Heterogenous Real-time Computing and some other stuff

John Ford

GUPPI Team: Patrick Brandt, Ron Duplain, Paul Demorest, Randy McCullough, Scott Ransom, Jason Ray

Outline

Heterogenous Real-time Systems

Computer Architecture

GUPPI

GBT Spectrometer

Heterogeneous Computing Systems

A system made up of architecturally diverse user-programmable computing units.

Typically a mixture of two or more of the following:

- Traditional RISC/CISC CPUs
- Reconfigurable Computing Elements (RCE)
- Graphical Processing Units (GPU)

Real-time Computing

A real-time computing system must operate while meeting fixed deadlines for completion of scheduled computations

- Computations finished correctly but late are not useful?
- Two cases:
  - Soft real-time systems
  - Hard real-time systems

Real-time does not mean "really fast"!

- ***But it is helpful!***

Heterogeneous Real-time Computing

Combines both ideas into one system

- Note that many heterogenous systems are not designed to be real-time systems:
  - SRC-{6,7}, Convey, CPU/GPU clusters, supercomputers
- Building a heterogeneous system is more difficult than building a homogenous system due to the need to master multiple toolsets, computing models, and hardware characteristics.
- Add in the requirement for real-time response, and it is even more challenging.

Then Why Bother?

- The data rates are too high for a general-purpose machine
  - Use an FPGA preprocessor to operate on the raw data streams for I/O management
  - Use a GPU to offload computations from the CPU for CPU load management
- A general purpose machine would demand too much power
  - FPGA's and GPU's are more efficient at computations that fit their abilities

Computer System Architecture

System interconnection heirarchy

- I/O – slowest, but still fast in today's systems
  - Networks (10-40 Gb/sec reasonably attainable)
  - Disk I/O (500 MB/sec streaming over single RAID)
  - CPU-GPU memory copying (8 GB/sec for x16 PCIe gen 2)

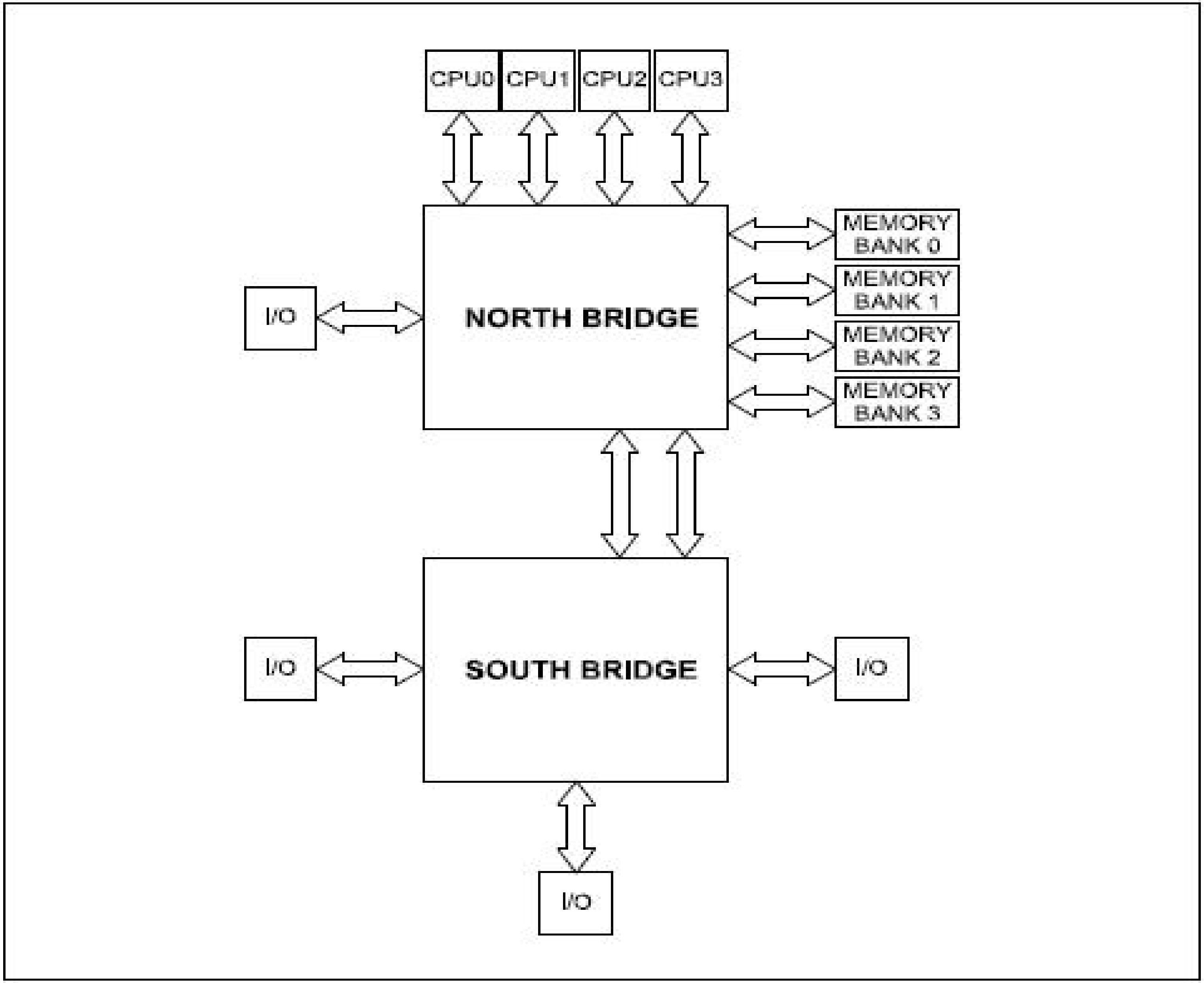
- Memory - Processor
  - Front-side bus
  - HyperTransport (51 GB/sec per link)
  - QuickPath Interconnect (25 GB/sec per link)

Computer System Architecture

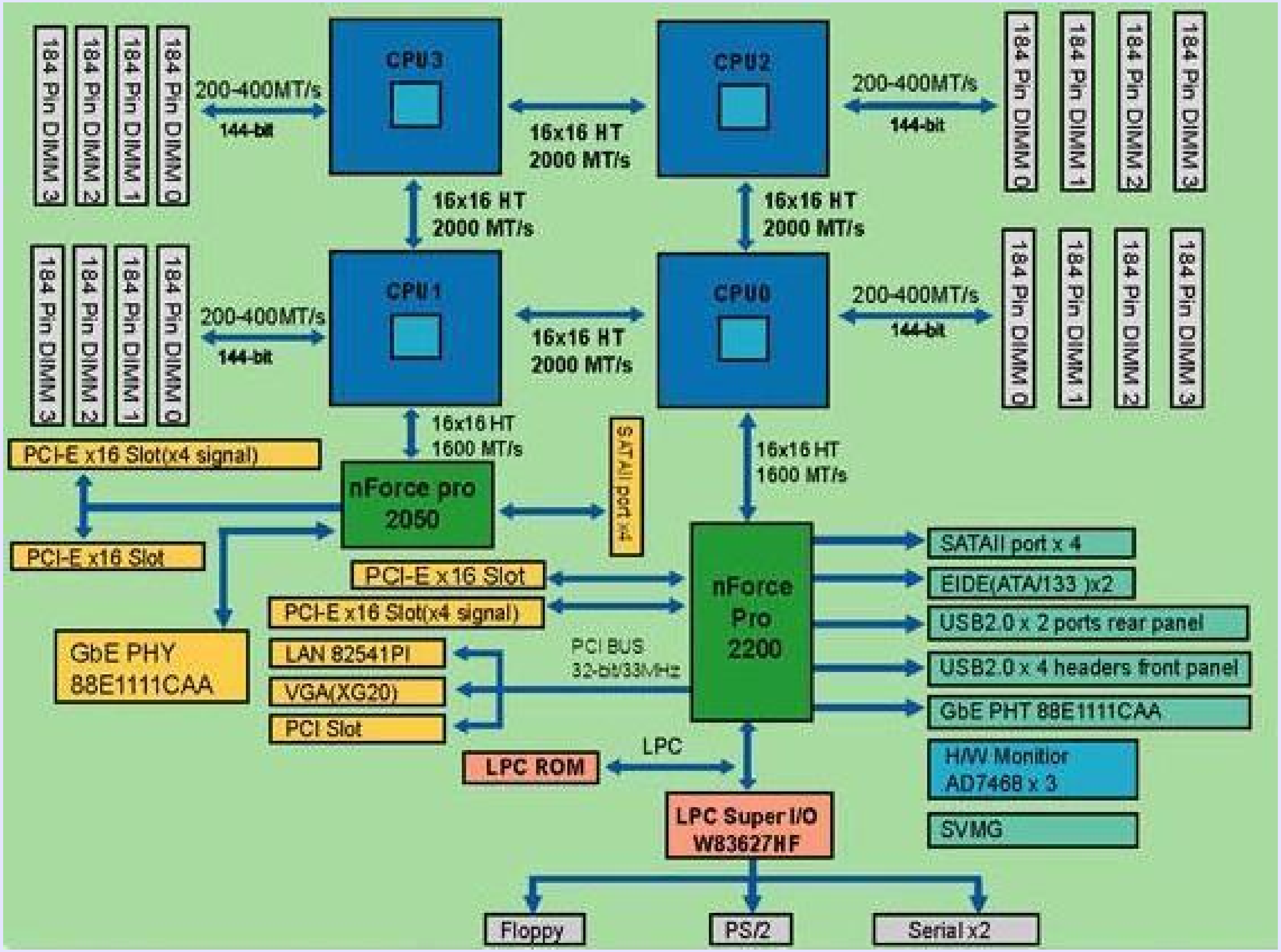
Memory organization

- Uniform Memory Access (UMA)
  - Older Intel Xeon systems
- Non-Uniform Memory Access (NUMA)
  - AMD Opteron systems
  - Newer Intel systems

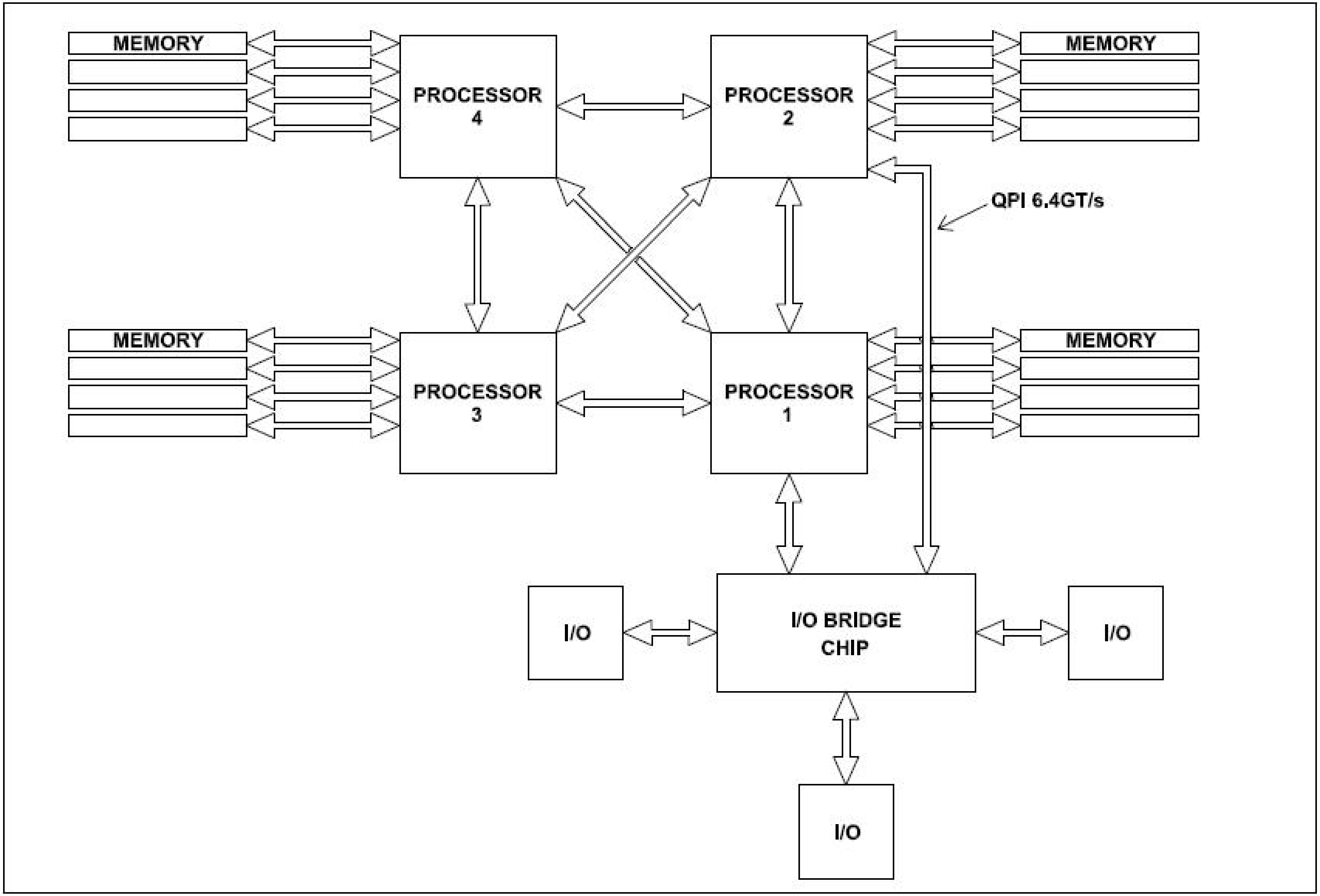
UMA Example


CPU0
CPU1
CPU2
CPU3
MEMORY BANK 0
MEMORY BANK 1
MEMORY BANK 2
MEMORY BANK 3
I/O
NORTH BRIDGE
I/O
SOUTH BRIDGE
I/O
I/O

HyperTransport NUMA Example



QPI NUMA Example

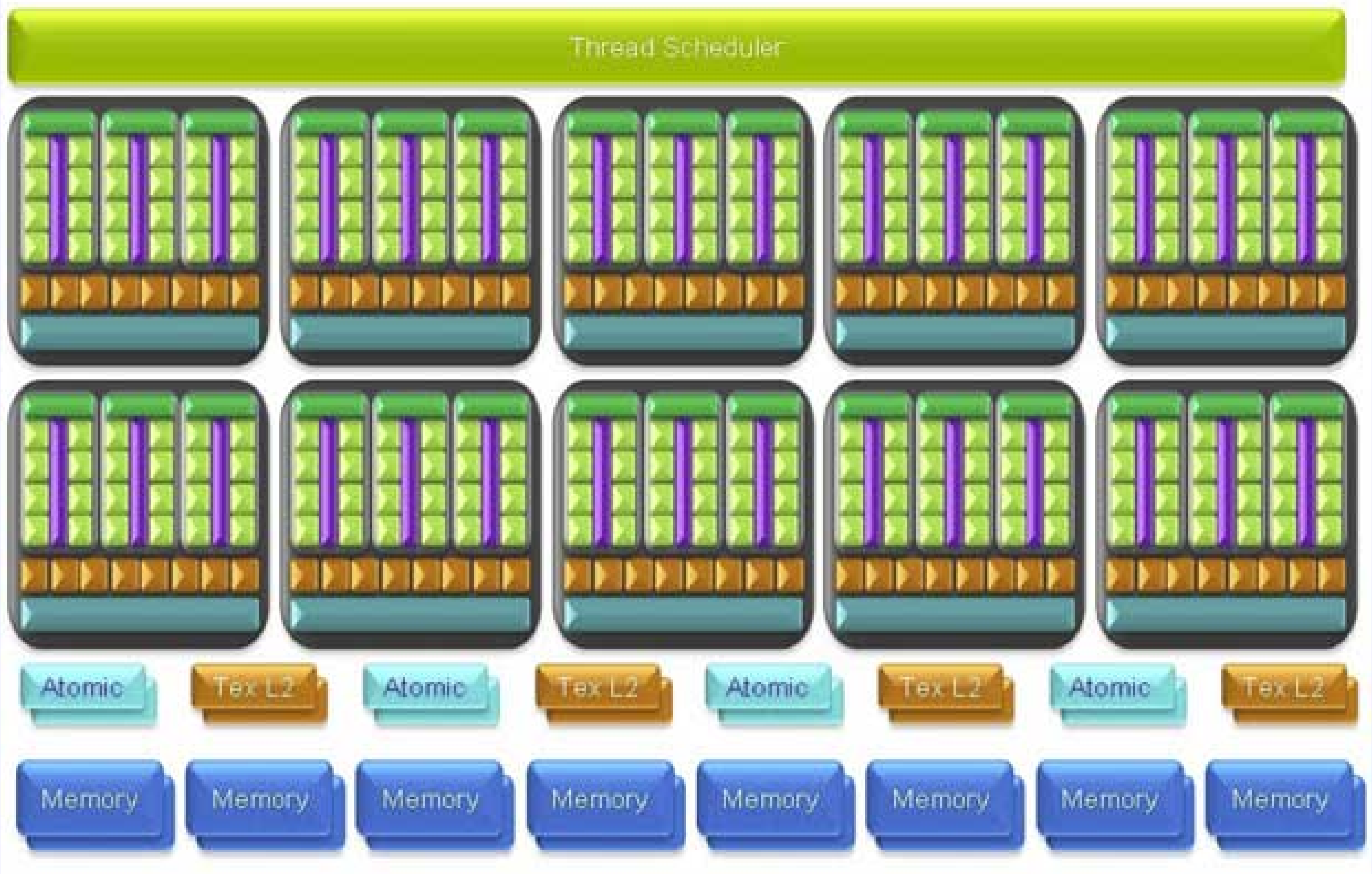


Computer Systems Architecture

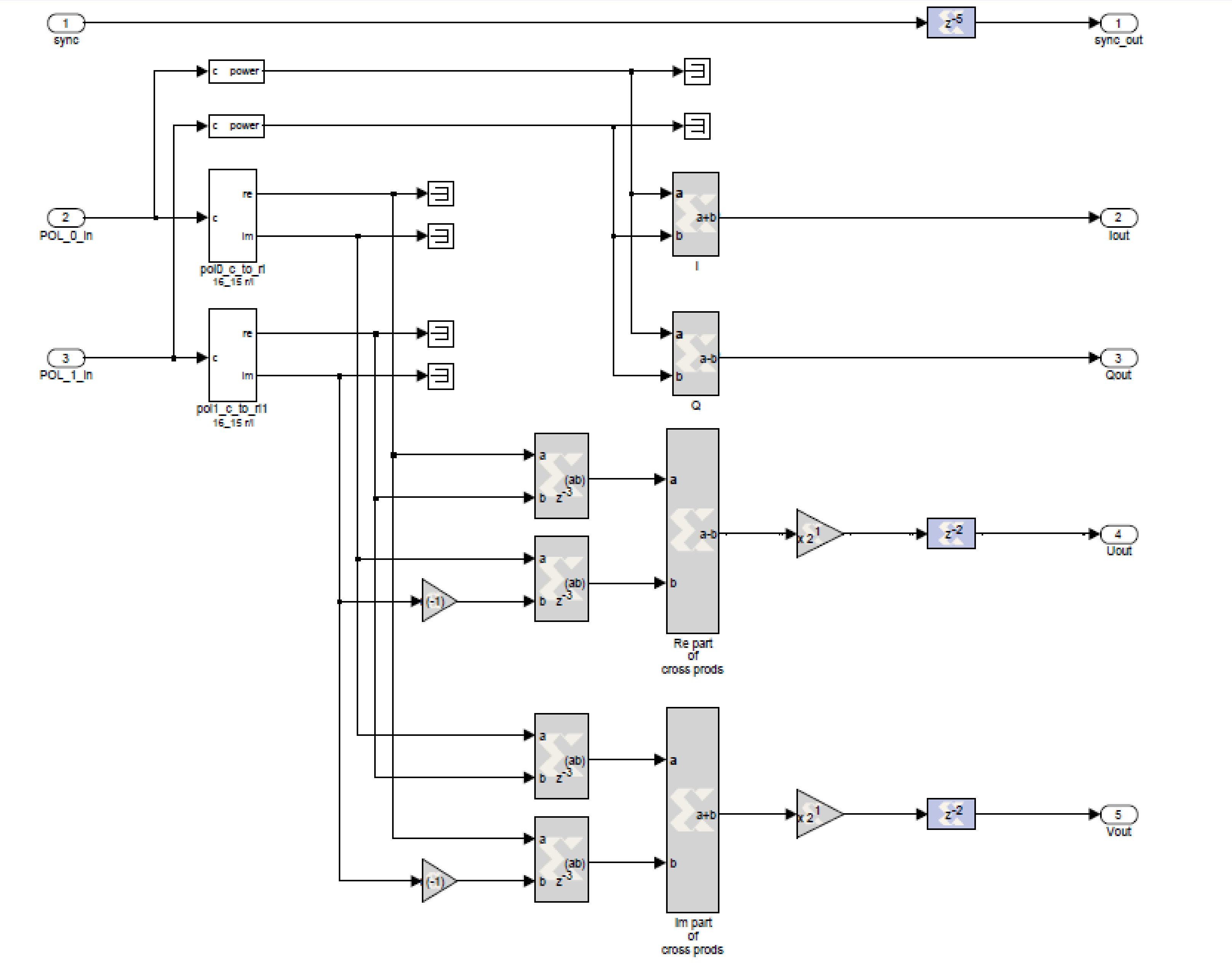
Processor architectures

- CPU
  - X86_64 – Intel and AMD
  - Everything else – IBM Power and Cell Broadband Engine
- GPU
  - NVidia
  - AMD/ATI
- Reconfigurable Computing Elements (FPGA-based processors)

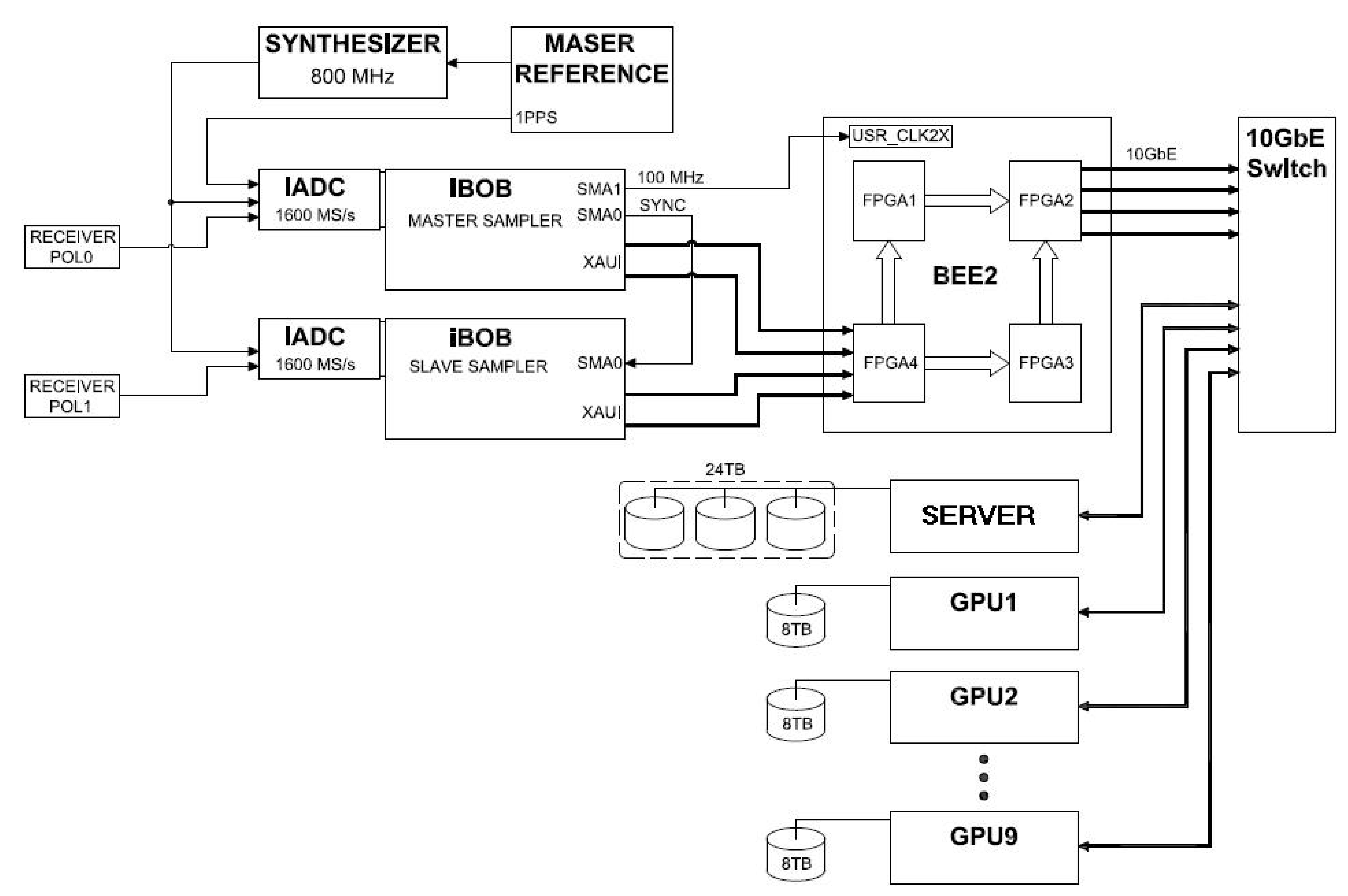
NVidia GPU Architecture



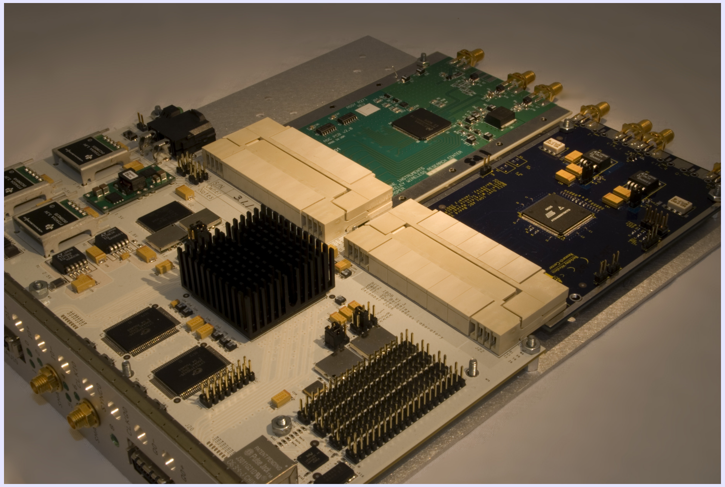
Reconfigurable Computing Elements



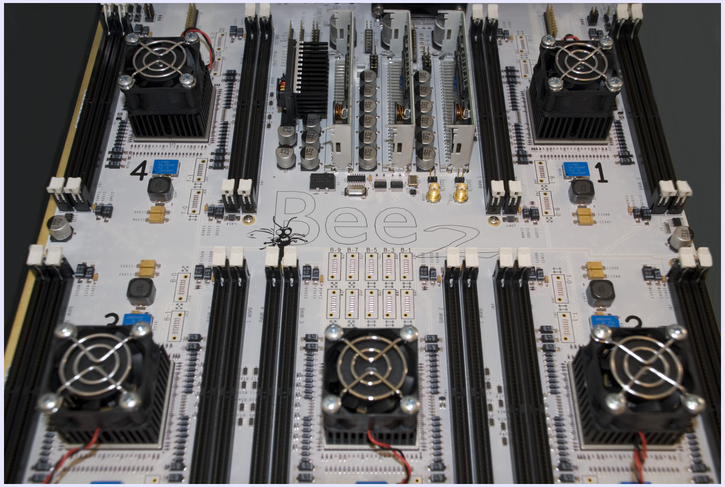
GUPPI Block Diagram



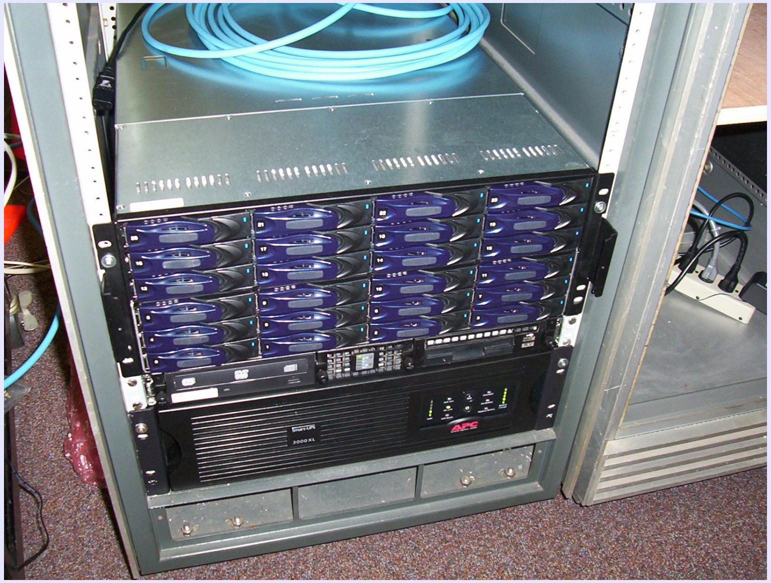
GUPPI architecture:
~1 MHz PFB in FPGAs
Coherent dedisp in GPUs



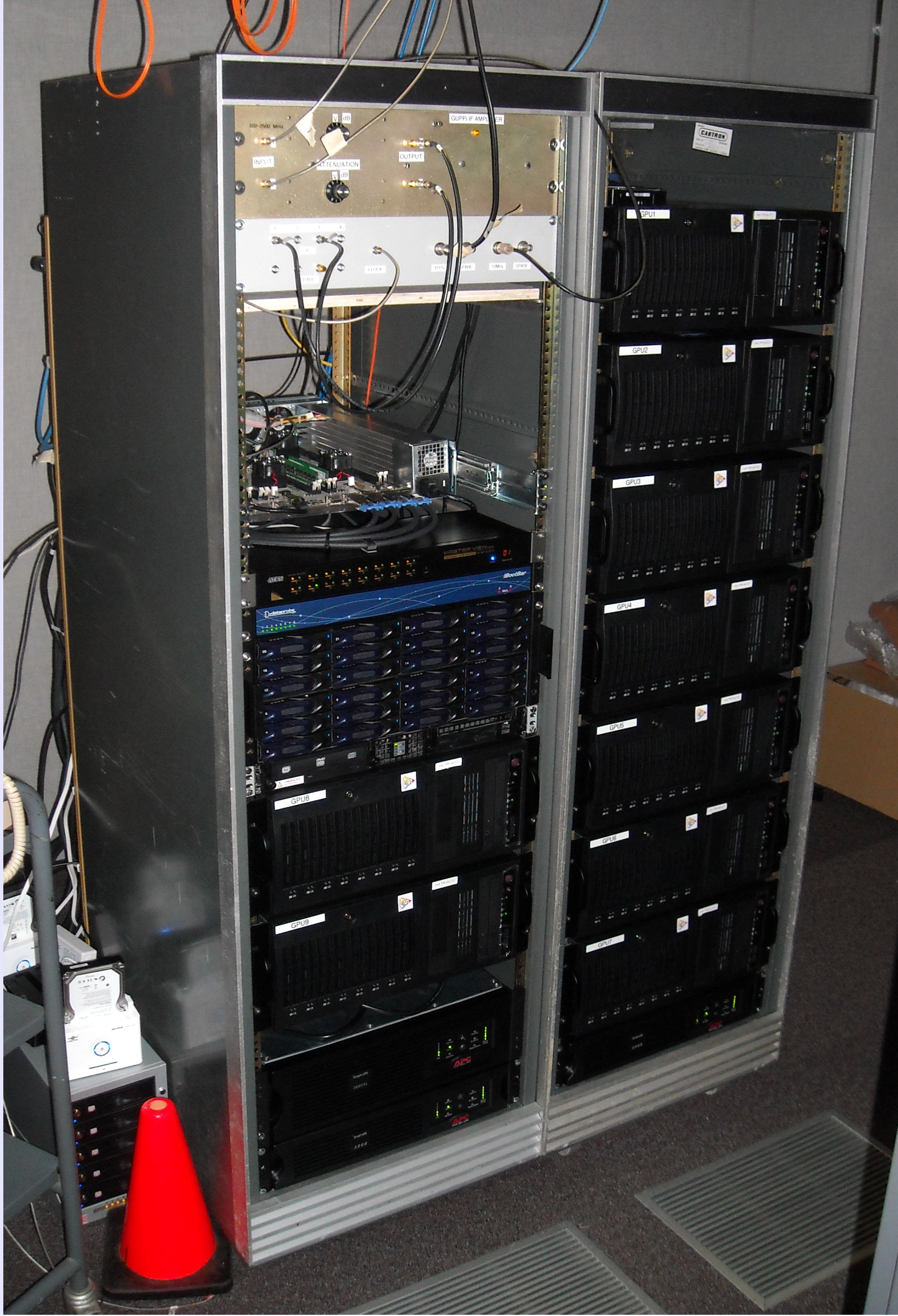
XAUI



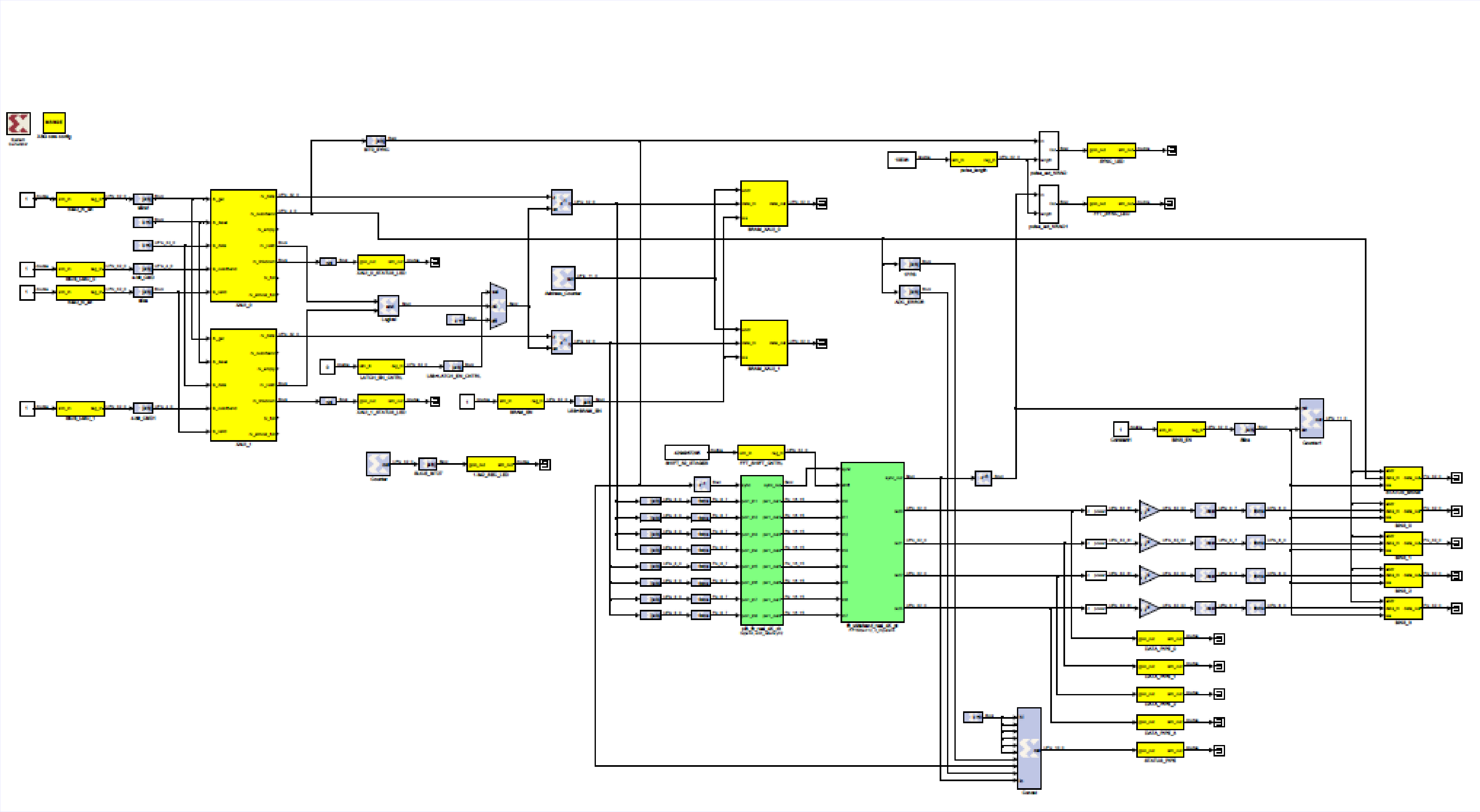
10 Ge
24 Gb/s



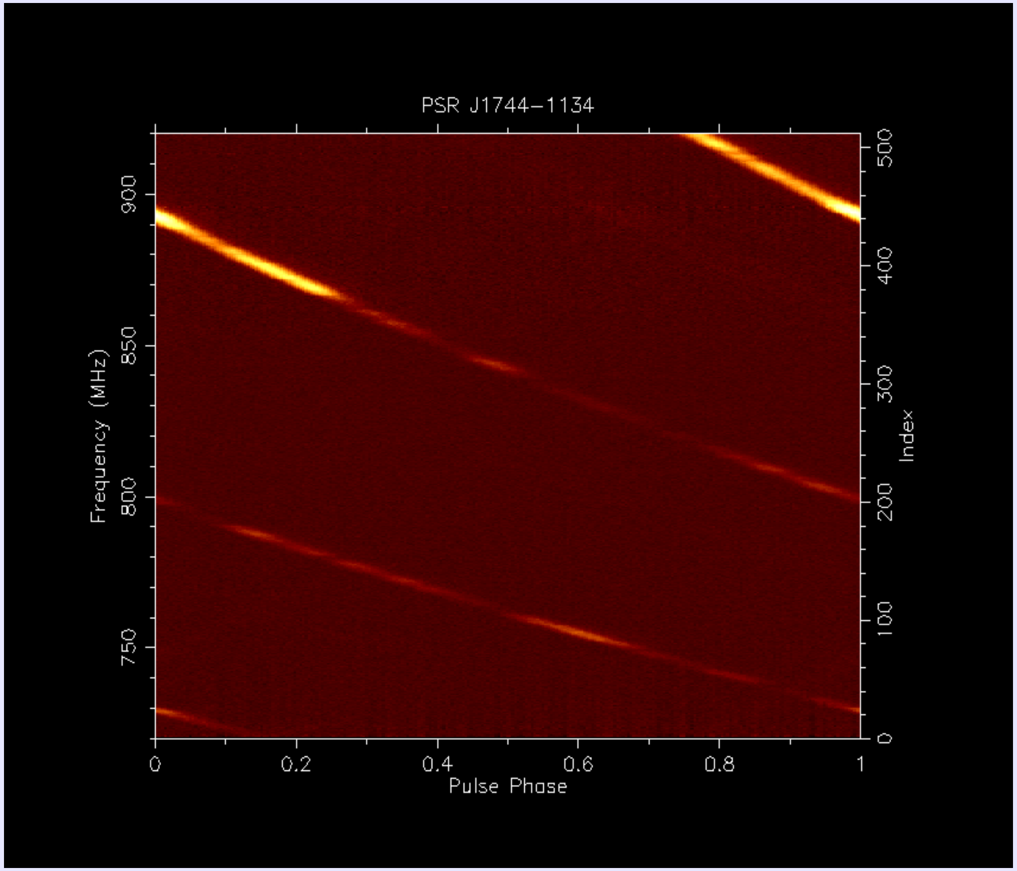
Complete GUPPI

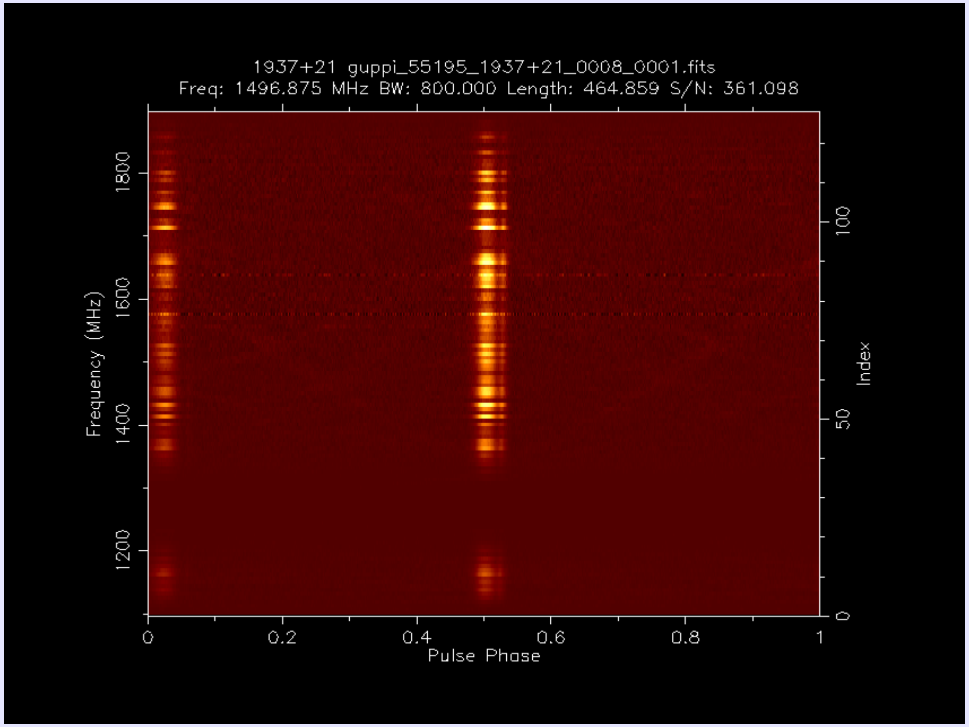


GUPPI Signal Processing



Dispersion


PSR J1744−1134
Frequency (MHz)
900
850
800
750
Index
500
400
300
200
100
0
0
0.2
0.4
0.6
0.8
1
Pulse Phase

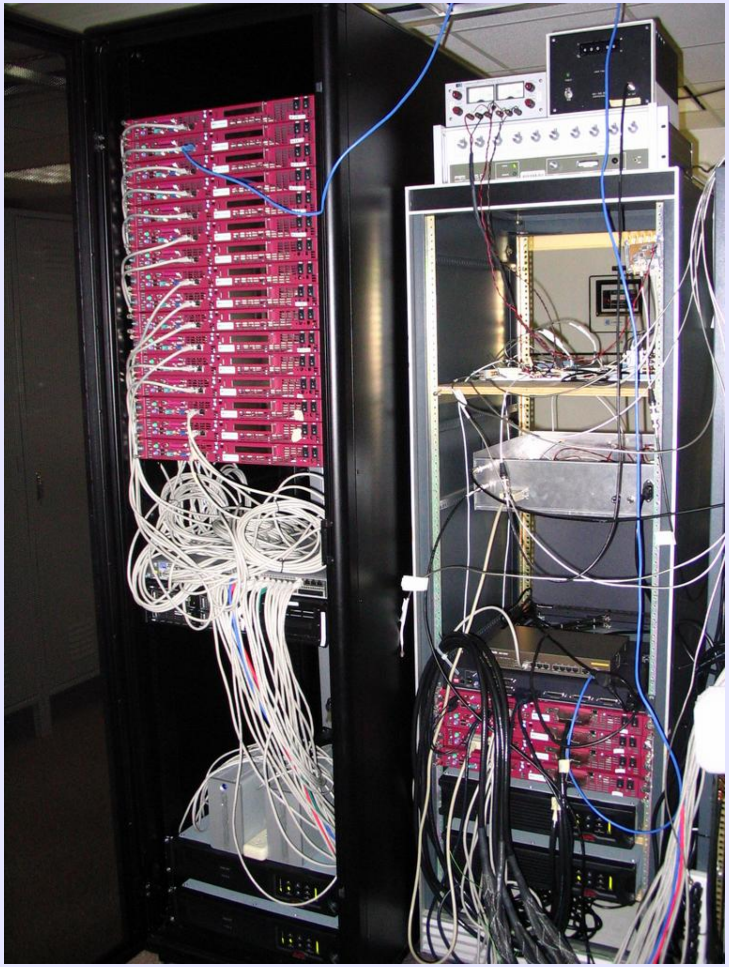

1937+21 guppi_55195_1937+21_0008_0001.fits
Freq: 1496.875 MHz BW: 800.000 Length: 464.859 S/N: 361.098
Frequency (MHz)
1800
1600
1400
1200
Index
100
50
0
0
0.2
0.4
0.6
0.8
1
Pulse Phase

Coherent Dedispersion

Removes ISM dispersion (for known DM) within a freq channel.

FIR filter applied pre-detection via FFT convolution; filter length ~k to ~M-point.

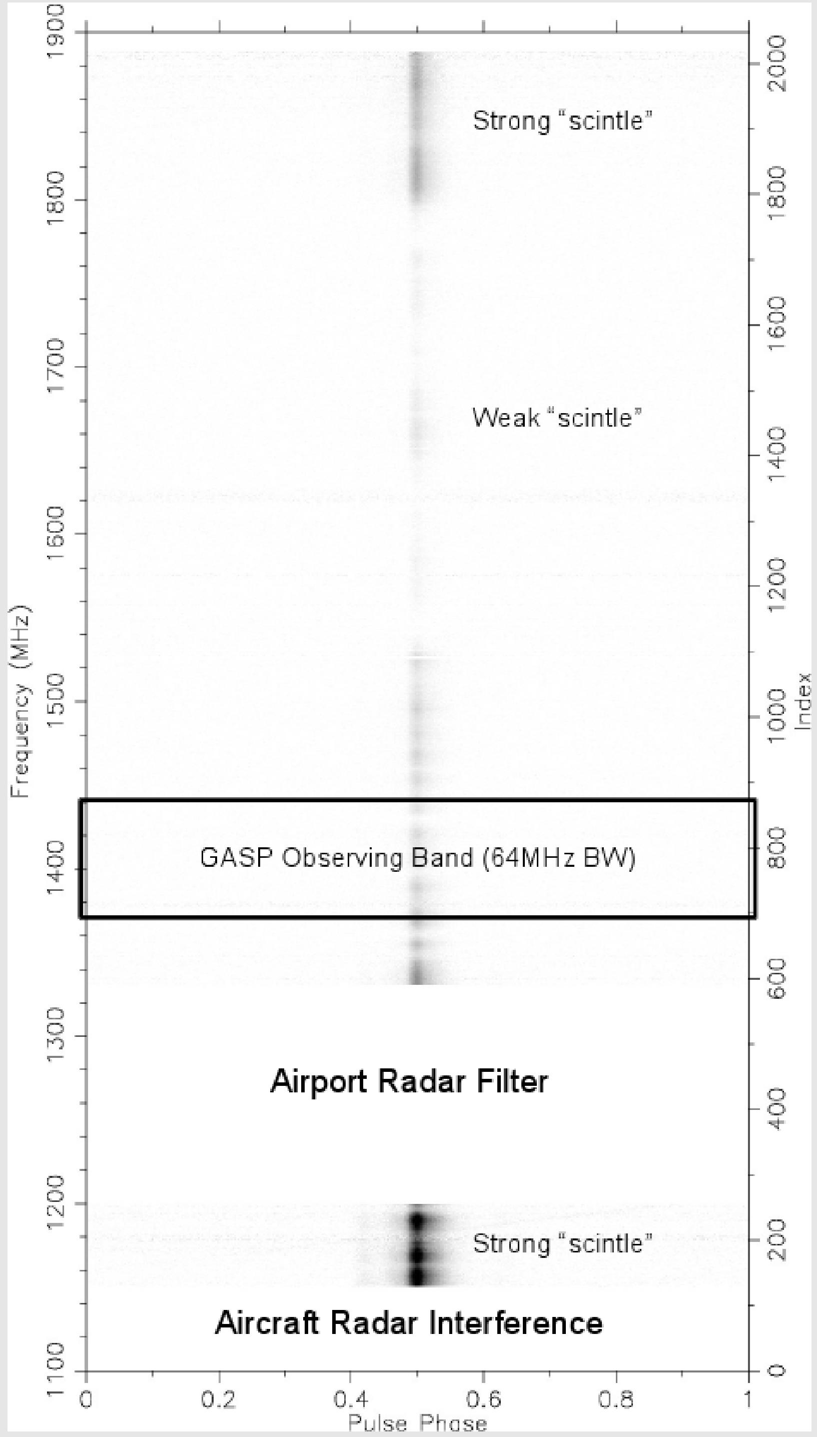
Most current systems are "software" based and handle ~100 MHz total BW.



GASP system @ GB

>10x improvement in BW!

Fully utilizes all GBT low-freq receivers.



PSR J1713+0747
plot: S. Ransom

New GBT Spectrometer

Son of GUPPI

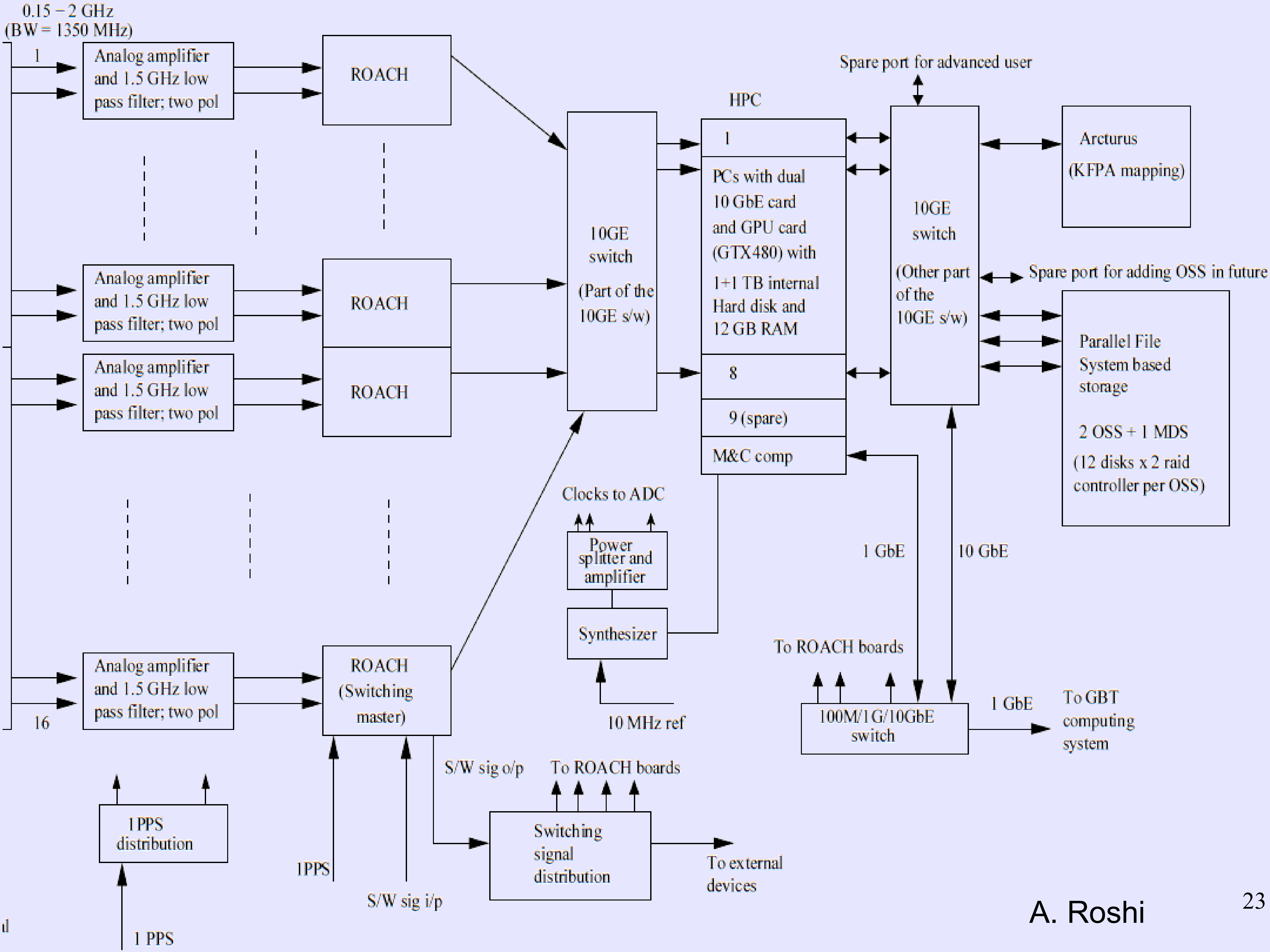
Joint effort with UC Berkeley – NSF ATI project

16 IF inputs

- 8 dual-pol beams, or 16 single-pol beams

3 GS/s sampling rate @ 8 bits/sample

Can be ganged to achieve ~10 GHz instantaneous bandwidth on 2 polarizations


0.15 – 2 GHz
(BW = 1350 MHz)
1
Analog amplifier and 1.5 GHz low pass filter; two pol
ROACH
Analog amplifier and 1.5 GHz low pass filter; two pol
ROACH
Analog amplifier and 1.5 GHz low pass filter; two pol
ROACH
Analog amplifier and 1.5 GHz low pass filter; two pol
16
ROACH (Switching master)
10GE switch (Part of the 10GE s/w)
HPC
1
PCs with dual 10 GbE card and GPU card (GTX480) with 1+1 TB internal Hard disk and 12 GB RAM
8
9 (spare)
M&C comp
Spare port for advanced user
10GE switch (Other part of the 10GE s/w)
Arcturus (KFPA mapping)
Spare port for adding OSS in future
Parallel File System based storage
2 OSS + 1 MDS
(12 disks x 2 raid controller per OSS)
Clocks to ADC
Power splitter and amplifier
Synthesizer
10 MHz ref
1 GbE
10 GbE
To ROACH boards
100M/1G/10GbE switch
1 GbE
To GBT computing system
1PPS distribution
1 PPS
1PPS
S/W sig i/p
S/W sig o/p
To ROACH boards
Switching signal distribution
To external devices

A. Roshi

Spectrometer Data Rates

ADC to FPGA: 3GS/s * 16 Polns * 8 bits = 384 Gb/sec

FPGA to GPU: ~ 9 Gb/sec * 8 GPUs = 72 Gb/s

Raw (mostly) Data Output Rate:

Maximum specified at 33 MB/s * 16 polns = 525 MB/s, or 15 TB/8 hours, or 1 PB every 66 hours.

Front-end (FPGA-GPU) hardware is capable of ~10x this rate…

Summary of Required Observing Modes

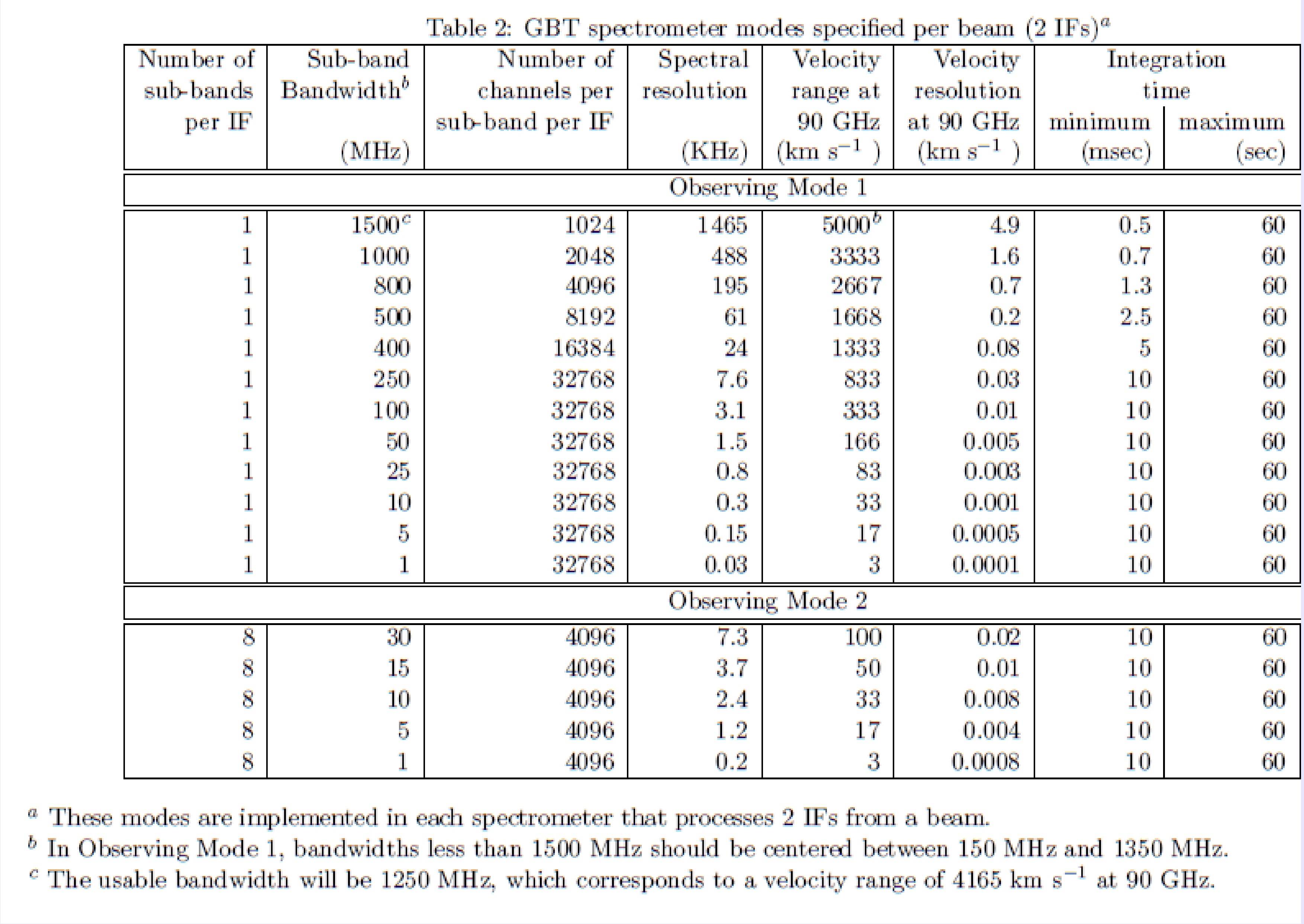


Table 2: GBT spectrometer modes specified per beam (2 IFs)[a]

Number of sub-bands per IF	Sub-band Bandwidth[b] (MHz)	Number of channels per sub-band per IF	Spectral resolution (KHz)	Velocity range at 90 GHz ($\mathrm{km\ s^{-1}}$)	Velocity resolution at 90 GHz ($\mathrm{km\ s^{-1}}$)	Integration time minimum (msec)	Integration time maximum (sec)
Observing Mode 1							
1	1500[c]	1024	1465	5000[b]	4.9	0.5	60
1	1000	2048	488	3333	1.6	0.7	60
1	800	4096	195	2667	0.7	1.3	60
1	500	8192	61	1668	0.2	2.5	60
1	400	16384	24	1333	0.08	5	60
1	250	32768	7.6	833	0.03	10	60
1	100	32768	3.1	333	0.01	10	60
1	50	32768	1.5	166	0.005	10	60
1	25	32768	0.8	83	0.003	10	60
1	10	32768	0.3	33	0.001	10	60
1	5	32768	0.15	17	0.0005	10	60
1	1	32768	0.03	3	0.0001	10	60
Observing Mode 2							
8	30	4096	7.3	100	0.02	10	60
8	15	4096	3.7	50	0.01	10	60
8	10	4096	2.4	33	0.008	10	60
8	5	4096	1.2	17	0.004	10	60
8	1	4096	0.2	3	0.0008	10	60

[a] These modes are implemented in each spectrometer that processes 2 IFs from a beam.

[b] In Observing Mode 1, bandwidths less than 1500 MHz should be centered between 150 MHz and 1350 MHz.

[c] The usable bandwidth will be 1250 MHz, which corresponds to a velocity range of 4165 $\mathrm{km\ s^{-1}}$ at 90 GHz.

Conclusions

Using the heterogenous combination works.

FPGAs for the fastest streaming integer signal processing tasks

GPUs for parallel floating point calculations

Use general-purpose CPUs for managing the system, assembling and transmitting data, and for parts of the computations that are not amenable to the GPU

IDIA Workshop

Workshop Summary

IDIA Workshop

Actions

- Look into creating an engineering REU program funded by CISE