



HAL
open science

Machine Learning Approaches for Application Mapping Optimization in Network-on-Chip Architectures

Zainab Ghrayeb, Sébastien Le Nours, Christophe Moy, Jordane Lorandel, Christine Sinoquet

► **To cite this version:**

Zainab Ghrayeb, Sébastien Le Nours, Christophe Moy, Jordane Lorandel, Christine Sinoquet. Machine Learning Approaches for Application Mapping Optimization in Network-on-Chip Architectures. Colloque du GDR SOC2, Jun 2025, Lorient, France. pp.sciencesconf.org:gdr-soc2-2025:641241, 2025. <hal-05069557>

HAL Id: hal-05069557

<https://hal.science/hal-05069557v1>

Submitted on 10 Jun 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Context and Objective

Real Applications Require Memory-Aware Mapping

AI models (e.g., CNNs, Transformers) involve:

- Millions or billions of parameters and large intermediate activations.
- **Shared memory** access is a common requirement for tasks

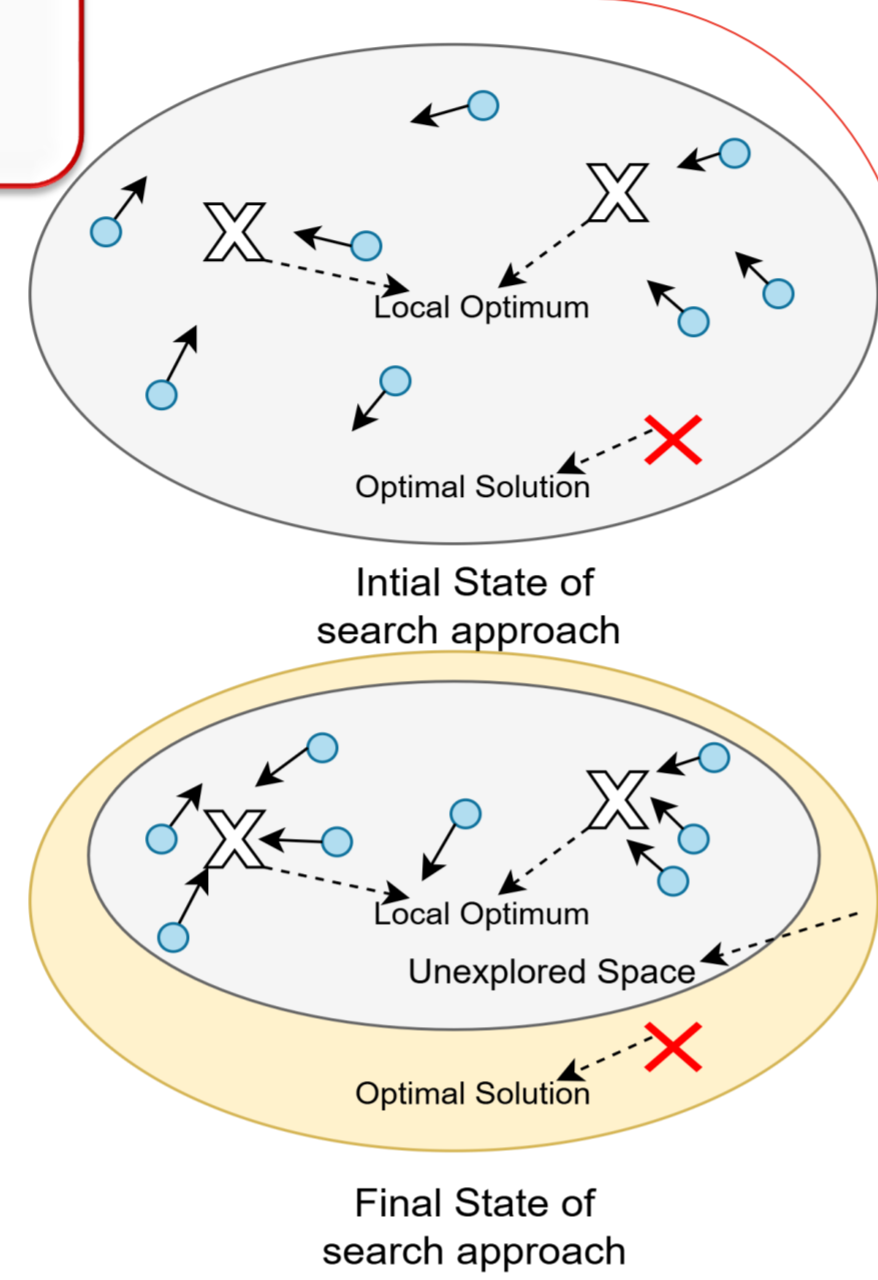
Aim of this work

Study the optimization process of application mapping on network on chip (NoC), including the memory hierarchy, using machine learning algorithms (ML).

Traditional Methods Limitation

Traditional methods generate initial mappings **randomly** thus:

The initial population is poor, converges to a sub-optimal solution, and have poor generalization



Traditional methods:

• Transformative heuristic methods:

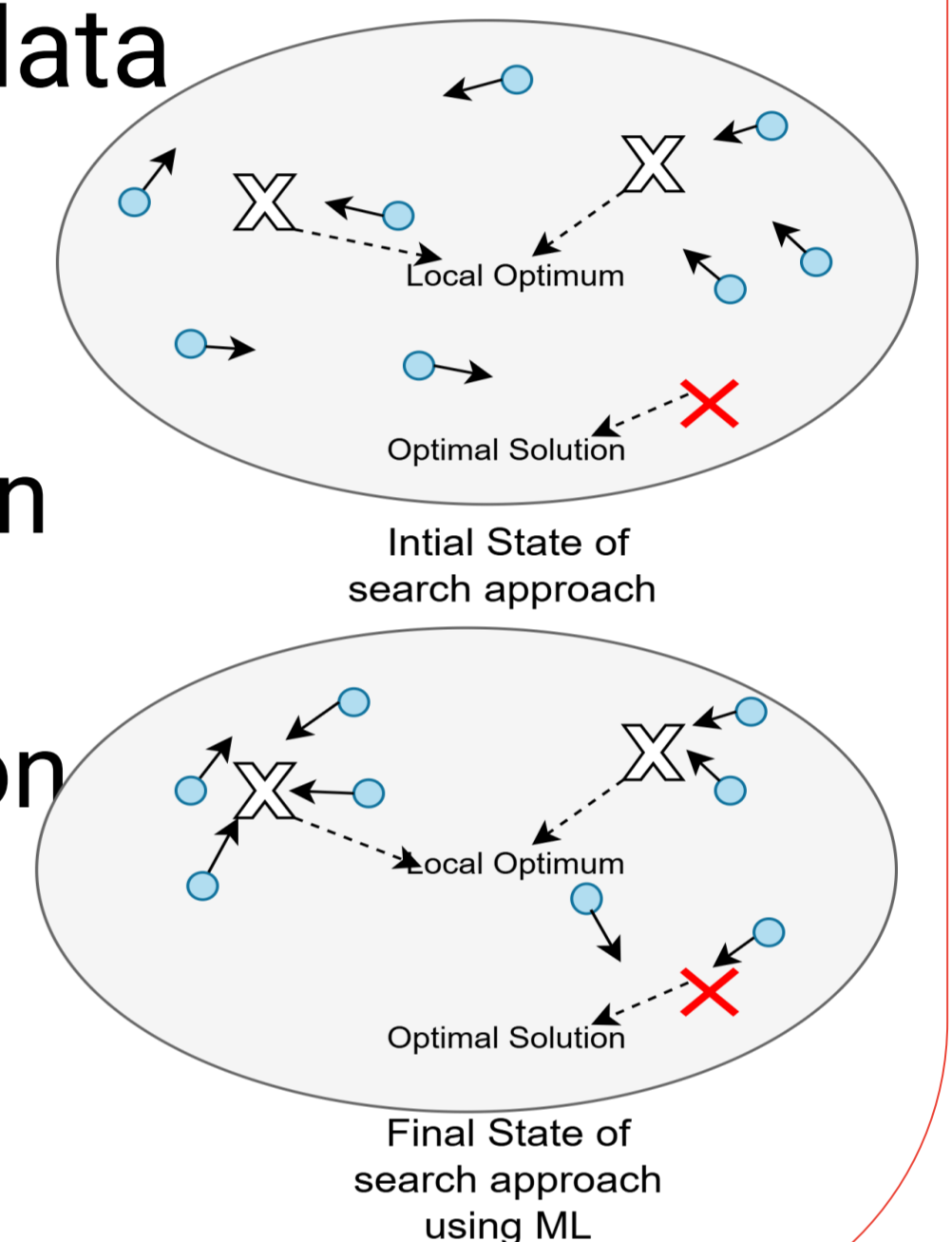
Genetic Algorithm(GA), Discrete Particle Swarm Optimization (DPSO)

• Constructive heuristic methods:

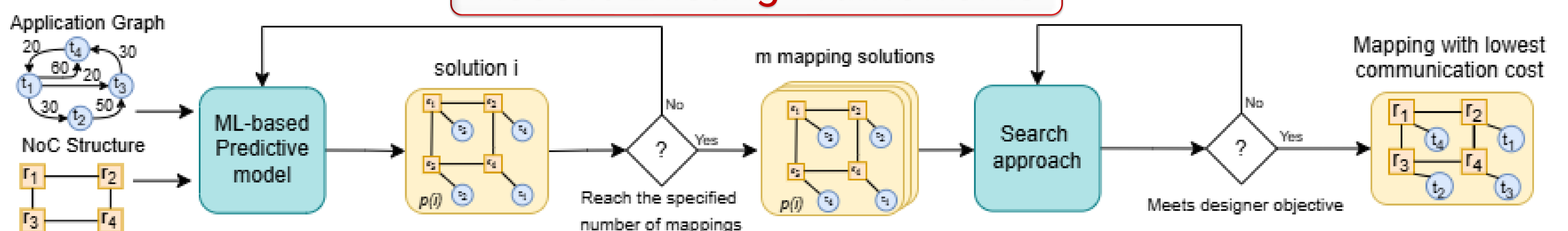
Greedy random adaptive search approach(GRASP), Simulated annealing(SA)

Potential Benefits of ML

- Enables computers to **learn from data** without being explicitly programmed.
- Uses **algorithms** to find patterns in data and handle large-scale mapping
- ML is able to:
 - Scalable for large graphs and unseen nodes.
 - Converge toward the optimal solution
 - Reduce exploration time
 - Adapt to topology changes



Recent Existing Frameworks



Generated by TGFF graph tool

Each solution is predicted from ML predictive probability model like MPNN-Ptr network. Model trained with reinforcement Learning

Take 80% of mapping with high probability and apply different search approach (e.g. GA, PSO, active search)

References:

- Chen, Qingkun, et al. "A reinforcement learning-based framework for solving the IP mapping problem." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 29.9 (2021): 1638-1651.
- Sambangi, Ramesh, et al. "Application mapping onto manycore processor architectures using active search framework." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 31.6 (2023): 789-801.
- Chen, Qingkun, et al. "An IP core mapping algorithm based on neural networks." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 29.1 (2020): 189-202.

Extension of the Existing Frameworks

The mapping of the application on the NoC directly affects:

- Communication cost (total data transfer).
- Bandwidth usage (ensuring no congestion in the network).
- Energy efficiency (minimizing power consumption in communication).

