

Usage of OpenStack Virtual Machine and MATLAB HPC Add-on leads to faster turnaround

Matthijs van Waveren, Antonio Arena, Arindam Chakraborty and Stanislav Flachs
IT Research Computing, King Abdullah University of Science and Technology

Maged Serag and Satoshi Habuchi

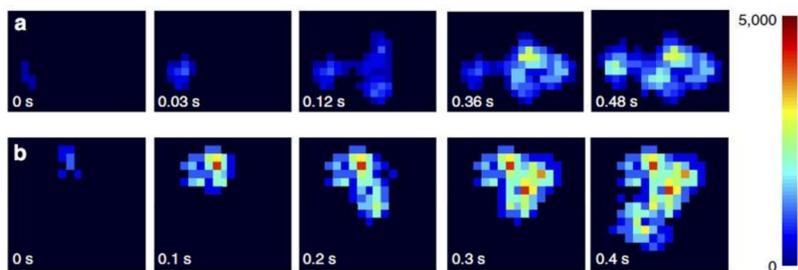
Lab for Single Molecule Spectroscopy and Microscopy, Department of Bioscience, King Abdullah University of Science and Technology

SUMMARY

We need to run hundreds of MATLAB® simulations while changing the parameters between each simulation. These simulations need to be run sequentially, and the parameters are defined manually from one simulation to the next. This makes this type of workload unsuitable for a shared cluster. For this reason we are using a cluster running in an OpenStack® Virtual Machine and are using the MATLAB HPC Add-on for submitting jobs to the cluster. As a result we are now able to have a turnaround time for the simulations of the order of a few hours, instead of the 24 hours needed on a local workstation.

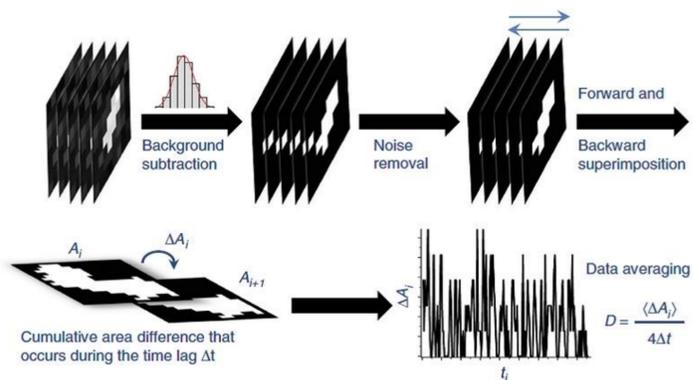
NUMERICAL PROBLEM

We have developed a method for measuring single-molecule diffusion that is superior to the traditional single-molecule localization and tracking (SMLT) method. Our method quantifies diffusion by analyzing the increase of the cumulative area (CA) occupied by the molecule in space over time. This is illustrated in the figure below.



Our method generates a great number of images. The images consist of up to 100 molecules in each data set and up to 20 datasets. Each single molecule movie contains up to 10,000 frames. From experimental data we fit the motion to a model.

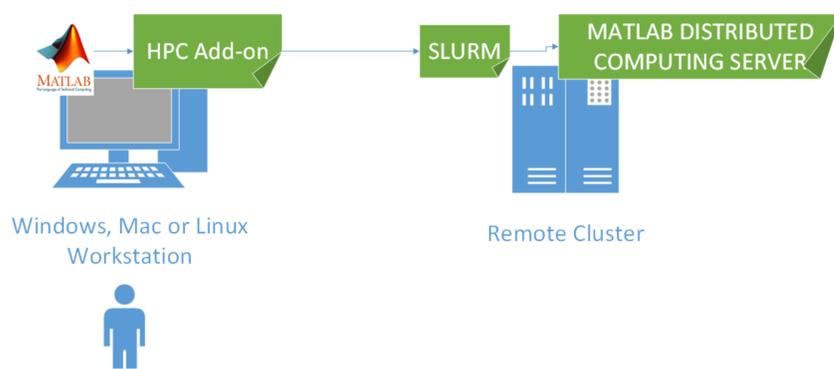
In parallel to the experiments, we simulate the motion of up to 1000 molecules, with 10,000 frames for each molecule. Using our new multi-parametric method, we need to change the simulation parameters up to 100 times before a final result is reached. Then from the final simulated data we fit the motion to a model. The workflow of our CA method is shown below.



Our type of workload is unsuitable for a large cluster with a scheduling system. We need to wait too long for the result of each simulation. For this reason we implemented an OpenStack Virtual Machine to run the simulations on, and an HPC Add-on to MATLAB in order to easily submit the calculations from our Windows workstation.

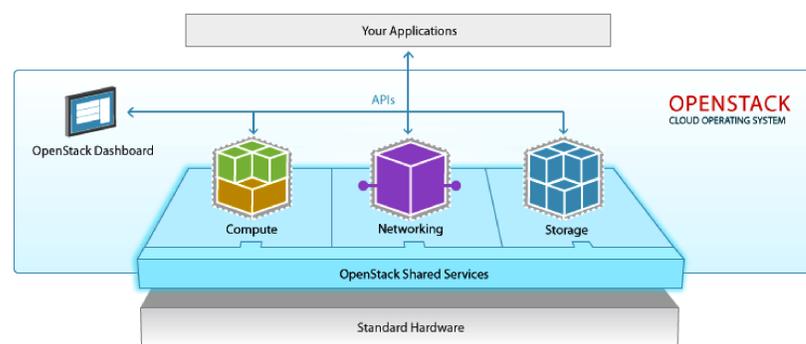
IMPLEMENTATION

We developed in collaboration with MathWorks® a software interface between MATLAB running on the user's workstations and the schedulers and MATLAB Distributed Compute Server (MDCS) running on the remote resource. The software interface, also called an HPC Add-on, is written in MATLAB code, and takes care of opening a session to the remote resource, creating and sending a job submission script to the scheduler and monitoring the status of the job.



When running on Noor II, we manage to run our jobs in 15 minutes instead of 24 hours using 512 cores, however due to the waiting times in the queues, the wall clock turnaround time is still too long for our multi-parametric method.

We then built a Virtual Machine (VM) using OpenStack on an Intel Xeon Sandy Bridge system with 6 nodes of 16 cores each. This VM includes the scheduler Simple Linux Utility for Resource Management (SLURM®) and MDCS. Our MATLAB HPC add-on was adapted to cover this VM. With this VM, we have a turnaround time of a few hours, depending upon input parameters.



RESULTS & BENEFITS

By submitting our parameterized jobs to the OpenStack VM, we are now able to have a turnaround time for the simulations of the order of a few hours depending upon input parameters instead of the 24 hours we had on our local workstation.

We are now able to perform multi-parametric simulations of molecular motion. The results of the simulations are compared to the experimental data.

This work is done in support of the development of the new multi-parametric single-molecule imaging method by the Laboratory for Single Molecule Spectroscopy and Microscopy.