Accelerating the Discovery Loop: Production-Ready Tools for GPU-Powered Materials Science

Piero Altoè[1]
[1]NVIDIA Corporation, 2788 San Tomas Expressway, Santa Clara, California 95051, USA

paltoe@nvidia.com

The integration of Machine Learning Interatomic Potentials (MLIPs) into molecular dynamics has significantly reduced the cost of first-principles accuracy. However, as the field moves toward high-throughput screening and large-scale data generation, traditional "pre-" and "post-inference" tasks—such as neighbor list construction, dispersion corrections, and long-range electrostatics—have become primary computational bottlenecks due to their historical reliance on CPU-centric implementations. We present NVIDIA ALCHEMI, a specialized, software ecosystem designed to address these disparities by providing an end-to-end, GPU-native workflow for atomistic modeling.

At the core of this stack is ALCHEMI Toolkit-Ops[1], a library of high-performance primitives built on NVIDIA Warp, an open-source Python framework for differentiable GPU computing. A defining technical feature of the Ops layer is its native batching architecture. Unlike traditional MD codes that process one system at a time, ALCHEMI[2] treats a "batch" of hundreds or thousands of independent molecules as a single contiguous memory layout. This allows for the simultaneous execution of kernels across the GPU's streaming multiprocessors (SMs), maximizing hardware occupancy even for small-to-medium systems that would otherwise underutilize the GPU.

ALCHEMI leverages Just-In-Time (JIT) compilation to deliver these batched kernels for neighborhood computation, periodic boundary conditions, and Becke-Johnson damped DFT-D3 dispersion. By utilizing the NVIDIA Warp runtime, these kernels achieve millions of atoms per second in throughput. Furthermore, the batching logic is built into the Geometry Optimizers and Integrators (NVT/NPT) within the ALCHEMI NIMs. This enables the parallel relaxation of massive chemical libraries—such as those encountered in data set generation or catalyst screening—without the overhead of sequential CPU-to-GPU data transfers.

This modularity is further extended to the cloud via ALCHEMI NIM microservices[3], providing containerized, scalable services for high-throughput conformer searches and molecular property mapping. The synergy between these layers results in speedups of up to 800x for complex geometry relaxations. This unified, batched stack establishes a production-ready infrastructure for researchers to navigate the vast chemical design space at the speed of modern GPU hardware.

References

[1] NVIDIA Technical Blog: Accelerating AI-Powered Chemistry and Materials Science Simulations with NVIDIA ALCHEMI Toolkit-Ops

[2] NVIDIA Blog: Revolutionizing AI-Driven Material Discovery Using NVIDIA ALCHEMI

[3] NVIDIA Developer: Faster Chemistry and Materials Discovery with AI-Powered Simulations Using NVIDIA ALCHEMI

Figures

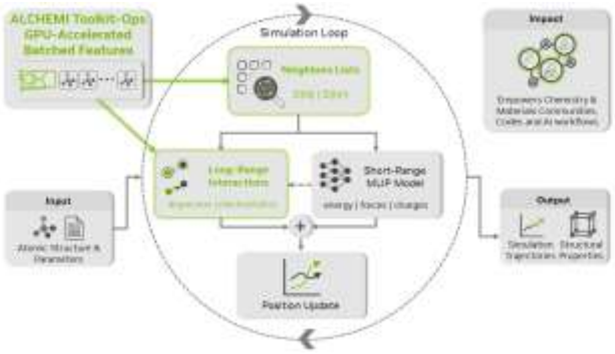

Figure 1. *NVIDIA ALCHEMI Toolkit-Ops is a repository of modules developed specifically for GPU-accelerated batched operations (one GPU, many systems) support for MLIPs and molecular dynamics engines.*