SAND2014-4196C

Exceptional service in the national interest

Sandia National Laboratories

Metrics for Evaluating Energy Saving Techniques for Resilient HPC Systems

Ryan E. Grant, Stephen L. Olivier, James Laros III,

Ron Brightwell and Allan K. Porterfield

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Images courtesy of: Oak Ridge National Laboratory; Argonne National Laboratory and Cray Inc.

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 NO. 2014-xxxx

Outline

Sandia National Laboratories

- Introduction to Extreme scale power/reliability concerns
- Interplay between power/energy and resilience
- Why you should care about resilience when analyzing runtime energy saving methods and how to consider it
- Impact of reliability on existing energy saving methods
- Case study
- Conclusions

Extreme Scale Power/Energy

Sandia National Laboratories

- Exascale computing is hitting a power/reliability wall
- We hit a power/reliability wall in the 1950s with vacuum tubes
  - Many thought ENIAC would never work due to the unreliability of vacuum tubes
- Solved the power-reliability problem with the transistor
- Assume we don't have revolutionary new technology coming to save us

Extreme Scale Power/Energy

Sandia National Laboratories

- We face a Power/Reliability wall again
  - Exotic technologies are not coming soon enough
  - Need to work on solutions for existing silicon technologies

- Known Issues with Extreme Scale Systems
  - Power caps
    - Practical power delivery issues
    - Energy Operating Expenses
  - Reliability
    - Systems need to be operational for > 1 day time frames
    - Multiple redundancy is expensive/impractical for HPC
      - Works for industry but not for scientific computing

Extreme Scale Resilience

Sandia National Laboratories

- Resilience methods are being researched for future extreme scale systems
- Fast burst buffer based traditional checkpoint/restore
  - Use a traditional checkpoint/restore method
  - Move fast storage to the node
- Uncoordinated checkpointing
  - Avoids issues with large synchronous network traffic
  - Difficult to implement and deploy correctly
- Replication
  - Have backup compute nodes that replicate the work of others
  - No stopping on failure
    - Unless all of the replicas of a process die

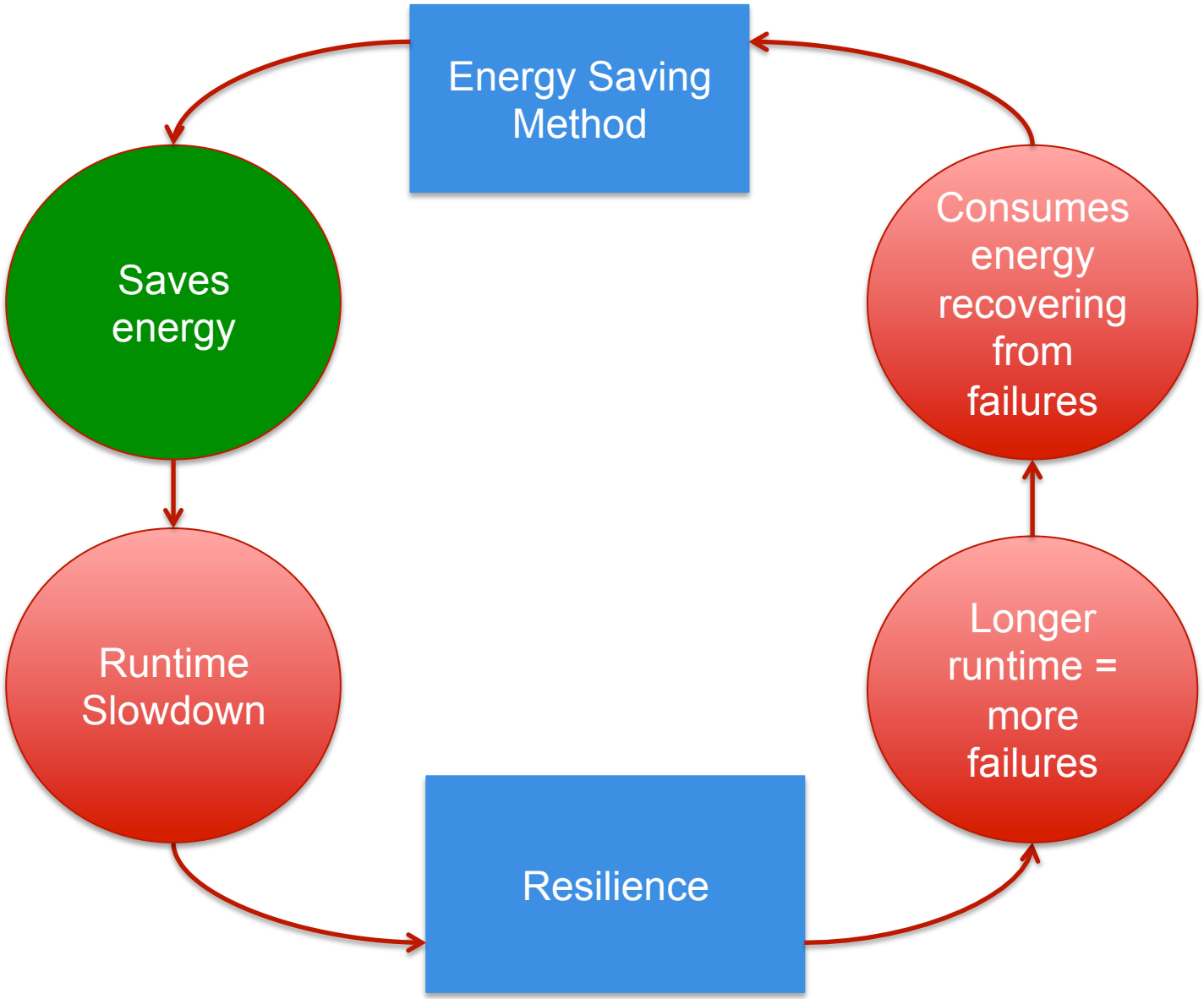
Extreme Scale Energy/Power

Sandia National Laboratories

- Proposed Methods to Conserve Energy
  - Online adjustment of CPU frequency/voltage
  - Application phase approaches
  - Scavenging energy during communications
- Resiliency
  - Burst buffers
  - New uncoordinated checkpoint methods
  - Proposed resilience aware middleware (MPI)

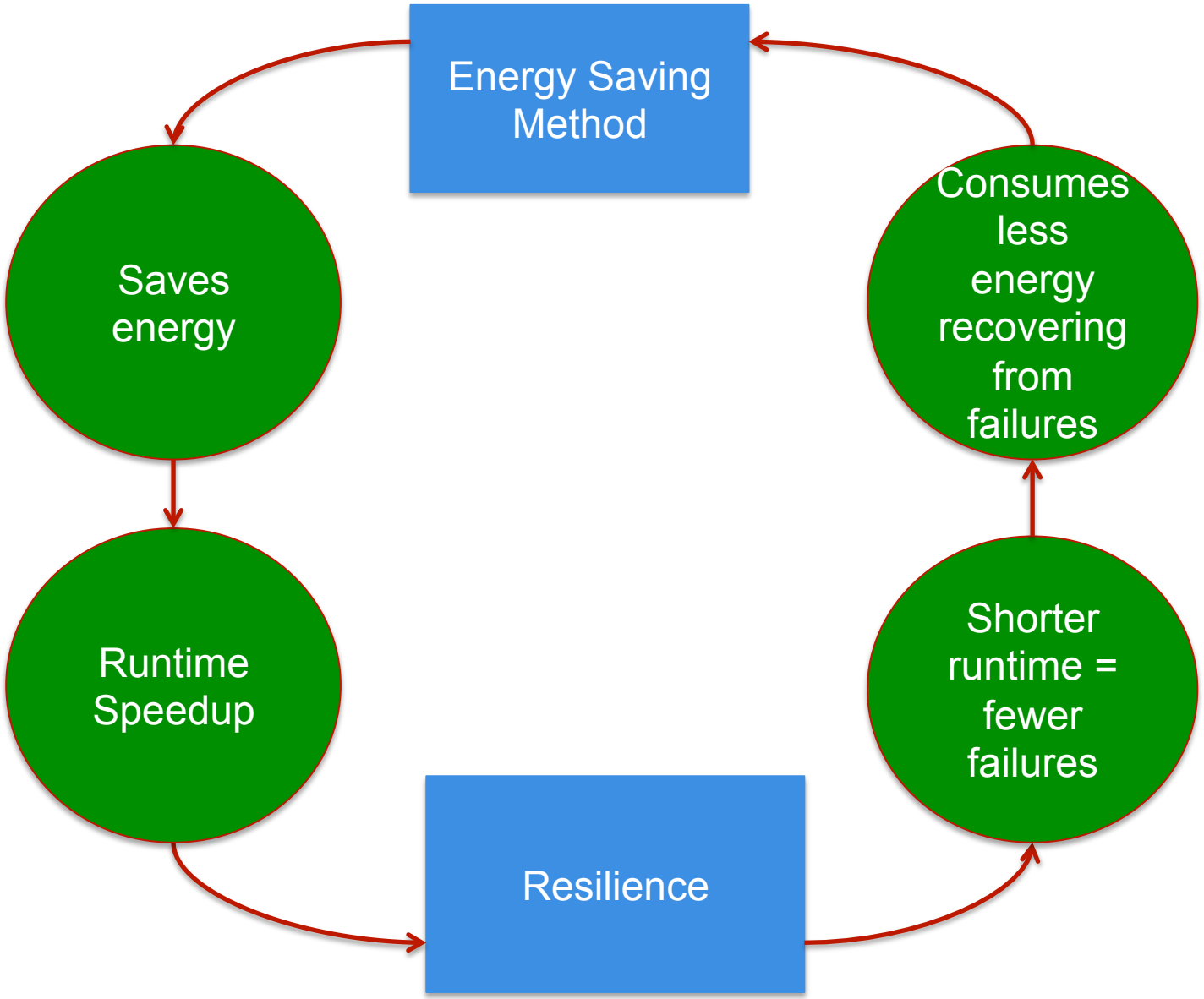
Interplay of Energy and Resilience

Sandia National Laboratories



Interplay of Energy and Resilience

Sandia National Laboratories



Impact of Reliability on Energy

Sandia National Laboratories

- When runtimes are lengthened the probability of a failure happening increases in relation to the additional runtime period

$$p(fail) = p(fail_{orig\_rt}) + p(fail_{add\_rt})$$

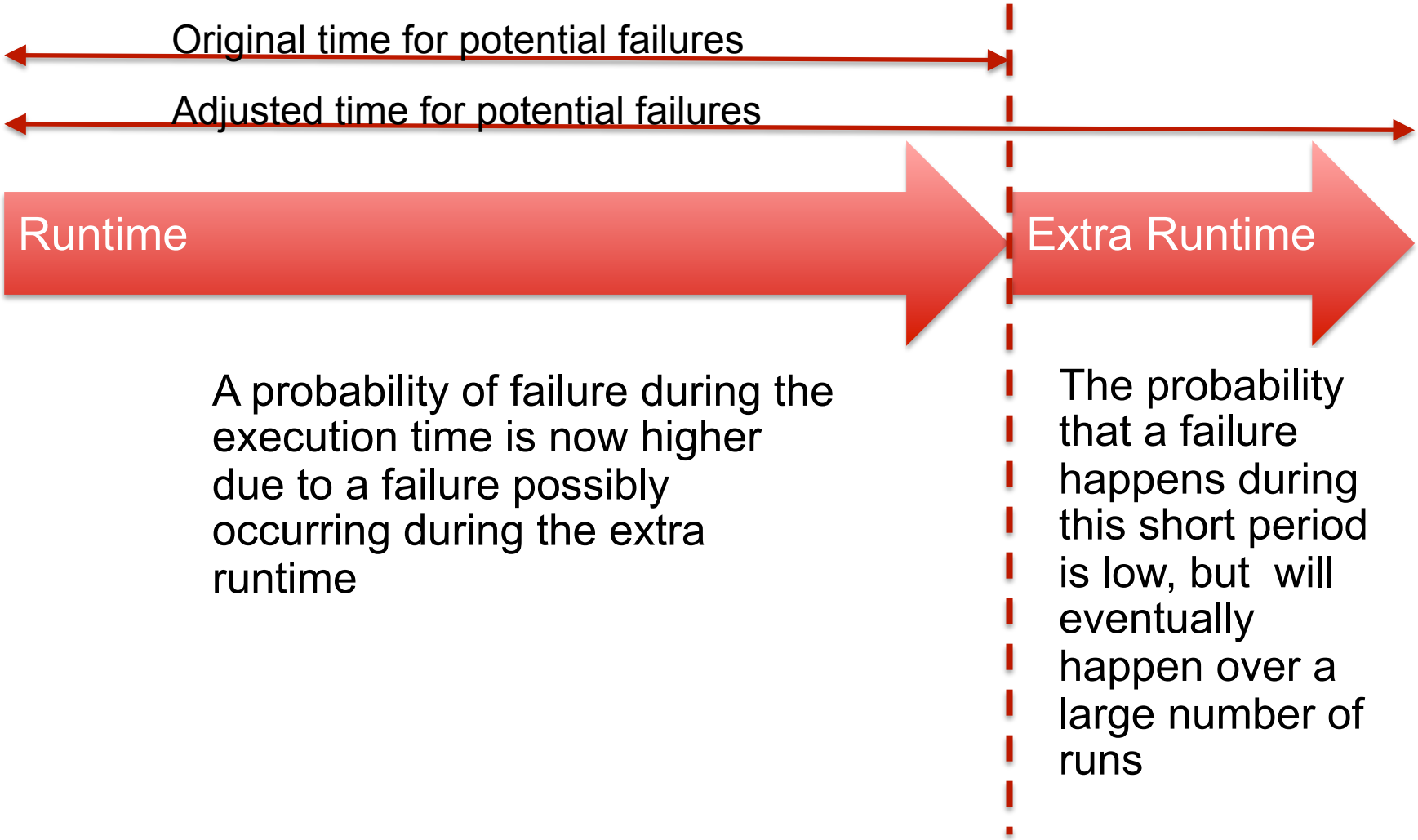
- Recovery requires energy, both to perform the recovery and then to re-compute lost work, E_{fail_recov}

$$E_{fail\_recov} = E_{recov\_operations} + (2 \times E_{lost\_work})$$

- Energy can be re-calculated as the energy consumed during runtime and the energy consumed for recovering from failures, both during the regular runtime and the extended runtime.

$$Energy = E_{successful\_runtime} + (E_{fail\_recov} \times (p(fail) + p(fail_{add\_rt}))$$

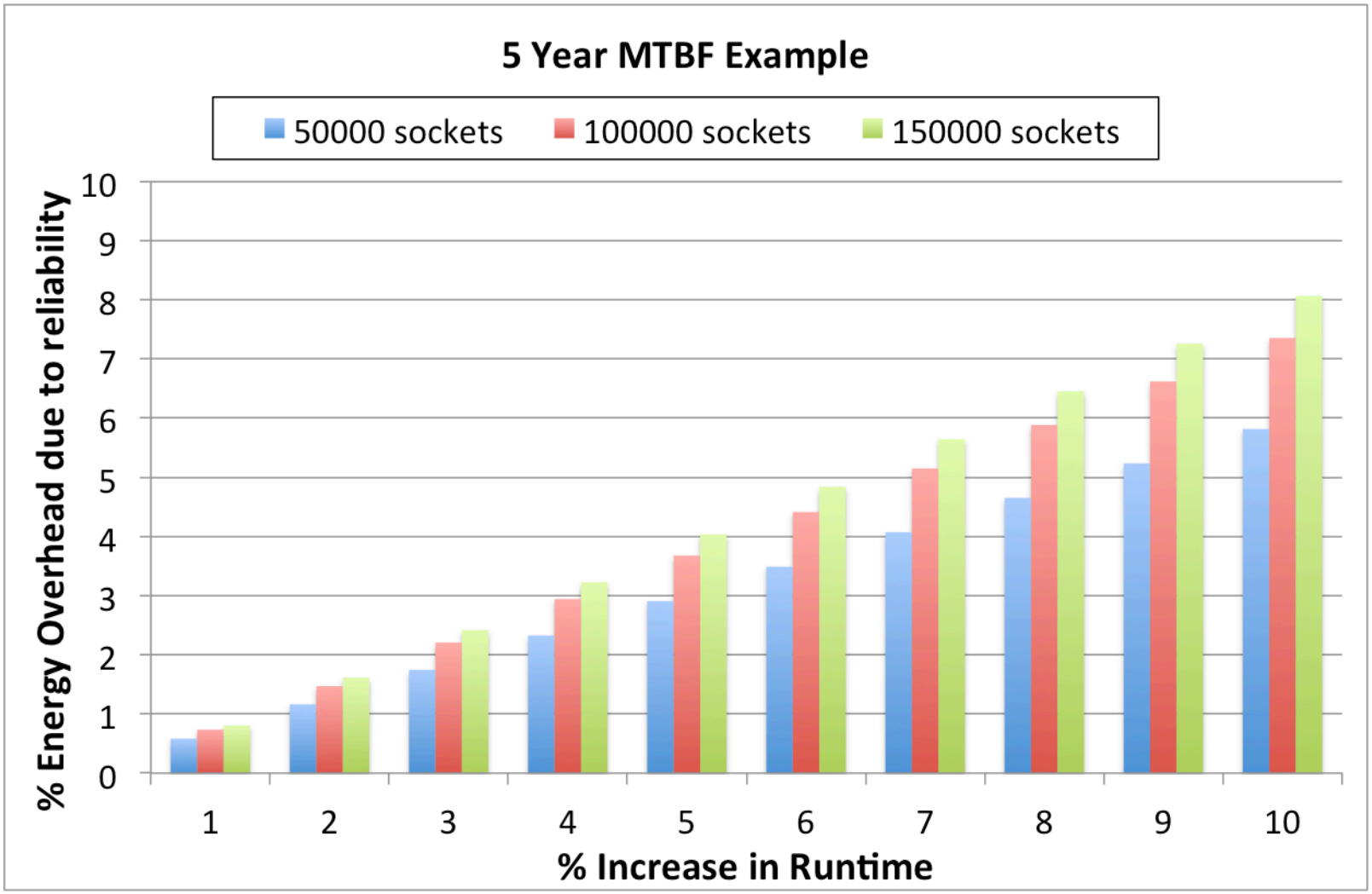
Impact of Reliability on Energy



Impact of Reliability on Energy

Sandia National Laboratories

Increases in runtime need to be offset by energy savings to break even



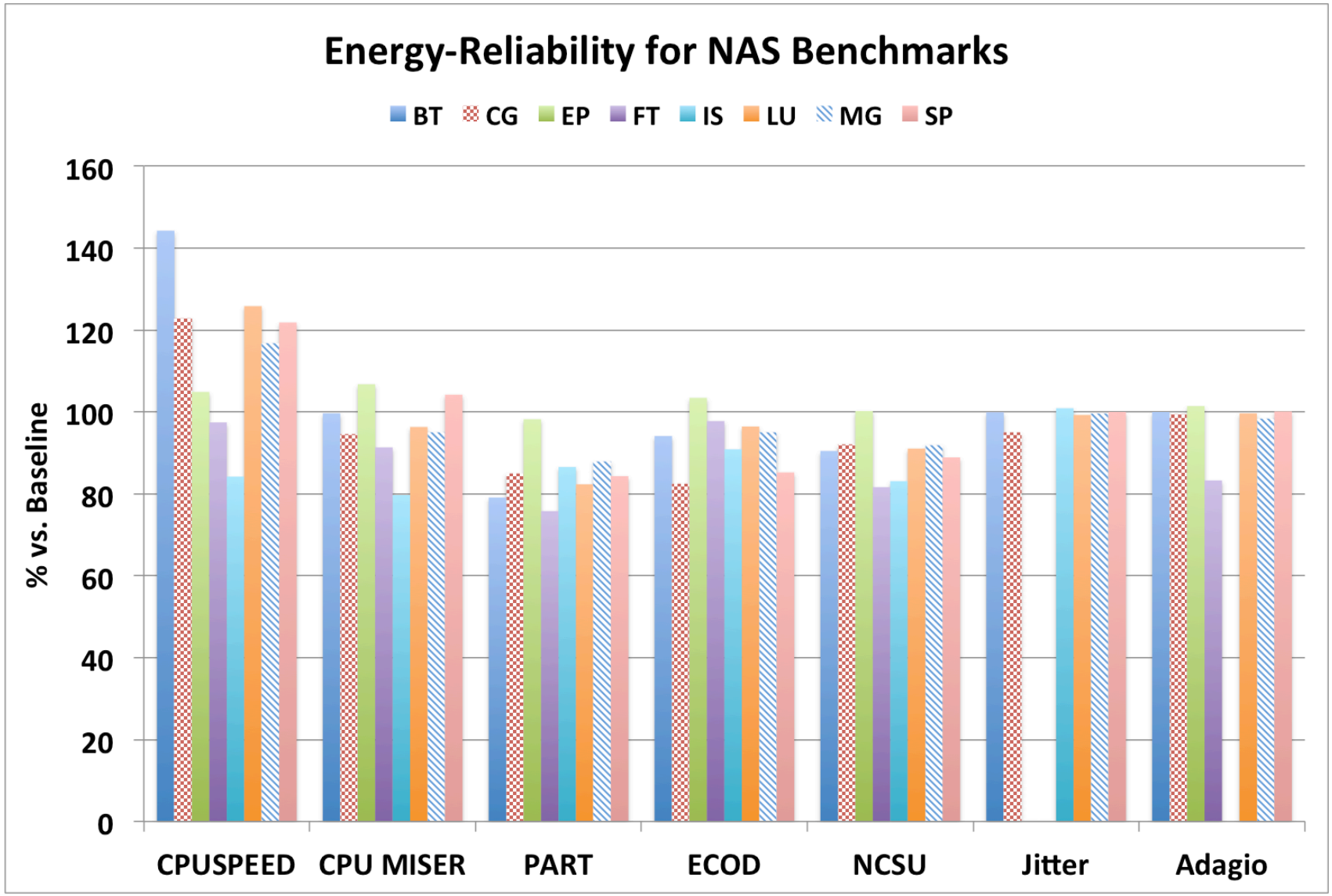
Retrospective on Techniques

Sandia National Laboratories

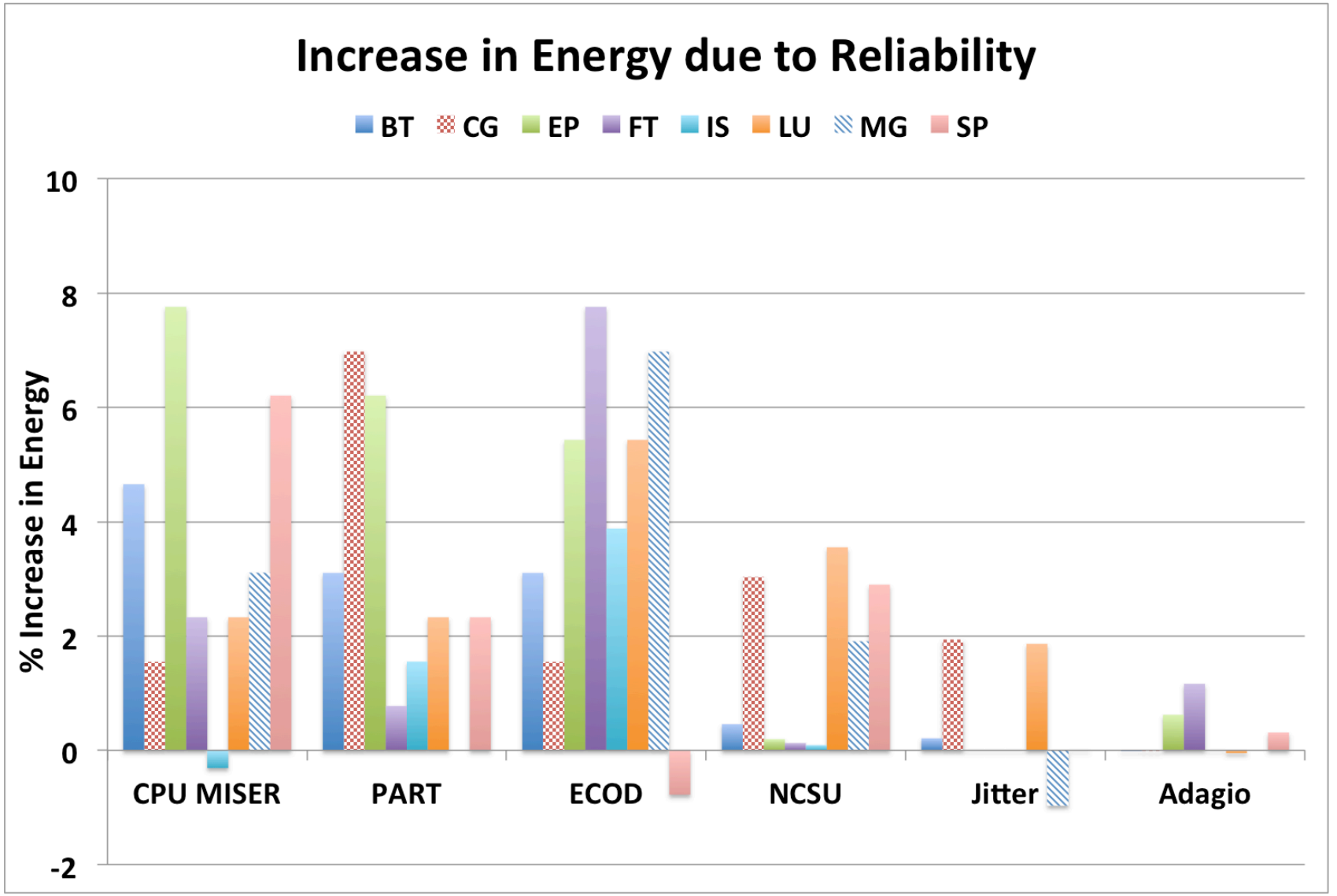
- Analyze history of energy saving methods
  - NAS benchmarks – common comparison point
- Study:
  - CPUSpeed (2005)
  - CPU Miser (2007)
  - PART (2005)
  - ECOD (2009)
  - NCSU method from 2006
  - Jitter (2005)
  - Adagio (2009)
  - Green Queue (2012)

Analysis of Techniques

Sandia National Laboratories

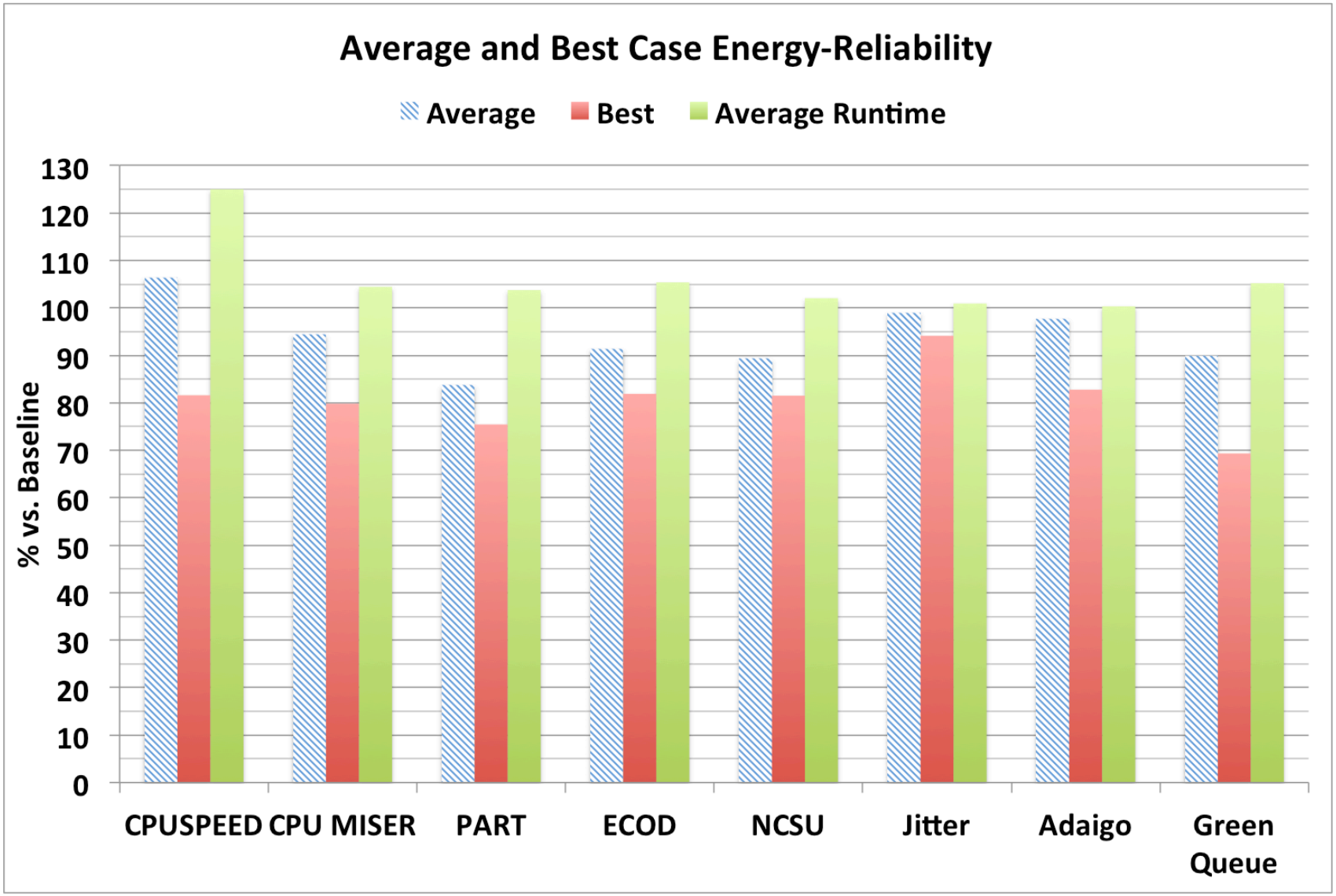


Analysis of Techniques



Analysis of Techniques

Sandia National Laboratories



Analysis of Techniques

Sandia National Laboratories

- Results show that techniques are still quite helpful
- The most aggressive energy saving techniques suffer the most from the reliability adjustment
  - Have longer runtimes in exchange for greater energy savings
  - Increases probability of failures happening during run
- Some techniques have applications that benefit from the addition of reliability concerns
- Benefits of running at lower temperatures not explored

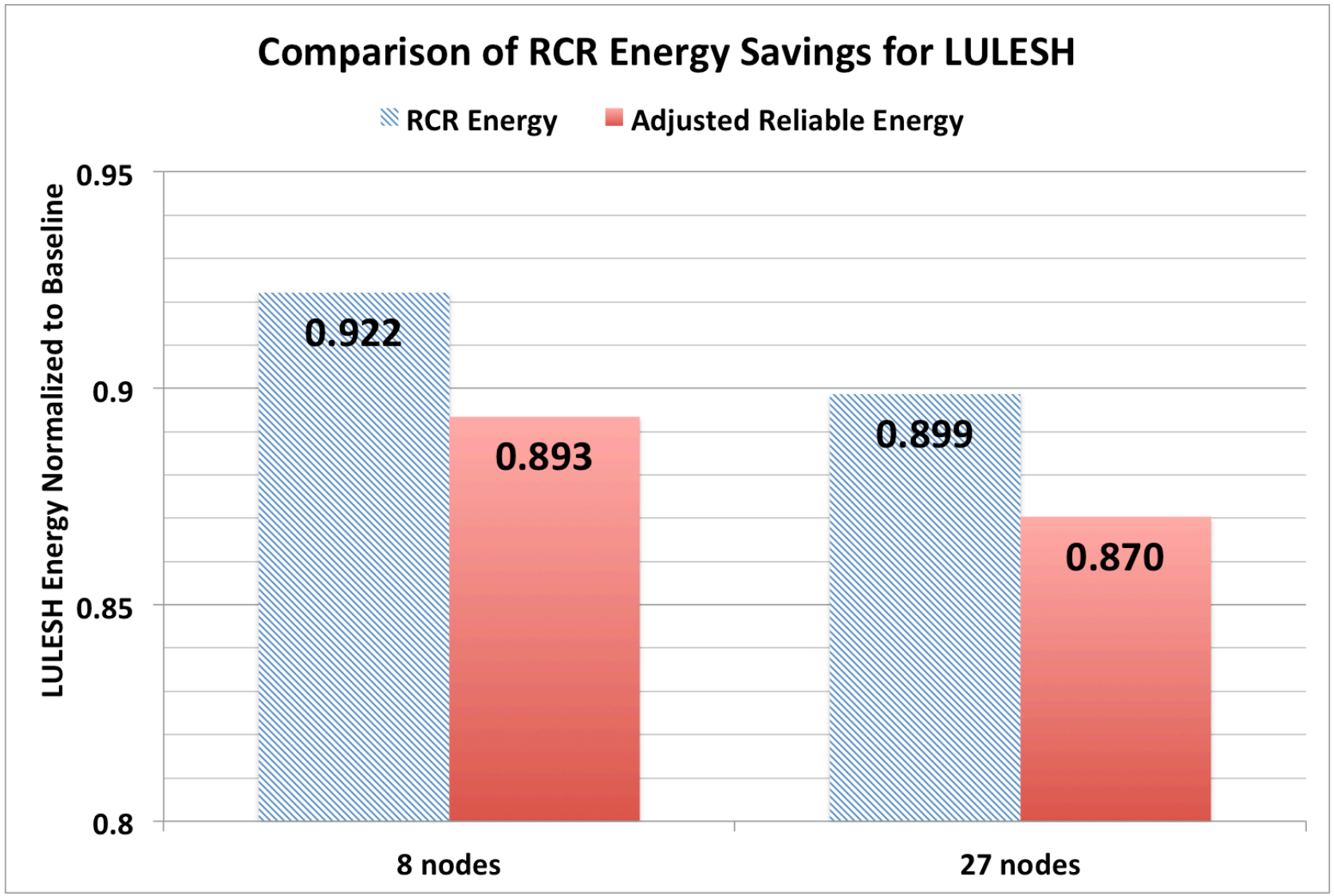
Case Study

Sandia National Laboratories

- Studied MAESTRO/RCR energy-efficiency techniques
  - Detects memory bandwidth saturation and reduces thread concurrency
  - Scales back processor frequency on some threads to reduce memory pressure
- Results improve by considering reliability as runtimes are improved
- Overall, a 2.9% average improvement in energy savings numbers due to reliability

Exascale Computing

Sandia National Laboratories



Conclusion

Sandia National Laboratories

- There is a need to take reliability into account when comparing energy saving techniques for extreme scale HPC
- For approaches that can reduce runtime, reliability considerations are beneficial to energy savings
- Methods that have very little performance overhead scale well with extreme scale reliability concerns
- This approach is resilience method agnostic
  - Works for different checkpointing and replication approaches

Thank you

Sandia National Laboratories

Questions?