Méthode de compression polynomiale

27/09/2016
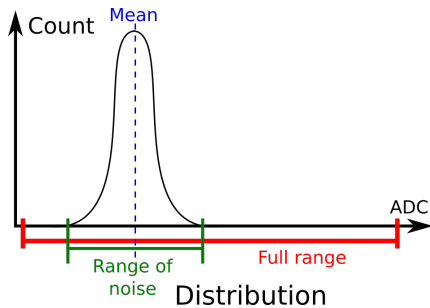
Pierre Aubert, Jean Jacquemier, Thomas Vuilllaume, Gilles Maurin, Armand Fiasson, Giovanni Lamanna

Data lossless compression

Digitalization : 16 - 24 bits
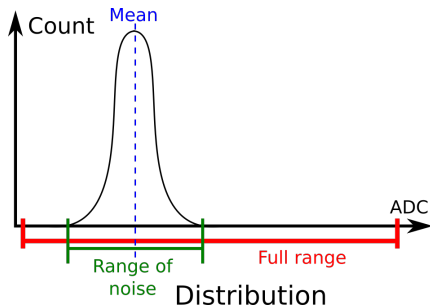short : 16 bits
unsigned int 32 bits
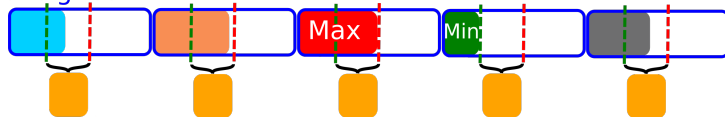
Data lossless compression

Digitalization : 16 - 24 bits
short : 16 bits
unsigned int 32 bits

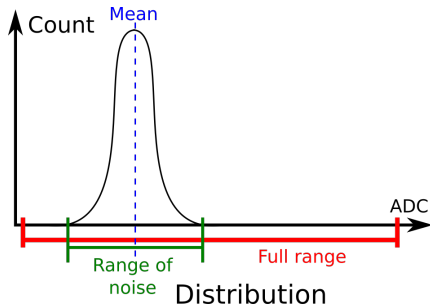# Data lossless compression

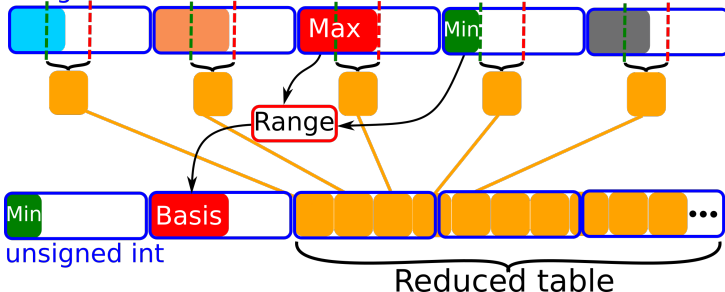
Digitalization : 16 - 24 bits
short : 16 bits
unsigned int 32 bits

Raw data Compression Ratio and speed (Lossless compression)

Test file run 497 (Paranal, Gamma Monte-Carlo CTA PROD_3)
- 475 MB
- Up to 99% ADC values

	Compression ratio	Time	File size (MB)
LZMA (7z)	4.84	7 min 48.636 s	98
Advanced Polynomial Compression	3.74	3.7 s	127

Raw data Compression Ratio and speed (Lossless compression)

Test file run 497 (Paranal, Gamma Monte-Carlo CTA PROD_3)

- 475 MB
- Up to 99% ADC values

	Compression ratio	Time	File size (MB)
LZMA (7z)	4.84	7 min 48.636 s	98
Advanced Polynomial Compression	3.74	3.7 s	127
Advanced Polynomial Compression + LZMA	**4.84**	**24.646 s**	**98**

Same ratio but faster ($\times 19$) compression !

- Vectorizable : possible faster reduction
- Can be used for any data format

Backups

Vectorizable ?

CPU Recent Architectures

Architecture	Instruction Set	CPU	Nb float Computed at the same time
SSE4	2006	2007	4
AVX	2008	2011	8
AVX 512	2013	2016	16

Easy adaptation for coming architectures

CPU Architecture
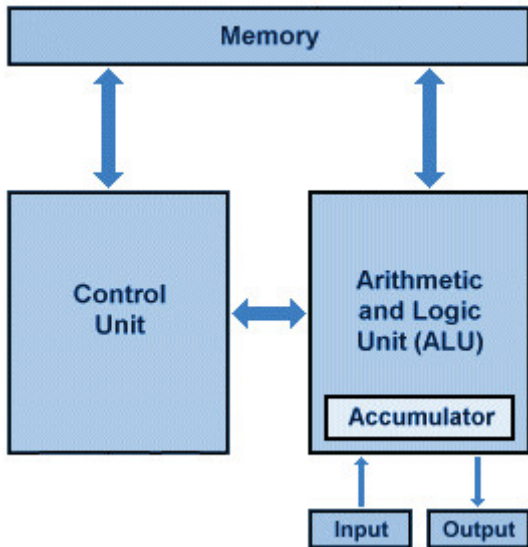
CPU Architecture

Von Neumann architecture 1945

Definition

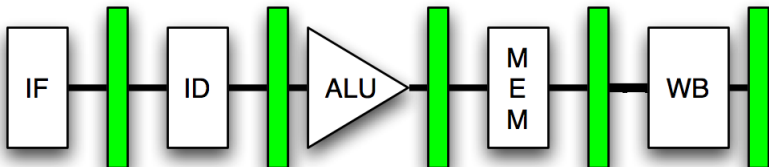Cycle : basis unit of time in a CPU

Time

- 1 cycle per elementary operation (load, store, add, ...)
- 4 cycles per whole operation ($c = a + b$)
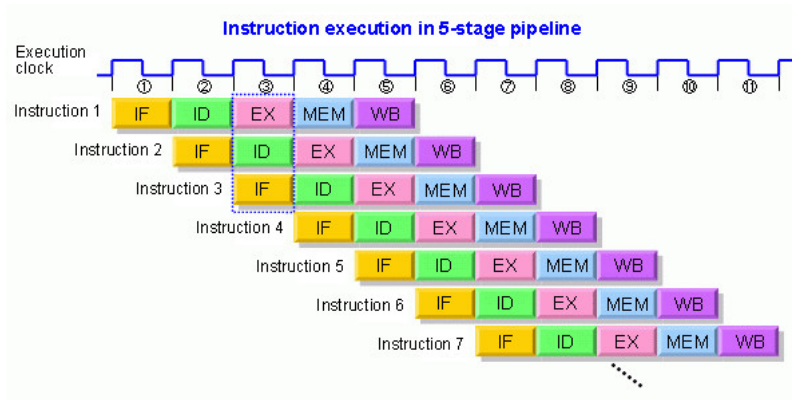
CPU Architecture

Pipeline approach

- IF : Instruction Fetch
- ID : Instruction Decode
- ALU : Execution
- MEM : Memory
- WB : Write Bytes

CPU Architecture evolution

Pipeline using

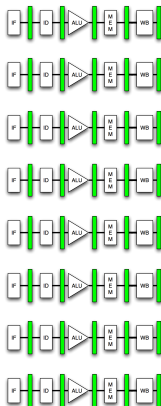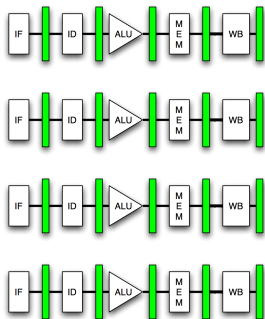Instruction execution in 5-stage pipeline

CPU Recent Architectures

SSE4   4 floats
Instruction set : 2006
CPU : 2007

AVX   8 floats
Instruction set : 2008
CPU : 2011

AVX 512   16 floats
Instruction set : 2013
CPU : 2016

Data format
Efficent only if data are contiguous