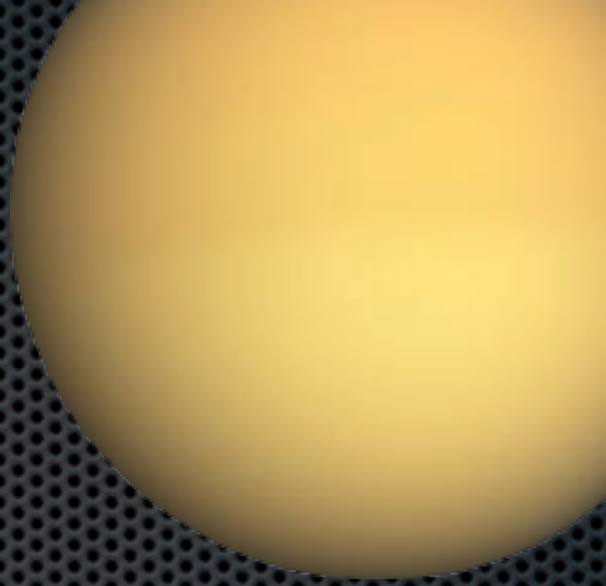


himera on Titan:

ploiting new parallelism for supernova simulations

ented by:

Lentz (U. Tennessee/ORNL)

son Messer (OLCF)

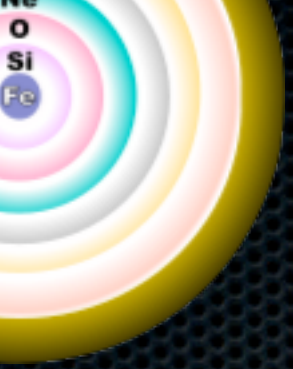
Lewkow (UConn)

nne Parete-Koon (UT -> NCCS)

the Chimera Team:

ruenn (FAU), J. Blondin (NCSU), A. Chertkow (UT), E. Endeve (ORNL), W. R. H

NL/UT), E. Lingerfelt (ORNL), P. Marronetti (FAU), K. Yakunin (FAU) and A.



Stellar collapse to explosion

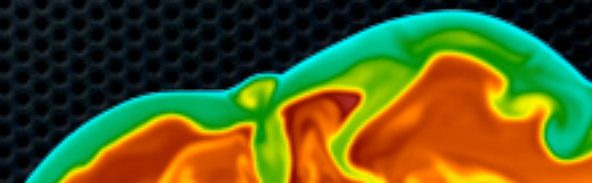
lassive stars (~8x solar mass) build inert Fe-core that collapses until center eaches nuclear density (2e14 g/cc)

ompressed proto-Neutron Star (pNS) launches shock that “stalls” expanding falling layers, leaving a Standing Accretion Shock (SAS) that must be revived xplode star

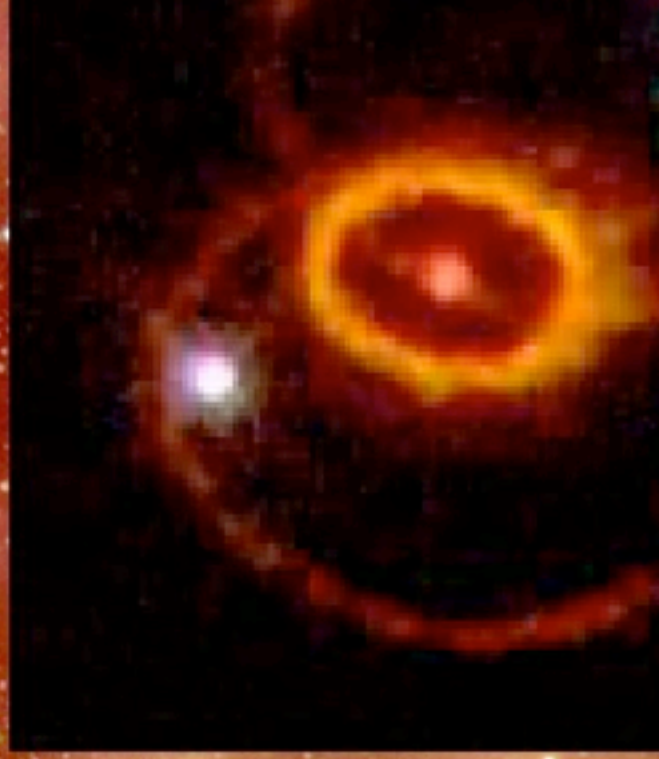
ver ~1 second, the “mantle” between the pNS and shock is heated by neutri nd neutrino-driven convection and builds up enough energy to drive the shoc rough the rest of the star. Bang!

evived shock drives out through rest of star. Shock expells stellar envelope a ggers nuclear burning, reaching the surface with a flash after a few hours.

xploding star appears in sky as a supernova



cientific Goals:



nderstanding the core-collapse supernova mechani

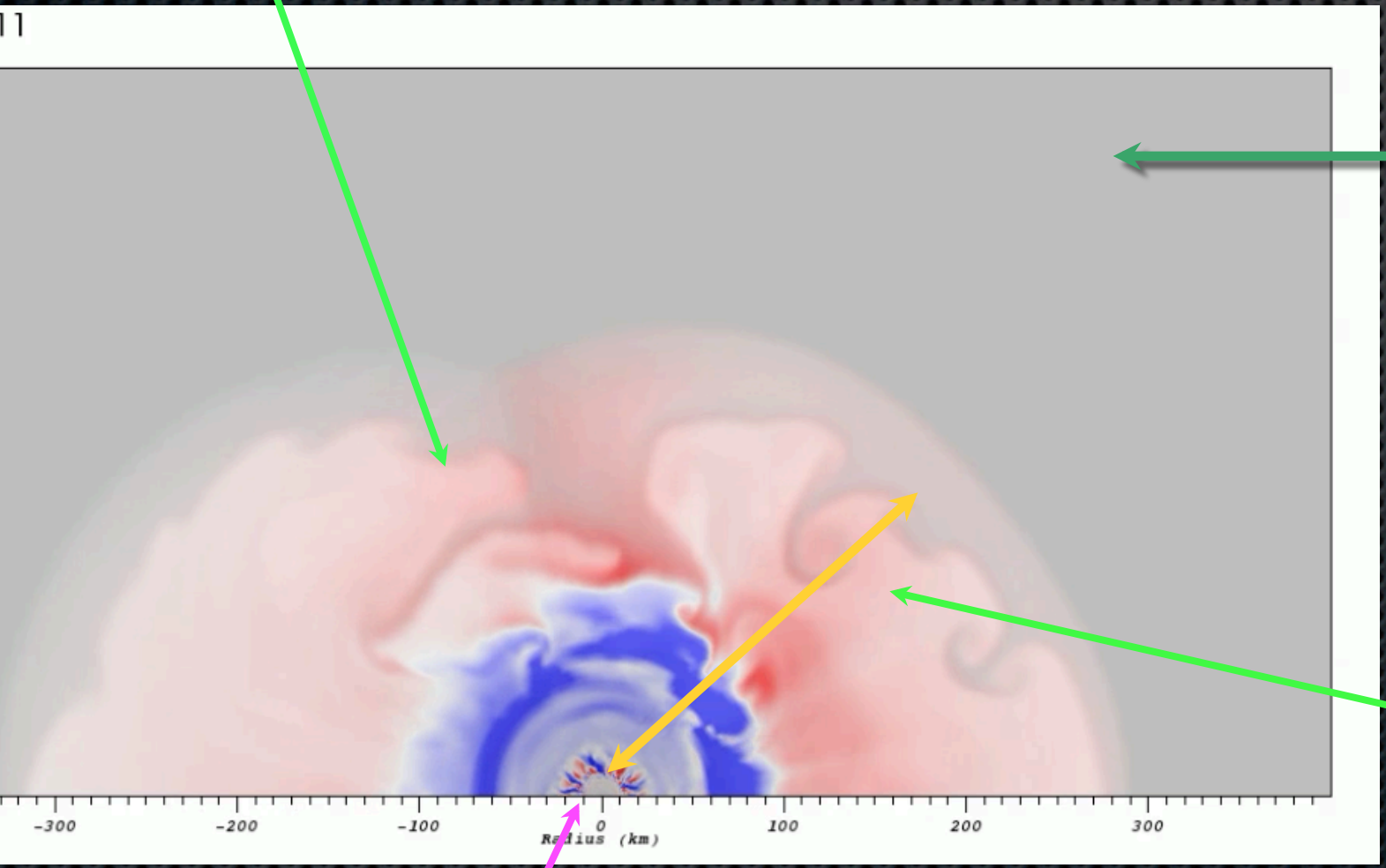
etermining the fates of massive stars (variation in
xplosions; neutron star/pulsar or black hole)

ptical properties of the ejecta (observed Supernovae

hemical element production in supernovae

nerical Challenges

mic “Hot mantle”: neutrino
ed material inside shock in
nuclear equilibrium



Small steps: Dense nuclear equation of state in

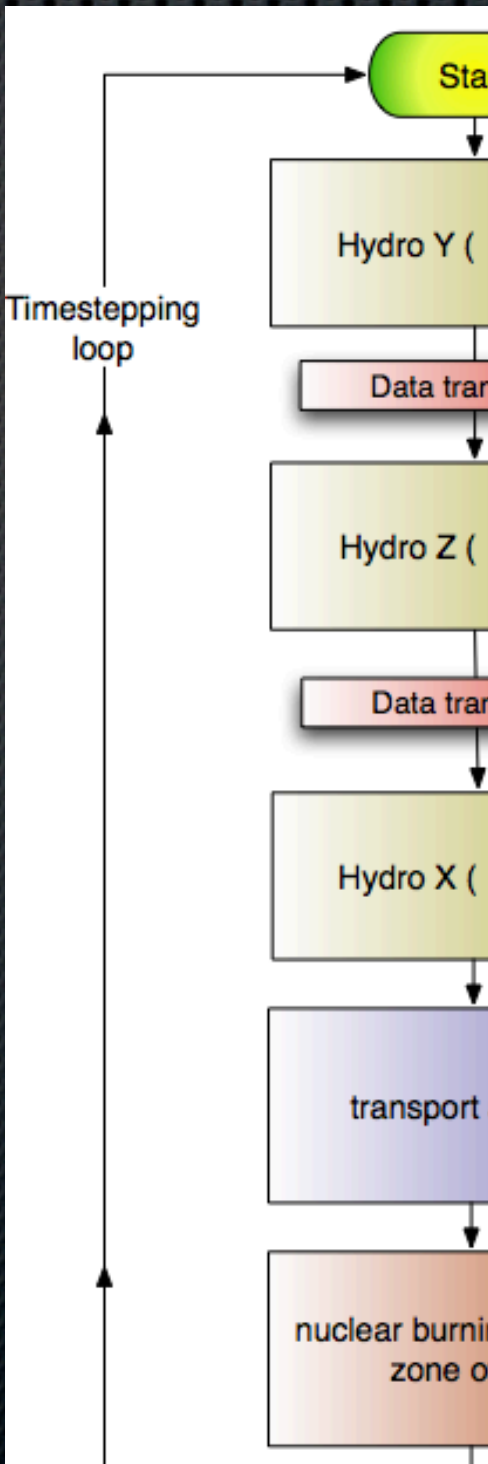
=O(10⁶) sequenti
- a few seconds p
=~2000 hours ru
- 2-/3-D Hydrody
- Radial neutrino t
- Nuclear Eq. of S
nuclear netw

Rxn. Network: I
material is not in
equilibrium. Use
network to comput
and nucleosynth
ejecta

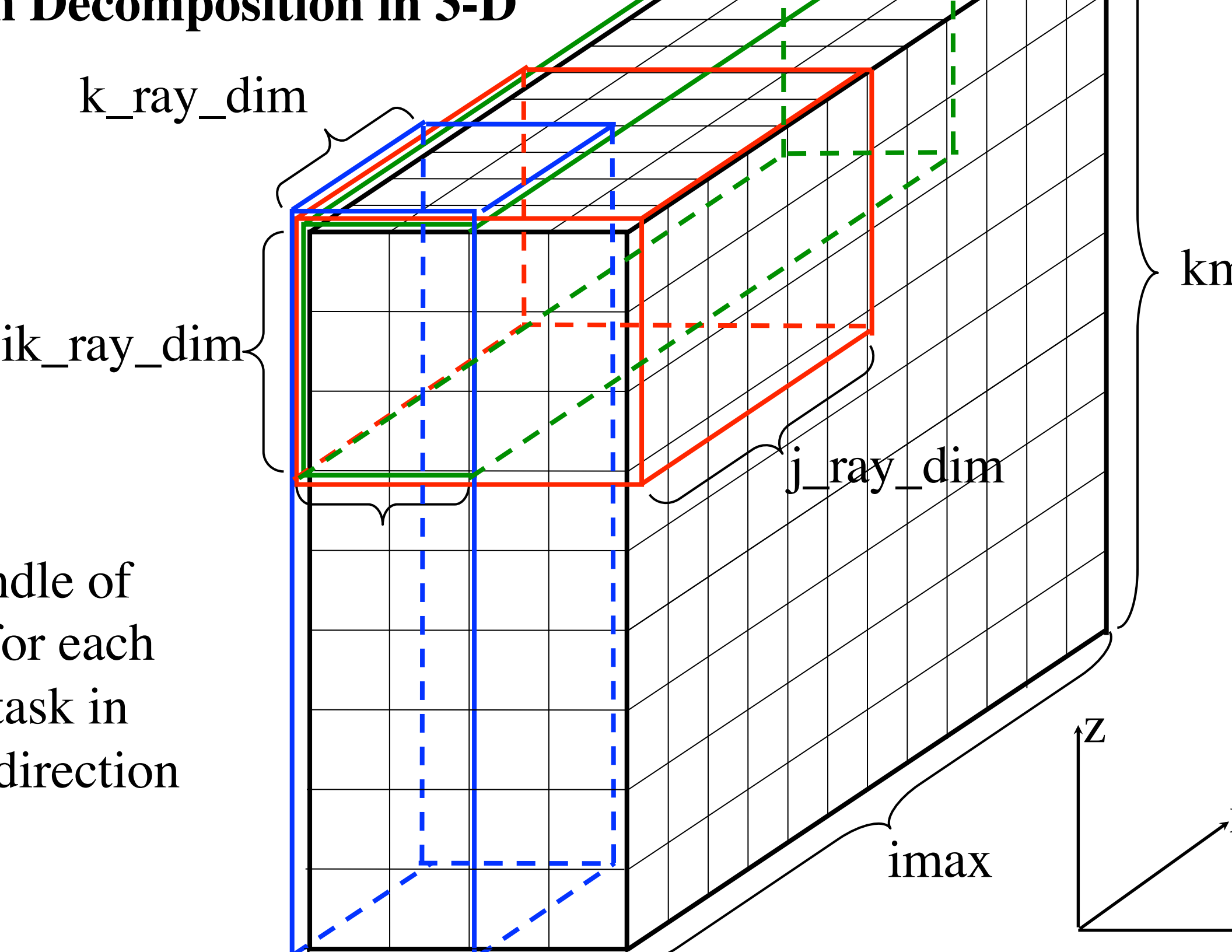
Neutrino transpor
energy from coolin
(blue) to heated
(Red) to energiz

himera domain decompositio

- Dimensional splitting: (x,y,z) or: (r,θ,ϕ)
  - alternating sweeps: XYZ-ZYX, XZY-YZX
- yz-rays: 1x1xN
- undles of several per MPI task
- ransfer data between sweeps (XY and XZ slabs") using MPI_alltoall on sub-communicators
- ransport, nuclear network, I/O performed in adial direction. This provides a natural load alancing

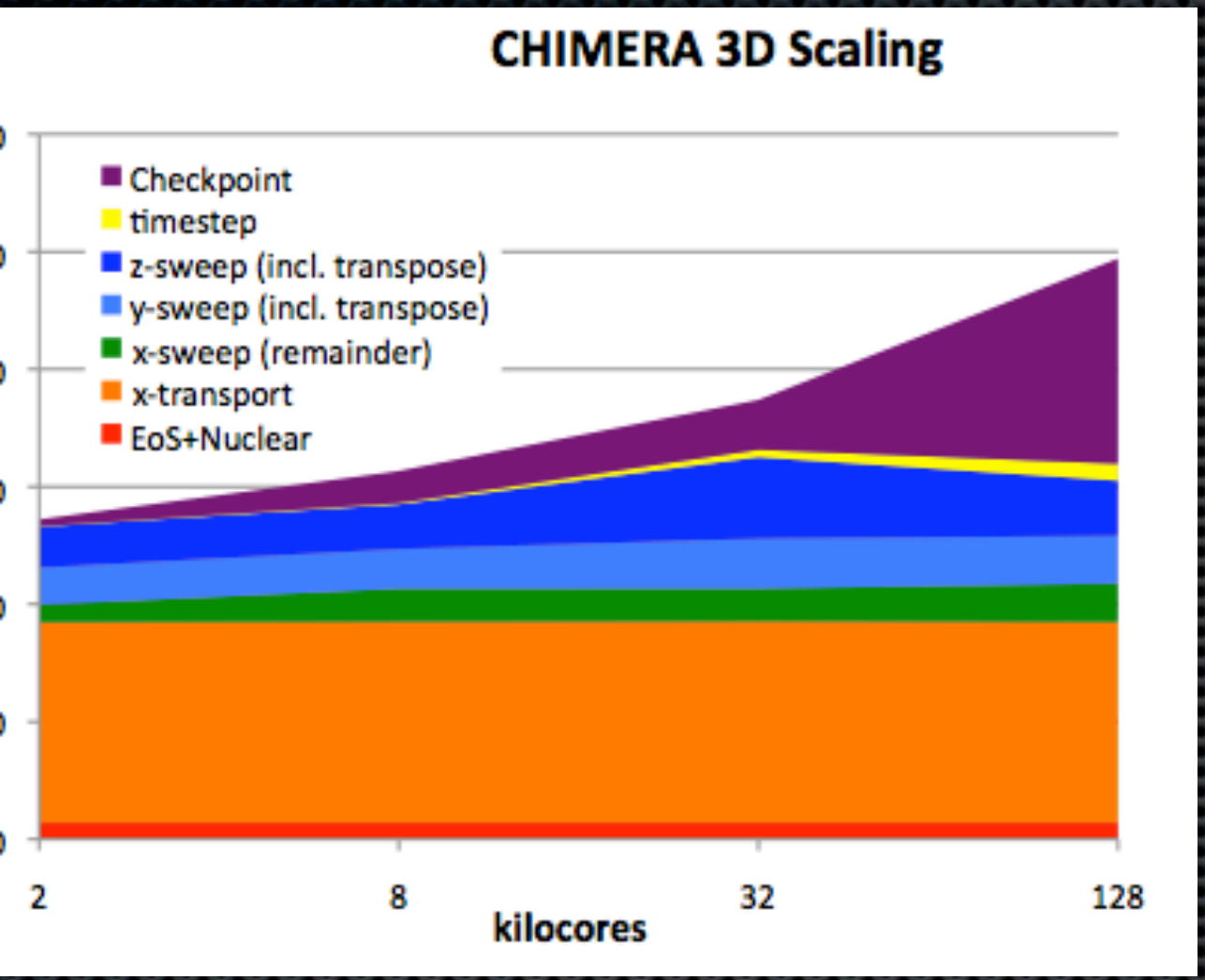


n Decomposition in 3-D



ndle of
for each
task in
direction

Weak Scaling



with increased
resolution

- Jaguar-XT5 ea
access
- 4 checkpoint
outputs, up to
1800 sec
- rest scale nice

proved collectives in MPI/O drop HDF5 time to ~100-200

Time step size goes down with each resolution increase

trong scaling

Loops over rays in each dimension

Current configuration

- 1 “x”-ray, or radial ray per MPI task
- typically 512 radial zones per task (Equivalent to 8
- 1 MPI task per core

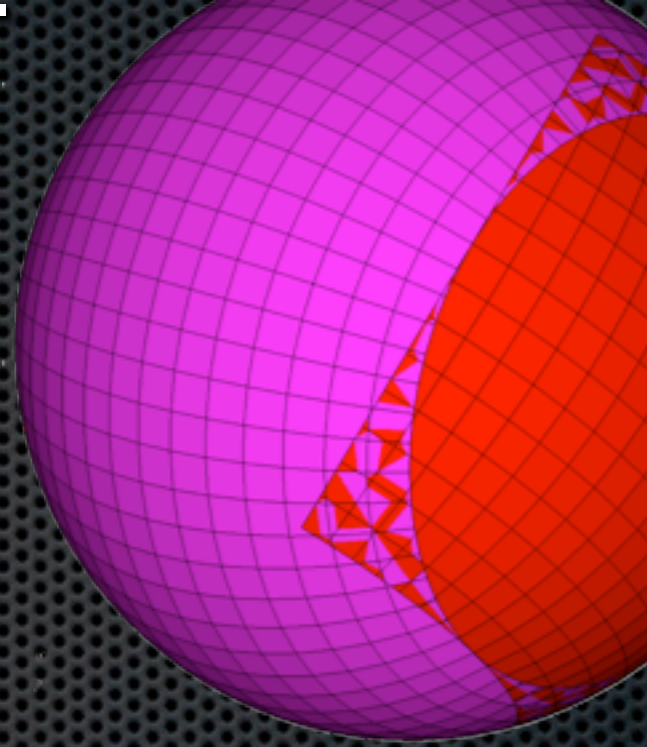
Reached limit of Strong scaling for pure MPI code

Next: more cores than radial rays + GPUs

ow to go faster...

- ewer time steps!
- mprove single thread performance (striding)
- Add OpenMP threading for multiple cores/task (x-ray
- Add GPU acceleration to computationally intense egions (transport, nuclear network)
- More efficient inter-process communication

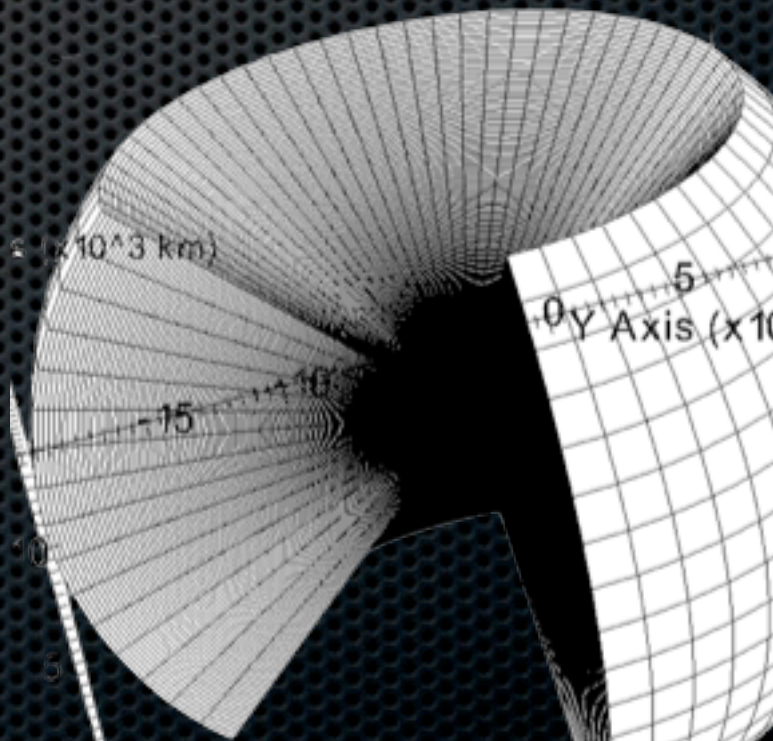
ne-step control = new grid



Yin/Yang or “Baseball” grid

Two spherical grids with poles removed and rotated together.

Eliminates short Courant times at poles and numerical artifacts.



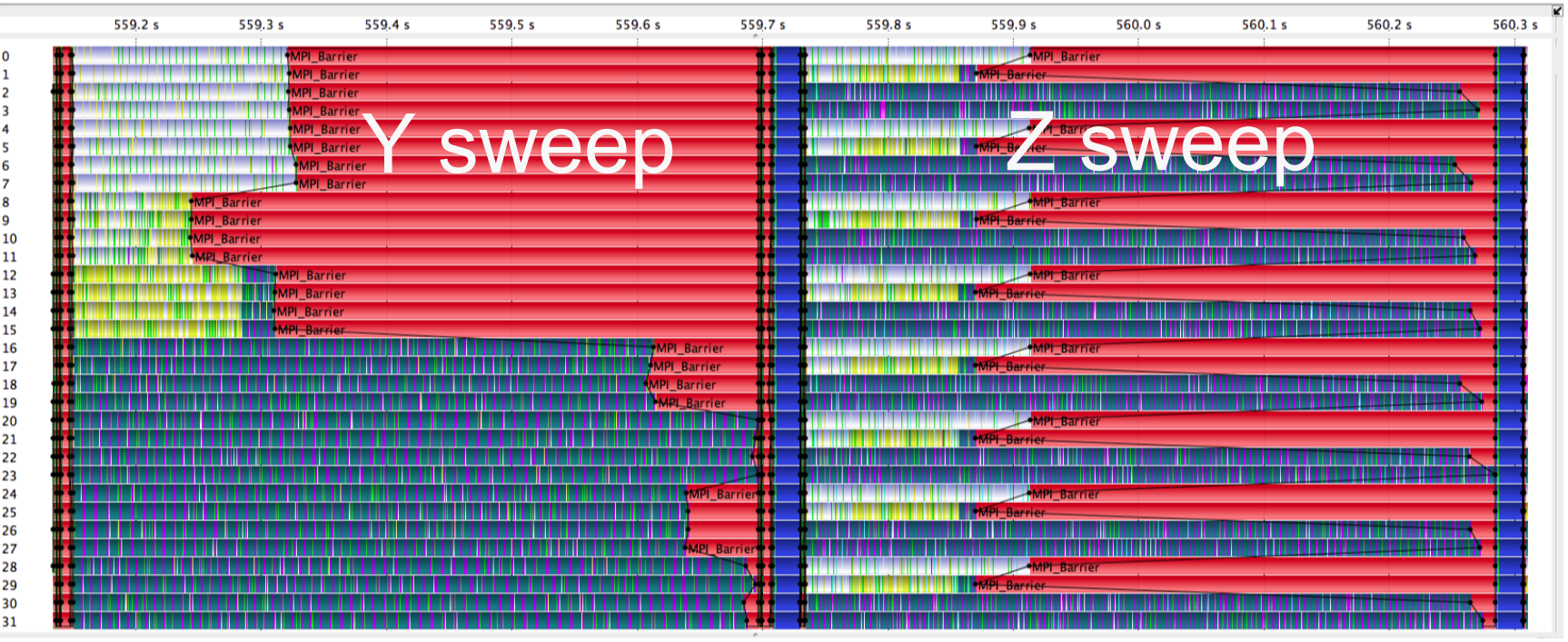
mproving "scalar" performance

- mproving array storage order
- Removing duplicate memory structures
- Pushing loops inside subroutines
- Example: Trimming cost of "lateral" computations
  - Typical problem size: 64(θ)x128(φ)x512(r) zones on 64*128=8192 MPI tasks (cores)
  - y-&z-sweeps each 5% of run time + 15% transpos time. The "low cost" lateral components are 1/4 of run seemingly out of proportion to expectations.

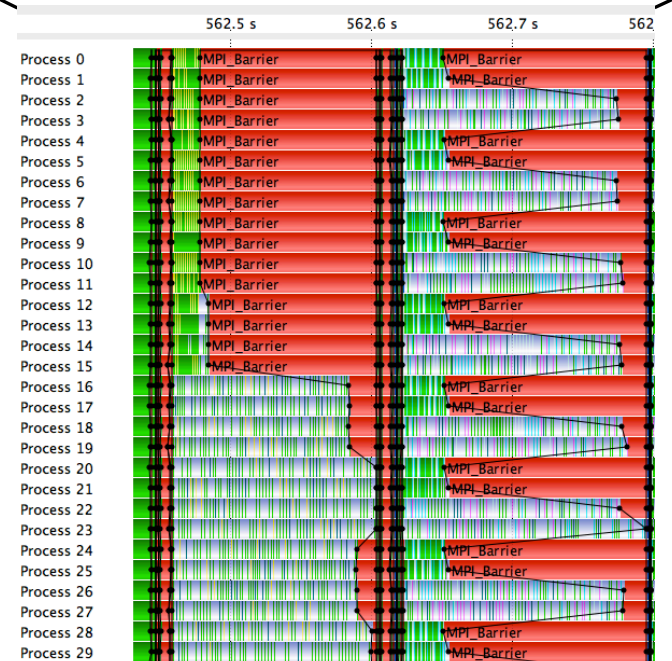
1 XY

befo

red: mp



$\Delta t = 0.36\,\text{sec}$



: 1/3 time
nb. % up
nb. Time down
size problem,
eeps are much
than transpose

Fixes:

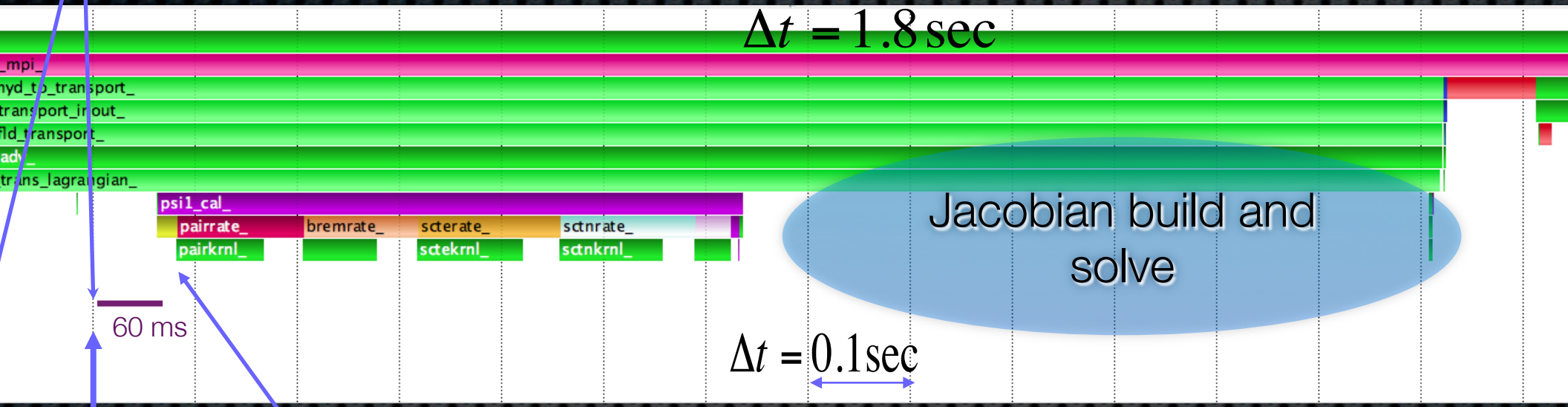
Refactoring subroutines

NSE: replace direct EoS call wi
interpolation

Network region: discovered bro
to skip calls to unneeded part o

terpolation Kernel

GPU Opacity Interpolation



uts

Use outputs

Copy outputs

out most intensive interpolation kernels for 1 or more opacities
inputs in asynchronously when updating table
olation can begin after temperature updated by nuclear burnin
cores do other preparation work while GPU interpolates
builds and solves Jacobian. GPU can start on next ray.

Net: Nuclear Network compone

outer regions, nuclei not in equilibrium–Solve nuclear
eaction network

Current usage: 14-species “alpha” network ~5% run

Newton-Raphson solution, Jacobian build and solve

Scales N^{2-3} with network size

Future: 150-species or more network

Est. 200-300% time cost of current run w/o accel.
numerical or physical)

OpenMP threaded, GPU work underway...

XNet: CULA Tools and the GPU

CULA is a GPU accelerated linear algebra library that works on CUDA enabled NVIDIA GPUs.

Host interface is very simple if you have the right setup.

A slow PCI bus offsets the speed of the GPU.

Table Solver Speed for CULA.

matrix order	LAPACK matrix time (s)	LAPACK matrix time (%)	LAPACK Step count	CULA LAPACK matrix time (s)	CULA LAPACK matrix time (%)	CULA LAPACK Step count
150	2.82	76	953	25.0	79	953
300	11.7	89	557	25.3	91	557
1072	724	87	664	224	67	668
2184	2837	92	547	500	66	547

penMP strategies

- n- or low-level threads
- ach sweep there is a loop over rays, but...
- omputational intensive radial orientation want $N_{cores} > N_{rays}$
- adial rays have lots of work internally to thread by zone, or tas
- Code is not yet thread-safe
- eading individual radial rays
- Network thread by zones
- hreading building of opacity table, interpolation

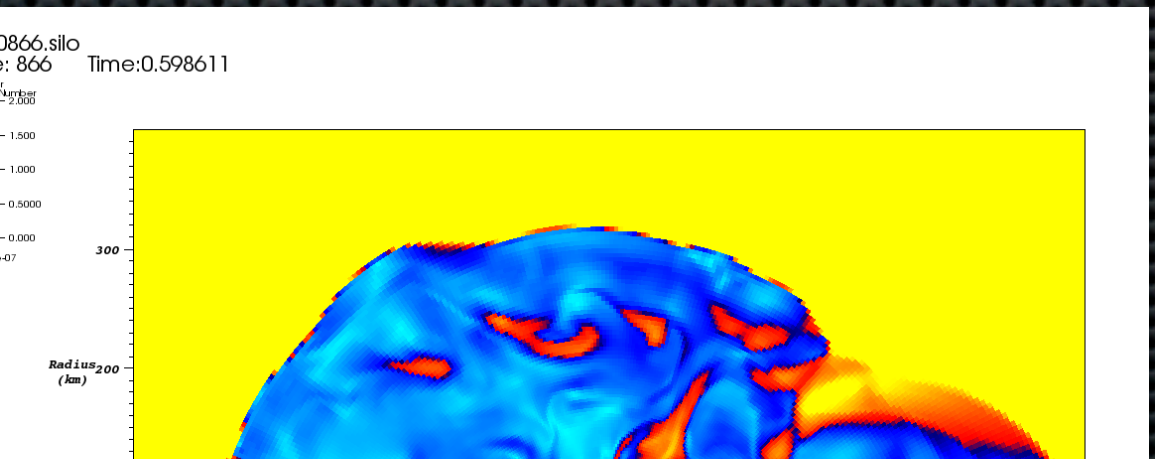
I/O - Analysis - Visualization

Checkpoint files written every 10-20 min, ~10-15 seconds each
or restart, analysis, and visualization

Automated plotting with Bellerophon automation tool

Checkpoint files are ~1% of memory footprint. More efficient to
move analysis offline.

Bellerophon also performs nightly automated code compilation
and testing. Future usage could include offline analysis, archivin
V&V, and performance testing.



himera on Titan

Current work on new grid and scalar performance
hould benefit immediately with transition to Bulldoze
- Gemini configuration of 2012 transition system.

Adding OpenMP + GPU acceleration will allow Chim
o take full advantage of final XK6 system with bigger
imulations that finish sooner.

Financial support: DOE, NASA, NSF
Computational support: OLCF, NICS, UT