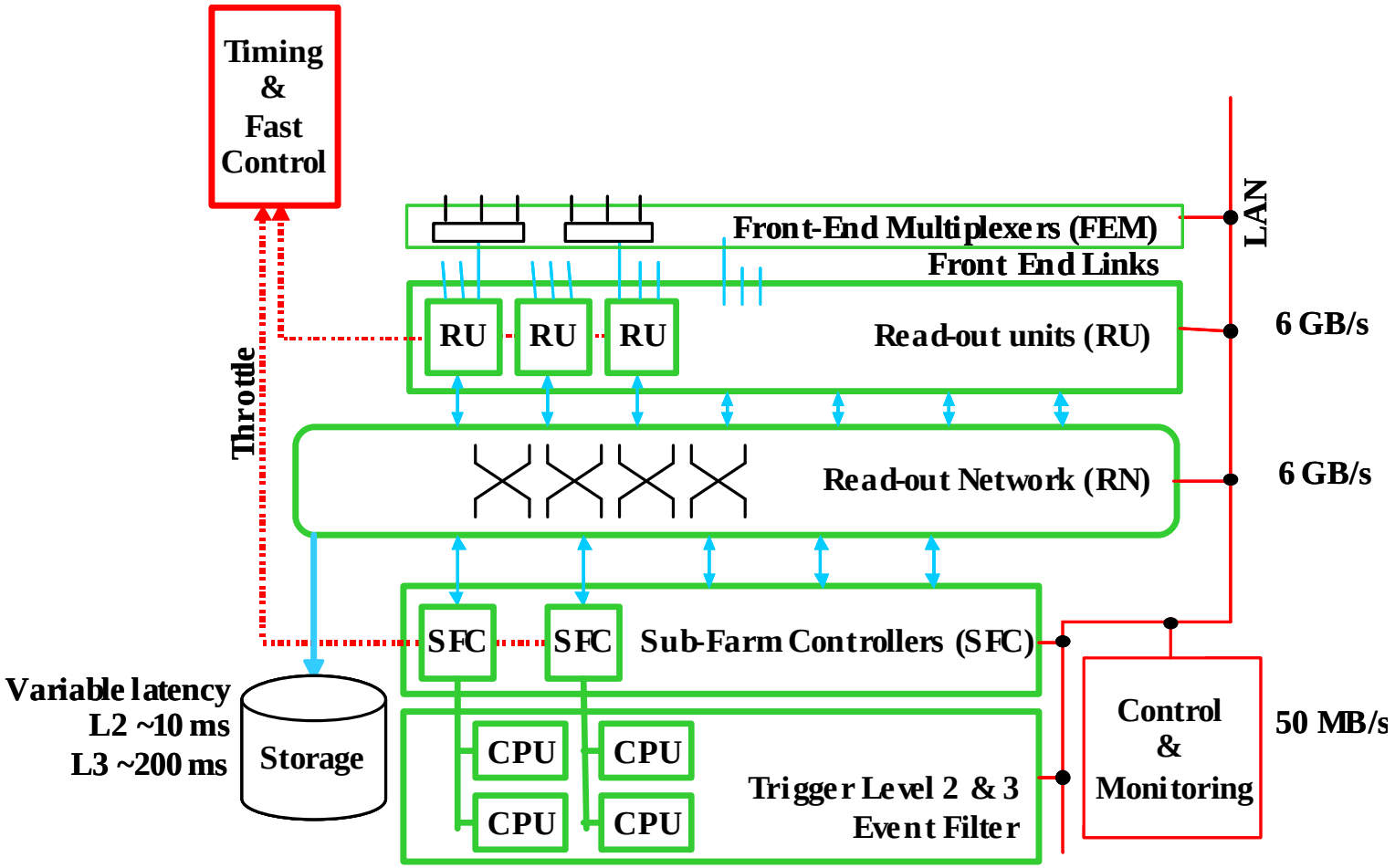
Event Building With Smart NICs

Jean-Pierre Dufey, Beat Jost,
Niko Neufeld & Marianna Zuin

DAQ 2000
Lyon, October 20, 2000

Recap: LHCb DAQ System



Event Building Components

- Readout units (RU): multiplexing of front-end links, destination assignment
- Switching read-out network
- Sub-farm controllers (SFC): event building and event dispatching

Event Building Properties

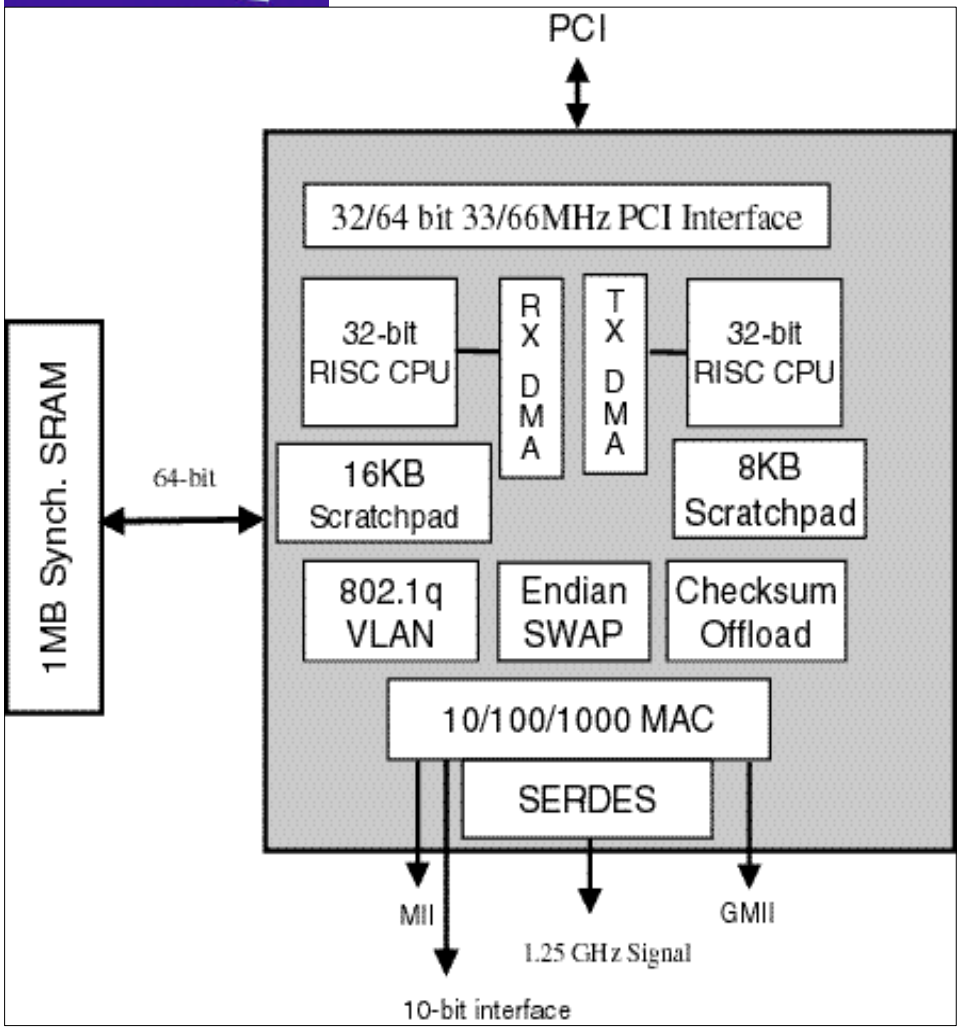
- Static load balancing among the SFCs
  - RUs send round robin to destinations

    destination = f(event_number)
    ***f* being the same for all RUs**
- Pure push protocol
  - congestions handled via flow control and more importantly by throttling
- Distributes the event data flow of 6 GB/s from *m* sources to *n* destinations, each of which has to handle O(1Kb) fragments at 80 kHz

Why Use Smart NICs?

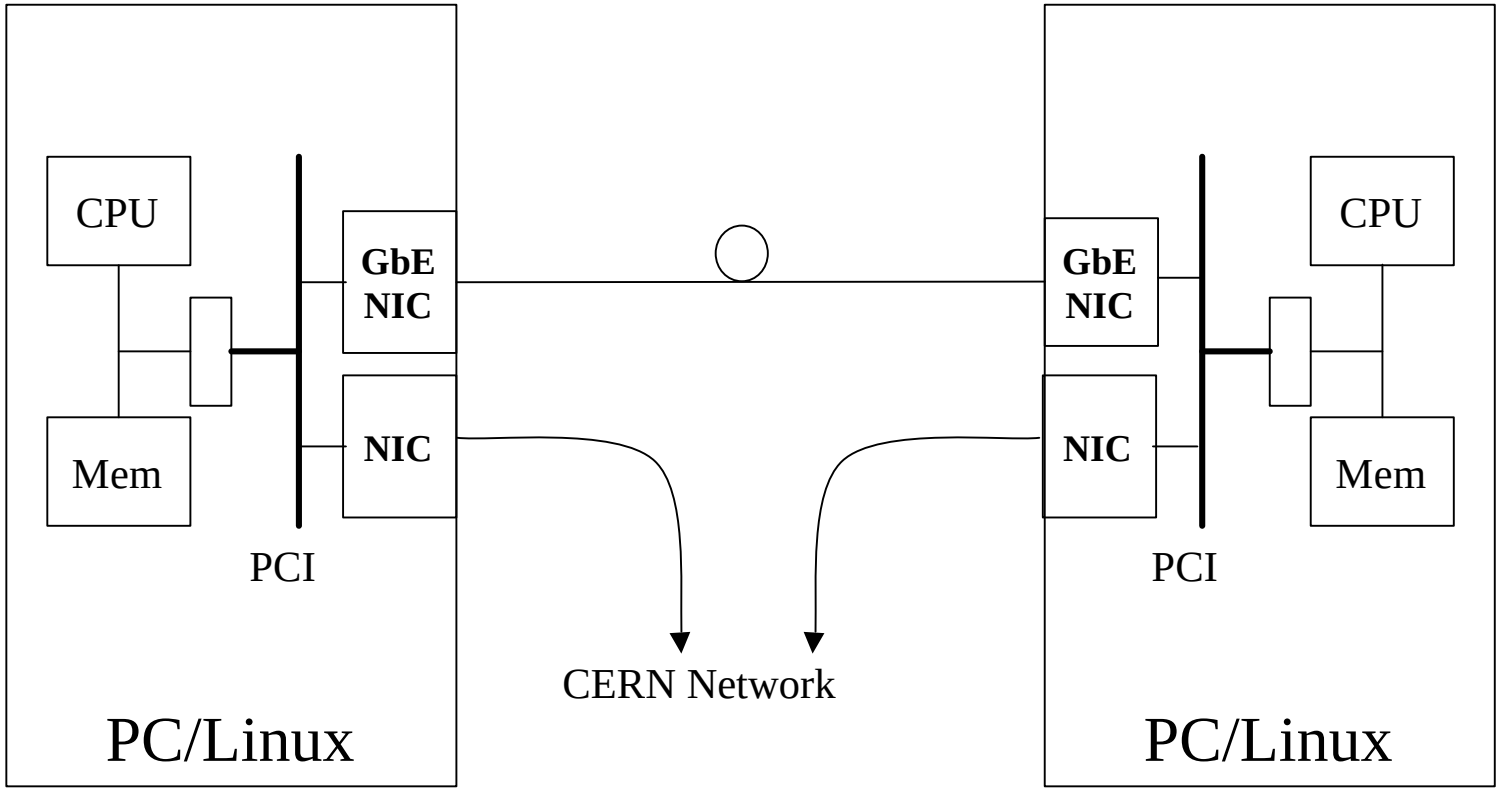
- ☺ Modern Smart NICs are powerful embedded computers
- ☺ Off-load general purpose CPU
- ☺ Take advantage of cheap CPU power on the NIC
- ☺ Facilitate hardware design of the RU
- 😐 (Yet) limited CPU power compared to commodity PC
- 😐 No guarantee that high-end NIC development will continue in this direction (firmware/CPU vs. ASIC/FPGA)

Alteon Tigon 2

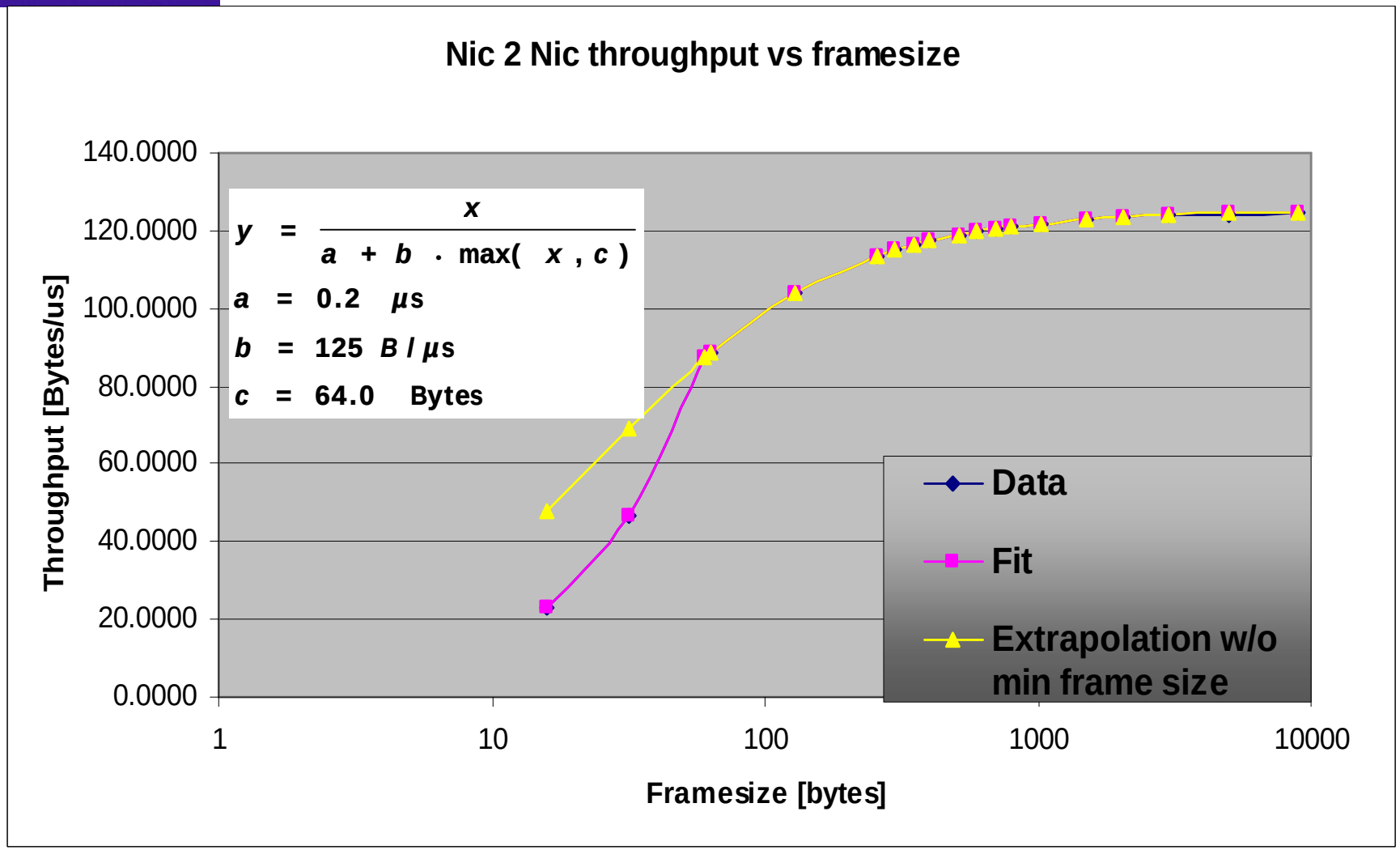


- Features
  - Dual R4000-class processor running at 88 MHz
  - Up to 2 MB memory
  - GigE MAC+link-level interface
  - PCI interface
- Development environment
  - GNU C cross compiler with few special features to support the hardware
  - Source-level remote debugger

Test Setup


CPU
GbE NIC
NIC
Mem
PCI
PC/Linux
CERN Network
GbE NIC
NIC
CPU
Mem
PCI
PC/Linux

NIC 2 NIC Performance



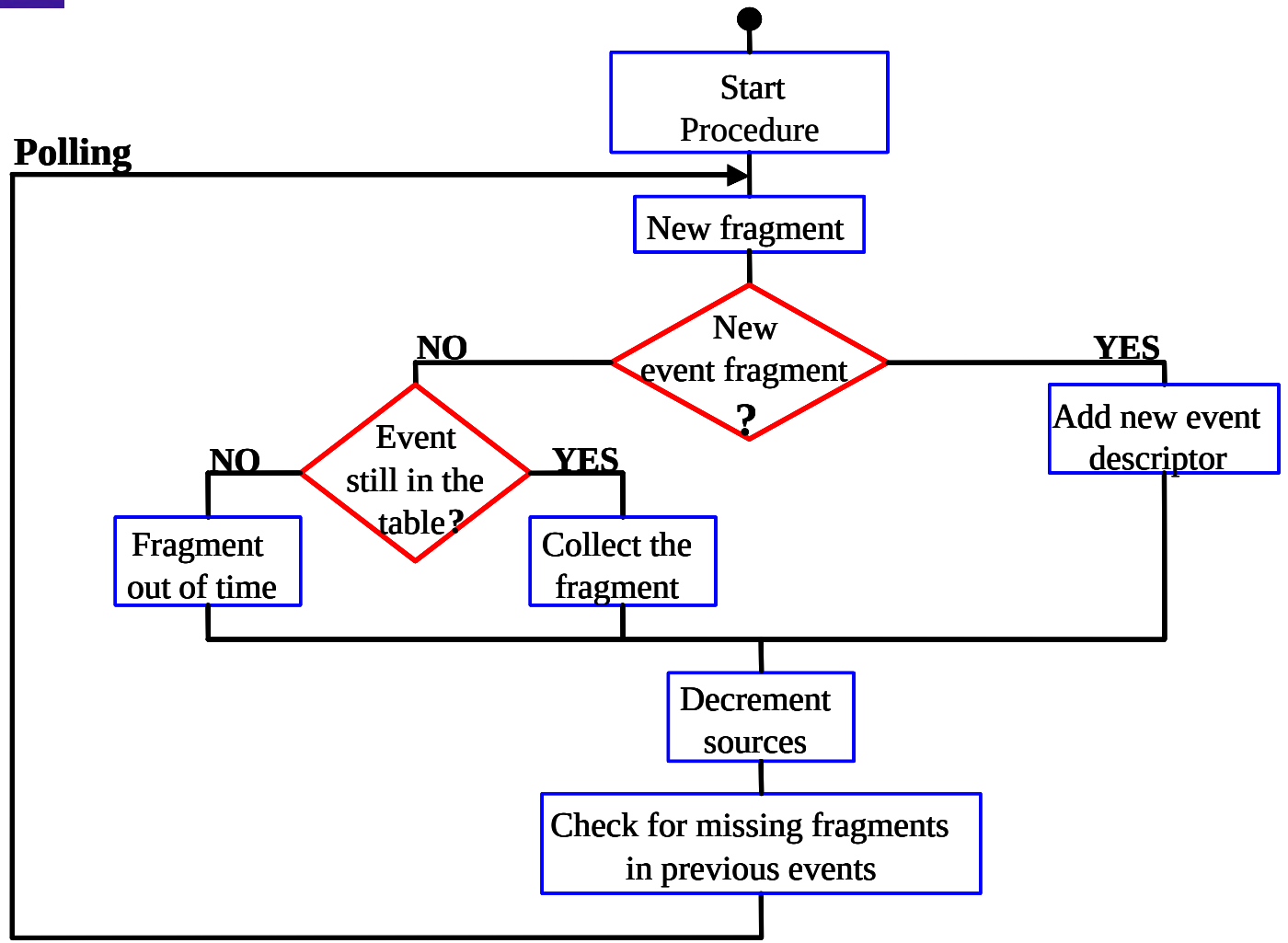
Performance of Alteon NIC

- Can fill the wire at any given frame size (from 64 to 9000 bytes)
- Can send out frames at a frequencies of up to 1.4 MHz
- For frames bigger than 512 bytes more than 95% of nominal bandwidth available for data (practically 100% for >8000 Jumbo frames)

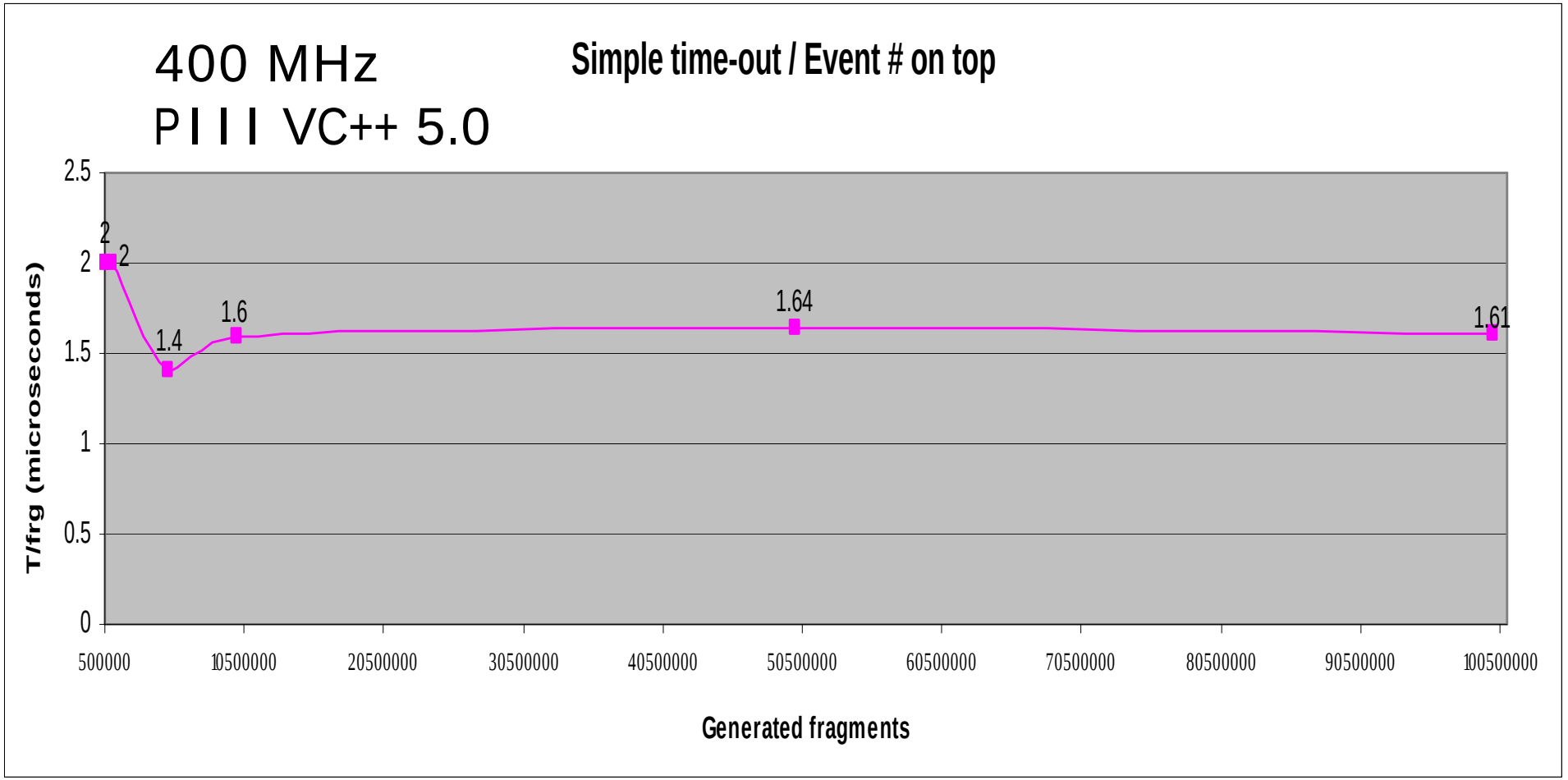
Event Building Algorithm

- Assembles events out of fragments from a *known* number of sources
- Handles an adjustable amount of events concurrently (limited only by buffer space)
- Implements "Implicit + Time-out Completion"
- Uses "scatter/gather" capabilities of NIC's DMA engine to concatenate the fragments into the host's memory

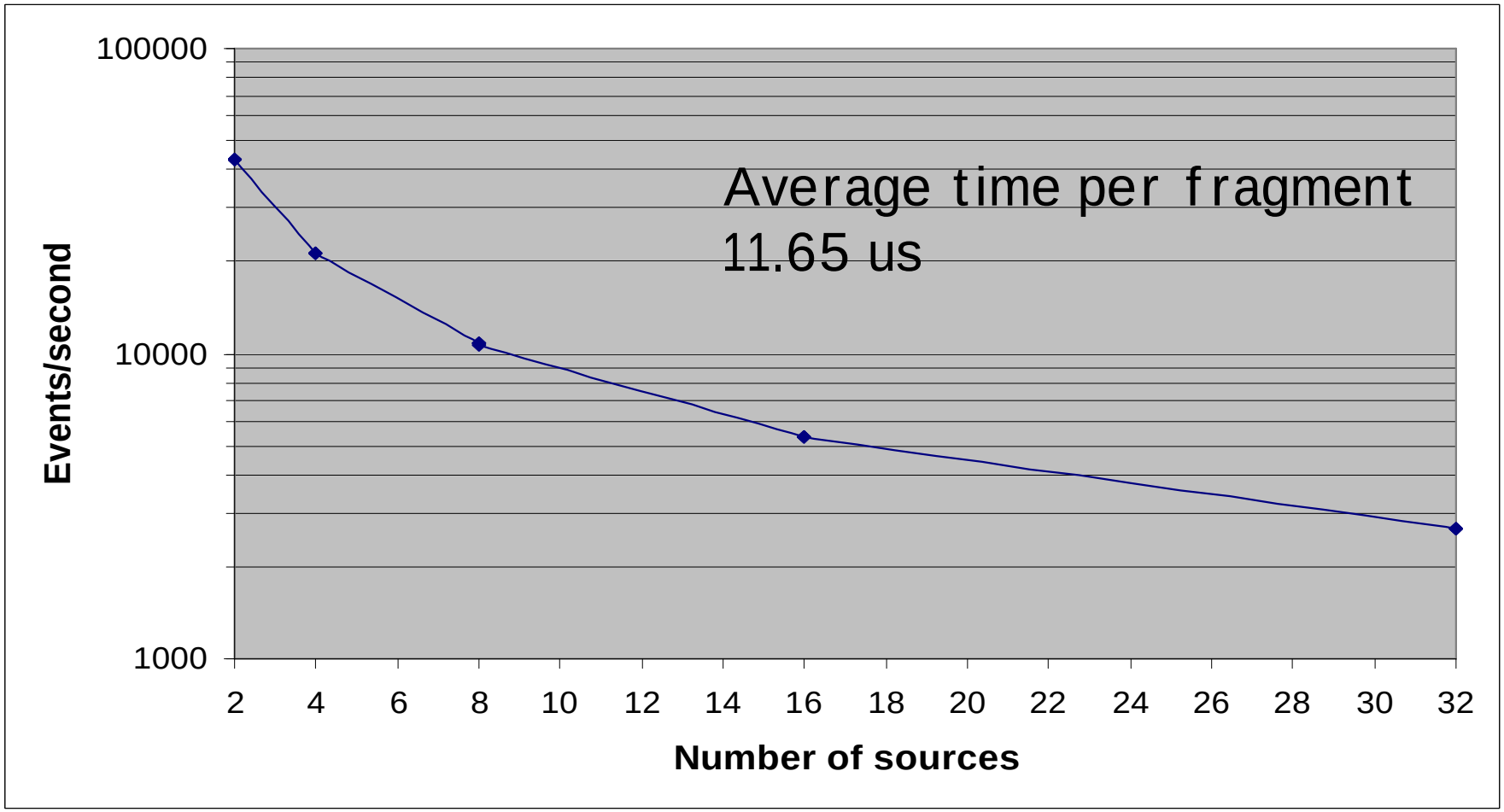
Algorithm


Start Procedure
Polling
New fragment
New event fragment ?
NO
YES
Add new event descriptor
Event still in the table?
NO
YES
Fragment out of time
Collect the fragment
Decrement sources
Check for missing fragments in previous events

PC Test Implementation



Performance NIC 2 NIC



Summary

- Event building on a smart NIC at a frequency of incoming fragments of almost 100 KHz has been demonstrated
- Event building at Gigabit speed for fragments bigger than ~1100 bytes
- Code Optimization ongoing (9 us/frag have already been achieved)

Program of Work

- Evaluate impact of interrupt coalescence on SFC performance
- Study possibility of handling some amount of TCP/IP traffic on the outgoing link of the SFC (events to storage)
- “Real world” tests on a Gigabit Ethernet switching network
- Use measured parameters in a detailed simulation of the readout network