ONSEN

(Online Selector Nodes)

Dennis Getzkow[2], Thomas Geßler[2], Wolfgang Kühn[2], Jens Sören Lange[2], Klemens Lautenbach[2], Zhen-An Liu[1], Björn Spruck[3], Jingzhou Zhao[1], (Leonard Koch[2], David Münchow[2]), [1]IHEP Beijing, [2]Univ. Giessen, [3]Univ. Mainz

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

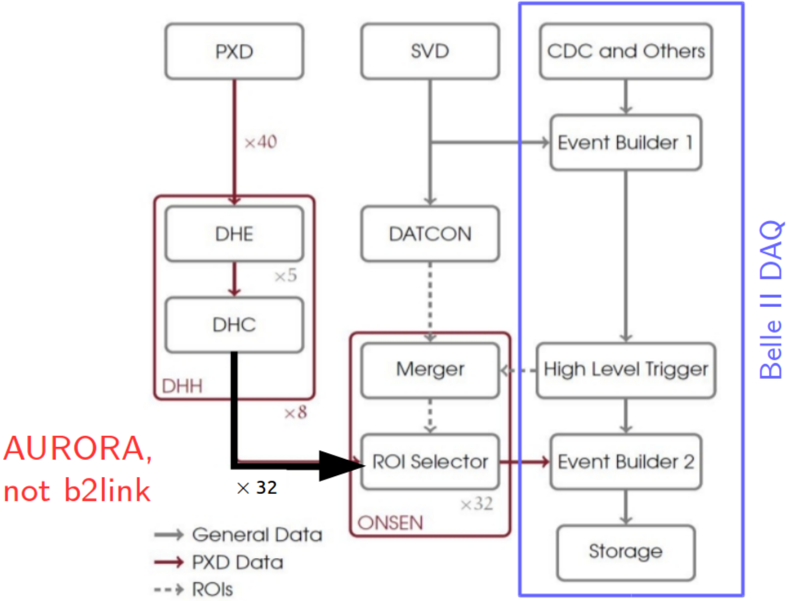
JUSTUS-LIEBIG-
UNIVERSITÄT
GIESSEN

JG|U
JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Outline

- Overview of PXD DAQ
- ONSEN
  - Hardware status
  - Full system test at Giessen, results
  - Processing basf2 events in ONSEN
  - Answer to questions, raised in BPAC report 10/2016

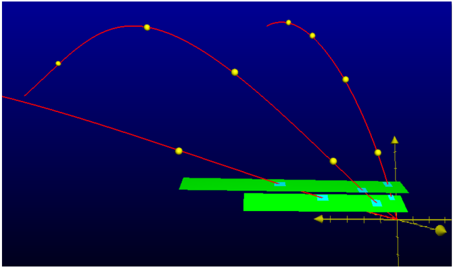
PXD DAQ Overview



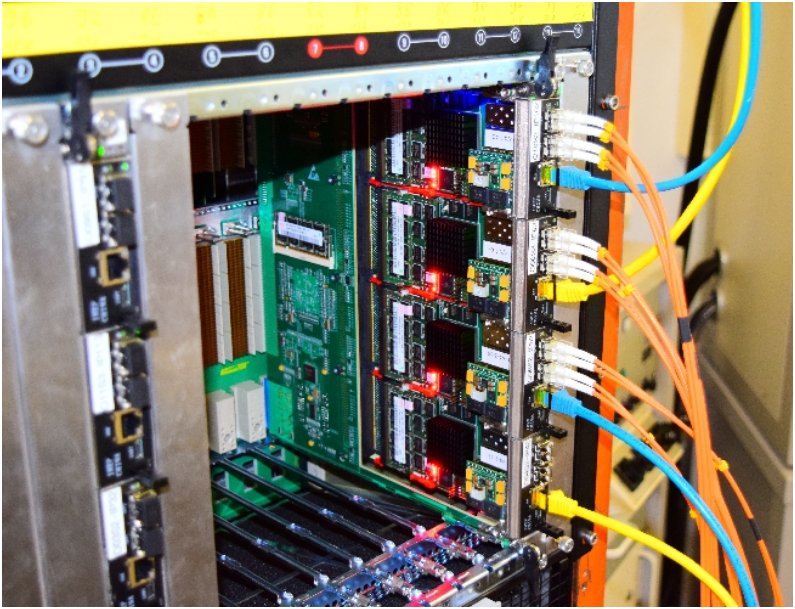
AURORA,
not b2link

PXD DAQ parameters

Trigger 30 kHz (1/3 accept, 2/3 reject)
≤3% PXD occupancy data input ≤21.6 GB/s
ROI selection (region of interest) HLT (SVD+CDC), PC farm DATCON (SVD only), FPGA logical OR (on ONSEN) data reduction factor ≥10



ONSEN 1/8 system



Status of ONSEN hardware

ONSEN AMC card
v4.0 (final)
Virtex-5 FX70T
2 optical links (6.25 Gbps)
GbE

DATCON AMC card
Virtex-5 LX50T
4 optical links (3.125 Gbps)

slow control / monitoring:
IPMI add-on boards (Mainz)

Status of ONSEN hardware

ONSEN xTCA carrier card

v3.3 (final)

Virtex-4 FX60
(switcher
to ATCA backplane)

GbE

add-on:
RTM board
power supply board

AMC card mass production

ONSEN hardware status

AMC v4.0	
10	KEK
8	DESY
4	IHEP (repair)
21	Giessen
43	(total)

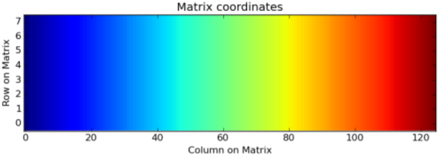
Carrier v3.3	
3	KEK
2	DESY
1	IHEP (repair)
6	Giessen
12	(total)

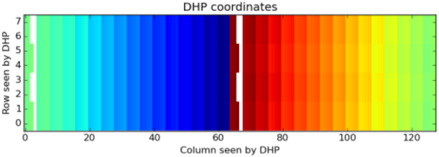
(status in VXD production database 12.10.2017)

- 33 AMC and 9 carrier to be sent to KEK for phase 3 will first be sent to DESY for PXD commissioning (testpattern and cosmic), then sent from DESY to KEK
- repair: 4+2 AMC cards, problem with flash must be fixed, no automatic bitstream booting
- repair: 1 carrier board, 1 backplane channel not working

ONSEN firmware: remapping

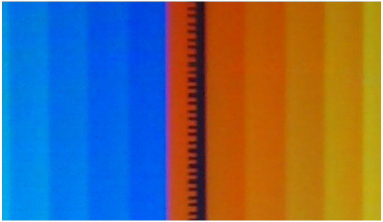
- introduced for PXD9 (1st time required in TB 04/2016)
- mirrored per 4 columns
- then mirrored per 64 columns
- 250 vs. 256 pixels
- different for PXD layer 1 and layer 2





ONSEN firmware: remapping

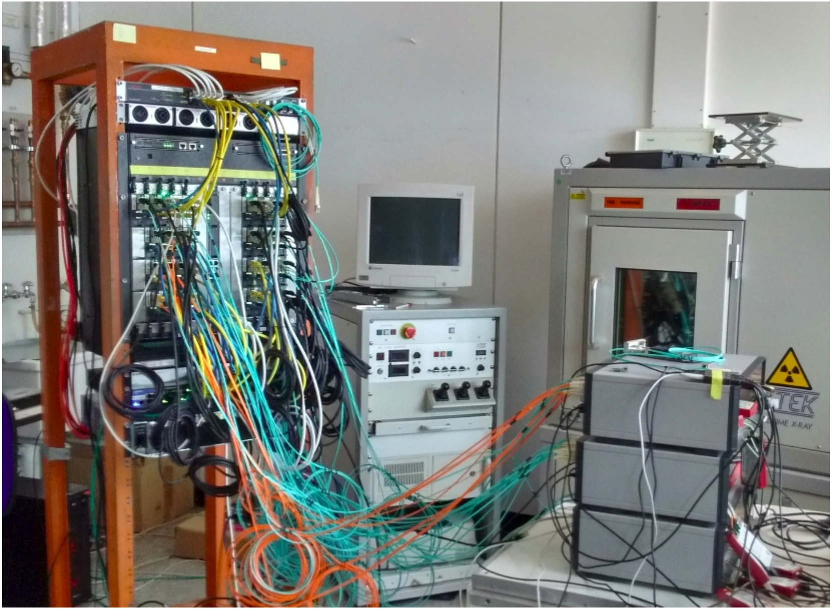
- implemented in basf2 unpacker (offline) in TB 04/2016
- implemented on Onsen (online) in TB 02/2017
  exact lookup tables on FPGA (no approximation)
  running stable in complete TB
- future: PXD online cluster finder will require remapping implemented on DHE (planned for phase 3)



There is one row alternating in DHP ID row-by-row

Full system test at Giessen

Simon Reiter



Full system test, results

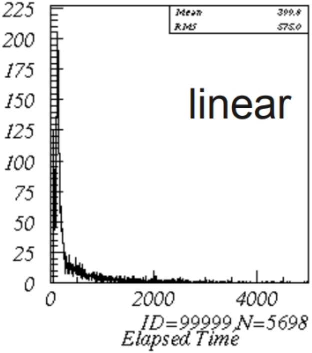
Simon Reiter

- 3 weeks testing (storing binary output data on SSD for crosscheck)
- 2 long runs over weekend
- Trigger rate $\leq$8 kHz (limited by DHC aurora line rate)
  requirement 30 kHz / 4 links/DHC = 7.5 kHz
- Data rate $\sim$595 MB/s
  540 MB/s is 3% occupancy
- Runs with HLT "send all" flag with reduced data rate of 600 Hz, send downscaled fraction of non-ROI processed (was problem in TB 2016)
  - No connection interrupts (backplane and external)
  - No buffer overflows (level $\leq$73%)
  - No framing errors, no data format errors
- Multiple start/stop without cold start
- Stable temperature in ATCA shelf ($\sim$60^o C at FPGA)

Full system test, HLT related results

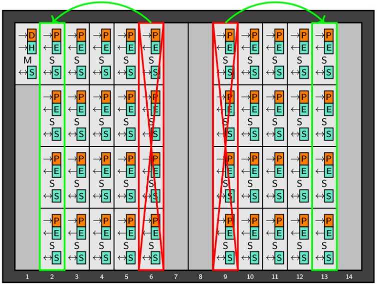
Simon Reiter

- ”send ROIs” flag in HLT data (write also ROIs into the data stream for offline check) → no error
- HLT reject trigger → no error non-triggered data are removed in ONSEN, buffer is freed
- HLT trigger unordered → no error
- HLT with fixed latency (τ=1 s) → no mismatch
- HLT latency according to Belle distribution, ~10^9 events (~8 hours, 30 kHz)
  → 7 mismatches
  → 111 “no DHC data” (but possibly HLT arrives <u>before</u> data)



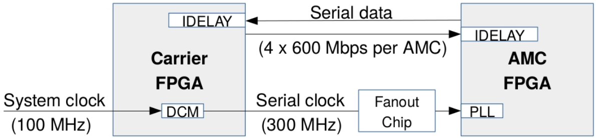
Full system test, backplane link problem

- phase 3 requires scaling of ONSEN carrier boards from 2 to 9
- problem: with merger firmware sending to multiple boards, all backplane links become unstable
- → crosstalk found between Ethernet IO and one MGT power supply (on the carrier board FPGA, not the backplane)
- solved by avoiding that link → use different ATCA slots (different FPGA pins)

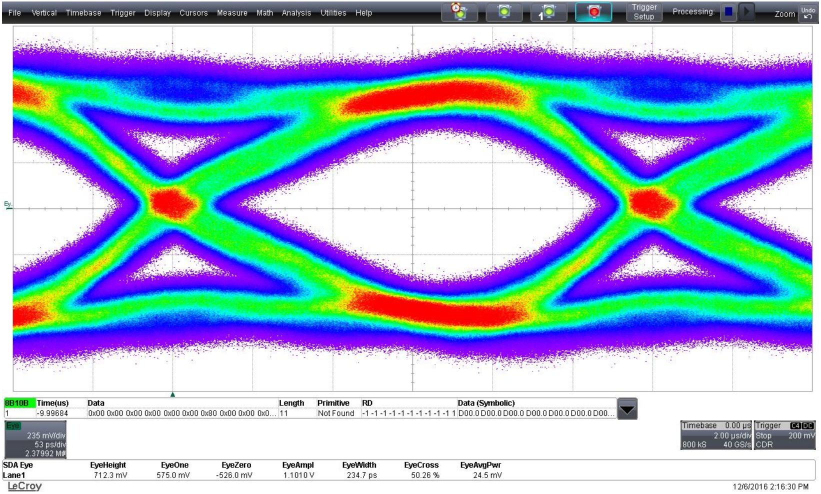


Full system test, links Between carrier and AMCs

- Connection Carrier FPGA AMC FPGA uses serial (LVDS) links
- Serial clock is distributed from Carrier to AMCs
- Clock/data phase shift is compensated by delay, determined by tuning
- Problem: strong delay difference between Carrier/AMC combinations (due to routing)
- Problem: small temperature drift of the delay
- Solution: online self-calibration mechanism
  vary delay, check if link is up or not



ATCA backplane eye diagram



8B10B	Time(us)	Data	Length	Primitive	RD	Data (Symbolic)
1	-9.99684	0x00 0x00 0x00 0x00 0x00 0x00 0x80 0x00 0x00 0x0…	11	Not Found	-1 -1 -1 -1 -1 -1 -1 -1 -1 -1 1	D00.0 D00.0 D00.0 D00.0 D00.0 D00.0 D00…

Eye
235 mV/div
53 ps/div
2.37992 M#

Timebase 0.00 µs
2.00 µs/div
800 kS 40 GS/s

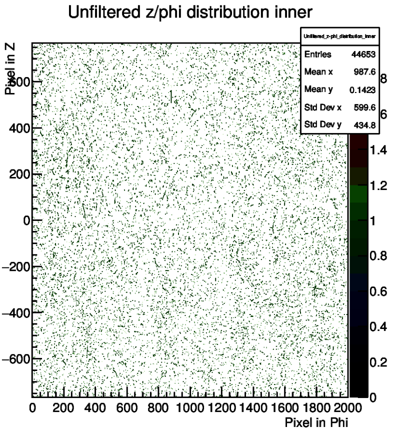
Trigger C4 DC
Stop 200 mV
CDR

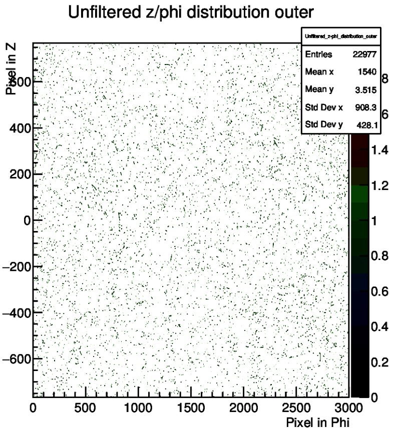
SDA Eye	EyeHeight	EyeOne	EyeZero	EyeAmpl	EyeWidth	EyeCross	EyeAvgPwr
Lane1	712.3 mV	575.0 mV	-526.0 mV	1.1010 V	234.7 ps	50.26 %	24.5 mV

LeCroy

12/6/2016 2:16:30 PM

Processing basf2 MC physics events in ONSEN



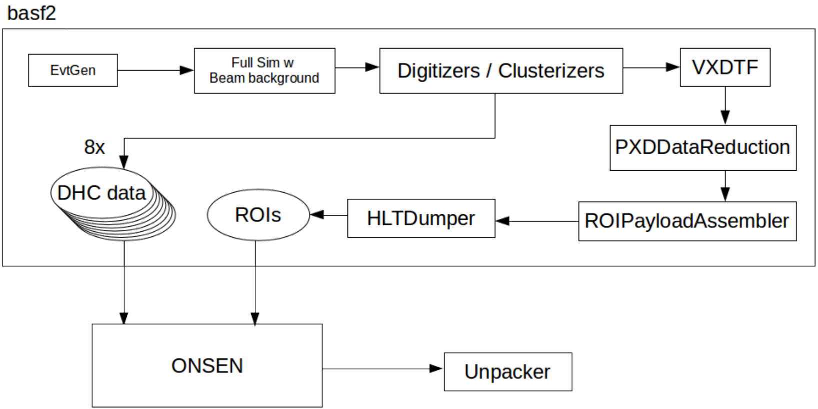


Avera

occupancy 0.8% (forward), 0.4% (backward), incl. background
BonnDAQ UDP limit 128 MB/s corresponds to 0.71%
(30 kHz)

Klemens Lautenbach

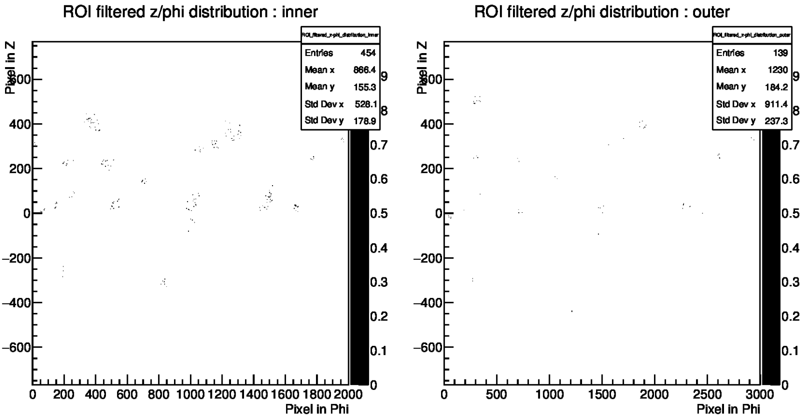
Processing basf2 MC physics events in ONSEN



Processing 5000 events (0.5 s of PXD data taking) and generate binary data required few days.
VXDTF1. Background MC8.

Klemens Lautenbach

Processing basf2 MC physics events in ONSEN



Reduc

factor 98.3 (inner), 121.6 (outer)
requirement ≥10.0 → may be released

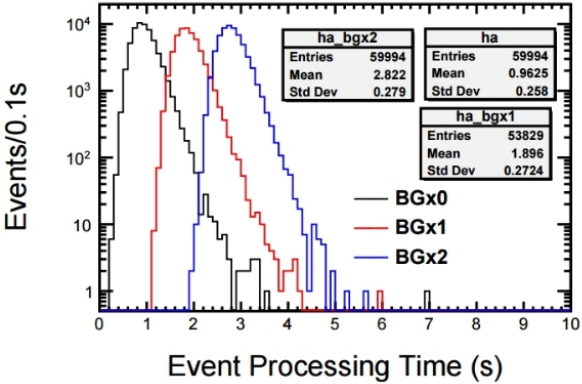
Klemens Lautenbach

BPAC Readout Integration Report, 10/2016, Question ♯1

- Line 363, 364, Section Event builder
  "The ONSEN buffering capabilities should checked against the maximum estimated fluctuations."
- HLT latency distribution from Belle ($\tau_{everage}$=1 s, τ_{max}=5 sec) confirmed by Chunhua Li (Melbourne) with MC for Belle II (see next slide)
- Full system test at Giessen
  Worst case scenario: full data rate (3% occupancy), full trigger rate (30 kHz)
  → no buffer overflows (level ≤73%)

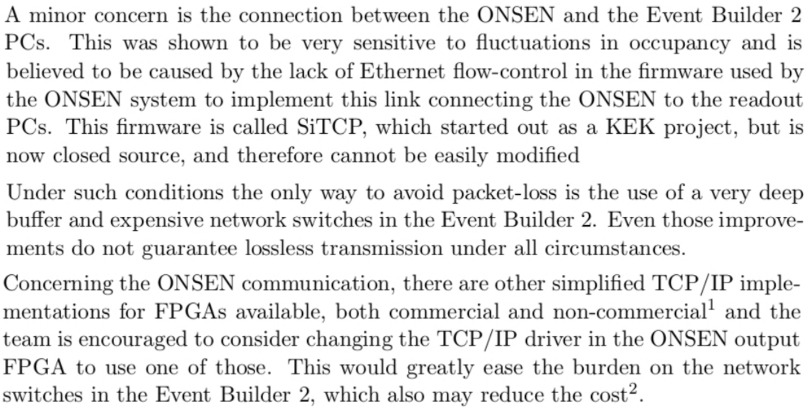
Belle II, HLT latency study

New results on HLT processing time, by Chunhua Li (Melbourne)
60.000 hadronic events ($B\bar{B}$ + udsc)
For BGx2 only 4 events with t>5 s



Chunhua Li (Melbourne)

BPAC Readout Integration Report, 10/2016, Question ♯2



A minor concern is the connection between the ONSEN and the Event Builder 2 PCs. This was shown to be very sensitive to fluctuations in occupancy and is believed to be caused by the lack of Ethernet flow-control in the firmware used by the ONSEN system to implement this link connecting the ONSEN to the readout PCs. This firmware is called SiTCP, which started out as a KEK project, but is now closed source, and therefore cannot be easily modified

Under such conditions the only way to avoid packet-loss is the use of a very deep buffer and expensive network switches in the Event Builder 2. Even those improvements do not guarantee lossless transmission under all circumstances.

Concerning the ONSEN communication, there are other simplified TCP/IP implementations for FPGAs available, both commercial and non-commercial[1] and the team is encouraged to consider changing the TCP/IP driver in the ONSEN output FPGA to use one of those. This would greatly ease the burden on the network switches in the Event Builder 2, which also may reduce the cost[2].

BPAC Readout Integration Report, 10/2016, Question ♯2

- We contacted BeeBeans Technology, and very kindly received an SiTCP version (v11.0) which should recognize PAUSE frames
- This SiTCP version is installed in the present ONSEN firmware (e.g. for phase 2)
- Not tested yet, because test non-trivial
  - provoke network congestion
  - monitor, if PAUSE frames arrive
  - monitor, if SiTCP stops sending in such a case (monitor backpressure by SiTCP in chipscope ?)
  - compare old and new version of SiTCP
- Yamagata-san provided a test program to send PAUSE frames from a PC

TB 02/2017, positive results

- final ONSEN hardware
- 2 ROI selectors parallel (2 DHCs connected)
- Onsen and DHH systems running stable for $\sim 10^9$ events
  per run up to 18 hours duration
  $\sim$1500 sroot files, 3.5 TB
  2$-$3kHz trigger rate (limited by DHC double trigger veto)
- online re-mapping (on Onsen) permanently switched on → basically permanently ROI selection in TB 04/2016 only 1 run ($\sim 10^5$ events)

TB 02/2017, negative results

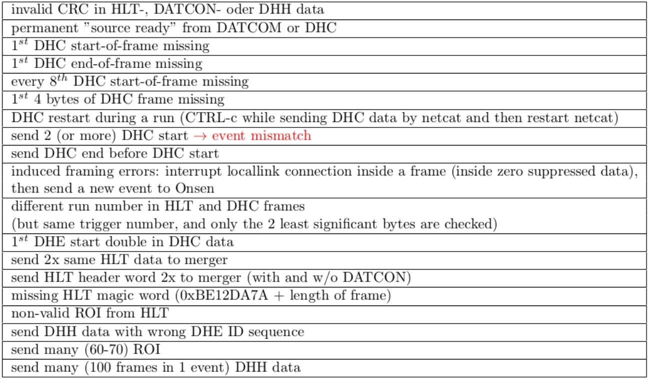
1. Onsen operation required cold restart for every run
   - re-upload FPGA bitstreams
   - otherwise trigger number mismatch
   - traced back to fragmented events from DHC, if ONSEN is reset, but DHC is not reset
     (DHC was not fully integrated in RC)
     not an ONSEN problem
2. Inconsistent states in PXD RC and global RC (READY or not-READY), in particular after Onsen cold restart
   - confusion for shift crew
   - traced back to 2 problems:
     - 2.1 software problem in global RC: updated state not interpreted in nsm-epics IOC
       not an ONSEN problem
     - 2.2 state of SiTCP connection between HLT or EB2 and ONSEN not clear ONSEN problem, but also HLT/EB2 problem

TB 02/2017, negative results

- Solutions to problem of unknown SiTCP connection status
- FIN ACK sequence implemented and tested on ONSEN
  SiTCP terminates the TCP connection correctly, if
  - run is terminated (by run control)
  - Linux (on Onsen embedded PowerPC) is shutdown
- RBCP sideband protocol
  - enables channel status monitoring
  - implemented in SiTCP (according to documentation and specification), but not tested yet
  - monitoring must be done from the receiver side (HLT or EB2), as SiTCP connection is initiated from receiver
  - agreed with DAQ group, on TODO list

ONSEN "sanitizer"

Protection of ONSEN against errors from other subsystems
Test system: copy of DESY setup with additional data fork inducing errors (intentionally) from other systems



invalid CRC in HLT-, DATCON- oder DHH data
permanent "source ready" from DATCOM or DHC
1^{st} DHC start-of-frame missing
1^{st} DHC end-of-frame missing
every 8^{th} DHC start-of-frame missing
1^{st} 4 bytes of DHC frame missing
DHC restart during a run (CTRL-c while sending DHC data by netcat and then restart netcat)
send 2 (or more) DHC start → event mismatch
send DHC end before DHC start
induced framing errors: interrupt locallink connection inside a frame (inside zero suppressed data), then send a new event to Onsen
different run number in HLT and DHC frames (but same trigger number, and only the 2 least significant bytes are checked)
1^{st} DHE start double in DHC data
send 2x same HLT data to merger
send HLT header word 2x to merger (with and w/o DATCON)
missing HLT magic word (0xBE12DA7A + length of frame)
non-valid ROI from HLT
send DHH data with wrong DHE ID sequence
send many (60-70) ROI
send many (100 frames in 1 event) DHH data

Dennis Getzkow

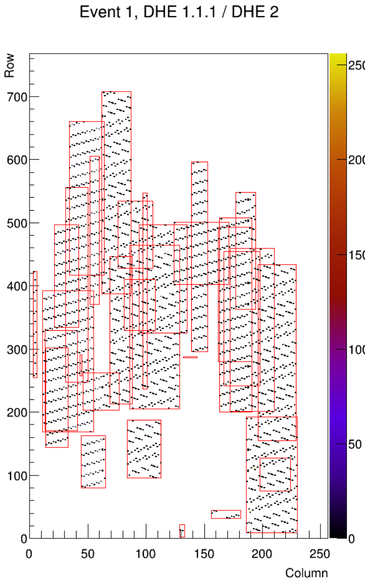
ONSEN "sanitizer"

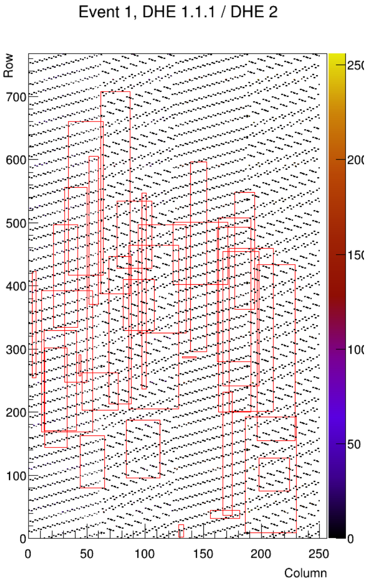
ONSEN firmware is now protected against 3 major external problems:

- invalid CRC in HLT frame
  → Onsen merger blocked any further incoming HLT data
- fragmented DHC data (cut in the middle of zero suppressed data block)
  → event fusion of 2 events (but no cold start required)
- double DHC start
  → event mismatch for all following events
  cold start required

Dennis Getzkow

Test: adding a 33^{th} ROI





ONSEN emulator

- C++-Program by S. Reiter, PXD data reduction in software
- Loads test data from file (PXD/HLT/DATCON) (requires 0xBE12DA7A header)
- Similar memory management as ONSEN
- Processing time example

  1000 events, 4% PXD occupancy = 780 MB pixel data
  ONSEN: ¡ 2 seconds after sending HLT (1 Selector node)
  Emulator: (Intel i7 @ 3.4 GHz, 16 GB RAM):
  11 min, 50 s with 1 thread (factor $\leq$355)
  2 min, 40 s with 8 threads (factor $\leq$80)

Simon Reiter

ONSEN firmware version number in bitstream

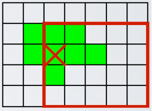
- uses 32 bits of commit-hash from the firmware git repository
- 2 files are generated: 1 bitstream, 1 linux kernel (contains epics PV definitions)
  1. hash is written into `USR_ACCESS` register (≥Virtex-5). ONSEN carrier board: reading on Virtex-4 non-trivial, only by JTAG (Impact).
  2. hash is written in bitstream at a fixed adress at the end of the block-RAM. Can be read easily from PowerPC. Version is printed on console when booting and exported into epics PV. Can be logged into database: for every run it is fixed which firmware version.
  3. hash is written into version string of Linux kernel, when compiled. Kernel ELF file is also tagged with version (in addition to bitstream).
- similar mechanism for DHH:
  - store timestamp and board number in `USR_ACCESS`
  - write the same timestamp to a git tag to identify the commit

Phase 3 preparations

- pedestal events (full frame events, in phase 2 recorded by BonnDAQ) requires FTSW-DHC communication (switch DHPT mode to memdump)
- load balancing, 5 → 4

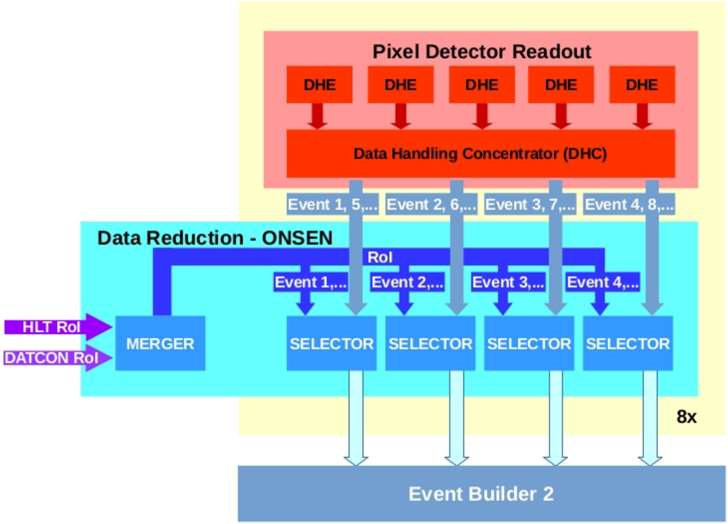
  requires RTM in DHC ATCA system

  requires ROI distribution system on ONSEN



- hit-based format → cluster-based format
  - non-trivial data format change: start-of-cluster adress requires in remapped coordinates 10 bits, but only 8 bits reserved
  - new logic in ONSEN: hit inside-cluster but outside-ROI → new cluster buffer in ROI selection
  - requires cluster finder on DHE
  - remapping must be changed from ONSEN to DHE (cluster finder needs remapped coordinates)

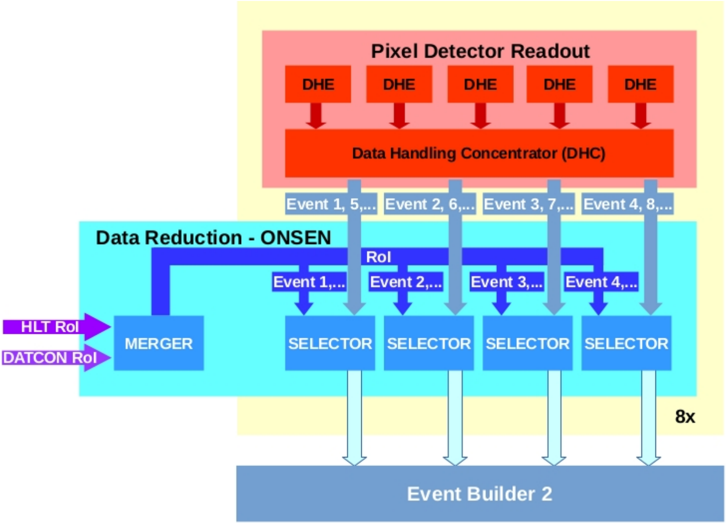
ROI distribution in phase 3



Uses additional "DHH ID filter" in front of ROI selector

(master thesis D. Getzkow)

ROI distribution in phase 2



ONSEN future development

- why?
  - almost no spares, but Virtex-4/5 at some point not available anymore
  - FPGA resources at limit
    e.g. presently no multiport memory controller for 2nd 2GB DDR2 RAM
- when? probably 2021 (planned PXD upgrade)
- new carrier board development for $\overline{\text{P}}$ANDA
  (IHEP Beijing and Univ. Giessen)
  remain compatible with existing AMC
  → Kintex Ultrascale, next slide
- upgrade link from DHC to ONSEN
  cluster-based format will increase required bandwidth by 30-50%
  (10 bit SOC adress)

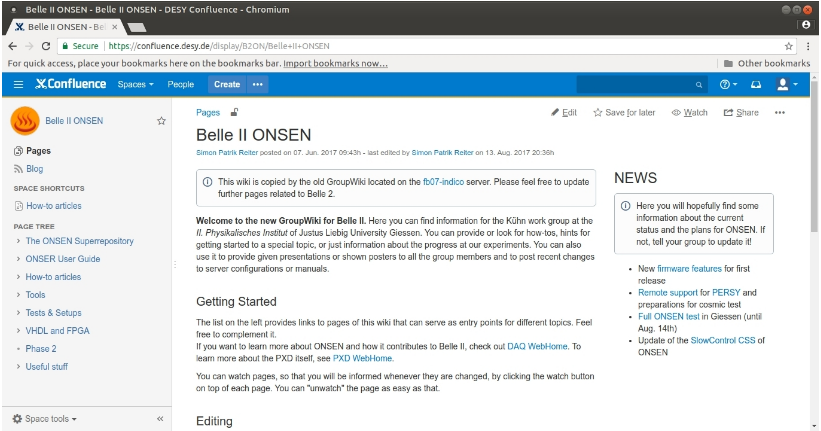
ONSEN future development

	Virtex-4 FX60 (CNCB)	Virtex-5 FX70T (xFP)	Kintex UltraScale 060 (Upgrade)
Registers	50k	44k	663k
LUTs	50k × 4-input	44k × 6-input	332k × 6-input
DSP Slices	128	128	2760
BRAM	4 Mb	5 Mb	38 Mb
MGT	16 × 6.5 Gbps	16 × 6.5 Gbps	32 × 16.3 Gbps
CPU	PPC405	PPC440	-

New physics rescue system

- CLUSTER RESCUE (high dE/dx → low p_T, no ROI)
  multilayer preceptron
  (input cluster size. cluster shape, seed charge, etc.)
  DHC or dedicated ONSEN carrier board

Belle II Onsen Confluence (wiki system)



Belle II ONSEN

Simon Patrik Reiter posted on 07. Jun. 2017 09:43h - last edited by Simon Patrik Reiter on 13. Aug. 2017 20:36h

ⓘ This wiki is copied by the old GroupWiki located on the fb07-indico server. Please feel free to update further pages related to Belle 2.

Welcome to the new GroupWiki for Belle II. Here you can find information for the Kühn work group at the *II. Physikalisches Institut* of Justus Liebig University Giessen. You can provide or look for how-tos, hints for getting started to a special topic, or just information about the progress at our experiments. You can also use it to provide given presentations or shown posters to all the group members and to post recent changes to server configurations or manuals.

Getting Started

The list on the left provides links to pages of this wiki that can serve as entry points for different topics. Feel free to complement it.
If you want to learn more about ONSEN and how it contributes to Belle II, check out DAQ WebHome. To learn more about the PXD itself, see PXD WebHome.

You can watch pages, so that you will be informed whenever they are changed, by clicking the watch button on top of each page. You can "unwatch" the page as easy as that.

Editing

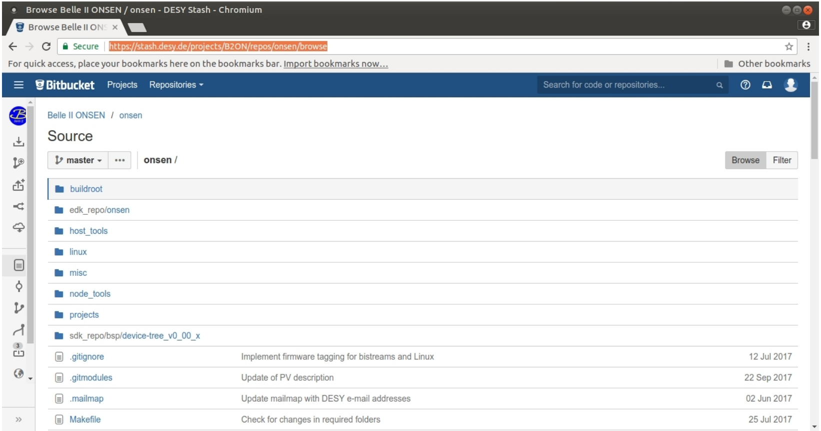
NEWS

ⓘ Here you will hopefully find some information about the current status and the plans for ONSEN. If not, tell your group to update it!

- New firmware features for first release
- Remote support for PERSY and preparations for cosmic test
- Full ONSEN test in Giessen (until Aug. 14th)
- Update of the SlowControl CSS of ONSEN

here: Onsen User Guide (not completely finished)
Onsen Ph. D. and master theses are on `https://belle2.docs.org`, googleable

Belle II Onsen Stash (git repository)


Browse Belle II ONSEN / onsen - DESY Stash - Chromium
Browse Belle II ONS ×
Secure
https://stash.desy.de/projects/B2ON/repos/onsen/browse
For quick access, place your bookmarks here on the bookmarks bar. Import bookmarks now...
Other bookmarks
Bitbucket
Projects
Repositories
Search for code or repositories...
Belle II ONSEN / onsen
Source
master
onsen /
Browse
Filter

buildroot		
edk_repo/onsen		
host_tools		
linux		
misc		
node_tools		
projects		
sdk_repo/bsp/device-tree_v0_00_x		
.gitignore	Implement firmware tagging for bistreams and Linux	12 Jul 2017
.gitmodules	Update of PV description	22 Sep 2017
.mailmap	Update mailmap with DESY e-mail addresses	02 Jun 2017
Makefile	Check for changes in required folders	25 Jul 2017

https://stash.desy.de/projects/B2ON/repos/onsen/browse

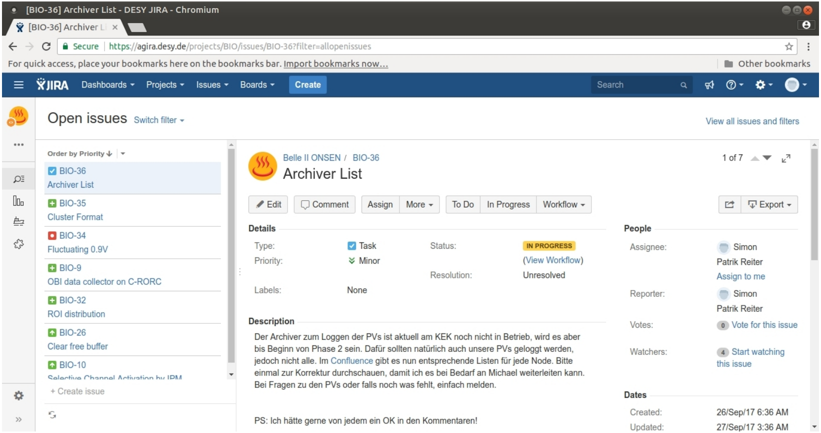
automatic bitstream build (Xilinx planAhead installed on DESY servers)

before phase 2: "release" (only event filter is missing)

"super onsen" git clone → checkout everything

firmware version encoded in bitstream.

Belle II Onsen JIRA (issue management system)


[BIO-36] Archiver List - DESY JIRA - Chromium
[BIO-36] Archiver L
Secure
https://agira.desy.de/projects/BIO/issues/BIO-36?filter=allopenissues
For quick access, place your bookmarks here on the bookmarks bar. Import bookmarks now...
Other bookmarks
JIRA
Dashboards
Projects
Issues
Boards
Create
Search
Open issues
Switch filter
View all issues and filters
Order by Priority
Edit
Comment
Assign
More
To Do
In Progress
Workflow
Export
+ Create issue
1 of 7

☑ BIO-36	Archiver List
BIO-35	Cluster Format
BIO-34	Fluctuating 0.9V
BIO-9	OBI data collector on C-RORC
BIO-32	ROI distribution
BIO-26	Clear free buffer
BIO-10	Selective Channel Activation by IPM

Belle II ONSEN / BIO-36

Archiver List

Details

Type: ☑ Task

Priority: Minor

Status: IN PROGRESS (View Workflow)

Resolution: Unresolved

Labels: None

Description

Der Archiver zum Loggen der PVs ist aktuell am KEK noch nicht in Betrieb, wird es aber bis Beginn von Phase 2 sein. Dafür sollten natürlich auch unsere PVs geloggt werden, jedoch nicht alle. Im Confluence gibt es nun entsprechende Listen für jede Node. Bitte einmal zur Korrektur durchschauen, damit ich es bei Bedarf an Michael weiterleiten kann. Bei Fragen zu den PVs oder falls noch was fehlt, einfach melden.

PS: Ich hätte gerne von jedem ein OK in den Kommentaren!

People

Assignee: Simon Patrik Reiter — Assign to me

Reporter: Simon Patrik Reiter

Votes: 0 Vote for this issue

Watchers: 4 Start watching this issue

Dates

Created: 26/Sep/17 6:36 AM

Updated: 27/Sep/17 3:36 AM

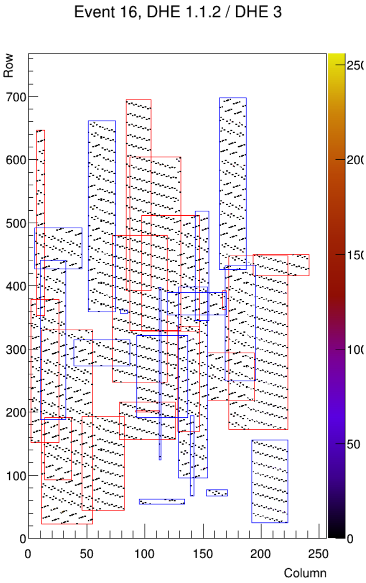
BACKUP

BPAC Readout Integration Report, 10/2016, Question ♯2

If new SiTCP version does not recognize PAUSE frames

- Problem is non-fatal
  - Communication is lossless, as siTCP includes retransmission
  - The problem is nonfatal: worst case in case of switch overload, reminder: there is 4 GB buffer on Onsen. If Onsen buffer full, back-pressure BUSY is issued (stop triggers), but there is no abort condition or data drop
- Other solutions?
  - CMS solution is only for sending, not receiving, but we need to receive HLT data
  - Advantage of siTCP: light weight, FPGA resources 15-20%, more complex protocol would require (non-available) resources
  - Long-term solution: use TCP on a PC with PCIe cards, input 32 optical links, output 10G uplink to event builder prototype existing and tested at Giessen (ALICE C-RORC)

Test: HLT and DATCON ROIs.


Event 16, DHE 1.1.2 / DHE 3
Row
Column