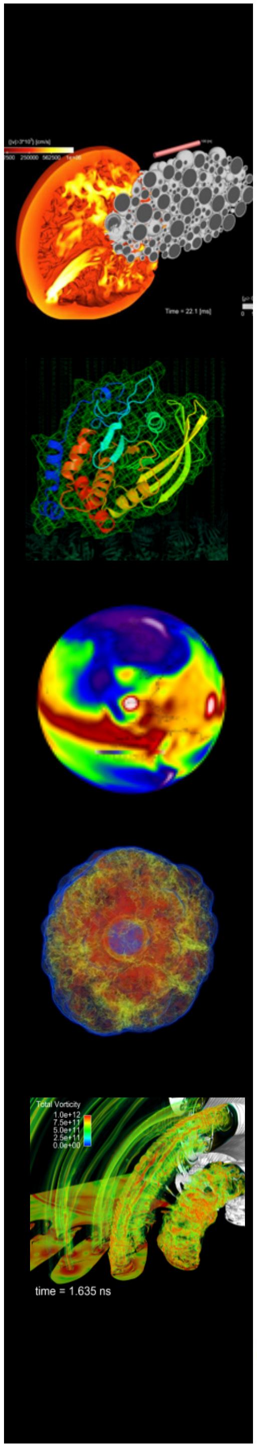

APEX

Alliance for application Performance at EXtreme scale

Vendor Roadmap Presentation Guidance

January, 2015

LA-UR-15-27422

What is APEX?

- ***Alliance for application Performance at EXtreme scale***
  - **As the acronym suggests our focus is on the Application**
  - **Performance has many dimensions**
- APEX is a collaboration between ACES (ASC) and NERSC (ASCR)
  - ACES (**A**lliance for **C**omputing at **E**xtreme **S**cale) is a collaboration between Sandia National Laboratories (SNL) and Los Alamos National Laboratory (LANL)
- The APEX collaboration is intended to result in the procurement of two platforms
  - NERSC/ASCR procurement of NERSC-9
  - ACES/ASC procurement of Crossroads (the 3rd Advanced Technology System)
    - ATS-1 = Trinity, ATS-2 = Sierra (CORAL), ATS-3 = Crossroads
- Both platforms will focus on meeting both mission needs and pursuing Advanced Technology concepts

High-level Design Philosophy

- Delivered application performance (as APEX suggests) is the primary driver in support of mission requirements
  - Peak FLOPS requirement will not appear in RFP
- APEX plans to purchase 2 platforms
  - Crossroads and NERSC-9
- Both target delivery in FY20
- Advanced technology development is assumed to be necessary to meet mission needs
  - Accelerate development of yet to be identified key technologies
  - 3rd round of NRE – (Trinity/NERSC-8, CORAL, APEX)
- Considered pre-exascale platforms
  - MUST support path to exascale programming models
    - While supporting existing mission needs
  - Support MPI+OpenMP (threads)
    - Matured on Trinity/NERSC-8 and CORAL platforms
  - Additional support for other, yet to be identified, MPI+X programming models

Capability Improvement

- An increase in predictive capability requires increases in the fidelity of both geometric and physics models
  - This implies usable large platform memory capacity
- APEX must demonstrate a significant capability improvement
  - Improvement measured relative to Trinity (ATS-1) and Cori (NERSC-8)
  - Improvement as a function of performance (total time to solution), increased geometries, increased physics capabilities, power/energy efficiency, resilience and other factors
- Previous DOE investments assumed to be an integral part of production computing for APEX.
  - Trinity/NERSC-8 NRE projects: Burst Buffer and Advanced Power Management
  - Fast Forward and Design Forward

Facility, Power & Cooling

- Crossroads will be located in the Nicholas C. Metropolis center (SCC) at Los Alamos National Laboratory
- NERSC-9 will be located in the Computational Research and Theory (CRT) facility at Lawrence Berkeley National Laboratory
- Estimated facility power and footprint
  - Crossroads
    - 15MW
    - 8000 square feet
  - NERSC-9
    - Power and floor space likely not primary platform constraints
- Liquid cooled
  - Is our assumption correct?
  - Warm water or chilled ? Direct or indirect?

Guiding Questions
guiding not exhaustive

- Basically we want to understand your roadmap(s) in the timeframe we anticipate taking delivery (FY20)
- Your roadmap presentations should NOT be limited to these guiding questions
- Tell us where and why our assumptions are wrong!
- We assume multi-level memory (storage) hierarchy
  - What will this look like?
  - Will it extend beyond the node?
  - Bandwidth and latency characteristics (between levels)?
  - Technologies?
  - Capacity?
  - Relative cost and energy trade-offs?
- What does a processor(s) look like on a node?
  - How many cores?
  - Heterogeneous or Homogeneous?
  - Core characteristics
  - NUMA characteristics?
  - Coherency?

Guiding Questions
(continued)

- NIC
    - Integrated or discrete?
    - Injection bandwidth?
    - Message injection rate?
        - At what message size(s)?
    - Offload characteristics?
    - Access to memory?
- Interconnect
    - Topology?
    - Physical layer?
    - Bisection bandwidth?

Guiding Questions
(continued)

- Software
  - Languages
  - Programming Environments
  - Programming Models
  - Profilers and Debuggers
  - Operating system(s)
  - Advanced Power Measurement and Control
  - RAS and/or System Management
  - Software to aid resiliency
  - Workload (and workflow) management

Guiding Questions
(continued)

- What will the file system look like?
  - Integrated into memory hierarchy?
  - Is traditional application driven check point restart still required?
  - How can we optimize for analysis usage models?
- Support for task based programming model(s)?
- What are the advanced resilience mechanisms?
  - Hardware and/or software
- What is the optimal way to support emerging data intensive computing workloads on the same platform as 'traditional' HPC ones?
- Will you have early test platforms / proxies available that we can explore these issues with?
- What are your proposed NRE areas?
  - and required lead times for each
- How can APEX best influence your roadmap?