Response to “86 FR 60021 - Stewardship of Software for Scientific and High-Performance Computing”

Hang Liu
Stevens Institute of Technology
hliu77@stevens.edu

It is the public consensus that general-purpose computing platforms, such as Central Processing Units (CPUs) and Graphics Processing Units (GPUs) are often more inefficient than Field-Programmable Gate Arrays (FPGAs) and Application-Specific Integrated Circuits (ASICs). **(1)** *The primary reason is that CPUs and GPUs are instruction-driven platforms while FPGAs and ASICs are data-driven ones.* That said, FPGAs and ASICs can avoid all the overheads related to instruction processing, such as fetching and decoding. Further, **(2)** FPGAs allow tailored hardware/architecture configurations for the particular requirements of specific applications which could boost the performance remarkably!

Unfortunately, my recent explorations in our field, i.e., High-Performance Computing (HPC), often treat FPGAs and ASICs as a general-purpose computing platform hence using general-purpose computing software infrastructures to program them, such as OpenCL and Vivado. The root cause of this phenomenon is partially attributed to **(i)** the complexity of using hardware description languages (HDL) to program reconfigurable hardware platforms, **(ii)** the lack of reconfigurable hardware-based HPC infrastructures, as well as **(iii)** the missing of encouraging success and well-documented lesions about how to DOE interested applications on FPGA-equipped HPC platforms.

In this RFI response, we believe that **there is a need of supporting a few selected experts who are passionate about promoting the hardware-system-algorithm codesigned platforms for DOE concerned applications.** **(i)** Regarding the hardware part, the expert should have the experience of computer architecture and can acquire the right hardware (FPGAs, servers, and FPGA interconnect fibers), put them into a dependable HPC platform, as well as possess the essential architecture and HDL knowledge. **(ii)** For the algorithm side, the experts should be extremely familiar with some very important DOE solvers, such as SuperLU_DIST, or other DOE concerned systems. We use SuperLU_DIST as an example because we have very close collaborations with Dr. Xiaoye Sherry Li, who is the major creator/developer of SuperLU_DIST, from Lawrence Berkeley National Laboratory. **(iii)** The system side expects the participants to design and develop a software infrastructure that can materialize the desired hardware-algorithm codesigned ideas into the aforementioned HPC platforms. Upon success, this project will offer invaluable lessons and encouragement to the community which will motivate more explorations thereafter.

When it goes to neck-to-neck turnaround time comparison, we anticipate that FPGA-based platforms should be able to achieve similar or even shorter turnaround time with orders of magnitude energy and power-saving when compared to general-purpose GPUs.