GTC Asia
Dec 14, 2011
Beijing, China

Large-scale Molecular Simulations of Bio-systems using GPUs

Ying Ren

The EMMS Group
State Key Laboratory of Multiphase Complex Systems (MPCS)
Institute of Process Engineering (IPE)
Chinese Academy of Sciences (CAS)

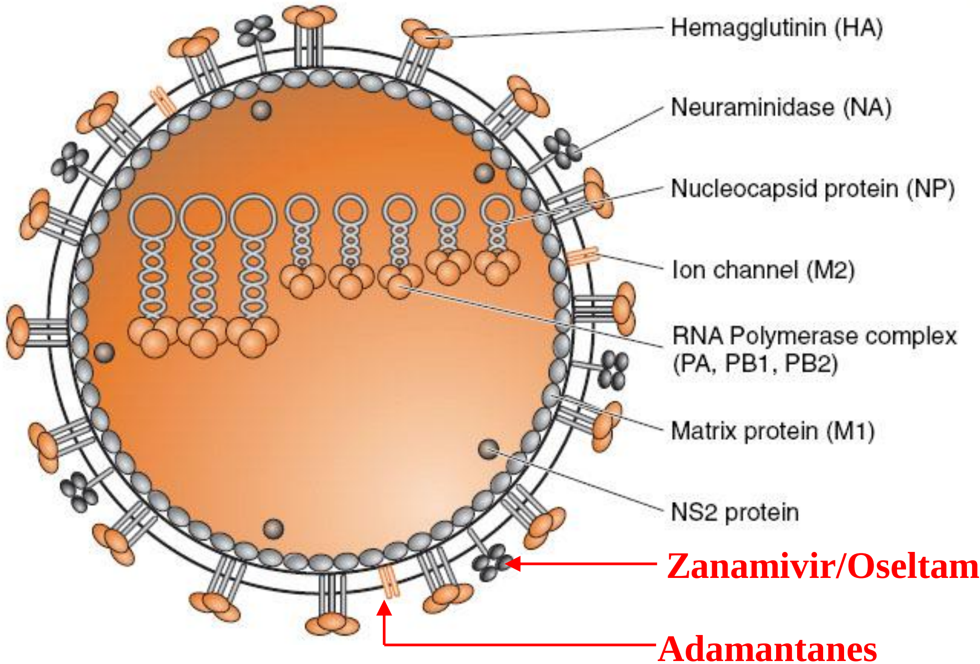
EMMS Group

Challenge

- Introduction to Influenza virus
- Challenge 1: Construction of static structure
- Challenge 2: MD simulation

EMMS Group

Influenza virus: *background*

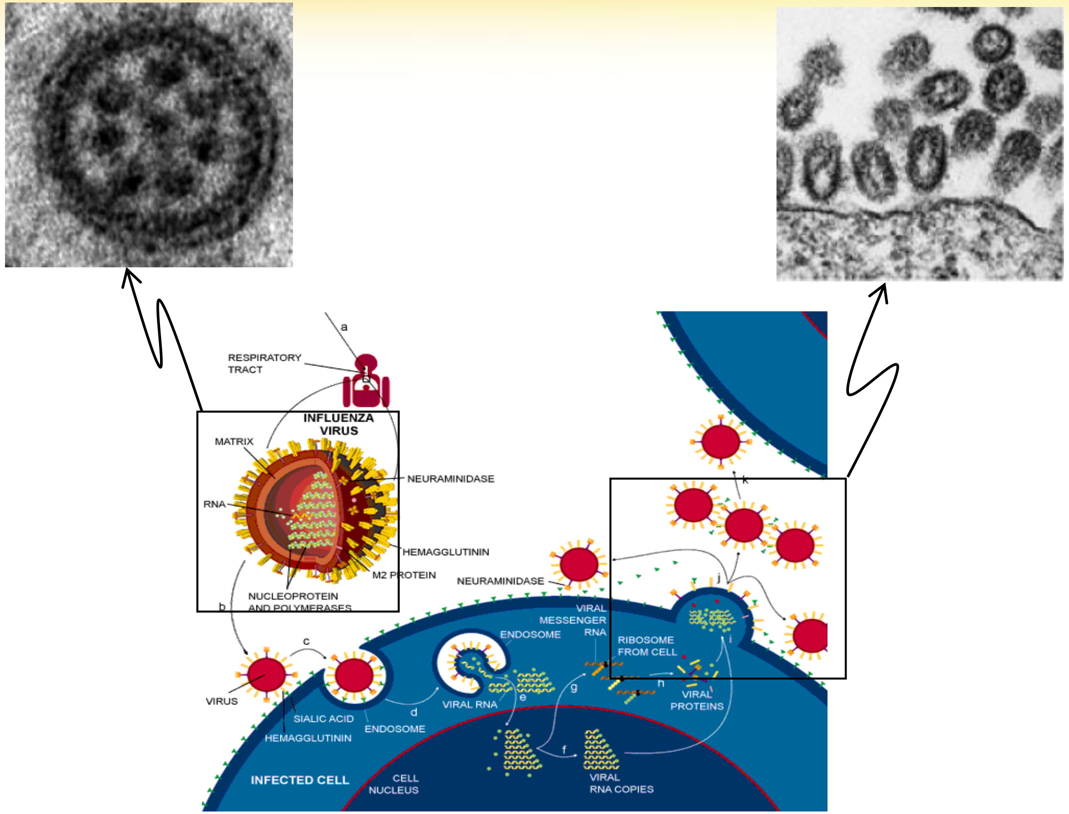


Zanamivir/Oseltamivir

Adamantanes

- Kills half a million people per year
- Particle level V.S. molecular level
- 80-200nm in diameter
- Mostly round
- 8 strands of RNA surrounded by NP
- Matrix: M1
- Membrane: embedded with HA,NA and M2

Influenza virus: *particle level*



Experimental understanding:

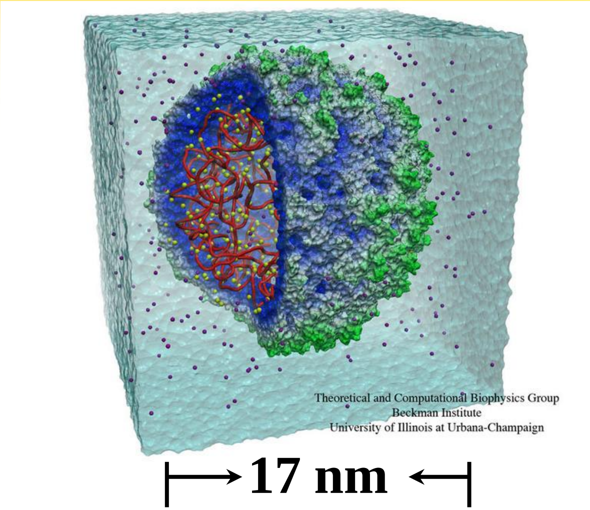
- 80-200nm in diameter
- Mostly round
- Lots of spikes on the surface
- 8 RNAs inside

⬆

Lifecycle of flu virus:

1. Enter cell membrane
2. Reproduce genetic material
3. Release new virus

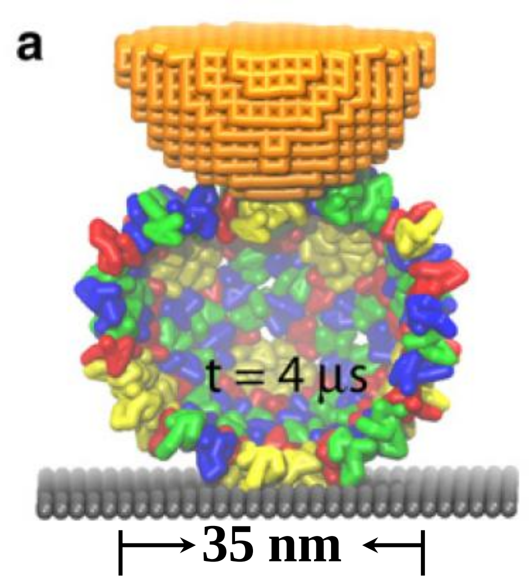
Influenza virus: *particle level*



STM Virus

RNA+capsid, ~1 million atoms
NAMD, Charmm force field, TIP3P water
128(256) Altix nodes at National Center for Supercomputing, 0.7 (1.1) ns/day

Structure, 2006(14): 437



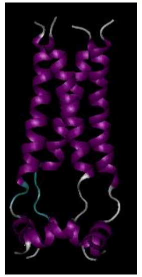
HB Virus

Capsid, 3600 beads for virus
NAMD, 1 CG bead/150 protein atoms

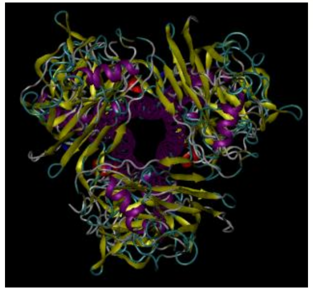
Biophys. J., 2009(97): 2061

Influenza virus: *molecular level*

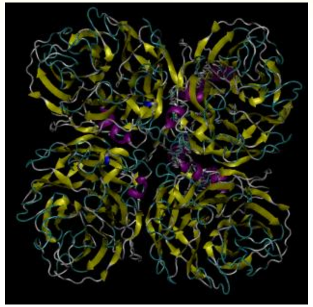
M2



HA

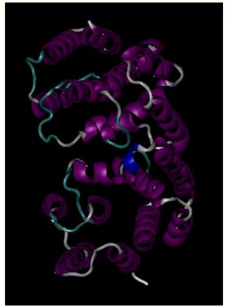


NA

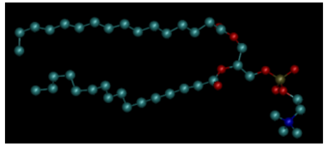


embedded in membrane

M1



envelope



dppc

membrane

Search for [ ] as complete name ☑

Display 3 levels using filter: none

Influenza A virus (A/Guangdong/SB1/2009(H1N1))

Taxonomy ID: 796093
Inherited blast name: **viruses**
Rank: no rank
Genetic code: Translation table 1 (Standard)

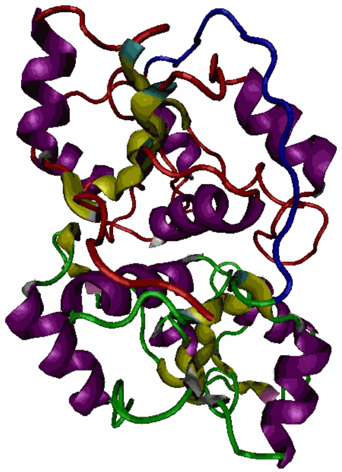
RNA sequence

NP

RNP complex

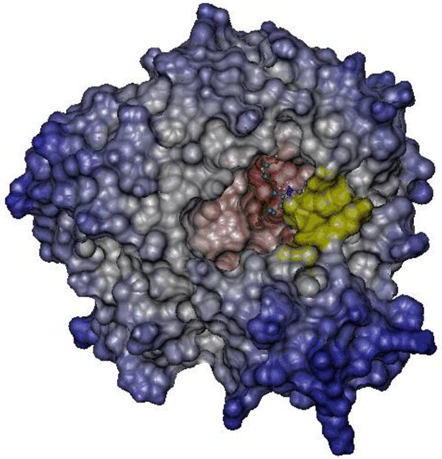
Influenza virus: *molecular level*

Fundamental research



Protein unfolding at high temperature

Drug development



Zanamivir binding to NA

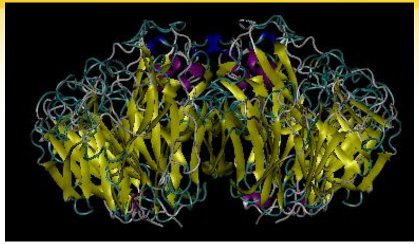
Can we study the dynamic structure of the whole virus particle at atomic level?

- Other factors affecting its life cycle
- New potential target for drug
- …

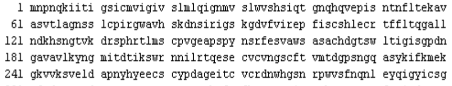
Challenge1：How to build a static H1N1 particle?

- **Append trans-membrane part of HA/NA…**
- **Construct helical structure of RNAs**
- **Coat RNAs with NPs**
- **Assemble all the components into a delicate virion**

Influenza virus: *construction*

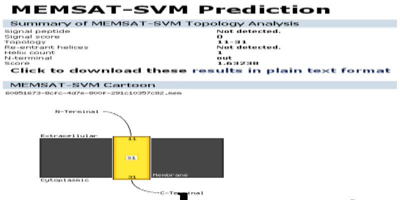


PDB: 386 residue



```
  1 mnpnqkiiti gsicmvigiv slmlqignmv slwvshsiqt gnqhqvepis ntnfltekav
 61 asvtlagnss lcpirgwavh skdnsirigs kgdvfvirep fiscshlecr tffltqgall
121 ndkhsngtvk drsphrtlms cpvgeapspy nsrfesvaws asachdgtsw ltigisgpdn
181 gavavlkyng mitdtikswr nnilrtqese cvcvngscft vmtdgpsngq asykifkmek
241 gkvvksveld apnyhyeecs cypdageitc vcrdnwhgsn rpwvsfnqnl eyqigyicsg
301 vfgdnprpnd gtgscgpvsp ngaygvkgfs fkygngvwig rtkspssrsg femiwdpngw
361 tetdssfsvk qdivaitdws gysgsfvqhp eltgldcirp cfwvelirgr pkestiwtsg
421 ssisfcgvns dtvswswpdg aelpftidk
```

NCBI: sequence of 449 residues



MEMSAT-SVM Prediction

Summary of MEMSAT-SVM Topology Analysis

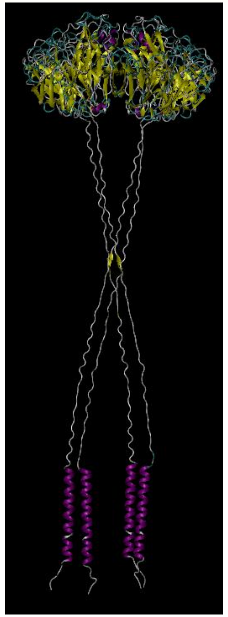
Signal peptide	Not detected.
Signal score	0
Topology	11-31
Re-entrant helices	Not detected.
Helix count	1
N-terminal	out
Score	1.63238

Click to download these results in plain text format

MEMSAT-SVM Cartoon

60851673-8cfc-4d7e-800f-291c10357c82.mem

Trans-membrane structure prediction



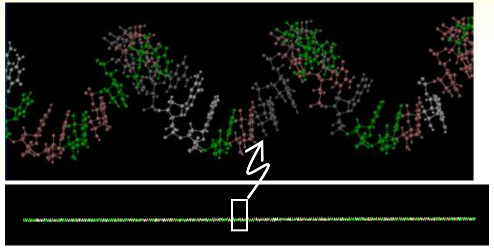
Molecular model of NA

MD simulation in solution to get a relaxed structure

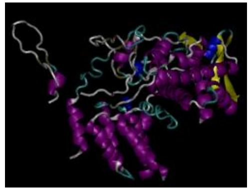
Virus particle

Influenza virus: *construction*

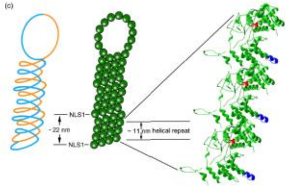
NCBI: sequence of RNA1
2366 necleotides



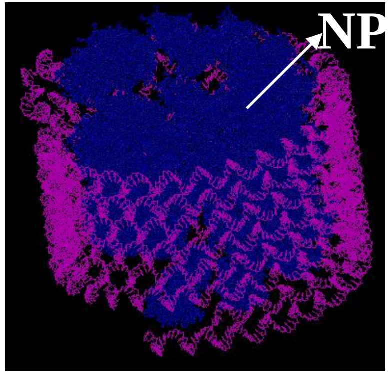
Linear structure



NP



J. Mol. Biol. 2007(374): 910–916



RNP1

MD simulation to relax

Virus particle

Influenza virus: *construction*

Pack 8 RNPs closely to form the interior of the virus

Coat RNPs with a sphere of M1 matrix

Coat M1 matrix with a sphere of membrane

$N_{HA}/N_{NA} \approx 4$

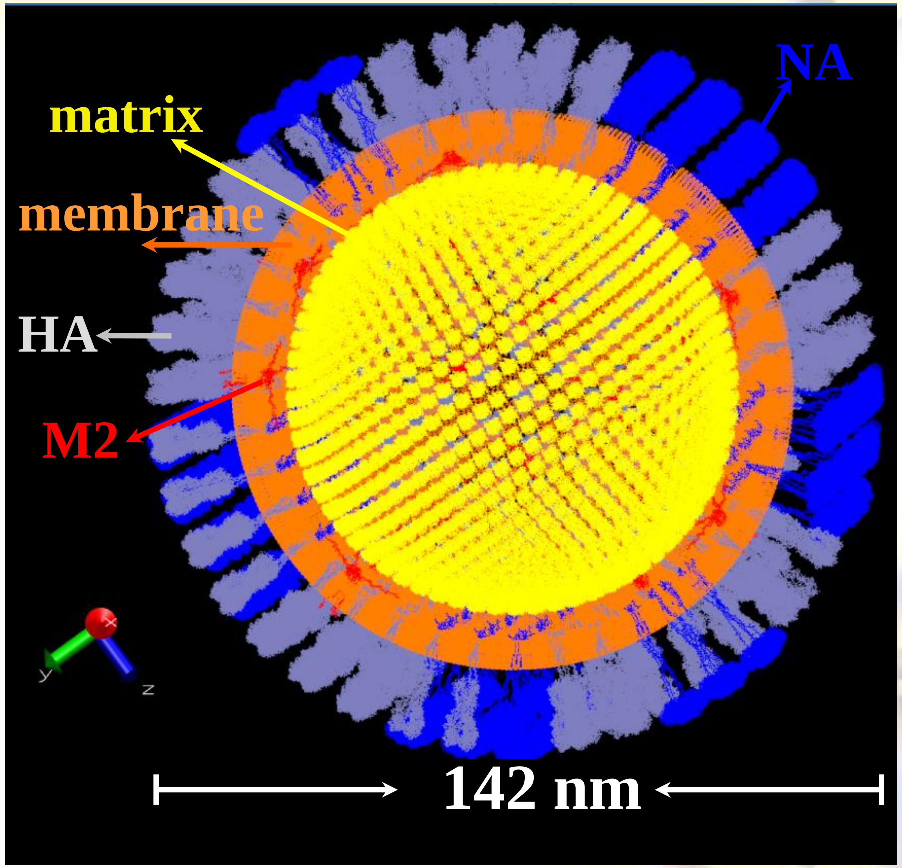
HA and NA in cluster

M2 randomly distributed

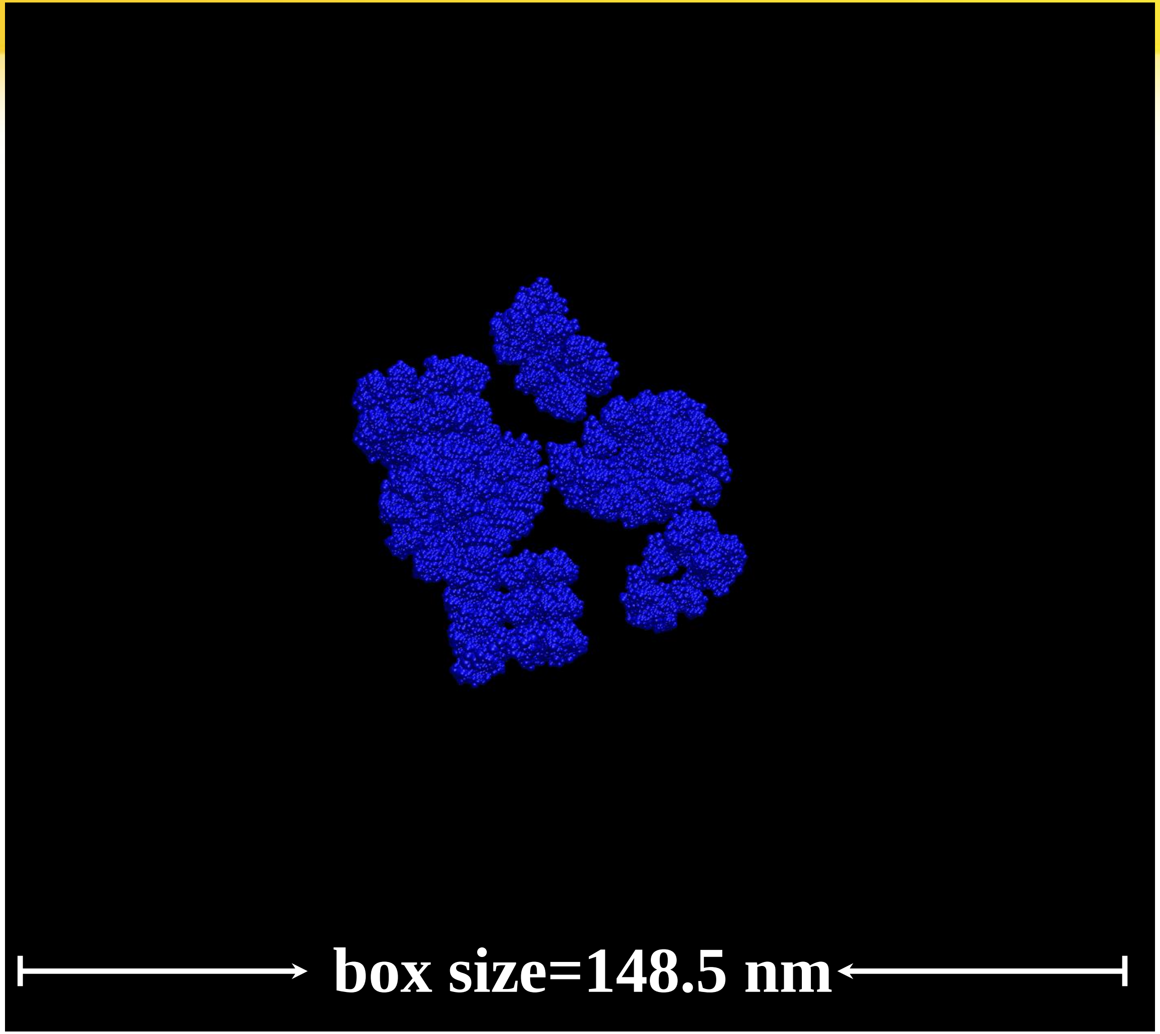
Embed trans-membrane proteins into membrane

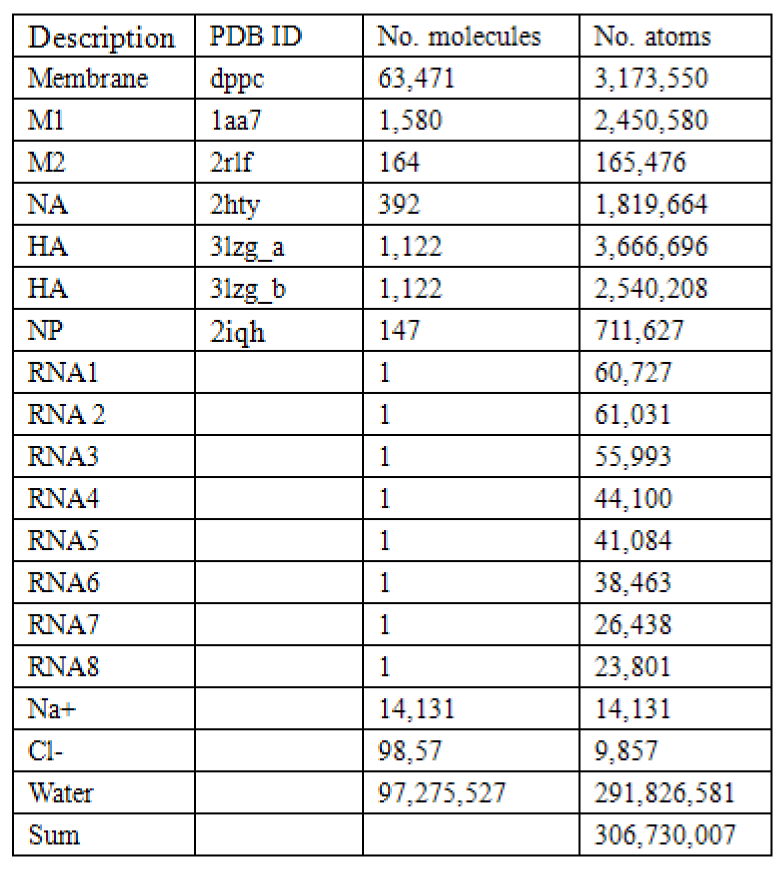
Fill the rest space with water and ions

Box size=148.5 nm



Influenza virus: *construction*





Description	PDB ID	No. molecules	No. atoms
Membrane	dppc	63,471	3,173,550
M1	1aa7	1,580	2,450,580
M2	2rlf	164	165,476
NA	2hty	392	1,819,664
HA	3lzg_a	1,122	3,666,696
HA	3lzg_b	1,122	2,540,208
NP	2iqh	147	711,627
RNA1		1	60,727
RNA 2		1	61,031
RNA3		1	55,993
RNA4		1	44,100
RNA5		1	41,084
RNA6		1	38,463
RNA7		1	26,438
RNA8		1	23,801
Na+		14,131	14,131
Cl-		98,57	9,857
Water		97,275,527	291,826,581
Sum			306,730,007

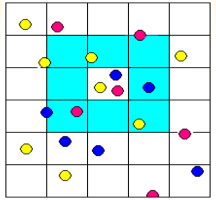
Challenge2：How to simulate such a large system?

- **Which software to use?**
- **Parameter preparation for the simulation**
- **Optimization of parallelization**

EMMS Group

Influenza virus: *simulation*

Neighbor search



Newtonian equation

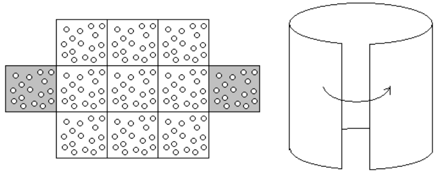
$$F = ma$$

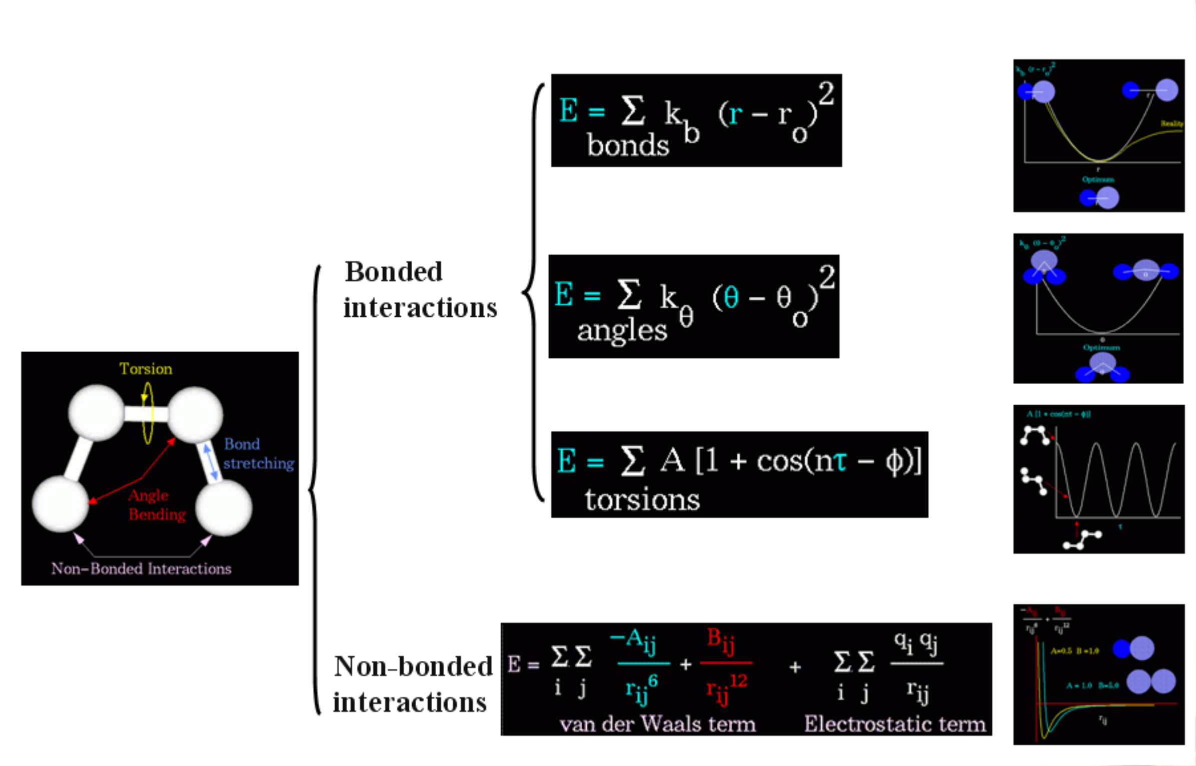
Intergration

$$rv_{t+1} = rv_t + ra_t \times \Delta t$$

$$r_{t+1} = r_t + rv_t \times \Delta t$$

Periodic boundary





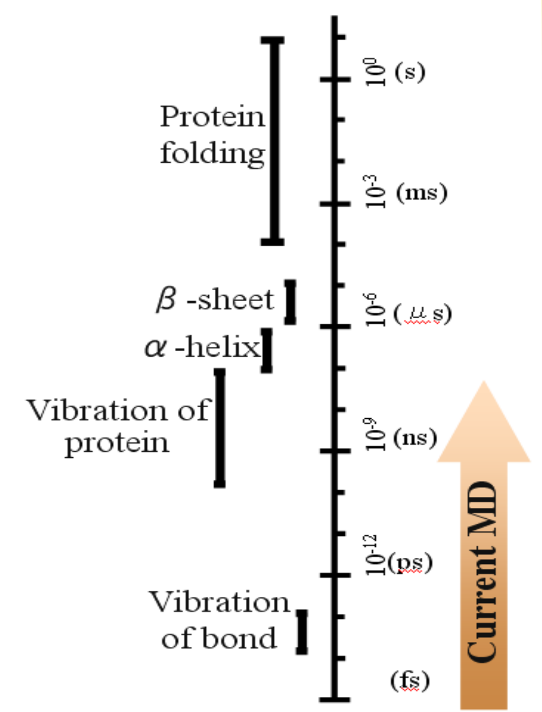
$$E = \sum_{\text{bonds}} k_b (r - r_o)^2$$

$$E = \sum_{\text{angles}} k_\theta (\theta - \theta_o)^2$$

$$E = \sum_{\text{torsions}} A\,[1 + \cos(n\tau - \phi)]$$

$$E = \sum_i \sum_j \frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \sum_i \sum_j \frac{q_i q_j}{r_{ij}}$$

Influenza virus: *simulation*



Huge gap between MD simulation and experiments

Mole-8.5

2.0 Petaflops (single precision)

GPGPU supercomputer, Tesla C2050

No. 21 on 38th Top500 list，No. 9 on Green500 list

Influenza virus: *simulation*

Gromacs (Netherlands)

- GMX-4.5(Sept. 2010)
- Use OpenMM library
- Only for single GPU
- Cannot simulate millions of atoms

NAMD (U.S)

- Only use GPU for nonbonded force
- Other parts are performed on CPU.

HOOMD (U.S)

- Current version: Dihedral and PPPM on GPU
- No force field for biomolecules.
- Not MPI enabled, only run on single node.

ACEMD (Acellera Ltd , commercial)

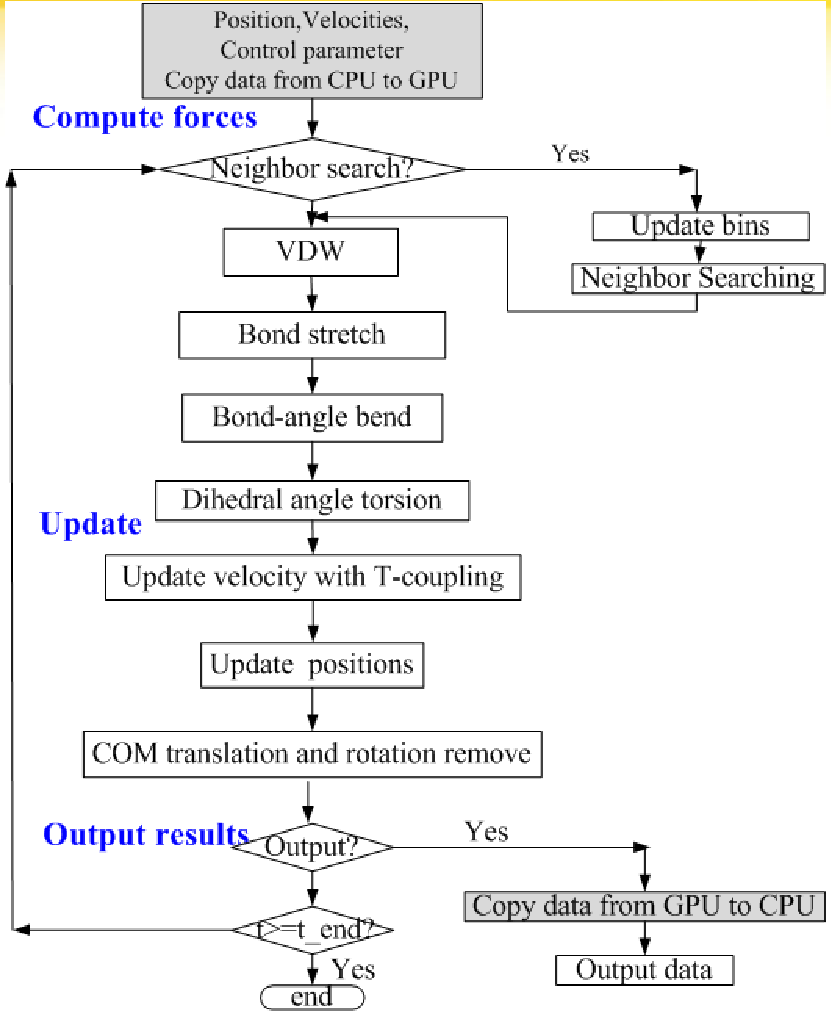
Can run on multiple nodes.

EMMS Group

GPU Solution

- MD implemented on single GPU
- MD implemented on multiple GPUs
- Virus simulation using 1728 GPUs

MD on Single GPU



- ■MD procedure fully implemented on GPU
- ■PME Electrostatics
- ■Constraints: LINCS, SETTLE
- ■Virtual site

Xu, J., Ren, Y., Ge, W., Yu, X., Yang, X. & Li, J. *Molecular Simulation*, 2010 , 36(14):1131

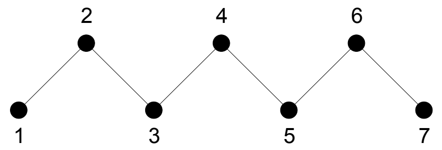
MD on Single GPU

Neighbor list generation

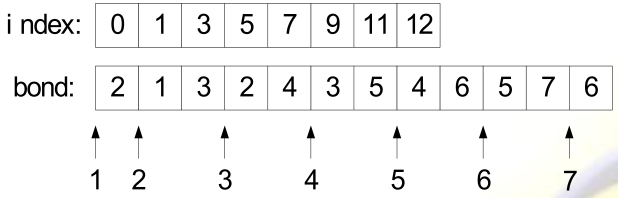
- **Assign one atom to one thread**
- **2D array in global GPU memory to store the neighbor list for each atom**
- **Neighbor search: one thread block corresponds to one bin, and one thread corresponds to one atom in the bin.**

Bonded interactions

- **Similar implementations for bond/angle/dihedral**
- **Copy to global GPU memory during initialization**
- **A thread computes all the bonds of the home atom**



Molecular model



i ndex:	0	1	3	5	7	9	11	12				
bond:	2	1	3	2	4	3	5	4	6	5	7	6

Contents of index and bond arrays in t_bond

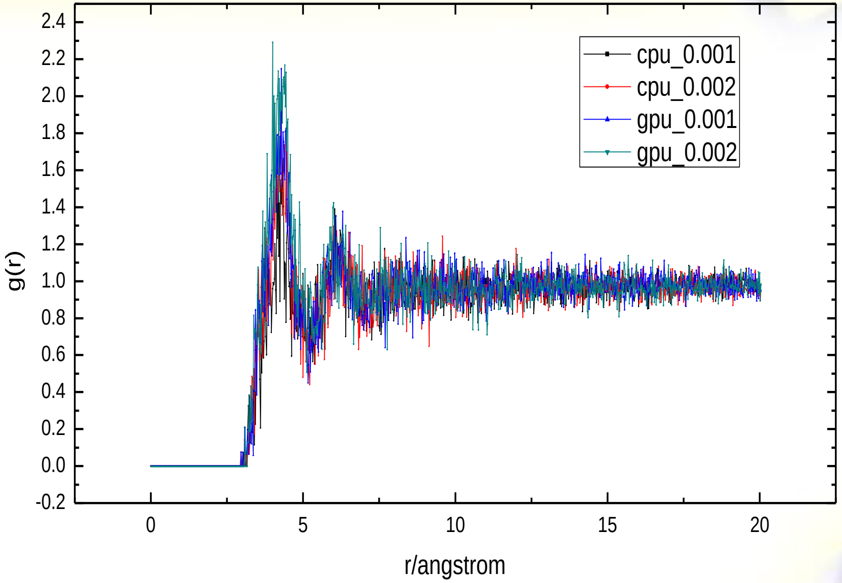
Xu, J., Ren, Y., Ge, W., Yu, X., Yang, X. & Li, J. *Molecular Simulation*, 2010 , 36(14):1131-1140.

MD on Single GPU: *PME*

$$E = E_{dir} + E_{rec} + E_{corr}$$

$$\tilde{E}_{rec} = \frac{1}{2\pi V}\sum_{m\neq 0}\frac{\exp(-\pi^2 m^2/\beta^2)}{m^2} B(m_1,m_2,m_3)\bullet F(Q)(m_1,m_2,m_3)F(Q)(-m_1,-m_2,-m_3)$$

1. **Set parameters for PME.**
2. **Distribute charged atoms onto FFT grids**
3. **FFT transformation of FFT grids**
4. **Solve the equation of energy**
5. **Inverse FFT**
6. **Calculate electro-force**

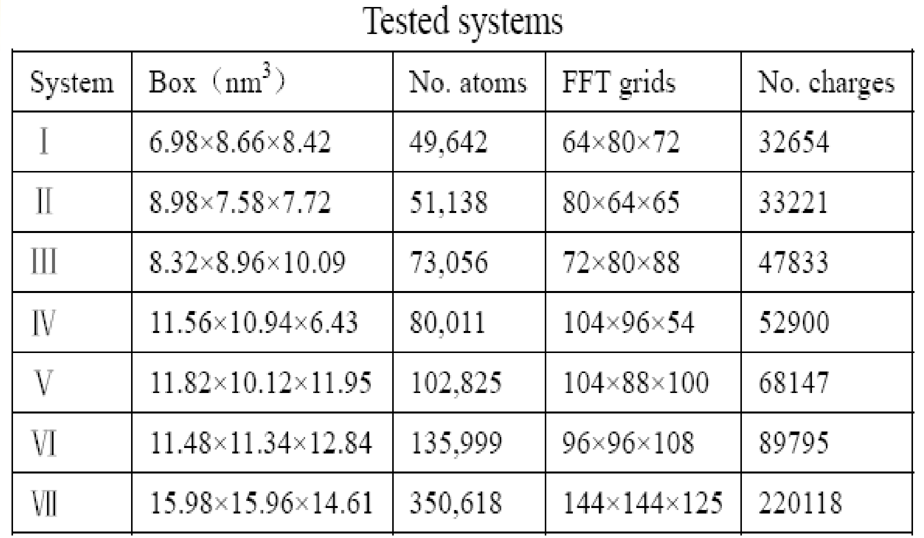


Radial distribution of Na+

1mol/L NaCl solution：40 Na+,40 Cl-,2146 water
NPT ：298K， 1 atm

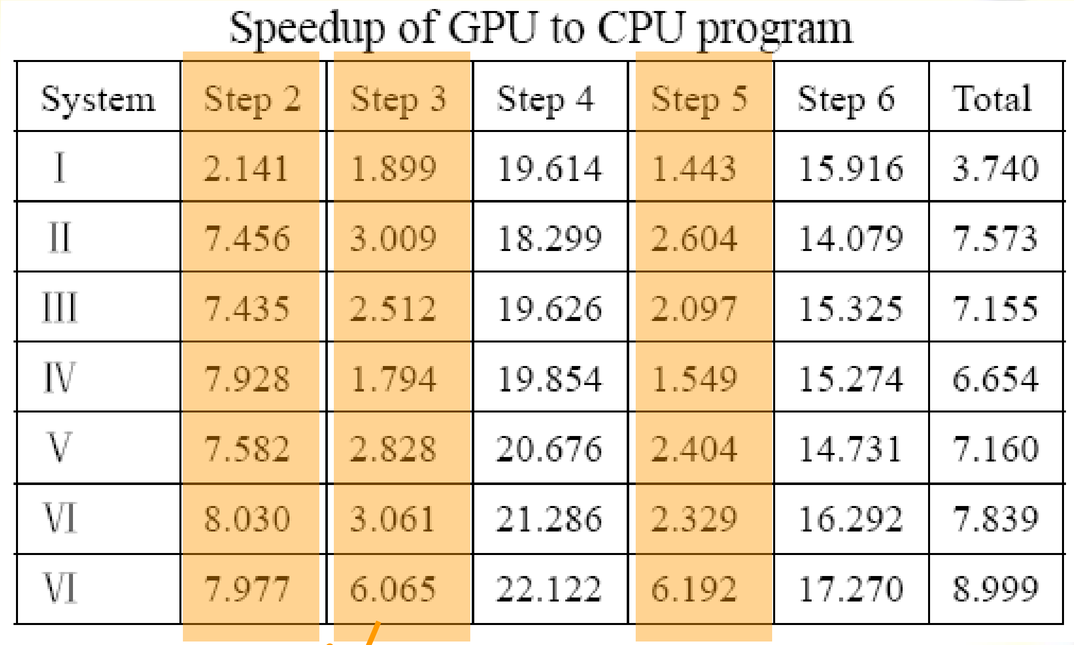
Xu, J., Ge, W., Ren, Y., Li, J. *Chinese Journal of Computational Physics*, 2010 , 27(4):548-554.

MD on Single GPU: *PME*



Tested systems

System	Box（nm^3）	No. atoms	FFT grids	No. charges
Ⅰ	6.98×8.66×8.42	49,642	64×80×72	32654
Ⅱ	8.98×7.58×7.72	51,138	80×64×65	33221
Ⅲ	8.32×8.96×10.09	73,056	72×80×88	47833
Ⅳ	11.56×10.94×6.43	80,011	104×96×54	52900
Ⅴ	11.82×10.12×11.95	102,825	104×88×100	68147
Ⅵ	11.48×11.34×12.84	135,999	96×96×108	89795
Ⅶ	15.98×15.96×14.61	350,618	144×144×125	220118



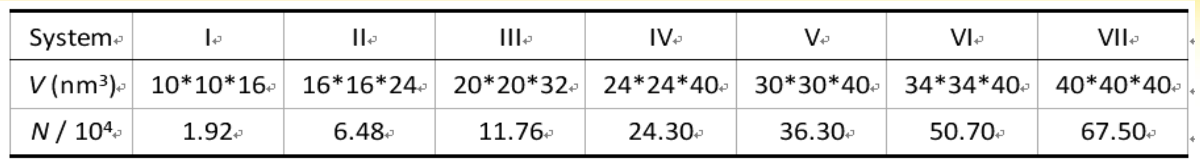
Speedup of GPU to CPU program

System	Step 2	Step 3	Step 4	Step 5	Step 6	Total
Ⅰ	2.141	1.899	19.614	1.443	15.916	3.740
Ⅱ	7.456	3.009	18.299	2.604	14.079	7.573
Ⅲ	7.435	2.512	19.626	2.097	15.325	7.155
Ⅳ	7.928	1.794	19.854	1.549	15.274	6.654
Ⅴ	7.582	2.828	20.676	2.404	14.731	7.160
Ⅵ	8.030	3.061	21.286	2.329	16.292	7.839
Ⅵ	7.977	6.065	22.122	6.192	17.270	8.999

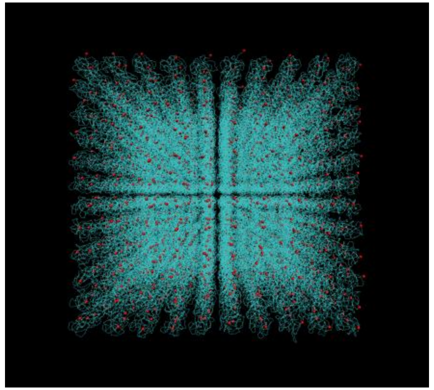
- Frequent data access to memory
- Dependence of data among GPU threads

Xu, J., Ge, W., Ren, Y., Li, J. *Chinese Journal of Computational Physics*, 2010 , 27(4):548-554.

MD on Single GPU: *performance*



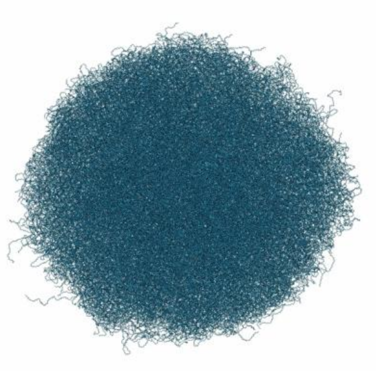
System	I	II	III	IV	V	VI	VII
V (nm^3)	10*10*16	16*16*24	20*20*32	24*24*40	30*30*40	34*34*40	40*40*40
N / 10^4	1.92	6.48	11.76	24.30	36.30	50.70	67.50



Entanglement

T=1000 K
1200 PE
300 beads per chain

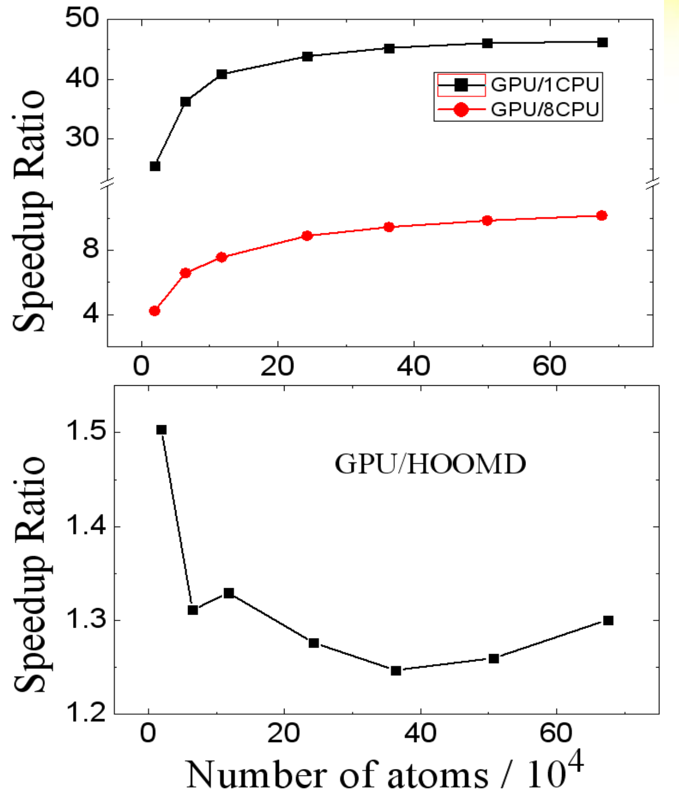
Lower T



Crystallization

T=600 K

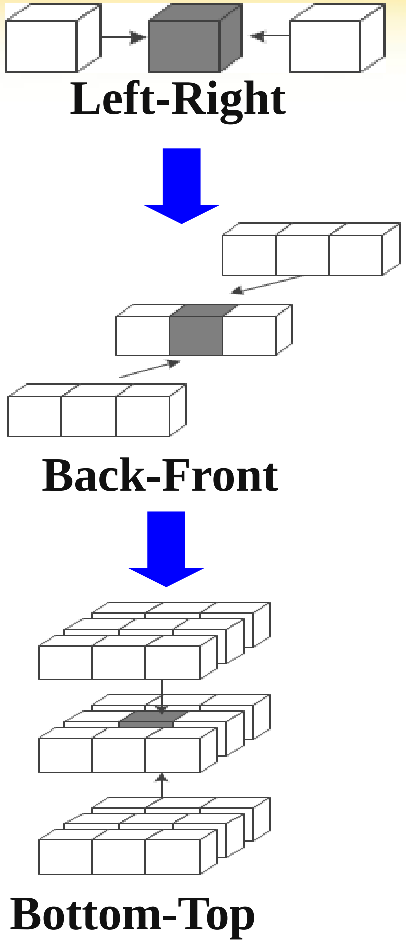
1.57ns/day on 1 GPU



Speedup ratio of GPU _MD-1.0.5 over GROMACS-4.0.5 on CPU(s) and HOOMD-0.9.0 on GPU

Xu, J., Ren, Y., Ge, W., Yu, X., Yang, X. & Li, J. *Molecular Simulation*, 2010 , 36(14):1134-1140.

GPU MD parallelization: *domain decomposition*



- **Normal method:**

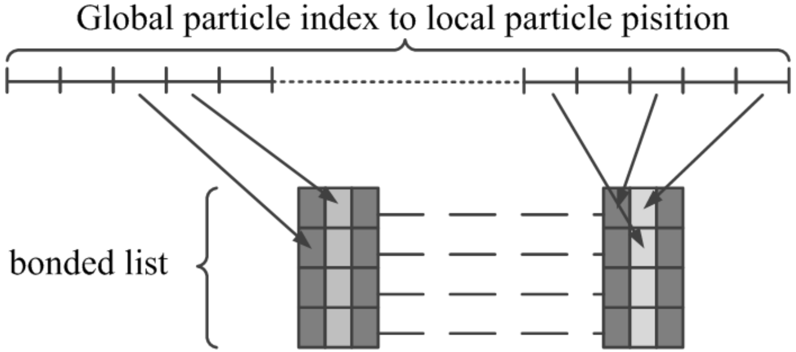
  1. Communicate position forward

  2. Communicate force backward
- **To reduce the memory copy times between GPU and CPU, one domain needs the positions of particles in neighboring domains, and can compute the interaction of the particles.**
- **No need to communicate force backward.**

GPU MD parallelization: *bonded list*

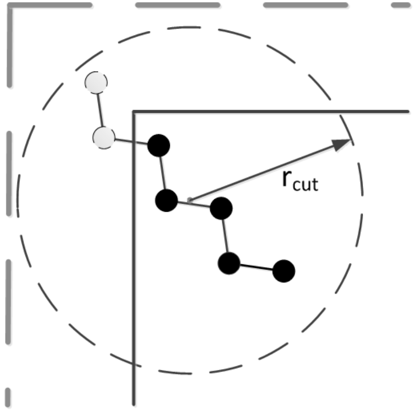
Set

- A huge 1D array to store the memory position of particles.
- **Advantage**: Fast
- **Disadvantage**: use more memory. System is too large, i.e. 0.3 billion particles — more than 1 GB, can not use it.



Search

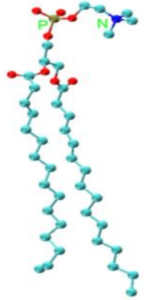
- Search the neighbor particles to generate the bond list
- **Advantage**: no use of memory
- **Disadvantage**: Slow



GPU MD parallelization: *performance*

Group1 **8 GPUs, $N_{DPPC}:N_{SPC}$ = 1:85**

Case	1	2	3	4	5	6	7
DPPC	256	864	2048	4000	6912	10976	16384
All(10^6)	0.08	0.26	0.62	1.22	2.11	3.35	5.00



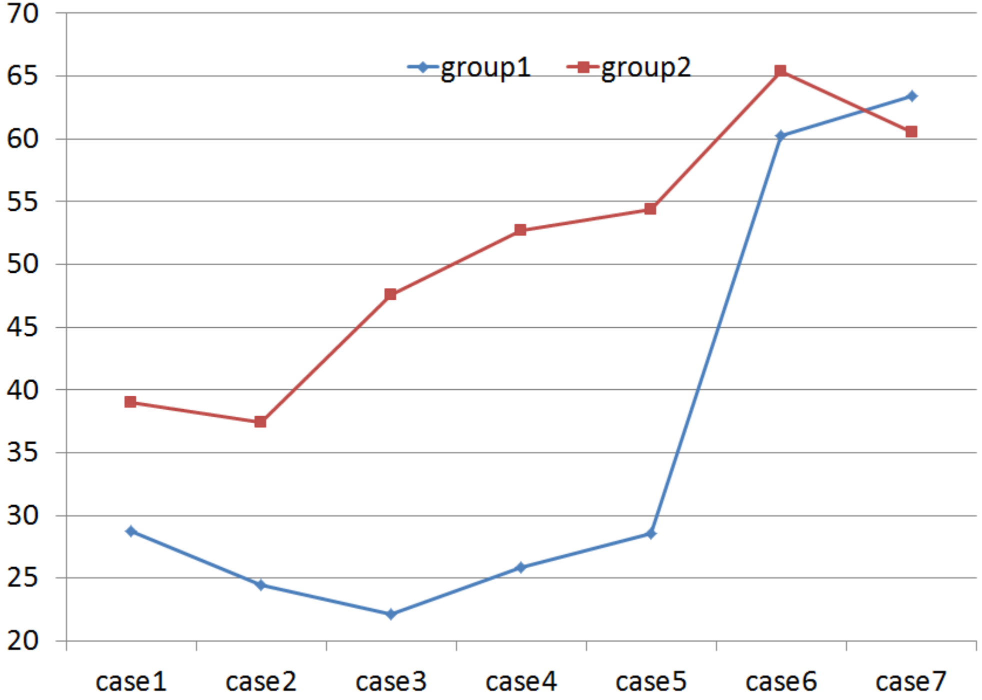
DPPC in water

Group2 **8-64 GPUs, $N_{DPPC}:N_{SPC}$ = 1:19**

Case	1	2	3	4	5	6	7	8	9	10
DPPC(10^3)	0.56	1.89	4.48	8.75	15.12	24.01	35.84	71.68	143.36	286.72
All(10^6)	0.06	0.20	0.48	0.94	1.62	2.57	3.84	7.68	15.36	30.7
No. GPUs	8	8	8	8	8	8	8	16	32	64

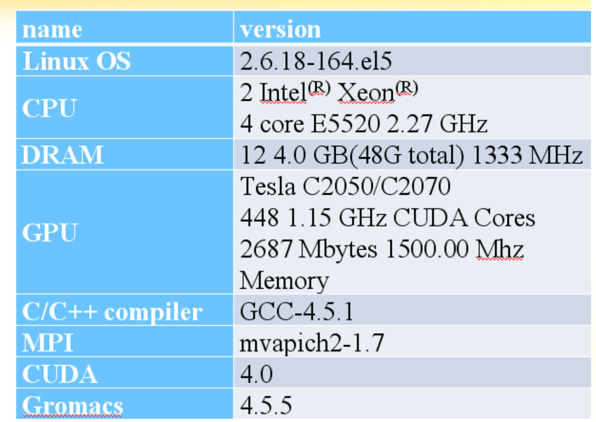
GPU MD parallelization: *performance*

If possible, we propose set method to build the bonded list.



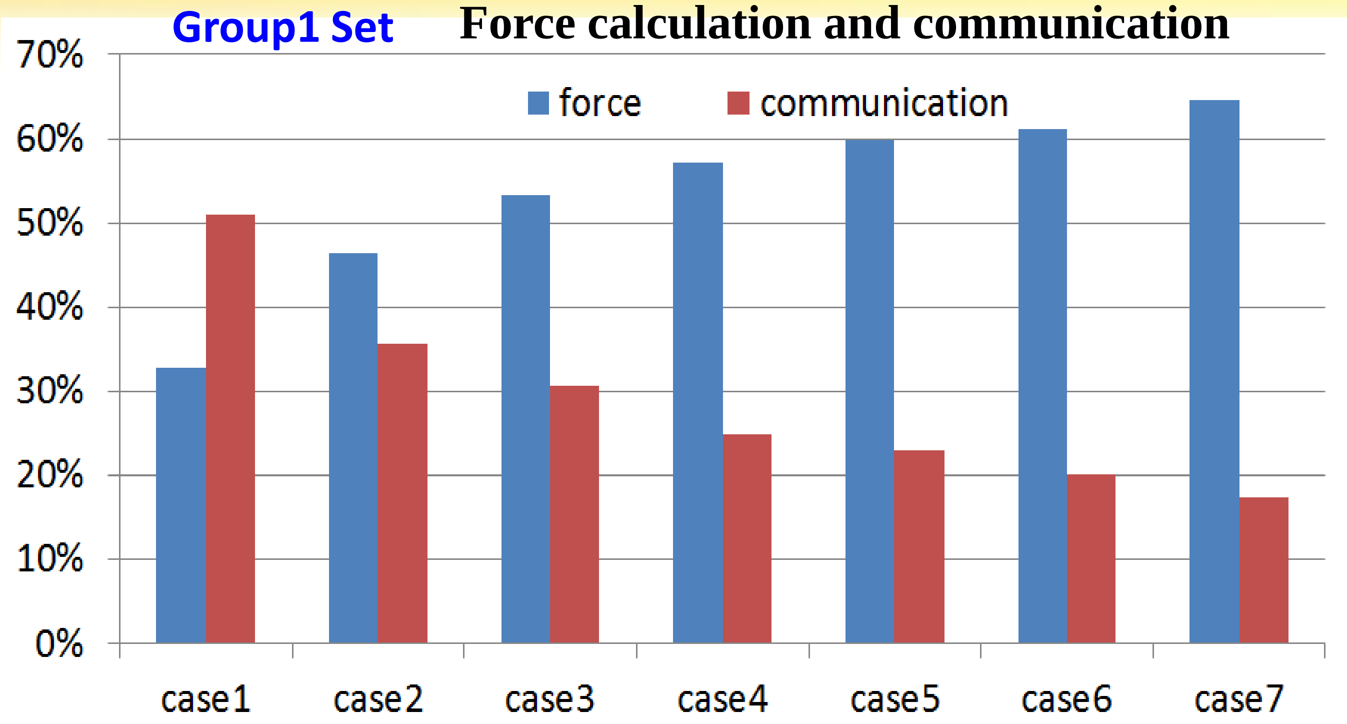
Speed up ratio of bond list generation set/search

Test environment



name	version
Linux OS	2.6.18-164.el5
CPU	2 Intel(R) Xeon(R) 4 core E5520 2.27 GHz
DRAM	12 4.0 GB(48G total) 1333 MHz
GPU	Tesla C2050/C2070 448 1.15 GHz CUDA Cores 2687 Mbytes 1500.00 Mhz Memory
C/C++ compiler	GCC-4.5.1
MPI	mvapich2-1.7
CUDA	4.0
Gromacs	4.5.5

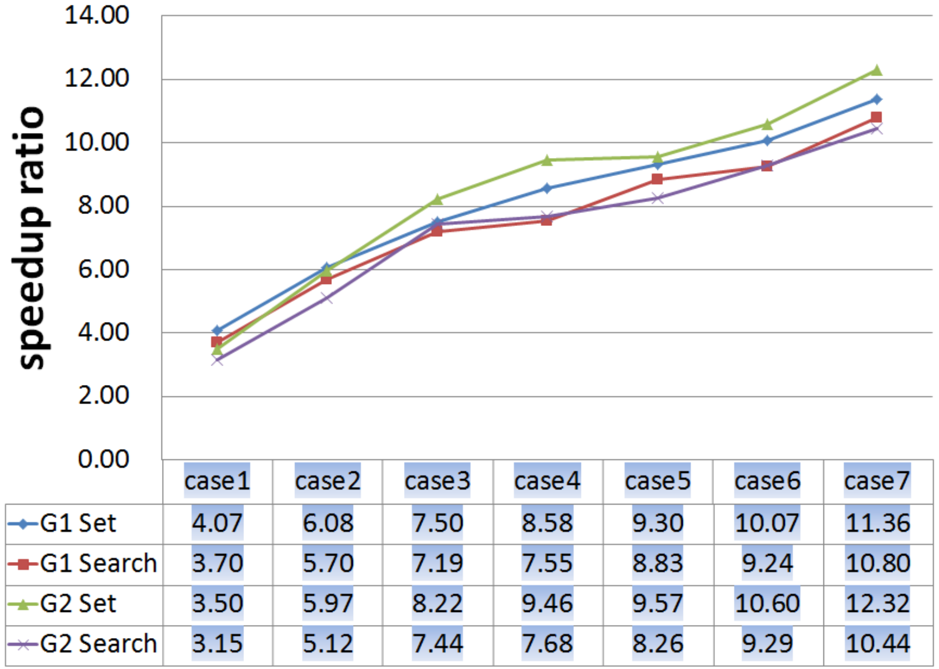
GPU MD parallelization: *performance*



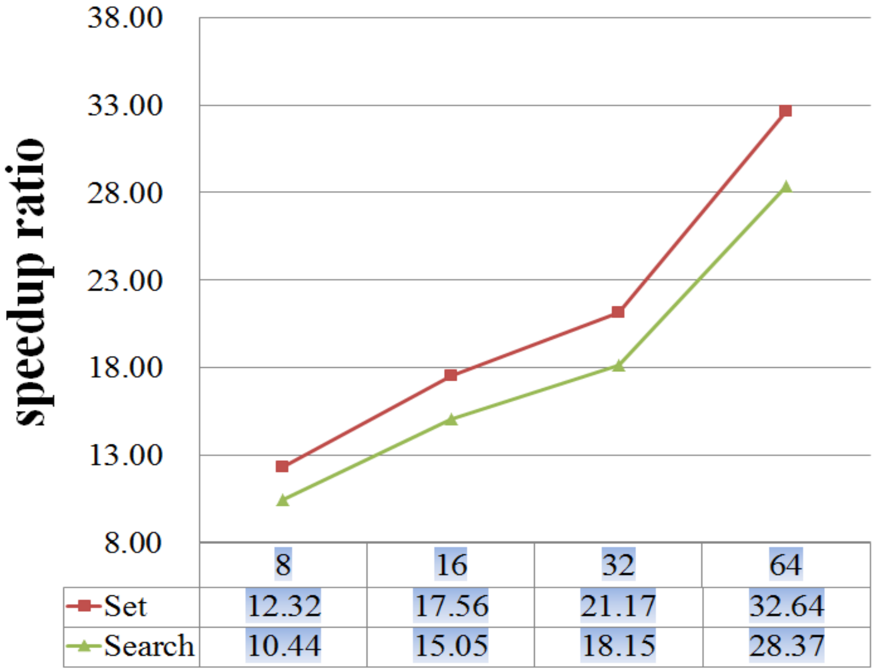
As system goes larger, the time-demanding part shifts from communication to force calculation.

EMMS Group

GPU MD parallelization: *performance*


No. of GPU(CPU) Process: 8

	case1	case2	case3	case4	case5	case6	case7
G1 Set	4.07	6.08	7.50	8.58	9.30	10.07	11.36
G1 Search	3.70	5.70	7.19	7.55	8.83	9.24	10.80
G2 Set	3.50	5.97	8.22	9.46	9.57	10.60	12.32
G2 Search	3.15	5.12	7.44	7.68	8.26	9.29	10.44


No. of GPU(CPU) Process: 8, 16, 32, 64

	8	16	32	64
Set	12.32	17.56	21.17	32.64
Search	10.44	15.05	18.15	28.37

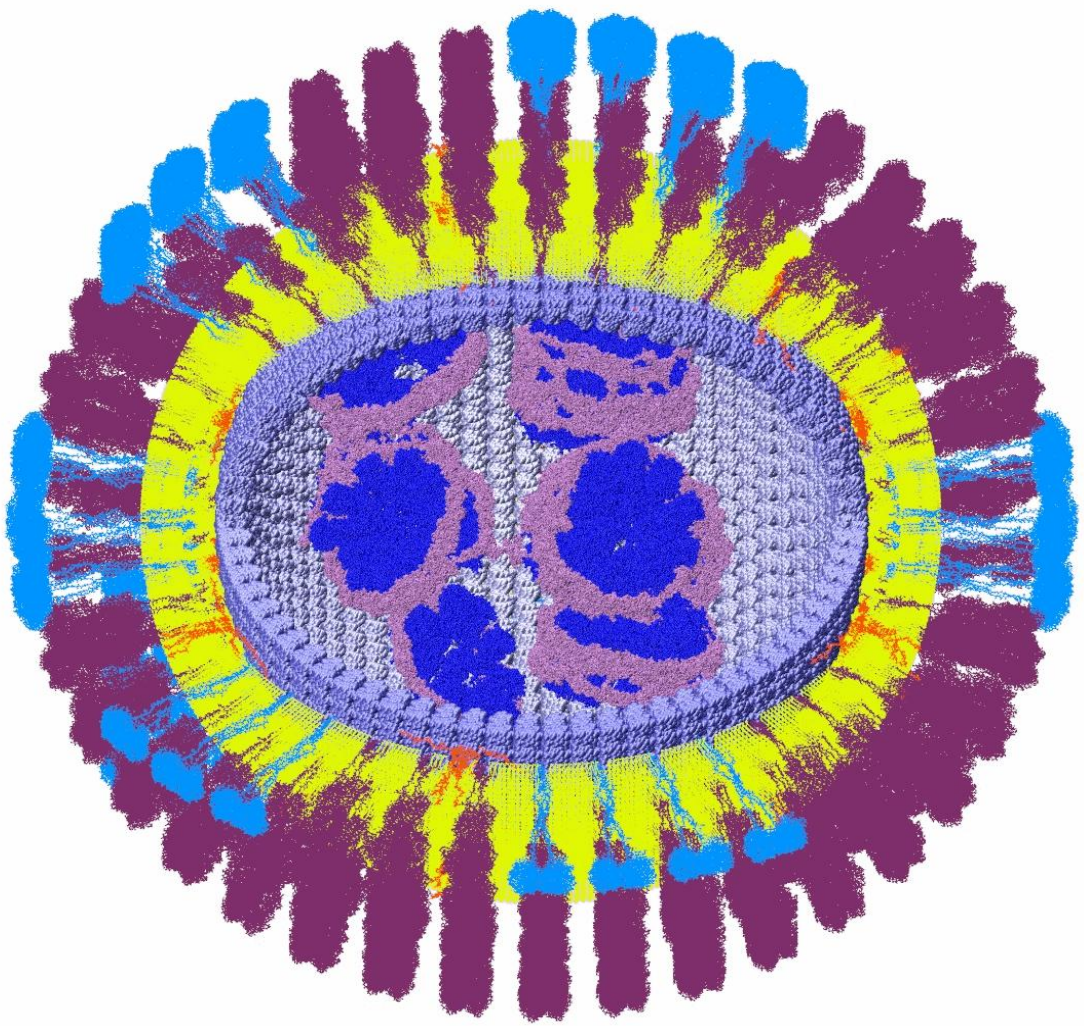
Our package runs even better for large systems, more GPUs.

Total speedup v.s. Gromcas-4.5.5

EMMS Group

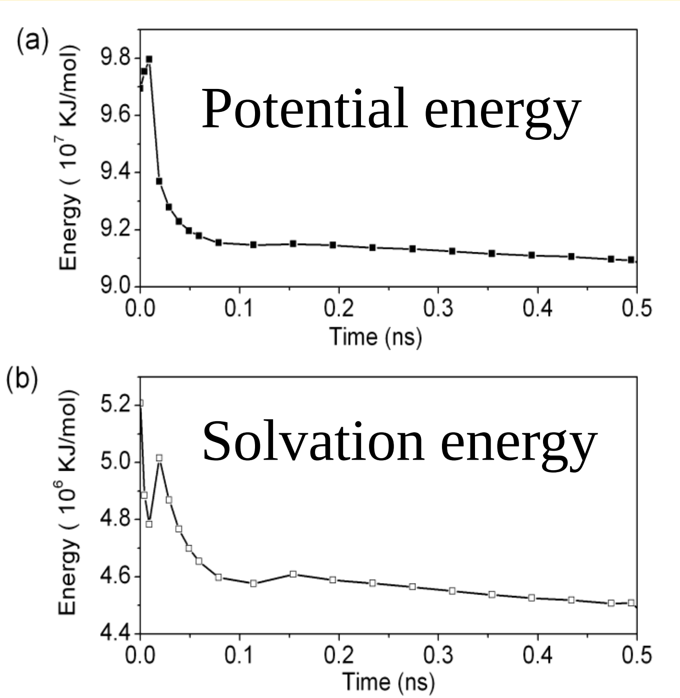
Virion simulation

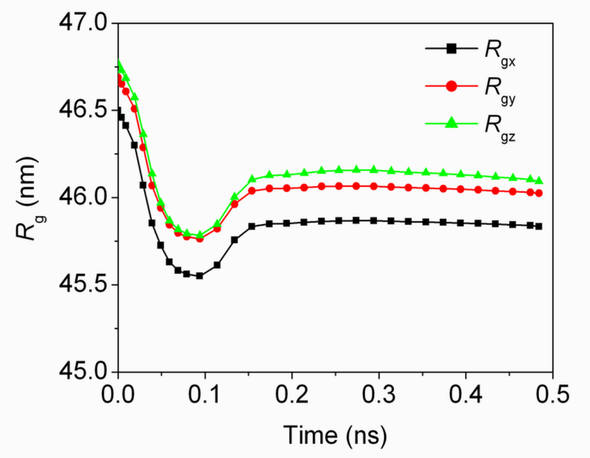
- ~300 million atoms, (148.5 nm)^3
- GROMOS 53a6 force field, SPC water
- NVT, T=300K
- Coulomb interaction with reaction field
- 3D domain decomposition:12^3=1728
- 288 nodes with 1,728 C2050 GPUs
- 0.77 ns/day with dt=1fs.



Xu J, Wang X, He X, Ren Y, Ge W & Li J. *China Science Bulletin, 2011,66(19):4426-4458.*

Virion simulation





Radius of the membrane

Starting from the predefined structure, the virus experiences a significant change to obtain a stable structure.

Conclusion and Prospect

EMMS Group

Conclusion

- A high-efficient MD package implemented on GPUs
- The first to produce a high-resolution picture of the complete influenza virus in solution.
- The most complicated biological system, with the largest number of atoms simulated at atomic level.
- GPU is a promising alternative for traditional CPU MD calculations.

Prospect

- **Optimization of software:**
  - GPU parallel implementation of PME
  - More force field inclusion
  - Better implementation, faster speed
- **Help to look for potential treatment targets**
- **Can be applied to polymer/emulsion…**

EMMS Group

Acknowledgment

People:

Institute of Process Engineering, CAS

Prof. Wei Ge, Xiaowei Wang, Xianfeng He

Dr. Ji Xu, Mengzhi Han

Institute of Chemistry, CAS

Prof. Xiaozhen Yang, Dr. Xiang Yu

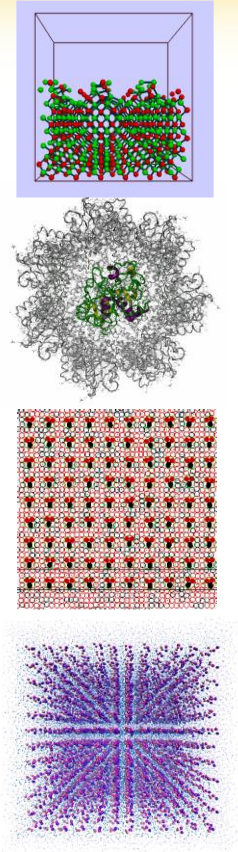
Institute of Microbiology, CAS

Prof. Fu Gao, Dr. Qing Li

Funding

- **Ministry of Finance (ZDYZ2008-2)**
- **National Key Science and Technology (2008ZX05014-003-006HZ)**
- **Chinese Academy of Sciences**
- **(KGCX2-YW-124)**
- **National Natural Science Foundation of China (20821092)**

EMMS Group: www.ipe.ac.cn/csms



Thanks for Your Attention!