A COMPOSABLE AUTOENCODER-BASED ALGORITHM FOR ACCELERATING NUMERICAL SIMULATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Numerical simulations for engineering applications solve partial differential equations (PDE) to model various physical processes. Traditional PDE solvers are very accurate but computationally costly. On the other hand, Machine Learning (ML) methods offer a significant computational speedup but face challenges with accuracy and generalization to different PDE conditions, such as geometry, boundary conditions, initial conditions and PDE source terms. In this work, we propose a novel ML-based approach, CoAE-MLSim (**Co**mposable **A**uto**E**ncoder **M**achine **L**earning **Sim**ulation), which is an unsupervised, lower-dimensional, local method, that is motivated from key ideas used in commercial PDE solvers. This allows our approach to learn better with relatively fewer samples of PDE solutions. The proposed ML-approach is compared against commercial solvers for better benchmarks as well as latest ML-approaches for solving PDEs. It is tested for a variety of complex engineering cases to demonstrate its computational speed, accuracy, scalability, and generalization across different PDE conditions. The results show that our approach captures physics accurately across all metrics of comparison (including measures such as results on section cuts and lines).

1 INTRODUCTION

Numerical solutions to partial differential equations (PDEs) are dependent on PDE conditions such as, geometry of the computational domain, boundary conditions, initial conditions and source terms. Commercial PDE solvers have shown a tremendous success in accurately modeling PDEs for a wide range of applications. These solvers generalize across different PDE conditions but can be computationally slow. Moreover, their solutions are not reusable and need to be solved from scratch every time the PDE conditions are changed.

The idea of using Machine Learning (ML) with PDEs has been explored for several decades (Crutchfield & McNamara, 1987; Kevrekidis et al., 2003) but with recent developments in computing hardware and ML techniques, these efforts have grown immensely. Although ML approaches are computationally fast, they fall short of traditional PDE solvers with respect to accuracy and generalization to a wide range of PDE conditions. Most data-driven and physics-constrained approaches employ static-inferencing strategies, where a mapping function is learnt between PDE solutions and corresponding conditions. In many cases, PDE conditions are sparse and high-dimensional, and hence, difficult to generalize. Additionally, current ML approaches do not make use of the key ideas from traditional PDE solvers such as, domain decomposition, solver methods, numerical discretization, constraint equations, symmetry evaluations and tighter non statistical evaluation metrics, which were established over several decades of research and development. In this work, we propose a novel ML approach that is motivated from such ideas and relies on dynamic inferencing strategies that afford the possibility of seamless coupling with traditional PDE solvers, when necessary.

The proposed ML-approach, CoAE-MLSim (**Co**mposable **A**uto**E**ncoder **M**achine **L**earning **Sim**ulation), operates at the level of local subdomains, which consist of a group of pixels in 2D or voxels in 3D (for example 8 or 16 in each spatial direction). As shown in Figure 1, the CoAE-MLSim has 3 main components: A) learn solutions on local subdomains, B) learn the rules of how a group of local subdomains connect together to yield

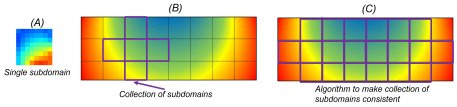

Figure 1: CoAE-MLSim components

locally consistent solutions & C) deploy an iterative algorithm to establish local consistency across all groups of subdomains in the entire computational domain. The solutions on subdomains are learnt into low-dimensional representations using solution autoencoders, while the rules between groups of subdomains, corresponding to flux conservation between PDE solution is learnt using flux conservation autoencoders. The autoencoders are trained *a priori* on local subdomains using very few PDE solution samples. During inference, the pretrained autoencoders are combined with iterative solution algorithms to determine PDE solutions for different PDEs over a wide range of sparse, high-dimensional PDE conditions. The solution strategy in the CoAE-MLSim approach is very similar to traditional PDE solvers, moreover the iterative inferencing strategy allows for coupling with traditional PDE solvers to accelerate convergence and improve accuracy and generalizability.

Significant contributions of this work:

1. The CoAE-MLSim approach combines traditional PDE solver strategies with ML techniques to accurately model numerical simulations.
2. Our approach operates on local subdomains and solves PDEs in a low-dimensional space. This enables generalization to arbitrary PDE conditions within a high-dimensional distribution.
3. The CoAE-MLSim approach training corresponds to training the various autoencoders on local subdomains and hence, it contains very few parameters and requires less data.
4. The iterative inferencing algorithm is completely unsupervised and allows for coupling with traditional PDE solvers.
5. Finally, we show generalization of our model for wide variations in sparse, high dimensional PDE conditions using evaluation metrics in tune with commercial solvers.

2 Related works

The use of ML for solving PDEs has gained tremendous traction in the past few years. Most of the literature has revolved around improving neural network architectures and optimization techniques to enable generalizable and accurate learning of PDE solutions. More recently, there has been a lot of focus on learning mesh independent, infinite dimensional operators with neural networks (NNs) Bhattacharya et al. (2020); Anandkumar et al. (2020); Li et al. (2020b); Patel et al. (2021); Lu et al. (2021b); Li et al. (2020a). The neural operators are trained from high-fidelity solutions generated by traditional PDE solvers on a mesh of specific resolution and do not require any knowledge of the PDE. From the perspective of mesh independent learning, Battaglia et al. (2018) introduced a graph network architecture, which has proven to be effective in solving dynamical systems directly on unstructured computational meshes. Sanchez-Gonzalez et al. (2020) and Pfaff et al. (2020) use the graph network for robustly solving transient dynamics on arbitrary meshes and to accurately capture the transient solution trajectories. For dynamical systems, the neural operators and mesh based learning strategies have shown reasonable prediction capabilities, however, the generalizability of these methods to different PDE conditions remains to be seen. Another direction of research falls under the category of training neural networks with physics constrained optimization. Research in this space involves constraining neural networks with additional physics-based losses introduced through loss re-engineering. PDE constrained optimization have shown to improve interpretability and accuracy of results. Raissi et al. (2019) and Raissi & Karniadakis (2018) introduced the framework of physics-informed neural network (PINN) to constrain neural networks with PDE derivatives computed using Automatic Differentiation (AD) (Baydin et al., 2018). In the past couple of years, the PINN framework has been extended to solve complicated PDEs representing complex physics (Jin et al., 2021; Mao et al., 2020; Rao et al., 2020; Wu et al., 2018; Qian et al., 2020; Dwivedi et al., 2021; Haghighat et al., 2021; Haghighat & Juanes, 2021; Nabian et al., 2021; Kharazmi et al., 2021; Cai et al., 2021a;b; Bode et al., 2021; Taghizadeh et al., 2021; Lu et al., 2021c; Shukla et al., 2021; Hennigh et al., 2020; Li et al., 2021). More recently, alternate approaches that use discretization techniques using higher order derivatives and specialize numerical schemes to compute derivatives have shown to provide better regularization for faster convergence (Ranade et al., 2021b; Gao et al., 2021; Wandel et al., 2020; He & Pathak, 2020). However, the use of optimization techniques to solve PDEs, although accurate, has proved to be extremely slow as compared to traditional solvers. There

is a large body of work related to training neural networks in tandem with differentiable PDE solvers in the loop to improve long range time predictions. The use of differentiable solvers to provide accurate estimates of adjoints has shown to improve learning and to provide better control of PDE solutions and transient system dynamics (Amos & Kolter, 2017; Um et al., 2020; de Avila Belbute-Peres et al., 2018; Toussaint et al., 2018; Wang et al., 2020; Holl et al., 2020; Portwood et al., 2019). These frameworks are useful in providing rapid feedback to neural networks to improve convergence stability and enable efficient exploration of solution space. Bar-Sinai et al. (2019), Xue et al. (2020) and Kochkov et al. (2021) train neural networks in conjunction with a differentiable solver to learn high-fidelity PDE discretization on coarse grids. Singh et al. (2017) and Holland et al. (2019) use differentiable solvers to tune model parameters in a live simulation. However, these methods require a differentiable solver in the training loop and commercial solvers that solve industrial scale problems may not be differentiable. All of the methods discussed above try to learn the learn the dynamics of the PDE in the solution space on entire computational domain, which can be a challenging task as domains can be high-dimensional with complicated physics. For many problems, local learning from smaller restricted domains have proved to accelerate learning of neural networks and provide better accuracy and generalization. Lu et al. (2021a) and Wang et al. (2021) learn on localized domains but infer on larger computational domains using a stitching process. Bar-Sinai et al. (2019) and Kochkov et al. (2021) learn coefficients of numerical discretization schemes from high fidelity data, which is sub-sampled on coarse grids. Beatson et al. (2020) learns surrogate models for smaller components to allow for cheaper simulations. Other methods compress PDE solutions on to lower-dimensional manifolds. This has shown to improve accuracy and generalization capability of neural networks (Wiewel et al., 2020; Maulik et al., 2020; Kim et al., 2019; Murata et al., 2020; Fukami et al., 2020; Ranade et al., 2021a). The ideas proposed in this work share the same goal with the aforementioned literature to accelerate numerical simulations without compromising on accuracy and generalizability. Our work draws inspiration from techniques such as local and latent space learning to solve PDEs in both transient and steady-state settings. More importantly, the approach proposed here relies heavily on unsupervised techniques such as autoencoders to solve PDEs Ranade et al. (2021a;b); Maleki et al. (2021) and mainly focus on leveraging knowledge from traditional solvers to develop a robust, stable, accurate and generalizable machine learning system.

3 CoAE-MLSim model details

3.1 Similarities with traditional PDE solvers

Consider a set of coupled PDEs with n solution variables. For the sake of notation simplicity, we take $n = 2$, such that $u(x, y, z, t)$ and $v(x, y, z, t)$ are defined on a computational domain Ω with boundary conditions specified on the boundary of the computational domain, Ω_b. It should be noted that extension to more solution variables is trivial. The coupled PDEs are defined as follows:

$$L_1(u, v) - F_1 = 0; L_2(u, v) - F_2 = 0 \tag{1}$$

where, L_1, L_2 denote PDE operators and F_1, F_2 represent PDE source terms. The PDE operators can vary for different PDEs. For example, in a non-linear PDE such as the unsteady, incompressible Navier-Stokes equation the operator, $L = \frac{\partial}{\partial t} + \vec{a}.\vec{\nabla} - \vec{\nabla}.\vec{\nabla}$

Traditional PDE solvers solve PDEs given in Eq. 1 by representing solutions variables, u, v, and their linear and non-linear spatio-temporal derivatives on a discretized computational domain. The numerical approximations on discrete domains are computed on a finite number of computational elements known as a mesh, using techniques such as Finite Difference Method (FDM), Finite Volume Method (FVM) and Finite Element Method (FEM). These solvers use iterative solution algorithms to conserve fluxes between neighboring computational elements and determine consistent PDE solutions over the entire domain at convergence. The CoAE-MLSim approach is designed to perform similar operations but at the level of subdomains with assistance from ML techniques.

3.2 CoAE-MLSim for steady-state PDEs

Steady-state PDEs correspond to partial differential equations without time dependencies. The PDE solutions in these equations are primarily governed by PDE conditions, which can have high-dimensional and sparse representations. The solution algorithm of CoAE-MLSim approach for

steady-state PDEs is shown in Fig. 2 and Alg. 1. Similar to traditional solvers, the CoAE-MLSim approach discretizes the computational domain into several subdomains. Each subdomain has a constant physical size and represents both PDE solutions and conditions. For example, a subdomain cutting across the cylinder in Fig. 2 represents the geometry and boundary conditions of the part-cylinder. In a steady-state problem, the computational domain is initialized with initial PDE solutions that are either randomly sampled or generated by coarse-grid PDE solvers. The initial solution is shown with randomly oriented flow vectors in Fig. 2 . Pre-trained encoders e are used to encode the initial solutions as well as user-specified PDE conditions into lower-dimensional latent vectors, η_u, η_v corresponding to PDE solutions and g, b, s corresponding to geometry, boundary conditions and sourceterms, respectively. The flux conservation is iteratively established by passing batches of neighboring subdomain latent vectors through a pre-trained flux conservation autoencoder (Θ). In each iteration, the algorithm loops over all the subdomains, gathers neighbors for a given subdomain and concatenates the neighboring latent vectors corresponding to PDE solutions and conditions. The concatenated latent vectors are evaluated through the pre-trained flux conservation autoencoder to obtain the new solution latent vector ($\eta'_{\vec{u}}$, $\eta'_{\vec{v}}$) state on each subdomain. Subdomains on the boundary have fewer neighbors and latent vectors are zero padded in such cases. The flux conservation autoencoder couples all the solution variables with PDE conditions to ensure that all the dependencies are captured. The iteration stops when the L_2 norm of change in solution latent vectors meets a specified tolerance, otherwise the latent vectors are updated and the iteration continues. The encodings of the PDE conditions are not updated and help in steering the solution latent vectors to an equilibrium state that is decoded to PDE solutions using pre-trained decoders (g_u) on the computational domain. The converged solution in Fig. 2 is represented with flow vectors that are consistent with neighboring subdomains. The iterative procedure used in the CoAE-MLSim approach can be implemented using several linear or non-linear equation solvers, such as Fixed point iterations (Bai et al., 2019), Gauss Siedel, Newton's method etc., that are used in commercial PDE solvers. Physics constrained optimization at inference time might be used to improve convergence robustness and fidelity with physics. The use of autoencoders for compressed solution and condition representation in tandem with iterative inferencing solution algorithm is inspired from Ranade et al. (2021a;b) but the main difference in this work is that the solution procedure is carried out on local subdomains as opposed to entire computational domains to align with principles used in traditional PDE solvers.

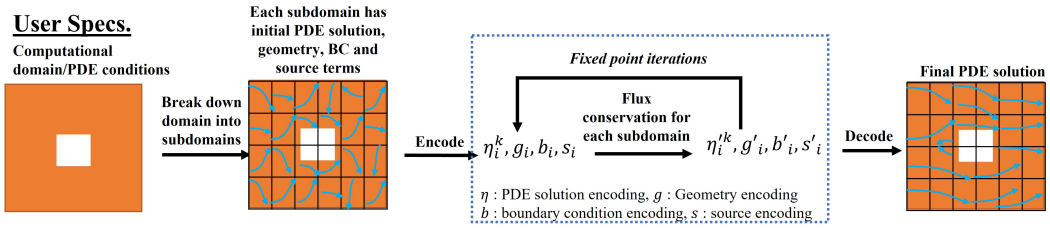

Figure 2: Steady state CoAE-MLSim algorithm

Algorithm 1: Steady state solution methodology of CoAE-MLSim approach

Domain Decomposition: Computational domain $\Omega \rightarrow$ Subdomains Ω_s
Initialize solution on all Ω_s: $\vec{u}(x) = 0.0$ for all $x \in \Omega_s$
Encode PDE solution and conditions on all Ω_s:
$\eta_{\vec{u}} = e_u(u(\vec{\Omega}_s)), \eta_g = e_g(g(\Omega_s)), \eta_b = e_b(b(\Omega_s)), \eta_s = e_s(s(\Omega_s))$
$\epsilon_t = 1e^{-4}$
while $\epsilon > \epsilon_t$ **do**
  for $\Omega_s \in \Omega$ **do**
    Gather neighbors of Ω_s: $\Omega_{nb} = [\Omega_s, \Omega_{left}, \Omega_{right}, ...]$
    $\eta'_{\vec{u}} = \Theta(\eta^{nb}_{\vec{u}}, \eta^{nb}_b, \eta^{nb}_g, \eta^{nb}_s)$
  Compute L_2 norm: $\epsilon = ||\eta_{\vec{u}} - \eta'_{\vec{u}}||_2$
  Update: $\eta_{\vec{u}} \leftarrow \eta'_{\vec{u}}$ for all $\Omega_s \in \Omega$
Decode PDE solution on all Ω_s: $\vec{u} = g_u(\eta_{u(\vec{\Omega}_s)})$

3.3 CoAE-MLSim for transient PDEs

The algorithm for transient CoAE-MLSim differs slightly from steady-state. Our approach models transient PDEs using two methods, fully coupled and loosely coupled. The fully coupled is analogous to the steady-state algorithm where time is considered as an additional dimension in addition to space and modeled using the algorithm described in Alg. 1. On the other hand, in the loosely coupled approach the spatial and transient effects are decoupled. The decoupling allows for better modeling of long range time dynamics and results in improved stability and generalizability. This claim has been validated in Appendix A.2.7 . The solution methodology shown in Figure 3 corresponds to the loosely coupled approach and differs from the steady-state methodology in Figure 2 in that it uses an additional pre-trained time integration network , which integrates the solution latent vectors in time. More details on the time integration network are provided in Section 3.6. Analogous to traditional PDE solvers, a flux conservation is applied after every time integration step. This is very important in establishing local consistency between neighborhood subdomains and minimizing error accumulation resulting from the transient process. Since, flux conservation and time integration networks are trained using the compressed latent vectors of PDE solutions and conditions, the PDE solution and condition autoencoders are required to be trained first.

3.4 PDE solution and condition autoencoders

Autoencoders are used to establish lower-dimensional latent vectors, η, η_s, η_g, for PDE solutions and conditions represented on local subdomains. Figure 4 shows a schematic of the autoencoder setup used in the CoAE-MLSim. Although the representative subdomains shown are 2-D, same concepts apply in higher dimensions. The geometry autoencoders encode a representation of the geometry into latent vector, η_g. In this work, we adopt a Signed Distance Field (SDF) representation of the geometry because it is smooth and differentiable (Maleki et al., 2021). Similarly, the PDE source terms are encoded to their respective latent vector, η_s. The solution autoencoders are conditioned upon the compressed latent vectors of the PDE conditions by concatenating them with the latent vector (η) of the PDE solutions (Ranade et al., 2021a). We recommend the training of each solution variable using a different autoencoder to improve accuracy.The autoencoders are trained using sample PDE solutions generated on entire computational domains and then divided into smaller subdomains for training.

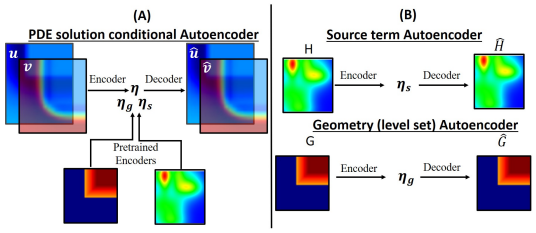

Figure 4: Autoencoders PDE solutions and condition

3.5 Flux conservation autoencoder

Flux conservation autoencoders are responsible for a bulk of the calculations in the solution algorithm. Their main function is to disperse solution information throughout the computational domain by transferring information between neighborhood subdomains and from boundaries and geometries. Figure 5 shows the schematic of the flux conservation autoencoder,

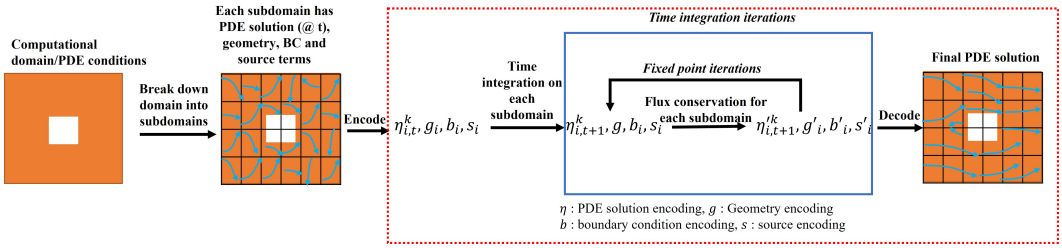

Figure 3: Transient CoAE-MLSim algorithm

which operates on the latent vectors of PDE solutions and conditions on a group of neighboring subdomains. The inputs and outputs to this network consist of concatenated latent vectors of all solution variables and PDE conditions on a group of neighboring subdomains. It uses a deep fully connected neural network to learn these relationships. The samples generated for autoencoders in Section 3.4 are used for training the flux conservation autoencoders as well.

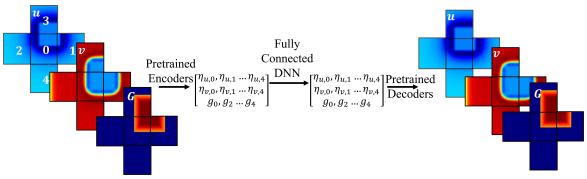

Figure 5: Flux conservation autoencoder

3.6 TIME INTEGRATION NETWORK

Figure 6 shows the schematic of the time integration network, which operates on the latent vectors of PDE solutions and conditions on a group of neighboring subdomains. The time integration network uses fully connected networks to transform the solution latent vector of the center subdomain (corresponding to 0 in Fig. 6) in time. The input of the time integration network can stack latent vectors from multiple previous time steps, $t, t-1, t-2...$, to predict the latent vectors for the next time step, $t+1$. We recommend the training of each solution variable using a different network to improve accuracy.Similar to other autoencoders, the time integration network also learns from randomly generate solution samples. More details on all the autoencoders are in appendix A.1.

3.7 WHY AUTOENCODERS?

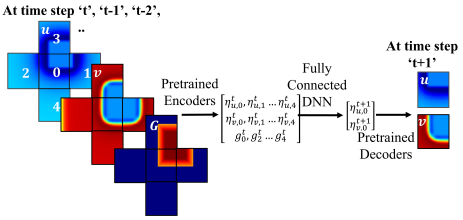

Figure 6: Time integration network

Solutions to classical PDEs such as the Laplace equation can be represented by homogeneous solutions, such as: $\phi(x,y) = a_0 + a_1 x + a_2 y + a_3(x^2 - y^2) + a_4(2xy) + ...$, where, $\vec{A} = a_0, a_1, a_2, ..., a_n$ are constant coefficients that can be used to reconstruct the PDE solution on any local subdomain. $\vec{A}$ can be considered as a compressed encoding of the Laplace solutions. Since, it is not possible to explicitly derive such compressed encodings for other high dimensional and non-linear PDEs, the CoAE-MLSim approach relies on autoencoders to compute them. It is known that non-linear autoencoders with good compression ratios can learn powerful non-linear generalizations (Goodfellow et al., 2016; Rumelhart et al., 1985; Bank et al., 2020). Thus, autoencoders enable efficient latent space computations. Autoencoders also have great denoising abilities, which improve robustness and stability, when used in iterative settings (Ranade et al., 2021a). Finally, they are data-efficient and result in a small number of learnable model parameters and much faster training. For these reasons, autoencoders form the fundamental building block of the CoAE-MLSim approach.

3.8 ATTRIBUTES OF COAE-MLSIM SOLUTION METHODOLOGY

1. **Unsupervised algorithm:** Although, the autoencoders are trained on PDE solutions generated for random PDE conditions, the iterative solution procedure described in Section 3.2 is never explicitly taught the process of computing PDE solutions and discovers solutions with a minimal knowledge about the rules of local consistency. This is remarkably similar to how traditional PDE solvers would operate.

2. **Local learning:** Since the autoencoders are trained on local subdomains, they have fewer trainable parameters and need very less data samples $(100 - 1000)$ to learn the dynamics between PDE solutions and high-dimensional PDE conditions.

3. **Latent space representation:** The iterative solution procedure is carried out in a compressed latent space to achieve solution speed-ups. Furthermore, compressed representations of sparse, high-dimensional PDE conditions improves generalizability.

4. **Coupling with traditional PDE solvers:** The iterative inferencing strategy and the solution algorithm is designed similar to traditional solver and allows for smooth coupling.

4 EXPERIMENTS AND RESULTS

This section demonstrates the CoAE-MLSim approach for two use cases, steady-state conjugate heat transfer and transient vortex decay. Additional use cases corresponding to different PDEs and applications are presented in the appendix sections, A.2.1, A.2.4 and A.2.6. All the experiments have varying levels of complexity across geometries, boundary and initial conditions & source terms imposed on the PDE. The source code and the datasets used in our experiments and analysis can be made available upon acceptance.

Data generation and training mechanics The data required to train the several autoencoders in the CoAE-MLSim approach is generated using Ansys Fluent (Fluent, 2015), except for a few cases where publicly available data sets are used. As stated earlier, the data requirements are minimal and each case requires about 100-1000 solutions, depending on the complexity of physics and dimensionality of PDE conditions. The network architectures and training mechanics are general and described in appendix A.1.

4.1 STEADY-STATE: INDUSTRIAL USE CASE OF ELECTRONIC-CHIP THERMAL COOLING

This experiment demonstrates the steady-state CoAE-MLSim approach. It consists of an electronic chip package surrounded by air and subjected to heat sources with random spatial distributions due to uncertainty in electrical heating. The temperature distribution on the chip is governed by a natural convection process, which is a balance between heating due to heat source and cooling because of flow. The heat sources are sampled from a Gaussian mixture model (up to 25 Gaussians with random mean and variances) and represent a high dimensional space (4096) with large variations, thereby making it incredibly hard to generalize across. The training data for this case corresponds to only 300 solutions generated for random heat sources using Ansys Fluent.

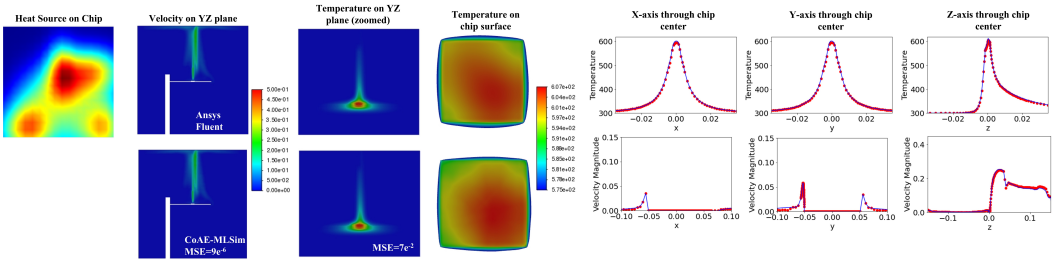

Figure 7: Comparisons: Contour (left) & Line (right, red: CoAE-MLSim & blue: Fluent)

In Fig. 7, we show the temperature and velocity magnitude contour and line plot comparisons between Ansys Fluent and CoAE-MLSim approach for an unseen heat source with very good agreement. Additional details are provided in appendix section A.2.3.

4.2 TRANSIENT: VORTEX DECAY OVER TIME

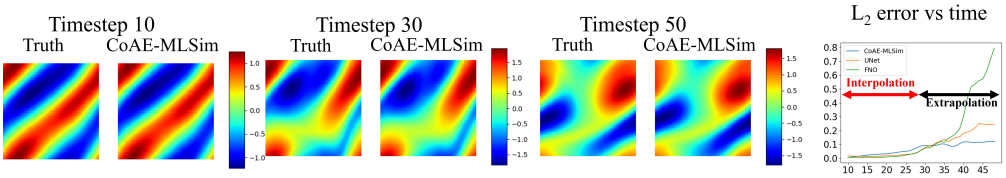

Figure 8: Vortex decay over time with CoAE-MLSim approach

This experiment demonstrates the transient CoAE-MLSim approach. In this case, the transient dynamics of flow are modeled for different choices of initial vorticity. The training data (generated by Li et al. (2020a)) used to train the CoAE-MLSim autoencoders comprises of 500 initial conditions and first 25 timesteps. The testing is carried out on 50 unseen initial conditions and extrapolated

up to 50 timesteps. It may be observed in Fig. 8 that the CoAE-MLSim predictions match well with the ground truth and the error accumulation is acceptable and significantly smaller than baselines such as UNet (Ronneberger et al., 2015) and Fourier neural operator (FNO) (Li et al., 2020a), especially in the extrapolation region for an unseen test sample. This is attributed to the flux conservation autoencoder, which provides a better control over solution trajectory in each time integration. Additional details are provided in appendix section A.2.5.

4.3 DISCUSSION

Analysis for different subdomain sizes:

We test the performance of our approach for varying subdomain resolutions. The autoencoders in the CoAE-MLSim approach are trained using the data corresponding to the case 4.1 with subdomain resolutions of 8^3, 16^3 and 32^3. The trained models are used to solve for PDE solutions for 50 different unseen heat sources and their results, shown in the table below, are compared with Ansys Fluent on 3 metrics, Error in maximum temperature (hot spots on chip), L_∞ error in temperature, Error in heat flux (temperature gradient) on the chip surface. Table 1 shows results for 3 randomly chosen cases and Fig. 9 plots the solve time averaged over all test cases.

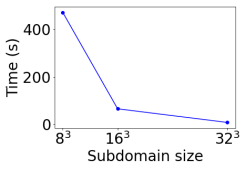

Figure 9: Computation time for different sub-domain sizes

Case ID	$L_\infty(\lvert T_{true} - T_{pred}\rvert)$			$Err.(T_{max})$			$Avg.\ Error(flux)$		
	8^3	16^3	32^3	8^3	16^3	32^3	8^3	16^3	32^3
553	16.93	20.36	16.90	8.35	5.09	5.50	0.65	1.26	0.10
555	19.45	14.35	12.89	-4.32	-1.04	-1.99	0.64	0.59	0.09
574	20.1	17.80	10.20	12.20	8.90	5.00	1.97	1.86	0.13

Table 1: Analysis of different subdomain sizes

The accuracy is very similar for different subdomain sizes, but the computational time is drastically different. Higher subdomain resolution corresponds to fewer subdomains in the entire domain and hence reduction in computational cost. The reduction in solve time is not linear because the latent vector compression is smaller for larger subdomains to represent a larger space of PDE solutions.

Stability:
A long standing challenge in the field of numerical simulation is to guarantee the stability and convergence of non-linear PDE solvers. However, we believe that the denoising capability of autoencoders (Vincent et al., 2010; Goodfellow et al., 2016; Du et al., 2016; Bengio et al., 2013; Ranzato et al., 2007) used in our iterative solution algorithm presents a unique benefit, irrespective of the choice of initial conditions. In this work, we empirically demonstrate the stability of our steady-state approach for case 4.1. In scenario A, we randomly sample 25 initial solutions from a uniform distribution and in scenario B we sample from 10 different distributions, such as Gumbel, Beta, Logistic etc. The mean convergence trajectory and the standard deviation bounds plotted in Fig. 10 show that the L_2 norm of the convergence error falls is acceptable for all cases and demonstrates the stability of our approach.

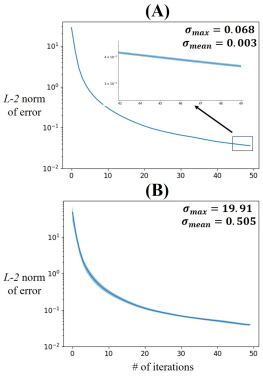

Figure 10: Robustness of CoAE-MLSim approach

Computational speed: We observe that the CoAE-MLSim approach is about 40-50x faster in the steady-state cases and about 100x faster in transient cases as compared to commercial PDE solvers such as Ansys Fluent for the same mesh resolution in all the experiments presented in this work. Both CoAE-MLSim approach and Ansys Fluent are solved on a single Xeon CPU with single precision. We expect our approach to scale to multiple CPUs as traditional PDE solvers but single CPU comparisons are provided here for benchmarking. Moreover, our algorithm is a python language interpreted code, whereas Ansys Fluent is an optimized, C language pre-compiled code. We expect the C/C++ version of our algorithm to further provide independent speedups (not included in current estimates).

Extent of generalization: Like other ML-based approaches, our approach also operates within certain bounds of generalization with respect to the high dimensional space of PDE conditions. In many cases, the PDE conditions also have sparse representations, which makes generalization tougher. Here, we have demonstrated that our approach can generalize within the space of high-dimensional and sparse PDE conditions without compromising on computational speed and solution accuracy while using limited training data for autoencoders.

Coupling with commercial PDE solvers using solution initialization:

In experiment 4.1, we initialize solution variables in iterative algorithm with solutions computed by commercial PDE solvers to demonstrate that our approach can be used as an accelerator to these solvers. In Fig. 11, we show that the coupling between our iterative algorithm and PDE solvers can result in significant convergence speedup for high fidelity solutions as compared to zero initialization. More details on this are in the appendix section A.2.3.

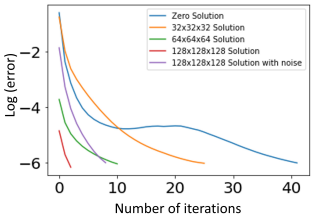

Figure 11: Error history

Comparison with other ML baselines: Although our method is significantly different from traditional ML methods, we have provided comparisons of experiments from our paper and appendix with ML baselines such as UNet (Ronneberger et al., 2015) and FNO (Li et al., 2020a), wherever available. The results are shown in the table 2. All the errors are computed with respect to results computed by a traditional PDE solver. The comparisons are carried over 50 unseen test cases. In the chip cooling experiment 4.1, the L_∞ error in temperature is compared as it is the most relevant metric in this experiment. For the vortex decay case 4.2, the mean absolute error is averaged over 50 timesteps. In both cases, the CoAE-MLSim performs better than the baselines and has small training parameter space. Since the CoAE-MLSim approach is unsupervised, local and lower-dimensional, it requires lesser amount of data. For consistency, we have used the same amount to train the ML baselines.

Table 2: Comparison with baselines

Experiment	CoAE-MLSim	UNet	FNO
Chip cooling (Temperature)	20.36	117.07	x
Vortex decay	0.04	0.08	0.09
Laplace Eq. (in Sec. A.2.1)	0.007	0.195	0.165

	# params
CoAE-MLSim	400 K
UNet	7.418 M
FNO	465 K

5 Conclusion

In this work, we introduced the CoAE-MLSim approach, which is an unsupervised, low-dimensional and local machine learning approach for solving PDE and generalizing across a wide range of PDE conditions randomly sampled from a high-dimensional distribution. Our approach is inspired from strategies employed in traditional PDE solvers and adopts an iterative inferencing strategy to solve PDE solutions. It consists of several autoencoders that can be easily trained with very few training samples. The proposed approach is demonstrated to predict accurate solutions for a range of PDEs and generalize across sparse and high dimensional PDE conditions.

Broader impact and future work: In this work we aim to combine the ideas developed by traditional PDE solvers with current advancements in machine learning and computational hardware and achieve faster simulations for industrial use cases that can range from weather prediction to drug discovery. Although the proposed ML-model can generalize across a wide range of PDE conditions, but extrapolation to PDE conditions that are significantly different still remains a challenge. However, this work takes a big step towards laying down the framework on how truly generalizable ML-based solvers can be developed. In future, we would like to address these challenges of generalizability and scalability by training autoencoders on random, application agnostic PDE solutions and enforcing PDE-based constraints in the iterative inferencing procedure. In this work, we also demonstrate the potential of a hybrid solver by coupling with traditional PDE solvers. This will be investigated further along with extensions to inverse problems and scale invariance of PDE solutions. Finally, one of the limitations of this work is that it cannot handle unstructured meshes and this will be addressed in a follow-up paper.

REFERENCES

Brandon Amos and J Zico Kolter. Optnet: Differentiable optimization as a layer in neural networks. In *International Conference on Machine Learning*, pp. 136–145. PMLR, 2017.

Anima Anandkumar, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Nikola Kovachki, Zongyi Li, Burigede Liu, and Andrew Stuart. Neural operator: Graph kernel network for partial differential equations. In *ICLR 2020 Workshop on Integration of Deep Neural Models and Differential Equations*, 2020.

Shaojie Bai, J Zico Kolter, and Vladlen Koltun. Deep equilibrium models. *arXiv preprint arXiv:1909.01377*, 2019.

Dor Bank, Noam Koenigstein, and Raja Giryes. Autoencoders. *arXiv preprint arXiv:2003.05991*, 2020.

Yohai Bar-Sinai, Stephan Hoyer, Jason Hickey, and Michael P Brenner. Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences*, 116(31):15344–15349, 2019.

Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of machine learning research*, 18, 2018.

Alex Beatson, Jordan Ash, Geoffrey Roeder, Tianju Xue, and Ryan P Adams. Learning composable energy surrogates for pde order reduction. *Advances in Neural Information Processing Systems*, 33, 2020.

Yoshua Bengio, Li Yao, Guillaume Alain, and Pascal Vincent. Generalized denoising auto-encoders as generative models. *arXiv preprint arXiv:1305.6663*, 2013.

Kaushik Bhattacharya, Bamdad Hosseini, Nikola B Kovachki, and Andrew M Stuart. Model reduction and neural networks for parametric pdes. *arXiv preprint arXiv:2005.03180*, 2020.

Mathis Bode, Michael Gauding, Zeyu Lian, Dominik Denker, Marco Davidovic, Konstantin Kleinheinz, Jenia Jitsev, and Heinz Pitsch. Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows. *Proceedings of the Combustion Institute*, 38(2):2617–2625, 2021.

Shengze Cai, Zhicheng Wang, Frederik Fuest, Young Jin Jeon, Callum Gray, and George Em Karniadakis. Flow over an espresso cup: inferring 3-d velocity and pressure fields from tomographic background oriented schlieren via physics-informed neural networks. *Journal of Fluid Mechanics*, 915, 2021a.

Shengze Cai, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Karniadakis. Physics-informed neural networks (pinns) for heat transfer problems. *Journal of Heat Transfer*, 2021b.

James P Crutchfield and BS McNamara. Equations of motion from a data series. *Complex systems*, 1(417-452):121, 1987.

Filipe de Avila Belbute-Peres, Kevin Smith, Kelsey Allen, Josh Tenenbaum, and J Zico Kolter. End-to-end differentiable physics for learning and control. *Advances in neural information processing systems*, 31:7178–7189, 2018.

Bo Du, Wei Xiong, Jia Wu, Lefei Zhang, Liangpei Zhang, and Dacheng Tao. Stacked convolutional denoising auto-encoders for feature representation. *IEEE transactions on cybernetics*, 47(4): 1017–1027, 2016.

Vikas Dwivedi, Nishant Parashar, and Balaji Srinivasan. Distributed learning machines for solving forward and inverse problems in partial differential equations. *Neurocomputing*, 420:299–316, 2021.

ANSYS Fluent. Ansys fluent. *Academic Research. Release*, 14, 2015.

Kai Fukami, Taichi Nakamura, and Koji Fukagata. Convolutional neural network based hierarchical autoencoder for nonlinear mode decomposition of fluid field data. *Physics of Fluids*, 32(9): 095110, 2020.

Han Gao, Luning Sun, and Jian-Xun Wang. Phygeonet: physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state pdes on irregular domain. *Journal of Computational Physics*, 428:110079, 2021.

Frederic Gibou, Ronald Fedkiw, and Stanley Osher. A review of level-set methods and some recent applications. *Journal of Computational Physics*, 353:82–109, 2018.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

Ehsan Haghighat and Ruben Juanes. Sciann: A keras/tensorflow wrapper for scientific computations and physics-informed deep learning using artificial neural networks. *Computer Methods in Applied Mechanics and Engineering*, 373:113552, 2021.

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

Haiyang He and Jay Pathak. An unsupervised learning approach to solving heat equations on chip based on auto encoder and image gradient. *arXiv preprint arXiv:2007.09684*, 2020.

Oliver Hennigh, Susheela Narasimhan, Mohammad Amin Nabian, Akshay Subramaniam, Kaustubh Tangsali, Max Rietmann, Jose del Aguila Ferrandis, Wonmin Byeon, Zhiwei Fang, and Sanjay Choudhry. Nvidia simnetˆ{TM}: an ai-accelerated multi-physics simulation framework. *arXiv preprint arXiv:2012.07938*, 2020.

Philipp Holl, Vladlen Koltun, and Nils Thuerey. Learning to control pdes with differentiable physics. *arXiv preprint arXiv:2001.07457*, 2020.

Jonathan R Holland, James D Baeder, and Karthikeyan Duraisamy. Field inversion and machine learning with embedded neural networks: Physics-consistent neural network training. In *AIAA Aviation 2019 Forum*, pp. 3200, 2019.

Xiaowei Jin, Shengze Cai, Hui Li, and George Em Karniadakis. Nsfnets (navier-stokes flow nets): Physics-informed neural networks for the incompressible navier-stokes equations. *Journal of Computational Physics*, 426:109951, 2021.

Ioannis G Kevrekidis, C William Gear, James M Hyman, Panagiotis G Kevrekidid, Olof Runborg, Constantinos Theodoropoulos, et al. Equation-free, coarse-grained multiscale computation: Enabling mocroscopic simulators to perform system-level analysis. *Communications in Mathematical Sciences*, 1(4):715–762, 2003.

Ehsan Kharazmi, Zhongqiang Zhang, and George Em Karniadakis. hp-vpinns: Variational physics-informed neural networks with domain decomposition. *Computer Methods in Applied Mechanics and Engineering*, 374:113547, 2021.

Byungsoo Kim, Vinicius C Azevedo, Nils Thuerey, Theodore Kim, Markus Gross, and Barbara Solenthaler. Deep fluids: A generative network for parameterized fluid simulations. In *Computer Graphics Forum*, volume 38, pp. 59–70. Wiley Online Library, 2019.

Dmitrii Kochkov, Jamie A Smith, Ayya Alieva, Qing Wang, Michael P Brenner, and Stephan Hoyer. Machine learning accelerated computational fluid dynamics. *arXiv preprint arXiv:2102.01010*, 2021.

Li Li, Stephan Hoyer, Ryan Pederson, Ruoxi Sun, Ekin D Cubuk, Patrick Riley, Kieron Burke, et al. Kohn-sham equations as regularizer: Building prior knowledge into machine-learned physics. *Physical review letters*, 126(3):036401, 2021.

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

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. *arXiv preprint arXiv:2006.09535*, 2020b.

Lu Lu, Haiyang He, Priya Kasimbeg, Rishikesh Ranade, and Jay Pathak. One-shot learning for solution operators of partial differential equations. *arXiv preprint arXiv:2104.05512*, 2021a.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021b.

Lu Lu, Xuhui Meng, Zhiping Mao, and George Em Karniadakis. Deepxde: A deep learning library for solving differential equations. *SIAM Review*, 63(1):208–228, 2021c.

Amir Maleki, Jan Heyse, Rishikesh Ranade, Haiyang He, Priya Kasimbeg, and Jay Pathak. Geometry encoding for numerical simulations. *arXiv preprint arXiv:2104.07792*, 2021.

Zhiping Mao, Ameya D Jagtap, and George Em Karniadakis. Physics-informed neural networks for high-speed flows. *Computer Methods in Applied Mechanics and Engineering*, 360:112789, 2020.

Romit Maulik, Bethany Lusch, and Prasanna Balaprakash. Reduced-order modeling of advection-dominated systems with recurrent neural networks and convolutional autoencoders. *arXiv preprint arXiv:2002.00470*, 2020.

Takaaki Murata, Kai Fukami, and Koji Fukagata. Nonlinear mode decomposition with convolutional neural networks for fluid dynamics. *Journal of Fluid Mechanics*, 882, 2020.

Mohammad Amin Nabian, Rini Jasmine Gladstone, and Hadi Meidani. Efficient training of physics-informed neural networks via importance sampling. *Computer-Aided Civil and Infrastructure Engineering*, 2021.

Ravi G Patel, Nathaniel A Trask, Mitchell A Wood, and Eric C Cyr. A physics-informed operator regression framework for extracting data-driven continuum models. *Computer Methods in Applied Mechanics and Engineering*, 373:113500, 2021.

Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W Battaglia. Learning mesh-based simulation with graph networks. *arXiv preprint arXiv:2010.03409*, 2020.

Gavin D Portwood, Peetak P Mitra, Mateus Dias Ribeiro, Tan Minh Nguyen, Balasubramanya T Nadiga, Juan A Saenz, Michael Chertkov, Animesh Garg, Anima Anandkumar, Andreas Dengel, et al. Turbulence forecasting via neural ode. *arXiv preprint arXiv:1911.05180*, 2019.

Elizabeth Qian, Boris Kramer, Benjamin Peherstorfer, and Karen Willcox. Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems. *Physica D: Nonlinear Phenomena*, 406:132401, 2020.

Maziar Raissi and George Em Karniadakis. Hidden physics models: Machine learning of nonlinear partial differential equations. *Journal of Computational Physics*, 357:125–141, 2018.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

Rishikesh Ranade, Chris Hill, Haiyang He, Amir Maleki, and Jay Pathak. A latent space solver for pde generalization. *arXiv preprint arXiv:2104.02452*, 2021a.

Rishikesh Ranade, Chris Hill, and Jay Pathak. Discretizationnet: A machine-learning based solver for navier–stokes equations using finite volume discretization. *Computer Methods in Applied Mechanics and Engineering*, 378:113722, 2021b.

Marc Ranzato, Christopher Poultney, Sumit Chopra, Yann LeCun, et al. Efficient learning of sparse representations with an energy-based model. *Advances in neural information processing systems*, 19:1137, 2007.

Chengping Rao, Hao Sun, and Yang Liu. Physics-informed deep learning for incompressible laminar flows. *Theoretical and Applied Mechanics Letters*, 10(3):207–212, 2020.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241. Springer, 2015.

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning*, pp. 8459–8468. PMLR, 2020.

Khemraj Shukla, Ameya D Jagtap, and George Em Karniadakis. Parallel physics-informed neural networks via domain decomposition. *arXiv preprint arXiv:2104.10013*, 2021.

Anand Pratap Singh, Karthikeyan Duraisamy, and Ze Jia Zhang. Augmentation of turbulence models using field inversion and machine learning. In *55th AIAA Aerospace Sciences Meeting*, pp. 0993, 2017.

Ehsan Taghizadeh, Helen M Byrne, and Brian D Wood. Explicit physics-informed neural networks for non-linear upscaling closure: the case of transport in tissues. *arXiv preprint arXiv:2104.01476*, 2021.

Marc A Toussaint, Kelsey Rebecca Allen, Kevin A Smith, and Joshua B Tenenbaum. Differentiable physics and stable modes for tool-use and manipulation planning. 2018.

Kiwon Um, Philipp Holl, Robert Brand, Nils Thuerey, et al. Solver-in-the-loop: Learning from differentiable physics to interact with iterative pde-solvers. *arXiv preprint arXiv:2007.00016*, 2020.

Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12), 2010.

Nils Wandel, Michael Weinmann, and Reinhard Klein. Learning incompressible fluid dynamics from scratch towards fast, differentiable fluid models that generalize. *arXiv preprint arXiv:2006.08762*, 2020.

Hengjie Wang, Robert Planas, Aparna Chandramowlishwaran, and Ramin Bostanabad. Train once and use forever: Solving boundary value problems in unseen domains with pre-trained deep learning models. *arXiv preprint arXiv:2104.10873*, 2021.

Wujie Wang, Simon Axelrod, and Rafael Gómez-Bombarelli. Differentiable molecular simulations for control and learning. *arXiv preprint arXiv:2003.00868*, 2020.

Steffen Wiewel, Byungsoo Kim, Vinicius C Azevedo, Barbara Solenthaler, and Nils Thuerey. Latent space subdivision: stable and controllable time predictions for fluid flow. In *Computer Graphics Forum*, volume 39, pp. 15–25. Wiley Online Library, 2020.

Jin-Long Wu, Heng Xiao, and Eric Paterson. Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. *Physical Review Fluids*, 3(7): 074602, 2018.

Tianju Xue, Alex Beatson, Sigrid Adriaenssens, and Ryan Adams. Amortized finite element analysis for fast pde-constrained optimization. In *International Conference on Machine Learning*, pp. 10638–10647. PMLR, 2020.

A APPENDIX

In the appendix, we provide additional details to support and validate the claims established in the main body of the paper. The appendix section is divided into 3 sections. In Section A.1, we provide details related to the network architectures, training mechanics and data generation for the autoencoders used in the CoAE-MLSim approach. In Section A.2, we provide additional details and results for the experiments described in the main paper. Additionally, we have demonstrated the CoAE-MLSim on a different use case. In Section A.3, we present details for reproducibility and for training the various autoencoders in the CoAE-MLSim approach.

A.1 NETWORK ARCHITECTURE, TRAINING MECHANICS AND DATA GENERATION

The training portion of the CoAE-MLSim approach proposed in this work corresponds to training of several autoencoders to learn the representations of PDE solutions, conditions, such as geometry and PDE source terms as well as flux conservation and time integration. The boundary conditions are manually encoded using the strategy explained in Section A.1.3. We train all the autoencoders with the NVIDIA Tesla V-100 GPU using TensorFlow. The autoencoder training is a one-time cost and is reasonably fast. In this section, we describe details related to the network architectures for the different autoencoders, as well as training mechanics and data generation.

A.1.1 GEOMETRY AUTOENCODER

The geometry autoencoder learns to compress signed distance field (SDF) representations of geometry. Mathematically, the signed distance at any point within the geometry is defined as the normal distance between that point and closest boundary of a object. More specifically, for $x \in \mathrm{R}^n$ and object(s) $\Omega \subset \mathrm{R}^n$, the signed distance field $\phi(x)$ is defined by:

$$\phi(x) = \begin{cases} +d(x, \partial\Omega) & x \in \Omega \\ -d(x, \partial\Omega) & x \notin \Omega \end{cases}.$$

where, $\phi(x)$ is the signed distance field for $x \in \mathrm{R}^n$ and objects $\Omega \subset \mathrm{R}^n$ Gibou et al. (2018). Maleki et al. (2021) use the same representation of geometry to successfully demonstrate the encoding of geometries. In this work, we use a CNN-based geometry encoder to encode SDF representation

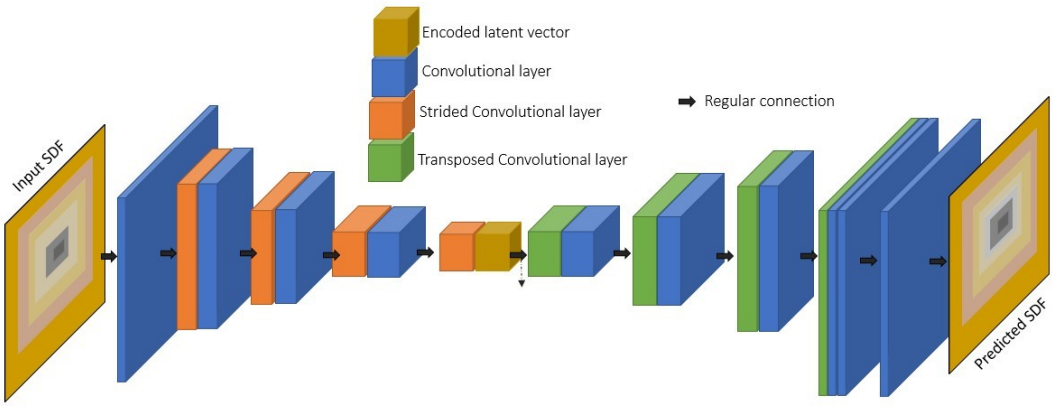

Figure 12: Network architecture for geometry autoencoder (Maleki et al. (2021))

of geometries for relevant use cases. The architecture of this network is shown in Figure 12. The geometry autoencoder has a CNN-based encoder-decoder structure. The encoder compresses the SDF representation to a latent vector, η_g and the decoder reconstructs the SDF representation. In the context of CoAE-MLSim approach, a trained geometry encoder is used to represent SDFs on local subdomains with latent vectors. In Section A.2.4, we present results to demonstrate the generalizability of the autoencoder to encode and decode unseen geometries on subdomains.

Training data and mechanics: The training data pertaining to the geometry autoencoder is constructed from random geometries. Arbitrary geometries on entire computational domains are generated and their SDFs are computed. The computational domain is divided into subdomains and the parts of geometry SDFs associated to every subdomains is used as part of the training samples. Since the training is carried out on subdomains, representations of complicated and arbitrary geometries can be learnt accurately. Moreover, we require only 200-400 geometries on entire computational domains to train the autoencoder on subdomains. The autoencoder is trained until a Mean Squared Error (MSE) of $1e^{-6}$ or Mean Absolute Error (MAE) of $1e^{-3}$ is achieved on a validation set. The latent vector length is of the lowest possible size that can result in training errors below these specified thresholds and can be determined through experimentation. Geometry encoders used in all experiments described in this work are trained as mentioned above.

A.1.2 PDE SOURCE TERM AUTOENCODER

The PDE source term autoencoder learns to compress the spatial distributions of source terms on each subdomain of the computational domain into a latent vector, η_s. The source term autoencoder uses the same architecture as in Figure 12, except that the inputs and outputs are the source term distributions.

Training data and mechanics: The training data for source terms is generated on entire computational domains by sampling from a Gaussian mixture model, where the number of Gaussian's, mean and variance of Gaussian's are arbitrary and span over orders of magnitude. The source terms are divided into subdomains, which are used as training samples for the autoencoder. In this case, we generate about 200-400 such source term distributions and train until the MSE or MAE of the validation set drop below their respective thresholds, $1e^{-6}$ and $1e^{-3}$. The latent vector length is chosen such that the training errors are below the specified thresholds. Source term encoders used in relevant experiments described in this work are trained as mentioned above.

A.1.3 REPRESENTATION OF BOUNDARY CONDITIONS

The boundary condition encoders can be learnt using the same autoencoders described in Figure 12. However, in this work we design a manual encoding strategy to establish the latent vectors for boundary conditions. This is because, the choice of boundary conditions in numerical simulations considered in this paper is very limited. The boundary condition encoding strategy is described in Figure 13. On each face of a subdomain, the boundary condition encoding, η_b, is a vector of size 2.

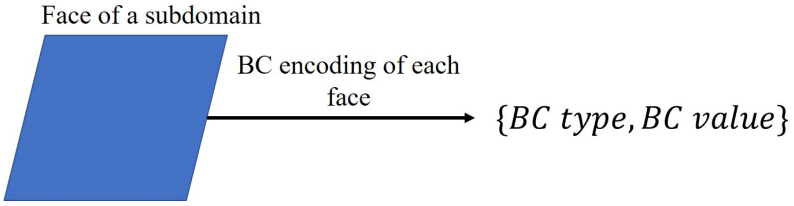

Figure 13: Strategy for BC encoding

The first element of this vector indicates the type of boundary condition, followed by the boundary condition value. In this work, 3 types of boundary conditions are considered, namely Dirichlet, Neumann and Open boundary condition and are indexed as $0, 1, 2$, respectively. Open boundary type refers to faces that do not have any boundary condition imposed on them and is suitable for faces of interior subdomains. For example, a subdomain placed on the inlet boundary, such that the left boundary of the subdomain aligns with the inlet boundary of the entire computational domain, will have a left face encoding of $0, 0.5$, where 0 refers to Dirichlet boundary and 0.5 is the BC value. As we move to applications in structural mechanics in future, new methods for encoding boundaries will be introduced.

A.1.4 PDE SOLUTION AUTOENCODER

Figure 14 shows the network architecture used for encoding the solutions, η, of all PDE solution variables on subdomains. It may be observed that the PDE solution autoencoders are also conditioned

on the geometry, source term and boundary latent vectors, that are associated to the subdomains. Since, the PDE solutions are dependent and unique to PDE conditions, establishing this explicit dependency in the autoencoder improves robustness. We recommend the training of each solution variable using a different autoencoder to determine a latent vector which is independent of the latent vector of other solution variables. This strategy of decoupling has shown to increase the accuracy of the solution autoencoders. The specific parameters used in the network architecture can vary based on the size of each subdomain, complexity of physics and extent of compression achieved.

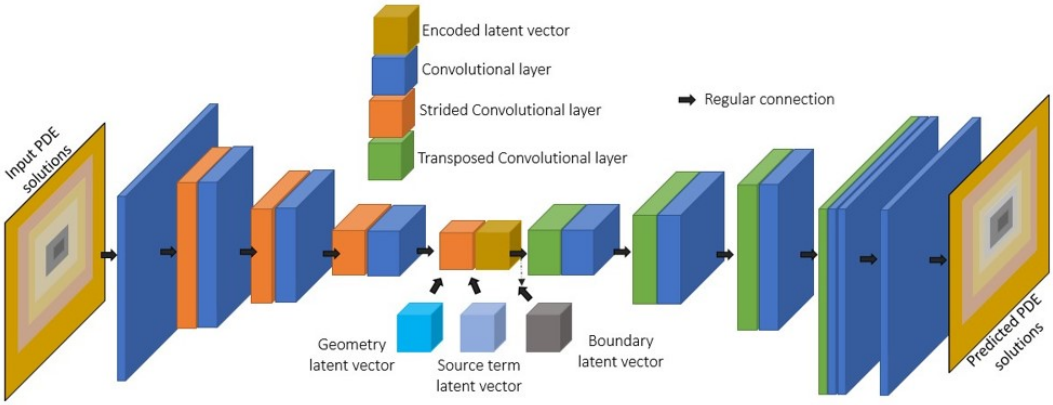

Figure 14: PDE solution autoencoder

Training data and mechanics: The PDE solution variables are very specific to the PDEs being solved and the engineering application being modeled. As a result, a different PDE solution autoencoder needs to be trained, when the application and the corresponding PDE is changed. Similar to the other autoencoders, the PDE solution autoencoder is trained on subdomains until an MSE of $1e^{-6}$ or an MAE of $1e^{-3}$ is achieved on a validation set. The compression ratio is selected such that the solution latent vector has the smallest possible size and yet satisfies the accuracy up to these tolerances. Finally, the data for training these autoencoders is generated by running CFD simulations on entire computational domains for arbitrary PDE conditions. The generated solutions are divided into subdomains and used for training the PDE solution autoencoder. Since, the learning is local the number of solutions required to be generated are about 100-1000 for different PDE conditions. The solution autoencoders used in all the experiments have been trained with the strategy described above.

A.1.5 Flux conservation autoencoders

The flux conservation autoencoder learns the local consistency conditions for a group of neighborhood subdomains in the latent space. Each subdomain is characterized by PDE solutions and conditions and each of these affects the flux conservation autoencoder. As a result, the inputs and outputs of this network are the latent vectors of solution (η), geometry (η_g), source term (η_s) and boundary (η_b) on a group of neighborhood subdomains. All the solution variables of system of PDEs are stacked together with PDE condition latent vectors and the learnt using the autoencoder architecture shown in Figure 15. This autoencoder implicitly learns to represent consistency conditions between neighboring subdomains. Since this autoencoder is only trained on locally consistent subdomains with continuous solutions across intersecting faces, it tries to establish this consistency in neighborhood subdomain solutions for arbitrary inputs. Subdomains at the boundaries may have fewer neighbors and we propose 2 ways to handle this. Firstly, the information related to the missing neighbors can be substituted with a vector of zeros. This would enable learning of all neighboring subdomain combinations with the same flux conservation network. Conversely, different flux conservation networks can be trained for subdomains with different number of neighbors. For example, a subdomain in the corner will have only 3 neighbors, while an interior subdomain has 6 neighbors. In this case, the interior and corner subdomains can be trained separately with different networks.

In our experience, both approaches work equally well but we have adopted the approach of zero padding in this work. The specific parameters used in the network architecture can vary based on the size of each subdomain, complexity of physics and extent of compression achieved.

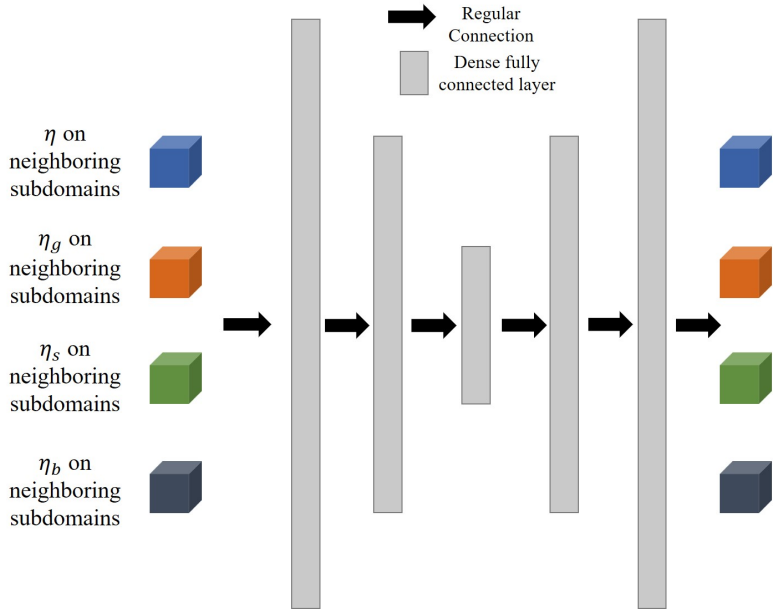

Figure 15: Flux conservation autoencoder

Training data and mechanics: The training data generated for training the PDE solution autoencoders is used to train these networks as well. The data is pre-processed such that groups of neighboring subdomains are collected together and the solutions and conditions associated with them are encoded using pre-trained autoencoders described in previous sections. This processed data is used for training the flux conservation autoencoder. This autoencoder is trained with an MSE loss and the training stops when the validation loss goes below $1e^{-6}$. The flux conservation autoencoders used in all the experiments have been trained with the strategy described above.

A.1.6 Time integration networks

The network architecture and training data generation and mechanics are very similar to the flux conservation autoencoder. The only difference is in the network architecture, where the time integration networks use latent vectors of PDE solutions at time t and conditions of neighborhood subdomains as the input but only predict the solution latent vectors of all solution variables on the center subdomain of the group of neighborhood subdomains at time t+1. Each PDE solution variable can be trained with a different time integration networks. The time integration networks used transient PDE related experiments have been trained with the strategy described above.

A.2 Experiments and additional results

In this section, we provide more results and details of the experiments discussed in the main paper. We have also demonstrated

A.2.1 Steady State: Laplace equations

The Laplace equation is defined as follows:

$$\nabla^2 \phi(\vec{x}) = 0.0 \tag{2}$$

subjected to a Dirichlet boundary condition, $\phi(\vec{x_b}) = f_b$ or a Neumann boundary condition, $\frac{\partial \phi}{\partial \vec{x_b}} = f_b$.

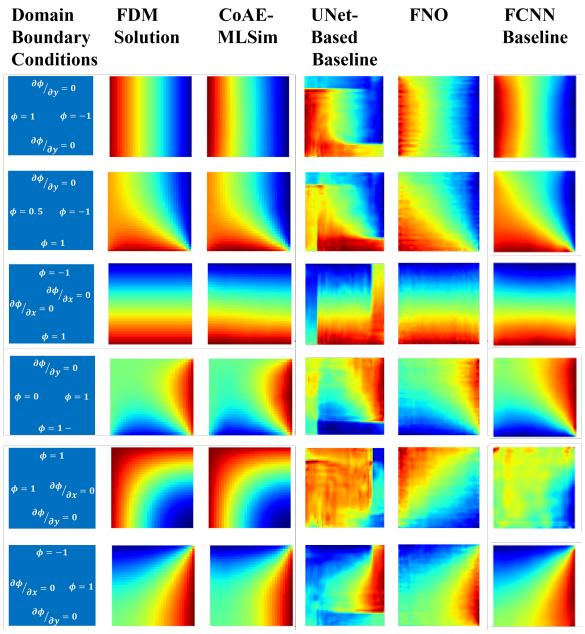

Figure 16: Laplace equation contour results

The Laplace equation is a second order, linear PDE that represents a canonical problem for benchmarking linear solvers. It is solved on a square computational domain with a resolution of 64x64 computational elements for random boundary conditions sampled from Dirichlet and Neumann. Here, we provide visual comparisons between the CoAE-MLSim approach and a second-order finite difference method (FDM) approach in Fig. 16. The PDE solutions are locally scaled with the Dirichlet boundary condition. We compare the CoAE-MLSim with 3 different ML approaches, Unet (Ronneberger et al., 2015), FNO (Li et al., 2020a) and FCNN. The training data for CoAE-MLSim corresponds to 300 solutions generated for arbitrary boundary conditions using a second order finite difference method, whereas all the baselines are trained with 6400 solutions. The baseline models are described below:

UNet: The input to this model is a two-channel grid of size 64 x 64. The first channel captures the boundary condition encoding on 2D dimensional grid. There are two boundary conditions chosen for this experiment, Dirichlet and Neumann. On a grid of zeros everywhere, boundaries are coded by replacing zeros with either a 1(Dirichlet) or a 2(Neumann) based on the boundary condition. Similar to the first channel, second channel is also a grid of zeros, and the edges' zeros are replaced with the magnitude of the boundary condition. The model is trained to output the solution again on a grid of 64 x 64. The UNet has 6 convolutional blocks, 2 at each down-sampled size. The bottleneck size is 8 x 8. The output of the bottleneck is again up-sampled in the usual fashion by concatenating the corresponding down-sampled output. The total number of learnable parameters in UNet baseline is equal to 1.946 million.

FNO: For the Fourier Neural Operator method as well, we have used the same input as in Unet. The FNO model is same as the original implementation in Li et al. (2020a). The FNO model has 1.188 million parameters.

FCNN: In this model, we consider the boundary condition encoding as the input as opposed to a representation of the boundary condition as grid. This network consists of a fully connected network and a convolutional neural network. The boundary condition encoding is first transformed into a vector of size 1024. This vector is then reshaped into 32 x 32 grid. The grid is then passed to convolutional network and the solution is then transformed to 64 x 64 grid. The model has 14.8 million learnable parameters.

It may be observed from Fig. 16 that the CoAE-MLSim approach outperforms the baseline ML models. The mean absolute errors for the CoAE-MLSim approach over 50 unseen testing samples

is $7e^{-3}$. On the other hand, Unet, FNO and FCN have mean absolute errors of $1.65e^{-1}$, $1.9e^{-1}$ and $3e^{-2}$ respectively. All errors are with respect to the second order FDM solution. It may be observed that the FCNN performs better than both UNet and FNO and this points to an important aspect about representation of PDE conditions and its impact on accuracy. The representation of boundary conditions on a 64x64 grid is very sparse and high-dimensional, making it very challenging for the networks to learn. On the other hand, the FCNN uses a low-dimensional, dense encoding as an input and hence is able to learn more effectively. Nonetheless, the CoAE-MLSim approach provides the best performance.

For the CoAE-MLSim approach, the computational domain is divided into 64 subdomains such that each subdomain has a resolution of 8x8 elements. The 64 dimensional PDE solution is represented by a latent vector of size 7 using a CNN-based autoencoder. The total number of parameters in the solution and flux autoencoders are $130,000$. The boundary conditions on each subdomain are encoded using the representation provided in Section A.1.3. Since, this is a steady state problem, the CoAE-MLSim iterative solution algorithm is initialized with a solution field equal to zero in all test cases.

A.2.2 Steady State: Darcy equations

The Darcy equation is defined as follows:

$$-\nabla.(\alpha\nabla\phi(\vec{x}) = f \tag{3}$$

It is subjected to different diffusivity conditions, $\alpha(\vec{x}) = f(\vec{x})$ in a 2-D bounded computational domain between (0, 1). The problem setup as well as the data to train the autoencoders is taken from Li et al. (2020a). We use about 400 samples for training the flux conservation autoencoders, and 50 randomly picked samples for testing from the remaining data. The computational domain in this problem is 2D and has a resolution of 241x241. The domain is divided into 241 subdomains of resolution 21x21 each. The solution on each subdomain is encoded into a latent vector of size 47 and the diffusivity is encoded into size 21. Since, this is a steady state problem, the CoAE-MLSim iterative solution algorithm is initialized with a solution field equal to zero in all test cases. The iterative algorithm convergence tolerance is stringent and set to $L_2 <$ 1e-7.

The main purpose of including this numerical experiment in the paper is to present an ablation study that evaluates the effect of autoencoder bottleneck layer size on the accuracy and computational speed of the solution algorithm used in the CoAE-MLSim approach. As our approach is based on autoencoders we have a rough idea that there must exist an optimum latent vector size which would provide the highest accuracy and a reasonable computational cost during inference time. However, in this experiment we validate this hypothesis. Here, we present results from our experiments to analyze the effect of flux-conservation autoencoder bottleneck layer size on the generalizability, accuracy, and performance of our approach. We choose the flux-conservation autoencoder for this analysis because it is the primary workhorse of our approach and is evaluated repeatedly during the iterative inferencing. We perform 6 experiments with bottleneck layer sizes, 8, 16, 32, 64, 128, 256 for an input size of 340.

The results from these experiments are presented in the table below:

Bottleneck Size (Compression ratio)	Testing error	Convergence iterations	Solve time
8 (42.5)	0.0241	15	0.23
16 (21.25)	0.0107	37	0.57
32 (10.62)	0.0067	45	0.706
64 (5.32)	0.0081	55	0.85
128 (2.65)	0.133	140	2.18
256 (1.32)	0.198	168	2.73

Table 3: Ablation study for different layer sizes

The testing error formulation is shown in Eq. 4.s

$$\epsilon = \frac{L_2(Y_{pred} - Y_{true})}{L_2(Y_{true})} \tag{4}$$

It may be observed from Table 1 3, that as the compression ratio of the flux conservation autoencoder decreases, it begins to overfit and the testing error as well as the number of convergence iterations and computational time significantly increase. On the other hand, if the compression ratio is too large the testing error increases. For all autoencoders used in this work, there exists an optimum bottleneck size compression ratio where the best testing error is obtained and the computational time is not too large.

A.2.3 Steady State: Electronic chip thermal problem

This is an industrial use case where the domain consists of a chip, which is sandwiched in between an insulated mold. The chip-mold assembly is held by a PCB and the entire geometry is placed inside a fluid domain. The geometry and case setup of the electronic chip cooling case can be observed in Fig. 17. The chip is subjected to electric heating and the uncertainty in this process results in random spatial distribution of heat sources on the on the surface of the chip. Fig. 18 shows an example of the various distributions of heat sources that the chip might be subjected to due to electrical uncertainty.

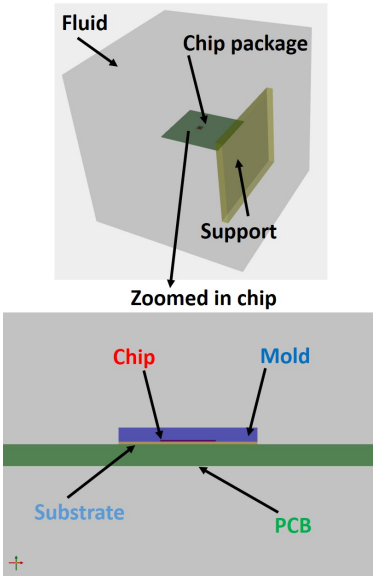

Figure 17: Electronic chip cooling geometry

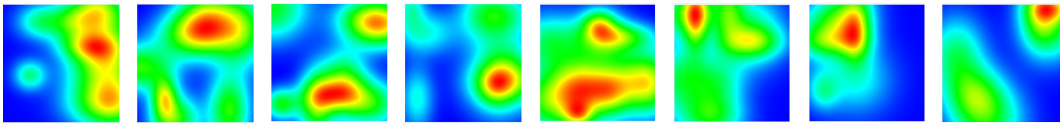

Figure 18: Different power map distributions

The physics in this problem is natural convection cooling where the power source is responsible for generating heat on the chip, resulting in an increase in chip temperature. The rising temperatures get diffused in to the fluid domain and increase the temperature of air. The air temperature induces velocity which in turn tries to cool the chip. At equilibrium, there is a balance between the chip temperature and velocity generated and both of these quantities reach a steady state. The objective of this problem is to solve for this steady state condition for an arbitrary power source sampled from a Gaussian mixture model distribution, which is extremely high dimensional with up to 25

Gaussian's, each with a different mean and variance, on a 4096 dimensional domain. The governing PDEs that represent this problem are shown in Eqs. 5.

$$\left.\begin{aligned} &\textbf{Continuity Equation:} && \nabla.\mathbf{v} = 0 \\ &\textbf{Momentum Equation:} && (\mathbf{v}.\nabla)\mathbf{v} + \nabla\mathbf{p} - \frac{1}{Re}\nabla^2\mathbf{v} + \frac{1}{\beta}\vec{g}\mathbf{T} = 0 \\ &\textbf{Heat Equation in Solid:} && \nabla.(\alpha\nabla\mathbf{T}) - P = 0 \\ &\textbf{Energy Equation in Fluid:} && (\mathbf{v}.\nabla)\mathbf{T} - \nabla.(\alpha\nabla\mathbf{T}) = 0 \end{aligned}\right\} \quad (5)$$

where, $v = u_x, u_y, u_z$ is the velocity field in x, y, z, p is pressure, T is temperature, Re and α are flow and thermal properties, P is the heat source term, $\frac{1}{\beta}\vec{g}T$ is the buoyancy term. P is the spatially varying power source applied on the chip center. The main challenges are in capturing the two-way coupling of velocity and temperature and generalizing over arbitrary spatial distribution of power.

The coupled PDEs with 5 solution variables, $v = u_x, u_y, u_z, P, T$ are solved on a fluid and solid domain with loose coupling at the boundaries. The fluid domain is discretized with 128^3 elements in the domain and the solid domain (chip) is modeled as a 2-D domain with 64^2 elements as it is very thin in the third spatial dimension.

The data to train the autoencoders in the CoAE-MLSim approach is generated using Ansys Fluent and corresponds to 300 PDE solutions. The computational domain is divided into 512 subdomains, each with 16^3 computational elements. The solution autoencoders for the 5 solution variables are trained independently on to establish lower dimensional latent vectors with size 29 on the subdomain level. The geometry and boundary conditions do not vary and hence an autoencoder is not trained for them in this experiment. The source term autoencoder is trained using randomly generated power source fields. Figure 19 shows a few results of the reconstruction capability of the autoencoders. The source term latent vector has a size of 39.

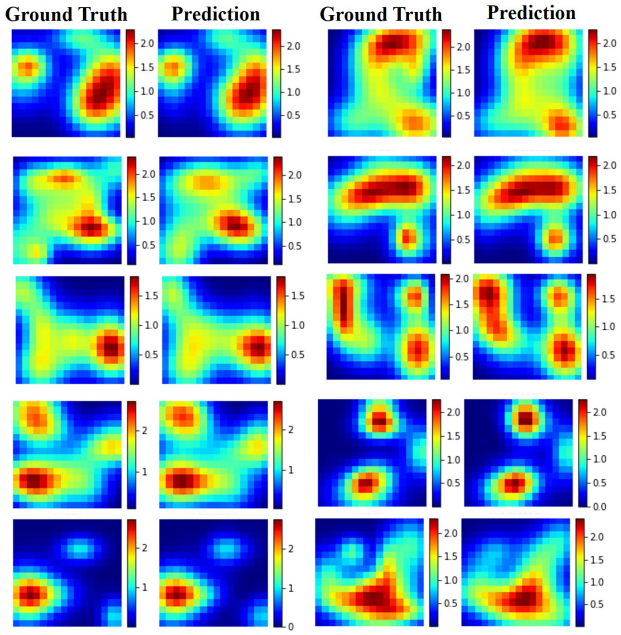

Figure 19: Power source autoencoder for CoAE-MLSim approach

In the main body of the paper, Section 4.3, as well as in this section we compare our approach with other baseline ML models such as UNet (Ronneberger et al., 2015) for this experiment and here we briefly explain the network architecture used.

UNet: The architecture of the UNet (Ronneberger et al., 2015) is as follows: Since the original dimension of the power map is 64^2, to construct a 3D chip power map, zero padding is applied on top and bottom of the 2D power map, so the input dimension of the Unet is 64^3. Each module

of the contracting path consists of 2 convolution layers, 1 max pooling layer and 1 dropout layer. The number of channels changes from 1 to 48 in the contracting path and the module is repeated 4 times. The module in the expansive path consists of 1 Upsampling layer, 1 concatenation layer and 2 convolution layers. Similar to that of the contracting path, the module is repeated 4 times and the last layer is a convolution layer with 1 filter. The activation function used is 'Relu', the convolution kernel size is 3 and the pooling window size is 2.

Additional results:

Here, we present comparisons between CoAE-MLSim approach and Ansys Fluent for 2 additional use cases. We compare velocity magnitude and temperature on plane contorus as well as line plots. Also, we compare additional parameters obtained from the simulation such as temperature on PCB-Fluid interface, pressure in the domain, total energy transfer between chip-fluid interface. It may be observed that the results of our approach agree well with a commercial PDE solver and this continues to work for other choices of power sources from the Gaussian mixture model distribution. Since, this is a steady state problem, the CoAE-MLSim iterative solution algorithm is initialized with a solution field equal to zero for all solution variables in all test cases.

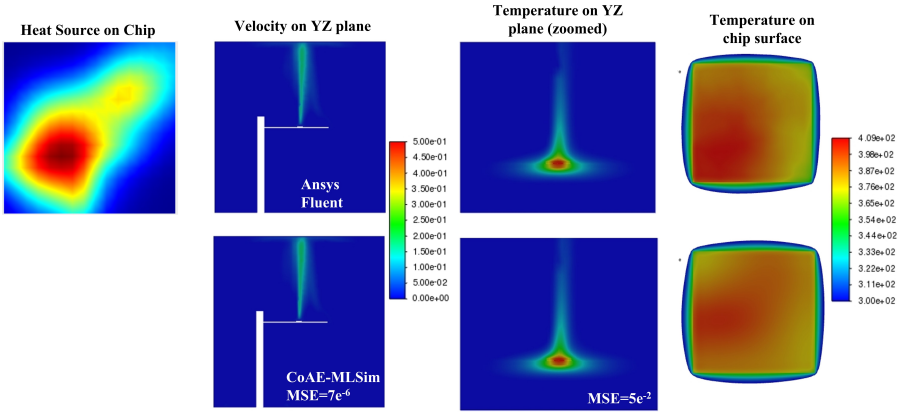

Figure 20: Contour plot comparisons of CoAE-MLSim and Ansys Fluent for test case 1

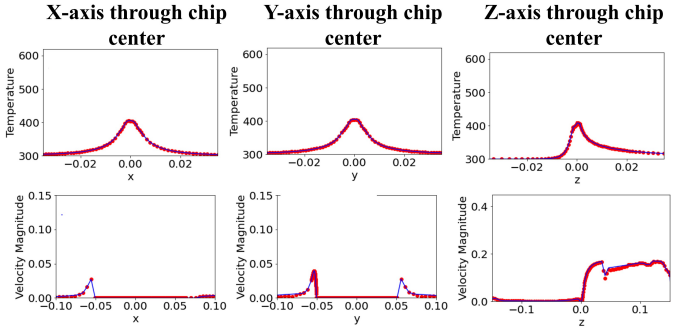

Figure 21: Line plot comparisons of CoAE-MLSim and Ansys Fluent for test case 1 (red: CoAE-MLSim & blue: Fluent)

Finally, we use this experiment to provide further analysis of the CoAE-MLSim approach. The different sub-experiments are listed below:

1. Comparison against UNet for various test cases.
2. Analysis of solution convergence during different iterations of the CoAE-MLSim approach
3. Analysis of coupling with commercial PDE solvers using solution initialization

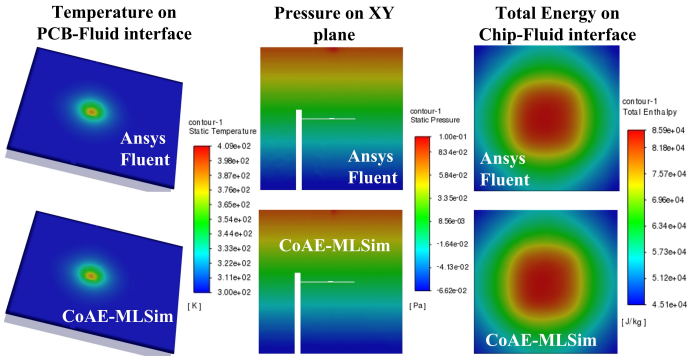

Figure 22: Other parameter contour plot comparisons of CoAE-MLSim and Ansys Fluent for test case 1

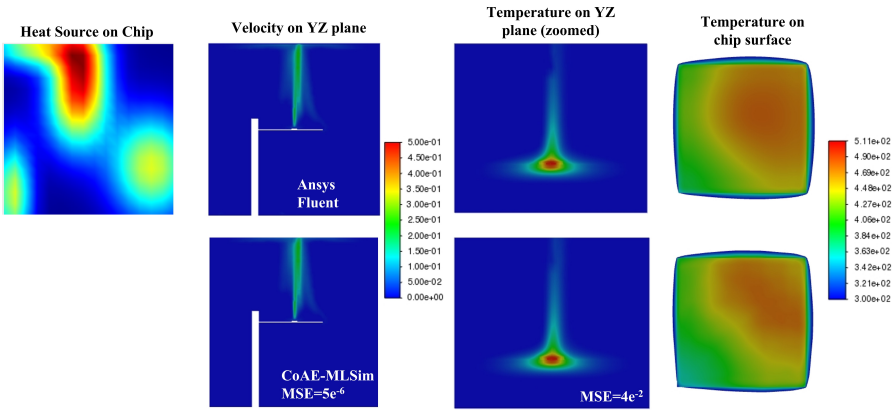

Figure 23: Contour plot comparisons of CoAE-MLSim and Ansys Fluent for test case 2

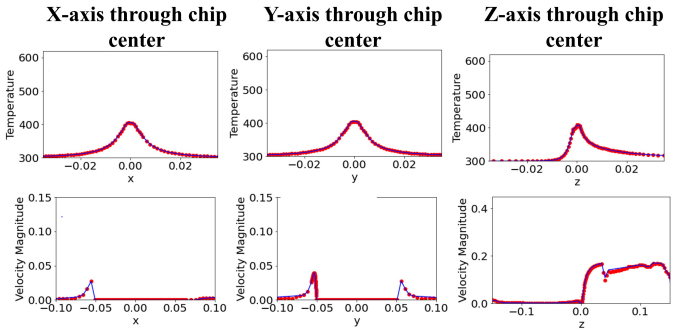

Figure 24: Line plot comparisons of CoAE-MLSim and Ansys Fluent for test case 2 (red: CoAE-MLSim & blue: Fluent)

Comparison of CoAE-MLSim and Unet:

Since we use very few training samples to train the CoAE-MLSim approach, it is important to demonstrate that our approach is not memorizing and that the physics represented by the experiments is non-trivial. Hence, we investigate this by comparing our approach with Unet (Ronneberger et al., 2015). We use both methods to solve for 5 unseen power sources and the results obtained are compared with Ansys Fluent with respect to the metrics discussed below.

1. Error in maximum temperature in computational domain (hot spots on chip),

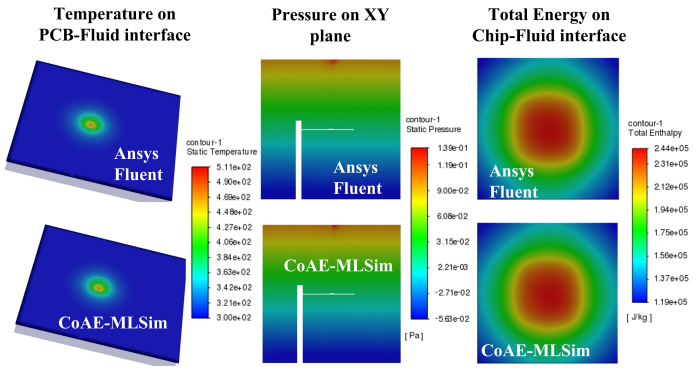

Figure 25: Other parameter contour plot comparisons of CoAE-MLSim and Ansys Fluent for test case 2

Case ID	$L_\infty(\lvert T_{true} - T_{pred}\rvert)$		$Error(T_{max})$		$Avg.\ Error(heat\ flux)$	
	CoAE-MLSim	Unet	CoAE-MLSim	Unet	CoAE-MLSim	Unet
553	20.36	117.07	5.09	-117.06	1.26	7.41
554	10.50	27.24	3.40	27.25	1.45	0.93
555	14.35	62.08	-1.04	-62.08	0.59	2.76
575	15.60	195.52	-0.75	-195.51	0.80	15.78
574	17.80	74.72	8.90	-74.72	1.86	3.41

Table 4: Comparison of UNet and CoAE-MLSim

2. L_∞ error in temperature in the computational domain,
3. Error in heat flux (temperature gradient) on the chip surface

These metrics are more suited for engineering simulations and provide a much better measure for evaluating accuracy and generalization than average based measures. The state-of-the-art Unet (Ronneberger et al., 2015) is trained on the same number of training samples as used by the CoAE-MLSim approach using the architecture described above. The results are compared in the table below.

It may be observed that the our approach outperforms the UNet for all unseen power sources.

Evolution of PDE solution during iterative inferencing:

Figure 26 shows the evolution of the CoAE-MLSim solution on a plane cut through the center of the chip and normal to the z direction, at different iterations until convergence. The results are shown for the y component of velocity for test case 1 presented in Section 4.1 and compared with a converged Ansys Fluent solution for the same case. The initial solution provided to the solver is sampled from a uniform random distribution. At iteration 1, it may be observed that the flux conservation autoencoder denoises the random signals. In the following iterations, the solution begins to develop based on the source term encoding constraint and finally converges at iteration 35. The stable progression of the solution points to the robustness and convergence of the CoAE-MLSim.

Convergence speedup after coupling with traditional PDE solvers:

As mentioned earlier, in all the steady-state experiments carried out in this work, all the solution variables are initialized to zero in the CoAE-MLSim solution algorithm. In this experiment, we argue that better initialization of the solution algorithm can result in faster convergence and the better initialization can be obtained by coupling with commercial PDE solvers. To test this hypothesis, we generate 5 different initial conditions listed below and compare their convergence trajectories with the original case of zero initialization. In each case the solution corresponding to the specified resolution is computed using Ansys Fluent and interpolated on to the 128x128x128x used by CoAE-MLSim approach.

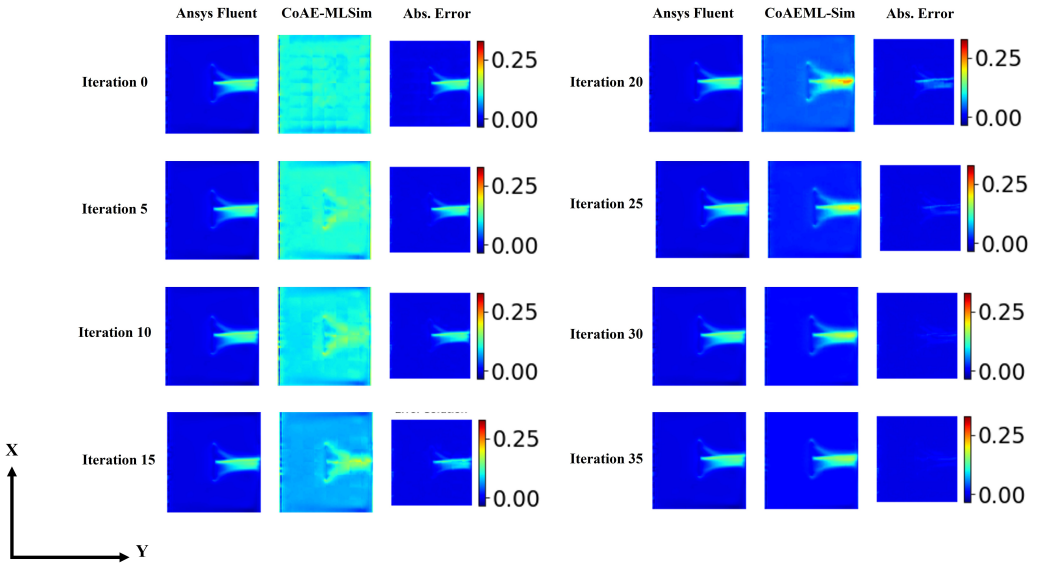

Figure 26: Convergence of CoAE-MLSim solution through the iterative procedure

(a) Initialization with zero solution

(b) Initialization with coarse resolution Ansys Fluent solution generated on 32x32x32 mesh

(c) Initialization with medium resolution Ansys Fluent solution generated on 64x64x64 mesh

(d) Initialization with fine resolution Ansys Fluent solution generated on 128x128x128 mesh

(e) Initialization with fine resolution Ansys Fluent solution generated on 128x128x128 mesh with added random Gaussian noise

The convergence history comparison on an unseen test geometry is shown in Fig. 27. It may be observed that with a zero solution initialization, the CoAE-MLSim iterative solution algorithm takes about 41 iterations to converge. On the other hand, when initialized with the best possible initial guess, which is the ground truth solution on fine grid resolution, it converges in 2 iterations. Next, we add a Gaussian random noise of 25 % maximum relative to the solution on each computational element. With the added noise, the solution converges in about 8 iterations. Finally, when initialized with coarse grid solution generated by Ansys Fluent on 32^3 and 64^3 resolutions, CoAE-MLSim take 27 and 11 iterations respectively. The convergence is still faster than zero solution initialization by 1.5x and 4x, respectively. Nonetheless, this experiment demonstrates that the convergence of the CoAE-MLSim approach can be improved with better initial guess and on the flip side our approach can be used as an accelerator to commercial PDE solvers.

Analysis of train and test errors:

In this section, we have carried out an experiment to compare the accuracy of the inference algorithm on both training and testing PDE conditions. In this problem, we use 300 solutions for training the autoencoders and another 300 for testing. The PDE conditions with respect to both the training and testing set are used in the iterative solution algorithm to generate a solution. In table 5, we compare the average results over the training set and the testing set for the infinity norm and error in maximum value of temperature in the computational domain. We also provide the best error that we obtain from all the training samples.

It may be observed that, the CoAE-MLSim approach performs same on the training set as it would perform on the testing set. The training portion of our model involves training several autoencoders. At inference time, the pretrained encoders cannot directly provide the solution directly for a particular setting of PDE conditions. We have to use the iterative algorithm to compute an equilibrium

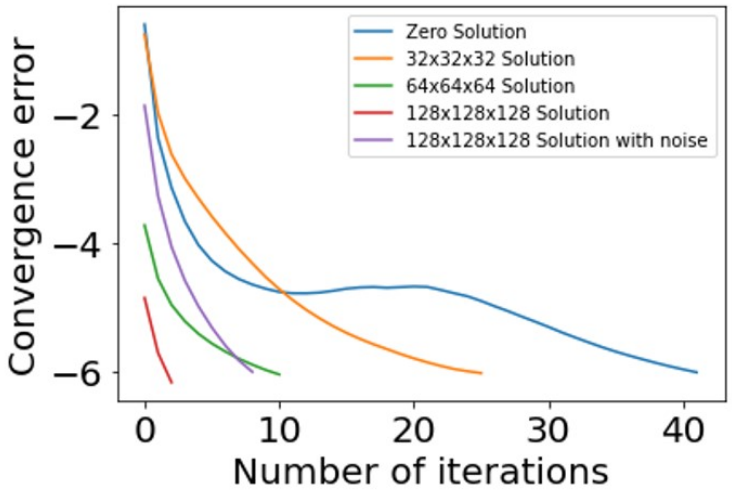

Figure 27: Convergence of CoAE-MLSim solution after coupling with commercial PDE solvers.

	Temperature L_∞	Error (T_{max})
Best training error	3.45	-0.3
Average training error	13.6	2.78
Average testing error	15.1	3.93

Table 5: Comparison of training vs testing errors

solution for a specified set of PDE conditions. The iterative algorithm must figure out the solution and is never taught anything about the trajectory to get there. As a result, the algorithm can perform the same on the training PDE conditions as it would on the testing conditions. This is different from traditional ML methods, which work better in training than in testing.

A.2.4 Steady State: Flow over arbitrary objects

In this use case, the CoAE-MLSim is demonstrated for generalizing across a wide range of geometry conditions. The geometry of objects are represented with a signed distance field representation and is extremely high-dimensional (512-4096). The use case consists of a 3-D channel flow over arbitrarily shaped objects as shown in Fig. 28. The domain has a velocity inlet specified at $1m/s$ and a zero pressure outlet boundary conditions on 2 surfaces, while the rest of the surfaces are walls with no-slip conditions. The shape of the object immersed in flow is arbitrary and the objective is to demonstrate the generalization of the CoAE-MLSim for such geometric variations.

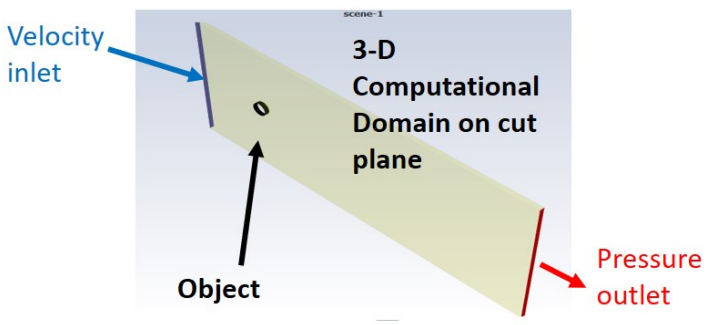

Figure 28: Computational domain

The computational domain, in this case, is elongated in the x direction, such that it consists of 320 computational elements in that direction, while the y and z directions have 160 and 80 computational elements each. The domain is decomposed into subdomains with 16 elements in each direction. Boundary conditions, geometry and PDE solutions are encoded on each subdomain using pre-trained encoders. The PDEs solved in this problem consist of 4 solution variables, include 3 components of velocity and pressure. The governing equations corresponding to these variables are shown in Eq. 6. The subdomain solution of each solution variable is compressed to a latent vector of size 31. The geometry is represented using signed distance fields, which are compressed to a latent vector of size 29. Boundary conditions are encoded using the boundary encoder described in Section A.1.3. There is no source term in this case and hence, it is excluded from the flux conservation autoencoder. The training data in this case corresponds to 300-400 solutions generated for arbitrary geometries using Ansys Fluent. Since, this is a steady state problem, the CoAE-MLSim iterative solution algorithm is initialized with a solution field equal to zero for all solution variables in all test cases.

$$\left.\begin{aligned} &\textbf{Continuity Equation:} && \nabla.\mathbf{v} = 0 \\ &\textbf{Momentum Equation:} && (\mathbf{v}.\nabla)\mathbf{v} + \nabla\mathbf{p} - \frac{1}{Re}\nabla^2\mathbf{v} = 0 \end{aligned}\right\} \tag{6}$$

where, $v = u_x, u_y, u_z$ is the velocity field in x, y, z, p is pressure, T is temperature, Re is Reynolds number representing flow properties.

Figure 29 shows the velocity, pressure and wall flux contour plots comparisons between CoAE-MLSim approach and Ansys Fluent for flow around an unseen arbitrary object. The results match to an acceptable accuracy. In fact, the mean absolute error for pressure and velocity magnitude over 50 unseen test samples is $2.3e^{-2}$ and $9.4e^{-4}$.

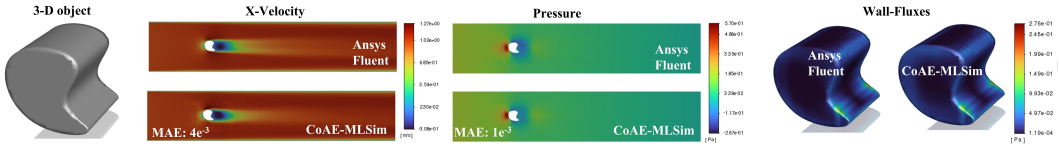

Figure 29: Flow over external object contour plot comparisons

Additional examples Figure 30 shows comparisons of CoAE-MLSim with Ansys Fluent for 4 unseen objects in addition to the example shown in Figure 29. The solver generalizes well and the errors fall within an acceptable range.

Geometry encoder performance

Figure 31 shows the comparison between the true signed distance field and the reconstructed field from the geometry autoencoder. In this example, the geometry autoencoder is evaluated on each subdomain, but all subdomains are reassembled using their connectivity information to obtain the SDF on the entire computational domain. The contour plots in Figure 31 are on planes cut through the center of the geometry along x direction. It may be observed that the autoencoder reconstructs agree well with the ground truth for unseen cases. The mean absolute error in all the cases presented here is less than $1e^{-3}$.

A.2.5 TRANSIENT: VORTEX DECAY FLOW

This is a similar use case from Li et al. (2020a) where a 2-d Navier-Stokes equation for a viscous, incompressible fluid in vorticity form on the unit torus. The PDE and case details corresponding to this may be found in Li et al. (2020a). In the main paper, we provided a single result for an unseen initial condition when integrated in time. Here, we will provide visual comparisons of additional results and additional details about the CoAE-MLSim approach used for this problem.

As described earlier, the mesh resolution considered in this experiment consists of 64x64 computational elements. The dataset considered in this experiment corresponds to a viscosity of $1e^{-3}$ and contains transient evolution of 5000 initial solution fields over 50 timsesteps. The autoencoders of the CoAE-MLSim approach are trained with data corresponding to only 500 initial conditions and first 25 timesteps. The computational domain is divided into 64 subdomains of 8^2 resolution. The

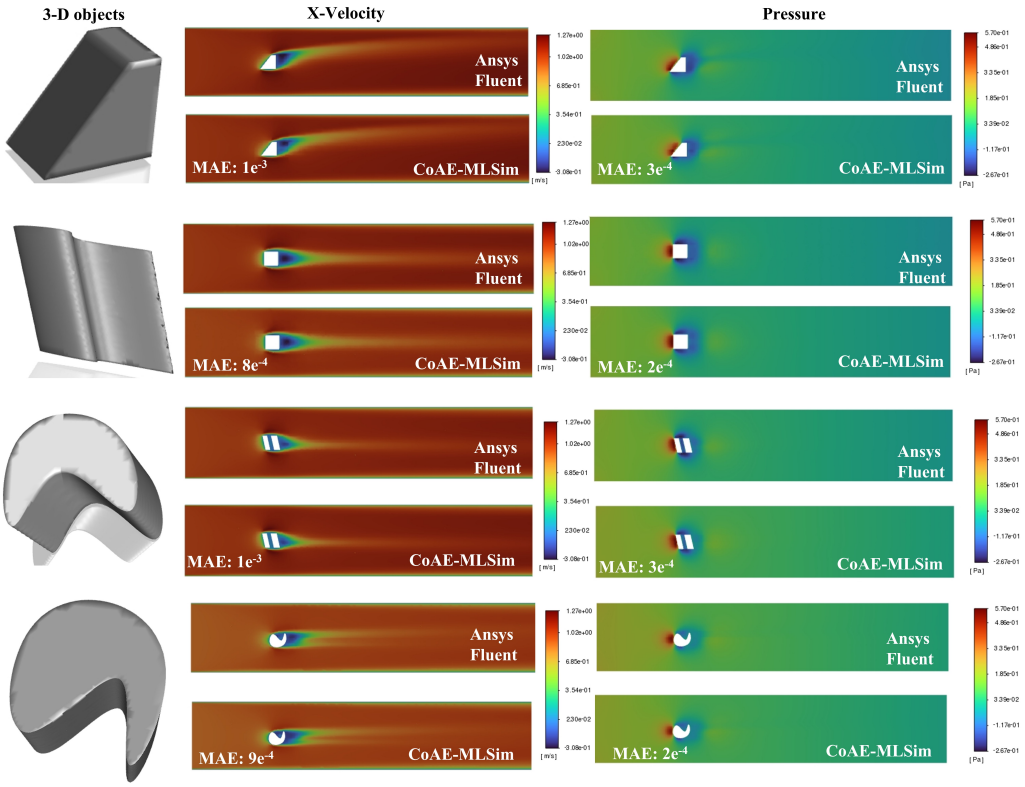

Figure 30: CoAE-MLSim vs Ansys Fluent comparisons for different objects

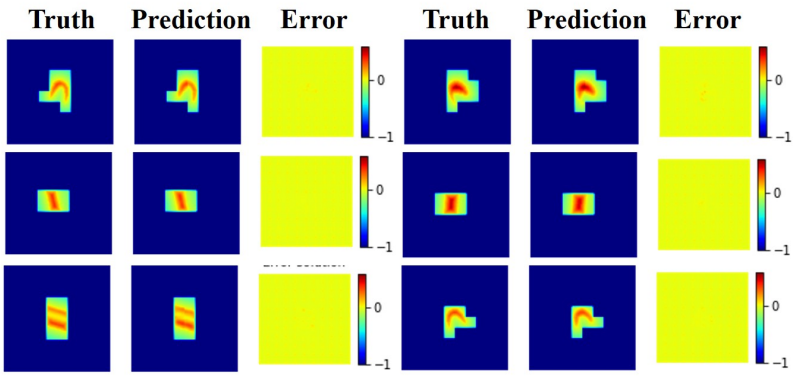

Figure 31: Geometry encoder generalization

solution autoencoder encodes vorticity on each subdomain into a latent vector of size 11. The flux conservation and time integration network are trained using the same data. The autoencoder networks combine to have a total of $400,000$ parameters. The time integration network is designed such that the input contains solution latent vectors on neighboring subdomains of 10 previous time steps to predict the latent vectors of the next time step. As a result, all the test runs start from the 10^{th} timestep. During each timestep integration during the solution inference, the flux conservation autoencoder is evaluated until an iterative convergence is achieved, similar to the steady state version of the CoAE-MLSim approach and this convergence is achieved to a specified tolerance of $1e^{-6}$ in about 2-4 iterations. As a result, the total evaluation time might be slower than other ML-approaches but is still faster than commercial PDE solvers by around 100x and moreover, it provides for a more stable and accurate transient dynamics.

In the main body of the paper, Section 4.3, as well as in this section we compare our approach with other baseline ML models such as UNet (Ronneberger et al., 2015) and Fourier Opertor Net (FNO) for this experiment and here we briefly explain the network architectures used.

UNet: Our UNet architecture includes 18 convolutional layers with channel sizes of 10 (input), 64, 64, 128, 128, 128, 128, 256, 256, 256, 512, 384, 384, 256, 256, 192, 192, 128, 64, 1 (output) with 9 before the bottleneck (encoder) and 9 after the bottleneck (decoder). Skip connections were used to connect layers before and after the bottleneck. In the encoder, the spatial resolution was reduced using four max-pooling layers with kernel size 2. In the decoder the original spatial resolution was recovered using four bilinear upsampling layers with kernel size 2. The total number of learnable parameters is 7.418M.

FNO: Our FNO architecture was that of the original implementation (Li et al., 2020a). The total number of learnable parameters is 465k.

Next, we compare results for 3 unseen initial conditions. It may be observed that the CoAE-MLSim predictions match well with the ground truth data and the error accumulation is acceptable, especially in the extrapolation range. It may be observed that our method outperforms FNO and UNet in terms of error accumulation.

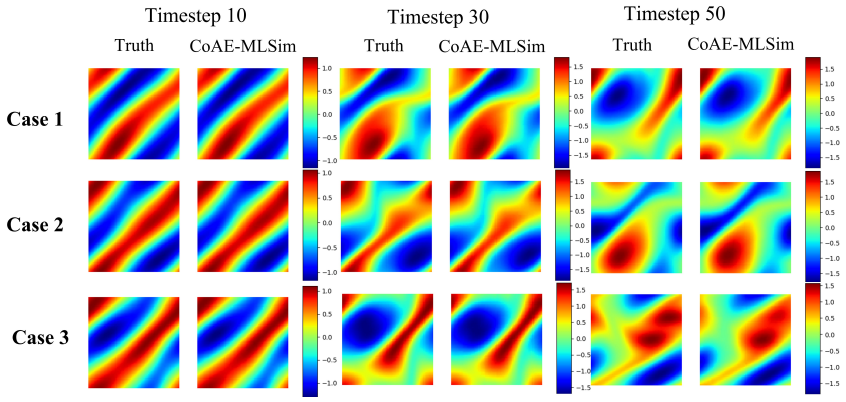

Figure 32: Contour comparisons of vortex decay at different time steps

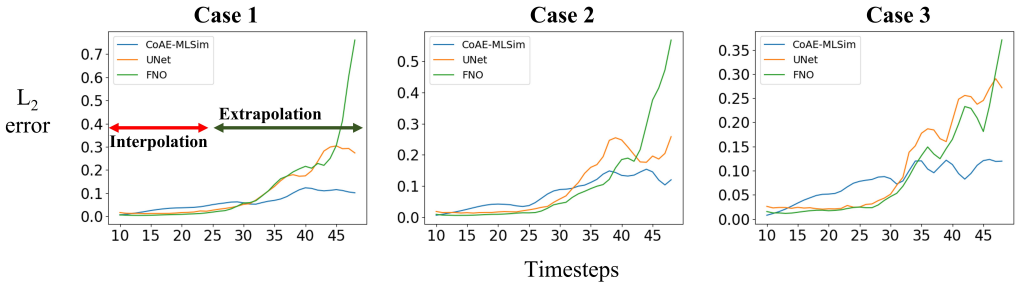

Figure 33: Error accumulation at different time steps

Finally, we perform an additional study where we average the predictions over 50 unseen testing samples from different initial conditions for each of the 50 timesteps. The results from this experiment are shown in Figure 34. It may be observed from the figure that our approach performs as good as FNO and UNet in the interpolation regime but outperforms both methods in the extrapolation regime by a large margin.

Additionally, in the table 6 we provide the average errors for the 3 approaches in the interpolation and extrapolation regions separately. For the interpolation regime the averages are computed over 50 testing samples and until 25 timesteps, while for extrapolation regime they are computed over 50 testing samples and for timesteps between 25 and 50.

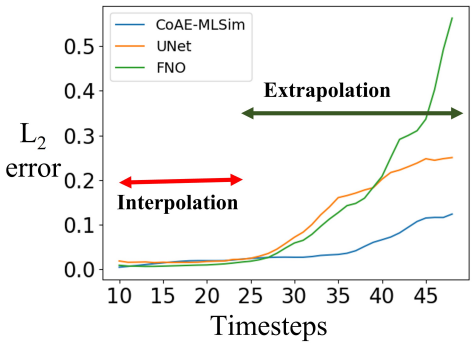

Figure 34: Error accumulation at different time steps averaged over 50 unseen test samples

Bottleneck Size (Compression ratio)	Testing error	Convergence iterations	Solve time
8 (42.5)	0.0241	15	0.23
16 (21.25)	0.0107	37	0.57
32 (10.62)	0.0067	45	0.706
64 (5.32)	0.0081	55	0.85
128 (2.65)	0.133	140	2.18
256 (1.32)	0.198	168	2.73

Table 6: Ablation study for different layer sizes

It may be observed that the CoAE-MLSim performs as well as FNO and UNet in the interpolation region but outperforms both of them for long time flow dynamics.

A.2.6 TRANSIENT: FLOW OVER A CYLINDER

Finally, we present a demonstration of the CoAE-MLSim approach in solving the flow around a cylinder problem in a transient setting. The case setup and geometry is similar to the case presented in Section A.2.4 except that the flow Reynolds number is much higher, equal to 200, in order to induce unsteady phenomenon in the flow, commonly known as vortex street. In this case, the CoAE-MLSim approach is trained on Reynolds number of 50 and 1000 and tested on 200. The timestep used for training is 20x larger than the one used by Ansys Fluent to generated the training and testing data. The complexity of the problem is increased by adding a constant heat flux to the cylinder, resulting in dissipation of temperature with the flow. The governing equations corresponding to this case are presented in 7:

$$\left.\begin{aligned} &\textbf{Continuity Equation:} && \nabla.\mathbf{v} = 0 \\ &\textbf{Momentum Equation:} && \frac{\partial v}{\partial t} + (\mathbf{v}.\nabla)\mathbf{v} + \nabla\mathbf{p} - \frac{1}{Re}\nabla^2\mathbf{v} = 0 \\ &\textbf{Energy Equation in Fluid:} && \frac{\partial T}{\partial t} + (\mathbf{v}.\nabla)\mathbf{T} - \nabla.(\alpha\nabla\mathbf{T}) = 0 \end{aligned}\right\} \quad (7)$$

where, $v = u_x, u_y, u_z$ is the velocity field in x, y, z, t, p is pressure, T is temperature, Re and α are flow and thermal properties.

Figs. 35 and 36 compare the vector plots on cut center planes for velocity magnitude and temperature with Ansys Fluent at 3 different time steps. The CoAE-MLSim approach captures the transient flow dynamics with an acceptable error in comparison to the Ansys Fluent. Furthermore, in Figure 37, we plot the total surface drag force of the cylinder as a function of time to evaluate the vortex shedding

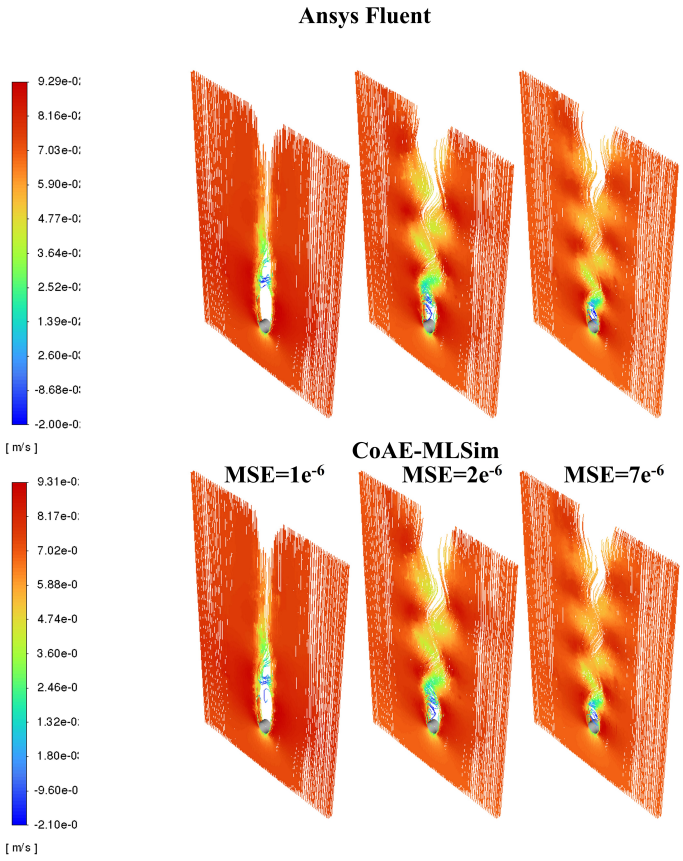

Figure 35: Transient flow comparisons of velocity magnitude

frequency of the flow. The CoAE-MLSim predictions are satisfactory and continues to predict well as the solution progresses ahead in time.

A.2.7 TRANSIENT: BURGER'S EQUATIONS

The 1-d Burger's equation is shown in Eq 8.

$$\frac{\partial u}{\partial t} + \nabla.(u^2/2) = \nu \nabla^2.(u) \tag{8}$$

It is subjected to different initial conditions, $u(\vec{x}) = u_o(\vec{x})$ in a 1-D bounded computational domain between (0, 1). The problem setup as well as the data to train the autoencoders is taken from Li et al. (2020a). The motivation of this experiment is to show the difference in performance between the two transient approaches mentioned in the paper: fully coupled and decoupled and validate the claim that the decoupled approach is better than the fully coupled approach in predicting long range time dynamics. The fully coupled transient approach is the one where space and time are coupled and in this example solved as a 2-D domain, where as the in the decoupled approach the time integration is handled separately from flux conservation as explained in the paper.

The 1-D computational domain has a resolution of 512 computational elements. The computational case in the case of fully coupled case is a 2d domain with 512 elements x 50 timesteps. We use 500 solutions and 50 timesteps per solution to train the autoencoders in both approaches. The 1-d domain in the decoupled case is divided into 16 subdomains with a resolution of 32 computational elements each. On the other hand, the fully coupled case is divided into 48 subdomains with a resolution of 32 elements vs 32 timesteps. In both cases, the testing is carried out for 30 unseen samples integrated all the way up to 200 timesteps, which is more than twice outside the training regime. The relative errors, from Eq. 4, averaged over the testing samples for each timestep in

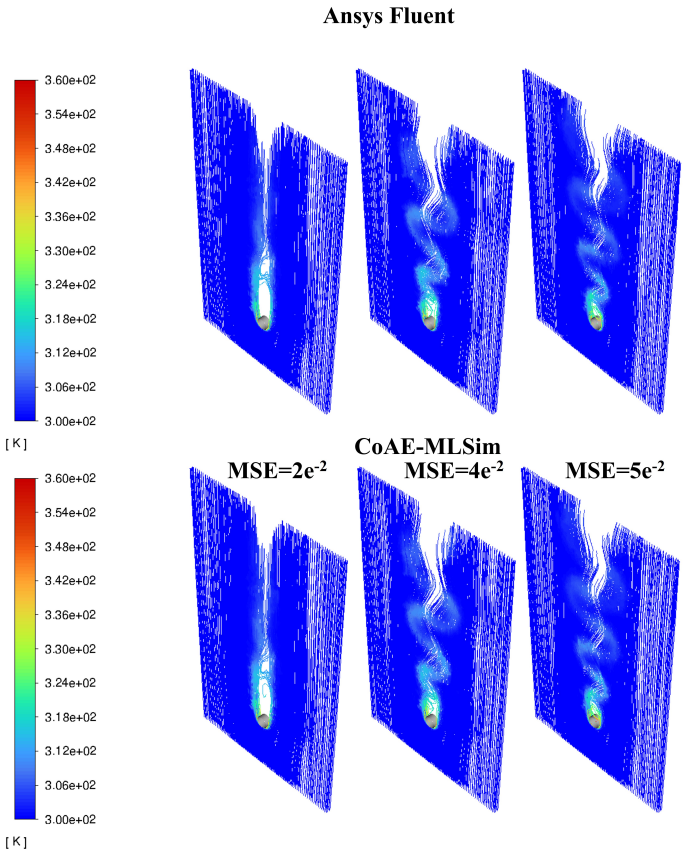

Figure 36: Transient flow comparisons of temperature

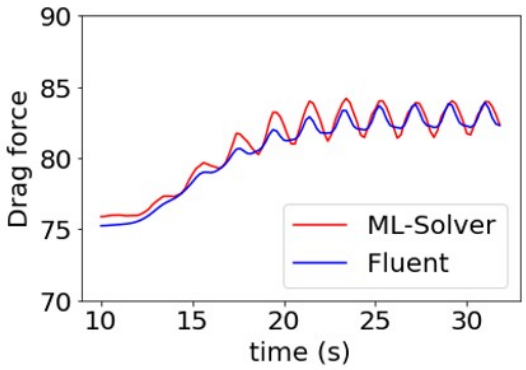

Figure 37: Total drag force comparison

the extrapolation regime (after 50 timesteps) are shown in Figure 38. It may be observed that the decoupled transient approach has a much smaller error accumulation than the fully coupled transient approach, therby validating our claim made in the paper.

A.3 Details for reproducibility

In this section, we provide the necessary details for training the CoAE-MLSim. As emphasized previously, the CoAE-MLSim training corresponds to training several autoencoders for PDE solutions, conditions, such as geometry, boundary conditions and source terms and for flux conservation. In

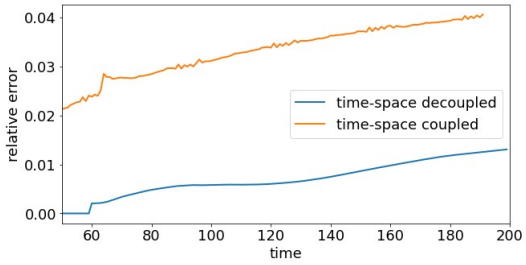

Figure 38: Comparison between fully coupled and decoupled transient approaches

Figure 39, we present a flow chart of the steps that can be followed to train each of these autoencoders. The specific training details and network architectures may be found in Section A.1.

For a given set of coupled PDEs, we start with 100-1000 sample solutions on a computational domain with n, m, p computational elements in spatial directions x, y, z, respectively. The solutions are divided into smaller subdomains of resolution, $n/16, m/16, p/16$. Each subdomain can has PDEs solutions and PDE conditions associated with it. The PDE solutions are used as training samples to the PDE solution autoencoder to learn a compressed encoding of all the variables on the subdomain. On the other hand, the autoencoders for PDE conditions can be trained with completely random samples, which may or may not be related to sample solutions. Once the PDE solution and condition autoencoders are trained, neighboring subdomains are grouped together and the solution and PDE condition latent vectors on groups of neighboring subdomains are stacked together. These groups of stacked latent vectors are used for training the flux conservation autoencoder. The trained PDE solution, condition and flux conservation autoencoders combine to form the primary components of the CoAE-MLSim.

The solution algorithm of the CoAE-MLSim has been described in detail in the main body of the paper and may be used for solving for unseen PDE conditions.

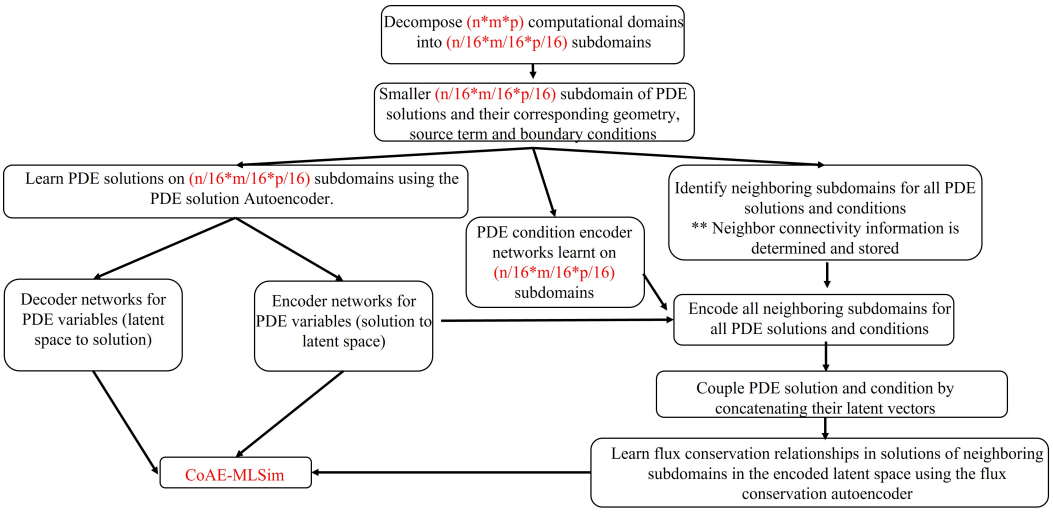

Figure 39: Flow chart for training the CoAE-MLSim approach