[BIOINFO 2025]

Optimized GPU Kernels for Training Large-Parameter Biomolecular Structure Prediction Models

Sanggeun Park[1], Minkyung Baek[1,2*]

[1] *School of Biological Sciences, Seoul National University*
[2] *Interdisciplinary Program in Artificial Intelligence, Seoul National University*
* *Corresponding author: minkbaek@snu.ac.kr*

Deep learning has revolutionized biomolecular structure prediction, setting the current state-of-the-art standard. However, training such models is severely constrained by practical limitations: extremely long runtimes and high memory demands. AlphaFold3, for instance, requires about 2,560 A100 GPU-days for initial training, a cost feasible only with access to vast clusters of high-end GPUs. In this work, we present optimized and fused GPU kernels implemented in Triton to address these bottlenecks. Kernel fusion was applied to operations including LayerNorm, TriangleMultiplication, Transition, and several smaller functions. In addition, we modified FlashAttention-2 to provide more efficient attention layers. We further designed an *atomic_add*-based all-to-all attention mechanism for the atom-level diffusion transformer, replacing the suboptimal masked attention. Collectively, these optimizations accelerate the full model by approximately 1.5×–2× compared to standard PyTorch kernels. This improvement reduces the estimated training cost to 960–1,920 A100 GPU-days, a far more manageable range, and enables large-parameter biomolecular structure prediction models to be trained with substantially fewer high-end GPUs.