YOU AIN'T SEEN NOTHING YET

Outline

- A (quick) look back
- Notable accomplishments (says me)
- Where is my … (unsolicited whining)

David J. Mountain, PhD
Senior Technical Director
NSA Advanced Computing Systems Research Program (ACS)
djmount@lps.gov

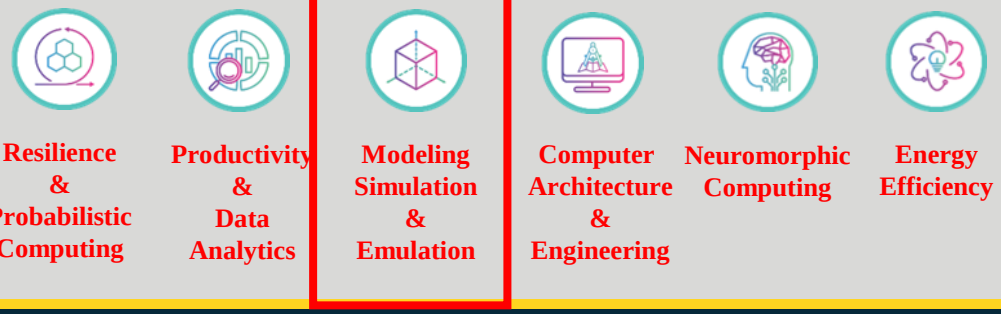
What is ACS?

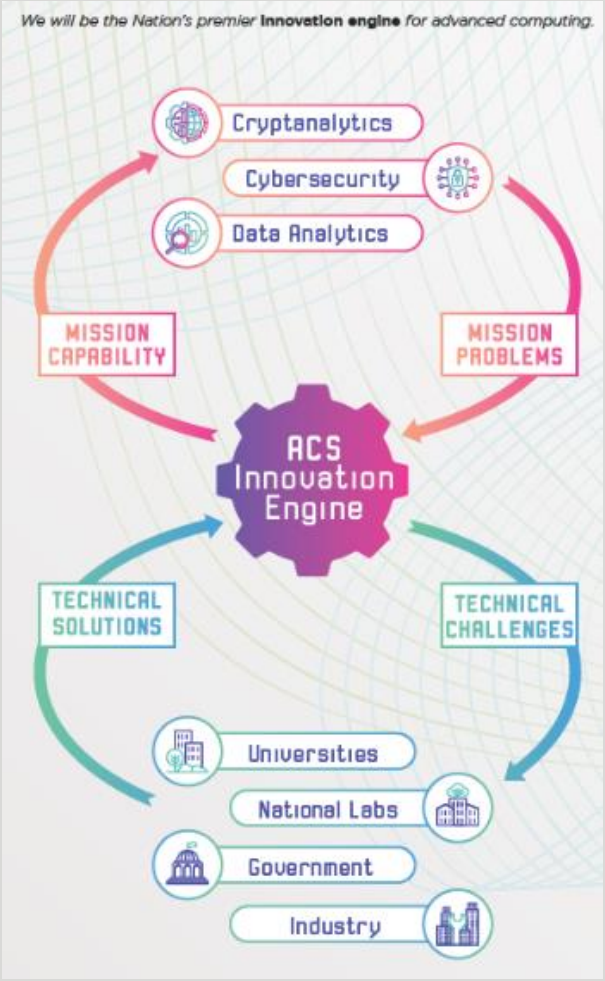
Advanced Computing Systems (ACS) Research Program

We will be recognized, both internally and externally, as the nation's premier **innovation engine** for advanced computing.

We conduct **exploratory research** that combines *algorithms*, *architectures and technologies* to demonstrate and/or develop advanced computing systems that provide **asymmetric advantage** for agency mission.

Our innovation engine is built upon **world class, mission oriented, participatory research.**





THESE WERE THE GOOD OLD DAYS?

Workshop on Modeling and Simulation of Exascale Systems and Applications

Organizing committee:

Bill Harrod, Adolfy Hoisie, Darren Kerbyson, Bob Lucas, Arun Rodrigues,
Sonia Sachs, John Shalf, Allan Snavely, Jeffrey Vetter

August 9-10, 2012; Seattle, WA

Current Issues

- Performance modeling only (not energy or resilience)
- Analysis is still dominated by simple metrics, back-of-the-envelope approaches
- A lot of use after the fact, less in design
- Tools-of-the-trade in industry and open R&D not coordinated/calibrated

ACS IN 2012/2013

Hired Noel Wheeler to lead our Modeling-Simulation-Emulation effort (creating a team of 2)

3 mandates – priority order

- Develop MSE tools
- Use MSE tools
- Create an MSE community

Looked at 50^{+} tools – found 2 worth further exploration (Manifold, SST)

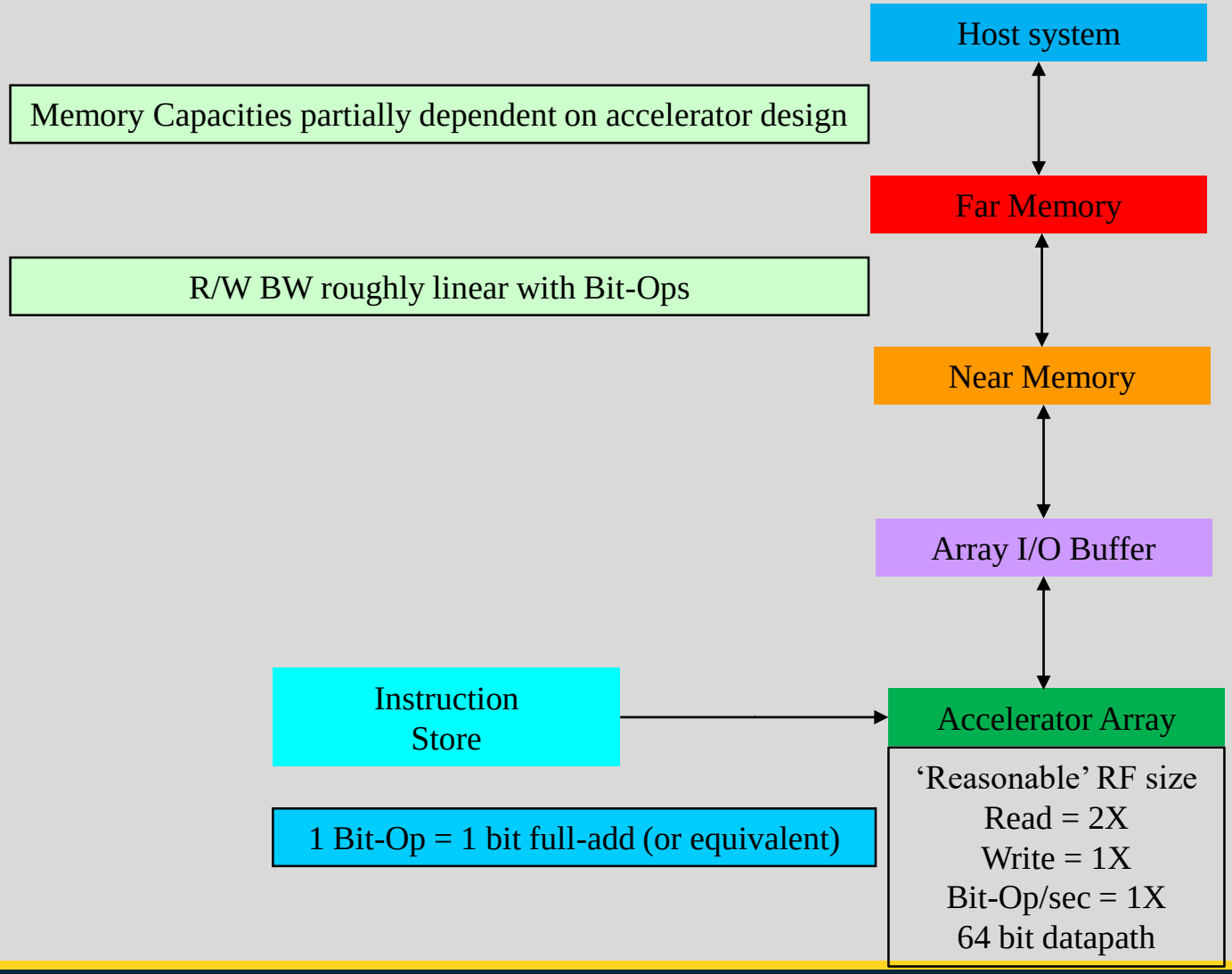
- PAL, OCCAM partnerships formed

ACS July 2013 workshop identified 3 priority efforts – <u>none of them were research</u>

- Usability – what a concept!
- API for interoperability and composability of SST models
- Documentation and training – Ben Payne, YouTube star (https://www.youtube.com/watch?v=_pXj8oGYgjU)

ModSim 2013 presentation – ACS becomes part of a nascent, national, coordinated effort

SUPERCONDUCTING TECHNOLOGY CHALLENGE -- 2013 ACS WORKSHOP



Use case (1 of 8) – Benchmarks/proxies guide a simulation to provide power/performance metrics that can be validly compared to alternate technology options.

Design questions:
Can we mod-sim this at all?

What are the power/performance trade-offs?

How do we know what goes where (differing temperature domains)?

How do we identify detrimental scaling effects?

TODAY -- THINGS ARE LOOKING (RELATIVELY) GOOD TO ME

NOEL, YOUR MIDTERM REPORT CARD IS READY FOR REVIEW

Develop MSE Tools: B+

SST version 12 is now out, with regular, scheduled updates

- Lots of interoperable models for memory, interconnect, processors; Python3 support
- ESSENT + SST Integration

ReGEM5 – Attaboy Jason – v21 available, ARM, RISC-V focus

Generalizable HW design in PyRTL, Verilog ‘containers’

SST + FPGA Emulation

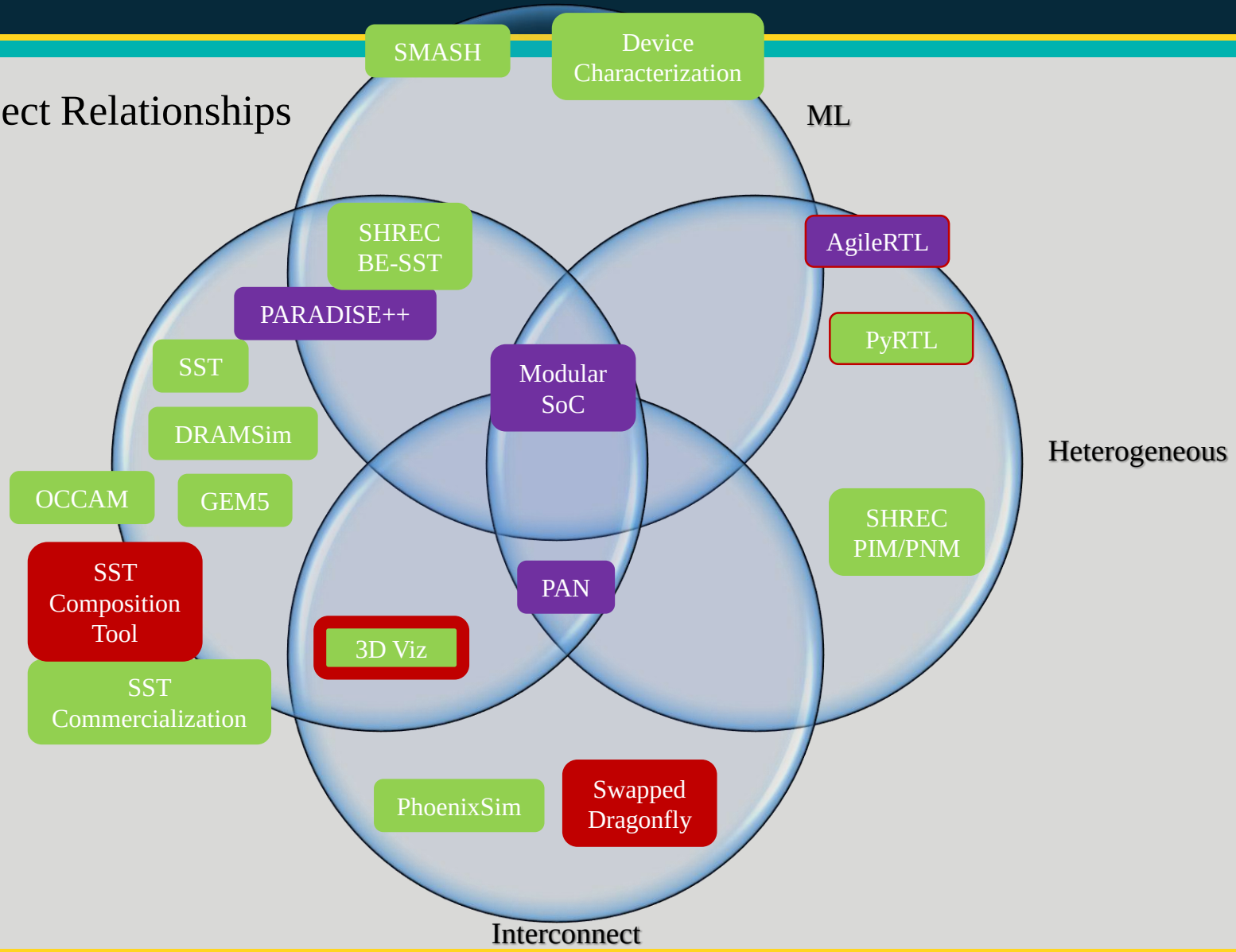
ML trained ISA simulation

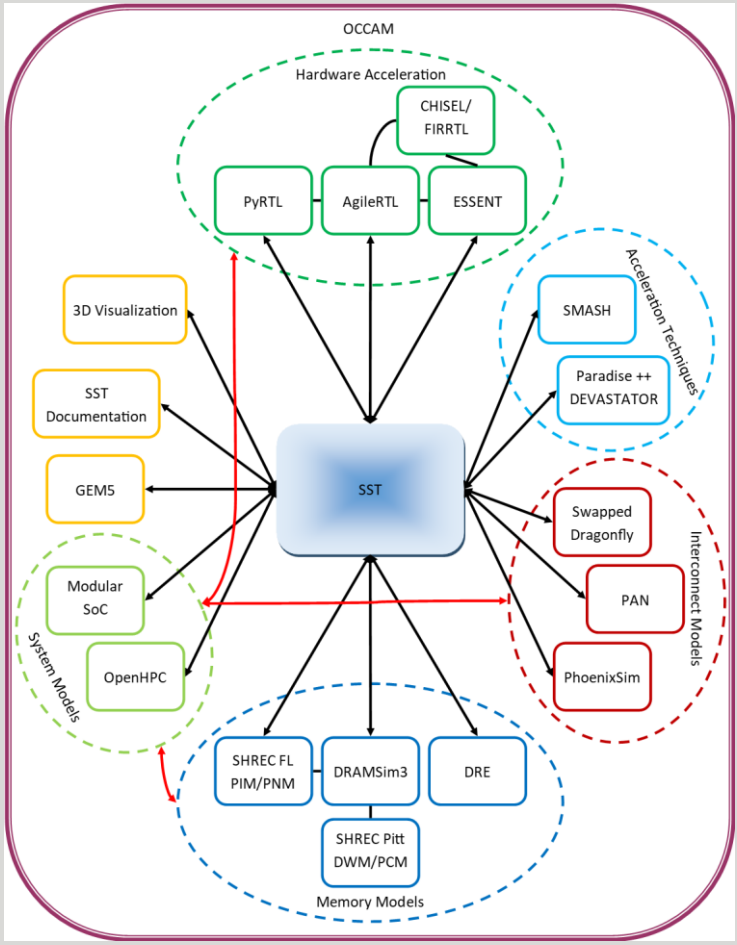
Documentation and training is much, much, much, …, much better

- Open source and usable out of the box is pretty close to the default

A STRUCTURED WAY TO ENSURE CONTINUED PROGRESS

MSE Project Relationships





USE TOOLS: A-

We have (at least) 4 major internal projects that use MSE tools extensively

Not research, these are designs expected to be built

Project 38

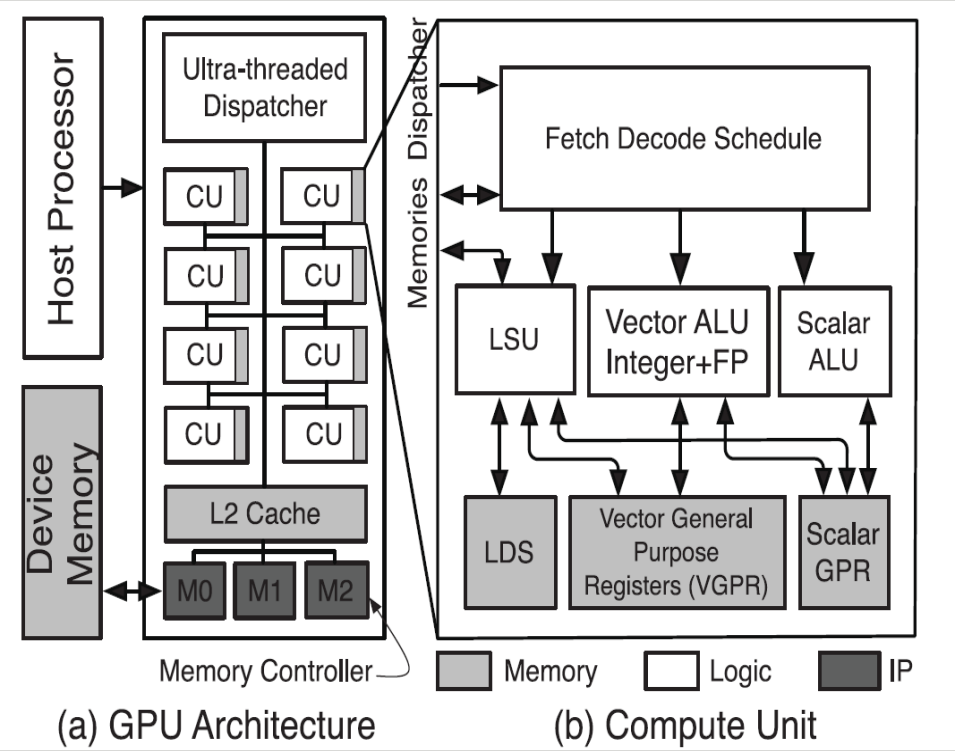
&

IARPA AGILE

rely heavily on MSE tools

And remember that superconducting technology challenge?

SUPERCONDUCTING ACCELERATOR – NOMINAL DESIGN POINT



- Model uArch in GEM5
  Evaluate performance impacts and energy efficiency of specialize pipelining
- Model Systems Level Architecture using SST
  Evaluate the performance of the system at scale
- Options for memories at different temperatures

CU
Ins Buffer: 64KB
LDS: 64KB
Data/VGPR: 384KB

(VGPR: 2Kb bits read)

3 clks
(32B cache line)

4K region

233K region

LLC
Hybrid
eDRAM

Up to 10 MB

Simulated access time ~40-60ns

DRAM
HBM

4GB

Develop a Community of Tool Users: C+

This is very positive, but I want so much more

How to get an A^+ , Noel

1. ACS (research) no longer needs to fund sustainment work

2. Computer Architecture labs are part of and generally available in graduate and undergraduate courses
   a. Number of schools with them
   b. Number of lab exercises they have in the curriculum

3. Computer Architecture research papers are rejected if they aren't based on standard, widely available mod-sim tools

4. Commercial computer systems routinely offer a precise, high fidelity simulation environment, based on how they designed the system

5. Get Arun to make me that SST video game!

I HAVE (MANY) DREAMS

What was it that Yogi Berra said?

“It's tough to make predictions, especially about the future.”

What's Happening?

[Photo credit: iStock.com/danbaily/Jessica Marx]

It's not only merely dead, it's really most sincerely dead.
~Computer engineer who shall remain anonymous~

For 3^{+} decades we had huge, 'free' computing gains by transistor scaling – faster, lower energy, cheaper, consistent programming environment

1971: 4004 microprocessor -- 4 bit data path, 740 kilohertz clock, 10 µm
2005: Pentium -- 64 bit data path, 3.8 gigahertz clock, 90 nm

80,000x raw compute improvement (40 percent increase per year); primarily driven by a 12,000x reduction in transistor area.

Over the last 2 decades, we lost faster, lost a consistent programming environment, may have lost cheaper, and energy wins are greatly reduced

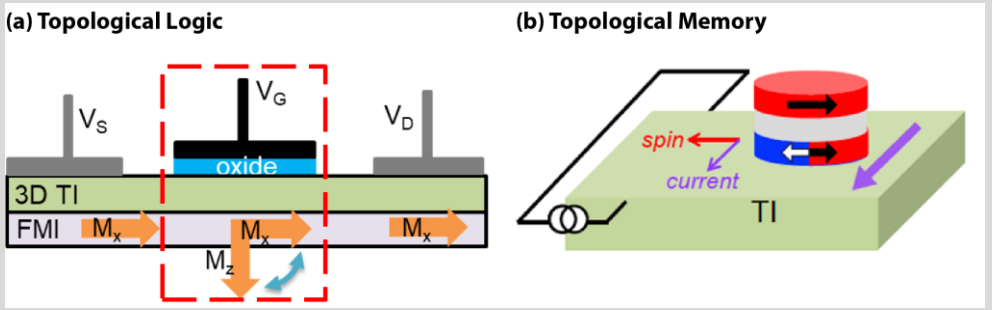
In these uncharted waters, Mod-Sim will become even more important

In-house, purpose built designs and systems are proliferating – both performance and $ wins
Co-design is a necessary methodology

Can mod-sim meet the challenges? Noel, this is your graduate coursework

The end of Moore's Law means lots of new ideas will be explored

- Increasing use of accelerators to provide energy efficient compute
- Moving compute to data – processors-in-memory, computational storage, smart NICs
- Advanced packaging (heterogeneous integration via chiplets, 3D stacking, wafer scale integration)
- Silicon photonics for interconnects
- New materials and devices
- Brand new computational paradigms
  - Quantum Computing
  - Neuromorphic Computing
  - Reversible Computing
  - Probabilistic Computing



From "Scalable energy-efficient magnetoelectric spin–orbit logic" by Manipatruni et al, Nature (2019).

Silicon Brain not required

Image: zmescience.com

Mod-Sim will need to provide extremely Usable, highly Capable, and easily Quantifiable tools

New resilience issues[1,2] are being created by extreme transistor scaling; they will need sophisticated capabilities to solve

1. H. Dattatraya et al, "Silent Data Corruptions at Scale," arXiv:2102.11245v1, Feb 2021.
2. P. Hochschild et al, "Cores that don't count," Workshop on Hot Topics in Operating Systems, 2021, Ann Arbor, MI.

USABILITY

I foresee tremendous opportunities for ML and Viz to make a real difference

NLP has the potential to vastly improve setting up simulations
Viz has the potential to vastly improve manipulating the simulation and analyzing the results

So I guess what I really want is … a holodeck

CAPABILITY

It needs to be much easier to develop a model – more ML please!

THE CONCEPT OF COMPATIBILITY/INTEROPERABILITY NEEDS TO BE MUCH MORE EXPANSIVE

Material Science leads to Devices/Circuits/Functions leads to Architectures leads to Systems & Applications that ultimately drive design, fabrication, and integration (HW + SW)

How do we link DFT with SPICE, SPICE with GEM5 and SST, and SST to CHISEL and PyRTL (plus …) to 'close the loop'?

Simple things to get you started

- Use SST simulation to create ISA documentation
- Use ISA documentation to drive a GEM5 architectural model

QUANTIFIABILITY

Sudhakar Yalamanchili, from the 2012 workshop

"If you can not measure it, you can not improve it.”

-Lord Kelvin

I propose ‘we’ create a contest

The details are left to the interested workshop attendee (perhaps his name is Bob)

Speed and Scale of the simulation are absolute musts

Performance is a minimum output

- Energy should be included if possible
- Resilience may still be a ‘bridge too far’

Fidelity and precision

Usability and interoperability should be considered

A contest needs a prize – that’s your problem, “Bob”

HEY, I'M NOT A COMPLETE CHEAPSKATE

If this becomes a reality (in the next 5 years)

I will host the creators of the contest (up to 4 people) to an Orioles Baseball game at Camden Yards – Tickets, food, and beverages (you supply your own transportation)

They should be a good team by then (I hope)

THANKS!

Image by David Mark from Pixabay


Laboratory for Physical Sciences
LPS

NOEL'S GOALS FROM 2013 – YOU BE THE JUDGE

1. Architectural Exploration – given an application or class of applications, how will architectural changes or new architectural approaches improve the system? System metrics of interest include energy, productivity, resilience, and cost.

2. Technology Impact – given a "reasonably" specific system concept(s), how will a new technology improve the system (energy, productivity, resilience, and cost)? Conversely, what capability thresholds does a technology need to meet to improve the system significantly?

3. Application evaluation – how well does a particular modification to an application (new algorithm, different programming model, new language) improve it (energy, productivity, resilience)?

4. System Optimization – given a fairly specific system concept and application, what is the best allocation of system resources (memory bandwidth, amount of storage, etc.) that achieves optimal system value (energy, productivity, resilience, and cost).

5. Early application development – given a specific system concept, enable applications to be examined and optimized in advance of system deployment.

6. SW stack validation – validate new SW (runtimes, OS, file system) developed in support of GP or SPD systems.

7. Vendor collaborations – provide quantifiable information to enable vendors to refine/optimize specific component R&D in support of future systems.

8. System acquisition and support – enable assessment of system capabilities as part of competitive selection or acquisition processes; verify correct system performance during operational lifetime and after upgrades.