ICECUBE
DELTA COMPRESSOR - COMPRESSION RATIOS AND DATA RATES
Version1.0

Joshua Sopher
jesopher@lbl.gov
Lawrence Berkeley National Laboratory

July 26, 2005

1 Method Of Delta Compression

Compressed data is obtained by calculating the differences between successive sample counts (after pedestal subtraction). Data is grouped into ranges of sample values (counts). The minimum number of ranges possible are used here, based on actual InIce waveforms and baseline noise. Examination of the waveforms show that samples to be grouped into two ranges, so that the overhead in terms of a range flag is minimized to 1 bit.

A typical waveform for tagged hits for near neighbor local coincidence, consists of two pulses and the baseline. Each sample of a digitized PMT waveform is 10 bits wide. The difference between counts of successive samples, provides a delta which is very often smaller that 10 bits. For baseline samples this difference is usually +/- three counts (3 bits, including the sign bit) or less, (and is associated with the noise). For pulses this difference is greater than +/- three counts. By using count sizes of 3 bits for the small deltas, and 11 bits (including the sign bit) for the large deltas, it is possible to obtain a compression ratio of about 2:1.

Calculations below show:

Compression ratio for ATWD ch0 data = 2.

Compression ratio for fADC data = 2.2

It is possible to obtain a better compression ratio using more than two delta ranges, but it adds complexity to the design. This is not necessary for In-Ice waveforms, since the compressed data rate = 11,912 bytes/s , which easily meets the data rate requirement of 20,000 bytes/s.

A comparison between using two ranges and multi-ranges suggested by Chris Wendt, for a number of waveforms. For single pulses the compression ratio of the dual is greater than 90% of the multi-range for 7 out of 9. These are single, double, complex, and saturated waveforms. This number falls for three (87%) and four (84%) pulses in the waveform.

2 Calculations For Compression Ratio Based *On Two Pulses*

Each of the two pulses is 8 samples wide, and are converted into big words. The number of big deltas for two pulses is 16, and the rest of the samples converted into small deltas. For example, a single ATWD channel will have 112 (=128-16) small deltas (baseline samples).

One bit is required in addition to the signed delta words. One bit will identify if the count size is small or large, and another is the negative/positive sign. This makes the total number of bits = 4 for small deltas, and = 12 for large deltas.

2.1 Compression Ratio For ATWD Ch0

Number of big delta words		=	16 words	
Number of bits	= 12 x 16	=	192 bits	
Number of small delta words	= 128 - 16	=	112 words	
Number of bits	= 4 x 112	=	448 bits	
Total compressed bits		=	640 bits	= 80 bytes
Number of uncompressed bits	= 128 x 10	=	1280 bits	= 160 bytes

Compression ratio = 1280/640 = 2.0

2.2 Compression Ratio For The fADC

Number of big delta words = 16 words
Number of bits = 12 x 16 = 192 bits
Number of small delta words = 256 - 16 = 240 words
Number of bits = 4 x 240 = 960 bits
Total compressed bits = 1152 bits = 144 bytes

Number of uncompressed bits = 256 x 10 = 2560 bits = 320 bytes
Compression ratio = 2560/1152 = 2.2

3 Calculations For The 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 + 144 + 80 = 236 bytes
Tagged Hit rate = 13 Hz
Tagged data rate = 226 x 13 = 3,068 bytes/s
Total compressed data rate = 11,912 bytes/s

3.2 Hard Local Coincidence

Only tagged data is transmitted to the DAQ.
The compressed data rate = the tagged data rate
= 3,068 bytes/s

3.3 Flabby Local Coincidence

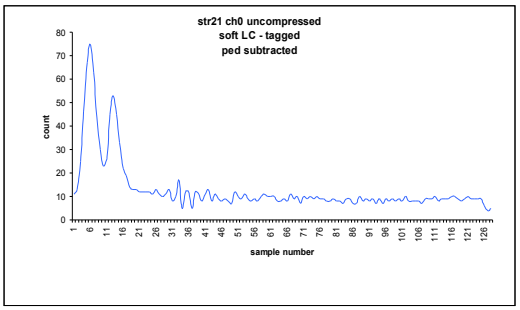
Flabby Local Coincidence includes the compressed fADC waveform (144 bytes) in the untagged data size. A better alternative is to add a Charge Stamp (4 bytes) to the Header for Soft Coincidence (David Nygren). This has two advantages over Flabby Local Coincidence. One is to reduce the size of untagged 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.

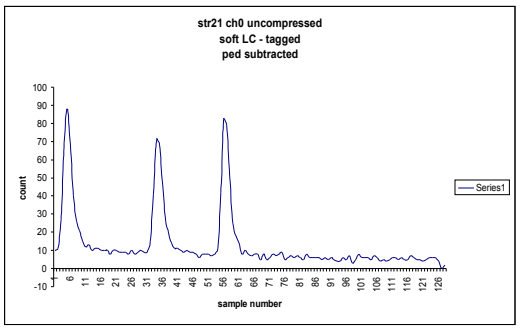
The Charge Stamp gives the pre-peak, peak, and post-peak counts of the first pulse. It also gives the number of peak sample. These parameters make it possible to reconstruct the pulse and obtain its charge and timing.

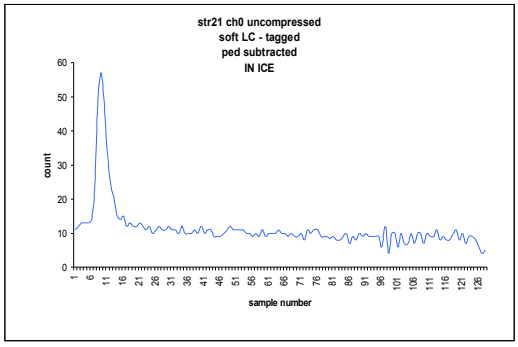
4 Comparison Between My Dual Range Method And The Multi Range Method Described By Chris, *Based On String21 Waveforms*

Waveforms are numbered right to left and top to bottom.

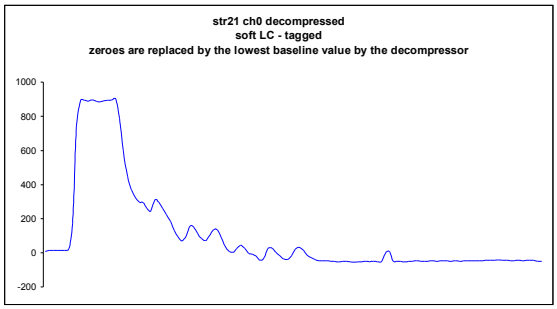
Waveform1

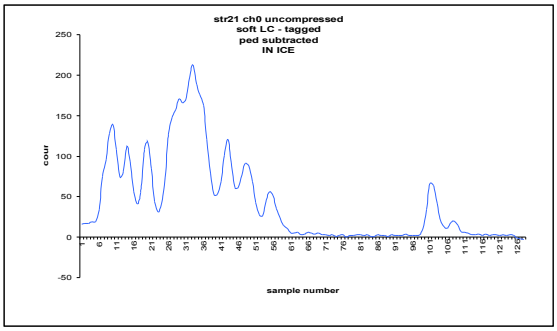


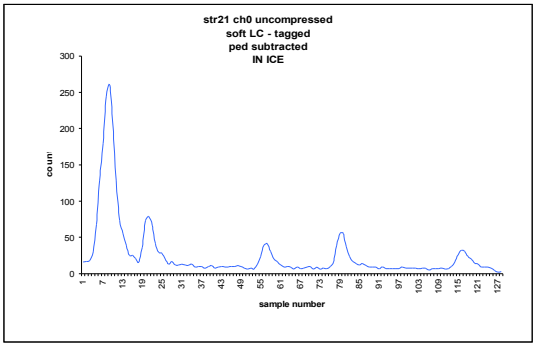




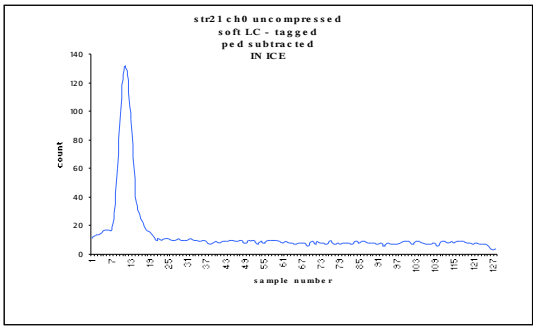
Waveform2



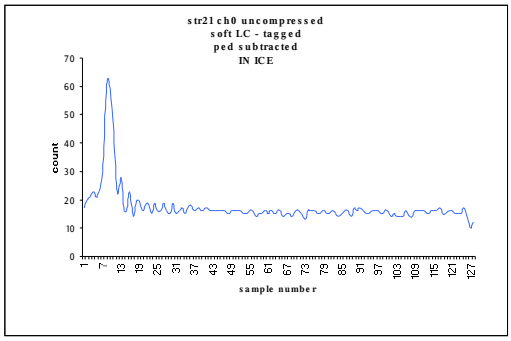




Waveform7



Waveform8



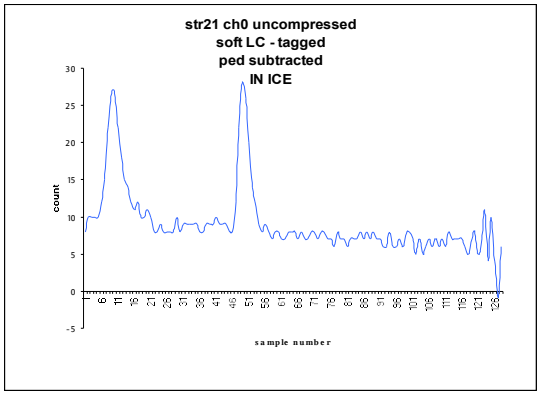


Table1. Compression ratios of the above nine waveforms are shown, for two types of delta compression.

	Dual range	Multi-range	compr: dual/multi range	Waveform number
total bytes	74	74		7
compr ratio	2.16	2.15	1.004651163	
total bytes	77	74		5
compr ratio	2.08	2.15	0.96744186	
total bytes	79	74		9
compr ratio	2.02	2.15	0.939534884	
total bytes	81	76		8
compr ratio	1.98	2.11	0.938388626	
total bytes	86	79		1
compr ratio	1.86	2.01	0.925373134	
total bytes	89	82		3
compr ratio	1.8	1.96	0.918367347	
total bytes	110	96		6
compr ratio	1.45	1.66	0.873493976	
total bytes	125	105		4
compr ratio	1.28	1.52	0.842105263	
total bytes	130	117		2
compr ratio	1.23	1.36	0.904411765	