GPSim
A framework to model current & future manycore architectures

A. Sankaranarayanan*, J.L. Briz†,
J. Renau*, G. Southern*

*Univ. California Santa Cruz

†Univ. Zaragoza, HiPEAC

gaZ

Outline

- Scope, motivation
- Project roadmap
- Simulation Framework, benchmarks
- Avenues for Research
- Summary

Scope, motivation

- **Current manycore processors (Nvidia)**
  - **100s of simple processing** elements (PEs), ~ SIMD
  - Pros:
    - TLP, DLP / ILP
    - OpenCL coming at rescue…
  - Cons:
    - Not suitable (yet) for general purpose applications
    - Communication overheads

Non-trivial **simulation issues**

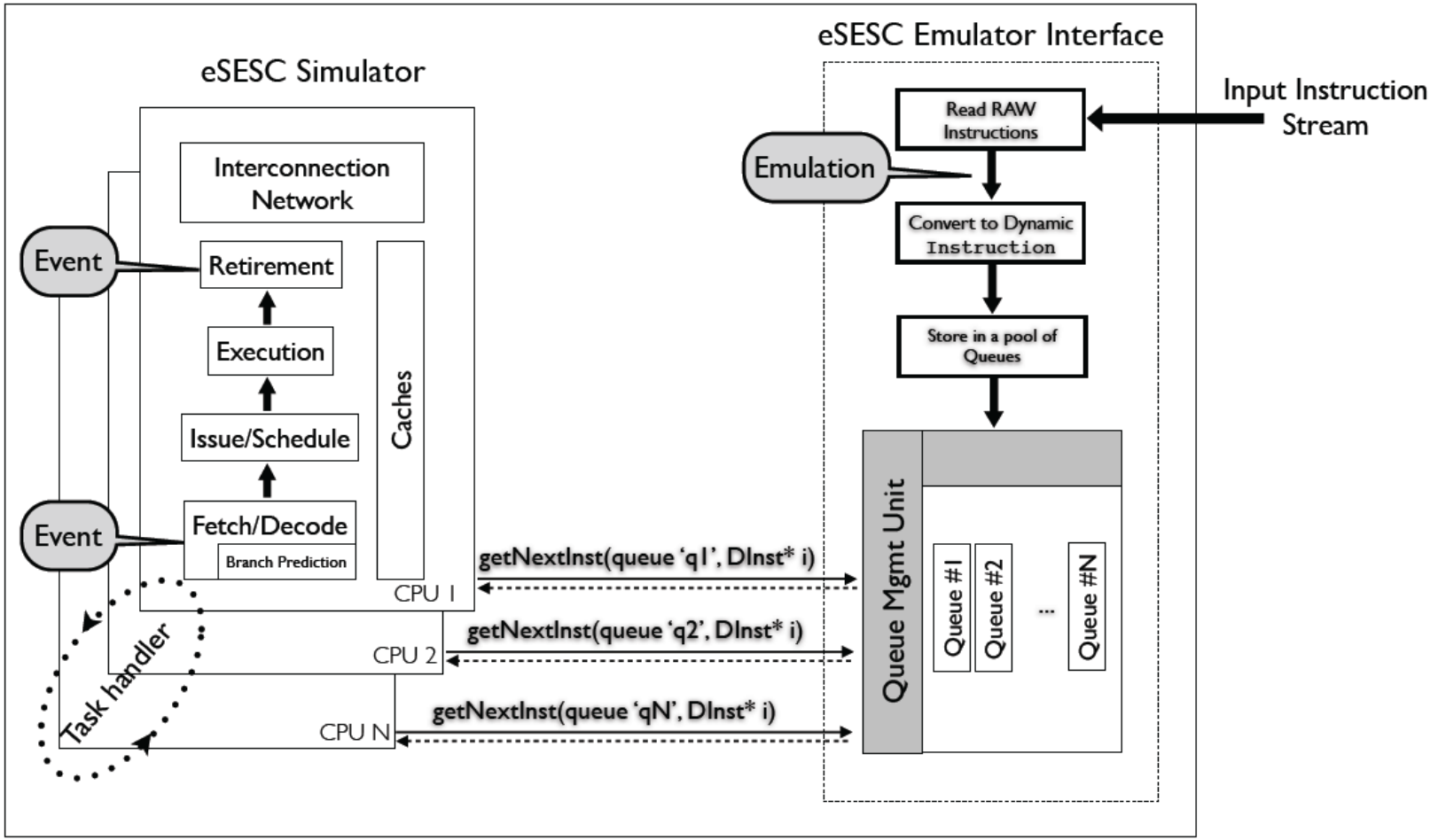
Project roadmap

- Use GPGPUs to model a manycore architecture
- Model a heterogenous system with CPUs and GPUs
- Extend this to a generic manycore system
- Propose micro-architectural or architectural modifications for a scalable, energy efficient system

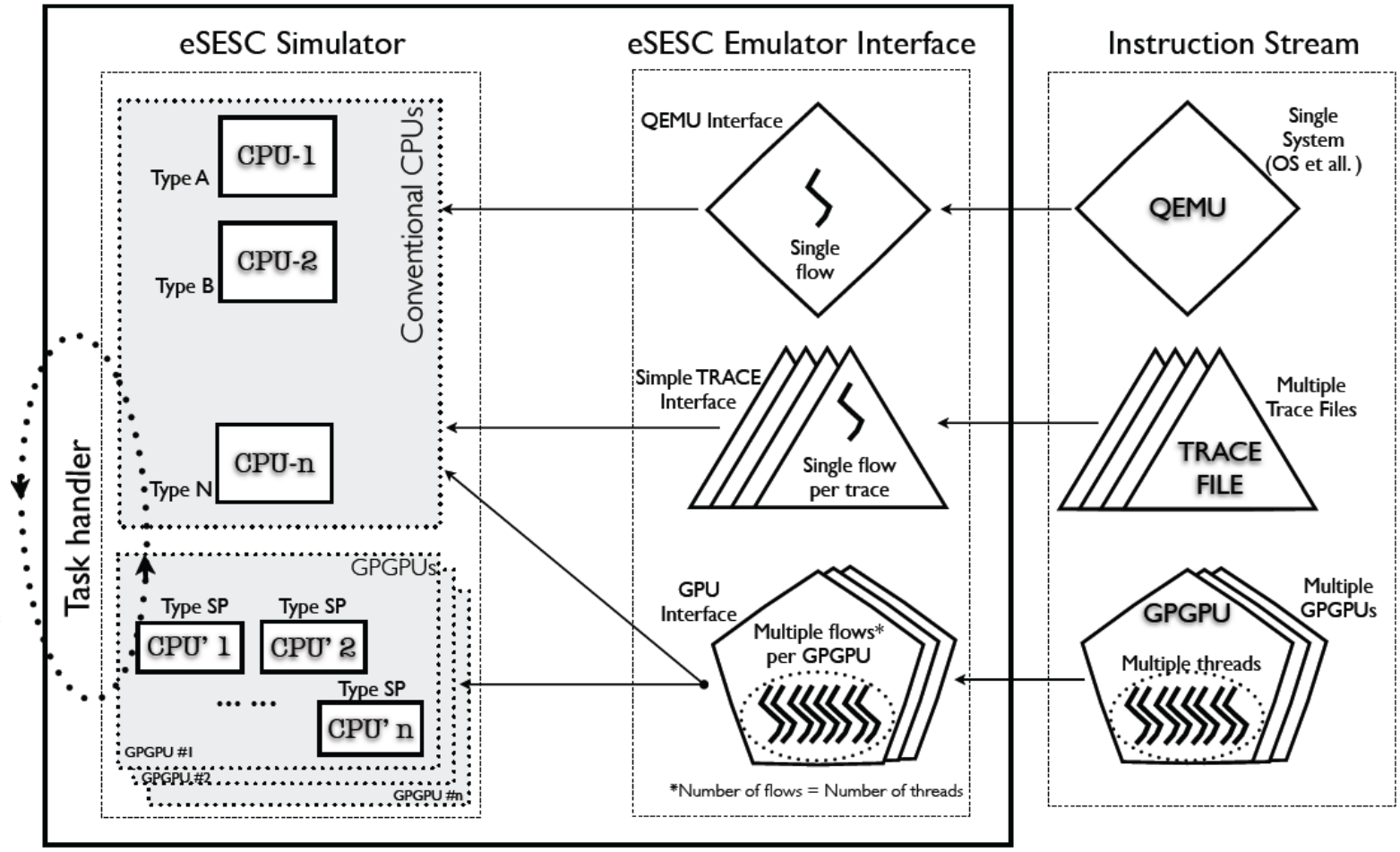
Simulation Framework

- Integrates CPU + GPGPU simulation
- Fast (enough)
- Allows full system simulation
- **eSESC** + **GPSim** + **Statistical Sampling**

eSESC Architecture



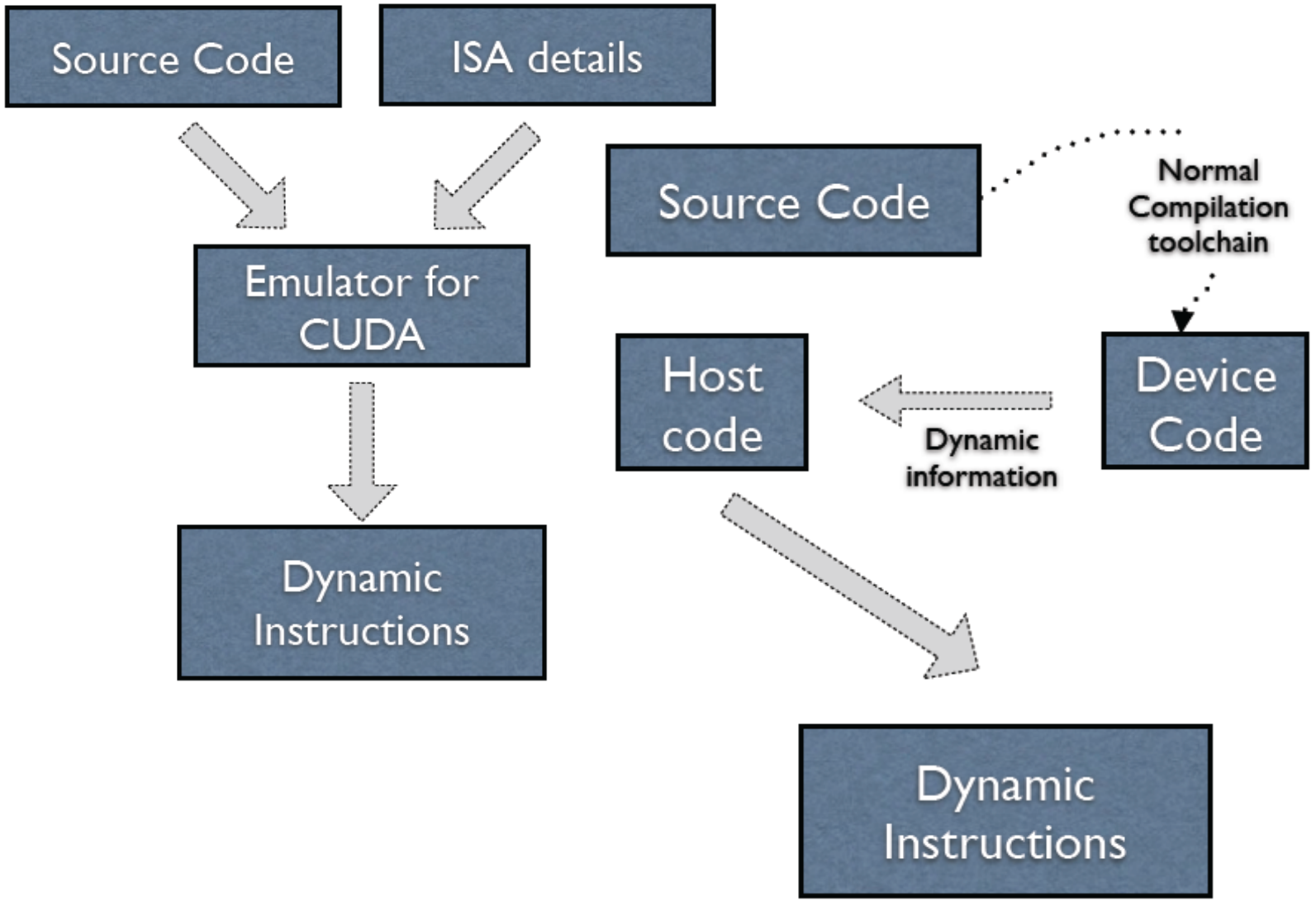
eSESC: going Full System



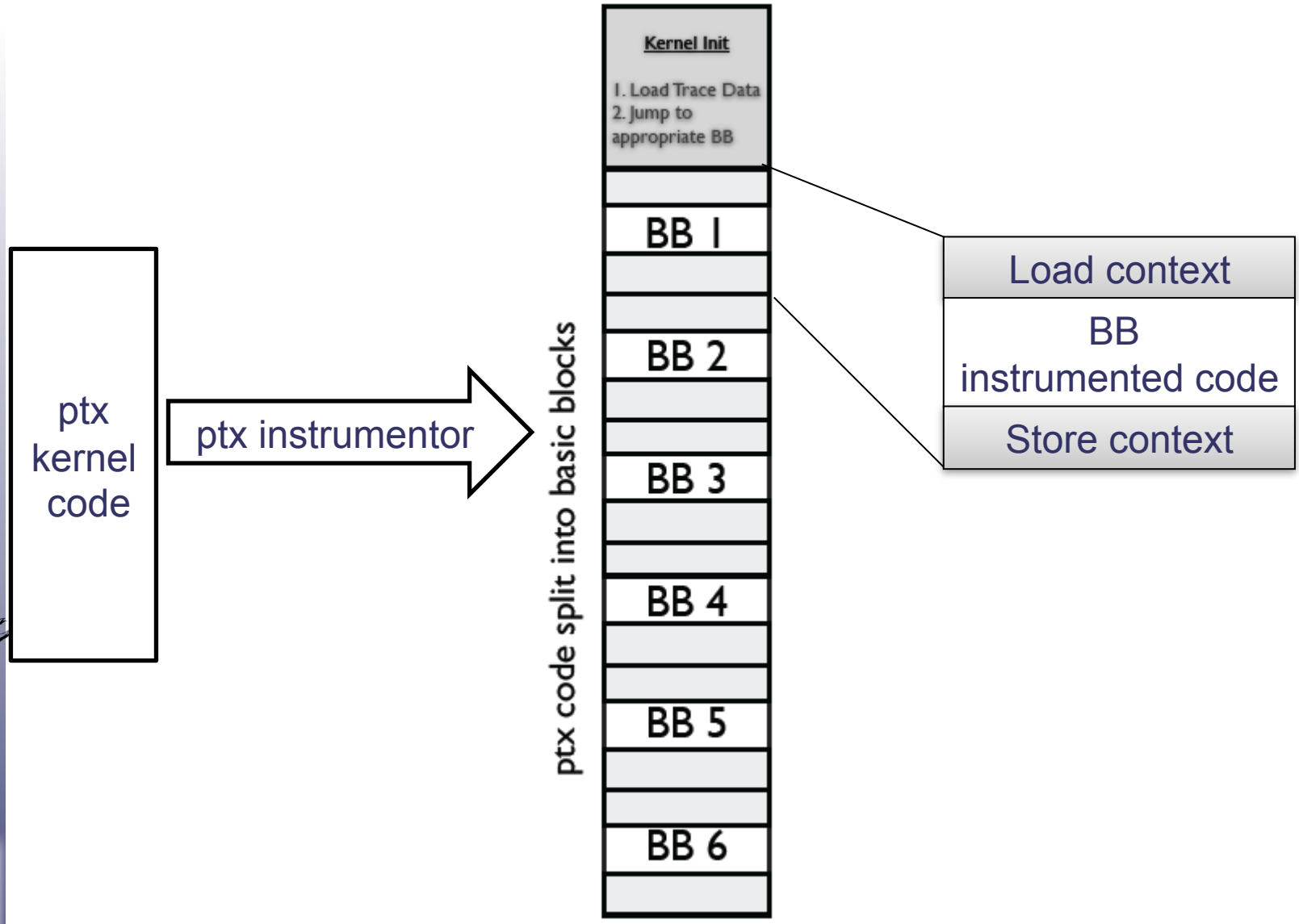
Modeling a GPGPU

- Native emulation
  - Accuracy
  - Speed
  - Other optimizations to reduce simulation time
- OpenCL compatibility

Native Emulation



Support for Native Emulation



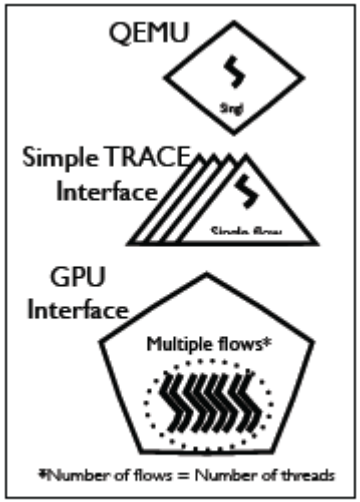
Reducing simulation time

- ~~Use a smaller input set~~
- ~~Profile based sampling~~
  - ~~Simpoint Approach~~
- Statistical Sampling
  - SMARTS/Simflex approach

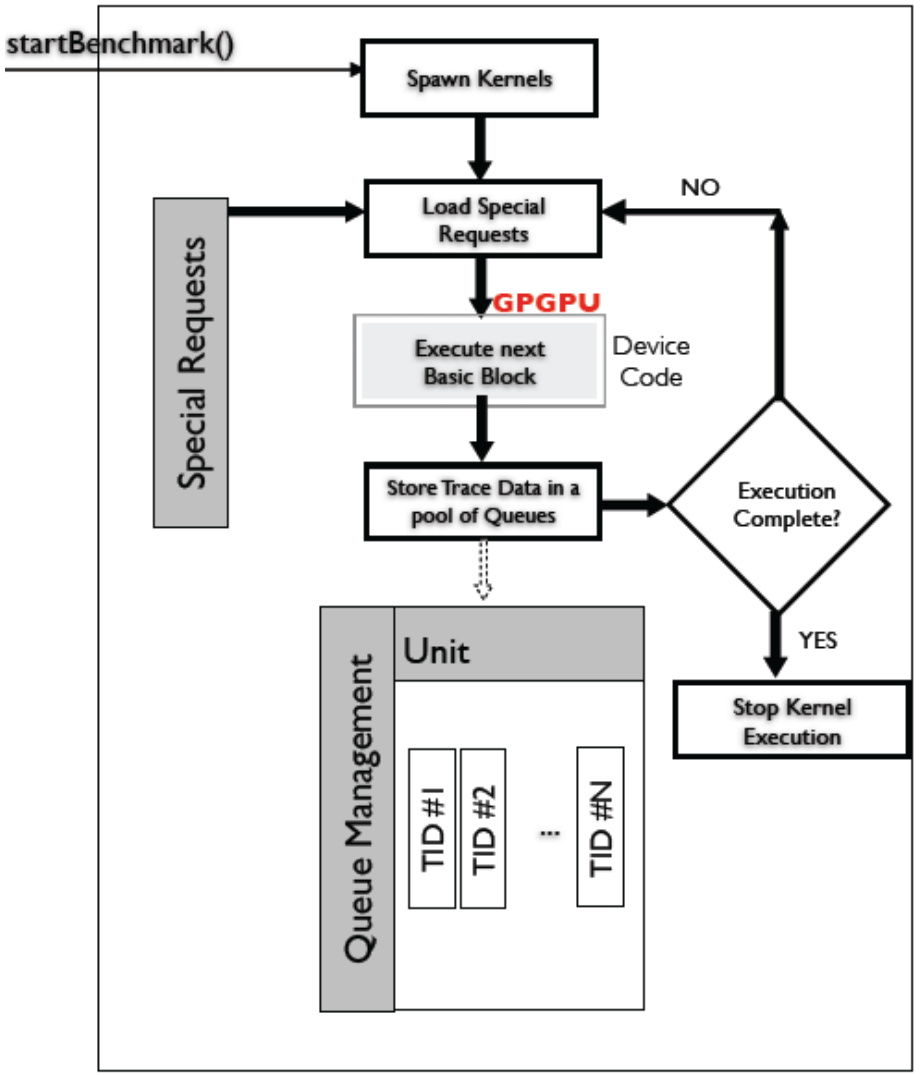
Statistical Sampling in GPGPUs

- Fast Runahead mechanism
- Regular code (one kernel, many threads)
- Ignore warm up *if* no caches within SMs

Putting it all toghether

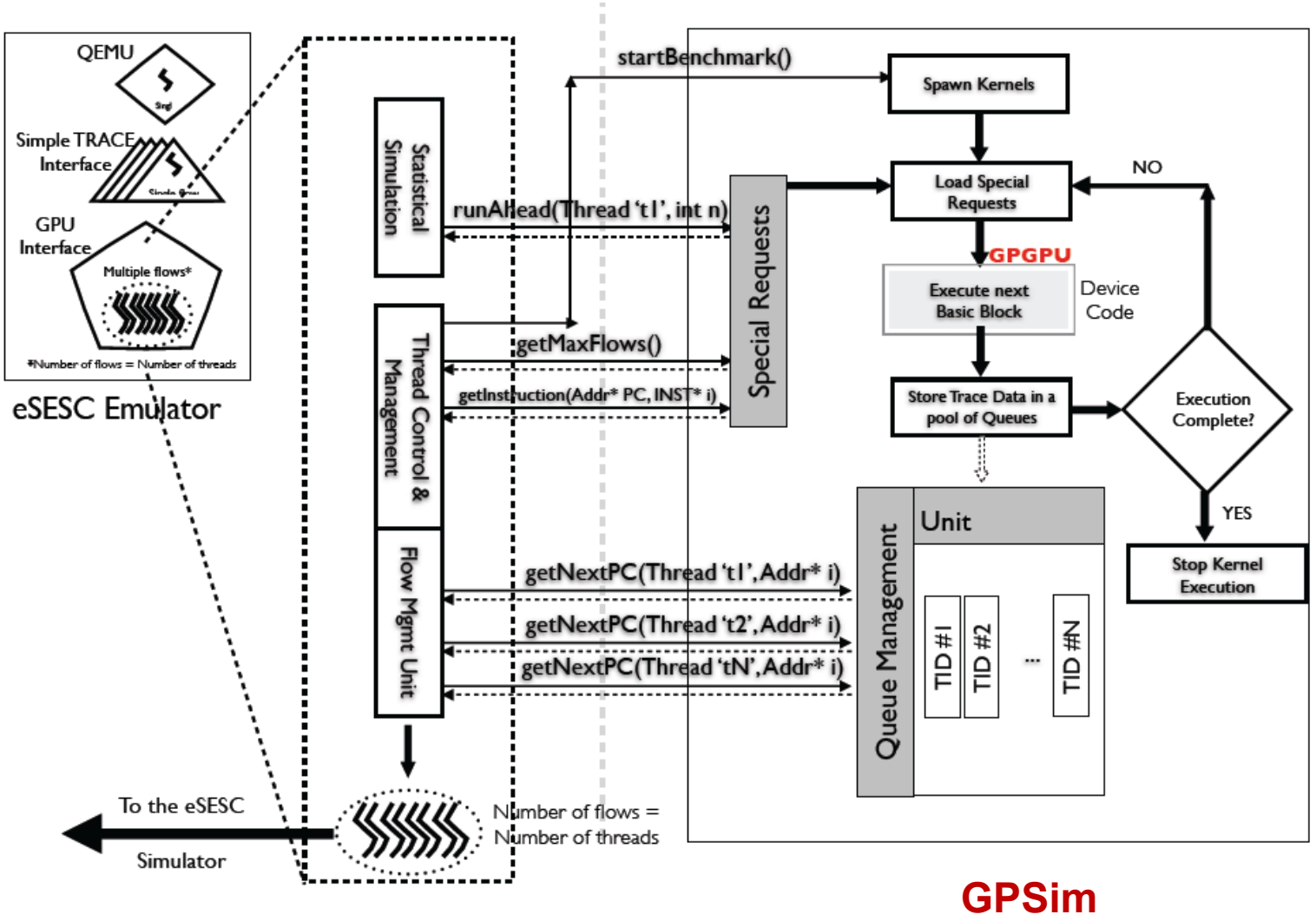


eSESC Emulator



GPSim

Putting it all toghether



Benchmarks

- No standard suite
- Mostly handpicked from CUDAZone and SDKs
- Choice as per the guidelines in *"The Landscape of Parallel Computing Research: A View From Berkeley"*
- Black-Scholes, Sparse matrix operations, Brownian particle movement, matrix multiplication…
  - DLP-friendly vs divergent behaviour
  - Easy memory coalescing vs. sparse or erratic behavior

Avenues for Research

- Components of a heterogenous manycore architecture
- Memory hierarchy issues
- Optimizations for area/energy/power efficiency

Summary

- Objective: To explore and propose high performance, scalable, energy efficient manycore architecture
- GPSim : A robust simulation framework
  - Currently working on the simulation infrastructure
- Target full system simulation and evaluation of at least one manycore system before the end of next year

Thanks…

Programming model and hdw