Zbigniew Piotrowski
Andrzej Wyszogrodzki
Piotr Smolarkiewicz

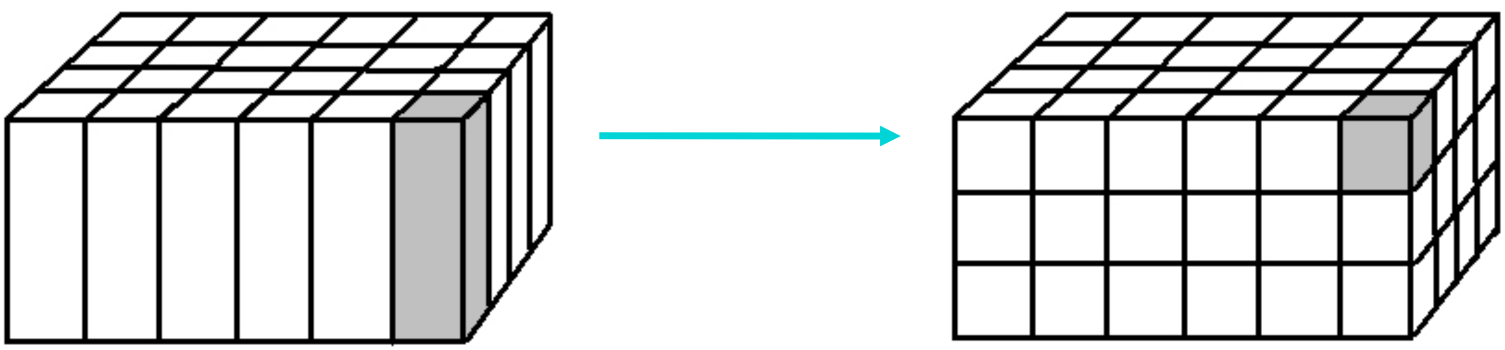
Towards petascale simulation of atmospheric circulations with soundproof equations

Institute for Meteorology and Water Management, Warsaw, Poland
National Center for Atmospheric Research, Boulder, CO

Extending MPI parallel formulation of multiscale anelastic model Eulag for geophysical flows

Traditional 2D decomposition

Newly developed 3D decomposition

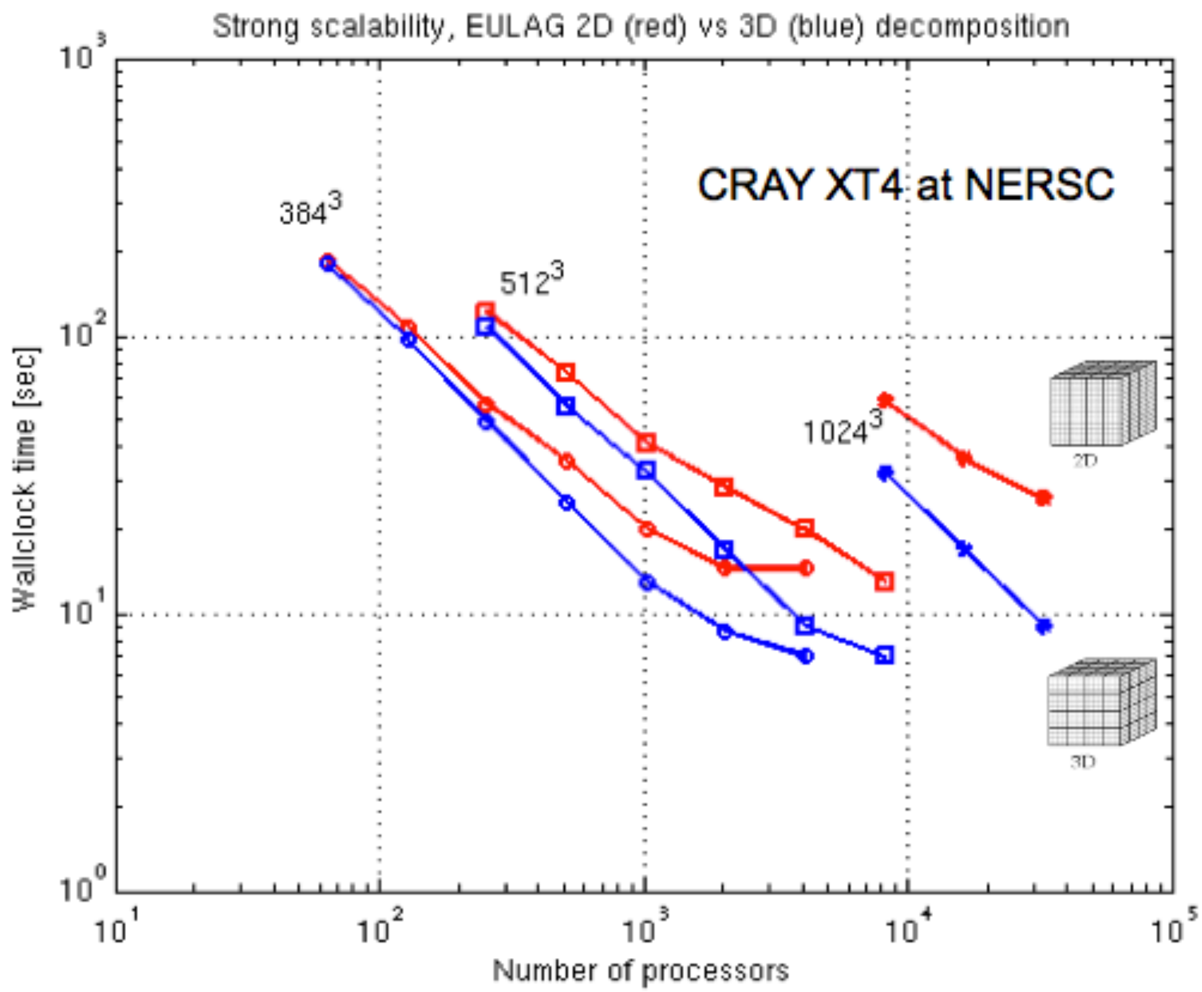


Test experiments in the range of scales from micro to planetary

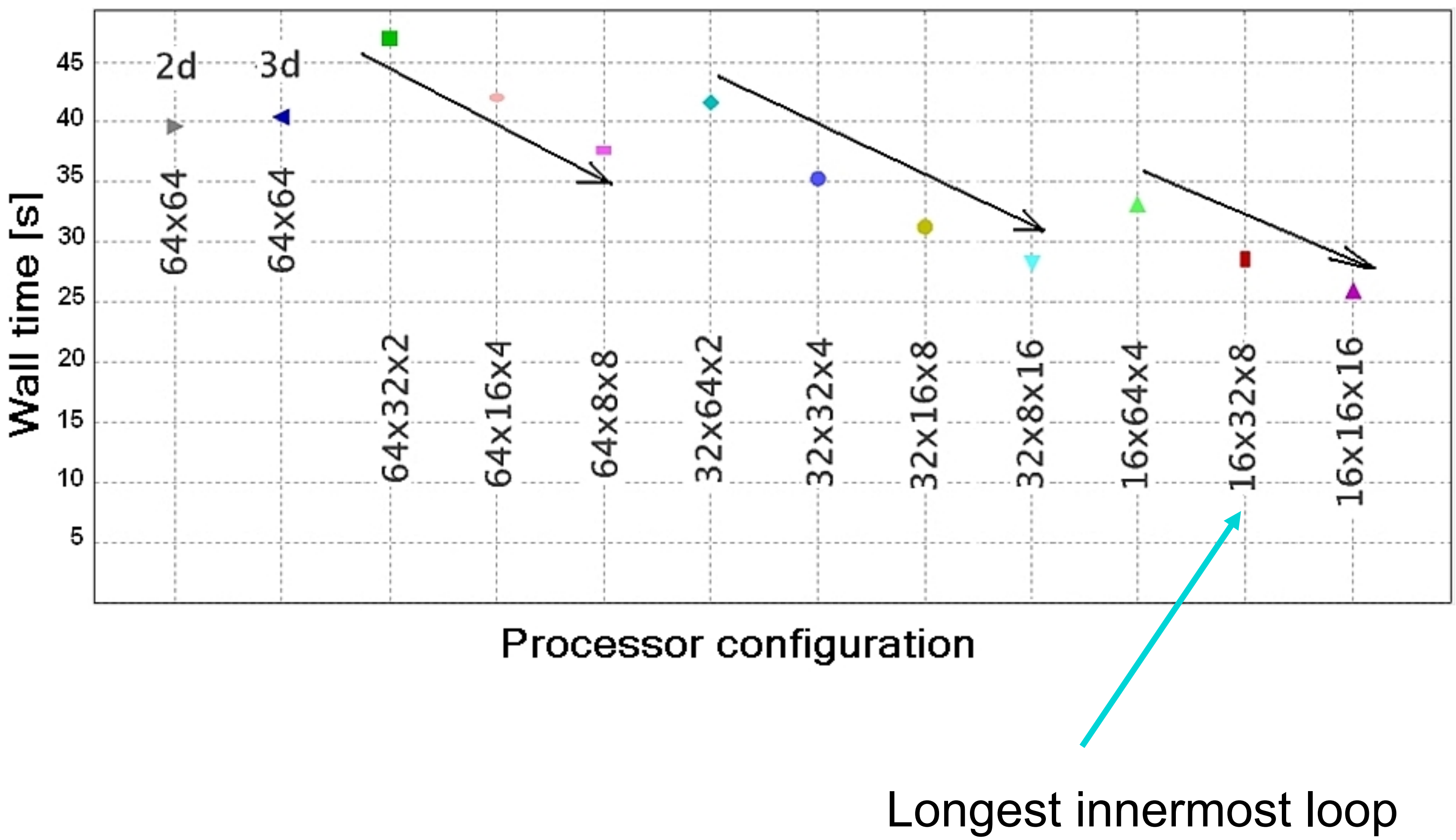
- Triply periodic decaying turbulence in the box (Taylor – Green vortex) – perfect candidate for three-dimensional parallelization

- Idealized climate simulations in a thin atmosphere shell (after Held Suarez, 1994) – traditionally decomposed in two horizontal dimensions

Strong scaling -
triply periodic decaying turbulence on Cray XT4



512^3 gridpoints decaying turbulence - dependence of performance on the processor configuration on Bluegene/L



Performance model for halo communication bandwidth usage

Examine $R = r_{3d}/r_{2d}$, where:

r3d = [(np3d + 2h) × (mp3d + 2h) × (lp3d + 2h) − V3d]/V3d
r2d = [(np2d + 2h) × (mp2d + 2h) × lp2d − V2d]/V2d .

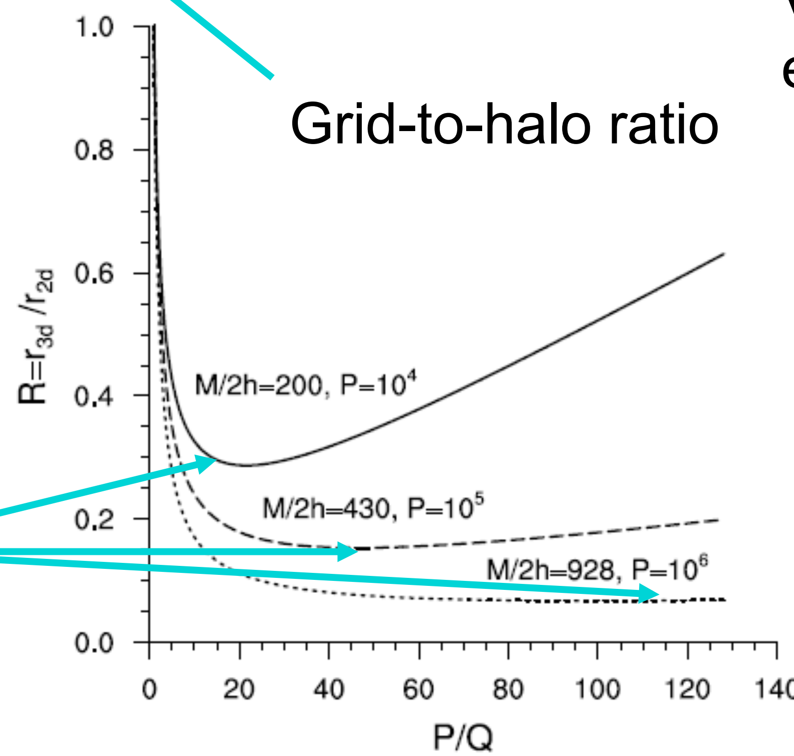
$$R\left(\frac{P}{bQ};\, P,\, \widetilde{M}\right) = \frac{\left(1 + \frac{\sqrt{P/b}}{\widetilde{M}}\sqrt{\frac{bQ}{P}}\right)^2 \left(1 + \frac{a}{\widetilde{M}}\frac{P}{bQ}\right) - 1}{\left(1 + \frac{\sqrt{P/b}}{\widetilde{M}}\right)^2 - 1},$$

No. of cores in vertical

Total no. of cores

Grid-to-halo ratio

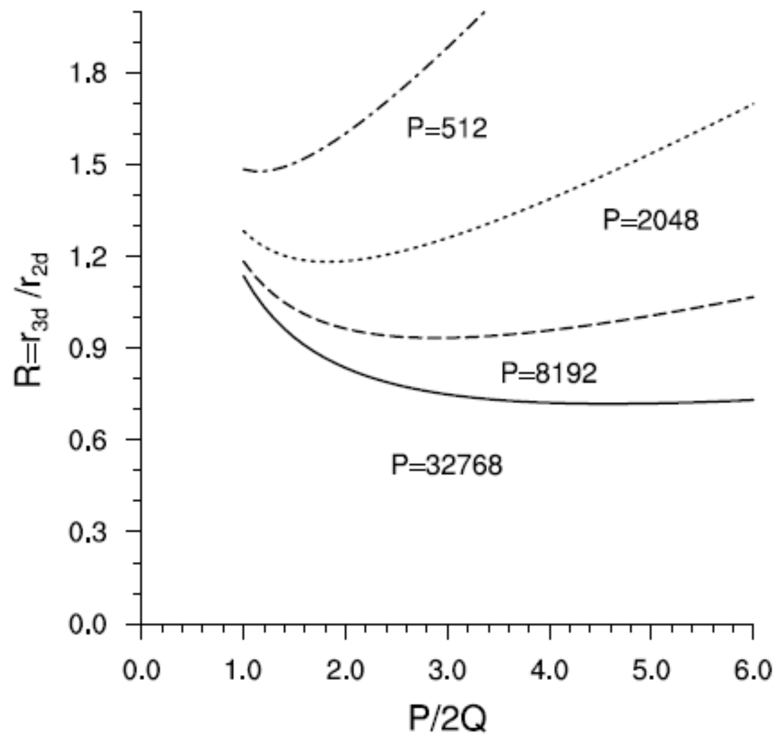
Vertical/horizontal extent



Optimal no. of cores in the vertical

$$P/bQ = \sqrt[3]{P/ba^2}.$$

Performance model for 1024 × 512 × 41 idealized climate simulation



Not always a performance gain from the 3D decomposition !

... but we can always TRY to use more cores to decrease time-to-solution !

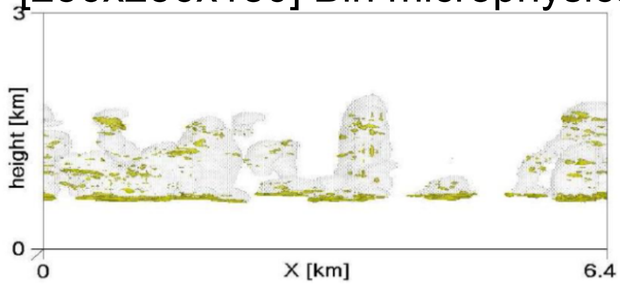
Table 1: Strong scaling of idealized climate simulation on a $256 \times 128 \times 64$ grid, using 512 processors in the horizontal with increasing number of processors in the vertical.

Total processor number	512	1024	2048	4096
Processor configuration	$32 \times 16 \times 1$	$32 \times 16 \times 2$	$32 \times 16 \times 4$	$32 \times 16 \times 8$
Wallclock time [sec]	52	30	20	15

Everything depends on the problem size and architecture, if communication takes 80 % of the wall time (e.g. if pressure solver works really hard and there is no physics), there is no way to achieve significant speedup

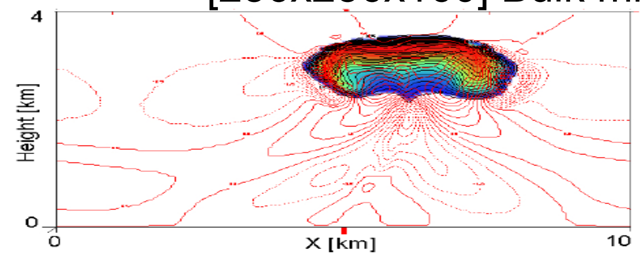
2D/3D *decomposition scalability*
(full model physics, Thomas preconditioner)

[256x256x180] Bin microphysics

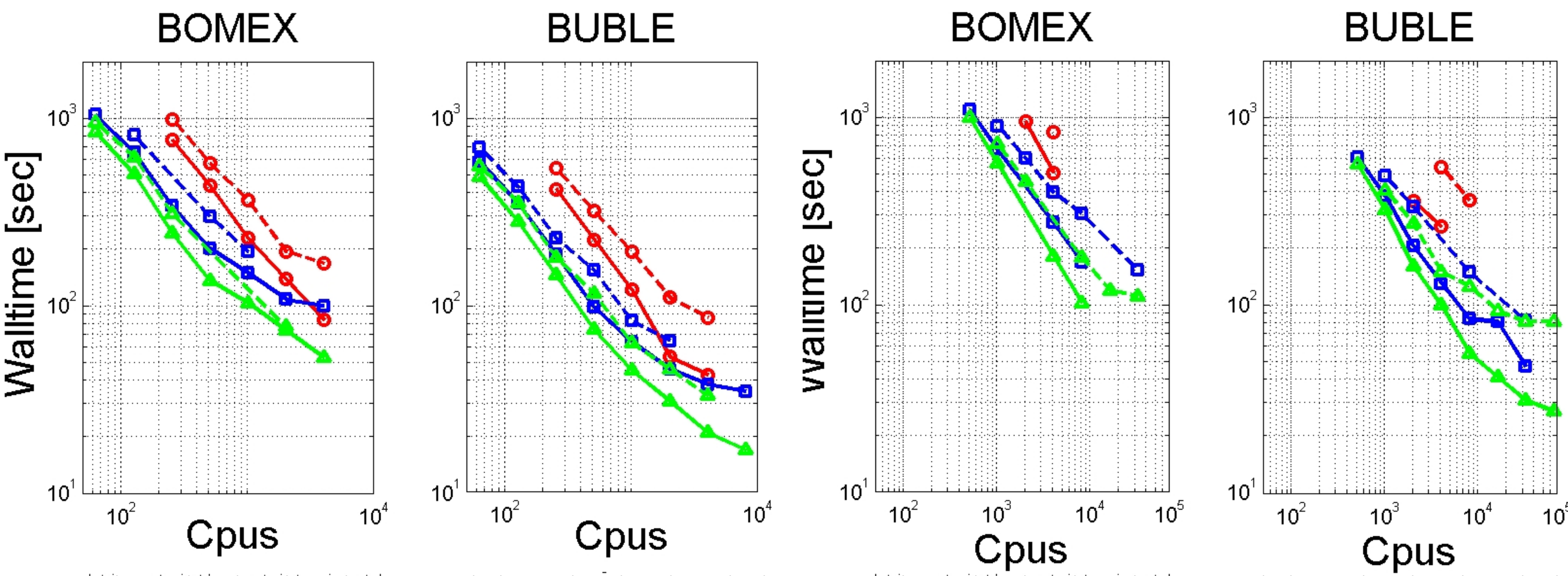


BOMEX

[256x256x160] Bulk microphysics



RISING
BUBBLE



Strong scalability results with full model physics. The red, blue, and green lines shows results from IBM BG//L, CRAY XT4 and Cray XE6 respectively, the dashed lines represent 2D decomposition, the continuous lines 3D decomposition. Left and right panels show default and double resolution problems, respectively.

Remarks on vertical algorithms

- Strong domain anisotropy (thin shell) results in very bad conditioning hurting the performance of iterative solvers
- Effective preconditioning is a key to the iterative solver convergence …
- … but it demands direct inversion of the tridiagonal matrix in the vertical direction (same for radiation)

- Thomas algorithm is a embarrassingly serial recurrence → special treatment necessary
- Possible solution is the recurrence doubling approach

$a(n+1) = Ba(n)+C$ is rewritten as:

$a(n+1)=F(B,C)a(1)$ + parallel part

+ pipelining or single GATHER/SCATTER in the vertical (depending on the machine and number of cores)

Additional benefits of 3D MPI parallelization

- Most part of the EULAG code is now symmetrical in x,y,z

- A number of long lasting bugs revealed and fixed

- For large class of experiments, time-to-solution significantly decreased for fixed number of cores

- Size of the innermost loop is more flexible – beneficial for vectorization

- Many optimizations introduced in process of coding and testing of the new code

More remarks …

With the new, 3D parallelization we can attempt to simulate much larger problems, BUT there is a memory wall ahead.

→ Need for improving memory locality and cache use efficiency

Also, we can decompose problem to use many more cores, BUT there is a communication wall ahead

→ Need for minimizing halo updates and, especially, reduce number of global MPI operations to minimum

Conclusions

- Three dimensional MPI parallel formulation, for symmetric (e.g. cubical turbulence) problems, can decrease time to solution for given number of cores used by factor of ~ 0.5.

- For thin-shell applications (weather and climate), it allows for decreasing time-to-solution by admitting much larger number of computing cores.