CONTRIBUTIONS to SCIENCE
OPEN ACCESS

CONTRIB SCI 12(1):5-11 (2016)
doi:10.2436/20.7010.01.238

FEATURE ARTICLE

Institut d'Estudis Catalans, Barcelona, Catalonia
www.cat-science.cat

RESEARCH CENTRES IN CATALONIA

Barcelona Supercomputing Center: Science accelerator and producer of innovation

Josep M. Martorell

Associate director, Barcelona Supercomputing Center, UPC-Barcelona Tech, Barcelona, Catalonia

Summary. Supercomputing has become an accelerator; its use is now essential in almost all scientific disciplines. Established a little more than ten years ago, the Barcelona Supercomputing Center (BSC) provides computing services for the European scientific community, carries out in-depth research in different fields, collaborates with multiple research centers and develops technologies for different sectors of society. Based on its capabilities and the skills of the responsible team, the BSC is planning to enlarge both its facilities and its penetration into the scientific and industrial communities as one of the most powerful tools to develop complex research. [Contrib Sci 12(1):5-11 (2016)]

Correspondence:
Josep Maria Martorell
martorell@bsc.es

Computation transforming the scientific process

The fact that scientific production is increasing exponentially because of many different causes is one which no one doubts. Bibliographic data suggest that the publication of journal articles verified by a peer review system grows at a rate of approximately 6–9% annually, which means that the rate almost doubles every nine years. The growing number of researchers and research centers, the pressure to publish and the ease of access to prior knowledge because of new technologies, partly explain this phenomenon.

But it is also true that it is changing how we do science. In October 2013, the influential *Financial Times* published an interesting article about how computing is transforming every aspect of the scientific process, also emphasized by the recent nominations of Peter W. Higgs and François Englert—Nobel laureates in Physics—and Martin Karplus, Michael Levitt and Arieh Warshel—Nobel laureates in Chemistry. Curiously, it is worth noting that the article was illustrated with a photograph of the MareNostrum supercomputer at the Barcelona Supercomputing Center.

Indeed, the 2013 Nobel Prize meant, probably unintentionally, an important recognition of the role of supercomputing and High Performance Computing as an accelerator of scientific progress. For decades, computers have increasingly

Keywords: supercomputing · Big Data · modellization

and subtly become a fundamental part of universities and research centers, changing the way knowledge is obtained and today their role is essential in almost all of our scientific disciplines.

Higgs' Nobel prize came half a century after he theorized the existence of the particle that bears his name. On one occasion, he explained how his theory took shape while walking in the Cairngorms, in Scotland, but as is well-known, it would not be proven until 2013, after several research sessions at the European Laboratory for Particle Physics at CERN. Specifically, experiments were done using the ATLAS and CMS particle detectors, which, when operating at full capacity, produce tens of millions of collisions per second; it was estimated that only one million of these collisions would contain useful information to be studied.

Interactions in the ATLAS and CMS detectors, such as those occurring in the other facilities at CERN, create a huge stream of data that is filtered as it flows, in an enormous massive computing task, to tell the detector which to register and which to ignore. Despite this filter, the volume of data saved is still huge—CERN produces about 30 petabytes of data each year— and, to analyze it all, the CERN computer center is used in addition to a network of distributed repositories and supercomputers over 42 countries.

There were also detectors, in this case LIGO's interferometers, which allowed gravitational waves to be located on September 14th, 2015, a century after Einstein's prediction and, according to some experts, the first day of a new era in astronomy.

Theory, a research team comprised of over a thousand researchers, access to the latest technology and computer resources from various centers and universities—also the MareNostrum—made this new discovery possible.

As regards the Nobel to Martin Karplus, Michael Levitt and Arieh Warshel, there are lots of people who, without failing to recognize the worthiness of these first-class researchers, regarded it as the Nobel to computational chemistry. This is not surprising and has been driven by the Swedish Academy itself, whose statement argued that the prize was awarded "for the development of multi-scale models of complex chemical systems" and stressed that it "laid the foundation for the powerful programs they used to understand and predict chemical processes." Today, Computational Chemistry is standard use for tasks such as nanotechnology and drugs design, where, thanks to *in silico* experiments, the behavior of target proteins when interacting with different molecules can be predicted very accurately, thereby saving time, money and materials.

Interdisciplinarity

Computational chemistry and computer science are different things—like computer science and physics or computer science and other scientific disciplines. A software engineer or computer scientist develops algorithms, software, hardware and visualization systems consistent with the purpose for which they are to be used. Computational chemists and physicists, however, work in conjunction with laboratories and theoretical specialties to apply computer tools to research and validate hypotheses in their respective fields of study.

The same occurs in many other areas of scientific research. Genomics, for example, is a clear point of convergence between biologists and computer scientists. A single strand of DNA, once sequenced and clean from any noise that may have occurred during sequencing, consists of 120 billion letters (C, T, A and G) spread over parts of a hundred letters. A jigsaw puzzle that has to be assembled, analyzed and compared, with tens and hundreds of similar chains that need to be unraveled in order to know what variables may be linked to a certain genetic trait or disease being studied. And DNA is just one part of this puzzle because to fully understand what happens in living organisms, epigenetic information is becoming increasingly important, as it modulates the expression of genes without changing the sequence of DNA.

Bioinformatics for genomic analysis, with the avalanche of data that entails, is considered one of the most disruptive areas in the life sciences because of its implications, among many other reasons, in what will soon be personalized or precision medicine. At present, in Barcelona—physically located at the Barcelona Supercomputing Center—we have a copy of the European Genome-Phenome Archive, containing the Genome and Phenome data of more than 100,000 people who have authorized their use for research. Analyzing and connecting these data with each other, and all those that will be produced exponentially over the next few years, is a daunting task for physicians and biologists, and will consist of the production, storage, data analysis and creation of software and hardware tools that are able to carry out these tasks accurately, efficiently and safely. Supercomputers, tools that were essential to the sequencing of the human genome, are now also key instruments in biomedical research; for example, they are used in research to better understand the complexity and the development of cancer, identify new treatments and find patterns in large and complex data sets.

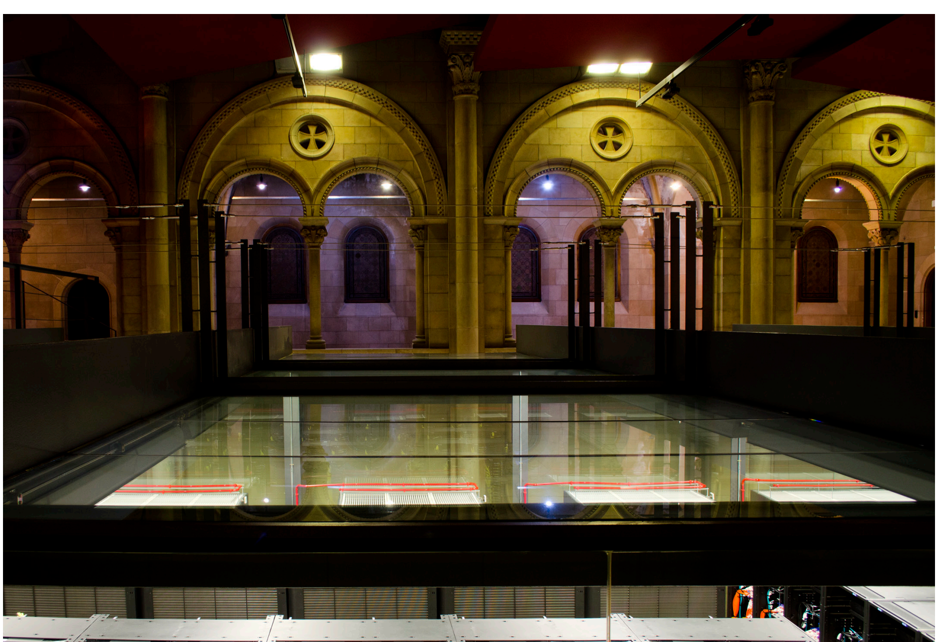
In a very different field of research, but also of great social interest, we find research on climate change. Studies on the weather and climate are a good example of the large amount of data needed to make accurate predictions. It is estimated that, in 2030, the mathematical models used to study climate change will require 350 petabytes of storage. These models have become more and more complex since the 1960s, adding more variables to the factors that must be considered if predictions are to be refined. The World Meteorological Organization is promoting a project, in which the Barcelona Supercomputing Center will participate, where thousands of years of simulations produced by 30 leading centers worldwide will be analyzed, comparing the data the different models produce and contrasting them with historical data and observations from satellites, ocean buoys and weather stations, among other devices, comparisons which will produce much more data. Whenever a model simulates the changes that will occur globally in a single decade, it generates between 135 terabytes and 5 petabytes of data. Running a ten-year simulation at an average resolution needs more than 150,000 hours of processing power. The project promoted by the World Meteorological Organization anticipates simulating approximately 300,000 years using different models and at different resolutions. These investigations could not be done without a network of High Performance Computing systems with distributed repositories throughout the entire globe that store the data generated by different centers; supercomputers, also distributed among different countries—including, again, the MareNostrum—which run the simulations; and extensive teams of experts in computer science and earth sciences.

The Barcelona Supercomputing Center as a service to universities and research

No one expects the disappearance of the old methods of short-term research, but researchers from different areas of science are expected to require increasingly more access to High Performance Computing to expand and accelerate their studies. One of the key missions of the Barcelona Supercomputing Center is providing that computing service to the scientific community.

The "MareNostrum". A supercomputer is a machine that is capable of performing many operations per second, thousands or millions of times superior to the operations a personal computer can make. Supercomputers are built with the same components as personal computers, but in larger quantities. All these components are connected together with high-speed networks and these networks, in addition to smart programming, result in them all functioning together as a single machine.

The MareNostrum stands out amongst our High Perfor-



Rubén Duro

Fig. 1. Upper view of the Mare Nostrum supercomputer located inside the deconsacrated Torre Girona's chapel (buildt in 1940).

mance Computing infrastructures, it is our most powerful machine and also the symbol of the center. The MareNostrum is currently in its third version and we are preparing for the arrival of the fourth; it is the most powerful supercomputer in Spain and one of the most powerful in Europe. With the latest update (made between 2012 and 2013), it has a capacity of 1.1 petaflop/s, or in other words it can carry out 1100 trillion operations per second.

MareNostrum is a supercomputer manufactured by IBM with iDataPlex architecture, which facilitates grouping all the components needed to reach the required capacity into very little space (120 m^2). It has 48,896 processors distributed among 6112 Intel SandyBridge 2.6 GHz chips and eight cores each, and 3056 nodes and 84 MIC chips. These components, chosen to configure a general purpose, not specialized machine, allow it to be useful for a variety of scientific disciplines.

Mare Nostrum's main memory is 100.8 terabytes and its elements communicate with each other via the high-speed network Mellanox Infiniband FDR10 and Gigabit Ethernet. It has a high-performance, 3 Petabyte file system and is connected to the Big Data BSC-CNS infrastructures. The supercomputer runs on a Linux operating system—SuSe Distribution—and the entire system is connected to universities and major European research centers through optical fiber, facilitating the transmission of data to researchers who need to use our infrastructure.

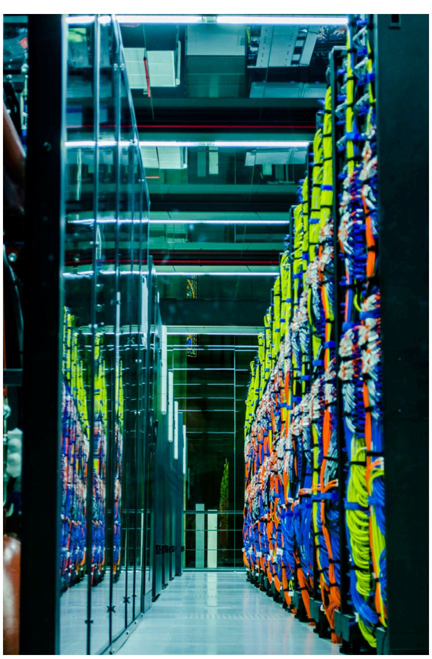


Fig. 2. Inner view of the Mare Nostrum supercomputer.

The MinoTauro. The MinoTauro is the second most powerful supercomputer in the BSC-CNS, with a peak performance of 343.74 teraflop/s (343 trillion operations per second). In 2011, a previous version was considered the most energy efficient supercomputer in Europe, according to the Green 500 ranking. The MinoTauro is manufactured by Bull and one of its features is the addition of NVIDIA GPU (graphics processing unit) cards to its architecture.

Big Data Infrastructures. Recently, due to a growing need for large servers and data centers, the BSC-CNS has adopted Big Data infrastructure with a total capacity of 24.6 petabytes. It is scalable architecture equipment, distributed on disks and tapes to store scientific data in the short, medium and long terms. The BSC-CNS has three types of storage systems: high-performance, an active file system and a tape robot.

The high-performance system has a 14 petabyte capacity, consisting of more than 2400 hard drives and about 18 servers and is used by all supercomputing machines directly for high-speed reading and writing of scientific data.

The file system, which has a capacity of 4.6 petabytes, is made up of about 3200 hard drives and about 20 servers, and is used to store scientific data generated from supercomputers until their subsequent study.

The tape robot consists of 7700, 800GB, uncompressed LTO4 magnetic tapes and is used for making backup copies of all the data generated in the center. The tapes also archive the data for which no short-term access is needed.

More than 3000 projects

Throughout the Barcelona Supercomputing Center's first 11 years, over 3000 scientific projects from a variety of disciplines have used the BSC's infrastructure to accelerate the results of their research. The center itself has coordinated and facilitated access for Spanish centers and universities' research teams to the Spanish Supercomputing Network (RES). Currently, access to 24% of the MareNostrum is done through this network, while 70% is accessed through the European network PRACE (Partnership for Advanced Computing in Europe), a center in which the BSC plays a primary role. The remaining 6% is for use by the BSC researchers. Both in the case of RES as well as PRACE, the use of the Barcelona Supercomputing Center services is coordinated through external access committees, consisting of researchers from different disciplines, which evaluate, prioritize and authorize the use of machinery, according to the importance and the needs of the projects presented by the research teams that require High Performance Computing services.

Fourth version of MareNostrum on the way

Because it is a publicly owned center, one of the BSC's maxims has always been to get the best possible performance of its available infrastructures and it is satisfying to note that our machines work 24 hours a day, 365 days a year and that more than 90 percent of their capacity is always occupied.

The continuous improvement of facilities is another of the center's maxims. The center currently works on the purchase and implementation of a new version of the MareNostrum planned to be operational at the end the first half of 2017. This new version is expected to have nearly 13.7 petaflop/s of computing power, tripling the current main memory and exceeding the 14 petabytes disk space.

Research of excellence

The BSC's ambition does not stop at being a service center of reference. Despite its short existence, it has already positioned itself as a research center of excellence, awarded the first Severo Ochoa Seal of Excellence and has become the center that receives the highest amount of international public funds for research per researcher.

Starting as a center that employed sixty employees, today it brings together more than 450. The 50% of its researchers have trained in computer science and/or telecommunications and the other 50% in disciplines as varied as mathematics, physics, chemistry, biology, geophysics, etc., in full accordance with the need to form interdisciplinary teams able to take full advantage of information technology as an accelerator of knowledge.

Because of this confluence of services and research, computer scientists and researchers from other disciplines, BSC has been able to position itself as a center of great uniqueness and interest. Our research is divided into four major areas: Computer Science—the department that has the highest number of researchers—Life Sciences, Earth Sciences and Computer Applications for Science and Engineering, resulting in a privileged position as researchers and creators of software and hardware and at the same time as users of this software and hardware in the field of advanced research.

This uniqueness translates into participation as leaders or as partners in research projects from the most diverse disciplines. We currently have one hundred ongoing projects and a presence in major international flagship projects such as the Human Brain Project, providing programming models and methods that help simulate the human brain at different levels, and the PanCancer Analysis of Whole Genomes, an international collaboration to identify common patterns of mutation in more than 2800 cancer whole genomes from the International Center Genome Consortium.

Included among the most significant projects that we develop are the following: The European Research Council Grants projects of the center's director, Mateo Valero, on micro-architectures, systems runtimes, compilers and programming languages (RoMoL); of David Carrera, on holistic integration of emerging supercomputing technologies (HiEST); and of Xevi Roca, to create new simulation methods intended for the aeronautical sector. They aim to expand the potential and field of applying High Performance Computing. It is also worth noting that we lead projects that support the European Commission in the alignment of the different sectors involved in Big Data research and High Performance Computing. An example is RETHINK big, a roadmap of opportunities to improve the EU's position in Big Data research.

Hub to boost research

Because of our characteristics—and without a doubt, our researchers' drive—, we have become collaborators in the research conducted in most centers and research units within our area of influence as a supercomputing center, in just over 10 years. From the beginning, we have participated in more than 1700 joint projects in collaboration with more than 800 universities, research centers and companies from 50 different countries. Needless to say, among these centers and universities are those closest to us, including most of those within the CERCA system and the majority of the Excelencia Severo Ochoa centers.

We also participate in numerous international initiatives and consortia that aim to support High Performance Computing, such as the Joint Laboratory initiative on Extreme Scale Computing (JLESC), an international association of six supercomputing centers to make the bridge between Petascale and Extreme computing, or the OpenMP, OpenPower or OpenFog consortia where trends in the fields of parallel programming, processor architecture and distributed computing, among others, are discussed.

Technology transfer

The BSC continues to go further than research by also acting as an accelerator for science. Our defining statement makes it clear that "the mission of BSC-CNS is to research, develop and manage information technology in order to facilitate scientific progress", emphasizing our third fundamental goal: to be drivers of innovation, at the service of science, business and social progress; in other words, to transfer technology to society.

We can now, after our first ten years of life, fully assess the work done on technology transfer and promote a real leap in this field which is, for us, essential.

Our collaboration with the business community and the government is structured through three main forms of cooperation: the development of joint long-term projects, timely collaboration through our service platform and the creation of spin-offs from our own technological capabilities.

Joint projects: IT, energy, banking

We have extensive experience in this first area, where the creation of joint centers with large IT (information technology) companies stands out. At present, we have centers with IBM, Intel, Microsoft, NVIDIA and Lenovo, with whom we conduct research into areas such as cognitive computing, computer systems behavior and prediction analysis, cloud computing, CPUs and energy efficient programming, among others.

Some of these collaborations are almost as old as the BSC itself—as is the case of IBM—and have adapted to the strategic needs of the company and the major trends in the field of High Performance Computing. Samsung and Lenovo are more recent collaborations, as is the influence and the increasingly important role played by large Asian companies in the world of supercomputing and IT in general. Note that the today's most powerful supercomputer is one built in China with mainly Chinese technology.

Our joint projects with companies are not limited only to the world of IT. We also have stable collaborations with other large companies. The energy sector was the first with which we formed a lasting and diverse activity. The best known success story is our collaboration with Repsol, with which we have developed and are constantly improving software that explores the presence of hydrocarbons in geographically complex areas. Using this software has greatly increased the company's success rate in finding hydrocarbon reserves in areas such as the Gulf of Mexico, where oil is located kilometers beneath seawater and under kilometers of layers of solidified salt. With this technology, the company has significantly increased its competitiveness when searching for new reserves which producing countries can exploit.

Another relevant research study in the field of energy, wind in this case, is our collaboration with Iberdrola. With this company we have developed tools to determine how to extract maximum resources from land designated to become a wind farm. With software developed jointly between Iberdrola and BSC we can determine the best way to distribute generators in a particular area in order to make the most of the available wind, and soon we will be able to include weather forecasting software that will predict how much energy a particular farm will produce over the following 24 or 48 hours. The industry believes that being able to predict how much wind energy will be generated in the short term will be a qualitative leap for companies' power distribution and performance.

And still in the energy field, we collaborate with the ITER project by conducting simulations of the behavior of fusion reactors, and also work on relatively new energies, such as that produced by biofuels, studying how to turn the energy generated from biomass combustion into an energy that has a homogeneous composition and can be distributable through networks.

As a supercomputer center, we want to extend our collaboration with companies to include the widest possible range of sectors. The jump from industry to services is increasingly plausible, especially with new tools and possibilities of cognitive computing applied to Big Data, an area in which, by our very nature, we have much experience. Along this line, in 2016 we have established a collaboration project with CaixaBank to explore cognitive computing in the world of finance, and we also work on smaller projects with small and medium-sized companies and governments that are increasingly aware of the value that cognitive computing, neural networks and machine learning can add to the administration of large volumes of data, and the value that High Performance Computing machines can add to the management of cities, the environment and public spaces.

Services platforms

The second way in whixh we transfer the technologies developed at the BSC to society is through our services platform: mainly using the Internet, we make our high-value applications developed at BSC available to users.

The Department of Life Sciences offers different tools for genomic analysis and medication explorations. We have one of the fastest and most complete and reliable software programs that exist to detect genetic mutations (Smufin) and the simulation method to predict interactions between proteins and ligands, PELE. This latter tool was created to be used in drug design and was considered the best of its kind in the last docking models comparative assessment funded by the National Institute of Medical Sciences of the United States and organized by the Community Structure-Activity Resource (CSAR) from the University of Michigan.

We also publish our daily predictions on air pollution across Europe and the transport of desert sand through the atmosphere. This last prediction system convinced the World Meteorological Organization to create the first regional center to address this matter in our facilities in Barcelona in collaboration with the Spanish State Meteorological Agency (AEMET).

In addition, our Computer Sciences experts offer different programming models and free software to be used in parallel computing, under the concept of "software as a service".

The creation of new businesses

In our first ten years of life we have had approved, or are currently under consideration for, about thirty patents. And after a hard journey through the bureaucratic mazes we have created our first spin-off, which will not be the last, called Nostrum BioDiscovery (BD). It is a company created to accelerate the development of drugs and molecules with different biotechnological applications based on supercomputing which emerged from a collaboration between IRB Barcelona, UB, ICREA and the Fundación Botín.

The two purposes of Nostrum BD are to collaborate with companies dedicated to the development of medication and molecules of biotechnological interest and to facilitate the launch of new products. It has state-of-the-art bioinformatics technologies related to protein modeling, the study of protein interactions with molecules of a therapeutic nature and enzyme engineering to develop biotechnology products, all through supercomputing. These technologies, when combined, increase the accuracy of pre-studies and facilitate a more rapid release of pharmaceuticals and biotechnology products.

In favorable cases, it is estimated that these advantages may advance the early stages of drug development (the discovery phase, prior to laboratory testing) by up to two years, and can reduce costs by between 15 and 20% in this first phase, with an average estimated saving of 40 million euros. These technologies have also proven useful for the discovery of medications in the context of precision medicine as they facilitate research conducted into the most appropriate medication for each therapeutic target, according to their genetic characteristics.

The great challenge of personalized medicine

It is precisely in the field of precision medicine where we are currently challenging ourselves to collaborate in the promotion and development of public and cutting-edge precision medicine in our country. Some countries have already launched very advanced pilot tests to regularly incorporate the possibilities that personalized medicine offers into the public network. Hospitals, biomedical research, sequencing, and data centers are the four key pieces of a cycle where the patients who enter the health system with certain complex diseases, are treated with a personalized investigation into their case, research that can develop more effective treatments for the patient and provide more accumulated knowledge about the disease itself.

The presence of precision medicine as a regular practice in the health system is almost a reality and Barcelona is in an excellent position to be a pioneer city in this field. The quality of health centers, the numerous and excellent research carried out in biomedicine and the presence in the city of two public centers able to sequence on a large scale (CNAG-CRG) and to process and analyze genomic data (BSC) offer a mix of knowledge, talent, infrastructure and technology that is capable of addressing this challenge.

Competing interests. None declared.