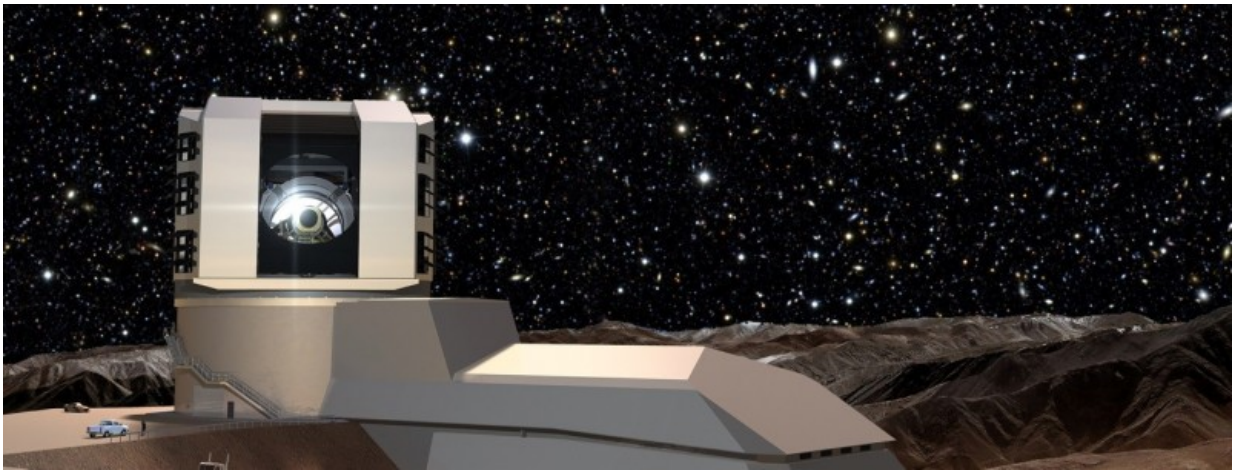# Shear brilliance: Computing tackles the mystery of the dark universe

November 24 2016

Credit: University of Manchester

Scientists from The University of Manchester working on a revolutionary telescope project have harnessed the power of distributed computing from the UK's GridPP collaboration to tackle one of the Universe's biggest mysteries – the nature of dark matter and dark energy.

Researchers at The University of Manchester have used resources provided by GridPP – who represent the UK's contribution to the computing grid used to find the Higgs boson at CERN – to run image processing and machine learning algorithms on thousands of images of galaxies from the international Dark Energy Survey.

The Manchester team are part of the collaborative project to build the Large Synoptic Survey Telescope (LSST), a new kind of telescope currently under construction in Chile and designed to conduct a 10-year survey of the dynamic Universe. LSST will be able to map the entire visible sky.

In preparation to the LSST starting its revolutionary scanning, a pilot research project has helped researchers detect and map out the cosmic shear seen across the night sky, one of the tell-tale signs of the dark matter and [dark energy](#) thought to make up some 95 per cent of what we see in the Universe. This in turn will help prepare for the analysis of the expected 200 petabytes of data the LSST will collect when it starts operating in 2023.

The pilot research team based at The Manchester of University was led by Dr Joe Zuntz, a cosmologist originally at Manchester's Jodrell Bank Observatory and now a researcher at the Royal Observatory in Edinburgh.

"Our overall aim is to tackle the mystery of the dark universe - and this pilot project has been hugely significant. When the LSST is fully operating researchers will face a galactic data deluge - and our work will prepare us for the analytical challenge ahead," said Sarah Bridle, Professor of Astrophysics.

Dr George Beckett, the LSST-UK Science Centre Project Manager based at The University of Edinburgh, added: "The pilot has been a great success. Having completed the work, Joe and his colleagues are able to carry out shear analysis on vast image sets much faster than was previously the case. Thanks are due to the members of the GridPP community for their assistance and support throughout."

The LSST will produce images of galaxies in a wide variety of frequency

bands of the visible electromagnetic spectrum, with each image giving different information about the galaxy's nature and history. In times gone by, the measurements needed to determine properties like cosmic shear might have been done by hand, or at least with human-supervised computer processing.

With the billions of galaxies expected to be observed by LSST, such approaches are unfeasible. Specialised image processing and machine learning software (Zuntz 2013) has therefore been developed for use with galaxy images from telescopes like LSST and its predecessors. This can be used to produce cosmic shear maps. The challenge then becomes one of processing and managing the data for hundreds of thousands of galaxies and extracting scientific results required by LSST researchers and the wider astrophysics community.

As each galaxy is essentially independent of other galaxies in the catalogue, the image processing workflow itself is highly parallelisable. This makes it an ideal problem to tackle with the kind of High-Throughput Computing (HTP) resources and infrastructure offered by GridPP. In many ways, the data from CERN's Large Hadron Collider particle collision events is like that produced by a digital camera (indeed, pixel-based detectors are used near the interaction points) – and GridPP regularly processes billions of such events as part of the Worldwide LHC Computing Grid (WLCG).

A pilot exercise, led by Dr Joe Zuntz while at The University of Manchester and supported by one of the longest serving and most experienced GridPP experts, Senior System Administrator Alessandra Forti, saw the porting of the image analysis workflow to GridPP's distributed computing infrastructure. Data from the Dark Energy Survey (DES) was used for the pilot.

After transferring this data from the US to GridPP Storage Elements,

and enabling the [LSST Virtual Organisation](#) on a number of [GridPP Tier-2 sites](#), the IM3SHAPE analysis software package ([Zuntz, 2013](#)) was tested on local, grid-friendly client machines to ensure smooth running on the grid. Analysis jobs were then submitted and managed using the [Ganga software suite](#), which is able to coordinate the thousands of individual analyses associated with each batch of galaxies. Initial runs were submitted using Ganga to local grid sites, but the pilot progressed to submission to multiple sites via the GridPP DIRAC (Distributed Infrastructure with Remote Agent Control) service. The flexibility of Ganga allows both types of submission, which made the transition from local to distributed running significantly easier.

By the end of pilot, Dr Zuntz was able to run the [image processing](#) workflow on multiple GridPP sites, regularly submitting thousands of analysis jobs on DES images.

Provided by University of Manchester