Ph.D. in Information Technology
Thesis Defense

December 11th, 2026
at 2:00 pm
Room 1B – building 20A

Gianmarco Accordi – XXXVIII Cycle

Titolo: OPTIMIZING EXTREME-SCALE VIRTUAL SCREENING ON HETEROGENEOUS AND EMERGING ARCHITECTURES

Supervisor: Prof. Palermo Gianluca

Abstract:

Extreme-scale virtual screening requires evaluating billions of ligand-protein interactions under strict time and resource constraints. At this scale, campaign-level utility depends not only on the accuracy of individual molecular docking models, but on the volume of chemical space evaluated within a fixed supercomputing allocation. This makes virtual screening a HPC problem in which efficiency must be sustained across increasingly heterogeneous systems. This thesis studies molecular docking from a campaign-scale PP perspective. In this context, PP acts as a structural design principle: the underlying application implementation must coordinate directly with workload scheduling and platform-level orchestrators to sustain high throughput across target execution environments. The first part of the thesis focuses on the architectural dimension of the problem, investigating how PP is affected by the execution characteristics of CPU and GPU. On CPU architectures, we investigate SIMD vectorization, showing that compiler auto-vectorization is highly sensitive to toolchain math libraries and memory access latency. Consequently, portable CPU performance depends strongly on compiler and vector-math support, while explicit SIMD provides a more controlled alternative when compiler-driven vectorization falls short. On GPU accelerators, execution must transition from latency-oriented single-molecule execution to throughput-oriented batched configurations. Structural and chemical irregularities across the ligand library induce divergent resource demands, necessitating out-of-kernel grouping strategies to align molecular inputs with device occupancy limits. Beyond core execution, we address the data and workflow bottlenecks of virtual screening. This includes designing a compact molecular compression format that preserves random-access capability, alongside scoring function approximations that trade precision for campaign throughput. These layers of optimizations maximize screening capacity while

satisfying full-path execution constraints. The thesis closes by evaluating a reduced quantum formulation of molecular docking in a hybrid classical-quantum setting. This investigation is intentionally exploratory; its purpose is to assess the feasibility and current limitations of available quantum backends for reduced docking formulations, rather than to present quantum computing as a competitor to classical HPC for production-scale virtual screening. Overall, the results show that PP in extreme-scale virtual screening cannot be obtained simply by adopting portable programming models. Functional portability is necessary, but it does not automatically translate into sustained efficiency. Sustaining useful throughput requires co-designing the entire virtual screening campaign, aligning core kernels and molecular representations directly with platform orchestration.

PhD Committee

Prof. Ferrandi Fabrizio, Politecnico di Milano

Prof. Peng, Bo

Prof. Weidendorfer, Josef