# Hybrid Monte Carlo Code User Documentation

Updated: 9/1/2014

by

Jeremy C. Palmer
Department of Chemical and Biological Engineering
Princeton University
Princeton, NJ 08544
jcpalmer@princeton.edu

This file contains basic user documentation for the accompanying Fortran-based Hybrid Monte Carlo (HMC) code that was used to generate some of the data presented in *Nature* **510**, 385 (2014) for the ST2 water model. The code implements the HMC algorithm [Duane *et al*., *Phys. Lett. B.* **195**, 216 (1987)] using the quaternion-based rigid-body molecular dynamics (MD) integrator of Miller *et al*. [*J. Chem. Phys*., 116, 8649 (2002)]. The MD integrator routines in the code closely follow the structure of the version implemented in DL_POLY Classic, which is described in DL_POLY Classic's user manual (v1.9, 2012). The user manual for DL_POLY Classic also provides an excellent overview of the rigid-body equations of motion in quaternion space.

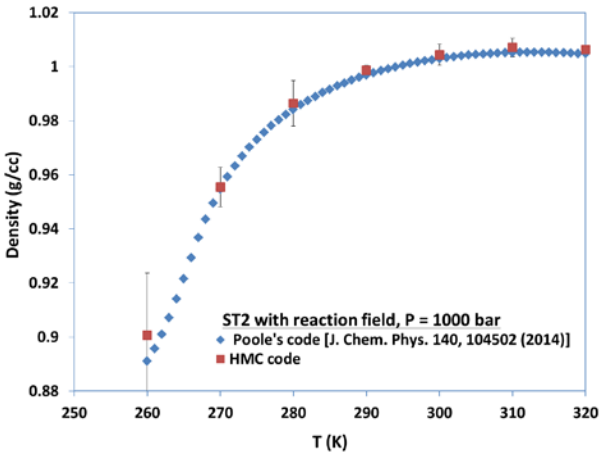
Two variants of the ST2 water model are implemented in the code. The first variant is described by Palmer et al. [*Faraday Discuss.* **167**, 77 (2013)] and uses the Ewald summation method with vacuum boundary conditions to treat the long-range electrostatic interactions. To use this variant, the option ***ewald_vacuum*** must be set (see discussion of the input below). If this command is not invoked, the program defaults to using the ST2 variant described by *Poole et al*. [*J. Chem. Phys.* **138**, 034505 (2013)], in which the long-range electrostatic interactions are treated using the reaction field method.

Message-passing functionality (via MPI) has been introduced into the code in order to make it more convenient to run on parallel machines where serial jobs receive low priority in the queueing systems. It is important to note that in parallel mode, each processor core runs an independent simulation, or replica, at the same temperature and pressure specified in the main input file. When performing umbrella sampling, however, each processor core may be assigned different parameters for the harmonic bias potential. This design is extremely convenient for performing many umbrella sampling simulations at the same temperature and pressure on parallel machines. Moreover, because the message-passing functionality has already been introduced, it is trivial to modify the code to run parallel replicas at different temperature and pressures, or event implement advance Monte Carlo move sets such as parallel tempering or Hamiltonian exchange, which require minimal communication between the independent replicas. If the code is run on Intel-based hardware, it is highly recommended that Intel Fortran compliers are used to compile the code because they perform hardware specific optimizations that will significantly accelerate the computations.

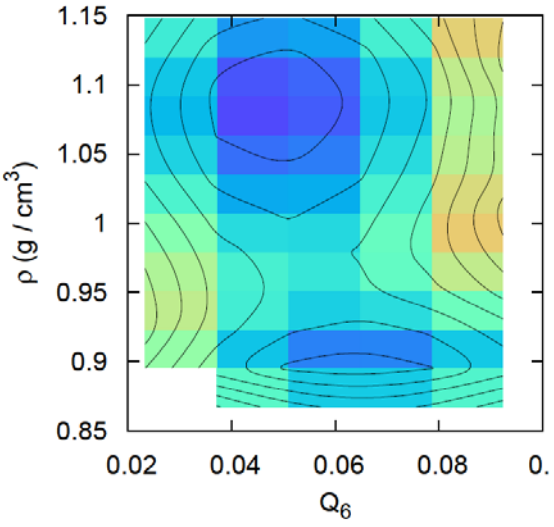
**Validation of the Code:**

Because the original version of the HMC program used to generate data for *Nature* **510**, 385 (2014) was derived from code that was initially developed for other applications besides simulating ST2 water, it contained many obsolete legacy options and functions. In order to make the code more user-friendly and reduce its length, obsolete functionalities have been removed, several options have been hardcoded, and the format of the input and output has been slightly modified. The integrity of the current version of the code has been verified by performing a number of tests.

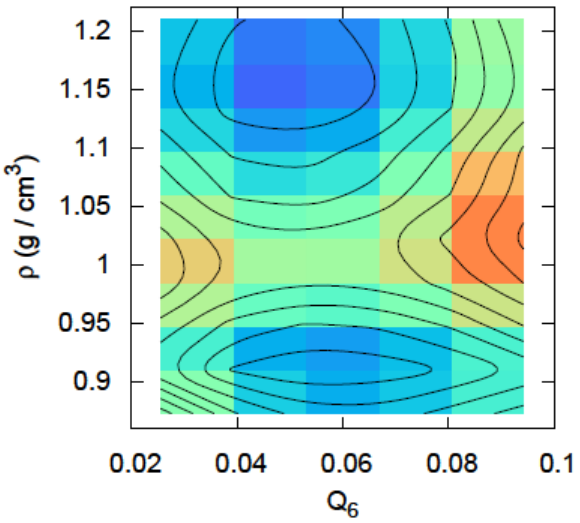
1. Comparison of the P=1000 bar isobar for ST2 with reaction field against equation of state data reported by Peter Poole [*J. Chem. Phys*. **140**, 104502 (2014)]




2. Calculation of the free energy surface for $N$ = 216 for the ST2 variant employing the reaction field method at 235 K and 2250 bar, which is a state condition previously reported by Poole *et al.* [*J. Chem. Phys.* **138**, 034505 (2013)] to exhibit liquid-liquid coexistence. Contours are 1kT. [Note that the free energy surface below was computed from relatively short simulations for the sole purpose of verifying that a low-density liquid basin is observed under these conditions. Thus, the free energy has not been computed to a high degree of accuracy (i.e., the error is >1 kT )]



3. Calculation of the free energy surface for *N*=216 for the ST2 variant employing the reaction field method at 228.6 K and 2400 bar, which is a state condition previously reported to exhibit liquid-liquid coexistence [*Nature* 510, 385 (2014)]. Contours are 1kT. [Note that the free energy surface below was computed from relatively short simulations for the sole purpose of verifying that a low-density liquid basin is observed under these conditions. Thus, the free energy has not been computed to a high degree of accuracy (i.e., the error is > 1 kT )]



**Inputs and Outputs:**

**Input**

**1. inputfile (Mantatory)**

**a. Mandatory Lines**

**Line 1:** ***NVT*** or ***NPT*** [NVT: NVT Hybrid MC, NPT: NPT Hybrid MC]
**Line 2:** $n_{proc}$ [Number of processor cores requested]
**Line 3**: ***T*** [K]
**Line 4:** ***P*** [bar]
**Line 5:** **st2.mol** [Name of molecular structure/parameter file]
**Line 6:** $t_{MD}$ [MD time step, ps], $n_{MD}$ [# of MD steps/HMC move], $dV_{max}$ [max log vol. displacement]

(*Note: all three are required, regardless of the simulation type. Use only spaces to separate. Ratio of HMC to volume perturbation MC moves for NPT ensemble is hardcoded as 2:1 (see trial_move.f90))*

**Line 7:** $n_B$ [Total # of MC Blocks], $n_{EB}$ [# of Equilibration Blocks], $n_{MCS}$ [# of MC Sweeps per block]

*(Note: all three are required, regardless of the simulation type. Use only spaces to separate. The total simulation duration will be* $n_{MCS,total} = (n_B \times n_{MCS})$*, where 1 MC Sweep = 1 attempted HMC or volume perturbation move; Definition of block are physically meaningless, they are just used to organize output and track the progression of the simulation)*

**b. Optional Commands (Lines 8 to N)**

i. ***dump_trajxyz***

Write a .xyz style trajectory file containing each atom on each water molecule (see **Output** section below). Frames are written with a frequency specified in terms of MC Sweeps on line immediately after the *dump_trajxyz* command. Thus, the following commands would tell the program to write a new frame every 10 MC sweeps:

**Line i**: dump_trajxyz
**Line i+1**: 10

ii. ***ewald_vacuum***

Treat long-range electrostatic interactions using the standard Ewald summation technique with vacuum boundary conditions. The relative precision of the summation is specified immediately after invoking this command. For example:

**Line i**: ewald_vacuum
**Line i+1**: 0.0001

Specifies a precision, *p* = 1e-4, or 1 part in 10,000 of the energy. The remaining Ewald parameters are determined using the following formulas:

$$\kappa = \frac{1.35 - 0.15\ln(p)}{r_{cut}}, \quad k_{\max} = \frac{\kappa L_{box}}{\pi}\sqrt{-\ln(p)}$$

where $\kappa$ is the width of the Gaussian screening function, $r_{cut}$ is the intermolecular cutoff distanced, $k_{\max}$ is the maximum number of wave vectors along each dimension, and $L_{box}$ is the edge length of the simulation cell.

***Important:*** If this option is not invoked, the reaction field version of ST2 described by Poole *et al*. [J. Chem. Phys. **138**, 034505 (2013)] is used by default. In this case, the oxygen-based cutoff distance specified in st2.mol (see below) is overridden and set to 7.8 Å and long-range dispersion corrections are automatically applied to the volume perturbation moves.

iii. ***ljlrc***

Apply standard long-range correction the Lennard-Jones interactions. Note that this is only applied to the volume perturbation moves – it is not tallied in the configurational energy.

iv. ***weighted***

Apply a harmonic umbrella bias to density ($\rho$) and $Q_6$. This command is immediately followed by an integer number specifying the frequency (in MC Sweeps) for writing density and $Q_6$ to file (see **Output** section below). The following example will apply the umbrella weighs and dump density and Q6 values to file every 50 MC sweeps.

**Line i**: weighted
**Line i+1**: 50

The umbrella sampling parameters for $\rho$ and $Q_6$ are read for each processor core from files 'rho' and 'Q6', respectively. Parameters are assigned to processor cores by rows in a sequential fashion. The number of requested cores must therefore match the number of entries (rows) in each file. Hence, core 1 reads from row 1, etc. The first column in each row contains the spring constants (either $k_\rho$ [Å$^6$/molecule$^2$] or $k_{Q6}$ [dimensionless]), while the second column contains the window's center (either $\rho^*$ [molecule/Å$^3$] or $Q_6^*$ [dimensionless]). Note that the spring constantans are multiplied by the factor kT inside the program. The "Q6" file, for example, would have the following format:

Line 1: 2000 0.05
.
.
.
Line $n_{proc}$: 3000 0.07

Thus, the first and last processors are assigned values of 2,000 and 3,000 for $k_{Q6}$ and 0.05 and 0.07 for $Q_6^*$ respectively. The format for the "rho" file is identical.

**2. configs (Mantatory)**

The file containing a list of initial configuration files that are assigned to the processor cores in a sequential fashion. Thus, the configuration file on row 1 is assigned to processor core 1, etc.

Each initial configuration file listed in **configs** has the following format

Line 1: *N* (i.e., # Molecules (1 integer))
Line 2: boxx,boxy,boxz (3 floats, x,y,z dimension of the unit cell in Å)
Line 3: "H2O", rx_1, ry, rz, q0, q1_1, q2,
.
.
.
Line N+2: "H2O", rx, ry, rz, q0, q1, q2, q3

"H2O" = String for molecule's name (must match name in .mol file, see below)
rx, ry, rz = Cartesian coordinates in Å for the center of mass positions of the molecule (3 floats)
q0,q1,q2,q3 = Elements of the unit quaternion describing the molecule's orientation (4 floats)

**3. st2.mol (Mandatory)**

The file specifying the geometry and potential parameters for the ST2 model. The geometry is specified in the body-centered frame with respect to the center of mass, which is defined to be located at the origin. This information is used in conjunction with the quaternions to compute the position of each atom on each molecule in the laboratory-centered frame during the simulations. Note that the oxygen-based cutoff specified in this file is overridden and set to 7.8 Å if the ***ewald_vacuum*** option is not invoked in **inputfile.** This is done so that the program defaults to the reaction field version of ST2 described by Poole *et al.* [J. Chem. Phys. **138**, 034505 (2013)]. The format of st2.mol is as follows:

| | |
|---|---|
| 'H2O' | [molecule name, max 4 chars] |
| 18.01528 | [molecular mass, g/mol] |
| 7.5 | [oxygen-based cutoff in Å; overridden to 7.8 Å for reaction-field ST2] |
| 5 | [number of interaction sites on each molecule] |
| 'O' | [name of 1$^{st}$ site, max 4 chars] |
| 15.9994 | [mass of 1$^{st}$ site, g/mol] |
| 0.000 -0.064604509 0.000 | [coordinates of 1$^{st}$ site relative to center of mass at origin] |
| 3.10 38.144 0.000 0.00 | [parameters for 1$^{st}$, sigma (Å), epsilon (K), dummy, charge (*e*)] |
| . | |
| . | |
| . | |
| 'Q' | [molecule name, max 4 chars] |
| 0.00 | [mass of 5$^{th}$ site, g/mol] |
| -0.6532 -0.526484509 0.000 | [coordinates of 5$^{th}$ site relative to center of mass at origin] |
| 0.000 0.000 0.000 -0.2357 | [parameters for 1$^{st}$, sigma (Å), epsilon (K), dummy, charge (*e*)] |

***Important:*** The st2.mol file contains internal coordinates specifying the position of each interaction site relative to the molecule's center of mass. There are three important restrictions: (1) the center of mass must be placed at the origin, (2) the body-centered coordinate system must be defined such that the off-diagonal elements of the inertial tensor are zero, and (3) molecule must be oriented such that the diagonal components of the inertial tensor ($I_{xx}$, $I_{yy}$ and $I_{zz}$) obey $I_{xx} > I_{yy} > I_{zz}$.

**Output**

**1. Screen output (for processor core 1 only)**

The following information is printed to screen after every block:

Block #, instantaneous configurational energy (J/mol), instantaneous density (g/cm$^3$), acceptance rate for HMC moves for the block, acceptance rate for volume perturbation moves for the block

**2. avgboxX.out**

File containing the block averaged thermodynamic data. A separate file is written for each processor core once all the simulations finish, substituting "X" with the processor core number. The quantities and units are listed in the header.

**3. config_bX_BLOCK.xyz**

File containing an instantaneous configuration in the quaternion-based format described for the input configuration files. A separate file is written for each processor core, replacing "X" with the processor core number. A new file is printed out every 1000 blocks, which is hardcoded in *mc_opt.f90*, with the current block number substituted for "BLOCK". These files serve as useful checkpoints for restarting simulations.

**4. config_bX_BLOCK_exp.xyz**

File containing an instantaneous configuration in an explicit all-atom format, which provides the Cartesian coordinates for each atomic site on each water molecule. A separate file is written for each processor core, substituting "X" with the processor core number. A new file is printed out every 1000 blocks, which is hardcoded in *mc_opt.f90*, with the current block number substituted for "BLOCK". These files are useful for visualization.

**5. restart_bX.xyz**

File containing an instantaneous configuration in the quaternion-based format described for the input configuration files. A separate file is written for each processor core, substituting "X" with the processor core number. The current files are overwritten with the new configurations every 10 blocks, which is hardcoded in *mc_opt.f90*. These files serve as useful checkpoints for restarting simulations.

**6. trajX.xyz (if *dump_trajxyz* option is invoked)**

Trajectory file containing configurations in an explicit all-atom format. The first line of each frame contains the total number of atoms, the total number of molecules, the block number, the MC sweep number (for the current block) and the umbrella bias energy (J/mol). This is followed the x,y,z unit cell dimensions in Å and then the Cartesian coordinates of the atoms in Å in standard xyz format. A separate file is written for each processor core, substituting "X" with the processor core number. These files are useful for post-simulation analysis.

**7. cvsX.out (if *weighted* option is invoked)**

Trajectory file containing instantaneous properties and values of rho and $Q_6$. A separate file is written for each processor core, substituting "X" with the processor core number. These files are useful for post-simulation analysis, including free energy calculations. The format of each line of the file is as follows:

Block #, MC Sweep # (for the current block), $U$ [J/mol], $N$, $T$ [K], $\rho$ [molecule/ Å$^3$], $k_\rho$ [Å$^6$/molecule$^2$], $\rho^*$ [molecule/ Å$^3$], $Q_6$ [dimensionless], $k_{Q6}$ [dimensionless], $Q_6^*$ [dimensionless]