**RELION/CryoEM questions:**

* *What exact version of RELION did you pull and build?*
  + *You can get this via “relion\_refine\_mpi –version”*

RELION version: 3.1-beta-commit-9090fd

Precision: BASE=double, CUDA-ACC=single

* *Did you make any modifications to the RELION source? If so, what?*

No.

* *What compiler and compiler version was used to build RELION?*

Intel compilers for cpu/MPI e.g. icc/2018.3.222-GCC-7.3.0-2.30

CUDA for GPU e.g. CUDA/10.1.105-GCC-8.2.0-2.31.1

Built on different gcc but no obvious problems (what was available in module).

* *What CMAKE options were used to build the CPU and GPU versions of RELION?*
  + *Was the CPU version built for Skylake using the following build recommendations*:

After “module load intel” and “module load CUDA” (default versions as above), no further options were given to cmake.

SCARF is a heterogeneous cluster, so no attempt was made to optimise for a particular CPU. Using the generic queue means that jobs start faster. We will explore this further as part of M5 (e.g. looking at -DALTCPU=ON as below).

* *Were you able to reproduce the performance results for Plasmodium Ribosome published here*:

This dataset is not part of the project (we note that Empiar 10028 has 105k shiny particles, while we have 451k particles as input to Class3D). So far, we have used GPU for Class3D jobs. We can explore this further as part of M5.

* *When running 2D and 3D classification for the two data sets, was RELION run with the “--cpu” flag, and with the thread and rank recommendations documented here*:

We have not looked at CPU acceleration options yet. The focus has been on developing the end-to-end benchmark with appropriate scientific parameters for the example dataset, and reference results. In M5, we will look at such optimisations (including checking the correctness of the results).

* *What are the details for the data sets used in this study?*
  + *Number of images in the image stacks used for 2D and 3D classification*

749k particle images into Class2D, 451k particles into Class3D, and 219k particles in Refine3D

* + *Dimensions of the images in the image stack*

128x128 at 4x binning, 256x256 at 2x binning, 512x512 at original (as in published protocol). Early steps use binned particles to save memory. Original pixel size only necessary as resolution improves.

* + *The size of the benchmark data set to be distributed in the container (or however the end-to-end example pipeline will be distributed)*

The initial movie data for the aldolase dataset is split into two collection sessions, of total size 337GB and 329GB (details are in the first presentation sent on 21/7/20). Benchmarks can be run on one session, though data from both sessions are needed to get to the highest resolution. There are several intermediate datasets we can include e.g. 749k particles at 4x binning for one session is 58GB.

* + *Runtimes currently getting on Skylake and GPU platforms on various numbers of nodes.*

Class3D took around 8 hours on 4 GPUs. Refine3D took between 13 hours and 37 hours on 4 GPUs (this depends on number and size of particles used). A table of all runtimes is being prepared.

* *What CUDA version and GPU hardware are used for GPU runs?*

K80 cards, see <https://www.scarf.rl.ac.uk/GPUs.html> CUDA/10.1.105

* *Are you consulting with Sjors Scheres or any of his team on how to get optimal performance on CPU and GPU?*

We are working with Sjors on pipeline development, but have not discussed performance. Our aim has been to get a well-documented benchmark which can be re-used several times (i.e. as hardware and compilers evolve). We have recorded performance, but so far these are “typical” rather than optimised.