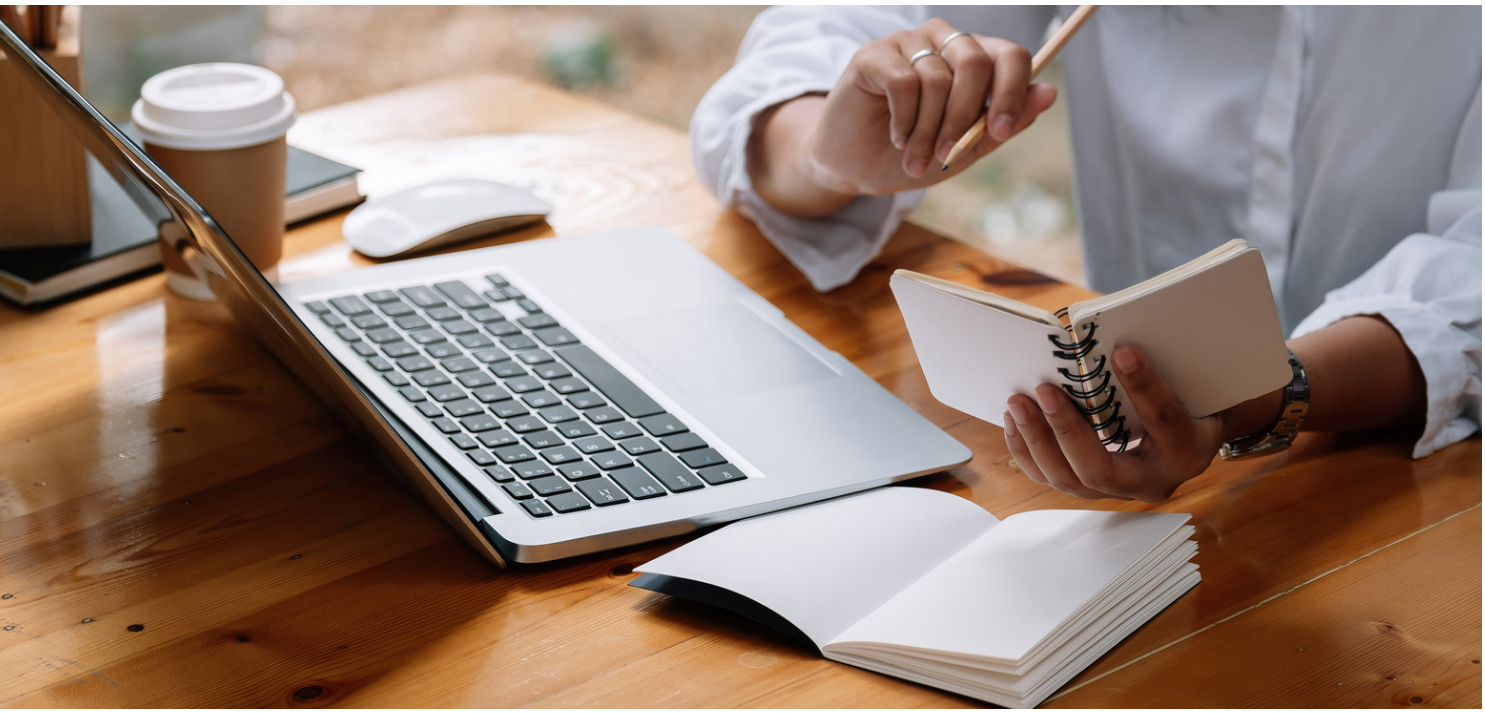
HPE Cray Supercomputing

Business white paper

Accelerating computational chemistry with high performance computing and artificial intelligence

Table of contents



Executive summary

Computational chemistry plays a vital role in fields such as drug discovery, molecular biology, and materials science by simulating molecular interactions and reactions. As the complexity of these systems increases, so do the computational challenges, necessitating the use of high performance computing (HPC) and artificial intelligence (AI). Together, these technologies have transformed computational chemistry, enabling simulations at unprecedented speed and scale. This white paper explores the role of HPC and AI in advancing computational chemistry, the challenges they help overcome, and their key applications, particularly in drug discovery, materials design, and protein folding.

Introduction

Computational chemistry allows researchers to simulate molecular and chemical processes to predict behaviors, interactions, and properties. In areas such as drug discovery, protein folding, and materials science, these simulations are critical for generating insights that would be difficult, costly, or time-consuming to achieve experimentally. However, the increasing complexity of chemical systems—ranging from biomolecular interactions to material design—requires substantial computational resources.

This is where HPC and AI come into play. HPC provides the necessary infrastructure to handle large datasets and perform complex algorithms while AI enhances the predictive capabilities of computational chemistry by learning from vast datasets and accelerating decision-making. Together, HPC and AI are pushing the boundaries of computational chemistry, particularly in health and life sciences, by significantly improving efficiency, speed, and ;acy.

The role of HPC and AI in computational chemistry

1. Enhancing large-scale simulations with HPC

In computational chemistry, large-scale simulations such as molecular dynamics (MD) and quantum mechanical modeling are essential for studying complex molecular systems. For example, MD simulations of large proteins or drug compounds often require solving millions of interactions between atoms, which results in an overwhelming number of computations. Traditional computing methods struggle with such tasks due to the sheer volume of data and time required.

HPC: HPC systems break down these simulations into parallel tasks, reducing the overall computation time from weeks or months to mere hours or days. For instance, in 2021, HPC-enabled simulations of large protein complexes generated **100 TB of data per simulation**, something that would be impractical to manage without supercomputers. HPC systems have the capability to simulate molecular systems with over 100 million atoms, offering unprecedented insights into molecular behaviors.[1]

AI: AI algorithms further optimize these simulations by predicting molecular behaviors more accurately and efficiently. For example, AI models can be trained to predict energy landscapes in MD, reducing the need for exhaustive computations. Additionally, AI-driven deep learning models are used to refine molecular structures, offering accurate predictions about how molecules fold, interact, or react—drastically reducing computational load and time.

[1] "Coupling streaming AI and HPC ensembles to achieve 100-1000x faster biomolecular simulations," ParslFest, October 19, 2023.

2. Accelerating drug discovery pipelines

Drug discovery is one of the most significant areas in computational chemistry, where the process often involves virtual screening of millions of compounds, molecular docking to identify how drug candidates interact with biological targets, and binding affinity predictions to optimize leads. These tasks are resource-intensive and time-consuming.

HPC: With the computational power of HPC, pharmaceutical companies can conduct large-scale virtual screening at a fraction of the time required for experimental methods. For example, during the COVID-19 pandemic, researchers leveraged HPC to screen over **2 billion compounds** in a matter of days, identifying potential candidates for COVID-19 therapies.[2] These simulations are critical for narrowing down large chemical libraries to the most promising compounds.

AI: AI plays a pivotal role by learning from these datasets and predicting binding affinities and drug-target interactions more efficiently than traditional computational approaches. AI models such as DeepChem have been used to screen compounds and rank them by potential effectiveness, allowing researchers to focus on the most promising candidates for experimental validation. Additionally, reinforcement learning algorithms are increasingly used to optimize molecular structures, ensuring that drug candidates have the desired properties while minimizing side effects.

3. Protein folding and structural biology

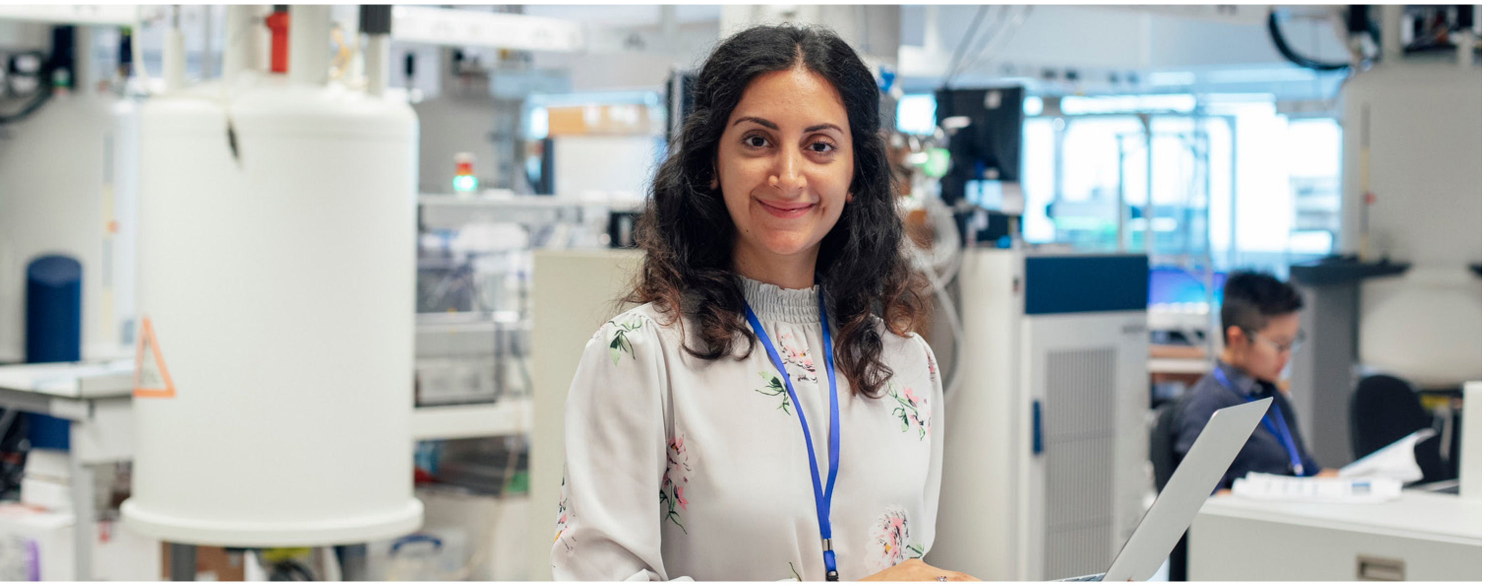
The study of protein structures and how they fold is critical to understanding biological processes and developing new therapies. Protein folding simulations are computationally demanding, involving the simulation of molecular interactions over time, which can require immense computational resources.

HPC: HPC systems have made it possible to simulate protein folding pathways for large biomolecules. These simulations not only provide insights into normal cellular processes but also help explain diseases caused by misfolded proteins, such as Alzheimer's disease and Parkinson's disease. Using HPC, researchers can simulate the folding process of proteins involving thousands of atoms and predict how environmental factors or mutations may affect their structure and function.

AI: The development of AlphaFold by DeepMind in 2021 showcased how AI can revolutionize protein structure prediction. AlphaFold uses deep learning models trained on known protein structures to predict the 3D conformation of proteins with near-experimental accuracy.[3] This has profound implications for drug discovery and biotechnology, as accurate protein structures are essential for designing therapies targeting specific proteins. The integration of AI models with HPC infrastructure allows for even faster and more accurate protein folding simulations, drastically reducing the time required to solve these complex biological problems.

[2] "Scalable HPC & AI Infrastructure for COVID-19 Therapeutics," arXiv, October 20, 2020.

[3] "AlphaFold 2 and the Next Revolution in Protein Folding. Nature," Jumper, J., and others, 2021



Key applications of HPC and AI in computational chemistry

1. Drug discovery and development

a. **Virtual screening and docking:** HPC enables large-scale virtual screening while AI refines these results by predicting molecular interactions more accurately. By combining HPC's computational power with AI's predictive capabilities, pharmaceutical companies can screen millions of compounds quickly and efficiently, identifying the most promising drug candidates.

b. **Lead optimization:** After identifying drug candidates, AI models use molecular simulations to optimize their properties, such as improving binding affinity or reducing toxicity. HPC systems handle the computational load of these simulations while AI models accelerate the optimization process through machine learning algorithms.

2. Materials science and catalysis

a. **Reaction mechanism modeling:** In materials science, HPC and AI are used to simulate chemical reactions at the molecular level, enabling researchers to design new catalysts for industrial processes. HPC systems simulate the atomic and electronic behavior of catalysts during reactions while AI models predict how small changes to molecular structures can impact reaction efficiency.

b. **Polymer and materials design:** AI is increasingly used to predict the properties of novel materials, such as polymers before they are synthesized. HPC accelerates these simulations by handling the large datasets and complex equations required for accurate predictions. Together, HPC and AI allow for faster and more efficient materials design.

3. Quantum chemistry and molecular dynamics

a. **Quantum mechanical simulations:** Quantum chemistry relies on solving the Schrödinger equation for molecular systems, a process that becomes exponentially more complex as the number of atoms increases. HPC systems allow for the parallel processing of these equations, enabling large-scale quantum mechanical simulations.

b. **AI models**, particularly those based on neural networks, are increasingly being used to approximate quantum mechanical behaviors, reducing computational time while maintaining accuracy. These AI-enhanced quantum simulations are critical for predicting the electronic structure of molecules, which is essential for understanding chemical reactivity and designing new materials.

Challenges and future directions

1. **Computational costs and infrastructure**

While HPC and AI offer transformative capabilities, the infrastructure required to run large-scale simulations remains costly. Access to supercomputers, high-speed networks, and data storage solutions is often limited to well-funded institutions or organizations.

2. **Data management and storage**

The data generated by HPC and AI simulations can be overwhelming. For example, a single MD simulation can generate **terabytes of data** that need to be stored, managed, and analyzed. Efficient data management systems and real-time analytics are necessary to handle the immense data loads produced by computational chemistry.[4]

3. **Algorithmic complexity**

Although AI is advancing rapidly, some algorithms used in quantum chemistry and MD are still highly complex and computationally expensive. Improvements in algorithmic efficiency, especially for parallelizing tasks across HPC clusters, will be crucial to further scaling up simulations.

Future directions: Integration of quantum computing into HPC infrastructures may provide a leap forward for computational chemistry, offering exponentially faster solutions to problems such as quantum mechanical modeling. AI-driven automation of workflows, combined with scalable cloud-based HPC, could democratize access to these advanced tools, allowing more researchers to benefit from the power of computational chemistry.

[4] "Insilico Medicine's AI-Driven Drug Discovery Platform: Applications and Case Studies," Journal of Chemical Information and Modeling, Ebert, M. C., and others, 2021

Conclusion

The integration of HPC and AI is revolutionizing computational chemistry, providing the computational power and intelligence needed to solve complex molecular problems in drug discovery, materials science, and structural biology. By leveraging HPC's ability to scale simulations and AI's predictive capabilities, researchers are accelerating breakthroughs in health and life sciences. As these technologies continue to evolve, the future promises faster, more accurate simulations, leading to transformative discoveries in both chemistry and biology.

Learn more at

HPE.com/us/en/supercomputing.html

Visit HPE.com

Chat now

a00143008ENW, Rev. 1