Trigger/DAQ design: from test beam to medium size experiments

ISOTDAQ 2019 Royal Holloway, University of London 6 April 2019

credit to Sergio Ballestrero most material from his talk at ISOTDAQ 2015



Trying to move ...

from here:



Basic DAQ: De-randomization





Friday, 5 February 2010

to here:

Trigger & DAQ in HEP



different issues \rightarrow different solutions no magic, unique solution for all cases

medium/large DAQ: constituents



aiming to cover the gap in 5 steps ...

- Step 1: Increasing the rate
- Step 2: Increasing the sensors
- Step 3: Multiple Front-Ends
- Step 4: Multi-level Trigger
- Step 5: Data-Flow control

Do we really need a trigger ?

Do we really need a trigger ?

not obvious ... triggerless DAQ systems do exist

Do we really need a trigger ?

not obvious ... triggerless DAQ systems do exist

even in high-energy physics

Do we really need a trigger ?

not obvious ... triggerless DAQ systems do exist

even in high-energy physics, e.g.:

a) LHCb upgrade 40 MHz readout → briefly described in Francesca's final talk

b) DUNE LAr TPC 2 MHz readout \rightarrow *Giovanna's talk on April 11*

but triggering may be crucial ...

https://en.wikipedia.org/wiki/Coincidence_circuit :

Walther Bothe (1924-1929): offline \rightarrow online coincidence (logic **AND**) of 2 signals

Bruno Rossi (Nature, 1930): "Method of Registering Multiple Simultaneous Impulses of Several Geiger Counters" → online coincidence of 3 signals (scalable)!

first modern trigger

https://en.wikipedia.org/wiki/Coincidence circuit :

"Rossi coincidence circuit was rapidly adopted by experimenters around the world. It was the first practical AND circuit, precursor of the AND logic circuits of electronic computers"



12

but even simple trigger systems ...



Fig. 1. Setup of the Na59 Experiment

... pose issues !

... pose issues !



(anti-)coincidence with veto \rightarrow simple, clear !

... pose issues !



(anti-)coincidence with veto \rightarrow simple, clear !

really doing what you think/need ?

(anti-)coincidence with veto



output signal may: a) jitter b) fluctuate in duration (or both)

because of relative timing of T1, T2, V

(anti-)coincidence with veto



much better !

first lesson(s)

trigger signal:

1) should be formed \rightarrow pulse with predefined duration

2) veto/busy should block pulse generation

step one: increase rate

Many issues:

- \rightarrow trigger latency
- \rightarrow readout latency
- \rightarrow throughput
- \rightarrow rate fluctuations (trigger bursts)
- \rightarrow throughput fluctuations

(correlated noise, ...)

step one: increase rate

Many issues:

- \rightarrow trigger latency
- \rightarrow readout latency
- \rightarrow throughput
- \rightarrow rate fluctuations (trigger bursts)
- \rightarrow throughput fluctuations

(correlated noise, ...)

$$\rightarrow$$
 dead-time

dead time (from Andrea's introduction)



dead time \rightarrow de-randomise

• Processing \rightarrow bottleneck ?



(f $\! \cdot \tau) \sim 1 \! \rightarrow \! dead$ time $\sim 50\%$

• Buffering \rightarrow decouple problems



What the impact ? (f $\cdot \tau$) ~ 1 \rightarrow dead time ?

buffering solve all problems ?

• FIFO

filling at very variable input flow
 emptying at smoothed output flow
 → the Leaky-Bucket problem

Q: how often may overflow ?



off-topic: some very basic queueing theory

• N-event buffer ... single queue size N:

 $P_k: \%$ time with k events in $\Rightarrow P_N = no space available \rightarrow dead time$

 $\sum P_{k} = 1 [k=0..N]$ rate($j \rightarrow j+1$) = f·P_i (fill at rate f) $rate(j+1 \rightarrow j) = v \cdot P_{_{i+1}} \quad (empty \text{ at rate } v \gtrsim f)$ steady state: $v \cdot P_{i+1} = f \cdot P_i \Rightarrow P_{i+1} = \rho \cdot P_i = \rho^{j+1} \cdot P_n$ where $\rho = (f/v) \leq 1$ for $\rho \sim 1 \Rightarrow P_{j} \sim P_{j+1} \Rightarrow \Sigma P_{k} \sim (N+1) \cdot P_{0} = 1 \Rightarrow \left(P_{0} \sim P_{N} \sim 1/(N+1) \right)$ \Rightarrow dead time $\sim 1/(N+1)$ want d.t. $\leq 1\% \Rightarrow N \geq 100$

off-topic: some very basic queueing theory



de-randomisation





• We can now attain a DAQ efficiency $\sim 100\%$ with:

 $-\tau \sim 1/f$

- "moderate" buffer size
- One more degree of freedom to play with
- Often managed by trigger system itself ("complex dead time")

complex dead time

1) simple dead time: avoid overlapping (conflicting) readout window

2) complex dead time: avoid overflow in front-end buffers (protection against trigger bursts)

e.g. ATLAS: simple leaky-bucket algorithms with 2 parameters:

max X triggers (X = FIFO depth) in any (sliding) time window = (X*readout time)



FIFO

First-In First-Out memory:

independent read/write (sequential) access
 may be hardware or over RAM

better with Dual-Port RAM

game over ?

many other possible limits even in a simple DAQ



\rightarrow sensor

- Sensors limited by physical processes such as:
 - drift times in gases
 - charge collection in Si
- (possibly) choose fast processes
- analog FE imposes limits as well
- split sensors, each gets less rate: "increase granularity"



$\rightarrow \mathsf{ADC}$

- A/D conversion also limited
- Fast ADC
 - \rightarrow # of bits (resolution)
 - \rightarrow power consumption
- Alternatives:

analog buffers

(e.g. switched capacitor arrays)

 You may need integration (or sampling) over quite some time



an example

- HPGe + Nal Scintillator High res spectroscopy and beta+ decay identification
- minimal trigger with busy logic
- Peak ADC with buffering, zero suppression
- VME SBC with local storage
- 3×12 bits data size
 + 32 bit timestamp
- Root for monitor & storage
- Rate limit ~14 kHz
 - HPGe signal shaping for charge collection
 - PADC conversion time

Ge crystal for isotope identification





\rightarrow trigger latency

- simple trigger: ~fast
- complex trigger logic: not obvious [even when all in hw]
- some trigger detectors may be far away / slow → latency
- trigger signal is one: all information at a single point
 - in one step: too many cables
 - in many steps:
 delays



 \rightarrow discrete modules: ~ 5-10 ns delay \rightarrow tot. latency \geq 20-30 ns \leftarrow

DREAM (2006 \rightarrow): a testbeam case

R&D on dual-readout calorimetry, setup:

- Crystals
- •Scintillating/cherenkov fibers in lead/copper matrices
- •Scintillator arrays as shower leakage counters
- Trigger/veto/muon counters
- Precision chamber hodoscope
- ... always evolving

Acquiring: waveforms, total charge, time information

DREAM (2006 \rightarrow): a testbeam case


readout system

1 PC \rightarrow readout of 2 VME crates (via CAEN optical interfaces) 1 PC \rightarrow storage

6 x 32 ch xDC.s (x = Q, T : CAEN V792, V862, V775)

1 x 34 ch (CAEN V1742) 5 Gs/s Digitizer (single event: ~34x1024x12bit)

1 x 4 ch Tektronix TDS7254B 20 Gs/s oscilloscope

... few VME I/O & discriminator boards

... all in the control room

dataflow

 Pull mode → FE electronics waiting for PC readout (self-blocking trigger, re-enabled after readout)

2) Block data transfer \rightarrow DMA (Direct Memory Access) data moved by specialised hw (not by CPU)

[Push mode \rightarrow FE electronics sending data as soon as available]

DAQ

DAQ logic spill-driven (no "real time", SLC desktops)

in-spill (slow extraction)

poll trigger signal ... if trigger present:

a) (block) read all VME boards

b) format & store on large buffers (FIFO over RAM)

c) re-enable trigger

out-of-spill

a) read scope (in case) → event size fixed at run start
b.1) flush buffers to disk (beam and pedestal files) over network
b.2) monitor data (produce root files)

rate \sim O(1 kHz)

spill-driven (asynchronous) trigger





a) crystals w/ fast PMT.sb) no analog buffering

\rightarrow low-latency trigger

first discrete, then FPGA (Xilinx Spartan 3AN evaluation board)

step two: increase # of sensors

- More granularity at the physical level
- Multiple channels (usually with FIFOs)
- Single, all-HW trigger
- Single processing unit
- Single I/O



multi-channel, single-PU system



- common architecture in test beams and small experiments
- often rate limited by (interesting) physics itself, not TDAQ system
- or by the sensors

bottlenecks: PU and storage



- a single PU can be a limit
 - collect / reformat / compress data can be heavy
 - simultaneously writing storage
- final storage too:
 - −VME up to 50MB/s
 → 1TB in 6h
 too many disks in a week!

Laptop SATA disk: \sim 100 MB/s USB2: \sim 60 MB/s

\rightarrow decouple storage from PU



- data transfer data → dedicated "Data Collection" unit to format, compress and store
- more room for smarter processing or decreased dead time on non-buffered ADCs

bottlenecks: trigger



- to reduce data rates (to avoid storage issues)
 → non-trivial trigger
- complexity may already hit manageability limits for discrete logic (latency!)
- integrated, programmable logic came to rescue (FPGA)

 \rightarrow latency may go down to O(few ns)

another example: NA43/63



- Radiation processes: coherent emission in crystals and structured targets, LPM suppression...
- 80/120 GeV e- from CERN SPS slow extraction
- 2s spill every 13.5s

- Needs very high angular resolution
- Long baseline + high-res, low material detectors
 → drift Chambers
- 10 kHz limit on beam for radiation damage

 \rightarrow 2-3 kHz physics trigger

NA43/63

- 30-40 TDC, 6-16 QDC, 0-2 PADC (depending on measurement)
- CAMAC bus
 1 MB/s, no buffers, no Z.S.
- single PC readout
- NIM logic trigger (FPGA since 2009)
 - pileup rejection
 - fixed deadtime



step three: multiple PUs (SBC)

- e.g.: CERN LEP experiments
- complex detectors, moderate trigger rate, very little background
- little pileup, limited channel occupancy
- simpler, slow gas-based main trackers



NOMAD (1995-1998)

- Search for $v_{\mu} \rightarrow v_{\tau}$ oscillations at the CERN WA neutrino facility (WANF)
- 2.4×2.4 m² fiducial (beam) area
- Two 4ms spills with 1.8×10¹³ P.o.T. each (v spills)
- One (2s) slow-extraction spill (µ spill)
- 14.4s cycle duration





 \rightarrow DAQ layout

WANF - SPS SuperCycle

14.4 s cycle length

 2×4 ms neutrino spills (f/s extractions)

 1×2 s muon spill (slow extraction)

f/s extractions





triggering once more ...

menu for NOMADs:



$triggering \rightarrow FPGAs \ at \ work$

MOdular TRIgger for NOmad (MOTRINO): 6 VME boards providing local and global trigger generation and propagation



DAQ

- FASTBUS digitisers:
 - \sim 200 (either 64 or 96 channel) xDC boards [x=Q,P,T]
 - $O(\geq 2 \text{ us})$ conversion time, 256 event buffers
- VME readout and processing:
 - Motorola 68040 FIC8234 (OS9 real-time system) VME PUs
 - 5 for readout + 1 for event building
- Typically
 - ~4 kHz of neutrino triggers (~15 evts in each 4ms spill)
 - ~30 Hz of muon triggers (~60 evts in each 2s spill)
 - 256-events in off-spill calibration cycles (calibration triggers)

readout sequence

- On-spill on-board buffering
- Off-spill (i.e. off-beam) data transfer and processing
 - on spill (or calibration cycle): on-board event buffering (no way to read event by event)
 - end of spill (or calibration cycle): block transfer to VME
 - then event building + storage
- monitoring and control on SunOs and Solaris workstations

 \rightarrow dead time in v spills: $\sim\!10\%$ due to digitisation

more bottlenecks ?



- trigger complexity ↔ storage
- single HW trigger not sufficient to reduce rate
- add L2 Trigger
- add HLT

step four: multi-level trigger



Typical Trigger / DAQ structure at LEP

more complex filters

 \rightarrow slower

see Trigger lectures

 \rightarrow applied later in the chain

LEP

- 10⁵ channels
- 22µs crossing rate
 –no event overlap
- single interaction
- L1 ~10³ Hz
- L2 ~10² Hz
- L3 ~10¹ Hz
- 100kB/ev \rightarrow 1MB/s



```
High-lumi pp collisions @ CERN SppS:

\sqrt{s} = 630 \text{ GeV}

L = 5 x 10<sup>30</sup> cm<sup>-2</sup> s<sup>-1</sup> (one order of magnitude increase)
```

Goal:

W/Z physics
 QCD
 top quark and SUSY particle discovery
 → robust theoretical prediction for this

Complex trigger signatures: em, jet and missing P_{τ}

Three-level trigger selection:

- L1 from on detector hardware
- L2 over a dedicated processor
- L3 over FASTBUS processors (ALEPH event builder)

DAQ readout & monitoring: CAMAC & FASTBUS \rightarrow VAX/VMS platforms

Three-level trigger selection: L1 from on detector hardware L2 over a dedicated processor L3 over FASTBUS processors (ALEPH event builder)

DAQ readout & monitoring: CAMAC & FASTBUS \rightarrow VAX/VMS platforms

 \rightarrow theorist were right but ...

Three-level trigger selection: L1 from on detector hardware L2 over a dedicated processor L3 over FASTBUS processors (ALEPH event builder)

DAQ readout & monitoring: CAMAC & FASTBUS \rightarrow VAX/VMS platforms

→ theorist were right but ... unfortunately nature was wrong!

Three-level trigger selection: L1 from on detector hardware L2 over a dedicated processor L3 over FASTBUS processors (ALEPH event builder)

DAQ readout & monitoring: CAMAC & FASTBUS \rightarrow VAX/VMS platforms

 \rightarrow theorist were right but ... unfortunately nature was wrong!

Nevertheless many new/better measurements and observations of SM processes

ATLAS (from run-1 to run-2)



\rightarrow Merge L2 and L3 into a single HLT farm

- preserve Region of Interest but dilute the farm separation and fragmentation
- increase flexibly, computing power efficiency

off-topic: event-selection latency

Typical (ATLAS) numbers:

- L1 : O(1 μ s in real-time) \rightarrow let say = 2.5 μ s
- L2 : O(10 ms) \rightarrow let say = 40 ms
- L3(HLT) : O(s) \rightarrow let say = 1 s

Q: do the 3 numbers mean the same thing ?

latency and real-time

real time: system must respond within some fixed delay → Latency = Max Latency → over fluctuations bad, will create dead time

non-real-time: system responds as soon as it's available → *Latency* = *Mean Latency*

 \rightarrow over fluctuations fine, shouldn't create dead time

real time o.s. :

very stable time delay in responding to events

standard unix kernels are not real time: a system call can in principle take any time

off-topic: real-time linux



April 6, 2019

Interruptible linux kernel

higher priority application

step five: dataflow control



- Buffers are not the <final solution> they can overflow due to:
 - bursts
 - unusual event sizes
- Discard
 - local, or
 - "backpressure", tells lower levels to discard

Who controls the flow? FE (*push*) or EB (*pull*)

a push example: KLOE

- DAΦNE e⁺e⁻ collider in Frascati
- CP violation parameters in the Kaon system
- "factory": rare events in a highrate beam



- 10⁵ channels
- 2.7 ns crossing rate
 - rarely event overlap
 - "double hit" rejection
- high rate of small events
- L1 ~10⁴ Hz
 - 2µs fixed dead time
- HLT ~10⁴ Hz
 - ~COTS, cosmic rejection only
- $5 \text{ kB/ev} \rightarrow 50 \text{ MB/s}$ [design]

KLOE



- deterministic FDDI network
- not real need for buffering at FE
- *push* architecture vs pull used in ATLAS *see DAQ Software lecture*
- try EB load redistribution before resorting to backpressure

Which LHC experiment has a somewhat similar dataflow architecture ?

LHCb: network is dataflow



more info in "TDAQ for the LHC experiments"

looking forward to LS2 and beyond

On some long term, all experiments looking forward to significant increase in L1 trigger rate and bandwidth. ALICE and LHCb will pioneer this path during LS2

DAQ@LHC Workshop



- First level trigger for Pb-Pb interactions 500 Hz → 50 kHz
- 22 MB/event
 - 1 TB/s readout \rightarrow 500 PB/month
- Data volume reduction
 - on-line full reconstruction
 - discard raw-data
- Combined DAQ/HLT/offline farm
 - COTS, FPGA and GPGPU

LHCb

- **1 MHz** \rightarrow **40 MHz** readout and event building \rightarrow trigger-less
 - trigger support for staged computing power deployment
- 100 kB/event
 - on-detector zero suppression \rightarrow rad-hard FPGA
 - 4 TB/s event-building
trends



- Integrate synchronous, low latency in front end
 - limitations do not disappear, but decouple (factorise)
 - all-HW implementation
 - isolated in replaceable(?)
 components
- Use networks as soon as possible



- Deal with dataflow instead of latency
- Use COTS network and processing
- Use "network" design already at small scale
 - easily get high performance with commercial components

take care, lot of issues not covered:

Hw configuration Sw configuration Hw control & recovery Sw control & recovery Monitoring

. . .

Thank you for your patience ...

Lost & Found (off-topics)

Appendix A: Cables and Transmission Lines

Spoken about signals, amp.s, digitisers, ... but ...

... almost nothing about how signals are transmitted over long distances. *Is there any issue ?*

Ok the full line must be properly matched:

Z(out) = Z(cable) = Z(in)

<u>That's all ?</u>

Cables and Transmission Lines

Lossless transmission line:



Lossy transmission line:



Cables

Cable element (dz):





R depends on the frequency (skin effect) G should be negligible

Z =
$$(L/C)^{\frac{1}{2}}$$

v_p = $(LC)^{-\frac{1}{2}}$ = $(\mu\epsilon)^{-\frac{1}{2}}$

Cables

Equation for standing waves:

$$\frac{\partial^2 V}{\partial z^2} = LC \frac{\partial^2 V}{\partial t^2} + (LG + RC) \frac{\partial V}{\partial t} + RGV$$

solution:

$$\frac{d^{2}V}{dz^{2}} = (R + i\omega L)(G + i\omega C)V = \gamma^{2}V$$
$$\gamma = \alpha + ik = \sqrt{(R + i\omega L)(G + i\omega C)}$$

R usually dominated by the skin effect: $R(\omega) = r^*D/(4^*\delta)$

 $\label{eq:r} \begin{array}{l} r = resistance \ per \ unit \ length \\ D = diameter \ internal \ conductor \\ \delta = skin \ depth \ \sim \ 1/\sqrt{\omega} \end{array}$

Cable Losses

Neglecting the transconductance G:

$$\alpha = R(\omega)/(2Z_0) \sim c \sqrt{\omega}$$
$$k = \omega \sqrt{RC} = \omega/(\beta c)$$
$$V(z,t) = V_1 \exp(-\alpha z) \exp[i(\omega t - kz)]$$

50-Ohm fast (v = 4 ns/m) CERN-store cables:

04.61.11.F - COAXIAL CABLE 50 OHM - TYPE C-50-6-1 04.61.11.H - COAXIAL CABLE 50 OHM - LOW LOSS - TYPE C-50-11-1

f(-3db, 40 m, cable C-50-6-1) ~ 120 MHz f(-3dB, 40 m, low loss cable) ~ 640 MHz

Signal Distortions

Time parameter:

$$\alpha \sim \mu \sqrt{f} \tau_0 = (\mu z)^2 / \pi$$

µz ~ 32*E-6 (C-50-6-1), 14E-6 (low loss cables)

$$\tau_0 \sim 320 \, ns \, (C - 50 - 6 - 1)$$

$$\tau_0 \sim 60 \, ns \, (low loss cables)$$

*** Take care: would like $\tau_0 \ll \tau$ (signal)

Digital Pulse Distortions



Bandwidth Effects – Analog Signals



Appendix B: backtrace

Segfaulting ? Have a look at backtrace:

https://www.gnu.org/software/libc/manual/html_node/Backtraces.html

BACKTRACE(3) BACKTRACE(3) Linux Programmer's Manual

NAME

backtrace, backtrace_symbols, backtrace_symbols_fd - support for application self-debugging

SYNOPSIS

#include <execinfo.h>

int backtrace(void **buffer, int size);

char **backtrace_symbols(void *const *buffer, int size);

void backtrace_symbols_fd(void *const *buffer, int size, int fd);

HowTo

1) file "my_segf.cxx" : install a signal handler to print the backtrace

```
#include <stdio.h>
#include <execinfo.h>
#include <signal.h>
#include <stdlib.h>
#include <unistd.h>
```

```
void handler(int sig) {
   void *array[10];
   size_t size;
```

```
// get void*'s for all entries on the stack
size = backtrace(array, 10);
```

```
// print out all the frames to stderr
fprintf(stderr, "Error: signal %d:\n", sig);
backtrace_symbols_fd(array, size, STDERR_FILENO);
exit(1);
```

```
void baz() {
  int *foo = (int*)-1; // make a bad pointer
  printf("%d\n", *foo); // causes segfault
  }
void bar() { baz(); }
void foo() { bar(); }
int main(int argc, char **argv) {
  signal(SIGSEGV, handler); // install our handler
  foo(); // this will call foo, bar, and baz. Baz segfaults.
}
```

}

2) compile with -g debug flag on:

```
g++ -g -rdynamic my_segf.cxx -o
my_segf
```

3) get the crash:



6) Take the Answer: baz() \rightarrow crash is at (baz+0x14)

7) crash is at (baz+0x14) ... open the debugger: gdb my_segf

(gdb) info address baz Symbol "baz()" is a function at address 0x400a55.

8) so crash is at address (0x499a55+0x14) ... then:

(gdb) info line *(0x400a55+0x14) Line 24 of "my_segf.cxx" starts at address 0x400a65 <baz()+16> and ends at 0x400a7c <baz()+39>.

9) got it ! That's not yet the reason but ...

Appendix C: Profiling

Take care: optimize your code – first of all - where it really needs. To get it, you may use of profiling.

for C/C++ code, look (for example) at this gprof tutorial: http://www.thegeekstuff.com/2012/08/gprof-tutorial/

Very simple, at least for standalone code ...