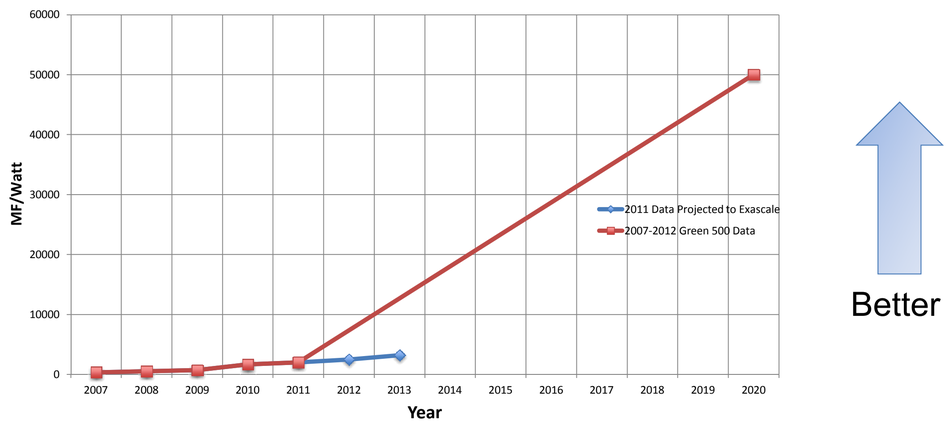
SAND2014-3241P

Addressing Power/Energy Challenges for Extreme Scale HPC

Sandia National Laboratories
James H. Laros III, Suzanne M. Kelly, Kevin T. Pedretti, Ryan E. Grant, Michael Levenhagen, David DeBonis, Stephen Olivier
Sandia National Laboratories, New Mexico 87123

Need Better Power Efficiency

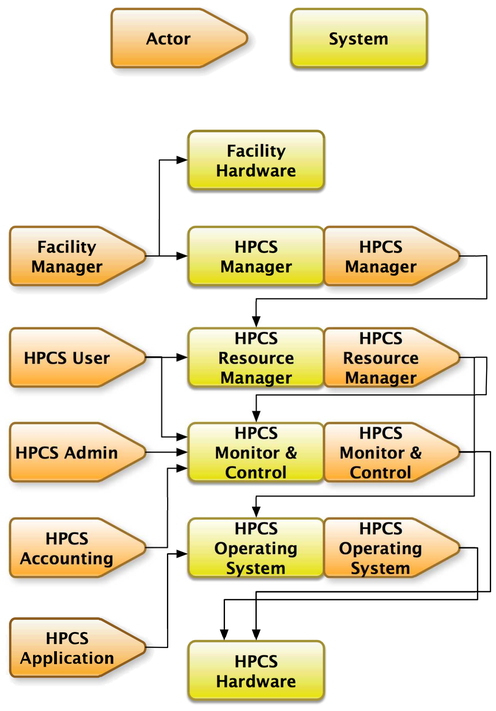
- Extreme scale HPC consumes *considerable* power
- Line in the sand: Exascale ≤ 20 MW
- Price of Power based on many factors (time of day etc.)
- Significant hardware advances needed
- Software controls and proactive management also required

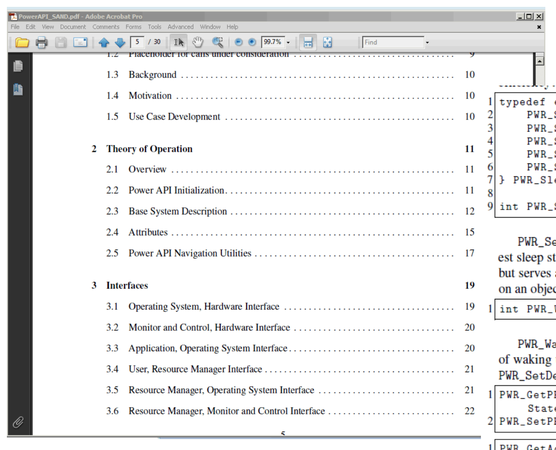


Current trajectory (Green 500) falls short of Exascale goal

Approach

- Continued research progress with an eye towards a Power API Specification
- A Use Case Approach to specifying Power API requirements – *Completed*
  - Very wide scope, from facility to hardware component, defined.
- Formal Power API Specification – *In Progress*
  - Prototyping targeted portions of the Spec.




PowerAPI_SAND.pdf - Adobe Acrobat Pro

1.3 Background 10
1.4 Motivation 10
1.5 Use Case Development 10

2 Theory of Operation **11**
2.1 Overview 11
2.2 Power API Initialization 11
2.3 Base System Description 12
2.4 Attributes 15
2.5 Power API Navigation Utilities 17

3 Interfaces **19**
3.1 Operating System, Hardware Interface 19
3.2 Monitor and Control, Hardware Interface 20
3.3 Application, Operating System Interface 20
3.4 User, Resource Manager Interface 21
3.5 Resource Manager, Operating System Interface 21
3.6 Resource Manager, Monitor and Control Interface 22

```
typedef enum {
    PWR_SLEEP_NO
    PWR_SLEEP_SHALLOW
    PWR_SLEEP_MEDIUM
    PWR_SLEEP_DEEP
    PWR_SLEEP_DEEPEST
} PWR_SleepState

int PWR_SetDeepestSleepState(PWR_Obj, PWR_SleepState)
```

PWR_SetDeepestSleepState allows the application to request that the OS restricts the deepest sleep state that the hardware can enter. This is not required to be honoured by the OS or HW, but serves as a hint to the OS that low latency sleep to active transitions are desired. Calling this on an object that does not provide sleep states will return an error PWR_ERR_NOT_PROVIDED.

```
int PWR_WakeUpLatency(PWR_Obj, PWR_SleepState)
```

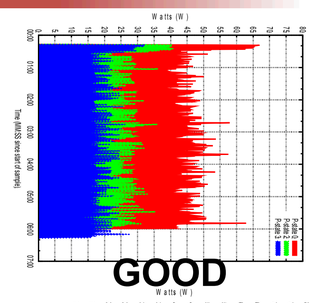
PWR_WakeUpLatency provides the application with the opportunity to determine the latency of waking up from a given sleep state. This information can be used to provide a better value to PWR_SetDeepestSleepState based on the application's sensitivity to latency.

```
PWR_GetPEState(); get and set call for some abstracted Power and/or Energy
    State, needs to be fleshed out
PWR_SetPEState()
```

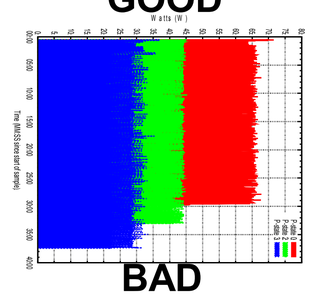
```
PWR_GetActiveCoreCount();
/* get number of cores active (not powered down and available for use by
   the application) */
PWR_GetCoreState();
/* get state of a core (or group of cores?) */
```

Research and Implementation

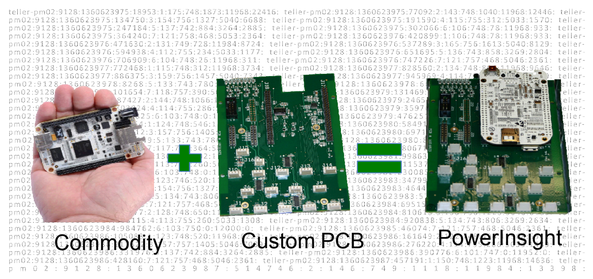
- Early research revealed power and energy saving opportunities are application specific
- CPU Frequency scaling can save energy with little impact to performance in some cases
- Network Bandwidth scaling shows similar application specific results



GOOD



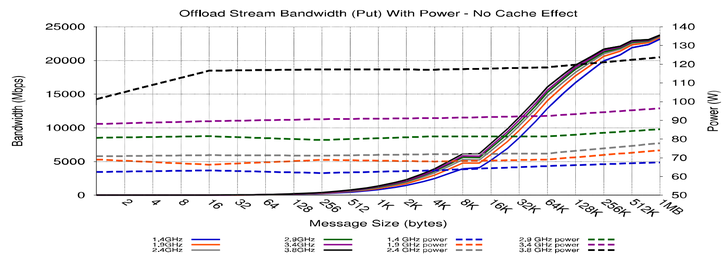
BAD



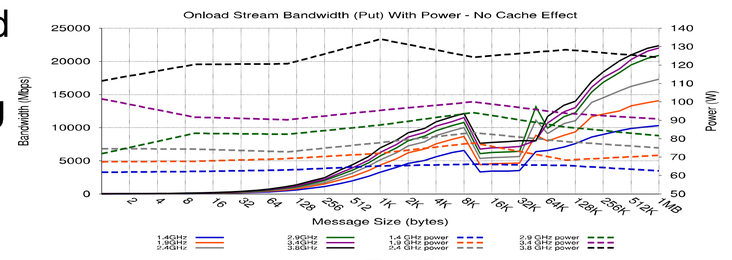
- PowerInsight
  - co-designed with Penguin Computing
- Commodity based component level high-frequency power measurement

**** PowerInsight enables ongoing Research and prototyping of evolving Power API Specification***

- CPU Frequency scaling benefit depends on NIC capabilities
- Onload vs Offload can be assessed for expected CPU frequencies



GOOD



BAD

Goals

- FY14 Completion of Power API Specification
- Refined commodity Power Measurement capabilities
  - PowerInsight Version 2
- Early Prototypes of Power API Specification
- Impact Trinity NRE – Power Measurement and Control
- Enable continued Research and Development

National Nuclear Security Administration
U.S. DEPARTMENT OF ENERGY

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.

SAND 2010-xxxP

Sandia National Laboratories