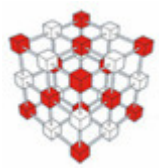
Project:

Constructing, testing, and utilizing a next-generation multi-TFLOP hybrid GPUCPU cluster

Dejan Vinković
Physics Department, Faculty of Natural Sciences and Mathematics, University of Split, Croatia



1st report to the National Foundation for Science, Higher Education and Technological Development of the Republic of Croatia (NZZ)

April 30, 2009
Split, Croatia

1. Project background

The project was approved by the Board of NZZ on May 29, 2008, for the total of 744,700.00 kunas. The contract for financing the project was signed on September 10, 2008, and the project officially started on October 1, 2008. The first payment of 631,900.00 kunas was received in October, 2008. Project associates include: Mario Jurić (Institute for Advanced Study, Princeton, USA), Günther Wuchterl (Thüringer Landessternwarte Tautenburg,Tautenburg, Germany), Petar Mimica (University of Valencia, Spain), Vanja Klepac-Ceraj (Harvard Medical School, Harvard University, USA), Kristian Vlahoviček (Department of Molecular Biology, University of Zagreb, Croatia), Leandra Vranješ Markić (Physics Department, University of Split, Croatia), Hrvoje Buljan (Department of Physics, University of Zagreb, Croatia), Željka Fuchs (Physics Department, University of Split, Croatia), Željko Ivezić (Department of Astronomy, University of Washington, Seattle, USA).

2. Project timetable for the first 6 months

The timetable for the first 6 months presented in the signed contract is as follows:

Phase	Activity	Results	Time (start-end months)
1. Initiation of the project	Purchase and installation of a workstation for the project leader.	Working conditions for the project leader established	1.
	Hiring of a student novice (PhD student). Purchase and installation of a workstation for the novice.	A novice (PhD student) hired and his/her working conditions established.	1. – 2.
2. Configuration of the cluster	Collecting offers for the purchase of cluster components. Purchasing the cluster components.	Cluster components purchased for the best price available.	1. – 2.
	Assembling the cluster	The cluster in basic operational mode.	2.-5.
3. The cluster in full production mode	Purchase and installation of a workstation for visitors (project partners).	Working conditions for visitors (project partners) established	6.
	Purchase of the basic technical equipment for the lecture room.	Secured conditions for providing training lectures	5.-6.
	Visits by project partners	Regular usage of the cluster	5.-12.
	Developing manuals for cluster and GPU usage	Informative instructions for cluster usage	5.-12.
	Development of CPU-GPU applications for the cluster	Speed up of science and technology software thanks to the cluster's GPU technology	5.-12.

3. Project activities in the first 6 months

The main activities of the project in this period were securing the infrastructural support for the project, purchasing the equipment, additional fundraising and expanding the collaborative network.

3.1. Infrastructure and equipment

Purchasing equipment and securing the space for the cluster and visualization lecture room turned out to be more difficult than initially envisioned. There were three major problems: slow response from NVIDIA regarding their new high-end GPUs, lack of interest of the University to give additional support to the project and undeveloped market of high-end computer solutions in Croatia. All together, the project is behind the schedule when it comes to assembling and utilizing the cluster by one or two months (depending on a particular activity).

NVIDIA's next generation high-end GPUs were released in the fall 2008 under the name Tesla C1060 (individual cards) and Tesla S1070 (4 GPUs configured into a 1U rack-mount) (figure 1). Note that we use the name G9x in the project proposal because we did not know the name at the time of writing the proposal. We were in contact with NVIDIA representatives in the US and asked them to provide us with the price quotes. After an unusually long waiting time, they responded and redirected us to their representative in Europe. After that we received the quotes from their reseller OCF in the UK on December 10, 2008. The price for S1070 was £4761 and for C1060 was £1026 (excluding VAT, delivery or any other taxes). At the same time we searched for a company in Croatia that has an NVIDIA reseller as a business partner. Once we had the quote from OCF we were able to compare it with the offer from the Croatian company. The conclusion was that the direct purchase from abroad would not have a price benefit after we include taxes, shipping and customs.

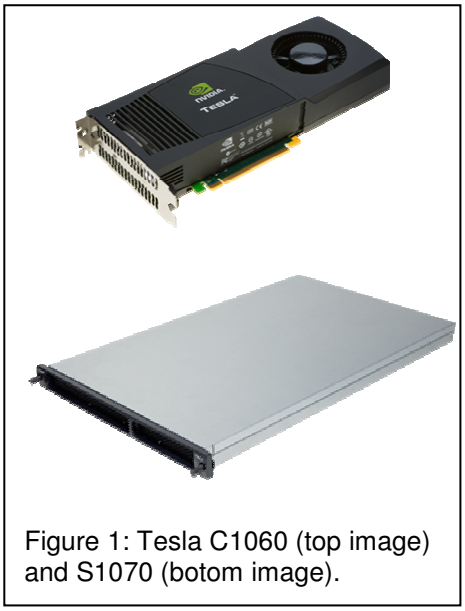


Figure 1: Tesla C1060 (top image) and S1070 (botom image).

Unfortunately, this long wait for the decision on Tesla cards slowed down the design of the whole cluster. Other cluster components had to be picked according to the selected GPU configuration. We had to decide if we were going to use only Tesla C1060 cards, only Tesla S1070 systems or a combination of both. Our final decision based on price quotes and project goals (the most efficient utilization of GPUs) was to order only Tesla S1070 systems. This configuration is also recommended by NVIDIA because S1070 is designed for cluster environments. In such a configuration one Tesla S1070 is connected to two CPU nodes, which enabled us to use 16 high-end GPUs connected to 8 nodes with 2 quad CPUs each, instead of originally planned 16 nodes. This way the higher-than-planned total cost of cluster's GPUs, UPS and cooling was compensated by the smaller number of nodes without loosing the targeted cluster performance. We keep open a possibility of buying Tesla C1060 in the

Date dd/mm/yy	Price	Receipt number	Description
08/12/08	9.999,00	8451	Computer for the project applicant
17/12/08	8.919,66	08-300-001700	Disk storage and backup
17/12/08	4.559,51	08-300-001703	The main server
22/12/08	57.981,72	08-300-001741	Tesla S1070 system components
26/01/09	17.496,02	01-09	Air-conditioning system and installation
26/01/09	6.771,00	0000127-1000043	36U frame rack
27/01/09	7.400,00	3/2009	Adaptation and installation of UPS
28/01/09	33.428,00	0000162-1000043	UPS
03/02/09	8.595,00	06/09	Installation of acoustic isolation
06/02/09	154,49	019-2009	Connectors for electric cables
06/02/09	59.353,00	09-300-000107	Tesla S1070 system components
09/02/09	56.012,64	09-300-000115	Tesla S1070 system components
10/02/09	62.693,36	09-300-000124	Tesla S1070 system components
10/02/09	3.384,39	0000286-1000043	Switch, cables
12/02/09	24.387,80	0000322-1000043	Components for the frontend
23/02/09	26.095,80	0000413-1000043	Components for the nodes
03/03/09	850,00	0000523-1000043	Monitor for the frontend
09/03/09	1.555,57	86	Small items for the cluster power setup
13/03/09	38.548,25	0000600-1000043	Components for the nodes
13/03/09	29.118,18	0000602-1000043	Components for the nodes
13/03/09	12.078,00	0000603-1000043	Memory for nodes
23/03/09	4.189,18	01163/2009	Monitor, keyboard
30/03/09	54,77	0000710-1000043	Extension cords
01/04/09	3.552,03	0000759-1000043	Disk, graphic card
02/04/09	915,00	0000771-1000043	Graphic card
10/04/09	466,65	038-2009	Electric multi-socket with a fuse
28/04/09	9.128,53	0000962-1000043	Monitors for workstations
28/04/09	6.441,60	0000961-1000043	Color printer
30/04/09	2.282,13	0001003-1000043	Monitor for a workstation
ordered	25.756,59	Ponuda: 000397	Components for workstations
ordered	12.106,06	Ponuda: 0000357	Components for workstations

Table 1: List of spendings in the first 6 months.

next project period for the purpose of testing the endurance and performance of Tesla C1060, but within a workstation and not in the cluster configuration.

Another slow down was caused due to adaptation of the dedicated cluster room at the Physics Department. University of Split does not have a data center or a general purpose server room that could host our cluster, except possibly at the new library building finished in the fall 2008. Hence, I sent a letter to the university Senate and the university Rector on June 10, 2008, in which I inquired into the possibility of hosting our cluster at the upcoming library's new server room. To this date I have not received any response from the rectorate or Senate. In the meantime, the Physics Department stepped in and provided us with a small dedicated room for the cluster and one dedicated room for our visualization lecture room. The drawback was that we had to

invest time into the adaptation of the cluster room (special wiring for strong currents, acoustic isolation, cooling, network upgrade).

Finally, we also did not expect that buying computer components can be a major difficulty in Croatia. We had problems finding some specific components, but what has been more surprising is that computer equipment resellers are often delivering incomplete orders or incorrect versions of the components and offering highly unrealistic prices. This required a constant renegotiations and scrutinization of each order. In the end we are quite satisfied with the equipment and budgetary results, except that the project was slowed down.

The final configuration of our cluster is presented in §4, while in Table 1 we list all spending on purchased equipment within the first 6 months (copies of receipts are attached to this report). Ordered equipment is also indicated in the table.

3.2. Personnel support

The key obstacle to this project is limited funds for technical and research support. As a scientist returnee to Croatia I can request a research assistant (PhD student) financed by the Ministry of Science and Education for my research projects in Croatia. I had submitted this request in October 2008., but it was not until the end of March that I received the approval from the Ministry. This position is currently officially opened and will be filled in mid May.

The project was successfully completed to this point thanks to the hard work of dipl.inž. Jurica Teklić and Dubravko Balić. Jurica has volunteered since the beginning of the project, while for Dubravko I have funds that I have been receiving monthly as a donation from the company Insako d.o.o. Additional key help came from dr.sc. Mario Jurić from the Institute for Advanced Study in Princeton, who is our coordinator of cluster development and co-PI on this project.

3.3. Fundraising and expansion of the collaborative network

Considering the lack of funds for project personnel, fundraising is becoming one of the key issues for the future sustainability of the project. Donation from Insako d.o.o. of 6.000,00kn monthly for a year (total of 72.000,00kn) is a major help to the project. In addition, I have applied for a grant from the Unity Through Knowledge fund to support project personnel (including a postdoc) and travel of collaborators to Split, but the project was rejected. The project was submitted in January 2009. and it included an expanded network of collaborators that would focus on four topic themes:

- **Large scale data mining**: handling large data sets (such as the LSST database, http://www.lsst.org)
- **Computational microscope**: simulation of macromolecules and materials on atomic scale in general
- **Particle dynamics**: N-body problems and dynamics of individual "particles" in various environments
- **Weather Forecasting for Split region**: improving weather forecasting models for Split region, which has a specific microclimate

New research collaborators include prof.dr.sc. Branko Grisogono from Andrija Mohorovičić Geophysical Institute, Department of Geophysics, University of Zagreb, doc.dr.sc. Mile Šikić from Faculty of electrical engineering and computing, University of Zagreb, dr.sc. Victor Debattista from Jeremiah Horrocks Institute for Astrophysics and Supercomputing, University of Central Lancashire, UK, and Ivica Vilibić from the Institute of Oceanography and Fisheries, Split, and prof.dr.sc. Ivan Slapničar from the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split.

More interestingly, we have established collaboration with Large Synoptic Survey Telescope (LSST) project. LSST will create the largest scientific database ever assembled and it requires many technological advances in data-intensive science and computing. This is the most ambitious ongoing project in the US astronomy and it has been identified as one of the national scientific priorities for the US. The project partner is also the Department of Astronomy at the University of Washington as the host of our collaboration with the LSST project.

Also, it is quite important that we have established collaborations with industry. We have interest from PLIVA, pharmaceutical industry, while Mirriad Limited from the UK and VEUS d.o.o. from Zagreb were project partners on the UKF grant proposal. Now, with the cluster running, we expect to continue our collaboration with these industry partners and negotiations are underway for future grant applications.

Figure 2: HYBRID in the rack

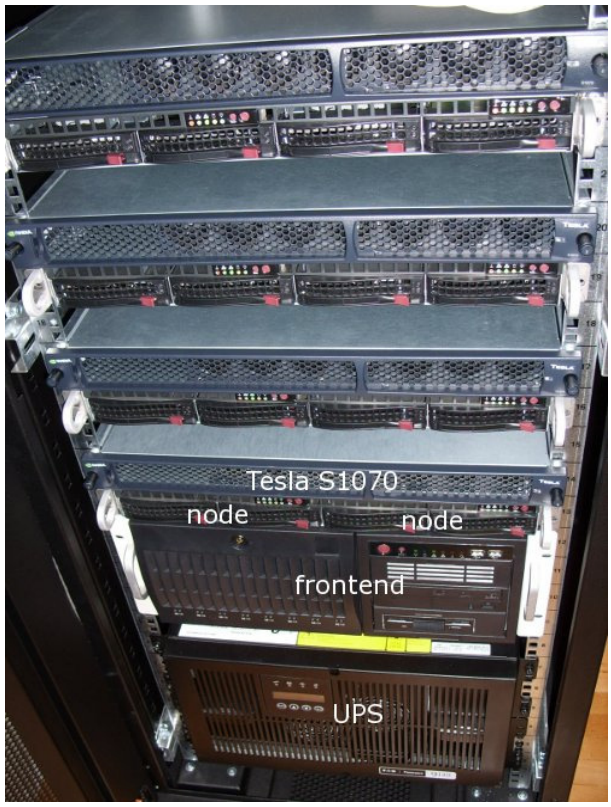


Figure 3: HYBRID with marked components: UPS, frontend, nodes and Tesla S1070.

4. Cluster configuration

The cluster saw its "first light" in the end of March, when it was physically assembled. The very first "burn-out" test of the equipment happened on March 27, 2009. Some fine tuning of the operating system and library requirements are still going on, but the first users are already using it.

One important initial problem that we encountered immediately is that the test code ran more than 5 times slower on our new high-end Tesla GPUs than on GTX 285 card used in the frontend. It turns out that the slow down was caused by using 64 bit integer variables in the code, instead of 32 bit. Once this was fixed in the code, the speed up was similar, as expected. In order to detect such problems, we keep one node with 32 bit version of CentOS and 32 bit drivers.

Photos of the cluster are shown in Figure 2 and 3. The final configuration is as follows:

Frontend:

Hardware: Supermicro motherboard with two Core 2 Quad Xeon processsors, 8 Gb RAM, and 3 Tb of disk storage

Software: Linux CentOS 5.2-x86_64, PBS job scheduler, distributed revision control software, compilers,...

Nodes:

Hardware: Supermicro motherboard with two Core 2 Quad Xeon processsors, 8 Gb RAM, and 500Gb disk connected to a half of Tesla S1070 (that is, 2 GPUs)

Software: Linux CentOS 5.2-x86_64

Network:

1GB ethernet switch

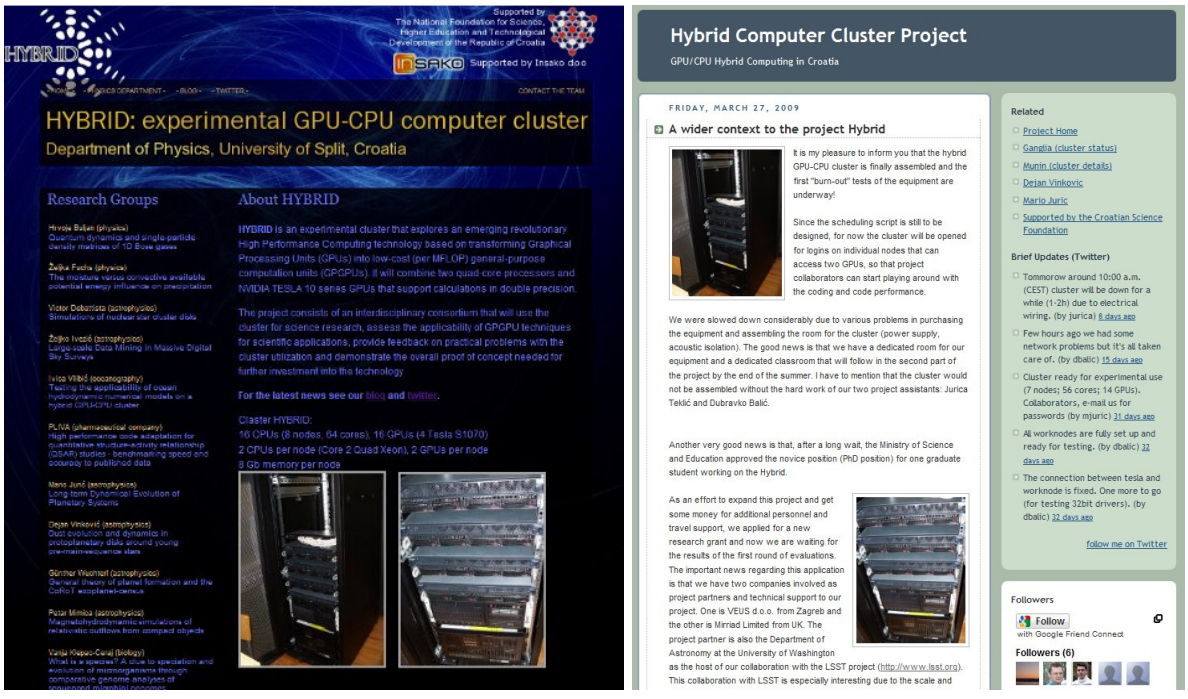


Figure 4: HYBRID web pages (fizika.pmfst.hr/hybrid) and blog (gpuhybrid.blogspot.com).

5. Dissemination efforts

The project has it own web pages (currently at fizika.pmfst.hr/hybrid, but we plan to buy our own domain name), blog (gpuhybrid.blogspot.com) and twitter site (twitter.com/gpuhybrid) (see Figure 4). These pages present the project not only to the general public, but also to the project collaborators and cluster users who can get the latest news about the project status and updates on the cluster status (see Figure 5). We plan to present results coming from the cluster to the public. The news about the cluster being assembled was already featured in the local daily newspaper Slobodna Dalmacija (see Figure 6).

Figure 5: Status of the cluster (hybrid.pmfst.hr/ganglia)

HYBRID SVJETSKI ZNAČAJAN PROJEKT PMF-a

Radi se superkompjuter za složene znanstvene izračune

Na Odjelu za fiziku splitskog PMF-a upravo je u tijeku eksperiment izrade jedinstvenog kompjutera u Hrvatskoj. Koristeći najnovije trendove u svijetu superračunala, ovim će se kompjuterom testirati granice današnje tehnologije u omjeru cijene i performanse (broja operacija u sekundi).

Kompjuter, pod nazivom Hybrid, stane u omanji ormar, ali računalne je snage kakvu su prije manje od 10 godina imali najveći svjetski superkompjuteri, koji su popunjavali veliku dvoranu i trošili 3 megavata struje i još 3 megavata za hlađenje.

"Tajna" tolikog skoka u "sažimanju" u Hybridu nisu samo brži procesori, nego i korištenje procesora koji originalno uopće nisu napravljeni u tu svrhu - grafičkih procesora. Konkretno, riječ je o grafičkim karticama Nvidia, koje će umjesto za igranje neke kompjuterske igrice sada biti iskorištene za znanstvene račune poput stvaranja i evolucije planeta, izgleda proteina, sekvencioniranja DNA, istraživanja ponašanja atoma, itd.

Partneri iz čitavog svijeta

Zanimljivo je da su projekt pokrenuli astrofizičari doc. dr. sc. Dejan Vinković s Odjela za fiziku, povratnik iz SAD-a gdje je i doktorirao, te dr. sc. Mario Jurić s prestižnog Instituta za napredna istraživanja u Princetonu. Projekt sa 100 tisuća eura financira Nacionalna zaklada za znanost, što uključuje i opremanje popratne učionice za znanstvenu vizualizaciju. Uz pomoć inženjera Jurice Teklića i Dubravka Balića projekt je upravo u fazi prvog testiranja instalirane opreme.

„Ovo je zoran primjer kako

Radi se o zornom primjeru kako znanstvene discipline poput astrofizike i astronomije mogu utjecati na razvoj novih tehnologija

Doc. dr. sc. Dejan Vinković i ing. Jurica Teklić ispred eksperimentalnog računala Hybrid u početku izgradnje.

znanstvene discipline poput astrofizike i astronomije mogu utjecati na razvoj novih tehnologija i generirati interdisciplinarna istraživanja. Stoga ne iznenađuje što je projekt privukao partnere ne samo iz Hrvatske, nego iz čitavog svijeta, a suradnja je uspostavljena i s nekoliko privatnih tvrtki koje zanima komercijalizacija znanja koja će proizaći iz projekta", ističe doc. dr. sc. Dejan Vinković s Odjela za fiziku PMF-a. *(M.P.)*

Figure 6: Article in Slobodna Dalmacija, April 03, 2009.