

UNIVERSITY OF HAWAII AT MĀNOA

Institute for Astronomy

Pan-STARRS Project Management System

GPC Architecture

Coop. Agreement No. : FA9451-06-2-0338
Prepared For : Pan-STARRS
Prepared By : Sidik Isani
Document No. : PSDC-730-004-00
Document Date : August 21, 2008
Revision : 00

DISTRIBUTION STATEMENT

Approved for Public Release – Distribution is Unlimited

Submitted By:

[Insert Signature Block of Authorized Developer Representative]

Date

Approved By:

[Insert Signature Block of Customer Developer Representative]

Date

The information here is available as HTML. The URL is: <https://svn.ifa.hawaii.edu/gpc/archive/psdc/gpc-architecture/>

Revision History

Revision Number	Release Date	Description
DR1	2005.02.15	Initial revision
DR2	2005.11.08	Summit Network v2.0 (as separate figure)
DR3	2006.02.01	Global Addressing started (as separate document)
DR4	2007.01.21	Incorporated Global Addressing and Summit Network
00	2008.08.21	Officially accepted into PSDC and added example GPC1 JPEG. Also updated Figure 17 (summit logical network) to reflect reality.

Contents

1	Physical Components	1
1.1	Pixel Representation in Software	1
1.2	The OTA Cell	2
1.3	The OTA CCD	3
1.4	The DAQ3U	4
1.5	The Modular Controller Subsystem	6
1.6	The Giga Pixel Camera	8
2	Global Addressing	10
2.1	Addressing Multiple Cameras	10
2.2	Addressing Slots/Chassis in a Camera	10
2.3	Addressing OTAs, Cells, and Pixels	11
2.4	Camera X,Y and Focal Plane Display	13
2.5	Translation Table	14
3	PS1 Computing Network	16
3.1	Logical Network (recommended)	16
3.2	Logical Network (actual, GPC1)	19
3.3	Logical Network (actual, TC3)	20
3.4	Physical Network (PS1)	21
3.4.1	Fiber Connections for STARGRASP Controllers	21
3.5	Summary of Fault Tolerances	23

List of Figures

1	OT Pixel	1
2	OTA cell	2
3	OTA CCD	3
4	DAQ3U Electronics	4
5	DAQ3U and 2 OTA	5
6	DAQ3U Internal Resources	6
7	Modular Controller Subsystem	6
8	OTAs and Modular Controller	7
9	Giga Pixel Camera	8
10	Giga Pixel Camera	9
11	Global Addressing : Multiple Cameras	10
12	Global Addressing : Test Camera 3	11
13	Global Addressing : One Giga Pixel Camera	12

14	Global Addressing : One OTA, One Cell	12
15	Image of the Sky on GPC1	13
16	Summit Logical Networks (Recommended)	16
17	Summit Logical Networks - GPC1	19
18	Summit Logical Networks - TC3	20
19	Dell 2748 Managing 1 STARGRASP Chassis	21
20	1 4-port Fiber Switch per Chassis with Redundancy	22
21	Pan-STARRS 1 Observatory Network	24

1 Physical Components

1.1 Pixel Representation in Software

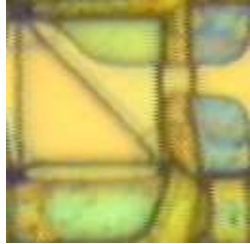


Figure 1: OT Pixel

The physical characteristics of an OT pixel have many implications on downstream software. A challenging property is that charge resulting from photons collected over time in one pixel can, under software command, can get shifted to a neighbor in any of four directions during the integration. Camera software is responsible for recording exact data necessary to derive the history of any pixel value it records.

Those pixel values shall be 16-bits for each sample point, and is the result of a combination of analog and digital steps, configured and managed by software and hardware:

- analog signal conditioning
- binning
- averaging of multiple ADC readings on one pixel
- differencing between the sample and a pedestal measurement (also averaged)
- linearity correction

Software interacts with this in the following ways:

- Value must always retain significance imposed by science requirements.
- Value from each pixel must be encoded within 16-bits of data per sample.
- Software must know all the options and provide access to each configuration mode, including binning factor, number of multi-samples.
- Software must provide diagnostic modes where it computes averages and differences itself.
- Linearity must be maintained within a range of output values.

1.2 The OTA Cell

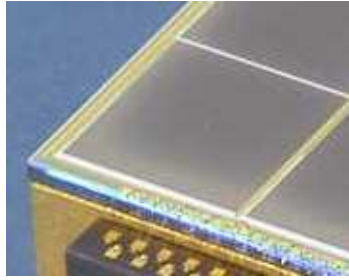


Figure 2: OTA cell

An OTA cell is manufactured as a square area of 6 x 6 mm filled with pixels. The smallest pixel size being considered is 10 microns, resulting in approximately 600 x 600 pixels per cell. This region of pixels has a single output amplifier at one corner. Cells will be operated by software in one of two ways (or ignored, if they are not usable due to expected manufacturing defects.)

Guide cells will be read out as video inputs during the time that the shutter is open. Rates between 10 and 30 Hz will need to be achieved, on boxes between 24x24 and 64x64 pixels in size. The location of the box must stay on one cell, but may be located anywhere within the boundaries of the cell and may be required to move to continue tracking an object.

Science Cells are read after the shutter closes, in the same manner as other Mosaic CCD science cameras.

Following common practice for constructing multi-extension FITS files, the pixel values resulting from the cell's amplifier shall be deposited linearly (with time, in the order seen by the amplifier) to a little-endian ("network") contiguous byte-ordered dump of 16-bit values. Thus, data provided to external systems by camera is always separated by cell, as represented on disk.

Science and controller hardware requirements dictate a maximum time of 2 microseconds per pixel. However, an expected rate of 1 microsecond per pixel could be reached. Software must never become a limiting factor, allowing the maximum performance of the detector and analog hardware always to be the driver. Therefore, software is expected to support solidly a 1 microsecond per pixel rate. Time to read a cell therefore depends on the pixel size used to fill the cell (but data rates remain the same.)

pixel size	geometry	read time	data rate
10 micron	600+x600	0.375 s/cell	16 Mbps/cell
12 micron	500+x500	0.25 s/cell	16 Mbps/cell

Note that because these numbers are based on the 1 microsecond per pixel baseline, these software performance numbers have two properties: (a) they assure that the system will never be software limited, and (b) they provide a factor of safety of 2 from having acquisition software infringe on science requirements (which result in a 2 microsecond per pixel read time.) Though it will not be reiterated, this statement applies to all derived requirements at the higher levels described below.

1.3 The OTA CCD

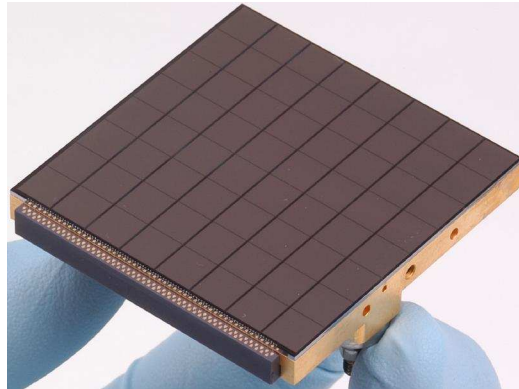


Figure 3: OTA CCD

An 8 by 8 grid of cells exists on a single piece of silicon with 8 amplifiers, capable of reading up to 8 cells (one from each column.)

The data rates listed in the previous section have other restrictions. These are not “burst” rates, as they must be sustainable, repeated over 8 rows of cells in succession. Between rows, small overheads exist as the hardware must be programmed to address a new row, but the bulk of the time is spent transferring pixels (at the maximum rate of 1 microsecond per pixel per cell.) Cell values are also not aggregate values. When applied per OTA, data rates must be multiplied by the number of amplifiers. This gives two minimum required times for software to read an OTA, again depending on pixel size, and a data rate per OTA, independent of pixel size.

pixel size	geometry	read time	data rate
10 micron	8 cells x 8 cells	3 sec/OTA	128 Mbps/OTA
12 micron	8 cells x 8 cells	2 sec/OTA	128 Mbps/OTA

These numbers apply to the data path between the clocking hardware generating pixel samples and on-board memory of the DAQ3U (described next.)

1.4 The DAQ3U



Figure 4: DAQ3U Electronics

The DAQ3U is the heart of STARGRASP detector controller. This board combines complex analog and digital functions to produce pixel output from a set of detector amplifiers. Each revision 2 boardset manages 16 amplifier outputs. As described in the previous section, one OTA device is capable of occupying half those channels at a time (because it can read one cell per column at a time, and has 8 columns of cells.) Thus, two OTA CCDs connect to one DAQ3U board as shown in the following figure.

Along the path of pixel data shown above, software must support maximum rates of twice that of one OTA from the previous section. I.e., $2 \times 128 \text{ Mbps} = 256 \text{ Mbps}$.

There are many resources on the board to meet this task. Those available to software depend partially on field-programmable logic in the Xilinx chip on the board, and on how this logic interconnects the other parts on the board. Revision 2 boardsets have these components, relevant to software:

- PowerPC 405 Embedded CPU
 - 300 MHz clock speed
 - 100 MHz FSB
 - Data cache
 - Instruction cache
 - NO floating point
- 256 MBytes of ECC DDR memory (SODIMM)
- 2 MBytes of slower SRAM memory (on-board)
- Analog to Digital Converters ([Analog Devices 9826 ADC](#))
- Digital to Analog Converters ([Analog Devices 5532-2 DAC](#))
- Clock chip configuration and control (TBD: link to doc?)
- Gigabit Ethernet MAC
- RS232 serial port (9600 N81)

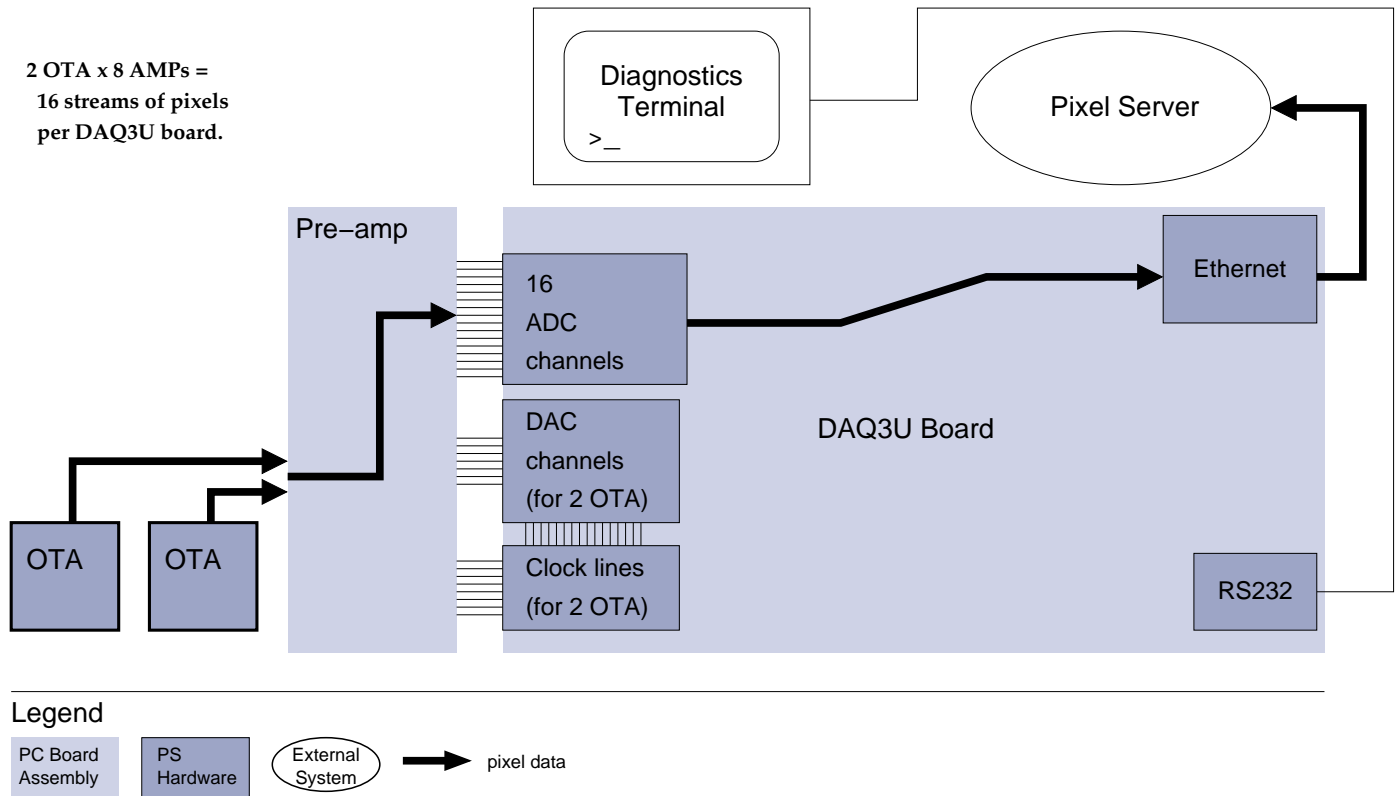


Figure 5: DAQ3U and 2 OTA

In addition, many of the resources on the board are field-programmable and their exact configuration (for example, the amount of memory) can depend on downloadable HDL implementations. These include:

- Configuration I/O ports to allow software to access CLK, DAC, ADC.
- DMA support for the Gigabit Ethernet
- Block RAM for initial PPC program data (BRAM)
- Small but fast dual-port on-chip memory (OCM)
- Access to a device configuration bus (DCR)

The DAQ3U is the primary hardware-software interface for pixel data, so it is also a component with many possibilities for cross-over. This is sometimes referred to as “off-loading.” For example, to meet the requirement of delivering a single 16-bit value representative of electrons on a single pixel, the PowerPC is capable of averaging multiple readout samples and subtracting a similarly averaged pedestal sample, just as field-programmable hardware in HDL can achieve this with accumulators and shift registers.

Software running on the PowerPC and HDL implemented in the Xilinx core both have access to the same busses on the board which interconnect all the hardware. Therefore there are a number of tasks which may be driven by either. Our strategy is to develop both in parallel, where possible. The typical things being traded by off-loading are speed, flexibility, or competition for other resources (FPGA space, memory, CPU cycles, etc.)

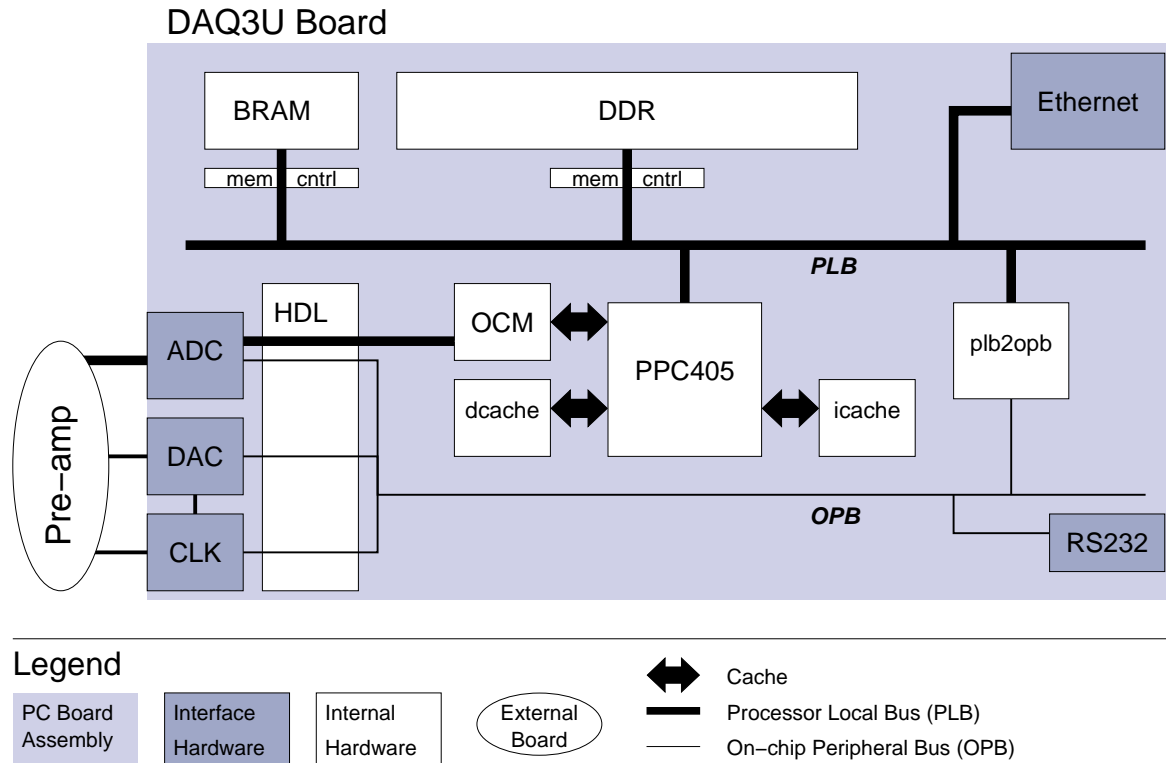


Figure 6: DAQ3U Internal Resources

1.5 The Modular Controller Subsystem

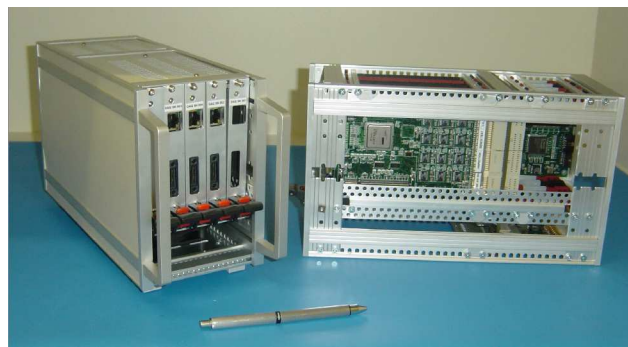


Figure 7: Modular Controller Subsystem

The controller design includes a chassis that holds a set of four DAQ3U boards on a common backplane. One such subsystem controls a total of 8 OTA devices in a 2x4 section of the final focal plane. There are four Ethernet links corresponding to this unit. The following figure illustrates the modular controller subsystem.

Each STARGRASP boardset has its own Ethernet as the exit for pixels, so from the point of view of software, there is nothing too significant about boards that are in the same subsystem. There may be an exception during bootstrapping, since it could be possible for one board to program another board in the same chassis, under software control. (We would rather avoid this complication, unless it significantly simplifies the hardware.) Another consideration is that software may wish to group IP node names or addresses by chassis for the convenience of hardware engineers (each boardset

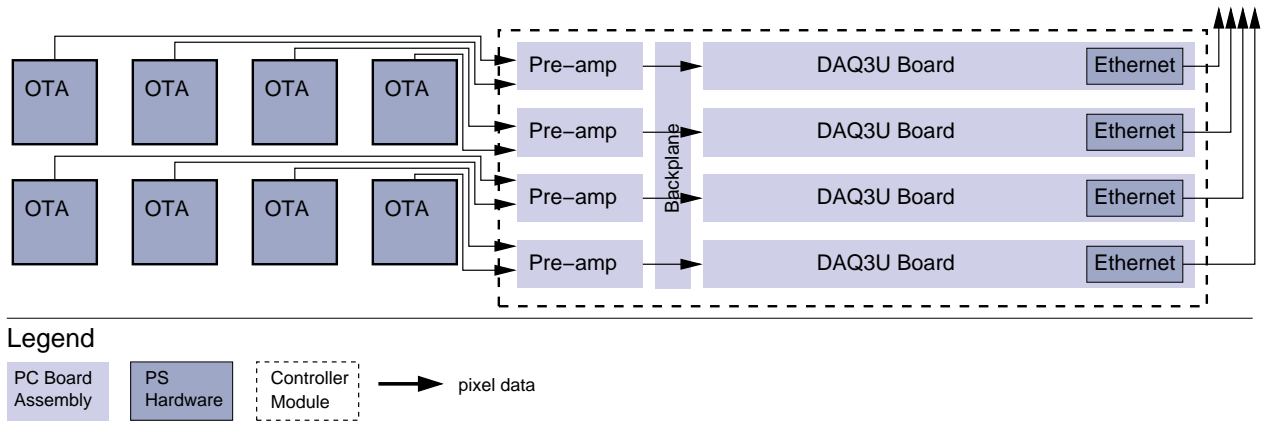


Figure 8: OTAs and Modular Controller

within a module has its own IP address on a private subnet.)

1.6 The Giga Pixel Camera

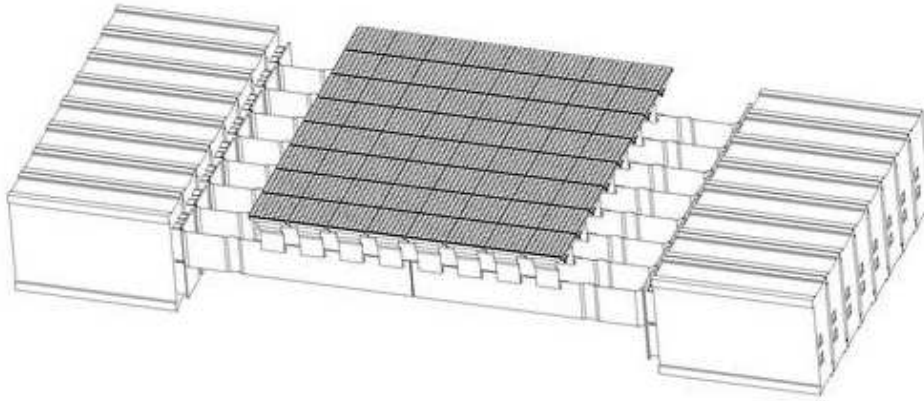


Figure 9: Giga Pixel Camera

The following diagram shows one of several possible network configurations. The switch shown may be located on the telescope, or at least located such that all the fibers rotate with the telescope. A switch with one or two 10 Gbps uplinks could be used, connected to another switch near the camera Pixel Servers. Finally, there is a possibility for no switch at all, in which case links from two DAQ3U would connect directly to one Pixel Server. This would require the installation of a third network interface on each Pixel Server. From a software level, these appear mostly to be cabling and logistics issues. Clearly, the choice of fiber versus copper is a physical layer difference only, when speaking of Ethernet. However, depending on exact constraints for guiding, having the DAQ3Us on a common switch has advantages compared to direct links from DAQ3U to Pixel Server:

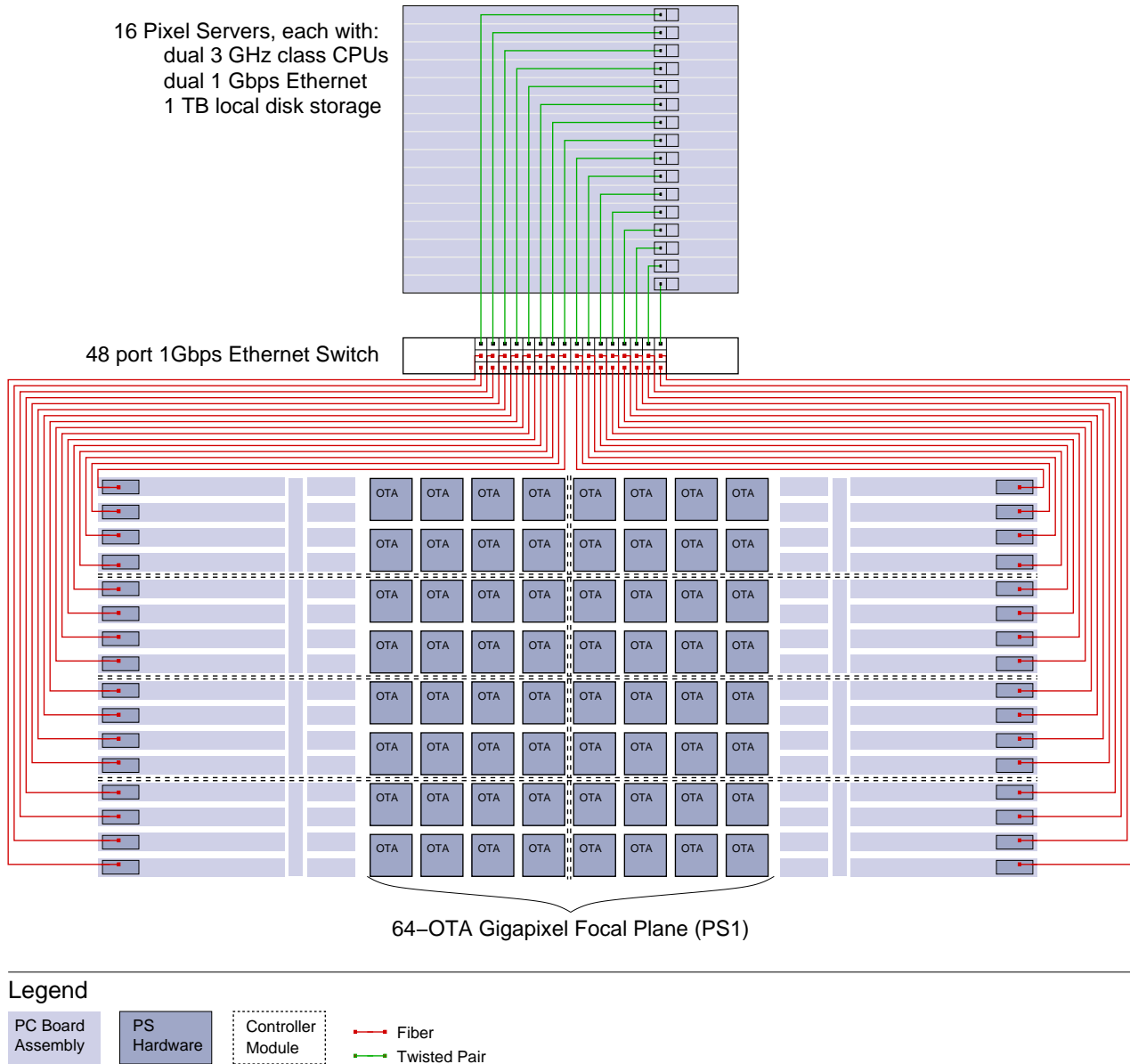


Figure 10: Giga Pixel Camera

- If Pixel Server hardware fails, its DAQ3Us can be re-assigned to another, on the fly.
- The possibility to exchange small amounts of information at very low latency directly between DAQ3U.
- The possibility to synchronize actions repeatably between DAQ3U.

The solution being implemented for Pan-STARRS 1 involves the 48-port switch in the support building, and one fiber pair per boardset passing through the wrap with no major switches in the dome itself. More details about this design are contained in the Summit Network descriptions later in this document.

2 Global Addressing

This section discusses the various machine readable bits and naming conventions used to locate and uniquely identify a camera, OTA, slot, and STARGRASP chassis. The identifiers may be found on silkscreens, etched into parts, burned into PROMS, and used as software names to connect to CCD controllers in the Giga Pixel Camera systems. These conventions should be used with TC3 (Test Camera 3), GPC1 (Giga Pixel Camera 1), and eventually GPC2, GPC3, and GPC4.

2.1 Addressing Multiple Cameras

The identifier for each camera we make should match its “common name” for simplicity. A camera will not be built with the knowledge of which camera it is. Instead, the DHCP service on the “Conductor” will assign a DNS domain name to a boardset on its subnet that will be indicative of the camera name. The camera name will be any combination of letters and numbers up to the first “.” (dot) in the domain name. The rest of the domain name might indicate the location of the camera (e.g., “.gpc.lab” or a telescope or site name.)

DHCP Server tells each IOTA which camera it is a part of, through "domain_name" field.

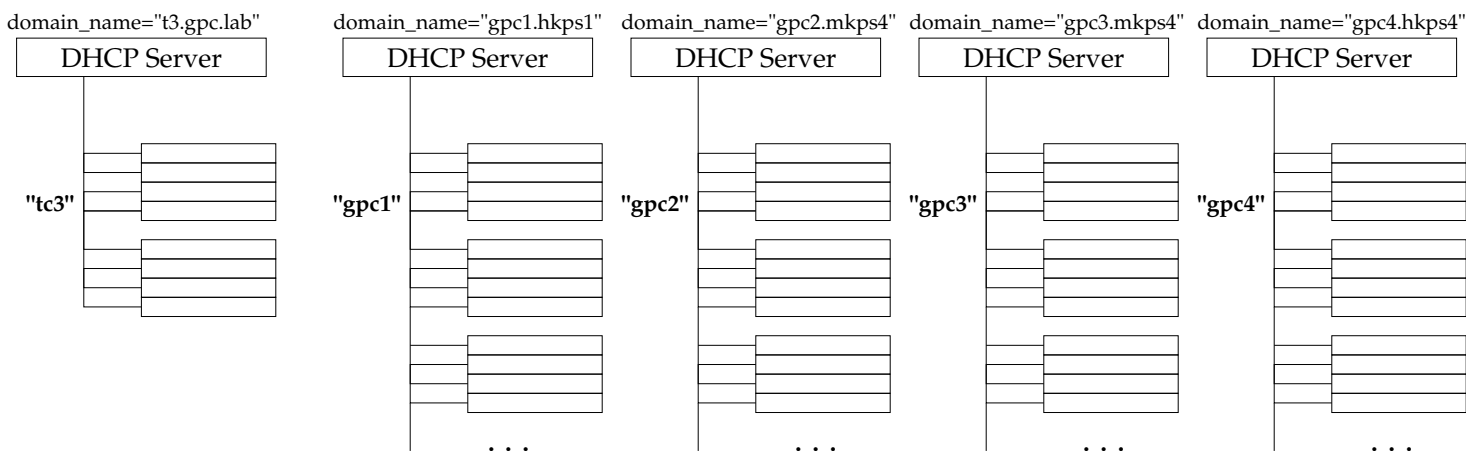


Figure 11: Global Addressing : Multiple Cameras

2.2 Addressing Slots/Chassis in a Camera

As described in the previous section, every boardset within one camera will be given the same DNS domain name by the Conductor DHCP server when the controllers are powered on. At the summit, Giga Pixel 3’s controller boards would all be assigned (pooled) IP addresses and given the domain name “gpc3.summit”, for example.

Each controller board will sense the following additional information:

- Whether it is a GPC, a TC, or a test bench based on the MSB of the **Chassis ID Bits** which software can read.
- A letter based on the three 3 LSBs of the 4 **Chassis ID Bits** which software can read.
- A number (1 through 4) based on the 2 **Slot ID Bits** which software can read.
- The package IDs of the two OTAs to which it is connected.

A complete Camera-Chassis-Slot identifier linked to each GigE interface is generated by appending all three together. Each card will connect to the shared namespace (also on Conductor) and use as its “working directory” a node by this Camera-Chassis-Slot. More on this can be found in the [Wiki Page on Resource Discovery](#).

In the shared namespace, it will automatically save:

- The OTA package ID it sensed for device position 0 (d0).
- The OTA package ID it sensed for device position 1 (d1).
- Whatever random IP address it was assigned by Conductor DHCP.

Chassis labelling, slot numbering, and the identifiers used by software are shown in the following figure, for TC3:

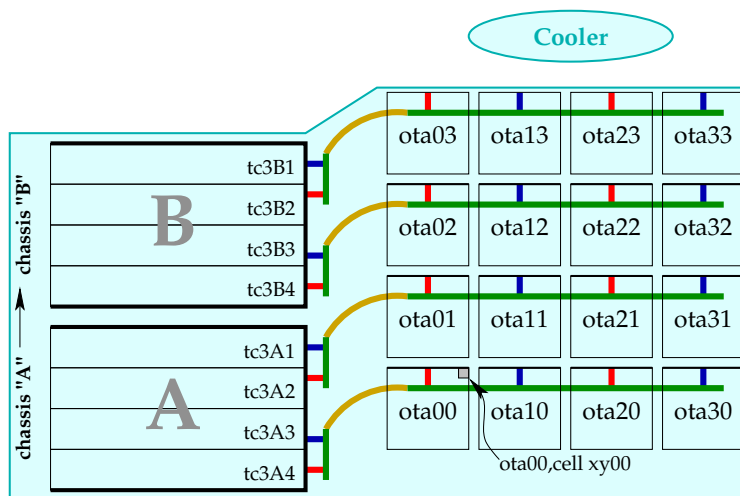


Figure 12: Global Addressing : Test Camera 3

Warning: Currently the TC3 Web interface displays images rotated 180 degrees relative to this specification.

The following figure shows a GPC (GPC4, just for the example, but 1,2,3 would be the same.)

Note that the chassis order is somewhat arbitrary, but we chose to put “A” in the lower left corner so that a TC3 setup would include only chassis “A” and chassis “B”. **The hole in the TC3 cryostat for the cooler comes out the top, when viewed in the orientation of this diagram then.** This means chassis go from “A” to “D” *right to left* when you’re looking at the side of the camera. Since the right half of the focal plane of a Giga Pixel Camera is rotated 180 degrees, chassis letter assignments go from “E” to “H” from top to bottom, or, again, *right to left* when looking at the side of the rack. Arrows in the figure are intended to show that.

2.3 Addressing OTAs, Cells, and Pixels

OTA and cell numbering are described in more depth in [the OTA FITS document](#). The thing to note here are that the 8x8 grid of OTAs in a GPC are numbered with the same Cartesian coordinates (“ota00” to “ota77”) as the 8x8 grid of cells found in each OTA (“xy00” to “xy77”).

Because half the focal plane is rotated 180 degrees in a GPC, and because the OTAs are back-side illuminated, pixel 0,0 is not in the lower left corner or cell 0,0!

Note: ID's like "gpc4A3" are the hostnames that the controller will request.

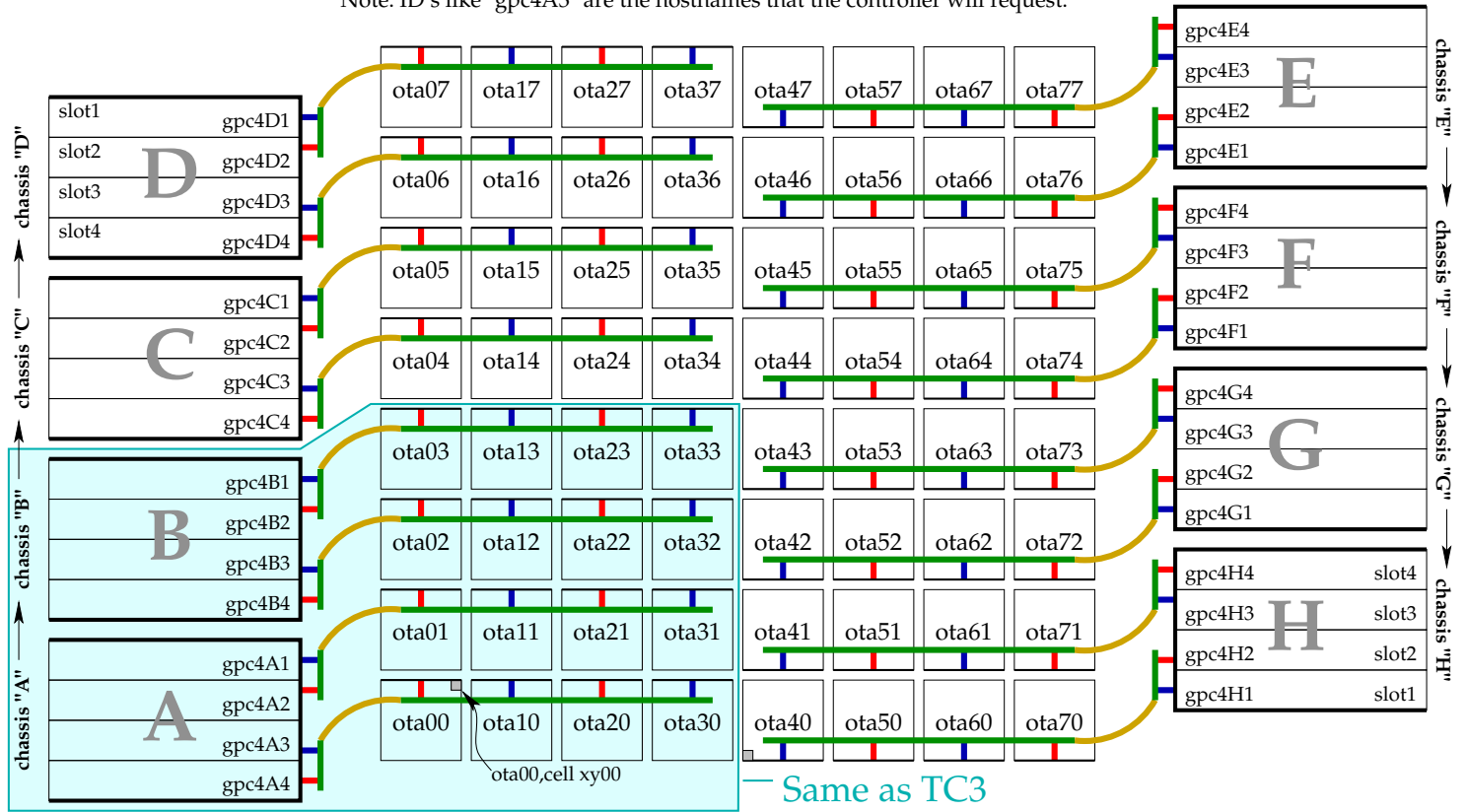


Figure 13: Global Addressing : One Giga Pixel Camera

Individual Cells:

One OTA:

xy07	xy17	xy27	xy37	xy47	xy57	xy67	xy77
xy06	xy16	xy26	xy36	xy46	xy56	xy66	xy76
xy05	xy15	xy25	xy35	xy45	xy55	xy65	xy75
xy04	xy14	xy24	xy34	xy44	xy54	xy64	xy74
xy03	xy13	xy23	xy33	xy43	xy53	xy63	xy73
xy02	xy12	xy22	xy32	xy42	xy52	xy62	xy72
xy01	xy11	xy21	xy31	xy41	xy51	xy61	xy71
xy00	xy10	xy20	xy30	xy40	xy50	xy60	xy70

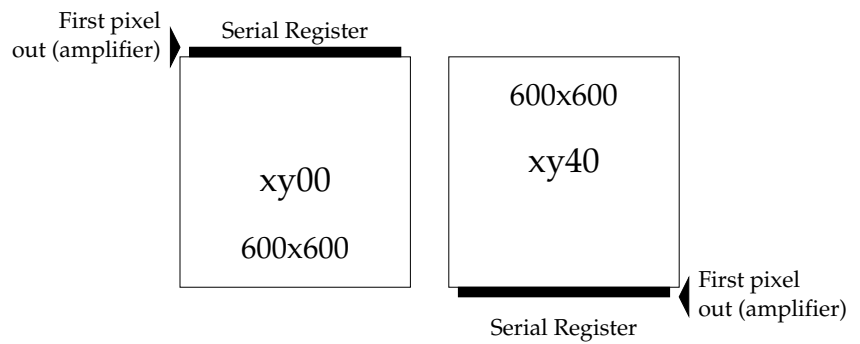


Figure 14: Global Addressing : One OTA, One Cell

2.4 Camera X,Y and Focal Plane Display

OTAs in the focal plane of a GPC are numbered so that Cartesian OTA00 is in the lower left, as seen through the dewar window. For the Pan-STARRS telescope, this is a mirror image of the sky, so a flip in X is made when displaying images for the user. The exception to this rule can be if the user wants to see North pointing up, but this is rarely at a convenient right angle on Pan-STARRS anyway.

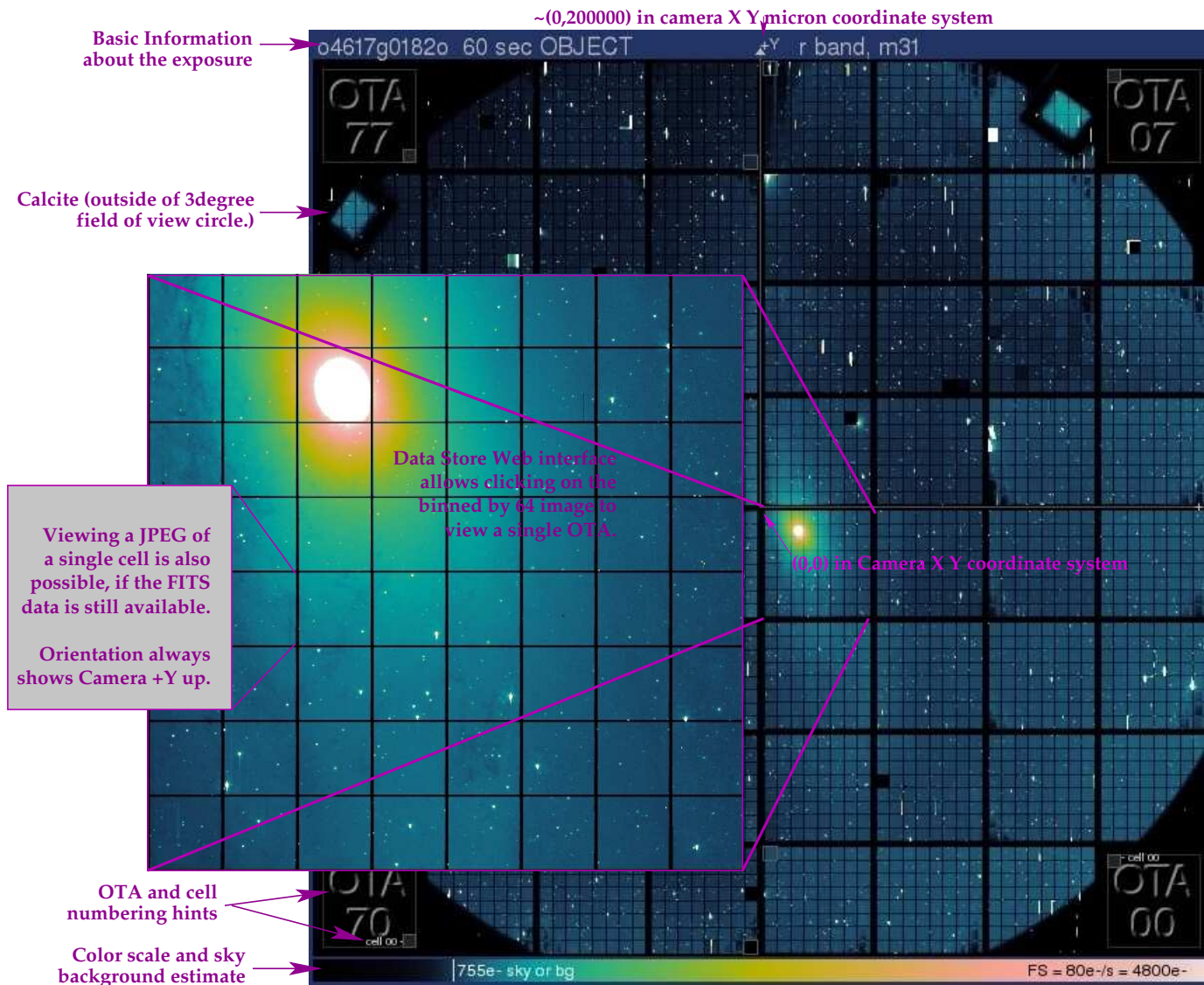


Figure 15: Image of the Sky on GPC1

A camera coordinate system has been defined for Pan-STARRS. This coordinate system uses units of **microns** and has its center at the boresite (which, incidentally, falls between the center four OTAs and not on a pixel.) Each binned by 64 quicklook image generated by **otatool** and a set of scripts that call **ImageMagick** documents this coordinate system, as well as the location of OTA00, and a few cell xy00. Figure 15 shows an example of the sky, seen through Pan-STARRS1 on GPC1.

2.5 Translation Table

The following tables entries for each OTA in a GPC or TC3. The first few columns are detected by the STARGRASP u-boot software:

- Domain name from DHCP server
- Chassis ID bits
- Slot ID bits

The subsequent columns are assigned based on the configuration of our cameras and controllers:

- The **Device bit** can be either 0 or 1, given that a single boardset can control up to two OTA devices.
- The unique Software Location is a combination of the boardset's name (e.g. "gpc4A1") which it has constructed from the previous columns) followed by the selection of device 0 or 1.
- Given Rigidflex geometry, a slot1or3-d0 always maps to ota1, slot1or3-d1 to ota3, slot2or4-d0 maps to ota2, and slot2or4-d1 maps to ota4 (as labelled on the rigidflex, "ota1 ota2 ota3 ota4" with ota1 being the farthest from the controller connector.
- Given focal plane configuration, both TC3 and a GPC will map to the two-digit otaNN identifiers given in the last column.

Chassis ID bits b0000 through b0111 are used by all Giga Pixel Cameras (The camera only figures out *which* GPC is it is from the DHCP server.) Chassis ID b1111 is special. It occurs by default if no chassis is present, and is also used by microcam. Other b1xxx IDs are assigned to testbenches and cameras as follows:

Chassis ID bits	Assignment
0 0 0 0	All GPCs, Chassis A or labA (use "laba" login)
0 0 0 1	All GPCs, Chassis B or labB (use "labb" login)
0 0 1 0	All GPCs, Chassis C or labC (use "labc" login)
0 0 1 1	All GPCs, Chassis D or labD (use "labd" login)
0 1 0 0	All GPCs, Chassis E or labE (use "labe" login)
0 1 0 1	All GPCs, Chassis F or labF (use "labf" login)
0 1 1 0	All GPCs, Chassis G or labG (use "labg" login)
0 1 1 1	All GPCs, Chassis H or labH (use "labh" login)
1 0 0 0	Test Camera 3, Chassis A (use "tc3" login)
1 0 0 1	Test Camera 3, Chassis B
1 0 1 0	Test Camera 2 use (use "daquser" login (TBD: make "tc2"))
1 0 1 1	Bench 2 (use "bench2" login)
1 1 0 0	Test Camera 1 use
1 1 0 1	Test Camera 1 use
1 1 1 0	Bench 1 (use "bench1" login)
1 1 1 1	μCam or no/default chassis ID

The most common use by software of this table would be to find out which SW Location to contact to control a given FP location. The table also shows the translation of Chassis ID bits to the Chassis letter (A through H) and the translation of Slot ID bits to slot number (which is the binary value *plus one*).

Here is the complete table for TC3:

Domain	Chassis ID bits	Slot ID bits	Device	OTA SW Location	Rigidflex Location	FP Location
tc3	1 0 0 0 = A(tc3)	0 0 = slot1	0	tc3A1 / d0	ota1	ota31
tc3	1 0 0 0 = A(tc3)	0 0 = slot1	1	tc3A1 / d1	ota3	ota11
tc3	1 0 0 0 = A(tc3)	0 1 = slot2	0	tc3A2 / d0	ota2	ota21
tc3	1 0 0 0 = A(tc3)	0 1 = slot2	1	tc3A2 / d1	ota4	ota01
tc3	1 0 0 0 = A(tc3)	1 0 = slot3	0	tc3A3 / d0	ota1	ota30
tc3	1 0 0 0 = A(tc3)	1 0 = slot3	1	tc3A3 / d1	ota3	ota10
tc3	1 0 0 0 = A(tc3)	1 1 = slot4	0	tc3A4 / d0	ota2	ota20
tc3	1 0 0 0 = A(tc3)	1 1 = slot4	1	tc3A4 / d1	ota4	ota00
tc3	1 0 0 1 = B(tc3)	0 0 = slot1	0	tc3B1 / d0	ota1	ota33
tc3	1 0 0 1 = B(tc3)	0 0 = slot1	1	tc3B1 / d1	ota3	ota13
tc3	1 0 0 1 = B(tc3)	0 1 = slot2	0	tc3B2 / d0	ota2	ota23
tc3	1 0 0 1 = B(tc3)	0 1 = slot2	1	tc3B2 / d1	ota4	ota03
tc3	1 0 0 1 = B(tc3)	1 0 = slot3	0	tc3B3 / d0	ota1	ota32
tc3	1 0 0 1 = B(tc3)	1 0 = slot3	1	tc3B3 / d1	ota3	ota12
tc3	1 0 0 1 = B(tc3)	1 1 = slot4	0	tc3B4 / d0	ota2	ota22
tc3	1 0 0 1 = B(tc3)	1 1 = slot4	1	tc3B4 / d1	ota4	ota02

The table for a GPC starts the same way, and adds chassis C,D,E,F,G, and H:

Domain	Chassis ID	Slot ID	Dev	OTA SW Location	Rigidflex	FP Loc.	Domain	Chassis ID	Slot ID	Dev	OTA SW Location	Rigidflex	FP Loc.
gpc1,2,3,4	0 0 0 0 = A	0 0 = slot1	0	gpc1A1 / d0	ota1	ota31	gpc1,2,3,4	0 1 0 0 = E	0 0 = slot1	0	gpc1E1 / d0	ota1	ota40
gpc1,2,3,4	0 0 0 0 = A	0 0 = slot1	1	gpc1A1 / d1	ota3	ota11	gpc1,2,3,4	0 1 0 0 = E	0 0 = slot1	1	gpc1E1 / d1	ota3	ota60
gpc1,2,3,4	0 0 0 0 = A	0 1 = slot2	0	gpc1A2 / d0	ota2	ota21	gpc1,2,3,4	0 1 0 0 = E	0 1 = slot2	0	gpc1E2 / d0	ota2	ota50
gpc1,2,3,4	0 0 0 0 = A	0 1 = slot2	1	gpc1A2 / d1	ota4	ota01	gpc1,2,3,4	0 1 0 0 = E	0 1 = slot2	1	gpc1E2 / d1	ota4	ota70
gpc1,2,3,4	0 0 0 0 = A	1 0 = slot3	0	gpc1A3 / d0	ota1	ota30	gpc1,2,3,4	0 1 0 0 = E	1 0 = slot3	0	gpc1E3 / d0	ota1	ota41
gpc1,2,3,4	0 0 0 0 = A	1 0 = slot3	1	gpc1A3 / d1	ota3	ota10	gpc1,2,3,4	0 1 0 0 = E	1 0 = slot3	1	gpc1E3 / d1	ota3	ota61
gpc1,2,3,4	0 0 0 0 = A	1 1 = slot4	0	gpc1A4 / d0	ota2	ota20	gpc1,2,3,4	0 1 0 0 = E	1 1 = slot4	0	gpc1E4 / d0	ota2	ota51
gpc1,2,3,4	0 0 0 0 = A	1 1 = slot4	1	gpc1A4 / d1	ota4	ota00	gpc1,2,3,4	0 1 0 0 = E	1 1 = slot4	1	gpc1E4 / d1	ota4	ota71
gpc1,2,3,4	0 0 0 1 = B	0 0 = slot1	0	gpc1B1 / d0	ota1	ota33	gpc1,2,3,4	0 1 0 1 = F	0 0 = slot1	0	gpc1F1 / d0	ota1	ota42
gpc1,2,3,4	0 0 0 1 = B	0 0 = slot1	1	gpc1B1 / d1	ota3	ota13	gpc1,2,3,4	0 1 0 1 = F	0 0 = slot1	1	gpc1F1 / d1	ota3	ota62
gpc1,2,3,4	0 0 0 1 = B	0 1 = slot2	0	gpc1B2 / d0	ota2	ota23	gpc1,2,3,4	0 1 0 1 = F	0 1 = slot2	0	gpc1F2 / d0	ota2	ota52
gpc1,2,3,4	0 0 0 1 = B	0 1 = slot2	1	gpc1B2 / d1	ota4	ota03	gpc1,2,3,4	0 1 0 1 = F	0 1 = slot2	1	gpc1F2 / d1	ota4	ota72
gpc1,2,3,4	0 0 0 1 = B	1 0 = slot3	0	gpc1B3 / d0	ota1	ota32	gpc1,2,3,4	0 1 0 1 = F	1 0 = slot3	0	gpc1F3 / d0	ota1	ota43
gpc1,2,3,4	0 0 0 1 = B	1 0 = slot3	1	gpc1B3 / d1	ota3	ota12	gpc1,2,3,4	0 1 0 1 = F	1 0 = slot3	1	gpc1F3 / d1	ota3	ota63
gpc1,2,3,4	0 0 0 1 = B	1 1 = slot4	0	gpc1B4 / d0	ota2	ota22	gpc1,2,3,4	0 1 0 1 = F	1 1 = slot4	0	gpc1F4 / d0	ota2	ota53
gpc1,2,3,4	0 0 0 1 = B	1 1 = slot4	1	gpc1B4 / d1	ota4	ota02	gpc1,2,3,4	0 1 0 1 = F	1 1 = slot4	1	gpc1F4 / d1	ota4	ota73
gpc1,2,3,4	0 0 1 0 = C	0 0 = slot1	0	gpc1C1 / d0	ota1	ota35	gpc1,2,3,4	0 1 1 0 = G	0 0 = slot1	0	gpc1G1 / d0	ota1	ota44
gpc1,2,3,4	0 0 1 0 = C	0 0 = slot1	1	gpc1C1 / d1	ota3	ota15	gpc1,2,3,4	0 1 1 0 = G	0 0 = slot1	1	gpc1G1 / d1	ota3	ota64
gpc1,2,3,4	0 0 1 0 = C	0 1 = slot2	0	gpc1C2 / d0	ota2	ota25	gpc1,2,3,4	0 1 1 0 = G	0 1 = slot2	0	gpc1G2 / d0	ota2	ota54
gpc1,2,3,4	0 0 1 0 = C	0 1 = slot2	1	gpc1C2 / d1	ota4	ota05	gpc1,2,3,4	0 1 1 0 = G	0 1 = slot2	1	gpc1G2 / d1	ota4	ota74
gpc1,2,3,4	0 0 1 0 = C	1 0 = slot3	0	gpc1C3 / d0	ota1	ota34	gpc1,2,3,4	0 1 1 0 = G	1 0 = slot3	0	gpc1G3 / d0	ota1	ota45
gpc1,2,3,4	0 0 1 0 = C	1 0 = slot3	1	gpc1C3 / d1	ota3	ota14	gpc1,2,3,4	0 1 1 0 = G	1 0 = slot3	1	gpc1G3 / d1	ota3	ota65
gpc1,2,3,4	0 0 1 0 = C	1 1 = slot4	0	gpc1C4 / d0	ota2	ota24	gpc1,2,3,4	0 1 1 0 = G	1 1 = slot4	0	gpc1G4 / d0	ota2	ota55
gpc1,2,3,4	0 0 1 0 = C	1 1 = slot4	1	gpc1C4 / d1	ota4	ota04	gpc1,2,3,4	0 1 1 0 = G	1 1 = slot4	1	gpc1G4 / d1	ota4	ota75
gpc1,2,3,4	0 0 1 1 = D	0 0 = slot1	0	gpc1D1 / d0	ota1	ota37	gpc1,2,3,4	0 1 1 1 = H	0 0 = slot1	0	gpc1H1 / d0	ota1	ota46
gpc1,2,3,4	0 0 1 1 = D	0 0 = slot1	1	gpc1D1 / d1	ota3	ota17	gpc1,2,3,4	0 1 1 1 = H	0 0 = slot1	1	gpc1H1 / d1	ota3	ota66
gpc1,2,3,4	0 0 1 1 = D	0 1 = slot2	0	gpc1D2 / d0	ota2	ota27	gpc1,2,3,4	0 1 1 1 = H	0 1 = slot2	0	gpc1H2 / d0	ota2	ota56
gpc1,2,3,4	0 0 1 1 = D	0 1 = slot2	1	gpc1D2 / d1	ota4	ota07	gpc1,2,3,4	0 1 1 1 = H	0 1 = slot2	1	gpc1H2 / d1	ota4	ota76
gpc1,2,3,4	0 0 1 1 = D	1 0 = slot3	0	gpc1D3 / d0	ota1	ota36	gpc1,2,3,4	0 1 1 1 = H	1 0 = slot3	0	gpc1H3 / d0	ota1	ota47
gpc1,2,3,4	0 0 1 1 = D	1 0 = slot3	1	gpc1D3 / d1	ota3	ota16	gpc1,2,3,4	0 1 1 1 = H	1 0 = slot3	1	gpc1H3 / d1	ota3	ota67
gpc1,2,3,4	0 0 1 1 = D	1 1 = slot4	0	gpc1D4 / d0	ota2	ota26	gpc1,2,3,4	0 1 1 1 = H	1 1 = slot4	0	gpc1H4 / d0	ota2	ota57
gpc1,2,3,4	0 0 1 1 = D	1 1 = slot4	1	gpc1D4 / d1	ota4	ota06	gpc1,2,3,4	0 1 1 1 = H	1 1 = slot4	1	gpc1H4 / d1	ota4	ota77

3 PS1 Computing Network

Three major logical networks exist to support the GPC Camera at the PS1 facility. These networks are:

- Internet
- External Camera Network
- Internal Camera Network

The design for a PS4 remains essentially the same, except that there are multiple, isolated Internal Camera Networks, and a subnetted External Camera Network with a DHCP server for each subnet.

3.1 Logical Network (recommended)

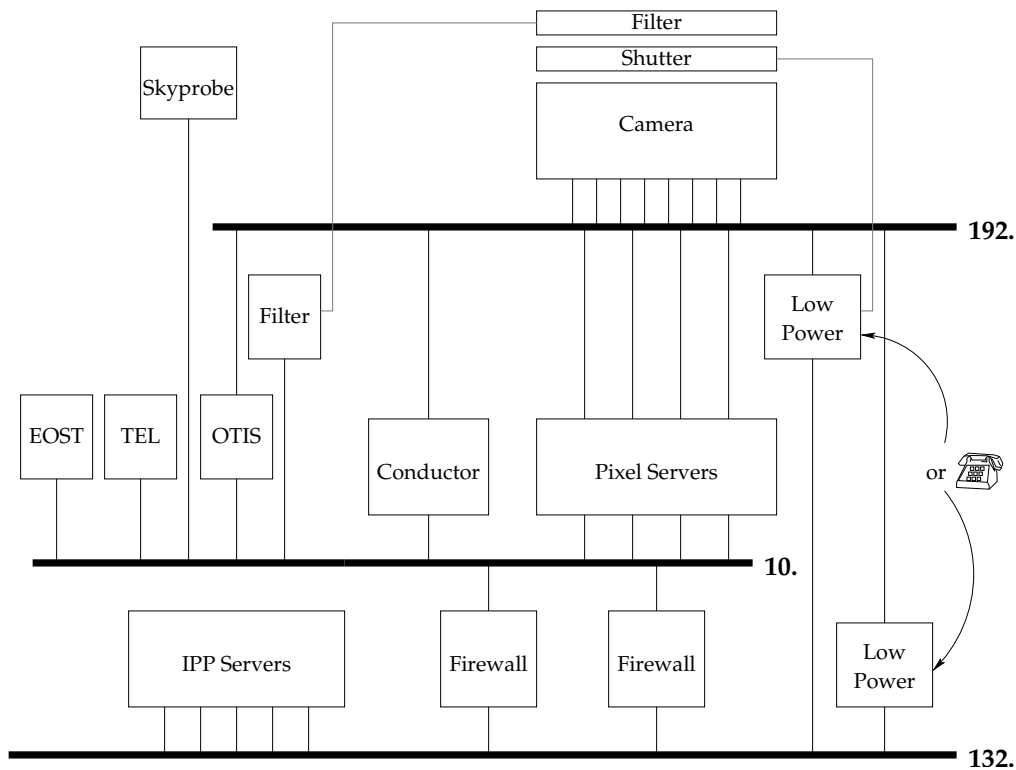


Figure 16: Summit Logical Networks (Recommended)

Ideally, the only hosts connected to the Internet would be the Firewall Servers. As necessary, these Firewall Servers may be configured to provide port forwarding and network address translation (NAT) to allow traffic in and out of hosts on the Camera Networks. Additional computers on the Internet increase security vulnerability. If they cannot be avoided, they should definitely be configured *not* to forward traffic between any other network to which they interface.

Since each box represents a computer or Ethernet client, and the heavy black lines represent a switch or group of switches, it is instructive to see what the backup path is for each possible outage, and what the impacts for performance and functionality might be.

- **Pixel Servers** represent a cluster of computers. No individual pixel server has any dedicated task, so a failure of one or two pixel servers has only one impact on operations. When a given pixel server is unavailable, if it contained the only copy of any previously taken files, those files will remain inaccessible until the pixel server is repaired. The data store disks in the pixel servers are redundant (RAID-5), so the chance of permanent loss of data is reduced. If temporary inability to access data due to a power supply or other such failure is an issue, the system must be modified to keep multiple copies of the data on different machines (for a shorter retention period, given the same total amount of disk space.)
- **Conductor** can be of the same hardware configuration as any Pixel Server. It is also connected to the network in the same way as a Pixel Server. It is the host chosen to make all decisions about which Pixel Servers to use for taking the next image, so its function is critical, but that function can be reassigned to any one of the Pixel Servers manually (but remotely) in the event of a failure. In contrast, failure of a Pixel Server at any time will ideally be handled automatically (by Conductor) without any intervention. To reduce the chances of a failure of Conductor, it makes sense to consider adding redundant power supplies and choosing a very reliable machine for this role.
- **Firewall Servers** will number at least 2 and at most 4 for PS1. They could be the same hardware as the Pixel Server, but without the large amount of disk space. Since the Firewall Servers are interchangeable, the failure of one should not present any problems as long as all clients of the Firewall (outside sources downloading Data Store files) try more than one.
- **Low-Power (Support Building)** exists only as a backup itself. If it alone has a failure, there is no impact. It will interface to two or more networks, answer incoming modem calls, and be powered off a UPS that will last a few hours. For GPC1, it will be given direct control of remote power outlets supplying each pixel server, conductor, and the switches.
- **Low-Power (PSE Rack)** controls the shutter. If it fails, someone has to physically swap it with the Low-Power from the Support Building rack. Alternatives to explore: there are multiple ways to trigger the shutter. A STAR-GRASP board in the camera could also be connected to the shutter to provide an alternate trigger. Impact to observing may only be somewhat degraded timestamp accuracy for shutter open/close events, since the PSE Rack Low-Power unit will also connect to a GPS to make it a stratum 1 time source.
- **Shutter**, and **Filter** are one of a kind and include no redundancy.
- **Camera** is comprised of many individual CPU/DAQ boardsets and could still operate with the loss of one. The cost of each non-functional boardset would be 2 OTA devices, until the boardset can be replaced with a spare.

The above covers the computers and hardware directly involved in operating the Giga Pixel Camera. Next, we consider what would happen during the failure of any of the interconnecting networks. This loss of a network is treated simplistically as the failure of the entire network. There may be cases where an single switch fails and the impact is less than what is described here, but if we can live with the following, these cases outline the most severe impacts related to network failures. Network failures may include a malfunctioning or misconfigured client that makes a network segment unusable, or a hardware failure that takes out a switch or damages copper cables or fibers. The root cause of the problem only affects the repair, not the way the design handles the failure to continue operating.

- **Internet.** In our recommended logical network design, problems with the internet, the switch on the internet subnet, or the cabling to that switch would all, at worst, prevent outside clients access to the observatory. The backup way of access would be a significantly slower dial modem connection. While this connection would be totally unsuitable for camera image transfers, it will still be sufficient to close shutters, turn off power, and even to keep taking new image data through Conductor's console interface, or by communicating with OTIS and having it continue to observe.
- **External Camera Network** is a private network used by Conductor to communicate with the Pixel Servers and provide them with their NFS root filesystems. It is also the path for Data Store files to reach the Firewall Servers and be transferred out via HTTP. In the worst case, the entire failure of this network would mean that the Firewall Servers have no path with sufficient bandwidth to extract Data Store files for an external client. New data can still be acquired because NFS-root and other services can still be provided either by Conductor, or by low-power, over

the Internal Camera Network. In this scenario there is a chance of slight performance degradation because this traffic will now compete with new data being transferred from the Camera. Such a failure will most likely require rebooting all of the Pixel Servers to force them to switch to the new NFS-root server (however, for NFS-read, there is a way to mount two servers at once and it should switch automatically. We should test this. Combined with the Status Server's ability to have a mirror, it could potentially make failover in the case of External Camera Network *and/or* Conductor failure seamless.) Risks of failure in the External Camera Network may be further mitigated by network topologies. A tentative physical implementation is included below, but we have yet to receive the switches and test several possible configurations.

- **Internal Camera Network** exists only to connect Camera to the Pixel Servers. Conductor is also on this network so it can assign DHCP addresses to the controller boardsets and receive status server updates from them. This network must be optimized for heavy bursts of data from Camera to Pixel Servers. The one and only impact this network could have is a serious one: inability to take any new data if it fails completely. Such a failure is unlikely because of the physical implementation of this network. At the camera end, this network consists of one fiber-pair run for each camera boardset connecting to a group of switches in the Support Building. If a fiber fails, contact is lost with one boardset and 2 OTAs become inaccessible. If a switch fails, multiple boardsets become inaccessible until those fibers and Ethernet cables are physically moved over to spare switches mounted next to the ones normally in use. The Internal Camera Network is also the path by which OTIS sends instructions to and reads recommended guide corrections from Conductor. A failure on this path will require OTIS to contact Conductor by the External Camera Network.
- **Other.** This is a Camera document. We have not considered what happens if EOST, Telescope, OTIS, Skyprobe, Filter, or their connections have a failure, other than that OTIS has two paths it can use to reach Conductor. (NOTE: If OTIS decides to connect directly the Internet instead of the External Camera Network with its other network interface, the backup OTIS-Conductor path then must involve the Firewall Servers. Guide corrections may begin to suffer from added latencies.)

3.2 Logical Network (actual, GPC1)

As discussed above, if OTIS desires to configure all of its computers with directly routable Internet addresses, the backup connection between OTIS and Conductor needs to pass through the Firewall Server.

A small Summit IPP cluster will also be located on the Internet and will access GPC1 data using the same path as a remote IPP cluster would.

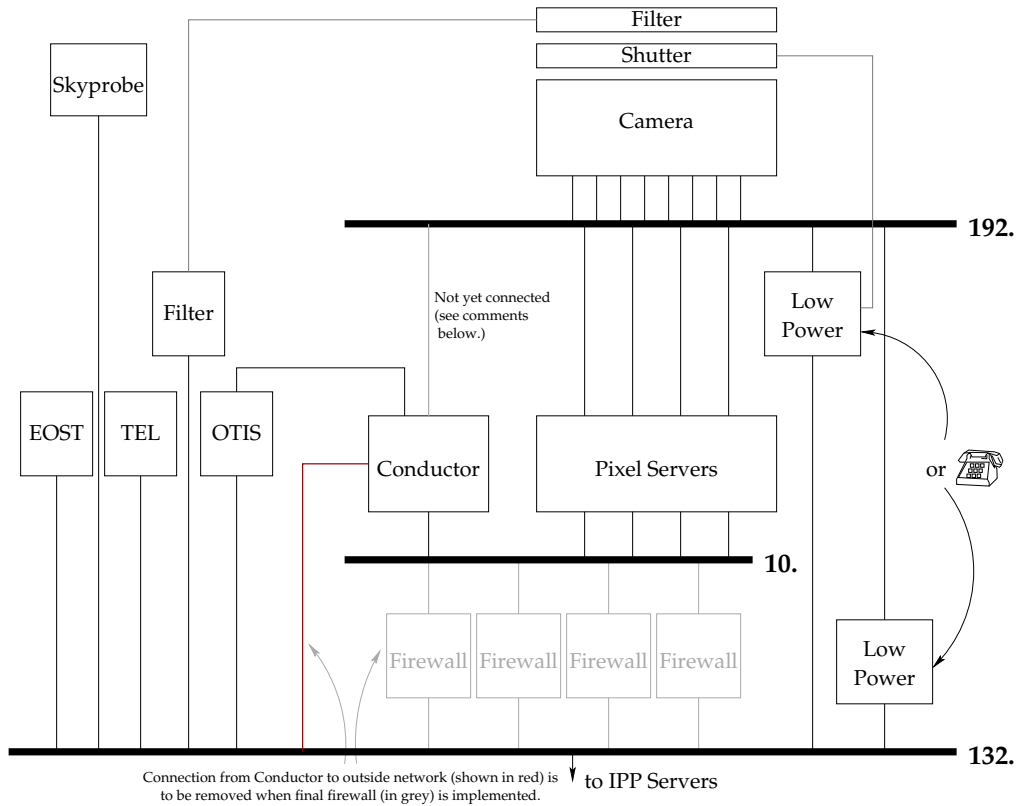


Figure 17: Summit Logical Networks - GPC1

3.3 Logical Network (actual, TC3)

TC3 will use between 2 to 4 active Pixel Servers. This could end up having no impact on the logical network, except that we also will not have the Firewall Servers in place. Instead, the Pixel Servers will take on their functions as well, connecting the Pixel Servers / Data Store directly to the Internet eliminating the entire External Camera Network for the TC3 phase. TC3 will allow direct NFS access (and possibly also HTTP to simulate what the Firewall Servers will provide) to IPP for all data store files with no bandwidth enforcement. Accessing the data could have a small impact on the speed at which new current data is acquired.

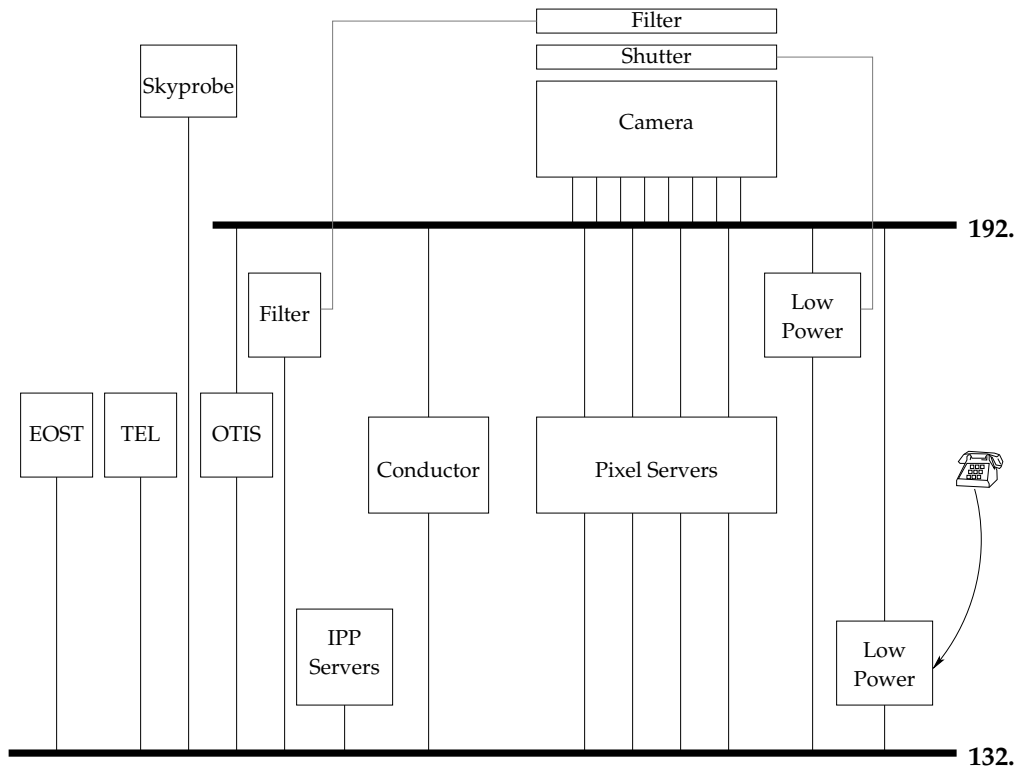


Figure 18: Summit Logical Networks - TC3

3.4 Physical Network (PS1)

The difference between TC3 and GPC1 networks will be that TC3 will have only a subset of the hardware installed, and slightly different configuration to implement the differences between the TC3 and GPC1 logical networks previously described. More details about this later.

3.4.1 Fiber Connections for STARGRASP Controllers

The most expensive component of the summit network is the fiber and switches that will implement the 192 network. We now have enough fibers (without any redundancy though, if we have only fibers now installed in the cable wrap) to keep all of the Internal Camera Network switch hardware in the support building. Other designs would have involved switches in both locations.

A new switch model from Dell might be a reasonable option for this network. Other manufacturers produce similar products with 4 fiber ports, and 44 or 48 copper gigabit ports. The **Dell PowerConnect 2748** switch is a “web managed” model with only minimal VLAN support and no spanning-tree capabilities. The current plan is to use two of these 48-port switches in their **unmanaged mode**, one each to run the Internal Camera Network and the External Camera Network. Since each 2748 switch only has 4 fiber ports, and the Giga Pixel Camera requires 32, one option might be to use 28 fiber media converters in front of the switch. This would only leave exactly 16 ports for Pixel Servers, but we may want as many as 18 permanently connected to have spares ready. Conductor also needs a connection. Also, 4 fiber media converters cost about the same simply buying another 2748 switch and 4 SFP modules for the job (and is a lot neater to mount in the rack.) Since each fiber link only needs to carry 256 Mbps of traffic, we can use the 2748’s to combine data from 2 to 4 fibers into one copper going into the unmanaged 2748. This means that the “media converter” 2748’s (there will be 8 of them so that the main 2748 can harbor spare SFPs ready to accept a fiber) will need to operate in managed mode. But Link Aggregation is not necessary, and neither is trunking. All that is required of the managed mode is to create some local VLANs on the switch that essentially chop the switch up into smaller separate switches, functionally. The following figure shows how one 2748 switch handles the fiber needs of an entire STARGRASP chassis (4 boardsets) and merges the data into 2 Gigabit copper links that will go to the main, unmanaged Internal Camera Network switch:

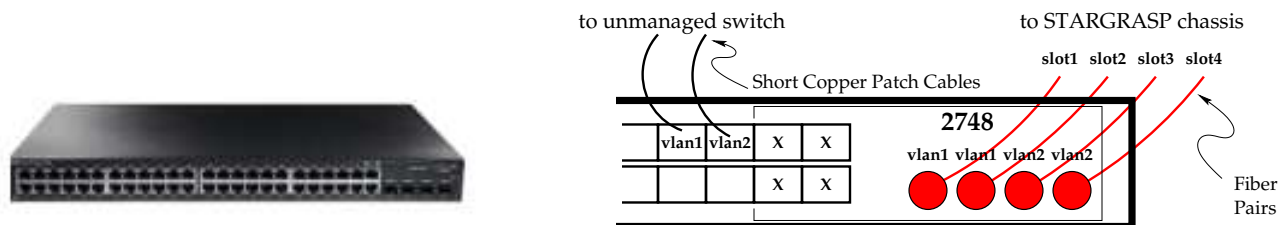


Figure 19: Dell 2748 Managing 1 STARGRASP Chassis

With this configuration, we should be able to achieve data rates close to 500 Mbps¹ per boardset when all are transmitting data at once. *This is the first thing we need to test as soon as we receive the switches.* It leaves 4 SFPs on the unmanaged 2748 free as backup, and 28 copper Gigabit ports on the same switch for Conductor and all the Pixel Servers.

If any of the 8 managed 2748’s fail, 8 OTAs in the focal plane become inaccessible until someone moves the fibers over to the unmanaged switch. No software reconfiguration is needed after this change to continue with a fully functional camera.

¹ 500 Mbps is almost twice the target (256 Mbps), though we may need this extra bandwidth to handle a 32 bit (2x) data stream between the controller and Pixel Servers. It is slightly below what we have been able to achieve with current STARGRASP implementations (624 Mbps). However, being able to handle the full 624 Mbps (or higher rates that may be possible with 9000-byte Ethernet packets) is not useful with our current Pixel Server platform, since a single Pixel Server will need to handle the data from at least 2 STARGRASP boardsets, and it would not be possible to accept more than 1000 Mbps of incoming data using only 1 Ethernet.

If the unmanaged 2748 switch fails, the following additions will make it so that the second unmanaged 2748 running the External Camera Network can take over: Each of the “media converter” managed 2748’s will have two more of their

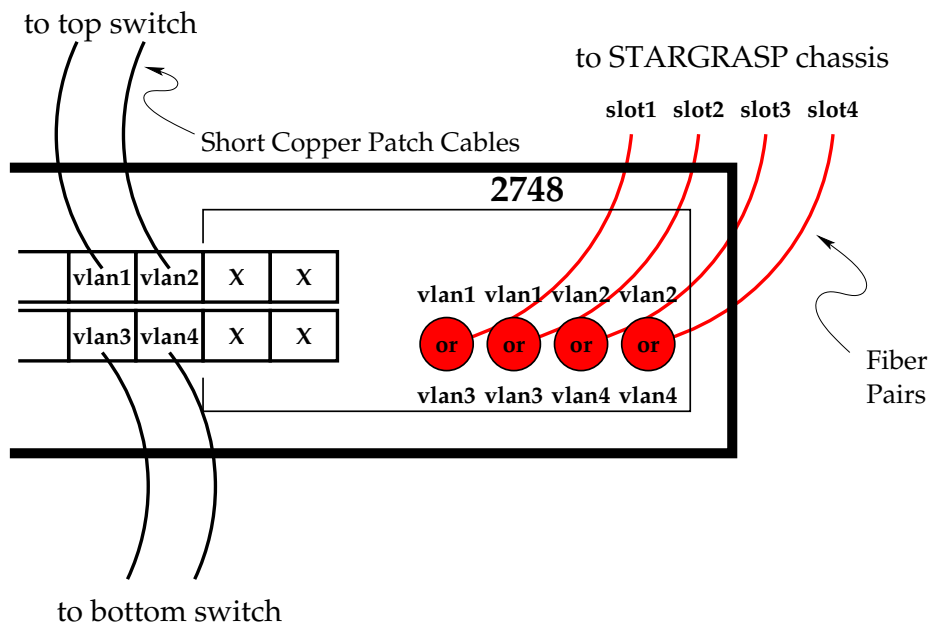


Figure 20: 1 4-port Fiber Switch per Chassis with Redundancy

copper ports connected to this External Camera Network in a symmetric fashion to the connections to the Internal Camera Network switch. All clients that are connected to the Internal are also connected by the second interface to the External Network. The only disadvantage to using the External Network for transferring pixel data from the camera is that this traffic will now compete with outgoing Data Store downloads. Since the 2748’s do not support any network topologies that include loops, the switch to the External network requires some intervention, but it can be done remotely: For each of the 8 “media converter” switches, membership of the two fiber ports in vlan1 must be switched to vlan3, and the two fiber ports in vlan2 must be reconfigured to be in vlan4. After this change, the camera must be rebooted so that each STARGRASP boardset will reacquire a DHCP on the new (external) camera subnet. All other communications are based on Status Server entries on Conductor and should switch automatically.

The following figure includes all of the switches and computers necessary to operate the Camera, and also the how the OTIS and Telescope computers will get their network connectivity through Camera’s fiber bundles and switches in the dome. All switches can operate as normal, unmanaged switches with no spanning tree support, except for the 8 4-fiber-port switches connected to the STARGRASP chassis. Before connecting these 8 switches as drawn, they must be configured with the necessary VLANs first.

3.5 Summary of Fault Tolerances

Failure of ...	Immediate Physical Intervention	Remote Intervention?	Down Time?	Loss of Camera/Data Store Performance?
Pixel Server	No	No	No	Percentage of Data Store affected
Conductor	No	Yes	Minimal	No
Firewall Server	No	No	No	No
Low Power (support)	No	No	No	No
Low Power (PSE)	Yes ²	Yes	Yes	Possible loss of GPS accuracy
Camera Boardset	No	No	No	Loss of 2 OTAs
132.x switch or Internet	No	No	No	No more Data Store access
10.x switch or cabling	No	Yes ³	No	Degraded Data Store speed
192.x switch or cabling	No	Yes ⁴	Minimal	Minimal
Fiber or fiber switch	Yes ⁵	No	No	Temporary Loss of 8 OTAs
Facility Power	No	No	Yes	–

“Immediate physical intervention” refers to someone at the facility moving things around to recover from the failure. “Remote intervention” can be performed over the internet, or over the dial-up connection. If both of these are “No”, then system recovery is intended to be fully automatic. All of the above require eventual repair of the failure hardware to maintain redundancy.

Power control and management during power outages is not covered by these figures. Separate power wiring diagrams indicating the locations of and connections to remote control power strips need to be drawn. The Low-Power servers will have their own UPSes, and will be in direct control of these power strips. Lost time is as long as the duration of the power outage since the UPSes are sized to provide only enough power to shut things down cleanly.

²Unless secondary means to trigger shutter with STARGRASP is available.

³Unless redundant NFS-root works

⁴Log in and reconfigure VLANs

⁵Move affected fibers to the unmanaged switch

