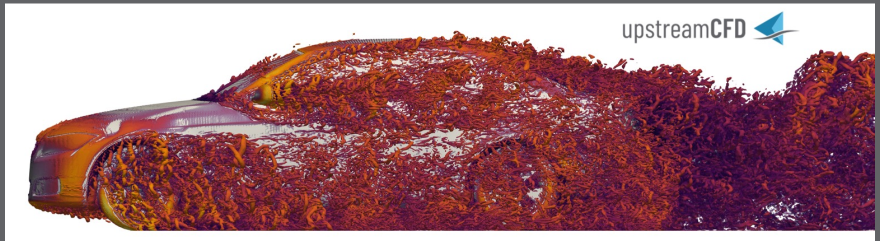
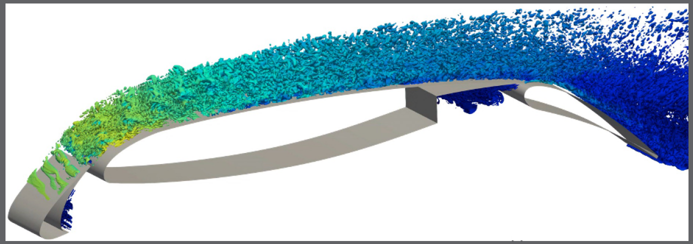
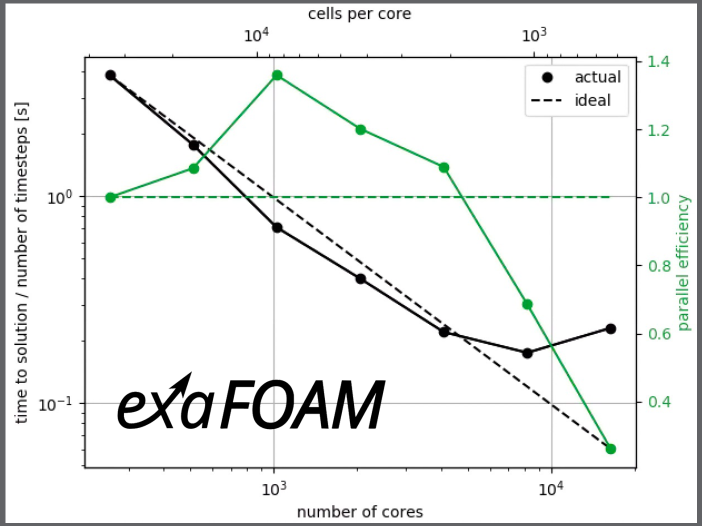
HPC Scalability as an Enabling Technology for High-Fidelity CFD

C. Mockett, J. Bergmann, L. Fliessbach, M. Fuchs, H. Hetmann,
F. Kramer, N. Schönwald

Upstream CFD GmbH, Berlin, Germany







2nd exaFOAM Workshop

Introduction

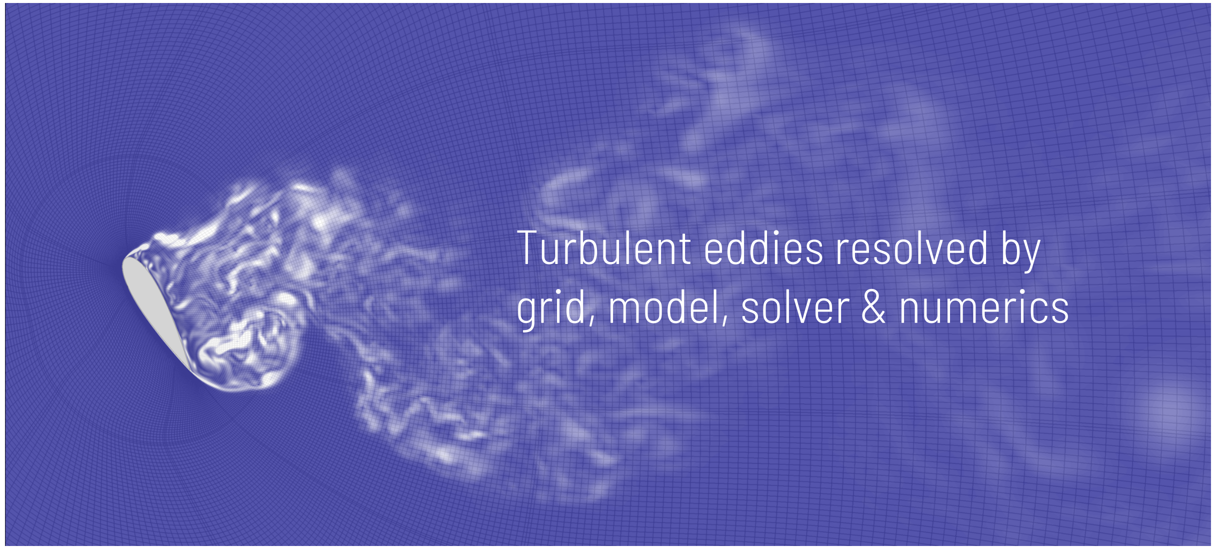
Objectives of talk

To convey:

- The perspective of an innovative SME specialising in high-fidelity CFD for industrial applications
  - Showing some of our work inside and outside of exaFOAM
- The value of high-fidelity CFD
  - Example of Machine Learning training dataset
- The inherent HPC challenges to the feasibility of high-fidelity CFD
  - And efforts within the exaFOAM project to overcome these
- The opportunities of combining super-linear scalability with cloud-HPC for CPU-based CFD
  - And the challenges to achieving this

…in an approachable manner

High-fidelity CFD



- Advantages of high-fidelity methods over RANS:
  - More accurate prediction of flows with large-scale separation
  - Resolving turbulent unsteadiness necessary for some applications, e.g. broadband aeroacoustics

"High-fidelity" a.k.a. turbulence-resolving methods:

- By definition 3D and time-resolved
- Different levels of turbulence resolution

RANS

Reynolds-averaged Navier-Stokes

All turbulence modelled

High empiricism

Ready: 1990

Hybrid RANS-LES

e.g. DES, SAS, WMLES, SBES...

LES

Large-eddy simulation

≈ 80% of turbulence resolved

Low empiricism

Ready: 2070

DNS

Direct Numerical Simulation

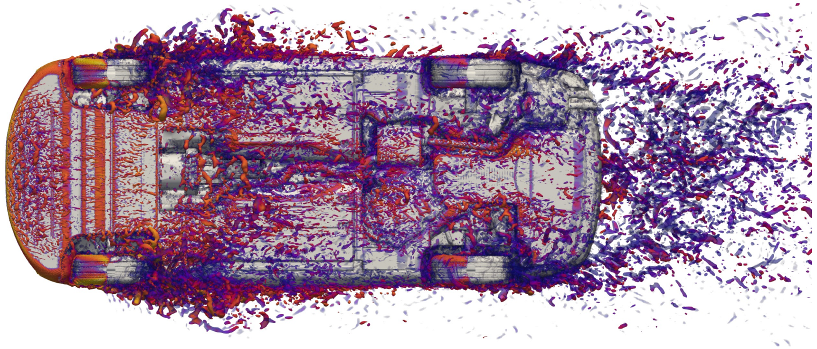
All turbulence resolved

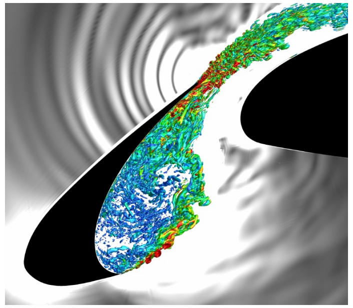
No empiricism

Ready: 2080

Increasing computational cost and method accuracy

(Readiness dates from the 1997 paper of Spalart et al. introducing DES)





Example high-fidelity CFD simulation: DES of DrivAer Notchback

- Case 2 from 2nd Automotive CFD Prediction Workshop
- OpenFOAM with advanced DES model and robust, low-dissipation solver/schemes
- Detached-Eddy Simulation (DES):
  - RANS in attached boundary layers
  - LES in separated flow regions
- DES widely used in industry for automotive external aerodynamics
- Computational cost varies widely depending on statistical convergence and settings used
  - Meshes of order 100M – 300M cells
  - Time step of order 10^{-4} s
  - Physical time of order 3s - 15s needed depending on statistical convergence

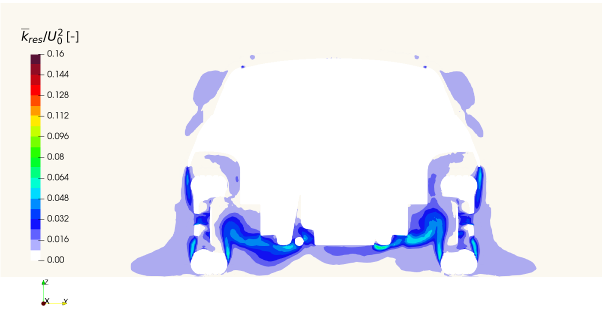
upstreamCFD

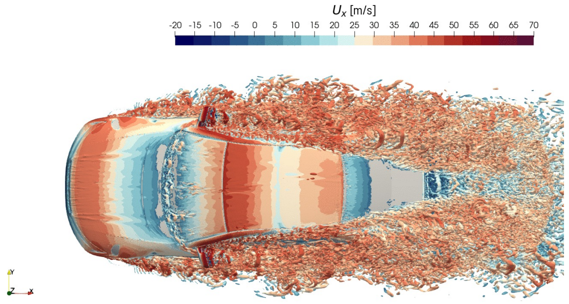
Example of high-fidelity CFD use case: A training dataset for Machine Learning

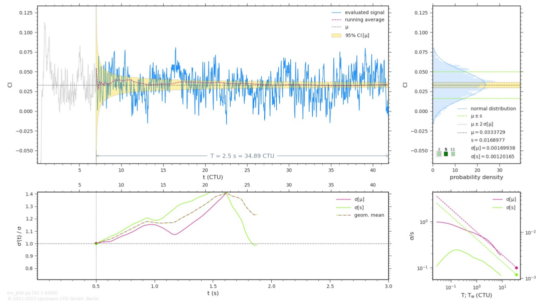
A high-fidelity, open-source dataset for ML in automotive aerodynamics

- Machine Learning has the potential to revolutionise automotive aerodynamics:
  - e.g. surrogate models, offering feedback on design iterations in seconds instead of overnight
- Challenges:
  - ML needs huge amounts of training data
  - Reliability of data determines reliability of model
- Collaborative initiative to generate a large, high-fidelity, open-source dataset:
  - Baseline DrivAer Notchback geometry and wind tunnel data provided by Ford (for workflow validation)
  - Geometry parametrisation by Siemens Energy and BETA-CAE
  - Morphing and automated meshing with ANSA by BETA-CAE
    - Consistent meshes of approx. 160M cells each
  - High-fidelity simulation workflow by Upstream CFD
  - Simulation execution, computing power and data storage/hosting by AWS

- RANS and DES solutions for 500 geometry variants:
  - Complete flow fields, surface data & slices
  - Aerodynamic quantities and flow visualisation images

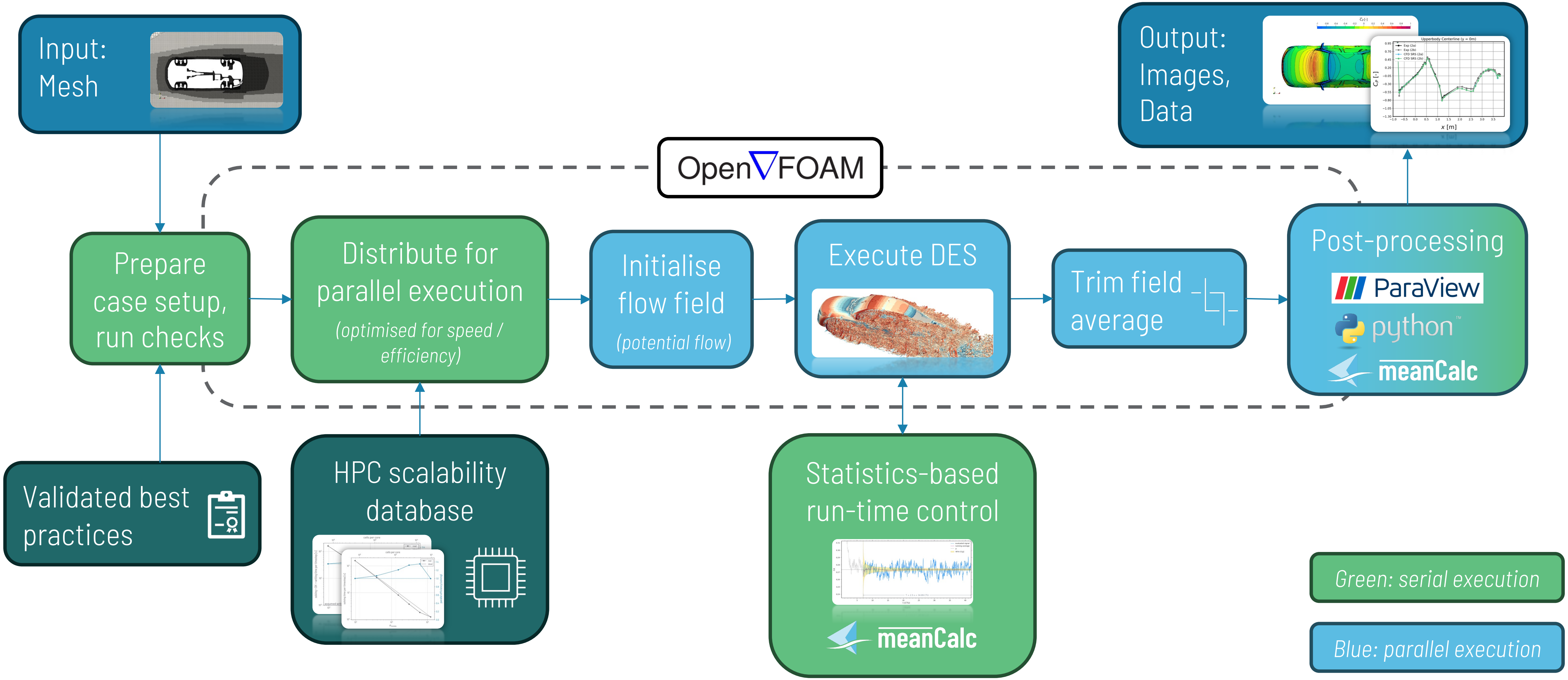






For more details:
C. Mockett, N. Ashton, L. Fliessbach, M. Fuchs, H. Hetmann, T. Knacke, N. Schönwald, V. Skaperdas & A. Walle: "High-fidelity, open-source training dataset for machine learning in automotive aerodynamics", ERCOFTAC Workshop on Machine Learning for Fluid Dynamics, Paris, 2024 (slides on LinkedIn)

Automated OpenFOAM-based workflow for automotive aerodynamics

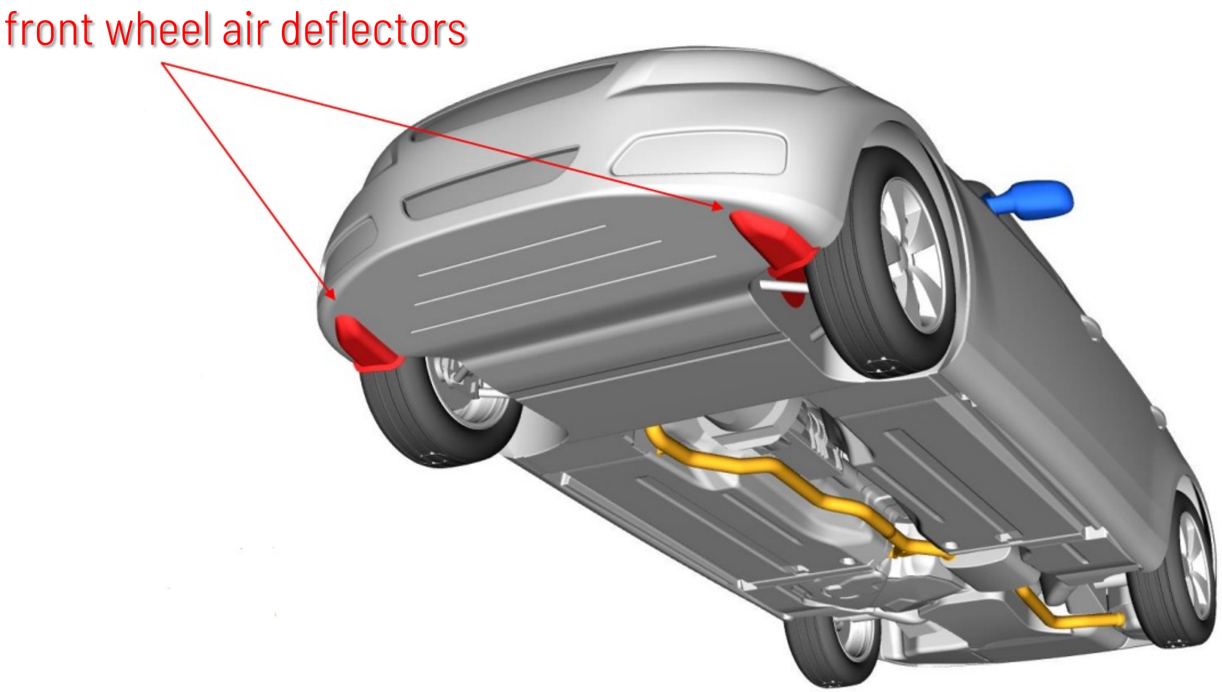


Validation cases: The Ford DrivAer Notchback

- Test cases 2a and 2b from the 4th [Automotive CFD Prediction Workshop]
- Full-scale, static wheels and ground, closed cooling
- Geometry and wind tunnel data provided by Ford
  - Hupertz et al. (2022): Towards a Standardized Assessment of Automotive Aerodynamic CFD Prediction Capability - AutoCFD 2: Ford DrivAer Test Case Summary. SAE Technical Paper 2022-01-0886, 2022, doi:10.4271/2022-01-0886.

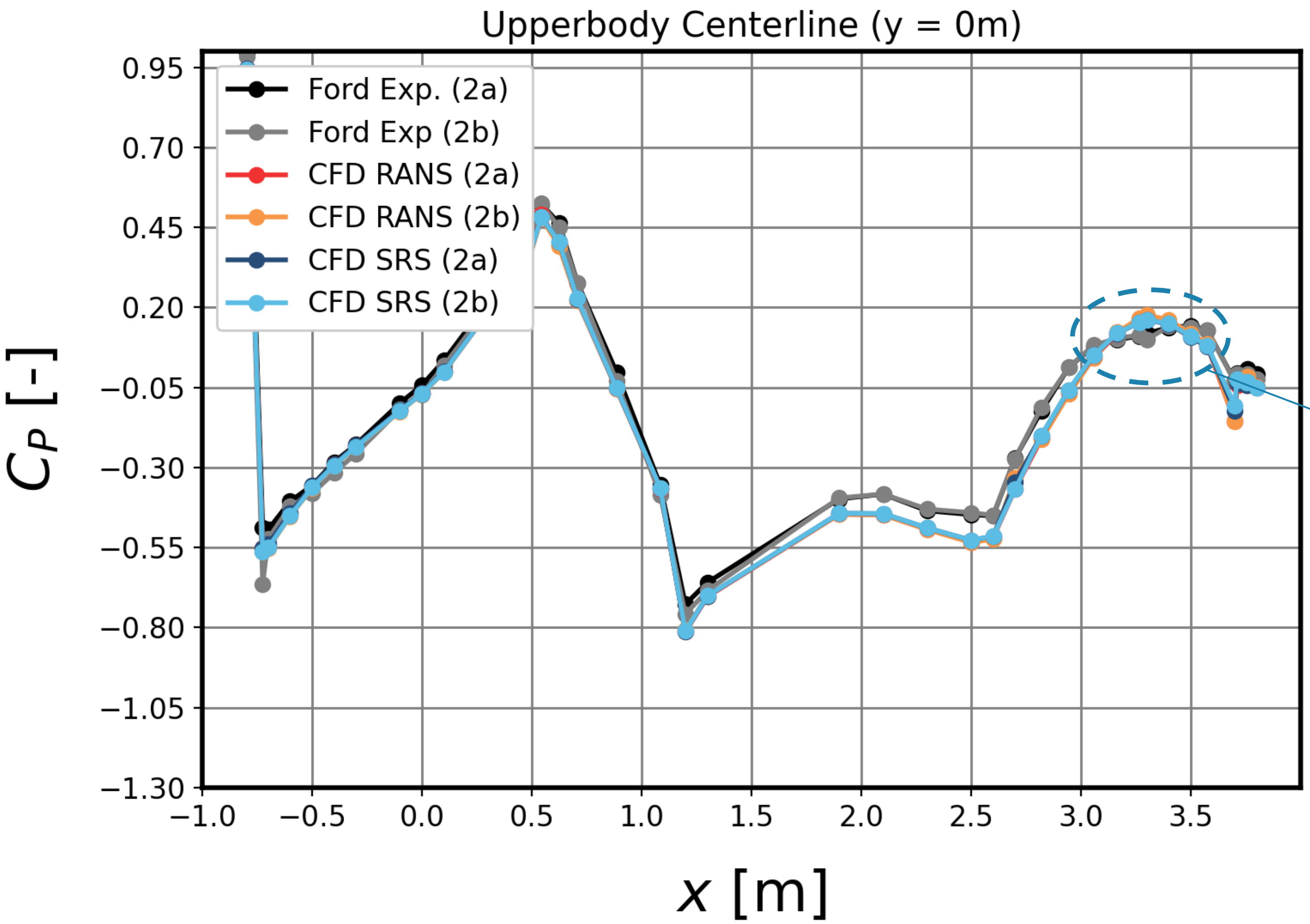
Ford DrivAer model (1:1 scale) investigated experimentally in Pininfarina wind tunnel

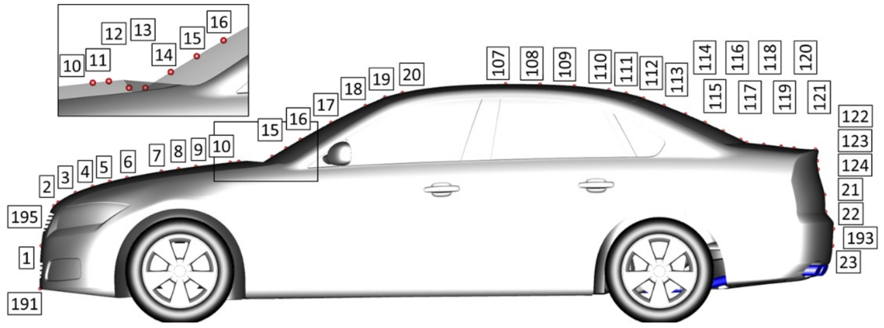
- Two geometries:
  - Case 2a: Baseline
  - Case 2b: With added frong-wheel air deflectors (FWAD)



Mean pressure coefficient, upper-body centreline

upstream CFD

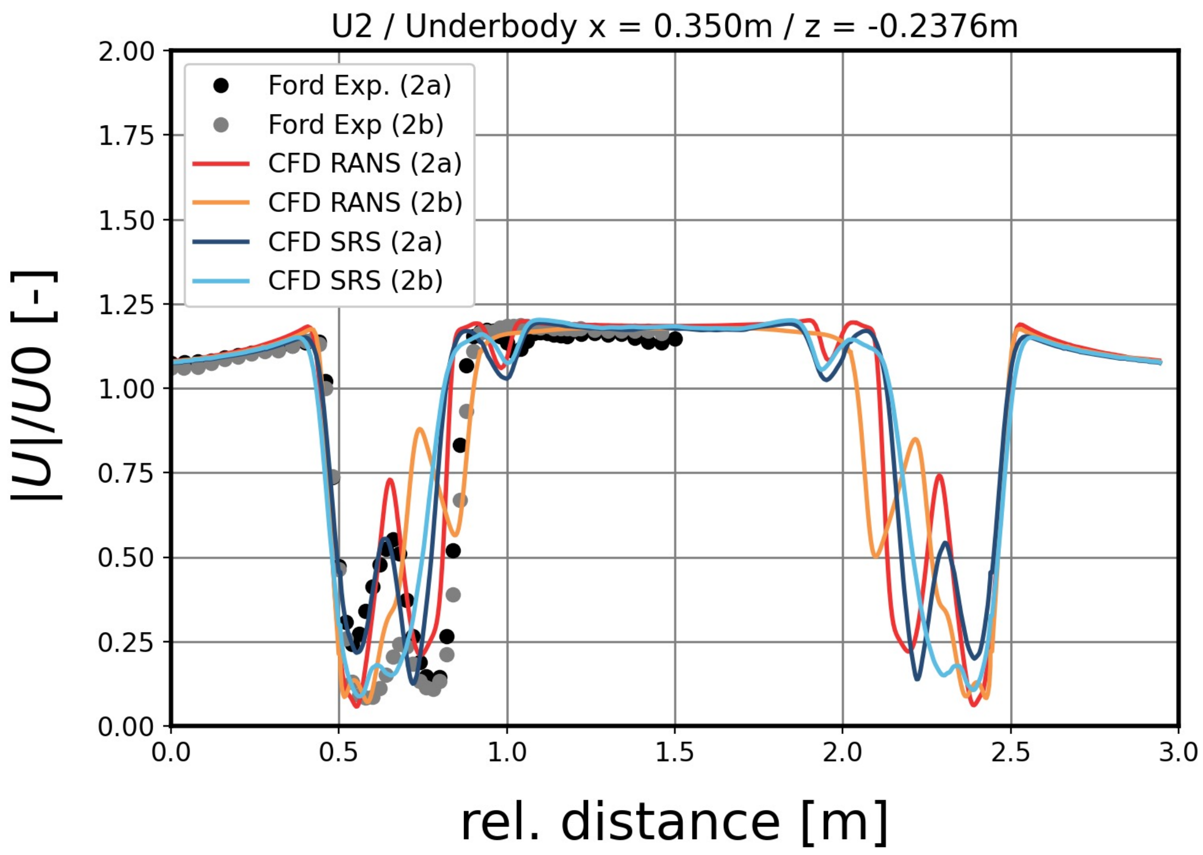


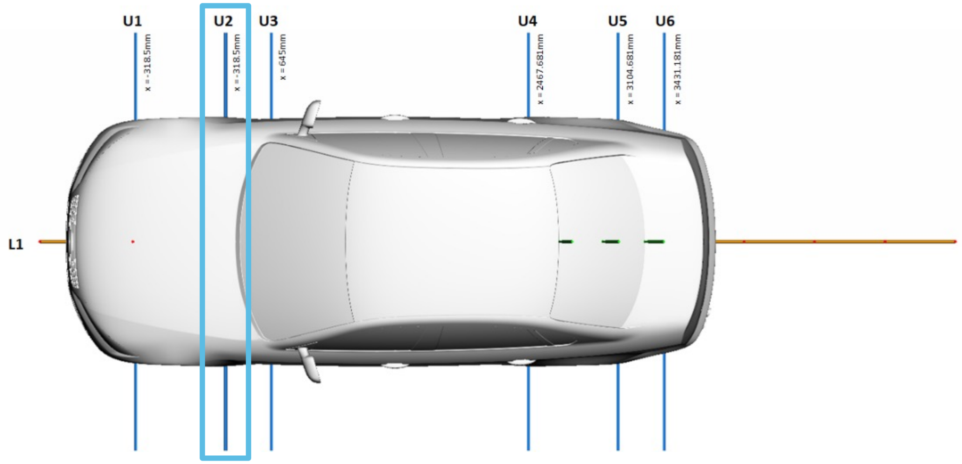


Separation plateau less pronounced in simulations

RANS and DES very similar here

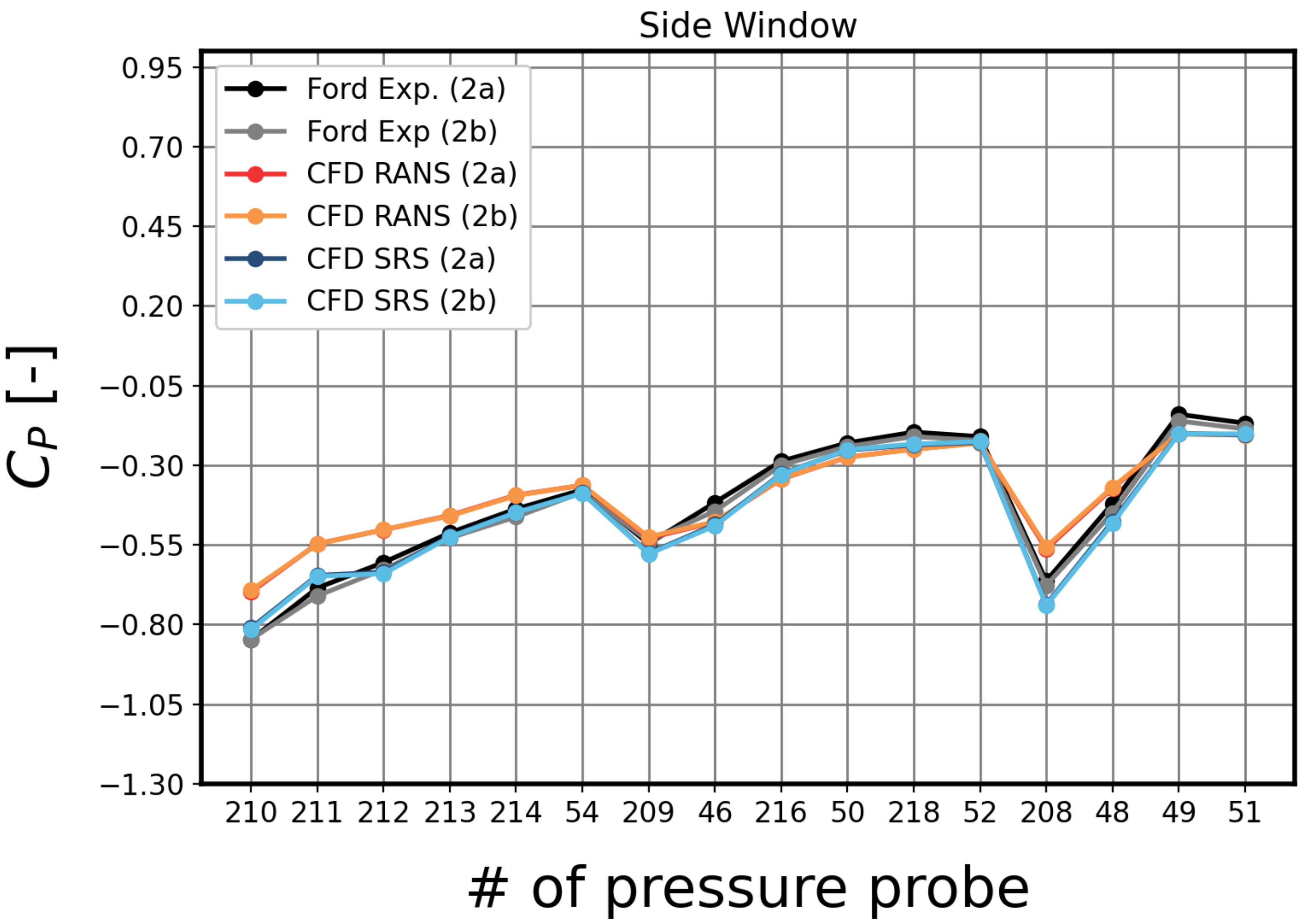
Mean velocity magnitude in front wheel wake

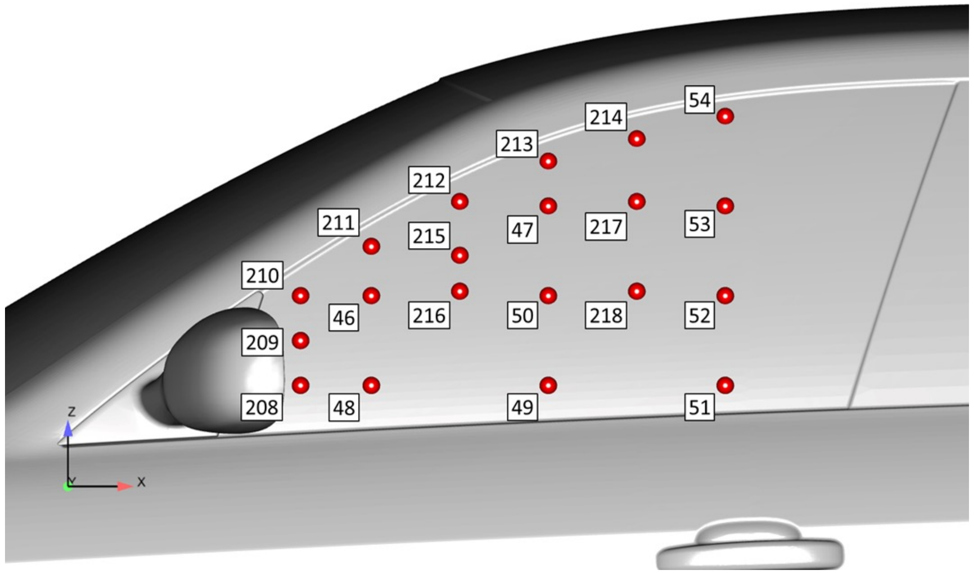




Tyre wake and effect of front wheel air deflectors reproduced much more accurately by DES than by RANS

Mean pressure on side window





Also here, DES much closer to wind tunnel measurements than RANS

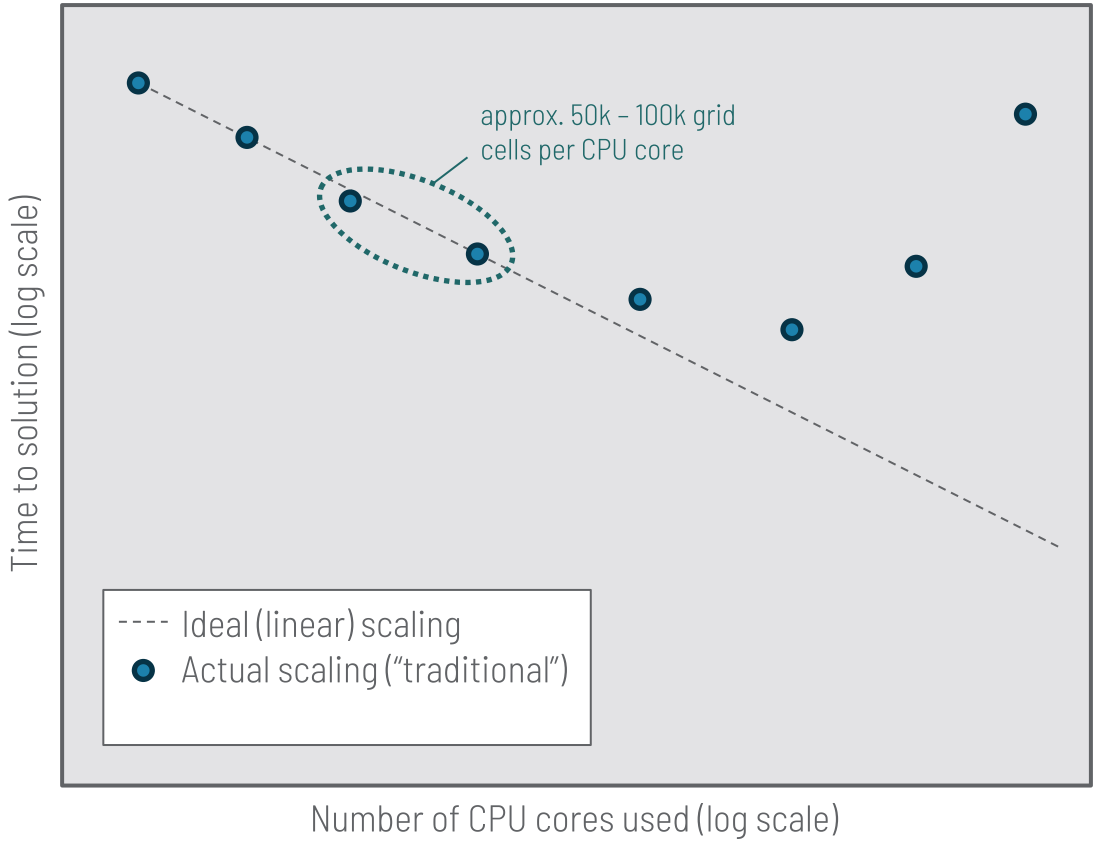
upstreamCFD

The potential for superlinear scalability

Strong scalability testing in CFD: What we thought we knew

- Strong scalability testing: Fixed problem size, change number of parallel ranks (CPU cores)
- We can shorten turnaround time by using more CPU cores
- ...until we reach the scalability limit:
  - Communication overhead starts to dominate
  - Parallel efficiency starts to drop off when we use less than around 50k – 100k grid cells per core
- So if we're in a hurry, we can speed up simulations, but this will cost more
- ...and anyway, we only have a few hundred cores on our in-house cluster

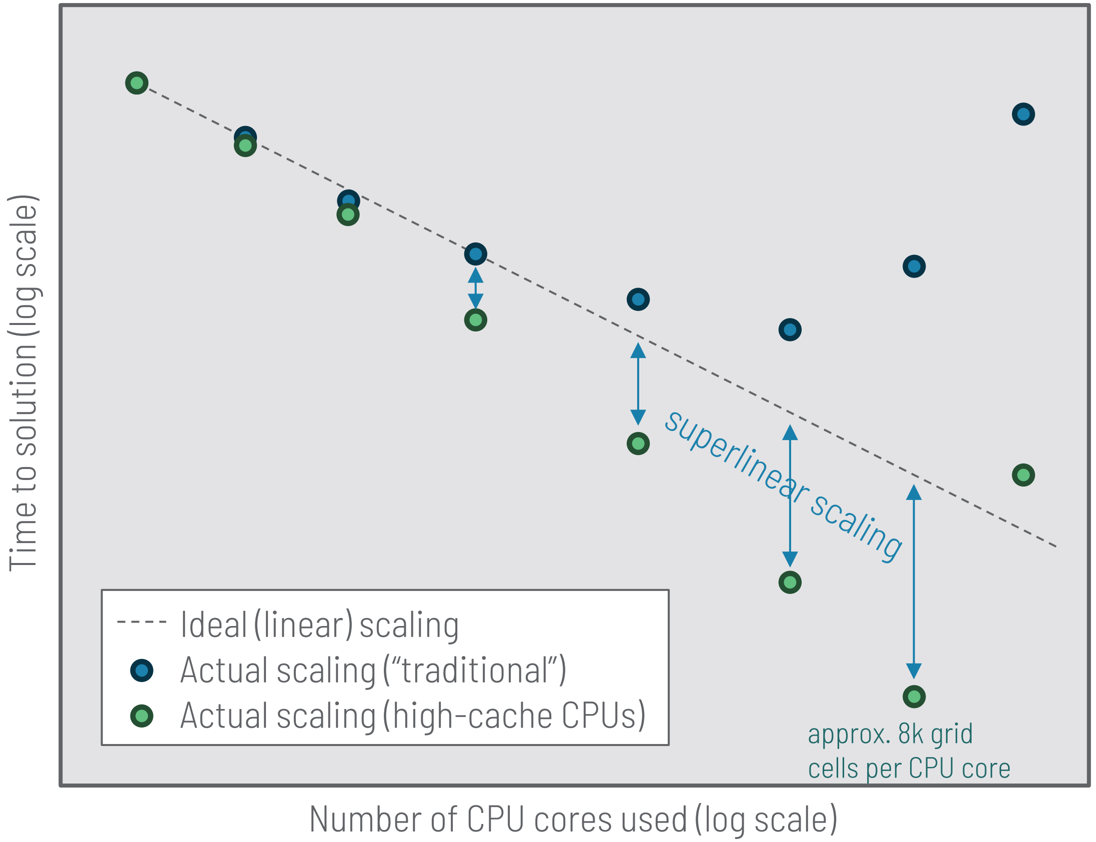
Strong scaling test:

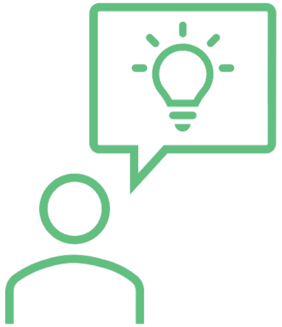


Important finding: Superlinear scaling

- Scalability testing on newer CPU generations reveals superlinear scaling phenomenon
- Explained by larger cache memory:
  - CFD performance limited by memory bandwidth
  - When a greater percentage of the problem fits in CPU cache, greater processing performance is unlocked
  - The effect is so strong that it more than compensates for the communication overhead, with optimum efficiency at approx. 8k grid cells per core

Strong scaling test:

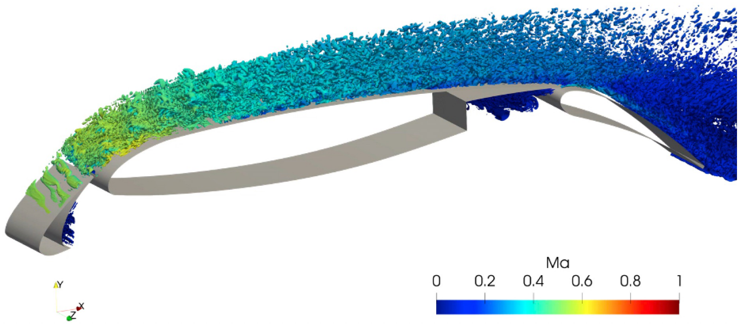


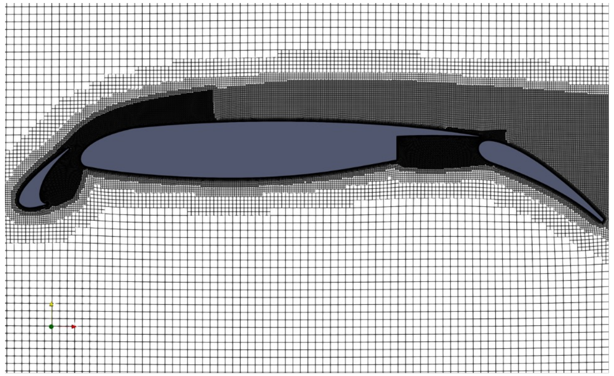


See e.g. presentation of Flavio Galeazzo (HLRS) at 18th OpenFOAM Workshop:
F.C.C. Galeazzo, R.G. Weiß, A. Ruopp, "Understanding Superlinear Speedup in Current HPC Architectures"
Link to YouTube video

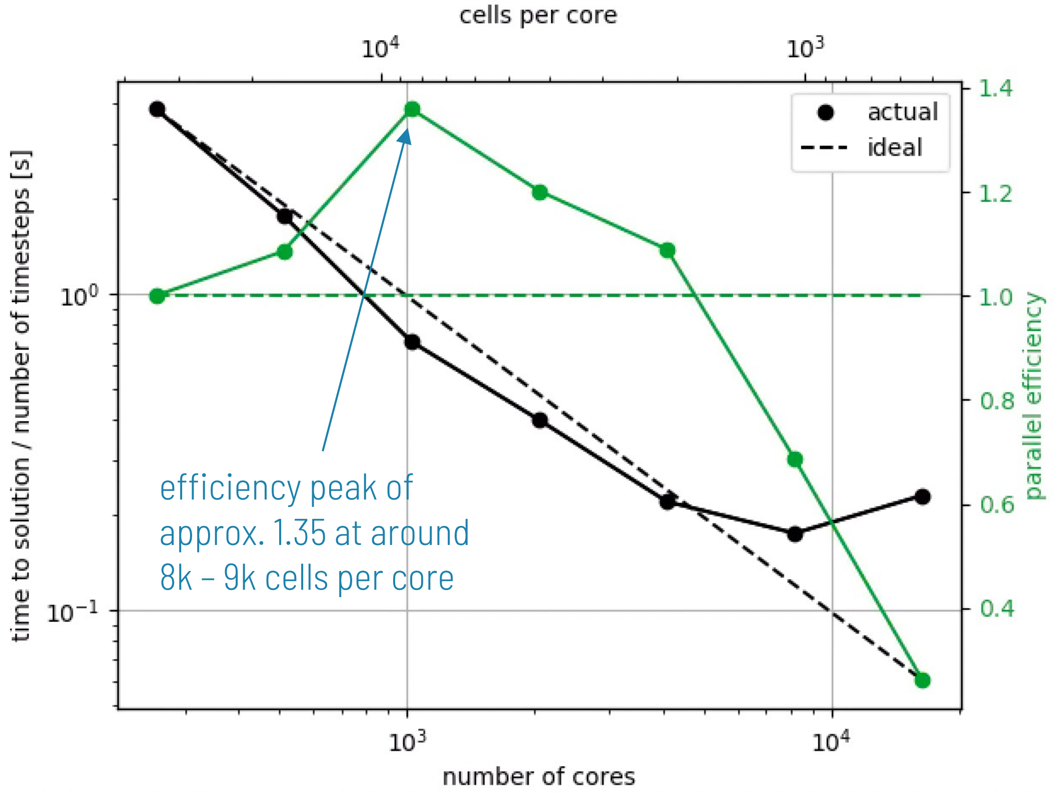
Superlinear scaling for high-fidelity CFD?

- Initial super-linear results seen for lid-driven cavity
- Turbulence-resolving simulations (LES, DES) have a very different "HPC profile" to the lid-driven cavity:
  - Unsteady simulations
  - Fine time steps (small changes in field from one time step to the next)
  - Very few linear solver iterations per time step
- exaFOAM Microbenchmark case MB9:
  - 30P30N high-lift airfoil
  - Wall-modelled LES
  - Unstructured grid of 19.5M cells
  - [Test case download link]





Strong scalability test for 30P30N high-lift airfoil on the VEGA supercomputer



The implications

- If this result holds for more practical cases, the implications are enormous
- Now the *most efficient* setup is also *much faster*
- Back-of-the-envelope calculation for a typical automotive DES case
  - 160M grid cells, baseline cost of ca. 60k core-hrs
  - Decomposing to 100k cells per core:
    - 1600 cores
    - 38 hr turnaround time
  - If we can go down to 8k cells per core:
    - 20 000 cores
    - 3 hr turnaround time
  - Factoring in the efficiency boost (+35%):
    - Cost reduced from 60k core-hrs to 45k core-hrs
    - 2.2 hr turnaround time
- Drastic shortening of turnaround time **and** lower cost!

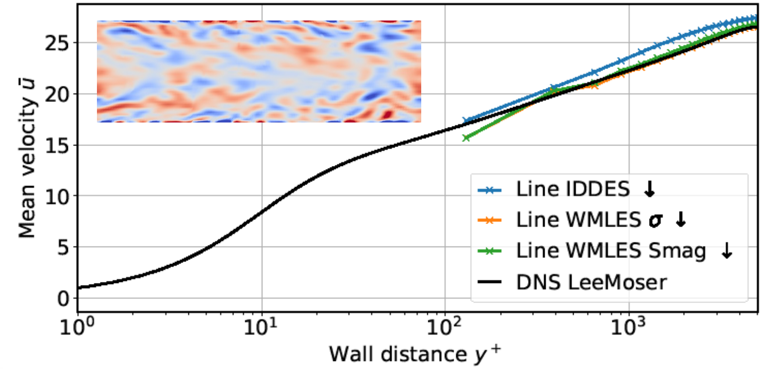
- So what's the catch?
  - Where can I find 20 000 cores?
    - On the cloud (but what if everybody does this at once...?)
  - Does the effect really apply for large cases and enormous core counts?
  - Decomposing the grid becomes the new bottleneck (details to follow)
- What does the future hold?
  - Chip manufacturers are adding increasing amounts of cache
  - As cache per core increases, the superlinear effect will start to kick in for lower numbers of cores (greater grid cells per core)

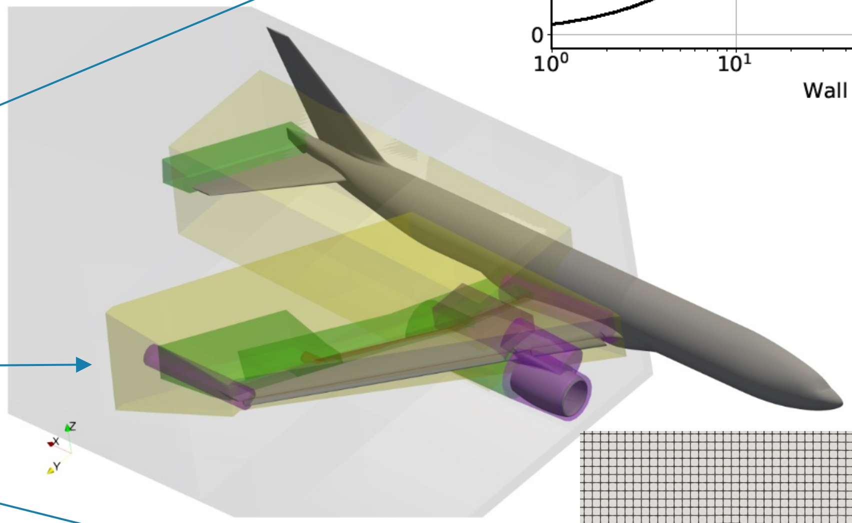
upstreamCFD

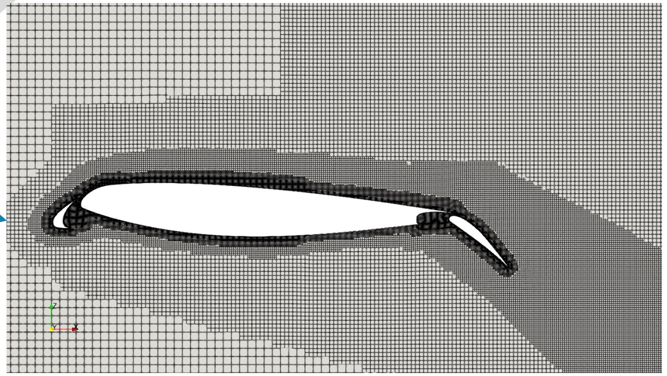
Exploring the limits: The high-lift HPC Grand Challenge case

HL-CRM Grand Challenge test case

- High-lift version of the NASA Common Research Model generic aircraft
  - Test case 2.4 from the NASA / AIAA High-Lift Prediction Workshop
- Compressible wall-modelled LES simulations
  - libWMLES library of T. Mukha
  - Validation and best practice derivation for channel flow and 30P30N three-element airfoil (exaFOAM MB9)
- Templated mesh setup using snappyHexMesh
  - Flow-adapted refinement zones
  - Refinement based on boundary layer thickness
  - Wide range of mesh sizes can be generated in parallel
- The objective:
  - Find maximum grid size and core count for which simulations can be successfully and efficiently conducted
  - Find out what the limiting factors are



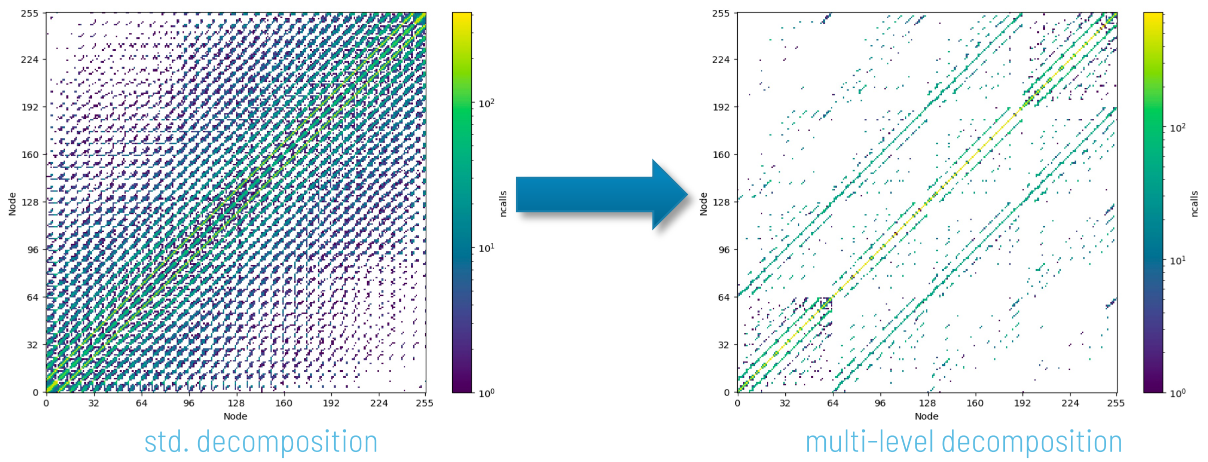


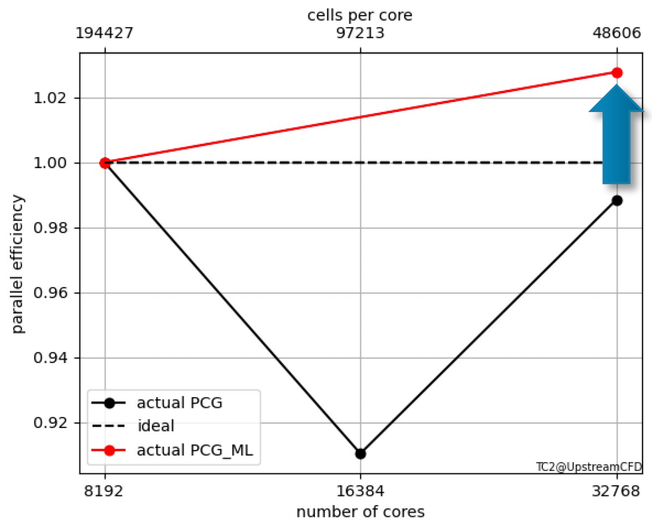


Outcome

- Testing on the LUMI-C supercomputer:
  - Simulation on mesh with 4.4 billion cells
  - Simulation on 65 536 CPU cores
- Domain decomposition was limiting factor
  - Domain decomposition stage exceeded queue limit for high core counts
  - Scotch & KaHIP decomposition methods failed for such mesh sizes / counts: Had to use hierarchical
  - A viable solution to this problem has been developed in exaFOAM:
    - The coherent mesh representation approach
    - See later presentation by Gregor Weiß (HLRS) for details
- Multi-level decomposition approach:
  - 3-levels (# nodes, # CPUs per node, # cores per CPU)
  - Strongly reduces communication density
  - Enables superlinear scaling on 32k cores

MPI calls between nodes (256 nodes):

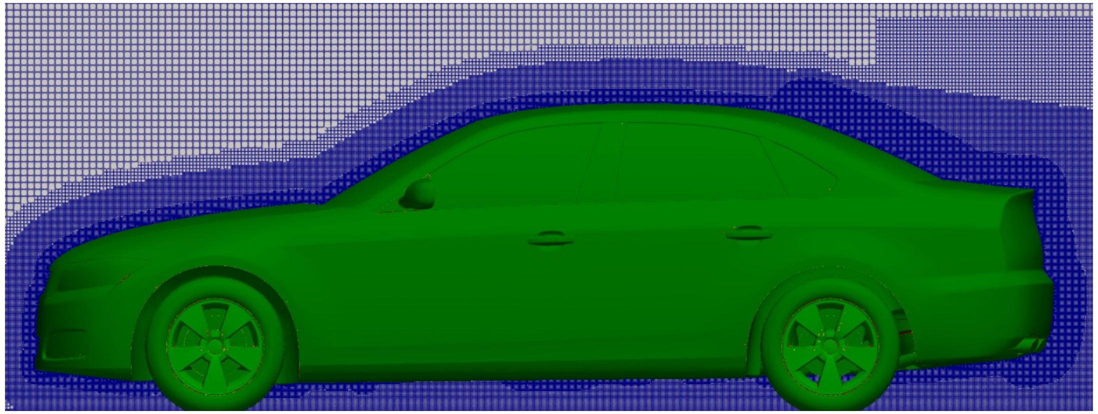


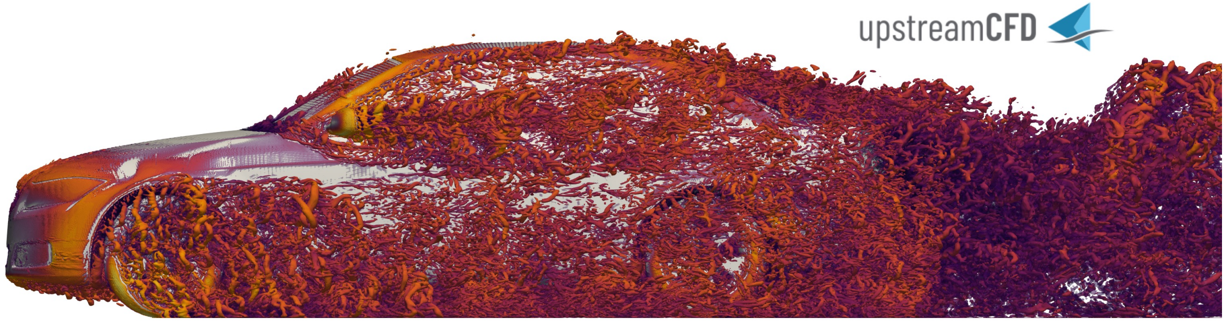


Another performance bottleneck: Rotating wheels

exaFOAM industrial application case B5: "DrivAer-rot"

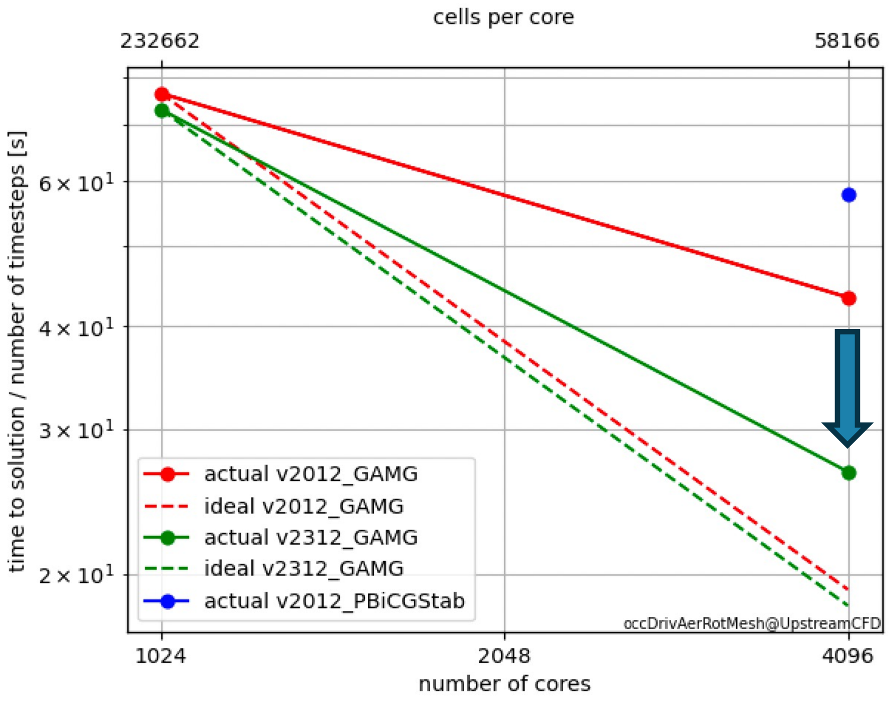
- The rotation of wheels has an important effect on the overall aerodynamics of a car and must be captured in simulations
- Modelling approach:
  - Non axisymmetric parts (wheel spokes) allocated to separate rotating-mesh zone
  - Zone interface:
    - Must be a body of rotation
    - Interpolation between rotating and stationary mesh zones using Arbitrary Coupled Mesh Interface (ACMI) approach
  - Axisymmetric rotating parts (e.g. tyre) treated with rotating wall velocity BC
- The mesh rotation and interpolation introduces a computational overhead and a significant bottleneck to scalability
- Test case (nominated by exaFOAM Stakeholder Volkswagen AG):
  - Rotating-wheel variant of DrivAer Notchback
  - Adapted from the Ford OCDA case geometry ([Hupertz et al., 2022]())
  - snappyHexMesh grid, 238M cells
  - Incompressible simulation with DES
  - [Test case download link]()





Scalability comparison: before/after exaFOAM

- Comparison of OpenFOAM release versions:
  - v2012, before start of project
  - v2312, most recent release
- Numerous improvements to A(C)MI method, e.g.:
  - Non-blocking approach communication approach implemented by ESI-OpenCFD
  - See OpenFOAM release notes for details
- Testing on the LUMI-C supercomputer using the DrivAer-rot case:
  - 1.63x improvement in Time-To-Solution when running on 4096 cores
  - Less significant improvement seen on lower core counts (1.05x on 1024 cores)



Conclusions and outlook

Conclusions and outlook

- High-fidelity CFD is a valuable contributor to innovation
  - e.g. source of reliable training data for Machine Learning
- HPC has always been a key limiting factor
- Enabling factors for high-fidelity CFD:
  - HPC in the cloud:
    - Access to 1000s of latest-generation CPU cores on demand
    - Particularly valuable for SMEs and startups
  - Superlinear scalability:
    - Potential for breakthroughs in turnaround time and efficiency
    - Requires improved technologies across the entire workflow to harness this
      - e.g. Coherent mesh representation approach developed by HLRS and Wikki within exaFOAM (see today's presentation by Gregor Weiß)
- Important to run benchmarks for scale-resolving simulations
  - Don't just focus on the lid-driven cavity
  - Numerous HPC benchmark cases for SRS published by the exaFOAM project: https://exafoam.eu/benchmarks/
- CPUs vs. GPUs:
  - Don't write off CPUs just yet
    - A lot more performance can be gained from <u>today's</u> fully-featured OpenFOAM
    - And chip manufacturers, please keep adding that cache!
  - However, a high-performance GPU implementation of OpenFOAM will be essential for future viability
    - Proprietary codes are demonstrating huge performance and energy efficiency gains
    - It's clear that "deep" integration is required to unlock the true performance potential
      - This is challenging, and will require intense collaboration from the open-source community
    - See today's talk by Simone Bnà (Cineca) for an overview
    - See also efforts in the ongoing EXASIM project (a partner project of exaFOAM)
      - https://exasim-project.com

Acknowledgments

upstream CFD

The computational time leading to these results were granted by multiple European HPC supercomputers

- EuroHPC Benchmark access at Vega:
  - The authors gratefully acknowledge the HPC RIVR consortium (www.hpc-rivr.si) and EuroHPC JU (eurohpc-ju.europa.eu) for funding this research by providing computing resources of the HPC system Vega at the Institute of Information Science (www.izum.si).
- EuroHPC Development access at LUMI
  - We acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LUMI, hosted by CSC (Finland) and the LUMI consortium through a EuroHPC Development Access call.
- Test account at HAWK (HLRS)
  - Some simulations were performed on the national supercomputer HPE Apollo Hawk at the High Performance Computing Center Stuttgart (HLRS) under a test account.

Computation of the ML training dataset was provided by Amazon Web Services, Inc. (https://aws.amazon.com)

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

EuroHPC Joint Undertaking

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 956416. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, United Kingdom, Germany, Italy, Croatia, Spain, Greece, Portugal

upstreamCFD

Thank you for your attention

Any questions?