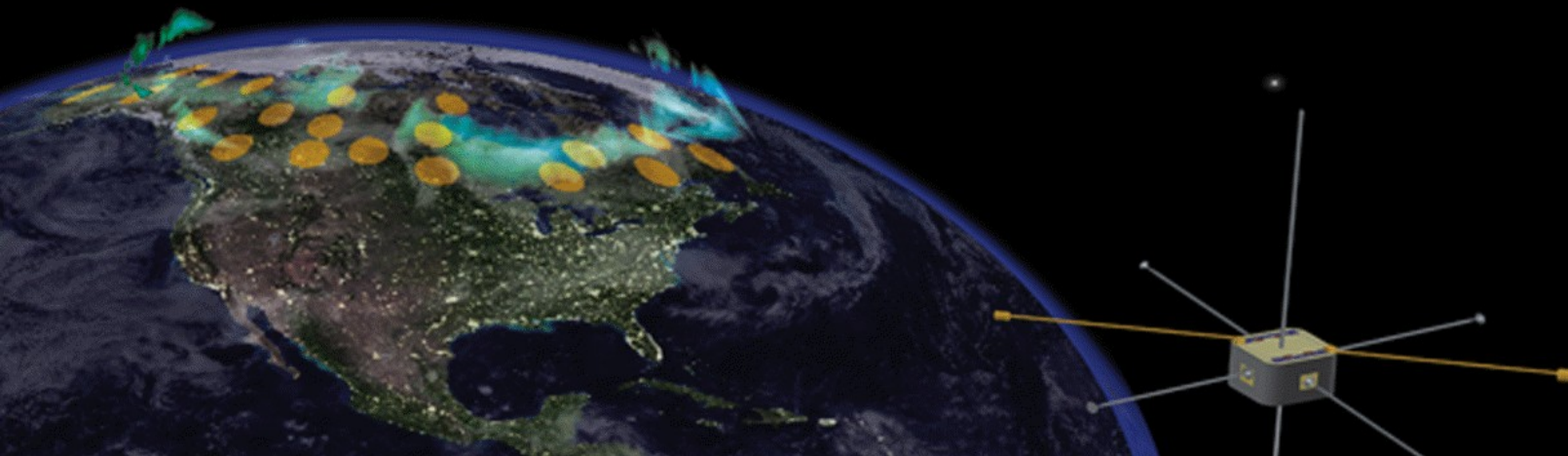


GBO Data Flow

Brian Jackel - University of Calgary

- Site level software
- Real-time data
- Full-res data retrieval & ingestion
- Mirroring/echoing
- Site monitoring



Data streams

1. Fluxgate magnetometer

- 0 Full resolution
- 1 Real time

2. All-sky imager

- 0 Full resolution
- 1 Real time
- 2 Summary

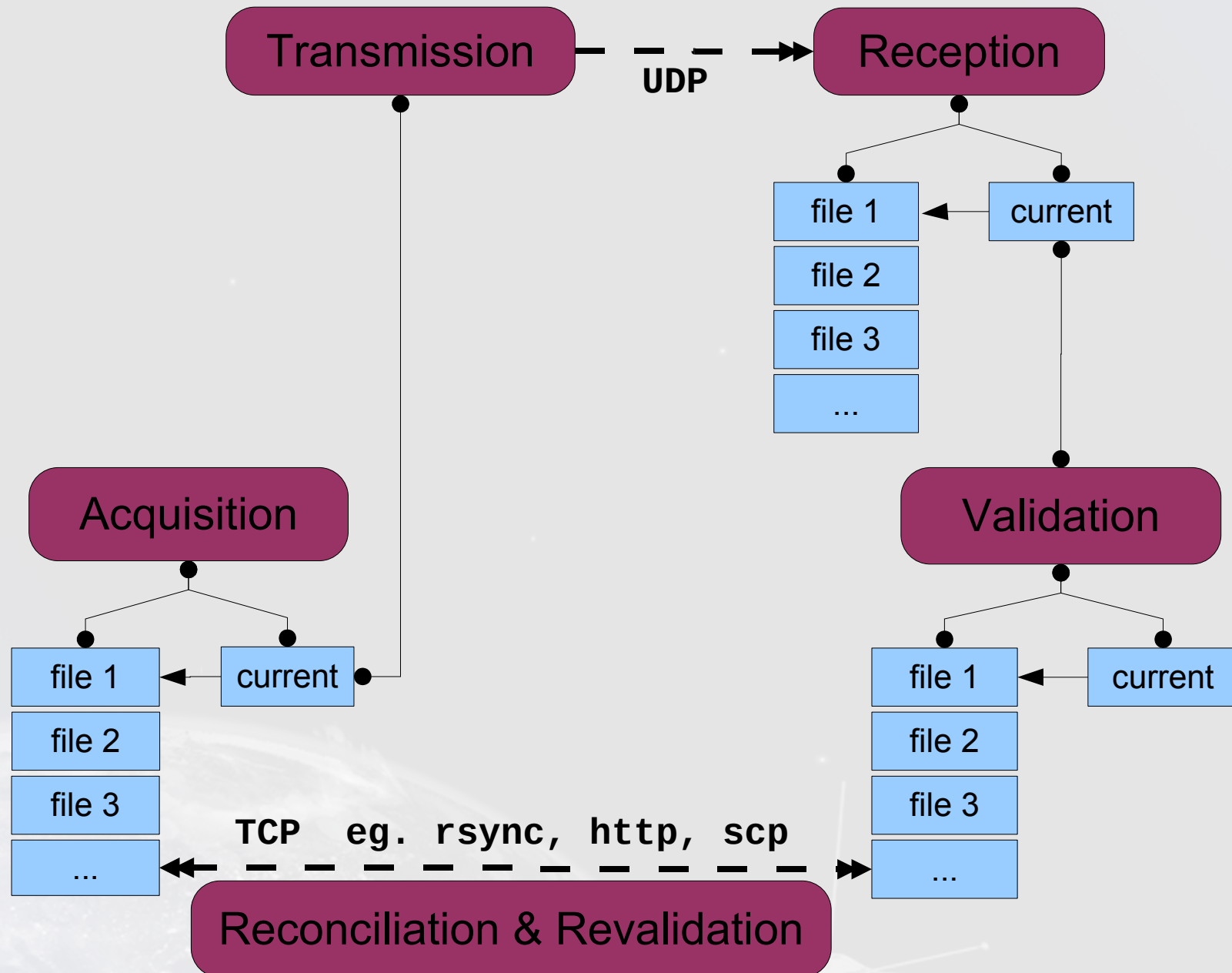
3. Health & housekeeping

- 0 Full resolution
- 1 Real time
- 2 Summary

Data transport

- Hard disks in USB enclosures
 - 160GB “transfer” shipped to Calgary every 4-6 months
 - 200GB “backup” remains on site; older files removed after ingestion from transfer disk
- Internet
 - Telesat HSi < 4.6 kbps
 - Local ISP (WHIT, RANK, GBAY, MCGR, FYKN)
 - Corporate VPN (EKAT)
 - Telesat HSe being tested: more bandwidth but dynamic IP
- Iridium: last resort

Network data flow

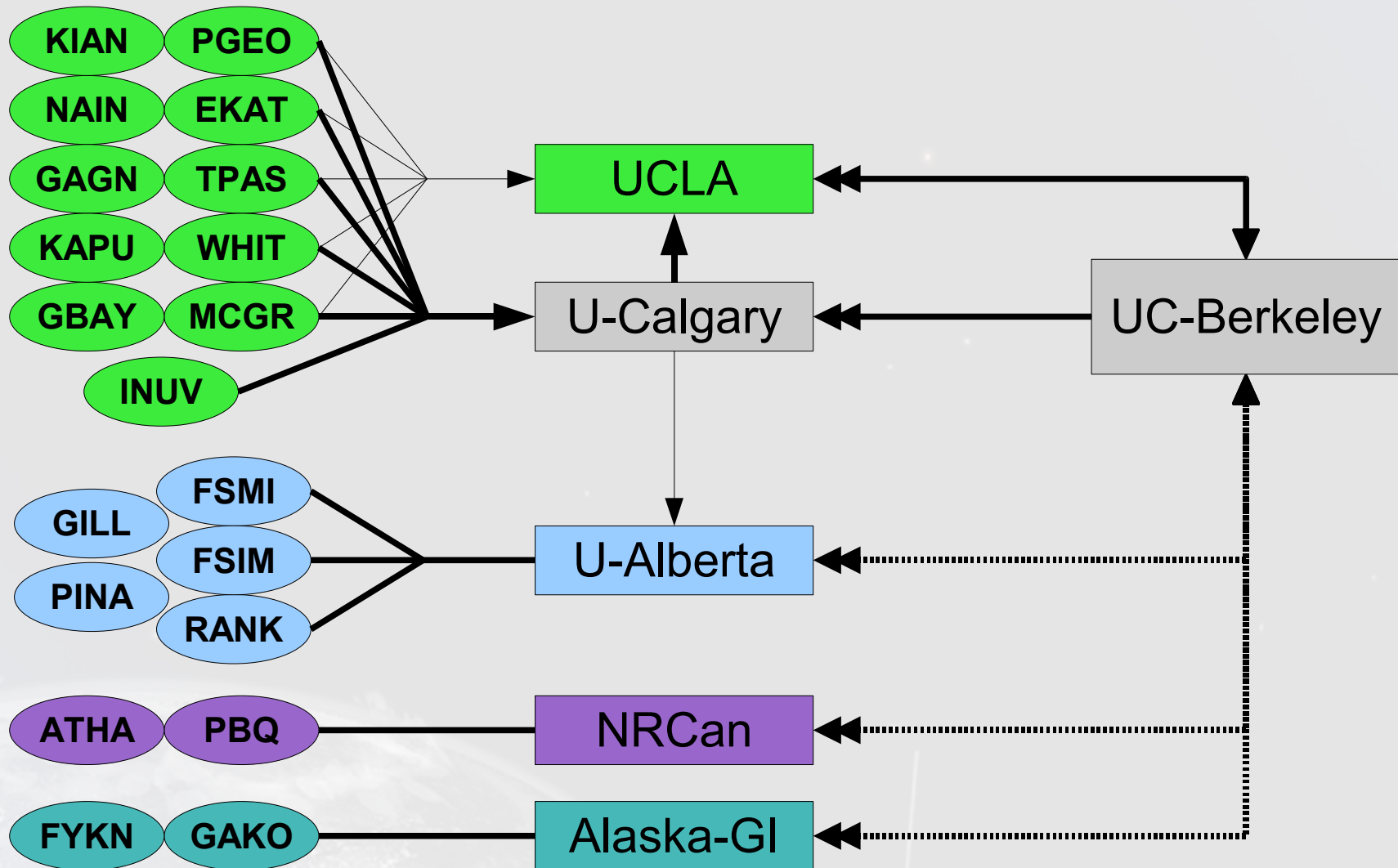


Data tools

- themis-rt.phys.ucalgary.ca
 - dual 2 Ghz AMD Athlon CPUs
 - three 160 GByte hard drives in RAID-1 with hot spare
- themis-data.phys.ucalgary.ca
 - dual 2.8 Ghz Xeon hyperthreaded CPUs
 - 1 GByte of RAM (expandable to 16G)
 - 3 TBytes of internal (SATA) storage
 - 6 TBytes of external (SCSI) storage

Both systems have dual 100/1000 ethernet ports. Data are provided internally to UCalgary via NFS and SMB, and externally to UC-Berkeley and UCLA via HTTP and rsync.

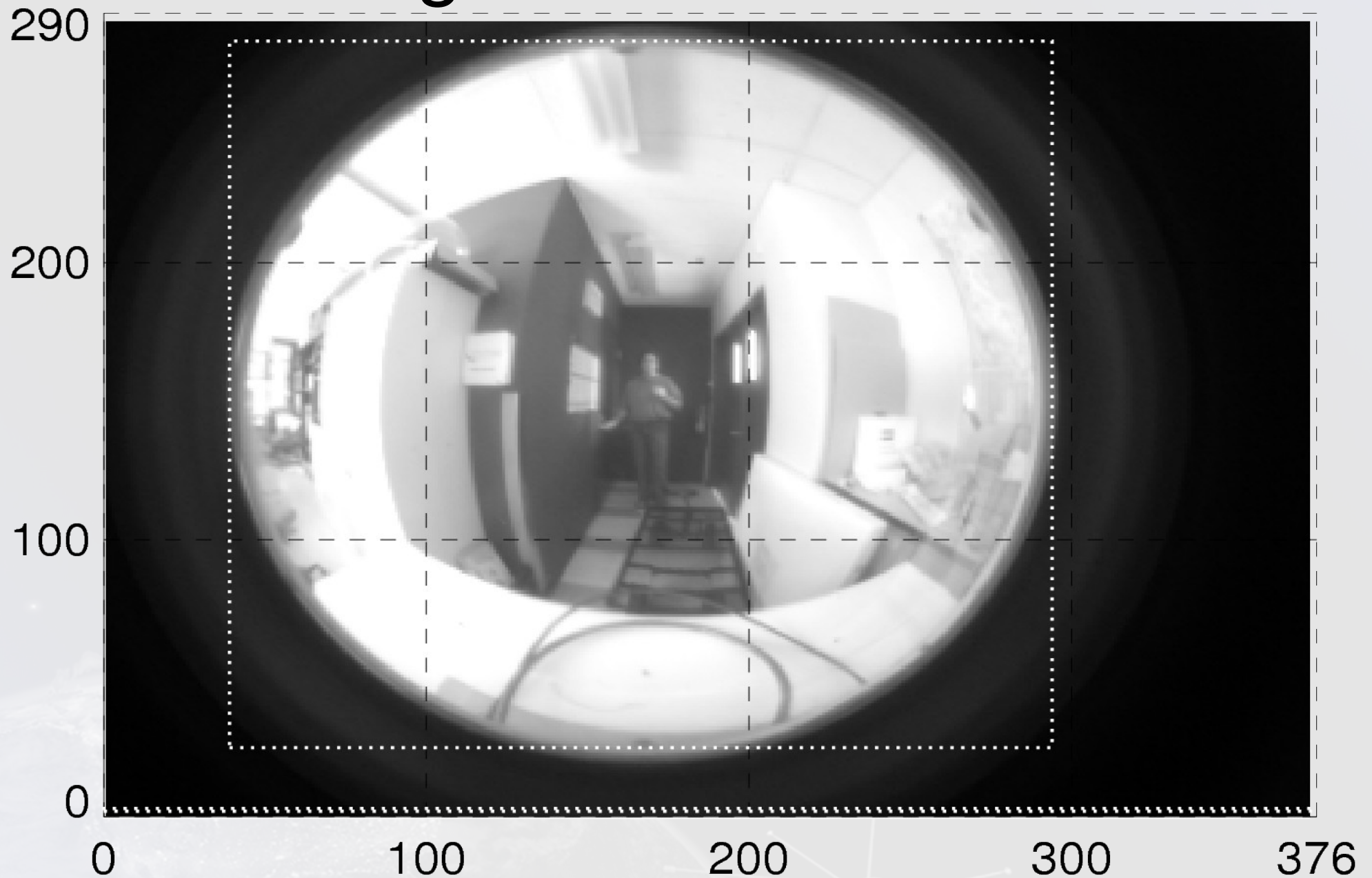
Magnetometer data flow



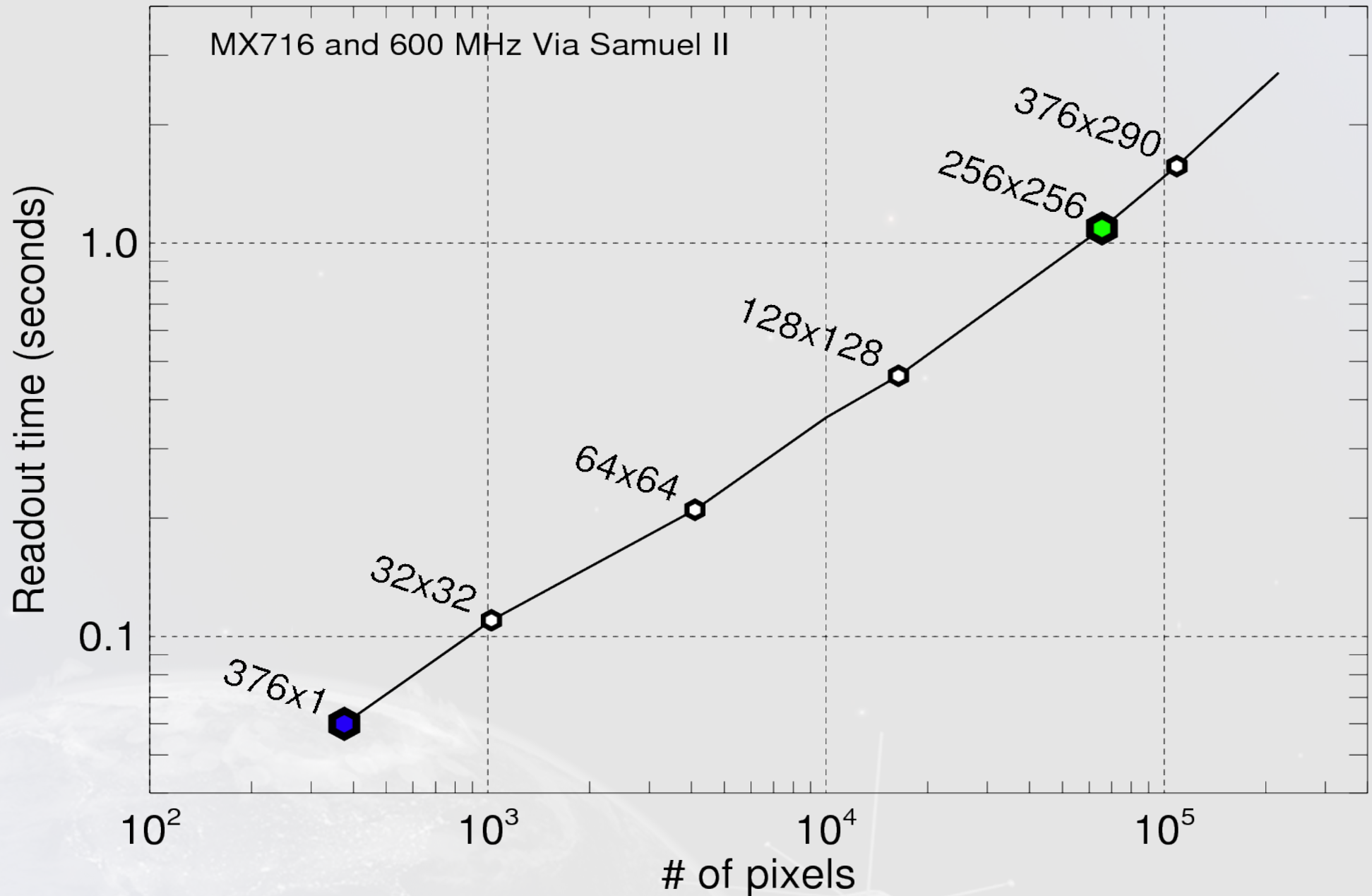
UCLA magnetometer data

- 2 samples/second @ 24-bits/sample in hourly binary (“RMD”) files
- 2.4 Mbytes/day of data stored at site and copied onto USB disks
- daily rsync pull from sites to UCalgary staging area
- UCLA pulls binary data from UCalgary for processing and validation, provides to UC-Berkeley
- Nearly complete UDP stream (0.5 kbps) with triple redundancy to UCalgary; echoed to UCLA and UAlberta

Image frames on CCD



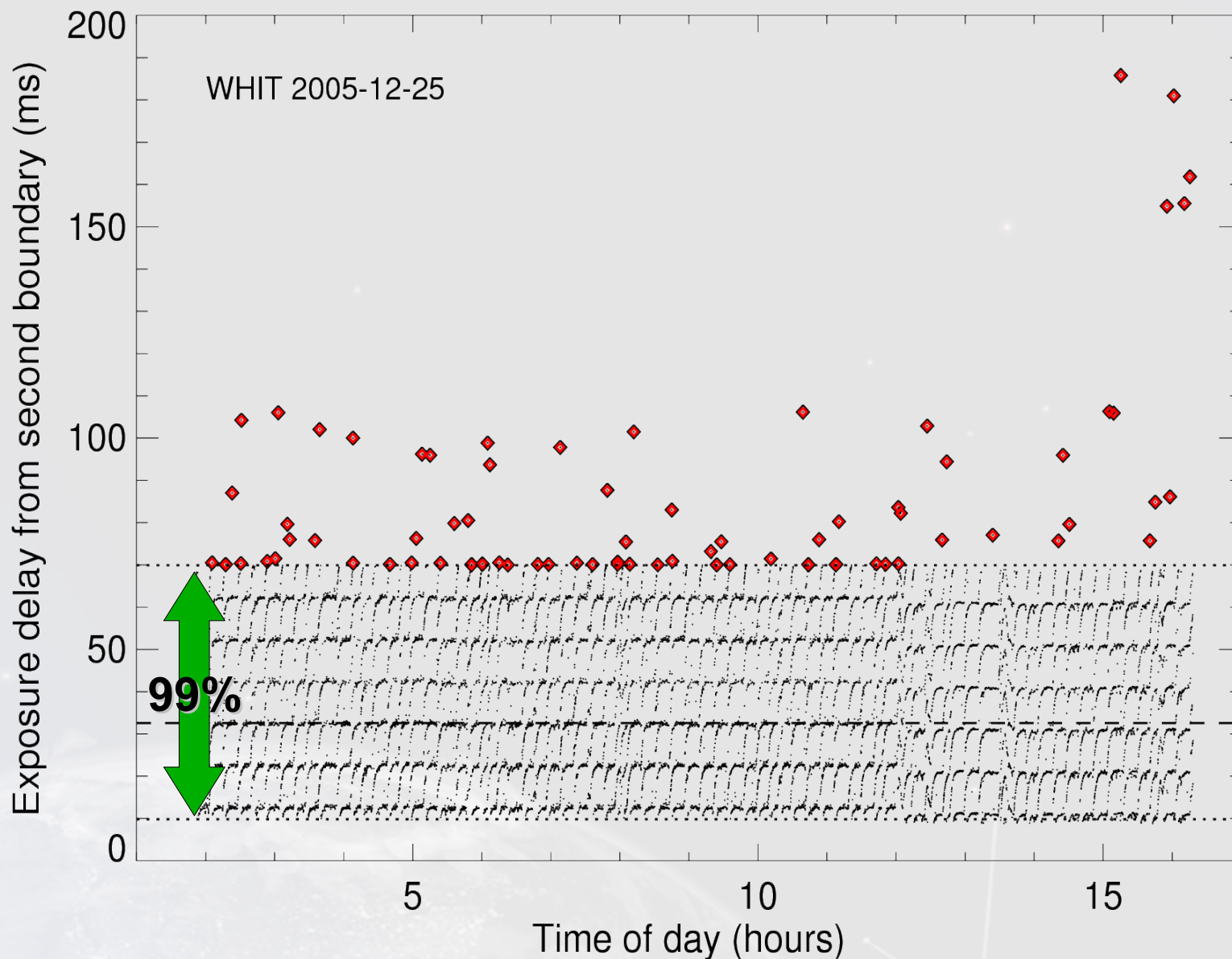
Frame readout times



Imager states

	Night	Day	Dusk
Solar zenith angle	$\geq 102^\circ$	$\leq 100^\circ$	$100^\circ - 102^\circ$
CCD power	on	off	on
Sun shade	open	closed	closed
Exposure mode	“full”@3s “row2”@3s	-	“wide”@60s

Exposure timing



Expose: full
 $1875 \pm 5 \text{ ms}$

Convert & store
 $40 \pm 5 \text{ ms}$

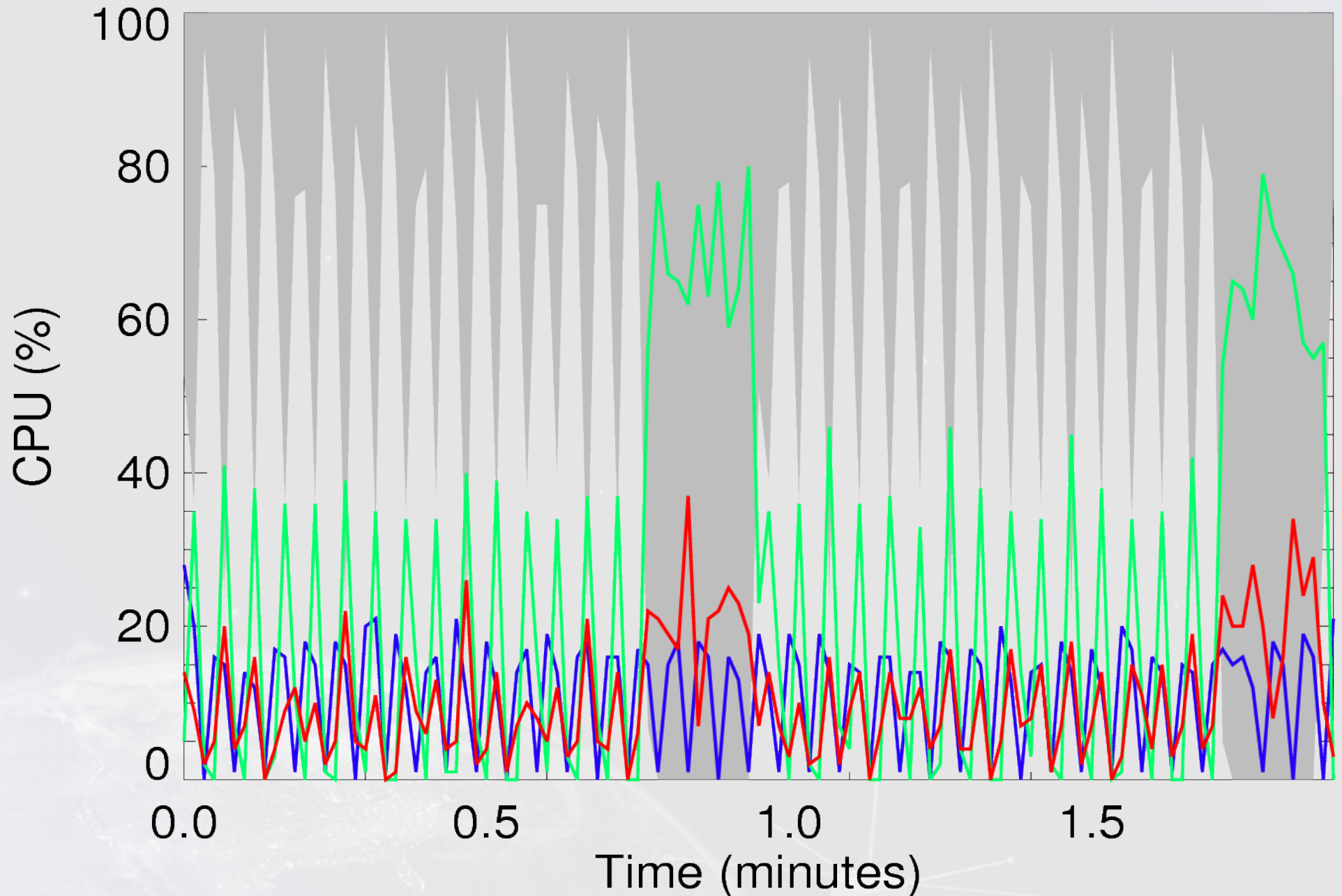
Wait
 $10 \pm 1 \text{ ms}$

Expose: row2
 $170 \pm 5 \text{ ms}$

Convert & store
 $25 \pm 5 \text{ ms}$

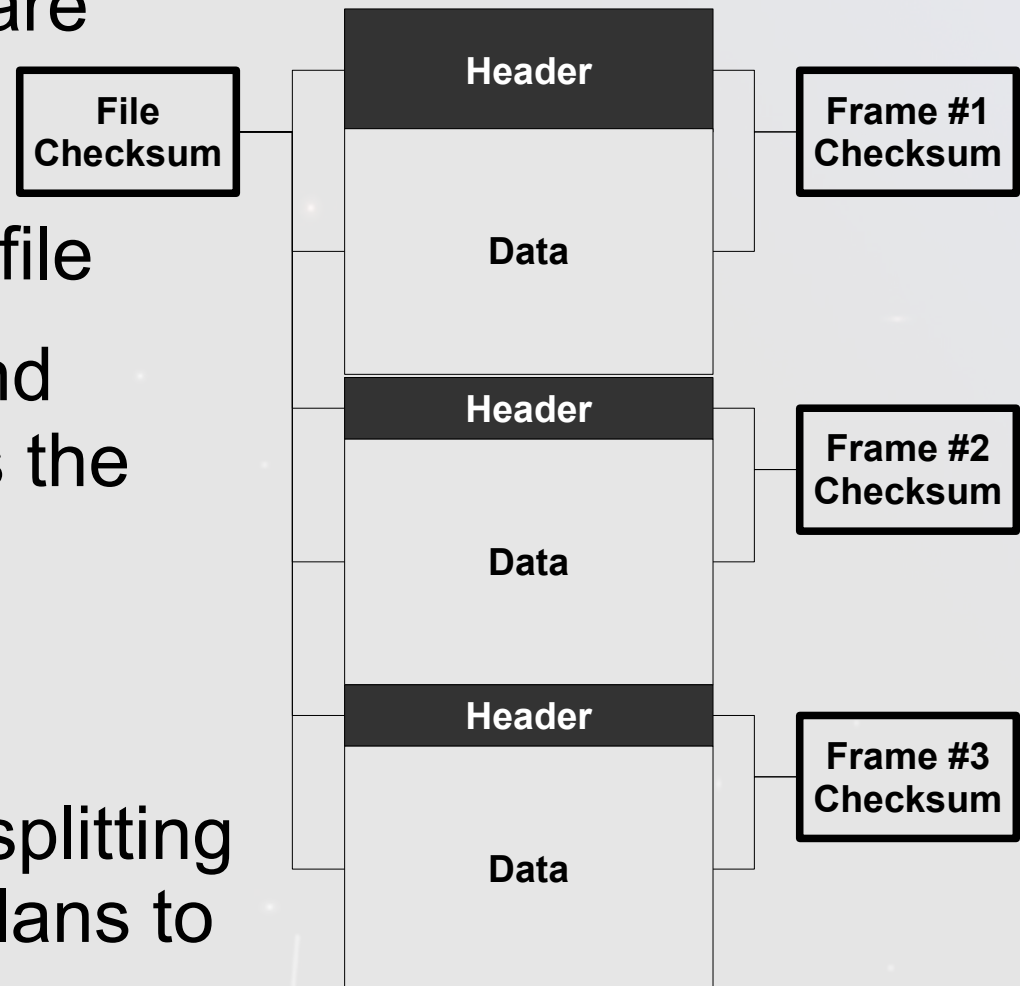
Wait
 $880 \pm 10 \text{ ms}$

Process scheduling

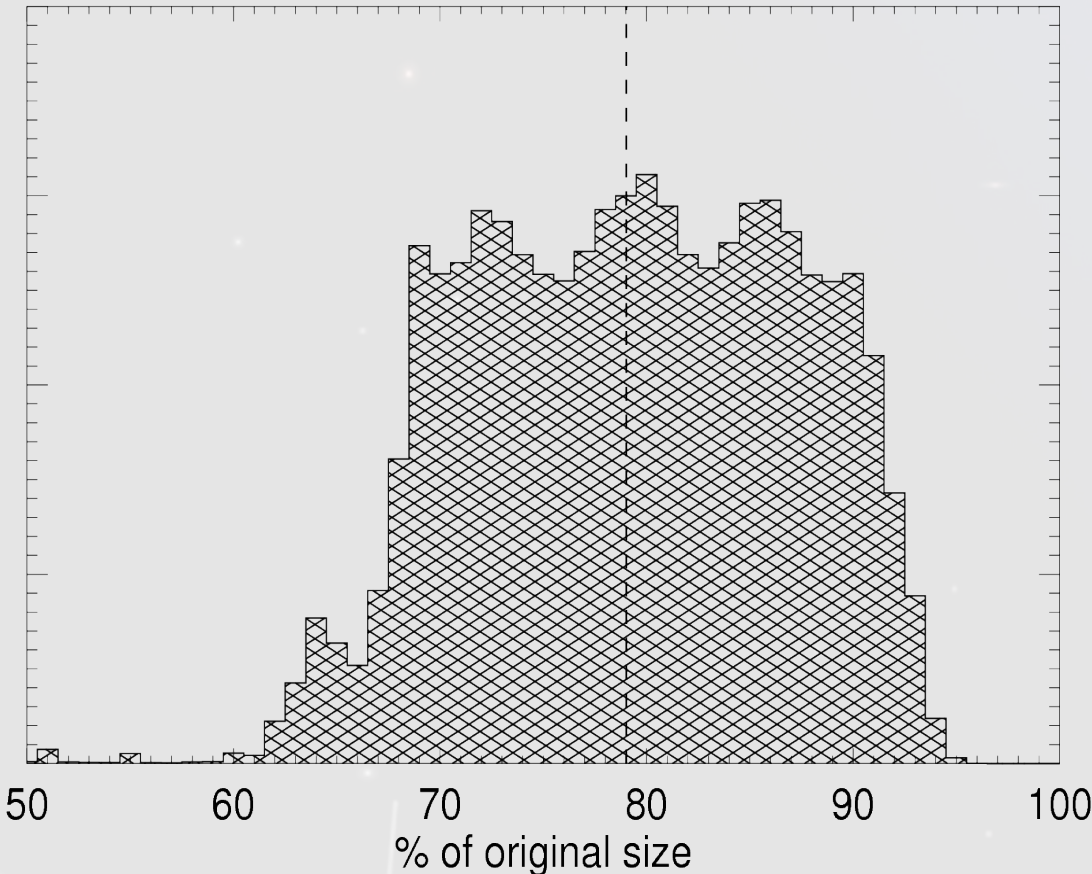


Checksums

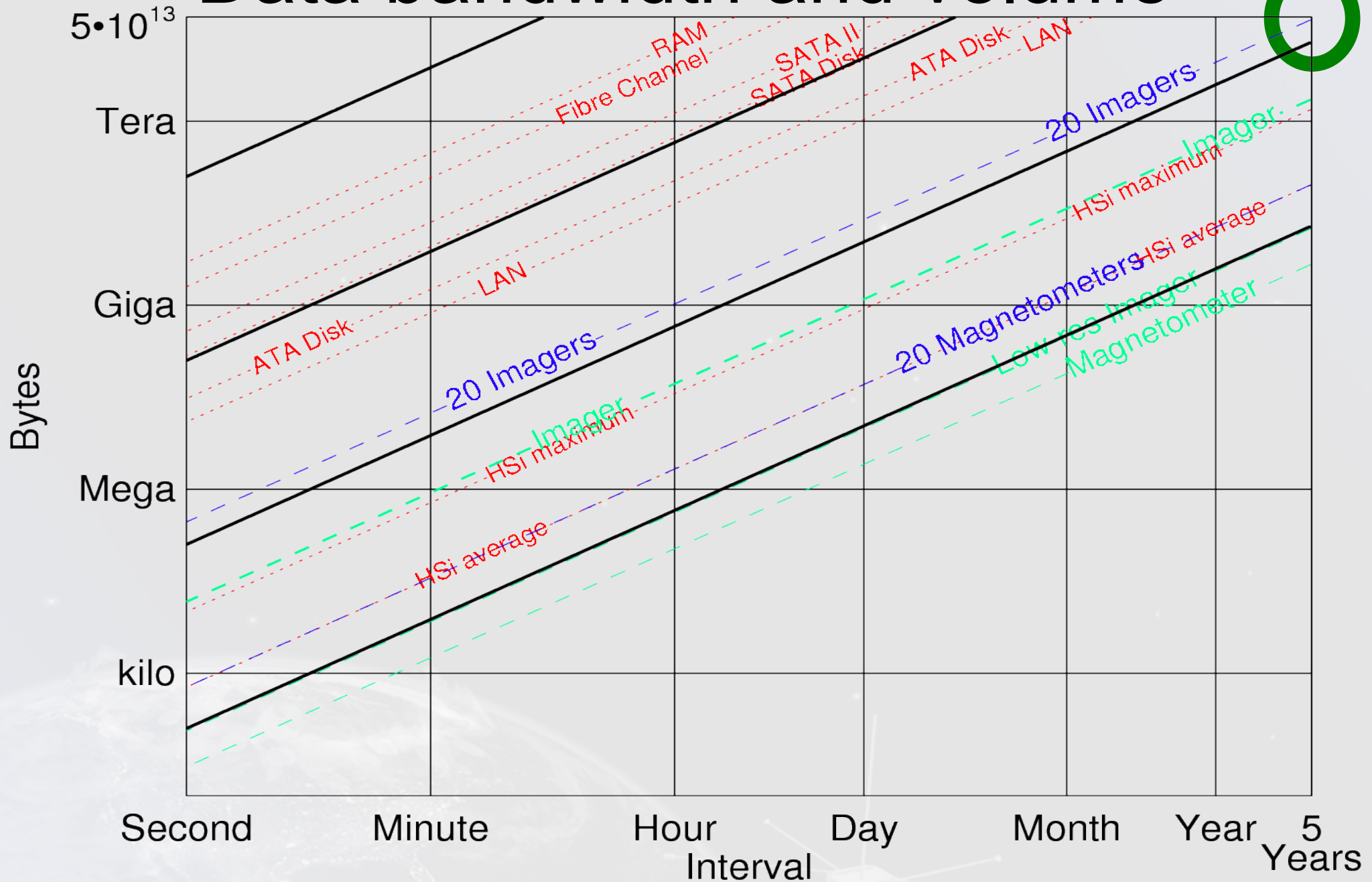
- 128-bit MD5 checksums are calculated and stored for each individual image frame and the aggregate file
- For $n=17.5$ million files and $d=2^{128}$ distinct checksums the collision probability is approximately 4×10^{-25} (birthday problem)
- Frame checksums allow splitting of files, but currently no plans to do so



Data file compression with “gzip”

- Average compression to 80% of original size reduces storage and bandwidth requirements
 - Also provides 32-bit checksum to detect data corruption
 - Compression at site requires 3 seconds for each file (20 frames)
 - Decompression is 10x faster
- 
- The histogram displays the frequency of compressed file sizes relative to the original size. The x-axis, labeled '% of original size', ranges from 50 to 100. The y-axis represents the frequency of files. The distribution is unimodal and centered around 80%, with a peak between 75% and 85%. A vertical dashed line is drawn at 80%.
- Alternate “bzip2” provides 60% reduction but 3x slower compression and 7x slower decompression

Data bandwidth and volume



Summary data

Several different summaries of imager data are produced in real-time at each site. These are all severely decimated in space and/or time for rapid downloading and examination.

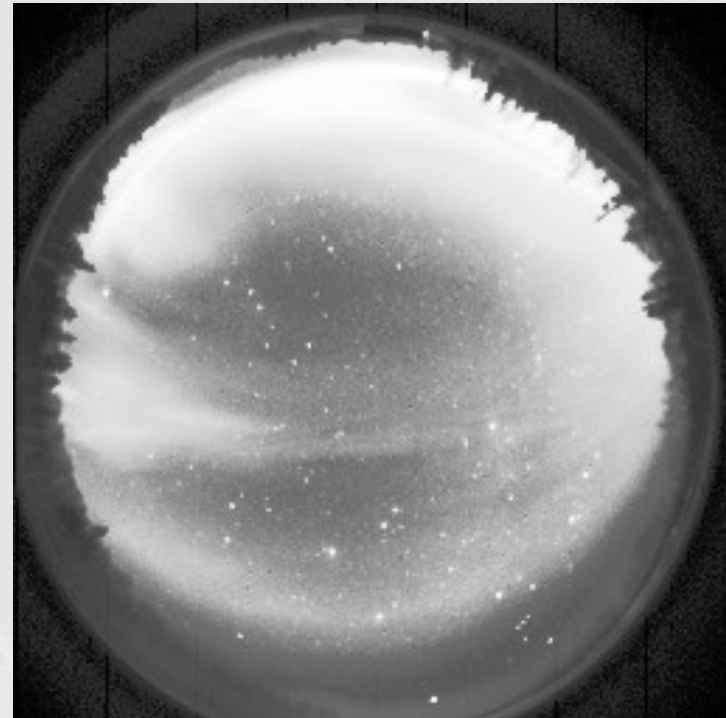
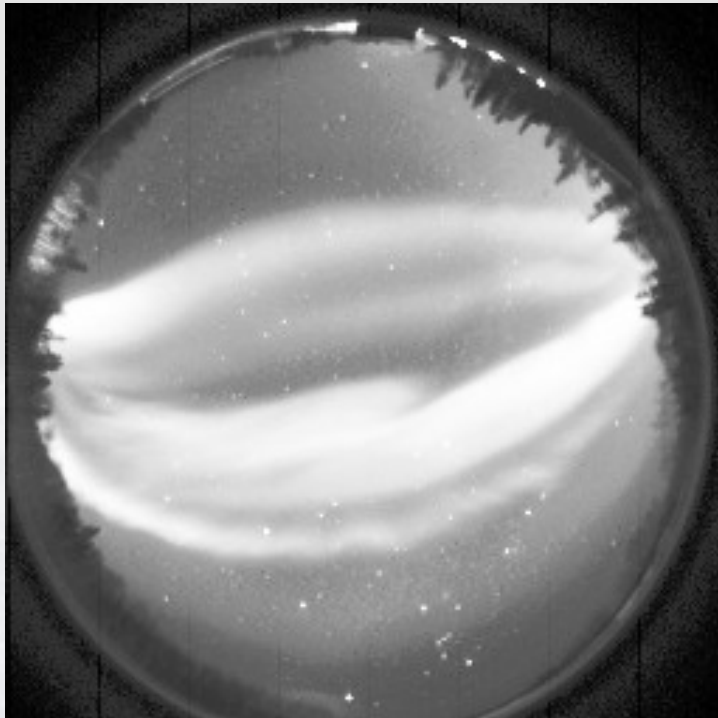
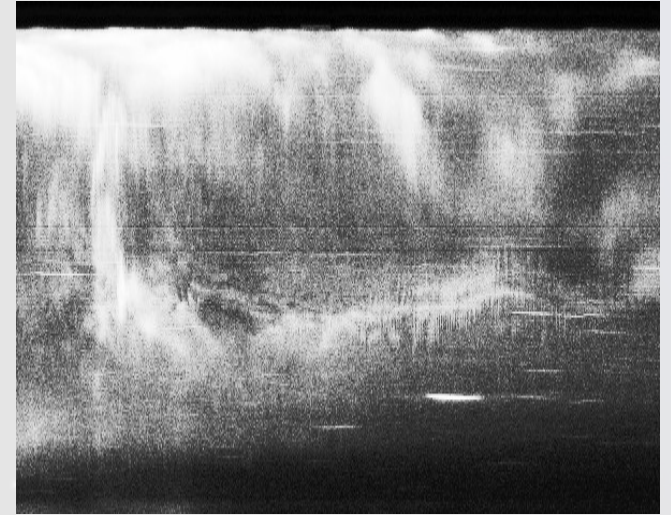
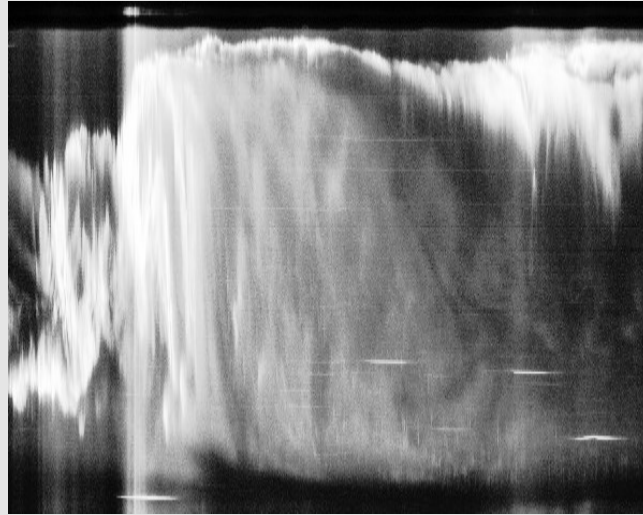
High time resolution is provided by “thumbnails” which are transmitted at a 6-second cadence with 3-second data recovered during the following day.

Spatial resolution in a north-south line is provided by keograms.

Averaging over 60-seconds (20 frames) provides high quality images that can be used for star detection.

Type	Width (pixels)	Height	Depth (bits)	Cadence (seconds)	Bandwidth (kbps)	Bandwidth (MB/hour)	Reduction
Full	256	256	16	3	341	150	
Thumb	32	32	8	3	2.67	1.17	128:1
Keogram (hourly)	1	256	16	6	0.67	0.29	512:1
Keogram (daily)	1	256	16	120	0.03	0.01	10240:1
Montage (hourly)	64	64	16	60	1.07	0.47	320:1
Montage (daily)	64	64	16	600	0.11	0.05	3200:1
Average (minute)	256	256	16	60	17.1	7.5	20:1

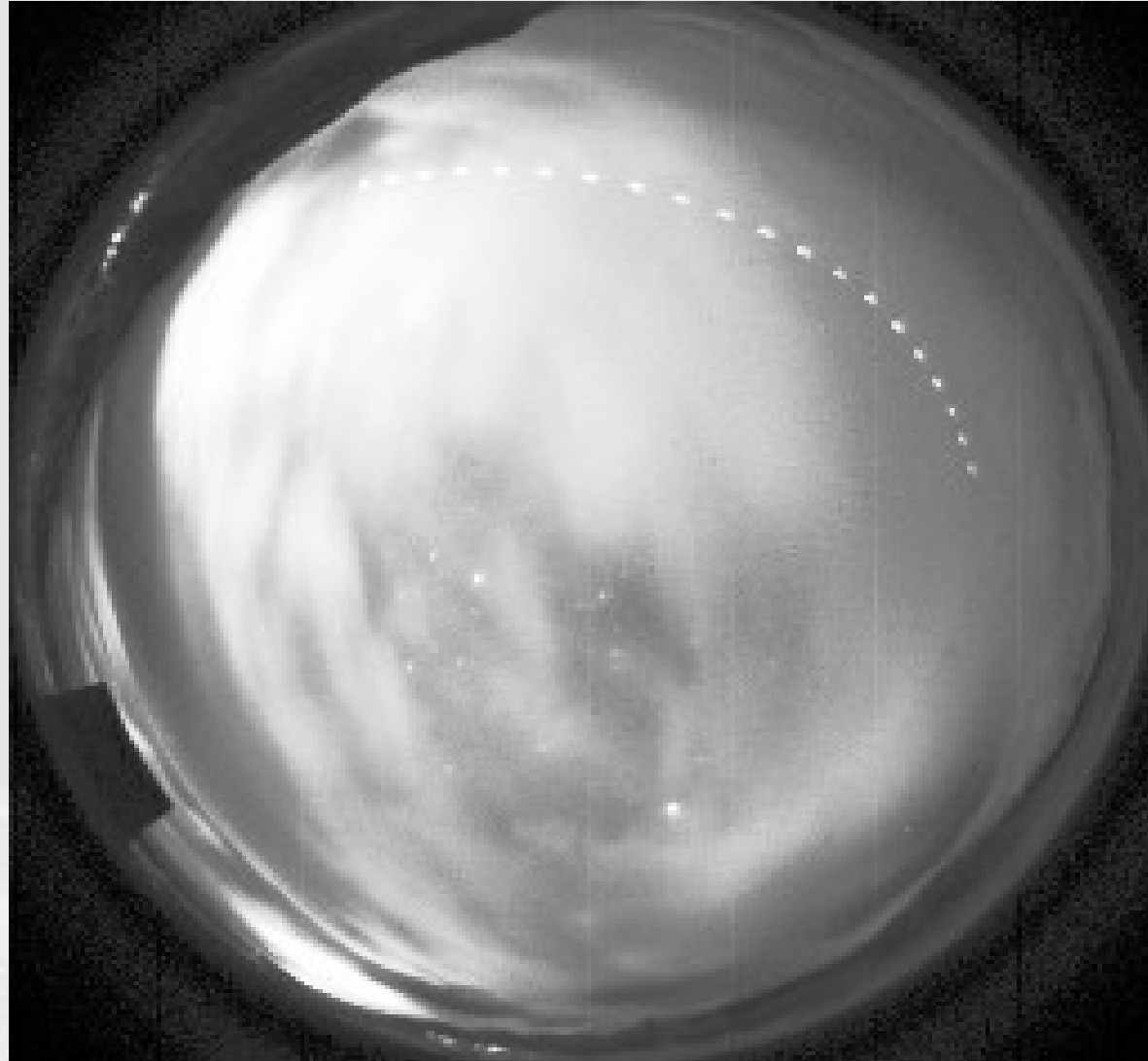
Hourly keograms



Hourly montage



Minute average (JPEG)



Health & housekeeping

- Unscheduled events are recorded remotely and transmitted to themis-rt using the standard syslog facility
- System parameters are polled every 5-minutes, stored remotely, and transmitted to themis-rt for display and reporting

themis_imager - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Telemetry Monitor Magnetometer Imager Syslog Ephemeris Help

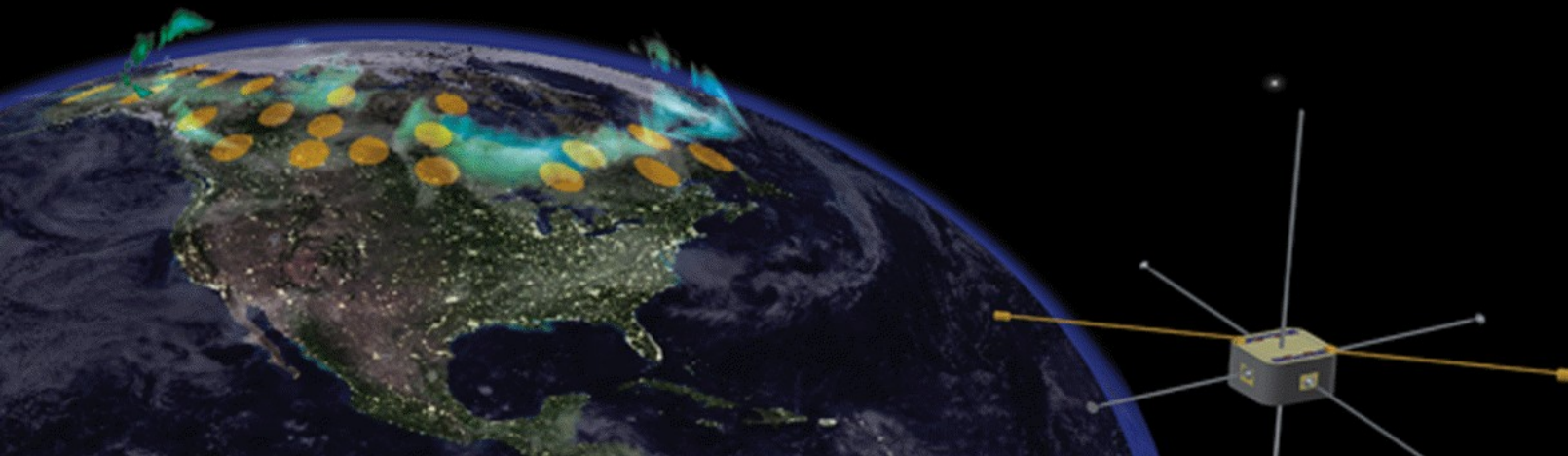
2006-03-23 03:39:28 UTC

Site	Imager	Age	Min	Mean	Max	St.Dev	Entropy	dEntropy	Sun angle
fykn	themis01	13:44:40	113	174.7	254	43.13	5.68	3.98	-1
gbo15	themis17	00:00:07	14	167.1	254	74.40	7.01	6.24	-1
berk	themis05	8:14:10.24	1	86.3	255	60.55	6.89	6.06	-1
mcgr	themis11	12:56:20	25	74.9	254	29.06	5.14	3.35	-1
whit	themis07	14:05:20	27	87.1	254	39.65	5.94	4.56	-1
inuv	themis08	14:53:11	100	160.1	254	32.03	4.68	3.07	-1
pgeo	themis03	14:37:20	43	85.5	254	33.40	5.70	4.05	-1
gbo13	themis13	8:14:10.24	31	204.8	254	71.48	4.42	3.97	-1
atha	themis02	00:00:07	20	64.9	254	29.09	4.54	3.16	-1
ekat	themis04	00:00:08	31	64.2	254	27.55	5.70	4.24	-1
fsmi	themis10	00:00:03	80	126.0	254	33.37	4.24	3.14	-1
tpas	themis06	00:00:08	187	214.9	254	14.29	2.59	1.64	-1
rank	themis09	00:00:08	58	126.8	254	48.90	5.22	4.31	-1

Time to create page: 41 milliseconds

Database

- Metadata
- Derived quantities
- Contextual information



Database Management

- PostgreSQL 7.4.11 with client/server architecture
- Currently hosted on themis-data.phys.ucalgary.ca

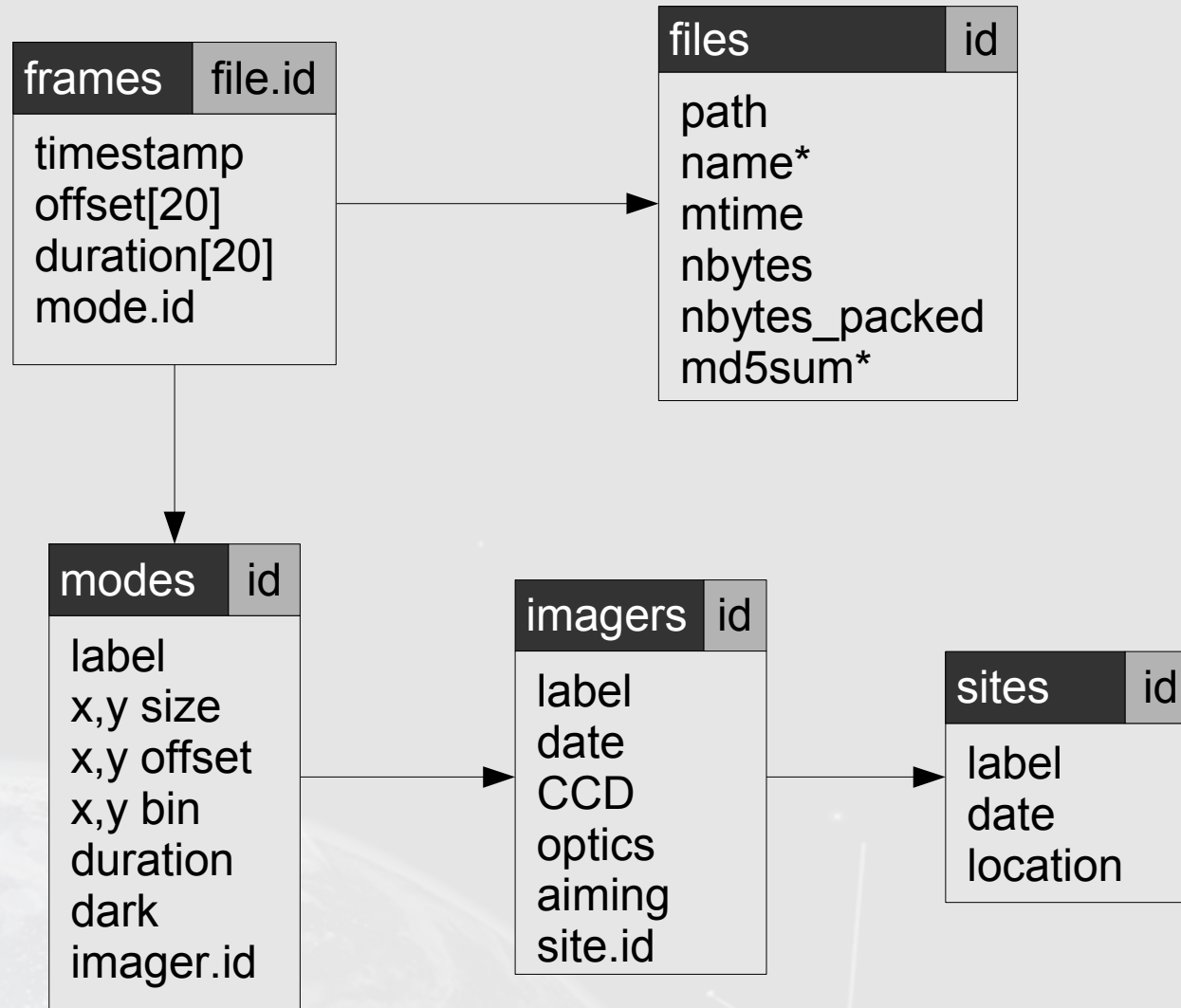
```
SELECT files.id
FROM frames, coordinates, ephemerides, statistics
WHERE statistics.maximum > 5000 AND
abs(frames.timestamp - coordinates.mlt_midnight)
    < '2 hours'
AND (ephemerides.lunar_elevation <= 20
    OR ephemerides.lunar_phase <= 0.2)
```

Image frame meta-data

```
#"Project unique ID" themis-gbo
#"Site unique ID" atha
#"Site name" Athabasca Geophysical Observatory
#"Imager unique ID" themis02
#"Ccd background" 2500
#"Optical projection" a1=1.5,a3=0.0,b2=0.0,b4=0.0
#"Optical center" x0=376, y0=142
#"Geodetic latitude" 54.72
#"Geodetic longitude" 246.72
#"Geodetic altitude" 676
#"Mode unique ID" row2
#"Mode sequence number" 291834
#"Exposure options" WIDTH=752 HEIGHT=1 XBIN=2 YBIN=1 XOFFSET=0
YOFFSET=2 MSEC=99
#"Image request start" 2006-02-26 11:00:02.014404 UTC
#"Image readout start" 2006-02-26 11:00:02.184967 UTC

#"Mode sequence number" 291835
#"Image request start" 2006-02-26 11:00:05.027394 UTC
#"Image readout start" 2006-02-26 11:00:05.198464 UTC
```

Core meta-data



Two stage meta-data entry

- 1) Register basic details (file path, name, modification time) in database: 1.5 milliseconds for each file.
 - 2) Complete (decompress, calculate checksum, extract meta-data header, identify exposure mode) and update record in database: at least 125 ms per file.
Multiple local or remote clients can work in parallel.
- Existing 750,000 files require 20 minutes for initial registration and 23 hours to complete.
 - Anticipated 17.5 million files would take 7 hours to register and 22 days to complete. *Five years of Moore's law will probably reduce this by a factor of 2^3 (8-fold)*

Derived quantities

statistics

file.id
frame_index

minimum
maximum
average
variance

textures

file.id
frame_index

contrast
correlation
entropy
...

stars

file.id
frame_index

polaris
zenith
quadrant
...

arcs

file.id
frame_index

latitude
width
intensity
...

detectors

file.id
frame_index

DC offset
dark count
...

Contextual information

ephemerides	file.id
zenith declination, right ascension solar azimuth, elevation lunar azimuth, elevation, phase	

coordinates	site.id
latitude longitude L-value MLT midnight meridian angle oval angle	

satellites	id
timestamp probe.id health GSM position regime T96 footprint	

T96	file.id
GSM x/y apex Bmin, GSM x/y/z	

geospace	id
timestamp IMF Bz, By dynamic P DsT AE Kp ...	

satellite_instruments	id
satellite.id electric field instrument fluxgate magnetometer search coil magnetometer electrostatic analyzer solid state telescope	

MHD models	id
timestamp ...	

Database status and plans

- Core file and frame information in database
 - 20 ms to retrieve an single frame
 - interface for basic processing and transformation
- Imager and site metadata partially integrated, should be done soon
- Preliminary statistics and ephemerides by late spring
- Improved interface for metadata by late spring

Stream	Frame				Compress	Total Bytes	Minute		Hour		Day		Month		Year 20 sites (GB)	Hourly	Bandwidth (kbps)			
	Bits/pixel	Width	Height	Header			#	KB	#	MB	#	Ave	Max	Ave			Max	Monthly (ave)		Monthly (max)
System status																				
1 Monitor						24272			1	0.02		0.56	0.56	16.7	16.7	3.96	0.05	0.05	0.05	1
1 Syslog						500000					1	0.48	0.48	11.4	11.4	3.40	0.05	0.05	0.05	1
Fluxgate at site and two remote targets																				
0 .RMD						106652			1	0.10		2.44	2.44	73	73	17.4	0.23	0.23	0.23	0
1 UDP: UCLA						234325			1	0.22		5.36	5.36	161	161	38.2	0.51	0.51	0.51	1
1 UDP: UCalgary						234325			1	0.22		5.36	5.36	161	161	38.2	0.51	0.51	0.51	1
Full raw frame every 3 seconds, bzip remotely, gzip centrally																				
0 PGM	16	256	256	380	1.00	131452	20	2567	1200	150		1203	2407	35467	70748	8579	342	112	224	0
.GZ					0.80	105162	20	2054	1200	120		963	1926	28374	56598	6864	274	90	179	
.BZ2					0.60	78871	20	1540	1200	90		722	1444	21280	42449	5148	205	67	134	
64 byte summary line for each image frame																				
1 Index						76800			1	0.07		0.6	1.2	17	34	4.2	0.17	0.05	0.11	1
Raw dark row every 3 seconds, bzip remotely, gzip centrally																				
0 PGM	16	376	1	380	1.00	1132			1200	1.30		10.4	20.7	305	609	73.9	2.95	0.97	1.93	0
.GZ					0.80	906			1200	1.04		8.3	16.6	244	487	59.1	2.36	0.77	1.54	
.BZ2					0.65	589			1200	0.67		5.4	10.8	159	317	38.4	1.53	0.5	1	
PACE thumbnail produced for each raw frame (3-second cadence), gzip hourly, rsync pull daily during non-imaging period																				
2 PGM	8	32	32	90	1.00	1114			1200	1.27		10.2	20.4	301	600	72.7	2.90	0.95	1.89	
.GZ					0.60	668			1200	0.76		6.12	12.24	180	360	43.6	1.74	0.57	1.14	
1 Rsync -z					0.50	334			1200	0.38		3.1	6.1	90	180	21.8	0.87	0.28	0.57	
Transmit alternate frames as PNG (6-second cadence)																				
1 PNG	8	32	32	90	0.85	947			600	0.54		4.33	8.67	128	255	30.9	1.23	0.4	0.81	1
2 CDF	8	32	32	90	0.80	891			600	0.51		4.08	8.2	120	240	29.1	1.16	0.38	0.76	2
Average of 20 raw frames once per minute; pull single JPEG each hour for operations																				
PGM	16	256	256	930	1.00	132002			60	7.55		60	121	1781	3552	431	17.2	5.63	11.2	
2 .BZ2					0.65	85801			60	4.91		39	79	1158	2309	280	11.2	3.66	7.3	2
2 CDF	16	256	256	930	0.80	105602			30	3.02		24	48	712	1421	172	6.9	2.25	4.49	2
1 JPG	8	256	256	930	0.75	49850			1	0.05		0.38	0.76	11.2	22	2.71	0.11	0.04	0.07	1
Hourly montage of 64x64 thumbs at 1 minute cadence, pull JPEG every hour																				
2 PNG	16	640	384		0.75	368640			1	0.35		2.81	5.63	82.9	165	20.1	0.80	0.26	0.52	2
1 JPG	8	640	384		0.25	61440			1	0.06		0.47	0.94	13.8	28	3.3	0.13	0.04	0.09	1
Hourly keogram at 6 second resolution, pull JPEG every hour																				
2 PNG	16	600	256		0.75	230400			1	0.22		1.76	3.52	51.8	103	12.5	0.50	0.16	0.33	2
1 JPG	8	600	256		0.25	38400			1	0.04		0.29	0.59	8.6	17	2.09	0.08	0.03	0.05	1
Daily montage of 64x64 thumbs at 10 minute cadence																				
2 JPG	8	768	768		0.25	147456					1	0.14	0.14	4.2	4.2	1.00	0.01	0.01	0.01	2
2 PNG	16	768	768		0.75	884736					1	0.84	0.84	25.3	25.3	6.02	0.08	0.08	0.08	2
Daily keogram at 1 minute resolution																				
2 JPG	8	1440	256		0.25	92160					1	0.09	0.09	2.6	2.6	0.63	0.01	0.01	0.01	2
2 PNG	16	1440	256		0.75	552960					1	0.53	0.53	15.8	15.8	3.76	0.05	0.05	0.05	2
Movies in hourly and daily chunks																				
2 AVI	8	256	256						1	0.5		12	12	360	360	86	1.14			2
2 AVI	8	256	256								1	1	1	30	30	7.13	0.09			2
All real-time telemetry data																				
HSi limits																				
All summary imager data																				
All high-resolution science data																				