GPU-based correlation for the EVN

Aard Keimpema (keimpema@jive.eu), Mark Kettenis, Marjolein Verkouter

Introduction

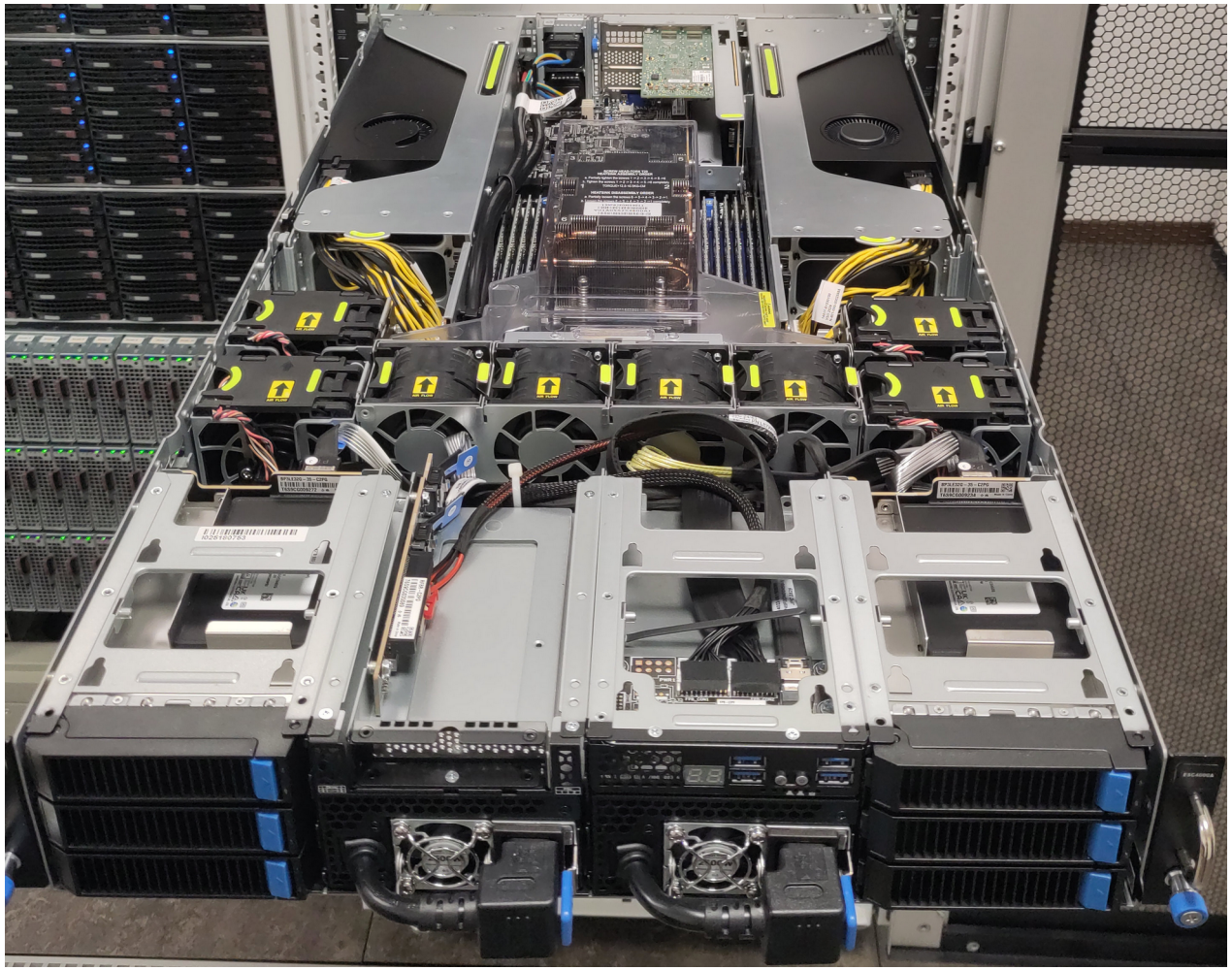
- Current EVN cluster: 840 Intel Xeon cores (2018)
  - 20 stations @ 2 Gpbs in realtime
  - Uses 7.5 kW of power!
- Next gen receivers, tri-band, C/X, … : 4 GHz+ bandwidth
- Next generation backends, DBBC3, DBBC4,.. : 32 Gbps+
- Moore's law for CPUs is dead
- Next generation EVN correlator needs to green
- GPUs can provide an order of magnitude of improvement in efficiency
- Funding in RADIOBLOCKS to develop GPU powered VLBI correlator

SFXGPU

- New correlator, not directly based on SFXC
- Same correlation algorithm as SFXC
- Can already correlate 95% of EVN experiments
- Implemented features include
    - Mixed-bandwidth correlation
    - Hanning smoothing
    - Multiple simultaneous phase centres
    - Coherent de-dispersion
- To-do
    - Pulsar folding
    - Geodetic VLBI support (phase-cal extraction)
    - eVLBI support
- Source code: **https://code.jive.eu/keimpema/sfxgpu**

GPU correlator

- **CPU:** 1x MD Epyc 9555P - 360W
- **GPU:** 4x NVIDIA RTX PRO 4000 PNY
  Blackwell 24GB GDDR7
  PCI Express x16 5.0 - 140W
- **RAM:** 768 GB (12x 64GB 4800MT/s
  DDR5 ECC Reg CL40 DIMM)
- **Storage:** 2x .9TB SSD U.2 NVMe
  PCIe 4.0 x 4 Enterprise
- **NIC:** ConnectX-6 DX EN adapter card,
  100GbE, Dual Port



Funded through the Radioblocks project

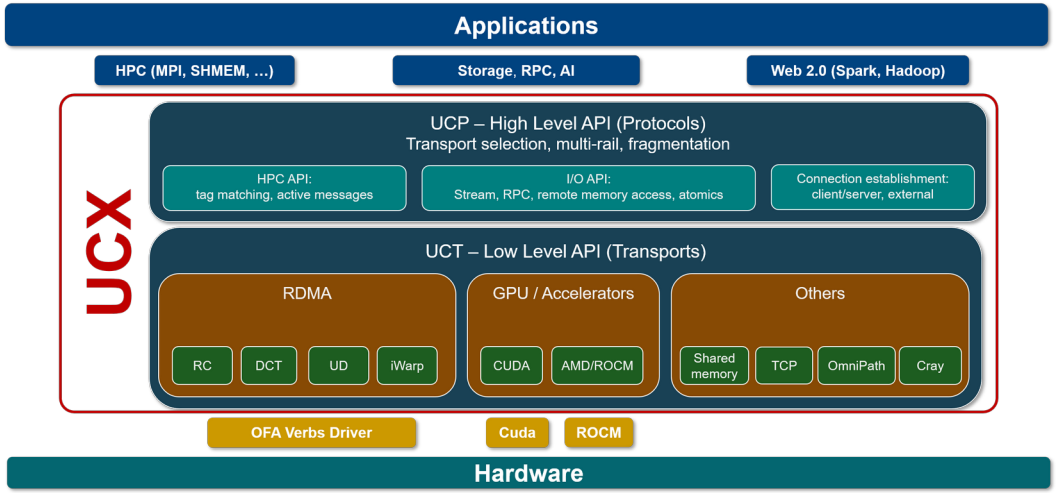
Bye Bye MPI; Welcome UCX

Abandoned MPI in favour of (lower-level) UCX library

- Easier to deploy in heterogeneous environment
- Use wakeup feature to save energy (no polling)

Split correlator application into separate reader and correlator processes

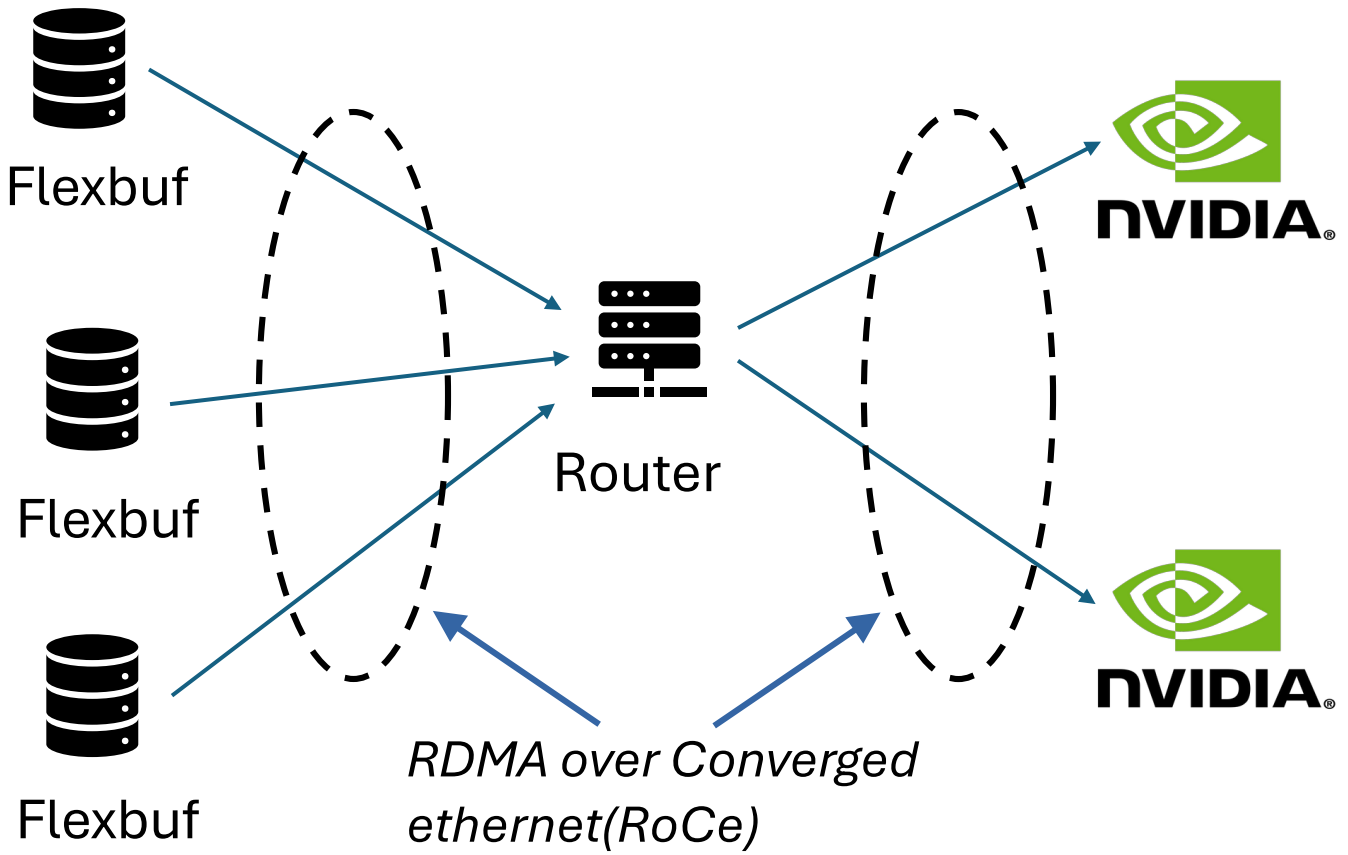
- Created a “ucxwrappers” abstraction
- Can use RDMA directly into GPU memory
  - Even works for “localhost”; using ConnectX NIC as a DMA engine!
- Correlator process consumes almost no CPU cycles!
- Reader process doesn’t need CUDA



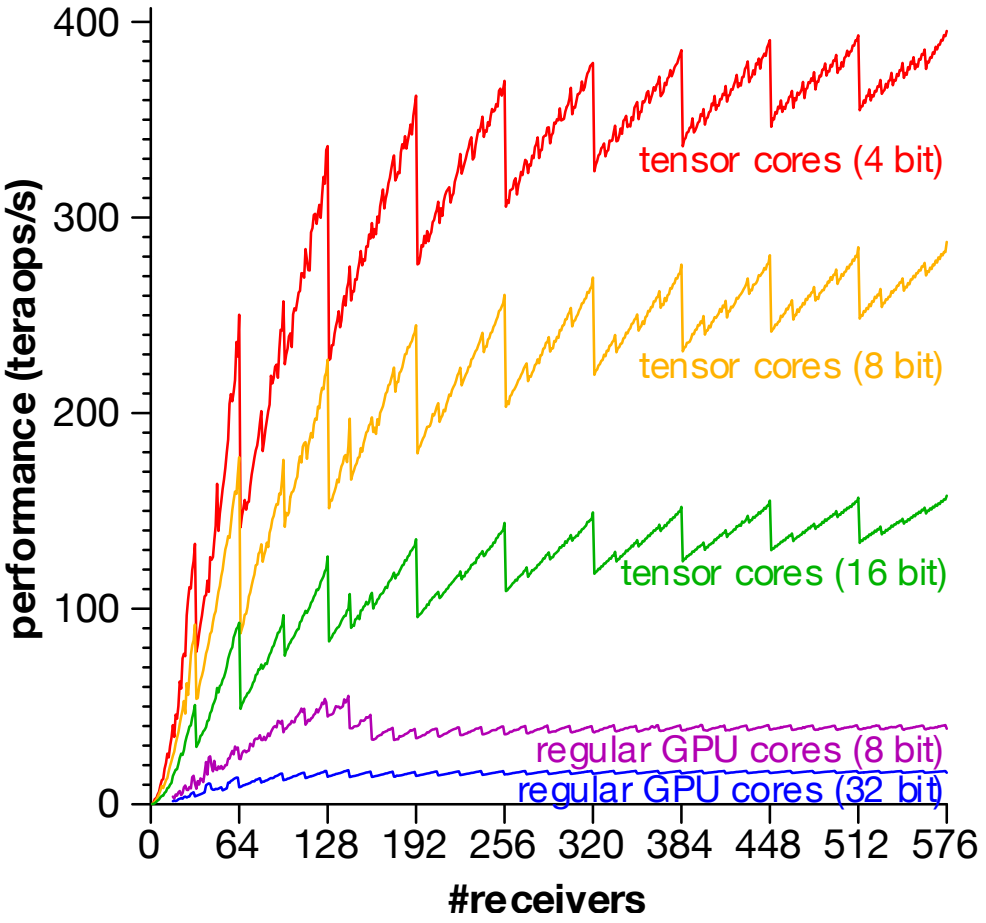
Unified Communication X (UCX) framework for HPC
https://openucx.org/

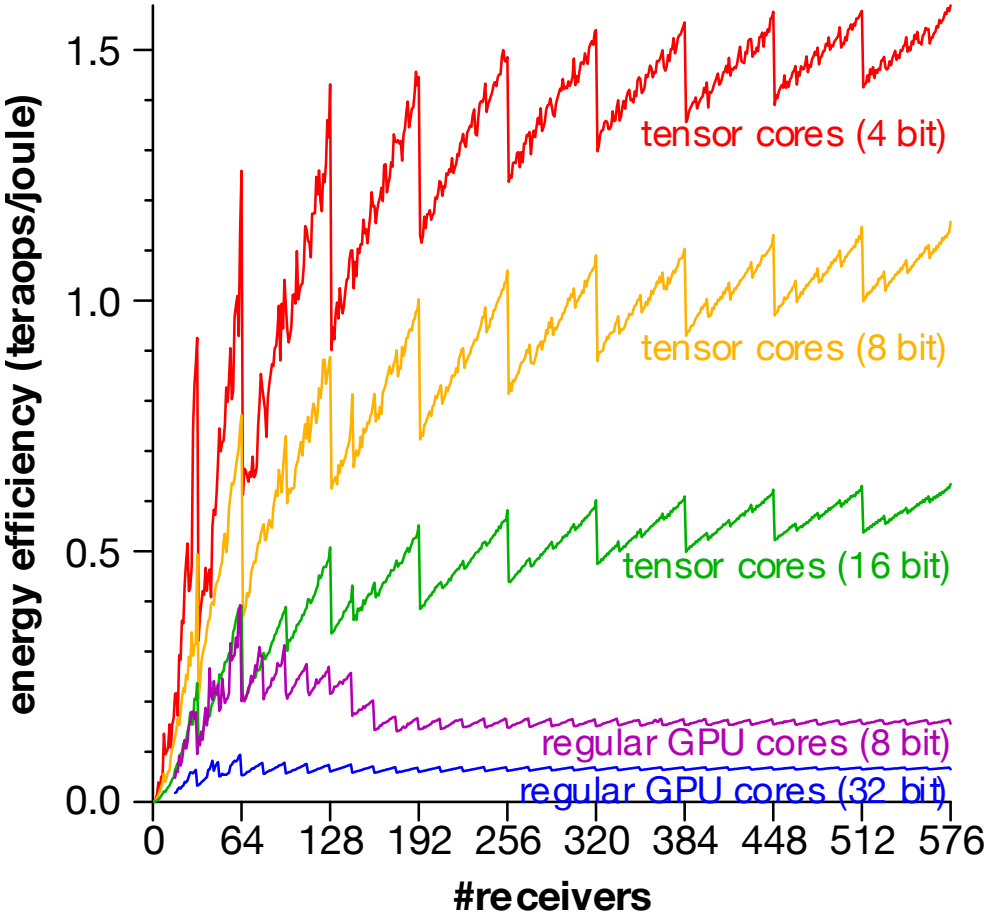
Architecture

- Reader processes on Flexbufs
- VDIF data decoded and corner-turned on Flexbuf
- Coarse integer delay correction is also applied on Flexbuf
- Correlation is parallelised in time and frequency
- Data is streamed directly into GPU memory using RDMA
- RDMA over Converged Ethernet (RoCE) through the ConnectX NICs



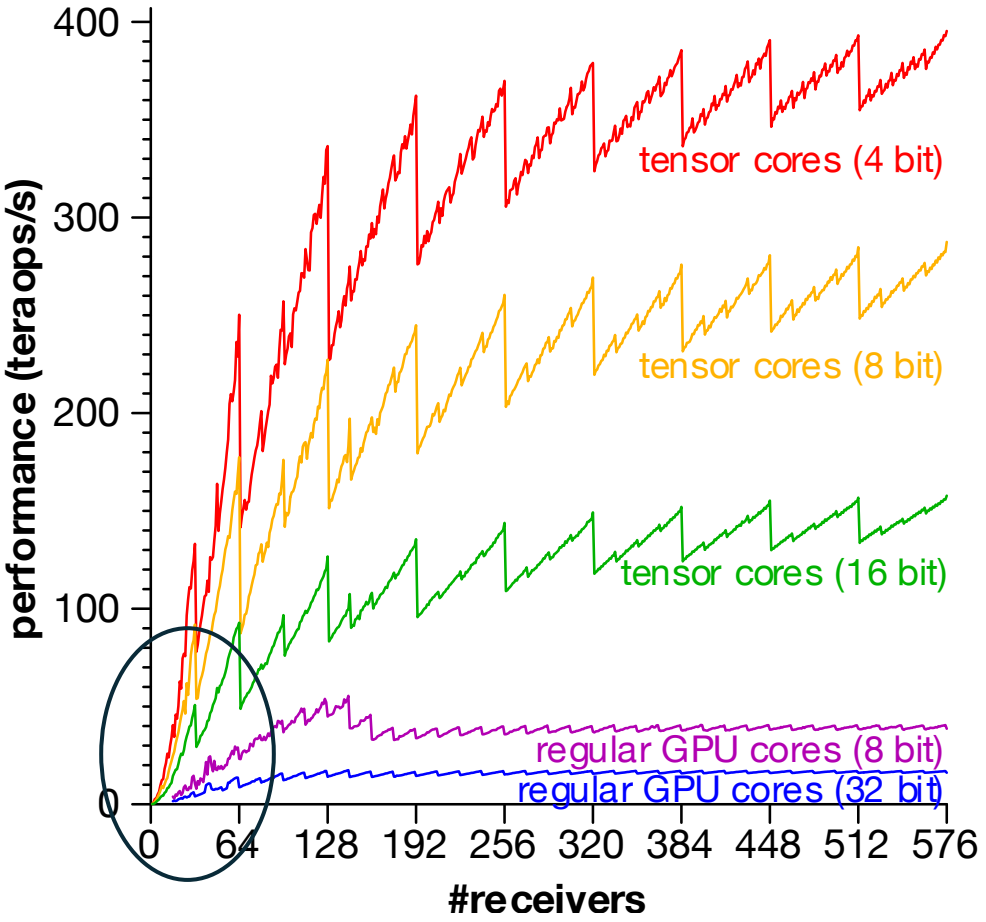
Tensor Core correlator

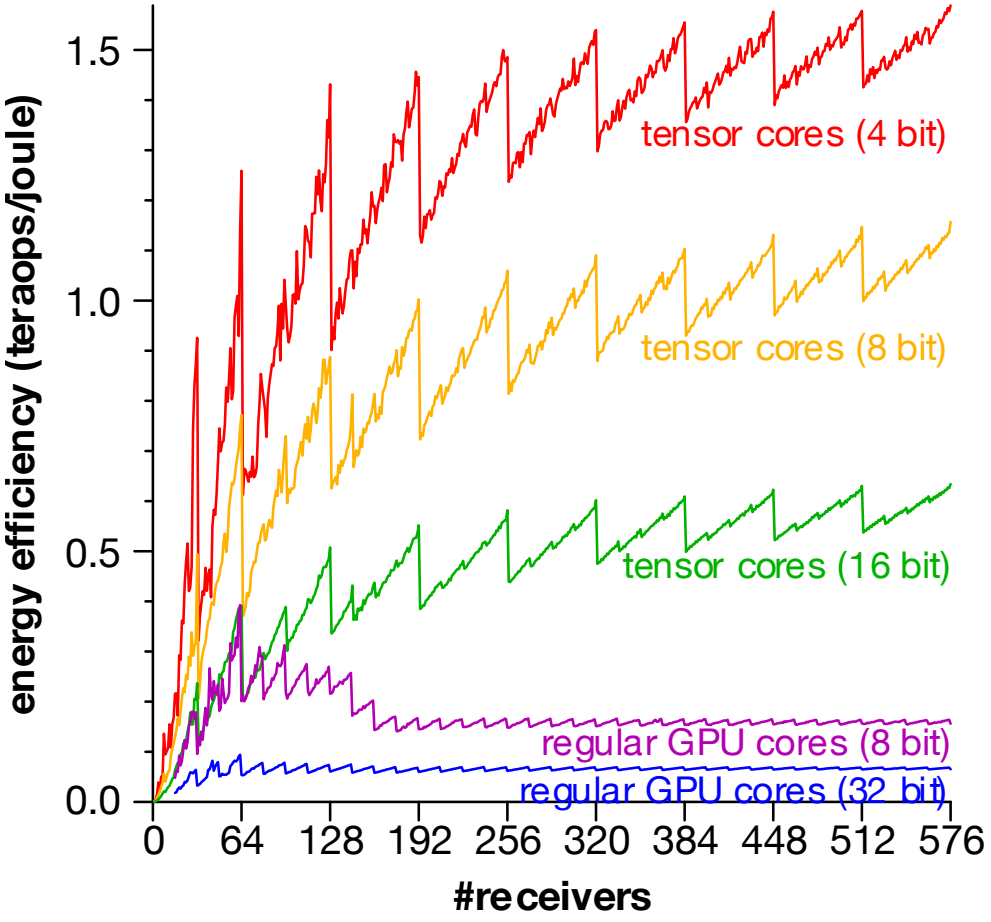




- John Romein: https://git.astron.nl/RD/tensor-core-correlator

Tensor Core correlator



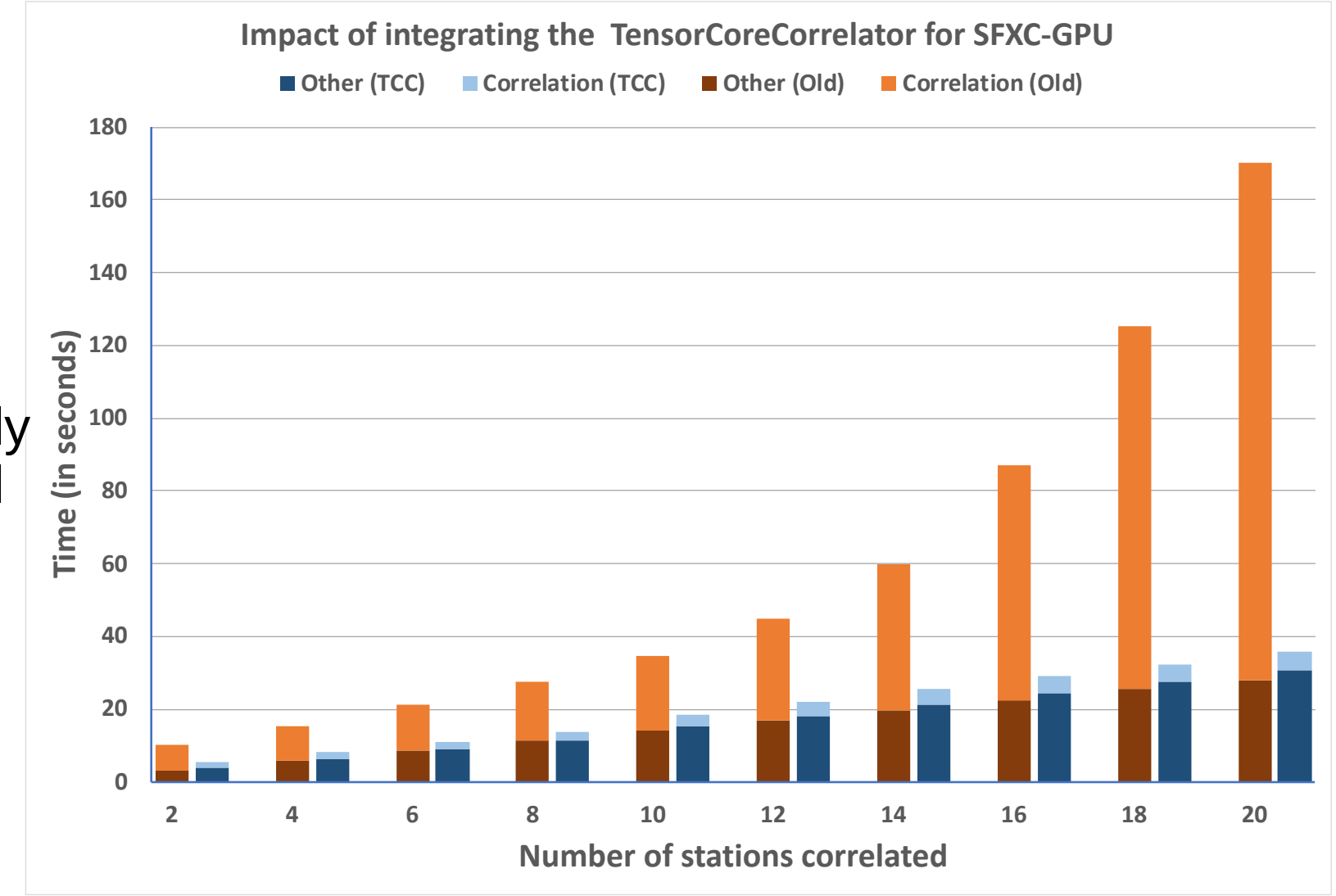


Significant speedup even in VLBI case

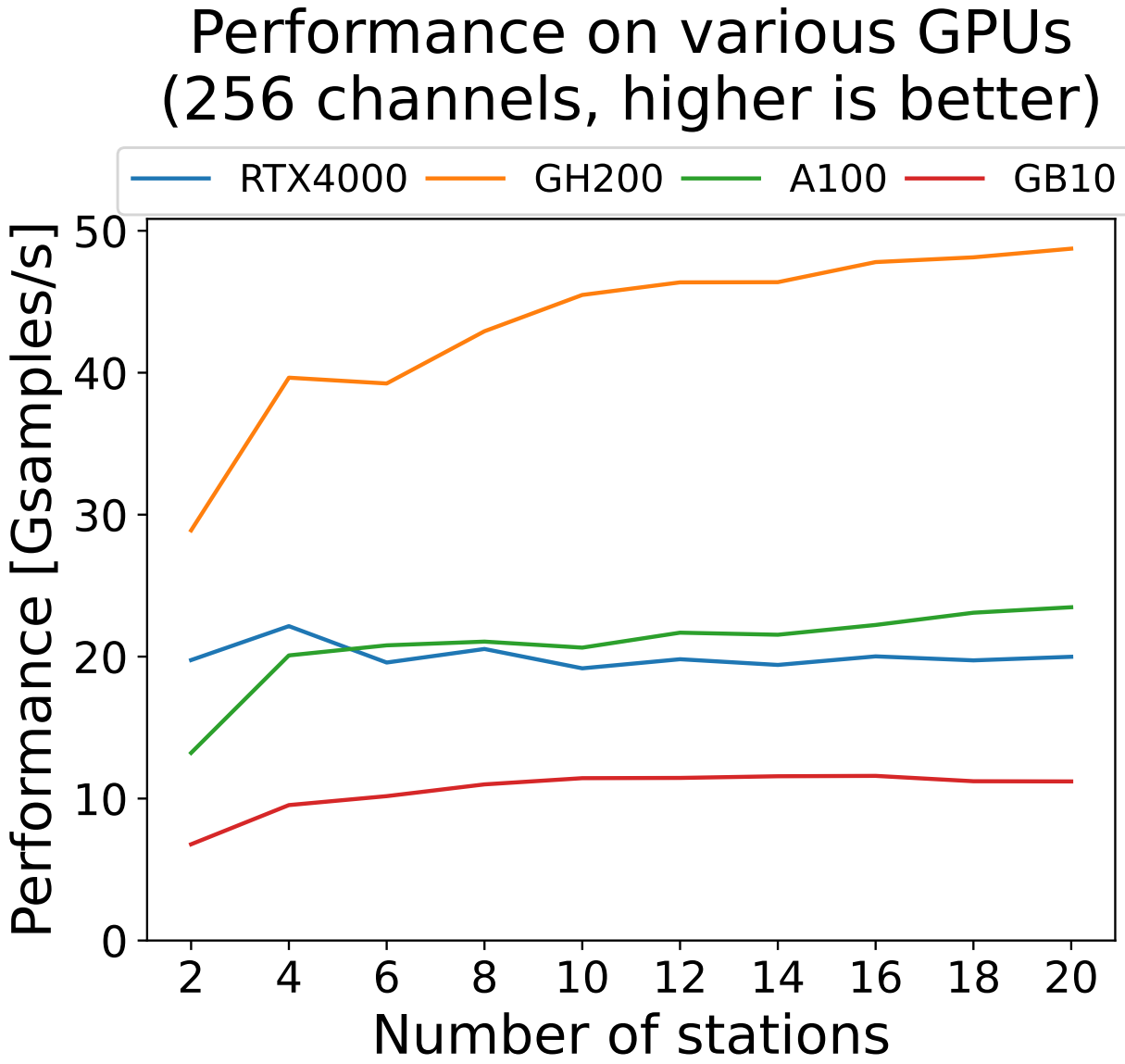
- John Romein: https://git.astron.nl/RD/tensor-core-correlator

Effect of TCC on SFXGPU

- Old correlation kernel used regular GPU cores
- Was based on xGPU developed by NVIDIA
- (cross-)correlation now is only a small part of computational load
- Delay-compensation + FFT now dominate computation

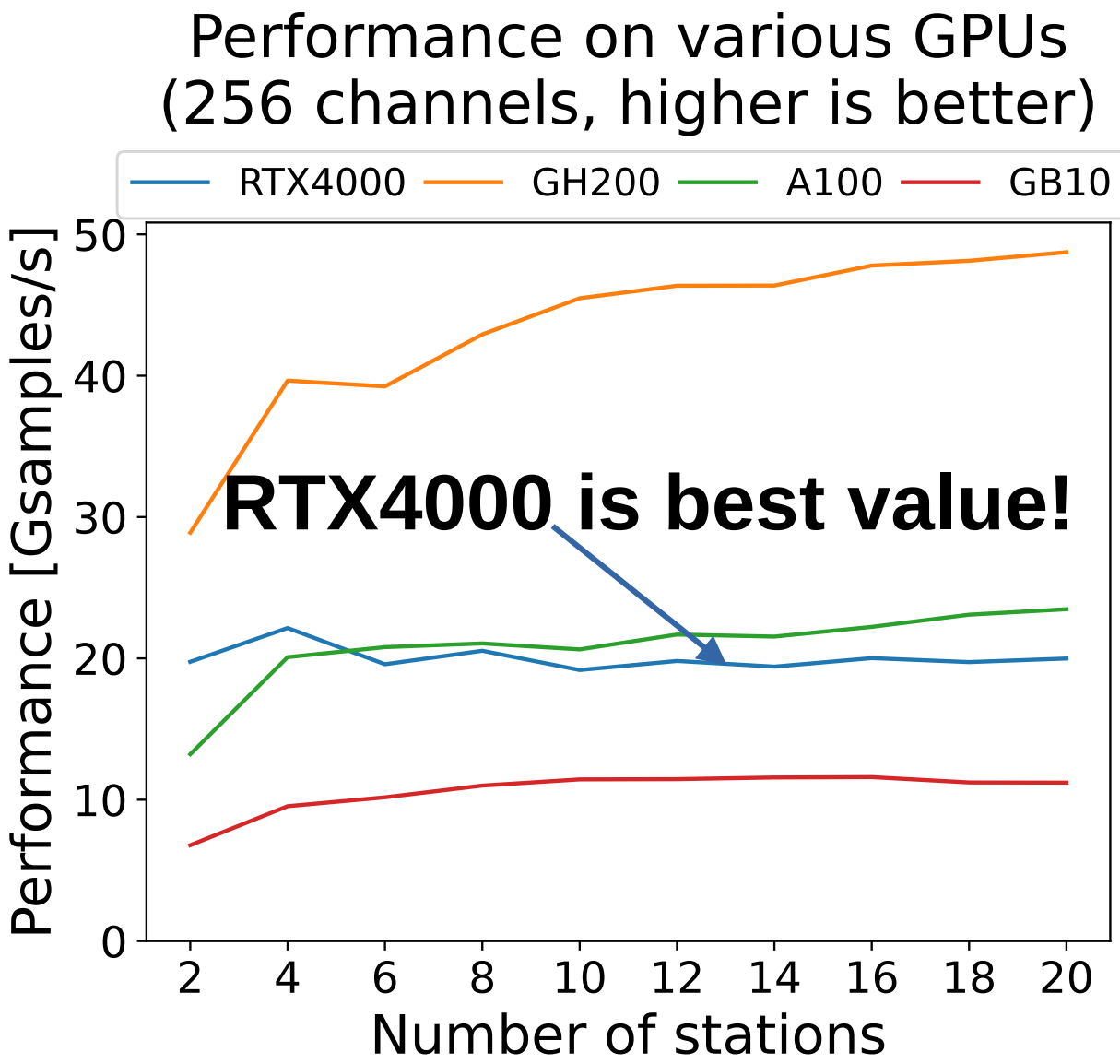


Benchmarks



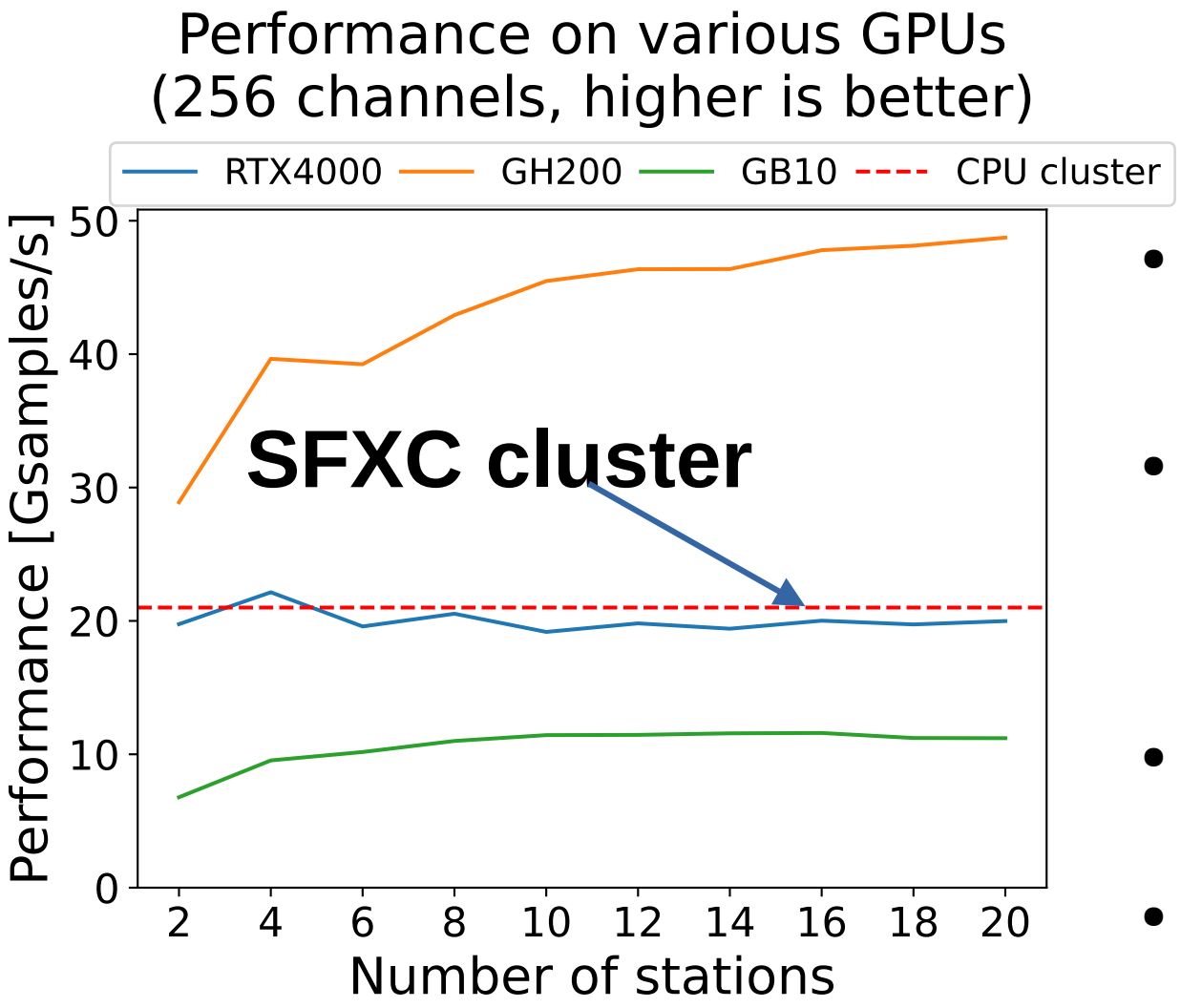
	RTX PRO 4000	GH200 (Grace Hopper)	A100	GB10 (DGX Spark)
Architecture	Blackwell	Hopper (GPU) + Grace (CPU)	Ampere	Blackwell (GPU) + Grace (CPU)
GPU memory	24 GB GDDR7	96–144 GB HBM3/HBM3e	80 GB HBM2e	128 GB unified LPDDR5x
Memory bandwidth	~672 GB/s	~4,000 GB/s (HBM) + 900 GB/s NVLink-C2C	~2,039 GB/s	~273 GB/s
TDP	145 W	~900W (system)	400W	~140 W
Price	~€3000	~€30000	~€12500	~€5000

Benchmarks



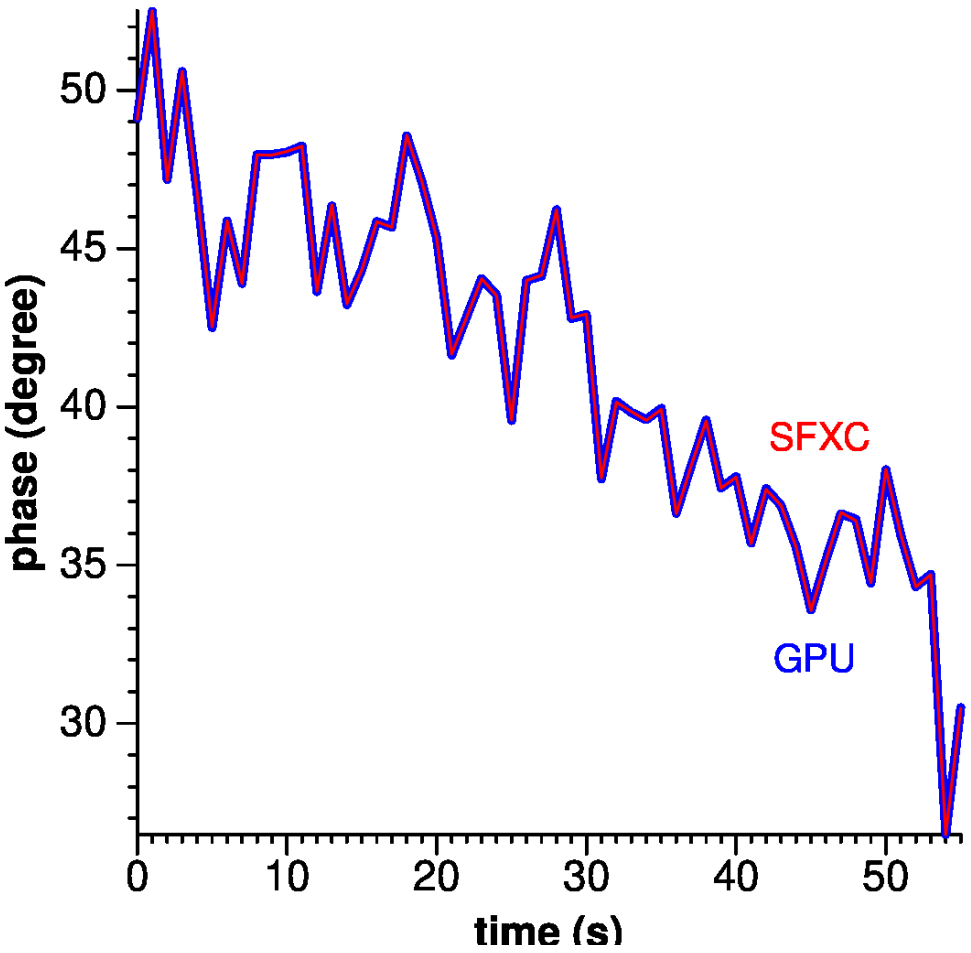
	RTX PRO 4000	GH200 (Grace Hopper)	A100	GB10 (DGX Spark)
Architecture	Blackwell	Hopper (GPU) + Grace (CPU)	Ampere	Blackwell (GPU) + Grace (CPU)
GPU memory	24 GB GDDR7	96–144 GB HBM3/HBM3e	80 GB HBM2e	128 GB unified LPDDR5x
Memory bandwidth	~672 GB/s	~4,000 GB/s (HBM) + 900 GB/s NVLink-C2C	~2,039 GB/s	~273 GB/s
TDP	145 W	~900W (system)	400W	~140 W
Price	~€1600	~€30000	~€12500	~€5000

Benchmarks

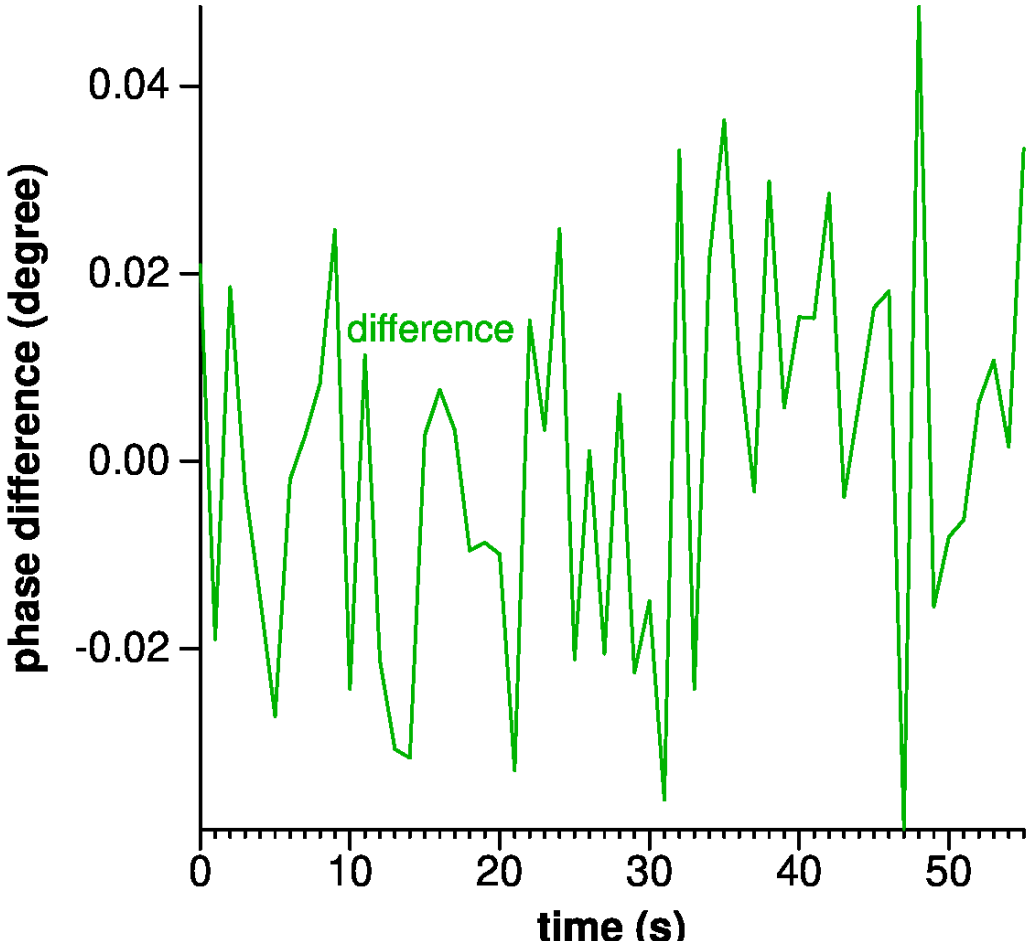


- Current SFXC cluster:
  - 840 Intel Xeon cores (2018)
- Power usage:
  - 3 kW (idle)
  - 7.5 kW (under load)
- Single RTX4000 matches SFXC cluster in performance
- Only 140W per GPU!

Effelsberg –Onsala @ 18cm



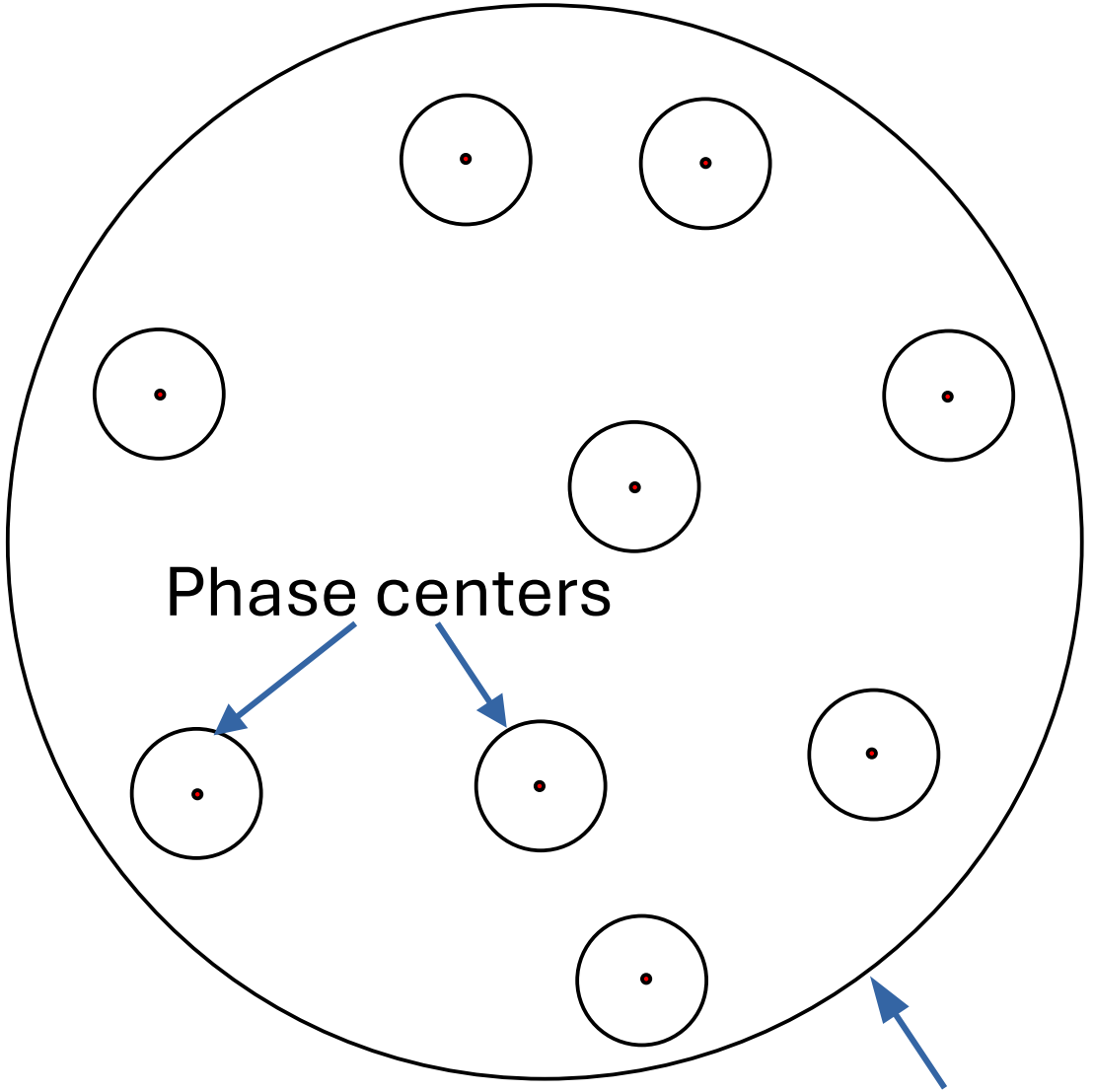
Visibility phase versus time for both SFXC and SFXGPU.



Phase difference between SFXC and SFXGPU

Multiple simultaneous phase centres

- It is possible to image the entire field of view of the individual antennas is the array
- Problem: Imaging the entire field of view yield very large data sets of mostly empty space
- Typically only a few dozen sources in 10’ field of view
- Solution: Correlate using high spectral / temporal resolution but only write small narrow field data sets for each source in the field.
- Implementation based on the Tensor Core Correlator

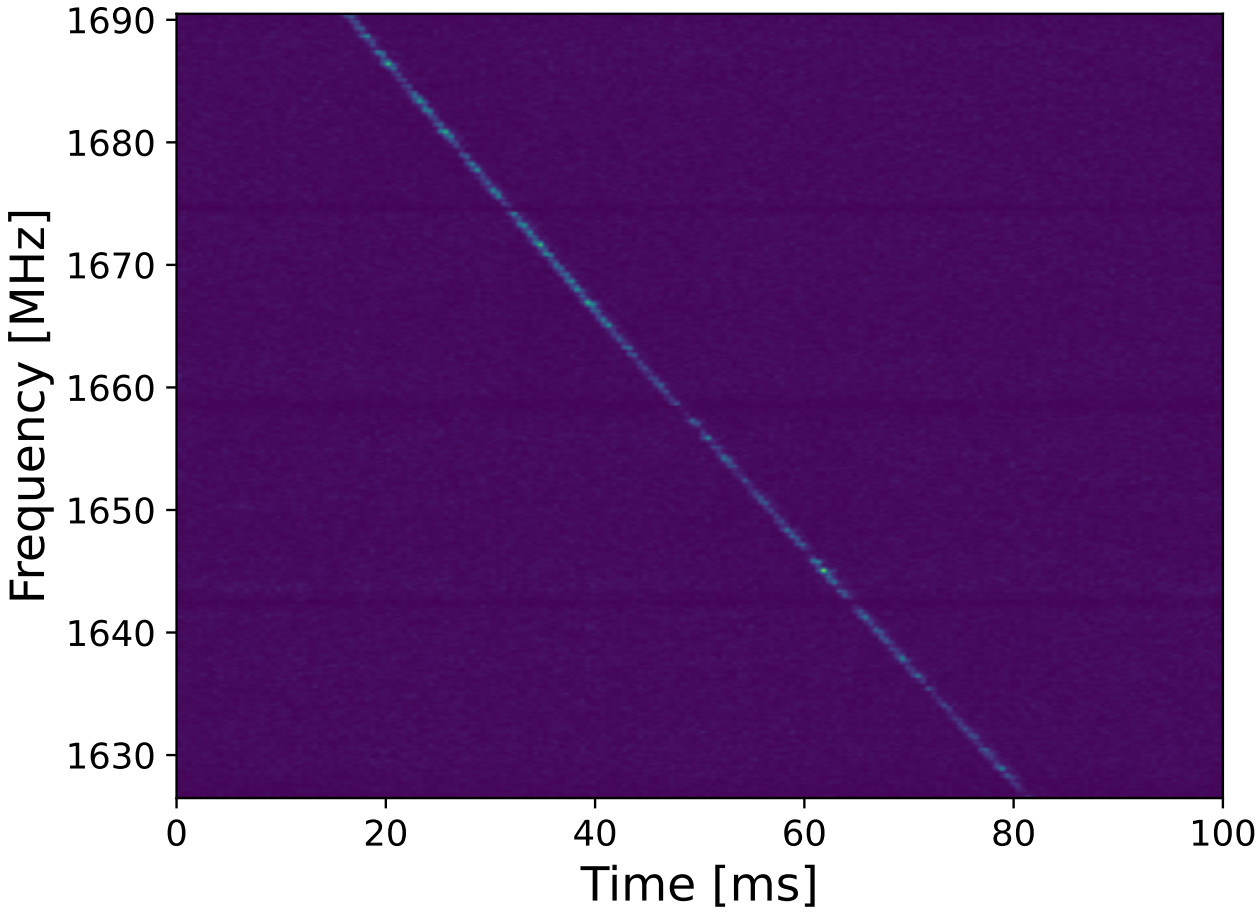


Coherent de-dispersion

- Emission from transients, such pulsar and fast radio bursts (FRBs) are dispersed by the ISM
- The dispersive delay between two frequencies v_1, v_2 [MHz]

$$\Delta t = 4.15 \times 10^{-6}\ DM\ (v_1^{-2} - v_2^{-2})\ [ms],$$

- DM is the dispersion measure
- Apply digital filter to remove dispersive delay from base-band signals
- Computational intensive
  - Very long FFTs, both forward and backward



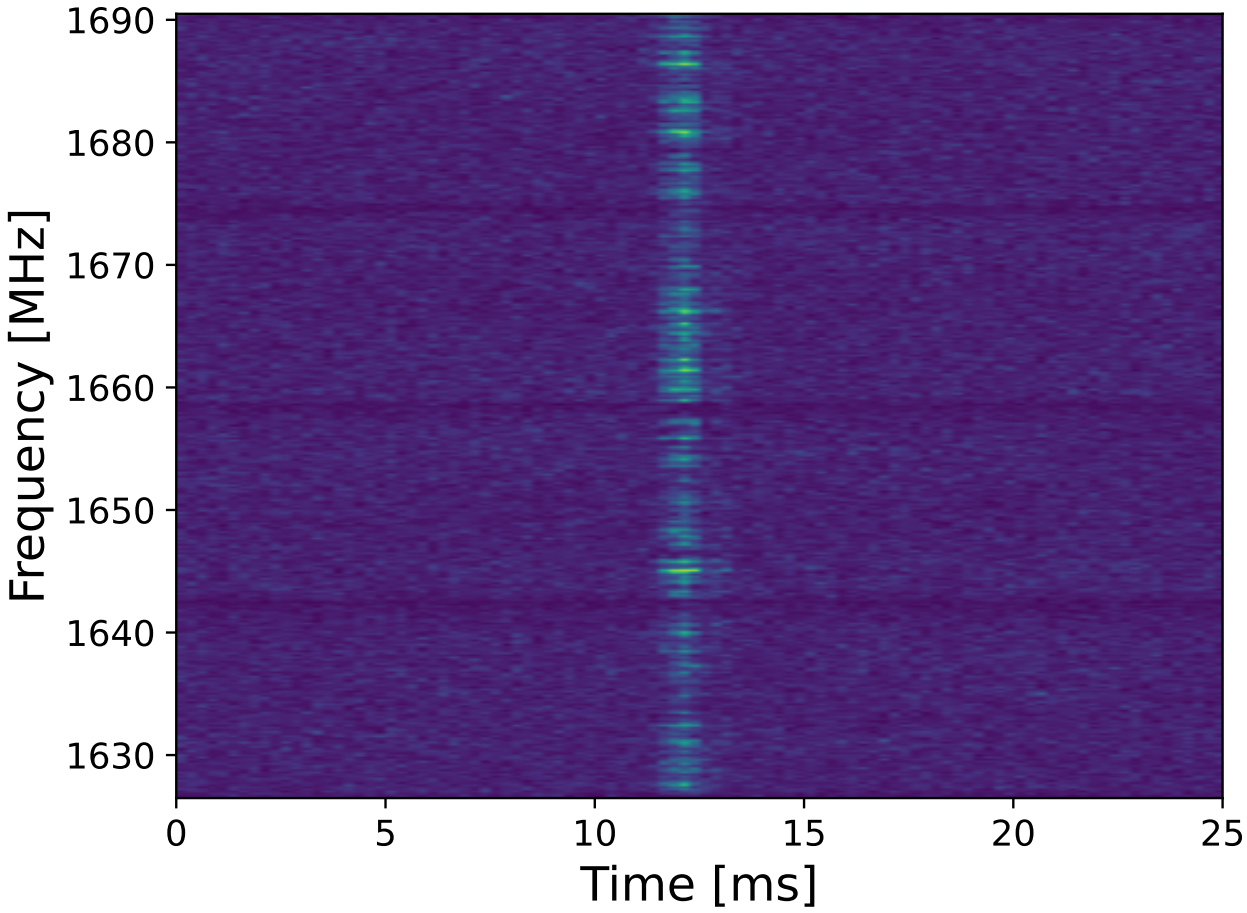
FRB 121102 on Ar-Ef baseline
DM = 561 pc.cm^{-3}

Coherent de-dispersion

- Emission from transients, such pulsar and fast radio bursts (FRBs) are dispersed by the ISM
- The dispersive delay between two frequencies v_1, v_2 [MHz]

$$\Delta t = 4.15 \times 10^{-6}\ DM\ (v_1^{-2} - v_2^{-2})\ [ms],$$

- DM is the dispersion measure
- Apply digital filter to remove dispersive delay from base-band signals
- Computational intensive
  - Very long FFTs, both forward and backward



FRB 121102 on Ar-Ef baseline

DM = 561 pc.cm^{-3}

Conclusion and Outlook

- Can already correlate 95% of EVN experiments
- Most important features are implemented
  - Mixed-bandwidth correlation
  - Hanning smoothing
  - Multiple simultaneous phase centres
  - Coherent de-dispersion
- Under active development to implement remaining features
- No exact time line yet when SFXGPU will be used in operations

https://code.jive.eu/keimpema/sfxgpu

Conclusion and Outlook

- Can already correlate 95% of EVN experiments
- Most important features are implemented
    - Mixed-bandwidth correlation
    - Hanning smoothing
    - Multiple simultaneous phase centres
    - Coherent de-dispersion
- Under active development to implement remaining features
- No exact time line yet when SFXGPU will be used in operations