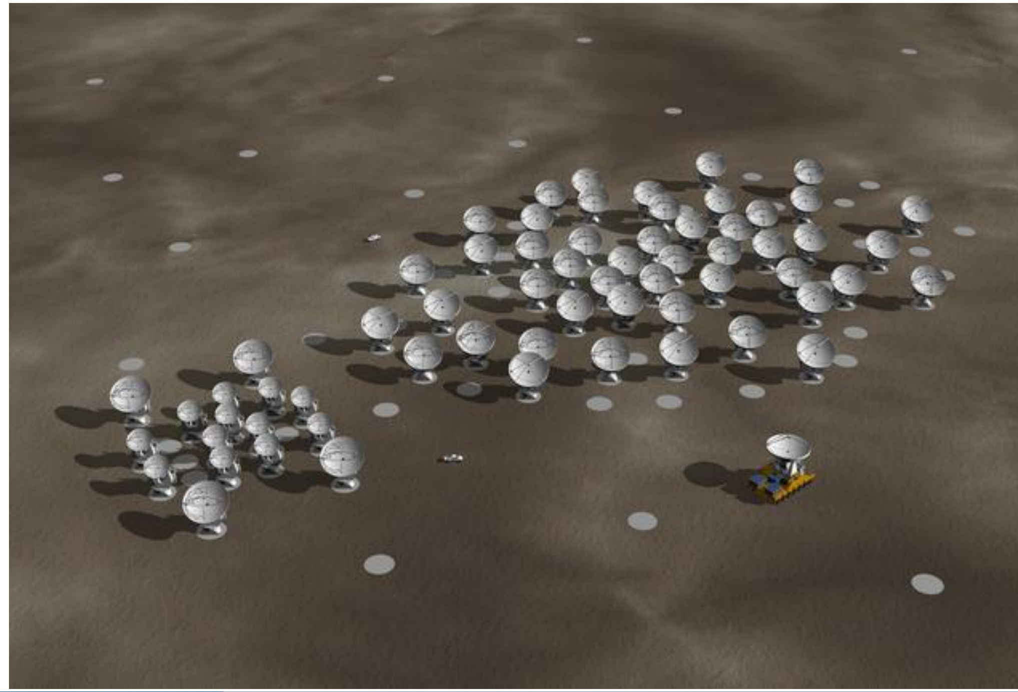
The development of GPU Spectrometer for the ALMA

Jongsoo Kim
Korea Astronomy and Space Science Institute and
University of Science and Technology

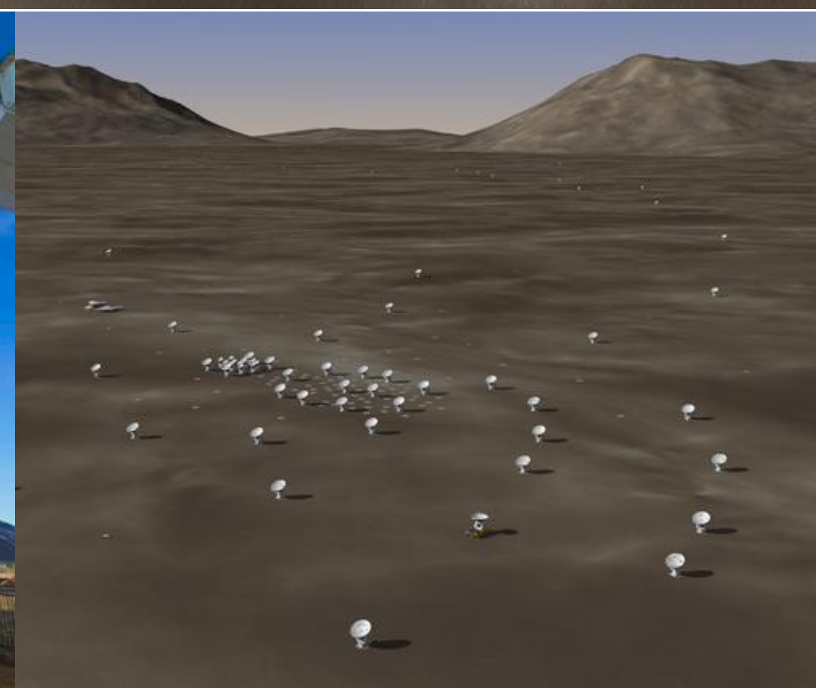
Antennas



50x12m-Array + 12x7m-Array + 4 12m-TP Array

Longest baseline: 16 km

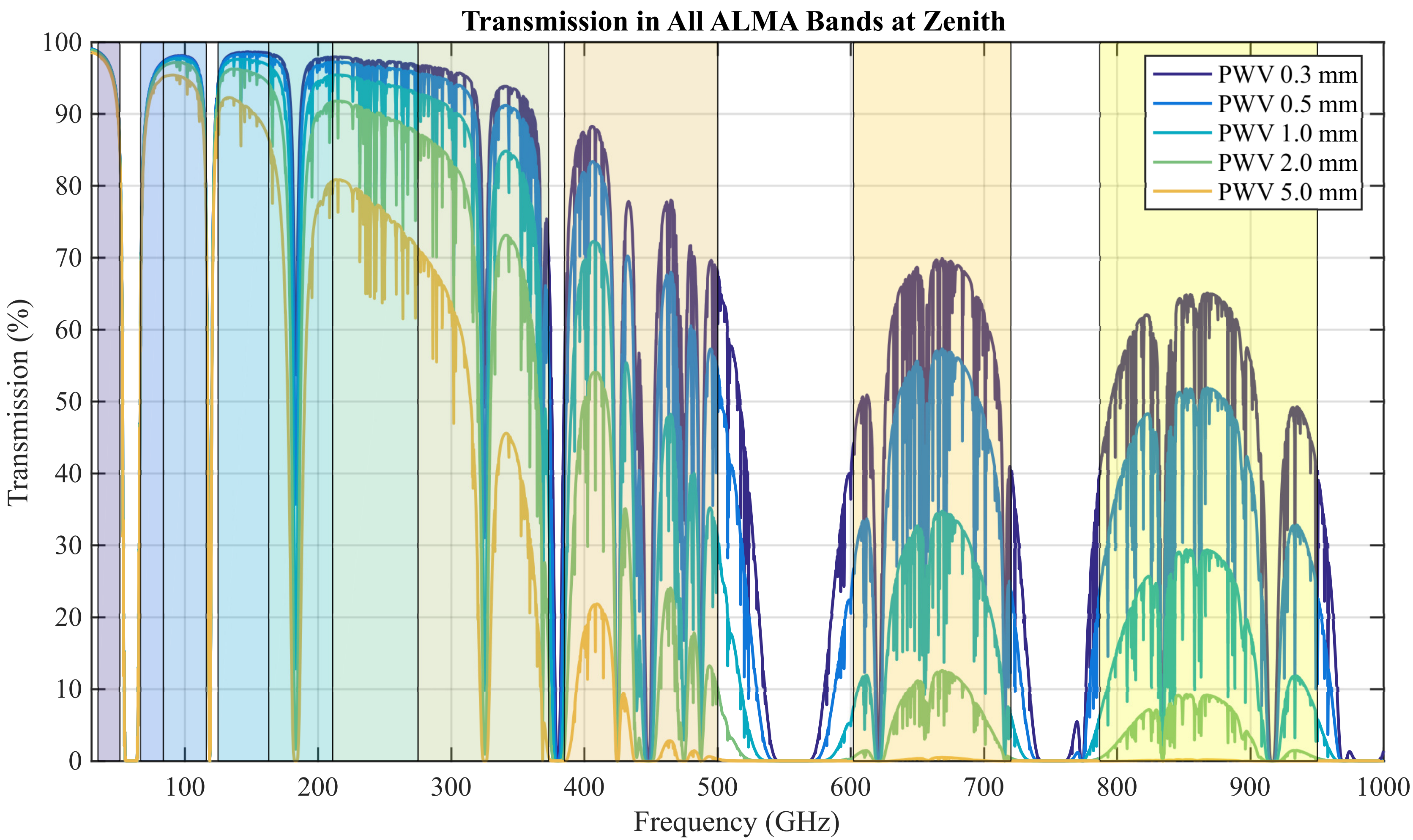
Construction Completed in 2013

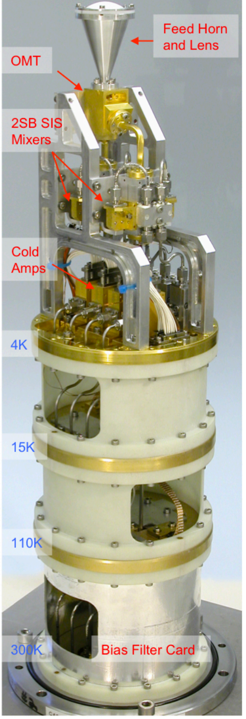


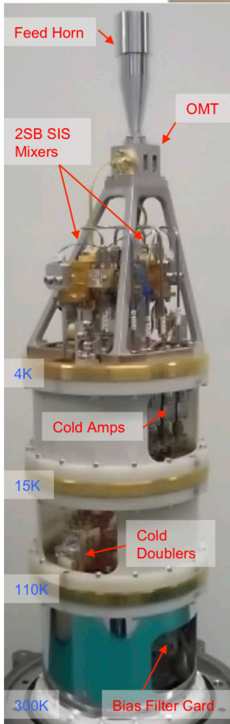
- 50x12m-Array
- 12x7m-Array + 4 12m-TP Array
- Longest baseline: 16 km

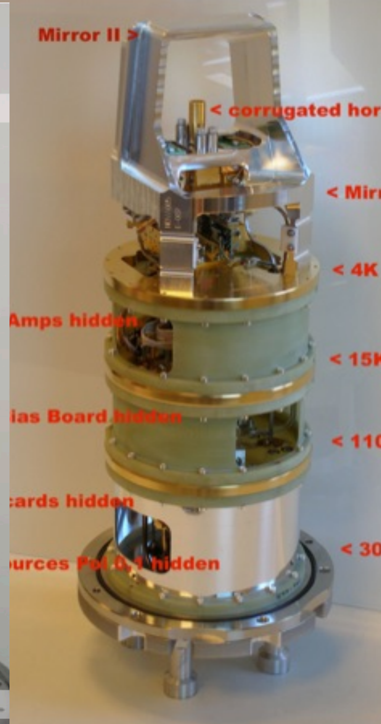
Locations of ALMA and GMT

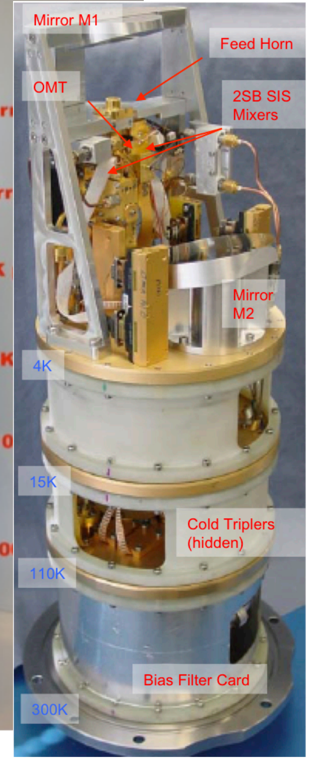
Atmospheric Transparency

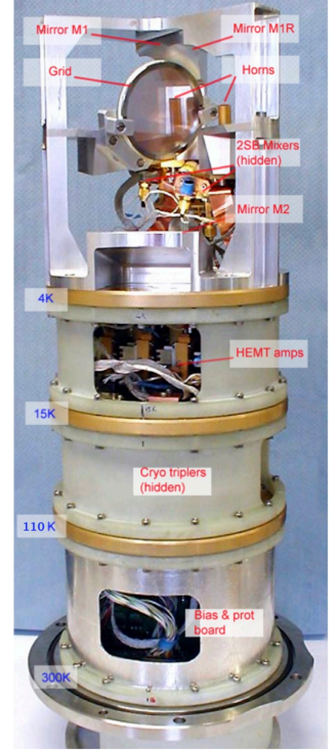


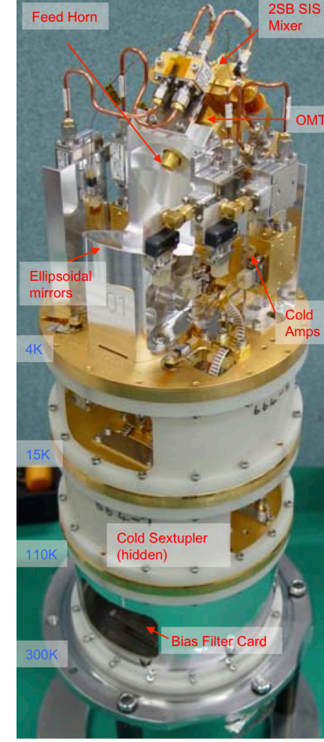


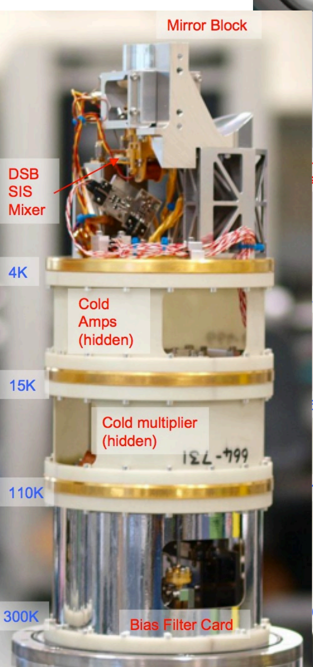


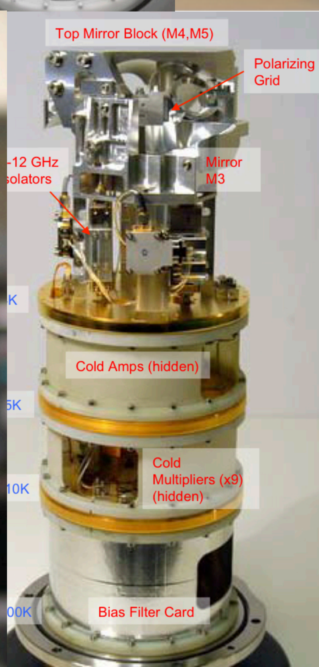






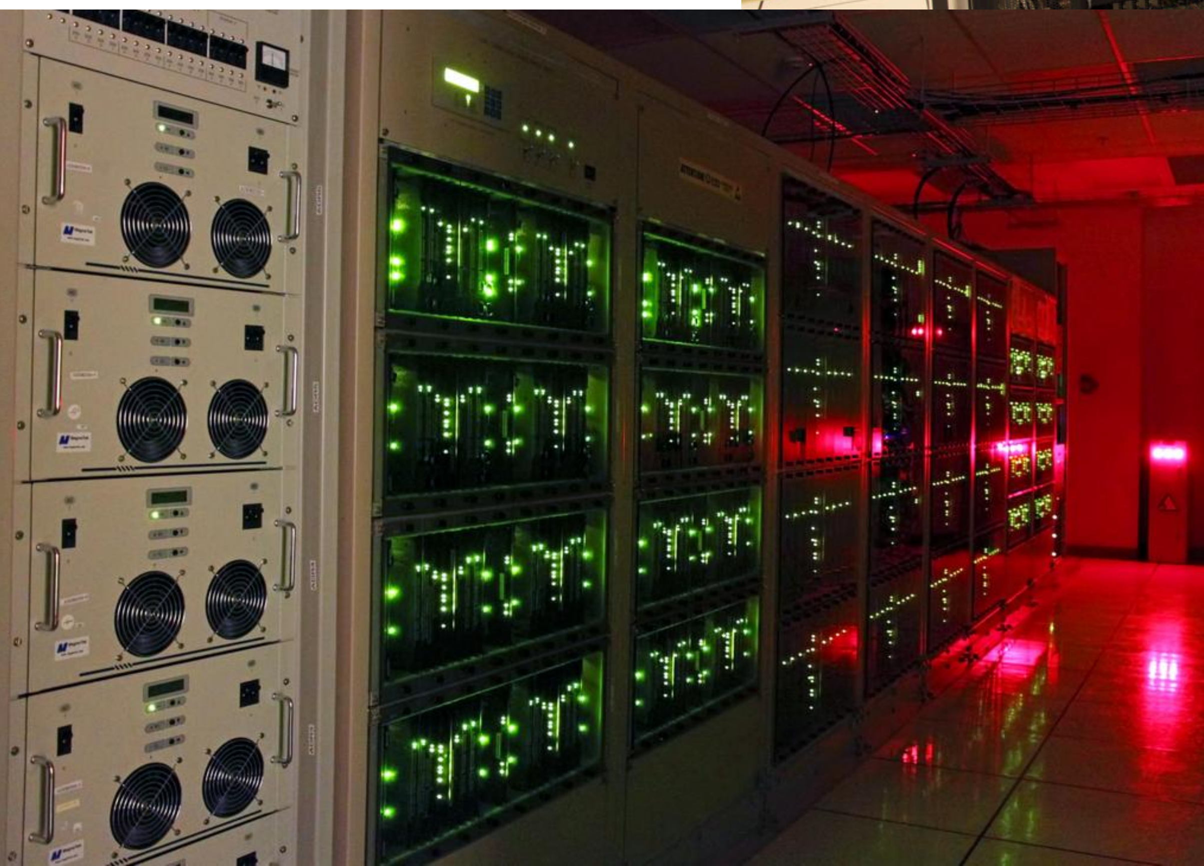






ALMA Band	Frequency Range (GHz)	Receiver Noise (K) over 80% of the RF band	Temperature (K) at any RF Frequency	To be produced by	Receiver Technology
1	31 - 45	17	26	tbd	HEMT
2	67 - 90	30	47	tbd	HEMT
3	84 - 116	37	60	HIA	SIS
4	125 - 163	51	82	NAOJ	SIS
5*	162 - 211	65	105	OSO	SIS
6	211 - 275	83	136	NRAO	SIS
7	275 - 373	147	219	IRAM	SIS
8	385 - 500	196	292	NAOJ	SIS
9	602 - 720	175	261	NOVA	SIS
10	787 - 950	230	344	NAOJ	SIS

ALMA correlators

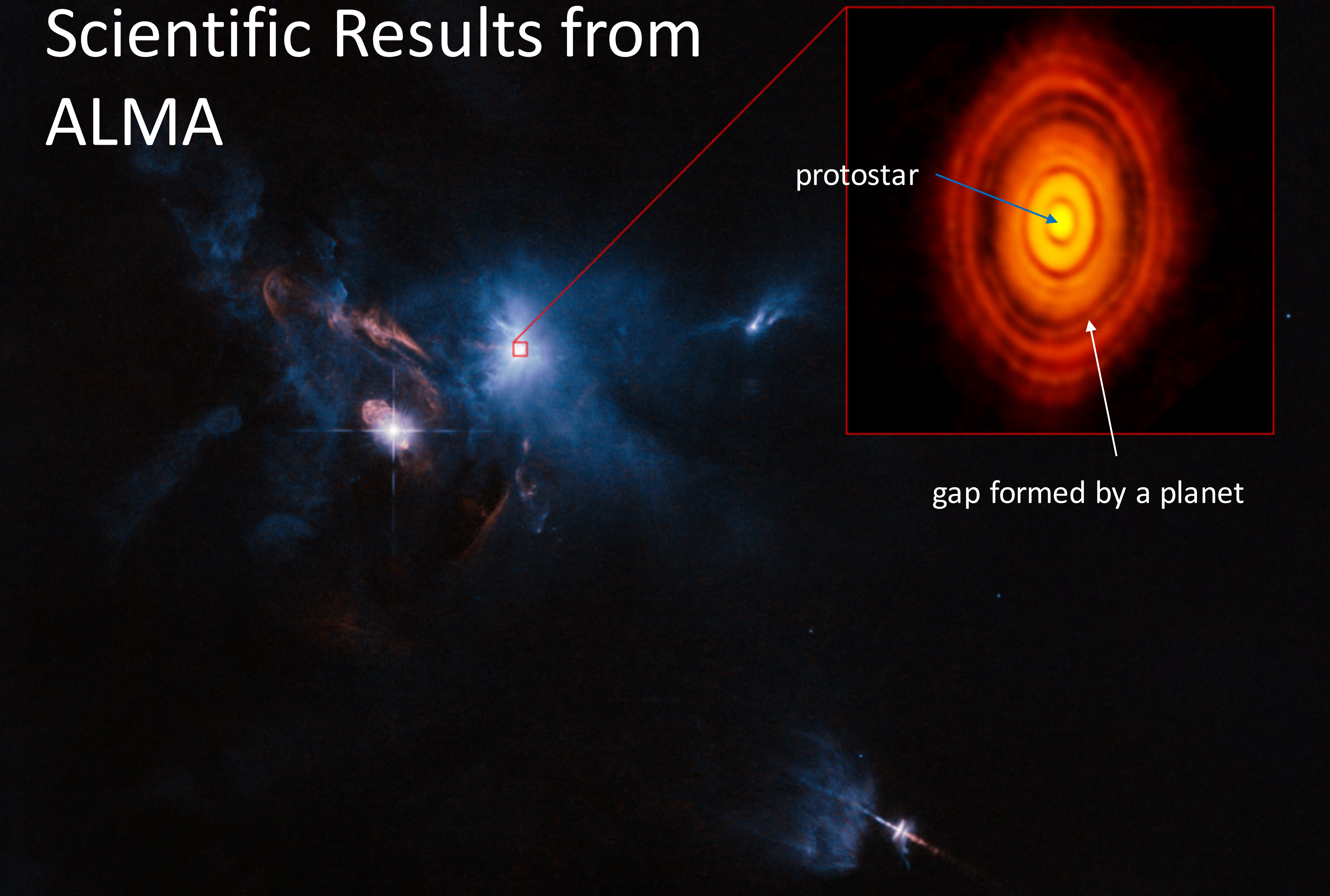


Korean Flag at OSF at 2900m

The Sun never sets on ALMA!

The Atacama Large Millimeter/submillimeter Array (ALMA), an international astronomy facility, is a partnership of the European Organisation for Astronomical Research in the Southern Hemisphere (ESO), the U.S. National Science Foundation (NSF) and the National Institutes of Natural Sciences (NINS) of Japan in cooperation with the Republic of Chile. ALMA is funded by ESO on behalf of its Member States, by NSF in cooperation with the National Research Council of Canada (NRC) and the National Science Council of Taiwan (NSC) and by NINS in cooperation with the Academia Sinica (AS) in Taiwan and the Korea Astronomy and Space Science Institute (KASI).

Scientific Results from ALMA

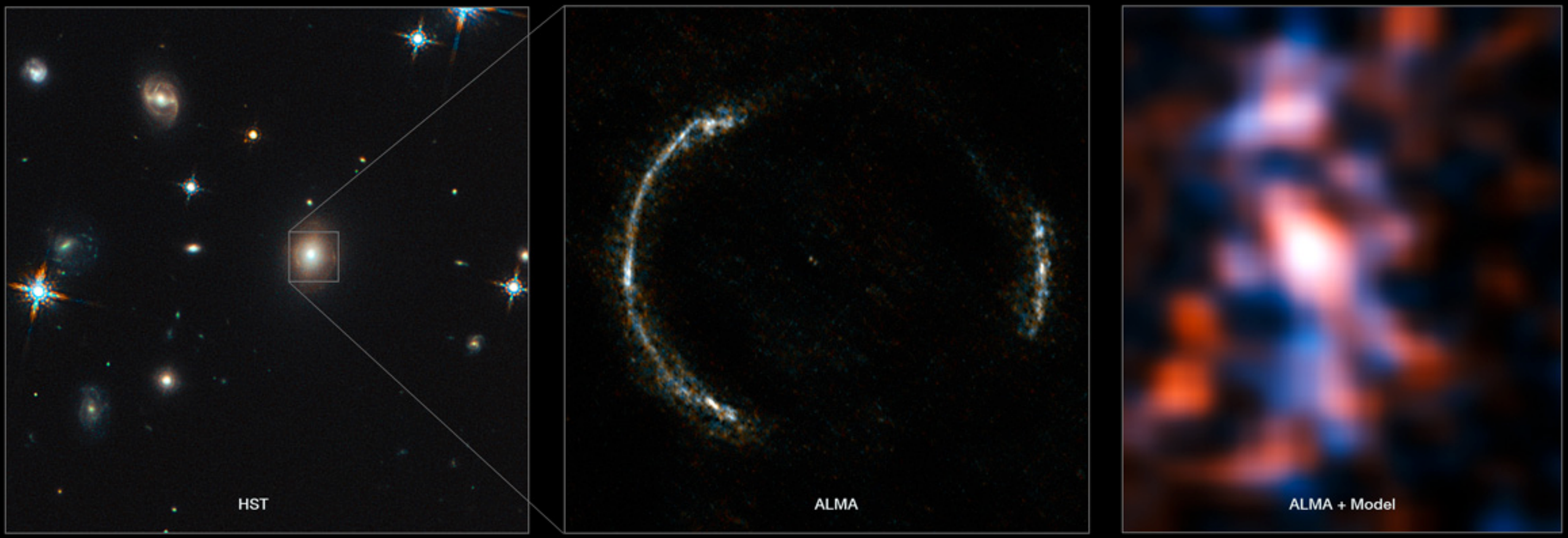


A protostellar disk of HL Tau in Constellation of Taurus at 140 pc.

Scientific Results form the ALMA

Einstein Ring HATLAS J090311.6+003906 (SDP.81)

Reveal most detailed view of molecular gas in the Distant Universe



Credit: ALMA (NRAO/ESO/NAOJ)/Y. Tamura (The University of Tokyo)/Mark Swinbank (Durham University)

Correlation Theorem
FX-correlator or FX-spectrometer

$$R_i(f) = \int_{-\infty}^{+\infty} r_i(t) \mathrm{e}^{2\pi i f t} dt$$

F-step (FT):
~$\log_2(N_c)$ operations per sample

$$\int_{-\infty}^{+\infty} r_i(\tau + t) r_j(\tau) d\tau \Leftrightarrow R_i(f) R_j^*(f)$$

X-step (CMAC):
~ N operations per sample

Technologies for Spectrometers or Correlators

- ASIC (Application-Specific Integrated Circuit)
  - e.g, ALMA 64-antenna Correlator
- FPGA (Field-Programmable Gate Arrays)
  - e.g, ALMA 16-antenna ACA Correlator
- Software (high level-languages, e.g., C/C++, MPI, CUDA/OpenCL)
  - Pros: rapid and easy development, flexibility, expandability
  - Cons: low performance/Watt

ALMA Board Approved Development of New Spectrometer for Morita Array

19 January, 2018

A group of antennas of the Morita Array in ALMA. Credit: ALMA (ESO/NAOJ/NRAO)

In November 2017, the ALMA Board approved the development of a new spectrometer for the Morita Array designed and developed by Japan for the ALMA telescope. The development will be undertaken by Korea Astronomy and Space Science Institute (KASI) and the National Astronomical Observatory of Japan (NAOJ) as part of the ALMA Future Development Program aiming to keep ALMA continuously producing remarkable scientific results for the future.

A principle investigator of the ACA spectrometer project, Jongsoo Kim, said: "The approval of the ACA spectrometer project by the ALMA board is a recognition of the cost-effective development plan and the successful collaboration between KASI and NAOJ." He also added that "An ACA spectrometer will become the first instrumental contribution to the ALMA community from Korea through the East Asia partnership."

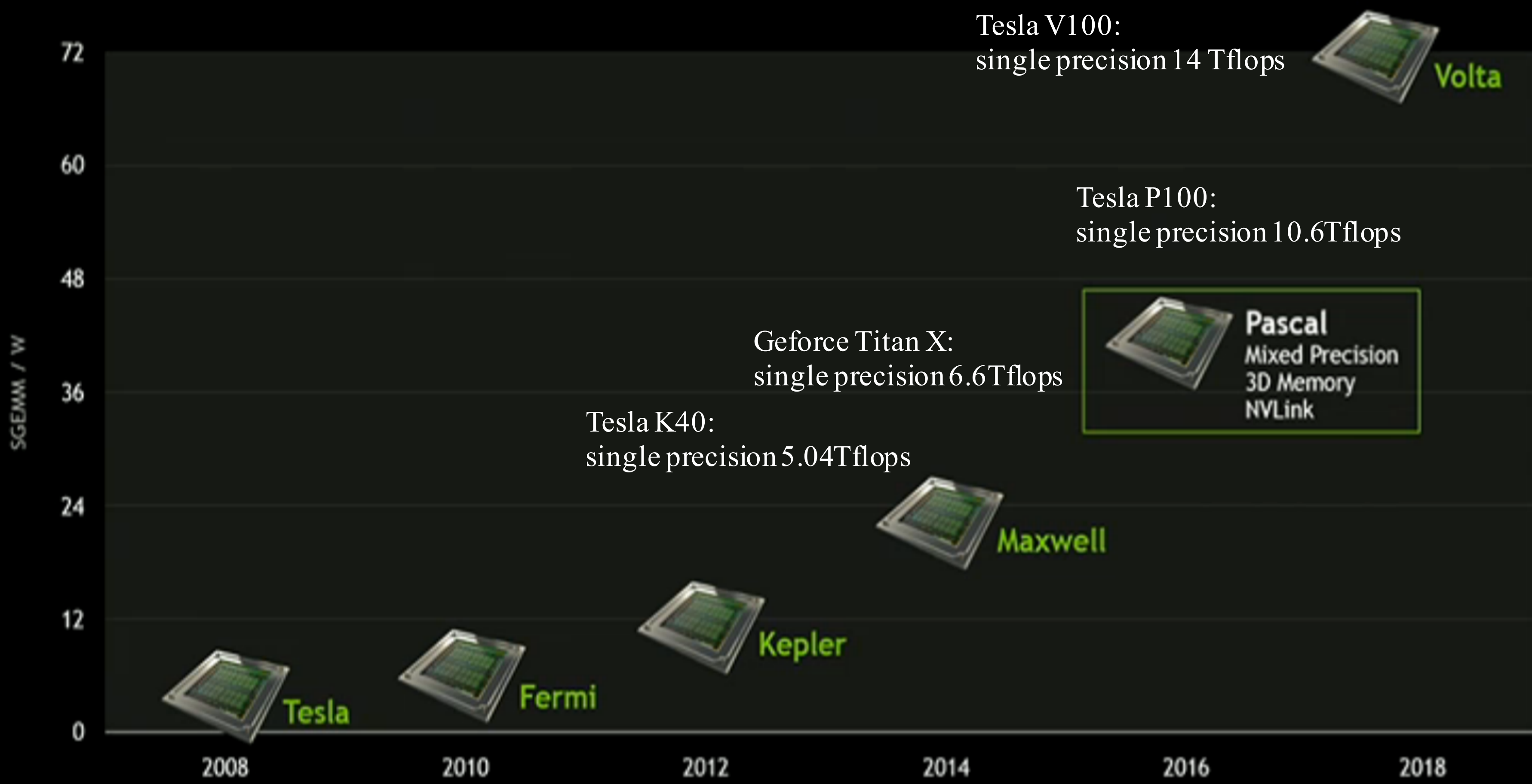
Three-stage development plan

- Stage 1: ACA Spectrometer for four antennas
  - Meet the current specification of the ACA Correlator
  - Composed of 4 GPU servers, 4 GPU cards per server, 2 DRXP cards per server, and 12 optical splitters per server
  - Will be installed on February 2020
- Stage 2: ACA Correlator with 16 antenna based on the GPU technology
  - Aligned with the upgrade plan of the 64-antenna ALMA Correlator
    - Resolution upgrade: 4096 (each pol) → 32768 (each pol)
    - Sampling bit upgrade: 3bit → 4bit
    - bandwidth upgrade: 2GHz → 4 GHz
  - Composed of 4 GPU servers, 8 GPU cards per server, 40GE or 100GE 16 ports
  - Will be installed on February 2022

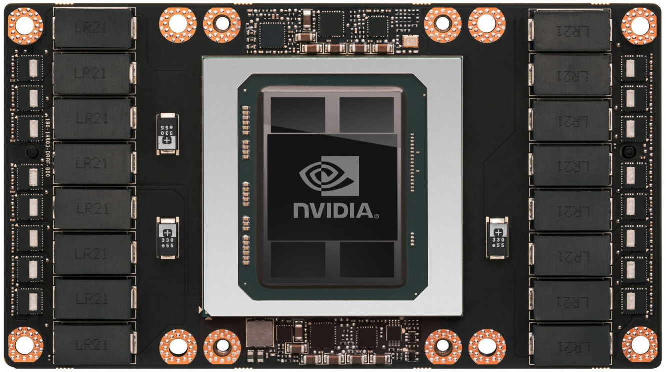
Three-stage development plan (continued)

- Stage 3: ALMA Correlator based on the GPU technology
  - all 66 ALMA antennas
  - 4bit or 8bit sampling
  - Poly-phase filter bank
  - Flexible sub array modes (including 12m, 7m, TP arrays, and any configuration)

NVIDIA GPU Roadmap

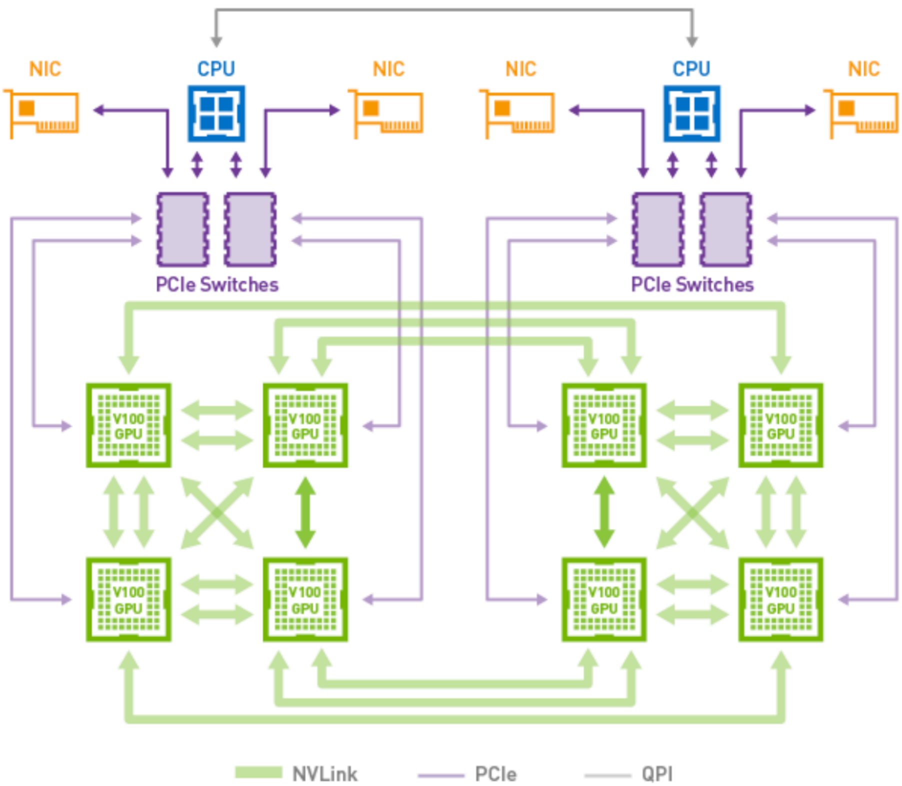


NVIDA TESLA V100 based on Volta Architecture



- NVLink: 25GB/s in each direction
- HBM2
- Unified Memory

- 7.8 TFLOPS of FP64
- 15.7 TFLOPS of FP32 (6.6 for Titan X)
- 125 TFLOPS for Deep Learning

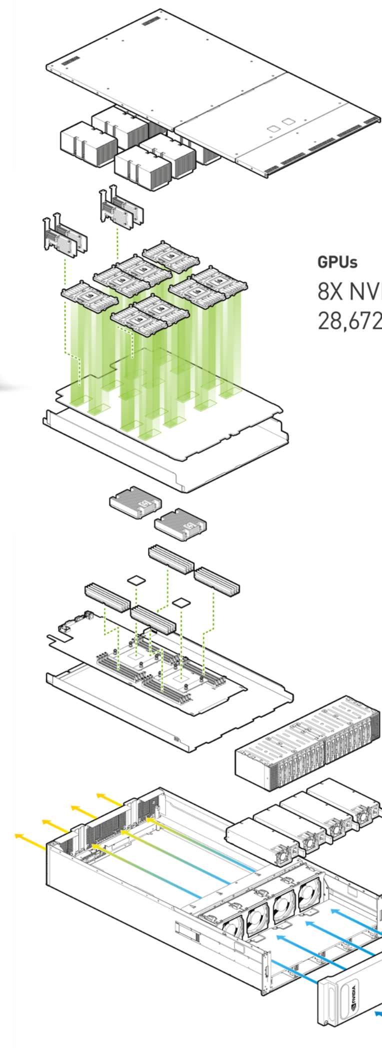


SPECIFICATIONS

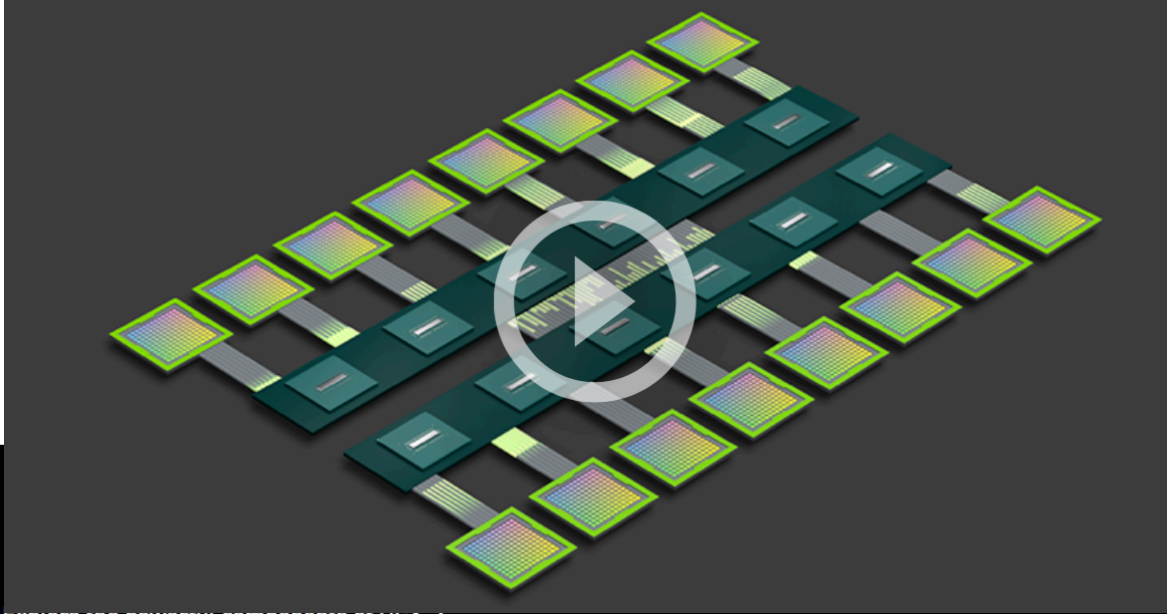
 

	Tesla V100 PCIe	Tesla V100 SXM2
GPU Architecture	NVIDIA Volta	
NVIDIA Tensor Cores	640	
NVIDIA CUDA® Cores	5,120	
Double-Precision Performance	7 TFLOPS	7.5 TFLOPS
Single-Precision Performance	14 TFLOPS	15 TFLOPS
Tensor Performance	112 TFLOPS	120 TFLOPS
GPU Memory	16 GB HBM2	
Memory Bandwidth	900 GB/sec	
ECC	Yes	
Interconnect Bandwidth*	32 GB/sec	300 GB/sec
System Interface	PCIe Gen3	NVIDIA NVLink
Form Factor	PCIe Full Height/Length	SXM2
Max Power Comsumption	250 W	300 W
Thermal Solution	Passive	
Compute APIs	CUDA, DirectCompute, OpenCL™, OpenACC	

NVIDIA DGX-1

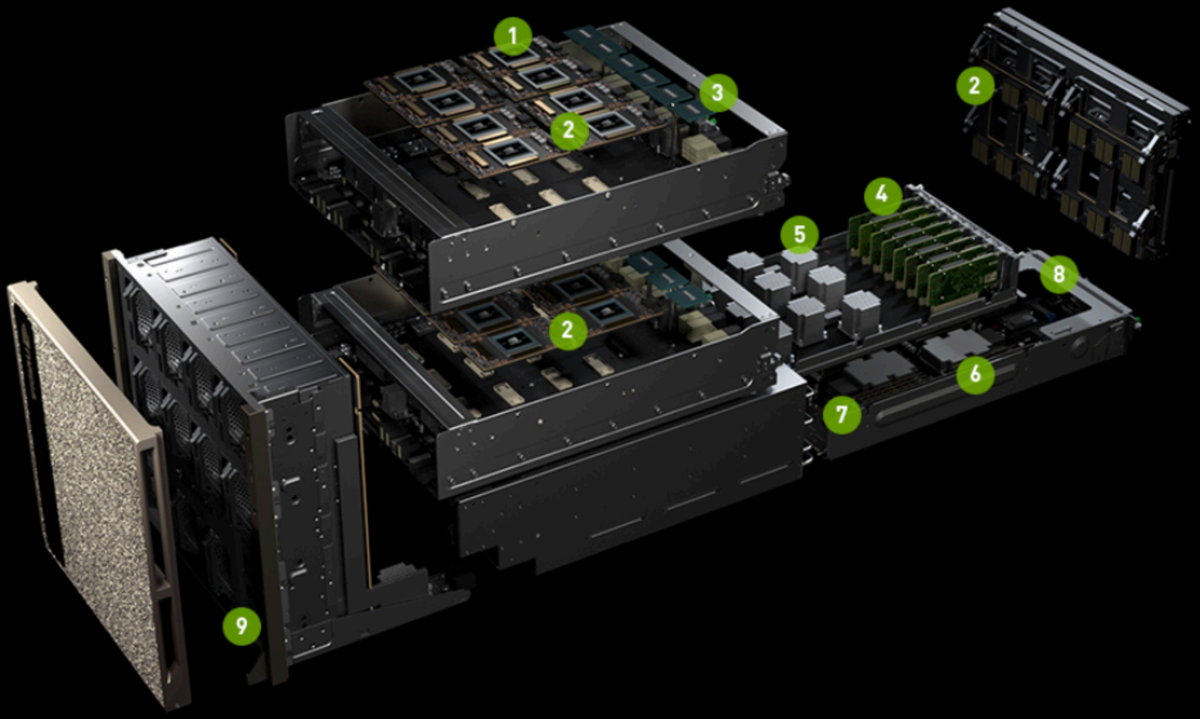


NVIDIA DGX-2

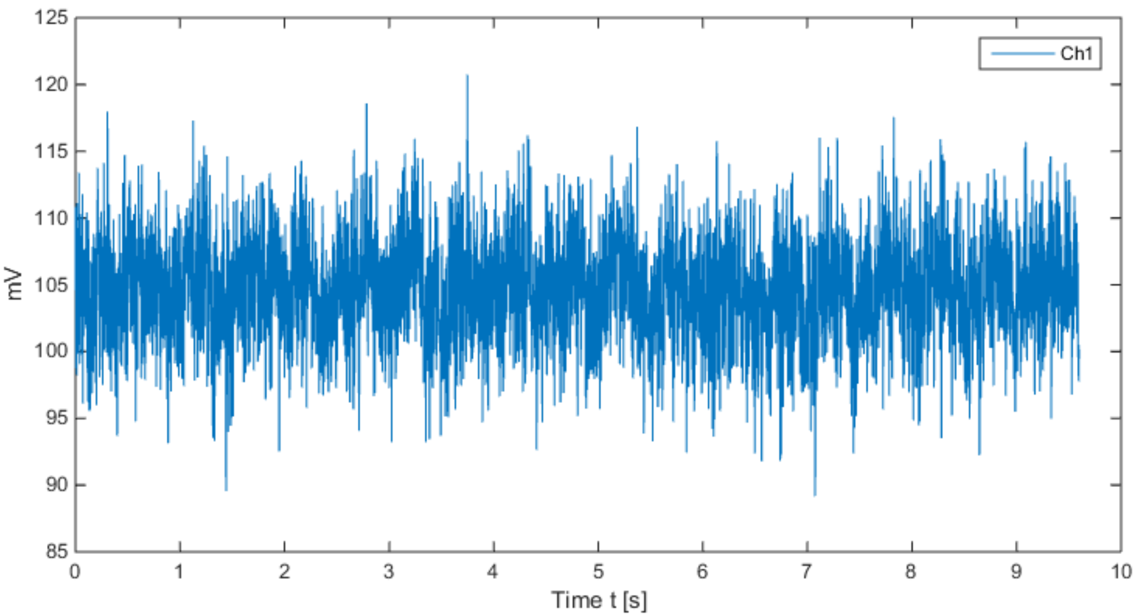


Explore the powerful components of DGX-2.

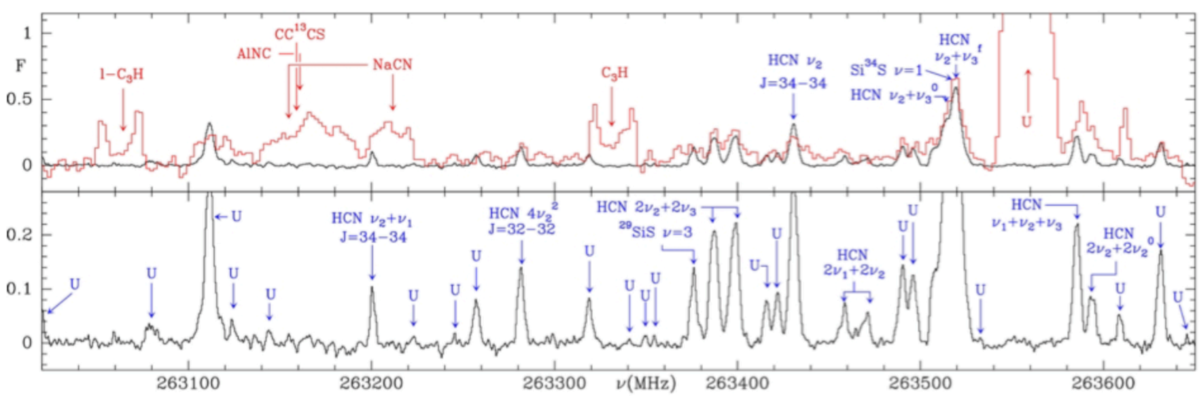
1. NVIDIA TESLA V100 32GB, SXM3
2. 16 TOTAL GPUS FOR BOTH BOARDS, 512GB TOTAL HBM2 MEMORY
   Each GPU board with 8 NVIDIA Tesla V100.
3. 12 TOTAL NVSWITCHES
   High Speed Interconnect, 2.4 TB/sec bisection bandwidth.
4. 8 EDR INFINIBAND/100 GbE ETHERNET
   1600 Gb/sec Bi-directional Bandwidth and Low-Latency.
5. PCIE SWITCH COMPLEX
6. TWO INTEL XEON PLATINUM CPUS
7. 1.5 TB SYSTEM MEMORY
8. DUAL 10/25 GbE ETHERNET
9. 30 TB NVME SSDS INTERNAL STORAGE

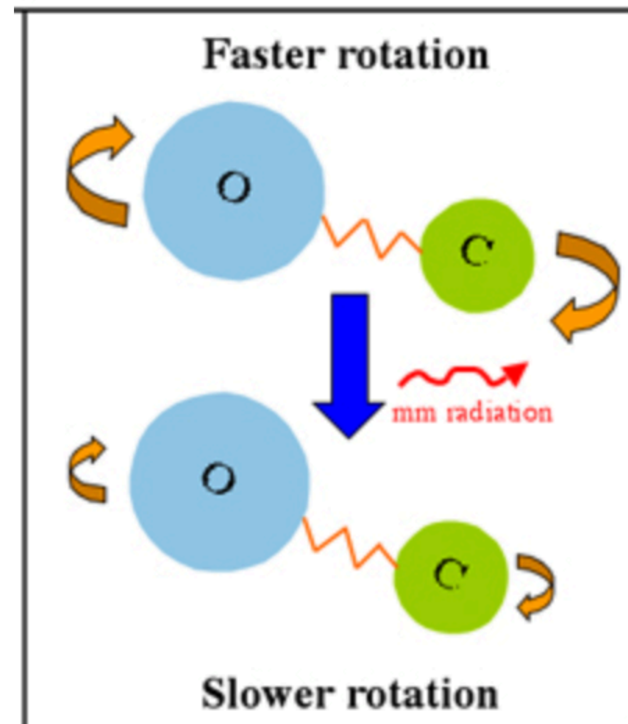


Radio astronomy 101

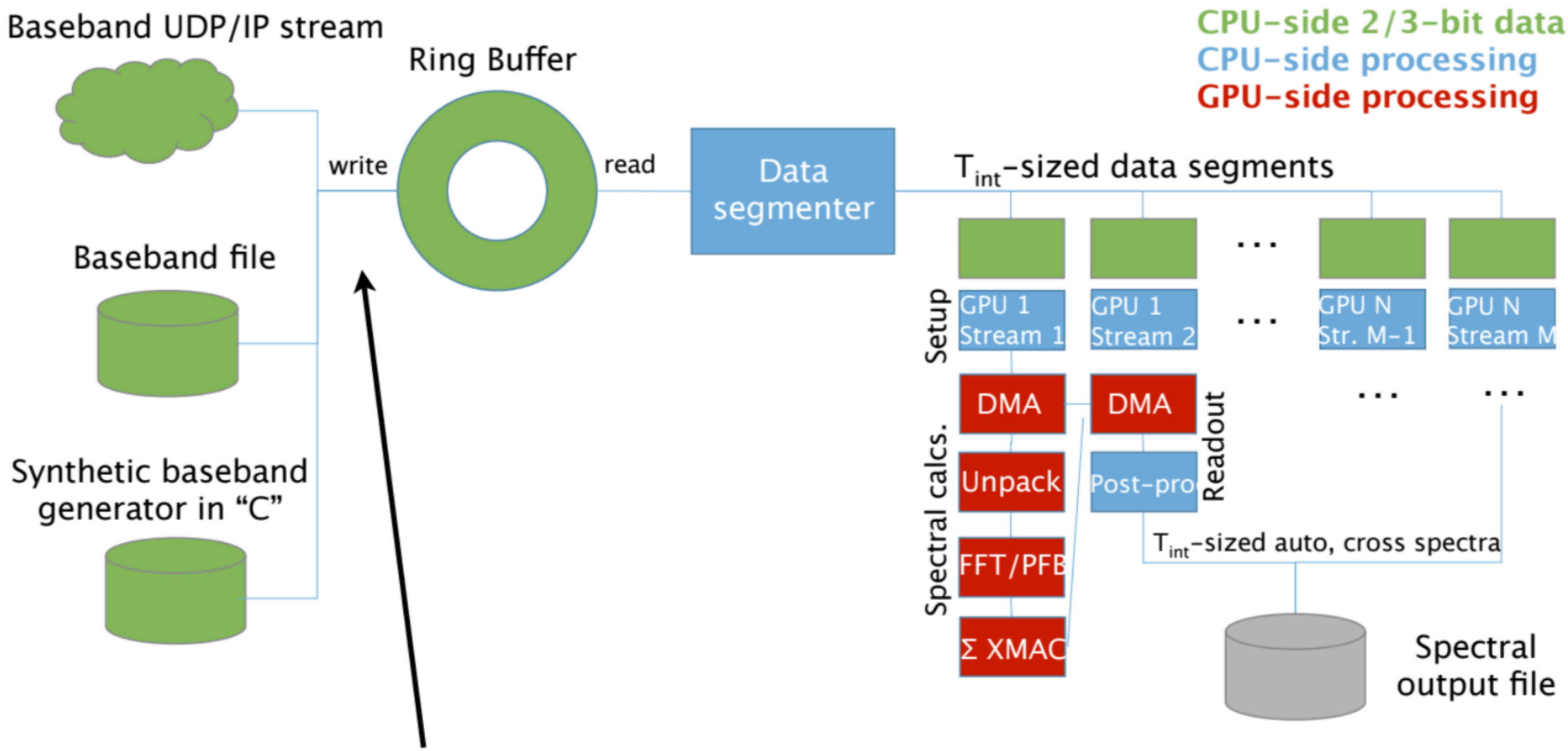


Fourier Transform





Data Flow in Spectrometer

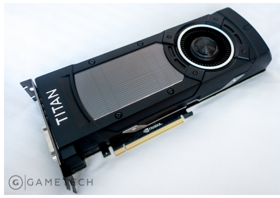


Input data are in VDIF format or in a yet to be specified ACA TP DXRP PCIe format. Sources: network, file, on-the-fly synthetic generator.

Data rates and required performance for FFT

- Total sampling rate of a single ALMA antenna
  - Each baseband: 4 Gs/sec x 2 (polarization) = 8 Gs/s [Giga samples/second]
  - Total sampling rate: 4 (baseband) x 8 Gs/s = 32 Gs/s
  - Input data rate = 32 Gs/s x 3 bit (sampling bit) = 96 Gbit/s
- Needed performance in units of FLOPS
  - 1M (2^{20}) point FFT
  - Number of floating point operations of 1M point FFT: $5\ N \log_2(N) = 5 * 10^6 * 20$ = 0.1 G floating point operations
  - Each baseband: $8 * 10^3$ * 1 M-point FFT /sec = 8k * 0.1 GFLOPS = 0.8 TFLOPS
  - Four basebands: 3.2 TFLOPS

GeForce Titan X



- Maxwell architecture
- CUDA cores: 3072
- base and boost clocks: 1000, 1075MHz
- performance: 6.14~6.6 Tflops single precision
- memory Bandwidth: 336.5 GB/sec
- memory: 12GB

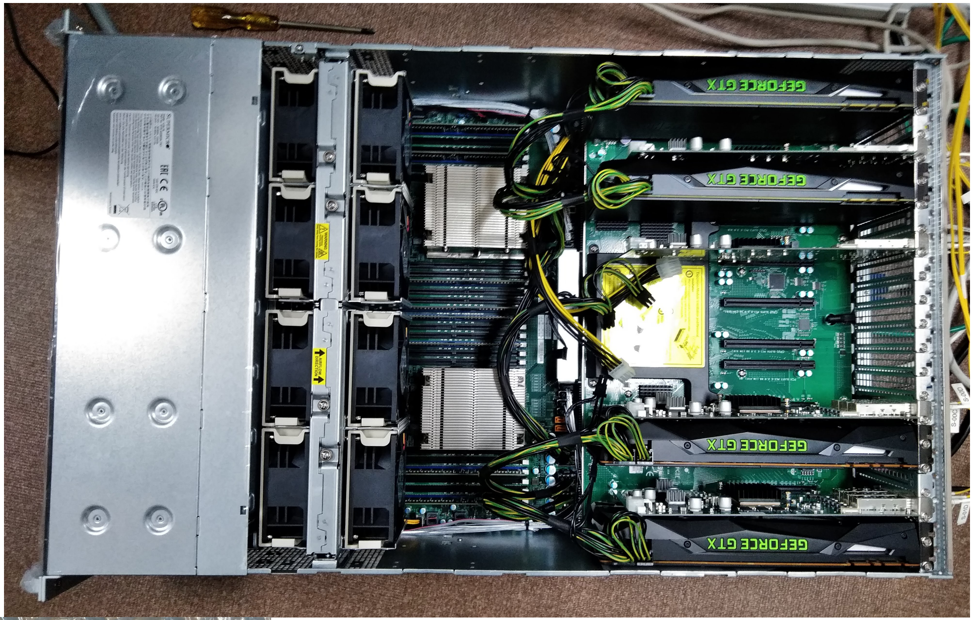
GeForce GTX 980

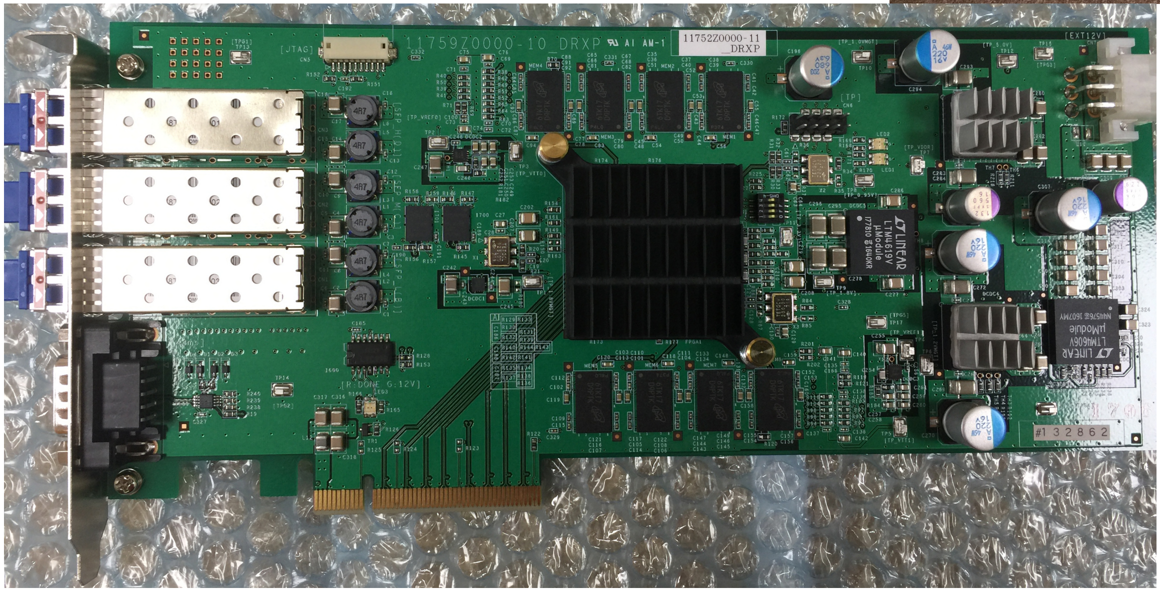
- Maxwell architecture
- CUDA cores: 2048
- base clock: 1064 MHz
- performance: 4.36 Tflops single precision
- memory Bandwidth: 224 GB/sec
- memory: 4 GB

Tesla K40m

- Kepler architecture
- cores: 2880
- base, boost clocks: 745 MHz, 810 MHz and 875 MHz
- performance: 4.29~5.04 Tflops single precision
- memory bandwidth: 288 GB/sec
- memory: 12 GB

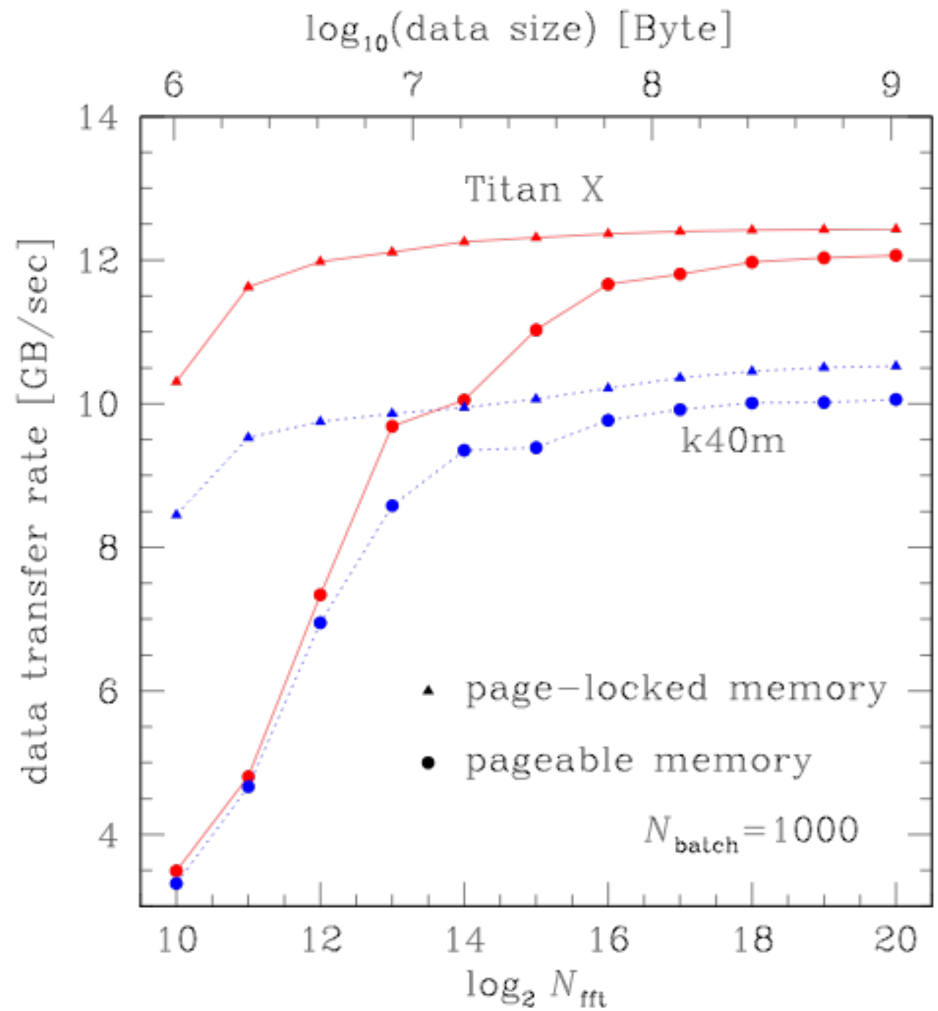
Prototype GPU spectrometer





data copy from CPU to GPU

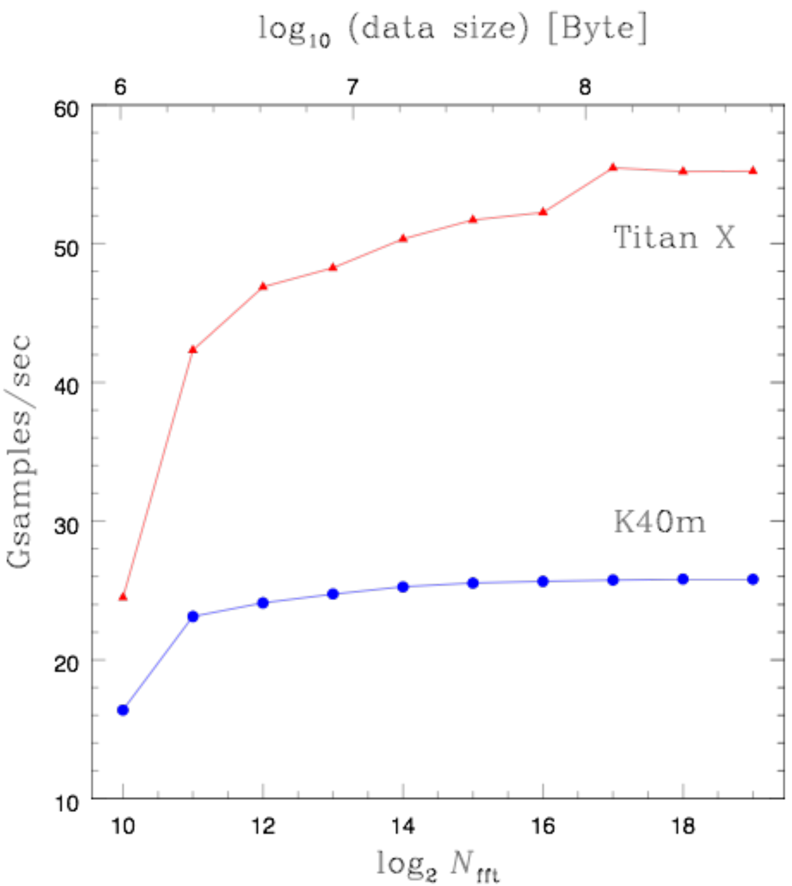
- converges to 12.5 GB/sec < 16GB/sec (PCIE 3)
- 2 bits/sample: 12GB/sec —> 48 Gsamples/sec (24 GHz bandwidth)
- 3 bits/sample: 12GB/sec —> 32 Gsamples/sec (16 GHz bandwidth)
- 4 bits/sample: 12GB/sec —> 24 Gsamples/sec (8 GHz bandwidth)



1024 16384 1048576

type conversion (2 or 3 bits to 32bits)

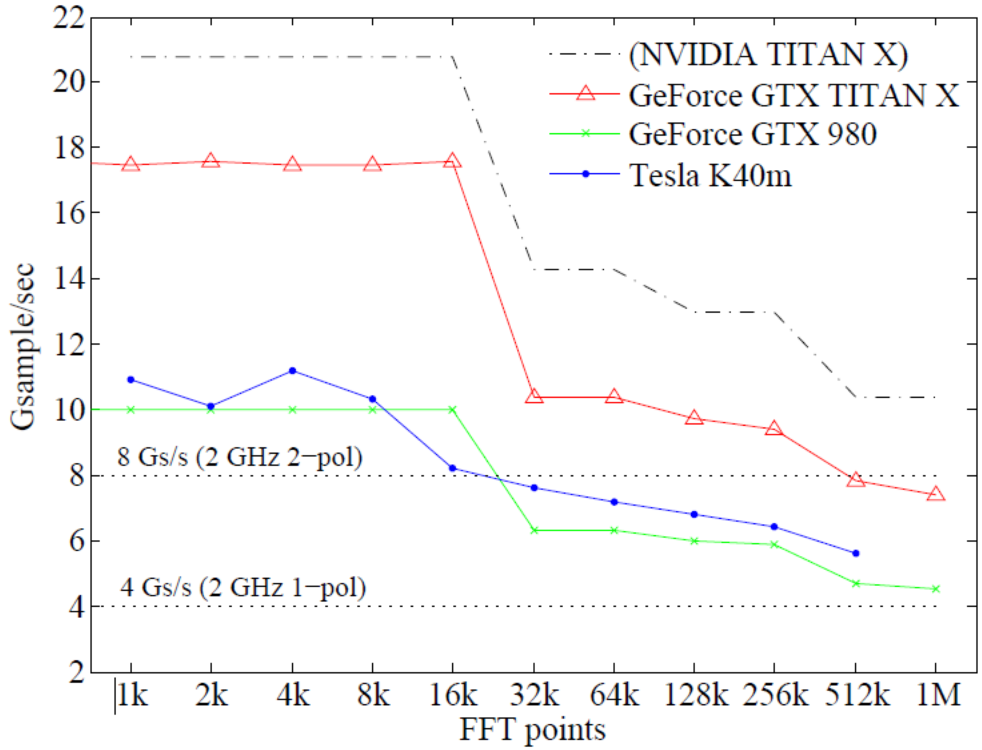
- lookup table for a 2 bit
  - {-3.3359, -1.0, +1.0, +3.3359}
- lookup table for a 3 bit sample
  - {-7.0f, -5.0f,-1.0f, -3.0f, +7.0f, +5.0f, +1.0f, +3.0f}
- ~ 50 Gsamples/s for 2bit and 3bit samples with GTX Titan X



cuFFT

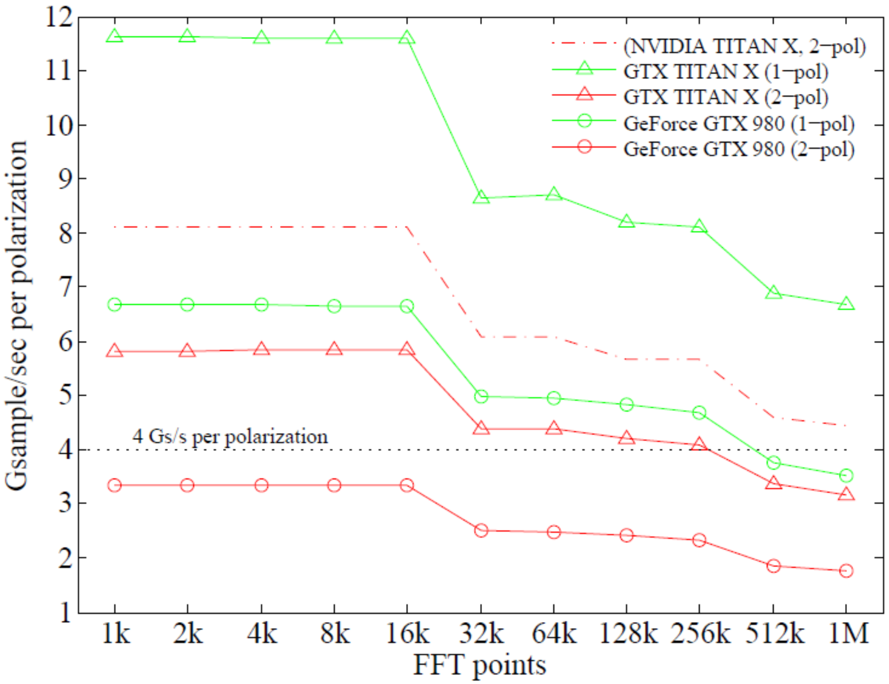
```
cufftPlanMany(….);
cufftExecR2C(plan, idata, odata)
```

- total number of samples: 250*2^20
- 7.5~17.5 Gsample/sec (3.8~8.8 GHz in single polarization)
- ACA correlation used 2^20 point FFT
- for a given number of samples, FFT performance is higher at small fft points



Performance Results

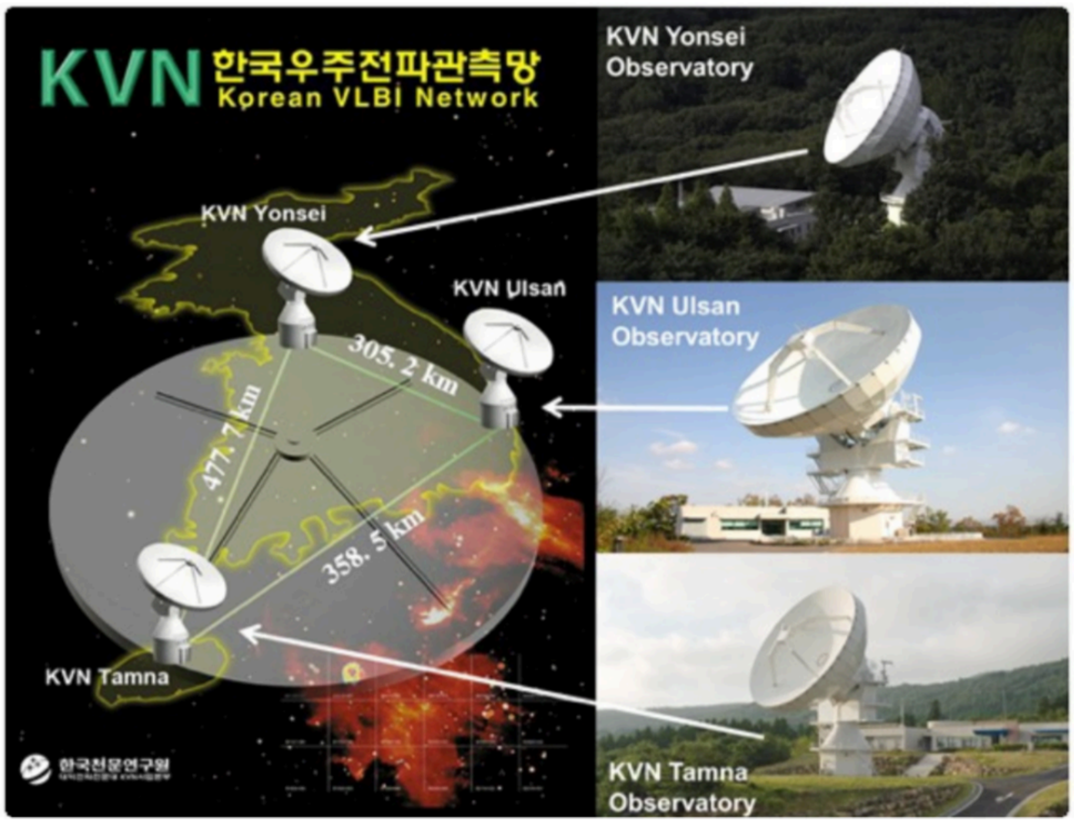
- Need > 4 Gsamples/sec per polarization
- How many GPUs per TP antenna
  - 1-pol: one GTX TITAN X (maxwell architecture) per baseband, in total 4 GPUs per antenna
  - 2-pol: ~1.3 GTX TITAN X (maxwell) per baseband, in total 6 GPUs per antenna
- The new GPU card (TITAN X Pascal architecture) could digest one baseband (2-pols) of a TP antenna.

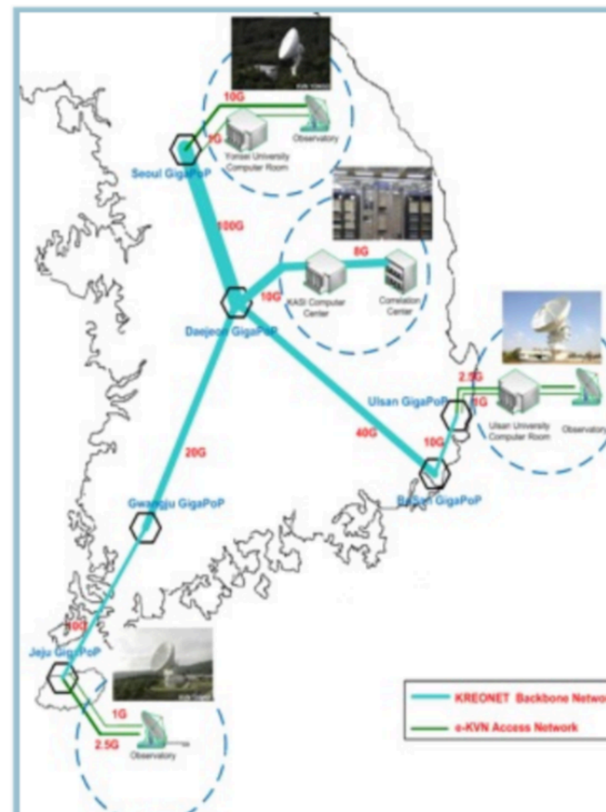


Preliminary On-Sky Test (15 March)

KVN Yonsei 21m telescope (Seoul) → **KVN correl. (Daejeon)**

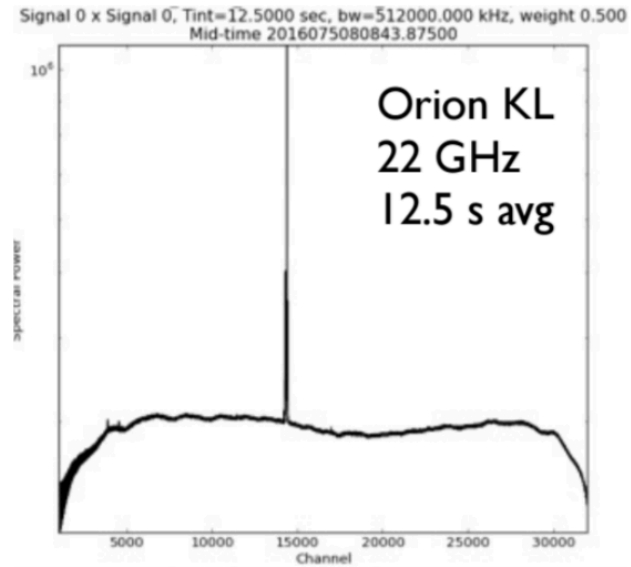
2 x 512 MHz IFs 22/43 GHz RXs → (analog) → ADS-1000 → (VSI-H digital) → FILA10G → (KREONET digital) → Dell T630 with 2 GPU

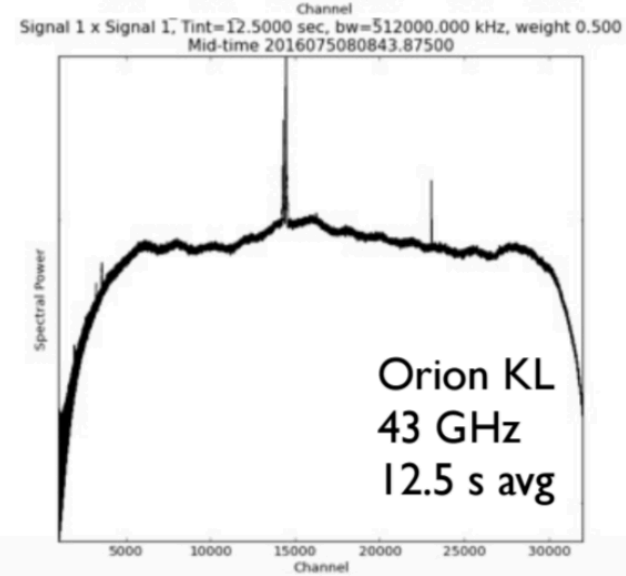




Preliminary On-Sky Test (15 March)

- KVN Yonsei, 2h, 22/43 GHz, 4 Gbps
- Currently "10G" link from Yonsei to Daejeon is 10x1G
  - Averages ~6 Gbps for parallel TCP
  - Very high loss (60%) for 4 Gbps UDP
- To be repeated! With real 10G, or after GPU spectrometer at KVN station!
- Similar tests late 2016 when ACA DXRP PCIe boards in hand!

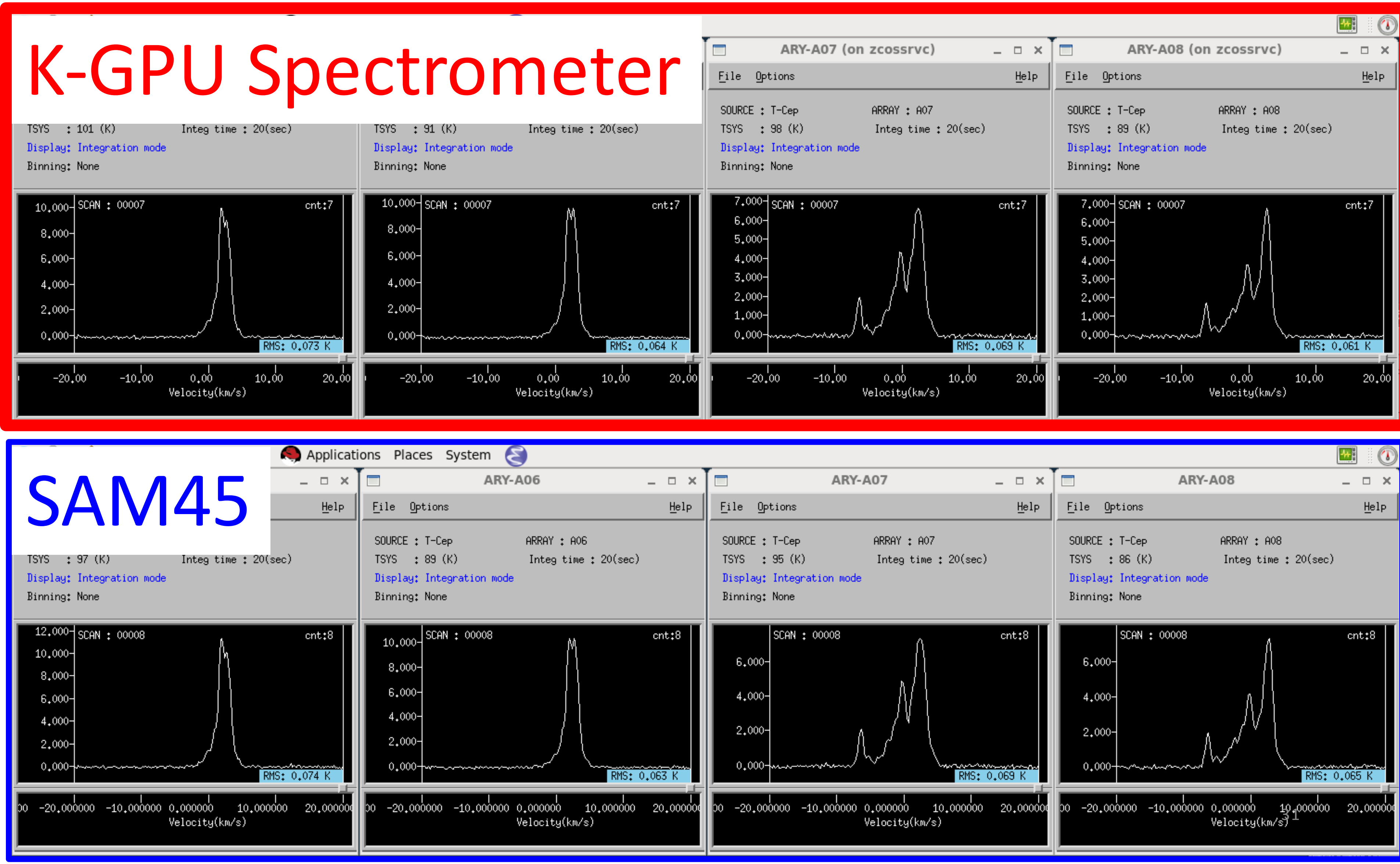




First Spectra with 45-m Telescope (Dec. 25, 2017)

SiO (v=2, J=1-0) @ 42.8GHz

SiO (v=1, J=1-0) @ 43.1GHz



Correlation Theorem
FX-correlator or FX-spectrometer

$$R_i(f) = \int_{-\infty}^{+\infty} r_i(t) e^{2\pi i f t} dt$$

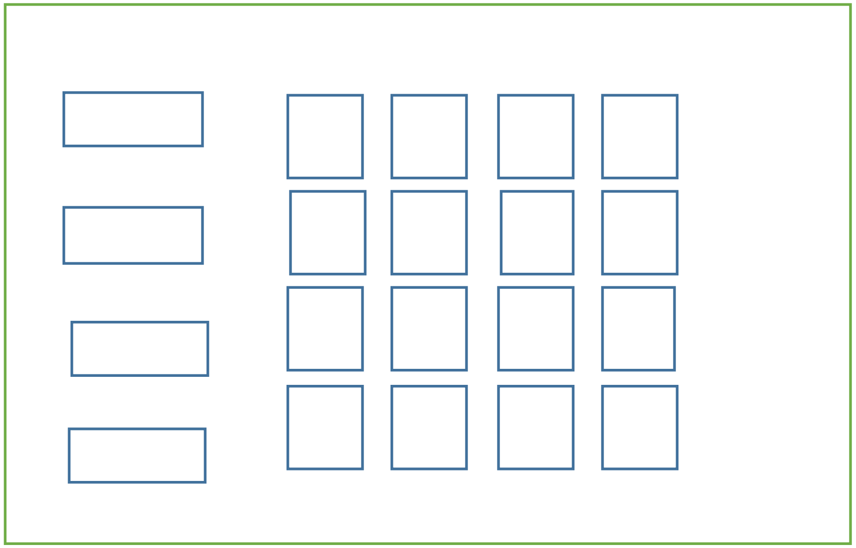
F-step (FT):
~$\log_2(N_c)$ operations per sample

$$\int_{-\infty}^{+\infty} r_i(\tau + t) r_j(\tau) d\tau \Leftrightarrow R_i(f) R_j^*(f)$$

X-step (CMAC):
~ N operations per sample

Conceptual Design of GPU correlator for the ACA array (16 antennas)

- 2GHz*2(Nyquest)*3(sampling bit) *2(dual pol)=24 Gb/s per antenna
- total input data rate: 24 * 16 = 384 Gb/s
- Need 4 100 GE ports



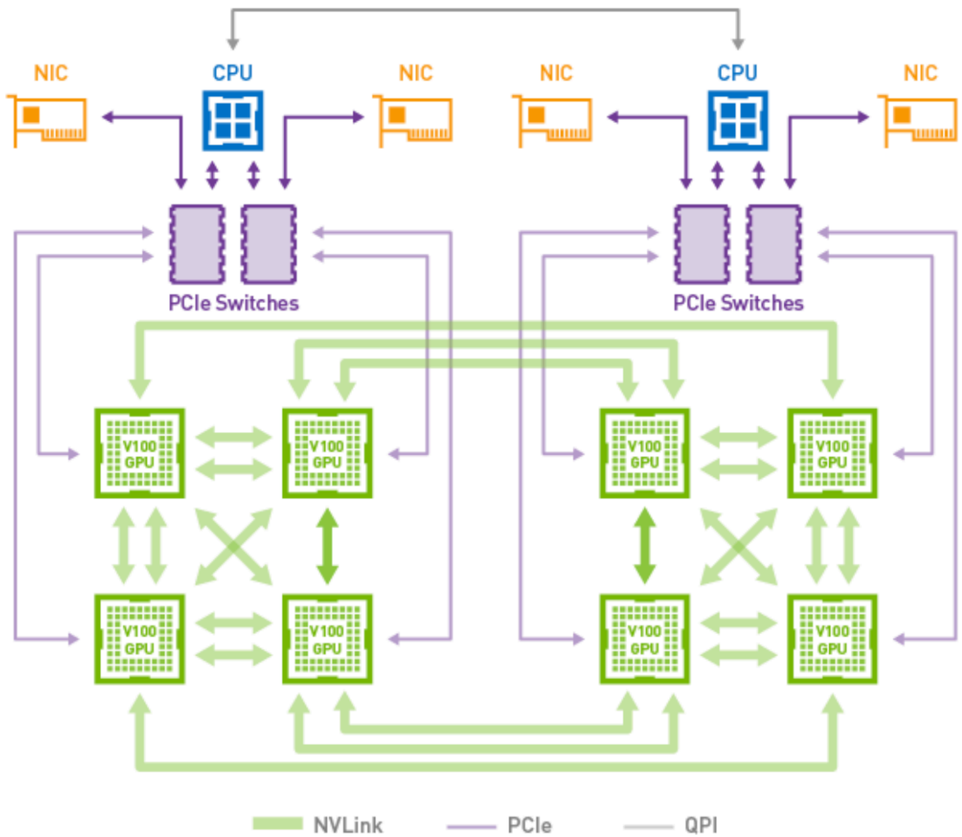
4 network adapters (8) 16 GPUs

Performance requirements for one server

- F-step
  - Assuming a server gets one BB (dual pol) from 16 antennas
  - $N=2^{20}$, dual pol, 16 antenna
  - Flops= 4e9/N*5*N*$\log_2$(N) * 2 (dual) * 16 (antenna) = 12.8 Tflops
- X-step
  - Assuming that a server collects fft data from 16 antennas
  - cross-correlations: 120 x 8 operations (complex multiplications) x 2e9 = 1.92 Tflops
  - # of auto-correlations: 16 x 4 operations x 2e9 = 0.128 Tflops
- Theoretical performance of one NVIDIA Tesla V100
  - Single precision (32bit): 15 Tflops x 8 = 120 Tflops

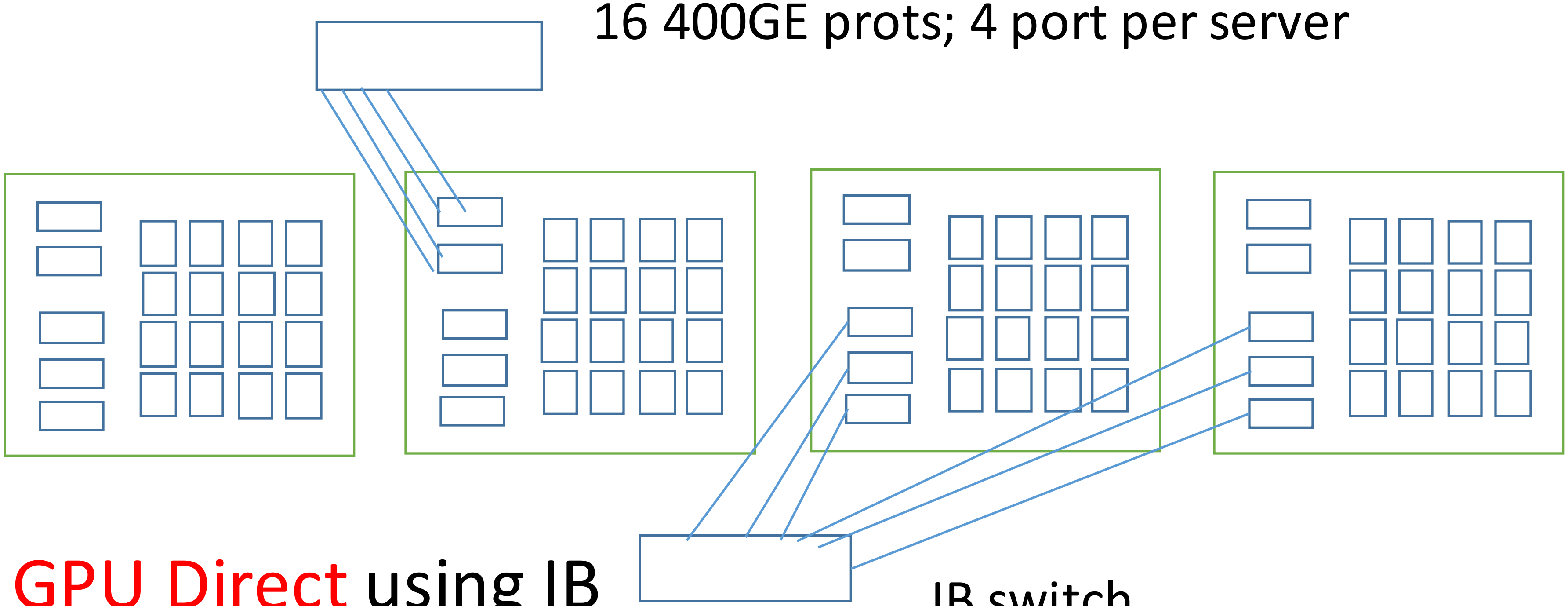
Conner turning

- amount input data per pol: 4 Gsamples/s
- Amount of data after fft per pol: 2 Gsamples/s*8 B = 16 GB/s
- Amount of data for both pols with 4 antenna per GPU: 128 GB/s
- Amount of data to send to nearby a GPU: 32 GB/s
- NVLINK with 25GB/s (one way) from one GPU to the other could be a promising for the conner turning.



ALMA correlator (for 64 antenna) based on the GPU technology

Ethernet switch
64 100GE ports from antenna
16 400GE prots; 4 port per server



GPU Direct using IB
For conner turning

IB switch
48 ports; 12 per baseband

Conclusions

- A GPU spectrometer (one server with four GPU cards) could process data streams of 32Gsamples/s from one ALMA antenna in real time.
  - The most time-consuming part is FFT, but cuFFT (in CUDA FFT library) is fast enough for our spectrometer
- A concept design of a GPU correlator for ACA array (16 antenna) is possible
  - It is composed of four servers. Each server is composed of 4 100 GEs and 8 GPU cards.
  - NVLINK is a good solution for transferring data among GPUs
- A concept design of a GPU correlator for 64 antennas is possible using future 400 GE and IB
  - GPU-direct and NVLINK are good solutions for transfering data among GPUs