Tutorial: Accelerating Simulations of Large Systems Using Virtual Sites

Abhishek Acharya

June 27, 2019

I. Introduction

The aim of this tutorial is to give a demonstration of how virtual sites can be used for accelerating all-atom MD simulations, particularly helpful for systems with large proteins. Of course, there are other applications of virtual sites such as construction of complex toplogies for molecules or in hybrid all-atom/coarse-grained simulations. The topologies for TIP4P or TIP5P water models (included with the GROMACS distribution) would be instructive in this regard. Interested readers may also look at the MARTINI tutorial for running hybrid simulations using virtual sites (http://cgmartini.nl/index.php/tutorials-general-introduction/tutorial-hybrid-model-using-virtual-sites).

The key idea is to increase the largest possible timestep used for integrating the equations of motion with reasonable accuracy. As we know, the timestep allowed is limited by the fastest degrees of freedom present in the system. In usual all-atom MD setups, the fastest degrees of freedom are the bond stretching vibrations involving H-atoms (10 fs). These are usually treated using constrainsts. The next fastest DOF is that of bond-angle vibrations involving H-atoms (13 fs), which currently limits the largest timestep to 2 fs. Although, constrainst can be used to set bond-lengths and angles as constant, and increase the limit to about 5 fs, a nice way to run stable simulations with larger timesteps is to altogether remove these high-frequency motions involving hydrogen from the system. In this alternative approach, hydrogen atoms are treated as Virtual Interaction Sites (simply, dummy atoms) that are NOT connected to the parent atom with a bond. Instead, the positions of the dummy atoms are calculated at every MD step from the positions of nearby connected heavy atoms using geometrical rules. The force acting on a dummy atom is redistributed to the heavy atoms, and the positions of these heavy atoms evolve under the action of these forces. The hydrogen atoms treated as dummy particles do not have an associated mass; therefore, to keep the momentum of the system umchanged, the mass is added to the bonded heavy atoms. The bond, angle and dihedral terms describing the natural connection of the hydrogen atom is removed. But, in case of specific functional groups requiring rotational freedom (amine, hydroxyl) care must be taken to keep the rotational degree of freedom intact. We will go over the finer details of these aspects in sections below. For more details about virtual sites, the reader is referred to the GROMACS manual and the orginal paper on virtual sites [Feenstra, Hess and Berendsen.(1999).JCC.Vol. 20, No. 8, 786-7989].

Here on, we will give a brief overview of the types of virtual sites implemented in GROMACS. We will look at an example protein (trp-cage) to see how the topology is constructed when treating hydrogens as virtual sites. Note that for proteins, pdb2gmx can automatically generate the virtual site topology file using the -vsite flag. But, it would be instructive to look at the protein topology first to get an understanding of the toplogy construction. Eventually, it would be helpful in the second example where we construct the topology of a ligand with dummy hydrogen atoms by hand.

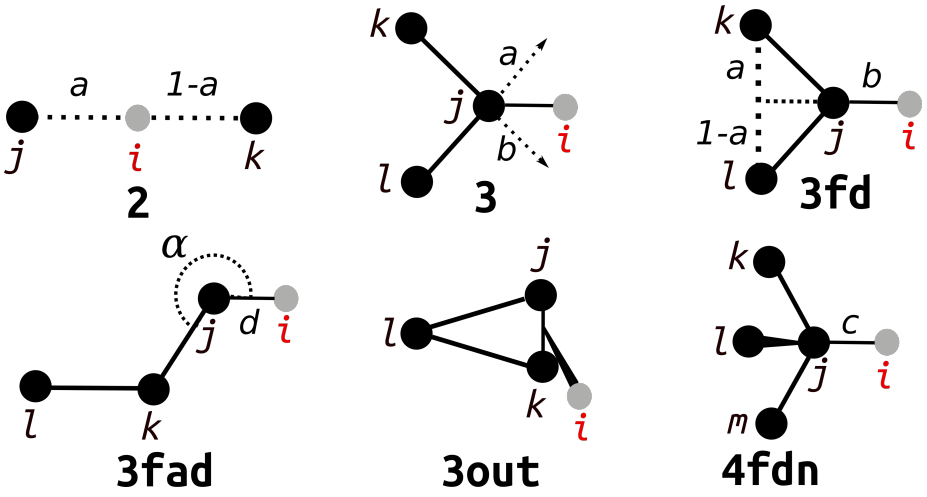

Figure 1: Types of virtual sites in GROMACS.The gray circles are dummy atoms and the black ones are the constructing atoms.

II. Types of Virtual Sites in GROMACS

There are six types of virtual site construction possible in GROMACS. For particular types of hydrogen atoms (polar connected to oxygen, non-polar connected to sp3 carbon etc.) particular constuctions can be used. This will be more clear with appropriate examples for each type given below.

Figure 1 below shows the different types of virtual site construction available in GROMACS.

a. Type 2 Virtual Site

This type is constructed as a linear combination of two atoms. As shown in Figure 1, the virtual site lies in the line connecting the two constructing atoms. In the topology file, this construction is specified under the topology directive *[virtual_sites2]*. The first column is the index of the virtual atom and the next two columns specify the two constructing atoms. Finally, the last column specifies the function to be used, which in this case is 1.

This type is not used for construction of hydrogen atoms in proteins but may be useful in construction of virtual sites for other molecules. See the virtual site tutorial by J. Lemkul for an example of this type used for a Carbon-dioxide molecule (http://www.mdtutorials.com/gmx/vsites/index.html).

b. Type 3 Virtual Site

This type is constructed as a linear combination of three atoms (Refer to Figure 1). The virtual site in this case lies in the plane containing the three constructing atoms (i,j and k). In the topology file, this construction is specified under the topology directive *[virtual_sites3]*. The first column is the index for the virtual atom, while the next three columns specify the index of the three constructing atoms. Again, the final column here is the function type 1.

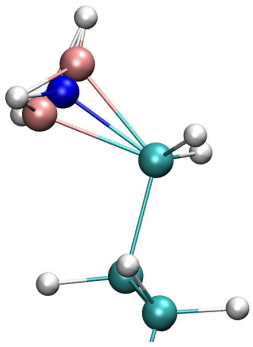

Figure 2: Construction of a NH3 group in Lys

In proteins, this type of construction can be used in case of umbrella-NH2 (freely-rotating amines), -NH3 and -CH3 groups.

c. Type 3fd Virtual Site

This type has the virtual site constructed in the plane of the three constructing atoms (*j,k* and *l*), but at a fixed distance b from the adjacent bonded atom j (Refer to Figure 1). In the topology file, this construction is specified under the topology directive *[virtual_sites3]*. The first column is the index for the virtual atom, while the next three columns specify the index of the three constructing atoms. Again, the final column here is the function type 2.

Examples of such a construction in proteins can be observed in case of backbone -NH- group, hydrogens of aromatic rings (Phe,Trp,His,Tyr) and Nε hydrogen in Arg.

d. Type 3fad Virtual Site

For this type the virtual site is constructed in the plane of the three constructing atoms (*j,k* and *l*) at a fixed angle α with $r_{j,k}$ and distance d from the adjacent bonded atom j (Refer to Figure 1). Like other three atom constructions, this construction is also specified under the topology directive *[virtual_sites3]*. First four columns are the index for the virtual site and the three constructing atoms, respectively, and the fifth column is the function type 3.

This construction is employed for groups such as planar -NH2 in Asn and Gln and the two -NH2 moieties in the guanidino group of Arg.

e. Type 3out Virtual Site

This type of virtual site is constructed out of the plane containing the three constructing atoms (Refer to Figure 1). In the topology file, this construction is specified under the topology directive *[virtual_sites3]*. The first column is the

index for the virtual atom, while the next three columns specify the index of the three constructing atoms. The final column here is the function type 4.

In proteins, this type of construction can be used in case of -NH3 and -CH3 groups, and Gly CA hydrogens.

f. Type 4fdn Virtual Site

Here, the virtual site is constructed from 4 atoms (*j,k,l* and *m*) at a fixed distance *c* from the adjacent bonded atom *j*. In the topology file, this construction is specified under the topology directive *[virtual_sites4]*. The first column is the index for the virtual atom, while the next four columns specify the index of the four constructing atoms. The final column here is the function type 2 (function type 1 is also available, but it is not recommended as it can be unstable).

This type of construction is observed in case of the Cα hydrogen (exception is Gly Cα hydrogens which are constructed using 3out type construction).

III. Case 1: Simulations of Trpcage Protein

Now that we have looked at the various virtual site types, we will go into the details of how each construction type is implemented in the protein toplogy file, by taking the example of the Trpcage protein.

a. Treatment of different types of hydrogens in protein

a.i. Umbrella-NH2, -NH3 and -CH3 groups

We are going to take the example of -NH3 group of Lysine, to go over the details of topology construction. Note in Figure 2 the two dummy mass centers that are added. In the toplogy file (relevant sections printed below) you would see the following for Lysine.

```
[ atoms ]
...
163         CT      8    LYS     CE    149    -0.0143     14.026
164         HP      8    LYS    HE1    150     0.1135          0
165         HP      8    LYS    HE2    151     0.1135          0
166       MNH3      8    LYS   MNZ1    152          0      8.517
167       MNH3      8    LYS   MNZ2    152          0      8.517
168         N3      8    LYS     NZ    152    -0.3854          0
169          H      8    LYS    HZ1    153       0.34          0
170          H      8    LYS    HZ2    154       0.34          0
171          H      8    LYS    HZ3    155       0.34          0
...

[ bonds ]
...
160   163     1
163   164     1
163   165     1
163   168     1
```

```
168   169     1
168   170     1
168   171     1
...

[ constraints ]
...
163   166     2
163   167     2
166   167     2
...

[ pairs ]
...
162   164     1
162   165     1
162   168     1
164   169     1
164   170     1
164   171     1
165   169     1
165   170     1
165   171     1
...

[ angles ]
...
157   160   163     1
161   160   163     1
162   160   163     1
160   163   164     1
160   163   165     1
160   163   168     1
164   163   165     1
164   163   168     1
165   163   168     1
163   168   169     1
163   168   170     1
163   168   171     1
169   168   170     1
169   168   171     1
170   168   171     1
...

[ dihedrals ]
...
  157   160   163   164     9
  157   160   163   165     9
  157   160   163   168     9
  161   160   163   164     9
```

```
  161   160   163   165     9
  161   160   163   168     9
  162   160   163   164     9
  162   160   163   165     9
  162   160   163   168     9
  160   163   168   169     9
  160   163   168   170     9
  160   163   168   171     9
  164   163   168   169     9
  164   163   168   170     9
  164   163   168   171     9
  165   163   168   169     9
  165   163   168   170     9
  165   163   168   171     9
  ...

 [ virtual_sites3 ]
...
  168   163   166   167     1
  169   163   166   167     1
...
 [ virtual_sites3 ]
...
  170   163   166   167    -4
  171   163   166   167     4
...
```

Here above are the topology directives used to define virtual sites in case of NH3 of Lys.

You will notice that in order to construct the -NH3 group, additional dummy sites and contraints are required in the topology. The two dummy masses are added and constrained with respect to the Cϵ atom in a triangular formation. The total mass on the two dummy sites is just the mass of the NZ atom. Based on the positions of these three points (one Cϵ and the two dummy masses), we can calculate on the fly the position of the NZ atom considering the fact that the new atom NZ must lie on the line passing through and equidistant from both dummy mass points. This is the type 3 virtual site construction and would appear under the topology directive *[virtual_sites3]*, as shown above. Additional virtual site directives are required to provide the rules for generating the positions of the three hydrogens connected to NZ atom. We use Type 3 virtual site construction for the position of the hydrogen (atom 169) in the plane of the three constructing atoms. For the other two hydrogens (atoms 170 and 171) we generate the positions using the Type 3out directive. The sign associated with the function number 4 is used to denote whether the virtual site is constructed above or below the plane containing the constructing atoms.

As an exercise, you may inspect the topology for -CH3 group.

a.ii. Planar-NH2 group

For planar -NH2, we take the example of the amide group of Asn. Relevant sections of the topology are printed below.

```
[ atoms ]
...
    12          C      1    ASN     CG     10     0.5833      12.01
    13          O      1    ASN    OD1     11    -0.5744         16
    14          N      1    ASN    ND2     12    -0.8634     16.026
    15          H      1    ASN   HD21     13     0.4097          0
    16          H      1    ASN   HD22     14     0.4097          0
...

[ bonds ]
...
    9    12     1
   12    13     1
   12    14     1
   14    15     1
   14    16     1
...

[ pairs ]
...
   12    17     1
   13    15     1
   13    16     1
...

[ angles ]
...
   10     9    12     1
   11     9    12     1
    9    12    13     1
    9    12    14     1
   13    12    14     1
   12    14    15     1
   12    14    16     1
   15    14    16     1
...

[ dihedrals ]
...
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult1
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult2
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult3
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult4
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult5
    7     9    12    14     9    torsion_ASN_CA_CB_CG_ND2_mult6
```

```
    7     9    12    13     9
   10     9    12    13     9
   10     9    12    14     9
   11     9    12    13     9
   11     9    12    14     9
    9    12    14    15     9
    9    12    14    16     9
   13    12    14    15     9
   13    12    14    16     9
...

[ dihedrals ]
...
    9    14    12    13     4
   12    15    14    16     4
...

[ virtual_sites3 ]
...
   15    14    12     9     3
   16    14    12     9    -3
...

Here above are the topology directives used to define virtual sites in
case of NH2 group of Asn.
```

You will notice in the topology that the two hydrogens (15 and 16) are constructed using the Type 3fad virtual site directives. Here, the function has a sign associated which denotes the direction in which the angle is measured.

a.iii. Secondary amine or amide (-NH-) and aromatic -CH groups

This is a very straightforward. The relevant section of the topology is given below. As you can see, the virtual site type used for construction of the hydrogen is Type 3fd.

```
[ atoms ]
...
    17          C      1    ASN      C     15     0.6163      12.01
    18          O      1    ASN      O     16    -0.5722         16
    19          N      2    LEU      N     17    -0.4157     15.018
    20          H      2    LEU      H     18     0.2719          0
    21         CT      2    LEU     CA     19    -0.0518     13.018
...

[ bonds ]
...
   17    18     1
   17    19     1
```

```
   19    20     1
   19    21     1
   21    22     1
   21    23     1
   21    40     1
...

[ pairs ]
...
...

[ angles ]
...
    7    17    18     1
    7    17    19     1
   18    17    19     1
   17    19    20     1
   17    19    21     1
   20    19    21     1
   19    21    22     1
   19    21    23     1
   19    21    40     1
   22    21    23     1
   22    21    40     1
   23    21    40     1
   21    23    24     1
   21    23    25     1
   21    23    26     1

...

[ dihedrals ]
...
    7    17    19    20     9
    7    17    19    21     9
   18    17    19    20     9
   18    17    19    21     9
   17    19    21    22     9
   17    19    21    23     9
   17    19    21    40     9
   20    19    21    22     9
   20    19    21    23     9
   20    19    21    40     9
   40    21    23    26     9    torsion_LEU_C_CA_CB_CG_mult1
   40    21    23    26     9    torsion_LEU_C_CA_CB_CG_mult2
   40    21    23    26     9    torsion_LEU_C_CA_CB_CG_mult3
   19    21    23    24     9
   19    21    23    25     9
   19    21    23    26     9
   22    21    23    24     9
```

```
   22    21    23    25     9
   22    21    23    26     9
   40    21    23    24     9
   40    21    23    25     9
   19    21    40    41     9
   19    21    40    42     9
   22    21    40    41     9
   22    21    40    42     9
   23    21    40    41     9
   23    21    40    42     9

...

[ dihedrals ]
...
   17    21    19    20     4
...

[ virtual_sites3 ]
...
  20    19    17    21     2
...
```

Here above are the topology directives used to define virtual sites in case of -NH group of backbone amide linkages.

a.iv. Hydroxyl (-OH) and sulfhydryl (-SH) groups

The -OH and -SH groups are treated a little differently. As you will notice in the topology printed below, in case of Ser the polar hydrogen is not treated as a dummy atom. Your will find that the mass of the polar hydrogen (atom 222) is not set to zero. In this case, we just add an additional constraint on the C-O-H angle by putting distance constraints on C—H distance. No virtual site directive is needed in this case.

```
[ atoms ]
...
   218         CT     13    SER     CB    202     0.2117     14.026
   219         H1     13    SER    HB1    203     0.0352          0
   220         H1     13    SER    HB2    204     0.0352          0
   221         OH     13    SER     OG    205    -0.6546         16
   222         HO     13    SER     HG    206     0.4275      1.008
...

[ bonds ]
...
  218   219     1
  218   220     1
```

```
  218   221     1
  221   222     1
...

[ constraints ]
...
218   222     2
...

[ pairs ]
...
...

[ angles ]
...
    217   216   218     1
    217   216   223     1
    218   216   223     1
    216   218   219     1
    216   218   220     1
    216   218   221     1
    219   218   220     1
    219   218   221     1
    220   218   221     1
    218   221   222     1
...

[ dihedrals ]
...
  217   216   223   224     9
  217   216   223   225     9
  218   216   223   224     9
  218   216   223   225     9
  216   218   221   222     9
  219   218   221   222     9
  220   218   221   222     9
...
```

b. Setting up the Trpcage simulation

Now once the topology is ready, it is generally trivial to run the simulations. During equilibration steps, it is advised to start the NVT equilibration step using 2fs as the timestep; this can be followed by 2fs of NPT. Thereafter, additional equilibration runs at increasing timestep values or 4fs and finally 5fs, should be run. Finally, production runs can be done using 5fs as a timestep, which leads to at least 2.5 times accelation in simulations.

IV. Case 2: Simulations of a Ligand

Unlike for proteins, virtual site topology for ligand are not generated automatically by gromacs, and need to be built by hand. This is where the undestanding of virtual site topology construction obtained from previous section would be helpful.

a. Treating ligand hydrogens as virtual interaction sites

Building virtual site topology for ligands typically involves modifying a normal full atom topology. So, the following procedure assumes that one already has an optimized topology obtained using methods routinely used for a given force-feild. Once we have a normal working topology for the ligand of interest, we can work through the following steps to edit the file and generate a virtual site topology. For the current demonstration, we take the example of N-Acetyl Glucosamine-1-phosphate (GlcNAc1P).

Step 1: Inspect the ligand structure. We study the ligand structure and make note of the different types of hydrogens that are present in the ligand. This will give us an idea of the different types of virtual site directives we need in the topology file. On inspection, you will find the following types of hydrogen (see Table 1). It is helpful to also note down the number of occurences of each and the gromacs virtual site type that could be used to treat these as virtual sites.

Table 1: Different types of hydrogens in GlcNAc1P.

Hydrogen type	**occurences**	**Virtual site type**
Polar hydrogen of -OH group	3	None
Hydrogen on CH3 group	3	Type 3 and 3out

b. Final notes