How to run TVREX:

Intro: TVREX is a novel algorithm developed by the Pomes lab that has higher sampling efficiency than conventional MD and lower computational cost than replica exchange.

The DRSSS code is used to run TVREX. It basically consists of a command script, a server, and multiple clients. The server interprets the command script and creates multiple clients, who each communicates with the server to determine whether swapping of temperatures is needed or not. For a more thorough introduction to TVREX and DRSSS, please see the Elastin TVREX report.

Files needed to run TVREX:

To run TVREX using DRSSS, there are multiple files needed in one directory. Minimal to no changes are needed in most files (see commented scripts themselves for details), except for the FILES directory and the .script file (aka command script). To determine how the different scripts interacts with each other, please see the schematic diagram.

Start.sh/restart.sh – the only script that the user submits to the supercomputer. Starts/restarts the whole process. Minimal changes necessary.

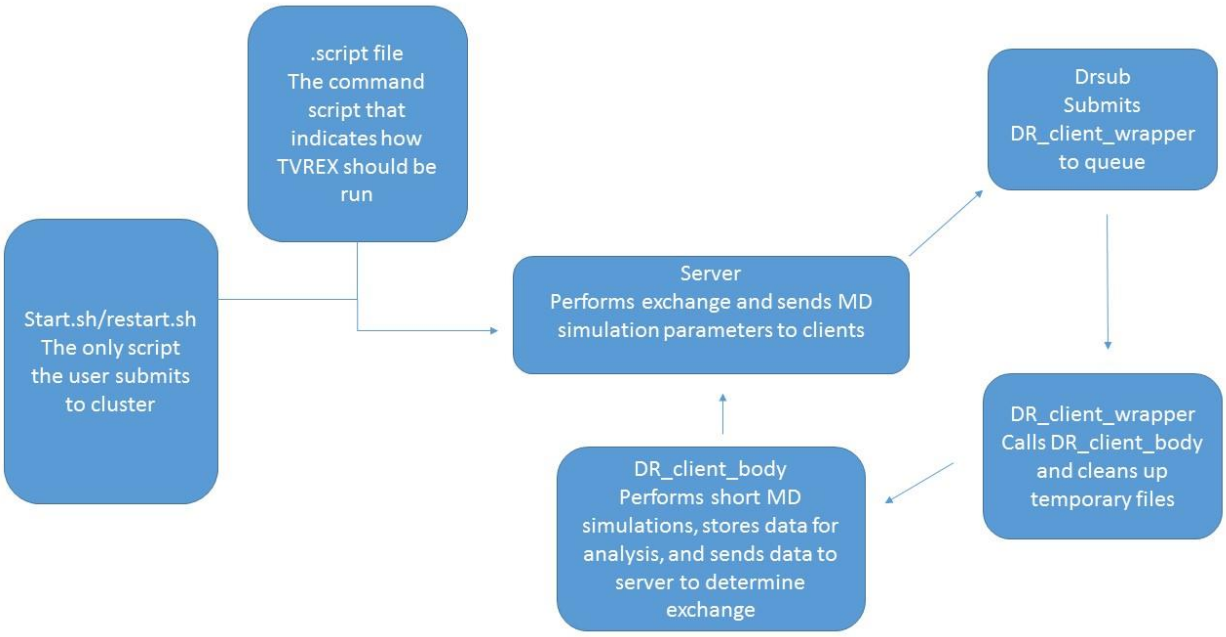
Drsub – the script that submits clients to the supercomputer. No change is necessary.

DR_client_wrapper – the client wrapper that creates files and calls DR_client_body and later cleans up temporary files. Minimal changes necessary.

DR_client_body – the client script that performs short MD simulations in between exchanges of temperature. The script performs grompp and mdrun_mpi each time and saves xtc files for analysis and potential energy files (.force files) that determine whether swapping of temperatures is needed. The user can also save additional analysis files called .add# files here (for more information, please see DR_guide.pdf). Minimal changes necessary.

.script file – the command script initially interpreted by the server. Mainly contains “meta data” information about TVREX holistically (eg. duration of simulation, rate of exchange) as well as “meta data” information about short MD simulations performed by the clients (number of initial conformations, list of temperatures to swap, length of short MD simulations). Major changes necessary to this file. For information on all the possible variables to include in this script, please consult DR_guide.pdf. For a minimal working example, see e0.script.

FILES directory – the directory where all the files needed to perform MD simulations is stored. This includes an .mdp file template (ie. variables substitute parameters that constantly change throughout TVREX, which in this case is mostly TEMPERATURE; please see elastin.mdp for reference), all the initial conformation gro files (stored in a GRO subdirectory, and each named start#.gro, where the range of the number is how many total replicas will be submitted according to the command script), and the topology files (as well as any files called by the topology files). The user needs to “manually” create this directory from scratch



Running TVREX:

Once all the files and parameters are tweaked, please add the following code to your .bashrc file in the $HOME directory:

export LD_LIBRARY_PATH=/scinet/gpc/mpi/openmpi/1.4.4-intel-v12.1/lib/:/scratch2/p/pomes/wuzhen3/GROMACS/gromacs-4.5.5/exec/lib/:$LD_LIBRARY_PATH

export GMXLIB=/scratch2/p/pomes/wuzhen3/GROMACS/gromacs-4.5.5/exec/share/gromacs/top

Then, all the user needs to do is to submit the start.sh/restart.sh script to the supercomputer. TVREX would then automatically run on the cluster until completion, PROVIDED that a server is always running on the cluster to connect with clients (see Troubleshooting). Data output would be tarred and stored in the output/data directory. A snapshot of the current condition of TVREX would be saved every 12 hours (can be modified in command script) so that the user can restart TVREX starting at a particular snapshot via restart.sh instead of running everything from the beginning.

If the user restarts TVREX from a certain snapshot, all the previous data obtained since the last restart or start script was submitted would be moved to a PRIOR_TO_RESTART directory. The contents of interest in that directory mainly exists in the output directory and your “username” directory.

Organizing XTC Data Output from TVREX:

The DRSSS framework can allow the user to analyze data on the go (ie. before simulation is finished) and store the analysis output in .add# files. However, the minimal working example of

TVREX does not include this. For more information, please consult DR_guide.pdf or contact the original authors.

This section will focus on analyzing xtc file output from TVREX. The xtc files for each replica is scattered across all the tarred files in output/data. Each xtc file in those tar files contains just a single frame of the replica. Thus, to analyze all xtc data for one replica, a few steps need to be made:

First, untar all tar files into one directory. The preferred directory is the "username" directory created by TVREX, located in the main directory of TVREX. To analyze a replica of interest, grep all the xtc files by the replica name (ie. e0w# in the example given). You should now have all the output xtc files for one replica. This step is already automated by the **extractXTCs.sh** script, which organizes all xtc files in their respective replica directory.

To retrieve the xtc files of a given replica at a given temperature, simply grep for the respective temperature out of all the filtered xtc files. Please note the floating point representations of temperatures in the file names when grepping (eg. 348 K may be expressed as 348.001 or 347.999). You should now have all the necessary xtc files for your analysis. Concatenate all xtc files into one via trjcat and you're good to go for further analysis!

Testing Temperature Exchange Rate and Temperature Convergence:

For TVREX to achieve maximal sampling efficiency, a random walk in temperature needs to occur and the temperature exchange rate needs to be somewhere between 0.2-0.8. To test both of these occurrences, simply untar all the xtc files and look at the distribution of temperatures for each replica (via the temperature parameter in the xtc file names) for the former and look at the frequency of changes in temperatures for each replica as sequence number increases for the latter. These processes have been automated by the **gettempdistribution.sh** and **getexchangerate.sh** scripts. The scripts depend on the xtc files to be already organized by replica directories done by **extractXTCs.sh** and performs analysis to all replicas by a for loop. To perform analysis on just a few of these replica directories, simply modify the directories traversed by the for loop.

Troubleshooting TVREX:

Sometimes when checking output tar files, you might see the tar files names have FAILURE appended to the end of it. Whenever this happens, please stop TVREX by the **make_snapshot.sh** file, kill all the queued jobs related to TVREX, and restart TVREX at a time point before these FAILURES occurred (check time stamp via ls –lrt).

The reason these FAILURES occur is because of "magic number errors", which occur ultimately due to cluster instability (files needed for downstream processes not written completely to user directory, causing corruption of file and termination of TVREX).

Another thing to watch out for is to ensure that one server is always running on the cluster when clients are queueing on the scheduler. If no server is running, then the whole process is terminated even though clients are still queueing to run. When that happens, TVREX will be abnormally terminated. To recover from this, restart TVREX at the latest snapshot file available, but first ensure all queued jobs related to TVREX are killed!

The reason this occur is because the queue time for clients to run is longer than the maximum walltime allocated to the server. Thus, even if the server is allowed to move to nodes with longer walltime remaining, TVREX will still be abnormally terminated if the client queue time is too long.