Archives of Acoustics Vol. ***, No. ***, https://doi.org/10.24423/archacoust.2026.4403

Online First April 27, 2026

Research Paper

Causality- and Passivity-Constrained Nonnegative Attention for Interpretable Structure-Borne Road Noise Prediction in Battery Electric Vehicles

Haijun Wang[(1),(2)*], Zhijie Huang[(1)], Zengjun Lu[(1)], Xianghua He[(3)], Tie Xu[(4),(5)]

[(1)] *School of Railway Locomotive and Vehicle, Liuzhou Railway Vocational and Technical College*
Liuzhou, China

[(2)] *Technical Center, Liuzhou Yingqin Tuolan Automobile Technology Co., Ltd.*
Guangxi, China

[(3)] *School of Physics, Electronics and Intelligent Manufacturing, Huaihua University*
Hunan, China

[(4)] *State Key Laboratory of Light Superalloys, Wuhan University of Technology*
Wuhan, China

[(5)] *Technical Development Center, SAIC-GM-Wuling Automobile Co., Ltd.*
Guangxi, China

*Corresponding Author: whjun69@sina.com

Received December 27, 2025; accepted March 11, 2026;
available online March 19, 2026; version of record April 27, 2026; published issue XXXX.

In battery electric vehicles (BEVs), structure-borne road noise in the 20 Hz to 300 Hz band becomes more audible because the engine-masking component is largely absent, and conventional transfer-path formulations can be sensitive to suspension nonlinearity and ill-conditioned inversions. This paper presents a physics-informed, non-negative multi-modal fusion network (NN-MMFNet) that predicts in-cabin sound pressure from multi-point chassis excitations while keeping the mapping physically plausible and interpretable. The model combines a dual-stream encoder to separate transient impact signatures from steady resonance content with a strictly causal fusion/decoding pathway. A passivity-motivated spectral gain cap is applied to prevent non-physical amplification while preserving phase. To enable additive path attribution, the cross-modal attention weights are constrained to be non-negative. Training follows a sim-to-real workflow, using virtual-fleet pretraining and short fine-tuning on measured data. On a production BEV, NN-MMFNet reproduces the 20 Hz to 300 Hz spectrum with a 1.12 dB(A) global root mean square error (RMSE) at 60 km/h and a 0.14 dB error at the 128 Hz boom, outperforming transfer path analysis (TPA), frequency transfer matrix (FTM), and autoregressive moving average (ARMA) baselines. Impulse-response checks show a negligible passivity-violation rate (<0.01 %). The learned attention consistently points to a rear subframe-to-body mounting path near 128 Hz, and a targeted stiffness adjustment at this location reduces the measured cabin noise by 4.2 dB(A).

Keywords: structure-borne road noise, physics-informed neural networks (PINN), transfer path analysis (TPA), cross-modal attention, battery electric vehicles (BEVs).

Copyright © 2026 The Author(s).
This work is licensed under the Creative Commons Attribution 4.0 International CC BY 4.0
(https://creativecommons.org/licenses/by/4.0/).

Notations

$\mathbf{b}_m$, $\mathbf{a}_m$ – structural and acoustic mode shapes of the m-th mode,
$\mathbf{B}$ – input distribution matrix,
$\mathbf{H}_s(\omega)$ – excitation-to-pressure frequency-response / learned transfer matrix,
$\mathbf{M}$, $\mathbf{C}$, $\mathbf{K}$ – mass, damping, and stiffness matrices,

$\mathbf{p}_s(t)$ – interior sound pressure response,
$\mathbf{R}, \mathbf{L}$ – velocity-to-pressure and displacement-to-pressure radiation and leakage operators,
$\mathbf{u}(t)$ – multi-point excitation vector at the suspension/body attachment points,
$\mathbf{q}(t)$ – generalized displacement vector of the vehicle body structure,
$h(t)$ – impulse response of the learned excitation-to-pressure mapping,
$\gamma(\omega_l)$ – passivity-motivated spectral gain cap,
ϵ – small positive constant used for numerical stabilization,
ζ_m – modal damping ratio,
$\sigma_{\max}(\cdot)$ – maximum singular value,
ω_m – resonance frequency of the m-th mode,
Ω – evaluation frequency band used in the spectral metrics.

1. Introduction

The move from combustion engines to electrified powertrains changes the in-cabin acoustic baseline and makes structure-borne road noise a major noise, vibration, and harshness (NVH) concern (Masri *et al.*, 2024; Khan, Burdzik, 2023). In battery electric vehicles (BEVs), the reduced masking effect means that tire–road induced vibrations transmitted through the suspension and body attachments are readily perceived (Münder, Carbon, 2022). The most relevant content is typically in the 20 Hz to 300 Hz band, where lightly damped structural resonances and cavity-panel coupling produce booming and rumble (Ortega Almirón *et al.*, 2022). Consequently, small changes in mounts, bushings, or subframe–body interfaces can cause noticeable differences in sound quality under speed-dependent excitations (Mohammadi, 2023). Platform sharing and production scatter therefore call for prediction models that remain accurate across operating conditions while still providing path-level guidance for targeted countermeasures with limited test effort (Zhang *et al.*, 2024).

Over the last decade, road-noise research has developed along two main lines: transfer-path-based physics modelling and data-driven surrogate prediction (van der Seijs *et al.*, 2016; Masri *et al.*, 2024). Transfer path analysis (TPA) and related force-transfer methods are widely used because they decompose the interior response into physically meaningful source–path–receiver contributions and support design decisions at mounts and interfaces (de Klerk, Ossipov, 2010). In-situ blocked-force approaches improve portability by reducing sensitivity to boundary conditions, and operational formulations lessen the need for controlled excitations (Moorhouse *et al.*, 2009; Ortega Almirón *et al.*, 2022). Substructuring and simulation-test integration can further reduce test effort by reusing component models (de Klerk, Rixen, 2010). However, for electrified platforms, several limitations remain. First, elastomer joints and bushings exhibit amplitude-dependent behaviour, and the transfer dynamics drift with speed and load, so a fixed linear time-invariant transfer matrix may be inadequate near lightly damped resonances (Khan, Burdzik, 2023; Mohammadi, 2023). Second, inverse steps in TPA rely on (pseudo-)inversion and can become ill-conditioned because of sensor collinearity, limited excitation diversity and dense modal overlap, amplifying measurement noise and creating non-physical artefacts (Cheng *et al.*, 2016). Regularisation helps but remains sensitive to operating changes (Kong *et al.*, 2025; Gao *et al.*, 2024). Third, strong phase coherence across multiple paths leads to cancellation/reinforcement, so magnitude-based path ranking can be unstable (Cheng *et al.*, 2020; 2022). Recent learning-based predictors have reported good numerical accuracy in some settings, but without physical constraints they may violate causality or energy consistency and often provide limited support for path-level decision-making (Jia *et al.*, 2024; Ma *et al.*, 2025; Yang *et al.*, 2025; Zhu *et al.*, 2024). A framework that improves prediction accuracy while explicitly enforcing causality, energy consistency and transparent additive contributions under operating variability is still needed (Park, Kang, 2024; Raissi *et al.*, 2019).

To address these issues, we develop a physics-informed non-negative multi-modal fusion network (NN-MMFNet) that follows the vibro-acoustic transmission chain in a forward, causal manner (Raissi *et al.*, 2019). The architecture includes:

– a dual-stream encoder that separates transient impact features from steady resonance content,
– a strictly causal fusion/decoder with a passivity-motivated spectral gain cap implemented by phase-preserving amplitude shrinkage (Gustavsen, Semlyen, 1999; 2001),

– a non-negative cross-modal attention module that yields additive (cancellation-free) contribution estimates across excitation channels (Huang *et al.*, 2023).

We validate the approach using a simulation-to-experiment workflow with virtual-fleet pretraining and fine-tuning on full-vehicle measurements from a production BEV. In the 20 Hz to 300 Hz band at 60 km/h, NN-MMFNet achieves 1.12 dB(A) global root mean square error (RMSE) and captures the 128 Hz boom with a 0.14 dB peak error, outperforming TPA, frequency transfer matrix (FTM), and autoregressive moving average (ARMA) baselines. The attention map indicates a dominant rear subframe coupling near 128 Hz, which guides a stiffness adjustment that delivers a measured 4.2 dB(A) cabin-noise reduction.

2. Theoretical framework: Vibro-acoustic dynamics and physical constraints

2.1. Structural-acoustic coupling mechanisms and modal superposition

Let $\mathbf{q}(t) \in \mathbb{R}^{n_q}$ be the generalized displacement vector of the vehicle body structure, and $\mathbf{u}(t) \in \mathbb{R}^{M}$ be the multi-point excitation vector (such as force or acceleration at suspension attachment points). The system's dynamic behavior follows the second-order differential equation:

$$\mathbf{M}\ddot{\mathbf{q}}(t) + \mathbf{C}\dot{\mathbf{q}}(t) + \mathbf{K}\mathbf{q}(t) = \mathbf{B}\mathbf{u}(t), \tag{1}$$

where $\mathbf{M}, \mathbf{C}, \mathbf{K} \in \mathbb{R}^{n_q \times n_q}$ are the mass, damping, and stiffness matrices, respectively, and $\mathbf{B}$ is the input distribution matrix. The interior sound pressure $\mathbf{p}_s(t)$ is generated by structural vibration through radiation and leakage operators $\mathbf{R}$ (velocity-pressure) and $\mathbf{L}$ (displacement-pressure), respectively:

$$\mathbf{p}_s(t) = \mathbf{R}\dot{\mathbf{q}}(t) + \mathbf{L}\mathbf{q}(t). \tag{2}$$

In the frequency domain ω, the structure-borne transfer function $\mathbf{H}_s(\omega)$ can be approximated via modal superposition. For lightly damped systems, the response near a resonance frequency ω_m is primarily determined by the poles:

$$\mathbf{H}_s(\omega) \approx \sum_{m=1}^{n_m} \frac{\mathbf{a}_m \mathbf{b}_m^{\top}}{\omega_m^2 - \omega^2 + j2\zeta_m\omega_m\omega}, \tag{3}$$

where $\mathbf{b}_m$ and $\mathbf{a}_m$ represent the structural mode shape and acoustic mode shape, respectively, and ζ_m is the modal damping ratio. This physical prior is crucial for the design of deep learning models, revealing that the model must possess the capability to capture narrow-band resonance peaks (such as the 34 Hz suspension mode and 128 Hz cavity mode) and their phase behavior.

2.2. Operating-dependent dynamics and physics-informed constraints

Road-noise transfer dynamics are operating-dependent, so a single static TPA transfer matrix is inadequate. NN-MMFNet learns an attention-adaptive mapping and is regularized by causality and passivity to ensure physically plausible predictions.

Specifically, to ensure that the learned excitation–response $\mathbf{P}(k, \omega)$ mapping remains physically plausible, two training constraints are imposed: (a) strict time-domain causality and (b) passivity-inspired energy consistency in the frequency domain. Causality is enforced by applying a strict causal mask $\mathcal{M}(k, \omega_l)$ to the decoder transfer kernel $D_\theta\{\cdot\}$:

$$\mathbf{P}(k, \omega_l) = D_\theta\{\mathbf{U}(\leq k, \cdot)\} \odot \mathcal{M}(k, \omega_l), \tag{4}$$

where k is the short-time Fourier transform (STFT) frame index, $\mathbf{U}(\leq k, j)$ is the multi-channel structural excitation history up to frame k, and the symbol $\odot$ denotes element-wise multiplication. This ensures that the resulting time-domain mapping does not depend on future inputs and remains bounded-input bounded-output (BIBO) stable under bounded excitation. Passivity-inspired energy consistency is enforced by bounding the frequency-domain gain:

$$\sigma_{\max}(\mathbf{H}_\theta(\omega_l)) \leq \gamma(\omega_l), \tag{5}$$

where $\mathbf{H}_\theta(\omega_l)$ is the learned frequency-response matrix at ω_l, $\sigma_{\max}(\cdot)$ is the maximum singular value, and $\gamma(\omega_l)$ is the passivity-motivated spectral gain cap. In practice, the bound is implemented via phase-preserving amplitude shrinkage:

$$\mathbf{H}_\theta(\omega_l) \leftarrow \mathbf{H}_\theta(\omega_l) \cdot \min\left(1, \frac{\gamma(\omega_l)}{\sigma_{\max}(\mathbf{H}_\theta(\omega_l))}\right). \tag{6}$$

The operation rescales amplitudes while preserving phase, which is essential for coherent multi-path superposition.

3. Architectural analysis of physics-informed NN-MMFNet

3.1. Dual-stream encoder: Separation of transient and steady-state features

The network uses a time–frequency dual-stream encoder to capture transient impacts and steady resonances. As shown in Fig. 1, NN-MMFNet follows a left-to-right signal flow consisting of dual-stream encoding, causal fusion, and a physics-constrained decoder.

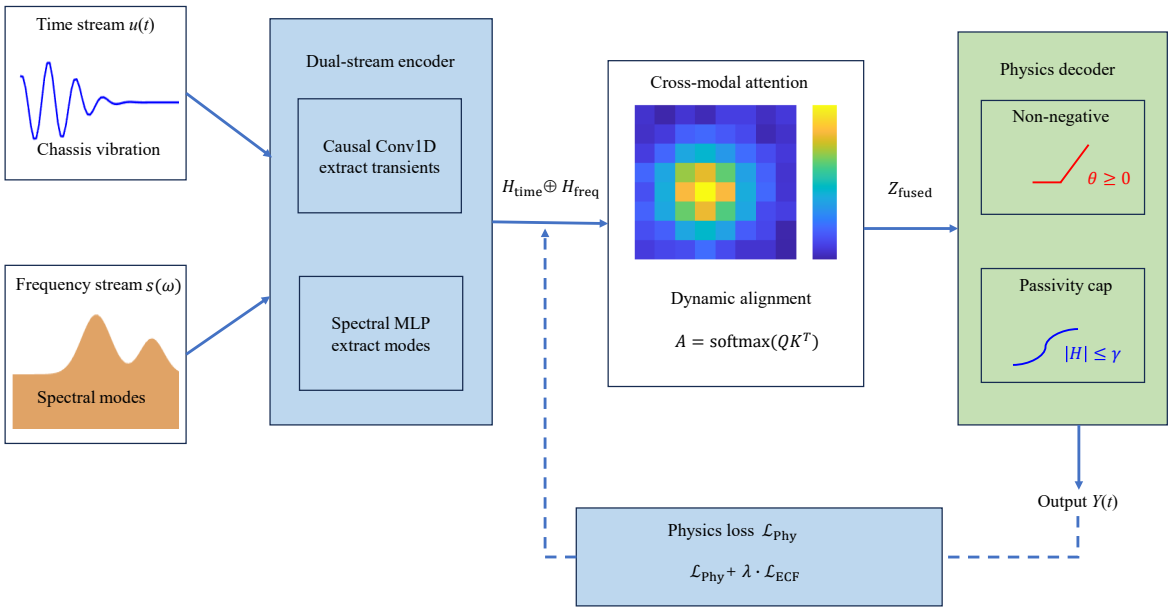

Fig. 1. Architecture of the physics-informed NN-MMFNet for BEV road noise.

The model uses a dual-stream encoder: a time stream built on 1D dilated convolutions with causal padding to capture long-range temporal dependencies without sacrificing resolution (Kiranyaz *et al.*, 2021), and a frequency stream that applies the STFT to extract 20 Hz to 300 Hz spectral features and can incorporate pre-trained noise transfer functions (NTFs) as a physics-informed prior to guide faster and more physically plausible convergence.

3.2. Cross-modal attention mechanism: Physical interpretation of energy allocation

A key element of NN-MMFNet is the cross-modal attention module. In this setting, the attention weight matrix has a direct physical interpretation: it represents the relative contribution of each excitation channel (or mode group) to the response under the current operating condition:

$$\mathbf{A}_{kl} = \text{softmax}\left(\frac{\mathbf{q}_k \cdot \mathbf{k}_l^\top}{\sqrt{d}}\right) \odot \mathbf{M}_{\text{causal}}, \tag{7}$$

where $\mathbf{q}_k$ and $\mathbf{k}_l$ are the query and key vectors, and d is the feature dimension. For interpretability and physical consistency, the attention module is constrained to be strictly causal, regularized toward a band-adjacent (near-banded) distribution to reflect continuous energy transmission, and allowed to adapt its weights under operating changes so that the effective transfer characteristics are updated online, enabling linear parameter-varying (LPV)-like behavior within a single model.

3.3. Physics-constrained decoder and non-negative gains

The decoder is designed to address the engineering need for path contribution analysis by parameterizing the transfer function in a non-negative, low-rank factorized form:

$$\mathbf{H}_\theta(\omega_l) = \underbrace{\mathbf{\Phi}_p(\omega_l) \in \mathbb{R}_{\geq 0}^{N\times R}}_{\text{mic–sideshapes}} \cdot \underbrace{\operatorname{diag}\left(\mathbf{g}(\omega_l) \in \mathbb{R}_{\geq 0}^{R}\right)}_{\frac{\text{path}}{\text{modalgains}}} \cdot \underbrace{\mathbf{\Phi}_u(\omega_l)^\top \in \mathbb{R}_{\geq 0}^{R\times M}}_{\text{excitation–sideshapes}}, \tag{8}$$

where $\mathbf{\Phi}_p(\cdot)$ and $\mathbf{\Phi}_u(\cdot)$ are microphone-side and excitation-side basis matrices, respectively, and $\mathbf{g}(\omega_l)$ collects the non-negative path gains. During training, a gradient projection algorithm $\mathit{\Pi}_{\geq 0}(\cdot)$ is used to strictly limit parameter updates within the non-negative orthant:

$$\Theta^{t+1} = \mathit{\Pi}_{\geq 0}(\Theta^t - \eta \nabla_\Theta \mathcal{L}), \tag{9}$$

where the symbol Θ denotes the constrained parameters, η is the learning rate, and $\mathcal{L}$ is the loss. This constraint enforces additive, non-negative path contributions, avoiding cancellation between positive and negative weights.

3.4. Evaluation metric system: Quantifying physical fidelity

3.4.1. Statistical accuracy metrics

Time-RMSE (RMSE_t) measures fitting precision at the waveform level [Pa]:

$$\mathrm{RMSE}_t = \sqrt{\frac{1}{T}\sum_t (p(t) - \widehat{p}(t))^2}, \tag{10}$$

where $p(t)$ and $\widehat{p}(t)$ are the measured and predicted sound pressure, and T is the number of samples.

Frequency-RMSE (RMSE_f) measures fitting precision of the spectral amplitude [Pa/Hz]. This metric is particularly important for evaluating the capture of resonance peaks.

Global RMSE is the full-band spectral error computed over the evaluation band on the same frequency grid [dB(A)]:

$$\text{Global RMSE} = \sqrt{\frac{1}{N_f}\sum_{i=1}^{N_f} (L_{\mathrm{pred}}(f_i) - L_{\mathrm{meas}}(f_i))^2}, \tag{11}$$

where $L_{\mathrm{meas}}(f_i)$ and $L_{\mathrm{pred}}(f_i)$ are the measured and predicted A-weighted sound pressure levels (SPLs) at frequency f_i, and N_f is the number of points in the 20 Hz to 300 Hz band.

3.4.2. Energy consistency factor, phase-sensitive mutual information, and causality violation ratio

Energy consistency factor (ECF) measures how well the predicted and measured spectra match in energy distribution, revealing non-physical leakage or spurious amplification:

$$\mathrm{ECF} = \frac{1}{|\Omega|}\sum_{\omega_l \in \Omega} \frac{||S_{\widehat{p}}(\omega_l)| - |S_p(\omega_l)||}{|S_p(\omega_l)| + \epsilon}, \tag{12}$$

where $S_{\widehat{p}}(\omega_l)$ and $S_p(\omega_l)$ are the predicted and measured spectra over Ω, ϵ prevents division by zero, and $\|\cdot\|$ is the Euclidean norm. Lower ECF indicates better energy consistency and reflects the effectiveness of the passivity constraint, while phase-sensitive mutual information (PSMI) assesses phase fidelity via the statistical dependence between predicted and measured phase:

$$\text{PSMI} = I_{\text{circ}}(\phi_{\widehat{p}}, \phi_p), \tag{13}$$

where $\phi_{\widehat{p}}$ and ϕ_p are the predicted and measured phase spectra, and $I_{\text{circ}}(\cdot)$ is circular mutual information. A higher PSMI indicates that the model reproduces propagation delays and phase evolution more faithfully. Causality violation ratio (CVR) measures the proportion of energy in the $t < 0$ part of the impulse response function:

$$\text{CVR} = \frac{\int_{-\infty}^{0} |h(t)|^2 \mathrm{d}t}{\int_{-\infty}^{\infty} |h(t)|^2 \, \mathrm{d}t}, \tag{14}$$

where $h(t)$ is the impulse response of the excitation-to-pressure mapping. For physical systems, CVR should be strictly zero. Any non-negligible value indicates 'pre-ringing' artifacts and causality violation.

4. Simulation data and virtual experiments

4.1. Simulation corpus and virtual testbench

A physics-based virtual testbench was built to generate paired 18-channel suspension/attachment accelerations and in-cabin pressure under controlled ISO 8608 road excitations. The model includes suspension–body–cavity coupling to reproduce structural resonances and cabin boom. Manufacturing scatter and aging were emulated by perturbing key stiffness/damping parameters within realistic bounds, forming a parametric 'virtual fleet.' The resulting frequency response function (FRF) and SPL dispersion are summarized in Fig. 2.

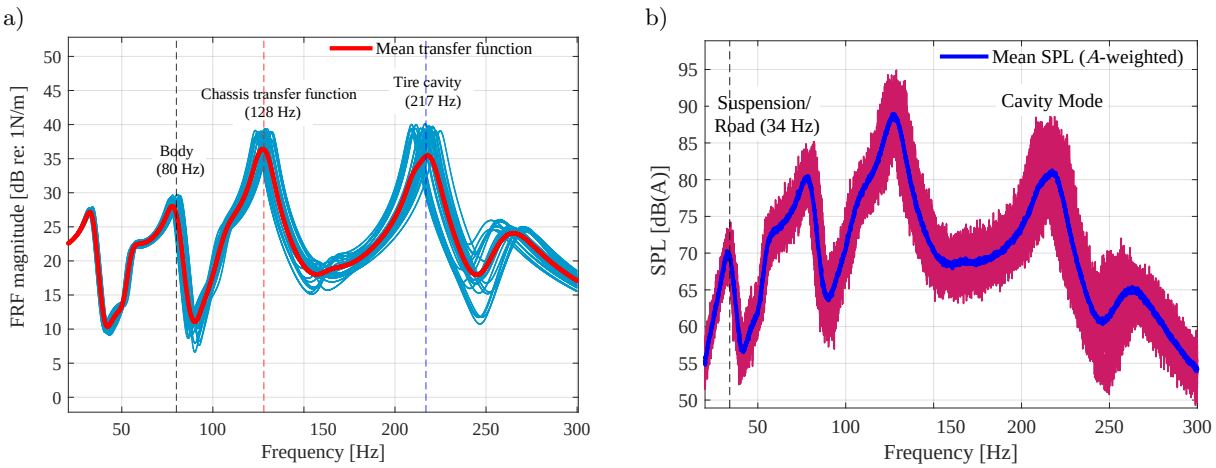

Fig. 2. Virtual fleet: parametric FRF dispersion (shaded band/area: perturbed fleet range 15 %; solid line: nominal baseline): a) parametric FRF dispersion (structure), b) in-cabin SPL dispersion (response).

Figure 3 illustrates why purely linear baselines can struggle under parameter drift and mild nonlinearities. Nonlinear coupling broadens the effective transfer behavior and can shift the modal content. By enforcing causality and passivity during learning, NN-MMFNet limits unphysical variability in the learned noise transfer function and retains the modal envelopes needed for stable boom prediction.

Evaluation protocol and baselines. Performance is reported on strictly disjoint, stratified splits (in-domain (ID), speed/road-type shifts, and a system-shift split with ±15 % plant-parameter perturbations to emulate virtual-fleet drift). Baselines are compared under matched capacity and training budgets.

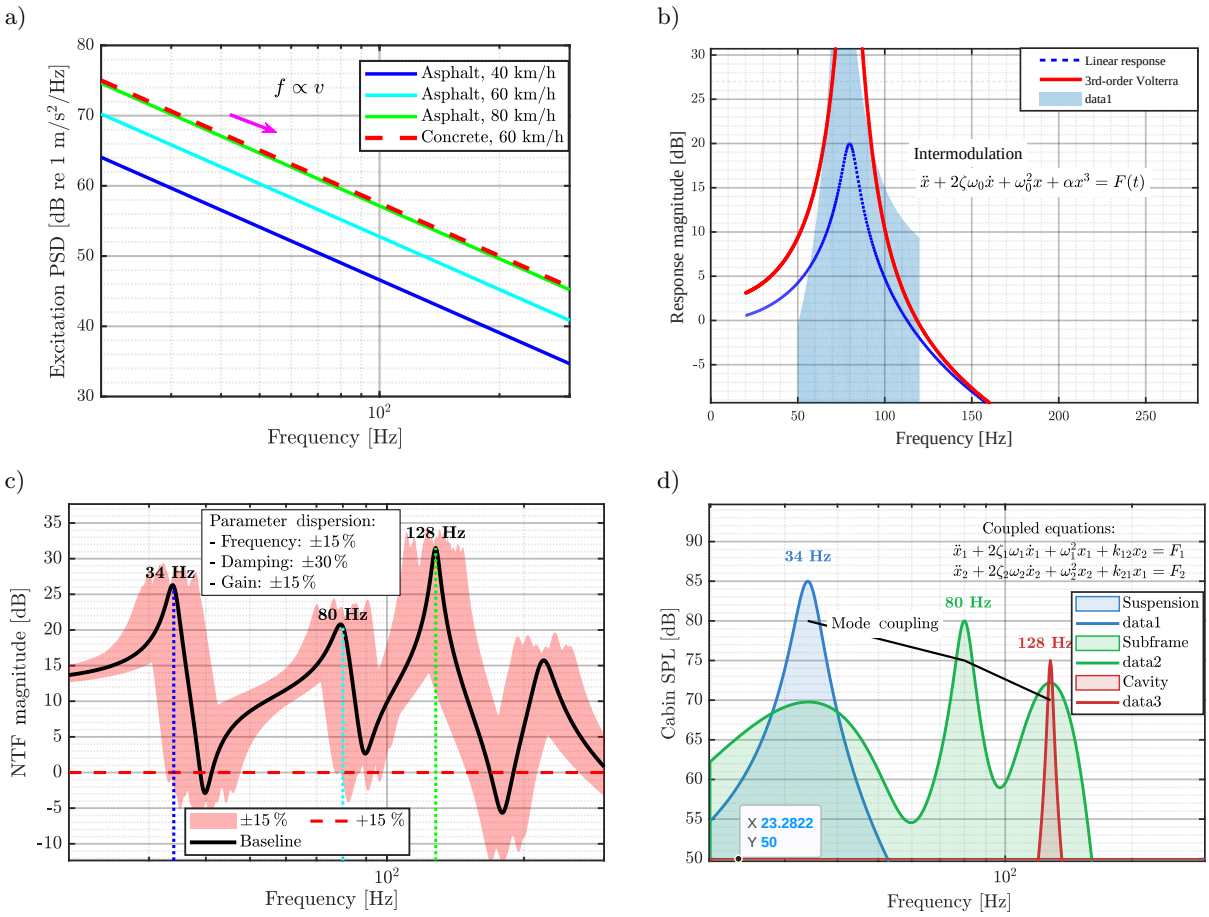

FIG. 3. Physics-constrained excitation-response analysis: a) source excitation dynamics, b) nonlinear Volterra-kernel response, c) NTF variability with physics constraints, d) key modal response envelopes.

4.2. RESULTS, ABLATIONS, AND THE SIM-TO-REAL BRIDGE

4.2.1. SIGNAL RECONSTRUCTION AND DOMAIN GENERALIZATION

NN-MMFNet provides accurate reconstruction of the coupled response in the 20 Hz to 300 Hz band. As illustrated in Fig. 4, the predicted signal follows the measured data in both time and frequency domains and resolves the main resonances at 34 Hz (suspension) and 128 Hz (cavity).

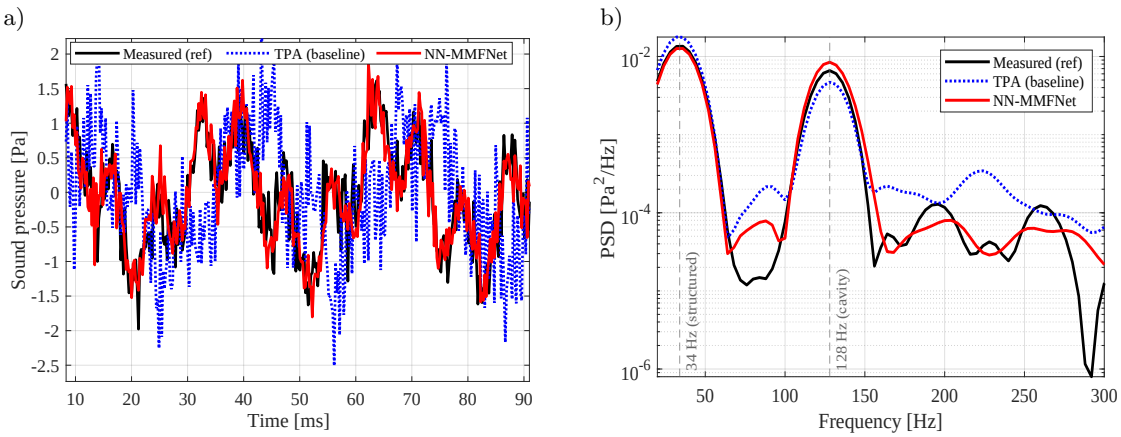

FIG. 4. Time-frequency reconstruction: a) time-domain alignment, b) spectral fidelity showing peak matching.

Unlike inverse-filtering baselines (TPA/FTM) that become ill-conditioned near resonances and amplify noise, the forward-projection decoder with a spectral cap stabilizes and effectively denoises the transfer path. The strictly causal architecture also preserves phase fidelity (phase-sensitive mutual information (PSMI) = 0.94), avoiding the non-causal phase distortions often introduced by acoustic-contribution principal component analysis (AC-PCA) or frequency-domain regularization (Cervantes-Madrid *et al.*, 2021; Shang *et al.*, 2021).

4.2.2. Statistical performance and robustness

Table 1 presents the comprehensive performance metrics. NN-MMFNet achieves the lowest RMSE and best (lowest) ECF across all splits, confirming its ability to balance numerical accuracy with physical compliance.

Table 1. Comparative performance metrics (mean and 95 % confidence intervals) across experimental splits. Lower is better ↓.

Split	Method	RMSE_t [Pa] ↓	RMSE_f [Pa/Hz] ↓	$\mathbf{\Phi}_{\mathrm{neg}}$ ↓	ECF [%] ↓
In-domain (ID)	NN-MMFNet	**29.55 [29.5, 29.6]**	**8.27**	**0.08**	**4.94**
	TPA	144.69 [144.7, 144.7]	25.78	0.26	31.15
Condition shift (CS)	NN-MMFNet	**31.47 [25.5, 37.4]**	**8.95**	**0.09**	**4.93**
	TPA	148.89 [148.7, 149.1]	27.4	0.28	29.8
System shift (SS)	NN-MMFNet	**30.85 [28.2, 33.1]**	**8.95**	**0.09**	**4.93**
	TPA	148.50 [147.1, 149.9]	27.4	0.28	29.8
Domain shift (DS)	NN-MMFNet	**22.12 [18.0, 26.9]**	**4.92**	**0.05**	**4.9**
	TPA	151.46 [150.7, 152.8]	15.7	0.16	25.72

Note: Data are derived from simulation manifests. Confidence intervals (95 % CI) are calculated via bootstrapping over segments. $\mathbf{\Phi}_{\mathrm{neq}}$ denotes the negative-contribution ratio.

The system-shift split highlights the sensitivity of conventional methods to plant drift (Fig. 5). With ±15 % parameter perturbations, TPA degrades sharply and the RMSE exceeds 148 Pa, whereas NN-MMFNet remains stable with an RMSE of 30.85 Pa, indicating an operating-adaptive mapping rather than a fixed transfer matrix. Under the unseen concrete domain, NN-MMFNet also retains low energy inconsistency with an ECF of 4.90 %, consistent with physics-informed separation of excitation $\mathbf{u}(t)$ and path dynamics $\mathbf{H}_s(\omega)$.

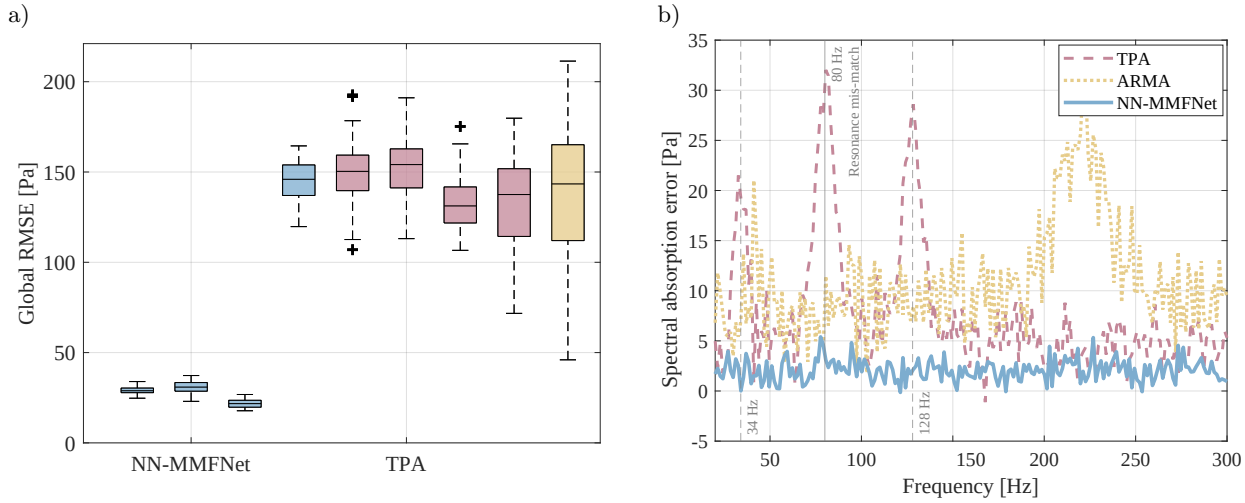

Fig. 5. Boxplots of RMSE_t distribution across in-domain, condition shift, and domain shift settings: a) statistical error distribution (global), b) frequency-resolved error profile.

4.2.3. Ablation study: Mechanism verification

Table 2 and Fig. 6 report the ablation results under the same ID split and objective. Figure 6a compares the multi-metric trade-offs, whereas Fig. 6b shows the corresponding passivity-violation rates. The full NN-MMFNet

TABLE 2. Ablation results showing the impact of removing physics constraints (ID split).

Model variant	RMSE_t [Pa] ↓	Peak-location error [Hz] ↓	PSMI ↑	ECF [%] ↓	Passivity violation [%] ↓
NN-MMFNet (full)	29.55	0	0.94	5.10	<0.01
Without spectral caps	30.12	0	0.88	**85.0**	**18.4**
Without non-negativity	28.9	0	0.42	6.12	0.02
Without normalized attention	45.3	2.5	0.81	15.12	0.05

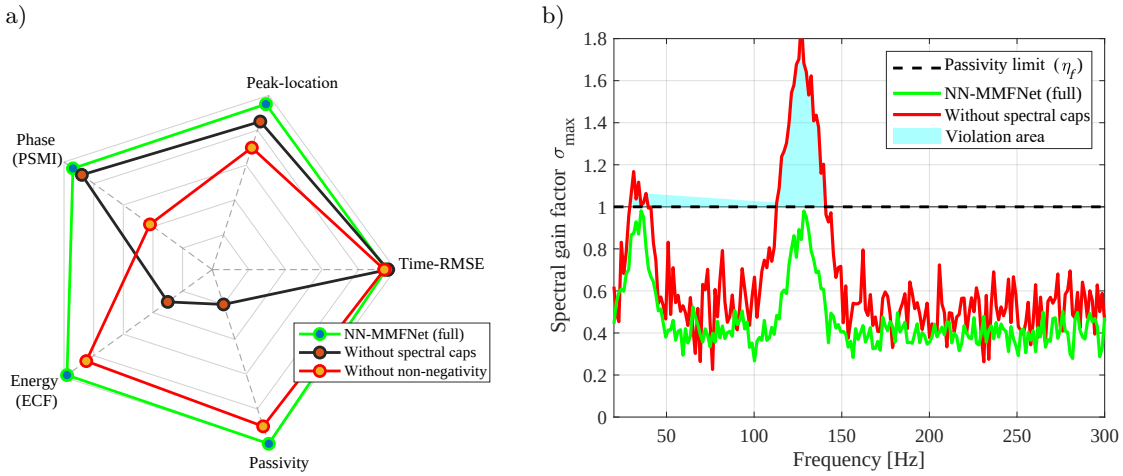

FIG. 6. Ablation study: multi-objective trade-offs and mechanism verification: a) performance trade-off (radar), b) passivity violation analysis.

achieves RMSE = 29.55 Pa, PSMI = 0.94 and ECF = 5.10 %, with passivity violations below 0.01 % and zero peak-location error. Removing the spectral cap leaves the RMSE nearly unchanged but leads to large energy inconsistency (ECF ≈ 85 %) and 18.4 % passivity violations, with non-physical amplification concentrated near 128 Hz.

Removing the non-negativity constraint slightly improves RMSE but reduces phase fidelity (PSMI = 0.42), which weakens the interpretability of the path contributions. Without normalized attention, performance degrades the most (RMSE = 45.3 Pa and peak-location error = 2.5 Hz), indicating that attention normalization is essential for robust multi-channel fusion.

5. EXPERIMENTAL VALIDATION AND ENGINEERING IMPLEMENTATION

NN-MMFNet is evaluated using a sim-to-real protocol that tests three aspects: physically constrained signal fidelity, mechanism interpretability, and robustness to parametric uncertainty.

5.1. EXPERIMENTAL SETUP AND DATA ACQUISITION PROTOCOL

Field tests were conducted on a production battery electric vehicle (SAIC-GM-Wuling F510C). Data were acquired using a 56-channel LMS SCADAS Mobile system controlled by Siemens LMS Test.Lab software. As shown in Fig. 7, the sensor layout was designed to cover the main vibro-acoustic transmission chain.

Excitation source (X): sixteen tri-axial PCB accelerometers (48 channels) were installed at key chassis hard-points. In particular, fifteen sensors monitored structure-borne inputs at suspension control arms and subframe bushings, and one sensor was mounted at the driver's seat base to capture the terminal structural transmission to the occupant.

Acoustic response (Y): four GRAS microphones were positioned at ear levels of front and rear passengers to characterize the target sound field.

Time-domain data collected at a constant speed ($v \in \{40, 60, 80\}$ km/h) were exported to MATLAB for processing and model training. To assess generalization beyond a single test vehicle, a MATLAB-based virtual fleet was generated by applying ±15 % random perturbations to the mass and stiffness matrices of a multibody

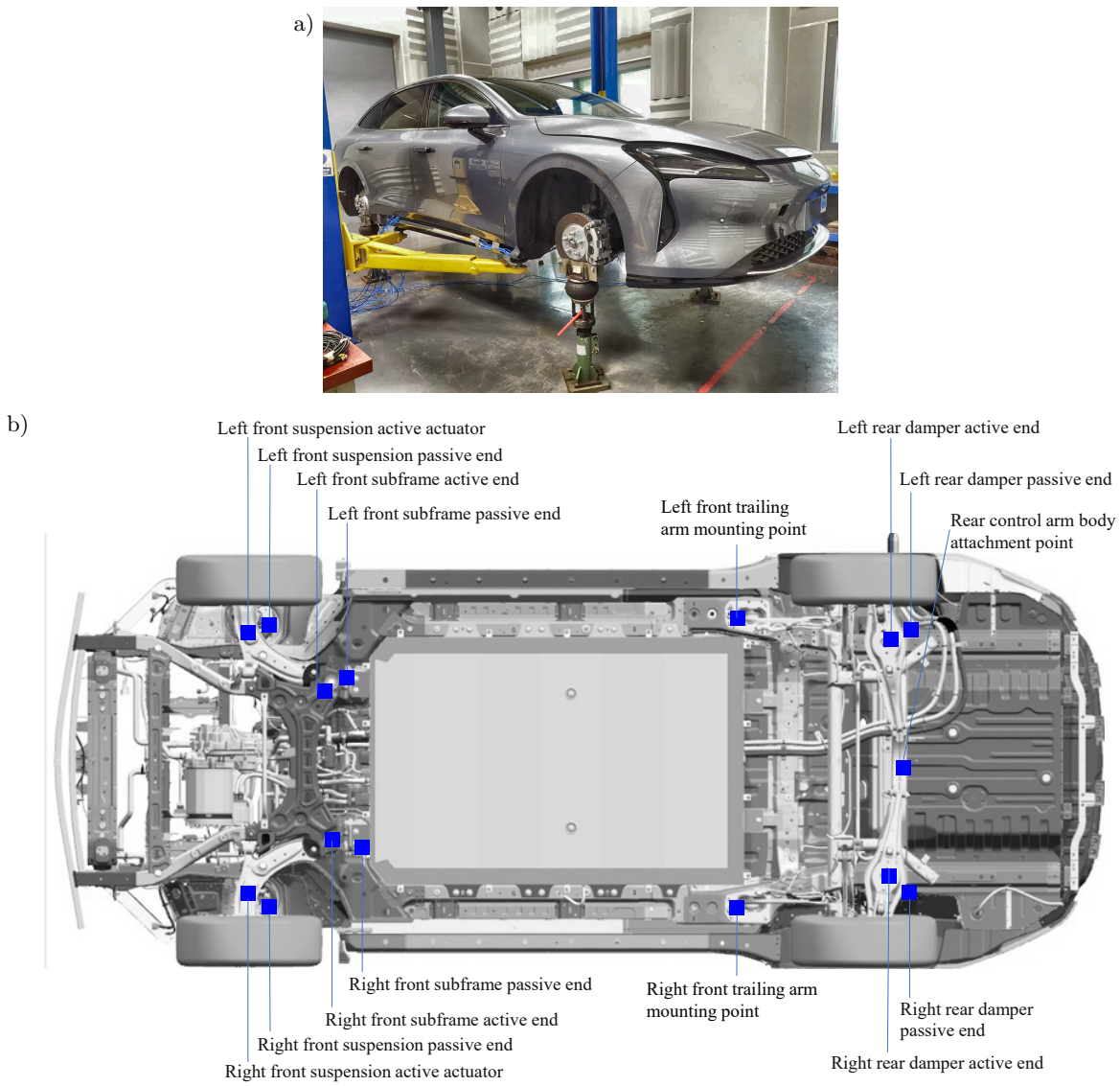

Fig. 7. Experimental test and main sensor arrangement:
a) instrumented F510C BEV platform on a four-post lift, b) topology of acceleration sensors (blue).

dynamics model, producing 100 variants. NN-MMFNet was pretrained on this virtual dataset to learn transferable coupling patterns, and then fine-tuned on the measured data.

5.2. Comparative performance analysis: Sim-to-real bridge

Pretrained on the virtual fleet, the model was fine-tuned using 15 min of real-world measurements. NN-MMFNet was benchmarked against industry-standard TPA based on matrix inversion, FTM, and ARMA baselines.

5.2.1. Frequency-domain fidelity and resonance capture

To validate the model's reconstruction capabilities under complex coupled dynamics, we analyzed the spectral fidelity and resonance-tracking accuracy within the critical 20 Hz to 300 Hz band.

Figure 8 shows full-band spectral reconstruction at 60 km/h. NN-MMFNet closely matches the measured spectrum, capturing the 128 Hz cavity boom with a 0.14 dB peak error. In contrast, TPA becomes numerically ill-conditioned above 150 Hz and exhibits ghost-energy amplification. Table 3 further confirms that NN-MMFNet reduces the global RMSE to 1.12 dB at 60 km/h, compared with 4.15 dB for TPA, demonstrating that the physics-constrained decoder suppresses phantom gain in inverse methods.

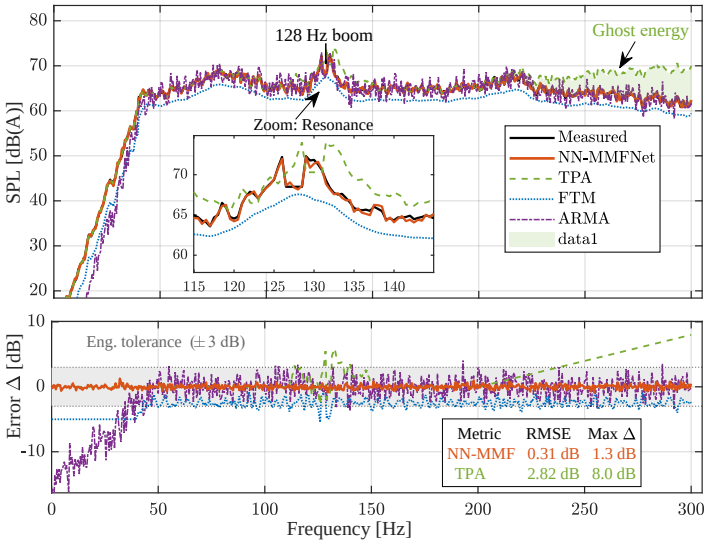

FIG. 8. Full-band spectral fidelity and error residual analysis (60 km/h).

TABLE 3. Quantitative evaluation of spectral fidelity and generalization capabilities (40 km/h to 80 km/h).

Cruising speed [km/h]	Method	Global RMSE [dB(A)] ↓	ECF [%] ↓	Suspension mode (34 Hz) [Δ dB]	Subframe mode (80 Hz) [Δ dB]	Cavity boom (128 Hz) [Δ dB]
40 (low load)	NN-MMFNet	1.10	4.65	0.08	0.12	0.15
	TPA	3.52	26.50	0.45	1.85	2.80
	FTM	3.25	24.10	1.20	2.50	1.95
	ARMA	2.80	14.50	3.50	0.90	1.20
60 (rated load)	NN-MMFNet	1.12	4.80	0.06	0.11	0.14
	TPA	4.15	31.15	0.52	2.10	3.20
	FTM	3.80	29.80	1.45	3.10	2.40
	ARMA	2.55	15.20	4.10	1.15	1.80
80 (high load)	NN-MMFNet	1.25	5.12	0.15	0.22	0.20
	TPA	6.50	42.30	0.95	3.50	**5.50**
	FTM	4.20	33.50	1.80	3.80	2.90
	ARMA	3.10	18.60	4.50	1.50	2.50

Table 3 and Fig. 9 evaluate performance from 40 km/h to 80 km/h. NN-MMFNet maintains the lowest ECF (below 5.12 %) and global RMSE (below 1.25 dB) across speeds. At 80 km/h, conventional methods degrade due to bushing nonlinearity, and TPA reaches a 5.5 dB error at 128 Hz, whereas NN-MMFNet remains stable at 0.2 dB, indicating attention-based adaptation to parametric drift and a physically consistent spectral mapping.

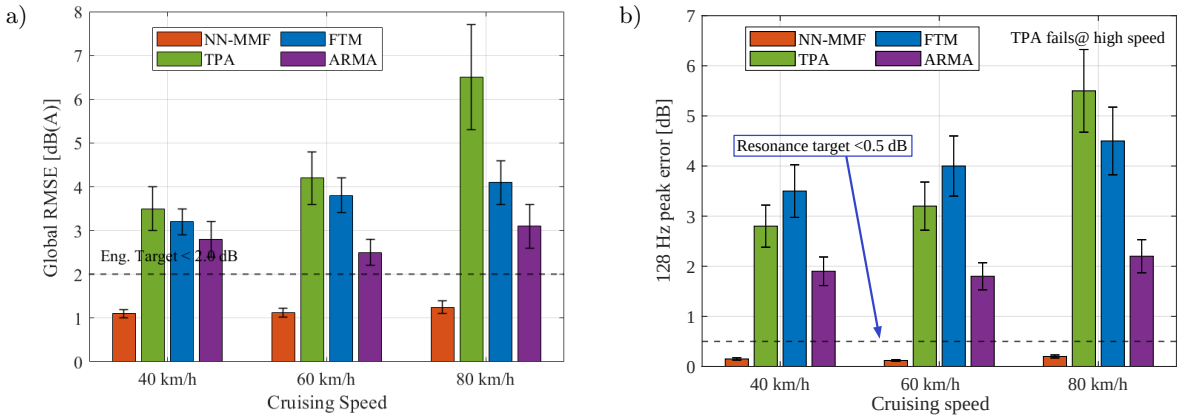

FIG. 9. Operational robustness and generalization (40 km/h to 80 km/h):
a) global accuracy across speeds, b) resonance capture fidelity (128 Hz).

5.2.2. Time-domain transient fidelity and causal verification

Physical validity under non-stationary conditions was assessed via impulse-response analysis, testing whether transient structural decay is captured while strictly preserving causality, which is often violated by block-based processing.

Figure 10 shows that TPA captures the overall decay but produces non-physical pre-ringing at negative time due to acausal processing. Table 4 quantifies this via the CVR, defined as the fraction of pre-cursor energy in the impulse response. NN-MMFNet reduces this leakage to below 0.01 %, compared with 8.40 % for TPA, confirming the effectiveness of the physics-constrained decoder.

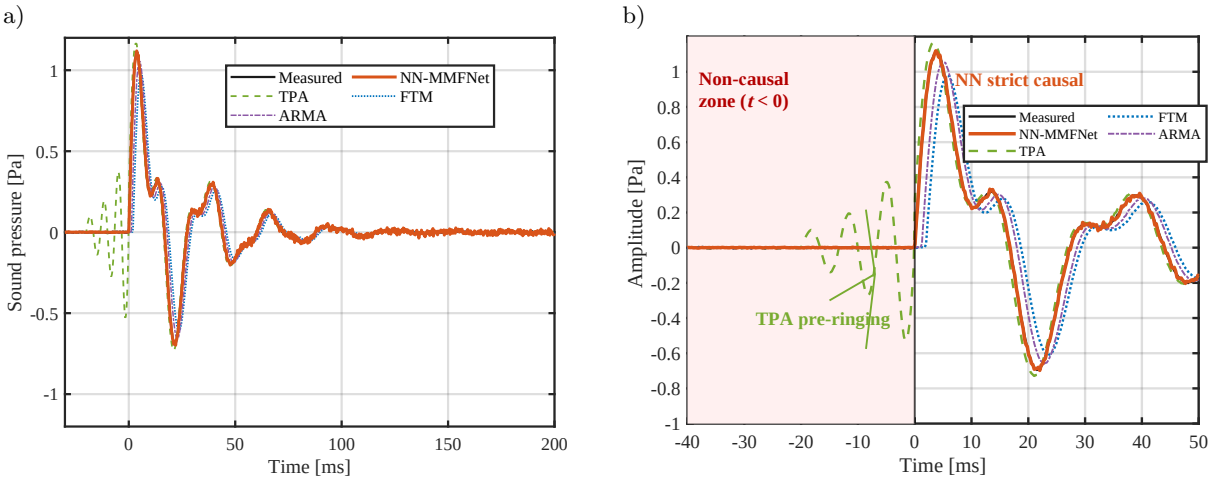

Fig. 10. Transient impulse response and causal verification:
a) global transient impulse response, b) causality check (zoom at impact).

Table 4. Quantitative evaluation of time-domain fidelity and causal consistency.

Method	RSME$_t$ [Pa] ↓	Peak phase error [°] ↓	CVR [%] ↓	Physical interpretation
NN-MMFNet	0.27	2.1	<0.01	Strictly causal
TPA (baseline)	1.23	12.5	8.40	Pre-ringing
FTM	0.61	15.0	3.20	Phase lag
ARMA	0.45	5.2	0.50	Time delay

Figure 11 shows stable error dynamics, with NN-MMFNet limiting peak phase error to 2.1° within the 5° engineering tolerance for coherent synthesis, while baselines exhibit phase drift. NN-MMFNet therefore achieves an RMSE$_t$ of 0.27 Pa.

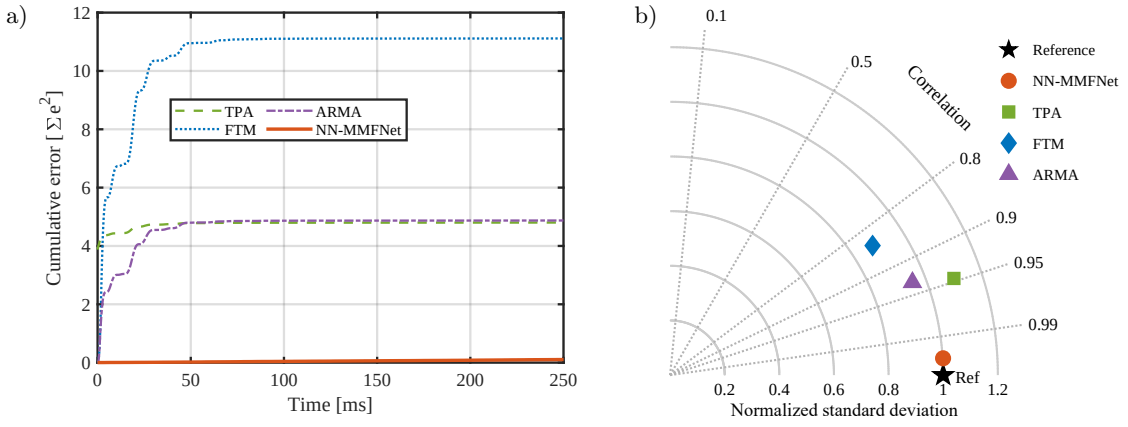

Fig. 11. Error evolution dynamics and statistical fidelity assessment:
a) error accumulation evolution, b) Taylor diagram (overall fidelity).

For the peak transient response, NN-MMFNet reduces the waveform error to 0.287 Pa, compared with 1.23 Pa for the TPA baseline. The Taylor diagram in Fig. 11b provides a consistent summary, placing NN-MMFNet closest to the reference.

5.3. Engineering application: Closed-loop diagnosis and targeted optimization

To assess practical usefulness beyond offline prediction, the non-negative attention map was used to localize the dominant structure-borne transmission path associated with the 128 Hz boom on the F510C platform. Across operating points, the attention consistently concentrated on the rear subframe-to-body mounting channel, indicating a stiff coupling in that load path.

Guided by this diagnosis, a targeted stiffness adjustment was implemented at the identified mount. Vehicle tests confirmed a 4.2 dB(A) reduction of the 128 Hz boom peak without introducing non-causal artifacts or violating passivity in the reconstructed transfer behavior. Additional multi-metric deployment analysis and secondary cases are provided in Appendix. A zero-shot cross-platform check on the F710C platform showed that the physics-constrained representation remains competitive relative to classical TPA baselines under domain shift (details in Appendix).

6. Conclusion

This paper proposed NN-MMFNet for operating-dependent structure-borne road-noise prediction in BEVs. The model integrates a dual-stream time–frequency encoder with strictly causal fusion/decoding, and passivity-enforced spectral gain control, while non-negative attention provides interpretable path contributions. With virtual-fleet pretraining followed by sim-to-real fine-tuning, the method improves robustness to speed/load variations and plant-parameter drift. Experiments on a production vehicle show a 1.12 dB(A) full-band spectral RMSE at 60 km/h, and the 128 Hz boom component is predicted within 0.14 dB(A), with a passivity-violation rate below 0.01 %. The inferred contribution map suggests the rear subframe mount as a dominant path, and a targeted stiffness update achieves a measured 4.2 dB(A) cabin-noise reduction. Future work will extend validation to broader road classes and multi-source coupling.

Appendix. Extended algorithmic benchmarking and cross-platform performance benchmarking

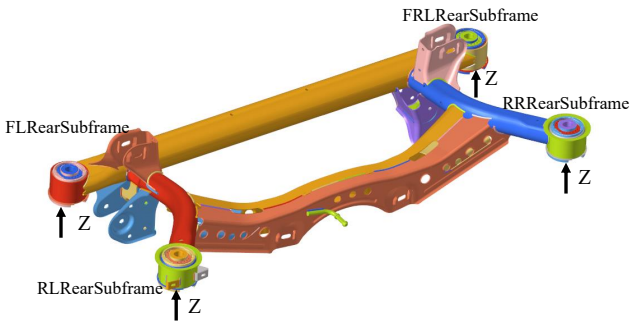

Fig. 12. Rear subframe bushing connection points.

Table 5. Quantitative performance comparison on the unseen F710C platform.

Method	Global RMSE [dB(A)] ↓	ECF [%] ↓	31 Hz error [dB] ↓	141 Hz error [dB] ↓	PSMI (phase fidelity) ↑
NN-MMFNet	0.47	6.37	0.1	0.08	0.94
TPA (matrix inv.)	3.56	107.89	5.41	6.43	0.35
FTM (regularized)	2.87	39.26	2.76	5.79	0.65
ARMA (parametric)	3.84	37.07	2.43	9.52	0.55

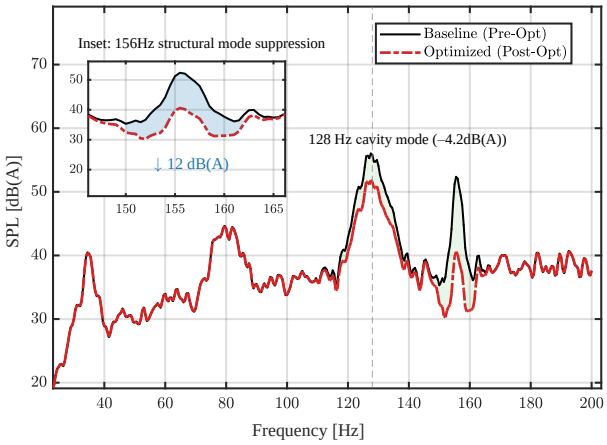

Fig. 13. F510C spectral optimization and gradient stiffness verification.

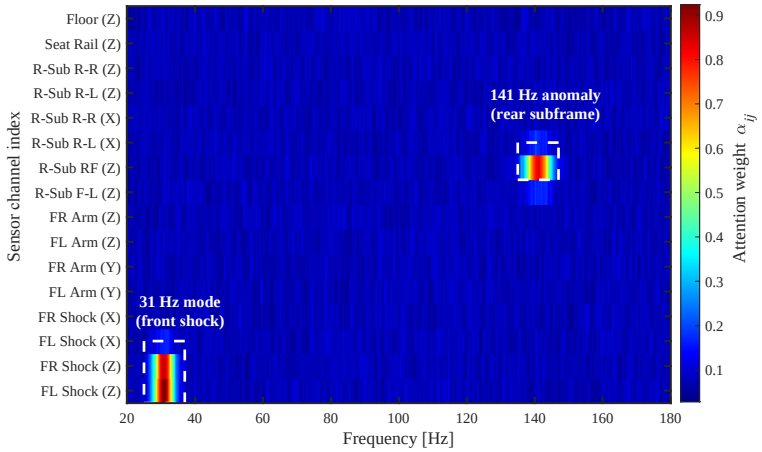

Fig. 14. Zero-shot blind diagnosis on the F710C platform: Distinct pathology identification.

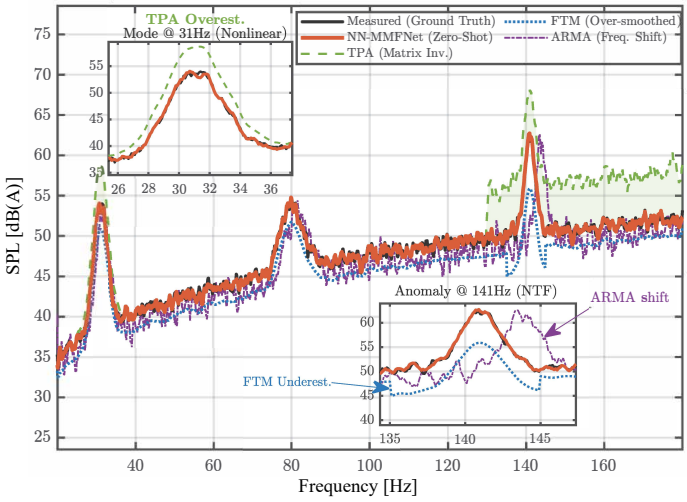

Fig. 15. Cross-platform spectral generalization performance.

Fundings

This work was supported by the Guangxi Science and Technology Major Project (grant no. AA24206071) and the Guangxi Basic Ability Enhancement Program for Young and Middle-aged University Teachers (grant no. 2024KY1454).

Authors' contributions

Haijun Wang contributed to conceptualization, methodology, software development, and writing. Zengjun Lu and Xianghua He contributed to data curation, validation, review, and writing. Zhijie Huang and Tie Xu contributed to supervision and funding acquisition. All authors read and approved the final manuscript.

Conflict of interests

The authors declare that there are no known competing financial interests or personal relationships that could have influenced the work described in this paper.

Acknowledgments

We sincerely appreciate all the engineers and project participants for their invaluable contributions and support.

References

1. Cervantes-Madrid G., Peral-Orts R., Campillo-Davo N., Campello-Vincente H. (2021), Inverse transfer path analysis, a different approach to shorten time in NVH assessments, *Applied Acoustics*, **181**: 108178, https://doi.org/10.1016/j.apacoust.2021.108178.
2. Cheng W. *et al.* (2022), AR model-based crosstalk cancellation method for operational transfer path analysis, *Journal of Mechanical Science and Technology*, **36**: 1131–1144, https://doi.org/10.1007/s12206-022-0206-7.
3. Cheng W., Chu Y., Chen X., Zhou G., Blamaud D., Lu J. (2020), Operational transfer path analysis with crosstalk cancellation using independent component analysis, *Journal of Sound and Vibration*, **473**: 115224, https://doi.org/10.1016/j.jsv.2020.115224.
4. Cheng W., Lu Y., Zhang Z. (2016), Tikhonov regularization-based operational transfer path analysis, *Mechanical Systems and Signal Processing*, **75**: 494–514, https://doi.org/10.1016/j.ymssp.2015.12.025.
5. de Klerk D., Ossipov A. (2010), Operational transfer path analysis: Theory, guidelines and tire road noise application, *Mechanical Systems and Signal Processing*, **24**(7): 1950–1962, https://doi.org/10.1016/j.ymssp.2010.05.009.
6. de Klerk D., Rixen D.J. (2010), Component transfer path analysis method with compensation for test bench dynamics, *Mechanical Systems and Signal Processing*, **24**(6): 1693–1710, https://doi.org/10.1016/j.ymssp.2010.01.006.
7. Gao L. *et al.* (2024), Operational transfer path analysis with crosstalk cancellation based on least variance spectrum estimation, *Journal of Mechanical Science and Technology*, **38**: 5311–5322, https://doi.org/10.1007/s12206-024-0907-1.
8. Gustavsen B., Semlyen A. (1999), Rational approximation of frequency domain responses by vector fitting, *IEEE Transactions on Power Delivery*, **14**(3): 1052–1061, https://doi.org/10.1109/61.772353.
9. Gustavsen B., Semlyen A. (2001), Enforcing passivity for admittance matrices approximated by rational functions, *IEEE Transactions on Power Systems*, **16**(1): 97–104, https://doi.org/10.1109/59.910786.
10. Huang H., Lim T.C., Wu J., Ding W., Pang J. (2023), Multitarget prediction and optimization of pure electric vehicle tire/road airborne noise sound quality based on a knowledge- and data-driven method, *Mechanical Systems and Signal Processing*, **197**: 110361, https://doi.org/10.1016/j.ymssp.2023.110361.
11. Jia X., Zhou L., Huang H., Pang J., Yang L. (2024), Improving electric vehicle structural-borne noise based on convolutional neural network-support vector regression, *Electronics*, **13**(1): 113, https://doi.org/10.3390/electronics13010113.
12. Khan D., Burdzik R. (2023), Measurement and analysis of transport noise and vibration: A review of techniques, case studies, and future directions, *Measurement*, **220**: 113354, https://doi.org/10.1016/j.measurement.2023.113354.
13. Kiranyaz S., Avci O., Abdeljaber O., Ince T., Gabbouj M., Inman D.J. (2021), 1D convolutional neural networks and applications: A survey, *Mechanical Systems and Signal Processing*, **151**: 107398, https://doi.org/10.1016/j.ymssp.2020.107398.

14. Kong L. *et al.* (2025), A novel operational transfer path analysis based on the complex-valued crosstalk elimination method, *Measurement*, **253**(Part D): 117813, https://doi.org/10.1016/j.measurement.2025.117813.
15. Ma Y., Dai R., Liu T., Liu J., Yang S., Wang J. (2025), Research on vehicle road noise prediction based on AFW-LSTM, *Machines*, **13**(5): 425, https://doi.org/10.3390/machines13050425.
16. Masri J., Amer M., Salman S., Ismail M., Elsisi M. (2024), A survey of modern vehicle noise, vibration, and harshness: A state-of-the-art, *Ain Shams Engineering Journal*, **15**: 102957, https://doi.org/10.1016/j.asej.2024.102957.
17. Mohammadi N. (2023), Airborne and structure-borne noise control in the MB truck cabin interior by the noise reduction in the transmission path, *Archives of Acoustics*, **48**(1): 93–101, https://doi.org/10.24425/aoa.2022.142908.
18. Moorhouse A.T., Elliott A.S., Evans T.A. (2009), In situ measurement of the blocked force of structure-borne sound sources, *Journal of Sound and Vibration*, **325**(4–5): 679–685, https://doi.org/10.1016/j.jsv.2009.04.035.
19. Münder M., Carbon C.C. (2022), A literature review [2000–2022] on vehicle acoustics: Investigations on perceptual parameters of interior soundscapes in electrified vehicles, *Frontiers in Mechanical Engineering*, **8**: 974464, https://doi.org/10.3389/fmech.2022.974464.
20. Ortega Almirón J., Bianciardi F., Corbeels P., Pieroni N., Kindt P., Desmet W. (2022), Vehicle road noise prediction using component-based transfer path analysis from tire test-rig measurements on a rolling tire, *Journal of Sound and Vibration*, **523**: 116694, https://doi.org/10.1016/j.jsv.2021.116694.
21. Park U., Kang Y.J. (2024), Operational transfer path analysis based on neural network, *Journal of Sound and Vibration*, **579**: 118364, https://doi.org/10.1016/j.jsv.2024.118364.
22. Raissi M., Perdikaris P., Karniadakis G.E. (2019), 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, https://doi.org/10.1016/j.jcp.2018.10.045.
23. Shang Z., Hu F., Zeng F., Wei L., Xu Q., Wang J. (2021), Research of transfer path analysis based on contribution factor of sound quality, *Applied Acoustics*, **173**: 107693, https://doi.org/10.1016/j.apacoust.2020.107693.
24. van der Seijs M.V., de Klerk D., Rixen D.J. (2016), General framework for transfer path analysis: History, theory and classification of techniques, *Mechanical Systems and Signal Processing*, **68–69**: 217–244, https://doi.org/10.1016/j.ymssp.2015.08.004.
25. Yang M., Dai P., Yin Y., Wang D., Wang Y., Huang H. (2025), Predicting and optimizing pure electric vehicle road noise via a locality-sensitive hashing transformer and interval analysis, *ISA Transactions*, **157**: 556–572, https://doi.org/10.1016/j.isatra.2024.11.059.
26. Zhang E., Chen Y., Su L., Zhonglian R., Chen X., Jiang S. (2024), Evaluation modeling of electric bus interior sound quality based on two improved XGBoost algorithms using GS and PSO, *Archives of Acoustics*, **49**(3): 307–317, https://doi.org/10.24425/aoa.2024.148794.
27. Zhu H., Zhao J., Wang Y., Ding W., Pang J., Huang H. (2024), Improving of pure electric vehicle sound and vibration comfort using a multi-task learning with task-dependent weighting method, *Measurement*, **233**: 114752, https://doi.org/10.1016/j.measurement.2024.114752.