Beyond Power Spectra: Cross-Frequency Interactions in Generative Dynamics

Amir Mehrpanah [1] **Mohammed Al-Jaff** [1] **Matteo Gamba** [2] **Hossein Azizpour** [1 3]

Abstract

Spectral analysis of deep generative models typically relies on marginal statistics, such as the power spectral density, implicitly assuming that per-frequency behavior suffices to characterize generative dynamics. We revisit this assumption and ask whether marginal spectral quantities are sufficient to describe the transformations learned by modern generative models. We introduce an interventional measure of spectral sensitivity that probes directional influence between frequencies and decomposes model behavior into diagonal (marginal) and off-diagonal (interaction) components. In a controlled Gaussian setting with known spectral structure, we verify that the proposed measure recovers the expected decoupling of frequencies. Applying this framework to flow matching models trained on frequency-masked CIFAR-10 and to denoising diffusion models (DDPM) trained on CelebA, we find that diagonal responses vary little across models despite substantial differences in the training data. In both cases, off-diagonal sensitivities clearly reflect the underlying perturbations, suggesting that marginal spectral statistics may be insufficient to fully characterize learned generative dynamics, and that cross-frequency interactions play a key role in their description.

1. Introduction

Generative modeling has witnessed remarkable progress in recent years, particularly through diffusion and flow-based methods, which achieve high-quality synthesis across a wide range of domains (Ho et al., 2020; Song et al., 2021). Despite this empirical success, the internal dynamics by which these models transform data distributions over time remain only partially understood (Falck et al., 2025).

A natural perspective for studying these dynamics is spectral analysis. In image settings, the Fourier basis provides a canonical representation in which data structure can often be organized by frequency (Freirich et al., 2021). This viewpoint has led to a growing literature on diffusion and flow models, including frequency-dependent noise schedules, spectral bias in learning dynamics, and frequency-aware guidance strategies (Benita et al., 2025; Wang & Pehlevan, 2026; Benita et al., 2026). These works suggest that generative processes may be understood through the evolution of spectral components over time.

Most existing spectral analyses focus on marginal quantities, such as the power spectral density (PSD) or per-frequency signal-to-noise ratios. These statistics provide a useful description of how energy is distributed across frequencies at a given time (Tivnan et al., 2023). However, they do not directly capture how frequencies interact under the learned transformation. As a result, they may offer only a partial characterization of model dynamics.

This raises a central question: *are marginal spectral statistics sufficient to characterize the dynamics of trained generative models?*

In this work, we argue that this is not generally the case. Rather than restricting attention to marginal summaries, we adopt an *interventional* view in the Fourier domain: we perturb one frequency at a time and measure how the output spectrum changes. This yields a dataset-level frequency-to-frequency sensitivity matrix J_t, which decomposes naturally into a diagonal component, corresponding to marginal response, and an off-diagonal component, corresponding to cross-frequency interactions.

Our main empirical finding is that off-diagonal interactions matter. Across both real-world image data and controlled synthetic settings, models can exhibit very similar diagonal spectral responses while differing substantially in their off-diagonal structure. In other words, marginal spectral statistics may fail to distinguish learned dynamics that are separated by their frequency interactions (see Figure 1).

To study this phenomenon, we consider two complementary regimes: 1) In real-image experiments, we examine models trained on datasets with structured spectral perturbations,

[1]Department of Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden [2]Department of Computer Science, Brown University, USA [3]Science for Life Laboratory, Stockholm, Sweden. Correspondence to: Amir Mehrpanah <amirme@kth.se>.

Published as a paper at the 1st FoGen workshop, ICML 2026, Seoul, South Korea, 2026.

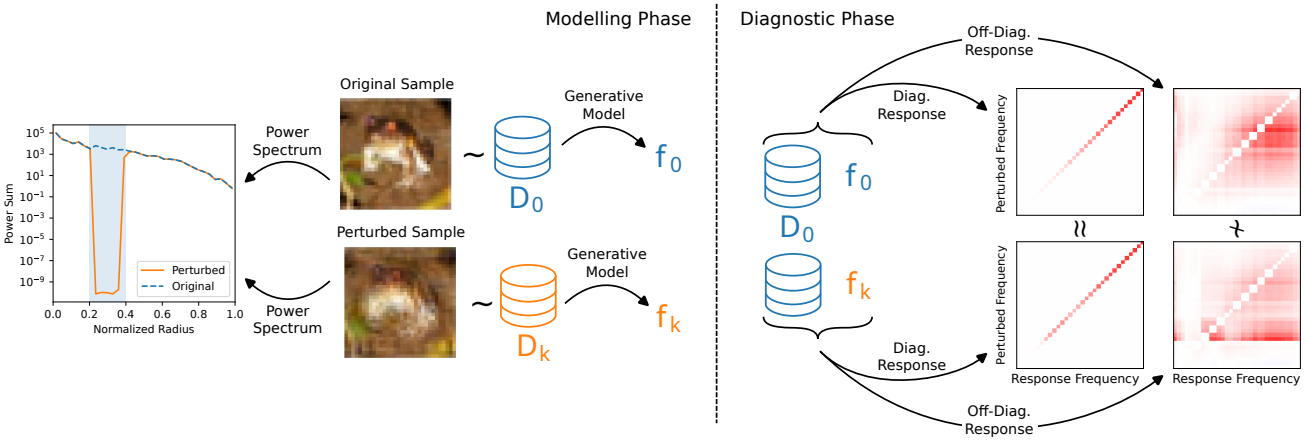

Figure 1. **Overview of our approach and main finding.** For a reference dataset D_0, we produce a family of spectrally perturbed variants D_k, via frequency masking, defined in Equation (9) (see Figures 7 and 8 for representative samples). We train generative models f_0, f_k trained on each dataset described in Section 5, and use a spectral sensitivity matrix J_t, defined in Equation (11), measured over $\mathcal{D}_0$, to decompose networks transformation in Fourier domain at time t into diagonal and off-diagonal components. We observe that while diagonal responses are nearly identical across models, off-diagonal responses differ markedly, indicating that marginal spectral statistics may be insufficient to characterize generative dynamics (see Figure 3).

and 2) In a synthetic Gaussian setting, we use a reference regime from prior work in which the covariance is diagonal in the Fourier basis and the optimal dynamics decouple across frequencies (Benita et al., 2025). This setting provides a baseline in which marginal spectral statistics are theoretically sufficient and off-diagonal sensitivity should vanish. We use it as a conceptual contrast to the more general settings of interest.

This perspective clarifies both the scope and the limitation of spectral analyses based on marginal quantities. When frequency-wise independence holds, diagonal statistics can accurately summarize the dynamics. However, outside that regime, distinct transformations may induce very similar marginal spectra while differing in how they couple frequencies. The resulting cross-frequency structure is therefore invisible to standard PSD-based analyses but is captured by our interventional measure.

The rest of the paper is organized as follows. Section 2 reviews related spectral analyses of diffusion and flow-based models, and Section 3 introduces notation. In Section 4, we define the spectral sensitivity measure and relate it to locally linearized dynamics, and briefly recall the Gaussian diagonal-covariance regime where prior work shows that frequency interactions decouple. We describe the experimental setups in Section 5, and present the resulting empirical analysis Section 6.

Contributions. Our work can be summarized as follows:

1. We empirically show that marginal spectral statistics may be insufficient to characterize the dynamics of trained generative models.
2. We introduce a dataset-level interventional spectral sensitivity matrix J_t that separates diagonal responses from cross-frequency interactions.
3. We demonstrate empirically, in both real-world and controlled settings, that off-diagonal structure exhibits differences invisible to marginal spectral analysis.

2. Related Work

Spectral Analysis of Diffusion Models. Recent work studies diffusion models in the Fourier domain through marginal spectral quantities, such as the power spectral density (PSD) or Signal-to-Noise Ratio (SNR) across frequencies (Falck et al., 2025; Benita et al., 2025). These analyses reveal characteristic frequency hierarchies, where low-frequency components are recovered earlier than high-frequency ones. More broadly, spectral formulations of diffusion have been explored in both Euclidean and non-Euclidean domains (Phillips et al., 2022; Brutti et al., 2026). These approaches analyze dynamics at the level of individual frequency components. In contrast, we investigate whether such marginal (diagonal) spectral statistics are sufficient to characterize the underlying generative dynamics.

Spectral Bias and Learning Dynamics. Several works analyze the learning dynamics of diffusion models through the lens of spectral bias, characterizing how different frequency components of the score function are learned at different rates (Wang & Pehlevan, 2026). Related analyses also study how posterior sampling and guidance mechanisms in-

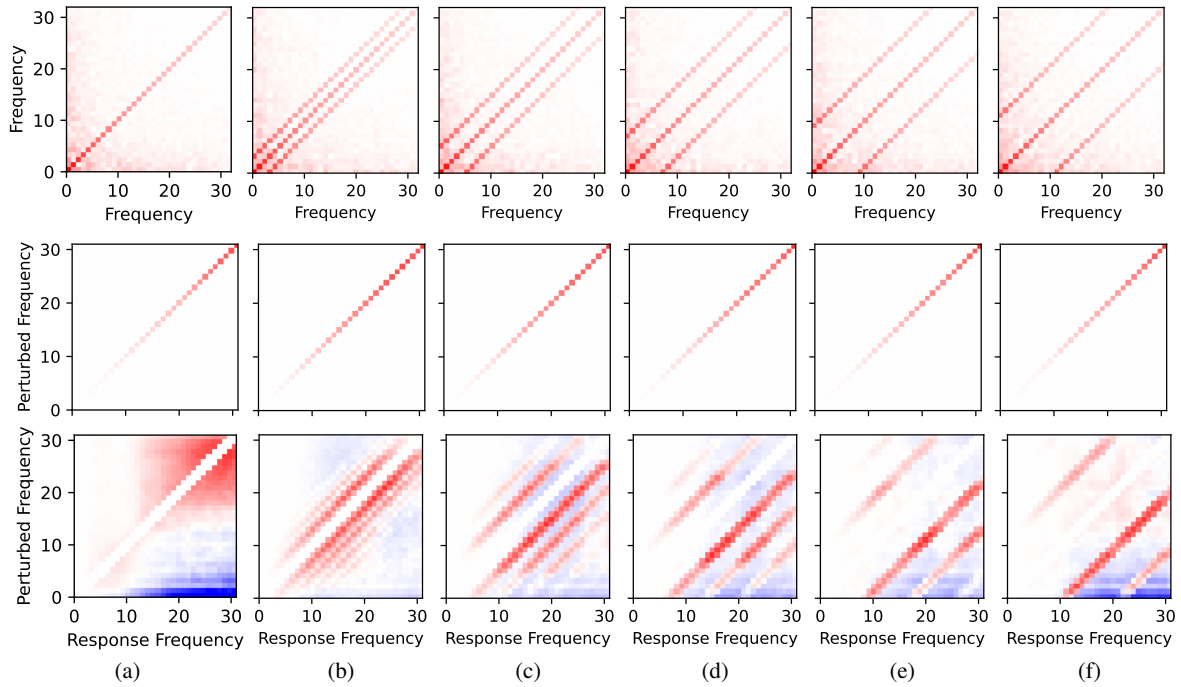

Figure 2. **Spectral Sensitivity Recovers Ground-Truth Cross-Frequency Structure in Gaussian Data.** Top row: prescribed Gaussian covariance in the Fourier domain (see Figure 9 for the corresponding covariance in the spatial domain). (**a**) shows the reference case $k = 0$, yielding a purely diagonal covariance. (**b**–**f**) introduce increasing spectral shifts $k \in \{3, 5, 7, 9, 11\}$, producing structured off-diagonal covariance with banded patterns at fixed frequency offsets. Middle and bottom rows: diagonal and off-diagonal components of the spectral sensitivity J_t (defined in Equation (11)) at time $t = 0.9$, evaluated on the reference diagonal case $\mathcal{D}_0$, where $k = 0$. Each column corresponds to a conditional flow matching model f_k trained on Gaussian data with the covariance structure shown in the top row. All quantities are averaged over 1024 samples. Despite substantial differences in the underlying covariance (top row), the diagonal responses (middle row) remain nearly invariant across models, indicating that marginal (per-frequency) sensitivity does not reflect the introduced cross-frequency dependencies. In contrast, the off-diagonal responses (bottom row) vary significantly across columns and closely reflect the off-diagonal band structure of the ground-truth covariance. This demonstrates that J_t^{off} recovers cross-frequency coupling induced by the data, highlighting its role as a diagnostic of interaction structure beyond marginal spectral statistics.

teract with frequency structure (Benita et al., 2026). While this line of work focuses on per-frequency behavior during optimization, our analysis is post-hoc and instead examines cross-frequency interactions in the learned dynamics.

Frequency-Aware Design and Modeling Assumptions. A complementary line of work incorporates spectral structure directly into model design, for example by defining diffusion processes in the Fourier domain to control image statistics (Tivnan et al., 2023) or by developing frequency-aware flow matching approaches for structured domains (Wang et al., 2025; Moghadas et al., 2025). While these methods leverage diagonalization in the Fourier basis to enable tractable modeling or control, we analyze standard time-domain models and study the extent to which their behavior can be captured by frequency-wise (diagonal) statistics, highlighting the role of cross-frequency interactions.

3. Preliminaries

Flow matching. We consider flow matching models (Tong et al., 2024) that learn a time-dependent velocity field $f_t(x)$ transporting a base distribution p_0 to a data distribution p_1. Given a linear interpolation process

$$x_t = tx_1 + (1-t)\epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \tag{1}$$

the optimal velocity field satisfies

$$f_t^{\text{opt}}(x) = \mathbb{E}[x_1 - \epsilon \mid x_t = x]. \tag{2}$$

In practice, f_t is parameterized by a neural network. This formulation admits the representation of noise at time t as

$$\epsilon_{\text{pred}} = x_t - tf_t(x_t). \tag{3}$$

Fourier representation and spectral power. Let $\mathcal{F}(\cdot)$ denote the discrete Fourier transform in spatial domain and Ω the set of discrete Fourier modes. For a signal z, we write its Fourier coefficient at frequency ω as $\widehat{z}(\omega)$. We define the

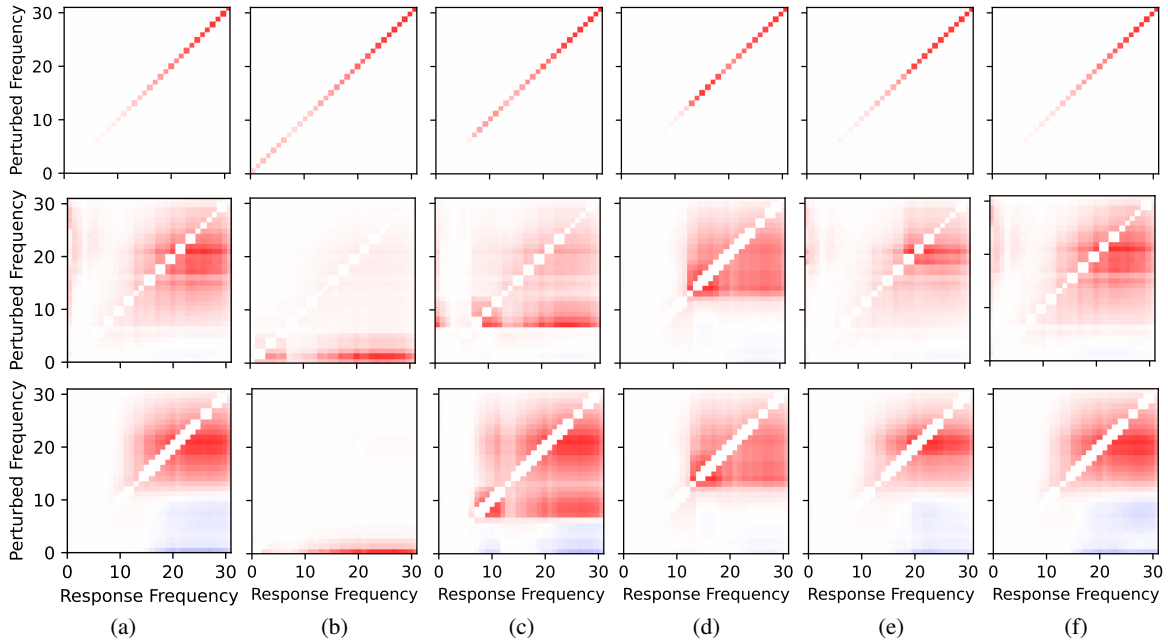

Figure 3. **Diagonal vs. Off-Diagonal Spectral Sensitivity at Late Time Reveals Interaction Structure.** Diagonal (top row) and off-diagonal (middle and bottom rows) components of the spectral sensitivity response (defined in Equation (11)) for generative models at time $t = 0.9$, evaluated on the reference dataset $\mathcal{D}_0$. Each visualization shows the average response over 1024 random samples. Each column corresponds to a model trained on a different frequency-perturbed version of CIFAR-10 (middle row) and CelebA (bottom row): **(a)** unperturbed reference; **(b)–(f)** correspond to band-stop masking over normalized radial frequency intervals $[0, 0.2]$, $[0.2, 0.4]$, …, $[0.8, 1.0]$ (see Figures 7 and 8 for representative samples). The diagonal responses exhibit nearly indistinguishable trends across columns, due to the dominance of the natural spectral decay of image PSD over the imposed masking. In contrast, the off-diagonal responses vary significantly across columns and better reflect the structure of the frequency-domain perturbations during training. A common pattern in both components is a high-frequency bias; diagonal sensitivity is concentrated at higher frequencies, while off-diagonal responses indicate that perturbations at frequency ω often influence frequencies $\omega' > \omega$. See Figures 12 to 15 for the evolution of diagonal and off-diagonal responses over time for CIFAR-10 and CelebA.

spectral power at frequency ω as

$$S(z, \omega) = |\widehat{z}(\omega)|^2. \tag{4}$$

Marginal spectral statistics. A standard summary of spectral structure is given by the power spectral density (PSD), which corresponds to the marginal (diagonal) component of the spectral covariance $S(x_t, \omega)$.

Prior work has shown that, under approximate stationarity, image data exhibit an (approximately) diagonal covariance structure in the Fourier basis, with the PSD capturing its marginal (diagonal) components (Freirich et al., 2021; Tivnan et al., 2023). This frequency-wise perspective is commonly used in analyses of diffusion dynamics (Benita et al., 2025). In particular, PSD describes how energy is distributed across frequencies at each time step.

In this work, we adopt the common perspective that PSD is a suitable descriptor of the *data distribution* at each time step. Our goal, however, is to understand the variety in *model transformations* while data assumed fixed.

Spectral covariance. To capture dependencies between frequencies, one may consider the spectral covariance

$$C_t(\omega, \omega') = \mathbb{E}\big[\widehat{x_t}(\omega)\,\overline{\widehat{x_t}(\omega')}\big]. \tag{5}$$

While $C_t(\omega, \omega')$ captures dependencies in the data distribution, it does not isolate the effect of the model transformation. A model-aware quantity is the input–output covariance

$$C_t^{\text{cross}}(\omega_{\text{out}}, \omega_{\text{in}}) = \mathbb{E}\big[\widehat{\epsilon_{\text{pred}}}(\omega_{\text{out}})\,\overline{\widehat{\epsilon}(\omega_{\text{in}})}\big], \tag{6}$$

which reflects how input frequencies relate to output frequencies. However, such correlational measure still conflate data and model effects. To isolate directional interactions induced by the model, we instead adopt the interventional sensitivity defined in the next section.

4. Spectral Structure of Generative Dynamics

To analyze how trained models transform spectral structure, we adopt a Fourier-domain representation of the generative dynamics and study how input frequencies are mapped to output frequencies.

Algorithm 1 Local Spectral Sensitivity Estimation

Input: sample $x \sim \mathcal{D}_0$, model f_k, time t, frequency set Ω
$x_t \leftarrow \textsc{Diffuse}(x, t)$ (Equation (1))
$S_x \leftarrow \mathrm{PSD}(x_t)$ (Equation (4))
$\epsilon_{\text{pred}} \leftarrow \textsc{Predict}(f_k, x_t)$ (Equation (3))
$S_\epsilon \leftarrow \mathrm{PSD}(\epsilon_{\text{pred}})$ (Equation (4))
Initialize $J_t^{\text{diag}} \leftarrow [\,]$, $J_t^{\text{off}} \leftarrow [\,]$
for each $\omega_{\text{in}} \in \Omega$ **do**
  $x_t^p \leftarrow \textsc{SpectralPerturb}(x_t, \omega_{\text{in}})$ (Equation (9))
  $S_x^p \leftarrow \mathrm{PSD}(x_t^p)$ (Equation (4))
  $\epsilon_{\text{pred}}^p \leftarrow \textsc{Predict}(f_k, x_t^p)$ (Equation (3))
  $S_\epsilon^p \leftarrow \mathrm{PSD}(\epsilon_{\text{pred}}^p)$ (Equation (4))
  $\text{num} \leftarrow S_\epsilon^p - S_\epsilon$
  $\text{den} \leftarrow S_x^p - S_x$
  $r \leftarrow \text{num}/\text{den}$ (Equation (11))
  Append $r(\omega_{\text{in}})$ to J_t^{diag}
  Append r to J_t^{off}
end for
$J_t^{\text{off}} \leftarrow J_t^{\text{off}} - J_t^{\text{diag}}$
Output: J_t^{diag}, J_t^{off}

Locally linearized spectral representation. We consider a local linearization of the model in the Fourier domain:

$$\widehat{\epsilon}_{\text{pred}} \approx W_t \widehat{x}_t, \tag{7}$$

where $W_t \in \mathbb{C}^{|\Omega|\times|\Omega|}$ is a linear operator describing the transformation between input and output frequencies at input x_t. The entry $W_t(\omega_{\text{out}}, \omega_{\text{in}})$ quantifies how input frequency ω_{in} influences output frequency ω_{out}.

Marginal spectral statistics and their limitations. A common approach to analyzing generative models is through marginal spectral quantities such as the power spectral density (PSD) (Benita et al., 2025). Under the locally linearized model the output PSD is given by

$$\mathbb{E}\big[S(\epsilon_{\text{pred}}, \omega)\big] = \big[W_t \Sigma_{x_t} W_t^*\big]_{\omega,\omega}, \tag{8}$$

corresponding to the diagonal of the matrix $W_t W_t^*$, weighted by the PSD of the intermediate state x_t.

However, this characterization identifies the model up to unitary mixing of input frequencies–see (Horn & Johnson, 2012). This invariance implies that marginal spectral statistics characterize only an equivalence class of transformations, and cannot distinguish between models that differ in their cross-frequency interactions.

Interventional setup. To probe these cross-frequency interactions, we introduce an interventional approach in the Fourier domain. Given a sample x_t, we construct a perturbed input along the direction of ω_{in}:

$$\widehat{x_t^{h(\omega_{\text{in}})}}(\omega) = \widehat{x}_t(\omega) + \delta \mathbb{1}_{\omega_{\text{in}}}(\omega), \tag{9}$$

where $\mathbb{1}_{\omega_{\text{in}}}$ denotes the indicator function, which injects a fixed fraction of energy at that frequency. The perturbed signal induces a corresponding prediction $\epsilon_{\text{pred}}^{h(\omega_{\text{in}})}$.

Spectral sensitivity of the network. We quantify how this perturbation affects the output spectrum at frequency ω_{out}, via the directional derivative:

$$J_t(\omega_{\text{out}}, \omega_{\text{in}}) := \frac{\partial\, S(\epsilon_{\text{pred}}, \omega_{\text{out}})}{\partial\, S(x_t, \omega_{\text{in}})} \tag{10}$$

$$= \lim_{\delta\to 0} \frac{S(\epsilon_{\text{pred}}^{h(\omega_{\text{in}})}, \omega_{\text{out}}) - S(\epsilon_{\text{pred}}, \omega_{\text{out}})}{S(\widehat{x}_t^{\,h(\omega_{\text{in}})}, \omega_{\text{in}}) - S(\widehat{x}_t, \omega_{\text{in}})}. \tag{11}$$

This quantity measures the sensitivity of output spectral power at ω_{out} w.r.t the input power at ω_{in}.

While J_t can be defined at the level of individual samples via spectral power differences, it constitutes a noisy estimate of $|W_t(\omega_{\text{out}}, \omega_{\text{in}})|^2$. Accordingly, we consider its expectation over the data distribution, $\mathbb{E}_x[J_t(x)]$, and, by a slight abuse of notation, denote this averaged quantity again by J_t.

The resulting spectral sensitivity matrix, J_t naturally decomposes into the following components:

- **Diagonal response (J_t^{diag}):** $\omega_{\text{in}} = \omega_{\text{out}}$, capturing marginal behavior.
- **Off-diagonal response (J_t^{off}):** $\omega_{\text{in}} \neq \omega_{\text{out}}$, capturing cross-frequency interactions.

Interpretation and motivation of J_t. The matrix J_t can be interpreted as an energy-based proxy for the squared magnitude of the Jacobian of the model in the Fourier basis, obtained via interventional perturbations. Specifically, under a local linear approximation of the network, the sensitivity $J_t(\omega_{\text{out}}, \omega_{\text{in}})$ approximates the squared magnitude of the corresponding transfer weights. Unlike spectral covariance, which conflates model effects with data statistics, our interventional setup enforces independence across input modes by perturbing one frequency at a time. This ensures that the observed spectral structure is indeed induced by the learned dynamics of the model rather than by the underlying training data distribution.

The following proposition formalizes the relationship between J_t and the underlying linearized dynamics.

Proposition 4.1 (Interventional sensitivity as squared transfer magnitude)**.** *Under the local linear approximation*

$\widehat{\epsilon}_{\text{pred}} \approx W_t \widehat{x}_t$ *and for sufficiently small perturbations, the spectral sensitivity satisfies*

$$\mathbb{E}[J_t(\omega_{\text{out}}, \omega_{\text{in}})] \approx |W_t(\omega_{\text{out}}, \omega_{\text{in}})|^2. \tag{12}$$

See Section C for derivations and theoretical considerations.

From marginal summaries to interactions. This interpretation highlights a fundamental distinction: while marginal spectral statistics access only the diagonal of $W_t W_t^*$, the interventional sensitivity J_t approximates the entrywise energy of W_t. As a result, J_t^{off} provides direct access to cross-frequency interactions that are invisible to marginal analyses.

By explicitly considering the off-diagonal component J_t^{off}, we capture cross-frequency interactions and obtain a richer description of the learned transformation. This motivates using this quantity as a measure of cross-frequency coupling beyond marginal spectral statistics.

A theoretically well-documented regime. A useful reference case is one in which frequency-wise independence holds exactly. When the data covariance is diagonal in the Fourier basis and the noise is isotropic, Gaussian conditioning yields optimal dynamics that factorize across modes,

$$\widehat{f_t^{\text{opt}}}(\omega) = a_t(\omega)\, \widehat{x_t}(\omega). \tag{13}$$

Such decoupling is formalized for diffusion models in Theorem 3.5 of (Benita et al., 2025), and follows analogously in flow matching under the same assumptions. In this regime, perturbations at one frequency do not affect any other, so $J_t^{\text{off}} \equiv 0$, and marginal spectral quantities such as the PSD suffice to characterize the dynamics.

The reference regime as a baseline. This regime provides a natural baseline for interpreting learned generative models. Moreover, it is not purely theoretical: early-time dynamics are often well-approximated by Gaussian behavior with weak interactions (Biroli et al., 2024). In the experiments below, we make this baseline concrete by constructing synthetic Gaussian data with prescribed covariance in the Fourier domain. In this setting, departures from $J_t^{\text{off}} \equiv 0$ indicate genuine cross-frequency coupling that is invisible to marginal statistics, allowing us to distinguish frequency-wise independent behavior from models with nontrivial off-diagonal structure.

5. Experimental Setup

We now describe the experimental protocol used to evaluate spectral sensitivity under controlled frequency-domain perturbations. Detailed implementation choices, including data construction, training hyperparameters, and numerical considerations, are deferred to Section B.

Gaussian data. We construct synthetic Gaussian data in the Fourier domain with a prescribed covariance structure designed to mimic natural image spectra possibly with controlled cross-frequency mixing. Specifically, we consider both diagonal covariances (frequency-independent baseline) and structured off-diagonal perturbations that couple frequencies at fixed spectral offsets. Samples are transformed back to the spatial domain for model training.

Frequency-masked natural image datasets. We construct variants of CIFAR-10 and CelebA by removing specific frequency bands in the Fourier domain and transforming back to the spatial domain. This yields datasets with systematically altered spectral content while preserving spatial structure.

Generative modeling. For each dataset variant, we train generative models using flow matching and diffusion objectives with shared architectures within each setting (UNet for images, MLP for Gaussian data). This ensures that differences in behavior are attributable to the data rather than model capacity.

Validation. We verify that all models adequately fit their respective datasets using FID, observing comparable performance across variants. This ensures that differences in spectral sensitivity reflect learned dynamics rather than failure to model the data.

Spectral sensitivity measurement. We estimate the spectral sensitivity matrix J_t across multiple time steps by averaging over dataset samples and report both diagonal and off-diagonal components.

6. Spectral Sensitivity Analysis

We report results on real-world image data and on controlled Gaussian constructions with prescribed spectral covariance. Our goal is to assess whether marginal spectral statistics suffice to characterize generative dynamics, and to identify the role of cross-frequency interactions.

In the main text, we visualize J_t as heatmaps over input frequency ω_{in} and response frequency ω_{out} at the fixed time step $t = 0.9$; the full temporal evolution of generative dynamics is provided in Figures 10 to 13.

6.1. Energy Distribution in the Components of J_t

To quantitatively assess the relative contribution of diagonal and off-diagonal components of J_t, we measure their Frobenius norms across models trained on datasets of increasing size. We consider CIFAR-10 and evaluate these norms separately on both training and test data, yielding a controlled comparison between in-distribution fitting and

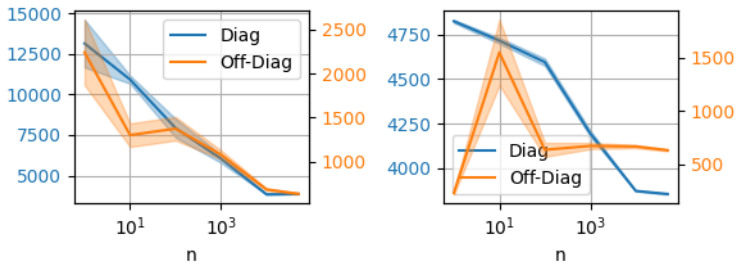

Figure 4. **Off-Diagonal Components Carry Less Energy but Distinguish Model Regimes** Frobenius norms of diagonal (blue) and off-diagonal (orange) components of J_t as a function of training set size of CIFAR-10 with $t = 0.9$. Left: test data. Right: training data. While the diagonal component carries substantially more energy and exhibits a similar decreasing trend in both regimes, the off-diagonal component behaves differently: it tracks the diagonal trend on test data (up to scale), but shows a pronounced peak at intermediate dataset sizes on training data. This highlights that off-diagonal structure, despite lower magnitude, may capture regime-dependent differences invisible to diagonal statistics.

generalization behavior.

The results are presented in Figure 4, with test data shown on the left and training data on the right. We observe that the diagonal component consistently dominates in magnitude and exhibits a smooth decreasing trend as the number of training samples increases. This trend is remarkably similar across both training and test regimes, suggesting that marginal (frequency-wise) responses alone provide a limited view of the learned dynamics.

In contrast, the off-diagonal component, while significantly smaller in magnitude, displays qualitatively different behavior across regimes. On test data, it broadly follows the same decreasing trend as the diagonal (up to scale). However, on training data it exhibits a pronounced peak at intermediate dataset sizes (around $n \approx 10$), deviating sharply from its test-time behavior. This indicates that off-diagonal structure may capture aspects of the model that differentiate training and test distributions, which are not reflected in marginal spectral statistics.

We conjecture that this peak is related to the transition between overparameterized and underparameterized regimes. A deeper theoretical understanding of this phenomenon is left for future work. Overall, these results provide quantitative support for our central claim: off-diagonal components of J_t encode complementary information about the learned dynamics that is invisible to diagonal, PSD-like summaries.

6.2. Diagonal Response: Low Variance Across Models

We first examine the diagonal component J_t^{diag}, corresponding to per-frequency sensitivity (top row in Figure 3 for CIFAR-10, and middle row in Figure 2 for the controlled Gaussian experiment).

Across all models f_k, the diagonal response on $\mathcal{D}_0$ exhibits highly similar structure, despite substantial differences in the training data. This indicates that marginal spectral responses capture broad trends, but may fail to distinguish models trained on structurally different distributions.

The diagonal response also follows a consistent temporal pattern shared across models. At early times, responses are approximately uniform across frequencies; as $t \to 1$, they progressively concentrate toward higher frequencies, while their overall magnitude increases. This yields more localized, high-frequency-dominated responses near the data distribution (see Figures 10, 12 and 15).

These behaviors are consistent with prior analyses based on marginal spectral quantities. At the same time, they reinforce the limitation of diagonal summaries: while they capture global spectral trends, they may not reflect the differences induced by the training distribution.

6.3. Off-Diagonal Response: Revealing Interactions

We now turn to the off-diagonal component J_t^{off}, which encodes cross-frequency interactions (middle and bottom rows in Figure 3 for CIFAR-10 and CelebA, and bottom row in Figure 2 for the controlled Gaussian experiment).

In contrast to the diagonal case, off-diagonal responses differ substantially across models, especially at larger time steps (see Figures 11, 13 and 15). While responses are nearly indistinguishable at $t \approx 0$, clear divergences emerge as t approaches to the late-time dynamics, reflecting the fact that the networks share the same noise distribution early in the trajectory and progressively move toward their respective data distributions.

At small t, perturbations at a given input frequency induce approximately uniform responses across output frequencies, indicating weakly structured early-time dynamics. As $t \to 1$, however, the responses become increasingly structured and concentrated. In the image dataset configurations, this structure aligns closely with the spectral bands removed in the training datasets $\mathcal{D}_k$. In the controlled Gaussian set-

ting, where off-diagonal covariance is introduced at a fixed spectral offset, J_t^{off} recovers the imposed structure, providing direct evidence that the sensitivity captures genuine cross-frequency coupling.

In the image dataset configurations, perturbations at frequency ω_{in} also induce responses that are systematically biased toward higher frequencies. In parallel, both the magnitude and dynamic range of off-diagonal responses increase with time, leading to more pronounced interaction patterns near the data distribution (see Figure 13).

Main takeaway

Diagonal responses exhibit consistent temporal trends and low variance across models, whereas off-diagonal responses are strongly model-dependent and encode dataset-specific structure. The empirical evidence suggests that the distinguishing features of learned generative dynamics are often concentrated in cross-frequency interactions rather than in marginal spectral behavior alone.

7. Limitations and Future Work

Our work highlights several directions to extend the analysis of cross-frequency interactions in generative models.

This study considers two classes of structured perturbations: frequency masking in real image data and controlled covariance modifications in a Gaussian setting. While these provide complementary perspectives—realistic and fully controlled—they remain relatively simple interventions. A natural extension is to consider richer classes of perturbations, such as varying bandwidths or more general covariance structures, to better understand how different forms of spectral structure influence learned dynamics.

Our analysis is conducted in the spatial Fourier domain, which is well-suited for image data under approximate stationarity. Extending this framework to other domains that admit analogous spectral representations, such as graphs or manifolds, would enable a broader characterization of interaction structure beyond Euclidean settings.

The sensitivity measure J_t is based on local perturbations and admits an interpretation through a linearization of the model. While this yields a tractable and interpretable probe of the dynamics, it does not capture higher-order nonlinear interactions. Developing measures that go beyond the local regime and characterize nonlinear cross-frequency effects remains an important direction for future work.

Recent analytical work on convolutional diffusion models has emphasized a patch-based perspective, where local spatial interactions give rise to structured generative behavior (Kamb & Ganguli, 2025). An interesting direction for future work is to relate this viewpoint to the frequency-domain analysis. While patch-based analyses characterize locality and compositional structure in the spatial domain, our results suggest that complementary insights can be obtained by studying cross-frequency interactions in the spectral domain. Establishing a precise connection between these perspectives may provide a more unified understanding of how generative models organize and transform structure across scales.

From a theoretical perspective, an important open question is to explain the emergence of structured off-diagonal interactions in realistic settings. While the Gaussian construction provides a setting where cross-frequency coupling can be precisely controlled and analyzed, a principled understanding of how such interactions arise during training—particularly the observed high-frequency bias—remains an open problem.

8. Conclusion

In this work, we showed that marginal spectral statistics may be insufficient to characterize the learned dynamics of deep generative models. Using an interventional sensitivity measure, J_t, we decomposed model transformations into diagonal and off-diagonal components, enabling a direct analysis of cross-frequency interactions.

Across two complementary settings—real-world image data with structured frequency masking and controlled Gaussian constructions with prescribed spectral covariance—we observed a consistent phenomenon: diagonal responses exhibit minimal variation across models and fail to reflect substantial differences in the underlying data distributions. In contrast, off-diagonal interactions clearly reflect these differences. Particularly, in the controlled Gaussian setting, J_t^{off} recovers the imposed covariance structure, providing direct evidence that the proposed measure faithfully captures cross-frequency coupling.

These results indicate that the defining characteristics of learned generative dynamics may emerge through cross-frequency interactions, precisely where idealized mode-independence breaks down. Consequently, this work challenges the reliance on marginal spectral summaries, such as power spectral density, and highlights the need for interaction-aware analyses to understand the internal mechanisms of deep generative models.

References

Benita, R., Elad, M., and Keshet, J. Spectral Analysis of Diffusion Models with Application to Schedule Design, December 2025. URL http://arxiv.org/abs/2502.00180. arXiv:2502.00180 [cs].

Benita, R., Elad, M., and Keshet, J. Analyzing and Guiding Zero-Shot Posterior Sampling in Diffusion Models, February 2026. URL http://arxiv.org/abs/2602.07715. arXiv:2602.07715 [cs].

Biroli, G., Bonnaire, T., Bortoli, V. d., and Mézard, M. Dynamical Regimes of Diffusion Models. *Nature Communications*, 15(1):9957, November 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-54281-3. URL http://arxiv.org/abs/2402.18491. arXiv:2402.18491 [cs].

Brutti, P., Durastanti, C., and Mari, F. Spectral Diffusion Models on the Sphere, January 2026. URL http://arxiv.org/abs/2601.20498. arXiv:2601.20498 [math].

Dhariwal, P. and Nichol, A. Diffusion Models Beat GANs on Image Synthesis, June 2021. URL http://arxiv.org/abs/2105.05233. arXiv:2105.05233 [cs].

Falck, F., Pandeva, T., Zahirnia, K., Lawrence, R., Turner, R., Meeds, E., Zazo, J., and Karmalkar, S. A Fourier Space Perspective on Diffusion Models, May 2025. URL http://arxiv.org/abs/2505.11278. arXiv:2505.11278 [stat].

Freirich, D., Michaeli, T., and Meir, R. A Theory of the Distortion-Perception Tradeoff in Wasserstein Space, July 2021. URL http://arxiv.org/abs/2107.02555. arXiv:2107.02555 [eess].

Ho, J., Jain, A., and Abbeel, P. Denoising Diffusion Probabilistic Models, December 2020. URL http://arxiv.org/abs/2006.11239. arXiv:2006.11239 [cs].

Horn, R. A. and Johnson, C. R. Matrix Analysis, October 2012. URL https://www.cambridge.org/highereducation/books/matrix-analysis/FDA3627DC2B9F5C3DF2FD8C3CC136B48.

Kamb, M. and Ganguli, S. An analytic theory of creativity in convolutional diffusion models, June 2025. URL http://arxiv.org/abs/2412.20292. arXiv:2412.20292 [cs].

Mehrpanah, A., Englesson, E., and Azizpour, H. On Spectral Properties of Gradient-based Explanation Methods. volume 15145, pp. 282–299. 2025. URL http://arxiv.org/abs/2508.10595. arXiv:2508.10595 [cs].

Moghadas, S. M., Cornelis, B., and Munteanu, A. FreqFlow: Long-term forecasting using lightweight flow matching, November 2025. URL http://arxiv.org/abs/2511.16426. arXiv:2511.16426 [cs].

Phillips, A., Seror, T., Hutchinson, M., Bortoli, V. D., Doucet, A., and Mathieu, E. Spectral Diffusion Processes, November 2022. URL http://arxiv.org/abs/2209.14125. arXiv:2209.14125 [stat].

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-Based Generative Modeling through Stochastic Differential Equations, February 2021. URL http://arxiv.org/abs/2011.13456. arXiv:2011.13456 [cs].

Tange, O. Gnu parallel 20260322 ('this is the last battle'), March 2025. URL https://doi.org/10.5281/zenodo.19321428. GNU Parallel is a general parallelizer to run multiple serial command line programs in parallel without changing them.

Tivnan, M., Teneggi, J., Lee, T.-C., Zhang, R., Boedeker, K., Cai, L., Gang, G. J., Sulam, J., and Stayman, J. W. Fourier Diffusion Models: A Method to Control MTF and NPS in Score-Based Stochastic Image Generation, March 2023. URL http://arxiv.org/abs/2303.13285. arXiv:2303.13285 [physics].

Tong, A. Conditional flow matching (code repository). https://github.com/atong01/conditional-flow-matching, 2023. Accessed: 2026-04-21.

Tong, A., Fatras, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., and Bengio, Y. Improving and generalizing flow-based generative models with minibatch optimal transport, March 2024. URL http://arxiv.org/abs/2302.00482. arXiv:2302.00482 [cs].

Wang, B. and Pehlevan, C. An Analytical Theory of Spectral Bias in the Learning Dynamics of Diffusion Models, April 2026. URL http://arxiv.org/abs/2503.03206. arXiv:2503.03206 [cs].

Wang, H., Pan, J., Wu, H., Zhang, F., and Wu, T. FourierFlow: Frequency-aware Flow Matching for Generative Turbulence Modeling, June 2025. URL http://arxiv.org/abs/2506.00862. arXiv:2506.00862 [cs].

A. Additional Visualizations of Generated and Perturbed Samples

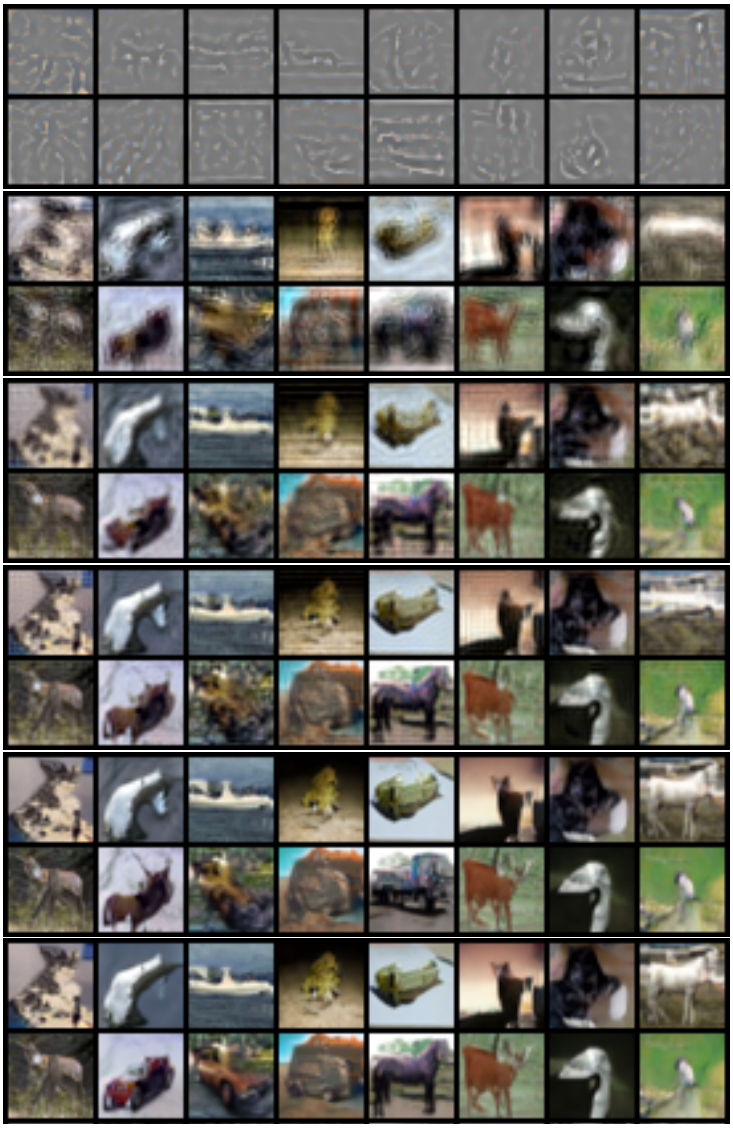

Figure 5. **Qualitative Effect of Frequency-Band Perturbations on Generated Samples Across Models (CIFAR-10).** Unconditional (classifier-free guidance) samples from conditional flow matching models trained on frequency-perturbed versions of CIFAR-10 together with the unperturbed version of CIFAR-10 (last two rows). For each model, two rows (of 8 samples) are visualized using the same random seed to facilitate qualitative comparison. Rows are organized by the perturbed frequency band: the first two rows correspond to masking $[0.0, 0.2]$, the next two to $[0.2, 0.4]$, followed by $[0.4, 0.6]$, and so on in the same order as Figure 7; the final two rows correspond to the unperturbed (reference) dataset. The most pronounced qualitative differences appear in the lowest-frequency perturbations (top rows), while higher-frequency perturbations yield subtler visual changes, despite the quantitative differences in their power spectral density (PSD) as shown in Figure 7.

B. Implementation Details

B.1. Gaussian Data Construction

We construct Gaussian data directly in the Fourier domain with a prescribed covariance. The marginal spectrum follows

$$S(\omega) = \frac{1}{1 + \|\omega\|^2}, \tag{14}$$

which mimics the decay observed in natural images.

To introduce cross-frequency interactions, we define the covariance

$$C = D + \kappa(T_k D + D T_k^\top), \tag{15}$$

where $D = \mathrm{diag}(S(\omega))$, T_k is a shift operator $(T_k z)(\omega) = z(\omega + k)$, $\kappa = 0.5$, and $k \in \{0, 3, 5, 7, 9, 11\}$. The case $k = 0$ corresponds to a purely diagonal covariance. Samples are drawn in the Fourier domain and mapped to the spatial domain using the inverse Fourier transform.

B.2. Frequency Masking for Natural Images

For CIFAR-10 and CelebA, we construct dataset variants by masking radial frequency bands in the Fourier domain. For each image x with transform $\hat{x}$, we apply a band-stop mask $h_k(\omega)$:

$$\mathcal{D}_k = \{x^{h_k(\omega)} : x \in \mathcal{D}_0\}, \tag{16}$$

where h_k removes frequencies in the interval $\omega \in [0.2(k-1),\ 0.2k]$. The modified spectrum is transformed back to the spatial domain via inverse Fourier transform.

B.3. Training Details

We train flow matching and diffusion models following (Tong, 2023; Dhariwal & Nichol, 2021). UNet architectures are used for image datasets and MLPs for Gaussian data.

Training uses the OT-CFM objective with linear interpolation between data and Gaussian noise. Models are trained for up to 200,000 steps with early stopping (patience 10,000, threshold 10^{-3}), batch size 128, and learning rate 10^{-3} with warmup. Gradient clipping is applied. For CIFAR-10, classifier-free guidance is used during training with dropout 0.1.

B.4. Sampling and Evaluation

Samples are generated using an Euler solver with 100 steps. For CIFAR-10, classifier-free guidance with scale 1.5 is used. Model quality is evaluated using FID, with scores of approximately 8.5 for CIFAR-10 variants and 13 for CelebA variants.

B.5. Spectral Sensitivity Estimation

We estimate the spectral sensitivity matrix J_t (defined in Algorithm 1) across multiple time steps. Estimates are obtained by averaging over 1024 dataset samples.

To ensure numerical stability, we replace the binary perturbation mask with a smoothed version and use a convex combination when injecting perturbations. When the denominator in J_t is zero, the corresponding entry is set to zero. Unless otherwise specified, perturbations inject 1% additional energy at a selected input frequency, and responses are measured via changes in output spectral power.

C. Derivations and Theoretical Considerations

We derive the relationship between the proposed interventional sensitivity and the squared magnitude of the Fourier-domain transfer weights under a local linear approximation.

We begin by approximating the model output in the Fourier domain by a linear operator:

$$\widehat{\epsilon}_{\mathrm{pred}}(\omega) \approx W_t \widehat{x_t}(\omega) \tag{17}$$

We then introduce a perturbation localized at the input frequency ω_{in}:

$$\widehat{x_t}^h := \widehat{x_t} + \delta \mathbb{1}_{\omega_{\mathrm{in}}} \tag{18}$$

Let $\mathbb{E}[\delta^2] = \sigma_\delta^2$, $\mathbb{E}[\delta] = 0$. We assume independence between δ and the input Fourier coefficients. That is, δ is statistically independent of $\widehat{x_t}$ (i.e., $\delta \perp \widehat{x_t}(\omega)$ for all ω).

Under the linear approximation, the perturbed prediction becomes

$$\widehat{\epsilon}^h_{\mathrm{pred}}(\omega) \approx \widehat{\epsilon}_{\mathrm{pred}}(\omega) + W_t \delta \mathbb{1}_{\omega_{\mathrm{in}}}(\omega) \tag{19}$$

$$:= \widehat{\epsilon}_{\mathrm{pred}}(\omega) + W(\omega_{\mathrm{out}}, \omega_{\mathrm{in}})\delta \tag{20}$$

where $W(\omega_{\text{out}}, \omega_{\text{in}})$ denotes the transfer from input frequency ω_{in} to output frequency ω_{out}.

We next consider the change in output spectral power induced by this perturbation. The numerator measures the change in the predicted output spectrum:

$$\begin{aligned}
\Delta^{\text{num}}_{\omega_{\text{in}}}(\omega_{\text{out}}) &:= S(\widehat{\epsilon}^{h}_{\text{pred}}, \omega_{\text{out}}) - S(\widehat{\epsilon}_{\text{pred}}, \omega_{\text{out}}) && (21)\\
&= |\widehat{\epsilon}^{h}_{\text{pred}}(\omega_{\text{out}})|^2 - |\widehat{\epsilon}_{\text{pred}}(\omega_{\text{out}})|^2 && (22)\\
&\approx |\widehat{\epsilon}_{\text{pred}}(\omega_{\text{out}}) + W_t \delta \mathbb{1}_{\omega_{\text{in}}}(\omega_{\text{out}})|^2 - |\widehat{\epsilon}_{\text{pred}}(\omega_{\text{out}})|^2 && (23)\\
&= |\widehat{\epsilon}_{\text{pred}}(\omega_{\text{out}}) + W(\omega_{\text{out}}, \omega_{\text{in}})\delta|^2 - |\widehat{\epsilon}_{\text{pred}}(\omega_{\text{out}})|^2 && (24)\\
&= \delta^2 |W(\omega_{\text{out}}, \omega_{\text{in}})|^2 + 2\Re\Big(\widehat{\epsilon}_{\text{pred}}(\omega_{\text{out}})\overline{\delta W(\omega_{\text{out}}, \omega_{\text{in}})}\Big) && (25)
\end{aligned}$$

Similarly, the denominator measures the perturbation energy injected at the input frequency:

$$\begin{aligned}
\Delta^{\text{denom}}_{\omega_{\text{in}}}(\omega_{\text{in}}) &:= S(\widehat{x}_t{}^{h}, \omega_{\text{in}}) - S(\widehat{x}_t, \omega_{\text{in}}) && (26)\\
&= |\widehat{x}_t{}^{h}(\omega_{\text{in}})|^2 - |\widehat{x}_t(\omega_{\text{in}})|^2 && (27)\\
&= |\widehat{x}_t(\omega_{\text{in}}) + \delta|^2 - |\widehat{x}_t(\omega_{\text{in}})|^2 && (28)\\
&= \delta^2 + 2\Re(\delta\overline{\widehat{x}_t}) && (29)\\
& && (30)
\end{aligned}$$

Taking expectations over the perturbation and the data distribution yields

$$\begin{aligned}
\mathbb{E}[\Delta^{\text{num}}_{\omega_{\text{in}}}(\omega_{\text{out}})] &= \mathbb{E}\Big[\delta^2 |W(\omega_{\text{out}}, \omega_{\text{in}})|^2 + 2\Re\Big(\sum_{\nu} W(\omega_{\text{out}}, \nu)\widehat{x}_t(\nu)\overline{\delta W(\omega_{\text{out}}, \omega_{\text{in}})}\Big)\Big] && (31)\\
&= \sigma_\delta^2 \mathbb{E}[|W(\omega_{\text{out}}, \omega_{\text{in}})|^2] && (32)
\end{aligned}$$

Applying the same argument to the denominator gives

$$\begin{aligned}
\mathbb{E}[\Delta^{\text{denom}}_{\omega_{\text{in}}}(\omega_{\text{in}})] &= \mathbb{E}[\delta^2 + 2\Re(\delta\overline{\widehat{x}_t})] && (33)\\
&= \sigma_\delta^2 && (34)
\end{aligned}$$

We now define the Jacobian-like sensitivity by taking the ratio of the expected variations and passing to the small-noise limit:

$$\begin{aligned}
J_t(\omega_{\text{out}}, \omega_{\text{in}}) &:= \lim_{\sigma_\delta^2 \to 0} \frac{\mathbb{E}[\Delta^{\text{num}}_{\omega_{\text{in}}}(\omega_{\text{out}})]}{\mathbb{E}[\Delta^{\text{denom}}_{\omega_{\text{in}}}(\omega_{\text{in}})]} && (35)\\
&= \lim_{\sigma_\delta^2 \to 0} \frac{\sigma_\delta^2 \mathbb{E}[|W(\omega_{\text{out}}, \omega_{\text{in}})|^2]}{\sigma_\delta^2} && (36)\\
&= \mathbb{E}[|W(\omega_{\text{out}}, \omega_{\text{in}})|^2] && (37)
\end{aligned}$$

The variance of the perturbation acts as a low-pass filter in the Fourier domain (Mehrpanah et al., 2025), yielding a smoother estimate of the transfer matrix J_t for finite values of σ_δ (we use $\sigma_\delta = 10^{-5}$ in practice).

While the theoretical formulation corresponds to the limit $\sigma_\delta \to 0$, we find that backpropagation-based estimation in this regime is numerically unstable and produces highly noisy estimates of J_t. Instead, we employ the finite-difference estimator described in Algorithm 1, and approximate the expectation by averaging the ratios $\frac{\Delta^{\text{num}}_{\omega_{\text{in}}}}{\Delta^{\text{denom}}_{\omega_{\text{in}}}}$ over samples.

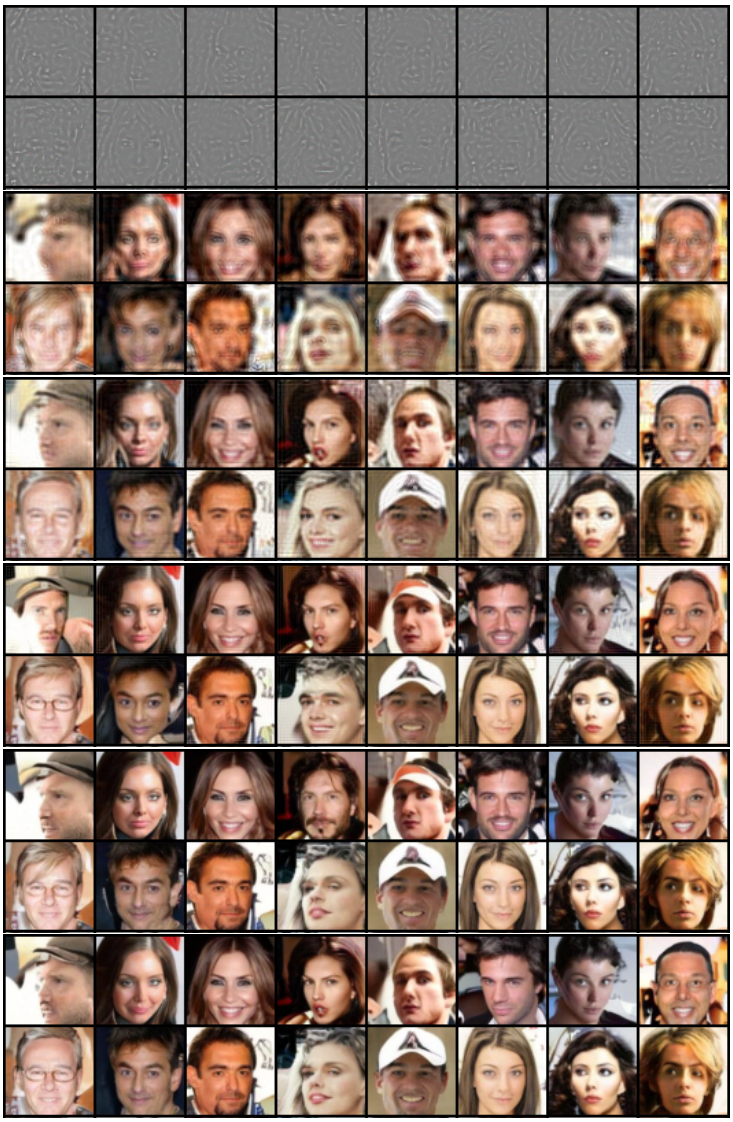

Figure 6. **Qualitative Effect of Frequency-Band Perturbations on Generated Samples Across Models (CelebA 64x64).** Unconditional (classifier-free guidance) samples from conditional flow matching models trained on frequency-perturbed versions of CelebA together with the unperturbed version of CelebA (last two rows). For each model, two rows (of 8 samples) are visualized using the same random seed to facilitate qualitative comparison. Rows are organized by the perturbed frequency band: the first two rows correspond to masking $[0.8, 1.0]$, the next two to $[0.6, 0.8]$, followed by $[0.4, 0.6]$, and so on in the same order as Figure 8; the final two rows correspond to the unperturbed (reference) dataset. The most pronounced qualitative differences appear in the lowest-frequency perturbations (top rows), while higher-frequency perturbations yield subtler visual changes, despite the quantitative differences in their power spectral density (PSD) as shown in Figure 8.

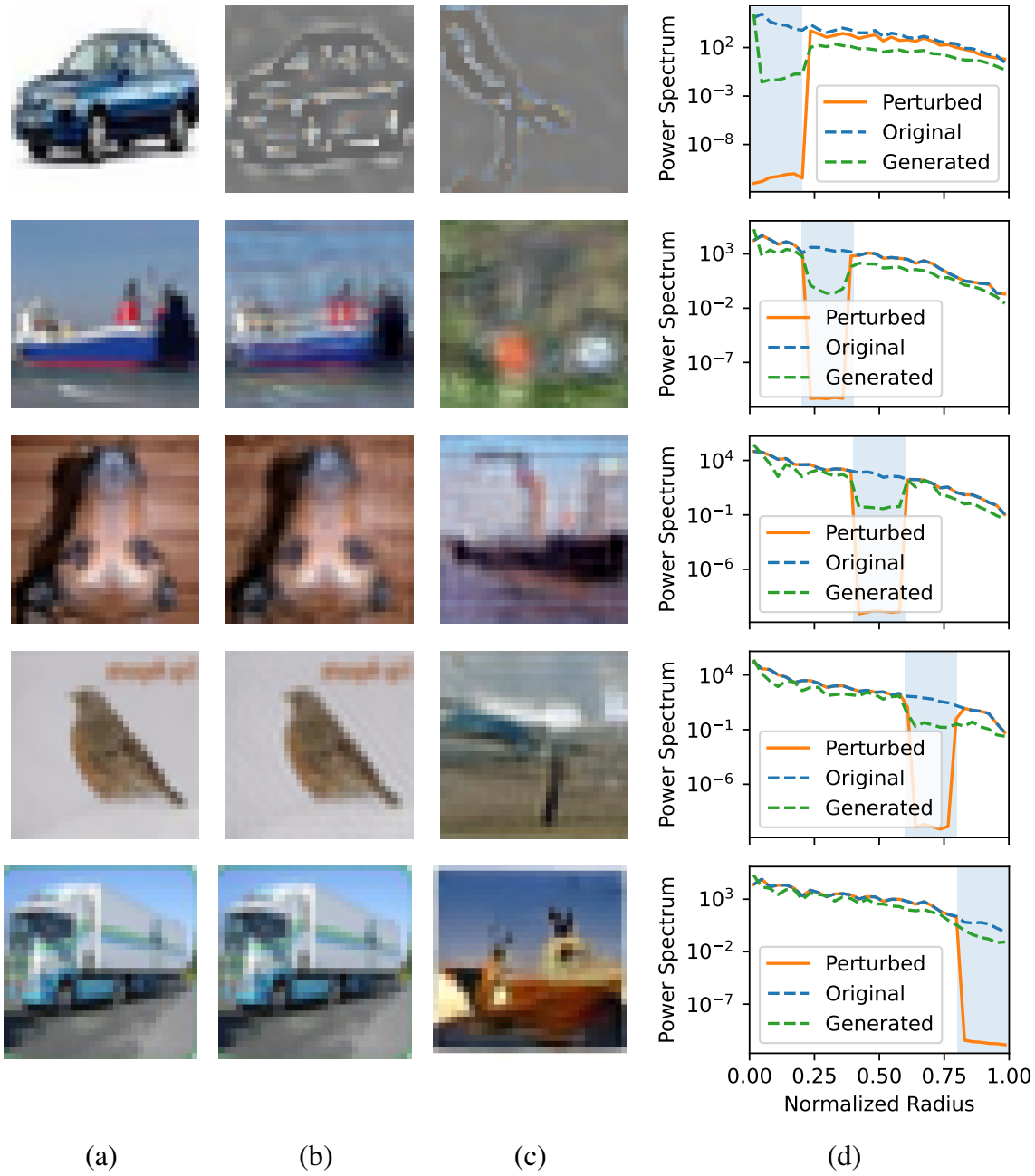

Figure 7. **Spectral Effects of Frequency Perturbations in Data and Generated Samples. (a)** Representative samples from CIFAR-10. **(b)** Corresponding frequency-filtered versions obtained via band-stop masking in the Fourier domain. **(c)** Unconditional samples (classifier-free guidance) from a conditional flow matching model trained on the corresponding perturbed dataset. **(d)** Power spectral densities (PSD) of original, perturbed, and generated samples. Notably, the generated samples exhibit a visible drop in spectral power within the filtered frequency band, often observable even at the level of individual samples.

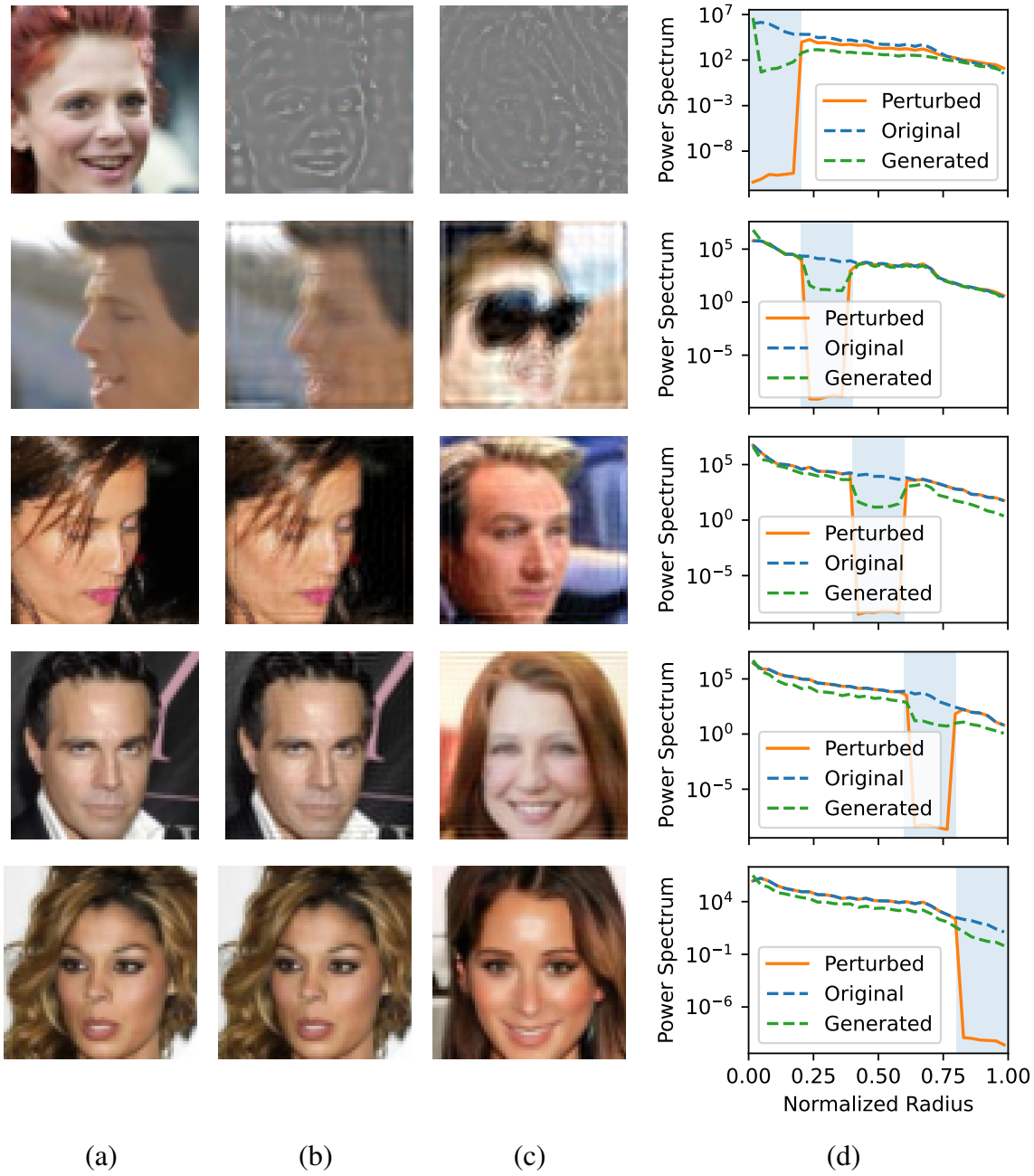

Figure 8. **Spectral Effects of Frequency Perturbations in Data and Generated Samples. (a)** Representative samples from CelebA. **(b)** Corresponding frequency-filtered versions obtained via band-stop masking in the Fourier domain. **(c)** Unconditional samples from a DDPM model trained on the corresponding perturbed dataset. **(d)** Power spectral densities (PSD) of original, perturbed, and generated samples. Notably, the generated samples exhibit a visible drop in spectral power within the filtered frequency band, often observable even at the level of individual samples.

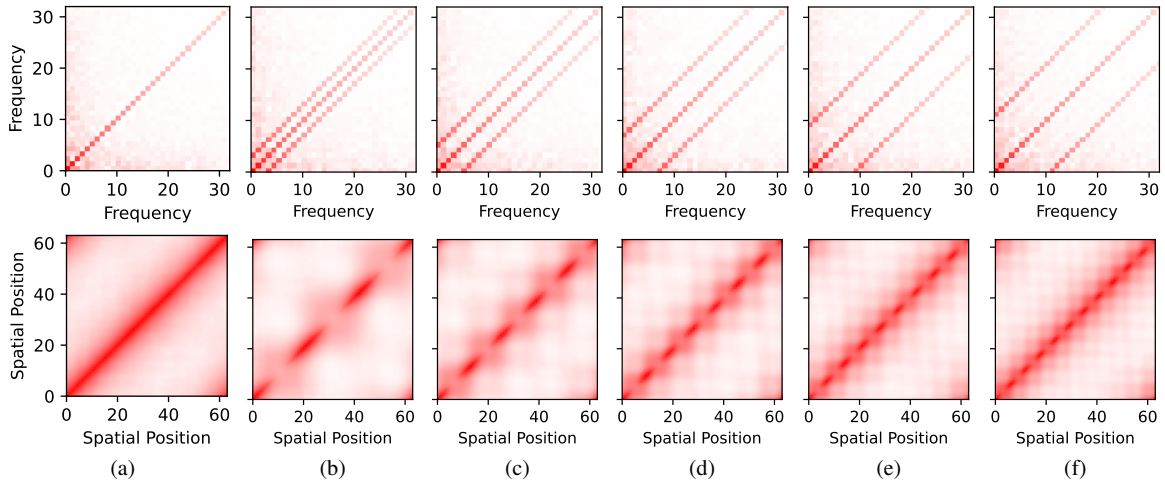

Figure 9. **Prescribed Covariance Structure in The Gaussian Setting.** Visualization of the covariance matrices used to generate Gaussian datasets with controlled spectral interactions. Top row: covariance in the Fourier domain, constructed according to **??**. Bottom row: corresponding spatial-domain covariance obtained via inverse Fourier transform. (**a**) Reference case $k = 0$, yielding a purely diagonal covariance in the Fourier domain. (**b–f**) Increasing shift values $k \in \{3, 5, 7, 9, 11\}$ introduce banded off-diagonal structure at a fixed spectral offset, visible as parallel bands away from the main diagonal. These structured cross-frequency correlations in the Fourier domain induce corresponding spatial correlations after transformation.

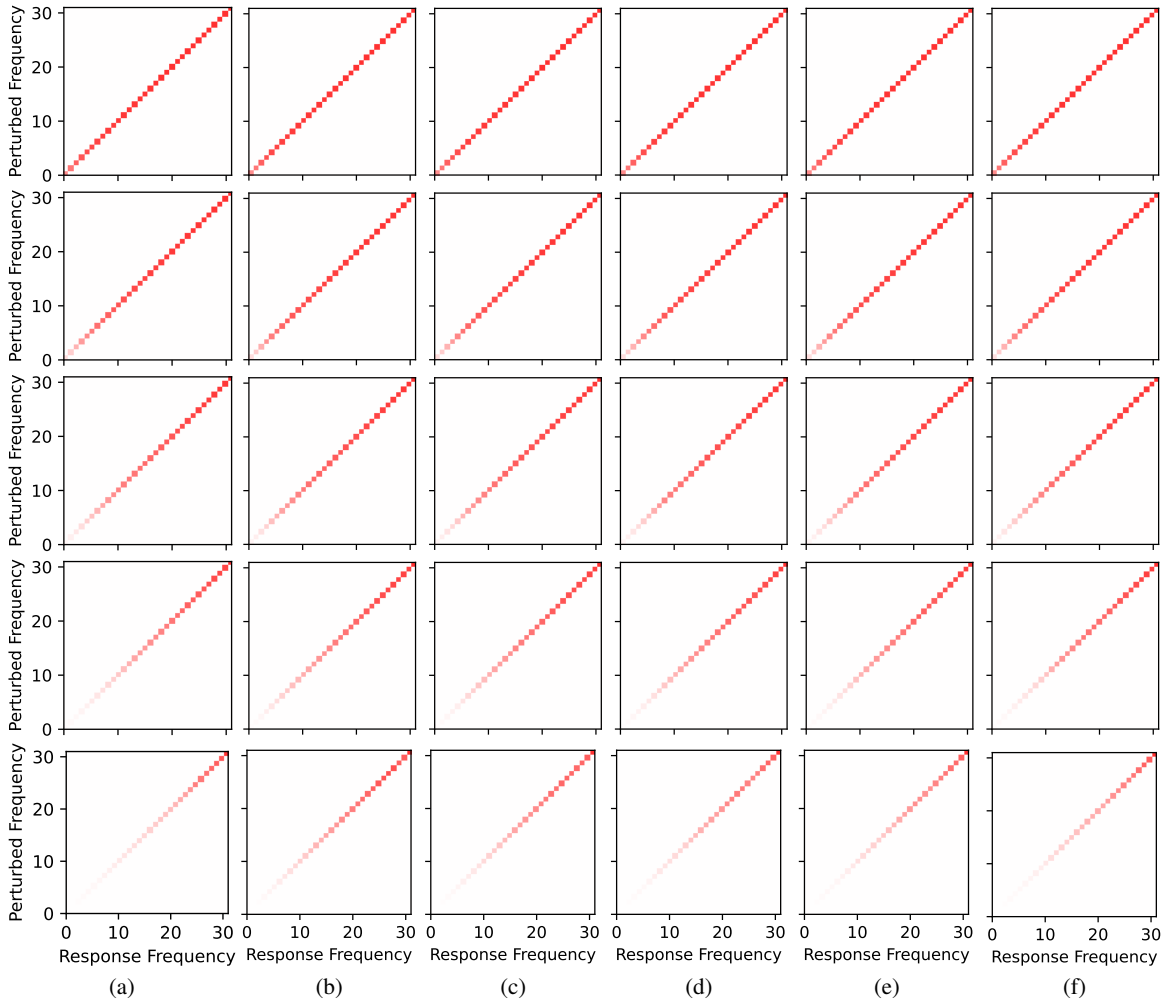

Figure 10. **Diagonal spectral sensitivity is largely invariant to controlled cross-frequency coupling (Gaussian setting).** Diagonal component of the spectral sensitivity response (defined in Equation (11)) evaluated on the reference dataset $\mathcal{D}_0$ ($k = 0$). Each column corresponds to a conditional flow matching (CFM) model f_k trained on Gaussian data with prescribed covariance structure. Each visualization shows the average response over 1024 samples. Rows correspond to time steps $t \in \{0.1, 0.3, \ldots, 0.9\}$ ordered from top to bottom. **(a)** Model trained on the reference dataset ($k = 0$), with purely diagonal covariance in the Fourier domain. **(b)–(f)** Models trained on datasets with increasing spectral shift $k \in \{3, 5, 7, 9, 11\}$, introducing structured off-diagonal covariance at fixed frequency offsets (see Figure 9). Across columns, the diagonal responses exhibit minimal variation despite substantial differences in the underlying covariance structure, indicating that marginal (per-frequency) sensitivity is largely insensitive to the introduced cross-frequency interactions. Similar to the CIFAR-10 setting (see Figure 12), the diagonal response follows a consistent temporal pattern: it is approximately uniform across frequencies at early times ($t \approx 0$), and progressively concentrates toward higher frequencies as $t \to 1$. Overall, these results confirm that diagonal spectral statistics may fail to capture structured deviations from frequency-wise independence, even in a controlled Gaussian setting where the ground-truth covariance is known.

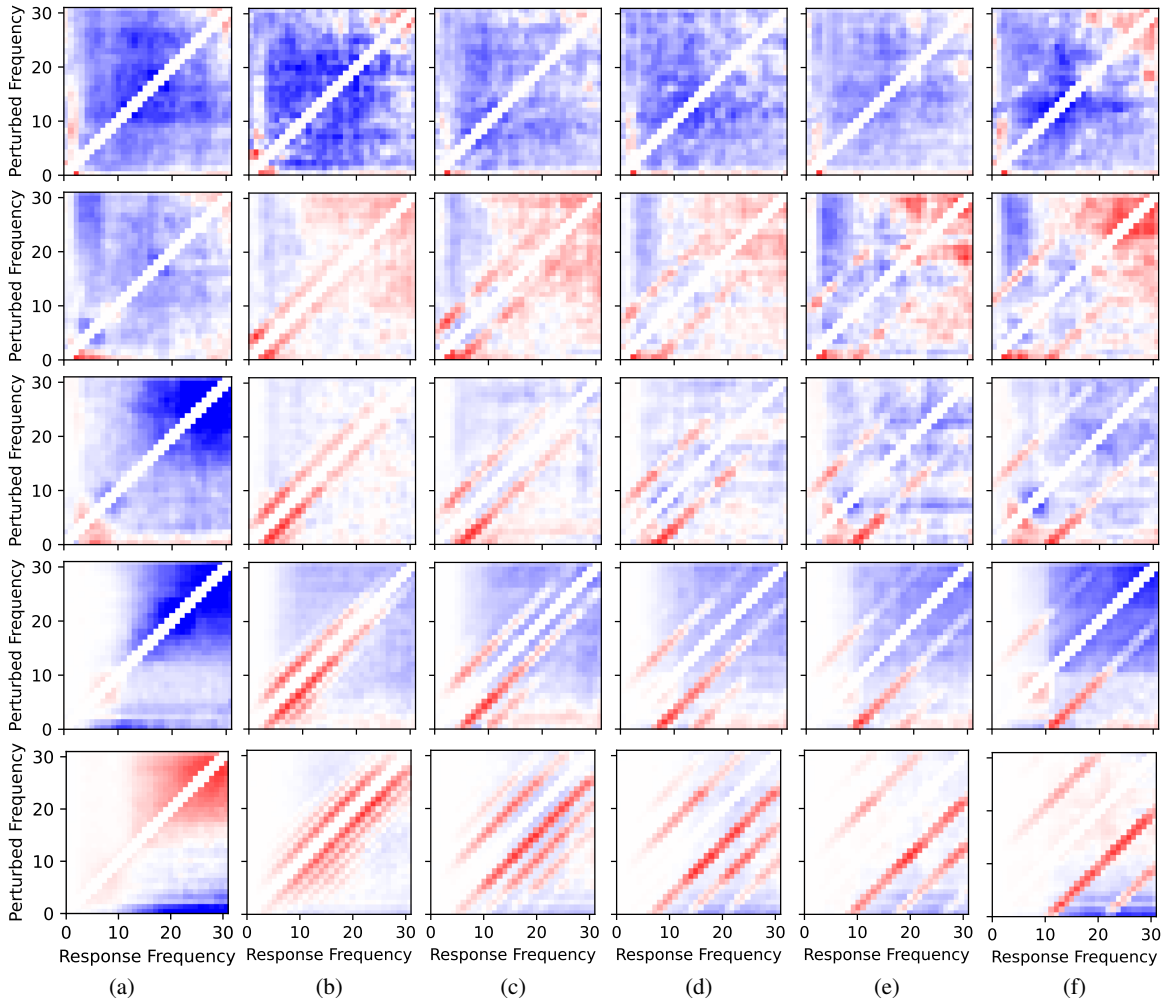

Figure 11. **Off-diagonal spectral sensitivity recovers prescribed cross-frequency structure (Gaussian setting).** Off-diagonal component of the spectral sensitivity response (defined in Equation (11)) evaluated on the reference dataset $\mathcal{D}_0$ ($k = 0$). Each column corresponds to a conditional flow matching (CFM) model f_k trained on Gaussian data with prescribed covariance structure (see Figure 9). Each visualization shows the average response over 1024 samples. Rows correspond to time steps $t \in \{0.1, 0.3, \ldots, 0.9\}$ ordered from top to bottom. **(a)** Model trained on the reference dataset ($k = 0$), with purely diagonal covariance in the Fourier domain. **(b)–(f)** Models trained on datasets with increasing spectral shift $k \in \{3, 5, 7, 9, 11\}$, introducing banded off-diagonal covariance at fixed frequency offsets. At early times ($t \approx 0$), the off-diagonal response is approximately uniform across frequencies, reflecting the independent structure of the dynamics near the Gaussian noise regime. As $t \to 1$, structured interaction patterns emerge and progressively align with the prescribed off-diagonal covariance of the training data. In particular, the location of the dominant response shifts with k, matching the imposed spectral offset. In contrast to the diagonal case (Figure 10), the off-diagonal sensitivity clearly differentiates between models and accurately recovers the known cross-frequency coupling. This provides direct evidence that J_t^{off} captures genuine interaction structure induced by the model, validating it as a diagnostic for cross-frequency dependencies.

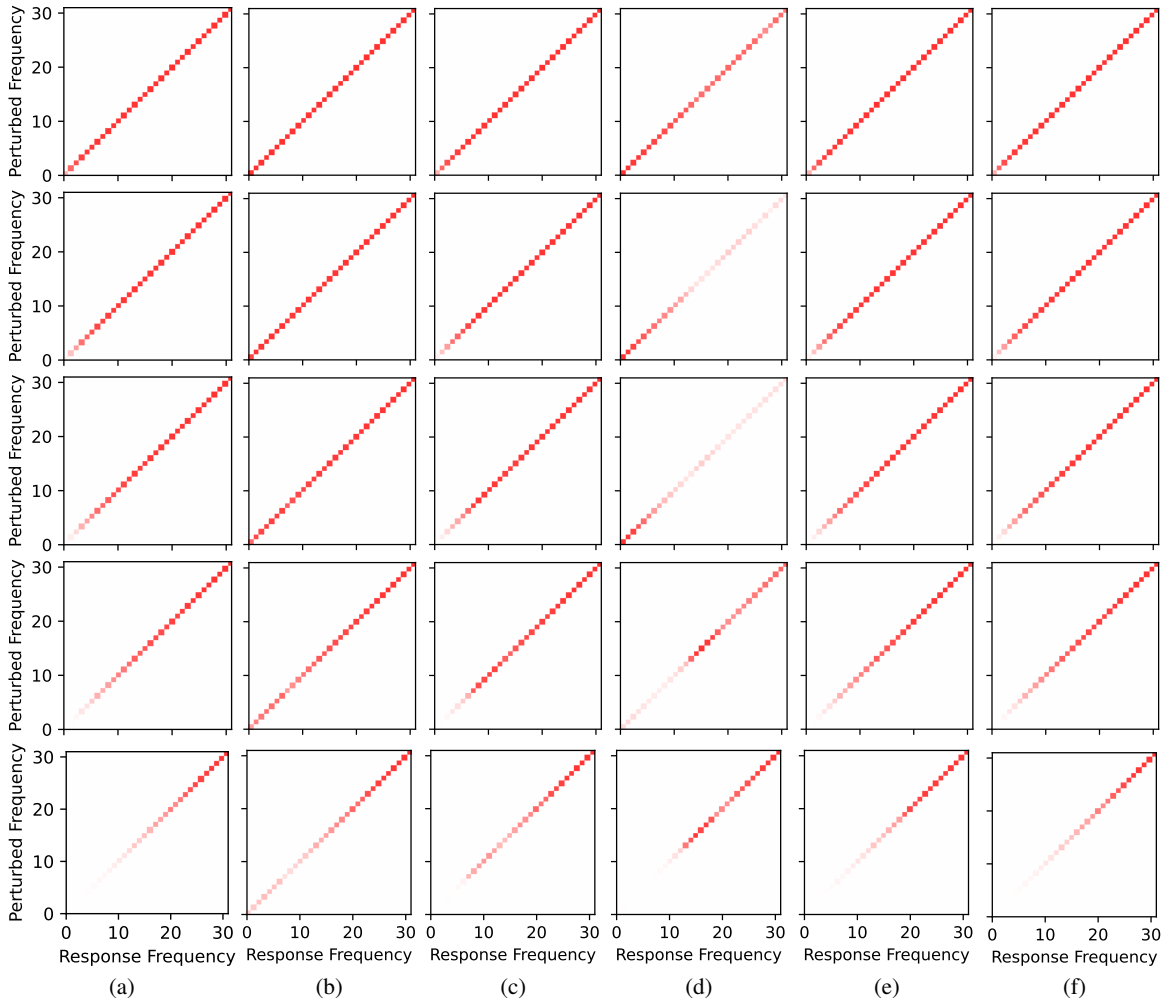

Figure 12. **Diagonal Spectral Sensitivity is Largely Invariant to Frequency-Domain Perturbations (CIFAR-10 32x32).** Diagonal component of the spectral sensitivity response (defined in Equation (11)) evaluated on $\mathcal{D}_0$. Each column corresponds to a conditional flow matching (CFM) model trained on a different frequency-perturbed version of CIFAR-10 (corresponding to rows in Figure 7). Each visualization shows the average response over 1024 samples, although some of the patterns are often visible at the level of individual samples. We have properly scaled the heatmaps individually for better visibility. **(a)** Unperturbed dataset (reference); **(b)–(f)** correspond to band-stop masking over normalized radial frequency intervals $[0, 0.2]$, $[0.2, 0.4]$, $\ldots$, $[0.8, 1.0]$, respectively. Rows correspond to time steps $t \in \{0.1, 0.3, \ldots, 0.9\}$ ordered from top to bottom. Across columns, the diagonal responses exhibit minimal variation despite substantial differences in training data, indicating that marginal (per-frequency) sensitivity is largely insensitive to the imposed spectral perturbations. A consistent trend across all models is an increasing sensitivity toward higher frequencies as t grows, resulting in a pronounced high-frequency bias at $t = 0.9$. While this aligns with the concentration of variance in higher frequencies for the reference dataset, the persistence of this bias even for the dataset (f), where high frequencies are masked is counterintuitive. We postpone a deeper analysis of this bias to future works. Another pattern that can be seen easily, is that at early times ($t \approx 0$), the sensitivity is approximately uniform across frequencies, consistent with the flat spectrum of white noise. Overall, the similarity across columns suggests that the natural spectral decay of image PSD dominates the effect of the perturbations in the diagonal response. In contrast, Figure 13 shows that off-diagonal sensitivities clearly differentiate between models trained on different perturbations.

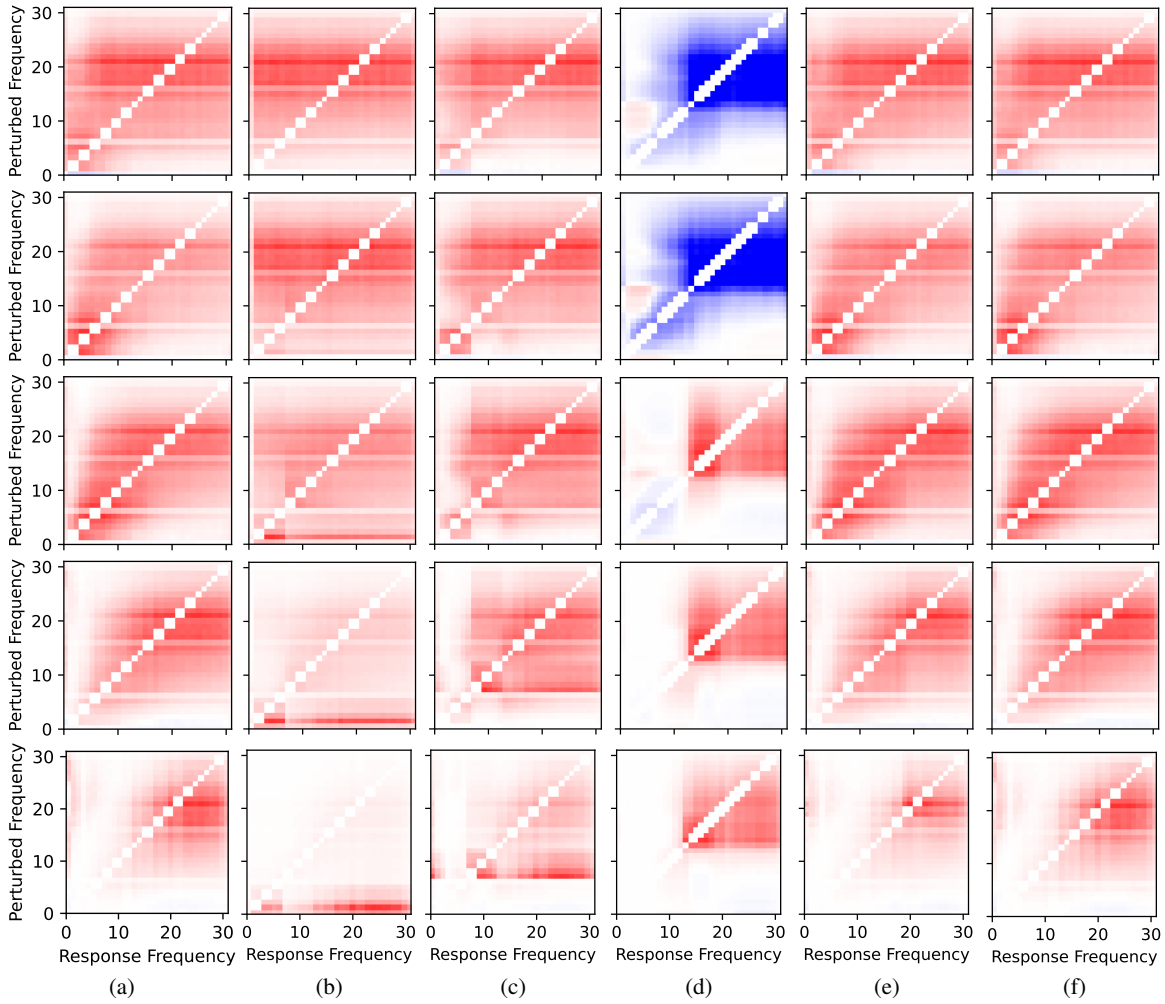

Figure 13. **Off-Diagonal Spectral Sensitivity Reveals Frequency-Dependent Model Differences (CIFAR-10 32x32).** Off-diagonal component of the spectral sensitivity response (defined in Equation (11)) evaluated on $\mathcal{D}_0$. Each column corresponds to a conditional flow matching (CFM) model trained on a different frequency-perturbed version of CIFAR-10 (corresponding to rows in Figure 7). Each visualization shows the average response over 1024 samples, although some of the patterns are often visible at the level of individual samples. We have properly scaled the heatmaps individually for better visibility. **(a)** Unperturbed dataset (reference); **(b)–(f)** correspond to band-stop masking over normalized radial frequency intervals $[0, 0.2]$, $[0.2, 0.4]$, $\ldots$, $[0.8, 1.0]$, respectively. Rows correspond to time steps $t \in \{0.1, 0.3, \ldots, 0.9\}$ ordered from top to bottom. In contrast to the diagonal case (Figure 12), the off-diagonal sensitivity exhibits clear and systematic variation across columns, reflecting the underlying frequency-domain perturbations used during training. The distinction is particularly pronounced at later times ($t = 0.9$), where the sensitivity patterns align with the masked frequency bands, indicating that cross-frequency interactions encode information about the training distribution. A consistent structural trend across all models is a high-frequency bias at larger t, whereby perturbations at a given frequency ω induce stronger responses at frequencies $\omega' > \omega$. Interestingly, column **(f)**, corresponding to masking the highest frequency band $[0.8, 1.0]$, is not clearly reflected in either diagonal or off-diagonal responses. This suggests that combined marginal and interaction-based spectral sensitivities may potentially be used as diagnostic tools for assessing whether specific frequency bands are faithfully modeled. These results demonstrate that off-diagonal spectral sensitivity captures interaction effects that are invisible to marginal (diagonal) analyses, while also exposing potential blind spots in the learned dynamics.

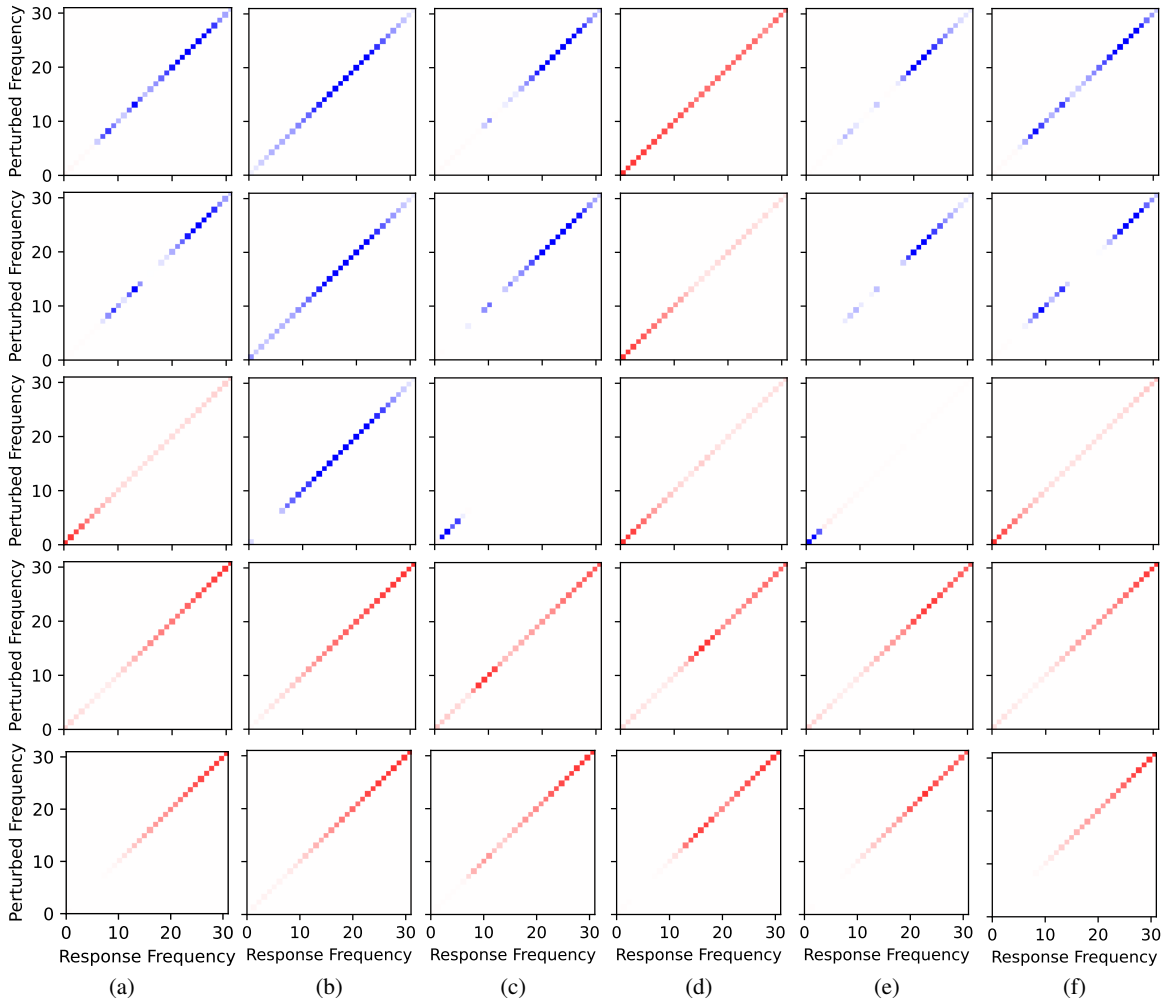

Figure 14. **Diagonal Spectral Sensitivity is Largely Invariant to Frequency-Domain Perturbations (CelebA 64x64).** Diagonal component of the spectral sensitivity response (defined in Equation (11)) evaluated on $\mathcal{D}_0$. Each column corresponds to a denoising diffusion model (DDPM) trained on a different frequency-perturbed version of CelebA (corresponding to rows in Figure 7). Each visualization shows the average response over 1024 samples, although some of the patterns are often visible at the level of individual samples. We have properly scaled the heatmaps individually for better visibility. **(a)** Unperturbed dataset (reference); **(b)–(f)** correspond to band-stop masking over normalized radial frequency intervals $[0, 0.2]$, $[0.2, 0.4]$, $\ldots$, $[0.8, 1.0]$, respectively. Rows correspond to time steps $t \in \{0.1, 0.3, \ldots, 0.9\}$ ordered from top to bottom. Across columns, the diagonal responses exhibit minimal variation despite substantial differences in training data, indicating that marginal (per-frequency) sensitivity is largely insensitive to the imposed spectral perturbations. A consistent trend across all models is an increasing sensitivity toward higher frequencies as t grows, resulting in a pronounced high-frequency bias at $t = 0.9$. While this aligns with the concentration of variance in higher frequencies for the reference dataset, the persistence of this bias even for the dataset (f), where high frequencies are masked is counterintuitive. We postpone a deeper analysis of this bias to future works. Another pattern that can be seen easily, is that at early times ($t \approx 0$), the sensitivity is approximately uniform across frequencies, consistent with the flat spectrum of white noise. Overall, the similarity across columns suggests that the natural spectral decay of image PSD dominates the effect of the perturbations in the diagonal response. In contrast, Figure 15 shows that off-diagonal sensitivities clearly differentiate between models trained on different perturbations.

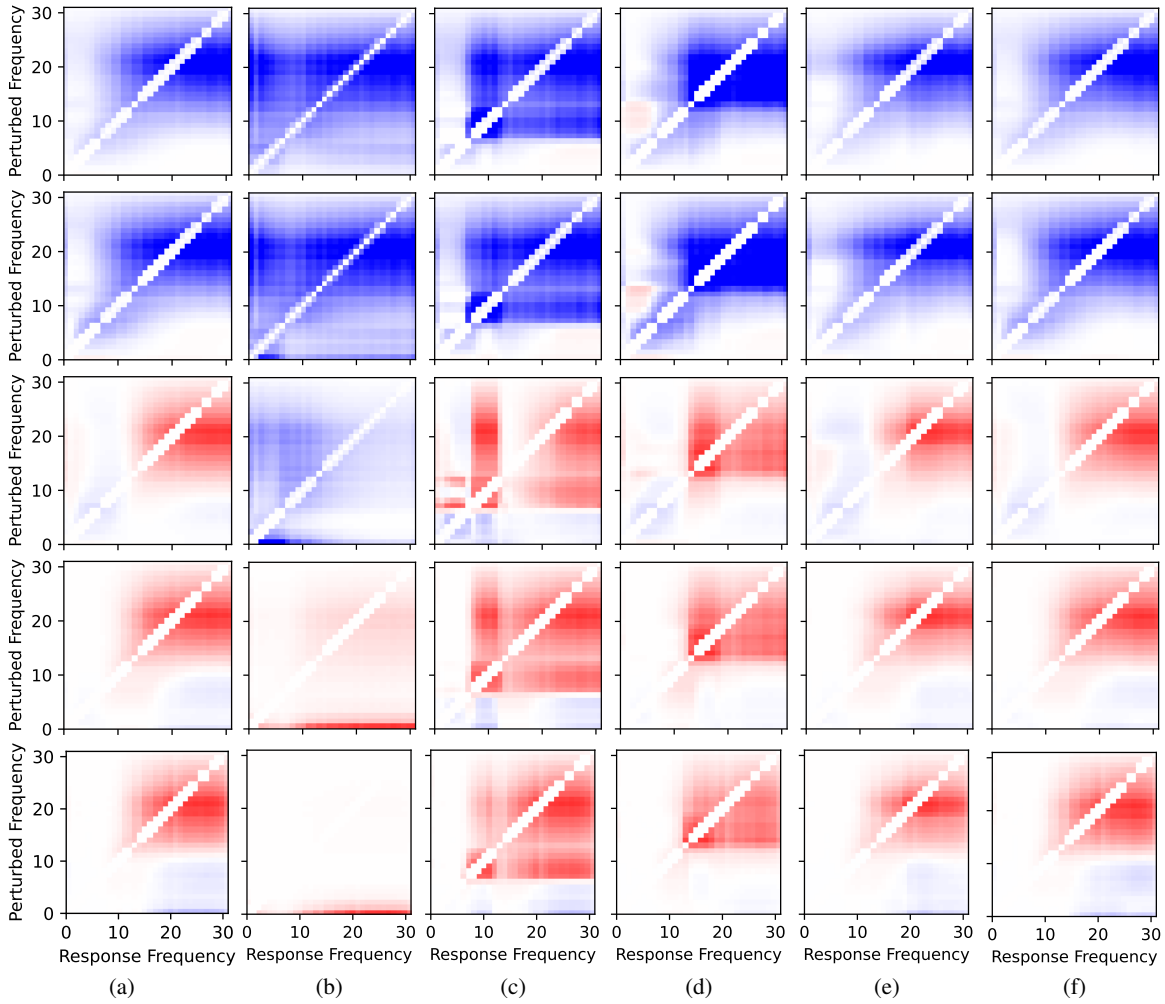

Figure 15. **Off-Diagonal Spectral Sensitivity Reveals Frequency-Dependent Model Differences (CelebA 64x64).** Off-diagonal component of the spectral sensitivity response (defined in Equation (11)) evaluated on $\mathcal{D}_0$. Each column corresponds to a denoising diffusion model (DDPM) trained on a different frequency-perturbed version of CelebA (corresponding to rows in Figure 7). Each visualization shows the average response over 1024 samples, although some of the patterns are often visible at the level of individual samples. We have properly scaled the heatmaps individually for better visibility. **(a)** Unperturbed dataset (reference); **(b)–(f)** correspond to band-stop masking over normalized radial frequency intervals $[0, 0.2]$, $[0.2, 0.4]$, $\ldots$, $[0.8, 1.0]$, respectively. Rows correspond to time steps $t \in \{0.1, 0.3, \ldots, 0.9\}$ ordered from top to bottom. In contrast to the diagonal case (Figure 14), the off-diagonal sensitivity exhibits clear and systematic variation across columns, reflecting the underlying frequency-domain perturbations used during training. The distinction is particularly pronounced at later times ($t = 0.9$), where the sensitivity patterns align with the masked frequency bands, indicating that cross-frequency interactions encode information about the training distribution. A consistent structural trend across all models is a high-frequency bias at larger t, whereby perturbations at a given frequency ω induce stronger responses at frequencies $\omega' > \omega$. Interestingly, column **(f)**, corresponding to masking the highest frequency band $[0.8, 1.0]$, is not clearly reflected in either diagonal or off-diagonal responses. This suggests that combined marginal and interaction-based spectral sensitivities may potentially be used as diagnostic tools for assessing whether specific frequency bands are faithfully modeled. These results demonstrate that off-diagonal spectral sensitivity captures interaction effects that are invisible to marginal (diagonal) analyses, while also exposing potential blind spots in the learned dynamics.