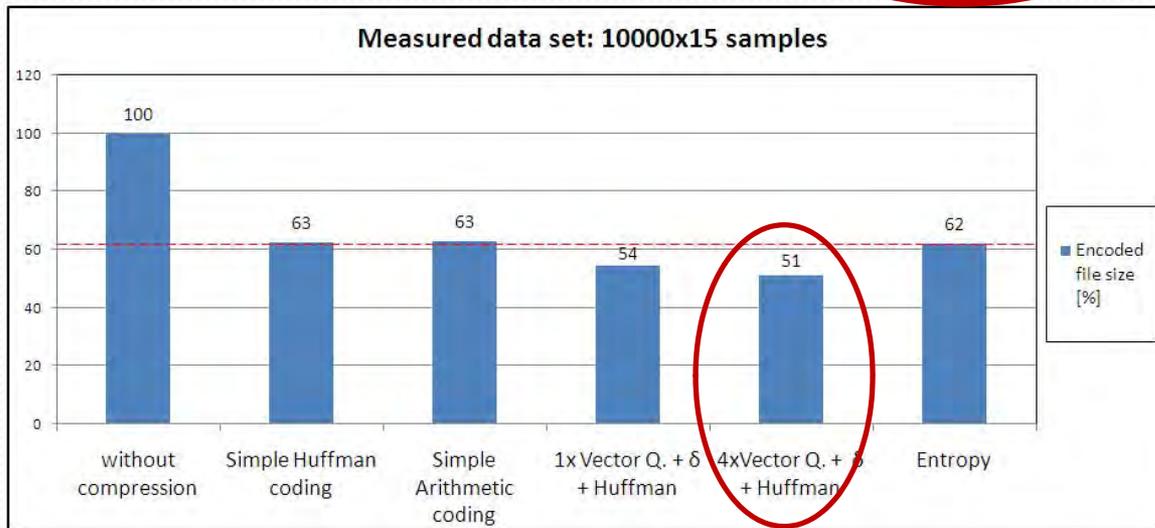
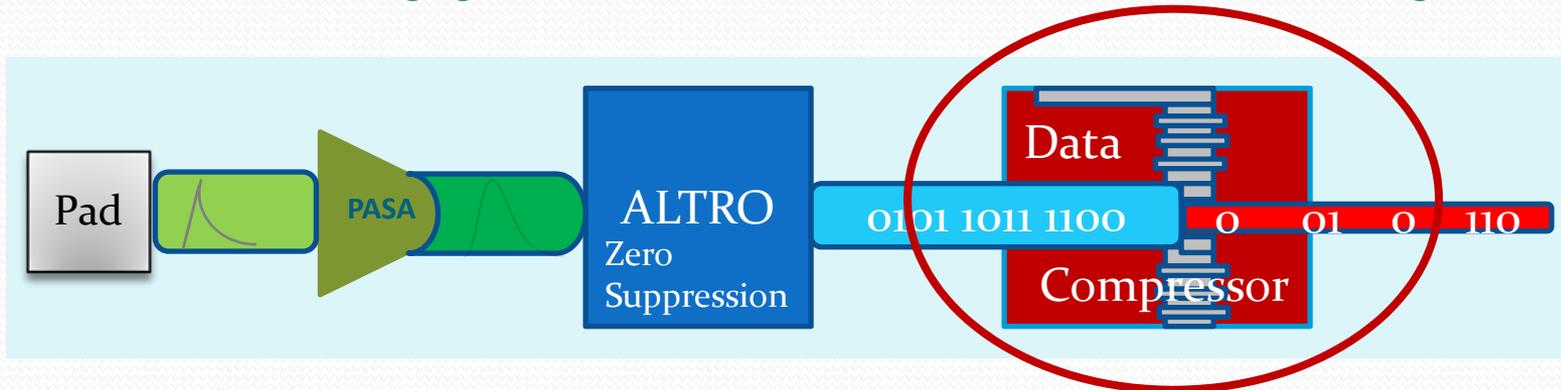


# A Lossless Data Compression System for a Real-Time Application in HEP Data Acquisition



# Poster

Motivation and Background Information

Data compression method

Results



## A Lossless Data Compression System for a Real-Time Application in HEP Data Acquisition



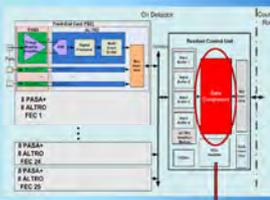
C. Patauner, A. Marchioro, S. Bonacini, Atiq Ur Rehman: CERN, Switzerland  
W. Pribyl: Institute of Electronics – Graz University of Technology, Austria

**Goal:** Realizing a data compression module in hardware to be implemented in front-end electronics of particle detectors. A lossless compression method was modeled in Verilog in a style suitable for automatic synthesis for a Field Programmable Gate Array (FPGA). The compression method was optimized for the reduction of data from a TPC such as the one of the ALICE experiment. The event size in case of the ALICE TPC for heavy ion collisions is estimated as 66MByte after the zero suppression and the event rate is around 300Hz which gives a data rate of:  
 $TPC \text{ data rate} = 66 \text{ Mbyte} \times 300 \text{ Hz} = 20 \text{ Gbyte/s}$   
 The proposed data compression algorithm is capable of reducing this data rate by 49%.

### 1. TPC Front-End Electronics

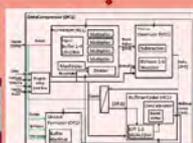
**Time Projection Chamber:** Particle detector used to produce 3D-images of the trajectories of charged particles from collisions.

TPC: Gas volume  $10^6 \text{ m}^3$ , No. tracks/collision 30,000,  
Pb-Pb Collision rate 4 Hz, Event trigger rate 300 Hz,  
Collisions: Event size 16 Mbytes, Data rate 20 Gbytes/s  
(at maximum collision).



**Front-end Electronics:**  
 Mapper + amplifier  
 ALIRO ADC + DSP  
 Zero suppression  
 FE1: 428 channels  
 FE2: PASA & ALIRO  
 ALIRO FE1: ALIRO-FCU  
 200Mbytes address  
 40M words  
 FE2: 18,33 FEU  
 The FCU interfaces the FEU to the data acquisition system (DAQ), the timing and trigger system (TTC) and the detector control system (DCS) in the computing farm.

### 2. Data Compressor



**Data Compressor:**  
 Lossless data compression consisting of three parts:  
 • Vector Quantization  
 • Delta Calculation  
 • Huffman Coding

Normally lossless compression methods use the frequency of occurrence of input symbols to compress them in variable length codes. This does not take into account the correlation between symbols. In the case of the TPC data, the sample values are correlated according to the shape of the pulses, which is defined by the analog shaper PASA.

In order to use this pulse shape, the presented compression method uses a **vector quantization**. To make the compression lossless, a second operation, the **delta calculation** is performed where the differences between the samples in the actual input vector and the chosen reference vector are calculated. The delta values are typically small numbers and can then efficiently be compressed using **Huffman coding**. The Huffman algorithm uses variable length codewords. Delta values present in the data with higher frequency result in shorter codewords, while delta values with lower frequency are coded in longer codewords.



**Delta Calculation**

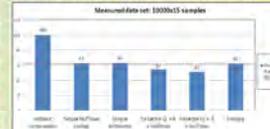


**Huffman Coding**

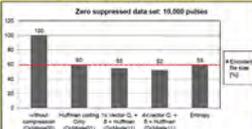
Huffman Code words:  
 - A: 0  
 - B: 10  
 - C: 110  
 - D: 111

### 3. Results

**Reduction Factor for Different Compression Methods**



**Verilog Model Performance**



**Error due to Normalization**



The performance of our lossless compression algorithm is first investigated by using a measured data set from the ALICE TPC containing 10,000 pulses obtained by cosmic ray measurements. Taking advantage of the vector quantization that uses the correlation of the samples allows us to achieve a **compression factor of 49%**. This is better than the Shannon entropy for this data set considered as uncorrelated data.

Number of Slices:	2224
Number of Slice Flip Flops:	1258
Number of 4 input LUTs:	3958
Number of BLOCKRAMs:	16
Number of MULTIPLYERs:	4

RT10 Real-Time 2010 Conference, 24 – 28 May, Instituto Superior Técnico, Lisboa, Portugal