STUDIES OF ATM FOR ATLAS HIGH LEVEL TRIGGERS

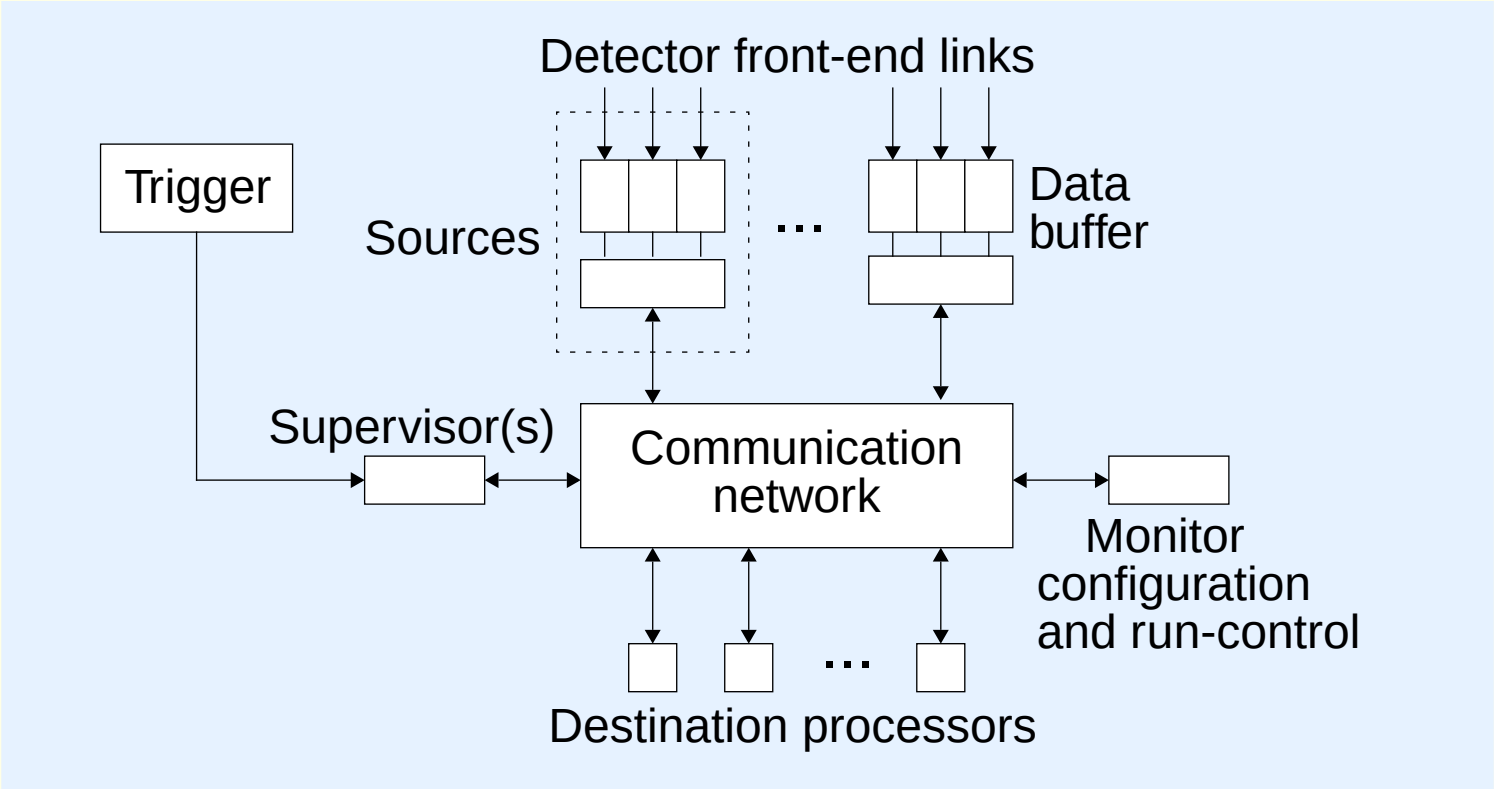
J. Bystricky, D.Calvet, M. Huet, P. Le Dû, I. Mandjavidze

CEA Saclay, 91191 Gif-sur-Yvette CEDEX, France

calvet@hep.saclay.cea.fr

OUTLINE

SCHEMATIC VIEW OF AN EVENT BUILDER



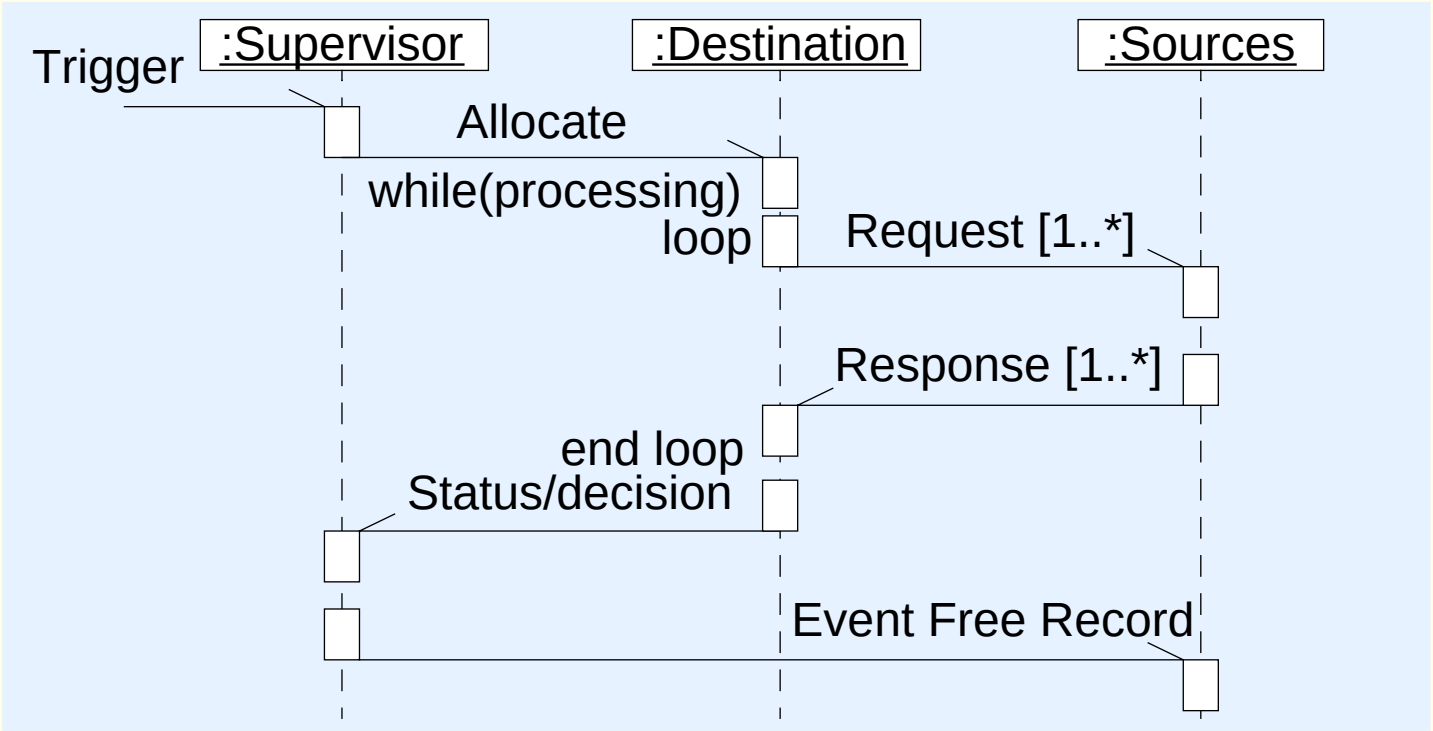
Supervisor(s): distribute events delivered by Trigger to destination processors

Processors: collect event data from relevant sources to assemble then process event

Sources: receive detector data, buffer it as long as needed, provide it to destinations

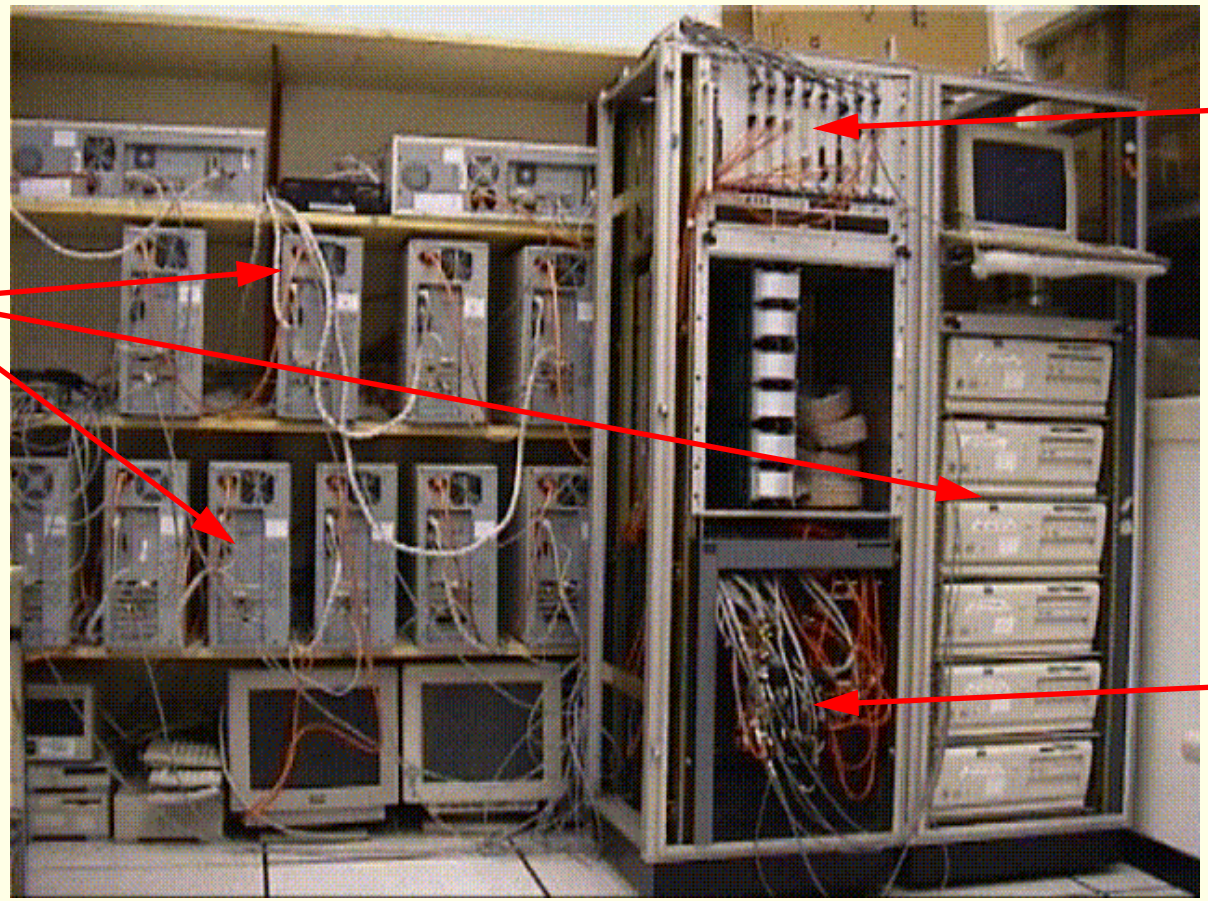
Communication network: transports protocol messages, blocks of data,...

A GENERIC PROTOCOL FOR EVENT BUILDING



- **Supports all types and combinations of event building:**
  - **full event building (all sources involved)**
  - **static/dynamic partial event building (fixed or variable subset of sources)**
  - **single step or multi-step event building (rejection can occur after each step)**
- **Pipe-lining several events per destination hides latency of data delivery**
  -> multi-threading / tasking in destinations
- **Flexible scheme equally suited for Trigger and DAQ systems; e.g. ATLAS HLT:**
  multi-step dynamic/static partial e.b. + single step static partial or full e.b.

DEMONSTRATOR SYSTEM (HARDWARE)



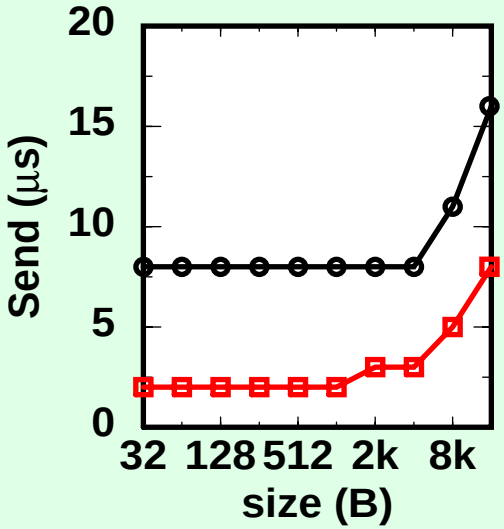
- 3 Fast Ethernet switches, 32 ports 100 Mbit/s or 24 ports 100 Mbit/s + 1 Gbit/s uplink
- PCs connected to ATM and Fast (Gigabit) Ethernet – VME cards only to ATM
- all machines also connected to CERN standard TCP/IP network

- Common effort of ATLAS LVL2 Trigger groups to build a large demonstrator at CERN

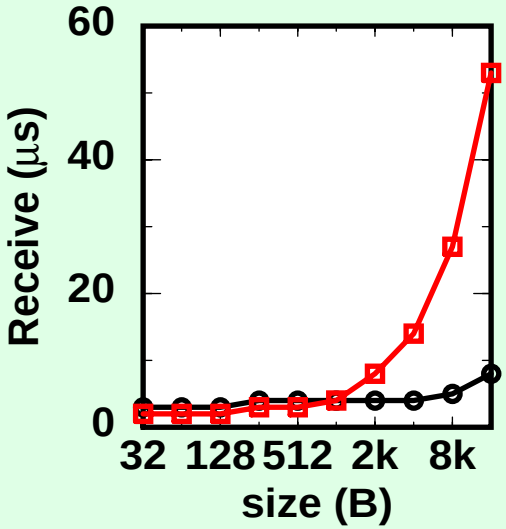
DEMONSTRATOR SYSTEM (SOFTWARE)

- **Custom made zero-copy driver/library for ATM and Fast Ethernet:**
  scatter/gather DMA, asynchronous send, blocking / non-blocking receive

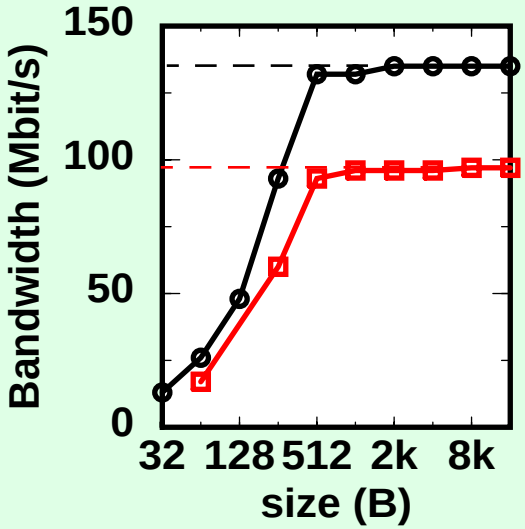
Message send



Message receive



User bandwidth



Message fill: **75 MB/s**

Pentium III 733 Mhz

Fast Ethernet / Linux

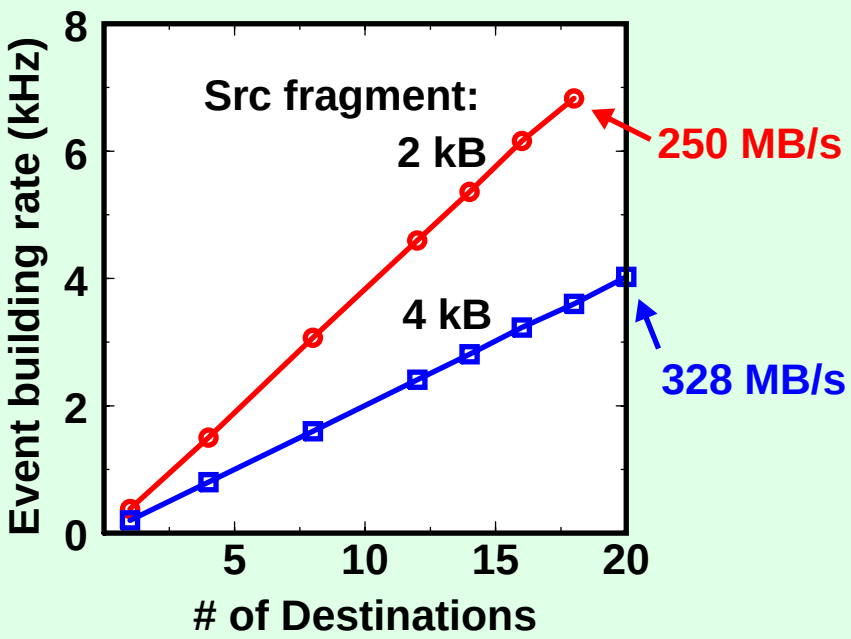
ATM / Windows2000

- **Our application software embodies the proposed event building protocol**
  Results on partial / full and single-step / multi-step event building:
  "An Integrated System for ATLAS High Level Triggers: Concept, General Conclusions on Architecture Studies, Final Results of Prototyping with ATM"
  ATLAS DAQ Note 2000-011 http://weblib.cern.ch

- **Testbed also run ATLAS LVL2 Trigger Pilot Project Reference Software**
  http://atlas.web.cern.ch/Atlas/project/LVL2testbed/www/

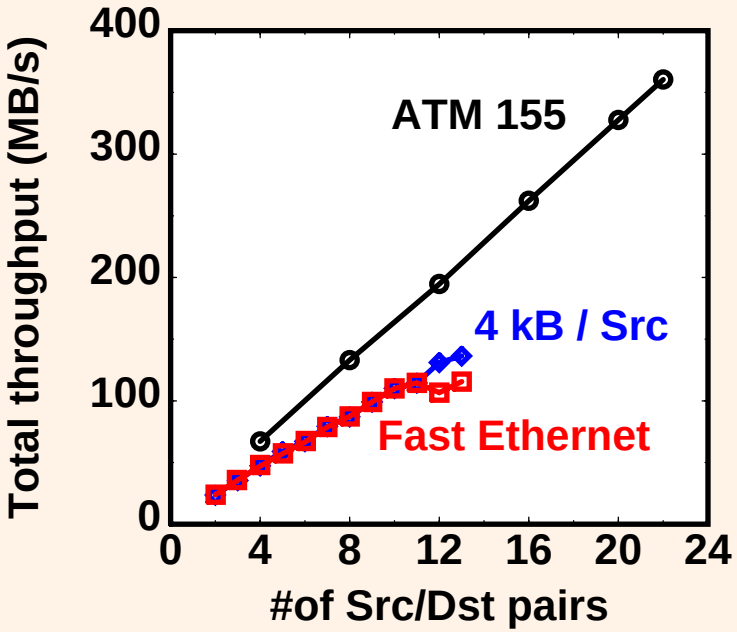
PERFORMANCE FOR FULL EVENT BUILDING

Throughput scaling when adding destinations



ATM event builder with 20 sources

Scalability of N x N event builder system



Fragment size per source: 8 kB

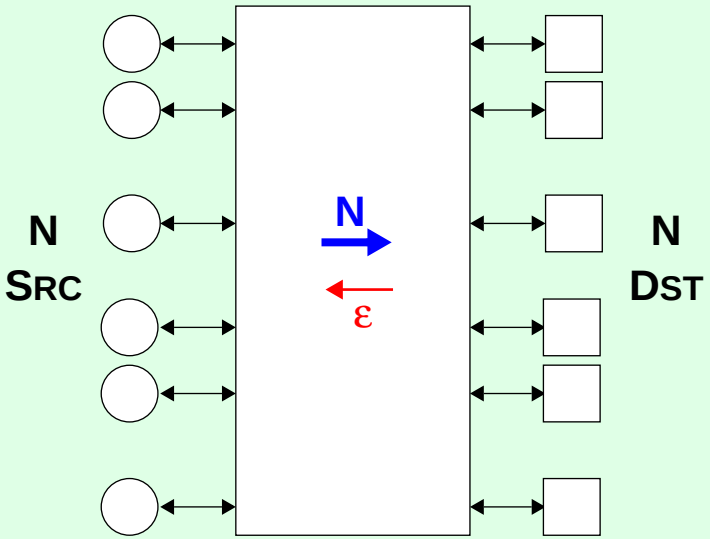
- **Use CBR channels for correct operation with ATM**
- **Enable PAUSE Flow Control for correct operation with Fast Ethernet**
- **Both ATM and Fast Ethernet offer performance close to theoretical limits and scalable**

INTERCONNECTING SWITCHES – PRINCIPLE

Large switches may not be economical / available / practical -> cascade switches

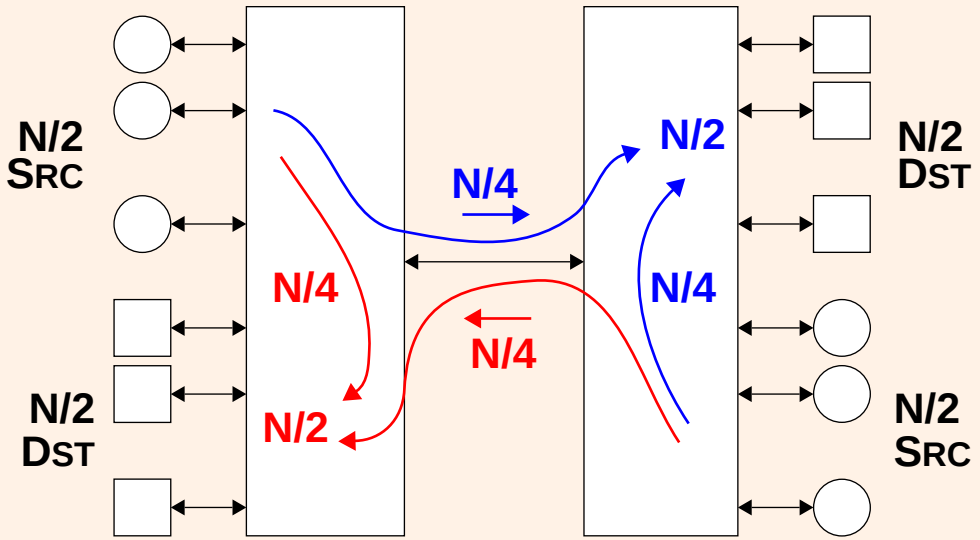
Asymmetric data flow
Symmetric bi-directional network links
Interleave sources and destinations

allow inter-switch port count reduction for a N x N event builder



1 switch with 2N ports

If such switch does not exist?



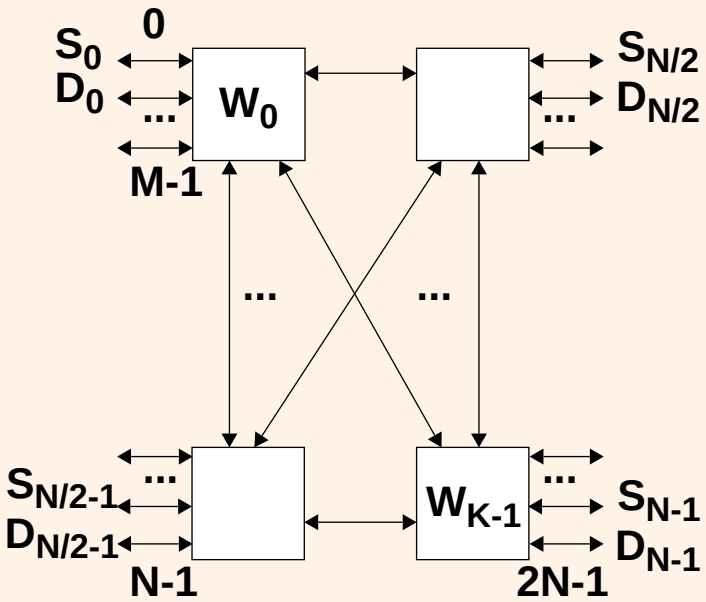
2 switches with 1.25N ports

25% more ports than single switch but smaller switches are used

- **Characteristics of event building traffic allow savings in switch cascading**

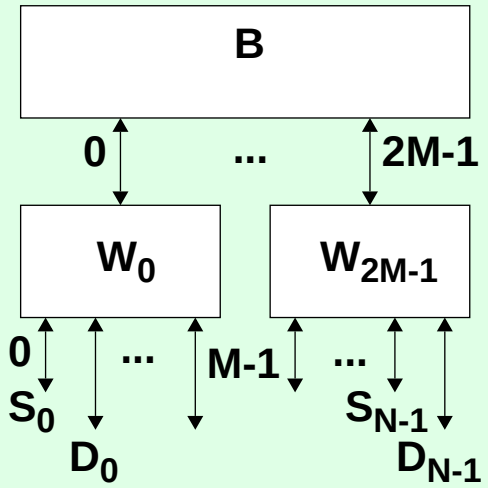
TOPOLOGIES OF INTERCONNECTED SWITCHES

Star network of K switches



Total port count = $2N\left(\frac{3}{2}-\frac{1}{2.K}\right)$

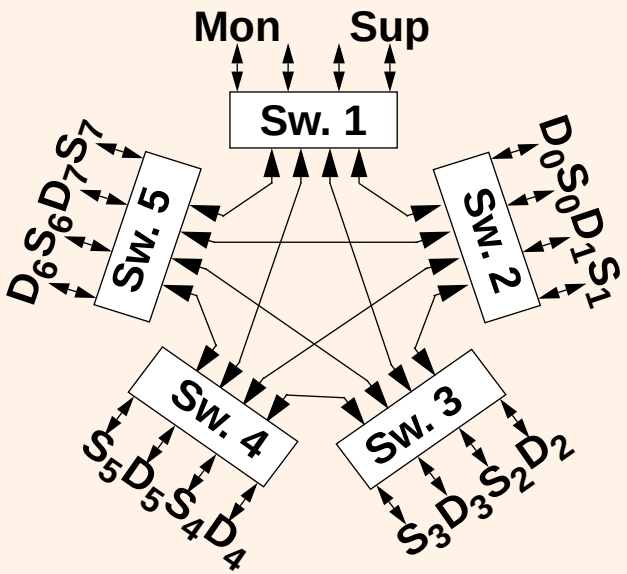
Backbone + K access switches



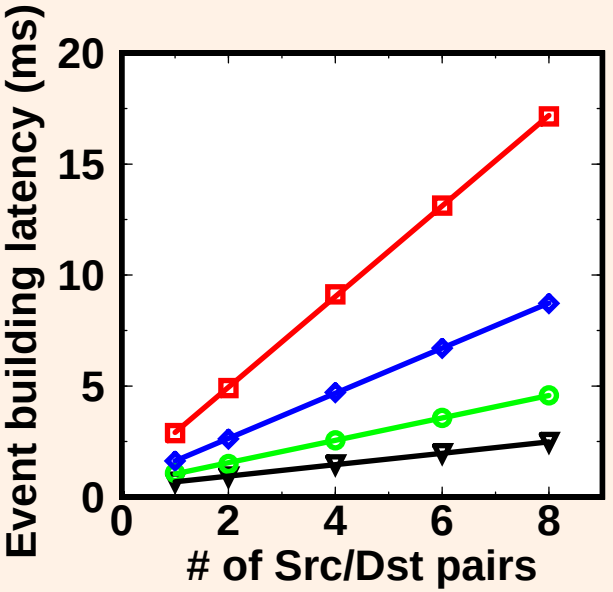
Total port count = $2N\left(2-\frac{1}{K}\right)$

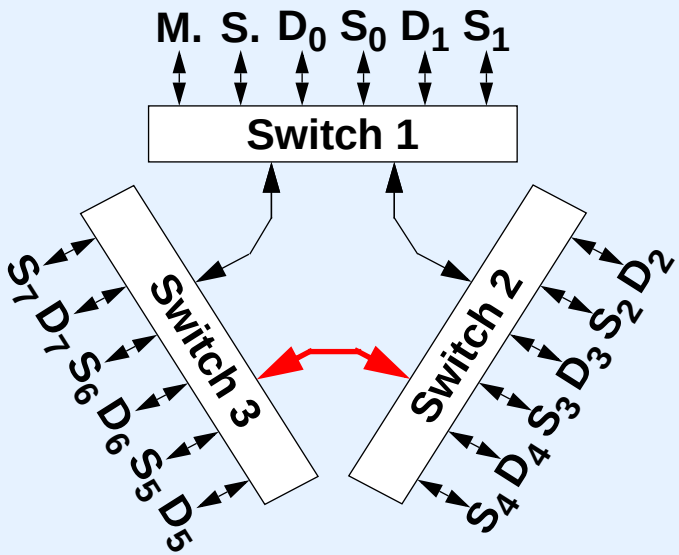
- Backbone + access switches is a single path network; star is multi-path
- Star network always needs smaller port count than backbone + access
- Port count for NxN event builder: Single switch: 2N – Star: 2.5N to 3N – Backbone: 3N to 4N

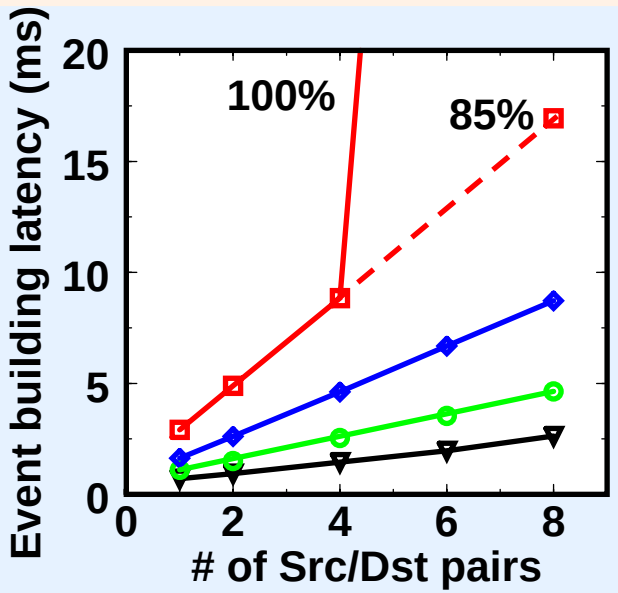
TESTS WITH INTERCONNECTED SWITCHES (ATM)



Fragment:
- 32 kB
- 16 kB
- 8 kB
- 4 kB







- **Linear scaling of performance**
- **As expected for 3-branch star, limit is fixed by link between switch 2 and 3**

BUILDING LARGE NETWORKS

Topology	Hardware switch x ports	Total ports	Usable ports (% of total)	System size	Max. theoretical throughput (GB/s)
Star	25 x 64 ATM 155 Mbit/s	1600	63	500 x 500	8.4
	8 x 224 ATM 155 Mbit/s	1792	69	615 x 615	10.4
Core + access	1 x 64 Gigabit Ethernet 64 x (24 Fast Ethernet + 1 Gigabit Ethernet uplink)	2816[a]	57[a]	768 x 768	7.7

a.1 Gigabit Ethernet port = 10 Fast Ethernet ports.

- **In theory event builders with > 1000 nodes and > 5 GB/s usable throughput can be built today**
- **Effective deployment, operation, management: engineering studies needed**

ATM VERSUS FAST/GIGABIT ETHERNET

Item	ATM	Fast/Gigabit Ethernet
Network configuration	PVC setup not easy	self learning switch, but multi-cast tree setup might still be needed
Qos features for event building	uses Constant Bit Rate channels scalable to any system size in theory	rely on switch buffers and flow control scalability needs to be studied
Multi-switch	using PVCs: - allows multi-path networks - eases link aggregation	multi-path networks possible? link aggregation: 1 Gigabit link = 10 Fast Ethernet
Price (switch port + NIC) Market	~1500 $ for 155 Mbit/s shrinking	<300 $ for 100 Mbit/s <1500 $ for 1 Gbit/s Healthy
Perspective	stay at 155 Mbit/s for NICs, 662 Mbit/s 2.5 Gbit/s for WAN trunks	1 Gbit/s becoming common 10 Giga to appear

- **ATM adequate but fading technology in the LAN**
- **Ethernet: huge sale volume, very low price, good perspective of evolution**

SUMMARY AND CONCLUSIONS

Event Building Protocol

Supports any combination of partial / full / single-step / multi-step event building

Event Builder Demonstrator

About 50 machines – ATM and Fast Ethernet networks
Custom made zero-copy driver
Generic software implements proposed event building protocol

Results related to full event building

Operation close to wire speed for both ATM and Fast Ethernet
System up to 22 x 22 and 13 x 13 with ATM and Fast Ethernet respectively
Need CBR channels for ATM; need Flow control for Fast Ethernet

Switch cascading

Star topology allows to reduce number of inter-switch links
Star with 20 usable ports tested with ATM
Systems over 500 x 500 seem feasible with today's components

Pending studies

Gigabit Ethernet not investigated in this work
Behavior of large systems with Ethernet type of flow control
Larger multi-switch configurations, link aggregation