# EVALUATION, OPTIMIZATION AND APPLICATION OF EXECUTION MODELS FOR EXASCALE COMPUTING

**Exascale computing for DOE mission-critical applications face daunting hardware and software challenges over the next decade. The application requirements for sustained performance, energy efficiency, dependability, and programmability are unlikely to be satisfied by incremental extensions to conventional practices. Ultimately, a paradigm shift in computing will be essential to enable practical exascale performance by 2020. A new paradigm for computation is manifested as an execution model, which provides the framework for deriving a consistent set of system elements and forms the basis for hardware/software co-design.**

The goals of our research are to define the salient components for an effective co-design process, devise a methodology of co-design for DOE exascale projects, and develop an experimental execution model as a proof-of-concept foundation for exascale computing systems and applications. We will examine potential exascale execution models and determine their impact on exascale system performance and energy for DOE mission-critical applications. This project will provide DOE with the essential techniques to derive, implement, deploy, and apply one or more revolutionary execution models to enable exascale computing before the end of this decade. The strategy for the research integrates two concurrent and complementary approaches within a single project focusing on both a bottom-up (EMBU) and a top-down (EMTD) approach.

EMTD is related to the development, validation, modeling, analysis, and quantitative comparison of advanced execution models for full DOE applications for specific Exascale architectures point design studies. EMBU will start from concrete examples of execution models and hardware and use architectural simulation to evaluate these models. In support of this the execution model toolkit (EXEMT) will be developed that will be a collection of both fine- and coarse-grained models for execution model components and will be rich enough to construct a variety of execution models of interest for exascale computing. Compact and skeleton applications will be developed utilizing EXEMT to implement candidate execution models and to compare their estimated time and energy to solution on Exascale machine models.

Custom hardware features will be explored as a mechanism to accelerate the execution models. A quantitative, predictive co-design methodology for comparing execution models will be derived through both EMBU and EMTD relying heavily on prior art and on the expertise of the team and will be applied in the context of full applications in production use on current leading ASCR supercomputers, and targeted to evolve to exascale through the work of the Co-Design Centers and other DOE SC Exascale projects. These findings will be used in part to engage processor vendors and the ASCR Exascale Co-design Centers.

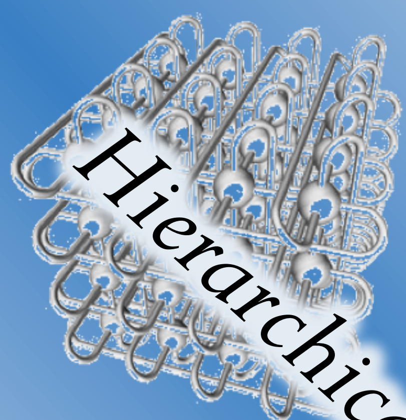
*This research is supported by Department of Energy Advanced Scientific Computing Research*

*Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.*

*The Pacific Northwest National Laboratory is operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RL01830.*

*LBNL work was supported by the Office of Science, Office and Advanced Scientific Computing Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.*

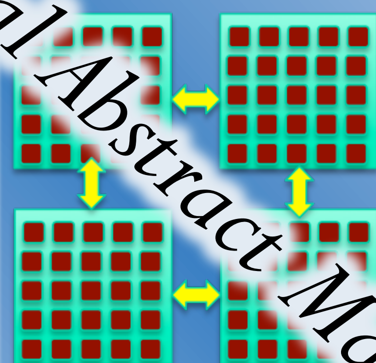
# TOP-DOWN APPROACH



Hierarchical Abstract Machine Model

The Top-Down approach aims at modeling Execution Models using accurate *modeling* methodologies. A distinguishing feature of EMTD is that the models will be based on full applications, as to preserve the workload characteristics of the representative DOE apps selected.

The Execution Models will be parameterized for a comprehensive set of attributes, and these parameters will serve as input to the models



Using this approach we will compare novel Execution Model concepts for the full applications optimally mapped onto Exascale architectures being considered by ASCR.

**SST/macro:**
**Simulation of different Interconnect Architectures**

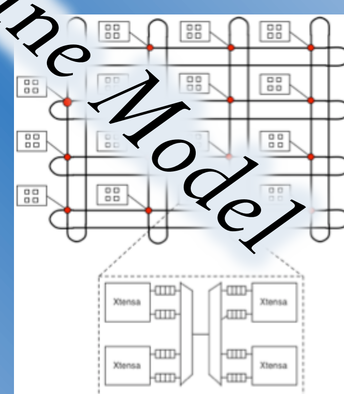
- Driven by traces collected from full application or skeletonized code (either manually or via ROSE)
- Determines degree of congestion on the network and the impact on application performance

**Application Surrogates for Co-Design**

- Use a simplified version of an application for exploration in a restricted manageable design space (inexpensive, fast turnaround), before trying out new approaches in a full application (expensive, multi-year effort).

**RAMP/GreenFlash:**
**Chip-level Simulation**

- Extend GreenFlash/RAMP simulation for more general proxy model.
- Provides model-checking for energy models offered by software simulators.
- Evaluate novel hardware support for alternative execution models.




# BOTTOM-UP APPROACH

***Project Team***

***Pacific Northwest National Laboratory:*** *Adolfy Hoisie (lead for the top-down approach), Kevin Barker, Daniel Chavarria, Darren Kerbyson, Joseph Manzano, Abhinav Vishnu*

***Sandia National Laboratories:*** *Robert Clay (lead for the bottom-up approach), Gilbert Hendry, Mike Heroux, Joe Kenny*

***Lawrence Berkeley National Laboratories:*** *John Shalf, Nick Wright*

***Indiana University:*** *Matt Anderson, Thomas Sterling*