Grant number: 101093934

WP4 – Data Transport & Correlator/BeamFormer

John W. Romein

Grant number: 101093934

Agenda

- technology assessment
- (GPU) Radio Blocks
- data transport
- applications
- cluster

Grant number: 101093934

Technology assessment

- studied GPUs, DPUs, TPUs, CPUs, FPGAs, NICs, …
- results:
  - near antennas: FPGAs
  - network: 400 GbE + 800 GbE switch
  - correlator/beam former: GPUs + (smart) NICs
- M4.1; D4.1

Grant number: 101093934

GPU Radio Blocks

Grant number: 101093934

GPU Radio Blocks

- channel filter
- VLBI filter + delay
- Tensor-Core Correlator
- Tensor-Core Beam Former
- dedispersion
- generic input handling
- ...

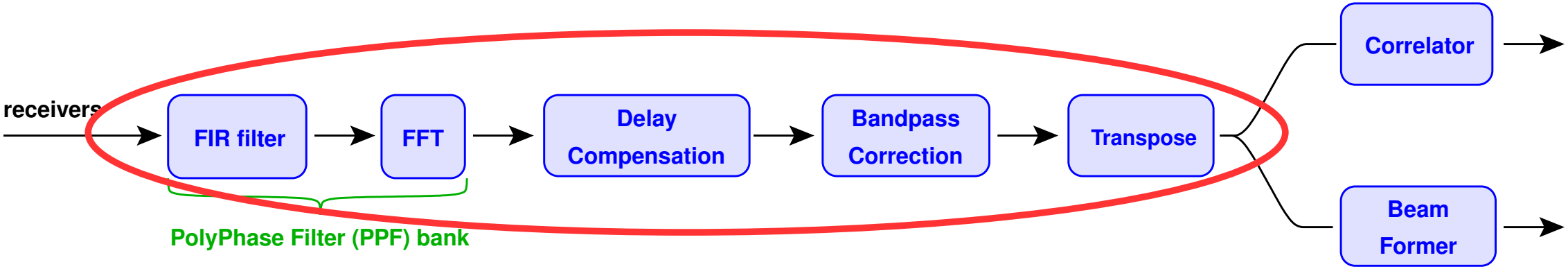
} later

- CUDA wrappers
- Power Measurement Toolkit

} generic

Grant number: 101093934

GPU channel filter

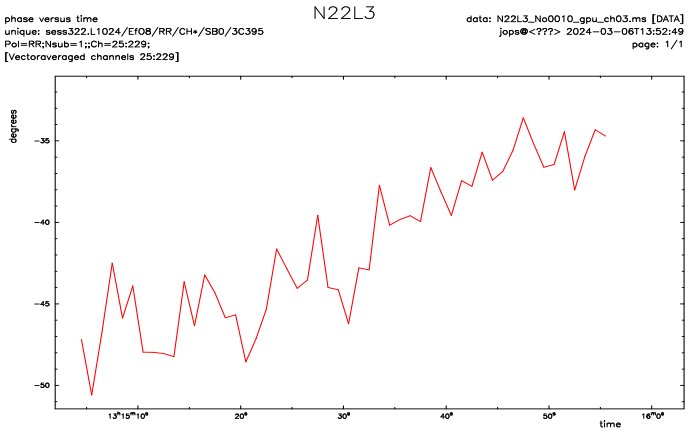


- single kernel → read and write memory ~~5x~~ 1x
- difficult to achieve *both* flexibility and high performance on all GPU memory types

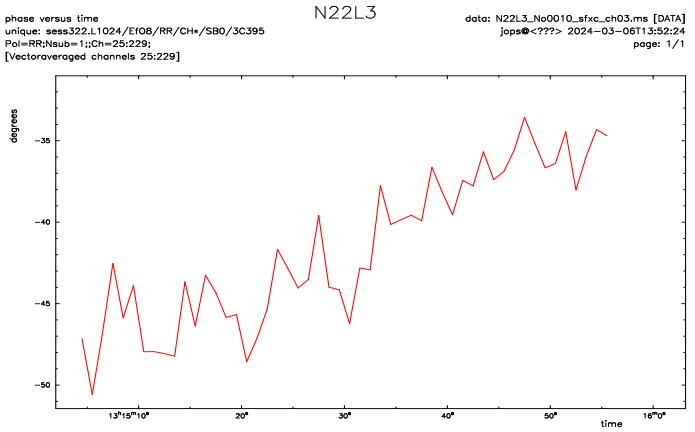
Grant number: 101093934

VLBI filter + delay

- working delay correction on GPU
- fp16 accuracy verified (on Grace ARM CPU)
- merge with channel filter?



GPU



CPU

Phase vs. Time on Ef-O8 baseline

Grant number: 101093934

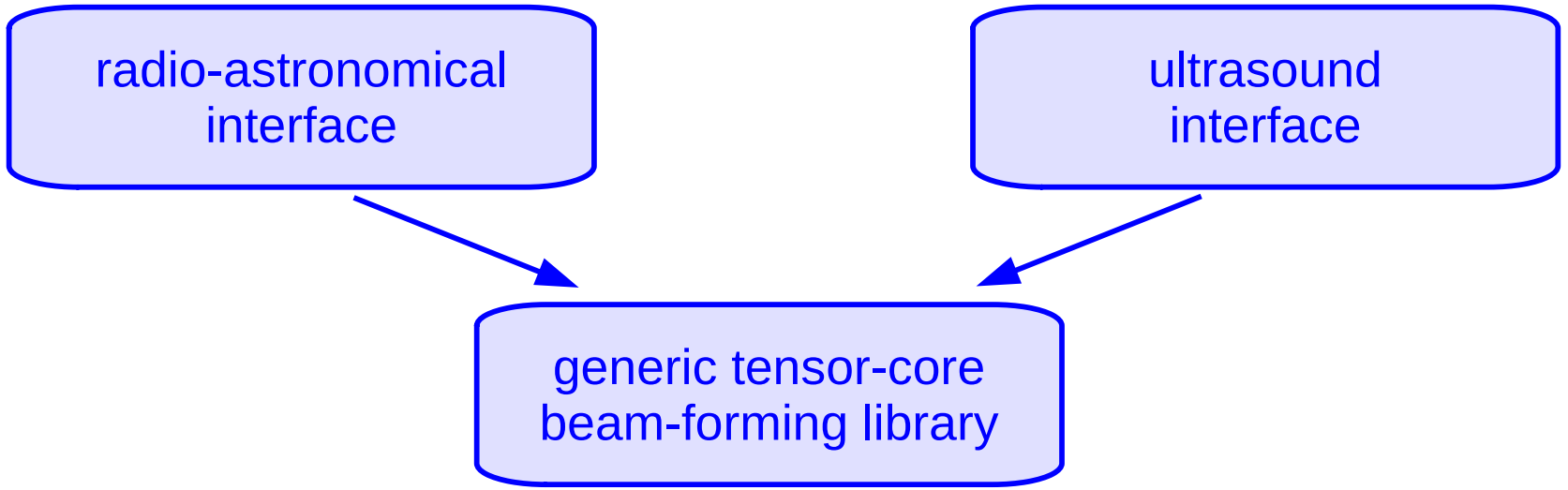
Tensor-Core Correlator

- see tomorrow's presentation

Grant number: 101093934

Tensor-Core Beam Former

- collaboration with Netherlands eScience Center + Erasmus Medical Center



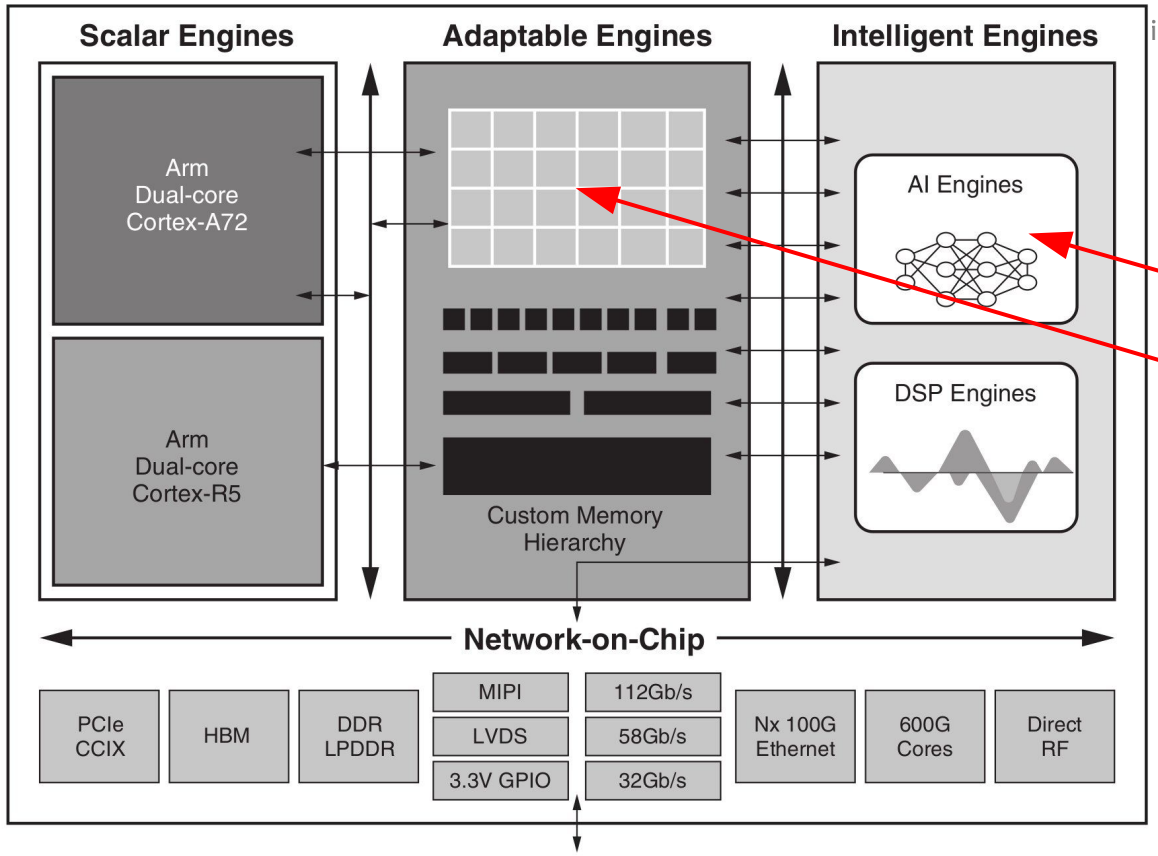
- promising results

Grant number: 101093934

FPGA technology exploration

Grant number: 101093934

ACAP exploration



vector processors

traditional FPGA logic

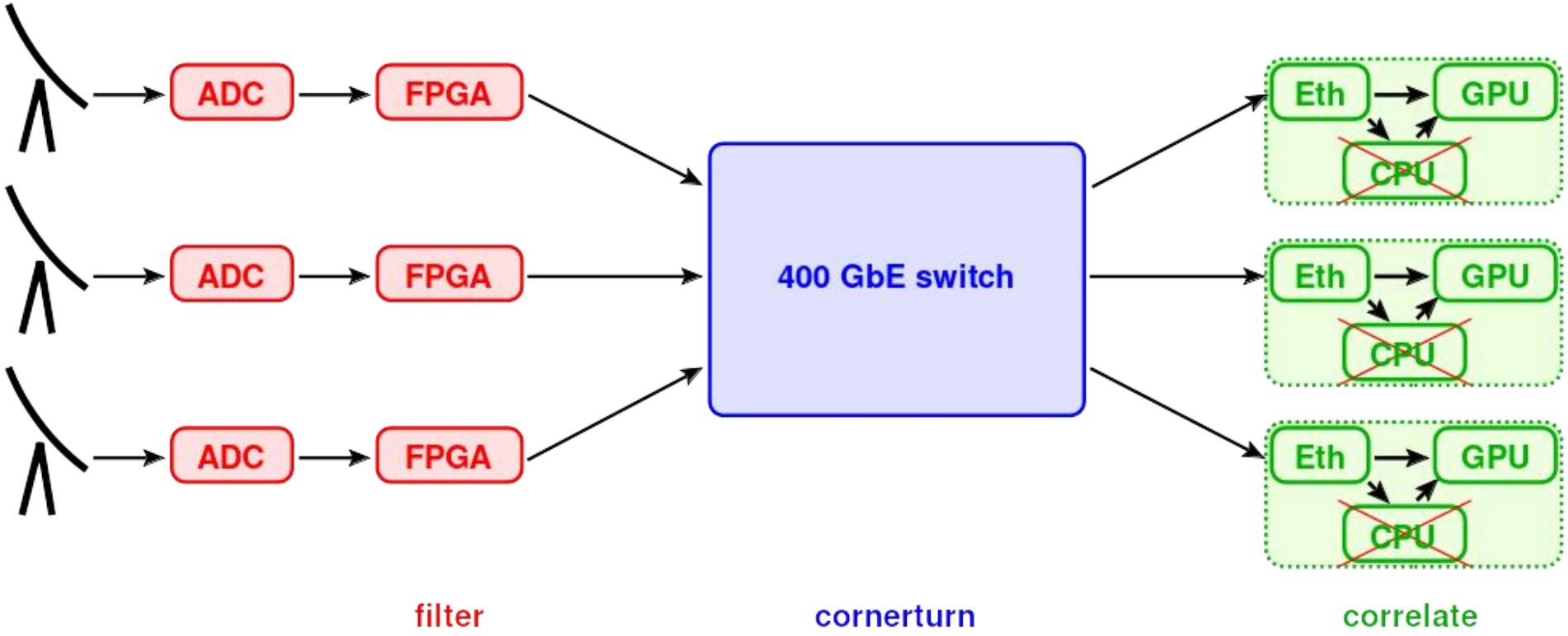
- Xilinx *Adaptive Compute Acceleration Platform*
- explore use of “AI engines” for signal processing
  - collaboration with universities
  - successful → FPL publication

Grant number: 101093934

Data Transport

Grant number: 101093934

Data Transport



- digitizer → corner turn → correlator
- stream data: FPGA → switch → GPU
- VLBI: disk → switch → GPU

Grant number: 101093934

Data Transport

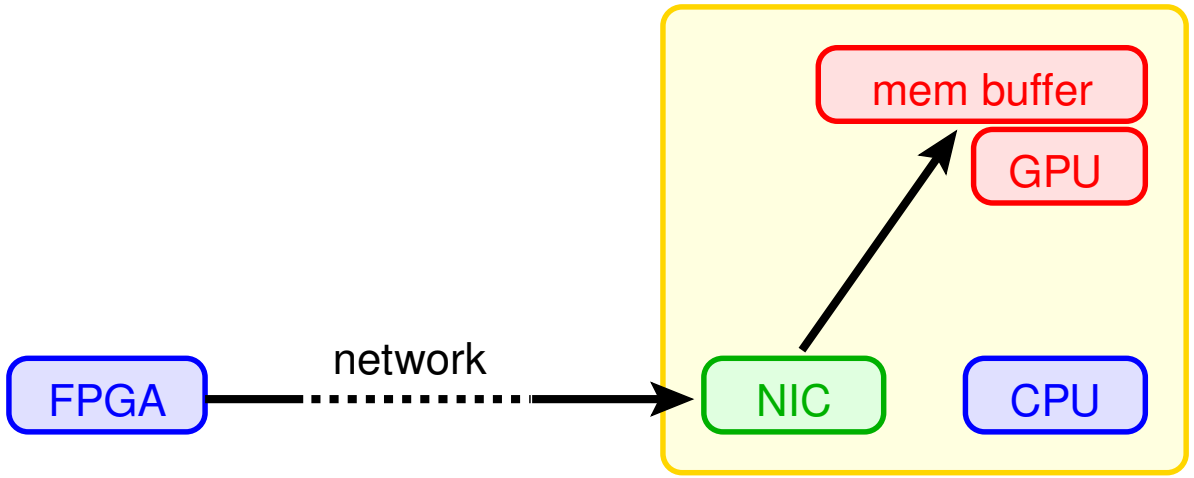
- >40 GbE: too much OS overhead → explore new methods
  - Remote Direct Memory Access (RDMA)
  - Data Plane Development Kit (DPDK)

  } different advantages & disadvantages
- D4.2

- implementing demo correlator 

Grant number: 101093934

RDMA



iWave Agilex 7 dev kit

- FPGA writes data in GPU memory
  - complex protocol (RoCE v2); complex FPGA firmware
- VLBI (offline): storage node → GPU

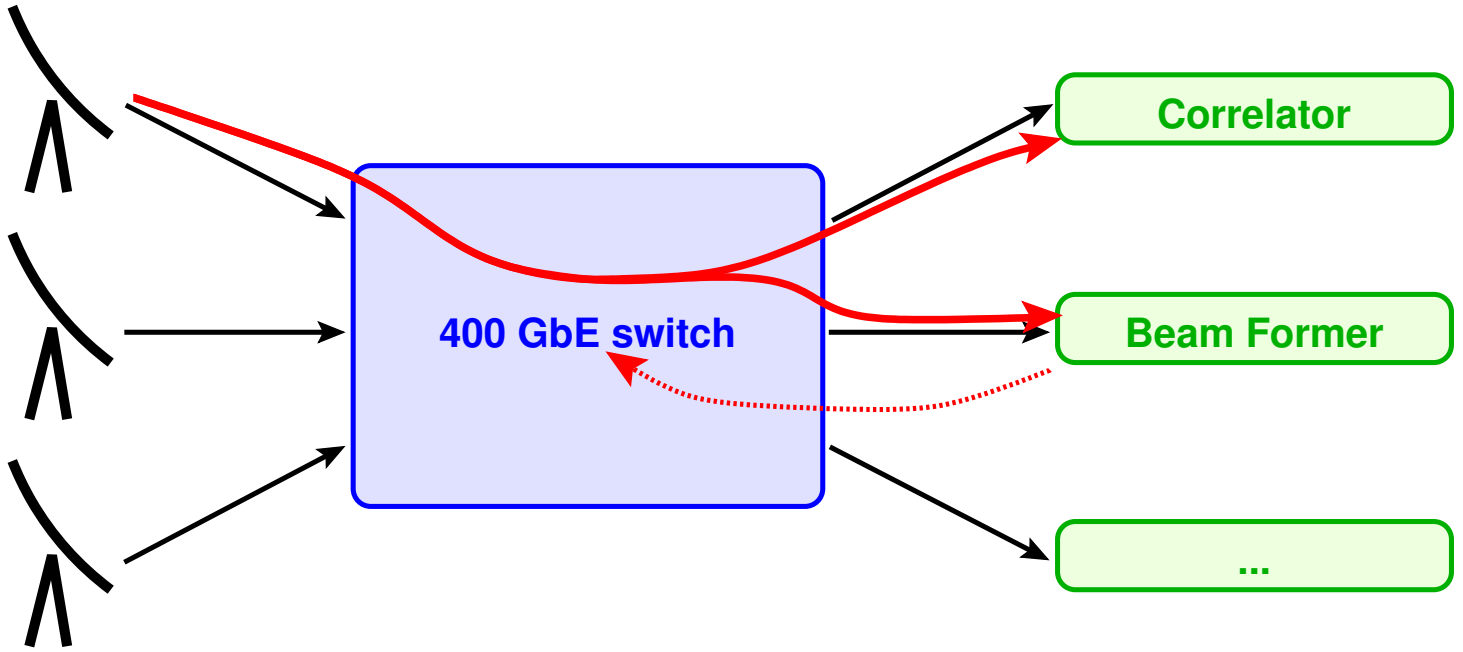
Grant number: 101093934

Data Plane Development Kit

- see tomorrow's presentation

Grant number: 101093934

IP multicast experiments



- correlator / beam-former applications subscribe to antenna streams
- switch replicates data
- ≤5 replications: 336 Gb/s on 400 GbE switch

Grant number: 101093934

Applications

Grant number: 101093934

Correlator applications

- AARTFAAC correlator
  - all relevant GPU Radio Blocks integrated
- DPDK/RDMA demo correlator
  - ALMA GPU correlator viability study
- EVN/(ng)EHT correlator
  - efficient 2/4/8/16 VDIF decoding (on CPU)
- LOFAR correlator
  - various GPU kernel optimizations
- VIRAC correlator
  - being ported from AARTFAAC correlator
  - input format changes

Grant number: 101093934

Applications' use of Radio Blocks

✔ ready
✘ planned
 work in progress

	channel filter	VLBI filter + delay	tensor-core correlator	tensor-core beam former	CUDA wrappers	Power Measurement Toolkit	DPDK	RDMA	IP multicast
AARTFAAC	✔		✔		✔	✘	✘		?
DPDK/RDMA demo			✔		✔	✘		✘	
VIRAC 	✔		✔		✔				
LOFAR 	✘		✘		✔	✔	✘		✘
EVN 		✔	✘	?	✘	✘			
eMERLIN		?	?						
KVN		?	✘	?	?		?		
ngEHT		✘	✘					✘	

Grant number: 101093934

Computer cluster

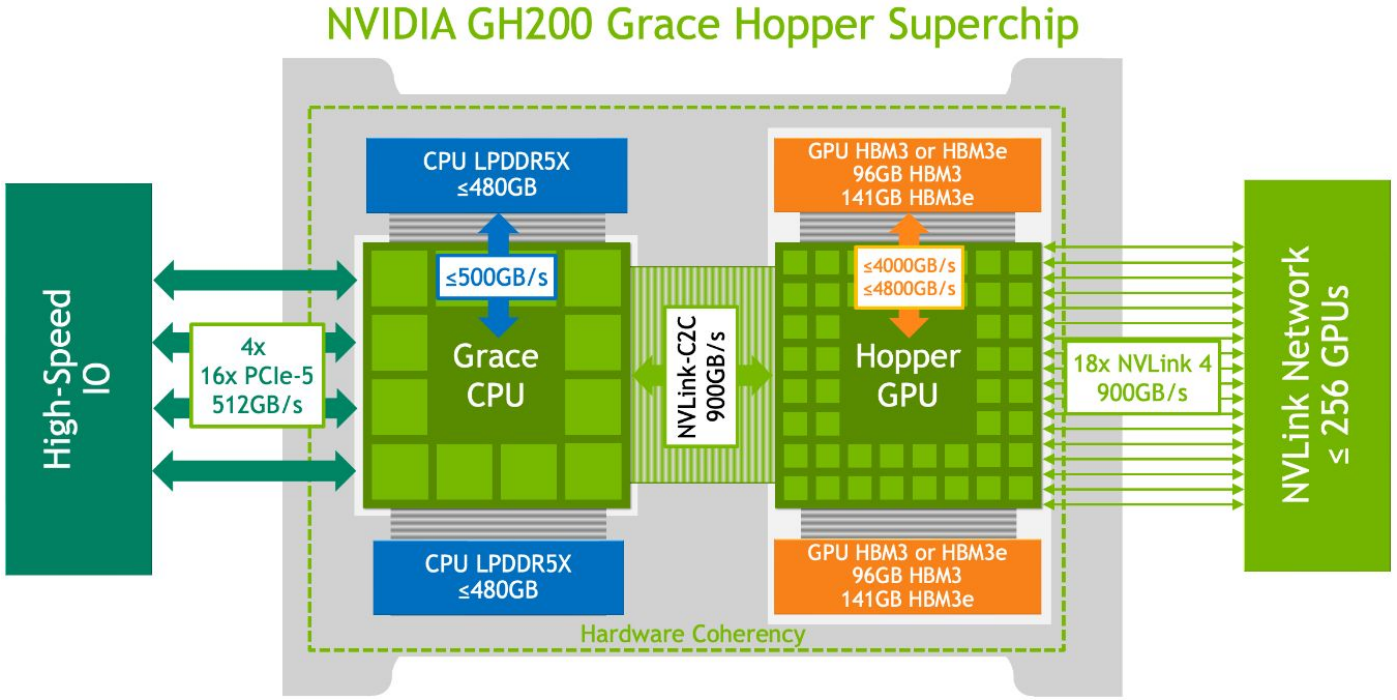
Grant number: 101093934

Computer cluster

- covers needs from WP4 + WP5
  - WP4: GPUs, network demands
  - WP5: GPUs, storage
- partly acquired

Grant number: 101093934

Available



- 2x NVIDIA Grace Hopper
  - most powerful GPU
  - 14x more CPU↔GPU bandwidth than PCIe gen4

- 2x Jetson AGX Orin
  - SoM with tightly coupled CPU/GPU
  - edge computing
  - not fast, but highly energy efficient

Grant number: 101093934

Rest of the cluster

- preparing acquisition
- plan:
  - head node (incl. SSD storage)
  - ≥4 “standard” GPU nodes
  - 2 fat GPU nodes
- (workstation-grade) GPUs
  - still PCIe gen4; reserve budget for next-gen GPUs
- 800 GbE switch + 400 GbE NICs
  - difficult to obtain with “right” connector type
- reserve budget for next-gen Orin

Grant number: 101093934

Milestones

number	milestone name	month	means of verification	
4.1	technology review	8	report of internal meeting	✔
4.2	verification of high-speed prototype	18	report of test results	
4.3	demonstrator benchmark defined	30	document	

Grant number: 101093934

Deliverables

number	deliverable name	lead	type	dis. lvl.	month	
4.1	assessment of the applicability of next-generation technology	ASTRON	R	PU	12	✔
4.2	high-speed data handling techniques, such as RDMA	ASTRON	R	PU	12	✔
4.3	prototype high-speed data transport	UBx	DEM	PU	24	
4.4	basic correlator on tensor-core architecture	ILT	OTHER	PU	24	
4.5	beam forming and coherent dedispersion modules	JIV-ERIC	OTHER	PU	36	
4.6	next-generation correlator demonstration	ILT	DEM	PU	48	

Grant number: 101093934

Conclusions

- many parallel efforts
  - (GPU) radio blocks
  - data transport
  - applications
  - cluster
- promising results