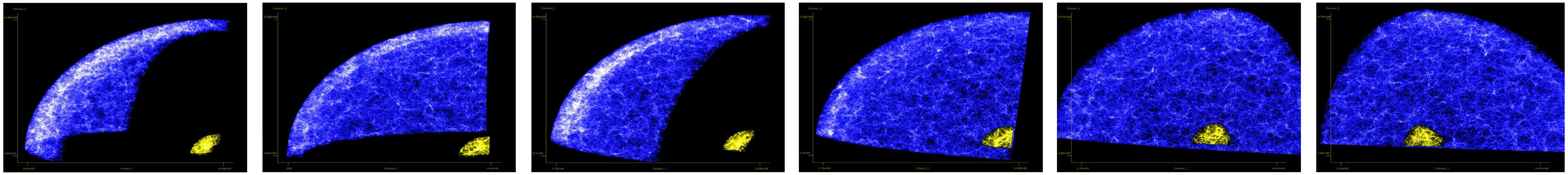
GPU-enabled Precision Measurements of the Structure of the Universe

Deborah Bard[1], Matthew Bellis[2], Christopher Godfrey[2]

1: SLAC National Accelerator Laboratory, Menlo Park, CA; 2: Siena College, Loudonville, NY



More than 1.5 million datapoints visualised in 3D using Viewpoints [1] on Nvidia GEForce GT 650M.

Measuring the structure of the universe

The past 20 years has seen the establishment of a new standard cosmological model, with a Universe dominated by dark matter and dark energy. Though constituting more than 95% of the mass-energy density of the Universe, the nature of dark matter and dark energy remain poorly understood. Future projects such as the Dark Energy Survey (DES) Large Synoptic Survey Telescope (LSST) will produce enormous datasets, with LSST characterizing tens of billions of galaxies. Calculating cosmological observables, such as correlation functions, over such vast datasets poses a significant computational challenge, with the requirements generally scaling with the square of the number of data points.

Cosmological measurements typically involve repeated calculations of the same quantities over many independent pairs (or groups) of objects, e.g. galaxies. As such, they are ideally suited to parallelization. We have recently implemented exact treatments of both the two-point correlation function and gravitational lensing aperture mass statistics onto GPUs [2]. Our work has shown that these calculations can be sped up by a factor of over 100 on GPUs compared to CPUs, whilst maintaining full accuracy. These algorithms can be parallelized yet further if they can be allocated to many GPUs simultaneously, allowing calculations over enormous datasets. This work compares existing fast approximation techniques (such as the kd-tree codes implemented by [3]) with the full calculations. We also take advantage of the scaling capabilities of GPUs to perform an in-depth exploration of systematic uncertainties; tests that can only be performed using the compute power of many GPUs.

Two-point angular correlation function on the GPU

The angular distance between two galaxies can be calculated using :

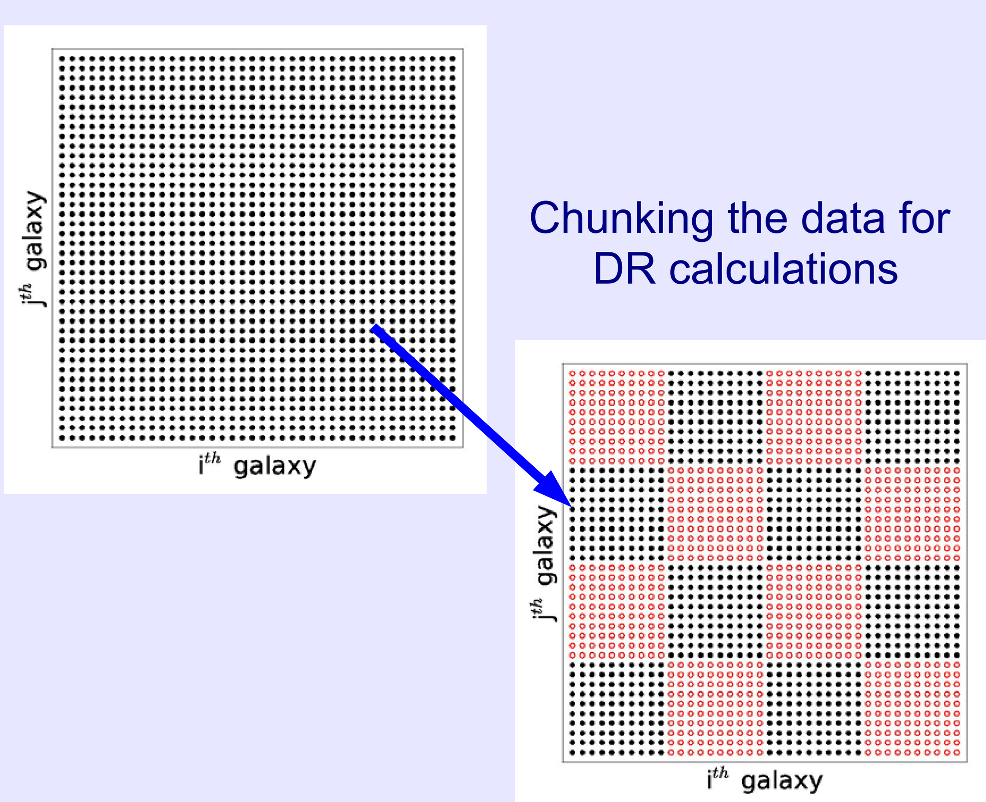
$$\theta = \frac{180^\circ}{\pi}\arctan\left(\frac{\sqrt{\cos^2\delta_2\sin^2(\alpha_2-\alpha_1)+[\cos\delta_1\sin\delta_2-\sin\delta_1\cos\delta_2\cos(\alpha_2-\alpha_1)]^2}}{\sin\delta_1\sin\delta_2+\cos\delta_1\cos\delta_2\cos(\alpha_2-\alpha_1)}\right)$$

where α and δ are the galaxy coordinates. The angular correlation function is calculated using the estimator [1]:

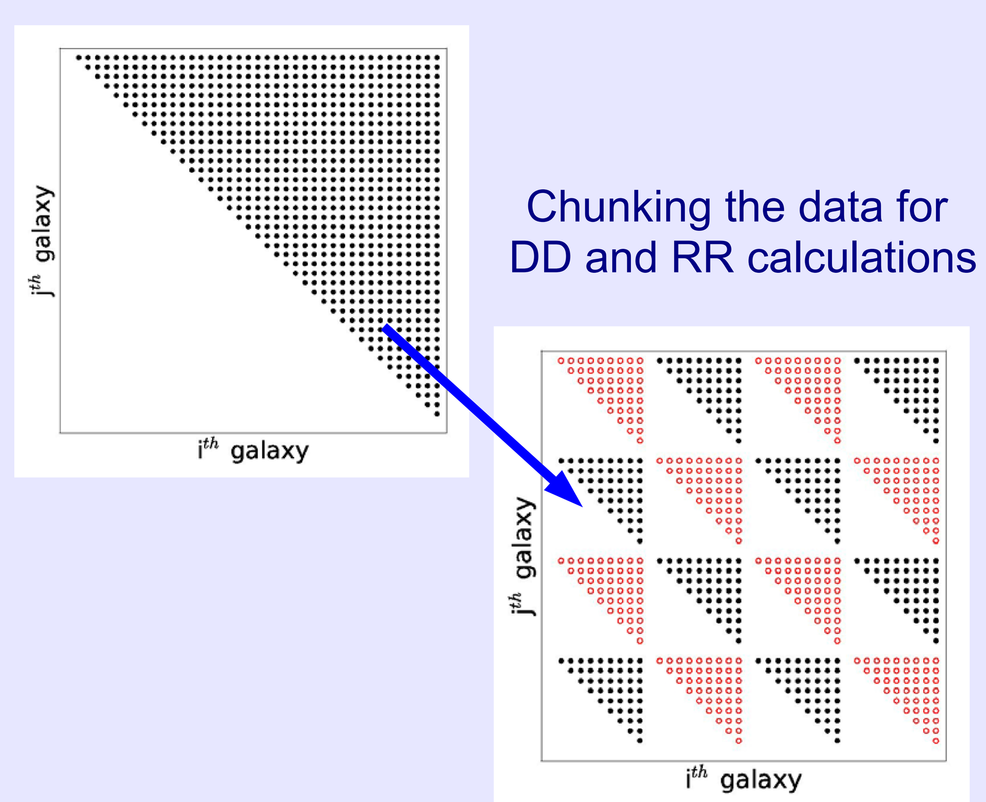
$$w(\theta) = \frac{DD - 2DR + RR}{RR}$$

where DD, DR and RR and pair counts of data-data, data-random and random-random objects, binned in angular separation θ.

The data and matrix of calculations are broken into chunks (see below), which are then passed to the GPU, distributed either as multiple kernel calls or to multiple GPUs for further parallelization. The histogramming is also performed on the GPU, with each thread given its own histogram array to increment. This array of histograms is then passed back to the CPU, where they are combined for the final answer.

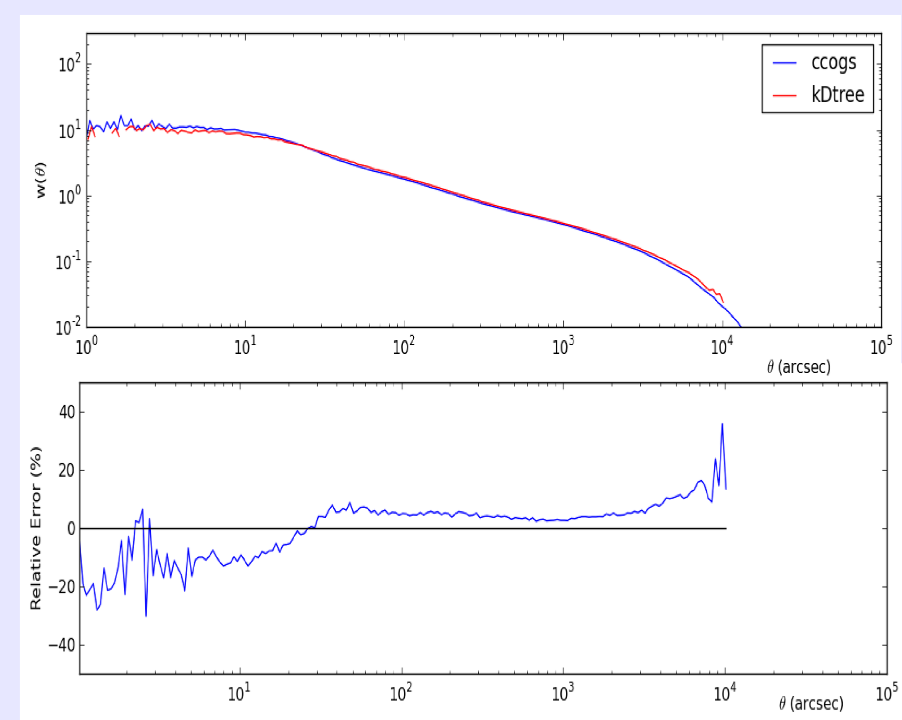


Chunking the data for DR calculations



Chunking the data for DD and RR calculations

Precision calculation through GPU computation

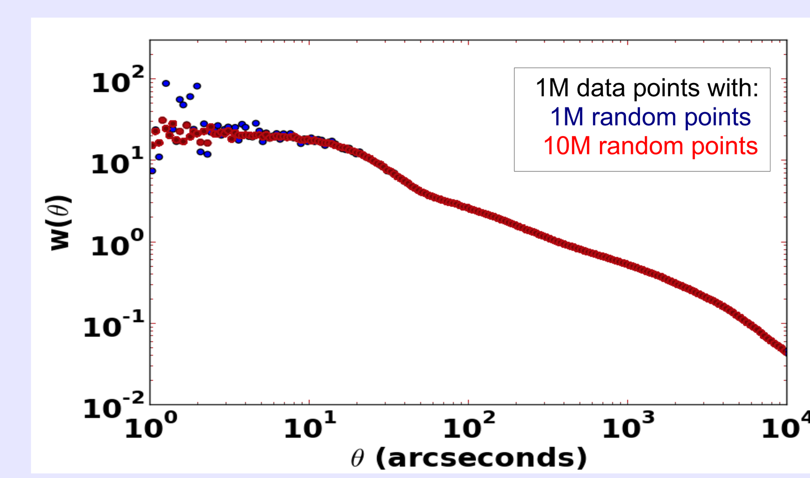


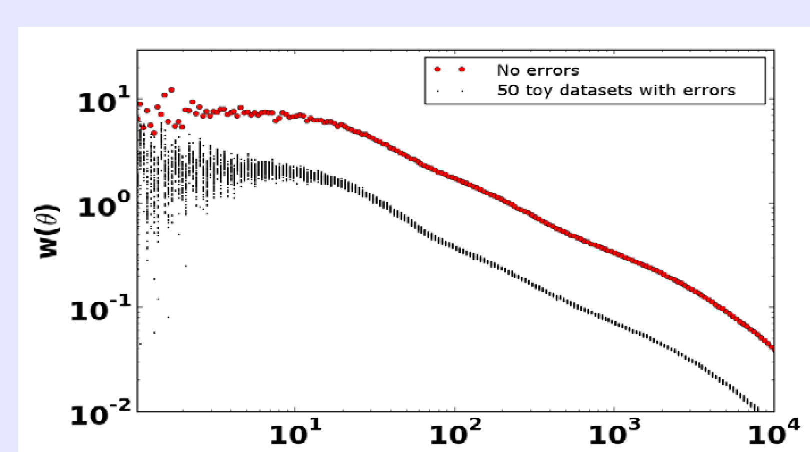
We can compare the full calculation made on the GPU with the approximation using the kDtree method (shown here for one redshift slice z=0.300–0.325). The kDtree approximation, although faster than the full calculation on the GPU, under-estimates the amount of structure in the universe at small distance scales by up to 20%, and over-estimates the structure at large scales by roughly 7%.

1.5M galaxies/5M random points. kDtree approximation - 94 min; GPU full calculation - 460 min.

The correlation function should be calculated using many times more random point than data points to reduce statistical error - the more random point the better, which increases the compute time significantly. GPUs enable us to make this kind of calculation in a reasonable time scale without sacrificing the accuracy of the calculation.

1M+1M: CPU full calculation - 128 hours; GPU (Fermi Tesla M2070) full calculation - 42 min.
1M + 10M: CPU full calculation - 324 days (extrapolated); GPU full calculation - 23 hours.





Using GPUs allows us to perform in-depth studies of uncertainties. Here we look at the effect on the angular correlation function of errors in redshift estimation, using 50 toy datasets produced by smearing simulated data by expected redshift errors. There is no analytic way to make these calculations – these effects must be estimated using numerical methods.

1.5M galaxies/5M random points, 50 toy studies: 907 trillion calculations in 332 GPU hours.

Ohio Supercomputer Center
This work was supported in part by an allocation of computing time from the Ohio Supercomputer Center.

https://github.com/djbard/ccogs

[1] Gazis, Levit & Way (2010) [2] Bard, Bellis et al, Astro. Comp. (2013), [3] Javis et al, (2004)