Porting a "Large" computational toolkit to the new generation of Processor

Porting and testing the Einstein Toolkit on the the generation of low-power architectures.

Roberto De Pietri
& Roberto Alfieri
INFN Parma e
Università di Parma

Abstract

Low-Power architectures are subject of much interest also as viable alternatives to traditional HPC platform. In this talk we will focus on

the performance that can now be obtained porting a large simulation toolkit (The EinsteinToolkit), widely used in Numerical Astrophysics to simulated matter coupled to the Einstein's equations, to Low Power Architectures.

We considered multicores / multi node cluster based on ARM and Intel low power processors and we compared results with a traditional HPC cluster, the Galileo system at CINECA. The work has been performed using the resources actually available for the INFN-COSA project.

WE NEED TO BE READY WHEN LOW-POWER SYSTEM WILL BE THE STANDARD HIGH PERFORMANCE ARCHITECTURE

Plan of the talk

- The scientific case: high resolution simulation of inspiral and merger phase of binary neutron stars system (one of source of the gravitational waves that are the observational target of the LIGO/VIRGO experiment)
- Computation performed using the The Einstein ToolKit
  - Description of the code.
- Performance of the code on Tier-1 system: Galileo at CINECA
- COSA low power systems
  - Basic performance analysis
  - Porting of the application
  - Comparative results analysis

More on scientifical motivations

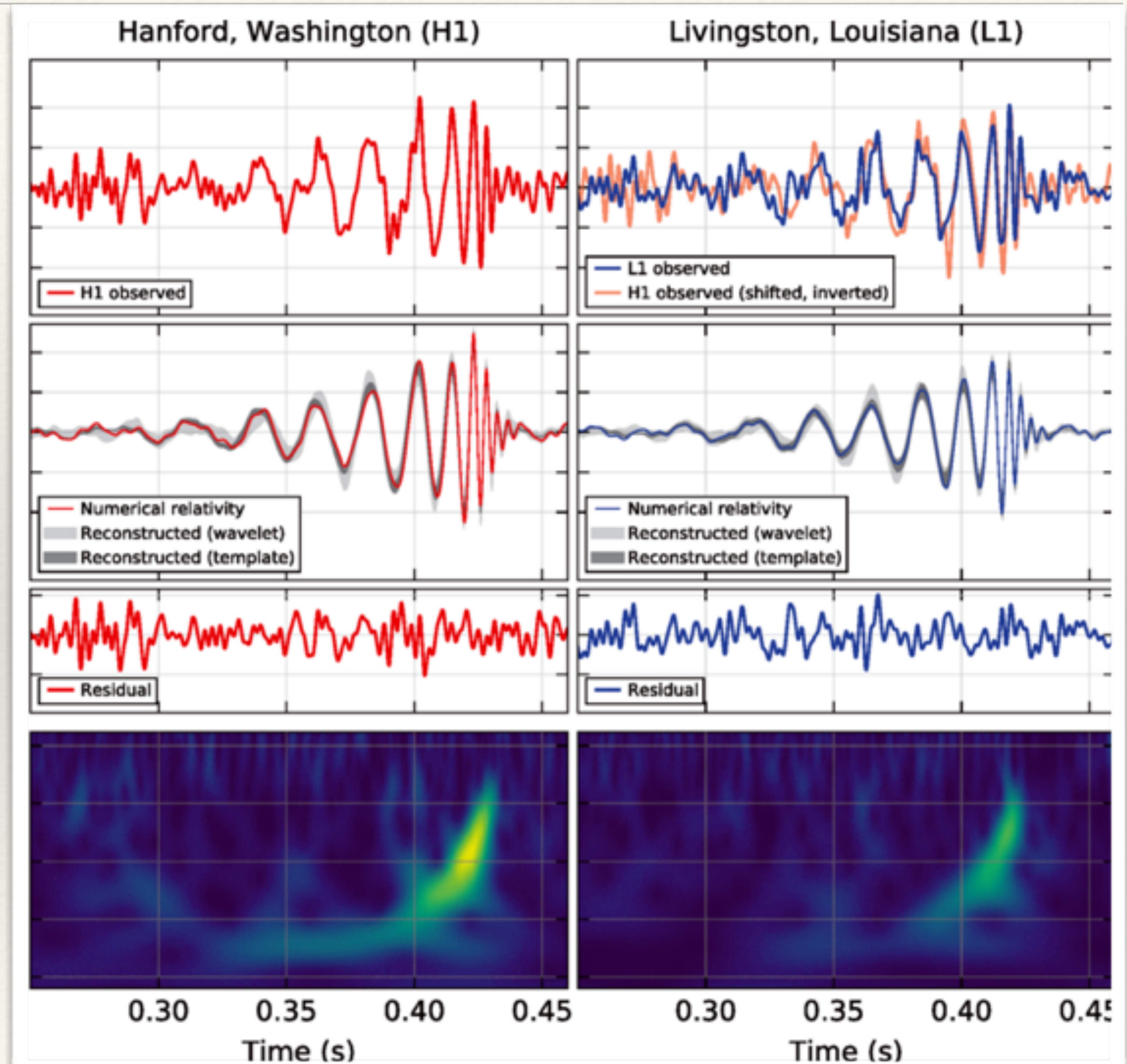
In the eve of Gravitational Wave physics the characterization of the gravitational wave signal emitted by compact binary source will be a prominent role.

We present results for three-dimensional simulations of the dynamics of binary neutron star (BNS) mergers from the late inspiral stage and the post-merger up to ~20 ms after the system has merged, either to form a hyper-massive neutron star (NS) or a rotating black hole (BH). We report here results for equal and un-equal-mass models and on the strength of the Gravitational Signal and its dependence on the EOS, the total ADM mass and the mass ratio of the two stars.

We use a semi-realistic descriptions of the equation of state (EOS) where the EOS is described by a seven-segment piece-wise polytropic with a thermal component given by Γth=1.8. One of the important characteristics of the present investigation is that it is entirely performed using only **publicly available open source software**, the Einstein Toolkit for the evolution and the LORENE code for the generation of the initial models.

Gravitational Wave Astronomy just begun!

- **The gravitational waves were detected on September 14, 2015 at 5:51 a.m. Eastern Daylight Time (09:51 UTC)** by both of the twin Laser Interferometer Gravitational-wave Observatory (LIGO) detectors, located in Livingston, Louisiana, and Hanford, Washington, USA.
- The signal was observed with a matched-filter signal-to-noise ratio of 24 and a false alarm rate estimated to be less than 1 event per 203 000 years, equivalent to a significance greater than **5.1σ**. The source lies at a luminosity distance of **410**(18) **Mpc** corresponding to a redshift z=0.09(4). In the source frame, the initial black hole masses are **36**(5)$\mathbf{M_\odot}$ and **29**(4)$\mathbf{M_\odot}$, and the final black hole mass is **62**(4)$\mathbf{M_\odot}$, with 3.0(5) $\mathbf{M_\odot c^2}$ radiated in gravitational waves. ***All uncertainties define 90% credible intervals***.



Observation of Gravitational Waves from a Binary Black Hole Merger B. P. Abbott et al. (LIGO Scientific Collaboration and **Virgo Collaboration**)
Phys. Rev. Lett. 116, 061102 – Published 11 February 2016

Numerical Relativity in a nutshell

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G\, T_{\mu\nu}$$ **Einstein Equations**

$$\nabla_\mu T^{\mu\nu} = 0$$ **Conservation of energy momentum**

$$\nabla_\mu (\rho\, u^\mu) = 0$$ **Conservation of baryon density**

$$p = p(\rho, \epsilon)$$ **Equation of state**

- Introduce a foliation of space-time
- write as a 3+1 evolution equation
- solve them on a computer !

Ideal Fluid Matter

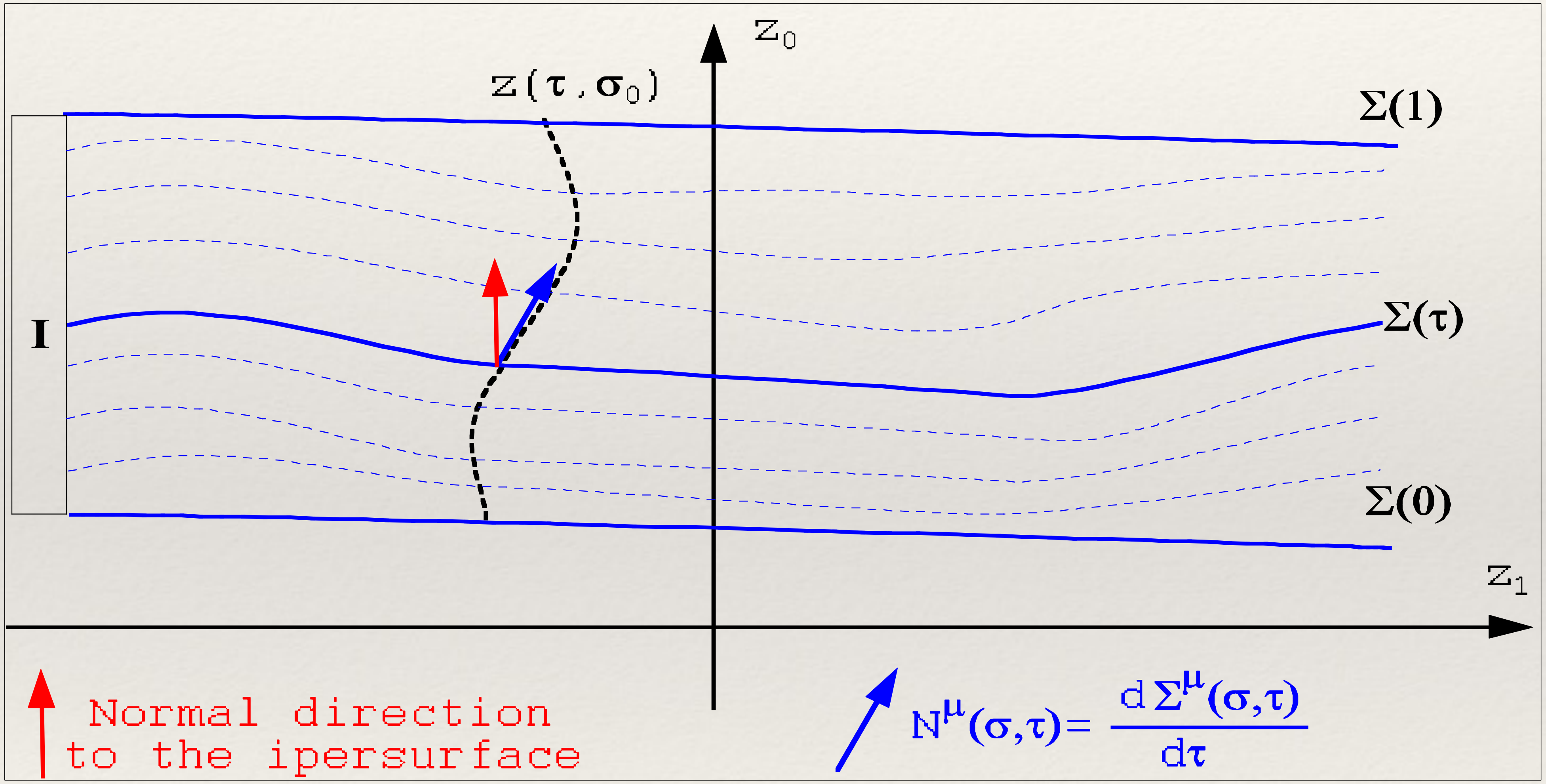
$$T^{\mu\nu} = (\rho(1+\epsilon) + p) u^\mu u^\nu + p g^{\mu\nu}$$

Why Numerical Relativity is hard!

- **No obviously "better" formulation of Einstein's equations**
  - ADM,
  - conformal decomposition,
  - first-order hyperbolic form,.... ???
- **Coordinates (spatial and time) do not have a special meaning**
  - this gauge freedom need to be carefully handled
  - gauge conditions must avoid singularities
  - gauge conditions must counteract "grid-stretching"
- **Einstein's Field equations are highly non-linear**
- **Physical singularity are difficult to deal with**

3+1 formulations of the metric.

$$ds^2 = -(\alpha^2 - \beta_i\beta^i)dt^2 + 2\beta_i dx^i dt + \gamma_{ij}dx^i dx^j$$



α :: lapse

β^i :: shift vector

γ_{ij} :: 3-metric

$$N^\mu \frac{\partial}{\partial x^\mu} = \frac{1}{\alpha}\left(\frac{\partial}{\partial t} - \beta^j \frac{\partial}{\partial x^j}\right)$$

ADM evolutions

$$\partial_t \gamma_{ij} = -2\alpha K_{ij} + \nabla_i \beta_j + \nabla_j \beta_i, \quad (2.1)$$

6 equations for the metric

$$\begin{aligned}\partial_t K_{ij} &= -\nabla_i \nabla_j \alpha + \alpha \Bigg[R_{ij} + K\, K_{ij} - 2K_{im} K^m_j \\ &\quad -8\pi \left(S_{ij} - \frac{1}{2}\gamma_{ij} S \right) - 4\pi \rho_{\mathrm{ADM}} \gamma_{ij} \Bigg] \\ &\quad + \beta^m \nabla_m K_{ij} + K_{im} \nabla_j \beta^m + K_{mj} \nabla_i \beta^m . \end{aligned} \quad (2.2)$$

+6 equations for the time-coordinate derivative of the metric (extrinsic curvature)

Hamiltonian + Momentum constraints

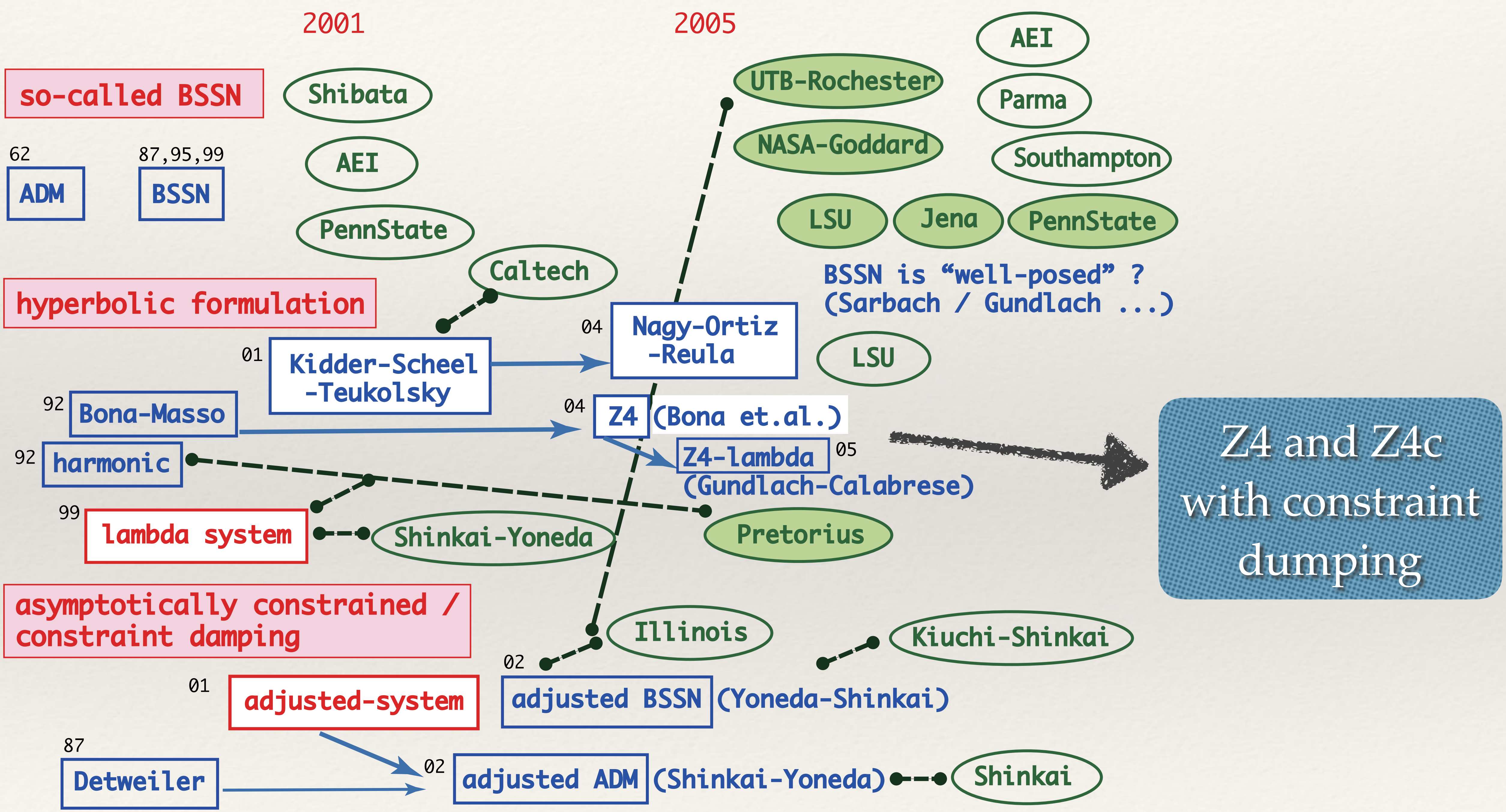
$${}^{(3)}R + K^2 - K_{ij} K^{ij} - 16\pi \rho_{\mathrm{ADM}} = 0$$

+1 constrain equation

$$\nabla_j K^{ij} - \gamma^{ij} \nabla_j K - 8\pi j^i = 0$$

+3 constrain equation

No better formulations……

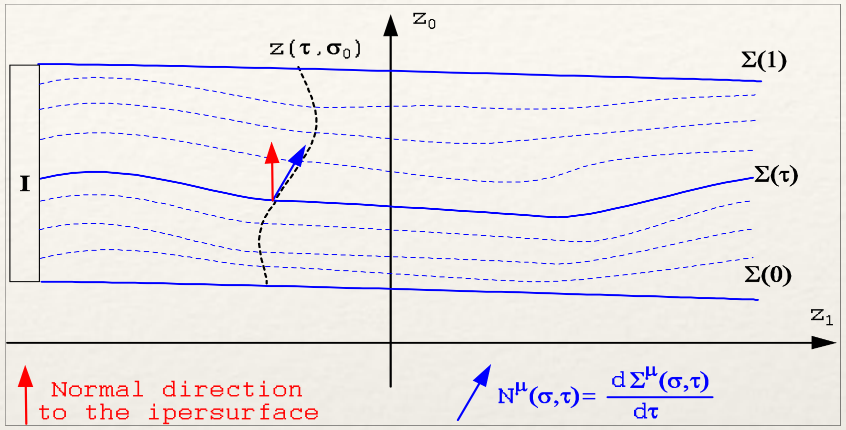


FROM: Hisa-aki Shinka, Formulations of the Einstein equations for numerical simulations, arXiv:0805.0068-

The Einstein EQUATIONS

$$ds^2 = -\alpha^2 dt^2 + g_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt)$$

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 0$$



- BSSN version of the Einstein's equations that introduce additional conformal variables:

$$R_{ij} = -\tfrac{1}{2}\tilde{g}^{lm}\tilde{g}_{ij,lm} - \tilde{g}_{k(i}\partial_{j)}\tilde{\Gamma}^k + \tilde{\Gamma}^k\tilde{\Gamma}_{(ij)k} + \tilde{g}^{lm}(2\tilde{\Gamma}^k_{l(i}\tilde{\Gamma}_{j)km} + \tilde{\Gamma}^k_{im}\tilde{\Gamma}_{klj})$$

$$R^{TF}_{ij} = R_{ij} - \tfrac{1}{3} g_{ij} R$$

$$\partial_t \varphi = -\tfrac{1}{6}\alpha K + \beta^i \partial_i \varphi + \tfrac{1}{6}\partial_i \beta^i$$

$$\partial_t K = -g^{ij}\nabla_i\nabla_j\alpha + \alpha(\tilde{A}_{ij}\tilde{A}^{ij} + \tfrac{1}{3}K) + \beta^i\partial_i K$$

$$\partial_t \tilde{g}_{ij} = -2\alpha K_{ij} + \tilde{g}_{jk}\partial_i\beta^k + \tilde{g}_{ik}\partial_j\beta^k - \tfrac{2}{3}\tilde{g}_{ij}\partial_k\beta^k$$

$$\partial_t \tilde{\Gamma}^i = -2\tilde{A}^{ij}\partial_j\alpha + 2\alpha(\Gamma^i_{jk}\tilde{A}^{jk} - \tfrac{2}{3}\tilde{g}^{ij}\partial_j K + 6\tilde{A}^{ij}\partial_j\varphi) +$$
$$+ \beta^k\partial_k\tilde{\Gamma}^i - \tilde{\Gamma}^k\partial_k\beta^i + \tfrac{2}{3}\tilde{\Gamma}^i\partial_k\beta^k + \tfrac{1}{3}\tilde{g}^{ij}\partial_j\partial_k\beta^k + \tilde{g}^{jk}\partial_j\partial_k\beta^i$$

$$\partial_t \tilde{A}_{ij} = e^{-4\varphi}(-(\nabla_i\nabla_j\alpha)^{TF} + \alpha R^{TF}_{ij}) + \alpha\left(\tilde{A}_{ij}K - 2\tilde{A}_{ik}\tilde{A}^k{}_j\right) - \partial_i\partial_j\alpha +$$
$$+ \beta^k\partial_k\tilde{A}_{ij} + \left(\tilde{A}_{ik}\partial_j + \tilde{A}_{jk}\partial_i\right)\beta^\kappa - \tfrac{2}{3}\tilde{A}_{ij}\partial_k\beta^k$$

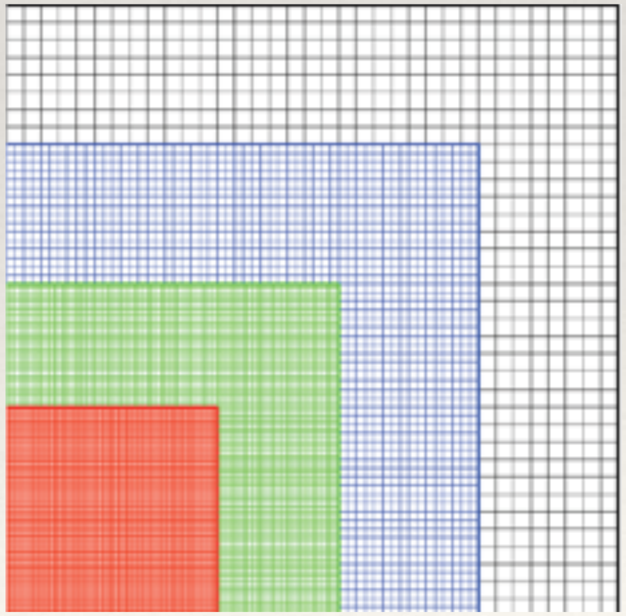
- Matter evolution (B set to zero) using shock capturing methods based on the GRHydro code

[4] M. Shibata, T. Nakamura: "**Evolution of three dimensional gravitational ..**", **Phys. Rev. D52(1995)5429**
[5] T.W. Baumgarte, S.L. Shapiro: "**On the numerical integration of Einstein..**", **Phys. Rev. D59(1999)024007**

+17 additional grid variables to evolves

The code: Einstein TOOLKIT + LORENE

- **Cactus** framework for parallel high performance computing (Grid computing, parallel I/O)
- **Einstein Toolkit** open set of over 100 Cactus thorns for computational relativity along with associated tools for simulation management and visualization
- Mesh refinement with **Carpet**
- Matter Evolution with **GRHydro:**
  (Magnetic+**CT evolution** of Magnetic Field)
  HLLE Riemann Solver
  WENO Reconstruction methods (*)
  PPM Reconstruction methods
- Metric evolution MacClacan:
  BSSN gravitational evolutions (*)
  Z4 gravitational evolutions
- Initial data computed using di **LORENE CODE**



How big is the code size

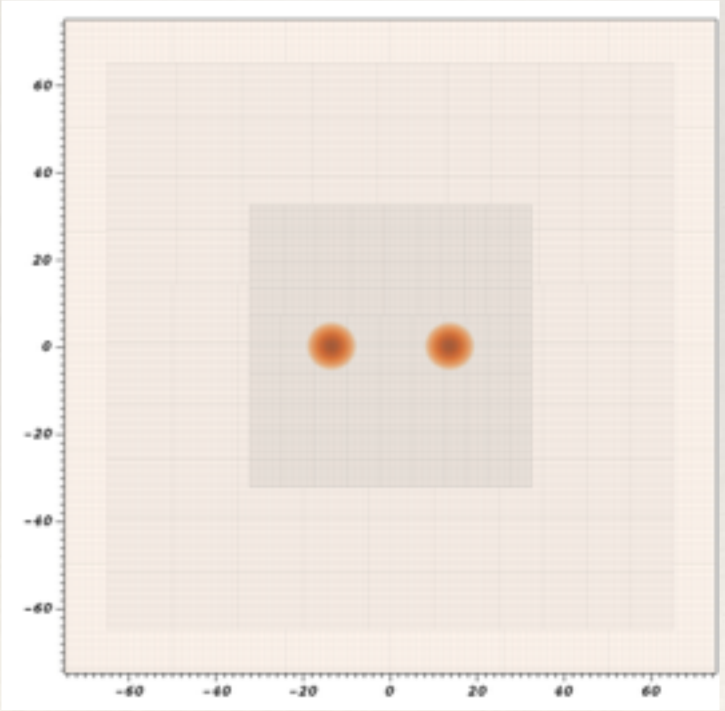
- The ET_2015_05 distribution (just counting the files):
  - F77 source file: 91 (25483 lines)
  - F90 source file: 508 (154515 lines)
  - C99 source file: 870 (419132 lines)
  - C++ source file: 391 (219645 lines)
- Plus the personalization and extra codes.
- **Impossible to rewrite the code for any new architecture !**
- **We do have to relay on COMPILER !**
- **The code must run in a machine neutral way (you can't know on which machine you will obtain an allocation)**

The computational challenge: **minimal requirement**.

- Cartesian grid with at-least 6 refinement levels.
- Standard Resolution in the finest grid **0.25** CU and up to 0.125 CU.
  => from 5,337,100 grid points and up
  to 42,696,800 for each refinement level.

Level	$\min(x/y)$ (CU)	$\max(x/y)$ (CU)	$\min(z)$ (CU)	$\max(z)$ (CU)	(N_x, N_y, N_z) $dx = 0.25$
1	-720	720	0	720	(185,185,96)
2	-360	360	0	360	(205,205,106)
3	-180	180	0	180	(205,205,106)
4	-90	90	0	90	(205,205,106)
5	-60	60	0	30	(265,265,76)
6	-30	30	0	15	(265,265,76)
(7	-15	15	0	7.5)	(265,265,76)

- Outer grid extends to 720M (1063Km) to extract gravitational waves far from the source.
- One extra refinement level added just before collapse to black hole.
- **17 spacetime variables + 4 gauge variables + 5** base variables evolved in each point + all the additional and derived variable needed to formulate the problem.
- **MPI+OpenMP code parallelization** already in place.



Computational environment

(This work would have not been possible without the support of the SUMA INFN project that provided the financial support of the work of AF and the computer resources of the CINECA “GALILEO” HPC Machine, where most of the simulations were performed. Other)

- Project developed almost entirely using computational resource provided by INFN:
- Development: **ZEFIRO** (INFN-PI)
- Production: **GALILEO** (SUMA)

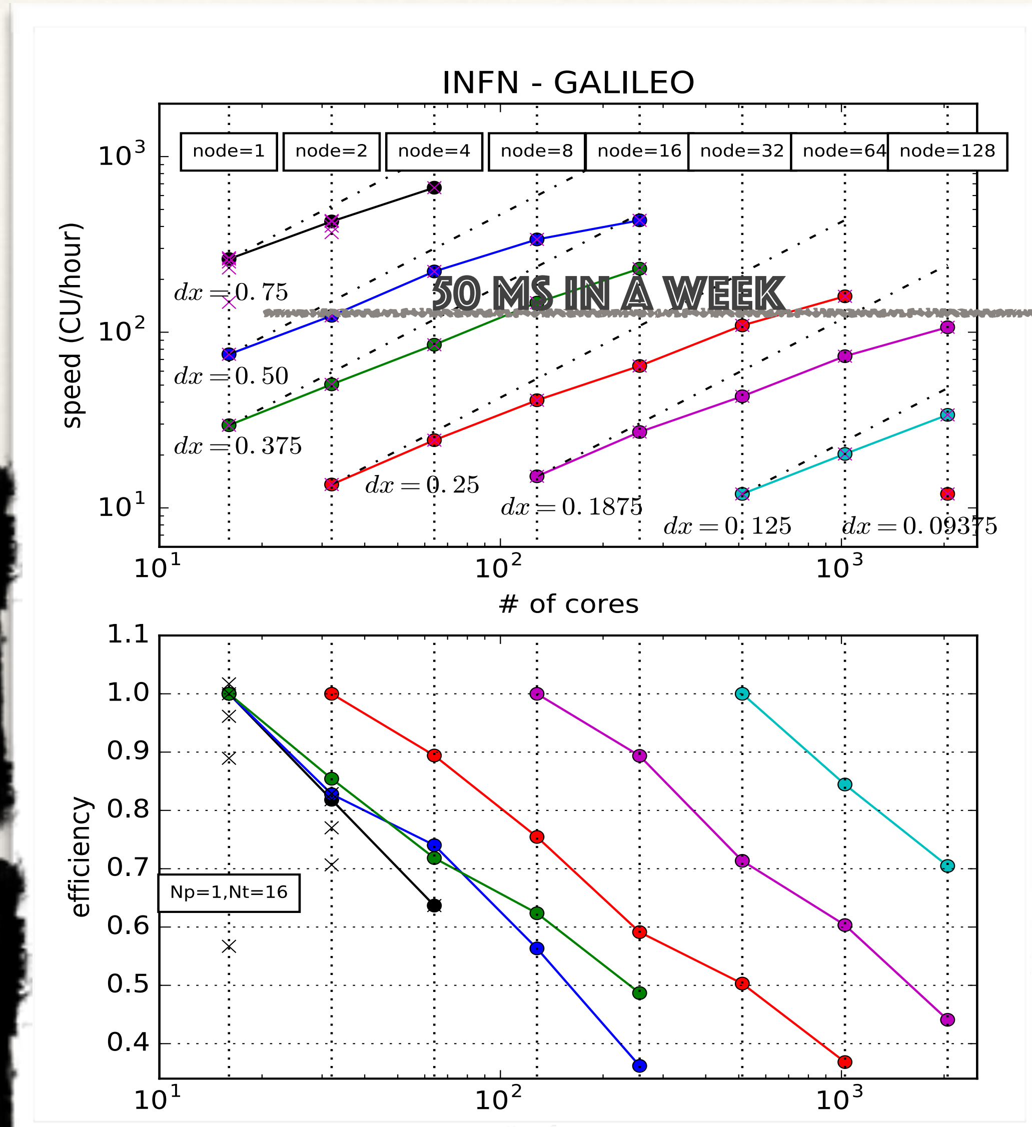
Zefiro is a cluster of the SUMA project SUMA Project,
It's made up of **32** computer each with **512** GB
4 processors (each with 16 cores).
Total of 2048 computing cores
AMD Opteron 6380 (2,5GHz)
Infiniband connection QDR
(switch: 36 ports Mellanox IS5025).

Model: IBM NeXtScale (Lenovo)
Architecture: Linux Infiniband Cluster
Nodes: **516** **RAM**: **128** GB/node, 8 GB/core
Processors: 2 8-cores Intel Haswell 2.40 GHz per node
Cores: 16 cores/node, 8256 cores in total
Accelerators:
2 Intel Phi 7120p per node on 384 nodes (768 in total);
2 NVIDIA K80 per node on 40 nodes
Internal Network: Infiniband with 4x QDR switches
Disk Space: 2.000 TB of local scratch
Peak Performance: 1.000 TFlop/s (to be defined)

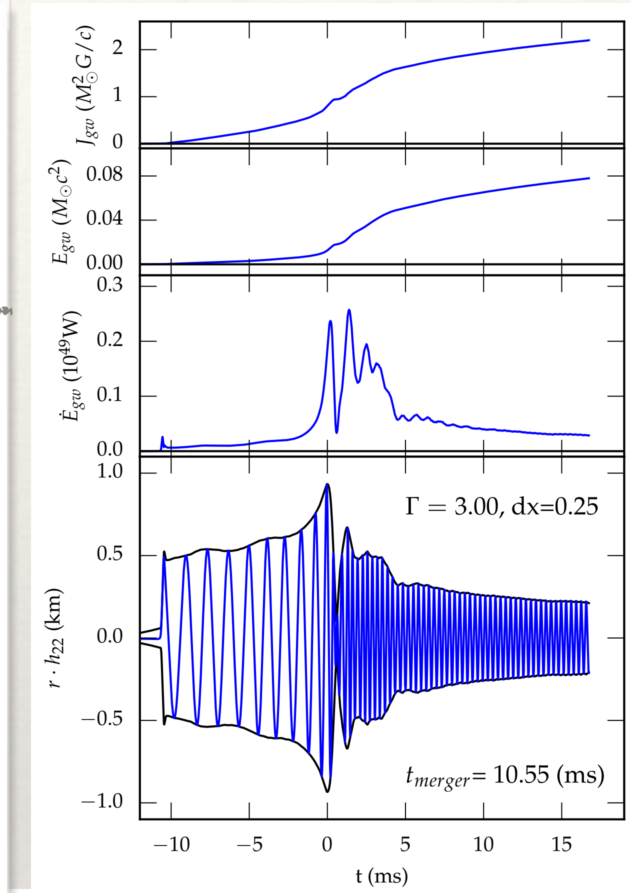
Modeling Equal and Unequal Mass Binary Neutron Star Mergers Using Public Codes

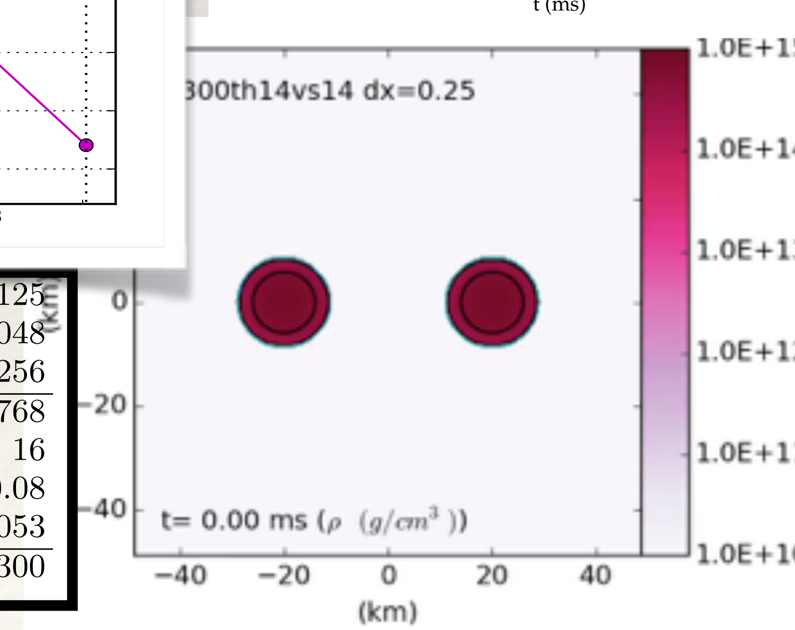
Roberto De Pietri, Alessandra Feo, and Francesco Maione
Parma University and INFN Parma, Parco Area delle Scienze 7/A, I-43124 Parma (PR), Italy

Frank Löffler
Center for Computation & Technology, Louisiana State University, Baton Rouge, LA 70803 USA
(Dated: February 13, 2016)



Δx (CU)	0.75	0.50	0.375	0.25	0.185	0.125
# threads	16	64	128	256	512	2048
# MPI	2	8	16	32	64	256
Memory (GBytes)	3.8	19	40	108	237	768
speed (CU/h)	252	160	124	53	36	16
speed (ms/h)	1.24	0.78	0.61	0.26	0.18	0.08
cost (SU/ms)	13	81	209	974	2915	26053
total cost (kSU, 50 ms)	0.65	4	10.5	49	146	1300

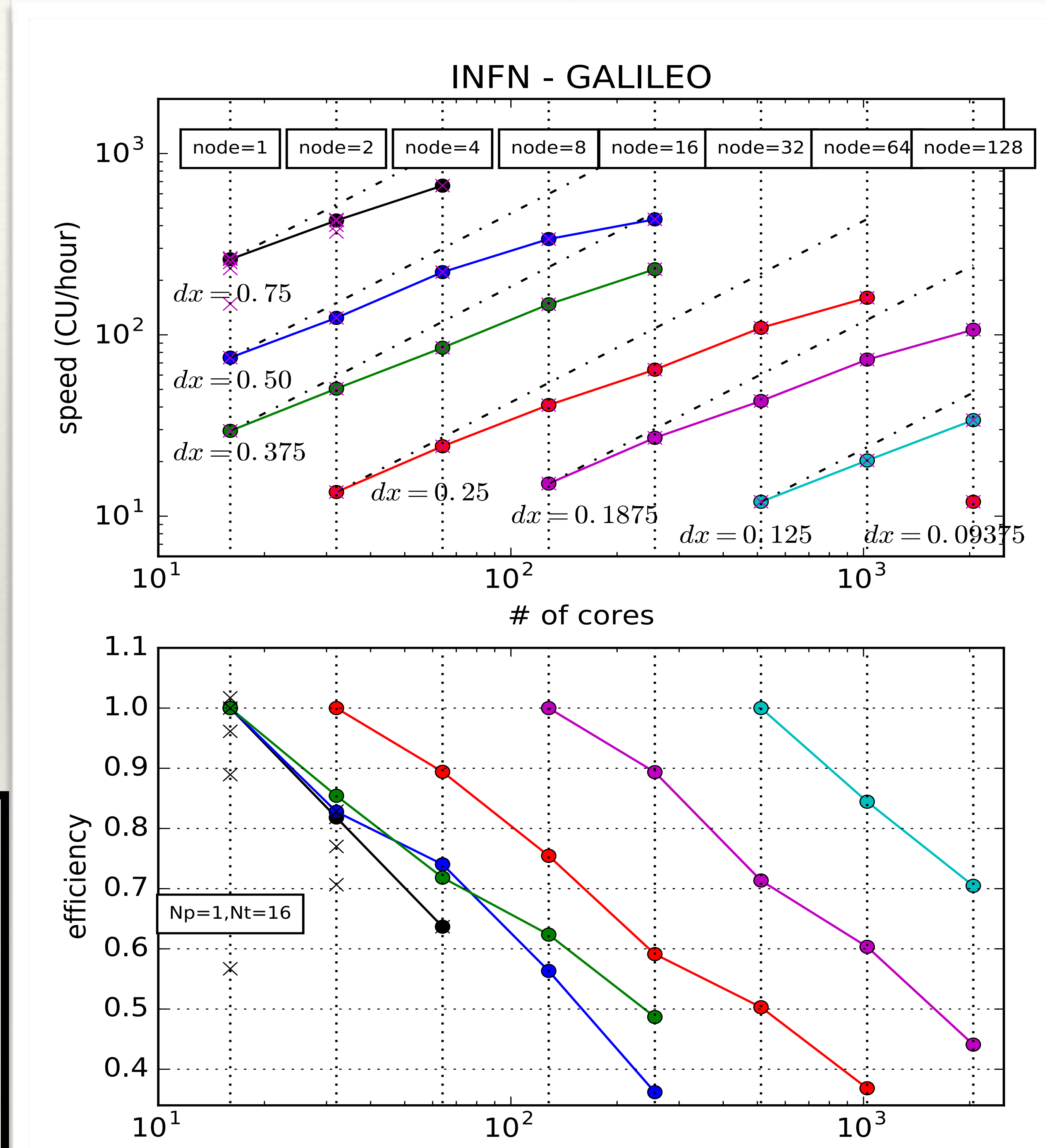




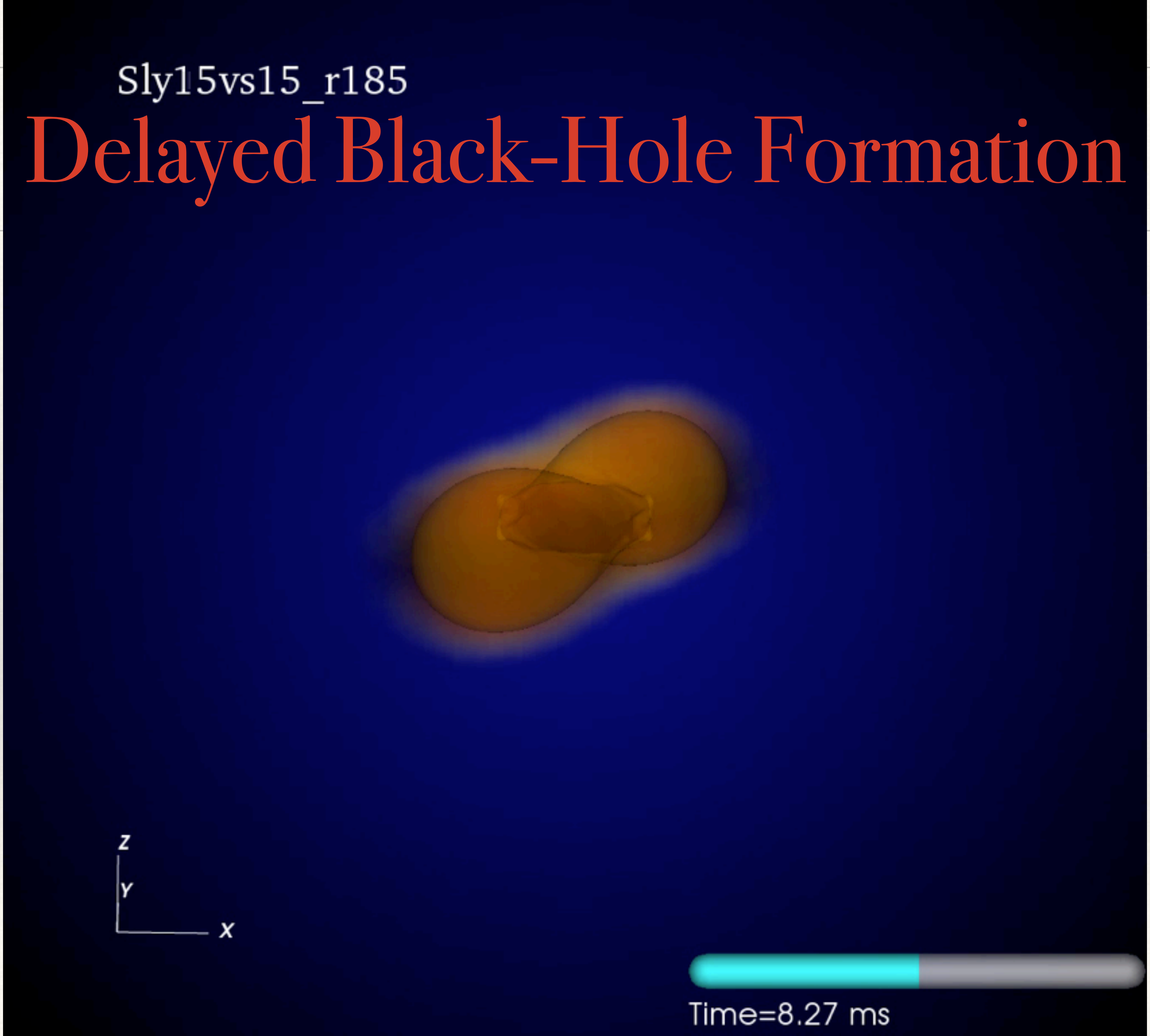
Computational Costs (2)

- Typical run requires at least 108 GByte of allocated RAM (dx=0.25)
- Even the coarser resolution would require 4 GByte of allocated RAM (dx=0.75)

Δx (CU)	0.75	0.50	0.375	0.25	0.185	0.125
# threads	16	64	128	256	512	2048
# MPI	2	8	16	32	64	256
Memory (GBytes)	3.8	19	40	108	237	768
speed (CU/h)	252	160	124	53	36	16
speed (ms/h)	1.24	0.78	0.61	0.26	0.18	0.08
cost (SU/ms)	13	81	209	974	2915	26053
total cost (kSU, 50 ms)	0.65	4	10.5	49	146	1300



Delayed Black-Hole Formation

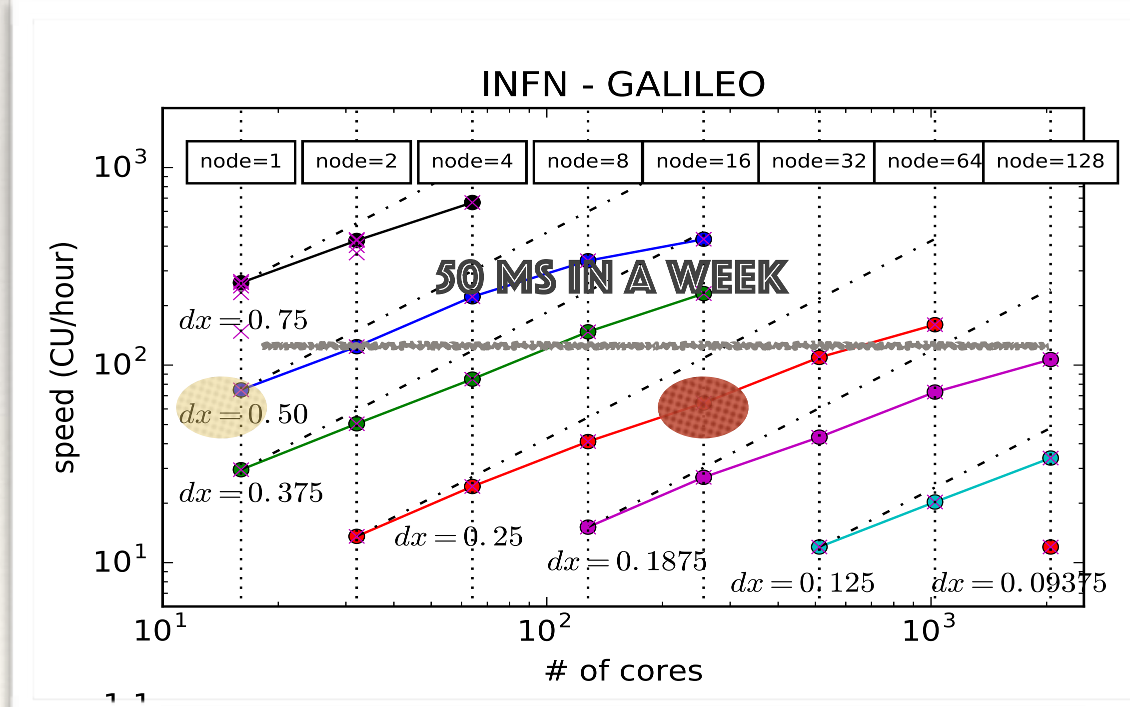


Production configuration used on COSA systems

- The same code, grid structure, initial data … used to perform the production runs on HPC computers (Fermi,**GALILEO**, queenbee, super mike, stampede,…) **but** additional symmetry imposed and use of of a very course resolution of dx=0.75 (not 0.25) to have a memory configuration below 2-GBytes.

- Performance on **GALILEO**:

```
statistics from CarpetLib:
Current allocated memory:   1715.999 MB
Maximum number of objects: 1146
Maximum allocated memory:   1722.232 MB
SPEED AT ITERATION 256 --->
    (1 CORE)   68.1962289 CU/hours
    (2 CORE) 112.825
    (4 CORE) 178.3600
    (8 CORE) 241.6999574
```



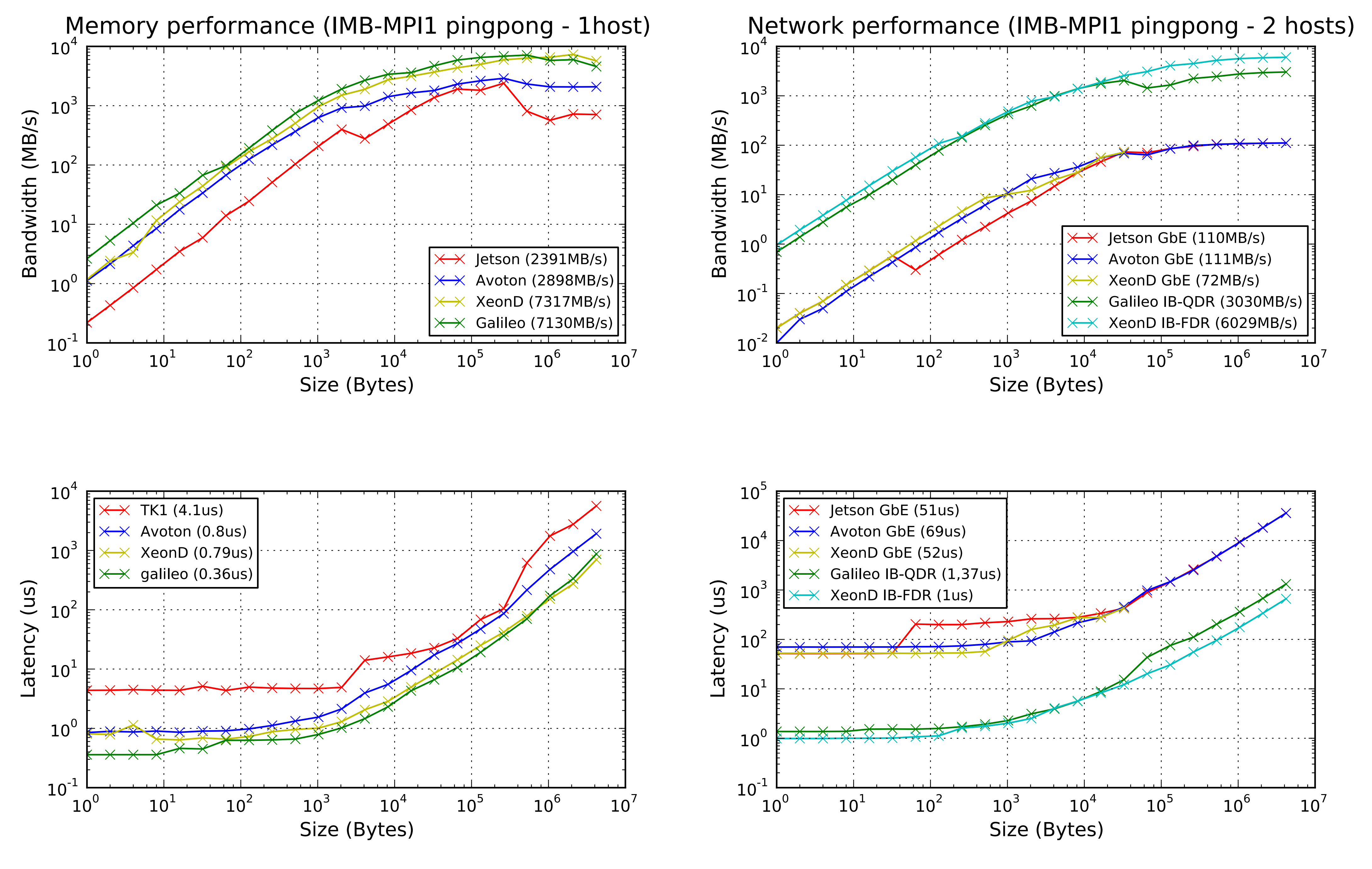
- Not a test code … but the actual code used in research !

Galileo and COSA architectures

System	Node	HS06 (thanks to D.Cesini)	GPU	Network
Galileo	2 XEON E5-2630, 64 bit, 2.4 GHz Cores: 16 cores (32 threads) Ram: 128GB TDP= 100 W x 2	359 (ht) **278** (ht off) 3.5 HS06/W (ht)	Nvidia K80 4992 Cuda cores 24GB GDDR5 8740 GFlops SP 2900 GFlops DP	**Infiniband** QDR 4x (16 Gb/s)
XEON-D	XEON D-1540, 64 bit, 2.0 GHz Cores: 8 (16 threads) Ram: 16 GB 90 W	151 (ht) **133** (ht off) 1.68 HS06/W (ht)		**Infiniband** FDR
Avoton	Atom C2750, 64 bit, 2.4 GHz Cores: 8 (no threads) Ram: 16 GB 24 W	**55** (ht) 2.2 HS06/W (ht)		GbE
Jetson TK1	ARM-A15, 32bit, 2.3 GHz Cores: 4 Ram: 2 GB 14 W	**28** 2 HS06/W	Nvidia Kepler SoC 192 Cuda cores 326 Gflops SP	GbE
Jetson TX1 (future)	ARM A57/A53, 64 bit, 2 GHz Cores: 4 A57 + 4 A53 Ram 4 GB	?	Nvidia Maxwell SoC 256 Cuda cores 512 GFlops SP	GbE

Basic performance analysis (MPI)

- **Memory and Network Bandwidth (experimental) of the systems.**

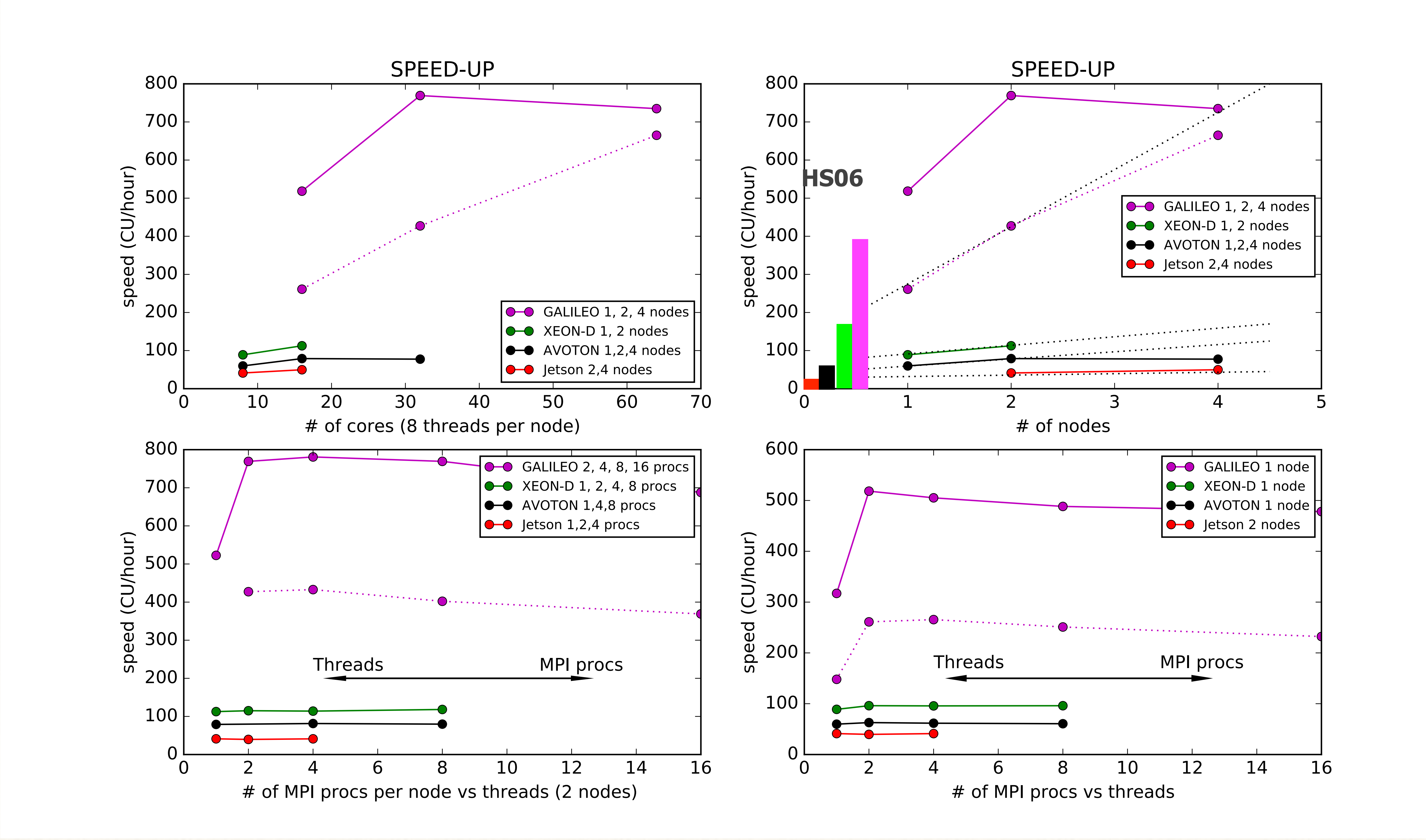


Memory and Network performance determined with the IMB-MPI intel's benchmark

Node Performance

System	Node	HS06 (thanks to D.Cesini)	GPU	Network	ET Speed (maximal on a node)
Galileo	2 XEON E5-2630, 64 bit, 2.4 GHz Cores: 16 cores (32 threads) Ram: 128GB TDP= 100 W x 2 (?- solo processore)	359 (ht) **278** (ht off) 3.5 HS06/W (ht)	Nvidia K80 4992 Cuda cores 24GB GDDR5 8740 GFlops SP 2900 GFlops DP	**Infiniband** QDR 4x (16 Gb/s)	**518** [261] (1x2x8=16 thread) 769 [427] (2x2x8=32 thread) 734 [665] (4x2x8=64 thread)
XEON-D	XEON D-1540, 64 bit, 2.0 GHz Cores: 8 (16 threads) Ram: **16** GB 90 W	151 (ht) **133** (ht off) 1.68 HS06/W (ht)		**Infiniband** FDR	**96,12** (1x2x4=8thread) 114.97 (2x2x4=8thread)
Avoton	Atom C2750, 64 bit, 2.4 GHz Cores: 8 (no threads) Ram: **16** GB 24 W	**55** (ht) 2.2 HS06/W (ht)		GbE	**62.86** (1x2x4=8 thread) 81.86 (2x4x2=16 thread) 79.07 (4x2x4=16 thread)
Jetson TK1	ARM-A15, 32bit, 2.3 GHz Cores: 4 Ram: **2** GB 14 W	**28** 2 HS06/W	Nvidia Kepler SoC 192 Cuda cores 326 Gflops SP	GbE	**41,22** (2x1x4=8 thread) 49,72 (4x1x4=16 thread)
Jetson TX1 (future)	ARM A57/A53, 64 bit, 2 GHz Cores: 4 A57 + 4 A53 Ram **4** GB	?	Nvidia Maxwell SoC 256 Cuda cores 512 GFlops SP	GbE	?

Speed-up



SPEED-UP
speed (CU/hour)
# of cores (8 threads per node)
GALILEO 1, 2, 4 nodes
XEON-D 1, 2 nodes
AVOTON 1,2,4 nodes
Jetson 2,4 nodes
SPEED-UP
HS06
speed (CU/hour)
# of nodes
GALILEO 1, 2, 4 nodes
XEON-D 1, 2 nodes
AVOTON 1,2,4 nodes
Jetson 2,4 nodes
speed (CU/hour)
# of MPI procs per node vs threads (2 nodes)
GALILEO 2, 4, 8, 16 procs
XEON-D 1, 2, 4, 8 procs
AVOTON 1,4,8 procs
Jetson 1,2,4 procs
Threads
MPI procs
speed (CU/hour)
# of MPI procs vs threads
GALILEO 1 node
XEON-D 1 node
AVOTON 1 node
Jetson 2 nodes
Threads
MPI procs

Conclusions

- **GOOD NEWS**: the framework works on LOW-POWER ARCHITECTURE.
- **BAD NEWS:** performance not up to the par of traditional High-End Processor. Memory limitation would require an even higher number of nodes interconnected with a high speed network.
- **FUTURE HARDWARE REQUIREMENTS**: In order to run our application on Low Power architectures at production level we need to exploit the accelerator present on the system (GPU) in order to speed up the computation; moreover we need nodes with larger amount of memory and a high speed network interconnection.
- **FUTURE PROGRAMMING REQUIREMENTS**: Hardware HPC architectures evolve too fast with respect the capacity of a large scientific collaboration to modify the code to support new features. We needs new programming paradigms able to transparently support new hardware features and to guarantee the portability of the code. Exploration of the OpenMP 4.0 framework just started.

Acknowledgment: the INFN computational resource had allowed to develop the present computational activity in numerical simulation of Einstein's Gravity Theory.

- The research activity in numerical relativity in ITALY benefit from the support of INFN to its computational efforts.
- Nel 2001 la commissione IV finanzia un cluster per ricerche di gravita numerica: il cluster **Albert100.** Installato presso il gruppo collegato di Parma.
- 2005 PRIN on numerical relativity by member of INFN-IS OG51 (now TEONGRAV). Realizzazione del cluster **Albert2** (a Parma).
- 2009 Cluster GRID-enabled **TRAMONTANA** (iniziativa di commissione IV)
- 2013 Cluster **ZEFIRO** (Commissione IV e SUMA).Our development platform.
- 2015 INFN Large prototype and Tier-1 **Galileo** system (Progetto Premiale SUMA)
- LOW POWER system research. Testing the code on the COSA system

- **Our thanks to INFN !**