


Fig. 4: Comparison of convergence in Ascend910 environment on normalized objective loss.

Table 1: Comparison of our method against state-of-the-art benchmarks on multi-GPU system.

| Multi-GPU | Step | T (°C) | WL (mm) | Objective |
|---|---|---|---|---|
| SA [2] | 99 | 96.3 | 178509 | 0.578 |
| BO [3] | 23 | 93.2 | 95421 | 0.292 |
| **BO (ours)** | **138** | **91.2** | **111068** | **0.216** |

Table 2: Comparison of our method against state-of-the-art benchmarks on Ascend910 system.

| Multi-GPU | Step | T (°C) | WL (mm) | Objective |
|---|---|---|---|---|
| SA | 17 | 74.1 | 33618 | 0.518 |
| BO | 38 | 74.8 | 23102 | 0.187 |
| **BO (ours)** | **70** | **75.0** | **18373** | **0.039** |

## Acknowledgments

This research is supported by the Agency for Science, Technology and Research (A*STAR) under its MTC Programmatic Funds (Grant No. M23M3b0064).

# Bayesian Optimization for Chiplet Placement on 2.5D System in Package Platforms: Tackling Rotation and Constraint Challenges

**Furen Zhuang [a], Partha Pratim Kundu [a], Sezin Kircali Ata[a], Yubo Hou [a], Mihai Dragos Rotaru [b], Rahul Dutta[b], Ashish James[a]**

[a] *Institute for Infocomm Research (I2R), A*STAR, Singapore, zhuangfr@i2r.a-star.edu.sg*

[b] *Institute for Microelectronics (IME), A*STAR, Singapore*

## References

[1] M. Molter et al., "Thermal-Aware SoC Macro Placement and Multi-chip Module Design Optimization with Bayesian Optimization," ECTC 2023.

[2] Y. Ma et al, "TAP-2.5D: A Thermally Aware Chiplet Placement Methodology for 2.5D Systems," DATE 2021.

[3] C. Oh et al., "Bayesian optimization for macro placement," ICML Workshop, 2022.

[4] F. Zhuang et al., "Efficient Thermal-Aware Floor-planning with Bayesian Optimization: A Simulation-Efficient Approach," in Proc. IEEE 26th Electronics Packaging Technology Conference (EPTC 2024), Singapore, 2024, pp. 1-4.

[5] Gardner, Jacob R., et al. "Bayesian optimization with inequality constraints." ICML. Vol. 2014. 2014.