

DOE's Fast Forward and Design Forward R&D Projects: *Influence Exascale Hardware*

James A. Ang, Ph.D.
Manager, Scalable Computer Architectures
Sandia National Laboratories
Albuquerque, NM

University of Florida
CCMT Exascale Deep Dive Workshop
Gainesville, FL
February 3-4, 2015

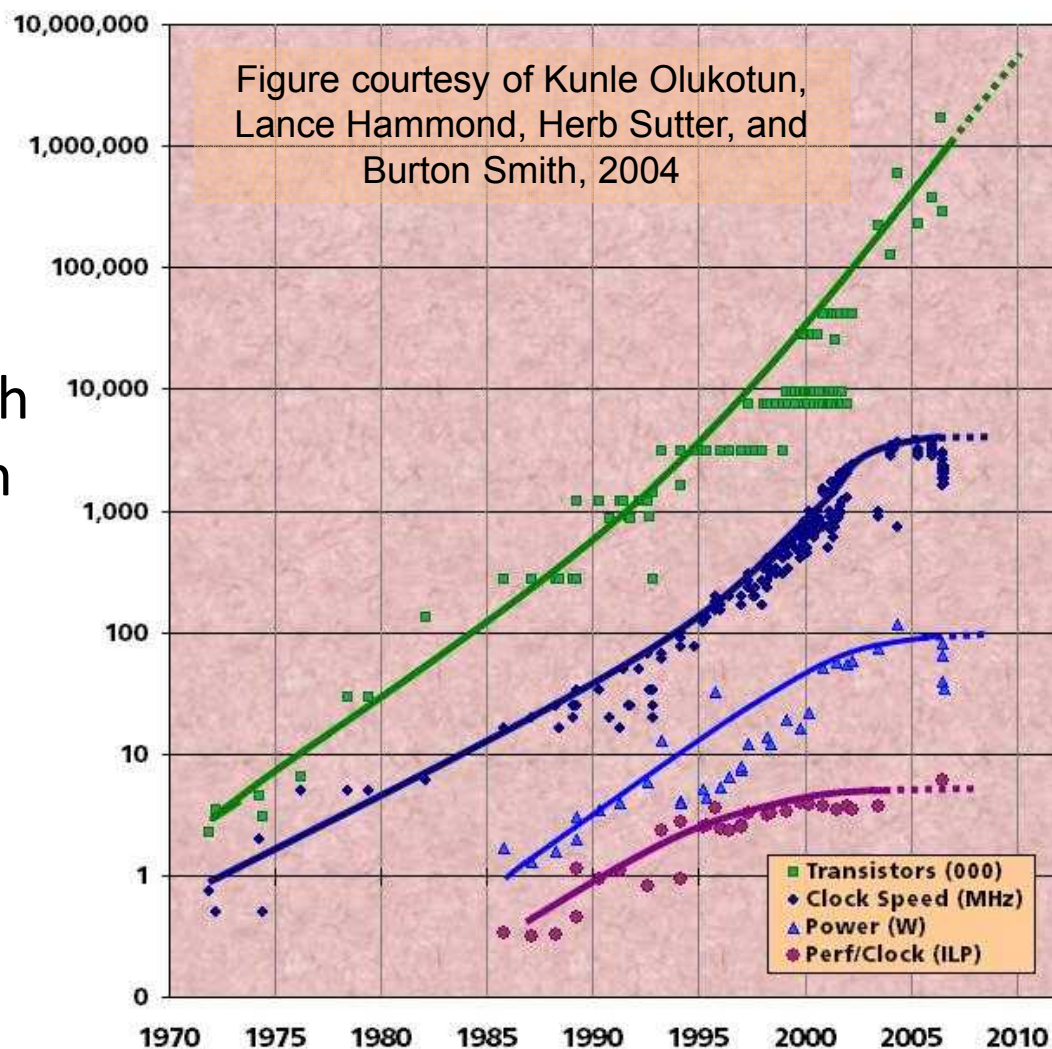


*Exceptional
service
in the
national
interest*



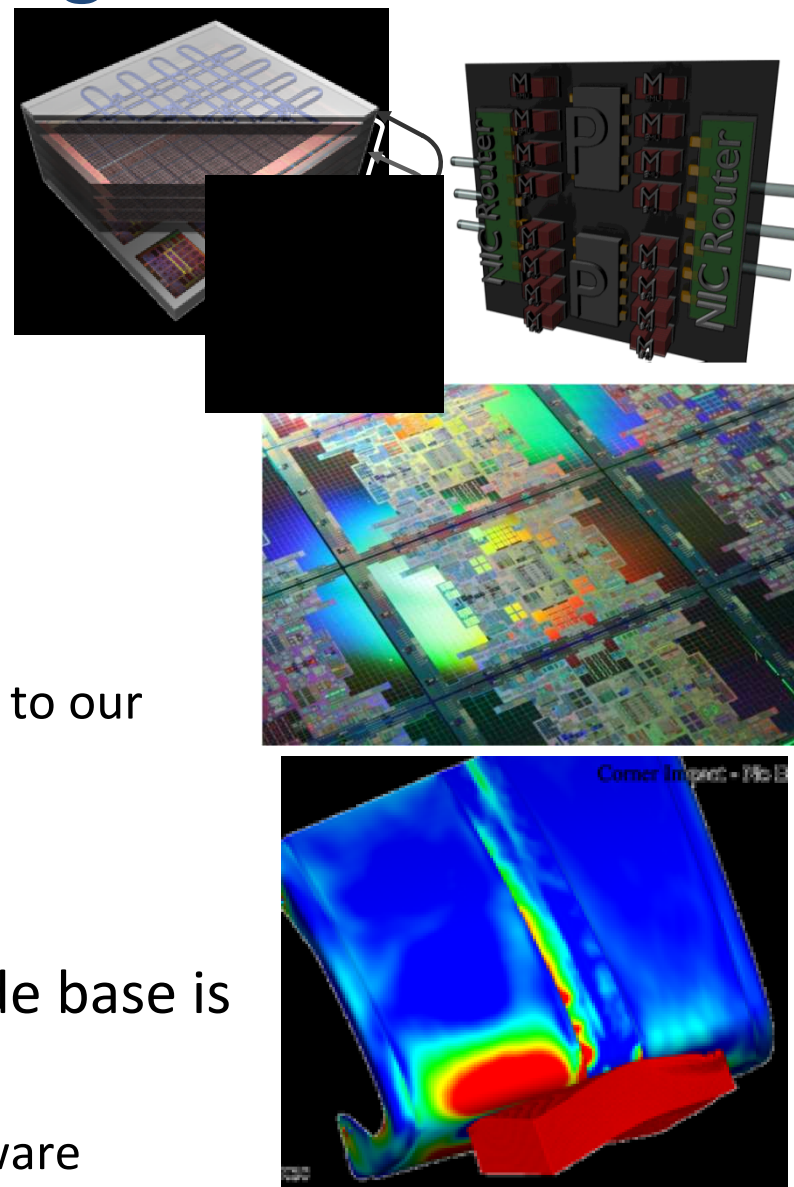
Exascale Hardware Challenges

- Left to the *Invisible Hand*
 - Industry follows an evolutionary path focused on Peak Flops
- In the Era of Dennard Scaling our *ad hoc* approach to integration of MPPs with COTS microprocessors was acceptable
- With the end of Dennard scaling, this is no longer able to meet DOE Mission Application Requirements



Exascale Hardware Challenges – cont.

- We need to Motivate and *Influence* Architectural Changes
 - Processor/node Architectures
 - System Architectures
- Our Investments are not only in Architectures
 - We cannot just develop new Exascale Architectures and *Throw it over the wall* to our application developers
 - We need Hardware/Software Co-design
- The transition of the DOE Legacy Code base is another important challenge
 - Challenge should influence future hardware thru Co-design



Industry Engagement is Vital

- **We need industry involvement**
 - Avoid one-off, stove-piped solutions
 - Continued “product” availability and upgrades beyond DOE support

- **Industry cannot and will not solve the problem alone**
 - Business model obligates industry to optimize for profit, beat competitors
 - Industry investments heavily weighted towards near-term, evolutionary improvements with small margin over competitors
 - Industry funding for long-term technology R&D is limited and constrained
 - Industry does not understand DOE Applications and Algorithms

- **How can we impact industry?**
 - Work with those that have strong advocate(s) within the company
 - Fund research, development and demonstration of long-term technologies that clearly show potential as future mass-market products (or product components)
 - Corollary: do not fund product development (as part of DOE R&D portfolio)
 - Industry will incorporate promising technologies into future product lines

NNSA/ASC and SC/ASCR are partnering to Influence Industry



- Aligned Hardware Architecture Efforts
 - April 2011 MOU signed between SC and NNSA
 - July 2011 Issued RFI on Critical Technologies for Exascale
 - July 2012 Established Fast Forward node-level Critical/Cross Cutting Technology R&D projects
 - October 2013 Established Design Forward interconnect R&D Projects
 - November 2014 Fast Forward 2: Exascale Node Designs
 - TBD: Design Forward 2: Conceptual Designs of Exascale Systems
 - Aligned joint Advanced Technology platform procurements
 - CORAL: Oak Ridge, Argonne, and Lawrence Livermore National Labs
 - APEX: Los Alamos, Lawrence Berkeley and Sandia National Labs

Fast Forward Program

- Objective: *Accelerate transition of innovative ideas from processor and memory architecture research into future products*
- Evaluate advanced research concepts and develop quantitative evidence of their benefit for DOE applications—using Proxy apps and collaborating on Co-design
 - Engage DOE application teams to understand technology trends/constraints (how it impacts their code development)
 - Understand how to *program* these new features
 - Quantitative evidence to lower risk to adoption of innovative ideas by product teams
- Critical Node Technologies and Designs for Extreme-scale Computing

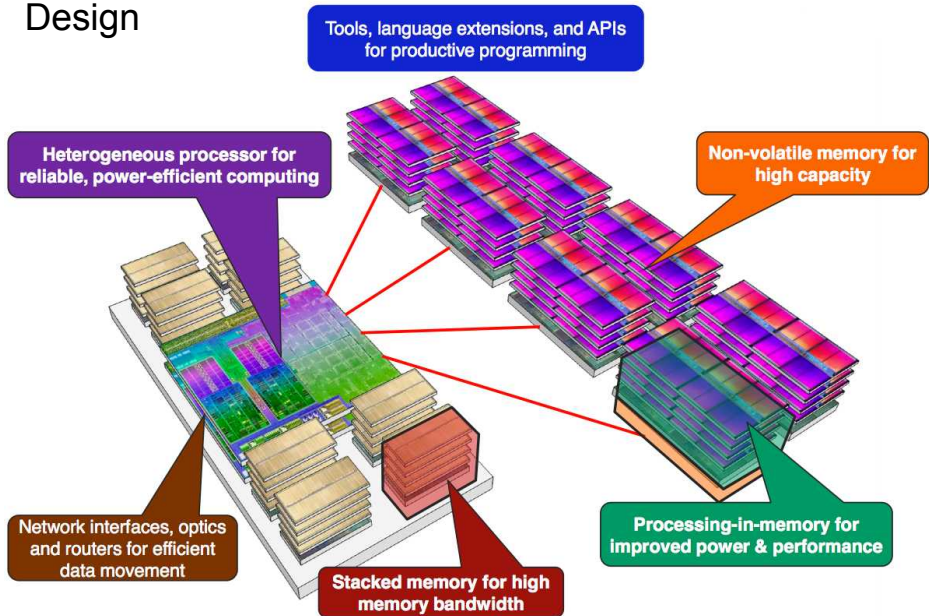
Fast Forward Program

- Fast Forward 1 (July 2012 – Sept. 2014)
 - AMD: Heterogeneous processor, Processing-in-memory and 2-level Memory
 - IBM: Advanced Memory Concepts
 - Intel: Core energy efficiency and Processing-near-memory
 - Intel/Whamcloud: Storage reliability, I/O API, burst buffer management
 - Nvidia: Memory hierarchy, processor/packaging/programming
- Fast Forward 2 (Nov 2014 – 2016)
 - AMD: Near threshold voltage logic, other low-power computing technologies, and new standardized memory interface
 - Cray: alternative processor design points including ARM microprocessors
 - IBM: investigate next-generation standardized memory interface
 - Intel: energy efficient node and system architectures, including software targeted at developing extreme scale systems
 - Nvidia: focus on energy efficiency, programmability and resilience

Design Forward Program

- Objective: R&D of interconnect architectures and conceptual designs for future extreme-scale computers:
 - Oct. 2013–2015, Design Forward 1: Interconnect Networks
 - Overall Interconnect Architecture
 - Interconnect Integration with Processor and Memory
 - Multiple Communication Library Progression and Interaction
 - Interconnect Fabrics and Management
 - Protocol Support
 - Scalability
 - Start is imminent, Design Forward 2: System design and integration
 - Overall System Architecture
 - Energy Utilization
 - Resilience and Reliability
 - Data Movement through the System
 - Packaging Density
 - System Software
 - Programming Environment

Concept Node Design



Processor Research

- Heterogeneous nodes which blend CPU and GPU cores
- Improved energy efficiency
- Efficient communication and data movement across the die
- Simplified programming models

Memory Research

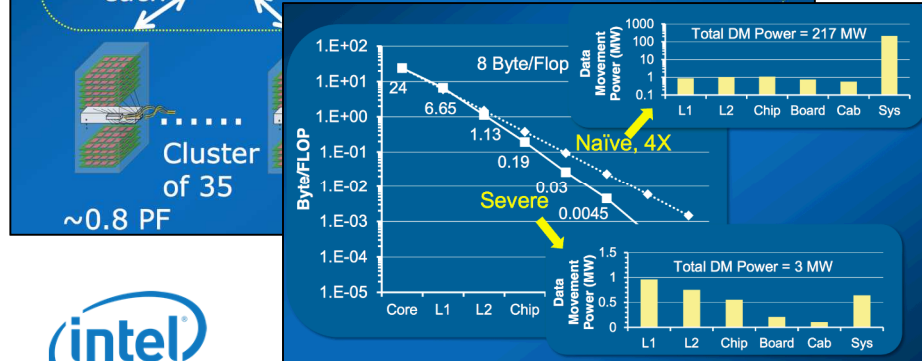
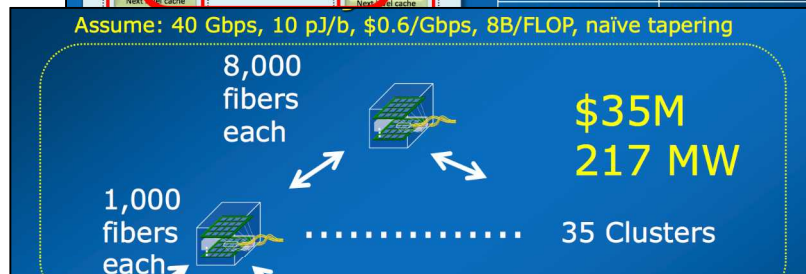
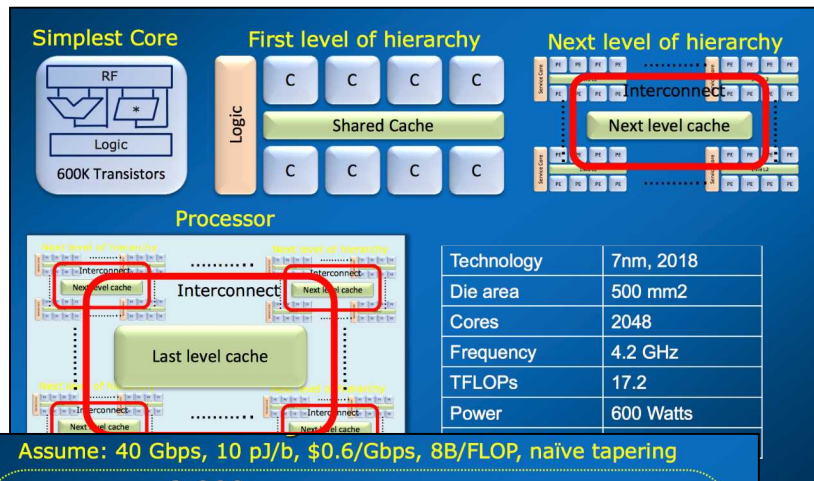
- Investigating new memory technologies
- Reduced data movement
- Higher performance
- Reduced energy consumption
- *New Memory Interface*: standardized, robust interface to support integration of heterogeneous memory and cores

Software Tools

- HSA Foundation



Source: AMD FastForward Project Overview (<https://asc.llnl.gov/fastforward/AMD-FF.pdf>)



Processor Research

- Lightweight processor cores
- Fast synchronization
- Specialized aspects of ISA and processor for data movement
- Tapered access to memory

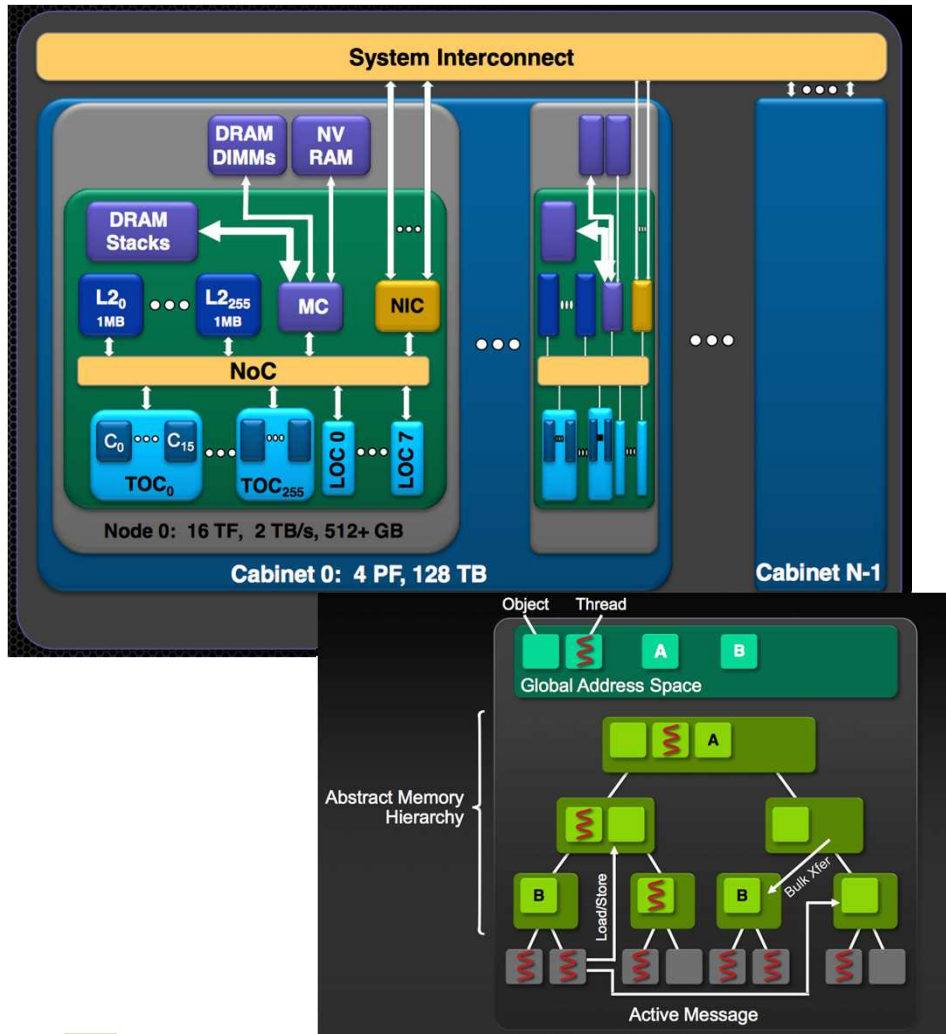
Interconnect Research

- Tapering bandwidth networks
- Integration of NICs into processor
- Intelligent data movement to reduce power

Software Tools

- Open Community Runtime (OCR)
- Exploration of OpenMP and MPI as legacy environment

Source: Intel FastForward Project Overview (<https://asc.lnl.gov/fastforward/Intel-FF.pdf> and IPDPS2013 talks)



■ Processor Research

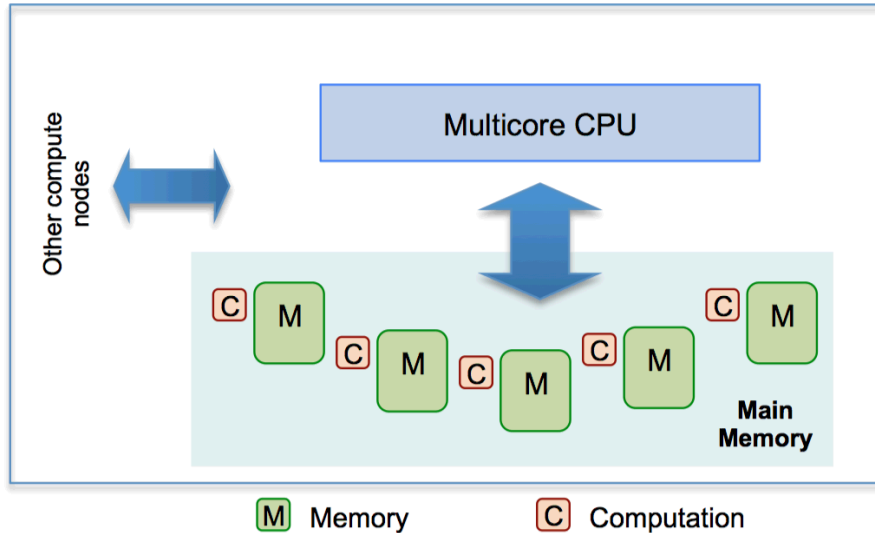
- Temporal SIMT and Scalarization
 - Reduce effect of wide vectors
- Coherency and consistency across system
- Hierarchical memory systems

■ Interconnect Research

- Open standards for the data center
- Support direct GPU messaging

■ Programmability

- Global address spaces (PGAS)
- Efficient cross machine collectives
- Fast synchronization
- Active messages
- Heterogeneous cores

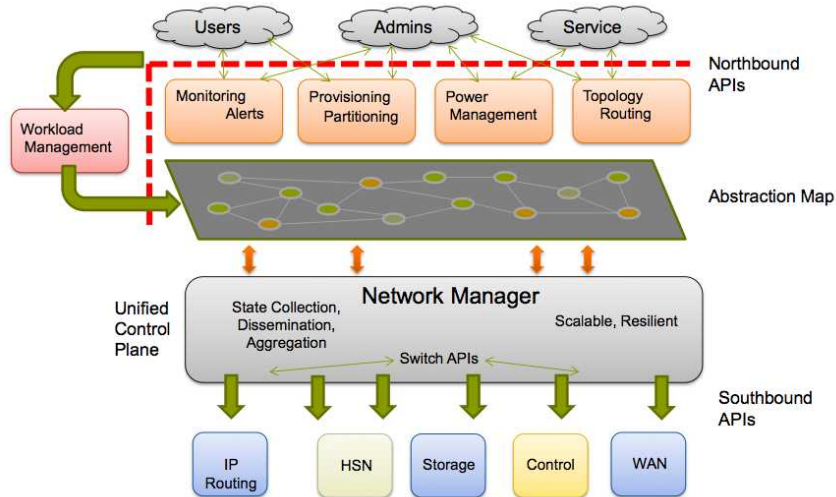


■ Memory Research

- Novel Computation near memory
- Reduction in data movement and associated overhead
- Advances in programming models, compiler and runtime environment
- Leverage of emerging memory technologies
- Advances in memory efficiency
- Advances in memory system integration, power and reliability management

■ Impact

- Large reduction of data movement
- Significant improvement at system level performance, power efficiency, and reliability
- Successful exploitation of novel architecture features while abstracting the hardware complexity, enables by evolutionary and revolutionary approaches



■ Network Management API

- What are the important management functions to provision?
- What structure of system management best serves those functions?
- Standardized APIs to allow management of a variety of high performance networks.

■ Network Communication API

- NIC functions to enable efficient execution of network API
- Structures required to achieve scalability of a diverse range of traffic patterns?
- Novel functions in future cores to facilitate efficient wakeup on the arrival of new data?

■ Network Protocol

- How can the NICs generate simple, small, HPC optimized packets at a sufficient rate?
- Interoperable protocols in support of heterogeneous, adaptive designs
- What flexibility is needed to allow vendor differentiation?

ASC and ASCR are partnering on Joint Advanced Technology System Procurements

- The APEX (LANL, LBNL, and SNL) collaboration is intended to result in the procurement of two platforms in ~2020
 - NERSC/ASCR procurement of NERSC-9
 - ACES/ASC procurement of ATS-3 (Advanced Technology System)
- Both platforms will focus on meeting both mission needs and pursuing Advanced Technology concepts
 - - We expect to use Non-Recurring Engineering investment to guide and improve system performance and productivity

High-level Design Philosophy for ATS3

- Delivered application performance is the primary driver in support of mission requirements
 - Peak FLOPS requirement will not appear in RFP
- 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)
- APEX are pre-exascale platforms
 - MUST support path to exascale programming models
 - While supporting existing mission needs
 - Support MPI + OpenMP (threads)
 - Matured on Trinity/Cori and CORAL platforms
 - Additional support for other, yet to be identified, MPI+X programming models

APEX 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 (ATS1) 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 Projects
 - Potential Path Forward project
 - NRE could take select technologies the final yards towards production

Fast Forward and Design Forward Impact

- APEX Team is performing Market Surveys
- Vendors visiting in phases starting in January 2015
 - **IBM, Intel, Cray, Nvidia, AMD**, SGI, HP, Micron, Broadcom, ARM, etc.
- Fast Forward and Design Forward Accomplishments and Progress have direct influence over the development of APEX technical requirements
- Developing NRE strategy
 - We started early to enable a richer range of NRE topics