Numerical Verification of Large Scale CFD Simulations: One Way to Prepare the Exascale Challenge

Christophe DENIS

Christophe.Denis@edf.fr
EDF Resarch and Development - EDF Lab Clamart

August 22, 2014

16^{th} IEEE International Conference on High Performance and Communications
HPC-CFD in Energy/Transport Domains Workshop

There is more to life than increasing its speed, Mahatma Gandhi.

Outline of the talk

1. On the road of the Exascale computing
2. Round-off error propagation analysis
3. Incidence of the floating point arithmetic in the numerical reproducibility
4. Incidence of the floating point arithmetic in the software quality insurance
5. Concluding remarks and future works

Outline of the talk

On the road of the Exascale computing

- Multiscale and multiphysics simulations require an increasing computing power [Duff, 2011]
  - The first Exascale supercomputers would appear between 2018 and 2020
- Architectures of supercomputer become more complex and heterogeneous to counteract the little power increase of processors.
- Parallel simulation codes running efficiently on SIMD clusters need to be adapted in order to maintain their performance on these supercomputers, for example:
  - by using asynchronous algorithms
  - by using a task based model

Possible non numerical reproducibility

- The order of the floating arithmetic operations becomes less and less deterministic from an execution to another one
- A parallel simulation code using the same data could produce different numerical results due to the lack of associativity of the floating point arithmetic.

On the road of the Exascale computing

EESI project, WG 4.3 Numerical libraries, solvers and algorithms, Iain Duff, Chair

"..a lower precision floating-point number will require less storage than one at higher precision. ...A further issue in floating-point arithmetic is the problem of reproducibility, particularly acute when computing in parallel. This problem is mainly caused by the lack of associativity in floating-point arithmetic so that the sequence of performing the arithmetic operations can influence the result. ..."

Extreme-Scale Solvers:Transition to Future Architectures, US DoE, 2012

"Research is needed to identify what classes of algorithms can benefit from the use of mixed-precision arithmetic. ...Also crucial in the presence of mixed precision arithmetic is a solid understanding of how round-off errors propagate through the software stack and how they may affect the accuracy achieved in application codes. ... "

One of (Exascale) Challenge

Ensuring jointly the numerical reproducibility, the accuracy and the performance of a parallel simulation code while minimizing its memory requirements

⇒ There is a need to estimate the effect of the round-off error propagation on a numerical simulation

Outline of the talk

Floating point computation

- Real numbers: infinite and continuous set of real values
- On a computer, the floating point computation is a model of a mathematical computation: discrete set of values.
- IEEE-754 standard (1985): arithmetic formats, interchange formats and rounding rules
- More details in *Handbook of Floating Point Arithmetic* [Muller et al., 2010]

Intrinsic problem of the floating point computation

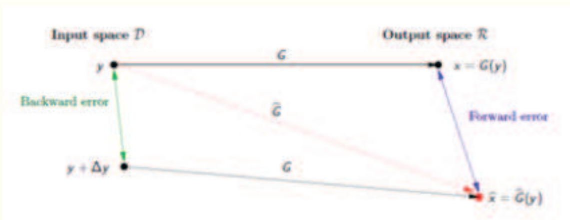
- Representation error of real values, for example the decimal value 0.1 has a representation error
- Loss of arithmetical properties: Arithmetic expressions are no longer associative and distributive.

The floating point computation as a model to be checked

- Estimation of the round-off error propagation effect on the computed results
- First step of the a rigorous V&V (Verification & Validation) process

Round-off error propagation analysis methods

- Inverse analysis [Wilkinson, 1986] [Higham, 1996]
  - the computed solution is assumed to be the exact solution of a nearby problem and provides error bounds
  - + some scientific libraries provides error bounds
  - - could not be used to perform a whole round-off error propagation analysis on a simulation code



- Interval arithmetic
  - The result of an operation between two intervals contains all values that can be obtained by performing this operation on elements from each interval.
  - + guaranteed bounds for each computed result
  - - overestimation of the error
  - - need to use specific algorithms and therefore to modify the simulation code

Round-off error propagation analysis

- Probabilistic approach
  - The CADNA library developped by LIP6 (University of Paris 6)
  - Estimation of the round-off error propagation effect on the computed results on Fortran77/90, C/C++ codes
    - Each floating point operation is computed three times using a random rounding modes
    - The number of exact significant digits is then estimated from these three samples
  - + does not need transform the code but to translate it to use CADNA datatypes and statements
  - - difficult to use CADNA on scientific libraries (BLAS, LAPACK ...)
  - The numerical verification of a whole numerical program could be found in [Scott et al., 2007].
  - Do not hesitate to contact me to have more information about using the CADNA library in an industrial context.
- After the round-off error analysis, it exists several ways to increase the precision:
  - Replace or modify algorithms leading to poor precision
  - Use compensated algorithms
    - A compensated summation uses a correction term cleverly designed to diminish the rounding errors
  - Use multiprecision arithmetic (for example MPFR)

Example provided by S. M. Rump in [Rump, 1988]

$$f(x,y) = 333.75y^6 + x^2(11x^2y^2 - y^6 - 121y^4 - 2)2 + 5.5y^8 + \frac{x}{2y} \quad (1)$$

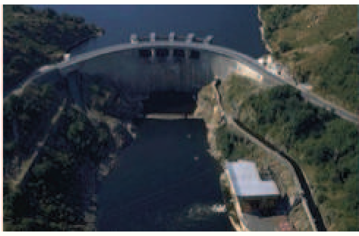
- Evaluation of $f(77617.0, 33096.0)$

Single precision	1.172603
Double precision	1.1726039400531
Extended precision	1.172603940053178
CADNA double precision	@.0 (numerical noise)
MPFI (multiple precision interval arithmetic)	[-0.827396059946821368141165..., -0.827396059946821368141165...]

Using extended precision does not always increase the accuracy of the computed results

The loss of accuracy during a numerical computation is independent of the initial precision used for floating point numbers.

Overview of the CFD TELEMAC-MASCARET software suite



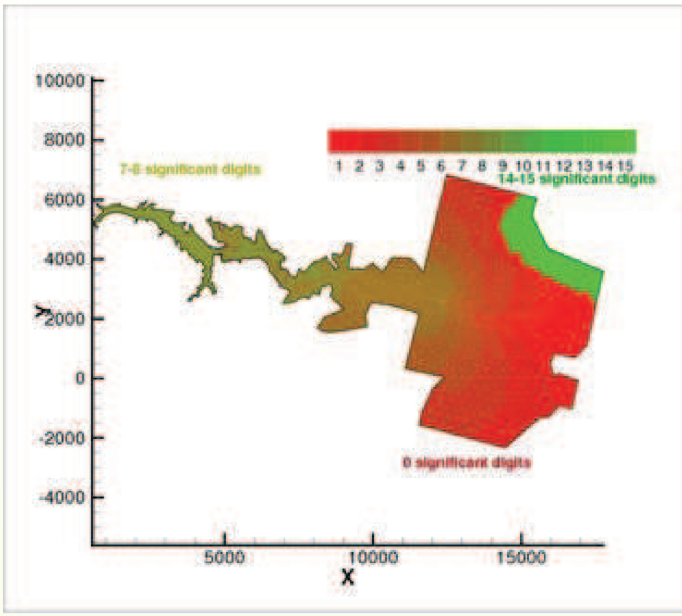
- Contains several CFD simulation codes dedicated to free-surface flows [Hervouet, 2007]
  - Developed since 1987 by EDF R&D through a scientific consortium
  - Open source license
- Example of simulation codes contained in this software suite
  - Telemac-2D
    - 2D free surface flows
    - Shallow water equations, finite element method
    - Parallel version based on a domain decomposition
  - Telemac-3D
    - 3D free surface flows
    - Navier-Stokes equations, finite element method
    - Parallel version based on a domain decomposition

Numerical verification framework

- The purpose is to indicate the confidence level that one can have in the results.
- The TELEMAC-2D code has been implemented with the CADNA library
  - TELEMAC-2D does not use external scientific libraries such as BLAS or LAPACK
  - the implementation of CADNA on a external scientific libraries could be difficult as
    - the overhead on computing time could be prohibitive on some optimized libraries
    - or even impossible if its source code is not available

Numerical verification framework: First example

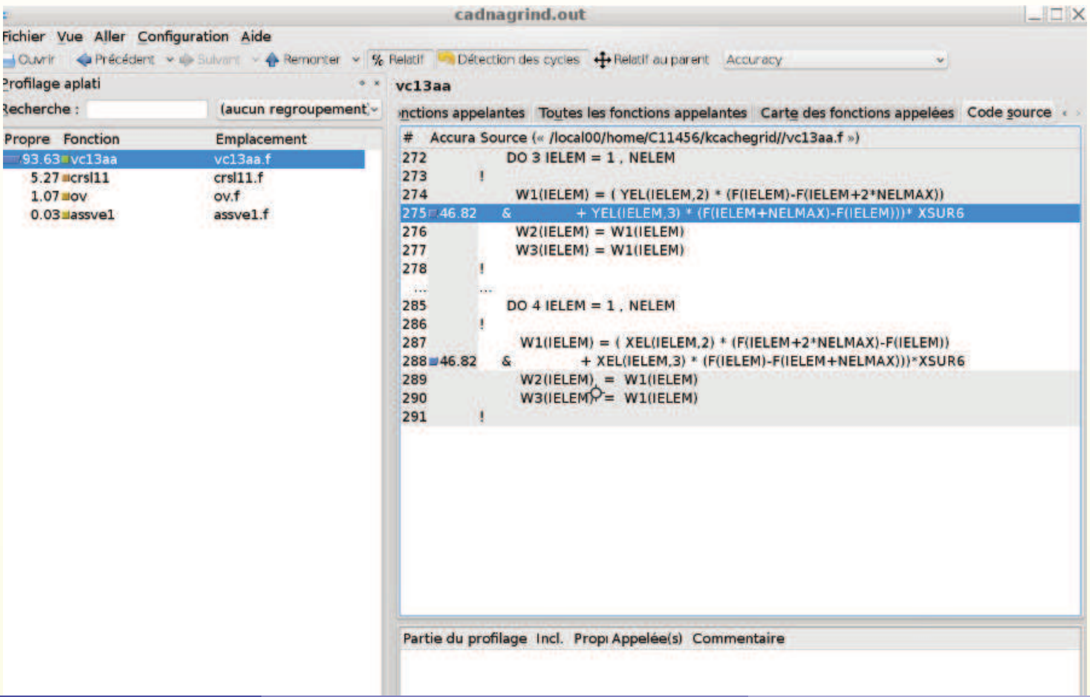
- Visualisation of the numerical quality of the computed values (xD+P approach [Denis, 2013])
  - Example: Precision of the water field computed with TELEMAC-2D (malpasset dambreak test case)



  - Here, zero significant digit does not highlight a numerical problem as the corresponding sample computed by CADNA is $(0, 0, \epsilon)$ where ϵ is a small value close to the precision machine.

Numerical verification framework: Second example

- Example of the report provided by the CADNA Library
  - There is 1 numerical instability ... 1 UNSTABLE CANCELLATION
- Development at EDF R&D of a CADNA post-processing tools (performed by R. Picot) based on a tool included in the Valgrind software suite
  - to indicate in the source code the location and the type of the numerical instability detected by the CADNA library



Outline of the talk

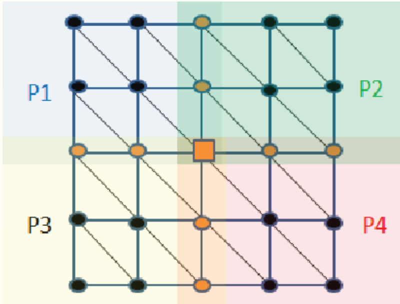
Incidence of the floating point arithmetic in the numerical reproducibility

An investigation has been performed to highlight the source of non reproducibility:

1. The first source of non reproducibility comes from the parallel dot product
   - Each processor computed its local dot product
   - This local value is then gathered with a MPI reduction.
   - The order of floating point operations to compute the global dot product differs from a parallel running to another one leading to different values.
2. The second source of non reproducibility come from the gathering of the value on interface nodes:
   - In the TELEMAC communication scheme, a subdomain adds successive contributions coming from the other processors by using MPI communication at each time step
   - Since the MPI communication is asynchronous, the order of summation depends on the network topology and activity.
   - Deadlock between processors during a Telemac-2D simulation with 4192 processors.

Incidence of the floating point arithmetic in the numerical reproducibility

- The problem could appeared during the gathering process on the interface node shared by four subdomains (the square node)



- A summation order has been arbitrarily fixed to suppress the deadlock
- Several sums has only one decimal significant digits (estimation coming from the CADNA library)
- Preliminary tests has shown that compensated summation algorithms could help to obtain better reproducibility

Outline of the talk

1. On the road of the Exascale computing
2. Round-off error propagation analysis
3. Incidence of the floating point arithmetic in the numerical reproducibility
4. Incidence of the floating point arithmetic in the software quality insurance
5. Concluding remarks and future works

Incidence of the floating point arithmetic in the software quality insurance

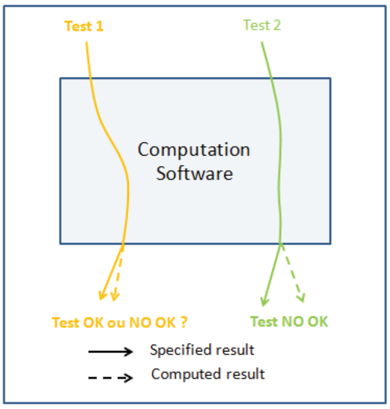


Figure: The dilemma of the orange test : OK or NO OK?

- The green test is not valid as the specified result is very different from the computed one: a correction in the source code is required.
- The main difficulty is to decide if the orange test is valid or not
- Solution
  - The functional specifications have to define the expected number of significant digits on the solution
  - This number of significant digits could then be estimated thanks the CADNA library.
- Some parts of the code will be able to be computed in lower precision if it leads to have the expected number of significant digits to reduce the amount of memory [Baboulin et al., 2009]

Outline of the talk

Concluding remarks and future works

- Concluding remarks
  - This talk has presented an overview of the numerical verification activities performed at EDF R&D
  - The CADNA library has been used to estimate the effect of the round-off error propagation
  - It helps to better understand the incidence of the floating point arithmetic on the numerical reproducibility of parallel simulation codes
  - There is also a need to improve the software development chain adding functional tests based on the numerical precision.
- Future works
  - Estimation of the round-off error propagation on external libraries without having instrumenting them
  - Definition of a methodology to improve the numerical reproducibility of a numerical code without both affecting its accuracy and its performance.
  - Numerical verification of mixed precision algorithm designed to reduce the amount of memory

References I

Baboulin, M., Buttari, A., Dongarra, J., Kurzak, J., Langou, J., Langou, J., Luszczek, P., and Tomov, S. (2009).
Accelerating scientific computations with mixed precision algorithms.
Computer Physics Communications, 180(12):2526–2533.

Denis, C. (2013).
Numerical verification of an industrial code to simulate accurately large scale hydrodynamics events.
International Water Journal, (2):46–51.

Duff, I. S. (2011).
European exascale software initiative: Numerical libraries, solvers and algorithms.
Technical Report TR/PA/11/118, CERFACS.

Hervouet, J.-M. (2007).
Hydrodynamics of Free Surface Flows.
Wiley-VCH Verlag GmbH & Co. KGaA.

References II

Higham, N. J. (1996).
Accuracy and Stability of Numerical Algorithms.
SIAM, Philadelphia.

Muller, J.-M., Brisebarre, N., de Dinechin, F., Jeannerod, C.-P., Lefèvre, V., Melquiond, G., Revol, N., Stehlé, D., and Torres, S. (2010).
Handbook of floating-point arithmetic.
Birkhäuser Boston Inc., Boston, MA.

Rump, S. (1988).
Reliability in Computing. The Role of Interval Methods in Scientific Computing.
Academic Press.

Scott, N. S., Jézéquel, F., Denis, C., and Chesneaux, J.-M. (2007).
Numerical 'health check' for scientific codes: the CADNA approach.
Computer Physics Communications, 176(8):507–521.

References III

Wilkinson, J. (1986).
Error analysis revisited.
IMA Bulletin, 22.