Statement of Purpose

of Yuejiang Yu (CS PhD applicant for Fall—2023)

As an undergraduate, I was lucky to participate in global **high-performance computing (HPC)** competition and several **HPC+Science** projects. I am deeply attracted by this fascinating research area with great potential in parallel computing, scientific applications, and heterogeneous accelerators. To further pursue my interest, I am excited to apply for PhD in Computer Science at Cornell. This will prepare me for fusing HPC and computational science concerning cutting-edge applications.

1 ASC20-21[1] Supercomputer Competition

In my first few semesters, I participated in the competition as the youngest representative of Peking University and won **First Prize**. My main contribution is to accelerate the pulsar searching challenge called PRESTO by more than 30 times on CPU cluster. I joined PKU Supercomputing Group as a freshman with the eagerness to understand the operation of supercomputers. I acquired system-level profiling tools (Nvidia Nsight, Intel VTune, etc.), as well as broad HPC skills such as CUDA GPU programming, profiling, and parallel programming in OpenMP and MPI. In the first round of ASC20-21, PRESTO is accelerated by 29 times using node-level and thread-level parallelism and I/O optimization such as Ramdisk. Further, I managed to detect and reduce specific hotspots with SIMD operations of floating point calculation of PRESTO before the final round. Teamwork is highly essential for on-site competition as we cooperated closely to deal with unexpected incidents. We also managed to build and accelerate the mystery application released on the spot with MPI in a few hours. Fortunately, we achieved 4th in the final out of 350+ teams, which the best of our school history.

This experience equips me with **extensive knowledge of HPC and scientific computing**, paving my way to accelerate specific applications in depth.

2 Academic Experience

Since the following semester of ASC20-21, I have been working as an algorithm research intern at DP Technology, with the motivation to **combine HPC with cutting-edge scientific tasks and bring changes to the industry**. I had the honor of being supervised by **Gorden Bell Prize** winner **Dr. Linfeng Zhang** to accelerate a scientific task called molecular docking. In recent years, a rapid increase in the size of commercial libraries [1] was observed. Consequently, widely used molecular docking methods [2, 3] may be hard to meet researchers' needs for the speed and cost of virtual screening, especially when facing urgent public health events such as the COVID-19 pandemic[4]. With skills in HPC, I designed and developed a **GPU-accelerated** program—Uni-Dock. Based on CUDA, Uni-Dock achieved more than **1600 times** acceleration on GPU compared to CPU and saved five times the cost without losing accuracy. Integrated into advanced industrial Computer-aided drug design (CADD) product, Uni-Dock enables ultra-large virtual screening of early-stage drug discovery in hours, used by hundreds of chemists. The key to acceleration is the architecture of multi-level parallelism on GPUs. Moreover, other CPU hotspots were migrated to GPU, and I/O was boosted using OpenMP. Uni-Dock is a stable, reliable, and versatile algorithm that has passed the test of computational chemists and engineered to be used in production environments.

The work leads to a paper [5] submitted to ACS JCTC, software copyright, and new product landing in the drug discovery industry.

To the border of hardware limits In developing Uni-Dock, I was deeply aware of the limits of hardware and I pushed Uni-Dock to the extreme under these constraints. When kernel-level parallelism was implemented, I found that the threads of NVIDIA V100 GPU are not fully utilized and memory copying between host and device is sequential and relatively sparse. Time spent on the CPU doing I/O operation

[1] 2020-2021 ASC Student Supercomputer Challenge

and data precalculations was greatly accumulated. After several iterations of profiling and tuning, I managed to stably utilize more than 95% of GPU memory, with the time of memory copy and I/O greatly reduced by 80%. Further, I migrated and tuned Uni-Dock on various hardware from NVIDIA, AMD and Sugon to find the most suited architecture for Uni-Dock in speed and cost–performance ratio. The migration of Uni-Dock to ROCm architecture GPUs leads to **the third prize** in *the 3rd Priority Research Application* competition held by Dawning Information Industry Co., Ltd.

Combine Uni-Dock with Machine Learning In DP Technology, I also worked closely with senior researcher **Guolin Ke**, leader of the machine learning (ML) group and creator of LightGBM, to combine ML-based docking model Uni-Mol [6] with Uni-Dock. This resulted in an outstanding result with molecular binding pose prediction success rate improved from 69% to 86%, tested on the CASF-2016 dataset. Uni-Mol is a universal molecular representation learning (MRL) framework that significantly enlarges the representation ability and application scope of MRL schemes, build on two SE(3)-equivariant transformers. Other ML models can only rescore the poses after traditional docking, while our approach utilizes predictions from Uni-Mol throughout Monte-Carlo searches of Uni-Dock. This innovation significantly improves force-field accuracy of both fine structures and atomic positions of molecules. This work leads to *an invention patent under review*.

3 Research Proposal & Why Cornell

I am applying to Cornell due to its leading position in machine learning, scientific computation, and HPC. There are three faculties at Cornell whose projects are particularly appealing to me. I am keen to work under the supervision of **Prof. Carla P. Gomes** to further my research in AI4Science and computational sustainability. From my research experiences, I would like to join **Prof. Guilia Guidi**'s projects on the intersection of HPC and large-scale computational biology software. Also, I'm very interested in Monte-Carlo searching and machine learning, which is the expertise of **Prof. Chris De Sa**. I am convinced that with the help of extraordinary minds and the top interdisciplinary research environment at Cornell, I can continue to contribute to the field of HPC+AI+Science.

References

[1] AN Shivanyuk et al. "Enamine real database: Making chemical diversity real". In: *Chemistry today* 25.6 (2007), pp. 58–59.

[2] Garrett M Morris et al. "AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility". In: *Journal of computational chemistry* 30.16 (2009), pp. 2785–2791.

[3] Oleg Trott and Arthur J Olson. "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading". In: *Journal of computational chemistry* 31.2 (2010), pp. 455–461.

[4] Andreas Luttens et al. "Ultralarge virtual screening identifies SARS-CoV-2 main protease inhibitors with broad-spectrum activity against coronaviruses". In: *Journal of the American Chemical Society* 144.7 (2022), pp. 2905–2920.

[5] Yuejiang Yu et al. "Uni-dock: A gpu-accelerated docking program enables ultra-large virtual screening". In: *ChemRxiv* (2022).

[6] Gengmo Zhou et al. "Uni-Mol: A Universal 3D Molecular Representation Learning Framework". In: *ChemRxiv* (2022).