SAND2014-20091C

Sandia National Laboratories

Exceptional service in the national interest

Extreme-Scale In-Situ Data Analysis

Janine C. Bennett

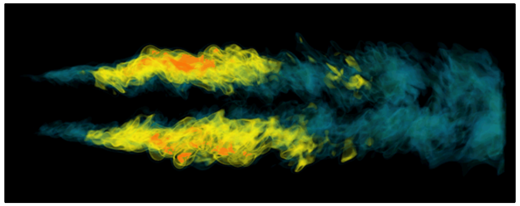
Sandia National Laboratories

Broader Engagement HPC Application Panel

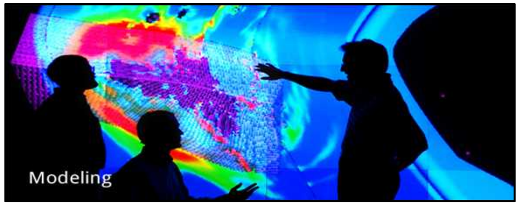
Nov 16, 2014

U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

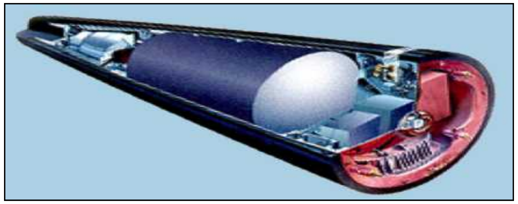
Leadership-class HPC capabilities are required for DOE policy and decision making



Energy: Reduce U.S. reliance on foreign energy, reduce carbon footprint



Climate change: Understand, mitigate, and adapt to the effects of global warming



National Nuclear Security: Maintain a safe, secure, and reliable nuclear stockpile

Exascale computing and beyond is required to simulate complex phenomena that characterize the DOE mission space

Simulations generate large, complex data sets

- Case study: Direct Numerical Simulations & turbulent combustion
- Data size
  - O(Billions) of grid points per time step
  - O(100K) time steps
- Data complexity
  - Multivariate
    - O(100) chemical species
    - Vector data
    - Particle data
  - Turbulence is a complex phenomenon
  - Length scales: microns to centimeters
  - Temporal scales: nanoseconds to milliseconds

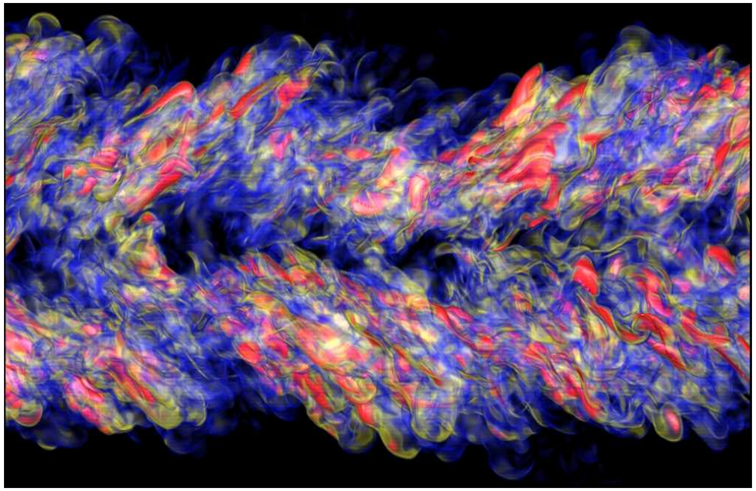
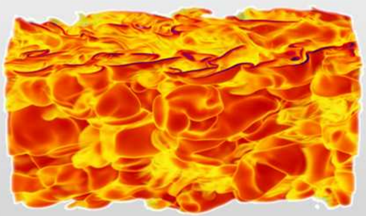
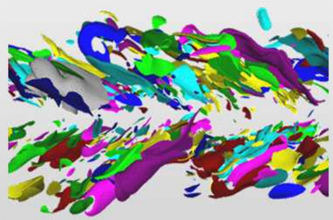
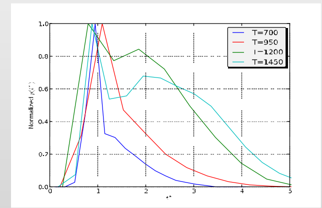


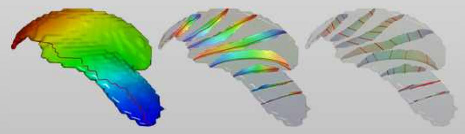
Image courtesy of Hongfeng Yu and Jacqueline Chen

Scientists are interested in analyzing their data in a variety of ways

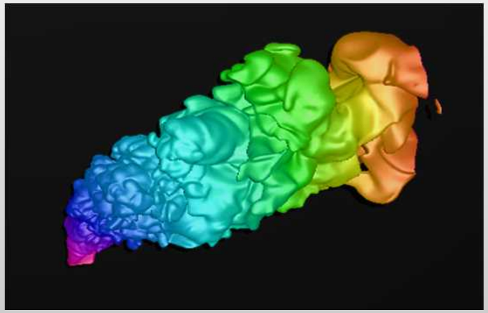




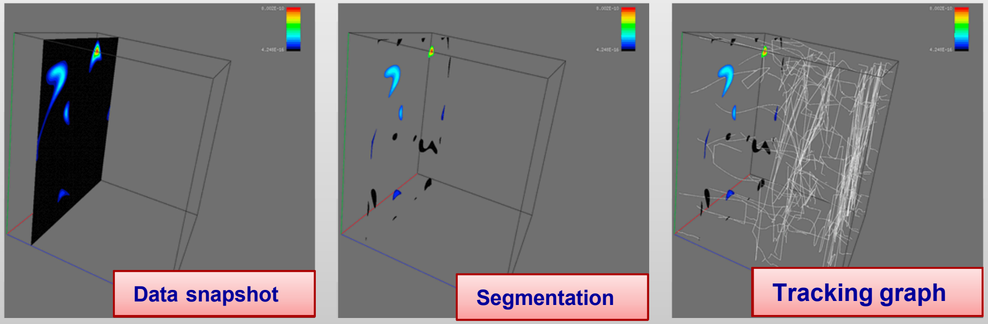




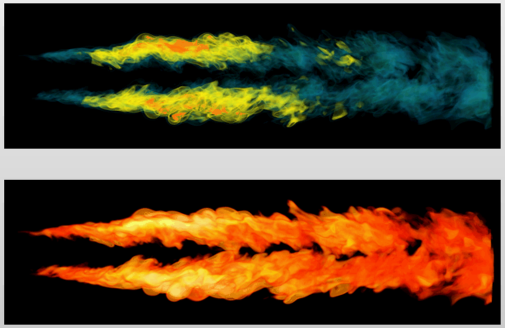
We want to identify features, characterize their shapes and analyze the behavior of other variables within these features



Jet-based coordinate systems allow for aggregation of statistics conditioned on bulk flame position

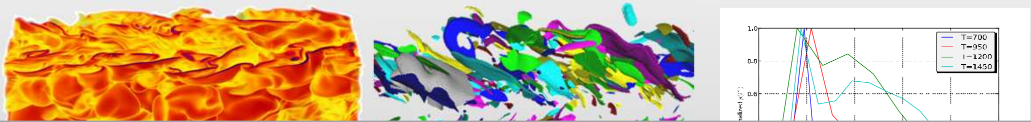


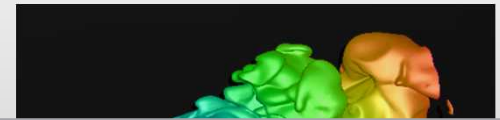
Tracking features in space and time



Visualization provides qualitative analysis results

Scientists are interested in analyzing their data in a variety of ways





- Jacqueline Chen
- Big Data and Combustion Simulation
- BE Plenary on Big Data and Exascale Challenges
- Monday 8:30-9:15, Room 288-289

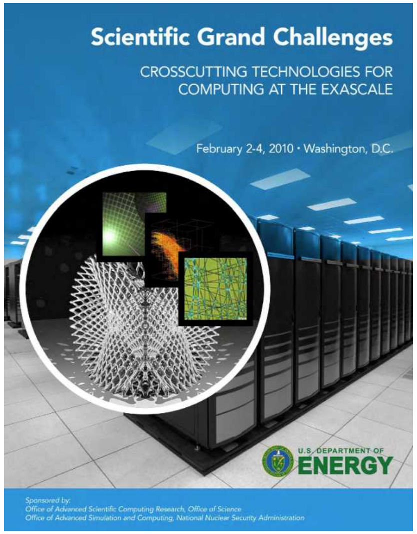
Tracking features in space and time

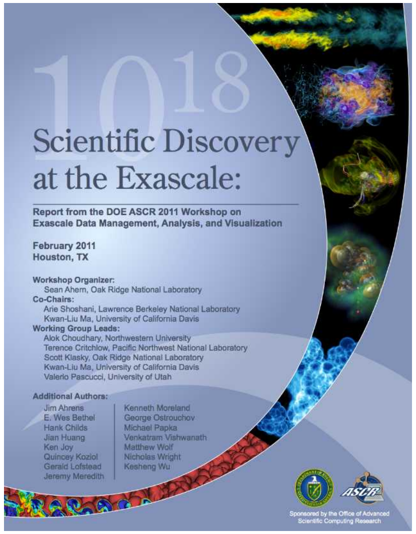
Visualization provides qualitative analysis results

Exascale ≠ Petascale x 1000

Sandia National Laboratories

System Parameter	2011	2018		Factor Change
System Peak	2 Pf/s	1 Ef/s		500
Power	6 MW	≤20 MW		3
System Memory	0.3 PB	32-64 PB		100-200
Total Concurrency	225K	1 BX10	1B X100	40000-400000
Node Performance	125 GF	1 TF	10 TF	8-80
Node Concurrency	12	1000	10000	83-830
Network Bandwidth	1.5 GB/s	100 GB/s	1000 GB/s	66-660
System Size (nodes)	18700	1000000	100000	50-500
I/O Capacity	15 PB	30-100 PB		20-67
I/O Bandwidth	0.2 TB/s	20-60 TB/s		10-30

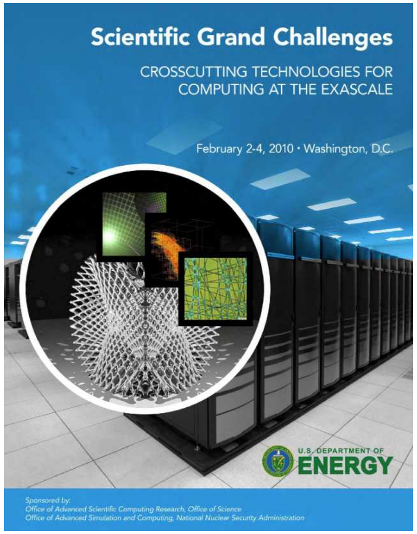


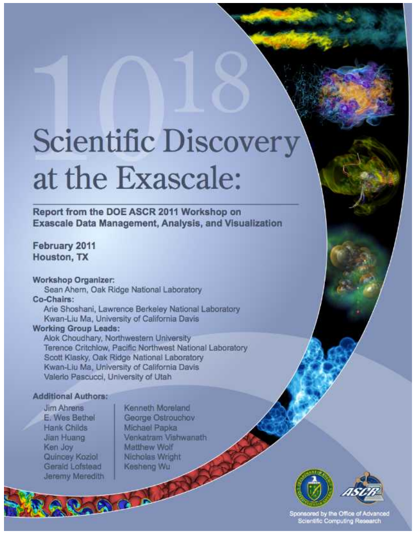


There is a widening gap between compute and I/O capabilities

Sandia National Laboratories

System Parameter	2011	2018		Factor Change
System Peak	2 Pf/s	1 Ef/s		500
Power	6 MW	≤20 MW		3
System Memory	0.3 PB	32-64 PB		100-200
Total Concurrency	225K	1 BX10	1B X100	40000-400000
Node Performance	125 GF	1 TF	10 TF	8-80
Node Concurrency	12	1000	10000	83-830
Network Bandwidth	1.5 GB/s	100 GB/s	1000 GB/s	66-660
System Size (nodes)	18700	1000000	100000	50-500
I/O Capacity	15 PB	30-100 PB		20-67
I/O Bandwidth	0.2 TB/s	20-60 TB/s		10-30



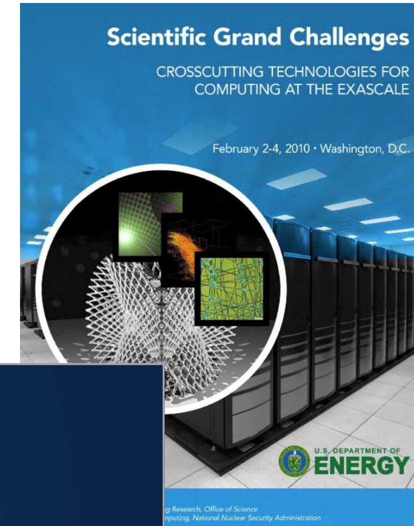


There is a widening gap between compute and I/O capabilities

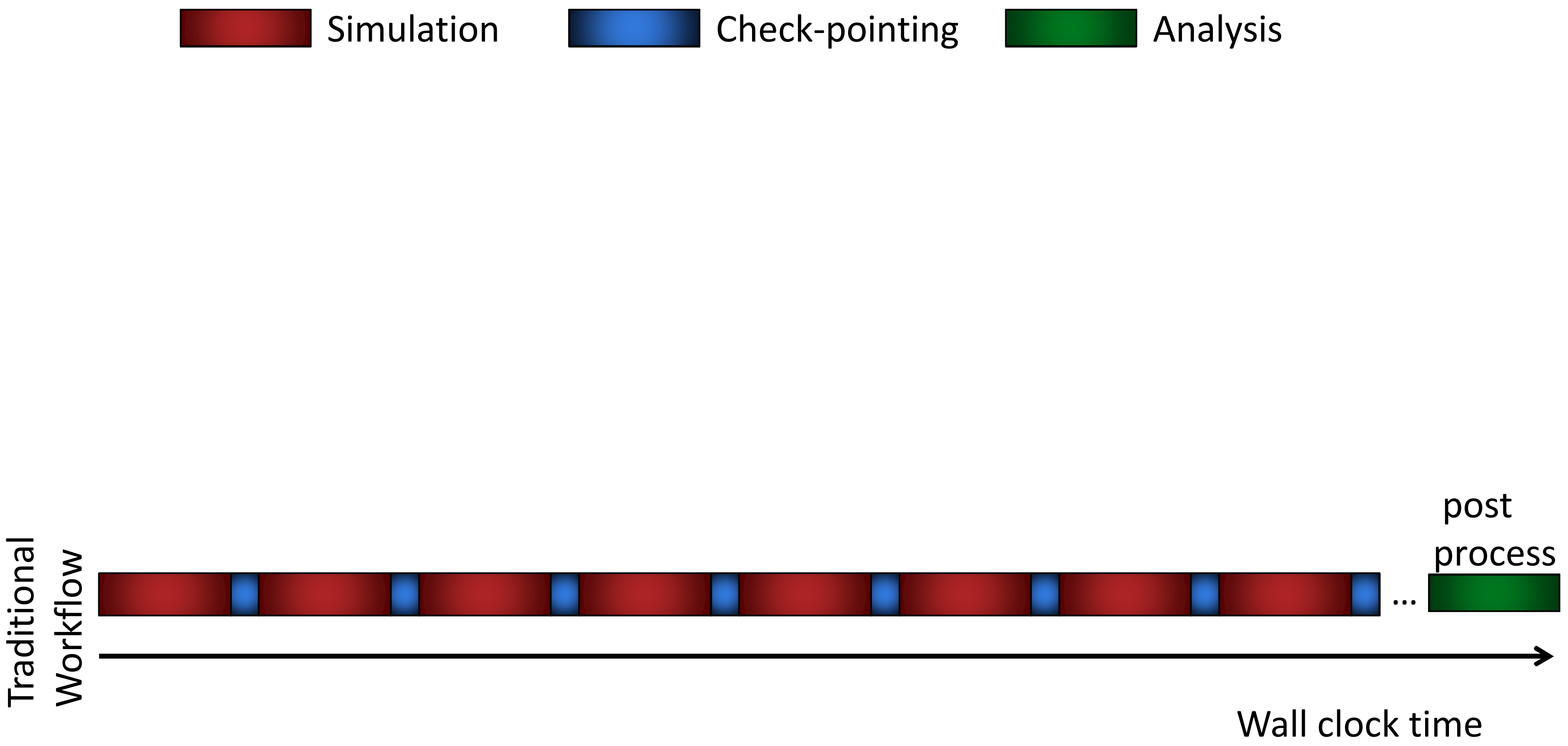
Sandia National Laboratories

System Parameter	2011	2018		Factor Change
System Peak	2 Pf/s	1 Ef/s		500
Power	6 MW	≤20 MW		3
System				
Total Cor				
Node Per				
Node Cor				
Network Bandwidth	1.5 GB/s	100 GB/s	1000 GB/s	66-660
System Size (nodes)	18700	1000000	100000	50-500
I/O Capacity	15 PB	30-100 PB		20-67
I/O Bandwidth	0.2 TB/s	20-60 TB/s		10-30

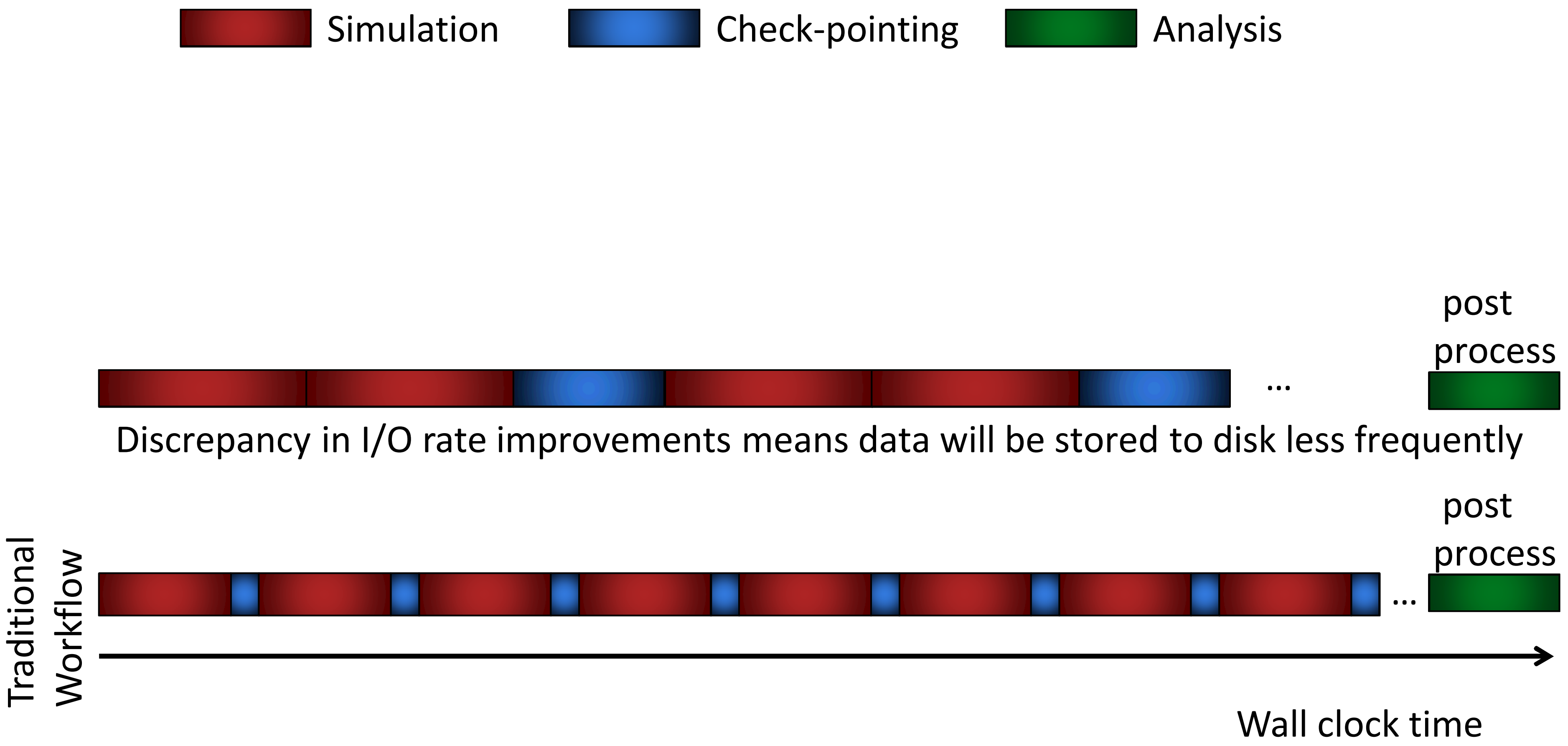
Scientific workflows are changing



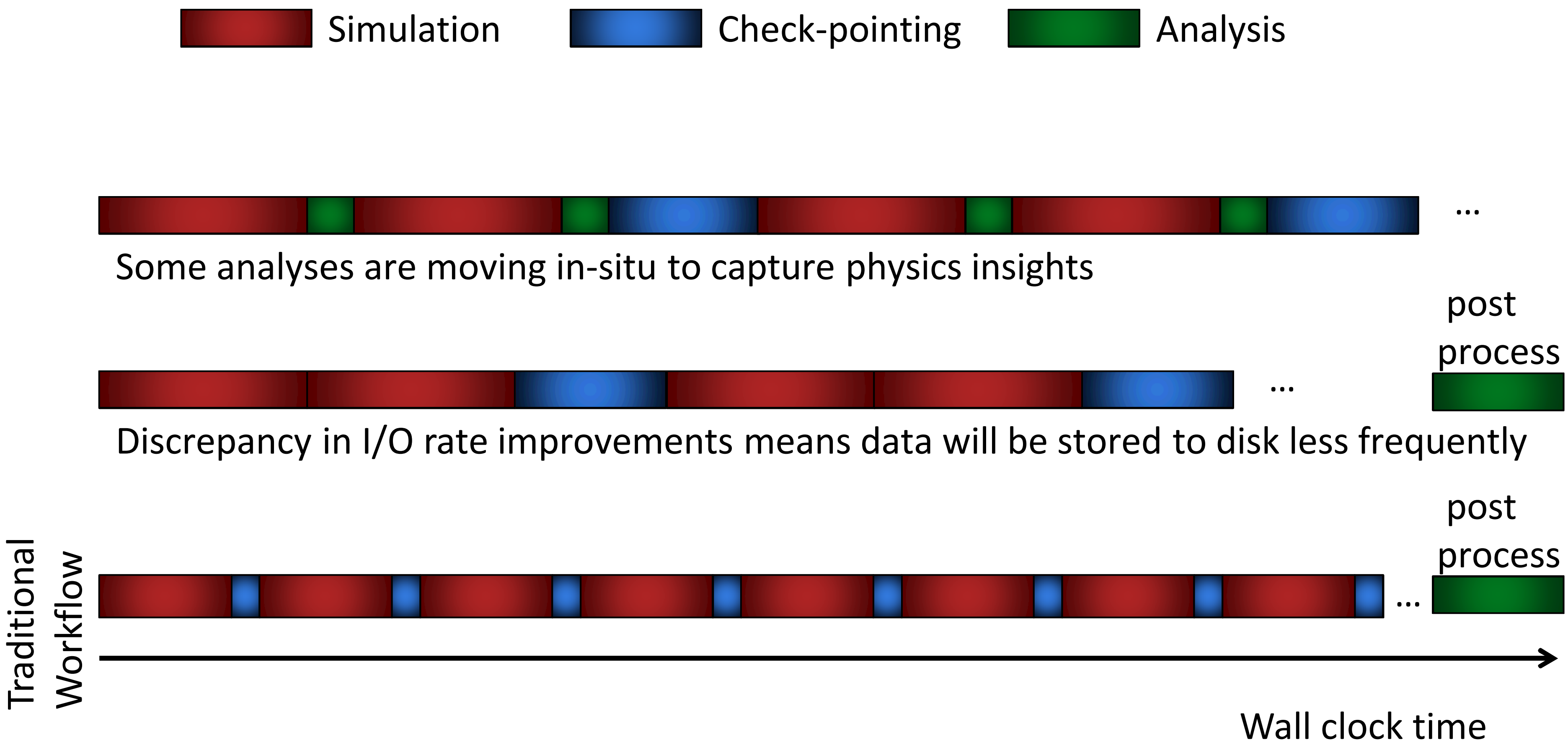
Data challenges are causing workflows to change



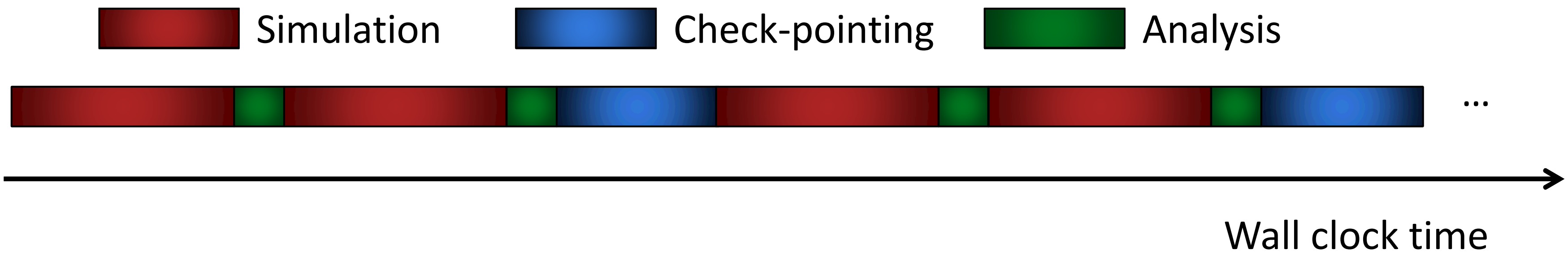
Data challenges are causing workflows to change



Data challenges are causing workflows to change



Workflow change introduces research challenges



- At what frequency should I/O or analysis be done?
  - Can we make this decision in an adaptive, data-driven fashion at runtime?
    - Avoid missing interesting science
    - Avoid costly I/O when simulation state is evolving slowly
  - How can we make these decisions quickly and efficiently?
- How do we change underlying analysis algorithms to be performant in situ?
- What programming models should we use to attain maximum performance, scalability, and resilience?

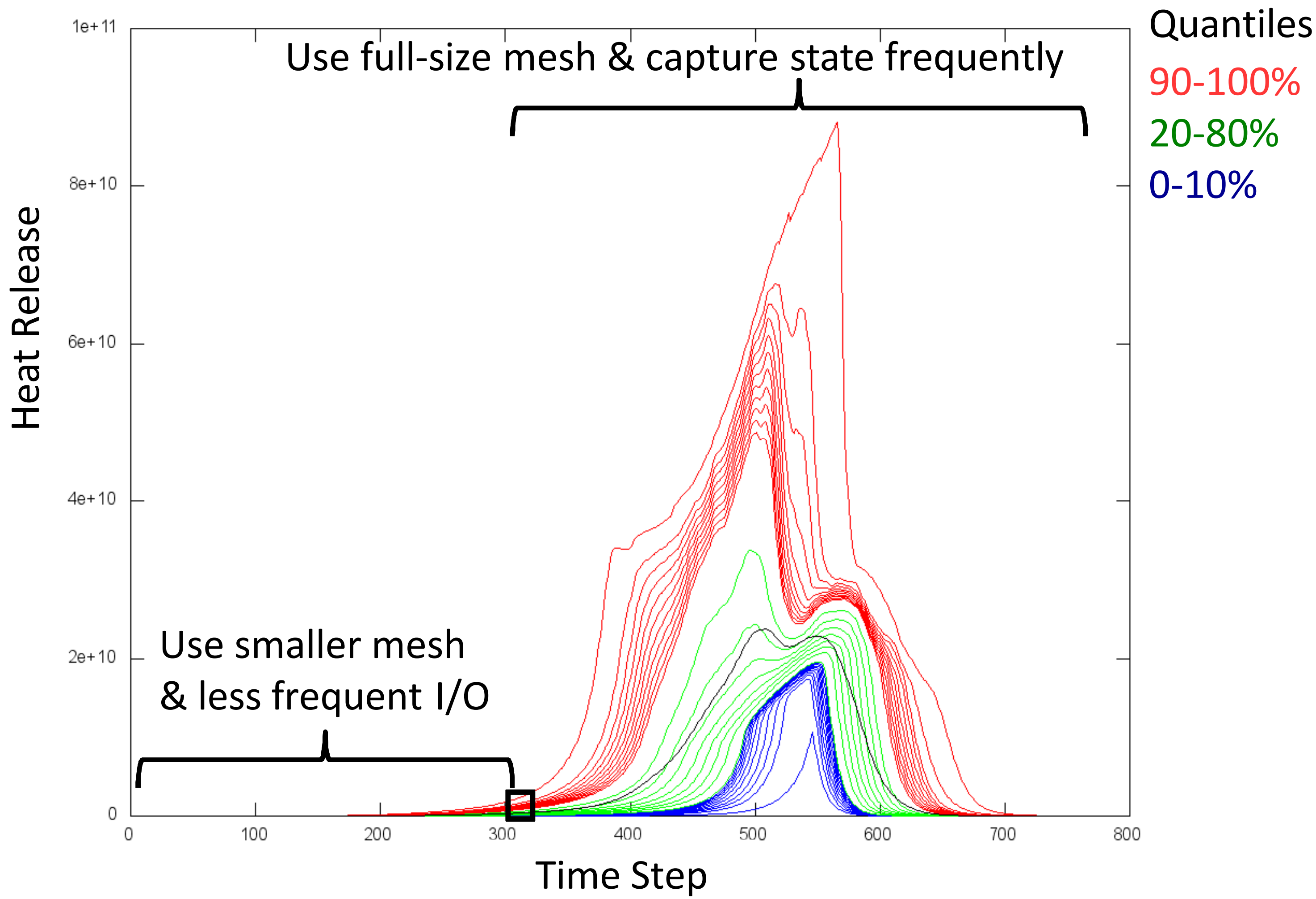
Sublinear analysis research to enable efficient, data-driven decisions at scale

Sublinear analysis is new theoretical subfield asking: how to determine properties of input by seeing tiny fraction

sublinear algorithms	in situ analysis challenges
• Small samples of data • Quantifiable time-error tradeoffs • Limited primitives for access	• Too much data to move • Constrained time budgets • Simulation dictates data structures

There is strong alignment between theory and challenges

Current research: Optimize mesh resolution and I/O frequencies in a data-driven manner



Fundamental algorithmic research can be required when moving analysis in situ

- Computation/communication profiles different than that of simulation
- Simulation dictates data structures/layout
- Strict time constraints

- To learn more:
  - Talk on Thursday 4:30-5:00
  - Room 391-392
  - Speaker: Aaditya Landge

In-Situ Feature Extraction of Large Scale Combustion Simulations Using Segmented Merge Trees

Aaditya G. Landge*, Valerio Pascucci*, Attila Gyulassy*, Janine C. Bennett‡, Hemanth Kolla‡, Jacqueline Chen‡, and Peer-Timo Bremer*†

*SCI Institute, University of Utah, Salt Lake City, UT
†Lawrence Livermore National Laboratory, Livermore, CA
‡Sandia National Laboratory, Livermore, CA

***Abstract*—The ever increasing amount of data generated by scientific simulations coupled with system I/O constraints are fueling a need for in-situ analysis techniques. Of particular interest are approaches that produce reduced data representations while maintaining the ability to redefine, extract, and study features in a post-process to obtain scientific insights.**

This paper presents two variants of in-situ feature extraction techniques using segmented merge trees, which encode a wide range of threshold based features. The first approach is a fast, low communication cost technique that generates an exact solution but has limited scalability. The second is a scalable, local approximation that nevertheless is guaranteed to correctly extract all features up to a predefined size. We demonstrate both variants using some of the largest combustion simulations available on leadership class supercomputers. Our approach allows state-of-the-art, feature-based analysis to be performed in-situ at significantly higher frequency than currently possible and with negligible impact on the overall simulation runtime.

Keywords*—*topological data analysis, feature extraction, in situ analysis, merge tree computation, segmented merge tree

I. Introduction

The continuing increase in available computing power allows scientists to simulate ever more complex phenomena at higher temporal and spatial resolutions. Correspondingly, the analysis of these datasets is becoming increasingly sophisticated, moving from global to local statistics and more recently to detailed studies of small, intermittent features of interest along with their characteristics and temporal evolution [1]–[3]. However, while the need for advanced data analysis techniques increases, the (relative) amount of data that can be permanently stored keeps decreasing. This can severely impede and may ultimately prevent an accurate and reliable analysis. State-of-the-art simulations are already reaching the point at which snapshots are stored too infrequently to accurately track fast moving or intermittent events, increasing the likelihood that potentially important phenomena are lost between snapshots.

While there exist a number of mitigating strategies such as compression [4] or advanced data management techniques [5], [6], the challenges discussed above will likely only be addressed by moving the analysis in-situ i.e., to perform it concurrently with the simulation. Since analysis results are typically orders of magnitude smaller than the original data, efficient in-situ algorithms would allow an effective analysis at much higher frequencies than otherwise feasible. To this end a number of in-situ visualization and analysis techniques have been proposed [7]–[11] either as stand alone tools or as part of existing systems. However, so far these efforts have been restricted to comparatively simple and largely data parallel operations and few solutions for more complex algorithms exist [12]. Furthermore, most of these analyses were designed in the context of a post-processing workflow, in which scientists test hypotheses by interactively adjusting input parameters to analysis algorithms that provide a single answer to a given question, to slowly converge to their results. In an in-situ setting, however, all parameters, spatial sub-domains, temporal windows, etc., must be specified *a priori*, making current algorithms ineffective at best and misleading at worst. Instead, a new kind of meta-analysis is required that can efficiently compute and encode a range of answers for an entire class of questions, effectively re-enabling a flexible and unbiased exploration of the results in post-processing.

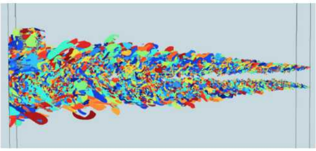


Fig. 1. Extinction regions in a lifted ethylene jet flame extracted using segmented merge trees and adaptive relevance thresholds.

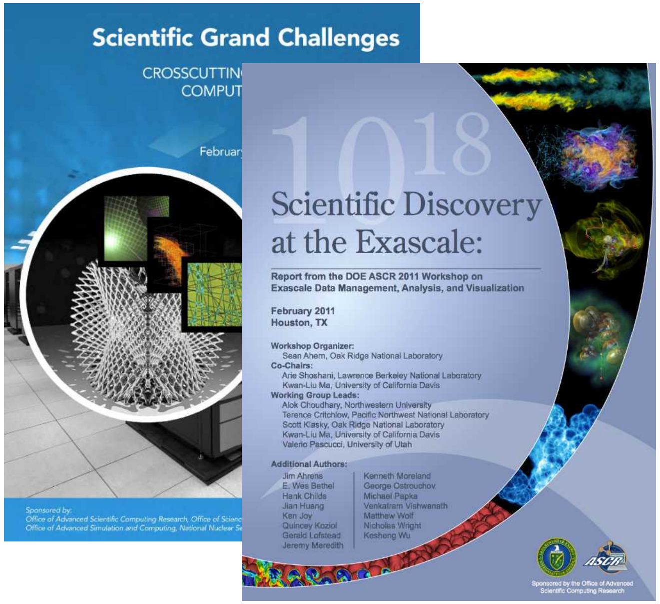
One promising class of techniques are topology-based segmentations based on merge trees [13], contour trees [14], or Morse-Smale complexes [15]. These techniques segment the domain into features according to either the level-set (e.g. thresholding) or gradient behavior of one of the simulation variables. In particular, segmentations of the domain derived from merge trees have been shown to efficiently encode threshold-based features. For example, as shown in Fig. 1, segmented merge trees can be used to extract extinction regions defined as areas of high scalar dissipation in turbulent

SC14, November 16-21, 2014, New Orleans, LA, USA

Programming models research aimed at portability, performance, scalability and resilience

Sandia National Laboratories

System Parameter	2011	2018		Factor Change
System Peak	2 Pf/s	1 Ef/s		500
Power	6 MW	≤20 MW		3
System Memory	0.3 PB	32-64 PB		100-200
Total Concurrency	225K	1 BX10	1B X100	40000-400000
Node Performance	125 GF	1 TF	10 TF	8-80
Node Concurrency	12	1000	10000	83-830
Network Bandwidth	1.5 GB/s	100 GB/s	1000 GB/s	66-660
System Size (nodes)	18700	1000000	100000	50-500
I/O Capacity	15 PB	30-100 PB		20-67
I/O Bandwidth	0.2 TB/s	20-60 TB/s		10-30



Shifts in programming models

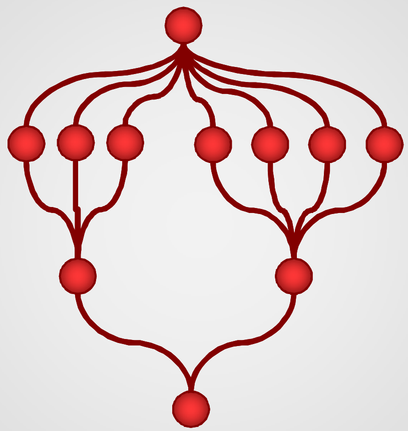
MPI+X: Cuda, OpenCL, Cilk+, OpenMP, Kokkos, …

Asynchronous Many-Task (AMT): Charm++, Uintah, Legion, Scioto, Dague, CnC, Dharma…

Research in asynchronous many-task (AMT) programming models at Sandia

- AMT programming models
  - + Data-flow model
  - + Show promise at sustaining performance
  - + Work stealing enables load balancing
  - + Failed tasks can be re-executed
- DHARMA project at Sandia (ASC)
  - Distributed asyncHronous Adaptive Resilient Management of Applications
- A Unified Data-Driven Approach for Programming In Situ Analysis and Visualization (ASCR)
  - Joint with LANL, Stanford, U. Utah, Kitware

Task Graph

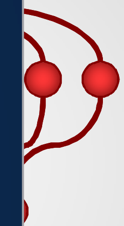


Nodes are tasks
Edges are data inputs/outputs

Research in asynchronous many-task (AMT) programming models at Sandia

- AN
  - +
  - +
  - +
  - +
- DH
  - ▪
- A
  Pr
  Vi
  - ▪

n



outputs

- BOF: Asynchronous Many-Task Programming Models for Next Generation Platforms
- Tuesday 12:15-1:15, Room 396
- Panel Members: Charm++, DHARMA, HPX, Legion, OCR, STAPL, Uintah

Questions?