SAND2015-0984PE

Data Analytics, Exascale Architectures and Computer Science: The Path to Tomorrow

Bruce Hendrickson

Director of Computing Research

Sandia National Laboratories, Albuquerque, NM

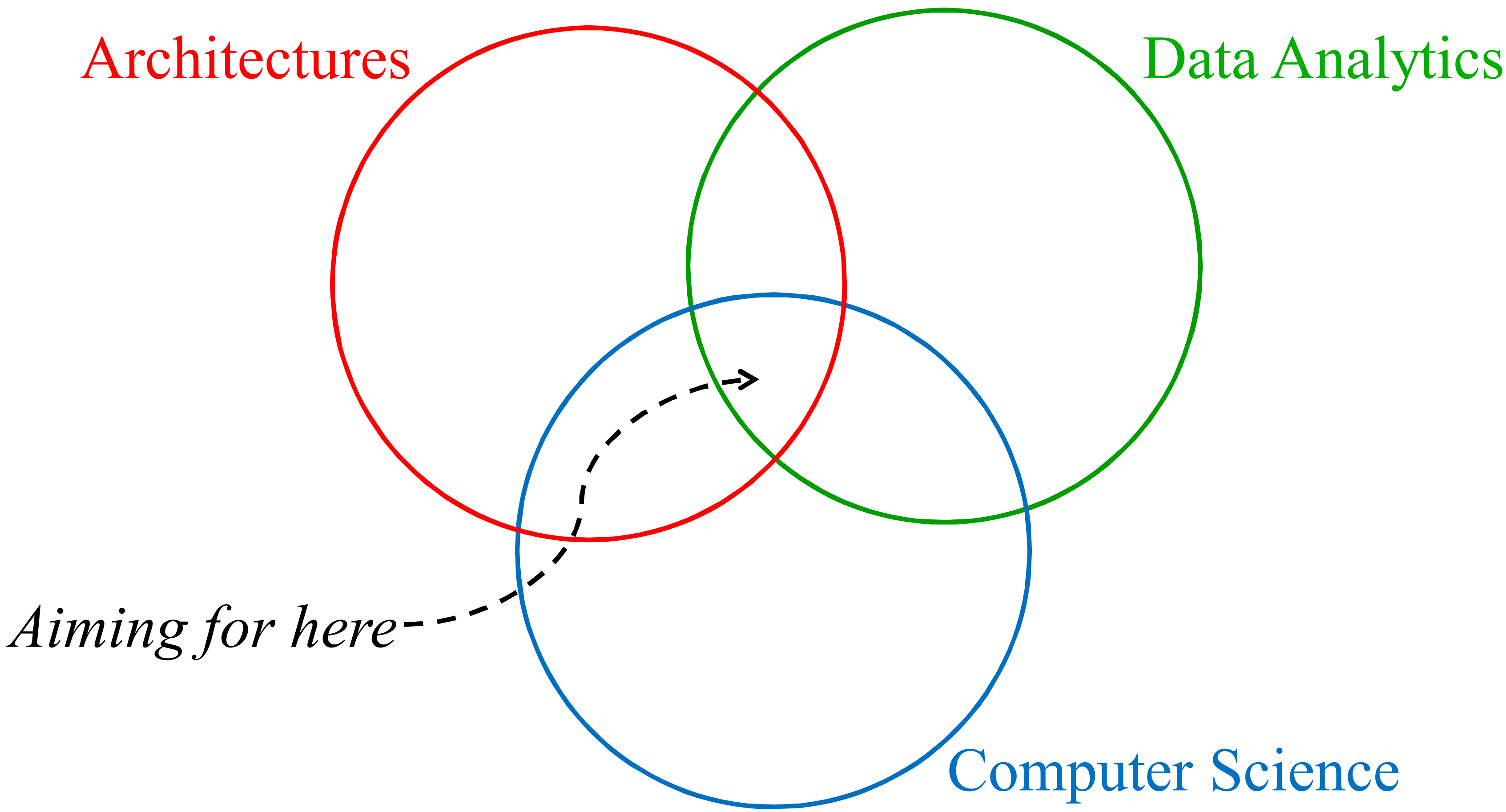
University of New Mexico, Computer Science Dept.

UNM

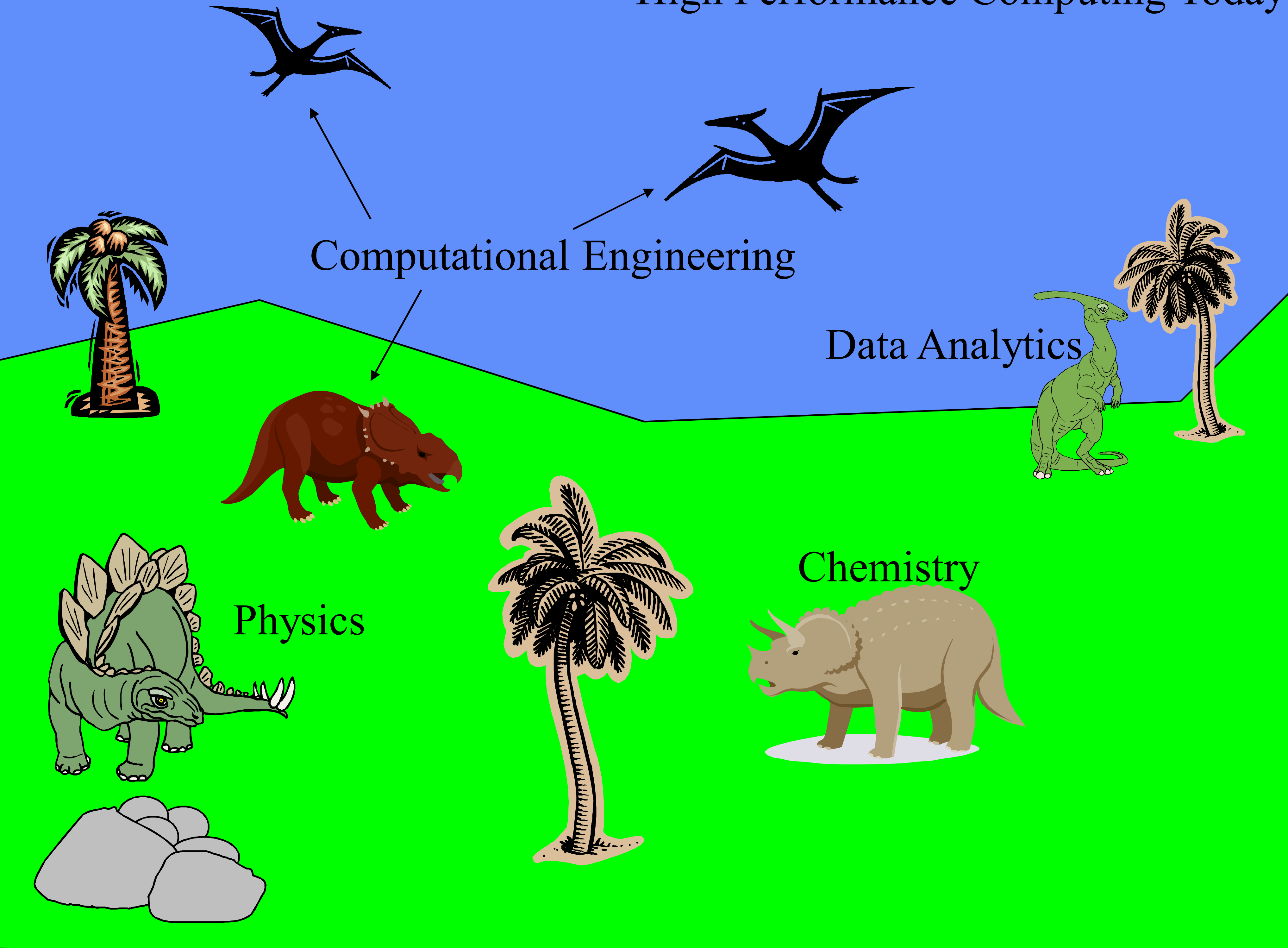
Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

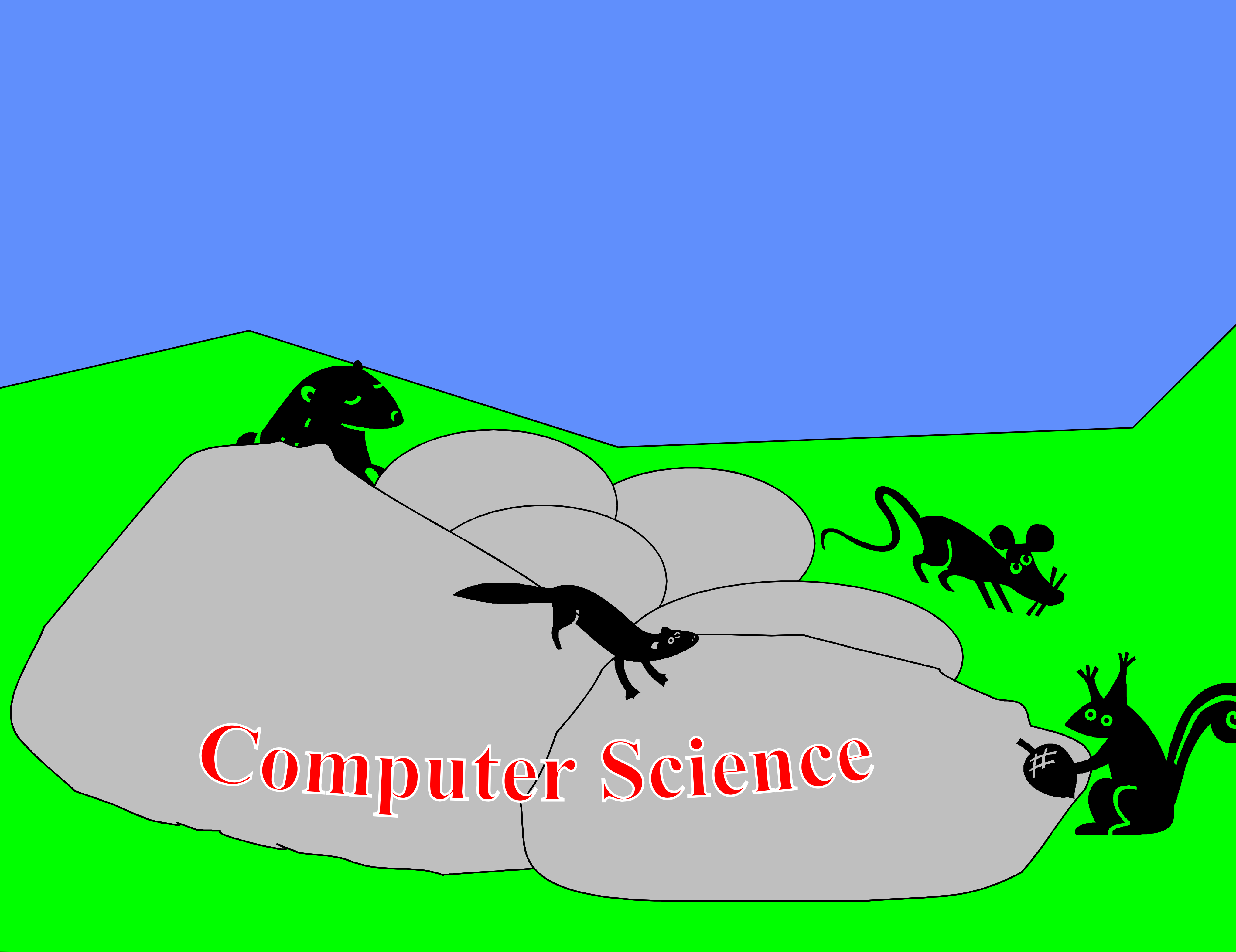
Sandia National Laboratories

Outline


Architectures
Data Analytics
Aiming for here
Computer Science

High Performance Computing Today




Computer Science

We Live in Exciting Times

- **Computing is changing the world**
  - How we learn and how we decide
  - How be design and how we build
  - How we consume and how we live
- **National Strategic Computing Initiative and DOE Exascale Initiative**
  - Great opportunities for DOE Labs to advance the frontiers, create new science, and advance national security
- **But great opportunities come with great challenges**
  - Tomorrow's computing will need to be very different from yesterday's

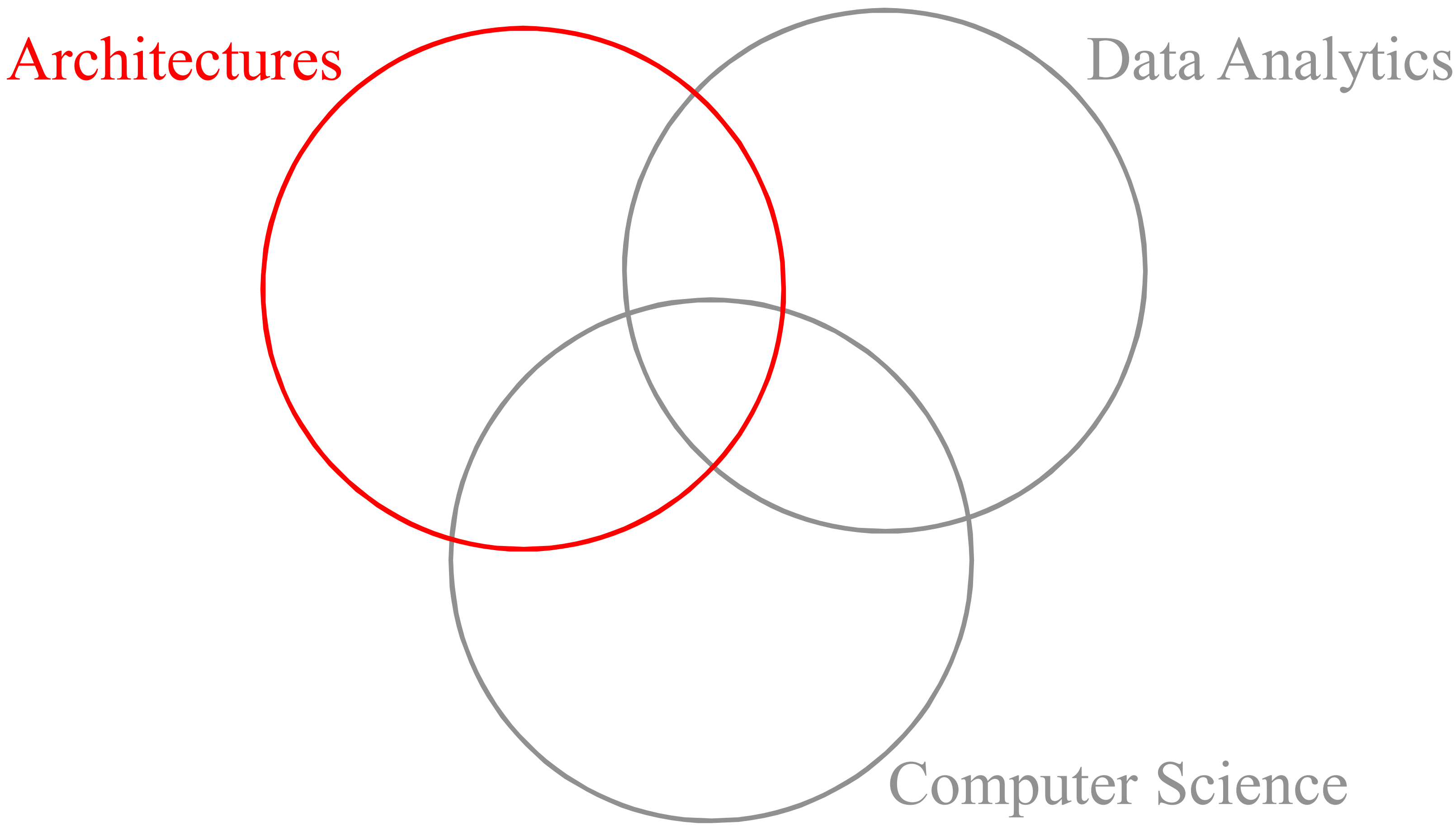
Sandia National Laboratories

Scientific Applications Getting More Complex

- **Higher fidelity simulations increasingly use:**
  - Unstructured and adaptive meshes
  - Multi-physics and multi-scale simulations
- **We're using simulation in more sophisticated ways**
  - Design optimization
  - Uncertainty quantification
- **These trends stress our memories and networks**
  - Data movement dominates performance and power consumption
- **What do we want our future machines to look like?**
  - And how do we get them built?

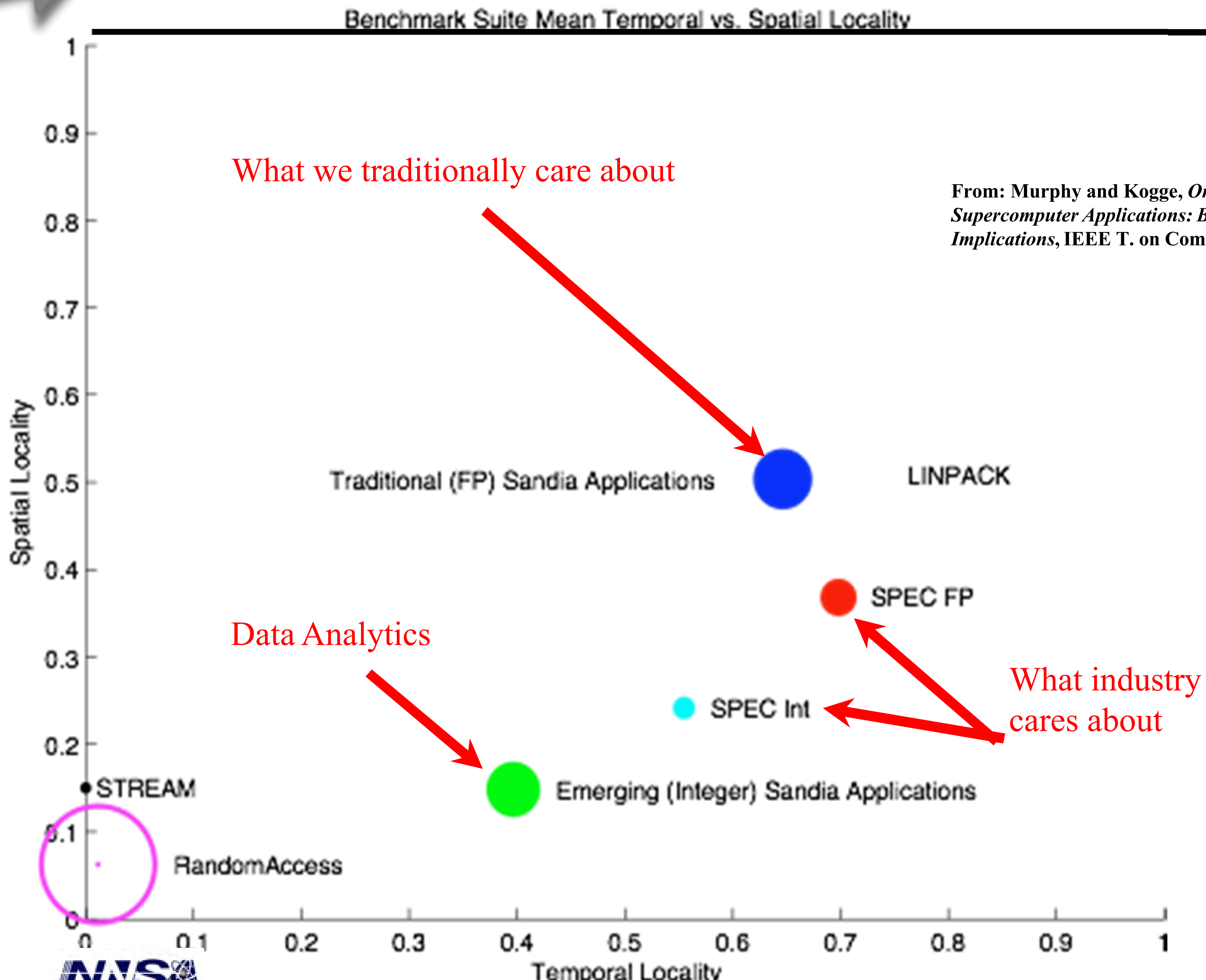
Sandia National Laboratories

Outline



Sandia National Laboratories

Memory Impact of Changing Applications



From: Murphy and Kogge, *On The Memory Access Patterns of Supercomputer Applications: Benchmark Selection and Its Implications*, IEEE T. on Computers, July 2007

NNSA National Nuclear Security Administration

Sandia National Laboratories

A Virtuous Circle…

Commodity Clusters

Architectures

Explicit Message Passing

Programming Models

Software

MPI

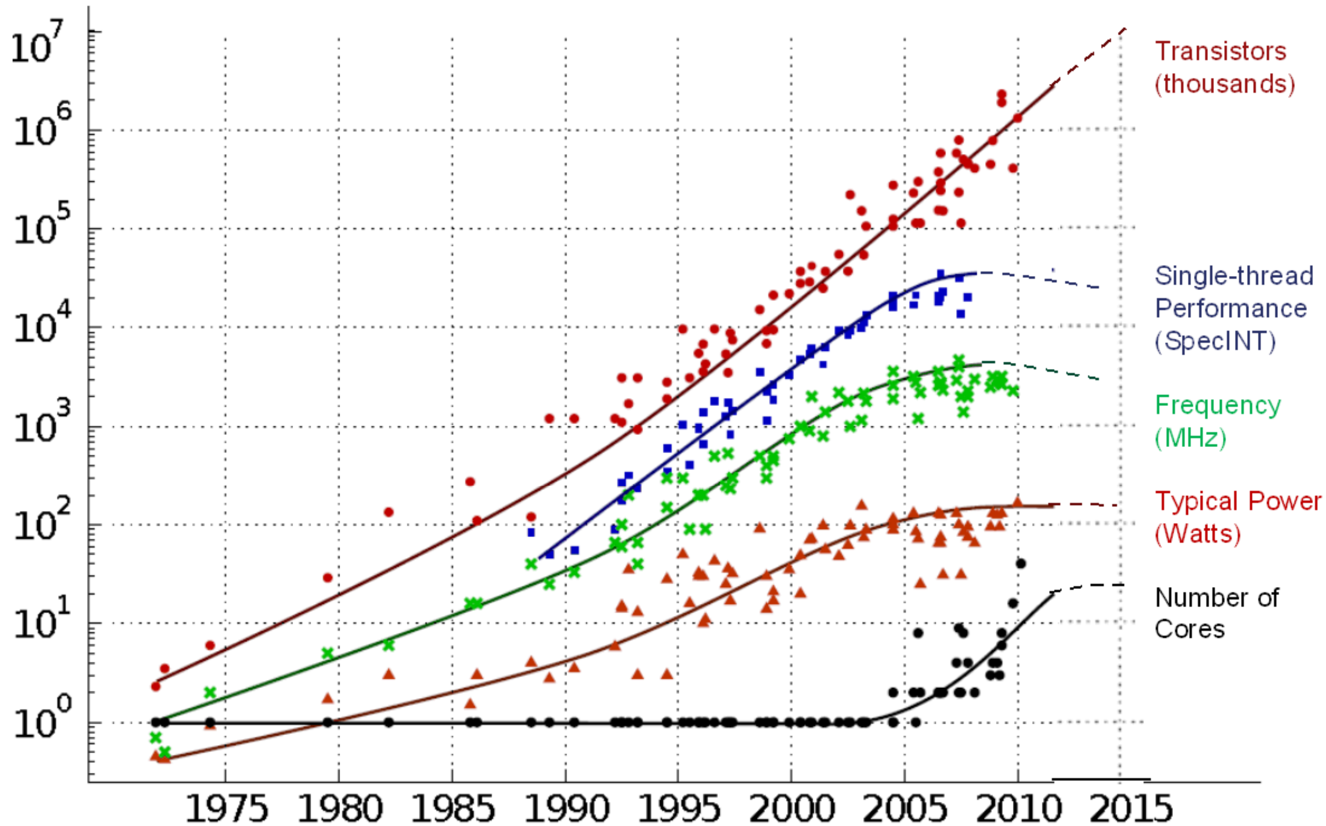
Algorithms

Bulk Synchronous Parallel

Sandia National Laboratories

… is Coming to an End

35 YEARS OF MICROPROCESSOR TREND DATA



Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

Moore's Law continues
Transistor count still doubles every 24 months

Dennard scaling stalls – key parameters flatline:
Voltage
Clock Speed
Power
Performance/clock

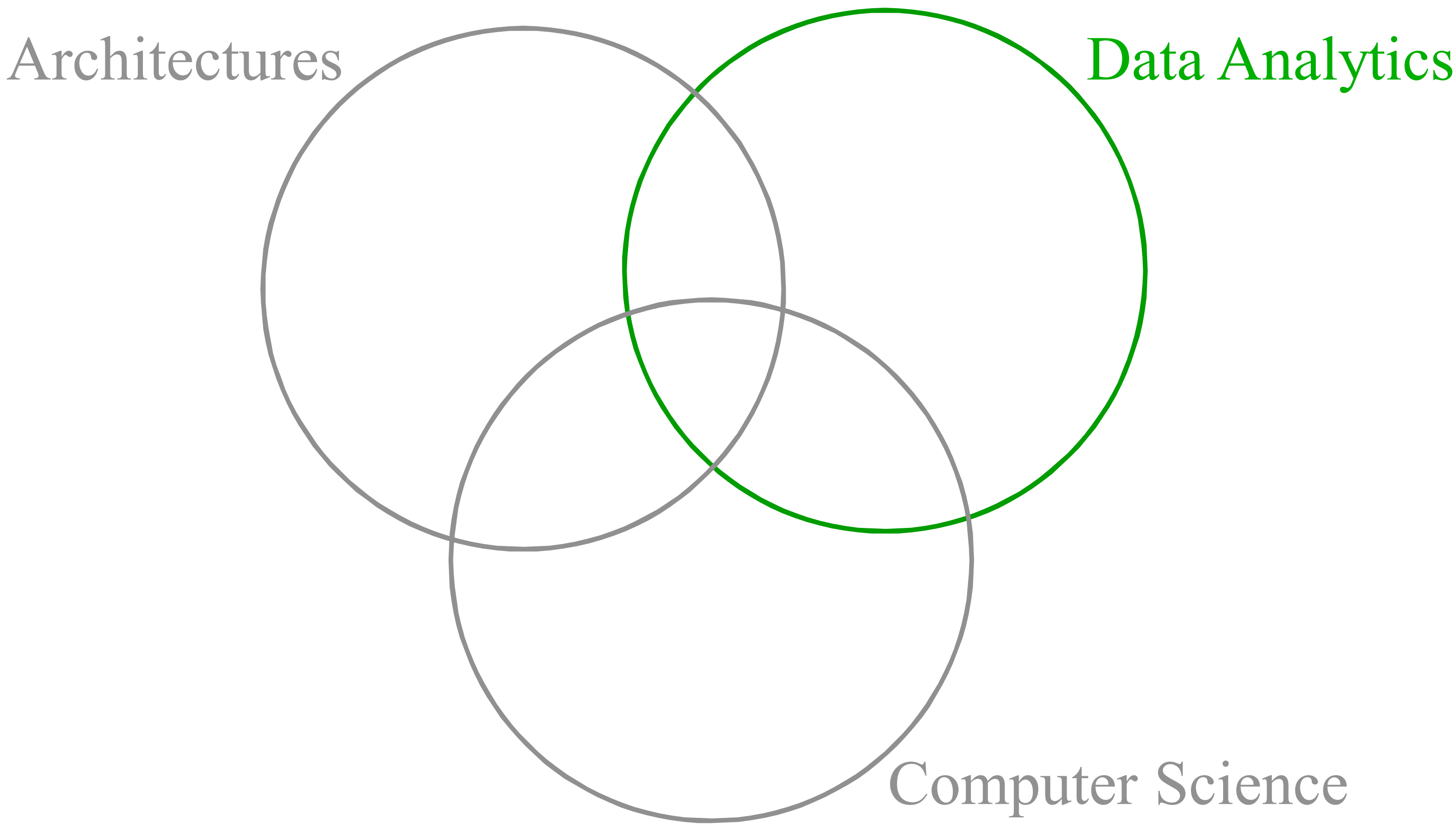
Power is the *Real* Enemy

- **From Kogge's exascale study for DARPA:**
  - Cannot "project any combination of currently mature technologies that will deliver sufficiently powerful systems … at the desired power levels."
  - "easier to solve the power problem associated with base computation than it will be to reduce the problem of transporting data"
- **Architectural constraints are driving technology changes**
  - Heterogeneous cores including simpler, power-efficient ones
  - 3D stacked memory
  - Need to move computation closer to memory

Implications of Architectural Changes

- **Need to avoid global synchronizations**
  - Move away from bulk synchronous computing
  - Adaptive, dynamic scheduling
- **Richer runtimes to support adaptivity and manage resilience**
- **Programming models that hide architectural complexity from application programmer**
- **Details of future machines are very murky**
  - How can we protect vast software base from this uncertainty?

Sandia National Laboratories



Sandia National Laboratories

Data Science is Driving the Bus

- **Computer vendors are (appropriately) focused on data-science opportunities**
  - Business analytics, social networking, biomedicine, mobile applications, etc.
  - Scientific computing will build machines out of parts designed principally for the data-centric market
- **Data is growing rapidly in volume, complexity and speed**
  - Need for faster machines and algorithms just to keep up
  - Growing use of sophisticated analytics and machine learning are outstripping computing capabilities
  - New paradigms are possible, e.g. neuro-inspired computing
- **How do these challenges drive computer architecture?**
  - In particular, how does this impact exascale computing?

Sandia National Laboratories

High Performance Data Analytics

- **Map-Reduce paradigm has been very successful, but has significant shortcomings**
  - Good for filtering, but struggles with advanced analytics
  - Lots of data movement – energy expensive

- **Some important problems require high performance**
  - Large, complex data sets
  - Need for high performance
    - Time-critical decision making (e.g. cyber security)
    - Expensive algorithms (e.g. deep learning)

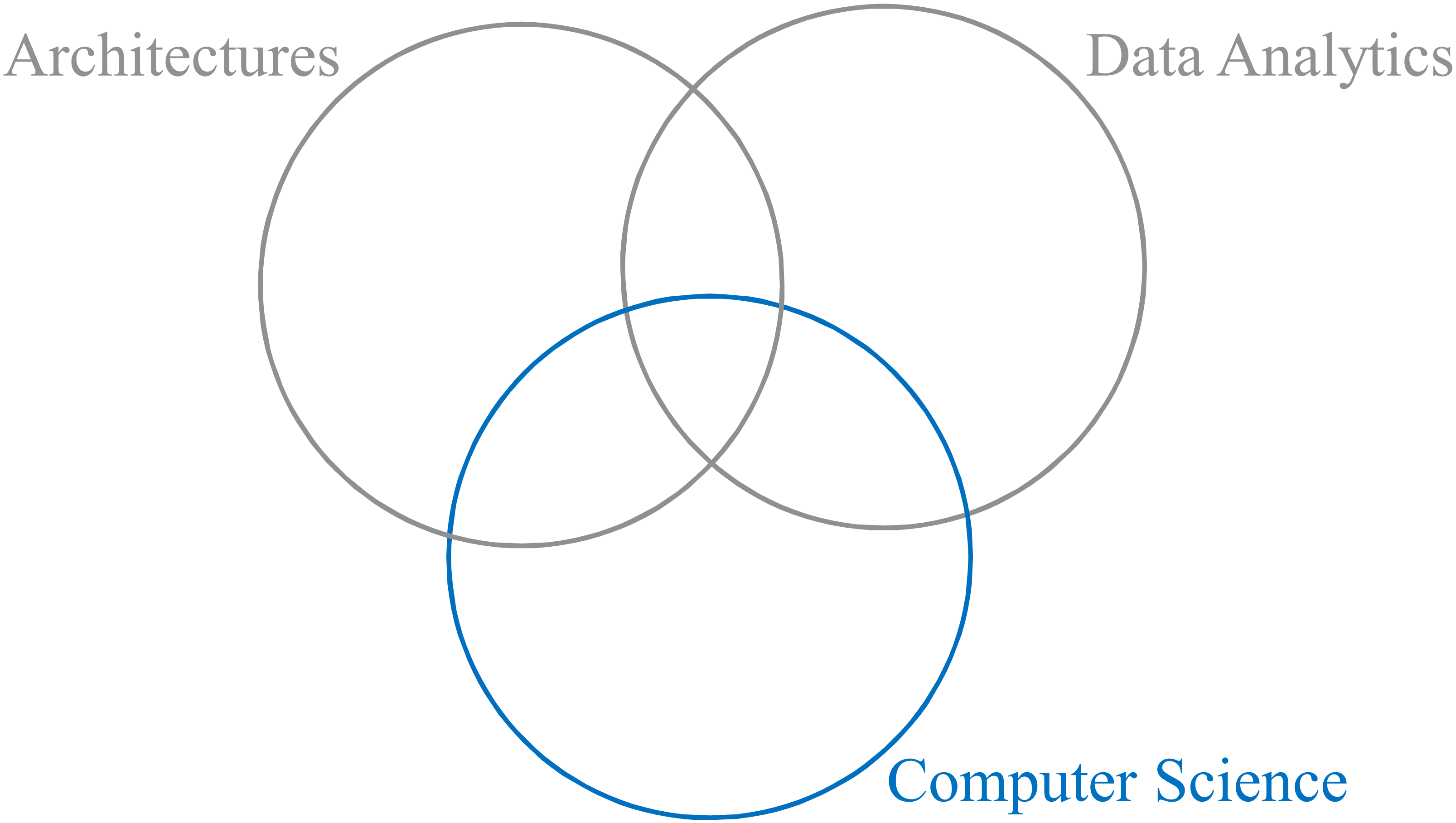
Sandia National Laboratories

Architecture Challenges for HPDA

- **Computations may be highly parallel, but often not bulk synchronous (e.g. graph algorithms)**
- **Data movement determines performance**
  - Movement within the memory hierarchy
  - Movement across a network
- **Power consumption is a major concern**

- **Architectural needs for HPDA are strongly aligned with those for scientific computing!**

- **These two communities will be confronting similar hardware and software challenges**



Sandia National Laboratories

Programming Future Architectures

- **A daunting litany of challenges:**
  - Manage heterogeneous cores and deep memory hierarchies
  - Move computation closer to data
  - Expose very high degree of light-weight parallelism
  - Insulate computation from growing frequency of errors
- **Application Programmers can't be expected to manage all this directly**
  - Need new ideas for hiding this complexity
- **Worse still, we don't yet know what these machines will look like!**
  - How do we write future-proof code now?
  - Key is finding the right conceptual and software abstractions

Sandia National Laboratories

Appropriate Abstractions Are Critical

- **Simplified machine model for programmers**
  - Simple interface for managing (or hiding) node heterogeneity
  - Managing resilience
  - Dealing with complex memory hierarchies

- **Performance portability across diverse architectures**

- **These need to intersect in a natural way with our application / library / runtime software stack**

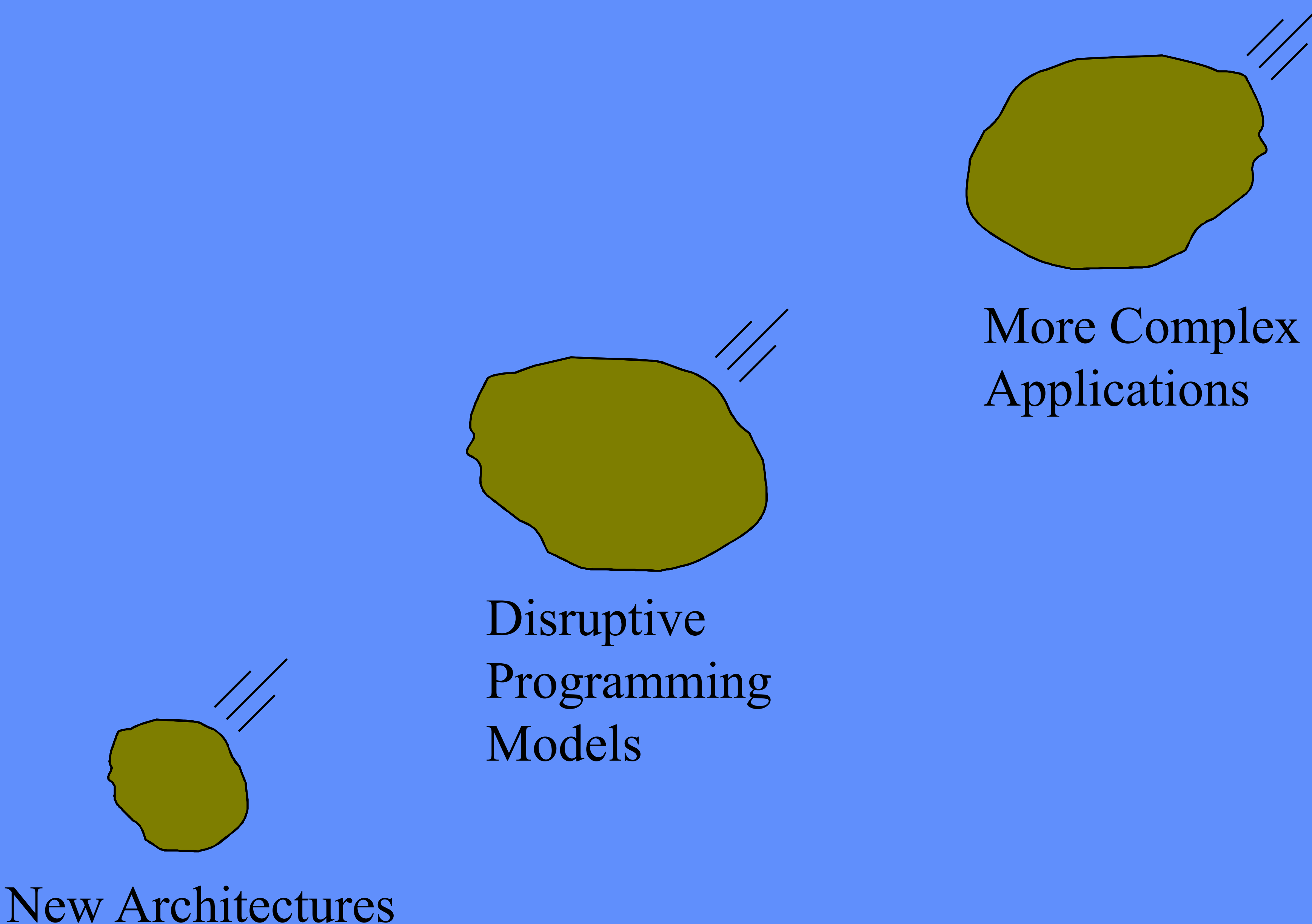
Sandia National Laboratories

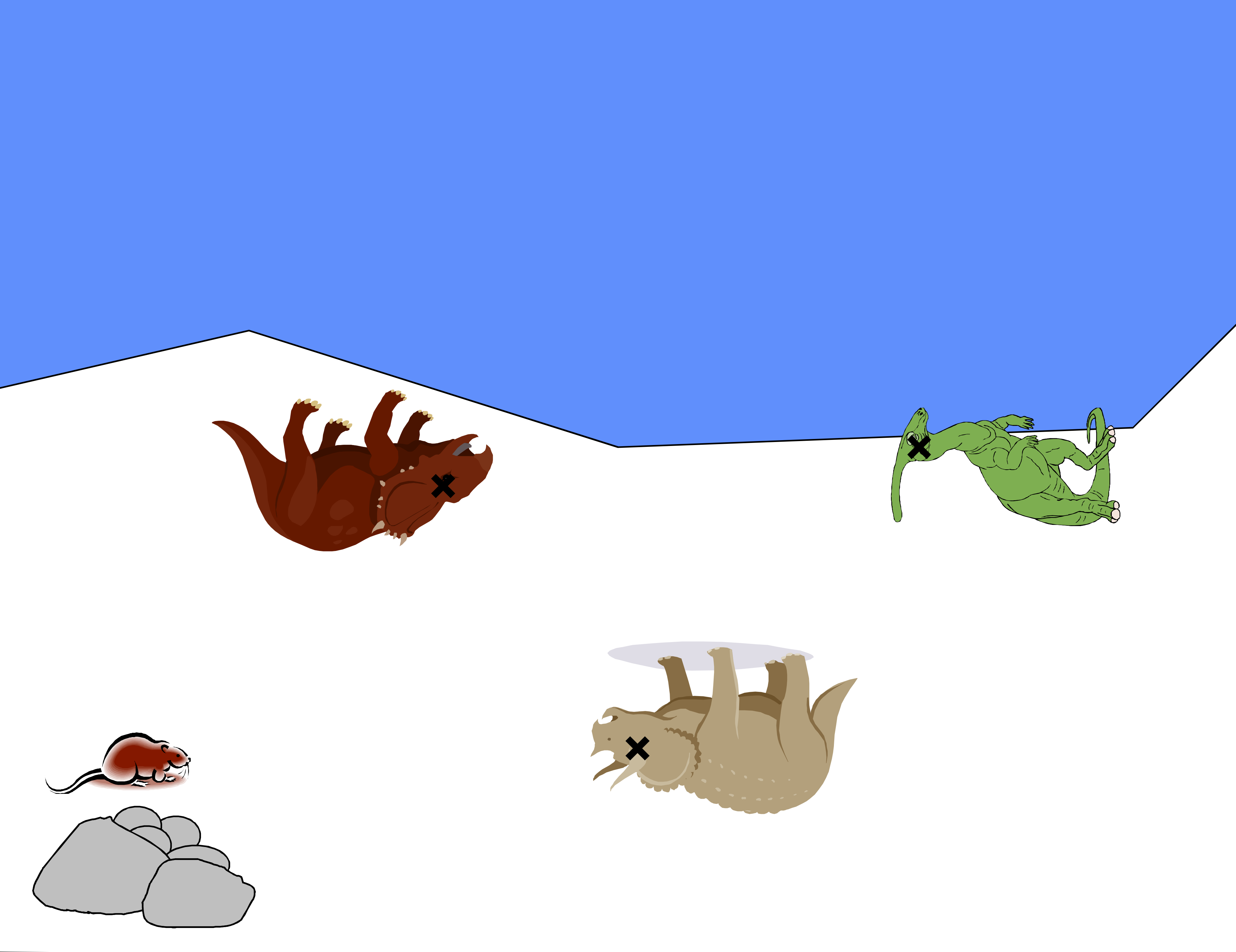
Promising Abstractions

- **Task-based programming models**
  - E.g. Charm++, Legion, Uintah, PaRSEC
  - Create many more tasks than processors
  - Schedule tasks at runtime
  - Allows for performance portability of high-level code

- **Kokkos memory abstractions**
  - Polymorphic Multidimensional Arrays
  - Decouple array layout and what memory space from algorithm
  - Match layout to architecture without modifying algorithm's implementation
  - Supports performance portability across different architectures
  - Employs template metaprogramming (another CS contribution)
  - By Carter Edwards and others at Sandia Labs

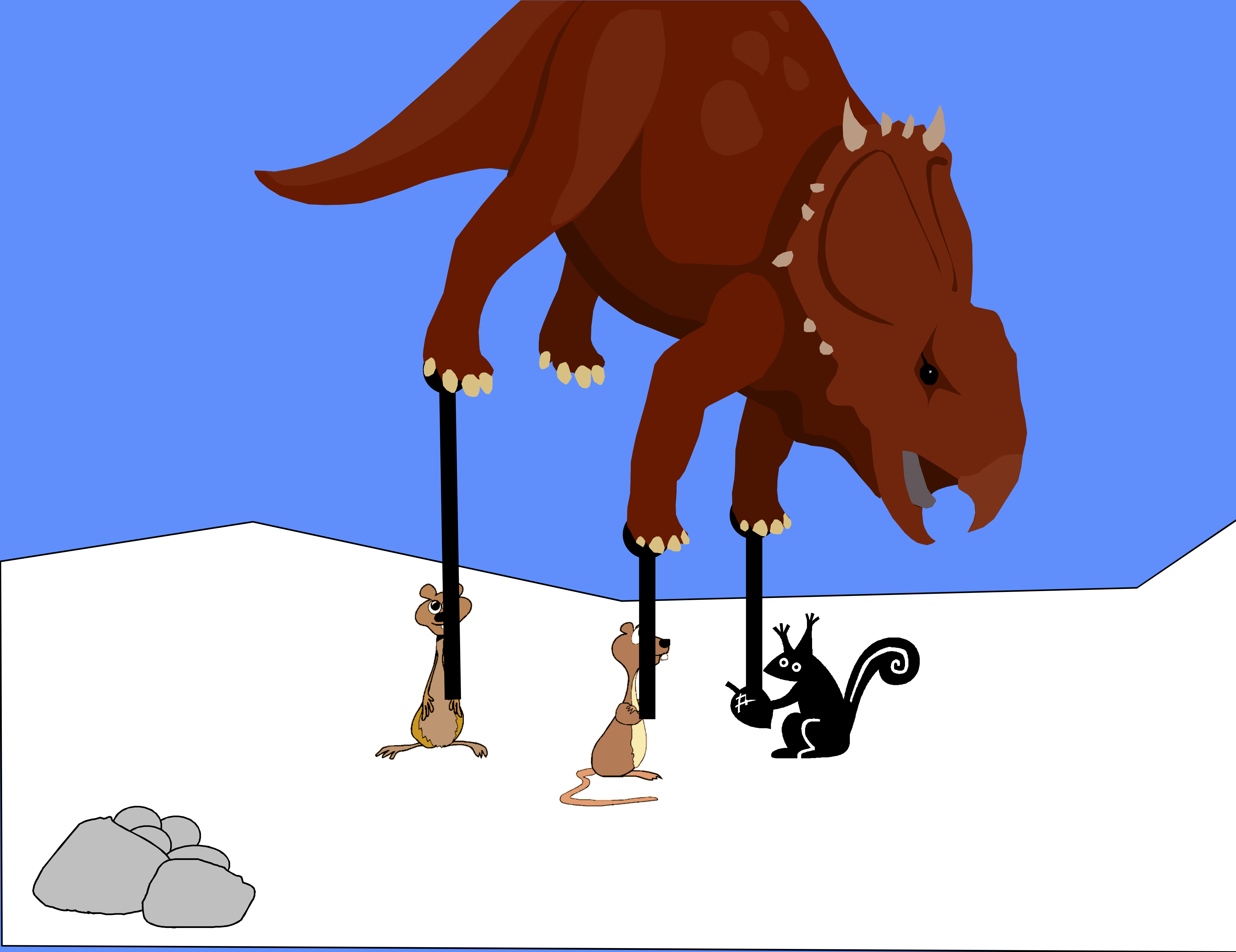
Sandia National Laboratories


More Complex Applications
Disruptive Programming Models
New Architectures



What Happens Next?

- **Virtuous circle will not survive the coming disruptions**
  - Our programming models will need to fundamentally change
- **But existing codes cannot be allowed to die**
  - Billions of dollars in investment in software
- **Computer science will have to play an ever larger role**
- **New programming models, algorithms and abstractions will be needed**
  - Need to support *both* scientific computing and data analytics



Conclusions

- **Physical, engineering and financial constraints have ended the era of Dennard scaling**
  - Future performance growth requires more complex architectures
- **Scientific computing and high performance data analytics will be confronting similar challenges**
  - And data-computing will drive the vendor decisions
- **Fresh thinking about programming abstractions is needed**
  - Computer science is crucial

- **Success will open the door to the next era in advanced computing**

- **We live (and work) in exciting times!**

Sandia National Laboratories

Acknowledgements

- Aspects of this talk have been informed and enriched by conversations with the following (and many more)

- Jim Ang
- Bill Carlson
- Carter Edwards
- John Gilbert
- Mike Heroux
- Pete Kogge
- Alex Pothen
- Keith Underwood

Jon Berry
Robert Clay
John Feo
Scott Hemmert
Simon Kahan
Rich Murphy
Steve Reinhardt

Sandia National Laboratories