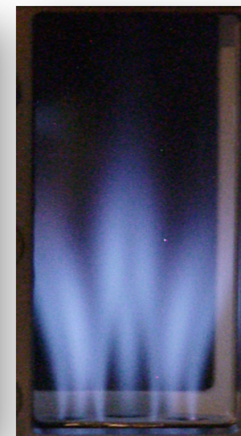
Challenges in the analysis of extreme-scale combustion simulation data

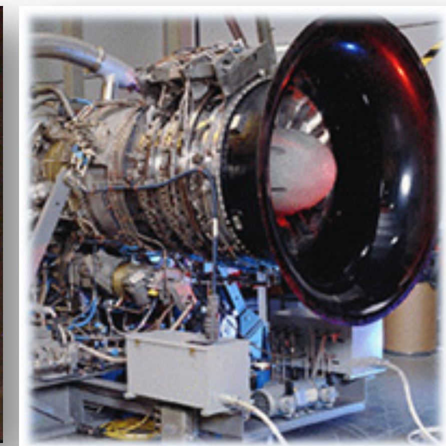
J. Bennett, J. Chen, H. Kolla, H. Yu @ Sandia National Laboratories
A. Gyulassy, V. Pascucci @ University of Utah
P.-T. Bremer@ Lawrence Livermore National Laboratory
R. Grout @ National Renewable Energy Laboratory

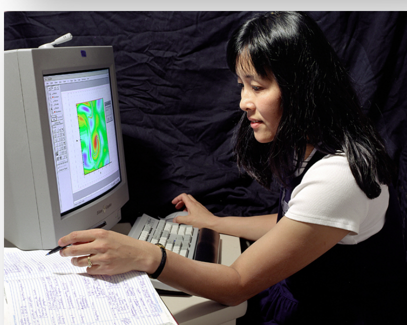
CScADS Workshop on Scientific Data & Analytics for Extreme Scale Computing
July 30, 2012

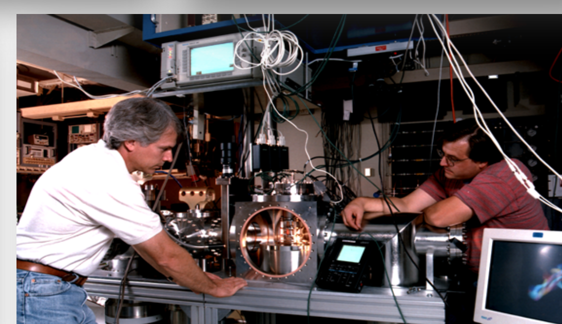
Combustion Research Facility at Sandia National Laboratories

A DOE user facility dedicated to energy science and technology for the twenty-first century









Sandia National Laboratories

Combustion plays an important role in energy security

- Combustion accounts for a majority of energy used in the United States

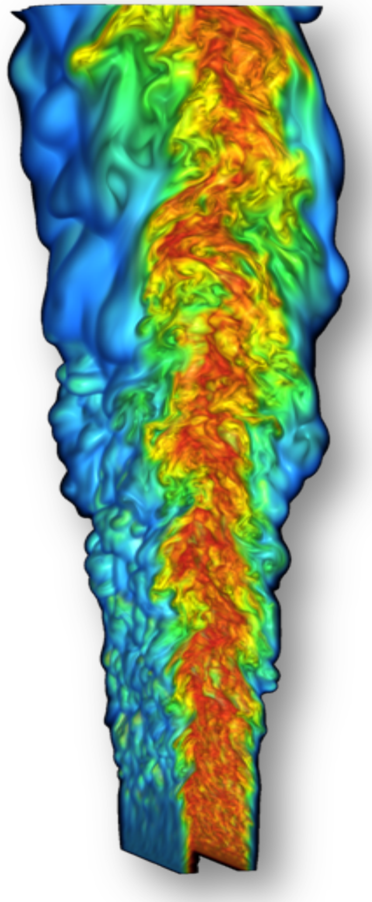
- Computer simulations provide tools for design of efficient clean burning devices

- Sound scientific understanding is necessary to develop predictive, validated multi-scale models

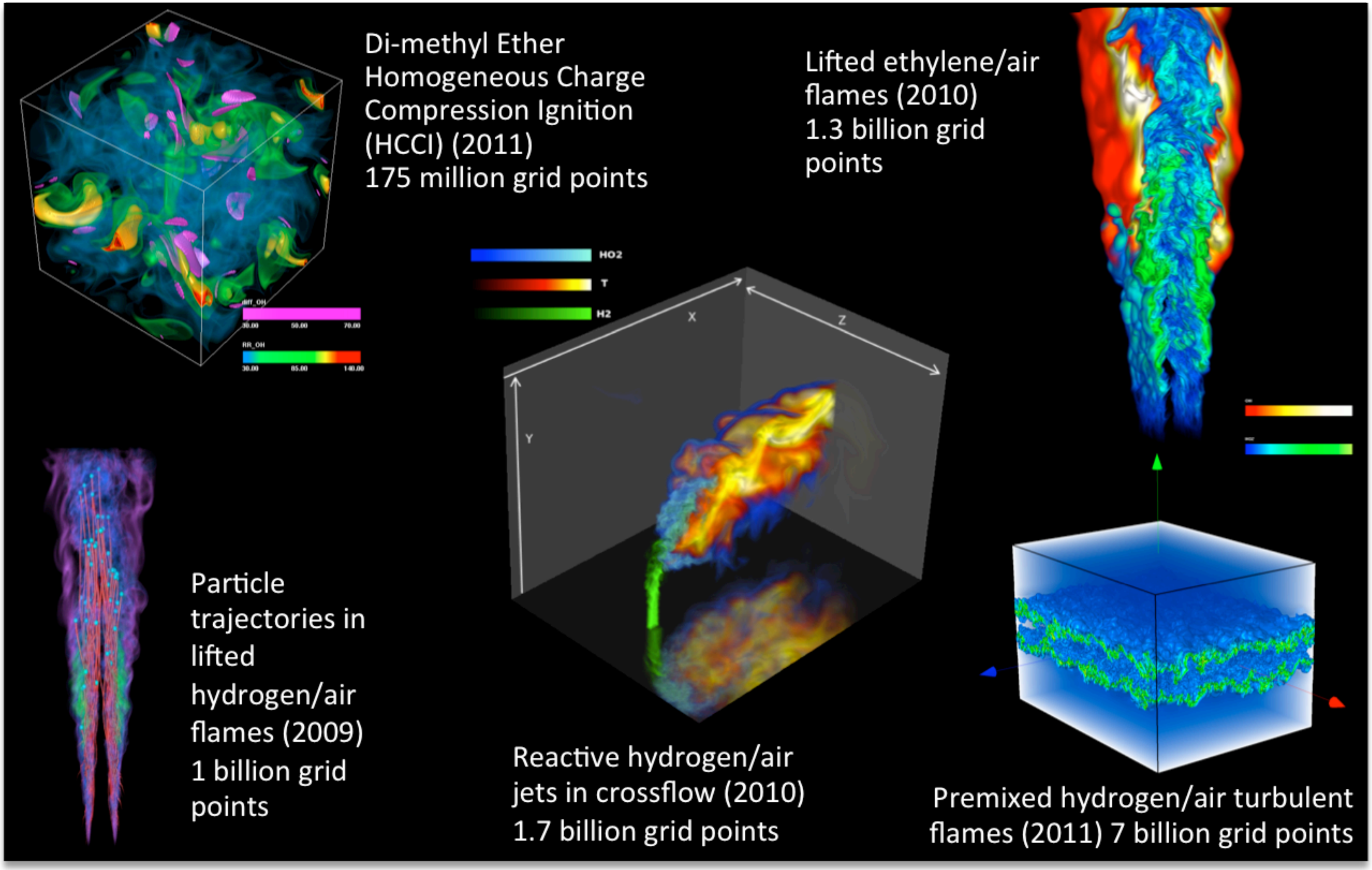
Sandia National Laboratories

Direct Numerical Simulations (DNS) are used to study fundamental turbulence-chemistry interactions

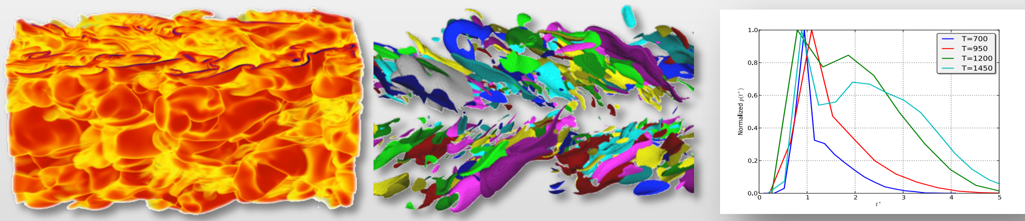
- Turbulence
  - Entrains, advects, strains and wrinkles a flame creating more area for burning
  - Causes molecular mixing of reactants
- Chemical reactions
  - Are enhanced with mixing to a limit → extinction
  - Create heat release
- Heat release, dilatation
  - Reduce turbulence intensity through density, and property changes

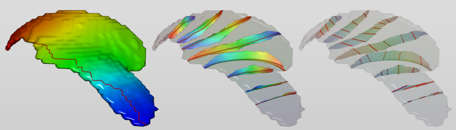


Simulation benchmark data generated by S3D is used for model development and validation

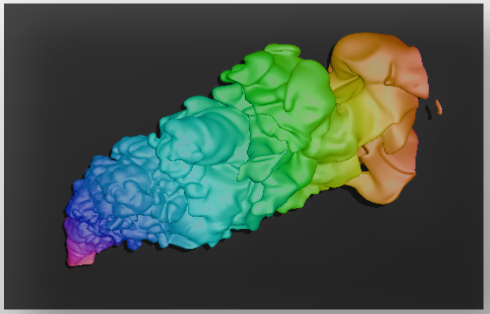


Combustion scientists are interested in analyzing this data in a variety of ways

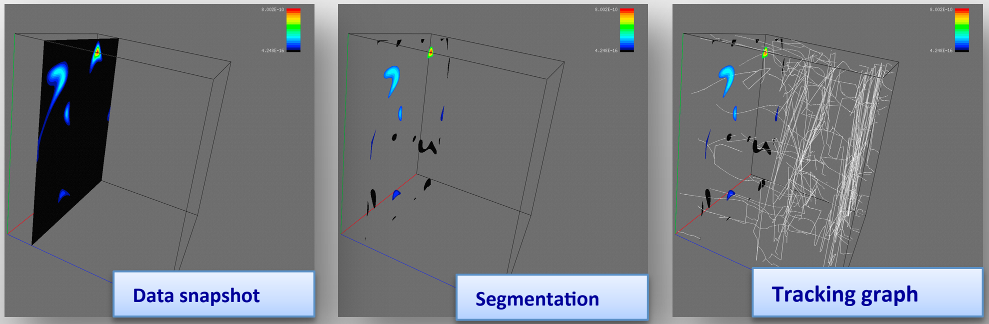




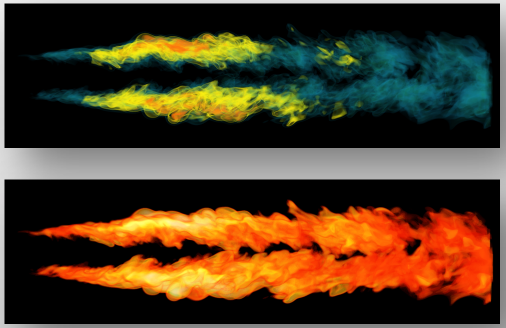
We want to identify features, characterize their shapes and analyze the behavior of other variables within these features



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



Tracking features in time

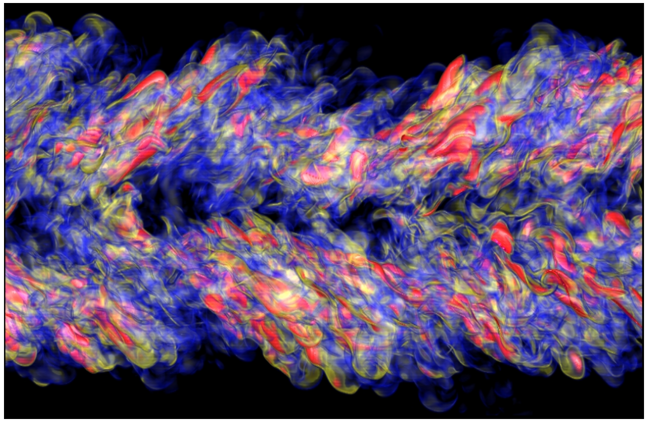


In-situ visualization allows for monitoring and provides qualitative analysis results

Data analysis is complicated by several factors

- Data size
  - Billions of grid points per time step
  - Hundreds of time steps written to disk
- Data complexity
  - Multivariate
  - Turbulence is a complex phenomenon
  - Length scales: microns to centimeters
  - Temporal scales: nanoseconds to milliseconds
- Example: Lifted Ethylene Jet
  - 1.3 bilion grid points
  - 22 chemical species, vector & particle data
  - 7.5 million cpu hours on 30,000 processors
  - 112,500 timesteps (data stored every 375th)
  - 240 TB of raw field data + 50 TB particle data

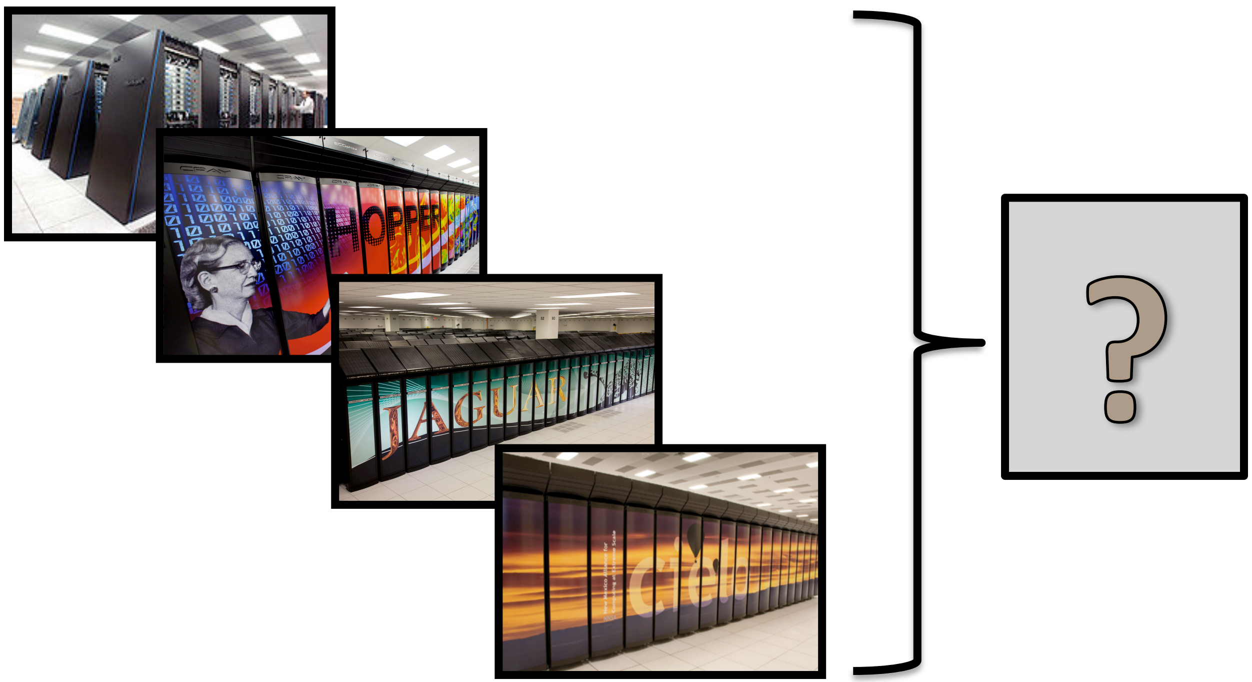
HPSS storage facility at NERSC



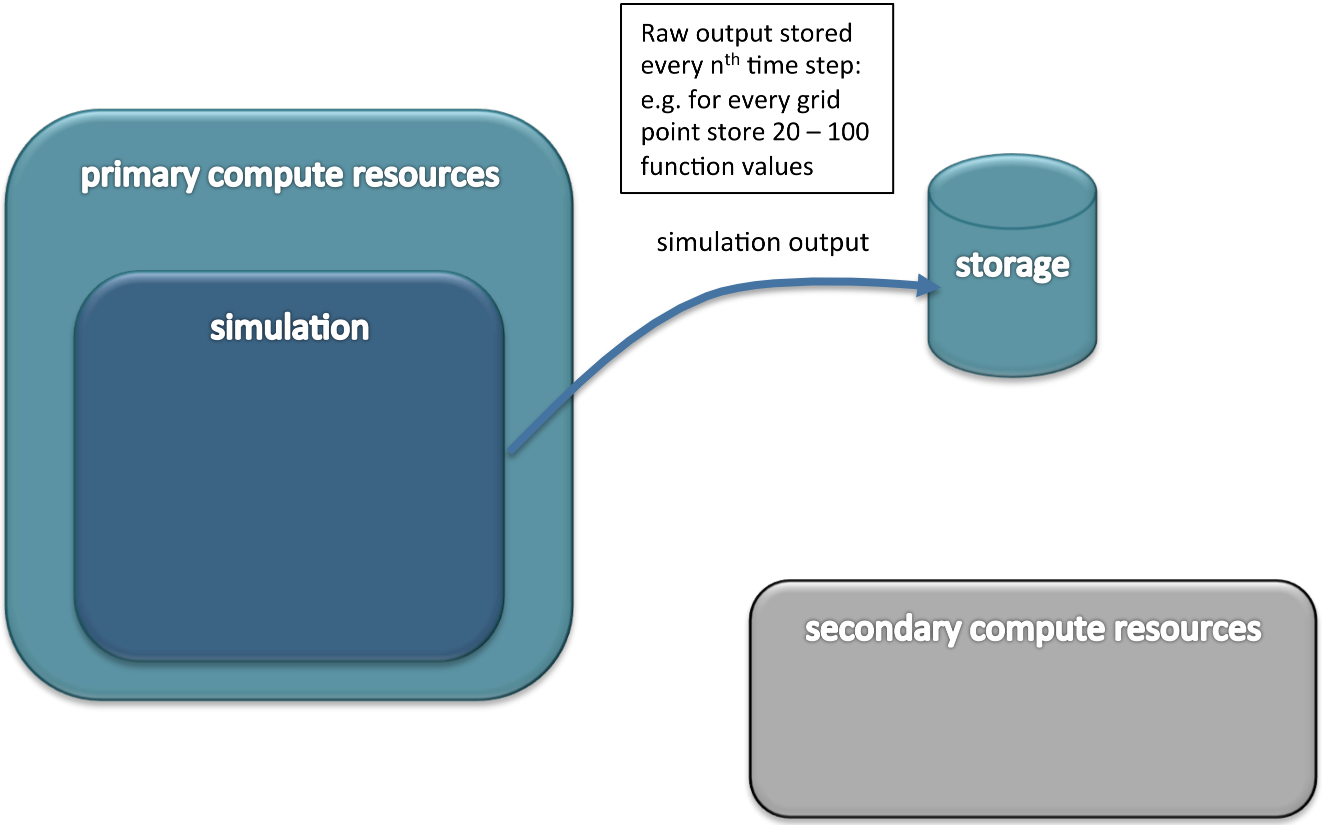
Sandia National Laboratories

The move to exascale computing adds additional complexity

ExaCT CENTER FOR EXASCALE SIMULATION OF COMBUSTION IN TURBULENCE

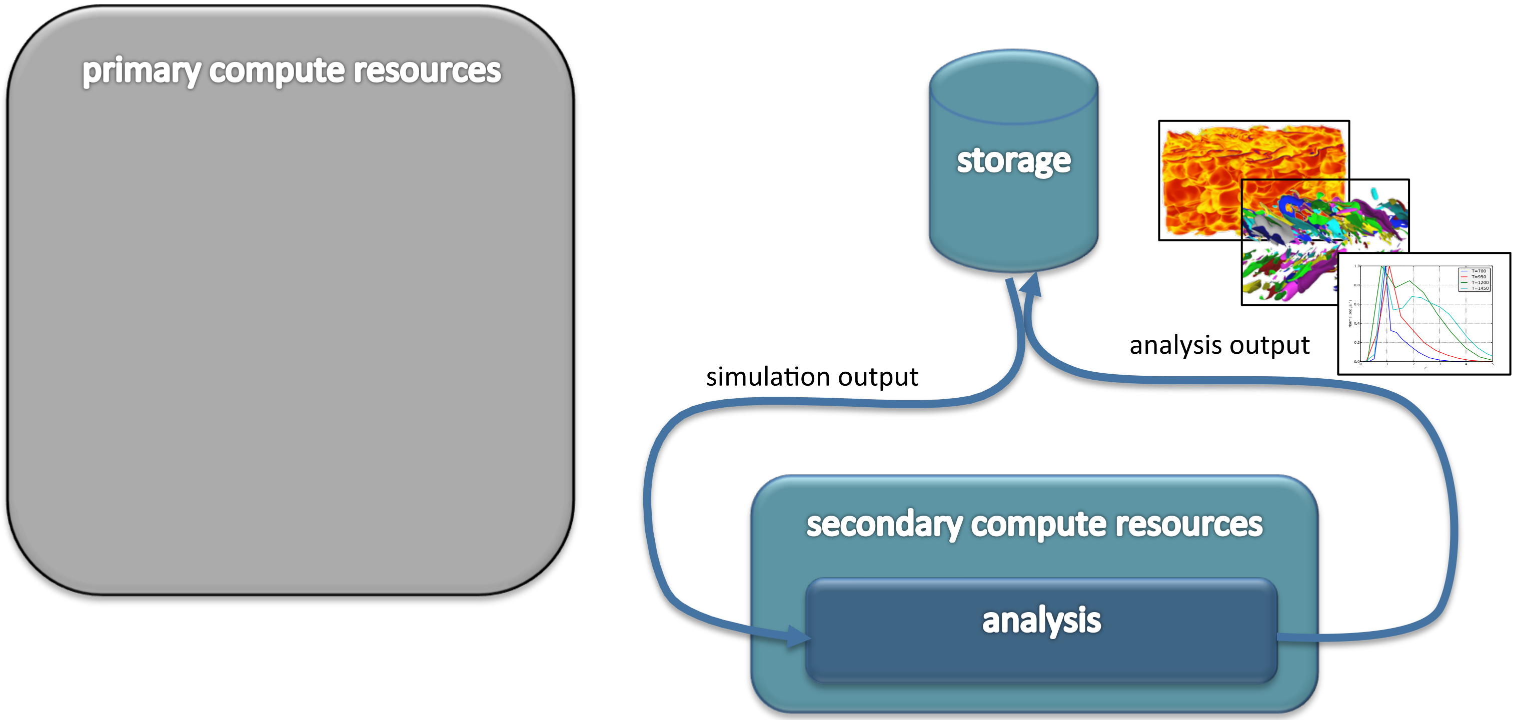


Existing data analysis paradigm comprises two stages
Stage 1: perform simulation

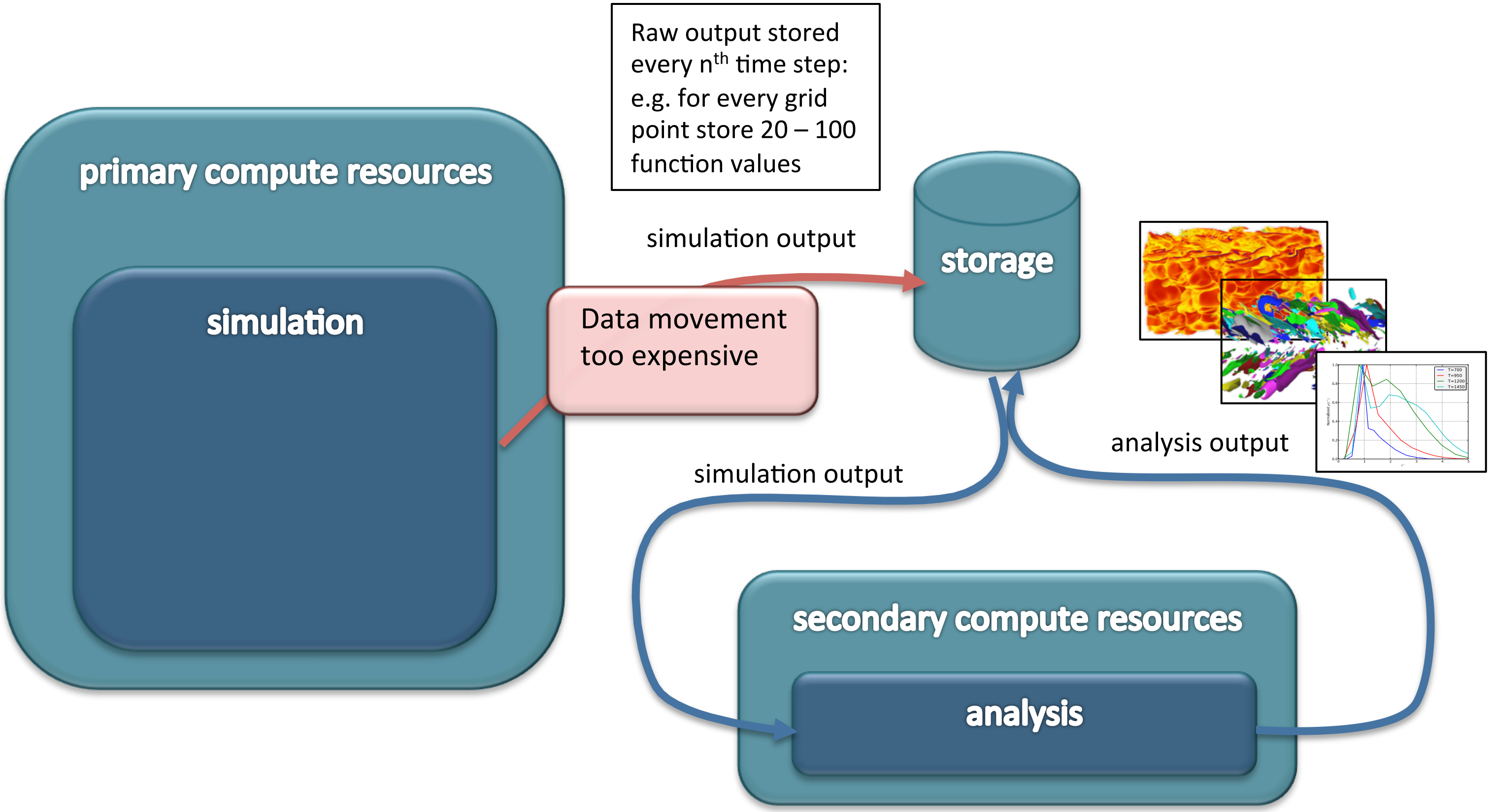


Sandia National Laboratories

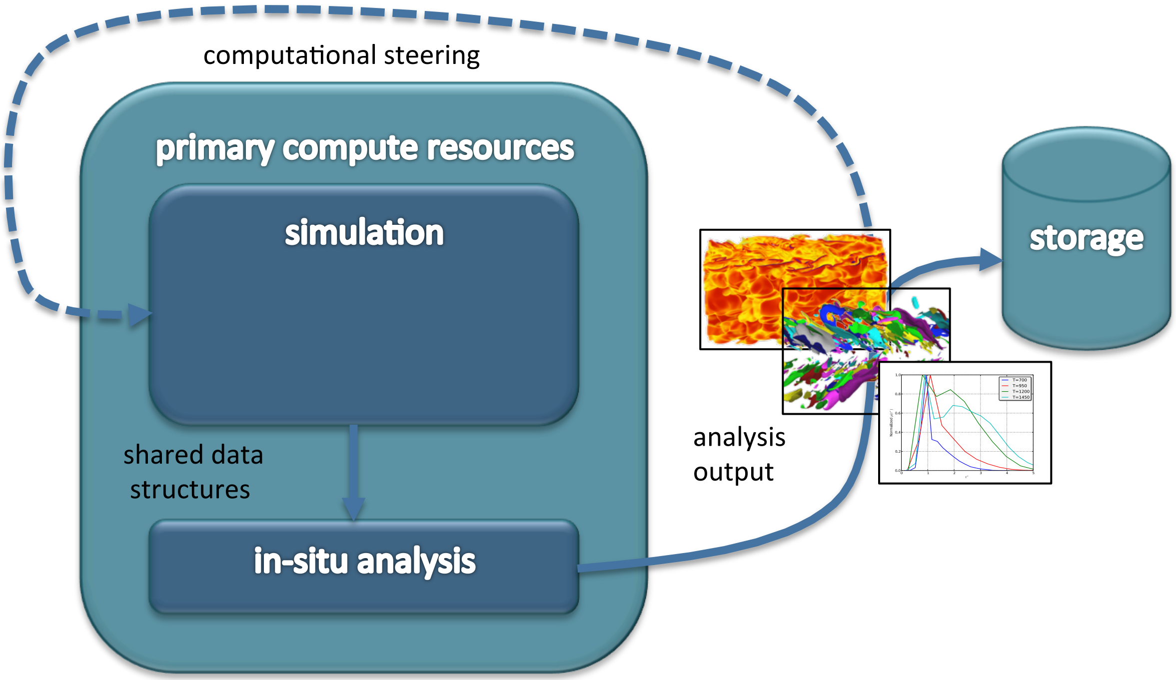
Stage 2: extraction of scientific insight is a post-process on secondary resources



This approach does not scale!

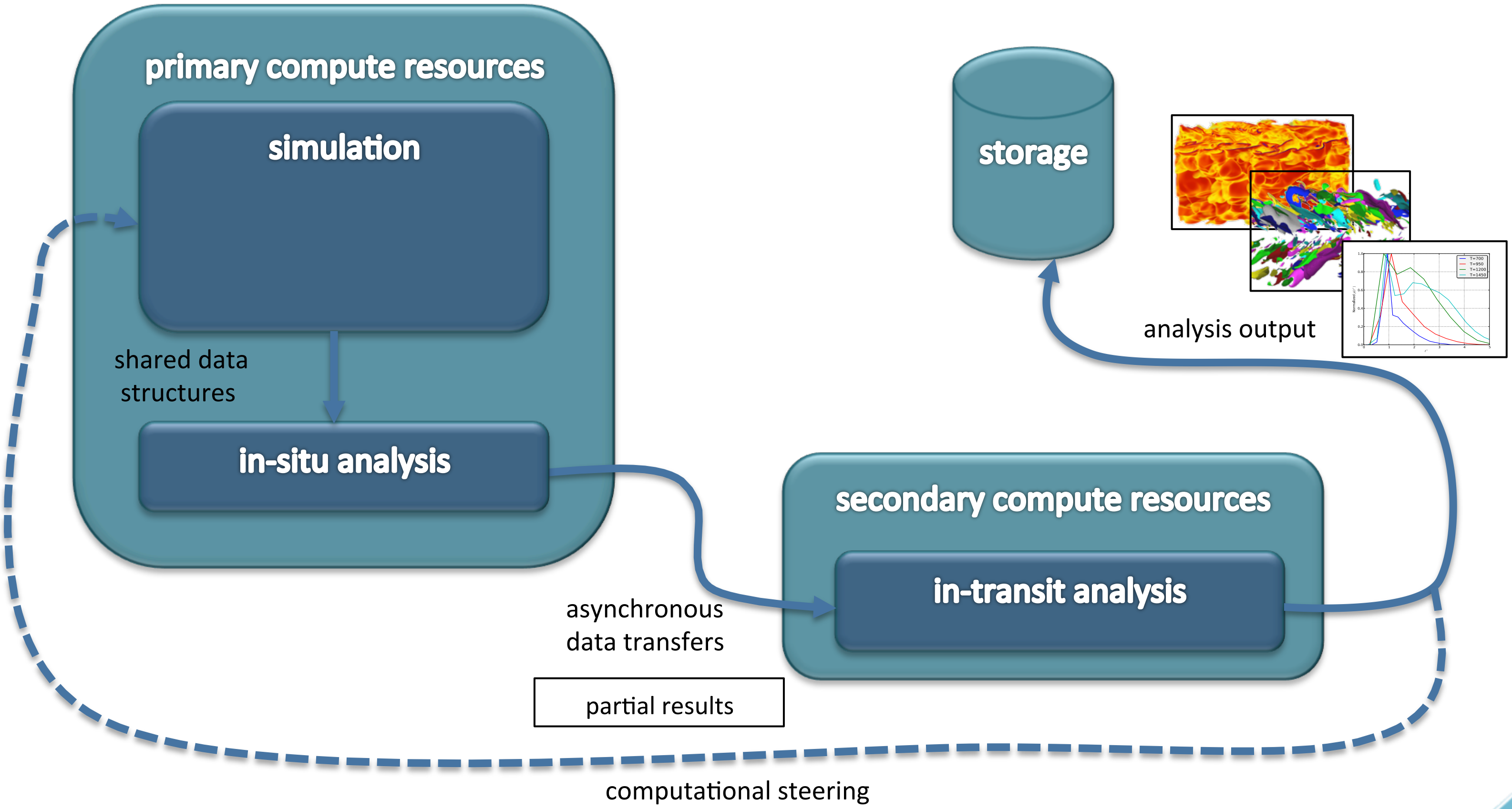


As a result we are seeing a shift towards a concurrent analysis paradigm



Sandia National Laboratories

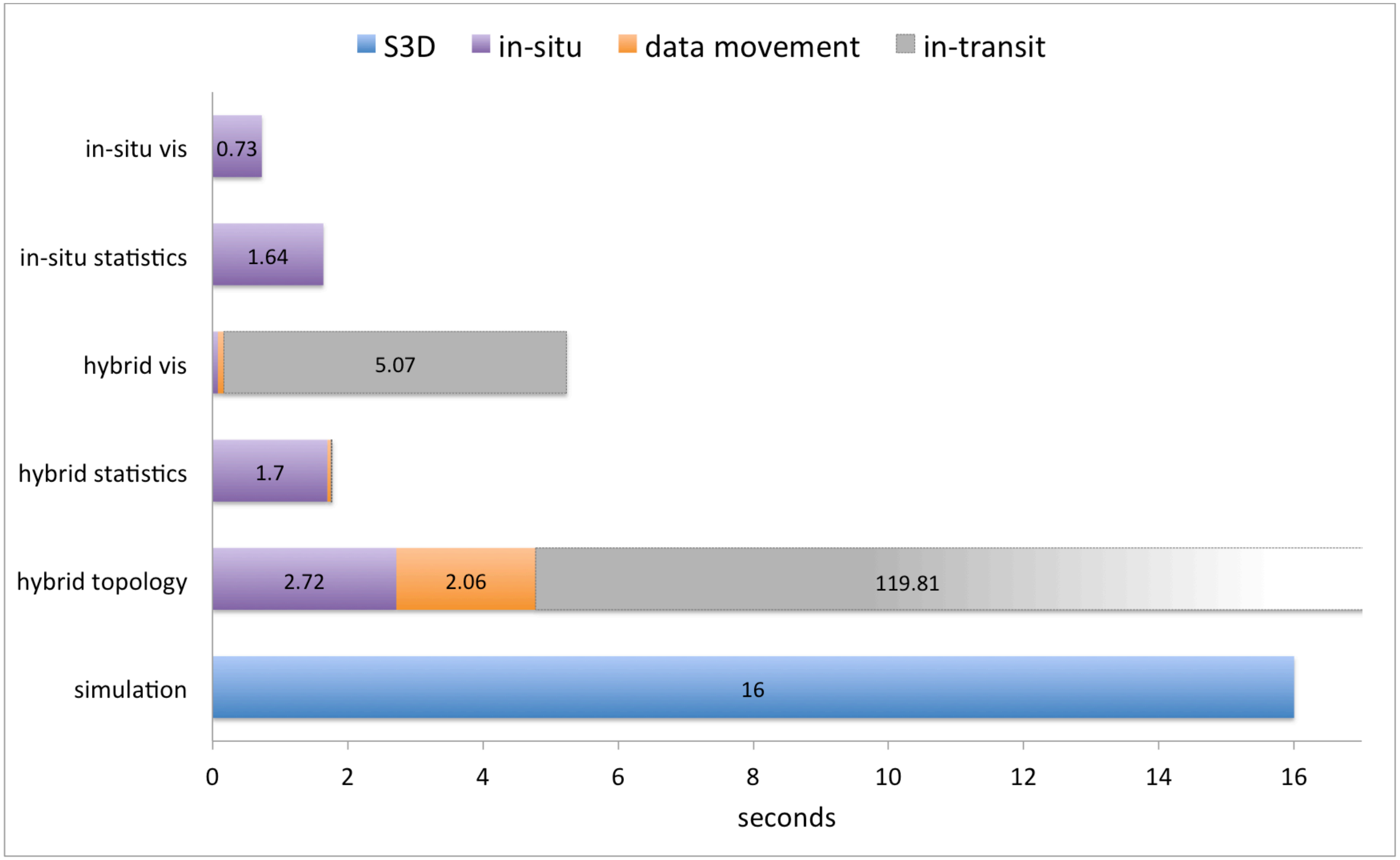
Secondary compute resources can be used to perform in-transit analysis



Sandia National Laboratories

Hybrid in-situ + in-transit framework shows promise

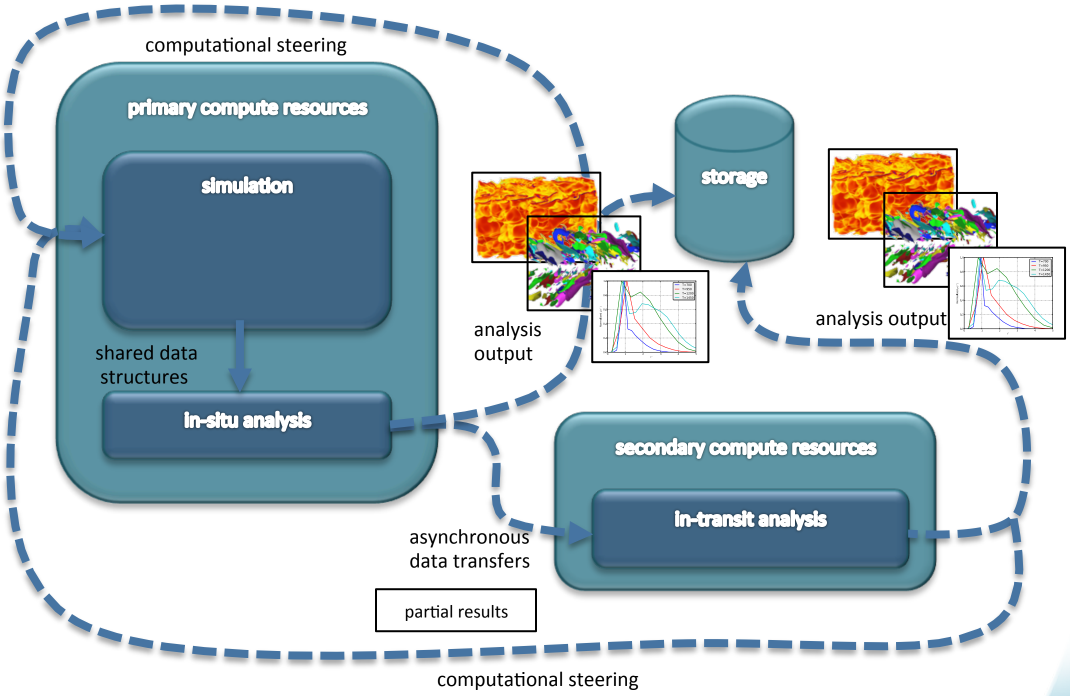
timing breakdown among simulation and analytics using 4896 cores



Combining In-situ and In-transit Processing to Enable Extreme-Scale Scientific Analysis
(J. Bennett, H. Abbasi, P-T Bremer, R. Grout, A. Gyulassy, T. Jin, S. Klasky, H. Kolla, M. Parashar, V. Pascucci, P. Pebay, D. Thompson, H. Yu, F. Zhang, and J. Chen, to appear in SC 2012)

Many research and development challenges remain

- In-situ/in-transit decomposition
- Shared data structures
- Strict time constraints
- Scheduling
- Input parameters
- Minimize communication
- Efficient data movement
- Data reduction
- Resilient analyses



Sandia National Laboratories

Questions?

Contact:

Janine Bennett

Sandia National Laboratories

jcbenne@sandia.gov

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.