

# VIRTUALIZATION OF FLOATING-POINT COMPUTATION

DALI, Université de Perpignan  
David Defour

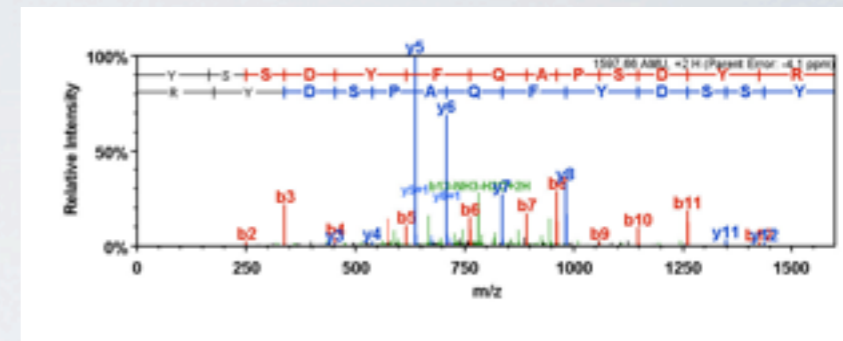
# WHAT IS SO SPECIAL ABOUT FP ?

- 3 fields with 3 different meanings:
  - Sign: Positive/Negative
  - Exponent: Do we deal with big/small numbers ?
  - Mantissa: What accuracy
- Various formats to address it (IEEE754 2008)
  - Binary32
  - Binary64
  - Binary16, Binary128 ...

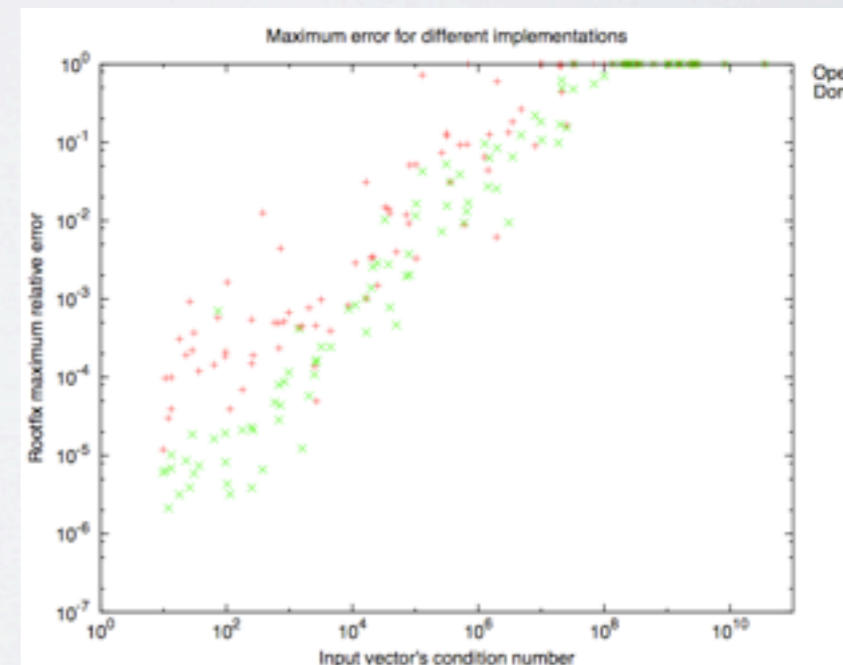


# EXAMPLES OF FP MISSUSE

- **Biology:** Data produced by MS/MS machine
  - Binary64 but analysis shows that binary32 is enough (divide storage & data transfert by 2)
- **Physics:** Power Flow Analysis
  - Fast simulation with low accuracy (eliminate irrelevant solution)
  - Accurate simulation to differentiate the remaining one
  - Accuracy depend on problem size



Joint work with Y. S. Dandass, MSU



Joint work with M. Marin, UPVD

# SOLUTIONS

- **real** format exposed to the programmer, transformed to one of the available format according to
  - Static program analysis
  - Statistical FP analysis
  - Computational time available
  - Free ressources



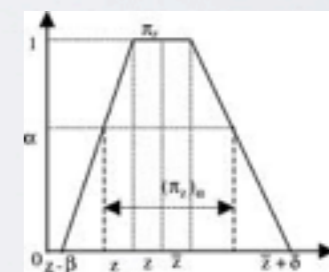
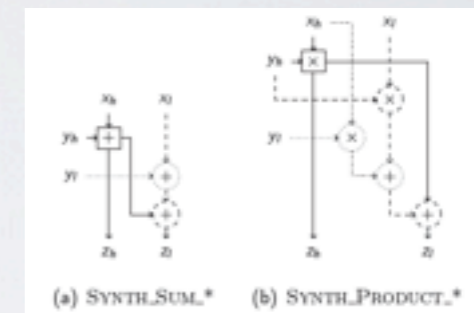
# EXAMPLES OF WORK DONE AT PERPIGNAN

- RangeLab
  - Static analysis of FP expression
- Error Free Transformation (+, x, -, /)
  - Same results in the same amount of time
  - Plus corrective terms if enough time/ressource is given
- Fuzzy Arithmetic on GPU
  - Quickly gather informations on FP distribution
- On the fly FP compression
  - Reduce data transfert without sacrificing precision

```
[ RANGE.LAB ]
Version 1.0
July 2011

RangeLab 1.0 Copyright (C) 2011 Mathieu Martel
This program comes with ABSOLUTELY NO WARRANTY; for details see
This is free software, and you are welcome to redistribute it
under certain conditions; type "show s" for details.

-> 2.140.2
```



# AEM: VIRTUALIZATION OF FP

- Master thesis subject:
  - Static translation of a generic FP format to IEEE-754 formats
- Work plan:
  - Propose different strategies  
(computation, allocated memory, mixed, ... )
  - Evaluate the impact of such modifications  
(Accuracy, Data transfert, Execution time, ...)
  -