Revolutionizing AI-Driven Material Discovery Using NVIDIA ALCHEMI

Wen Jie Ong[1], **Piero Altoè[1]**, Justin S. Smith[1], Dallas Foster[1], Melisa Alkan[1], Harry Petty[1]

[1]*NVIDIA Corp., Santa Clara, CA*

wong@nvidia.com

The discovery of novel chemicals and materials will revolutionize various industries such as energy storage, environmental remediation, and manufacturing. Traditional discovery methods have been laborious and time-consuming, often taking years or even decades to move from hypothesis to production. This lengthy process involves extensive experimentation, trial and error, and significant resource investment. Traditional methods are not only slow but also costly, limiting the pace of innovation in chemistry and materials science. However, with the advent of AI, this paradigm is rapidly changing. AI technologies have the potential to revolutionize the way materials are discovered and developed, making the process faster, more efficient, and more cost-effective. For example, graph neural network interatomic potentials constrained to preserve the physical symmetries of atomic systems provide highly accurate and efficient alternatives to density functional theory. NVIDIA ALCHEMI (AI Lab for Chemistry and Materials Innovation) is at the forefront of this revolution, enabling AI to accelerate chemical and materials discovery by accelerating the components needed to deploy these AI methods in real-world workloads with maximum efficiency and usability.

NVIDIA ALCHEMI aims to employ a comprehensive set of AI-accelerated microservices and the required software tooling to enable efficient and user-friendly solutions for hypothesis generation, solution space definition, property prediction, and experimental validation. By utilizing chemistry-informed large language models (LLMs) and machine learning interatomic potential (MLIP) AI models, our platform aims to enable the synthesis of vast amounts of chemical literature, formulate and refine hypotheses, and predict material properties with unprecedented speed and accuracy. In this presentation, we will introduce our suite of batched geometry relaxation (BGR) tools which provide 100x+ speedup for the task of geometry relaxation, a common inference workload in materials and chemical discovery.

NVIDIA Alchemi's Batched Geometry Relaxation (BGR) tool is a cutting-edge solution designed to accelerate the process of geometry relaxation in materials and chemical discovery. Geometry relaxation is a critical step in computational chemistry and materials science, where the atomic positions of a system are optimized to find the lowest energy configuration. This process is essential for predicting material properties and understanding chemical stability. The BGR tool leverages advanced AI models, such as MACE[1] and AIMNet2[2], to perform geometry relaxation with unprecedented speed and accuracy.

Table 1. Accelerated geometry relaxation using the NVIDIA Batched Geometry Relaxation NIM with MACE-MP-0 model for 2,048 periodic materials systems with 20-40 atoms.

Batched Geometry Relaxation NIM	Batch size	Total time (s)	Speedup
Off	1	874	1x
On	1	36	25x
On	128	9	100x

Table 1 presents the benchmark results for accelerated geometry relaxation using the NVIDIA Batched Geometry Relaxation tool with the MACE-MP-0 model for 2,048 periodic materials systems with 20-40 atoms. When the BGR tool is turned off, the total time for relaxation is 874 seconds. However, with the BGR NIM turned on and a batch size of 1, the total time is reduced to 36 seconds, resulting in a 25x speedup. Further increasing the batch size to 128 reduces the total time to 9 seconds, achieving a 100x speedup1.

Table 2. Accelerated geometry relaxation using the NVIDIA Batched Geometry Relaxation NIM with AIMNet2 model for 851 small organic molecules with an average of ~20 atoms.

Batched Geometry Relaxation NIM	Batch size	Total time (s)	Speedup
Off	1	678	1x
On	1	12	60x
On	64	0.9	800x

Table 2 shows the benchmark results for accelerated geometry relaxation using the NVIDIA Batched Geometry Relaxation NIM with the AIMNet2 model for 851 small organic molecules with an average of ~20 atoms. With the BGR NIM turned off, the total time for relaxation is 678 seconds. When the BGR NIM is turned on and a batch size of 1, the total time is reduced to 12 seconds, resulting in a 60x speedup. Increasing the batch size to 64 further reduces the total time to 0.9 seconds, achieving an 800x speedup.

These unprecedented efficiency gains are primarily due to the use of inference batching and the elimination of CPU to GPU communication overhead during batched simulation. Batching allows multiple tasks to be processed simultaneously, significantly reducing computation time and increasing throughput. This is achieved using the NVIDIA Warp framework which has the flexibility to directly implement accelerated GPU kernels in machine

learning frameworks such as PyTorch and Jax. Additionally, optimizing AI inference for chemical simulations with NVIDIA CUDA libraries such as cuEquivariance further enhances efficiency of models such as MACE by optimizing specific slow operations. These innovations enable researchers to achieve faster, more accurate results, paving the way for next-generation innovations in various industries.

References

[1] I. Batatia, D. P. Kovacs, G. N. C. Simm, C. Ortner, and G. Csanyi. MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields. *In NeurIPS 35*, 2022.

[2] D. Anstine, R. Zubatyuk, and O. Isayev. AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs*, Preprint at ChemRxiv*, 2024.

Figures

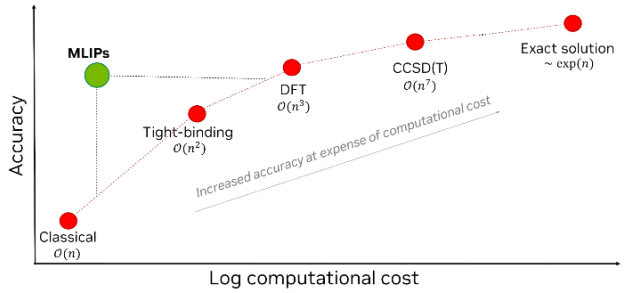

Figure 1. Breaking the accuracy-cost conundrum with MLIPs.