

A Physics-Aware Deep Learning Model for Shear Band Formation Around Collapsing Pores in Shocked Reactive Materials

Xinlun Cheng,¹ Bingzhe Chen,¹ Joseph Choi,¹ Yen T. Nguyen,² Pradeep Seshadri,²
Mayank Verma,² H.S. Udaykumar,² and Stephen Baek^{1,3, a)}

¹⁾*School of Data Science, University of Virginia*

²⁾*Department of Mechanical Engineering, University of Iowa*

³⁾*Department of Mechanical and Aerospace Engineering,
University of Virginia*

Modeling shock-to-detonation phenomena in energetic materials (EM) requires capturing complex physical processes such as strong shocks, rapid changes in microstructural morphology, and nonlinear dynamics of chemical reactions fronts. These processes participate in energy localization at hotspots, which initiate chemical energy release leading to detonation. This study addresses the formation of hotspots in crystalline EMs subjected to weak-to-moderate shock loading which, despite its critical relevance to the safe storage and handling of EMs, remains underexplored compared to the well-studied strong shock conditions. To overcome the computational challenges associated with direct numerical simulations, we advance the Physics-Aware Recurrent Convolutional Neural Network (PARCv2), which has been shown to be capable of predicting strong shock responses in EMs. We improved the architecture of PARCv2 to rapidly predict shear localizations and plastic heating which play important roles in the weak-to-moderate shock regime. PARCv2 is benchmarked against two widely used physics-informed models, namely Fourier neural operator (FNO) and neural ODE; we demonstrate its superior performance in capturing the spatiotemporal dynamics of shear band formation. While all models exhibit certain failure modes, our findings underscore the importance of domain-specific considerations in developing robust AI-accelerated simulation tools for reactive materials.

^{a)}Electronic mail: mwn4yc@virginia.edu

Topics: Deep learning, Artificial intelligence, Machine learning, Energetic materials, Shear modulus

I. INTRODUCTION

Shock initiation of energetic materials (EM) results from a combination of physico-chemical mechanisms that convert the mechanical energy of a shock into chemical energy release.¹ When the timescale of this chemical energy release becomes comparable to the shock passage time, a detonation wave can form, leading to rapid and catastrophic energy release. Ensuring the safety of EMs therefore requires careful design and control of the energy release process during storage, transportation, or development of EMs. Typical EMs are composed of mixtures of organic energetic crystals (e.g., HMX), polymeric binders, and other additives.² The microstructures of processed EMs are inherently heterogeneous, with energetic crystals and additives interspersed with unavoidable flaws, defects, and porosity. These defect sites act as zones of energy localization and initial chemical reactions when a shock propagates through the heterogeneous energetic material (HE).² Such energy localization sites are known as hotspots³, and the ignition and growth of these hotspots play a crucial role in the shock-to-detonation transition (SDT) or the deflagration-to-detonation transition (DDT). A variety of mechanisms have been identified for hotspot formation.^{4–7} Under strong shock conditions — where shock pressures are in the range of several gigapascal (GPa), far exceeding the Hugoniot Elastic Limit of the HE — the predominant hotspot mechanism is pore collapse.⁵ However, at lower shock strengths, plastic dissipation begins to play a role, and shear bands appear to contribute to energy localization.^{8,9} Under even weaker loading conditions, mechanisms such as friction between crystals can also play a role.

While the strong shock regime and resulting pore collapse have been widely studied, the intermediate shock strength range—where plasticity (deviatoric stress) competes with pore compression and collapse (pressure effects)—remains underexplored. This is largely due to the lack of simulation models that can accurately capture the dynamics of shear bands in a physically consistent manner. Only recently have works begun to address this gap, and the development of appropriate material models is still ongoing.^{10–13}

In the weak-to-moderate shock regime, pore collapse is coupled with plastic dissipation

and localization in shear bands. The resulting hotspot shapes and temperature distributions are more complex and difficult to predict. The initiation and evolution of shear bands have been simulated using molecular dynamics (MD).^{10,11,14} As shock strength increases, the density of shear bands rises, significantly affecting the plastic response of the EM.¹² At the continuum level, modeling the shear- and shear-rate-dependent dynamics of shear bands and their impact on energy localization remains a challenge. In recent work, Herrin *et al.*¹³ proposed an atomistically-consistent modified Johnson-Cook material model for HMX and RDX materials. They demonstrated that continuum simulations of hotspot formation over a wide range of shock strengths show good agreement with corresponding atomistic calculations. While continuum simulations greatly reduce computational cost ($O(1)$ CPU hour on a computing cluster) compared to atomistic simulations ($O(10^3)$ CPU hours on computing clusters), performing ensembles of simulations required for constructing microstructure-informed reactive burn models remains computationally intensive.¹⁵

In recent years, advances in artificial intelligence (AI) techniques have shown promise in significantly accelerating simulations, potentially predicting shock-generated hotspot formation within seconds on a GPU workstation. In general, physics-informed machine learning (PIML) models have demonstrated significantly higher accuracy than traditional physics-agnostic models, especially for complex dynamical systems in fluid flows and solid mechanics.¹⁶ An early and influential class of such methods is Physics-Informed Neural Networks (PINNs).¹⁷ In this approach, a deep neural network (DNN) approximates the solution of governing equations as a function of spatial coordinates (x, y, z) and time (t) variables. Automatic differentiation¹⁸ is employed to compute spatial and temporal derivatives in the governing equations, and the residuals of these equations are minimized during PINN training. However, the nature of PINNs — fitting a highly nonlinear function from space-time coordinates to equation solutions while minimizing a combination of equation residuals and data loss — makes it difficult to generalize to new initial and/or boundary conditions, even when the governing equations remain unchanged.

To address this limitation, alternative approaches such as neural operators (NOs) have been explored.¹⁹ NOs learn mappings between function space operators and have demonstrated improved generalizability. Variants such as the Fourier Neural Operator (FNO)²⁰ and Physics-Informed Neural Operator (PINO)²¹ further enhance this generalizability to

“nearby” physical systems with different initial conditions. Another class of PIML models is Neural Ordinary Differential Equations (NODEs)²², where a neural network approximates the mapping between the system state at time t and its time derivative. The predicted derivatives are then numerically integrated to obtain the state at the next time step, $t + \Delta t$. The network is not trained on ground-truth time derivatives; instead, the loss is computed based on the predicted states, and gradients are backpropagated through the numerical integrator — either directly or via the adjoint method — to update the network weights. This approach has seen widespread use beyond physical system dynamics, including cases where governing equations are unknown^{23,24}, or where inference across varying time steps is essential²⁵.

AI-accelerated or AI-supported simulations are also emerging in EM research, as summarized in the review articles by Nguyen *et al.*²⁶, Choi *et al.*¹⁵ and Udaykumar *et al.*²⁷. Previous research was focused on generation of synthetic microstructures^{28–31}, and predicting the meso- and/or micro-scale evolution of EM under a strong shock loading regime^{32,33}. For example, a novel deep learning architecture known as the Physics-Aware Recurrent Convolutional Neural Network (PARCv2; Nguyen *et al.*³⁴), in which the structure of the advection-diffusion-reaction equations is embedded into the network design, has successfully modeled pore collapse under strong shock loading conditions. However, PARCv2’s performance in predicting the dynamics of pore collapse under weak-to-moderate shock conditions was been found to be less satisfactory. Direct application of PARCv2 with its original architecture leads to deformation of pore geometry, spurious high-temperature regions near boundaries, and numerical instability within 20 roll-out steps across a broad range of impact velocities. The application of machine learning to directly model EM under weak-to-moderate shock conditions has been largely understudied, and this work represents a pioneering work in identifying, investigating and proposing solutions to the challenges associated with this problem, namely spectral bias, boundary condition enforcement, and performance deterioration under low impact velocity.

In this paper, we present substantial improvements to the PARCv2 architecture for modeling shear band formation in materials subjected to weak-to-moderate shock loading. In Section II, we describe data generation, model setup, and training strategies. In Section III, we present our findings, demonstrating that while PARCv2 achieves state-of-the-art accu-

racy compared to two other popular PIML models, all three deep learning models struggle with extrapolation and capturing finer details. In Section IV, we analyze the likely causes of these limitations and propose several refinements to the training strategy and network architecture. We investigated whether spectral bias reduction methods proven successful in computer vision community can translate to similar improvements on physics modeling problem. We examined the theoretical and computational foundations for enforcing various types of boundary conditions through different padding operations in convolutional neural networks. We also broke down the performance decrease in low velocity cases into three likely reasons and unique challenges associated with modeling complex dynamical processes. Finally, in Section V, we draw our conclusions and outline directions for future work.

II. DATA GENERATION

A. Simulation Setup and Dataset

Simulation data are acquired from Herrin *et al.*¹³, capturing the dynamics of pore collapse under the influence of shock waves. The calculations are performed using the continuum dynamics code SCIMITAR3D³⁵, in the configuration shown in FIG. 1, in which a rectangular block of RDX containing an initial circular shaped pore of diameter $D_{void} = 50\text{ nm}$ is subjected to normal-incidence shock loading in a reverse-ballistic configuration wherein the sample impacts with velocity U_p onto a rigid and stationary wall at $y = 0$. This results in a supported shock wave that initially travels with velocity U_w in the frame of the stationary piston, anti-parallel to U_p . The shock velocity U_s in the material frame (i.e., a reference frame moving with the sample) is $U_s = U_w + |U_p|$. For governing equations and simulation setup details, readers are referred to the Appendix of this paper, as well as the supplementary materials of Nguyen *et al.*¹² and Herrin *et al.*¹³, which provide extended formulations and MD-informed model parameters.

The simulation data are rasterized onto a 128×256 grid, corresponding to a spatial resolution of $\Delta x = 1.1719$ nanometers per pixel (nm/pixel), with a temporal resolution of $\Delta t = 2.5$ picoseconds (ps) between adjacent snapshots, starting from the initial condition at $t = 0$. The training-validation sets cover initial impact velocities from 720 m/s to 2880 m/s in 20 m/s increments. The training and validation sets are then randomly split from

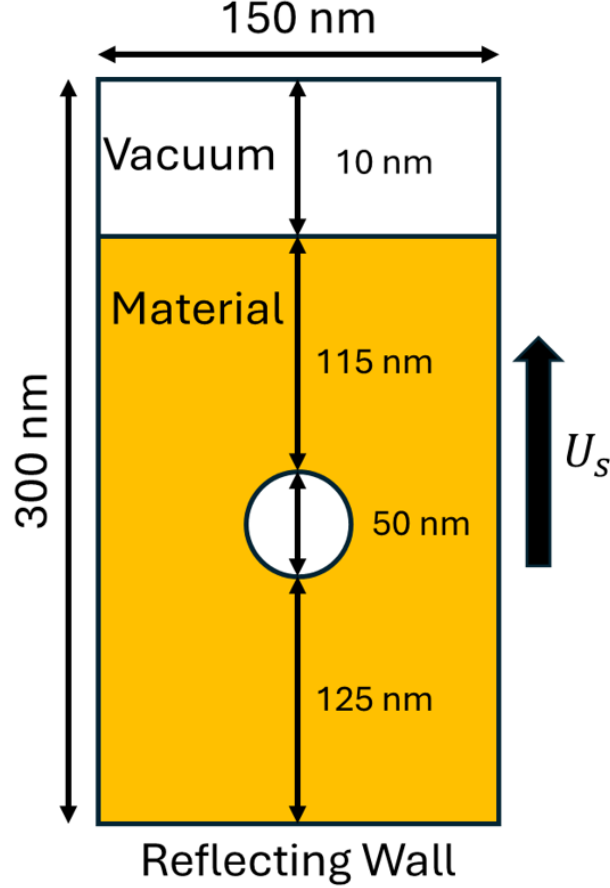


FIG. 1: Simulation configuration for this work. A rectangular block of RDX with a single 50 nm pore impacts a rigid lower wall, generating a supported shock wave that propagates upward and initiates pore collapse and shear localization.

the combined set, yielding 70 training simulations and 18 validation simulations.

The hold-out testing set includes all cases with initial impact velocities divisible by 100 within the aforementioned range, as well as cases in the ranges of 500–700 m/s and 2900–3080 m/s. There are 26 simulations in this hold-out testing set. These cases are selected to assess both in-distribution interpolation and out-of-distribution extrapolation performance of the deep learning models. To ensure fair model comparison, the same training-validation-test split is used across all three deep learning models evaluated in this study.

The dataset is normalized prior to training. Temperature and pressure channels are min-max normalized, while velocity fields are normalized as vectors such that each component lies between -1 and 1 . Pore morphology does not require normalization, as its values

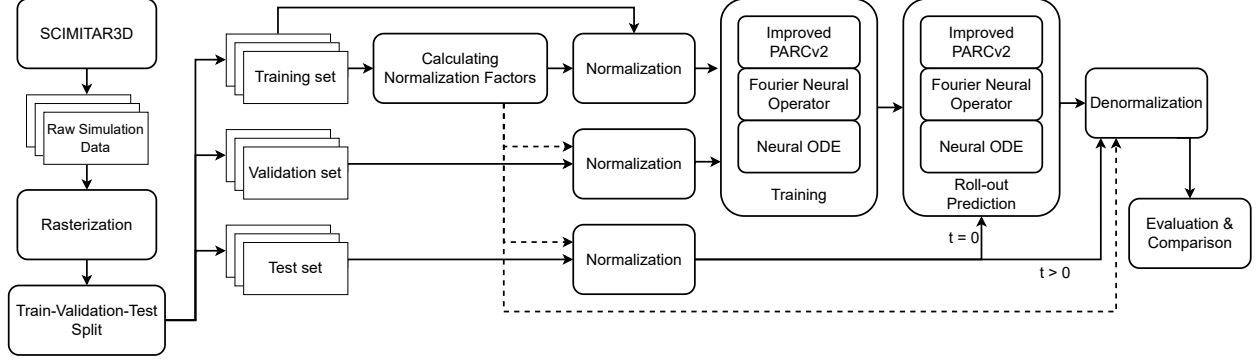


FIG. 2: Flowchart of this work.

are inherently bounded between 0 (vacuum/pore) and 1 (material). Normalization constants are computed using only the training set. The input channels consist of temperature, pressure, pore morphology, and velocity fields, with outputs provided in the same configuration but at subsequent time steps. Because the dataset is uniformly spaced in time and none of the models explicitly depend on time, the actual timestep value does not affect prediction accuracy; we therefore supply a fixed $\Delta t = 0.1$ to all models. All three ML models are trained and evaluated with a single time step as input but, owing to their auto-regressive nature, can generate arbitrarily many future time steps as specified by the user. Model architecture, training hyperparameters and strategies for the three models are presented in the following sections, and their key differences are summarized in TAB. I. Flowchart representing the entire workflow of this paper is presented in FIG. 2.

Model	Architecture Highlights	Physics Law Enforcement
Improved PARCv2	Differentiator-integrator	Embedding governing eqn structures
Fourier Neural Operator	Operator learning	Fourier mode operators
Neural ODE	Differentiator-integrator	Learned from data

TABLE I: A comparison of the ML models presented in this work.

B. Improved PARCv2

We adopted a novel architecture, PARCv2, to address this problem due to its demonstrated capability to model complex transient physical phenomena, such as strong shock

loading of complex material systems with non-linear reactions³⁴ and supersonic flow³⁶. The overall network design is similar to the architectures presented in the two cited works, but includes several notable improvements, which are detailed below:

- In the original work, advection was computed using right derivative regardless of the velocity direction. We have updated this to use an upwind numerical scheme for physical consistency with the advection operator and improved stability.
- Boundary conditions in the original implementation were limited to zero padding. We have extended this to support constant, periodic, and zero-gradient boundary conditions. For this particular problem, we enforced zero-gradient boundary conditions across all boundaries of the domain.
- We adopt a multi-step curriculum training strategy, beginning with 3000 epochs of training to predict the next timestep using a constant learning rate of 10^{-4} , followed by 1000 epochs of training to predict the next three timesteps using a learning rate of 10^{-5} .

The advection of the five state variables — temperature (T), pressure (p), microstructure (μ), horizontal (U) and vertical (V) velocities — is explicitly calculated using an upwind numerical scheme, while additional couplings between them are modeled using a U-Net³⁷. The fourth-order Runge–Kutta (RK4^{38,39}) method is employed for temporal integration, and no data-driven integrator is used. The model contains approximately 15.7 million trainable parameters.

A mean absolute error (MAE, L1) loss function and the Adam optimizer⁴⁰ are employed for both training stages, and no regularization techniques are applied. The model is trained on four Nvidia A100 GPUs using the PyTorch⁴¹ Distributed Data Parallel (DDP) framework, with a total wall-clock training time of 72 hours. The improved PARCv2 package and project code are publicly available at <https://github.com/baeklab/PARCtorch> and https://github.com/chengxinlun/shear_bands, respectively.

C. Fourier Neural Operator

Fourier Neural Operator (FNO) is a deep learning architecture designed to efficiently solve partial differential equations (PDEs) by learning mappings between function spaces. FNO is publicly available under the package *neuraloperator*, which can be found at <https://github.com/neuraloperator/neuraloperator>; version 0.3.0 was used in this work. The architecture leverages Fourier transforms to operate in the frequency domain, enabling it to capture global structures more effectively than many other machine learning methods. FNO integrates neural networks with Fourier transforms to directly learn operators in the Fourier space, making it particularly effective for a range of computational fluid dynamics problems^{42–44}, as spectral and pseudo-spectral method have seen wide usage of application in solving Poisson, Burgers, and incompressible Navier-Stokes equations⁴⁵.

In this work, we employ Tensorized Fourier Neural Operator (TFNO2d) recursively in time to model the shear band formation process, following the recommendation of Li *et al.*²⁰, who advocated the use of recursive FNO2d over FNO3d in data-constrained scenarios. We use four Fourier blocks, retaining the first 24 Fourier modes due to computational resource limitations. Additionally, 78 channels are used for lifting, projection, and projection layers, resulting in a model with a number of trainable parameters comparable to the previously discussed PARCv2 model. All other architectural hyperparameters are kept at their default values.

The model is trained to perform single-step prediction for 10,000 epochs with a constant learning rate of 10^{-4} . Attempts to extend training to predict the next three timesteps showed no significant improvement after 100 epochs and were therefore not pursued further.

D. Neural ODE

Contrary to PINN or FNO, neural ODE approximates the PDE system directly with a neural network and uses numerical integration methods to evolve from the predicted temporal derivatives of physical variables to their values at subsequent snapshots²². The trainable weights in the network can be updated either by backpropagating directly through the numerical integrator or via reverse-time integration of the adjoint equation, as the temporal derivatives are often difficult or impossible to measure directly.

Neural ODEs have gained popularity in the computer vision community for their ability to continuously transform one probability distribution into another (continuous normalizing flow²³) and are attractive for modeling physical systems due to their similarity to conventional numerical simulation routines.

In this work, we use a ResNet⁴⁶ architecture with approximately 15.5 million trainable parameters to approximate the five temporal derivatives from a given input state. The predicted derivatives are then passed through an RK4 numerical integrator to obtain the predicted field values at the next snapshot. The input is first processed through two convolution layers to lift it to 460 channels, followed by four ResNet blocks. Each ResNet block contains two paths: one where the input undergoes a sequence of convolution–batch normalization–ReLU–convolution–batch normalization operations, and another where the input remains unchanged. The outputs of these two paths are added together and passed through a final ReLU activation to produce the output of the ResNet block. An additional 1×1 convolution is applied after the final ResNet block. All convolution operations use 3×3 kernels unless otherwise specified. This model has a similar number of trainable parameters to the PARCv2 model described earlier.

The model is trained on single-step prediction for 3000 epochs with a constant learning rate of 10^{-4} . Attempts to train the model to predict the next three timesteps using a reduced learning rate did not yield significant improvements beyond 100 epochs, likely due to the more complex loss landscape and diminishing gradients encountered during backpropagation through the numerical integrator.

III. RESULTS

In this section, we present the results from applying the three deep learning techniques described above to learn and predict the dynamics of pore collapse in the reverse ballistic impact configuration shown in FIG. 1. A detailed study of the physics underlying this phenomenon has recently been presented in Herrin *et al.*¹³. The top row of FIG. 4 illustrates a typical example of pore collapse for an impact velocity of 1800 m/s. Upon impact, a shock wave travels upward toward the pore, causing it to collapse. The primary features of the resulting hotspot include a high-temperature core and surrounding shear bands that

exhibit energy localization and moderately elevated temperatures. Additionally, a blast wave propagates outward from the site of pore collapse, as seen in the pressure plots in the fourth column of the top row of FIG. 5.

The performance of each AI model is evaluated based on its ability to accurately reproduce the evolution of key field variables – temperature, pressure, and velocity – as well as the pore shape during collapse, relative to the ground truth. All results presented below are based on roll-out prediction over the test set, in which the models are provided only the initial conditions at $t = 0$ and tasked with predicting all subsequent snapshots for test cases that were not seen during training. The number of predicted steps varies depending on the initial impact velocity, ranging from 15 to 50 steps.

A. Roll-out RMSE Comparison

We first examine the model accuracy and compare the three models using root mean squared error (RMSE) against the ground truth (DNS simulation data in the test set). RMSE values reported are averaged over the entirety of the simulation for each channel for each impact velocity. As previous research indicates that PIML model accuracy is related to initial conditions (e.g., Cheng *et al.*³⁶), we present the RMSE in FIG. 3 as a function of initial impact velocities to examine whether a similar dependency exists in the modeling of shear bands. The results show that PARCv2 consistently achieves comparable or lower RMSE values regardless of the physical quantity predicted, indicating that it outperforms the other two models in terms of accuracy across all five fields. The reduction in RMSE achieved by PARCv2 is particularly notable in low-velocity extrapolation cases.

Interestingly, we do not observe significant improvement when comparing PARCv2 predictions to those of FNO for the pore profile or the horizontal (X-) component of the velocity field. We suspect that the pore profile evolution may not be sufficiently complex to fully demonstrate the advantages of PARCv2, while the patterns in the X-velocity field are highly disordered, leading to reduced accuracy in both models. Compared to the neural ODE, PARCv2 achieves nearly a tenfold decrease in RMSE across the five fields, indicating that the inclusion of explicitly calculated advection and diffusion information enables the model to better learn the physical system dynamics by directing attention to regions with strong

advection and gradients.

Nevertheless, we find that even for PARCv2, the RMSE of temperature (25–50 K when within the training set coverage) is comparable to or sometimes greater than the temperature difference between shear bands and the surrounding material, suggesting that non-dominant shear bands may be absent from the predictions of all AI models. Given the physical importance of accurately predicting the evolution and formation of shear bands in this study, we now focus on a more detailed examination of the rollout-predicted fields in this context.

B. Visual Examination of Roll-out Prediction

1. *Interpolation*

We begin our evaluation of model performance with test cases that fall within the training set coverage. FIG. 4—FIG. 8 show the ground truth and predicted evolution for an initial velocity of 1,800 m/s. Throughout the simulation, the predicted sequences from PARCv2 (second row) and FNO (third row) are in close agreement with the ground truth. Neural ODE with a ResNet differentiator (last row), on the other hand, fails to produce reasonable predictions. In the early stage of the simulation (first two columns), when the shock travels through the material before contacting the circular pore, there are no visible differences. During this stage, all three models correctly predict the trivial all-zero solution for the X-velocity, as shown in the first two columns of FIG. 7.

During pore collapse (the third and fourth columns of FIG. 4 — FIG. 8), while the overall patterns are well captured by both PARCv2 and FNO, some finer shear bands present in the ground truth pressure field — such as the alternating blue-cyan regions just below the pore in the top panel of the third column of FIG. 5 — are smoothed out in the predictions, particularly in the bottom panel of the same column. A similar phenomenon is observed in the predicted X and Y velocity fields in the third and fourth columns of FIG. 7. Nevertheless, the dominant shear bands predicted by PARCv2 are noticeably sharper than those produced by FNO, and therefore more closely match the ground truth.

After the pore has fully collapsed and a region of high temperature has developed (fourth column of FIG. 4), both PARCv2 and FNO successfully capture the shape of the high-temperature region (the central bird-shaped pattern), the shock heating phenomenon (a light

blue circle centered on the pattern), and several of the most prominent shear bands (cyan bands radiating from the tips of the central structure). In the pressure field after pore collapse (FIG. 5), both models capture most of the complex shear bands (red bands radiating from the center) and the shock wave generated by the high-temperature region (green circle). However, the predicted pressure fields are less sharp than the ground truth. The earlier observation that PARCv2 produces sharper predictions that more closely resemble the ground truth remains valid at this stage.

Lastly, in the later stages of the simulation (last column of FIG. 4), when reverse ballistics dominate, the reflected pressure wave from the bottom boundary is predicted with high accuracy by both PARCv2 and FNO. At this point in the simulation, the X-velocity field becomes highly disordered, and the predicted field loses some finer details, as seen in FIG. 8. However, the anti-symmetric pattern around the central vertical axis remains well preserved in the predictions from both PARCv2 and FNO. For the Y-velocity, we observe a behavior similar to that of the pressure field: the overall pattern closely resembles the ground truth, with prominent features — such as shock waves generated by the central high-temperature region and those reflected from the bottom boundary — being accurately predicted, though finer structures appear smeared out.

We then visually examine the predictions at $V_0=800$ m/s and $V_0=2,800$ m/s in FIG. 9 and FIG. 10, respectively — the lowest and highest velocity test cases within the training set coverage. For brevity, we present only the temperature fields, as the shear banding phenomenon is most prominent in this field.

For the 800 m/s impact velocity case, the PARCv2 prediction is satisfactory in that three dominant shear bands are accurately predicted at $t = 70$ ps. However, the FNO prediction begins to deviate rapidly from the ground truth as the localized high-temperature region starts to form. This clearly reveals a limitation of commonly used PIML models: fast transient features are considerably more difficult to learn, likely due to their infrequent appearance in the training data. Indeed, the ResNet-based neural ODE is seen to fail in this case. Additionally, most of the weaker shear bands are completely smeared out in the predictions produced by all three methods.

For the 2,800 m/s impact velocity case, as shown in FIG. 10, PARCv2 outperforms FNO, particularly in the prediction of both the localized region of high temperature and

the dominant shear bands. Although the quantitative RMSE loss shown in the top panel of FIG. 3 indicates that the RMSE value for the 2800 m/s case is nearly twice that of the 800 m/s case, the predictions appear visually closer to the ground truth. We suspect that this is because the highest temperature values in the 800 m/s case are significantly lower in value than those in the 2800 m/s case, making deviations more noticeable in the lower-velocity predictions despite the lower RMSE. This finding not only underscores the importance of visually examining model predictions but also contrasts with typical numerical calculations, where high-velocity cases usually pose greater challenges for coarse-grid simulations. We provide further discussion on this behavior in the following sections.

2. *High Velocity Extrapolation*

A few test cases were specifically selected to evaluate the ability of the three models to extrapolate beyond the coverage of the training set. We begin with an examination of high-velocity extrapolation cases ($V_0 > 2,900$ m/s). Quantitatively, we observed a sharp increase in RMSE values, as shown in FIG. 3. Examination of the predicted fields revealed no noticeable artifacts for PARCv2; the increased error is primarily due to inaccuracies in predicting the shape of the central high-temperature region and an overall blurriness in the predicted fields. In the later stages of the simulation (4th and 5th columns of FIG. 11), while the ground truth displays a blunt-shaped high-temperature region, the PARCv2 prediction retains a relatively sharp shape. Although the most prominent shear bands—wide blue bands radiating from the sides of the central high-temperature region—are well predicted, many smaller and weaker structures are lost in the prediction.

In contrast, FNO exhibits significant numerical artifacts beyond $t = 32.5$ ps, rendering its predictions unreliable for further analysis. This observation reinforces our earlier conclusion that FNO struggles to learn the dynamics of nonlinear reactions occurring within localized high-temperature regions. This comparison further highlights that PARCv2 not only achieves better accuracy than popular state-of-the-art PIML models but also generalizes more effectively to out-of-distribution initial conditions.

3. *Low Velocity Extrapolation*

In Section III B 1, we presented our finding that all three models exhibit deteriorated accuracy in lower-velocity cases. We now further investigate this failure mode when extrapolating beyond the coverage of the training set toward smaller velocities.

We first observe that, while the RMSE indicates some decrease in performance for extrapolation to lower initial velocities—specifically, the error at $V_0 = 500$ m/s is approximately twice the lowest error achieved by the models, as shown in the top panel of FIG. 3—a more concerning issue arises when average temperature is considered. In this case, a nearly tenfold increase in relative error is observed. Notably, for the temperature, pressure, microstructure, and Y-velocity fields, the increase in RMSE during low-velocity extrapolation is much more pronounced for FNO (often exceeding a tenfold increase) compared to PARCv2 (typically around a twofold increase). Visual inspection of the predicted fields (e.g. FIG. 12 for temperature) reveals several numerical artifacts, including pore deformation, material interface warping, and anomalously high temperatures around the pore boundary.

Despite these issues, our observation of declining model performance at low velocities is consistent with the failure mode reported in Cheng *et al.*³⁶, where PARCv2 was used to model supersonic flow around a cylinder. However, unlike that case, we do not observe numerically diverging values, persistent large-scale artifacts, or evidence that the failure is due to a shift in the underlying physical processes that are dominant in the system.

C. *Physics-driven Metrics*

1. *Distribution Functions of Temperature and Pressure*

Following the approach of Nguyen *et al.*¹², we compared the distribution functions of the temperature and pressure fields at the time of complete pore collapse for each initial velocity, as shown in FIG. 13, to evaluate the performance of PARCv2 on more physics-driven metrics. We observed that the model correctly captures the general trend of increasing peak temperature distribution with higher initial impact velocities. However, several systematic deviations are also apparent. First, while the ground truth distribution displays a clear gap between 0 K (representing vacuum at the top of the physical domain) and 200 K (the initial

temperature of the material block), this gap is less prominent in the temperature distribution predicted by PARCv2. This suggests that the model struggles with discontinuous distributions. Additionally, the model consistently underpredicts the highest temperatures in the domain, as indicated by the blue region in the bottom panel of FIG. 13 not extending as far as in the top panel. A similar trend is observed in the pressure field: although the predicted pressure PDF closely matches the ground truth overall, PARCv2 consistently underpredicts the maximum pressure, with no predicted values exceeding 100 GPa.

2. *Shear Band Temperature Profile*

As the shear band phenomenon is of high interest in this work, we conducted a more focused study of the region exhibiting the strongest shear band signals. For each simulation, we selected the timestep at which shear bands were most prominent and compared the temperature profile along a vertical section at $X = 29.30$ nm and $29.30 < Y < 175.79$ nm for several evenly distributed test cases, as shown in FIG. 14. We found that PARCv2 consistently predicts the location and width of the most prominent shear bands, though the peak temperatures within the bands and their surrounding background are underestimated by 10–40 K. FNO performs well for the 1,800 m/s and 2,400 m/s cases, but it consistently misses more shear bands and exhibits larger RMSE values within the regions of interest.

This comparison confirms four key behaviors observed in our previous visual examinations: (a) PARCv2 outperforms FNO in predicting shear band formation and evolution; (b) all models achieve their best prediction accuracy within the 1,200 m/s to 2,400 m/s range; (c) weaker shear bands are often missed by all models, including PARCv2; and (d) predictions from the ResNet-based neural ODE are unsatisfactory.

We further analyzed the temperature profile within the region of interest for the high-velocity extrapolation test case, as shown in the bottom panel of FIG. 15. Notably, the ground truth sequence indicates a general absence of strong shear band effects in higher-velocity cases. Unlike other cases within the training set, where the ground truth (solid lines) exhibits frequent but weak temperature variations, the $V_0 = 3000$ m/s case features one large, wide drop in temperature. PARCv2 accurately captures this feature, although it underpredicts the peak temperature. Nevertheless, compared to FNO, the PARCv2 predic-

tion yields an RMSE value approximately 50% lower and is free of any nonphysical increases in the temperature profile, such as a peak in temperature at $y \sim 110$ nm in the FNO prediction when $V_0 = 3000$ m/s in the bottom panel of FIG. 15.

A lack of accurately predicted shear bands at low velocity is also evident from the top panel of FIG. 14. While PARCv2 predicts some temperature increase at the expected locations of shear bands, the predicted peak temperatures show the greatest deviation among all cases, and the background continuum is poorly captured. The predicted continuum is consistently lower than the ground truth, peak temperatures in the shear bands are more than 50 K lower, and the predicted bands are significantly wider than those in the ground truth.

3. Dominant Shear Band Location, Width and Temperature Increase

We further examined the prediction quality of the dominant shear band by extracting its location, width, and temperature increase from FIG. 14 and FIG. 15, with results summarized in TAB. II. Models achieving the smallest deviation from ground truth in each physics-driven metric and impact velocity are underscored in the corresponding column. For $V_0 = 1800$ m/s and $V_0 = 2400$ m/s, both PARCv2 and FNO predicted the peak-temperature location with zero error and reproduced the dominant shear band width with only a one-pixel deviation from ground truth. However, in terms of temperature increase relative to the surrounding continuum, both models underpredict compared to ground truth—indicating a systematic tendency to miss higher-frequency variations, a telltale sign of spectral bias. Further inspection of the $V_0 = 2400$ m/s case in the top panel of FIG. 15 shows that FNO (dashed line) consistently underpredicts the temperature near strong shear bands. Although this effect is not captured by the physics-driven metrics in TAB. II, the elevated RMSE values correctly reflect the evident deviation from ground truth.

For lower impact velocity cases, PARCv2 again demonstrates superior accuracy compared to FNO. At $V_0 = 1200$ m/s, PARCv2 outperforms in predicting both the width and temperature increase of the dominant shear band in TAB. II, while trailing slightly behind FNO in the predicted location. However, this simple descriptive metric is misleading: as shown in the middle panel of FIG. 14, the dominant shear band exhibits a double-peak structure that FNO completely misses. Although PARCv2 shows some misalignment in the higher-temperature

V_0 [m/s]		Ground Truth	Improved PARCv2	FNO	Neural ODE
500	y_{peakT} [nm]	154.6908			
	Δy [nm]	15.2348	Fail	Fail	Fail
	ΔT [K]	70.4020			
1200	y_{peakT} [nm]	144.1437	141.7999	<u>142.9718</u>	
	Δy [nm]	18.1644	<u>18.1645</u>	14.6487	Fail
	ΔT [K]	155.1000	<u>111.9133</u>	52.05890	
1800	y_{peakT} [nm]	145.3156	<u>145.3156</u>	<u>145.3156</u>	
	Δy [nm]	11.1330	<u>9.96115</u>	<u>9.96115</u>	Fail
	ΔT [K]	132.975	<u>129.8824</u>	107.6409	
2400	y_{peakT} [nm]	114.8462	<u>114.8462</u>	<u>114.8462</u>	
	Δy [nm]	11.1330	<u>13.4768</u>	<u>13.4768</u>	Fail
	ΔT [K]	140.7800	90.4980	<u>111.3630</u>	

TABLE II: Dominant shear band location y_{peakT} , width Δy and temperature increase ΔT at $X = 29.30$ nm for a number of impact velocities. Model with the prediction closest to ground truth are underline for each metrics, and models failed to predict shear band patterns are labeled with “Fail” in the corresponding column. $V_0 = 3000$ m/s was skipped due to the lack of shear band phenomenon.

peak, it correctly captures the existence of the double-peak structure. Consistent with this observation, the elevated RMSE values for FNO reflect its larger deviation from ground truth.

Similar to previous cases, the $V_0 = 500$ m/s impact velocity proves challenging for all three models, as none of the predictions exhibit a clear shear-band pattern. Consequently, all three models are labeled as Failed in TAB. II. The $V_0 = 3000$ m/s case cannot be assessed using these metrics, since the shear-band patterns are statistically indistinguishable from temperature oscillations present in the continuum, and is therefore omitted from the table.

IV. DISCUSSION

While PARCv2 demonstrates the best overall performance among popular PIML models and achieves reasonable accuracy for many dynamical processes — including the formation and evolution of dominant shear bands and hotspots, shocks generated by the reverse ballistic impacts, and distortion of material interfaces — we also observed several limitations of current PIML models, which we address in more detail below.

A. Spectral Bias

As observed in previous sections, the predicted fields display diffused shock fronts and shear bands, leading the model to miss some of the smaller and weaker shear bands. This limitation can be attributed to the low-frequency bias of neural networks (e.g., Xu, Zhang, and Luo⁴⁷). In our problem, this bias manifests in the model learning large-scale patterns (e.g., regions of high temperature, shock waves, reverse ballistics) earlier in training, while smaller-scale patterns (e.g., shear bands, shape and asymmetry of the high-temperature regions) are learned later. We examined our model at various training epochs and confirmed this hypothesis in FIG. 16. The evolution of the material interface and the motion of the high-temperature region were learned within the first 1,500 epochs. The ignition of the high-temperature region began to appear with reasonable detail around epoch 1,750, and the shear band patterns became predictable after 2,000 training epochs.

To further investigate the spectral bias of PARCv2, we present the relative error of the temperature spectrum as a function of time in FIG. 17. As previously discussed, blurry predictions and missing weaker shear bands persist across impact velocities; therefore, we focus on the case with $V_0 = 1800$ m/s. At the start of the roll-out, shown in the first two panels of FIG. 17, the relative prediction error remains approximately constant across spatial scales. However, at $t = 27.5$ ps, during pore collapse, a sharp increase in error appears at the high-frequency end (corresponding to smaller spatial scales), as seen in the third panel. When shear band patterns become most prominent at $t = 40$ ps, the deviation for temperature structures smaller than $10nm$ rapidly increase, reaching 50% at the smallest spatial scale resolvable by the simulation. From this snapshot alone, it is unclear whether the elevated high-frequency error originates from inaccuracies in predicting weaker, non-

dominant shear bands or from the fine-scale hotspot geometry, as both features exist below the $10nm$ scale. Examination of the final snapshot, however, rules out the hotspot geometry as the primary source: the hotspot shape remains largely unchanged over time, while the error drops significantly as shear-band-induced temperature rises dissipate. Together, these results provide evidence that the misprediction of weaker shear bands is primarily due to the spectral bias of the machine learning model.

To address this issue and improve prediction accuracy, we turned to progress made in the computer vision community. It has been suggested that using outputs from shallow layers of a pretrained network in image synthesis and super-resolution tasks can significantly enhance image sharpness⁴⁸. Given the distinct nature of image classification and physics simulation, we first examined features extracted by an ImageNet⁴⁹ pretrained VGG-16 classifier⁵⁰ to ensure that patterns of interest could be identified using this pretrained model. We selected the *relu1_2* layer features for use as additional perceptual loss terms, since shear band patterns were clearly extracted at this specific network depth. Some shear band structures appeared in the extracted features from this layer, as shown in FIG. 18. We then fine-tuned our network with this added perceptual loss term for an additional 1,000 epochs using a learning rate of 10^{-5} . While we observed a non-trivial decrease in both data and perceptual loss, we did not see noticeable improvements in the predictions. Visual inspection of the rollout sequence revealed no significant changes, even when comparing the VGG-extracted features in the second and third rows of FIG. 18. Additionally, improvements in temperature prediction for shear bands in FIG. 19 were minimal. In terms of mean absolute difference from the ground truth (indicated in each panel), we found that the fine-tuned model actually performed worse in both high- and low-speed extrapolation cases.

One major drawback of the VGG perceptual loss is the increased training time, as it requires backpropagation through part of another deep learning model. Given that our model already demands substantial computational effort to train, we explored an alternative approach by using discrete wavelet transforms to extract high-frequency features, which we then used as an additional perceptual loss term. We selected the Haar wavelet and constructed the wavelet perceptual loss from the highest-frequency wavelet coefficients. We repeated the fine-tuning process described above, substituting the VGG perceptual loss with the wavelet perceptual loss. However, we again observed no improvement in the shear band

temperature profiles.

Concerned that the added loss term may have drastically altered the loss landscape, making the previous L1-only trained model poorly situated in the new space, we retrained the model from scratch using the wavelet perceptual loss. Although this did not significantly improve prediction performance, we observed a drastically faster convergence, with the model producing usable results within just 900 epochs. We therefore conclude that while perceptual loss does not enhance the sharpness of predictions in our problem, its emphasis on high-frequency components is nonetheless beneficial for accelerating training. We hypothesize that this acceleration could extend to many other mesh-based PIML models across a range of physical systems.

Ng, Wang, and Lai⁵¹ proposed that using a generalized Lp norm with large p values could improve learning of high-frequency components and increase predictive accuracy. In theory, reducing per-step error should yield improved rollout performance. To test this, we fine-tuned PARCv2 using $p = 10$ for 1000 epochs with a constant learning rate of 10^{-6} . However, our results show no benefit for our problem. Despite a significant decrease in Lp loss values on both training and validation sets, the rollout predictions exhibited degraded performance, including numerical artifacts and signs of instability. Many of these artifacts originated from dragging of the circular pore and horizontal banding in regions of constant field values. Compared to the much smoother field and stationary setup in Ng, Wang, and Lai⁵¹, we conclude that this method is less effective for time-recursive PIML models dealing with complex, dynamic physics.

Our investigation into improving prediction sharpness highlights that, in the context of physics-informed machine learning, addressing low-frequency bias remains an open challenge. Although several techniques have been proposed for deep neural networks—such as phase shifting⁵², Fourier features⁵³, and multi-stage networks with sinusoidal activation functions⁵⁴—these methods all explicitly rely on spatial coordinates. This dependency can compromise shift invariance and limit spatial extrapolation capabilities. Moreover, it has been shown that emphasizing high-frequency features can impair a model’s generalization ability despite reducing prediction error⁵⁵. While DNNs are not generally expected to generalize across different initial conditions, this loss of generalizability (a) undermines a key strength of mesh-based PIML models like ours, which can infer across a range of initial con-

ditions with reasonable accuracy, and (b) introduces the risk of numerical instability due to accumulated errors in temporally recursive models.

We hypothesize that addressing the sharpness issue effectively will require entirely new neural network architectures. Most modern architectures rely on hierarchical information summarization: shallow layers learn local features, which are then aggregated by deeper layers to capture global dynamics. We believe this structure inherently deprioritizes the accurate modeling of fine-scale features, as training gradients initially flow to layers responsible for large-scale dynamics. A promising solution may lie in designing architectures that can learn both large and small scales at the same network depth. Such a design could help resolve the issue of missing smaller and weaker shear bands and enhance the overall contrast of predicted fields.

B. Enforcing Boundary Conditions for Mesh-based PIML Models

While numerical solvers can handle all types of boundary conditions with minimal user intervention, this is not the case in AI/ML based simulations for physics applications. Physics-naïve models generally apply no special treatment to boundary conditions—unsurprising, given that such considerations are of limited importance in the computer vision community. In the realm of physics-informed machine learning (PIML), however, enforcing boundary conditions depends heavily on model architecture, and a variety of methods have been proposed.

The most straightforward approach involves placing collocation points along the boundary and applying a soft constraint via additional loss terms. This method is the most commonly used for meshless PIML models (e.g., Raissi, Perdikaris, and Karniadakis¹⁷, Berg and Nyström⁵⁶, Sirignano and Spiliopoulos⁵⁷). Other strategies include using variational principles^{58,59} or Petrov-Galerkin domain decomposition^{60,61} to impose boundary conditions indirectly. However, multiple studies (e.g., Lu *et al.*⁶², Chen, Du, and Wu⁶³, Wang, Teng, and Perdikaris⁶⁴) have shown that the inexact nature of these methods can hinder convergence and reduce prediction accuracy, especially in problems with complex geometries. To address this, recent efforts have focused on exactly enforcing boundary conditions (e.g., Lyu *et al.*⁶⁵, Sukumar and Srivastava⁶⁶). Unfortunately, these exact methods often require

explicit coordinate dependencies, making them difficult to translate to mesh-based models such as convolutional or graph neural networks.

For mesh-based PIML models, custom solutions have emerged. Ren *et al.*⁶⁷ introduced a practical and user-configurable strategy for enforcing boundary conditions using built-in padding modes from popular deep learning frameworks. In their method, Dirichlet boundary conditions (e.g., zero pressure at an outlet) are enforced via constant padding, Neumann conditions (e.g., zero velocity gradient) via reflective padding, and periodic boundaries via circular padding. We adopted this approach in our model. For graph neural networks, Pfaff *et al.*⁶⁸ proposed marking boundary nodes with a non-updating attribute, effectively freezing their values to enforce Dirichlet conditions. However, extending this to Neumann boundaries is significantly more difficult.

In our work, we apply zero-gradient (Neumann) boundary conditions across all fields following Ren *et al.*⁶⁷. However, we encountered several practical caveats. While the method is easy to implement, most deep learning libraries apply the same padding rule to all boundaries. Enforcing distinct boundary conditions (e.g., fixed inlet velocity and zero-gradient outlet) requires low-level customization of padding before each convolution. Moreover, constant padding for Dirichlet boundaries is straightforward at the input layer but becomes problematic in deeper layers. For instance, padding the inlet with a fixed velocity value works at the first layer, but the corresponding values after subsequent layers are no longer the original inlet velocity — they are transformed by layer weights. This implies that the padded values must be updated dynamically after each weight update to remain consistent, which is not supported natively by any existing deep learning code frameworks.

Enforcing Neumann boundary conditions also poses challenges. Reflect padding enforces zero gradients only under symmetric schemes like central differencing. For asymmetric schemes such as upwinding, zero-gradient conditions instead require constant padding. However, because the specific behavior of each convolution filter is unknown ahead of time, modifying padding based on filter orientation becomes infeasible. Therefore, robust and generalizable methods for *hard* enforcement of boundary conditions in mesh-based models remain an important area for further research and development.

C. Prediction of Low Impact Velocity/ Low shock strength cases

In Section IIIB 3, we highlighted a puzzling behavior of our deep learning model: extrapolations to lower initial velocities result in significantly higher loss values and a failure to capture the dynamics. This is not the first observation of such behavior. While similar trends have been reported in Cheng *et al.*³⁶ — where reduced accuracy at lower velocities was attributed to the complex transition from subsonic to supersonic wake flow — no such change in dynamics is present in our problem. In fact, Weger *et al.*⁶⁹ showed that a power law can describe the spacing of dominant shear bands across the entire initial impact velocity range, suggesting that the underlying physical processes remain consistent.

We note that lower initial impact velocities correspond to smaller magnitudes of physical quantities (e.g., temperature, pressure, velocity), which we hypothesize is a key contributor to the model’s degraded performance. As discussed in FIG. 16, smaller variations in data are typically learned at later stages of training. While such deviations are negligible at higher impact velocities, they can be comparable in scale to the temperature contrast between shear bands and their surroundings at lower velocities. Consequently, the model’s inability to resolve these small differences may lead to an overall failure to learn the dynamics in these cases.

Although the universal approximation theorem assures that deep learning models can represent any function in theory, in practice, prediction errors often plateau even with larger model sizes and prolonged training, as shown in Wang and Lai⁵⁴. This can be attributed to the bias–variance tradeoff, which limits the model’s generalization capability beyond its training distribution. We believe that normalizing data through a learnable normalization flow, parameterized by initial conditions or physical time, may help mitigate this failure mode. We are currently exploring such methods.

A second likely cause is the distribution shift associated with different initial conditions. As with any machine learning application, model performance depends critically on the similarity between the training and inference data distributions. While this assumption typically holds in computer vision tasks, it is violated in the modeling of shear band formation. Different impact velocities produce vastly different distributions in temperature, pressure, and velocity. For example, maximum temperatures exceed 10,000 K at high velocities, while

they barely reach 500 K at the lowest. This wide variation adds further complexity and exacerbates the learning challenge.

Finally, the mathematical structure of our model — based on learning time derivatives — may further worsen the issue. Lower velocity impact cases exhibit slower dynamics, which reduces the signal-to-noise ratio for learning time evolution and makes the model more susceptible to this specific failure mode.

V. CONCLUSION

An improved implementation of physics-aware recurrent convolution algorithm called PARCv2 was employed to predict the temperature, pressure and interfacial dynamics in a shock-induced pore collapse problem. The physics of this problem is relevant to hotspot formation and growth in energetic materials, which is the initiating feature leading from shock to detonation. We show that the PARCv2 algorithm provides superior predictions when compared to two other popular PIML approaches, namely FNO and neural ODE. When tested for cases within the range of the training dataset PARCv2 consistently provides predictions in close agreement with ground truth.

The improved PARCv2 model achieves state-of-the-art performance in modeling pore collapse in solid materials under weak-to-moderate shock regimes, capturing most physical processes with satisfactory accuracy across a range of initial conditions. However, persistent issues remain — namely, blurred predictions, missing weaker shear bands, and degraded performance at lower impact velocities — which affect all PIML models tested. Although the computer vision community has made significant progress on similar challenges, our attempts to incorporate their techniques — such as perceptual losses and high-frequency feature learning — into physics-informed machine learning architectures and training strategies yielded only limited improvements. These methods failed to fully resolve the observed failure modes.

We conclude that such limitations can hinder the broader adoption of deep learning models in complex computational fluid dynamics and solid mechanics applications, despite the substantial reductions in simulation time and computational cost they offer. Deeper investigation into these failure modes is essential for advancing PIML model development

and will have broad impact on the computational physics community.

ACKNOWLEDGMENTS

This work was supported by the Army Research Office (ARO) Energetics Basic Research Center (EBRC) program under Grant No. W911NF-22-2-0164.

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- ¹J. E. Field, N. Bourne, S. Palmer, and S. Walley, Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences **339**, 269 (1992).
- ²C. Handley, B. Lambourn, N. Whitworth, H. James, and W. Belfield, Applied Physics Reviews **5** (2018).
- ³J. E. Field, Accounts of chemical Research **25**, 489 (1992).
- ⁴C. M. Tarver, S. K. Chidester, and A. L. Nichols, The Journal of Physical Chemistry **100**, 5794 (1996).
- ⁵H. Springer, C. Tarver, and S. Bastea, in *AIP Conference Proceedings*, Vol. 1793 (AIP Publishing, 2017).
- ⁶H. K. Springer, S. Bastea, A. L. Nichols III, C. M. Tarver, and J. E. Reaugh, Propellants, Explosives, Pyrotechnics **43**, 805 (2018).
- ⁷N. Rai and H. Udaykumar, Physics of Fluids **31** (2019).
- ⁸C. Coffey, Le Journal de Physique Colloques **48**, C4 (1987).
- ⁹B. W. Hamilton and T. C. Germann, npj Computational Materials **10**, 147 (2024).

- ¹⁰A. Pereverzev and T. D. Sewell, Crystals **10**, 1120 (2020).
- ¹¹T. D. Sewell, R. Menikoff, D. Bedrov, and G. D. Smith, The Journal of chemical physics **119**, 7417 (2003).
- ¹²Y. T. Nguyen, C. Okafor, P. Zhao, O. Sen, C. R. Picu, T. Sewell, and H. Udaykumar, Journal of applied physics **136** (2024).
- ¹³J. Herrin, G. Tow, J. Brennan, J. Larentzos, C. R. Picu, and H. Udaykumar, Journal of applied physics **136** (2024).
- ¹⁴R. Chitsazi, M. P. Kroonblawd, A. Pereverzev, and T. Sewell, Modelling and Simulation in Materials Science and Engineering **28**, 025008 (2020).
- ¹⁵J. B. Choi, P. C. Nguyen, O. Sen, H. Udaykumar, and S. Baek, Propellants, Explosives, Pyrotechnics **48**, e202200276 (2023).
- ¹⁶G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Nature Reviews Physics **3**, 422 (2021).
- ¹⁷M. Raissi, P. Perdikaris, and G. E. Karniadakis, Journal of Computational physics **378**, 686 (2019).
- ¹⁸A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, Journal of machine learning research **18**, 1 (2018).
- ¹⁹N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar, Journal of Machine Learning Research **24**, 1 (2023).
- ²⁰Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, arXiv preprint arXiv:2010.08895 (2020).
- ²¹Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar, ACM/JMS Journal of Data Science **1**, 1 (2024).
- ²²R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, Advances in neural information processing systems **31** (2018).
- ²³W. Grathwohl, R. T. Q. Chen, J. Bettencourt, I. Sutskever, and D. Duvenaud, International Conference on Learning Representations (2019).
- ²⁴S. Du, Y. Luo, W. Chen, J. Xu, and D. Zeng, in *proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2022) pp. 12570–12580.
- ²⁵E. De Brouwer, J. Simm, A. Arany, and Y. Moreau, Advances in neural information processing systems **32** (2019).

- ²⁶P. C. Nguyen, J. B. Choi, H. Udaykumar, and S. Baek, *Journal of Computing and Information Science in Engineering* **23**, 060808 (2023).
- ²⁷H. Udaykumar, S. Wurster, D. Tomaschewski, and S. Baek, *Propellants, Explosives, Pyrotechnics* **49**, e202400300 (2024).
- ²⁸S. Chun, S. Roy, Y. T. Nguyen, J. B. Choi, H. S. Udaykumar, and S. S. Baek, *Scientific reports* **10**, 13307 (2020).
- ²⁹P. C. Nguyen, N. N. Vlassis, B. Bahmani, W. Sun, H. Udaykumar, and S. S. Baek, *Scientific reports* **12**, 9034 (2022).
- ³⁰S. Roy, Y. T. Nguyen, C. Neal, S. Baek, and H. Udaykumar, *Journal of Applied Physics* **131** (2022).
- ³¹I. Fang, S. Roy, P. Nguyen, S. Baek, and H. Udaykumar, *APL Machine Learning* **3** (2025).
- ³²P. C. H. Nguyen, Y.-T. Nguyen, P. K. Seshadri, J. B. Choi, H. Udaykumar, and S. Baek, *Propellants, Explosives, Pyrotechnics* **48**, e202200268 (2023).
- ³³P. C. Nguyen, Y.-T. Nguyen, J. B. Choi, P. K. Seshadri, H. Udaykumar, and S. S. Baek, *Science advances* **9**, eadd6868 (2023).
- ³⁴P. C. Nguyen, X. Cheng, S. Azarfar, P. Seshadri, Y. T. Nguyen, M. Kim, S. Choi, H. Udaykumar, and S. Baek, in *Proceedings of the 41st International Conference on Machine Learning* (2024) pp. 37649–37666.
- ³⁵N. K. Rai and H. Udaykumar, *Physical Review Fluids* **3**, 033201 (2018).
- ³⁶X. Cheng, P. C. Nguyen, P. K. Seshadri, M. Verma, Z. J. Gray, J. T. Beerman, H. Udaykumar, and S. S. Baek, *International Journal of Multiphase Flow* , 104877 (2024).
- ³⁷O. Ronneberger, P. Fischer, and T. Brox, in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (Springer, 2015) pp. 234–241.
- ³⁸C. Runge, *Mathematische Annalen* **46**, 167 (1895).
- ³⁹W. Kutta, *Beitrag zur näherungsweise Integration totaler Differentialgleichungen* (Teubner, 1901).
- ⁴⁰D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” (2017), arXiv:1412.6980 [cs.LG].
- ⁴¹J. Ansel, E. Yang, H. He, N. Gimelshein, A. Jain, M. Voznesensky, B. Bao, P. Bell, D. Berard, E. Burovski, G. Chauhan, A. Chourdia, W. Constable, A. Desmaison, Z. De-

- Vito, E. Ellison, W. Feng, J. Gong, M. Gschwind, B. Hirsh, S. Huang, K. Kalambarkar, L. Kirsch, M. Lazos, M. Lezcano, Y. Liang, J. Liang, Y. Lu, C. K. Luk, B. Maher, Y. Pan, C. Puhersch, M. Reso, M. Saroufim, M. Y. Siraichi, H. Suk, S. Zhang, M. Suo, P. Tillet, X. Zhao, E. Wang, K. Zhou, R. Zou, X. Wang, A. Mathews, W. Wen, G. Chanan, P. Wu, and S. Chintala, in *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, ASPLOS '24 (Association for Computing Machinery, New York, NY, USA, 2024) p. 929–947.
- ⁴²K. Qi and J. Sun, *Computers & Fluids* **274**, 106239 (2024).
- ⁴³A. Tran, A. Mathews, L. Xie, and C. S. Ong, arXiv preprint arXiv:2111.13802 (2021).
- ⁴⁴K. Zhang, Y. Zuo, H. Zhao, X. Ma, J. Gu, J. Wang, Y. Yang, C. Yao, and J. Yao, *Spe Journal* **27**, 1815 (2022).
- ⁴⁵S. Qin, F. Lyu, W. Peng, D. Geng, J. Wang, N. Gao, X. Liu, and L. L. Wang, arXiv e-prints , arXiv (2024).
- ⁴⁶K. He, X. Zhang, S. Ren, and J. Sun, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 770–778.
- ⁴⁷Z.-Q. J. Xu, Y. Zhang, and T. Luo, *Communications on Applied Mathematics and Computation* , 1 (2024).
- ⁴⁸J. Johnson, A. Alahi, and L. Fei-Fei, in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14* (Springer, 2016) pp. 694–711.
- ⁴⁹J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, in *2009 IEEE conference on computer vision and pattern recognition* (Ieee, 2009) pp. 248–255.
- ⁵⁰K. Simonyan and A. Zisserman, arXiv preprint arXiv:1409.1556 (2014).
- ⁵¹J. Ng, Y. Wang, and C.-Y. Lai, in *ICML 2024 AI for Science Workshop* (2024).
- ⁵²W. Cai, X. Li, and L. Liu, *SIAM Journal on Scientific Computing* **42**, A3285 (2020).
- ⁵³M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, *Advances in neural information processing systems* **33**, 7537 (2020).
- ⁵⁴Y. Wang and C.-Y. Lai, *Journal of Computational Physics* **504**, 112865 (2024).
- ⁵⁵A. A. Abello, R. Hirata, and Z. Wang, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021) pp. 863–871.

- ⁵⁶J. Berg and K. Nyström, *Neurocomputing* **317**, 28 (2018).
- ⁵⁷J. Sirignano and K. Spiliopoulos, *Journal of computational physics* **375**, 1339 (2018).
- ⁵⁸B. Yu *et al.*, *Communications in Mathematics and Statistics* **6**, 1 (2018).
- ⁵⁹J. Han, A. Jentzen, and W. E, *Proceedings of the National Academy of Sciences* **115**, 8505 (2018).
- ⁶⁰E. Kharazmi, Z. Zhang, and G. Karniadakis, arXiv preprint arXiv:1912.00873 (2019).
- ⁶¹E. Kharazmi, Z. Zhang, and G. E. Karniadakis, *Computer Methods in Applied Mechanics and Engineering* **374**, 113547 (2021).
- ⁶²L. Lu, X. Meng, Z. Mao, and G. E. Karniadakis, *SIAM review* **63**, 208 (2021).
- ⁶³J. Chen, R. Du, and K. Wu, arXiv preprint arXiv:2005.04554 (2020).
- ⁶⁴S. Wang, Y. Teng, and P. Perdikaris, arXiv preprint arXiv:2001.04536 (2020).
- ⁶⁵L. Lyu, K. Wu, R. Du, and J. Chen, arXiv preprint arXiv:2008.01491 (2020).
- ⁶⁶N. Sukumar and A. Srivastava, *Computer Methods in Applied Mechanics and Engineering* **389**, 114333 (2022).
- ⁶⁷P. Ren, C. Rao, Y. Liu, J.-X. Wang, and H. Sun, *Computer Methods in Applied Mechanics and Engineering* **389**, 114399 (2022).
- ⁶⁸T. Pfaff, M. Fortunato, A. Sanchez-Gonzalez, and P. Battaglia, in *International Conference on Learning Representations* (2020).
- ⁶⁹L. Weger, J. Herrin, C. Okafor, J. Larentzos, J. Brennan, T. Sewell, C. R. Picu, and H. S. Udaykumar, *Journal of Applied Physics* **137**, 085901 (2025).
- ⁷⁰N. K. Rai, *Numerical Framework for Mesoscale Simulation of Heterogeneous Energetic Materials*, Ph.d. dissertation, University of Iowa (2015).
- ⁷¹N. K. Rai and H. S. Udaykumar, *Journal of Applied Physics* **118**, 245901 (2015).
- ⁷²A. S. Khan and S. Huang, *Continuum Theory of Plasticity* (John Wiley & Sons, New York, 1995).
- ⁷³R. Huang and Y. Li, *Materials Today Communications* **26**, 102086 (2021).
- ⁷⁴S. Sambasivan, A. Kapahi, and H. S. Udaykumar, *Journal of Computational Physics* **235**, 334 (2013).
- ⁷⁵J.-P. Ponthot, *International Journal of Plasticity* **18**, 91 (2002).
- ⁷⁶M. L. Wilkins, “Calculation of elastic-plastic flow,” Tech. Rep. (Lawrence Radiation Laboratory, 1963).

- ⁷⁷Y. B. Zel'dovich and Y. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, reprint ed. (Dover Publications, Mineola, NY, 2002) p. 916.
- ⁷⁸R. C. Batra, *Elements of Continuum Mechanics*, AIAA Education Series, Vol. 17 (American Institute of Aeronautics and Astronautics, Reston, VA, 2006) p. 325.
- ⁷⁹L. B. Munday, P. W. Chung, B. M. Rice, and S. D. Solares, The Journal of Physical Chemistry B **115**, 4378 (2011).
- ⁸⁰O. Sen, P. K. Seshadri, N. K. Rai, J. Larentzos, J. Brennan, T. Sewell, C. R. Picu, and H. S. Udaykumar, Journal of Applied Physics **135**, 145901 (2024).
- ⁸¹S. Izvekov and B. M. Rice, The Journal of Chemical Physics **155** (2021).
- ⁸²M. J. Cawkwell, D. J. Luscher, F. L. Addessio, and K. J. Ramos, Journal of Applied Physics **119**, 185106 (2016).
- ⁸³J. J. Dick, D. E. Hooks, R. Menikoff, and A. R. Martinez, Journal of Applied Physics **96**, 374 (2004).
- ⁸⁴M. P. Kroonblawd and H. K. Springer, Propellants, Explosives, Pyrotechnics **47**, e202100328 (2022).
- ⁸⁵D. C. Sorescu, J. P. Larentzos, B. M. Rice, and J. K. Brennan, Propellants, Explosives, Pyrotechnics **47**, e202100338 (2022).
- ⁸⁶R. Perriot, M. S. Powell, J. D. Lazarz, C. Bolme, S. D. McGrane, D. S. Moore, M. Cawkwell, and K. J. Ramos, arXiv preprint arXiv:2103.11950 (2021).

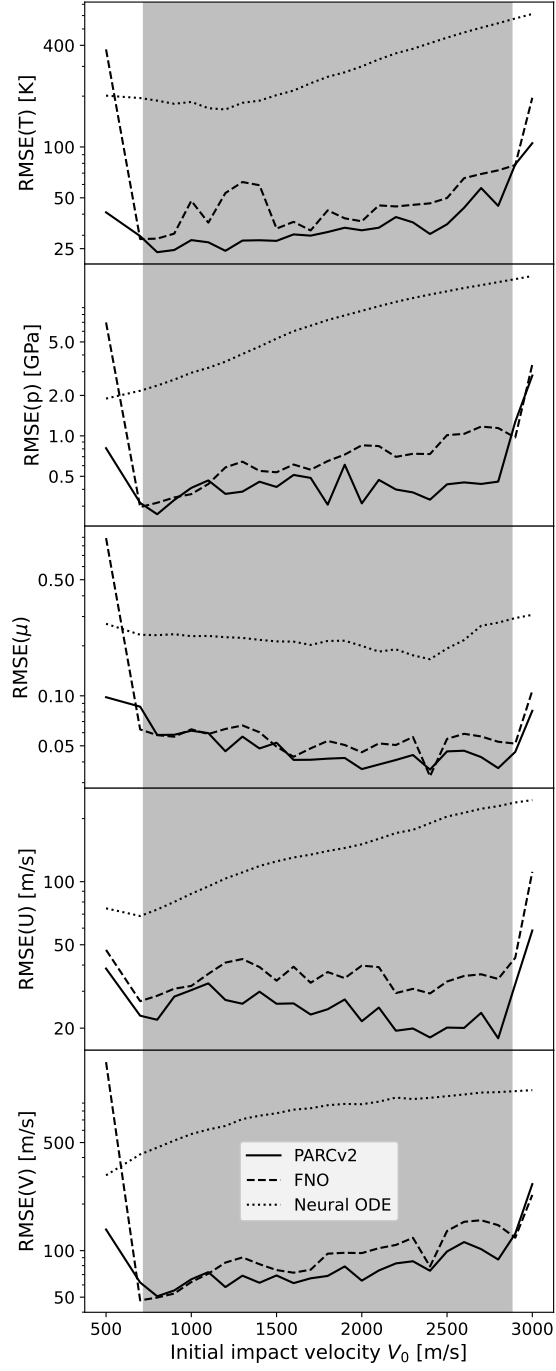


FIG. 3: Initial impact velocity versus root mean squared error (RMSE) of model predictions. Shaded areas indicate coverage of training set.

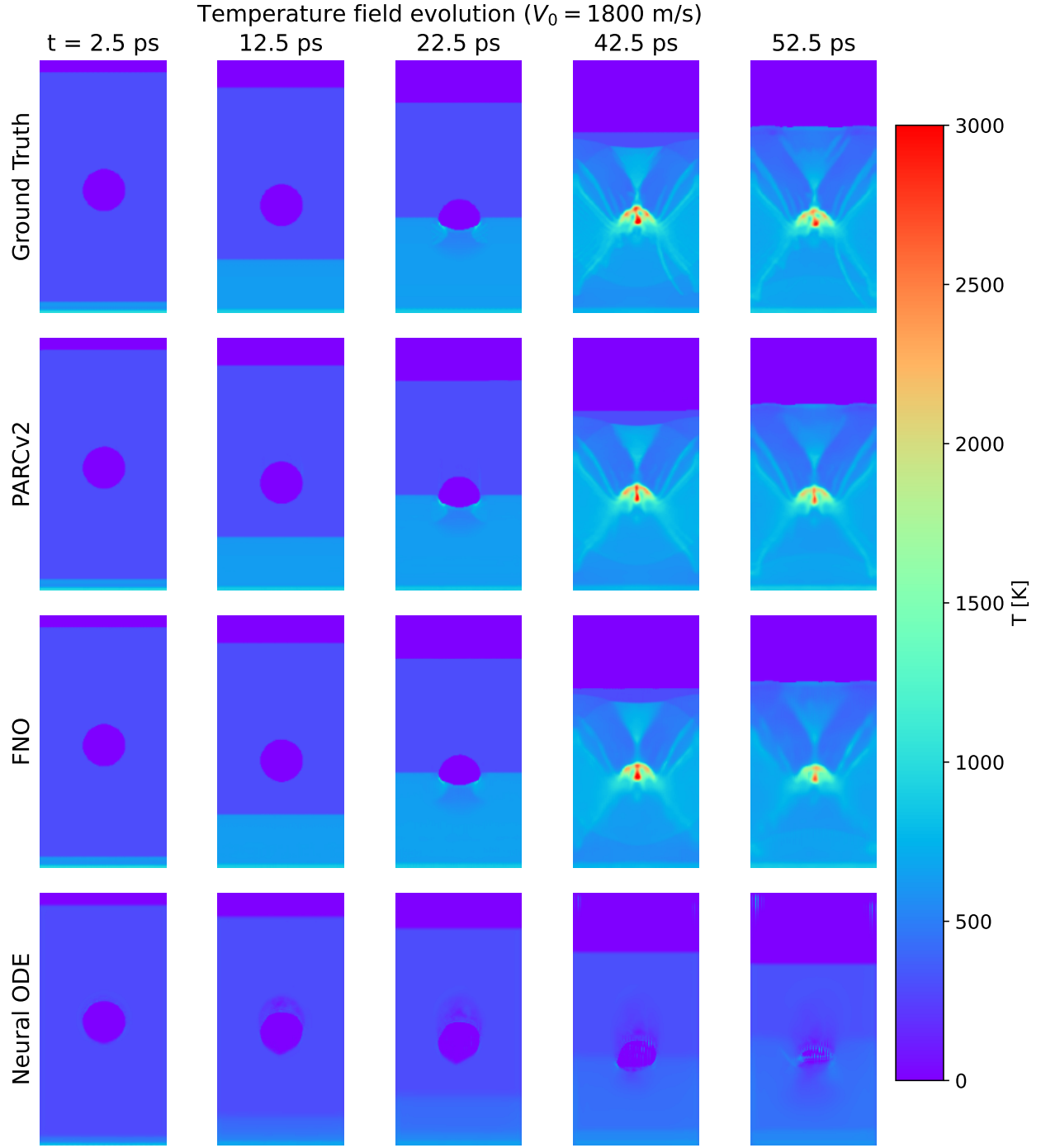


FIG. 4: Temperature field evolution of initial velocity of 1800 m/s.

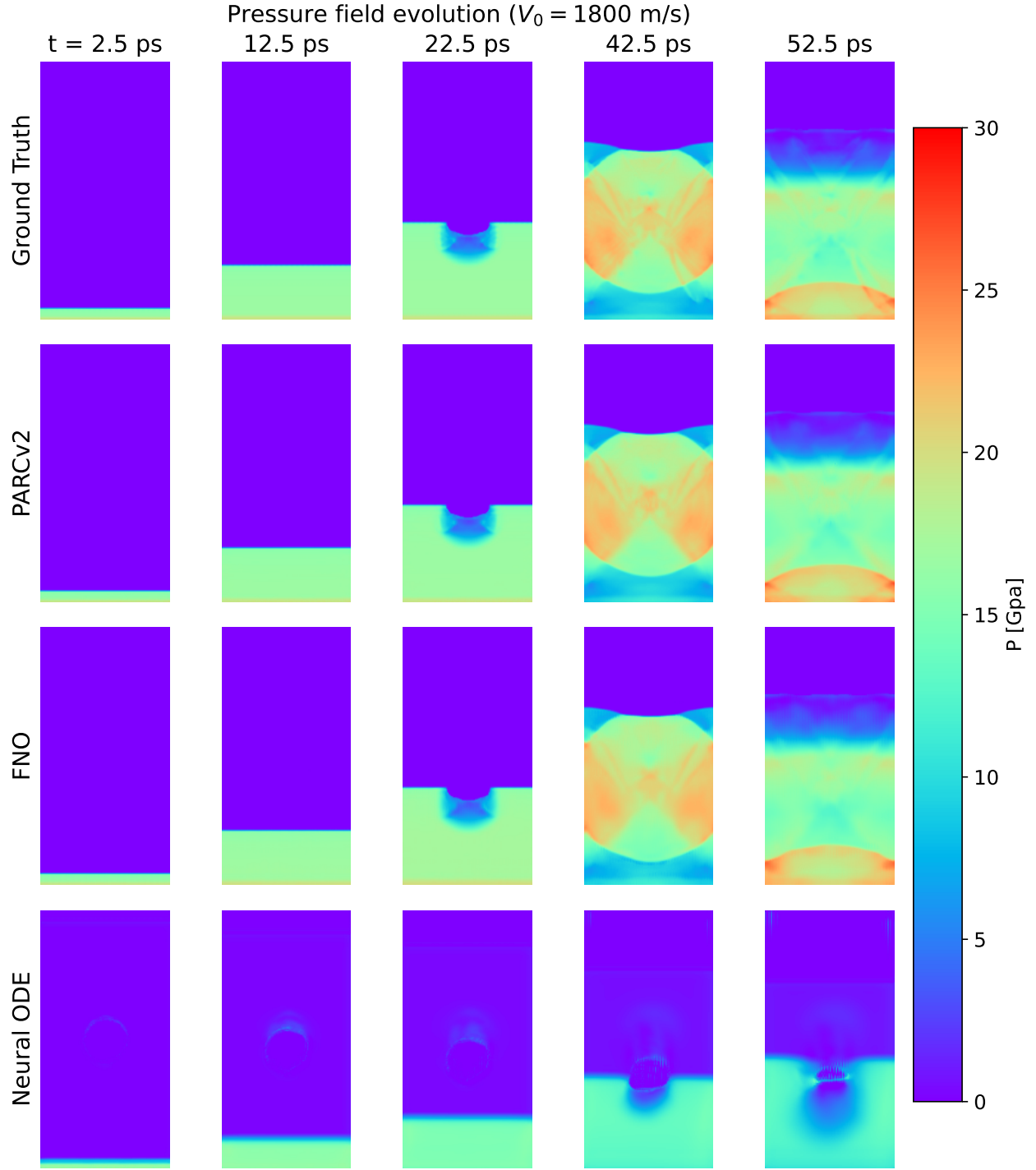


FIG. 5: Pressure field evolution of initial velocity of 1800 m/s.

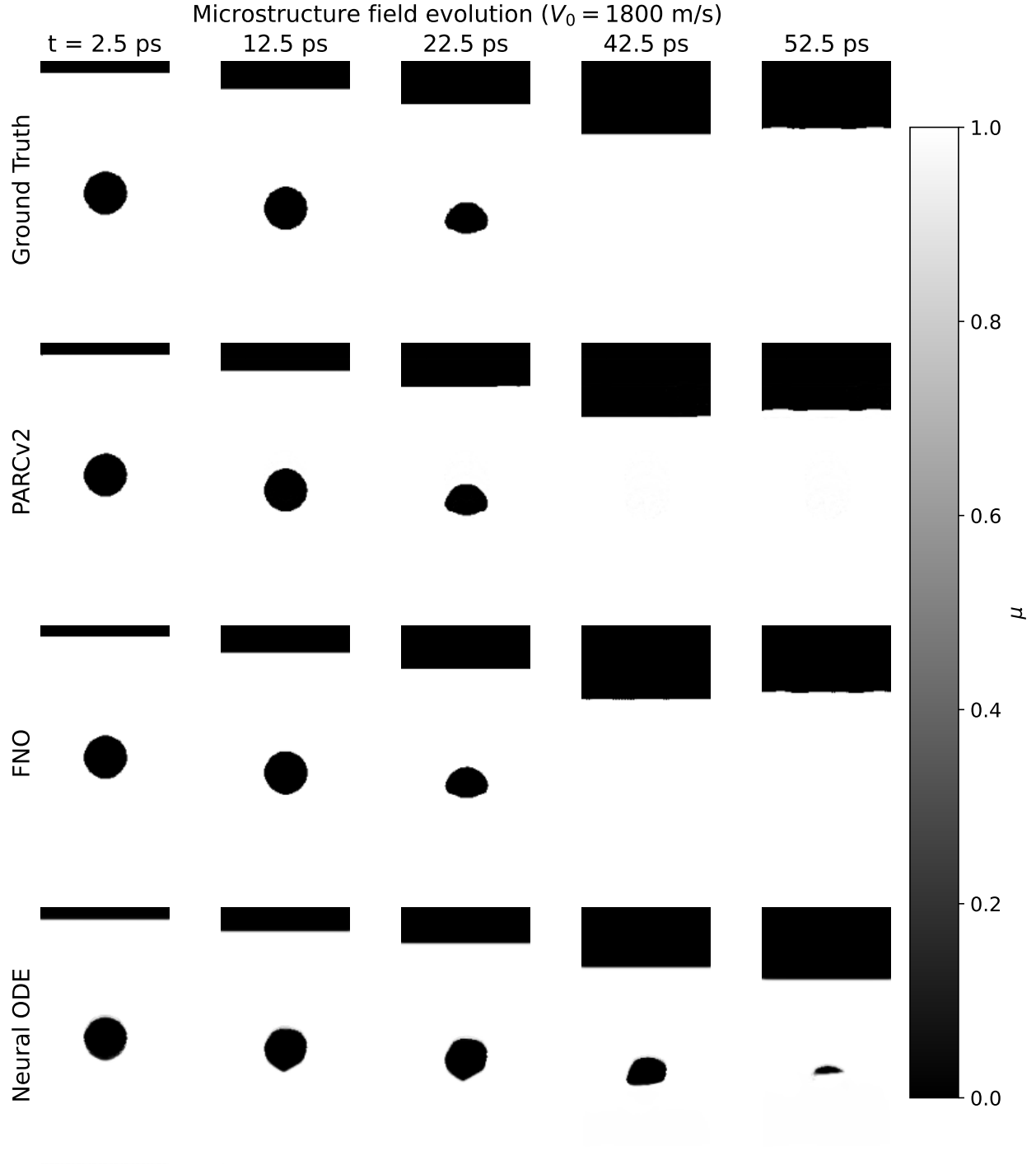


FIG. 6: Microstructure field evolution of initial velocity of 1800 m/s.

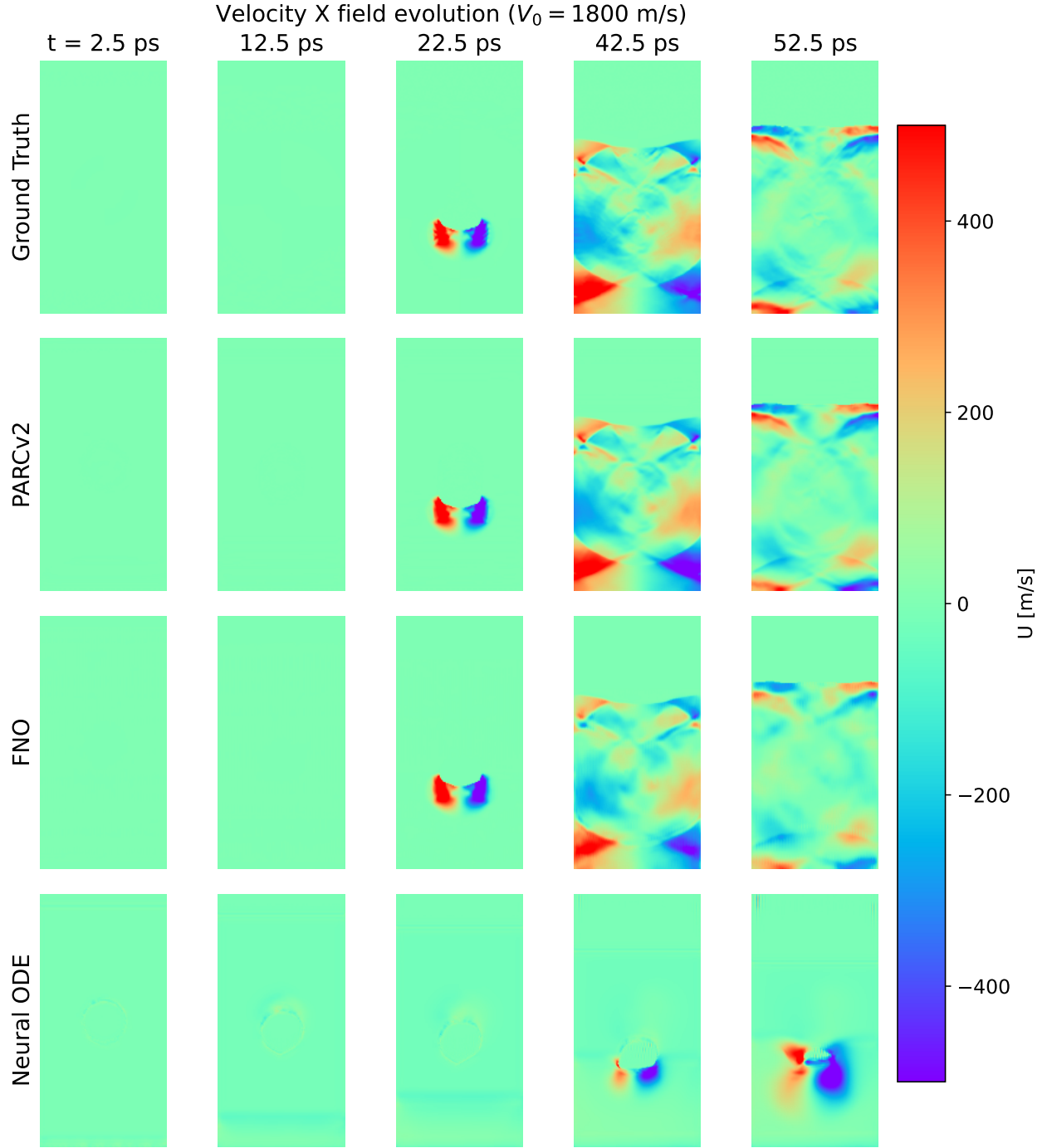


FIG. 7: X-direction velocity field evolution of initial velocity of 1800 m/s.

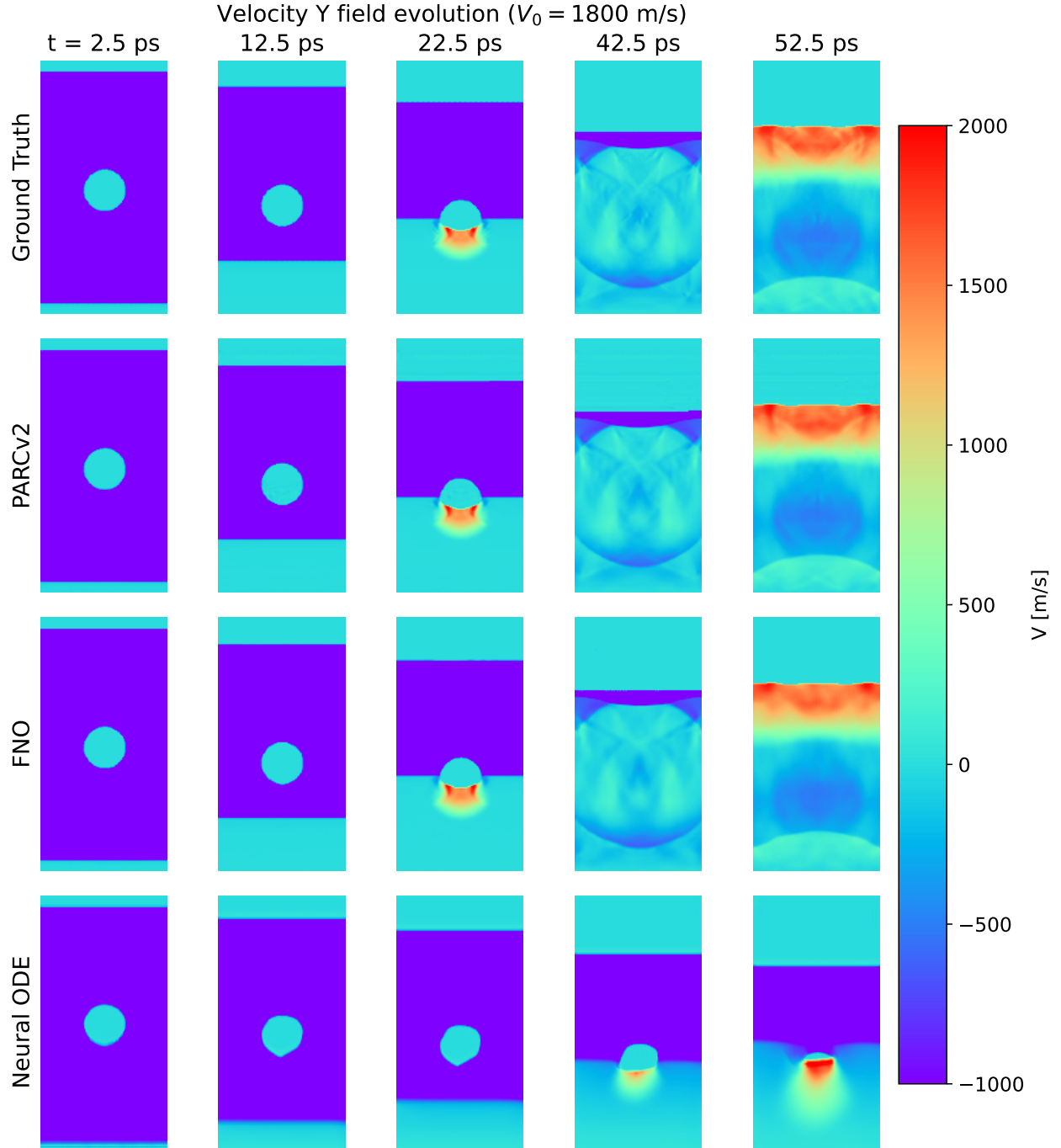


FIG. 8: Y-direction velocity field evolution of initial velocity of 1800 m/s.

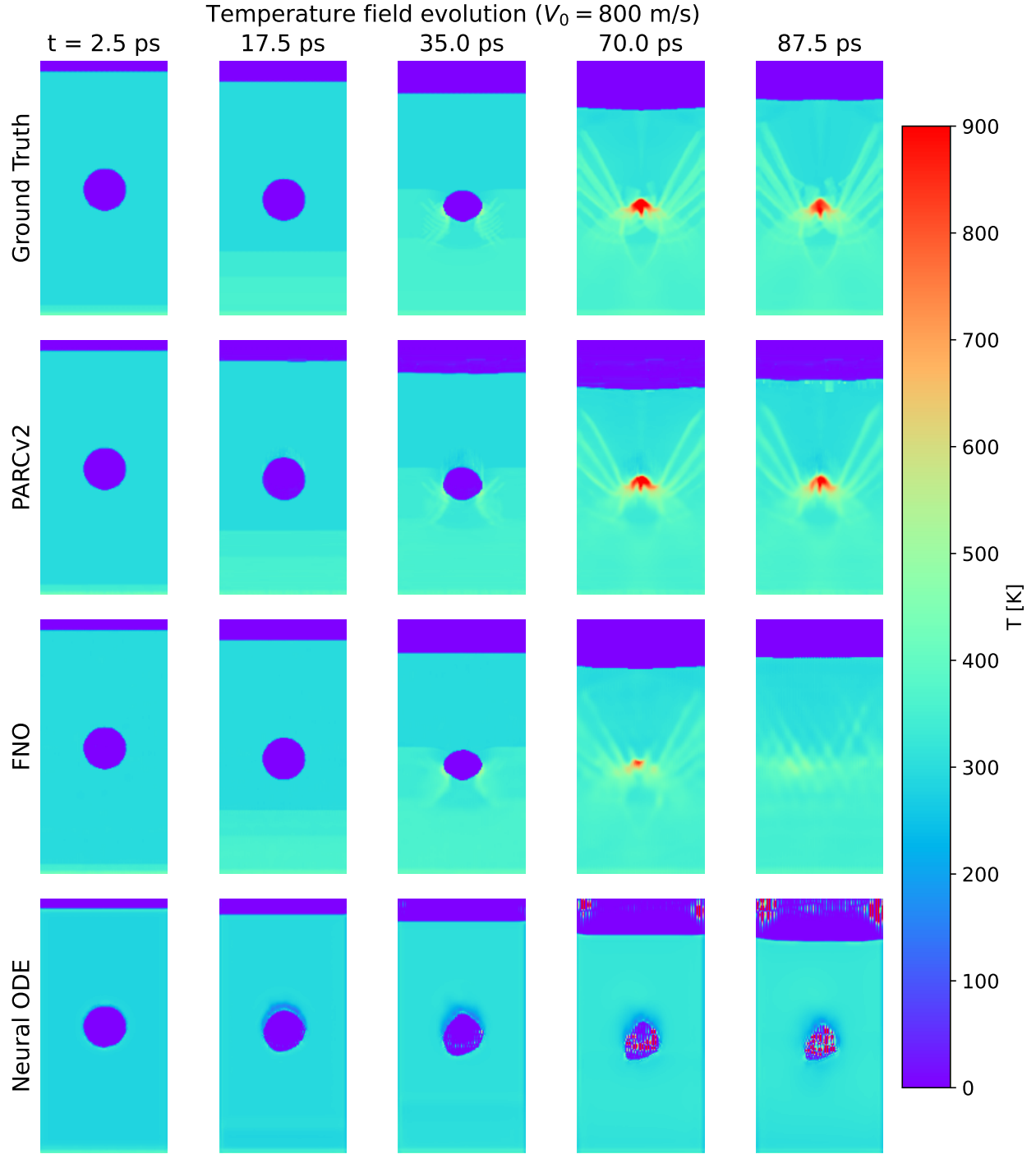


FIG. 9: Temperature field evolution of initial velocity of 800 m/s.

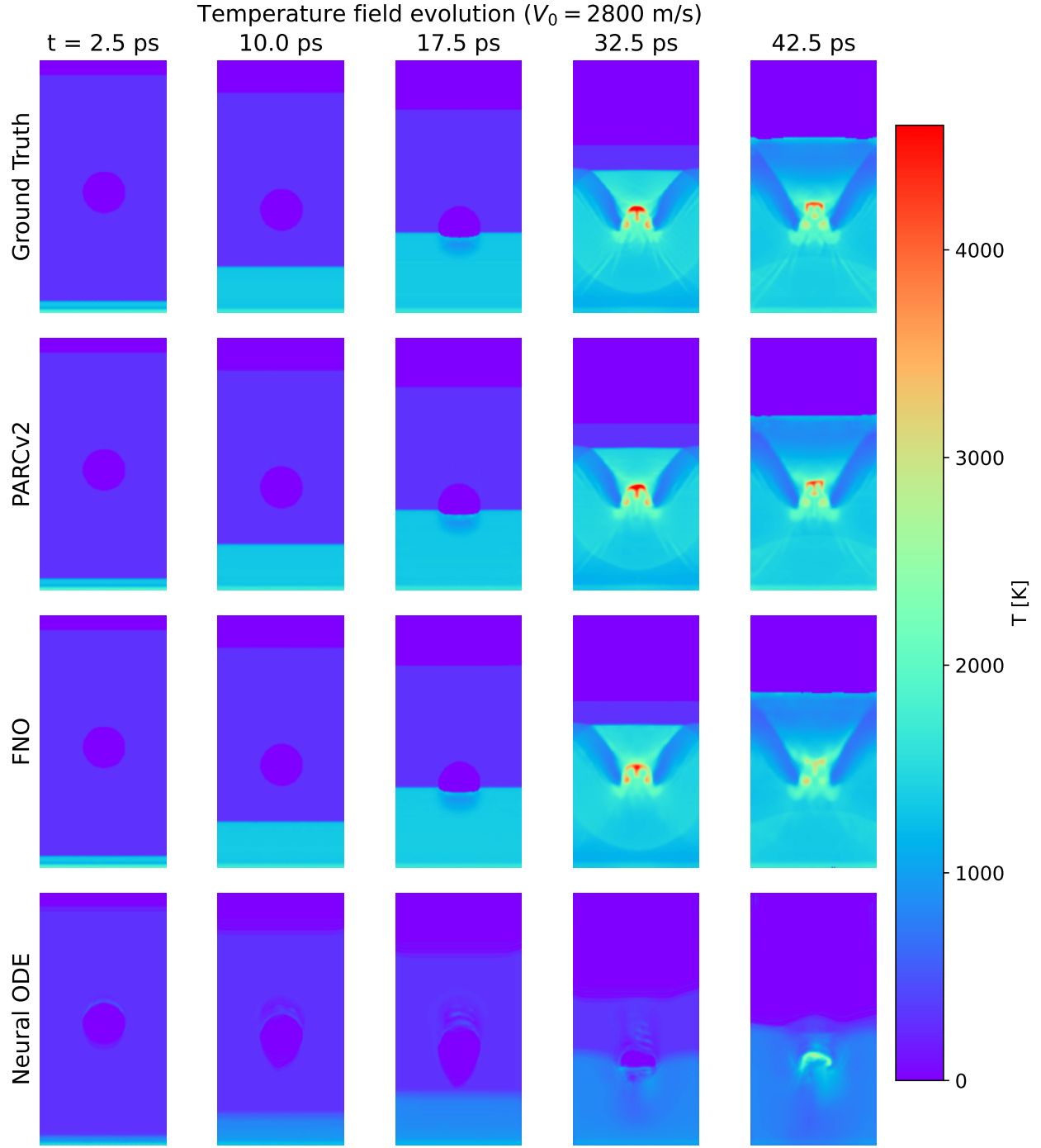


FIG. 10: Temperature field evolution of initial velocity of 2800 m/s.

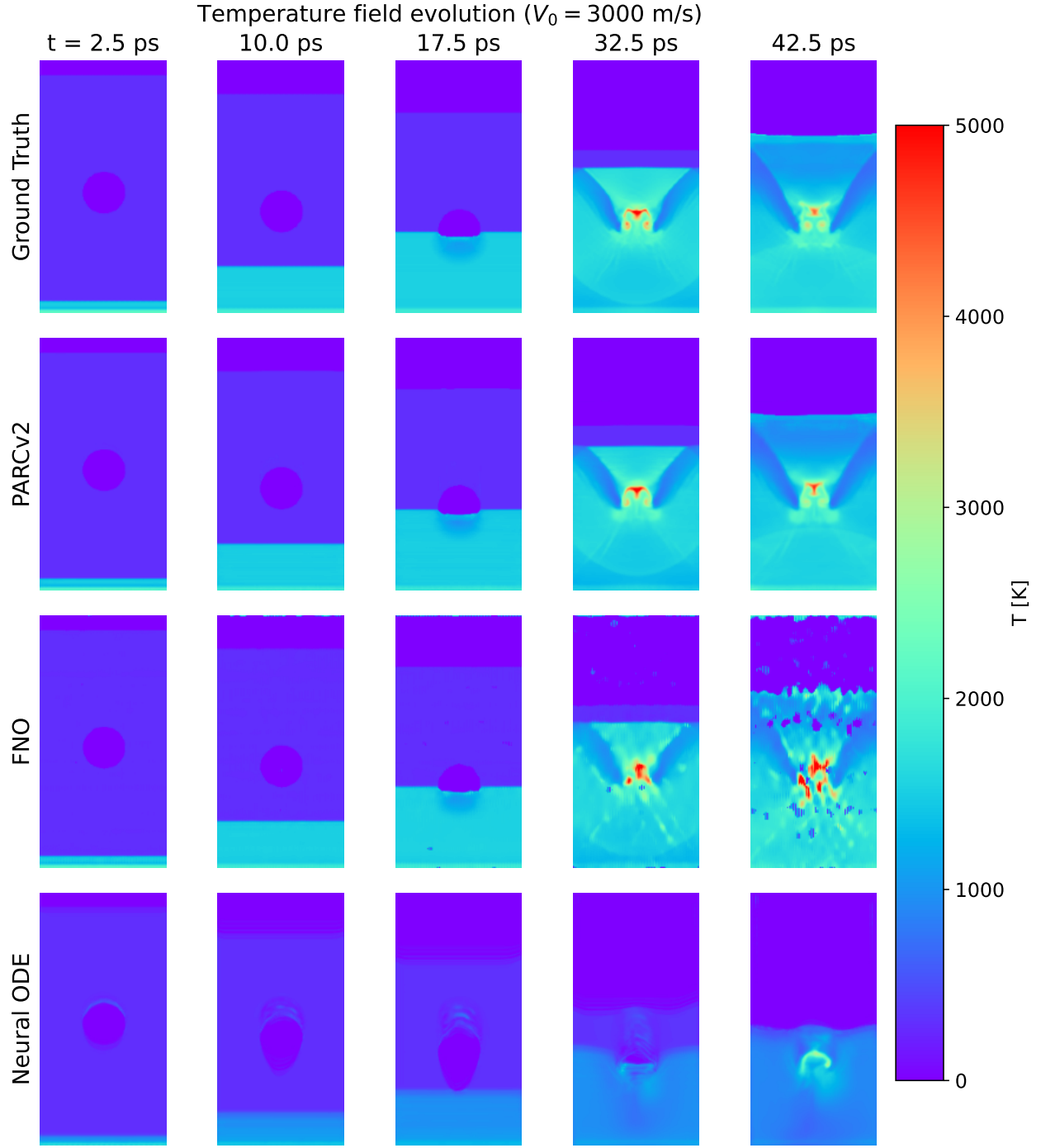


FIG. 11: Temperature field evolution of initial velocity of 3000 m/s.

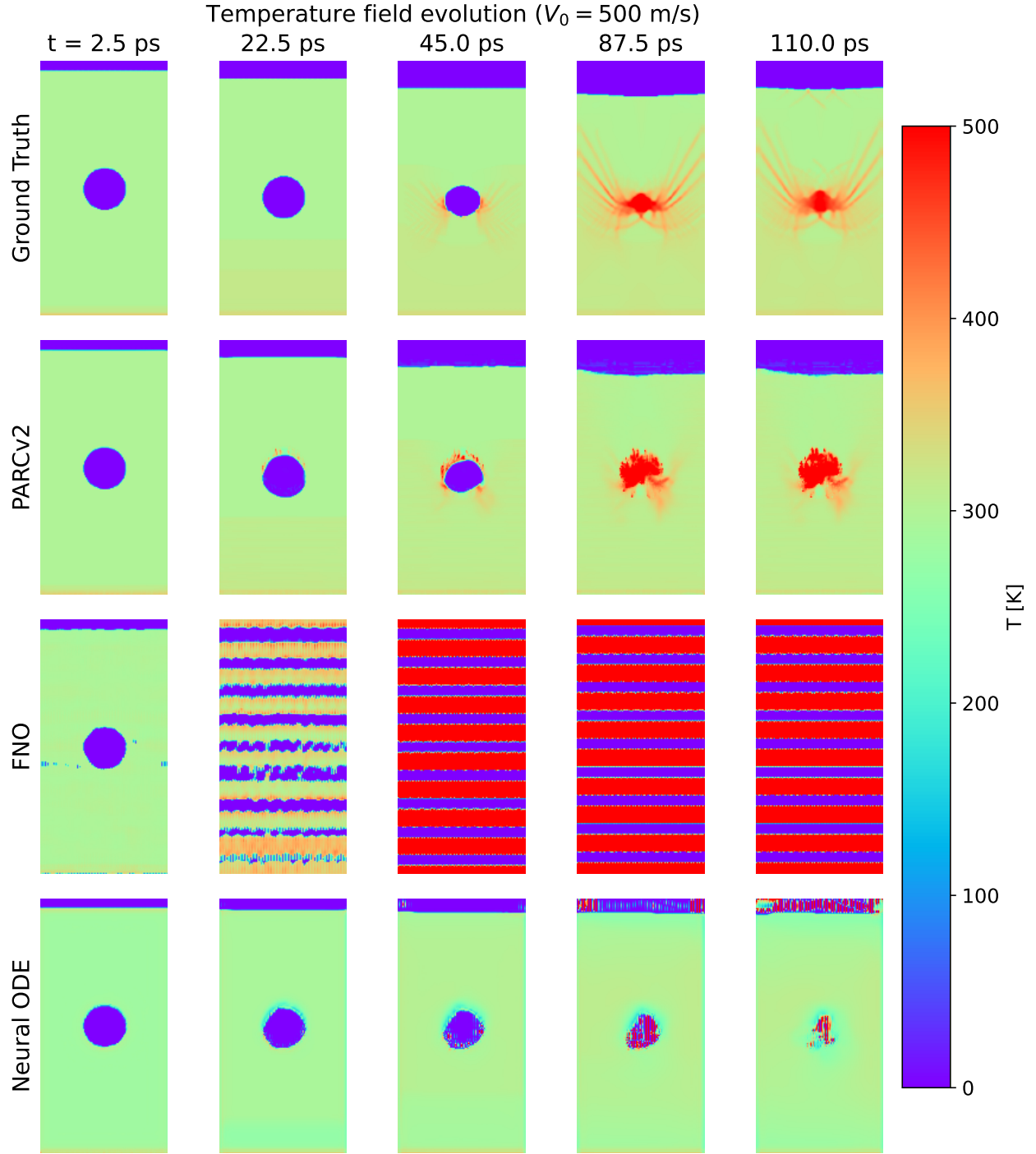
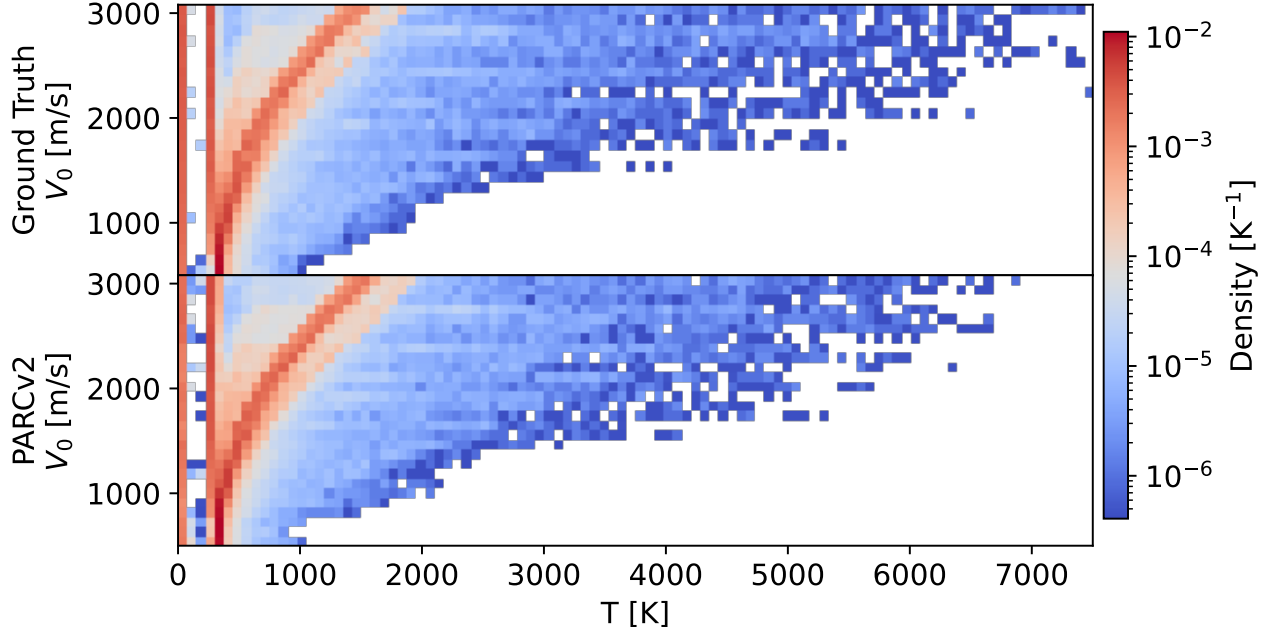
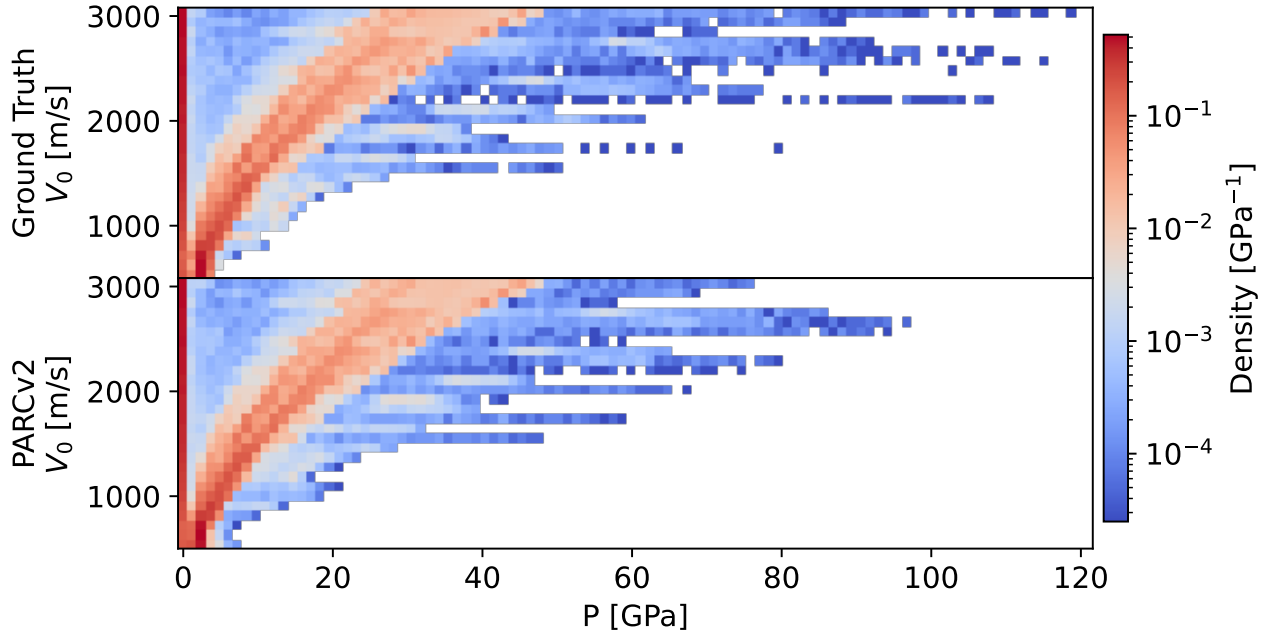


FIG. 12: Temperature field evolution of initial velocity of 500 m/s.



(a) Temperature distribution function



(b) Pressure distribution function

FIG. 13: Temperature (top) and pressure (bottom) distributions at the time of complete pore collapse for different initial velocity test cases.

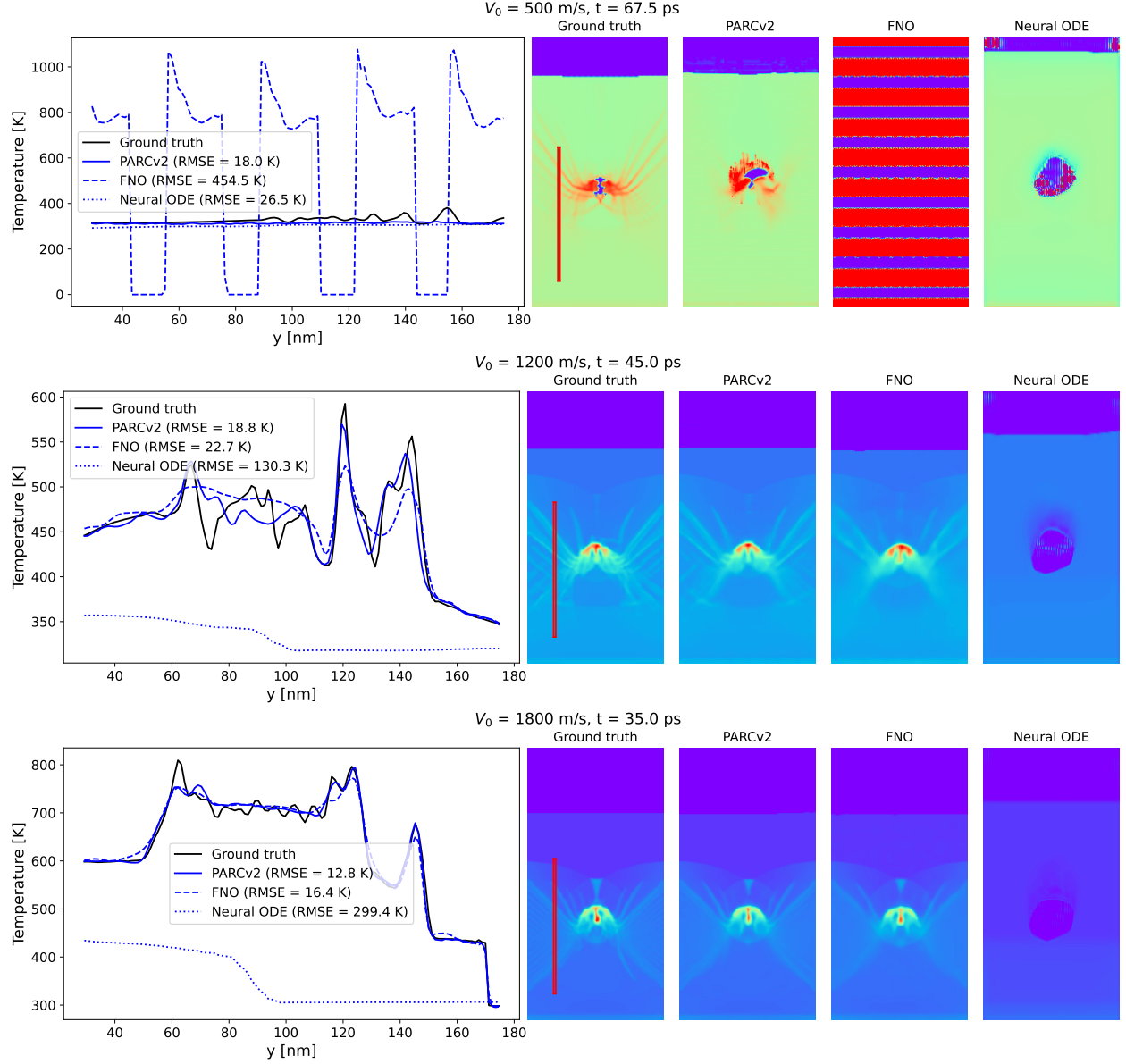


FIG. 14: Shear band profile in temperature for initial impact velocity $V_0=500$, 1200 , 1800 , 2400 and 3000 m/s. We examine the temperature profile along a vertical slice inside the red rectangle. Snapshots at the time of complete pore collapse, when the shear bands is most significant, are chosen to be examined.

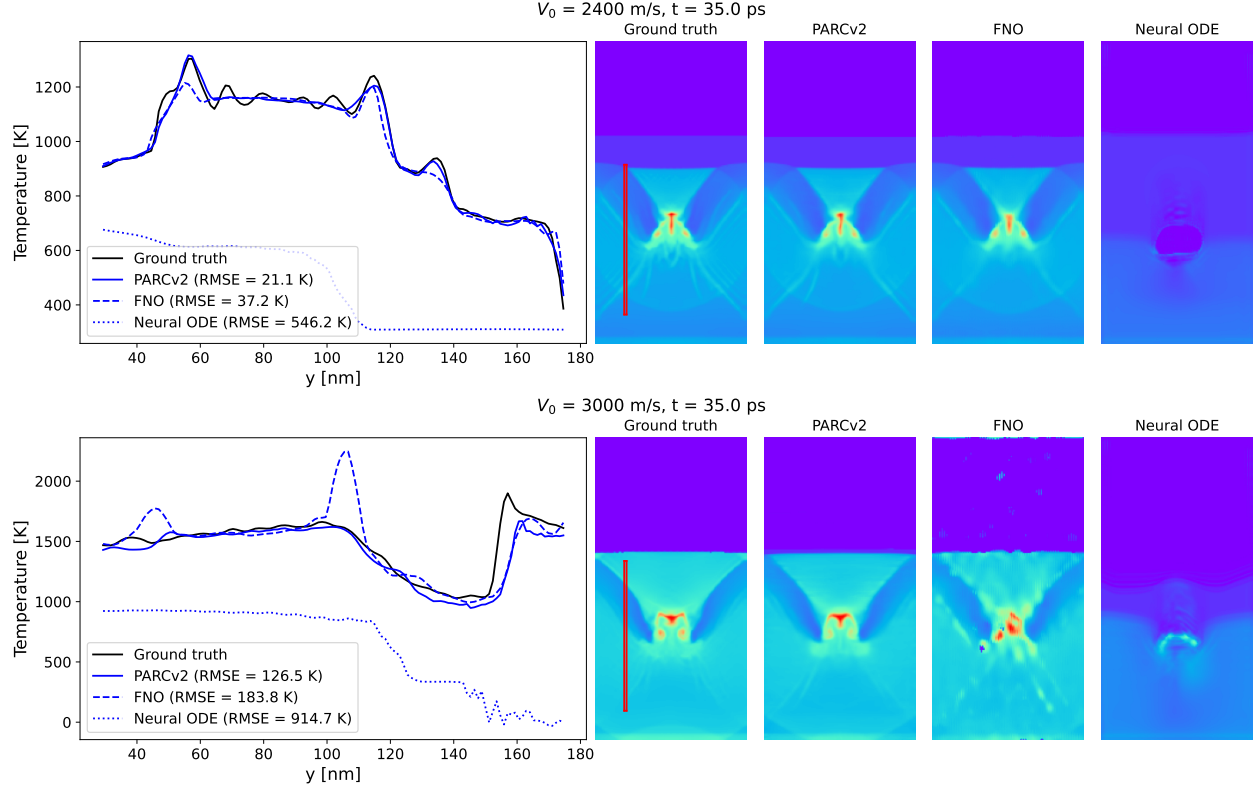


FIG. 15: Continuation of FIG. 14

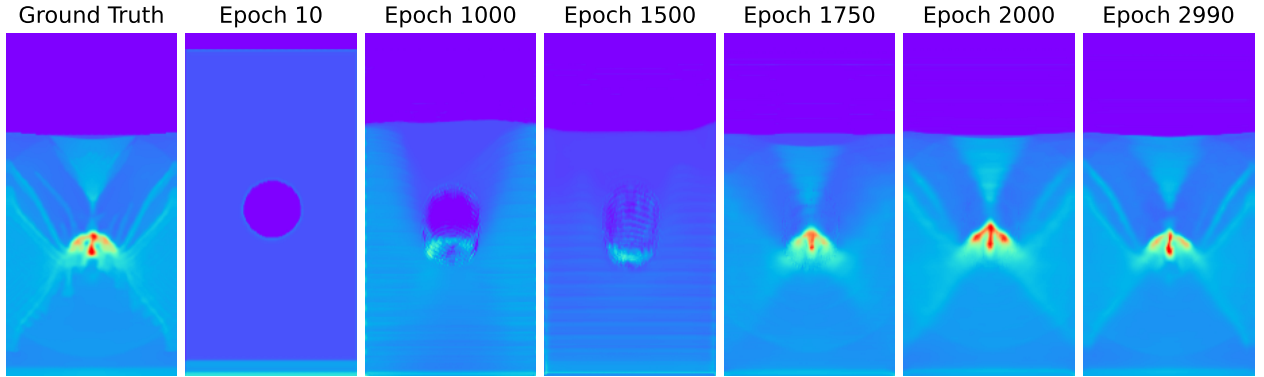


FIG. 16: 16-step roll-out prediction at various epoches during training. We conclude that the model tends to learn the low frequency dynamics first and finer details later, a telltale sign of spectral bias.

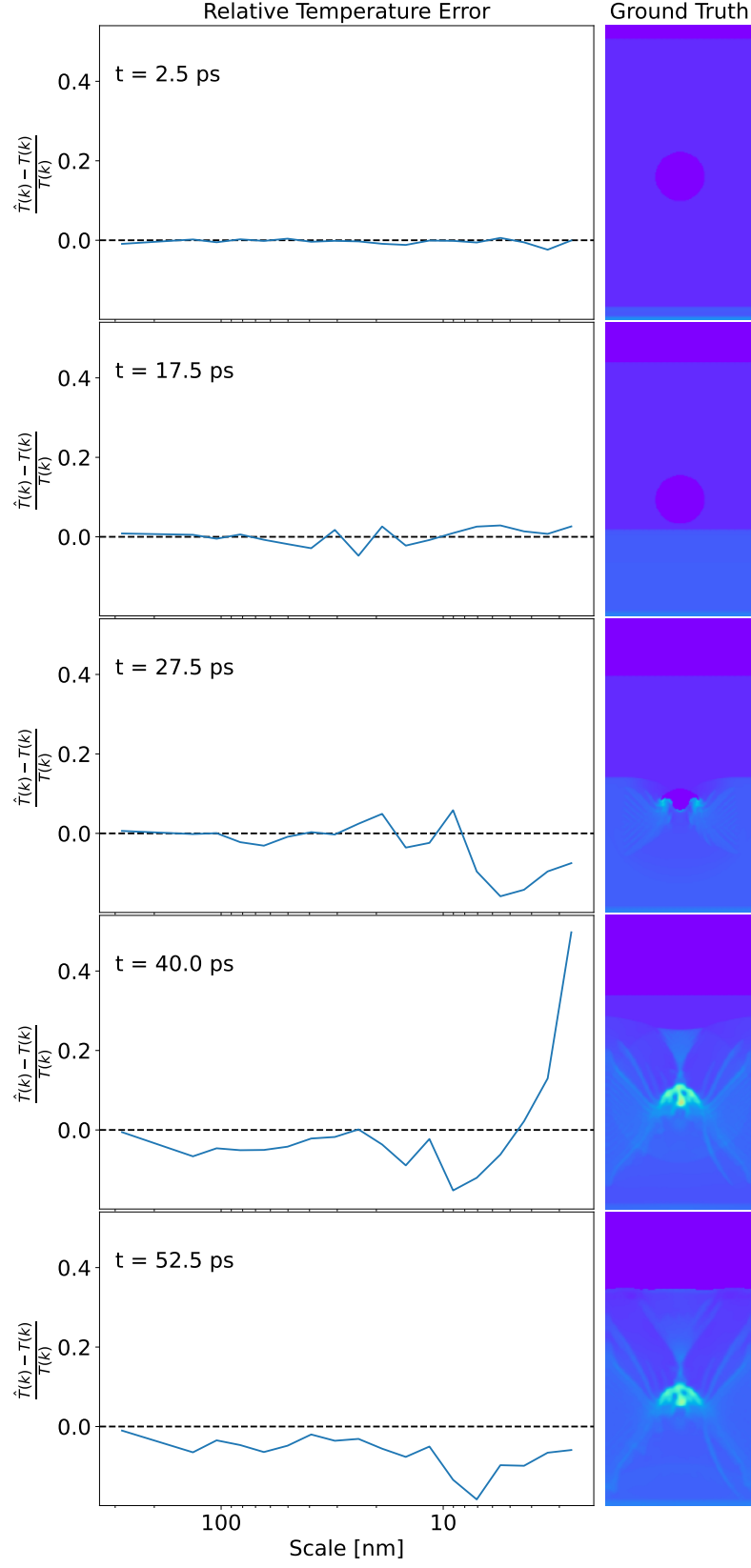


FIG. 17: Relative error of temperature spectrum at various time for $V_0 = 1800$ m/s.

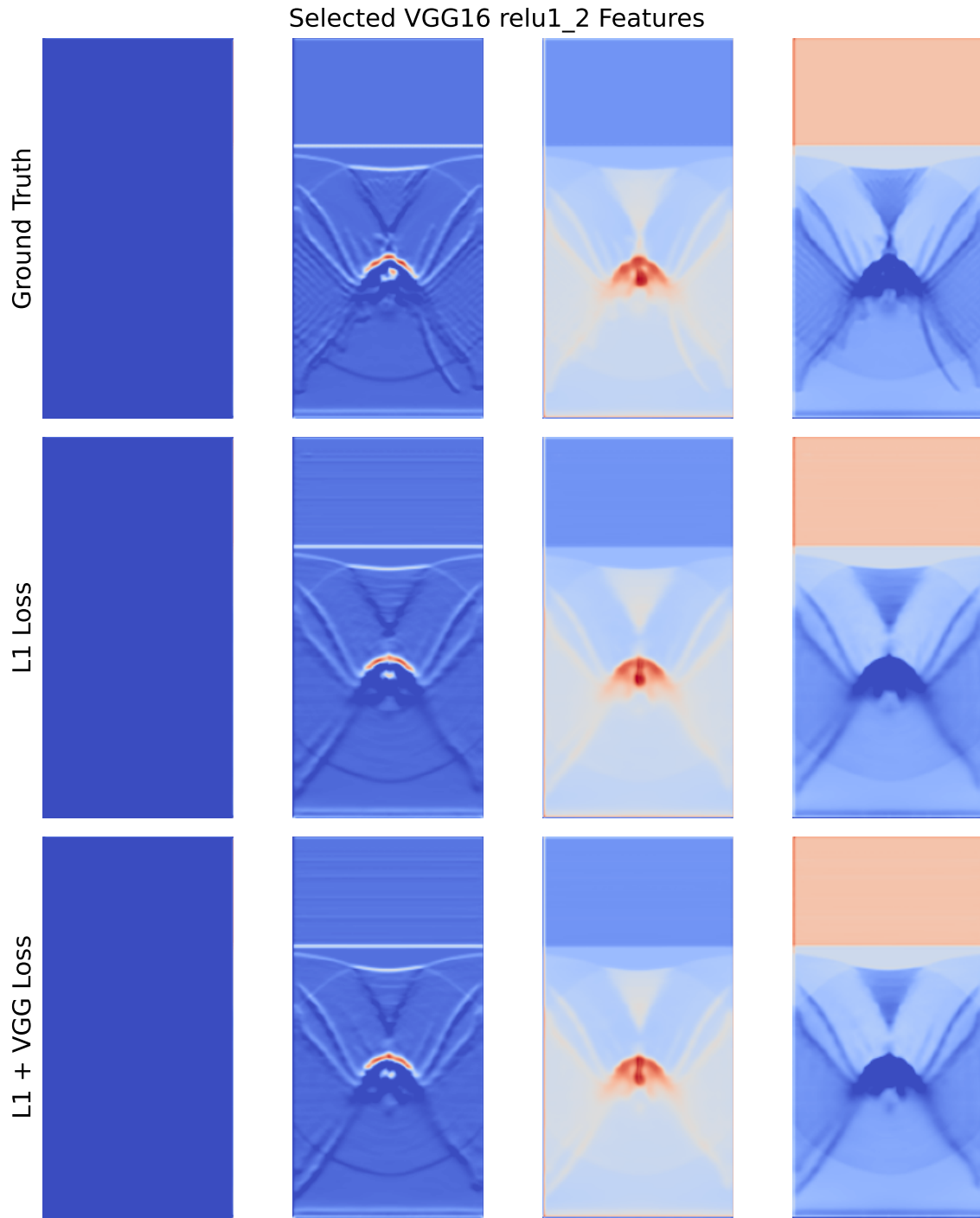


FIG. 18: Feature extraction with pretrained VGG model on ground truth simulation, prediction from model trained with L1 data loss only and prediction from model further finetuned with perception loss. The pretrained VGG16 model can indeed extract the shear band features in some channels but not others. Nevertheless, this indicates that these features can indeed be utilized as perception loss.

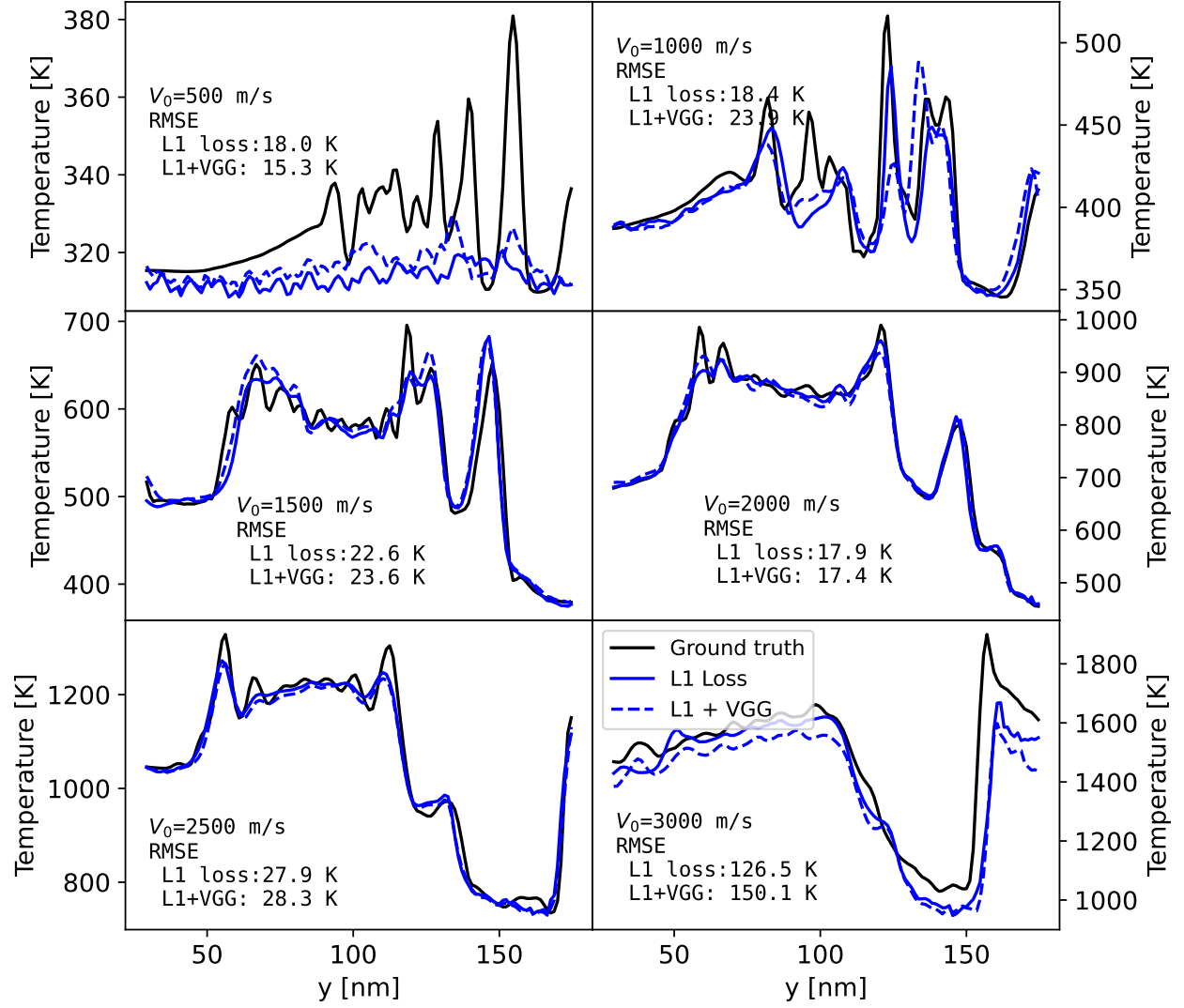


FIG. 19: Comparison of shear band profile with model trained on L1 data loss only, finetuned with VGG loss and finetuned with wavelet loss. We did not observe any meaningful increases in performance in the region of scientific interests.

Appendix A: Continuum Governing Equations and Constitutive Laws

1. Continuum Model

The continuum-scale conservation laws and material models are as follows. The mass, momentum, and energy equations are cast in Eulerian form^{70,71}:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (\text{A1})$$

$$\frac{\partial(\rho \mathbf{u})}{\partial t} + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u} + p \mathbf{I}) = \nabla \cdot \mathbf{S} \quad (\text{A2})$$

$$\frac{\partial \rho \left(e + \frac{1}{2} \mathbf{u} \cdot \mathbf{u} \right)}{\partial t} + \nabla \cdot \left(\rho \left(e + \frac{1}{2} \mathbf{u} \cdot \mathbf{u} \right) + p \right) \mathbf{u} = \nabla \cdot (\mathbf{S} \cdot \mathbf{u}) - \nabla \cdot \mathbf{q} \quad (\text{A3})$$

Here, ρ is the material density; $\mathbf{u}$ is the velocity vector; e is the specific internal energy; T is the temperature; p and $\mathbf{S}$ are respectively the volumetric and the deviatoric components of the Cauchy stress tensor: $\sigma = -p\mathbf{I} + \mathbf{S}$ where $\mathbf{I}$ is the identity tensor. The heat flux $\mathbf{q}$ is given by Fourier's law $\mathbf{q} = -\chi \nabla T$, where χ is the thermal conductivity.

The above equations are supplemented with constitutive models to close the system, as follows.

2. Stress evolution equation

The deviatoric stress tensor $\mathbf{S}$ is obtained from the following hypo-elastic constitutive model in a rate formulation (Prandtl-Reuss^{72,73}):

$$\frac{\partial \rho \mathbf{S}}{\partial t} + \nabla \cdot (\rho \mathbf{S} \otimes \mathbf{u}) + \rho (\mathbf{S} \mathbf{W} - \mathbf{W} \mathbf{S}) = 2\rho G (\mathbf{D}' - \mathbf{D}_{pl}) \quad (\text{A4})$$

where the left hand side is the objective Jaumann rate for the evolution of the deviatoric stress $\mathbf{S}$, $\mathbf{D}' = \frac{1}{2}(\nabla \mathbf{u} + (\nabla \mathbf{u})^T) - \frac{1}{3}(\nabla \cdot \mathbf{u})\mathbf{I}$ is the deviatoric strain-rate tensor, $\mathbf{W} = \frac{1}{2}(\nabla \mathbf{u} - (\nabla \mathbf{u})^T)$ is the spin tensor, $\mathbf{D}_{pl}$ is the plastic component of the deviatoric strain-rate tensor, and G is the shear modulus of the material. Eq. A4 for the evolution of deviatoric stress $\mathbf{S}$ is solved using a two-step operator-splitting algorithm⁷⁴. First, the deviatoric stress is evolved assuming a purely elastic deformation [i.e., setting the plastic component of the deviatoric

strain-rate tensor, $\mathbf{D}_{pl}$ in Eq. A4] in a predictor step:

$$\frac{\partial \rho \mathbf{S}}{\partial t} + \nabla \cdot (\rho \mathbf{S} \otimes \mathbf{u}) + \rho (\mathbf{S} \mathbf{W} - \mathbf{W} \mathbf{S}) - 2\rho G \mathbf{D}' = 0 \quad (\text{A5})$$

This is followed by a correction step to remap the predicted stress onto the yield surface using the radial return algorithm [7-9]. In this step, $\mathbf{D}_{pl}$ is modeled assuming isotropic J_2 plasticity, with a von Mises yield criterion and associated flow-rule. The consistency condition is enforced explicitly and $\mathbf{D}_{pl}$ is computed using explicit plastic integration [ref] where the deviatoric stress is corrected to conform to the yield surface through the radial return algorithm^{73,75,76}.

3. Pressure and Temperature equation of state

The Mie-Gruneisen equation of state relates hydrostatic pressure to specific energy and density⁷⁷ as shown below,

$$p = p_c(\rho) + \Gamma \rho (e - e_c(\rho)) \quad (\text{A6})$$

in which e_c is the cold compression energy, p_c is the corresponding cold pressure, evaluated at 0 K. In Eq. A6 the “cold” (athermal) part of the specific internal energy e_c is obtained from:

$$e_c = e_{c,hydro} + e_{el} + e_{c,pl} \quad (\text{A7})$$

where the subscripts *hydro*, *el*, and *pl* denote respectively hydrodynamic, elastic, and plastic contributions. The hydrodynamic contribution to the athermal part of the internal energy e_c is given by:

$$\dot{e}_{c,hydro} = \frac{p_c(\rho)}{\rho^2} \dot{\rho} \quad (\text{A8})$$

in which $p_c(\rho)$ is the cold compression contribution to pressure, $e_{c,hydro}$ is obtained from Eq. A8 by integration,

$$e_{c,hydro} = e_0 + \int_{\rho_0}^{\rho} \frac{p_c(\rho)}{\rho^2} d\rho \quad (\text{A9})$$

where e_0 is the reference internal energy at T_0 , taken as 0 kcal/kg. The plastic energy $e_{c,pl}$ contribution to cold specific internal energy is due to plastic work and is obtained by solving:

$$\dot{e}_{c,pl} = (1 - \beta) \mathbf{S} : \mathbf{D}_{pl} = (1 - \beta) S_{vm} \dot{\epsilon}_{pl} \quad (\text{A10})$$

In the above equation, β is the Taylor-Quinney parameter S_{vm} is the effective (von Mises) stress and $\dot{\epsilon}_{pl}$ is the effective plastic strain rate. In the last term on the R.H.S. of Eq. A10, the rate of plastic work $\mathbf{S} : \mathbf{D}_{pl}$ is written as $S_{vm}\dot{\epsilon}_{pl}$ for coaxial plasticity (J2 plasticity with Drucker's postulate⁷⁸).

The elastic energy contribution e_{el} is obtained from the difference between the total deviatoric stress work and plastic work, by solving:

$$\dot{e}_{el} = \mathbf{S} : \mathbf{D}_{el} = \mathbf{S} : \mathbf{D}' - S_{vm}\dot{\epsilon}_{pl} \quad (\text{A11})$$

Temperature is computed from the calorific equation of state as:

$$T = T_0 + (e - e_c)/c_v, \quad (\text{A12})$$

where e is computed from A3, T_0 is the reference temperature (298 K), c_v is the isochoric specific heat and the cold energy e_c is given by Eq. A7

TABLE III: RDX Material Properties

Model	Expression	Parameters	Ref	Method
EOS (Pressure (Pa))	$p_c(\rho) = \frac{3}{2}K_0 \left[\left(\frac{\rho}{\rho_0} \right)^{7/3} - \left(\frac{\rho}{\rho_0} \right)^{5/3} \right] \times$ $\left[1 + \frac{3}{4}(K'_0 - 4) \left\{ \left(\frac{\rho}{\rho_0} \right)^{2/3} - 1 \right\} \right]$	$K_0 = 13 \times 10^9 \text{ Pa}$ $K'_0 = 9.2$	79	MD
Shear Modulus (Pa)	$G(P, T) = G_0 + a_1 P + a_2 (T - T_0)$	$G_0 = 5.314 \times 10^9 \text{ Pa}$ $a_1 = 3.3774$ $a_2 = -10.356 \times 10^6$ $\text{Pa} \cdot \text{K}^{-1}$	80	CP informed by MD ⁸¹
Gruneisen coefficient	$\Gamma = \Gamma_0 + \gamma_1 \left(\frac{\rho_0}{\rho} \right) + \gamma_2 \left(\frac{\rho_0}{\rho} \right)^2$	$\Gamma_0 = 0.667$ $\gamma_1 = 2.00878$ $\gamma_2 = -0.805669$	82	MD
Johnson- Cook Strength Model	$S^y = (A + B \epsilon_{ps}^n) \left[1 + C \ln \left(\frac{\dot{\epsilon}_{ps}}{\dot{\epsilon}_0} \right) \right] \times$ $\left[1 - \left(\frac{T - T_{\text{ref}}}{T_{\text{melt}} - T_{\text{ref}}} \right)^m \right]$	$A = 0.3 \times 10^9 \text{ Pa}$ $B = 0.1 \times 10^9 \text{ Pa}$ $m = 3$ $n = 0.1$ $C = 1.8$ $\dot{\epsilon}_0 = 4.36 \times 10^4 \text{ s}^{-1}$	80	CP informed by experiments ⁸³
Melt curve	$T_m(P) = T_{m,\text{ref}} \left[1 + \frac{P - P_{\text{ref}}}{a} \right]^{1/c}$	$T_{m,\text{ref}} = 478 \text{ K}$ $P_{\text{ref}} = 0.0001 \text{ GPa}$ $a = 0.9631 \text{ GPa}$ $c = 2.8855$	84	MD
Heat Capacity	$c_v = 1980 \text{ J kg}^{-1} \text{ K}^{-1}$	N/A	85	MD
Thermal conductivity	$\chi = 0.178 \text{ W m}^{-1} \text{ K}^{-1}$	N/A	86	MD, averaged over < 100 >, < 010 >, and < 001 > crystal orientations