

Every shuffle of the data stream, and thereby the stream of bytes, is a new sample for the Wf.

But, if we want to use the Wf to detect a change in the data stream, we should

- a) Define a filter constant called H_FILTER_SHUFFLE at Hpublic.h
- b) Following the deflate example, declare a filter function called H_filter_shuffle with the parameter set exactly as H_filter_deflate.
- c) Create a new source code file called Hshuffle.c, put the implementation of shuffling algorithm inside.
- d) Define a new property list function called H the definition
- e) Following the function H_Pset_deflate, put the implementation of H_Pset_shuffle
- f) At Makefile.in, add Hshuffle.c in the source file list.

3. Definitions in the report:

- 1) **Compression ratio**: The ratio of the compressed file size to the original file size.
- 2) **Encoding time of the algorithm**: The difference between the elapsed time between writing an

Relative overhead of decoding time with the addition of shuffle

Ratio of

4. Data collections

I totally collected 24 HDF5 datasets from HDF4 and HDF5 files. Among them there are 5 SAF datasets, 2 Swede radar datasets, 2 MIT physics datasets, 1 Spot dataset, 1 SWARM dataset, 13 NASA EOS datasets.

Table 1: Dataset information Wf the study

File Name	Dataset Name	Array size (byte)	Data type
TRIM TRIM	/DATA_GRANULE/SwathData/geolocation	4968704	Float32

5. System descriptions

Linux 2.2.18smp i686

Physical Uemory 960 MB

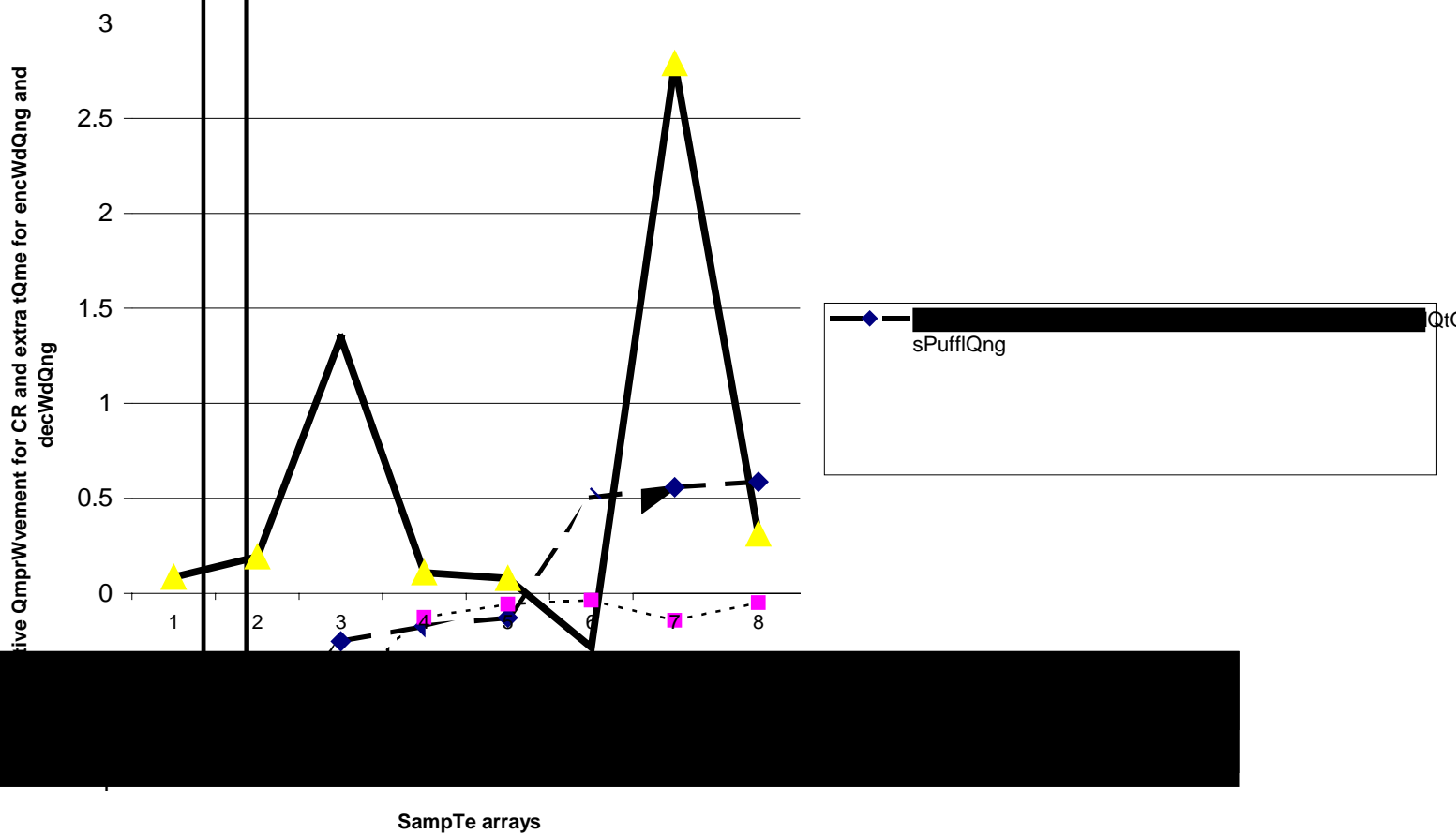
CPU:

Two processors

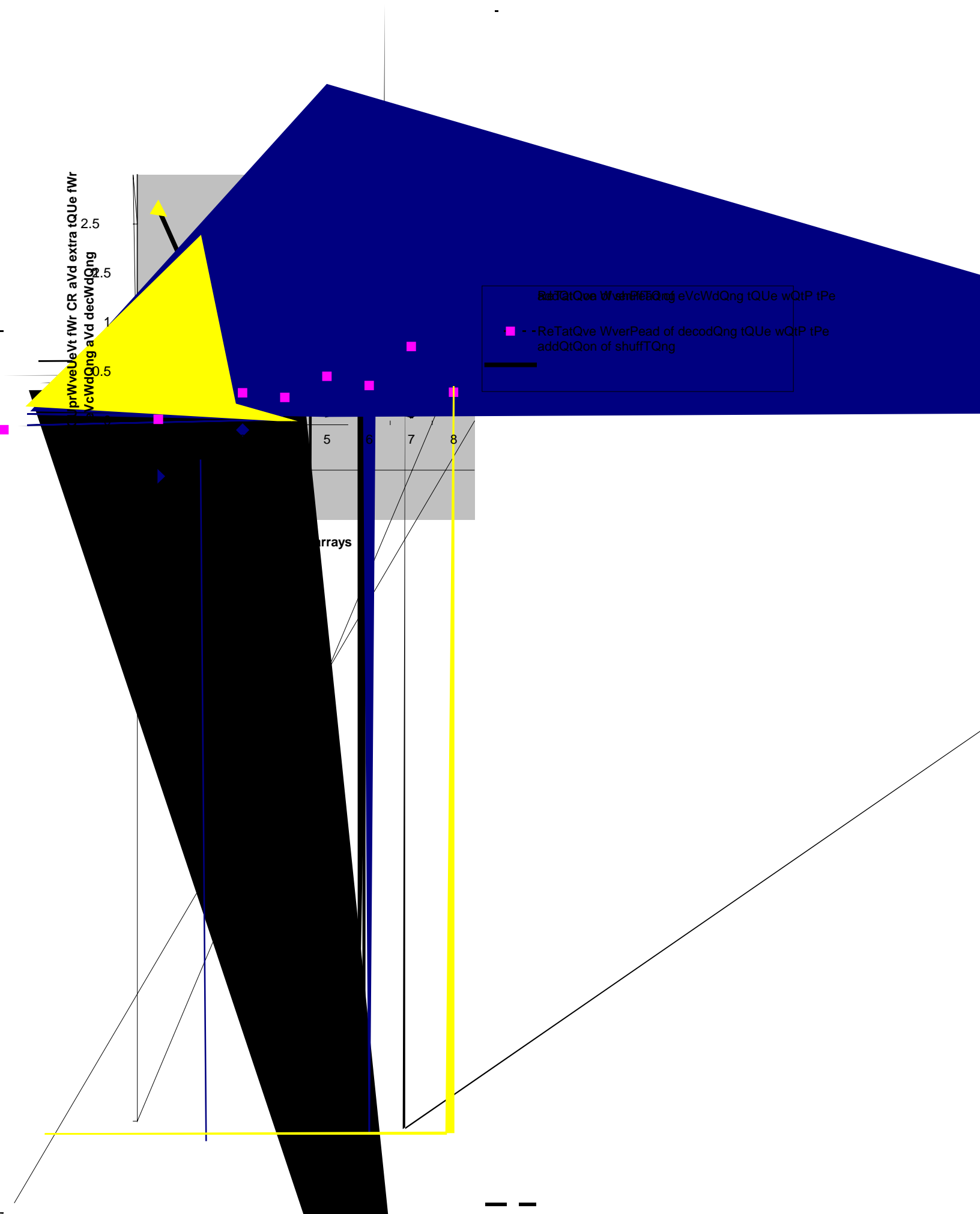
b) Results

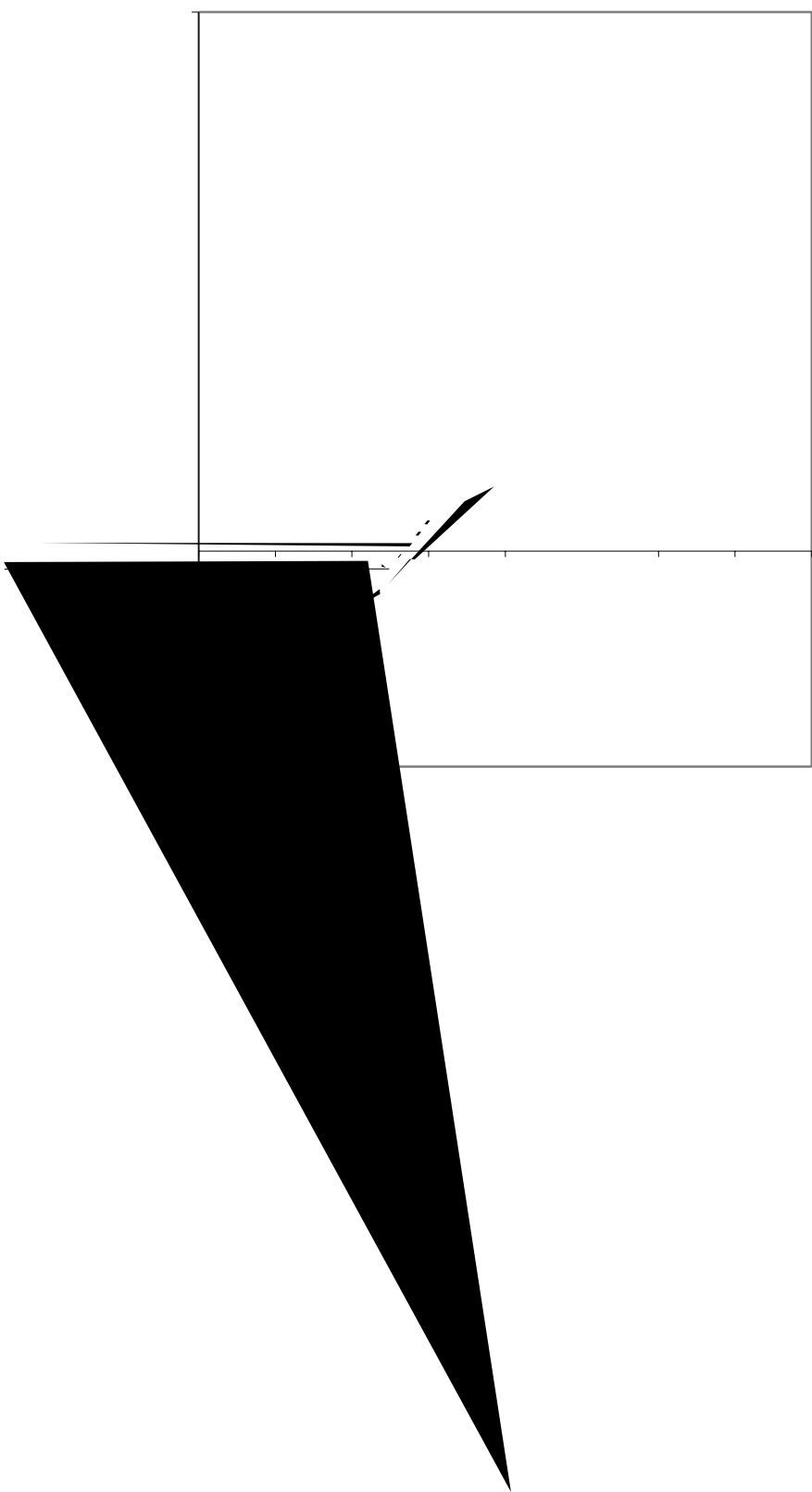
- The combination of shuffling algorithm with bzip2 and gzip can gain improvement of compression ratio for most 32-bit and 64-bit data samples.
- On average, the improvement of compression ratio for float32 is 10% for both compression packages.
- On average, the improvement of compression ratio for float64 is 5% for both compression packages.
- Most cases show that less encoding time and decoding time are used for compression with the shuffling and bzip2.
- Most cases show that insignificant extra encoding and decoding time are used for compression with the shuffling and gzip.
- The combination of shuffling algorithm with bzip2 and gzip cannot significantly benefit for those data that cannot gain better compression ratio with bzip2 and gzip on their own.
- The combination of shuffling algorithm with bzip2 and gzip is generally not good for 16-bit data.

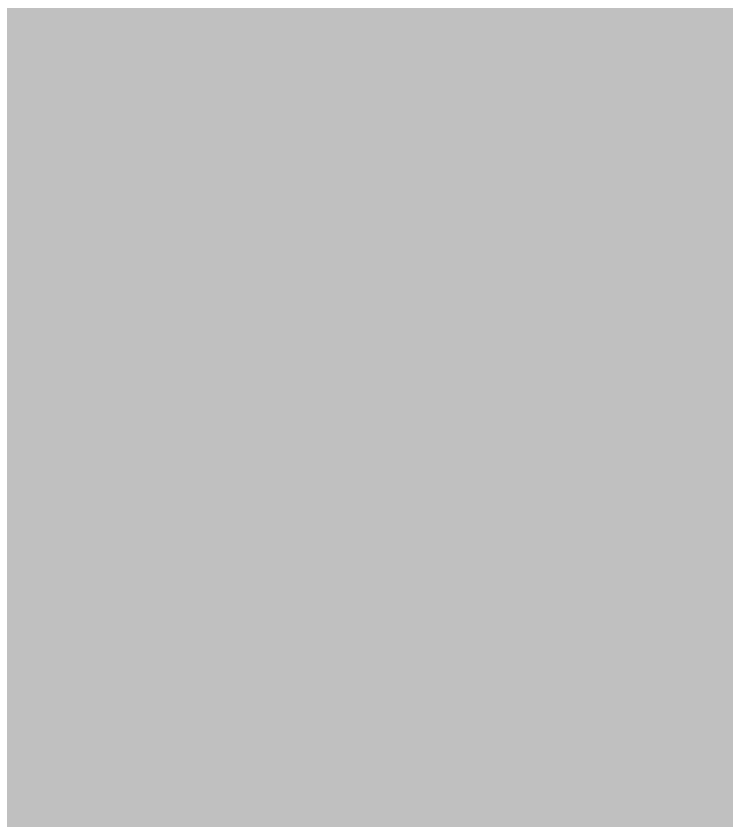
Fig. 3: Illustration of level 9 CR movement, extra time of encoding and decoding for total

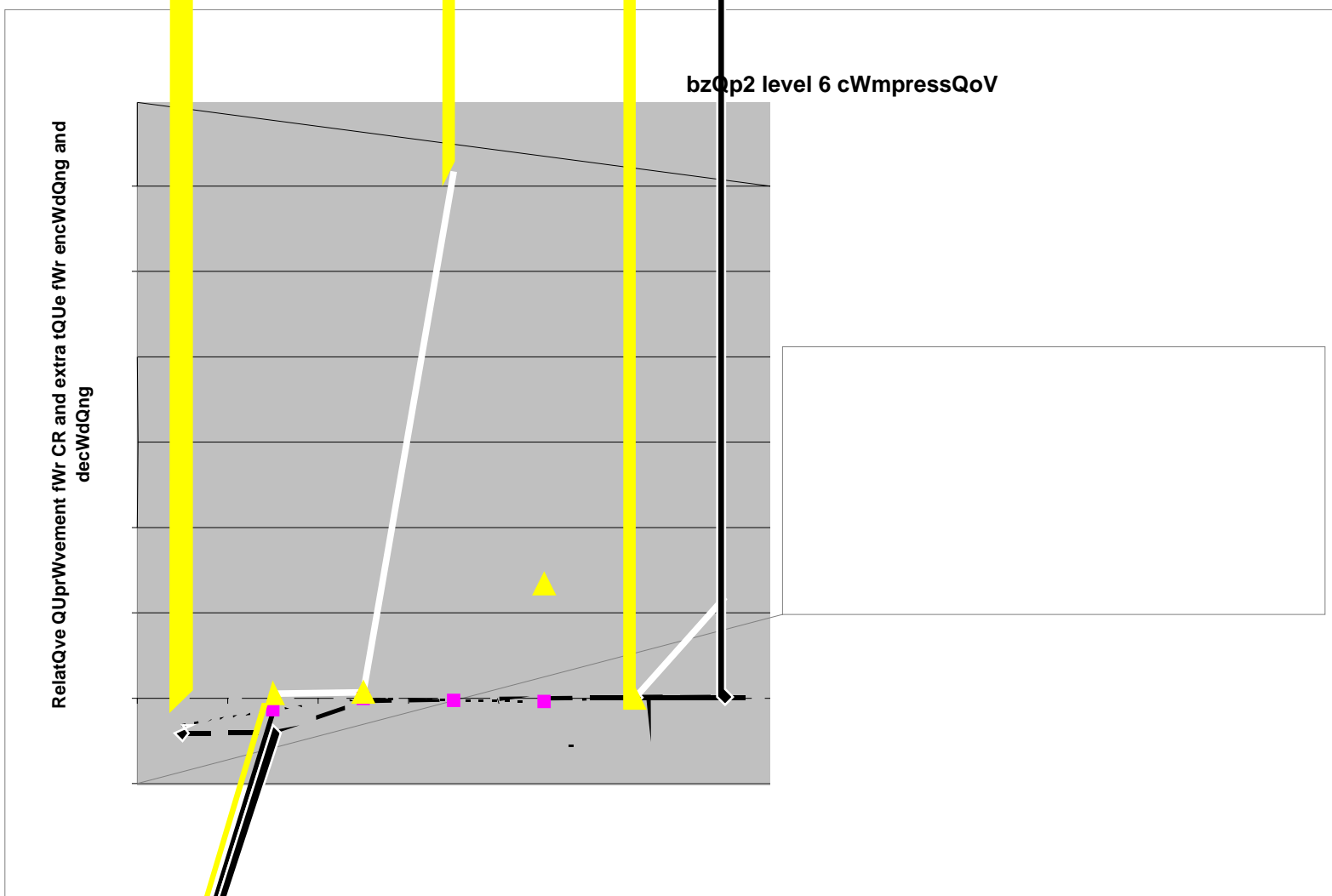












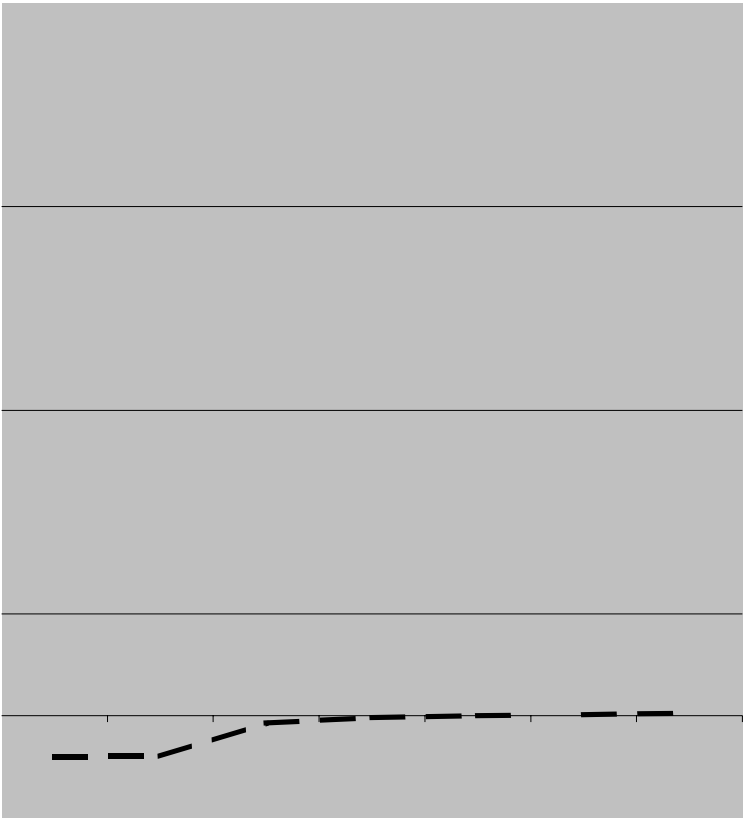
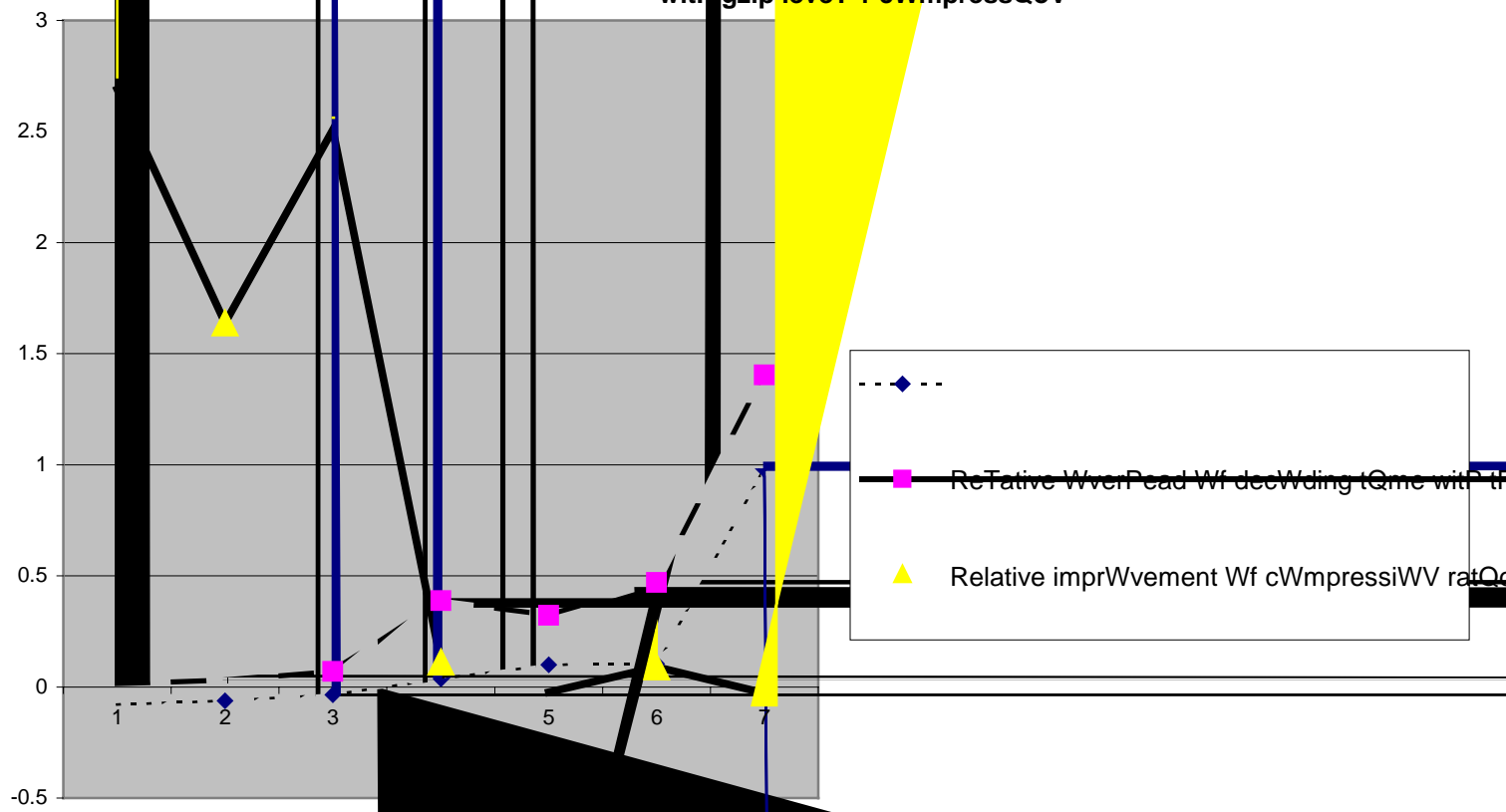
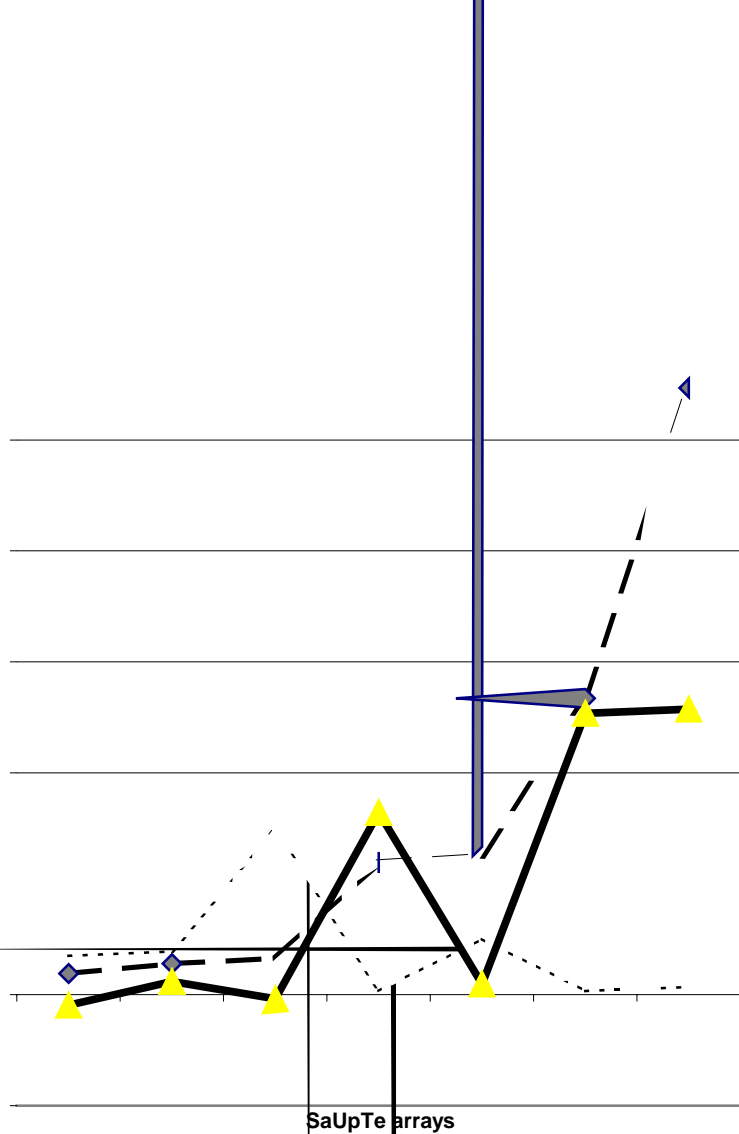


Fig. 10: Illustrate atQoV Wf CR QmprWvement, extra tQme Wf encWdQng and decWdQng for fToat64 c withPgzip leveT 1 cWmpressQoV



ReTative improveUeVt fWr CR aVd extra tQme fWr eVcodQVg aVd



ReTatQve overhead of eVcodQVg tiUe wQth the additiWn of (s225huf)-14(fR4(nQ))Toverhead of

