SW-eMD: Molecular Dynamics Simulation Framework on Sunway TaihuLight Supercomputer

Shun Xu[1,2]; **Yi Zhao**[1]; **Zhong Jin**[1,2]
[1]Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190
[2]Center of Scientific Computing Applications & Research, Chinese Academy of Sciences , Beijing 100190

Abstract

Molecular dynamics (MD) simulation is an important method for modeling systems in atomic and molecular scale, and is widely used in physics, chemistry, materials and other related fields. Calculation of atom's interaction potential, particularly non-bonded interaction, is always a time-consuming task. In this work, we present an efficient MD simulation framework on Sunway TaihuLight supercomputer, which is ranked the top of World's TOP500 Supercomputers in recent two years. Domain decomposition of atom's 3D system is first designed in MPI parallel level cross nodes, and hard tasks such as potential calculation and neighbor list construction can be offloaded to heterogeneous many-core processor, in which 256bit-SIMD vectorization is also available. Our current implementation and optimization efforts provide a foundation for the high performance of large scale MD simulation.

Introduction

Molecular Dynamics (MD) programs simulate the behavior of atomic and molecular systems, leading to understanding of their functions. However, the computational complexity of such simulations pose a challenge to software application. Parallel machines , especially in heterogeneous architecture, provide the potential to meet this challenge for large temporal and spatial scale. To harness this potential, program needs to be well-designed with high computational efficiency and scalability. The mainstream molecular dynamics software such as NAMD[1], LAMMPS[2] and HOOMD-blue[3], take heterogeneous computation device like GPU to accelerate the MD calculation with a good acceleration effect. In our work, we design a MD software framework, named SW-eMD, to harness the computing power of TaihuLight which is the world's first system with a peak performance greater than 100 PFlops[4,5]. SW-eMD can offload hard tasks such potential calculation and neighbor list construction into the Computer Processing Element (CPE) composed of an 8x8 mesh on SW26010. Here we present the framework architecture of SW-eMD, and optimizing strategies of implementation on SW26010 processor.

Methods

The software architecture of SW-eMD (Fig. 1) is provided for high performance Molecular Dynamics simulations in biosystems and nanomaterials. The main features can be summarized as follows:

- Flexible MD simulation models
- Custom force fields integration
- Advanced modeling and simulating algorithms
- Heterogeneous computing on Sunway CPU.

To fully exploit the computing power of Sunway CPU of TaihuLight, our efforts focus on some strategies:

- The communication between Local Data Memory (LDM) and main memory and the CPE calculation are overlaid to meet the imbalance between computation and data communication.
- Point-Point and Collective communication are combined to use, For example MPI Allreduce operation collaborate with 64 CPE of SW26010
- Vectorization of floating point calculation is realized to exploit the instruction parallelism,

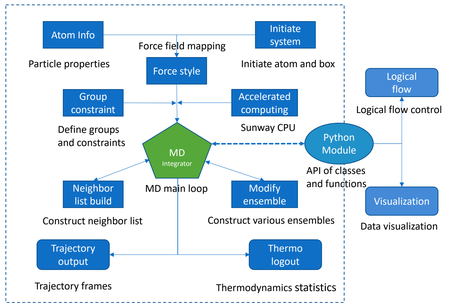

Figure 1. SW-eMD framework for high performance Molecular Dynamics simulations.

Results

The classical system of Lennard-Jones potential is taken as a test case, in which reduction unit is used. We implement the Replica Exchange method (also known as parallel tempering method) over MPI parallelism in SW-eMD. The Replica Exchange simulation demonstrates the parallell evolution of argon of Lennard-Jones potential with separated temperatures. Simulations are performed in core-group private mode of SW26010 CPU. The input file of SW-eMD is given as Fig. 2. three temperature points is parallel tempered shown as Fig. 3. And finally system temperature and potential evolution in separate replica are refactored as Fig. 4, in which energy trajectories are, as expected, overlapped with each other.

```
{
  "comm": {
    "world": ["3x2"],  "factor": [ "0.7",0.78, 0.86]
  },
  "force": {"file": "n250.myd",
    "type": "moly", "frame_used":0,  "newton": false, "ghost_vel": false,
    "neighbor": {"every": 2,"skin": 0.3,"type": "nsq"}
  },
  "log_out": {
    "every": 100,"file": "stdout",  "fields": "step temp epot ekin etotal press vol"
  },
  "update": {
    "steps": 50000,"timestep": 0.01, "type": "md_verlet"
  },
  "traj_out": {
    "every": 0, "file": "traj_out.myd"
  },
  "vel_init": {
    "rndseed": 78985,"temp": 1.2
  },
  "modify": {
    "type": "nvt",  "style": "langevin",  "tstart": 1.44,  "tstop": 1.44,
    "damp": 10, "seed": 3434,
    "bias": {
      "enable": false, "type": "beta",  "beta_g": 0.7,
      "alpha": 0.2,  "pe_g": -645,   "scale": "force"    }
  }
}
```

Figure 2. Configure file in JSON format as SW-eMD input.

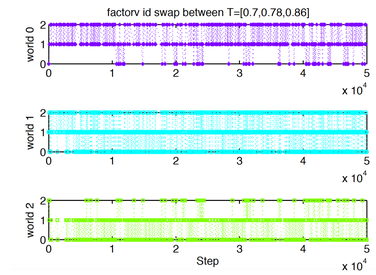

Figure 3. The temperature factor of MD system is tempered in Replica Exchange method.

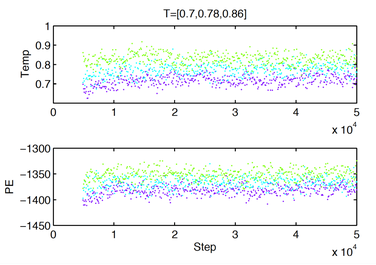

Figure 4. The temperature and Potential Energy of MD system is reconstruction in Replica Exchange method.

Discussion

In our Replica Exchange implementation, one replica use multiple MPI ranks, and each MPI rank can harness a 64-PE group of SW26010 CPU in core-group private mode.
The architecture of many-core SW26010 CPU with 64K LDM cache-like memory per thread is very different from other CPU and GPU device. Atom data is loaded into LDM by DMA before potential calculation in PE thread. A limited 64K LDM in fully occupation have a capacity of about 5K atoms of XYZ position data in single precision. If velocity is involved in potential calculation such as Dissipative Particle Dynamics, less atoms can be cached in LDM for calculation threads. This lead to low efficiency in neighbor list computation. In this case, software-simulated cache in LDM is a efficient and common used way. In our implementation, a fine-grained communication of registers buffer between 8x8 cores within the same column and row group is used to enhance the domain communication between PE threads.

Conclusions

We have implemented an basic framework for Molecular dynamics on Sunway TaihuLight. A Multilevel structure parallelism, including MPI, CPE and SIMD parallel levels , is proposed. The common used interaction potentials such as Lennard-Jones and Dissipative Particle Dynamics are included, moreover, A parallel MD algorithm of Replica Exchange method is designed in cooperated with domain decomposition by MPI parallelism.

Future Directions

Ongoing work:

- To support computation of long-range Coulombic interactions in k space, such as PME, in which CPE of SW26010 CPU is used as k-space calculation accelerator.
- A strategy of CPE dynamic partitioning for computation and IO tasks will be introduced to improve the overall utility of many-core processor.

Contact Information

Zhong Jin, professor
Computer Network Information Center,
Chinese Academy of Science
zjin@sccas.cn
http://english.cnic.cas.cn

References

1. Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Schulten, K. (1999). NAMD2: Greater Scalability for Parallel Molecular Dynamics. Journal of Computational Physics, 151(1), 283–312.
2. Brown W M, Wang P, Plimpton S J, et al. Implementing molecular dynamics on hybrid high performance computers–short range forces [J]. Computer Physics Communications, 2011, 182(4): 898--911.
3. Glaser, J., Nguyen, T. D., Anderson, J. A., Lui, P., Spiga, F., Millan, J. A., ... Glotzer, S. C. (2015). Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Computer Physics Communications*, *192*(July), 97–107.
4. Fu, H., Liao, J., Yang, J., Wang, L., Song, Z., Huang, X., ... Yang, G. (2016). The Sunway TaihuLight supercomputer: system and applications. *Science China Information Sciences*, *59*(7), 72001.
5. Zhang, W., Lin, J., Xu, W., Fu, H., & Yang, G. (2017). *SCStore: Managing scientific computing packages for hybrid system with containers*. *Tsinghua Science and Technology* (Vol. 22).

Acknowledgements

We would like to thank the technical support from National Supercomputing Center in Wuxi. This work was partially supported by the National Key Research and Development Program of China (Grant No. 2016YFB0201700) and the Research and Application of Information Engineering Program of Chinese Academy of Science (Grant No. XXH13506-404).