peralex
the sense of signals

SKARAB Agile Extreme-Scale Networked Compute Platform

peralex
the sense of signals

Overview of Presentation

- What is SKARAB?
- Designing/manufacturing SKARAB
- Demo of SKARAB platform
- Extreme scale intelligent storage (a SKARAB enabler)
- Data Acquisition Prototype

What is the “SKARAB”?

- SKARAB is an energy-efficient, agile, extreme scale, networked, FPGA-centric cluster computing node, with 19” 1U dimension
- SKARAB = Square Kilometre Array Reconfigurable Application Board
- SKARAB is the next generation/successor to the ROACH2 platform, conceptualised by SKA-SA
- Peralex won the nationwide tender to design and manufacture 300 SKARAB units for the MeerKAT precursor to the SKA radio-telescope in South Africa
- SKARAB is not application-specific

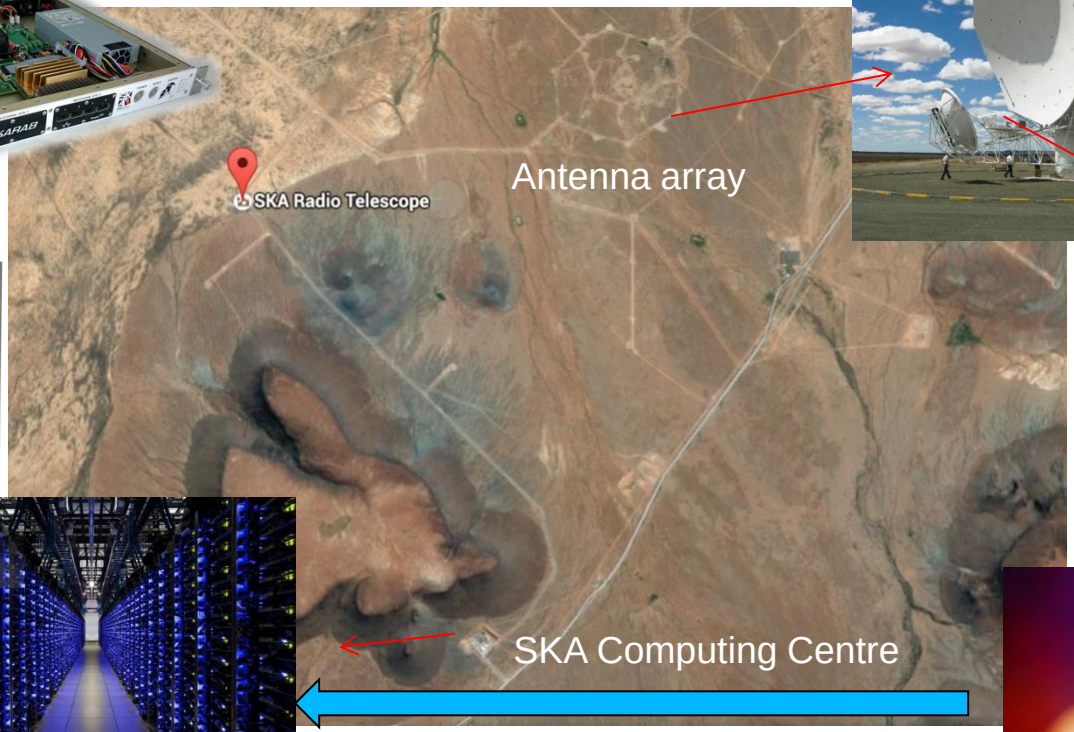
What is the “SKARAB”?

peralex
the sense of signals

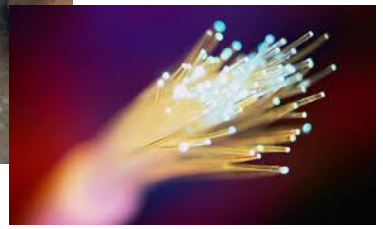
x 300

MeerKAT Radio-telescope

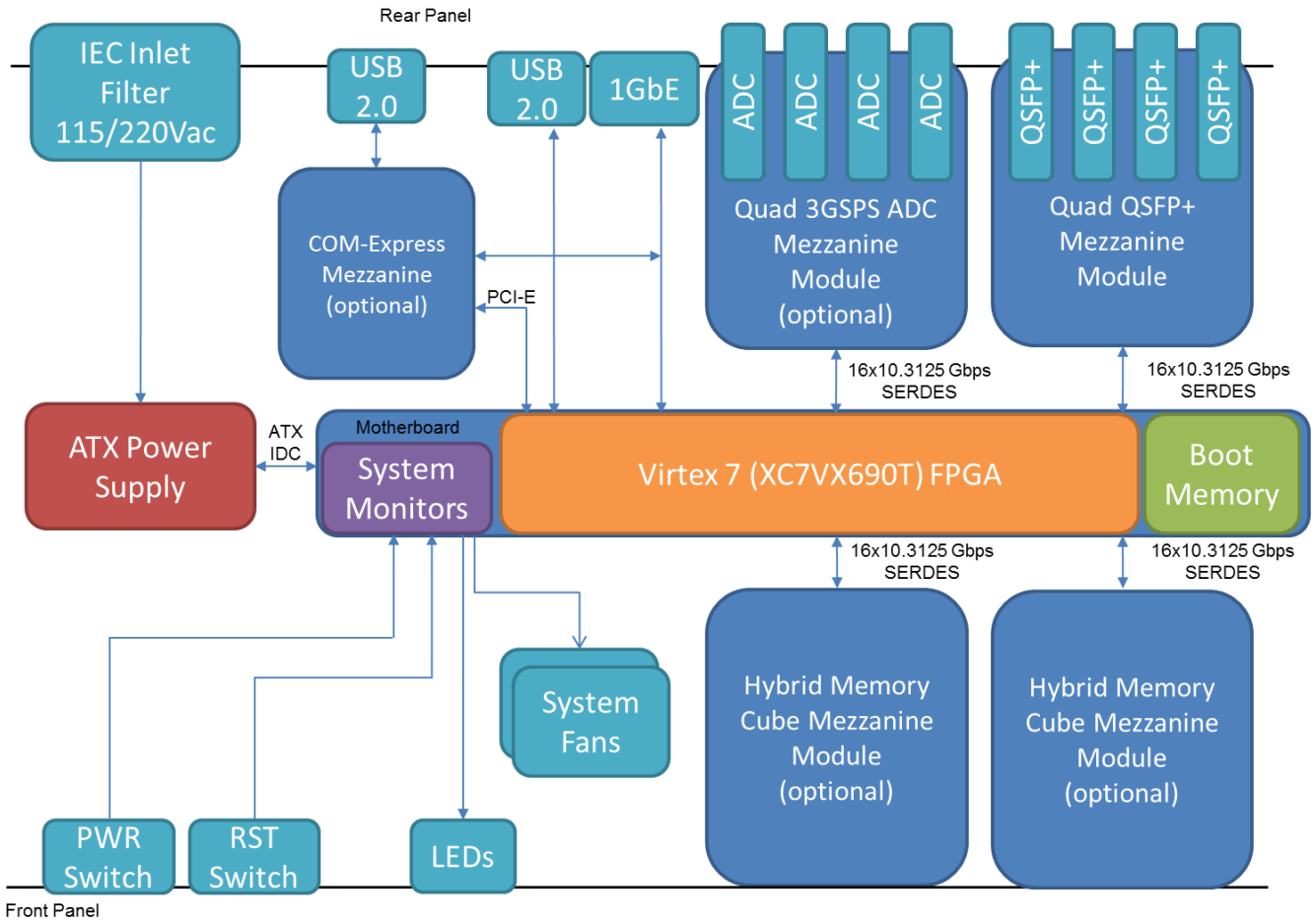
x 64



x 4

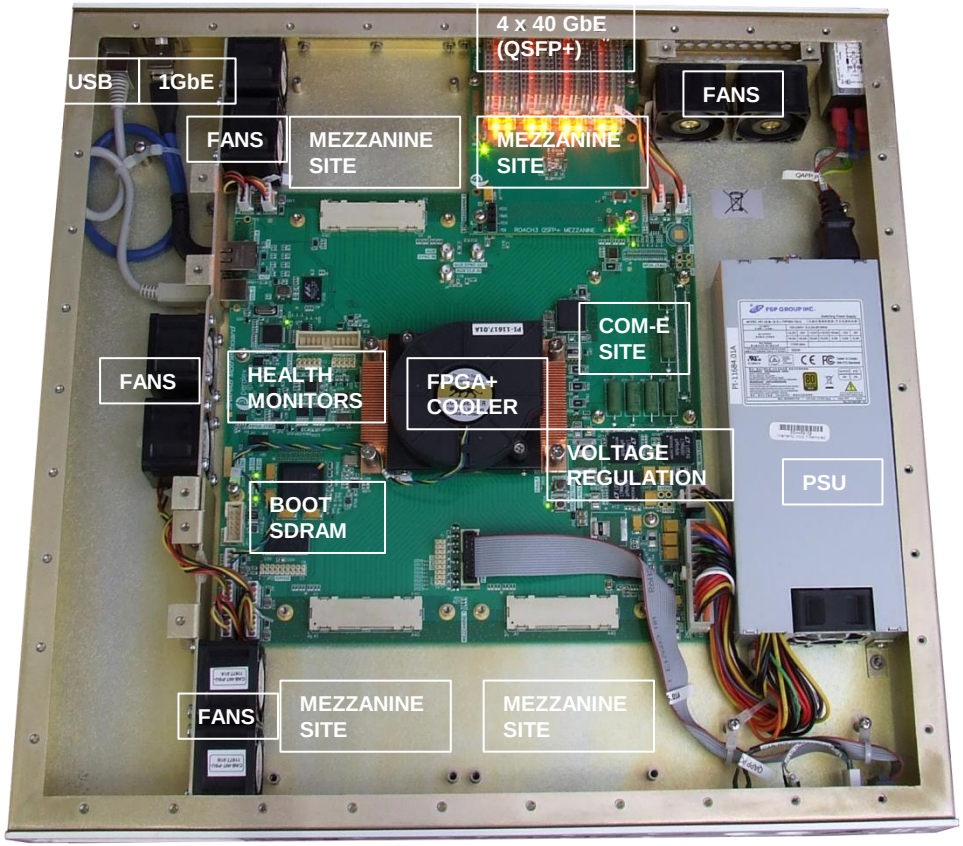


SKARAB Block Diagram



SKARAB: Under The Hood

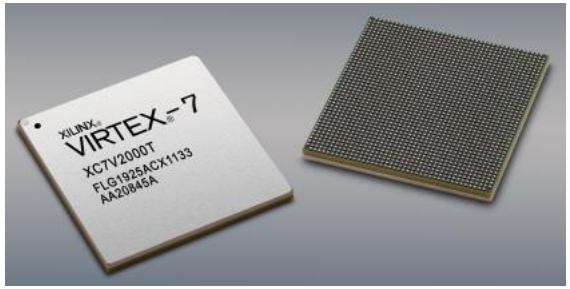
peralex
the sense of signals



SKARAB Compute: Virtex 7 FPGA

peralex
the sense of signals

- 693120 Logic Cells
- 80 x SERDES
  - 64 used for I/O mezzanine sites @ 10 Gbps
  - 1 used for PCI-E to COM-E module site
- 1470 x 36Kb RAM Blocks (~52 Mb)
- 3600 DSP Slices
- 1927 pins

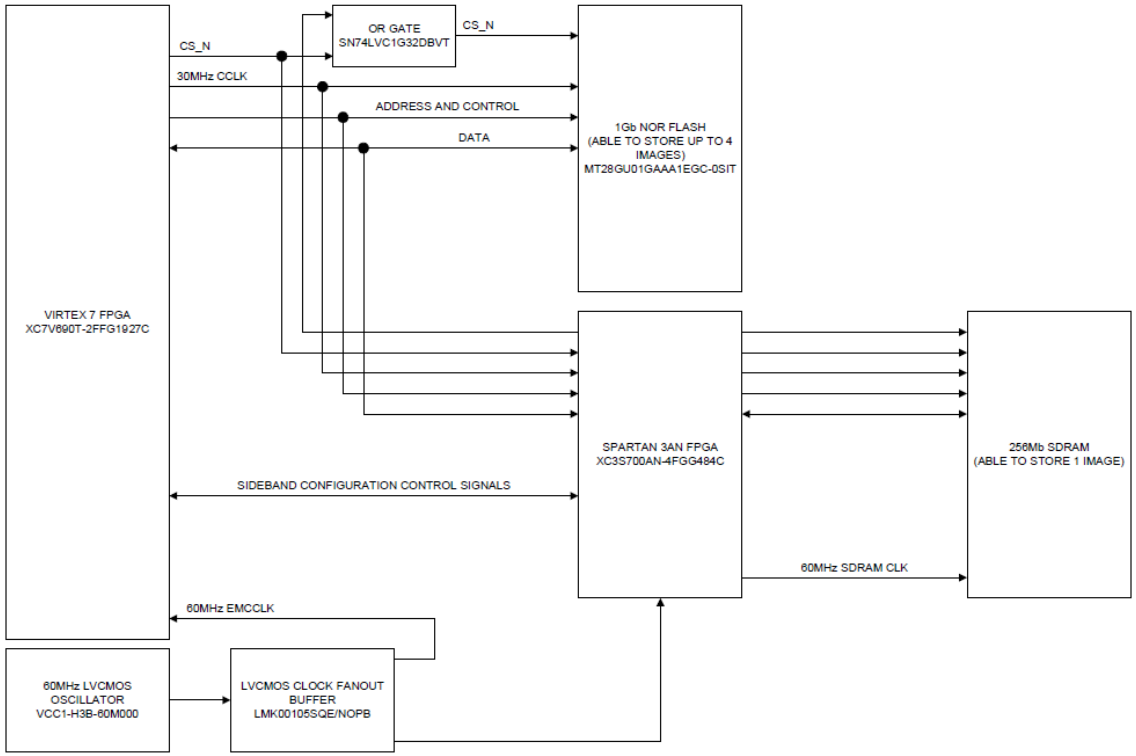


	Part Number	XC7VX690T
	EasyPath™ Cost Reduction Solutions[1]	XCE7VX690T
Logic Resources	Slices	108,300
	Logic Cells	693,120
	CLB Flip-Flops	866,400
Memory Resources	Maximum Distributed RAM (Kb)	10,888
	Block RAM/FIFO w/ ECC (36 Kb each)	1,470
	Total Block RAM (Kb)	52,920
Clocking	CMTs (1 MMCM + 1 PLL)	20
I/O Resources	Maximum Single-Ended I/O	1,000
	Maximum Differential I/O Pairs	480
Integrated IP Resources	DSP Slices	3,600
	PCIe® Gen2[2]	—
	PCIe Gen3	3
	Analog Mixed Signal (AMS) / XADC	1
	Configuration AES / HMAC Blocks	1
	GTX Transceivers (12.5 Gb/s Max Rate)[3]	—
	GTH Transceivers (13.1 Gb/s Max Rate)[4]	80
	GTZ Transceivers (28.05 Gb/s Max Rate)	—
Speed Grades	Commercial	-1, -2
	Extended[5]	-2L, -3
	Industrial	-1, -2

	Package[6]	Dimensions (mm)	GTH)
	FFG1157 / FFV1157[7]	35 x 35	0, 600 (0, 20)
Footprint Compatible	FFG1761 / FFV1761[7]	42.5 x 42.5	0, 850 (0, 36)
	FHG1761	45 x 45	
	FLG1925	45 x 45	
	FFG1158 / FFV1158[7]	35 x 35	0, 350 (0, 48)
Footprint Compatible	FFG1926	45 x 45	0, 720 (0, 64)
	FLG1926	45 x 45	
	FFG1927 / FFV1927[7]	45 x 45	0, 600 (0, 80)
Footprint Compatible	FFG1928	45 x 45	
	FLG1928	45 x 45	
Footprint Compatible	FFG1930	45 x 45	0, 1000 (0, 24)
	FLG1930	45 x 45	

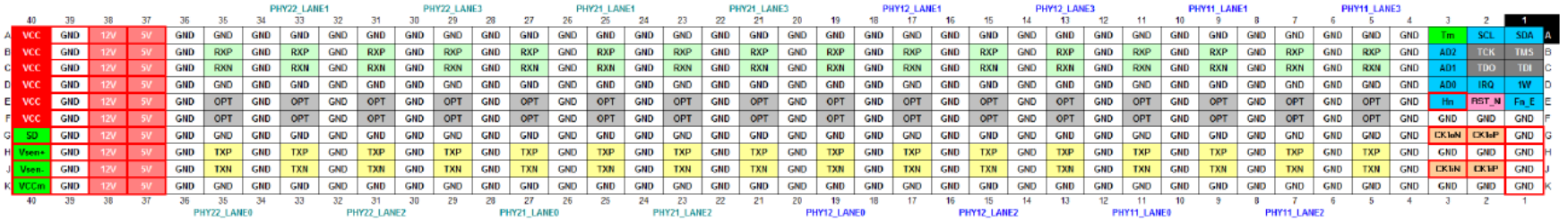
SKARAB: FPGA Configuration

- Requirement to boot in under 1 second
- On power-up, FPGA boots from NV on-board boot flash (boot image and backup image)
- After power-up FPGA can be rebooted over 1GbE network interface using SDRAM-based high-speed boot mode



SKARAB Mezzanine Sites

- Four 400 pin FCI Megarray connector (28 Gbps capable)
- 16 x ~10 Gbps SERDES from FPGA
- 1-Wire Interface (Configuration PROM)
- Low speed I2C Management interface, PD, Fault
- JTAG interface (test)
- 12V (bulk), 5V, 3.3V
- High speed Clocks (in/out)



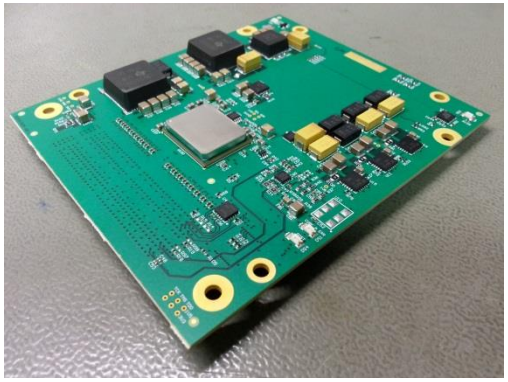
SKARAB Mezzanine: 4 x 40 GbE QSFP+

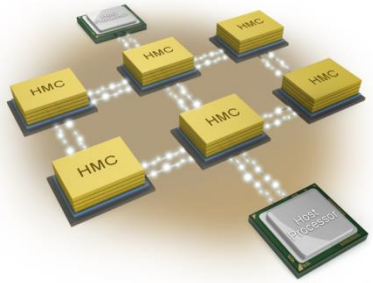
- Two variants (direct vs retimed)
- 4 x QSFP+ interfaces (16 x 10 Gbps)
  - Support for Copper, AOC, SR, LR fibre cables
- 32-bit ARM uC (management, PHY boot)
- Thermal sensor
- Configuration PROM
- Clock generation

SKARAB Mezzanine: Hybrid Memory Cube

peralex
the sense of signals

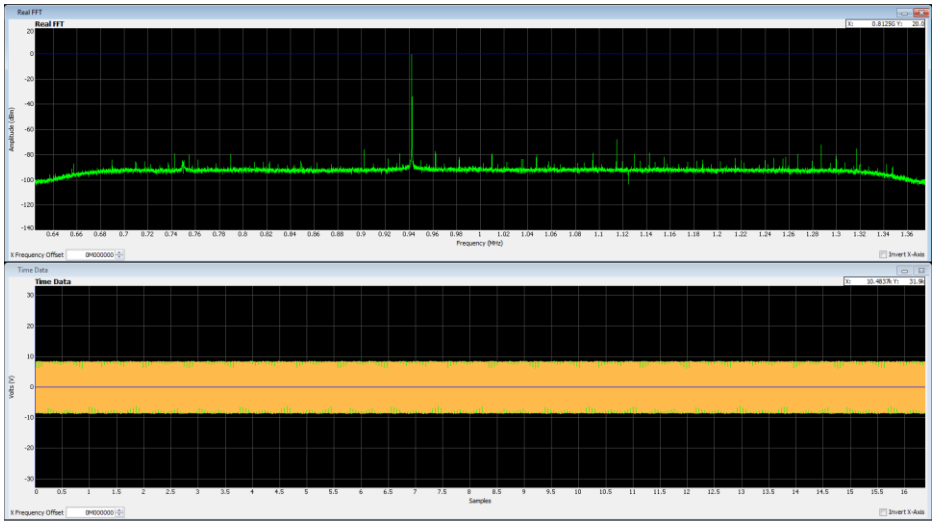
- Mezzanine designed by SKA-SA, manufactured by Peralex
- 10 Gbps x 16 SERDES interface to FPGA
- 2 or 4 GB Micron HMC devices
- Energy efficient (relative to DDR3/4)
- High bandwidth (relative to DDR3/4)
- Large capacity (relative to QDR SRAM/Bandwith Engine)





Skarab Mezzanine: SKARAB ADC32RF45X2 (New Kid On The Block)

peralex
the sense of signals

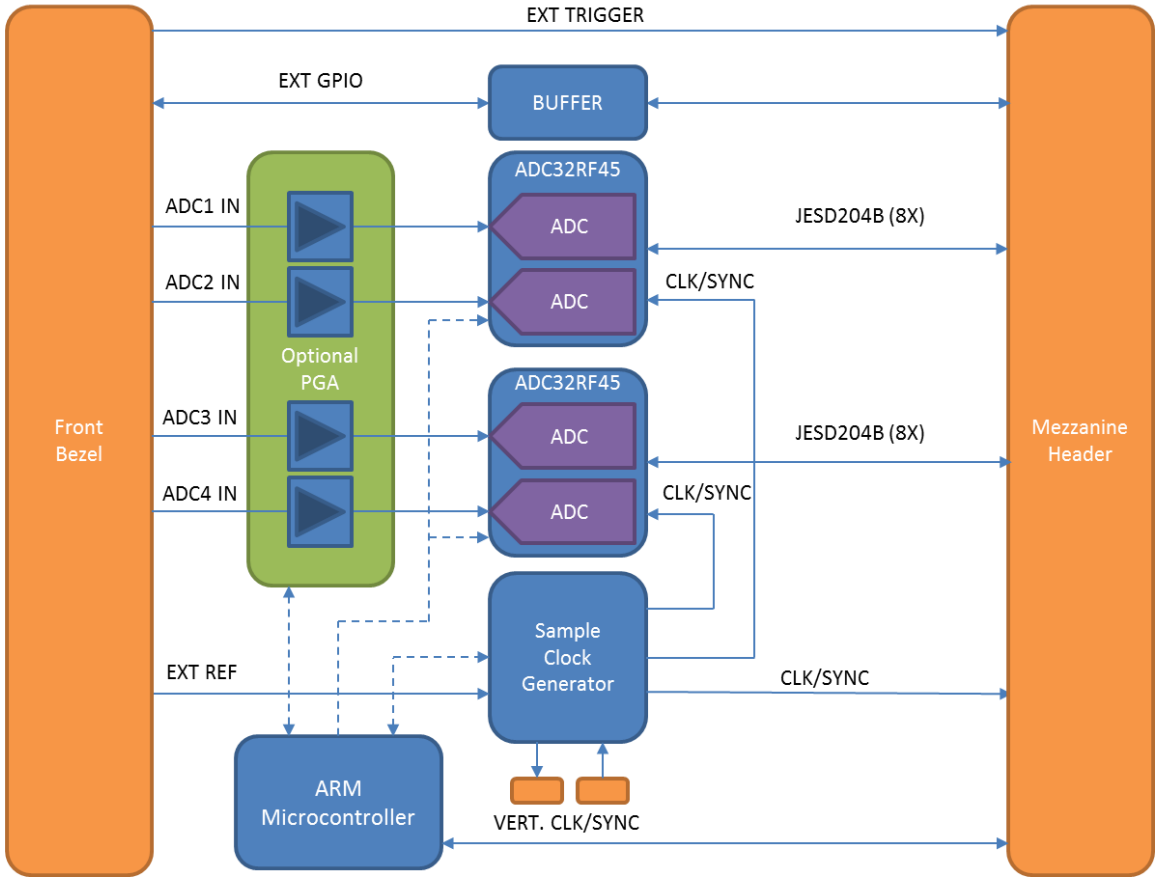


SKARAB ADC32RF45X2 Features

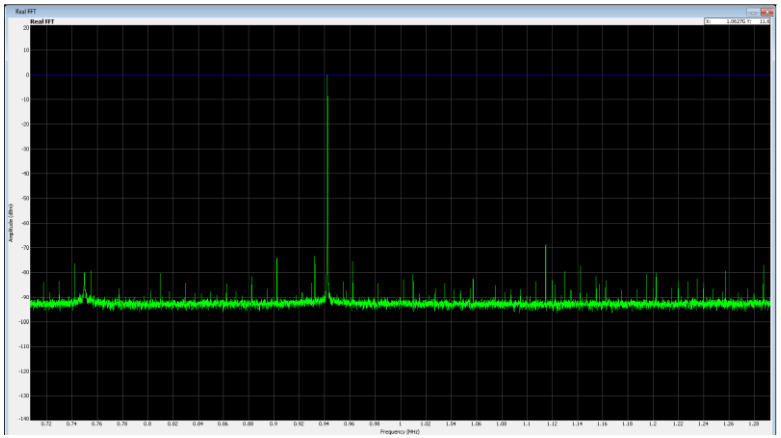
- Four 14-Bit, 3.0 GSPS ADC channels
- On-chip dither, on-board PGA (optional)
- On-board digital down-converters:
  - Up to 8 DDCs (two per ADC; dual-band mode)
  - 3 independent NCOs per DDC
- High performance 3 GHz sample clock generator
- External trigger and GPIO
- Dedicated ADC sub-system management processor
- Phase-synchronous data acquisition across multiple channels/boards
- Power consumption: 22 W @ 3 GSPS, 600 MHz BW

SKARAB ADC32RF45X2 Block Diagram

peralex
the sense of signals



SKARAB ADC32RF45X2 Performance



Test Case 1:

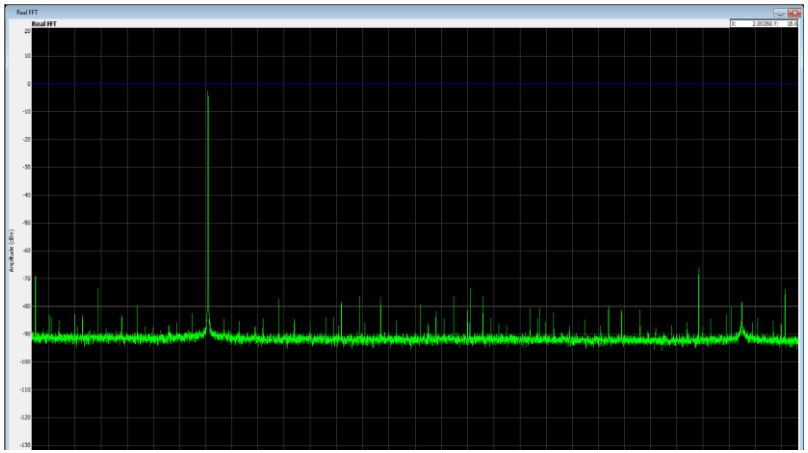
- Input Frequency: 942.5 MHz
- Input amplitude: -10 dBFS
- Sample rate: 3.0 GSPS
- Full Scale: 9.375 dBm
- DDC: Decimate-by-4
- FFT: 16384
- Average: 5

Result:
Single-tone spurious: -71.725 dBc
IMD3: 76.435 dBc @ -8dBFS
FFT Noise Level: -94.125 dBm

Test Case 2:

- Input Frequency: 1842 MHz
- Input amplitude: -1 dBFS
- Sample rate: 3.0 GSPS
- Full Scale: 12.525 dBm
- DDC: Decimate-by-4
- FFT: 16384
- Average: 5

Result:
Single-tone spurious: -63.125 dBc
IMD3: 68.375 dBc @ -8dBFS
FFT Noise Level: -81.025 dBm



Zooming in: ADC32RF45 Dual ADC Features

- 14-Bit, Dual-Channel, 3.0-GSPS ADC
- Noise Floor: –155 dBFS/Hz
- RF Input Supports Up to 4.0 GHz
- Aperture Jitter: 90 fS
- Channel Isolation: 95 dB at Fin = 1.8 GHz
- Spectral Performance (Fin = 900 MHz, –2 dBFS):
  - – SNR: 60.9 dBFS
  - – SFDR: 67-dBc HD2, HD3
  - – SFDR: 77-dBc Worst Spur
- Spectral Performance (fIN = 1.78 GHz, –2 dBFS):
  - – SNR: 58.8 dBFS
  - – SFDR: 66-dBc HD2, HD3
  - – SFDR: 75-dBc Worst Spur
- On-Chip Digital Down-Converters:
  - – Up to 4 DDCs (Dual-Band Mode)
  - – Up to 3 Independent NCOs per DDC
- On-Chip Input Clamp for Overvoltage Protection
- Programmable On-Chip Power Detectors with
- Alarm Pins for AGC Support
- On-Chip Dither
- On-Chip Input Termination
- Input Full-Scale: 1.35 VPP
- Support for Multi-Chip Synchronization
- JESD204B Interface:
  - – Subclass 1-Based Deterministic Latency
  - – 4 Lanes Per Channel at 12.5 Gbps
- Power Dissipation: 3.2 W/Ch at 3.0 GSPS

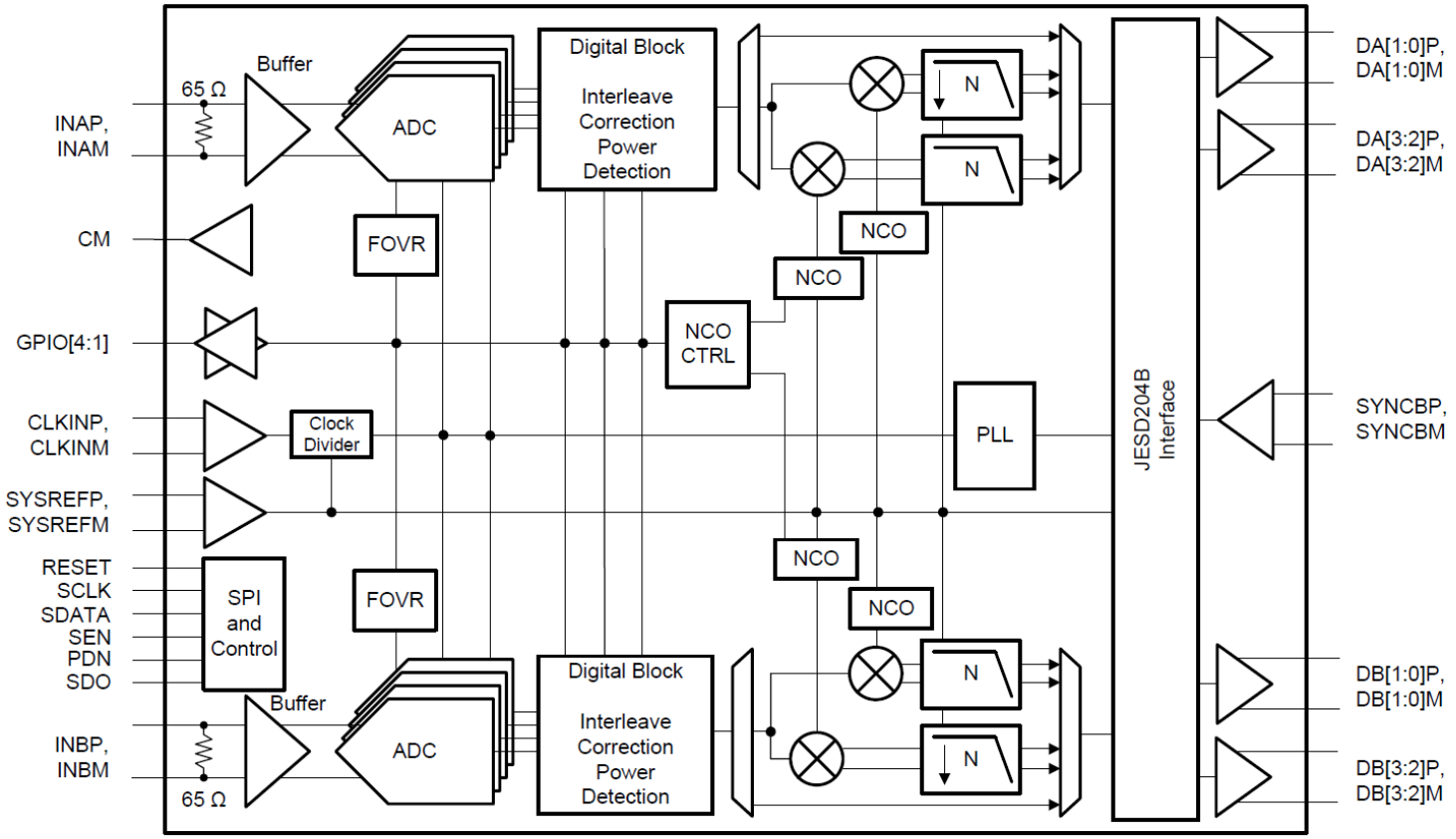
ADC32RF45 Dual ADC Applications

peralex
the sense of signals

Applications

- ❑ Multi-Band, Multi-Mode 2G, 3G, 4G Cellular Receivers
- ❑ Phased Array Radars
- ❑ Electronic Warfare
- ❑ Cable Infrastructure
- ❑ Broadband Wireless
- ❑ High-Speed Digitizers
- ❑ Software-Defined Radios
- ❑ Communications Test Equipment
- ❑ Microwave and Millimeter Wave Receivers

ADC32RF45 Dual ADC Block Diagram



SKARAB ADC32RF45X2

Configuration Options:

- Programmable Gain Amplifier (PGA) vs balun-coupled input
  - PGA gain range: ~ -6 to +15 dBm
  - PGA lowers required full scale drive strength (at the expense of noise figure)
  - Balun coupling recommended when operating at higher analogue input frequencies e.g. 2.0 GHz
- ADC
  - ADC32RF45 ($2500) – Full 1.5 GHz Nyquist bandwidth using DDC bypass (at 3.0 GSPS, this requires Skarab FPGA speed grade bump)
  - ADC32RF80 ($1250) – DDC not bypassable – maximum bandwidth of 600 MHz

SKARAB Platform Management

- Consists of:
  - FPGA configuration
  - Autonomous/programmable fan control
  - Autonomous voltage, current and fan protection/monitoring (automatic shutdown on fault)
  - Blackbox recording of faults
  - Diagnostic “always-on” USB interface allows access to fault logs even when unit is off.
- Network Platform Management
  - DHCP, PING available on ALL Ethernet interfaces
  - Health monitoring functions available on ALL Ethernet interfaces
  - High speed (<1s) FPGA configuration (through 1 GbE interface only)
- Diagnostics
  - USB/JTAG access to JTAG-compliant devices on motherboard and mezzanines
  - Serial port interface to Microblaze processor

SKARAB Board Support Package

peralex
the sense of signals

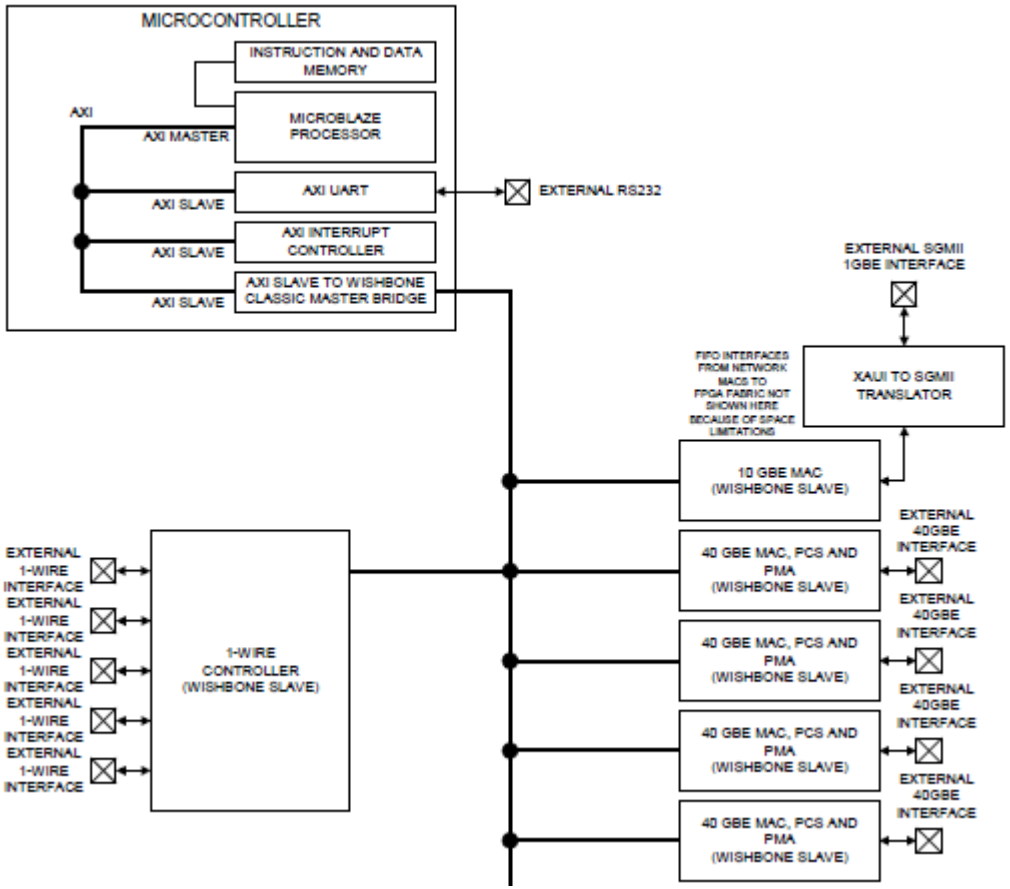
- Consists of:
  - FPGA firmware infrastructure (HDL) required for remote platform/cluster management and standard interfaces (I/O, memory)
  - Windows/Linux C++ host library and executables for platform management (health monitoring, FPGA reconfiguration over LAN etc)

SKARAB Board Support Package Firmware

- 1 Gb Ethernet core (health monitoring, control)
  - MAC only (external PHY)
- 40 Gb Ethernet core (high speed data)
  - MAC
  - PHY (XLAUI/XLPPI to external PHY/QSFP+)
- Hybrid Memory Cube controller core (SKA-SA)
- Management microcontroller
  - MicroBlaze uC with Wishbone peripheral bus
  - 1-Wire Configuration PROM access
  - Voltage/current monitors (12 rails)
  - Fan control/status
  - High speed network-based FPGA boot
  - Network setup/management for 1 GbE and 40 GbE interfaces (PING, DHCP,etc)

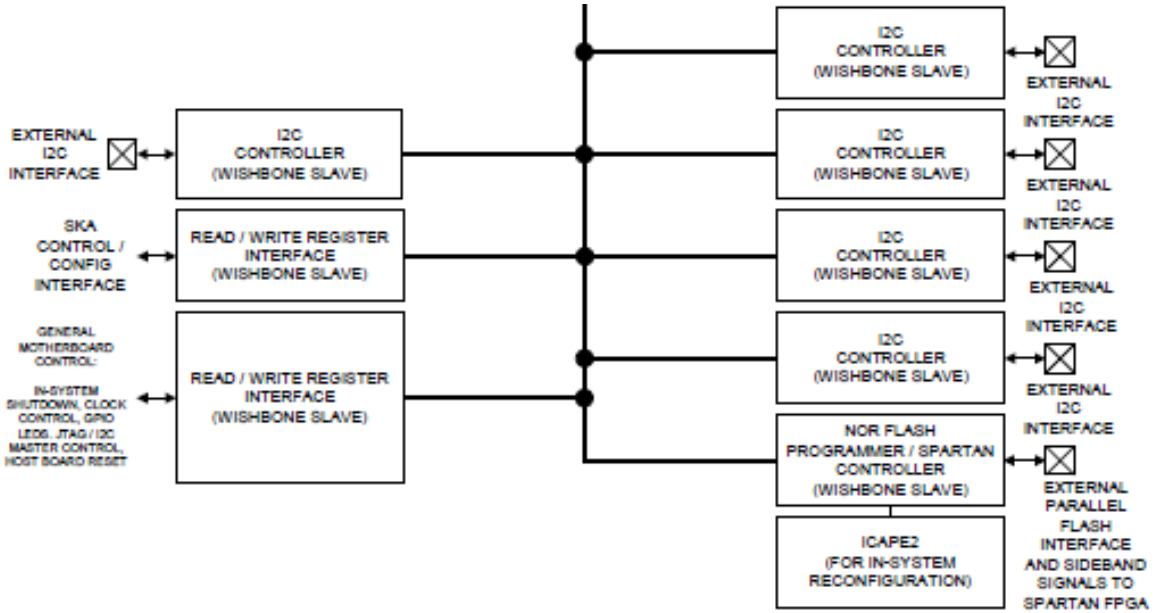
SKARAB BSP Firmware Block Diagram (1)

peralex
the sense of signals



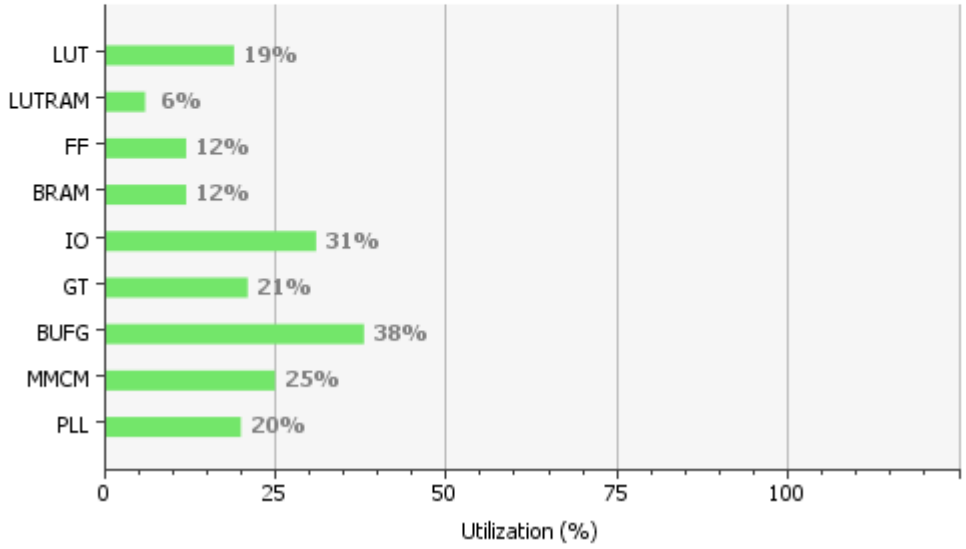
SKARAB BSP Firmware Block Diagram (2)

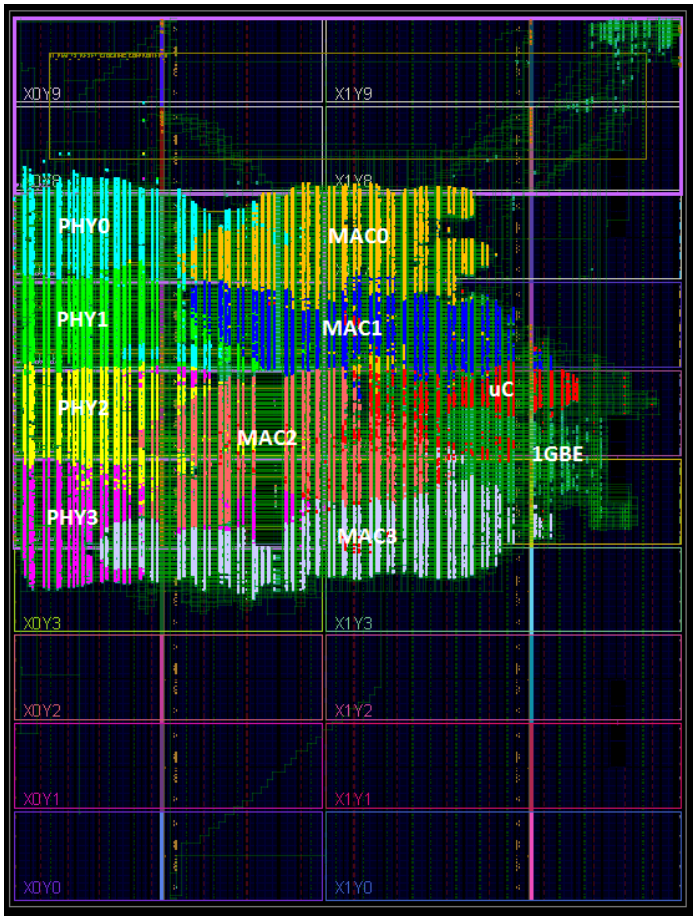
peralex
the sense of signals



SKARAB Firmware Footprint

Resource	Utilization	Available	Utilization %
LUT	81084	433200	18.72
LUTRAM	10163	174200	5.83
FF	99707	866400	11.51
BRAM	176	1470	11.97
IO	184	600	30.67
GT	17	80	21.25
BUFG	12	32	37.50
MMCM	5	20	25.00
PLL	4	20	20.00





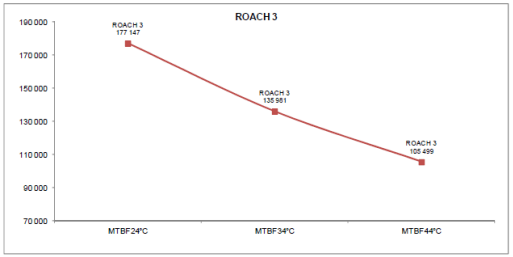
SKARAB: Some Interesting Platform Features

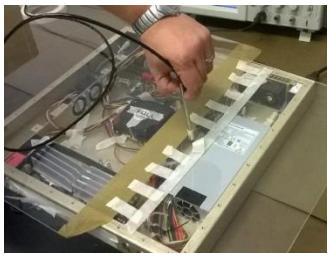
peralex
the sense of signals

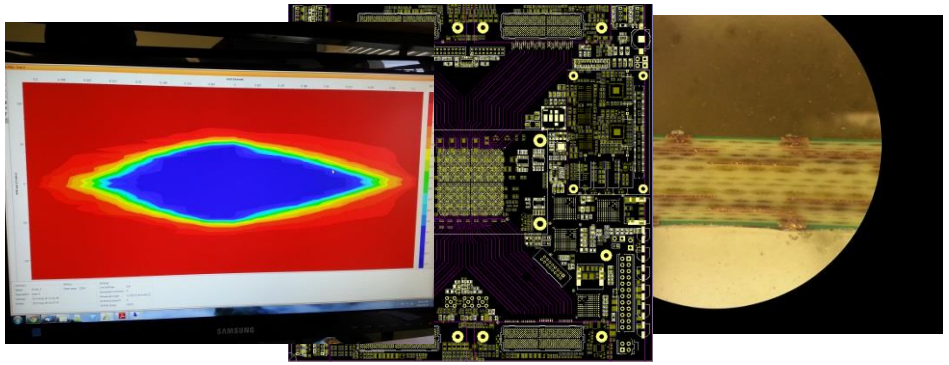
- FPGA-centric (FPGA-only!) compute platform
  - Low power consumption (relative to GPU/CPU) = low operational cost - ~45W (100-125W with memory)
  - Reduced product complexity = simpler development (software layering/configuration management)
- Flat architecture
  - Uniform, highly scalable networked processing cluster
  - Easy adoption, “bite-size” unit scaling
  - Scalable platform management
  - Fault tolerance through over-provisioning/network traffic rerouting
- Configurable Memory versus I/O Bandwidth
  - Mezzanine tile architecture
  - Semi-independent upgrade path (memory vs I/O vs compute)

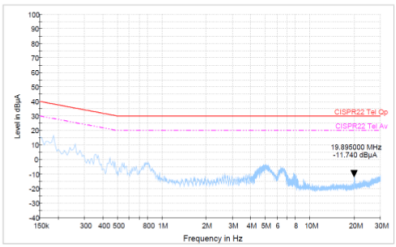
Designing a SKARAB

- MTBF analysis fed back into design to maximise reliability
- Thermal analysis (CFD and real-world models)
- Extensive Signal Integrity analysis and verification (IBIS-AMI, TDR, eye diagrams, BER) to meet 10.3125 Gbps over >7m on 128 links.
- Thorough environmental qualification (EMC/RFI, temperature, shock and vibration)









Manufacturing a SKARAB

- Internal and supply chain quality management
- PCB delamination stressing
- X-RAY/AOI of all boards
- Thermal stress screening (eliminate infant mortality)
- Automated ATP
- Optimised assembly and test harnesses.

Data Storage Pods

- Hardware developed by Peralex under a joint venture between Peralex (HW) and SKA-SA Science Data Processing (SDP) group
- Specific configurations tailored to SKA-SA MeerKAT SDP requirements:
  - High speed recording
  - High speed buffering to tape drive arrays
  - High capacity network-centric bulk storage of SDP visibility data

Data Storage Pods

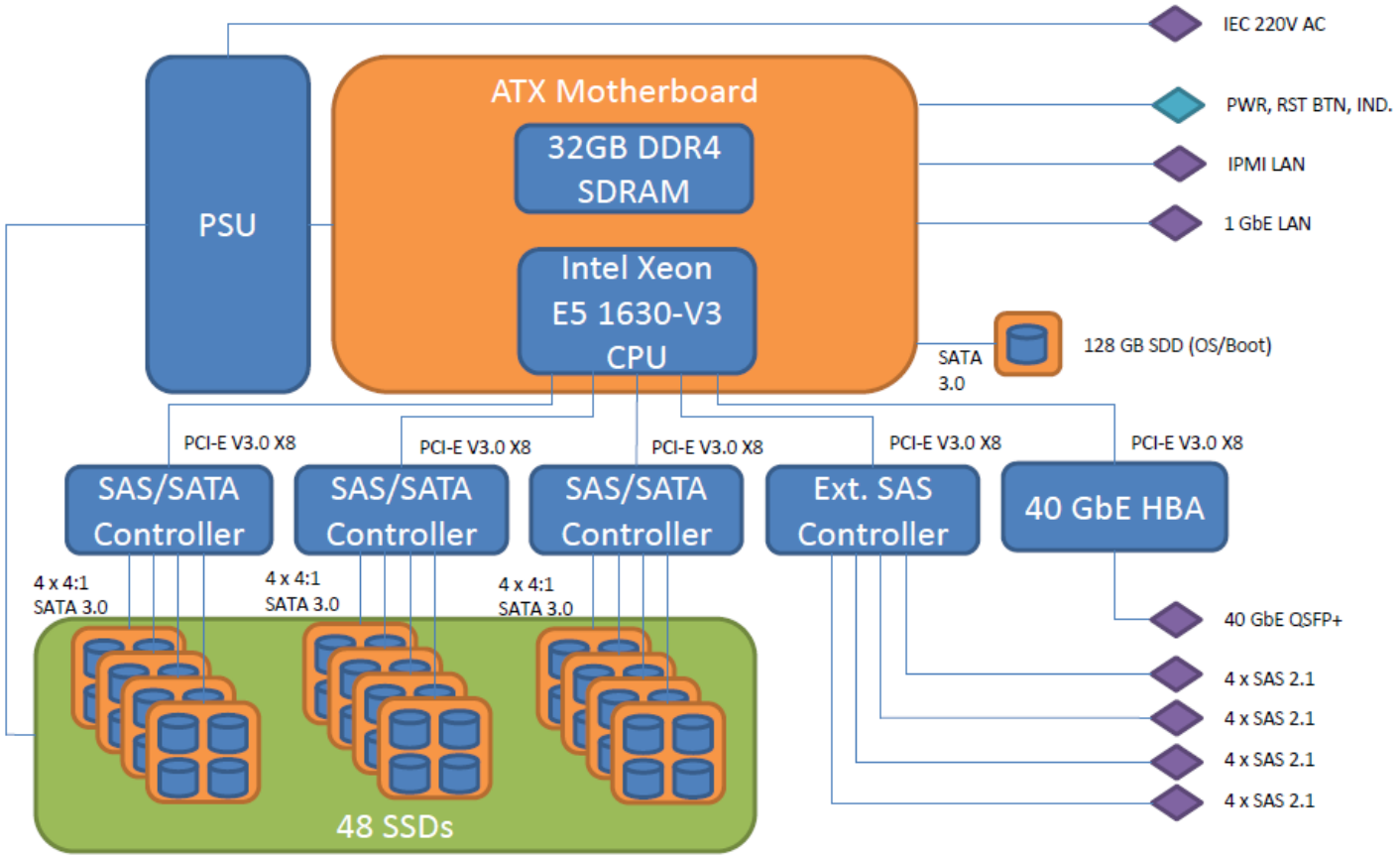
- Building block: Cost-optimised 4U, 48-drive intelligent “Storage Pods”
  - Loosely based on Backblaze open source design – more performance-focussed.
  - Configurations to supports high capacity 3.5” SAS/SATA HDDs (up to 384 TB) or high speed 2.5” SSDs (up to 96 TB)
  - Extreme performance and redundancy scaling through Ceph object store protocol.
  - Aim to provide near-40 Gbps streaming to/from disk array (record/replay)
  - Stream to tape store at 160 MB/s x 8 = 1.28 GB/s
- Possible applications:
  - Wideband recording/replay
  - Large digital delays/buffering
  - Bulk storage

SP4000C (High Speed Buffer)

- Storage: 48 x 2.5” SATA bays
  - E.g. SSD storage 2TB x 48 = 96TB
- External Interfaces:
  - 1 x 40 Gb Ethernet
  - 16 x external SAS (storage extension, tape drive interface)
  - IPMI network management interface
- Internals:
  - Xeon E5-1630V3 processor
  - 32GB DDR4-2133 SDRAM
  - 3 x 16i SAS controllers
- Applications:
  - High speed buffer (up to 38 Gbps)
  - Data recorder

SP4000C Block Diagram

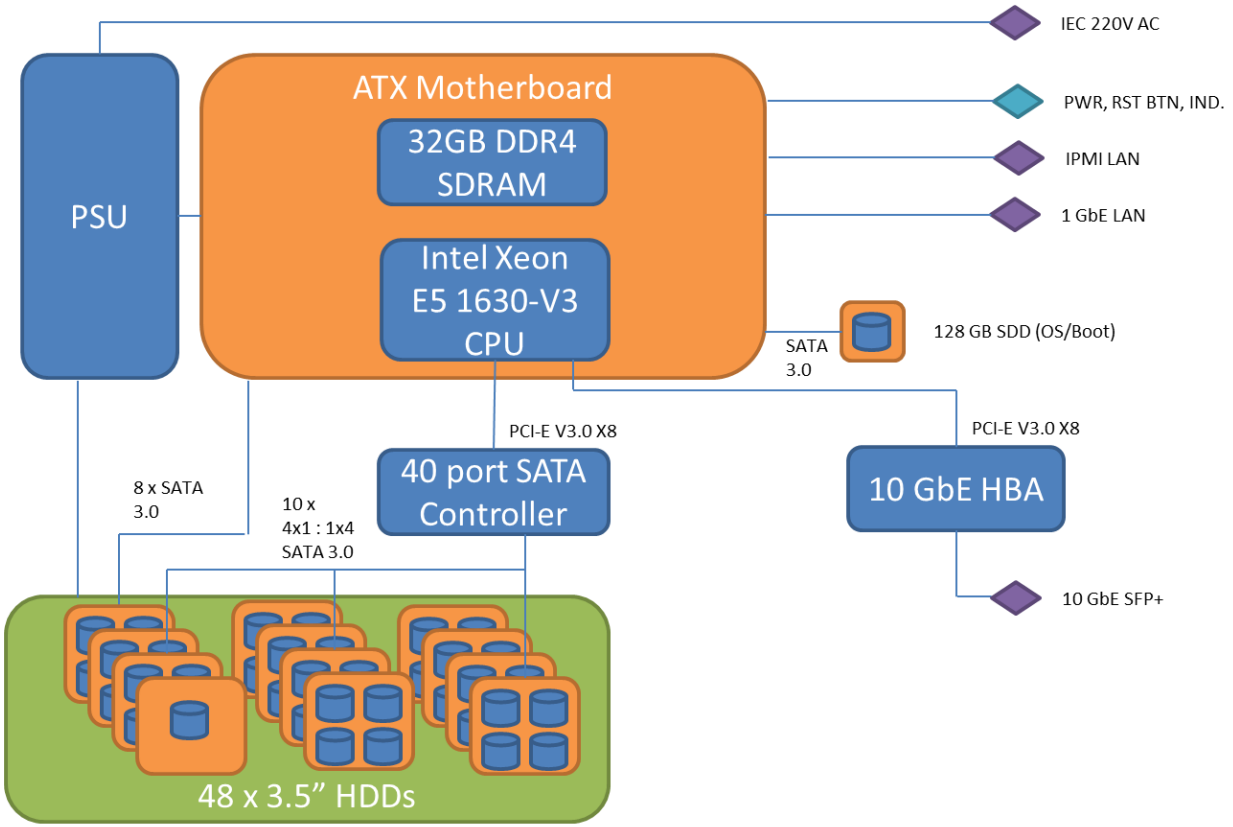
peralex
the sense of signals



SP4000L (High Capacity)

- Storage:
  - 48 x 3.5” HDDs
  - E.g. 48 x 8 TB = 384 TB
- External Interfaces:
  - 1 x 10 Gb Ethernet
  - IPMI network management interface
- Internals:
  - Xeon E5-1630V3 processor
  - 64GB DDR4-2133 SDRAM
  - 40 + 5 SATA controllers
- Applications:
  - High capacity, scalable network data store
  - Ceph Object Storage Daemon (OSD) node

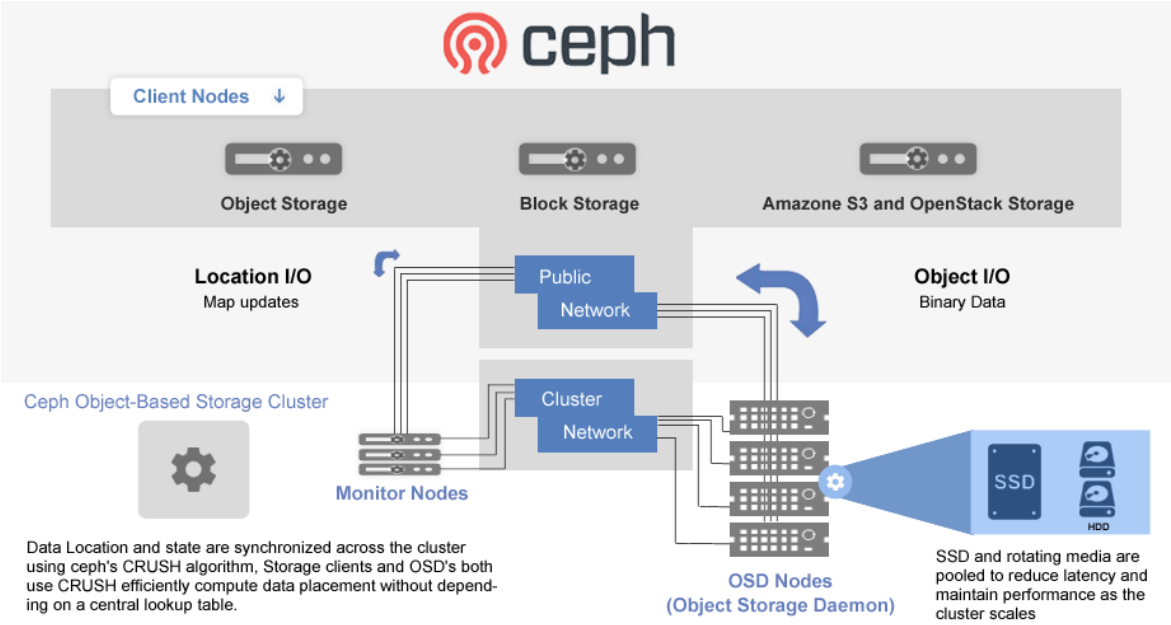
SP4000L Block Diagram



Ceph

peralex
the sense of signals

- Ceph is a software storage platform designed to present object, block, and file storage to network attached clients from a distributed storage cluster. Ceph's main goals are to be completely distributed without a single point of failure, scalable to the exabyte level, and freely-available.



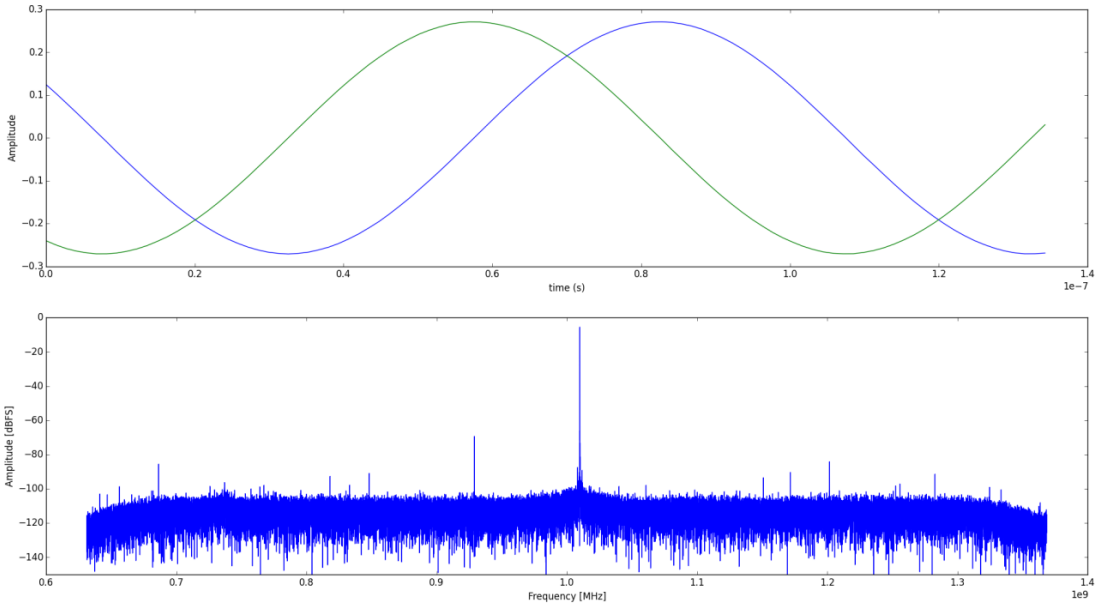
Bringing it all together (2016): ADC32RF45 Data Acquisition Proof of Concept

peralex
the sense of signals

- 3 GSPS sample rate
- 600 MHz BW (ADC internal DDC)
- ~23 Gbps raw data rate over 40 GbE (using Skarab 40 GbE core)
- Continuous streaming to SP4000C storage pod (SSD array).

ADC32RF45 Data Acquisition Proof of Concept: ADC First Light (2016)

- Performance was encouraging.
- Full real-time acquisition to storage path demonstrated (continuous storage for >1 hour @ 600 MHz BW)
- Precursor to SKARAB ADC32RF45X2 mezzanine



Conclusion

- SKARAB volume production (including HMC and 40 GbE mezzanines) is in full swing. To date, ~100 Skarab units deployed.
- ADC32RF45 proof-of-concept (2016) demonstrated full Skarab data acquisition infrastructure integration (ADC32RF45 + Virtex 7 FPGA + SP4000C). This is now application-ready as Skarab (with integrated ADC32RF45X2) + SP4000C.
- SKA-SA “CasperFPGA” Radio Astronomy tool suite is in active use by SKA-SA and early adopters (Harvard Smithsonian CfA).
- Both Storage Pod variants are verified and in series production. To date, ~4.6PB deployed in a Ceph test cluster of 12 x SP4000L (SKA-SA MeerKAT project) that will soon scale out to ~23PB.

peralex
the sense of signals

Thank You For Listening!