FAIR Facility for Antiproton and Ion Research

#### GSÅ

# FutureDAQ for CBM: On-line Event Selection

About FAIR About CBM About FutureDAQ About Demonstrator

CHEP06, Mumbai

# FAIR Facility for Antiproton and Ion Research





### FAIR in 2014





CHEP06, Mumbai

Hans G. Essel, GSI / FAIR, CBM collaboration www.gsi.de



- Nuclear Structure Physics and Nuclear Astrophysics with RIBs
- Hadron Physics with Anti-Proton Beams



Physics of Nuclear Matter with Relativistic Nuclear Collisions





Plasma Physics with

highly Bunched Beams

- Atomic Physics and Applied Science with highly charged ions and low energy Anti-Protons
- + Accelerator Physics



U+U 23 AGeV



CHEP06, Mumbai

Hans G. Essel, GSI / FAIR, CBM collaboration www.gsi.de

### **CBM** physics topics and observables



enhanced strangeness production ? multi guark states? measure: K,  $\Lambda$ ,  $\Sigma$ ,  $\Xi$ ,  $\Omega$ 

#### 4. Critical endpoint of deconfinement phase transition

event-by-event fluctuations measure:  $\pi$ , K

- High data rates
- Short latency (µsec)
- Complex (displaced vertices)
- Most of the data needed





- 1. A conventional LVL1 trigger would imply full displaced vertex reconstruction within fixed (short) latency.
- 2. Strongly varying complex event filter decisions needed on almost full event data

### No common trigger! Self triggered channels with time stamps! Event filters

- 10 MHz interaction rate expected
- 1 ns time stamps (in all data channels, ~10 ps jitter) required
- **1 TByte/s** primary data rate (Panda < 100 GByte/s) expected
- **GByte/s** maximum archive rate (Panda < 100 MByte/s) required
- Event definition (time correlation: multiplicity over time histograms) required
- Event filter to 20 KHz (1 GByte/s archive with compression) required
- On-line track & (displaced) vertex reconstruction required
- Data flow driven, no problem with latency expected
- Less complex communication, but high data rate to sort





# FutureDRQ

### European project 2004 (FP6 $\rightarrow$ I3HP $\rightarrow$ JRA1)

FP6: 6<sup>th</sup> Framework Program on research, technological development and demonstration I3HP: Integrated Infrastructure Initiative in Hadron Physics JRA: Joint Research Activity

### Participants from

- GSI (Spokesperson: Walter F.J. Müller)
- Kirchhoff Institute for Physics, Univ. Heidelberg
- University of Mannheim
- Technical University Munich
- University of Silesia, Katowice
- Krakow University
- Warsaw University
- Giessen University
- RMKI Budapest
- INFN Turino

### Studying AA-collisions from 1 - 45 AGeV



**CBM Detector: 8 - 45 AGeV** 

## The CBM detectors

#### At 10<sup>7</sup> interactions per second!





#### At 10<sup>7</sup> interactions per second!



- Radiation hard Silicon (pixel/strip) tracker in a magnetic dipole field
- Electron detectors: RICH & TRD & ECAL pion suppression up to 10<sup>5</sup>
- Hadron identification: RICH, RPC
- > Measurement of photons,  $\pi^0$ ,  $\eta$  and muons electromagn. calorimeter ECAL

Multiplicities:	160	р
	400	π-
	400	π+
	44	K+
	13	Κ
	800	γ
	1817	total at 10 MHz





CHEP06, Mumbai

### **DAQ** hierarchy



### TNet: Clock/time distribution



#### Challenge of time distribution:

- TNet must generate GHz time clock with ~10 ps jitter
- must provide global state transitions with clock cycle precise latency
- Hierarchical splitting into 1000 CNet channels

#### Consequences for serial FEE links and CNet switches:

- bit clock cycle precise transmission of time messages
- low jitter clock recover required
- FEE link and CNet will likely use custom SERDES (i.e. OASE)

### **CNet: Data concentrator**



### **BNet: Building network**



Has to sort parallel data to sequential event data Two mechanisms, both with traffic shaping

- switch by time intervals
  - all raw data goes through BNet
  - + event definition is done behind BNet in PNet compute resources
- switch by event intervals
  - event definition done in BNet by multiplicity histogramming
  - some bandwidth required for histogramming
  - + suppression of incoherent background and peripheral events
  - + potentially significant reduction of BNet traffic

Functionality of *data dispatcher* and *event dispatcher* implemented on one *active buffer* board using bi-directional links.

Simulations with mesh like topology

### **BNet: Factorization of 1000x1000 switch**



### Simulation of BNet with SystemC

#### Modules:

- event generator
- data dispatcher (sender)
- histogram collector
- tag generator
- BNet controller (schedule)
- event dispatcher (receiver)
- transmitter (data rate, latency)
- switches (buffer capacity, max. # of package queue, 4K)

Running with 10 switches and 100 end nodes. Simulation takes 1.5 \*10<sup>5</sup> times longer than simulated time. Various statistics (traffic, network load, etc.)

### BNet: SystemC simulations 100x100



CHEP06, Mumbai

Hans G. Essel, GSI / FAIR, CBM collaboration www.gsi.de

## PNet: Structure of a sub-farm



- A sub-farm is a collection of compute resources connected with a PNet
- Compute resources are
  - programmable logic (FPGA)
  - processors
- Likely choice for the processors are high performance SoC components
  - CPUs, MEM, high speed interconnect on one chip
  - optimized for low W/GFlop and high packing density
  - see QCDOC, Blue Gene, STI cell, ....
- PNet uses 'build-in' serial links connected through switches
- PCIe-AS is a candidate for a commonly used serial interconnect
- A plausible scenario for the low level compute farm
  - O(100) sub-farms with O(100) compute resources each
  - one sub-farm on O(10) boards in one crate
- Consequences
  - only chip-2-chip and board-2-board links in PNet
  - thus only short distance (<1m) communication</li>

### PNet: First & second level computing

#### Event selection level 1 (FPGA): 1% Event selection level 2 (CPU): 10%



64-128 sub-Farms, each with 32 FPGA and 32 CPU

1 GByte/s





#### Five different networks with very different characteristics

#### CNet (custom)

- Capture hit clusters, communicate geographically neighboring channels
- Distribute time stamps and clock (from TNet) to FEE
- Low latency bi-directional optical links (OASE)
- Eventually communicate detector control & status messages
- connects custom components (FEE ASICS, FPGAs)
- TNet (custom)
  - generates GHz time clock with ~10 ps jitter
  - provides global state transitions with clock cycle precise latency
- BNet (standard technology, i.e. Ethernet or Infiniband)
  - switch by time intervals: event definition is done behind BNet in PNet compute resources
  - switch by event intervals: event definition done in BNet by multiplicity histogramming
- PNet (custom)
  - short distance, most efficient of already 'build-in' links (i.e. PCIe-AS)
  - connects standardized components (FPGA, SoCs)
- HNet
  - general purpose, to archive

# DAQ Demonstrator



# Prototype for DC and DD

### Data Collector Board (DC)

### Data Dispatcher (DD)



- bi-directional (optical) link
  - data, trigger, Rol, control, clock
- FPGA
  - logic for data (/protocol) processing
  - processor for control
    - external memory (DDR)
    - Ethernet as main control interface
  - external memory
    - for data storage
    - •
  - PC interface (PCIexpress)
    - Interface to Bnet



V4FX Testboard for DC and DD (Joachim Gläss, Univ. Mannheim)

- PPC with external DRAM and Ethernet (Linux): Software for control
- Test and classification of MGTs
  - optical, copper, backplane, ...
- Test of OASE
- PCIexpress interface to PC
- Develop and test RAM controllers
  - DD prototype for demonstrator Develop and test algorithms (shrinked down)

DC prototype for demonstrator

**MGTs** P (miniGBIC) Virtex4 mezzanines A ER ..... 0 . 0 **PClexpress** 

4 x OASF on mezzanine

2 x RAM on mezzanine

Probe on mezzanine





- Final FE, DC, DD boards (to do)
- FPGA codes (to do)
- Link management (to do)
- Data formats (to do)
- Interfaces (to do)
- Framework
  - xDaq (CMS) under investigation
- Controls
  - EPICS (+ xDaq) under investigation
- InfiniBand cluster with 4 double opteron PCs installed
  - MPI and uDAPL tested (msg, RDMA) (900KB/s with 20KB buffers)
- In production end of 2007 (hopefully)