Sub-Grid Scale Modelling at Scale with Deep Learning and up to 60 Billion Degrees of Freedom

M. Bode, D. Denker, J. Jitsev, H. Pitsch

published in

NIC Symposium 2020
M. Müller, K. Binder, A. Trautmann (Editors)

Forschungszentrum Jülich GmbH,
John von Neumann Institute for Computing (NIC),
Schriften des Forschungszentrums Jülich, NIC Series, Vol. 50,
ISBN 978-3-95806-443-0, pp. 379.
http://hdl.handle.net/2128/24435

© 2020 by Forschungszentrum Jülich
Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

Sub-Grid Scale Modelling at Scale with Deep Learning and up to 60 Billion Degrees of Freedom

Mathis Bode[1], **Dominik Denker**[1], **Jenia Jitsev**[2], **and Heinz Pitsch**[1]

[1] Institute for Combustion Technology, RWTH Aachen University, 52056 Aachen, Germany
E-mail: {m.bode, d.denker, h.pitsch}@itv.rwth-aachen.de

[2] Jülich Supercomputing Centre, Institute for Advanced Simulation, Forschungszentrum Jülich, 52425 Jülich, Germany
E-mail: j.jitsev@fz-juelich.de

This work presents fully resolved direct numerical simulations (DNSs) of a turbulent reactive planar temporally non-premixed jet configuration with up to 60 billion degrees of freedom. As scalar mixing is of utmost importance for this kind of configuration, a novel deep learning (DL) approach in the context of large-eddy simulation is presented which results in predictive mixing statistics on underresolved grids. The usability of the mixing model is approved by applying it to the DNS data. Furthermore, node performance measurements for the training of the DL networks are shown for different computing clusters.

1 Introduction

Turbulent fluid flows pose some of the most difficult and fundamental problems in classical physics as turbulence is a complex, strongly non-linear, multi-scale phenomenon.[1] Consequently, the prediction of the statistics of fluctuating velocity and scalar fields and the development of models for a precise statistical prediction of these fields even in scale-resolved simulations is one of the main challenges in turbulence research.[2, 3] An important example is the prediction of scalar mixing in turbulent flows as it controls many processes of industrial relevance.

One often used modelling approach for turbulent flows is large-eddy simulation (LES). LES solves for the larger, flow-dependent scales of the flow and models all scales below a particular filter width.[4, 5] By assuming that the smaller, unresolved scales reveal certain universal features and decouple from the larger non-universal scales, models for LES can be built from relatively simple, semi-empirical algebraic relations. A systematic approach for developing such models is to perform fully resolved direct numerical simulations (DNSs), filter the resulting data with a given filter kernel, and find relations between the filtered data and the original DNS results. The objective of the present work is to develop a LES model for scalar mixing in turbulent flows based on deep learning (DL). This data-driven approach reconstructs the subfilter statistics, which can be used to close the filtered Navier-Stokes equations.

DL is one of the most emerging research topics in the age of big data. Examples of DL applications include image processing,[6–9] voice recognition,[10] or website customisation.[11] One reason for this is the continued growth of computational power (especially GPUs). Another reason is the availability of exceptionally large labelled experimental datasets. Data-driven methods and DL have also become more popular over the last years in the field of fluid mechanics including turbulence research.[12–15] However, simple networks or small and artificial datasets have often limited these studies in the past.

The remainder of this article is organised as follows. Sec. 2 describes the used reactive datasets. In Sec. 3, details about the data-driven reconstruction methodology are given, and results for inert data are discussed. The scalar mixing LES model is applied to the reactive data in Sec. 4. The paper finishes with conclusions.

2 Reactive DNS

Combustion in turbulent flows can be classified in several regimes based on the level of interaction of the turbulent flow field and the combustion chemistry. These combustion regimes have severe implications to the choice of adequate combustion models employed in the simulation of real world engineering applications. However, in the limit of intense scale interaction between turbulence and flame scales, no satisfactory combustion model exists. In this combustion regime, the well defined and understood flame structure is disintegrated by turbulent eddies. Unfortunately, the lack of suitable combustion models in this regime limits the quality of the prediction of LESs of novel and promising combustion processes, which rely on low temperatures and consequently feature slow combustion time scales. As both turbulence and chemistry scales must be fully resolved for an appropriate investigation of these combustion conditions, the conduction of high fidelity direct numerical simulations (DNS) is inevitable. To this end, a series of five DNSs of planar temporally evolving non-premixed jet flames was performed on the supercomputers JURECA, JUQUEEN, and JUWELS. The DNSs feature methane as fuel and air as oxidiser. Different levels of dilution of the fuel, as well as increasing turbulence intensities, allow for a precise determination of combustion regime boundaries. The configurations of the DNSs, as well as important non-dimensional numbers are summarised in Tab. 1. As simulating real world engineering configurations is currently impossible due to the immense computational resource requirements, the idealised configuration of the planar temporally evolving jet was chosen for maximising the flame surface as well as for the ease of obtaining statistics. In this configuration, the fuel stream is situated in stream-wise centre slab of the domain and the oxidiser is moving in the opposing direction at the upper and lower part of the domain. Two highly turbulent flame fronts form in the shear layers between the fuel and oxidiser streams.

	Low Re low dilution case	Low Re high dilution case	Intermediate Re case	High Da case	High Re case
$\mathrm{Re}_{\mathrm{jet},0}$	4500	4500	6000	6000	10 000
Da_τ	0.125	0.150	0.150	0.450	0.150
Z_{st}	0.20	0.45	0.45	0.45	0.45
$n_{\mathrm{gridpoints}}\,[10^9]$	0.4	0.3	0.6	1.6	1.2
DOF $[10^9]$	15	11	23	60	45

Table 1. Important non-dimensional numbers and numerical parameters of the DNSs of planar temporally evolving non-premixed jet flames conducted on the supercomputers JUQUEEN, JURECA, and JUWELS. The initial jet Reynolds number $\mathrm{Re}_{\mathrm{jet},0}$ indicates the ratio of the largest to the smallest turbulent length scales, the Damköler number Da_τ indicates the inverse level of flame extinction and the stoichiometric mixture fraction Z_{st} reflects the level of the dilution of the fuel stream. DOF indicates the degrees of freedom of the system of equations solved in each time step.

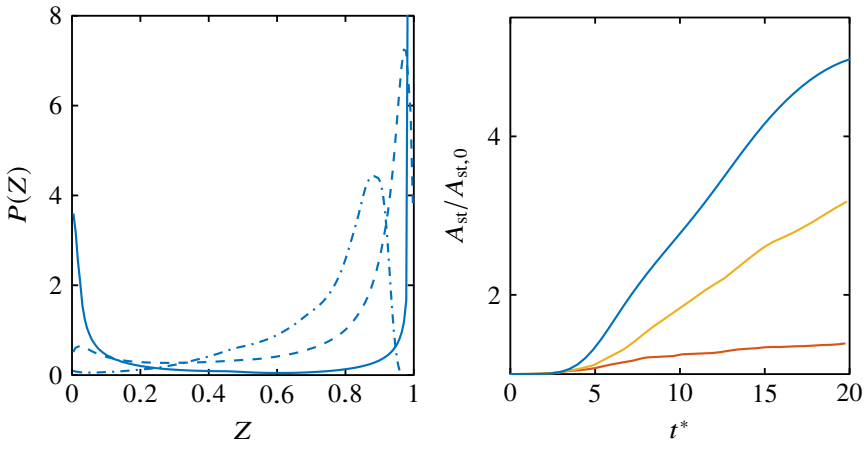

Figure 1. Left: PDF of the mixture fraction $P(Z)$ in the centre slab of the high Re case for three different time steps (solid line: early time step, dashed line: intermediate time step, dashed-dotted line: last time step). Right: temporal evolution of the normalised flame surface area $A_{\mathrm{st}}/A_{\mathrm{st},0}$ (red line: Low dilution case, yellow line: intermediate Re case, blue line: high Re case).

As fuel and oxidiser are initially separated and a combustable mixture must be formed prior to combustion, applications using non-premixed combustion rely heavily on turbulent mixing to achieve high efficiencies. Therefore, knowledge of the turbulent mixing process is of upmost importance for the reliable modelling of non-premixed combustion. The turbulent mixing process is illustrated by the example of the probability density function (PDF) of the mixture fraction $P(Z)$ in the centre slab of the domain of the High Re case in Fig. 1 (left). In the early time step, $P(Z)$ resembles a double-δ PDF with peaks at $Z = 0$ (pure oxidiser) and $Z = 1$ (pure fuel), which highlights the initial separation of the two streams. In later time steps, the PDF changes shape as turbulent transport continuously causes fuel and oxidiser to mix. In the final time step of the simulation, neither pure fuel nor pure oxidiser is present in the central slab.

The consequences of increased turbulent mixing for the combustion is demonstrated by the temporal evolution of the normalised flame surface area $A_{\mathrm{st}}/A_{\mathrm{st},0}$, with the initial flame surface area $A_{\mathrm{st},0}$, in Fig. 1 (right). A_{st} quickly rises with higher Reynolds numbers and, consequently, the overall consumption of fuel in the domain is increased significantly as higher turbulence intensities are reached.

The configuration of the DNS and the highly turbulent nature of the combustion is illustrated in Fig. 2. The exceedingly corrugated iso-surface of the stoichiometric mixture fraction is shown for the highest turbulence intensity case. The stoichiometric mixture fraction indicates the position of the reaction zone in non-premixed flames. Additionally, the local value of the temperature is displayed to show spots of local extinction.

The DNS were performed solving the reactive, unsteady Navier-Stokes equations in the low Mach number limit using the in-house solver CIAO.[16, 17] The momentum equations are spatially discretised with a fourth-order scheme and time advancement is performed using a semi-implicit Crank-Nicholson time integration.[18] The transport of species mass fractions is described using the Hirschfelder and Curtiss approximation to the diffusive

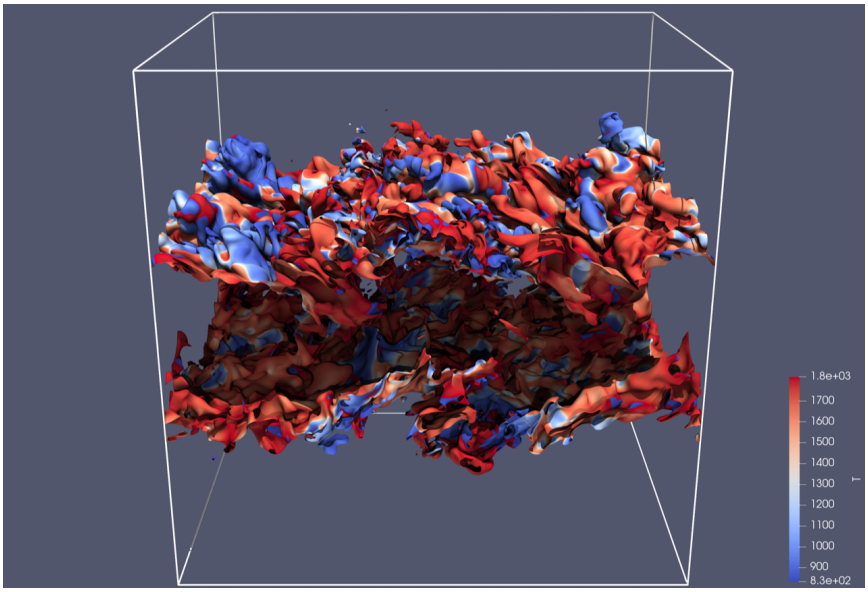

Figure 2. Iso-surface of the stoichiometric mixture fraction in the High Re case. The colour of the iso-surface indicates the local value of the temperature.

fluxes together with a velocity-correction approach for mass conservation. The temperature and species equations are advanced by introducing the symmetric operator split of Strang.[19] The chemistry operator uses a time-implicit backward difference method.[20] The chemistry is included using finite rate chemistry employing a mechanism for the oxidation of methane comprising 30 species and 102 reactions capable of capturing the anticipated effects of localised extinction and re-ignition.

To resolve all scales, from the detailed flame structure to the integral scales of the turbulent flow fields, computational meshes containing 0.3×10^9 to 1.6×10^9 grid points were employed. This results in a system of equations with up to 60 billion degrees of freedom which must be solved for each of the approximately $11\,000$ time steps needed for the transition from an initial laminar solution to a fully turbulent flame. The largest case in this series, the High Da case, consumed a total of 42 million CPU-h on the supercomputer JUQUEEN.

3 Modelling

The objective of this work is to move beyond conventional models for LES and develop a data-driven approach, which is able to predict instantaneous, fully-resolved three-dimensional flow realisations including scalar fields, such as the mixture fraction, by knowing solely the coarse-grained data fields. This approach generalises deconvolution techniques as no explicit knowledge of the filter kernel is required.

A neural network called turbulence reconstruction generative adversarial network (TR-GAN) is used in this work and predominantly motivated by the previous single image

super-resolution (SISR) works. A pioneer work for SISR was the super-resolution convolutional neural network (SRCNN) designed by Dong *et al.*[6] Great efforts towards SISR had been made through the community since SRCNN first emerged. Yet most of these SISR networks showed the deficiency of producing over-smoothed results, owing to focusing on the peak-signal-to-noise-ratio (PSNR) as the primary metrics. Resultantly, the high frequency details occurring in the ground truth could not be adequately reconstructed. This makes such networks especially fatal for turbulence applications, as the small-scale (high frequency) structures at high wave numbers are exactly what is intended to be restored. Amelioration regarding the PSNR-based SISR networks was elucidated by Johnson *et al.*,[21] who presented the concept "perceptual loss", *i. e.* a criterion to optimise the performance in the VGG-features space that substitutes the PSNR which optimised in the pixel space.

The other fundamental framework named generative adversarial network (GAN) was presented by Ref. 22. A GAN is composed of two models, a generator that captures the data distribution and generates new data, and a discriminator that learns to distinguish whether a sample stems from the original data distribution (genuine) or the generator (fake). During training, the generator learns to produce samples that are indistinguishable for the discriminator, while the discriminator learns to more accurately judge the genuineness.

For this work, the state-of-art ESRGAN was adapted to three-dimensional (3-D) turbulence data and a problem-specific, physically-based loss functions was introduced. The resulting network is called turbulence reconstruction GAN (TRGAN). The TRGAN is supposed to recuperate the fully-resolved DNS results from coarse turbulence data, such as LES data.

The perceptual loss proposed for the ESRGAN based on VGG-feature space is apparently not as suitable for the turbulence data as the geometrical features from VGG19 are not representative for turbulent flows. Hence, a new formulation for the cost function was developed inspired by physical flow constraints.

Before training the TSRGAN as a combined model, the generator is pretrained with root-mean-square error (RMSE) due to the complexity of the residual-in-residual-dense-block (RRDB). For the combined model, the loss function for reconstructing velocity fields is proposed as

$$l = \beta_1 l_{\mathrm{RADG}} + \beta_2 l_{\mathrm{pixel}} + \beta_3 l_{\mathrm{gradient}} + \beta_4 l_{\mathrm{continuity}} \tag{1}$$

with $\beta_1, \beta_2, \beta_3$, and β_4 being coefficients weighting the different loss term contributions. l_{RADG} is the 'realistic average' discriminator/generator loss, which is the accuracy feedback between discriminator and generator as given by Wang *et al.*[7] For a scalar ϕ, the pixel loss l_{pixel} is defined as

$$l_{\mathrm{pixel}} = \mathrm{MSE}(\phi^{\mathrm{predicted}}, \phi^{\mathrm{DNS}}) \tag{2}$$

The mean-square error (MSE) operator is given by

$$\mathrm{MSE}(\{\cdot\}_1, \{\cdot\}_2) = \frac{1}{N_{\mathrm{samples}}} \sum_{i=1}^{N_{\mathrm{samples}}} (\{\cdot\}_1^i - \{\cdot\}_2^i)^2 \tag{3}$$

with N_{samples} as the number of all samples, *i. e.* the total number of grid points of the reconstructed field. If the MSE operator is applied on tensors including vectors, it is applied

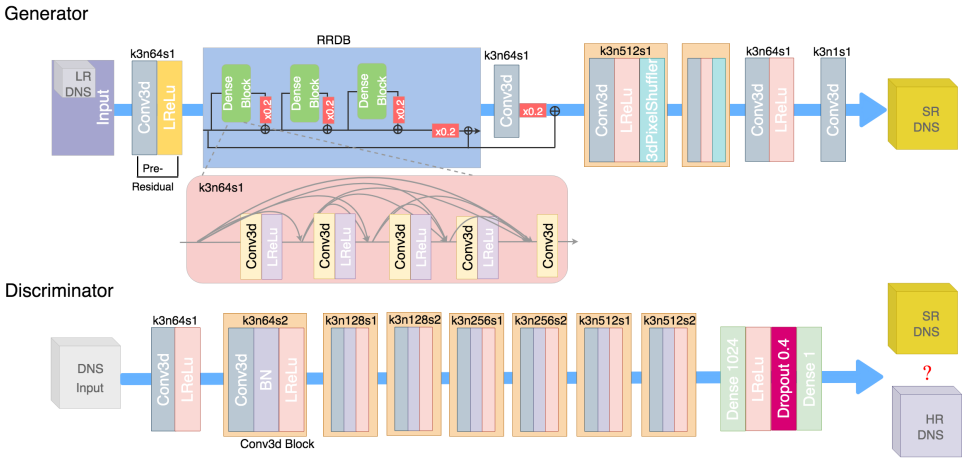

Figure 3. TRGAN network structure. Image from Bode *et al.*[13]

to all elements separately. Afterwards the resulting tensor is mapped into a scalar using the 1-norm. The gradient loss l_{gradient} is defined as

$$l_{\text{gradient}} = \text{MSE}(\nabla\phi^{\text{predicted}}, \nabla\phi^{\text{DNS}}) \tag{4}$$

with ∇ as del operator. $l_{\text{continuity}}$ is the continuity loss, which enforces the continuity equation in the reconstructed field and reads

$$l_{\text{continuity}} = \text{MSE}(\nabla \cdot \mathbf{u}^{\text{predicted}}, \mathbf{0}) \tag{5}$$

with $\mathbf{u}$ as velocity vector. The TRGAN network is graphically presented in Fig. 3. Note that differently from the state-of-art SR networks, the TRGAN does not involve up-/ or downsampling layers. The high-/ and low-resolution turbulence data refer to the energy spectral property of the discussed turbulence, not the trivial image resolution.

The TRGAN was implemented by using Keras API with TensorFlow backend, and the training was performed on the JURECA GPU nodes due to the large size of the DNS dataset. The results for a reconstructed velocity and scalar field are shown in Figs. 4 and 5 for an inert case. Homogeneous isotropic turbulence data were used, and the quality of the DL closure can be seen.

Training with large data, such as the reactive DNS dataset used in this work, is not possible without computing clusters. The GPU performance of JURECA, JUWELS (both at Jülich Supercomputing Centre, Forschungszentrum Jülich), and Claix18 (at IT Center, RWTH Aachen University) and their scaling performance are depicted in Fig. 6. Here, the numerical values give the average number of subboxes with dimension $16 \times 16 \times 16$, which were trained to the TRGAN within one minute. The better performance of the Tesla V100 GPUs (JUWELS, CLAIX18) compared to the Tesla K80 GPUs (JURECA) is obvious.

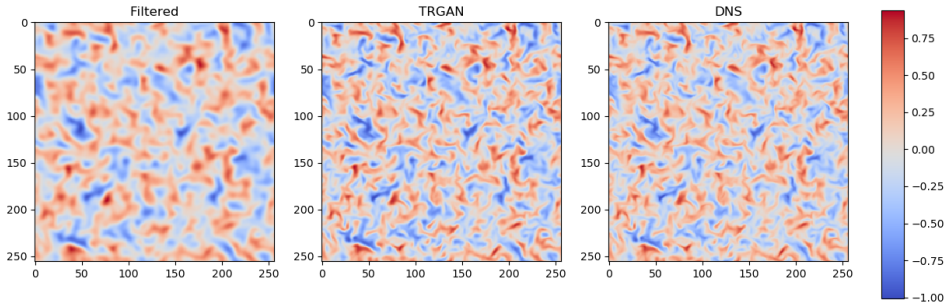

Figure 4. Velocity reconstruction.

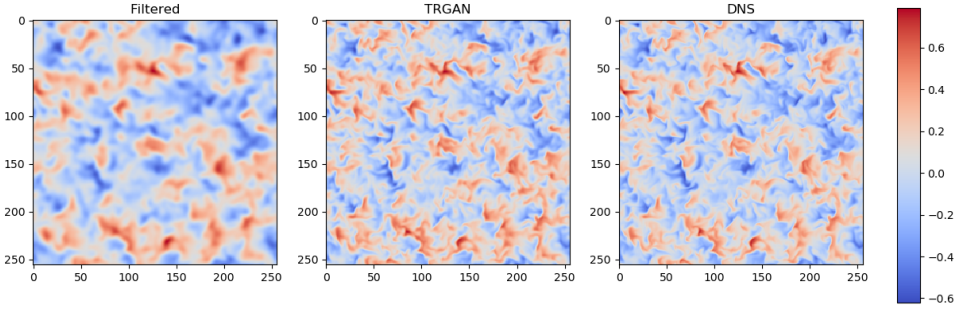

Figure 5. Passive scalar reconstruction.

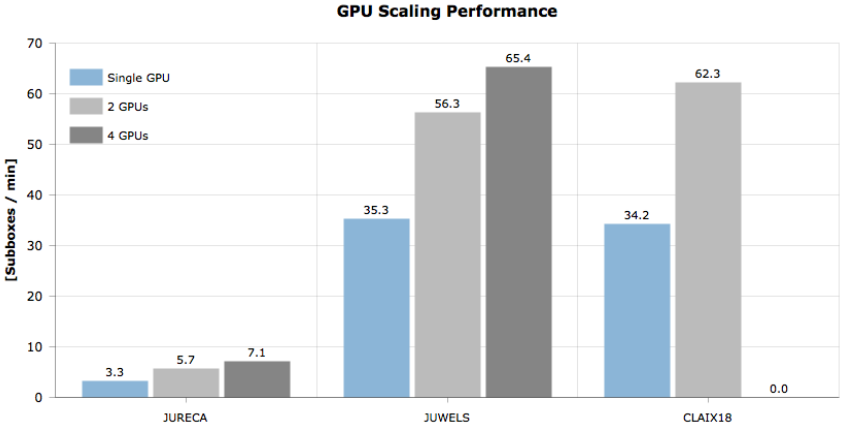

Figure 6. GPU performance.

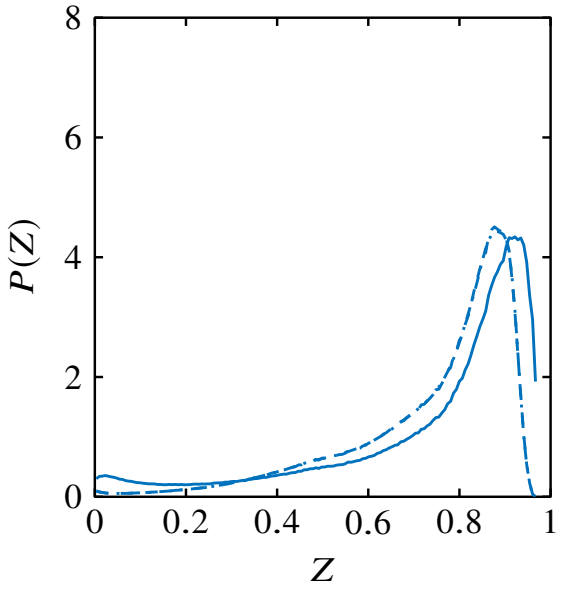

Figure 7. PDF of the mixture fraction $P(Z)$ in the centre slab of the high Re case for the last time step from Fig. 1 (solid line: no model, dashed line: DL model, dashed-dotted line: DNS).

4 Application

In order to validate the data-driven DL modelling approach presented in Sec. 3, the high Reynolds number DNS data of an early time step are filtered and interpolated to a coarser mesh with only one third of the number of points per direction compared to the corresponding DNS. Using this coarser mesh, the mixing results shown in Fig. 1 are recomputed with and without the DL model. The result is compared to the DNS result. As can be seen in Fig. 7, the DNS and modelled results match very well while the underresolved simulation without model mixes too slow.

5 Conclusions

Fully resolved DNSs of a turbulent reactive non-premixed planar temporal jet configuration and a data-driven DL modelling approach are presented in this work. It is shown how the combination of computationally expensive simulations and novel LES modelling methods can be used to accurately predict scalar mixing even on underresolved grids. An estimation for training cost is given.

Acknowledgements

The authors gratefully acknowledge the computing time granted for the projects JHPC09, JHPC22, and JHPC55 by the JARA Vergabegremium and provided on the JARA Partition part of the supercomputers JUQUEEN and JURECA at Forschungszentrum Jülich. In

addition, the authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding project JHPC09 by providing computing time on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). Furthermore, the authors want to thank Michael Gauding, Zeyu Lian, and Antonio Attili for their help.

References

1. D. Ruelle and F. Takens, *On the nature of turbulence*, Les rencontres physiciens-mathématiciens de Strasbourg-RCP25 **12**, 1–44, 1971.
2. U. Piomelli, *Large-eddy simulation: achievements and challenges*, Progress in Aerospace Sciences **35**, 335–362, 1999.
3. A. R. Kerstein, *Turbulence in combustion processes: Modeling challenges*, Proceedings of the Combustion Institute **29**, 1763–1773, 2002.
4. M. Germano, U. Piomelli, P. Moin, and W. H. Cabot, *A dynamic subgrid-scale eddy viscosity model*, Physics of Fluids A: Fluid Dynamics **3**, 1760–1765, 1991.
5. A. Leonard, *Energy cascade in large-eddy simulations of turbulent fluid flows*, in Turbulent Diffusion in Environmental Pollution, Advances in Geophysics **18**, 237–248, 1975.
6. C. Dong, C. C. Loy, K. He, and X. Tang, *Learning a deep convolutional network for image super-resolution*, in Computer Vision, Lecture Notes in Computer Science **8692**, Springer, 184–199, 2014.
7. X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*, Lecture Notes in Computer Science **11133**, 63–79, 2019.
8. H. Greenspan, B. Van Ginneken, and R. M. Summers, *Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique*, IEEE Transactions on Medical Imaging **35**, 1153–1159, 2016.
9. N. Wang and D.-Y. Yeung, *Learning a deep compact image representation for visual tracking*, in Advances in Neural Information Processing Systems **26**, 809–817, 2013.
10. G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, B. Kingsbury *et al.*, *Deep neural networks for acoustic modeling in speech recognition*, IEEE Signal processing magazine **29**, 82–97, 2012.
11. M. Langheinrich, A. Nakamura, N. Abe, T. Kamba, and Y. Koseki, *Unintrusive customization techniques for web advertising*, Computer Networks **31**, 1259–1272, 1999.
12. M. Bode, M. Gauding, J. H. Göbbert, B. Liao, J. Jitsev, and H. Pitsch, *Towards Prediction of Turbulent Flows at High Reynolds Numbers Using High Performance Computing Data and Deep Learning*, Lecture Notes in Computer Science **11203**, 614–623, 2018.
13. M. Bode, M. Gauding, K. Kleinheinz, and H. Pitsch, *Deep learning at scale for subgrid modeling in turbulent flows*, 2019, arXiv:1910.00928 [physics.comp-ph].
14. J. N. Kutz, *Deep learning in fluid dynamics*, Journal of Fluid Mechanics **814**, 1–4, 2017.
15. R. Maulik and O. San, *A neural network approach for the blind deconvolution of turbulent flows*, Journal of Fluid Mechanics **831**, 151–181, 2017.

16. M. Bode, M. Davidovic, and H. Pitsch, *Towards Clean Propulsion with Synthetic Fuels: Computational Aspects and Analysis*, in High Performance Computing in Science and Engineering '18, W. Nagel W., D. Kröner D., M. Resch M. (Editors), Springer, 185–207, 2018.
17. M. Bode, N. Collier, F. Bisetti, and H. Pitsch, *Adaptive chemistry lookup tables for combustion simulations using optimal B-spline interpolants*, Combustion Theory and Modelling **23**, 674–699, 2019.
18. O. Desjardins, G. Blanquart, G. Balarac, and H. Pitsch, *High Order Conservative Finite Difference Scheme for Variable Density Low Mach Number Turbulent Flows*, J. Comp. Phys. **227**, 7125–7159, 2008.
19. G. Strang, *On the construction and comparison of difference schemes*, SIAM J. Numer. Anal. **5**, 506–517, 1968.
20. A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, and C. S. Woodward, *SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers*, ACM Transactions on Mathematical Software (TOMS) **31**, 363–396, 2005.
21. J. Johnson, A. Alahi, and L. Fei-Fei, *Perceptual losses for real-time style transfer and super-resolution*, in European Conference on Computer Vision, Springer, 694–711, 2016.
22. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets*, in Advances in Neural Information Processing Systems **27**, 2672–2680, 2014.