



# **Present and Future Computing Requirement for Combustion of Alternative Fuels for Transportation and Power – Fundamental Investigation using Direct Numerical Simulations**

Jacqueline H. Chen

Combustion Research Facility

Sandia National Laboratories

[jhchen@sandia.gov](mailto:jhchen@sandia.gov)

NERSC BES Requirements for 2017

October 8-9, 2013

Gaithersburg, MD



## Team Members (MP241)

Jacqueline Chen (Sandia)

Hemanth Kolla (Sandia)

Sgouria Lyra (Sandia)

Ankit Bhagatwala (Sandia)

John Bell (LBNL)

Marc Day (LBNL)

Damir Valiev (Princeton U.)

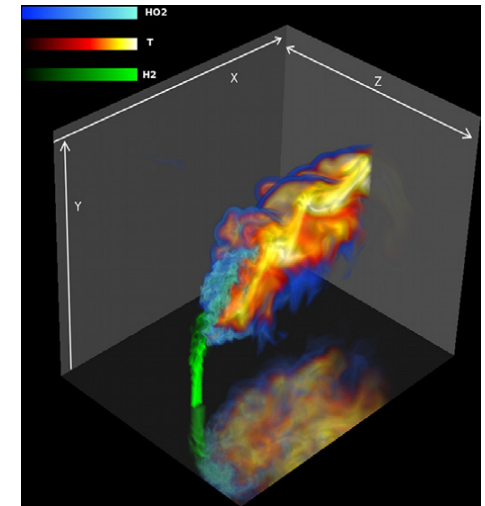
Andrea Gruber (SINTEF)

Ray Grout (NREL)

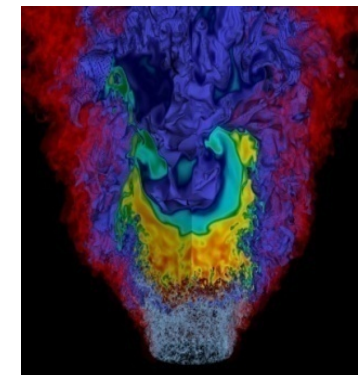


# Why Combustion

- **83% of U.S. energy comes from combustion of fossil fuels**
- **National goals**
  - Reduce greenhouse gas emissions by 80% by 2050
  - Reduce petroleum usage by 25% by 2020
- **New generation of high-efficiency, low emission combustion systems**
  - New designs for IC engines such as HCCI
  - Fuel flexible turbines for power generation
- **Rapidly evolving fuel streams**
  - Biodiesel for transportation
  - Syngas from gasification processes
  - Alcohols
- **These factors significantly increase the design space for new combustion technologies**



Nitrogen-diluted hydrogen jet in crossflow

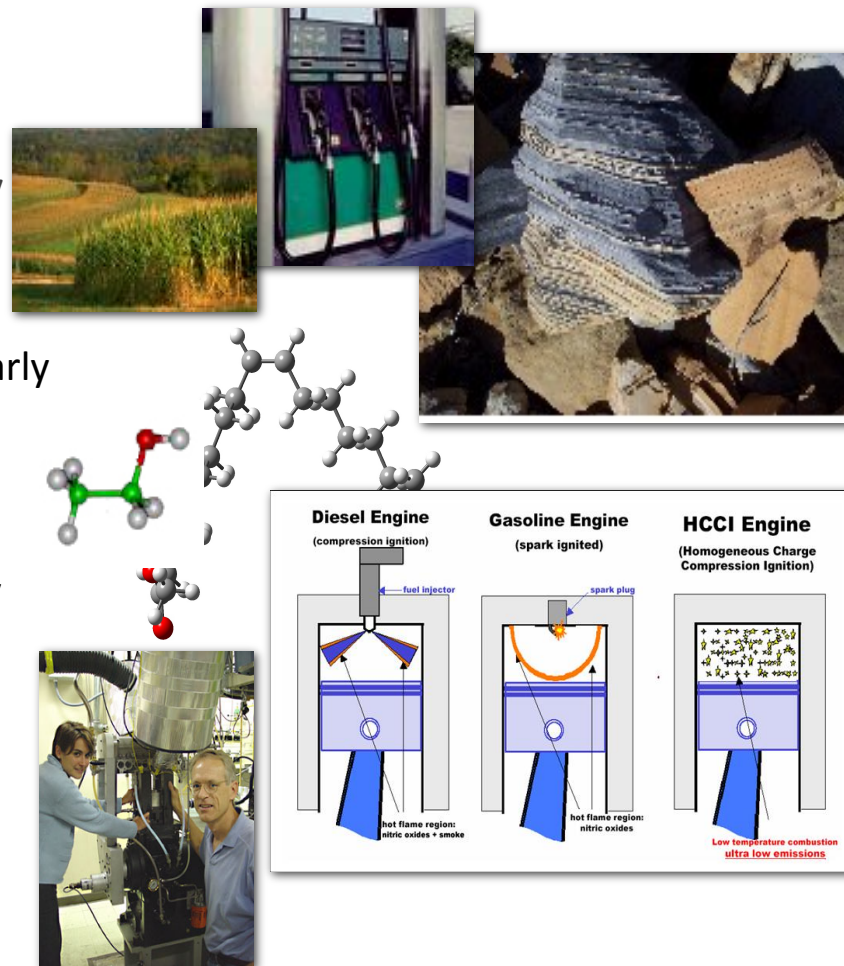


NOx emissions from a low swirl injector fuels by H2



# Why Exascale

- **Current design methodologies are largely phenomenological**
  - Combustion models based on simplified theory
  - Global-step simple chemical kinetics
- **Significant increase in computational capability will dramatically reduce design cycle for new combustion technologies and new fuels**
  - Simulations with higher physical fidelity, particularly chemistry at high pressure
  - More predictive science-based turbulence/chemistry interaction models
    - Address new mixed-mode regimes driven by efficiency and emissions
    - Differentiate effects of alternative fuels
- **Focus is on first principles direct numerical simulation methodologies**
  - Science base for novel fuels at realistic pressures
  - Not addressing complexity of geometry in engineering design codes

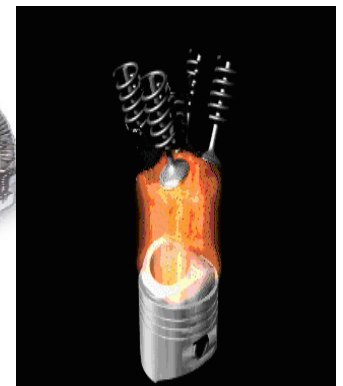
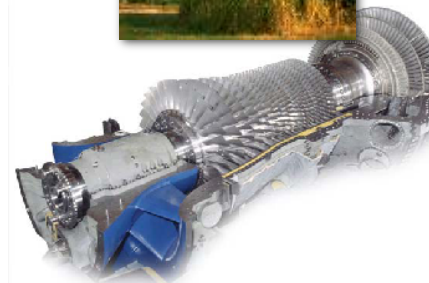
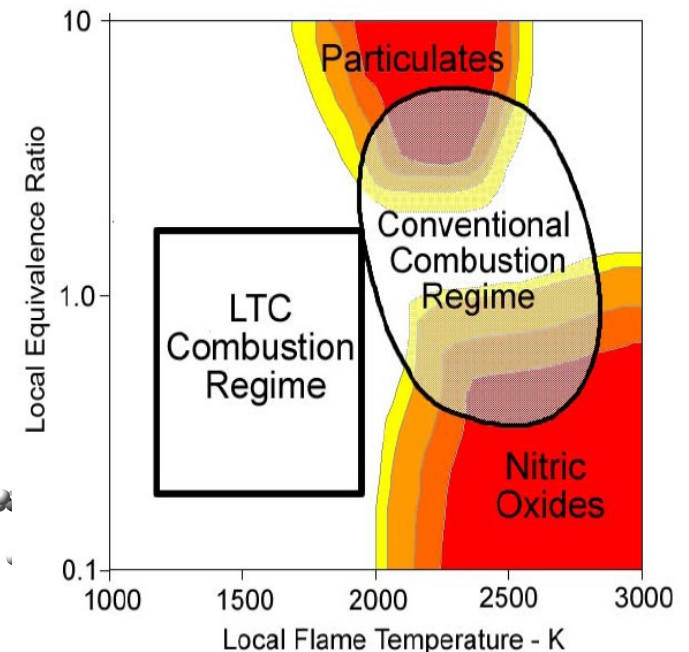
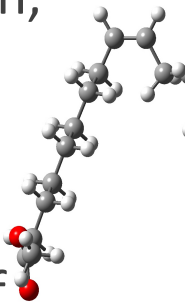






# Fundamental Turbulence-Chemistry Interactions Motivated by Advanced Engines and Gas Turbines

- Higher fuel efficiency and lower emissions driving combustion towards more **dilute, fuel lean, partially-premixed conditions**
- New **mixed-mode combustion regimes**: stratified premixed flames, autoignition, extinction exhibiting strong finite-rate kinetic effects
- Strong **sensitivities to ignition fuel chemistry** – differentiate the effects of fuels, e.g. oxygenated biofuels
- **Preferential diffusion effects** – synthesis gases enriched with hydrogen for carbon capture storage in gas turbines for power

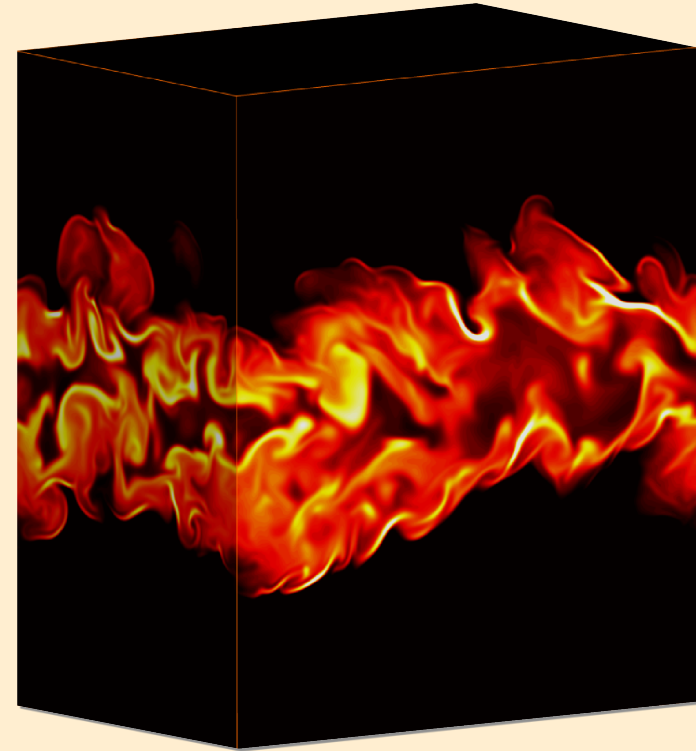




## Direct Numerical Simulation – S3D

Chen *et al.*, *Comp. Sci. Disc.*, 2009

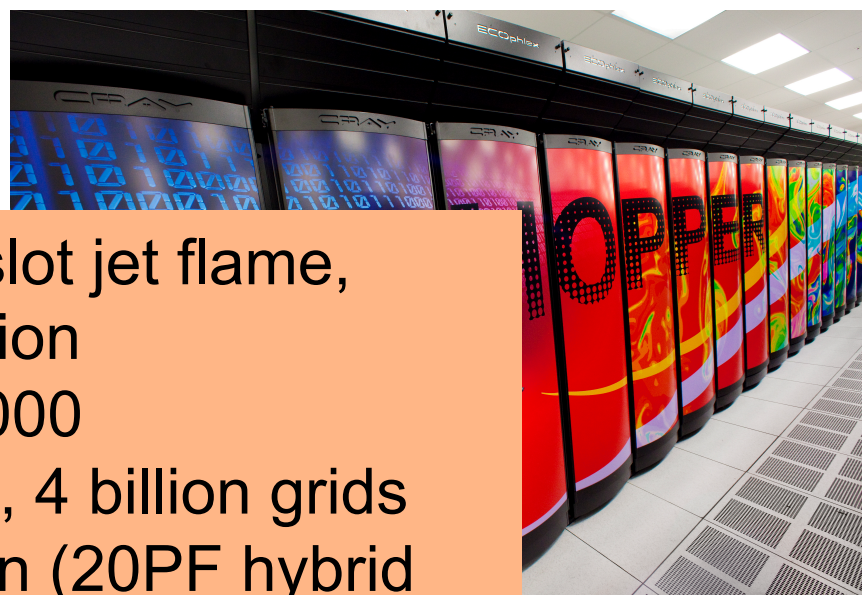
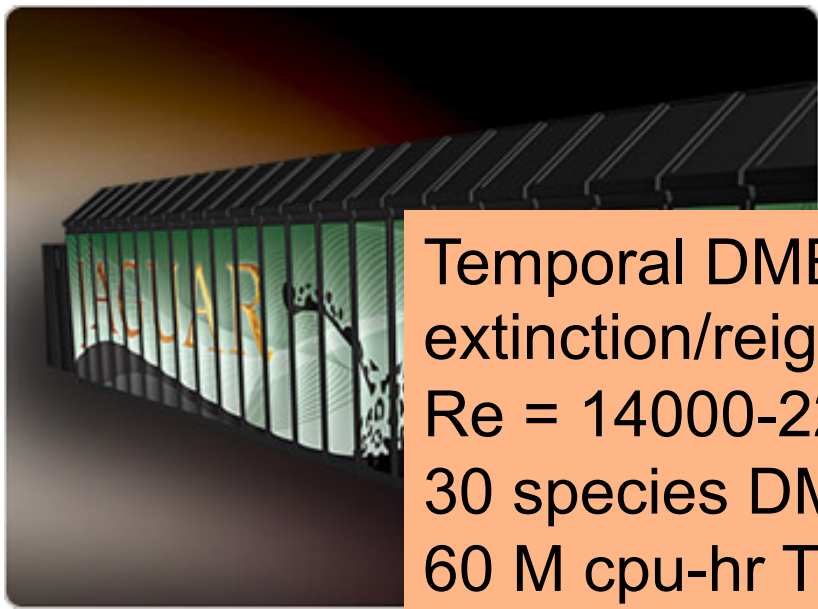
- Used to perform first-principles-based DNS of reacting flows
- Solves compressible reacting Navier-Stokes equations
- High-fidelity numerical methods
- Detailed reaction kinetics and molecular transport models
- Multi-physics: sprays, radiation and soot through SciDAC
- Ported to all major petascale platforms
- Particle and flame element tracking
- In situ analytics and visualization
- Refactored for multi-threaded, many core heterogeneous architectures



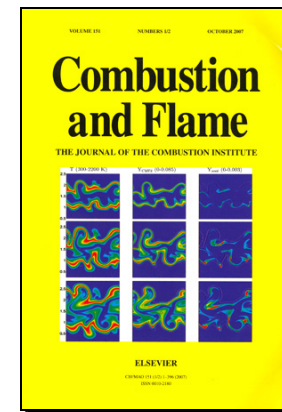
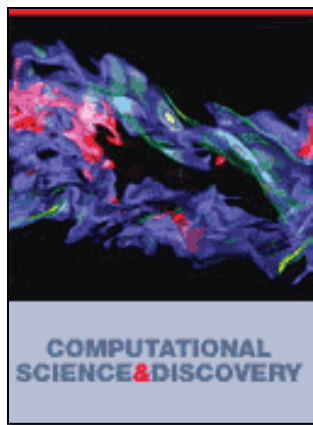
DNS provides unique fundamental insight into the chemistry-turbulence interaction



# Combustion DNS Enabled by Large Computer Allocations



Temporal DME slot jet flame,  
extinction/reignition  
 $Re = 14000-22,000$   
30 species DME, 4 billion grids  
60 M cpu-hr Titan (20PF hybrid  
CrayXK7)





## Roles of DNS

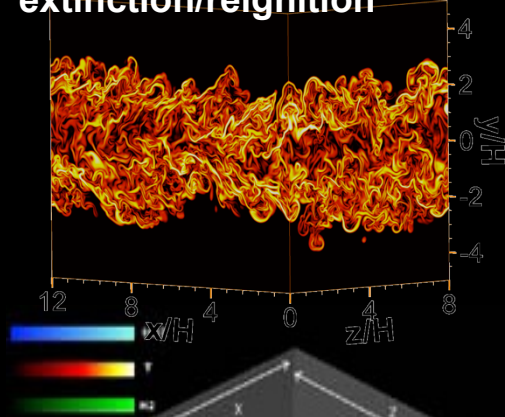
- Glean fundamental insight into ‘turbulence-chemistry’ interactions in canonical configurations – understanding causality between turbulence, mixing and reaction
- Evaluate RANS and LES subgrid models
- Evaluate experimental assumptions: 1D->3D,
- Identify empirical reduced manifolds for complex chemical processes
- Validate chemical mechanisms in the presence of unsteady transport



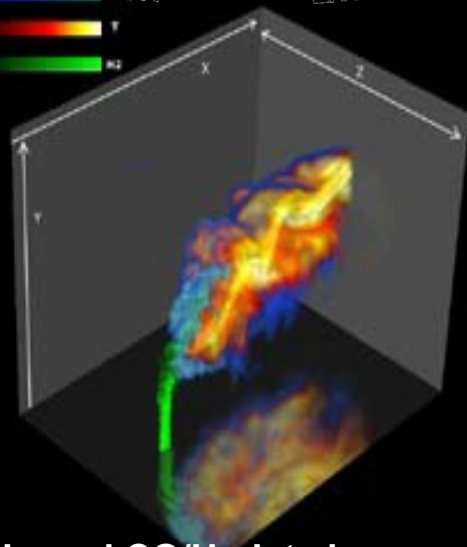


# DNS Benchmarks for Model Development

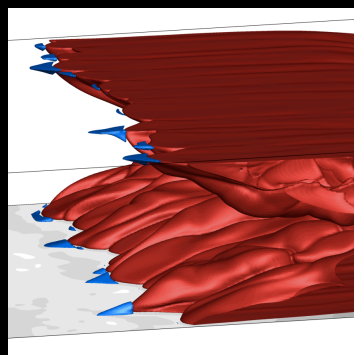
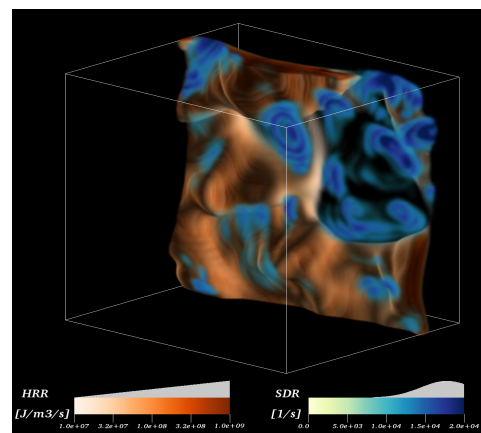
CO/H<sub>2</sub>, C<sub>2</sub>H<sub>4</sub> and DME jet flames extinction/reignition



H<sub>2</sub> and CO/H<sub>2</sub> Jets in Crossflow

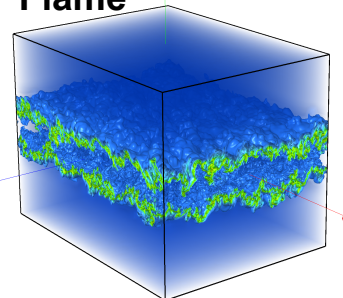


Turbulent Counterflow H<sub>2</sub>/air Stratified Flames

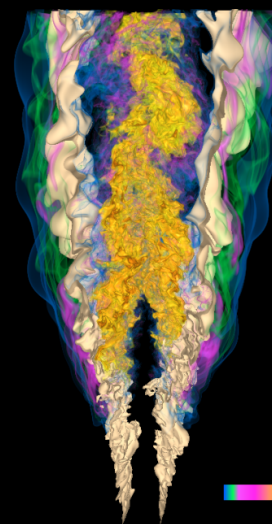


H<sub>2</sub>/air Flame-Boundary Layer Interaction

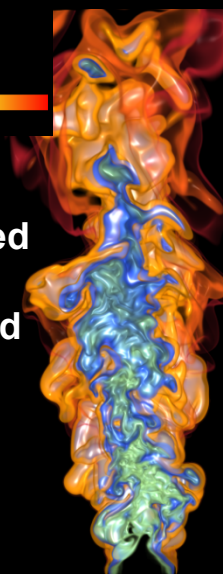
H<sub>2</sub> Premixed Jet Flame



Lifted H<sub>2</sub> and C<sub>2</sub>H<sub>4</sub> jet flames in hot coflow



Lean premixed and stratified CH<sub>4</sub>/air Bunsen flames

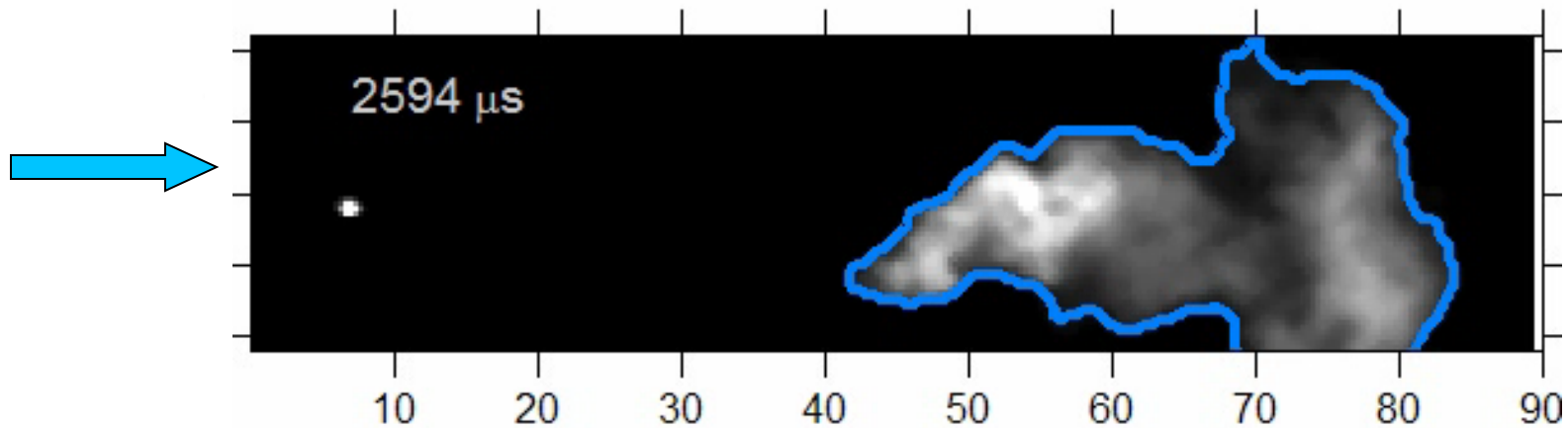




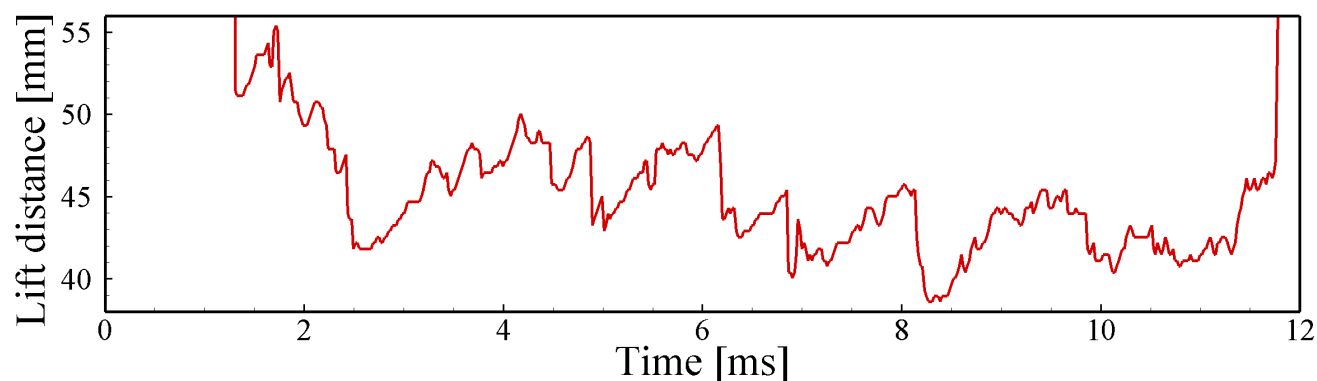


# Motivation: Understanding Stabilization of Lifted Flames in Heated Coflow

What is the role of ignition in lifted flame stabilization?



Chemiluminescence from diesel lift-off stabilization for #2 diesel, ambient 21%  $\text{O}_2$ , 850K, 35 bar courtesy of Lyle Pickett, SNL

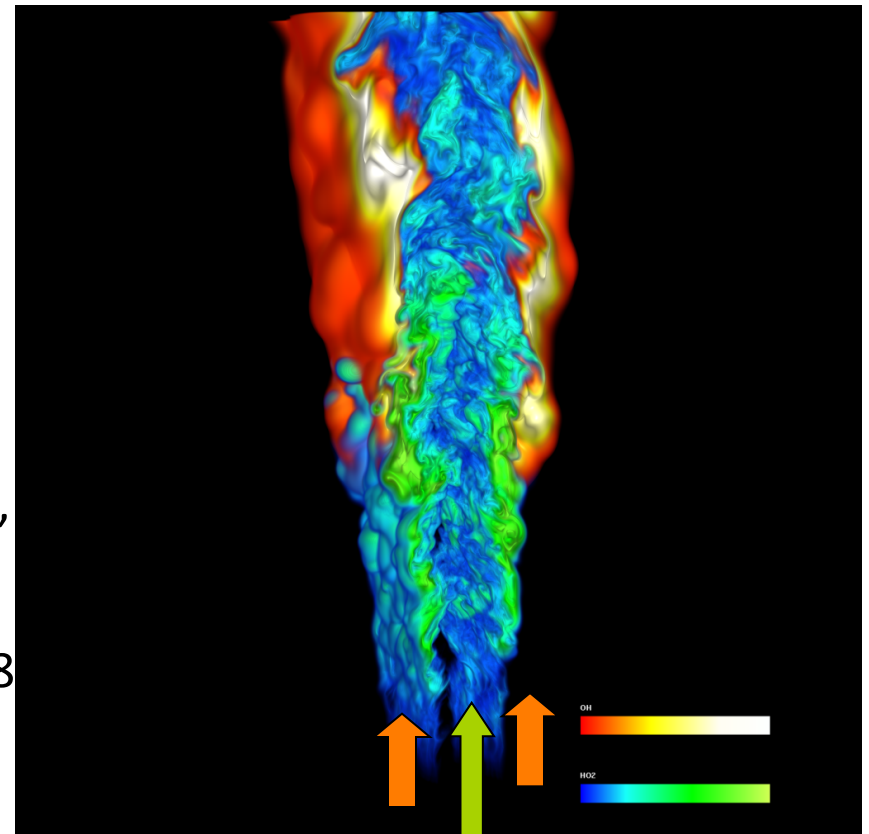




# DNS of Lifted Ethylene-air Flame in a Hot Coflow

- 3D slot burner configuration:
  - $L_x \times L_y \times L_z = 30 \times 40 \times 6 \text{ mm}^3$  with
  - 1.28 billion grid points
  - High fuel jet velocity (204m/s); coflow velocity (20m/s)
  - Nozzle size for fuel jet,  $H = 2.0\text{mm}$
  - $Re_{\text{jet}} = 10,000$
  - Cold fuel jet (18%  $\text{C}_2\text{H}_4$  + 82%  $\text{N}_2$ ) at 550K,  $\eta_{\text{st}} \approx 0.27$
  - Detailed  $\text{C}_2\text{H}_4/\text{air}$  chemistry, 22 species 18 global reactions, 201 steps
  - Hot coflow air at 1,550K

C. S. Yoo, et al. *Proc. Comb. Inst.* 2011

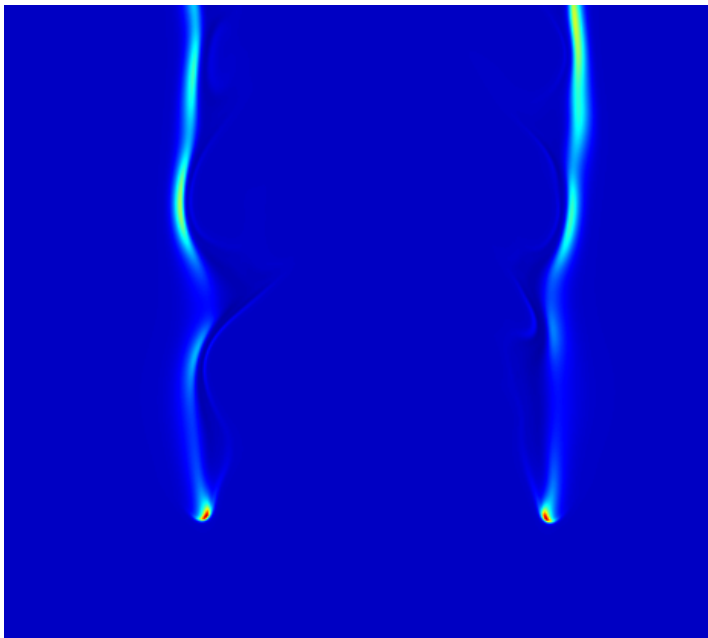


Ethylene-air lifted jet flame at  $Re=10000$

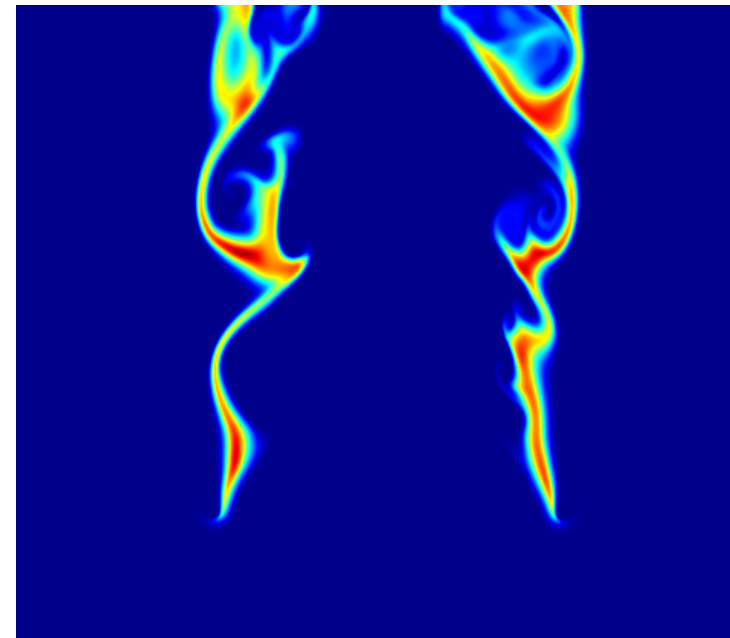


## Lifted Turbulent DME Jet Flame at 40 atm

- AMR and S3D comparisons
- AMR to adapt around thin flame front and turbulent jet
- Run in progress on Edison



Heat release rate



$C_2H_4$



# Turbulent Premixed H<sub>2</sub>/air Counterflow Flame with Product Stratification

- Investigate the effect of turbulent strain and enthalpy stratification on H<sub>2</sub>/air premixed flame structure and local extinction/reignition in turbulent counterflow in the thin/broken reaction zones regime
- Perform joint DNS (S. Lyra, J. Chen, H. Kolla) and OH LIF experiments (J. Frank and B. Coriton) of *highly turbulent*, H<sub>2</sub>/air flames stabilized against non-adiabatic stoichiometric combustion products
- The configuration provides a compact cylindrical volume delimited by the two co-axial nozzles & promotes local flame quenching and re-ignition



# Experimental Counterflow Configuration

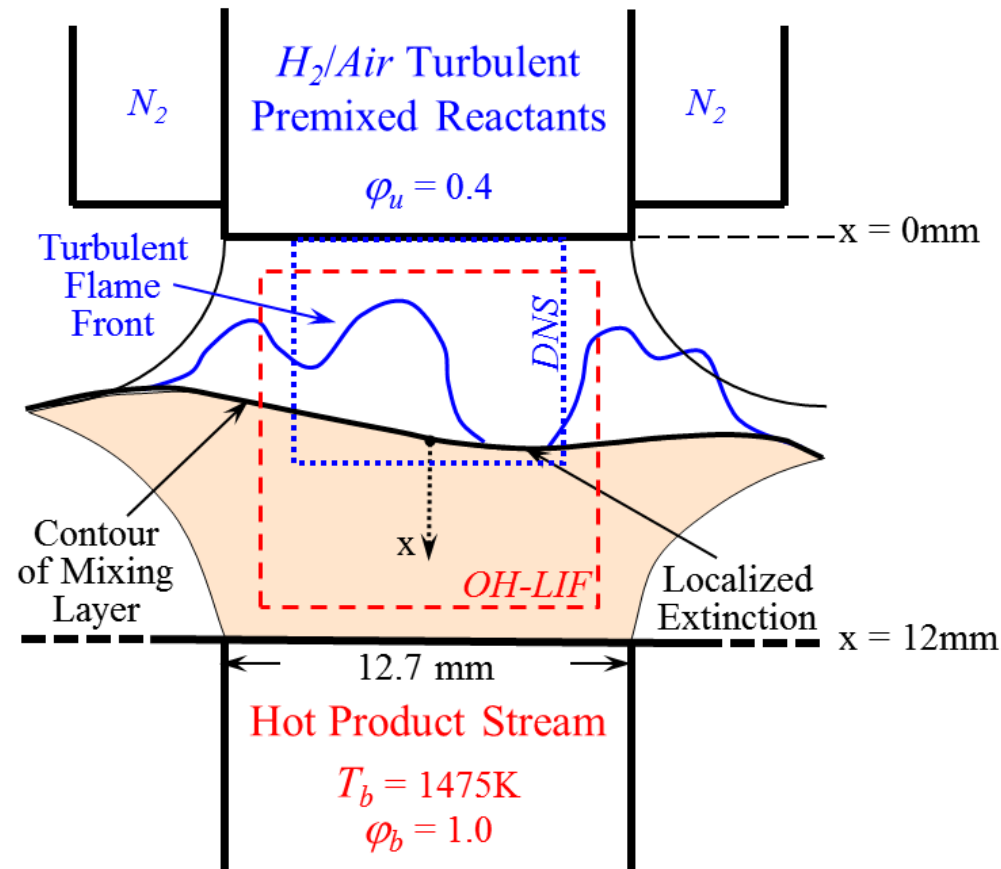
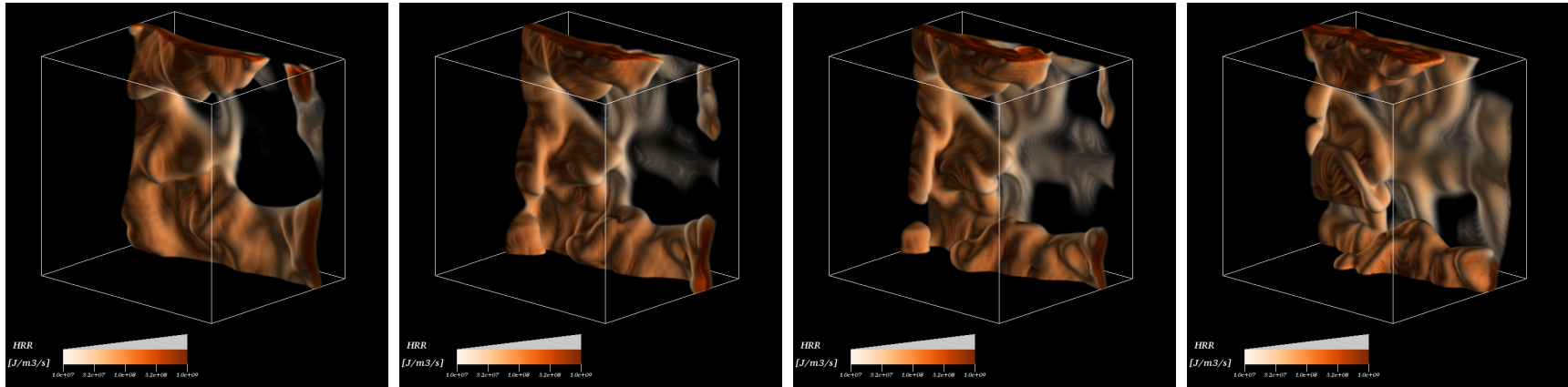


Figure 1: Schematic of a turbulent counterflow flame. The *OH-LIF* signal was analyzed in the  $10 \times 10 \text{ mm}^2$  red dashed square box and the *DNS* data in a  $6.6 \times 8 \times 8 \text{ mm}^3$  rectangular volume denoted by the blue dotted box in the *xy*-plane of the schematic. The full *DNS* domain includes the co-flow and is  $12 \times 17.5 \times 17.5 \text{ mm}^3$ .

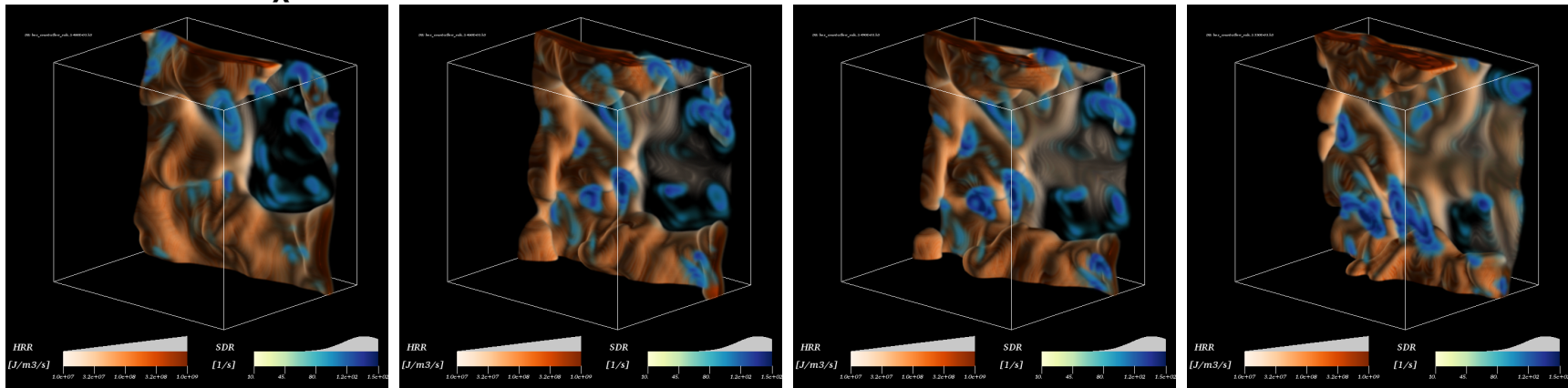


# Flame Recovery by Ignition

Heat Release



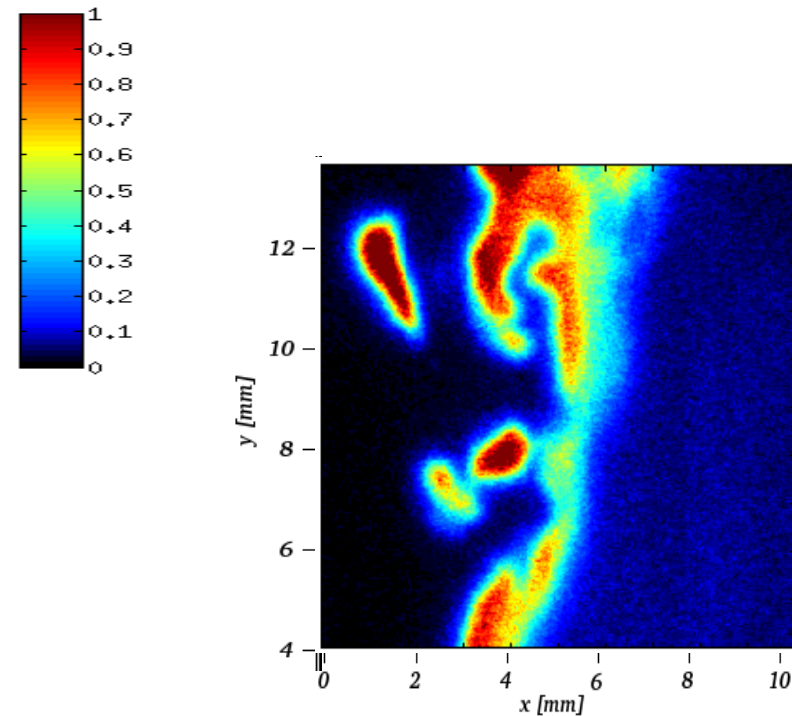
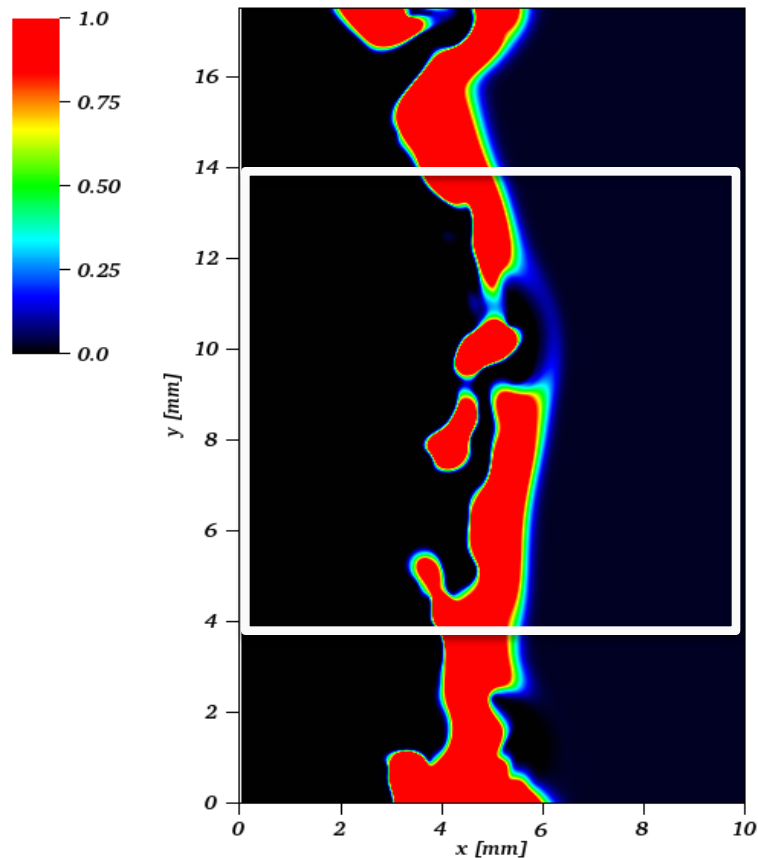
Heat Release and  $\chi$



- Ignition kernel forms within the extinguished region in a low  $\chi$  region
- Relight starts a cusp close to the hot products and forms a thin reacting surface
- Eventually the burning intensity of the surface increases & the flame re-ignites supported by the continuous supply of reactants and the addition of heat from the products



# Comparison with Experiments



- Similar flame structure as experimentally measured
- Flame thickness, expressed in terms of mean progress variable  $c$ , is 2-3 mm, consistent with experimental data measured at similar turbulence & strain

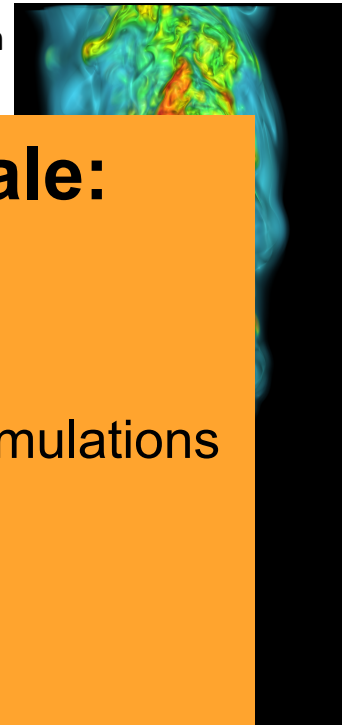


# Petascale Codes

- S3D

- Compressible formulation

S3D simulation  
of HO<sub>2</sub> ignition

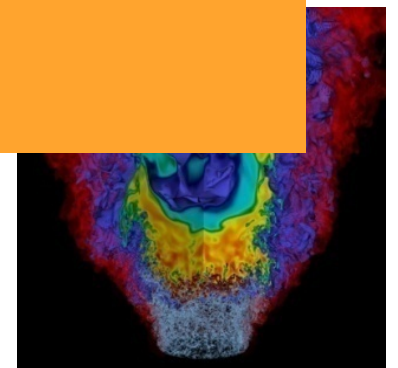


## Target computational capability at exascale:

- High-fidelity physics
  - Detailed chemical kinetics
  - Multicomponent species transport
- Supports both compressible and low Mach number formulations
- Block-structured adaptive mesh refinement
- Higher-order spatial discretizations
- Higher-order temporal integration
- **Support for embedded UQ**
- ***In situ* analytics**

- Hybrid parallel model with MPI + OpenMPI

LMC simulation  
of NO<sub>x</sub> emissions  
from a low swirl  
injector



**ExaCT co-design is working on a new code base**



# S3D Solution Algorithm

- Method of lines solution:
  - Replace spatial derivatives with finite-difference approximations to obtain coupled set of ODEs
  - 8<sup>th</sup> order centered approximations to first derivative
  - Second derivative evaluated by repeated application of first derivative operator
- Integrate explicitly in time
- Thermochemical state and transport coefficients evaluated pointwise
- Chemical reaction rates evaluated point-wise
- Block spatial parallel decomposition between MPI ranks



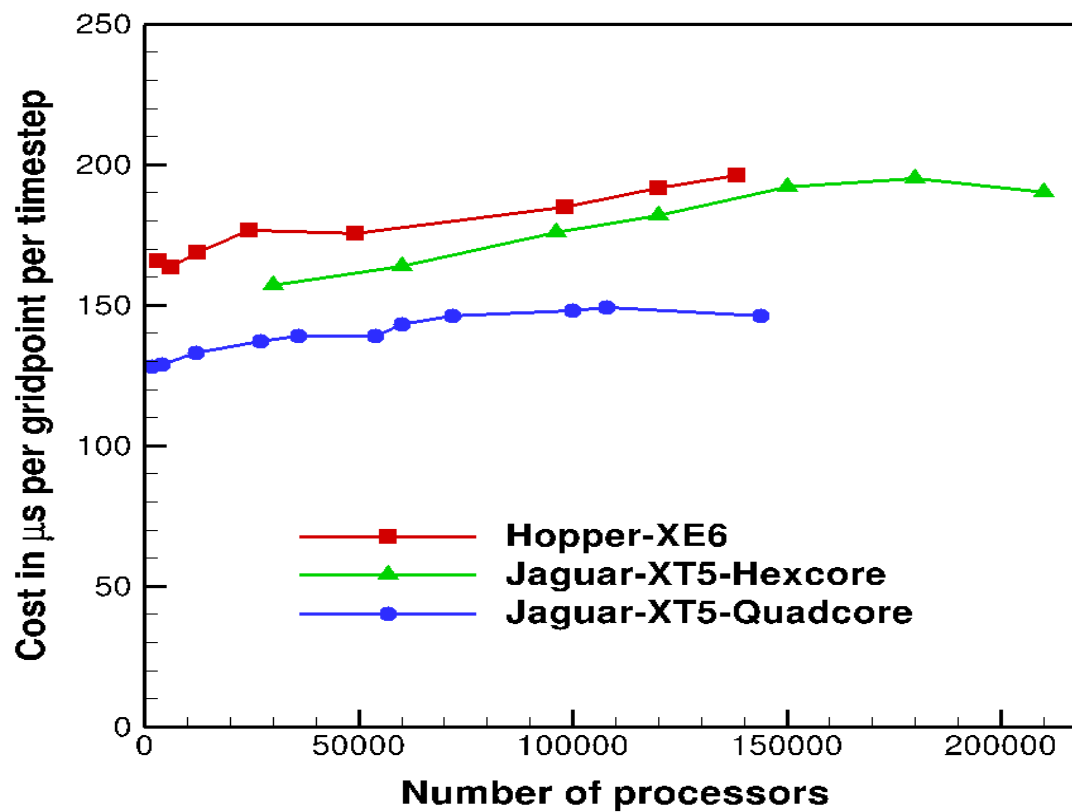


## S3D Solution Algorithm

- Fully compressible formulation
- Fully coupled acoustic/thermochemical/chemical interaction
- No subgrid model: fully resolve turbulence-chemistry interaction
- Total integration time limited by large scale (acoustic, bulk velocity, chemical) residence time
- Grid must resolve smallest mechanical, scalar, chemical length-scale
- Time-step limited by smaller of chemical timescale or acoustic CFL



# S3D Parallel Weak Scaling Performance





## Current HPC Usage

- NERSC and LCF machines: Hopper, Edison and Titan
- 2012-2013 cpu-hrs used: Hopper: 29M, Edison: 2M+, Titan: 50M+
- Typical parallel concurrency and run time: 20-150K cores, ~10-20 M cpu-hrs
- Number of runs: 3-4
- Data read/written per run: ~300TB (200 restart files)
- Memory used per (node/core/globally) 2.3TB (7 billion grids, 14 dof, 3 times size of state= 2.3 TB) global
- Needed software (MPI, ADIOS, VisIt, VISUS, ParaView, MatLab, HDF5)
- Data resources used (scratch/HPSS/GFS) 300TB scratch, ~1 PB HPSS
- Amount of data stored: 3.7M SRU's



## HPC Requirement 2017

- Compute hours needed (in units of Hopper hours): ~1 billion cpu-hrs
- Changes to parallel concurrency, run time, number of runs per year: 100-400K cores, 4 runs per year, 3 weeks on 400K cores
- Changes to data read/written: rate at which data is written out won't scale with problem size, I/O rate limited – move analysis/UQ/viz in situ, end-to-end workflow, staging, NVRAM
- Changes to memory needed per ( core | node | globally ) limited by memory bandwidth/bisection bandwidth and not aggregate memory, communication with stencils is expensive, looking at exploiting both data and task fine-grain parallelism
- Changes to necessary software, services or infrastructure – staging middleware, in situ workflow, dynamic runtime, DSL compiler+ low level IR, combustion portal for sharing data/software with community



# Strategies for New Architectures

- Hybrid Architectures e.g. Titan
- Deferred Dynamic Execution and DSL
- In situ analytics and UQ





# OLCF Center for Accelerated Application Readiness on Titan

- S3D Migration strategy: (2-3 FTE's for 1.5 years, Cray, Nvidia, Sandia, ORNL, NREL)
  1. Requirements for host/accelerator work distribution
  2. Profile legacy code
  3. Identify key kernels for optimization
    - Chemistry, transport coefficients, thermochemical state (pointwise)
    - Derivatives (reuse)
  4. Prototype and explore performance bounds using Cuda
  5. “Hybridize” legacy code: MPI for inter-node, OpenMP intra-node
  6. OpenACC for GPU execution
  7. Restructure to balance compute effort between accelerator and host



## Summary of Titan port

- Significant restructuring to expose node-level parallelism
- Resulting code is hybrid MPI+OpenMP and MPI+OpenACC (-DGPU only changes directives)
- Optimizations to overlap communication and computation
- Changed balance of effort
- For small per-rank sizes, accept degraded cache utilization in favor of improved scalability



## Lessons Learned

- Reworked code is 'better': more flexible, well suited to both manycore and accelerated
  - GPU version required minimal overhead using OpenACC approach
  - Potential for reuse in derivatives favors optimization (chemistry not easiest target despite exps)
- Majority of work is done by GPU: extra cycles on CPU for new physics (including those that are not well suited to GPU)
- We already have 'Opteron + GPU' performance exceeding by factor of 2 over Opteron performance
- We have the 'hard' performance
- Specifically moved work back to the CPU
- Production run on Titan with OpenACC S3D



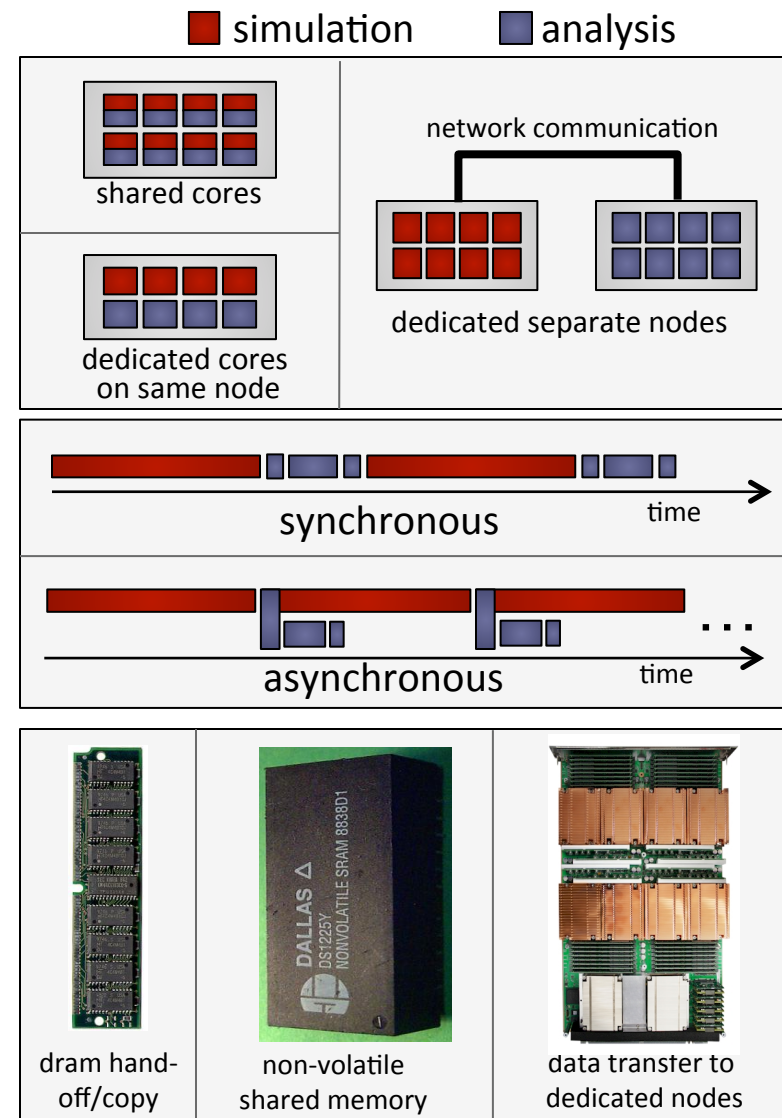
# Future Algorithmic Improvements

- Second Derivative approximation
- Chemistry network optimization to minimize working set size
- Time integration schemes - coupling, semi-implicit chemistry, spectral deferred correction
- Several of these are being looked at by ExaCT co-design center, where the impacts on future architectures are being evaluated
  - Algorithmic advances can be back-ported to this project



# Rich Design Space of Workflows at Exascale

- **Location of analysis compute resources**
  - Same cores as the simulation (in-situ)
  - Dedicated cores on the same node (in-situ)
  - Dedicated nodes on the same machine (in-transit)
  - Dedicated nodes on external resource (in-transit)
- **Synchronization and scheduling**
  - Execute synchronously with simulation every  $n^{\text{th}}$  simulation time step
  - Execute asynchronously
- **Data access, placement, and persistence**
  - Shared memory access via hand-off / copy
  - Shared memory access via non-volatile near node storage (NVRAM)
  - Data transfer to dedicated nodes or external resources

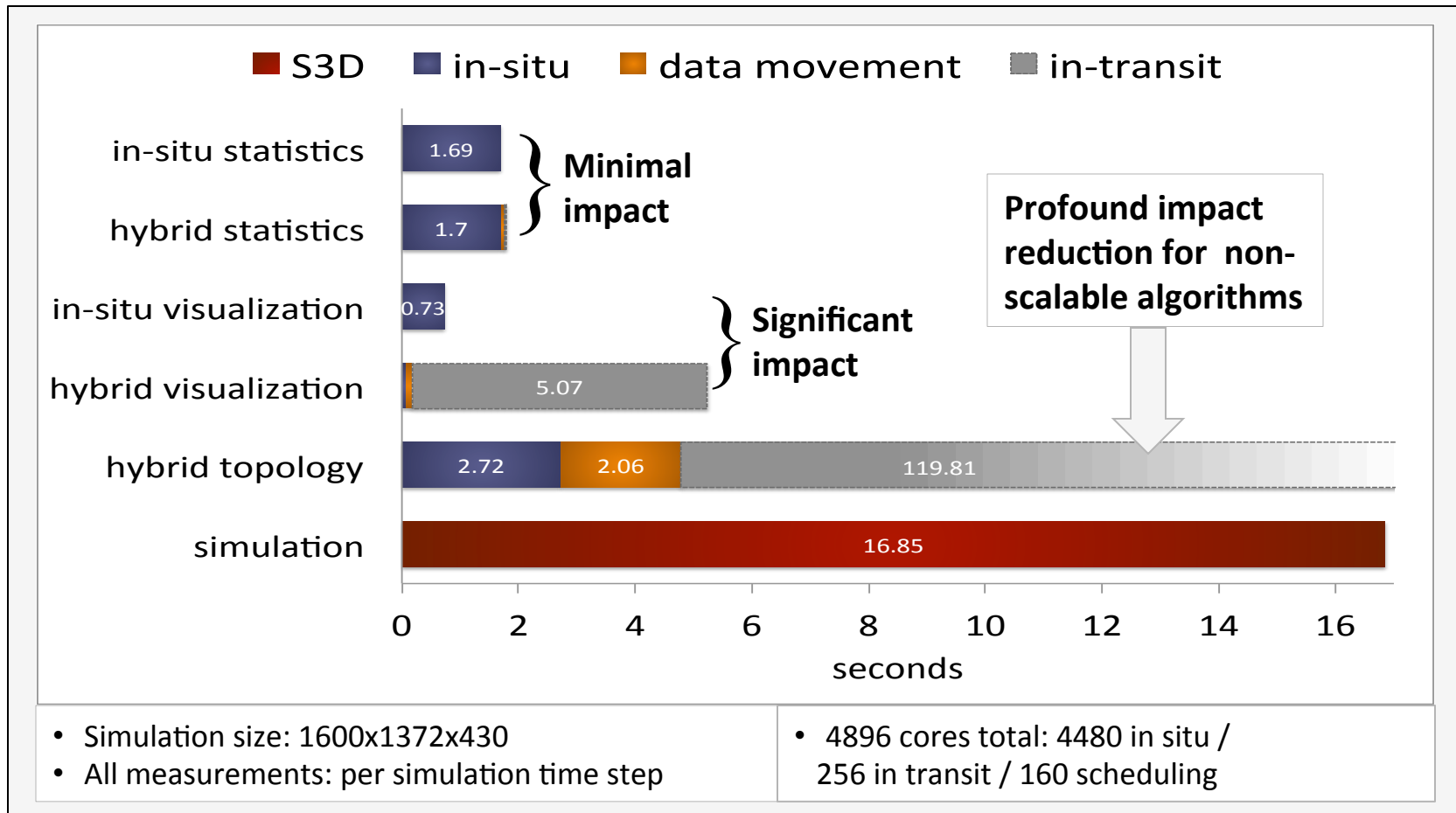




## An empirical study on Titan demonstrates hybrid workflow architectures show promise for minimizing impact to the simulation

- Primary resources: execute main simulation & in situ computations
- Secondary resources: staging area for in transit computations

*Bennett et al. SC12*

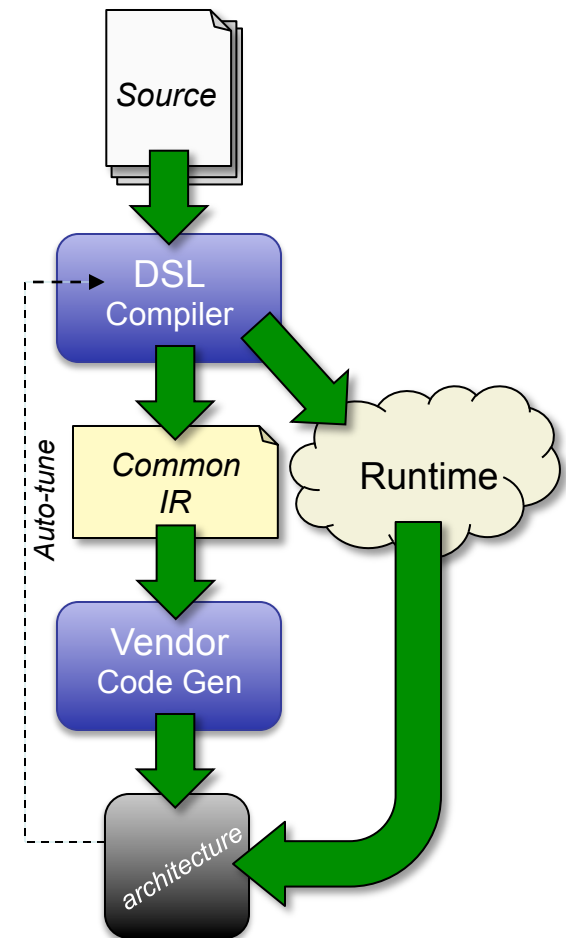






# Domain-Specific Languages

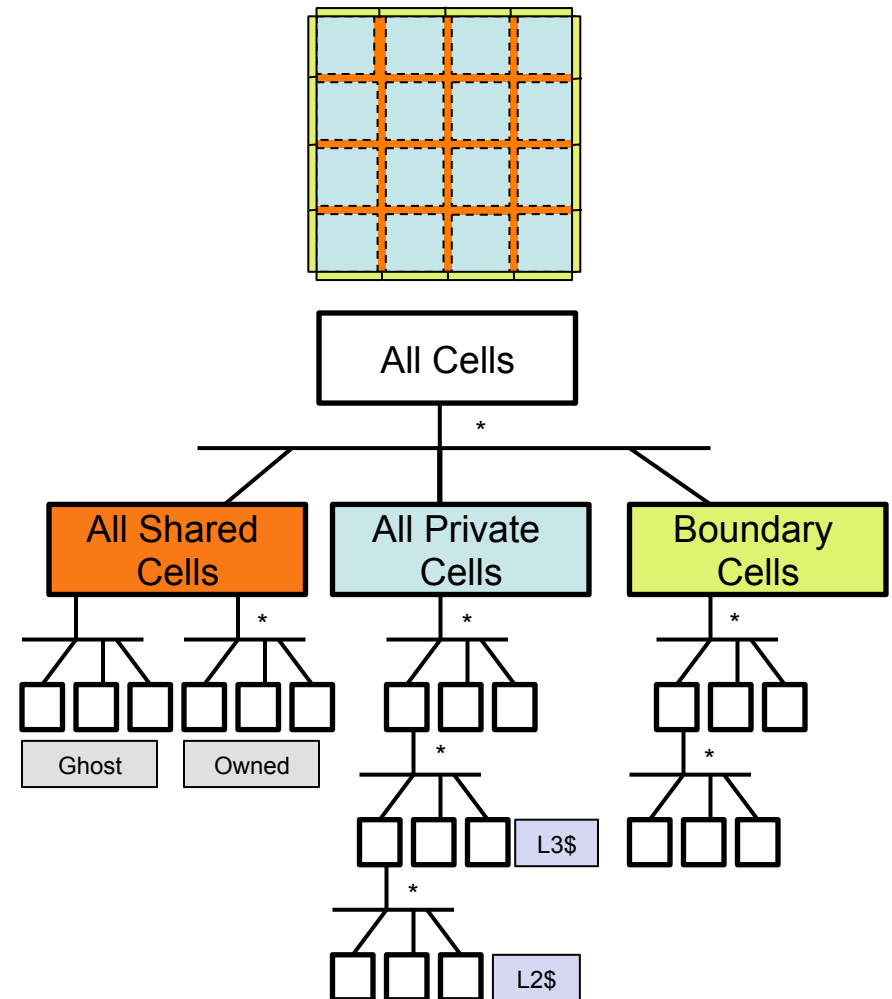
- A domain-specific language (DSL) is a language of **reduced expressiveness** targeted at developers in a specific, focused problem domain
- **Perfect fit for hybrid many-core architectures.**
  - Influenced and driven by both the **domain** and **architecture**
  - Contributes directly as an enabling technology that insulates applications from the complexity of the architecture. *Specialized software that works well with complex hardware.*
- Starting points:
  - Team has previous experience in DSLs via ASC PSAP (Stanford) – unstructured CFD, DSL with LLVM (LANL)
  - **LLVM Common IR** for vendor toolchain interactions and supporting infrastructure (not building compiler from scratch). Using as FastForward interface for code generation (with AMD and NVIDIA). Intel? IBM?
  - Orion Stencil DSL
  - Memory-hierarchy-aware programming model and runtime (*Legion – leveraging initial DARPA funding*)
  - Terra Language (DSL building blocks)





# Legion: Programming Locality and Independence

- Programming based on logical **regions** to hierarchically describe organization of data
  - Legion tasks operate on logical regions, specifying required privileges (read-only, read-write, reduction) and coherence (exclusive, atomic, etc.)
  - **Mapping** of tasks and regions onto processors and memories. *Programmable from both **domain** and **architecture** points of view*

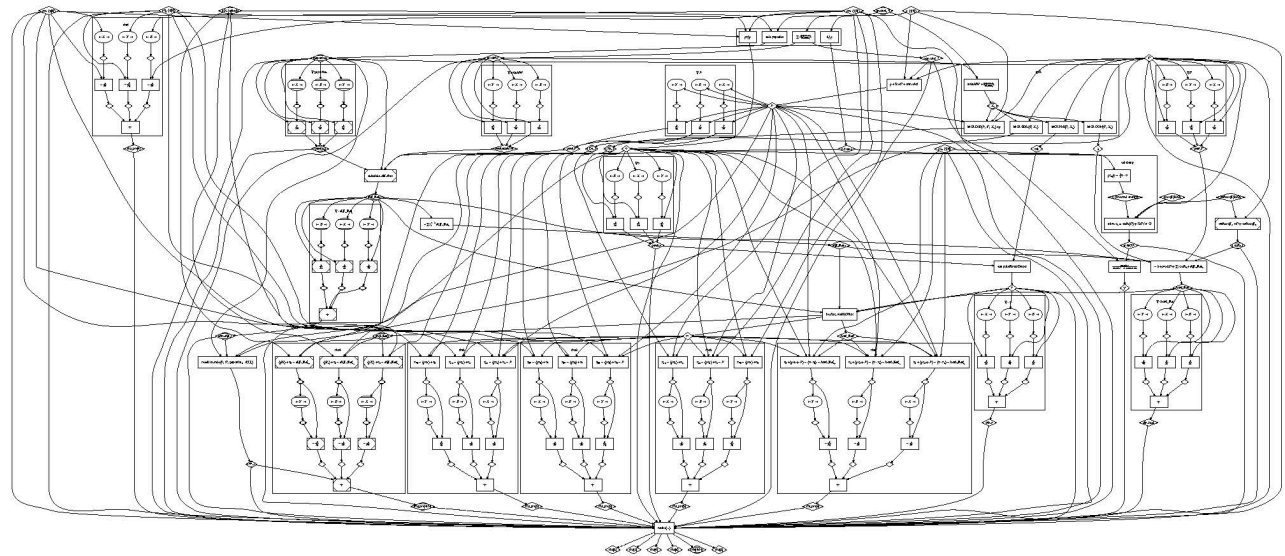




# Exposing Locality and Expressing Parallelism in S3D

- Tasks are coded in familiar sequential style (subtasks look like function calls).
- Legion runtime uses region information to automatically extract parallelism and map tasks using the same data together to benefit from locality.
- Current S3D Fortran has no task parallelism (interleaves 6 inter-node transfers at a time).
- Straightforward port of S3D to Legion:
  - Results in identification of 100's of independent tasks by runtime
  - Allows interleaving of up to 100's of inter-node transfers to hide communication latencies.
  - Has a code structure that is very similar to current Fortran code

## S3D Data Flow Graph in Legion

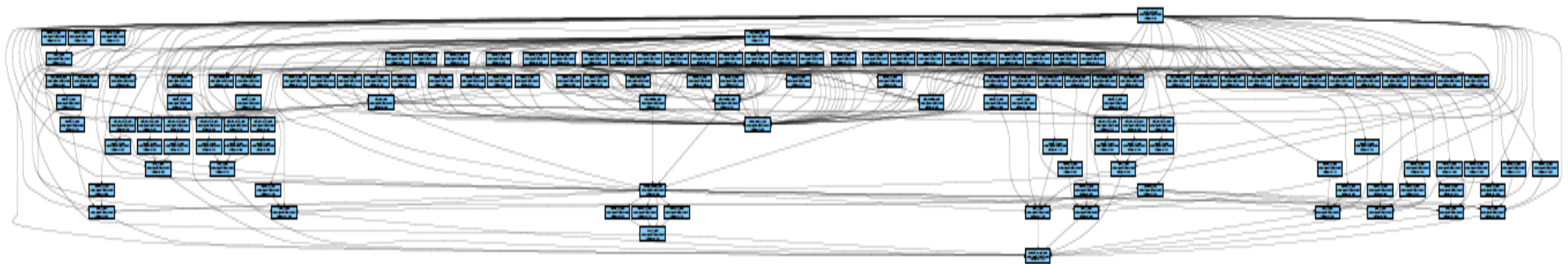


*Aiken, Bauer, Treichler*



# S3D Task Parallelism

- One call to RHSF as seen by the Legion runtime
  - Arrows indicate true data dependences
  - Width == task parallelism
  - H2 Mechanism (9 species)
  - Heptane (52 species) is even wider



5/14/2013

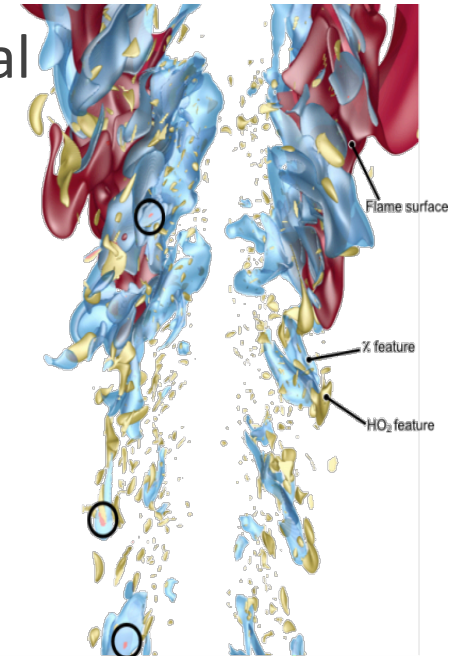
33



# Intrusive Uncertainty Quantification - Chemistry

## Models are Source of Error in DNS:

- **Motivation:** Inform chemical kineticists and chemical engineers in the fuels industry (BP, Exxon) on what chemical properties need to be pinned down more accurately for optimal utilization of a given fuel
- Embedded chemistry model source of uncertainty
  - Reaction rates
  - Missing reactions
  - Transport coefficients
- Combustion intermittency characterized by space-time localized phenomena of interest, tractable for UQ
- Solve adjoint equations backward in time: need the primal state at all times (data management opportunity for co-design)





# Future Directions

- Develop more complete models for exascale combustion simulation
  - Architecture aware AMR algorithms
    - Incorporate interconnect topology into AMR data layout
    - Dynamic load-balancing that incorporates communication costs, data movement and dynamic machine behavior
  - Capture model for complete work flow – solver, analytics, UQ
    - Encapsulate all aspects of analysis into performance model
- Explore hardware tradeoffs with vendors and CS collaborators
  - Refined analysis of node architectures
  - Analysis of network behavior for AMR at scale
  - Explore tradeoffs for machine balance
- Focused interaction with programming environment community to ensure that future programming models will support effective expression of methodology needed for combustion simulation





## **30X faster machine in 2017 will enable us to push closer to aero-thermo-chemical regime of compression ignition engines and gas turbines**

1. Higher pressure (1 atm vs. 30 atm.)
  2. Higher Turbulent Reynolds number (1000 vs. 3000)
  3. Complex Fuel Chemistry (methane (C1) vs. n-heptane (C7))
- Lifted DME or n-heptane diesel jet flames at high pressure and turbulent Reynolds numbers of 3000 (jet Reynolds numbers of 30000)
  - Natural gas jet-in-crossflow into a vitiated cross-stream for flashback safety compared to laboratory gas experiments
  - Spherically expanding turbulent premixed flames at moderate pressure (5-10 atm) with lean DME compared with experiments as part of CEFRC



**Thank you!**  
**Questions**  
**[jhchen@sandia.gov](mailto:jhchen@sandia.gov)**