BlueWaters Symposium 2016, Sunriver, OR

BlueTides Simulation

Modeling the First Quasars and Galaxies on BlueWaters

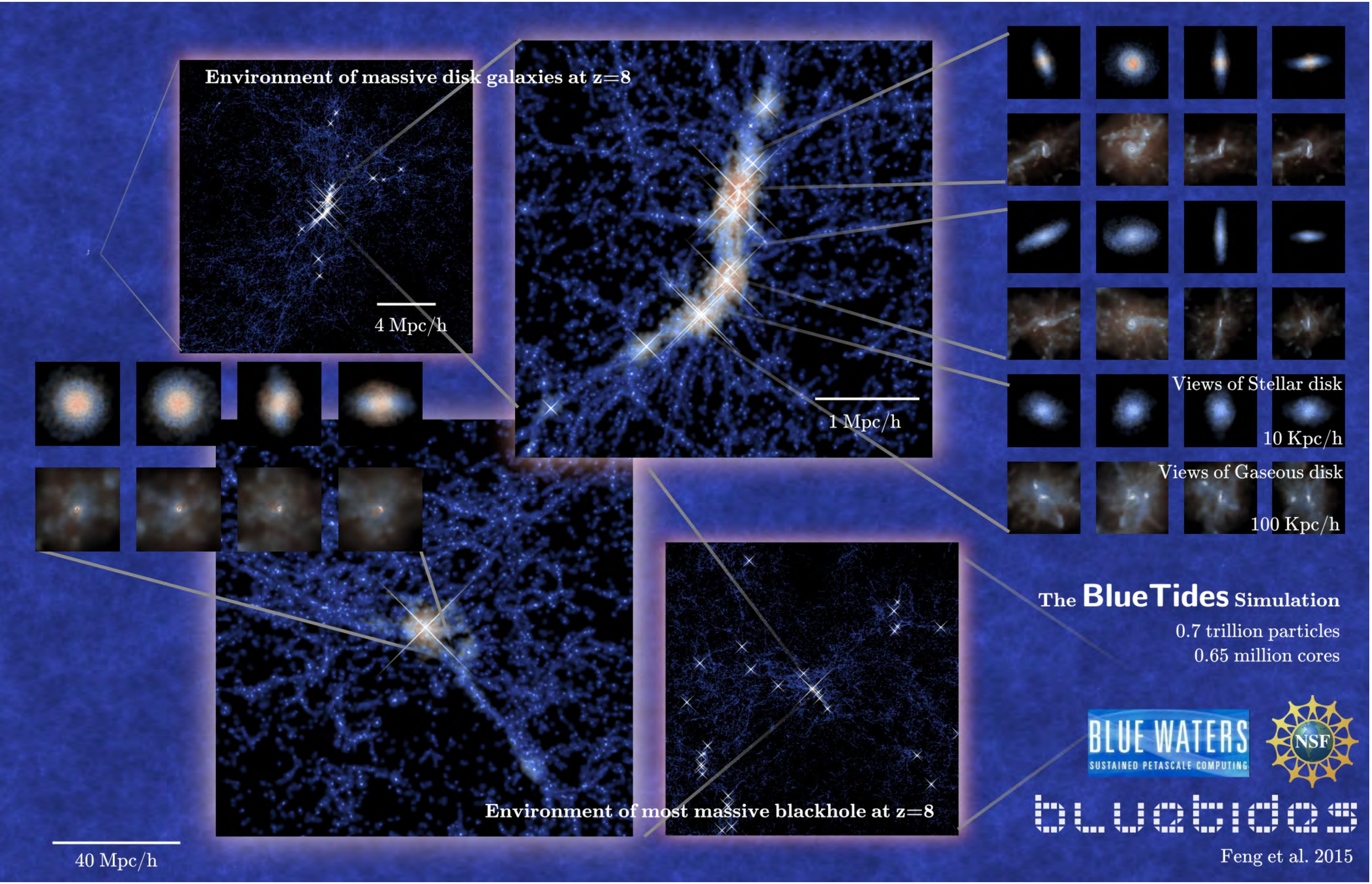
With the simulation software MP-Gadget

Yu Feng
Berkeley Center for Cosmological Physics
Berkeley Institute for Data Science

Science Team:
Nicholas Battaglia, Princeton University
Simeon Bird, John Hopkins University
Rupert A Croft, Carnegie Mellon University
Tiziana Di Matteo, Carnegie Mellon University
Ananth Tenneti, Carnegie Mellon University
Dacen Waters, Carnegie Mellon University
Steven Wilkins, University of Sussex

Introduction of BlueTides on the Public Website:

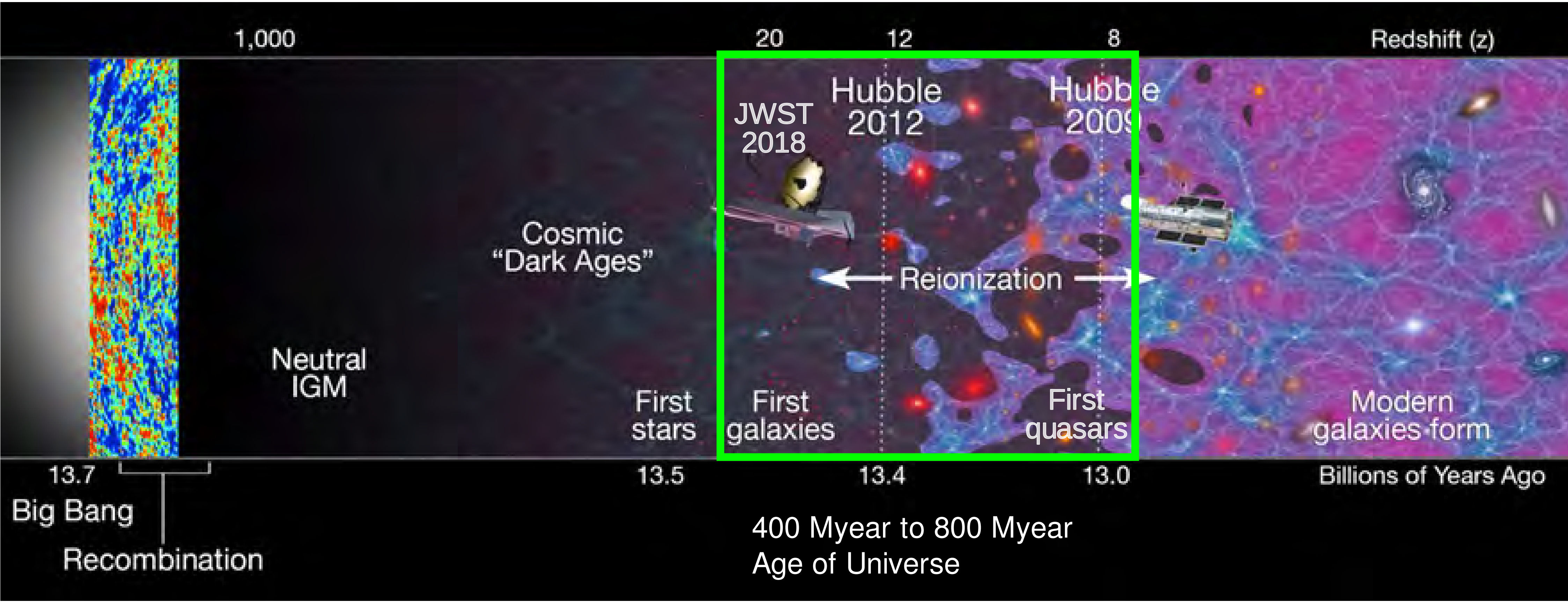
http://bluetides-project.org/

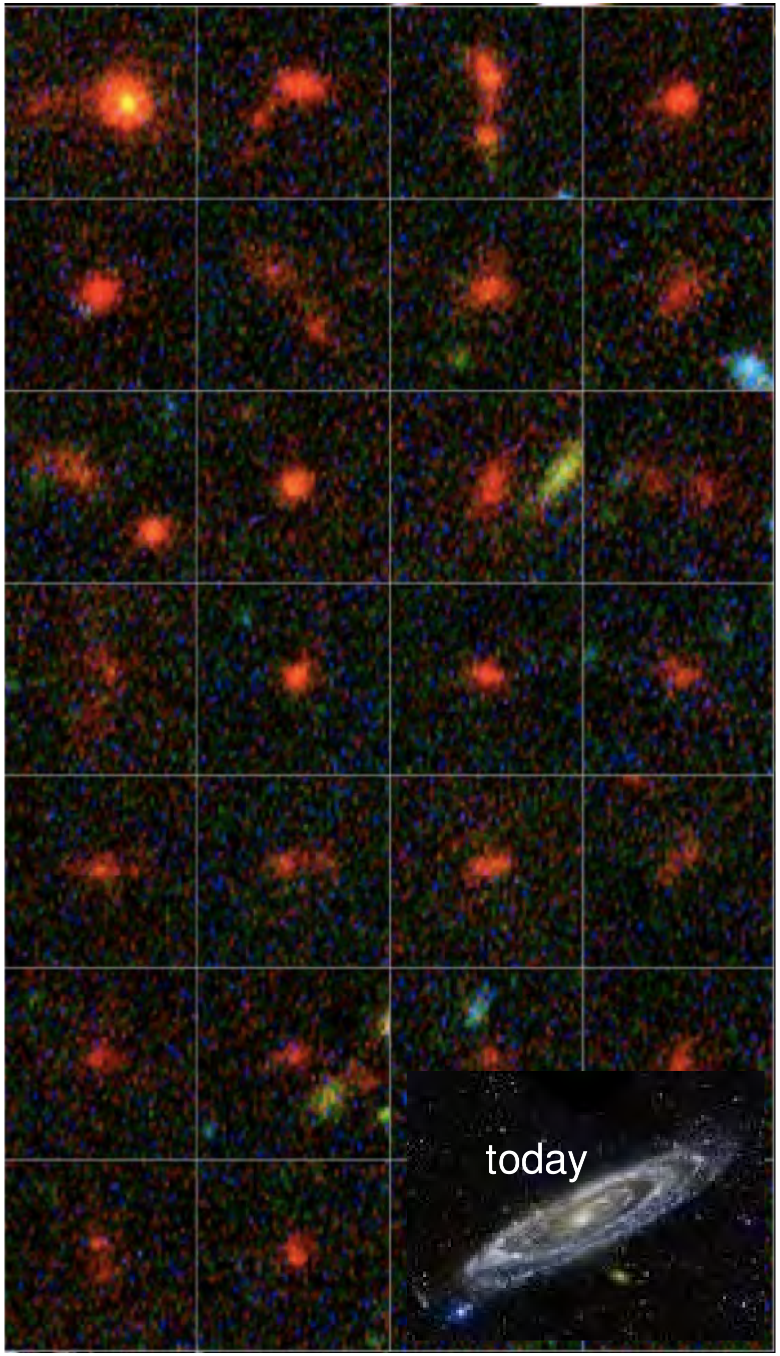

Environment of massive disk galaxies at z=8
4 Mpc/h
1 Mpc/h
Views of Stellar disk
10 Kpc/h
Views of Gaseous disk
100 Kpc/h
Environment of most massive blackhole at z=8
40 Mpc/h
The BlueTides Simulation
0.7 trillion particles
0.65 million cores
BLUE WATERS
SUSTAINED PETASCALE COMPUTING
NSF
bluetides
Feng et al. 2015

The stage of first galaxies and quasars: Epoch of Reionization

Interplay of multiple physics:
dark energy and dark matter, neutrinos, radiation,
hydrodynamics, atomic physics, stellar formation, black-holes.

A rapid build up of observational data in the next decade.





Current observations

- A handful of galaxies irregular and clumpy

- A single Quasar

LETTER TO NATURE

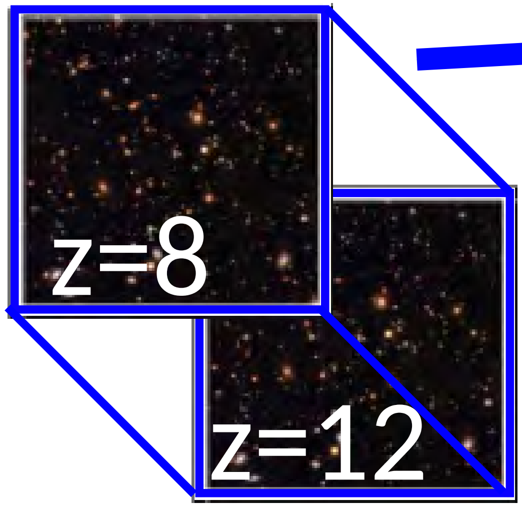
A luminous quasar at a redshift of $z = 7.085$

Daniel J. Mortlock[1], Stephen J. Warren[1], Bram P. Venemans[2], Mitesh Patel[1], Paul C. Hewett[3], Richard G. McMahon[3], Chris Simpson[4], Tom Theuns[5,6], Eduardo A. Gonzáles-Solares[3], Andy Adamson[7], Simon Dye[8], Nigel C. Hambly[9], Paul Hirst[10], Mike J. Irwin[3], Ernst Kuiper[11], Andy Lawrence[9] & Huub J. A. Röttgering[11]

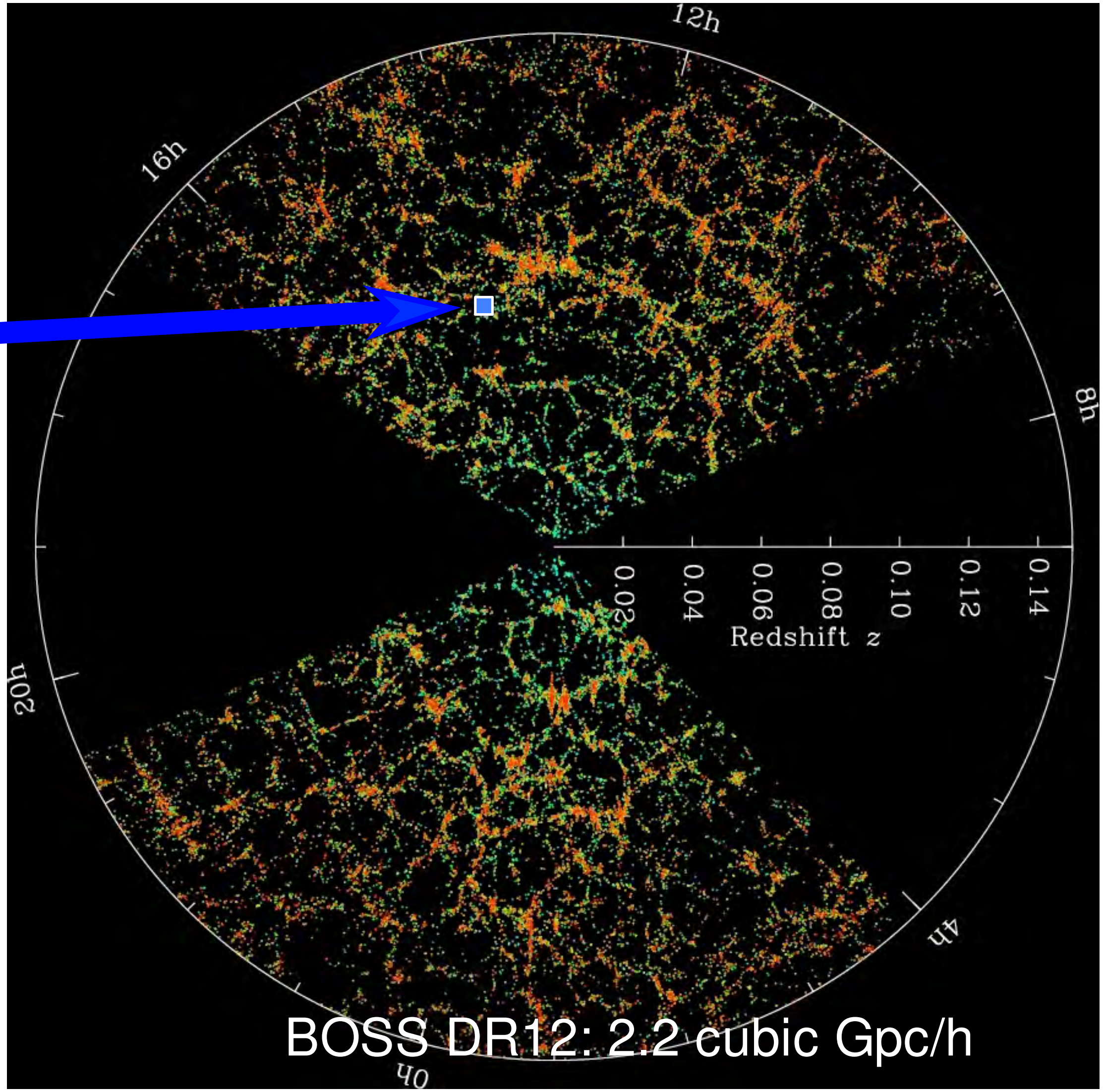
The intergalactic medium was not completely reionized until approximately a billion years after the Big Bang, as revealed[1] by observations of quasars with redshifts of less than 6.5. It has been difficult to probe to higher redshifts, however, because quasars have historically been identified[2-4] in optical surveys, which are insensitive to sources at redshifts exceeding 6.5. Here we report observations of a quasar (ULAS J112001.48+064124.3) at a redshift of 7.085, which is 0.77 billion years after the

troscopically confirmed to have even higher redshifts, two are faint $J_{AB} \gtrsim 26$ galaxies[10,11] and the other is a γ-ray burst which has since faded[12]. Indeed, it has not been possible to obtain high signal-to-noise ratio spectroscopy of any sources beyond the most distant quasars previously known: CFHQS J0210−0456[13] ($z = 6.44$), SDSS 1148+5251[3] ($z = 6.42$) and CFHQS J2329+0301[14] ($z = 6.42$). Follow-up measurements of ULAS J1120+0641 will provide the first opportunity to explore the 0.1 Gyr between $z = 7.08$ and $z = 6.44$, a significant cosmological epoch about which little is cur-

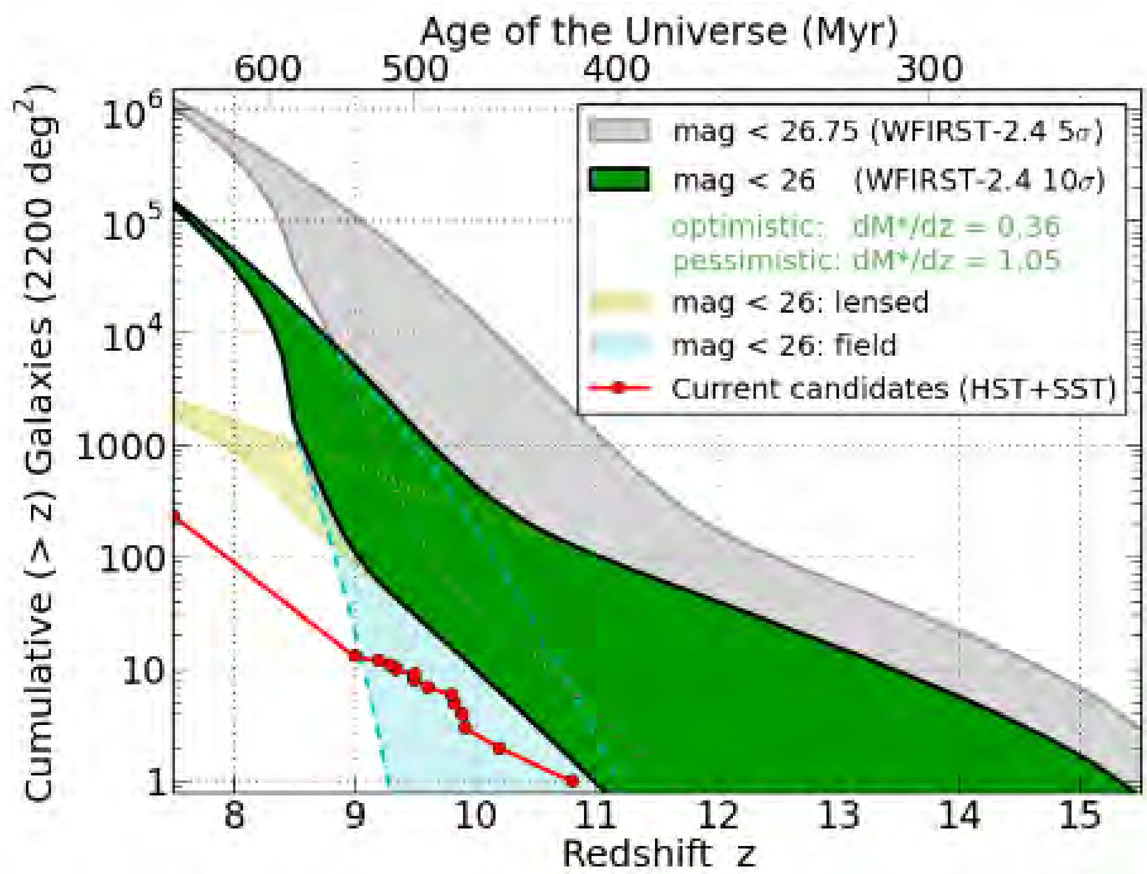
A 1 / 200,000th tip of the iceberg



HUDF: A 23 Mpc/h cube

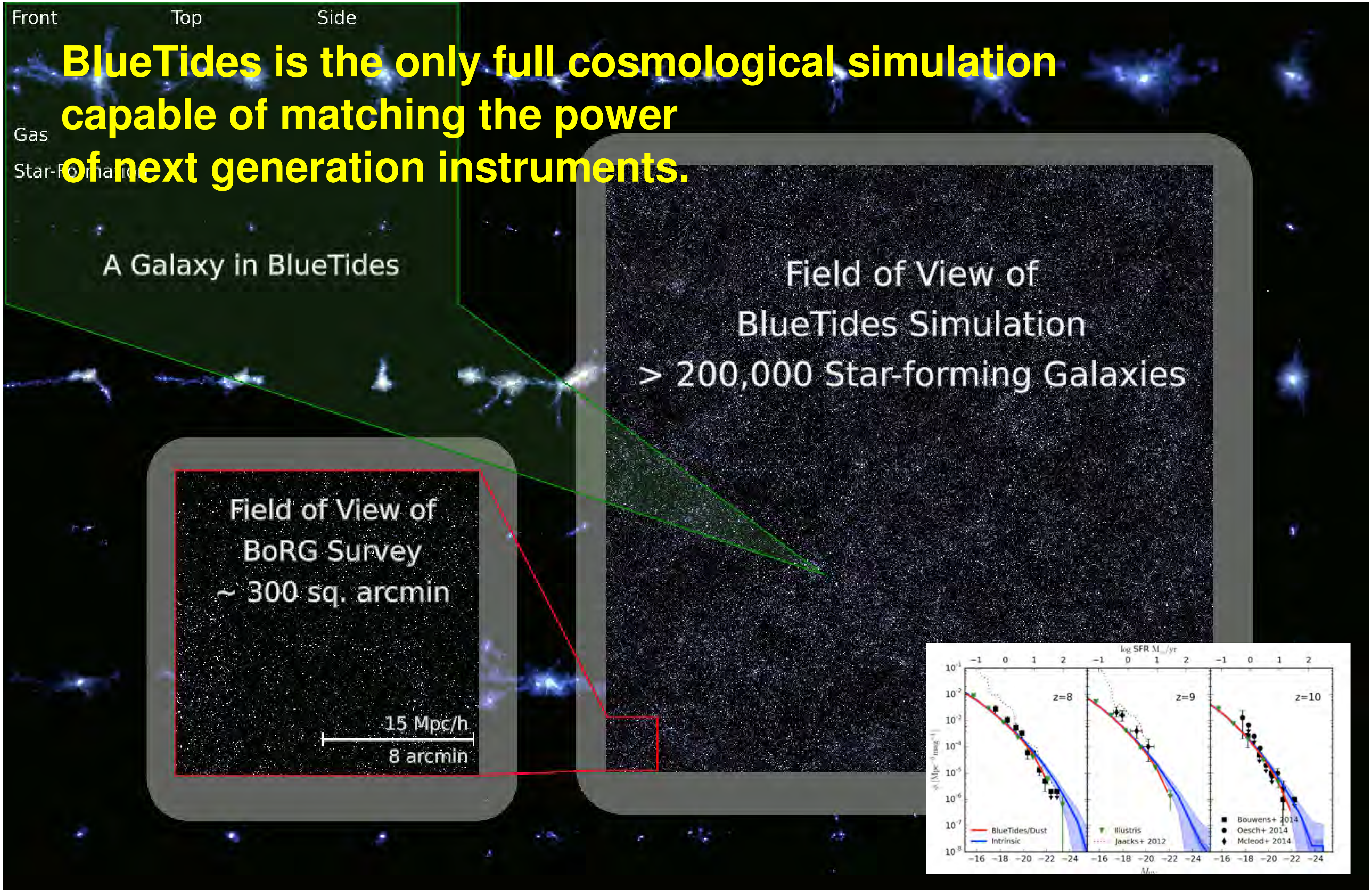


New major space based programs brings 1000 times more data, covering significantly large volume than current data.



Challenges to theoretical modelling:
Need Large volume and high resolution computer simulations

BlueTides is the only full cosmological simulation capable of matching the power of next generation instruments.



Scientific Achievements

A unique position: large volume and high resolution.

- 11 papers submitted / accepted / published

Made possible by super computers as big and as accessible as BlueWaters:
Lots of memory, lots of cpus and large IO bandwidth.

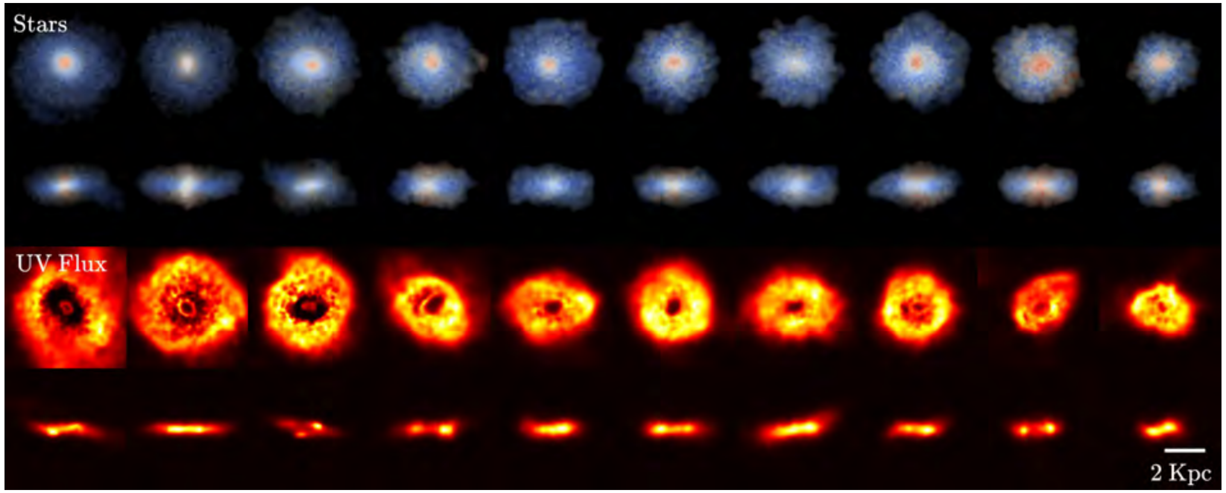
BlueTides Simulation Visualize Data Code People Publications Acknowledgements

Publications

- *Petascale cosmology: simulation of structure formation* R. Croft, T. Di Matteo, N. Khandai, Y. Feng, Computing in Science and Engineering, 2015, Volume 17, Issue 2, DOI:10.1109/MCSE.2015.5
- *Sorting At Scale on BlueWaters* Y. Feng, M. Straka, R. Croft, T. Di Matteo, 2015, Proceedings of Cray User Group Conference 2015; best paper finalist.
- *MP-Gadget: Gadget with half of million CPUs* Y. Feng, T. Di Matteo, R.A.C. Croft, 2015, in prep
- *The formation of Milky Way-mass disk galaxies in the first 500 million years of a cold dark matter universe* Y. Feng, T. Di Matteo, R.A.C. Croft, A. Tenneti, S. Bird, N. Battaglia, S. Wilkins, 2015, APJ Letters http://http://adsabs.harvard.edu/abs/2015ApJ...808L..17F
- *BlueTides: First galaxies and reionization* Y. Feng, T. Di Matteo, R.A.C. Croft, S. Bird, N. Battaglia, S. Wilkins; 2015, Monthly Notices of Royal Astronomical Society, http://adsabs.harvard.edu/abs/2016MNRAS.455.2778F
- *Monsters in the Dark: Predictions for Luminous Galaxies in the Early Universe from the BlueTides Simulation* Waters, Dacen; Wilkins, Stephen; Di Matteo, Tiziana; Feng, Yu; Croft, Ruper; Nagai, Daisuke, http://adsabs.harvard.edu/cgi-bin/bib_query?arXiv:1604.00413
- *The Lyman-continuum photon production efficiency in the high-redshift Universe* Wilkins, Stephen M.; Feng, Yu; Di-Matteo, Tiziana; Croft, Rupert; Stanway, Elizabeth R.; Bouwens, Rychard J.; Thomas, Peter, 2016 Monthly Notices of Royal Astronomical Society, http://adsabs.harvard.edu/abs/2016MNRAS.458L...6W
- *Forecasts for the WFIRST High Latitude Survey using the BlueTides Simulation* Waters, Dacen; Di Matteo, Tiziana; Feng, Yu; Wilkins, Stephen M.; Croft, Rupert A. C. submitted to MNRAS May 13th, 2016 http://adsabs.harvard.edu/abs/2016arXiv160505670W
- *The Photometric Properties of Galaxies in the Early Universe* Wilkins, Stephen M.; Feng, Yu; Di-Matteo, Tiziana; Croft, Rupert; Stanway, Elizabeth R.; Bunker, Andrew; Waters, Dacen; Lovell, Christopher, accepted in MNRAS http://adsabs.harvard.edu/abs/2016arXiv160505044W

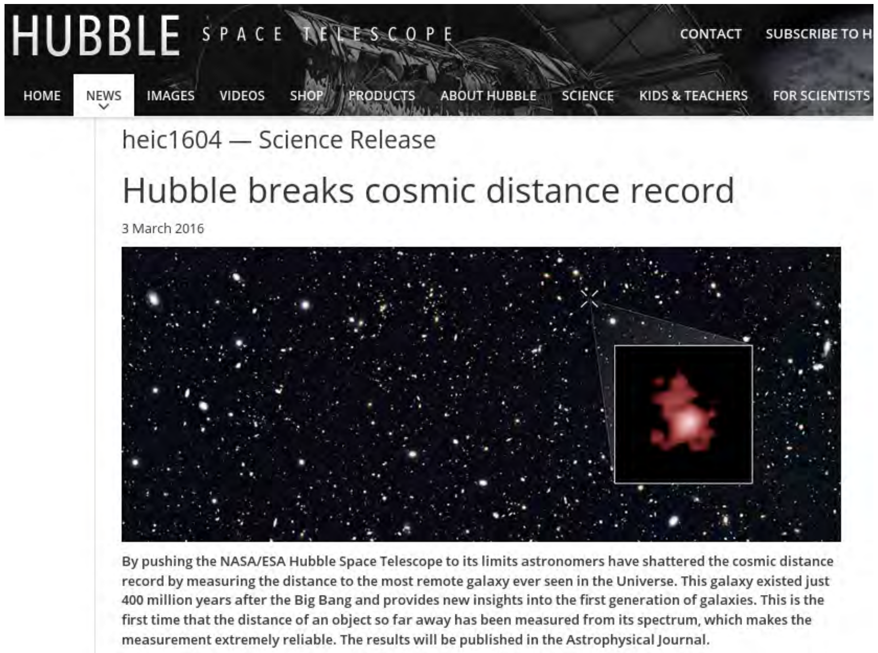
Example 1:

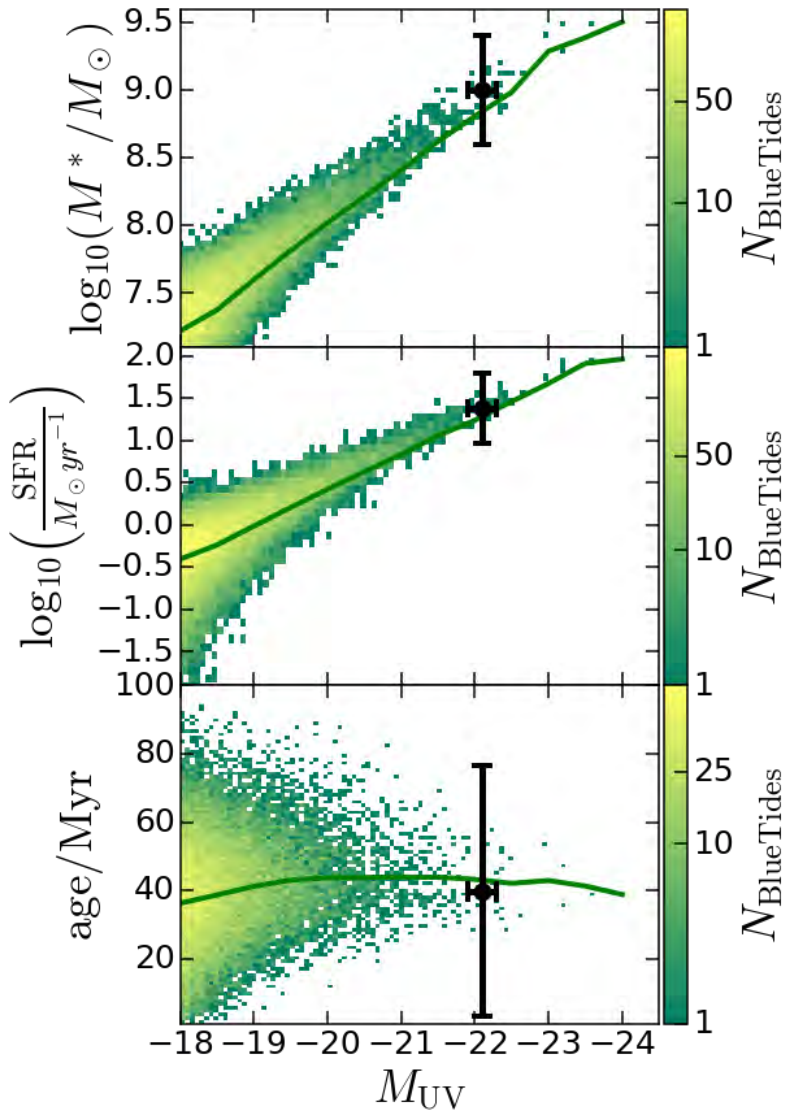
The existence of massive disky high redshift galaxies;
(Feng et al 2016)



Example 2:

The earliest galaxy known,
discovered in March 2016 (Oesch et al.)
comparing to BlueTides predictions.



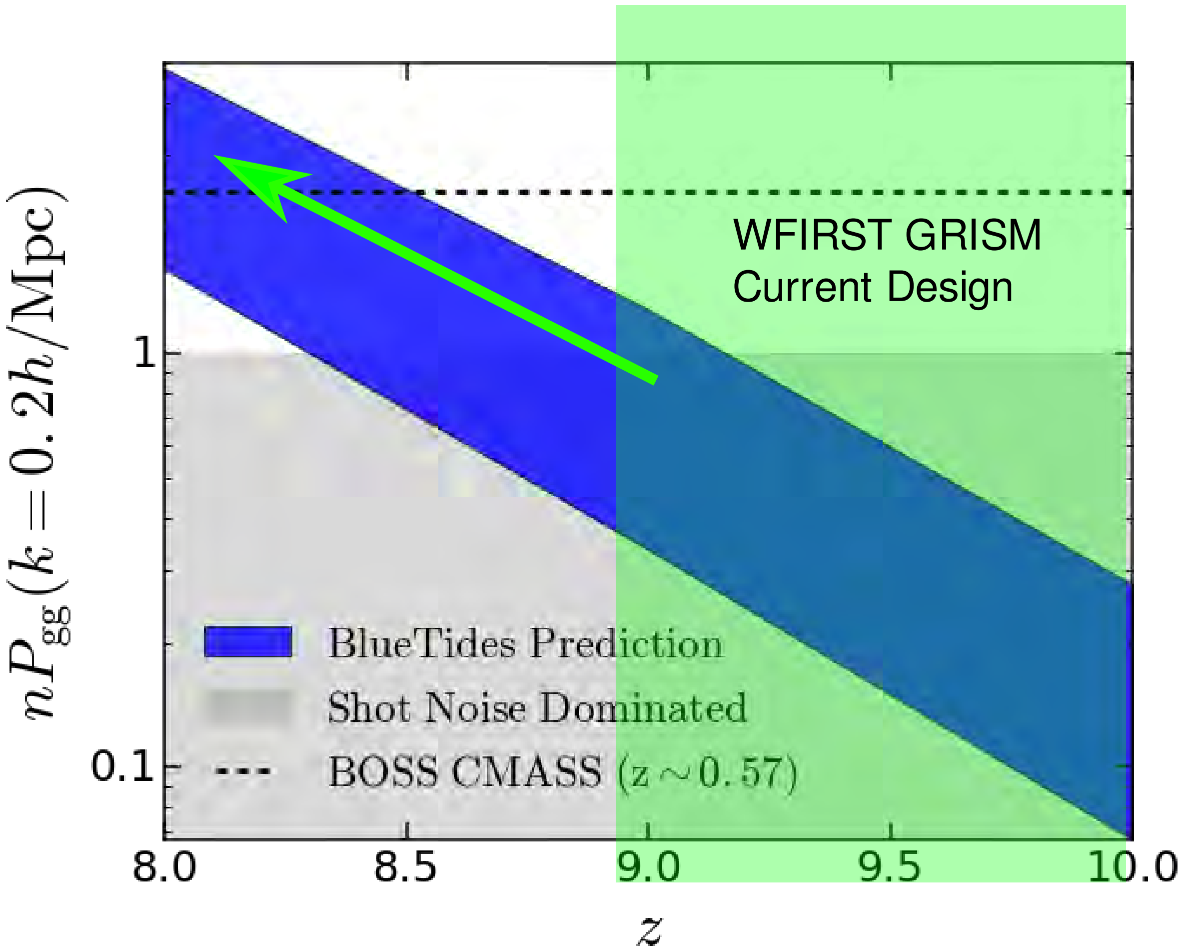


Waters et al, 2016

Example 3:

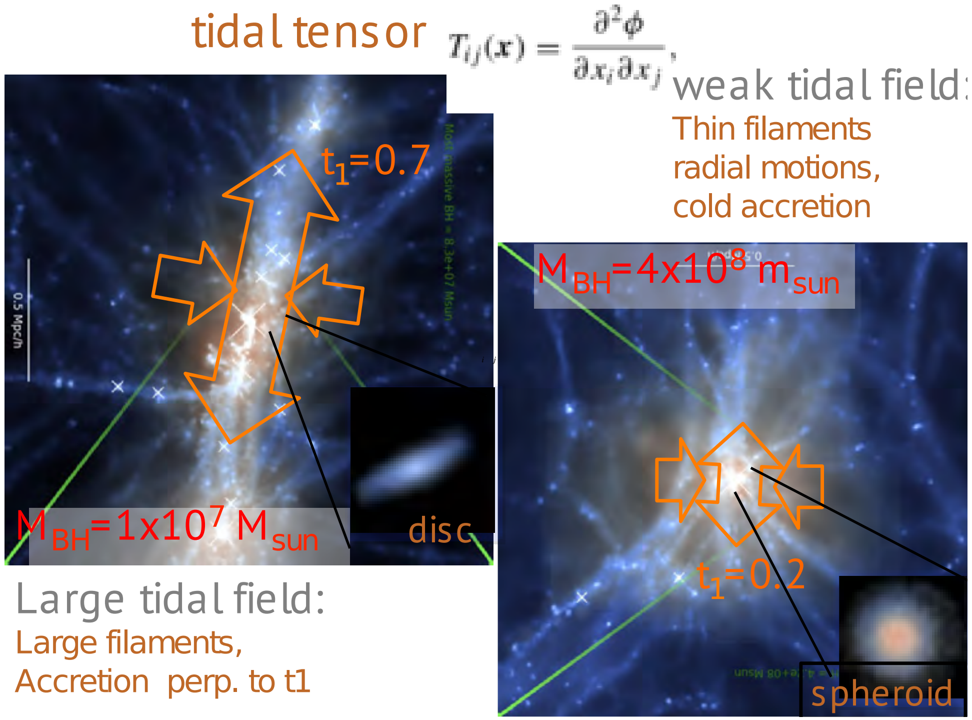
BlueTides predicts a measurement of
galaxy clustering from WFIRST galaxies; (Waters et al, 2016)

First dark energy measurement from first half of Universe!
(Baryon Acoustic Oscillation)



Example 4:

Co-evolution of quasars and their hosting galaxies;
(Di Matteo et al 2016, in prep)

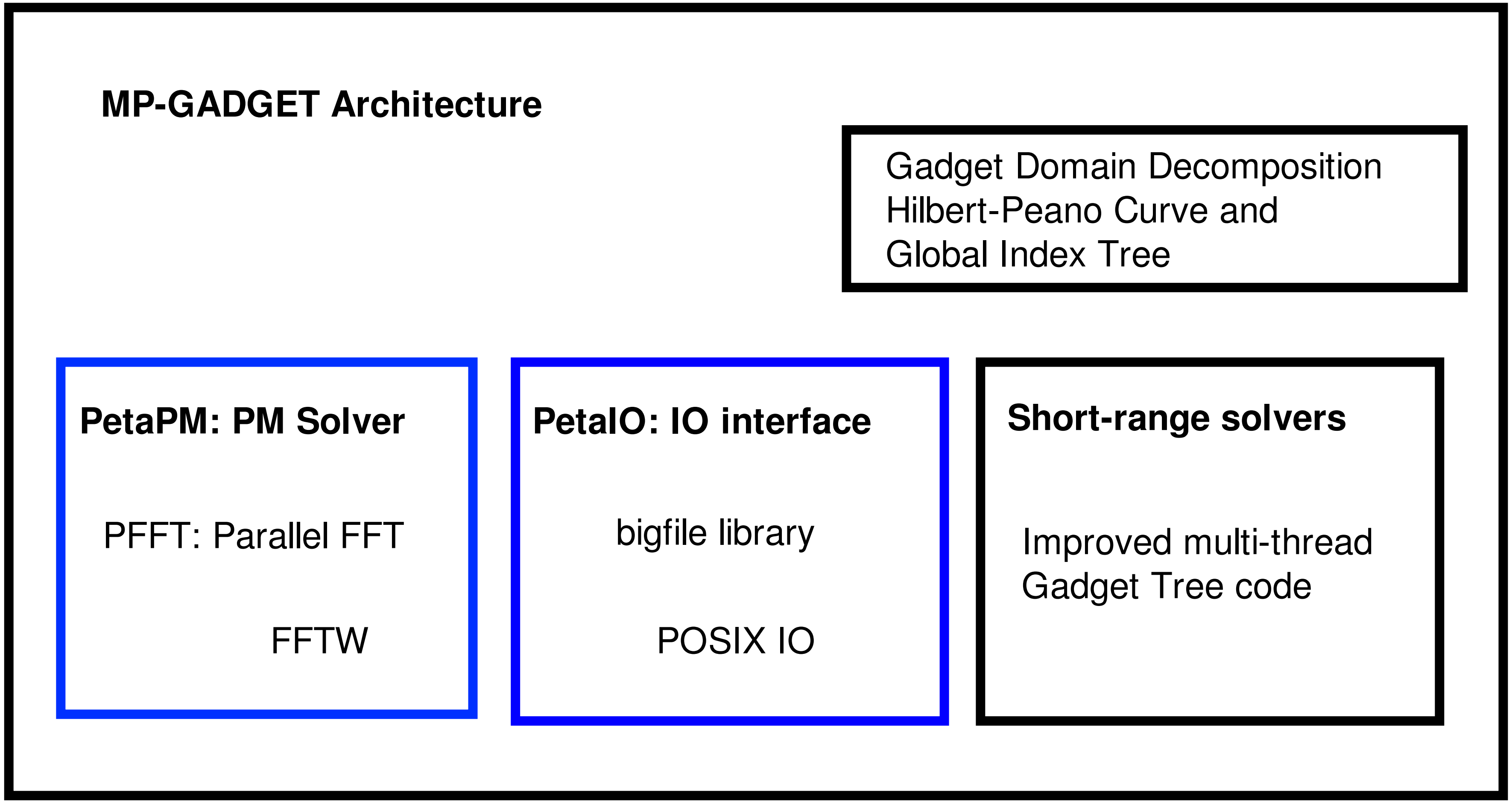


Software Product: Continued Developments on MP-Gadget

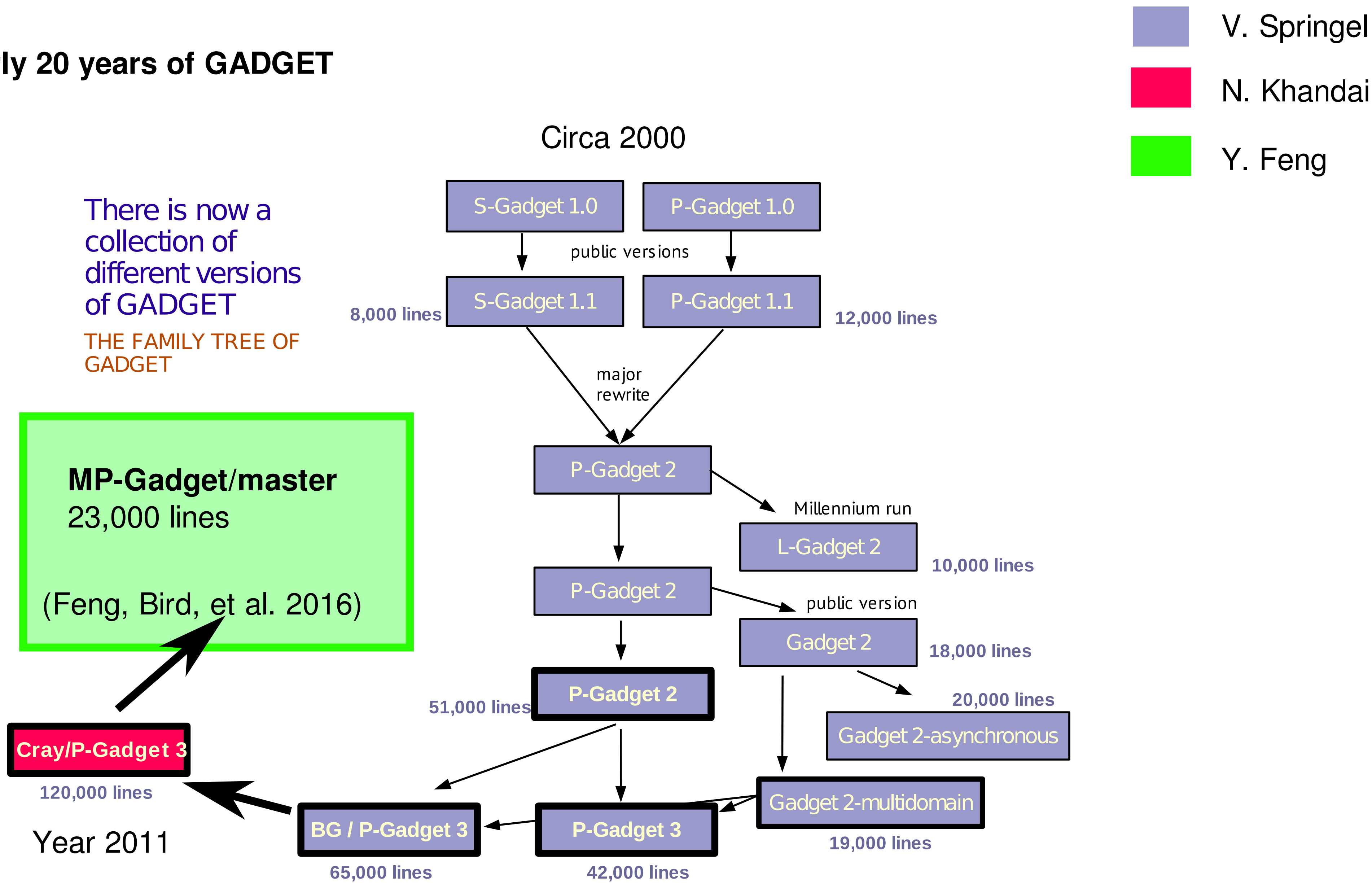
MP-Gadget:
- software behind BlueTides;
- a hydrodynamics and gravity solver for cosmological simulations;
- multiple phyiscs models for baryon physics;
- scales from a single laptop to BlueWaters full capability.

Pushing the envolope in cosmology and astrophysics
that involves full machine capability scale computing of facilities.

Access to BlueWaters resources (computing and support) is crucial.

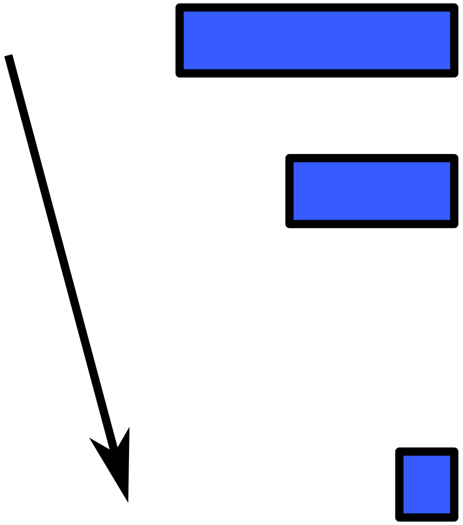

MP-GADGET Architecture
Gadget Domain Decomposition
Hilbert-Peano Curve and
Global Index Tree
PetaPM: PM Solver
PFFT: Parallel FFT
FFTW
PetaIO: IO interface
bigfile library
POSIX IO
Short-range solvers
Improved multi-thread
Gadget Tree code

Nearly 20 years of GADGET



Preparing the MP-Gadget code base for a sustainable future.

With a much leaner code base



Cray/P-Gadget3: 120,000 lines

MP-Gadget/BlueTides-I: 72,000 lines

Improving existing capabilities.
- Improved PM scaling
- Improved Threading efficiency

Current MP-Gadget: 22,000 lines

A better foundation for new features.

Specifically, since MP-Gadget/BlueTides-I

Removed 18,000 lines of unused, deprecated, or duplicated code from the code base;
- rewriting and cleaning up unlicensed code;
- eliminating nested *#ifdef* blocks from the source code;
- Zel'dovich initial condition solver now uses the main PM solver.
- Blackhole, star formation and Halo finder uses the upgraded Tree Evaluator.

Improved build system
- off-tree compilation;
- A flat list of MAKEFILE flags;

A self-describing parameter file parser;
- Demand the modules to be self-documenting.

For longer term sustainability:
Encouraging a flatter, low redundancy, self-contained coding style

Additional features since MP-Gadget/BlueTides-I

on the new code base

Implemented:

- Modeling radiation, and neutrino components in Hubble expansion;
- A quick galaxy formation model for inter-galactic medium physics;

In-progress:

- A low resolution feedback model for inter-cluster medium physics;

Planned:

- A new domain decomposition scheme that uses less memory to improve over-decomposition and reduce load imbalance in short range solvers.

PAID: Improving the IO performance of MP-Gadget

IME Team: Parallel IO (Darren Adams, William Gropp, Luu Huong, Edward Karrels)

Science Team: BlueTides team + Markus Scheucher (CMU Student)

Peak IO throughput of BlueTides simulation:

554.8 GB/s.

Can we do better?

The bigfile benchmark tool
http://github.com/bluetides-project/IO_simulator

- Simulating the IO pattern of BlueTides without running the simulation;

- General purpose benchmarking;

IO of MP-Gadget : bigfile

bigfile is the IO library of MP-Gadget.

Exposed data avoid container overhead;

Splicing avoids Lustre magics;

Plain text meta data - schema and attributes;

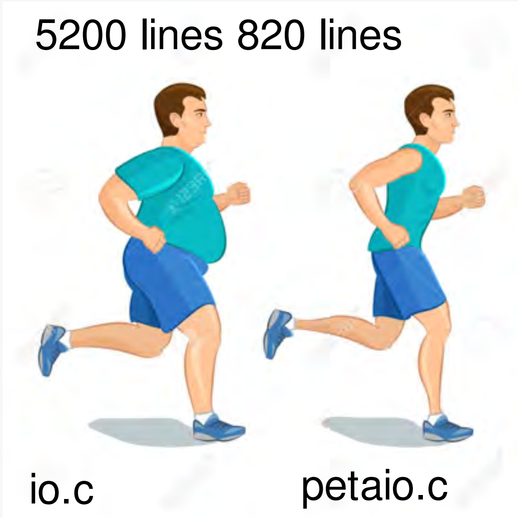
High availability:
- Reusable by linking or including source code.

http://github.com/rainwoodman/bigfile

A factor of 10 faster than Compressed HDF5 used in previous benchmarks of BlueTides

This is the directroy structure of an example file:

```
/scratch1/scratchdirs/sd/yfeng1/example-bigfile
  block0
    header
    attrs-v2
    000000
  group1
    block1.1
      header
      attrs-v2
      000000
      000001
    block1.2
      header
      attrs-v2
      000000
      000001
  group2
    block2.1
      header
      attrs-v2
      000000
      000001
```



Initial benchmark runs identified three hotspots in pre-IO states:
- Sequential truncation of files;
- Updating the checksums;
- Casting data types;

Need to investigate further.

Reusable software components in MP-Gadget

PFFT, bigfile, mpsort are reusable components of MP-Gadget;
Updated with full Python bindings and C++ compatible headers.

FastPM:: Fast particle mesh N-body gravity solver/library:
- PFFT, bigfile

(Feng et al 2016, MNRAS)

nbodykit : massively parallel n-body data analysis toolkit with Python
- PFFT, bigfile, mpsort

(Hand et al, 2016, in prep)

Software tested and developed from BlueWaters directly improves the downstream applications.

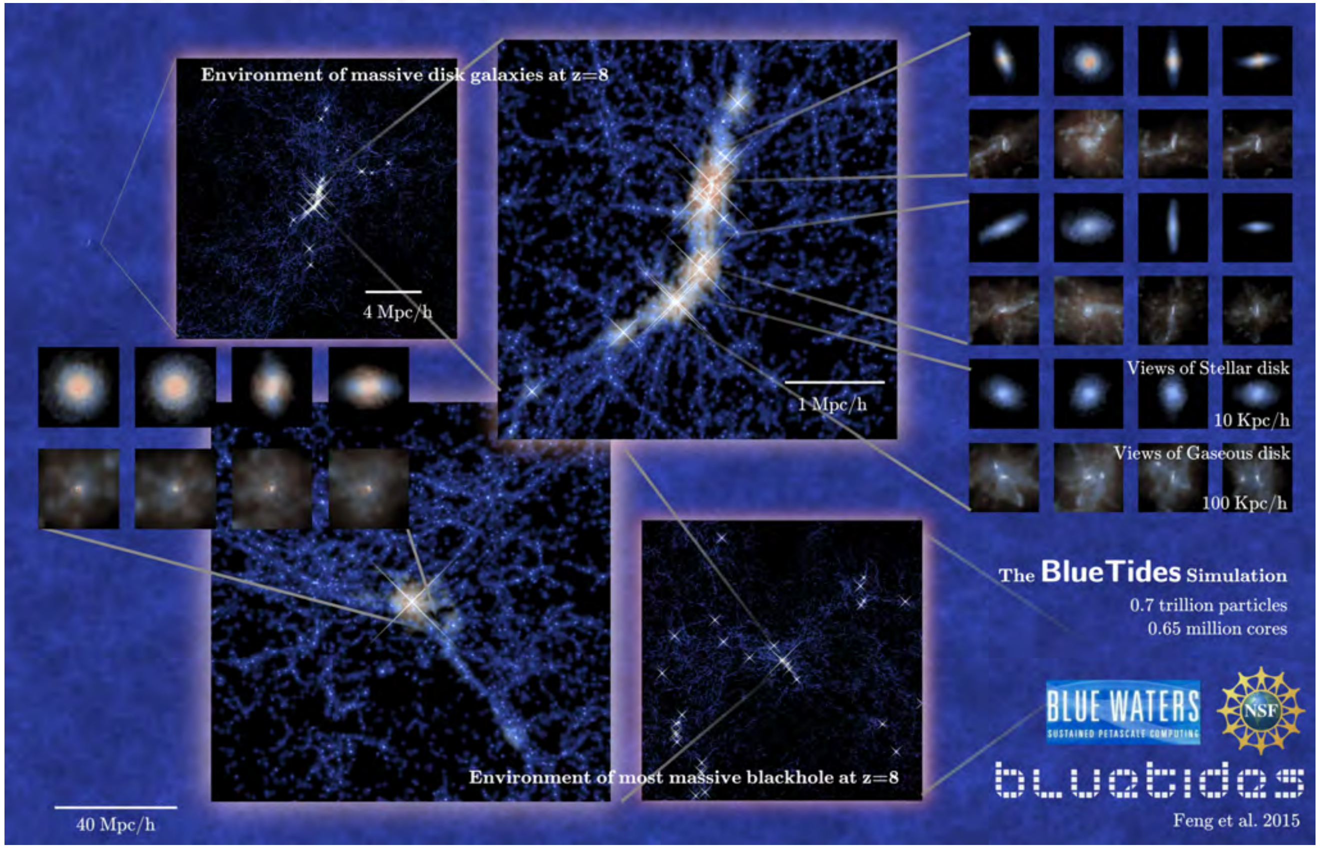
Summary

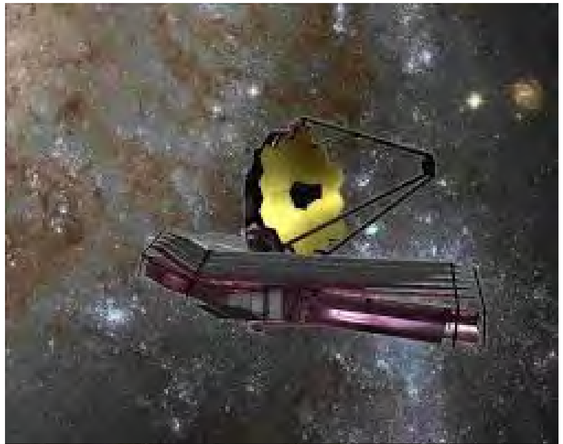
Wide spectrum of impact in science and scientific computing that are enabled by BlueWaters.

Interesting science about first galaxies and quasars from BlueTides simulation.
- Predictions on WFIRST and JWST satellite programs
- Physics of first quasars and galaxy

Support from BlueWaters allows to continue the development of MP-Gadget:
- leaner code base for sustainable development;
- new models and directions;
- understanding and improveing I/O performance from the PAID program;
- component reuse enables shared improvements;


BLUE WATERS

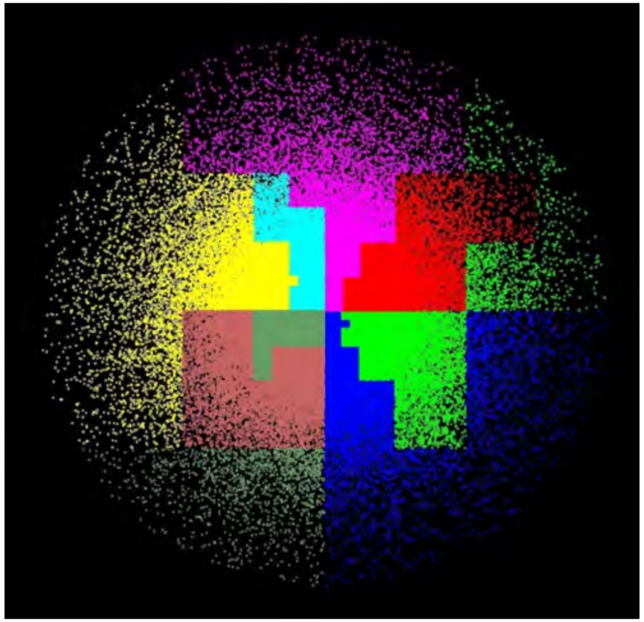

Environment of massive disk galaxies at z=8
4 Mpc/h
1 Mpc/h
Views of Stellar disk
10 Kpc/h
Views of Gaseous disk
100 Kpc/h
The BlueTides Simulation
0.7 trillion particles
0.65 million cores
BLUE WATERS
SUSTAINED PETASCALE COMPUTING
NSF
bluetides
Environment of most massive blackhole at z=8
40 Mpc/h
Feng et al. 2015

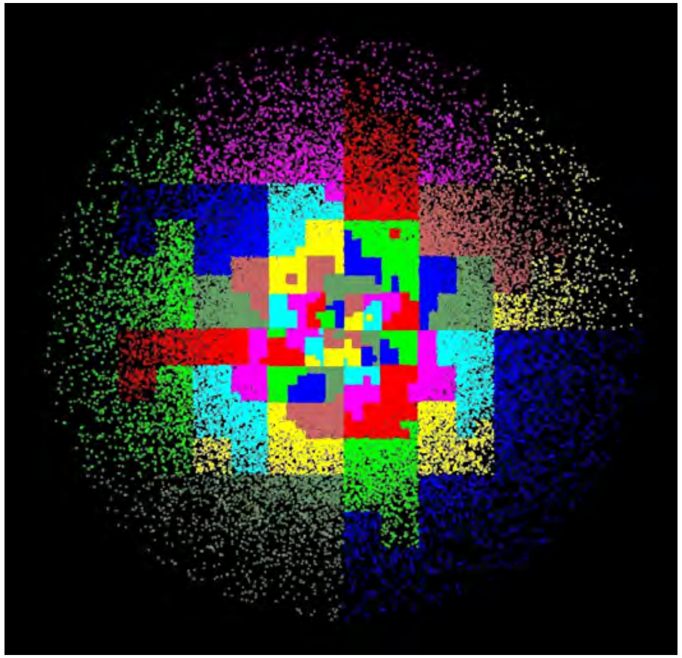


Plans ahead: Hydrodynamics and Domain decomposition

Short range interaction:
 - Shorter time steps
 - Particle-particle interaction
Highly clustered distribution of computation in simulation volume.

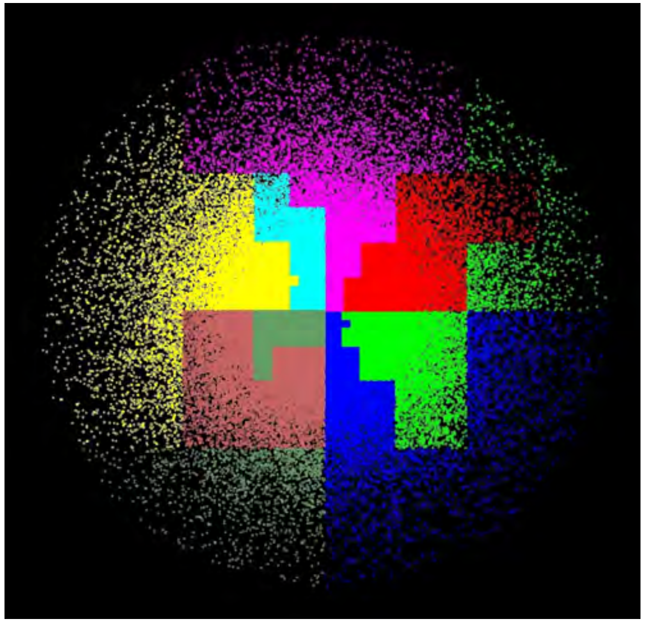
Over-decomposition improves efficiency of short range interaction,
by spliting local hot-spots into chunks that can be scheduled to other ranks,
improving balance and CPU utilization.

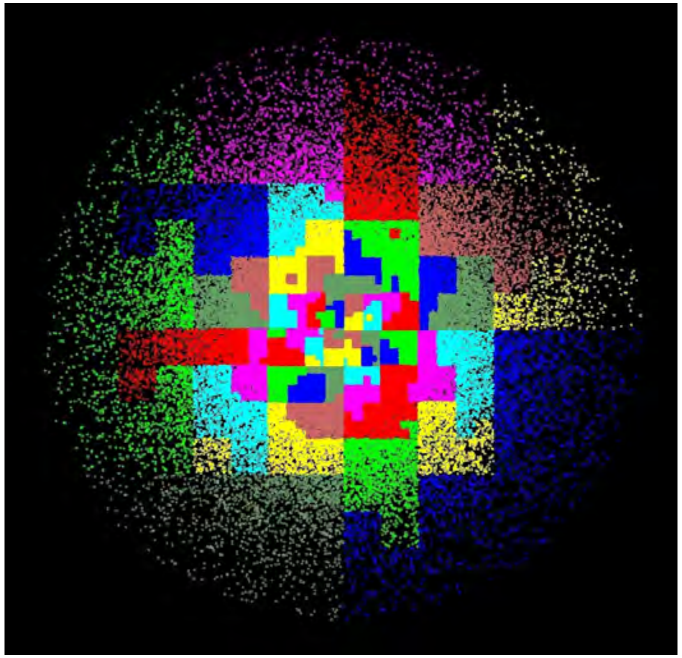


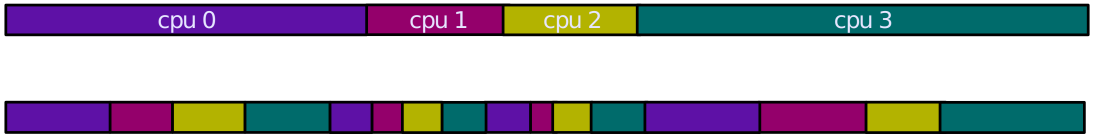


Highly clustered distribution of computation in simulation volume.

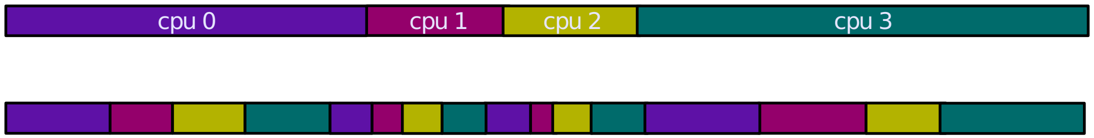
Over-decomposition improves efficiency of short range interaction,
by spliting local hot-spots into chunks that can be scheduled to other ranks,
improving balance and CPU utilization.







But a domain in Gadget is expensive, costing about 1000 bytes per domain due to the Hilbert-Peano decomposition and BH-tree coverage.



But a domain in Gadget is expensive, costing about 1000 bytes per domain due to the Hilbert-Peano decomposition and BH-tree coverage.

On BlueWaters scale, this becomes prohibitively expensive. BlueTides uses 81000 MPI ranks, and 4 sub-domains per ran, totalling about 300 MB / each rank for domain storage, competing memory storage for the state vector, tree, and particle mesh.

We are looking into two possible solutions:

1. Compress the domain data structure, reducing memory usage per domain.
2. Switch to a simpler domain decomposition scheme.