Benchmarking the Robustness of Neural Network-based Partial Differential Equation Solvers

Jiaqi Gu[†], Mohit Dighamber[‡], Zhengqi Gao[‡], Duane S. Boning[‡],
[†]Arizona State University, [‡]MIT

jiaqigu@asu.edu, mdighamb@mit.edu, zhengqi@mit.edu, boning@mtl.mit.edu

ABSTRACT

The application of artificial intelligence (AI) to scientific domains has witnessed significant growth, notably in the realm of solving partial differential equations (PDEs). Customized neural network models, including convolutional neural networks (CNNs), physics-informed neural networks (PINN), DeepONet, and Fourier neural operator (FNO), have emerged as promising tools for addressing complex scientific computing tasks. Despite their increasing prevalence, the robustness of these models against non-ideal noises and errors in training data remains a relatively unexplored area of inquiry. Such robustness is worth investigating as it ultimately impacts the convergence, training cost, and generalization of the model. In this paper, for the first time, we comprehensively investigate the data robustness of NN-based solvers on different benchmarks. Specifically, we target a state-of-the-art data-driven model, Fourier neural operator, and evaluate its robustness on Burgers' equation, Darcy flow, Navier-Stokes equation, and frequency-domain Maxwell equations, considering training data errors due to non-ideal simulation and data storage process, e.g., random Gaussian noises, low-resolution data downsampling errors, and numerical quantization errors. We believe this comprehensive benchmarking will contribute to a deeper understanding of the learning dynamics and robustness of AI-driven approaches in PDE solving.

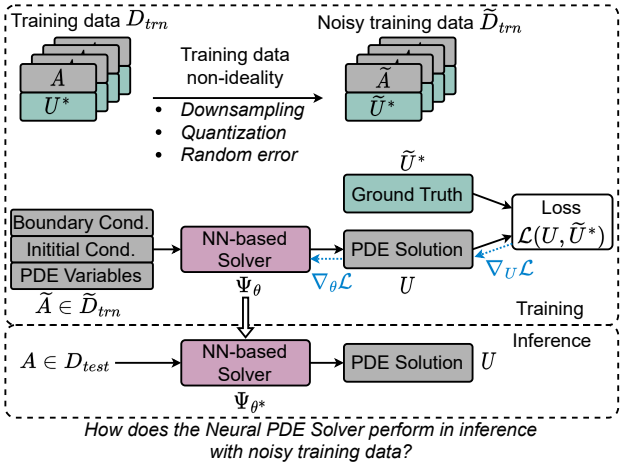

Figure 1: Overview of the robustness investigation for neural PDE solver with noisy training data.

1 INTRODUCTION

Recent research has shown a growing interest in applying artificial intelligence (AI) to scientific domains [2, 4, 5, 7, 8, 10, 12, 13]. This interest has led to the development of tailored machine learning models for mapping within function spaces, specifically for solving partial differential equations (PDEs) in scientific computing tasks such as global weather prediction [9], turbulent flow modeling, and optical device simulation [1, 2, 6, 11].

In comparison to conventional finite element-based numerical solvers, customized neural networks, including convolutional neural networks (CNNs), physics-informed neural networks (PINN) [10], DeepONet [7, 8], and Fourier neural operator (FNO) [4, 5, 12, 13], exhibit substantial speed improvements and competitive solution quality. However, these data-driven or physics-informed neural network models often require a substantial volume of high-quality training data.

Obtaining such high-quality training data, particularly for data-driven models with supervised learning, entails significant costs associated with simulating high-resolution and high-fidelity ground-truth PDE solutions. The resilience of these models against non-ideal noises or errors in training data remains an under-explored area of investigation. Given the importance of high-quality simulation data to neural PDE solvers, the data robustness is worth investigating as it directly impacts the training set acquisition cost, model training cost, training convergence, and model generalization.

For the first time, this paper aims to benchmark the data robustness of NN-based solvers. Among all neural PDE solvers, we target the state-of-the-art data-driven neural operator, Fourier neural operator (FNO), as a case study and evaluate its data robustness on Burgers' equation, Darcy flow, Navier-Stokes equation, and frequency-domain Maxwell equations, considering training data errors including random Gaussian noises, simulation data downsampling errors, and numerical quantization errors. We visualize and quantitatively demonstrate the relative errors when perturbing the training data quality with noise injection. To investigate the impacts on the model robustness, we monitor the training dynamics by comparing the noise-free model gradients with the degraded ones calculated on noisy training data, which serve as a good indicator of the impacts on the final model performance. Furthermore, we compare the training error and inference error to finally evaluate the model generalization across different benchmarks under various noise injection intensities. We draw the conclusion that high-resolution data with low-frequency field/flow patterns demonstrate better tolerance to our injected data errors, especially the downsampling error. Regularization techniques, e.g., dropout and data augmentation, improve the smoothness of the loss landscape, thus enhancing the resilience of neural PDE solvers against non-ideal training data errors. We believe this comprehensive benchmarking will contribute to a deeper understanding of the learning dynamics and robustness of AI-driven approaches in PDE solving and provide insights for future data-efficient and reliable neural-network-based PDE solvers.

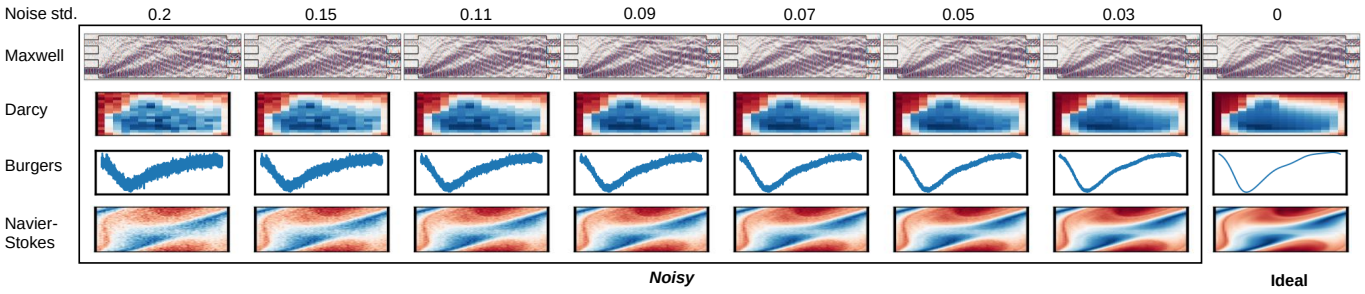

Figure 2: Visualization of the effect of Gaussian errors on the target PDE solutions $\widetilde{U}$ with different noise intensities.

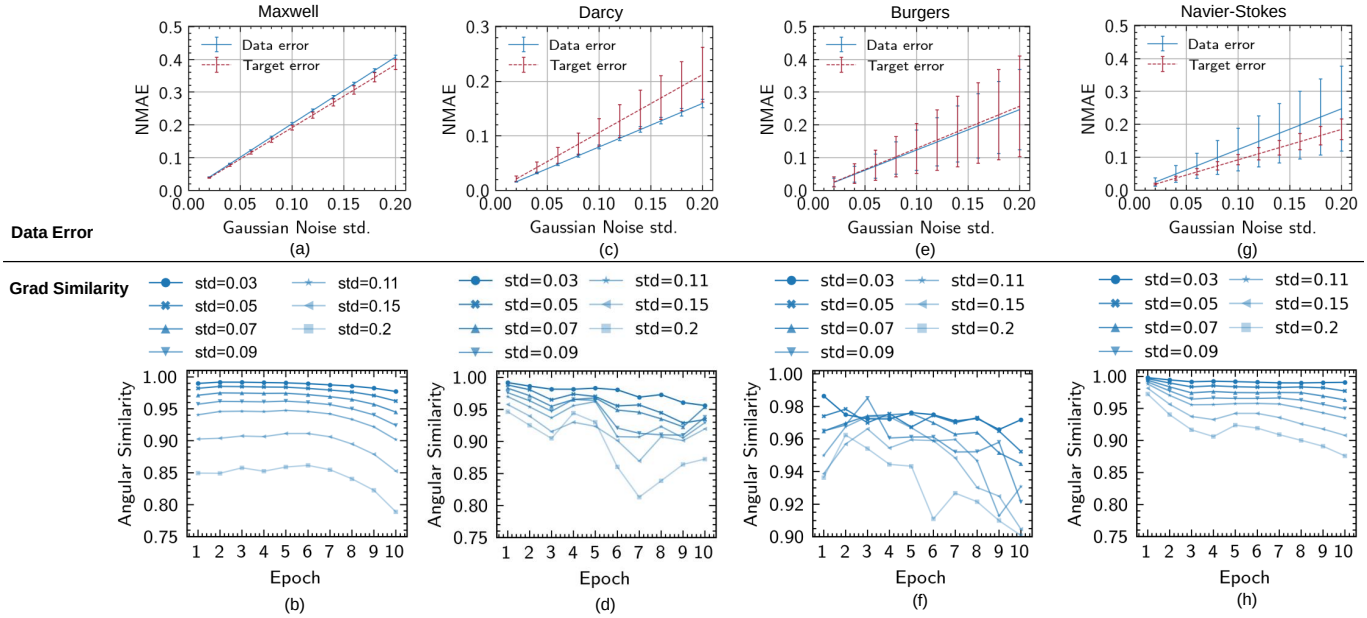

Figure 3: (a)(c)(e)(g) Training input data and target NMAE with different Gaussian noise intensities. (b)(d)(f)(h) Gradient alignment (angular similarity) when training FNO on clean and noisy data with different noise intensities.

2 ROBUSTNESS BENCHMARKING

We evaluate the impact of random Gaussian noises, data downsampling errors, and quantization errors on the training performance and generalization to test dataset. We focus on the Fourier neural operator (FNO) model with 4 different representative PDE solving tasks, including frequency-domain Maxwell equations, Darcy flow, Burgers equation, and Navier-Stokes equations. Note that the errors ϵ are assumed to be pre-injected onto the training input $\widetilde{A} \leftarrow A + \epsilon$ and target field/flow $\widetilde{U} \leftarrow U + \epsilon$ in a *static* fashion to mimic the non-ideality when obtaining the noisy training dataset $\widetilde{\mathcal{D}}_{trn} = \{\widetilde{A}, \widetilde{U}\}$ from the clean dataset $\mathcal{D}_{trn} = \{A, U\}$, not dynamically injected during each learning step. This helps to benchmark the robustness of the neural PDE solvers against low-quality training examples and their ability to generalize the learned mapping to noise-free high-quality queries during inference $\mathcal{D}_{test}$.

2.1 Robustness to Random Noises

We first explore the impact of random Gaussian noise on the training input data and target field/flow, which emulates uniform high-frequency errors existing in the training examples. On normalized target field/flow, we randomly inject static Gaussian noises to the examples before training to create a noisy training set, i.e., $\epsilon \sim \mathcal{N}(0, \sigma^2)$.

Training Data Error. We sweep the Gaussian noise standard derivation from 0.03 to 0.2 and visualize the resultant training target field/flow in Fig. 2. We can see the overall global pattern of the field or flow does not change significantly, while the noises only corrupt the local fine-grained features. In Fig. 3, we quantitatively show the normalized mean-absolute error (NMAE) as the relative discrepancy compared to the noise-free data, i.e., $\|A - \widetilde{A}\|_1 / \|A\|_1$ for input data and $\|U - \widetilde{U}\|_1 / \|U\|_1$ for target field/flow with different noise intensities. In general, the Gaussian noises will lead to 2%-30% errors. It shows a much higher relative error on the electromagnetic field in the frequency-domain Maxwell equations since there are quite some regions with very low light intensities.

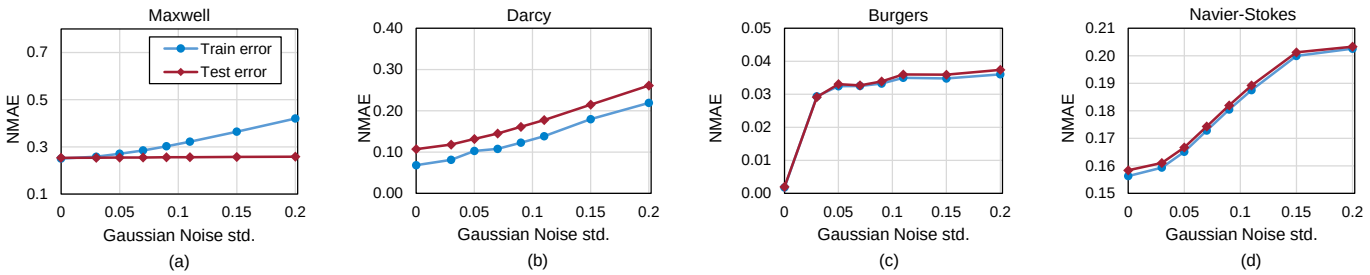

Figure 4: Train and test NMAE of FNO on Maxwell, Darcy, Burgers, and Navier-Stokes simulation tasks with various Gaussian noise std.

Training Dynamics. To deeply investigate how these random noises affect the training performance of the neural PDE solver, we compare the alignment of the gradient when training the same model with noise-free and noisy training examples. We use angular similarity as the metric for the gradient alignment, i.e.,

$$\texttt{Similarity} = 1 - \arccos\Big(\frac{\nabla_\theta \mathcal{L}(\mathcal{D}_{trn}) \cdot \nabla_\theta \mathcal{L}(\widetilde{\mathcal{D}}_{trn})}{\|\nabla_\theta \mathcal{L}(\mathcal{D}_{trn})\| \cdot \|\nabla_\theta \mathcal{L}(\widetilde{\mathcal{D}}_{trn})\|}\Big)/\pi. \quad (1)$$

In Fig. 3, we show the gradient alignment for each epoch averaged across all steps in one epoch. With larger noises, we observe more gradient misalignment across all benchmarks. As training proceeds, the gradient discrepancy becomes larger, which represents that the noisy training data can guide the model to learn the mapping at the beginning but still has different minima in the loss landscape in the later optimization stage compared to its noise-free counterpart, which is the fundamental reason for the reduced inference accuracy for the neural PDE solvers. An interesting observation is when comparing Darcy flow and Burgers equations. Training data of Burgers equation have larger NMAE both for inputs and target fields than that of Darcy flow. However, the noisy gradient shows high alignment with the noise-free ones with >0.9 angular similarity, while the alignment metric in Darcy flow drops to 0.8. One of the reasons is that the inputs/solutions of Burgers equation are 1-D length-8K vectors shown in Fig. 2, which is much easier to learn than the Darcy flow. Therefore, the random high-frequency Gaussian noise will not corrupt the information too much, which is still enough to train the NNs toward the smooth minima in the loss landscape. In contrast, more complicated fields, especially Maxwell equations, suffer from low gradient alignment since the high-frequency periodic wave pattern is more sensitive to Gaussian noises.

Model Robustness and Generalization. Besides the training dynamics investigation, we demonstrate the inference performance of the trained Neural PDE solver to evaluate its generalization and robustness to training data noises. Figure 4 shows the training and test errors on 4 benchmarks with different Gaussian noise intensities. On Darcy, Burgers, and Navier-Stokes equations, the NMAE increases simultaneously on the noisy training set and clean test set, where the prediction performance on Burgers equation suffers from significant degradation with even small noises. It is worth noting that we observe almost the same inference performance on the Maxwell equations, even with degraded training performance. This can be attributed to the regularization techniques applied to the Maxwell training process to avoid overfitting. Due to the application of superposition-based Mixup for data augmentation [2] and dropout layers before the final prediction head, the FNO model for Maxwell equations has better generalization and shows significantly higher tolerance to random Gaussian noises on the training data even when with large gradient misalignment we observe in Fig. 3(b).

2.2 Robustness to Data Downsampling Error

Another possible error source is the simulation data downsampling error. Typically, the raw simulation data is high-resolution, which takes hundreds of GB spaces to store the data, which becomes intractable when we require more data with higher dimensions. Therefore, the simulated training data will be downsampled to smaller sizes that can be efficiently fed into the neural network for mini-batch-based training flow. This downsampling process will inevitably introduce errors to the training examples.

Training Data Error. Different from the element-wise Gaussian noises, the downsampling errors are structural and related to the local patterns. As shown in Fig. 5, we adopt bicubic interpolation to 2D images and linear interpolation to 1D sequences by different scaling factors shown on the top of the figure and scale it back to its original resolution before feeding into the NNs to emulate the downsampling error, i.e., $\widetilde{U} = \texttt{Interp}_{1/s}\big(\texttt{Interp}_s(U)\big)$. In Fig. 6, for Burgers with an 8K vector length and smooth patterns, the downsampling introduces negligible errors (~0 NMAE). For Navier-Stokes equations, the flow is of high resolution and shows good pixel-wise local smoothness such that a 3× downsampling only leads to 4-8% relative errors on input data and target flows.

However, for benchmarks like the 16×16 Darcy flow that already has a relatively low resolution, further downsampling will significantly increase the errors, e.g., 10-30%. Besides the case, another kind of benchmark with high-frequency local features, e.g., oscillating light wave in Maxwell equations, downsampling can be detrimental, leading to almost 80% relative absolute error on the target optical fields. Typically, the pixel size is no shorter than 1/15 of the wavelength in optical simulation [3]. An overly small downsampling factor ruins the validity of Maxwell equations and all interference patterns. In conclusion, for mid-resolution and high-frequency flows/fields, downsampling errors will considerably deteriorate the training data quality.

Training Dynamics. Figure 6 visualizes the training dynamics for the first 10 epochs with different scaling factors. In general, the gradient alignment shows the similar trends as the Gaussian noise cases, more misalignments as training proceeds. For Burgers and Navier-Stokes equations, the gradients are well-aligned. For Maxwell equations, due to the large downsampling errors, we observe large angles between the noisy gradients and the noise-free ones. With 3× downsampling, the gradients are almost orthogonal to each other with a 90-degree angle. Unlike the pixel-wise Gaussian noises with

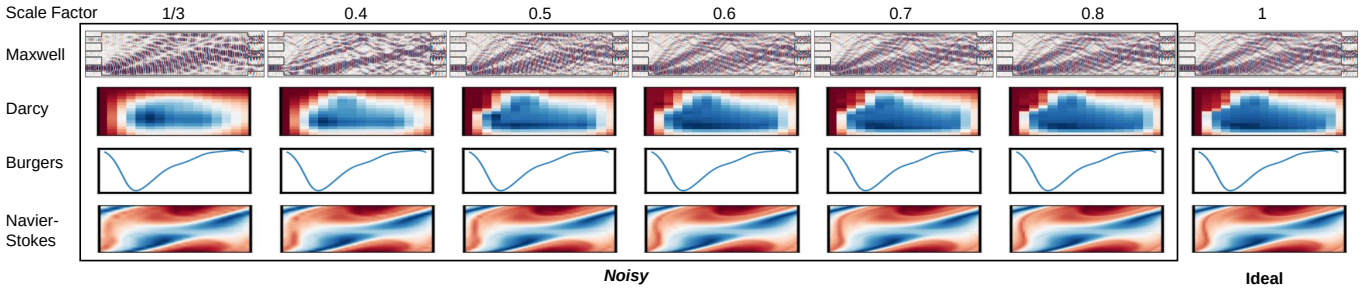

Figure 5: Visualization of the effect of downsampling on the target PDE solutions $\widetilde{U}$ with different scale factors.

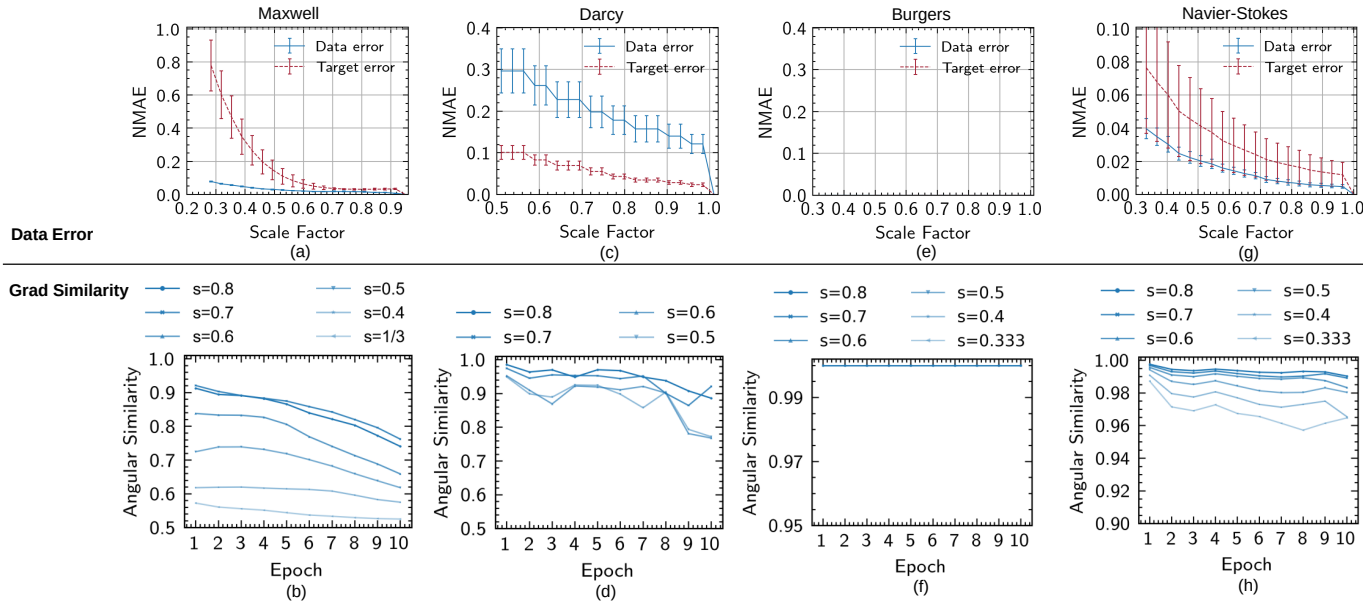

Figure 6: (a)(c)(e)(g)Training input data and target NMAE with different downsampling scale factors. (b)(d)(f)(h) Gradient alignment (angular similarity) when training FNO on clean and noisy data with different downsampling scale factors. Darcy flow data is 16×16 in size, so we only investigate up to 2× downsampling.

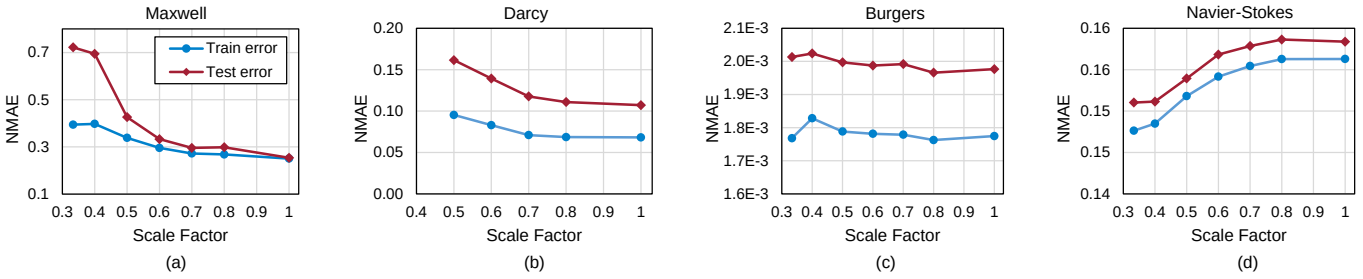

Figure 7: Train and test NMAE of FNO on Maxwell, Darcy, Burgers, and Navier-Stokes simulation tasks with various downsampling scale factors.

uniform intensity, this structured downsampling error misled the model towards a far-away solution space with a high bias compared to the noise-free counterpart.

Model Robustness and Generalization. We observe almost the same test performance on Burgers due to its negligible downsampling error. On Maxwell equations, due to the large gradient misalignment, we observe worse inference performance with smaller scaling factors. Even dropout and data augmentation cannot help regain the robustness. Unlike the pixel-wise Gaussian perturbations that only lead to higher variance without a systematic bias and thus can be tolerated by a locally smooth loss landscape, the regional correlated errors from downsampling cause a systematic bias on the data distribution and thus cannot be fully countered by regularization techniques. A counterintuitive experimental result happens

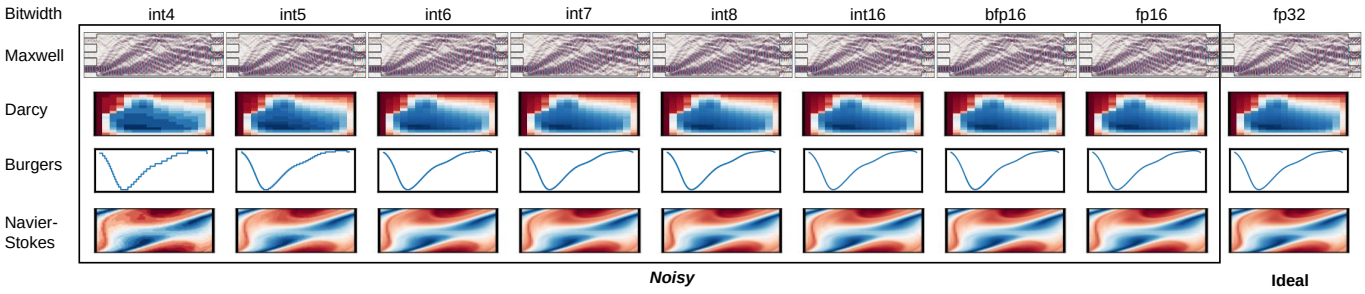

Figure 8: Visualization of the effect of quantization on the target PDE solutions $\widetilde{U}$ with different bitwidths.

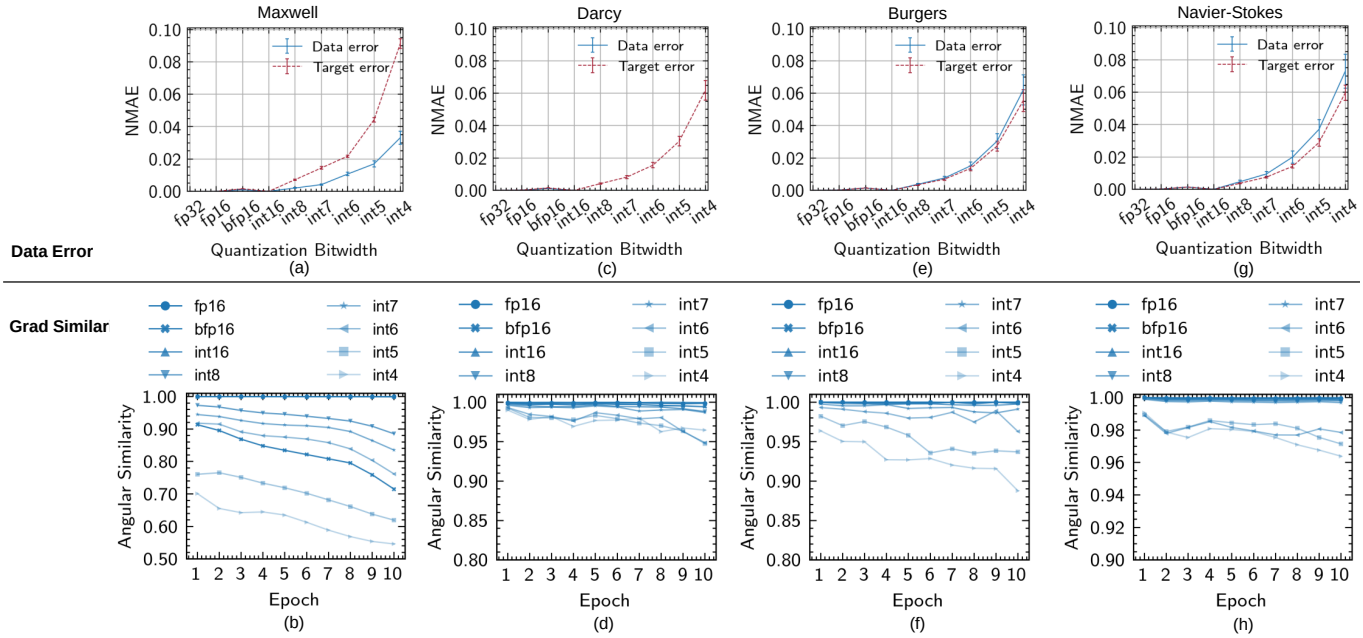

Figure 9: (a)(c)(e)(g) Training input data and target NMAE with different bitwidths. (b)(d)(f)(h) Gradient misalignment (angular similarity) when training FNO on clean and noisy data with different bitwidths.

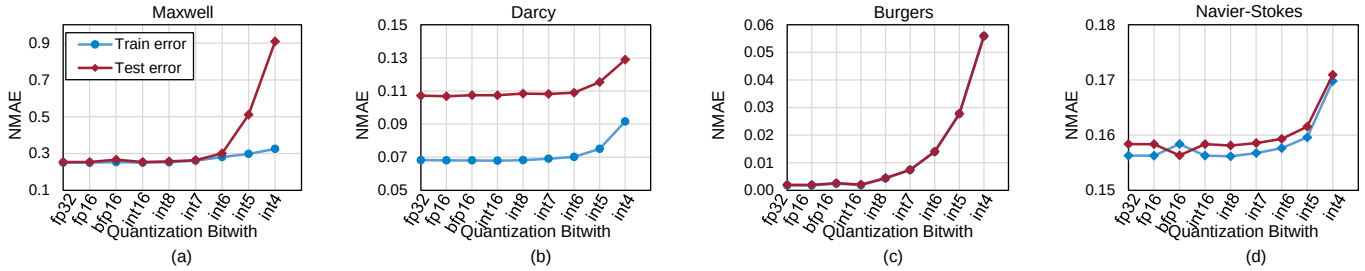

Figure 10: Train and test NMAE of FNO on Maxwell, Darcy, Burgers, and Navier-Stokes simulation tasks with various quantization bitwidth.

on Navier-Stokes equations. The small (4-8%) downsampling error consistently improves the PDE solution prediction quality, as shown in Fig. 7(d). This is not a regularization mechanism, as the training performance also gets improved. The smoothing effects of bicubic interpolation help the model to better converge.

To summarize, downsampling errors are detrimental to high-frequency patterns. The local smoothing effects from downsampling might help the NN PDE solver better converge and thus improve both training performance and generalization on the test dataset.

2.3 Robustness to Numerical Quantization Error

Quantization is a common error that happens in high-precision, high-dynamic range simulation tasks. Typically, double precision or complex-128 is used as the default bitwidth for numerical simulation,

which casts significant storage and processing cost. In this section, we explore how quantization affects the simulation data quality and the training/inference performance on 4 benchmarks.

Training Data Error. In Fig. 8, we can see how discretization impacts the solution quality. We evaluate representative floating point representations, including FP16 and BFP16, widely used in efficient NN training/inference, and integer representations with various bitwidths. For integer quantization, we adopt a min-max quantizer to uniformly discretize tensors U within (U_{min}, U_{max}). Visually, low-bit quantization has only subtle impacts on global patterns. For pixel-wise features, quantization perturbs the field value but maintains the relative magnitude. There are no significant differences across four benchmarks, all leading to 4-8% relative absolute errors compared to float32.

Training Dynamics. From the gradient alignment analysis in Fig. 9, only Maxwell equation benchmark shows high sensitivity to quantization errors, while other benchmarks all maintain high gradient alignment. It is worth noting that Maxwell equations have complex-valued scalar fields in the frequency domain. Quantizing real and imaginary parts separately, though maintaining relative magnitude patterns, can lead to significant phase rotation since a small real component or imaginary component does not mean small intensity, which partially explains the large angles between the noise-free and noisy parameter gradients. Besides, due to a larger range and fewer fraction bits, BFloat16 shows worse gradient fidelity than Float16 as it introduces more errors to small field values, which are also important to the overall phase information.

Model Robustness and Generalization. As shown in Fig. 10, we observe significant impacts on the training error on Maxwell equations with 4-bit quantization. Besides the gradient misalignment, another important reason is the intrinsic sensitivity of Maxwell equation solutions to the input variables, i.e., material permittivity ϵ. A subtle change on the ϵ will lead to a significant change in the resultant fields. Therefore, the large quantization error on the input variable ϵ makes it difficult to learn. While on inference, the model still maintains high prediction fidelity, which shows that the model has good robustness to training quantization error. The regularization and data augmentation techniques applied to Maxwell equation cases improve the model's generalization, such that models trained on low-bit data still perform well on high-precision test data. In general, across different benchmarks, quantization with higher than 6-bit has almost negligible impacts on the training/inference performance, which justifies 8-bit data storage to save memory and disk space in training example collection.

3 CONCLUSION

In this paper, we conduct a comprehensive evaluation of the robustness of several PDE-learning models, including FNO and DeepOnet, when applied to Burgers' equation, Darcy flow, Navier-Stokes equations, and Maxwell's equations. Our evaluation takes into account training data errors arising from non-ideal simulations and data storage processes, such as random Gaussian noises, low-resolution simulation errors, and numerical quantization errors. We aspire for our benchmarking efforts to illuminate the critical concern of model robustness and provide key insights for future data-efficient and high-reliability neural network-based PDE solvers in the realm of AI for scientific applications.

REFERENCES

[1] Mingkun Chen, Robert Lupoiu, Chenkai Mao, Der-Han Huang, Jiaqi Jiang, Philippe Lalanne, and Jonathan Fan. 2021. Physics-augmented deep learning for high-speed electromagnetic simulation and optimization. *Nature* (2021).

[2] Jiaqi Gu, Zhengqi Gao, Chenghao Feng, Hanqing Zhu, Ray T. Chen, Duane S. Boning, and David Z. Pan. 2022. NeurOLight: A Physics-Agnostic Neural Operator Enabling Parametric Photonic Device Simulation. In *Proc. NeurIPS*.

[3] Tyler W. Hughes, Momchil Minkov, Ian A. D. Williamson, and Shanhui Fan. 2018. Adjoint method and inverse design for nonlinear nanophotonic devices. *ACS Photonics* (2018).

[4] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. 2020. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint, arXiv:2003.03485* (2020).

[5] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. 2021. Fourier Neural Operator for Parametric Partial Differential Equations. In *International Conference on Learning Representations (ICLR)*.

[6] Joowon Lim and Demetri Psaltis. 2022. MaxwellNet: Physics-driven deep neural network training based on Maxwell's equations. *Appl. Phys. Lett.* (2022).

[7] Lu Lu, Pengzhan Jin, and George Em Karniadakis. 2019. DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint, arXiv:1910.03193* (2019).

[8] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. 2021. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature machine intelligence* 3, 3 (2021), 218–229.

[9] Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. 2022. FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators. arXiv:2202.11214 [physics.ao-ph]

[10] M. Raissia, P. Perdikarisb, and G. E. Karniadakisa. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comp. Phys.* (2019).

[11] Yingheng Tang, Jichao Fan, Xinwei Li, Jianzhu Ma, Minghao Qi, Cunxi Yu, and Weilu Gao. 2022. Physics-Guided and Physics-Explainable Recurrent Neural Network for Time Dynamics in Optical Resonances. *Nat. Compu. Sci.* (2022).

[12] Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. 2021. Factorized Fourier Neural Operators. In *NeurIPS Workshop*.

[13] Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M. Benson. 2021. U-FNO: an enhanced Fourier neural operator based-deep learning model for multiphase flow. *arXiv preprint, arXiv:2109.03697* (2021).