

# ECP Hardware Technology (HT)

## PathForward Information Meeting

April 6, 2016

James A. Ang (Sandia) HT Director

John Shalf (LBNL) HT Deputy Director

### Overview of ECP Hardware Technology Portfolio

- *Goals and Objectives for PathForward*
- *Overview of Draft PathForward Technical Requirements*



EXASCALE COMPUTING PROJECT

# Hardware Technology Focus Area Objectives

- Leverage our window of time to support advances in both system and node architectures
- Close gaps in vendor's technology roadmaps or accelerate time to market to address ECP performance targets *while* affecting and intercepting the 2019 Exascale System RFP
- Provide an opportunity for ECP Application Development and Software Technology efforts to *influence* the design of future node and system architecture designs

# Hardware Technology Overview

**Objective:** Fund R&D to design hardware that meets ECP Targets for application performance, power efficiency, and resilience

Issue competitively awarded PathForward Hardware Architecture R&D contracts that deliver:

- Conceptual system designs, and associated node designs
- Analysis of performance improvement relative to level of effort required to migrate software to the conceptual system design
- Simulation, emulation or test hardware, technology demonstrations to quantify performance gains over existing roadmaps
- Support for active engagement in ECP holistic co-design efforts
- Credible productization plan for proposed hardware designs

DOE labs engage to:

- Evaluate PathForward RFP responses
- Participate in reviews and evaluation of PathForward deliverables
- Develop Architectural Analysis, Abstract Machine Models, Proxy Architectures of PathForward designs to support ECP Co-Design
- Organize Co-design and Integration Workshops

# PathForward RFI/SOW Team

- Technical Requirements Team:
  - Ray Bair, ANL
  - Jeff Kuehn, LANL
  - Doug Doerfler, LBNL
  - Bronis de Supinski, LLNL
  - Al Geist, ORNL
  - Si Hammond, SNL
- Procurement Team:
  - Brandt Esser, LLNL
  - Willy Besancenez, ORNL

# Overarching Goals for PathForward

- High Level Approach

- Close gaps in vendor's technology roadmaps or accelerate time to market to address ECP performance targets
- Provide an opportunity for Application Development and Software Technology to *influence* the design of future node and system architecture designs
- Deliver hardware technology analysis and (where appropriate), demonstrations to increase confidence in node and system design performance benefit, programmability and ability to affect a 2019 Exascale System RFP

- Goals

- Ensure that laboratory platform acquisition teams have quantitative information to identify the most promising technology options to include in the 2019 Exascale System RFP
- Improve vendor's confidence in the value and feasibility of aggressive advanced technology options that they may propose for 2019 Exascale System RFP

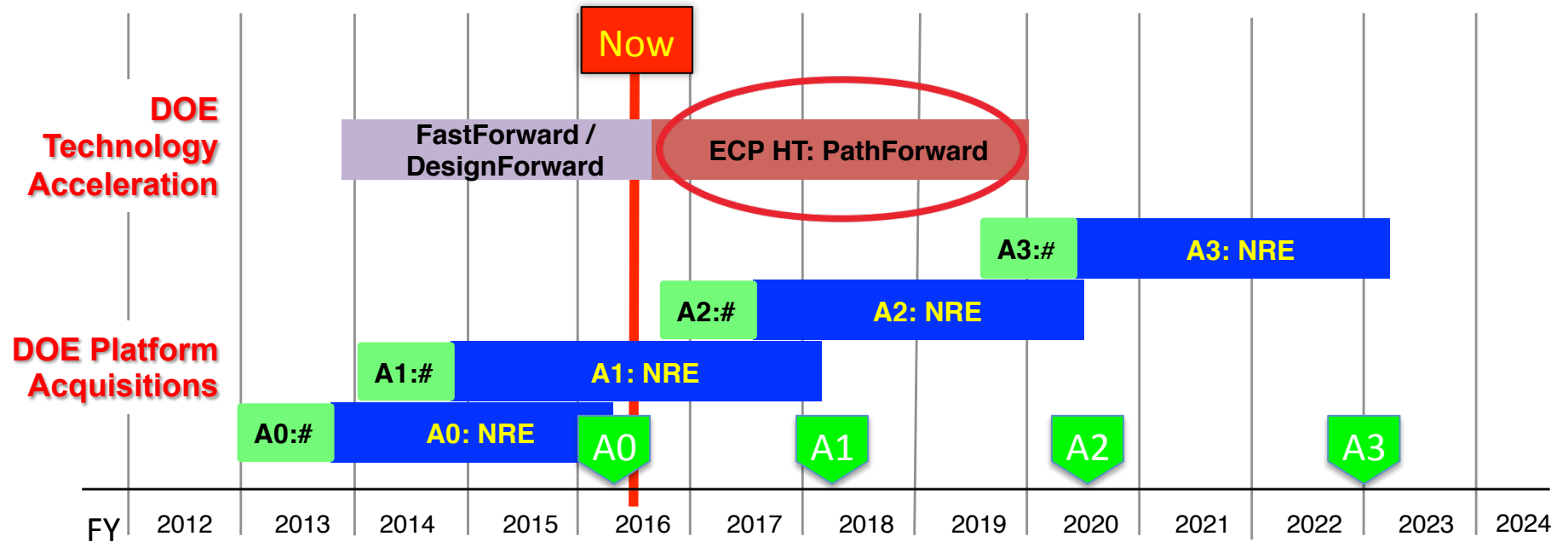
# PathForward: Vendor Node and System Design

End State	Approximately five vendors have viable candidates for exascale system designs with associated node and computing component designs that meet or exceed the ECP target requirements
Scope & Requirements	<ul style="list-style-type: none"><li>• Pursue Hardware Technology investments in innovative node and system architecture concepts that enhance DOE application performance to meet mission needs, and improve Power and Resilience performance</li><li>• Application Development focus area's challenge problems provide application drivers and inputs to the ECP co-design efforts</li><li>• Co-design opportunity for Software Technology focus area requirement to support ECP applications on diverse system architectures</li><li>• PathForward projects are organized with independently scoped (and priced) work-packages</li></ul>

# PathForward: Vendor Node and System Design

Assumptions	<ul style="list-style-type: none"><li>• PathForward will use ECP goals to define new designs for commercial-off-the-shelf (COTS) computing components for integration into future node and system architectures</li><li>• This new generation of COTS components changes the computing eco-system and helps ensure PathForward hardware designs are affordable</li></ul>
Objectives	<ul style="list-style-type: none"><li>• Influence hardware architectures that are designed to meet ECP performance objectives</li><li>• This supports previously off-roadmap architecture innovations that could not or would not otherwise be investigated</li></ul>

# Platform Acquisition and \*Forward Timelines



# - Procurement Phase

A0: Trinity/Cori

A1: CORAL

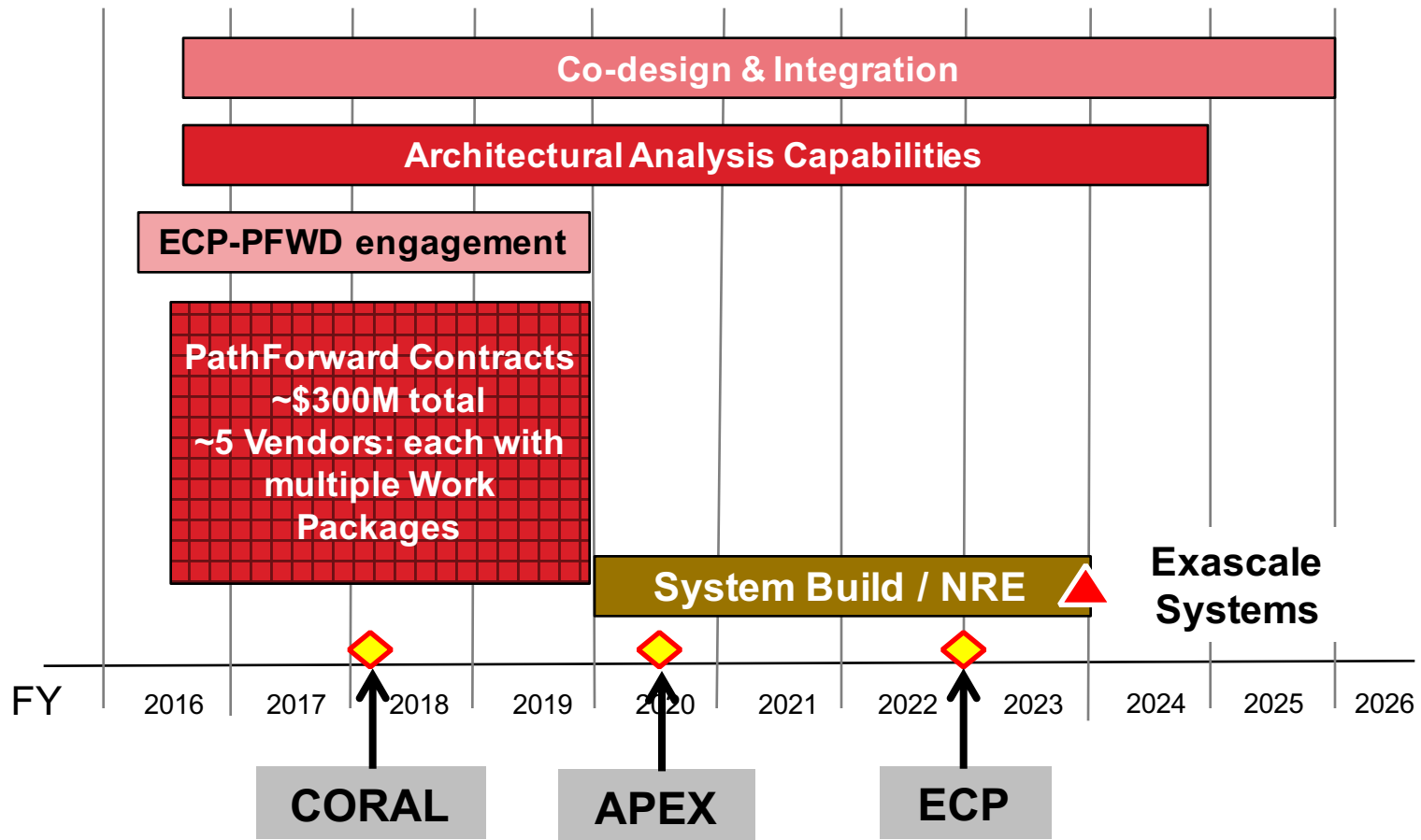
A2: APEX

A3: Exascale

*All Dates are Approximate*



# Hardware Technology & Exascale System Timeline

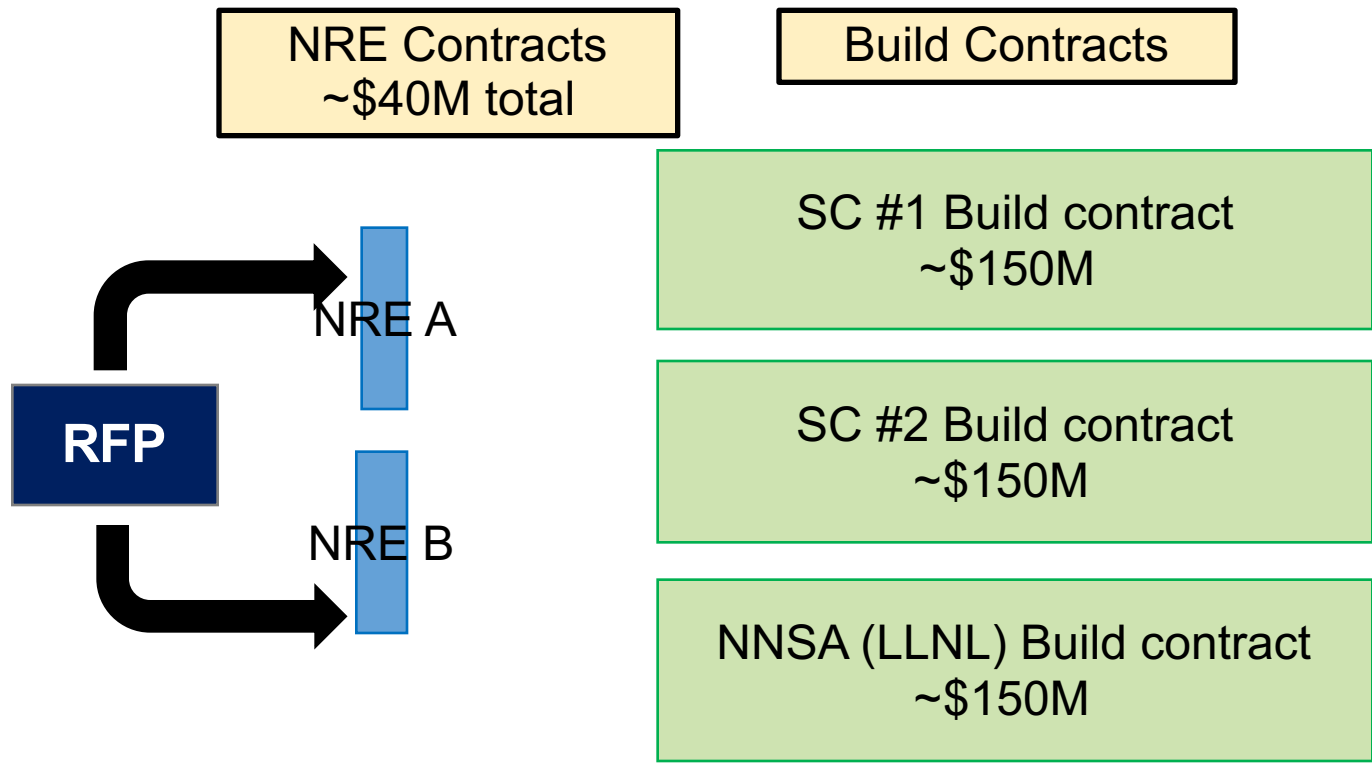


*All Dates are Approximate*

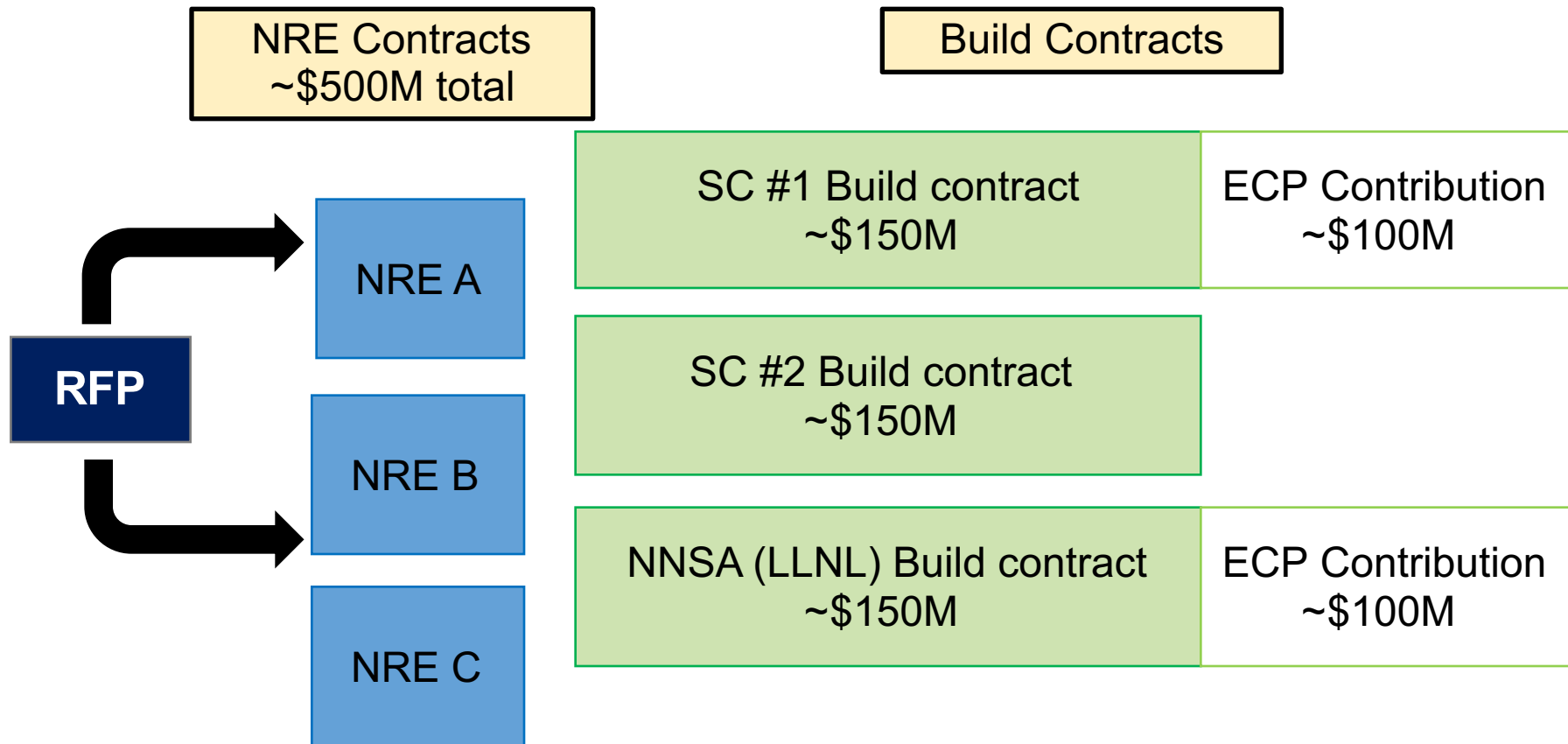
*◇ Approximate System Dates*

# Baseline 2022-2023 DOE Office of Science and NNSA system acquisition plan:

At least 2 diverse architecture paths for SC, 2 NRE, and 3 system contracts



# ECP additions to the 2022-2023 DOE Office of Science and NNSA system acquisition plan: A third NRE contract and at least two systems will be capable exascale



# For PathForward, A *Capable* Exascale System meets the following **ECP Performance Targets\***

- **Application Performance Improvement vs. Sequoia and Titan baseline**
  - 50x improvement in application Figure of Merit (FOM)
  - FOM is defined by each ECP Application Development team to increase with problem size/complexity as well as computation rate
  - $FOM = [\text{Problem Size or Complexity Increase}] \times [\text{Speedup}]$
- **Total System Power**
  - 20-30 MW peak power
- **System Resilience**
  - 1 week on average between *human intervention* due to hardware or system faults

\* Note, the Performance Targets that appear in the 2019 Exascale System RFP will be updated based on Application Development, Software Technology and PathForward accomplishments in the first 3-4 years of ECP

# Current PathForward RFP Plan

- Will have Firm Fixed Price contracts with milestone deliverables and payments
- Will seek DOE Advance IP Waivers for vendors that provide 40% cost share
- Number and scale of contracts is determined after proposal reviews and source selection
- Project Duration – 3 years
  - Interim deliverable: after 12-18 months
  - Interested in work packages that support opportunities for ECP application development or software technology needs to *influence* hardware designs

# ECP Hardware Technology (HT)

## PathForward Information Meeting

April 6, 2016

James A. Ang (Sandia) HT Director

John Shalf (LBNL) HT Deputy Director

### Overview of ECP Hardware Technology Portfolio

- *Goals and Objectives for PathForward*
- *Overview of Draft PathForward Technical Requirements*



EXASCALE COMPUTING PROJECT

# Overarching Goals for Path Forward RFP

- Improve the quality of offeror responses to the 2019 Exascale Systems RFP
- Improve the offeror's confidence in the value and feasibility of aggressive advanced technology options that would be bid in response to the 2019 Exascale Systems RFP
- Improve DOE confidence confidence in technology performance benefit, programmability and ability to integrate into a credible system platform acquisition

# PathForward Proposals

## Submit separate technical and cost proposals

- **Technical Proposal – SOW**

- Proposals that are focused on system architectures must describe a conceptual exascale system design including the node design
- One technical proposal per conceptual system design – can have multiple technical proposals from one offeror
- Proposals that are focused on node or component level architectures must still provide a conceptual system design
- Detailed review will be performed by the Technical Evaluation Team

- **Cost Proposal is a separate document**

- PathForward Buying Team will evaluate this cost information and inputs from the Technical Evaluation Team
- Cost proposal shall have for each Work Package: schedule/quarterly milestones, costs for each work package deliverable, dependencies on other work packages or technologies, and exit conditions



# Overview of Draft PathForward Technical Requirements

- **Mandatory Requirements (MRs)** – Essential Requirements. The Offeror must address ALL MRs to have its proposal considered responsive
- **Target Requirements (TRs)** – Important Requirements, but will not result in a non-responsive determination if omitted from a proposal

# Mandatory Requirements

- **Exascale System Description**
  - Describe an exascale system design including the conceptual node design
  - Describe the innovative nature of the proposed R&D and describe where it differs from your company roadmap
- **Prioritized list of work-packages with a high-level synopsis of the following factors for each work package**
  - **Technology Description:** Aspect of the system that this technology targets (e.g., parallelism, resilience, energy consumption, memory and storage)
  - **Area of innovation:** Short (a few words) description of the technology innovation that is proposed in this work package
  - **Impact:** Degree to which the exascale system design and associated COTS computing components will be impacted by successful completion of the work package
  - **Risk:** The likelihood of success/failure of this technology option
  - **Cost** of each work package in the *separate cost proposal*

# Prioritize List of Work Packages across Proposals (TR1)\*

- If the Offeror has more than one Exascale System Design and submits more than one technical proposal
- A separate document shall prioritize all work-packages across the proposals
- This separate document does not count against the page limit

\* Only needed if Offeror is submitting more than one technical proposal

# PathForward Work Packages (TR1)

## (For Each Work Package)

- **Technical challenge:** Describe the performance opportunity or risk in conceptual design that requires DOE investment
- **Proposed Remedy:** Provide a description of the specific technological remedy to the identified challenge
- **Value proposition:**
  - In recognition of the diversity of potential R&D options, the Offeror shall define a Figure of Merit (FOM) and describe how much the FOM can be improved through PathForward investments
  - The FOM must be **specific, measurable, and attainable** with quantitative goals
- **Work Plan:** Describe concrete steps that will be required in development of the technological solution that meets the FOM and describe any co-design requirements

# Impact on Software Environment (TR2)

- Describe impact of the conceptual system or node design on the programming environment and/or describe the programming interface to access the new hardware capability
- Describe for any hardware technology that requires software changes, the level of effort required to:
  - Move existing applications and components of the software stack to the new system design
  - Bridge to evolutionary applications development with existing programming models, while meeting ECP performance targets
  - Meet or exceed ECP performance targets if more revolutionary applications development is required with new programming models

# Productization Strategy (TR2)

- Describe how and when the proposed conceptual system design will be commercialized, productized, or otherwise made available to customers
- Identify target customer base/market(s) for hardware technology designs
- Describe the impact of system and node designs on the HPC market, as well as the potential for broad adoption
- Indicate projected timeline for productization

# Technology Demonstrations (TR2)

## Simulators, Emulators, Test Hardware

- **Provide confidence in the impact of the proposed advanced architecture concept**
  - Appropriate level of tech demonstration to make the case that this is viable
  - Must be a specific and independently evaluable artifact
  - Ideally provide ability to run application kernels and software technology and collect performance data (stretch goal)
- **Also provide baseline application performance data for Status Quo Hardware Technology**
- **ECP application development will be focused on exascale challenge problem needs, hence must balance software costs against benefits**
  - Some application changes and additions will necessarily be more intrusive
  - R&D grade system software must bridge to DOE applications
  - CORAL and APEX RFP benchmarks can serve as useful proxies for actual exascale applications
  - The suite of ECP applications will be determined by Q4 2016

# Overall Outline of RFP Responses

**MR** **TR1** **TR2**

- **Solution Description: Conceptual design of Exascale node and system (baseline)**
- **R&D Plan: High level outline of prioritized workpackages**
  - Outline independent work-packages that can directly address challenges identified
  - Itemize costs for each work package *in the cost proposal*
- **Describe details of each work package (one vignette per workpackage)**
  - Technology Description: Summary
  - Area of innovation: Summary
  - Impact: Summary
  - Risk Assessment:
  - Technical challenge: Detail
  - Proposed Remedy: Detail
  - Value proposition (FOM): Detail
  - Work Plan:
  - Productization Strategy
  - Impacts on Software Environment
  - Technology Demonstration



# Differences between PathForward and FastForward / DesignForward:

- **Successful PathForward Projects:** System and Node Designs that will be incorporated in Exascale System bids to benefit from further development with NRE support
- **Must have impact on fixed schedule**
  - Must improve quality of a 2019 response for Exascale RFP
- **Funding:** Higher than FF/DF
- **Project Duration:** Longer than FF/DF – 3 years
  - Interim deliverable: at 12-18 month point
  - Work Package option(s) for hardware demonstrations that integrate PathForward conceptual designs
  - Need to intersect the 2019 Exascale System RFP

# Timeline

- **Now:** *Draft RFP posted on March 13*
  - *Provide as much feedback as you can now!*
- **June:** Official RFP Release
- **July:** Responses Due (30 days after RFP Release)
- **August:** Selection
- **End of September:** First Round awards
  - Remaining awards depend on SOW negotiations and funding availability



# Extreme-Scale Technology Challenges

- **Parallelism:** System design must enable applications to exploit the extreme levels of parallelism effectively that will be necessary at exascale.
- **Resilience:** System-level resilience to both permanent and transient faults and failures must enable applications software to “work through” these problems to achieve successful, accurate, reliable execution and completion.
- **Energy Consumption:** Energy efficiencies must enable the entire system to operate within affordable power budgets when run at the targeted computational rates.
- **Memory and Storage Challenge:** Memory and storage architectures must enable applications to access/store information at high capacities and with low latencies to support anticipated computational rates.

# Target Requirements

- **Work Packages**
- **Prioritized list of work-packages across proposals**
- **Impact and Risks**
- **Productization Strategy**