Accelerated Astrophysics:

Using NVIDIA GPUs to Simulate and Understand the Universe

Prof. Brant Robertson
Department of Astronomy and Astrophysics
University of California, Santa Cruz
brant@ucsc.edu, @brant_robertson

UC Santa Cruz: a world-leading center for astrophysics

- Home to one of the largest computational astrophysics groups in the world.
- Home to the University of California Observatories.
- World-wide top 5 graduate program for astronomy and astrophysics according to US News and World Report.
- Many PhD students in our program interested in professional data science.
- http://www.astro.ucsc.edu

#1 California Institute of Technology
United States Pasadena, CA
#5 – Best Global Universities

#2 Harvard University
United States Cambridge, MA
#1 – Best Global Universities

#3 University of California--Berkeley
United States Berkeley, CA
#4 – Best Global Universities

#4 University of California--Santa Cruz
United States Santa Cruz, CA
#27 (tied) – Best Global Universities

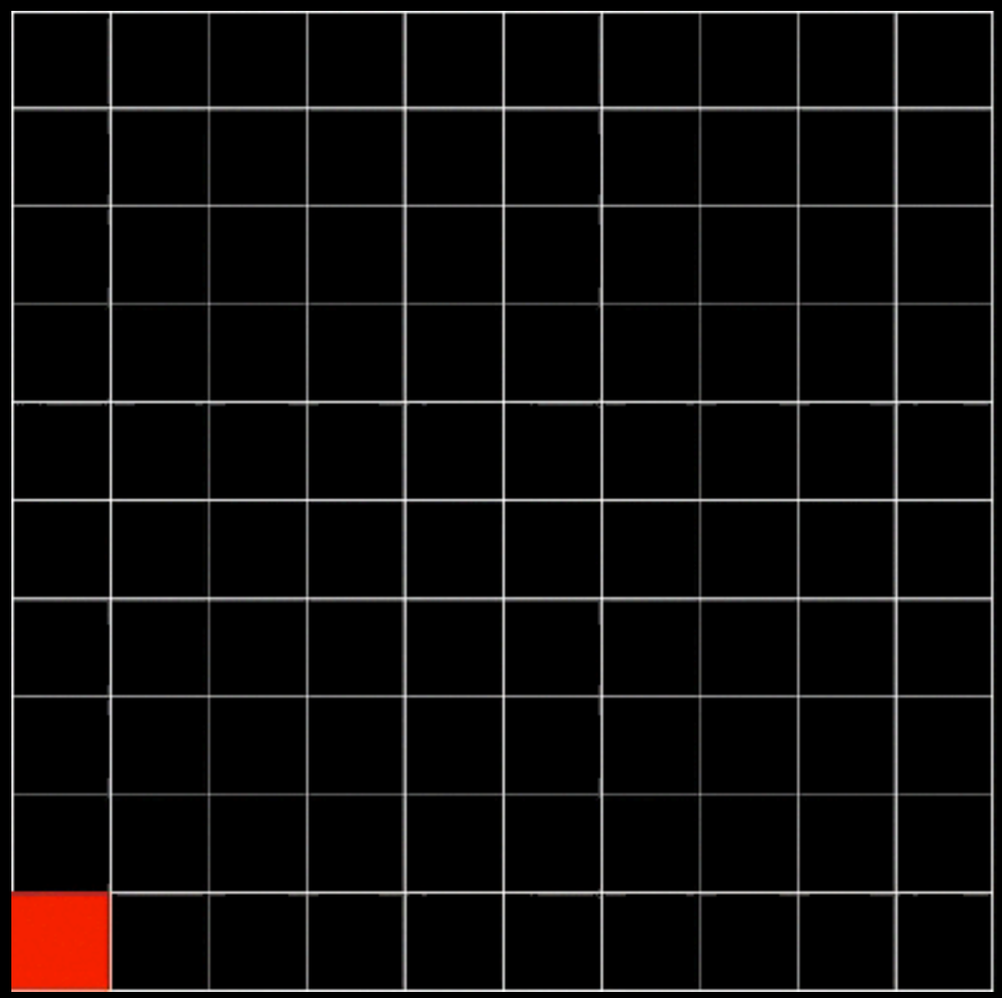
#5 Princeton University
United States Princeton, NJ
#8 – Best Global Universities

#6 University of Cambridge

https://www.usnews.com/education/best-global-universities/space-science

GPUs as a scientific tool

Grid code on a CPU



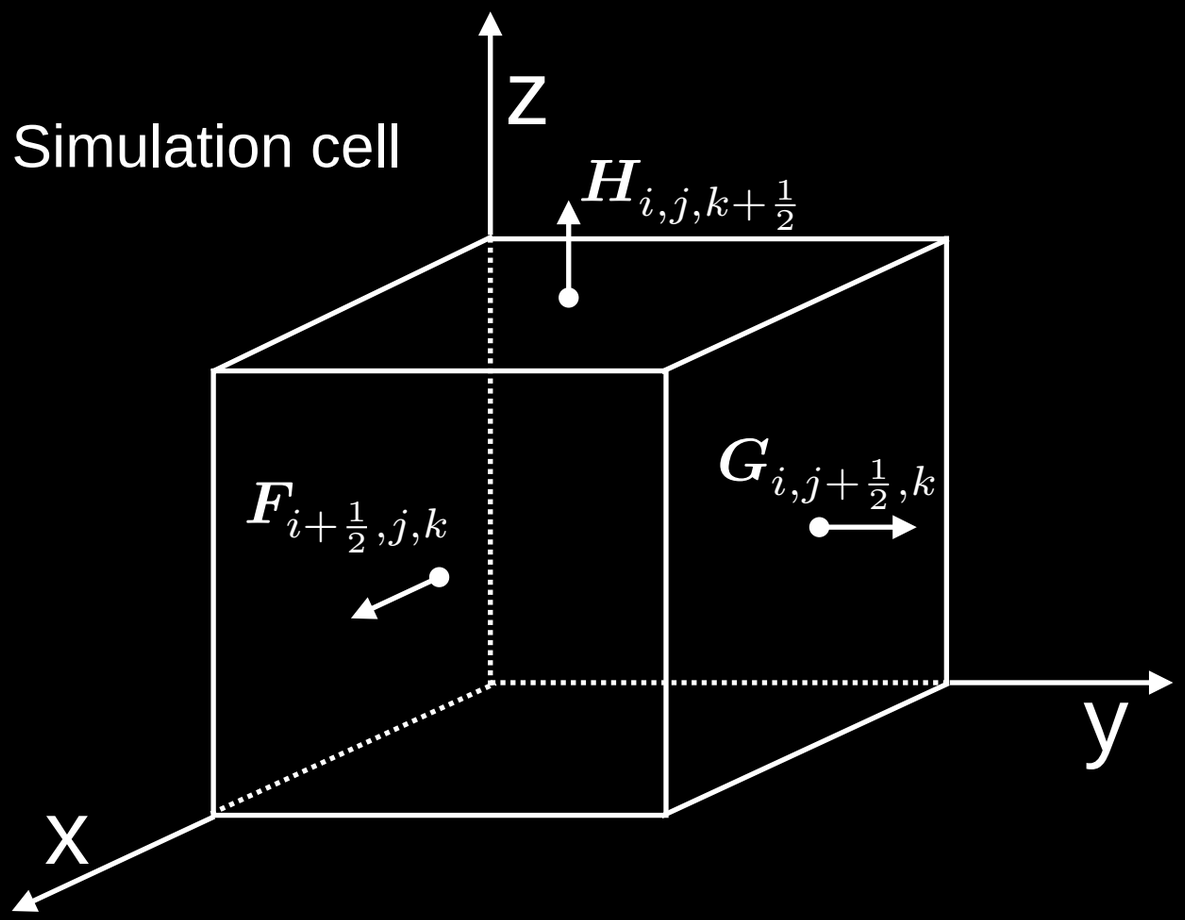
A (brief) intro to finite volume methods

conserved quantity at time n+1

conserved quantity at time n

$$\boldsymbol{u}_{i,j,k}^{n+1} = \boldsymbol{u}_{i,j,k}^{n} - \frac{\delta t}{\delta x}\left(\boldsymbol{F}_{i-\frac{1}{2},j,k}^{n+\frac{1}{2}} - \boldsymbol{F}_{i+\frac{1}{2},j,k}^{n+\frac{1}{2}}\right) - \frac{\delta t}{\delta y}\left(\boldsymbol{G}_{i,j-\frac{1}{2},k}^{n+\frac{1}{2}} - \boldsymbol{G}_{i,j+\frac{1}{2},k}^{n+\frac{1}{2}}\right) - \frac{\delta t}{\delta z}\left(\boldsymbol{H}_{i,j,k-\frac{1}{2}}^{n+\frac{1}{2}} - \boldsymbol{H}_{i,j,k+\frac{1}{2}}^{n+\frac{1}{2}}\right)$$

“fluxes” of conserved quantities across each cell face



Conserved variable update in standard C

```
for (i=0; i<nx; i++) {
  density[i]    += dt/dx * (F.d[i-1]  - F.d[i]);
  momentum_x[i] += dt/dx * (F.mx[i-1] - F.mx[i]);
  momentum_y[i] += dt/dx * (F.my[i-1] - F.my[i]);
  momentum_z[i] += dt/dx * (F.mz[i-1] - F.mz[i]);
  Energy[i]     += dt/dx * (F.E[i-1]  - F.E[i]);
}
```

Simple loop; potential for loop parallelization, vectorization.

Conserved variable update using CUDA

```
// copy the conserved variable array onto the GPU
cudaMemcpy(dev_conserved, host_conserved, 5*n_cells*sizeof(Real),
           cudaMemcpyHostToDevice);


// call cuda kernel
Update_Conserved_Variables<<<dimGrid,dimBlock>>>(dev_conserved, F_x, nx, dx, dt);


// copy the conserved variable array back to the CPU
cudaMemcpy(host_conserved, dev_conserved, 5*n_cells*sizeof(Real),
           cudaMemcpyDeviceToHost);
```

Memory transfer, CUDA kernel, memory transfer…

Conserved variable update CUDA kernel

```
void Update_Conserved_Variables(Real *dev_conserved, Real *dev_F, int nx, Real dx,
Real dt)
{
  // get a global thread ID
  id = threadIdx.x + blockIdx.x * blockDim.x;

  // update the conserved variable array
  if (id < nx) {
    dev_conserved[       id] += dt/dx * (dev_F[       id-1] - dev_F[       id]);
    dev_conserved[  nx + id] += dt/dx * (dev_F[  nx + id-1] - dev_F[  nx + id]);
    dev_conserved[2*nx + id] += dt/dx * (dev_F[2*nx + id-1] - dev_F[2*nx + id]);
    dev_conserved[3*nx + id] += dt/dx * (dev_F[3*nx + id-1] - dev_F[3*nx + id]);
    dev_conserved[4*nx + id] += dt/dx * (dev_F[4*nx + id-1] - dev_F[4*nx + id]);
  }
}
```

Mapping between CUDA thread and simulation cell;
memory coalescence for transfer efficiency.

Cholla: Computational hydrodynamics on ll (parallel) architectures

Cholla are also a group of cactus species that grows in the Sonoran Desert of southern Arizona.

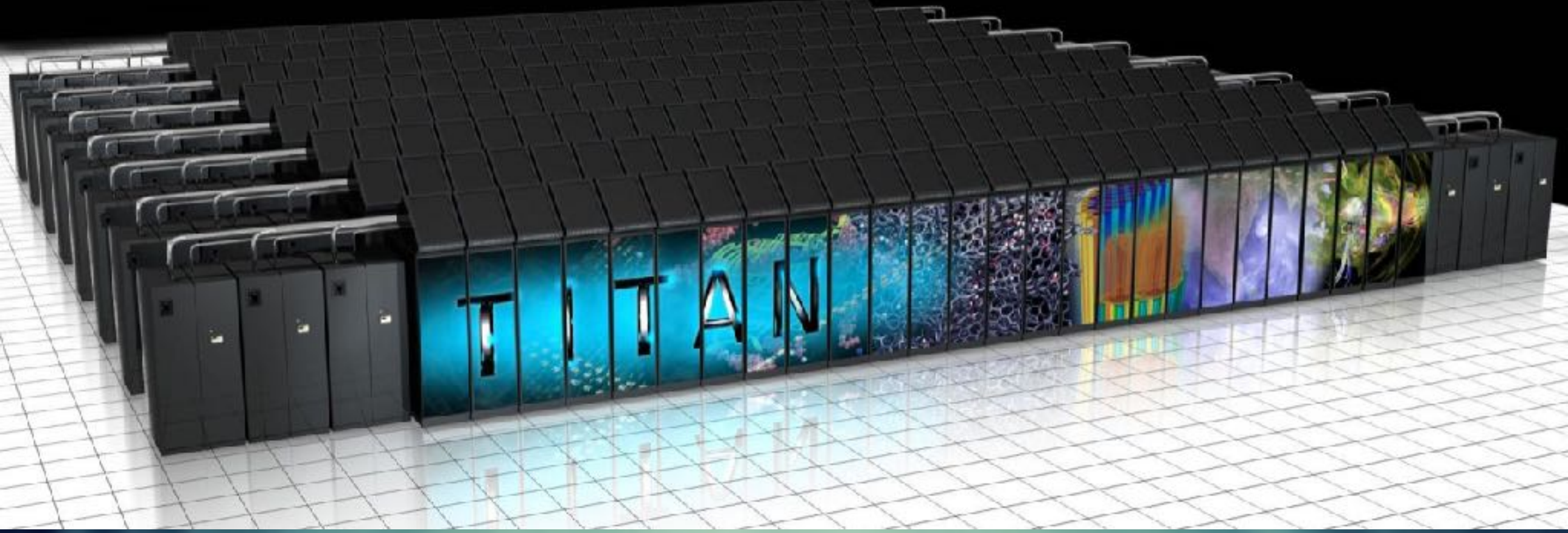
- A GPU-native, massively-parallel, grid-based hydrodynamics code written by Evan Schneider for her PhD thesis.
- Incorporates state-of-the-art hydrodynamics algorithms (unsplit integrators, 3rd order spatial reconstruction, precise Riemann solvers, dual energy formulation, etc).
- Includes GPU-accelerated radiative cooling and photoionization.
- github.com/cholla-hydro/cholla

Schneider & Robertson (2015)

Cholla leverages the world's most powerful supercomputers

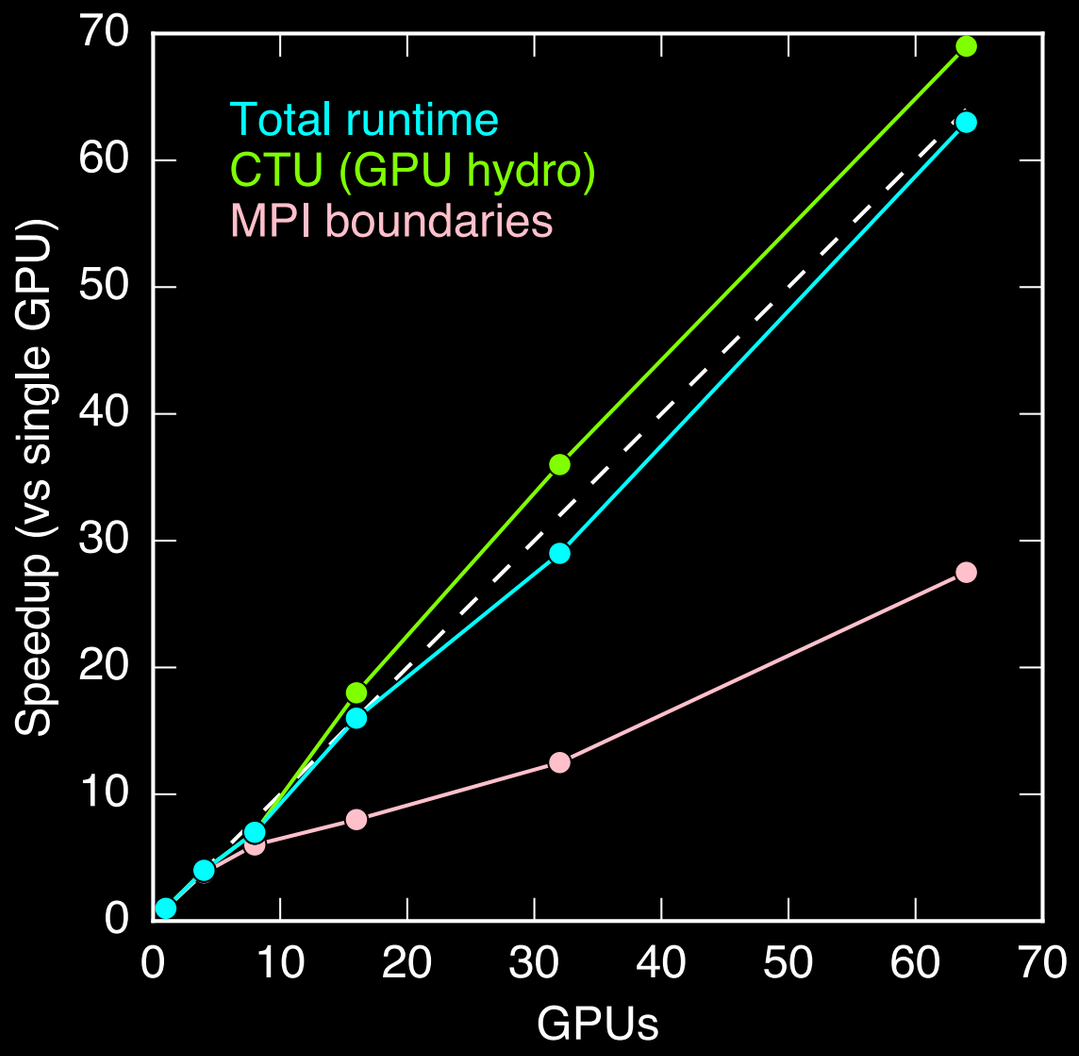
Titan: Oak Ridge Leadership Computing Facility

Flagship accelerated computing system | 200-cabinet Cray XK7 supercomputer |
18,688 nodes (AMD 16-core Opteron + NVIDIA Tesla K20 GPU) |
CPUs/GPUs working together – GPU accelerates | 20+ Petaflops

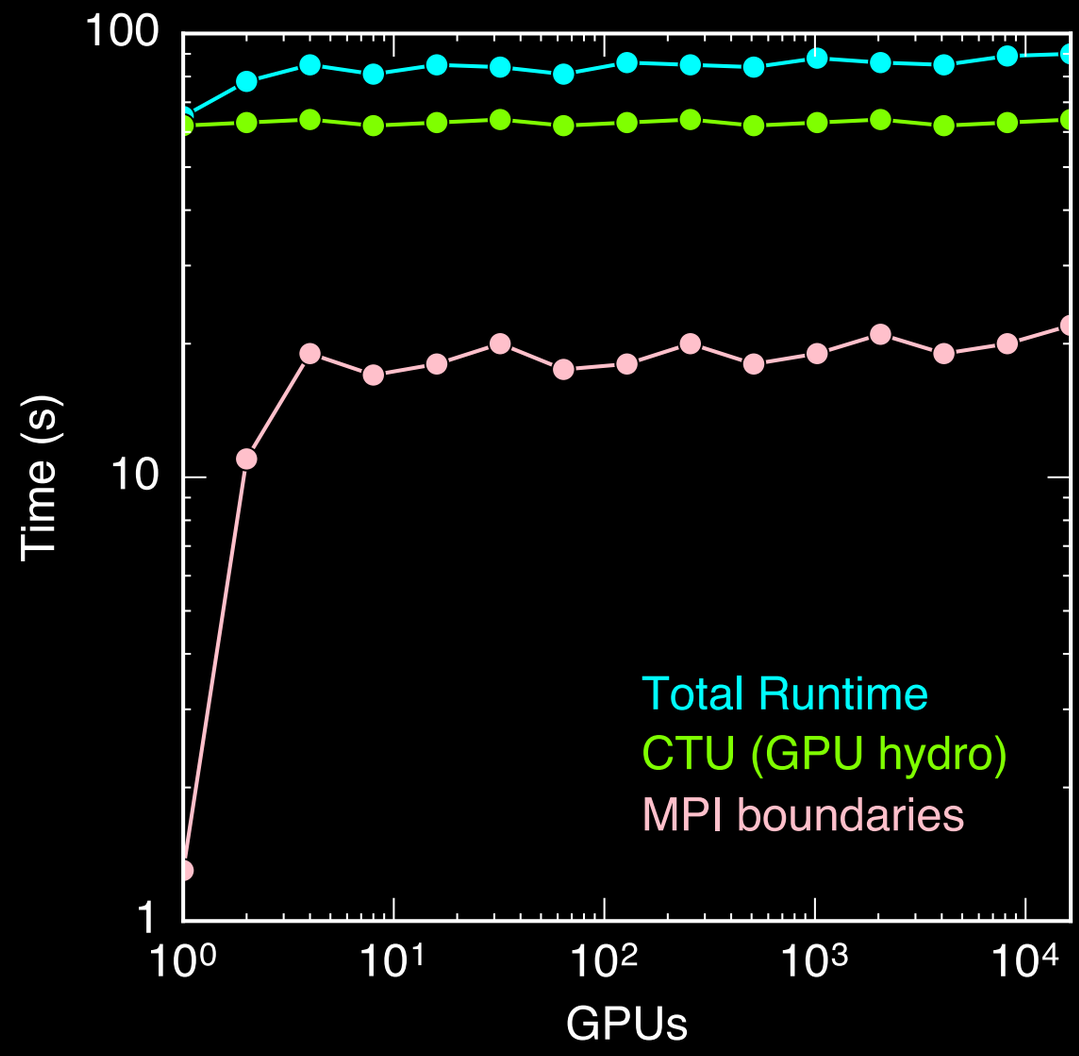


Cholla achieves excellent scaling to >16,000 NVIDIA GPUs

Strong Scaling test, 512^3 cells



Weak Scaling test, ~322^3 cells / GPU



Tests performed on ORNL Titan (AST 109, 115, 125).

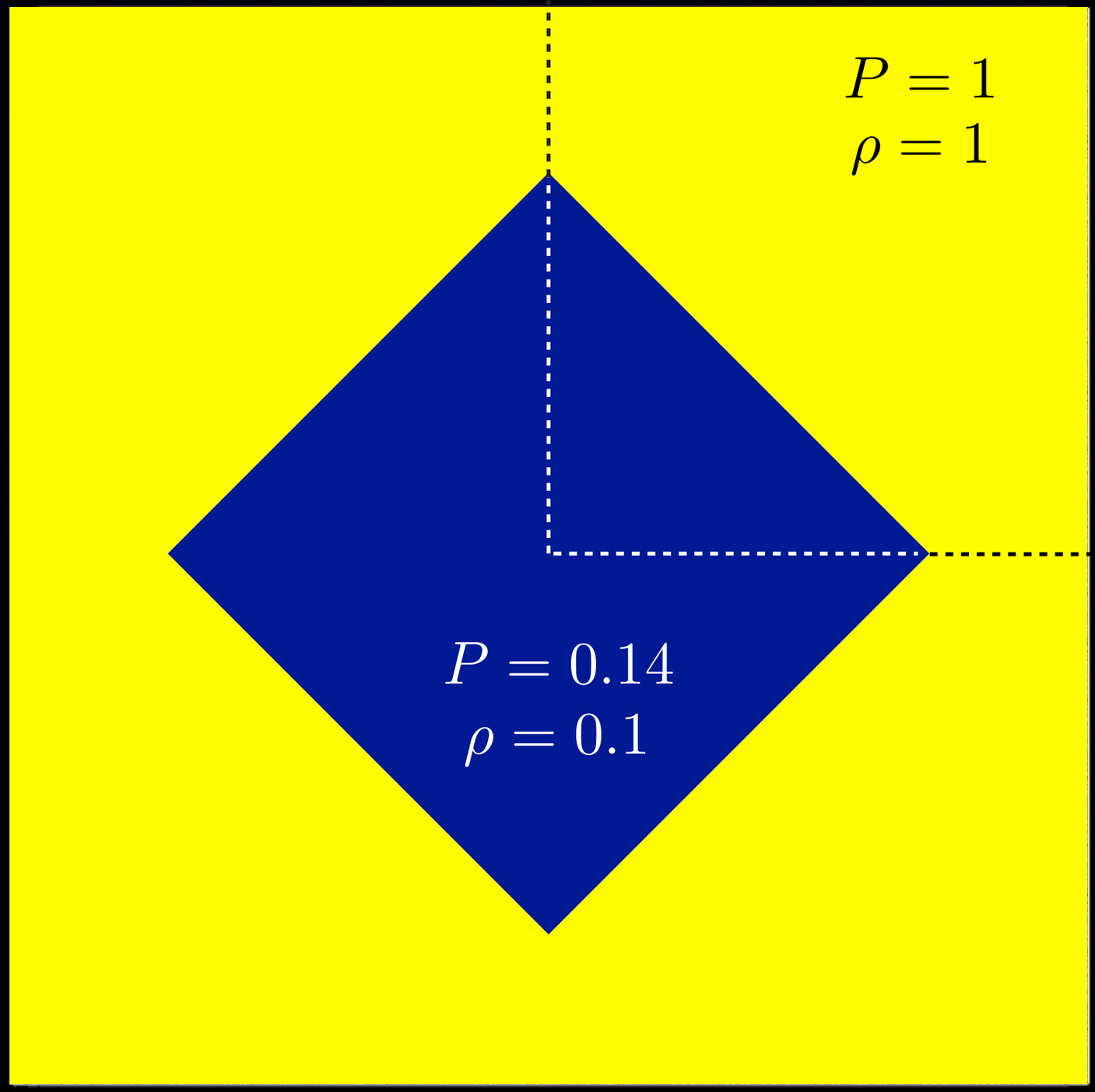
Schneider & Robertson (2015, 2017)

2D implosion test with *Cholla* on NVIDIA GPUs

Example test calculation: implosion (1024^2)

55,804,166,144 cell updates

symmetric about y=x to roundoff error



Application: modeling galactic outflows

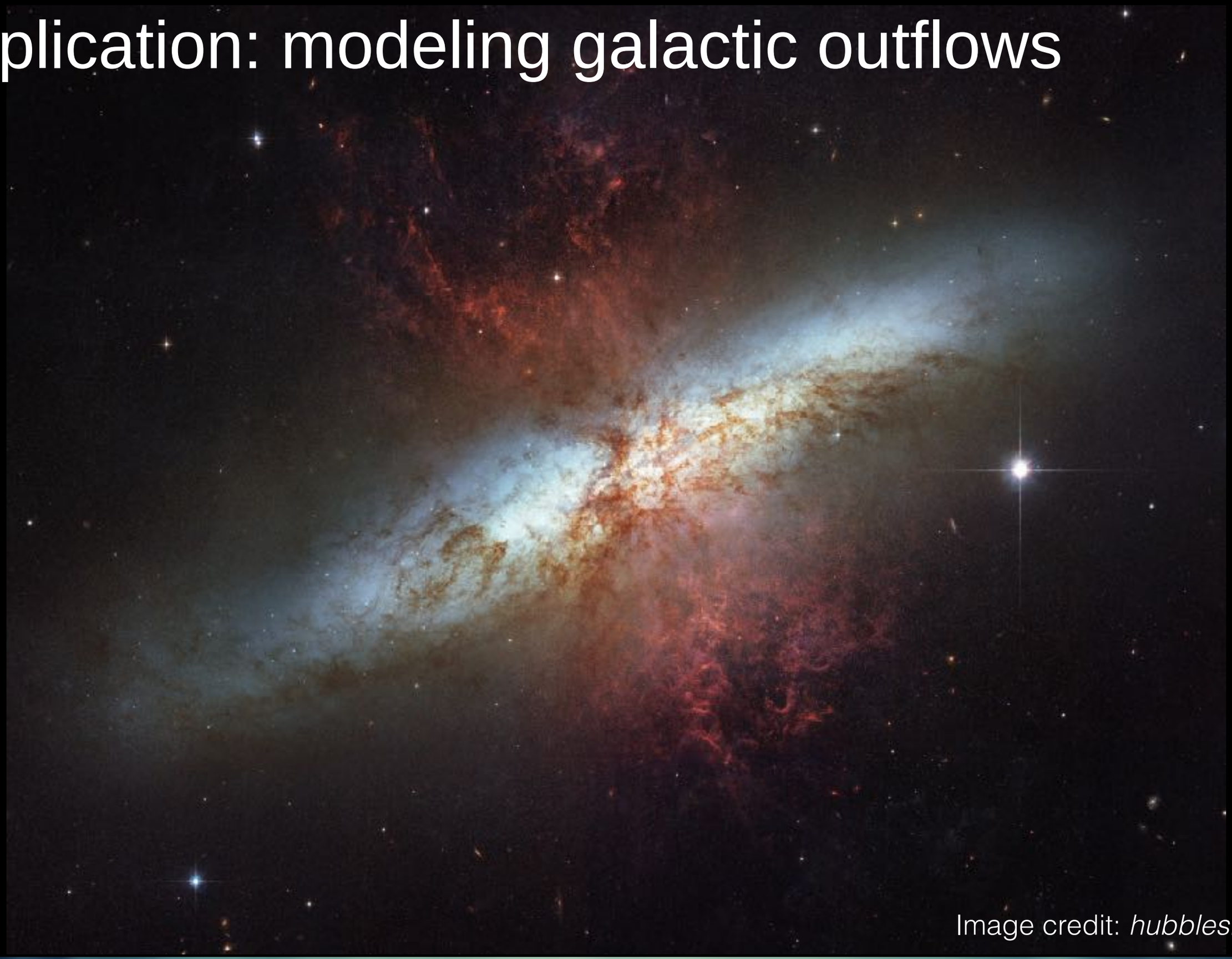
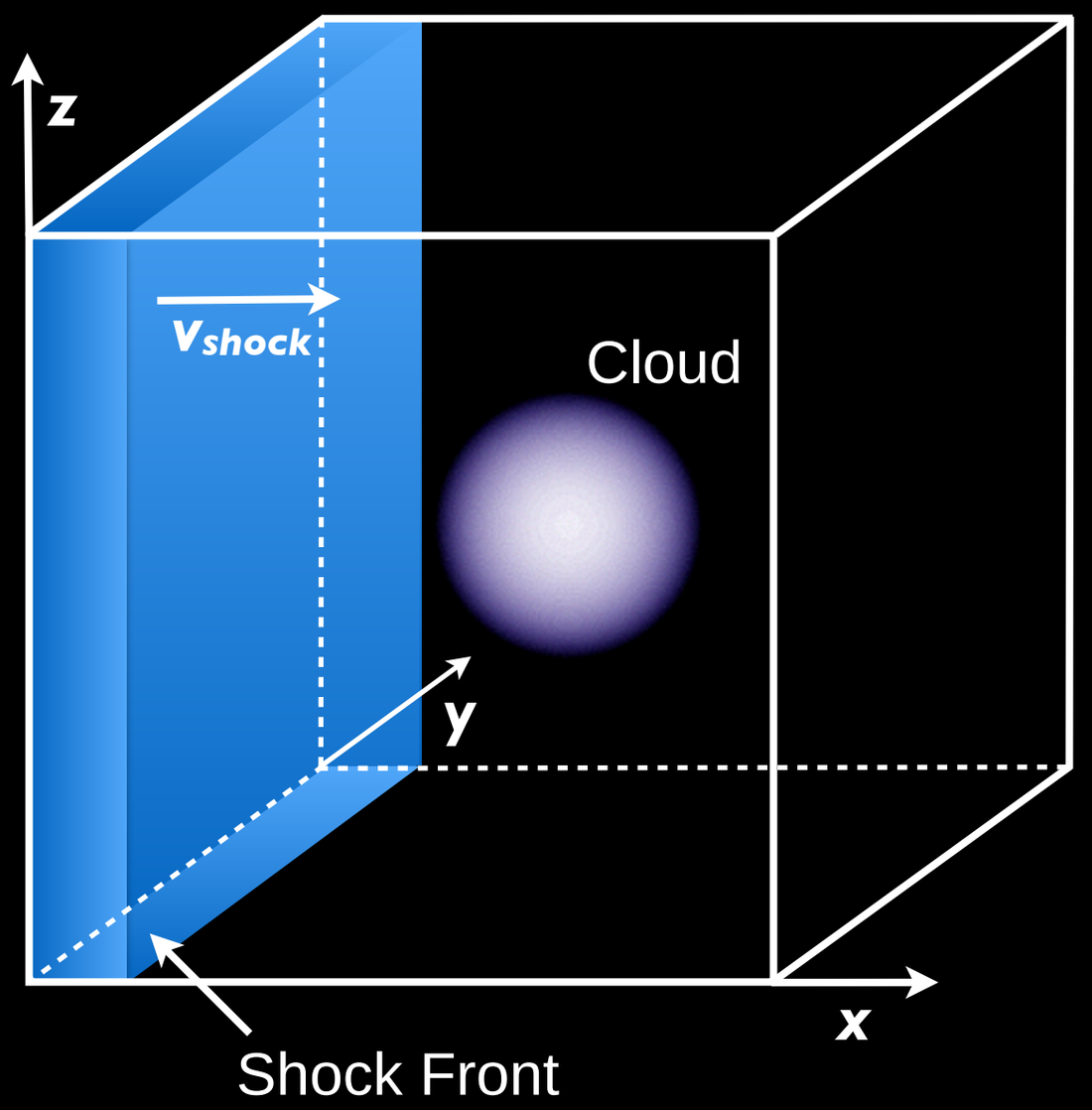


Image credit: *hubblesite.org*

Cholla can simulate the structure of galactic winds

Important questions:

- How does mass and momentum become entrained in galactic winds?
- How does the detailed structure of galactic winds arise?



Cholla + NVIDIA GPUs form a unique tool simulating astrophysical fluids.

Cholla can simulate the structure of galactic winds

Schneider, E. & Robertson, B. 2017, ApJ, 834, 144

1.25e9 cells, 512 NVIDIA K20X GPUs on ORNL *Titan*

Leveraging the NVIDIA DGX-1 for astrophysical research

NVIDIA DGX-1

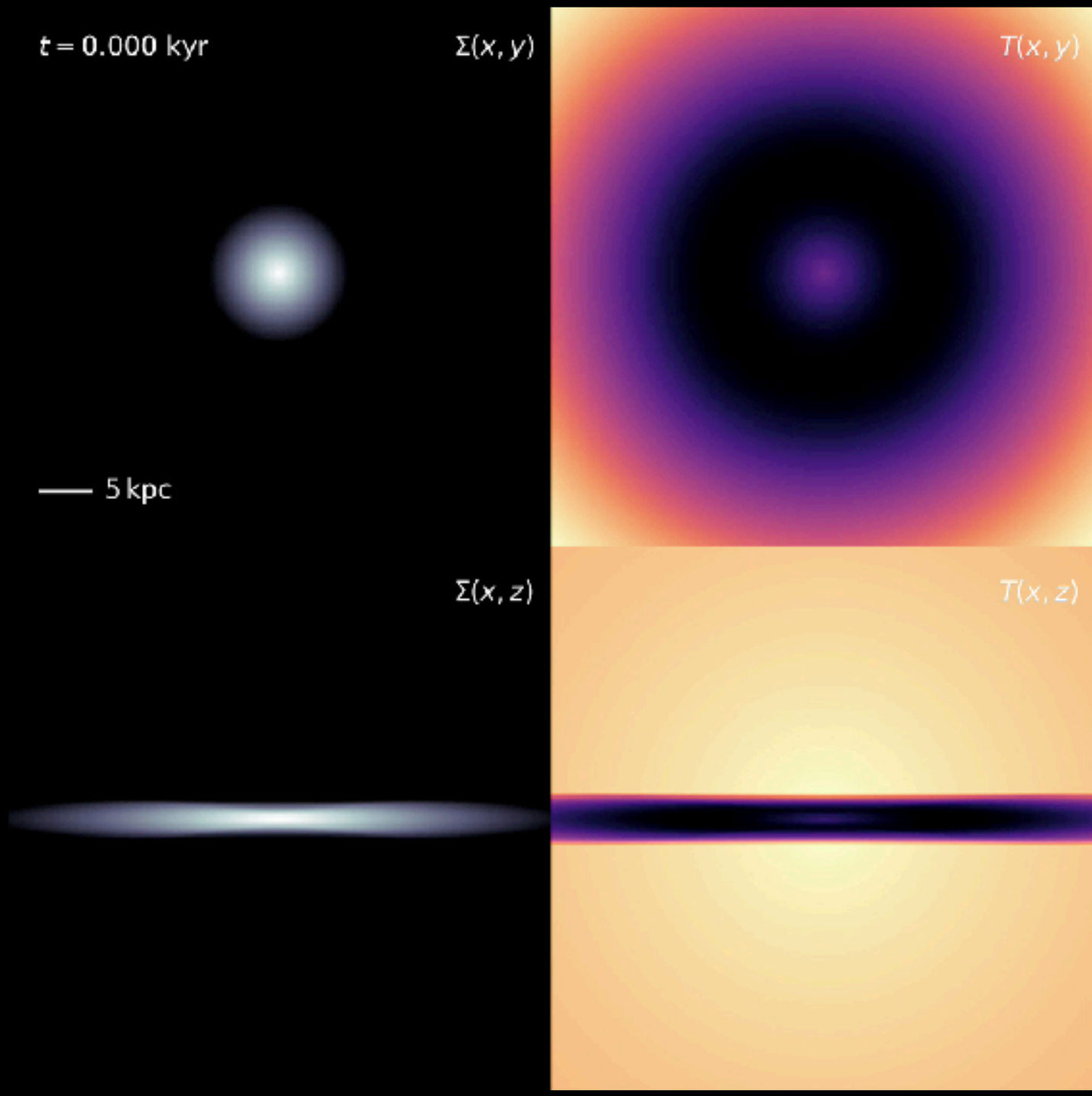
2x 20-core Intel E5-2698 v4 CPUs,
8x NVIDIA P100 GPUs, 768 GB/s Bandwidth,
4x Mellanox EDR Infiniband NICs

- Unlike risk-adverse mission-critical astronomical software, pipeline and high-level analysis software can leverage new and emerging technologies.
- Utilize investments in software from Silicon Valley, data science, other industries.
- UCSC Astrophysicists use the NVIDIA DGX-1 for astrophysical simulation and astronomical data analysis.

Accelerated simulations of disk galaxies

- The UCSC Astrophysics DGX-1 system is our development platform for constructing complex initial conditions.

- The DGX-1 system is powerful enough to perform high-quality *Cholla* simulations of disk galaxies.

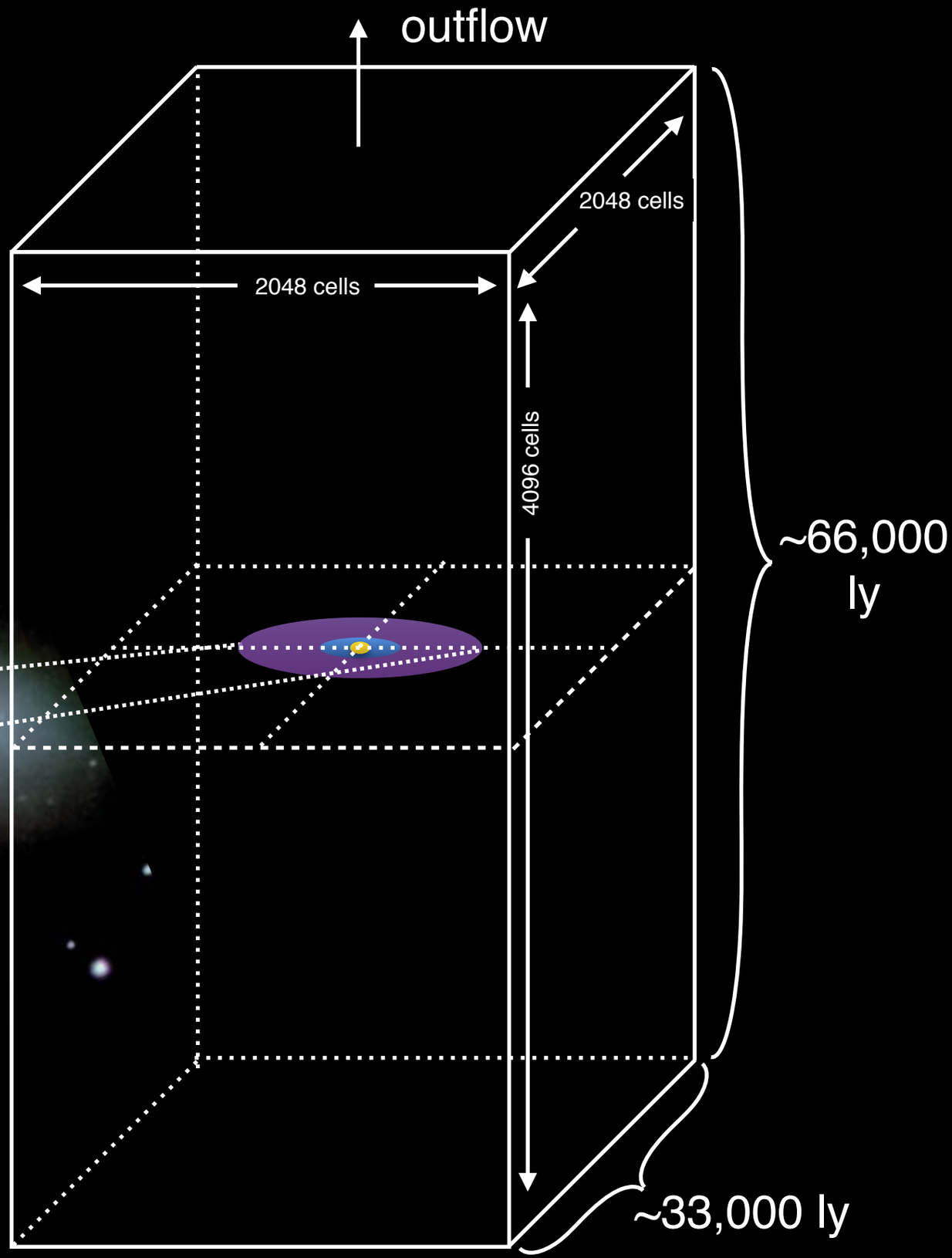


256^3, single P100, 2hrs

Cholla + Titan global simulations of galactic outflows

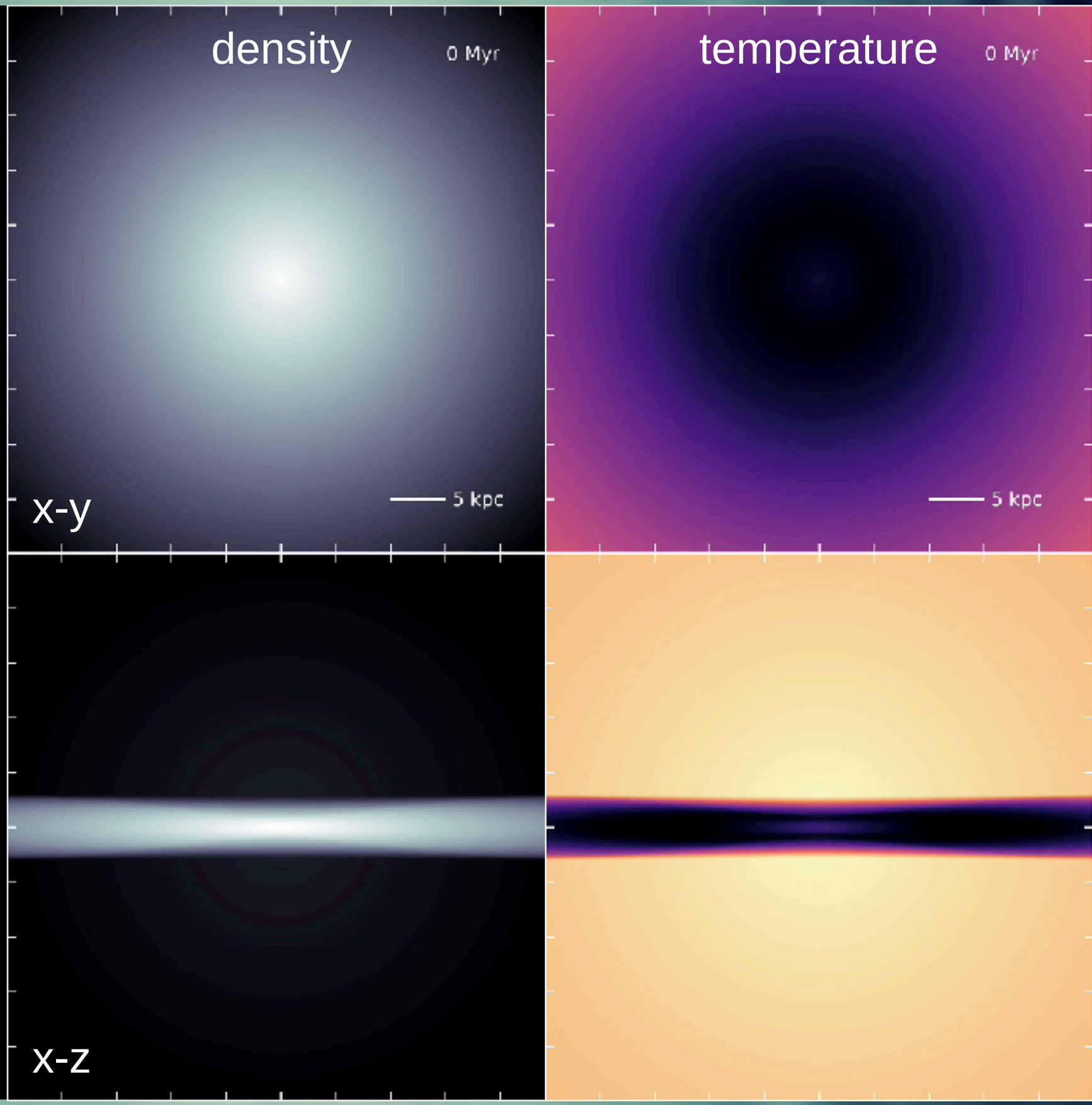
Cholla simulations of M82 initial conditions

gain region

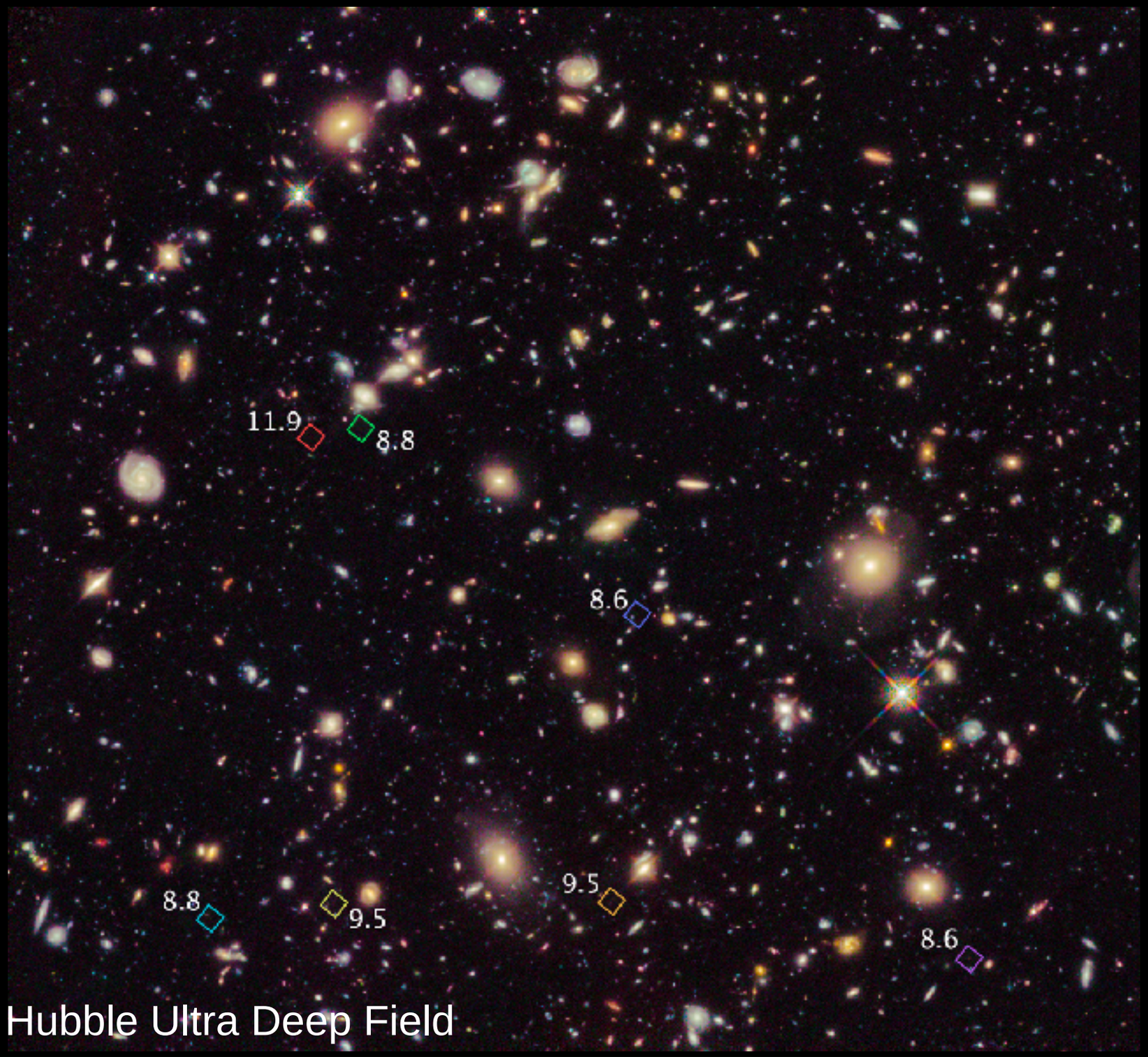


Cholla + ORNL Titan global simulations of galactic outflows

- Test calculation on Titan - 1024^3, largest hydro simulation of a single galaxy ever performed.
- 512 K20X GPUs, 6hours, ~90K core hours
- ~47M core hour allocation (AST-125)



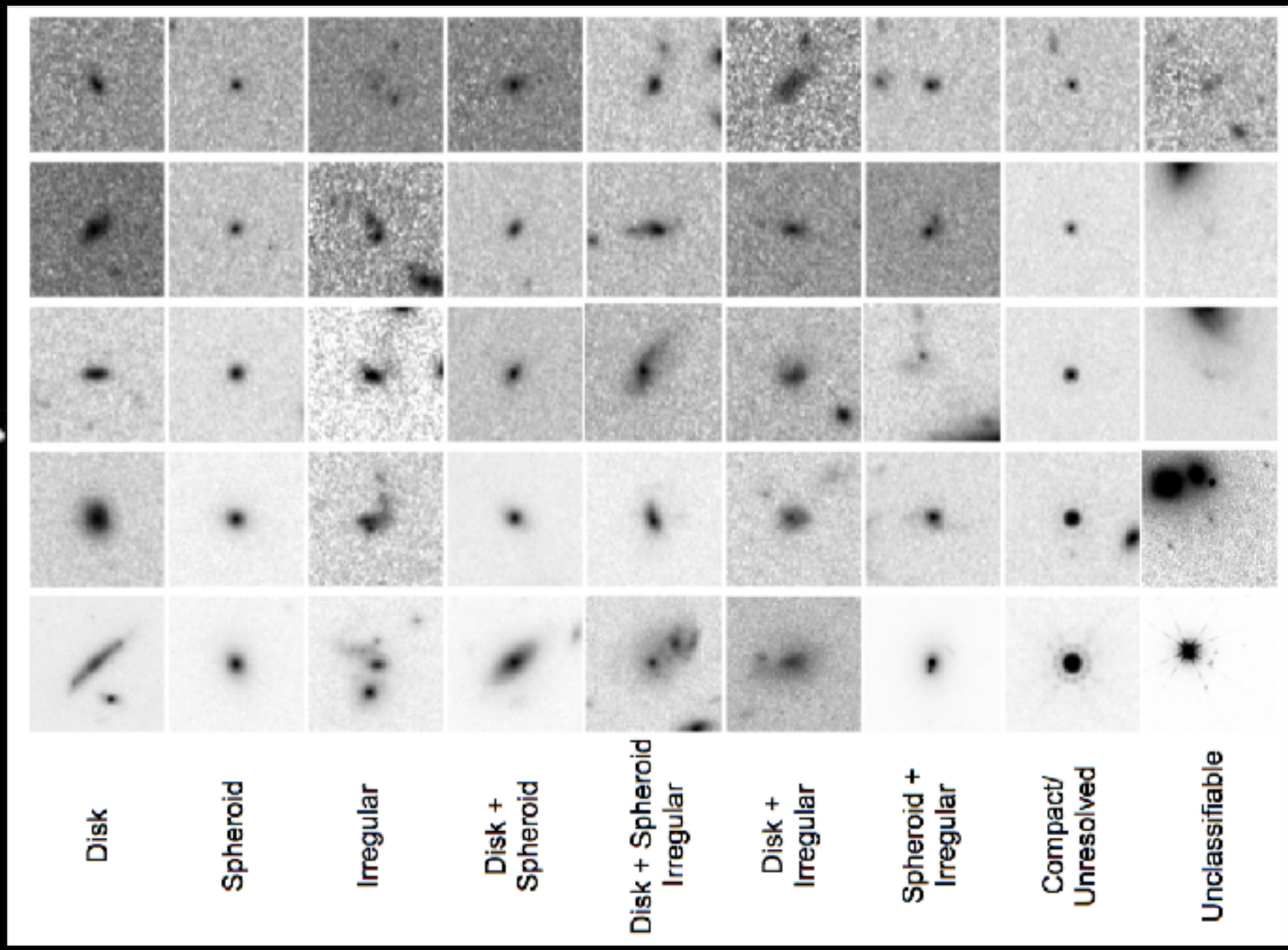
Using NVIDIA GPUs for astronomical data analysis



Hubble Ultra Deep Field

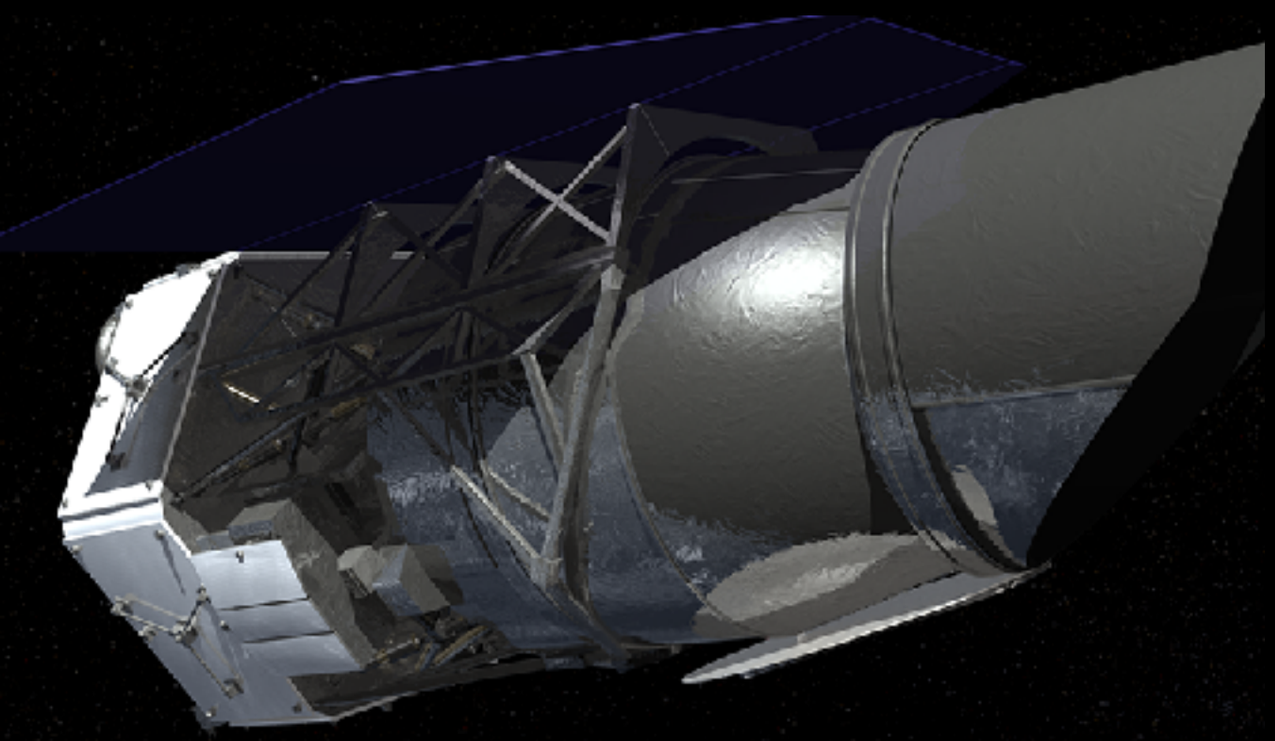
Human galaxy classification….

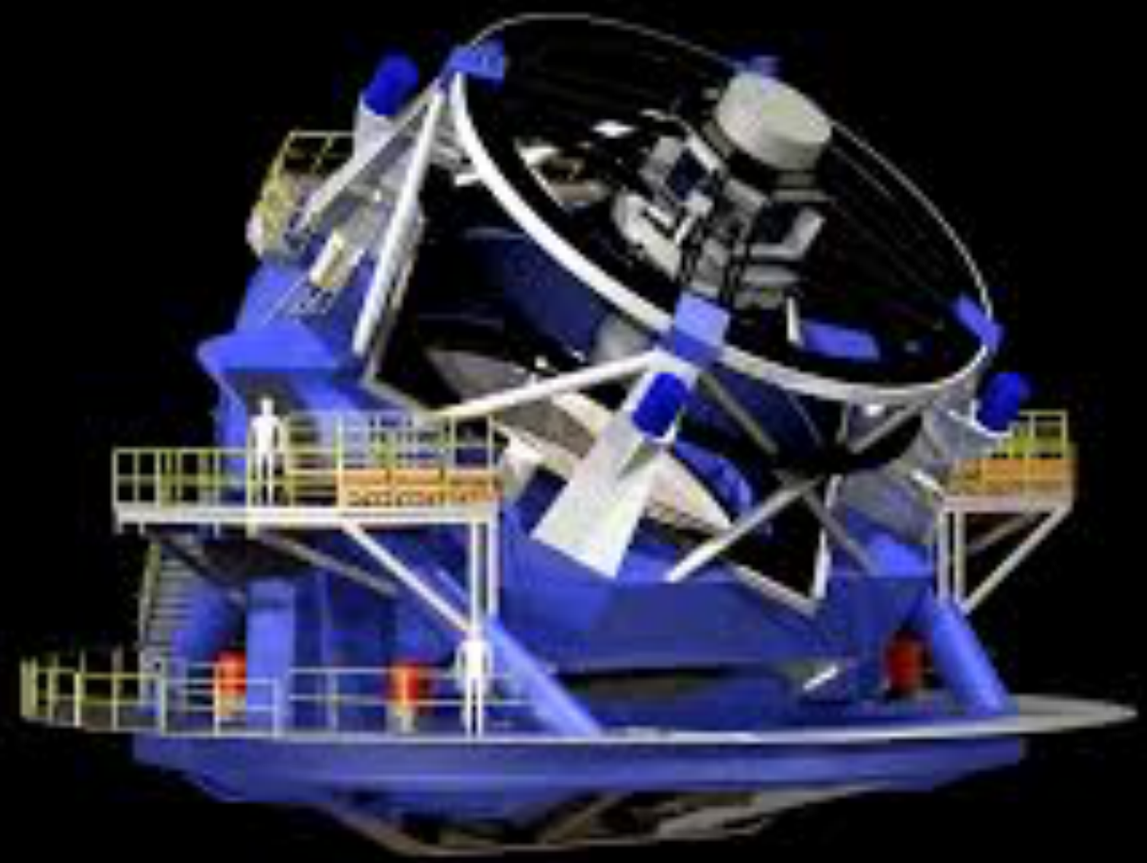
Expert classifications of Hubble images from the CANDELS survey.



Kartaltepe et al., ApJS, 221, 11 (2015)

Human galaxy classification does not scale.

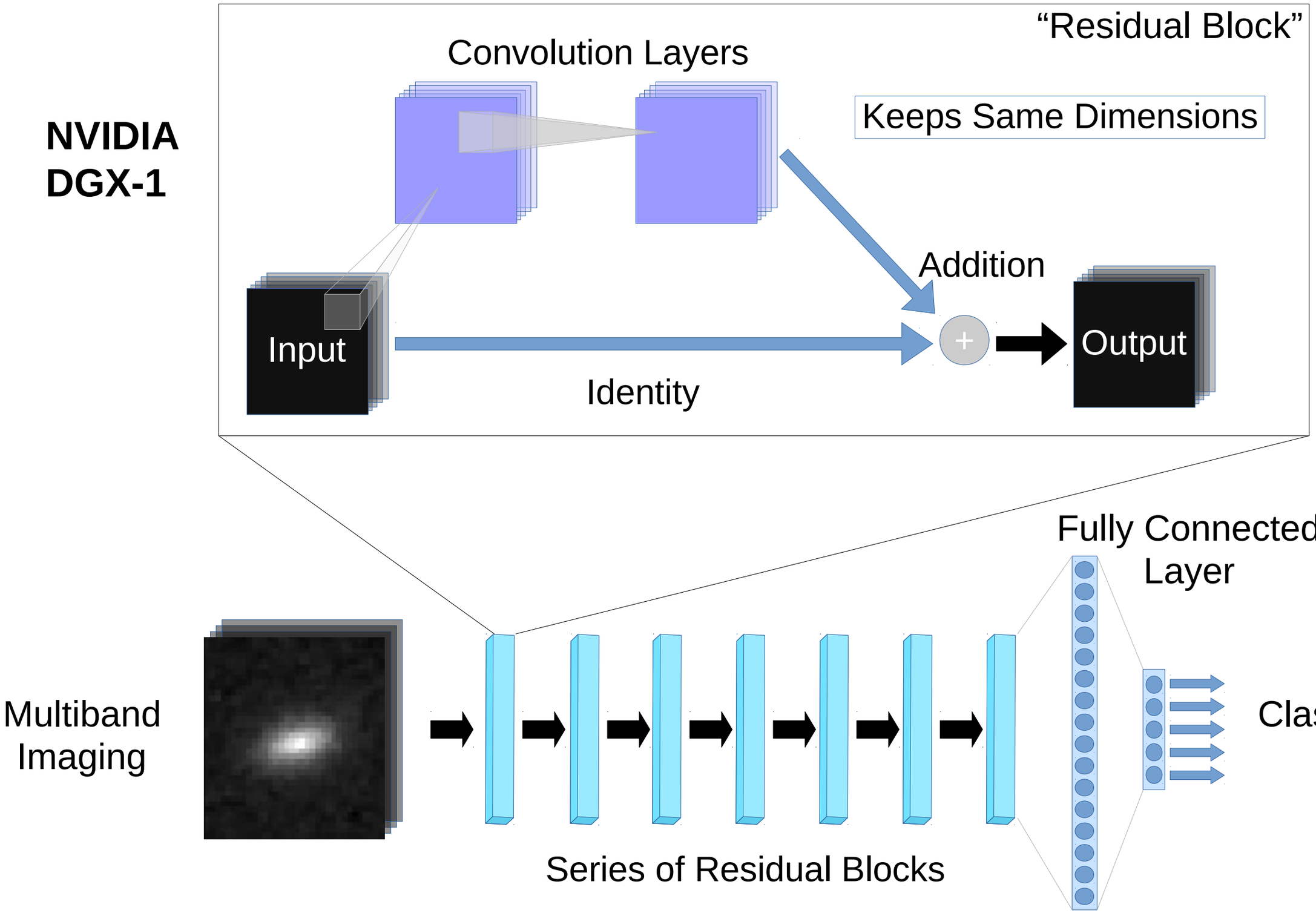




New observatories will image >10 billion galaxies.

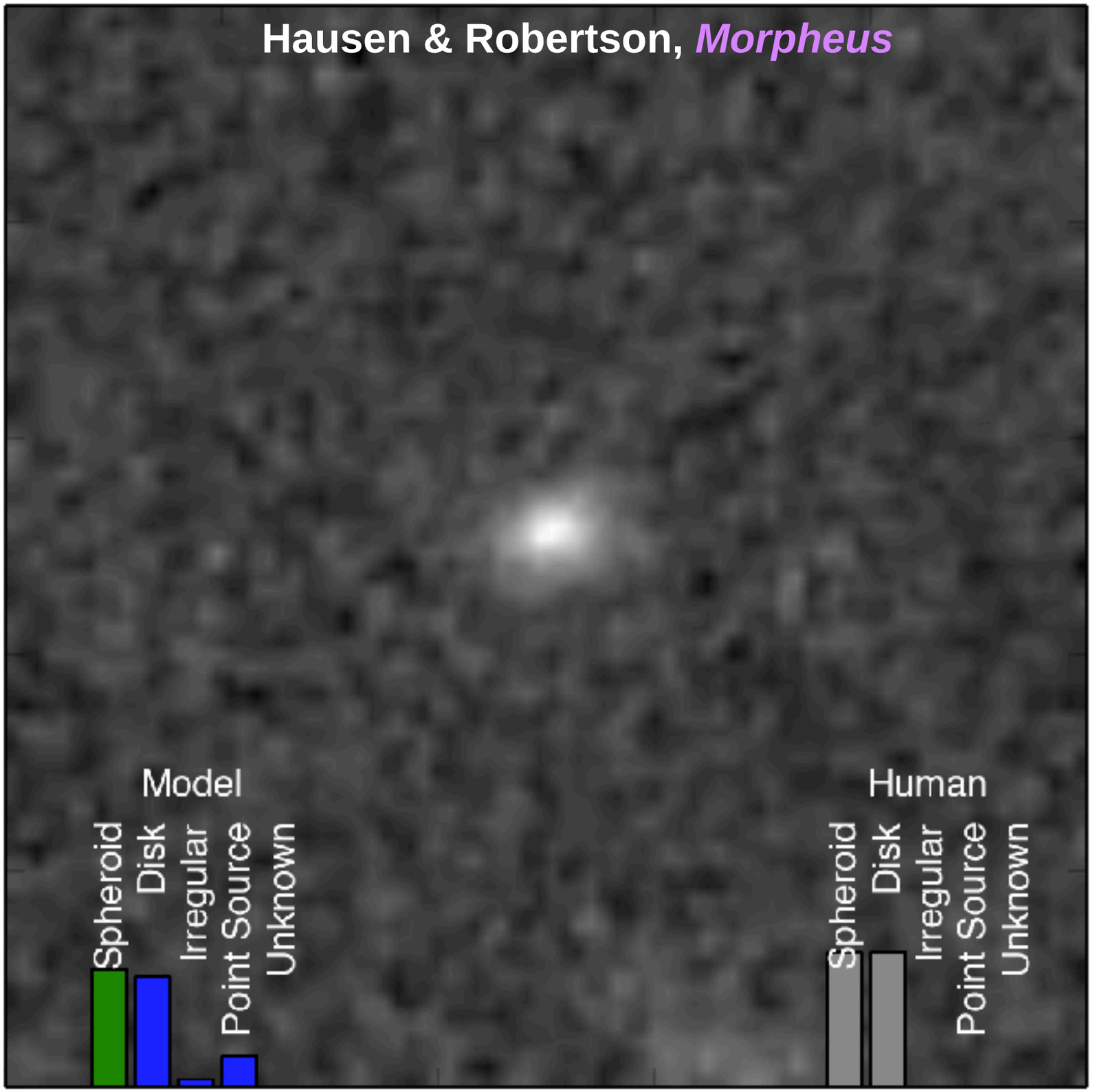
Morpheus — a UCSC deep learning model for astronomical galaxy classification by Ryan Hausen

NVIDIA DGX-1



Hausen & Robertson, *(in preparation)*

Hausen & Robertson, *Morpheus*



preliminary

Summary

- The *Cholla* hydrodynamical simulation code uses NVIDIA GPUs to model astrophysical fluid dynamics, written by Evan Schneider for her PhD thesis supervised by Brant Robertson.

- UCSC Astrophysics is using the ORNL *Titan* supercomputer and DGX-1 system, each powered by NVIDIA GPUs, for astrophysical simulation and astronomical data analysis.

- The *Morpheus* Deep Learning Framework for Astrophysics is under development by Ryan Hausen at UCSC for automated galaxy classification and other astrophysical machine learning applications.