170,000 watts of power, 120 tons of cooling

for

17,308,672 control/monitor bits

in

1,473,536 registers

in

24,832 FPGAs

in

256 boards

in

16 racks

in

1 room

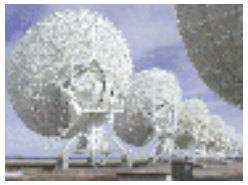
Behind Closed Doors

Operator's Overview of the EVLA
WIDAR Correlator

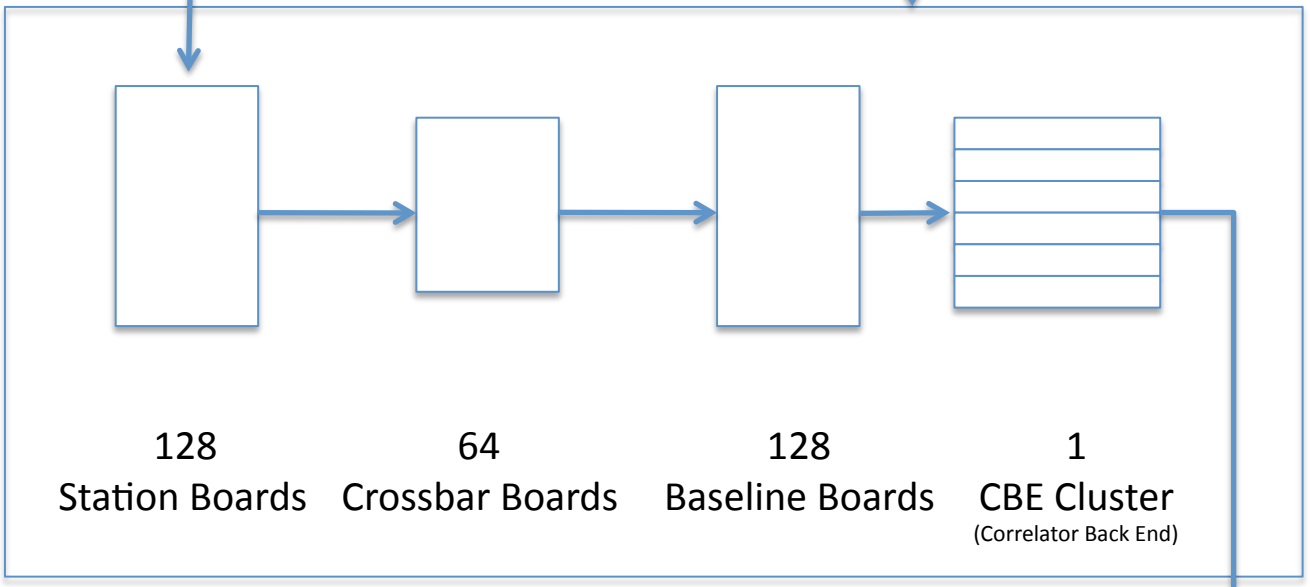
Wideband **I**nterferometric **D**igital **AR**chitecture

- Concept began about 11 years ago (1998)
- NRAO's Ray Escoffier was busy with ALMA's correlator - could not commit to EVLA's also
- Canada needed an 'in' into ALMA (a share in its use)
- Rick Perley contacted Peter Dewdney of DRAO, Penticton, BC, Canada; they go back to college days
- Peter, after much effort, got funding ~$25M by the Canadian Government to supply the correlator in return for access to NRAO antennas including our share in ALMA.
- Primary architect of WIDAR is Brent Carlson of DRAO
- Hardware was built in Canada
- Software is being developed as a joint effort between NRAO and DRAO.

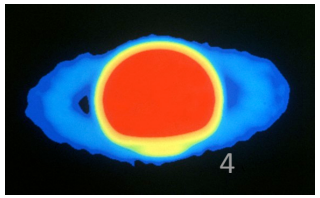
WIDAR In a Nutshell

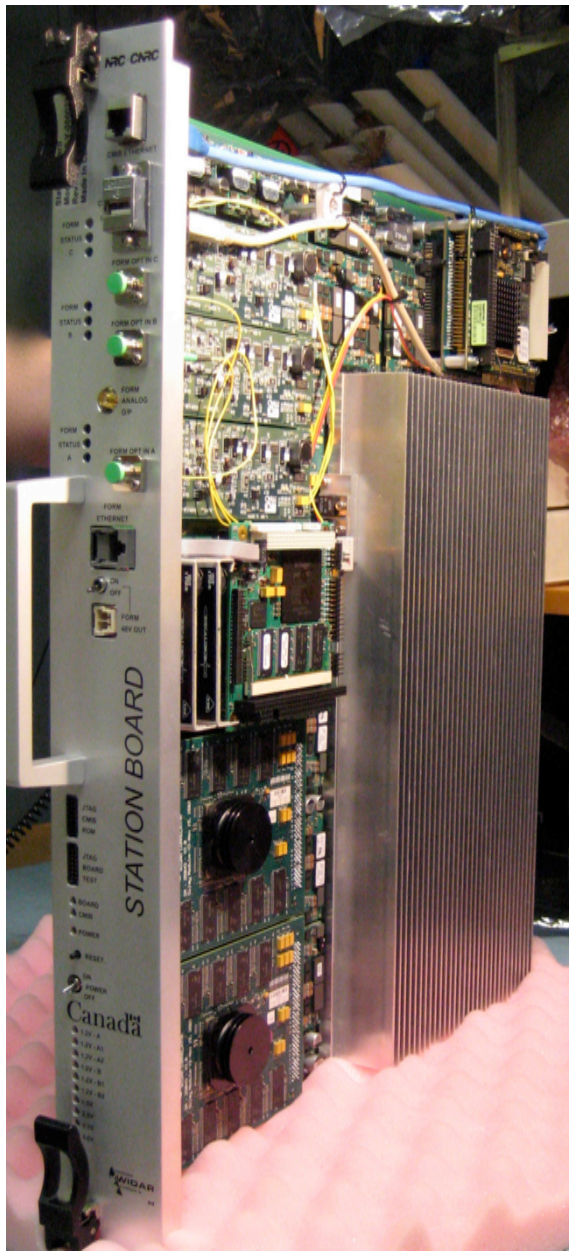


170,000 Watts

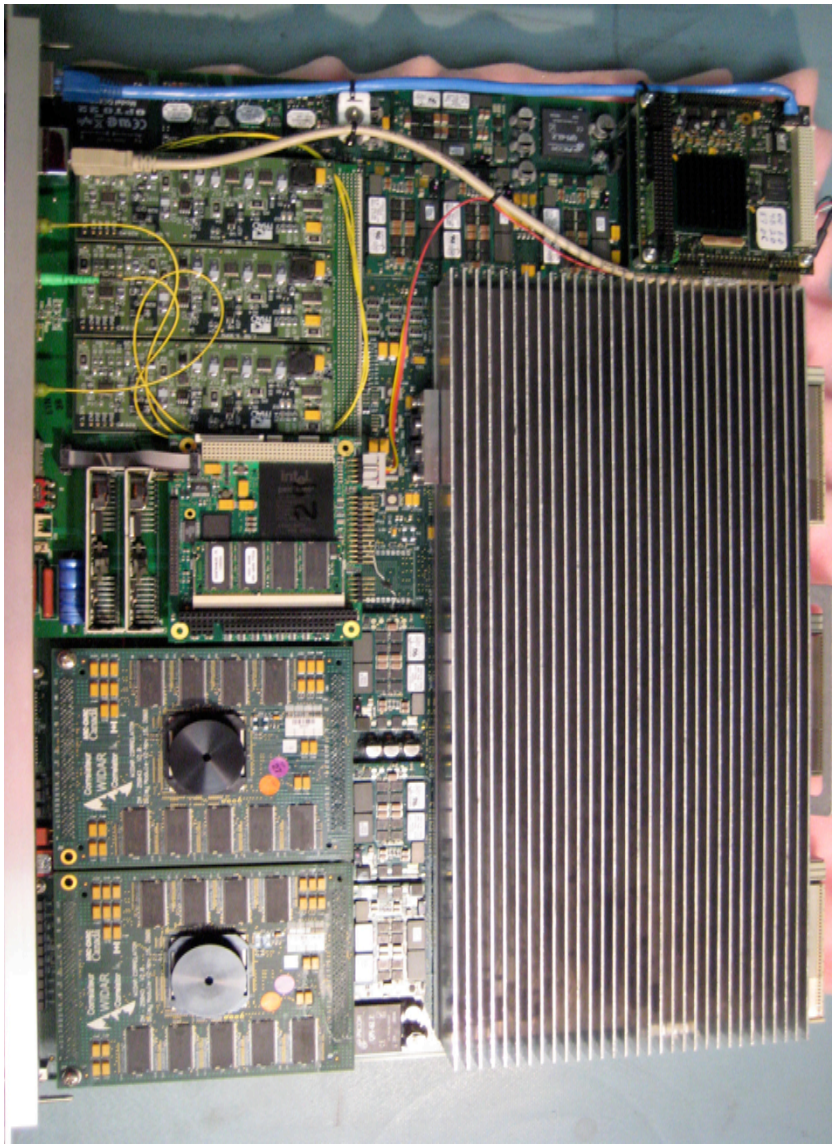


hmmm



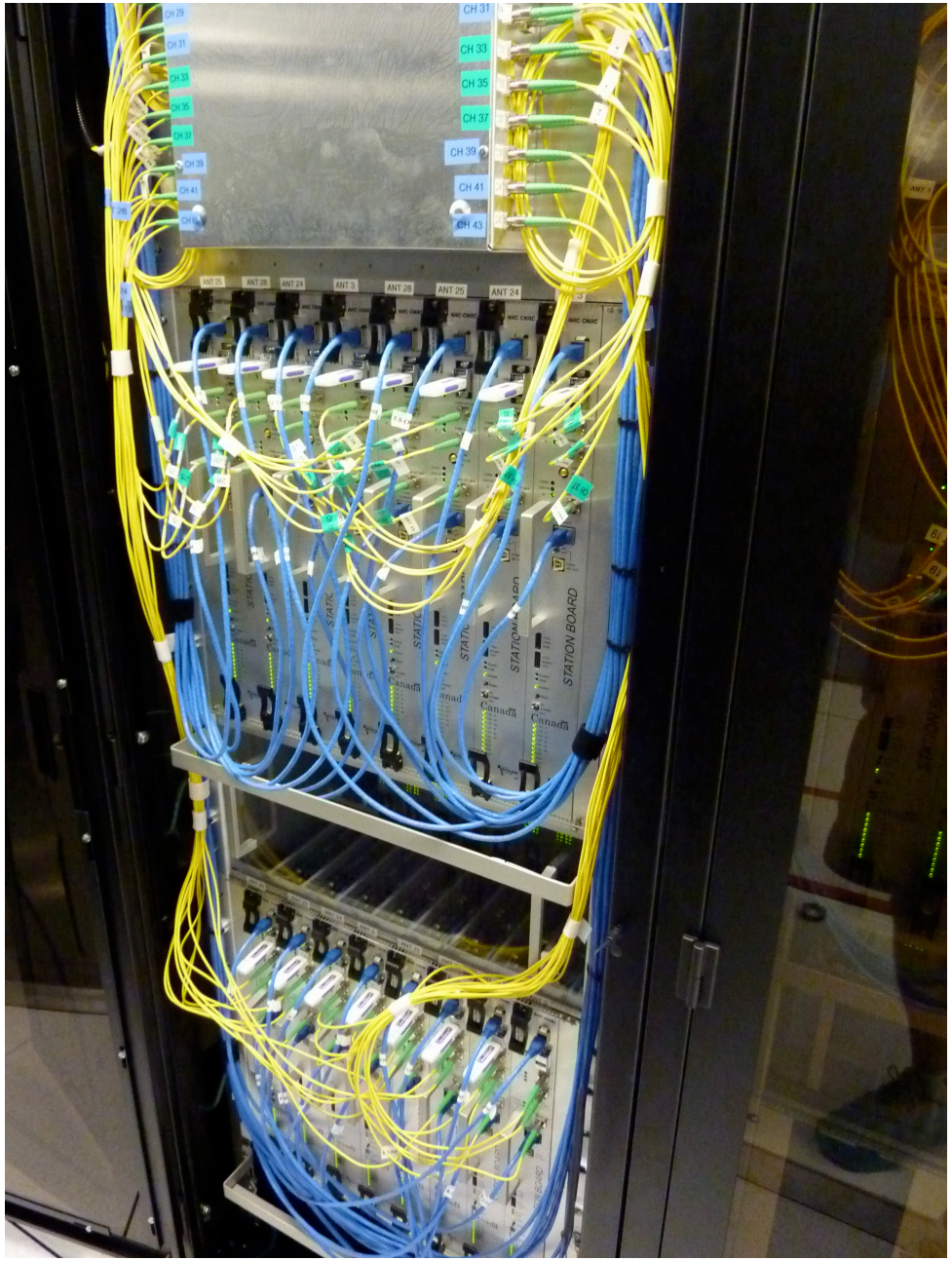


Station Board

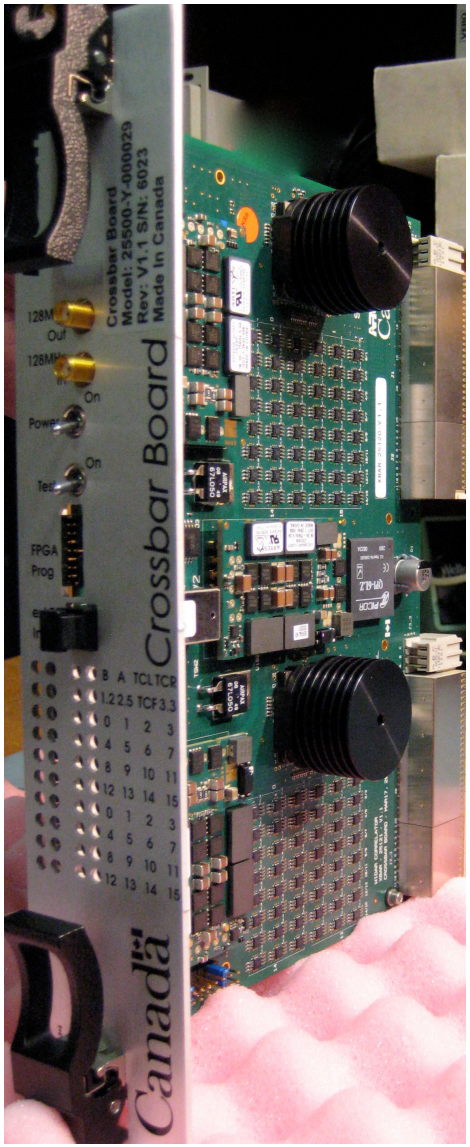


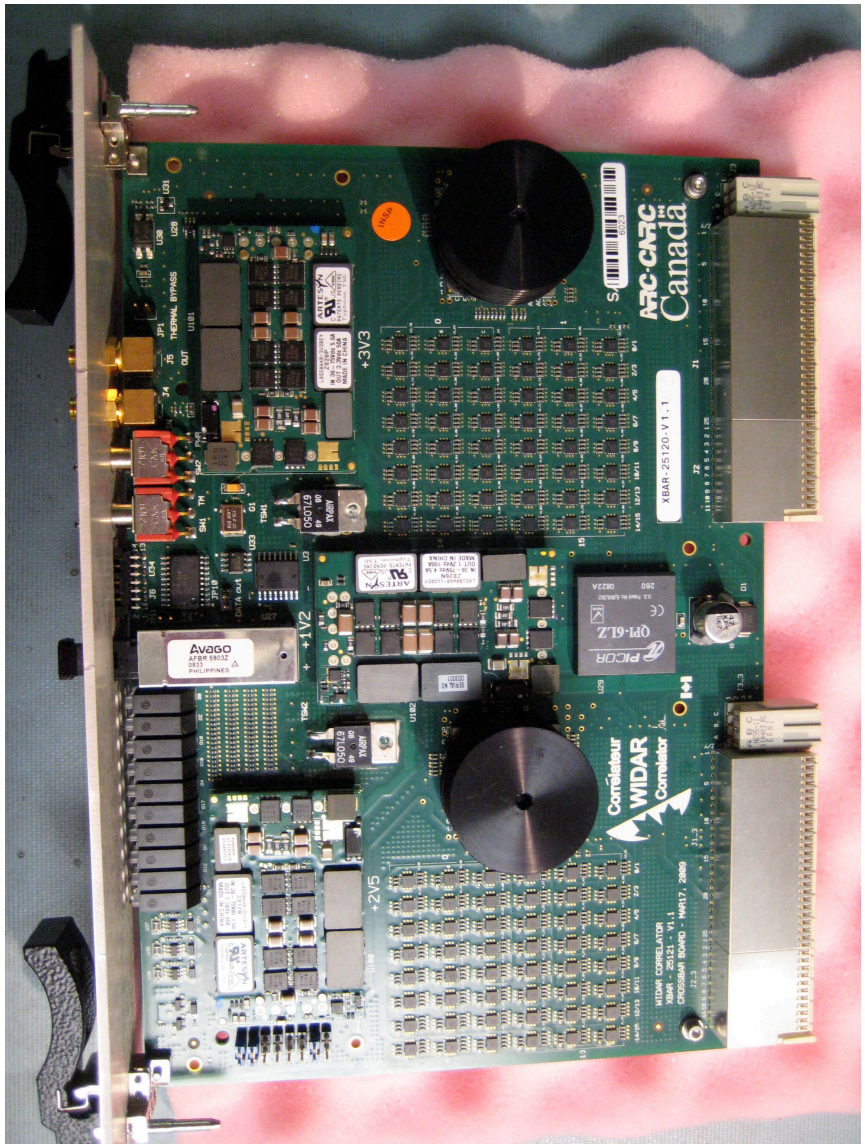
15” x 19” • 28 layers • 1km of traces • 8,000 components • 50,000 solder points • $20k • x 128 boards

Two Station Racks each with 4 Station Boards and 1 Crossbar Board



Fully populated
Station Rack

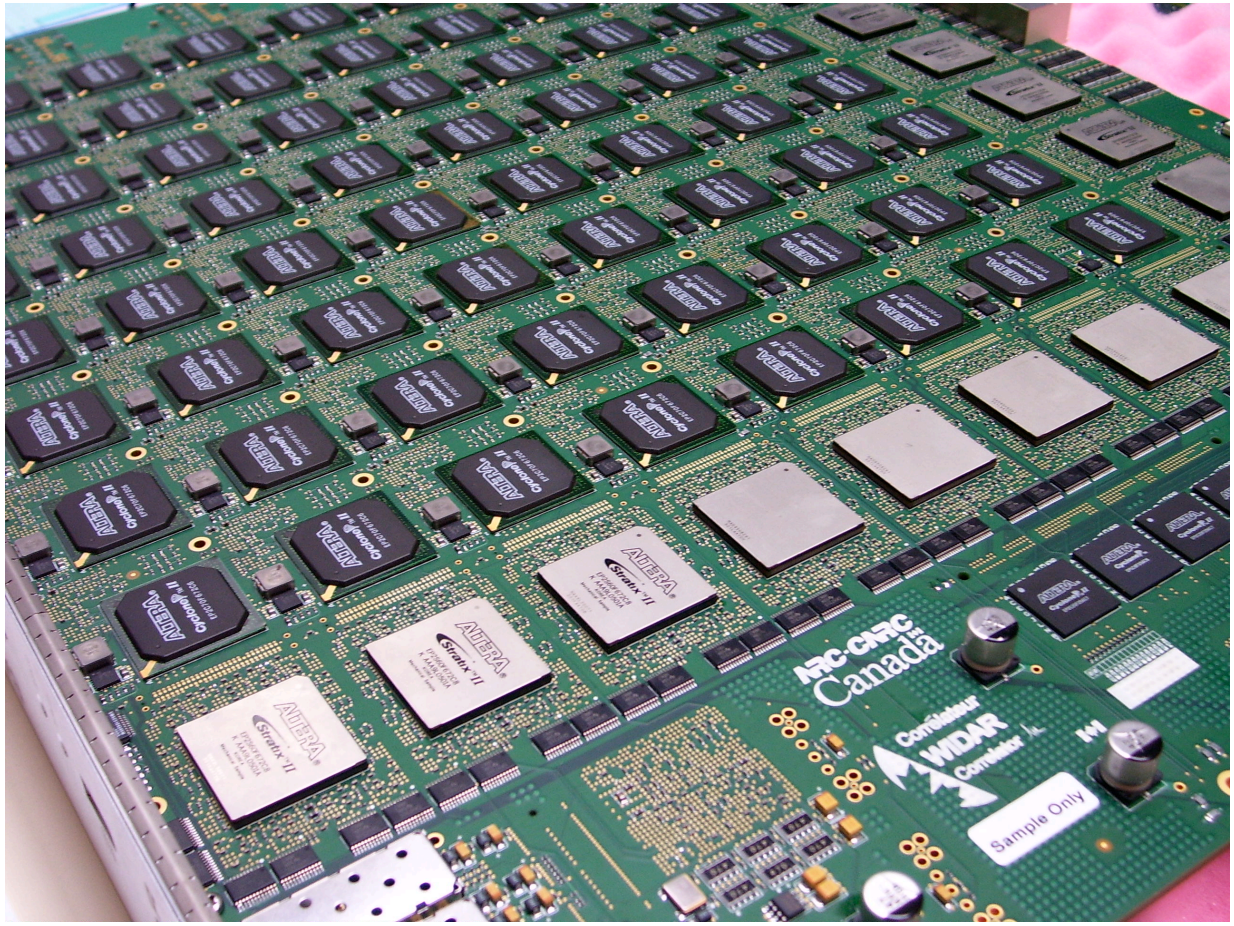


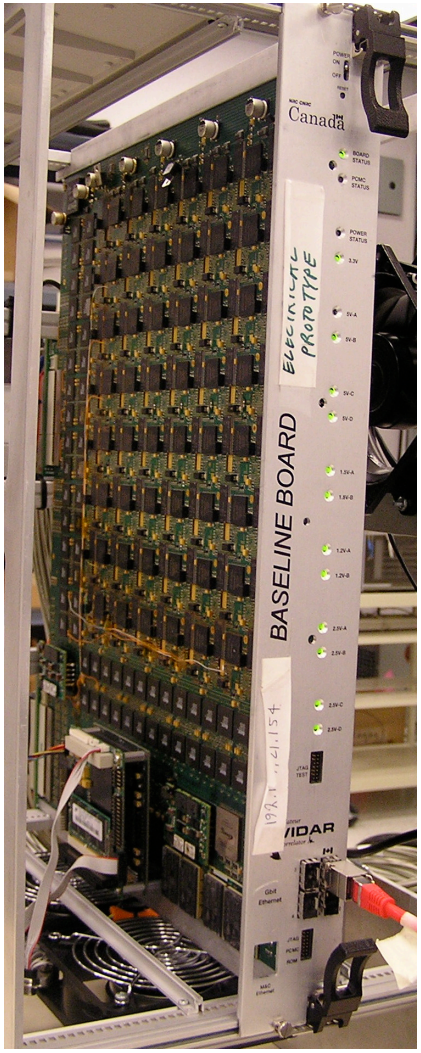


64 Crossbar Boards route Station Boards Outputs to Baseline Board inputs

8 Crossbar Boards reside in the ‘6U’ crate between the Top and Bottom Station Board crates in each Station Rack.

Baseline Board

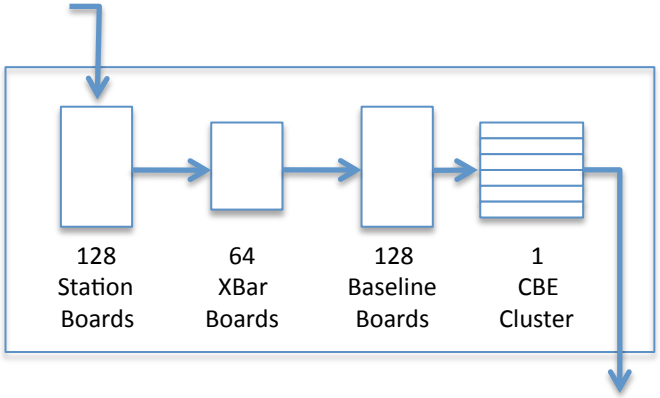




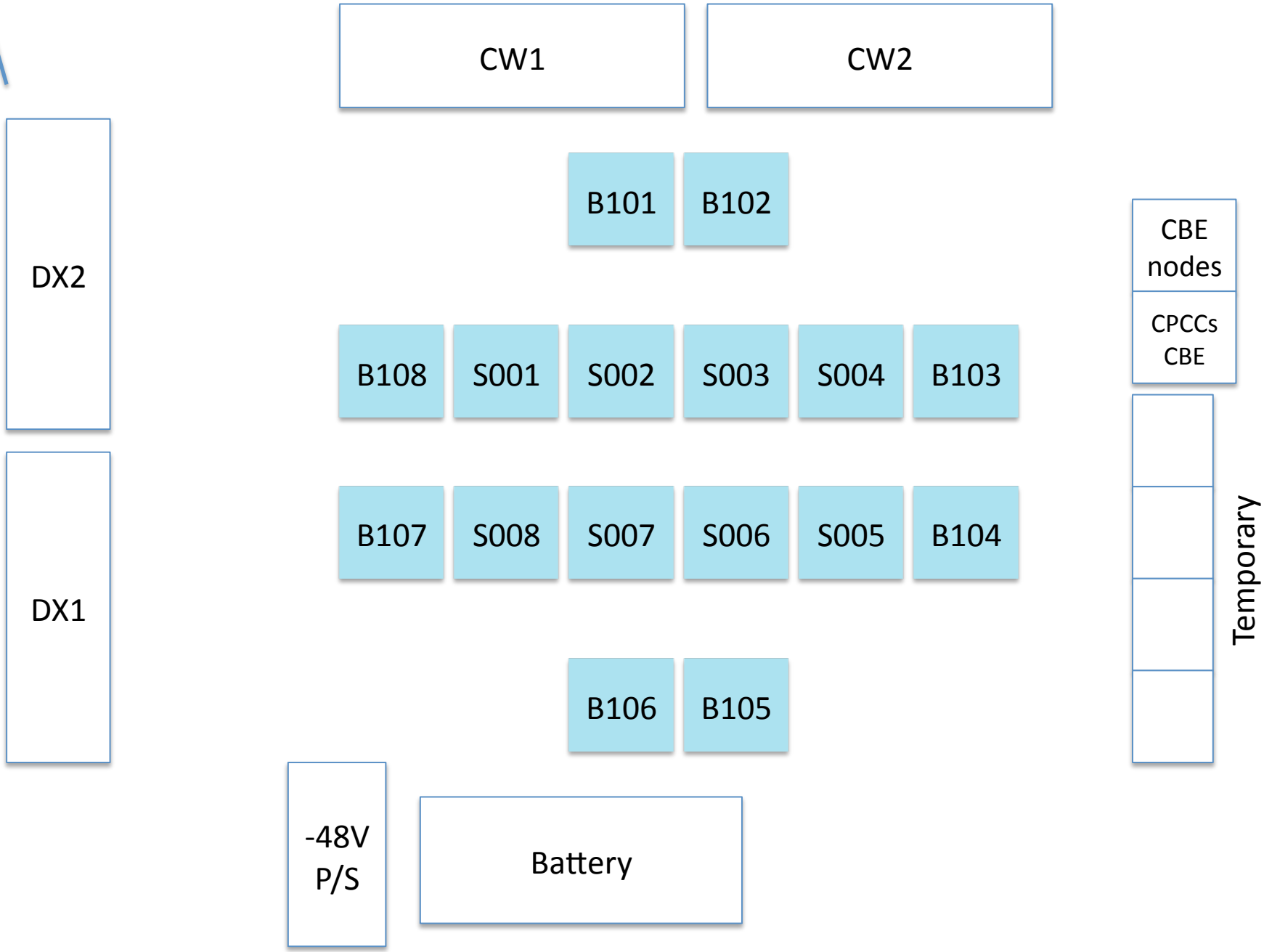
- 15” x 19”, • 28-layers, • 2km of traces, • 11,000 components
- 100,000 solder joints, • 32 Gbits of RAM, • $30k • 128 boards

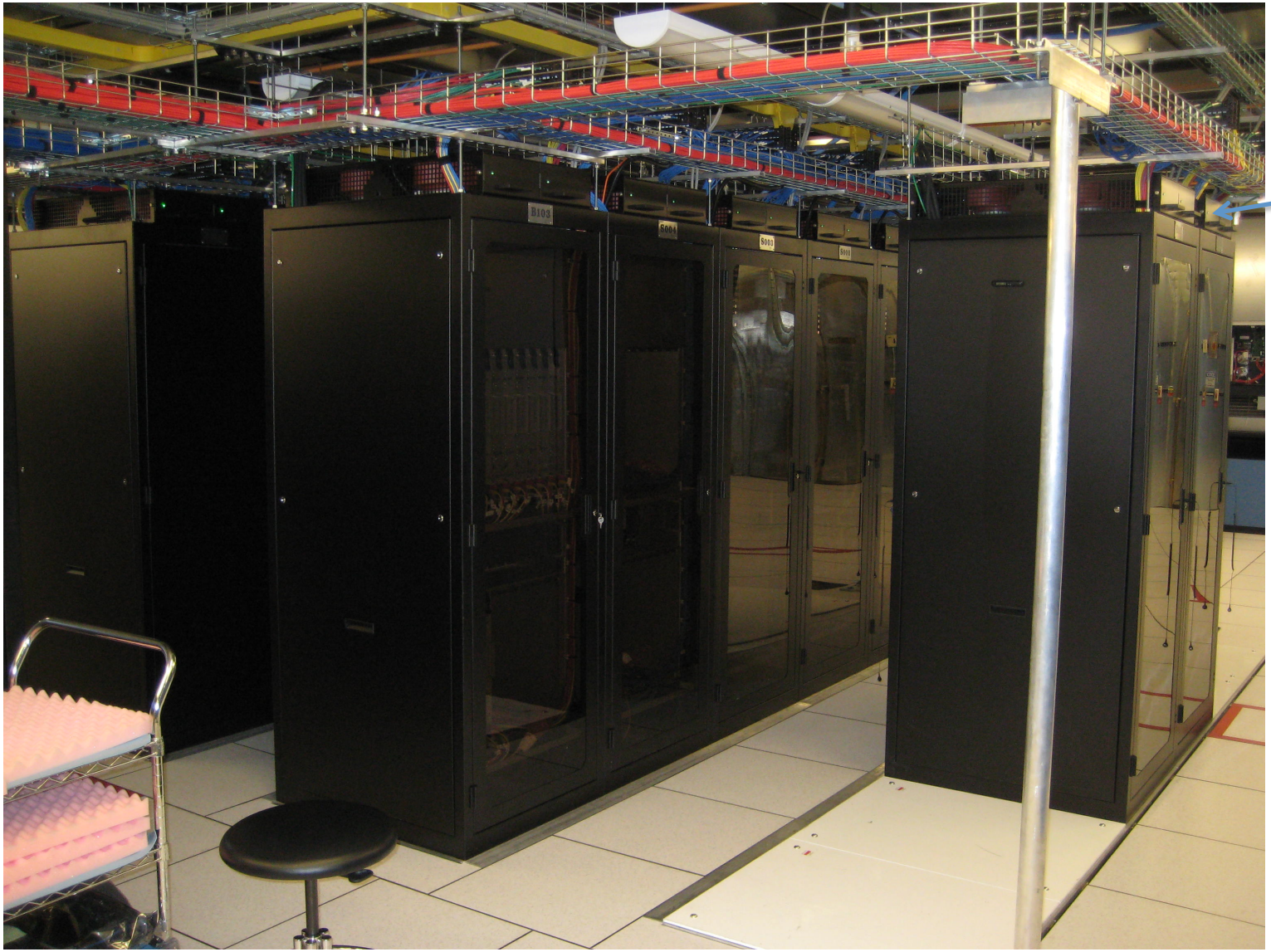
Two fully populated Baseline Racks

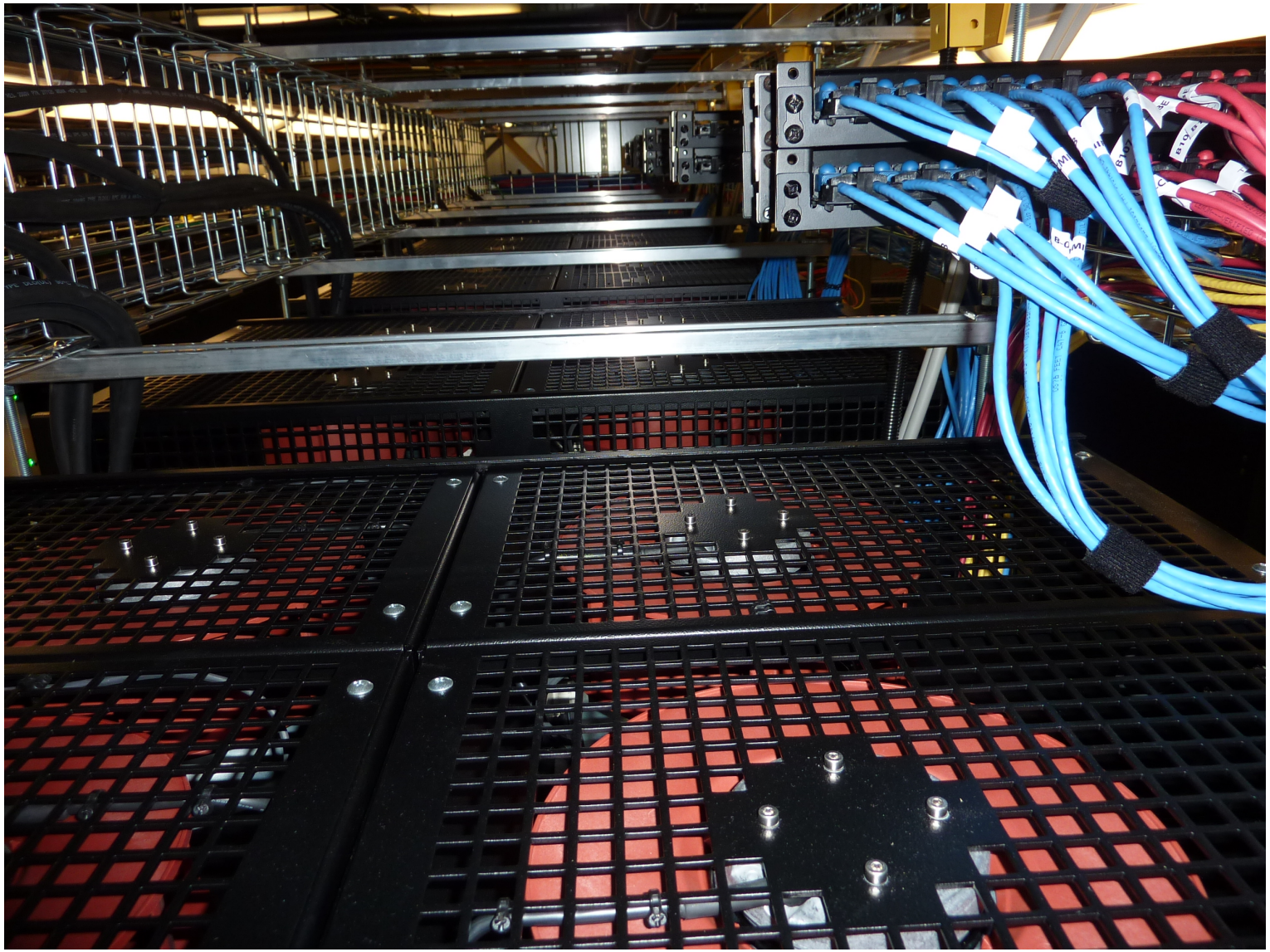
- Boards are identified by a unique **rack-crate-slot** name.



- The Station, Crossbar and Baseline Boards are housed in 16 Racks
- Eight Station Racks numbered S001 – S008
  - Each containing 16 Station and 8 Crossbar boards
- Eight Baseline Racks numbered B101 – B108
  - Each containing 16 Baseline Boards (no Crossbar Boards)


CW1
CW2
DX2
B101
B102
CBE nodes
CPCCs CBE
B108
S001
S002
S003
S004
B103
DX1
B107
S008
S007
S006
S005
B104
Temporary
B106
B105
-48V P/S
Battery


Fans

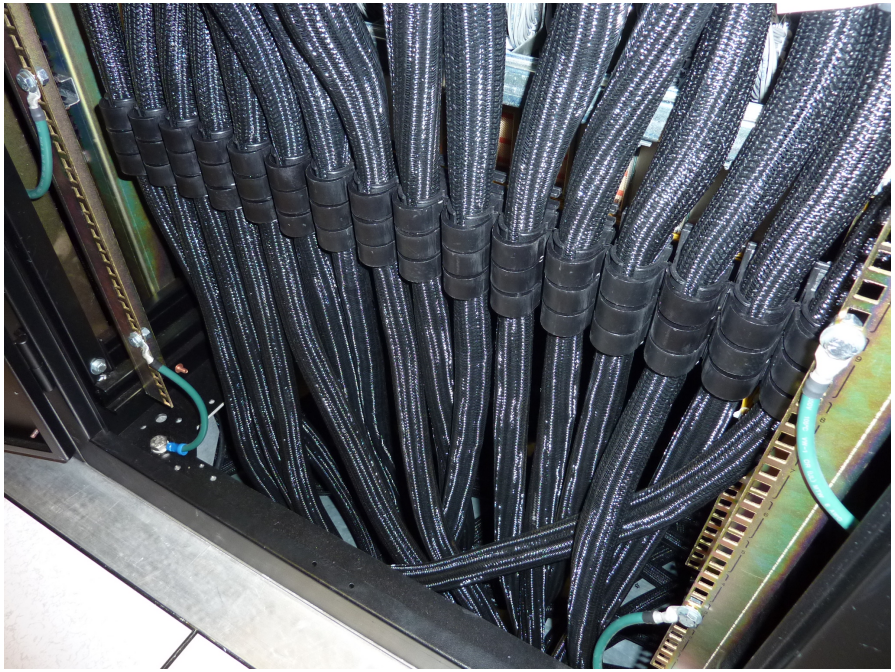


The top of each rack houses 4 adjustable-speed squirrel-cage fans.

- Each rack contains a panel with a breaker for each Station/Baseline Board
- Each rack also contains an RPMIB (Rack Power Monitor/Control Interface Board) for remote control and monitor by software.

Intra-rack cabling between Station Boards and Crossbar Boards.

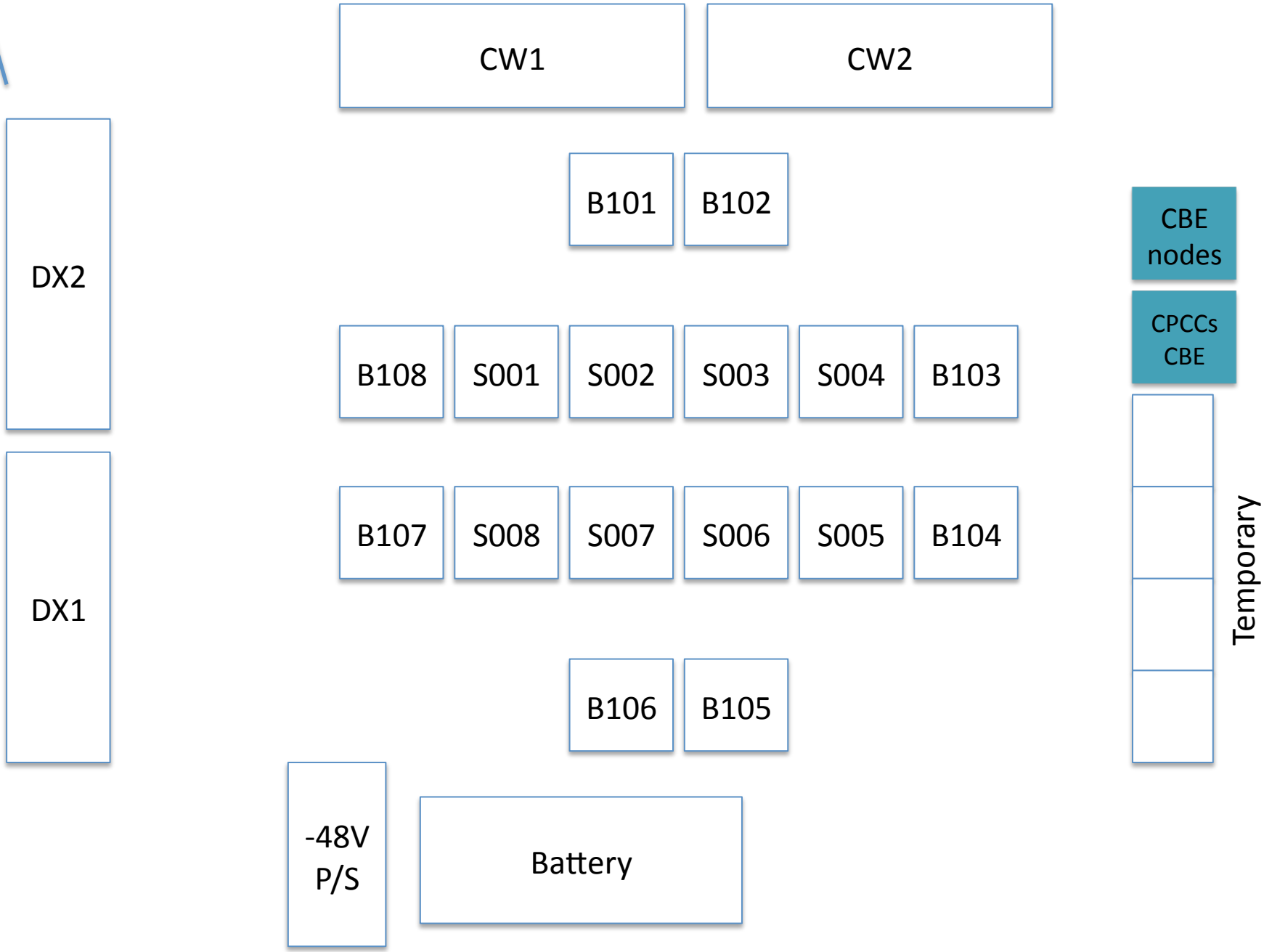
Inter-rack cabling between Station Racks and Baseline Racks are under the floor.
(**some 30 miles of it**)



Correlated Output and Monitor/Control data travels overhead

- Blue: Monitor/Control data to/from CMIBs (100Mb/s)
- Red: Correlated Output Data from Baseline Boards to CBE (currently 1Gb/s)
- Green: Output Data for VLBI use
- The yellow channel houses fiber optic cable from Antennas to Station Boards

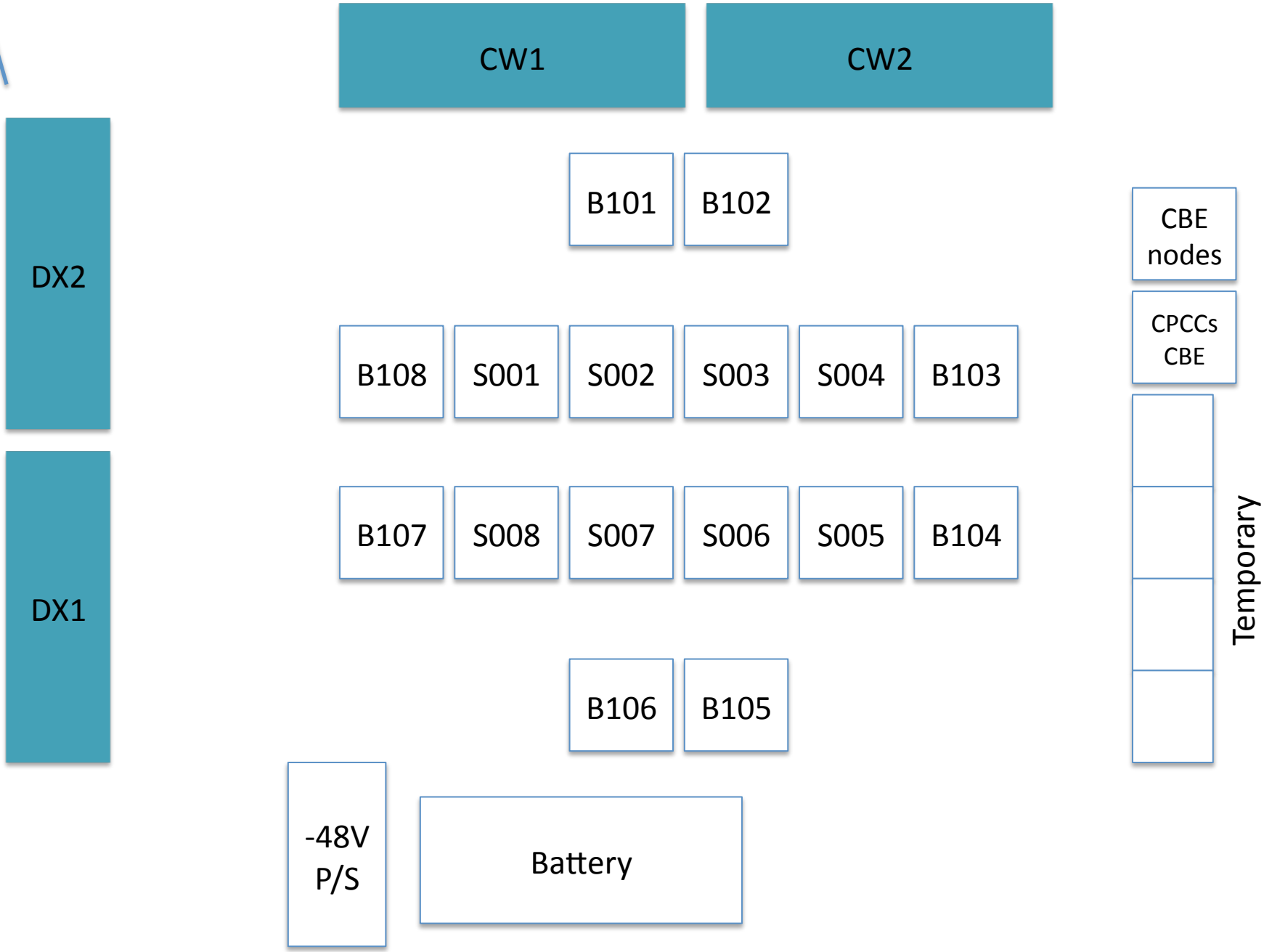
- Two auxiliary racks house additional equipment
  - Correlator Backend (CBE) computer cluster
  - Correlator Power Control Computers (CPCCs)
  - Next Generation Archive Server (NGAS)
  - Master Correlator Control Computer (MCCC)
  - CMIB Bootservers


CW1
CW2
DX2
B101
B102
CBE nodes
CPCCs CBE
B108
S001
S002
S003
S004
B103
DX1
B107
S008
S007
S006
S005
B104
Temporary
B106
B105
-48V P/S
Battery

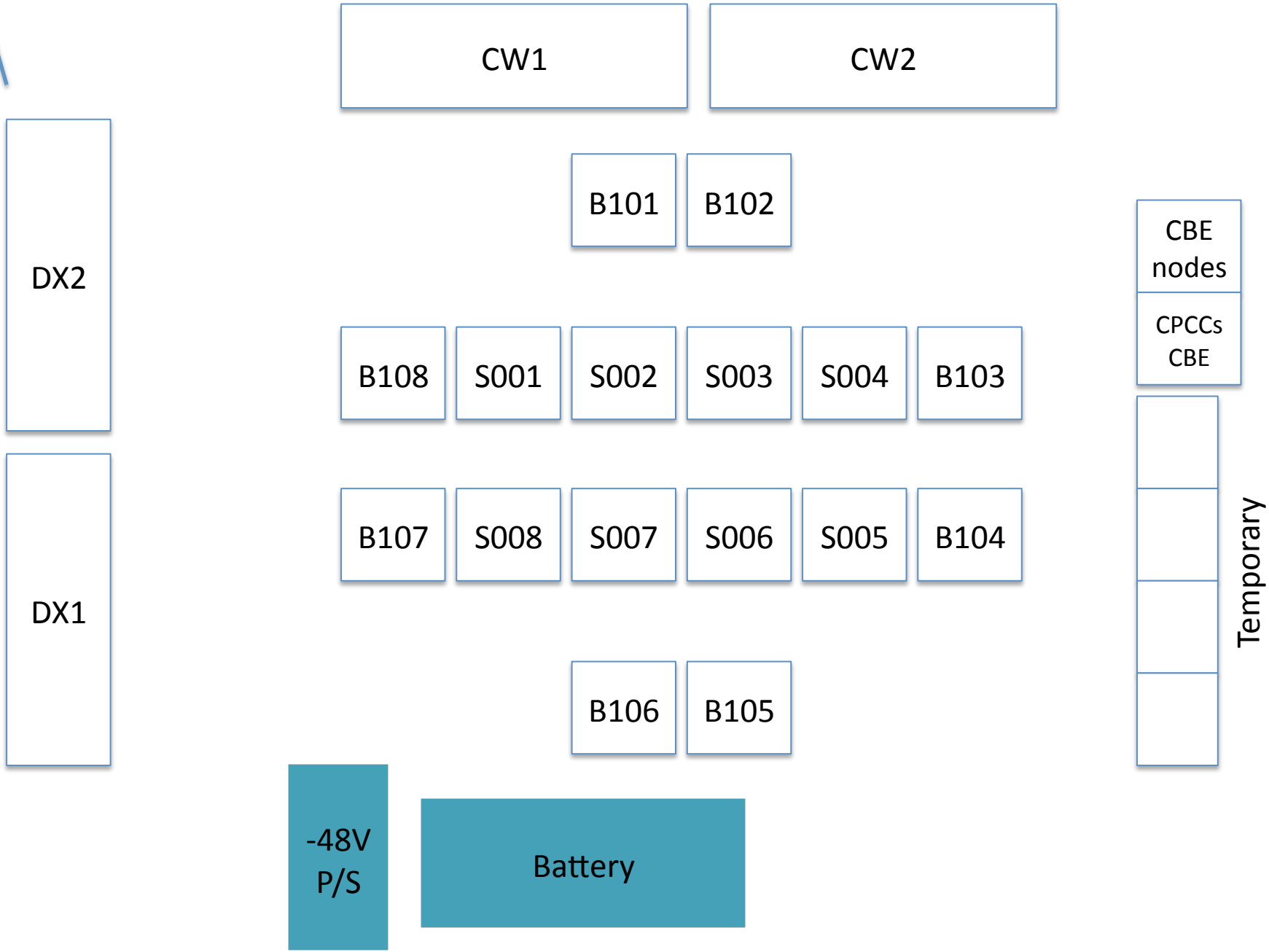
CPCCs, NGAS, CBE Master, CBE nodes, staging area file system (two views of same rack)

- CBE (Correlator Back End)
  - Assembles raw data from Baseline Boards - in the proper order - into what is called the BDF (Binary Data Format)
  - The BDF is written to the Staging Area which is a file system that temporarily holds the data until the Archive System takes it.
- Each Baseline Board has a 1 Gbit/s Ethernet link to the CBE (expandable to 10 Gbit/s)

- CPCCs are the Correlator Power Control Computers.
  - Two of them run in parallel for redundancy
    - Every command sent to one is mirrored to the other
  - At least one CPCC must be running to keep power applied to the correlator boards
    - If they both should ‘die’, power to all boards will be removed at once (bad), all rack fans will go to high speed and all 4 HVACs will be turned ON.
    - One CPCC is powered by A/C mains via UPS
    - The other is powered by -48VDC via inverter


CW1
CW2
DX2
DX1
B101
B102
B108
S001
S002
S003
S004
B103
B107
S008
S007
S006
S005
B104
B106
B105
CBE nodes
CPCCs CBE
Temporary
-48V P/S
Battery

Two DX – Direct Exchange units. Have their own compressors and refrigerant.
Two CW – Chilled Water units. Work off the building's chilled water loop.

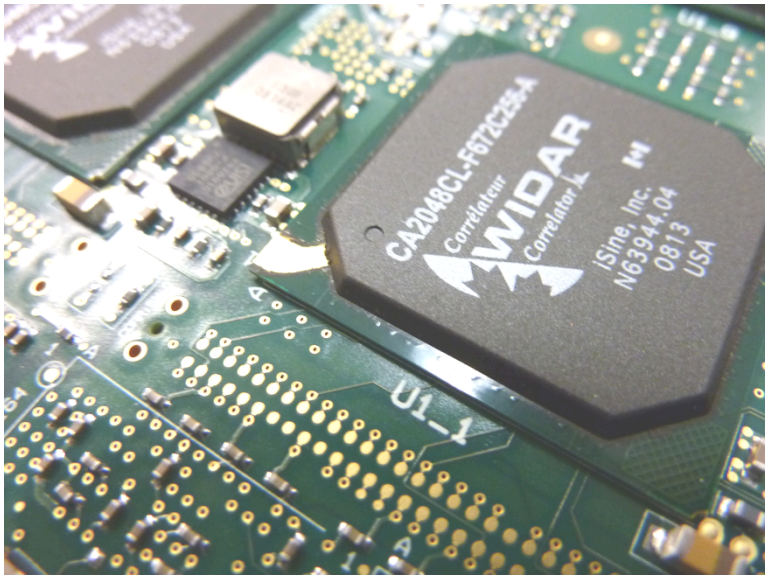

CW1
CW2
DX2
B101
B102
CBE nodes
CPCCs CBE
B108
S001
S002
S003
S004
B103
DX1
B107
S008
S007
S006
S005
B104
Temporary
B106
B105
-48V P/S
Battery

202,000 watts max, nominal 170,000
Battery backup rated for 5 minutes
-48VDC is what is used in the telephone industry for human safety

So, how does it all work?

- The correlator is ‘configured’ into one of many (many, many) different modes of operation as specified by the Observation Executor.
- Each Station, Baseline and Crossbar board and the CBE too, is sent a set of configuration instructions at the beginning of the observation.
- Presently this is done manually but will eventually be accomplished by Sonja’s Configuration Mapper in conjunction with the Obs Prep Tool.
- Okay, so now how does a board ‘get configured’ …

Correlator boards contain Field Programmable Gate Array (FPGA) devices*



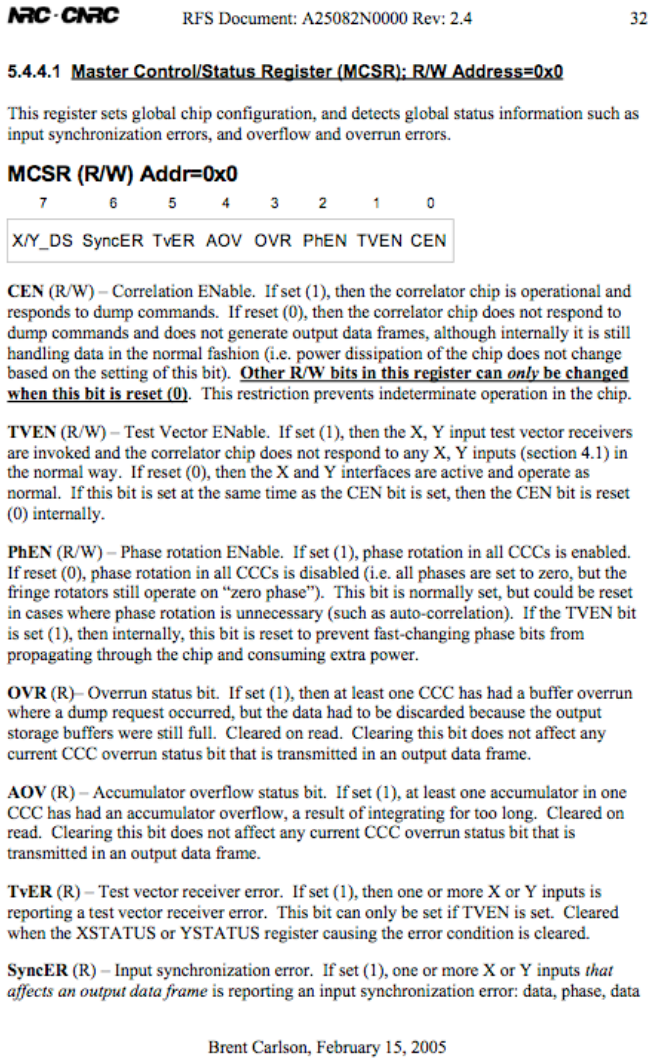
* The Correlator Chips themselves are actually ASICs custom factory made for WIDAR

17,308,672 control/monitor bits in
1,473,536 registers in
24,832 FPGAs in
256 boards in
16 racks in
1 room

- The FPGAs are generic and get their 'personalities' loaded into them at system startup. (Filters, recirculators, LTAs, Ethernet, etc. …)
- They contain sets of registers that may be written to to configure the chip (cause it to behave in a specific manner).
- At the beginning of an observation, thousands of registers are set to cause the correlator to become configured as required by the observation.
- Conversely, the registers may be read to obtain the state of the system.

• Each and every register of every FPGA/ ASIC is documented in a set of Requirements and Functional Specifications (RFS).
• That can be found here: https://mitr.drao.nrc.ca/widar/private/Private.html

Access to these registers is through the **CMIB**



NRC·CNRC RFS Document: A25082N0000 Rev: 2.4 32

5.4.4.1 Master Control/Status Register (MCSR): R/W Address=0x0

This register sets global chip configuration, and detects global status information such as input synchronization errors, and overflow and overrun errors.

MCSR (R/W) Addr=0x0

7	6	5	4	3	2	1	0
X/Y_DS	SyncER	TvER	AOV	OVR	PhEN	TVEN	CEN

CEN (R/W) – Correlation ENable. If set (1), then the correlator chip is operational and responds to dump commands. If reset (0), then the correlator chip does not respond to dump commands and does not generate output data frames, although internally it is still handling data in the normal fashion (i.e. power dissipation of the chip does not change based on the setting of this bit). **Other R/W bits in this register can *only* be changed when this bit is reset (0)**. This restriction prevents indeterminate operation in the chip.

TVEN (R/W) – Test Vector ENable. If set (1), then the X, Y input test vector receivers are invoked and the correlator chip does not respond to any X, Y inputs (section 4.1) in the normal way. If reset (0), then the X and Y interfaces are active and operate as normal. If this bit is set at the same time as the CEN bit is set, then the CEN bit is reset (0) internally.

PhEN (R/W) – Phase rotation ENable. If set (1), phase rotation in all CCCs is enabled. If reset (0), phase rotation in all CCCs is disabled (i.e. all phases are set to zero, but the fringe rotators still operate on "zero phase"). This bit is normally set, but could be reset in cases where phase rotation is unnecessary (such as auto-correlation). If the TVEN bit is set (1), then internally, this bit is reset to prevent fast-changing phase bits from propagating through the chip and consuming extra power.

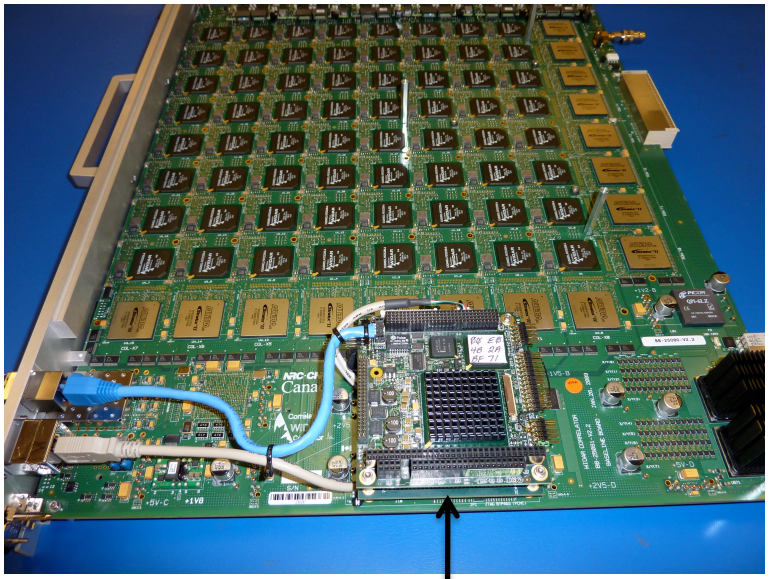
OVR (R)– Overrun status bit. If set (1), then at least one CCC has had a buffer overrun where a dump request occurred, but the data had to be discarded because the output storage buffers were still full. Cleared on read. Clearing this bit does not affect any current CCC overrun status bit that is transmitted in an output data frame.

AOV (R) – Accumulator overflow status bit. If set (1), at least one accumulator in one CCC has had an accumulator overflow, a result of integrating for too long. Cleared on read. Clearing this bit does not affect any current CCC overrun status bit that is transmitted in an output data frame.

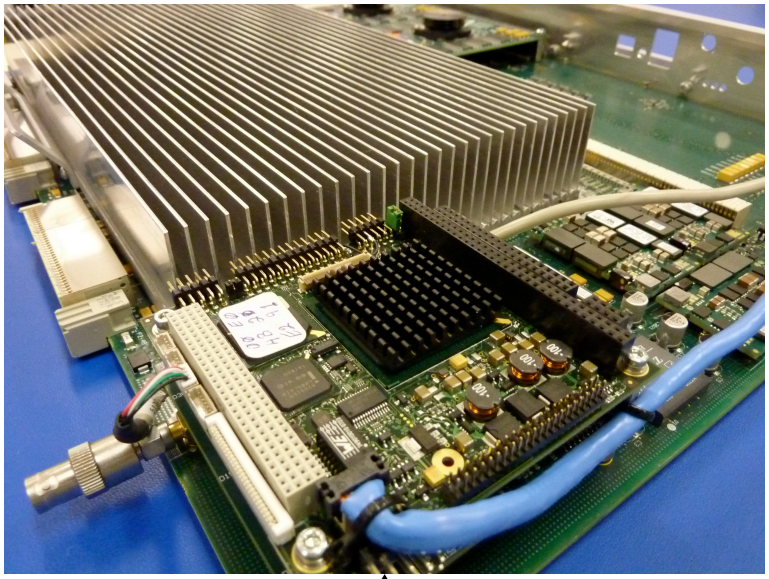
TvER (R) – Test vector receiver error. If set (1), then one or more X or Y inputs is reporting a test vector receiver error. This bit can only be set if TVEN is set. Cleared when the XSTATUS or YSTATUS register causing the error condition is cleared.

SyncER (R) – Input synchronization error. If set (1), one or more X or Y inputs *that affects an output data frame* is reporting an input synchronization error: data, phase, data

Brent Carlson, February 15, 2005



Baseline Board CMIB

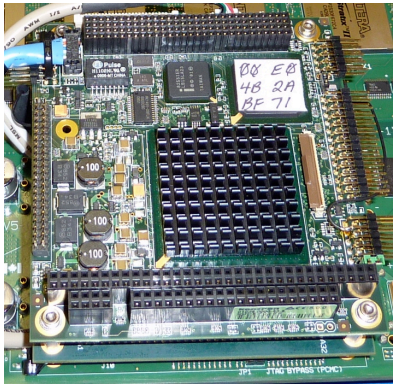


Station Board CMIB

Correlator **M**odule **I**nterface **B**oard

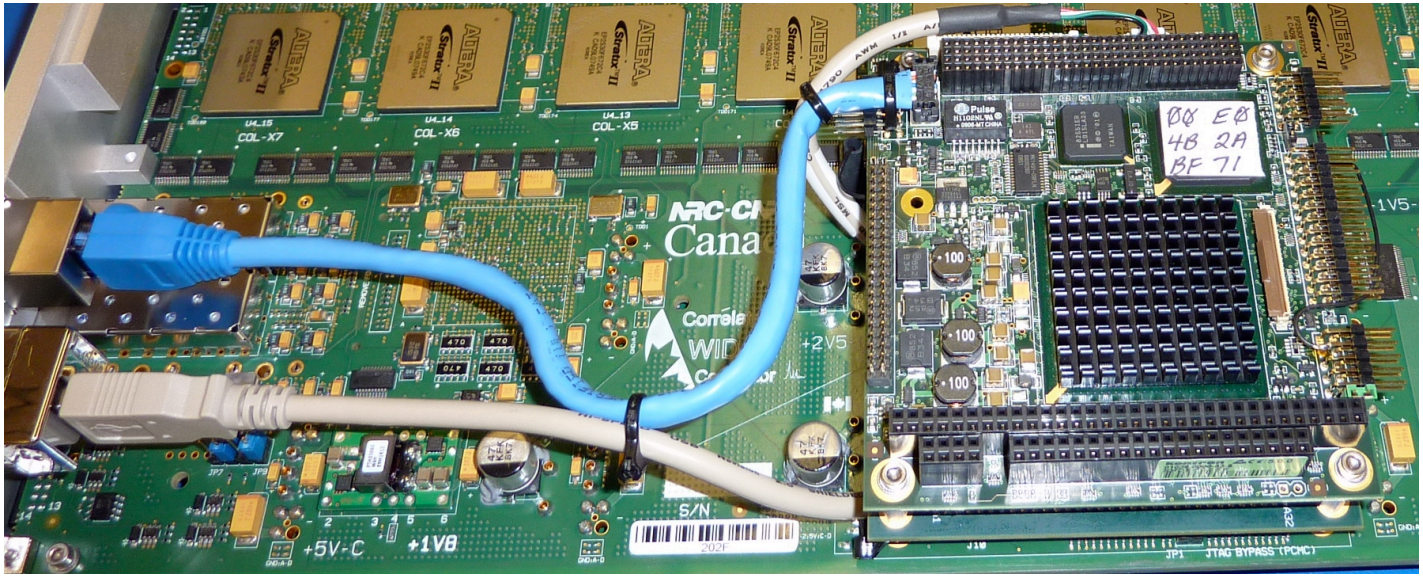
EVLA Antenna MIB on steroids

- The CMIB is a single board computer in the PC-104 form factor
  - COTS (Commercial Off The Shelf), ~ $300/ea
- It contains an Intel/AMD type processor with ½ Gbyte of RAM
- Runs regular Linux, Java, C, and scripting languages
- It interfaces to the board hardware via a Monitor/Control Bus (MCB)
- And to the 'outside' via Ethernet



```
0000000 fe ed fa ce 00 00 00 12 00 00 00 00 00 00 00 02
0000010 00 00 00 0a 00 00 05 00 00 00 00 85 00 00 00 01
0000020 00 00 00 38 5f 5f 50 41 47 45 5a 45 52 4f 00 00
0000030 00 00 00 00 00 00 00 00 00 00 00 10 00 00 00 00
0000040 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0000050 00 00 00 04 00 00 00 01 00 00 01 8c 5f 5f 54 45
0000060 58 54 00 00 00 00 00 00 00 00 00 00 00 00 10 00
0000070 00 00 10 00 00 00 00 00 00 00 10 00 00 00 00 07
0000080 00 00 00 05 00 00 00 05 00 00 00 00 5f 5f 74 65
0000090 78 74 00 00 00 00 00 00 00 00 00 00 5f 5f 54 45
00000a0 58 54 00 00 00 00 00 00 00 00 00 00 00 00 19 20
00000b0 00 00 04 e4 00 00 09 20 00 00 00 02 00 00 00 00
00000c0 00 00 00 00 80 00 04 00 00 00 00 00 00 00 00 00
00000d0 5f 5f 70 69 63 73 79 6d 62 6f 6c 5f 73 74 75 62
```

- CMIBs speak binary to the FPGAs. It reads and writes to its registers.
- Processes in the CMIB run continuously to monitor/control the FPGAs.
  - All real-time processing is done in the CMIB eliminating the need for attempting it over the network.



- The ‘other side’ of the CMIB speaks to the outside world via Ethernet.
- The binary register data is abstracted by processes in the CMIB into human (almost) readable XML.
- Higher level Java processes in the CMIB present this XML via a standard World Wide Web interface using HTTP and URLs.

WIDAR contains some 260 Web Servers!

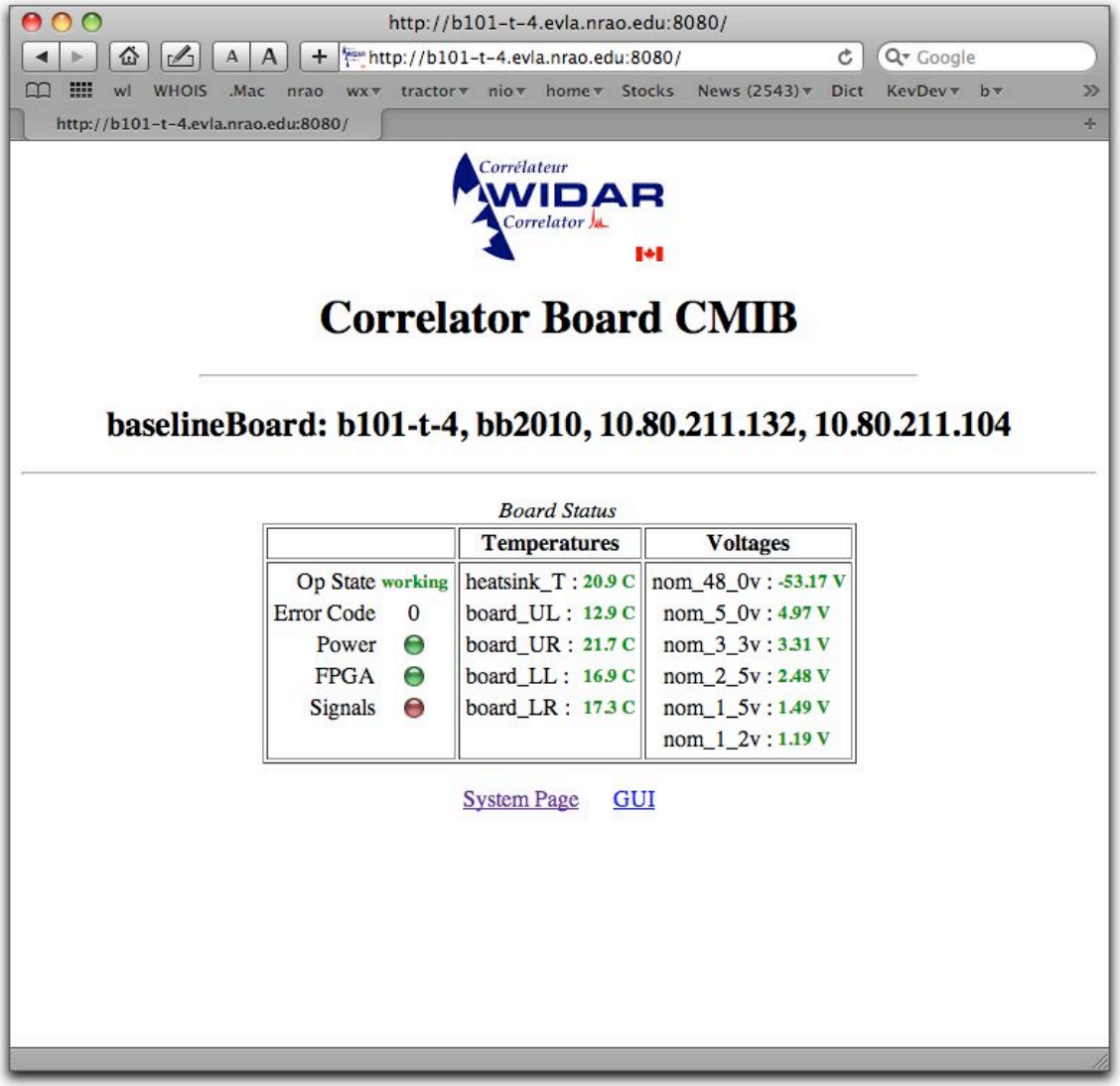
(Every CMIB and the CBE/CPCCs/MCCC)

- Each CMIB has three network addresses:
  - location name (b101-t-4)
  - serial number (bb2010)
  - IP address (10.80.211.104)

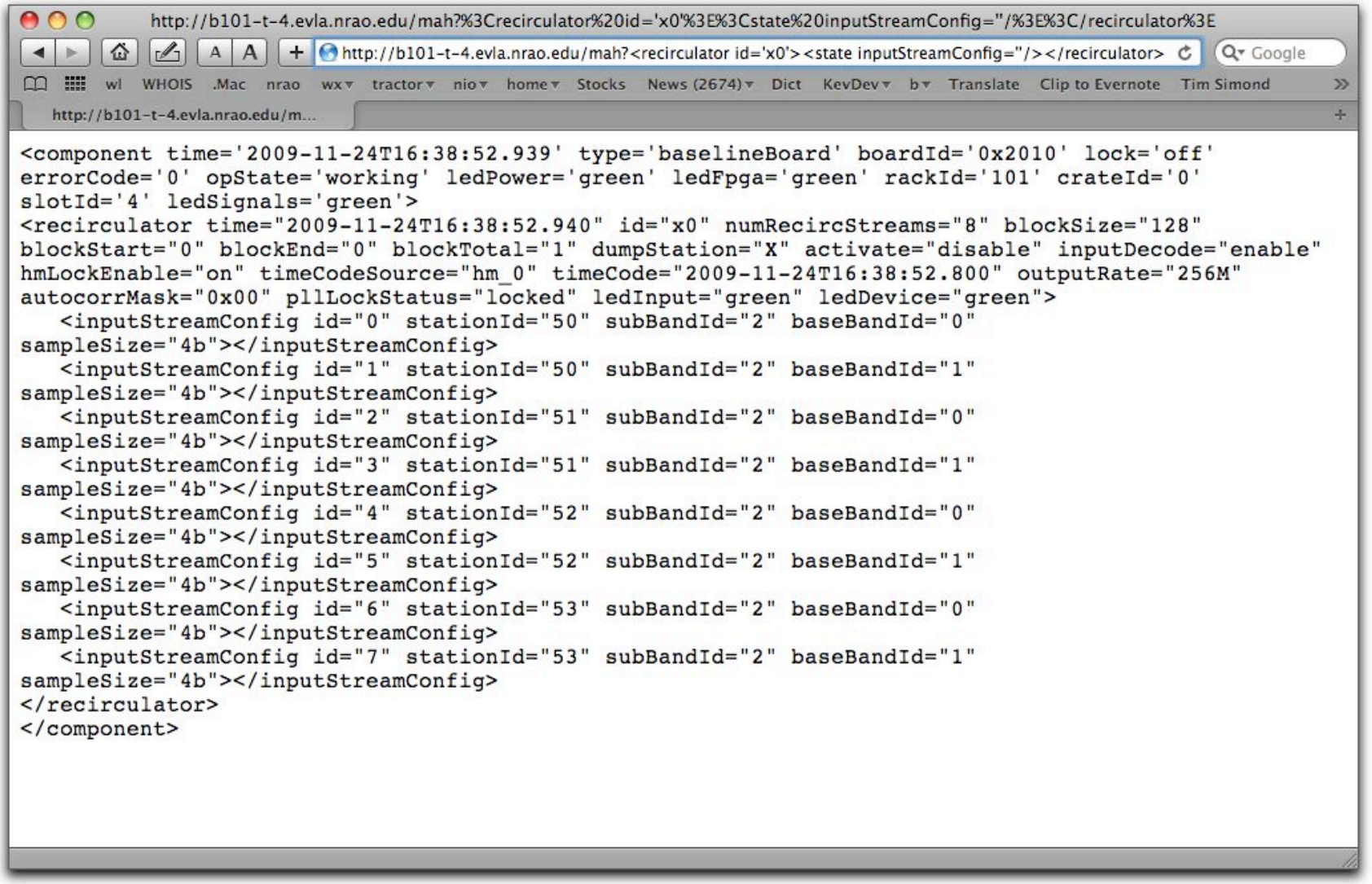
10.80.211.104

100 – 107 top crate, slots 0 – 7
200 – 207 bottom crate, slots 0 - 7

201 – 208: station racks S001 – S008
211 – 218: baseline racks B101 – B108

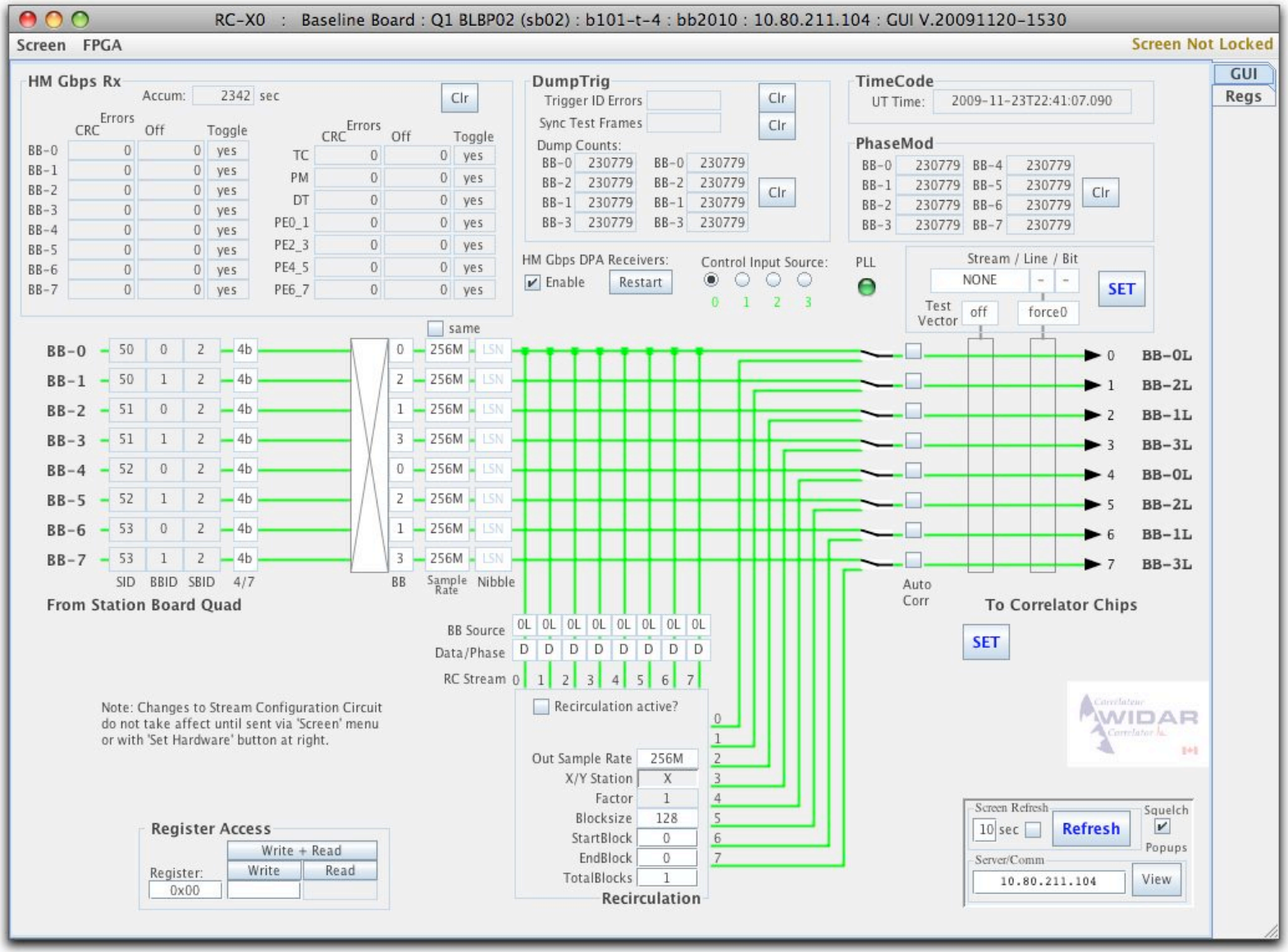


https://mitr.drao.nrc.ca/widar/private/Memos/NRC_EVLA_Memo030_MappingNames_Locations_IP_May19-07.pdf



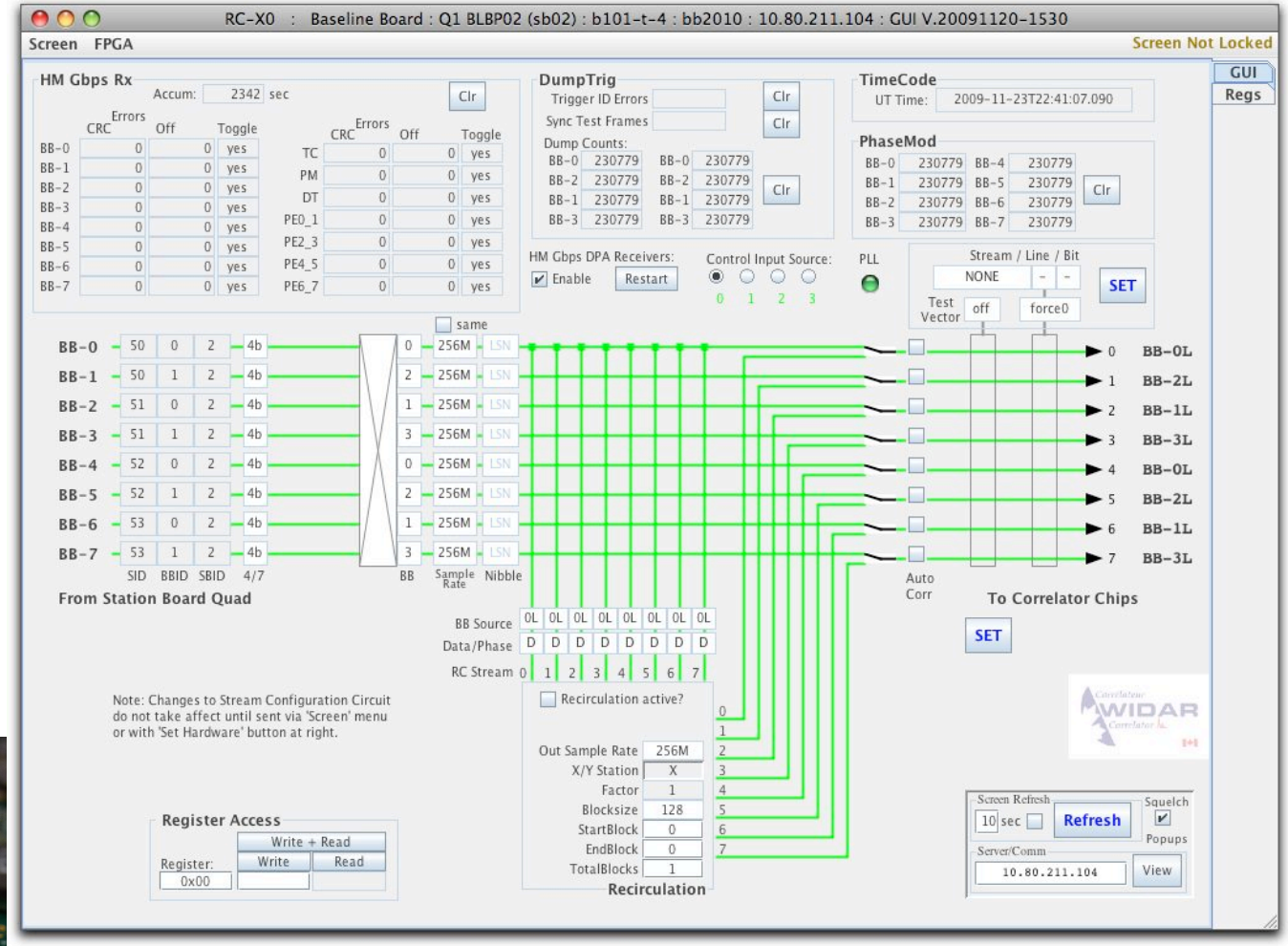
- We can use regular ol' URLs to obtain information from them:
http://b101-t-4.evla.nrao.edu/mah?<recirculator id='x0'><state inputStreamConfig=''/></recirculator>

- But that is still rather painful for all but the most trivial things; so we made GUIs …



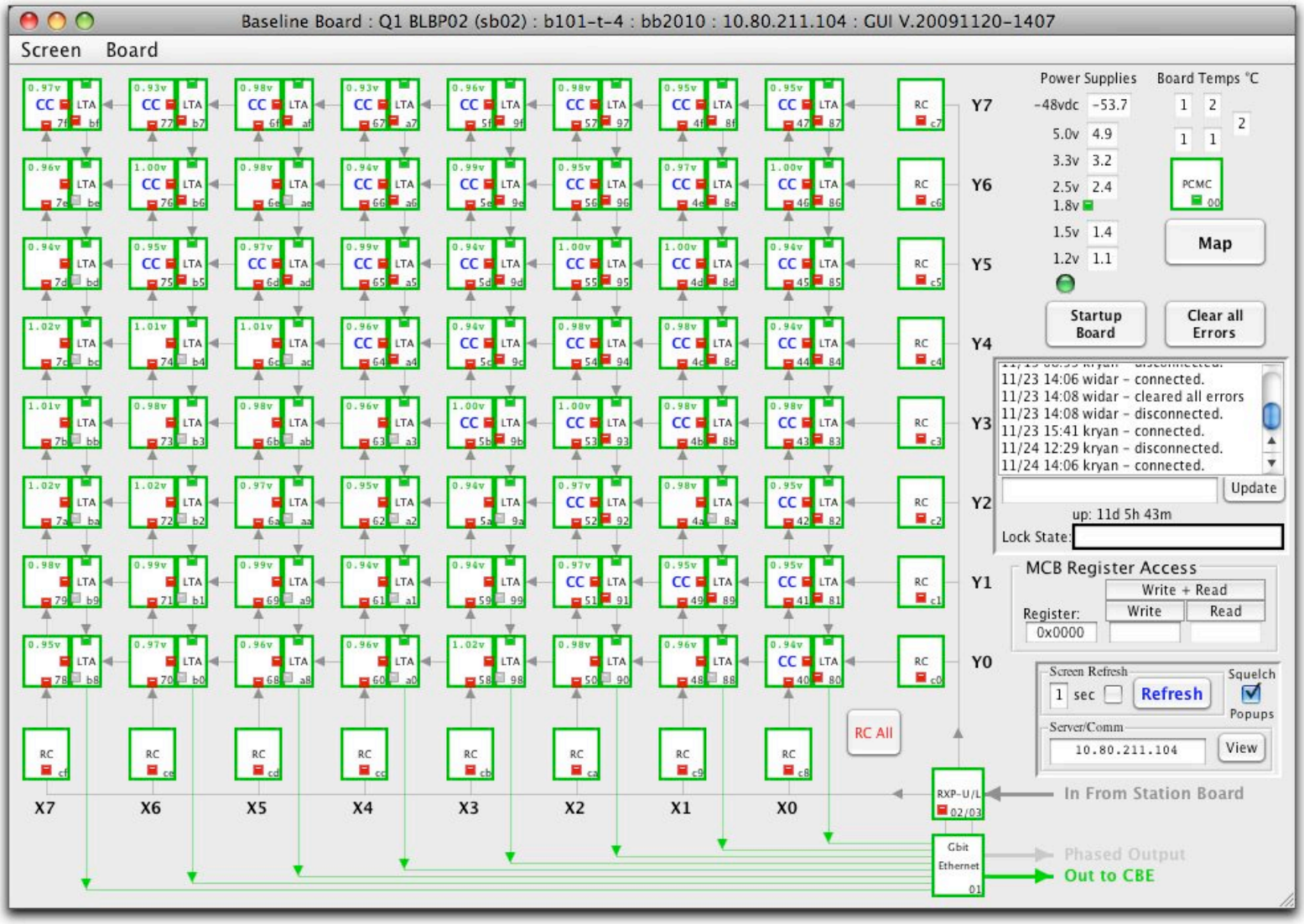
- GUI's send XML back and forth to the CMIBs (still using URLs like a web browser).
- The XML is translated into something better understood by a human.

GUIs give us a picture of what's going on inside an FPGA.



- CMIBs translate register data to XML, GUIs translate XML to a picture
- GUIs can also translate a picture to XML …

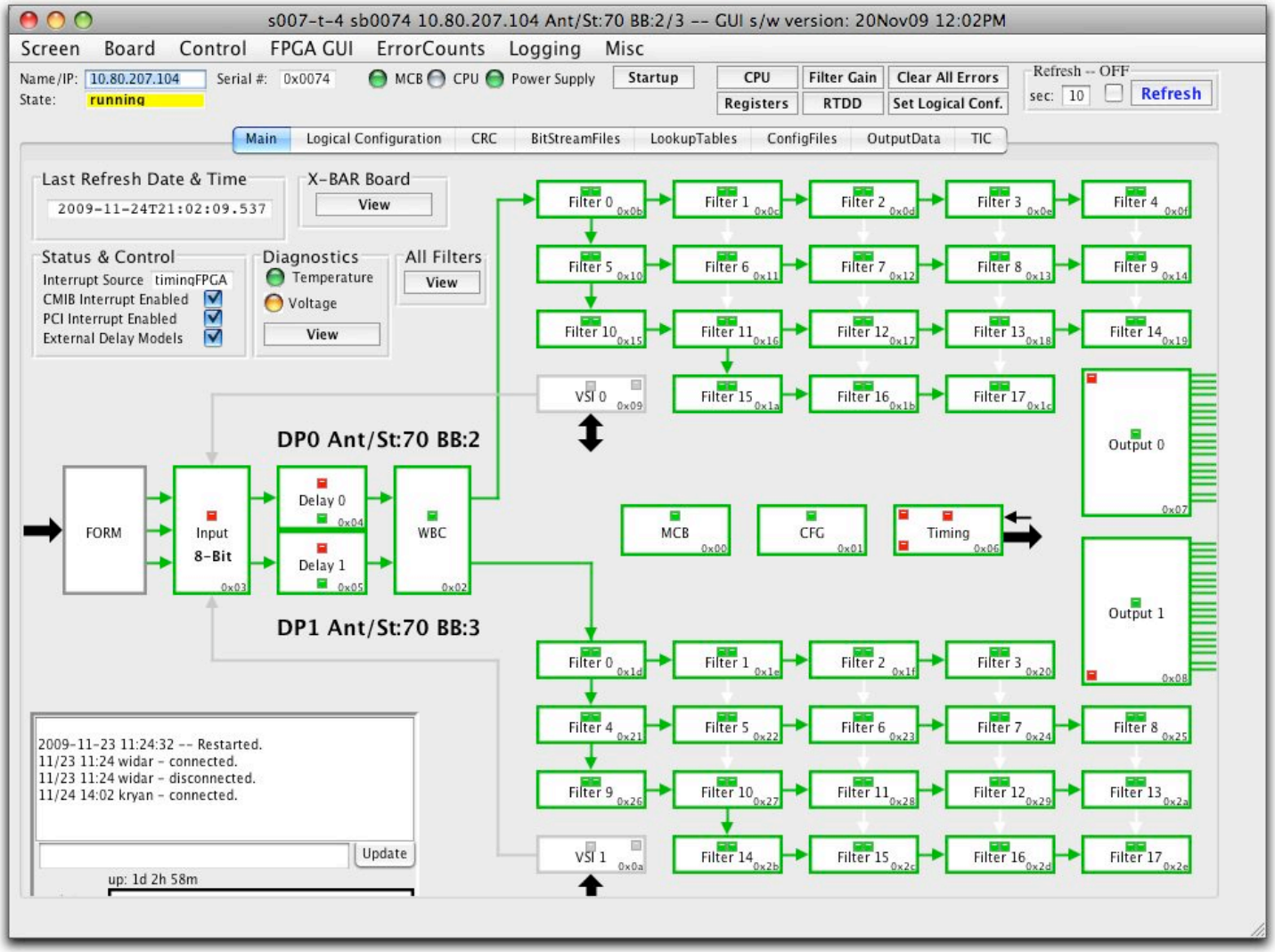
- GUIs can also read and write XML from and to files that can then be used by the CMIB to make a configuration.
  - This is an important feature. The whole correlator may be configured by loading files of XML into the CMIBs (as opposed to configuring them manually from the various GUIs).



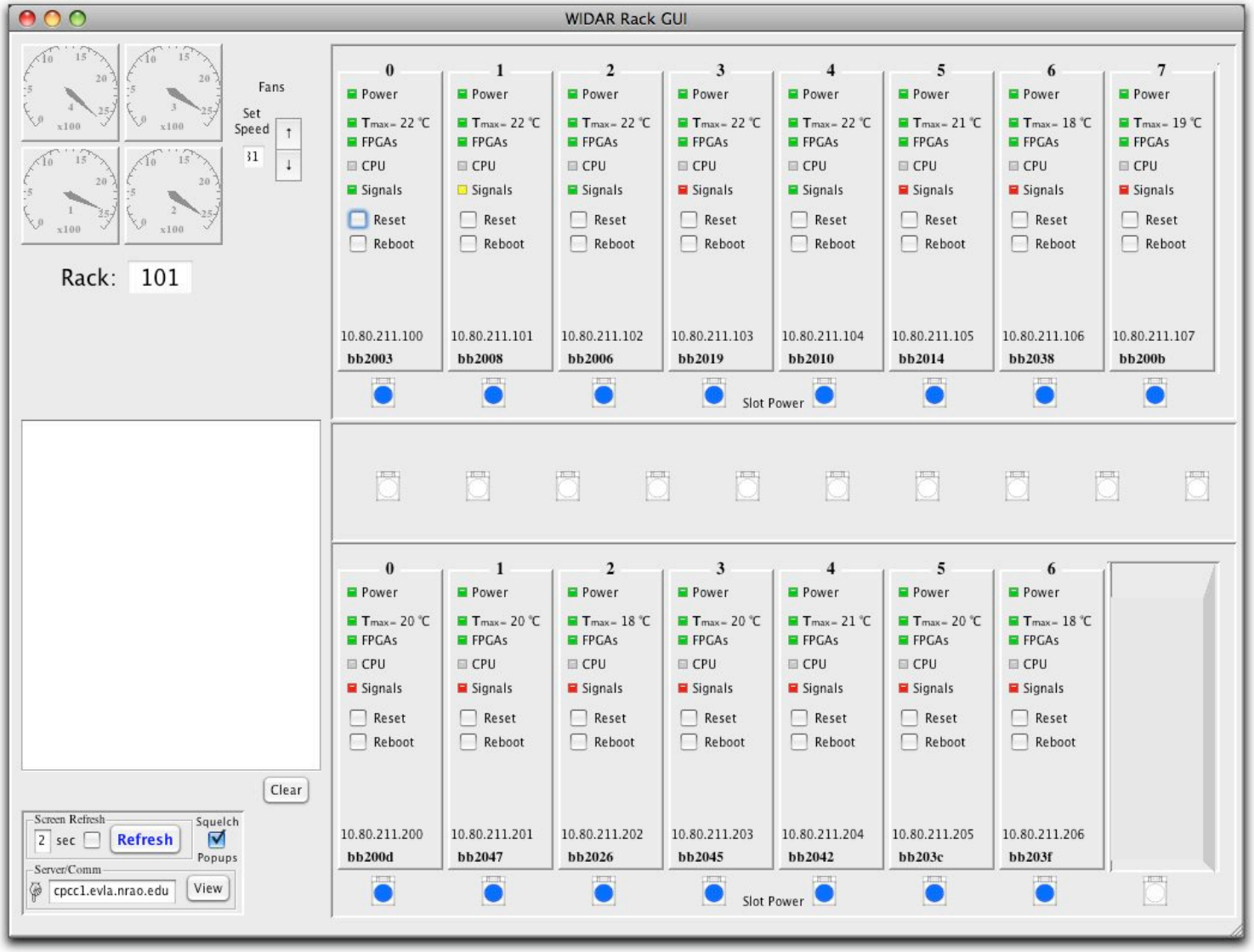
Baseline Board Top-level GUI

GUIs were designed to be hierarchical – this shows all the FPGAs at once

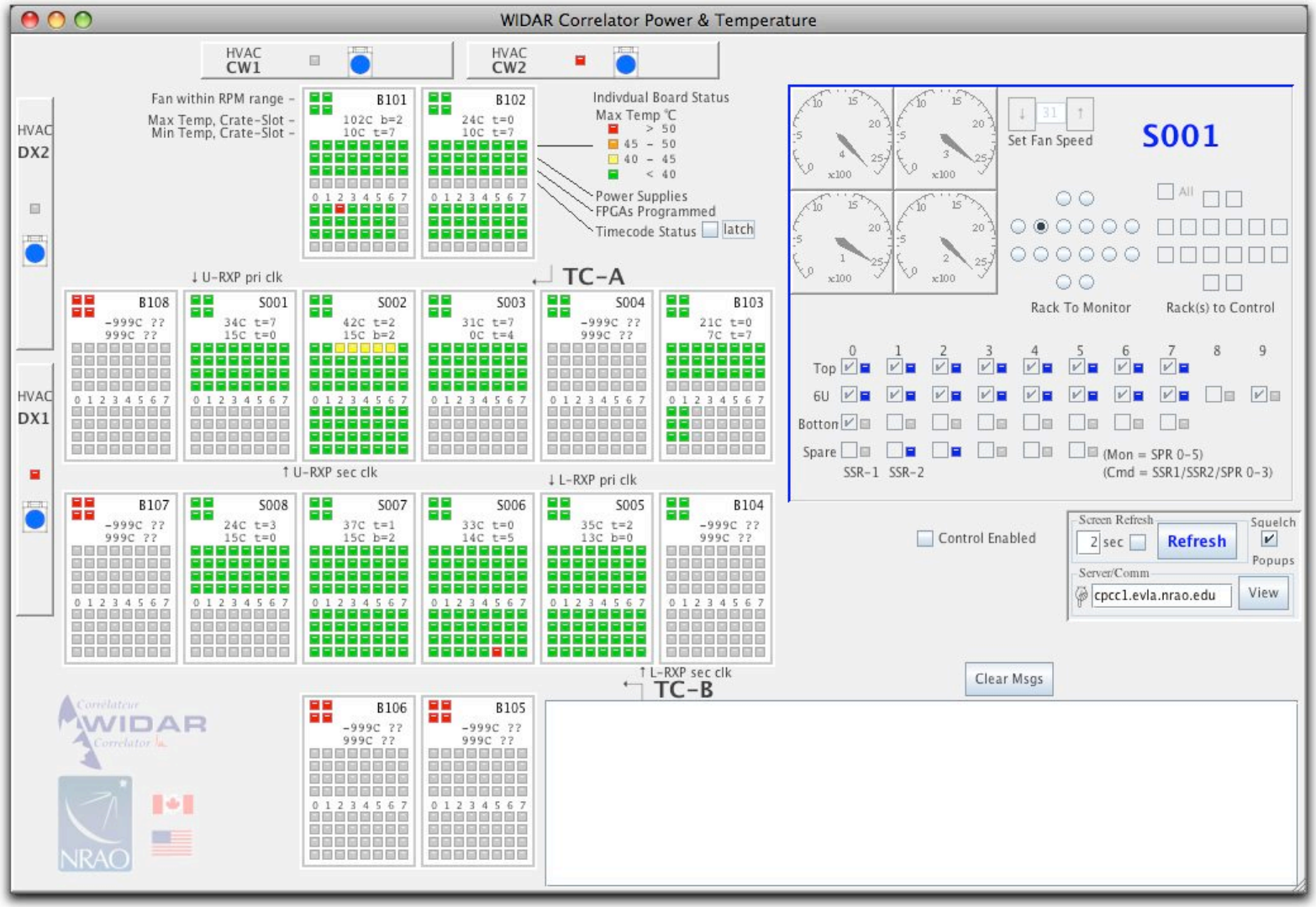
- With much higher-level summary info about each



Station Board Top-level GUI



Rack-level GUI
A Rack's worth of even higher-level board status



Room-level GUI

And, finally the highest-level status of all boards in the room.

- This last GUI will be the first one delivered for Operator use.
- This GUI communicates with the CPCCs for room status information.
- Continually running processes in the CPCC obtain monitor information from various sources:
  - CMIBs for board temperatures/voltages
  - RPMIBs for rack power and fan control
  - HVACs for power/alarm
  - Smoke Detection System
  - Power Plant

- CPCCs will attempt to protect the system without the need for human intervention
  - Reduce board thermal cycling (temperature swings) by controlling rack fans and HVAC
  - Prevent spikes by staging power to the boards (when turning the whole system both on or off)
  - Reduce power consumption while running on battery backup to preserve battery
  - Remove power to boards at 3rd stage of smoke detection (before all power is removed to the room and fire suppressant dumped).

- When the operator does need to be alerted, the CPCCs will do so via the present EVLA alert system.
- These alerts and associated operator actions have not yet been defined.
- Likewise, many of the correlator's interfaces to the EVLA are still under development.
- The correlator is in its infancy and is still not even fully populated with all its boards.
  - So expect growing pains as the EVLA and WIDAR become integrated.

- NRAO People of Possible Interest to Operators
  - Mike Revnell
    - Fiber optic receiver module (FORM)
  - Kerry Shores/Ray Ferraro
    - Correlator H/W Maintenance
  - Kevin Ryan
    - GUIs/Web, Inter-process comm s/w, CPCC system
  - Bruce Rowen
    - CMIB software
  - Martin Pokorny
    - Correlator Backend system software
  - Michael Rupen/Ken Sowinski
    - Correlator Operation