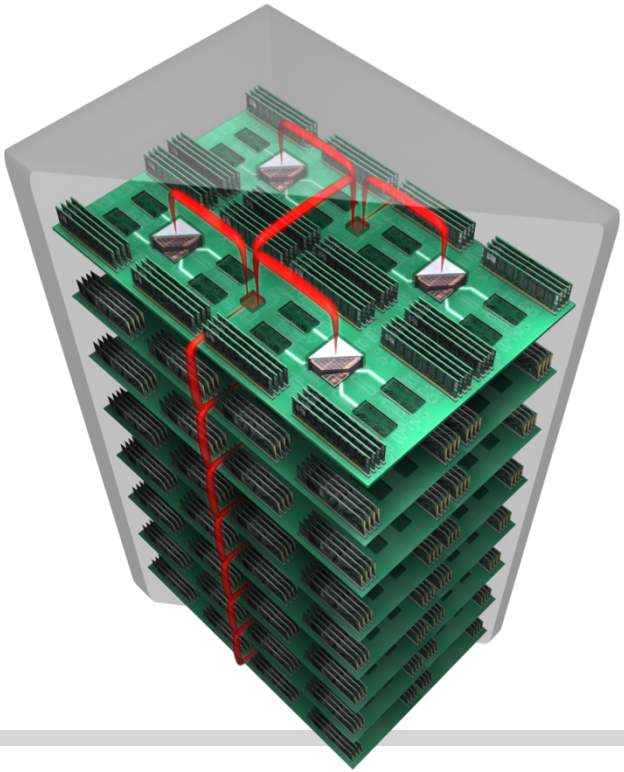
# Modeling Silicon Photonic Networks Using Cache Models

ASCR ModSim Workshop, 2014

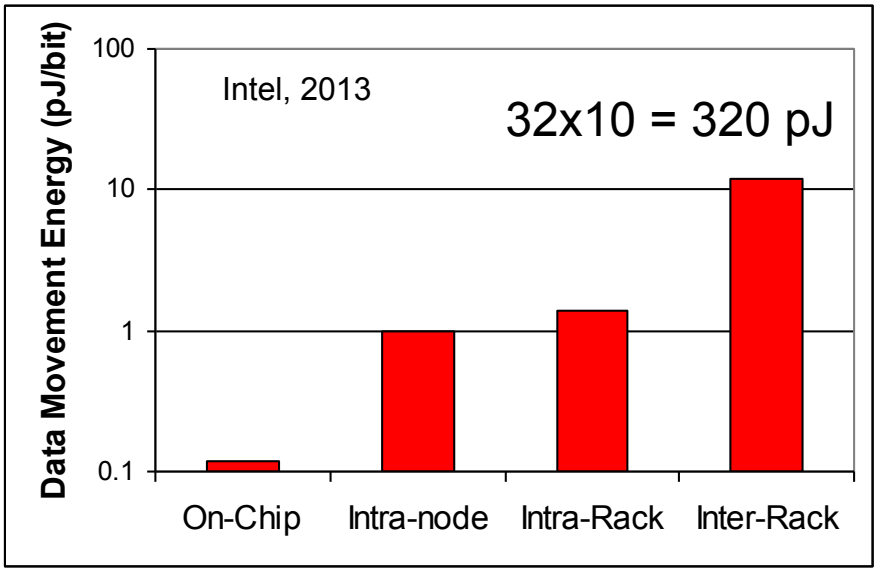
Ke Wen, Sébastien Rumley and Keren Bergman
*Lightwave Research Lab, Columbia University*

Jeremiah J. Wilke
*Sandia National Laboratories, Livermore, CA*

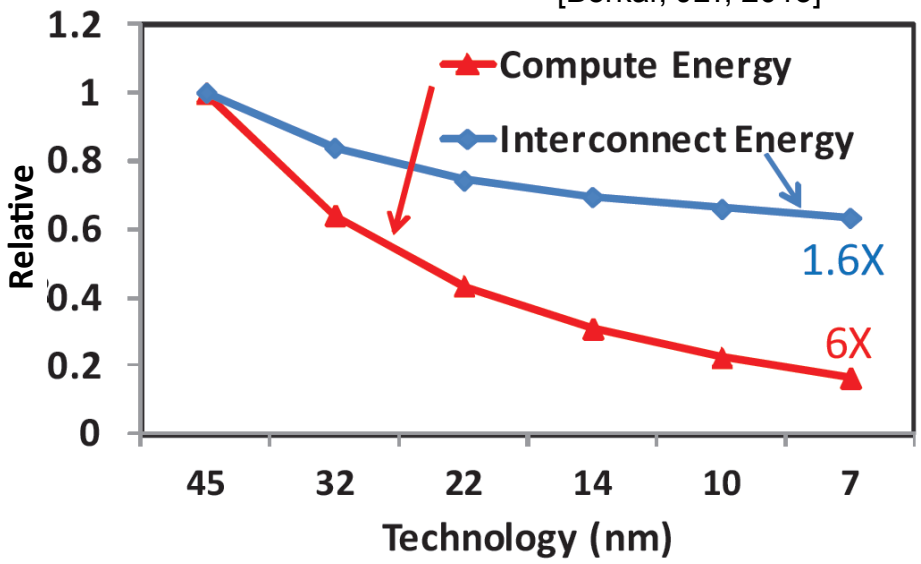


# Data movement is becoming a major challenge

- Dictates power consumption
  - 32-bit ADD: **0.5 pJ** for ALU while **4.5 pJ** for instruction & data fetching along with result storage (on chip, 2010 technology [Balfour]).
  - 20 MW/1 ExaFLOPs = **20 pJ/flop**

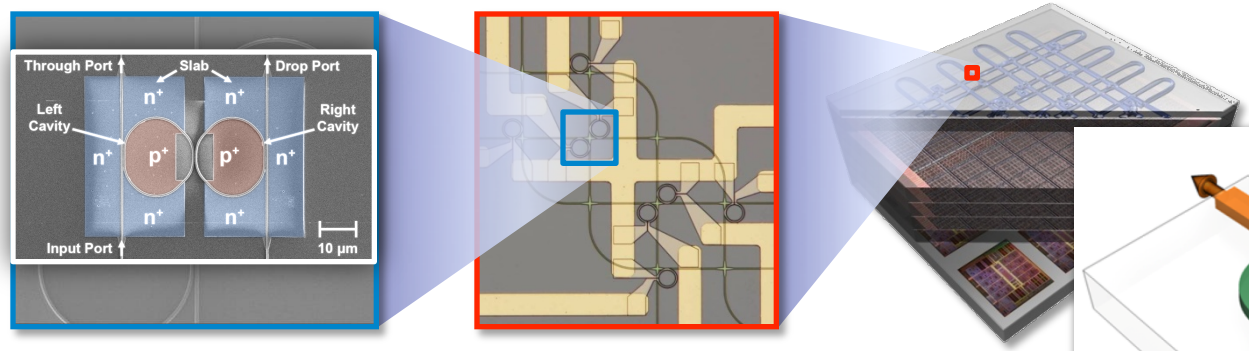


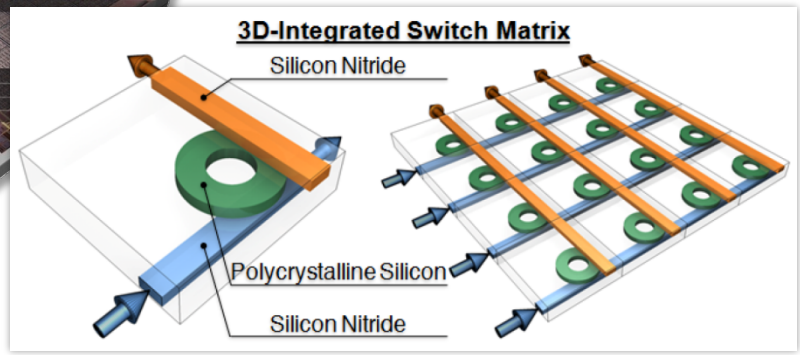

[Borkar, JLT, 2013]



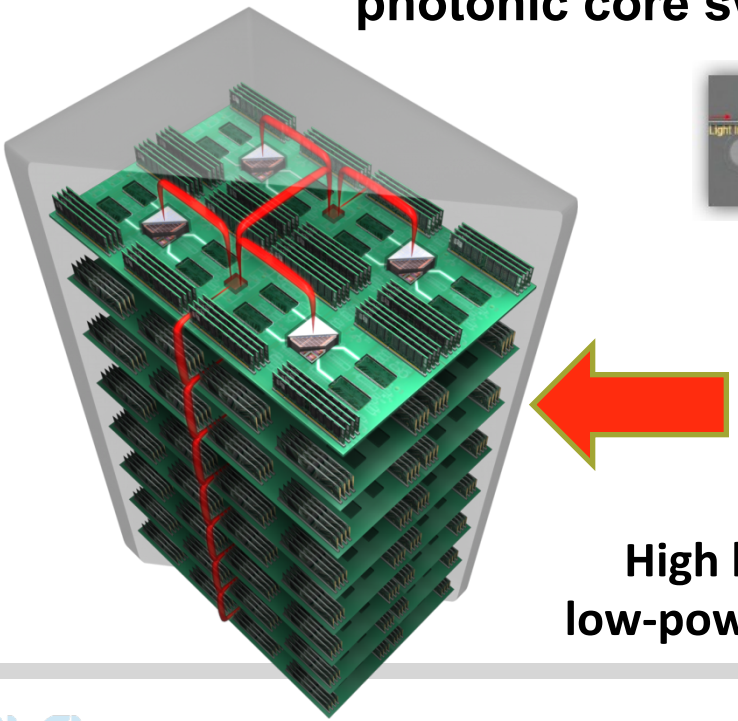

- Limits compute efficiency

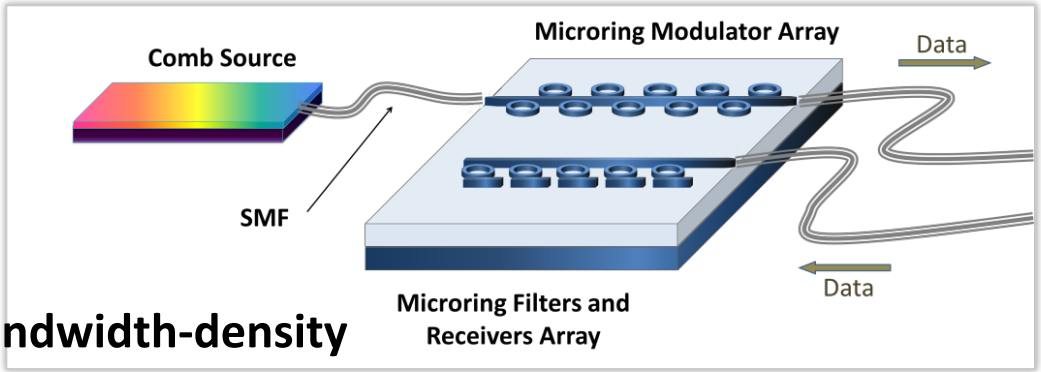
# Systems Impact with Si-photonic interconnected Data Movement




**High radix fast-reconfigurable photonic core switches**

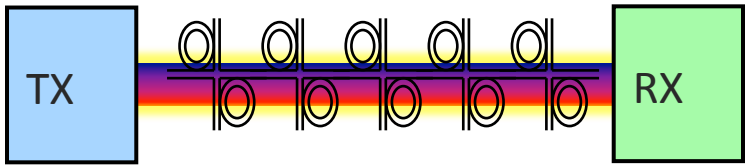


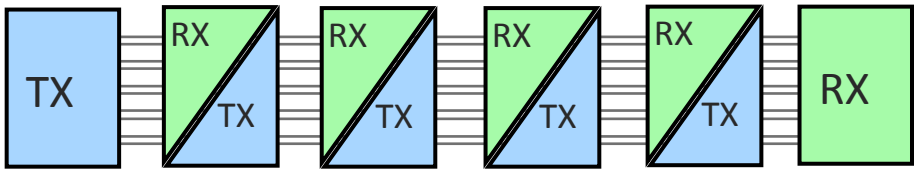




**High bandwidth-density low-power WDM Transceivers**


3

## Change the Rules for Bandwidth-per-Watt




### PHOTONICS

- Modulate/receive data once
- *Wavelength Parallelism :*
  - Broadband switch routes multi-wavelength stream
- Distance Independence
  - Off-chip BW ≈ on-chip BW for nearly same power

### ELECTRONICS

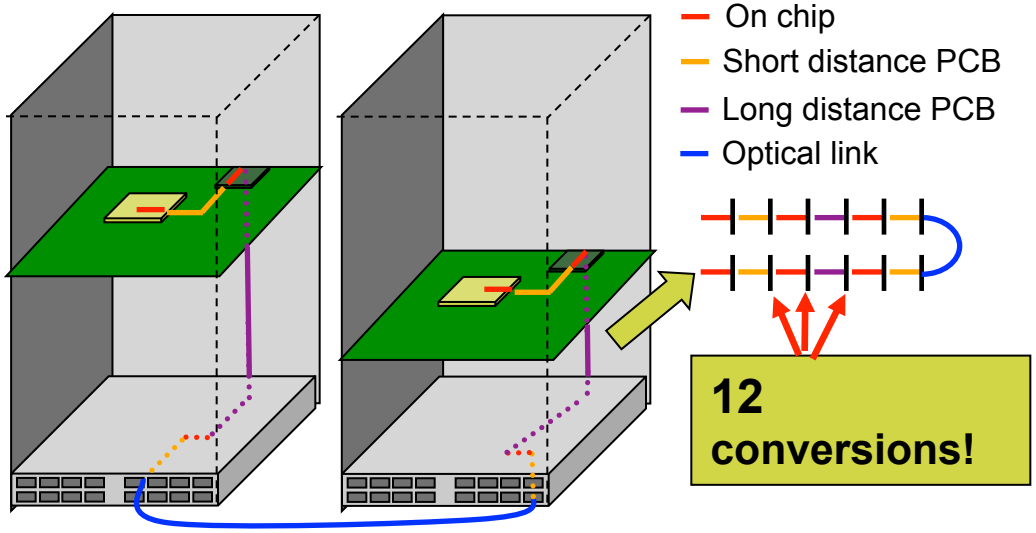
- Buffer, receive, and re-transmit at every router
- *Space Parallelism :*
  - Each bus lane routed independently ($P \propto N_{LANES}$)
- Off-chip BW requires much more power than on-chip BW

**only by rethinking how to leverage its unique data-movement capabilities to realize new system architectures**

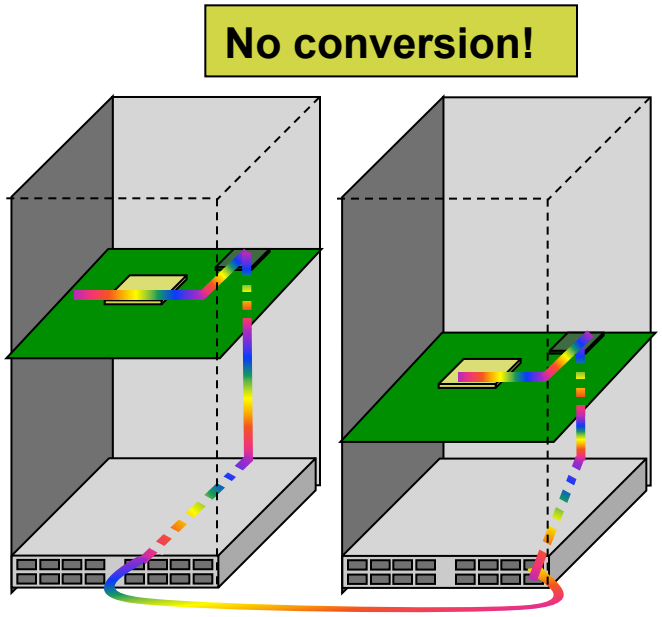

4

# Optical Data Movement Beyond Wire Replacement

- Optics-enabled system architecture transformations:
  - distance-independent, cut-through, bufferless




Conventional inter-node data-movement

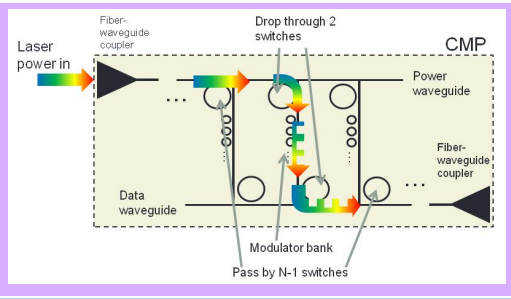



**Fully flattened optical data-movement**

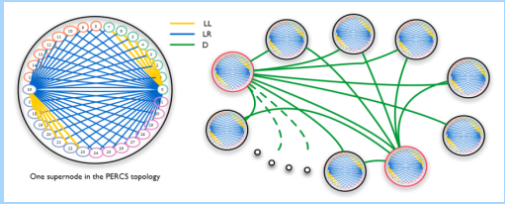
# PhoenixSim: Multi-Level Design

## Design, Modeling and Simulation Environment

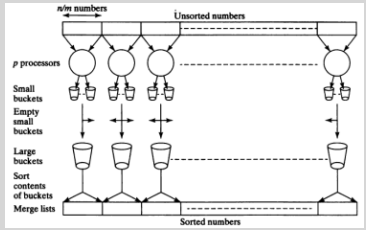
- **Physical link layer:**
  - SiP components modeling
  - Link bandwidth maximization
  - Optical power budget validation




- **Network layer**
  - Optical data flow, switching, routing protocols
  - Network performance analysis

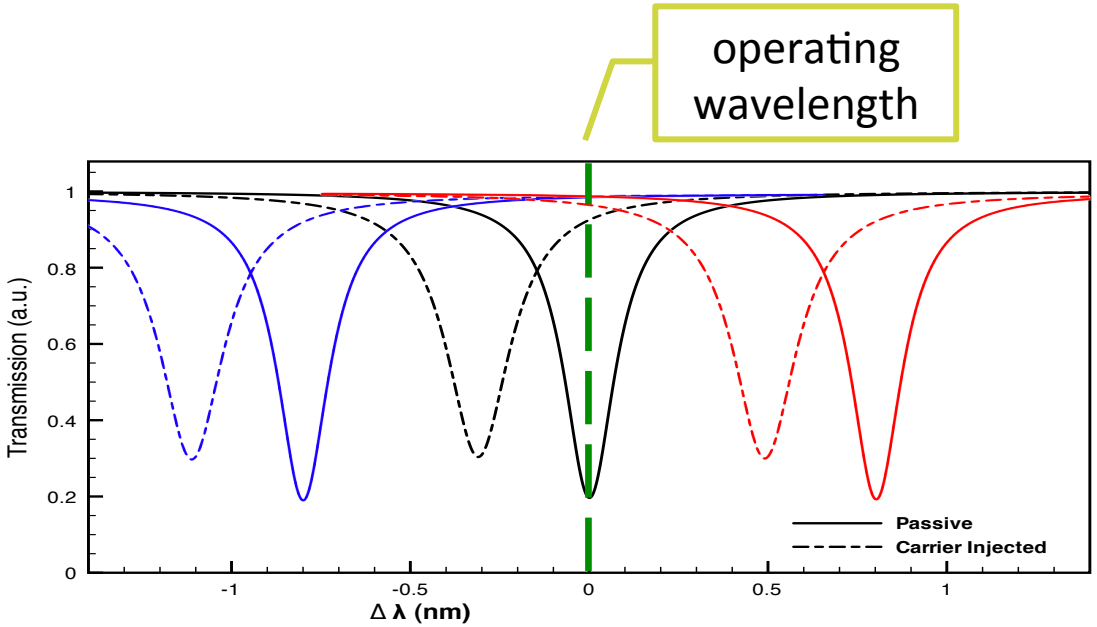



- **Application layer**
  - Intelligent management of photonic resources based on application characteristics
  - Optically enabled algorithm re-design

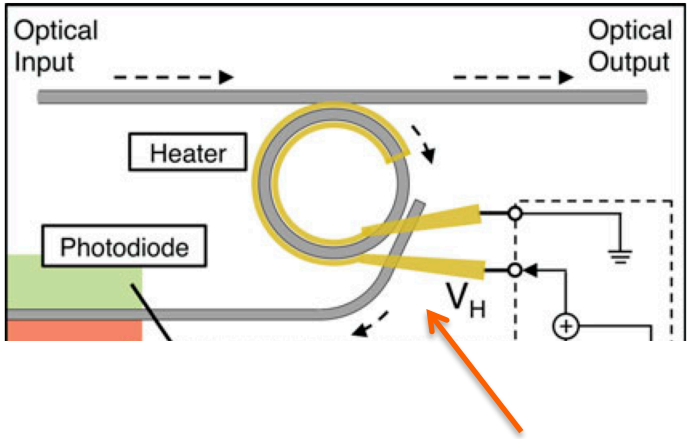


# Challenge: Toll to pay before using SiP links

- Optics relies on circuit switching → path setup cost
- Microrings are sensitive to temperature → Need to thermally (re)-initialize microrings to work on correct wavelengths



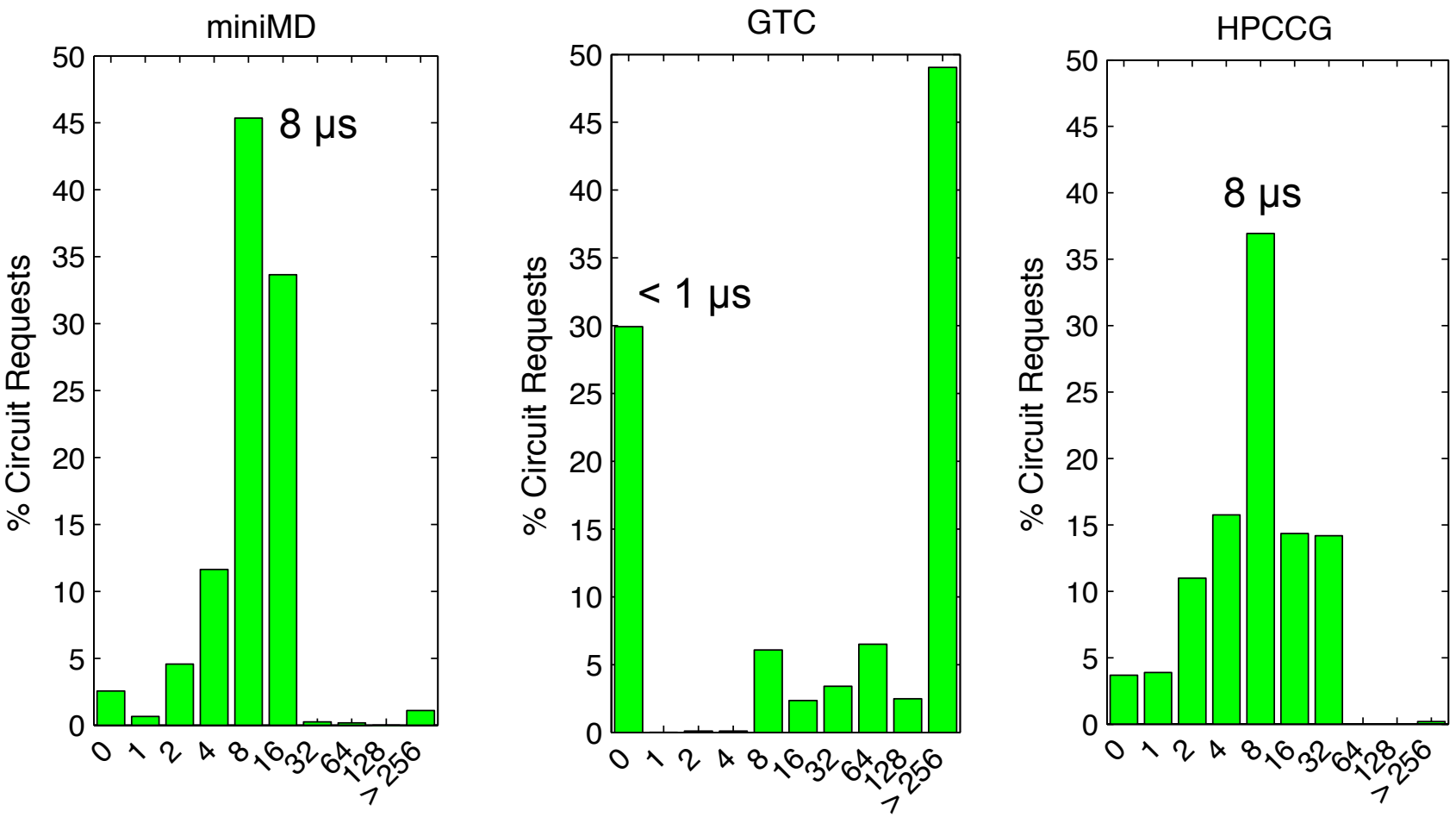

(Re)-initialization latency:
**~ 10 µs**




Heater

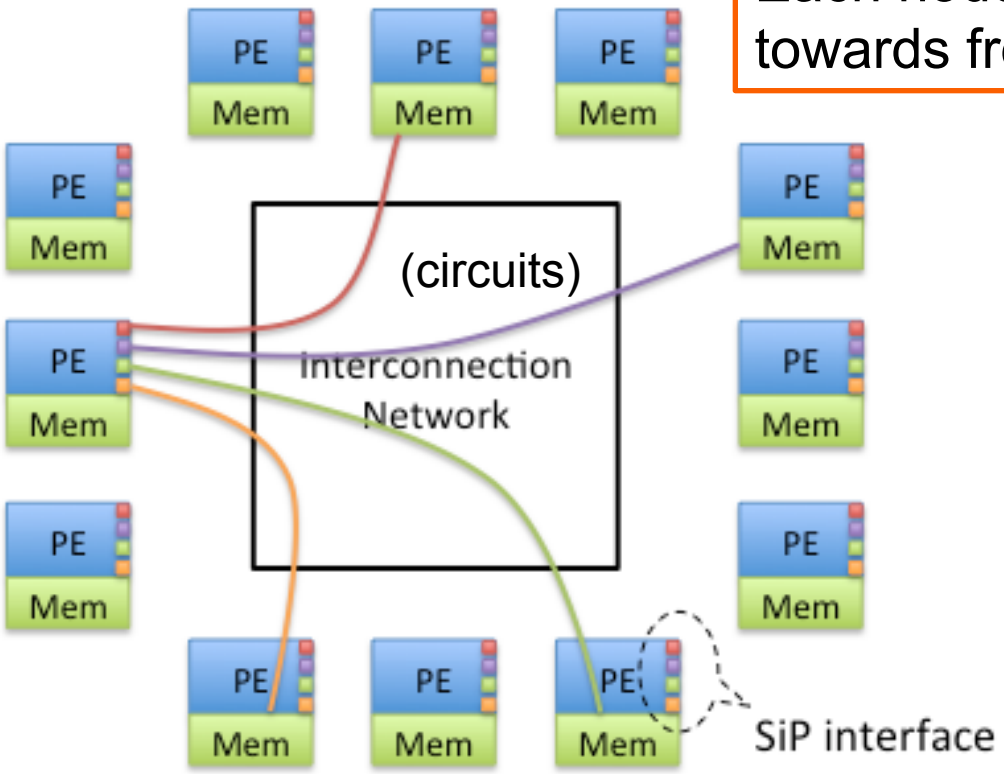
**How can we avoid the toll?**

# Time Elapsed Before Reusing a Circuit




**Confirm the need for maintaining a circuit after uses**

# Circuit Maintaining Architecture: Cache-Like Performance Model




Each node maintains a set of circuits towards frequently accessed destinations.

- Similar to cache:
- Circuit hit
  → time of flight
- Circuit miss
  → penalty from setup + initialization
- Need intelligent policies for circuit replacement

**Modeling circuit management as modeling caches**

# Design Questions to Answer

**by Simulation & Modeling**

1. How many circuits do we need to achieve X% hit rate?
2. What replacement strategies are best for which applications?
3. What is the relationship between performance and cost?
4. What is the relationship between performance and energy consumption?
5. Is there room for improvement by exposing circuit management to the programming model?
6. How does application mapping change the reuse pattern? How does it interact with the circuit management scheme?

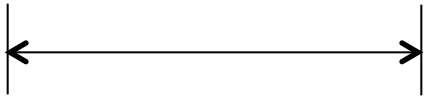
# Inspired by cache modeling

- ***Reuse Distance*** captures the number of circuits used between two consecutive accesses to the same circuit.

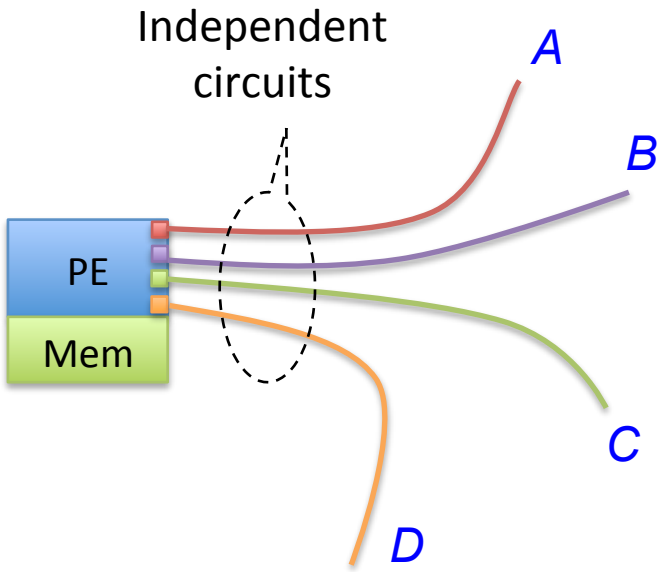
- Example:

  Circuits used by a node in order:

  *C, A, B, D, B, C*

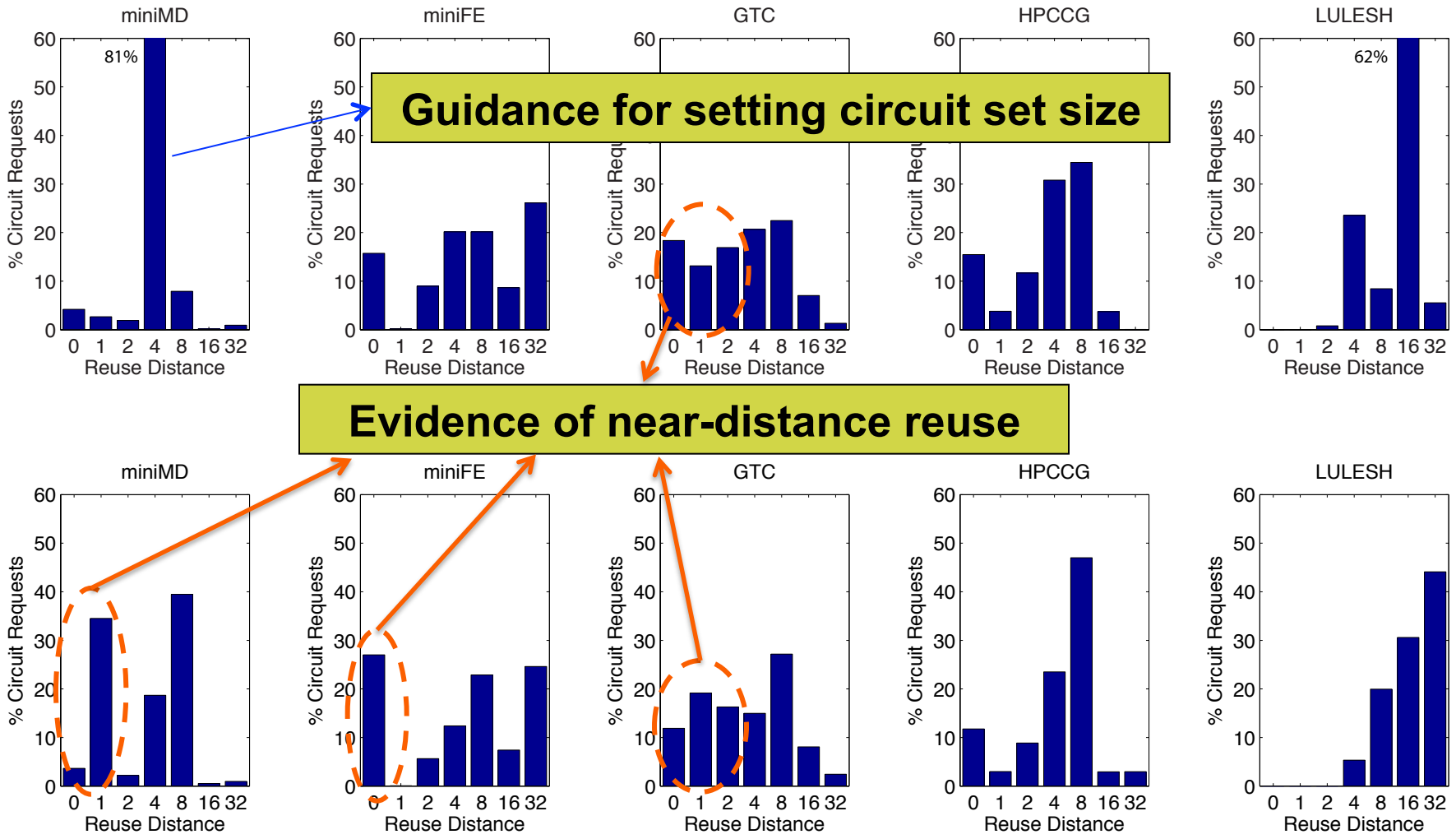


  Reuse Distance = 4 for circuit *C*




**Strategy: optimize circuit management (setup/maintain/ tear down) based on observed reuse distances**

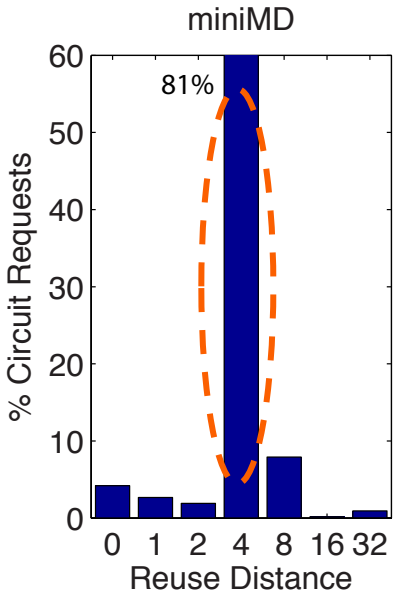
# Distribution of Reuse Distances
## (top: 64 nodes, bottom: 256 nodes)



## Observation is not enough, prediction is what matters

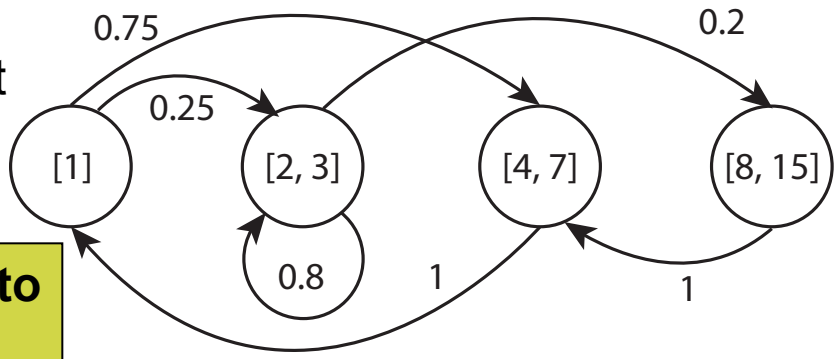
- ***Goal***: optimize circuit replacements to maximize *circuit hit rate*
- ***Idea***: replace the circuit to be reused in the farthest future → prediction needed for next reuse distance of replacement candidates

1. Maximum Likelihood (ML) prediction

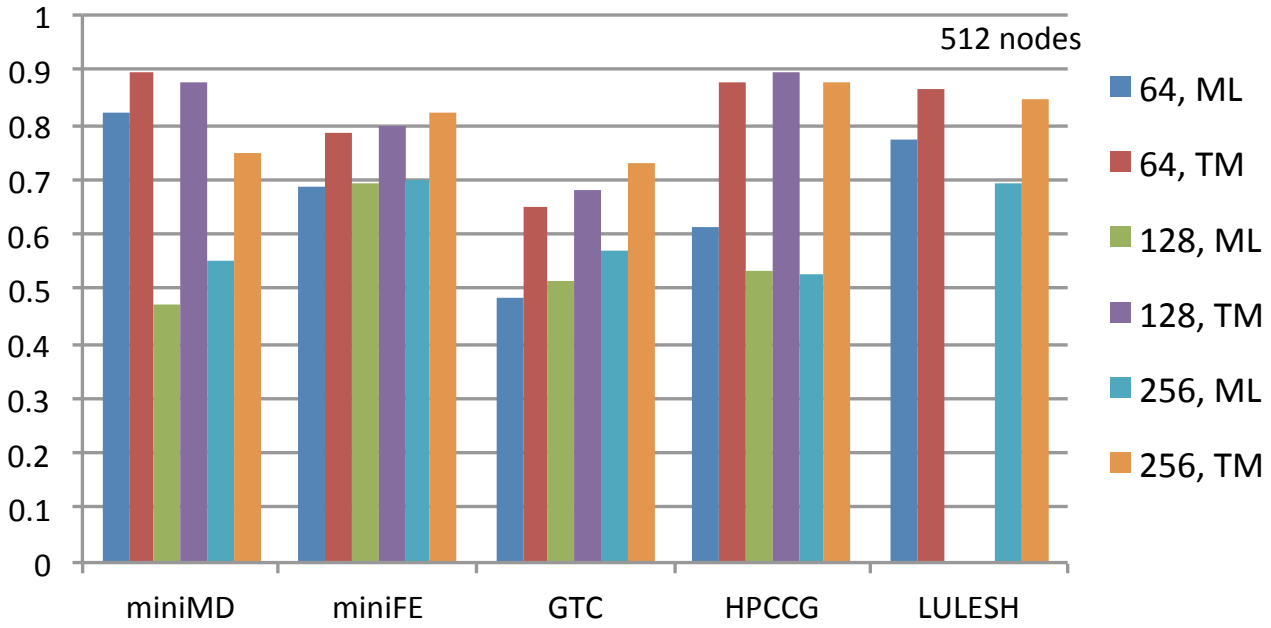



2. Transition Matrix (TM) based prediction

Reuse distance sequence of a circuit:
9 6 1 6 1 6 1 2 2 2 2 2 9 6 1 6 . . .

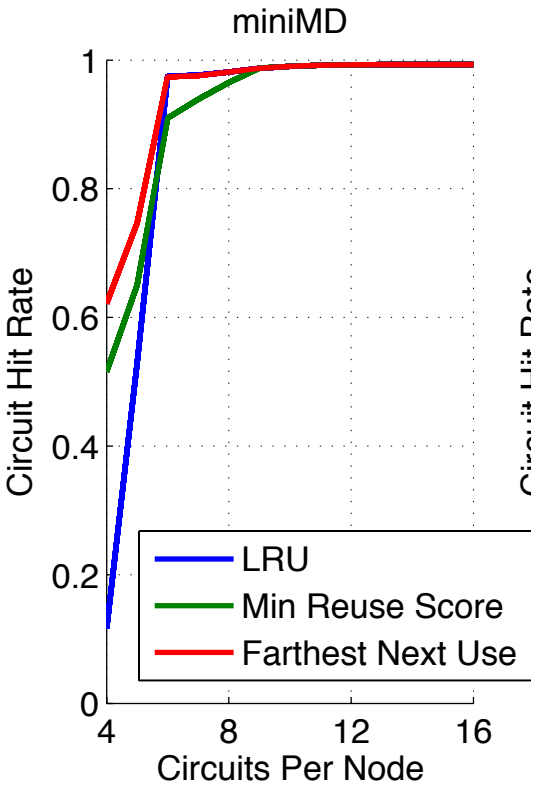


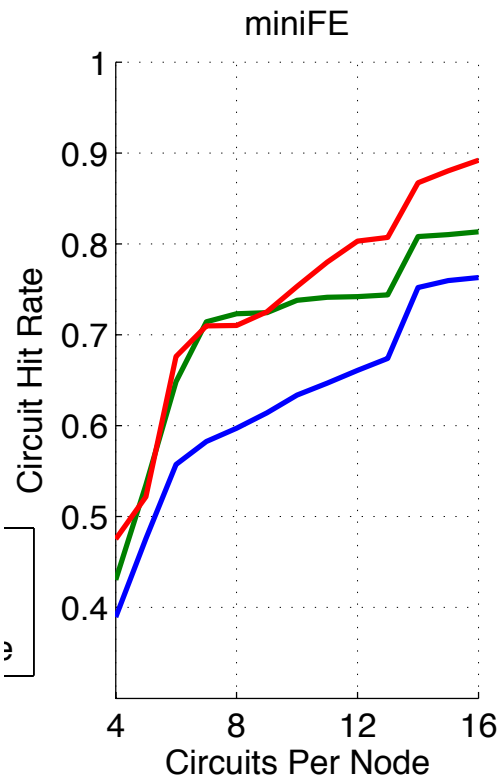

**And many more to explore…**

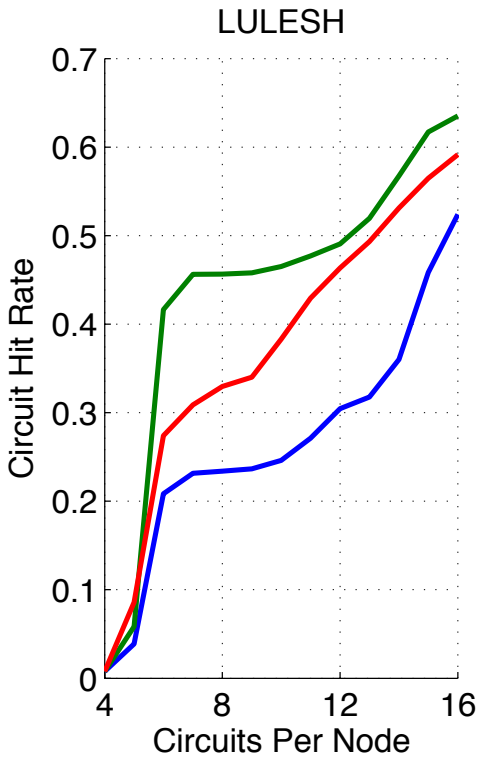


# Simulation based on *SST Macroscale* components

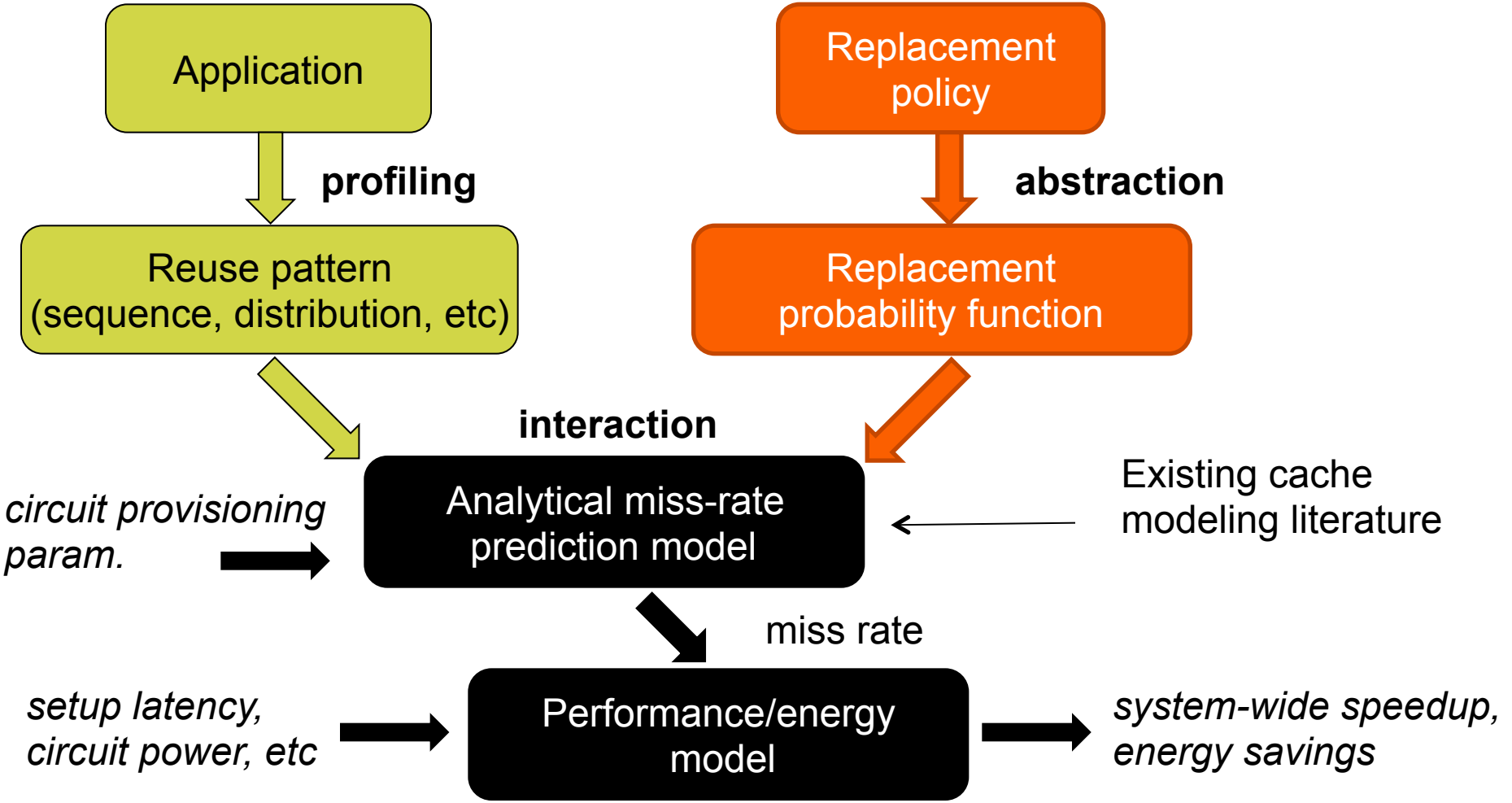
- Co-simulate with apps, leverage app skeletons for reduced sim time
- Implement different prediction and replacement strategies in *SST Macroscale* components
- Simulation should be scalable to 100,000 end points



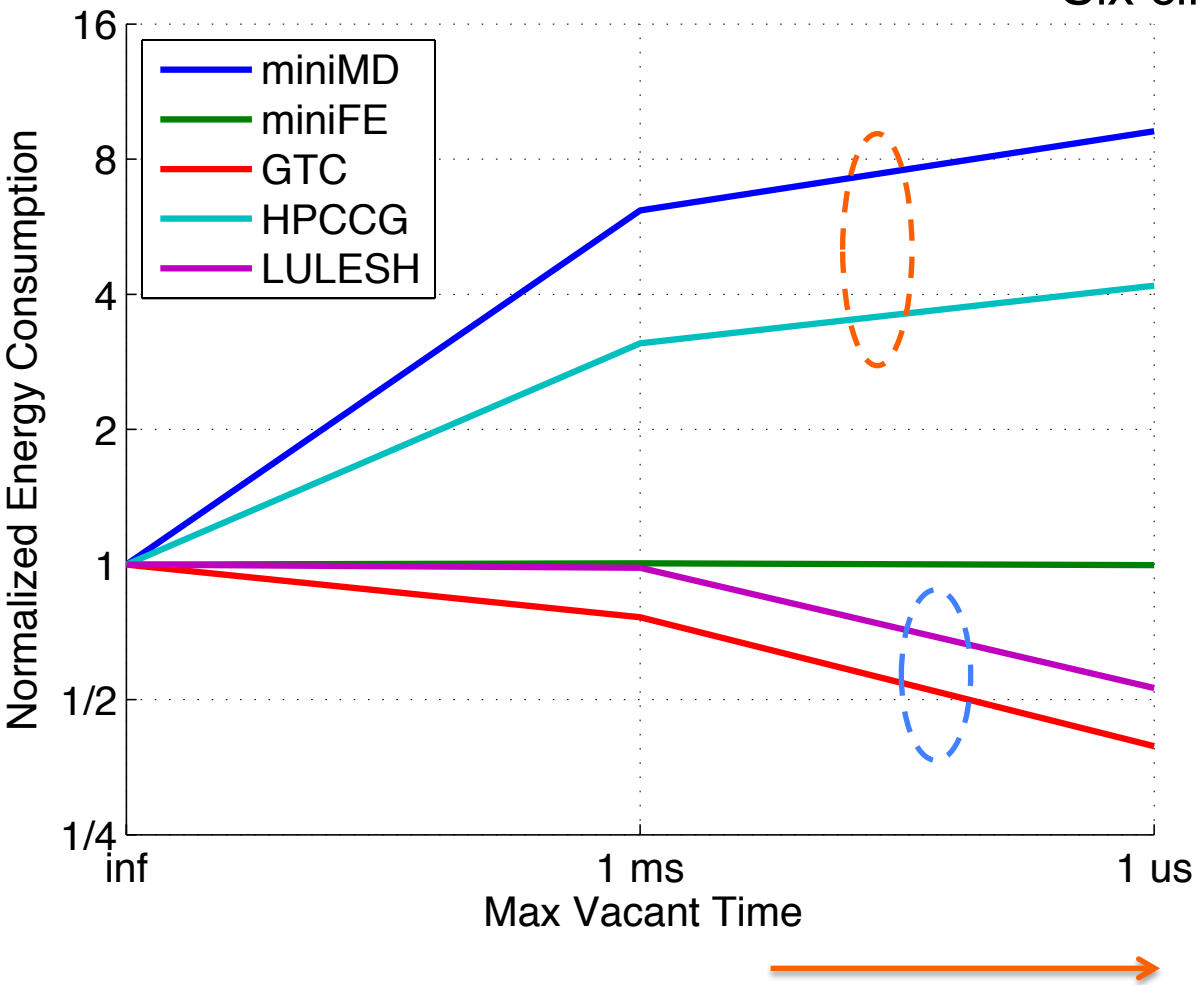



# Faster policy space exploration via abstraction + modeling



# Performance-Energy Tradeoff

* Six circuits per node




Increased circuit misses cause more power consumption

Turning off circuits saves energy

- What is the major contribution of your research?
  - We investigate architectural solutions to inefficiencies incurred in silicon photonic circuit switching;
  - We establish a circuit management framework and ModSim approaches for performance/energy characterization
- What are the gaps you identify in the research coverage in your area?
  - Co-design of communication substrate and application/algorithm
- What is the bigger picture for your research area? (i.e., identify synergistic projects, complementary projects in technical sense, etc)
  - Flattened global memory architecture (enabled by high-performance interconnects, which provides virtual uniformity)
- What major opportunities do you see for cross-pollination?
  - Cache modeling/management techniques