A GPU-based survey for millisecond radio transients using ARTEMIS

W. Armour, A. Karastergiou, M. Giles, C. Williams, A. Magro, K. Zagkouris, S. Roberts, S. Salvini, F. Dulwich and B. Mort.

Oxford e-Research Centre

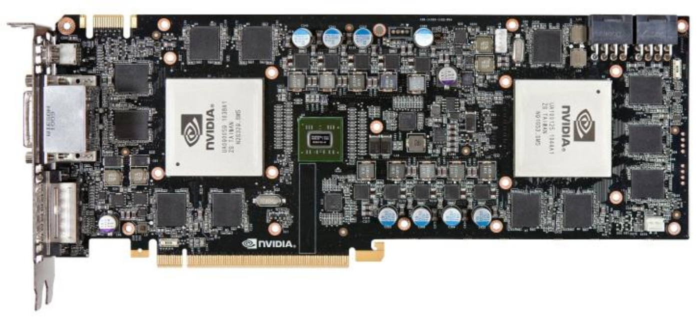
7th November 2011

Overview of current GPUs - Fermi

- Fermi released 2010
- 32/48 cores (stream processors or SP) per Streaming Multiprocessor (SM) (sm_20/sm_21)
- Configurable 16/48K or 48/16K L1 cache / shared memory (per SM)
- New L2 cache – 768K
- GigaThread – Concurrent kernel execution (16)
- Simple to use extensions - CUDA for C, C++ & FORTRAN

Latest Fermi based cards

- GeForce GTX 590 – 2x GF110 GPUs
- 6.8 GFLOPS / watt
- Price point ~ €600



- M2090 – 1x GF110 GPU, 16 SMs, 512 cores.
- 6.8 GFLOPS / watt
- Price point ~ € 2.5K

OXFORD e-Research CENTRE

Aims of our project…

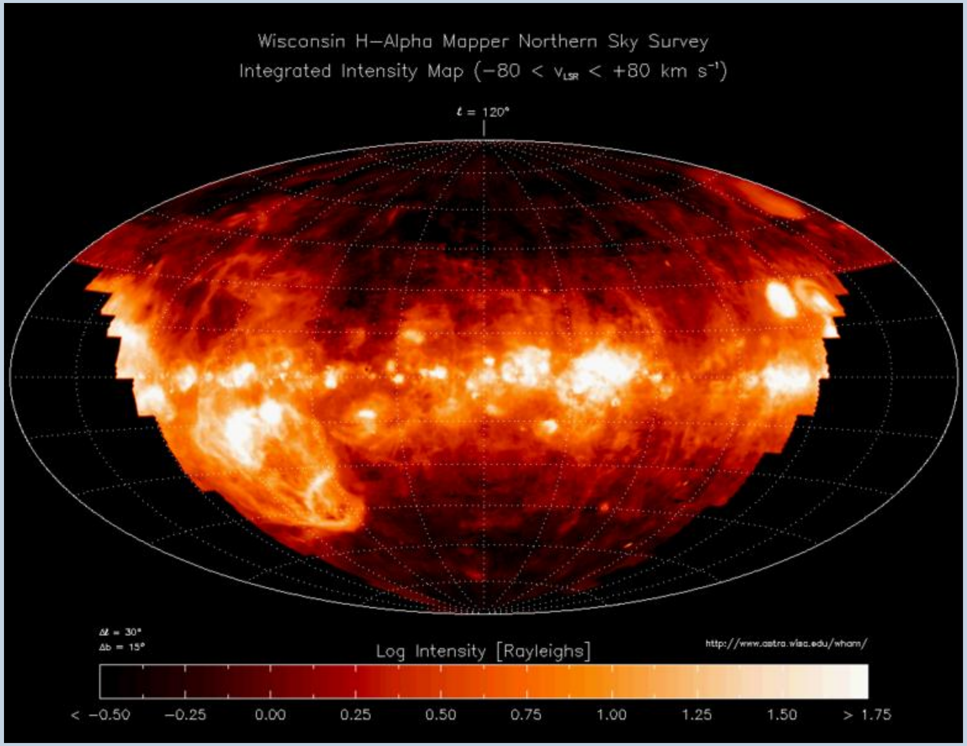
- Process a 3.2 Gb/s radio telescope data stream on a single GPU, de-dispersing and detecting transient events.
- Allowing for a vast reduction in the cost (capital/maintenance) of compute.
- Identify appropriate areas of parallelism on both CPU and GPU in the de-dispersion pipeline.

We have chosen to use the MDSM software as a wrapper to out GPU kernel

A. Magro, et.al. Monthly Notices of the Royal Astronomical Society, 417, 2642-2650.

Dispersion of Radio waves by the ISM

The interstellar medium (ISM) is the matter that exists between stars in a galaxy.



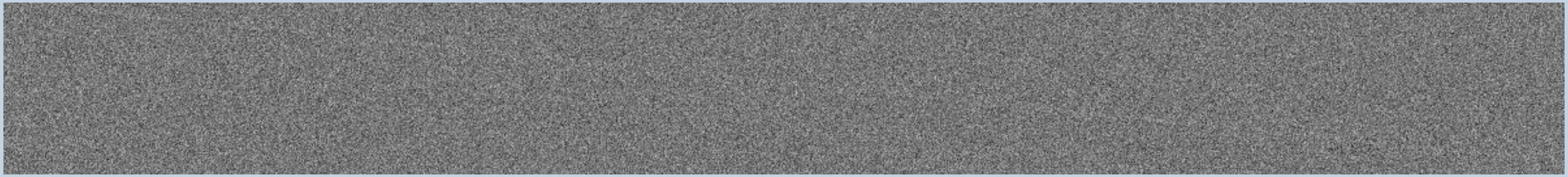
Haffner et al. 2003

In warm regions of the ISM (~8000K) electrons are free and so can interact with and effect radio waves that pass through it.

OXFORD e-Research CENTRE

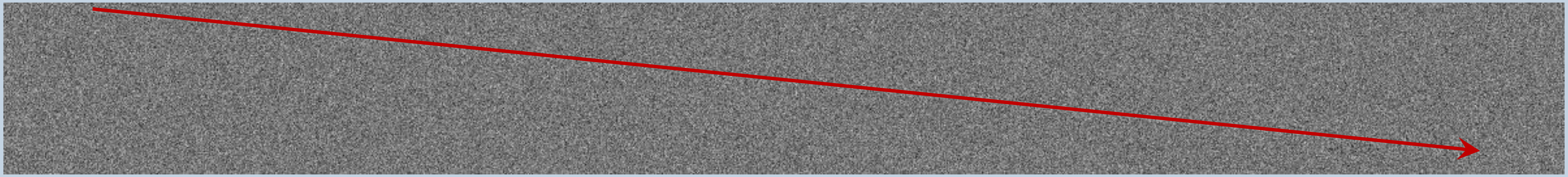
Experimental data

Most of the measured signals live in the noise of the apparatus.

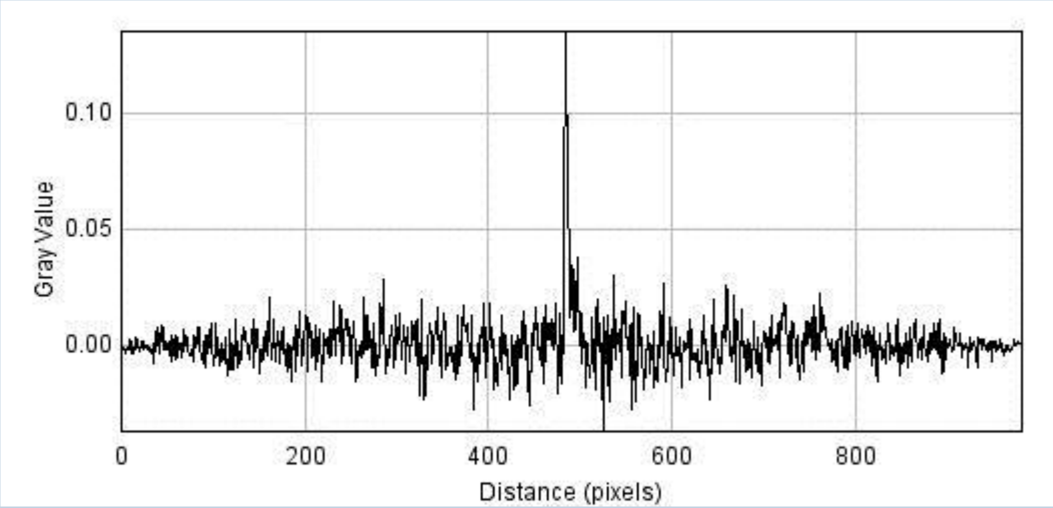


Experimental data

Most of the measured signals live in the noise of the apparatus.



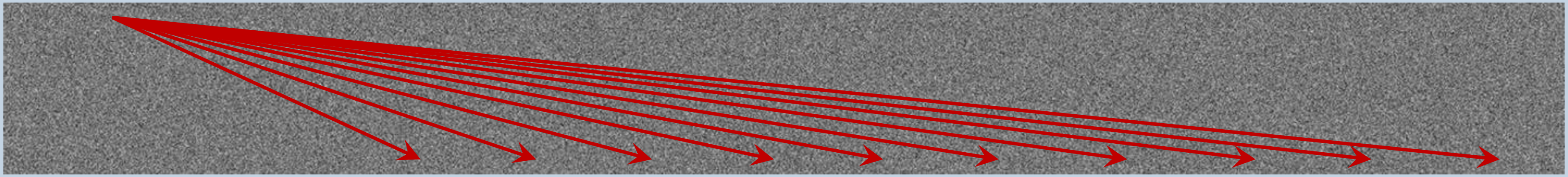
Hence frequency channels have to be “folded”



OXFORD e-Research CENTRE

Brute force algorithm

Every DM is calculated to see if a signal is present.



- In a blind search for a signal many different dispersion measures are calculated.
- This results in many data points in the (f,t) domain being used multiple times for different dispersion searches.
- This allows for data reuse in a GPU algorithm.

OXFORD e-Research CENTRE

A new brute force algorithm

Three key features…

- Each thread processes a variable number of dispersion measures in local registers.

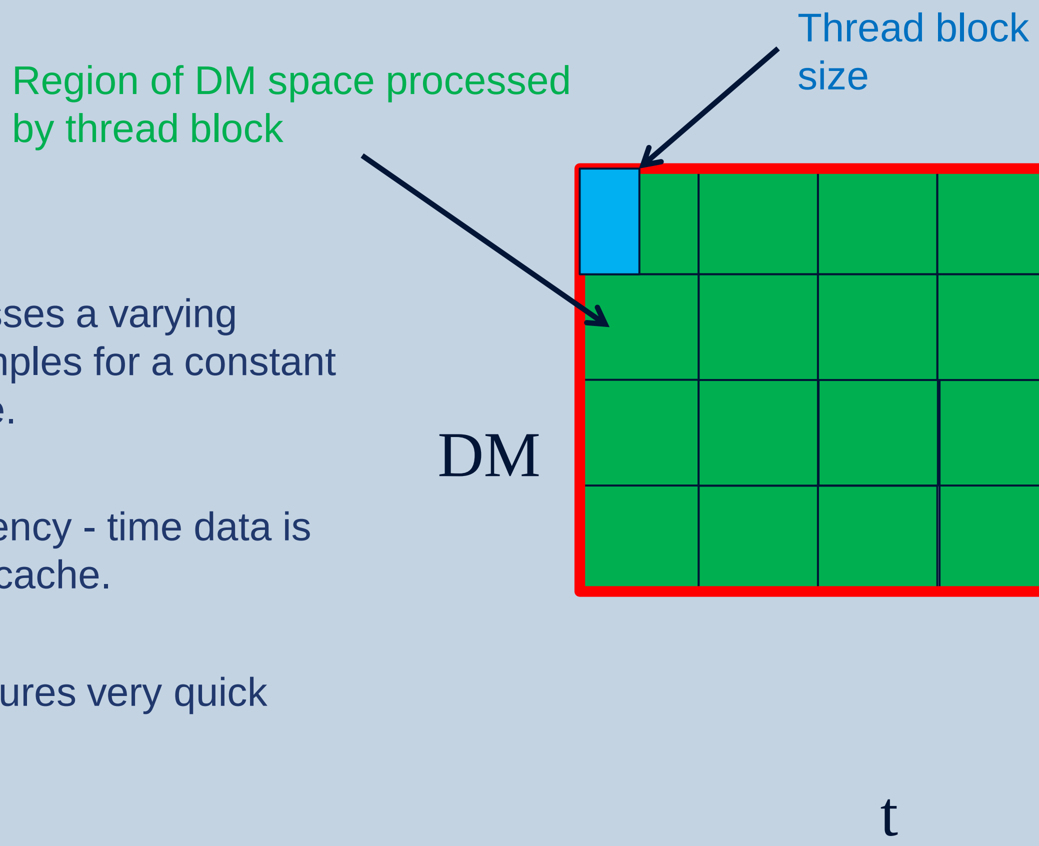
- Exploit the L1 cache present on the Fermi architecture.

- Optimise the region of dispersion space being processed (thread blocksize).

Web page : http://www.oerc.ox.ac.uk/research/wes

Processing several DM's per thread

New Algorithm works in the DM - t space rather than frequency – time space.

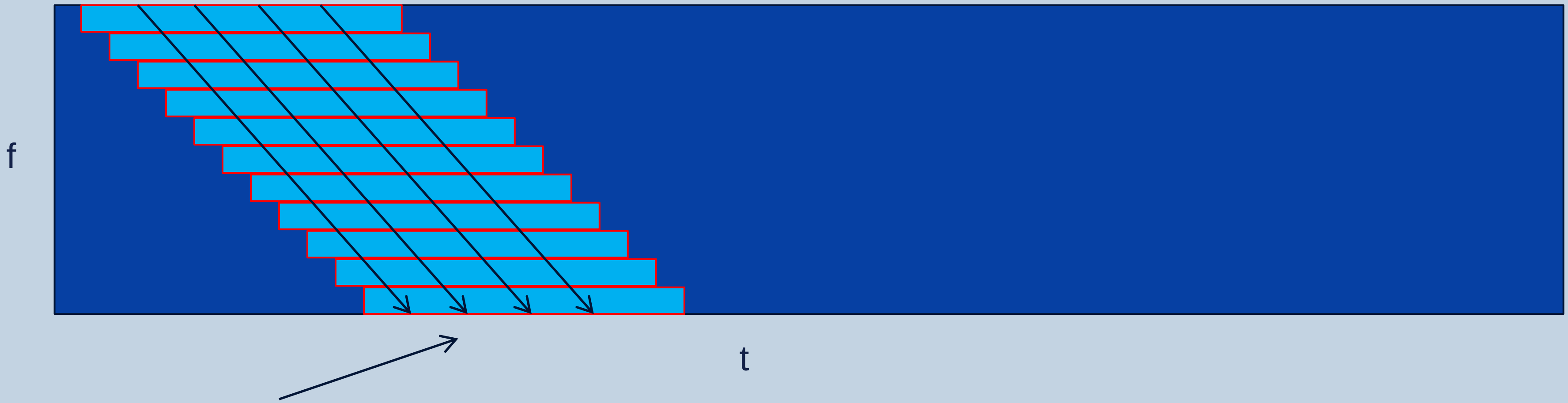


- Each thread processes a varying number of time samples for a constant dispersion measure.
- This ensures frequency - time data is loaded into fast L1 cache.
- Using registers ensures very quick memory access.

OXFORD e-Research CENTRE

Exploiting the L1 cache…

Each dispersion measure for a given frequency channel needs a shifted time value.

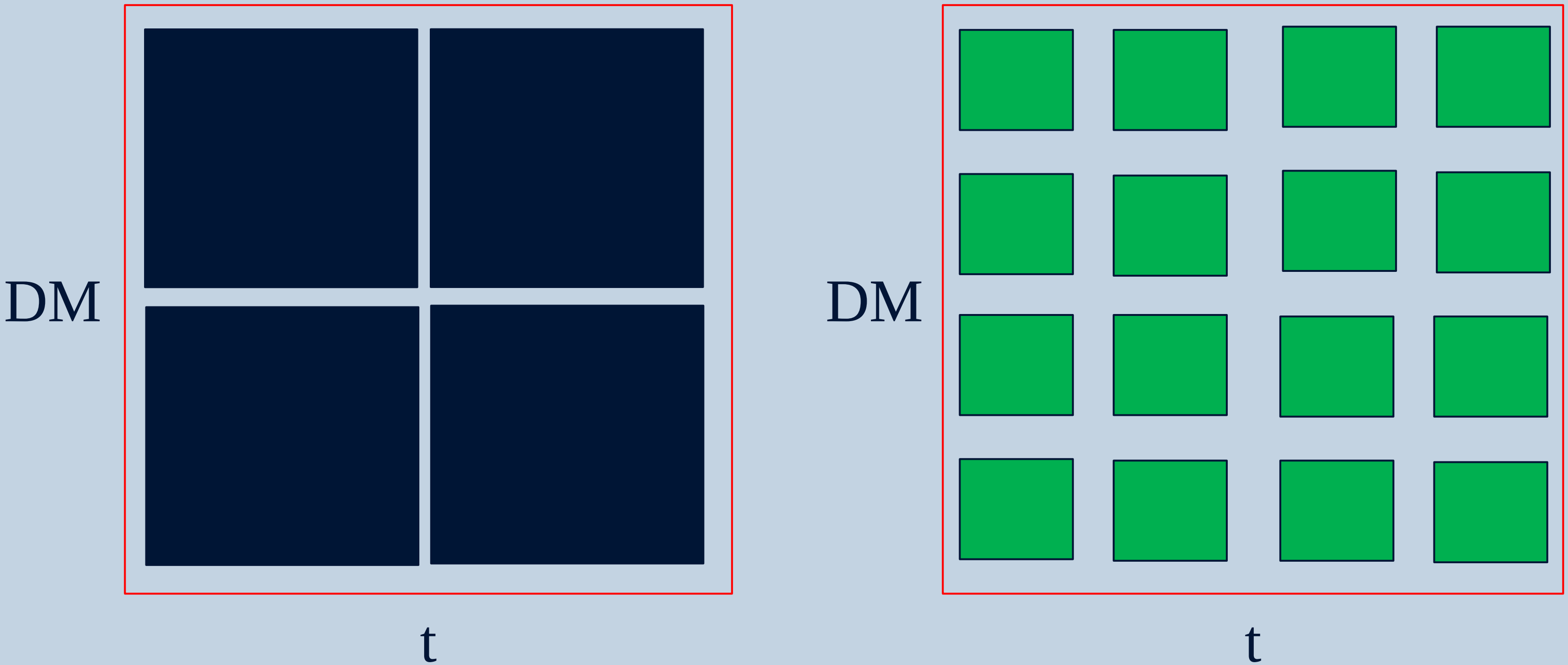


Constant DM's with varying time.

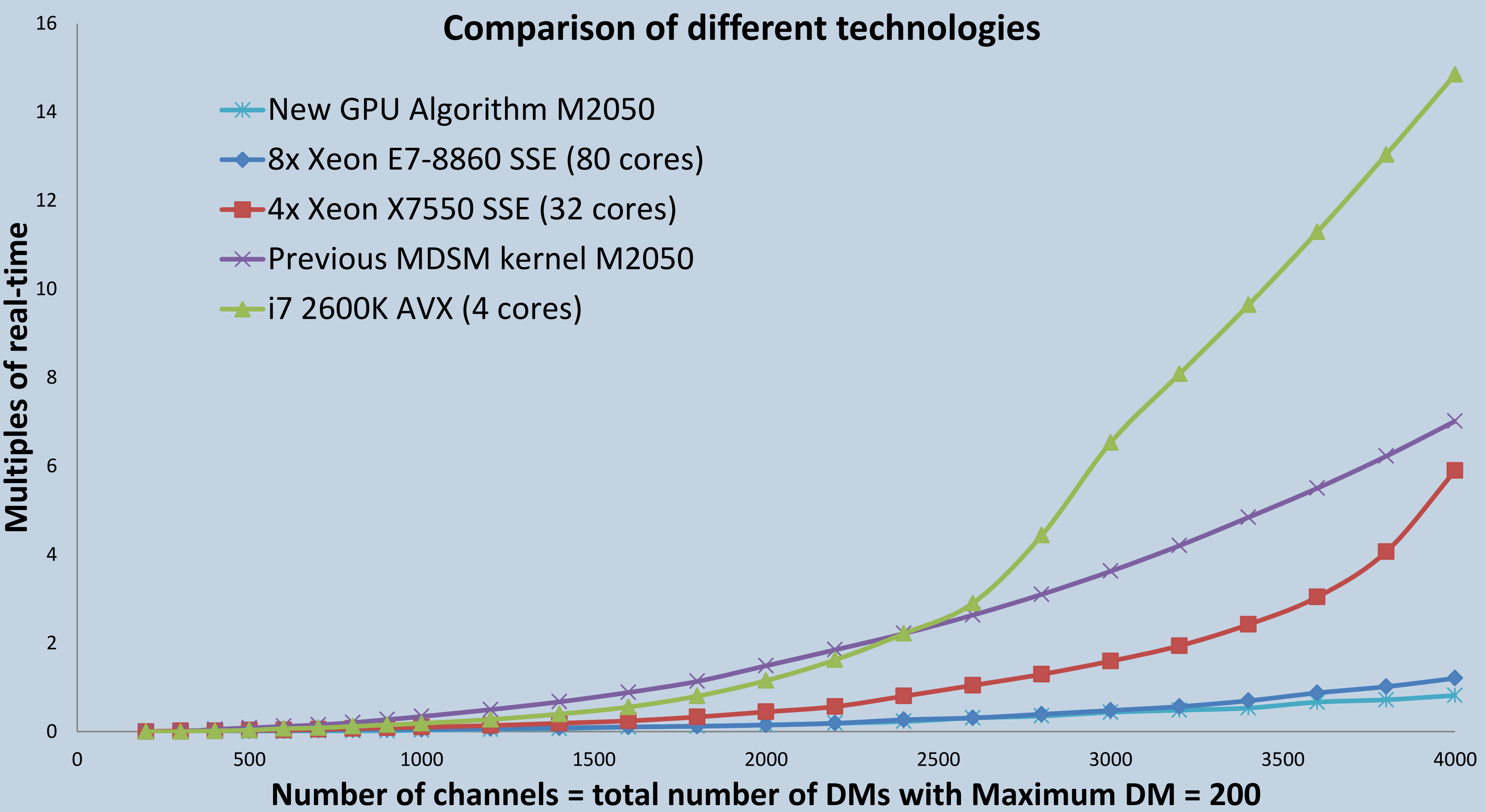
Incrementing all of the registers at every frequency step ensures a high data reuse of the stored frequency time data in the L1 cache.

Optimising the parameterisation.

The GPU block size of the new algorithm can take on any size that is integer multiples of the size of a “data chunk”…



Results…



OXFORD e-Research CENTRE

Conclusions and Future Work

- GPU wins hands-down. **At the moment!**
- AVX puts up a good fight.
- Watch out for Intels MIC (Many Integrated Core) chip – 32 in-order cores, 4 threads per core 512 bit SIMD units running a 1024 bit ring bus.

- Shared memory algorithm ensures more predictable data reuse and is about 15% quicker for some maximum DM and number of channels combinations – results to come.
- Myself, Ben and Mike are working on AVX vectorisaton of the Polyphase filter.

Acknowledgments and Collaborators

GPU de-dispersion : http://www.oerc.ox.ac.uk/research/wes

ARTEMIS : http://www.oerc.ox.ac.uk/research/artemis

University of Malta

Alessio Magro – MDSM

University of Oxford

Mike Giles (Maths) – Cuda, GPU algorithms.

Aris Karastergiou (Physics) – ARTEMIS, Astrophysics, Experimental Work.

Kimon Zagkouris (Physics) – Astrophysics, Experimental Work.

Chris Williams (OeRC) – RFI Clipper, Data pipeline.

Ben Mort (OeRC) – Data Pipeline, pelican.

Fred Dulwich (OeRC) – Data Pipeline, pelican.

Stef Salvini (OeRC) – Data Pipeline, pelican.

Steve Roberts (Engineering) – Signal processing/detection algorithms.