Predicting Eigenmode Decompositions in Vibroacoustic Systems

Christian Libner[1], **Jan van Delden**[1], **Julius Schultz**[2], **Alexander S. Ecker**[1,3], **Timo Lüddecke**[1]
[1]Institute of Computer Science and Campus Institute Data Science (CIDAS), University of Göttingen
[2]Institute for Acoustics and Dynamics, TU Braunschweig
[3]Max Planck Institute for Dynamics and Self-Organization, Göttingen
Correspondence: jan.vandelden@uni-goettingen.de

Abstract

Neural surrogate models can predict the dynamics of physical systems and generalize across different structures. However, they typically rely on generic representations that are not physically motivated. The dynamics of many physical systems, in contrast, can be efficiently described through a *mode superposition*. Accordingly, any system state can be decomposed into fundamental modes, where the influence of each mode is determined by a time-varying amplitude. Directly predicting such a decomposition with neural surrogate models could serve as an efficient and meaningful representation of the system's dynamical behaviour. In this work, we present an implementation of this idea – the Mode Operator Network (MODEONET). We apply it to predicting velocity fields of rectangular plates with indentation patterns, a dynamical system in which vibrations can be described through a superposition of eigenmodes of the plate. Our network learns the structure-dependent mode shapes and frequency dependent coefficients to construct the plate's velocity field. MODEONET achieves on-par accuracy with the state-of-the-art method, is 18 times faster, and allows for easier interpretability of the system's behaviour, which could enable future applications to design optimization. Code can be found here [1].

1 Introduction

Structural vibrations can affect material integrity and are prone to radiate noise. Therefore, engineers work on minimizing them in systems such as cars, trains or airplanes. Structural vibrations are typically numerically simulated via the finite-element method (FEM), which is compute-intensive. Performing many-query tasks like design space exploration, optimization or uncertainty quantification poses a high computational cost. To mitigate this high cost, researchers have explored using neural surrogate models, trading the expense of data generation and training for fast inference [1].

In this paper, we consider linear structural vibrations induced by a harmonic excitation. In such a system, the vibrations can be factorized into a spatial component (mode shapes) and a frequency-dependent component (modal frequency responses). The modes can be identified by resonance peaks in the system's frequency response function. If the modes and modal frequency responses are known, the velocity field for any excitation frequency can be computed via mode superposition [2]. We introduce the Mode Operator Network, a neural network architecture for predicting a factorized representation of physical system behaviour, where both the spatial component (mode shapes) and the frequency component (modal frequency responses) depend on the structure. We evaluate the MODEONET architecture on the vibrating plates benchmark dataset, which consists of 50,000 rectangular aluminum plates with varied indentations alongside their corresponding velocity fields

[1] https://github.com/ecker-lab/modeonet

1st Workshop on Differentiable Systems and Scientific Machine Learning @ EurIPS 2025.

given an harmonic excitation [1, 3]. MODEONET reaches on-par accuracy to the current state-of-the-art method, speeds up inference by more than an order of magnitude and facilitates interpretation.

2 Related work

Discovering low-dimensional patterns in the behaviour of systems is a long-time goal of data-driven engineering research. For single systems, powerful data-driven decomposition methods have been developed, such as proper orthogonal decomposition (POD) [4], and dynamic mode decomposition (DMD) [5], that aim to discover modes [6]. Some deep learning based extensions have been proposed to address issues such as complex domains or partial observations [7, 8]. These methods are constrained to a single structure or domain. In contrast, a range of deep learning methods have been developed that are able to predict field quantities for a variety of dynamical systems and generalize across structures. [1, 9–13]. Few methods aim to combine factorization as well as generalization across structures based on the DEEPONET framework [14], that factorizes the surrogate modelling task into separate networks for general geometry encoding and spatial/temporal querying (like predicting modal responses or mode shapes) [15–18]. However, in these methods either the predicted quantity was not factorized into a spatial and temporal part [15–17] or one of the factorization components is independent of the structure [18], limiting the transferability to vibroacoustic systems.

3 Physical Model

In the following, we give a brief overview of the physical model. Please refer to van Delden et al. [3] for a more detailed definition. We consider thin aluminum rectangular plates that are harmonically excited at a point $p \in \mathbb{R}^2$ by a force with an excitation frequency $\omega \in \mathbb{R}_+$. The plate is discretized by a mesh with N vertices. A beading pattern $\mathbf{g} \in \mathbb{R}^N$ is pressed into the plate. As a boundary condition, a rotational stiffness of $r \in \mathbb{R}$ is applied. The vibration characteristics are modelled with the Mindlin plate theory [19]. By solving a linear partial differential equation, we can derive the velocity field $v(\omega) \in \mathbb{C}^N$. Let us consider the i^{th} vertex of the mesh. It vibrates with an angular frequency ω, where the amplitude and direction of the velocity orthogonal to the plate are given by $v_i(\omega)$. Due to the linearity of the partial differential equation, we can represent the velocity field v by a mode superposition [2]:

$$v(\omega; \mathbf{g}, r, p) = \sum_{k=1}^{N} \Phi_k(\mathbf{g}, r)\, q_k(\omega; \mathbf{g}, r, p) \tag{1}$$

where $\Phi_k(\mathbf{g}, r) \in \mathbb{C}^N$ denotes the k^{th} mode shape, and $q_k(\omega; \mathbf{g}, r, p) \in \mathbb{C}$ are frequency dependent coefficients, in the following referred to as modal frequency response. Both the mode shapes Φ_k and the modal frequency response functions q_k depend on the geometry $\mathbf{g}$ (and boundary condition r), whereas only the modal frequency response depends on the excitation point p. The modal response corresponds to the shape of a harmonically driven oscillator and has a single peak at the eigenfrequency [2]. If a mode is excited, the eigenfrequency of this mode can be identified as a peak in the system frequency response function $L_v(\omega)$. The system frequency response is defined by the spatial average over the velocity field and a transformation to the decibel scale:

$$L_v(\omega; \mathbf{g}, r, p) = \log_{10} \|v(\omega; \mathbf{g}, r, p)\|^2 + \text{const} \tag{2}$$

4 Mode Operator Network

MODEONET predicts a set of K mode shapes $\hat{\Phi}(\mathbf{g}, r) \in \mathbb{C}^{N \times K}$ and their corresponding modal frequency response functions $\hat{q}(\omega; \mathbf{g}, p, r) \in \mathbb{C}^K$ (Fig. 1). The number of predicted modes, K, was chosen based on an empirical analysis of our dataset. Analytically, the velocity field contains as many modes as vertices in the mesh (denoted by N). However, within the data frequency range the solution is usually dominated by a small number of modes. When inspecting the system frequency responses in our dataset, we found a maximum of 12 distinct resonance peaks within the 1–300 Hz range. We therefore chose to predict $K = 16$ modes to provide a conservative buffer.

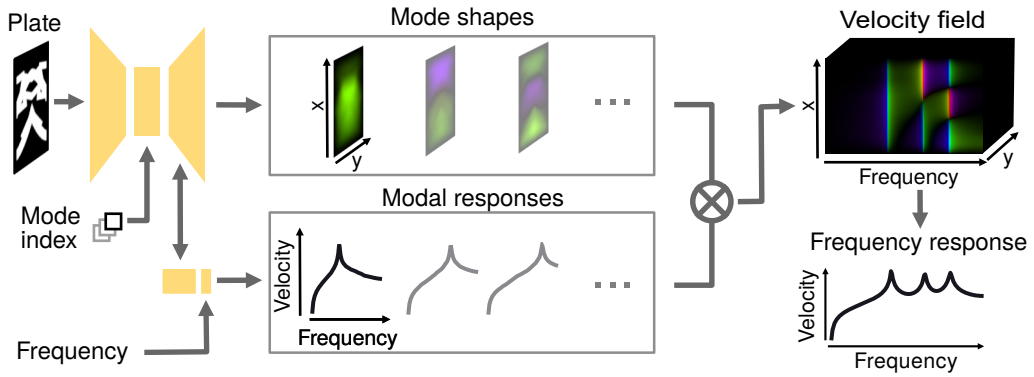

Figure 1: Our MODEONET predicts mode shapes and modal responses, from which the velocity field can be computed.

Architecture. MODEONET uses an encoder-decoder architecture based on the Frequency Query Operator (FQO) [1]. The encoder, identical to that of the FQO, takes the plate height map, excitation point, and rotational stiffness as input, encoding them into a latent feature representation. The decoder is conditioned on a mode index $k \in \{1, \dots, K\}$ via a FiLM layer [20] to predict a single mode, consisting of a mode shape $\hat{\Phi}_k$ and the corresponding modal response $\hat{q}_k$. To generate all K modes, the decoder is evaluated separately for all indices. The decoder consists of two interconnected branches. The mode shape branch, structured as a U-Net decoder, predicts the spatial mode shape $\hat{\Phi}_k$. Concurrently, the modal response branch is an MLP and predicts a feature vector representing the modal frequency response $\hat{q}_k$. This feature vector is then conditioned on a specific query frequency ω via a FiLM layer to predict the final value $\hat{q}_k(\omega)$. This design does not enforce the analytical structure of the modal frequency response functions q to provide more flexibility during the training. Both branches are interconnected, since mode shapes and eigenfrequencies are interdependent. The modal response branch conditions the mode shape branch via a FiLM layer, while the features of the modal response branch cross-attend to the feature volumes of the modal shape branch.

Loss. MODEONET learns to predict the mode shapes and modal frequency responses indirectly using a loss on the velocity field computed via Eq. 1 and an additional sparsity constraint loss. We do not provide ground truth mode shapes and responses during training. Our loss for the velocity field is inspired by the STFT loss in speech models [21]. We split the loss of reconstructing the velocity field into a magnitude $\mathcal{L}_{\text{mag}}$ and phase $\mathcal{L}_{\text{phase}}$ loss. In addition, we encourage the model to use the smallest possible number of modes by adding a sparsity loss $\mathcal{L}_{\text{sparse}}$, which is a group lasso on the modal frequency response functions. The final loss is given by:

$$\mathcal{L} = \lambda_{\text{mag}}\mathcal{L}_{\text{mag}}\left(\hat{v}, v\right) + \lambda_{\text{phase}}\mathcal{L}_{\text{phase}}\left(\hat{v}, v\right) + \lambda_{\text{sparse}}\mathcal{L}_{\text{sparse}}(\hat{q}) \tag{3}$$

where $\hat{v}^{(z)}$ are the predicted velocity fields, constructed with the predicted mode shapes $\hat{\Phi}$ and modal responses $\hat{q}$ via the mode superposition (Eq. 1). For the sparsity loss to be effective, we normalize the mode shapes, s.t. $\|\hat{\Phi}_i\| = 1$. A detailed definition of all loss terms is given in the Appendix A.

5 Experiments

Dataset. We make use of the vibrating plates benchmark dataset [1], in the version employed in [3]. It contains a training dataset of 50,000 plates with indentations, where for each plate 15 frequencies are sampled in an interval of 1 Hz to 300 Hz. The dataset includes variations in indentation patterns, excitation points, and rotational stiffness at the boundary of the plates. For testing, an additional dataset is available with 500 plates and full solutions from 1 Hz to 300 Hz.

Training. We train the model using the AdamW optimizer [22] for 500 epochs with a weight decay of 0.001. The learning rate is varied via cosine scheduling [23]. After a warm-up of 50 epochs, the learning rate decays from $5 \cdot 10^{-4}$ to 10^{-5}.

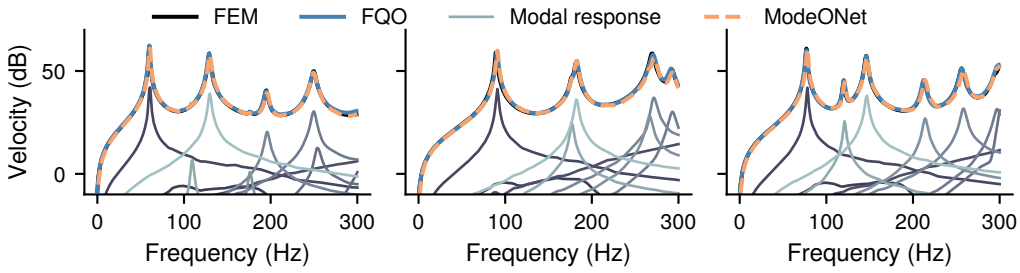

Figure 2: Frequency response curves with predictions from our method as well as the FQO method for three random samples. We observe a close alignment between ground truth (FEM) and prediction. Our MODEONET produces an excitation curve per mode, whose peaks directly match to the peaks in the ground truth frequency response function. The residuals to the FEM solution are shown in Fig. 5

Table 1: Test results for frequency response prediction. The peak error $\mathcal{E}_{\text{PEAKS}}$ quantifies whether the correct number of peaks is predicted, while the peak frequency distance error $\mathcal{E}_F$ measures the deviation between the predicted and actual peak locations. For detailed definitions, please refer to [1].

Model	$\mathcal{E}_{\text{MSE}}$	$\mathcal{E}_{\text{PEAKS}}$	$\mathcal{E}_{\text{F}}$	FLOPs	FLOPs ratio
FQO	0.028	0.037	1.16	$1.4 \cdot 10^{12}$	18x
MODEONET	0.028	0.043	1.04	$7.7 \cdot 10^{10}$	1x

Evaluation. We compare our results to the current state-of-the art method, the Frequency Query Operator (FQO) van Delden et al. [1, 3], which directly predicts velocity fields instead of a modal decomposition. To assess predictive performance, we compute the mean squared error (MSE) between the predicted and ground-truth frequency response functions. The data is normalized such that an MSE of 1 can be achieved by always predicting the mean frequency response. To estimate the number of floating point operations (FLOPs) required, we predict with both models for one plate, the velocity field for 300 frequencies. We calculate the FLOPs with pytorch's internal FLOPs counter.

6 Results

ModeONet matches FQO performance. We first investigate the prediction accuracy of MODEONET on the vibrating plates dataset by comparing to the state-of-the-art model, the FQO [1]. MODEONET achieves comparable accuracy to FQO (Table 1), demonstrating that the factorization does not compromise predictive quality. MODEONET requires on the order of 18 times less FLOPs for predicting a full frequency response functions than FQO. This is the case since our method directly produces the low dimensional mode decomposition, while the FQO method needs to produce a prediction for every single frequency separately.

Interpretability. Next, we take a closer look at the learned factorization. In a classical eigenmode decomposition, the modal frequency response functions resemble the response of a harmonically driven oscillator with a single peak. We observe that the learned modal frequency response functions follow this expected shape (Fig. 2, Appendix, Fig. 3). The eigenfrequencies of a system, which correspond to the peaks in the modal response function, are independent of the position of the excitation point. We investigate whether MODEONET has learned this relationship by evaluating the model for different excitation points for a single geometry. As expected, the peak positions remain unchanged, while only their amplitudes vary (Appendix, Fig. 4).

We further observe that MODEONET predicts modal frequency responses with very small magnitudes, which have little effect on the overall system response. Some of these still show distinct peaks. We hypothesize that these peaks correspond to eigenfrequencies of unexcited modes. When we vary the excitation point, these peaks appear in the system frequency response (Appendix, Fig. 4). This indicates that the model also predicts unexcited modes.

Author contributions. **Christian Libner:** Writing – original draft, Visualization, Software, Conceptualization, Methodology. **Jan van Delden:** Writing – original draft, Supervision, Conceptualization, Methodology. **Julius Schultz:** Writing - review & editing, Methodology. **Alexander S. Ecker:** Writing - review & editing. **Timo Lüddecke:** Supervision, Writing - review & editing.

Acknowledgements. This research is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 501927736, within the DFG Priority Programme 2353: Daring More Intelligence - Design Assistants in Mechanics and Dynamics'. The authors gratefully acknowledge the computing time granted by the Resource Allocation Board and provided on the supercomputer Emmy/Grete at NHR-Nord@Göttingen as part of the NHR infrastructure. The calculations for this research were conducted with computing resources under the project nii00194.

References

[1] Jan van Delden, Julius Schultz, Christopher Blech, Sabine Langer, and Timo Lüddecke. Learning to predict structural vibrations. *Advances in Neural Information Processing Systems*, 37:103706–103736, 2024.

[2] Franz G Kollmann. *Maschinenakustik: Grundlagen, Meßtechnik, Berechnung, Beeinflussung*. Springer-Verlag, 2011.

[3] Jan van Delden, Julius Schultz, Sebastian Rothe, Christian Libner, Sabine C Langer, and Timo Lüddecke. Minimizing structural vibrations via guided flow matching design optimization. *arXiv preprint arXiv:2506.15263*, 2025.

[4] Lawrence Sirovich. Turbulence and the dynamics of coherent structures. i. coherent structures. *Quarterly of applied mathematics*, 45(3):561–571, 1987.

[5] Peter J Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of fluid mechanics*, 656:5–28, 2010.

[6] Steven L Brunton and J Nathan Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022. URL `https://www.databookuw.com/`.

[7] Shaowu Pan, Steven L Brunton, and J Nathan Kutz. Neural implicit flow: a mesh-agnostic dimensionality reduction paradigm of spatio-temporal data. *Journal of Machine Learning Research*, 24(41):1–60, 2023.

[8] Ali SaraerToosi, Renbo Tu, Kamyar Azizzadenesheli, and Aviad Levis. Neural dynamic modes: Computational imaging of dynamical systems from sparse observations. *arXiv preprint arXiv:2507.03094*, 2025.

[9] Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, et al. Geometry-informed neural operator for large-scale 3d pdes. *Advances in Neural Information Processing Systems*, 36:35836–35854, 2023.

[10] Mohamed Elrefaie, Angela Dai, and Faez Ahmed. Drivaernet: A parametric car dataset for data-driven aerodynamic design and graph-based drag prediction. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 88360, page V03AT03A019. American Society of Mechanical Engineers, 2024.

[11] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter Battaglia. Learning mesh-based simulation with graph networks. In *International conference on learning representations*, 2020.

[12] Chris Choy, Alexey Kamenev, Jean Kossaifi, Max Rietmann, Jan Kautz, and Kamyar Azizzadenesheli. Factorized implicit global convolution for automotive computational fluid dynamics prediction. *arXiv preprint arXiv:2502.04317*, 2025.

[13] Louis Serrano, Lise Le Boudec, Armand Kassaï Koupaï, Thomas X Wang, Yuan Yin, Jean-Noël Vittaut, and Patrick Gallinari. Operator learning with neural fields: Tackling pdes on general geometries. *Advances in Neural Information Processing Systems*, 36:70581–70611, 2023.

[14] 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, 2021.

[15] Junyan He, Seid Koric, Shashank Kushwaha, Jaewan Park, Diab Abueidda, and Iwona Jasiuk. Novel deeponet architecture to predict stresses in elastoplastic structures with variable complex geometries and loads. *Computer Methods in Applied Mechanics and Engineering*, 415:116277, 2023.

[16] Junyan He, Seid Koric, Diab Abueidda, Ali Najafi, and Iwona Jasiuk. Geom-deeponet: A point-cloud-based deep operator network for field predictions on 3d parameterized geometries. *Computer Methods in Applied Mechanics and Engineering*, 429:117130, 2024.

[17] Ahmad Peyvan, Varun Kumar, and George Em Karniadakis. Fusion-deeponet: A data-efficient neural operator for geometry-dependent hypersonic and supersonic flows. *arXiv preprint arXiv:2501.01934*, 2025.

[18] Varun Kumar, Jing Bi, Cyril Ngo Ngoc, Victor Oancea, and George Em Karniadakis. Learning nonlinear responses in pet bottle buckling with a hybrid deeponet-transolver framework. *arXiv preprint arXiv:2509.13520*, 2025.

[19] RD0044 Mindlin. Influence of rotatory inertia and shear on flexural motions of isotropic, elastic plates. 1951.

[20] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[21] Shinji Takaki, Toru Nakashika, Xin Wang, and Junichi Yamagishi. Stft spectral loss for training a neural speech waveform model. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7065–7069. IEEE, 2019.

[22] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[23] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

[24] Qianxao Li, Felix Dietrich, Erik M. Bollt, and Ioannis G. Kevrekidis. Extended dynamic mode decomposition with dictionary learning: A data-driven adaptive spectral decomposition of the koopman operator. *Chaos*, 27 10:103111, 2017. URL `https://api.semanticscholar.org/CorpusID:41957686`.

Appendix

A Loss Definitions

Learning a sparse factorization is closely related to dictionary learning (or sparse coding). In this setting, the mode shapes act as the dictionary, and the modal frequency responses act as the atoms [24]. The dictionary is fixed for a single plate. However, because the plates vary across the dataset, the dictionary also changes from sample to sample. This differs from the classical dictionary learning formulation, where the dictionary remains constant.

As described in Section 4, the reconstruction loss is divided into two components: a magnitude loss and a phase loss. This allows us to compute the mean squared error (MSE) between the logarithm of the magnitude of the predicted velocity field $\hat{v}(\omega; \mathbf{g}, p) \in \mathbb{C}^N$ and the ground-truth velocity field $v(\omega; \mathbf{g}, p) \in \mathbb{C}^N$:

$$\mathcal{L}_{\text{mag}}(\hat{v}, v) = \mathbb{E}_{\mathbf{g},p,\omega}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\log\left(|\hat{v}_i^{(z)}(\omega; \mathbf{g}, p)|^2 + \epsilon\right) - \log\left(|v_i^{(z)}(\omega; \mathbf{g}, p)|^2 + \epsilon\right)\right)^2\right] \tag{4}$$

with $\epsilon > 0$. Our evaluation metric is the mean squared error between the frequency response functions, which are measured in the logarithmic decibel scale. The logarithm in the magnitude loss ensures that the loss operates on a similar scale as this evaluation metric.

For small magnitude values, the phase is not meaningful. Therefore, instead of computing the cosine similarity per vertex (i.e., per pixel), we compute it over the entire spatial domain:

$$\mathcal{L}_{\text{phase}}(\hat{v}, v) = \mathbb{E}_{\mathbf{g},p,\omega}\left[\text{Re}\left(\frac{\hat{v}^T(\omega; \mathbf{g}, p)v^*(\omega; \mathbf{g}, p)}{\|\hat{v}(\omega; \mathbf{g}, p)\|\,\|v(\omega; \mathbf{g}, p)\|}\right)\right] \tag{5}$$

where v^* denotes the complex conjugate of v. This term encourages the predicted and ground-truth velocity fields at the same frequency to align in phase.

The network should explain the velocity field of a single plate using as few modes as possible. Our sparsity loss is applied to the modal frequency response function. Instead of the classical ℓ_1 sparsity used in dictionary learning, we employ a group lasso regularizer, which drives the entire predicted modal response $\hat{q}_k$ for a given plate towards zero. Again, we apply a logarithm to keep the loss on the same scale as the log-magnitude loss:

$$\mathcal{L}_{\text{sparse}} = \mathbb{E}_{\mathbf{g},p}\left[\sum_{k=1}^{K}\log\left(\sqrt{\mathbb{E}_\omega\left[\hat{q}_k^2(\omega; \mathbf{g}, p)\right]} + 1\right)\right] \tag{6}$$

Combining all these terms yields the final loss function (Eq. 3):

$$\mathcal{L} = \lambda_{\text{mag}}\mathcal{L}_{\text{mag}} + \lambda_{\text{phase}}\mathcal{L}_{\text{phase}} + \lambda_{\text{sparse}}\mathcal{L}_{\text{sparse}} \tag{7}$$

where we choose the weights as $\lambda_{\text{mag}} = 1$, $\lambda_{\text{phase}} = 0.1$, and $\lambda_{\text{sparse}} = 0.002$.

B Additional Figures

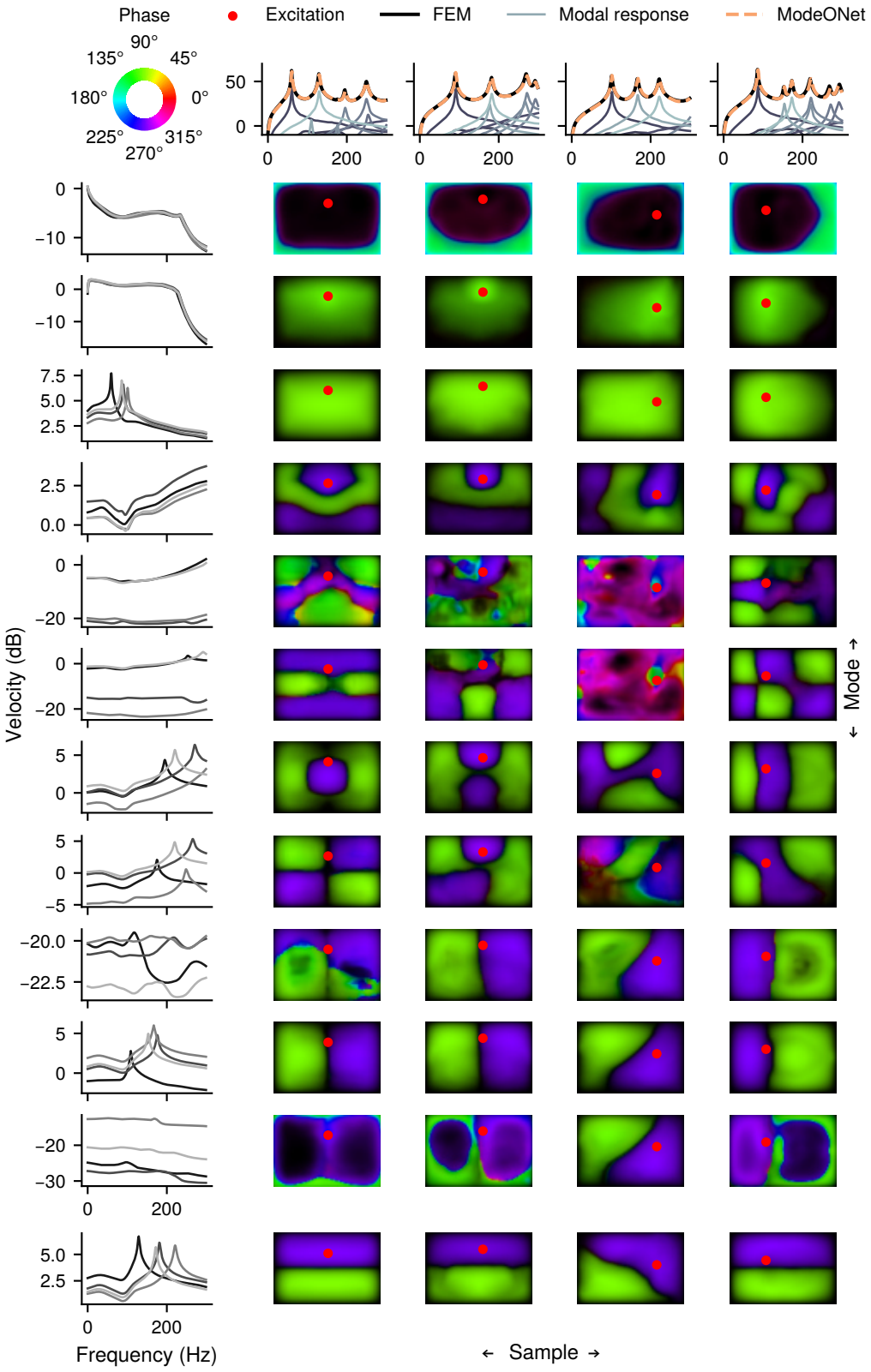

Figure 3: A subset of predicted modes

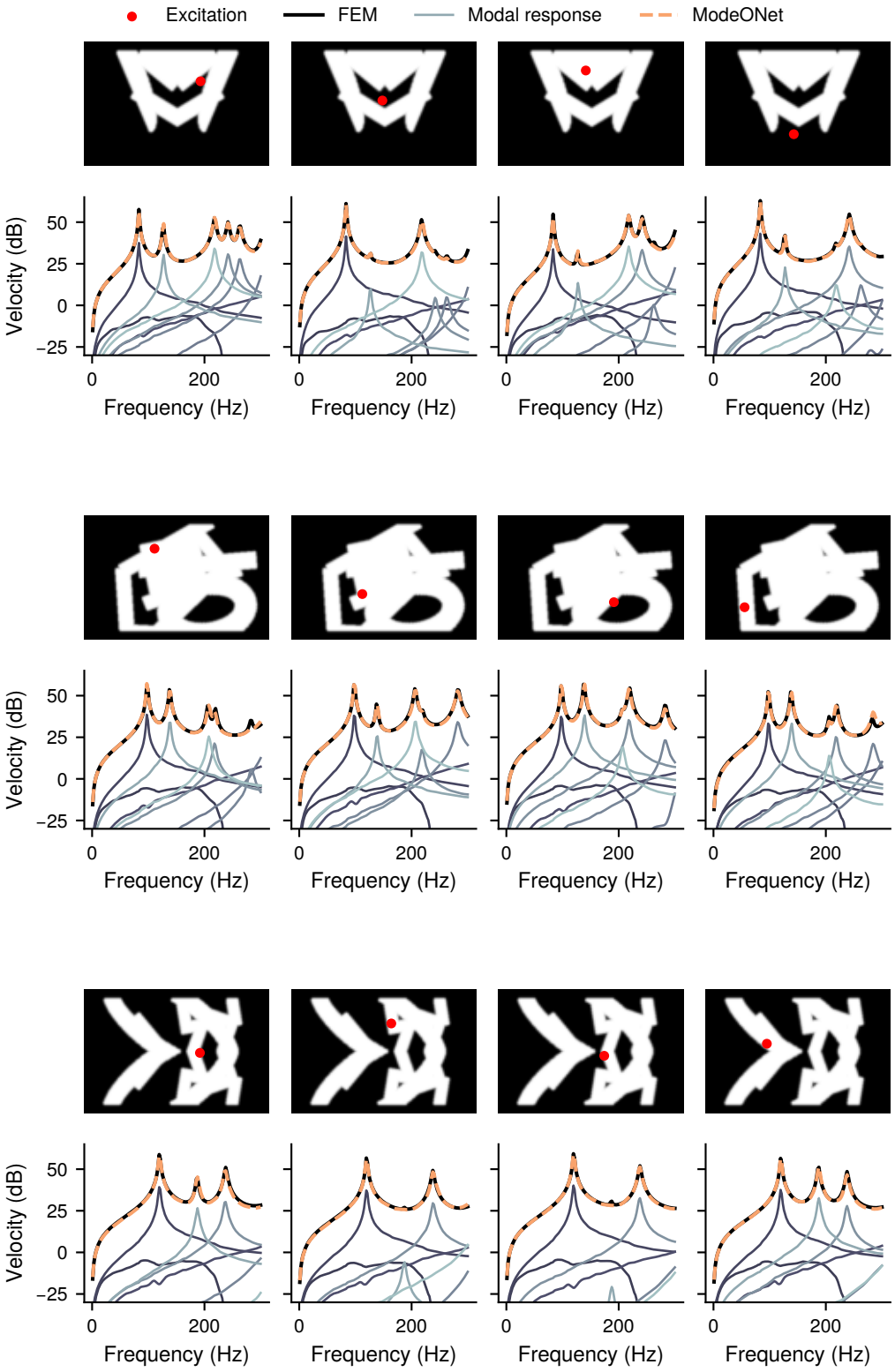

Figure 4: Variation of the excitation point

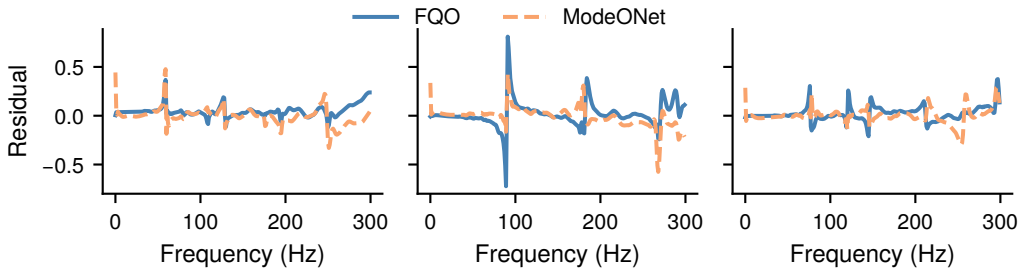

Figure 5: The normalized residual for the MODEONET and FQO. The normalization is the same as used for the MSE error $\mathcal{E}_{\text{MSE}}$