EXPERIENCE SCALING NEK5000 COMBUSTION CODE FOR MIRA

Ammar Abdilghanie

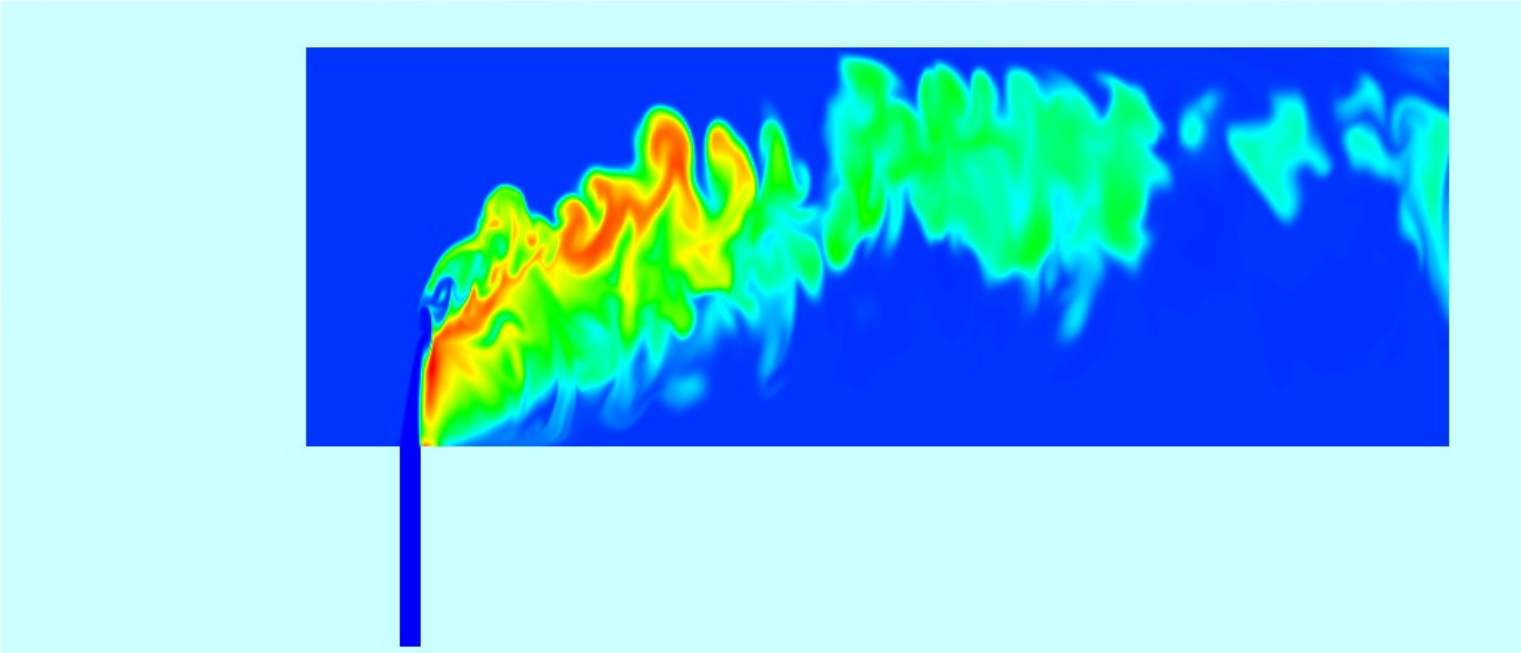
Mira Community Conference - March 4-8, 2013

Argonne Leadership Computing Facility

NEK5000

- **Multi-physics** DNS/LES code for fluid flow, heat transfer, MHD & combustion.
- **100k lines of code**: Mainly f77 O(70k) , & C O(30k)
- **Interface** w/ state of the art visualization & grid generation: **VisIt, CUBIT-MOAB & Gambit.**
- **High order** spectral element method (locally structured, globally unstructured)
- **Highly scalable** variant of Nekton (first commercial distributed memory computer code marketed by Fluent in mid 90s)

Autoignition of a JICF



- Mixture tendency to Autoignite & stabilize.

- Understanding Flame Anchoring Mechanism.

- Flash-back hazard in stationary Gas turbines

Computational Challenges for Autoignition of a JICF

- **Disparate length and time scales:**
  - **turbulence & mixing caused by myriad of vortical structures (Horse-shoe, CVP , wake vortices..etc)**
  - **Resolve Kolmogorov & Batchelor time scales**
  - **Flame/reaction zone thickness.**
  - **Flame time scale.**
  - **Chemical species reaction rates (highly nonlinear functions of local composition).**

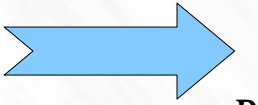
- **Numerical stiffness!**
- **DNS is a necessity for high fidelity simulations.**

Computational Challenges for Autoignition of a JICF II

- **Detailed chemistry (account for chain branching and terminating elementary steps)>> Can capture local extinction & reignition.**

- **Accurate account of differential diffusion effects through detailed multi-component transport model.**

Grid Resolution for DNS of turbulent combustion



$$N = (R)^{9/4} (\eta/\delta)^3$$

$$Flame\,Thickness\,(\delta) = D/S_L$$

$$R = Reynolds\,Number\,,\ D\ Diffusivity$$

$$S_L\ Laminar\,Flame\,Speed$$

Spectral Element Method

- **Variational formulation** , like FEM with GLL quadrature.

- **Domain decomposition** into E high order, deformed, quadrilateral/hexahedral elements with h & p-type refinements.

- **High order accuracy** (resolve fine scale and high frequency events typical of turbulent flows)

- **Converges exponentially fast** with polynomial order (N) for smooth solutions.

- **Typically used with third order accuracy time steppers.**

Computational Efficiency

- **High-order:**
  - Minimal numerical dispersion (suitable for turbulence and wave-dominated phenomena)

- **Efficiency:**
  - Recast tensor products into mxm products
  - Solvers: MG-preconditioned CG or GMRES

- **Stability:**
  - Spectral filters
  - Dealiasing (eliminate spurious eigenmodes)

Computational Efficiency

- **For n=E N^3 grid points, require**
  - **O(n) memory accesses**
  - **O(nN) work in the form of matrix-matrix products**

- **Scalability:**
  - **Scalable coarse-grid solvers (sparse-basis projection or AMG)**
  - **Design for Procs > 10^6**
  - **Low iteration counts for typical simple domains**
  - **Iteration counts higher (still bound) for complex geometries**

NEK5000 Implementation I

- Eliminate Loop clutter (minimize conditionals and function evals inside a loop and minimize number of nested loops
- Increase data reuse
- Use unit stride data access
- Tensor products converted to highly optimized matrix-matrix products (dgemm , loop unrolling and assembly code)
- Use non-blocking MPI sends and receives

NEK5000 Implementation II

- **Efficient communication strategies for inter-element data exchange.**

- **Efficient, automated domain decomposition strategies to ensure load balancing**

- **Parallel IO**
  - **MPIIO**
  - **Subset of ranks**

Parallel Scaling

- **Strong scaling :** Increasing number of cores/ranks for a constant problem size--->> Ideally, “Linear Speedup”

- **Weak scaling :** Double problem size & double number of cores --->> Ideally, “Constant run time”

- Scalability issues may result from :
  - Computations
  - Communication
  - I/O

Most Efficient Problem Size ?

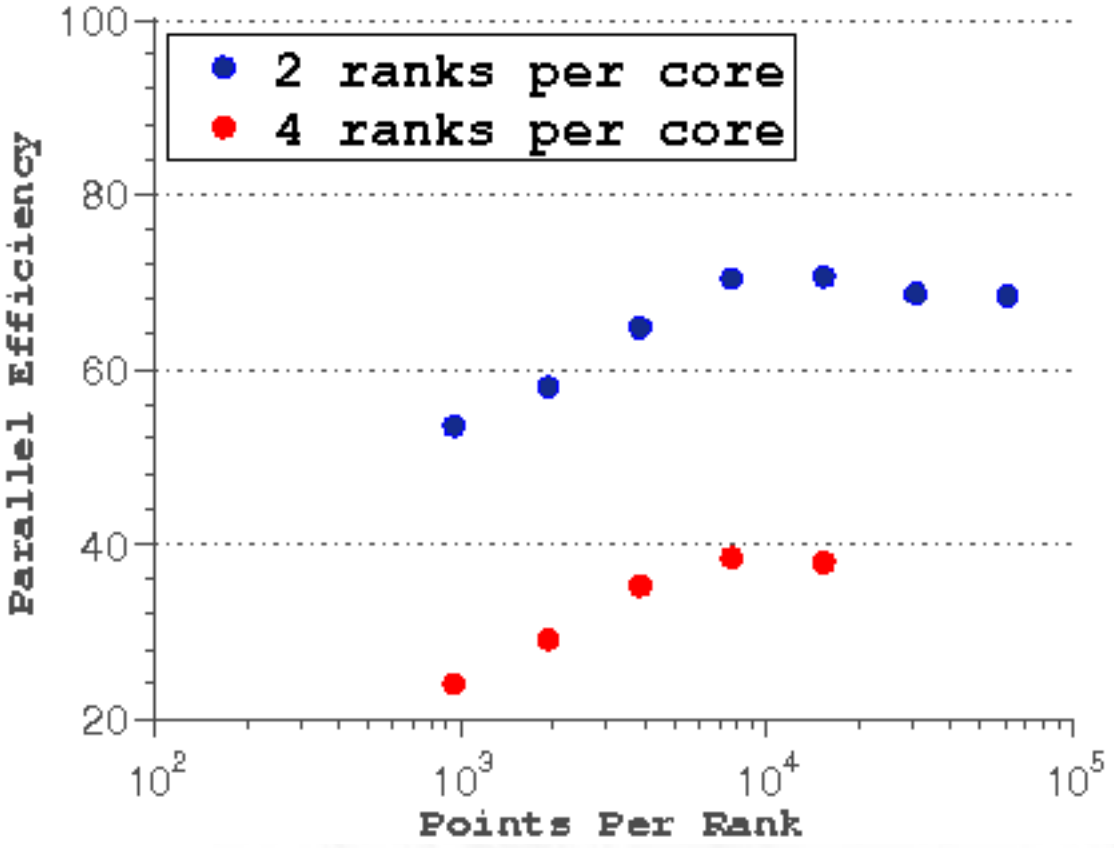
$$N = n^d * P$$

- **Increase core utility by threading.**
- **Minimize resource contention**

Parallel Efficiency $$E = \frac{N1\ t(N1)}{N2\ t(N2)}$$

$$N2 > N1$$

Core-level/Threading Efficiency



Best parallel efficiency achieved with 2 ranks per core
When each rank has >= 7000 grid points!

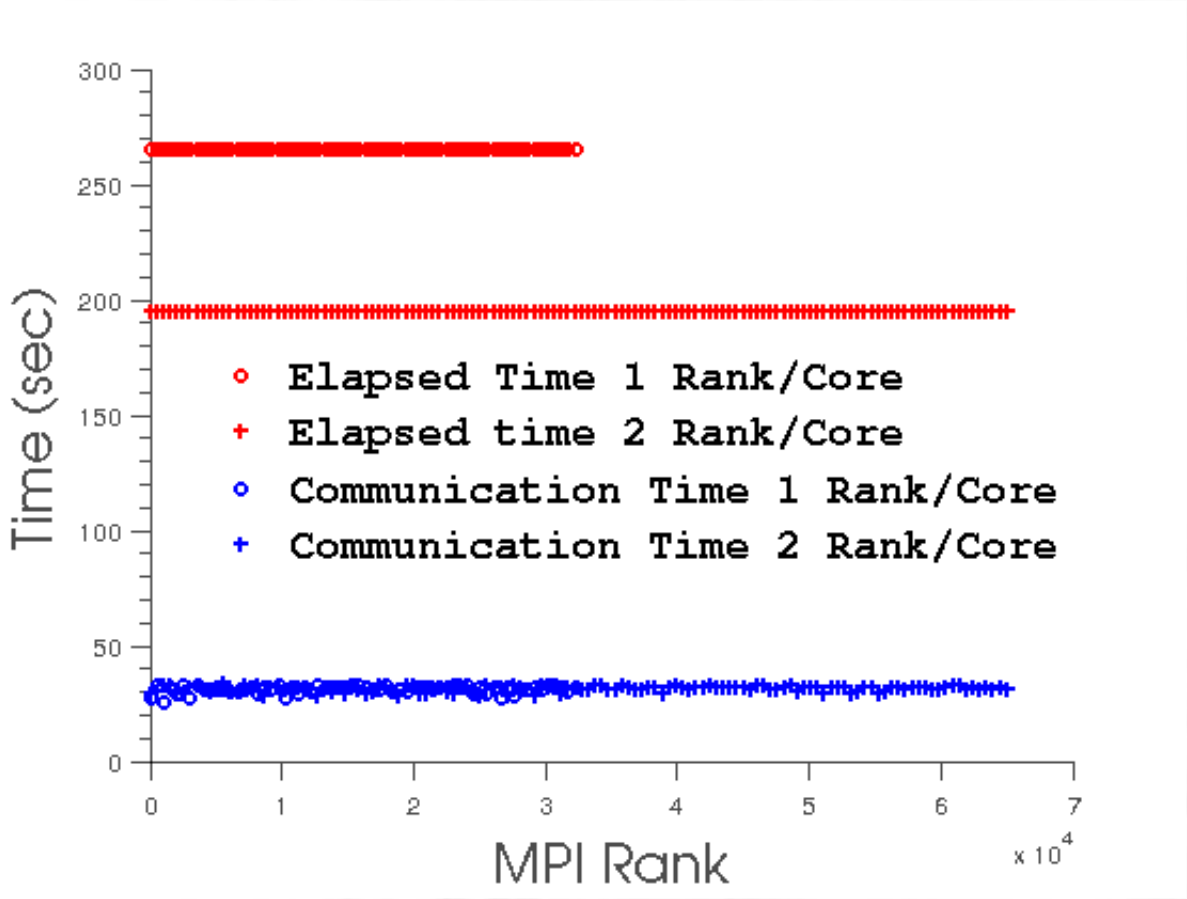
MPI profile using IBM HPC toolkit: (Snapshot for a 1 rank/core case, @ Nek_advance

Data for MPI rank 0 of 32768
Times and statistics from summary_start() to summary_stop().

MPI Routine	#calls	avg. bytes	time(sec)
MPI_Isend	519129	1290.6	1.653
MPI_Irecv	519129	1295.9	0.394
MPI_Waitall	153783	0.0	9.204
MPI_Bcast	1	4.0	0.000
MPI_Barrier	81	0.0	0.002
MPI_Allreduce	19290	2717.6	15.658
MPI_File_close	1	0.0	0.087
MPI_File_open	1	0.0	0.293
MPI_File_seek_shared	27	0.0	0.002
MPI_File_write_all	29	24547.6	11.239

total communication time = 26.911 seconds.
total elapsed time = 265.360 seconds.
heap memory used = 108.109 MBytes.

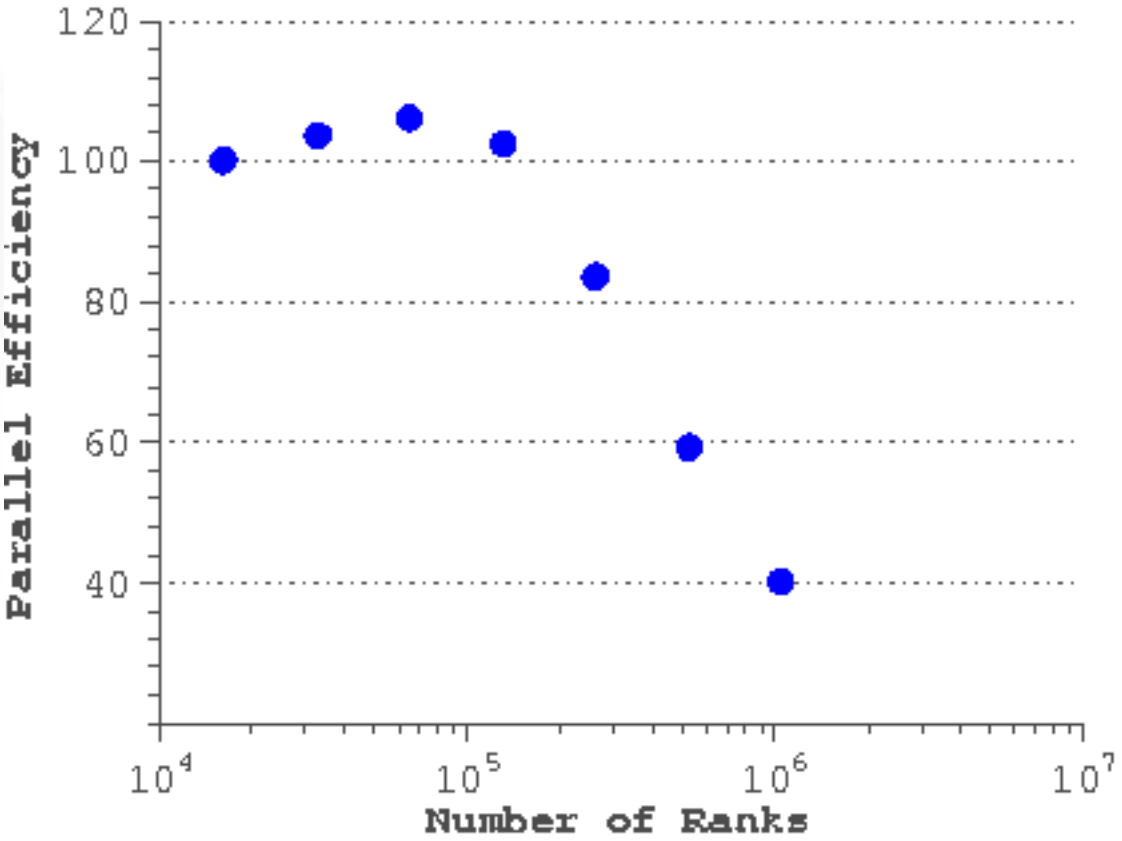
MPI Load Balancing



Hardware performance counter using IBM hpm

	1 Rank/Core	2 Rank/Core
FPU% , FXU %	20.59% - 79.41%	17.7% - 82.29%
Inst/cycle/core	0.3586	0.5794
GFLOPS/node	4.398	6.057
% of max issue rate/core	35.86%	47.68%
L1 d-cache hit	96.37%	95.52 %
L1P buffer hit	1.46%	2.09%
L2 Cache hit	1.85%	2.0%
DDR hit	0.32%	0.4%
DDR total traffic	2.264 Bytes/cycle	3.588 Bytes/cycle

Strong Scaling



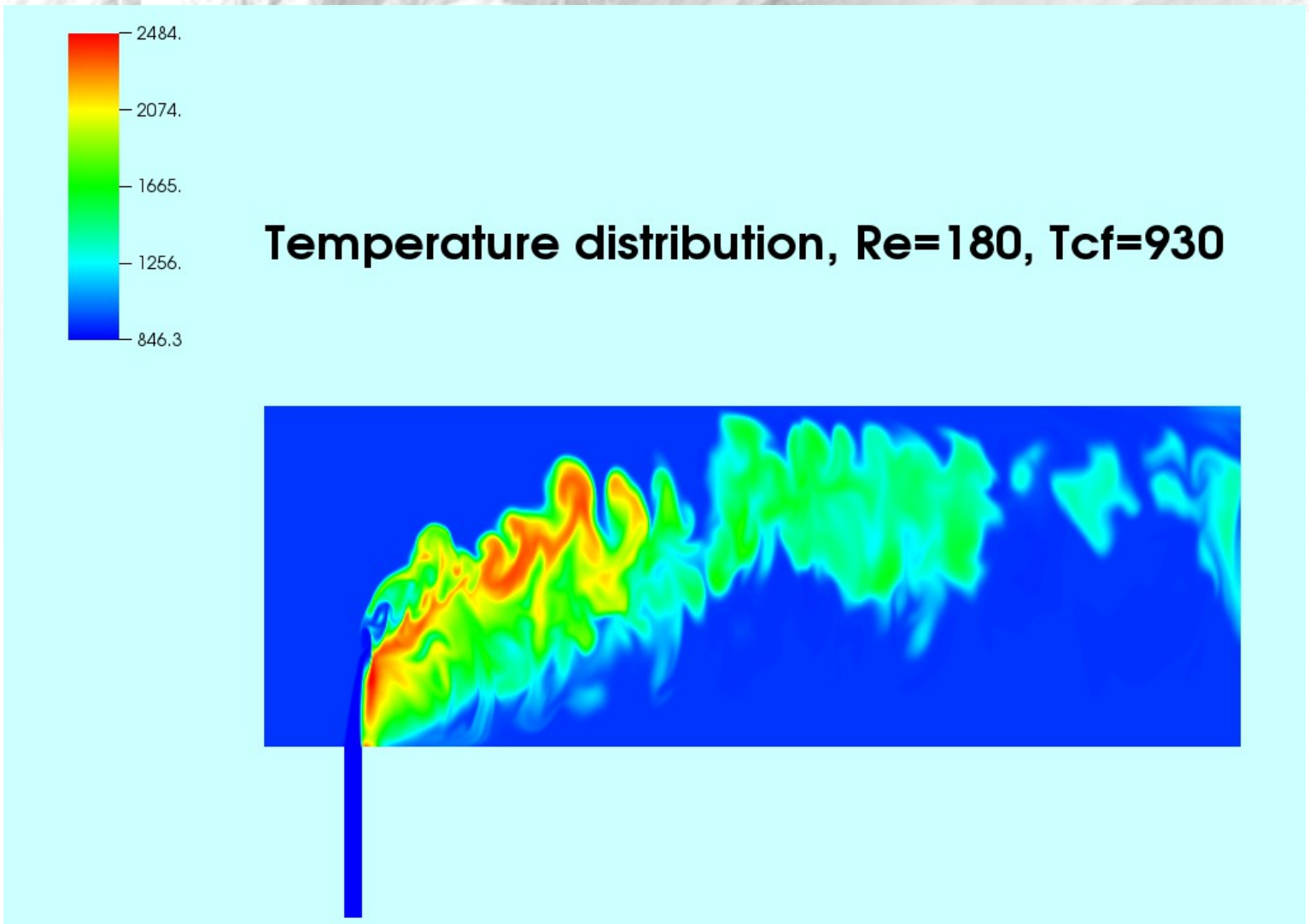
3 million elements
Polynomial order 7
(~ 1 billion grid points)

➔ 100% parallel efficiency up to ~ 130k ranks.

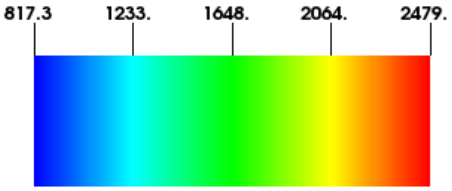
➔ 40% on 1 million ranks!

Visualization with VisIt's Parallel Architecture:

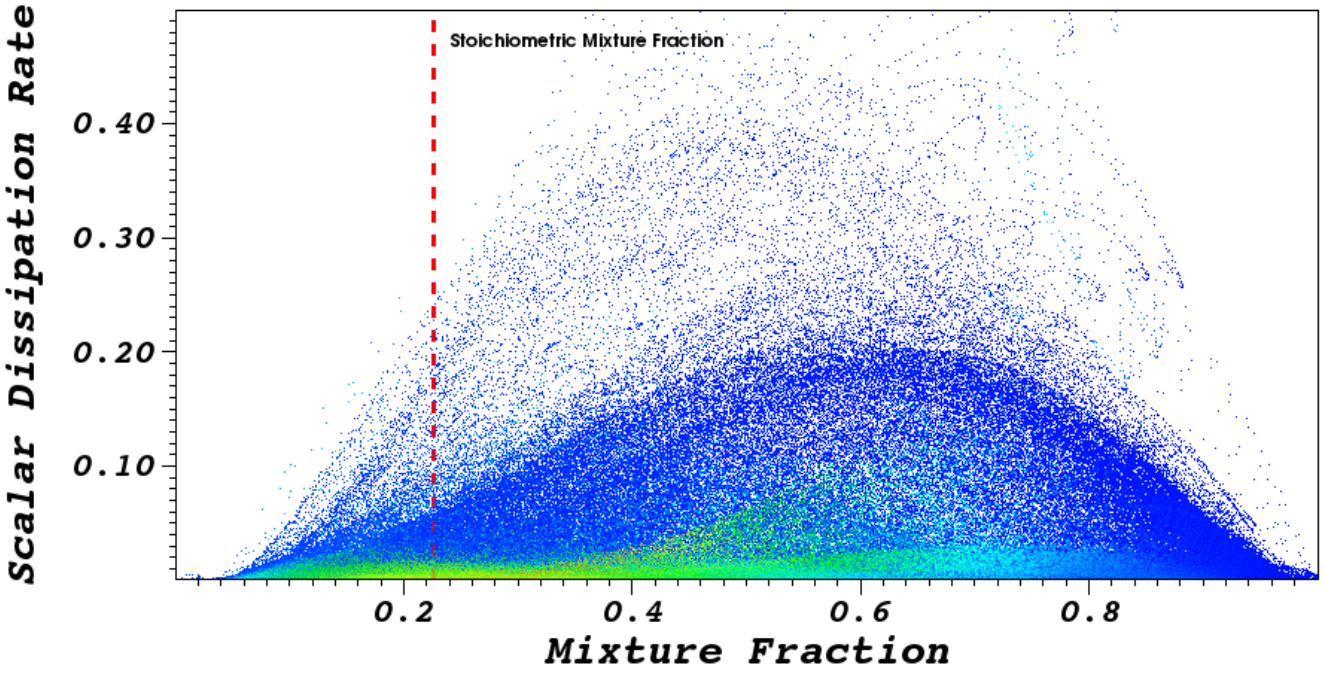
- Snapshot size 22 GB at low Re & 84 GB at high Re
- 200-500 snapshots are needed to simulate full transition to stable flame>>> 10-40 TB of data/simulation
- Globus online to transfer files to Eureka.



Pseudocolor plot of temperature distribution on a x-z plane in the Center of the jet at flame stabilization (t=30.225)



Temperature (K)



Scatter plot of mixture fraction & scalar dissipation rate of the stable flame (t=30.225) [colored by temperature]

Thank You!