Profiling Arpege, Aladin and Arome … and Alaro !

R. El Khatib
with contributions from CHMI

Aladin workshop & Hirlam all staff meeting
Utrecht, 12-15 May 2009

METEO FRANCE
Toujours un temps d'avance

Outlines

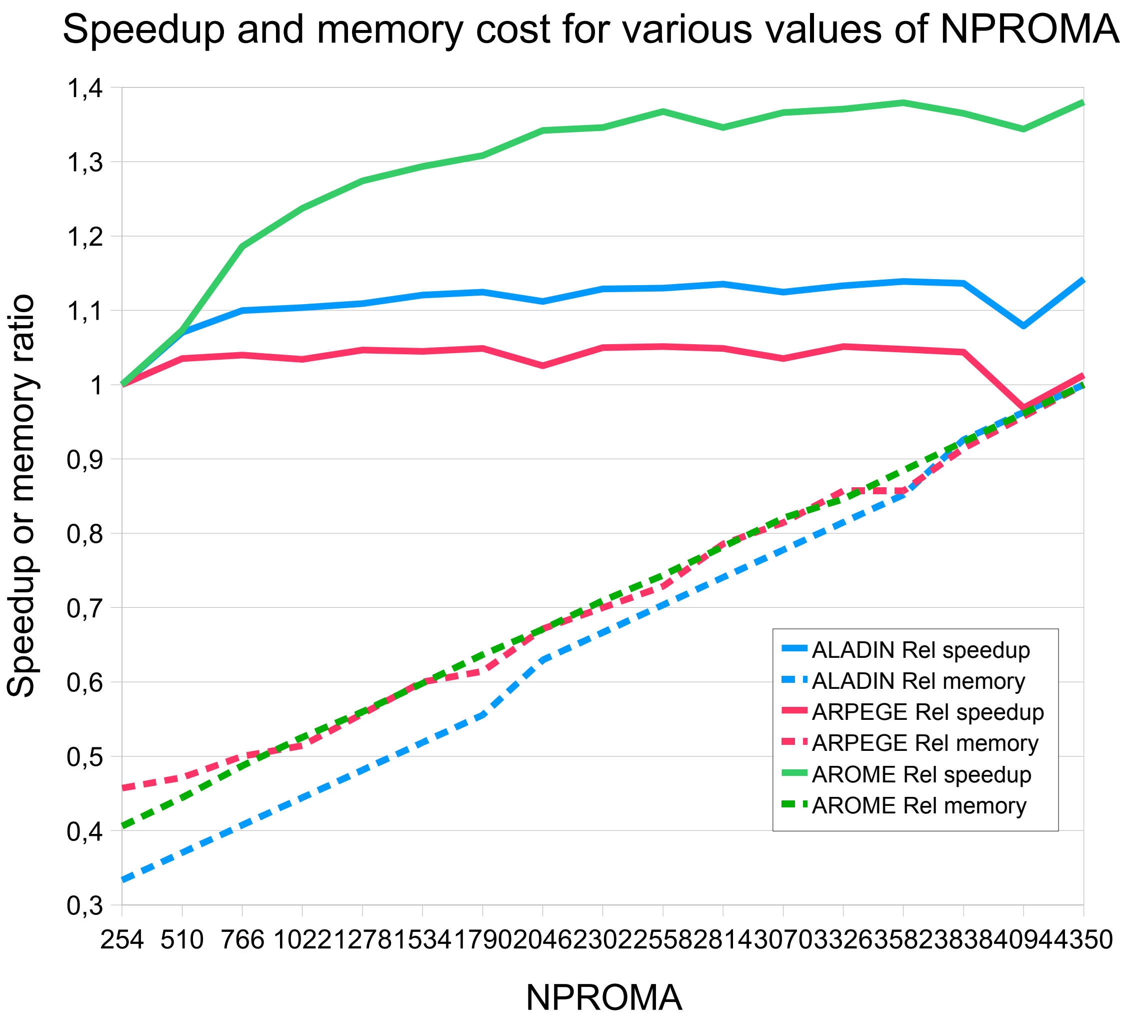
- **What's new regarding computational performances since Brussels 2008 ?**
  - About the results shown last year
  - Optimisations progress
  - Discussion about profiling tools
  - ECMWF HPC workshop (Nov, 2008)
  - What's new in the environment at and nearby Météo-France
- **A benchmarkers' « Mitraillette »**
  - Purposes
  - Overview of the procedure (as it is for now)
  - Some results
  - Incoming developments
- **Conclusions**

What's new … : About the results shown last year

- Results shown last year have been quickly revisited
  - using proper benchmarking conditions
  - updating with a more recent cycle (cycle 33 or 35)
  - updating with the latest operational namelists

- All results confirmed, except :
  - Tuning of Communication buffer length (NCOMBFLEN) : impact too weak
  - North-South distribution : always better on vector machine, even for LAMs.
  - NPROMA retuned (for vector machines)

NPROMA retuning for vector processors


Speedup and memory cost for various values of NPROMA

Recommended
Value :
NPROMA=3582

To be avoided :
1022, 2046, 3072
(banks conflicts in RRTM)

No inflation
of the memory used

Arome :
a problem of overhead
localized in Surfex

What's new … : Optimisations progress

- **Progress :**
  - Fullpos on-line
    - Miscellaneous bugfixes (cycle 35T1)
    - Saves up to 5 % elapse time for the same enveloppe of ressources (<=> « ressource sharing »)

      Operational at Météo-France for ARPEGE, soon for AROME, maybe later for ALADIN
  - Improved support for using different file system (cycle 35T2)
    - Namelist variables to setup full path of output files
    - => jobs can run on a local file system
- **Stand-by :**
  - Surfex initial file reading :
    - optimisation still bugged.

      Interface with ALADIN should be deepely revisited.

What's new … : about profiling tools

- **DrHook used as a *basic* profiler (code-embedded) :**
  - Seems to work on any platforms
  - But implementation missing on 'externalized' software
  - Inplementation uncomplete in the internal parts uti/, xla/, xrd/

  Are we able to *assume a semi-automatic instrumentation of the code where it is missing ?*

- **Machine-specific profilers :**
  - OK, but *do not profile the message passing library (mpl) !*
  - *Complementary to DrHook and often more informative*

- **GSTAT specific profiler (code-embedded) :**
  - Developped & used on IFS at ECMWF
  - Is it worth investing human ressources in a non-automatic profiler ?

What's new ... : ECMWF HPC workshop (Nov, 2008)

- **Tremendous increasement of electricity and cooling needed if we follow the Moore's law**
  - *>> What will the next generation computers look like ? Possible concepts :*
    - « Many-core »
    - Hybrid scalar-vector architectures
    - Heterogenous CPUs (specialized processors ...)
- **Bottlenecks in source code are : scalability and I/Os (volume, access)**
  - *Many projects launched to work on super-scalability of softwares*
  - *Langages possible evolutions, like :*
    - PGAS (one-sided communications)
    - Fortran coarrays concept (« virtualisation » of message passing)

What's new … : Environment nearby Météo-France

- **Computers :**
  - New computer at Météo-France : NEC SX9 (previous : SX8R)
  - New computer at ECMWF : IBM power6 (previous : Power5)
  - Update of a Linux cluster in the research center of Météo-France
  - Access to a IBM Blue Gene at CERFACS (4096 processors)
  - Access to other kinds of machines may be possible

Interesting opportunities, isn't it ?

- **Source code novelties :**
  - RTTOV9 since cycle 35 : expected to be better optimised
  - AEOLUS (lidar project) : performance to be investigated

Benchmarker's « Mitraillette » : Purposes

A testbed ready for continuing benchmarking :

- To control the evolution of computational performances from one source code release to another
- To find out the optimal namelist tuning for computational performances
- To anticipate optimisations problems at higher resolutions

- To anticipate the adequation of the software on the latest generation machines (RAPS or other projects)
- To be prepared for the coming Invitations To Tender

Benchmarker's « Mitraillette » : Overview (as it is for now)

A tree of data files and basic shell scripts :

build/ tools/ run/ data/

run/ : release_1/ release_2/ … release_$n

release_$n : conf_1/ conf_2/ … conf_$n

conf_$n : data@../../data namelists/ Job_1/ Job_2/… Job_$n

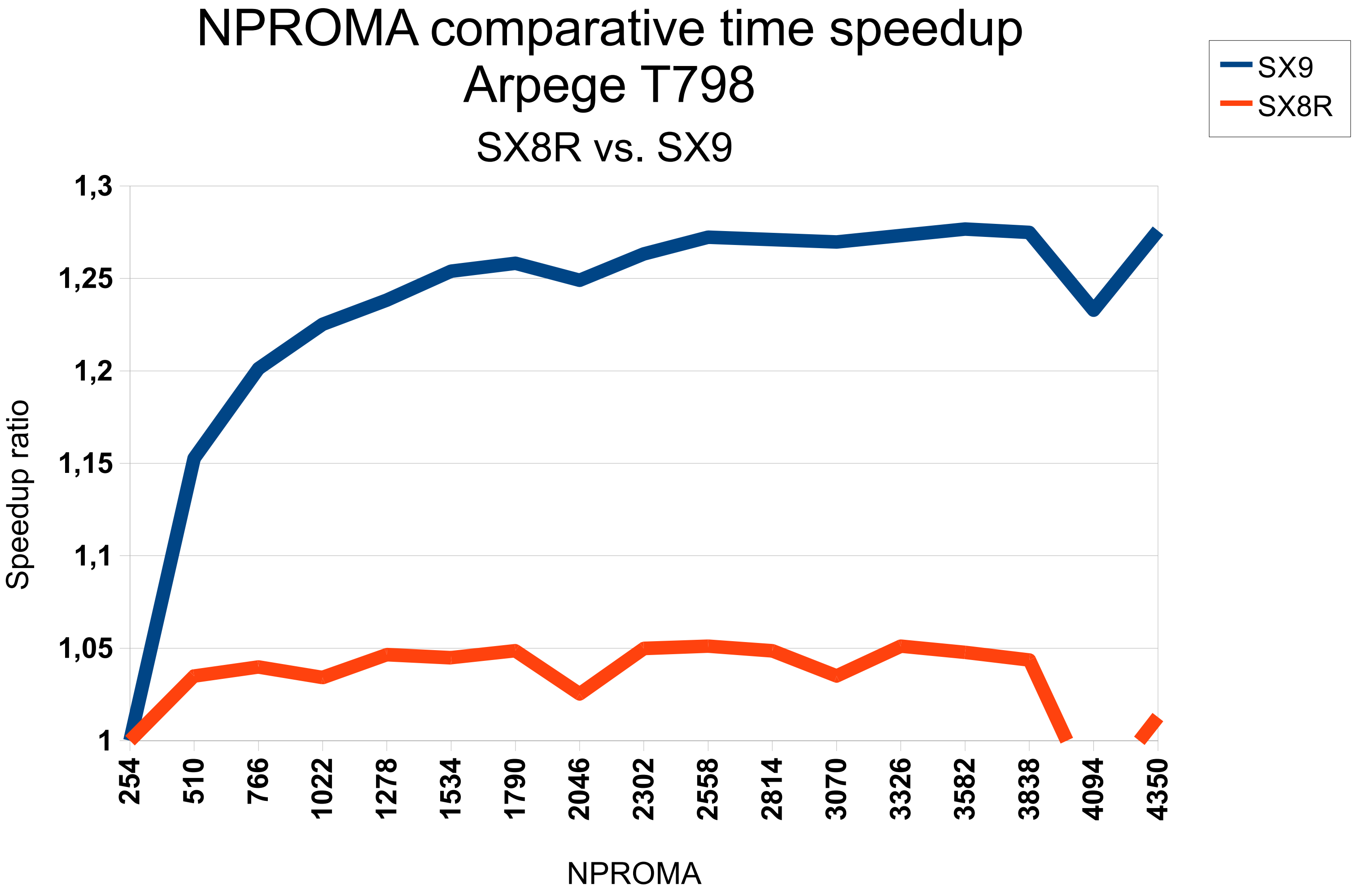
Job_$n/ : script_1 script_2 … script_$n

Building executable is not (or not yet ?) part of this procedure

Benchmarker's « Mitraillette » : jobs implemented

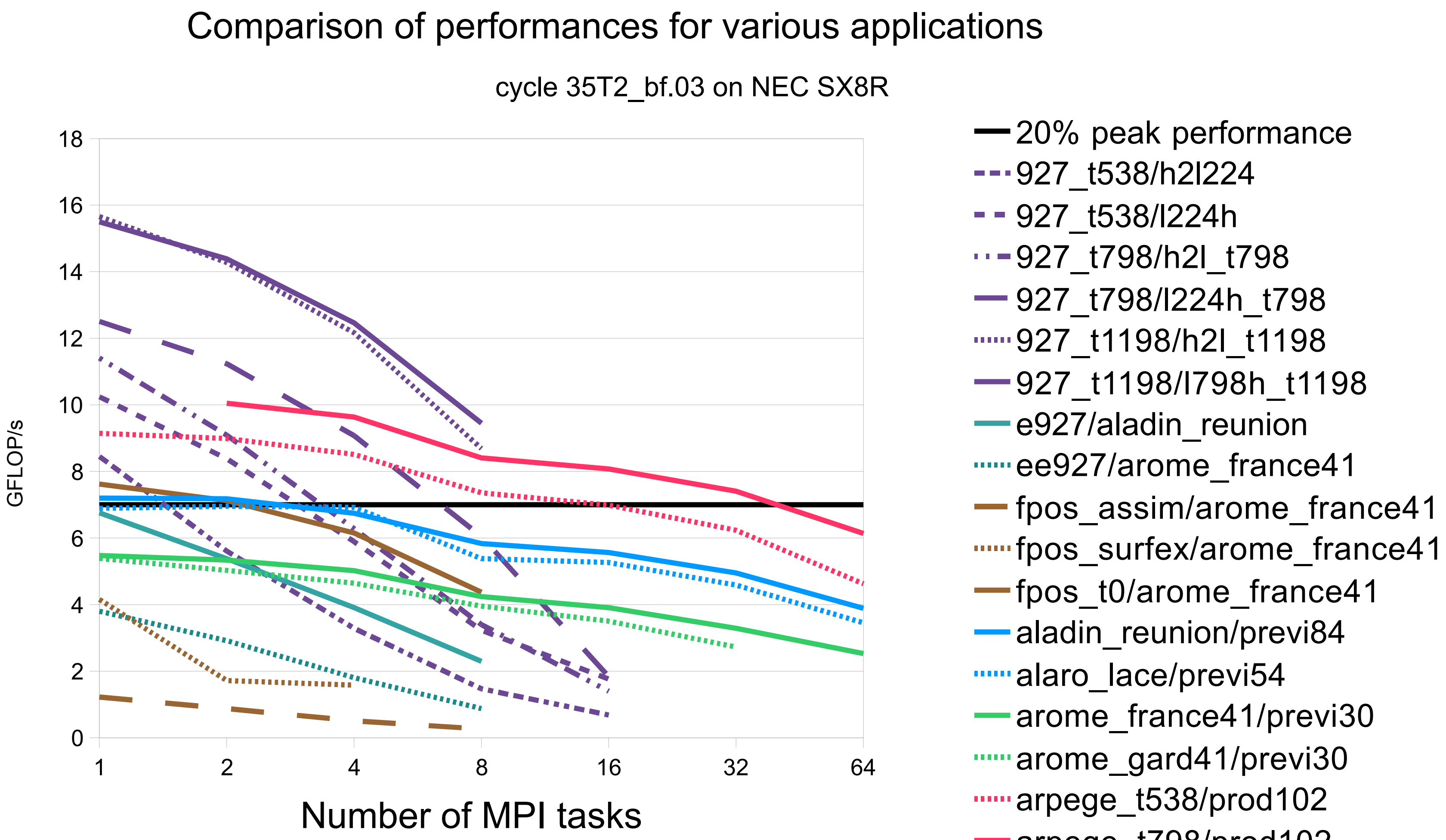
- Supported for : NEC SX8R and SX9
- Releases :
  - cycle 33T1
  - cycle 35T2
- Configurations :
  - ALADIN-Reunion incl. Fullpos on-line
  - ALARO-LACE extended domain
  - AROME-France & AROME-Gard (small size) incl. Fullpos on-line
  - ARPEGE T538 & T798 incl. Fullpos on-line
  - Various Fullpos conf. 927, e927, ee927
- Jobs :
  - DrHook profiler
  - 'Ftrace' = specific profiler on NEC
  - Scalability test (running from 1 to n processors)
  - NPROMA tuning on ARPEGE

Some results : NPROMA (again !)



METEO FRANCE
Toujours un temps d'avance

Intrinsic performance : Number of floating point operations per second

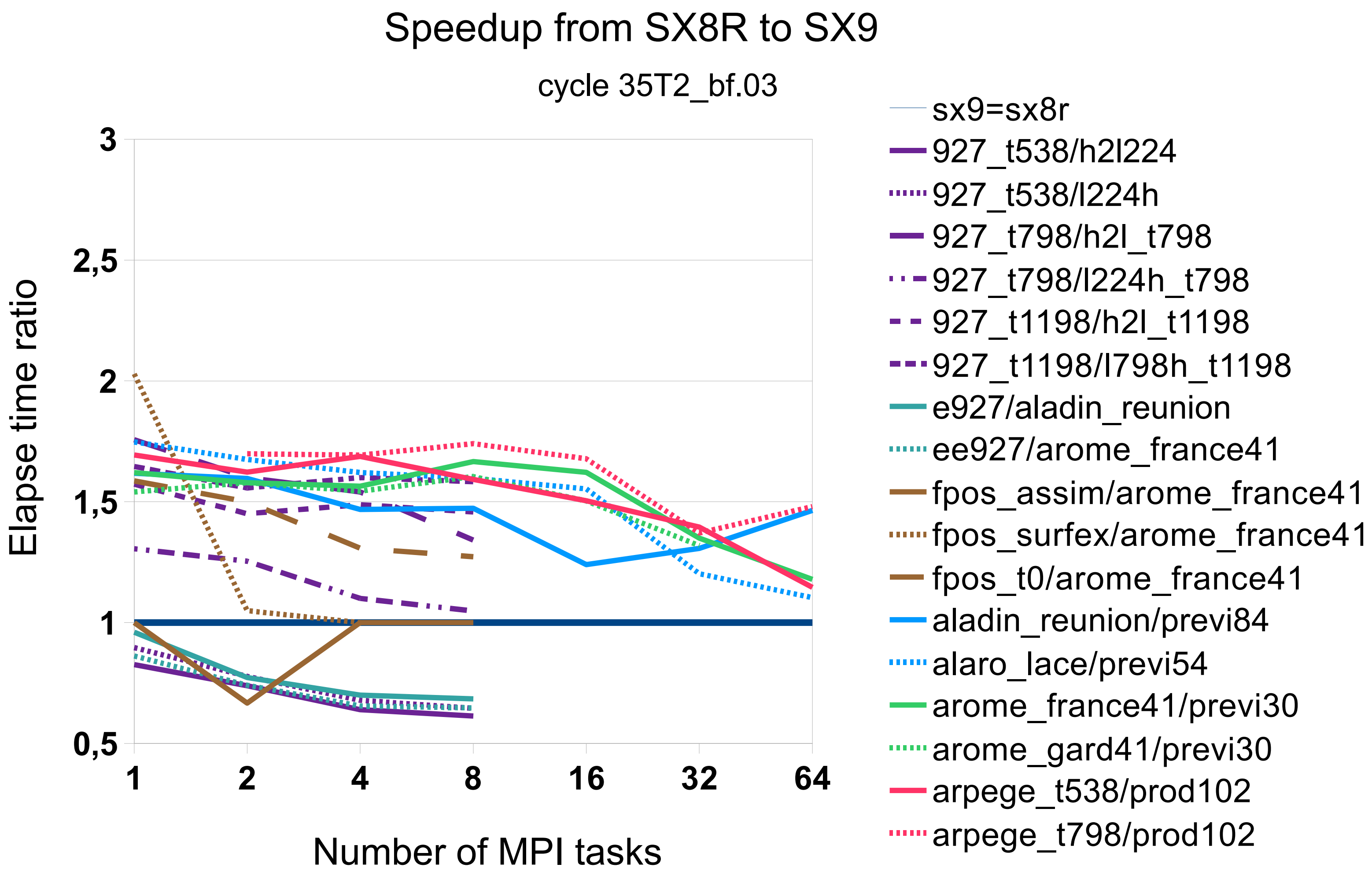


<u>Number of processors used in operations at MF :</u>

ARPEGE=8 ALADIN=4 AROME=56

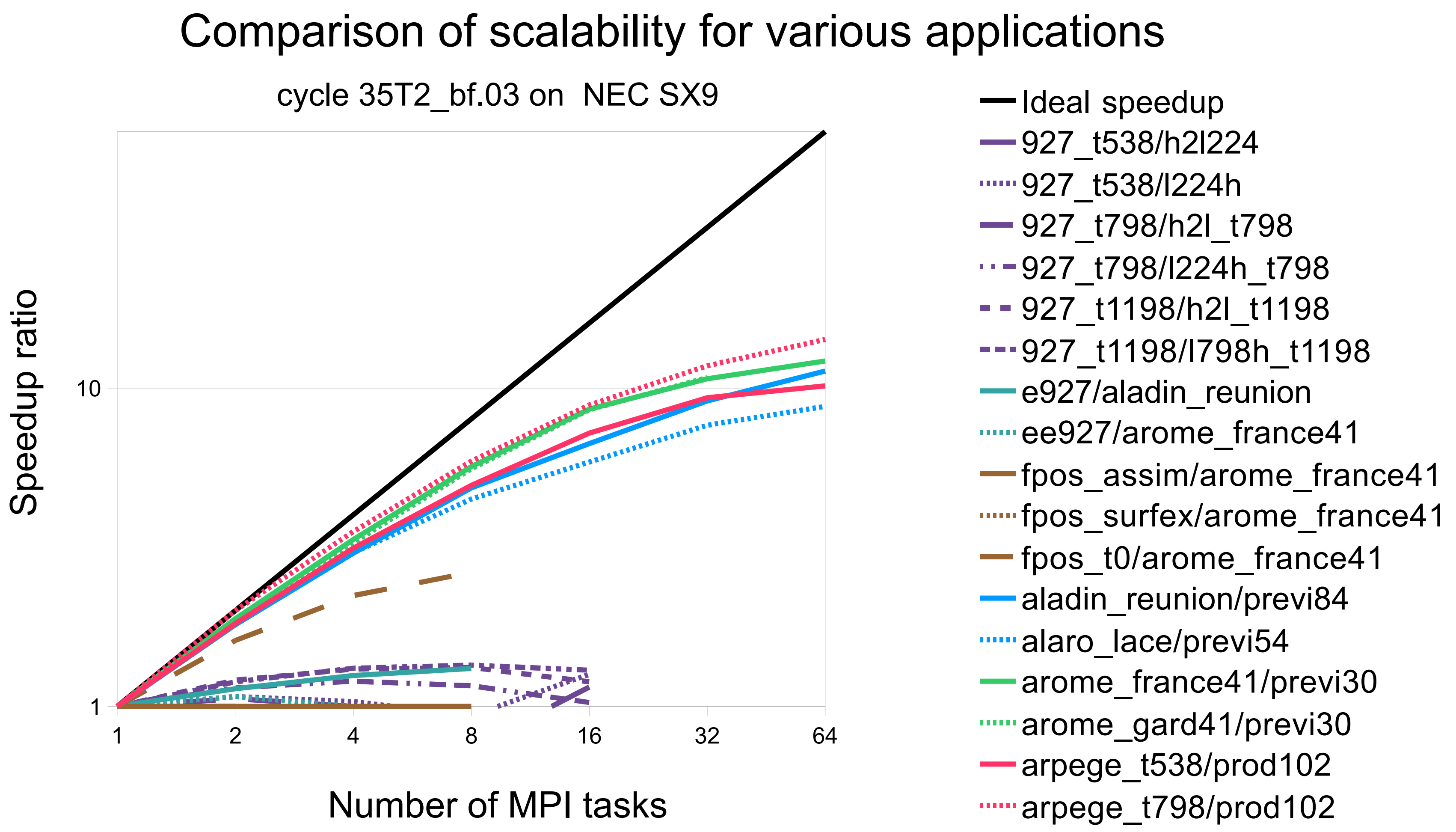
METEO FRANCE
Toujours un temps d'avance

Relative speedup



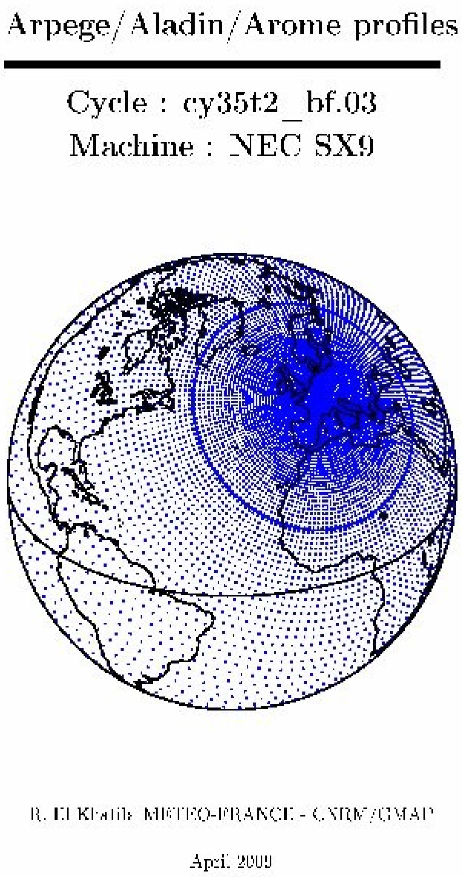
METEO FRANCE
Toujours un temps d'avance

Scalability



METEO FRANCE
Toujours un temps d'avance

Detailed (per routine) profiles



Arpege/Aladin/Arome profiles

Cycle : cy35t2_bf.03
Machine : NEC SX9

R. El Khatib METEO-FRANCE - CNRM/GMAP

April 2009

- Profilers notebooks
  - per release
  - per machine
  - For all the configurations of the software
  - generated by automatic extraction of the profiles

Benchmarker's « Mitraillette » : next steps

- Extend this procedure for
  - At least one scalar machine (IBM Power6 at ECMWF)
  - All variational configurations (3DVar, 4DVar)
  - More namelist parameters tuning (MPI distributions, OMP parallelisation)
  - Incoming cycle 36
- Prepare a package for vendors (RAPS)
  - Eventually easier to port to various platforms
  - Containing Arome 3DVar + forecast
- Find new graphical representations of performances
  - « camemberts » from DrHook profiles ?

Besides :

- … Find time to study the detailed profiles and optimise !
- And play tennis again ;-)

Conclusions

- Optimisations work has progressed very slightly. Still a lot of things to.
- Profilers are helpful but require maintenance effort
- Something is happening on the computers side : keep an eye on this.
- Source-code and machines are perpetually changing

  => Optimisation is a never-ending story

  => Better anticipate than cure
- A counterpart of « mitraillette » is proposed to control the computational performances
  - Still under developement
  - Any idea & contributions welcome !

METEO FRANCE

Toujours un temps d’avance