CORAL COLLABORATION
OAK RIDGE • ARGONNE • LIVERMORE

Oak Ridge National Laboratory

Argonne National Laboratory

Lawrence Livermore National Laboratory

Acquisition of the next generation Leadership Systems at Oak Ridge, Argonne, and Lawrence Livermore National Laboratories

Presented to
Advanced Scientific Computing Advisory Committee
Washington DC
March 31, 2014

Susan Coghlan
Project Director, ALCF

Arthur Bland
Project Director, OLCF

DOE Leadership Computing Facility

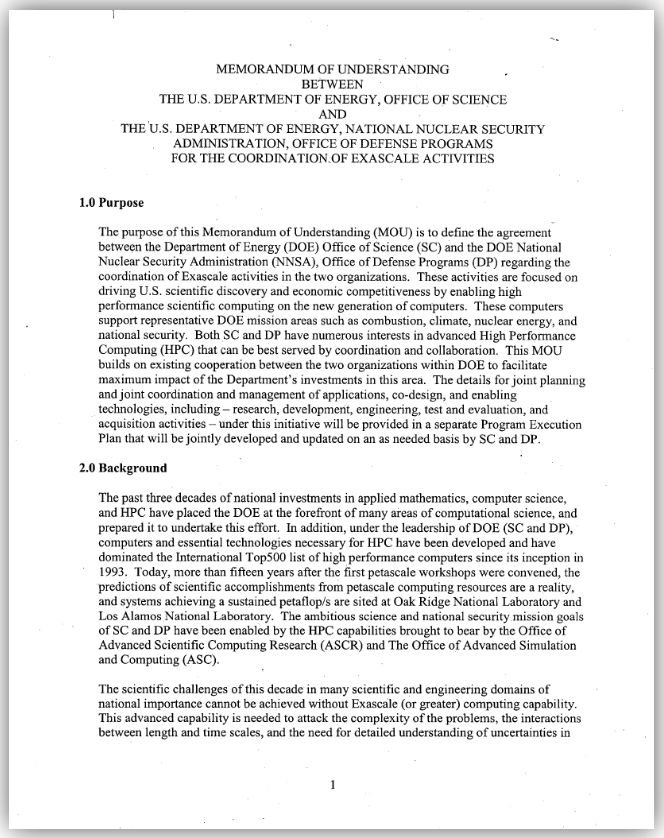
- Mission: Provide the computational and data science resources required to solve the most challenging scientific & engineering problems.
- Two centers with two architectures to address diverse and growing computational needs of the scientific community.
- Highly competitive user allocation programs (INCITE, ALCC). 3x oversubscribed.
- Projects receive computational resources typically 100x greater than generally available.
- LCF centers partner with users to enable science & engineering breakthroughs (Liaisons, Catalysts).

- Peak Performance: 10 Petaflops
- 49,152 Compute Nodes, 768 TB system memory
- File system: 35 PB, 240 GB/s bandwidth
- System accepted December 2012

- Peak Performance: 27 Petaflops
- 18,688 Compute Nodes, 710 TB system memory
- File System: 32 PB, 1 TB/s Bandwidth
- System Accepted May 2013

What is CORAL

- DOE's Office of Science (DOE/SC) and National Nuclear Security Administration (NNSA) signed an MOU agreeing to collaborate on exascale research and pre-exascale & exascale acquisitions
- **CORAL** is a **C**ollaboration of **O**ak **R**idge, **A**rgonne, and **L**awrence Livermore Labs **to acquire three systems for delivery in 2017/2018.**
- Collaboration grouping was done based on common acquisition timings and is a win-win.
  - Reduces the number of RFPs for vendors
  - Allows pooling of R&D funds
  - Supports sharing technical expertise among labs
  - Strengthens SC/NNSA alliance for exascale



MEMORANDUM OF UNDERSTANDING
BETWEEN
THE U.S. DEPARTMENT OF ENERGY, OFFICE OF SCIENCE
AND
THE U.S. DEPARTMENT OF ENERGY, NATIONAL NUCLEAR SECURITY ADMINISTRATION, OFFICE OF DEFENSE PROGRAMS
FOR THE COORDINATION OF EXASCALE ACTIVITIES

1.0 Purpose

The purpose of this Memorandum of Understanding (MOU) is to define the agreement between the Department of Energy (DOE) Office of Science (SC) and the DOE National Nuclear Security Administration (NNSA), Office of Defense Programs (DP) regarding the coordination of Exascale activities in the two organizations. These activities are focused on driving U.S. scientific discovery and economic competitiveness by enabling high performance scientific computing on the new generation of computers. These computers support representative DOE mission areas such as combustion, climate, nuclear energy, and national security. Both SC and DP have numerous interests in advanced High Performance Computing (HPC) that can be best served by coordination and collaboration. This MOU builds on existing cooperation between the two organizations within DOE to facilitate maximum impact of the Department's investments in this area. The details for joint planning and joint coordination and management of applications, co-design, and enabling technologies, including – research, development, engineering, test and evaluation, and acquisition activities – under this initiative will be provided in a separate Program Execution Plan that will be jointly developed and updated on an as needed basis by SC and DP.

2.0 Background

The past three decades of national investments in applied mathematics, computer science, and HPC have placed the DOE at the forefront of many areas of computational science, and prepared it to undertake this effort. In addition, under the leadership of DOE (SC and DP), computers and essential technologies necessary for HPC have been developed and have dominated the International Top500 list of high performance computers since its inception in 1993. Today, more than fifteen years after the first petascale workshops were convened, the predictions of scientific accomplishments from petascale computing resources are a reality, and systems achieving a sustained petaflop/s are sited at Oak Ridge National Laboratory and Los Alamos National Laboratory. The ambitious science and national security mission goals of SC and DP have been enabled by the HPC capabilities brought to bear by the Office of Advanced Scientific Computing Research (ASCR) and The Office of Advanced Simulation and Computing (ASC).

The scientific challenges of this decade in many scientific and engineering domains of national importance cannot be achieved without Exascale (or greater) computing capability. This advanced capability is needed to attack the complexity of the problems, the interactions between length and time scales, and the need for detailed understanding of uncertainties in

1

CORAL Joint NNSA & SC Leadership Computing Acquisition Project

Current DOE Leadership Computers

Titan (ORNL)
2012 - 2017

Sequoia (LLNL)
2012 - 2017

Mira (ANL)
2012 - 2017

Objective - Procure 3 leadership computers to be sited at ANL, ORNL, and LLNL in CY17-18

Leadership Computers run the most demanding DOE mission applications and advance HPC technologies to assure continued US/DOE leadership

Approach:

Competitive process - one RFP (issued by LLNL) leading to 2 Non-Recurring Engineering (NRE) contracts and 3 computer procurement contracts

For risk reduction and to meet a broad set of requirements,
2 architectural paths will be selected – one at each of the LCF centers.

Once selected, multi-year, lab-awardee relationship to deliver the best performance

Both NRE contracts jointly managed by the 3 Labs

Each lab manages and negotiates its own computer procurement contract, and may exercise options to meet their specific needs

Understanding that **long procurement lead time may impact architectural characteristics and designs** of procured computers

CORAL

Using a Combined-RFP benefits DOE and the labs

- **Leverages NRE investments across DOE and procurements**
  - Long-term, ongoing relationship with all three labs actively involved in managing and reviewing milestones of both NRE contracts
  - Jointly identify, assess, and frequently develop innovative solutions through co-design backed by NRE funding

- **Improves quality of technical requirements by having larger group of experts to provide input**
- **Facilitates responding to vendors with one voice**
- **Provides risk mitigation option of switching to other system architecture if something catastrophic happens to one**

CORAL

Advantages of Diversity

Having system diversity provides many advantages.

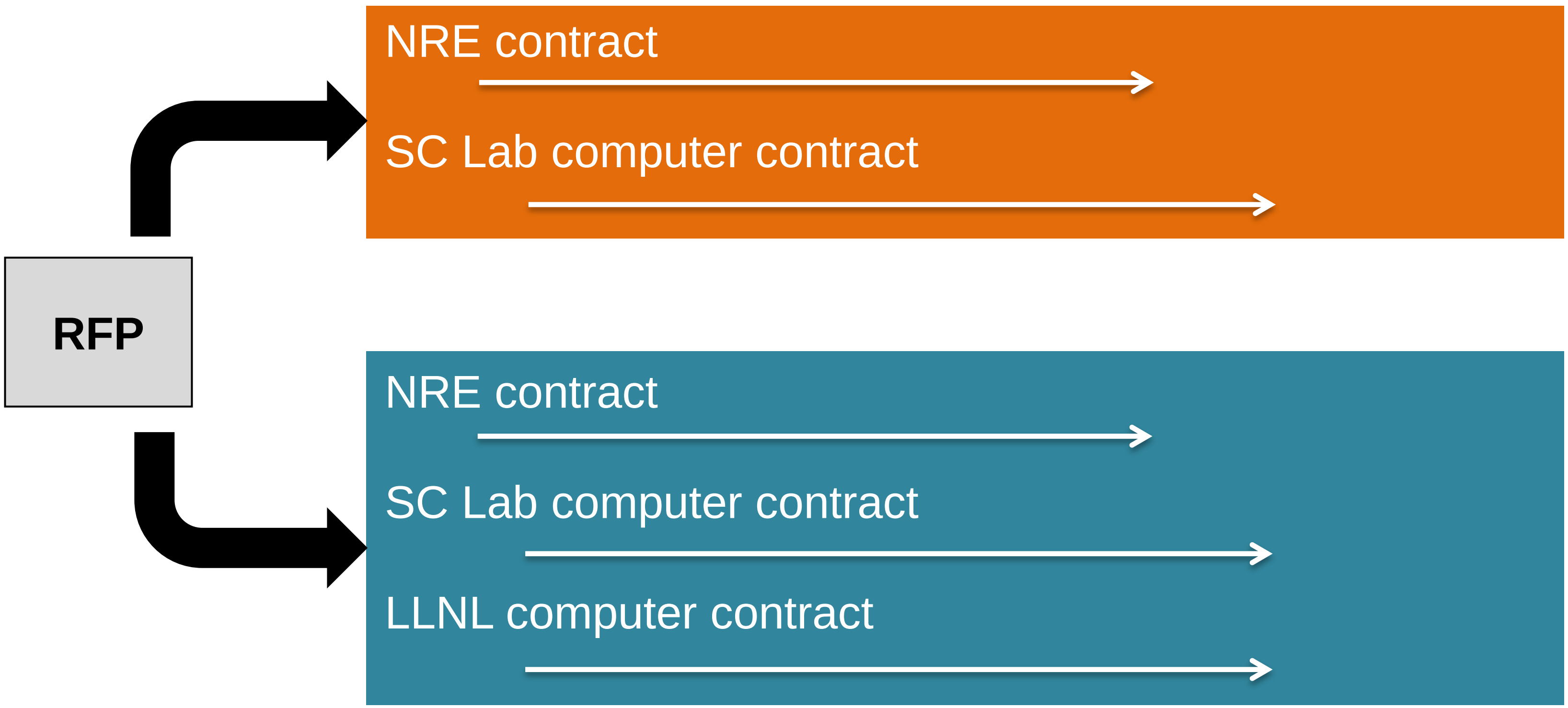
- It promotes price competition, which increases the value to DOE.
- It promotes a competition of ideas and technologies, which helps provide more capable systems for DOE's mission needs.
- It helps promote a rich and healthy high performance computing ecosystem, which is important for national competitiveness and DOE's strategic plan.
- Leading HPC is a priority goal in DOE's strategic plan.
- It reduces risk that may be caused by delays or failure of a particular technology or shifts in vendor business focus, staff or financial health.

The CORAL team worked with the HPC vendor community to ensure that the responses had sufficient diversity to provide the above advantages.

CORAL

The CORAL Procurement Model Supports Siting of Three Large Systems in 2017-2018

Two Diverse Architecture Paths



CORAL

CORAL RFP Terminology

- **RFP Documents**
  - Statement of Work: SOW
  - Proposal Evaluation & Proposal Preparation Instructions: PEPPI
- **SOW Requirement Categories**
  - Mandatory Requirements (MR) – essential and must be bid to be considered
  - Mandatory Options (MO) – Options that must be bid to be considered
  - Technical Options (TO-1) – Important, but not required to be responsive
  - Target Requirements (TR-1, TR-2, TR-3) – features, components, performance characteristics, or other properties that are important, but will not result in nonresponsive bid if omitted

High Level System Requirements

- Target speedup over current systems of 4x on Scalable benchmarks and 6x on Throughput benchmarks
- Peak Performance ≥ 100 PF
- Aggregate memory of 4 PB and ≥ 1 GB per MPI task (2 GB preferred)
- Maximum power consumption of system and peripherals ≤ 20MW
- Mean Time Between Application Failure that requires human intervention ≥ 6 days
- Data centric capabilities
- Delivery in 2017 with acceptance in 2018

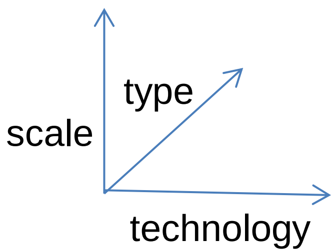
Technical requirements guiding principles

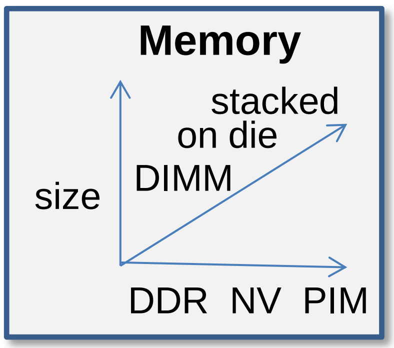
- **Minimize number of Mandatory Requirements (MRs) and allow consideration of widest range of architectural solutions.**
  - Word those requirements to allow architecturally diverse solutions

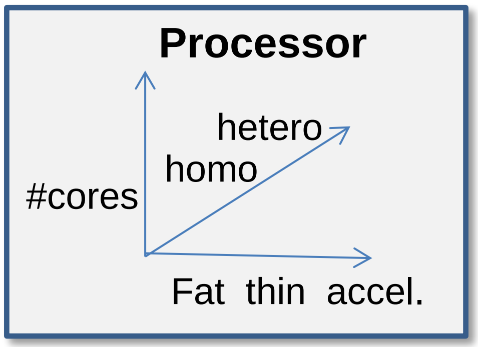
- **Focus on requiring science and throughput performance. Avoid overly prescriptive explicit speeds and feeds.**
- **Agree on common technical requirements across all three Labs – not three separate sets of requirements.**
  - Teams concur on requirements in each technical area
- **Require vendors to describe available options to adjust system size and configuration to meet individual site needs and/or budgetary constraints.**

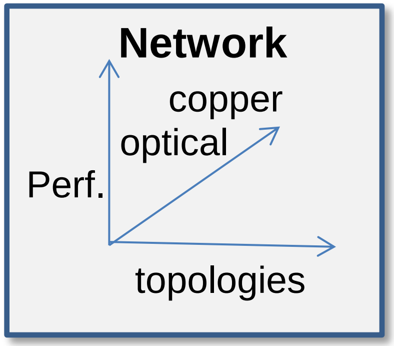
Many Types of System Diversity

- **Systems can vary from one another in many different dimensions**
  - System (architecture, interconnect, IO subsystem, density, resilience, etc.)
  - Node (heterogeneous, homogeneous, memory and processor architectures, etc.)
  - Software (HPC stack, OS, IO, file system, prog. environment, admin tools, etc.)
  - Hardware e.g.







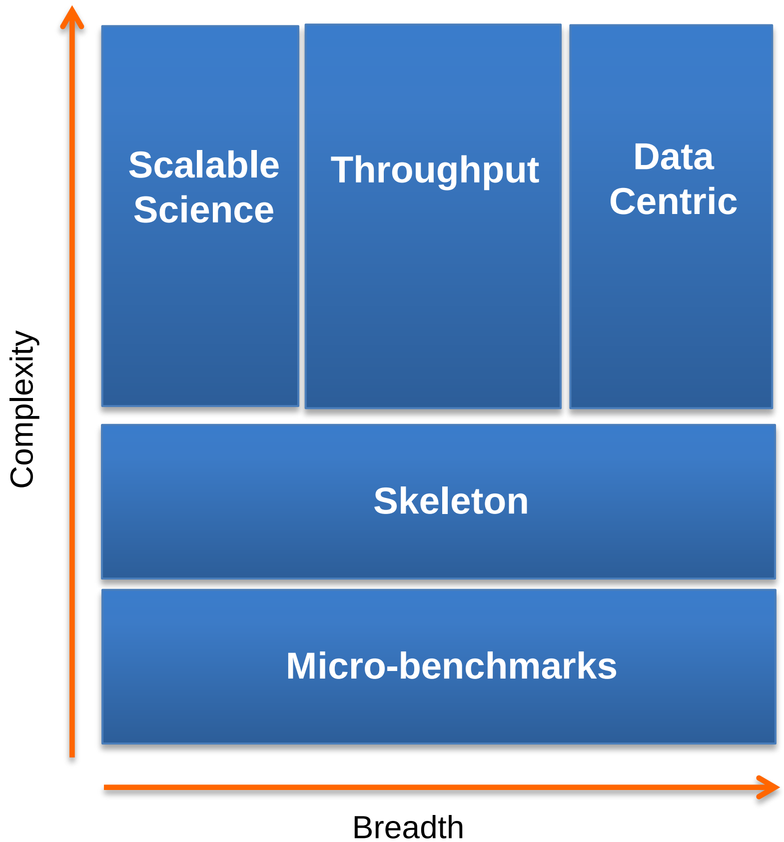


- **Ways Systems can be diverse**
  - Few big differences
  - Many little differences
  - Different technologies
  - Different ecosystems, i.e., vendors involved

CORAL benchmark categories represent DOE workloads and technical requirements

- **Scalable science benchmarks**
  - Expected to run at full scale of the CORAL systems
- **Throughput benchmarks**
  - Represent large ensemble runs; may be subsets of full applications
- **Data centric benchmarks**
  - Represent emerging data intensive workloads
  - Integer operations, instruction throughput, indirect addressing
- **Skeleton benchmarks**
  - Investigate various platform characteristics including network performance, threading overheads, I/O, memory, memory hierarchies, system software, and programming models
- **Micro benchmarks**
  - Small code fragments that represent expensive compute portions of some of the scalable science and throughput applications
  - Useful for testing programming methods and performance at the node level & for emulators and simulators

Benchmark Categories



CORAL

CORAL benchmarking suite uses mini-apps and a few larger applications

Categories	Scalable Science	Throughput	Data Centric	Skeleton
Marquee (TR-1)	LSMS QBOX NEKbone HACC	CAM-SE UMT2013 AMG2013 MCB	Graph500 Int sort Hashing	CLOMP IOR CORAL MPI Memory CORAL loops
Elective (TR-2)		QMCPACK NAMD LULESH SNAP miniFE	SPECint_ peak2006	Pynamic HACC I/O FTQ XSBench miniMADNESS
Elective Micro-Benchmarks (TR-3)	NEKbonemk HACCmk	UMTmk AMDmk MILCmk GFMCmk		

CORAL

Application performance requirements are the highest priority to CORAL

The goal is an average performance improvement over today's systems of:

- 4-8x for scalable science apps
- 6-12x for throughput apps

Two Step Evaluation Process
Step 1: Technical Evaluation

We created **eight teams of technical experts** with three people from each of the CORAL labs plus one from either LANL or Sandia to represent their users of the Livermore system.

1. **Project Management**
2. **System Hardware**
3. **System Software**
4. **System Performance**
5. **Programming Environment**
6. **File System**
7. **Facilities and Operations**
8. **NRE**

These 8 technical teams reviewed the proposals for three weeks ahead of a two day face-to-face meeting at ORNL to assess each of the proposals against the DRAFT SOW and PEPPI criteria

CORAL

Two Step Evaluation Process
Step 2: Buying Team

A **Buying team** consisting of the management, technical, and procurement leadership of the three CORAL labs met to select the set of two proposals that provided the best value to the government

Evaluation Criteria:

- DOE mission requirements - the best overall combination of solutions
- Technical proposal excellence; projected performance on the applications is the single most important criterion
- Feasibility of schedule and performance
- Diversity
- Overall Price
- Supplier attributes
- Risk evaluation

CORAL

Procurement approach designed specifically for acquiring leadership computers

- **Long-term contractual partnership with vendors**
- **CORAL is a partnership for the long-term. All three labs are collaborating on both NRE contracts and Build contracts.**
- **Few mandatory requirements; many targets**
  - Targets are performance levels that both parties reasonably believe can be achieved depending on the NRE; converted to traditional mandatory requirements after GO/NO-GO
  - Evaluate the level of achievement of the system as a whole rather than an assessment of performance of each target individually
- **NRE contracts coupled to build contracts**
  - NRE results are basis for GO/NO-GO decisions in build contracts
  - Modeled on the successful Sequoia/Mira partnership
- **Announcement of Evaluation Results cannot be made until contracts have been negotiated (expected late FY14)**

CORAL

Questions?

CORAL
COLLABORATION
OAK RIDGE • ARGONNE • LIVERMORE

CORAL RFP

- https://asc.llnl.gov/CORAL
- **Provides all documents the bidders need**
  - PEPPI – Proposal Evaluation & Proposal Preparation Instructions
  - DRAFT CORAL Build Statement of Work
  - Specific requirements for each laboratory
  - Worksheets for bidders to fill in to make it easier to compare information in a standard way
  - Bidder Questions and our Answers

- https://asc.llnl.gov/CORAL-benchmarks/
- **Provides all information on the benchmark codes**

CORAL