Project Title: Modern Exascale Acceleration of Amoeba Polarized Force Fields for the Tinker-9 Molecular Modeling Software

By: Hussain Kadhem, Joyce Kim
Collaborator: Dr. Jay Ponder

Practical Content and Creativity

Replica-Exchange Molecular Dynamics (REMD) is a prominent computer simulation method employed to explore the lowest energy state of complex systems composed of interacting particles. This technique effectively addresses the issue of entrapment in metastable states at low temperatures or the emergence of distinct stable states at higher temperatures. In the realm of computational chemistry and biophysics research, the Tinker software package has gained widespread popularity as a molecular modeling and simulation tool, offering diverse functionalities for investigating molecular behavior and properties.

Tinker9 is a significant advancement and extension of the original Tinker software and represents a notable progression in terms of implementation, as it is now developed in C++ and integrates advanced features aimed at enhancing performance and functionality. Notably, Tinker9 harnesses the immense computational power of GPUs leading to accelerated calculations and heightened efficiency in molecular simulations.

The primary objective of Tinker9 is to streamline the process of molecular dynamics simulations, enabling more efficient exploration of the conformational space of complex molecular systems. Through the integration of replica-exchange molecular dynamics (REMD) with the Tinker9 software package, we seek to leverage the benefits of REMD in addressing the challenges associated with metastability and temperature-dependent stable states. By running multiple replicas of the system at different temperatures in parallel and periodically changing configurations based on the Metropolis criterion, REMD facilitates effective sampling of both low and high energy configurations, leading to more accurate depiction of the system's energy landscape.

Implementation and Tuning Effort

For the implementation and tuning effort, we developed a parallel algorithm for the Replica exchange method in molecular dynamics simulations using Tinker9 software. Instead of directly permuting the replica configurations, we propose permutations by permuting the temperatures, reducing data transfer. To parallelize the process, we decomposed permutations into transpositions of neighboring replicas, following a bipartite exchange scheme based on the replica index and exchange steps. This ensures a well-defined permutation without overlapping transpositions. The acceptance rate for a proposal depends on the temperatures and potential energies of the replicas, with communication and decision-making involved. We have synchronized the exchange steps with the save steps of the Tinker integrator to minimize disruption of GPU kernels.

In our project, we aim to investigate the effectiveness and optimize the replica exchange method in accelerating molecular dynamics simulations using Tinker9 software. Molecular dynamics simulations are essential for studying the physical and chemical properties of molecules and materials. However, they can be computationally demanding, especially for larger systems and longer simulation times. The replica exchange method helps overcome this challenge by allowing replicas to exchange configurations, exploring higher-energy states and avoiding energy minima traps.

Our research specifically focused on parallelizing replica exchange simulations using Tinker9 software. We assessed the scalability and efficiency of the parallel implementation and compared its performance with non-parallelized simulations. This study provides valuable insights into the potential of replica exchange simulations for accelerating molecular dynamics simulations and optimizing computational resources in the field of molecular modeling and simulation.

The implementation comprises three key files: "Mdintg.cpp," "Mdsave.cpp," and "Xdynamic.cpp." Each file serves a specific purpose within the overall implementation and tuning effort. "Mdintg.cpp" is responsible for integrating MD data, handling MD-related tasks, and configuring MD integrators and time scaling. "Mdsave.cpp" focuses on saving MD simulation states and introducing asynchronous data-saving functionality. Finally, "Xdynamic.cpp" acts as the main function for running the MD simulation, handling initialization, reading input coordinates, executing mechanics-related tasks, propagating dynamics, and performing necessary cleanup.

Throughout the implementation and tuning process, we have emphasized attention to detail, efficient memory management, and flexibility through user-defined inputs. The code has been optimized to enhance the performance of MD simulations, improve data-saving processes, and incorporate parallelization techniques. By conducting this research and evaluating the parallel implementation of replica exchange simulations, we gained valuable insights into accelerating molecular dynamics simulations and optimizing computational resources in the field of molecular modeling and simulation.

Experimental Data: Scaling and Performance Analysis

The experimental data provides valuable insights into the scaling and performance analysis of the simulation with different numbers of replica exchange times. The results reveal that the simulation with eight replica exchange times has the highest wall times, indicating increased computational requirements compared to configurations with fewer exchange times. This is attributed to the added complexity and inter-replica communication involved in handling eight replicas, resulting in longer wall times. Additionally, when comparing simulations with the same number of steps, the wall times for the eight replica exchange cases are consistently higher than those for simulations with one, two, or four replica exchange times.

However, despite these variations in wall times, the data consistently demonstrates a linear scaling trend. As the number of simulation steps increases, the wall times also increase proportionally, indicating a consistent computational requirement throughout the simulations. It is important to consider that increasing the number of replica exchange times can provide

benefits such as enhanced sampling and exploration of conformational space. However, it comes at the cost of escalated computational costs and longer wall times due to the increased complexity and inter-replica communication.

In conclusion, the experimental data highlights the scaling behavior of the simulation with eight replica exchange times in terms of wall times, where the wall time increases proportionally with the number of simulation steps. It emphasizes the computational costs associated with more replica exchange times and provides insights into the trade-offs between enhanced sampling and computational performance in replica exchange simulations.

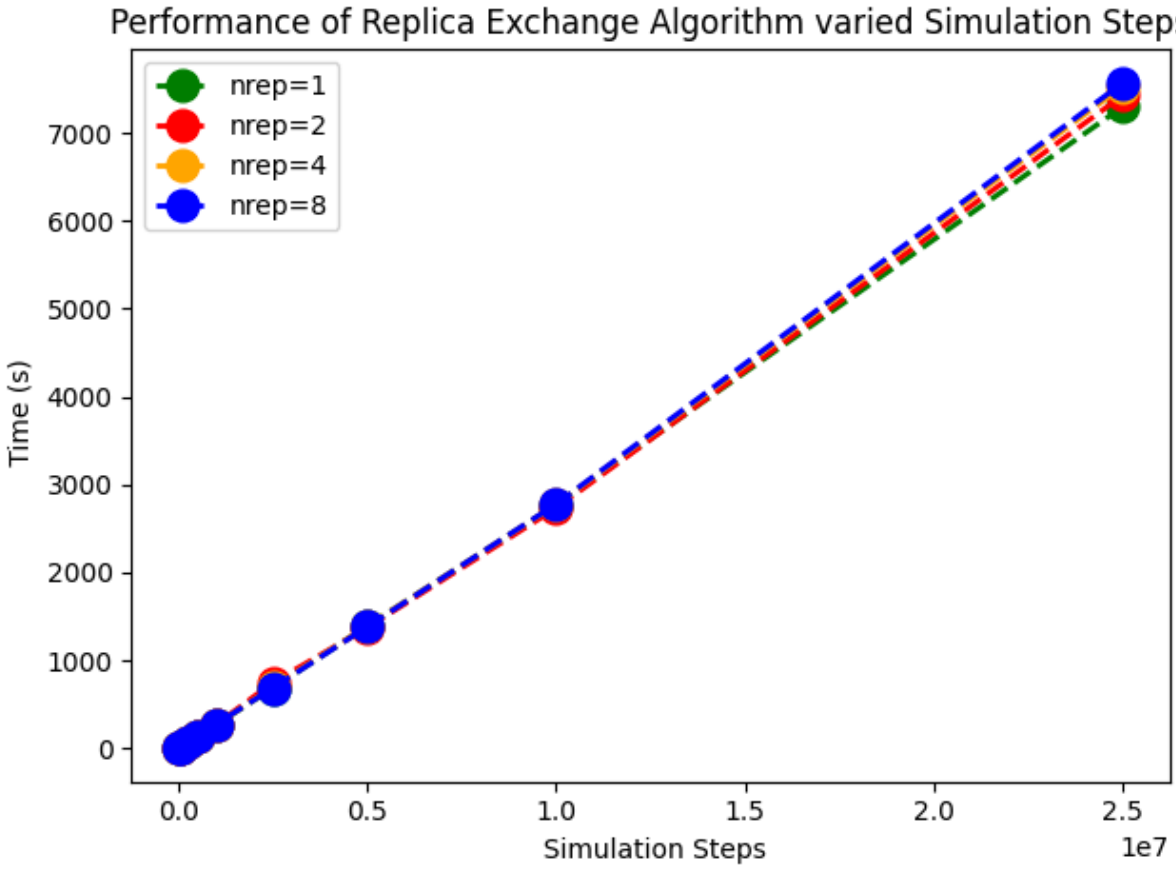

Figure 1: This figure presents a performance evaluation of simulation steps and time for different numbers of replica exchange occurrences. The data is presented for four scenarios: one replica exchange, two replica exchanges, four replica exchanges, and eight replica exchanges.

The experimental data illustrates scaling and performance analysis of a simulation using different numbers of replica exchanges (1, 2, 4, and 8) and their corresponding root mean square deviation (RMSD) values. The RMSD values reflect the degree of structural deviation between the simulated structure and the reference structure.

The data reveals that with a single replica exchange, the simulation generates an RMSD value of 6.14, indicating a certain level of deviation from the reference structure. When the number of replica exchanges is increased to two, the RMSD value rises to 12.313, suggesting a greater level of structural deviation during the simulation. However, with four replica exchanges, the RMSD value slightly decreases to 10.693, indicating an improved agreement with the reference structure compared to the case with two replica exchanges. The most noteworthy result is observed with eight replica exchanges, where the simulation achieves the lowest RMSD value of 3.507 among all the tested configurations. This signifies a higher accuracy and similarity between the simulated structure and the reference structure.

These findings highlight the potential for increased accuracy and agreement with the reference structure by employing a higher number of replica exchanges. The data demonstrates that as the number of replica exchanges increases, the simulation can achieve a higher fidelity in reproducing the desired structural characteristics. This suggests that replica exchange simulations with a greater number of exchanges can effectively explore the conformational space and improve the accuracy of the results.

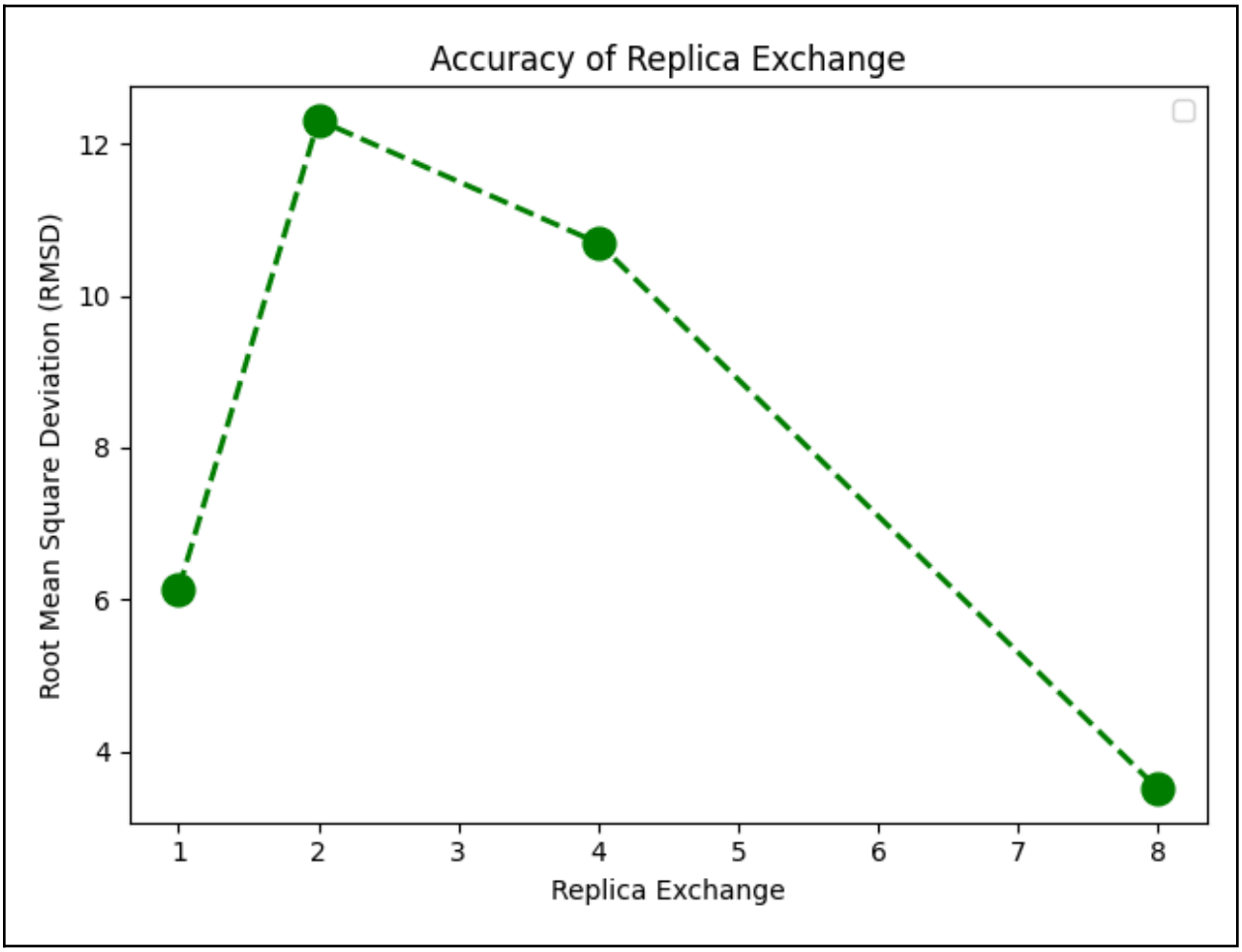

Figure 2: This demonstrates the accuracy of replica exchange events measured by the root mean square deviation (RMSD) values. The x-axis represents the number of replica exchange occurrences, while the y-axis represents the corresponding RMSD values.

Experimental Data Analysis: Interesting Inputs or Outputs

The experimental data provides valuable inputs regarding the scaling and performance analysis of the simulation with varying numbers of replica exchange times. Notably, the simulation utilizing eight replica exchange times exhibits the longest wall times compared to configurations with fewer exchange times. This observation can be attributed to the increased complexity and inter-replica communication associated with handling eight replicas. Moreover, the data consistently demonstrates a linear scaling trend, whereby an increase in the number of simulation steps corresponds to a proportional increase in wall times across all configurations. However, it is crucial to consider the additional computational costs and longer wall times incurred when employing more replica exchange times.

In terms of the accuracy of replica exchange events, the experimental data provides valuable outputs in the form of root mean square deviation (RMSD) values, which measure the structural deviations from the reference structure. The simulations conducted with different numbers of replica exchanges exhibit varying levels of agreement with the reference structure. The simulation employing a single replica exchange shows a moderate RMSD value of 6.14, indicating some degree of deviation. Increasing the number of replica exchanges to two results in a higher RMSD value of 12.313, signifying a greater level of structural deviation during the simulation. However, with four replica exchanges, the RMSD value slightly decreases to 10.693, suggesting an improved agreement compared to the case with two replica exchanges. Notably, the simulation with eight replica exchanges achieves the lowest RMSD value of 3.507 among all the tested configurations, indicating the highest accuracy and similarity to the reference structure.

All in all, the experimental data highlights the scaling behavior in terms of wall times and sheds light on the impact of replica exchange events on the accuracy of the simulations. Increasing the number of replica exchanges can lead to enhanced structural agreement, as evidenced by lower RMSD values. However, this improvement comes at the cost of increased computational requirements and longer wall times.

Theoretic Content and Creativity

The integration of Tinker9 and replica exchange in molecular simulations and computational drug design brings numerous advantages. Tinker9's compatibility with widely recognized forcefield models like CHARMM and AMBER allows for a more accurate representation of physics, particularly concerning electrostatic interactions. This means that Tinker9 can capture the intricate details of molecular interactions and provide reliable results.

Additionally, Tinker9's capability to handle multipole interactions, including anisotropic interactions, further enhances simulation accuracy. Many molecular systems exhibit complex

charge distributions and non-spherical shapes, and Tinker9's ability to handle these interactions enables a more realistic representation of these systems. This is particularly important in drug design, where accurate representation of molecular properties is crucial for understanding drug-target interactions.

In contrast, Replica exchange is a Monte Carlo method that plays a crucial role in improving sampling efficiency. By running multiple replicas of the simulation at different temperatures, replica exchange enables the exchange of particle configurations between replicas. The exchange is governed by a Metropolis criterion, which ensures that configurations are exchanged with a probability based on the energy difference between the replicas. This approach enables the replica at the target temperature to explore higher-energy configurations, facilitating faster convergence and avoiding entrapment in potential energy minima.

Design and Analysis of Algorithms

Thus, our plan to achieve parallelization in replica exchange was to assign a separate rank to each simulation replica. During the exchange step, replicas communicate by exchanging messages containing their current particle configurations. Each replica evaluates the acceptance criterion and, if satisfied, exchanges its particle data with the received buffer. Typically, the exchanges occur in a left-to-right manner, where replicas exchange configurations with their neighboring replicas to the left. This parallelization approach significantly speeds up the simulation process by distributing the computational load across multiple processors or nodes, allowing for the exploration of larger and more complex molecular systems.

The combination of Tinker9 and replica exchange, with their compatibility with established forcefield models, ability to handle multipole interactions, and support for parallelization, empowers researchers to comprehensively investigate molecular systems and advance computational drug design efforts. By leveraging these capabilities, researchers can achieve higher simulation accuracy, understand complex molecular behavior, and facilitate the development of novel drugs.

The integration replica exchange in Tinker9 offers a powerful approach for molecular simulations and computational drug design. The compatibility with forcefield models, handling of multipole interactions, and parallelization capabilities enhance simulation accuracy and efficiency. This combination of theoretical tools opens up avenues for in-depth exploration of molecular systems and the advancement of drug discovery efforts, leading to more accurate computational drug design with potential applications in various fields.

Impact: Difficulty and Timeliness of the Contribution

The integration of Tinker9 and replica exchange in molecular simulations and computational drug design presents both difficulty and timeliness in its contribution. From the perspective of design and analysis of algorithms, this integration introduces several algorithmic challenges that need to be addressed for efficient and accurate simulations.

One key challenge is optimizing the replica exchange process. Replica exchange employs a Monte Carlo method to accelerate molecular dynamics simulations. Running multiple simulation replicas at different temperatures and exchanging particle configurations between them can enhance sampling and improve convergence. However, determining the optimal acceptance ratio and tuning it for different temperatures is crucial. Finding the right balance between exploration of higher-energy configurations and convergence towards the desired target temperature requires careful algorithmic design and analysis.

Parallelization is another significant challenge in replica exchange. Assigning each simulation replica to a separate rank enables parallel execution. During the exchange step, replicas communicate by exchanging messages containing their particle configurations. The challenge lies in designing efficient communication and synchronization algorithms to minimize overhead and maximize computational throughput. Optimizing the left-to-right exchange strategy can further improve the efficiency of the parallelization process.

The integration of Tinker9 and replica exchange also requires addressing algorithmic considerations specific to molecular simulations. Efficient forcefield calculations for intermolecular interactions, particularly electrostatic interactions, need to be developed, taking advantage of established forcefield models like CHARMM and AMBER. Handling multipole interactions, including anisotropic interactions and computing forces and torques, adds complexity to the algorithmic design. Leveraging the computational power of GPUs can be instrumental in increasing performance and optimizing models.

Considering the timeliness of this contribution, it is essential to recognize the significance of computational drug design, especially in the context of urgent public health needs such as the COVID-19 pandemic. The added replica exchange to Tinker9 software provides a powerful approach to accelerate drug discovery efforts. The ability to compute binding interactions of ligands to proteins, interpolate between different ligands, and explore various applications opens up new possibilities in computational drug design. Potentially achieving high simulation accuracy, on the order of 1 kilocal/mol, can have a significant impact on the development of novel drugs.

Reference

Wang, Z. and Ponder, J. (no date) TinkerTools/Tinker9: Tinker9: Next generation of Tinker with GPU support, GitHub. Available at: https://github.com/TinkerTools/tinker9 (Accessed: 09 May 2023).

Kadhem, H. (no date) Tinker9, GitHub. Available at: https://github.berkeley.edu/hmk/tinker9 (Accessed: 09 May 2023).