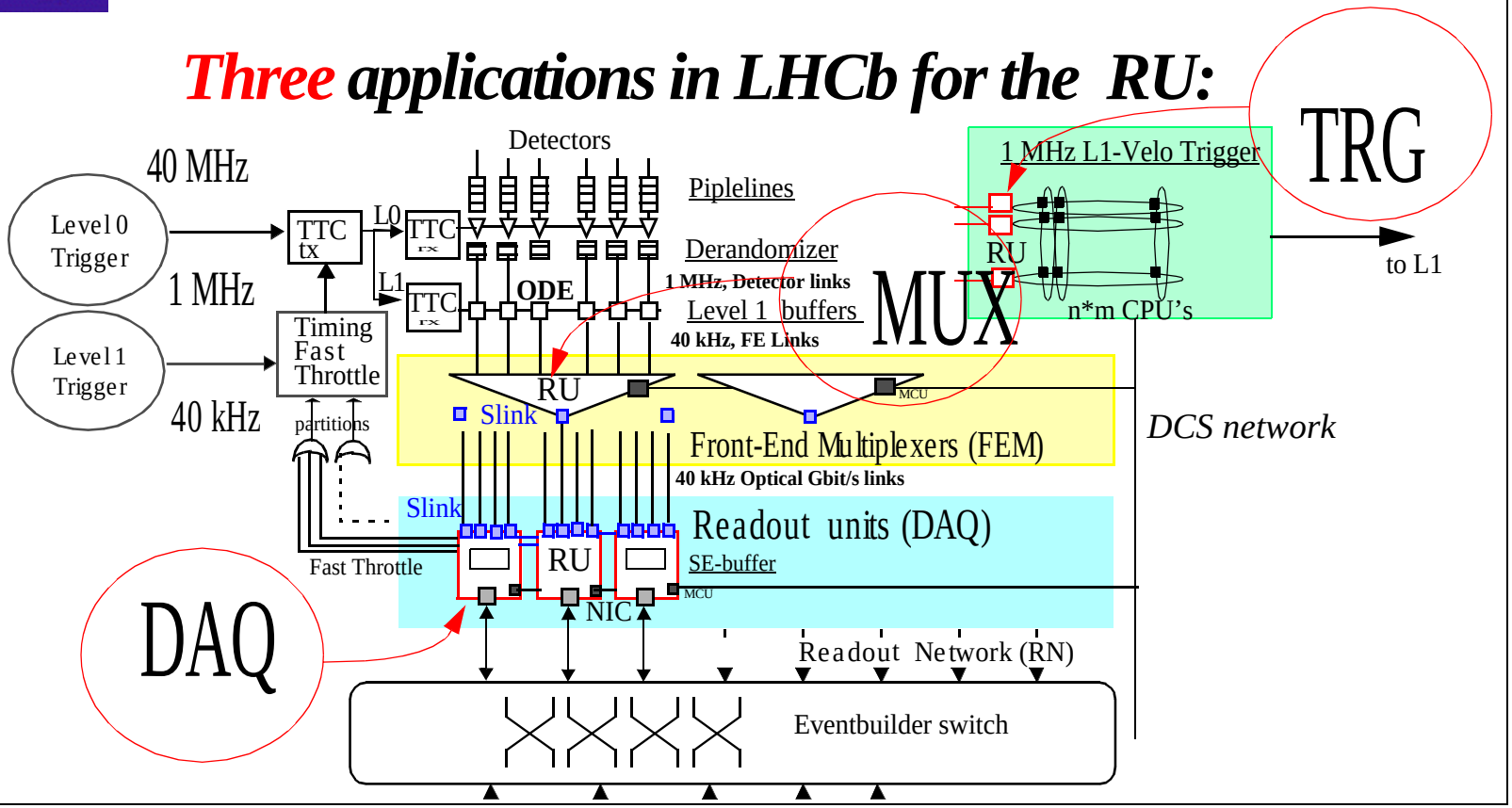
# Readout Unit-II

## *Three applications in LHCb for the RU:*




## RU' Scope

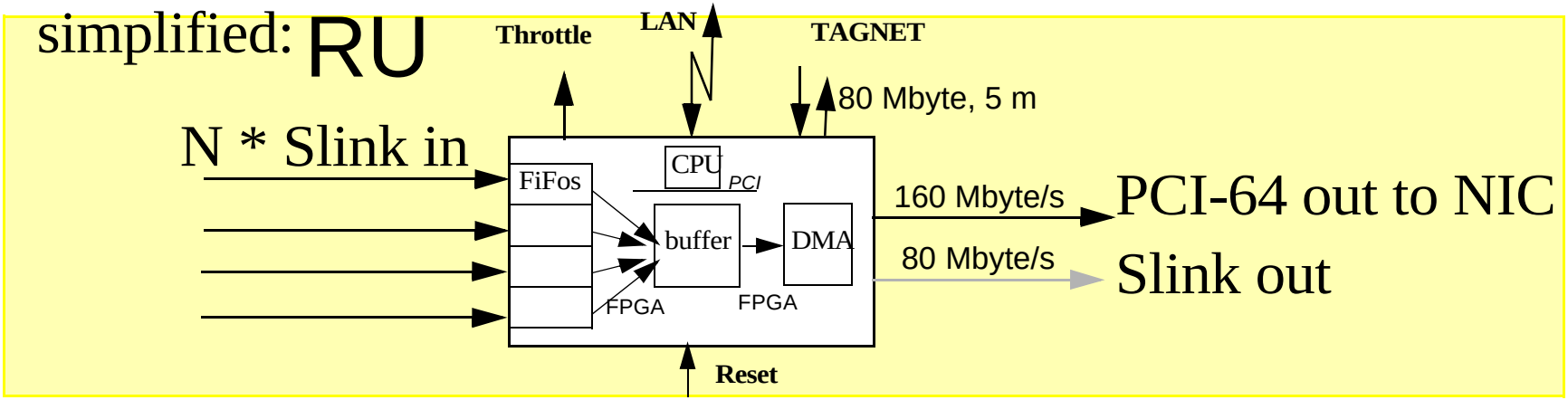
- DAQ/ Mux 40 kHz, VELO 1Mhz subevent building, **Transport Format (STF), links (Slink),** PCI bus, performance ( VELO, DAQ ), design&simulation (VHDL), crate environment/cost/connectivity, FPGA remote configuration, buffer management, L1 Traffic scheduling (Tagnet), NIC protocols ...

RU TEAM: *Jose Toledo, David Dominguez, Bertrand Bruder, Francois Bal, Angel Guiraro, Hans Muller*

# Ru Team / support for LHCb


- Curro: ( leaving + 2 weeks) designer RU-1 + RU -2, HW architectures and applications
- Bert: (leaving + 2weeks ) MCU card (PCI) and remote RU configuration software (Linux)
- David: (leaving + 2 month) programmable logic, VHDL , PCI
- Francois: Gbit links, Slink cards, Tagnet, Lvds
- Angel: (New) commissioning of VHDL applications for RU-II: Mux, VELO, DAQ .

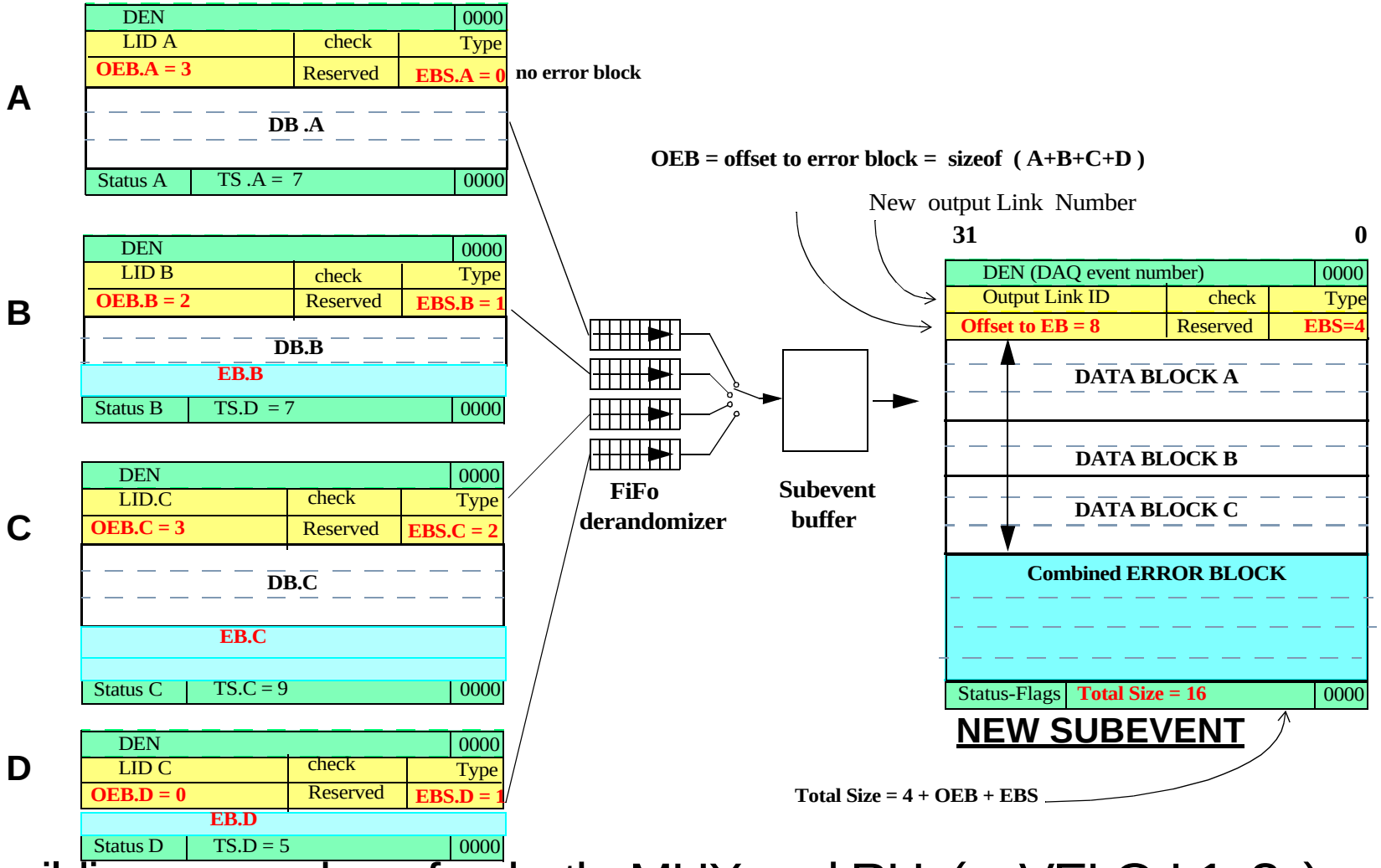
# General RU operation : convert raw link input data




# N* Slink -> Subevents on PCI or Slink output

- STF- format based subevent building process
- Application-specific FPGA logic (state machines -> VHDL -> binary configuration file)
- Slink input, derandomization -> buffer->SE
- queuing / building /errorchecking of SE
- making of $DMA_{able}$ “subevents” ( 2-16 inputs)
- ordered xmission of SE’s to PCI or Slink
- remote control via diskless PCI host card (MCU)
- output scheduling option via L1-TAGnet
- buffer fillstate monitoring ( throttle)

# Subevent Building using STF format

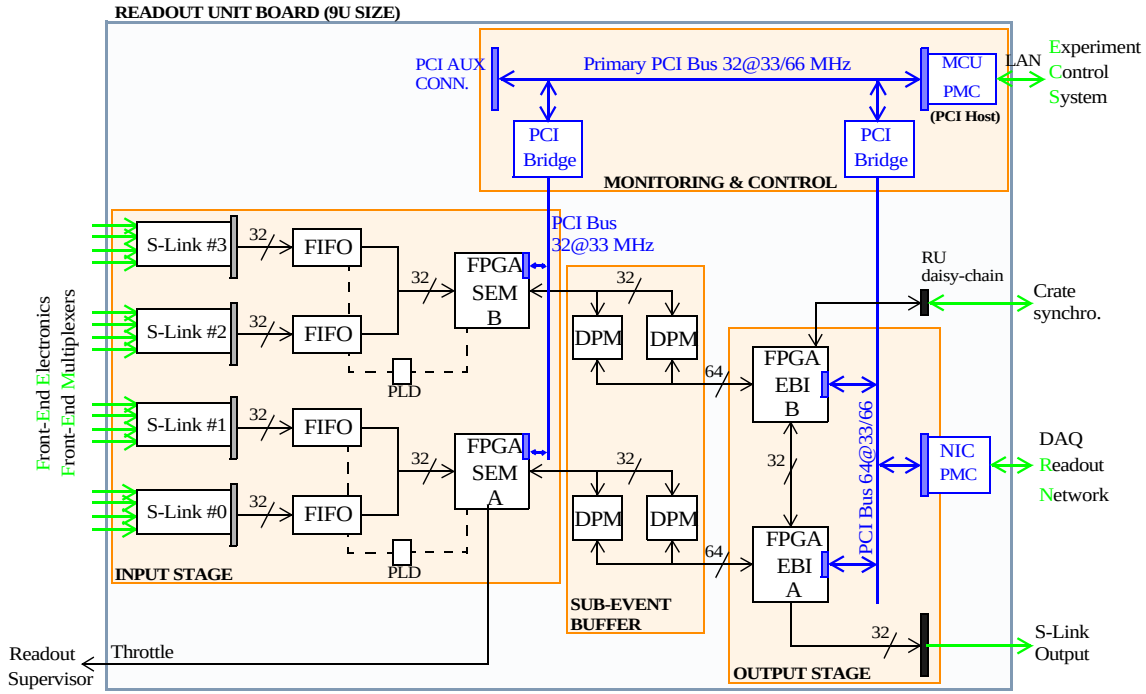



SE-building procedure for both MUX and RU ( + VELO-L1 ? )

# RU, "second edition:

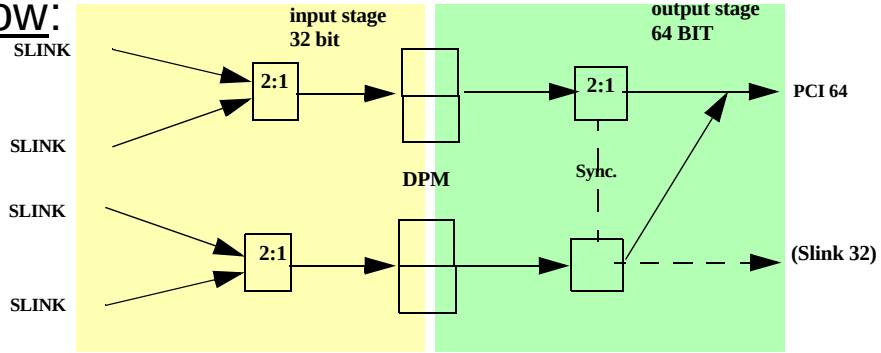
- Cheaper ( less components, less PCB layers, less connectors, no more non-standard PC card slot )
- More performant ( 2-channel, parallel architecture, faster FPGA)
- L1-Velo compliant ( Tagnet, 160 Mbyte/s S.E.building, PCI write combining for making subevents )
- More standard ( networked MCU is a VITA-36 mezzanine )
- Less power ( less chips, 2.5 Volt FPGAs )
- Symmetric architecture ( all FPGAs same type + inter-connected via PCI to networked MCU )
- MCU driven Jtag chain ( programmable devices + PCI )

# revised RU Architecture ( RU-II)




## Event data flow:

2 * dual RU




## Functional blocks

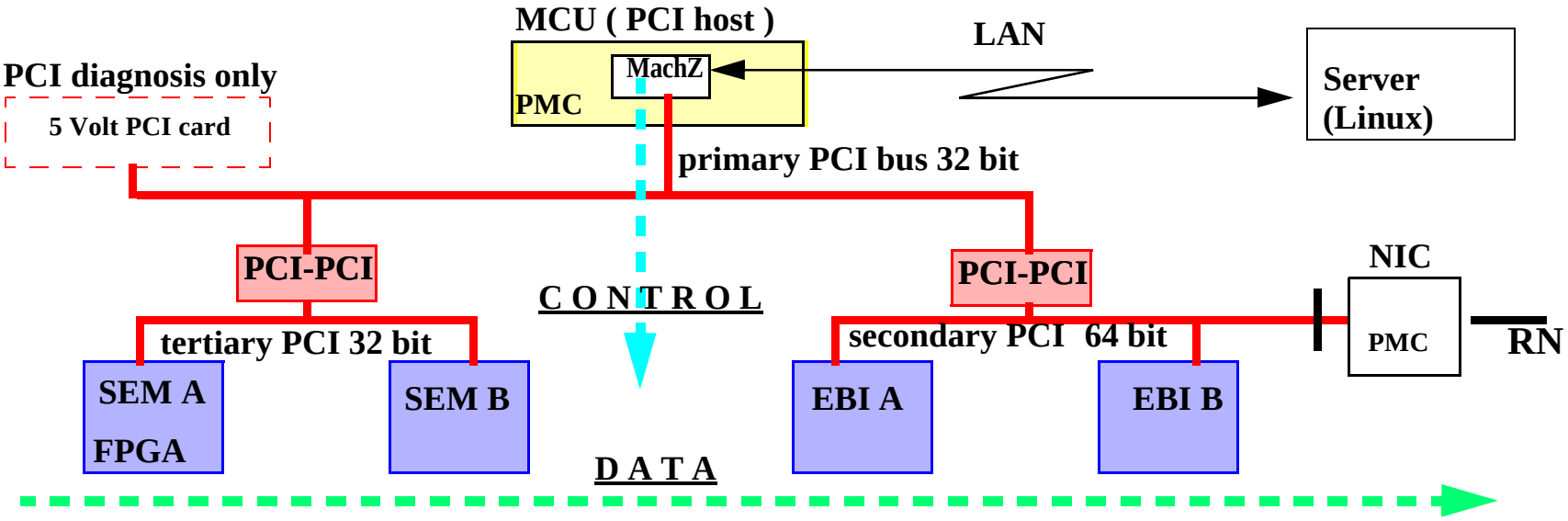
- SEM ..... 2:1 Input merger
- SEB ..... Subev. buffer
- EBI ..... 2:1 subev.builder
- MCU..... Mon.&Ctrl Unit
- NIC ..... PCI Network IF
- Slink ..... link Interface card

## Bus view

- SEM ..... 2*32 bit ->DPM
- SEB ..... DPM 2*64 bit, 1M
- EBI .....2*DPM->PCI-64
- MCU..... PMC: PCI host
- NIC ..... PCI-64 -> RN
- Slink ..... 16/32 bit links

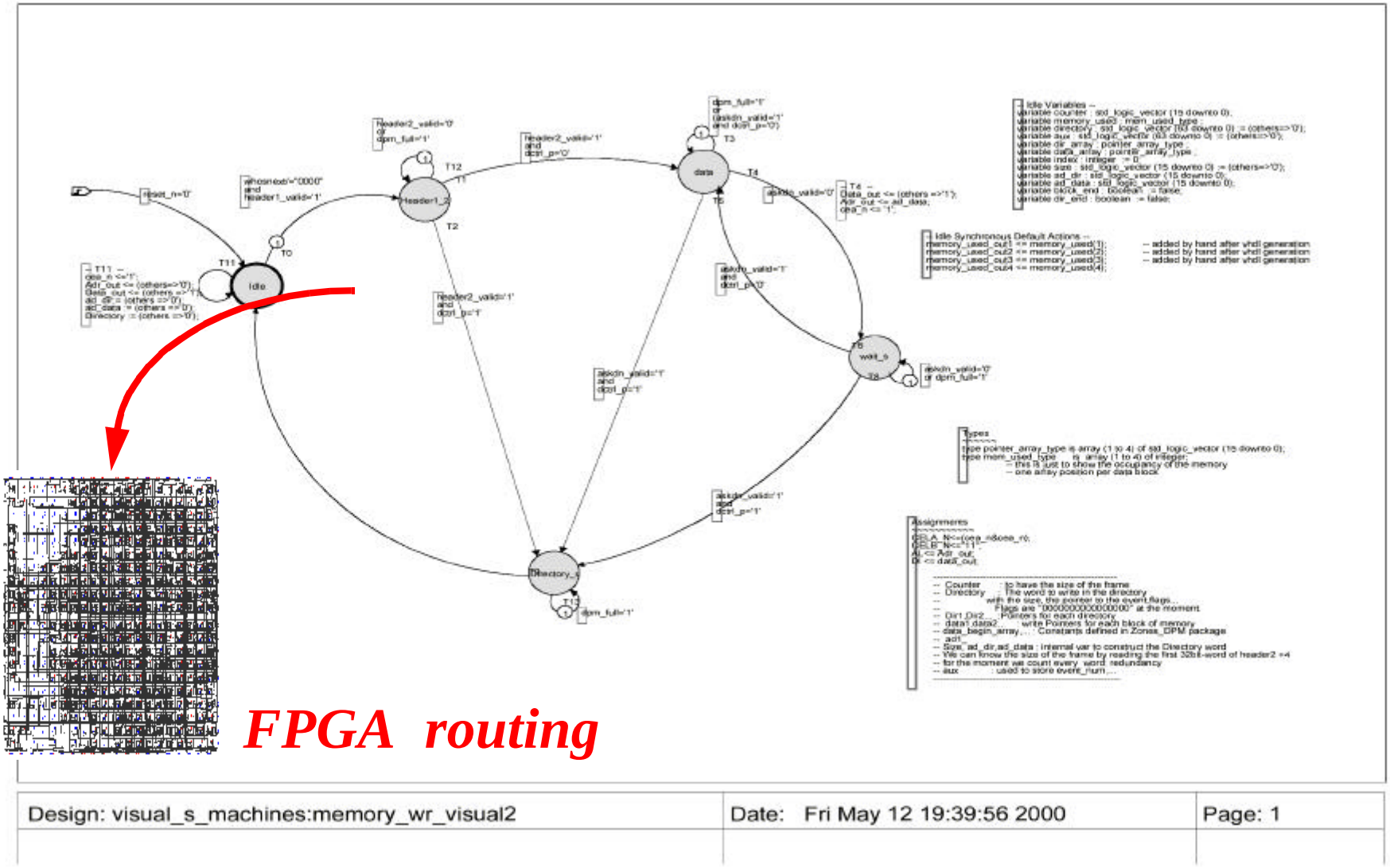
# PCI bus:

1.) Initilatization, Control & Monitoring, error handling
2.) Interface bus for data to RN (Readout Network) card
3.) Diagnostics & Development

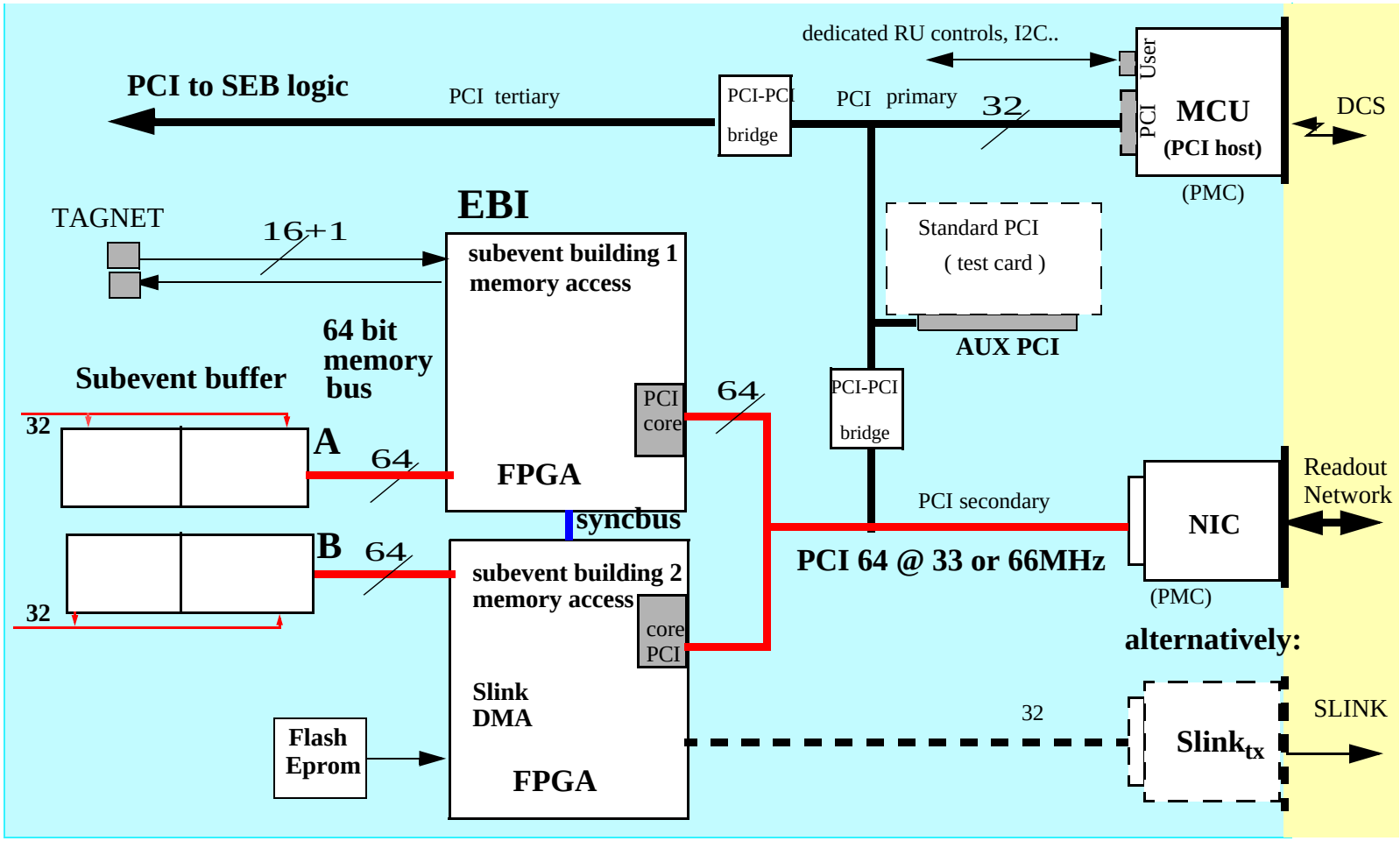



- primary PCI bus : PCI host (MCU ). No event data flow here. Available for control and monitoring, error reporting etc. any time
- secondary PCI bus: main I/O bus to NIC. Unshared high BW ( 64 bit @ 66 MHz ). FPGAs generates PCI master write + write combining to NIC for highest performance ( VELO application).
- tertiary PCI bus: FPGA configuration, Input link control ( CSRs via PCI)

# Example: Writing to SEB -buffer ( Visual HDL )



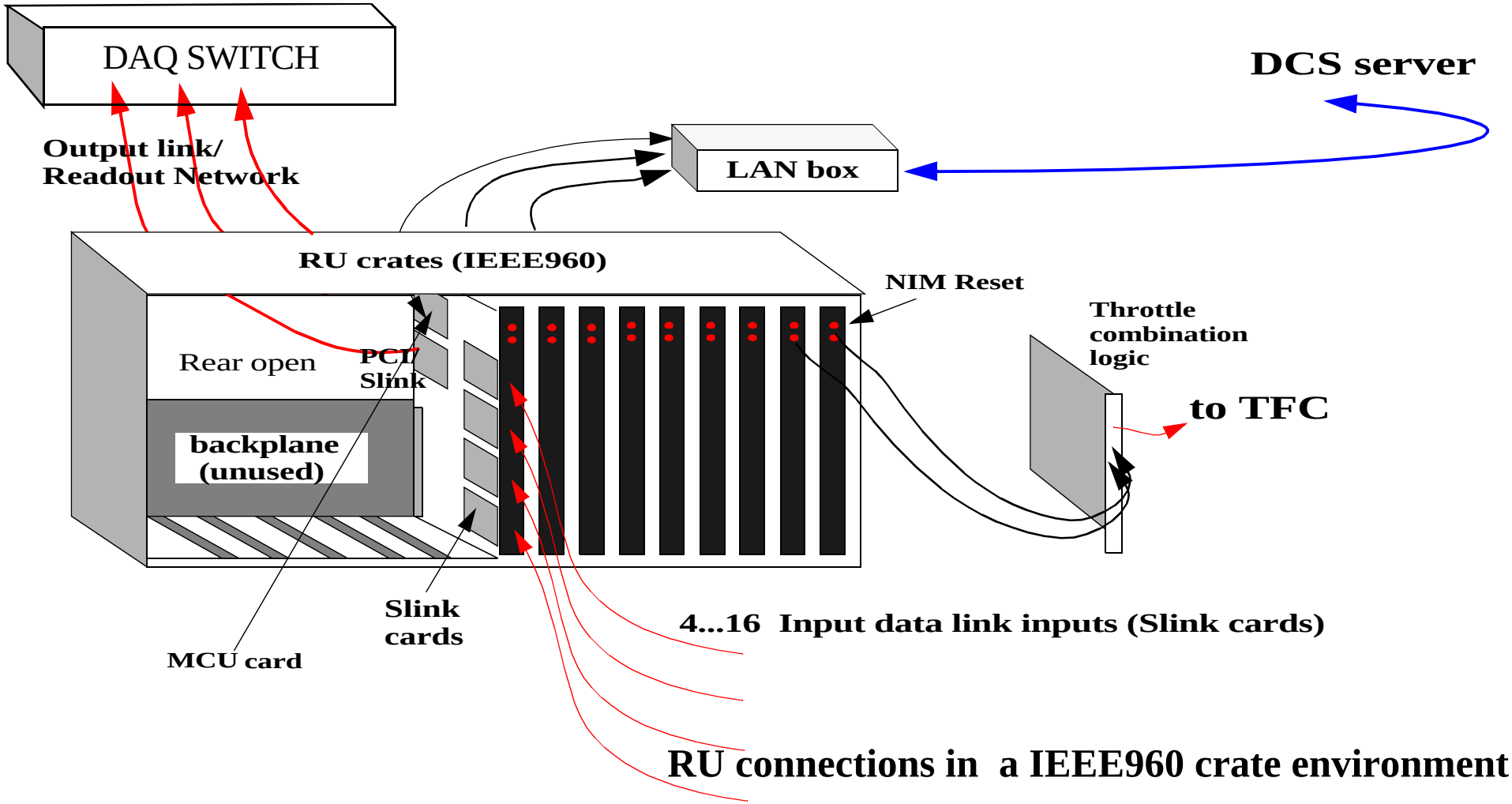
# RU-II output stage




- **MCU is now a PMC**
- **Standard PCI card: diagnostics**
- **each FPGAs configurable remotely via PCI**
- **Dedicated RU controls ( I2C, JTAG ) via PMC connector**
- **Slink or NIC share same PMC slot**

- 64 bit PCI cores of Lucent FPGAs up 66 MHz . Tandem master operation for L1-VELO
- PCI-64 high speed data path between FPGA and NIC is decoupled by PCI bridge
- “TAGnet” connector 16 bit in / 16 bit out for L1 high rate traffic scheduling ( or other )

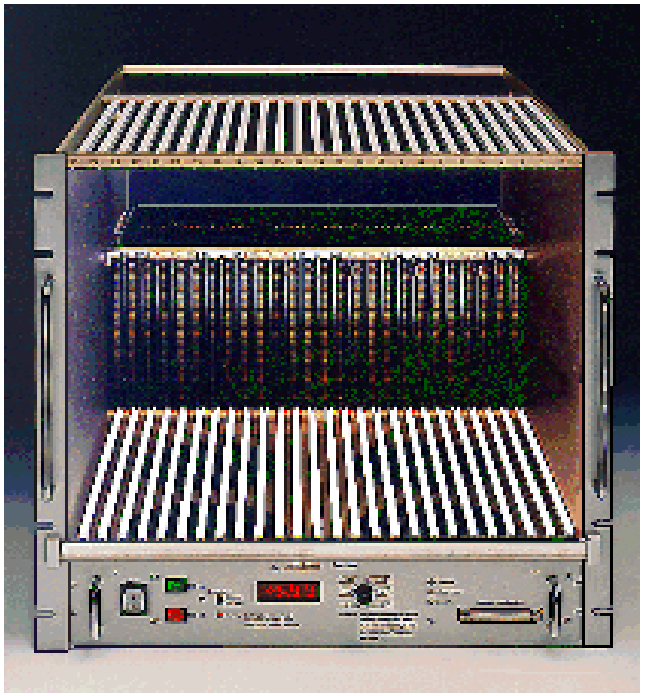
# Crate environment: FB rack/power/cooling ( exists)




- Crate for power & cooling only
- Mezzanine cards for I/O & Ctrl

# RU in IEEE 960 crates: 3 crates /rack, 3 kWatt/rack, 33 RUs/rack

- **13 RU's per crate: 1 KW**

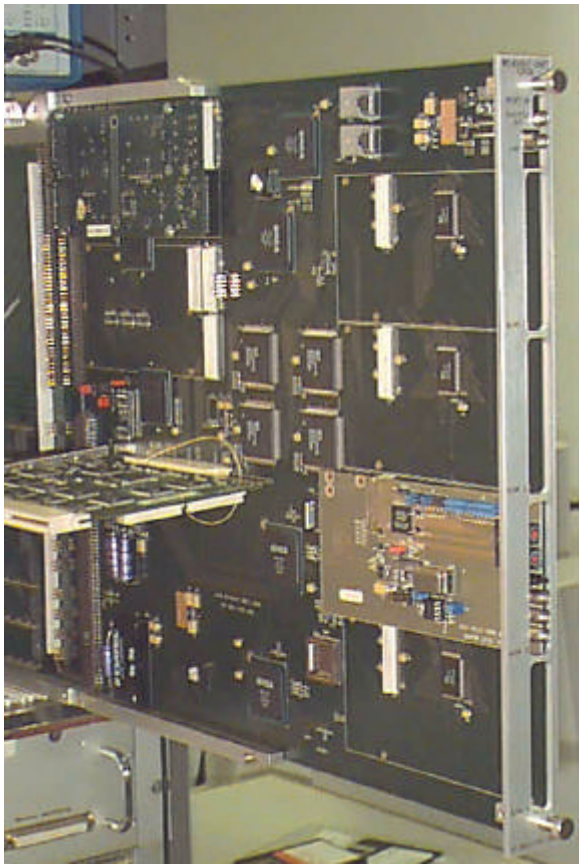


- 26 slots/C
- 4 KW max
- Backside 50% free
- Proven water/air cooling
- crates from LEP

**Table 1: RU full crate power module requirement**

| power modules | Maximum current | No of modules needed |
|---|---|---|
| +5 V | 100 A | 2 |
| +15 V | 25 A, 30 A | 1 |
| -15 V | 25 A, 30 A | 1 |

- RU-II on a FB crate extender



**RU-II: 65 Watt**

# Monitoring & Control Unit (MCU)

- Networked PC-chip on a PCI mezzanine (VITA-36 standard PMC)
- MachZ from ZFlinux: (x86 *100 MHz "PC-on-a-chip" with PCI and PIO)
- RJ45 LAN connection ( Intel chip )
- 64 Mbyte SDRAM ( "Ram disk" )
- I/O connector: JTAG, I2C, IDE-8, USB, PIO
- Diskless operation: Linux image via BootP / from server tested OK
- Local operation options: Mouse, Keyboard, Floppy, IDE connectors

I/o connector controls used by RU-II

- **i2C master for clock chip programming**
- **PIO for "remote jumper options"**
- **PCI arbiter**
- **System reset to RU mainboard**
- **IRQ ( soft reset ) from RU**
- **JTAG master ( via PIO )**

Spare controls on I/o

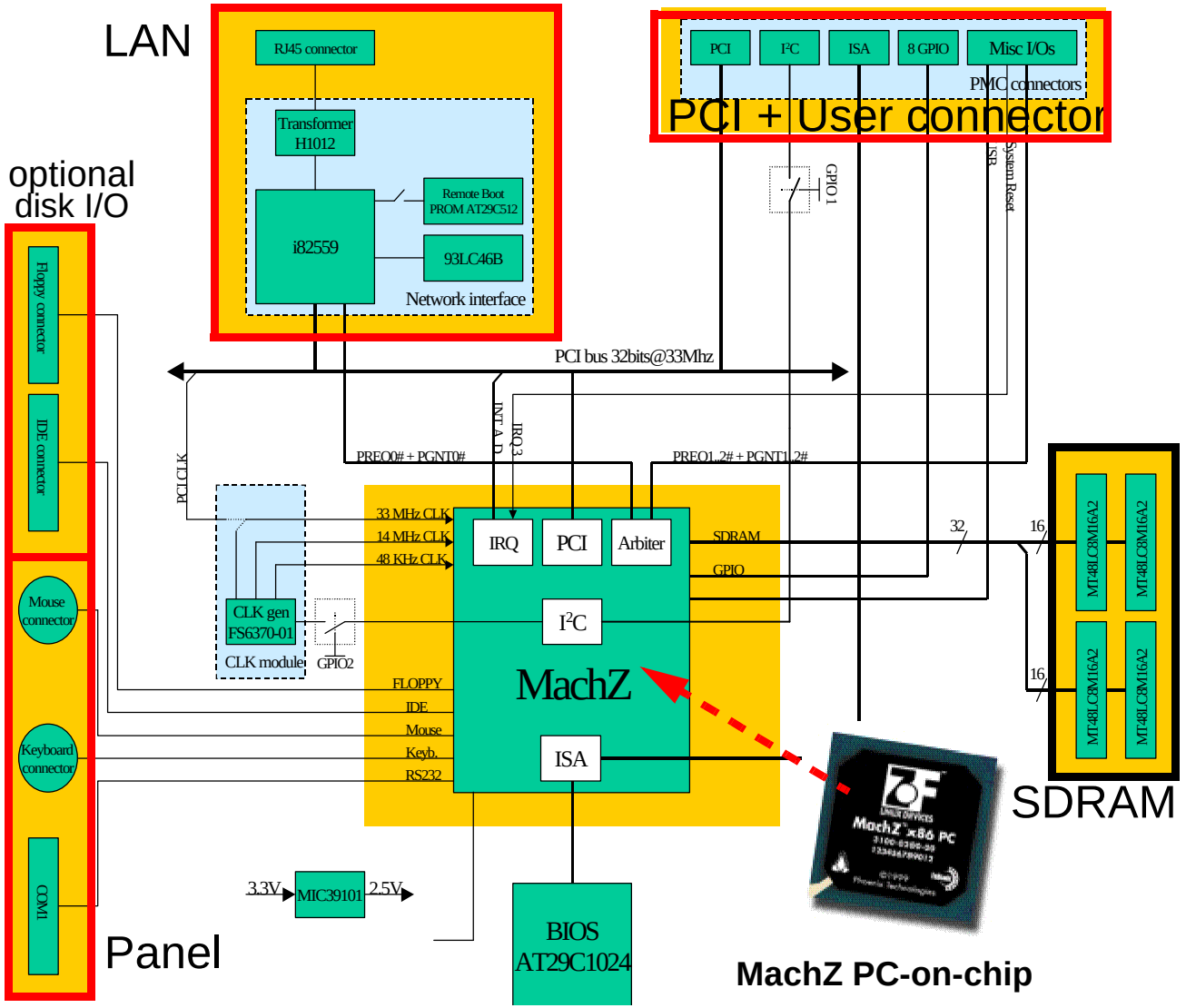
- **USB**
- **ISA -8 bit**

**Tasks for MCU**

- **FPGA loading ( .bit files ) via PCI**
- **Jtag chain via PIO, B.Bruder**
- **SCI Linux driver for VELO-L1 appl.**
- **CSR register control (FPGAs) PCI**
- **Error messages from FPGAs to LAN**
- **Clock chip programming via I2C**
- **Buffer R/write access**
- **and more..**

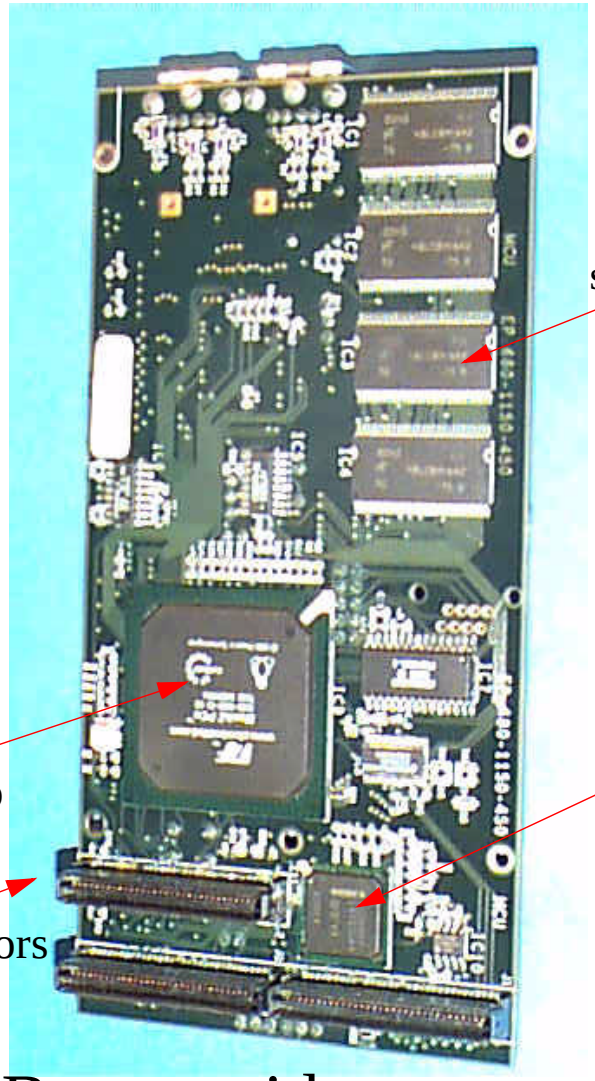
# MCU block diagram

- Very easy design ( 2 month ) combined effort of Bertrand and Curro.
- 8 layer PCB
- MachZ chip single Qty price ca 80 USD.

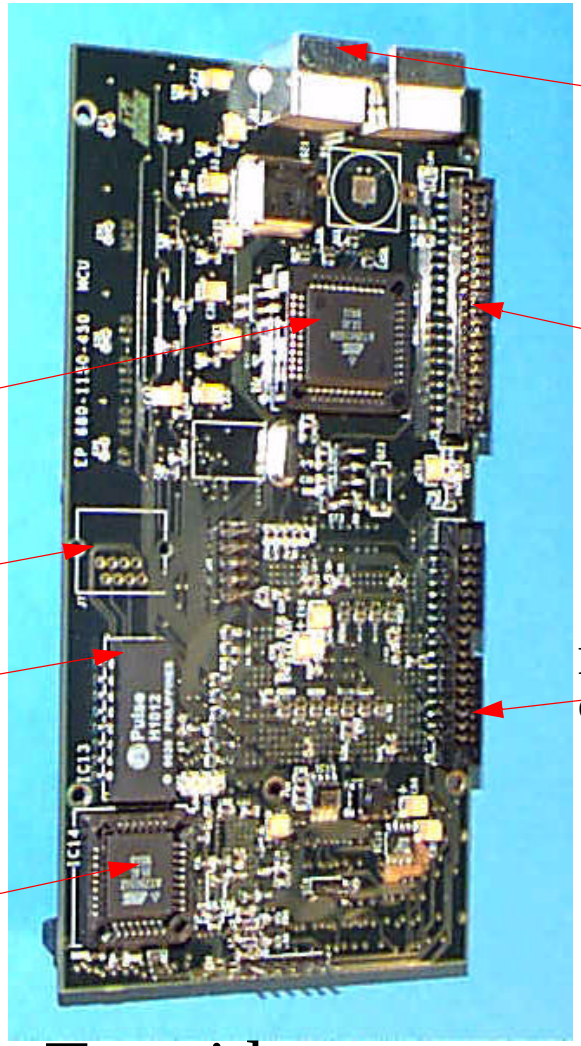



- PMC target price~ 500 FS
- Also for general purpose ( VME etc )
- Also for general purpose w.out PCI
- Remote Linux Boot and error Recovery

# MCU card for RU-II ( PMC 149 * 74 mm )




Bottom side




Top side

* Preboot Execution Environment ( intel)

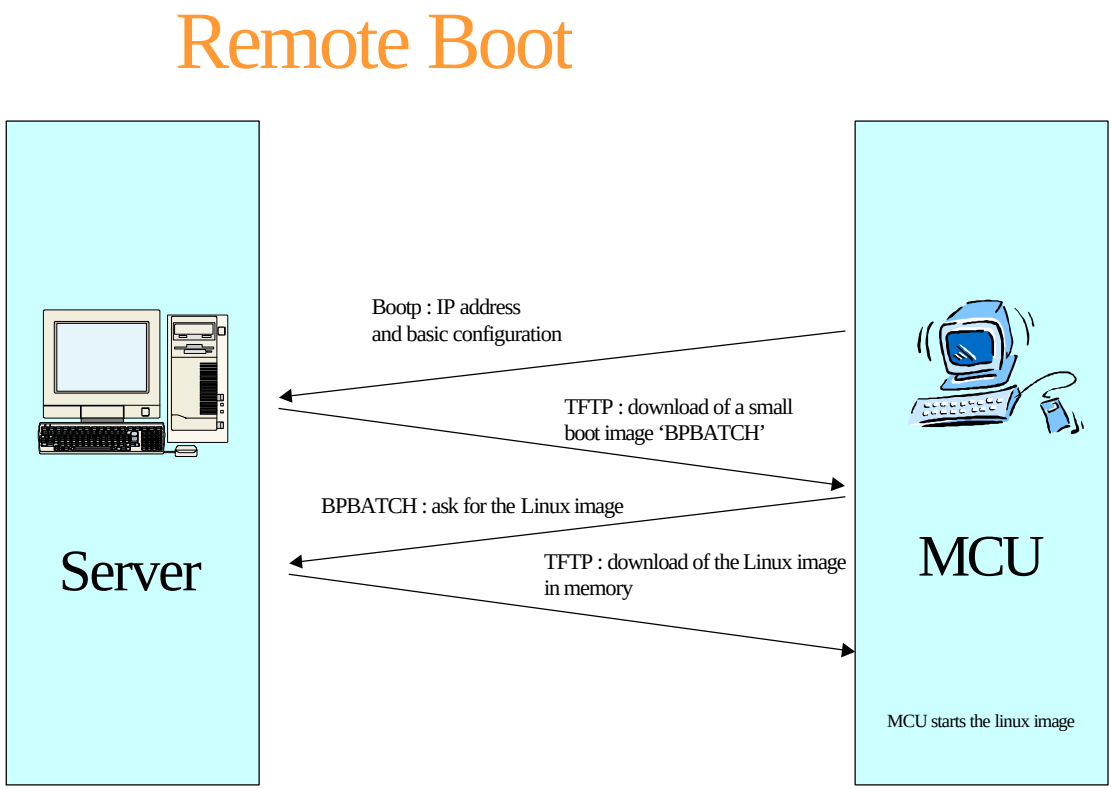
# MCU remote boot (Linux )

Harware requirements:

- Remote Boot compliant BIOS
- Ethernet interface with a flash EEPROM PXE 2.0

Software requirements

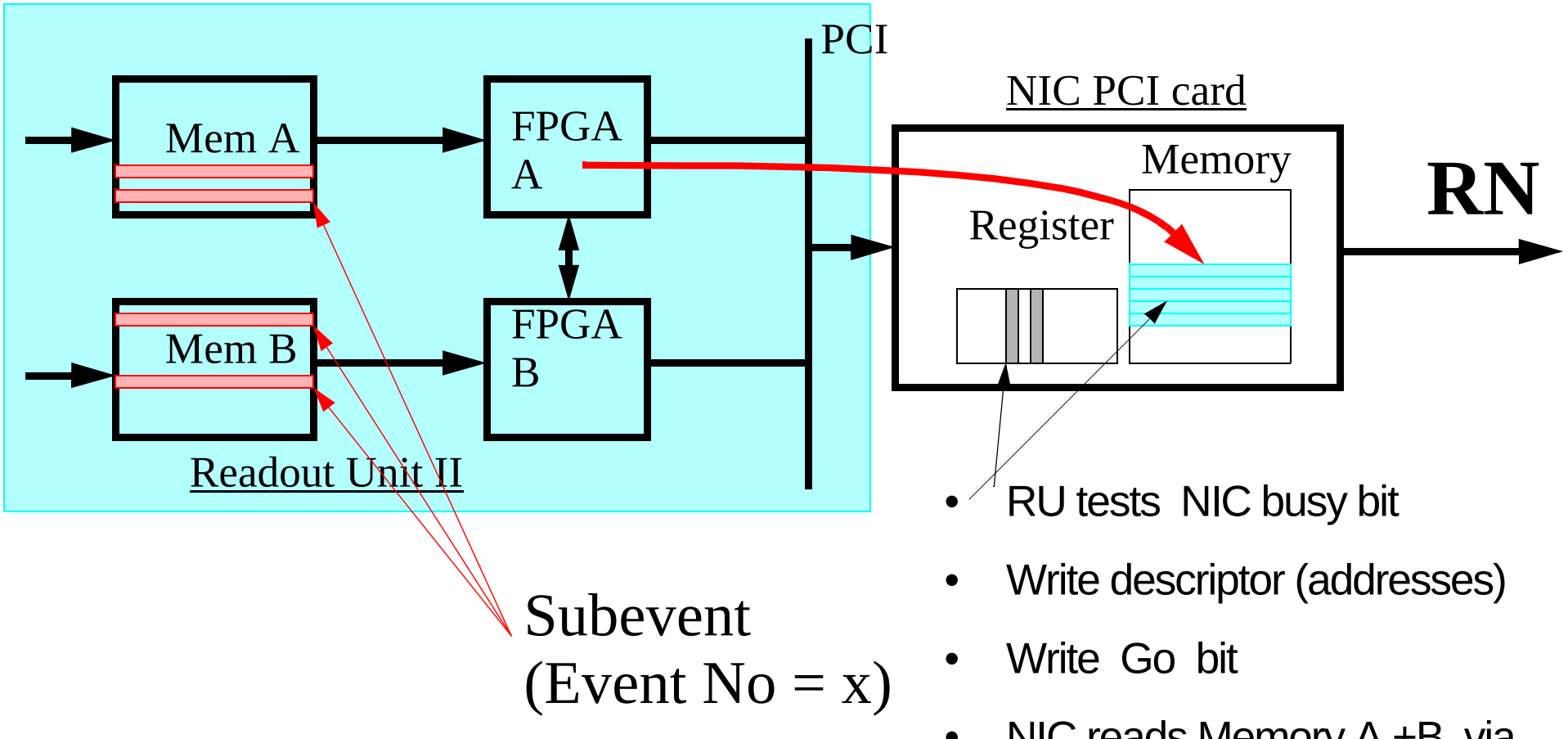
- BOOTP or DHCP server
- TFTP server
- DNS server
- Remote Root file system




C-Programs and remote boot has been tested on a MachZ development system (after removal of disks) : Linux Boot in RAM from a specially configured Server PC.

Boot time: Order < 5 sec. However watch broadcast traffic on the Network if many MCU"s are used.. Multicast should be a better way to avoid overloading the LAN.

# Output protocols NIC card (DAQ Appl.)

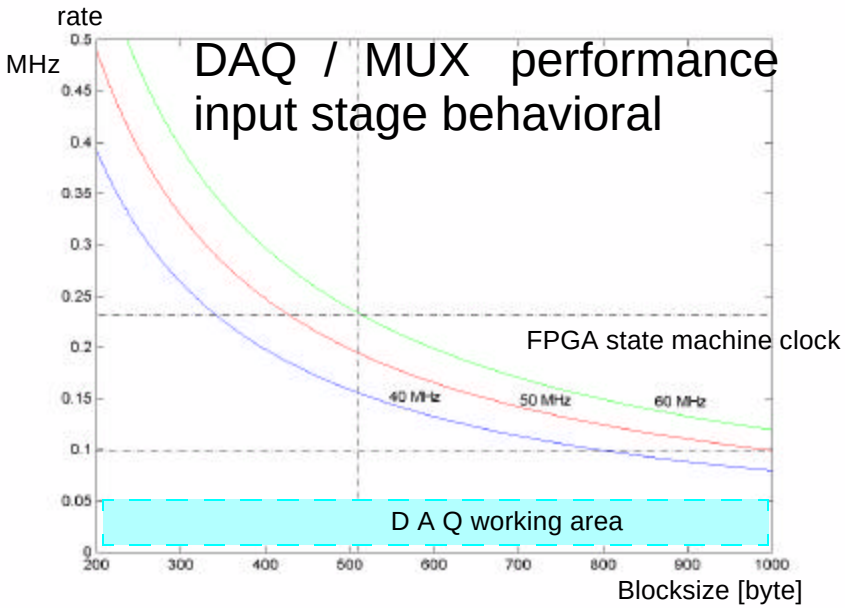


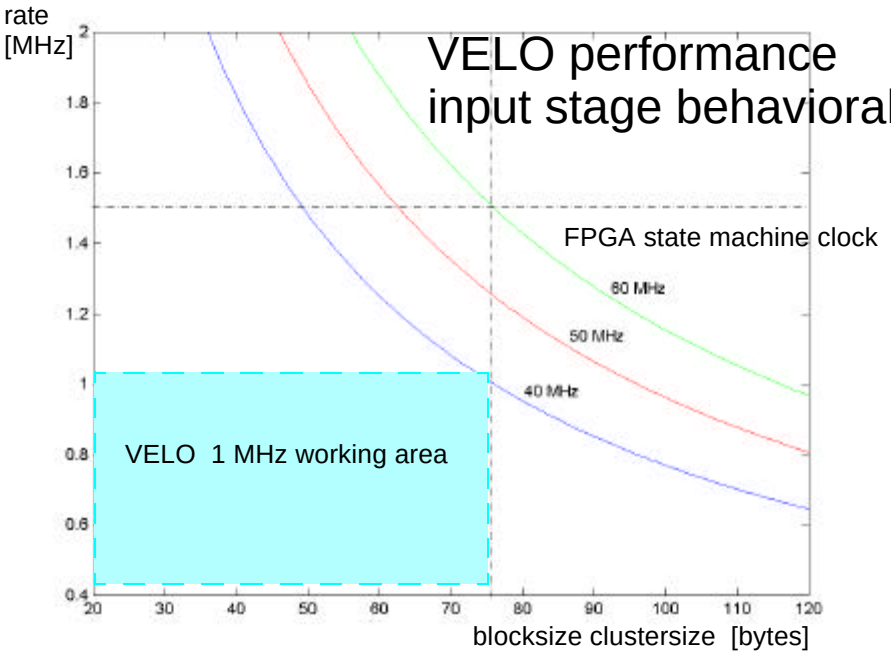

- RU tests NIC busy bit
- Write descriptor (addresses)
- Write Go bit
- NIC reads Memory A +B via PCI. Result: 1 subevent
- NIC transmits SE + resets bits

# RU performance Rate versus Blocksize

## Achieved improvements in RU-II:

- faster FPGA state machine clocks, (SEM in Prototype: 28 MHz, now factor 2 ( up to 60 MHz )
- less FiFo switching overhead ( 6 clocks in Prototype, now down to 1 ! )
- PCI tandem master operation now part of Subevent building to SCI card ( VELO)
- totally parallel working dual input stage
- 66 MHz operation on output PCI bus



# Overall performance RU-II (with output protocols)

Output protocols with PCI network cards limit output BW. Various scenarios estimated below:

**Table 1: RU-II overall performance**

| Scenario | Total output payload[a] | Input stage max. instant. trigger rate | Output stage max. sustained trigger rate | Sustainable output bandwidth |
|---|---|---|---|---|
| **DAQ**[b]: 4 * In , 1 PCI out | 2 Kbytes | 230 kHz | 60 kHz | **120 MB/s** |
| **FEM** 16 * In, 1 PCI out | 2 Kbytes | 230 kHz | 60 kHz | **120 MB/s** |
| **MUX** 16 * In,1 Slink out | 0.5 Kbytes | 2.3 MHz | 130 kHz | **70 MB/s** |
| **VELO**[c] 3 * In, PCI out | 256 bytes[d] | 1.5 MHz | 1-1.5 MHz | **256-375 MB/s** |

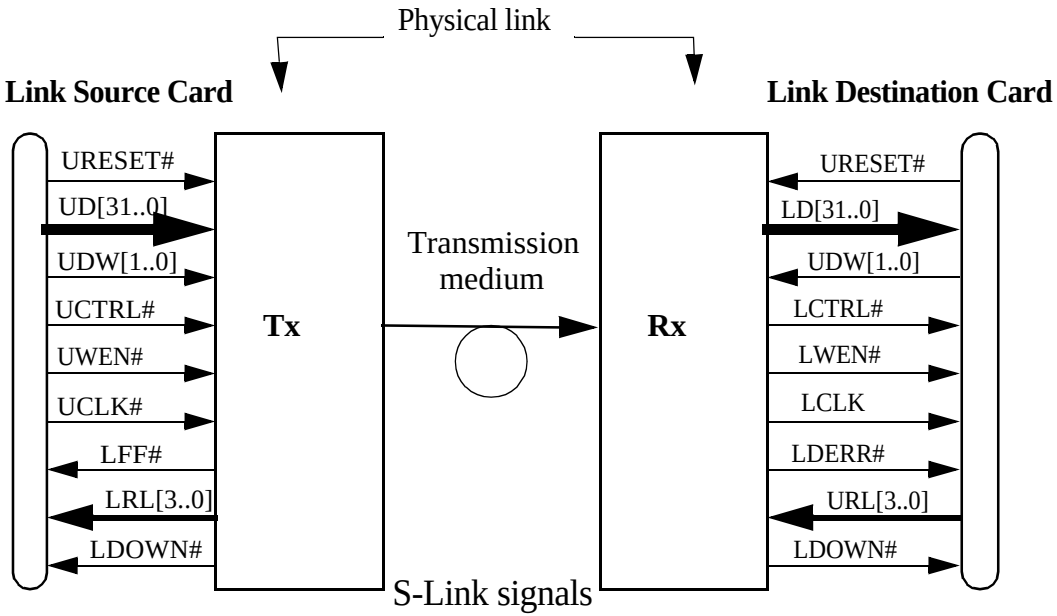
a.i sub-event size without transport framing
b. Assuming Gbit ethernet as NIC technology, descriptor based Readout by NIC card
c.Assuming SCI as NIC technology, DMA'ed Master-write to NIC, PCI write combining
d. "PCI subevent building" ( write combining) assumed with 64 bit @ 66 MHz PCI cards

BUFFER: 1/2 Mbyte allows for instantaneous higher input rates

Buffer fill state is monitored and generates Throttle signal

at programmable low and high "water" marks.

# Slink: link I/O mezzanine cards on RU




## Slink choice:

- Make link technology “transparent”.
- Use Slink design knowhow

RU: Four Slink interfaces on input, one (optional ) on output

## Slink connector relativ to PMC-type mezzanines:

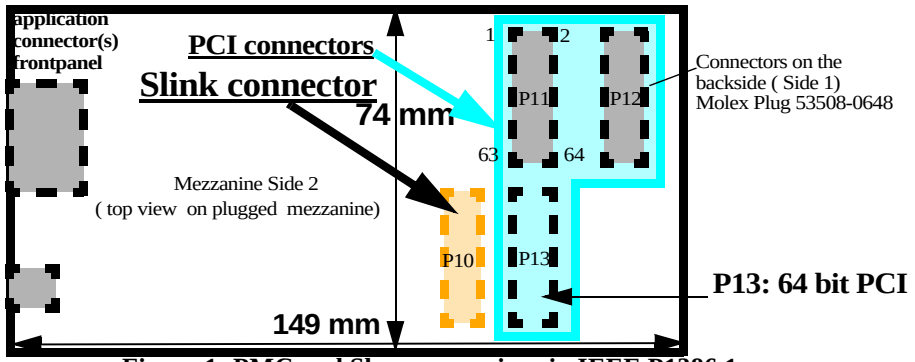


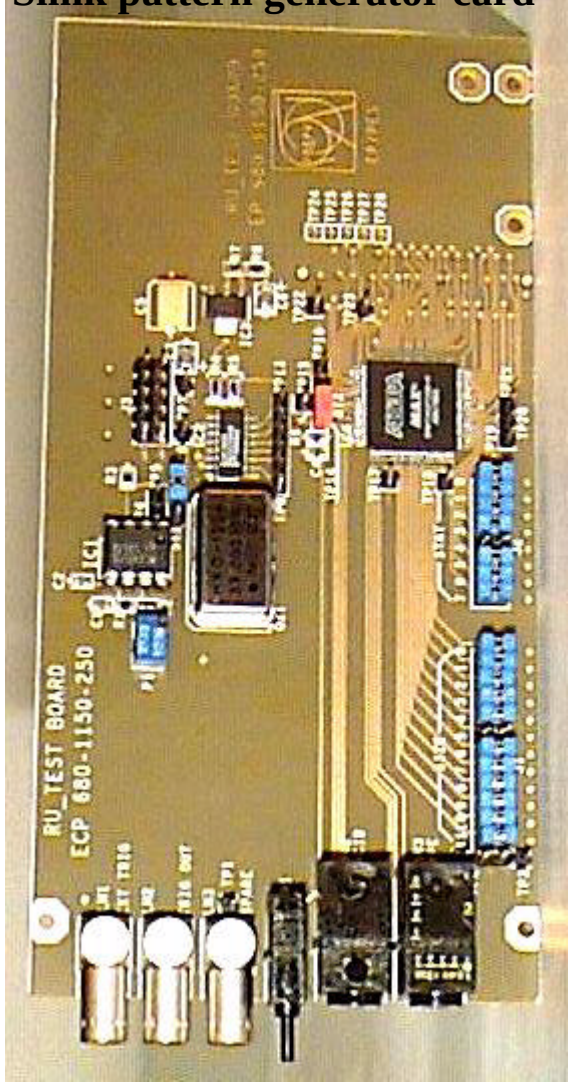

**Figure 1: PMC and Sbus mezzanines in IEEE P1386.1**

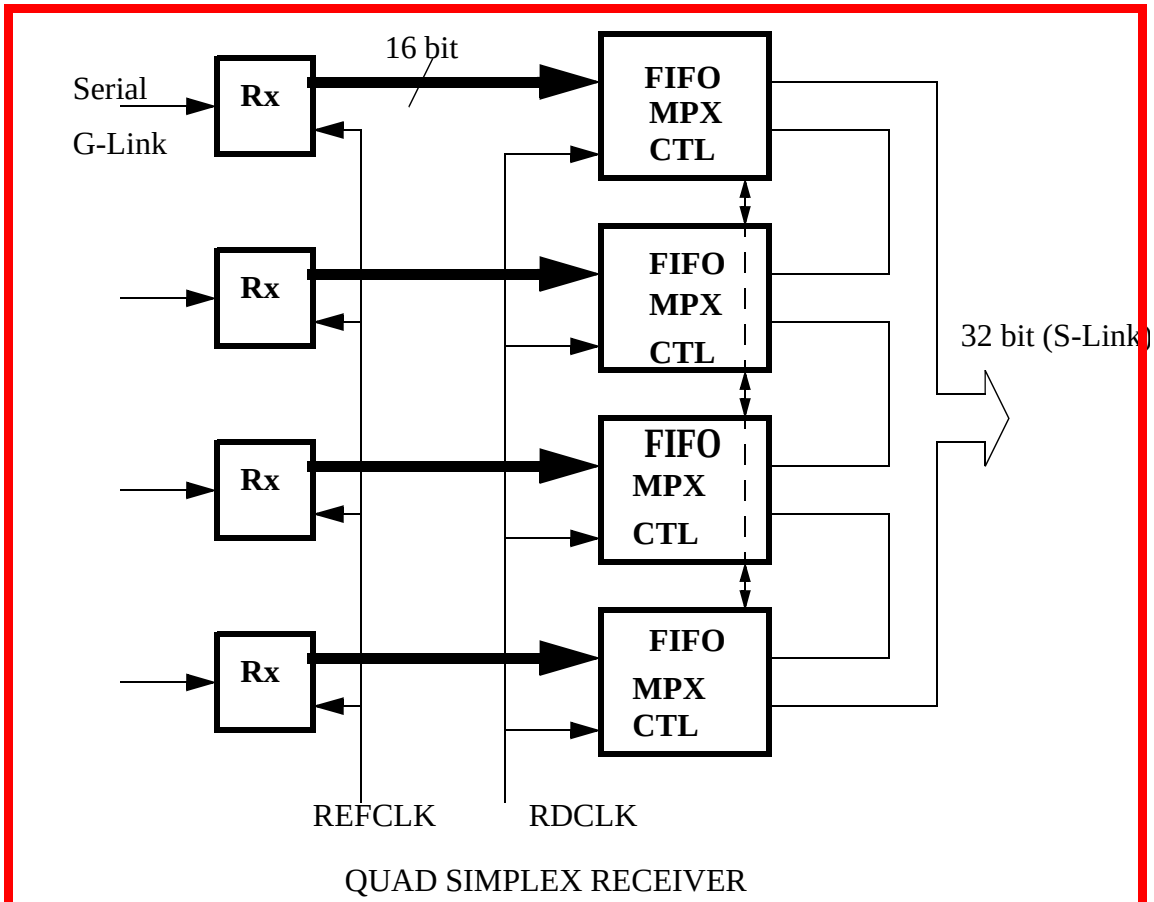
Slink mezzanine cards ~ complementary to PMC

RU: use NIC xor Slink out-put in same PMC slot

# Slink cards:

**Slink pattern generator card**

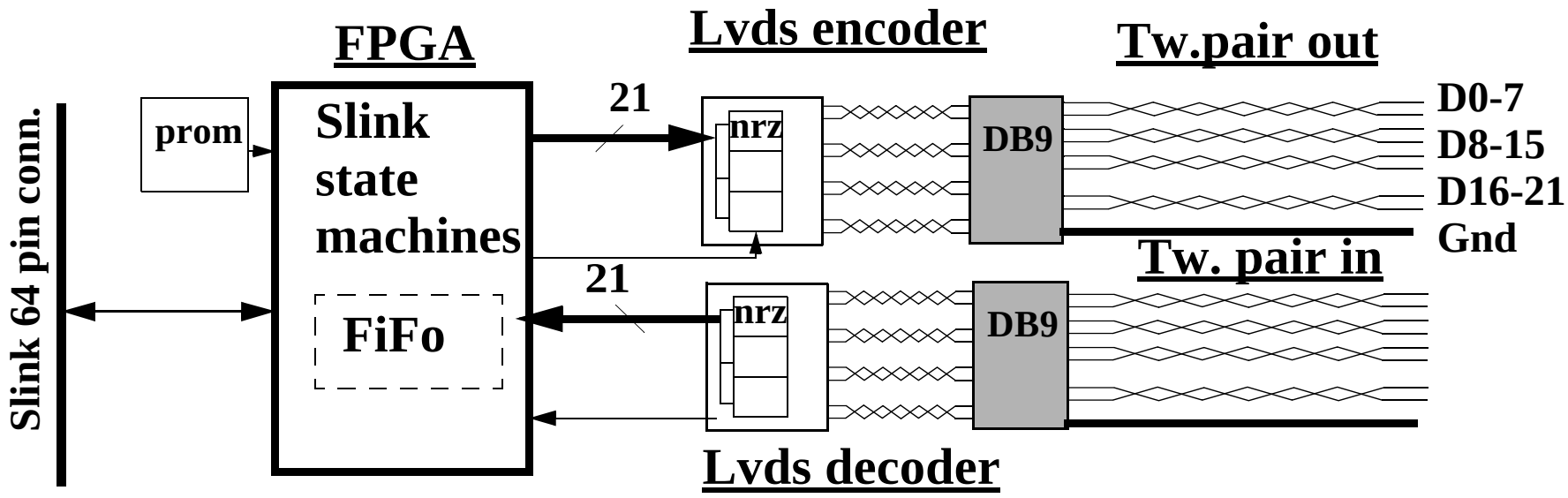


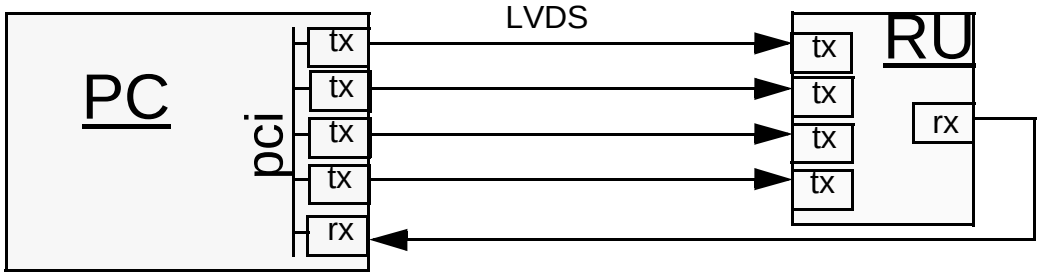

**in preparation, decision on Link Techchnology?**

**STF format generation at 40 kHz**

# Universal Slink card ( LVDS) Tx or Rx




- Card being produced ( F. Bal )
- Only Tx or Rx at a time ( Prom )
- For loop testing of RU-II with Slink-PCI card

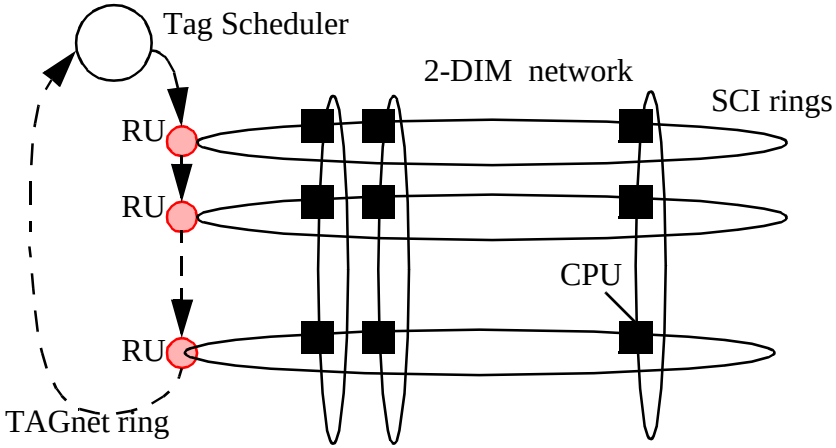



planned SE Building loop tests with "LHCb-like" data

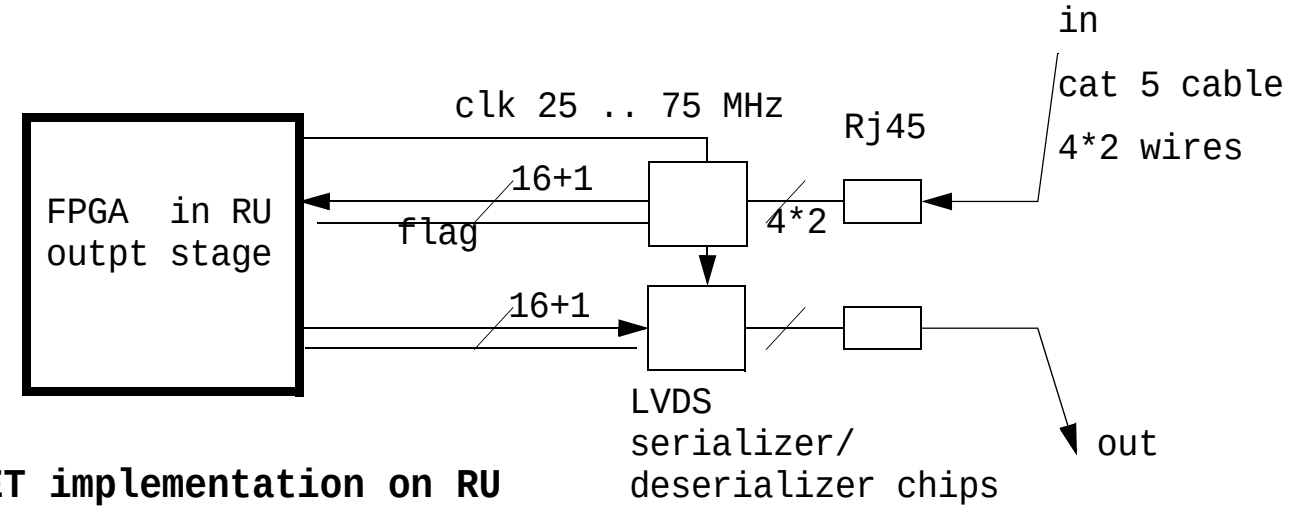
# Status

- RU applications: DAQ, MUX and L1-VELO
- STF transport format for subevent building MUX, DAQ
- New RU-II architecture: less cost and higher performance
- 6 new RU-II ( 8layer 9U boards) received, 1 RU fully under tests
- Subeventbuilding 4*Slink->1* Slink test planned this week
- MCU card ( PMC) completed /under test, remote boot tested OK
- 4 Slink LHCb pattern generators ( Slink card F.Bal) completed and in use
- Bidirectional Slink LVDS card ( Rx / Tx ) commissioned for tests
- PC-piloted loop tests ( subeventbuilding) with RU-II planned
- Docs / status http://hmuller.home.cern.ch/hmuller/lhcb_projects.htm
- EDMS archive ( schematics,docs etc) edms.cern.ch see under electronics design archive
- RU distribution / discussion <lhcb-daq-ru@listbox.cern.ch>

# A.1 Tagnet I/O scheduling ( Level 1 VELO Trigger )

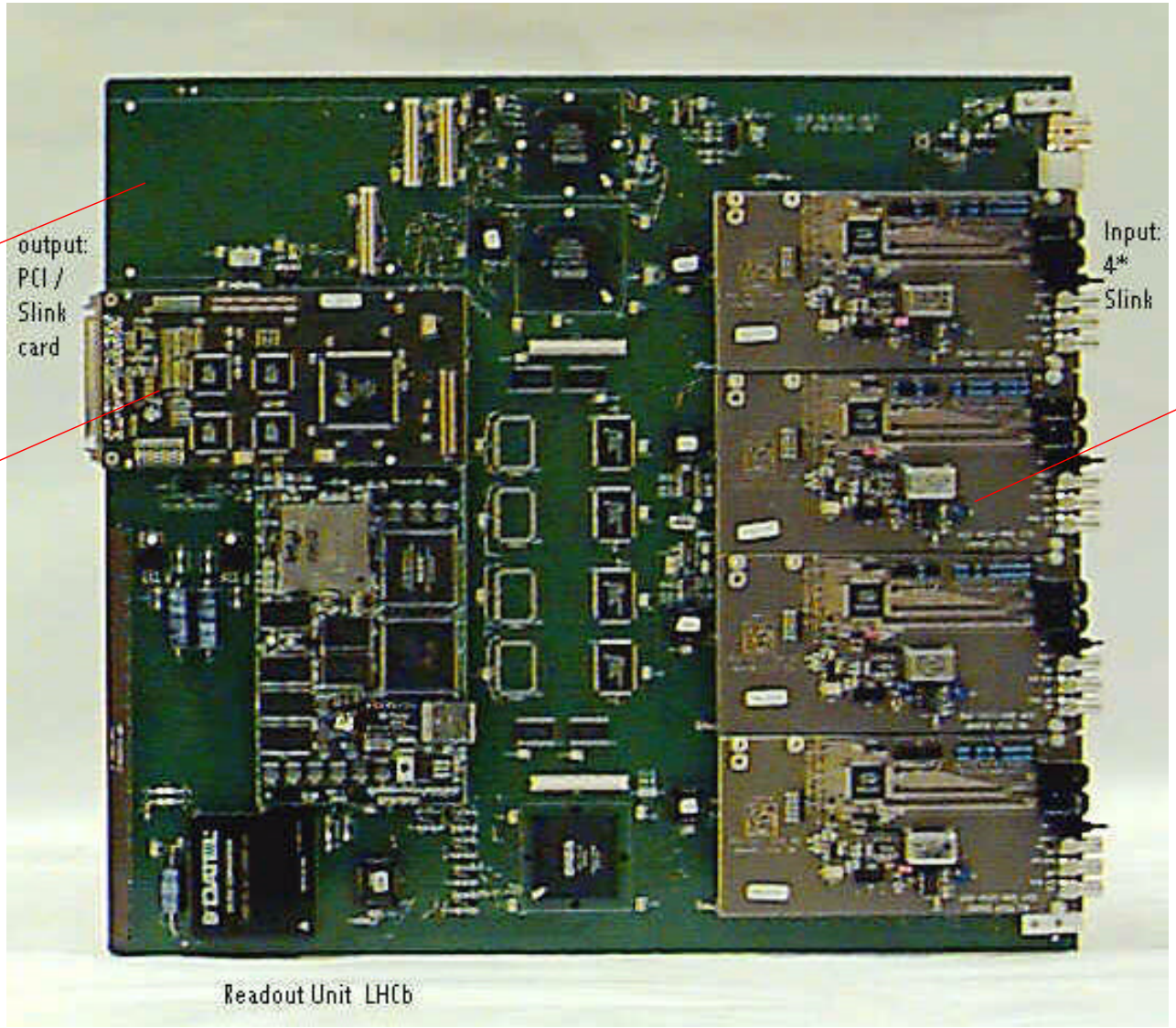



L1 2D mesh: at any moment, only 1 row and one column is fed with data. There is 1 TAG every 1 us. A “scheduler” which keeps a list of all TAGNET tokens




**TAGNET implementation on RU**

# A.3 RU prototype card



# A.4 Readout Unit “second edition