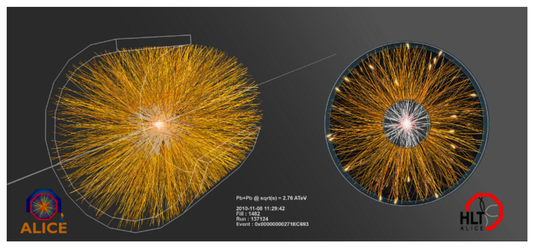
Data compression in ALICE by on-line track reconstruction and space point analysis

Matthias Richter[1] for the ALICE collaboration

[1]Department of Physics, University of Oslo, Norway
Matthias.Richter@fys.uio.no

Introduction

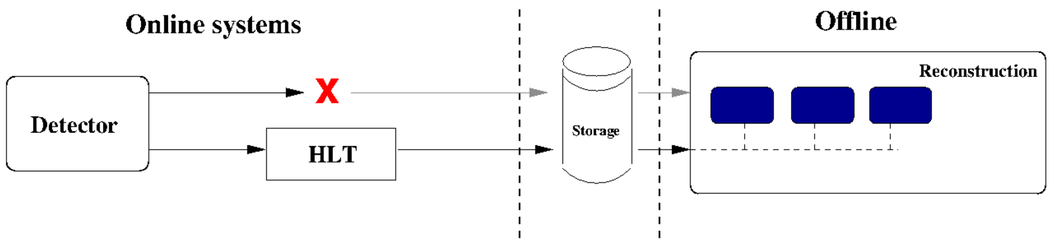
High resolution detectors in high energy nuclear physics deliver a huge amount of data which is often a challenge for the data acquisition and mass storage.
Lossless compression techniques on the level of the raw data can provide compression ratios up to a factor of 2. In ALICE [1], a data compression technique has been developed for the Time Projection Chamber (TPC) [2] to reach an overall compression factor suited for data taking in Heavy Ion collisions. Studies on simulated TPC data have earlier shown the potential for data compression [3, 4]. The application on real data is presented for the first time. With a combination of on-line cluster reconstruction from raw data and a subsequent lossless data compression, the data volume can be reduced by a factor 4 to 6 depending on the data sample.



One of the first recorded Pb-Pb events in the ALICE TPC in 2010

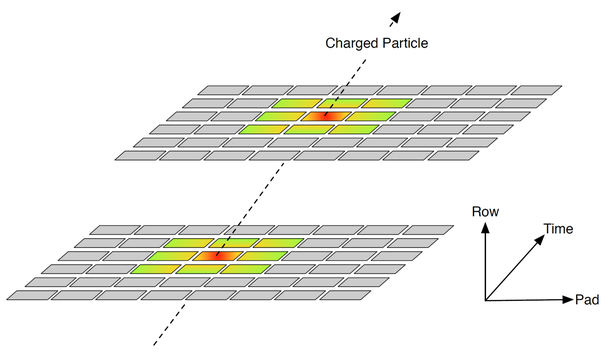
Data compression scheme

Approach: Replacing detector raw data by pre-processed data from the High Level Trigger (HLT)



→ HLT data can enter offline detector reconstruction at different stages

Model for TPC raw data - Clusterization



Time Projection Chamber provides simultaneous measurement of three-dimensional track information:

- 2D readout of segmented pad planes
- Sampled drift time of induced charges

Ionizations form clusters, which are calculated from the raw data as first step in the reconstruction.

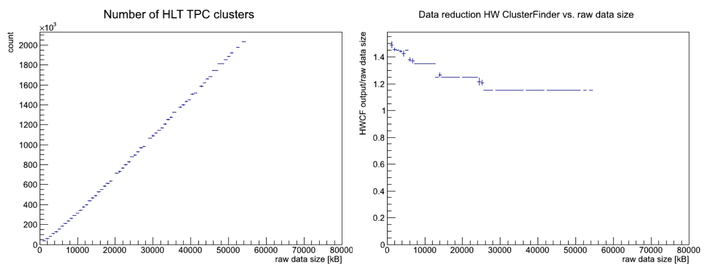
normal clusters: distributed over several readout pads fulfill required precision for tracking

single pad clusters: deteriorated pad resolution, not used in tracking but useful charge information for PID

In the HLT the cluster reconstruction is performed in an FPGA co-processor in real-time [5, 6].

Data reduction in clusterization

The reconstruction of clusters from raw data is a lossy transformation, the original raw data can not be restored. However, the loss in the required information for the event reconstruction is negligible for the physics analysis. A small data reduction factor is achieved in this step.



⇒ data reduction by factor 1.1 to 1.5 depending on the event size

Raw cluster format

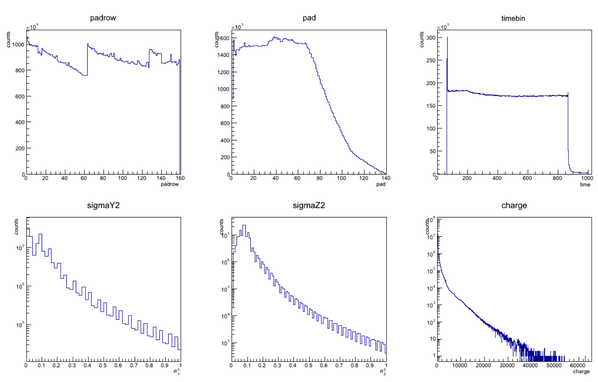
The reconstructed clusters are stored in raw coordinates which makes the recorded compressed data independent of drift time calibration and further corrections.

Clusters are characterized by seven parameters:

	HLT raw cluster format		compressed format	
padrow number	unsigned short	16	local padrow in partition	6
pad position	float	32	pitch 0.4/0.6 cm → 1/60	14
timebin	float	32	0.25cm/timebin → 1/25	15
sigmaY2	float	32		8
sigmaZ2	float	32		8
total charge	short	16		10
max charge	short	16		16
		176		77
		22 Byte		10 Byte

The size for each parameter in the raw cluster format and a fixed point format of sufficient precision has a big impact of the achievable compression ratio and is indicated in the table.
The figure shows the distribution for six of the parameters, the maximum charge is similar to charge.

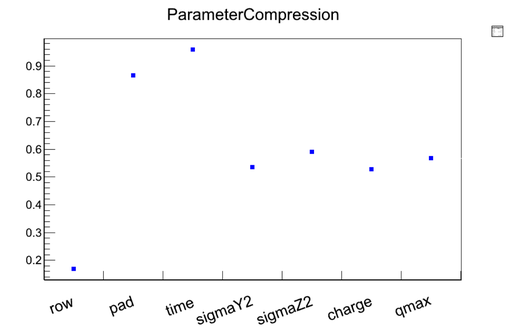
Note the logarithmic scale in the lower panels.



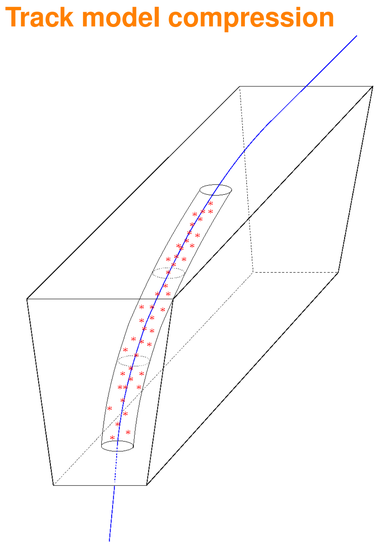
Data Deflation - Lossless data compression

- Store data in a bit stream → prerequisite for variable value length
- Data deflater interface between data structure and data buffer
- Used in 2011: Huffman implementation of DataDeflater

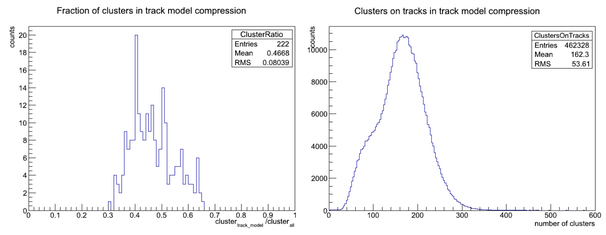
- row numbers stored differentially for sorted clusters in ascending row number → effectively 1 bit
- Huffman coding works best for sigma, charge and max charge
- Other parameters can be converted to format suited for Huffman coding



Track model compression

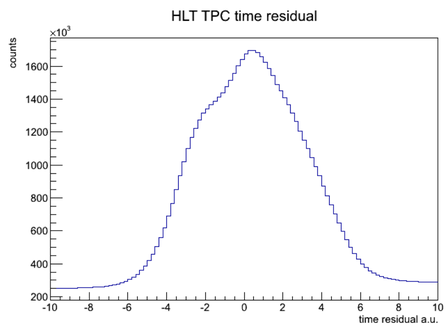


∼ 50% of all clusters can be associated with tracks:
→ assigned by tracking
→ unassigned clusters associated to nearest track



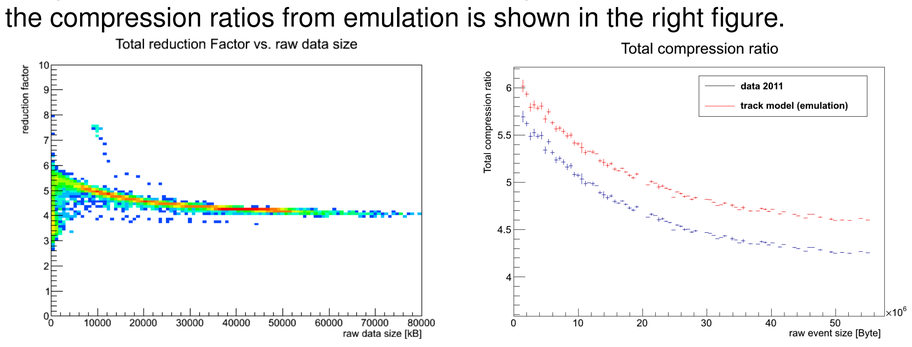
But: Overall compression factor is affected by clusters not fitting into the track model- ∼50% (left panel)
→ kept in separate data blocks

- pad and time residuals have a smaller parameter range which allows to reduce the required size
  → pad from 14 to 10 bit
  → time from 15 to 9 bit
- Distributions of pad and time in residual coordinates are suited for entropy encoding.
- Clusters are transformed from raw coordinates into Cartesian space for tracking, including calibration and correction. Reconstructed tracks are then transformed back to raw coordinates to calculate residuals. The process broadens the distribution and has an impact to the possible compression by entropy encoders.



Compression ratio

Left figure: compression factor in the 2011 Pb-Pb data taking, with an average compression factor of 4.4. Track model compression has not been used, a comparison of the compression ratios from emulation is shown in the right figure.



Pb-Pb data 2011

Conclusions

- Effective data compression is achieved by a combination of lossy transformation (cluster reconstruction) and lossless data compression
- Data compression solution for TPC data operational in ALICE since 2011 Pb-Pb data taking.
- 2011 data dominated by large events, average compression factor of 4.4 has been measured.

References

[1] K. Aamodt et al. The ALICE Collaboration. The ALICE Experiment at the CERN LHC.
JINST 3 S08002, 2008. doi: 10.1088/1748-0221/3/08/S08002.

[2] J. Alme et al. (ALICE TPC collaboration). The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events.
Nucl. Instrum. Meth. A 622 (1) pp. 316-367, 2010. DOI: 10.1016/j.nima.2010.04.042

[3] D. Rohrich, A. Vestbo, Efficient TPC data compression by track and cluster modeling
Nucl. Instrum. Meth. A 566 (2), pp. 668-674 2006. DOI: 10.1016/j.nima.2006.06.056

[4] M. Richter, Data Compression for Large Tracking Detectors for High Energy Nuclear Physics
presented at Nuclear Science Symposium, Valencia, Spain 2011.

[5] G. Grastveit (ALICE Collaboration). FPGA Co-processor for the ALICE High Level Trigger.
Proceedings CHEP 2003, La Jolla, California, USA, 2003. eConf C0303241 (2003) THHT001.

[6] T. Alt, High-speed algorithms for event reconstruction in FPGAs
presented at 17th Real Time Conference, Lisbon, 2009.

UiO : Department of Physics
University of Oslo