Quantum ESPRESSO on GPU:
accelerating a complex code combining performance with portability

Ivan Carnimeo

The new features recently implemented in the Quantum ESPRESSO code and presented herein will allow to efficiently simulate a wide range of molecular and material properties on different accelerated hardware architectures, provided by the most advanced HPC centers worldwide.

Abstract

Quantum ESPRESSO **[1-4]** is one of the most popular open-source suite of computer codes for atomistic materials and molecular modelling, based on Density Functional Theory, plane waves, and pseudopotentials.

These kinds of computations, called “from first principles” or “ab initio” are able to predict and give fundamental insights about many properties of materials, molecular systems, micro and nanodevices, biological systems, in many fields, providing a huge amount of data for data driven science applications.

The Quantum ESPRESSO suite is developed using a modular approach: a plurality of independent modules, devoted to specific tasks and features, is used to assemble executables. In the last years a large effort **[3]** has been devoted to port the most important and numerically intensive modules to accelerated architectures based on hybrid CPU-GPGPU hardware, for large scale applications.

The core of Quantum ESPRESSO, a code named `pw.x`, is devoted to the solution of the Kohn-Sham equations. `pw.x` and the modules thereof connected have been fully ported to hybrid architectures, making extensive use of accelerated libraries for Fast Fourier Transform, matrix algebra and eigenvalues computations and efficient schemes for host-device data movement and device memory allocations. Generally excellent performance is achieved, as stated by extensive benchmarks **[3]** and users experience.

Afterwards, the most recent advances in the Quantum ESPRESSO code are nowadays mainly focused on (i) extending the typology of chemico-physical properties that can be computed on accelerated supercomputers, and (ii) improving the portability of the code, in order to effectively exploit the broad range of hardware architectures currently made available by the HPC centers, based on different technologies developed by different vendors.

Regarding the first point (i), work is in progress to fully port ab initio Car-Parrinello molecular dynamics and linear response calculations to accelerated hardware, boosting the simulation of the time evolution and of vibrational properties of molecular and condensed matter systems of technological interest.

Regarding the second point (ii), the main modules of Quantum ESPRESSO have been refactored using a flexible scheme where a unique high-level layer of code manages calls to hardware specific GPU and CPU procedures, allowing acceleration on a broad range of hardware.

These features allow to simulate a wide range of molecular and material properties in an efficient way on a broad range of hardware architectures, ranging from personal computers to the most advanced HPC centers worldwide.

References

[1] P. Giannozzi et al., J. Phys.: Condens. Matter 21, 395502 (2009);
[2] P. Giannozzi et al., J. Phys.: Condens. Matter 29, 465901 (2017);
[3] P. Giannozzi et al., J. Chem. Phys. 152, 154105 (2020);
[4] S. Scandolo et al., Z. Kristallogr. 220, 574-579 (2005).