

Application of the Mole-8.5 supercomputer: Probing the whole influenza virion at the atomic level

XU Ji^{1,2}, WANG XiaoWei¹, HE XianFeng¹, REN Ying^{1*}, GE Wei^{1*} & LI JingHai¹

¹ State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China;

² Graduate University of the Chinese Academy of Sciences, Beijing 100049, China

Received March 18, 2011; accepted April 7, 2011

Discrete computer simulations are quite helpful in understanding dynamic structures in complex systems. Recently, using the Mole-8.5 supercomputer and molecular dynamics simulations as a “computational microscope”, we simulated the dynamic structure of a whole H1N1 influenza virion in solution for the first time at the atomic level. In total, 300 million atoms in a periodic cube with an edge length of 148.5 nm were simulated. Using 288 low level hybrids with 1728 C2050 GPUs and a software package developed specifically for the hardware, the simulation executed 770 ps/d with an integration time step of 1 fs, and analyzed the dynamic structure. With the tremendous computational power of GPUs, efficient software packages for various hardware designs, and consistent physical models, more challenging applications will be carried out in the near future.

influenza virion, Mole-8.5, supercomputer, GPU, molecular dynamics simulation

Citation: Xu J, Wang X W, He X F, et al. Application of the Mole-8.5 supercomputer: Probing the whole influenza virion at the atomic level. Chinese Sci Bull, 2011, 56: 2114–2118, doi: 10.1007/s11434-011-4543-7

During the past few decades, discrete simulation has been used as a “computational microscope” to probe the dynamic structures of complex systems, especially when inspection instruments are insufficient in resolving capability, or operational conditions are difficult to realize in experiments. In chemical engineering, detailed information of individual particles is necessary to understand the heterogeneity structure in both gas-solid and liquid-solid systems. In life science, an understanding of the dynamic structure of the influenza virion *in vivo* is quite useful in preventing influenza epidemics and designing anti-influenza drugs. However, these processes are too fast and too delicate to capture using current experimental facilities. On the other hand, computer simulations of these systems usually involve billions of particles, which is beyond the reach of existing supercomputers. In a wide range of scientific fields where such challenges occur, urgent demands for large-scale simulations are calling for the provision of

more powerful computational capability.

1 Computational facility

A supercomputer called Mole-8.5 with about 2.0 Petaflops peak performance in single precision and 1.0 Petaflops in double precision was formally established in April 2010 [1]. It was designed and implemented by the Institute of Process Engineering (IPE), Chinese Academy of Sciences (CAS), one of the NVIDIA CUDA Centers of Excellence (CCOE) (CUDA, Computed Unified Device Architecture, a prevailing programming infrastructure for general purpose graphic processing units, or GPUs). The Mole-8.5 embodies a design philosophy utilizing the structural similarity between hardware, software, and the problems to be solved [2], based on the multi-scale method and discrete simulation approaches developed at IPE [3–5]. This supercomputer, which includes 362 nodes, is the first GPGPU supercomputer using NVIDIA Tesla C2050 boards in the world. It

*Corresponding authors (email: yren@home.ipe.ac.cn; wge@home.ipe.ac.cn)

contains a multi-level structure, with high level CPU computing nodes, mid-level visualization-computing nodes connected to a 3×6 DID LCD array for online visualization, and low level hybrid computing nodes. The simulation work is mainly carried out by the low level nodes, which use 2 Intel(R) Xeon(R) E5520 2.26 GHz CPUs, 48 GB DDR3 1333 MHz memory, and 6 C2050 GPU boards to accelerate the computation. All the nodes are connected via a Gigabit Ethernet and QDR Infiniband network. If computation of the application running on the Mole-8.5 is fast enough, the results can be displayed online on the display wall of 3×6 DID LCD panels at the same time as the computation, thereby realizing realtime simulation. The general structure of the whole system is given in Figure 1 [1]. The Mole-8.5 has some unique advantages over a HPC system with the same performance based on CPUs; that is, the high performance/price ratio, and the fact that it occupies only about 150 m². Though the system is mainly designed for multi-scale discrete simulation, the Linpack performance of its 320 low level nodes still reaches 207 TFlops, with a power efficiency of 431 Mflop/s/Watt, debuting at No. 8 on the Green500 list in June 2010 (<http://www.green500.org/lists/2010/11/top/list.php?from=1&to=100>), No. 19 on the 35th TOP500 list in June 2010 (<http://www.top500.org/list/2010/06/100>), and No. 3 on the China TOP100 list in October 2010 [6].

With the specifically developed software package for the hardware structure described above, the Mole-8.5 has exhibited significant acceleration compared with traditional CPU-based clusters and has been used to perform large-

scale simulations in chemical engineering, such as a continuum-scale direct numerical simulation of a gas-solid flow system with up to 672 GPUs¹⁾ and a quasi-realtime simulation of an industrial scale rotating drum using the discrete element method with more than 200 GPUs²⁾. Recently, using a molecular dynamics (MD) simulation software package developed from our previous work [7], it has become feasible to probe the dynamic structure of the whole H1N1 virion at the atomic level.

2 Stationary structure

The influenza virion, an enveloped RNA virus belonging to the Orthomyxoviridae family, is a major cause of global infection and mortality. The H1N1 virus, with the numbers denoting the subtypes of hemagglutinin (HA) and neuraminidase (NA), respectively, is a subtype of the influenza A virus and was responsible for the flu pandemic in 2009. Influenza virions are highly pleomorphic, with sizes ranging from spherical particles with a diameter of approximately 100–150 nm to filamentous particles with a length of several μm [8]. As shown in Figure 2, a rough model of the general structure of influenza virions and crystal structures of single macromolecules comprising the virions have been provided by electron microscopes, X-ray diffraction, etc. The influenza A virion is studded with glycoprotein spikes of a large number of HA and NA, at a ratio of approximately four to one, projecting from a lipid membrane [9]. HA is evenly spaced, while NA is located in clusters. The

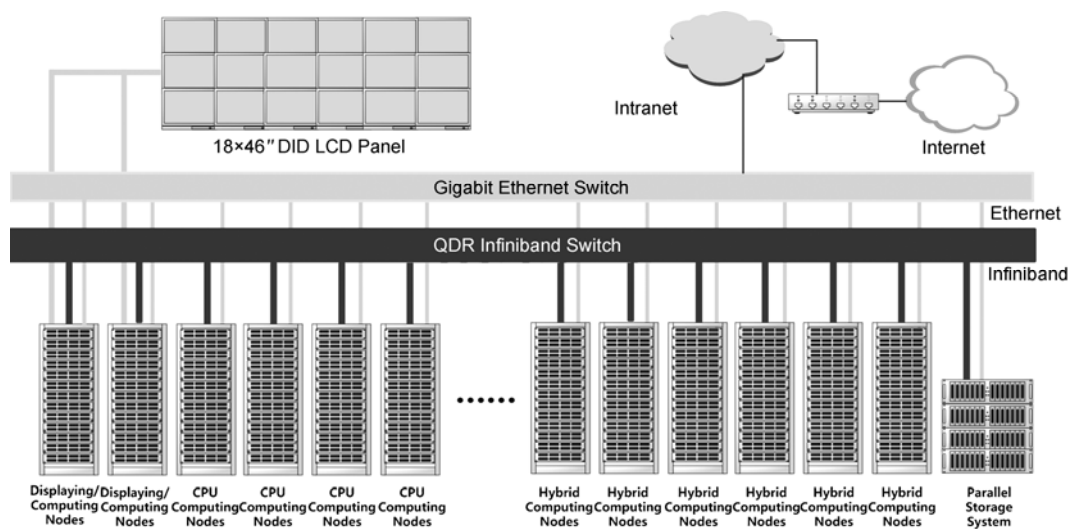


Figure 1 Global architecture of the entire Mole-8.5 system [1].

1) The EMMS group (Principal authors: Wei Ge, Wei Wang, Ning Yang, Jinghai Li, Mooson Kwauk). Meso-scale oriented simulation towards virtual process engineering (VPE)—the EMMS Paradigm. Chem Eng Sci, 2011, doi: 10.1016/j.ces.2011.05.029

2) Xu J, Qi H, Fang X, et al. Quasi-realtime simulation of rotating drum using discrete element method with parallel GPU computing. Accepted by *Particology*, 2011

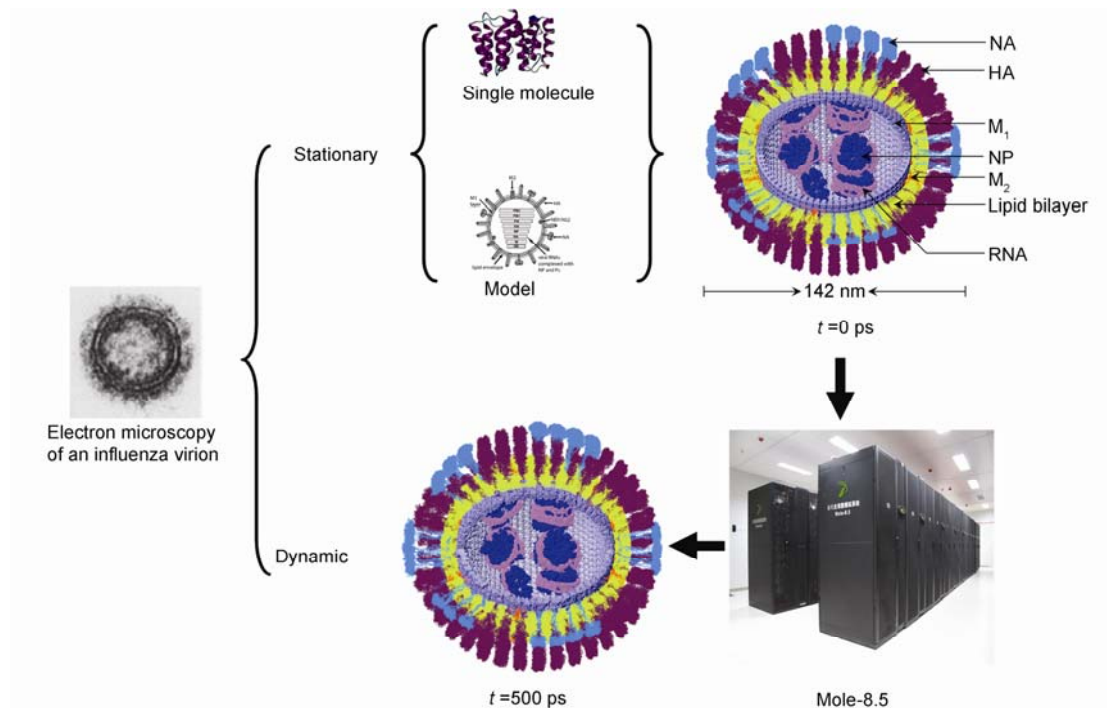


Figure 2 The main procedure for probing H1N1 dynamic structure at the atomic level. Figures of the electron microscopy and the model were taken from ref. [9]. A quarter of the outer sphere is not shown, and components are shown in different colors for better visualization.

membrane also contains the matrix protein M_2 , although in much lower quantities (1:10–100 M_2 :HA) [10]. Matrix protein M_1 , the most abundant component of the virion, lies just beneath the lipid envelope and its three integral membrane proteins, resulting in an enclosure of the virion core [11]. The viral core comprises eight helical viral ribonucleoproteins (vRNPs), each containing a segment of single-stranded negative-sense RNA, multiple copies of nucleoprotein (NP), and three polymerase proteins (PB1, PB2, PA).

The information above has already provided a static picture of the 3D structure of influenza virions, though the model is rough and the structures of some componential molecules are still unknown or not completely known. Computer simulations can be utilized as an important tool to complement the current theoretical and experimental understanding. First, the missing details should be reconstructed to provide an atomic structure of the virion, and then, an explicit solvent MD study should be performed to detect further dynamics of the virion *in vivo*. The globular lipid membrane is made up of a double layer of dipalmitoylphosphatidylcholine (DPPC), resulting in a globular membrane with an external diameter of 106 nm. Three transmembrane proteins, NA, HA, and M_2 are constructed following similar procedures, that is, reconstruction of single macromolecular structure at atomic-scale followed by preparation of a matrix of the macromolecules on a spherical surface. A total of 374 HA and 98 NA molecules are thus located on a spherical sphere with their tails embedded in the lipid membrane as appropriate. Forty-seven proton

channels penetrate the membrane, with each proton channel formed by 4 parallel monomers of M_2 . The spherical protein layer of the M_1 matrix is constructed in the same way as the membrane. A vRNP is constructed by placing NP monomers in a helical structure and locating a packed single-stranded negative-sense RNA strand on the surface of the protein scaffold.

The simulated H1N1 virion is finally constructed with 2363 proteins, 63471 DPPC molecules and 8 RNA strands, with details of the components given in Table 1. Each component has been equilibrated before being assembled into the virion. The complex is then solvated into SPC [12] water with an appropriate concentration of ions, resulting in 300 million atoms in a periodic cube with each side 148.5 nm long. The entire system is equilibrated again, and then simulated at 300 K with a slightly modified GROMOS 53a6 force field [13] against RNA strands. Berendsen's temperature coupling method [14] is adopted in this work.

Table 1 Components of the H1N1 virion

Component	PDB ID	No. of molecules
HA	3LZG	374
NA	2HTY	98
M1	1AA7	1580
M2	2RLF	164
NP	2IQH	147
Lipids	–	63471
RNA	–	8

3 Software

The software package used here is based on GPU_MD-1.0.5 [7], which is the single GPU version software for simulating macromolecules with high efficiency. Parallel simulation on multiple GPUs *via* three-dimensional domain decomposition [15] is implemented to probe the whole virus structure at the atomic level. One domain also needs the particle data of its surrounding space, usually called the “virtual boundary”, besides its own to perform the calculations in CPU_MD-1.0.5. The particle data of the “virtual boundary” are, in turn, constructed along the direction of decomposition [5], i.e., X, Y, Z. All particles reside in the same process in a single GPU computation, while they transfer among the processes in the parallel implementation. After the particles have moved in or out, the original arrays of particles will have “empty spaces” in them. To make full use of the memory bandwidth and avoid vacant memory access, the particles should be rearranged in memory after a given number of simulation steps as some particles move out to other processes, executed by the CPU, and others move in, that is, transfer back to the GPU. The bonded interaction relations are changed after the rearrangement of particles, so the bonded interaction lists should be reconstructed correspondingly. To maximize GPU efficiency, the amount and times of data transfer between GPU and CPU should be minimized. Only the positions of the particles in the boundary space are transferred at each step. Regarding this point, all the particle data needed by one process is constructed and stored locally. Then the computational task in each GPU can be performed in the same way as those of the GPU_MD-1.0.5. The computations are carried out on the low level hybrid computing nodes, the main computational resources of the Mole-8.5, where GPUs deal with simple computational tasks and CPUs deal with the more complex and high level decision-making tasks. With 288 nodes comprising 1,728 C2050 GPUs, the simulation reached a speed of 770 ps/d with an integration time step of 1 fs.

4 Computational results and discussion

Starting from the predefined structure, the virus experiences a significant change to obtain a stable structure. As shown in Figure 3, the 3 components of the radius of gyration (R_{gx} , R_{gy} and R_{gz}), calculated for all the atoms in the lipid membrane and transmembrane proteins, show similar behaviors during the 0.5 ns simulation. The profiles decrease sharply during the first 0.1 ns, suggesting a quick adjustment of the unfavorable structures; then the profiles slowly increase in the next 0.05 ns and remain comparatively stable thereafter. In accordance with the structural changes, the potential energy (Figure 4(a)), including both intra- and inter-molecular Van der Waals interactions and columbic interactions, to-

gether with the solvation energy of the virion (Figure 4(b)), show a significant decrease in the first 0.1 ns, but later become relatively stable.

The plots of the structural changes and energetic properties suggest the virion slowly evolves to a stable structure in the aqueous solution. The hardware and software used in this work serve as nutrients to keep the virion alive. This endeavor may help us to investigate the biological behaviors of the influenza virus, find its weakness, and thus provide valuable knowledge to prevent future influenza epidemics.

5 Conclusions

For a long time, molecular simulations have focused on

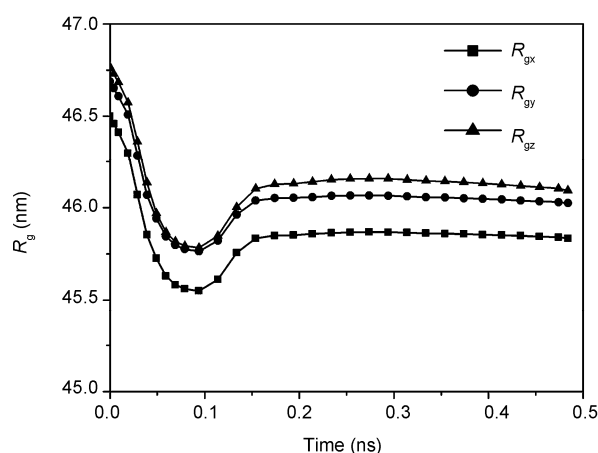


Figure 3 R_{gx} , R_{gy} and R_{gz} of the lipid membrane and transmembrane proteins of a virus as a function of simulation time.

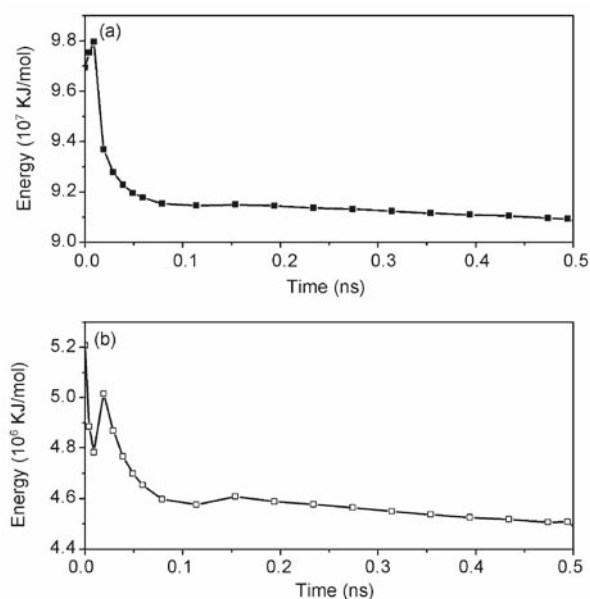


Figure 4 Potential energy (a) and solvation energy (b) of the virion as a function of simulation time.

individual biological molecules. In recent years, the capacity of computational resources has increased to macro-biomolecular complexes. Now, it is encouraging to see that the new era of simulating biological agents has arrived. With the tremendous computational power of GPUs, efficient software packages for various hardware designs, and consistent physical models, more challenging applications will be carried out in the near future.

This work was supported by Ministry of Finance (ZDYZ2008-2), the National Key Science and Technology Project (2008ZX05014-003-006HZ), the Knowledge Innovation Project of Chinese Academy of Sciences (KGCX2-YW-124), and the National Natural Science Foundation of China (20821092).

- 1 Wang X, Ge W, He X, et al. Development and application of a HPC system for multi-scale discrete simulation-Mole-8.5. International Supercomputing Conference, Germany, 2010
- 2 Chen F, Ge W, Guo L, et al. Multi-scale HPC system for multi-scale discrete simulation—Development and application of a supercomputer with 1 Petaflops peak performance in single precision. *Particuology*, 2009, 7: 332–335
- 3 Ge W, Li J. General approach for discrete simulation of complex systems. *Chinese Sci Bull*, 2002, 47: 1172–1175
- 4 Tang D. A general method of parallel computation of particle method and its preliminary applications. Doctor Dissertation. Beijing: Chinese Academy of Sciences, 2005
- 5 Wang X. A framework for parallel simulation of particle systems with pairadditive local interactions—toward a general approach. Doctor Dissertation. Beijing: Chinese Academy of Sciences, 2004
- 6 Zhang Y, Sun J, Yuan G, et al. Perspectives of China's HPC system development: A view from the 2009 China HPC TOP100 list. *Front Comput Sci China*, 2010, 4: 437–444
- 7 Xu J, Ren Y, Ge W, et al. Molecular dynamics simulation of macromolecules using graphics processing unit. *Mol Simulat*, 2010, 36: 1131–1140
- 8 Roberts P C, Compans R W. Host cell dependence of viral morphology. *Proc Natl Acad Sci USA*, 1998, 95: 5746–5751
- 9 Lamb R A, Krug R M. Orthomyxoviridae: The viruses and their replication. In: Knipe D M, Howley P M, eds. *Field's Virology*. Philadelphia: Lippincott-Raven Publishers, 2001. 1487–1531
- 10 Zebedee S L, Lamb R A. Influenza A virus M2 protein: Monoclonal antibody restriction of virus growth and detection of M₂ in virions. *J Virol*, 1988, 62: 2762–2772
- 11 Yasuda J, Bucher D J, Ishihama A. Growth control of influenza A virus by M₁ protein: Analysis of transfectant viruses carrying the chimeric M gene. *J Virol*, 1994, 68: 8141–8146
- 12 Berendsen H J C, Postma J P M, Van Gunsteren W F, et al. Interaction models for water in relation to protein hydration. *Intermol Forces*, 1981, 11(Suppl 1): 331–338
- 13 Oostenbrink C, Villa A, Mark A E, et al. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem*, 2004, 25: 1656–1676
- 14 Berendsen H J C, Postma J P M, Van Gunsteren W F, et al. Molecular dynamics with coupling to an external bath. *J Chem Phys*, 1984, 81: 3684–3690
- 15 Shaw D E. A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions. *J Comput Chem*, 2005, 26: 1803

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.