SAND2009-5499C

Structural Simulation Toolkit

Arun Rodrigues

.

Many Cores + Memory

X

Many Many Nodes

X

Many Many Many Threads

- Audiences.....

X

Application writers
purchasers
designers

X

system procurement
algorithm co-design
architecture research
language research

X

present
future

xity.....

Physics Apps
matics Apps

X

Communication Libraries
Run-Times
OS Effects

X

Existing Langu
New Languag

ints.....

Performance

Power

Cooling

Risk

NI's Simulation More Challenges

work/Application Feedback: A static trace or simple statistic
el will not capture the causal relationships between message

ability: Many network effects only become apparent at hund
ousands of nodes.

able Processor/Memory/Network Systems: Local interacti
have global performance implications.

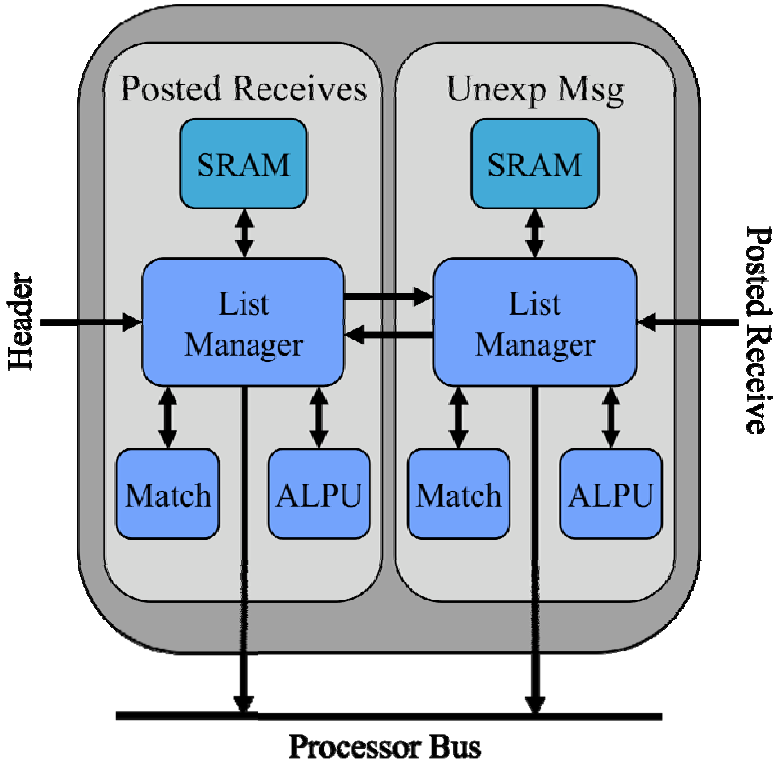
ty to Model Message Overheads: Overheads in the netwo
packetization, protocol overhead) and messaging library (e.g
hing, message assembly) can have a major effect on
rmance.

ty to Explore Programming Models: Novel hardware will r
l programming techniques and capabilities.

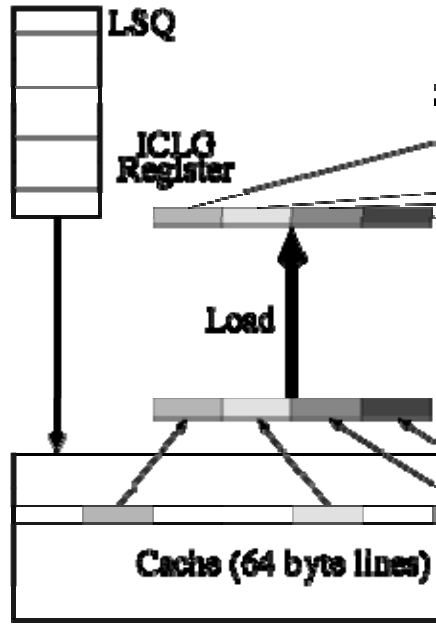
er and Economic Effects: Power and cost are the key limiti
rs on system design. Any system model must be able provid
back on the power and cost implications of new architectural

Several Projects

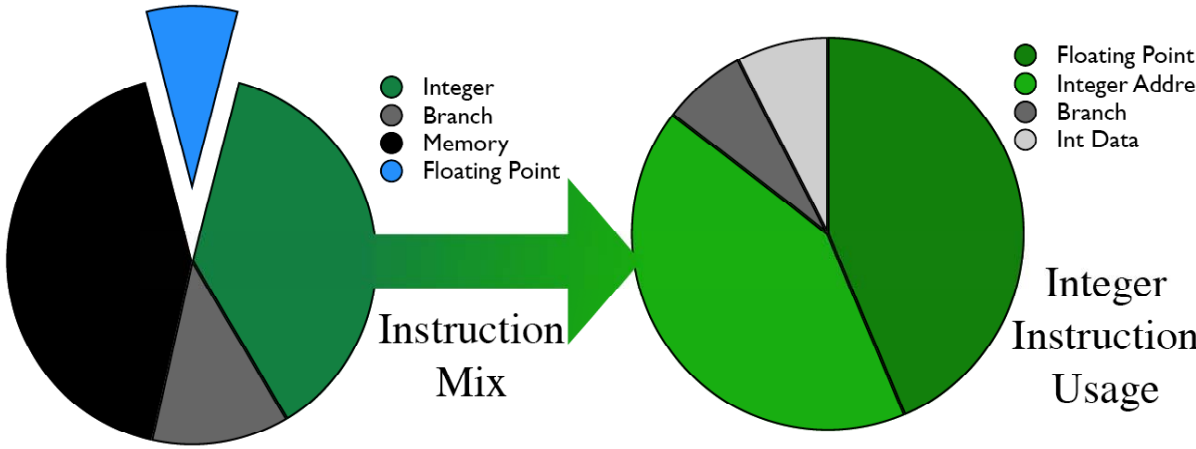
- architecture
- r CacheLine Gather (ICGL)
- on. FU (Wisc./SNL)
- Aggregates
- emory Ops
- cation analysis
- nory Footprint
- ruction Usage
- ork/MPI
- Tradeoffs
- Acceleration
- amming Models
- Compiler work (SNL/Rice)
- allX (LSU/SNL)(FastOS)
- nsactional Memory (ORNL)



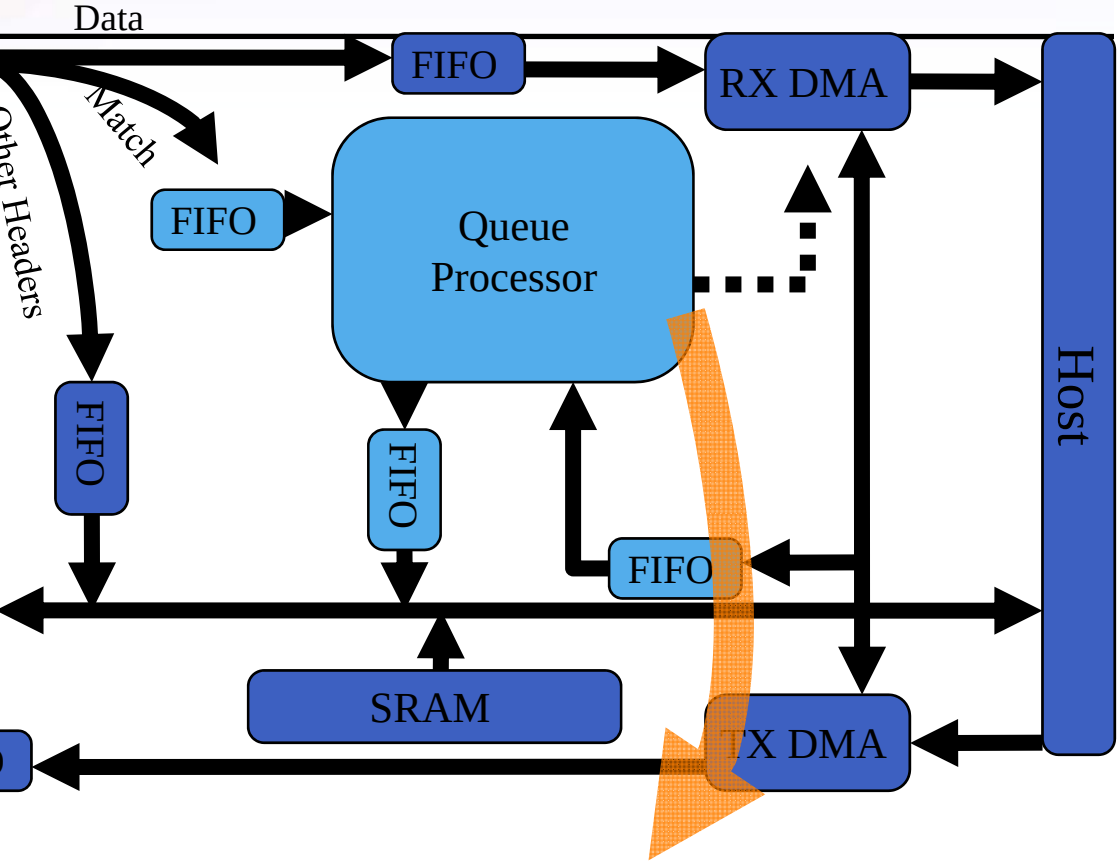
MPI Accelerator Tested with SST

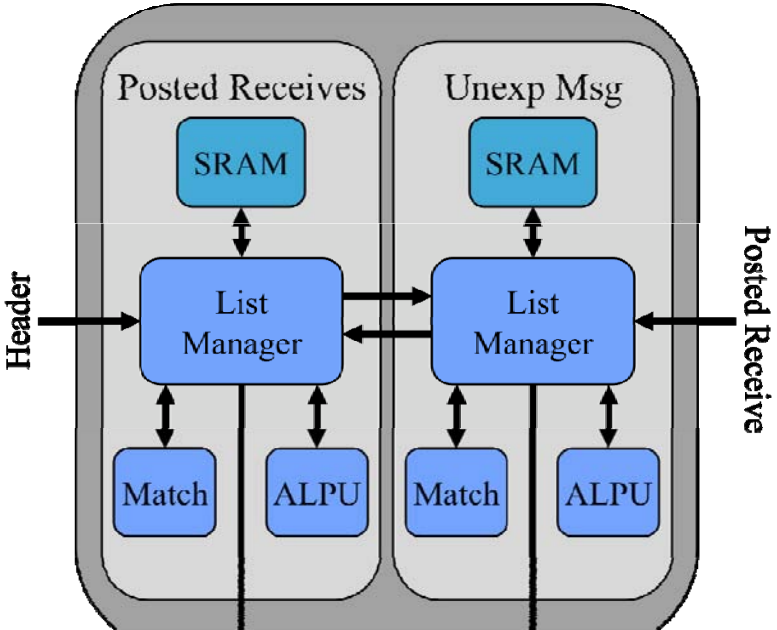


Inter CacheL
Scatter/Gath
Microarchitec



Instruction Mix & Usage for





- **Problem: Message rate determines effective banc**
- **SST Analysis: MPI matchi limits performance**
- **Solution: Accelerate list management with hardwa**
- **SST Simulation**
  - **Hardware**
    - **Implement NIC/Router base RedStorm SeaStar**
    - **Implement List Manager, A Match Unit, integrate with N**
      - **"Translated" from FPGA**
    - **Validate against FPGA & R**
  - **Software**
    - **Create baseline offload MP**
    - **Modify MPI to use accelera**
- **Impact: Collaboration w/ I**

Goals

- Become the standard architectural simulator for the HPC community
- Be able to evaluate future systems on DOE workloads
- Use supercomputers to design supercomputers

chnical Approach

ale

ccurate to analytic

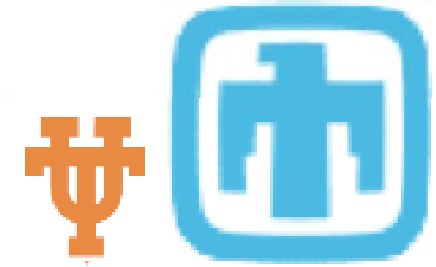
tion-based to message-based

f simulated nodes on 100s of des

ted Tech. Models

Consortium

- "Best of Breed" simulation suite
- Combine Lab, academic, & indus

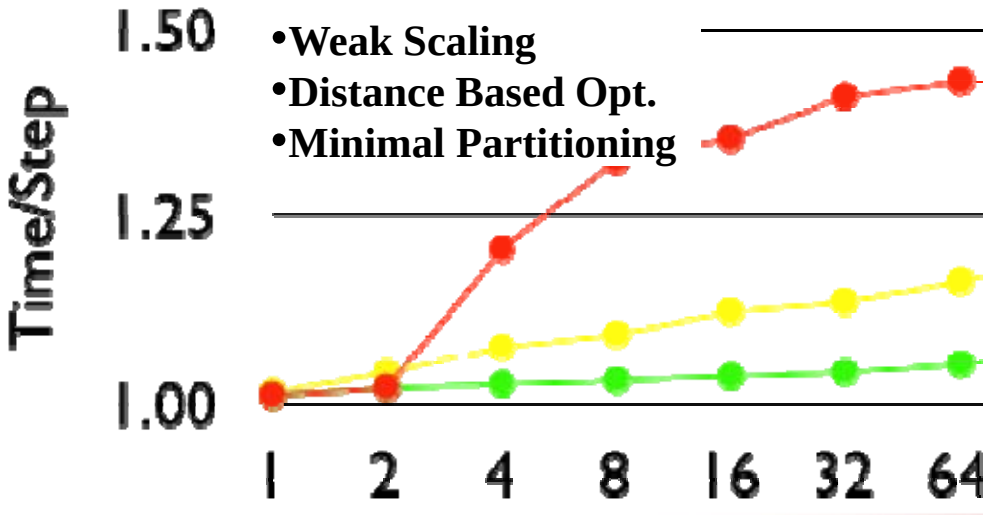


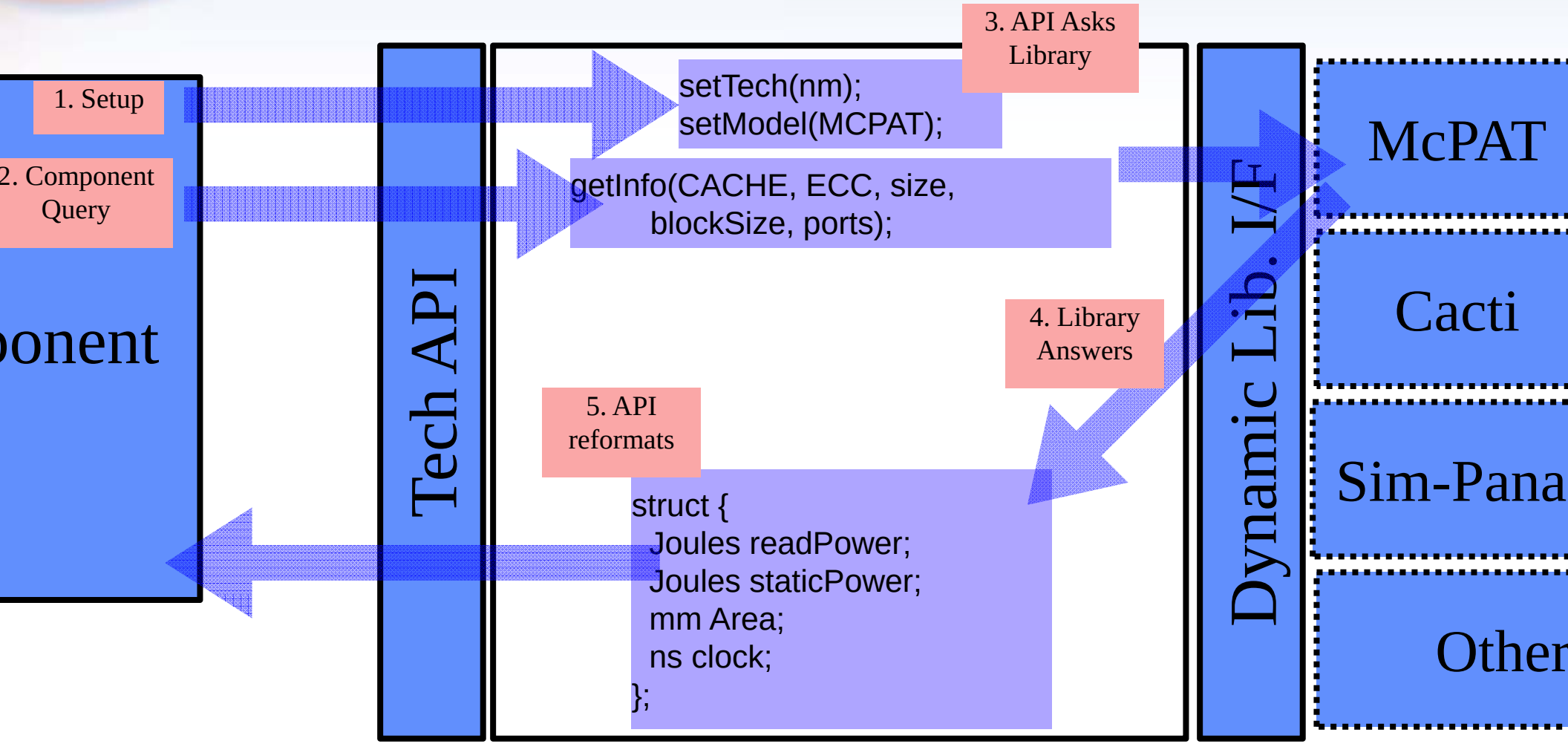
```
= getNextEvent(queues)) {
nt->time;
sClockEvent) {
mponent->preTic();
ent->exchange) {
nds();

ends();
ent->checkpoint) {
oint();

mponent->handleEvent();
```

Parallel Core Pseudocode

- **Requirements**
  - **High speed, Parallel, Scalable**
  - **Multiple clock domains**
  - **Checkpointing**
- **Implementation**
  - **Conservative distance-based D optimization**
  - **Multi-criteria partitioning**
  - **Built on MPI**
  - **Future: FPGA acceleration**

Message Traces, Symbolic Workload Descriptions	Execution Based	Execution w/ FPGA Acceleration
100s-1000s	100-1000s	1-10
10000s-100000s	100s-1000s	1-10
System Scaling Behavior	Cycle-level system performance	Co-design





ate interface to multiple technology libraries

wer/Energy

liability

ea/Timing estimation

e it easier for components to model technology

nterface component reuse at
nt scales

& Low-level interfaces (more?)

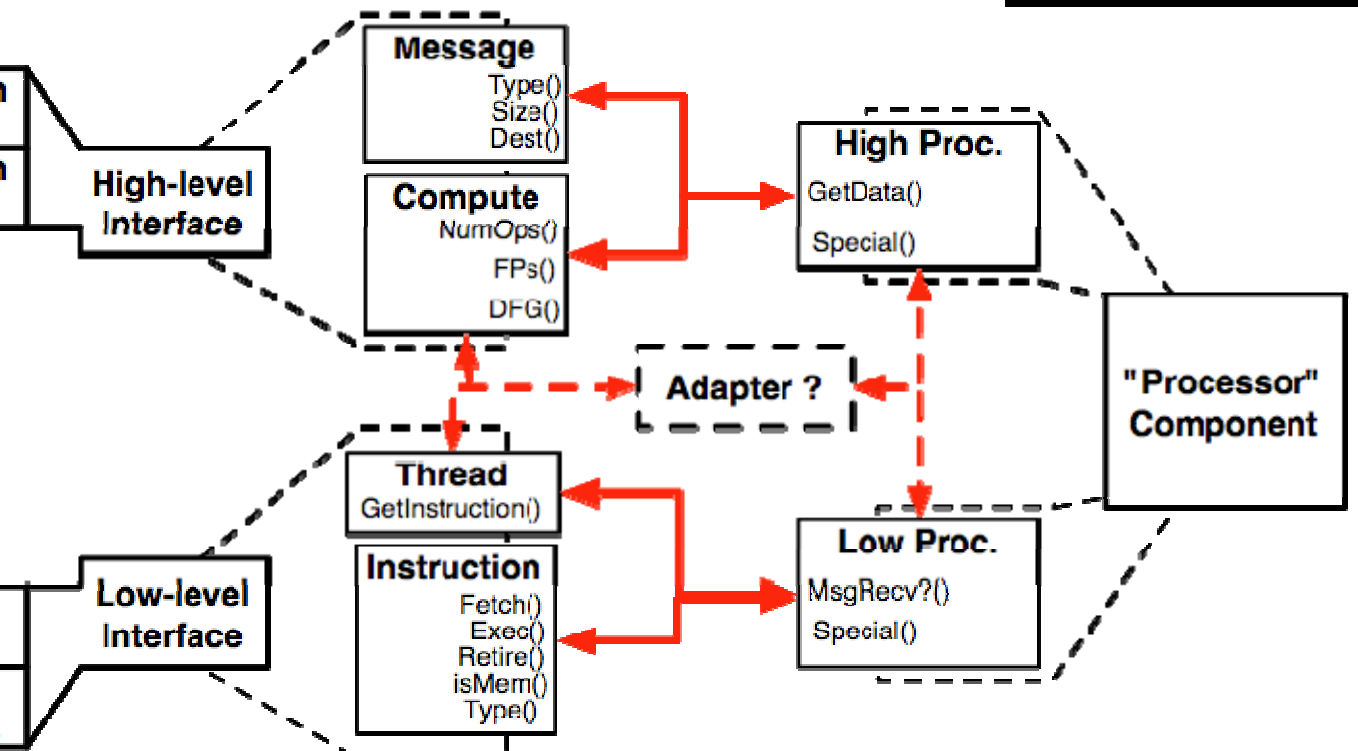
ows multiple input types

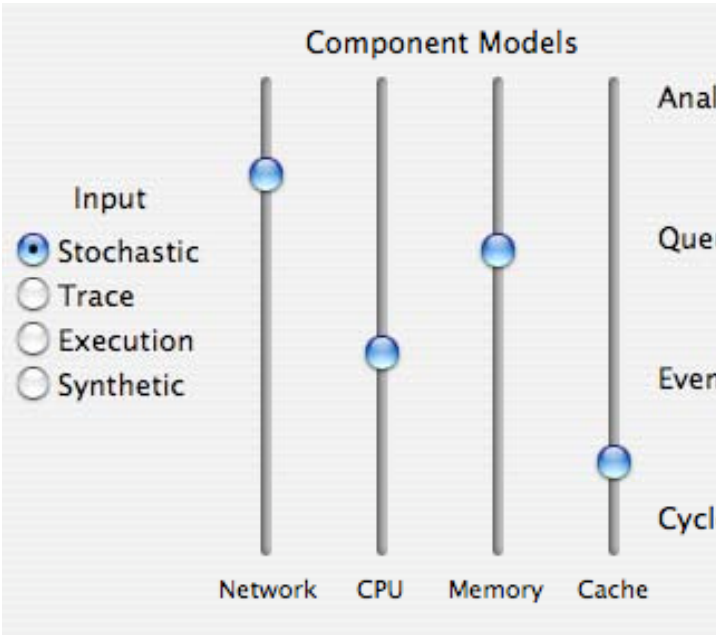
ows multiple input sources

races, stochastic, state-
nachines, execution...

apter objects to translate?

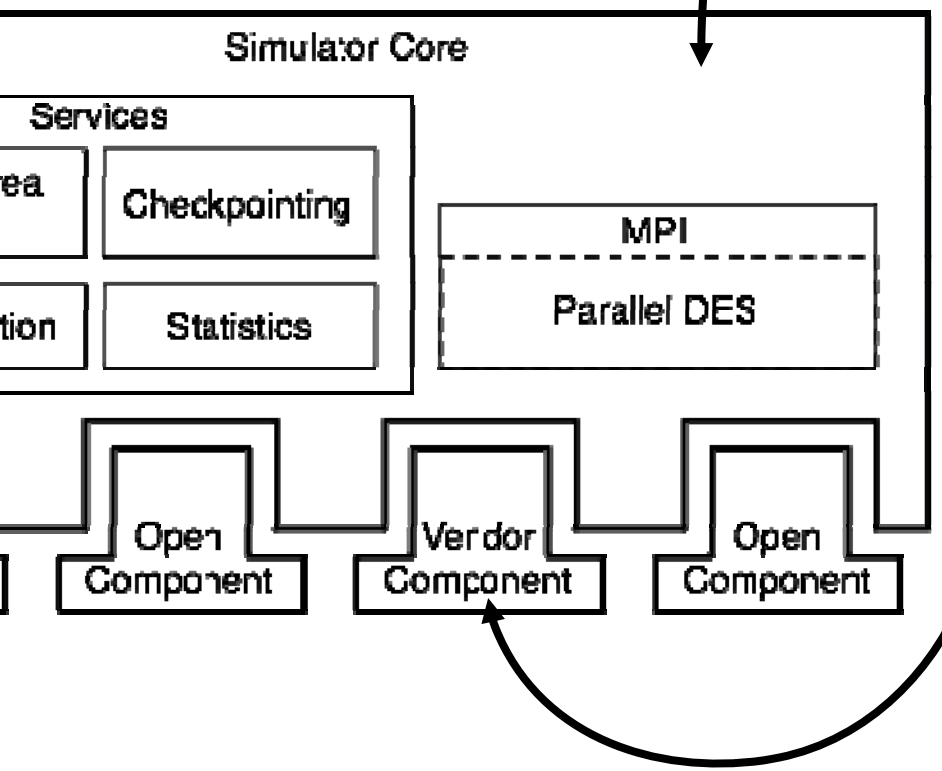
	High-Level	Low
Detail	Message	Instr
Fundamental Objects	Message, Compute block, Process	Instr Th
Static Generation	MPI Traces, MA Traces	Instr Tr
Dynamic Generation	State Machine	Exe





Multiscale Parameters

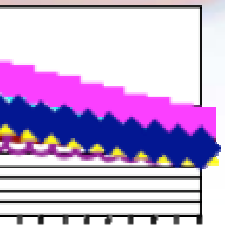
Simulator Framework Model



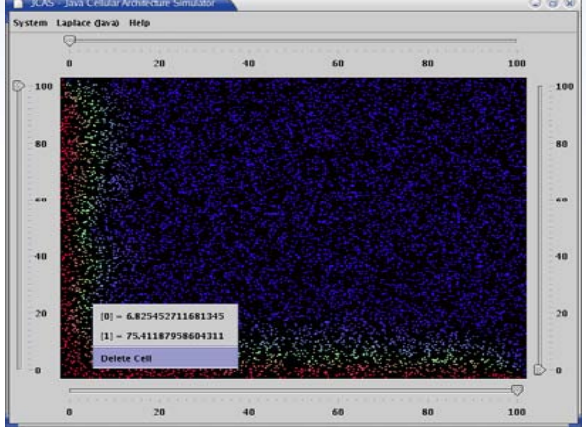
- **Simulator Core will provid**
  - **Power, Area, Cost modeling**
  - **Checkpointing**
  - **Configuration**
  - **Statistics gathering**
  - **Parallel Component-Based Discrete Event Simulation**
    - **MPI hidden from user**
    - **Multiple clocks**
- **Components**
  - **Ships with basic set of open components**
  - **Industry can plug in their ow models**

Planned Releases

ponent	September	Dec 09	Future
CORE	DES, Checkpointing, Configuration	Improved Power Models, Optimized DES	Manifo Compatib
essor	SimpleScalar-based	Stochastic NMSU Models	Detailed
work	Off-load NIC, Simple Router, Macroscale Models	Improved NIC, multiple topologies	Improv Router
mory	DRAMSim II Integration	Improved	Improv



g Assertions

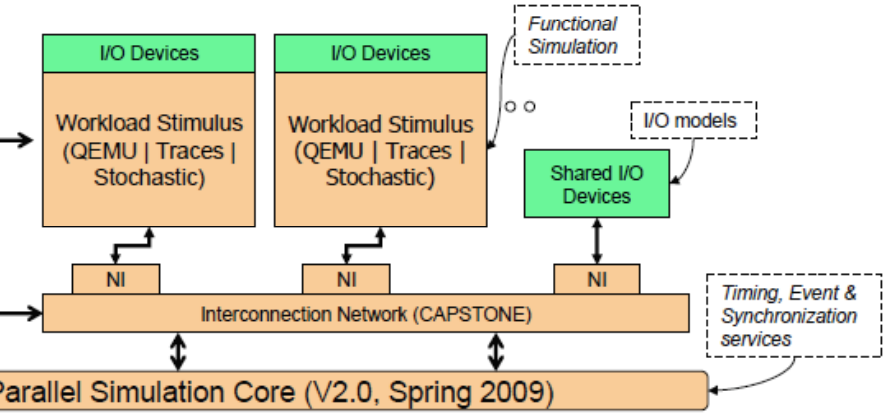


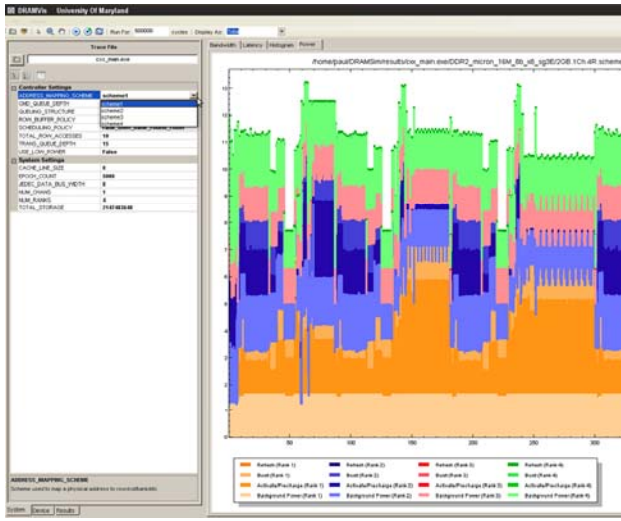
JCAS Vizualizer

- IAA Simulation effort is a community effort
- Seeking more partners...

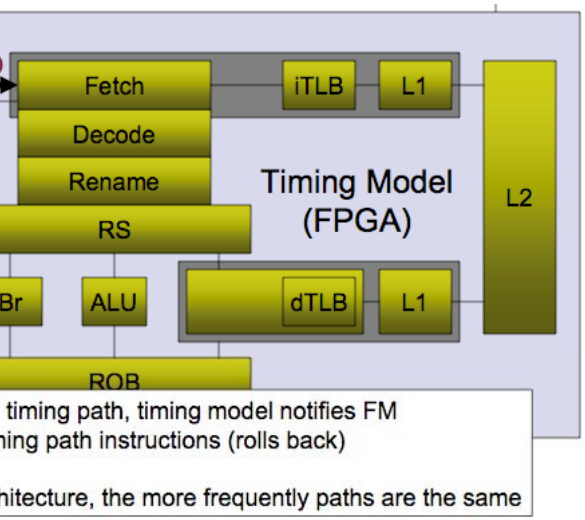
- Current consortium
  - Sandia (Structural Simulation Too
  - ORNL (Scalable application model
  - U. Maryland (DRAMSim II)
  - U.Texas-Austin (FAST)
  - Georgia Tech
  - JCAS (ORNL)
  - Seshat (SNL)

Overview

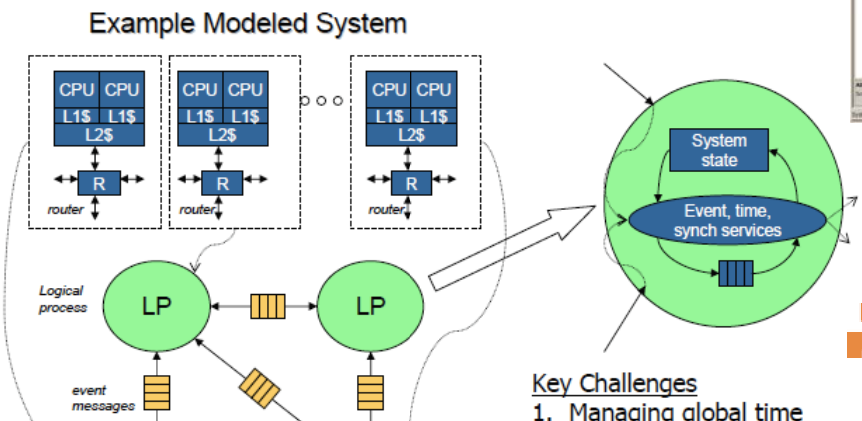




DRAMSim II



FAST



Big Question

Seeks to be...

ltiscale

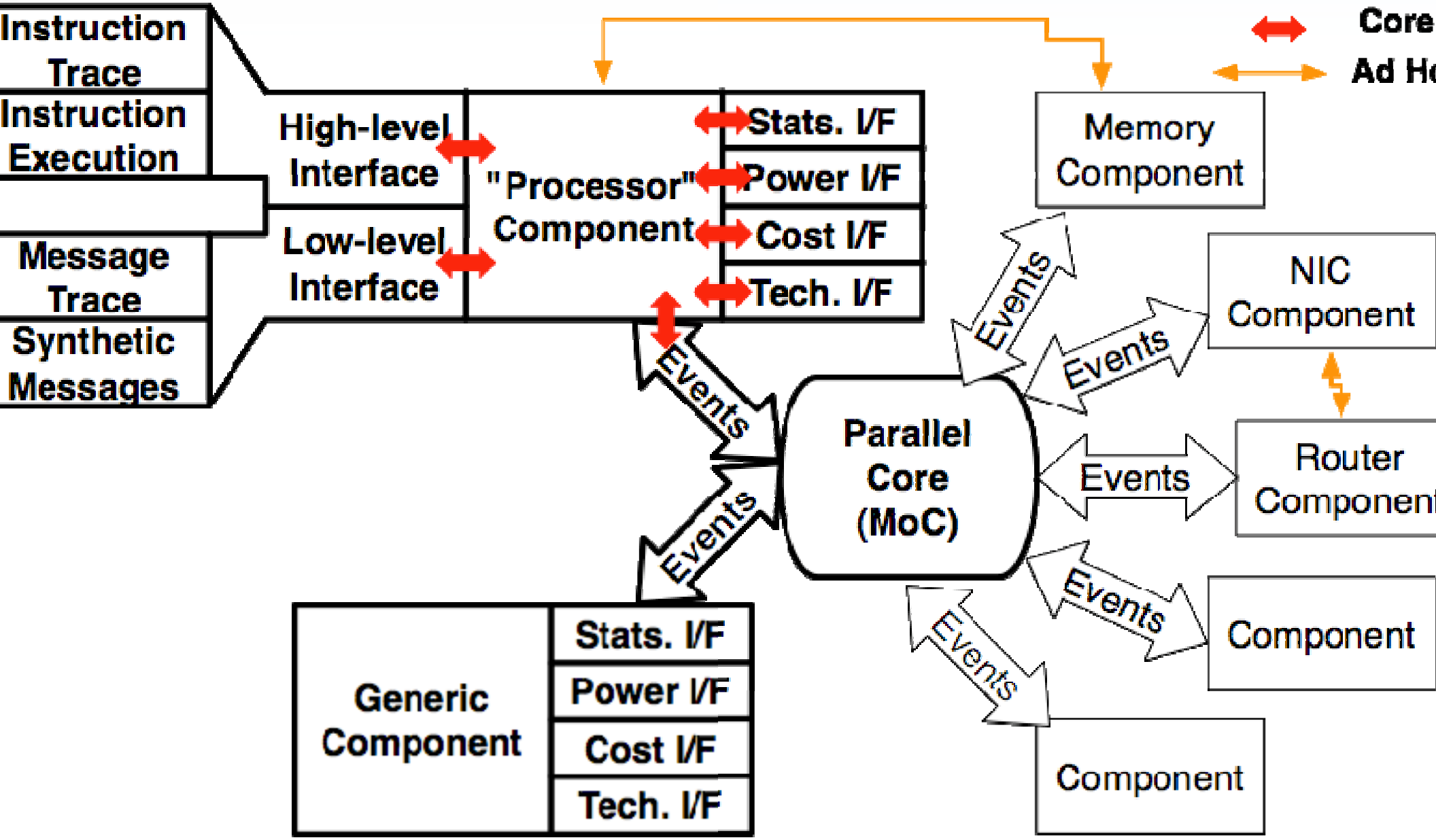
rallel

listic

dia is actively seeking partners, requirements, an
ce on the SST

w can the SST be useful for YOU?

Bonus



Separate Software/Front-End from Hardware/Timing/Back-End

Standard interfaces for power, area, cost?