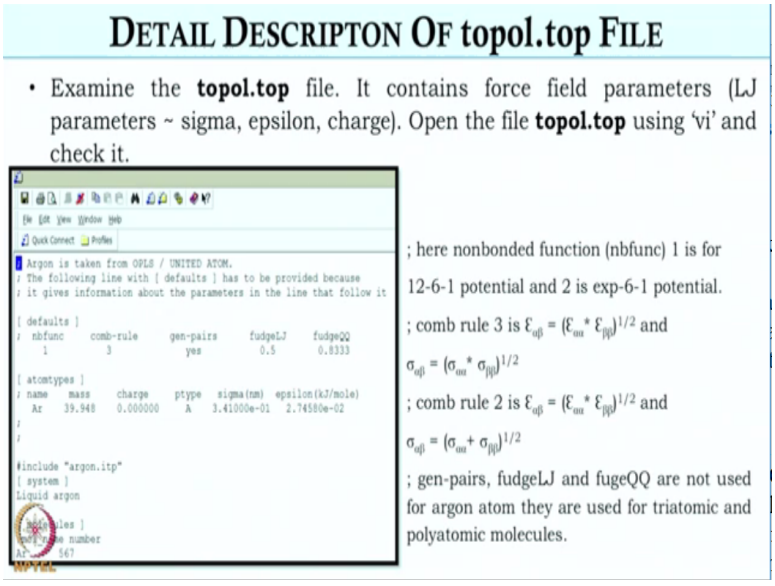
Computational Chemistry & Classical Molecular Dynamics
Prof. B. L. Tembe
Department of Chemistry
Indian Institute of Technology- Bombay

Lecture – 34
Molecular Dynamics using Gromacs-2 Simulating Liquid Argon

Hello again, this is our second lecture on Gromacs software, what we have done so far is that we have told you what kinds of files are needed for example, we started with a pdb file, we started with a top file, we talked of a gro file, now we will resume our discussion on the top file and then go to all other files that are needed and then Sonanki will tell us how to actually run the Gromacs package, so you go ahead with top file again.

Okay, so this is a topology file and topology file has a, said earlier it contain all be force fill parameters, the Leiningen parameter which are sigma, epsilon, charge, mass of the molecule, how many number of molecules are been stimulated, all these file, all this parameters are to be included in the topology file.

(Refer Slide Time: 01:04)



This is a topology file for liquid argon, so here we see that the first 3 lines are comment and the line were it is written, nb function, the nb function it stands for non-bonded function and if it is 1, then it means that we are using a Leiningen's potential that is a 1261 potential and if it is 2, then

it is exponential 61 potential, so it is one includes the coulomb also, yeah, the one include the coloumb also and it is user specified.

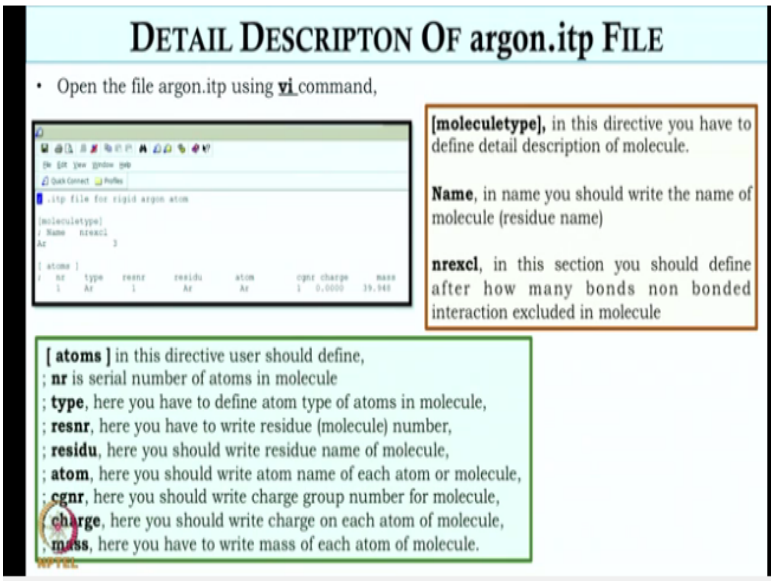
The kind of potential we want to use, we can specify it here, then comb rule is the combination rule, right, right that we will come when we come to water and other natures, yeah and the last line is your what is that; gen pairs, fudge LG and fudge QQ, it is used for poly atomic molecule and tri atomic molecule. For mono atomic molecule like argon, we do not use it and then the second type is the atom type, where we have to define the type of atoms like your; it is just only one that is argon atom.

Then, we have to specify charge of argon atom and mass, then sigma and epsilon in sigma in nanometres and epsilon in kilojoules per mole and all this parameters will get from OPLS directory, the file name is ff bonded and ff non-bonded. Itp, for each molecule and each atom we have sigma, epsilon, charge defined there and for each; different, different force will the parameters were. So, when you say include argon .itp, yeah, in which directory will that be argon.itp.

It will be in the same directory, as argon in the working directory, yeah, it has to be in the argon directory, okay and then we will have molecules where we have to define the residue name of the molecule and the number of; total number of molecule or atom that is been simulated, which are already chosen for 30 angstrom; box or 3 nanometre box like, yeah that is 567, so it has to be defined by the user, okay and it has to be changed.

And the residue name which is defined here that is ar, it has to be same throughout, it has to be same in the pdb file, gro file, topology file, it has to be same otherwise, it will give an error. Suppose, somewhere, somebody a instead of ar, then what will happen, it will get some error, it will give some error saying that residue name ar will not found, okay, whatever the name, okay, yeah it will show the error, yeah you can go to the next slide.

(Refer Slide Time: 03:30)



So, the next is the; we will have the detailed description of argon.itp file, so we can open the file using the command vi; vi argon .itp, this is the file for liquid argon, so here the molecule type; it is the directory, which where we have to give a detailed description of the molecule or the solvent or anything that we are simulating, then in name we have to write the residue name of the molecule or atom, here it is ar, since we have use a residue name ar for liquid argon.

Then, there is nrexcl, in this section, we have to define after how many bonds, the non-bonded interactions are included in the molecule, since this is a mono atomic molecule, it does not matter but in case of water or methanol, ethanol, it has to be chosen properly, so instead of 3, can I give 2 here, yeah or 4 here, yeah, yeah, does not matter, it does not matter because every pair is interacting, yeah and what is that last line back after the atoms.

After the atoms, I will just say, in this directive, we have to define nr; nr is a serial number of atom in the molecules, since there is only 1 argon atom like, okay, argon is a mono atom molecules, so serial number is 1 but if you have water, there will be 3 types, so we will have 3 serial number; 1, 2 and 3, then in type, we have to define the type of atom in the molecule, in resnr; that is the residue number.

We have to write the residue number of the molecule in residue, we have to give the residue name of the molecule, which we have used throughout, in atom, we have to write the name of;

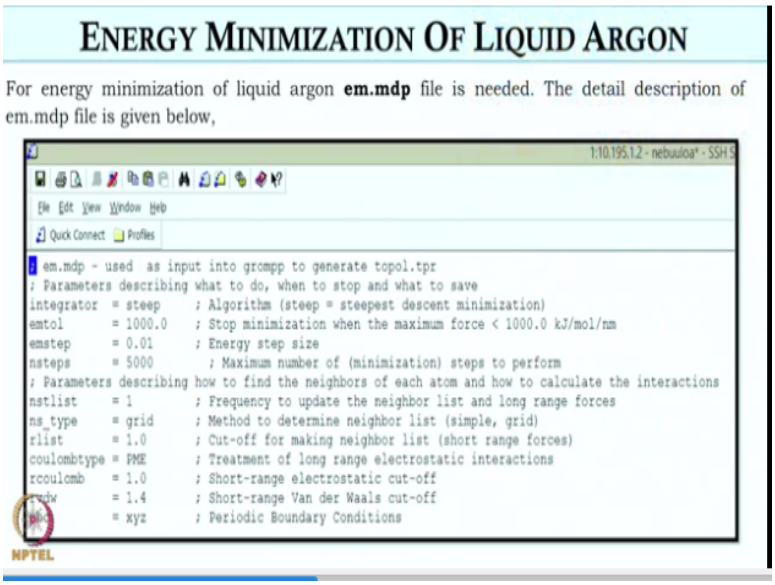
atom name of each molecule, then cgnr; this is the charge group number, okay, since there is it is a mono atomic molecule, so charge group number will be 1, then there is charge; we have to define the charge of the molecule and then there is mass.

So, this mass has to be the same that we started with, yeah, suppose this mas is slightly different from the original, it will take this as the final, yeah, it will take this is a final because this we are including in the topology file, okay and this we are using in a simulation. Now, what will be the the atom type, what is the meaning of atom type? Atom type; it will give a detailed description of the charge, mass, residue, name, each and everything of atom.

Then, how many residues are there? So, for example, suppose I am doing something like benzene, okay or say toluene okay, there the charge on one carbon in the ring could be different from the; so, we have to define each and every carbon atom there, so there will be 2 types of carbon then, yeah, so in case of benzene, we will have 6 nr; here nr is the serial number of atom, right, so in case of benzene, we will have 12 carbon, since there are 6 carbon and 6 hydrogen, okay.

So, we will have 12 residues, then each type of carbon atom will have a different charge, okay, so the type of atoms really means the total atoms then not just the type, yeah, it means; type plus the total; the total yeah. Suppose, I have something like (()) (06:27) with 60 carbons, it will have 60, there will be 60 different carbons, type of carbon atoms, yes and each one could have the same charge or they could have different charge, okay, depending on the environments, single bond or double bond, okay.

(Refer Slide Time: 06:40)



So, next we will perform energy minimisation of liquid argon, since we started; why do we need energy minimisation? Because in packmol, we started with the random configuration, we have just give a tolerance like minimum distance between 2 atoms should be < 3 angstrom or more than 3 angstrom or something but this is not properly equilibrated or energy minimised configuration, so we need an energy minimised configuration to start with.

So that we will get all the properties which we want to calculate from the simulation, it will correspond to a equilibrated configuration., okay, so for energy minimisation, we will need mdp file, mdp file; what instead of minimisation in principle, we could do direct simulation, but it may not be affective because if we energy minimised, we will get a global minimum, we may not get a local minimum butt we will get actually we will get a local minimum, right.

So, may be with minimisation, if you go ahead, the chance of a successive simulation is much higher, yeah and the properties like diffusion or potential energy which we want to calculate will have a better value of that okay, so is it like when students come to an institute, they do not know what it is and you give 1 or 2 days for orientation, yeah exactly, now this is like they orient nicely so that you can do the simulation, they orient nicely in the simulation box, okay.

So, we need; say for example if all the waters are in a strange arrangement, when you minimise, lot of hydrogen bonds will be form, so it is a more comfortable starting point for the actual run,

yeah, okay, so we need an mdp file for this and mdp file; mdp stands for a molecular dynamics parameter file and it contain all the information which is needed, so here is the information given for energy minimisation, we will need em.mdp, em for energy minimisation.

Now, already there are parameters like Leiningen's parameter, charge parameter, mass parameter, yeah, so mdp file is having many more parameters than the old ones, yeah, like you are giving some; I see somewhere .01 that is the time step, yeah, correct, so you will give many; so these includes more parameter than the starting one, yeah definitely, so this is let us say it is super file compared to the old one, yeah.

Because it will contain information like which is an algorithm we want to use for minimising the energy, then what is the total number of steps we required, then so is this the last parameter file?

No, will have, this is just for energy minimisation, now to perform an md simulation also, we will need a mdp file that is a more; that is even a bigger file than this, okay that will have maybe additional parameters, yeah that will have some additional parameters.

Okay, so, let us see what these parameters are, so here the first line is integrated that is steep, so that is the algorithm which we are using for energy minimisation that is a steepest descent algorithm, so, what are the; are there some other types, yeah, there is lfb, g algorithm, there are different other algorithm but for simple molecule or atoms (()) (09:44) the algorithm is fine to be like sufficient.

Did you use in your work other than steepest descent? no, I have used steepest descent algorithm throughout. Actually, steepest descent means, you have a high energy level, low energy, it comes fastest to your lowest energy, so this is supposed to be a very good method, yeah, so the next line is emtol that = 1000, so what is mean is that we need to stop minimisation when the maximum force or one atom or molecule is < 1000 kilo joule per mole per nanometre.

So, when the maximum forces 1000, it is reached, it will stop the minimisation, this should be between all pairs, is not it? Yeah, it should be between all pairs, so starting forces could be 100, 000 lakh, it can be any number because starting forces will be very high and is it possible that

some starting forces will be so high that you have to reject the configuration, yeah, it may be possibility.

Okay, so in that case, you have to do the pack more once again, yeah, okay, yeah, the next is the emstep that is 0.01 that is the energy step size, so after 0.01 step, it will perform again energy minimisation, so this is a step size for a energy minimisation. Then, this is not a time step, no this is not a time step, this is just after 0.01 timeframe, it will yeah, it is kind of a time step only, okay because we are given total number of steps that is n steps is 5000 here that is the maximum number of minimisation step.

So like, suppose I started from the step 0.0; step zero then it will go to step 0.01, then 0.02, like that it will perform the energy minimisation, okay. So, next is nstlsit that is a neighbour list which we need to update, so it can be, it is user-specified, it can be between 1 to 10, okay so for example, for liquid argon, could I have started with 5? Yeah, okay but if it is molecule; the preferred neighbour list is 10 for molecule, for water, for methanol, for poly atomic molecule, we prefer 10, okay.

Then, in principle, there is nothing to stop me for taking 12, no, there is nothing to stop but generally, it will take longer time to calculate, yeah, okay, so then is ns type which is grid, it is the method to calculate the neighbour lists, the neighbour of nearing; nearby argon atom, okay, so we use the grid method to calculate it, then is rlists that is the cut-off for the neighbour lists which is 1 nanometre then to calculate, then, so that means anything which is beyond 1 nanometre is not the neighbour of a given atom, yeah.

We have to define the cut-off for neighbour list also, then is coulomb type; coulomb type is for the treatment of long-range electrostatic interaction, we have to use certain algorithms, so here we are using particle mesh Ewald algorithm that is pme algorithm to treat the long-range electrostatic interactions. So, there are instead of pme, there is what is called reaction field, there is what is called ewald, may be when we have time, we can actually do this.

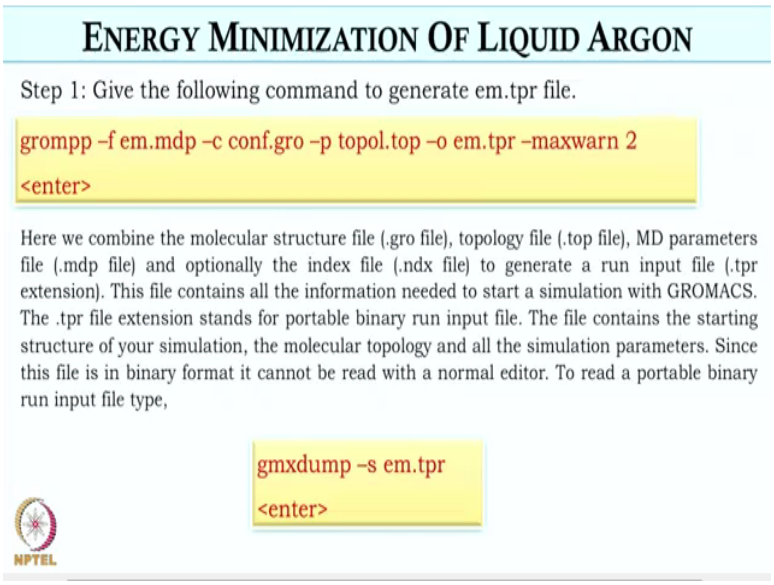
Or we can give one as an assignment, yeah to check which is more efficient, correct, then there is r coulomb; r coulomb is the cut-off for a short range electrostatic that is 1 nanometre, then rvdw is the Vander Waals cut off which is 1.4 nanometre and all this cut-off that rlist, r coulomb and rvdw, it should be < = 1/2 of the box length, correct, either < the box length or = box length, it cannot be more than the box length.

And then there is pbc, which is xyz direction that is a periodic boundary condition, it is applied in all the 3 directions. So, you said that all these cut off should be < the box length, yes, it cannot be more or it can be equal to the box length but it cannot be more than the box length, correct, so one of the reasons why we use cut off is that suppose, there are 500 molecules, it is not necessary to consider interaction in 1 molecule with all the 500.

Because whichever are very near it like say as sphere of 1 angstrom, sphere of 1 nanometre diameter, so instead of 500, this is around 250 and that 250 is enough to do a calculation, yes, we need not consider all, so this is just to increase the speed of computations and also if you have; if we start calculating between 500 atom, it will take much longer; much long time and then computation space will also; and plus many of those computations are wasted, know, yeah.

Exactly, it does not matter we can do all the calculations, if they are really needed, yes, but if it is unnecessary, cut off, yes, okay, yeah, so this is the mdp file that will be using for energy minimisation. So, now to perform energy minimisation, we need a tpr file, which is a binary file and it will contain gro file, topology file, an mdp file, so these are the 3 file that is contained in one single em.tpr file.

(Refer Slide Time: 14:44)



Now, to generate em.tpr file, we have type the command; grompp space –fem.mdp space –c conf.gro space p topol.top space –o em.tpr space maxwarn space 2, so grompp is the command which will combine the mdp, gro and topology file into one single binary input file which is the tpr file, so that is your –o refers to the output; -o refers to the output and –f –c –p are the flag for different files.

And if we just type grompp space -h which is the help command, it will display that what is the flag that is required before an mdp file, before a gro file, so one need not need to memorise the whole command lke grow grompp –f em.mdp, one will get if they just; after practice, you will know all that yeah, but now do you mean to say all the 4 will be needed in this; all the; mdp, gro, top and tpr; mdp, gro, top these are the main input file that is required.

And we will use all these 3 files in one; to generate; to generate a tpr files and grompp is the command to do that and in that single em.tpr file, it will contain information of all the 3 files, so that is the final thing that is needed to done; do the energy minimisation, yeah, this is the final thing which will need to do that energy minimisation.

So, suppose instead of maximum warning 2, I give 5, does not matter, suppose I give 0; 0 it would not take, 0 if you give them there is no point of giving a maximum command, maximum command is basically, right sometimes by running a simulation, we get warnings, correct and

warnings you can ignore, okay, so we generally get 1 or 2 warnings, so that is maximum 2 we give, or but suppose I do not give that command at all, max 1, 2, then what will happen?

Then, it will still work? Yeah, it will still work, if you get warnings say, sometimes while running the command grompp, you get like there are; there is 1 warning, there is 2 warning, so to ignore those warnings, we use maximum command because those warnings are not so important to consider but are these warnings because when the force is very large for example, not exactly the force is very large.

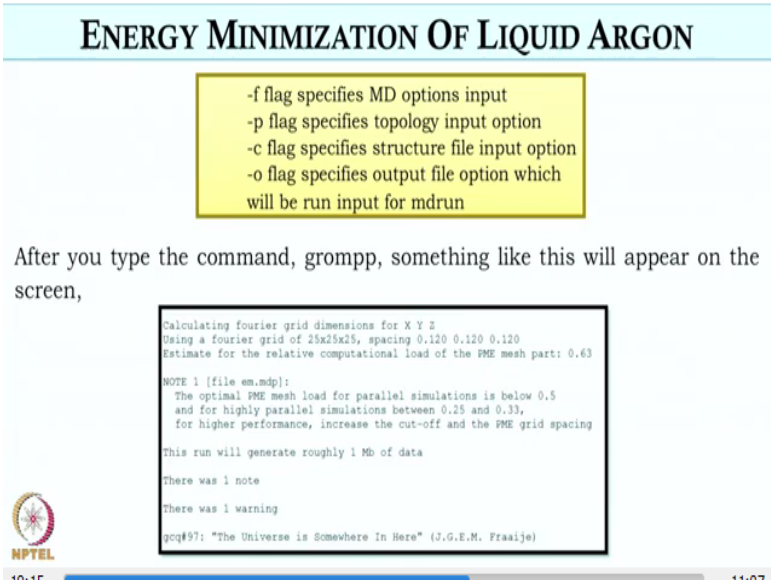
But sometimes like, this algorithm is not working for this molecule, we need to use different algorithm but if at all does not work, it will give a segmentation fault but warning we can ignore, okay, segmentation fault we cannot. So, now we will get a; it will; but then segmentation fault is not a warning? No, segmentation fault you cannot even run it, that means it will just crash; it will just crash, okay, you have to properly rectify and identify what is the error.

And then only it can work, so we combine all these 3 files using grompp command to generate em.tpr file and then since it is a binary file, we cannot view it using the vi command, so we have to have a different command to view the size, so gmxdump is the command, so you just type gmxdump space –s space em.tpr and then enter, so what this gmxdump s is reads the binary tpr file, yes and writes the output on the screen, yes.

So, is there a way to write the output not on the screen but into some file? No, or you can open the tpr file using a text pad, okay, so it will give all the detail, but then the text pad will give you too much information know, yeah it will have all; it will have the gro file first then it will have the topology file then it will have the mdp file, so suppose I do gmxdump –s em.tpr, yes, -o something, maybe I will be able to write it on some file, we can try that.

Yeah, we can try that but we cannot read the tpr file, correct, no, no, I am not saying, you are using gmxdump, yeah, the default is dumping on the screen yes, now, it should be possible to dump somewhere else, so that I can see it at leisure, yeah, maybe, okay maybe we can try it using –go command, correct.

(Refer Slide Time: 18:55)

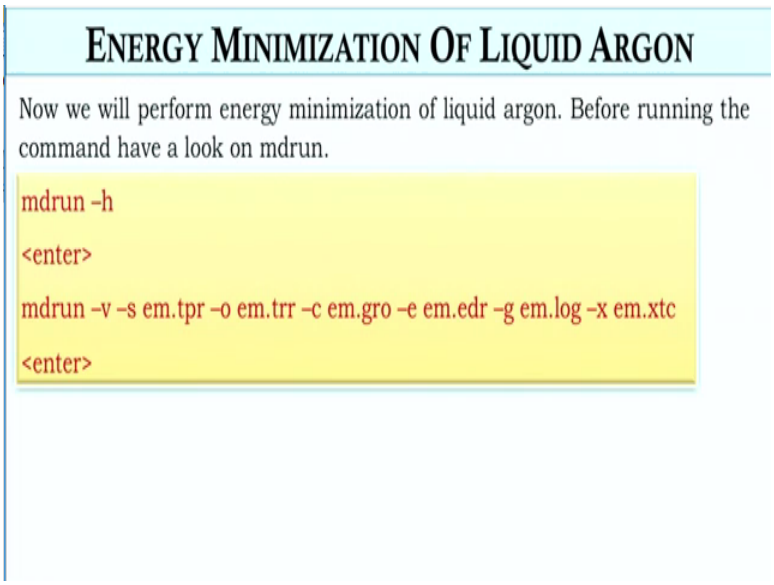


So, here I have given the detail that –f; flag –f; it specifies the MD options unit, -p will specified topology file units, -c is called configuration, -o is for output, okay. Now, after; so the configuration is what files; gro file, yeah, gro file, okay, so after we type the command grompp –f em.mdp something like this will appear on the screen, okay, so there is one warning here, yeah there is one warning.

So, you can ignore it, yeah that is okay and what are all those commands; the universe is something; yeah, that you when you run a command in Gromacs, you will always get a code, so that is just for some humour, yeah, it is just for some humour, so that you do not get exhausted, yeah, okay, so now; but we cannot change those command? No, we cannot change, it comes automatically.

Now, but it is not automatic because somebody has put all those commands, yeah, somebody has put, right, okay,

(Refer Slide Time: 19:51)

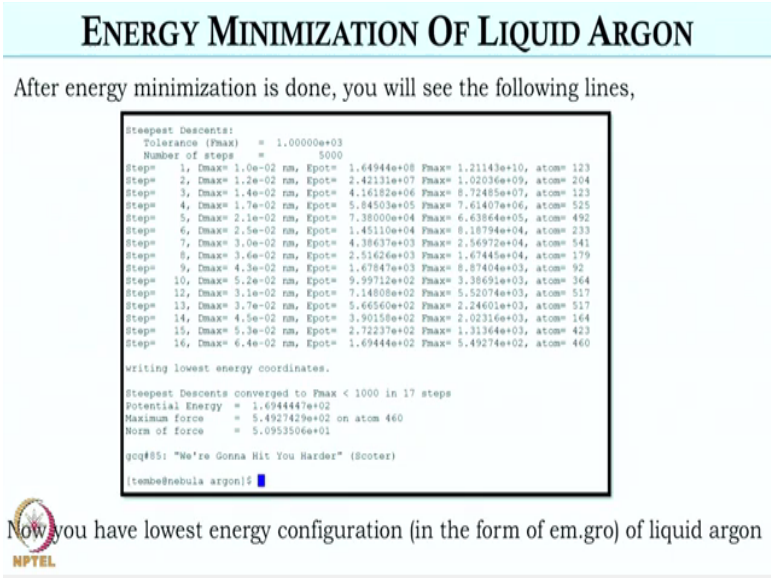


so now to perform energy minimisation using the tpr file, we need to use the command mdrun, so if we type mdrun –h, we can get; we can have a look on what it does, details of all the option; detail of all the options, so this is the full command that you have to type on screen to get the output file from mdrun but instead of using such a long command, we can as well use mdrun space -v space -deffnm space em.

So, it will create all the output file using initialise em, so it will create em.trr, it will create em.gro; em.gro is the energy minimise configuration, then em.edr is the energy file, it contains all the information about energy like potential energy, kinetic energy, temperature, pressure, it will contain all the files, em.log is the energy file in a log format, then em.xtc is the tragictory file at each time step what is the position coordinate velocity, it will contain all those information.

So, now how do I know that this mdrun is only minimisation and not final dynamics? Because we are using the tpr file of energy minimised; from energy minimisation, so if we use tpr file of full md, then we have to type mdrun space -v space –deffnm space full, okay, since we are using em, which is the tpr file from energy minimisation, so it will give us the output of energy minimisation, okay this is the command which we can use instead of the large command previously, it will give the same output, okay.

(Refer Slide Time: 21:30)



So, now after energy minimisation of liquid argon, we will see the following lines, where it will write the tolerance, number of steps and then steeps descent, converge to f max < 1000 kilo Joule per mole per nanometre, in how many steps it is getting converge, then what is the potential energy, what is the maximum force on which atom, it will give us all these and now, we will have the lowest energy configurations in the form of em.gro.

So, now our limit for the force was 10 to the 4, yeah, correct, now, this is saying maximum force 10 to the 2, yes, so that means it did better than what we expected, yes, correct, so instead of 10 to the 4 for the maximum force, we could have given 10 to the 3 also, yeah but maximum 10 to the power 4 is used because in case, we have a large molecule or large system, so in that case, so the other thing that is noticed here is that you did that only 17 steps are used.

Yes, okay and I have given 5000 steps and but the potential energy is still positive you know, yeah it is positive in case of neutral molecule, it generally comes as a positive point but that 10 to the 2 is not a very large number, no, not a very large number but when we do for charge molecules in water or ions, we will get a negative value of free; but if it is not negative then, you cannot really start that.

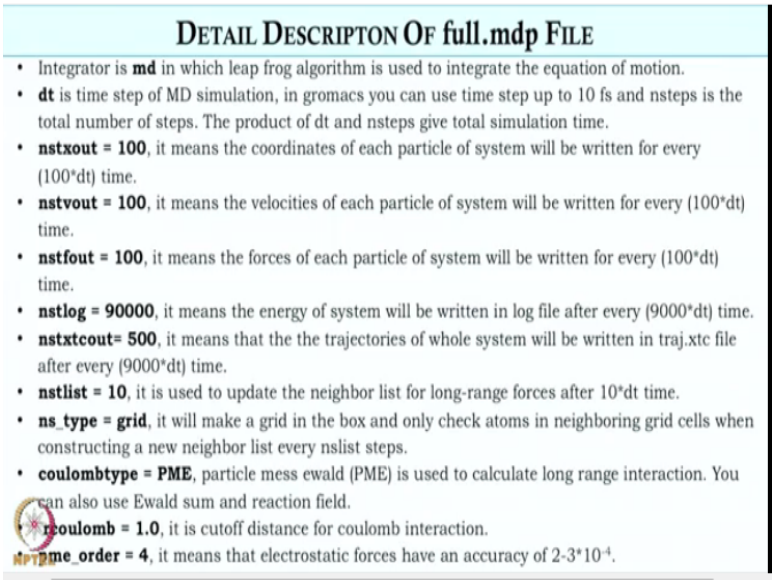
Yeah, because in case of neutral molecule, it is generally found to be positive but if we have a charge molecule, we will get negative quantity. But instead of 10 to the 2, suppose it was 10 to

the 4, then you may have to minimise again, yeah, then it would not get even minimised, it will show us error that fmax did not converged to; no, no fmax could be small yet the potential large can that happen or not; generally, not; generally, not.

Because if the force is small, the potential is also small, yes, okay, so now we are having the em.gro file which is the energy minimised configuration file and now taking this configuration file, we will start of full molecular dynamics simulation, okay. So, in full.mdp file, we will have all these details thing, so I will just give a brief information about all these; all the parameters, yes.

So, integrated; in the first line, we will have integrated; integrated is given as MD; MD which specifies that which is the algorithm that we are using to integrate Newton's equation of motion, so here we are using leap frog algorithm, there are many other algorithm; verlet algorithm, velocity verlet, there are some Runge Kutta method, yeah there are different methods, so in that case, we have to use a different notation; yeah, different notation.

(Refer Slide Time: 24:04)



In case, we are using MD which signifies, we are using leap frog algorithm, okay. The dt; it is the time step of MD simulation and it has to; the time step is generally taken as 2 femtoseconds then is nstxout, which mean, it is = 100, it means that it will save the coordinate of the particle after

every 100 * dt times step. Then, so now why do you think we need this 100, instead of 100, suppose I did 1000, what will happen?

It would not happen anything but you will write much fewer coordinates, it will write much fewer coordinates but if we need to analyse very carefully and if we need to get good results, it is better to have few like get the coordinates and velocity saved after fewer number of time step, okay. Then is nstvout; is 100, it means it will saved the velocities of each particle after every 100 * dt time.

So that nstout and nstvout; xout and vout, they can be different? 100, suppose, you are giving 100 as xout and 200 as vout, right, yeah that is user defined, it can be different, see that is suppose you do not want too much information on velocities, yeah then you may give that nstvout as very large number, yeah you can give as a 10,000 but it has to be less than the number of steps in the simulation, right, yeah.

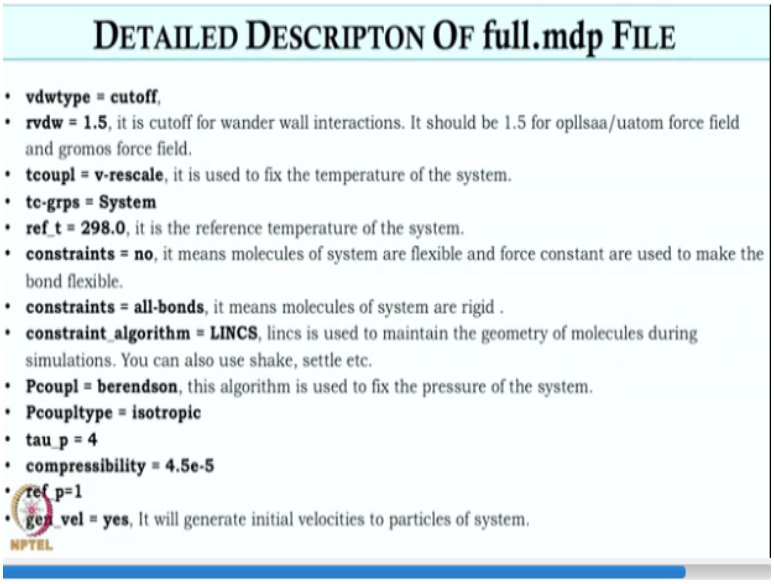
Then there is nstfout, it will write the force on each particle after every 100 8 dt time step, then there is nstlog, the energy of the system will be written in a log file, okay, then there is nstxtcout that is 500, if your nstlog, it is showing 90,000, yeah because we do not need, whereas on the right side, it is 9000 times dt, I think; I guess it will be 9000 times dt, it is a typo, okay, so both should be the same, yeah, both should be the same.

Then the next file is; but the more important is nst log value not what is written on the right side, no, okay, then is nstxtcout that is 500, it means that the trajectory of the whole system will be written after every 500 * dt times, let here also there is a typo, it has to be 500 * dt, not 9000, that is okay but main thing is your bold phase value, these are the parameters, yeah and every like this trajectory file is very important, so lesser the number, better we get the result, correct.

Then, nstlist is neighbour list as I have already said earlier, okay which is used to make update the neighbour list, then ns type is grid that is it will calculate the neighbour list using the grid method, okay, then coulomb type is PME that is the long range electrostatic interaction will be;

particle mesh ewald, yeah, then we are using coulomb cut off is 1 nanometer and PME order is the interpolation order for the electrostatic interaction which is taken as 4.

(Refer Slide Time: 26:57)



Then, there is Vander Waals type, here also we are using cut-off for van der Waals interaction that is 1.5 nanometre, then there tcouple that is the temperature coupling, so we have to; since we are performing simulation in npt or nvt ensemble, so we have to fix the temperature; we have to keep it constant throughout the simulations. So, if there is any fluctuation for some random reason, it will rescale to a temperature of our designed value, yeah, so we are using here velocity rescaling thermostat to fix the temperature of the system.

TC groups mean temperature coupling group, so here you can specify which are the group that you need to couple or you need to fix a temperature also, so here we are fixing the temperature; so for example, we have water, I need not use hydrogen's, is that what you are saying? No, no system is like you have water and methanol, okay, right, so you have to fix the temperature of water and not methanol.

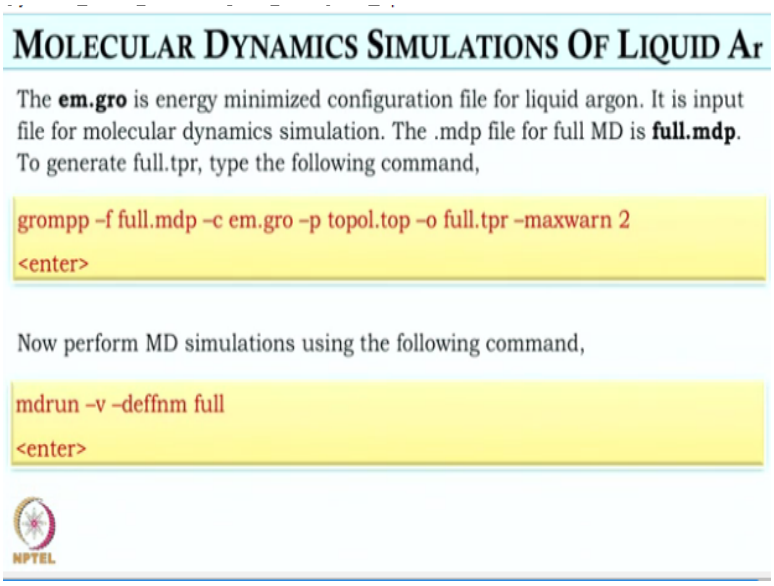
So, you can specify water as different coupling group but generally, we fix the temperature of the whole system that is what is desired, then there is reference temperature, which is 298, so that; to which all things are scaled, yes, they are all to 298, yeas and then there is constraint known that

is because you are; we are using just mono atomic molecules, so we do not need to fix any bond length or bond angle.

And then there is P couple that is Berendson, so here we are using pressure coupling; power pressure, yeah Berenson algorithm is being used and Pcouple type is isotropic that is, it is taking as isotropic system, then tau p is the pressure coupling constant which is use as 4 picoseconds, then compressibility is 4.5, it is in bar; bar inverse, then reference pressure is one bar and gen velocity means, it will generate velocity using the Maxwell Boltzmann distribution, okay, so these are the parameters that is specified in the mdp file, right.

So, now the next command is for your actual MD or minimisation , we have already done the minimisation, so the next command will be for full MD, so we will just show the command and conclude this lecture, okay.

(Refer Slide Time: 29:09)



So, we have already em.gro, which is the energy minimised configurations and then we will have full.mdp, instead of em.mdp, so this is the command to do the final MD simulation, yes, it is the same command as em.mdp, just instead of em.mdp, we will be using full.mdp, will be using the output gro file of energy minimisation that is em.gro, topol.top file is the same, it is the common, yeah, it is the common and then output value we are giving as full.tpr instead of em.tpr.

So, the max one here need not be the same as earlier, it could be different, right and now we will be doing MD simulation using the command mdrun space –v space –deff nm full, full is the tpr file that we are getting from the grompp command, so what we will do; we will conclude this with this mdrun statement and next time, again we will begin with mdrun and then analyse all the trajectories and if there is time we will also do the installation, okay, thank you.