Update on Wavelet Compression

Presented by S.Klimenko

University of Florida

- Outline
  - Wavelet compression
    - concept
    - E2 data
    - artifacts
    - software
  - Conclusion

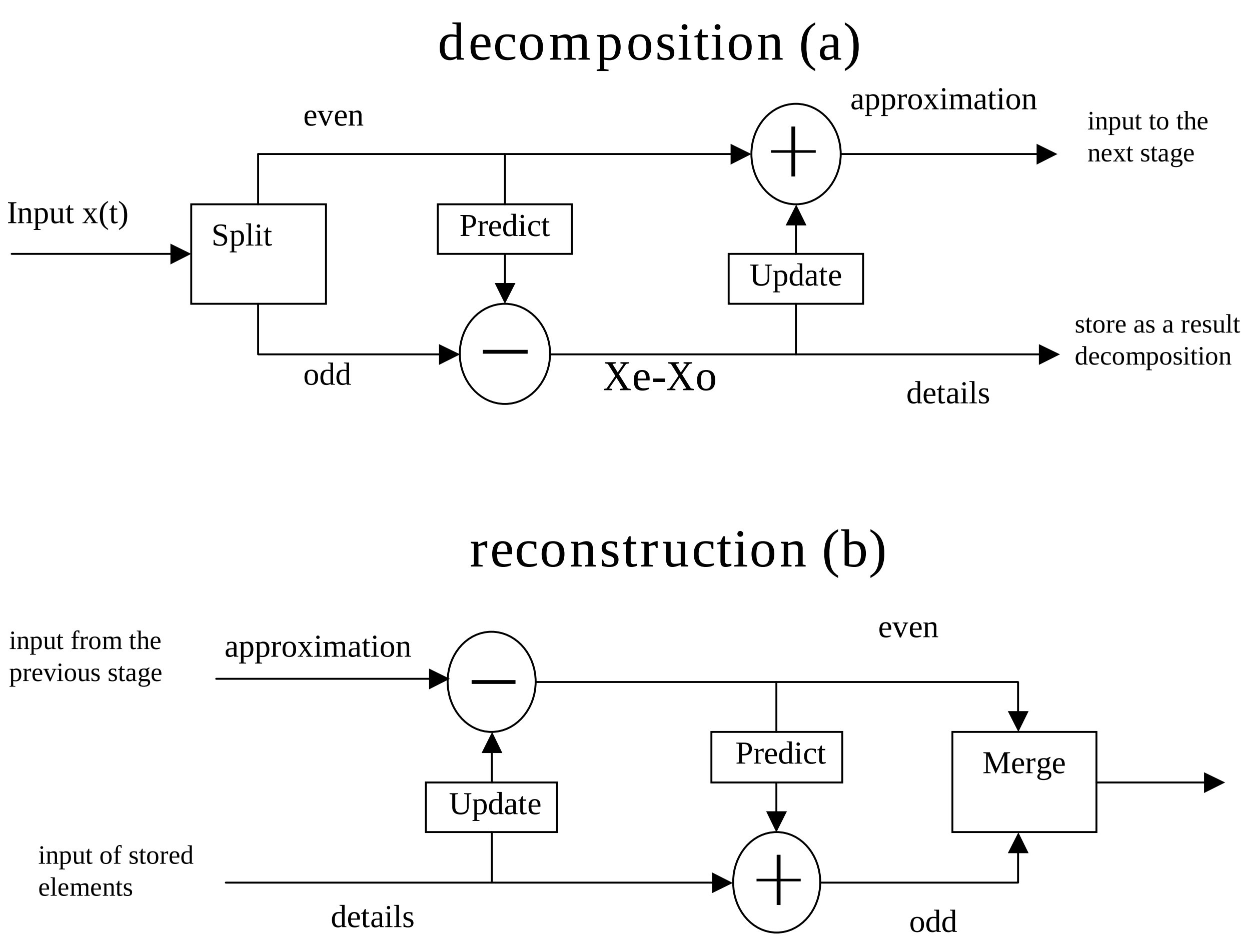
LIGO-G010139-00-Z

Data Compression with Wavelets

- Data de-correlation & reduction
  - apply transforms that allow ***more compact*** data representation
  - wavelet are used to de-correlate data
  - wavelets allow data reduction
- Pack data using lossless encoder (gzip, 0 suppression, ….)
  - many LIGO signals are mainly random Gaussian noise with admixture of non-Gaussian components
  - wavelet transform makes data even “more random”

➡ use data encoder optimized for compression of Gaussian noise

Lifting Wavelet Transform



→ 0 - f / 2

→ f / 2 - f

Haar:
P = Xe
U=1 / 2

- Wavelets de-correlate data

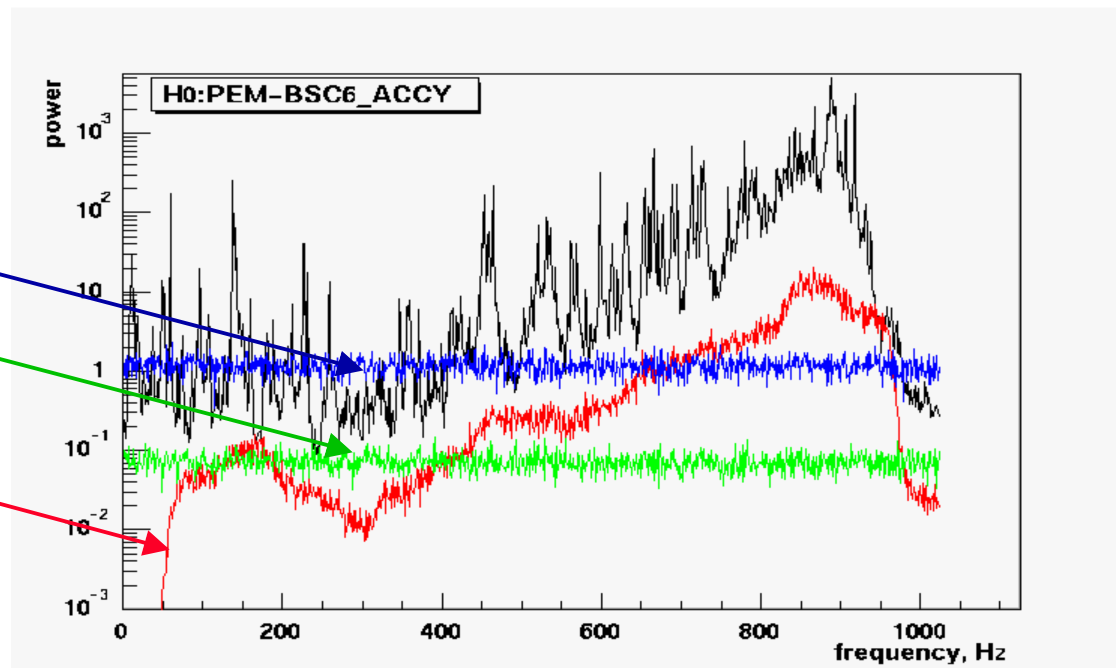
Compression With Losses

- Goal
  - considerably reduce the data ***bps*** - average # of ***bits per sample***
- Applications
  - compress environmental and control channels, where very detail information may not be important
  - produce archived data sets using *quasi-lossless* compression
  - generate reduced data sets for the data analysis & investig. task
- Main problems
  - possibly loss of “useful information” (how to control it?)
  - artifacts can be added to compressed signal
  - different channels may require different compression options
- Possible solution
  - reduction of the data dynamic range in wavelet domain.

Data Dynamic Range Reduction

- compression: $x \rightarrow w_i \rightarrow int(w_i/K_i) \rightarrow encode\ (rdc, gzip, \ldots)$
- reconstruction: $decode \rightarrow w_i' \rightarrow w_i' K_i \rightarrow x'$
  - i - wavelet layer number, K_i - scaling factors
  - x – initial data set, x' - reconstructed data set
- noise for compression in time domain (K_i=*constant*) limits compression at 6-7bps
- compression noise δ
  - noise generated by random process *int*: $x' = x + \delta$
  - small correlation between δ and x if $\delta^2 << x^2$, no artifacts
  - losses $\varepsilon = \delta^2 / x^2$

δ_T (4.0bps)
δ_T (6.0bps)
vs
δ_W (4.0bps)

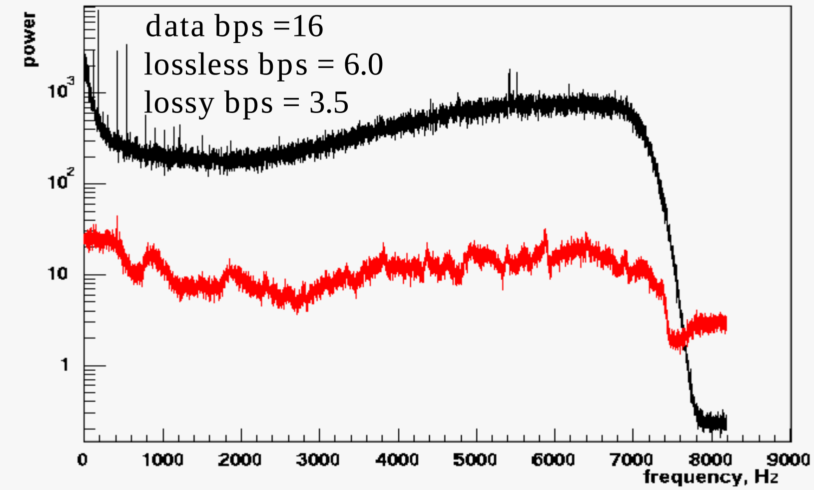


S.Klimenko, LSC March, 2001

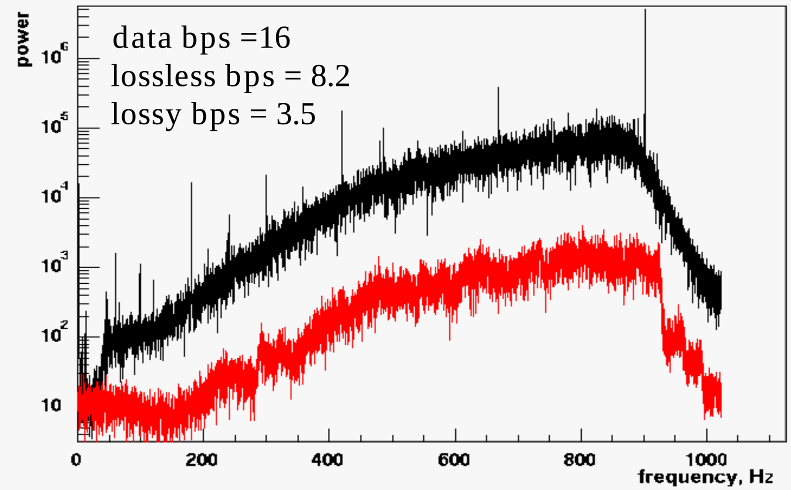
Short Channels (ε=1%)

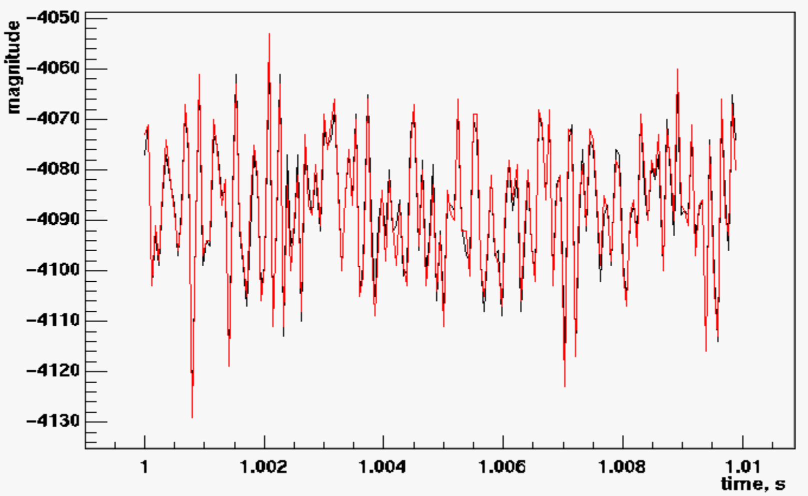
black – original data, red – δ & reconstructed data (E2 run)

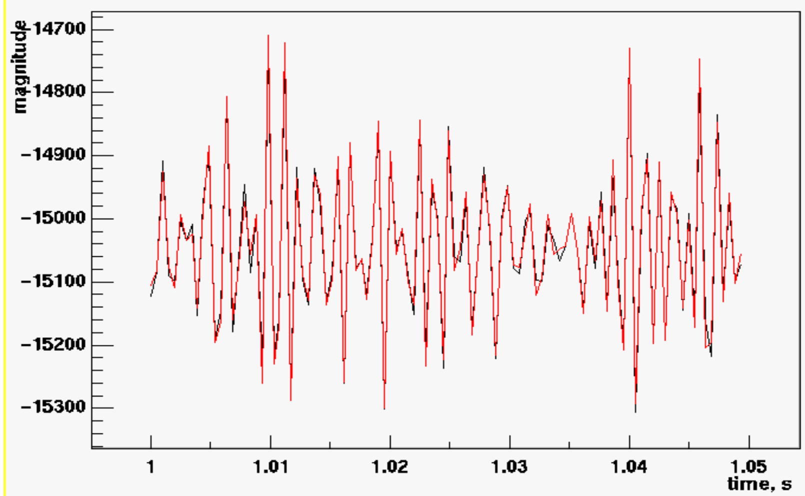
H2:IOO-MC_I



H2:SUS-ETMX_SENSOR_UR

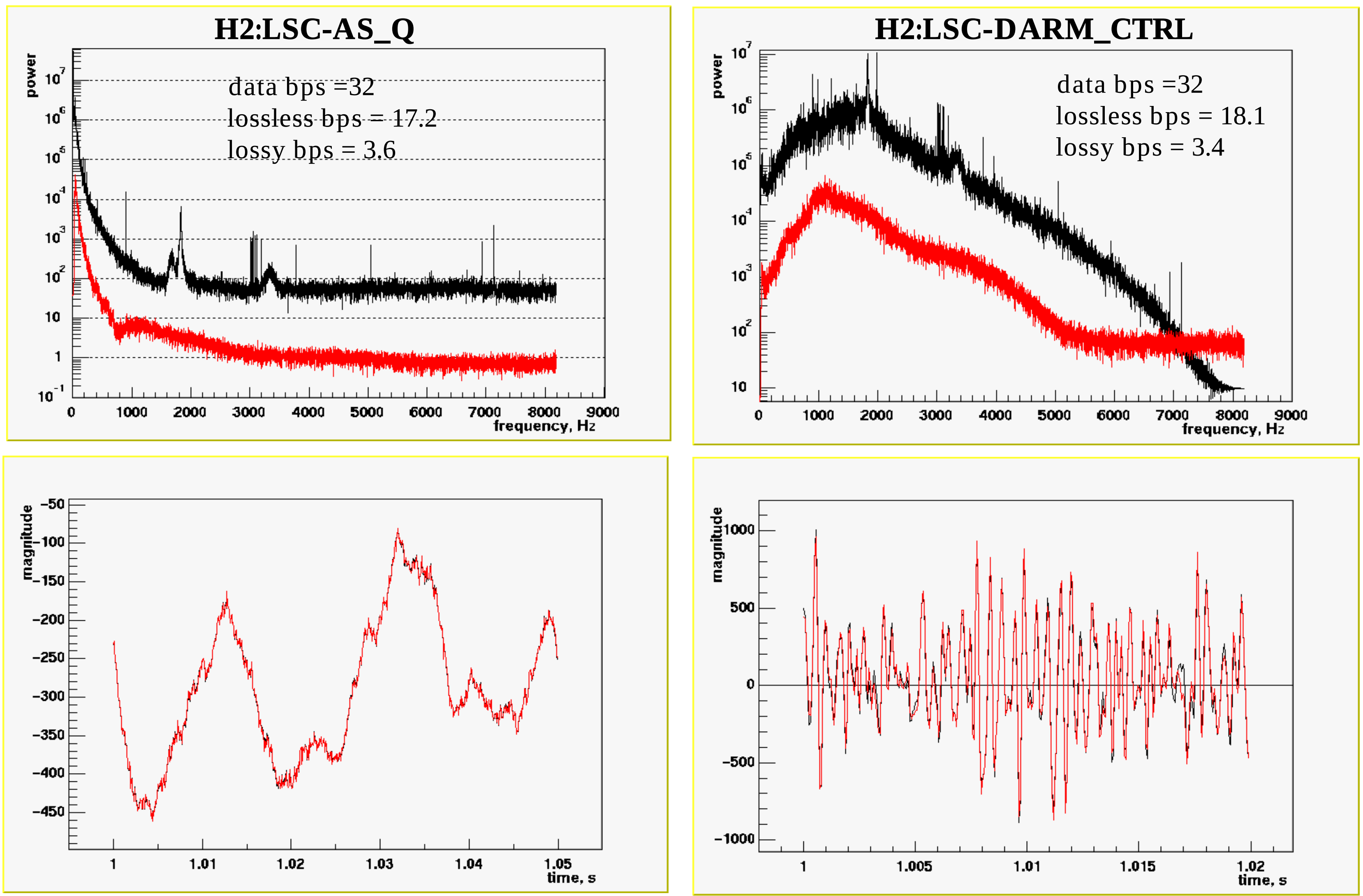






Float Channels (ε=1%)

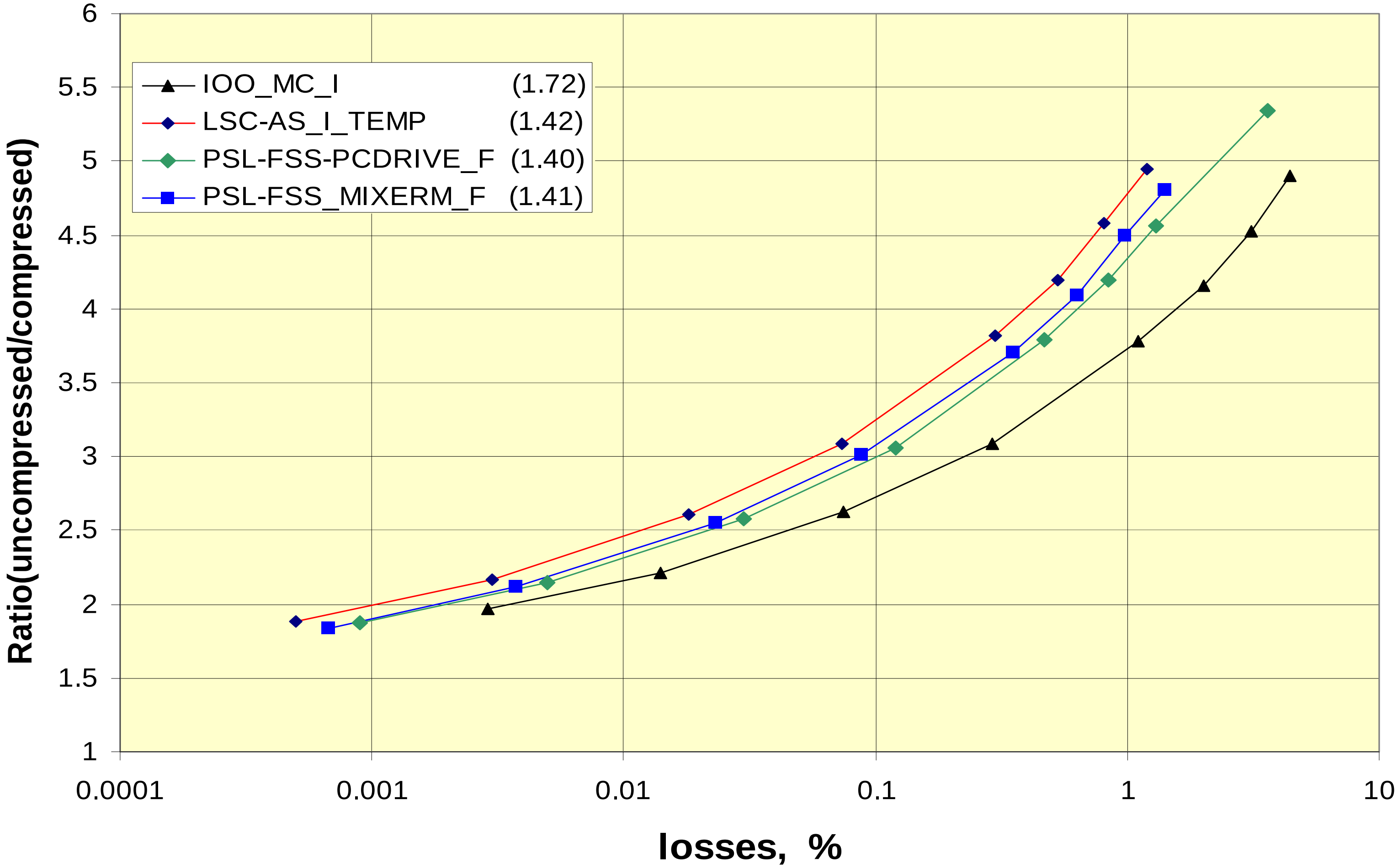
black – original data, red – δ & reconstructed data (E2 run)



S.Klimenko, LSC March, 2001

compression vs losses (E1)

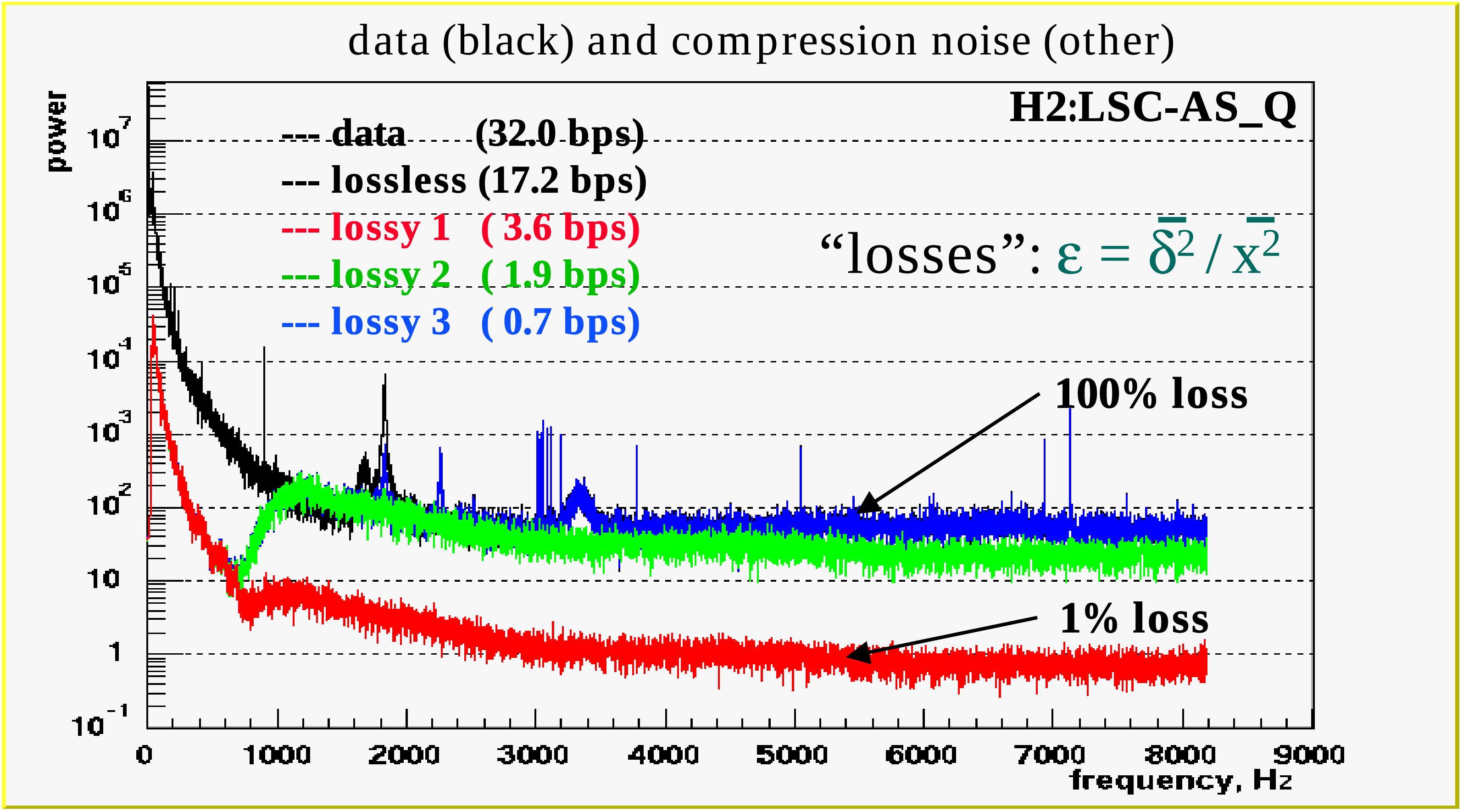
Wavelet compression allows to work in terms of “losses”: $\varepsilon = \delta^2 / x^2$



S.Klimenko, LSC March, 2001

Wavelet Data Reduction

- Varity of options: lossless, quasi-lossless, aggressive or severe compression for different channels and frequency bands.
  - currently losses specified separately for high and low frequencies ε_1 & ε_2 – losses for frequency bands <1kHz & >1kHz

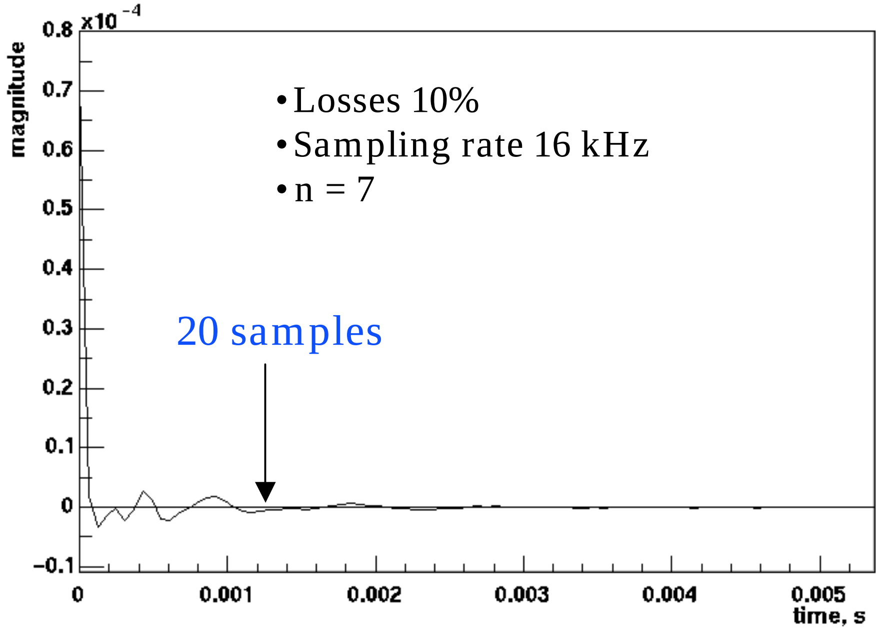


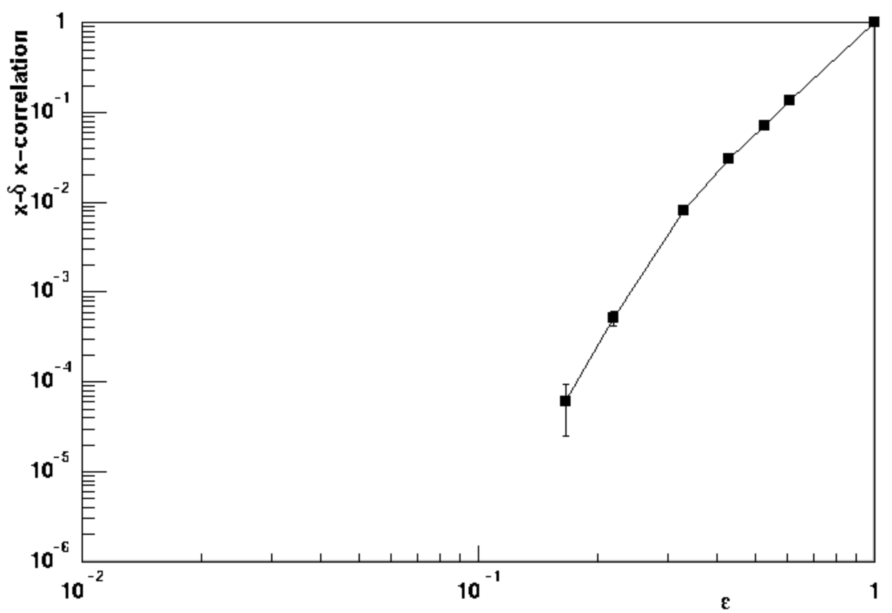
S.Klimenko, LSC March, 2001

Compression Noise

- x' auto-correlation function: $R(\tau) = R_{xx}(t)+2R_{x\delta}(\tau)+R_{\delta\delta}(\tau)$
  - cross-correlation term: $r_{x\delta}(\tau) = R_{x\delta}(\tau)/R_{xx}(0)$
  - noise auto-correlation term: $r_{\delta\delta}(\tau) = R_{\delta\delta}(\tau)/R_{xx}(0)$

Induced correlations for white Gaussian noise





- Is perturbation local?

  $r_{\delta\delta}(\tau) = 0$ if $\tau > 2^{n+1}/f_s$

- Is there a correlation between x & δ?

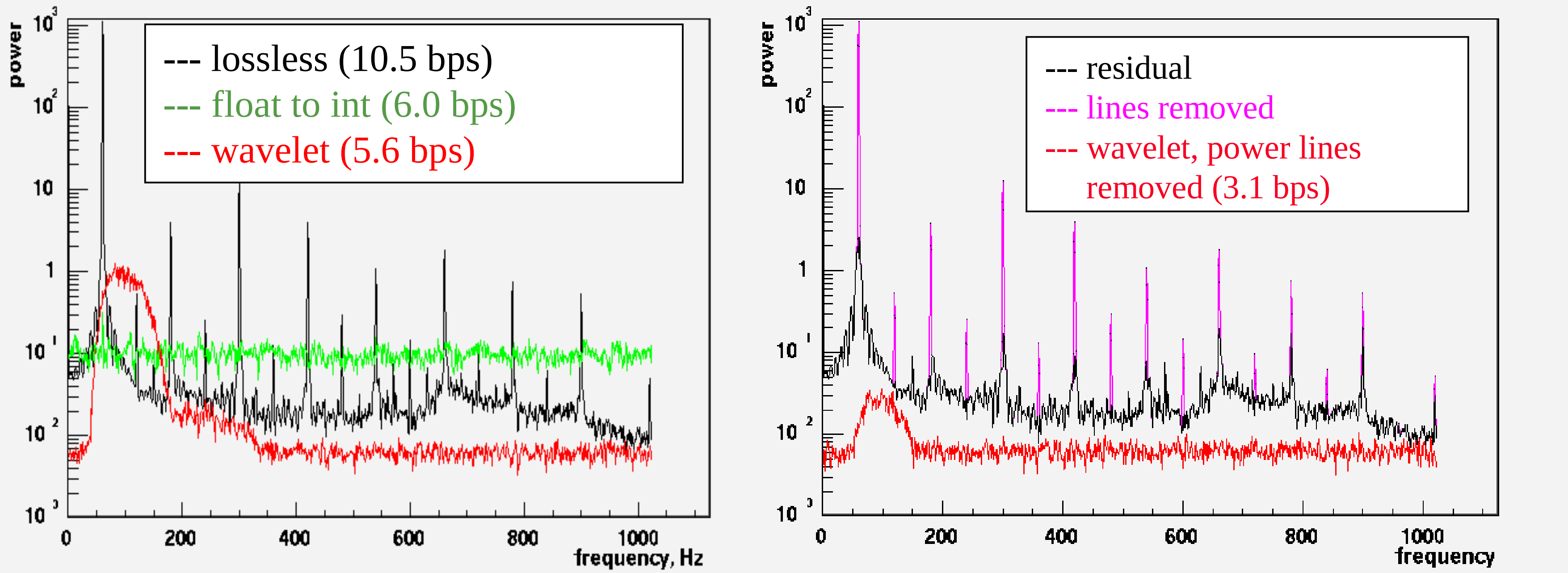
  $r_{x\delta}(0)=max(r_{x\delta}) < 5.e\text{-}5$ for $\varepsilon < 15\%$

S.Klimenko, LSC March, 2001

Artifacts from Narrow Lines

- If compress signals is dominated by lines – artifacts can be added
  - limitation for blind use of wavelet compression (need some thinking)
  - need to improve strategy for selection of scale factors (current strategy $K_i = a\ rms$) to make wavelet compression robust (doesn't need thinking)
- Several options available: strong narrow lines (power) should be removed to minimize signal *rms* and δ, the residual signal can be safely compressed
  - line removal is lossless and requires to store 2 float numbers per line.

PEM-EX_V1 (RDS ~10channels – 40k / s)



Wavelet Compression Software

- Standalone utilities (S.Klimenko, B.Mours, A.Sazonov)
  - WatFrComp input output [parameters] --- copy the entire structure of input file, replacing original time series with compressed time series. Headers and low rate channels (<2kHz) are not wavelet compressed.
  - WatFrUnComp input output [frame compression option] --- recover time series from compressed data. The output file is as big as original data file and can be read by standard means.
  - WatFrTest input --- view compression summary statistics
  - Makefile —-- make script for automated processing of large number of files
- Reading of compressed data
  - Wavelet compressed file is an ordinary frame file that is readable with FrLib, but data is stored in wavelet (time-scale) domain.
  - currently readable from ROOT using DMT means.
  - adding uncompress function in FrLib removes all limitations for use
- Documentation (installation, usage, examples)
  - http://www.phys.ufl.edu/LIGO/wavelet/compress.html

E2 run statistics

- frame data (total 1.52MB)
  - 2kHz & 16kHz channels : **92.0%** --- lossy WAT
- Lossless compression

  gzip – **1.03MB,** frame gzip – **1.06MB,** frame gzip+zero **–1.01MB**

- WAT compression performance
  - frame file compression

default

Losses, %	0.1	1	10	0.1(<1kHz) ,10(>1kHz)
WAT, kB	430	341	250	317
headers/losses, %	16 / 20	20 / 25	27 / 34	21 / 27
+gzip, kB	348	258	165	234

    - ✓ for losses >10%, fraction of service data can be as large as 30%.
    - ✓ Add ~40kb if dark port signal is not wat compressed.
  - computation efficiency
    - ✓ SUN ultra 300 – 0.7-2 MB/s
    - ✓ one 1GHz PC (~2-3 times faster) will compress all LIGO data

Summary

- Wavelets is a powerful and flexible compression tool.
- UNIX (Solaris) utilities are available for practical use.
- Wavelets can be used to de-correlate and reduce data
  - for lossy compression the data dynamic range reduction in wavelet domain is used – next step compare to dynamic range reduction in time domain
- Combination of wavelets and rdc encoder offers a universal tool both for lossless and lossy compression.
- range of options between lossless, quasi-lossless, lossy and aggressive compression (like decimation), 2 parameters to specify.
- Data reduction down to the level of 1bps is achievable.
- Computational efficiency is good and can be improved.
- Plan: develop robust strategy for selection of scale coefficients to allow blind use of wavelet compression at rates ~ 3-4bps