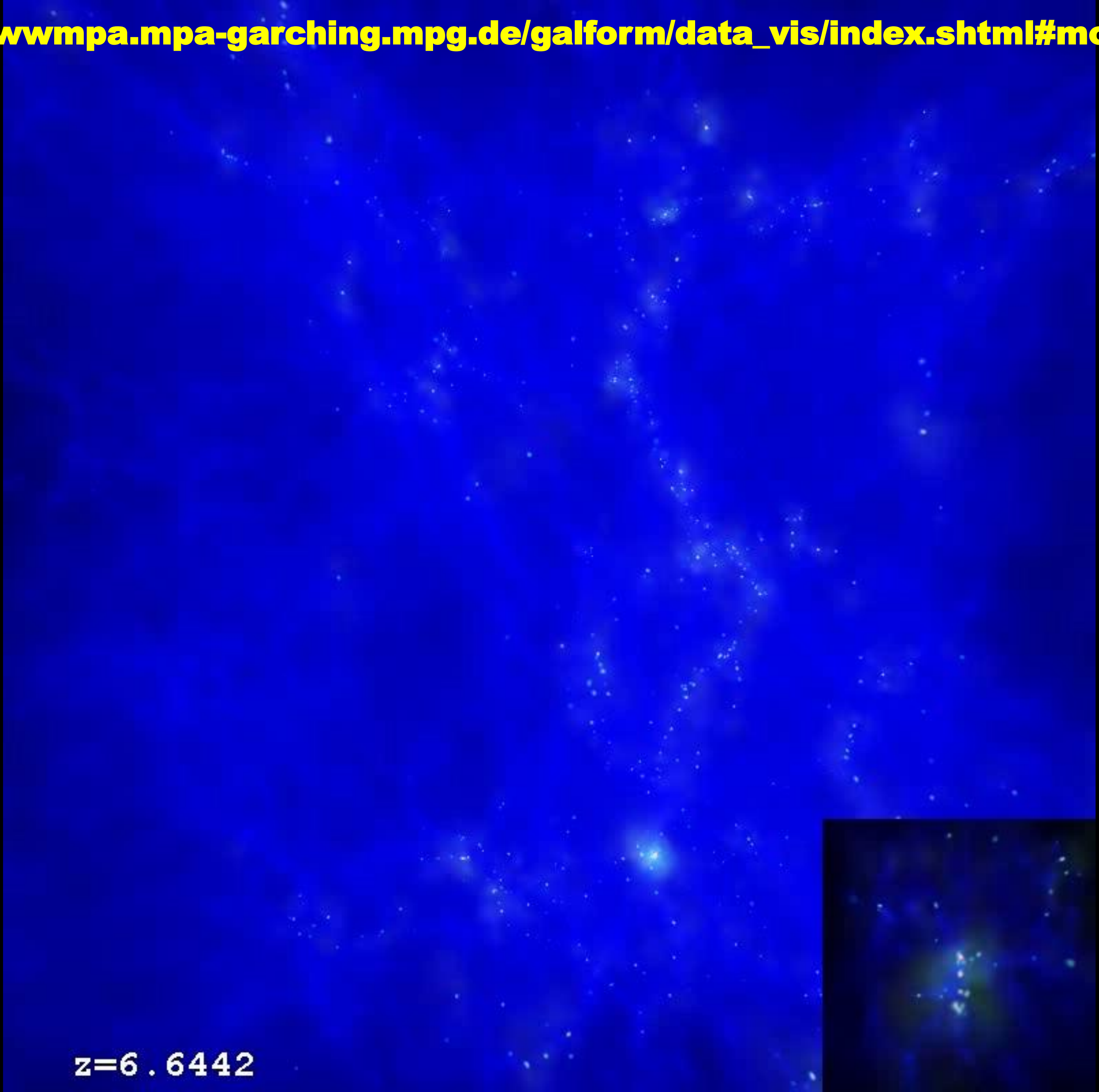
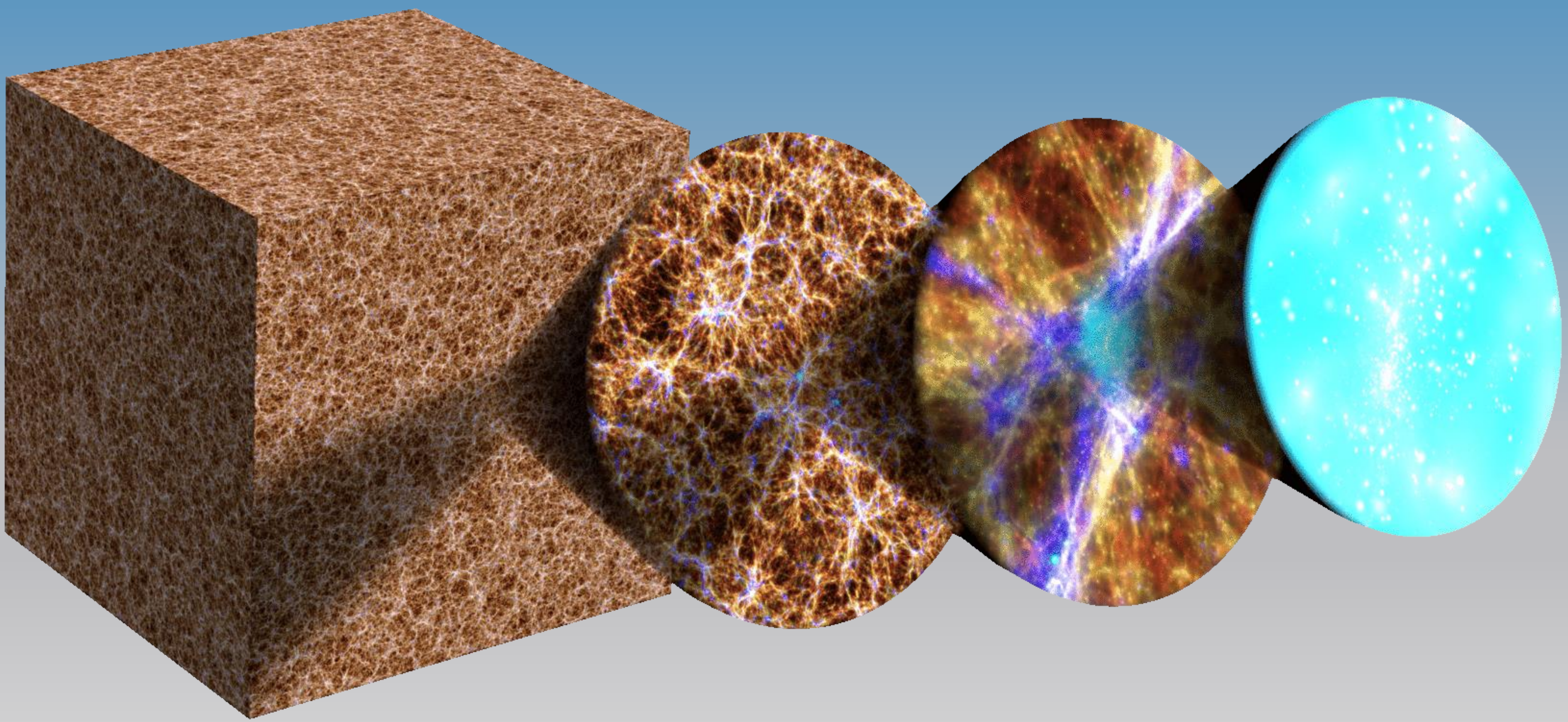
http://wwwmpa.mpa-garching.mpg.de/galform/data_vis/index.shtml#movie8



HPC challenges in simulation Cosmological Struture Formation

Klaus Dolag, Universitäts Sternwarte München



What do we aim for ?

Dave (2001: Space Odyssee):
„My God, it‘s full of stars !"

Bild: APOD 23. Aug 2010, Alex Cherney, Terrastro

What do we aim for ?

Not only, but it is even full of galaxies !

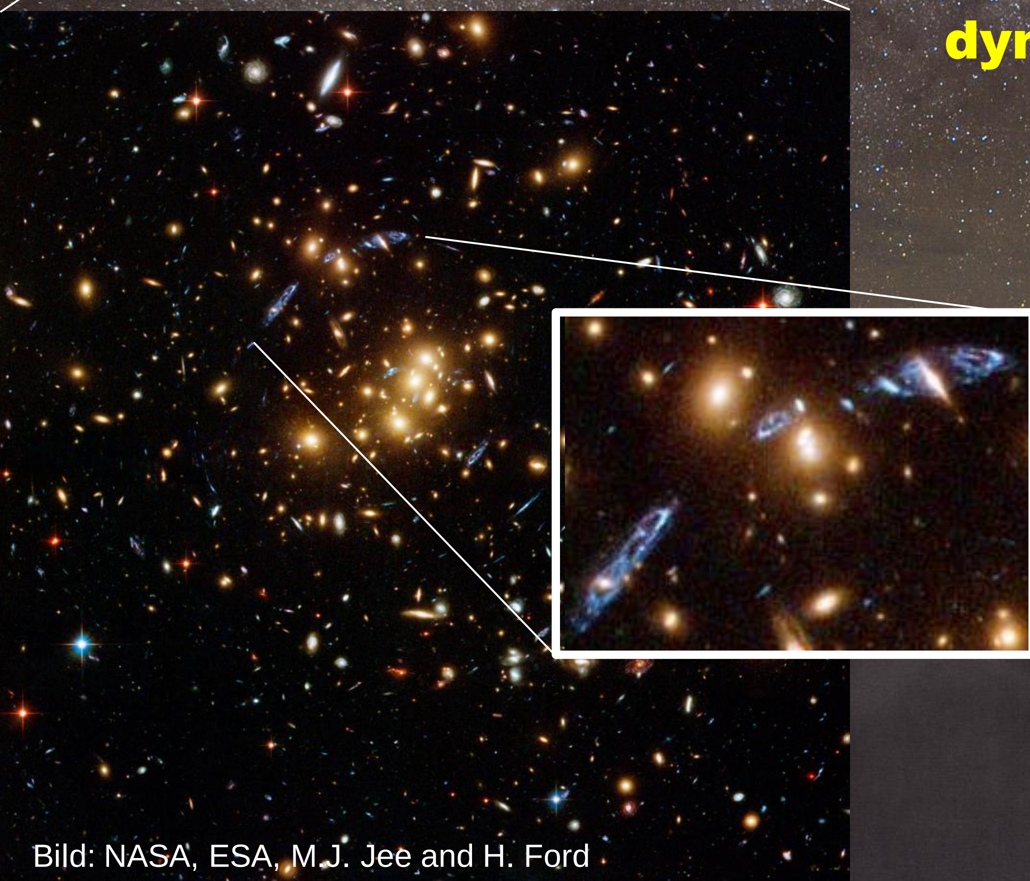


Bild: NASA, ESA, M.J. Jee and H. Ford

Galaxies show us even the dynamic of the universe !

Bild: APOD 21. Juli 2008, Gemini Observatory

How many stars / galaxies ?

Stars (200.000.000.000)
Galaxies (100.000.000.000)

~10^{22}, so ca. the numbers of atoms in a dice. Impossible to dircetly simulate !

Bild: APOD 23. Aug 2010, Alex Cherney, Terrastro

Importance of Galaxy Clusters

Ein merkwürdiger Haufen von Nebelflecken.

Auf zwei mit dem Bruce-Teleskop genommenen Aufnahmen vom 24. März dieses Jahres, welche die Umgebung von 31 Comae Berenices darstellen, findet sich eine sehr interessante Gegend des Himmels. Um die Stelle

$$\alpha = 12^h\,52^m\!.6 \qquad \delta = +28^\circ\,42' \; (1855.0)$$

stehen nämlich zahlreiche kleine Nebelflecken so dicht beisammen, dass man beim Anblick der Gegend förmlich über das merkwürdige Aussehen dieses »Nebelhaufens« erschrickt. Ich habe die Anzahl der Nebel in einem Kreis von 30′ Durchmesser um die angegebene Stelle bestimmt und finde, dass mindestens 108 Nebelflecken auf dieser Fläche beisammen stehen, also auf einer Fläche etwa von der Grösse des Vollmondes. Darunter sind vier oder fünf grössere ausgedehnte und centralverdichtete Nebel, sowie mehrere langgestreckte. Die weitaus meisten haben aber rundliche Form und sind kleiner. *)

Heidelberg, 1901 März 27.

Max Wolf.

	0m	59m	58m	57m	56m	55m	54m	53m	52m	51m	50m	49m	48m	47m	46m	45m	44m	43m	42m	41m	40m	39m	38m	37m	36m	35m	
59° 15′	—	—	—	—	—	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	—	—	—	—	—
30	—	—	—	—	0	0	2	1	3	1	2	1	3	2	1	1	2	0	0	2	1	0	0	—	—	—	—
45	—	—	—	0	1	4	2	0	1	1	3	1	0	1	2	0	1	2	0	1	1	0	0	1	—	—	—
60° 0′	—	—	0	0	2	1	0	1	1	2	2	1	0	2	3	0	1	3	5	0	0	2	0	0	0	—	—
15	—	0	0	2	3	3	7	3	5	5	2	1	4	1	3	1	2	0	0	0	3	5	0	1	0	0	—
30	—	0	2	3	4	3	2	3	5	3	1	3	0	0	0	2	4	1	0	0	0	3	2	0	0	2	0
45	0	1	4	5	9	16	12	15	5	3	1	4	2	1	4	1	2	1	1	0	1	2	4	0	1	0	0
61° 0′	0	1	5	15	19	10	23	15	19	8	4	3	4	2	1	0	1	1	1	0	1	1	1	2	0	0	4
15	0	0	9	17	11	14	36	68	10	7	3	7	0	2	1	2	3	1	1	3	0	4	4	3	0	2	1
30	1	2	2	9	6	12	13	17	20	16	6	7	1	2	1	1	1	1	4	1	1	5	3	1	6	3	1
45	0	5	5	10	8	8	12	9	10	11	4	5	4	2	5	2	6	5	2	1	2	2	1	3	2	2	3
62° 0′	0	2	1	3	6	8	3	10	7	3	5	4	2	4	6	8	3	2	2	5	0	3	9	10	10	2	2
15	0	3	1	6	5	10	11	9	1	10	7	1	5	3	4	4	3	2	3	3	6	4	1	5	2	3	1
30	—	0	1	4	4	1	2	4	8	4	2	1	2	1	2	3	4	2	1	3	9	4	3	2	4	5	3
45	—	0	1	5	2	3	1	3	6	4	6	2	0	6	2	4	3	5	2	6	10	5	3	1	1	7	6
63° 0′	—	0	2	2	2	3	0	0	0	0	1	1	1	1	2	4	0	2	4	2	7	5	0	4	0	2	2
15	—	1	1	2	0	3	0	8	1	1	0	0	0	2	2	1	0	5	3	5	3	8	4	1	2	1	—
30	—	—	0	0	4	0	0	8	0	2	0	0	0	2	4	3	3	4	4	2	8	2	2	2	0	0	—
45	—	—	—	0	2	0	1	4	1	1	1	6	4	2	0	0	5	4	2	1	5	1	1	3	0	—	—
64° 0′	—	—	—	—	0	3	0	0	0	0	1	7	1	2	6	4	3	7	2	4	0	0	0	—	—	—	—
15																											

…uch close
sight of the
of nebulae.

4 or 5 with large extend and
densities, as well as several
s. However most of them are
pared to other observations).

Max Wolf (1863 – 1932)

Quelle: Wikipedia

Importance of Galaxy Clusters

Ein merkwürdiger Haufen von Nebelflecken.

Es ist sofort zu sehen, wenn man die Tabelle oder die Tafel betrachtet, dass das Zusammendrängen der Nebel immer stärker wird, je weiter man in's Innere der Hauptinsel eindringt. Je näher man dem Puncte grösster Dichtigkeit kommt, umso dichter treten auch die Nebel an einander, so dass auf dem innersten Quadratgrad mehr als 320 einzelne Nebelflecken beisammen stehen. An der dichtesten Stelle dieses »Weltpoles« finden sich mehr als 70 Nebel auf der Fläche von $^1/_{16}$ Quadratgrad.

Wir finden also hier ein völlig gesetzmässiges Verhalten in der Anordnung dieser fernen Welten; und dieser ungeheure Reichthum führt uns so eine Ordnung im Weltsystem vor Augen, die sicher für die Erkenntniss des Universums von allergrösster Bedeutung ist, von der wir uns aber auch zugestehen müssen, dass wir noch lange keine erschöpfende Erklärung für sie werden finden können. *)

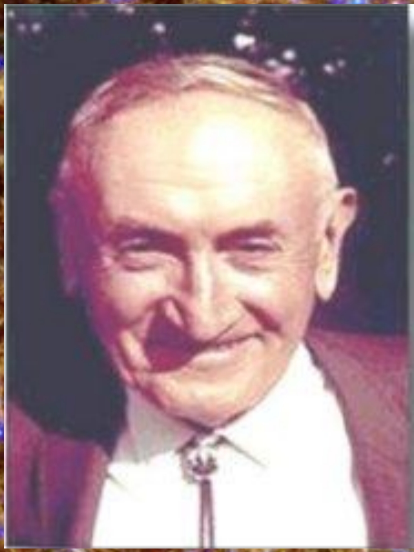
Nebelvertheilung um den Pol der Milchstrasse

the regular behavior within the arrangement of these distant worlds ... of greatest significance for understanding the universe !

Max Wolf (1863 – 1932)

Quelle: Wikipedia

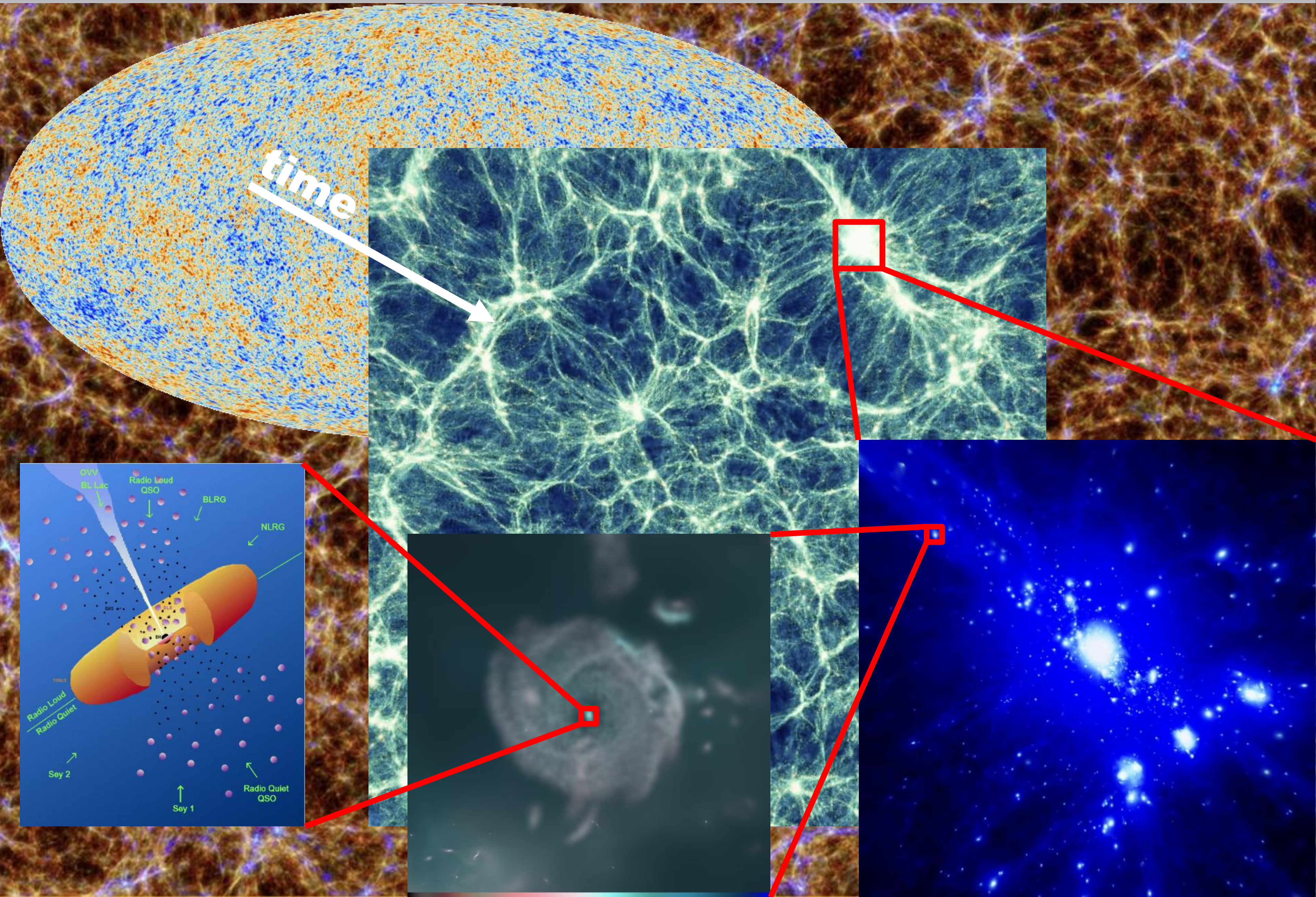
And there is even something more!



Fritz Zicky
(1863 – 1932)
Quelle: Wikipedia

- **1933: Clusters of galaxies: Dark Matter (Dynamics of the galaxies)**

Multi-scale & physics problem!



The N-body problem:

The dynamics of stellar systems can be followed applying newton's law of gravity.

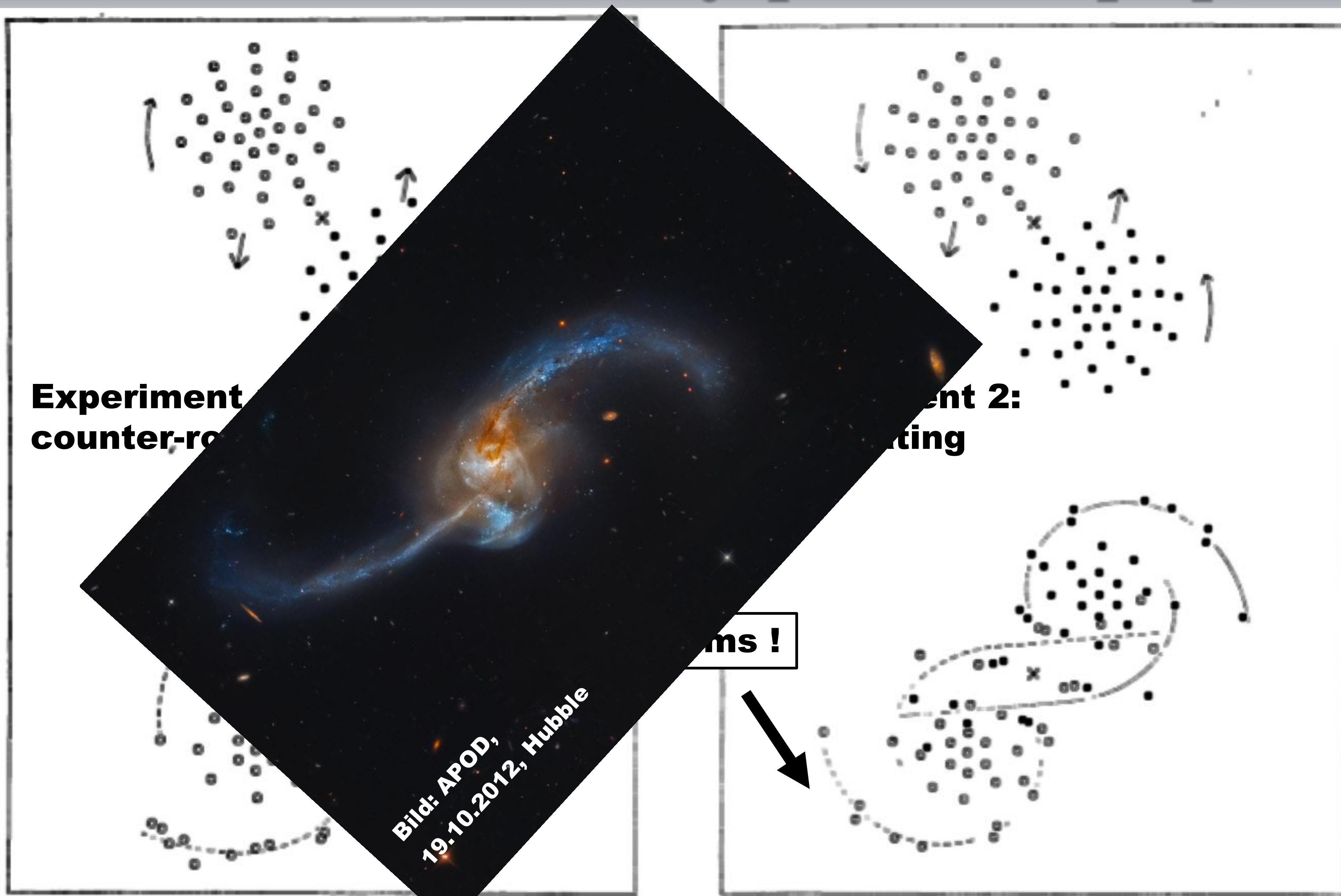
Messier 92 (INT)

Arp 87 (NASA)

Coma
(Dean Riwe, APOD)

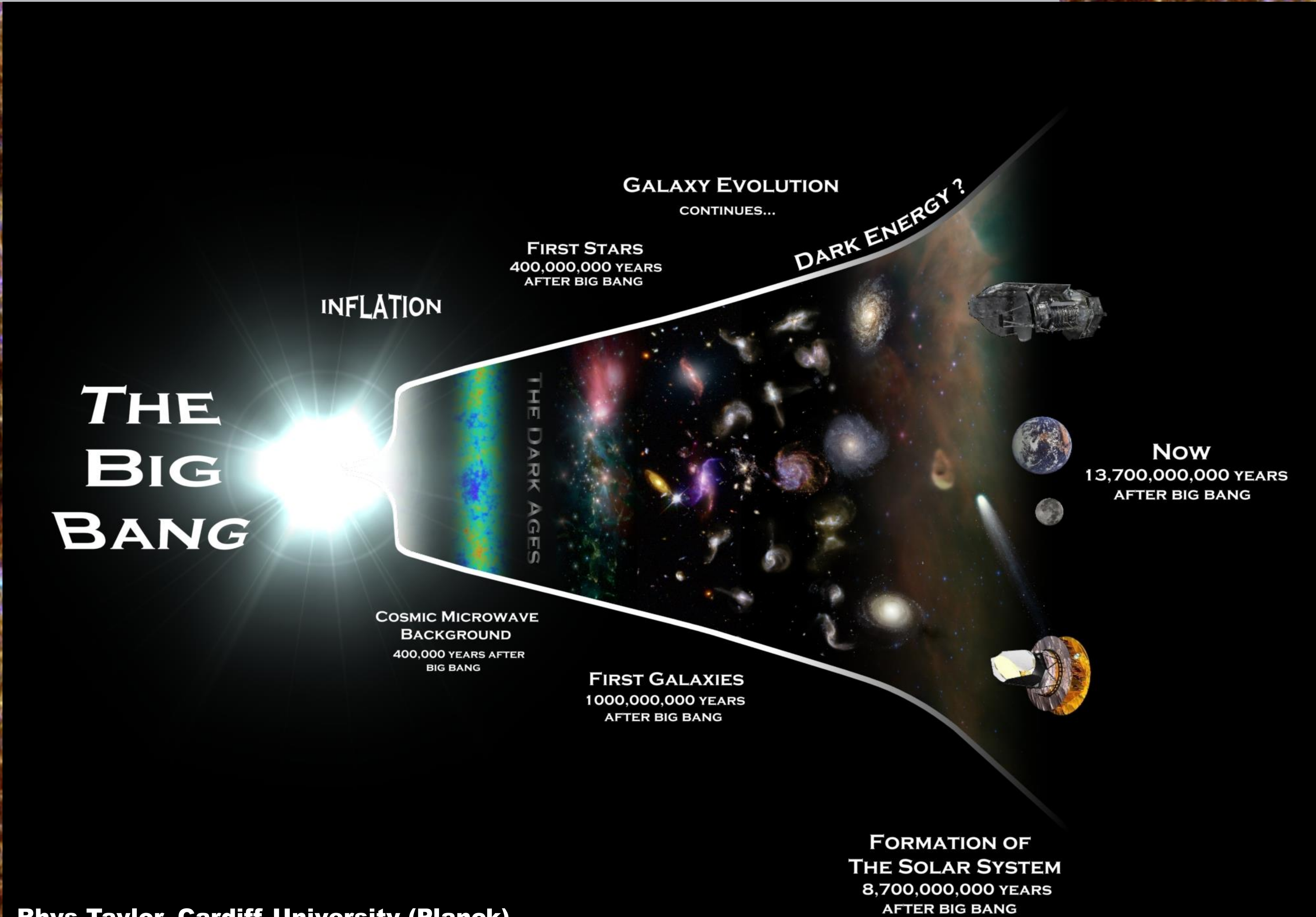
$$\mathbf{F}_i = -\sum_{j=0,\, j\neq i}^{j=N} \frac{G\, m_i\, m_j}{r_{ij}^2}\, \hat{\mathbf{r}}_{ij}$$

First simulation by pen and paper



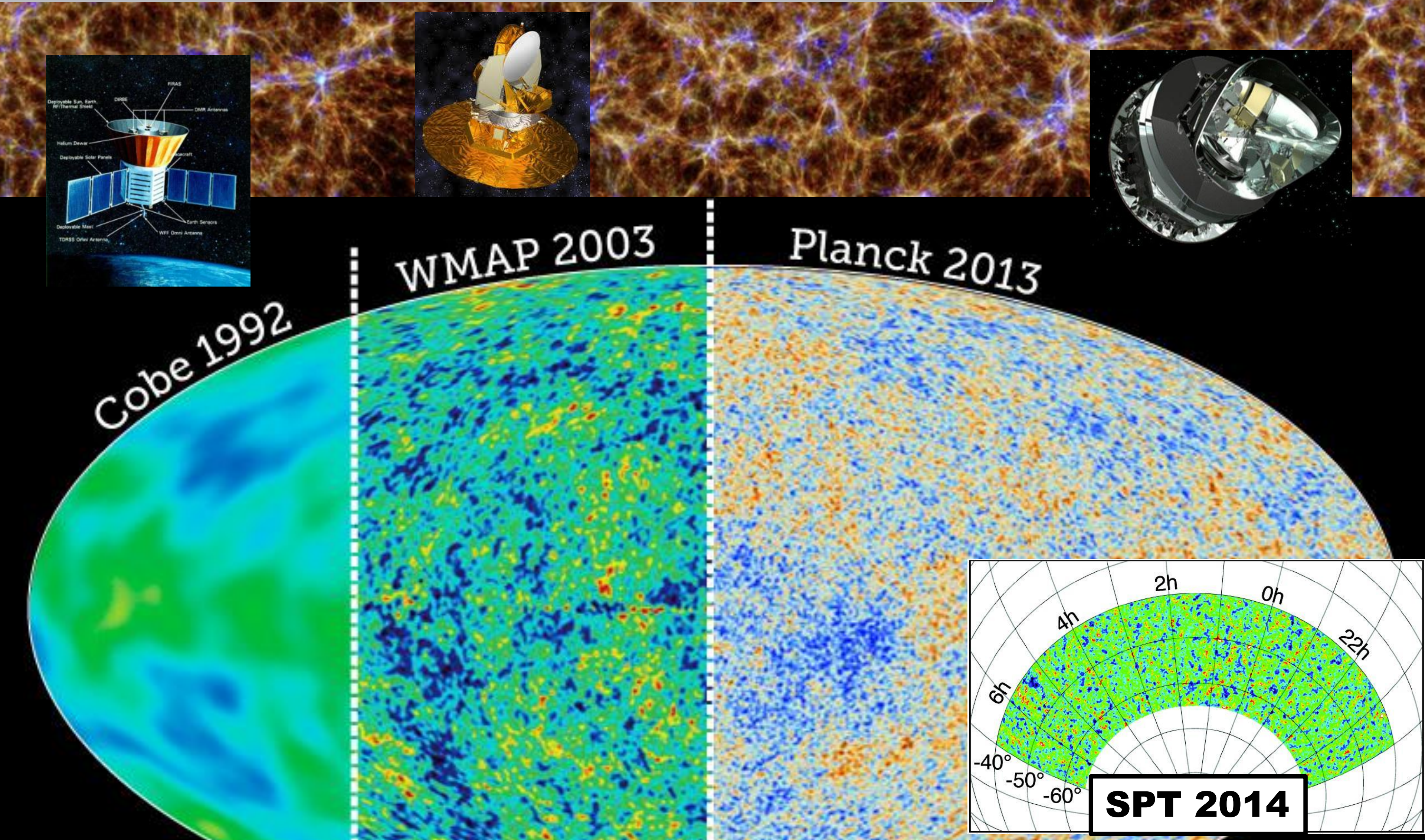
Source: D. ubber

Evolution of the Universe



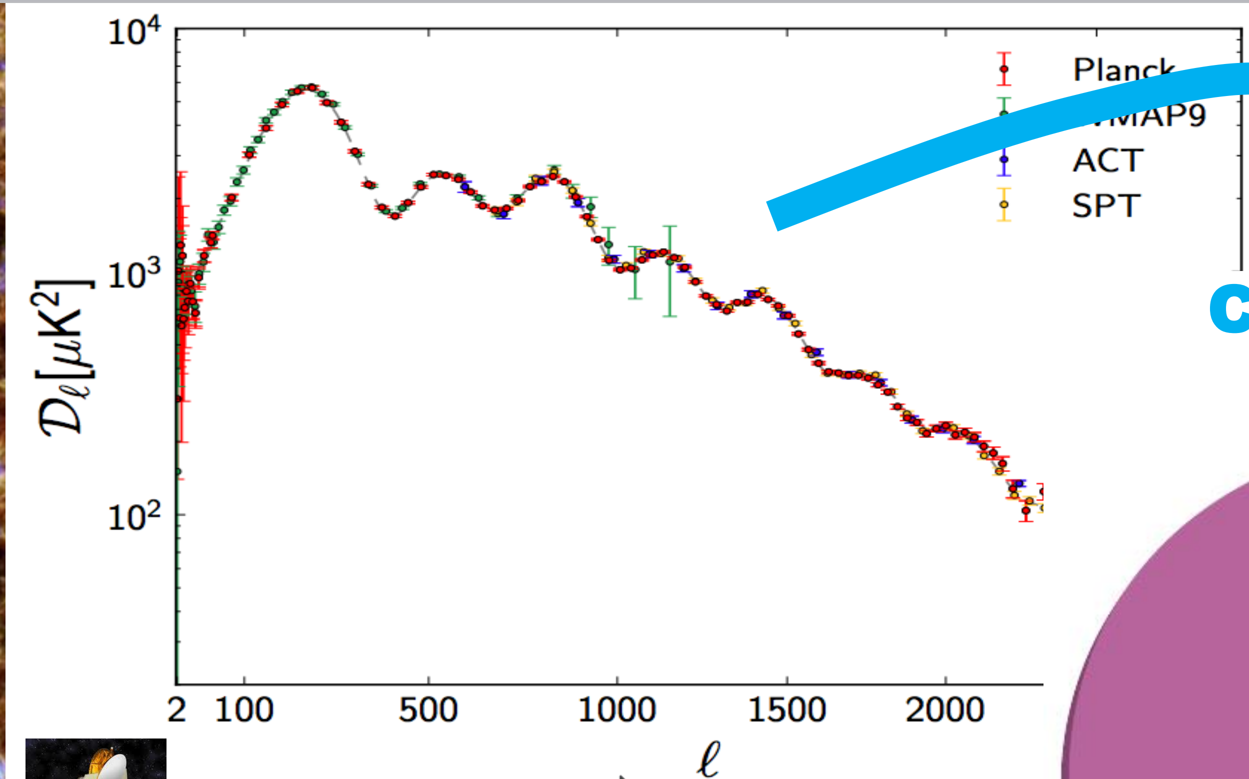
Rhys Taylor, Cardiff University (Planck)

Setting the Framework

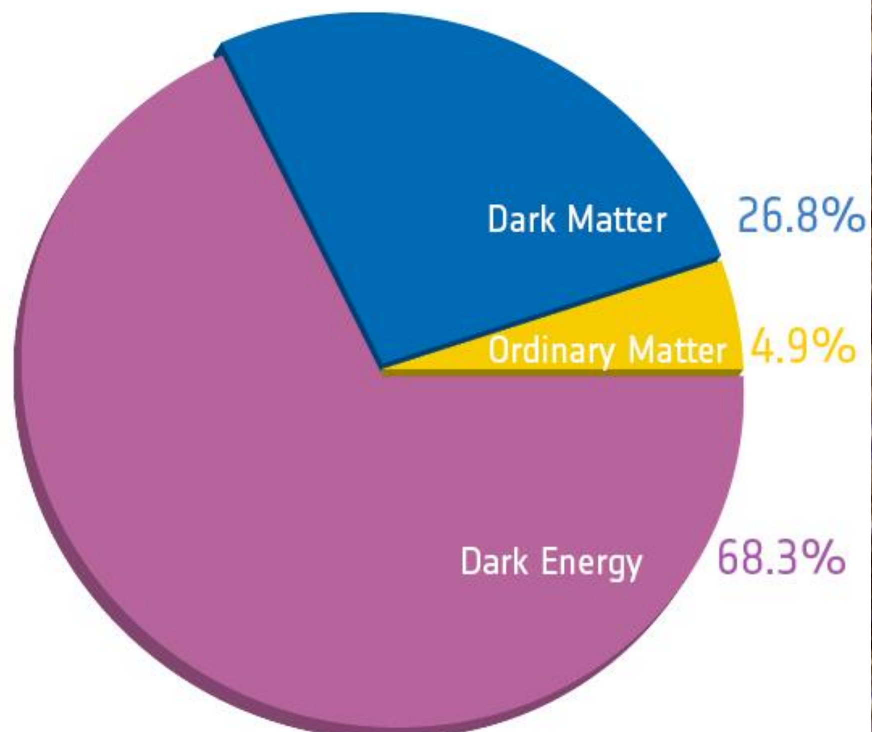


After two decades we can map the initial fluctuations of the structures in the universe with very high precission.

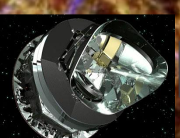
Setting the Framework



Composition of the Universe



WMAP

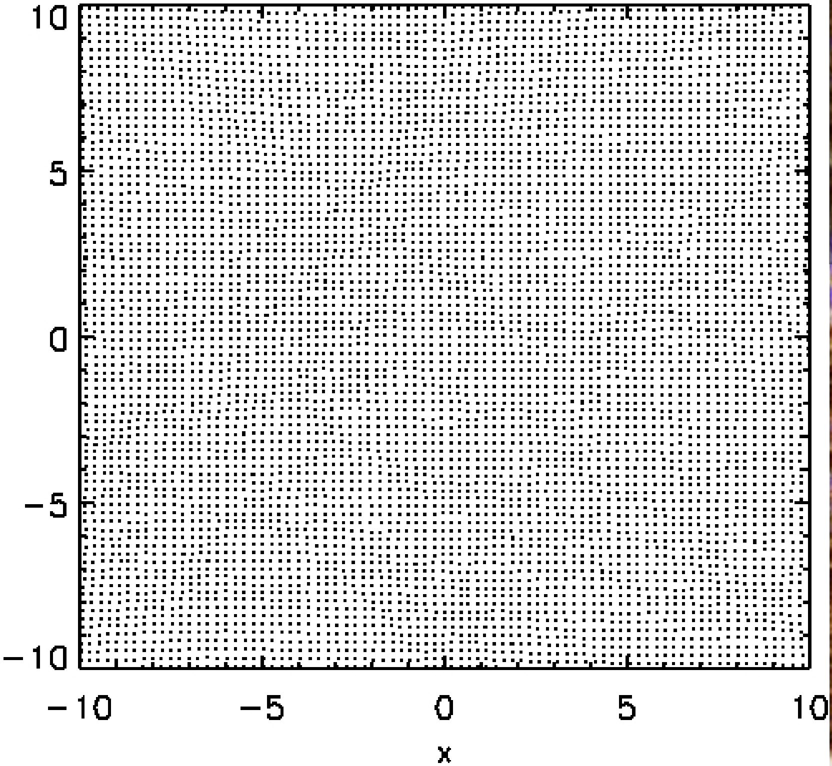


Planck

SPT und ACT

+ Expansion history of the Universe

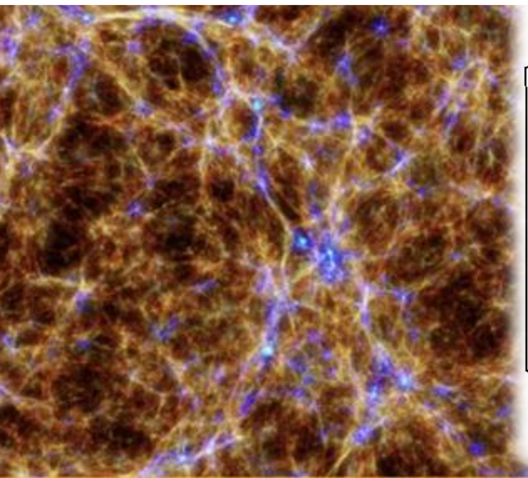
The growth of structures

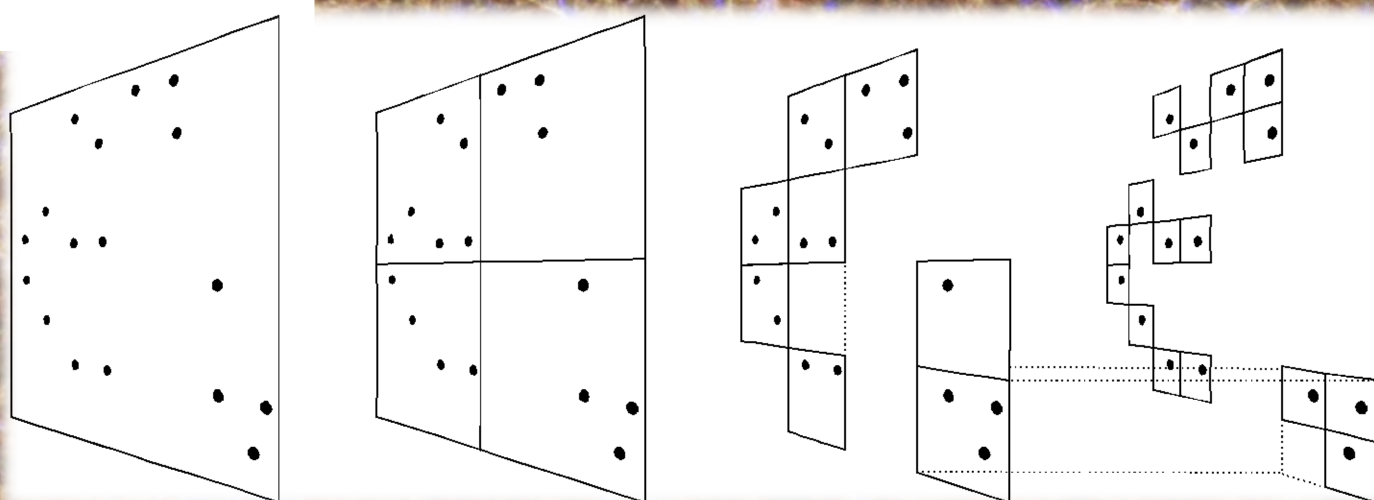


Discretizing the matter distribution and starting from initial Gaussian density perturbation solving the N-body problem:

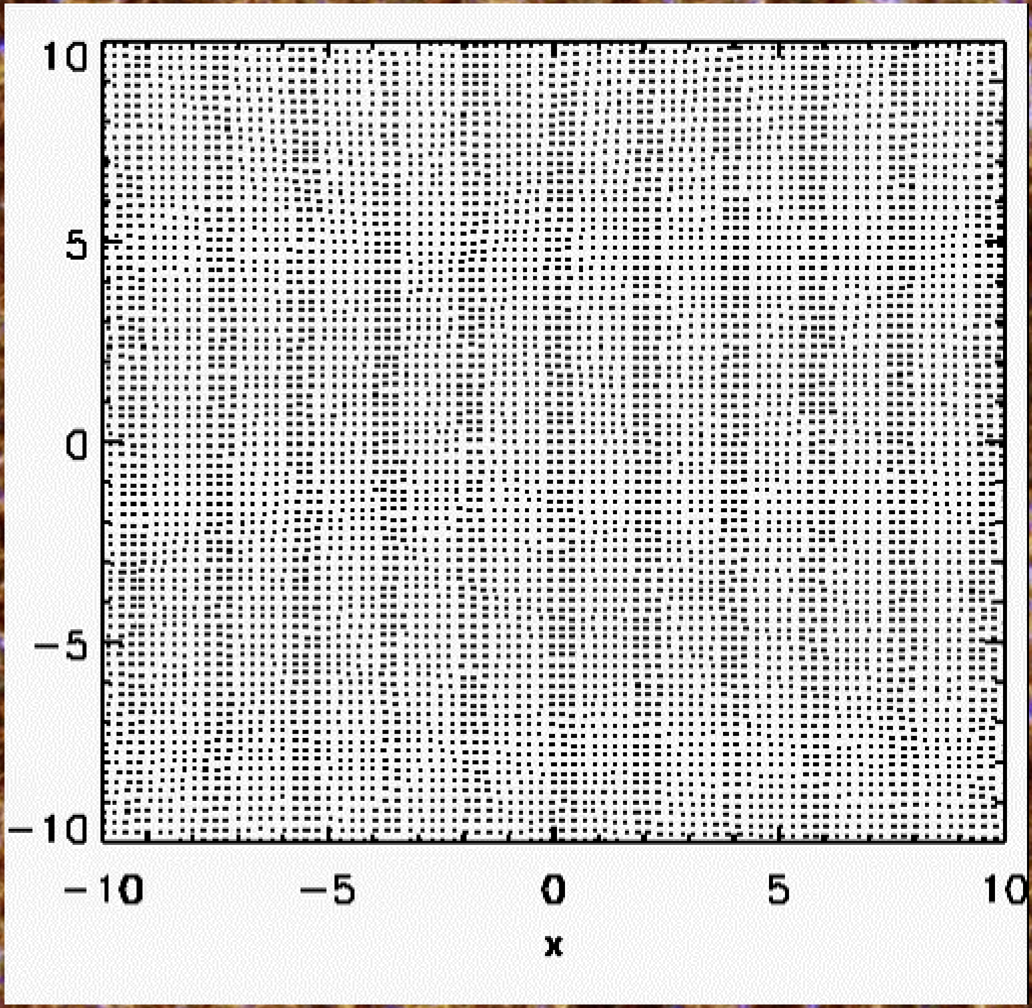
$$\mathbf{F}_i = -\sum_{j=0,\, j\neq i}^{j=N} \frac{G\, m_i\, m_j}{r_{ij}^2}\, \hat{\mathbf{r}}_{ij}$$

Speeding up the n^2 problem by building a Tree structure to approximate the forces.





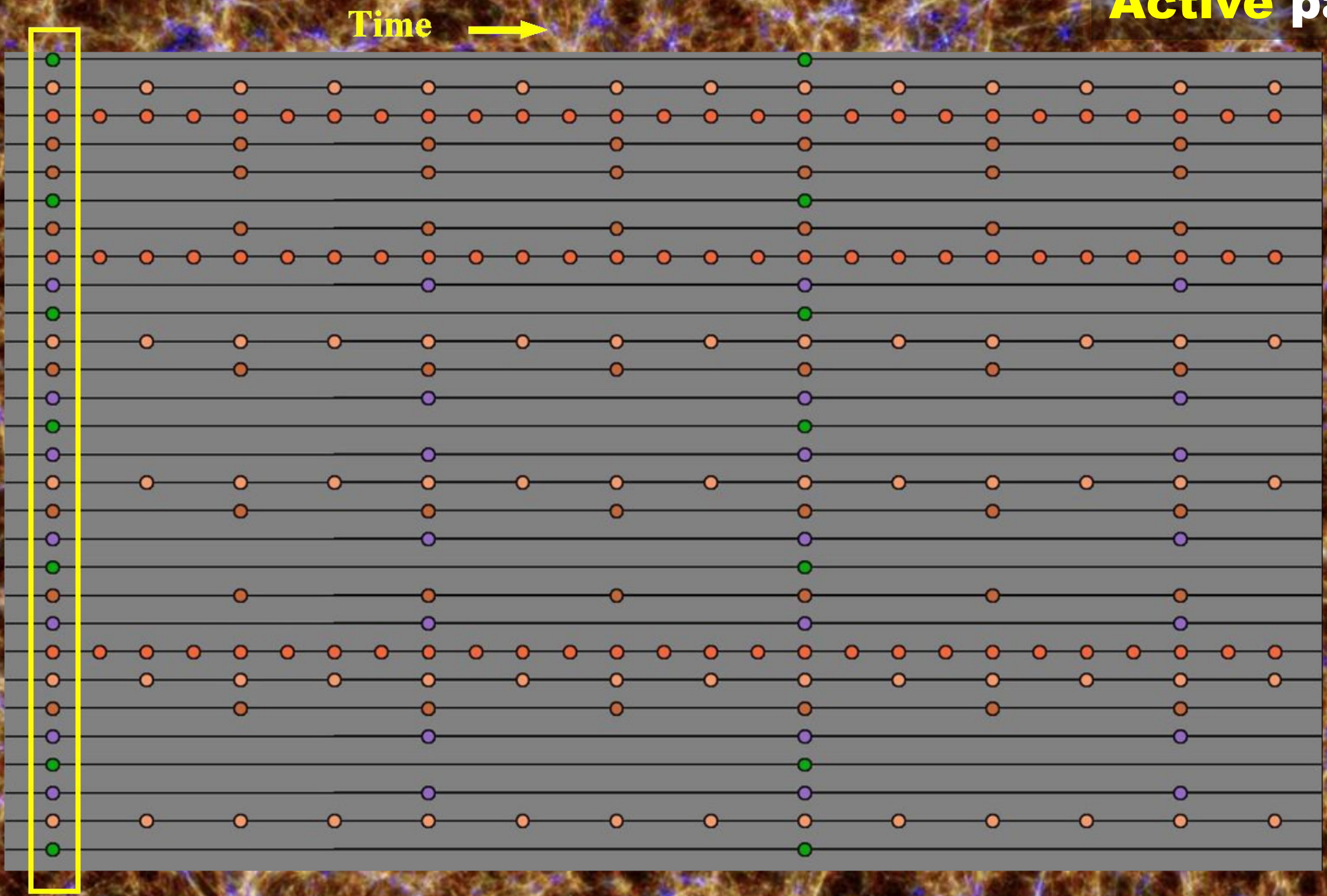
The growth of Structures



⇒ **formation of typical, cosmic structures like voids, filaments and collapsed objects (e.g. galaxies and galaxy clusters)**

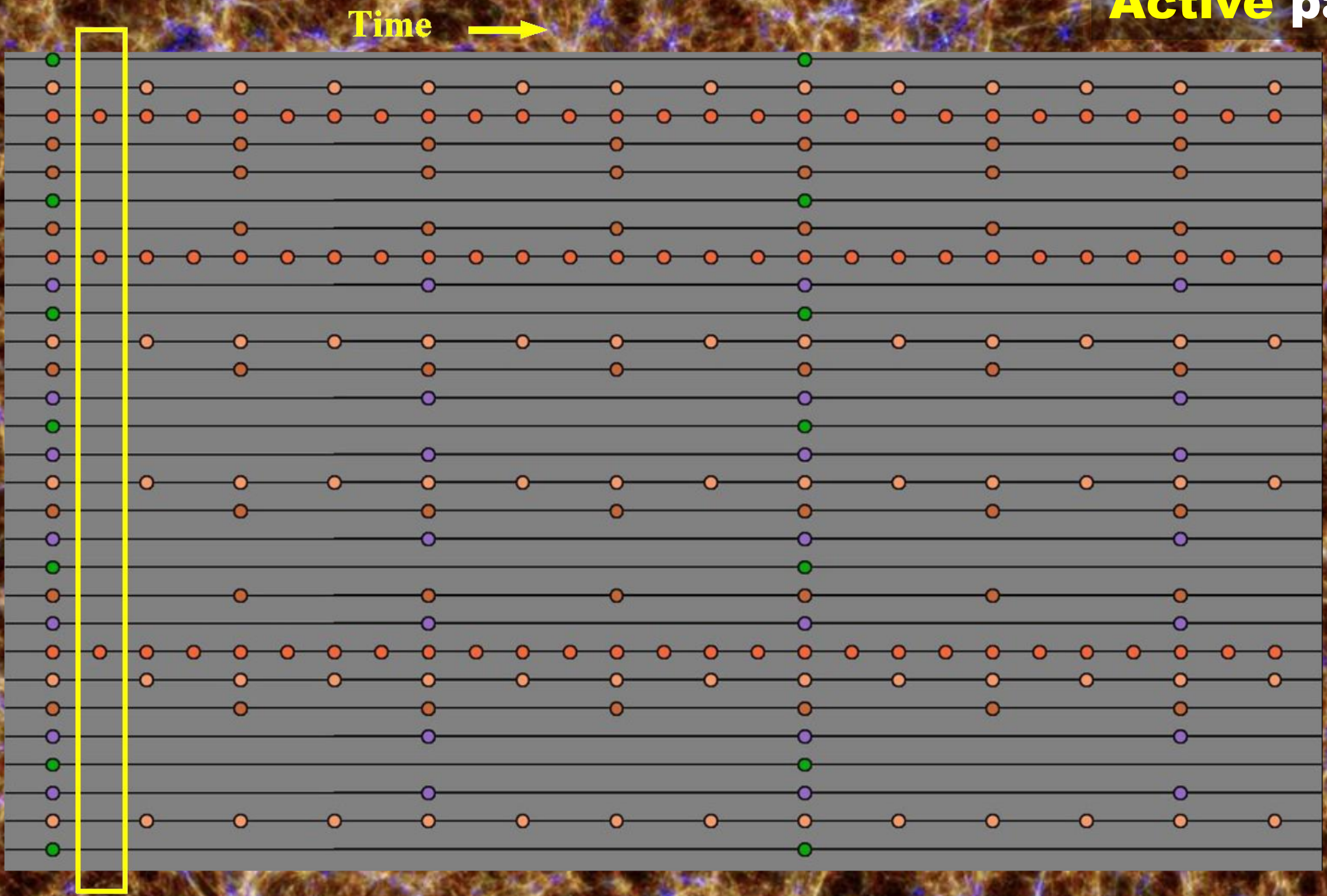
Speedup I: individual time-steps

Active particles:



Speeding up the problem by individual time steps. Force needs to be calculated only for active particles.

Speedup I: individual time-steps

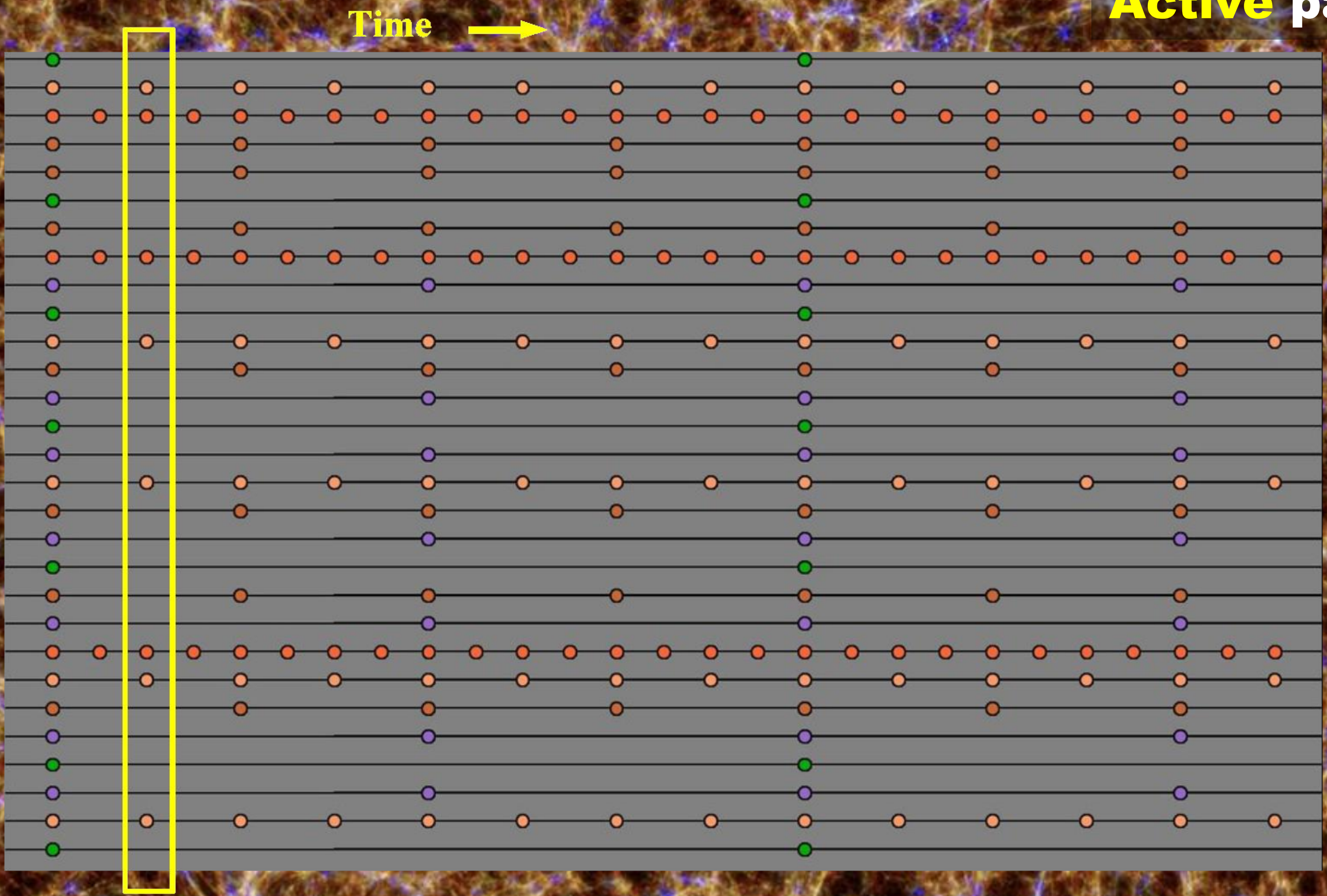


Active particles:

Speeding up the problem by individual time steps. Force needs to be calculated only for active particles.

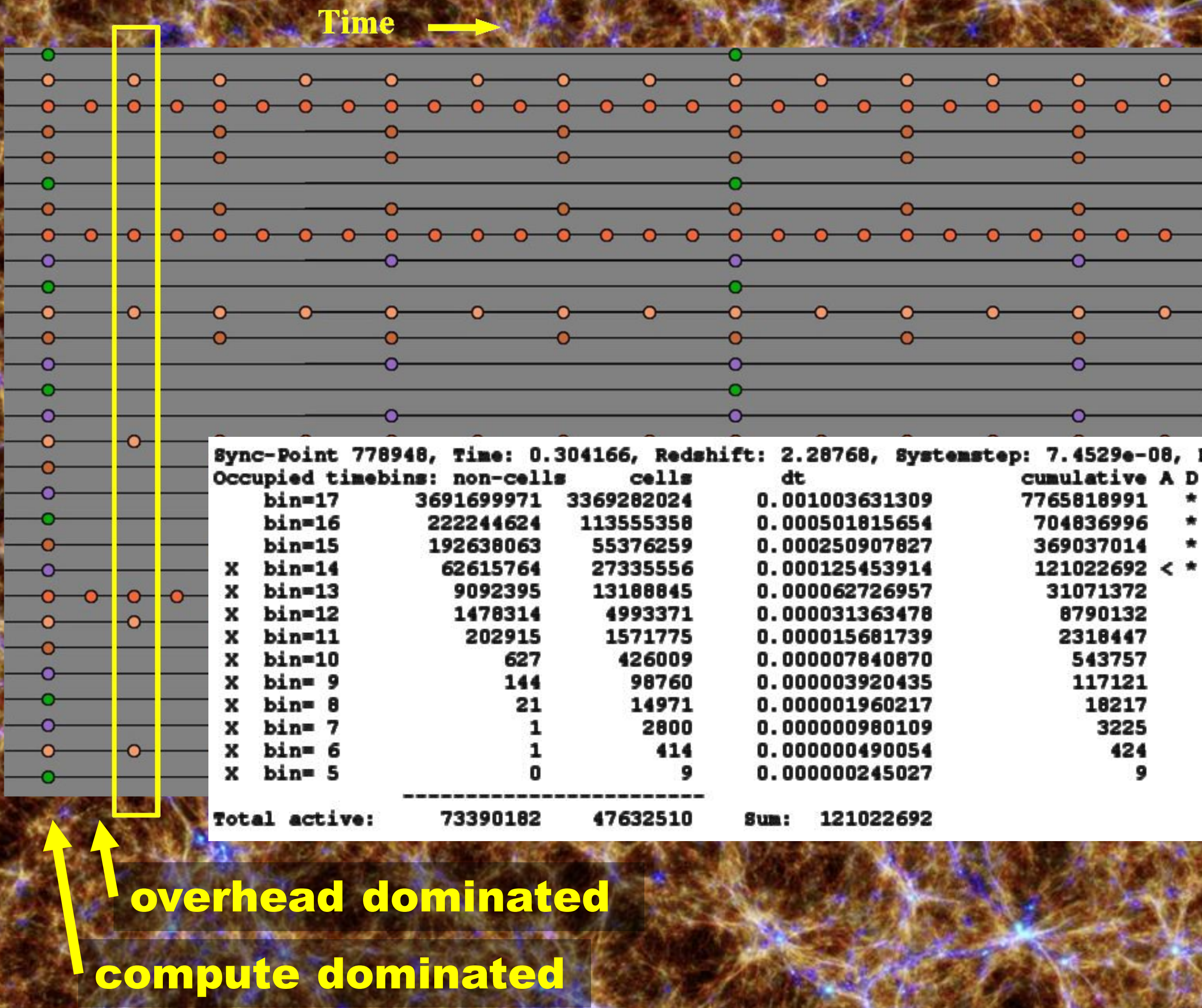
Speedup I: individual time-steps

Active particles:



Speeding up the problem by individual time steps. Force needs to be calculated only for active particles.

Speedup I: individual time-steps

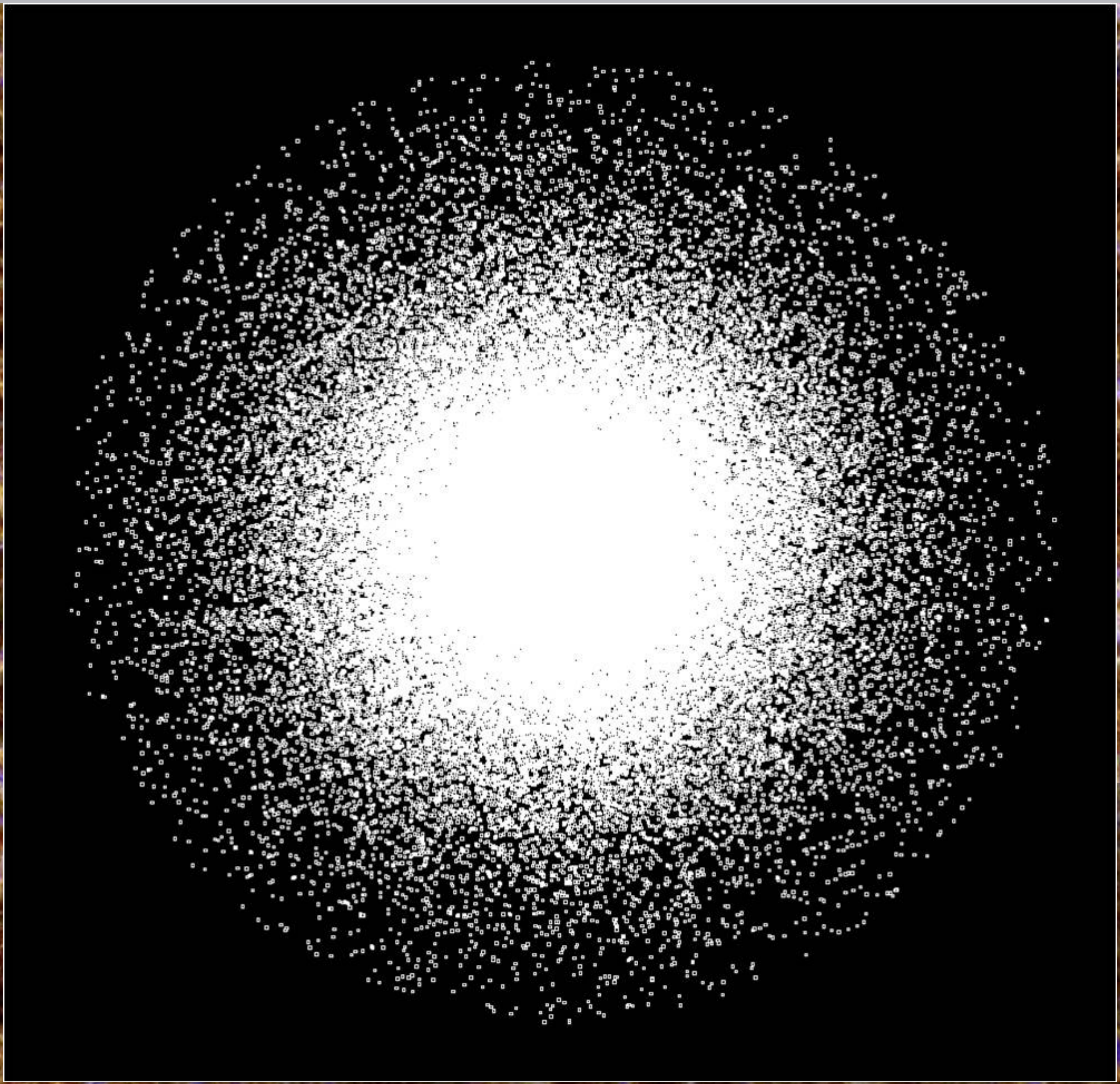


Sync-Point 778948, Time: 0.304166, Redshift: 2.28768, Systemstep: 7.4529e-08, Dloga: 2.45027e-07

Occupied timebins:		non-cells	cells	dt	cumulative	A	D	avg-time	cpu-frac
	bin=17	3691699971	3369282024	0.001003631309	7765818991		*	595.00	7.8%
	bin=16	222244624	113555358	0.000501815654	704836996		*	324.98	4.3%
	bin=15	192638063	55376259	0.000250907827	369037014		*	260.81	6.9%
X	bin=14	62615764	27335556	0.000125453914	121022692	<	*	273.60	14.4%
X	bin=13	9092395	13188845	0.000062726957	31071372			130.15	13.7%
X	bin=12	1478314	4993371	0.000031363478	8790132			68.56	14.4%
X	bin=11	202915	1571775	0.000015681739	2318447			31.35	13.2%
X	bin=10	627	426009	0.000007840870	543757			1.60	1.3%
X	bin= 9	144	98760	0.000003920435	117121			1.16	1.9%
X	bin= 8	21	14971	0.000001960217	18217			0.95	3.2%
X	bin= 7	1	2800	0.000000980109	3225			0.60	4.1%
X	bin= 6	1	414	0.000000490054	424			0.39	5.2%
X	bin= 5	0	9	0.000000245027	9			0.35	9.5%

Total active: 73390182 47632510 Sum: 121022692

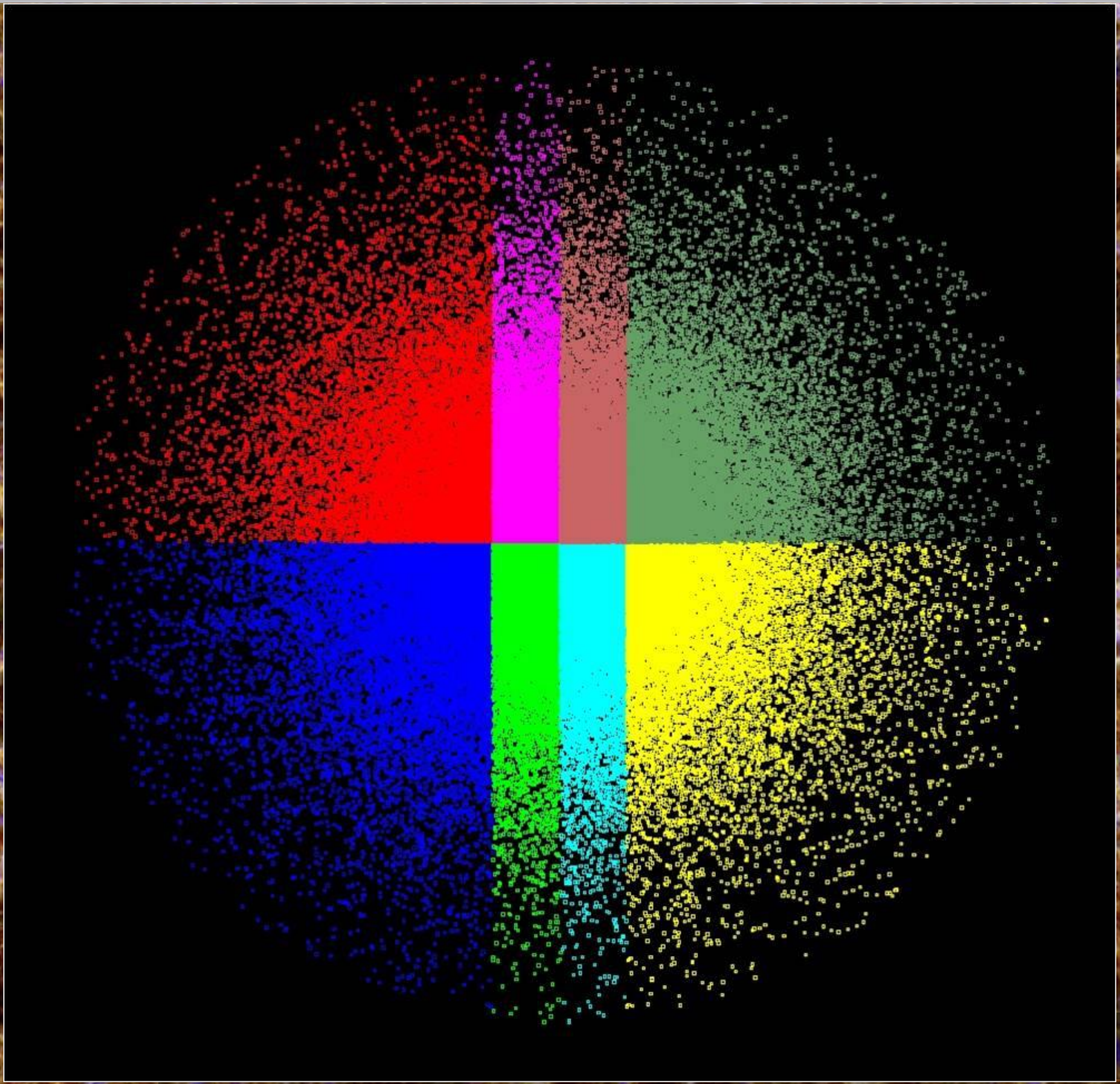
Speedup II: distributing the work



Particles to be distributed among MPI ranks:

Distributing particles (memory) and work load (balance) simultaneously is a big challenge !

Speedup II: distributing the work

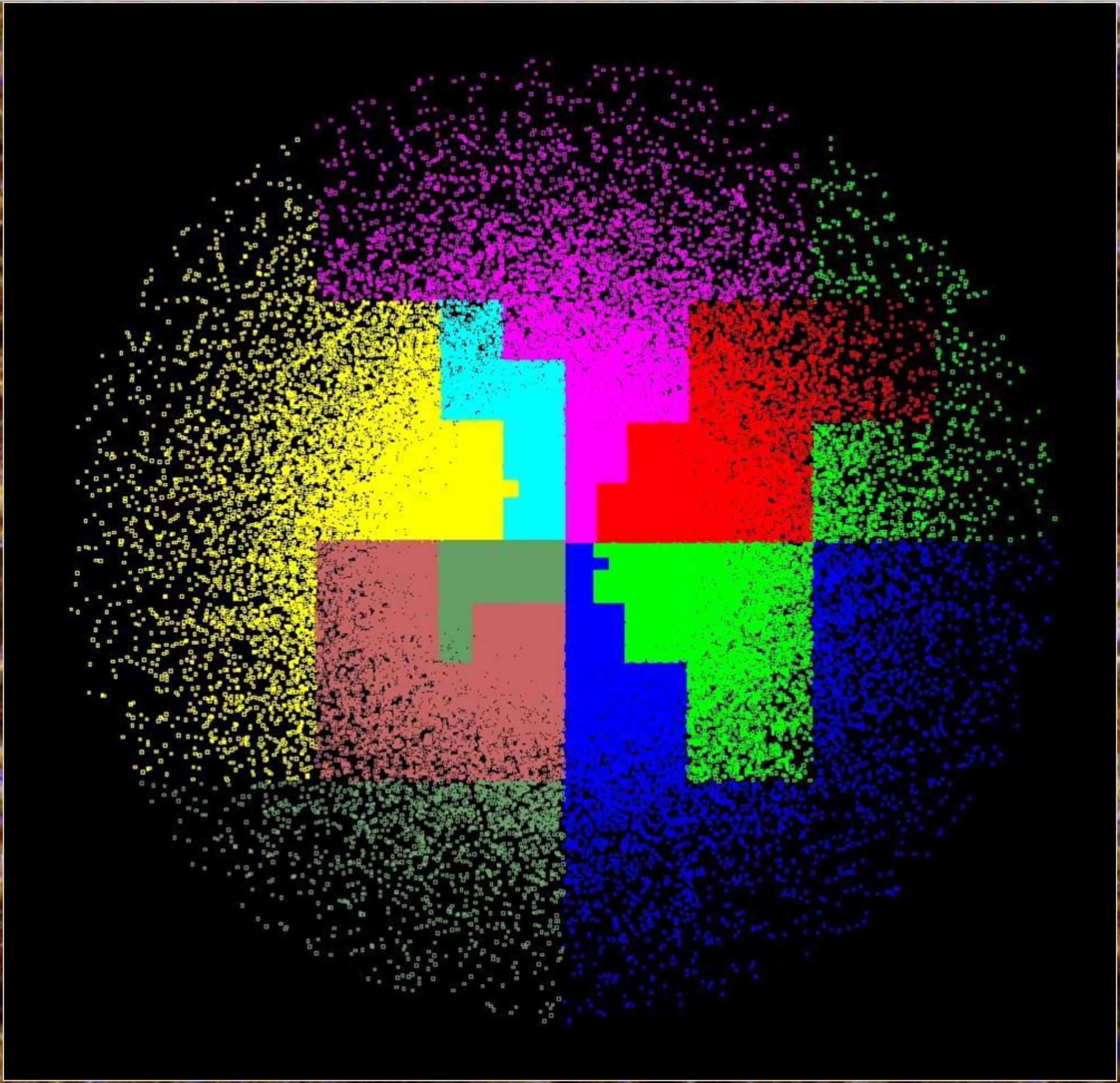


Example:
Gadget1

Large memory imbalance!

Distributing particles (memory) and work load (balance) simultaneously is a big challenge !

Speedup II: distributing the work

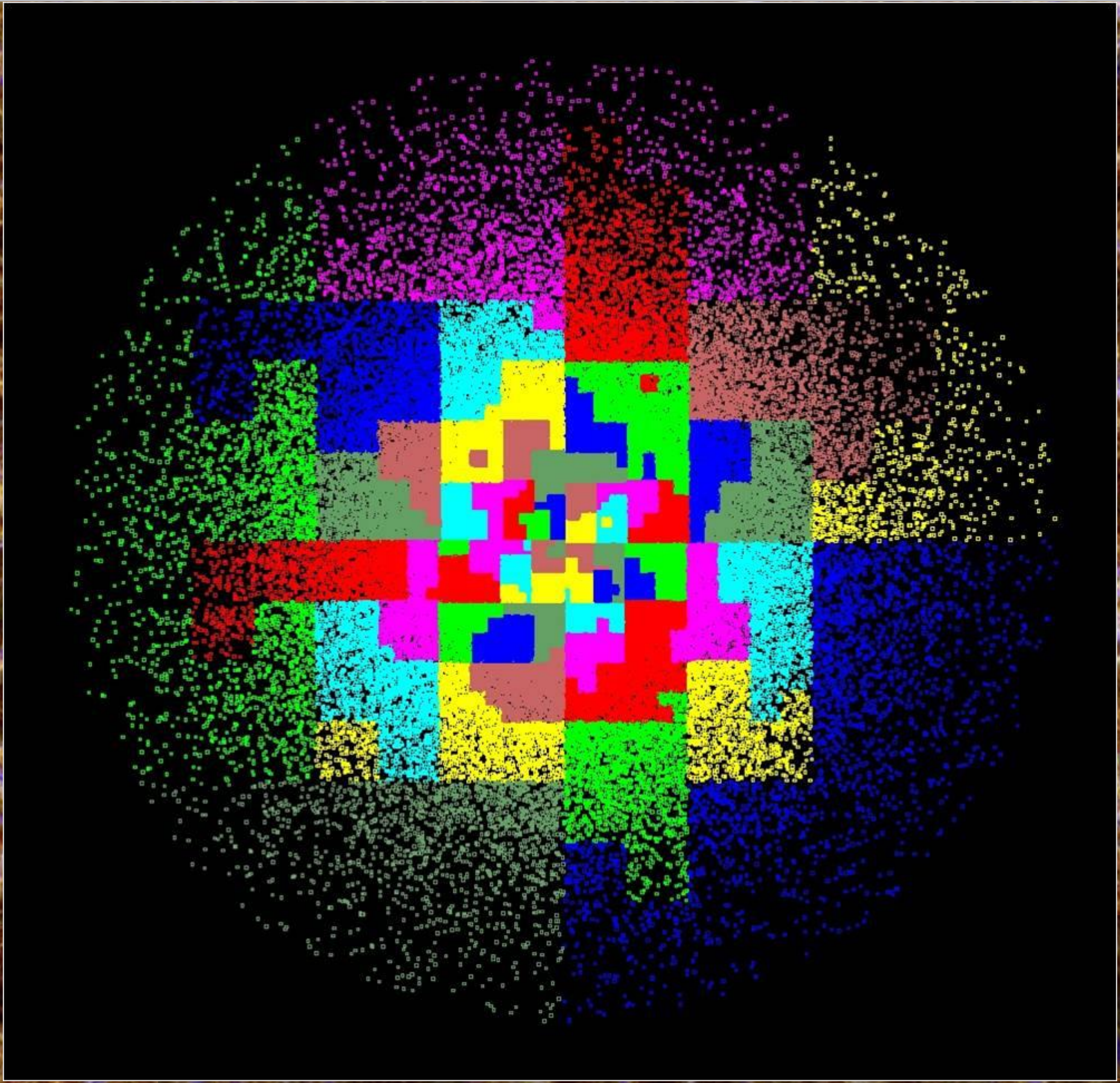


Example:
Gadget2

Perfect memory balance, but imperfect workload!

Distributing particles (memory) and work load (balance) simultaneously is a big challenge !

Speedup II: distributing the work



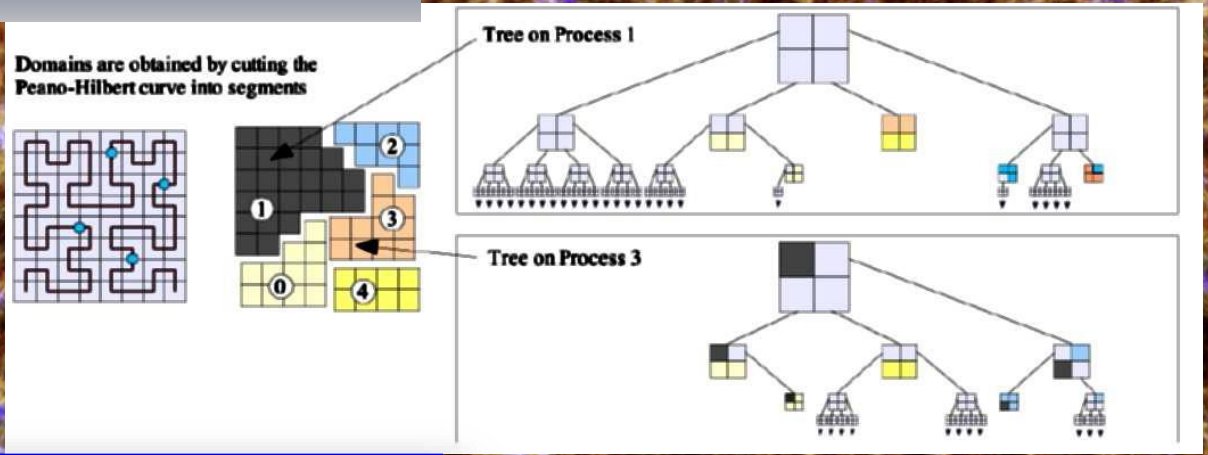
Example: Gadget3

Good balance (work/load) but communication overhead!

Distributing particles (memory) and work load (balance) simultaneously is a big challenge !

Speedup III: Tree-PM

Splitting gravity in short range (Tree) and long range (FFT)



$$\Phi_{\vec{k}} = \Phi^{\text{long}}_{\vec{k}} + \Phi^{\text{short}}_{\vec{k}}$$

$$\Phi^{\text{short}}(\vec{x}) = -G \sum_i \frac{m_i}{\vec{r}_i} \operatorname{erfc}\left(\frac{\vec{r}_i}{2r_s}\right)$$

$$\Phi^{\text{long}}_{\vec{k}} = \Phi_{\vec{k}} \exp(-\vec{k}^2 r_s^2)$$

- density on the grid

$$\rho_m = \frac{1}{h^3} \sum_i m_i W(\vec{x}_i - \vec{x}_m).$$

- solve for Φ using FFT methods

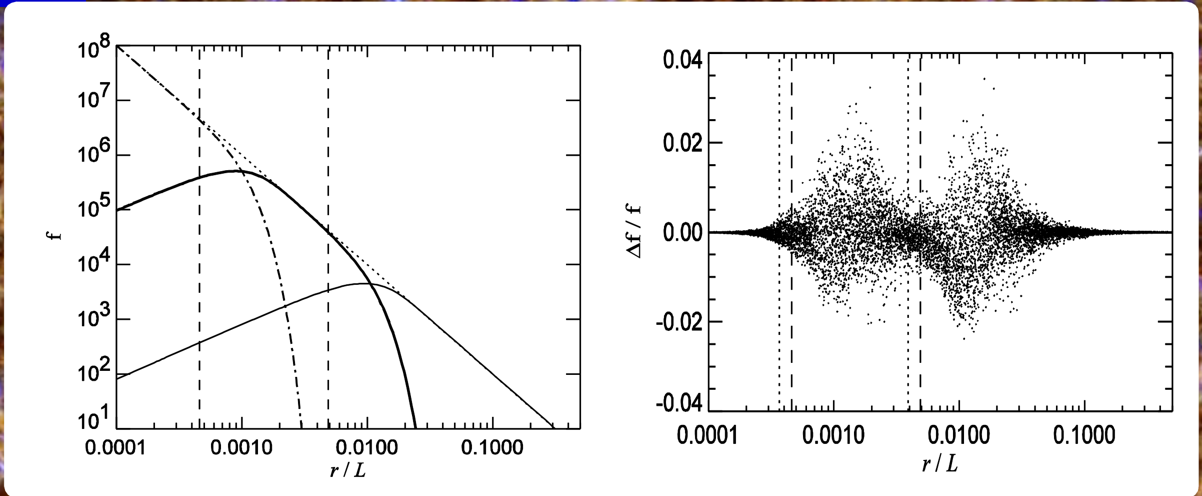
$$\Phi(\vec{x}) = \int g(\vec{x} - \vec{x}')\rho(\vec{x}')\mathrm{d}\vec{x}'$$

- calculate force using finite differences

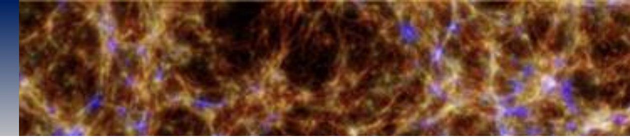
$$f^{(x)}_{i,j,k} = -\frac{\Phi_{i+1,j,k} - \Phi_{i-1,j,k}}{2h}.$$

- interpolate force back to particles

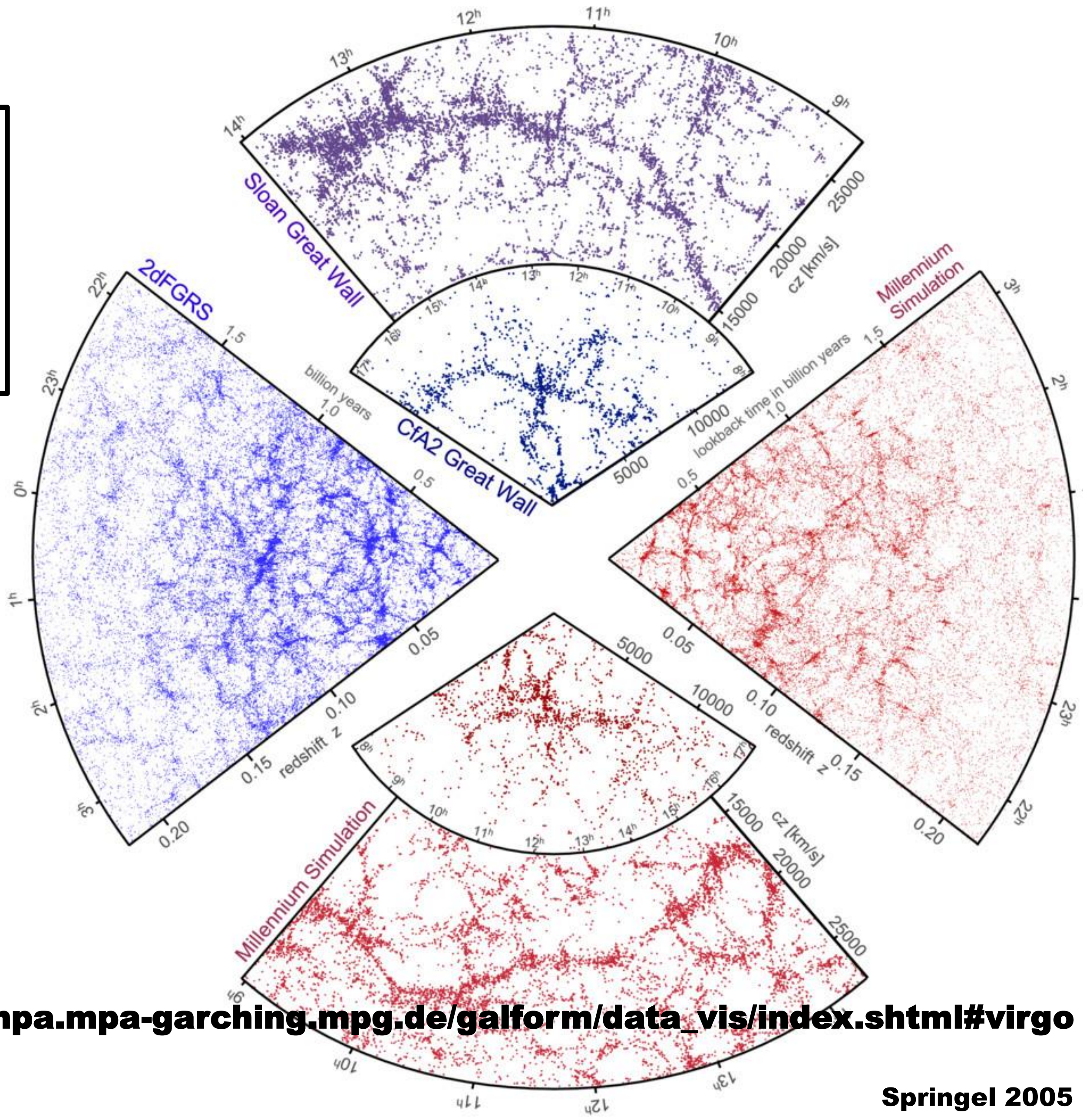
$$\vec{f}(\vec{x}_i) = \sum_m W(\vec{x}_i - \vec{x}_m)\vec{f}_m,$$



The growth of Structures



The distribution of galaxies in the universe and in simulations.



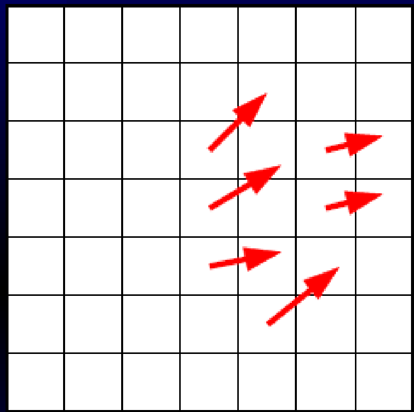
http://wwwmpa.mpa-garching.mpg.de/galform/data_vis/index.shtml#virgo

Springel 2005

How to treat the visible Universe:

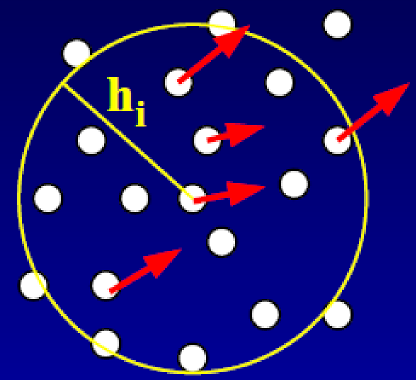
Eulerian

discretized space



Lagrangian

discretized mass



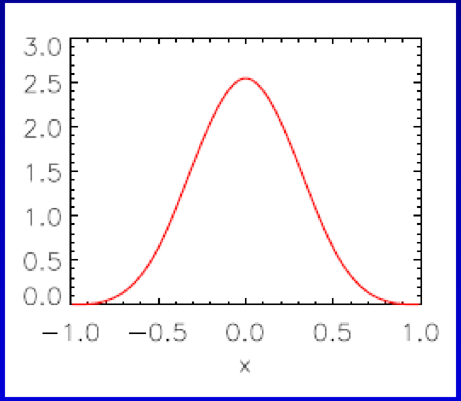
SPH

kernel estimate

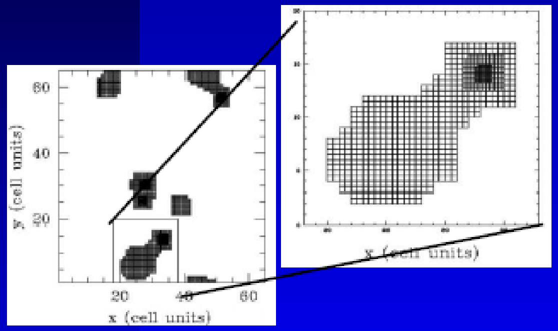
$$\langle A(\mathbf{r})\rangle = \int W(\mathbf{r}-\mathbf{r}',h)\,A(\mathbf{r}')\,\mathrm{d}^3r'$$

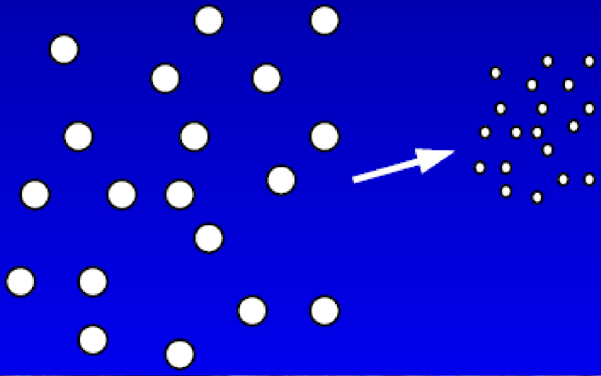
$$\mathrm{d}^3r' \mapsto \frac{m_j}{\rho_j}$$

$$\langle A_i\rangle = \sum_{j=1}^{N}\frac{m_j}{\rho_j}A_j\,W(\mathbf{r}_{ij};h_i)$$



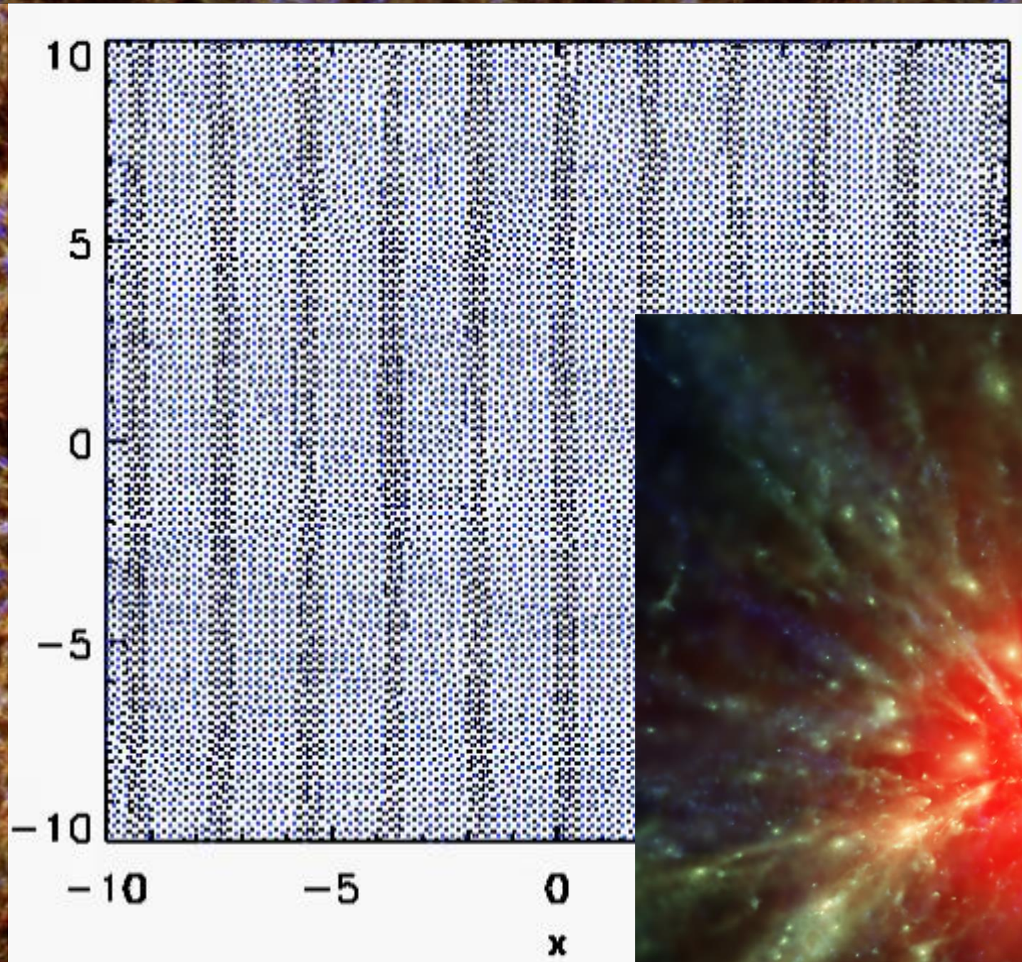
Collapse:

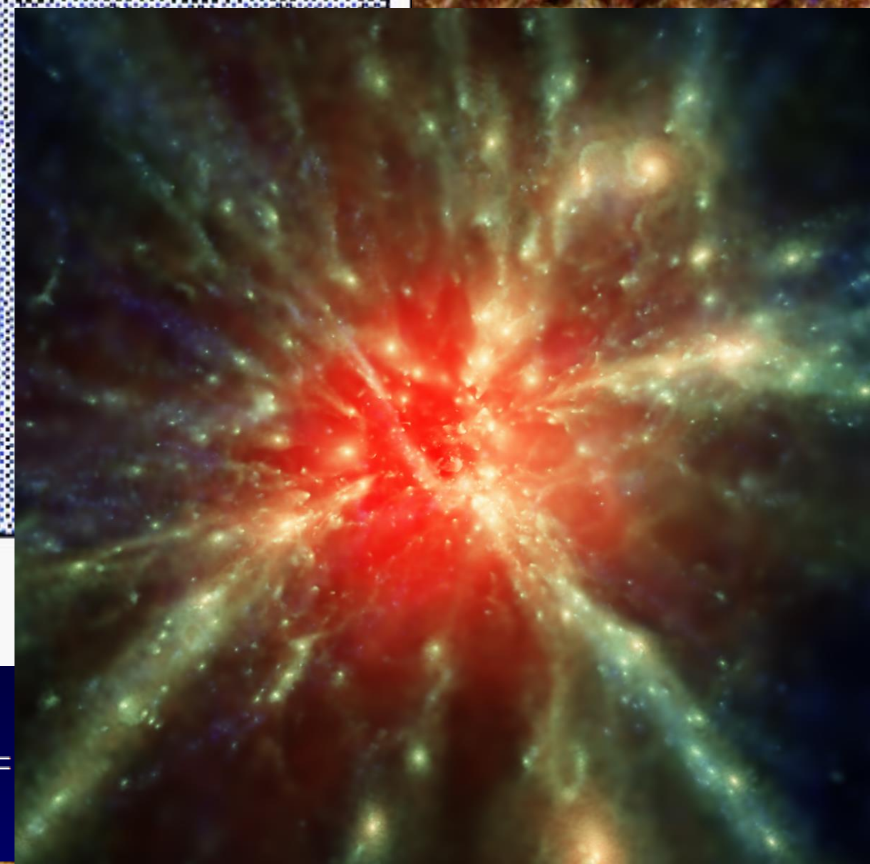




How to treat the visible Universe:

Now we can follow the equation of motion of two species of particle, but for the gas additional terms appear.



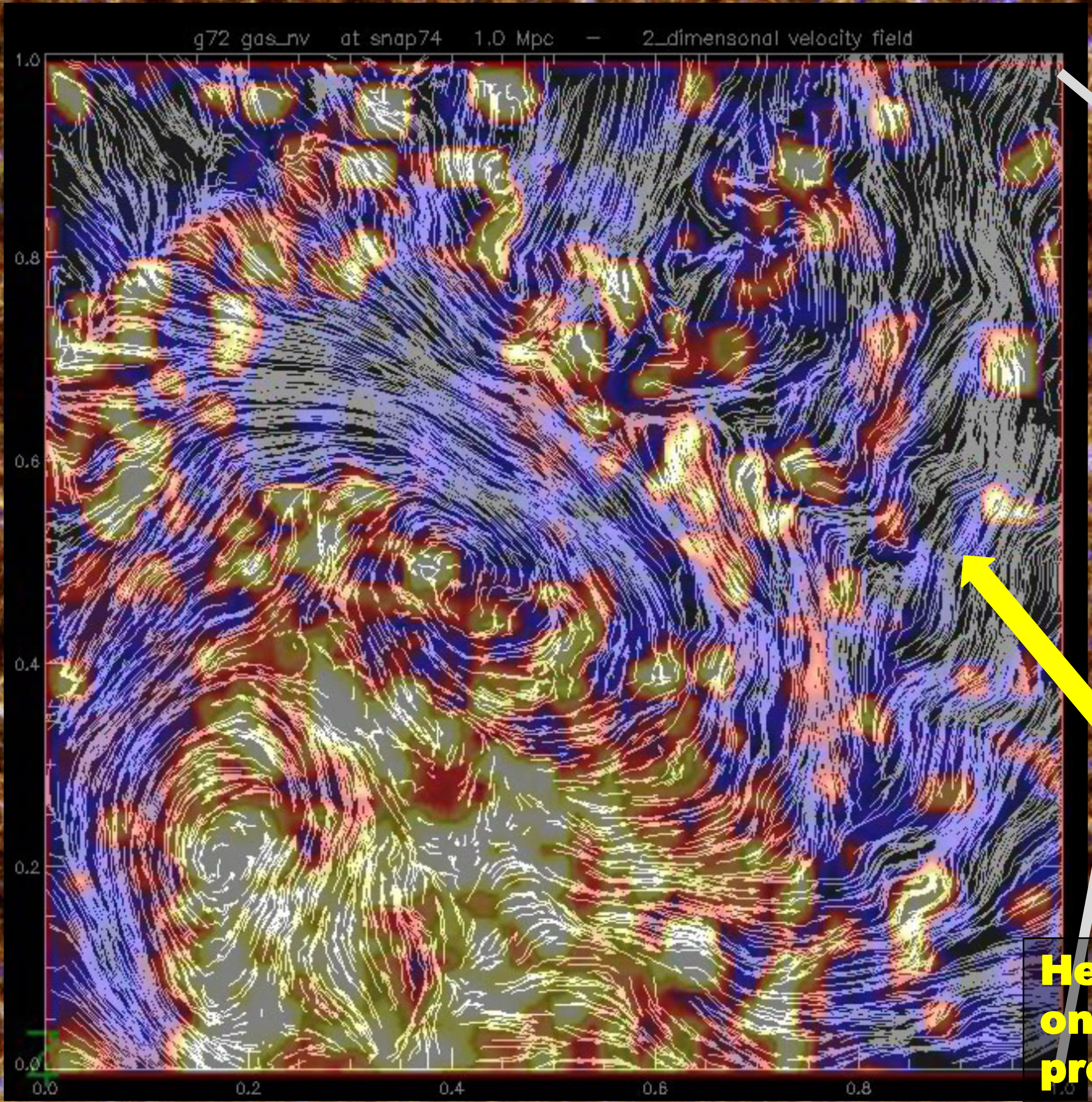


$$\left(\frac{\mathrm{d}\vec{v}}{\mathrm{d}t}\right)_{\mathrm{DM}} = -\vec{\nabla}\Phi, \qquad \left(\frac{\mathrm{d}\vec{v}}{\mathrm{d}t}\right)_{gas} =$$

How to treat the baryons:

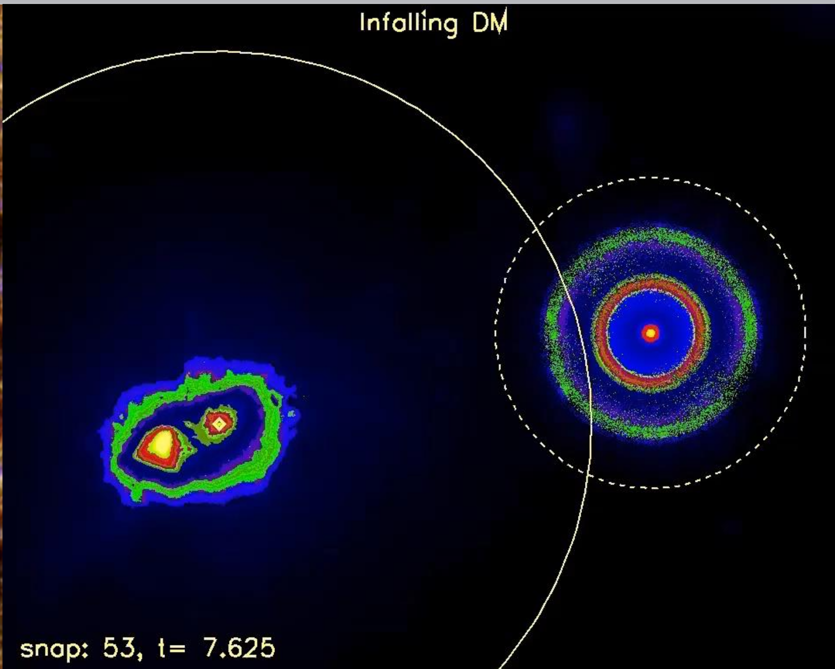
How to treat the baryons:

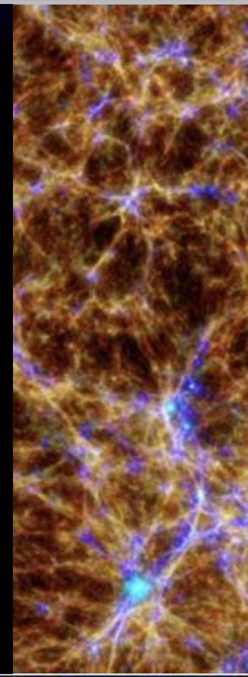
1 Mpc

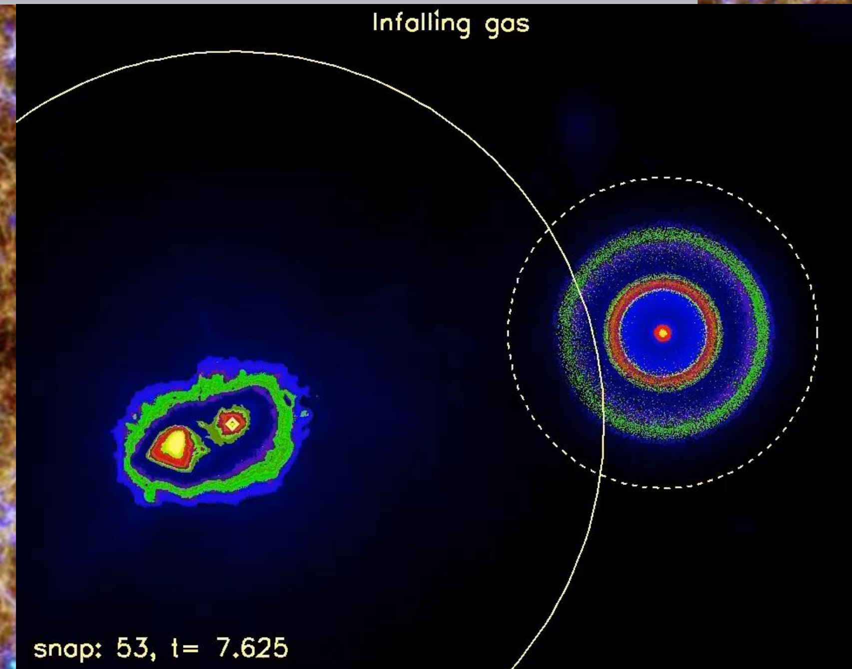


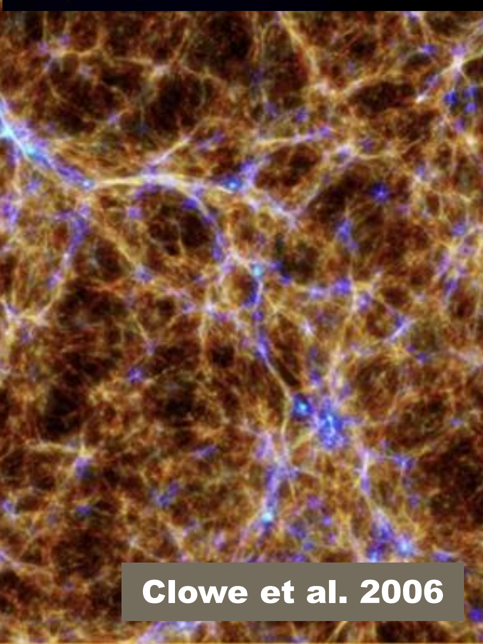
Here, many details on the actual fluid properties go in.

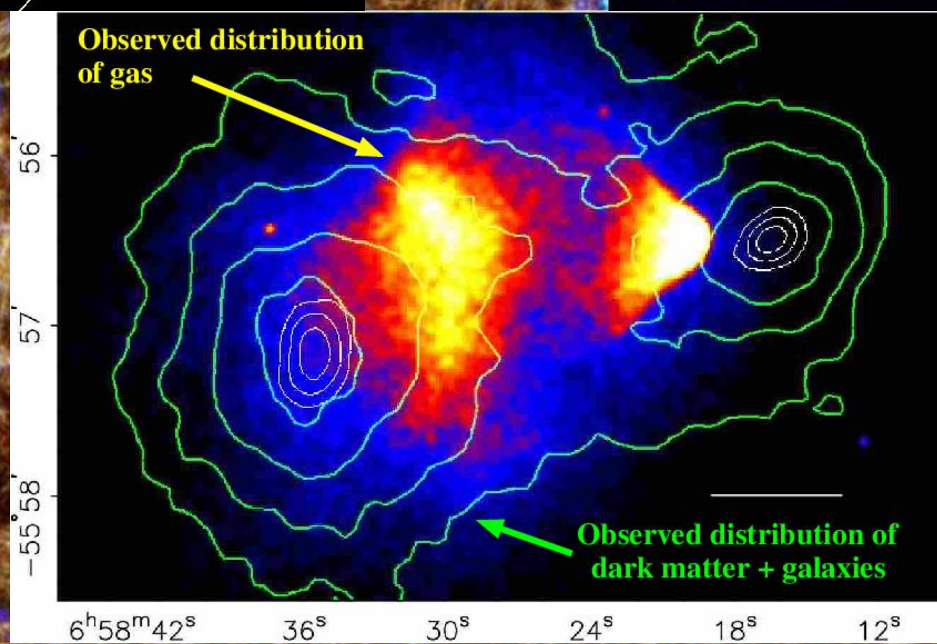
Direct observation of Dark Matter:

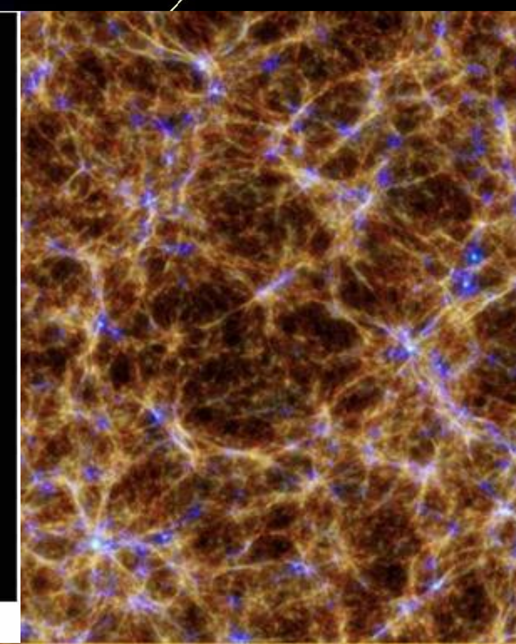






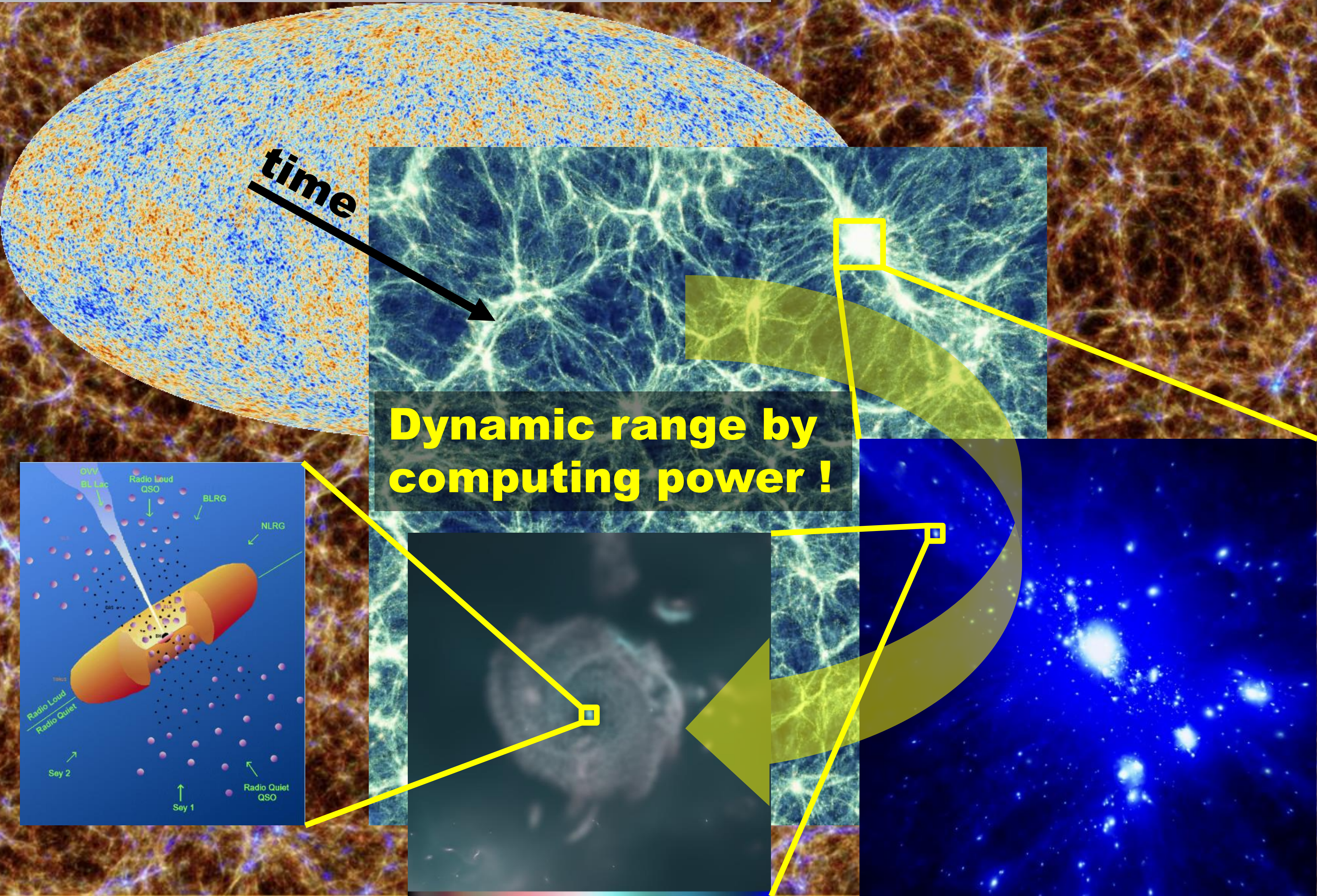




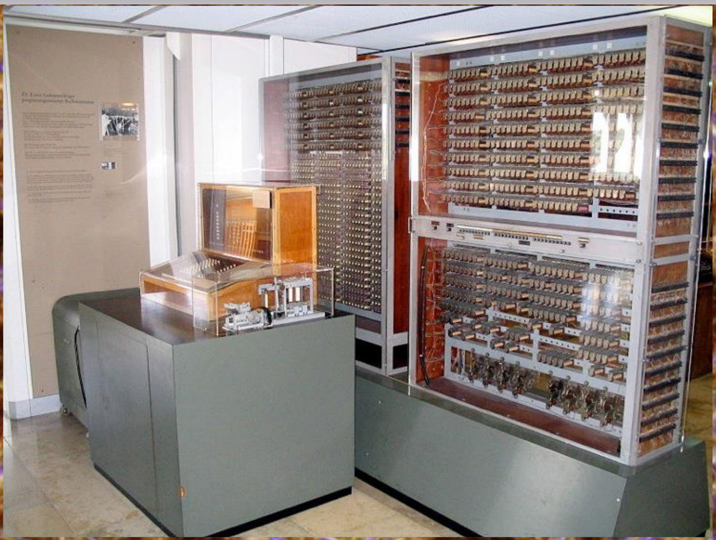


Clowe et al. 2006

How to get bigger ?

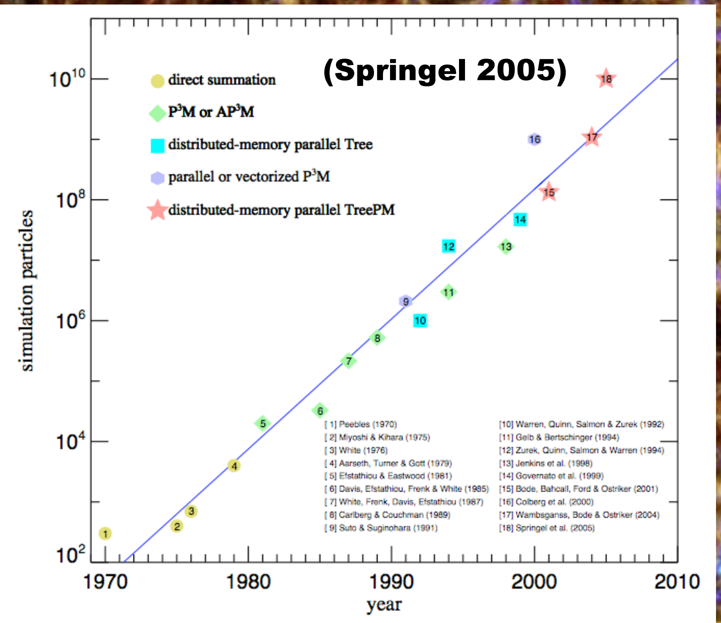


Growing computing power helps



Z3, 1941-1943 (Germany)
5 Computations per second

SuperMUC, 2013- (München)
3 Petaflops ($3*10^{15}$ per second)

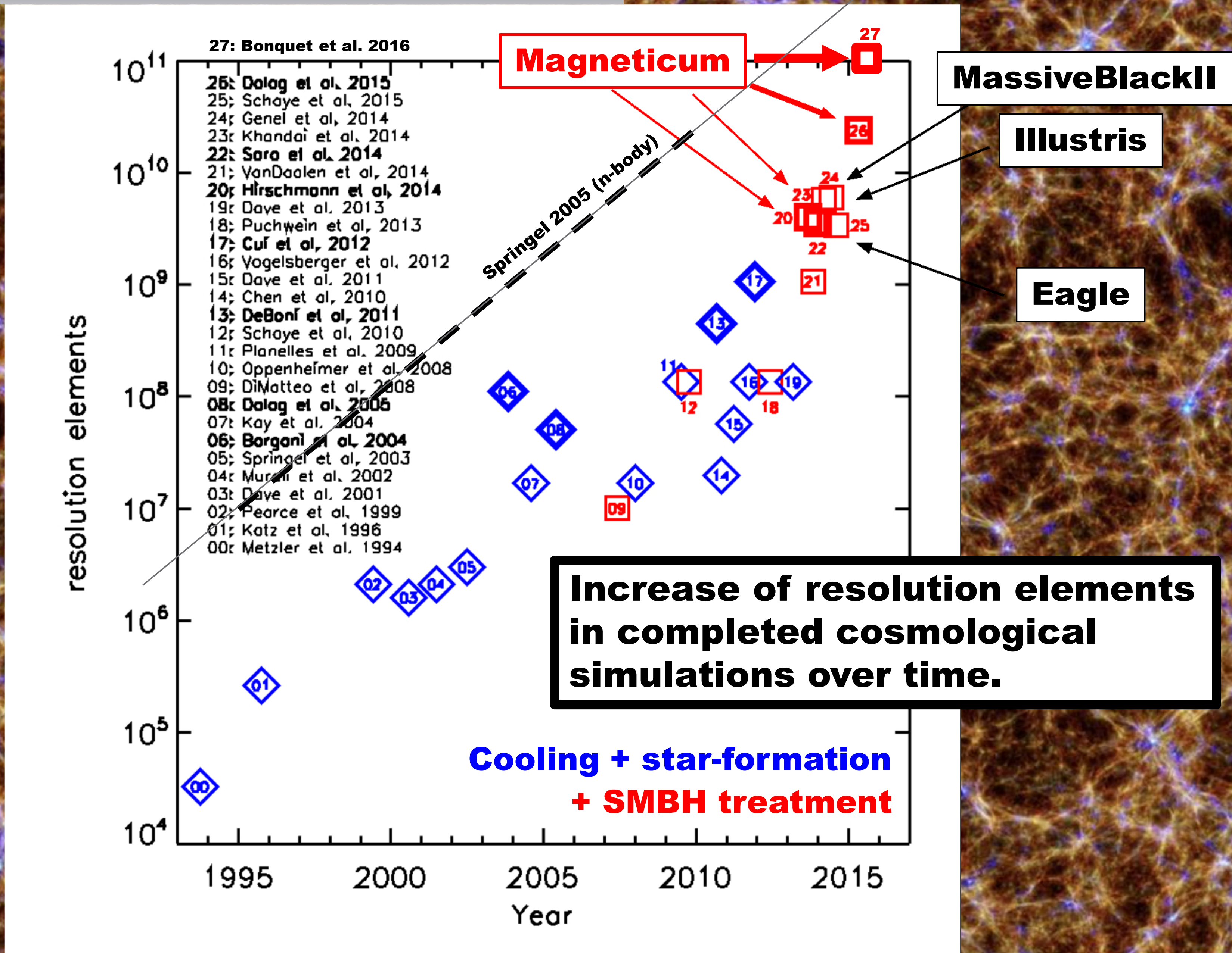


Moore's Law: Doubles all 1.5 Years

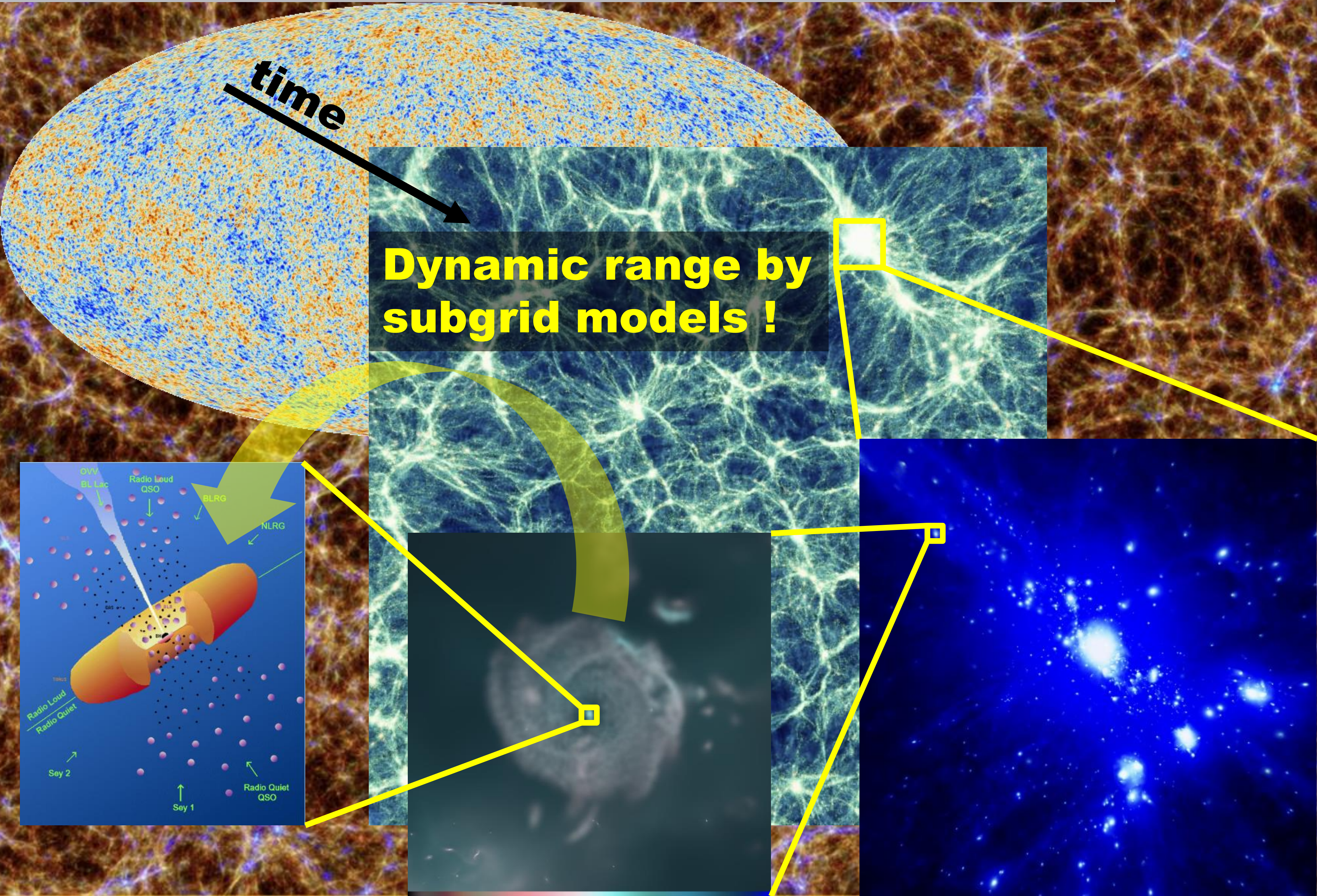
Z3 – SuperMUC: 1.4 Years

N-Body: 1970-2010: 1.3 Years

Hydro Simulations



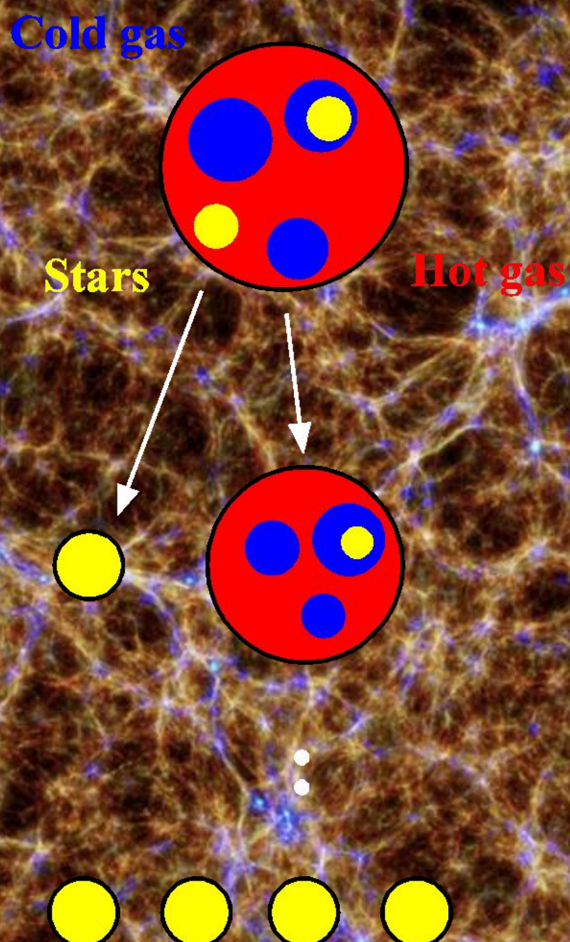
Extending scales to sub-grid



Sub-resolution star-formation:

Multi phase model (sub-scale)

Springel & Hernquist 2002



Star formation

supernova mass fraction

$$\frac{\mathrm{d}\rho_\star}{\mathrm{d}t} = (1-\beta)\frac{\rho_c}{t_\star}$$

star formation timescale

Cloud evaporation

$$\left.\frac{\mathrm{d}\rho_h}{\mathrm{d}t}\right|_{\mathrm{evap}} = A\beta\frac{\rho_c}{t_\star}$$

cloud evaporation parameter

Growth of clouds

cooling function

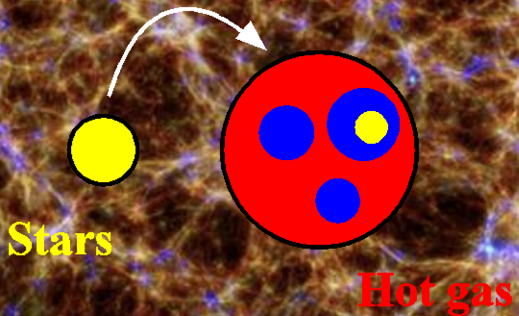
$$\left.\frac{\mathrm{d}\rho_c}{\mathrm{d}t}\right|_{\mathrm{TI}} = -\left.\frac{\mathrm{d}\rho_h}{\mathrm{d}t}\right|_{\mathrm{TI}} = \frac{\Lambda_{\mathrm{net}}(\rho_h, u_h)}{u_h - u_c}$$

Sub-resolution stellar evolution:

Stellar evolution model (sub-scale)

Tornatore et al. 2003/2007

Energy: SNIa, SNII
Metals: SNIa, SNII, AGB winds
H,He,C,Ca,O,N,Ne,Mg
S,Si,Fe,Na,Al,Ar,Ni



IMF:
Salpeter, Kroupa, Chabrier, Arimoto & Yoshii

Life-time:
Maeder & Meynet 1989
Padovani & Matteucci 1993

Stellar yields:
AGB: Groenewegen, Karakas
SN1a: Thielemann
SNII: Woosly & Weaver
Romano, Kobayashi, ...

star-formation rate

fraction of stars in binary systems

SNIa rate:

$$R_{\rm SN\,Ia}(t) = A \int_{M_{\rm B,inf}}^{M_{\rm B,sup}} \phi(m_{\rm B}) \int_{\mu_{\rm m}}^{\mu_{\rm M}} f(\mu)\, \psi(t - \tau_{m_2})\, {\rm d}\mu\, {\rm d}m_{\rm B}$$

distribution of mass-ratios in binary systems

mass range of SN1a binary systems (0.8-8Msol)

SNII and AGB rate:

$$R_{\rm SN\,II|ILMS}(t) = \phi(m(t)) \times \left(-\frac{{\rm d}m(t)}{{\rm d}t}\right)$$

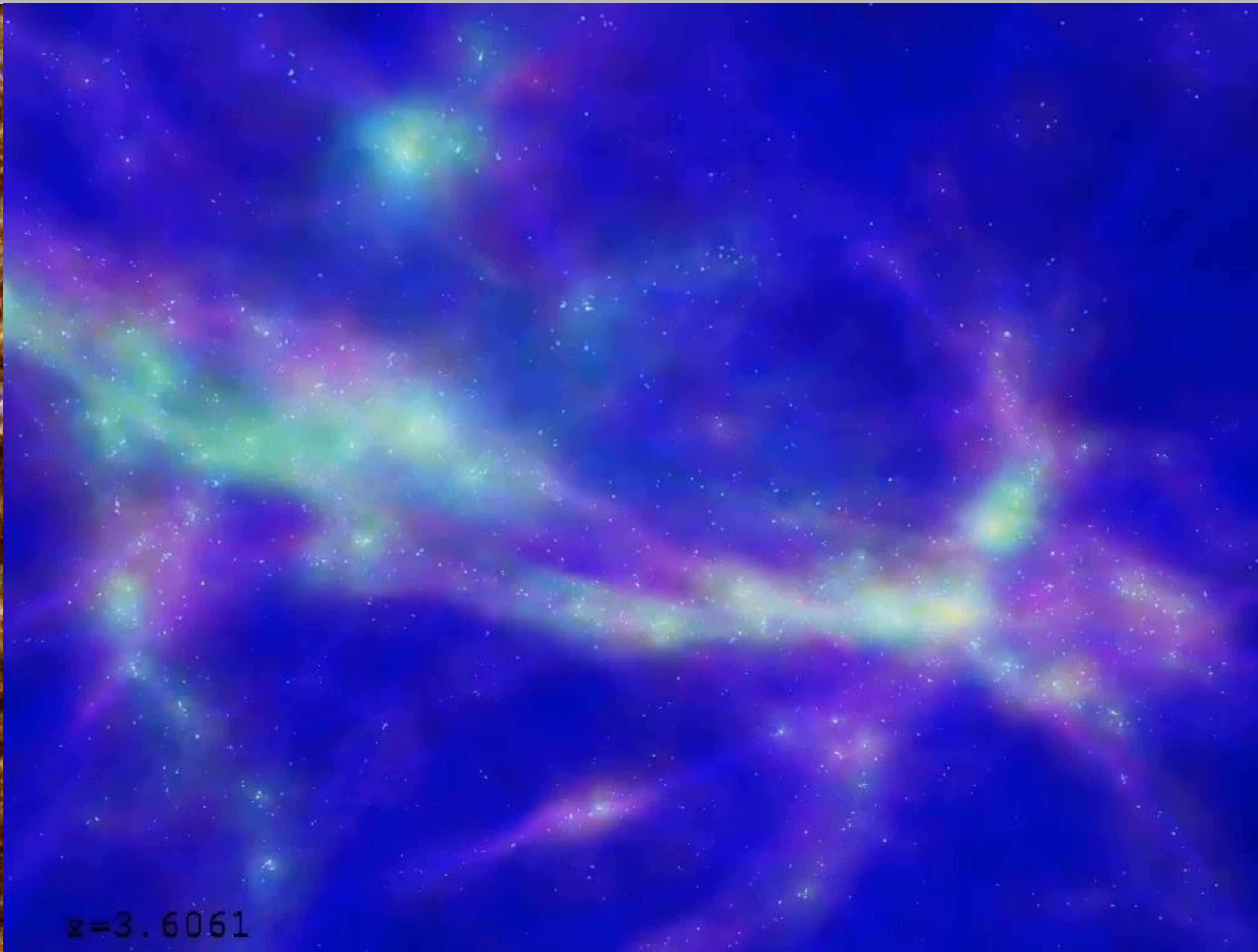
Initial mass function (IMF):

$$\phi(m) = dN/d\log m$$

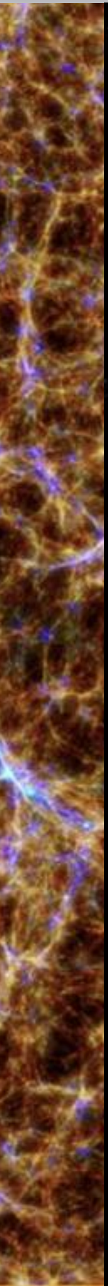
Life-time of stars

$$\tau(m) = \begin{cases} 10^{[(1.34-\sqrt{1.79-0.22(7.76-\log(m))})/0.11]-9} & \text{for } m \le 6.6\ {\rm M}_\odot \\ 1.2m^{-1.85} + 0.003 & \text{otherwise.} \end{cases}$$

Formation of galaxy clusters: gas



Formation of galaxy clusters: stars



Sub-resolution SMBH-formation:

Black Hole model (sub-scale)

Springel & Di Matteo 2006

Seeding

Constant seeding
Seeding on m-sigma

Accretion on BH

α -Bondi (Springl & Di Matteo 06)
β -Bondi (Booth & Schaye 09)
cold/hot (Bachmann et al. 14)
....

Feedback

Thermal (Springel & Di Matteo 06)
Bubbles (Sijacki et al. 07)
Mass dependent (Steinborn 2015)
....

Merging

Instant merging
Based on velocity
....

Growth of Black Hole

$$\dot{M}_{\mathrm{B}} = \alpha \times 4\pi R_{\mathrm{B}}^2 \rho c_s \simeq \frac{4\pi\alpha G^2 M_\bullet^2 \rho}{(c_s^2 + v^2)^{3/2}}$$

gas density

sound speed

$$\dot{M}_\bullet = \min(\dot{M}_{\mathrm{B}}, \dot{M}_{\mathrm{Edd}})$$

Feedback by Black Holes

$$L_{\mathrm{bol}} = 0.1 \times \dot{M}_\bullet c^2$$

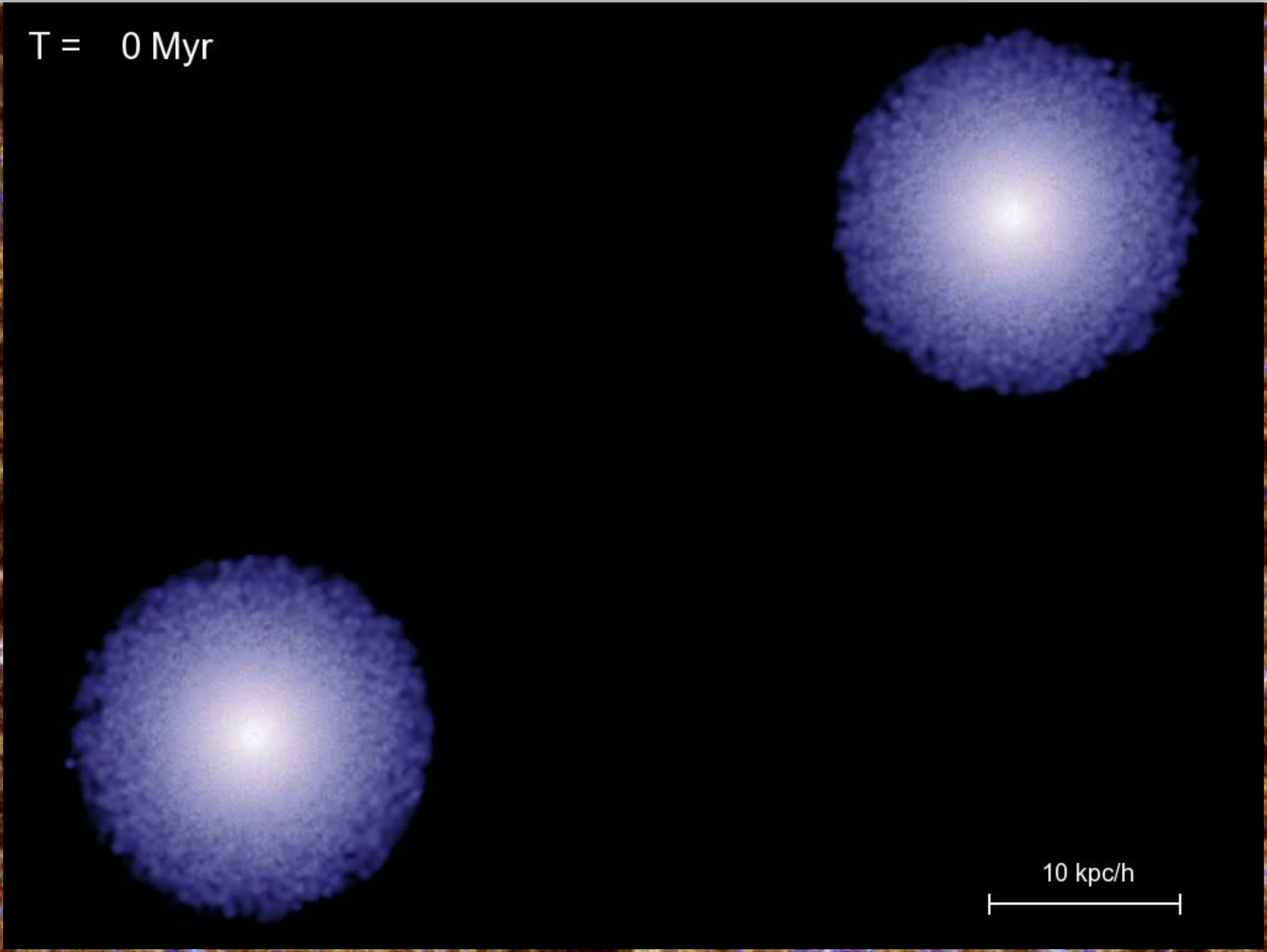
$$\dot{E}_{\mathrm{feedback}} = f \times L_{\mathrm{bol}}$$

efficiency

Positioning:

Pinning to min. Potential
Free floating

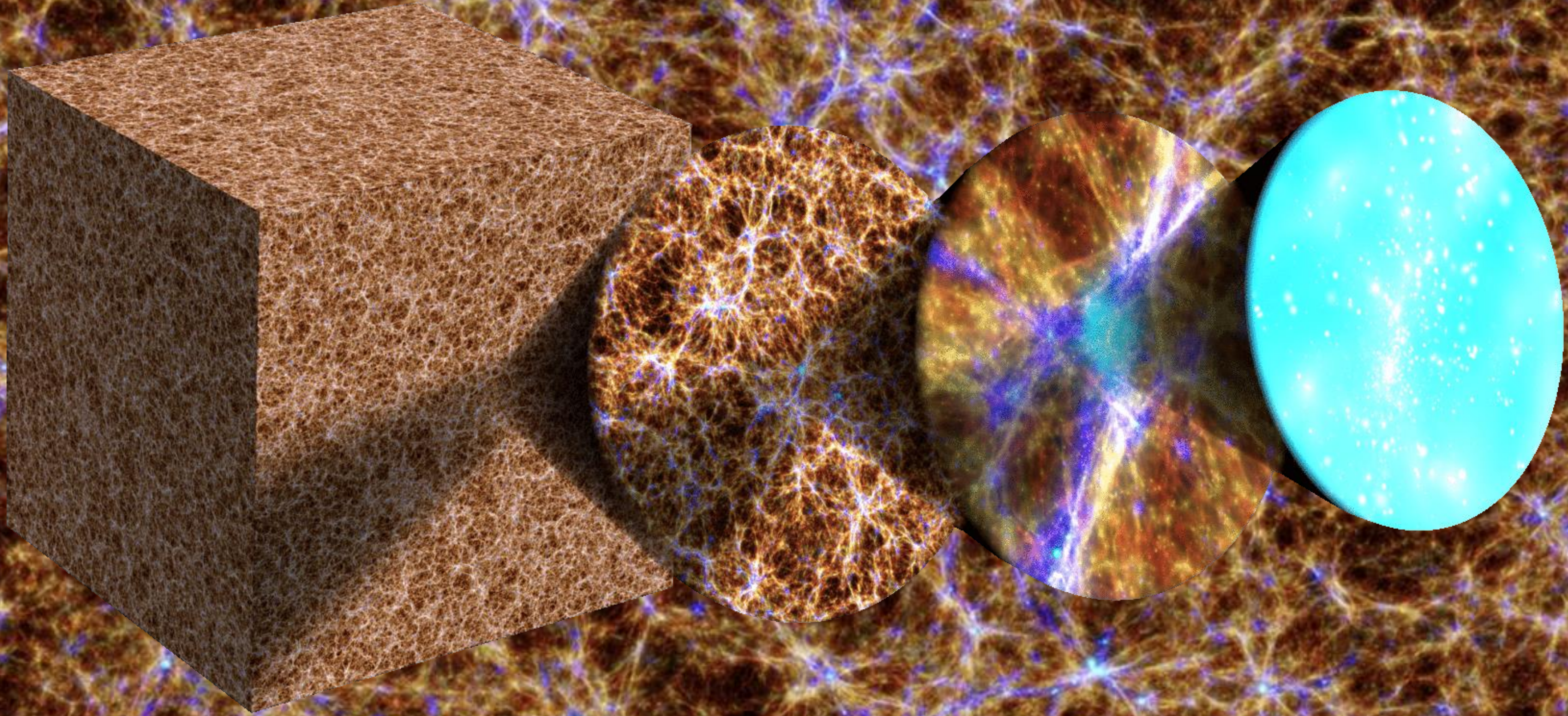
Merging galaxies with black holes



http://wwwmpa.mpa-garching.mpg.de/gadget/right.html#Movies

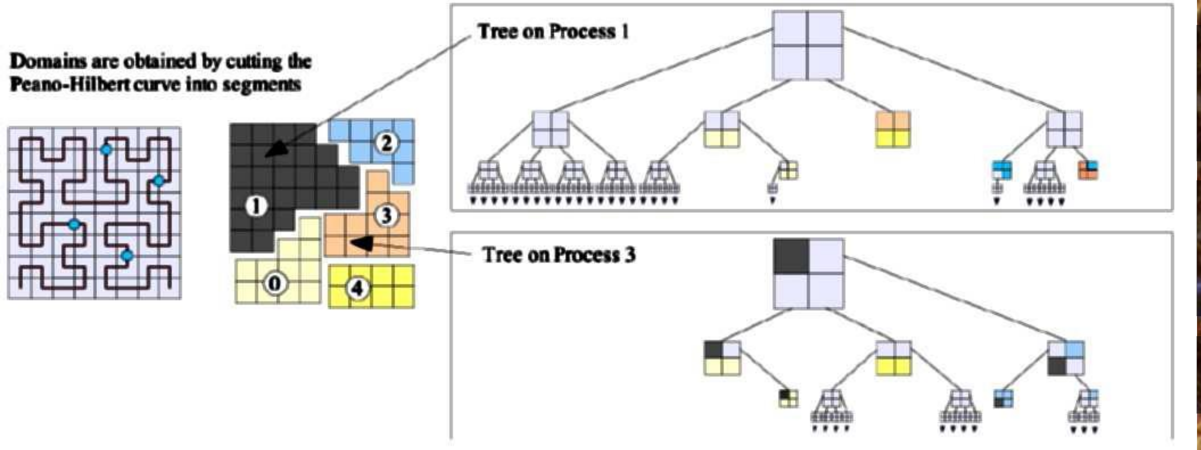
Bigger: the challenge

Extending to the largest simulation currently possible, e.g. production runs on entire HPC systems !



General Constrains:

1) Fill significant fraction of memory with real particle/cell data instead of overhead !



80 bytes per Particle (MILLENNIUM XXL)

650 bytes per Particle (Magneticum)

=> 10^7 per socket

Often still limiting factor in simulations which store full tree or cell structure (AMR) on every MPI task.

(Springel 2005)

2) Formulate algorithm facing limitations in memory

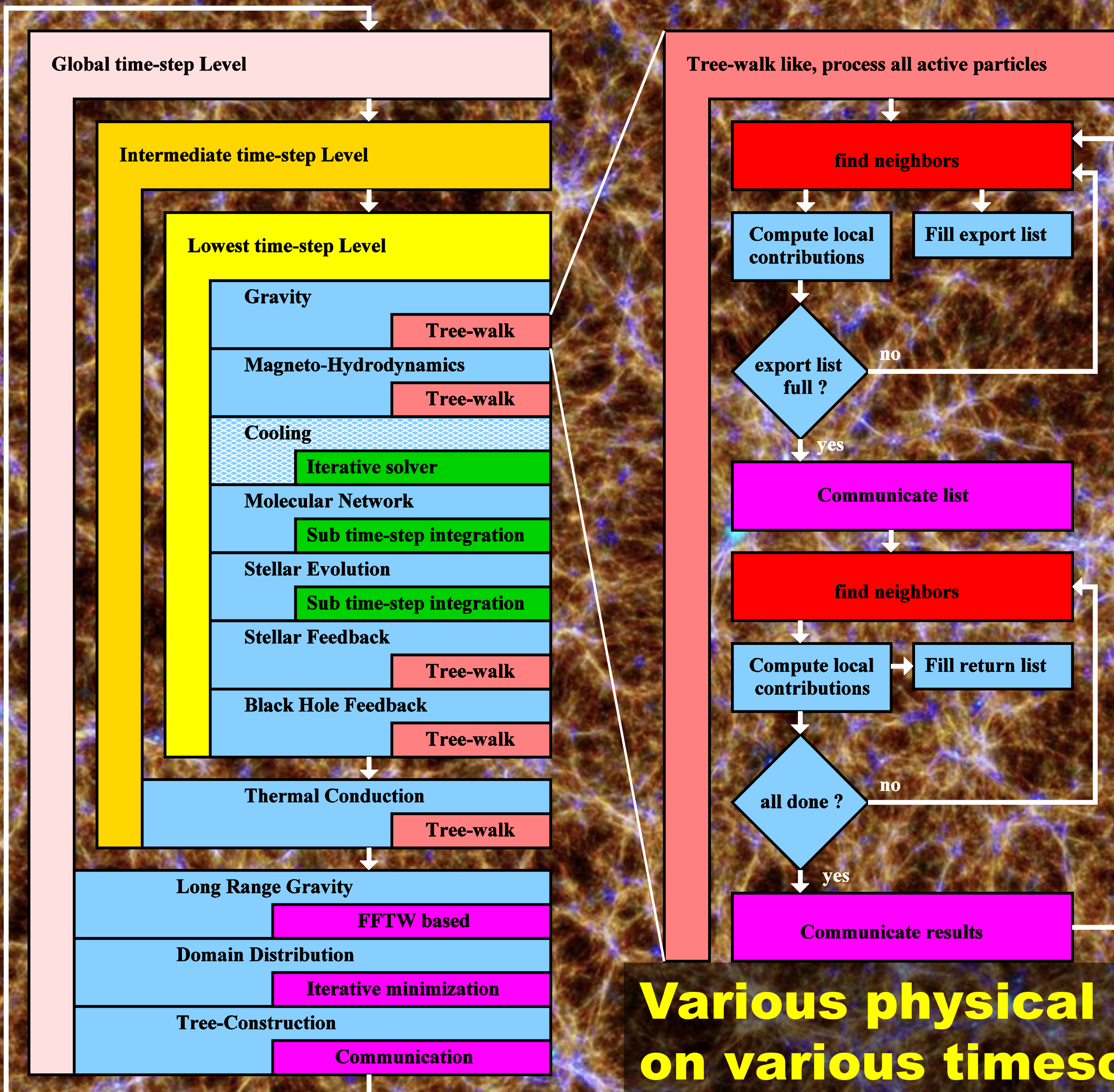
Using buffers and checking their state major hassel for massive node parelelism (LOCKs !)

3) Formulate algorithm to be reproducable ?

Conserving order of operation is not always trivial, but make strongly non linear system difficult to compare (but this is almost a philosophical question) !

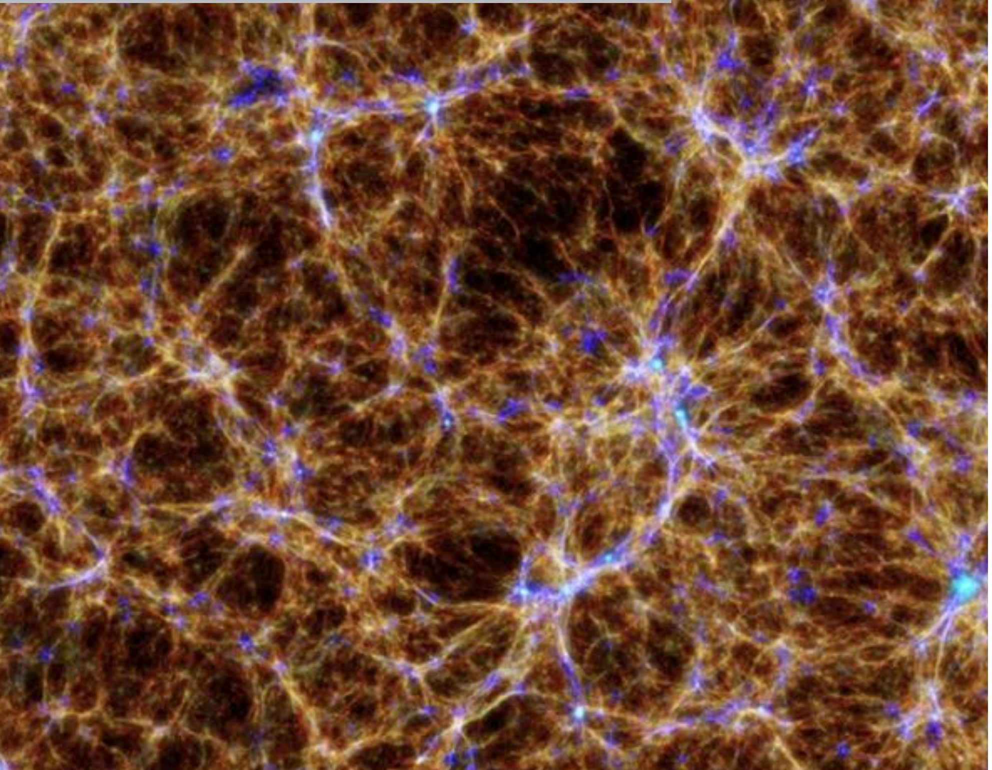
Gadget

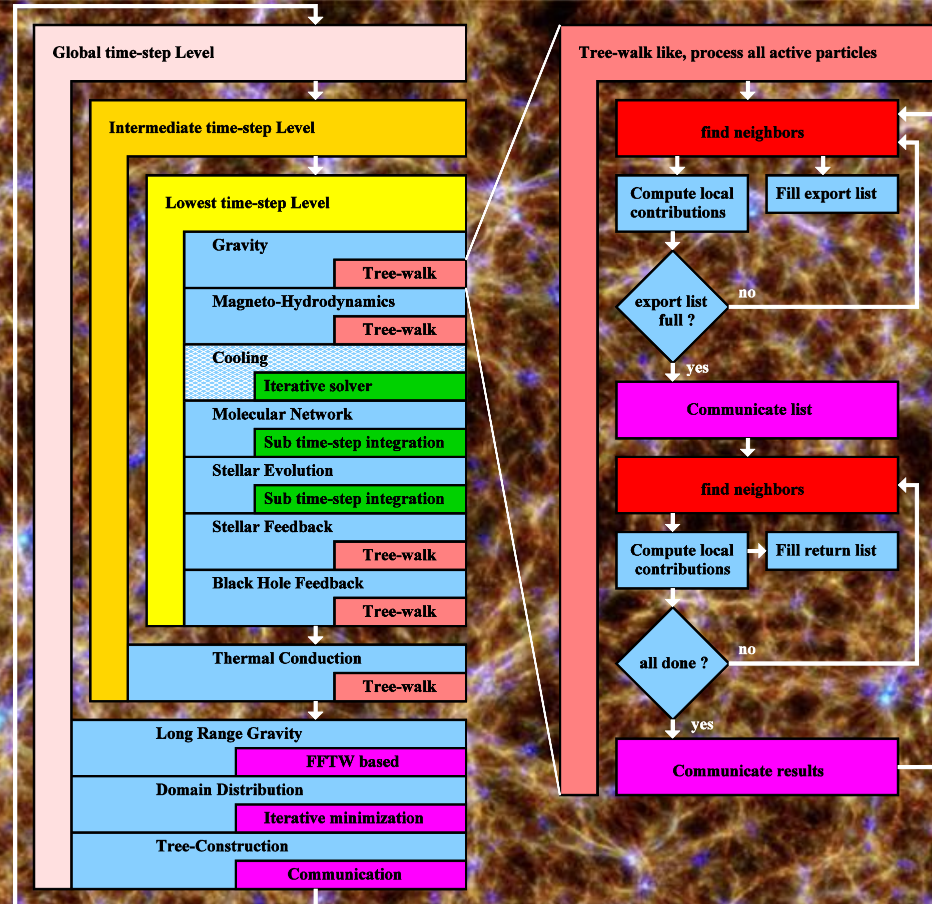
Interplay of different physics modules in Gadget3



Various physical proceses on various timescales ...

Time-steps






```
Sync-Point 35762, Time: 0.354961, Redshift: 1.81721, Systemstep: 5.56636e-06, Dloga: 1.56817e-05
Occupied timebins: non-cells     cells       dt                  cumulative A D    avg-time  cpu-frac
 X  bin=20      88467428683  83736416883    0.008029050471      187287803634 < *      712.64      22.0%
 X  bin=19      4161559798  7573611723   0.004014525236     15083958068    *     219.80      6.8%
 X  bin=18       689640843  1347123564   0.002007262618     3348786547    *      63.43      3.9%
 X  bin=17       114047063   493679540   0.001003631309     1312022140          100.57     12.4%
 X  bin=16       424984562    98886753   0.000501815654      704295537           14.78      3.7%
 X  bin=15       149882337    22957206   0.000250907827      180424222           24.04     11.9%
 X  bin=14         1436369     5551598   0.000125453914        7584679            5.78      5.7%
 X  bin=13           10827      562559   0.000062726957         596712            3.46      6.8%
 X  bin=12             453       22095   0.000031363478          23326            2.38      9.4%
 X  bin=11              10         768   0.000015681739            778            2.20     17.4%
               ------------------------
PM-Step. Total: 94008990945  93278812689    Sum: 187287803634
```

Time-steps

Simple Math:

0.5sec per step, 10 Million steps:

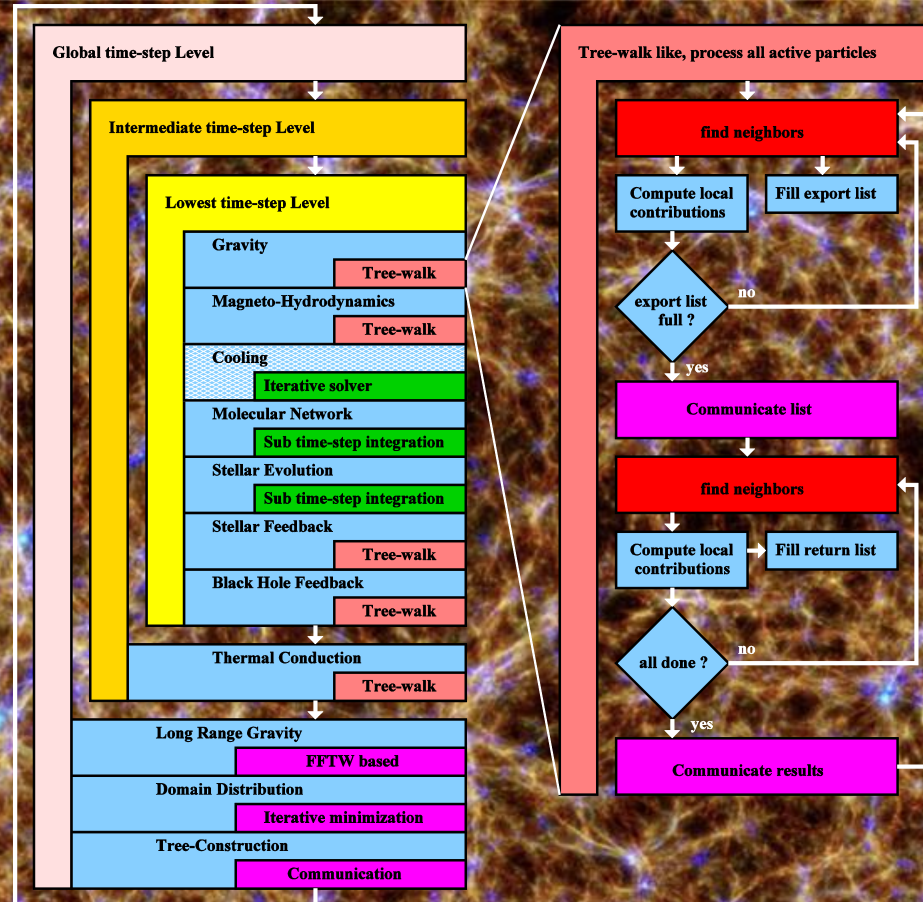
5.000.000 sec

1 day ~ 86.400 sec

1 week = 604.800 sec

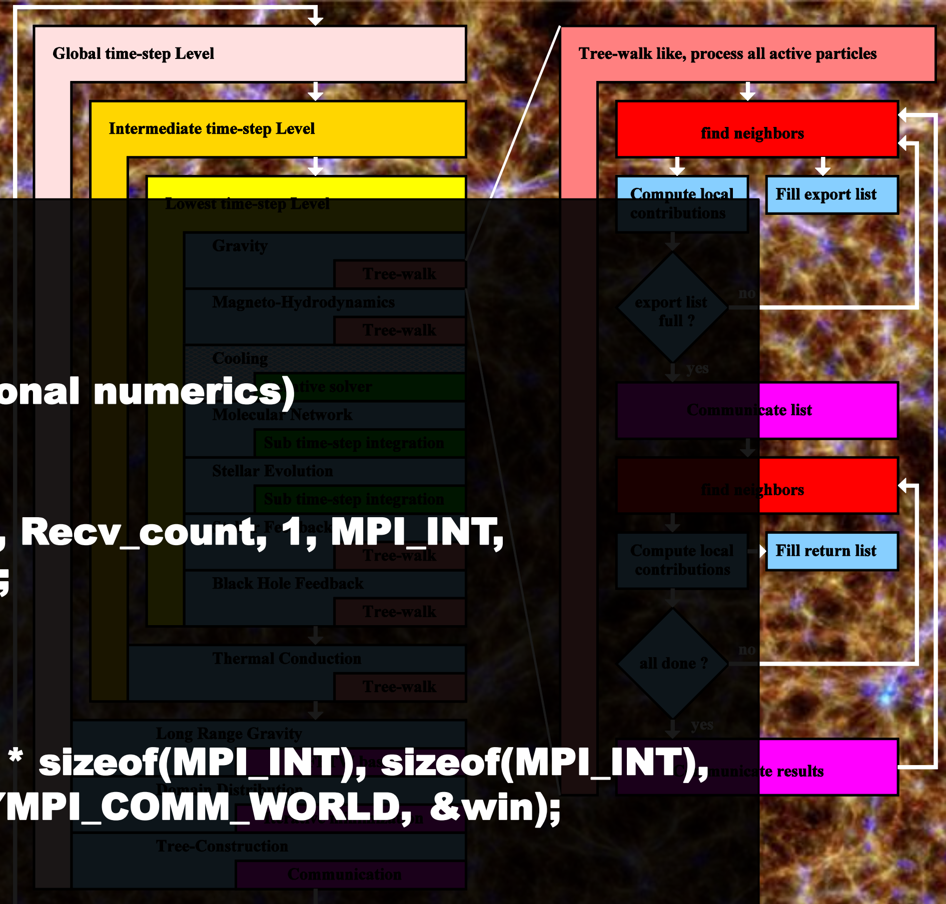
1 month = 2.678.400 sec

$\Rightarrow$ 2 month to hanlde 778 out of $1.7\text{x}10^{11}$




```
Sync-Point 35762, Time: 0.354961, Redshift: 1.81721, Systemstep: 5.56636e-06, Dloga: 1.56817e-05
Occupied timebins: non-cells      cells        dt                   cumulative A D    avg-time  cpu-frac
  X  bin=20      88467428683  83736416883   0.008029050471      187287803634 < *      712.64     22.0%
  X  bin=19       4161559798   7573611723   0.004014525236       15083958068   *       219.80      6.8%
  X  bin=18        689640843   1347123564   0.002007262618        3348786547   *        63.43      3.9%
  X  bin=17        114047063    493679540   0.001003631309        1312022140           100.57     12.4%
  X  bin=16        424984562     98886753   0.000501815654         704295537            14.78      3.7%
  X  bin=15        149882337     22957206   0.000250907827         180424222            24.04     11.9%
  X  bin=14          1436369      5551598   0.000125453914           7584679             5.78      5.7%
  X  bin=13            10827       562559   0.000062726957            596712             3.46      6.8%
  X  bin=12              453        22095   0.000031363478             23326             2.38      9.4%
  X  bin=11               10          768   0.000015681739               778             2.20     17.4%
               ------------------------
PM-Step. Total: 94008990945  93278812689    Sum: 187287803634
```

Time-steps



What can be done ?

- **Avoid loops over all particles**
  (but: drift, additional physics, additional numerics)

- **Avoid almost empty All to All**

```
MPI_Alltoall(Send_count, 1, MPI_INT, Recv_count, 1, MPI_INT,
                     MYMPI_COMM_WORLD);
```

↔

```
for(j = 0; j < NTask; j++)
  Recv_count[j] = 0;
MPI_Win_create(Recv_count, NTask * sizeof(MPI_INT), sizeof(MPI_INT),
                     MPI_INFO_NULL,MYMPI_COMM_WORLD, &win);
MPI_Win_fence(0, win);
for(j = 0; j < NTask; ++j)
   if(Send_count[j] > 0)
      MPI_Put(&Send_count[j], 1, MPI_INT, j, ThisTask, 1, MPI_INT, win);
MPI_Win_fence(0, win);
MPI_Win_free(&win);
```

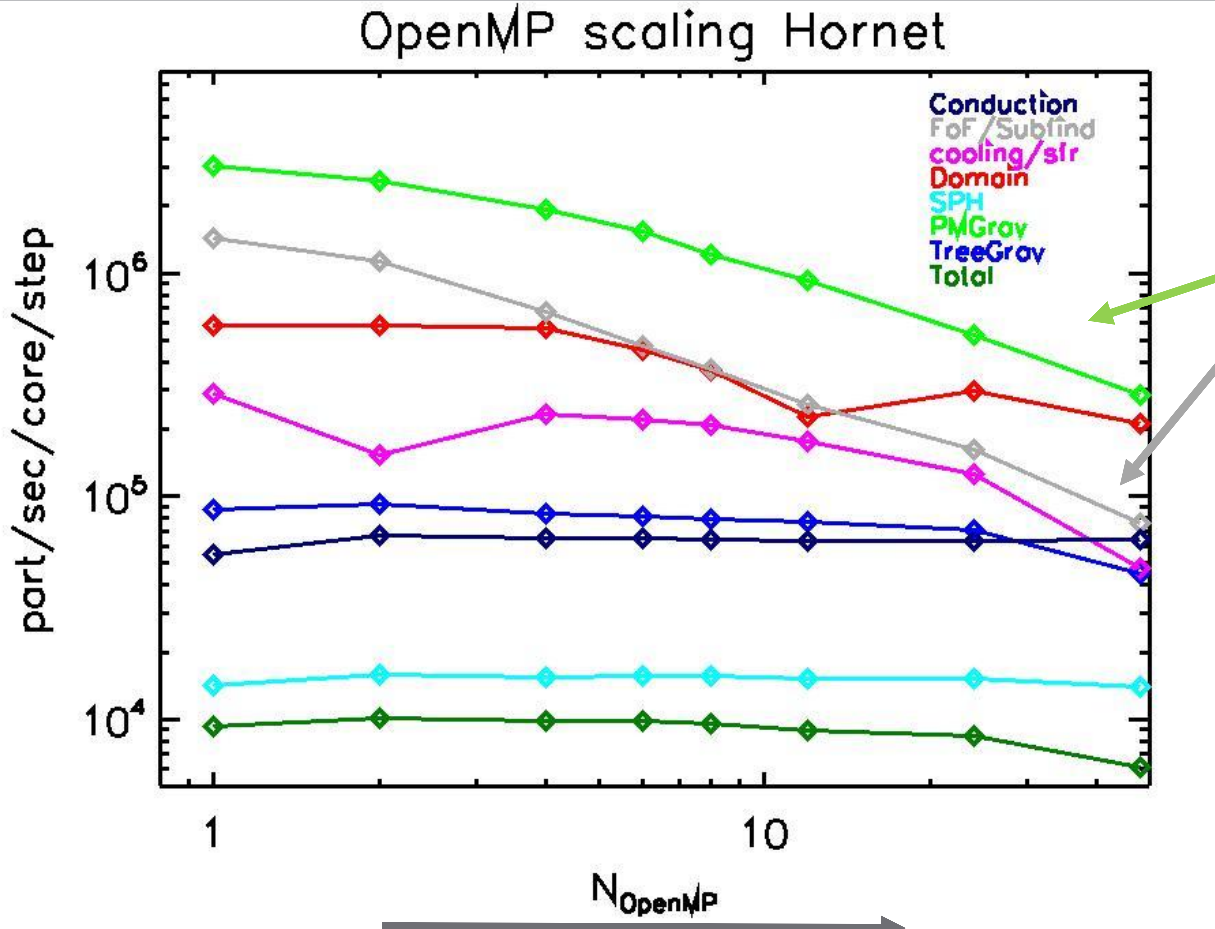
Sync-P[illegible] 357[illegible] [illegible]hift: 1.81721, Systemstep: 5.56636e-06, Dloga: 1.56817e-05

Occup[illegible] [illegible]rac

X	bi[illegible]=20	88467428683	827364?6883	0.008029050471	187287803634 < *	712.64	22.0%
X	bi[illegible]=19	[illegible]	[illegible]	0.004014525236	15083958068 *	219.80	[illegible]6.8%
X	bi[illegible]=18	[illegible]	[illegible]	0.002007262618	3348786547 *	63.43	3[illegible].9%
X	bin=17	114047063	493679540	0.001003631309	1312022140	100.57	12.4%
X	bin=16	424984562	98886753	0.000501815654	704295537	14.78	3.7%
X	bin=15	149882337	22957206	0.000250907827	180424222	24.04	11.9%
X	bin=14	1436369	5551598	0.000125453914	7584679	5.78	5.7%
X	bin=13	10827	562559	0.000062726957	596712	3.46	6.8%
X	bin=12	453	22095	0.000031363478	23326	2.38	9.4%
X	bin=11	10	768	0.000015681739	778	2.20	17.4%

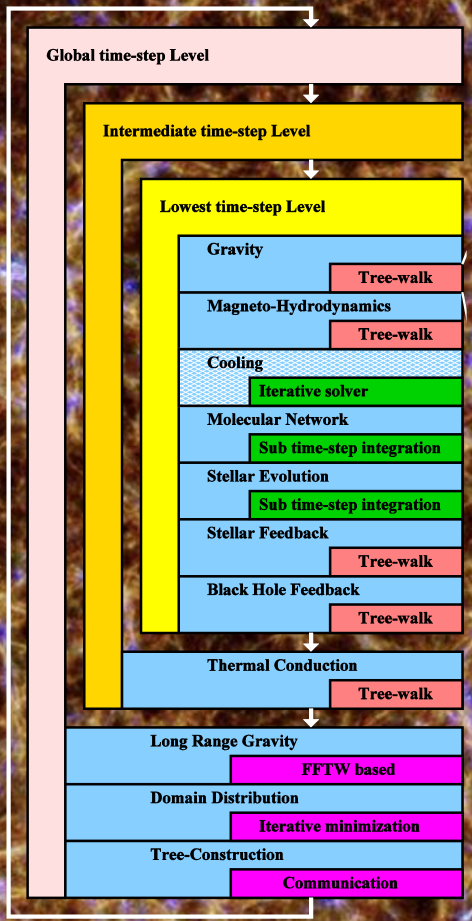
PM-Step. Total: 94008990945 93278812689 Sum: 187287803634

A must: MPI/OpenMP hybrid



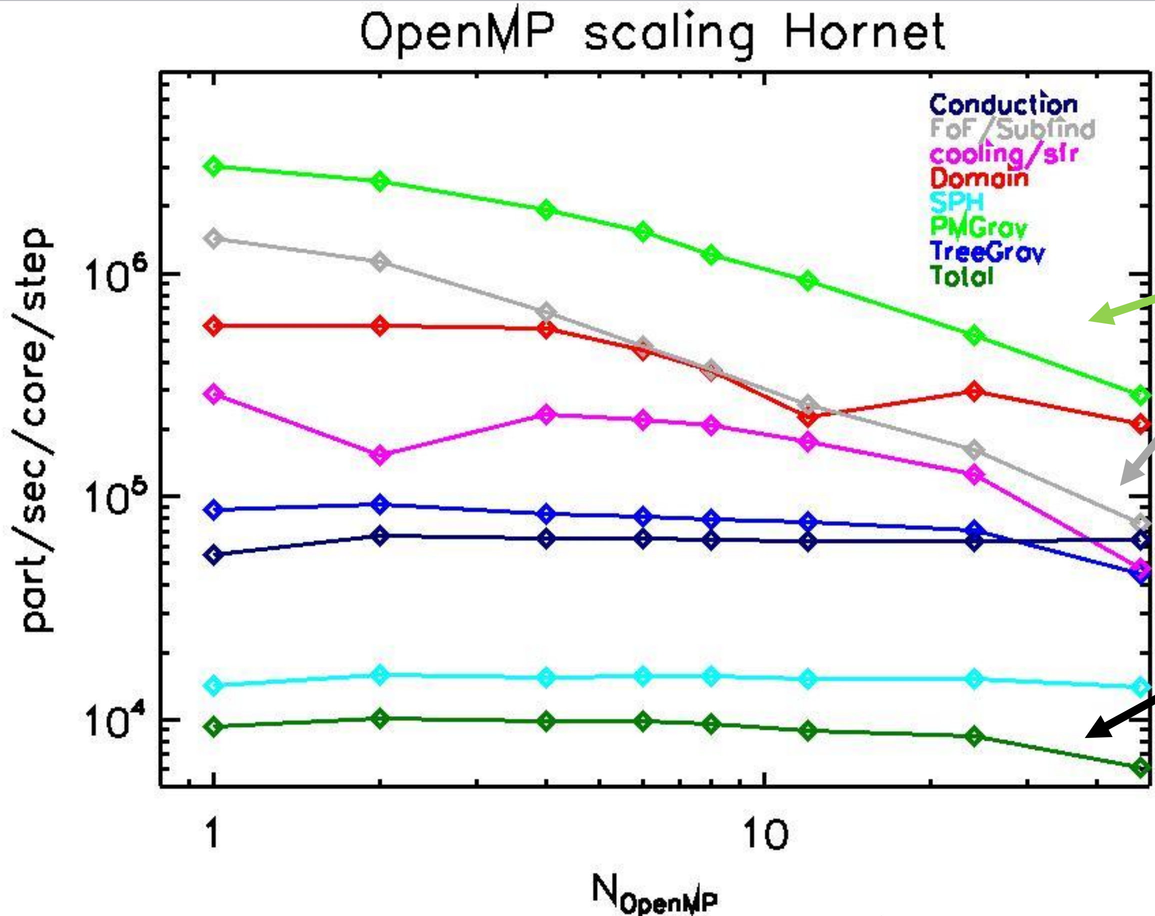
Replacing MPI by OpenMP threads

Every part count !



For Gadget3 started 2012 with KONWHIR III project, but still ongoing effort …

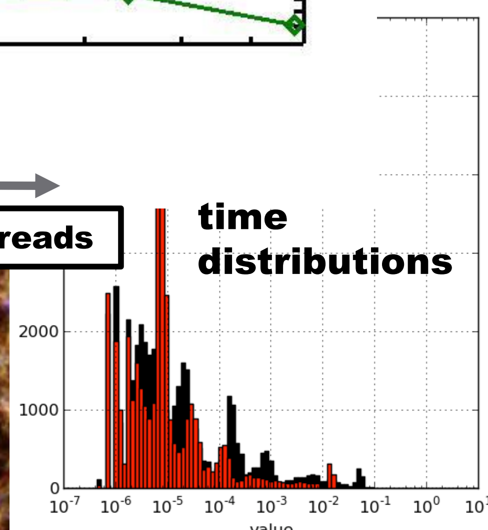
A must: MPI/OpenMP hybrid

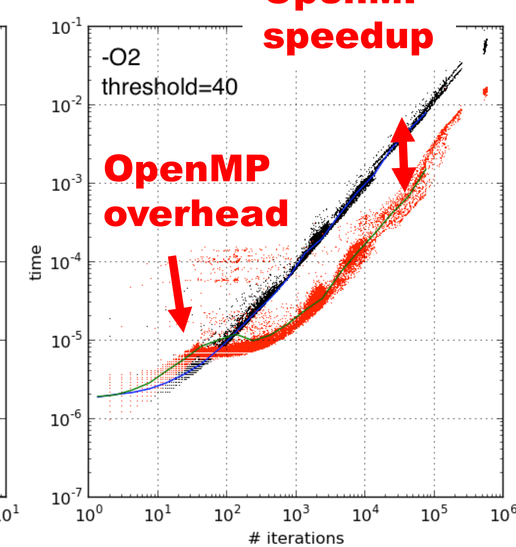


Not fully OpenMP parallelised ...

Crossing sockets

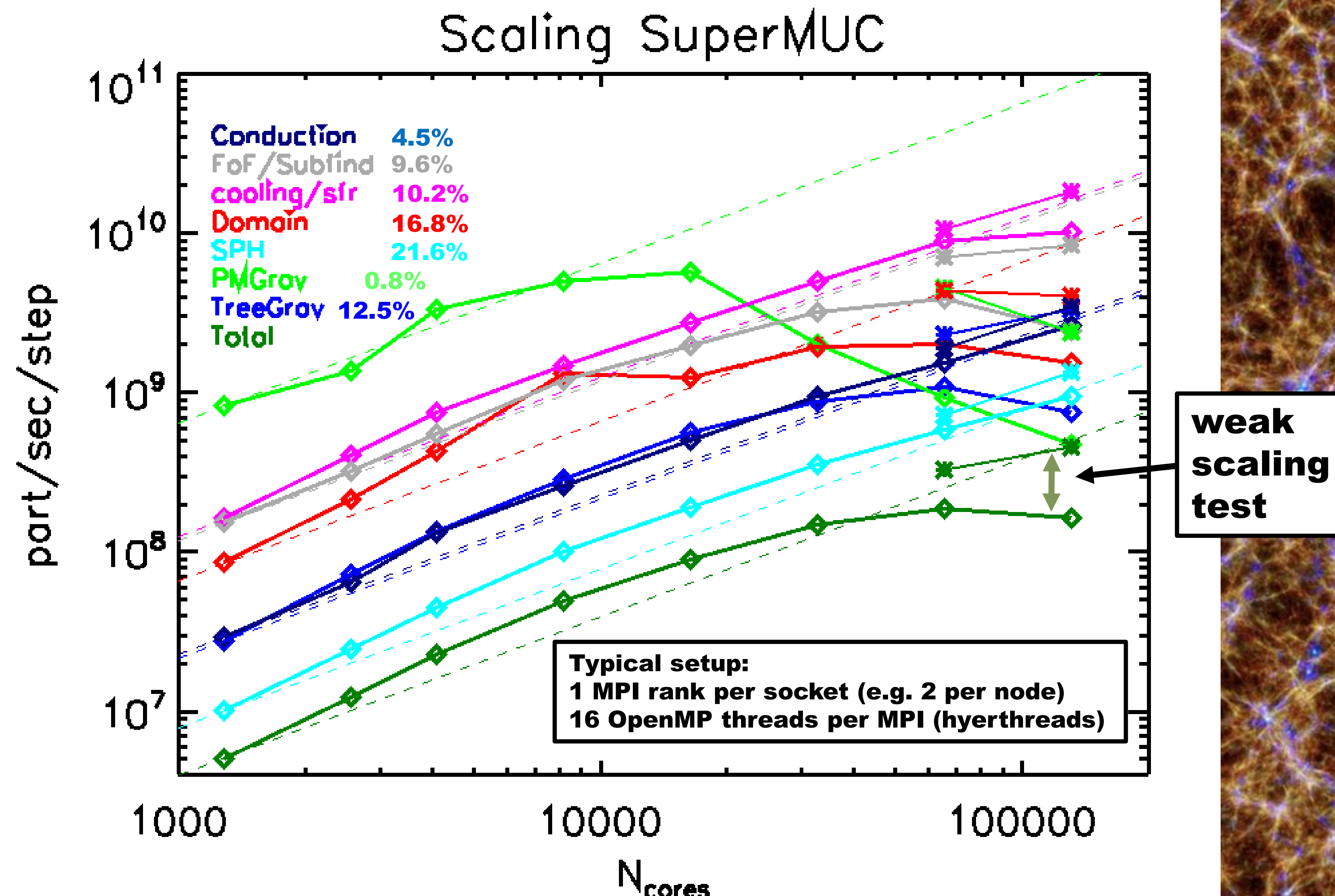
Replacing MPI by OpenMP threads



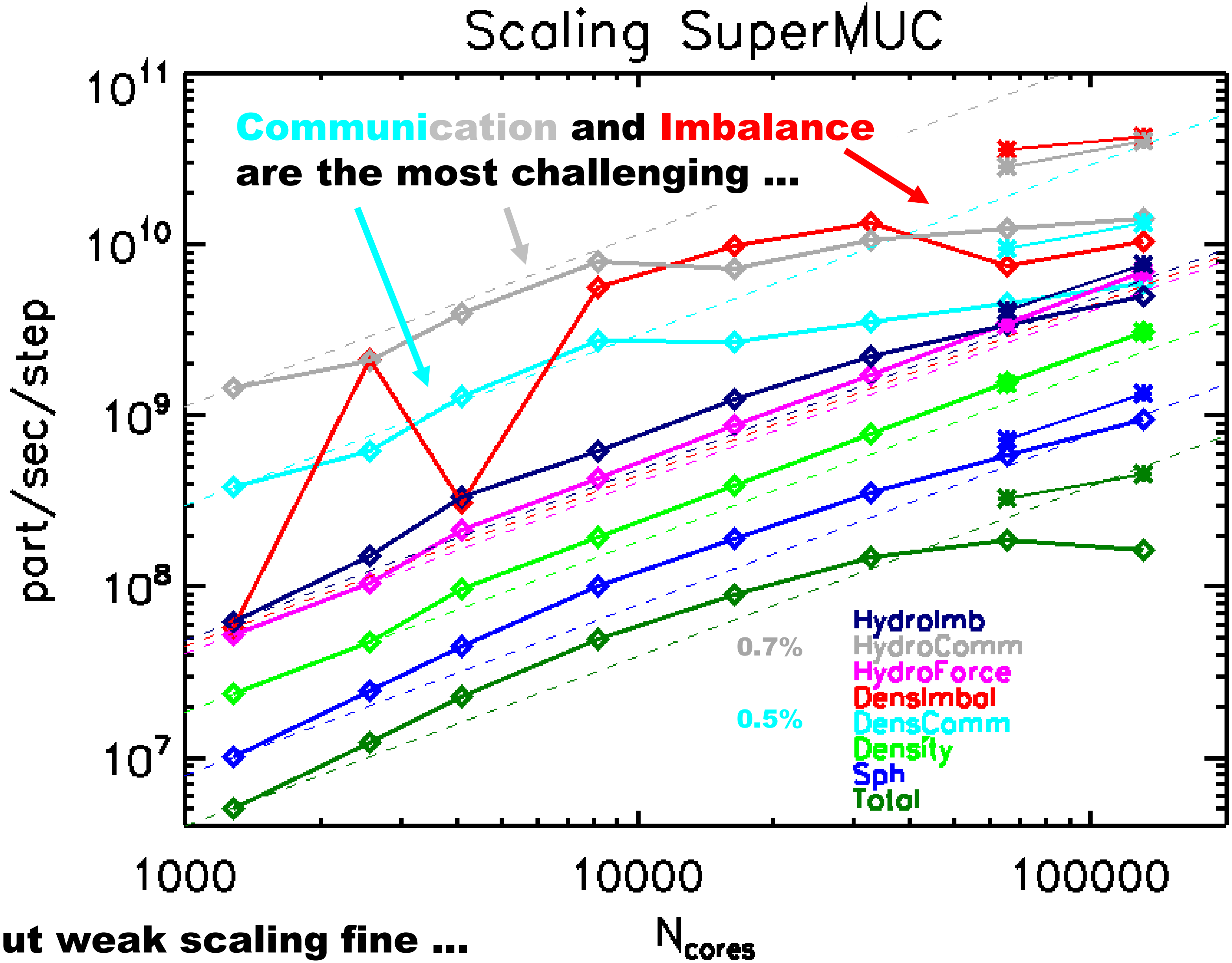


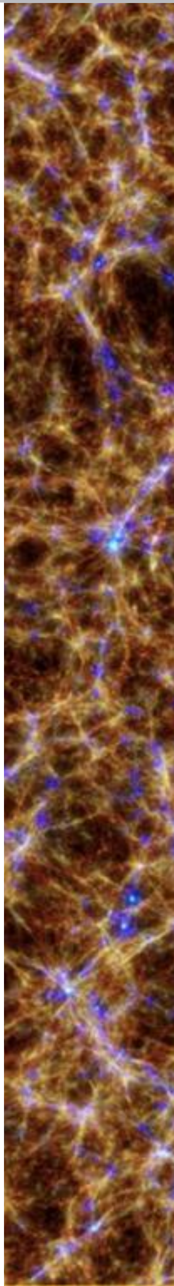
For Gadget3 started 2012 with KONWHIR III project, but still ongoing effort ...

Strong improvement for scaling



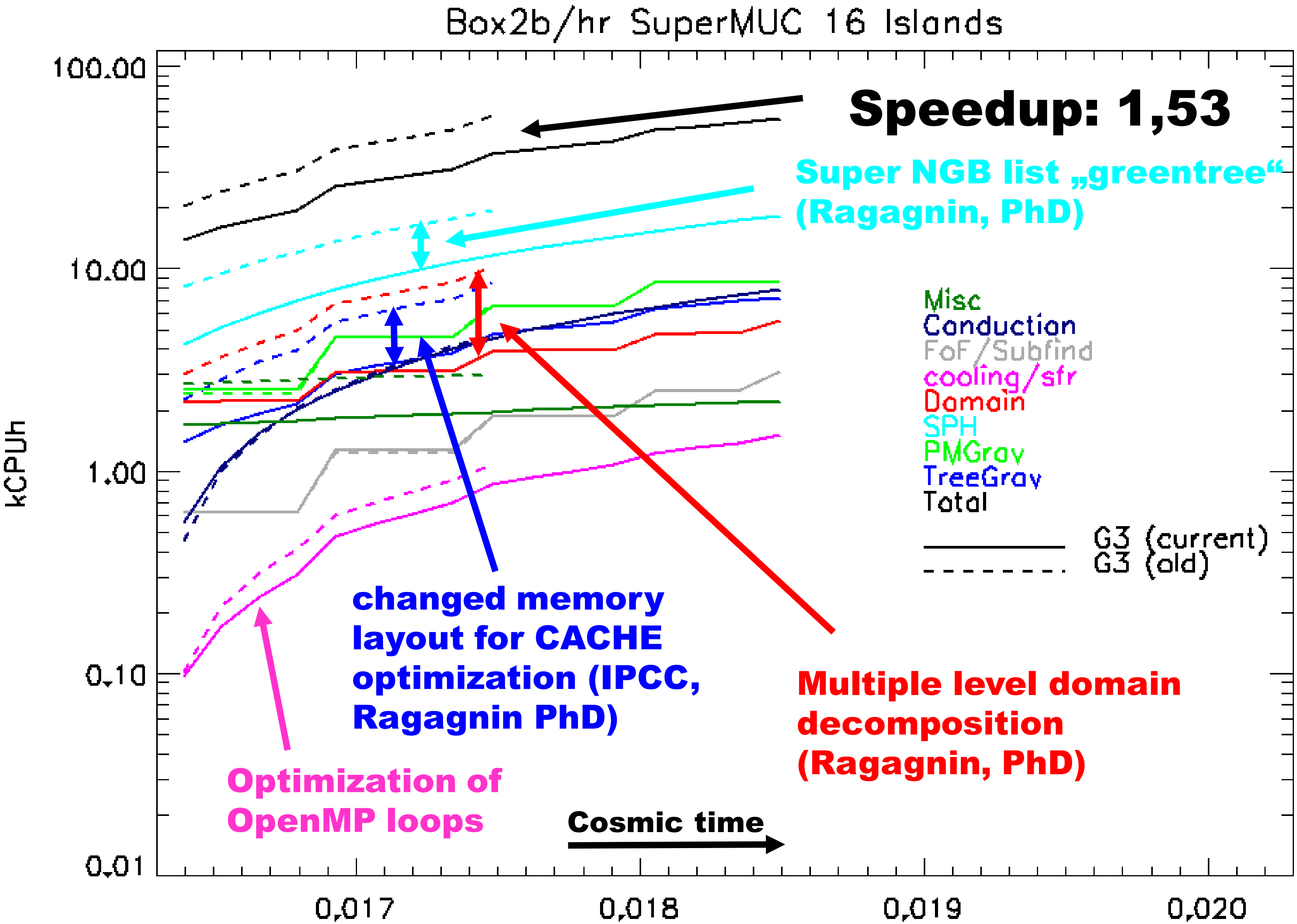
Actual scaling hydro computation

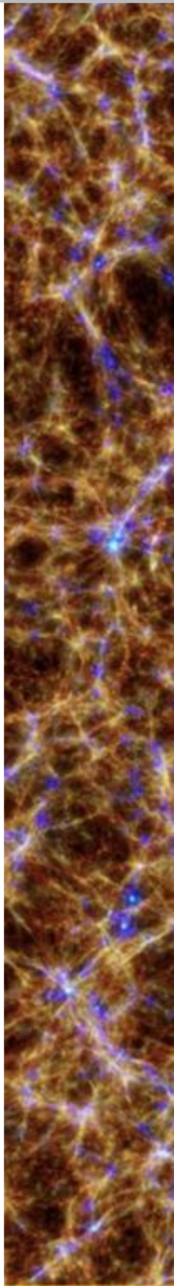




but weak scaling fine ...

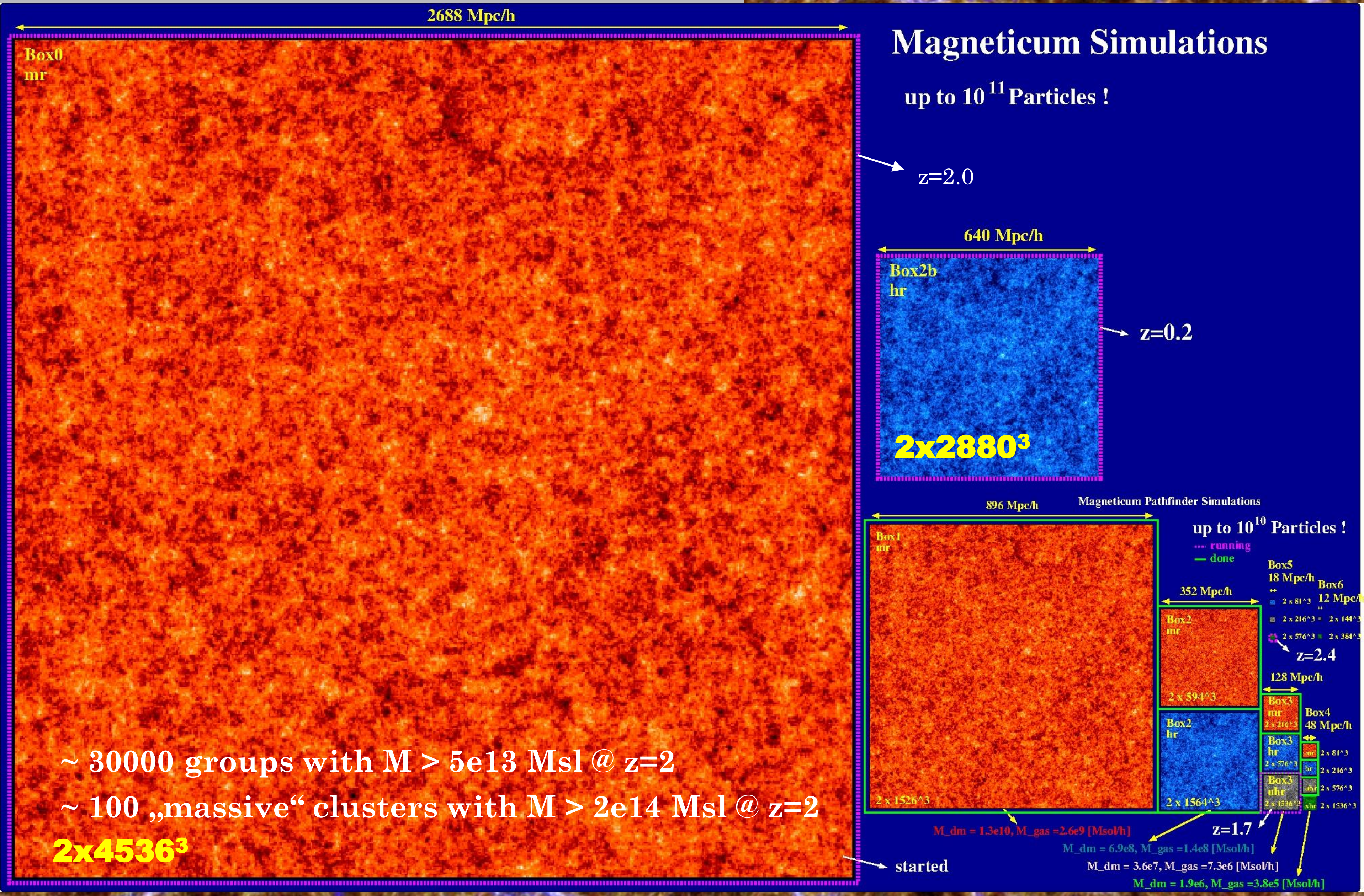
Recent improvements ...



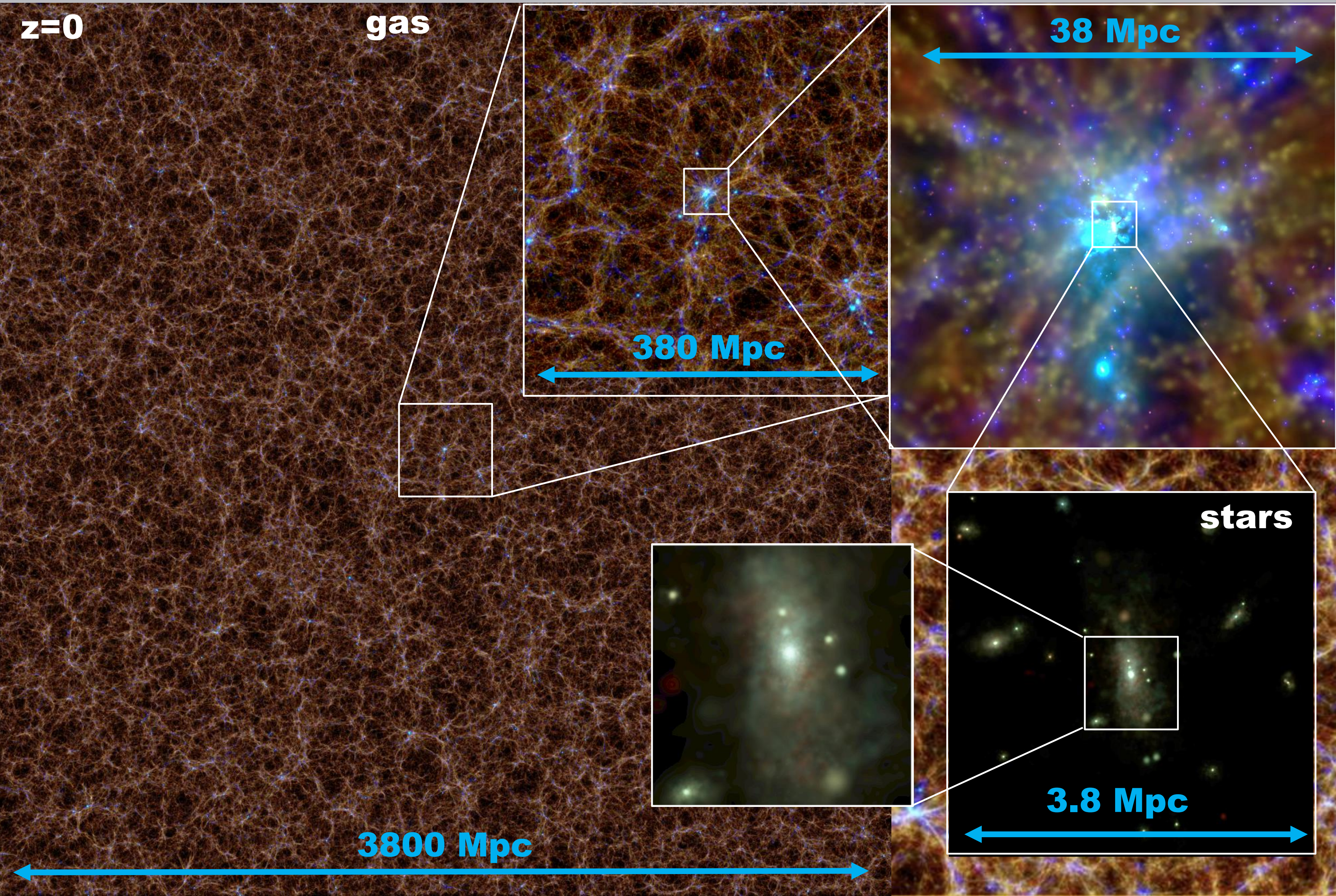


Recent test run on full SuperMUC Phase I (262144 threads) ...

What we reached ...



Largest Simulation (Box0/mr)



MAGNETICUM

	Box0	Box1	Box2b	Box2	Box3	Box4
[Mpc/h]	2688	896	640	352	128	48
mr	2×4536^3	2×1526^3	–	2×594^3	2×216^3	2×81^3
hr	–	–	2×2880^3	2×1584^3	2×576^3	2×216^3
uhr	–	–	–	–	$2 \times 1536^3 (z = 2)$	2×576^3

Setup:

$2x4536^3$ = 186.659.085.312 particle

Almost 20 times size of ILLUSTRIS or EAGLE

Full Physics + improved SPH:

200 bytes per DM particle, 456 bytes per GAS particle

Complete SuperMUC Phase II:

6 x 512 x 2 x 28 = 172032 tasks

1 MPI task per socket, 28 OpenMP per MPI

68.5 TB for single checkpointing

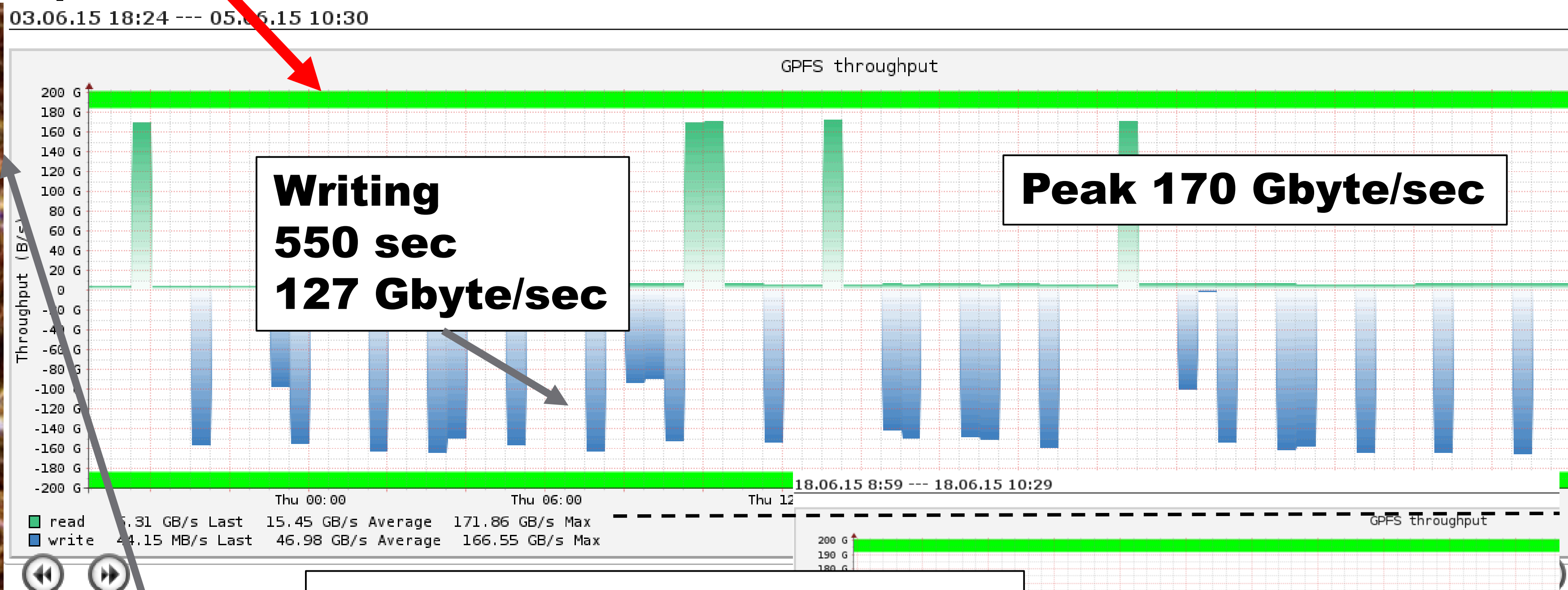
reaching 170 Gbyte/sec, more than 3 Peta byte written

Dying nodes are the main hassle !

12h longest contineous run, checkpoints all 1.5h

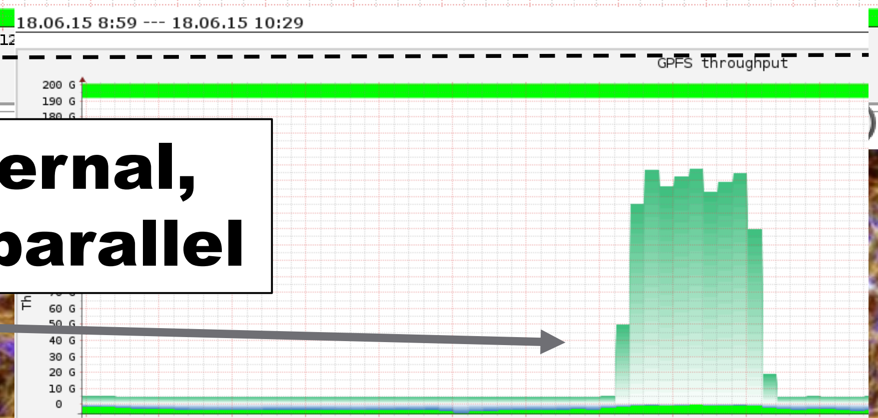
MAGNETICUM

	Box0	Box1	Box2b	Box2	Box3	Box4
[Mpc/h]	2688	896	640	352	128	48
mr	2×4536^3	2×1526^3	–	2×594^3	2×216^3	2×81^3
hr	–	–	2×2880^3	2×1584^3	2×576^3	2×216^3
uhr	–	–	–	–	$2 \times 1536^3 (z = 2)$	2×576^3



Writing
550 sec
127 Gbyte/sec

Peak 170 Gbyte/sec



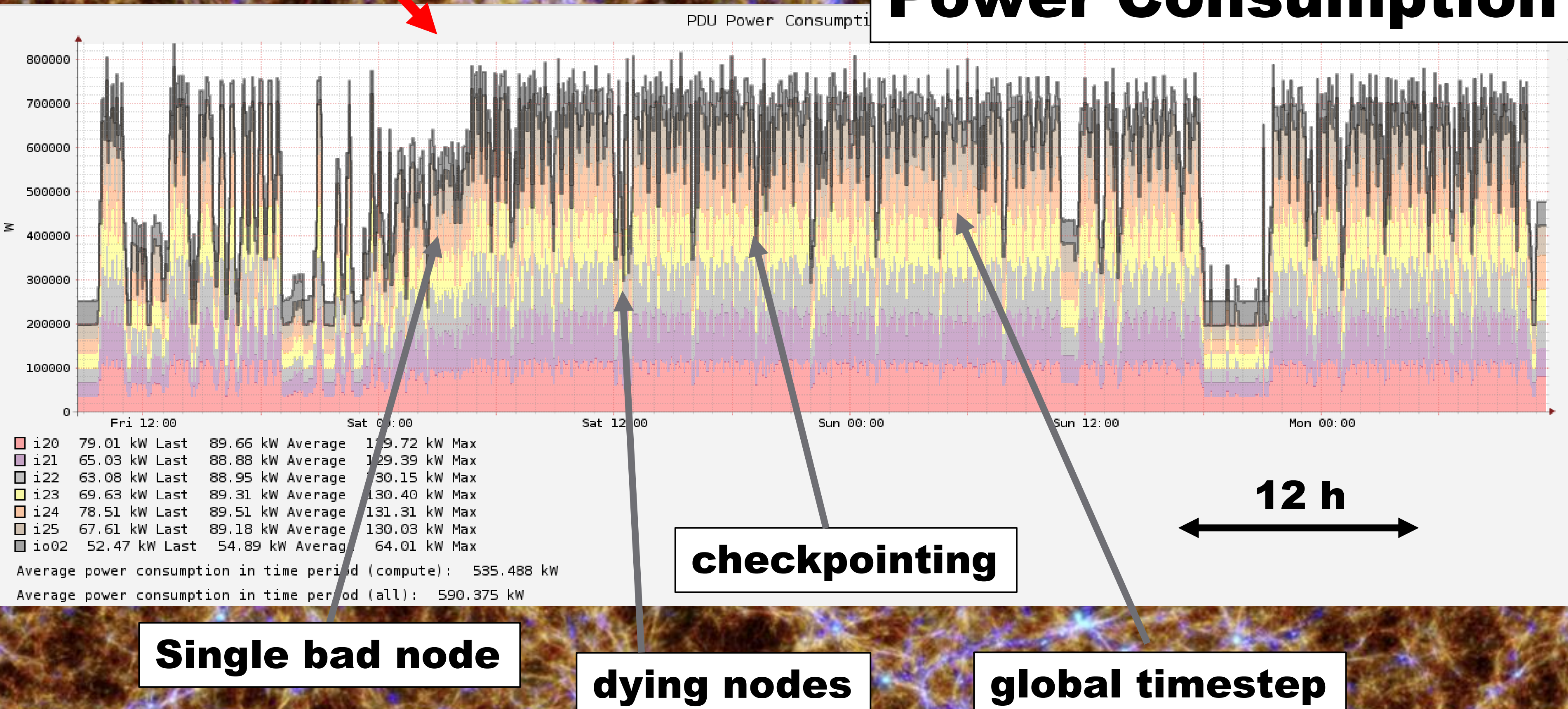
I/O: Gadget internal,
2048 tasks in parallel

Reading
460 sec
152 Gbyte/sec

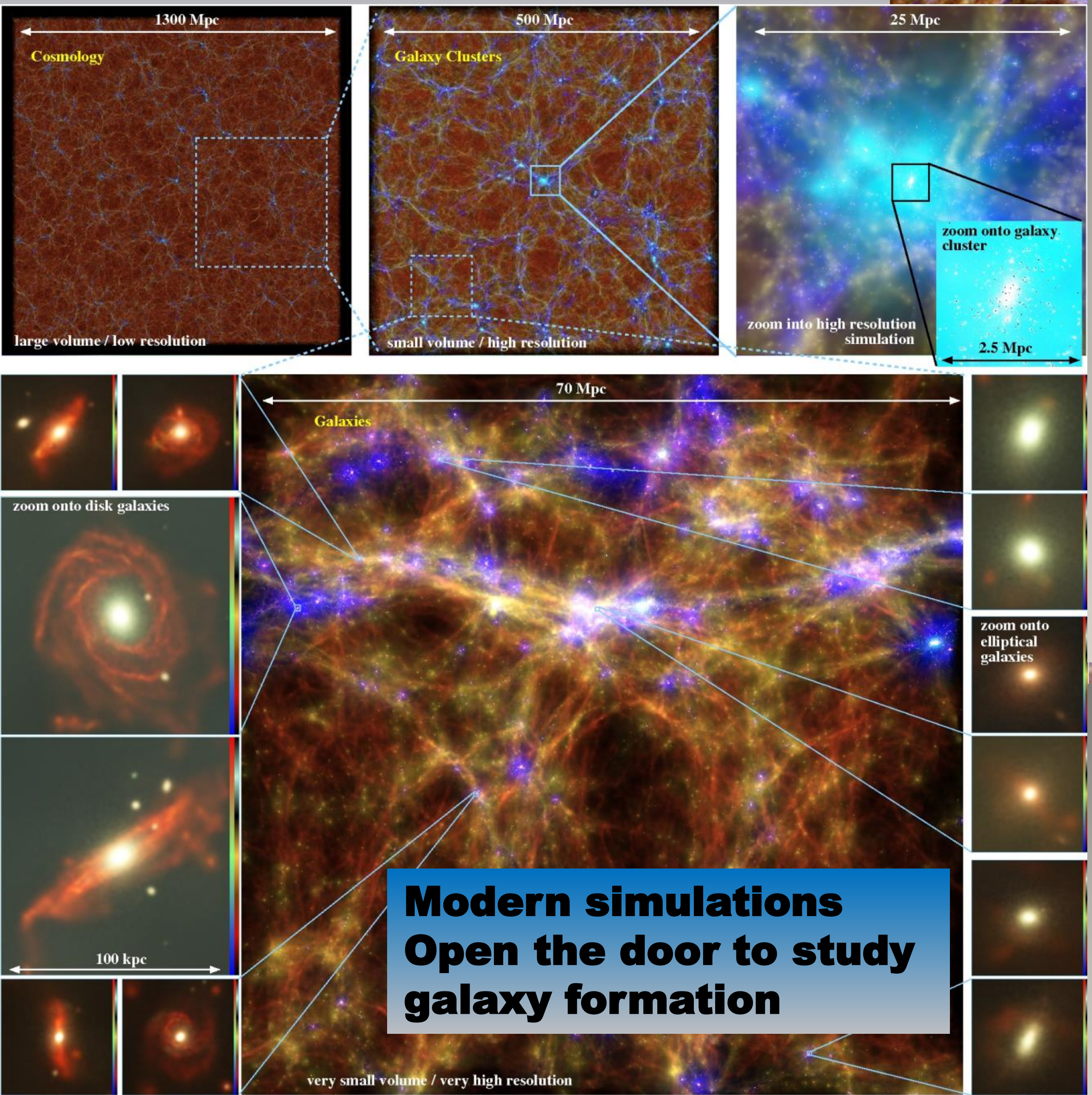
MAGNETICUM

	Box0	Box1	Box2b	Box2	Box3	Box4
[Mpc/h]	2688	896	640	352	128	48
mr	2×4536^3	2×1526^3	–	2×594^3	2×216^3	2×81^3
hr	–	–	2×2880^3	2×1584^3	2×576^3	2×216^3
uhr	–	–	–	–	$2 \times 1536^3 (z = 2)$	2×576^3

Power Consumption



And the rest …



Physics:

cooling+sfr+winds

Springel & Hernquist 2002/2003

Metals cooling

Wiersma et al. 2009

SNIa,SNII,AGB

Tornatore et al. 2003/2006

BH+AGN feedback

Springel & Di Matteo 2006
Fabjan et al. 2010
Hirschmann et al. 2014 (std)
Steinborn et al. 2015 (new)

Thermal conduction

1/20th Spitzer

Dolag et al. 2004

Numerics:

New Kernels: WC6

Dehnen et al. 2012

Low visc. scheme

mr/hr (time dep. alpha)

Dolag et al. 2005

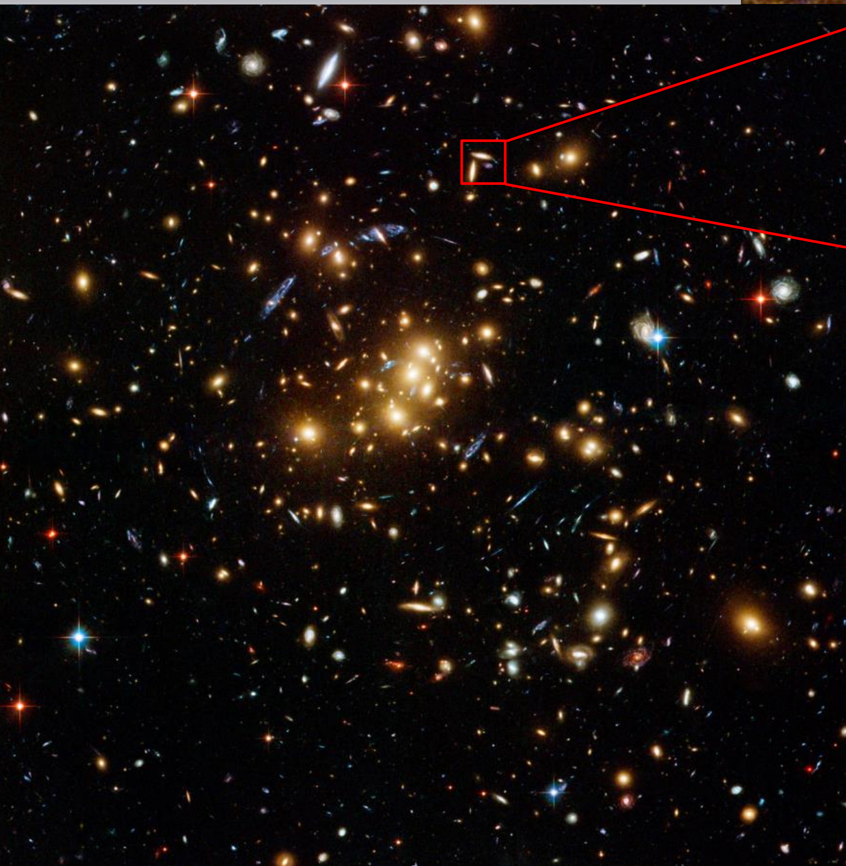
uhr (high order grad.)

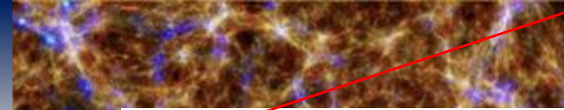
Beck et al. 2015

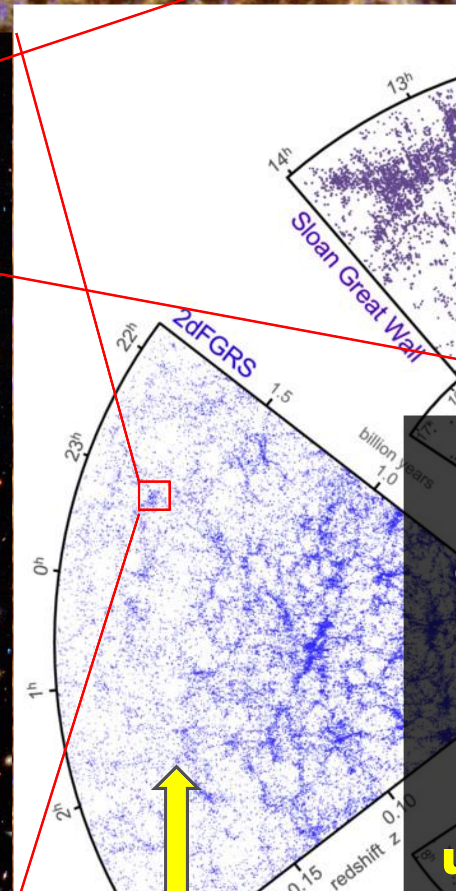
The many ways to form galaxies

https://www.youtube.com/watch?v=Je53JQhpTfk

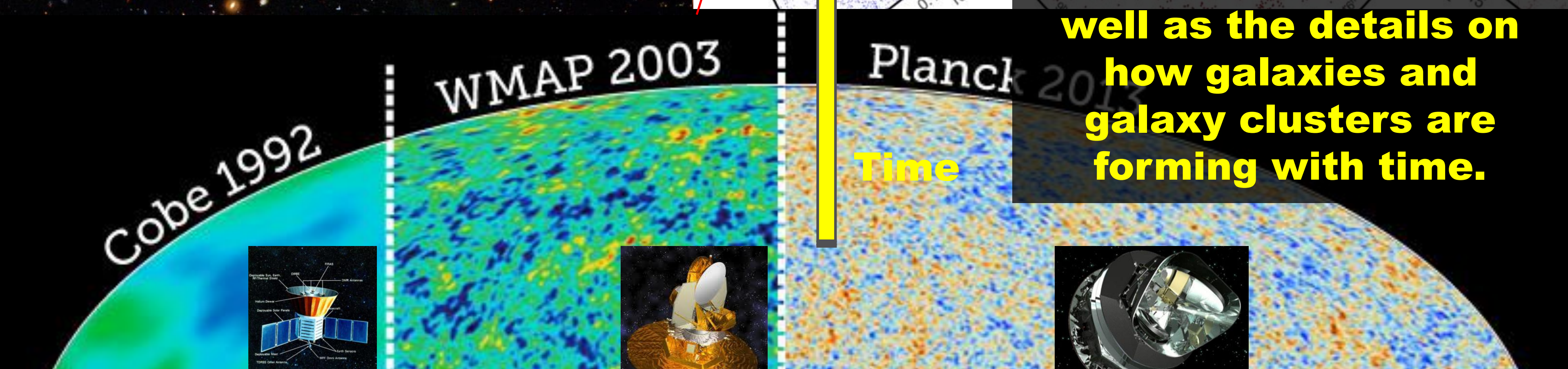
Conclusions







Improvements in computing facilities, code optimization and incorporating various important physical processes are helping to understand the universe throughout as well as the details on how galaxies and galaxy clusters are forming with time.



MAGNETICUM

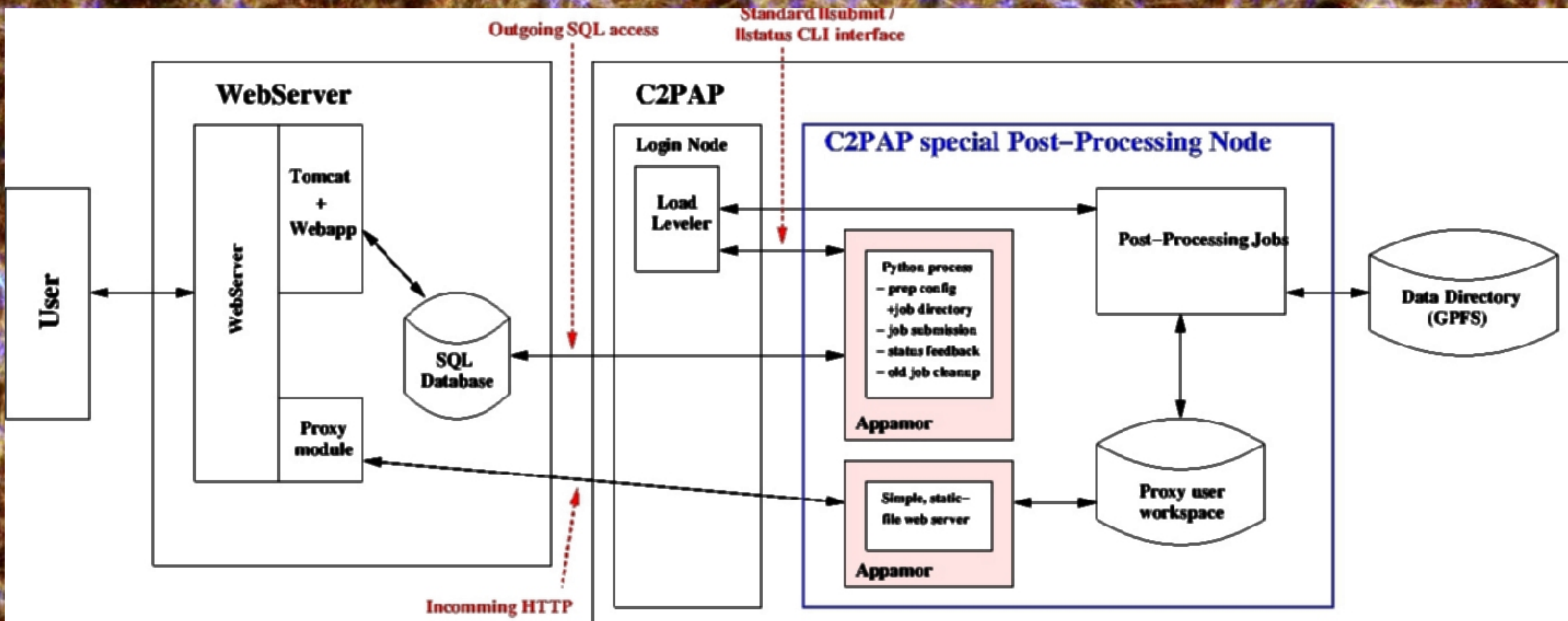
Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften
CAST Computational Astrophysics Group Munich University Observatory
C²PAP Excellence Cluster Universe

Magneticum/Box2/hr
redshift=9.428
g=1.600 c=1.500 b=-0.000

https://www.youtube.com/watch?v=HHh_BcQ6fbQ

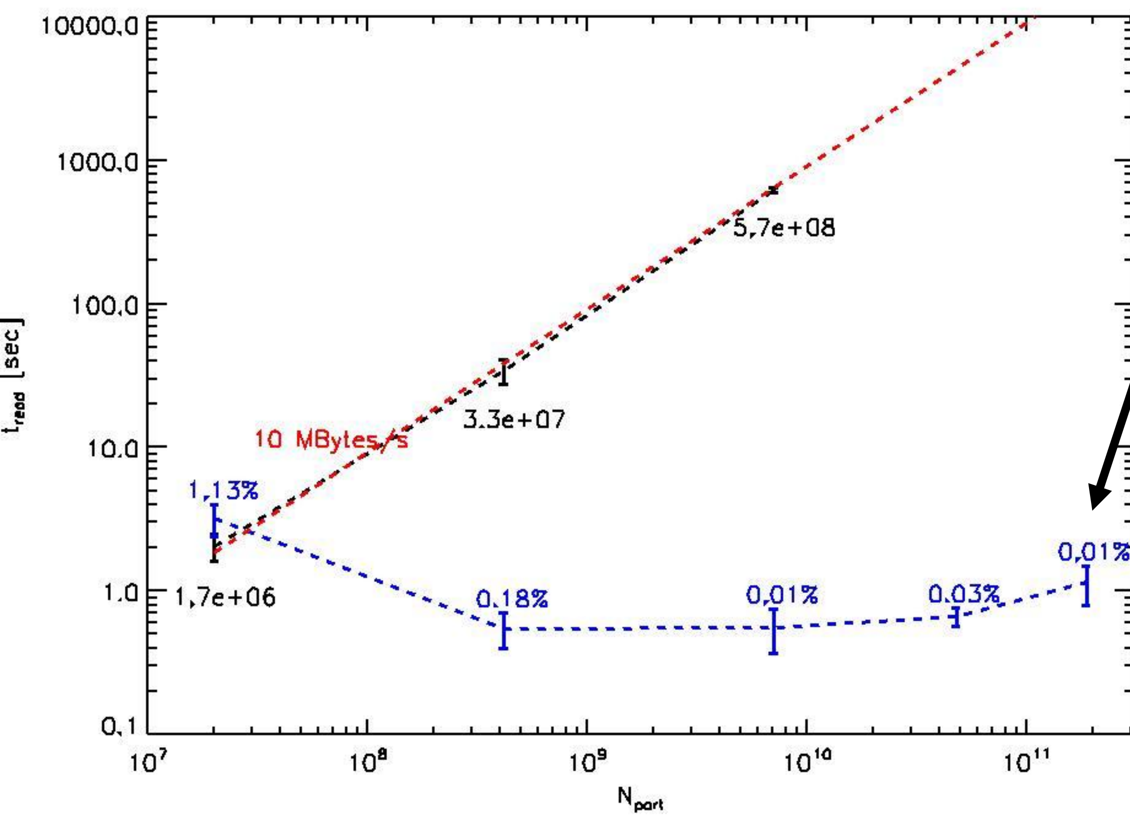
Distributing the Data

General concept of the Cosmosim Web Portal



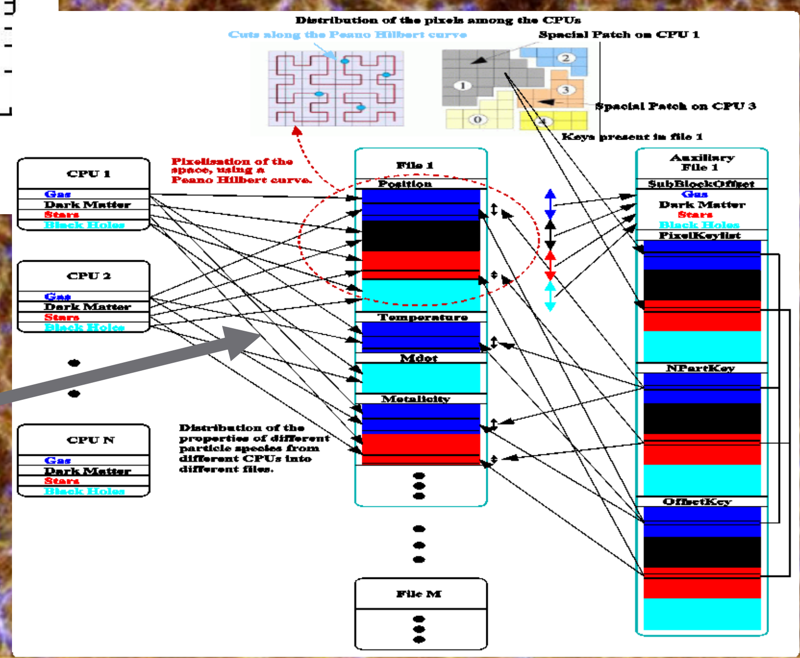
Ragagnin et al, in prep

Testing indexing for file access



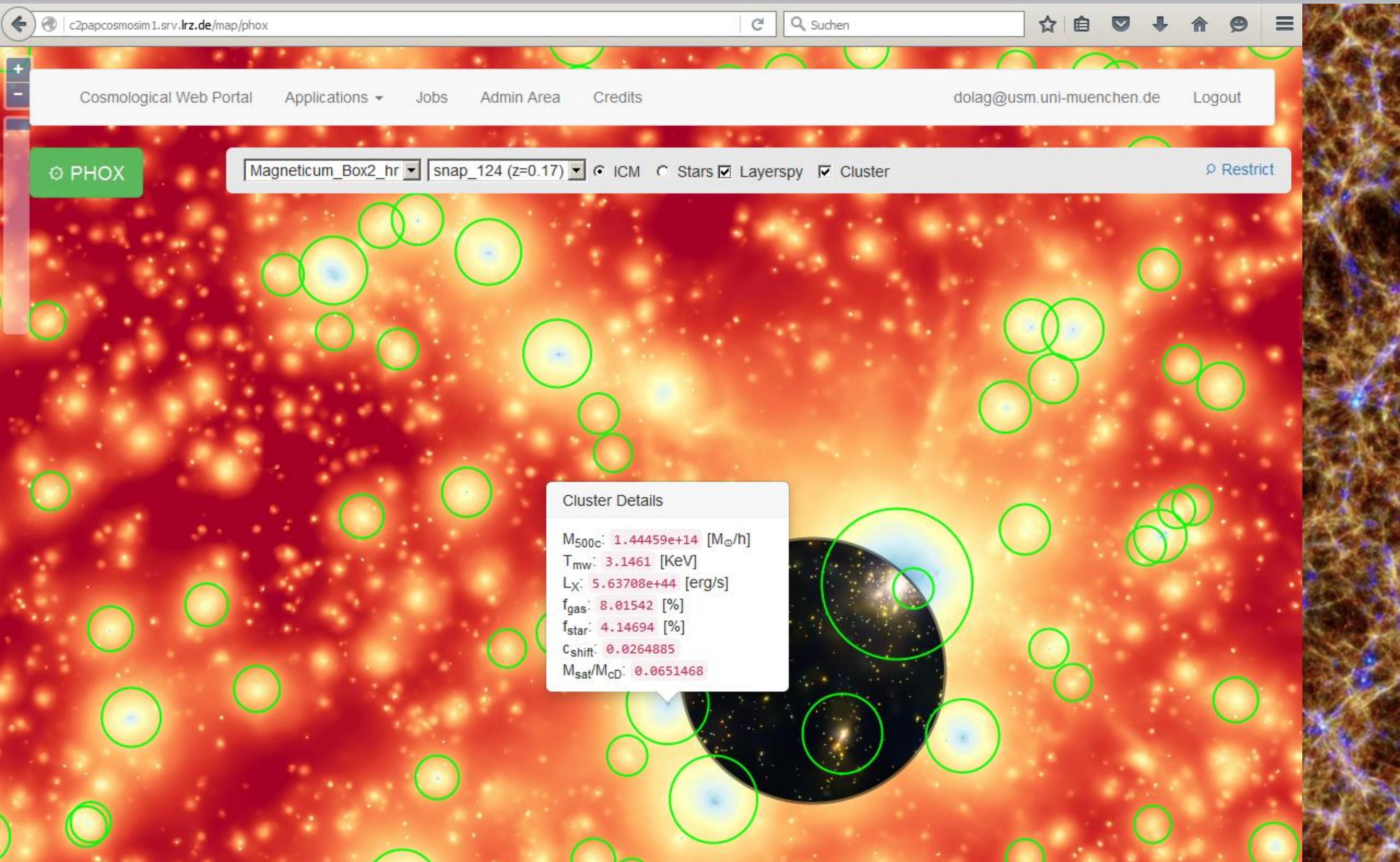
Finding and reading single object out of 20TB particle data.

Would need several hours as single thread !



Indexing utelizing Peano-Hilbert keys

Cosmological Simulation Portal



c2papcososim1.srv.lrz.de

Ragagnin et al, in prep

Conclusions (II):

- **Direct dynamical range of 10^6 almost reached**
  Combination of optimization and growing computing power
- **Further increased by „resonable“ sub-scale models**
  But need to be better validated and improved
- **Next simulations need new strategies**
  work distribution and massive vectorization (memory layout)
- **First success across various scales**
  Global properties of Galaxy Clusters
  AGN properties well reproduced in many respects
  Diversity of CC and non CC clusters
  Intra Cluster light and outskirts of Galaxies
  Morphology of Galaxies
  Internal dynamics of Galaxies
- **Data federalization for hydro sims is challenging**
  Need complex infrastructure