Making Supercomputing Available to All Cuban Researchers

Hector Cruz-Enriquez
Universidad Central "Marta Abreu" de Las Villas

Scientists in Cuba must get up to date with the latest trends and best practices in scientific computing. As a consequence of the socioeconomic and geopolitical situation in and around Cuba, a lack of financial means restricts Cuban researchers' access to the supercomputing infrastructures, techniques, and tools that are urgently needed to support the increasing use of computational tools and efficient data processing in different areas, especially those related to public health.

In the following article, I recount my story as a young researcher and my efforts to make nationwide high-performance computing (HPC) services available to all Cuban scientists. This effort was motivated by my research experiences 15 years earlier. I began working with complex images acquired with a low signal-to-noise ratio (SNR), and found myself faced with the problem of phase residues, a major drawback for the bidimensional phase unwrapping process that is usually performed when processing phase images in several health-related applications. My experiments applied de-noising algorithms to complex images combining nonlinear filters based on order statistics methods in the wavelet domain, which demanded high computational power. The unavailability of HPC facilities imposed severe challenges for the research process required to achieve my proposed objectives.

DE-NOISING COMPLEX IMAGES

I graduated in 1999 from the School of Electrical Engineering of the Universidad Central "Marta Abreu" of Las Villas (UCLV) in Cuba. Since 2003, I have been linked to the Center for Studies on Electronics and Information Technologies (CEETI), where I did my PhD studies on de-noising complex images, meanwhile I served as head of the Telecommunications Department in the Electrical Engineering School. Upon receiving my PhD in 2007, I was nominated to and accepted the role of director of the UCLV Central ICT Department.

  Copublished by the IEEE CS and the AIP
1521-9615/18/$33 ©2018 IEEE

My research involved formulating and testing a set of algorithms using simulated, phantom, and real images, which showed improvements in the SNR (see Figure 1) and image quality (see Figure 2) when using combined methods (nonlinear ordering statistics and wavelet domain algorithms such as NLSUFM1AD, and NLSUFD1AD).[1–5]

To work with raw real data, the phantom images used during the research were obtained with an experimental tomograph installed at the Centro de Biofísica Médica (Medical Biophysical Center) in Santiago de Cuba. Those images had to be processed in Santa Clara, 650 km away, on a small cluster of 12 power desktop computers.

The usage of nonlinear ordering statistics methods implies an extra computational burden inherent in calculating angular distances. The simulations started to delay more than a week with the available hardware, making experiments nonviable. Additionally, the electrical supply suffered shortages and the lack of electrical backups and limited air conditioning systems in conjunction with the low-quality hardware available imposed extreme conditions on my research.

Finding a solution in Cuba to shorten the time of the experiments and progress more rapidly in my research was not possible, because there were no available super computation infrastructures in the country. Many experiments had to be performed remotely in HPC clusters in Canada, Spain, and Belgium through narrow bandwidth links. In 2007, UCLV had an Internet connection of 1 Mbps, shared across all university services. This link was congested during the day, so early morning (from 1 AM to 6 AM) became the ideal time for experimentation.

Despite the passing of time and technological advances, the situation continues to repeat itself. Investigators at important research centers in Cuba, such as the Centro de Neurociencias de Cuba (Cuban Neuroscience Center, or CNEURO), have a 3-T magnetic resonance tomograph that allows them to obtain functional magnetic resonance imaging (fMRI) to follow up on drug treatments of cancer-related brain lesions, among other applications. The lack of sufficient supercomputing capabilities to process this data within tolerable times means that the researchers must send high-resolution images to Canada to be processed—and deal with the storage limitations and Internet congestion that still confront Cuba.

Since 2007, I have led international collaboration projects related to ICT infrastructure. From 2013 to date, I have coordinated the Cuban side of the ICT-NETWORK program between Flemish universities in Belgium and five Cuban universities. Within this collaboration's framework, I have explored opportunities to strengthen computational research capacities for Cuban scientists.

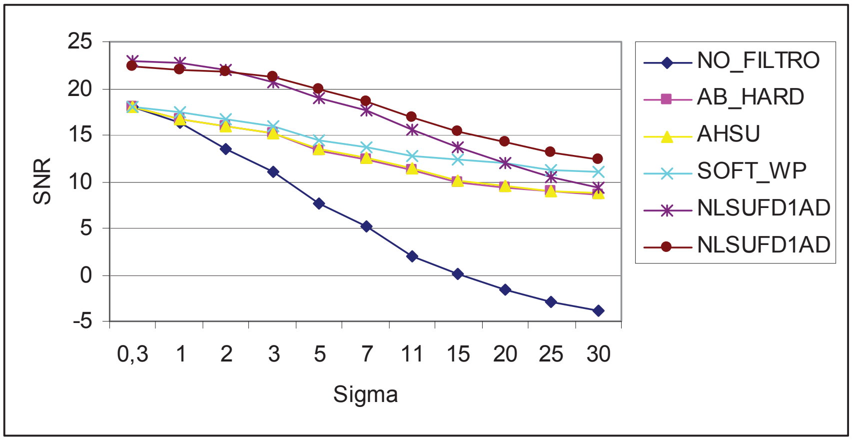

Figure 1. Algorithms with better performance SNR for gradual increments of noise (σ) over the phase images shown in Figure 2.[5]

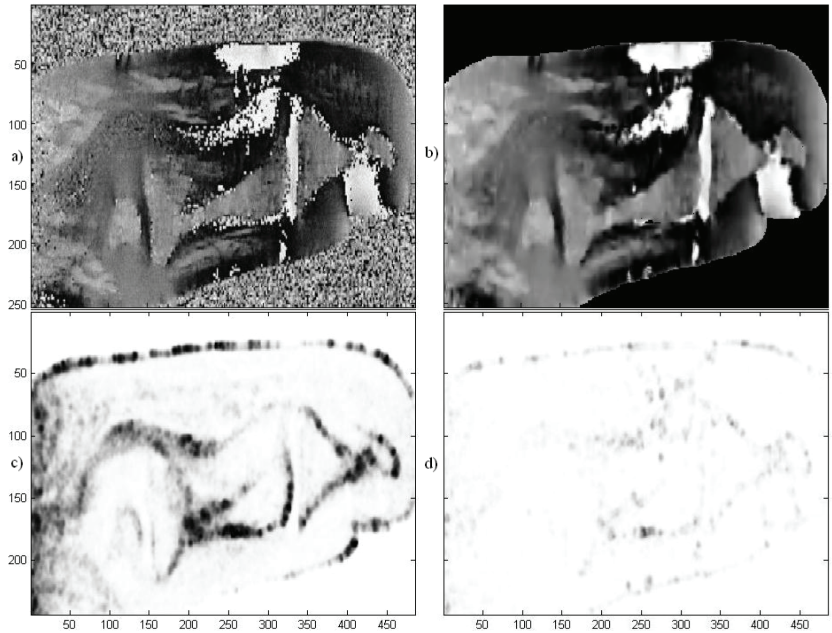

Figure 2. Sagittal cross section of a human finger: (a) noisy (wrapped) phase image, (b) filtered phase image with NLSUFD1AD, (c) noisy structural similarity map, and (d) filtered structural similarity map.[1,5]

BIRTH OF THE HPC-CUBA PROJECT

Researchers rely more and more on scientific computing. In their efforts to transform research data into knowledge, Cuban researchers are restrained by their limited knowledge of and access to technologies and tools for efficient data processing. These limitations can be alleviated by providing expertise and access to scientific computing infrastructures for researchers from a wide range of institutions, which would in turn provide opportunities for synergy and application license sharing. It was discovered that difficulties arise mainly in the following aspects: access, autonomous management of research applications, data science knowledge, and effective coding for parallel/distributed processing.

The idea of making supercomputing available to all Cuban scientists was born in 2015 when I met Dieter Roefs in Belgium (Datacenter Coordinator of Ghent University), starting with two small HPC clusters installed in Santa Clara (UCLV) and Santiago de Cuba (UO). Thanks to big donations from Ghent University in Belgium (besides VLIR-UOS funded purchases), various supercomputer clusters were shipped to Cuba. Two IBM clusters were installed in Havana (UCI) and Santa Clara (UCLV), with 84 nodes, 600 cores, and 2 Tbytes of RAM each. Two GPU Nvidia K80 cards were also installed in Santa Clara. Similar capacities with Dell technology were installed in Santiago de Cuba (UO). This signaled the start of the HPC-Cuba project (see Figure 3).[6,7]

Although this infrastructure is available at some universities, it is not accessible to all potential users from surrounding research groups and scientific institutes. Nationwide access will remain problematic as long as there is no established national center for scientific computing to provide a broad user base of researchers with a platform and portal to access to HPC and big data services. The HPC-Cuba project has a geographic distribution of HPC infrastructure at three points, the strongest Cuban universities in terms of ICT infrastructure: UCI in Havana (west), UCLV in Santa Clara (central), and UO in Santiago de Cuba (east). Several free software-management platforms have been installed for services orchestration (foreman), configuration management (Puppet/Ansible), and management of computing resources (slurm) with CentOS, as well as a national Lightweight Directory Access Protocol (LDAP) system for access to distributed infrastructure.

Figure 3. Supercomputing cluster at UCLV in Santa Clara (left) and UO in Santiago de Cuba (right).

Technology and knowledge transfer are attracting industry (for example, BioCubaFarma) to merge their computational infrastructures with the proposed network. Industry and its associated research centers play a key role in the sustainability of the project. We have identified a group of enterprises and associated research centers producing cutting-edge technology products (CNEURO, FINLAY, CIM, CIGB, all of them belonging to BioCubaFarma). ETI-BioCubaFarma, a hub that provides supercomputing services to the most important research centers in Havana (Biocubafarma consortium), will acquire a supercomputing infrastructure and share it through the HPC-Cuba project. Other centers and research laboratories in the country, especially those associated with in silico experimental modeling, are already benefiting from this initiative, for example, Centro de Bioactivos Químicos (Chemical Bioactive Center) at UCLV and Laboratorio de Química Comutacional y Teórica (Computational Chemistry Lab at Havana University). So Cuban scientists are now closer to accessing a national center for scientific computing (see Figure 4).

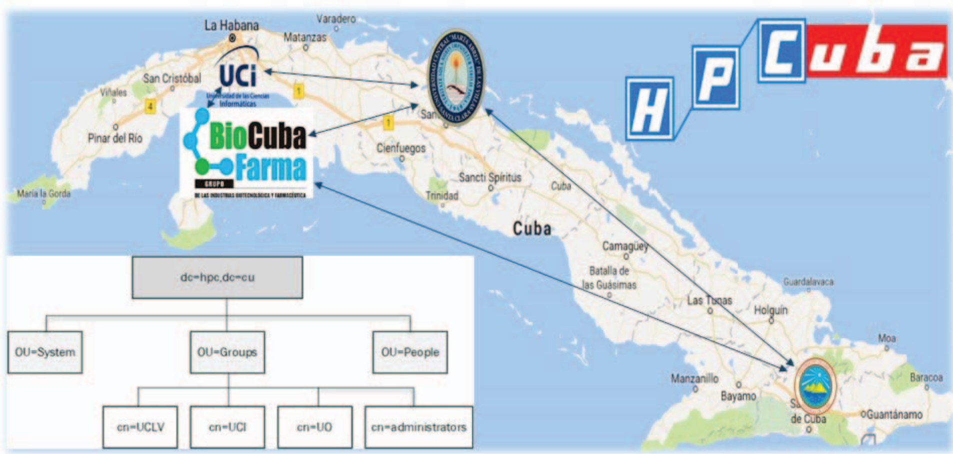

Figure 4. Cuban Center for Academic Supercomputing.

The national strategy includes capacity building to professionalize our HPC staff. To draw the attention of international HPC experts, the project was presented at the 2017 Free and Open Source Software Developers' European Meeting held in Brussels. Enthusiastic international experts from supercomputer centers in Jülich, Barcelona, and Flanders and academics from Spain, Argentina, and Belgium have shown their willingness to participate and visit Cuba to support this project initiative.[8,9]

As scientists are often not IT experts, they face difficulties managing the applications they require; there is a need to adopt solutions to ease deployment and configuration of applications. UCLV HPC staff adopted EasyBuild, which is based on open source project. We have even submitted contributions to the EasyBuild community.[10]

There is no lack of challenges to face, as researchers are often dealing with difficulties when analyzing large amounts of heterogeneous and dynamic data (for example, big data and deep learning). There is a need to provide users with techniques for intelligent data analysis and code optimization for parallel applications. This would allow them to effectively use the available compute capacity—rather than depending only on generic analysis techniques and single-node computations—and thus convert their data into highly valuable knowledge.

CONCLUSIONS

The HPC-Cuba project initiative enables, supports, guides, and trains Cuban researchers to maximize their scientific output based on scientific computing services. In this way, we are influencing the quality and quantity of research conducted at various institutions at the national level, thus contributing to the general sphere of interest—science contributing to national development.

ACKNOWLEDGMENTS

I would like to thank the University of Ghent (Ugent) for its donation of HPC infrastructure, and VLIR-UOS for its support through international collaboration projects; Dieter Rose, datacenter coordinator at Ugent, as one of the founders of the HPC-Cuba project; Prof. Dr. Dirk Rose, KU Leuven, for supporting this project idea; Prof. Dr. Esteban Mocskos, Universidad de Buenos Aires, for opening a door to the regional HPC community; and Kenneth Hoste and Ugent HPC staff, VSC/ Flemish Supercomputer Centre, and Dr. Damian Alvarez, Jülich Supercomputing Centre, for training our HPC staff and all the support offered. Special thanks to the Cuban HPC staff for all of their valuable efforts.

REFERENCES

1. H. Cruz-Enríquez and J.V. Lorenzo-Ginori, "Combined Wavelet and Nonlinear Filtering for MRI Phase Images," *Int'l Conf. Image Analysis and Recognition* (ICIAR 09), 2009, pp. 83–92.
2. H. Cruz-Enríquez and J.V. Lorenzo-Ginori, "Wavelet-based Methods for Improving Signal to Noise Ratio in Phase Images," *Proc. Int'l Conf. Image Analysis and Recognition* (ICIAR 05), 2005, pp. 247–254.
3. J.V. Lorenzo-Ginori and H. Cruz-Enríquez, "De-noising Method in the Wavelet Packets Domain for Phase Images," *Proc. Iberoamerican Congress on Pattern Recognition* (CIARP 05), vol. CIARP 2005, 2005, pp. 593–600.
4. H. Cruz-Enriquez and J.V. Lorenzo Ginori, "Filtrado de Imágenes para Aplicaciones de Resonancia Magnética," *Proc. 4th Congreso Latinoamericano de Ingeniería Biomédica* (CLAIB 07), 2007, pp. 263–266.
5. H. Cruz-Enriquez, J.L. Ginori, and C.L. Dueñas, *Filtrado de Ruido en Imágenes de Fase para Aplicaciones Médicas*, Publicia, 2015.
6. "UCLV HPC," Universidad Central "Marta Abreu" of Las Villas, 2018; www.uclv.hpc.cu.
7. "Portal Wiki," Red Académica de Supercómputo, 2017; wiki.uclv.hpc.cu/doku.php.
8. "The Birth of HPC Cuba," FOSDEM 2017, 2017; archive.fosdem.org/2017/schedule/event/hpc_cuba/.
9. R. Brueckner, "Video: The Birth of HPC Cuba," *Inside HPC*, 2018; insidehpc.com/2018/03/video-birth-hpc-cuba/.
10. "EasyBuild," 2017; twitter.com/easy_build.

ABOUT THE AUTHOR

Hector Cruz-Enriquez is a professor of telecommunications and the director of the Central ICT Department at the Universidad Central "Martha Abreu" Las Villas. He coordinates the Cuban side of the ICT-Network program between Flemish universities in Belgium and five Cuban universities. As part of this collaboration, Cruz-Enriquez implemented one of the biggest datacenters in Cuba. He is one of the founders of the HPC-Cuba project, which aims to make supercomputing available to all Cuban researchers. Contact him at hcruz@uclv.edu.cu.