

XTC decompression in pyana: performance study

April 17, 2013

Igor Gaponenko

The code

- **OFFLINE release ana-0.9.3**
 - pdsdata V06-07-00 (optimizations -O4 and -O5)
 - pypdsdata (modified)
- **Compression algorithms:**
 - *Hist16* - original version for 16-bits CSPad
 - *HistN* - better compression, but slower
- **pyana configurations:**
 - 1 through 9 processes

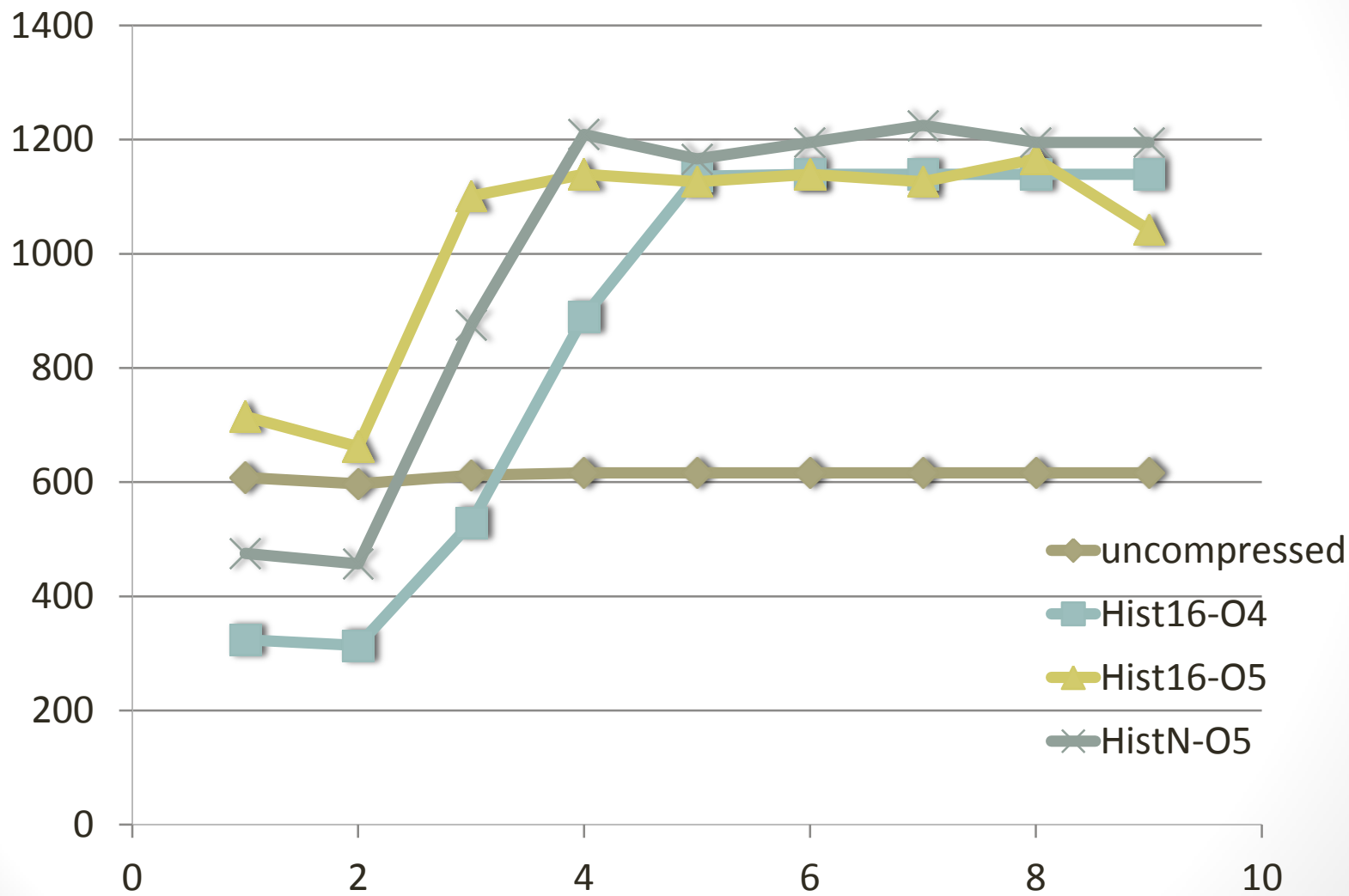
The machine

- **Psana201**
 - 2 CPUs, Intel X5675, 3.07 GHz
 - 12 MB cache/CPU
 - 12 cores (24 cores with HT)
 - 24 MB RAM
 - IB connection to storage

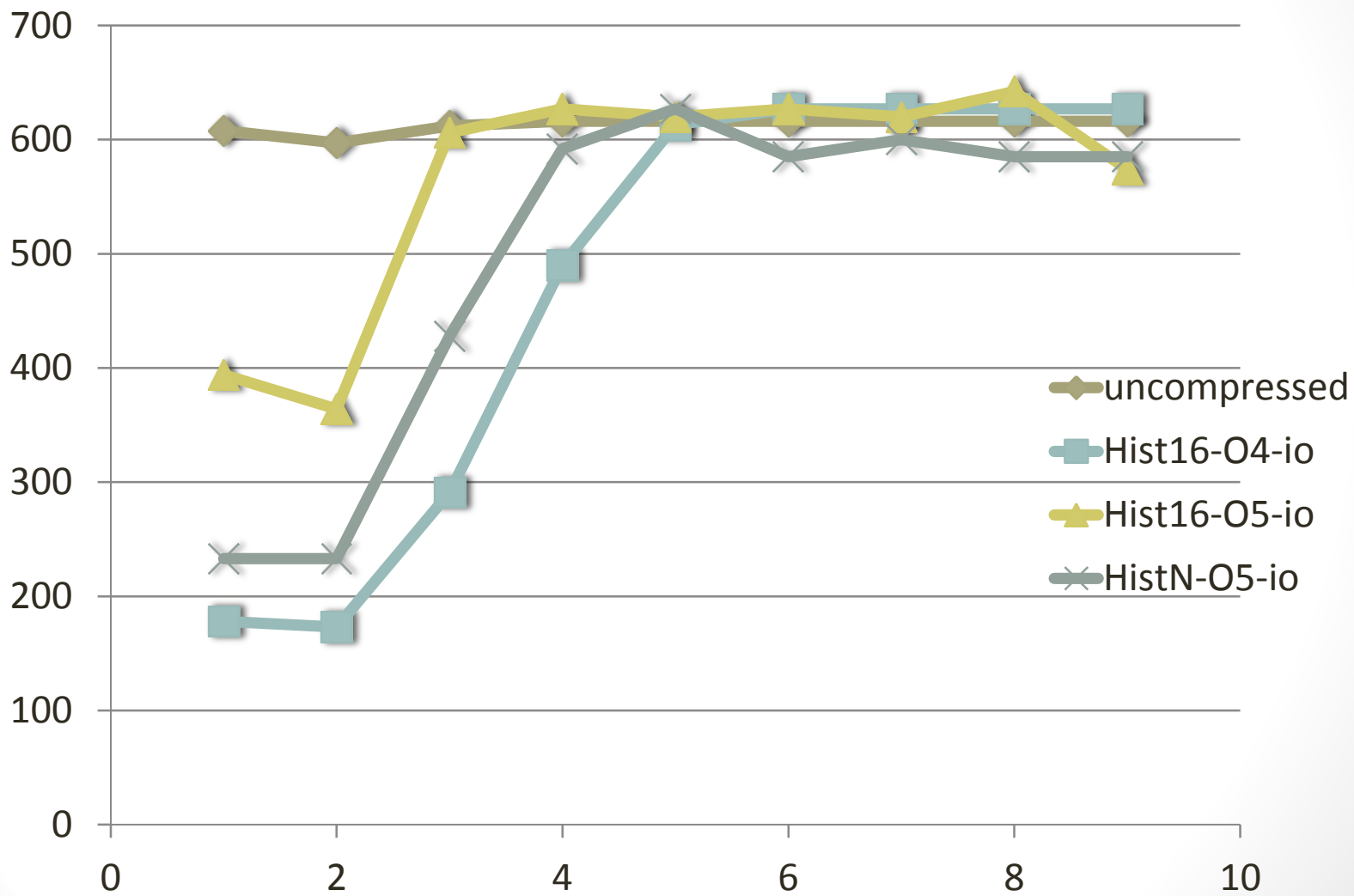
The dataset

- **1 file of experiment cxis0413**
 - e236-r0003-s04-c00.xtc (**98 GB**)
- **Compressed files:**
 - e236-r0003-s04-c00-compressed-V06-07-00-Hist16.xtc (**54 GB**)
 - e236-r0003-s04-c00-compressed-V06-07-00-HistN.xtc (**48 GB**)
- **Location for all 3 files:**
 - /reg/data/ana11/test/gapon
- **Files compressed in OFFLINE (psana202) using:**
 - xtccompress (DAQ utility from pdsdata)
 - **Hlst16:** 55% compression, 208 MB/s
 - **HlstN:** 49% compression, 187 MB/s

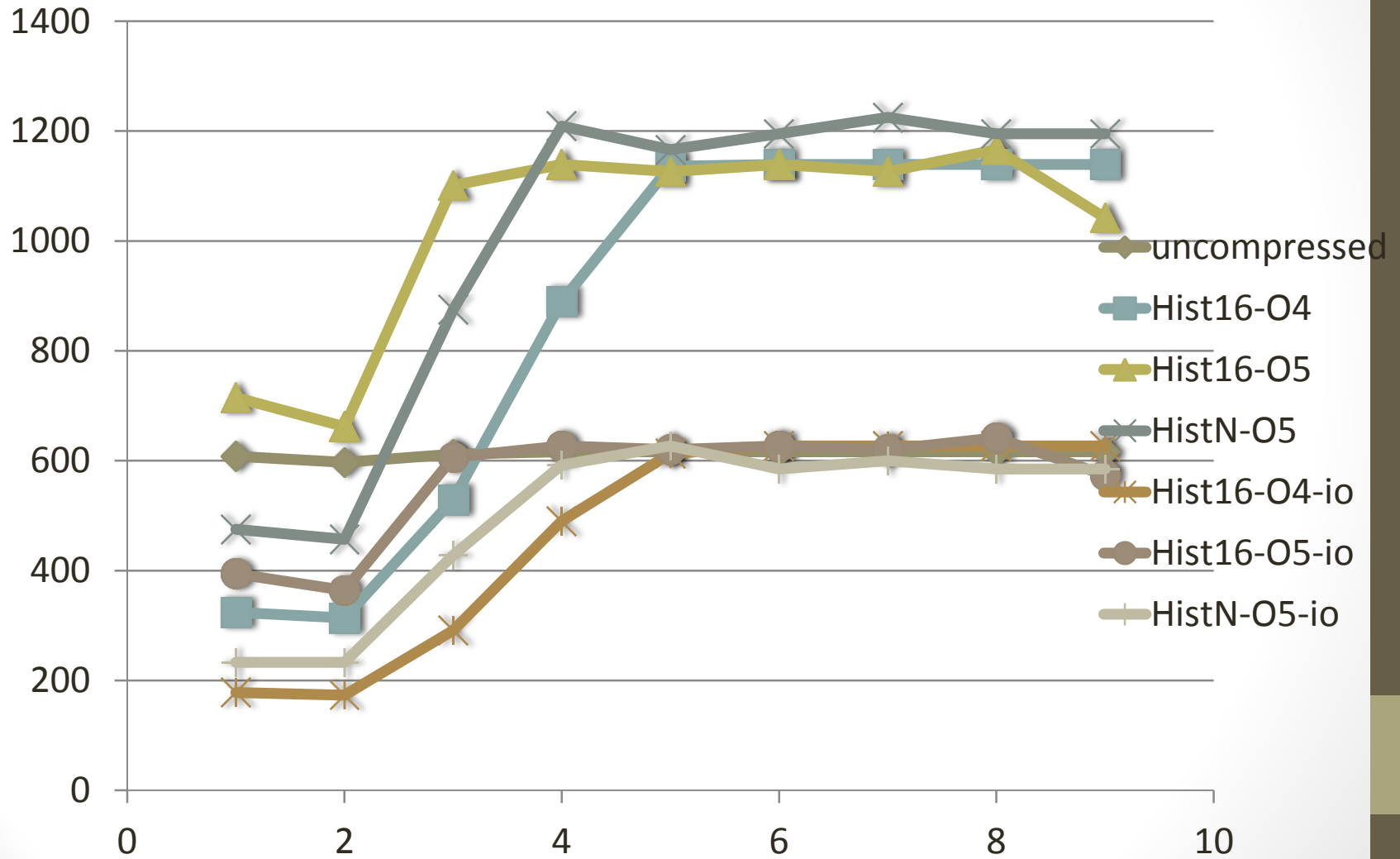
Effective Data Rate [MB/s]



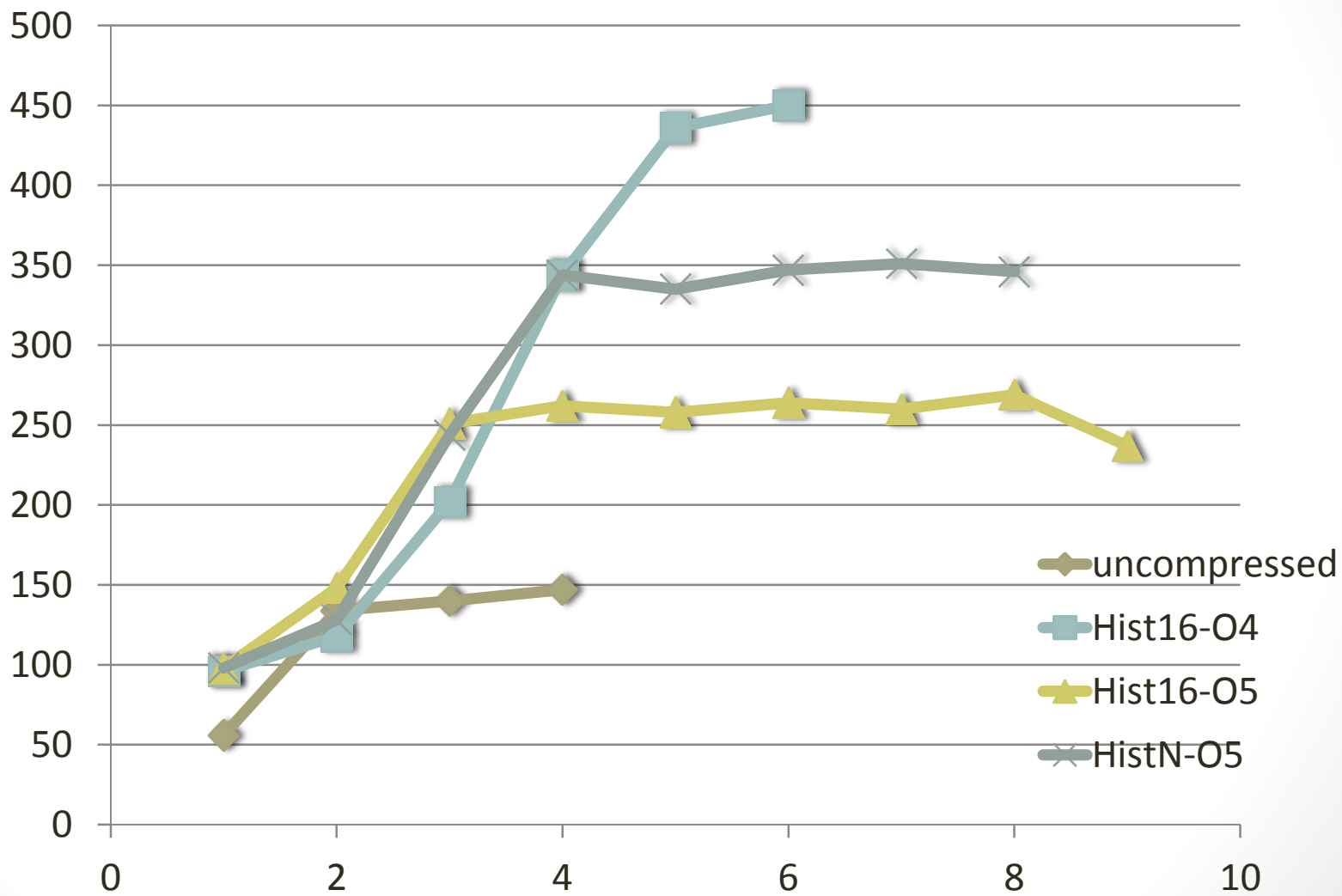
Disk I/O [MB/s]



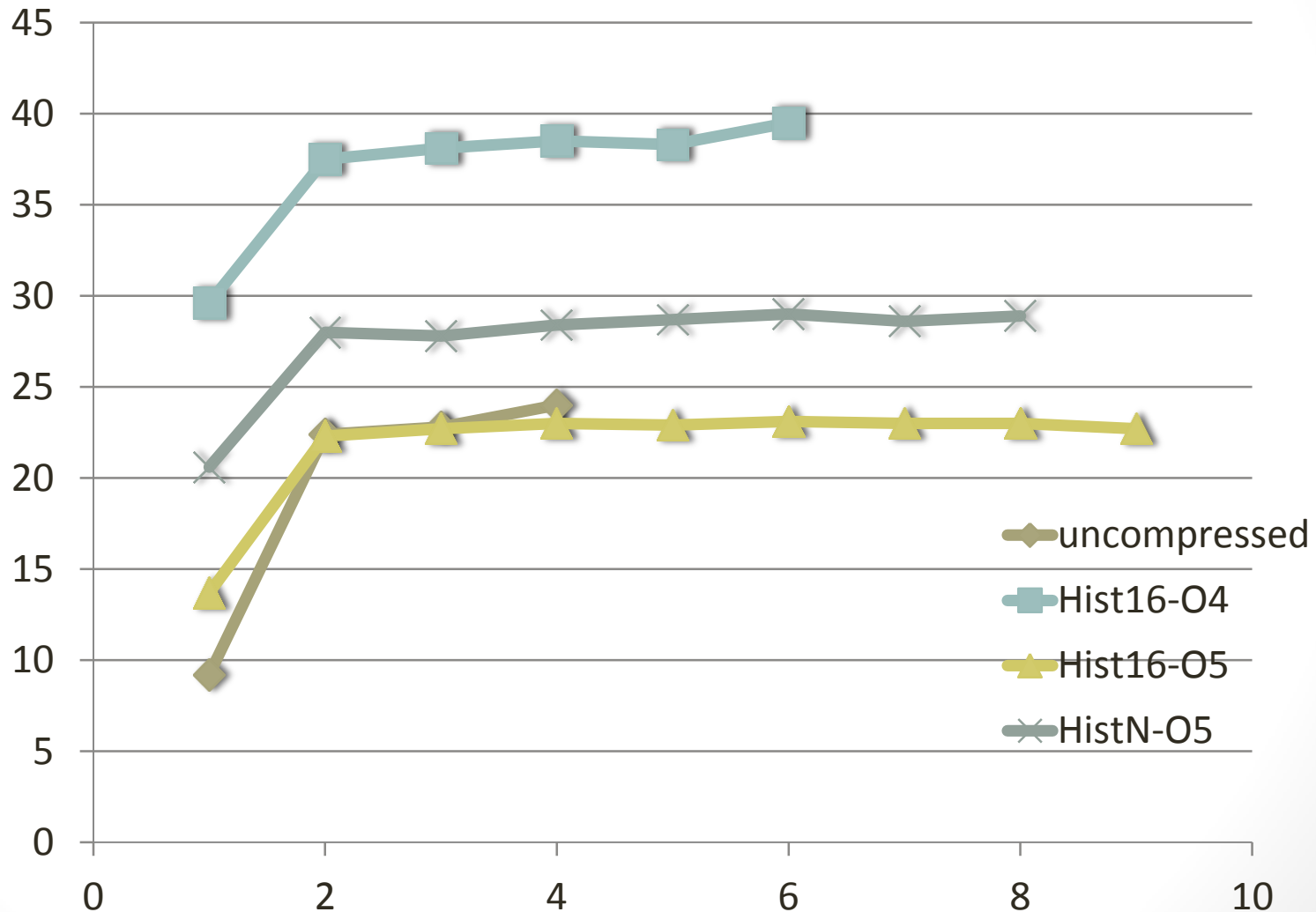
Data Rate and I/O



CPU Utilization [%]



CPU [%] / Data Rate [100 MB/s]



Summary of observations

- **Compiler optimization level has a dramatic effect on the decompression performance:**
 - Always compile with `-O5`
- **CPU consumption “price” for decompression:**
 - x2 smaller file size “costs” 50% of a CPU core
- **Compression performance:**
 - Too low for 1 thread
 - DAQ should consider a multithreaded version
- **Benefiting from multiprocessing in pyana:**
 - For both algorithms
 - 3 processes or higher (depending on an algorithm)
- **Single-threaded psana may require extra work:**
 - Asynchronous I/O, or
 - Multi-threaded decompression