ICECUBE
ROAD GRADER - COMPRESSION RATIOS AND DATA RATES
Version1.0

Joshua Sopher
jesopher@lbl.gov
Lawrence Berkeley National Laboratory

July 26, 2005

1 Method Of Compression For The Road Grader

Data are compressed in three stages: zero-suppression, run length encoding, and "Huffman-lite "compression.
Uncompressed fADC and ATWD data words (10 bits wide) are zero-suppressed as they are read out of buffers. Data at or below a threshold are converted to zeros, and data above a threshold are left unchanged.
Zero-suppressed data are run-length encoded. The number of times data is *repeated* in succession is the run length. Each zero suppressed word is encoded as pairs of data, with the data value first.
The compressor creates a flag bit to indicate if the word is non-zero or zero. (1/0 = non-zero/zero). It replaces 10 bit zero words with a zero (non-zero flag) bit. A non-zero flag is added to the non-zero words, so that an eleven bit word is produced: Bit10 to Bit1 is the non-zero value and Bit0 is the 1 bit nonzero flag. Thus, non-zero (11 bit) and zero (1 bit) words are generated.

2 Calculations For Compression Ratio Based On Two Pulses

A typical waveform for tagged hits for near neighbor local coincidence, consists of two pulses and the baseline for the ATWD or the fADC. Typically only ATWD Ch0 is compressed. Compression ratios in this paper are therefore calculated for two pulses.

2.1 Zero Suppressor Process

Each of the two pulses is 8 samples wide, which are converted into non-zero words, in the zero suppressor. In addition pre-pulse baseline samples are not zero-suppressed, and are also converted into non-zero words. The average number of these samples is 4. so that the total number of the non-zero words is 20. The rest of the samples are "zero" words. For example, a single ATWD channel will have 108 (=128 - 20) "zero" words (baseline samples).

2.2 Calculating The Size Of Non-Zero Compressed Words In The Run Length Encoder And The Hoffmann Compressor

Each non-zero pulse sample occurs once so that its run length = 0. We assume that each pre-pulse baseline sample has a run length of 0, as a worst case.
Each uncompressed non-zero word will generate a compressed data value + run length pair. The number of bits for this pair = 11 (non-zero data) + 1 (zero run-length) = 12 bits.

2.3 Calculating The Size Of Zero Compressed Words For The Baseline In The Run Length Encoder And The Hoffmann Compressor

There are two zero-suppressed baseline sections in the waveforms of the ATWD or fADC. The data value for baselines is a "zero" word. The run length corresponding to each section is a non-zero word.
Each uncompressed baseline section will generate a compressed data value + run length pair. The number of bits for this pair = 1 (zero data) + 11 (non-zero run-length) = 12 bits.

2.4 Compression Ratio For ATWD Ch0

Number of non-zero word pairs = 20 words

Number of bits	= 12 x 20	= 240 bits	
Number of zero word pairs		= 2 words	
Number of bits	= 12 x 2	= 24 bits	
Total compressed bytes		= 264 bits	= 33 bytes

Number of uncompressed bits	= 128 x 10	= 1280 bits	= 160 bytes
Compression ratio	= 1280/216		= 4.84

(Total compressed bytes for 1 pulse = 12 x 12 (non-zero) + 12 x1 (zero) = 156 bits)
(Compression ratio for 1 pulse = 1280/156 = 8.2)

2.5 Compression Ratio For The fADC

Number of non-zero word pairs	= 20		
Number of bits	= 12 x 20	= 240 bits	
Number of zero word pairs	= 2		
Number of bits	= 12 x 2	= 24 bits	
Total compressed bytes		= 264 bits	= 33 bytes

Number of uncompressed bits	= 256 x 10	= 2560 bits	= 320 bytes
Compression ratio	= 2560/216		= 11.85

(Total compressed bytes for 1 pulse = 12 x 12 (non-zero) + 12 x1 (zero) = 156 bits)
(Compression ratio for 1 pulse = 2560/156 = 16.41)

3 Calculations For Data Rate Based *On Two Pulses*

Waveforms for tagged coincidences are typically from the fADC and ATWD ch0.
Tagged and untagged Hit rates are obtained from string 21 data (Dima Chirkin).
The header size for soft and hard local coincidence is 12 bytes, and includes 4 bytes of the Charge Stamp.

3.1 Soft Local Coincidence

Untagged data size = header only	= 12 bytes
Untagged Hit rate	= 737 Hz
Untagged data rate = 12 x 737	= 8,844 bytes/s

Tagged data size = header + number of fadc bytes + number of atwd ch0 bytes.

= 12 + 33 + 33	= 78 bytes
Tagged Hit rate	= 13 Hz
Tagged data rate = 78 x 13	= 1,014 bytes/s
Total compressed data rate	= 9,858 bytes/s

3.2 Hard Local Coincidence

Only tagged data is transmitted to the DAQ.

The compressed data rate	= the tagged data rate
	= 1,014 bytes/s

3.3 Flabby Local Coincidence

Flabby Local Coincidence includes the compressed fADC waveform (33 bytes) in the untagged data size. A better alternative is to add a Charge Stamp (4 bytes) to the Header for Soft Coincidence (David Nygren). This has two advantages over Flabby Local Coincidence. One is to reduce the size of untagged data. This is important, since the total data rate is dominated by the untagged Hit rate. The other is to allow untagged data to be separated into signal or noise, before data is decompressed. This is a useful property, since untagged fADC waveforms are mostly due to scintillation noise (SPE), which should be identifiable at an early stage.
The Charge Stamp gives the pre-peak, peak, and post-peak counts of the first pulse. It also gives the number of peak sample. These parameters make it possible to reconstruct the pulse and obtain its charge and timing.