

EXPR_eS FABRIC WP1

Updates: WP1.1.1/2 (MRO)

- 2007 Sep 20 in MPIfR, Bonn

Ari Mujunen, Metsähovi Radio Observatory,
Helsinki University of Technology TKK

EXPR_eS is an Integrated Infrastructure Initiative (I3), funded under the European Commission's Sixth Framework Programme (FP6), contract number 026642 EXPR_eS.



Updates on WP1.1.1/2

- WP1.1 Data Acquisition...
 - .1 ...Architecture
 - .2 ...Prototype
- First: demo 0.5Gbps over 1GE (using VSIBs)
- Then: demo nGbps (4?) over 10GE (using iBOBs/iADCs)

Already Demonstrated

- 896Mbps over 1GE with Tsunami (2x VSIB PCs)
- Piggybacking onto regular Mark5A+FS experiments
 - Tested with regular IVS geo experiments (On, Mh, Bonn)
 - Mark5A takes disk backup, VSIB sends data from Mark5A outputs (via VSIC) in real-time /w Tsunami to correlator
 - Automated scripts to use regular FS schedule files
 - Demonstrates data acq control and FS compatibility
 - Should work as well with Mark5B VSI outputs

Already Demonstrated 2

- 256Mbps on-the-fly Mark5B formatting for VSIB
 - Demoes COTS PC can do Mark5B format
 - Demoed during 28-Aug-2007 APAN real-time experiment
 - Software contributed by Chris Phillips et al @CSIRO
- On, Mh, Tsukuba, and Kashima "long-distance" transfers
 - Geo near-real-time UT1 observations as motivation
 - Tsunami to K5 software correlator, UT1 results out 30 minutes after transfer
 - Simultaneous real-time from both On and Mh to Ts at 2x256Mbps attempted; failed; will retry
 - Worked during testing...

Already Demonstrated 3

- Tested using USB2.0+FireWire+Ethernet external disks in PC data acquisition
 - <http://www.metsahovi.fi/vlbi/instr/boards/vsib2usbdisk.html>
 - 1—2TB commercial "cubes" with 1GE appearing at less than 500e
 - 48 cubes + 48-port 1GE-to-10GE switch => 96TB of buffer space at 30—40ke and 4Gbps for 53 hours
 - Would there be a way to use a "fuzzy forest" of networked disks?

iBOB/iADC Status

- The first quotation from Digicom, Inc. (Mo Ohady) received 19-Feb-2007, the second 15-Mar-2007
- Indicated "willing to order" 27-Mar-2007 upon receiving a detailed quote
- Revised quotes 30-May2007, 01-Jun-2007, 04-Jun-2007
- Purchase order & four sets of Xilinx chips sent, confirmed 02-Jul-2007
- Chips arrived 09-Jul-2007 at Digicom
- 04-Sep-2007 confirmed number of boards; Digicom say they are building iADCs, in about 10 days iBOBs, warnings about testing being slow (done at UCB)

iBOB/iADC Status 2

- So: end-of-September On should get 1 iBOB and 1 iADC and Mh 3 iBOBs and 3 iADCs
 - Jb already has 10 iBOBs but no iADCs
- Tested 2048.0/1024.0MHz iADC clock synthesizer boards
 - Locked to station 5MHz
 - Synth chip evaluation boards from ADI and NSC
 - ~100e each, web-ordered, quick delivery
 - Both seem to perform just fine (good phase noise), report being prepared by Guifré
 - ADI might be slightly better
 - Still need to check output power compatibility with iADC and if phase changes due to temperature changes are a problem

DiFX on PS3 in the meantime...

- Ported i386-based DiFX (Swinburne, Adam Deller) into PowerPC/Cell-based PlayStation3 with Linux
 - Bought PS3 in January from US
 - Initial "direct port" mainly utilized the 3GHz PowerPC main processor -> many times slower than i386 DiFX
 - Because about 90% of the Cell processor SPE processing power was unused

DiFX on PS3

- "Fast Matrix Multiplication on Cell" by Daniel Hackenberg (Technische Universität Dresden) inspired greatly the ways to achieve closer to 80% SPE utilization
- 10 stations, 16ch x 128Mbps==1Gbps each in real-time with about 32 PS3s looks feasible
 - All other I/O in PS3 quite limited except 1GE
 - Does about 600--680Mbps
 - One of motivations behind tsunamifs (see later slides)
 - Estimating one PS3 correlating power being equivalent of five 3GHz Xeons (32 PS3s <=> 160 i386s)

fuseMk5A

- A “user space” file system layer on top of StreamStor API
- Makes 8-pack recordings readable as regular Linux files
- Enables using all Linux file utilities on 8-packs
 - Including Tsunami server
- Enables Mark5A/B/C 8-pack recordings to take part of regular software grid correlation
 - Such as DiFX; DiFX can read Mark5 8-pack recordings as well as regular VSIB/NAS/SAN/Raid recording files
- Already implemented by Jan Wagner; seems faster than Disk2File...

tsunamifs

- A "user space" file system layer being developed on top of connections to Tsunami server(s)
- Makes remote Tsunami server files (or real-time streams which look like files named with the standard VLBI file name convention) available as local files
 - But no local files are actually created
 - Instead a largish memory ring buffer makes file data available to all regular Linux programs as they demand
 - Ring buffer filled in the background, with Tsunami retransmissions (if required)
- Obliterates data transfer code from data acquisition and correlation processing (Linux) software

tsunamifs Uses

- Makes distributed disk buffers feasible
 - At stations, at correlator, anywhere
- Near-realtime reading of a remote recording
 - (The opposite of real time recording to correlator disks)
- Can be combined with fuseMark5A to enable remote (near-realtime and/or software correlator) reading of 8-packs, too
- Near-realtime mode of correlation might prove more appealing with software correlators than hard real-time
 - As it delivers results nearly as fast
 - And if software correlator is fast enough, "near" becomes "hard" automatically!

Distributed disk buffers

- Need to keep an open eye for alternative ways to have network-accessible disk buffers around
 - Industry NAS, SAN, iSCSI, Raid -> 10GE
- 1TB SATA2 disks (no PATA version...)
- fuseMark5A amalgamates 8-packs to this concept

4U!



http://www.verkkokauppa.com - Verkkokauppa.com

Tuote 2695 [täällä sivu](#)



1 2

AIC, 4U 19" serverikotelo, 950W ESP12V, SATA-takalevy 24 HDD.

- Optimoitu 4U RAID-kovalevykotelo.
- 24x HDD.
- Emolevy max 12"x13" E-ATX / SSI EEB 3.5.
- Tilaa myös kahdelle 3.5" järjestelmäkovalevyille.
- SATA-takalevy ja 24kpl hot-swap kovalevykasetteja 3.5" kovalevyille, joiden korkeus on korkeintaan 1".

Asennuskiskot mukana.

Takuu:	12 kk
Linkit:	Tuoteinfo
AIC tuotekoodi:	RMC4E2-QI-XPSS
Saatavuus:	heti
Hinta:	1499.90 € (sis alv.)

http://www.verkkokauppa.com/productimages/orig/38292_01.jpg

10GE Networking

- To enable iBOB and PC tests we need some 10GE networking equipment
 - 10GE switch(es)
 - 10GE network boards for PCs
- Need also fast PCIe-enabled PCs
- These can/should be used to test 10GE behavior as it becomes available at On and Jb, too

10GE Networking, Switches

- Small 10GE switches to connect iBOBs to each other and PCs
- iBOBs have CX4 copper ports; no CX4 to fiber converters found (yet) nor "mixed" switches
- E.g. HP models:
- To mix media, the more expensive modular switch must be populated with CX4 and fiber modules

ProCurve Switch 6400cl series

	Mallit	Yleiskatsaus	Tekniset tiedot	Tarvikkeet ja lisälaitteet
	Yhdellä silmäyksellä			
	J8433A	J8474A		
Ulkoiset I/O-portit	6 CX4 10-GbE ports; 1 RS-232C DB-9 console port; Supports a maximum of 8 10-GbE ports	6 10-GbE ports; 6 open transceiver slots; 1 RS-232C DB-9 console port; Supports a maximum of 8 10-GbE ports		

10GE Networking, Boards



- Recommendations from Jb/RHJ that Myrinet PCIe boards perform well; backed up with test results
 - They are not too expensive, either

Myri-10G 10-Gigabit Ethernet Solutions - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.myrinet.com/Myri-10G/10gbe_solutions.html

These NICs are available in the standard low-profile PCI-Express add-in card form factor with either 10GBase-CX4 copper or [10GBase-R](#) fiber network ports.

PHY	10GBase-CX4	10GBase-R
Photo	 Specifications	 Specifications
Product Codes with the Ethernet software bundle	10G-PCIE-8A-C+E 10G-PCIE-8AL-C+E	10G-PCIE-8A-R+E 10G-PCIE-8AL-R+E
List Price	\$695	\$795 XFP sold separately
Cables	10GBase-CX4 cables	Duplex multimode fiber cables

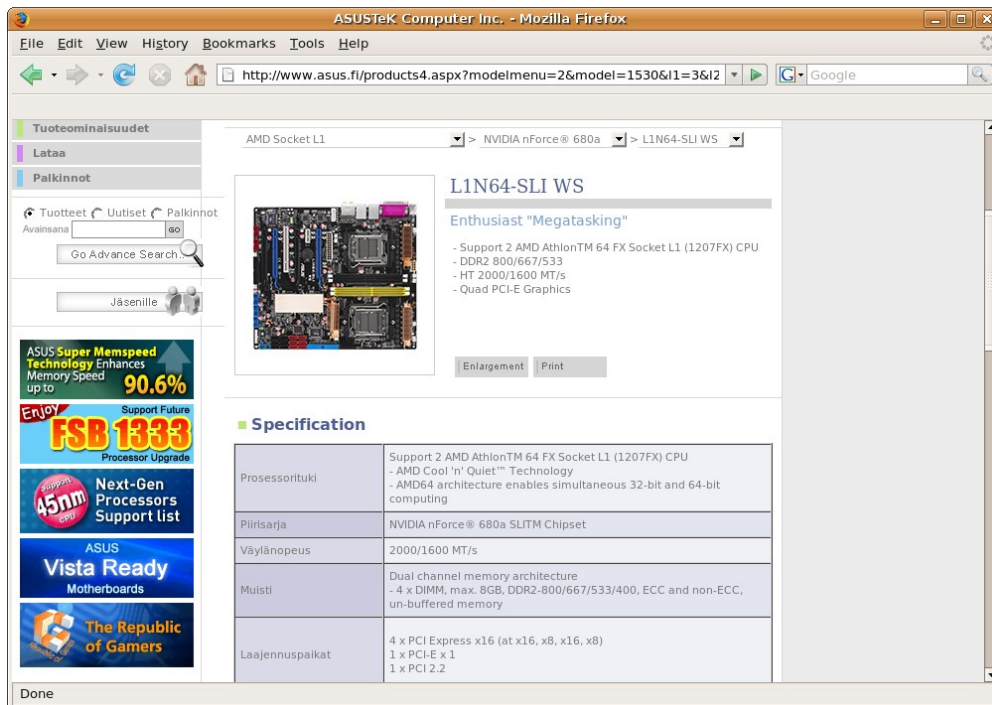
Software. The driver and NIC firmware for 10-Gigabit Ethernet operation is currently available for Linux, Windows, Solaris 10, Mac OS X, and FreeBSD. You can download the drivers (which include the NIC firmware) from this [Myri-10G 10-Gigabit Ethernet Driver Download](#) page. The driver was contributed to and accepted in the Linux kernel, and is included in the 2.6.18 and later kernels. However, the driver available by download will be more up-to-date. The driver is also included in FreeBSD 7.0.

Offloads. The driver and NIC firmware implement zero-copy on the send side with all supported operating systems.

Done

PCIe x8 capable PCs?

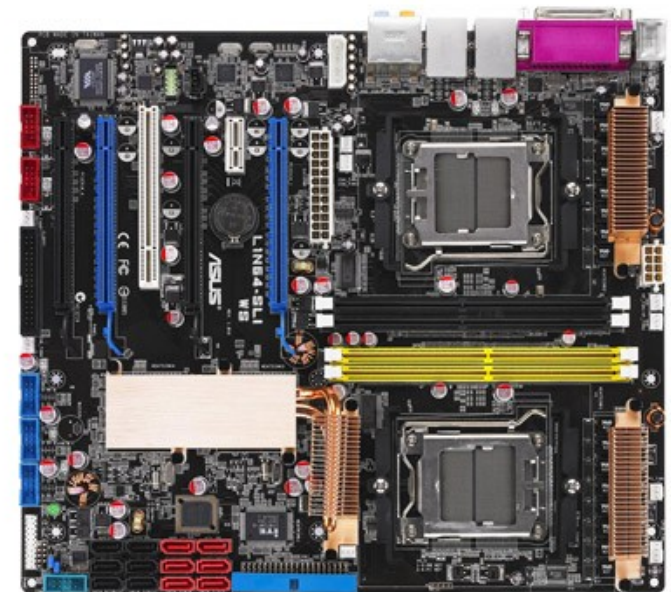
- 10GE board is not of much use unless there are PCs which can pump data at >>4Gbps around
- Just starting to appear; first ones might not be too good in practice
- For instance Asus L1N64-WS:



The screenshot shows the ASUS website for the L1N64-SLI WS motherboard. The page includes a navigation menu, a search bar, and a sidebar with promotional banners for ASUS Super Memspeed Technology, Next-Gen Processors, Vista Ready Motherboards, and The Republic of Gamers. The main content area features a product image, a title "L1N64-SLI WS", a sub-header "Enthusiast 'Megatasking'", and a list of features: "Support 2 AMD AthlonTM 64 FX Socket L1 (1207FX) CPU", "DDR2 800/667/533", "HT 2000/1600 MT/s", and "Quad PCI-E Graphics". Below this is a "Specification" table.

Specification	Details
Prosessorituki	Support 2 AMD AthlonTM 64 FX Socket L1 (1207FX) CPU - AMD Cool 'n' Quiet™ Technology - AMD64 architecture enables simultaneous 32-bit and 64-bit computing
Piirisarja	NVIDIA nForce® 680a SLITM Chipset
Väylänopeus	2000/1600 MT/s
Muisti	Dual channel memory architecture - 4 x DIMM, max. 8GB, DDR2-800/667/533/400, ECC and non-ECC, un-buffered memory
Laajennuspaikat	4 x PCI Express x16 (at x16, x8, x16, x8) 1 x PCI-E x 1 1 x PCI 2.2

L1N64-SLI WS



12 SATA ports!

© 2007 ASUSTeK Computer Inc. All rights reserved.

Current 10GE Plan

- Purchase two modular HP switches with
 - 4 10GBASE-CX4 modules
 - 4 10GBASE-SR fiber modules
- Extend our Funet connection (Extreme switch) with 10GBASE-SR module
- Purchase two Myri-10GE boards
 - One CX4 and one SR
- 4 CX4 cables; 4 SR fibers of various length
- Thus can have one switch near VLBI rack and another in FPGA lab with fiber connects
 - And still can connect iBOBs and Extreme directly to Myri boards (for instance, to debug switch problems)
- Purchase just one Asus MB to see if it is any good

Current Plan with iBOBs

- Primarily interested to get the iADC going
 - Not so interested in advanced filtering/channeling
 - Might want to try Mileura 1024-ch fw
- Main 10GE mode seen as a very simple (though controllable) "set-rate UDP packet emitter"
 - Assuming Jb is more interested in trying more advanced protocols on iBOB
 - Not so interested in attempting to run Linux or similar on on-FPGA PowerPC
 - iBOB memory is too scarce for this
- Interested in trying various ways to offload iBOB control to a regular Linux computer
- Assuming Jb is doing hw connection to eMERLIN

FrankeniBOB resurrects...

- The design document referred to the idea of connecting iBOB to a Linux computer with a similar parallel bus that is used on the BEE2
 - Would enable using BEE2 firmware, perhaps
 - Could perhaps be implemented with VSIB board connected to the iBOB VSI connector
 - Not so many VSIB boards left...
- Other approaches to connect a Linux PC to iBOB being considered:
 - Control packets via
 - 100Mbps extra Ethernet
 - embedded in 10GE stream of packets
 - Rumor of a Linux expansion board for iBOB?
 - Ethernet "proxying"; Linux computer doing the "diffcult" things like ARP for iBOB

VSIB2 resurrects...

- ...no, not really. Still want to mention Xilinx evaluation board for PCIe x8 with Virtex 5 chip:
- Could emulate VSI or take iADC data directly or connect to iBOB as "transplant"



Xilinx Bus I_F IO and other Connectivity : Virtex-5 LXT ML555 FPGA Development Kit for PCI Express

File Edit View History Bookmarks Tools Help

http://www.xilinx.com/xlnx/xebiz/designResources/lp_product_details.js

Technology Solutions | Products & Services | Support | Online Store | About Xilinx

Silicon Devices | Design Tools | Intellectual Property | Boards & Kits | Training | Services | Third Party

Xilinx : Products & Services : Boards & Kits : Bus Connectivity and IO : Virtex-5 LXT ML555 FPGA Development Kit for PCI Express and PCI

ASIC Emulation
Accessory Boards
Bus Connectivity and IO
Digital Signal Processing
Embedded Processing
General Device Solutions
Memory Interfacing

Virtex-5 LXT ML555 FPGA Development Kit for PCI Express and PCI (HW-V5-ML555-G)
Xilinx Boards & Kits

The Virtex-5 LXT FPGA Development kit for PCI-Express® supports PCIe®/PCI-X™/PCI™. This complete development kit passed PCI-SIG compliance for PCI Express v1.1 and enables you to rapidly create, and evaluate designs using PCI Express, PCI-X and PCI interfaces.

Buy Online from:

- [Avnet](#)
- [Nu Horizons](#)
- [Local distributor](#)

Click to see enlarged image.

\$2,950.00

Device Family Support

- ◆ Virtex-5 LXT

Key Features

- ◆ Virtex™-5 LXT XC5VLX50T-1FF1136 FPGA
- ◆ Serial Connectivity:
 - PCIe: 8-lane add-in card connector
 - SFP: 2 transceiver module ports (SFP modules not included)
 - SATA: 1 Serial ATA disk drive interface connector (SATA cable not included)
 - SMA: 1 SMA port for off-board GTP transceiver connectivity
 - USB: 1 USB 2.0 port (USB interface cable not included)
- ◆ Parallel Connectivity:
 - PCI / PCI-X: 64-bit 3.3V add-in card connector

Related Information

- > Xilinx On Board Information
- > Reference Designs for PCIe Development Kit
- > Access PCI/ PCIX Evaluation IP Cores

Related Products

- > ISE Design Tools Center
- > Xilinx IP Center
- > Conversion Module, SMA to SFP
- > Conversion Module, SMA to SATA

http://www.xilinx.com/xlnx/xebiz/utills/image_window_product.jsp?valueA=HW-V5-ML555-G

So, what next?

- Get 10GE networks up and running
- Get PC-based UDP transmitters and receivers running
- Test in lab, test in triangle On-Mh-Jb
- Get iBOB-based UDP transmitters (Jb+Mh; deploy at On, too) and receivers (Jb) running
- Test all four "UDP boxes" against each other, in lab and in real networks
- Add real iADC data into UDP packets (Mh)
 - See if there are any useful channelization firmware available that can be added
- Add more advanced protocols to iBOB (Jb)
- See what we can and want to demo at 4(?)Gbps