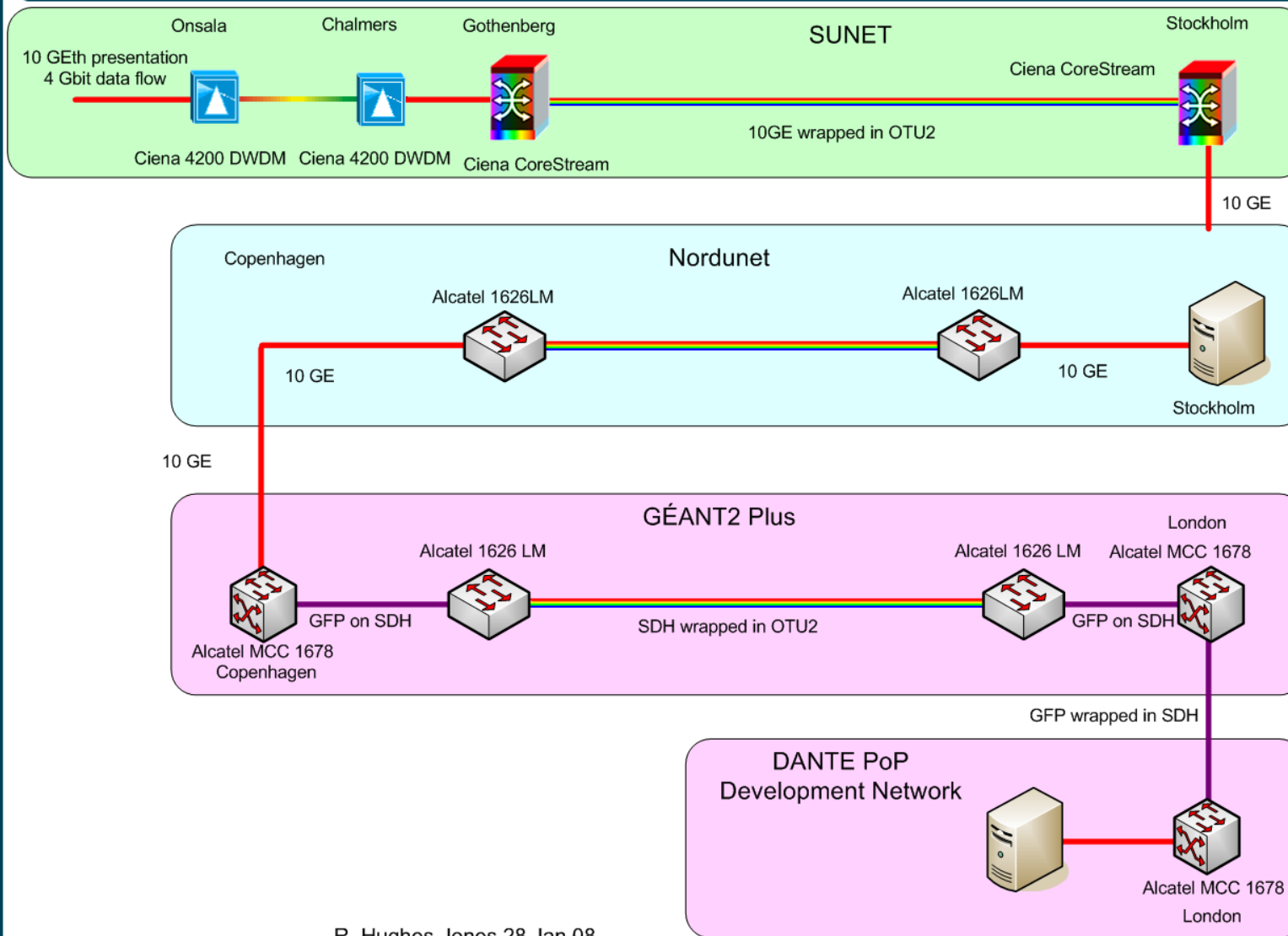


Commissioning and Using the 4 Gigabit Lightpath iNetTest & eMerlinIN

Richard Hughes-Jones

DANTE Delivery of Advanced Network Technology to Europe
8th International eVLBI Workshop, Madrid , 22-26 June 2009

The First Test Path



Multi

- Vendor
- Technology
- Domain

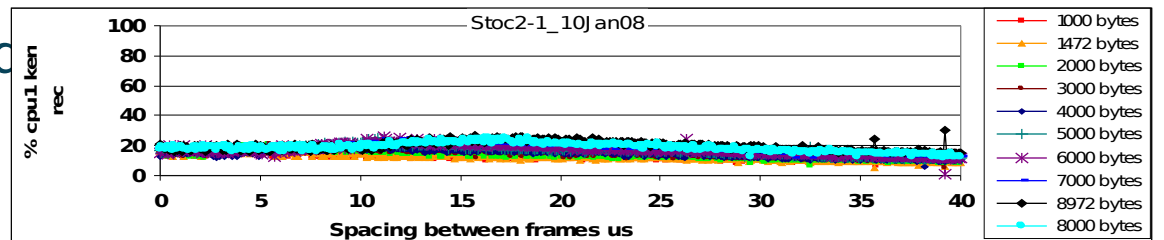
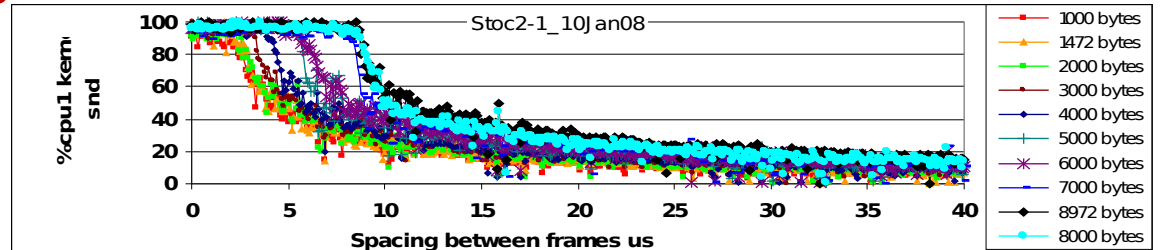
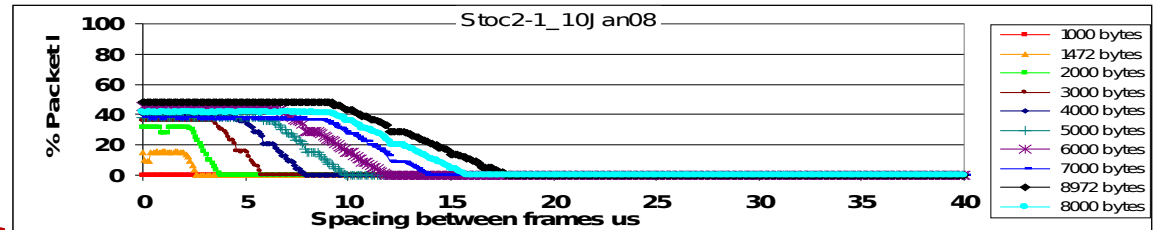
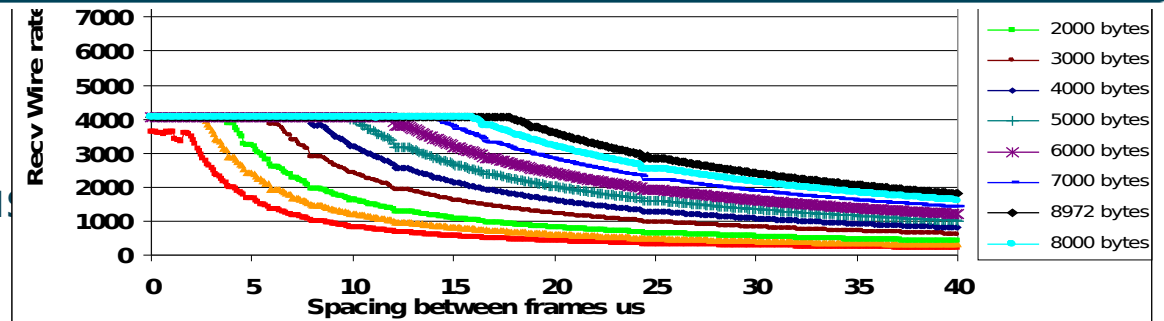
10 Gigabit Ethernet uses a Lambda over NORDUnet

4 Gig TDM/SDH Lighpath over GÉANT2+

R. Hughes-Jones 28 Jan 08

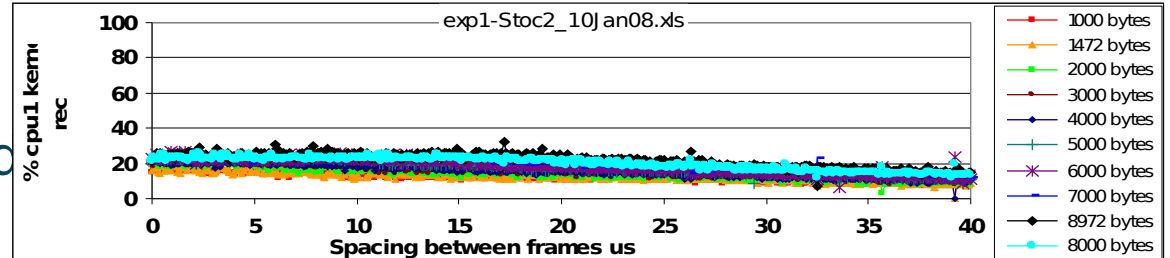
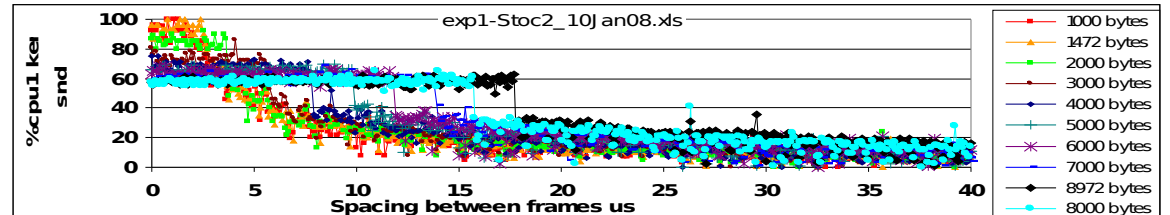
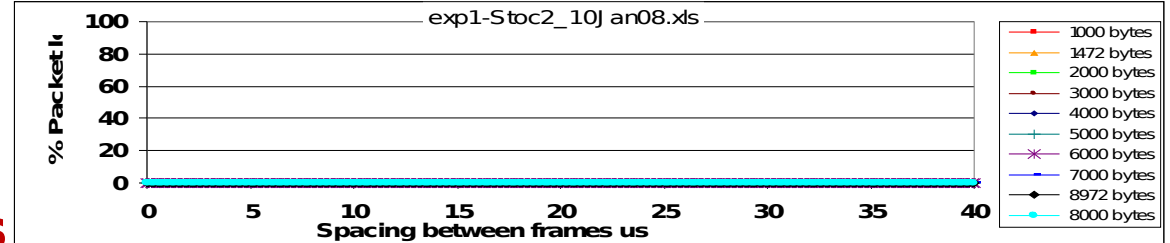
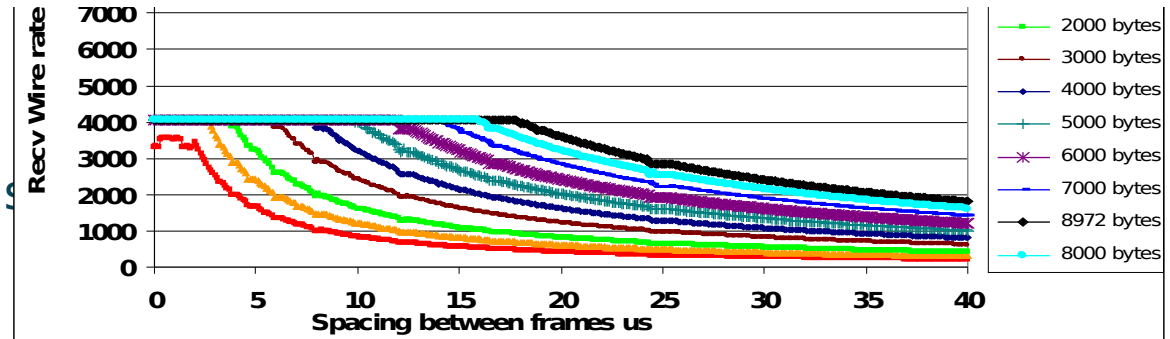
UDP Throughput Stockholm-London

- Alcatel Metro Core Connect
MCC Flow control OFF
- Kernel 2.6.20-web100_pktd-plu
- Myricom 10G-PCIE-8A-R Fibre
 - rx-usecs=25
 - Coalescence ON
- MTU 9000 bytes
- Max throughput **4.05 Gbit/s**
- Packet loss as expected**
- Sending host, 3 CPUs idle
- For $<18 \mu\text{s}$ spacing, 1 CPU is $>90\%$ in kernel mode
sending at 10 Gbit
- Receiving host 3 CPUs idle
- For $<20 \mu\text{s}$ packets, 1 CPU is $\sim 30\%$ in kernel mode



UDP Throughput London-Stockholm

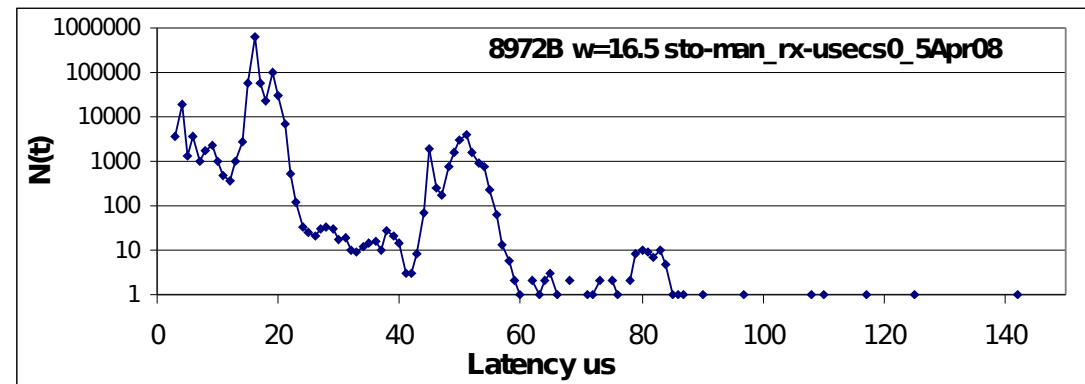
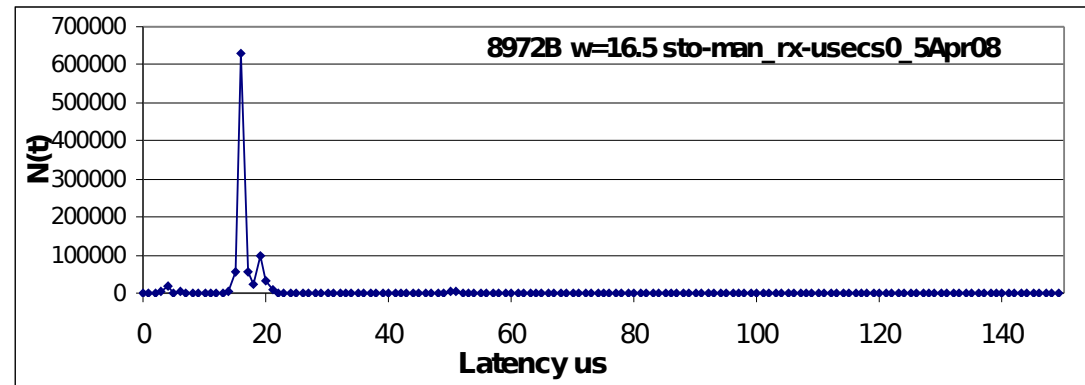
- Alcatel Metro Core Connect
MCC Flow control ON
- Kernel 2.6.20-web100_pktd-plus
- Myricom 10G-PCIE-8A-R Fibre
 - rx-usecs=25
 - Coalescence ON
- MTU 9000 bytes
- Max throughput **4.05 Gbit/s**
- NO Packet loss – flow control**
- Sending host, 3 CPUs idle
- For $<8 \mu\text{s}$ packets,
1 CPU is $\sim 60\%$ in kernel mode
sending at 4 Gbit
- Receiving host 3 CPUs idle
- For $<20 \mu\text{s}$ packets,



UDP Jitter Stockholm-London

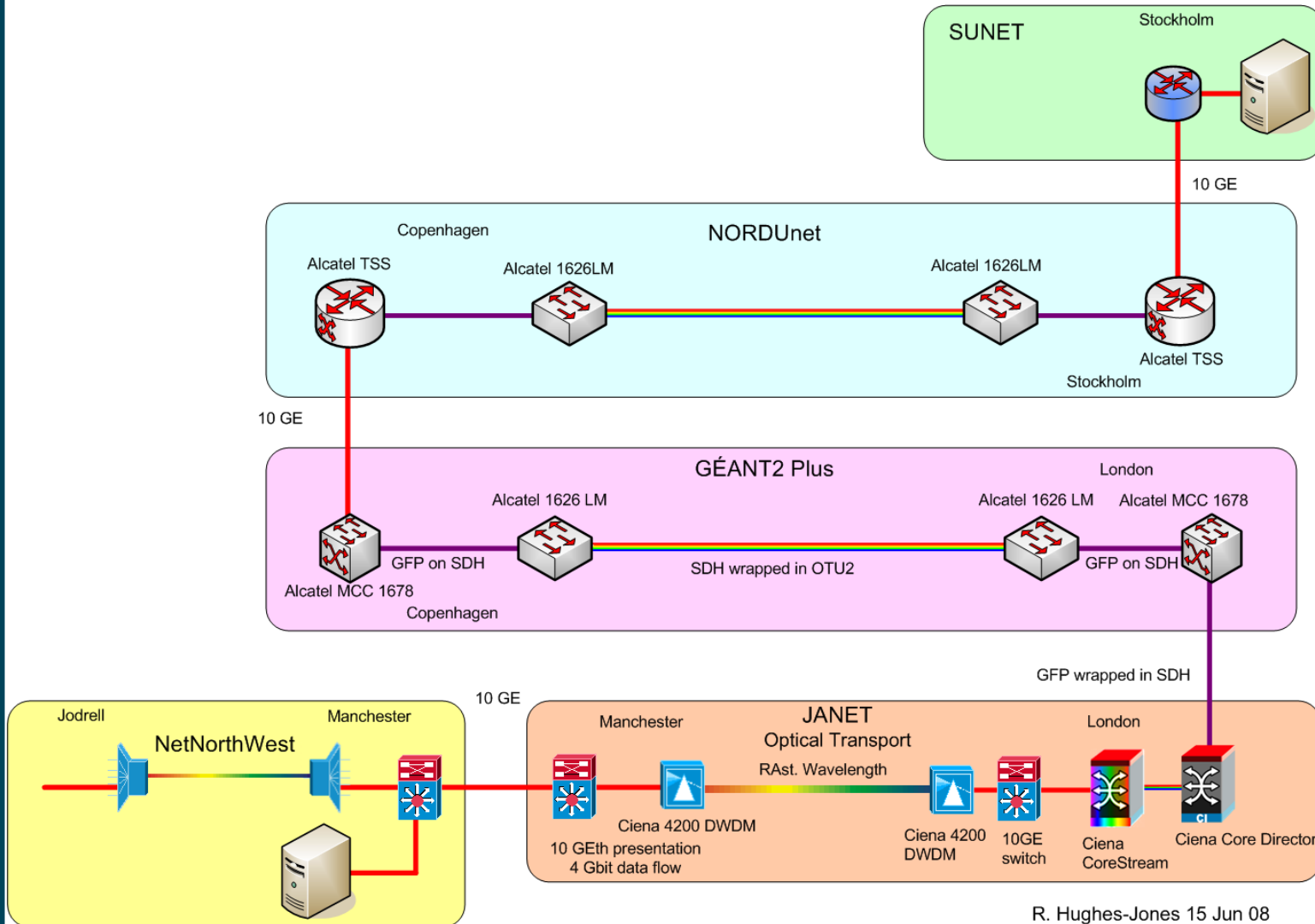


- Alcatel Metro Core Connect
MCC Flow control OFF
- Kernel 2.6.20-web100_pktd-plus
- Myricom 10G-PCIE-8A-R Fibre
 - rx-usecs=0
Coalescence OFF
- 8972 byte packets
- 16.5 μ s spacing
- **Peak in packet spacing at 16 μ s as expected**
- Other peaks at 10 and 51 μ s



NORDUnet install TSS

Test Stockholm – Manchester



4 Gigabit
Ethernet
Path over TSS
NORDUnet

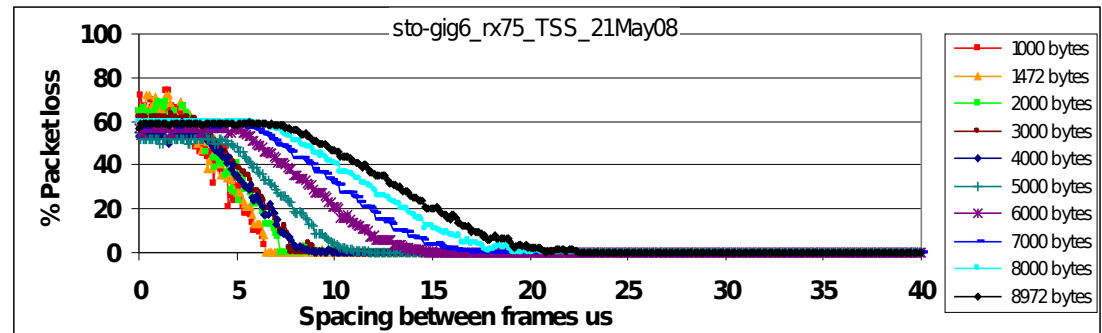
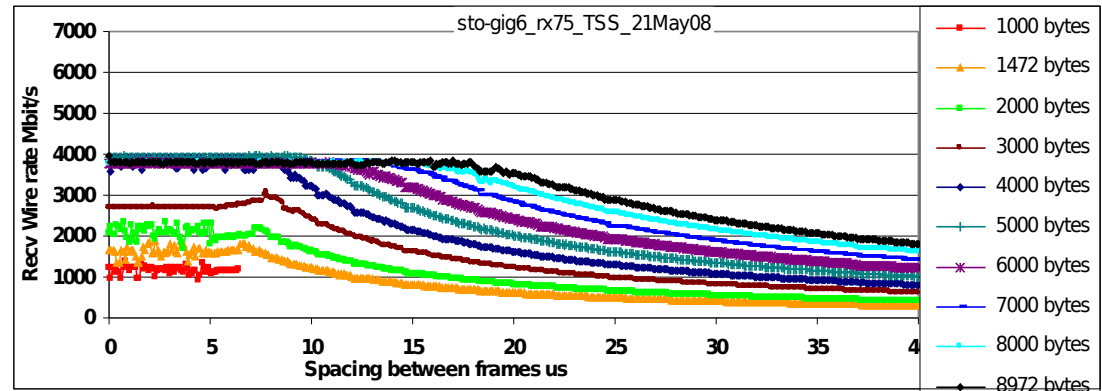
4 Gig TDM/SDH
Lighpath
over GÉANT2+

R. Hughes-Jones 15 Jun 08

UDP Throughput Stockholm-London



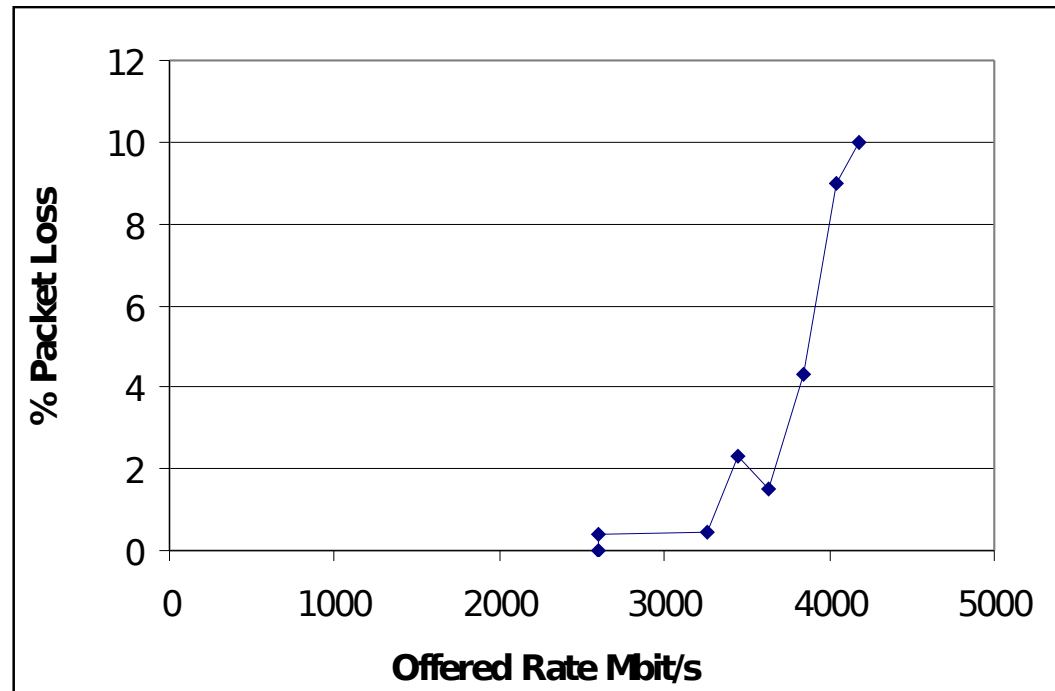
- Alcatel TSS – 10GE – MCC
- Kernel 2.6.20-web100_pktd-plus
- Myricom 10G-PCIE-8A-R Fibre
 - rx-usecs=75
 - Coalescence ON
- MTU 9000 bytes
- Max throughput **3.78 Gbit/s**
(Previously was 4.05 Gbit/s)
- **Packet loss 10%**
at packet spacing for 4.096 Gbit/s
with a 8129 byte payload.
- Manchester – Stockholm
NO packet loss at 4.096 Gbit/s



Packet loss Stockholm-Manchester



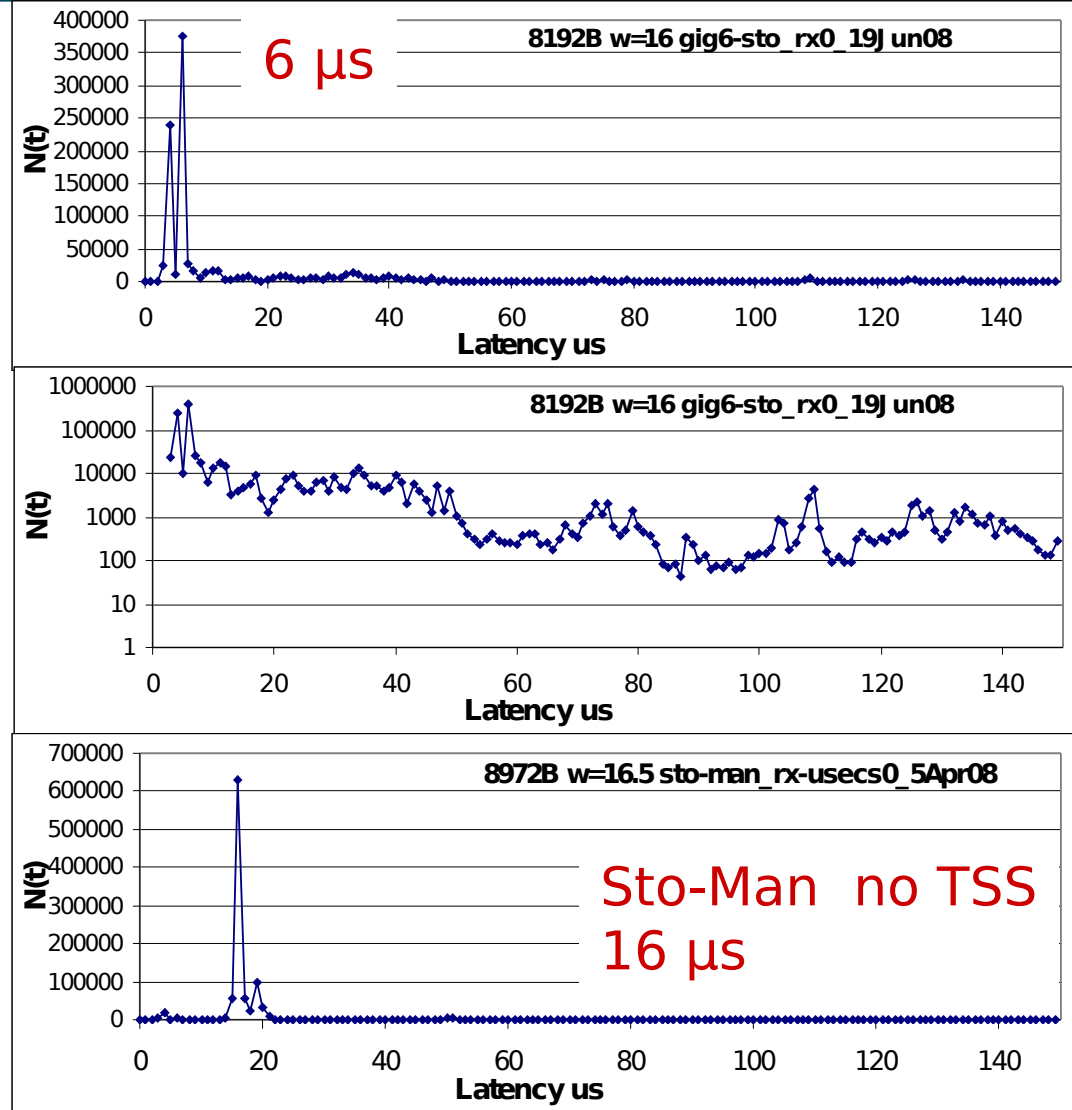
- Alcatel TSS – 10GE – MCC
- Send 1M packets from PC in Stockholm
- 8192 Bytes
- All cross the TSS cloud.
- All enter Alcatel MCC as 10 GE frames
- **Packet loss was in MCC Card**
- **~10% at 4096 Mbit/s**
- **Classic “bottleneck” performance.**
- **But WHY ??**



UDP Jitter Manchester-Stockholm



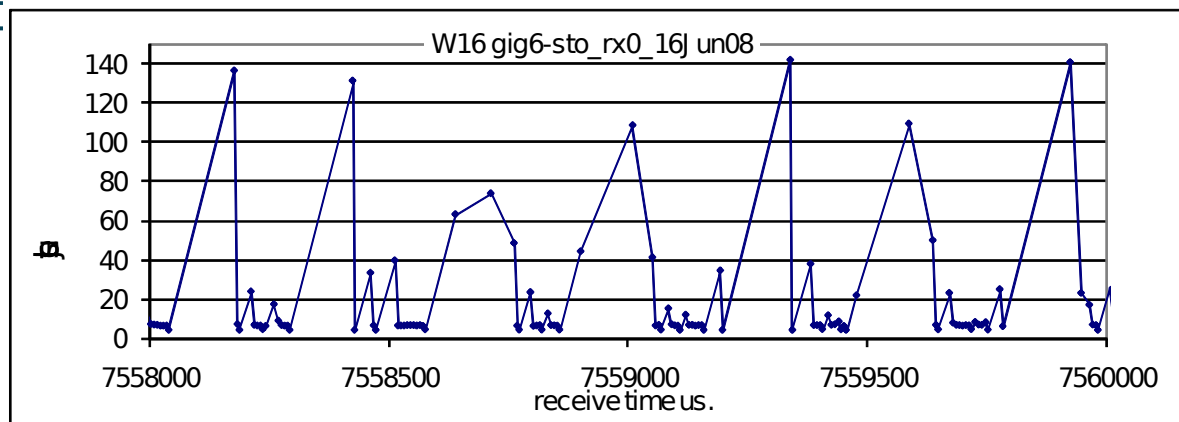
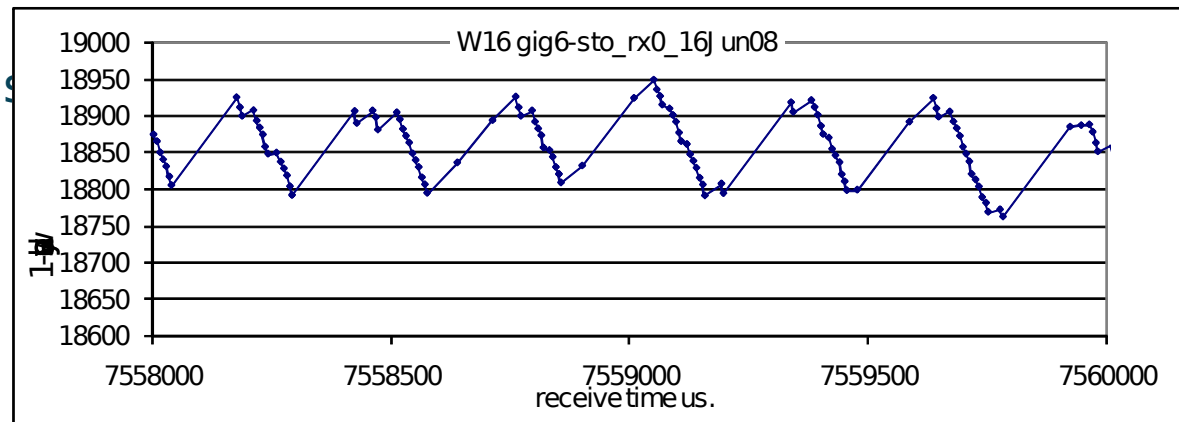
- Alcatel MCC – 10GE – TSS – PC
- Kernel 2.6.20-web100_pktd-plus
- Myricom 10G-PCIE-8A-R Fibre
 - rx-usecs=0
- Coalescence **OFF**
- 8192 Byte packets
- Alcatel MCC is TDM 28 VC-4. Can only send packets to the TSS at 16 μ s **not at 10 Gbit/s.**
- Peak packet spacing is ~ 6 μ s ie packets arriving at 10 Gbit/s
- Suggests packet bunching in the TSS network



UDP 1-way delay Manchester-Stockholm with



- Alcatel MCC – 10GE – TSS – PC
- Kernel 2.6.20-web100_pktd-plus
- Myricom 10G-PCIE-8A-R Fibre
 - rx-usecs=0
 - Coalescence **OFF**
- 8192 Byte packets
- Structure similar to interrupt coalescence but this is OFF
- **Jitter as function of recv time**
see delays of ~ 130 μ s
then packets at ~6 μ s
These packets at ~ 10Gbit/s
- This confirms packet bunching in the TSS network



Why do we get packet loss in Copenhagen?

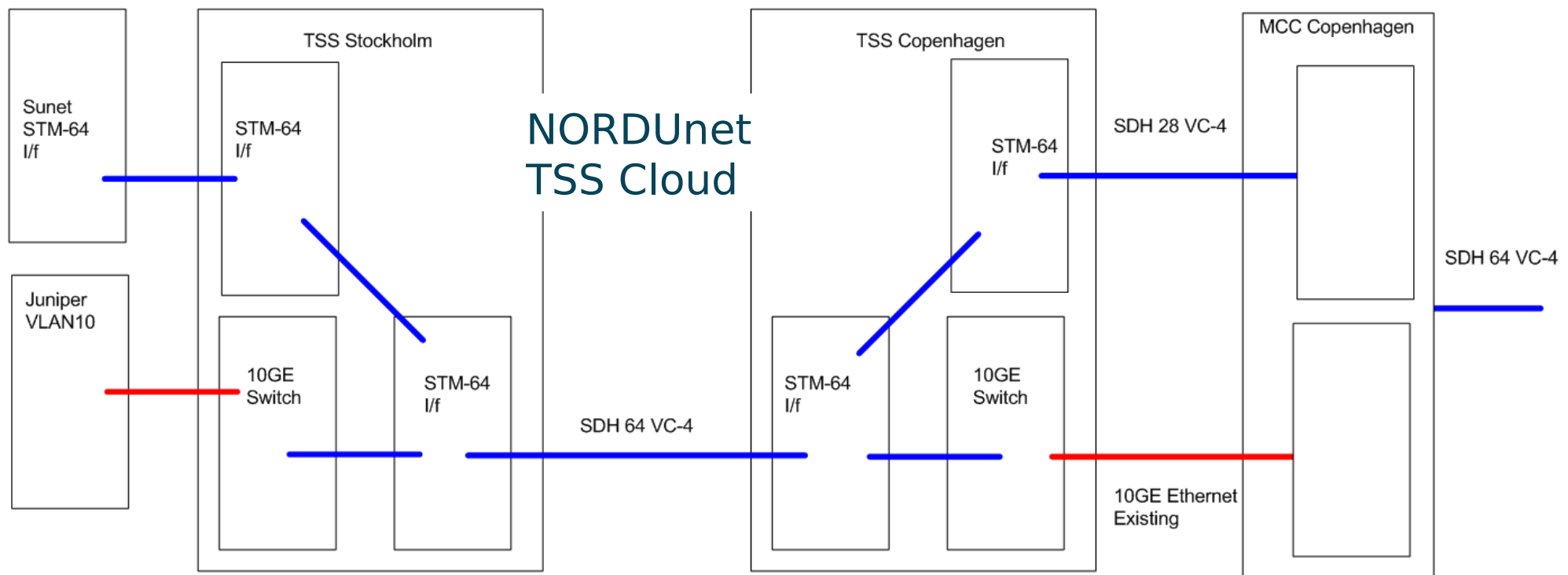


- The Alcatel TSS delivers the packets over the 10GE link to the Alcatel MCC in bunches at line-speed (with suitable pauses to get 4Gbit/s).
- The SDH circuit over GÉANT from the Alcatel MCC runs at 28 VC-4 (~4.2Gbit/s).
- The Alcatel MCC ingress buffering is only about 57 k Byte.
- These factors together cause buffer overflow and the packet loss.

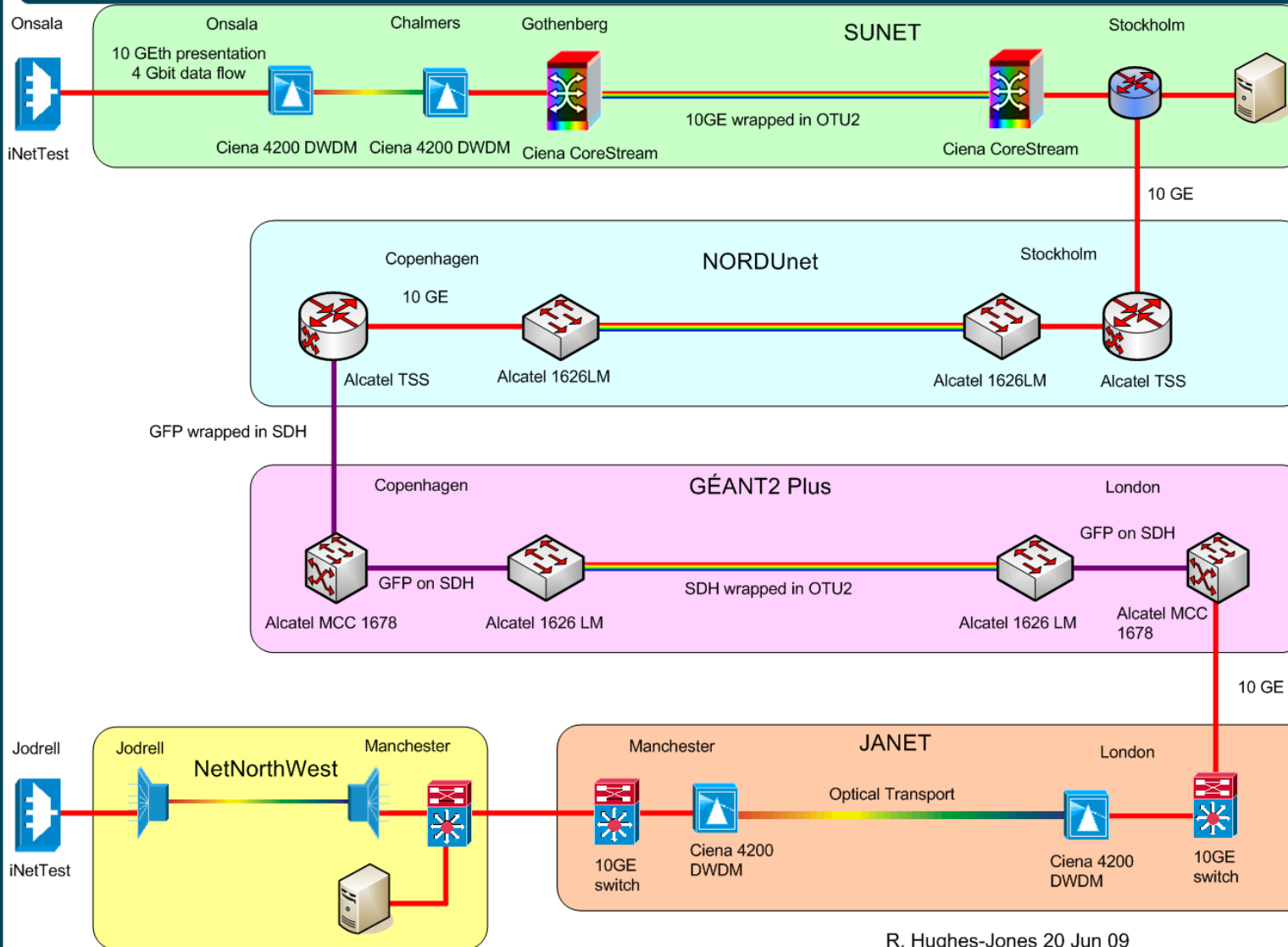
NORDUnet & GÉANT change the TSS – MCC link



- A work-around !
- VLBI uses SDH to GÉANT
- And buffering in the TSS



The Current 4 Gigabit link



28 VC-4 SDH
Links
TSS to MCC

4 Gig on
Ethernet
Over DWDM
JANET & NNW

R. Hughes-Jones 20 Jun 09

INETTEST

iNetTest: Simulink design

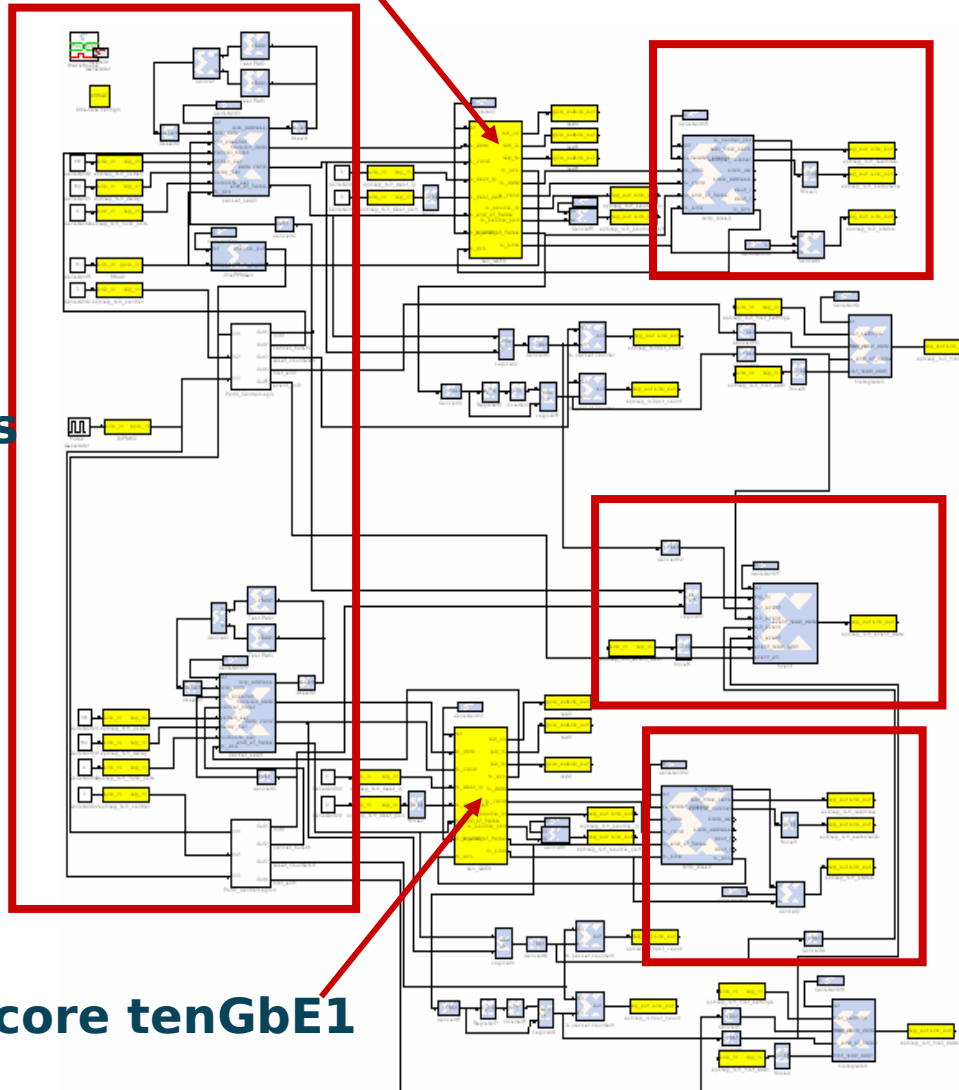


10GE MAC core tenGbE0

**Network
parameters
CSRs**

10GE MAC core tenGbE1

Jonathan Hargreaves JBO



**Inter-packet time
Histogram tenGbE0**

**Transmit & receive
event time log**

**Inter-packet time
Histogram tenGbE1**

iNetTest: Details

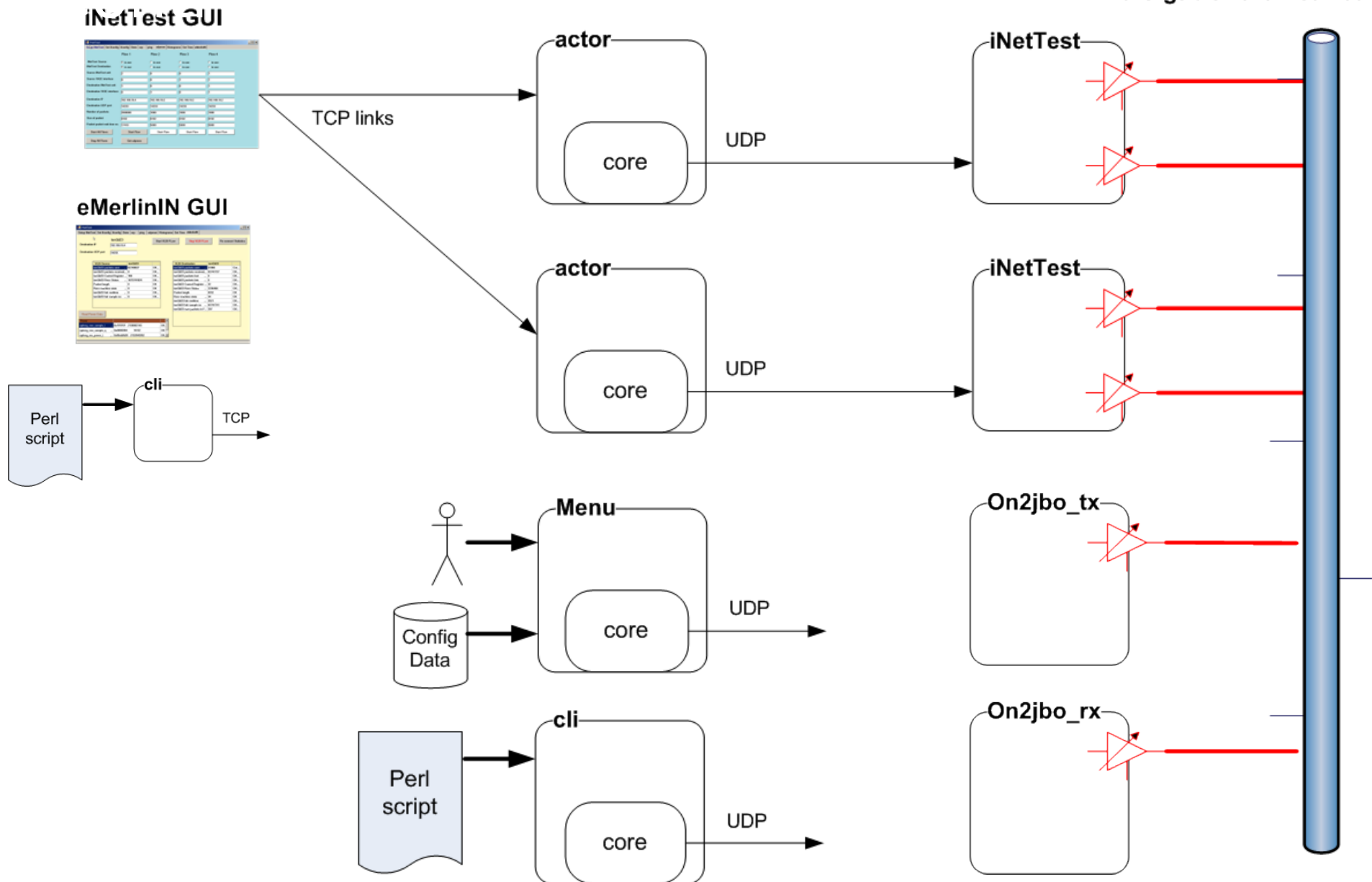


- iNetTest FPGA runs with a 200 MHz (5 ns) clocks
- Two iNetTest units can be controlled over ethernet IP from a PC
- iNetTest Ethernet IP address and MAC preset in code
 - last digit selected by jumper
- Ten GE IP address, Port and Gateway can be user configured
- Automatic or manual ARPing. ARP tables for each port can be examined
- iNetTest responds to and generates PING
- UDP packets can be sent by firmware between two iNetTests or between iNetTest and PCs
- User selects the number of packets to send, their length, and the time between them
- iBOBs count received and transmitted packets, and store up to 2048 time stamped events per port
- iBOBs generate histograms of the arrival time distribution for received packets

iNetTest & eMerlinIN: Architecture for Operation &



10 Gigabit Ethernet Network



iNetTest & eMerlinIN: GUI



VLBI Source

☒ in use

IP address of eMerlinIN Actor130.88.9.61

UDP port of eMerlinIN Actor14233

IP address of eMerlinIN Unit192.168.80.70

UDP port of eMerlinIN Unit14224

Connect to Actors

Update eMerlinIN

VLBI Source

VLBI Destination

MAC

IP address

Gateway

UDP port

pkts rcv

pkts rcv arp

pkts rcv ip

pkts rcv icmp

pkts rcv udp

pkts rcv my udp

pkts rcv dataack udp

pkts rcv ip unknown

VLBI Source

Clear Stats on Source

	tenGbE0	
tenGbE0 packets sent ...	74640839	OK...
tenGbE0 packets received...	0	OK...
tenGbE0 Control Register ...	160	OK...
tenGbE0 Recv Status ...	1073741824	OK...
Packet length	0	OK
Recv machine state ...	0	OK
tenGbE0 hdr realtime ...	0	OK...
tenGbE0 hdr sample no ...	0	OK...

Reset State Machine

VLBI Destination

Clear Stats on Destination

	tenGbE0	
tenGbE0 packets sent ...	51966	Cor...
tenGbE0 packets received...	74781389	OK...
tenGbE0 packets lost ...	0	OK...
tenGbE0 packets late ...	0	OK...
tenGbE0 Control Register ...	32	OK...
tenGbE0 Recv Status ...	466984	OK...
Packet length	8232	OK
Recv machine state ...	7	OK
tenGbE0 hdr realtime ...	3714	OK...
tenGbE0 hdr sample no ...	74781394	OK...
tenGbE0 num packets in F...	99	OK...

Reset State Machine

VLBI Source

VLBI Destination

List arp Cache

List tenGbE0 arp Cache

List 100 Mbit arp Cache

Perform a set of Class C arps

Send arp Requests

Send arp Requests

ArpForm

Close arp list

Unit 1

tenGbE0

IP host	MAC
0	00:00:00:00:00:00
.1	00:00:00:00:00:00
.2	00:00:00:00:00:00
.3	00:00:00:00:00:00
.4	02:50:c2:72:00:02
.5	00:00:00:00:00:00
.6	00:00:00:00:00:00
.7	00:00:00:00:00:00

ping IP destination

192.168.10.4

size of ping packet

0

sequence number

1

identifier

14233

VLBI Source

ping tenGbE0

ping 100 Mbit

VLBI Destination

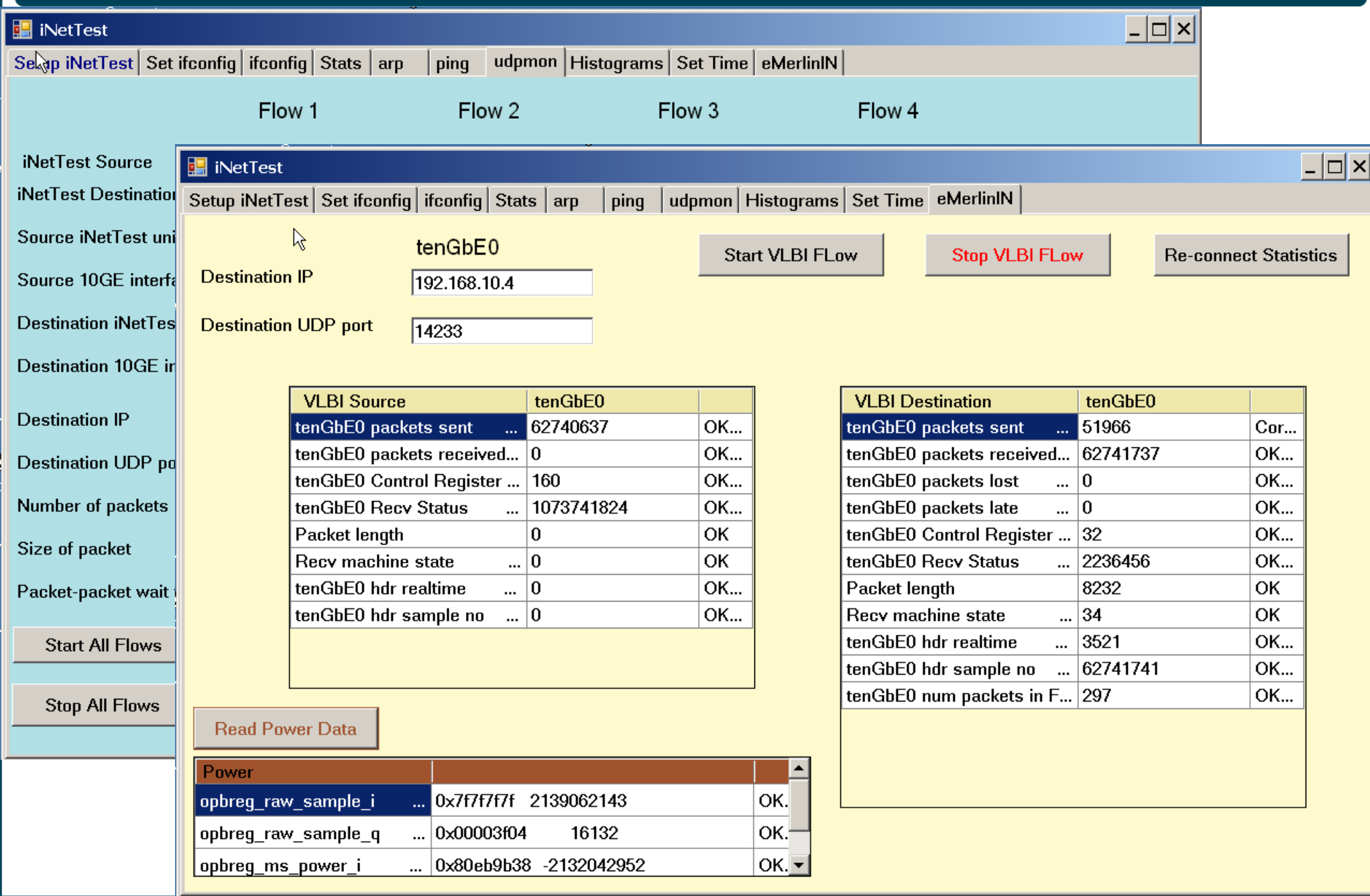
ping tenGbE0

ping 100 Mbit

ping icmp response for Unit 1 tenGbE0

Reply from 192.168.10.4: bytes=0 icmp_seq=1 time=21 ns

GÉANT



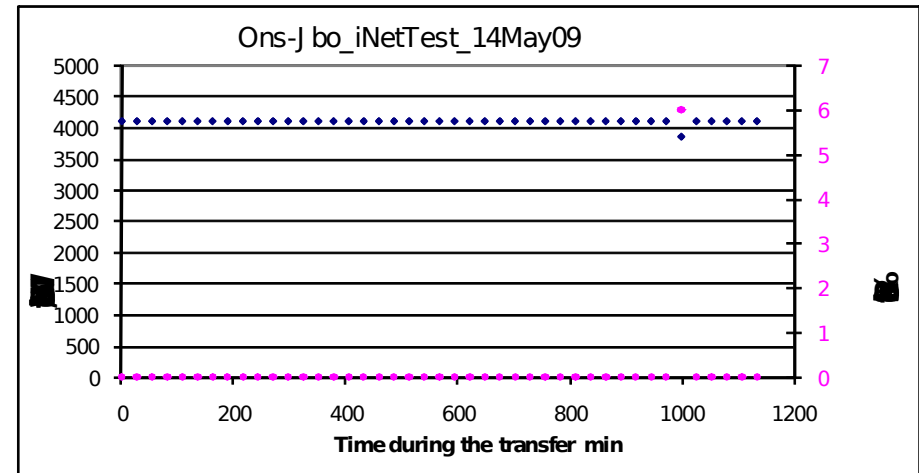
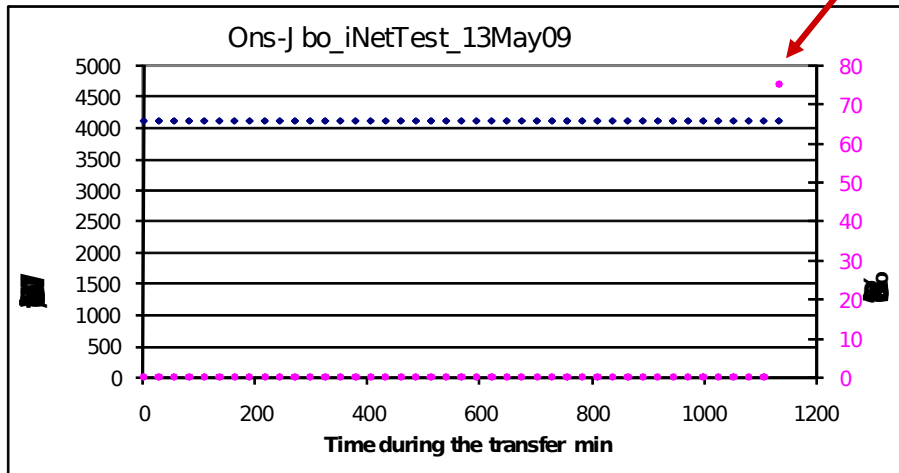
Testing Long Term Stability

UDP Throughput Onsala-Jodrell



- Use iNetTest to send trials of 100M 8192 Byte packets.
- Spacing 16 μ s, Trial takes ~27 min.
- Measure:
 - Throughput 4.094 Gbit/s
 - Packet loss Zero
 - jitter

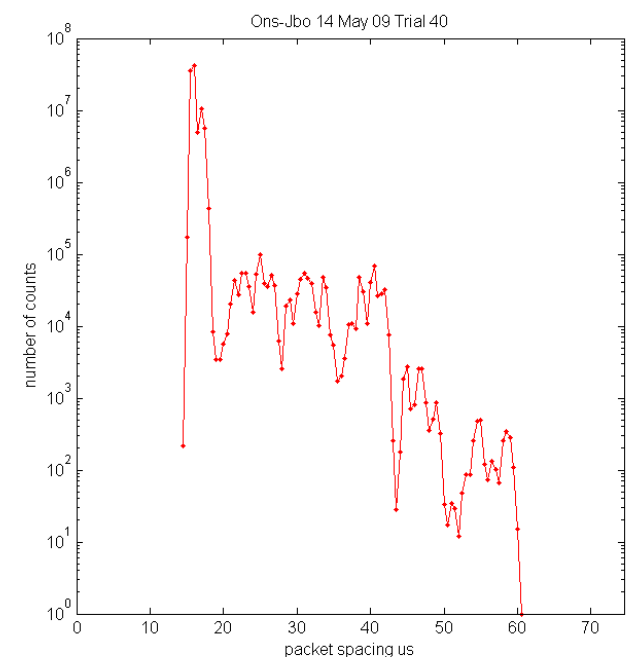
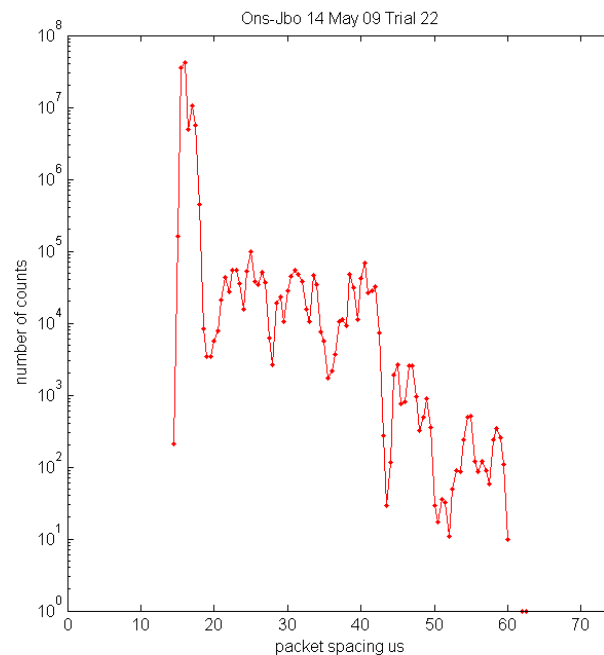
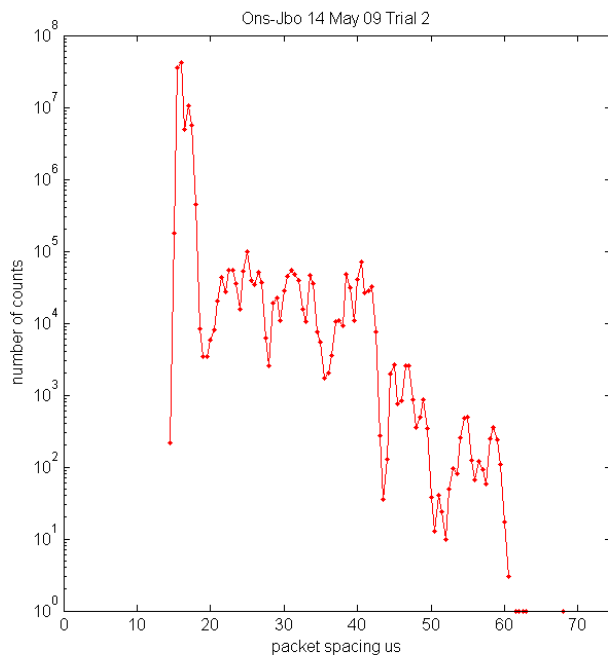
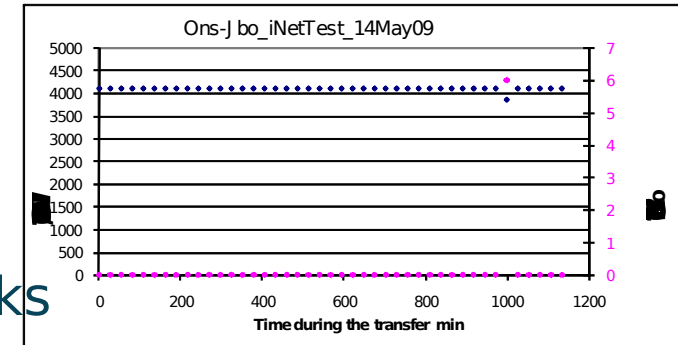
iBoB tenGbE0 locked up



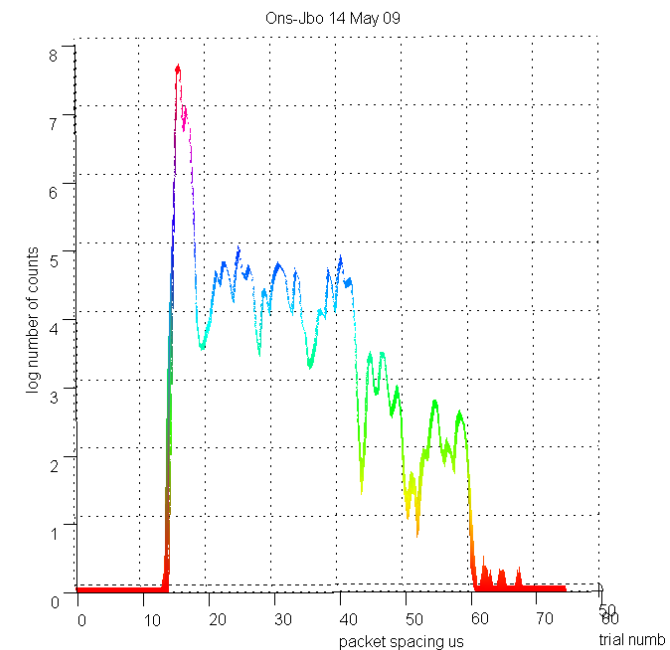
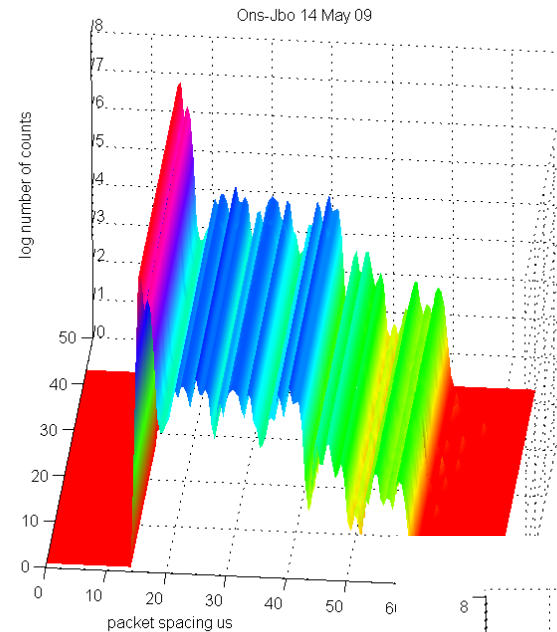
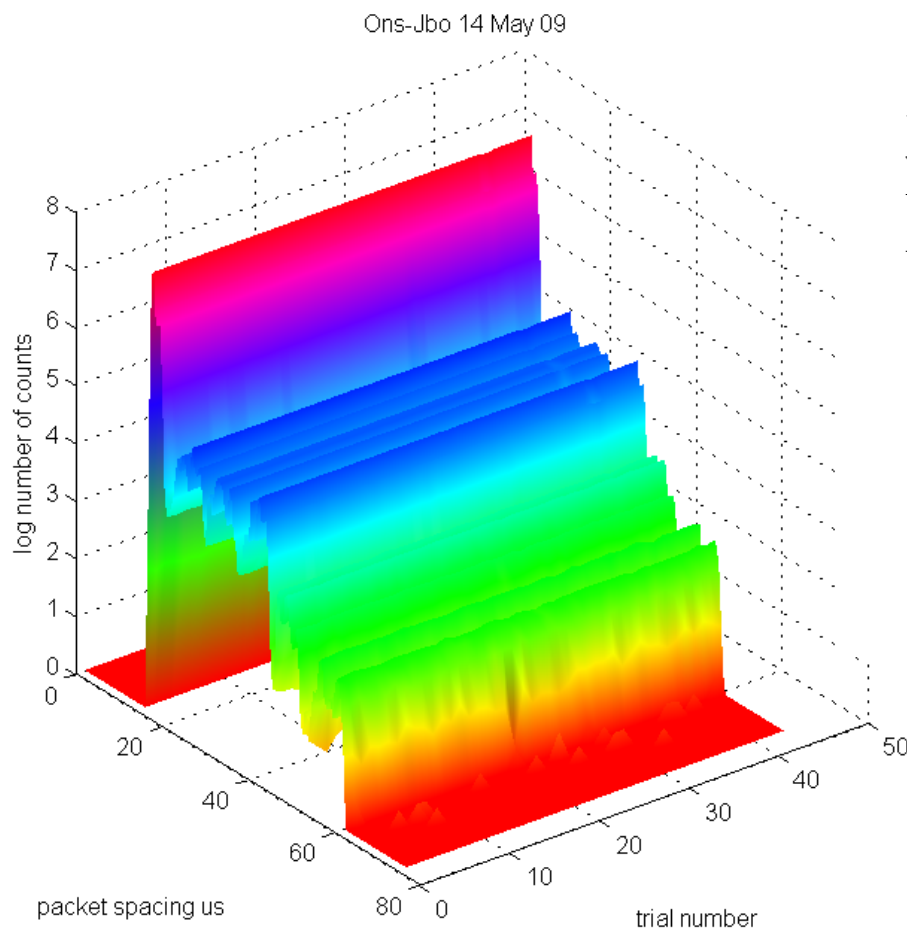
Testing Long Term Stability UDP Jitter Onsala-Jodrell



- Plot the inter-packet arrival times for 3 trials
- No noticeable differences
- Same distribution for tests over several weeks



Testing Long Term Stability UDP Jitter Onsala-Jodrell



- NO change in Jitter
- Tail $\sim 10^{-3}$ down
- Peak at 16 μs , tail extends to $\sim 70 \mu\text{s}$

What did we learn?

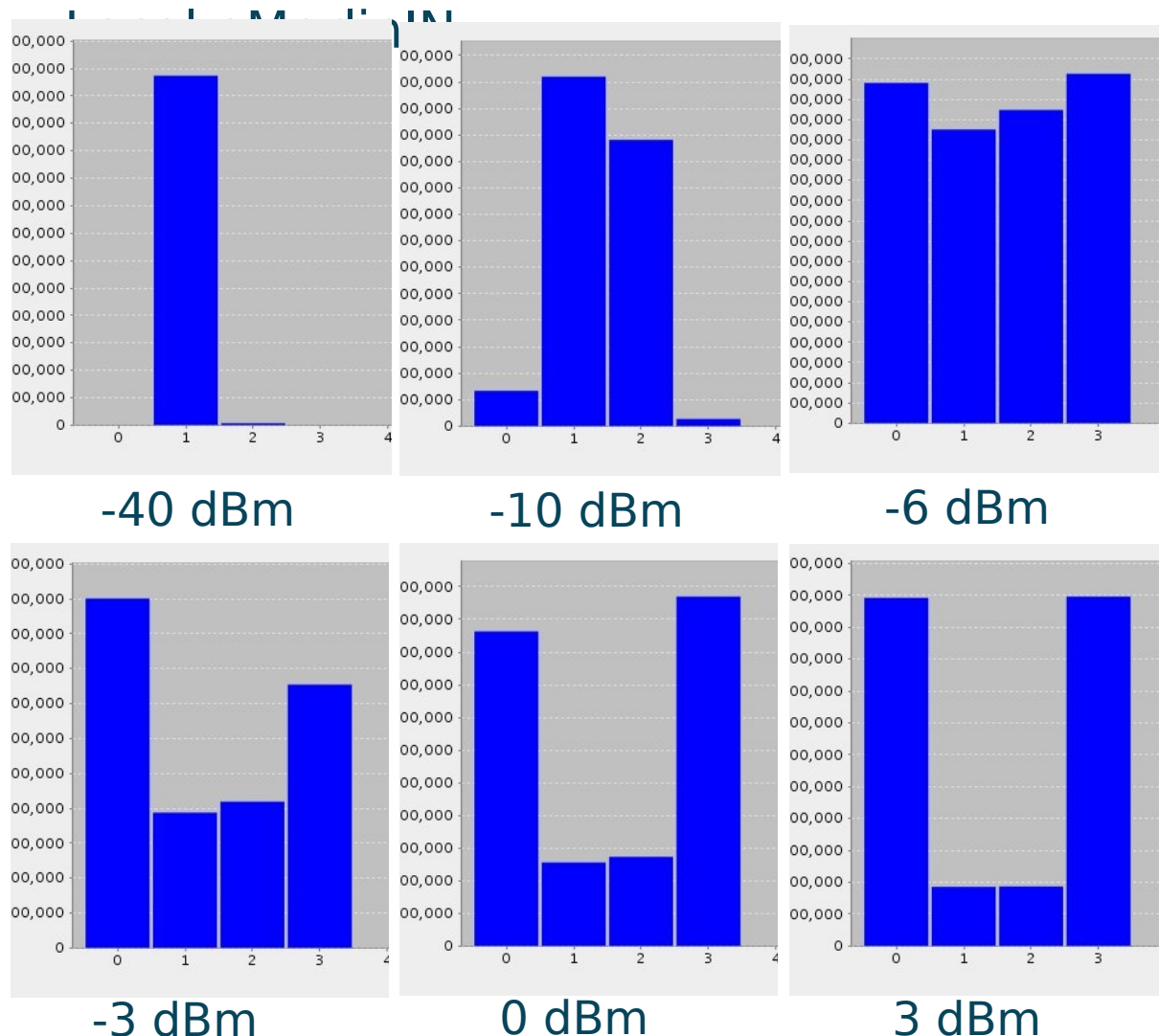


- The network is very stable
- iBOB state machine can lock the Ethernet transmit & receive (?)
- Added a reset function to the FPGA.
- Ethernet switches do age their MAC tables (as expected) SO:
 - they loose the port of the destination iBOB
 - flood packets at 4 Gbit/s to all ports in the VLAN
- iNetTest & on2jbo_rx send “data ACK” packets back to the source every n packet received.
- iNetTest and eMerlinIN operate as required.
- Signals from signal generator in Onsala have been sampled and transmitted to the Station board in the

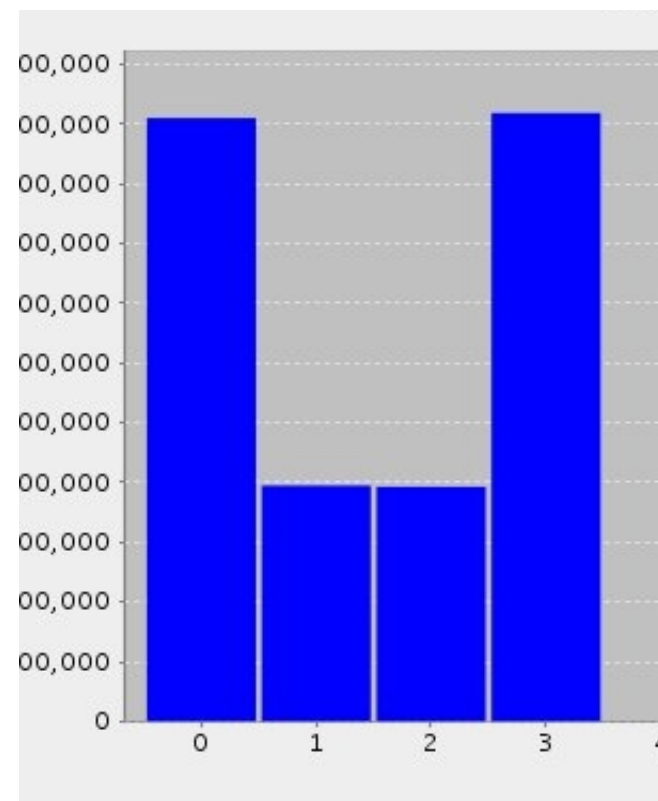
State histograms from the e-Merlin Station Board



- 88 MHz Signal, 1024 MHz clock, 2 bits, 1 pol.



- 88 MHz Signal, 1024 MHz clock, 2 bits, 1 pol.
eMerlinIN from Onsala



ANY QUESTIONS ?