D77: Data Acquisition Prototype at Telescope

Jan Wagner, jwagner@kurp.hut.fi
Guifré Molera, gofrito@kurp.hut.fi
Jouko Ritakari, jr@kurp.hut.fi
Minttu Uunila, minttu@kurp.hut.fi
Ari Mujunen, amn@kurp.hut.fi
Metsähovi Radio Observatory

October 8th 2008

Contents

This work is done as a part the Express Production Real-time e-VLBI Service, EXPReS, project. This work has received financial support from the European Commission (DG-INFSO), within the Sixth Framework Programme (Integrated Infrastructure Initiative contract number 026642, EXPReS).

1 Introduction

This report describes the 4 Gbps recorder designed for the EXPReS project, hardware purchased and used, the iBOB data acquisition board, tests performed and possible applications for the software spectrometer designed. The tests performed within the work were run with normal iperf, with Tsunami [1] UDP data transfer, File Transfer Protocol (FTP) and other usual tools.

First, a couple of explanations; Maximum Transmission Unit (MTU) refers to the size (in bytes) of the largest packet that a given layer of a communications protocol can pass onwards. The User Datagram Protocol (UDP) is one the core protocols of the Internet protocol suite. By using UDP, programs on networked computers can send short messages to one another. UDP is used for, for example, sending real-time videos and sound. Transmission Control Protocol (TCP) is also one of the core protocols of the Internet protocol suite. TCP provides realible, in-order delivery of a stream of bytes, making it suitable for applications like file transfer and e-mail. Iperf is a commonly used network testing tool that can create TCP and UDP data streams and measure the throughput of a network carrying them.

In Table 1 the test hardware is listed. It includes the following: the internet setup, switches, network cards, iBobs, PCs and settings.

Table 1: Hardware.

Hardware	
Internet:	10 Gbit/s over XR fibre to Funet/CSC
Switches:	1 x Extreme Networks Summit X450-24t with a XR and SR Xenpak
	1 x HP 6400cl CX4 with dual SR fibre
Network cards:	1 x Chelsio N320E-CX dual-CX4 eval kit (2 cards and a cable)
	1 x Myrinet CX4
	1 x Myrinet XFP with SR fibre
iBobs:	2 x iBob with ADC, connected to the HP switch
PCs:	Adibal: Asus LIN64-SLI WS, 2 dual-core AMD 2212 CPUs, 4 GB RAM, 12 x SATA in RAID 0
	Juliano: Asus P5WD2-E Premium, Core 2 Duo 3.2 Ghz, 2 GB RAM, 4+4 x SATA in RAID 0
Settings:	Locally 9000 MTU, Funet 4470 MTU

The fibre connection from the Finnish University and Research network (FUNET) to Metsähovi ends in a 10 Gbit/s ZR Xenpak fibre module in the Extreme Networks Summit X450-24t switch [2]. The switch has two 10G Xenpak module ports and 24 1G Ethernet ports. PC-EVN test computers have been connected to these 1G Ethernet ports. The test computers have one or two 1 Gbit/s network cards. Two network cards can be combined into a single virtual network card using ling aggregation, which in theory can achieve an almost two-fold transfer rate. The transfer rate of the network connection between Metsähovi and the outside world was and can be tested using iperf, which is a program that can be used to measure the transfer rate of a network between two computers for both TCP and UDP protocols.

Tsunami is a UDP-based fast aggressive file transfer protocol that divides data into large UDP packets and can also be used as a data transmission protocol. Tsunami was developed at Pervasive Technology Labs research center at the University of Indiana. Data is transferred in UDP/IP packets at significantly higher rates than plain TCP would allow especially with large round-trip delays.

The basic idea behind Tsunami is that the transmission data is segmented into large sequentially numbered packets of equal size. The default size of the packets is 32 kB and the goal transfer rate is 650 Mbit/s. The maximum packet size is about 64 kB due to the UDP/IP protocol limitation. In the next Tsunami version that we developed in 2007 and released on SourceForge, theoretical rates up to several Tbit/s are possible.

The report is organized as follows: in Section 2 we report on the Metsähovi 4 Gbps Data Recorder for VLBI and e-VLBI. In Section 3 the iBOB Data Acquisition and Network Streaming is described and in Section 4 we present Evaluation of New Multi-bits Sampling System.

2 Metsähovi 4 Gbps Data Recorder for VLBI and e-VLBI

The Metsähovi correlator and spectrometer were developed for the Cell processor (Sony Playsation 3) in 2007. There are two correlators: the Swinburne DiFX and the Cell minicorrelator. The minicorrelator is about ten times more powerful than Intel-based solutions and copes with high-speed input data. Our generic-platform version of DiFX on the other hand is slower but is easier to optimally move onto a new platform using only small gradual modification steps.

There are a few problems that need to be solved, like getting data in and out of specialised disk controllers, and the too slow speed of old parallel-ATA (PATA) disks. VLBI data sources are migrating from custom interfaces to 10 Gbps Ethernet, for example when the iBOB and DBBC are used. Testing and analyzing 4 Gbps streams and data is easier with a computer that can write these network streams to disk without loss.

As a solution a 4 Gbps recorder can be constructed from only commercial off-the-shelf components. The recorder uses 10 Gbps Ethernet for I/O and with the Tsunami protocol and other new software tools it can do high-speed international data streaming. The recorder harware is listed in Table 2.

Table 2: Recorder hardware.

Asus L1N64-SLI WS motherboard, nForce 680A chipset, 12 SATA-II ports
Two dual-core AMD 2212 CPUs
4GB of DDR2-1066 RAM
Chelsio N320-CX 10 Gbps Ethernet NIC
Myrinet XFP 10 Gbps NIC with SR fibre
12 Samsung Spinpoint F1 750 GB disks

2.1 Tests

The 4 Gbps recorder was tested with four different sets of tests. In the following subsections the disk subsystem tests, 10 Gbps Ethernet tests, combined 10 GbE and disk tests and Tsunami tests are described.

Earlier tests last year were made using a 12-disk RAID array and one-year-old Seagate 250 GB disks. Best write speed of 5.5 Gbps was achieved with ext2 fs but it decreases to half of that on inner tracks. The disks were replaced to the newest and fastest Samsung F1 750 GB disks and 4 Gbps performance was achieved, see Figure 1.

For the 10 Gbps Ethernet tests Myrinet and Chelsio NICs were used. Initial iperf tests gave 2-3 Gbps and 100% of CPU usage. With the use of jumbo frames the data treansfer rate improved to 8 Gbps when also UDP checksum was turned off. The final UDP results are given in Table 3.

Both tasks are CPU-limited. We used two dual-core AMD 2212 processors and threads to divide the load. Optimum performance was achieved using two six-disk RAID arrays and two UDP streams. 4 Gbps tranfer from 10 GbE to disks was sustained.

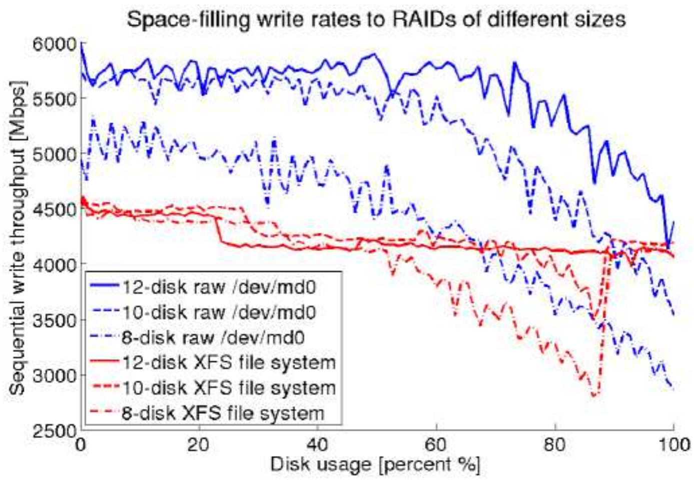

Figure 1: Space-filling write rates to RAIDs of different sizes.

Table 3: Final iperf UDP results.

MTU 1500	4.88 Gbps
MTU 4470	8.55 Gbps
MTU 9000	9.04 Gbps

By using Tsunami a file transfer to the correlator is possible. When disks are used a single Tsunami transfer runs at 4 Gbps, two parallel transfers run a little faster, without disks Tsunami rates of over 7 Gbps are possible. In the disk transfers we could achieve zero packet loss. One typical lossless 4 Gbps Tsunami transfer is shown in Figure 2.

2.2 4 Gbps recorder conclusions

A 4 Gbps network-disk recording system is feasible now. However, at this time the recorder is CPU-limited. The situtation will improve to 8 Gbps or more within a short time or with the use of four-core processors. Of the currently available disks, only the newest and fastest SATA-II consumer disks have the required performance. On the networking side, both Chelsio and Myrinet boards are fast enough. It should also be noted that at least with current CPU and often poor implementation of UDP fragmentation in software instead of in hardware, jumbo frames are needed for speeds of 4 Gbps and higher.

3 iBOB Data Acquisition and Network Streaming

The iBOB (the internet Break-out Board) is developed at Center for Astronomy Signal Processing and Electronics Research (CASPER) at University of Berkeley. The goal of CASPER is to streamline and reduce the current radio astronomy instrumentation design flow through the developement of an open-source, platform-independent design approach [3]. The iBOB was designed as the way to take data from a daughter board and pack it onto XAUI, InfiniBand or 10Gbit Ethernet to be interfaced with for example BEE2 FPGA boards. Though originally intended to be only a break-out board, it is very capable as a pre-processing board, or as a standalone platform. It has been used to interface to ADC boards to bring in digitized data for processing, output digital data to DAC boards, as well as ASIC

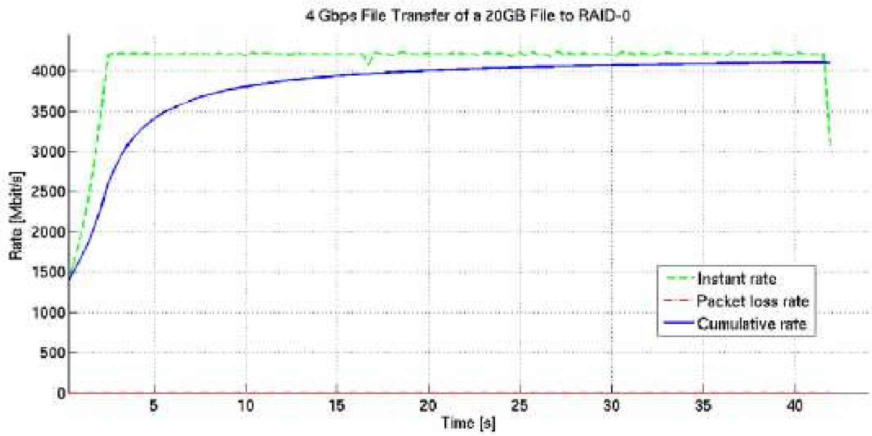

Figure 2: Lossless Tsunami transfer.

test setups as an integrated stimulus source and data capture platform. iBOB has a Xilinx Virtex-II Pro 2VP50 FPGA, and is used in radio astronomy applications primarily for digitizing data, performing downconversion, filtering and FFT operations, and outputting this data over XAUI/10GbE, enabled by two CX4 connectors on the board. [4]

Xilinx programmable solutions are designed in a wide range of applications. Xilinx customers can change or upgrade product features and functions "on the fly" - adapting to new standards and reconfiguring the hardware for a specific application. This "on the fly" technology enables faster time-to-market, product differentiation and reduced cost. [5] The hardware used in iBOB data acquisition and network streaming is listed in Table 4.

Table 4: UC Berkeleys iBOB FPGA board, iADC board and Nat LMX2531 or ADI ADF4360.

iBOB FPGA board
Virtex II Pro 2vp50 with PPC CPUs two copper 10GbE ports (CX4) two expansion ports (Z-DOK)
Berkeley iADC board
Atmel AT84AD001 dual-channel, 8-bit, 1 Gs/s/ch = 1 GHz MHz BW, 16 Gbit/s / board
Nat LMX2531 or ADI ADF4360
950..1080 MHz PLL for ADC fs

A similar board ROACH (meerKAT) is available soon and it has 4 x 10GbE, iBOB designs should drop in. For iBOB several DAQ firmware test designs were created.

3.1 iBOB DAQ and protocol for 10 GbE

In real-time e-VLBI data integrity is less important than timely transmission. Data loss factors f≤10% are tolerable although the final correlation product degredes. Packet loss and reordering can be coped with in various ways when it occurs and packet payload integrity is non-critical. It is not necessary to retransmit lost packets, nor are UDP checksums necessary. The research network links have a very low bit error rate and of out of these incorrect bits the CRC32 frame checksum will miss only 2*10-9.

Congestion control is not a real option at Gbps rates - either we can have real-time transfer, or we can not and will get some data loss. The only method is rate reduction. Rate reduction has to be controlled by the receiving side, the correlator or schedule management, to work in a meaningful way. This is identical to cancelling the current data acquisition altogether and restarting at a new time point with a new DAQ setup, perhaps with less channels or a lower sampling rate. For this reason no special congestion control protocol extensions are required in the DAQ system itself.

The two essential points implemented in Test-DAQ UDP are the following. First we must be able identify which RF band or channel each sample comes from. An obvious way is to place samples of the same channel into the same UDP packet and use a packet header that has a packet sequence number (PSN). Meta-information such as the sampling start time and sampling rate are external to the samples and it is not necessary to integrate this as duplicate information into some UDP user header field. A channel ID field in the header may be useful, though typically VLBI experiments use much less than 60.000 channels. This allows us to map every channel to a unique UDP port number and with extra no effort have all advantages of the Linux network stack and packet filtering. Hence the DAQ UDP packets have to consist only of a PSN field followed by samples from the same RF channel.

The second essential point in the Test-DAQ is that the timecode of each samples must deducible. Our solution is not to use an explicit time stamp - this information is already part of the PSN number. This has pre-requisites that are already met in VLBI: during the same acquisition process the UDP packets must have constant size (e.g. 4kB, the size can be changed when acquisition is restarted), the acquistion has to be triggered at the station 1PPS synchronization second, the 1PPS and sampling clock have to be derived from the same stable clock source, and packets must have a large enough sequence number (e.g. 64-bit) which should preferrably start from 0. The required 1PPS input trigger and PSN counters have been implemented in the iBOB DAQ design.

With the proposed method, assuming fixed 4kB packets and a 64-bit PSN, the timecode of every sample can be computed directly using the a priori sampling frequency of the current acquisition process. Count overflow wrapping of the 64-bit PSN is straight forward to detect and causes a timestamp ambiguity after every 4.5e15 samples. This corresponds to sampling at 10 Terasamples/second for 245 years continuously.

3.2 iBOB DAQ tests

In local 10 Gbps LAN tests analog signal is sampled with iBOB, streamed as UDP to 10G LAN. UDP stream recording is done with udp2raid, raw2raid (tcpdump, gulp) and udp2raided to RAID-0 at up to 4 Gbps (current RAID 4.3G, XFS), see Figure 3. Data and spectrum checks are performed with Matlab and software spectrometer.

3.2.1 Mh-On streaming/network Tests

First tests were performed on May 14/15th 2008 when Onsala was supplied with test firmware: a 100 MHz clock and 1024 Ms/s sampling. I/Q: VLBI IF 90..540 MHz, -10dBm Onsala streamed to Metsähovis RAID-PC. Unfortunately Metsähovis PC was busy with other computations and data was recorded at 1.6 Gbps, rest with slow-PC loss, checked for and found PCAL.

Second test was performed on June 11th 2008. adc2tsunami was used and proto/test design sent 2 channels (I/Q) to Onsala, see Table 5 and Figures 4 and 5. In Figure 6 $H_2 0$ in Orion is shown.

3.3 iBOB conclusions

Basic Metsähovi iBOB designs allow to digitize antenna signal and different rates stream as UDP to local or remote IP.

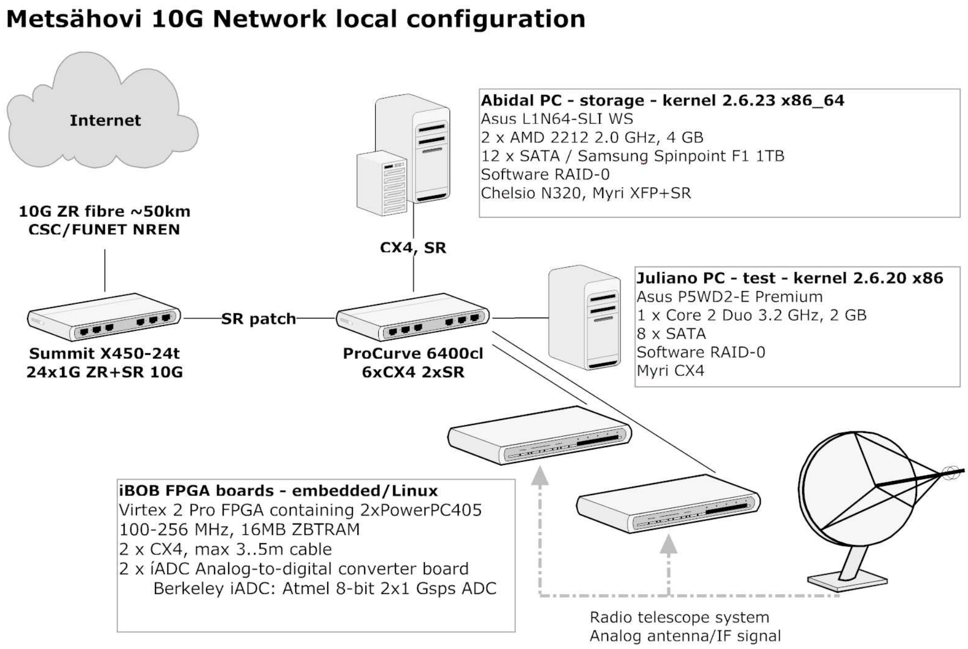

Figure 3: Network topography.

Table 5: Mh-On streaming/network test on June 11th 2008.

IBOB % regwrite dest_ip_ich 0x82F20A05	# 130.242.10.5, Onsala
IBOB % regwrite dest_port_ich 46330	# I-channel to port 46330
IBOB % regwrite dest_ip_qch 0x82F20A05	# ...
IBOB % regwrite dest_port_qch 46331	# ...
IBOB % regwrite teng_wordsperframe 512	# 512x8 : 4kB UDP
IBOB % regwrite teng_decfactor 0	# sample rate reduction 1/(1+n)

Onsala to Metsähovi RAID recording was lossless only at 1.6 Gbps. Tests should be performed in the future when the PC is really free. In local tests 4 Gbps is achieved when recording to disk.

Metsähovi to Onsala network streaming had no problems with 0 - 7.8 Gbps traffic. Credit goes to Nordunet, Sunet, Funet: it is straightforward and welcomed to stream multi-Gbps over Internet/RN using their services.

For a future test or demo the following could be planned: stream and record a maser source with 2-bit / 8-bit, 2 x 512 MHz BW and analyze it in software spectrometer.

4 Evaluation of New Multi-bits Sampling System

Storage capacity, bandwidth length and transmission rate have been limiting the VLBI recording systems to a 2-bit encoding format. These are not critical issues anymore and it is worthwhile to develop a new 8/12 bits sampler equipment. Hence, the Metsähovi research group has been experimenting with systems which offer these capabilities. The new iBOB board and an ADC converter have been the basis of building the new equipment. A Mark IV formatter and PC-EVN were used to compare all three

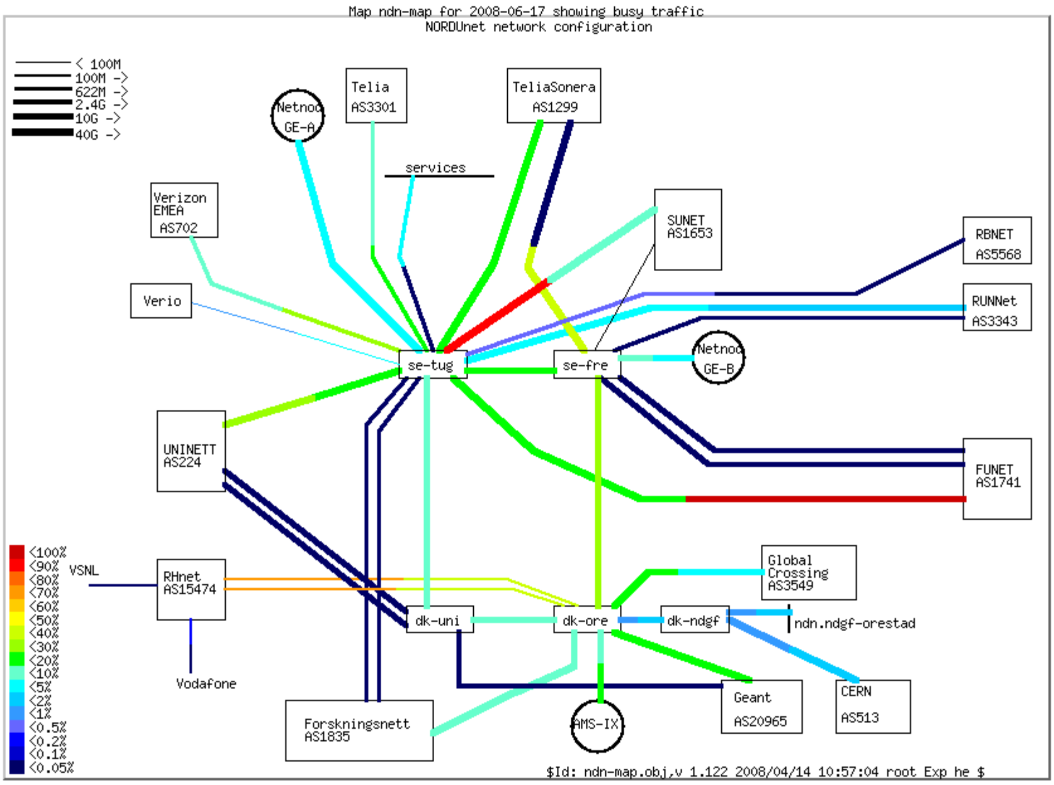

Figure 4: Mh-On Streaming test route.

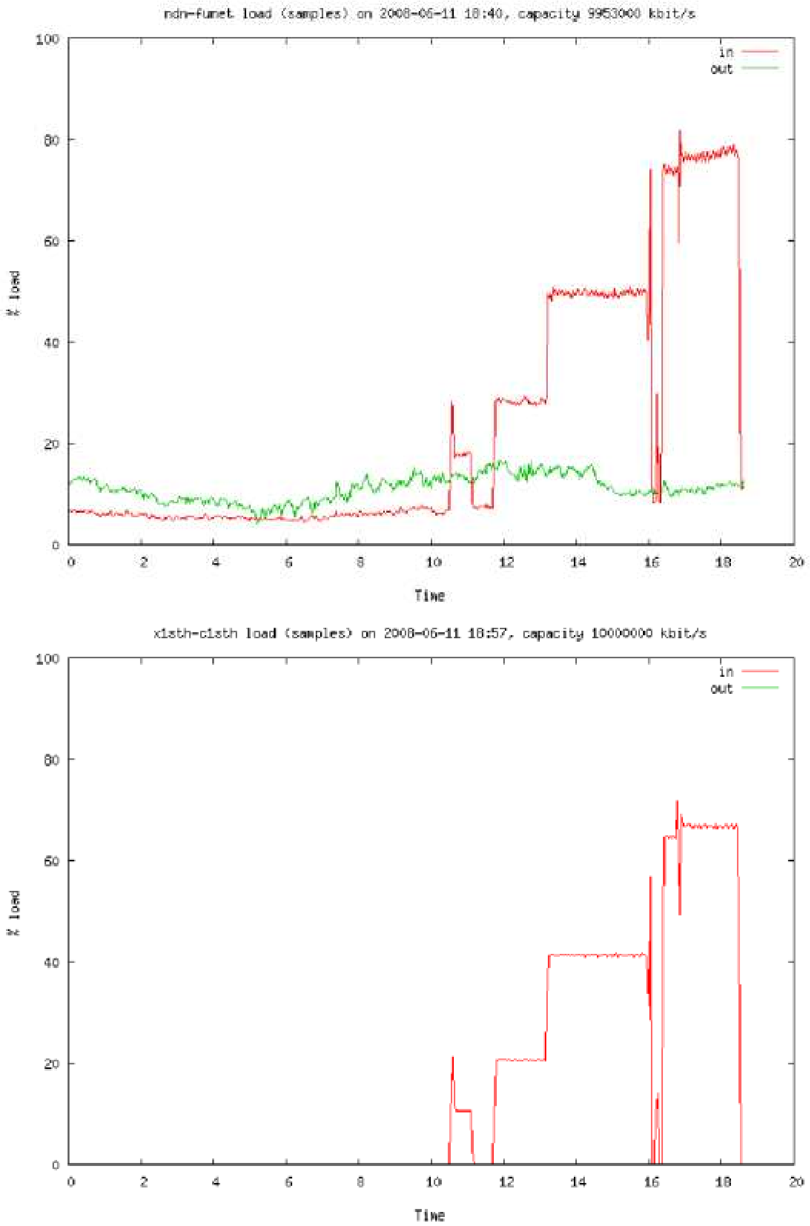

Figure 5: Mh-On Streaming test traffic.

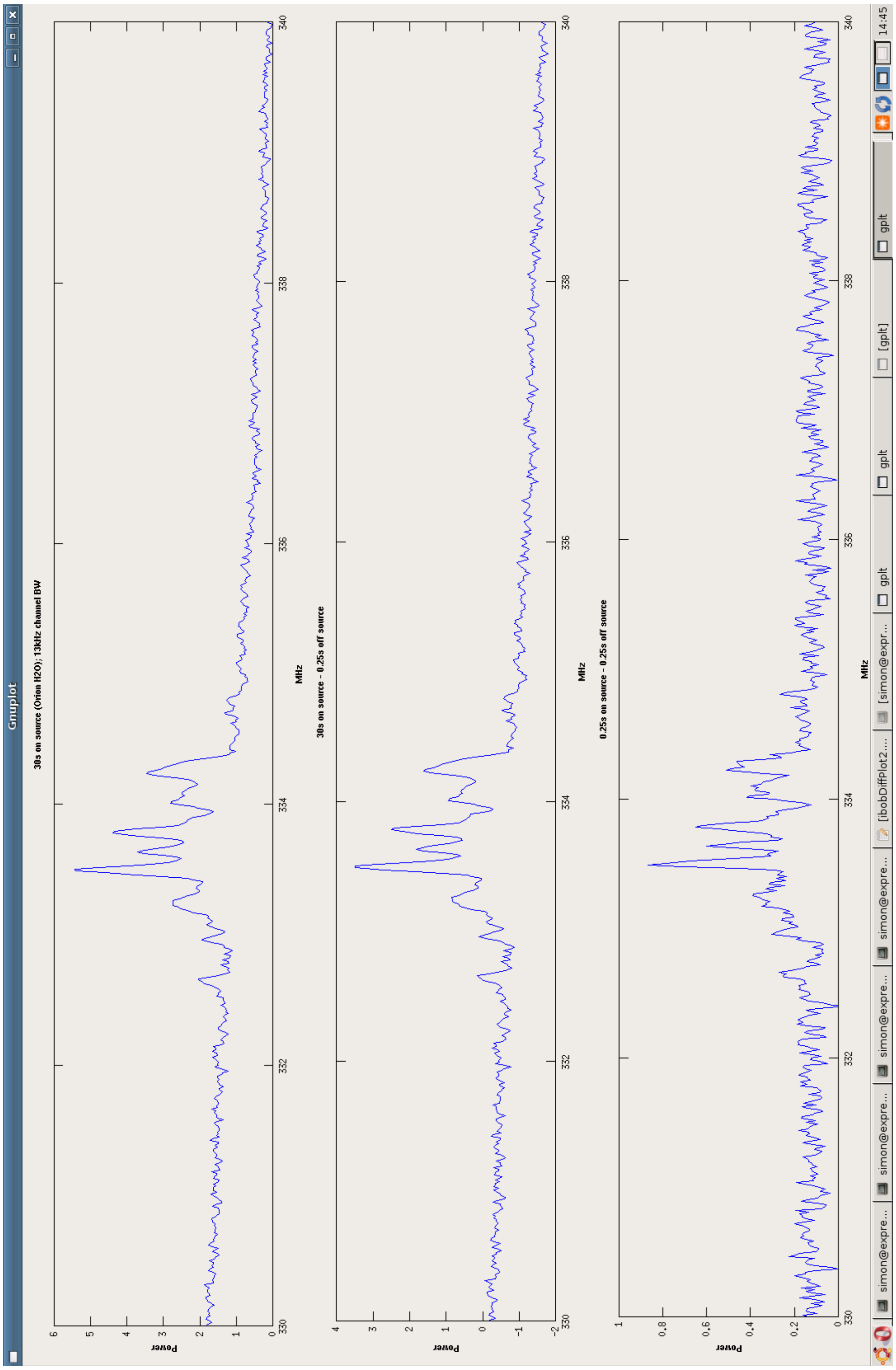

Figure 6: Gnuplot figure of Orion H2O.

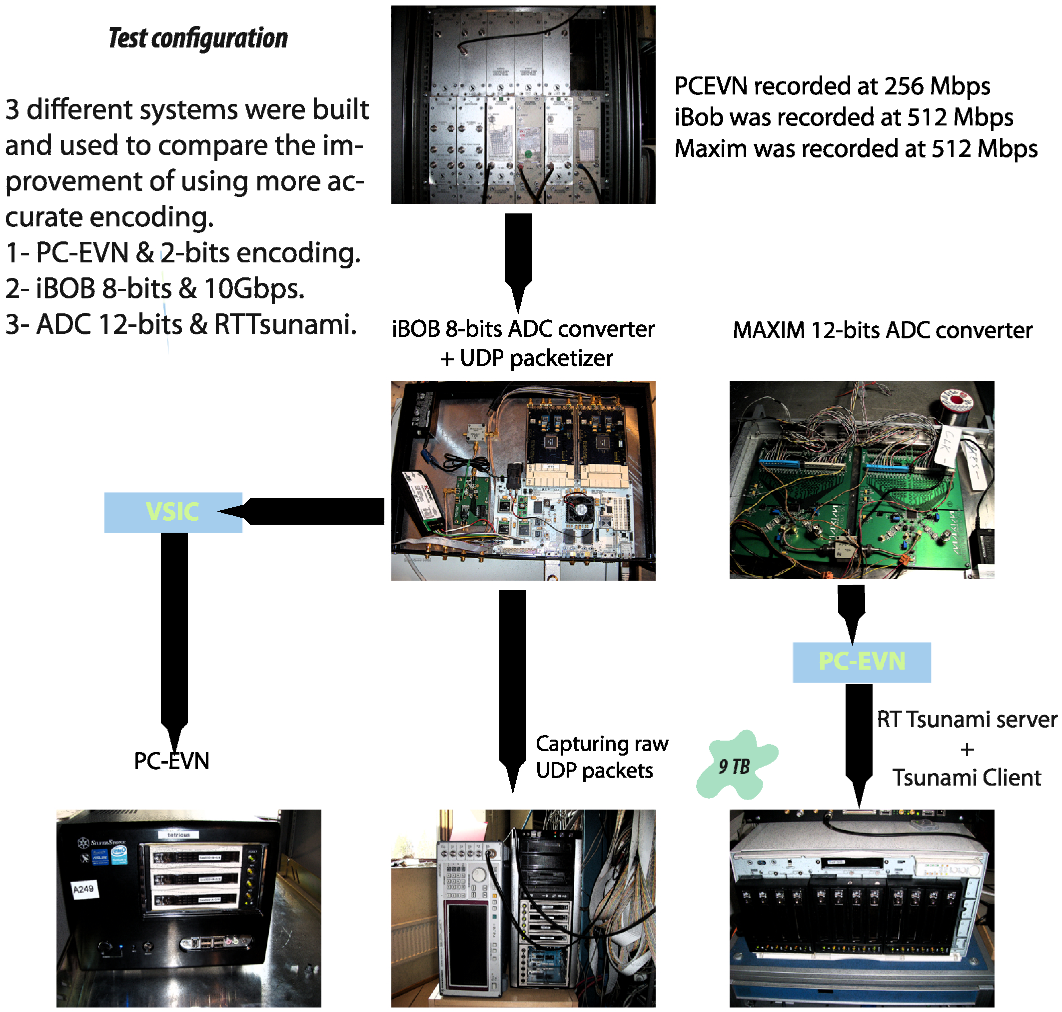

Figure 7: Configuration used for the latest tests, to record simultaneously using all three DASs.

systems. They have satisfactorily been tested in the latest mm and EVN session. Possible applications include spacecraft-tracking software, search for water on planets and a real-time spectrometer.

4.1 System configuration

Three different systems, as shown in Figure 7, were built, set up and compared for this study. Firstly, six years after the release of the VSIB new applications are still emerging for low-cost and easy usage. Secondly, iBOB is a board with enormous potential for RF applications. A simple IF-to-Ethernet sampler design has been one of the projects to test its performance. Finally, Maxim ADC evaluation kits allowed us to produce a high-sampler rate device at low cost.

PC-EVN uses a custom VLBI configuration with a Mark IV formatter and recording to a PC-EVN. Data is dual-channel and the recording rate is 128 Mbps. For the iBOB RF data for two bi-polarized channels is filtered using MarkIV baseband converters and sampled at 512 MHz. The firmware designed discards samples at a rate of 1/32. Total effective throughput is equal to 256 Mbps (16 Msps). The VHDL design also includes a UDP packetizer to send the raw packets through both 10 Gbps interfaces. Finally the data is acquired in a recording computer equipped with a dual port 10 Gb Chelsio Ethernet board. Figure 8 shows the VHDL Simulink code used in our iBOB firmware.

Two Maxim ADC boards were mounted, wired and connected to digitize the analog filtered signal similar to what was done with the iBOB. The mode can be set to either 8 or 12 bits resolution. Data up to 1 Gsps can be recorded by using VSIB or any kind of VSI bus interface. In our case, a system with 9 TB of storage capacity was used to handle the high recording rates.

4.2 Building process

iBOB was purchased and brought to Metsähovi at mid-October 2007 and it took approximately one month to set up the whole environment. The system includes the iBOB board version 1.3, with its chip Virtex-2 and two 10 Gbps Ethernet CX4 sockets. In addition, it included an ADC board which can process 1 Gsps at 8 bits resolution and a LMX 2351 frequency synthetizer to provide a reference clock of 2048/1024 GHz. At present two iBOB systems are ready at the laboratory for testing purposes. Figure 9 shows the enclosure designed and its contents.

The VHDL design for this target is simple. The input comes from two channels for the filtered data provided by the BBCs. With an external reference frequency it was sampled at 512 Msps in both channels. Data was dropped by a factor of 1/8 and then packetized in 64 bits packet format to feed the CX4 drivers. Finally UDP packets were sent via both Ethernet connections to a Data Acquisition Server.

In order to build a low-cost multi-bit ADC sampler a couple of Maxim ADC 1217 evaluation kits were purchased. With each board it is possible to sample two analog channels at 126 Msps and 8/12 bits resolution. The sampler designed consists of two ADC boards which enabled us to use two or four channels up to 512 Msps. Special wiring was required to convert the manufactorer's LVDS output style to the 50/40 connectors for recording with the VSIB. The wiring added to the system is shown in Figure 10.

The digital data can be easily recorded using the VSIB interface and a PC-EVN. Detailed information of the Maxim enclosure and the wiring set is available at the Metsähovi web site [6].

4.3 Results

The first astronomical tests were performed during the geodetic session in the beginning of March 2008 using the systems for spacecraft tracking. All three systems recorded simultaneously for several days and the results were posteriorly compared. Later, the test was repeated during the EVN session at the beginning of June 2008. The systems were used again to observe several sources to explore the existance of water and the water level.

At first glance both experiments concluded satifactorily and each system operated as expected. On the other hand, the noise from the baseband back-end or from the extra wiring, required us to insert a low pass filter to decrease the out-of-band noise. In any case, the results showed a strong need for large capacity storage systems due to the high rates involved in the amount of data stored. Finally, Figure 11 shows a comparison spectra of five minute scans of the maser W75NB at 22 GHz measured with the iBOB and PC-EVN. In addition, Figure 12 compares all the three devices recording VEX signal with PCal enabled.

4.4 Applications

The spectrometer software [7] written at Metsähovi continues the work around the Cell processor. The aim is to develop real-time spectra by using the Cell processor included in PlaySation 3. The software runs on both the Cell and Intel processors. By now, it has been succesfully tested to search for the

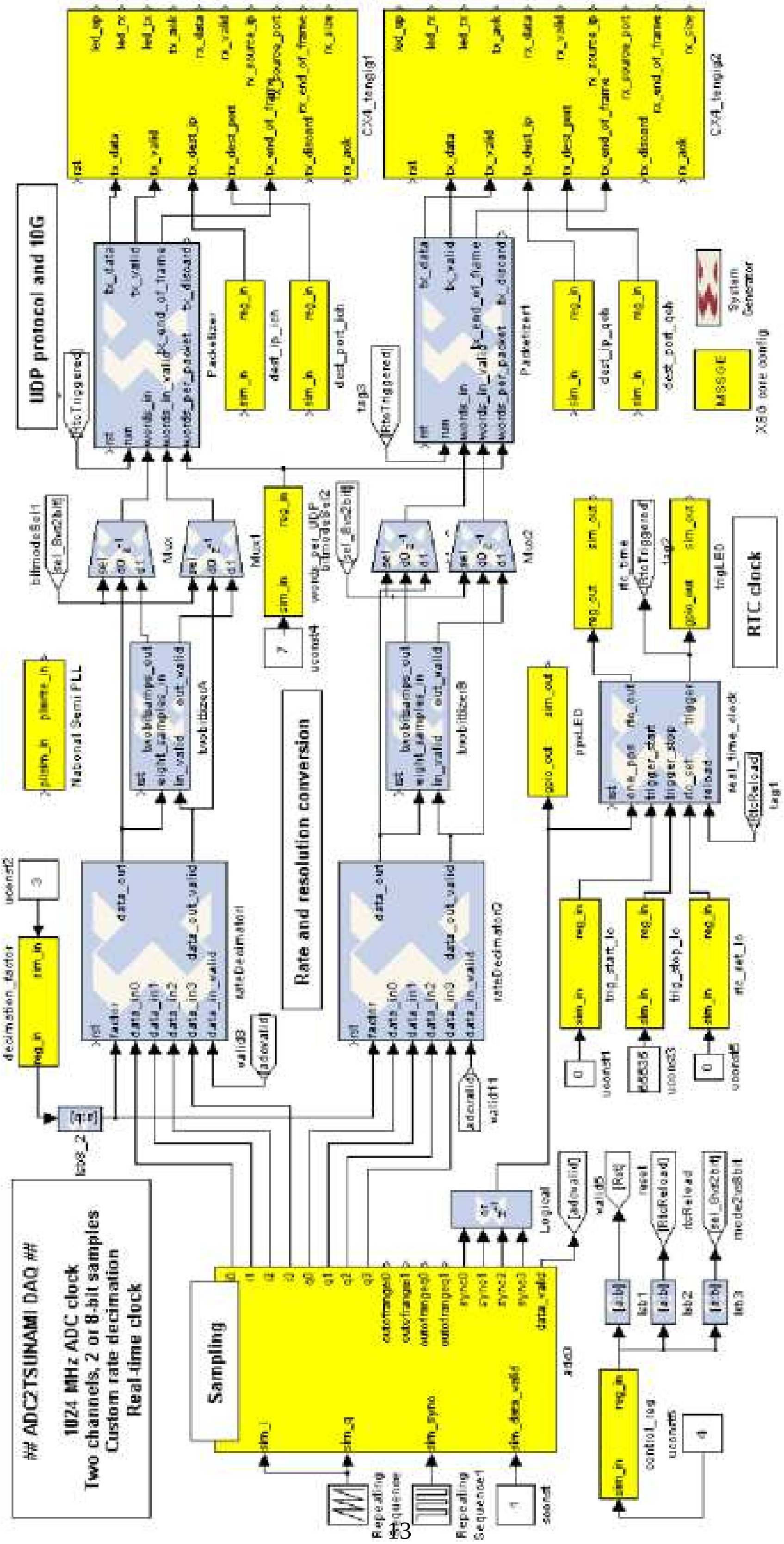

Figure 8: The VHDL code for the iBOB firmware consists of an ADC, resampler, packetizer and finally the two 10 Gbps modules.

Figure 9: A current picture of the iBOB, two ADC sampler modules and the frequency synthetisizer.

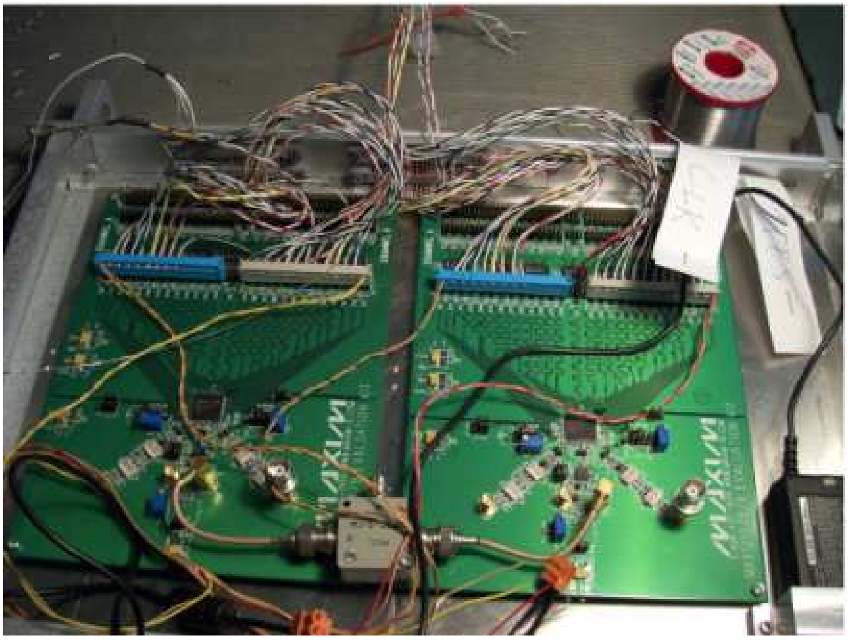

Figure 10: Building the enclosure to allocate both Maxim ADC samplers and the new wiring to convert the LVDS output mode to VSIB compatible.

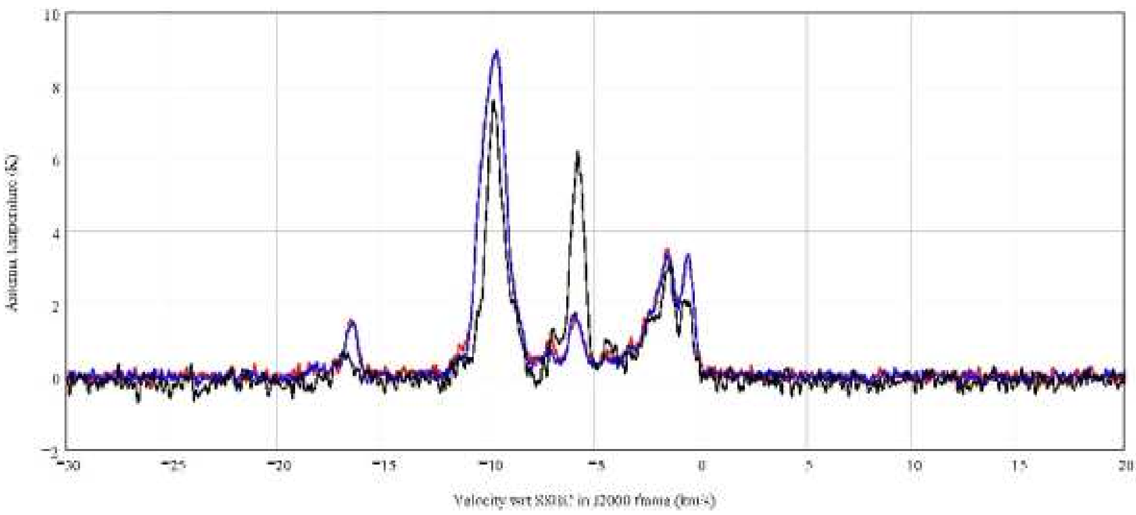

Figure 11: W75NB spectra during 06/02 (red), 06/03 (black) and 06/04 (blue) 2008. In the two first cases data was acquired with PC-EVN and for 06/04 iBOB (8-bits) was used.

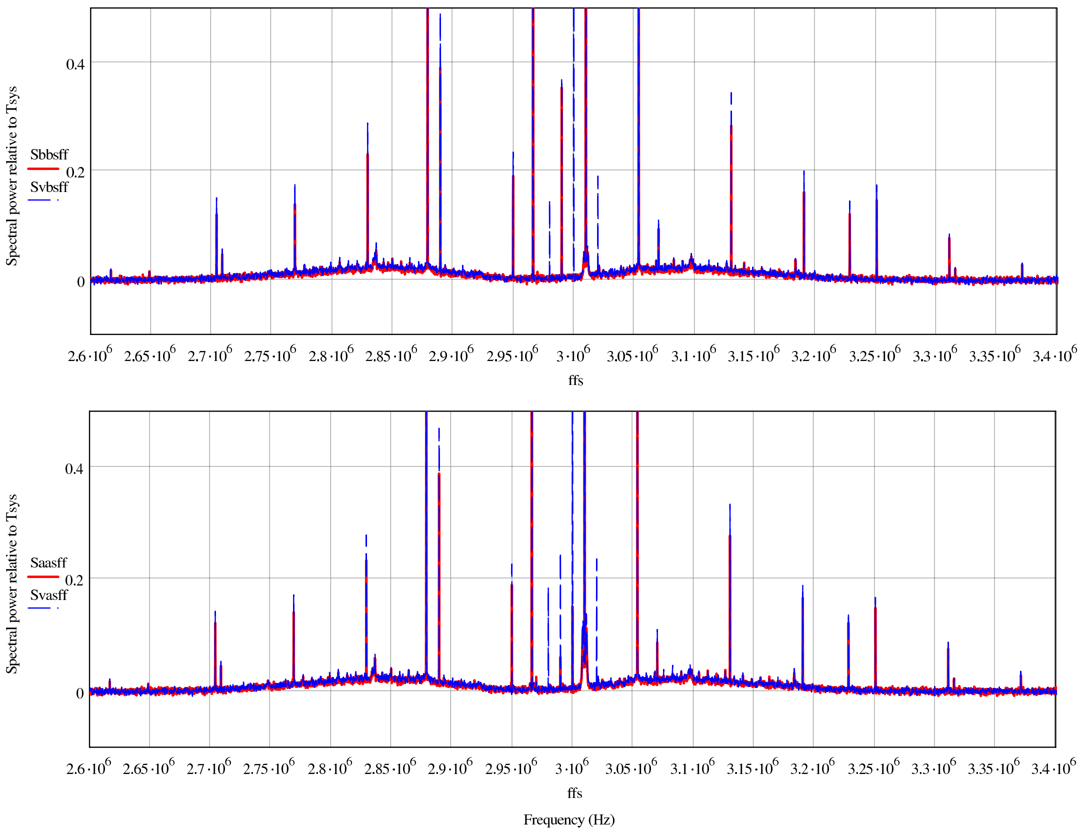

Figure 12: Upper image shows iBOB (red) and VSIB (blue) spectra and lower image shows Maxim (red) and VSIB (blue) data. S/C carrier signal is centered at 2.97 mHz and PCal is at 3.01 MHz.

existance of water in maser systems. The work completes the VSIB real-time spectrometer, a fast tool to check the performance instantaneously from each BBC [8].

Lately, a Tracking Spacecraft software has been started and it is currently in alpha stage. The systems were tested to record succesfully more than 50 hours of data form MEX and VEX S/Cs and now they are being analyzed to exract channel information. The project is based on the experience gained through Huygens and SMART-1 adventures by S. Pogrebenko.

References

[1] The Tsunami web page, http://tsunami-udp.sf.net/ July 2008

[2] The Extreme Networks web page, http://www.extremenetworks.com/ July 2008

[3] CASPER, http://casper.berkeley.edu/

[4] IBOB, http://casper.berkeley.edu/wiki/index.php?title=IBOB July 2008

[5] Xilinx, http://www.xilinx.com/

[6] Maxim 1217 ADC, http://www.metsahovi.fi/en/vlbi/misc-hardware/ July 2008

[7] Cell task library, http://cellspe-tasklib.sourceforge.net/ July 2008

[8] http://www.metsahovi.fi/en/vlbi/vsib-tools/ July 2008