Pacific Northwest
NATIONAL LABORATORY
*Proudly Operated by* **Battelle** *Since 1965*

# CENter for Advanced Architecture Evaluation (CENATE): A Computing Proving Ground


Adolfy Hoisie
Chief Computer Scientist and Laboratory Fellow
Pacific Northwest National Laboratory

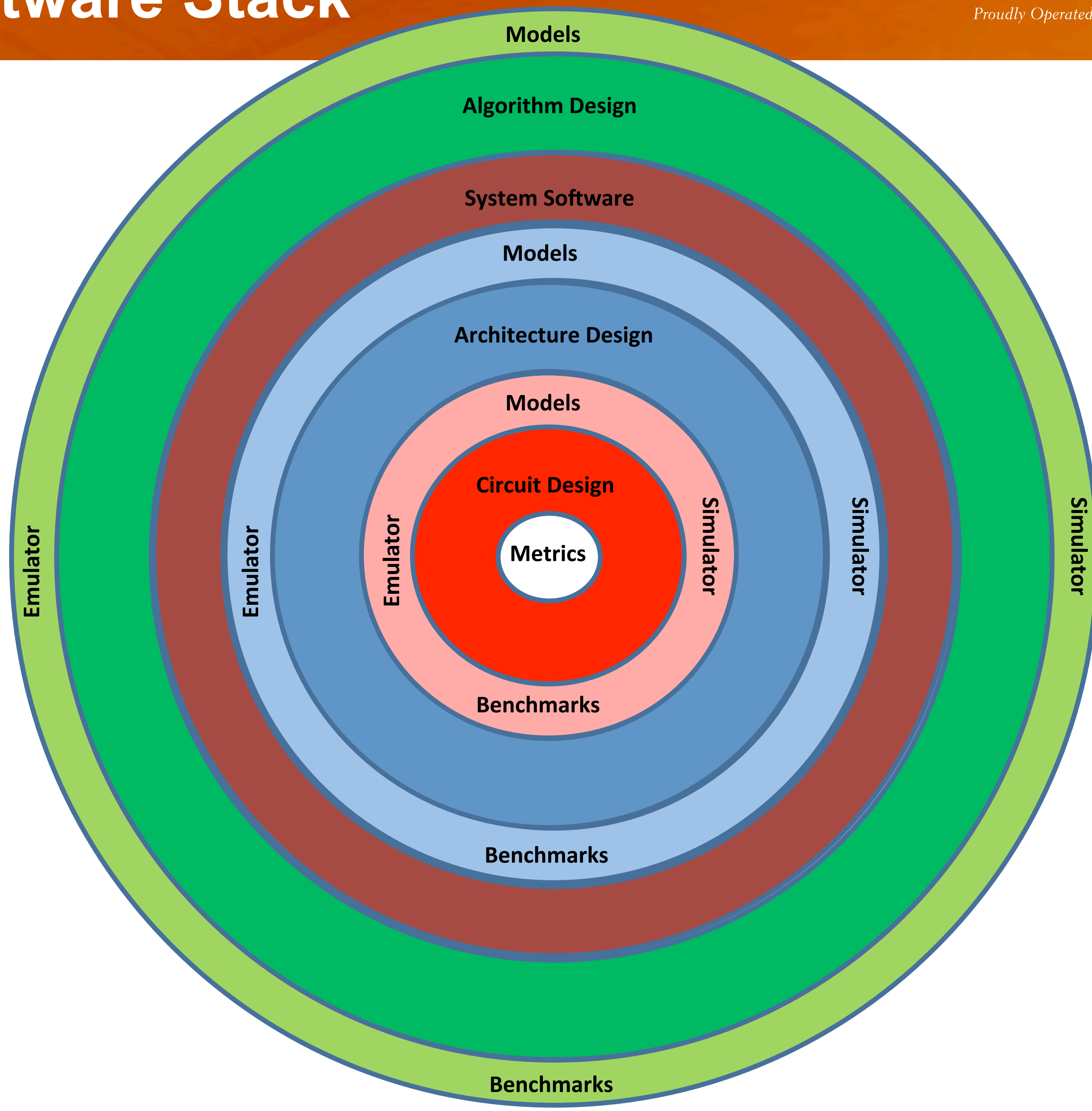
Work with: Kevin J. Barker, Ryan Friese, Roberto Gioiosa, Nitin Gawande, Darren J. Kerbyson, Gokcen Kestor, Andres Marquez, Matthew Macduff, Shuaiwen Leon Song, Nathan Tallent, Antonino Tumeo


ModSim Workshop, August 2016, Seattle

# Technology Fragmentation Across the Hardware/Software Stack

Pacific Northwest NATIONAL LABORATORY
*Proudly Operated by Battelle Since 1965*

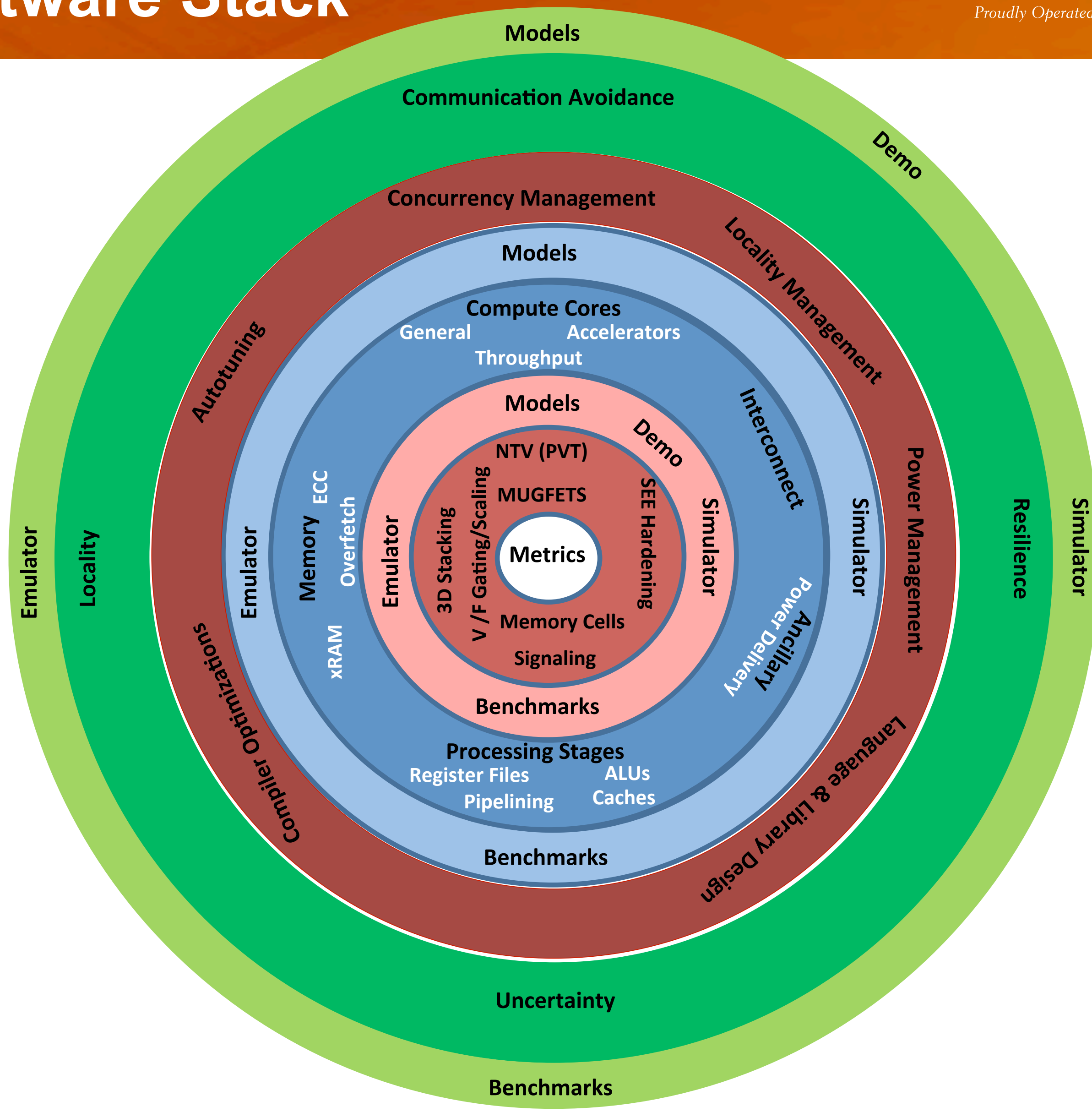
- Circuit Design
- Circuit Modeling & Simulation
- Architecture Design
- Architecture Modeling & Simulation
- System Software
- Algorithm Design
- Algorithm Modeling & Simulation



# Technology Fragmentation Across the Hardware/Software Stack

Pacific Northwest
NATIONAL LABORATORY
*Proudly Operated by* **Battelle** *Since 1965*

- Circuit Design
- Circuit Modeling & Simulation
- Architecture Design
- Architecture Modeling & Simulation
- System Software
- Algorithms
- Algorithm Modeling & Simulation



# Center of Advanced Technology Evaluation (CENATE)

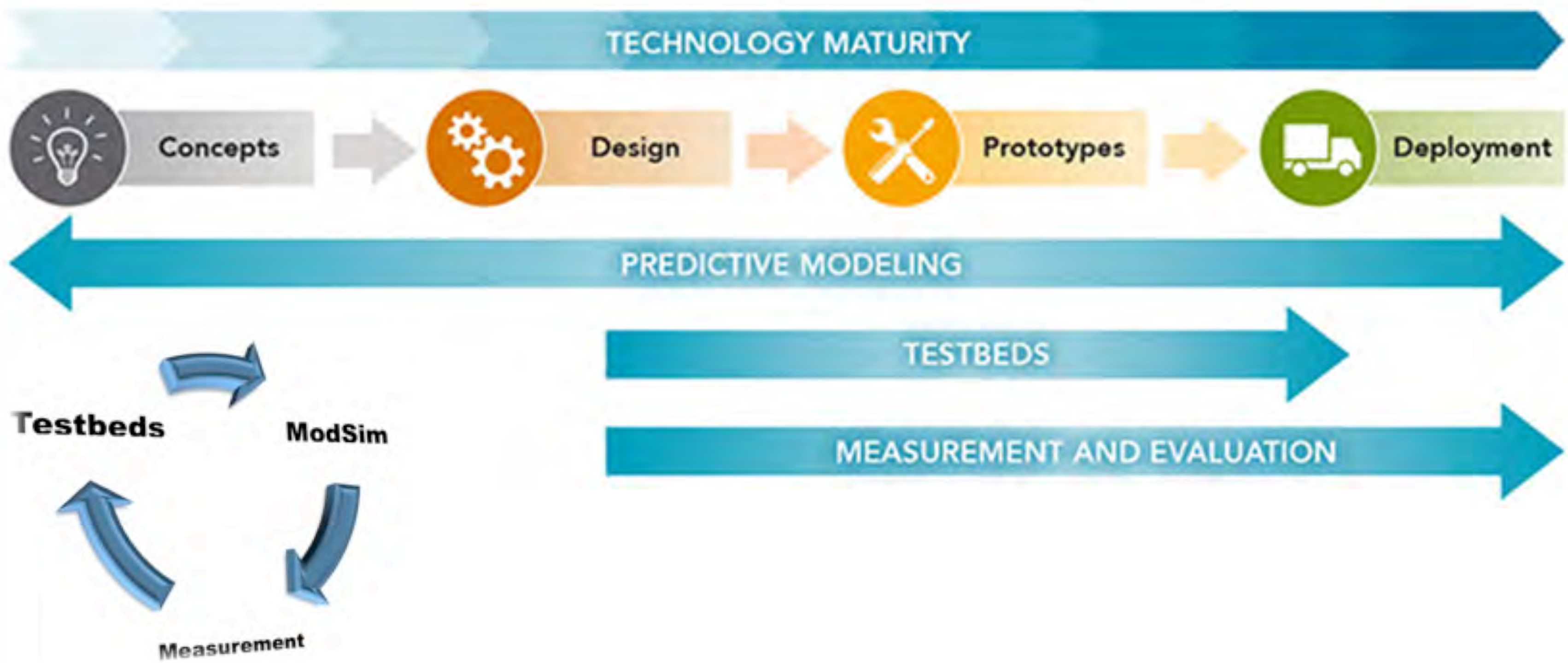
Pacific Northwest NATIONAL LABORATORY
*Proudly Operated by* **Battelle** *Since 1965*

- Advanced technology evaluations
- Instrumentation for power and performance
- Testbed infrastructure for high-throughput evaluation of technologies
- Predictive exploration: integration of results from empirical evaluations with modeling and simulation
  - Impact at scale; "what-ifs"

# CENATE Covers a Multidimensional Technology Space

Pacific Northwest NATIONAL LABORATORY

*Proudly Operated by* **Battelle** *Since 1965*

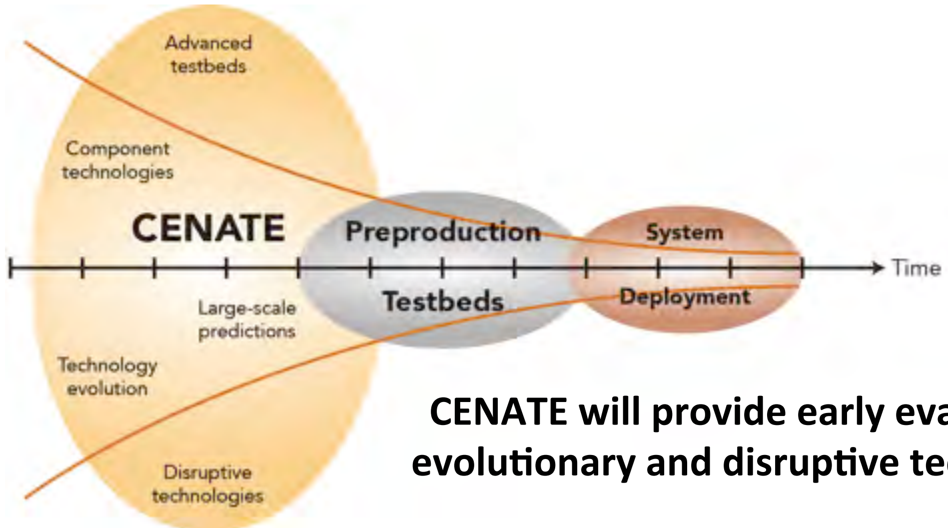



| | | | | |
|---|---|---|---|---|
| **Sub-systems** | Processing | Memory | Network | Storage and I/O |
| **Emerging Paradigms** | Approximate | Quantum | Neuromorphic | Superconductive |
| **Workloads** | Numeric | Machine Learning | Data Analytics | Graph Analytics |

# CENATE: Potential to Leverage

Pacific Northwest NATIONAL LABORATORY
*Proudly Operated by* Battelle *Since 1965*

- Testbeds targeted at several areas
  - System Technologies with future high impact to HPC
  - Novel Processing Paradigms beyond traditional computing
  - Emerging Technologies beyond Moore's Law
- Evaluation of workloads of interest
  - Scientific Computing
  - Irregular Applications
  - DOE-focused
- Integrated measurement and ModSim




**CENATE will provide early evaluation of evolutionary and disruptive technologies.**

# Advanced Measurement Laboratory (AML)

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- AML provides infrastructure to measure
  - Early Engineering Boards
  - Subsystem Prototypes (e.g., HMC)
  - Small Systems
- AML measures
  - Performance
    - Time-to-solution
    - Performance counters
  - Power
    - System Wall Power
    - Internal Shunts and Hall-sensors
  - Temperature
    - Thermo-Couples
    - Thermal cameras
- Building up FPGA capabilities
  - Xilinx and Altera Toolkits
  - Mentor Graphic's Modelsim

# SEAPEARL: Integrated Power, Performance, and Thermal Measurement



- Critical needs
  - Ability to study power consumption and thermal effects ***at scale***
  - Correlation of measurements to workload features (not steady state)
  - Platform for development of modeling and optimization capabilities
- SEAPEARL: A Unique Resource
  - High-fidelity power measurement
    - Spatial: separate CPU from memory
    - Temporal: low sampling period of 1 ms
  - Coupled thermal information
  - Advanced architectures: x86 multicore and AMD Fusion (integrates CPU and GPU)
- Offline analysis and potential for online (dynamic) optimization

# CENATE: Establishing Best Practices for Measurements

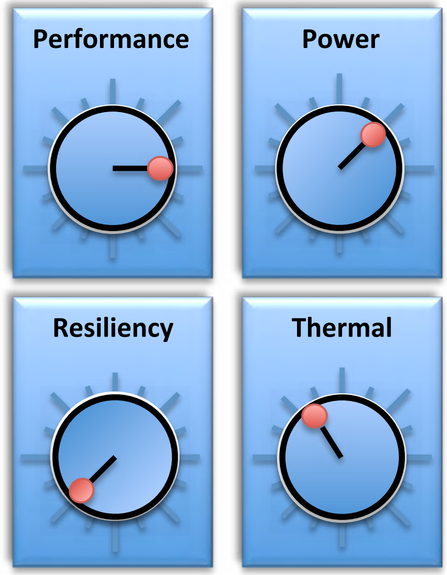
Pacific Northwest NATIONAL LABORATORY
*Proudly Operated by* **Battelle** *Since 1965*

- Instrumenting systems requires knowledge of what “instrumentation hooks” are available:
  - Integrating state-of-the-art measurements into idiosyncratic systems
  - Best practices: Measurement is a science and a craft
- Multi-tier approach:
  - Tier 1: external, low-resolution, available to all systems
  - Tier 2: internal, system specific, provided by vendor (e.g., RAPL, Amester, Data Vortex thermal)
  - Tier 3: external, high-resolution, invasive, need vendor support (e.g., Penguin Power Insights)
- Measurements:
  - In-band: synchronous with the application (e.g., performance counters)
  - Out-of-band: asynchronous with the application (e.g., power meters)

# Instrumentation Space of Interest

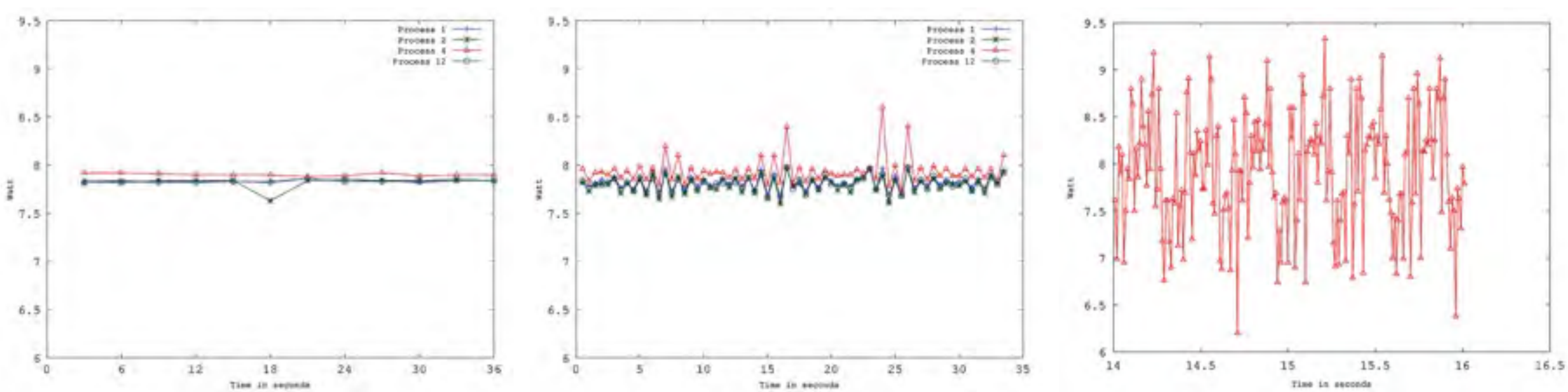


- Measurements of interest:
  - Performance (e.g., system: FLOPS, Memory Access/s, Application: TEPS…)
  - Power/Energy
  - Thermal
  - Reliability: Soft/Hard errors
- Interplay across dimensions
  - Measurements are not independent
  - System components are not isolated
  - Workload impacts interplay
    - Power/thermal throttling -> performance
    - Performance (e.g., high IPC) -> power
    - Temperature -> soft errors
    - Hard errors -> performance (e.g., reduced bandwidth)
  - Need to isolate effects through carefully designed experimentation
- Application behavior is ultimate goal



# Instrumentation Granularity Affects Insight

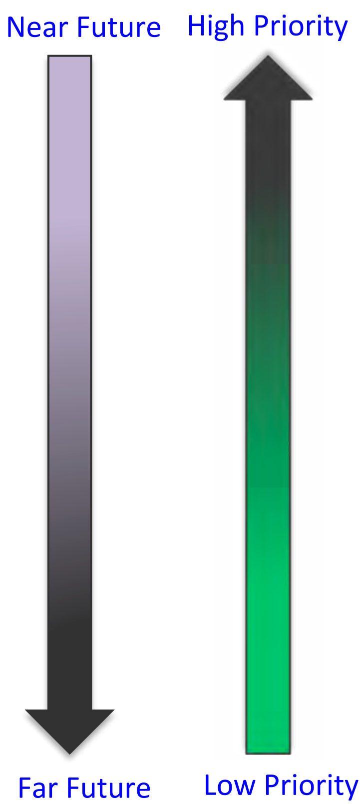
- Coarse *spatial* and *temporal* instrumentation may hide important information
  - e.g., peak power/temperature consumption
- Example for scalar pentadiagonal solver with 32 parallel threads
  - Peak power measured with 0.1 second granularity is much higher (9.7 W/core) than the one measured with 1 second granularity (7.8 W/core)



1 second | 0.5 second | 0.1 second

# Testbed Classification






- Testbed Classes
  - System Technologies with future high impact factor to HPC
    - Computing (Throughput or Latency Optimized Computing , Processing Near Memory)
    - Memory (Phase-Change, Spin Transfer Memory, Memristor, 3D DRAM, 3D Flash)
    - Networks (single packet support *Data Vortex*, collectives support *Mellanox*, Software Defined Networks)
    - Storage (high bandwidth/low latency buffers, fast container support *netCDF/ HDF5*, Software Define Storage)
  - Novel Processing Paradigms beyond traditional computing paradigms
    - Neuromorphic Processing (Spatio-Temporal Dentrite Processing (STDP))
    - Approximate Computing
    - Quantum Computing (topological quantum computing)
  - Emerging Technologies (Materials -> Devices) beyond Moore's Law transitioning towards rudimentary logic/memory cell and circuit basic blocks
    - Superconductive Processing (Single Flux Quantum (SFQ), Adiabatic Quantum Computing (AQC))

# Memory: Novel Memory Architectures



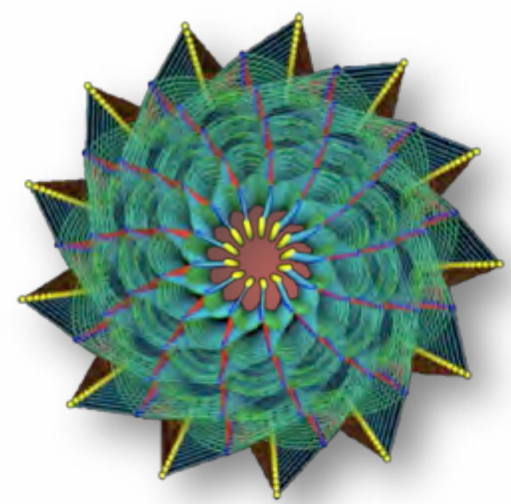
- Novel memory technology benefiting existing systems
  - Exploration of alternative memory solutions – in many cases persistent
    - Phase Change Memory
    - Spin Transfer Memory
    - 3D Memory
    - Nvram
    - Memristor
  - Analyze memory performance/power for memory execution patterns that can leverage fast, persistent memory

# Networks: Data Vortex



- Novel network technology based on topology fundamentals
  - Ensures high probability of contention-free transport
  - Enables very small packet transport without paying an overhead premium
  - Analyze network performance/power for applications of interest, *s.a.*, high-dimensional PDE or FFT solvers or finely partitioned, data-intensive applications
  - Four end-point system expected July 2016

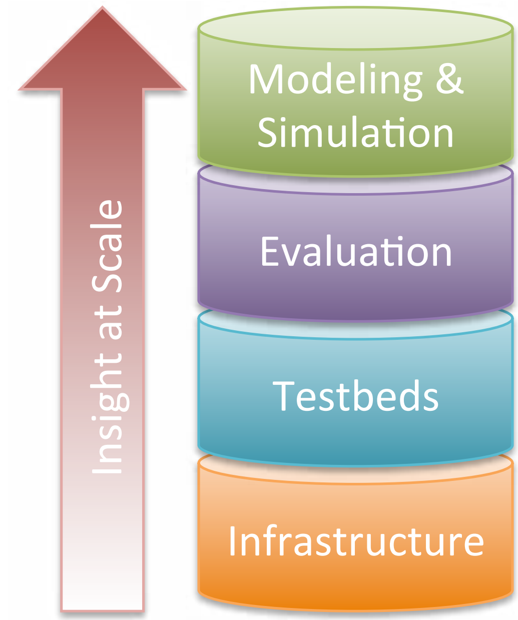




# ModSim within CENATE

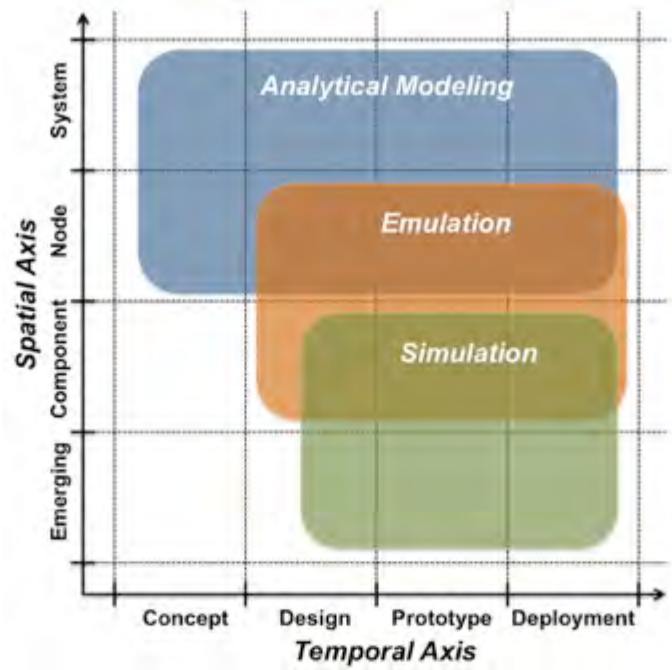


- Modeling and Simulation will be used to explore:
  - System scales that cannot be directly measured
  - Systems integrating disparate technologies
  - Multiple alternative system configurations
- Quantify trade-offs between multiple metrics of interest:
  - Performance
  - Power and energy consumption
  - Impact of thermal variation, faults, and fault mitigation
- Modeling builds on the CENATE foundation:
  - Application-centric models are derived from workload applications
  - Models are parameterized using measurements taken on instrumented testbeds (micro-benchmarks isolate “atomic” performance characteristics)
  - Models are validated at small-scale
- Key contribution of modeling is insight:
  - Rapid turnaround from system specification to performance quantification
  - Issues in performance can be traced to root causes
  - Quantify interplay between application characteristics and system



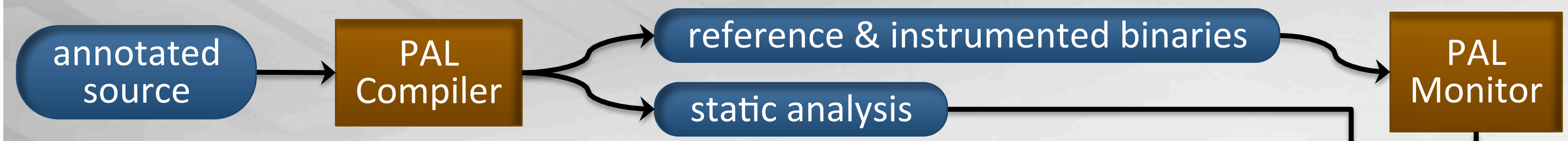
# Integrated ModSim Approach

- Across the spatial axis
  - Tools tradeoff between rapid evaluation and high precision
  - High-fidelity, low-level simulations serve as input into more abstract system models
- Across the temporal axis
  - More abstract techniques are able to use partial or incomplete specification information
  - Greater certainty in the design allows for the use of more precise prediction tools
- Employ a “bag of tools” approach in which ModSim technologies are applied where they are most appropriate
- CENATE will employ:
  - PALM: Automated analytical modeling tool for full applications at large scale
  - Cross-Roofline Model: Node-level modeling of throughput-oriented architectures
  - P-McPAT: Power modeling at the micro-architectural level
  - Prometheus: Emulation tool for node-level non-deterministic workloads

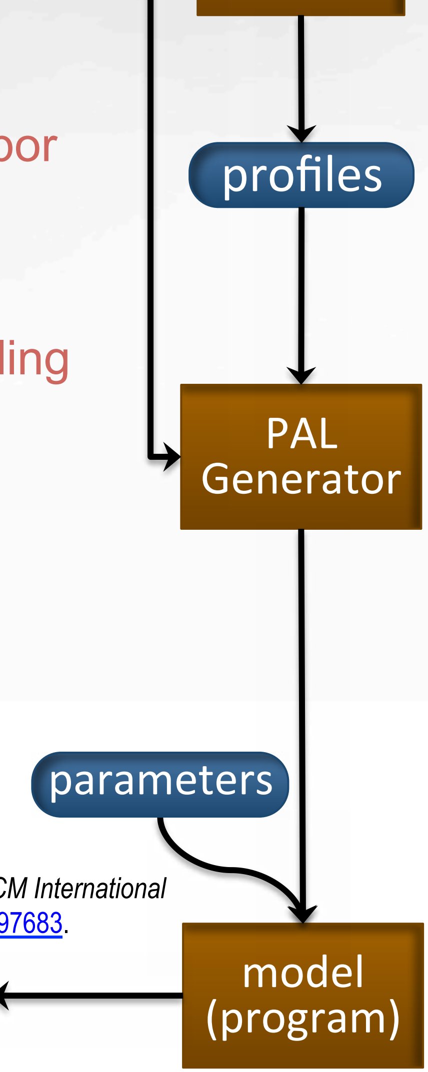


# Tools for Analytical Modeling of Performance

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965




- Application modeling is difficult
  - Creating and validating models requires expertise and labor
  - Reproducing and distributing models is ad hoc
- Palm: PAL Modeling tool
  - Annotation language expresses insight and guides modeling
    - Develop model and application in tandem
    - Decompose modeling task into sub-problems
  - Generate model from static/dynamic annotation structure
    - Name, capture, and use dynamic application structure
  - Generate same model given same input (reproducible)
  - Generated model is executable program (distributable)

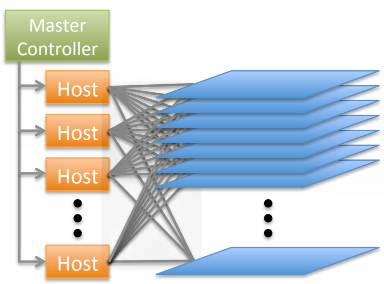



Tallent NR and A Hoisie. 2014. “Palm: Easing the Burden of Analytical Performance.” In *ICS'14 Proceedings of the 28th ACM International Conference on Supercomputing*, pp. 221-230. Association for Computing Machinery, New York. DOI: 10.1145/2597652.2597683.

# Modeling Possible Future Silicon Photonics Networks

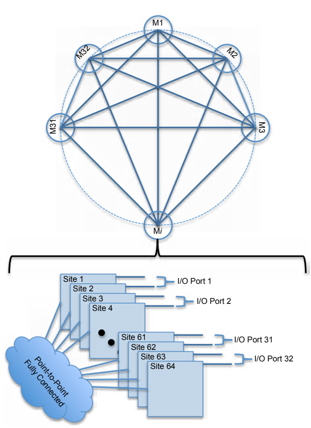
Pacific Northwest NATIONAL LABORATORY
*Proudly Operated by* **Battelle** *Since 1965*

- Disparate technologies from IBM (internode) and Oracle (intranode)
- Modeling enabled:
  - Possible “marriage” options to be explored overcoming separation barriers
  - Quantified advantages over expected future electrical networks
  - Analyzed in the context of key graph analytic applications

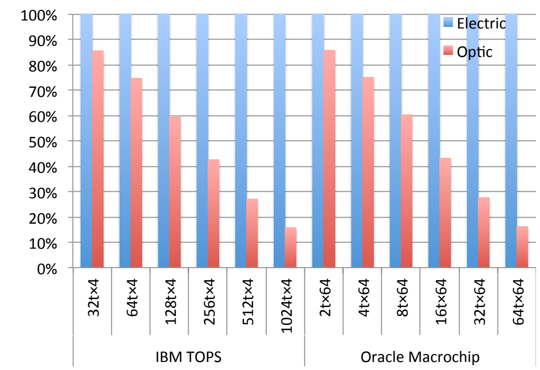



64 Compute Nodes

256 64×64 Optical Switch Planes




- IBM TOPS inter-node network
  - 64 node system
  - 256 of 64 x 64 optical switch planes
    - 16 wavelengths per fiber
    - 20 GB/s BW per wavelength
- Oracle Macrochip intra-node network
  - 64 compute/memory sites fully connected
  - 2 GB/s per site pair (128 GB total)
  - 32 ports I/O per macrochip (for internode)
- Improvement due to:
  - Improved link bandwidth
  - Greater link *concurrency*
  - Varied topological routing




**Parallel Community Detection Relative Execution Time**

# ModSim as an Integrating Methodology



PAL

- Data Vortex network fabric
  - Provides contention-free routing for small packet sizes
  - Currently, nodes have small core counts due to limitations on memory bandwidth to feed the network
- Micron's Hybrid Memory Cube (HMC)
  - 3D stacked memory architecture results in high levels of memory bandwidth
- CENATE's modeling approach will allow for the "virtual" integration of technologies such as these
  - e.g., How many cores can a Data Vortex system equipped with HMC support per node?
- Integrated modeling approach requires
  - Detailed understanding of application characteristics (e.g., what are the data movement patterns exhibited by a particular workload?)
  - Detailed understanding of architecture characteristics and capabilities
  - Integration of ModSim techniques across scales

# Conclusions

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Adaptivity of systems and system and application software requires an “introspective” approach to design and optimization
- Integrated measurement and ModSim for performance, power, and thermal aspects through CENATE
- Measurements serve as input for models, and as the validation means for them
- Tackling a wide space of technologies in terms of maturity and architectural coverage
- Testbeds of technologies
- ModSim for “union of technologies”
- Invited to access and use CENATE resource



- **Work sponsored by the U.S. Department of Energy, Office of Advanced Scientific Computing Research, CENATE contract, 2015.**