Mark 6 VLBI Data System

Alan Whitney
Roger Cappallo
Chet Ruszczyk
Jason SooHoo
MIT Haystack Observatory

6 May 2013
VLBI Technical Operation Workshop
MIT Haystack Observatory

Why Mark6? - drivers for ever-increasing data rates

-Sensitivity!

-VLBI2010 – enables smaller antennas & shorter scans for better sampling of the atmosphere

-EHT – enables coherent detection of Sgr A* at mm wavelengths (through a fluctuating atmosphere)

Mark 6 goals

- **16Gbps sustained record capability**
  - >=32Gbps burst-mode capability
- **Support all common VLBI formats**
  - possibly general ethernet packet recorder
- **COTS hardware**
  - relatively inexpensive
  - upgradeable to follow Moore's Law progress
- **100% open-source software**
  - Linux O/S

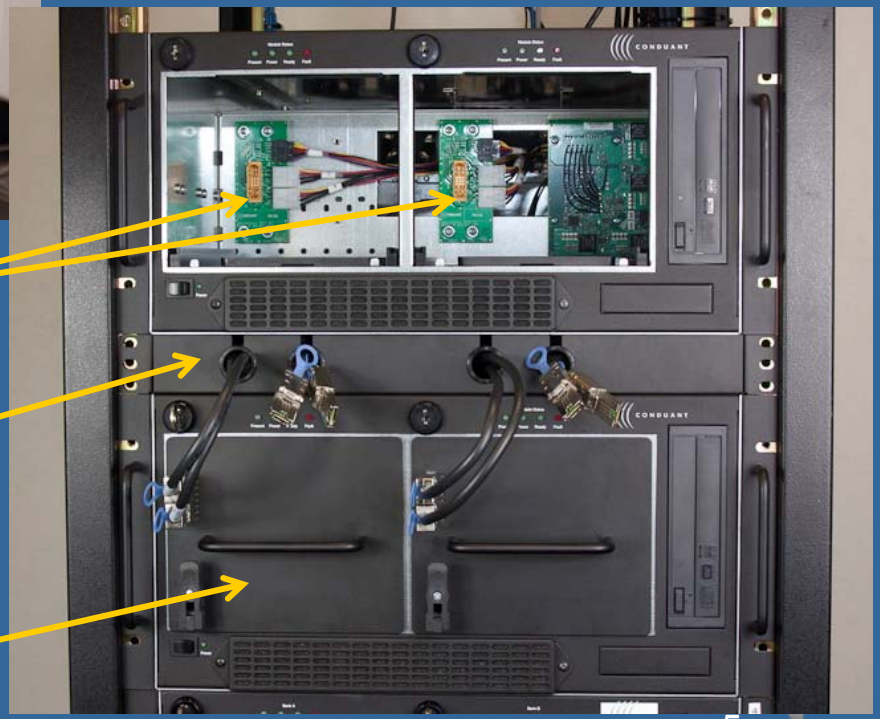
- Other considerations
  - playback as standard Linux files
  - e-VLBI support
  - smooth user transition from Mark 5
  - preserve Mk5 hardware investments, where possible

Mark 6 Mug Shot

Prototype Mark 6 hardware

High-speed data connections to module front-panel via two standard SAS cables

Existing Mark 5 chassis is upgradeable to Mark 6



New chassis backplanes for disk power management

Cable-management panel (unused cables retract into panel)

Existing Mark 5 SATA disk modules are upgradeable to Mark 6 (new backplane and front panel)

Timeline

-Dec 2011: v.0 prototype achieved 16 Gb/s to four 8 disk RAID arrays

-July 2012: v.1 operational dataplane code using RAID array

-Sept 2012: integration of v.1 control and data plane codes

-Sept 2012: v.2 dataplane code with scattered filesystem

-Oct 2012 - now: performance testing, assessment, & tuning

-Feb 2013: bistatic radar observations of asteroid DA14

recorded at Westford

continuous 8 Gb/s on 2 modules

-Mar 2013: VLBI2010 stand-alone testing

-Apr 2013: start of operational VLBI2010 use

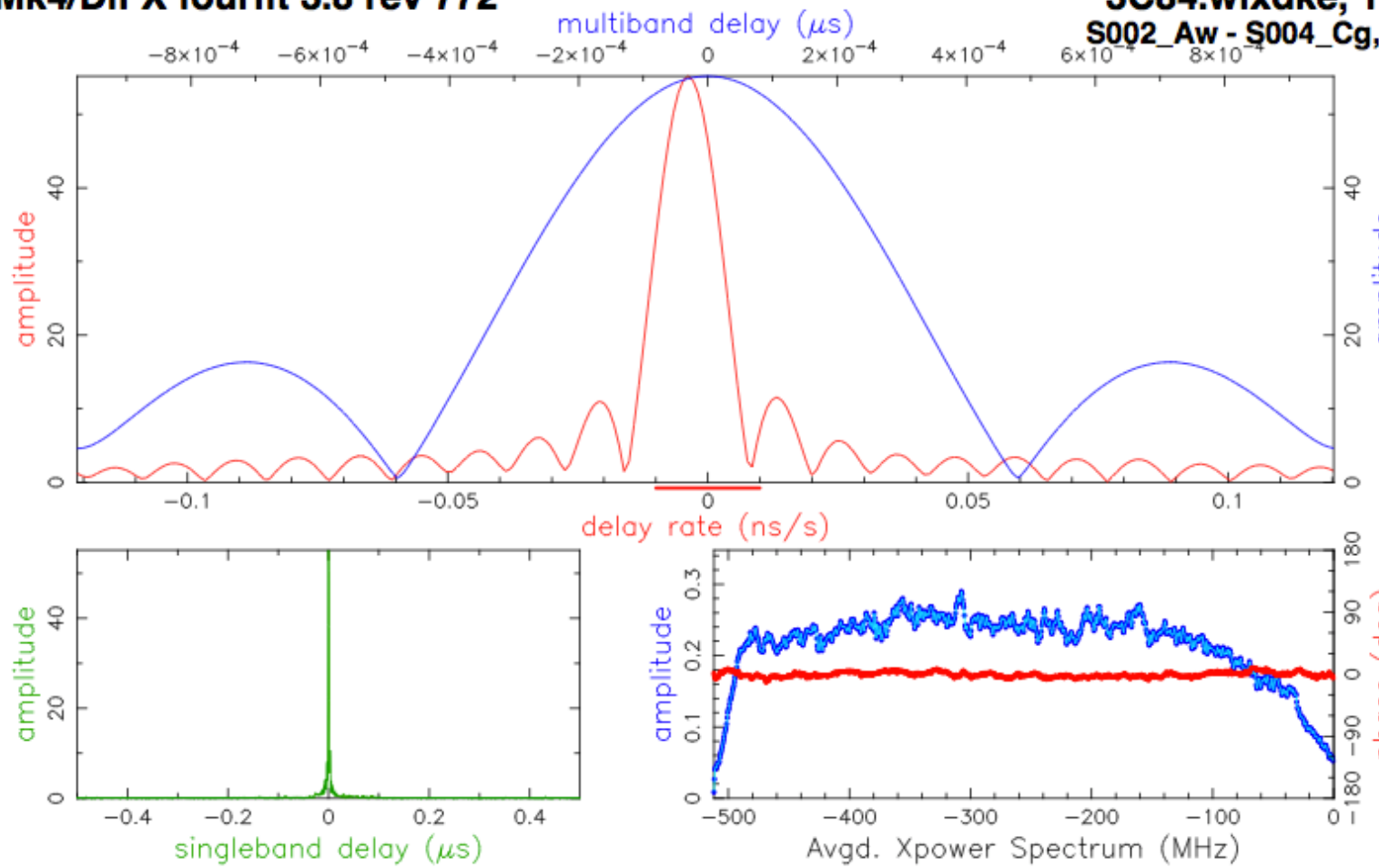
Proof of Concept Experiment

-done with prototype software (v.0)
-June 2012
-Westford – GGAO
-technical details
  VDIF format
  16 Gb/s onto 32 disks
  4 GHz bandwidth on the sky
  dual polarization with 2 GHz IF's
  processed as four 512 MHz channels

Mk4/DiFX fourfit 3.8 rev 772

3C84.wixdke, 171-1927, AC

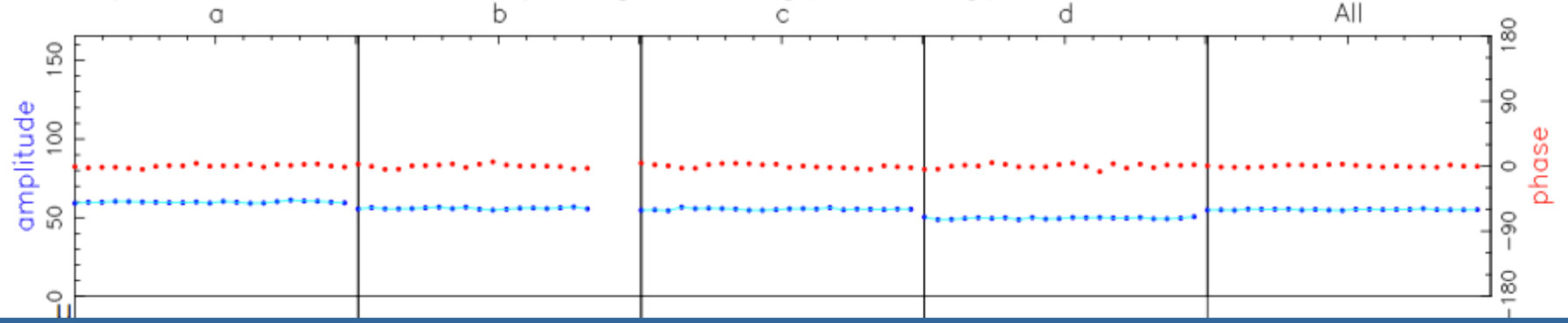
S002_Aw - S004_Cg, fgroup X, pol LL



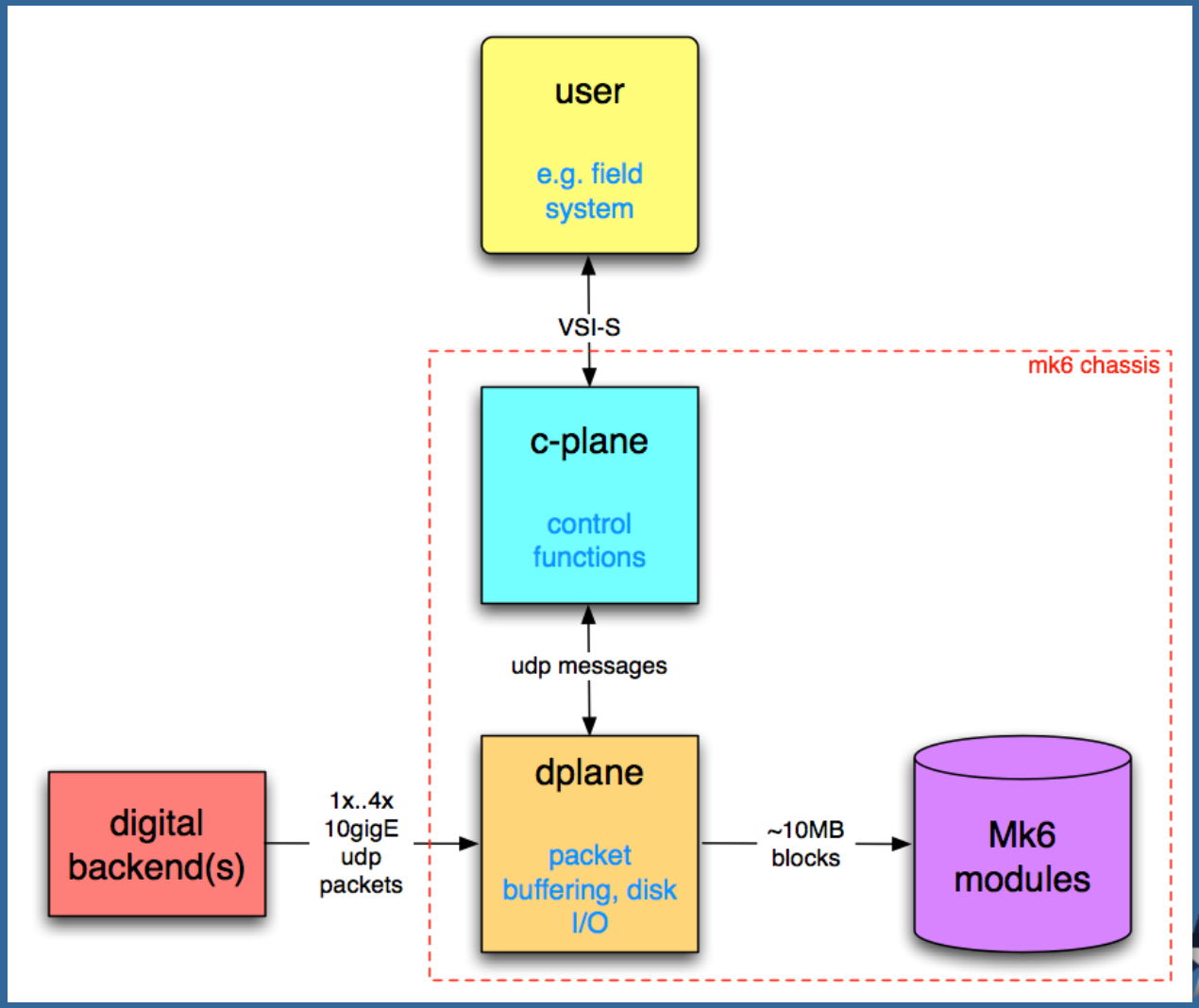
Fringe quality 9

SNR	940.9
Int time	9.720
Amp	55.201
Phase	0.0
PFD	0.0e+00
Delays (us)	
SBD	-0.000007
MBD	-0.000000
Fringe rate (Hz)	
	-0.030985
Ion TEC	0.00
Ref freq (MHz)	
	8592.0000
AP (sec)	0.480
Exp.	y2
Exper #	3407
Yr:day	2012:171
Start	192730.00
Stop	192740.08
FRT	192735.00
Corr/FF/build	
	2012:235:191323
	2012:237:152123
	2012:237:083218
RA & Dec (J2000)	
	03h19m48.1601s
	+41°30'42.104"

Amp. and Phase vs. time for each freq., 21 segs, 1 APs / seg (0.48 sec / seg.), time ticks 1 sec



Mark6 block diagram



c-plane

-control plane
-author: Chet Ruszczyk
-written in python
-interface to user (e.g. field system)
    VSI-S protocol
    command set ver. 3.03
-responsible for high-level functions
    -disk module management
        creating
        mounting & unmounting
    -scan-based recording
    -status, error-checking, etc.

c-plane integration & test

-focused on the VLBI2010 system

-using/controlling 4 RDBE-H's

-8 Gb/s, mk5b format

-RM6_CC master control

- temporary – until FS is ready for mk6
- converts .skd to xml format
- simple time sequencing of scan-based obs.

dplane

-data plane

-author: Roger Cappallo

-written in C

-implements the high-speed data flow

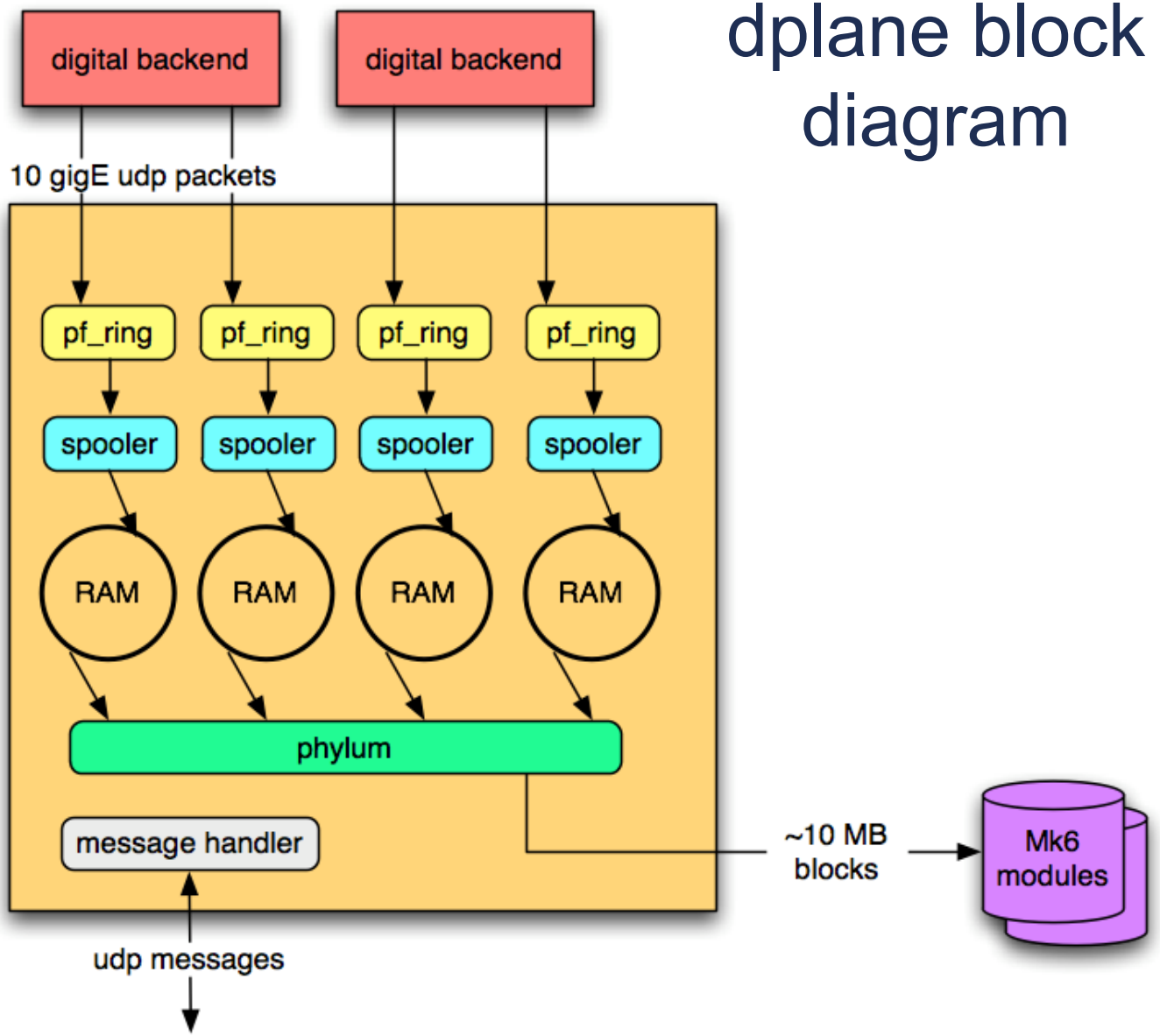
-input from NIC's

-output to disks within mk6 modules

-manages:

- start and stop of data flow via packet inspection
- organization of data into files addition of metadata to files

dplane block diagram



dplane - Technical Highlights

- pf_ring used for high-speed packet buffering
- efficient use of multiple cores – based on # of available cores
  - smp affinity of IRQ's
  - thread binding to cores
- most of physical RAM (16 of 24 GB) grabbed for large ring buffers and locked in
  - one large ring buffer per stream
  - can be changed dynamically from 1 to 4 streams

dplane – file modes

scatter mode

- **~10 MB blocks scattered to files resident on different disks**
  - prepended block# for ease of reassembly
  - uses faster disks to keep up with flow, but balances disk usage as much as possible
  - requires reconstitution of datastream

standalone program _gather_

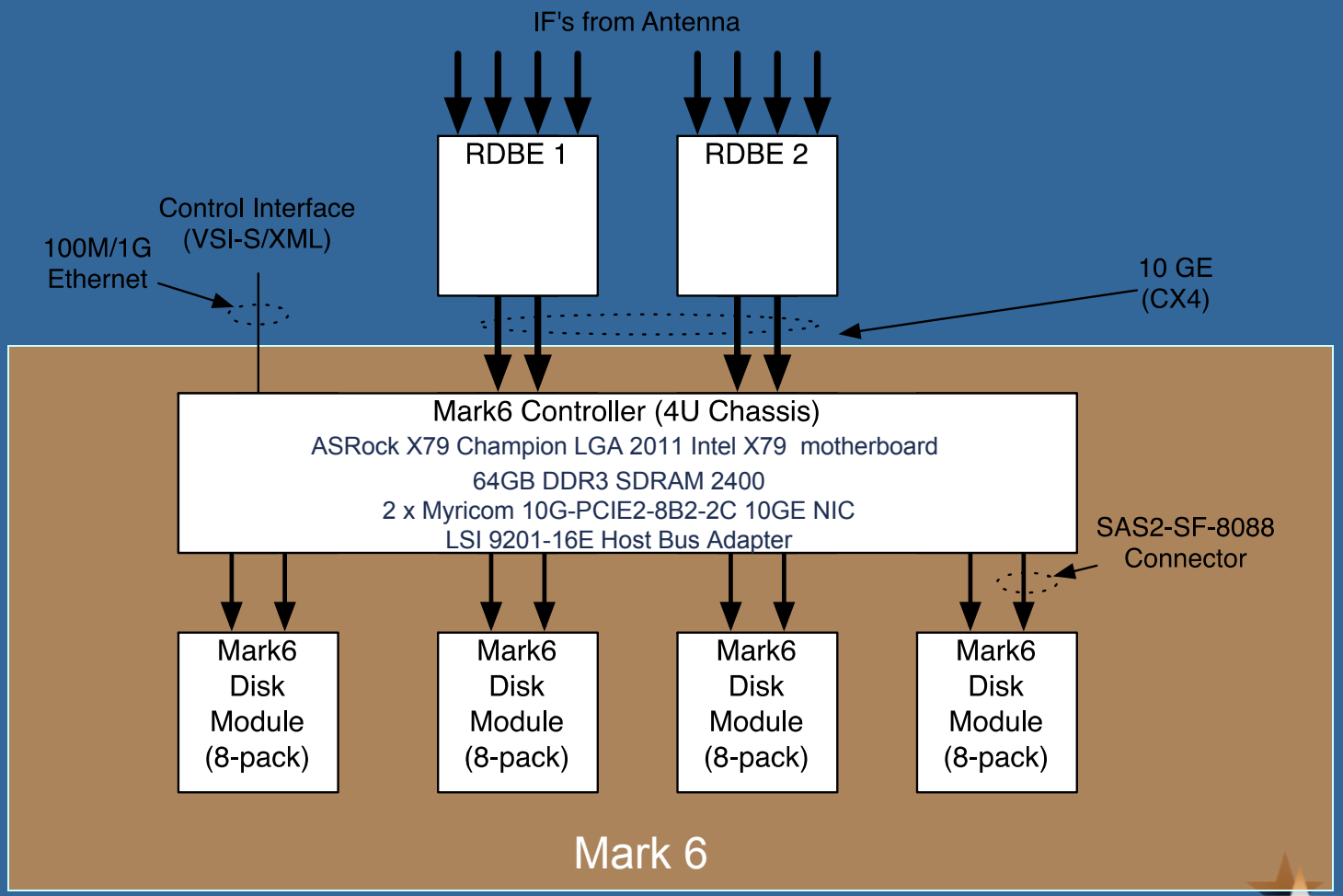
- efficiently writes data in correct order to single file
- not necessary for single-file (RAID) mode
- front end merging software planned for difx

RAID mode

data written to single file

- typically on a RAID array
- good mode for single module of SSD's

Mark 6 16Gbps demonstration system



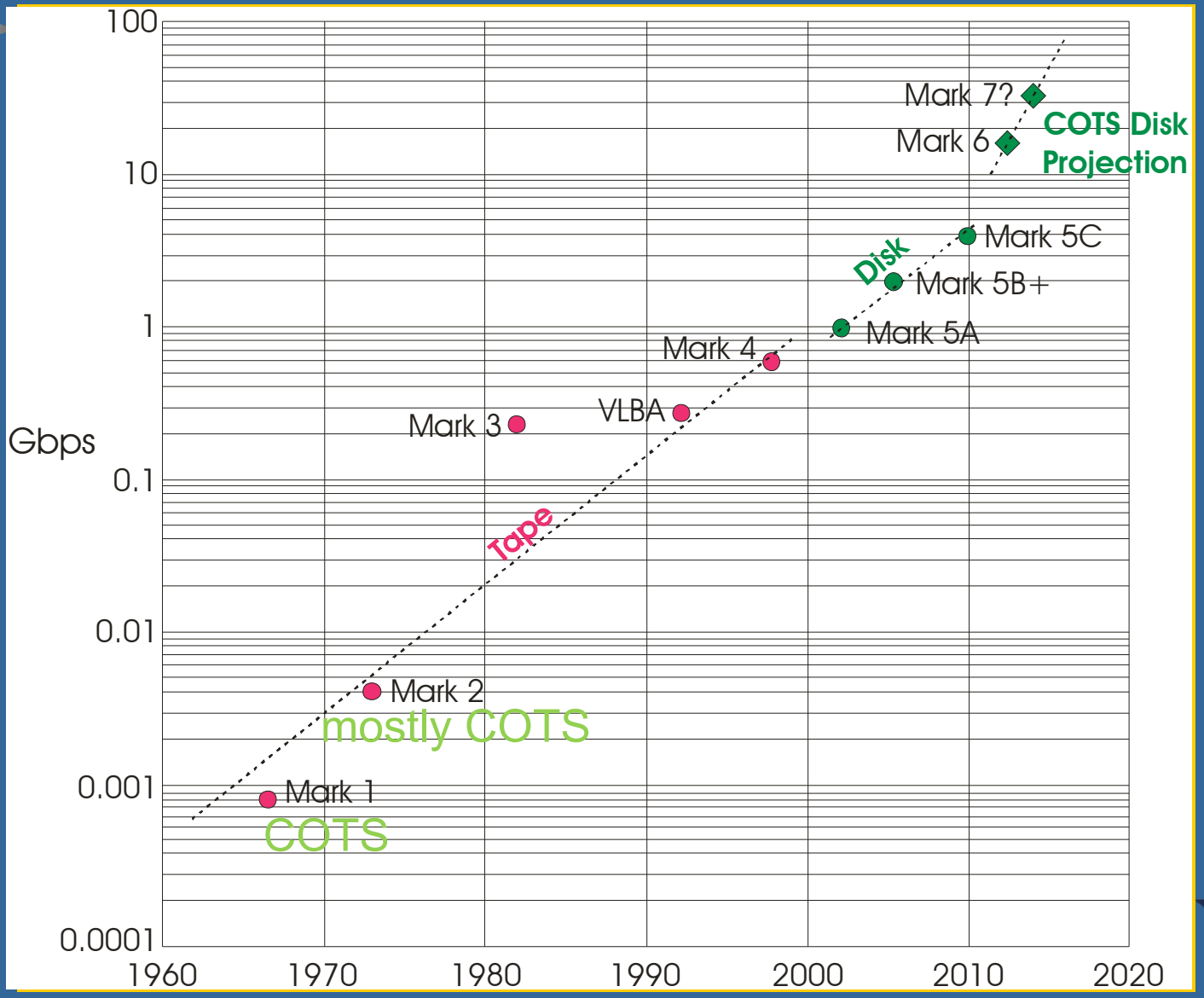
Additional Features

- capture to ring buffers is kept separate from file writing
  - helps to facilitate e-VLBI
- FIFO design decouples writing from capturing (e.g. keep writing during slew)
- mk5b format packets converted “on the fly” into vdif packets
- all Mk6 software is open source for the community
- Mk6 electronics hardware is also non-proprietary & openly published
  - Conduant hardware components known to work
  - Conduant modules are community standard for data transport

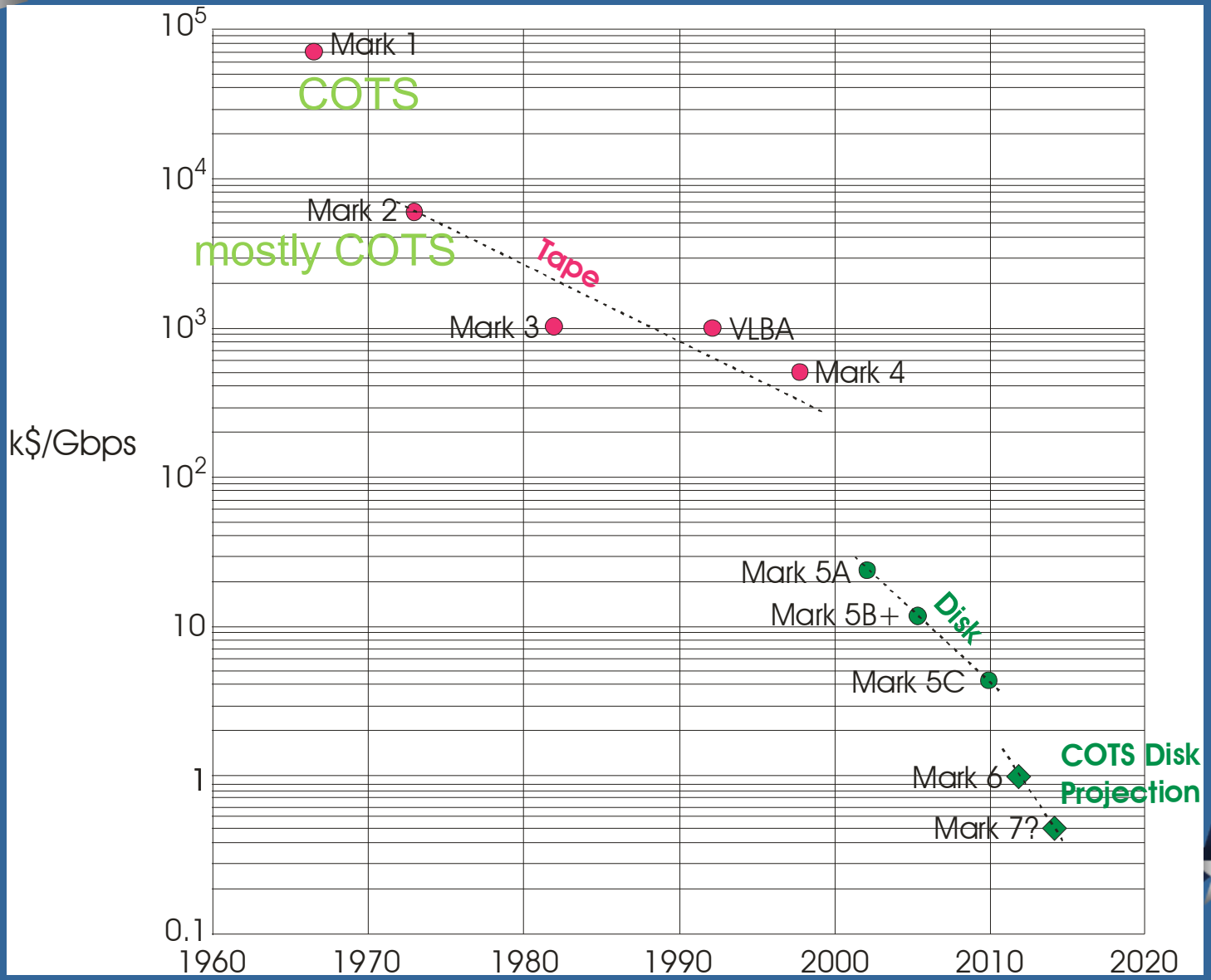
Recent progress

- continuous 16Gbps error-free operation onto 32 disks using ‘scatter’ file system
- ‘scatter’ file writing upgraded to support writing operations as high as 31 Gbps onto 32 disks
- testing of high-level interface software
- long duration simulated experiment usage
- Planned:
  - testing with SSD-equipped Mark 6 modules

Recording-rate capability vs time



Recording-rate cost vs. time



How is Mark 6 available?

- Several options:
  - Purchase new Mark 6 system from Conduant
  - Upgrade existing Mark 5 system (either yourself or with kit from Conduant)
  - Upgrade Mark 5 SATA-modules (with upgrade kits from Conduant)
  - Purchase Mark 6 modules (with or without disks)

Greg Lynott of Conduant will be at Haystack ~10am to 1pm with additional information for anyone interested.

For those interested:

Informal Mark 6 demonstrations will be held Mon and Wed after sessions end (and before dinner); limited to ~12 people at a time; see Chet Ruszczyk or Alan Whitney

Mark 5 SATA Drive Module Upgrade to Mark 6

New Front Panel

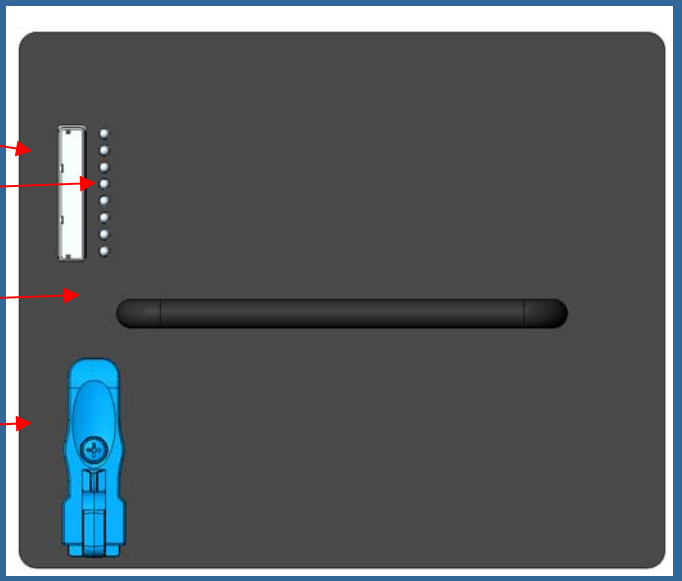
Connectors for two eSATA cables

8x LED (1 per drive)

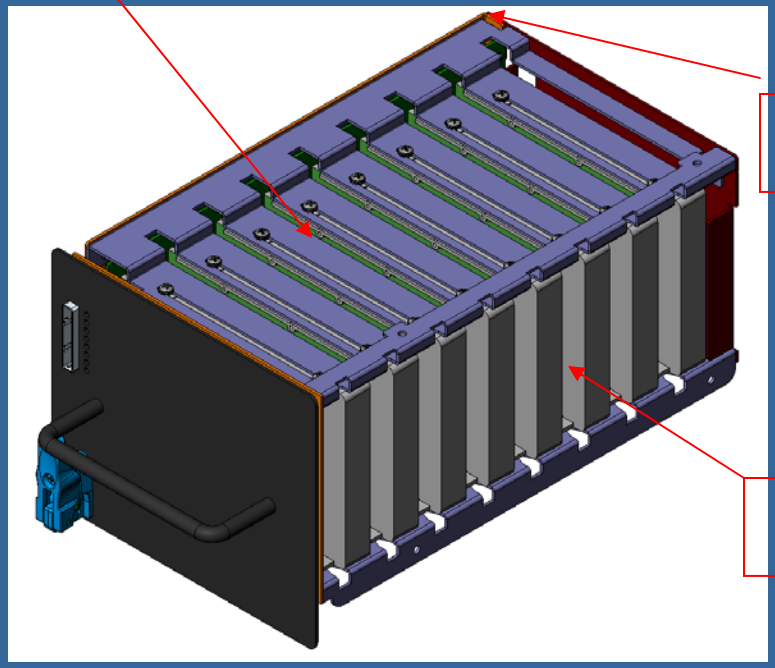
Re-use Handle from old Module

pre-installed

Front Panel

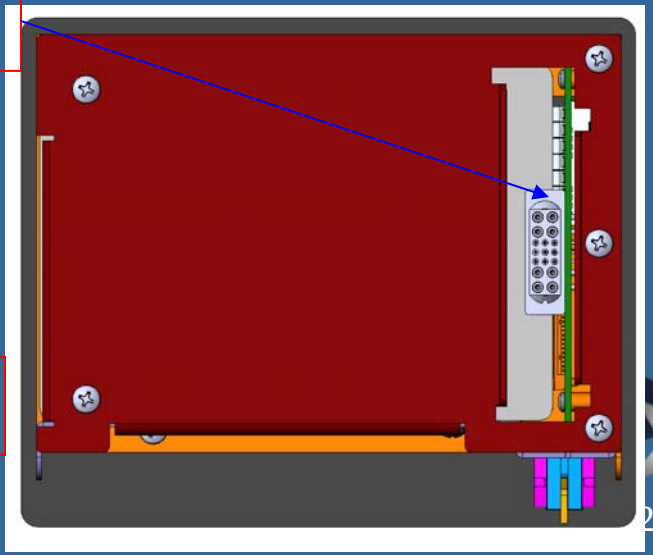


Cooling slots



New PCB and power connector

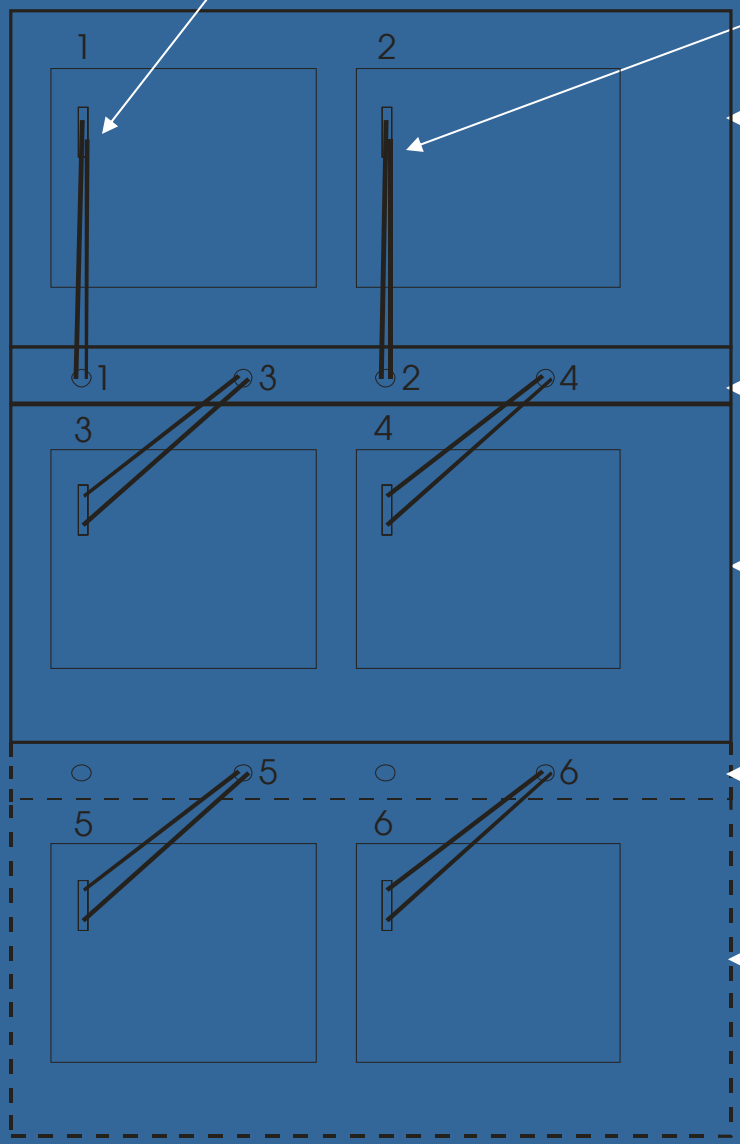
Rear Panel



Easily removable disks

Thank you

Mark 6 Physical Layout



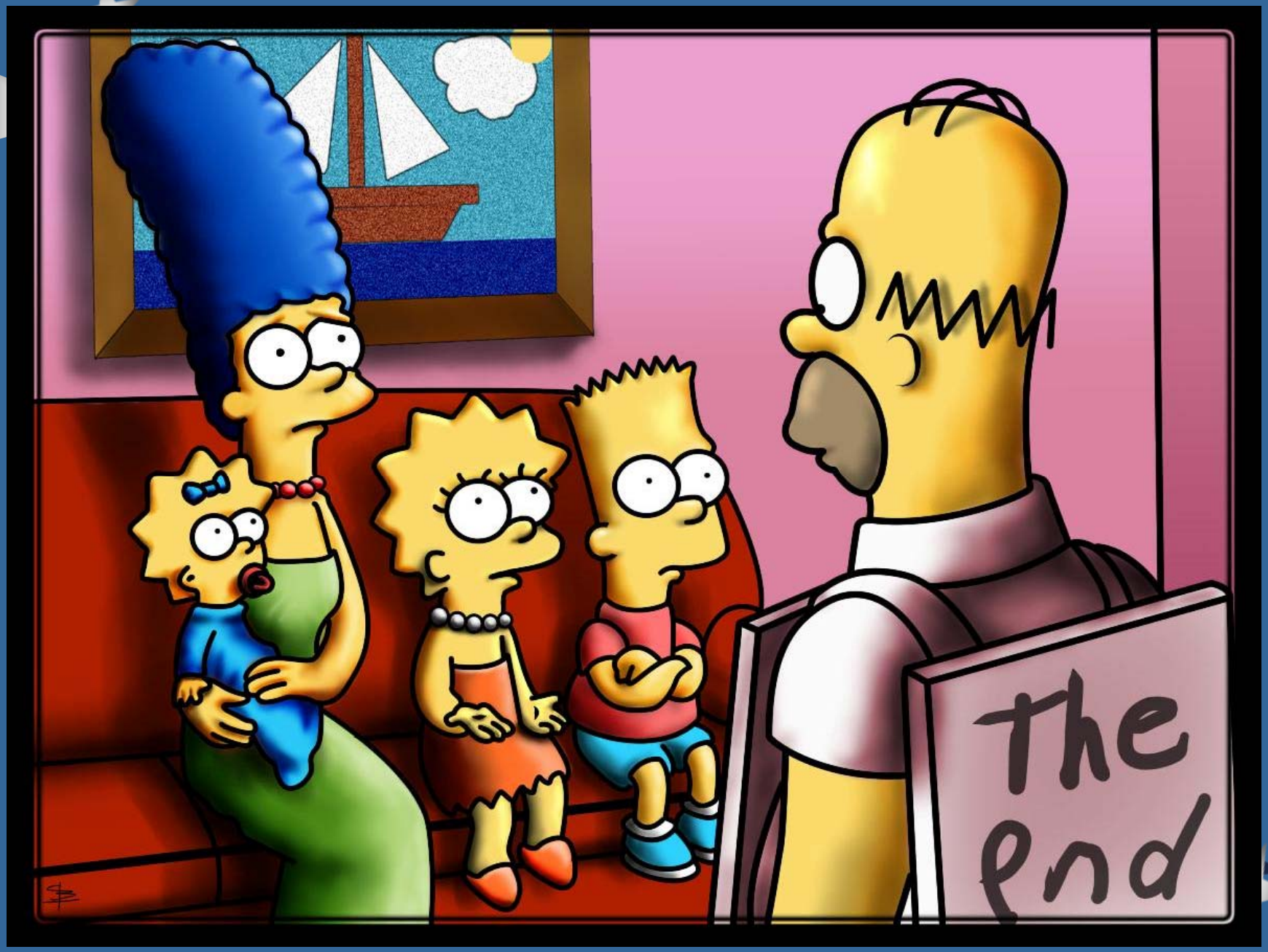
System chassis (supports 8Gbps by itself)
- Mark 5/6 enclosure
- Power supply
- Chassis backplane kit
- Data-electronics hardware

Retractable cable panel
(for easy management of eSATA data cables)

Expansion chassis (needed for 16Gbps)
- Mark 5/6 enclosure
- Power supply
- Chassis backplane kit

(Optional) 2nd cable panel

(Optional) 2nd Expansion chassis


The
End

Software versions and strategies

v.0 prototype (RAID 0)
- command line one-off control

v.1 operational RAID-based code
- continuous operation
- control via messages

v.2 with single output file per stream:
- write (ordinary) Linux files on RAID arrays
- can use normal file-based correlation directly

v.2 with multiple files data need to be reconstructed:
- *gather* program – interim solution
  - does so very efficiently
  - requires an extra step
- FUSE/mk6 interface could be written
- in difx: will likely implement native mk6 datastream

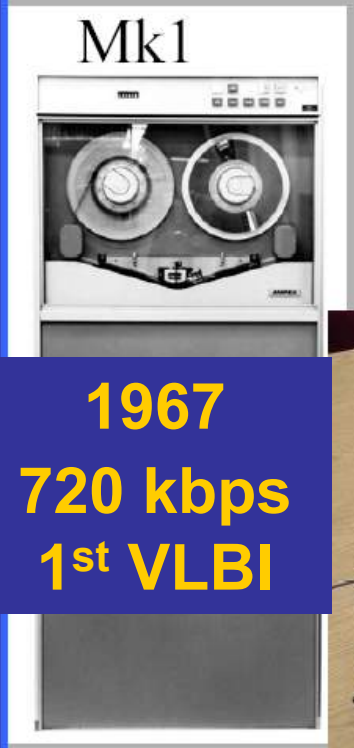
Data rates for VLBI2010

- VLBI2010 data rates are dictated by
  - Small antennas (12m class)
    - Antennas must be able to move very quickly around sky
  - Weak sources
    - Sources need to be ~uniformly distributed in the sky and have simple or no structure, which severely constrains available sources
  - Short observations
    - VLBI2010 on-source observations will be ~10 secs each
    - antenna must move around sky quickly to map temporal atmospheric fluctuations
    - most of observation period is spent moving antenna from source-to-source

All these factors conspire to dictate very high data rates (both instantaneous and average)

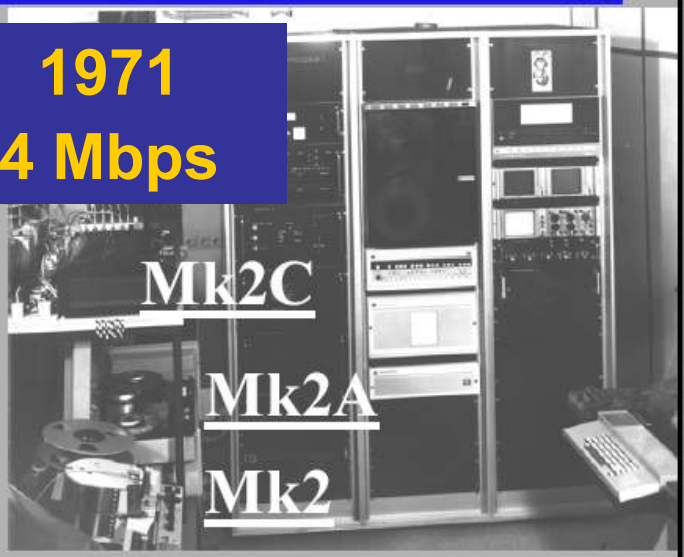
VLBI Data Rates and Volume

- VLBI2010 at 4 Gbps/station average, 4 to 20 stations
  - ~5-40 TB/station/day
  - Global 10-station experiment @ 4 Gbps/station up to ~400 TB/day
  - Single 10-day experiment can produce up to ~4 PB
- Higher data rates (16-32 Gbps) are already being demanded for higher sensitivity – ALMA phased array produces 64Gbps!
- Available disk supply can support only few days of observations at these rates

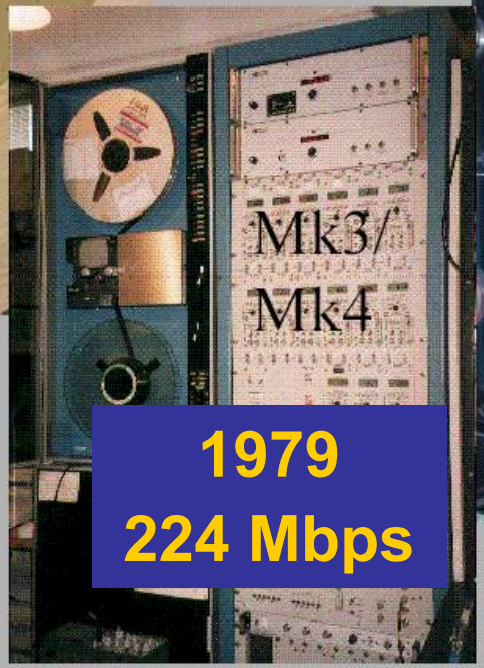


1967
720 kbps
1st VLBI

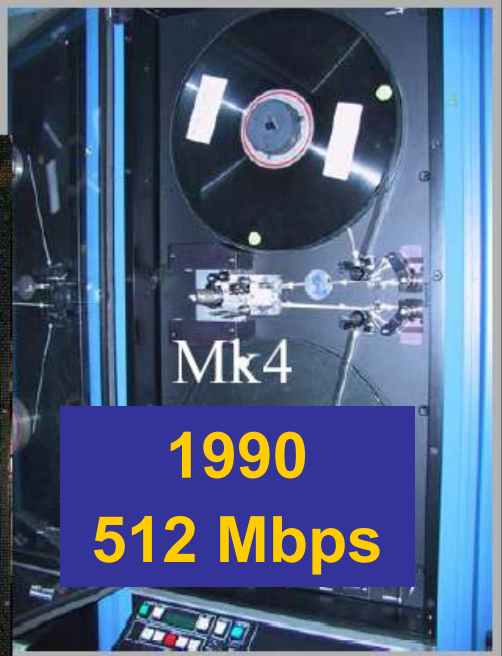
1971
4 Mbps



2002
1 Gbps
2006
2 Gbps
2011
4 Gbps



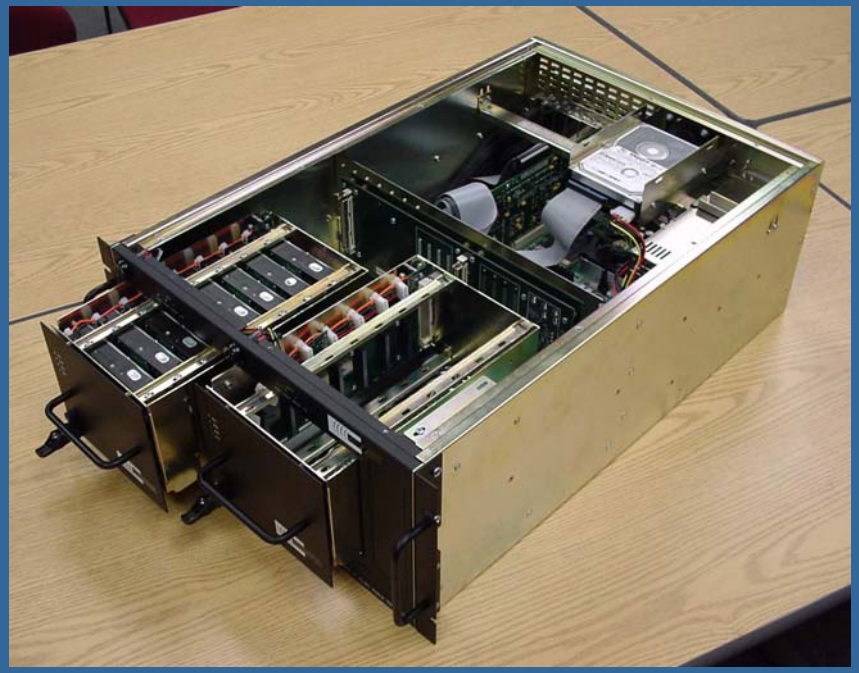
1979
224 Mbps



1990
512 Mbps

Mark 5 Data Acquisition System

(Mark 5A/B/B+/C all look the same)



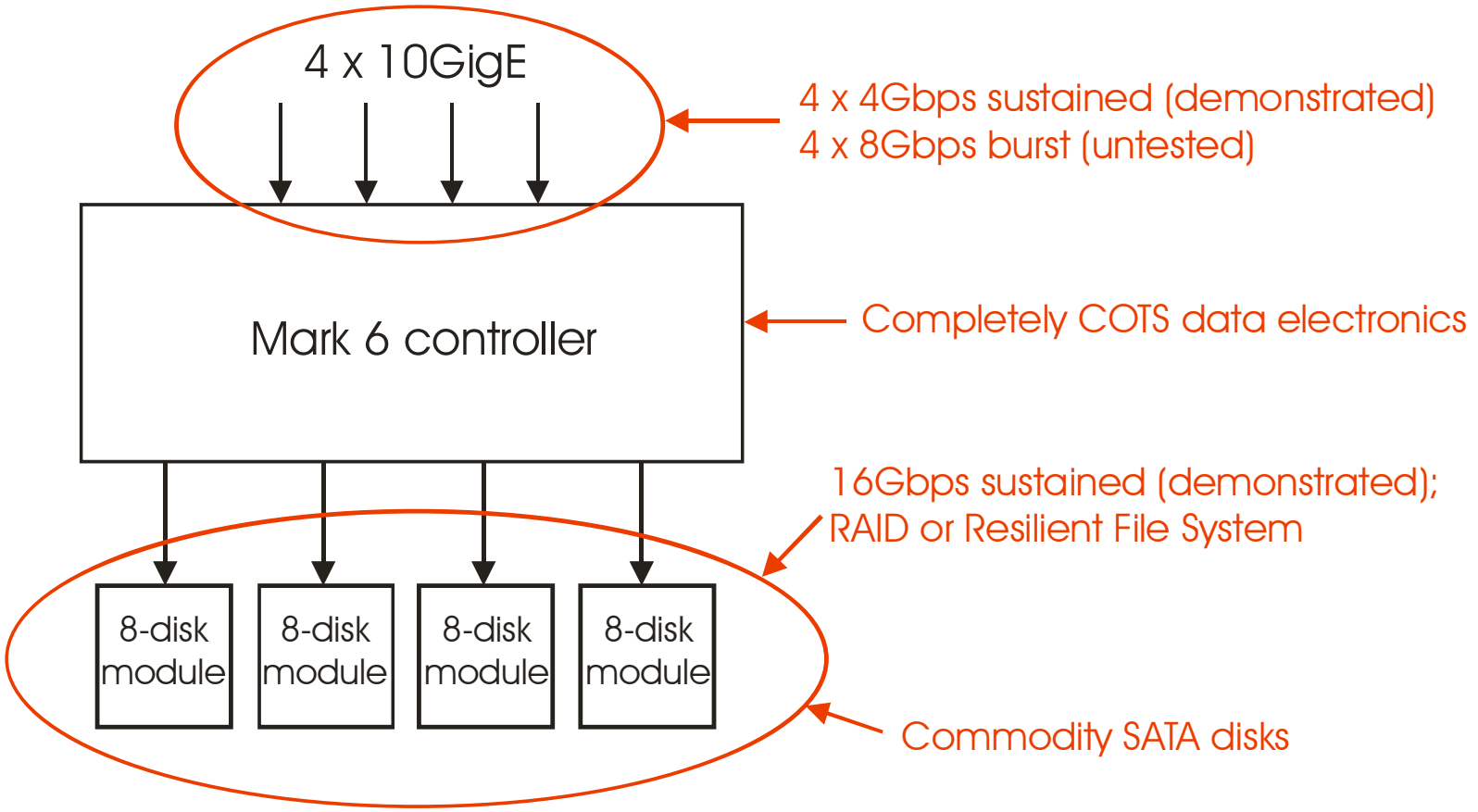
	Year introduced	Record rate (Mbps)	Interface	Cost (USk$)	#deployed
Mark 5A	**2002**	**1024**	**Mk4/VLBA**	**21**	**~130**
Mark 5B	**2005**	**1024**	**VSI-H**	**22**	**~40**
Mark 5B+	**2006**	**2048**	**VSI-H**	**23**	**~30**
Mark 5C	**2011/12**	**4096**	**10GigE**	**21**	**~20**

Mark 5 includes a significant amount of proprietary technology

Next up – Mark 6

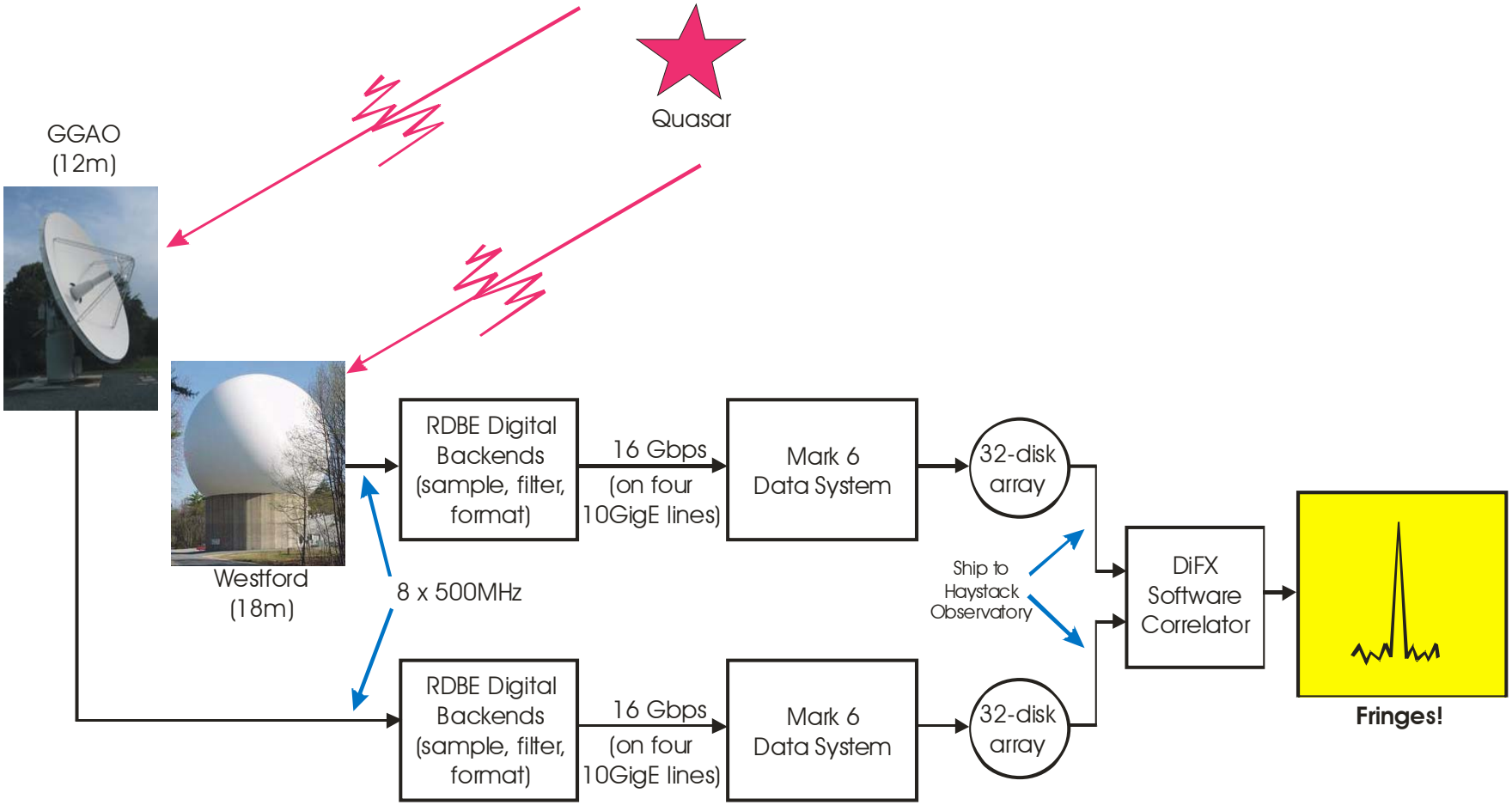
- 16Gbps sustained record/playback
- 4 x 10GigE input ports
- Based on inexpensive high-performance COTS hardware
- Linux OS w/open-source software
- Resilient ‘scatter/gather’ file system to manage slow and failed disks
- VDIF/VTP compliant
- Goal is to preserve as much investment in existing Mark 5 systems and disk libraries as possible
- Mark 6 collaboration:
  - Haystack Observatory – all software and software support
  - NASA/GSFC High-End Network Computing group – consultation on high-performance COTS hardware
  - Conduant Corp – Mark 6 disk module and power backplane

Basic Mark 6 System

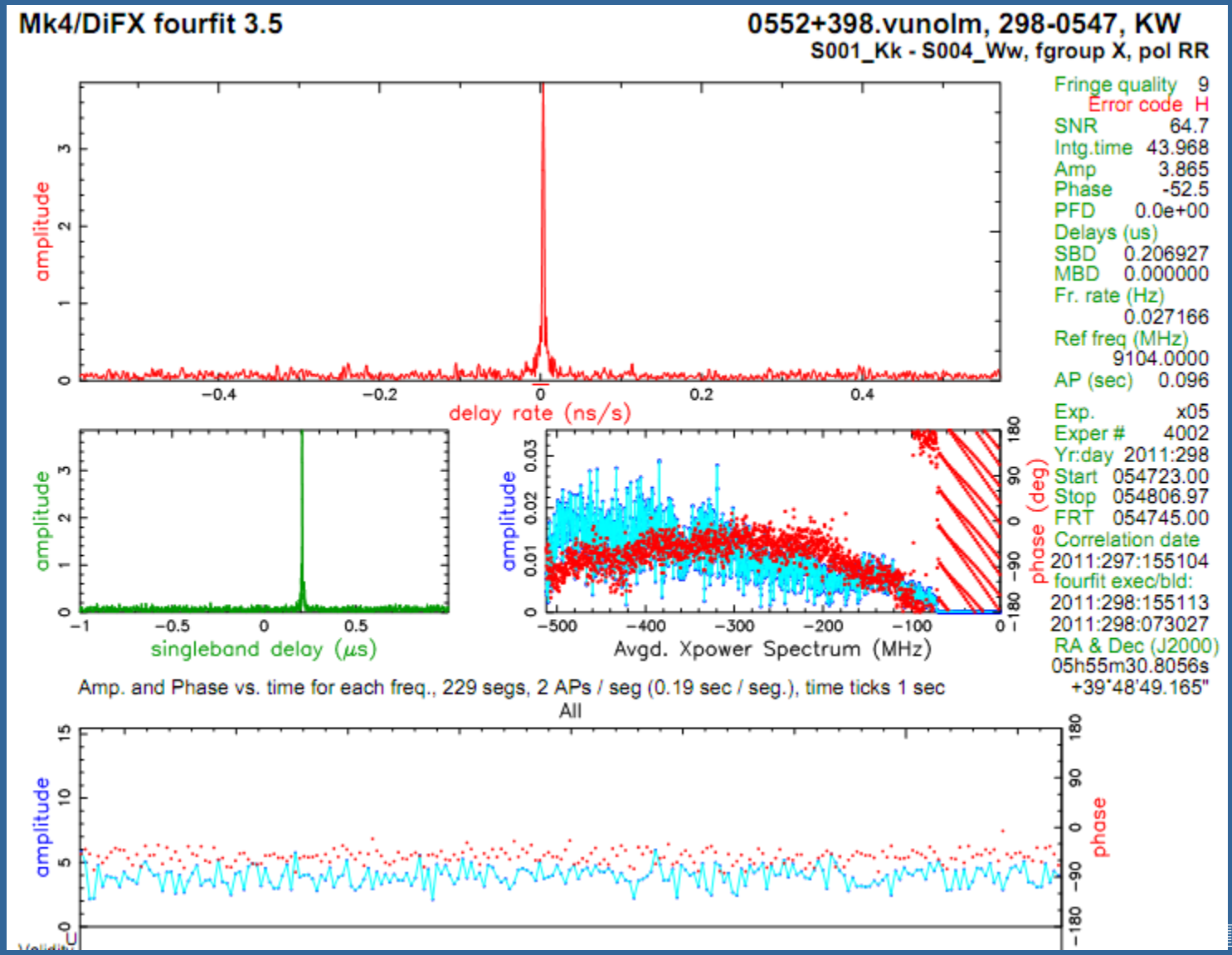


16 Gbps VLBI demonstration with Mark 6

24 October 2011



Correlation results (single 500MHz channel)



Mark 6 Project Status

- Sustained 16Gbps from four 10GigE interfaces to disk is now readily achieved
- 'Scatter/gather' file system manages around slow disks
- Mark 6 systems now being routinely used in VLBI2010 development work (at 8Gbps, replacing 4 Mk5C units at each antenna)
- To be completed:
  - Full VSI-S command set
  - Playback as standard Linux files
- Prototype Mark 6-specific hardware pieces arrived at Haystack last week from Conduant
  - New Mark 6 chassis-backplane power-management boards
  - Mark 5-to-Mark 6 SATA disk module upgrade kit
  - New cable-management panel

Projected Mark 6 schedule

- Mar 2012 – GGAO/Westford Mark 6 test with broadband VLBI2010 system (dual-pol with 2GHz BW/pol)
- Mar 2012 – Test Conduant prototype hardware; integrate complete hardware system; begin integration with Field System
- mid/late 2012 – System complete and fully tested; (new complete Mark 6 system <$15k)

VDIF
(VLBI Data Interchange Format)

- Standardized format for raw time-sampled VLBI data
- Compatible with both VLBI data-recording systems and e-VLBI data transmission
- Highly flexible to accommodate a large variety of channel and frequency configurations, including mixed sample-rate data
- VDIF being implemented for all new VLBI2010 systems
- Accompanying VLBI Transport Protocol (VTP) specifies e-VLBI data-transmission protocol for VDIF-formatted data stream

For details: www.vlbi.org

VTP
(VLBI Transmission Protocol)

- Companion specification to VDIF
- Specifies e-VLBI data-transmission protocol for VDIF-formatted data streams
  - Normally must use UDP or UDP-like protocol to maintain necessary data rate
  - Addition of Packet Serial Numbers (PSNs) helps to keep packets organized and identify missing packets (a few missing packets are not normally a problem)

Backup slides

Thank You's

Haystack/Westford –
Chris Beaudoin, Pete Bolis, Roger Cappallo, Shep Doeleman, Geoff Crew, Rich Crowley, Dave Fields, Alan Hinton, David Lapsley, Arthur Niell, Mike Poirier, Chet Ruszczyk, Jason SooHoo, Ken Wilson

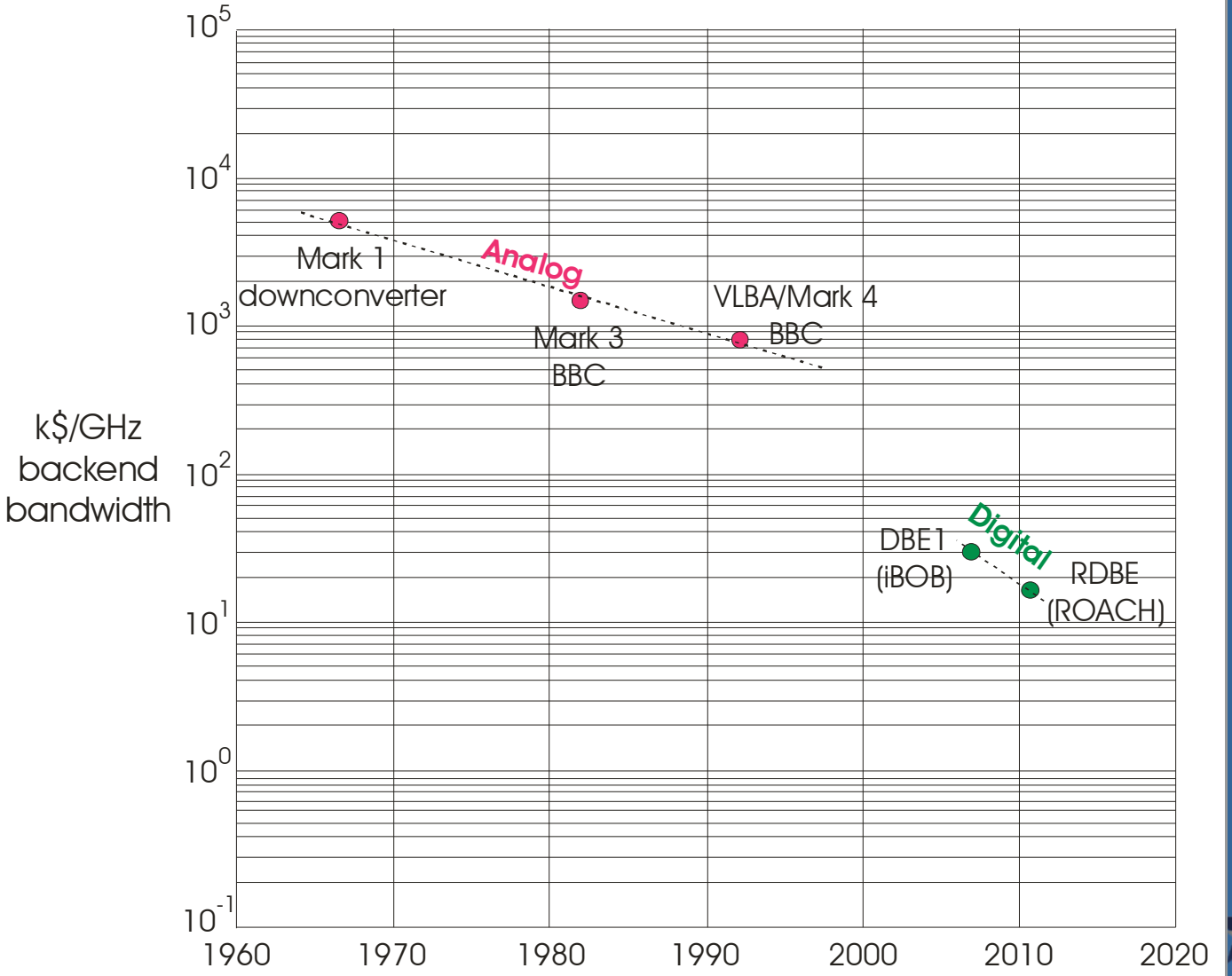
NASA/GSFC VLBI Group –
Tom Clark, Ed Himwich, Chopo Ma

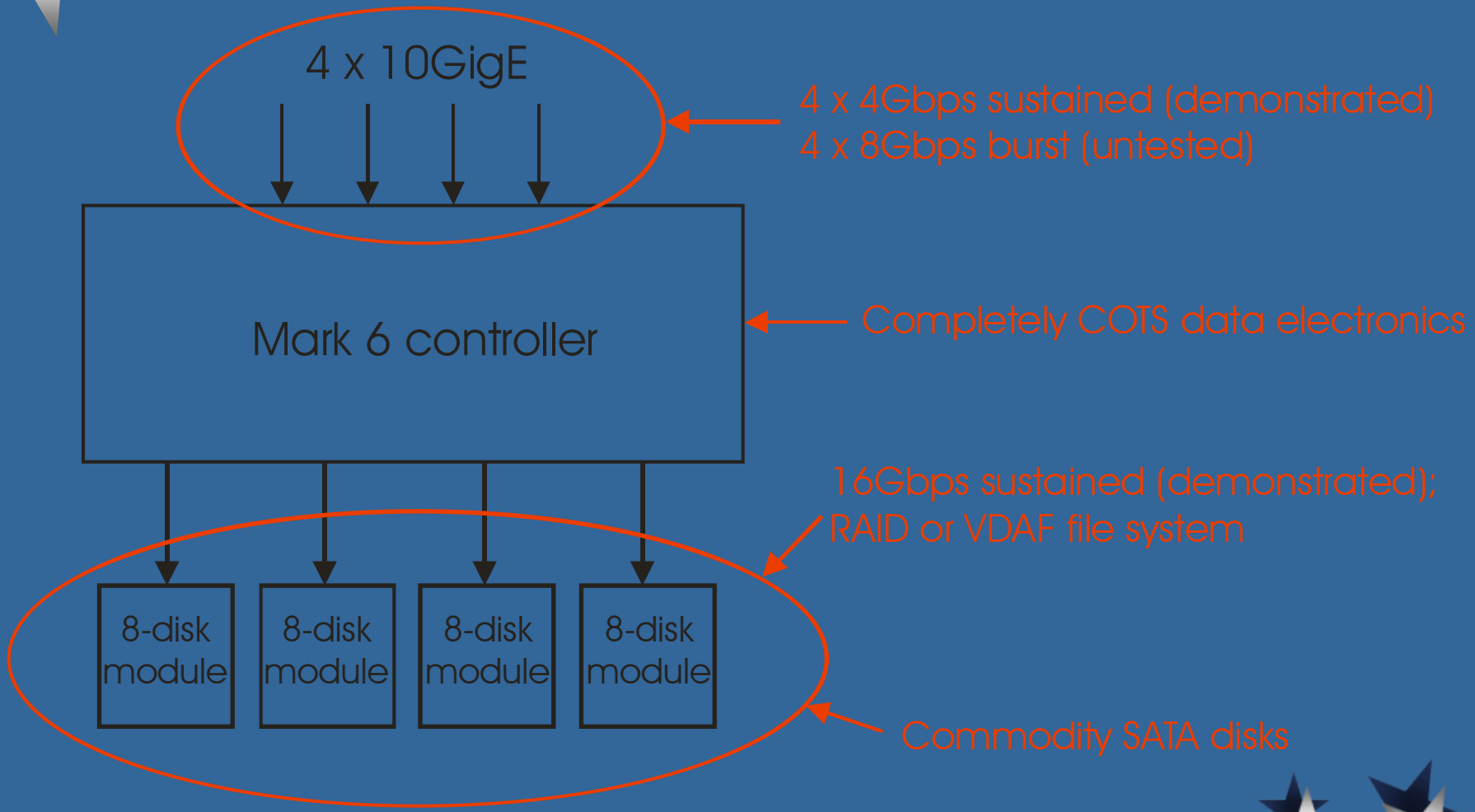
NASA/GSFC GGAO –
Roger Allshouse, Wendy Avelar, Jay Redmond

NASA/GSFC High-End Computer Networking Group –
Bill Fink, Pat Gary (recently deceased), Paul Lang

Conduant –
Phil Brunelle, Greg Lynott, Ken Owens

Backend-bandwidth cost vs. time



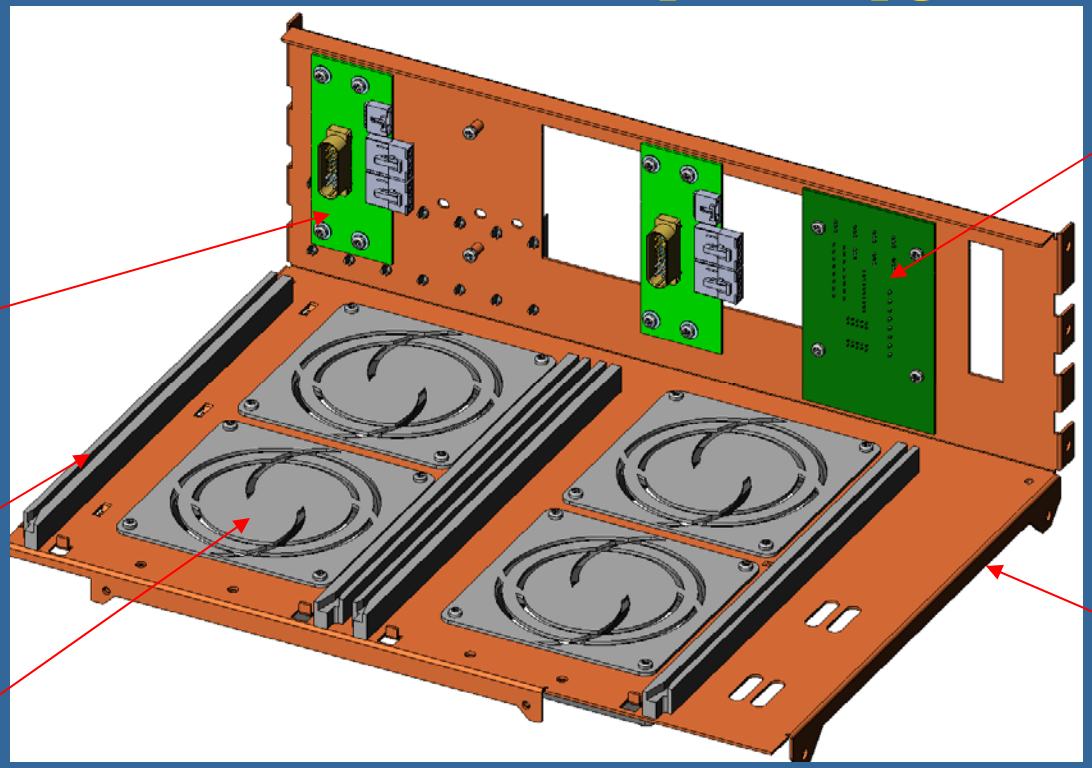

4 x 10GigE
4 x 4Gbps sustained (demonstrated)
4 x 8Gbps burst (untested)
Mark 6 controller
Completely COTS data electronics
16Gbps sustained (demonstrated);
RAID or VDAF file system
8-disk module
8-disk module
8-disk module
8-disk module
Commodity SATA disks

Mark 5 Chassis Backplane Upgrade

New Drive Module Backplane (x2):
-Sequences power to disks
-Regulates voltage at disk power **pins**

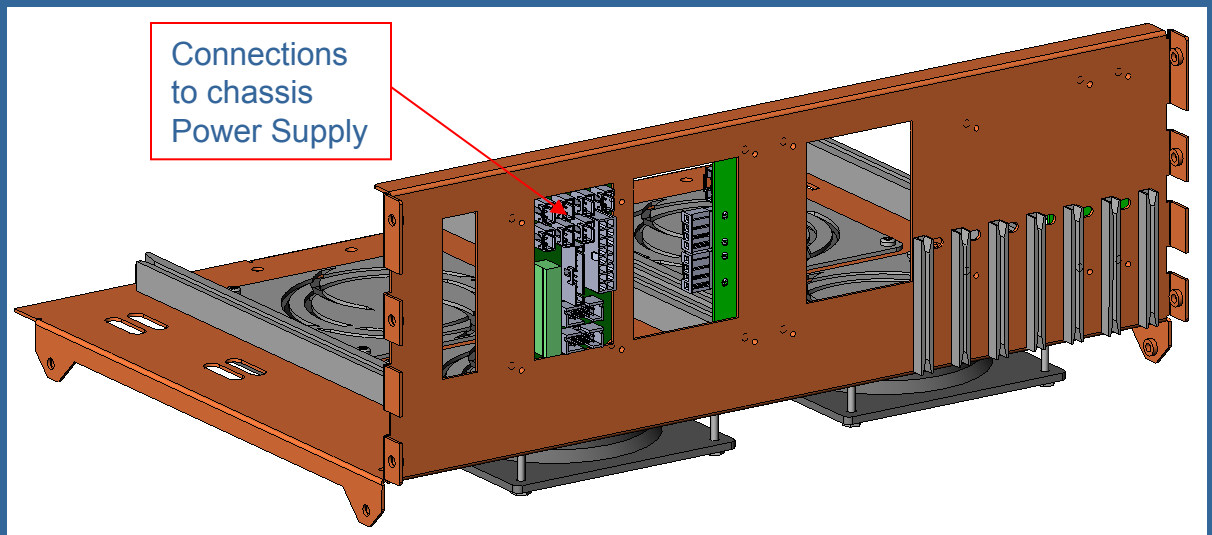
Module guide rails

Cooling fans



New Connector Board:
-simple disconnect to allow easy removal of Module Tray from chassis

Module Tray



- Choose best hardware (our partners at NASA High End Computer Networking generously provided the entire hardware specification based on extensive NASA/GSFC testing)
- Optimize settings such as interrupt-to-processor mapping and process-to-processor mapping
- Control-plane integration
  - Implement full-set of operational controls
  - Minimize stress of transition from Mark 5 to Mark 6
- Thorough testing in real-world environment

Mark 6 M&C and concepts

- VSI-S command set
- Recording units are defined as ‘volumes’, each of which consists of one or more physical disk modules
  - Multi-module volumes are required for recording rates >~4Ggps
  - Multi-module volumes retain identity thru correlation processing, then are returned to single-module volumes
- Volumes are managed on an ordered ‘Volume Stack’ that allows multiple volumes to be mounted simultaneously
  - Allows volumes to be queued in specific order for usage
  - Supports automated switchover to next volume in Volume Stack when current module becomes full; switchover takes place between scans
- Disk statistics gathered during recording allow easy identification of slow/failing disks by disk serial number