PROCEEDINGS OF SCIENCE

Commissioning and Using the 4 Gigabit Lightpath from Onsala to Jodrell Bank

Richard Hughes-Jones[1]
DANTE
City House,126-130 Hills Road, Cambridge CB2 1PQ , UK
And
Visiting Fellow, School of Physics and Astronomy, The University of Manchester,
Oxford Rd, Manchester UK
E-mail: Richard.Hughes-Jones@dante.net

Jonathan Hargreaves[2]
Jodrell Bank Centre for Astrophysics, The University of Manchester,
Oxford Rd, Manchester, UK
E-mail: hargreaves@jive.nl

This paper describes the installation, commissioning and testing of the 4 Gigabit Lightpath from Onsala to Jodrell bank which is being used as part of the EXPReS project to send data from the telescope at Onsala to the new WIDAR correlator at Jodrell Bank. The network path is truly multi-domain, it crosses multiple administrative domains, uses equipment from different manufacturers, and both Ethernet and SDH framing technologies are used on different portions of the path. Measurements performed using both PC and deterministic Field Programmable Gate Array techniques are described and results on the performance and stability of the 4 Gigabit path are shown. Finally, plots demonstrating the successful sampling and movement of data from the telescope to the correlator are presented.

Science and Technology of Long Baseline Real-Time Interferometry:
The 8th International e-VLBI Workshop - EXPReS09
Madrid , Spain
June 22-26, 2009

[1] Speaker
[2] Now at: JIVE, Oude Hoogeveensedijk 4, 7991 PD, Dwingeloo, NL

PoS(EXPReS09)035

1.Introduction

The FABRIC/JRA1 work package of EXPReS investigated moving VLBI data at 4 Gigabit/s from the telescope at Onsala to the new WIDAR correlator at Jodrell Bank. The aim was to correlate data from Onsala with that from eMERLIN telescopes [1][2]. To do this, a multi-domain 4 Gigabit path had to be established from Onsala to Jodrell Bank using lightpaths supplied by the National Research Networks (NRENs), NORDUnet, and GÉANT, the international backbone.

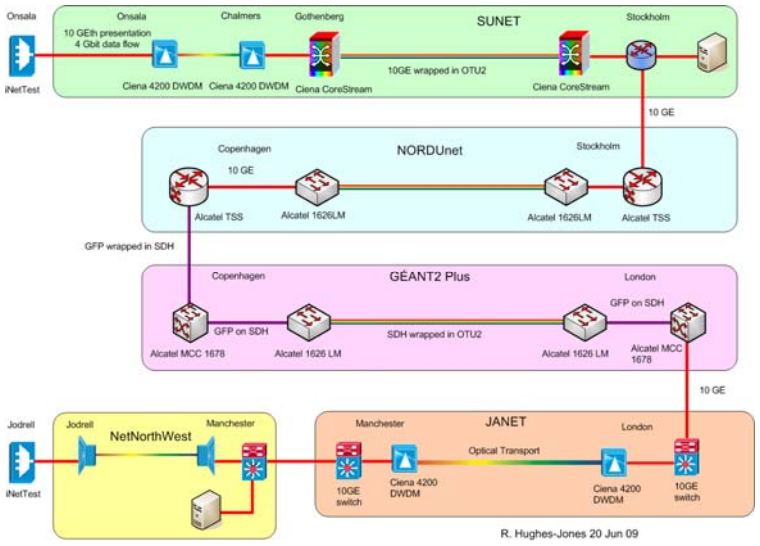
Section 2 describes the network path and the various stages of the implementation, which started with the setting up of a "test path". Earlier e-VLBI work [3], confirmed by the studies in ESLEA [4] and EXPReS [5] [6], indicated that it is best to use the UDP/IP network protocol for moving real-time VLBI data. The UDP based tests used to characterise the link are described in Section 3 and results discussed in Section 4. Finally, some plots from the WIDAR correlator are shown indicating that test signals sampled at Onsala can be sent over the network at 4 Gigabits and successfully received by the correlator.

2.Details of the Network Path

This part of the EXPReS project required VLBI signals from two polarisations each sampled at 1024 MHz with 2-bit resolution. This gives a data rate of 4.096 Gbit/s. Encapsulation of this data in application, UDP, IP and Ethernet headers resulted in a requirement to send 8274 bytes over the Ethernet every 16 us giving a wire rate of 4.137 Gbit/s. 28 VC-4s were provisioned on the SDH sections of the path, give a possible transfer rate of 4.193 Gbit/s. This was sufficient to carry the Ethernet data as well as the GFP wrapping and VCAT overheads. On the 10 Gigabit Ethernet sections of the path the ingress policing was set at 4.2 Gbit/s. All the VLBI equipment connected to the network used 10 Gigabit Ethernet physical interfaces.

The final multi-domain end-to-end network path is shown in Figure 1. It crosses multiple administrative boundaries: the local campuses at Onsala and Jodrell Bank, the regional networks at the Universities of Manchester and Gothenburg, the NRENs SUNET, and JANET, the NORDUnet regional network and the GÉANT Plus international backbone. On different portions of the path Ethernet or SDH framing technologies are used and equipment from many different manufacturers is involved.

The international connection was provisioned in several stages starting with the "test path" from Stockholm to London to gain confidence that everything would actually work. For this, NORDUnet supplied a dedicated Lambda, framed as 10 Gigabit Ethernet, at the optical layer. At Copenhagen, this was connected to a GÉANT Plus circuit to London provisioned as 28 VC-4 over SDH running on the Alcatel MCC cross-connect platform, as shown in Figure 2. Test PCs were installed at the PoPs in Stockholm and London



.***Figure 1:*** *Diagram showing transport and connectivity details of the final path Onsala to Jodrell Bank.*

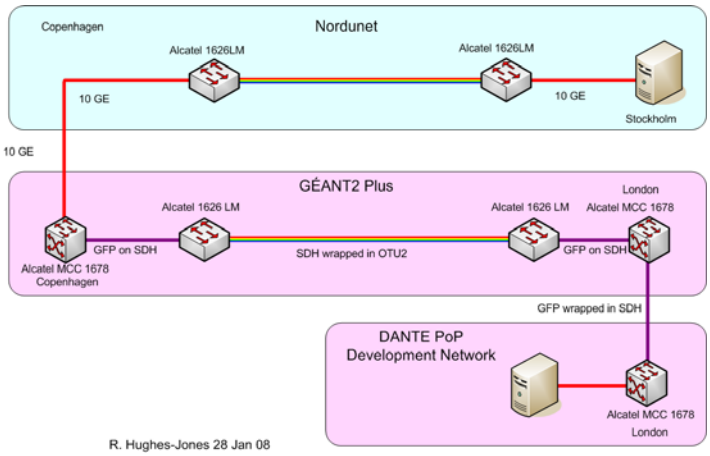


Figure 2: *Diagram of the" test path" from Stockholm to London.*

During 2008 NORDUnet transitioned their backbone to an Alcatel TSS cloud which allowed more flexible provision of Ethernet or SDH circuits. JANET initially used SDH to supply the path from London to Manchester over UKLIGHT and then transitioned UKLIGHT to a 10 Gigabit Ethernet backbone. Further network tests were performed as this work progressed.

3.Testing Methodology

First, lab tests were performed to establish the performance of the PCs, built using the Supermicro X7DBE motherboard, two Dual Core Intel Xeon Woodcrest 5130 2GHz CPUs, 4 G Bytes of memory, and the PCI-Express 10G-PCIE-8A-R 10 Gigabit Ethernet NIC from Myricom. It was found that they could successfully operate at 9.8 Gigabit/s.

A program call udpmon [7] was run on the PCs at either end of the path to send a stream of carefully spaced UDP packets between the two hosts. The UDP throughput and packet loss were measured as a function of the spacing between the frames with the interrupt coalescence on the network interface cards (NIC) set to 25 μs, the standard value for the Myricom 10 GE

cards. The interrupt coalescence was turned off for measurements of the inter-packet arrival times, which were histogramed, and when the relative one-way delay of each packet from sender to receiver was recorded for a set of packets. Prior to the jitter and one-way delay measurements, the frequency difference and phase offset between the two PC CPU clock signals was determined. This information was used to relate the measurements of time made on the two PCs.

While creating the firmware for the units that would move the VLBI data, a design called iNetTest [8] was produced to allow the iBoB [9] Field Programmable Gate Array (FPGA) hardware to perform as a two port 10 Gigabit Ethernet network test device. The iNetTest FPGA was designed to transmit streams of UDP/IP packets at regular intervals and the throughput and packet arrival times were measured at the iNetTest receiver. This operation is similar to the udpmon program, but unlike a PC, the FPGA is deterministic and had a time resolution of 5 ns. iNetTest to iNetTest measurements were made on the path between Onsala and Jodrell.

4.Results from Testing the Network Path

4.1Data Obtained from the Test Path

Figure 3 shows the received "wire" rate UDP throughput and packet loss as a function of the transmitted packet spacing for UDP for various packet sizes. The left hand plots show the data for packet sent from Stockholm to London with no Ethernet flow control, and the right hand plots show packets from London to Stockholm with flow control enabled. The packet size refers to the user data and the "wire rate" makes allowance for the UDP, IP and Ethernet frame overheads and the minimum inter-frame gap. (This corresponds to an extra 66 bytes.) On the right hand side of both throughput plots, the curves show a 1/t behaviour, where the delay, t, between sending successive packets is most important. When the frame transmit spacing is such that the data rate would be greater than the available bandwidth, one would expect the curves to be flat, as is the case.

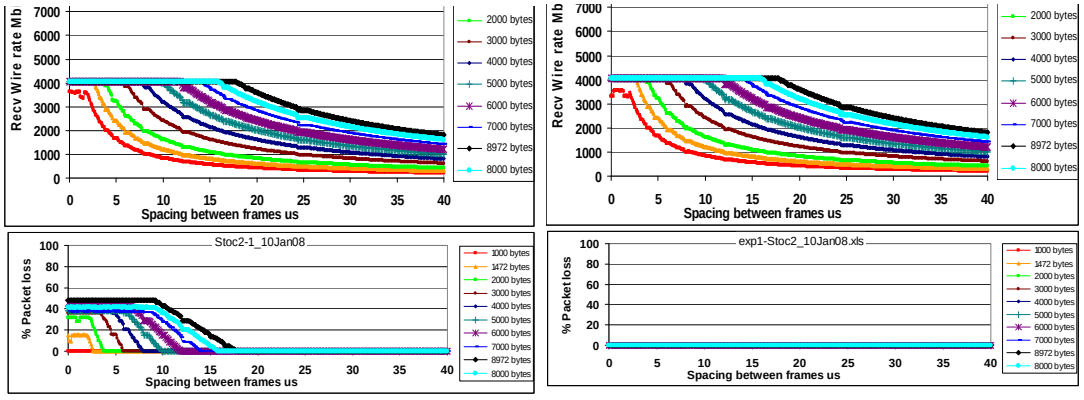


Figure 3: *Throughput and packet loss as a function of packet spacing for various packet sizes. Left: Stockholm to London. Right: London to Stockholm.*

With no flow control, there is packet loss for spacing less than ~15 µs. This is expected as more packets are being sent than can be carried by the 28 VC-4 circuit. When the Ethernet flow

control is enabled, the sending host is prevented from sending too fast and there is no packet loss. There was no packet loss for 8192 byte packets at 16 µs spacing in either direction.

The udpmon program was used to investigate the packet jitter. Figure 4 shows histograms of the received inter-packet spacing for 8972 byte packets sent from Stockholm to London with a spacing of 16.5 µs with the interrupt coalescence turned off. The main peak is in the 16 µs bin, as expected, but there are smaller secondary peaks at ~19 and 51 µs and a low level tail out to ~90 µs.

These measurements gave encouragement that the 4 Gigabit path would meet the requirements of EXPReS.

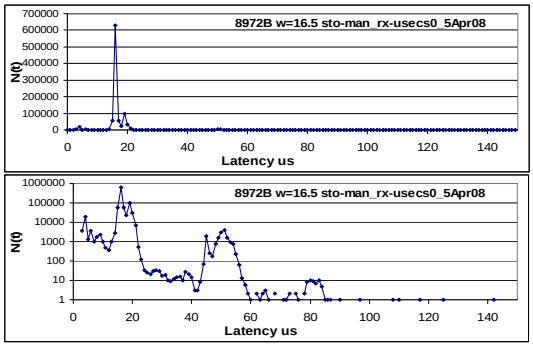


Figure 4: *Histograms of the received inter-packet spacing for 8972 byte packets sent from Stockholm to London with a spacing of 16.5 µs on the "Test Path". The main peak is at 16 µs as expected.*

4.2The Performance with TSS

The throughput, packet loss, and packet jitter tests described in section 4.1 were repeated as the 4 Gigabit path was extended and similar results were obtained when the path was established over JANET & NetNorthWest from London to Jodrell Bank. When NORDUnet transitioned their backbone to an Alcatel TSS cloud however, over 10% of the 8192 byte packets were lost when sending at 16 µs, the spacing required for user data throughput of 4096 Mbit/s.

The device responsible for the packet loss was located by using udpmon to send 1 million packets from Stockholm and checking the number entering the TSS cloud in Stockholm, the number leaving the TSS cloud in Copenhagen, the number of packets entering the 10 GE interface Alcatel MCC in Copenhagen and the number of packets being passed to the SDH circuit in the Copenhagen MCC. All packets, for every offered rate, traversed the Alcatel TSS and were received by the Alcatel MCC without loss. However not all were passed to the SDH circuit section inside the MCC at Copenhagen, hence causing the packet loss. The loss as a function of the offered rate suggested a classic bottleneck in the MCC.

In order to investigate if the reason for the packet loss was due to packet bunching, udpmon was used to send UDP flows from Manchester to the PC in Stockholm and the packet jitter and relative 1-way delay of the packets received was recorded. This direction was chosen because the PC in Stockholm had a full 10 Gigabit Ethernet presentation to the TSS cloud in

Stockholm and because the use of 28 VC-4s between the GÉANT Plus MCCs meant that packets could not leave the MCC in Copenhagen with spacing closer than ~15.7 µs, which corresponds to 4.2 Gbit/s. When using the TSS, the packet jitter histogram changed dramatically as shown in Figure 5. There was no peak at the expected 16 µs, as measured on the "Test Path" and shown in Figure 4, however there were peaks at ~4 &6 µs which correspond to frames arriving at line rate i.e. 10 Gbit/s, and a very long tail.

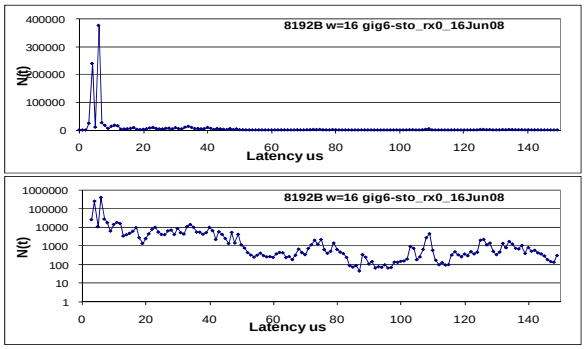


Figure5: *Histograms of the received inter-packet spacing for 8192 byte packets sent over the path using TSS from Manchester to Stockholm with a spacing of 16.5 µs. The main peaks are at 5&7 µs indicating packets arriving at line speed i.e. 10 Gigabit/s.*

Using the data from the relative one-way delay measurements for 8192 byte packets sent from Manchester to Stockholm with a spacing of 16 µs, the difference between the arrival times of successive packets was calculated and is shown in Figure 6 as a function of the time the packet was received. There are periods of time, about 130 µs long when no packets arrive which are followed by periods were the packets arrive with spacing ~ 6 µs. This is consistent with bursts of packets at the 10 Gigabit line speed, confirming the conclusions derived from the jitter plots.

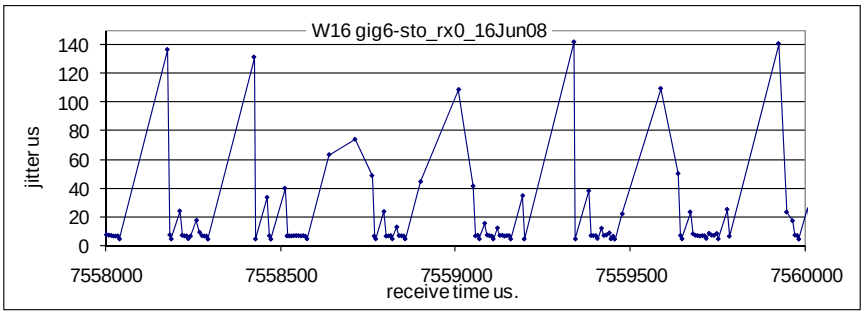


Figure 6 *Separation between received packets as a function of the time the packet was received. The peaks indicate the gaps, about 130 µs long, were no packets arrive.*

The tests suggest that extracting the Ethernet frames from the SDH transport in the Alcatel TSS caused bursts of packets to be transmitted at 10 Gigabit line speed. These bursts exceeded the buffer capability of the Alcatel MCC unit which was next device in the path. As a work around, NORDUnet and DANTE used spare interfaces to configure a SDH path all the way from Stockholm to London. This avoided the TSS SDH to Ethernet transition and tests indicated that now there was no packet loss.

4.3Stability of the Final 4 Gigabit link

The stability of the multi-domain network path was determined by using the iNetTest devices to send trials of 100M 8192 Byte packets with a spacing of 16µs from Onsala to Jodrell and measuring the achievable UDP throughput, the packet loss and the inter-packet jitter. Each trial took about 27 minutes and consecutive trials were repeated immediately. For each trial the data throughput was measured as 4.094 Gbit/s, with no variation between trials. Over a typical set of about 40 trials, the loss rate was $\leq 10^{-9}$ (approx bit error rate better than10^{-13}).

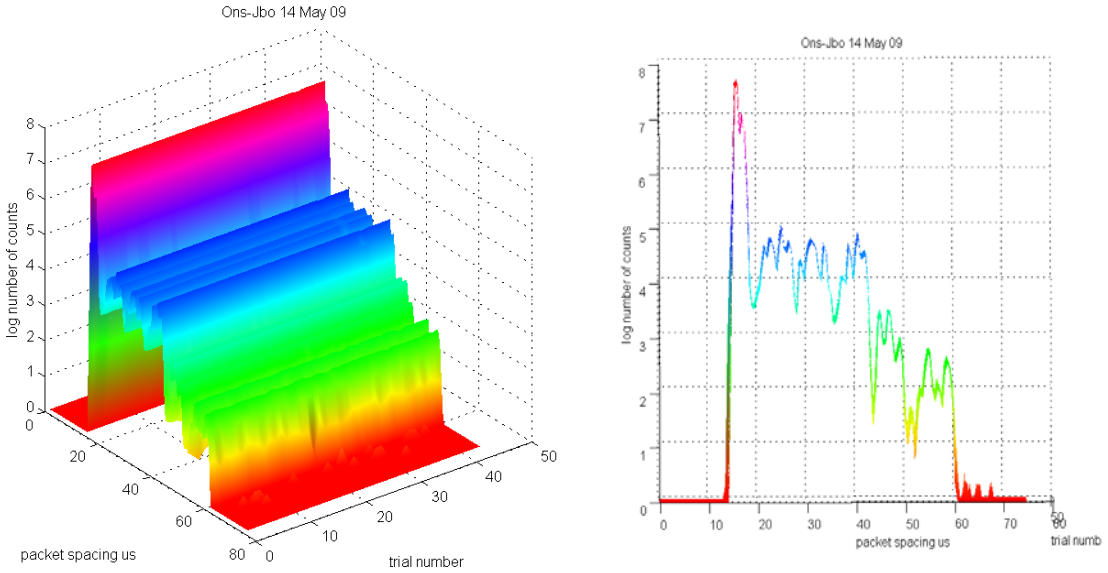


Figure 7: *Left: a three dimensional plot the inter-packet arrival times for 43 trials of 100M UDP packets from Onsala to Jodrell sent and measured by iNetTest units. Right: end projection of the plot. The plots indicate no variation in the distribution.*

Figure 7 shows a three dimensional plot the inter-packet arrival times for the trials described above and their projection. The main peak at 16 µs has a full width half maximum of ~1µs and there are tales extending to ~70 µs, but the tails are a factor of 10^{-3} smaller. There was no change in the shape of the distributions of the inter-packet arrival times for these trials and very similar distributions were observed for tests made over several weeks. The throughput, packet loss and jitter measurements indicate that the link, with it's one-way delay of 18.8 ms, is extremely stable. Also no out-of order packets were detected.

5.Moving Data from Onsala to the Correlator

Signals from regular e-MERLIN antennas are sampled at the telescope and then sent to the WIDAR correlator where they are received by the Station Boards (SBs) [10] before being passed to Baseline Boards for cross-correlation. The SBs de-formats the incoming data stream, and splits the wideband data into several sub-bands through the use of filter banks. Data from antennas external to e-MERLIN, such as that sent from Onsala using the iBOBs, also passes through a SB in a similar fashion.

For diagnostic purposes, it is possible to examine histograms of the state counts at the input of the SB, and also in the filter banks. The state counts give a count of the number of data

points in the signal which were detected at a certain voltage level. For a regular sinusoid, one would expect to see a 'U' shaped histogram, where the outer peaks represent the time spent in the highest and lowest parts of the waveform. For white noise, the histogram has a bell shape. Figure 8 shows a histogram of the state counts when the input data was an 88 MHz sinusoid, sampled by an iBOB at Onsala, and transmitted to an iBOB at Jodrell, where it was fed into the SB. The clear 'U' shape confirms successful transmission of the data over the network.

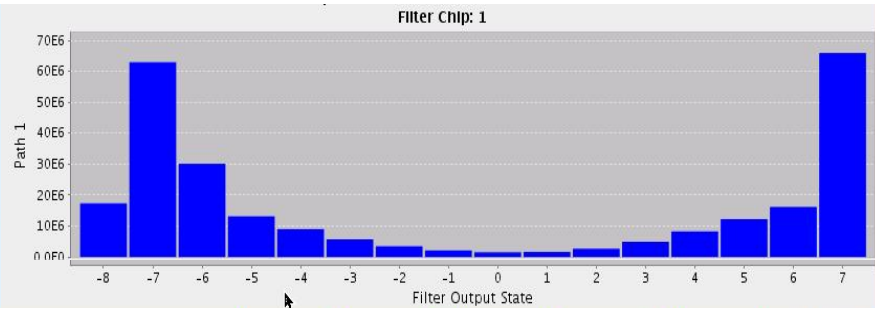


Figure 8: *State count histogram from e-MERLIN Station Board filter chip. The Input signal was a 88 MHz sinusoid generated at Onsala and transmitted over the network using iBOBs to Jodrell.*

As the amplitude of input signal to the sampler is strengthened, more counts would be detected at the maxima since the wave would be clipped and would resemble a square wave. Likewise if the signal is attenuated, counts in the outer bins would decrease. Figure 9 shows the histograms obtained when the amplitude of the sine wave input signal to the digitising iBoB was altered. The changes in the histograms are as expected, again confirming the successful operation of moving the data from the sampler over the network to the correlator.

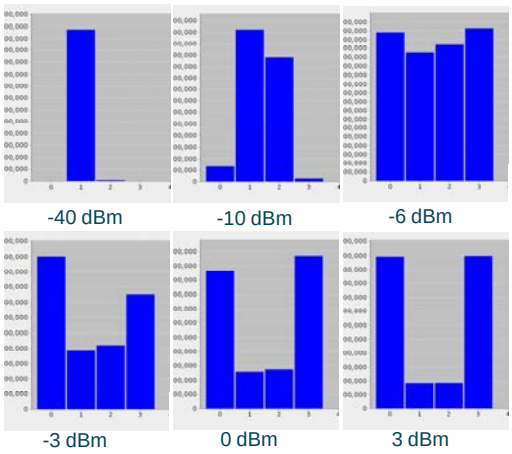


Figure 9: *State count histograms from the WIDAR Station Board input chip, showing the variation of the state count histogram as the amplitude of the input sine wave was altered.*

6.Conclusions

The measurements on the "Test Path" gave encouragement that the 4 Gigabit path would meet the requirements of EXPReS, and allowed the times of the arrivals of successive packets to be estimated. This allowed specification of the buffer sizes in the FPGA designs for transmitting and receiving the VLBI data.

The iNetTest units have been used to make extensive tests of the 4 Gigabit network between Onsala and Jodrell Bank. A packet loss problem, caused by the bunching of packets by one system causing buffer overflow on the following network hardware, was fully investigated and understood. Once DANTE and NORDUnet devised a work around, the network has proved extremely stable with reproducible packet jitter, no packet loss and no packets out of order.

Tests using sine wave signals injected locally at Jodrell and sine wave signals injected at Onsala and transmitted over the 4 Gigabit link to Jodrell have been successfully received in the Station board, and further work is in progress to enable the correlator to process incoming data from Onsala and the e-MERLIN telescopes.

7.Acknowledgements

Throughout the development of the 4 Gigabit link for the EXPReS project we have had encouragement, close co-operation and valuable technical discussions with network personnel from the Jodrell, Manchester & Onsala campuses, the Metropolitan Area Networks, the National Research Networks JANET, SUNET, NORDUnet, and DANTE. We would like to thank all who have been involved, and this support has been vital in making the project an operational success.
EXPReS is funded by the European Commission (DG-INFSO), under the Sixth Framework Programme, Contract #026642.

References

[1] e-MERLIN Web site http://www.e-merlin.ac.uk/ , viewed 27/7/09.

[2] R. Spencer, *Progress and Status of e-MERLIN*, POS (EXPReS09) 029

[3] Richard Hughes-Jones, Steve Parsley , Ralph Spencer, *High Data Rate Transmission in High Resolution Radio Astronomy - vlbiGRID*, FGCS Special Issue: IGRID2002 Vol 19 (2003) No 6, August 03

[4] R.E. Spencer, P. Burgess, S. Casey, R. Hughes-Jones, S. Kershaw, A. Rushton, M. Strong, A. Szomoru and C. Greenwood, *The contribution of ESLEA to the development of e-VLBI* in "Lighting the Blue Touchpaper for UK e-Science - Closing Conference of ESLEA Project", PoS, March 2007 (http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=41#session-5)

[5] EXPReS Protocol Strategic Report D3 – interim report.

[6] EXPReS Protocol performance report D150.

[7] udpmon: a Tool for Investigating Network Performance, http://www.hep.man.ac.uk/~rich/net viewed September 2009.

[8] R. Hughes-Jones & J. Hargreaves, *iNetTest a 10 Gigabit Ethernet Test Unit*, POS (EXPReS09) 043

[9] IBOB web page http://casper.berkeley.edu/wiki/IBOB viewed August 2009

[10] http://www.drao-ofr.hia-iha.nrc-cnrc.gc.ca/science/widar/private/Station_Board.html Then select Station Board VSI Test FPGA.