Strategies for increasing the arithmetic intensity on ensemble averaged Parallel-in-time simulations

J. Plana-Riu[1], F.X. Trias[1], À. Alsalti-Baldellou[1,2], A. Oliva[1]

[1]Heat and Mass Transfer Technological Centre
Technical University of Catalonia
[2]Termo Fluids S.L.

34th International Conference on Parallel Computational Fluid Dynamics

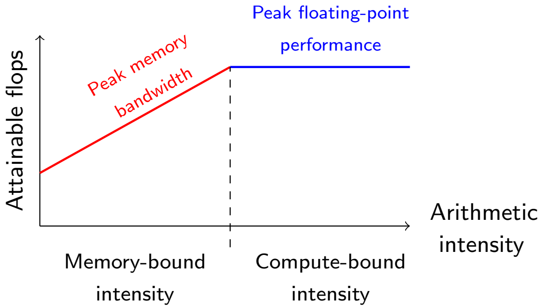
May 29th-31st, 2023
Cuenca, Ecuador

Centre Tecnològic de Transferència de Calor
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Contents

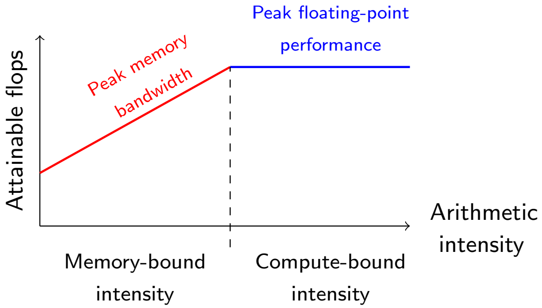
Motivation

- CFD is a memory-bound computational process...



Motivation

- CFD is a memory-bound computational process...



- **Research question 1:** How can this be improved?

Motivation

- Turbulence is (pretty much) all about statistics...

Motivation

- Turbulence is (pretty much) all about statistics...

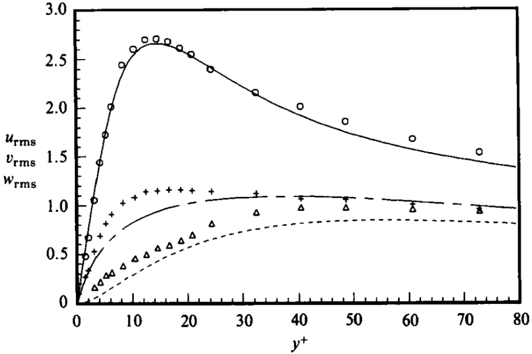


Figure: Root-mean-square velocity fluctuations in wall cooordinates for a $\mathrm{Re}_\tau = 180$ channel flow.
J. Kim, P. Moin and R.Moser (1987). Turbulence statistics in fully developed channel flow at low Reynolds number. Journal of Fluid Mechanics(177) pp. 133-166

Motivation

- Turbulence is (pretty much) all about statistics...

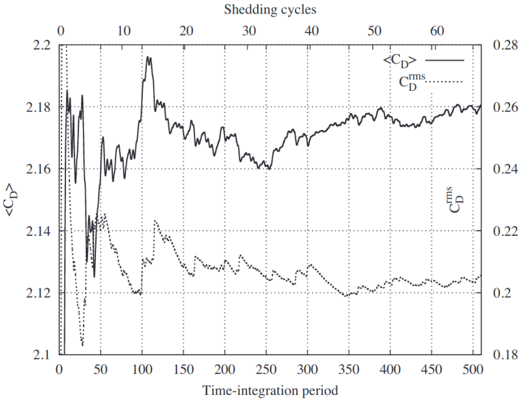


Figure: Temporal convergence of drag coefficient in a Re=22,000 square cylinder DNS.
F.X. Trias, A.Gorobets and A.Oliva (2015). Turbulent flow around a square cylinder at Reynolds number 22,000: A DNS study. Computers and Fluids (123) pp. 87-98

Motivation

- Turbulence is (pretty much) all about statistics...

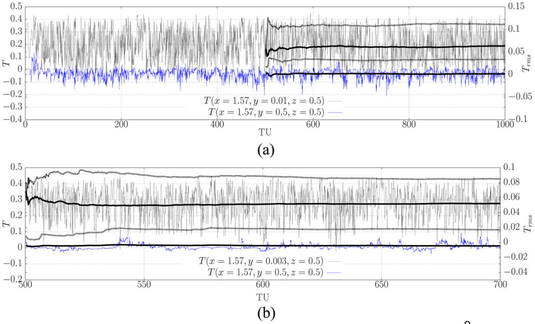


Figure: Time series and cumulative statistics for temperature at two points for Ra $= 10^8$ and 10^{10} Rayleigh-Bénard DNS. *F.Dabbagh, F.X. Trias, A.Gorobets and A.Oliva (2017). A priori study of subgrid-scale features in turbulent Rayleigh-Bénard convection. Physics of Fluids (29) 105103*

Motivation

- Turbulence is (pretty much) all about statistics...

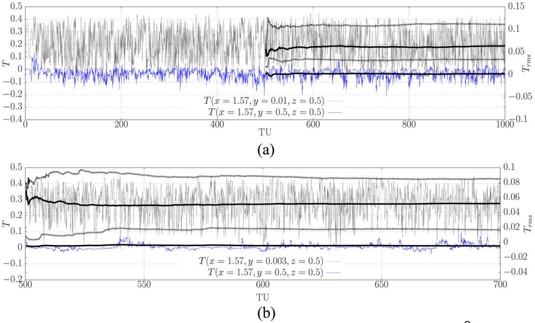


Figure: Time series and cumulative statistics for temperature at two points for Ra $= 10^8$ and 10^{10} Rayleigh-Bénard DNS. *F.Dabbagh, F.X. Trias, A.Gorobets and A.Oliva (2017). A priori study of subgrid-scale features in turbulent Rayleigh-Bénard convection. Physics of Fluids (29) 105103*

But statistics require extremely long simulations...

Motivation

- Turbulence is (pretty much) all about statistics...

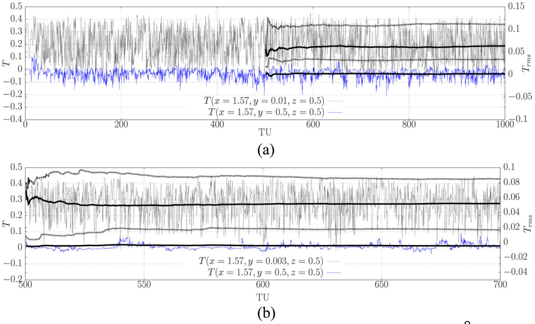


Figure: Time series and cumulative statistics for temperature at two points for Ra $= 10^8$ and 10^{10} Rayleigh-Bénard DNS. *F.Dabbagh, F.X. Trias, A.Gorobets and A.Oliva (2017). A priori study of subgrid-scale features in turbulent Rayleigh-Bénard convection. Physics of Fluids (29) 105103*

But statistics require extremely long simulations...

- **Research question 2:** Is there a way to shorten the averaging?

Ensemble averaging

Multiple RHS can be exploited in multiple ways (i.e. exploiting domain symmetries[1]), yet in this case it will be by **ensemble averaging**.

Ensemble averaging

- Running *m parallel-in-time* simulations, and then average the results.

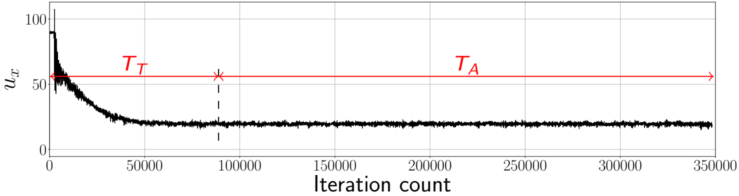


Figure: Regular averaging for calculation of statistics.

[1] *À.Alsalti-Baldellou, X.Álvarez-Farré, F.X. Trias, A.Oliva (2023). Exploiting spatial symmetries for solving Poisson's equation. Journal of Computational Physics (486) 112133*

Ensemble averaging

Multiple RHS can be exploited in multiple ways (i.e. exploiting domain symmetries), yet in this case it will be by **ensemble averaging**.

Ensemble averaging

- Running *m parallel-in-time* simulations, and then average the results.

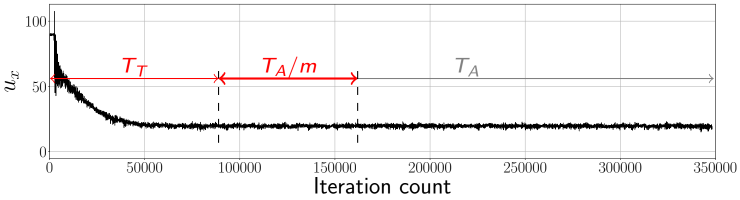


Figure: Ensemble averaging.

Ensemble averaging

Where does T_T end?

- Steady-state?
- Statistical decoupling between all cases run
  - Rolling Pearson correlation coefficient

$$\rho_{xy}(\tilde{t}) = \frac{\text{Cov}(x,y)(\tilde{t})}{\sigma_x(\tilde{t})\sigma_y(\tilde{t})}$$

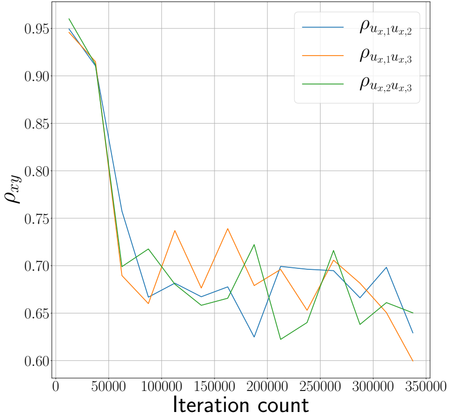


Figure: Rolling Pearson correlation coefficient for a $\text{Re}_\tau = 180$ channel flow for $m = 3$ with a window of 1500 iterations.

Memory-bounded computation

Memory-boundedness is due to SpMV operations… yet its arithmetic intensity (AI) can be improved if multiple RHS are used (leading to sparse matrix-matrix products (SpMM)).

Arithmetic intensity of an SpMM

$$\mathsf{AI}_{\mathsf{SpMM}}(d) = \frac{(2nnz(A)+1)d}{12nnz(A)+4(m+1)+8(m+n+1)d}$$

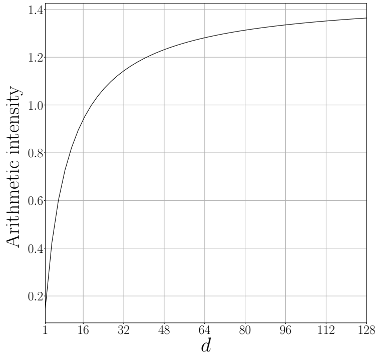


Figure: Arithmetic intensity of SpMM up to 128 RHS vectors, with 17 non-zero entries per row.

Making use of SpMM in ensemble averaging

Previous work...

- First developed by Krasnopolski[2]... the method was called generalized sparse matrix-vector product (GSpMV).
- Only applied in the solution of the Poisson equation for ensemble averaging.

- But... there are plenty of SpMV's that can be exploited:
  - Diffusive operator
  - Convective operator
  - ...
  - Up to 18 SpMV per iteration (in an AB2 setup) + Poisson

[2] *B.I. Krasnopolski (2018). An approach for accelerating incompressible turbulent flow simulations based on simultaneous modeling of multiple ensembles. Computer Physics Communications (229) pp.8-19*

Making use of SpMM in ensemble averaging

In-house unstructured collocated code +

HPC^2 framework[3][4]

Fully-portable, algebra-based framework

- BLAS-like kernels
- SpMM computation capabilities
- Fully-portable (CPU, GPU)
- Poisson solution

[3] *X. Álvarez, A. Gorobets, F.X. Trias, R. Borrell, G. Oyarzun (2018). HPC^2-A fully-portable, algebra-based framework for heterogeneous computing. Application to CFD. Computers and Fluids (173) pp.285-292*

[4] *X. Álvarez, A. Gorobets, F.X. Trias (2021). A hierarcical parallel implementation for heterogeneous computing. Application to algebra-based CFD simulations on hybrid supercomputers. Computers and Fluids (214) 104768*

Theoretical speed-ups

- Increment on the AI will only be achieved in the SpMV (now replaced by SpMM)... thus let θ be the fraction of the iteration in which an SpMV is computed.
- Times ratio, $\beta = T_A/T_T$.

According to Krasnopolski [1]...

$$P_m = \frac{1+\beta}{m+\beta}\frac{5m}{5m-3\theta(m-1)}$$

- Optimal value found for

$$m_{\text{Opt}} = \sqrt{\frac{3\beta\theta}{5-3\theta}}$$

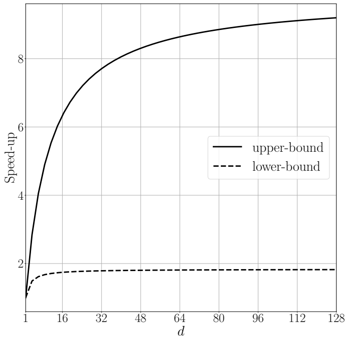


Figure: Theoretical speed-up bounds for a sparse matrix A with $nnz(A)/n = 17$

Results

A speed-up analysis...

Conditions

- $Re_\tau = 180$ channel flow
- 128^3 and 256^3 meshes
  - Uniform in x and z.
  - Hyperbolic tangent stretching in y.
- 7 non-zero entries per row
- AB2 + CFL (0.35) integration
- Runs for $1, 2, 4, 8$ rhs
- Algebraic approach for in-house code+HPC[2]
  - 100 non-preconditioned CG iterations
  - 2 MPI tasks, 20 OpenMP threads
  - 1 JFF fourth-generation compute node:
    - 2x Intel Xeon 6230

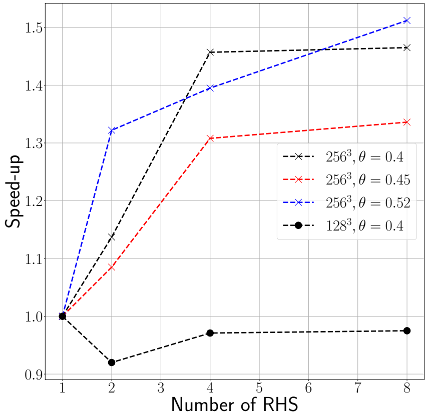


Figure: Time iteration speed-ups, τ, for 2, 4 and 8 RHS in 128^3 and 256^3 meshes, averaged for 40 time-steps.

Results

A speed-up analysis...

Simulation speed-up extrapolation

$$P_m = \tau \frac{1+\beta}{m+\beta}$$

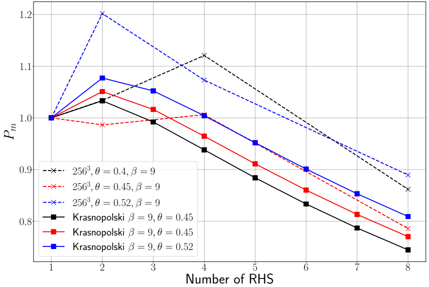


Figure: Simulation speed-up for 2, 4 and 8 RHS in a 256^3 mesh, compared against the theoretical expression from Krasnopolski.

Results

A speed-up analysis...

Simulation speed-up extrapolation

$$P_m = \tau \frac{1+\beta}{m+\beta}$$

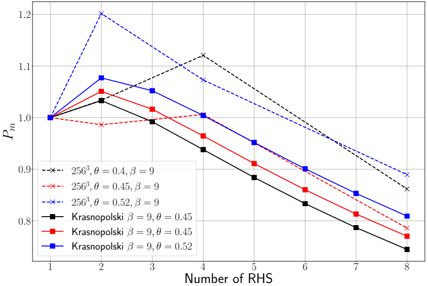


Figure: Simulation speed-up for 2, 4 and 8 RHS in a 256^3 mesh, compared against the theoretical expression from Krasnopolski.

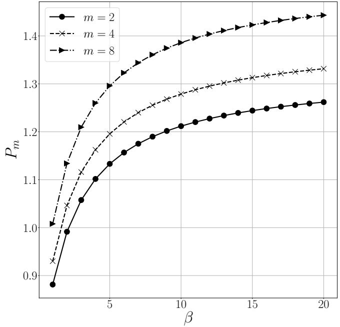


Figure: Simulation speed-up for different β values for the 256^3 mesh, $\theta = 0.4$.

- $m = 8$ has speed-up $\forall \beta$
- $m = 4$ speeds-up for $\beta \geq 2$, $m = 2$ for $\beta \geq 3$

Concluding remarks

Take-away messages

- Ensemble averaging as a technique for computing statistics.

Concluding remarks

Take-away messages

- Ensemble averaging as a technique for computing statistics.
- Original work from Krasnopolski can be extended to all SpMVs in the simulation.

Concluding remarks

Take-away messages

- Ensemble averaging as a technique for computing statistics.
- Original work from Krasnopolski can be extended to all SpMVs in the simulation.
- Method works properly under certain conditions: 1-rhs-case should not fit in cache memory (128^3-mesh case not working properly)

Concluding remarks

Take-away messages

- Ensemble averaging as a technique for computing statistics.
- Original work from Krasnopolski can be extended to all SpMVs in the simulation.
- Method works properly under certain conditions: 1-rhs-case should not fit in cache memory (128^3-mesh case not working properly)
- Leads to notable improvements compared to Krasnopolski's theoretical speed-ups (+11% for $\beta = 9, \theta = 0.52$).

Concluding remarks

Future work

- Full simulation speed-up calculation, with ensemble averaging statistics.
- Testing with multiple platform HPC systems.