



Development of an Optimized Real-Time Radio Transient Imager for LWA-SV

Hariharan Krishnan¹
e-mail : hari.krish@asu.edu

The EPIC Collaboration : James Kent², Jayce Dowell³, Matthew Kolopanis¹, Adam P. Beardsley^{5,1}, Judd D. Bowman¹, Greg B. Taylor³, Nithyanandhan Thyagarajan⁴ & Daniel Jacobs¹

¹School of Earth & Space Exploration, Arizona State University

²Cavendish Laboratory, University of Cambridge, UK

³Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, USA

⁴National Radio Astronomy Observatory, Socorro, NM, USA

⁵Winona State University, Winona, Minnesota, USA

Outline

- Motivation for direct real-time imager
- Radio Interferometry & Conventional Correlator
- Direct Radio Imager - EPIC
- GPU Implementation
- Optimization & Results
- Summary

Motivation

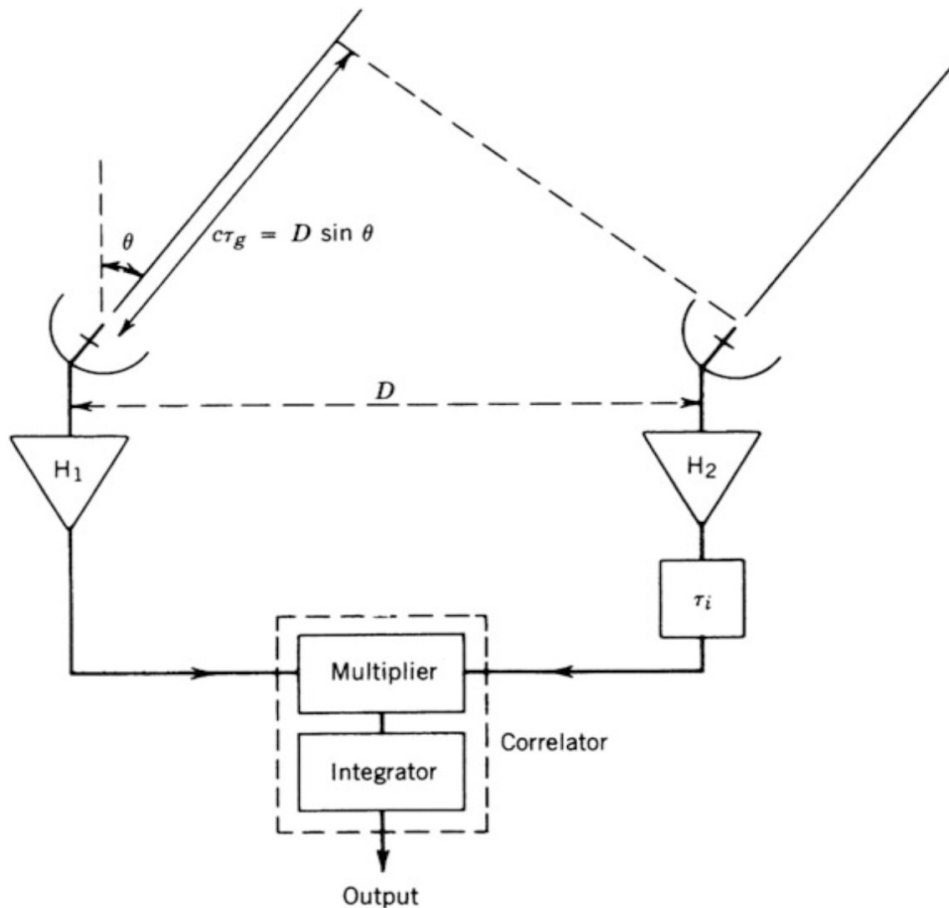
Scientific -

- Physics of radio transient phenomena like Fast Radio Burst (FRBs), Meteor Radio Afterglows (MRAs), planetary lightening, cosmic-ray air showers
- Observational study of stellar flares on sun-like stars in the exo-space weather context

Technical -

- Requirements of sensitivity, wide field-of-view and high angular resolution
- Real-time imaging across a wide frequency band at very high temporal resolution
- Current and next-generation radio telescopes rely heavily on digital signal-processing techniques

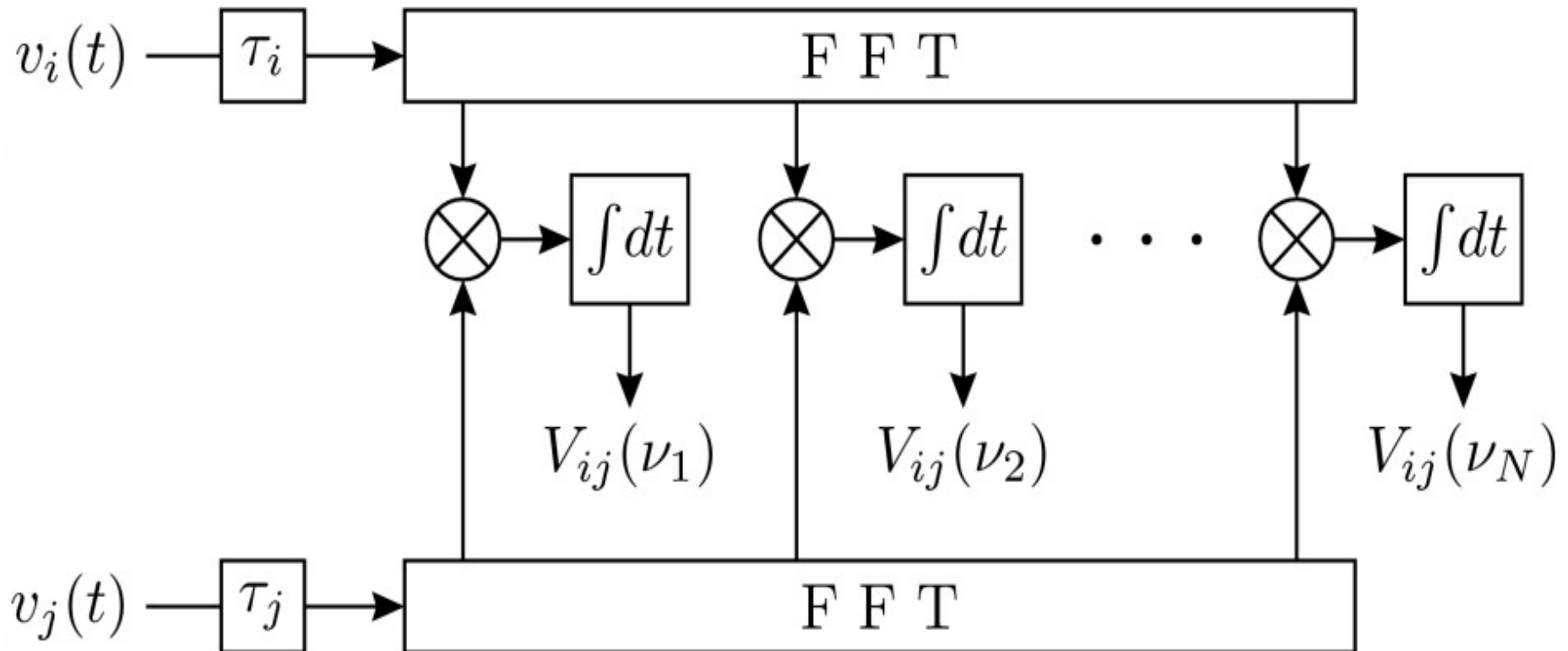
Radio Interferometry



- Two-element interferometer – Fundamental Unit of a radio telescope
- Cross-correlation – Multiplication & Integration of voltages to measure visibilities
- Baseline separation between antennas decides the spatial sampling of the sky

Image : Thomson, Moran & Swenson, 2017

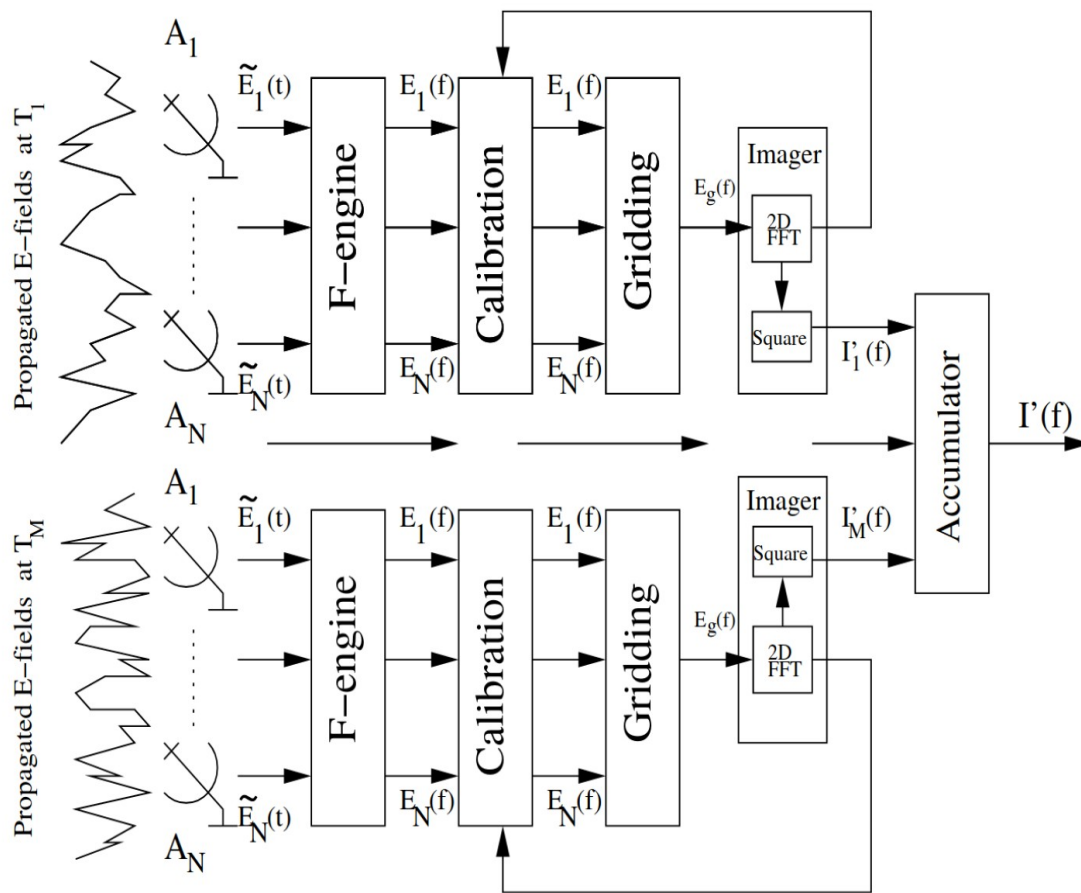
Conventional FX Correlators



E-Field Parallel Imaging Correlator (EPIC)

- Generic correlator implementation for real-time imaging in large-N dense arrays (viz. HERA, HIRAX, CHORD, PUMA etc.)
- Based on the Modular Optimal Frequency-Fourier (MOFF : Morales 2011) mathematical formalism for direct Fourier imaging
- Grid electric fields from individual antennas and spatially Fourier transform to sky image : synthesizing the aperture on-the-fly
- Significant reduction in computational scaling from $O(n_a^2)$ to $O(n_g \log_2 n_g)$ (where n_a is the number of antennas and n_g is the number of grid points)

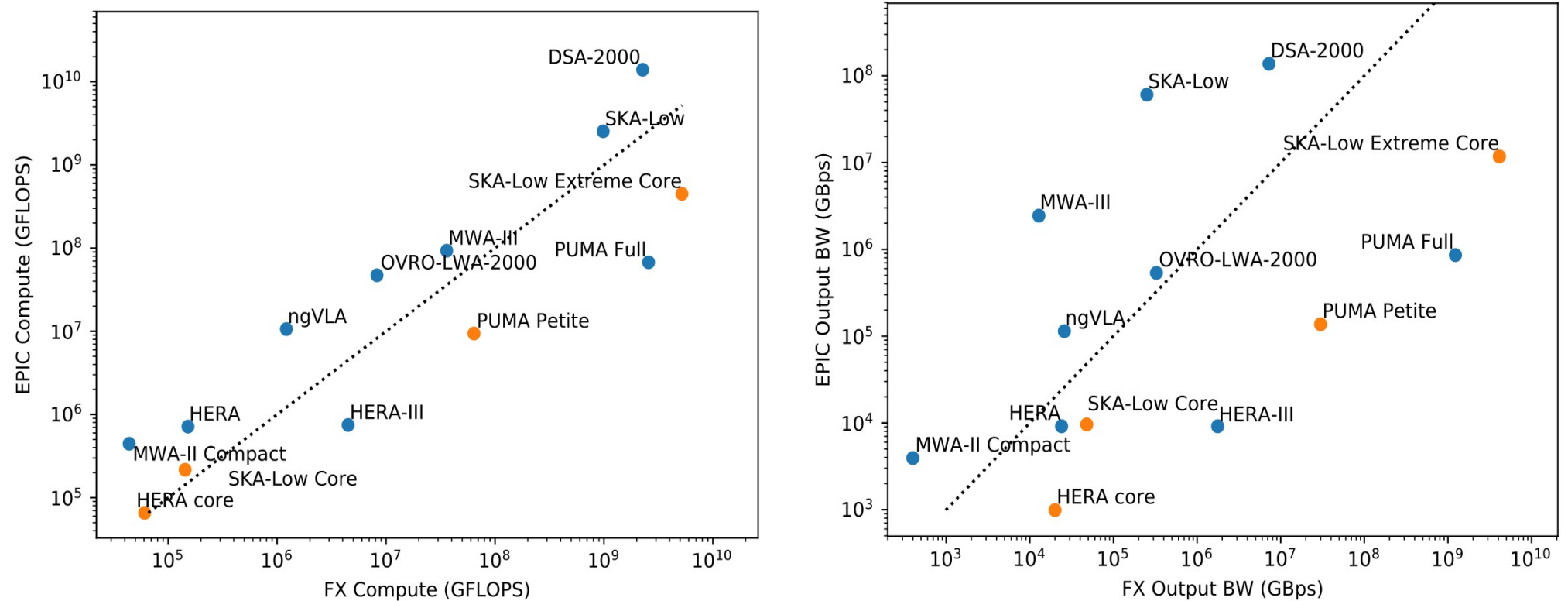
Direct Imager - EPIC



Flowchart of MOFF imaging in EPIC
(Thyagarajan et al. 2017)

- Propagated electric fields ($E(t)$) are measured as time-series from individual antennas
- $E(t)$ transformed by the F-engine to produce electric field spectra ($E(f)$)
- $E(t)$ is calibrated and gridded
- The gridded electric fields $E_g(f)$ from each time series are imaged
- Images are time-averaged to obtain final image $I'(f)$

EPIC vs FX

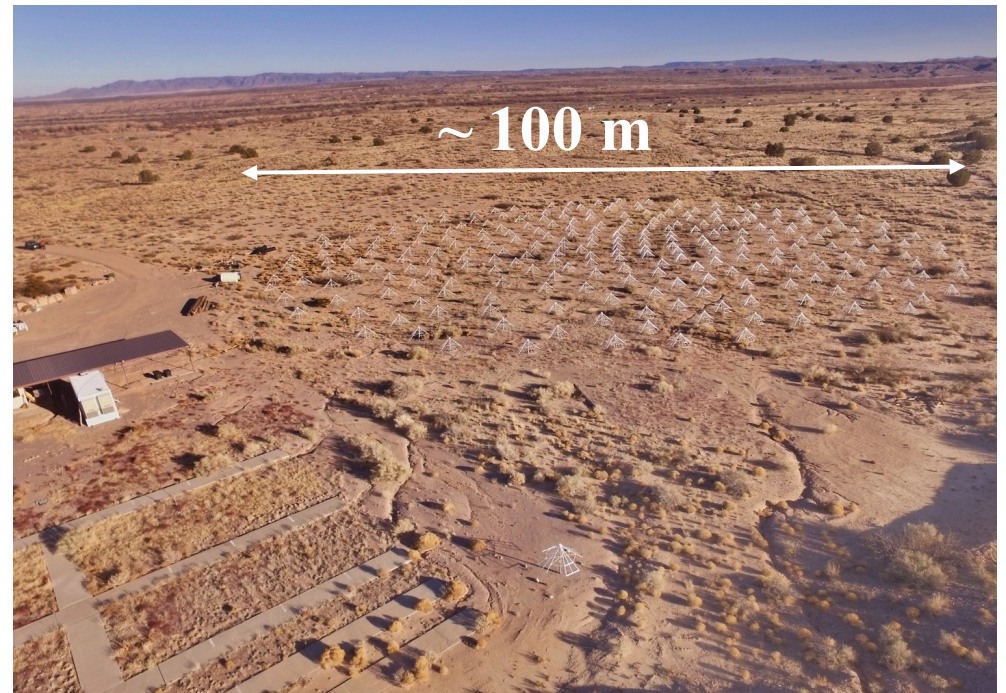


Comparison of computational cost (left) and output bandwidth (right) with EPIC and FX-based approaches for a fast transient campaign (images at 0.5 ms cadence) using various interferometer arrays. The dotted line denotes where performances are equal (Thyagarajan et. al., 2019 – APC White paper)

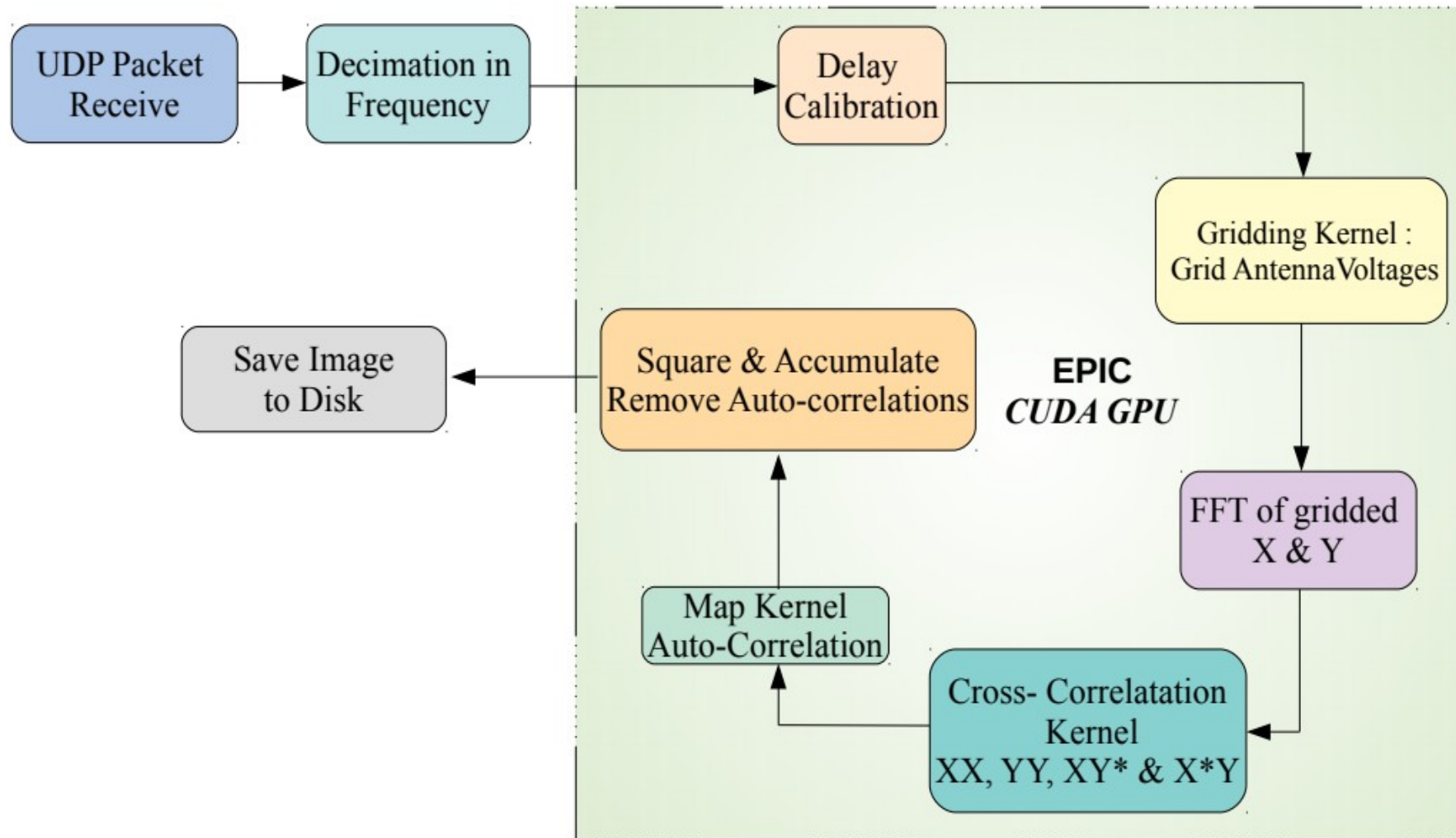
Deployment of EPIC

- Implemented on a GPU-accelerated architecture and integrated with a python/C++ based high-performance streaming framework, Bifrost (Cranmer et al. 2017)
- Successfully deployed and tested on the Long Wavelength Array station (Taylor et al. 2012) located at the Sevilleta National Wildlife Refuge (LWA-SV) in New Mexico, USA
- LWA-SV is a compact array of 256 antennas arranged in an elliptical footprint spanning ~ 100 m
- LWA-SV operates in the frequency range 10-88 MHz

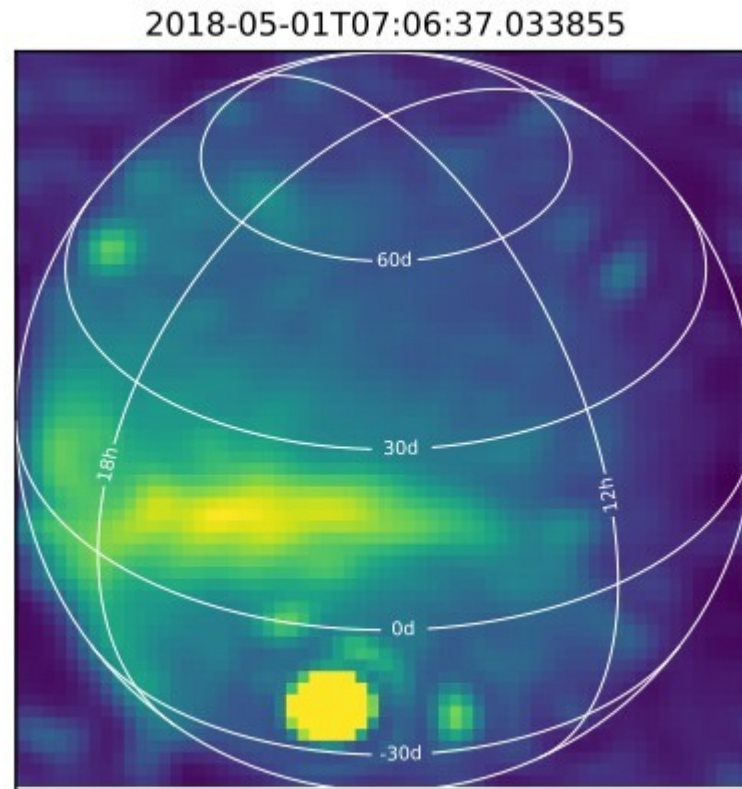
(Image Courtesy : Greg Taylor)



GPU Implementation



Real-time Images with EPIC



All-sky pseudo-Stokes-I image showing a meteor reflection detection during an observation on the LWA-SV site

Need for Optimization

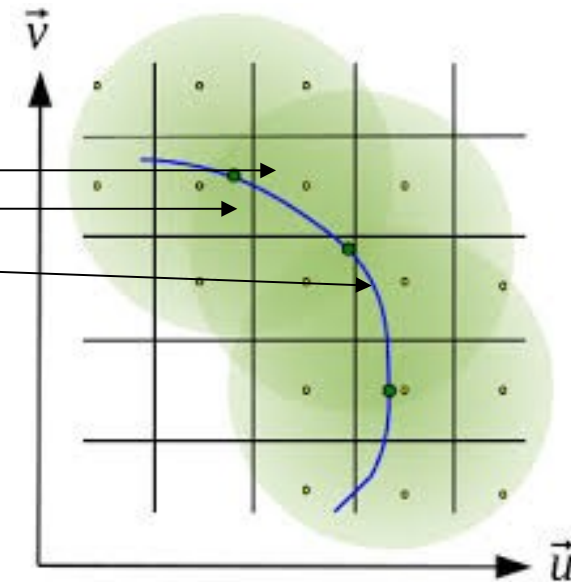
- Instantaneous bandwidth for initial deployment limited to ≈ 400 kHz per GPU
- Optimize the GPU-code of the correlator for better performance in order to increase the bandwidth processable per node in real-time
- It was decided to begin with low-level CUDA coding modifications to the voltage gridding module
- CUDA thread configuration and memory access pattern rearrangement
- Optimizing memory accesses has a huge effect on GPU code efficiency
- Introducing new modules for cross-correlation and auto-correlation removal

Optimization Strategy

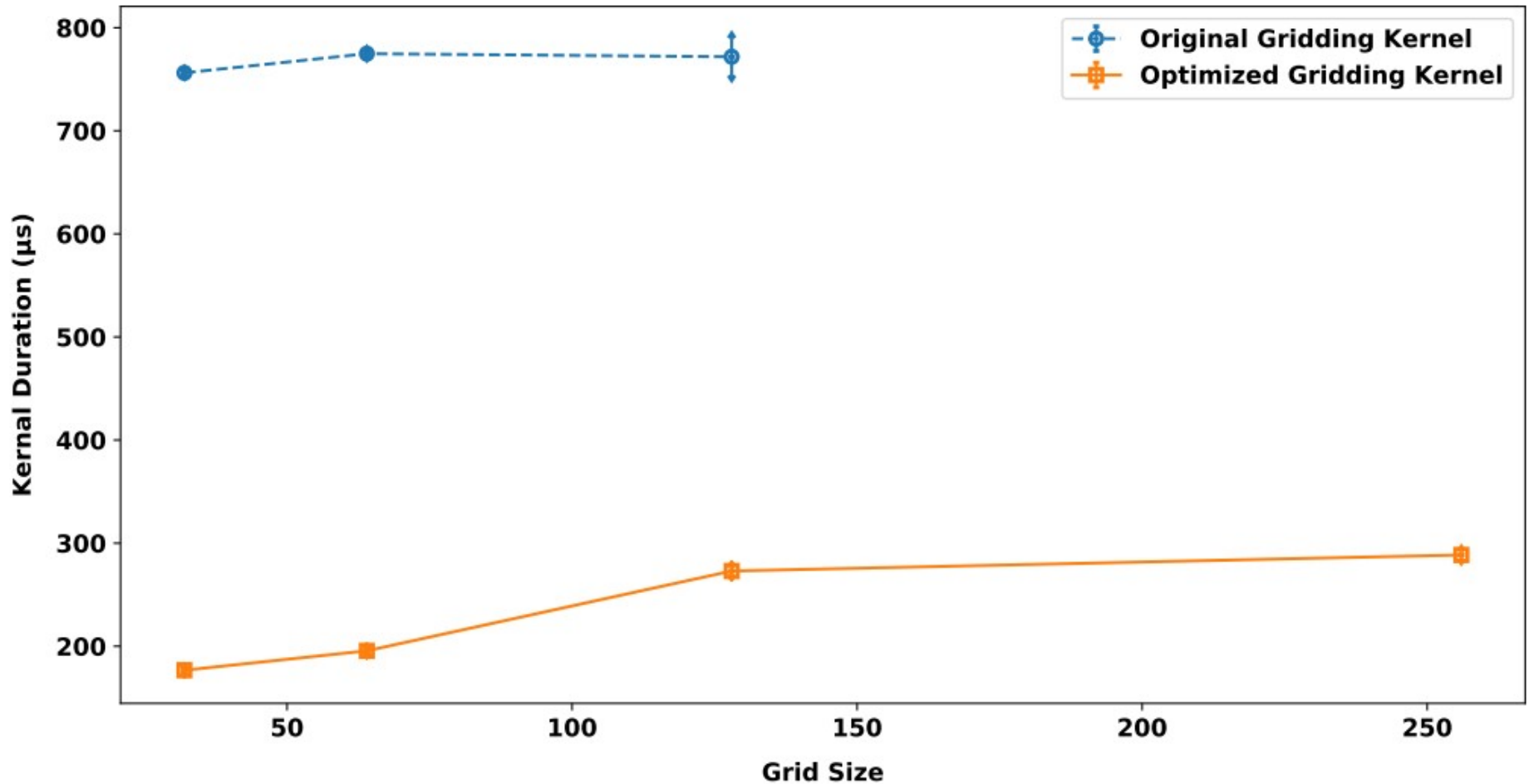
- Memory Optimization
 - Reduce redundant memory access
 - Memory Coalescing for improved memory access pattern
 - Shared memory usage to reduce global memory access
- Choice of thread block size for increased concurrency to hide latency
- Achieve optimal thread occupancy
- Instruction level optimization with high-throughput instructions and reduced branch-divergences.
- CPU-GPU interaction optimization through overlapped execution

Gridding Module

- One of the critical blocks of the EPIC, that is based on a GPU-accelerated convolution algorithm (Romein 2011)
- Delay corrected frequency domain signals are convolved with an antenna illumination pattern/Convolution function and gridded with a spacing of $< \lambda/2$ on to a 2-D grid



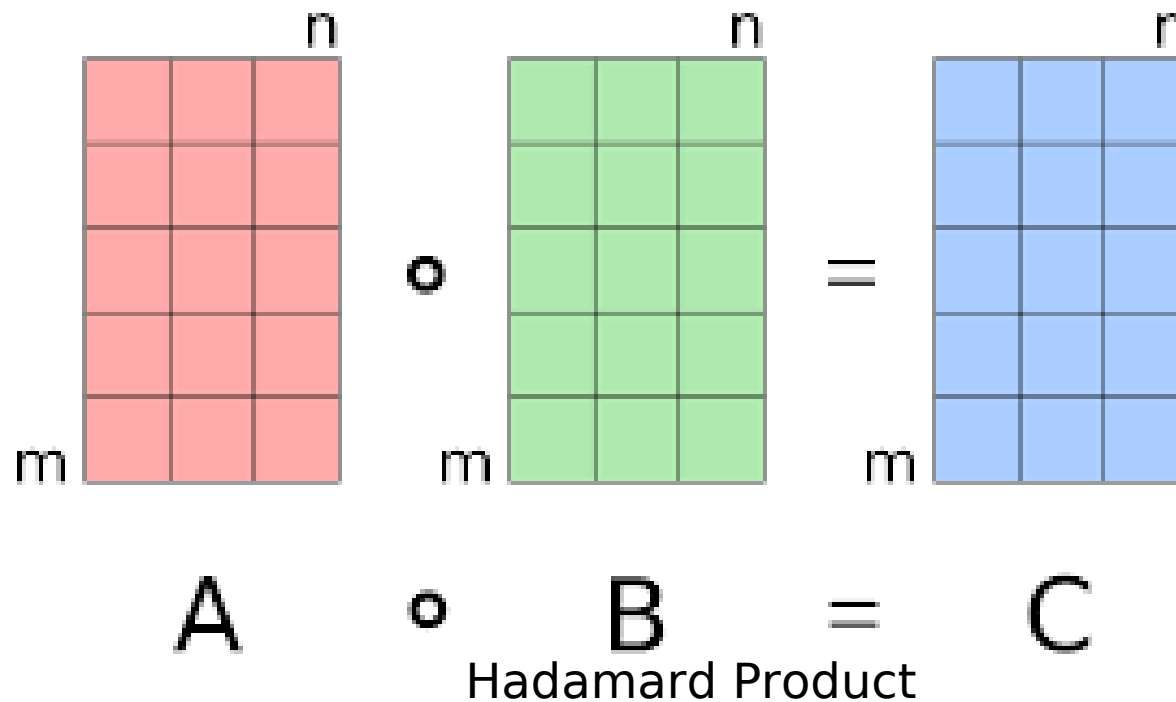
Kernel Duration



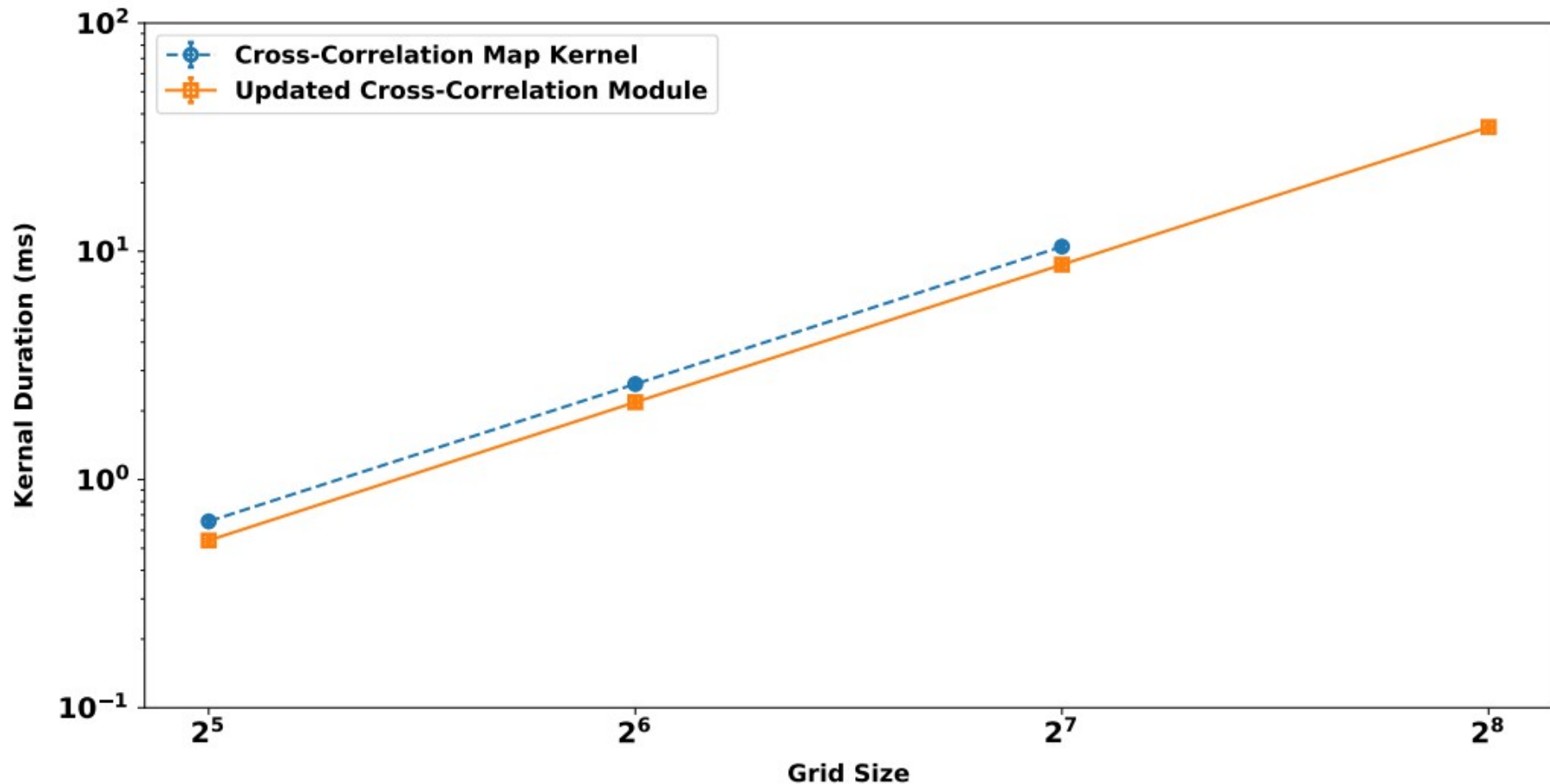
Comparison of the kernel run-time duration vs Grid-size Dimension: original (blue) and modified (orange) Gridding kernel (Hariharan et al. 2020)

Cross-Correlation Module

- Cross-correlation of gridded X & Y voltages
- Full Polarization estimator - XX^* , YY^* , XY^* & YX^*



Kernel Duration



Comparison of the kernel run-time duration vs Grid-size Dimension:
Bifrost map kernel (blue) and Cross-correlation kernel (orange) (Hariharan et al. in preparation)

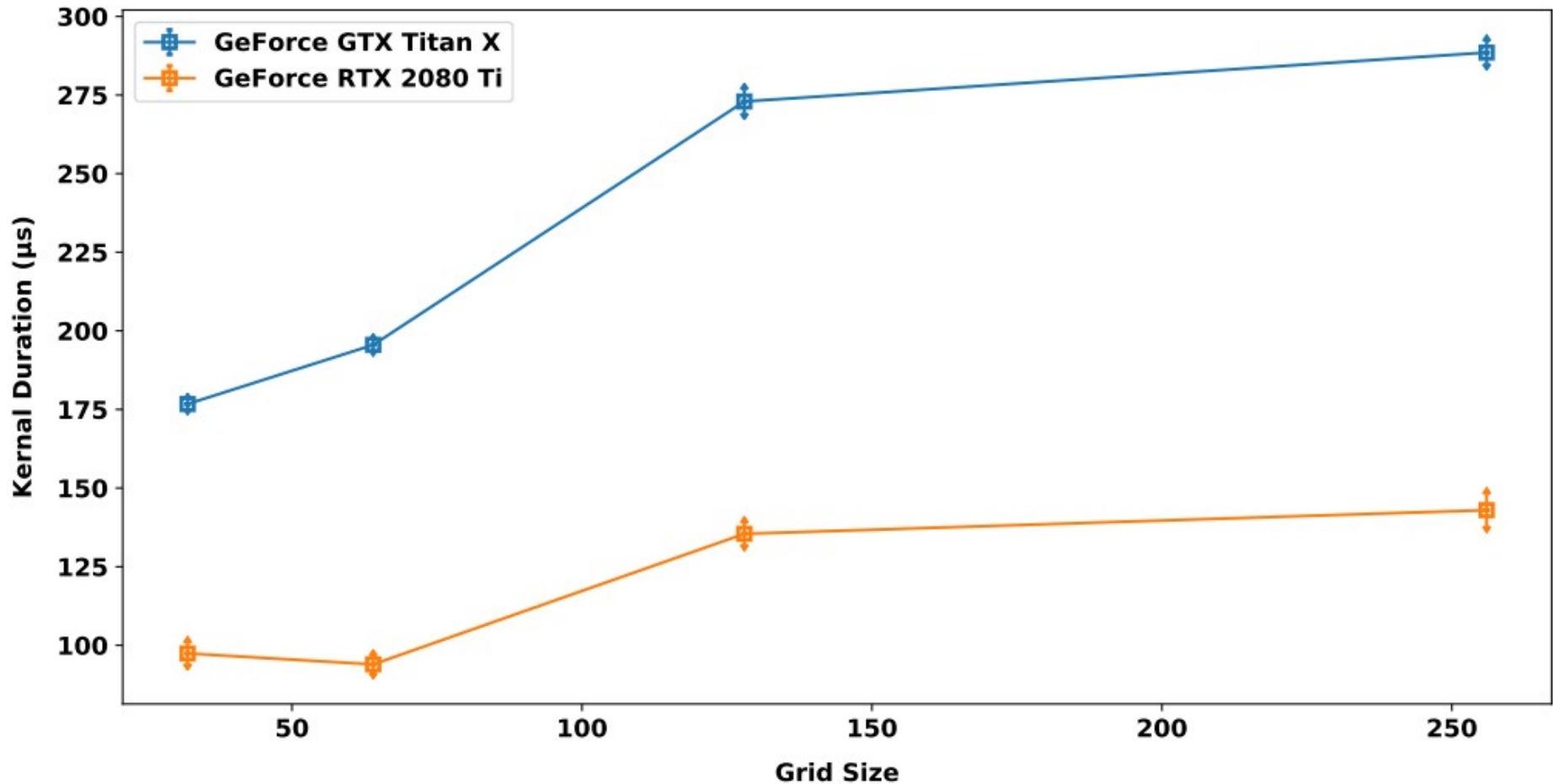
Hardware & System Modifications

- Hardware and system changes can drastically improve performance of software-processing
- Commensal machine upgraded to 40Gbps from earlier 10 Gbps

Comparison of specs for GPU

	GeForce GTX 980 (Old)	GeForce GTX Titan X(ASU)	GeForce RTX 2080 Ti (Commensal)
Number of Cores	2048	3072	4352
GPU Clock (MHz)	1127 MHz	1000 MHz	1350
Number of SM	16	24	68
Global Memory- Bandwidth	224.4 GB/s	336.6 GB/s	616 GB/s
Texture Rate	155.6 GTexel/s	209.1 GTexel/s	420.2 GTexel/s
FP32 (float) performance	4.981 TFLOPS	6.691 TFLOPS	13.45 TFLOPS
FP64 (double) performance	155.6 GFLOPS	209.1 GFLOPS	420.2 GFLOPS

Gridding Kernel Duration - Hardware Change



Expected Theoretical Capability for EPIC : ~ 6.4 MHz per node @ 2.5 ms & 25 kHz

Image Comparison

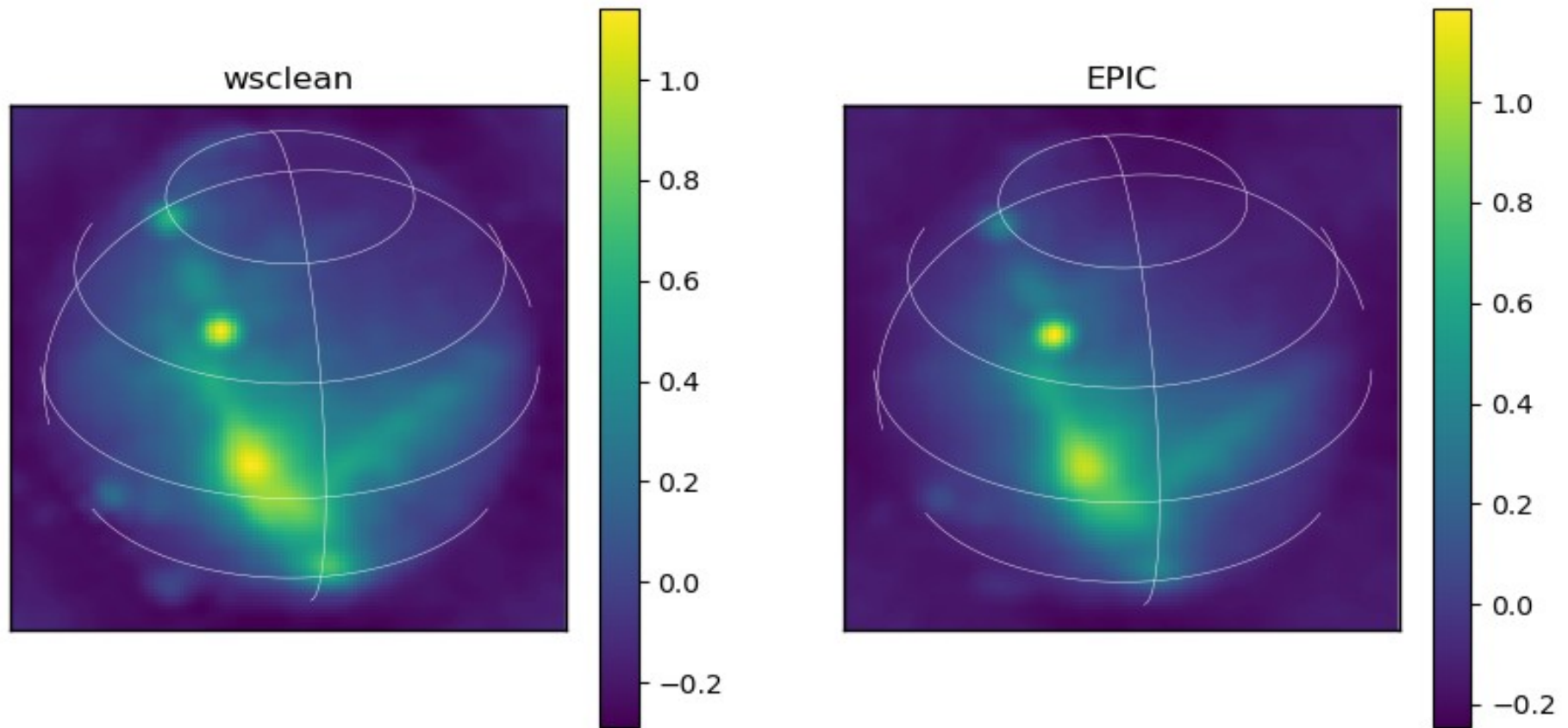


Image Courtesy : Adam Beardsley

Summary & Future Perspectives

- EPIC is a generic / fast / efficient version of a direct imager and inherently a science-ready interferometric imaging architecture
 - Potential for usage in current and next-generation densely packed radio arrays
 - Optimization of the gridding and cross-correlation modules through low-level code modifications and memory management was performed for improved performance.
 - Evaluation of optimizations and addition of new modules are currently being carried out.
 - Current theoretical capability of EPIC is : ~ 3.2 MHz per GPU @ ~ 2.5 ms integration (Real-time Implementation Underway)
 - Further Advancements to deploy a transient detector with EPIC as a commensal imaging back-end is planned

For Discussions, Questions/Comments – hari.krish@asu.edu