The Well: a Large-Scale Collection of Diverse Physics Simulations for Machine Learning

Ruben Ohana [1,2,*], **Michael McCabe** [1,*], **Lucas Meyer** [1], **Rudy Morel** [1,2], **Fruzsina J. Agocs** [2,3,†], **Miguel Beneitez** [4,†], **Marsha Berger** [2,5,†], **Blakesley Burkhart** [2,6,†], **Stuart B. Dalziel** [4,†], **Drummond B. Fielding** [2,7,†], **Daniel Fortunato** [2,†], **Jared A. Goldberg** [2,†], **Keiya Hirashima** [1,2,8,†], **Yan-Fei Jiang** [2,†], **Rich R. Kerswell** [4,†], **Suryanarayana Maddu** [2,†], **Jonah Miller**[9,†], **Payel Mukhopadhyay** [10,†], **Stefan S. Nixon** [4,†], **Jeff Shen** [11,†], **Romain Watteaux** [12,†], **Bruno Régaldo-Saint Blancard** [1,2], **Liam H. Parker**[1,2,10], **Miles Cranmer** [1,4], **Shirley Ho** [1,2,5,11]

[1] Polymathic AI [2] Flatiron Institute [3] University of Colorado, Boulder [4] University of Cambridge
[5] New York University [6] Rutgers University [7] Cornell University [8] University of Tokyo
[9] Los Alamos National Laboratory [10] University of California, Berkeley
[11] Princeton University [12] CEA DAM

Abstract

Machine learning based surrogate models offer researchers powerful tools for accelerating simulation-based workflows. However, as standard datasets in this space often cover small classes of physical behavior, it can be difficult to evaluate the efficacy of new approaches. To address this gap, we introduce *the Well:* a large-scale collection of datasets containing numerical simulations of a wide variety of spatiotemporal physical systems. The Well draws from domain experts and numerical software developers to provide 15TB of data across 16 datasets covering diverse domains such as biological systems, fluid dynamics, acoustic scattering, as well as magneto-hydrodynamic simulations of extra-galactic fluids or supernova explosions. These datasets can be used individually or as part of a broader benchmark suite. To facilitate usage of the Well, we provide a unified PyTorch interface for training and evaluating models. We demonstrate the function of this library by introducing example baselines that highlight the new challenges posed by the complex dynamics of the Well. The code and data is available at `https://github.com/PolymathicAI/the_well`.

1 Introduction

Simulation is one of the most ubiquitous and important tools in the modern computational science and engineering toolbox. From forecasting [1–3], to optimization [4, 5], to parameter inference [6, 7], practitioners lean heavily on simulation to evaluate how physical systems will evolve over time in response to varying initial conditions or stimuli. For many physical phenomena, this evolution can be described by systems of *partial differential equations* (PDEs) which model fundamental physical behavior aggregated to the continuum level under different material assumptions. Unfortunately, finding analytical solutions is infeasible for all but restricted classes of PDEs [8]. As a result, *numerical methods* which solve discretized versions of these equations with well-understood convergence and approximation properties have become the preeminent approach in this space. However, in some cases, numerical methods can provide accuracy in excess of what is needed for applications at significant computational cost while lower resolution direct simulation

*Equal contribution. Contact: `{rohana, mmccabe}@flatironinstitute.org`
†Domain expert, alphabetical order.

D3S3: Data-driven and Differentiable Simulations, Surrogates, and Solvers @ NeurIPS 2024.

may not resolve key features of the dynamics. This has spurred the development of faster, simplified models referred to as *surrogate models* that resolve only the essential features for a given scale of simulation [9, 10].

It is this surrogate modeling space where deep learning is poised to make a significant impact [11–13] with tangible results already demonstrated across diverse sets of fields and applications [3, 14–18]. Yet despite these successes, deep learning based surrogates face significant challenges in reaching broader adoption. One reason for this is the gap between the complexity of problems of practical interest and the datasets used for evaluating these models today. Scaling analysis has shown that deep learning-based surrogates can require large amounts of data to reach high accuracy [19, 20]. Meanwhile, even at resolutions accessible to modern machine learning architectures, high-quality scientific simulation can require the combination of specialized software, domain expertise, and months of supercomputer time [21]. On the other hand, from the perspective of scientists running these simulations, even just storing the frequent snapshots necessary for conventional off-line deep learning training is a significant and unnecessary expense [22–24]).

To address this gap, we introduce *the Well*, a diverse 15 TB collection high quality numerical simulations produced in close collaboration with domain scientists and numerical software developers. The Well is curated to provide challenging learning tasks at a scale that is approachable to modern machine learning but where efficiency remains an important concern. It contains 16 datasets ranging across application domains, scales, and governing equations from the evolution of biological systems to the growth of galaxies. Each dataset contains temporally coarsened snapshots from simulations of a particular physical phenomenon across multiple initial conditions or physical parameters, while providing a sufficiently large number of snapshots to explore simulation stability. Furthermore, the Well provides machine learning researchers with complex, demanding benchmarks that will inform the development of the next generation of data-driven surrogates.

Related Work. Modern machine learning relies on massive, curated and diverse datasets [25–28]. Natural language processing is built on internet-scale datasets [29–32], while vision models have grown to utilize sets containing billions of text-images pairs [33]. These datasets are sufficiently diverse that model improvement can be derived from sophisticated filtering approaches [32, 34, 35].
On the other hand, datasets designed for physical dynamics prediction are still growing. Early datasets featured a variety of common reference simulations [36–38]. While these datasets have seen rapid adoption, the broader trend has moved towards more complex but specialized simulation datasets [39–45]. These have opened new application areas for deep learning but have typically been limited to a small number of tasks. Other datasets have tackled more ambitious high-resolution problems [46–48], but the limited number of snapshots and scale of individual samples often restricts their usage. New datasets which offer complexity, volume, and diversity simultaneously are necessary for holistic evaluation of individual models and for the emerging trend of multiple physics foundation models [49–54]. The Well provides unified access to a collection of physical scenarios and benchmarking tools that are both diverse and challenging.

2 Diving into the Well

Format. The Well is composed of 16 datasets totaling 15TB of data with individual datasets ranging from 6.7GB to 4.9TB. The data is provided on uniform grids and sampled at constant time intervals. Data and associated metadata are stored in self-documenting `HDF5` files [55]. All datasets use a shared data specification described in the supplementary materials and a PyTorch [56] interface is provided. These files include all available state variables or spatially varying coefficients associated with a given set of dynamics in `numpy` [57] arrays of shape `(n_traj, n_steps, coord1, coord2, (coord3))` in single precision `fp32`. We distinguish between scalar, vector, and tensor-valued fields due to their different transformation properties. Each file is randomly split into training, testing, and validation sets with a split of 0.8/0.1/0.1 * `n_traj`. Details of individual datasets are given in Table 2.

Extensibility. The PyTorch interface can process any data file following the provided specification without any additional modification to the code base. Scripts are provided to check whether `HDF5` files are formatted correctly. This allows users to easily incorporate third-party datasets into pipelines using the provided benchmarking library.

Description of the datasets. The rest of this section provides physical intuition and background for the scenarios contained in the datasets. Visualizations are in Appendix in Figures 1–5. Technical details on the underlying physics, fields, physical parameters, and the generating processes for the datasets are given in Supplementary Material.

`acoustic_scattering`. Acoustic scattering possesses simple linear dynamics that are complicated by the underlying geometry. In this dataset, we model the propagation of acoustic waves through a domain consisting of substrata with sharply variable density in the form of maze-like walls (Figure 1,

top) or pockets with vastly differing compositions. These simulations are most commonly seen in inverse problems including source optimization and inverse acoustic scattering in which sound waves are used to probe the composition of the domain.

`active_matter`. Active matter systems are composed of agents, such as particles or macromolecules, that transform chemical energy into mechanical work, generating active forces or stresses. These forces are transmitted throughout the system via direct steric interactions, cross-linking proteins, or long-range hydrodynamic interactions, leading to complex spatiotemporal dynamics (Figure 1, middle). These simulations specifically focus on active particles suspended in a viscous fluid leading to orientation-dependent viscosity with significant long-range hydrodynamic and steric interactions.

`convective_envelope_rsg`. Massive stars evolve into red supergiants (RSGs), which have turbulent and convective envelopes. Here, 3D radiative hydrodynamic (RHD) simulations model these convective envelopes, capturing inherently 3D processes like convection (Figure 1, bottom). The simulations give insight into a variety of phenomena: the progenitors of supernovae (SN) explosions and the role of the 3D gas distribution in early SN [58]; calibrations of mixing-length theory (used to model convection in 1D [59–61, 21]); the granulation effects caused by large-scale convective plumes and their impacts on interferometric and photometric observations [62–65].

`euler_multi-quadrants`. The Euler equations model the behavior of inviscid fluids. These simulations specifically describe the evolution of compressible gases in a generalization of the classical Euler quadrants Riemann problem [66]. In these problems, initial discontinuities lead to shocks and rarefactions as the system attempts to correct the instability. This dataset is adapted to include multiple initial discontinuities (Figure 2, top) so that the resulting shocks and rarefactions experience further interactions.

`helmholtz_staircase`. Scattering from periodic structures (Figure 2, middle) occurs in the design of e.g. photonic and phononic crystals, diffraction gratings, antenna arrays, and architecture. These simulations are the linear acoustic scattering of a single point source from an infinite, periodic, corrugated, sound-hard surface, with unit cells comprising two equal-length line segments.

`MHD_64` and `MHD_256`. An essential component of the solar wind, galaxy formation, and of interstellar medium (ISM) dynamics is magnetohydrodynamic (MHD) turbulence (Figure 2, bottom). This dataset consists of isothermal MHD simulations without self-gravity (such as found in the diffuse ISM) initially generated with resolution 256^3 and then downsampled to 64^3 after anti-aliasing with an ideal low-pass filter.

`gray_scott_reaction_diffusion`. Pattern formation is the spontaneous assembly of ordered structures from a seemingly disordered system (Figure 3, top). It occurs across a wide range of biological and chemical systems, often taking place when chemical reactions are coupled to spatial diffusion. For example, reaction–diffusion systems are thought to underpin many of the self-assembly processes present in the early development of organisms [67]. These simulations model the Gray–Scott reaction–diffusion equations [68] describing two chemical species, A and B, whose scalar concentrations vary in space and time.

`planetswe`. The shallow water equations approximate incompressible fluid flows where the horizontal length scale is significantly larger than the vertical as a depth-integrated two-dimensional problem. They have played an important roll in the validation of dynamical cores for atmospheric dynamics as seen in the classical Williamson problems [69]. These simulations can be seen as a refinement of Williamson 7 as they are initialized from the hPa500 level of the ERA5 reanalysis dataset [42] with bathymetry corresponding to the earth's topography and featuring forcings with daily and annual periodicity (Figure 3, middle).

`post_neutron_star_merger`. After the in-spiral and merger of two neutron stars, a hot dense remnant is formed. These events, central to gamma ray bursts and heavy element formation, produce a reddening glow called a *kilonova* [70–77]. Accurate predictions require modeling neutrino interactions, which convert neutrons to protons and vice versa. These simulations model the accretion disk driving the gamma ray burst and the hot neutron-rich wind causing the kilonova (Figure 3, bottom).

`rayleigh_benard`. Rayleigh-Bénard convection [78, 79] is a phenomenon in fluid dynamics encountered in geophysics (mantle convection [80], ocean circulation [81], atmospheric dynamics [82]), in engineering (cooling systems [83], material processing [84]), in astrophysics (interior of stars and planets [85]). It occurs in a horizontal layer of fluid heated from below and cooled from above. This temperature difference creates a density gradient that can lead to the formation of convection currents, where warmer, less dense fluid rises, and cooler, denser fluid sinks (Figure 4, top).

`rayleigh_taylor_instability`. The Rayleigh-Taylor instability [86] is comprised of two fluids of different densities initially at rest. The instability arises from any perturbation that will displace a parcel

of heavier fluid below a parcel of lighter fluid (Figure 4, middle). Pressure forces are then not aligned with density gradients and this generates vorticity, increasing the amplitude of the perturbations. Eventually, these amplitudes become so large that non-linear turbulent mixing develops.

`shear_flow`. Shear flow phenomena [87–89] occurs when layers of fluid move parallel to each other at different velocities, creating a velocity gradient perpendicular to the flow direction (Figure 4, bottom). This can lead to various instabilities and turbulence, which are fundamental to many applications in engineering (e.g., aerodynamics [90]), geophysics (e.g., oceanography [91]), and biomedicine (e.g. biomechanics [92]).

`supernova_explosion_64` and `supernova_explosion_128`. Supernova explosions happen at the end of the lives of some massive stars. These explosions release high energy into the interstellar medium (ISM) and create blastwaves. The blastwaves accumulate in the ISM and form dense, sharp shells, which quickly cool down and can be new star-forming regions (Figure 5, top). These small explosions have a significant impact on the entire galaxy's evolution.

`turbulence_gravity_cooling`. Within the interstellar medium (ISM), turbulence, star formation, supernova explosions, radiation, and other complex physics significantly impact galaxy evolution. This ISM is modeled by a turbulent fluid with gravity. These fluids make dense filaments (Figure 5, middle), leading to the formation of new stars. The timescale and frequency of making new filaments vary with the mass and length of the system.

`turbulent_radiative_layer_2D` and `turbulent_radiative_layer_3D`. In astrophysical environments, cold dense gas clumps move through a surrounding hotter gas, mixing due to turbulence at their interface. This mixing creates an intermediate temperature phase that cools rapidly by radiative cooling, causing the mixed gas to join the cold phase as photons escape and energy is lost (Figure 5, bottom). Simulations and theories show that if cooling is faster (slower) than mixing, the cold clumps will grow (shrink) [93, 94]. These simulations [95] describe the competition between turbulent mixing and radiative cooling at a mixing layer. These simulations are available in 2D and 3D.

`viscoelastic_instability`. In two-dimensional dilute polymer solutions, the flow exhibits four coexistent attractors: the laminar state, a steady arrowhead regime (SAR), a chaotic arrowhead regime (CAR), and a (recently discovered) chaotic regime of elasto-inertial turbulence (EIT). SAR corresponds to a simple traveling wave, while CAR and EIT are visually similar but differ by a weak polymer arrowhead structure across the mid-plane in CAR. These simulations [96] are snapshots of the four attractors and two edge states. Edge states exist on the boundary between two basins of attraction and have a single unstable direction, marking the boundary between different flow behaviors.

3 Benchmark

To showcase the dataset and the associated benchmarking library, we provide a set of simple baselines time-boxed to 12 hours on a single NVIDIA H100 to demonstrate the effectiveness of naive approaches on these challenging problems and motivate the development of more sophisticated approaches. These baselines are trained on the forward problem - predicting the next snapshot of a given simulation from a short history. The models used here are the Fourier Neural Operator [97, FNO], Tucker-Factorized FNO [98, TFNO], U-net [99] and a modernized U-net using ConvNext blocks [100, CNextU-net]. The neural operator models are implemented using `neuralop` [101]. Full training and hyperparameter details are included in the appendix.

Table 3 reports the one-step Variance Scaled Root Mean Squared Error (VRMSE) which is defined as the standard RMSE normalized by the standard deviation of the target field, averaged over all physical fields. We report the top performing run for each model across all hyperparameters. The CNextU-net architecture outperforms the others on 8 of the 17 experiments. However, what is very interesting is that there is a noticeable split between problems which favor spatial domain handling and those which prefer the spectral approach. At the one-step level, 9/17 favor U-net type models while 8 favor spectral models. While in some cases, the results are close, in others, one class of models has a clear advantage. The reason for this is not immediately clear. Boundary conditions would be a natural hypothesis as the boundary condition are handled naively according to model defaults which vary between the U-net and FNO-type models, but there is no clear trend in this direction.

Limitations. These datasets are not without their limitations. They focus largely on uniformly sampled domains at manageable resolutions while many engineering problems require higher resolutions and more complicated meshes than most conventional architectures can feasibly process. As available VRAM increases or more efficient architectures are developed, the current version of the Well may no longer be challenging and new datasets may be needed to push the community forward.

References

[1] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016.

[2] Marsha J. Berger and Randall J. LeVeque. Implicit adaptive mesh refinement for dispersive tsunami propagation, 2024.

[3] Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, et al. Graphcast: Learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794*, 2022.

[4] Lorenz T Biegler, Omar Ghattas, Matthias Heinkenschloss, and Bart van Bloemen Waanders. Large-scale pde-constrained optimization: an introduction. In *Large-scale PDE-constrained optimization*, pages 3–13. Springer, 2003.

[5] Bijan Mohammadi and Olivier Pironneau. Shape optimization in fluid mechanics. *Annu. Rev. Fluid Mech.*, 36:255–279, 2004.

[6] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

[7] Pablo Lemos, Liam Parker, ChangHoon Hahn, Shirley Ho, Michael Eickenberg, Jiamin Hou, Elena Massara, Chirag Modi, Azadeh Moradinezhad Dizgah, Bruno Regaldo-Saint Blancard, and David Spergel. Simbig: Field-level simulation-based inference of galaxy clustering, 2023.

[8] Lawrence C Evans. *Partial differential equations*, volume 19. American Mathematical Society, 2022.

[9] Nestor V. Queipo, Raphael T. Haftka, Wei Shyy, Tushar Goel, Rajkumar Vaidyanathan, and P. Kevin Tucker. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1):1–28, 2005.

[10] A. Forrester, A. Sóbester, and A. Keane. *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley, 2008.

[11] Luning Sun, Han Gao, Shaowu Pan, and Jian-Xun Wang. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data. *Computer Methods in Applied Mechanics and Engineering*, 361:112732, 2020.

[12] Jun Tao and Gang Sun. Application of deep learning based multi-fidelity surrogate model to robust aerodynamic design optimization. *Aerospace Science and Technology*, 92:722–737, 2019.

[13] Ehsan Haghighat, Maziar Raissi, Adrian Moure, Hector Gomez, and Ruben Juanes. A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics. *Computer Methods in Applied Mechanics and Engineering*, 379:113741, 2021.

[14] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447–450, 2018.

[15] Kevin Ryczko, David A Strubbe, and Isaac Tamblyn. Deep learning and density-functional theory. *Physical Review A*, 100(2):022512, 2019.

[16] Kamal Choudhary, Brian DeCost, Chi Chen, Anubhav Jain, Francesca Tavazza, Ryan Cohn, Cheol Woo Park, Alok Choudhary, Ankit Agrawal, Simon JL Billinge, et al. Recent advances and applications of deep learning methods in materials science. *npj Computational Materials*, 8(1):59, 2022.

[17] Ali Siahkoohi, Rudy Morel, Randall Balestriero, Erwan Allys, Grégory Sainton, Taichi Kawamura, and Maarten V de Hoop. Martian time-series unraveled: A multi-scale nested approach with factorial variational autoencoders. *arXiv preprint arXiv:2305.16189*, 2023.

[18] Vignesh Gopakumar, Stanislas Pamela, Lorenzo Zanisi, Zongyi Li, Ander Gray, Daniel Brennand, Nitesh Bhatia, Gregory Stathopoulos, Matt Kusner, Marc Peter Deisenroth, Anima Anandkumar, JOREK Team, and MAST Team. Plasma surrogate modelling using fourier neural operators, 2023.

[19] Shashank Subramanian, Peter Harrington, Kurt Keutzer, Wahid Bhimji, Dmitriy Morozov, Michael W. Mahoney, and Amir Gholami. Towards foundation models for scientific machine learning: Characterizing scaling and transfer behavior. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[20] Bogdan Raonić, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel de Bézenac. Convolutional neural operators for robust and accurate learning of pdes, 2023.

[21] Jared A. Goldberg, Yan-Fei Jiang, and Lars Bildsten. Numerical Simulations of Convective Three-dimensional Red Supergiant Envelopes. *The Astrophysical Journal*, 929(2):156, April 2022.

[22] Yuichi Kodama, Martin Shumway, and Rasko Leinonen. The sequence read archive: explosive growth of sequencing data. *Nucleic acids research*, 40(D1):D54–D56, 2012.

[23] Tony Hey and Anne Trefethen. The data deluge: An e-science perspective. *Grid computing: Making the global infrastructure a reality*, pages 809–824, 2003.

[24] Lucas Thibaut Meyer, Marc Schouler, Robert Alexander Caulk, Alejandro Ribes, and Bruno Raffin. Training deep surrogate models with large scale online learning. In *International Conference on Machine Learning*, pages 24614–24630. PMLR, 2023.

[25] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[26] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

[27] Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Mérouane Debbah, Étienne Goffinet, Daniel Hesslow, Julien Launay, Quentin Malartic, et al. The falcon series of open language models. *arXiv preprint arXiv:2311.16867*, 2023.

[28] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. Scaling rectified flow transformers for high-resolution image synthesis. *arXiv preprint arXiv:2403.03206*, 2024.

[29] Pedro Javier Ortiz Suárez, Benoît Sagot, and Laurent Romary. Asynchronous pipeline for processing huge corpora on medium to low resource infrastructures. In *7th Workshop on the Challenges in the Management of Large Corpora (CMLC-7)*. Leibniz-Institut für Deutsche Sprache, 2019.

[30] Hugo Laurençon, Lucile Saulnier, Léo Tronchon, Stas Bekman, Amanpreet Singh, Anton Lozhkov, Thomas Wang, Siddharth Karamcheti, Alexander Rush, Douwe Kiela, et al. Obelics: An open web-scale filtered dataset of interleaved image-text documents. *Advances in Neural Information Processing Systems*, 36, 2024.

[31] Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*, 2020.

[32] Guilherme Penedo, Quentin Malartic, Daniel Hesslow, Ruxandra Cojocaru, Alessandro Cappelli, Hamza Alobeidli, Baptiste Pannier, Ebtesam Almazrouei, and Julien Launay. The refinedweb dataset for falcon llm: outperforming curated corpora with web data, and web data only. *arXiv preprint arXiv:2306.01116*, 2023.

[33] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems*, 35:25278–25294, 2022.

[34] Jack W Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Sarah Henderson, Roman Ring, Susannah Young, et al. Scaling language models: Methods, analysis & insights from training gopher. *arXiv preprint arXiv:2112.11446*, 2021.

[35] Yuanzhi Li, Sébastien Bubeck, Ronen Eldan, Allie Del Giorno, Suriya Gunasekar, and Yin Tat Lee. Textbooks are all you need ii: phi-1.5 technical report. *arXiv preprint arXiv:2309.05463*, 2023.

[36] Makoto Takamoto, Timothy Praditia, Raphael Leiteritz, Daniel MacKinlay, Francesco Alesiani, Dirk Pflüger, and Mathias Niepert. Pdebench: An extensive benchmark for scientific machine learning. *Advances in Neural Information Processing Systems*, 35:1596–1611, 2022.

[37] Jayesh K Gupta and Johannes Brandstetter. Towards multi-spatiotemporal-scale generalized pde modeling. *arXiv preprint arXiv:2209.15616*, 2022.

[38] Zhongkai Hao, Jiachen Yao, Chang Su, Hang Su, Ziao Wang, Fanzhi Lu, Zeyu Xia, Yichi Zhang, Songming Liu, Lu Lu, et al. Pinnacle: A comprehensive benchmark of physics-informed neural networks for solving pdes. *arXiv preprint arXiv:2306.08827*, 2023.

[39] Georg Kohl, Li-Wei Chen, and Nils Thuerey. Benchmarking autoregressive conditional diffusion models for turbulent flow simulation. *arXiv*, 2023.

[40] Florent Bonnet, Jocelyn Mazari, Paola Cinnella, and Patrick Gallinari. Airfrans: High fidelity computational fluid dynamics dataset for approximating reynolds-averaged navier–stokes solutions. *Advances in Neural Information Processing Systems*, 35:23463–23478, 2022.

[41] Artur Toshev, Gianluca Galletti, Fabian Fritz, Stefan Adami, and Nikolaus Adams. Lagrangebench: A lagrangian fluid mechanics benchmarking suite. *Advances in Neural Information Processing Systems*, 36, 2024.

[42] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

[43] Sungduk Yu, Walter Hannah, Liran Peng, Jerry Lin, Mohamed Aziz Bhouri, Ritwik Gupta, Björn Lütjens, Justus Christopher Will, Gunnar Behrens, Julius Busecke, Nora Loose, Charles I Stern, Tom Beucler, Bryce Harrop, Benjamin R Hillman, Andrea Jenney, Savannah Ferretti, Nana Liu, Anima Anandkumar, Noah D Brenowitz, Veronika Eyring, Nicholas Geneva, Pierre Gentine, Stephan Mandt, Jaideep Pathak, Akshay Subramaniam, Carl Vondrick, Rose Yu, Laure Zanna, Tian Zheng, Ryan Abernathey, Fiaz Ahmed, David C Bader, Pierre Baldi, Elizabeth Barnes, Christopher Bretherton, Peter Caldwell, Wayne Chuang, Yilun Han, Yu Huang, Fernando Iglesias-Suarez, Sanket Jantre, Karthik Kashinath, Marat Khairoutdinov, Thorsten Kurth, Nicholas Lutsko, Po-Lun Ma, Griffin Mooers, J. David Neelin, David Randall, Sara Shamekh, Mark A Taylor, Nathan Urban, Janni Yuval, Guang Zhang, and Michael Pritchard. Climsim: A large multi-scale dataset for hybrid physics-ml climate emulation, 2024.

[44] Steeven Janny, Aurélien Benetteau, Madiha Nadri, Julie Digne, Nicolas Thome, and Christian Wolf. Eagle: Large-scale learning of turbulent fluid dynamics with mesh transformers. In *International Conference on Learning Representations (ICLR)*, 2023.

[45] Sheikh Md Shakeel Hassan, Arthur Feeney, Akash Dhruv, Jihoon Kim, Youngjoon Suh, Jaiyoung Ryu, Yoonjin Won, and Aparna Chandramowlishwaran. BubbleML: A multi-physics dataset and benchmarks for machine learning. In *Advances in Neural Information Processing Systems*, 2023.

[46] Milad LEYLI ABADI, Antoine Marot, Jérôme Picault, David Danan, Mouadh Yagoubi, Benjamin Donnot, Seif Attoui, Pavel Dimitrov, Asma Farjallah, and Clement Etienam. Lips-learning industrial physical simulation benchmark suite. *Advances in Neural Information Processing Systems*, 35:28095–28109, 2022.

[47] Wai Tong Chung, Bassem Akoush, Pushan Sharma, Alex Tamkin, Ki Sung Jung, Jacqueline H. Chen, Jack Guo, Davy Brouzet, Mohsen Talei, Bruno Savard, Alexei Y. Poludnenko, and Matthias Ihme. Turbulence in focus: Benchmarking scaling behavior of 3D volumetric super-resolution with BLASTNet 2.0 data. *Advances in Neural Information Processing Systems (NeurIPS)*, 36, 2023.

[48] Yi Li, Eric Perlman, Minping Wan, Yunke Yang, Charles Meneveau, Randal Burns, Shiyi Chen, Alexander Szalay, and Gregory Eyink. A public turbulence database cluster and applications to study lagrangian evolution of velocity increments in turbulence. *Journal of Turbulence*, (9):N31, 2008.

[49] Michael McCabe, Bruno Régaldo-Saint Blancard, Liam Holden Parker, Ruben Ohana, Miles Cranmer, Alberto Bietti, Michael Eickenberg, Siavash Golkar, Geraud Krawezik, Francois Lanusse, et al. Multiple physics pretraining for physical surrogate models. *arXiv preprint arXiv:2310.02994*, 2023.

[50] Liu Yang, Siting Liu, Tingwei Meng, and Stanley J. Osher. In-context operator learning with data prompts for differential equation problems. *Proceedings of the National Academy of Sciences*, 120(39), September 2023.

[51] Md Ashiqur Rahman, Robert Joseph George, Mogab Elleithy, Daniel Leibovici, Zongyi Li, Boris Bonev, Colin White, Julius Berner, Raymond A Yeh, Jean Kossaifi, et al. Pretraining codomain attention neural operators for solving multiphysics pdes. *arXiv preprint arXiv:2403.12553*, 2024.

[52] Jingmin Sun, Yuxuan Liu, Zecheng Zhang, and Hayden Schaeffer. Towards a foundation model for partial differential equation: Multi-operator learning and extrapolation. *arXiv preprint arXiv:2404.12355*, 2024.

[53] Junhong Shen, Tanya Marwah, and Ameet Talwalkar. Ups: Towards foundation models for pde solving via cross-modal adaptation. *arXiv preprint arXiv:2403.07187*, 2024.

[54] Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for pdes, 2024.

[55] Andrew Collette, Thomas Kluyver, Thomas A Caswell, James Tocknell, Jerome Kieffer, Aleksandar Jelenak, Anthony Scopatz, Darren Dale, Thomas VINCENT, Matt Einhorn, et al. h5py/h5py: 3.8. 0-aarch64-wheels. *Zenodo*, 2023.

[56] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019.

[57] Charles R Harris, K Jarrod Millman, Stéfan J Van Der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J Smith, et al. Array programming with numpy. *Nature*, 585(7825):357–362, 2020.

[58] Jared A. Goldberg, Yan-Fei Jiang, and Lars Bildsten. Shock Breakout in Three-dimensional Red Supergiant Envelopes. *The Astrophysical Journal*, 933(2):164, July 2022.

[59] E. Böhm-Vitense. Über die Wasserstoffkonvektionszone in Sternen verschiedener Effektivtemperaturen und Leuchtkräfte. Mit 5 Textabbildungen. *Zeitschrift für Astrophysik*, 46:108, January 1958.

[60] J. P. Cox and R. T. Giuli. *Principles of stellar structure*. 1968.

[61] Meridith Joyce and Jamie Tayar. A Review of the Mixing Length Theory of Convection in 1D Stellar Modeling. *Galaxies*, 11(3):75, June 2023.

[62] A. Chiavassa, R. Norris, M. Montargès, R. Ligi, L. Fossati, L. Bigot, F. Baron, P. Kervella, J. D. Monnier, D. Mourard, N. Nardetto, G. Perrin, G. H. Schaefer, T. A. ten Brummelaar, Z. Magic, R. Collet, and M. Asplund. Asymmetries on red giant branch surfaces from CHARA/MIRC optical interferometry. *Astronomy & Astrophysics*, 600:L2, April 2017.

[63] A. Chiavassa, B. Freytag, and M. Schultheis. Heading Gaia to measure atmospheric dynamics in AGB stars. *Astronomy & Astrophysics*, 617:L1, September 2018.

[64] A. Chiavassa, K. Kravchenko, F. Millour, G. Schaefer, M. Schultheis, B. Freytag, O. Creevey, V. Hocdé, F. Morand, R. Ligi, S. Kraus, J. D. Monnier, D. Mourard, N. Nardetto, N. Anugu, J. B. Le Bouquin, C. L. Davies, J. Ennis, T. Gardner, A. Labdon, C. Lanthermann, B. R. Setterholm, and T. ten Brummelaar. Optical interferometry and Gaia measurement uncertainties reveal the physics of asymptotic giant branch stars. *Astronomy & Astrophysics*, 640:A23, August 2020.

[65] Andrea Chiavassa, Kateryna Kravchenko, and Jared A. Goldberg. Signatures of convection in the atmospheres of cool evolved stars. *Living Reviews in Computational Astrophysics*, 10(1):2, March 2024.

[66] Peter D. Lax and Xu-Dong Liu. Solution of two-dimensional riemann problems of gas dynamics by positive schemes. *SIAM Journal on Scientific Computing*, 19(2):319–340, 1998.

[67] Alan M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society B*, 237(641):37–72, 1952.

[68] P. Gray and S. K. Scott. Autocatalytic reactions in the isothermal, continuous stirred tank reactor: Oscillations and instabilities in the system $A+2B \rightarrow 3B$; $B \rightarrow C$. *Chemical Engineering Science*, 39(6):1087–1097, 1984.

[69] David L. Williamson, John B. Drake, James J. Hack, Rüdiger Jakob, and Paul N. Swarztrauber. A standard test set for numerical approximations to the shallow water equations in spherical geometry. *Journal of Computational Physics*, 102(1):211–224, 1992.

[70] J.M. Lattimer and D.N. Schramm. Black-Hole-Neutron-Star Collisions. *Astrophysical Journal Letters*, 192:L145, 1974.

[71] J. M. Lattimer and D. N. Schramm. The tidal disruption of neutron stars by black holes in close binaries. *The Astrophysical Journal*, 210:549–567, December 1976.

[72] Li-Xin Li and Bohdan Paczyński. Transient Events from Neutron Star Mergers. *Astrophysical Journal Letters*, 507(1):L59–L62, 1998.

[73] B. P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Physical Review Letters*, 119(16):30–33, 2017.

[74] B. P. Abbott et al. Multi-messenger Observations of a Binary Neutron Star Merger. *Astrophysical Journal Letters*, 848:L12, 2017.

[75] K. D. Alexander et al. The electromagnetic counterpart of the binary neutron star merger ligo/virgo gw170817. vi. radio constraints on a relativistic jet and predictions for late-time emission from the kilonova ejecta. *The Astrophysical Journal Letters*, 848(2):L21, oct 2017.

[76] P. S. Cowperthwaite et al. The electromagnetic counterpart of the binary neutron star merger LIGO/VIRGO GW170817. II. UV, optical, and near-ir light curves and comparison to kilonova models. *Astrophysical Journal Letters*, 848:L17, 2017.

[77] V. A. Villar et al. The combined ultraviolet, optical, and near-infrared light curves of the kilonova associated with the binary neutron star merger gw170817: unified data set, analytic models, and physical implications. *Astrophysical Journal Letters*, 851(1):L21, 2017.

[78] Lord Rayleigh. Lix. on convection currents in a horizontal layer of fluid, when the higher temperature is on the under side. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 32(192):529–546, 1916.

[79] Baole Wen, David Goluskin, and Charles R Doering. Steady rayleigh–bénard convection between no-slip boundaries. *Journal of Fluid Mechanics*, 933:R4, 2022.

[80] Gerald Schubert, Donald Lawson Turcotte, and Peter Olson. *Mantle convection in the Earth and planets*. Cambridge University Press, 2001.

[81] Gerold Siedler, John Gould, and John A Church. *Ocean circulation and climate: observing and modelling the global ocean*. Elsevier, 2001.

[82] James R Holton and Gregory J Hakim. *An introduction to dynamic meteorology*, volume 88. Academic press, 2013.

[83] Sadik Kakaç, Hafit Yüncü, and Kunio Hijikata. *Cooling of electronic systems*, volume 258. Springer Science & Business Media, 2012.

[84] David R Poirier and G Geiger. *Transport phenomena in materials processing*. Springer, 2016.

[85] Carl J Hansen, Steven D Kawaler, and Virginia Trimble. *Stellar interiors: physical principles, structure, and evolution*. Springer Science & Business Media, 2012.

[86] Geoffrey Ingram Taylor. The instability of liquid surfaces when accelerated in a direction perpendicular to their planes. i. *Proceedings of the Royal Society of London. Series A, Mathematical and physical sciences*, 201(1065):192–196, 1950.

[87] Pijush K Kundu, Ira M Cohen, and David R Dowling. *Fluid mechanics*. Academic press, 2015.

[88] Ting Wu and Guowei He. Space-time energy spectra in turbulent shear flows. *Physical Review Fluids*, 6(10):100504, 2021.

[89] Prathmesh Vinze and Sebastien Michelin. Self-organization of autophoretic suspensions in confined shear flows. *Physical Review Fluids*, 9(1):014202, 2024.

[90] Arthur Rizzi. Separated and vortical flow in aircraft aerodynamics: a cfd perspective. *The Aeronautical Journal*, 127(1313):1065–1103, 2023.

[91] William D Smyth and James N Moum. Ocean mixing by kelvin-helmholtz instability. *Oceanography*, 25(2):140–149, 2012.

[92] Romana Perinajová, Joe F Juffermans, Jonhatan Lorenzo Mercado, Jean-Paul Aben, Leon Ledoux, Jos JM Westenberg, Hildo J Lamb, and Saša Kenjereš. Assessment of turbulent blood flow and wall shear stress in aortic coarctation using image-based simulations. *Biomedical engineering online*, 20(1):84, 2021.

[93] Max Gronke and S. Peng Oh. The growth and entrainment of cold gas in a hot wind. *Monthly Notices of the Royal Astronomical Society: Letters*, 480(1):L111–L115, October 2018.

[94] Matthew W. Abruzzo, Drummond B. Fielding, and Greg L. Bryan. Taming the TuRMoiL: The Temperature Dependence of Turbulence in Cloud–Wind Interactions. *The Astrophysical Journal*, 966(2):181, May 2024.

[95] Drummond B. Fielding, Eve C. Ostriker, Greg L. Bryan, and Adam S. Jermyn. Multiphase Gas and the Fractal Nature of Radiative Turbulent Mixing Layers. *The Astrophysical Journal Letters*, 894(2):L24, May 2020.

[96] Miguel Beneitez, Jacob Page, Yves Dubief, and Rich R Kerswell. Multistability of elasto-inertial two-dimensional channel flow. *Journal of Fluid Mechanics*, 981:A30, 2024.

[97] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[98] Jean Kossaifi, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, and Anima Anandkumar. Multi-grid tensorized fourier neural operator for high resolution PDEs, 2023.

[99] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[100] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s, 2022.

[101] Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

[102] Mubashara Akhtar, Omar Benjelloun, Costanza Conforti, Pieter Gijsbers, Joan Giner-Miguelez, Nitisha Jain, Michael Kuchnik, Quentin Lhoest, Pierre Marcenac, Manil Maskey, Peter Mattson, Luis Oala, Pierre Ruyssen, Rajat Shinde, Elena Simperl, Goeffry Thomas, Slava Tykhonov, Joaquin Vanschoren, Jos van der Velde, Steffen Vogler, and Carole-Jean Wu. Croissant: A metadata format for ml-ready datasets. DEEM '24, page 1–6, New York, NY, USA, 2024. Association for Computing Machinery.

[103] Clawpack Development Team. Clawpack software. http://www.clawpack.org, 2021.

[104] James M. Stone, Kengo Tomida, Christopher J. White, and Kyle G. Felker. The Athena++ Adaptive Mesh Refinement Framework: Design and Magnetohydrodynamic Solvers. *The Astrophysical Journal Supplement Series*, 249(1):4, July 2020.

[105] Keaton J Burns, Geoffrey M Vasil, Jeffrey S Oishi, Daniel Lecoanet, and Benjamin P Brown. Dedalus: A flexible framework for numerical simulations with spectral methods. *Physical Review Research*, 2(2):023068, 2020.

[106] Jonah M. Miller, Ben. R. Ryan, and Joshua C. Dolence. νbhlight: Radiation grmhd for neutrino-driven accretion flows. *The Astrophysical Journal Supplement Series*, 241(2):30, apr 2019.

[107] Romain Watteaux. *Détection des grandes structures turbulentes dans les couches de mélange de type Rayleigh-Taylor en vue de la validation de modèles statistiques turbulents bi-structure*. Theses, École normale supérieure de Cachan - ENS Cachan, September 2011.

[108] Masaki Iwasawa, Ataru Tanikawa, Natsuki Hosono, Keigo Nitadori, Takayuki Muranushi, and Junichiro Makino. Implementation and performance of FDPS: a framework for developing parallel particle simulation codes. *Publications of the Astronomical Society of Japan*, 68(4):54, August 2016.

[109] Barath Ezhilan, Michael J Shelley, and David Saintillan. Instabilities and nonlinear dynamics of concentrated active suspensions. *Physics of Fluids*, 25(7), 2013.

[110] Tong Gao, Meredith D Betterton, An-Sheng Jhang, and Michael J Shelley. Analytical structure, dynamics, and coarse graining of a kinetic model of an active fluid. *Physical Review Fluids*, 2(9):093302, 2017.

[111] David Saintillan and Michael J Shelley. Theory of active suspensions. *Complex Fluids in Biological Systems: Experiment, Theory, and Computation*, pages 319–355, 2015.

[112] Scott Weady, David B Stein, and Michael J Shelley. Thermodynamically consistent coarse-graining of polar active fluids. *Physical Review Fluids*, 7(6):063301, 2022.

[113] Suryanarayana Maddu, Scott Weady, and Michael J Shelley. Learning fast, accurate, and stable closures of a kinetic theory of an active fluid. *Journal of Computational Physics*, page 112869, 2024.

[114] Yan-Fei Jiang. An Implicit Finite Volume Scheme to Solve the Time-dependent Radiation Transport Equation Based on Discrete Ordinates. *The Astrophysical Journal Supplement Series*, 253(2):49, April 2021.

[115] Carlos A. Iglesias and Forrest J. Rogers. Updated Opal Opacities. *The Astrophysical Journal*, 464:943, June 1996.

[116] Yan-Fei Jiang, Matteo Cantiello, Lars Bildsten, Eliot Quataert, and Omer Blaes. Local Radiation Hydrodynamic Simulations of Massive Star Envelopes at the Iron Opacity Peak. *The Astrophysical Journal*, 813(1):74, November 2015.

[117] Yan-Fei Jiang, Matteo Cantiello, Lars Bildsten, Eliot Quataert, Omer Blaes, and James Stone. Outbursts of luminous blue variable stars from variations in the helium opacity. *Nature*, 561(7724):498–501, September 2018.

[118] K.T.Mandli, A.J. Ahmadia, M.J. Berger, D.A. Calhoun, D.L. George, Y. Hadjimichael, D.I. Ketcheson, G.I Lemoine, and R.J. LeVeque. Clawpack: building an open source ecosystem for solving hyperbolic PDEs. *PeerJ Computer Science 2:e68*, 2016. https://doi.org/10.7717/peerj-cs.68.

[119] Fruzsina J Agocs and Alex H Barnett. Trapped acoustic waves and raindrops: high-order accurate integral equation method for localized excitation of a periodic staircase. *arXiv preprint arXiv:2310.12486*, 2023.

[120] Armin Lechleiter and Ruming Zhang. A convergent numerical scheme for scattering of aperiodic waves from periodic surfaces based on the Floquet–Bloch transform. *SIAM J. Numer. Anal.*, 55(2):713–736, 2017.

[121] Ruming Zhang. A high order numerical method for scattering from locally perturbed periodic surfaces. *SIAM J. Sci. Comput.*, 40(4):A2286–A2314, 2018.

[122] Graeme Fairweather and Andreas Karageorghis. The method of fundamental solutions for elliptic boundary value problems. *Advances in Computational Mathematics*, 9:69–95, 1998.

[123] Alexander HD Cheng and Yongxing Hong. An overview of the method of fundamental solutions—solvability, uniqueness, convergence, and stability. *Engineering Analysis with Boundary Elements*, 120:118–152, 2020.

[124] Alex H Barnett and Timo Betcke. Stability and convergence of the method of fundamental solutions for helmholtz problems on analytic domains. *Journal of Computational Physics*, 227(14):7003–7026, 2008.

[125] Carlos JS Alves and Svilen S Valtchev. Numerical comparison of two meshfree methods for acoustic wave scattering. *Engineering Analysis with Boundary Elements*, 29(4):371–382, 2005.

[126] RF Millar. The Rayleigh hypothesis and a related least-squares solution to scattering problems for periodic surfaces and other scatterers. *Radio Science*, 8(8-9):785–796, 1973.

[127] William C Meecham. On the use of the Kirchhoff approximation for the solution of reflection problems. *Journal of Rational Mechanics and Analysis*, 5(2):323–334, 1956.

[128] J B Keller. A geometrical theory of diffraction. In L M Graves, editor, *Calculus of Variations and its Applications*, number 8 in Proceedings of Symposia in Applied Mathematics, pages 27–52, New York, 1958. McGraw-Hill.

[129] Joseph B Keller. Geometrical theory of diffraction. *Josa*, 52(2):116–130, 1962.

[130] Edward L Richards, HC Song, and WS Hodgkiss. Acoustic scattering comparison of Kirchhoff approximation to Rayleigh-fourier method for sinusoidal surface waves at low grazing angles. *The Journal of the Acoustical Society of America*, 144(3):1269–1278, 2018.

[131] Jungyeon Cho and A. Lazarian. Compressible Magnetohydrodynamic Turbulence: mode coupling, scaling relations, anisotropy, new regime and astrophysical implications. *MNRAS*, 345:325–339, jan 2003.

[132] G. Kowal, A. Lazarian, and A. Beresnyak. Density fluctuations in mhd turbulence: Spectral properties and intermittency. *The Astrophysical Journal*, 658:423–445, 2007.

[133] BICEP2 Collaboration. BICEP2 I: Detection Of B-mode Polarization at Degree Angular Scales. *The Astrophysical Journal Letters*, 785:L1, April 2014.

[134] Blakesley Burkhart, A. Lazarian, and B. M. Gaensler. Properties of interstellar turbulence in the galactic halo from structure function analysis of rotation measure maps. *The Astrophysical Journal*, 785(1):123, 2014.

[135] S. K. N. Portillo and D. P. Finkbeiner. Applying the Wasserstein Distance to Astronomical Catalogs. *The Astrophysical Journal*, 862:119, August 2018.

[136] Alex S. Hill, Robert A. Benjamin, Grzegorz Kowal, Ronald J. Reynolds, L. Matthew Haffner, and Alex Lazarian. The turbulent warm ionized medium: Emission measure distribution and mhd simulations. *The Astrophysical Journal*, 686:363–378, 2008.

[137] Christopher F. McKee and Eve C. Ostriker. Theory of star formation. *Annual Review of Astronomy and Astrophysics*, 45:565–687, 2007.

[138] Blakesley Burkhart, Diego Falceta-Goncalves, Grzegorz Kowal, and A. Lazarian. Density studies of mhd interstellar turbulence: Statistical moments, correlations and bispectrum. *The Astrophysical Journal*, 693:250–266, 2009.

[139] B. Burkhart, S. M. Appel, S. Bialy, J. Cho, A. J. Christensen, D. Collins, C. Federrath, D. B. Fielding, D. Finkbeiner, A. S. Hill, J. C. Ibáñez-Mejía, M. R. Krumholz, A. Lazarian, M. Li, P. Mocz, M. M. Mac Low, J. Naiman, S. K. N. Portillo, B. Shane, Z. Slepian, and Y. Yuan. The Catalogue for Astrophysical Turbulence Simulations (CATS). *The Astrophysical Journal*, 905(1):14, December 2020.

[140] Robert P. Munafo. Reaction-diffusion by the Gray-Scott model: Pearson's parametrization. https://www.mrob.com/pub/comp/xmorphia/.

[141] B.J. Walker, A.K. Townsend, A.K. Chudasama, and A. L. Krause. VisualPDE: Rapid interactive simulations of partial differential equations, 2023.

[142] T. A Driscoll, N. Hale, and L. N. Trefethen. *Chebfun Guide*. Pafnuty Publications, 2014.

[143] Aly-Khan Kassam and Lloyd N. Trefethen. Fourth-Order Time-Stepping for Stiff PDEs. *SIAM J. Sci. Comput.*, 26(4):1214–1233, 2005.

[144] Uri M. Ascher, Steven J. Ruuth, and Raymond J. Spiteri. Implicit-explicit runge-kutta methods for time-dependent partial differential equations. *Applied Numerical Mathematics*, 25(2):151–167, 1997. Special Issue on Time Integration.

[145] Michael McCabe, Peter Harrington, Shashank Subramanian, and Jed Brown. Towards stability of autoregressive neural operators. *Transactions on Machine Learning Research*, 2023.

[146] Jonah M. Miller, Benjamin R. Ryan, Joshua C. Dolence, Adam Burrows, Christopher J. Fontes, Christopher L. Fryer, Oleg Korobkin, Jonas Lippuner, Matthew R. Mumpower, and Ryan T. Wollaeger. Full transport model of GW170817-like disk produces a blue kilonova. *Physical Review D*, 100(2):23008, 2019.

[147] Jonah M Miller, Trevor M Sprouse, Christopher L Fryer, Benjamin R Ryan, Joshua C Dolence, Matthew R Mumpower, and Rebecca Surman. Full transport general relativistic radiation magnetohydrodynamics for nucleosynthesis in collapsars. *The Astrophysical Journal*, 902(1):66, 2020.

[148] Sanjana Curtis, Jonah M Miller, Carla Fröhlich, Trevor Sprouse, Nicole Lloyd-Ronning, and Matthew Mumpower. Nucleosynthesis in outflows from black hole–neutron star merger disks with full gr (ν) rmhd. *The Astrophysical Journal Letters*, 945(1):L13, 2023.

[149] Kelsey A Lund, Gail C McLaughlin, Jonah M Miller, and Matthew R Mumpower. Magnetic field strength effects on nucleosynthesis from neutron star merger outflows. *The Astrophysical Journal*, 964(2):111, 2024.

[150] E.P. Velikhov. Stability of an ideally conducting liquid flowing between rotating cylinders in a magnetic field. *Zhur. Eksptl'. i Teoret. Fiz.*, 36(5), 5 1959.

[151] S. A. Balbus and J. F. Hawley. A powerful local shear instability in weakly magnetized disks. I - Linear analysis. II - Nonlinear evolution. *The Astrophysical Journal*, 376:214–233, July 1991.

[152] N. I. Shakura and R. A. Sunyaev. Black holes in binary systems. Observational appearance. *Proceedings of the American Astronomical Society*, 24:337–355, January 1973.

[153] Charles F. Gammie, Jonathan C. McKinney, and Gabor Toth. Harm: A numerical scheme for general relativistic magnetohydrodynamics. *The Astrophysical Journal*, 589(1):444–457, May 2003.

[154] Joshua C. Dolence, Charles F. Gammie, Monika Moś cibrodzka, and Po Kin Leung. grmonty: A MONTE CARLO CODE FOR RELATIVISTIC RADIATIVE TRANSPORT. *The Astrophysical Journal Supplement Series*, 184(2):387–397, oct 2009.

[155] B. R. Ryan, J. C. Dolence, and C. F. Gammie. bhlight: General relativistic radiation magnetohydrodynamics with monte carlo transport. *The Astrophysical Journal*, 807(1):31, jun 2015.

[156] Oliver Porth et al. The event horizon general relativistic magnetohydrodynamic code comparison project. *The Astrophysical Journal Supplement Series*, 243(2):26, aug 2019.

[157] Jonathan C. McKinney and Charles F. Gammie. A Measurement of the Electromagnetic Luminosity of a Kerr Black Hole. *The Astrophysical Journal*, 611(2):977–995, August 2004.

[158] Alexander Tchekhovskoy, Jonathan C. McKinney, and Ramesh Narayan. WHAM: a WENO-based general relativistic numerical scheme - I. Hydrodynamics. *Monthly Notices of the Royal Astronomical Society*, 379(2):469–497, August 2007.

[159] A. Mignone and Jonathan C. McKinney. Equation of state in relativistic magnetohydrodynamics: variable versus constant adiabatic index. *Monthly Notices of the Royal Astronomical Society*, 378(3):1118–1130, July 2007.

[160] S. M. Ressler, A. Tchekhovskoy, E. Quataert, M. Chandra, and C. F. Gammie. Electron thermodynamics in GRMHD simulations of low-luminosity black hole accretion. *Monthly Notices of the Royal Astronomical Society*, 454(2):1848–1870, 10 2015.

[161] Roy P. Kerr. Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics. *Physical Review Letters*, 11(5):237–238, September 1963.

[162] L. G. Fishbone and V. Moncrief. Relativistic fluid disks in orbit around Kerr black holes. *The Astrophysical Journal*, 207:962–976, August 1976.

[163] A. W. Steiner, M. Hempel, and T. Fischer. Core-collapse supernova equations of state based on neutron star observations. *Astrophysical Journal*, 774(1), 2013.

[164] M. Aaron Skinner, Joshua C. Dolence, Adam Burrows, David Radice, and David Vartanyan. Fornax: A flexible code for multiphysics astrophysical simulations. *The Astrophysical Journal Supplement Series*, 241(1):7, feb 2019.

[165] A. Burrows, S. Reddy, and T. A. Thompson. Neutrino opacities in nuclear matter. *Nuclear Physics A*, 777:356–394, October 2006.

[166] Evan O'Connor and Christian D Ott. A new open-source code for spherically symmetric stellar collapse to neutron stars and black holes. *Classical and Quantum Gravity*, 27(11):114103, 2010.

[167] Erik Lindborg. Reynolds-number scaling and convergence time scale in two-dimensional rayleigh–bénard convection. *Journal of Fluid Mechanics*, 973:A9, 2023.

[168] Guy Dimonte, D. L Youngs, A Dimits, S Weber, M Marinak, S Wunsch, C Garasi, A Robinson, M. J Andrews, P Ramaprabhu, A. C Calder, B Fryxell, J Biello, L Dursi, P MacNeice, K Olson, P Ricker, R Rosner, F Timmes, H Tufo, Y.-N Young, and M Zingale. A comparative study of the turbulent Rayleigh–Taylor instability using high-resolution three-dimensional numerical simulations: The Alpha-Group collaboration. *Physics of fluids (1994)*, 16(5):1668–1693, 2004.

[169] P. Ramaprabhu, Guy Dimonte, P. Woodward, C. Fryer, G. Rockefeller, K. Muthuraman, P.-H. Lin, and J. Jayaraj. The late-time dynamics of the single-mode Rayleigh-Taylor instability. *Physics of Fluids*, 24(7):074107, 07 2012.

[170] Malcolm J Andrews and Stuart B Dalziel. Small Atwood number Rayleigh-Taylor experiments. *Philosophical transactions of the Royal Society of London. Series A: Mathematical, physical, and engineering sciences*, 368(1916):1663–1679, 2010.

[171] R. Peyret and T.D. Taylor. Computational methods for fluid flow. *The International journal of heat and fluid flow*, 4(3):182–182, 1983.

[172] A Harten. On a class of high resolution total-variation-stable finite-difference schemes. *SIAM journal on numerical analysis*, 21(1):1–23, 1984.

[173] A Harten, B Engquist, S Osher, and Chakravarthy. Uniformly high order accurate essentially non-oscillatory schemes .3. (reprinted from journal of computational physics, vol 71, pg 231, 1987). *Journal of computational physics*, 131(1):3–47, 1997.

[174] S. Gottlieb, C.-W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM review*, 43(1):89–112, 2001.

[175] William L Briggs, Van Emden Henson, and Steve F McCormick. *A multigrid tutorial, second edition*. Society for Industrial and Applied Mathematics, Philadelphia, Pa, 2nd ed. edition, 2000.

[176] William H. Press and Saul A. Teukolsky. Multigrid methods for boundary value problems. i. *Computers in physics*, 5(5):514, 1991.

[177] William H. Press and Saul A. Teukolsky. Multigrid methods for boundary value problems. ii. *Computers in physics*, 5(6):626, 1991.

[178] S.B. Pope. Turbulent flows, cambridge university press, cambridge, u.k., 2000, 771 pp. *Combustion and flame*, 125(4):1361–1362, 2000.

[179] Olivier Soulard and Jérôme Griffond. Inertial-range anisotropy in Rayleigh-Taylor turbulence. *Physics of fluids (1994)*, 24(2):025101–025101–25, 2012.

[180] William Cabot and Andrew Cook. Reynolds number effects on Rayleigh–Taylor instability with possible implications for type Ia supernovae. *Nature Physics*, 2:562–568, 07 2006.

[181] Michael Chertkov. Phenomenology of Rayleigh-Taylor Turbulence. *Physical review letters*, 91(11):1150011–1150014, 2003.

[182] Thomas Berlok and Christoph Pfrommer. On the kelvin–helmholtz instability with smooth initial conditions–linear theory and simulations. *Monthly Notices of the Royal Astronomical Society*, 485(1):908–923, 2019.

[183] L. I. Sedov. *Similarity and Dimensional Methods in Mechanics*. 1959.

[184] Takayuki R. Saitoh, Hiroshi Daisaka, Eiichiro Kokubo, Junichiro Makino, Takashi Okamoto, Kohji Tomisaka, Keiichi Wada, and Naoki Yoshida. Toward First-Principle Simulations of Galaxy Formation: I. How Should We Choose Star-Formation Criteria in High-Resolution Simulations of Disk Galaxies? *Publications of the Astronomical Society of Japan*, 60(4):667–681, August 2008.

[185] Keiya Hirashima, Kana Moriwaki, Michiko S. Fujii, Yutaka Hirai, Takayuki R. Saitoh, and Junichiro Makino. 3D-Spatiotemporal forecasting the expansion of supernova shells using deep learning towards high-resolution galaxy simulations. *Monthly Notices of the Royal Astronomical Society*, 526(3):4054–4066, December 2023.

[186] Takayuki R. Saitoh and Junichiro Makino. A Density-independent Formulation of Smoothed Particle Hydrodynamics. *The Astrophysical Journal*, 768(1):44, May 2013.

[187] Takayuki R. Saitoh and Junichiro Makino. A Necessary Condition for Individual Time Steps in SPH Simulations. *The Astrophysical Journal Letters*, 697(2):L99–L102, June 2009.

[188] Keiya Hirashima, Kana Moriwaki, Michiko S. Fujii, Yutaka Hirai, Takayuki R. Saitoh, Junichiro Makino, and Shirley Ho. Surrogate Modeling for Computationally Expensive Simulations of Supernovae in High-Resolution Galaxy Simulations. *arXiv e-prints*, page arXiv:2311.08460, November 2023.

[189] S. Portegies Zwart, S. L. W. McMillan, E. van Elteren, I. Pelupessy, and N. de Vries. Multi-physics simulations using a hierarchical interchangeable software interface. *Computer Physics Communications*, 184(3):456–468, March 2013.

[190] F. I. Pelupessy, A. van Elteren, N. de Vries, S. L. W. McMillan, N. Drost, and S. F. Portegies Zwart. The Astrophysical Multipurpose Software Environment. *Astronomy & Astrophysics*, 557:A84, September 2013.

[191] Simon Portegies Zwart and Steve McMillan. *Astrophysical Recipes*. 2514-3433. IOP Publishing, 2018.

[192] G. J. Ferland, K. T. Korista, D. A. Verner, J. W. Ferguson, J. B. Kingdon, and E. M. Verner. CLOUDY 90: Numerical Simulation of Plasmas and Their Spectra. *Publications of the Astronomical Society of the Pacific*, 110(749):761–778, July 1998.

[193] G. J. Ferland, R. L. Porter, P. A. M. van Hoof, R. J. R. Williams, N. P. Abel, M. L. Lykins, G. Shaw, W. J. Henney, and P. C. Stancil. The 2013 Release of Cloudy. *Revista mexicana de astronomía y astrofísica*, 49:137–163, April 2013.

[194] G. J. Ferland, M. Chatzikos, F. Guzmán, M. L. Lykins, P. A. M. van Hoof, R. J. R. Williams, N. P. Abel, N. R. Badnell, F. P. Keenan, R. L. Porter, and P. C. Stancil. The 2017 Release Cloudy. *Revista mexicana de astronomía y astrofísica*, 53:385–438, October 2017.

[195] Nadim Saad, Gaurav Gupta, Shima Alizadeh, and Danielle Maddix Robinson. Guiding continuous operator learning through physics-based boundary constraints. In *ICLR 2023*, 2023.

[196] N. Sukumar and Ankit Srivastava. Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks. *Computer Methods in Applied Mechanics and Engineering*, 389:114333, February 2022.

[197] Antonio Alguacil, Wagner Gonçalves Pinto, Michael Bauerheim, Marc C. Jacob, and Stéphane Moreau. Effects of boundary conditions in fully convolutional networks for learning spatio-temporal dynamics, 2021.

[198] Derek Hansen, Danielle Maddix Robinson, Shima Alizadeh, Gaurav Gupta, and Michael Mahoney. Learning physical models that can respect conservation laws. In *ICML 2023*, 2023.

[199] Nick McGreivy and Ammar Hakim. Invariant preservation in machine learned pde solvers via error correction, 2023.

[200] Kim Stachenfeld, Drummond Buschman Fielding, Dmitrii Kochkov, Miles Cranmer, Tobias Pfaff, Jonathan Godwin, Can Cui, Shirley Ho, Peter Battaglia, and Alvaro Sanchez-Gonzalez. Learned simulators for turbulence. In *International Conference on Learning Representations*, 2022.

[201] Phillip Lippe, Bastiaan S. Veeling, Paris Perdikaris, Richard E. Turner, and Johannes Brandstetter. Pde-refiner: Achieving accurate long rollouts with neural pde solvers, 2023.

[202] Jean Kossaifi, Nikola Kovachki, Kamyar Azizzadenesheli, and Anima Anandkumar. Multi-grid tensorized fourier neural operator for high-resolution pdes. *arXiv preprint arXiv:2310.00120*, 2023.

Appendix

A How to build The Well

A.1 Initial Construction

The Well was built using the following organization method:

- Domain scientists and numerical software developers were contacted. Individuals working with simulations that were sufficiently distinct from existing datasets, non-trivial for learning, and did not require excessive resolution were brought into the collaboration.
- Domain experts were asked to generate data across a sensible range of simulation parameters or initial conditions given the complexity of their simulations. They generated the data on the clusters associated with their home institution.
- The data was then transferred to a cluster for storage and processing.
- Data was analyzed to ensure there were no `NaN`, that the grid and time steps were uniform, and that the data files were consistent.
- A data specification was created for storage, distribution, and programmatic access for machine learning users.
- The data was processed into this common format. A PyTorch Dataset was constructed to read this data for machine learning usage.
- Once processed, a compute budget was allocated to benchmarking based on the size of the data and typical workloads in the space.
- Preliminary benchmarking was performed and results were reported in the paper.

A.2 Data Availability

We are in discussion for making subsets available on HuggingFace upon release. The code and the data will be released when the paper will be published.

A.3 Data Specification

We provide the data with a unified data specification and PyTorch-based interface. The data resides in HDF5 archives with a shared format.

The specification is described below with example entries for a hypothetical 2D ($D=2$) simulation with dimension B x T x H x W. Note that this uses HDF5 Groups, Datasets, and attributes (denoted by "@"):

```
root: Group
  @simulation_parameters: list[str] = ['ParamA', ...]
  @ParamA: float = 1.0
  ... # Additional listed parameters
  @dataset_name: str = 'ExampleDSet'
  @grid_type: str = 'cartesian' # "cartesian/spherical currently supported"
  @n_spatial_dims: int = 2 # Should match number of provided spatial dimensions.
  @n_trajectories: int = B # "Batch" dimension of dataset

  -dimensions: Group
    @spatial_dims: list[str] = ['x', 'y'] # Names match datasets below.
    time: Dataset = float32(T)
      @sample_varying = False # Does this value vary between trajectories?
    -x: Dataset = float32(W) # Grid coordinates in x
      @sample_varying = False
      @time_varying = False # True not currently supported
    -y = float32(H) # Grid coordinates in y
      @sample_varying = False
      @time_varying = False
```

```
-boundary_conditions: Group # Internal and external boundary conditions
  -X_boundary: Group
    @associated_dims: list[str] = ['x'] # Defined on x
    # If associated with set values for given field.
    @associated_fields: list[str] = []
    # Geometric description of BC. Currently support periodic/wall/open
    @bc_type = 'periodic'
    @sample_varying = False
    @time_varying = False
    -mask: Dataset = bool(256) # True on coordinates where boundary is defined.
    -values: Dataset = float32(NumTrue(mask)) # Values defined on mask points

scalars: Group # Non-spatially varying scalars.
  @field_names: list[str] = ['ParamA', 'OtherScalar', ...]
  ParamA: Dataset = float32(1)
    @sample_varying = False # Does this vary between trajectories?
    @time_varying = False # Does this vary over time?
  OtherScalar: Dataset = float32(T)
    @sample_varying = False
    @time_varying = True

t0_fields: Group
  # field_names should list all datasets in this category
  @field_names: list[str] = ['FieldA', 'FieldB', 'FieldC', ...]
  -FieldA: Dataset = float32(BxTxHxW)
    @dim_varying = [ True  True]
    @sample_varying = True
    @time_varying = True
  -FieldB: Dataset = float32(TxHxW)
    @dim_varying = [ True  True]
    @sample_varying = True
    @time_varying = False
  -FieldC: Dataset = float32(BxTxH)
    @dim_varying = [ True  False]
    @sample_varying = True
    @time_varying = True
  ... # Additional fields

-t1_fields: Group
  @field_names = ['VFieldA', ...]
  -VFieldA: Dataset = float32(BxTxHxWxD)
    @dim_varying = [ True  True]
    @sample_varying = True
    @time_varying = True
  ... # Additional fields

-t2_fields: Group
  @field_names: list[str] = ['TFieldA', ...]
  - TFieldA: Dataset = float32(BxTxHxWxD^2)
    @antisymmetric = False
    @dim_varying = [ True  True]
    @sample_varying = True
    @symmetric = True # Whether tensor is symmetric
    @time_varying = True
  ... # Additional fields
```

We did not generate Croissant [102] descriptions of the datasets because the specification does not currently support HDF5 files and would have required converting the 15TB of data to another format that is handled

by the standard. Our provided specification is self-documenting and contains sufficient metadata for machine processing.

For usage purposes, the `GenericWellDataset` outputs all fields as a dictionary giving users the option of how to arrange the input and output for their goals. We include default data processors which add all time-invariant fields as model inputs, but not as targets.

Dataset	Size (GB)	Run time (h)	Hardware	Software
acoustic_scattering_discontinuous	157	0.25	64 C	Clawpack [103]
acoustic_scattering_inclusions	283	0.25	64 C	Clawpack [103]
acoustic_scattering_maze	311	0.33	64 C	Clawpack [103]
active_matter	51.3	0.33	A100 GPU	Python
convective_envelope_rsg	570	1460	80 C	Athena++ [104]
euler_multi-quadrants	5170	80*	160 C*	ClawPack [103]
helmholtz_staircase	52	0.11	64 C	Python
MHD_128	4580	48	64 C	Fortran MPI
MHD_64	72	–	–	–
gray_scott_reaction_diffusion	154	33*	40 C	Matlab
planetswe	186	0.75	64 C	Dedalus [105]
post_neutron_star_merger	110	505*	300 C*	νbhlight [106]
rayleigh_benard	358	60*	768 C*	Dedalus [105]
rayleigh_taylor_instability	256	65*	128 C*	TurMix3D [107]
shear_flow	115	5*	448 C*	Dedalus [105]
supernova_explosion_128	754	4*	1040 C*	ASURA-FDPS [108]
supernova_explosion_64	268	4*	1040 C*	ASURA-FDPS [108]
turbulence_gravity_cooling	829	577*	1040* C	ASURA-FDPS [108]
turbulent_radiative_layer_2D	6.9	2*	48 C	Athena++ [104]
turbulent_radiative_layer_3D	745	271*	128 C	Athena++ [104]
viscoelastic_instability	66	34*	64 C	Dedalus [105]

Table 1: Information about the different dataset generation. In the running time and hardware columns, * denotes a total for all the runs. Otherwise, these figures are given for running one simulation only. For hardware, C denotes the number of Cores. Computation was performed on nodes equipped with either 2 48-core AMD Genoa or 2 32-core Intel Icelake.

B Dataset Details

All numerical simulations are on a uniform grid, uniform time-steps and in single precision `fp32`.

B.1 `acoustic_scattering`

Description of the physical phenomenon. We include three variants of an acoustic scattering problem to showcase the challenges introduces by sharp discontinuities and irregular structure. The acoustic equations describe the evolution of an acoustic pressure wave through materials with spatially varying density. The specific modeling equations used here are:

$$\frac{\partial p}{\partial t} + K(x,y)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0 \tag{1}$$

$$\frac{\partial u}{\partial t} + \frac{1}{\rho(x,y)}\frac{\partial p}{\partial x} = 0 \tag{2}$$

$$\frac{\partial v}{\partial t} + \frac{1}{\rho(x,y)}\frac{\partial p}{\partial v} = 0 \tag{3}$$

with ρ the material density, u,v the velocity in the x,y directions respectively, p the pressure, and K the bulk modulus. ρ and K jointly define the speed of sound and so only ρ is varied in these simulations while K is maintained at a constant value of 4.

These equations are most prevalent in inverse problems like source optimization of a signal or inverse scattering where the underlying material densities are inferred from observed dynamics. These are simple

Dataset	CS	Resolution (pixels)	n_steps	n_traj
acoustic_scattering	(x,y)	256×256	100	8000
active_matter	(x,y)	256×256	81	360
convective_envelope_rsg	(r,θ,ϕ)	$256\times128\times256$	100	29
euler_multi-quadrants	(x,y)	512×512	100	10000
gray_scott_reaction_diffusion	(x,y)	128×128	1001	1200
helmholtz_staircase	(x,y)	1024×256	50	512
MHD	(x,y,z)	64^3 and 256^3	100	100
planetswe	(θ,ϕ)	256×512	1008	120
post_neutron_star_merger	$(\log r,\theta,\phi)$	$192\times128\times66$	181	8
rayleigh_benard	(x,y)	512×128	200	1750
rayleigh_taylor_instability	(x,y,z)	$128\times128\times128$	120	45
shear_flow	(x,y)	128×256	200	1120
supernova_explosion	(x,y,z)	64^3 and 128^3	59	1000
turbulence_gravity_cooling	(x,y,z)	$64\times64\times64$	50	2700
turbulent_radiative_layer_2D	(x,y)	128×384	101	90
turbulent_radiative_layer_3D	(x,y,z)	$128\times128\times256$	101	90
viscoelastic_instability	(x,y)	512×512	variable	260

Table 2: Dataset description: coordinate system (CS), resolution of snapshots, n_steps (number of time-steps per trajectory), n_traj (total number of trajectories in the dataset).

linear dynamics, but the sharp discontinuities in the underlying material density lead to interesting behavior that can be challenging for learned models.

The three datasets vary in the families of material density configurations they consider:

- *Single Discontinuity* - The simplest example consisting of two continuously varying subdomains with a discontinuous interface. The intitial conditions consist of a flat pressure static field with 1-4 high pressure rings randomly placed in domain. The rings are defined with variable intensity $\sim\mathcal{U}(.5,2)$ and radius $\sim\mathcal{U}(.06,.15)$. The subdomain densities are generated from one of the following randomly selected functions:
  - Gaussian Bump - Peak density samples from $\sim\mathcal{U}(1,7)$ and $\sigma\sim\mathcal{U}(.1,5)$ with the center of the bump uniformly sampled from the extent of the subdomain.
  - Linear gradient - Four corners sampled with $\rho\sim\mathcal{U}(1,7)$. Inner density is bilinearly interpolated.
  - Constant - Constant $\rho\sim\mathcal{U}(1,7)$
  - Smoothed Gaussian Noise - Constant background sampled $\rho\sim\mathcal{U}(1,7)$ with IID standard normal noise applied. This is then smoothed by a Gaussian filter of varying sigma $\sigma\sim\mathcal{U}(5,10)$
- *Inclusions* - In this dataset, we first generate a background from the single discontinuity set and further add randomly generated potentially overlapping "inclusions" containing wildly different material properties. This is akin to a geoscience setting with interfaces and mineral deposits. The inclusions are added as 1-15 random ellipsoids with center uniformly sampled from the domain and height/width sampled uniformly from [.05, .6]. The ellipsoid is then rotated randomly with angle sampled [-45, 45]. For the inclusions, $Ln(\rho)\sim\mathcal{U}(-1,10)$.
- *Maze* - This dataset explores complex arrangements of sharp discontinuities. We generated a maze with initial width between 6 and 16 pixels and upsample it via nearest neighbor resampling to create a 256 x 256 maze. The walls are set to $\rho=10^6$ while paths are set to $\rho=3$. The initial sources are generated as a flat pressure static field with 1-6 high pressure rings randomly placed along paths of maze. The rings are defined with variable intensity $\sim\mathcal{U}(3.,5.)$ and radius $\sim\mathcal{U}(.01,.04)$. Any overlap with walls is removed.

Simulation details. The simulations are performed using the total variation diminishing solvers in Clawpack [103], a framework for solving hyperbolic conservation laws using an explicit finite volume scheme, with a monotonized central-difference flux limiter with step-size determined by the CFL condition. The simulation occurs on a domain that is open in the y direction and closed (reflective) in the x direction. Each simulation took approximately 15 minutes of wall time on 64 Icelake CPU cores. Parallelization is done using domain decomposition with ghost node padding for internal boundaries. As the maze simulations are run for more steps, they each required 20 minutes.

Varied Physical Parameters. We vary ρ while keeping K constant to control the material speed of sound c.

Fields present in the data. $\boldsymbol{u}$ or u,v the vector-valued velocity field, p the pressure, and constant fields ρ and c (the material speed of sound).

References to cite when using these simulations: [103].

B.2 `active_matter`

Description of the physical phenomenon. We are interested in studying the dynamics of N active particles of length ℓ and thickness b (aspect ratio $\ell/b \gg 1$) immersed in a Stokes fluid with cubic volume V. In large particle limit, continuum kinetic theories describing the evolution of the distribution function $\Psi(\mathbf{x},\mathbf{p},t)$ have proven to be useful tools for analyzing and simulating particle suspensions [109, 110]. The Smoluchowski equation governs Ψ's evolution, ensuring particle number conservation,

$$\frac{\partial \Psi}{\partial t} + \nabla_{\mathbf{x}} \cdot (\dot{\mathbf{x}}\Psi) + \nabla_{\mathbf{p}} \cdot (\dot{\mathbf{p}}\Psi) = 0, \tag{4}$$

where the conformational fluxes $\dot{\mathbf{x}}$ and $\dot{\mathbf{p}}$ are obtained from the dynamics of a single particle in a background flow $\mathbf{u}(\mathbf{x},t)$. The moments of Ψ yield the concentration field $c = \langle 1 \rangle$, polarity field $\mathbf{n} = \langle \mathbf{p} \rangle / c$, and nematic order parameter $\mathbf{Q} = \langle \mathbf{pp} \rangle / c$, with $\langle f \rangle = \int_{|\mathbf{p}|=1} f\Psi \; d\mathbf{p}$. For dense suspensions, the conformational fluxes are

$$\dot{\mathbf{x}} = \mathbf{u} - d_T \nabla_{\mathbf{x}} \log \Psi; \quad \dot{\mathbf{p}} = (\mathbf{I} - \mathbf{pp}) \cdot (\nabla \mathbf{u} + 2\zeta \mathbf{D}) \cdot \mathbf{p} - d_R \nabla_{\mathbf{p}} \log \Psi. \tag{5}$$

Here d_T and d_R are dimensionless translational and rotational diffusion constants, ζ is the strength of particle alignment through steric interactions, and $\mathbf{D} = \langle \mathbf{pp} \rangle$ is the second-moment tensor. The Smoluchowski equation is coupled to the Stokes flow as

$$-\Delta \mathbf{u} + \nabla P = \nabla \cdot \mathbf{\Sigma}, \nabla \cdot \mathbf{u} = 0, \tag{6}$$

$$\mathbf{\Sigma} = \alpha \mathbf{D} + \beta \mathbf{S} : \mathbf{E} - 2\zeta\beta(\mathbf{D} \cdot \mathbf{D} - \mathbf{S} : \mathbf{D}). \tag{7}$$

Here $P(\mathbf{x},t)$ is the fluid pressure, α is the dimensionless active dipole strength, β characterizes the particle density, $\mathbf{E} = [\nabla \mathbf{u} + \nabla \mathbf{u}^\top]/2$ is the symmetric rate-of-strain tensor, and $\mathbf{S} = \langle \mathbf{pppp} \rangle$ is the fourth-moment tensor. The stress tensor $\mathbf{\Sigma}$ in Eq. (7) includes contributions from active dipole strength, particle rigidity, and local steric torques. Despite the fact that kinetic theories are consistent with microscopic details and are amenable to analytical treatment, they are not immune from computational challenges. For instance, in dense suspensions with strong alignment interactions (high ζ), the cost to resolve the orientation field $\mathbf{p}$ is prohibitively high even in 2D. Though approximate coarse-grained models that track only low-order moments exist, they rely on phenomenological [111][112] or learned corrections [113] to close the system. This underscores the need for fast, high-fidelity, data-efficient physical surrogate models to track and predict the evolution of few low-order moments. An autoregressive surrogate model can efficiently screen the high-dimensional parameter space of complex active matter systems and help design self-organizing materials that switch between nontrivial dynamical states in response to external actuation or varying parameters.

Simulation details. We numerically close the system of equations (4)-(6) using pseudo-spectral discretization where Fourier differentiation is used to evaluate the derivatives with respect to space and particle orientation. We use the second order implicit-explicit backward differentiation time-stepping scheme (SBDF2), where the linear terms are handled implicitly and the nonlinear terms explicitly with time-step $\Delta t = 0.0004$. The numerical simulations are performed in a periodic square domain of length $L = 10$ with 256^2 spatial modes and 256 orientational modes. The simulation code will be available. The approximate time to generate the data is 20 minutes per simulation on an A100 80GB GPU in `fp64` precision. In total, this is about 75 hours of simulation.

Varied Physical Parameters. $\alpha \in \{-1,-2,-3,-4,-5\}$ $\beta = 0.8$; $\zeta \in \{1,3,5,7,9,11,13,15,17\}$.

Fields present in the data. concentration (scalar field), velocity (vector field), orientation tensor (tensor field), strain-rate tensor (tensor field).

References to cite when using these simulations: [113].

B.3 convective_envelope_rsg

Description of the physical phenomenon. The 3D radiation hydrodynamic (RHD) equations are [114]:

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{v}) = 0 \tag{8}$$

$$\frac{\partial (\rho \boldsymbol{v})}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{v}\boldsymbol{v} + \mathsf{P}_{\rm gas}) = -\mathbf{G}_r - \rho \boldsymbol{\nabla} \Phi \tag{9}$$

$$\frac{\partial E}{\partial t} + \boldsymbol{\nabla} \cdot [(E + P_{\rm gas})\boldsymbol{v}] = -cG_r^0 - \rho \boldsymbol{v} \cdot \boldsymbol{\nabla} \Phi \tag{10}$$

$$\frac{\partial I}{\partial t} + c\boldsymbol{n} \cdot \boldsymbol{\nabla} I = S(I, \boldsymbol{n}) \tag{11}$$

where ρ is the gas density, $\boldsymbol{v}$ is the flow velocity, $\mathsf{P}_{\rm gas}$ and $P_{\rm gas}$ are the gas pressure tensor and scalar, respectively, E is the total gas energy density, with $E = E_g + \rho v^2/2$, where $E_g = 3P_{\rm gas}/2$ is the gas internal energy density, G_r^0 and $\mathbf{G}_r$ are the time-like and space-like components of the radiation four-force, and I is the frequency integrated intensity, which is a function of time, spatial coordinate, and photon propagation direction $\boldsymbol{n}$. $\boldsymbol{\nabla}\Phi$ is defined as $\boldsymbol{\nabla}\Phi = -Gm(r)/r^2$, where $m(r)$ is the mass inside the radial coordinate r including the mass contained within the simulation inner boundary. The source term describing the interaction between the gas and radiation in a comoving frame as given by

$$S_0(I_0, \boldsymbol{n_0}) = c\rho\kappa_{aP}\left(\frac{ca_rT^4}{4\pi} - J_0\right) + c\rho(\kappa_S + \kappa_{aR})(J_0 - I_0), \tag{12}$$

where κ_{aP} and κ_{aR} are Planck and Rosseland mean absorption opacities from OPAL [115], and and κ_S is the electron scattering opacity, all evaluated in the comoving frame. These simulations neglect stellar rotation and magnetic fields. Similar setups have been used by [116, 117].

Simulation details. The RHD equations are solved using the standard Godunov method in `Athena++` [104], available at `https://www.athena-astro.app/`. The simulation grid is in spherical-polar coordinates with 128 uniform bins in polar angle θ from $\frac{\pi}{4} - \frac{3\pi}{4}$ and 256 bins in azimuth f from $0 - \pi$ with periodic boundary conditions in θ and f. Outside of the simulation domain, `Athena++` uses ghost zones to enforce its boundary conditions. For the "periodic" boundary in θ, the ghost zones from $\pi/4$ ($3\pi/4$) are copied from the last active zones around the $3\pi/4$ ($\pi/4$) boundary so that the mass and energy flux across the θ boundary is conserved. The radial direction is covered by a logarithmic spaced grid consisting of 384 (256) zones, with $\delta r/r \approx 0.01$, extending far out enough to capture any wind structure or extended atmosphere. The simulations were generated during 2 months on 80 nodes of NASA PLeiades Skylake CPU nodes.

Varied Physical Parameters. All simulations are cuts of a larger simulation. They have all the same physical parameters, but are different times of the same simulation.

Fields present in the data. : energy (scalar field), density (scalar field), pressure (scalar field), velocity (vector field).

References to cite when using these simulations: [21].

B.4 euler_multi-quadrants

Description of the physical phenomenon. This particular set of simulations solves the compressible inviscid Euler equations, which in two dimensions in integral form are

$$\frac{d}{dt}\iint_\Omega U dA + \oint_{\partial\Omega} (F\hat{i} + G\hat{j}) \cdot \hat{n} dS = 0 \tag{13}$$

where $U = (\rho, \rho u, \rho v, \rho E)^T$ and

$$F = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ u(\rho E + p) \end{pmatrix} \quad G = \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ v(\rho E + p) \end{pmatrix}. \tag{14}$$

Here, ρ is the density, u and v are the Cartesian velocities, p is the pressure, and $\rho E = p/(\gamma - 1) + \frac{1}{2}\rho(u^2 + v^2)$ is the total velocity.

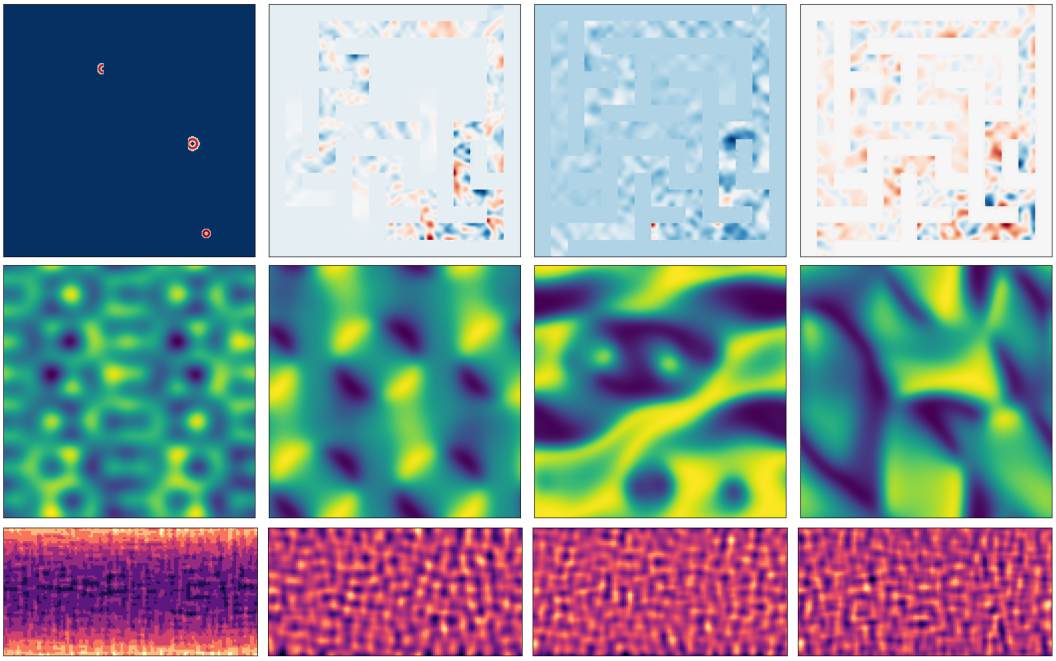

Figure 1: Top to bottom row: snapshots at $t = \{0, \frac{T}{3}, \frac{2T}{3}, T\}$ of `acoustic_scattering`, `active_matter` and `convective_envelope_rsg`.

Simulation details. These simulations used the open source software `CLAWPack` [103, 118], a general framework for solving hyperbolic conservation laws using an explicit finite volume scheme . The simulations use different sets of piecewise constant initial data, which is known as a Riemann problem [66]. The possible solutions are then a combination of shocks, rarefaction waves, or contact discontinuities that sometimes interact as the simulation proceeds in time. The data was generated in `fp64` in 80 hours on 160 CPU cores.

Varied Physical Parameters. $\gamma \in \{1.3, 1.4, 1.13, 1.22, 1.33, 1.76, 1.365, 1.404, 1.453, 1.597\}$ and boundary conditions are either open or periodic.

Fields present in the data. density (scalar field), energy (scalar field), pressure (scalar field), momentum (vector field).

References to cite when using these simulations: [103, 118].

B.5 `helmholtz_staircase`

Description of the physical phenomenon. We simulate linear acoustic scattering of a single point source from an infinite, periodic, corrugated, sound-hard surface. The region $\Omega \in \mathbb{R}^2$ above the boundary $\partial\Omega$ is simply connected and filled with a constant-density gas with sound speed $c > 0$. We define $\mathbf{x} = (x_1, x_2)$. The boundary $\partial\Omega$ extends with spatial period d in the x_1 direction and unbounded in the perpendicular x_2 direction. The current example is a right-angled staircase whose unit cell consists of two equal-length line segments at $\pi/2$ angle to each other, see Fig. 1 in [119]. This geometry models a 3D staircase which extends infinitely in the third direction pointing into the plane of the paper. While we solve the problem in the frequency domain, the original time-domain problem is described by the wave equation sourced by a point excitation at $t = 0$ and $\mathbf{x} = \mathbf{x}_0 \in \Omega$,

$$\frac{\partial^2 U(t,\mathbf{x})}{\partial t^2} - \Delta U(t,\mathbf{x}) = \delta(t)\delta(\mathbf{x} - \mathbf{x}_0) \qquad t \in \mathbb{R}, \quad \mathbf{x} \in \Omega, \tag{15}$$

where $\Delta = \nabla \cdot \nabla$ is the spatial Laplacian, and time t is rescaled such that the sound speed $c = 1$. We assume quiescence before the point excitation: $U \equiv 0$ for $t < 0$, and that the normal component of the fluid velocity vanishes at the staircase's surface, yielding Neumann boundary conditions

$$U_n(t,\mathbf{x}) = \mathbf{n} \cdot \nabla U(t,\mathbf{x}) = 0 \qquad t \in \mathbb{R}, \quad \mathbf{x} \in \partial\Omega, \tag{16}$$

where $\mathbf{n}$ is the unit boundary normal pointing into Ω. Taking the Fourier transform with respect to t of Eqs. (15)–(16), we get the inhomogeneous Helmholtz Neumann boundary value problem (BVP) that is the focus of this simulation,

$$-(\Delta+\omega^2)u=\delta_{\mathbf{x}_0} \quad \text{in } \Omega, \tag{17}$$

$$u_n=0 \quad \text{on } \partial\Omega, \tag{18}$$

where $\omega \in \mathbb{R}$ is the emission frequency of the source. We solve for the acoustic pressure u, which is additionally subject to radiation conditions as described in [119].

Scattering from periodic structures occurs in real-life applications such as the design of waveguides on various lengthscales: photonic and phononic crystals, diffraction gratings, antenna arrays, and architectural elements. These applications often involve numerical simulations performed repeatedly in an optimization or inference loop, calling for fast and robust numerical methods. This setting, however, presents some challenges to accurate numerical modelling. The solution domain is unbounded in both the vertical direction and along the surface; truncation in the vertical direction requires satisfying the correct radiation conditions, and naive truncation in the horizontal direction would result in large artificial reflections (and hence errors) due to the possibility of waves being guided along the surface. Periodization—reducing the computation to the unit cell—is seemingly impossible, since the point source breaks the periodicity of the problem. It is possible, however, to express the nonperiodic solution in terms of a family of quasiperiodic solutions via the Floquet–Bloch transform (also referred to as the array scanning method). The current geometric setup involves corner singularities that must be dealt with if high-order accuracy is to be achieved. Finally, as the input frequency ω grows, the computation will become more expensive due to the need for a finer discretization grid to resolve oscillations.

Simulation details. Our simulation combines the Floquet–Bloch transform with a high-order boundary integral equation (BIE) method to solve each of the quasiperiodic BVPs. The main advantage of this approach is a reduction of the number of discretization nodes (and hence computational cost) by conversion of the 2D PDE to an integral to be evaluated on a 1D boundary. High-order accuracy is then achievable via appropriately chosen quadrature rules, which can easily handle the corner singularities. In contrast, finite difference (FD) and finite element (FEM) schemes require finer meshing of the domain near the corners and implement radiation conditions explicitly. The Floquet–Bloch transform has previously been paired with both FD and FEM methods to tackle scattering from a nonperiodic source, but only to low-order accuracy [120, 121]. Other approaches include meshfree methods such as the method of fundamental solutions [122–124] and the plane waves method [125], as well as tools based on the Rayleigh hypothesis [126]. In the high-frequency limit, fast methods exist that exploit approximations including the Helmholtz–Kirchhoff approximation [127] and geometric acoustics [128–130].

The Helmholtz staircase dataset consists of 25600 images generated from 512 distinct input parameter combinations; the parameters are the source frequency ω (takes 16 different values) and the source position $\mathbf{x}_0$ (takes 32 values). All input frequencies lie in the "low-frequency" regime in the sense that there exists a trapped acoustic mode at that frequency, meaning that the input wavelength is of the same lengthscale as the staircase period. For each parameter combination, we generate 50 timesteps spanning one temporal period, $T=2\pi/\omega$, analytically via $U(t,x)=u(t,x)\exp(-i\omega t)$. The simulations are accurate to around 7–8 digits.

We chose the low-frequency regime for the purposes of training due to the existence of trapped modes in this limit. One can identify two distinct spatial frequencies in the generated images: the input frequency ω, which dominates the outgoing waves far away from the boundary, and the *distinct* spatial frequency of the trapped mode visible along the boundary. The prediction algorithm needs to learn that out of the two, it is ω that determines the time-dependence of the acoustic waves, and correctly identify it from the image despite the presence of a trapped mode. This gets increasingly difficult as ω rises (due to the two frequencies growing more disparate), until a cutoff above which trapped modes no longer exist. In the future, it would be of interest to also learn the dispersion relation of trapped modes, i.e. infer the relationship between their spatial frequency and the input frequency based on the boundary geometry. On 64 CPU cores, the simulation takes $\sim 400s$ per input parameter, total ~ 50 hours.

Varied Physical Parameters. frequency of the source $\omega \in \{0.062, 0.251, 0.439, 0.626, 0.813, 0.998, 1.182, 1.363, 1.541, 1.715, 1.882, 2.042, 2.191, 2.323, 2.433, 2.511\}$, with the sources coordinates being all combinations of $x \in \{-0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4\}$ and $y \in \{-0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4\}$.

Fields present in the data. real and imaginary part of accoustic pressure (scalar field), the staircase mask (scalar field, stationary).

References to cite when using these simulations: [119].

B.6 `MHD` (magnetohydrodynamic simulations)

Description of the physical phenomenon. These simulations employ a third-order-accurate hybrid essentially non-oscillatory (ENO) scheme [131] to solve the ideal MHD equations:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0, \tag{19}$$

$$\frac{\partial \rho \boldsymbol{v}}{\partial t} + \nabla \cdot \left[\rho \boldsymbol{v}\boldsymbol{v} + \left(p + \frac{B^2}{8\pi} \right) \mathbf{I} - \frac{1}{4\pi} \mathbf{B}\mathbf{B} \right] = \mathbf{f}, \tag{20}$$

$$\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\boldsymbol{v} \times \mathbf{B}) = 0. \tag{21}$$

Here, ρ is the density, v the velocity, $\mathbf{B}$ denotes the magnetic field, p represents the gas pressure, and $\mathbf{I}$ is the identity matrix. These simulations utilize periodic boundary conditions and an isothermal equation of state, $p = c_{\mathrm{s}}^2 \rho$, where c_{s} is the isothermal sound speed. For the energy source term $\mathbf{f}$, we assume a random large-scale solenoidal driving at a wave number $k \approx 2.5$ (i.e., 1/2.5 of the box size), with continuous driving. The simulations are executed with a 256^3 grid resolution and have been referenced and utilized in numerous prior studies [131–135].

The main control parameters of these MHD simulations are the dimensionless sonic Mach number, $\mathcal{M}\mathrm{s} \equiv |\boldsymbol{v}|/c\mathrm{s}$, and the Alfvénic Mach number, $\mathcal{M}_{\mathrm{A}} \equiv |\boldsymbol{v}|/\langle v_A \rangle$, where $\boldsymbol{v}$ is the velocity, $c\mathrm{s}$ and v_A are the isothermal sound speed and the Alfvén speed respectively, and $\langle \cdot \rangle$ signifies averages over the entire simulation box. A range of sonic Mach numbers is provided for two different regimes of Alfvénic Mach number (see below varied physical parameters). The simulations are either sub-Alfvénic with $\mathcal{M}_{\mathrm{A}} \approx 0.7$ (indicating a strong magnetic field) or super-Alfvénic with $\mathcal{M}_{\mathrm{A}} = 2.0$. The initial Alfvén Mach number in the super-Alfvénic runs is 7.0, but after the small-scale dynamo saturates, the final $\mathcal{M}_{\mathrm{A}}$ value is around 2. These simulations are non-self-gravitating, and the file units are in code units. The MHD simulations are scale-free, allowing users to assign a physical scale to the box length and density [136, 137]. Rescaling these simulations requires maintaining the sonic and Alfvén Mach numbers constant, though other physical quantities (e.g., density, velocity) may be converted to physical units. On 64 CPU cores, it takes 48 hours per simulation.

Varied Physical Parameters. dimensionless sonic Mach number $\mathcal{M}\mathrm{s} \in \{0.5,\ 0.7,\ 1.5,\ 2.0,\ 7.0\}$ and dimensionless Alfvénic Mach number $\mathcal{M}_{\mathrm{A}} \in \{0.7,\ 2.0\}$.

Fields present in the data. Density (scalar field), velocity (vector field), magnetic field (vector field).

References to cite when using these simulations: [131, 138, 135, 139].

B.7 `gray_scott_reaction_diffusion`

Description of the physical phenomenon. The Gray–Scott equations [68] are a set of coupled reaction–diffusion equations describing two chemical species, A and B, whose scalar concentrations vary in space and time:

$$\frac{\partial A}{\partial t} = \delta_A \Delta A - AB^2 + f(1 - A), \tag{22}$$

$$\frac{\partial B}{\partial t} = \delta_B \Delta B + AB^2 - (f + k)B. \tag{23}$$

The two parameters f and k control the "feed" and "kill" rates in the reaction, respectively; specifically, f controls the rate at which species A is added to the system and k controls the rate at which species B is removed. The two diffusion constants δ_A and δ_B govern the rate of diffusion of each species. A zoo of qualitatively different static and dynamic patterns in the solutions are possible depending on the two parameters f and k [140]. There is a rich landscape of pattern formation hidden in these equations.

Simulation details. Many numerical methods exist to simulate reaction–diffusion equations. If low-order finite differences are used, real-time simulations can be carried out using GPUs, with modern browser-based implementations readily available [140, 141]. We choose to simulate with a high-order spectral method here for accuracy and stability purposes. We simulate (22)–(23) in two dimensions on the doubly periodic domain $[-1,1]^2$ using a Fourier spectral method implemented in the MATLAB

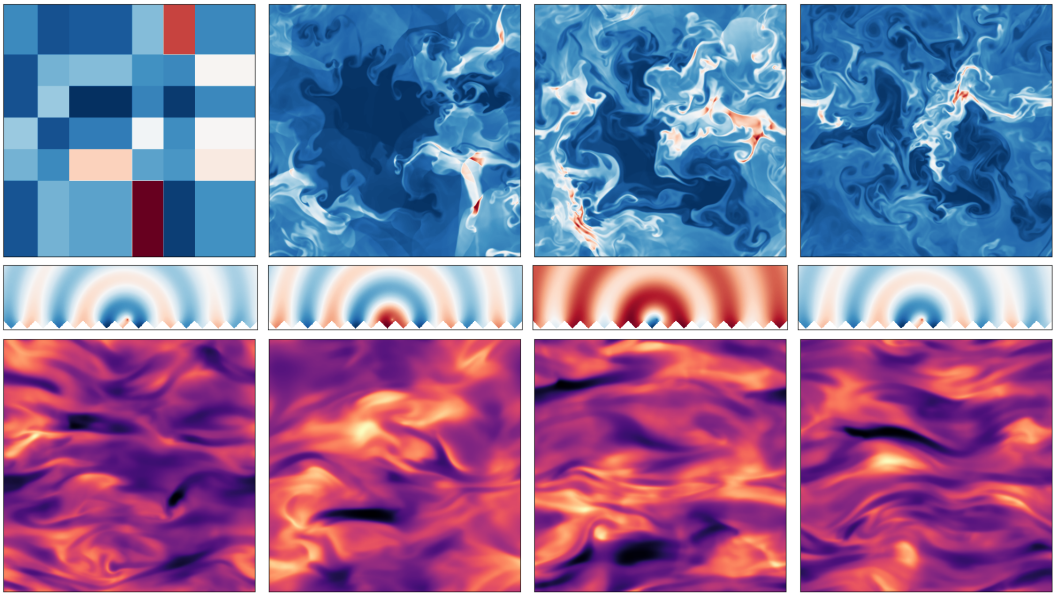

Figure 2: Top to bottom row: snapshots at $t = \{0, \frac{T}{3}, \frac{2T}{3}, T\}$ of `euler_multi-quadrants`, `helmholtz_staircase` and `MHD`.

package Chebfun [142]. Specifically, we use the implicit-explicit exponential time-differencing fourth-order Runge–Kutta method [143] to integrate this stiff PDE in time. The Fourier spectral method is used in space, with the linear diffusion terms treated implicitly and the nonlinear reaction terms treated explicitly and evaluated pseudospectrally. Simulations are performed using a 128×128 bivariate Fourier series over a time interval of 10,000 seconds, with a simulation time step size of 1 second. Snapshots are recorded every 10 time steps. We seed the simulation trajectories with 200 different initial conditions: 100 random Fourier series and 100 randomly placed Gaussians. In all simulations, we set $\delta_A = 0.00002$ and $\delta_B = 0.00001$. Pattern formation is then controlled by the choice of the "feed" and "kill" parameters f and k. We choose six different (f,k) pairs which result in six qualitatively different patterns, summarized in the following table:

	f	k
Gliders	0.014	0.054
Bubbles	0.098	0.057
Maze	0.029	0.057
Worms	0.058	0.065
Spirals	0.018	0.051
Spots	0.030	0.062

On 40 CPU cores, it takes 5.5 hours per set of parameters, 33 hours in total for all simulations.

Varied Physical Parameters. All simulations used $\delta_u = 2.10^{-5}$ and $\delta_v = 1.10^{-5}$. "Gliders": $f = 0.014$, $k = 0.054$. "Bubbles": $f = 0.098$, $k = 0.057$. "Maze": $f = 0.029$, $k = 0.057$. "Worms": $f = 0.058$, $k = 0.065$. "Spirals": $f = 0.018$, $k = 0.051$. "Spots": $f = 0.03$, $k = 0.062$.

Fields present in the data. Two chemical species A and B.

References to cite when using these simulations: None.

B.8 `planetswe`

Description of the physical phenomenon. The shallow water equations are a 2D approximation of a 3D flow in the case where horizontal length scales are significantly longer than vertical length scales. They are derived from depth-integrating the incompressible Navier-Stokes equations. The integrated dimension then

only remains in the equation as a variable describing the height of the pressure surface above the flow. In this case, we specifically explore the rotating forced hyperviscous spherical shallow water equations defined as:

$$\frac{\partial \boldsymbol{u}}{\partial t} = -\boldsymbol{u}\cdot\nabla\boldsymbol{u} - g\nabla h - \nu\nabla^4\boldsymbol{u} - 2\Omega\times\boldsymbol{u} \tag{24}$$

$$\frac{\partial h}{\partial t} = -H\nabla\cdot\boldsymbol{u} - \nabla\cdot(h\boldsymbol{u}) - \nu\nabla^4 h + F \tag{25}$$

where ∇^4 denotes a hyperviscosity term. Hyperviscosity is largely non-physical but is commonly used in atmospheric modeling to maintain stability of under-resolved simulations without effecting large scales to the same degree as conventional diffusion. $\nu = 1.76\times10^{-10}$ is therefore selected for simulation stability - equivalently to matching at wave number 224. F is a forcing term designed to introduce seasonality.

These equations have long been used as a simpler approximation of the primitive equations in atmospheric modeling of a single pressure level, most notably in the Williamson test problems. The scenario in this dataset can be seen as similar to Williamson Problem 7 as we derive initial conditions from the hPa 500 pressure level in ERA5. These are then simulated with realistic topography and two levels of periodicity. Since this is supposed to present a simplified version of the challenges in atmospheric prediction, F is constructed to be a time-dependent forcing term with annual and daily seasonality giving the simulation a sense of "days" and "years", though these are defined in simulation time rather than in physical units. The logic for F is defined in code as:

```
def find_center(t):
    time_of_day = t / day
    time_of_year = t / year
    max_declination = .4
    lon_center = time_of_day*2*np.pi
    lat_center = np.sin(time_of_year*2*np.pi)*max_declination
    lon_anti = np.pi + lon_center
    return lon_center, lat_center, lon_anti, lat_center

def season_day_forcing(phi, theta, t, h_f0):
    phi_c, theta_c, phi_a, theta_a = find_center(t)
    sigma = np.pi/2
    coefficients = np.cos(phi - phi_c) \
                   * np.exp(-(theta-theta_c)**2 / sigma**2)
    forcing = h_f0 * coefficients
    return forcing
```

Simulation details. The simulations are performed using the spin-spherical harmonic pseudospectral method in Dedalus [105] with initial conditions derived from the u,v,z fields in the hPa 500 level of ERA5 [42]. The spatial grid is oversampled by a factor of 3/2 relative to the spectral grid as an anti-aliasing measure following Orszag's rule. To ensure stable initialization, these unbalanced initial conditions are repeatedly simulated for short sequences then projected into hydrostatic balance. The resulting inititial conditions are then burned-in for half a model year. The next three model years are then recorded at an interval of one model hour resulting in a total of 3024 recorded steps per initial condition. The simulation time-step varies according to the CFL condition and is performed using a second-order IMEX Runge-Kutta scheme [144]. The resulting data was interpolated onto a equiangular grid by resampling from the spectral representation.

Each simulation took approximately 45 minutes of wall time on 64 Icelake CPU cores.

Varied Physical Parameters. This data varies only in initial conditions as it is intended to roughly approximate the challenges associated with a specific physical object (the earth).

Fields present in the data. $\boldsymbol{u}$ or u,v the vector-valued velocity field and h the surface height.

References to cite when using these simulations: [145]

B.9 `post_neutron_star_merger`

Description of the physical phenomenon. These simulations are of the disk of hot, dense gas formed after the in-spiral and merger of two neutron stars. These cataclysmic events are now known to be the central engines of gamma ray bursts—some of the most energetic events in the universe—and a primary source of heavy elements in the universe [70–72]. The radioactive decay of heavy elements fused in these systems

produces a reddening glow that can be seen from earth, a *kilonova,* the first observation of which was made in 2017 [73–77]. Of key importance in predicting these events is capturing accurately the interaction of neutrinos, subatomic particles that interact with nuclei and nucleons to convert neutrons to protons and vice versa. The models here are the most advanced simulations in the world of the accretion disk that drives the relativistic jet that powers the gamma ray burst and of the hot neutron-rich wind that sources one component of the kilonova. We solve the general relativistic equations of ideal magnetohydrodynamics coupled to lepton conservation and the neutrino transport equation:[2]

$$\partial_t\left(\sqrt{g}\rho_0 u^t\right)+\partial_i\left(\sqrt{g}\rho_0 u^i\right) = 0 \tag{26}$$

$$\partial_t\left[\sqrt{g}\left(T^t_{\ \nu}+\rho_0 u^t\delta^t_\nu\right)\right]+\partial_i\left[\sqrt{g}\left(T^i_{\ \nu}+\rho_0 u^i\delta^t_\nu\right)\right] = \sqrt{g}\left(T^\kappa_{\ \lambda}\Gamma^\lambda_{\nu\kappa}+G_\nu\right)\ \forall\nu=0,1,...,4 \tag{27}$$

$$\partial_t\left(\sqrt{g}B^i\right)+\partial_j\left[\sqrt{g}\left(b^j u^i-b^i u^j\right)\right] = 0 \tag{28}$$

$$\partial_t\left(\sqrt{g}\rho_0 Y_e u^t\right)+\partial_i\left(\sqrt{g}\rho_0 Y_e u^i\right) = \sqrt{g}G_{\rm ye} \tag{29}$$

$$\frac{D}{d\lambda}\left(\frac{h^3\mathcal{I}_{\nu,f}}{\varepsilon^3}\right) = \left(\frac{h^2\eta_{\nu,f}}{\varepsilon^2}\right)-\left(\frac{\varepsilon\chi_{\nu,f}}{h}\right)\left(\frac{h^3\mathcal{I}_{\nu,f}}{\varepsilon^3}\right), \tag{30}$$

where here ρ_0 is the rest mass density, g is the absolute value of the determinant of the metric tensor, u^μ the is the fluid four-vector, T^μ_ν, the stress energy tensor, δ^μ_ν, the Kronecker delta, $\Gamma^\lambda_{\mu\nu}$ the Christoffel symbols, B^i the magnetic field 3-vector, b^μ the magnetic field four-vector, Y_e the electron fraction (ratio of electrons to baryons, G_ν the radiation field four-fource, G_{ye} the lepton exchange source term. $\mathcal{I}_{\nu,f}$ is the neutrino intensity as a function of position, energy $\varepsilon=h\nu$, and flavor f. $d/d\lambda$ is the total derivative along null geodesics of $\mathcal{I}$. $\eta_{\nu,f}$ is the emissivity and $\chi_{\nu,f}$ the opacity. h is Planck's constant.

Roughly, equation (26) is conservation of mass or particle number. Equation (27) is conservation of energy and momentum. Conservation of momentum is of course Newton's second law, but in general relativity this is combined with conservation of energy. Equation (28) is conservation of magnetic flux. In ideal hydrodynamics, conductivities are assumed to be infinite and thus electric fields can be ignored. Magnetic field lines then get advected with the fluid flow. Equation (29) is conservation of lepton number and controls how neutrons and electrons are advected with the fluid. Equation (30) evolves the motion of neutrinos, which are binned into three flavors: electron neutrinos, their antiparticles, and "heavy neutrinos" which include muon and tau neutrinos and their antiparticles. We assume neutrino mass is negligible and approximate the neutrinos as travelling at the speed of light. Thus we are able to use the radiative transfer equation for photons with some modification. For more details, see [106]. The simulations provided in the Well are from a series of papers, [146–149].

These simulations are computationally expensive and challenging. They require sufficiently high resolution and short time scales to capture the magnetorotational instability, which drives fluid motion [150–152]. But they must also be run for sufficiently long times to track the motion of outgoing material. The electron fraction, Y_e is a critical parameter for heavy element nucleosynthesis which ultimately determines the kilonova signal. ML algorithm that captures bulk fluid motion and tracks the electron fraction Y_e without requiring detailed modeling of magnetohydrodynamic turbulence would be a powerful tool in modeling these systems.

Simulation details. These simulations were produced using the open source ν`bhlight` code, available at `https://github.com/lanl/nubhlight` and first described in [106]. This code builds on a long history of methods spanning more than two decades [153–156]. It solves the equations of ideal general relativistic magnetohydrodynamics via finite volume methods with constrained transport, and uses Monte Carlo methods to perform neutrino radiation transport. The two are coupled via first-order operator splitting. The code uses a radially logarithmic quasi-spherical grid in horizon penetrating coordinates, as first described in [157], the WENO reconstruction first described in [158], the primitive variable recovery scheme described in [159], and the drift-frame artificial atmosphere treatment described in [160].

Simulations were generated using the `torus_cbc` problem generator, which constructs a torus of gas in hydrostatic equilibrium around a rotating black hole, as first detailed in [161, 162]. Initial conditions must specify a black hole mass and and angular momentum, an initial disk entropy, electron fraction, inner radius and radius of maximum pressure, and the preferred units of density (usually chosen so that the peak density is close to 1 in code units). A ratio of gas pressure to magnetic pressure at the point of maximum pressure must also be chosen. (This parameter is called plasma β.) Parameters to reproduce can be found in the cited papers. Finally a finite temperature nuclear equation of state and neutrino opacities must be

[2]Below we use Einstein summation notation. Repeated indices are summed. Greek indices range from 0 to 3 inclusive. Latin indices range from 1 to 3 inclusive.

chosen. The equation of state is the SFHo [163] model. The opacities are the `Fornax` opacities [164] first described in [165]. Both opacities and equation of state are tabulated in Stellar Collapse format [166] and may be found on the web at `https://stellarcollapse.org/`. Each simulation takes ~ 3 weeks to be generated using 300 CPU cores.

Varied Physical Parameters. Black hole spin parameter a, ranges 0 to 1. Initial mass and angular momentum of torus. In dimensionless units, evaluated as inner radius Rin and radius of maximum pressure Rmax. Torus initial electron fraction Ye and entropy kb. Black hole mass in solar masses.

Fields present in the data. fluid density (scalar field), fluid internal energy (scalar field), electron fraction (scalar field), temperate (scalar field), entropy (scalar field), velocity (vector field), magnetic field (vector field), contravariant tensor metric of space-time (tensor field, no time-dependency).

References to cite when using these simulations: [146–149].

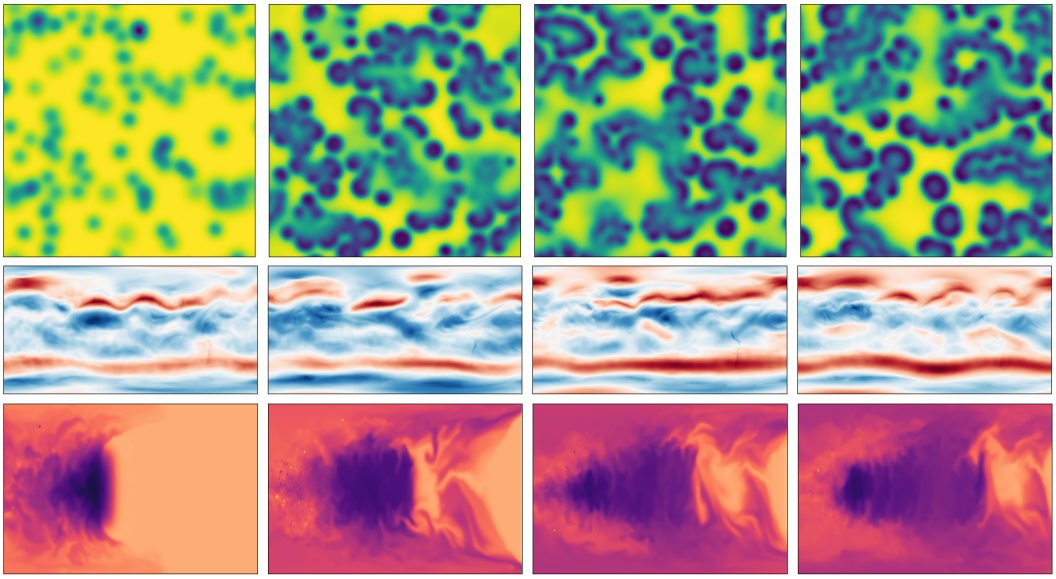

Figure 3: Top to bottom row: snapshots at $t = \{0, \frac{T}{3}, \frac{2T}{3}, T\}$ of `gray_scott_reaction_diffusion`, `planetswe` and `post_neutron_star_merger`.

B.10 `rayleigh_benard`

Description of the physical phenomenon. We consider a 2D horizontally-periodic fluid. We write $u = (u_x, u_z)$ its velocity (horizontal and vertical), b its buoyancy which is the upward force exerted on the fluid due to differences in density, themselves caused by difference in temperature, and p the pressure. The fluid is governed by the equations:

$$\frac{\partial b}{\partial t} - \kappa \Delta b = -u \cdot \nabla b,$$
$$\frac{\partial u}{\partial t} - \nu \Delta u + \nabla p - b \boldsymbol{e}_z = -u \cdot \nabla u,$$

where $\Delta = \nabla \cdot \nabla$ is the spatial Laplacian and $\boldsymbol{e}_z$ is the unit vector in the vertical direction, with the additional constraints $\int p = 0$ (pressure gauge). The first equation rules the convection and diffusion in the fluid, while the second equation is a Navier-Stokes equation augmented by the buoyancy force. The fluid is periodic in the horizontal direction but it has boundary conditions in the vertical direction at the bottom $z = 0$ and at the top $z = Lz$ as follows $u(z = 0) = 0$, $b(z = 0) = Lz$ and $u(Lz = 0) = 0$, $b(Lz = 0) = 0$.

The fluid equations are parameterized by the Rayleigh and Prandtl numbers through the thermal diffusivity κ and viscosity ν

$$\kappa = \left(\text{Rayleigh} \times \text{Prandtl}\right)^{-\frac{1}{2}},$$
$$\nu = \left(\frac{\text{Rayleigh}}{\text{Prandtl}}\right)^{-\frac{1}{2}}.$$

The Rayleigh number is a dimensionless parameter that measures the relative importance between the effect of the buoyancy forces and the effect of the viscosity forces and thermal conduction. The Prandlt number is a dimensionless parameter that measures the relative importance between momentum diffusivity and thermal diffusivity [79].

Simulation details. The data is simulated by solving these PDE through spectral methods using the Dedalus software [105]. The solution is evolved over time with adaptive time-steps. High Rayleigh simulations are very time-consuming because they require very small time-step to prevent the solution from diverging [167]. Generation of simulations take between $\sim 6000s$ and $\sim 50000s$ (high Rayleigh numbers take longer to be generated, 60h in total for all simulations.

Varied Physical Parameters. Rayleigh $\in \{1e6,\ 1e7,\ 1e8,\ 1e9,\ 1e10\}$, Prandtl $\in \{0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0\}$. For initial conditions $\delta b_0 \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$.

Fields present in the data. buoyancy (scalar vield), pressure (scalar field), velocity (vector field).

References to cite when using these simulations: [105]

B.11 `rayleigh_taylor_instability`

Description of the physical phenomenon. The key dimensionless parameter for RTI is the dimensionless density difference or Atwood number ($A = (\rho_h - \rho_l)/(\rho_h + \rho_l)$). As RTI is found to be self-similar, the growth rate (α) of the mixing can be characterised by

$$\alpha = \frac{\dot{L}^2}{4AgL}, \tag{31}$$

where L is the width of the turbulent mixing zone.

The flow is governed by equations for continuity, momentum and incompressibility in the case of miscible fluids with common molecular diffusivity:

$$\partial_t \rho + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{u}) = 0, \tag{32}$$
$$\partial_t (\rho \boldsymbol{u}) + \boldsymbol{\nabla} \cdot (\rho \boldsymbol{u}\boldsymbol{u}) = -\boldsymbol{\nabla} p + \boldsymbol{\nabla} \cdot \boldsymbol{\tau} + \rho \boldsymbol{g}, \tag{33}$$
$$\boldsymbol{\nabla} \cdot \boldsymbol{u} = -\kappa \boldsymbol{\nabla} \cdot \left(\frac{\boldsymbol{\nabla}\rho}{\rho}\right). \tag{34}$$

Here, ρ is density, $\boldsymbol{u}$ is velocity, p is pressure, $\boldsymbol{g}$ is gravity, κ is the coefficient of molecular diffusivity and $\boldsymbol{\tau}$ is the deviatoric stress tensor

$$\boldsymbol{\tau} = \rho\nu\left(\boldsymbol{\nabla}\boldsymbol{u} + (\boldsymbol{\nabla}\boldsymbol{u})^T - \frac{2}{3}(\boldsymbol{\nabla}\cdot\boldsymbol{u})\boldsymbol{I}\right), \tag{35}$$

where ν is the kinematic viscosity and $\boldsymbol{I}$ is the identity matrix.

From a fundamental standpoint, we would expect a good machine learning-based model or emulator to advect and mix the density field rather than create or destroy mass to give appropriate statistics. Our simulations are of comparable spatial resolution to simulations run by a large-scale study of the growth rate of RTI [168]. Therefore, we would consider a good emulator to produce a comparable value for the growth rate as reported in their paper for an appropriately similar set of initial conditions. In addition, during the non-linear regime, as turbulence develops, we would expect to observe typical energy spectra of the inertial cascade where energy is distributed following an appropriate $k^{-5/3}$ slope. From a structural perspective, we would expect that for an initialisation with a large variety of modes in the initial spectrum to observe a range of bubbles and spikes (upward and downward moving structures). In the other limit (where there is only one mode in the initial spectrum) we would hope to observe a single bubble and spike [169]. Finally, a good emulator would exhibit a statistically symmetric mixing width for low Atwood numbers in the Boussinesq regime (defined as $A < 0.1$ [170]) and asymmetries in the mixing width for large Atwood number.

Simulation details. We use TURMIX3D [107] to solve the governing equations (32), (33) and (34) on a staggered 'Marker and Cell' type mesh [171] using a 'Lagrange + remap' method with a Helmholtz–Hodge type decomposition. The domain is discretised such that each cell is a cube (i.e. $\Delta x = \Delta y = \Delta z = h$) and parallelized in all three directions using MPI.

The code is second-order in space using an upwind total variation diminishing approach with Van Leer flux limiters [172, 173] and second-order in time using a strong stabilisation preserved Runge-Kutta [174]. Our discretised pressure equation is modified to account for the non-zero divergence of velocity fields and large density difference and reads as

$$\boldsymbol{\nabla}\cdot\left[\frac{1}{\rho^{n+1}}\boldsymbol{\nabla}\left(\frac{\rho_l}{\rho_h-\rho_l}p^n\right)\right]=\frac{\rho_l}{\Delta t(\rho_h-\rho_l)}\boldsymbol{\nabla}\cdot\left(\frac{(\rho^{int}\boldsymbol{u}^{int})}{\rho^{n+1}}+\kappa\frac{\boldsymbol{\nabla}\rho^{n+1}}{\rho^{n+1}}\right), \tag{36}$$

where indices n and $n+1$ refer to times t^n and t^{n+1} and the index int refers to an intermediate time incorporating all remaining forces of the momentum equation. Equation (36) is then solved using a 'red and black' relaxation method coupled with a 'V-cycle' multigrid convergence method [175–177]. The coefficient $\rho_l/(\rho_h-\rho_l)$ normalises the diffusion term to make the pressure solver quasi-independent of the Atwood number [107]. Finally, we must comment on the treatment of viscosity in the code. The kinematic viscosity, ν, is re-scaled to keep the Kolmogorov scale

$$\eta=\nu^{3/4}\langle\varepsilon\rangle^{-(1/4)}, \tag{37}$$

on the order of the mesh resolution. Here $\langle\varepsilon\rangle$ is the mean dissipation rate per unit mass found using the large-scale energy budget rather than the small-scale shear average. Therefore, we define ν as

$$\nu(t)=\left[\left(\frac{h}{2.1}\right)^4\langle\varepsilon\rangle\right]^{1/3}, \tag{38}$$

where, the dissipation rate is determined using the average potential energy $\langle E_p\rangle$ and kinetic energy $\langle K\rangle$ as follows:

$$\langle\varepsilon\rangle=\frac{1}{\langle\rho\rangle L}\frac{d}{dt}\Big(\langle\rho\rangle L\big[\langle E_p\rangle-\langle K\rangle\big]\Big). \tag{39}$$

The coefficient 2.1 is a classical value given by Pope[178] to limit the pile-up of energy on small scales. The use of η here is justified by the presence of a Kolmogorov cascade in RT-driven flows [179–181]. On 128 CPU cores, it takes 1 hour to obtain 1 simulation, $\sim$65 hours in total.

Varied Physical Parameters. We run simulations with 13 different initializations for five different Atwood number $At \in \{\frac{3}{4},\ \frac{1}{2},\ \frac{1}{4},\ \frac{1}{8},\ \frac{1}{16}\}$. The first set on initial conditions considers varying the mean μ and standard deviation σ of the profile $A(k)$ with $\mu\in\{1,\,4,\,16\}$ and $\sigma\in\{\frac{1}{4},\,\frac{1}{2},\,1\}$, the phase (argument of the complex Fourier component) ϕ was set randomly in the range $[0,2\pi)$. The second set of initial conditions considers a fixed mean ($\mu=16$) and standard deviation ($\sigma=0.25$) and a varieed range of random phases (complex arguments $\phi\in[0,\phi_{max})$) given to each Fourier component. The four cases considered are specified by $\phi_{max}\in\{\frac{\pi}{128},\,\frac{\pi}{8},\,\frac{\pi}{2},\,\pi\}$.

Fields present in the data. Density (scalar field), velocity (vector field).

References to cite when using these simulations: [180]

B.12 `shear_flow`

Description of the physical phenomenon. We consider a 2D-periodic incompressible shear flow whose velocity $u=(u_x,u_z)$ (horizontal and vertical) and pressure p are governed by the following Navier-Stokes equation:

$$\frac{\partial u}{\partial t}-\nu\Delta u+\nabla p=-u\cdot\nabla u.$$

where $\Delta=\nabla\cdot\nabla$ is the spatial Laplacian, with the additional constraints $\int p=0$ (pressure gauge). In order to better visualize the shear, we consider a passive tracer field s governed by the advection-diffusion equation

$$\frac{\partial s}{\partial t}-D\Delta s=-u\cdot\nabla s.$$

We also track the vorticity $\omega=\nabla\times u=\frac{\partial u_z}{\partial x}-\frac{\partial u_x}{\partial z}$ which measures the local spinning motion of the fluid. The shear is created by initializing the velocity u at different layers of fluid moving in opposite horizontal directions.

The fluid equations are parameterized by the Reynolds and Schmidt numbers through the viscosity ν and the tracer diffusivity D

$$\nu = (\text{Reynolds})^{-1},$$
$$D = \left(\text{Reynolds} \times \text{Schmidt}\right)^{-1}.$$

The Reynolds number is a dimensionless parameter that measures the relative importance of inertial forces to viscous forces. The Schmidt number measures the relative importance of momentum diffusivity and mass diffusivity.

Shear flows are challenging to model and predict due to their inherent instability and the potential for turbulent transition, which is highly sensitive to initial conditions and external perturbations. This instability leads to complex flow phenomena such as Kelvin-Helmholtz instabilities [182], turbulent eddies, and vortex formation, all of which require high-resolution simulations to capture accurately.

Simulation details. The data is simulated by solving these PDEs through mixed Fourier-Chebychev pseudospectral methods using the Dedalus software [105]. The solution is evolved over time with adaptive time-steps. With 7 nodes of 64 CPU cores, each with 32 tasks running in parallel on each node, it takes ~ 5 hours to generate all the data.

Varied Physical Parameters. Reynolds $\in \{1e4, 5e4, 1e5, 5e5\}$, Schmidt $\in \{0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0\}$. For initial conditions $n_{\text{shear}} \in \{2, 4\}$, $n_{\text{blobs}} \in \{2, 3, 4, 5\}$, $w \in \{0.25, 0.5, 1.0, 2.0, 4.0\}$.

Fields present in the data. Tracer (scalar field), velocity (vector field), pressure (scalar field).

References to cite when using these simulations: [105].

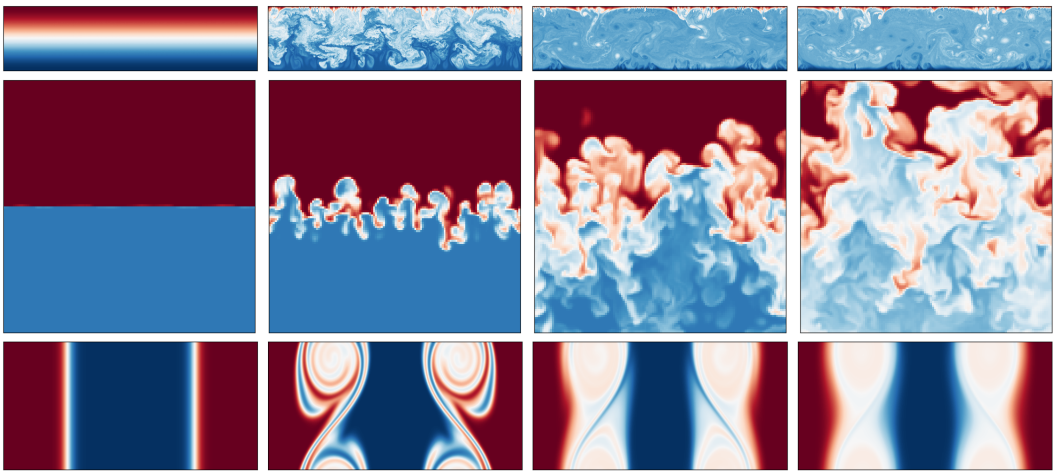

Figure 4: Top to bottom row: snapshots at $t = \{0, \frac{T}{3}, \frac{2T}{3}, T\}$ of `rayleigh_benard`, `rayleigh_taylor_instability` and `shear_flow`.

B.13 `supernova_explosion`

Description of the physical phenomenon. The simulations solve an explosion inside a compression of a monatomic ideal gas, which follows the equation of state with the specific heat ratio $\gamma = 5/3$:

$$P = (\gamma - 1)\rho u, \tag{40}$$

where P, ρ, and u are the pressure, smoothed density, and specific internal energy. The adiabatic compressible gas follows the following equations:

$$\frac{d\rho}{dt} = -\rho \nabla \cdot \boldsymbol{v}, \tag{41}$$

$$\frac{d^2\boldsymbol{r}}{dt^2} = -\frac{\nabla P}{\rho} + \boldsymbol{a}_{\rm visc} - \nabla \Phi, \tag{42}$$

$$\frac{du}{dt} = -\frac{P}{\rho} \nabla \cdot \boldsymbol{v} + \frac{\Gamma - \Lambda}{\rho}, \tag{43}$$

where r is the position, $a_{\rm visc}$ is the acceleration generated by the viscosity, Φ is the gravitational potential, Γ is the radiative heat influx per unit volume, and Λ is the radiative heat outflux per unit volume.

Under a one-dimensional spherical symmetry model [183], an analytic solution describes the propagation of blastwaves in a uniform medium. The time evolution of the radius of the SN shell is written as

$$R(t) = \xi \left(\frac{E}{\rho} \right)^{1/5} t^{2/5}, \tag{44}$$

where E, ρ, and ξ are the energy injected by SN, the density of the surrounding ISM, and the dimensionless similarity variable, respectively. However, ISM has a large density contrast. Turbulence and cooling form a dense filamentary structure, especially in star-forming regions where SN often occurs. Such structure prevents the blastwave's propagation, and the SN remnants' shells become anisotropic.

Simulation details. The simulations are implemented with N-body/SPH code, ASURA-FDPS [184, 108, 185]. To solve the hydrodynamic interaction, a DISPH [186] is employed. SPH methods may encounter difficulties resolving contact discontinuities caused by shock waves (such as SN shells) with low mass resolution. Integration timesteps are determined by the resolution and thermal energy [187] so that the blastwave is resolved. Nevertheless, the code has been tested and verified to resolve the shock wave accurately. It can capture the formation of SN shells caused by thermal energy when the mass resolution is finer than 1 solar mass [185, 188]. The gas in simulations has 1 solar metallicity to mimic the environment around the solar system, which causes a strong radiative cooling. For the 128^3 data, it takes $\sim$3500 CPU hours on up to 1040 CPU cores to generate all data. For the 64^3 data, it takes $\sim$3800 hours on up to 1040 CPU cores to generate all data.

Varied Physical Parameters. Initial temperature T_0={100K}, Initial number density of hydrogen ρ_0={44.5/cc}, metallicity (effectively strength of cooling) $Z = \{Z_0\}$.

Fields present in the data. Pressure (scalar field), density (scalar field), temperature(scalar field), velocity (vector field).

References to cite when using these simulations: [185, 188]

B.14 `turbulence_gravity_cooling`

Description of the physical phenomenon. Similar to `supernova_explosion`, the simulations solve a compression of a monatomic ideal gas, which also follows the equations (40) - (43). To explore different evolutions of ISM under several conditions, simulations are performed with variant initial density, initial temperature, and metallicity with a similar setup to [185]. Metallicity refers to the effectiveness of radiative cooling and heating. In this dataset, richer metallicity mostly has a stronger radiative cooling.

Simulation details. Simulations are implemented with N-body/SPH code, ASURA-FDPS [184, 108, 185].A Density-Independent Smoothed Particle Hydrodynamics (DISPH) [186] is employed to solve the hydrodynamic interaction.

The simulations are performed with two resolutions (1 solar mass and 0.1 solar mass) to capture detailed structures in turbulence. First, gas spheres with a total mass of 10^6 solar mass are generated to make initial gas clouds with turbulence following $\propto v^{-4}$ mimicking star-forming regions. By changing radius, uniform densities are varied in three levels. The initial conditions are constructed using the Astrophysical Multi-purpose Software Environment [189–191]. Radiation is included using the metallicity-dependent cooling and heating functions from 10 to 10^9 K generated by CLOUDY version 13.5 [192–194]. Assuming the environment of the Milky Way Galaxy, dwarf galaxies, and the early universe, 1 solar metallicity, 0.1 solar metallicity, and 0 metallicity (adiabatic) are adopted. The turbulent spherical clouds are initialized

at three different temperatures: 10 K, 100 K, and 1000 K. Details about each simulation time are available on the `README.md` of the dataset.

Varied Physical Parameters. Random seeds for generating an initial turbulence velocity field, Initial temperature T_0={10K, 100K, 1000K}, Initial number density of hydrogen ρ_0 ={44.5/cc, 4.45/cc, 0.445/cc}, metallicity (effectively strength of cooling) $Z = \{Z_0, 0.1Z_0, 0\}$.

Fields present in the data. Pressure (scalar field), density (scalar field), temperature(scalar field), velocity (vector field).

References to cite when using these simulations: [185].

B.15 `turbulent_radiative_layer_2D` and `turbulent_radiative_layer_3D`

Description of the physical phenomenon. The simulations solve the standard fluid equations with an additional energy source term, which removes thermal energy at a rate t_{cool} which is fastest for intermediate temperatures between the hot and cold phase. The full system of equations solved is given by:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0 \tag{45}$$

$$\frac{\partial \rho \boldsymbol{v}}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}\boldsymbol{v} + P) = 0 \tag{46}$$

$$\frac{\partial E}{\partial t} + \nabla \cdot ((E+P)\boldsymbol{v}) = -\frac{E}{t_{\text{cool}}} \tag{47}$$

$$E = P/(\gamma - 1) \quad \text{where} \quad \gamma = 5/3 \tag{48}$$

The major result from these simulations and the corresponding analytic theory is that the total volume integrated radiative cooling is proportional to the net rate of transfer of mass from the hot phase to the cold phase, and that both are proportional to the relative velocity of the phases risen to the 3/4 and the cooling time to the -1/4 power, i.e. $\dot{E}_{\text{cool}} \propto \dot{M} \propto v_{\text{rel}}^{3/4} t_{\text{cool}}^{-1/4}$.

Simulation details. 2D data takes 100 CPU core hours on nodes of 48 CPUs to generate all data, while 3D data was generated on 128 core nodes, taking 34560 CPUhours for all simulations.

Varied Physical Parameters. $t_{\text{cool}} = \{0.03, 0.06, 0.1, 0.18, 0.32, 0.56, 1.00, 1.78, 3.16\}$.

Fields present in the data. Density (scalar field), pressure (scalar field), velocity (vector field).

References to cite when using these simulations: [95].

B.16 `viscoelastic_instability`

Description of the physical phenomenon. This dataset contains results from two-dimensional direct numerical simulations between two parallel walls with periodic boundary conditions in the streamwise (horizontal) direction and no velocity at the walls. The governing equations of the problem read,

$$Re(\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u}) + \nabla p = \beta \Delta \mathbf{u} + (1-\beta) \nabla \cdot \mathbf{T}(\mathbf{C}), \tag{49a}$$

$$\nabla \cdot \mathbf{u} = 0. \tag{49b}$$

We consider FENE-P fluids, where the polymeric stress is related to the conformation tensor $\mathbf{C}$ - an ensemble average of the product of the end-to-end vector of each polymer molecule - via

$$\mathbf{T}(\mathbf{C}) := \frac{1}{Wi}\left(\frac{\mathbf{C}}{1 - (\text{tr}(\mathbf{C}) - 3)/L_{\text{max}}^2} - \mathbf{I}\right). \tag{49c}$$

We consider the evolution equation for the polymer conformation tensor $\mathbf{C}$,

$$\partial_t \mathbf{C} + (\mathbf{u} \cdot \nabla)\mathbf{C} + \mathbf{T}(\mathbf{C}) = \mathbf{C} \cdot \nabla \mathbf{u} + (\nabla \mathbf{u})^T \cdot \mathbf{C} + \varepsilon \Delta \mathbf{C}. \tag{49d}$$

In these equations $\mathbf{u} = (u, v)$ is the velocity with u and v the streamwise and wall-normal velocity respectively, p is the pressure, $\beta := \nu_s/\nu$ is a ratio of kinematic viscosities, where ν_s and $\nu_p = \nu - \nu_s$ are the solvent and polymer contributions respectively, and L_{max} is the maximum extensibility of the polymer chains. The half-distance between the plates h and the bulk velocity U_b are used to make the

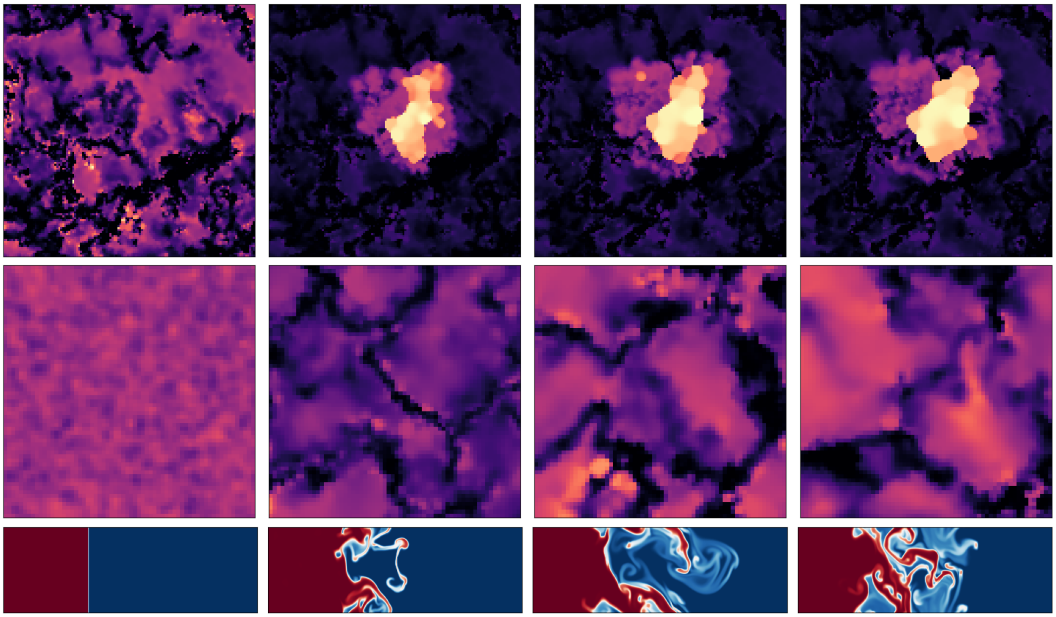

Figure 5: Top to bottom row: snapshots at $t = \{0, \frac{T}{3}, \frac{2T}{3}, T\}$ of `supernova_explosion`, `turbulence_gravity_cooling` and `turbulent_radiative_layer_2D`.

system non-dimensional. The remaining non-dimensional parameters are the Reynolds, $Re := U_b h/\nu$, and Weissenberg, $Wi := \tau U_b/h$, numbers, where τ is the polymer relaxation time, along with the parameter $\varepsilon := D/U_b h$ which is the dimensionless polymer stress diffusivity.

Simulation details. The edge states in the present data set are obtained by bisecting between initial conditions known to reach each attractor. This is done between the laminar state and EIT and between EIT and SAR. The data is generated using the Dedalus codebase [105]. It takes $\sim$1 day to generate $\sim$50 snapshots on 32 or 64 CPU cores, 3 months in total.

Varied Physical Parameters. Reynold number $Re = 1000$, Weissenberg number $Wi = 50$, $\beta = 0.9$, $\epsilon = 2.10^{-6}$, $L_{max} = 70$.

Fields present in the data. pressure (scalar field), velocity (vector field), positive conformation tensor (c^*_{xx},c^*_{yy},,c^*_{xy} are in tensor fields, c^*_{zz} in scalar fields).

References to cite when using these simulations: [96].

C Additional Tasks of Interest

C.1 Moving Beyond the Baselines

The baseline models employed here are powerful but naive models employed en masse without accounting for the specific physical characteristics of the datasets. These are just a starting point for analysis with the Well. Areas for further exploration include:

Physical constraints. Conservation laws and boundary conditions are both key physical properties that can often be directly controlled by a model [195–199]. The Well features a variety of conserved quantities and diverse boundary conditions that can vary within a single dataset, making it well-suited to advance such research.

Long-term stability. Several prior studies have highlighted the difficulty and importance of stable surrogate models [20, 49, 200, 201]. The Well is designed with these studies in mind with most datasets including at least 50 snapshots per trajectory while some include a thousand or more.

Further challenges. While our example baselines target the forward problem, the Well can be used for a variety of other tasks. Several datasets, such as `acoustic_scattering` and `helmholtz_staircase` are well-suited for inverse scattering tasks. Others like `MHD` are coarsened representations of high resolution simulations and could be used for studies of super-resolution. Many contain wide parameter ranges valuable for generalization studies.

The Well contains an enormous diversity of data and can be used for more than forecasting dynamics. We propose a list of additional challenges to be tackled within the Well:

- *Super-resolution:* `MHD` has been downsampled and is available at two resolutions. `supernova_explosion` has been generated at two resolutions. For `MHD` which is downsampled, infer the unresolved scales from the remaining scales. For either, explore generalization from lower resolution training to higher resolution.
- *Transfer across dimensionality:* The same physical phenomenon is represented in 2D and 3D in `turbulent_radiative_layer_2D` and `turbulent_radiative_layer_3D`. Identify approaches for generalizing from cheaper 2D training to more expensive 3D dynamics.
- *Time-steps generalization:* `rayleigh_taylor_instability` simulations for different Atwood numbers have different simulation time-steps. Develop a model trained at a given time-step that can generalize to others.
- *Transfer across a physical parameter range:* Develop a model trained on a restricted range of physical parameters that can generalize to unseen ones which can have different physics behavior. Datasets: `active_matter`, `gray_scott_reaction_diffusion`, `rayleigh_benard`, `viscoelastic_instability`, `shear_flow`, `euler_multi-quadrants` generate data across ranges of parameters that can easily be filtered in the provided dataset object.
- *Steady-state prediction:* `convective_envelope_rsg` and `gray_scott_reaction_diffusion` eventually reach a steady-state. Predict this steady-state from initial conditions.
- *Stable long-term forecasting:* Each trajectory of `planetswe` is rolled out for three model years. Develop models that can produce stable predictions in the sense that the forecasted states follow the same distribution as the simulated system at long time horizons.
- *Inverse-scattering problem:* `acoustic_scattering` and `helmholtz_staircase` contain forward simulations of acoustic waves scattering in response to different material densities. Try instead predicting the material densities from the evolution of the pressure fields.
- *Simulation acceleration:* `post_neutron_star_merger` and `turbulence_gravity_cooling` are enormously expensive simulations taking months to generate. Accurate predictions here can constitute an enormous speed-up relative to the generating process.

D Benchmarking

D.1 Results

D.2 Standard Methodology

The preliminary benchmarks included in the Well are intended to demonstrate the value of new, more challenging tasks for pushing the field forward. As the focus of this work is on the data, our benchmarking methodology is designed to be representative of a generic standard practice in the field both in terms of design choices and computational resources. With that in mind, all benchmarks were performed with the following procedure:

- Baseline models were scaled to approximately 15-20 million parameters.
- Batch size was chosen to maximize GPU memory consumption for a given dataset.
- AdamW was used for all experiments with the PyTorch default WD of .01. We performed a coarse learning rate search over $\{1 \times 10^{-4}, 5 \times 10^{-4}, 1 \times 10^{-3}, 5 \times 10^{-3}, 1 \times 10^{-2}\}$. The run/epoch with the best validation $VRMSE$ was used for subsequent reporting (see Table. 5).
- All models and datasets were trained using Mean Squared Error averaged over fields and space during training.
- Boundary conditions were handled naively according to model architecture. Fourier domain convolutions implicitly used periodic boundaries while spatial domain convolutions utilized standard zero padding.

Dataset	Model			
	FNO	TFNO	U-net	CNextU-net
acoustic_scattering (maze)	0.5062	0.5057	0.0351	**0.0153**
active_matter	0.3691	0.3598	0.2489	**0.1034**
convective_envelope_rsg	**0.0269**	0.0283	0.0555	0.0799
euler_multi-quadrants (periodic b.c.)	0.4081	0.4163	0.1834	**0.1531**
gray_scott_reaction_diffusion	**0.1365**	0.3633	0.2252	0.1761
helmholtz_staircase	**0.00046**	0.00346	0.01931	0.02758
MHD_64	0.3605	0.3561	0.1798	**0.1633**
planetswe	0.1727	**0.0853**	0.3620	0.3724
post_neutron_star_merger	0.3866	**0.3793**	—	—
rayleigh_benard	0.8395	**0.6566**	1.4860	0.6699
rayleigh_taylor_instability (At = 0.25)	>10	>10	>10	>10
shear_flow	0.1567	**0.1348**	0.5910	0.2037
supernova_explosion_64	0.3783	0.3785	**0.3063**	0.3181
turbulence_gravity_cooling	0.2429	0.2673	0.6753	**0.2096**
turbulent_radiative_layer_2D	0.5001	0.5016	0.2418	**0.1956**
turbulent_radiative_layer_3D	0.5278	0.5187	0.3728	**0.3667**
viscoelastic_instability	0.7212	0.7102	0.4185	**0.2499**

Table 3: **Model Performance Comparison:** VRMSE metrics on test sets (lower is better). Best results are shown in **bold**. VRMSE is scaled such that predicting the mean value of the target field results in a score of 1.

- All runs were time-limited to 12 hours on a single Nvidia H100 GPU. Due to the size of these datasets, this intentionally gave an advantage to faster models. As such, we used recent, optimized libraries wherever possible and avoided cutting-edge architectures without optimized GPU kernels.
- Single precision was used for all experiments as several datasets encountered stability issues with mixed or low precision training.

D.3 Models

We opted to stick with time-tested models that are widely used in applications and that natively extend to 3D. This is not intended to be an exhaustive baseline, but rather provide a starting point for the community to use in their own studies. The Fourier Neural Operator [97, FNO] and U-net [99] are among the most widely used models for data driven surrogates. While neither can fairly be called state of the art at this point, they have demonstrated robustness across many problems and are common starting points for practitioners. The TFNO [202] is a more recent tensor-factorized variant of the FNO that improves scalability. We additionally felt that the 2015 variant of the U-net with MaxPool layers and Tanh activations was lacking many recent improvements and so replaced the convolutional blocks with a modern ConvNext [100] architecture for fairer evaluation.

As mentioned in the previous section, all models were scaled to obtain approximately 15-20 million parameters for 2D models. We prioritized reaching this with adjustments to depth or width rather than filter size or downsampling rates. The hyperparameter settings that allowed us to reach these are as follows:

- FNO
  - Spectral filter size (modes) - 16
  - Hidden dimension - 128
  - Blocks - 4
- TFNO
  - Spectral filter size (modes) - 16
  - Hidden dimension - 128
  - Blocks - 4
- U-net Classic
  - Spatial filter size - 3
  - Initial dimension - 48

  - Blocks per stage - 1
  - Up/Down blocks - 4
  - Bottleneck blocks - 1
- CNextU-net
  - Spatial filter size - 7
  - Initial dimension - 42
  - Blocks per stage - 2
  - Up/Down blocks - 4
  - Bottleneck blocks - 1

D.4 Metrics

We evaluate the performance of our models using a diverse set of spatial metrics, namely:

- The mean squared error (MSE): for two spatial fields u and v it is defined as:

$$\mathrm{MSE}(u,v) = \langle |u-v|^2 \rangle,$$

  where $\langle \cdot \rangle$ denotes the spatial mean operator. We also consider its variant the root mean squared error (RMSE) that is the square root of the MSE.

- The normalized mean squared error (NMSE): it corresponds to the MSE normalized by the mean square value of the truth, that is:

$$\mathrm{NMSE}(u,v) = \langle |u-v|^2 \rangle / (\langle |u|^2 \rangle + \epsilon),$$

  where $\epsilon = 10^{-7}$. The term ϵ prevents division by zero in cases where $|u|^2\rangle$ reaches zero. We also consider its square root variant called the NRMSE.

- The variance scaled mean squared error (VMSE): it is the MSE normalized by the variance of the truth

$$\mathrm{VMSE}(u,v) = \langle |u-v|^2 \rangle / (\langle |u-\bar{u}|^2 \rangle + \epsilon).$$

  We chose to report its square root variant, the VRMSE.

- The maximum error (L^∞):

$$L^\infty(u,v) = \max |u-v|$$

  .

- The binned spectral mean squared error (BSMSE): it is the MSE after bandpass filtering of the input fields on a given frequency band $\mathcal{B}$, that is:

$$\mathrm{BSMSE}_{\mathcal{B}}(u,v) = \langle |u_{\mathcal{B}} - v_{\mathcal{B}}|^2 \rangle,$$

  where $u_{\mathcal{B}} = \mathcal{F}^{-1}[\mathcal{F}[u]\mathbf{1}_{\mathcal{B}}]$, with $\mathcal{F}$ the discrete Fourier Transform and $\mathbf{1}_{\mathcal{B}}$ the indicator function over the set of frequencies $\mathcal{B}$. For each dataset, we define three disjoint frequency bands $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$ corresponding to low, intermediate, and high spatial frequencies, respectively. In practice, these bands are defined by partitioning the frequencies based on the magnitudes of their wavenumbers, which are split evenly on a logarithmic scale.

- The binned spectral normalized mean square error is a variant of the previous metric normalized to bin energy of the target:

$$\mathrm{BSNMSE}_{\mathcal{B}}(u,v) = \langle |u_{\mathcal{B}} - v_{\mathcal{B}}|^2 \rangle / \langle |v_{\mathcal{B}}|^2 \rangle,$$

  thus a value of 1 or more indicates that the model would have performed better if it had predicted coefficients of zero corresponding to that scale. This is used in Figure 6 for instance to make the rollout quality more immediately visually interpretable.

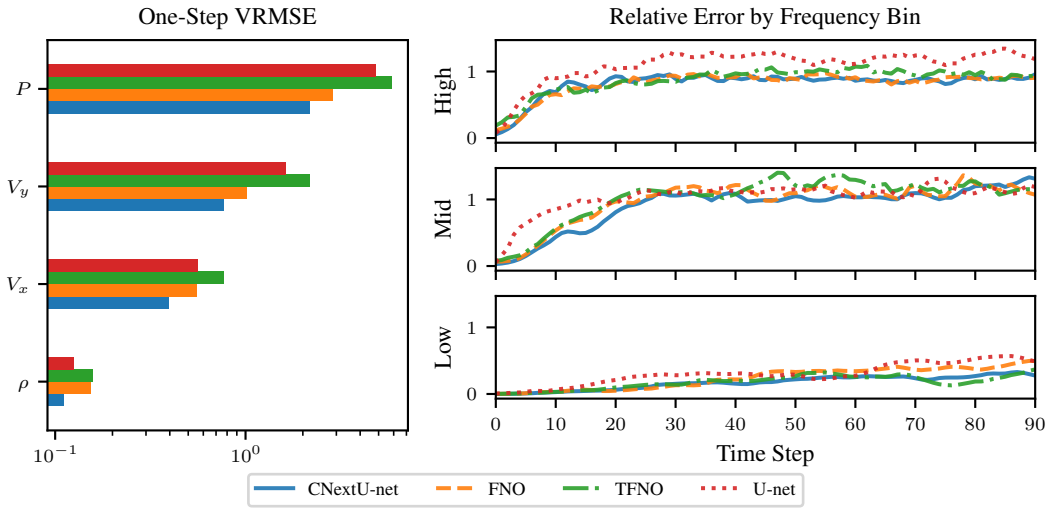

Figure 6: The benchmark library included with the Well includes both coarse and fine level metrics. On the left, we can see the model's performance in VRMSE across state variables. On the right, we divide the isotropic power spectrum into three bins whose boundaries are evenly distributed in log space and evaluate the growth of RMSE per bin normalized by the true bin energy to examine the model's ability to consistently resolve the problem scales.

D.5 Evaluation Metrics

The benchmark library comes equipped with a variety of metrics to inform architecture design in physically meaningful ways. For instance, if we look at the aggregate score for `turbulent_radiative_layer_2D`, most baselines appear to have performed quite poorly. However, in Figure 6, using per-field metrics, we can see that several fields did quite well and that loss is concentrated in the pressure (P) field. Similarly, looking at one-step performance, it appears that CNextU-net has a large advantage, but when we look at longer time horizons, this advantage quickly dissipates and all models apart from the original U-net becomes largely interchangeable. The binning of this error over frequency bins provides further insight as we see all models effectively predict low frequency modes in the long run, but high frequency modes diverge more quickly. The full collection of metrics is described in the supplementary material.

D.6 Results

Due to space limitations, we report the one-step VRMSE in Table 2. In many cases, the simple, generic training approach works quite poorly. We choose VRMSE as the reporting metric as it has the clear interpretation that scores above 1.0 indicates one could have improved the result by predicting the non-spatially varying mean of the target. This is not the same as predicting the population mean, but it is a significantly easier task that predicting the spatially varying target. However, we see in many cases, 12 hours of training is insufficient to reach this level of performance.

When we dig deeper into individual datasets as we do in Figure 6, we can see that this poor performance is not uniform across fields. Even when overall performance is poor, individual fields may obtain good accuracy. Perhaps this is in part due to the use of unnormalized losses during training which could support the use of normalized losses for general surrogate modeling tasks.

Interestingly, though also predictably, we see the model is better able to track the evolution of low frequency modes over time while high frequency modes diverge relatively quickly. The metrics included in the Well pipeline provide valuable insights like this into training and developing new architectures.

More generally, certain datasets proved particularly challenging due to either computational limitations or inherent complexities in their dynamics. For the following datasets, the training could only be done on less than 5 epochs within 12 hours (see Table 5): `convective_envelope_rsg` (544GB), `euler_multi-quadrants` (4.9TB), `turbulence_gravity_cooling` (793GB), `turbulent_radiative_layer_3D` (711GB). Non-time limited training could improve the results.

Dataset	FNO	TFNO	U-net	CNextU-net
acoustic_scattering (maze)	0.5033	0.5034	0.0395	**0.0196**
active_matter	0.3157	0.3342	0.2609	**0.0953**
convective_envelope_rsg	0.0224	**0.0195**	0.0701	0.0663
euler_multi-quadrants (periodic b.c.)	0.3993	0.4110	0.2046	**0.1228**
gray_scott_reaction_diffusion	0.2044	**0.1784**	0.5870	0.3596
helmholtz_staircase	0.00160	**0.00031**	0.01655	0.00146
MHD_64	0.3352	0.3347	0.1988	**0.1487**
planetswe	**0.0855**	0.1061	0.3498	0.3268
post_neutron_star_merger	0.4144	**0.4064**	-	-
rayleigh_benard	0.6049	0.8568	0.8448	**0.4807**
rayleigh_taylor_instability (At = 0.25)	0.4013	**0.2251**	0.6140	0.3771
shear_flow	0.2963	**0.2087**	0.5799	0.3258
supernova_explosion_64	0.3804	0.3645	0.3242	**0.2801**
turbulence_gravity_cooling	0.2381	0.2789	0.3152	**0.2093**
turbulent_radiative_layer_2D	0.4906	0.4938	0.2394	**0.1247**
turbulent_radiative_layer_3D	0.5199	0.5174	0.3635	**0.3562**
viscoelastic_instability	0.7195	0.7021	0.3147	**0.1966**

Table 4: Dataset and model comparison in VRMSE metric on the validation sets, best result in **bold**. VRMSE is scaled such that predicting the mean value of the target field results in score of 1.

Dataset	FNO	TFNO	U-net	CNextU-net
acoustic_scattering (maze)	1E-3 (27)	1E-3 (27)	1E-2 (26)	1E-3 (10)
active_matter	5E-3 (239)	1E-3 (243)	5E-3 (239)	5E-3 (156)
convective_envelope_rsg	1E-4 (14)	1E-3 (13)	5E-4 (19)	1E-4 (5)
euler_multi-quadrants (periodic b.c.)	5E-4 (4)	5E-4 (4)	1E-3 (4)	5E-3 (1)
gray_scott_reaction_diffusion	1E-3 (46)	5E-3 (45)	1E-2 (44)	1E-4 (15)
helmholtz_staircase	5E-4 (132)	5E-4 (131)	1E-3 (120)	5E-4 (47)
MHD_64	5E-3 (170)	1E-3 (155)	5E-4 (165)	5E-3 (59)
planetswe	5E-4 (49)	5E-4 (49)	1E-2 (49)	1E-2 (18)
post_neutron_star_merger	5E-4 (104)	5E-4 (99)	-	-
rayleigh_benard	1E-4 (32)	1E-4 (31)	1E-4 (29)	5E-4 (12)
rayleigh_taylor_instability (At = 0.25)	5E-3 (177)	1E-4 (175)	5E-4 (193)	5E-3 (56)
shear_flow	1E-4 (91)	1E-4 (91)	1E-3 (93)	1E-4 (39)
supernova_explosion_64	1E-4 (40)	1E-4 (35)	5E-4 (46)	5E-4 (13)
turbulence_gravity_cooling	1E-4 (13)	5E-4 (10)	1E-3 (14)	1E-3 (3)
turbulent_radiative_layer_2D	5E-3 (500)	1E-3 (500)	5E-3 (500)	5E-3 (495)
turbulent_radiative_layer_3D	1E-3 (12)	5E-4 (12)	5E-4 (13)	5E-3 (3)
viscoelastic_instability	5E-3 (205)	5E-3 (199)	5E-4 (198)	5E-4 (114)

Table 5: Optimal learning rate and number of training epochs (in parenthesis) to obtain the VRMSE validation loss reported in Table 4.