

Multi-FPGA Infrastructure at PC2: Architectural Design and Integration of FPGAs in Modern HPC

Authors:

Heinrich Riebler, Tobias Kenter

Field-Programmable Gate Arrays (FPGAs) offer energy efficiency and tailored hardware acceleration for specialized high-performance computing (HPC) workloads. However, integrating FPGAs into production-grade supercomputers historically suffers from fragmented vendor ecosystems, rigid network topologies, and high programming barriers, limiting their widespread adoption among scientific communities.

This work presents the architectural design and operational model of the multi-vendor, multi-FPGA infrastructure at the Paderborn Center for Parallel Computing (PC2). We leverage the dedicated FPGA partition within the Noctua 2 supercomputer, which uniquely deploys hardware from competing industry leaders: AMD/Xilinx Alveo U280 and Intel Stratix 10 cards. Crucially, rather than treating these accelerators as isolated PCIe peripherals, the infrastructure implements an independent, direct FPGA-to-FPGA network. This allows low-latency, high-bandwidth communication across nodes bypassing the host CPU. The software ecosystem bridges the vendor divide by providing standardized high-level synthesis (HLS) frameworks and optimized open-source runtime environments, lowering the entry barrier for domain scientists.

The infrastructure successfully hosts a variety of production workloads, spanning quantum chemistry, fluid dynamics and machine learning. The multi-vendor approach allows users to match specific algorithmic traits - such as high-bandwidth memory (HBM2) requirements - to the hardware architecture best suited for them. Furthermore, the custom FPGA-to-FPGA routing enables spatial parallelization, allowing large-scale simulations to scale tightly across multiple reconfigurable nodes.