CYCLOSTATIONARY PHASE CONDITIONING FOR MEDICAL TIME SERIES DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Many physiological time series, such as cardiac and brain recordings, exhibit cyclostationarity: their statistics vary periodically with an underlying cycle phase. Corruption from motion, poor contact, and physiological interference obscures morphology needed for diagnosis, making signal restoration essential. Existing diffusion approaches condition on corrupted observations alone and must learn cyclic structure implicitly. We instead propose two inductive biases which encode cyclostationarity: a shift-covariant wavelet representation and dense per-sample phase conditioning inferred from the corrupted input. We further introduce a training-free cyclostationarity index that quantifies phase structure and predicts when phase conditioning will help. Finally, we propose antithetic coupling of reverse trajectories to reduce sampling variance while achieving comparable performance with fivefold fewer network evaluations. Across modalities, our results show that explicitly encoding measurable cyclic structure improves physiological time-series restoration.

1 INTRODUCTION

Recurrent structure is a common property of signals arising from natural processes. Physiological time series are a prominent example, reflecting the cyclic biological processes that generate them. A heartbeat appears in the electrocardiogram (ECG) and, moments later, as a pulse of blood volume at the wrist in photoplethysmography (PPG) (Xie et al., 2020; Nitzan & Ovadia-Blechman, 2022); a muscle contraction appears as a burst of electrical activity in electromyography (EMG) (Khorrami Chokami et al., 2021); and the brain's electrical activity rises and falls at characteristic frequencies in the electroencephalogram (EEG) (Brenner & Schaul, 1990). These signals support cardiac diagnosis, wearable monitoring, prosthetic control and brain–computer interfaces, yet are routinely corrupted by movement, poor electrode contact, and interference from other physiological sources (Chatterjee et al., 2020; Ismail et al., 2021; Boyer et al., 2023; Jiang et al., 2019). Restoration methods aim to remove these corruptions and recover the underlying clean signal.

These time series are not stationary: their mean and covariance change over time, and do so approximately periodically as a function of a latent cycle phase. This property is known as wide-sense *cyclostationarity* (Gardner, 2018; Napolitano & Gardner, 2025). Despite their recurring structure, existing restoration methods do not model this explicitly. This limitation is especially important for diffusion-based approaches, which have become the state of the art for recovering signals from corrupted observations. Diffusion models have been adapted across physiological modalities (Li et al., 2024; Liu et al., 2024; Huang et al., 2025; Ruipérez-Campillo et al., 2025; 2026a) and to spectral representations (Crabbé et al., 2024; Li et al., 2026; Qin et al., 2026), yet all condition the reverse process solely on the corrupted observation. Therefore, the denoiser must learn from data that samples separated by one cycle correspond to the same position within the underlying cycle.

A natural question arises: can modeling cycle structure improve restoration? To answer this, we introduce PTSDIFF, to include direct cycle-phase information into the diffusion process, and investigate whether this benefit can be predicted before training. We encode this structure through two complementary inductive biases. First, instead of operating directly in the time domain, we perform diffusion in a shift-covariant wavelet representation, making the model insensitive to the arbitrary starting position of the input window, which we hypothesize should provide a consistent benefit across modalities. Second, we condition the denoiser on a phase field estimated from the corrupted input, a dense per-sample representation of position within the current cycle, which should be most useful when the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

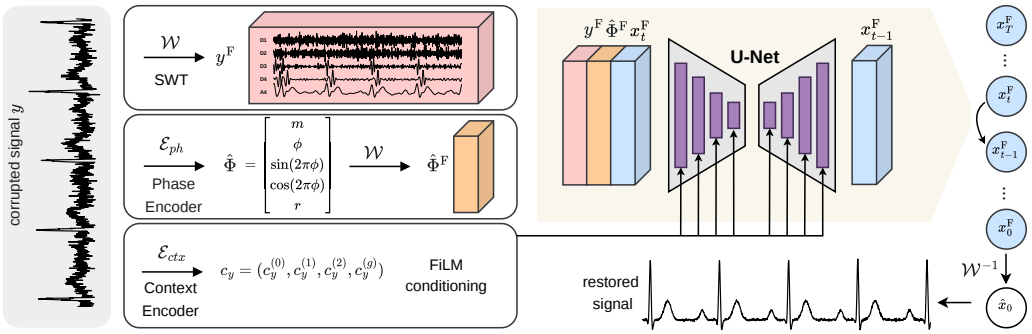


Figure 1: Overview of PTSDIFF. The corrupted observation y feeds the context encoder and the phase encoder; it is frame-transformed (top) and concatenated with the frame-transformed phase channels (middle) and the latent x_t^{F} as U-Net input. The context encoder (bottom) modulates the backbone through FiLM embeddings, and the U-Net drives the update to x_{t-1}^{F}. Block-level detail is in Figure 5.

signal is strongly structured by its cycle. Together, the two components remove sensitivity to absolute window position while providing explicit information about position within the underlying cycle.

To understand the effect of phase conditioning we analyze it along two axes: how phase information is estimated and how strongly the underlying signal is structured by its cycle. We find that estimation quality is critical: a classical detector applied to the corrupted input performs no better than omitting phase conditioning, whereas a learned estimator substantially improves restoration. Moreover, we study whether the usefulness of phase conditioning can be anticipated prior to training. To this end, we introduce a cyclostationarity index that measures the fraction of signal variance explained by cycle phase. This provides a training-free estimate of which signals benefit from phase conditioning, avoiding unnecessary model complexity and training when cycle structure is weak.

Beyond improving the conditional model with both inductive biases, diffusion restoration remains stochastic at inference: different reverse trajectories can yield different estimates, while averaging several trajectories improves the estimate but is computationally expensive. We therefore study antithetic coupling (Hammersley & Morton, 1956; Ruipérez-Campillo et al., 2026a) as an inference-time variance-reduction strategy and show that one antithetic pair can match or outperform ten independent trajectories, with fivefold fewer network evaluations. Overall, our contributions are summarized as:

- We propose PTSDIFF, a novel formulation of physiological signal restoration under a cyclostationary prior. We address a limitation of existing diffusion-based methods that overlook this structure, through two inductive biases: shift-covariant wavelet diffusion and cycle-phase conditioning.
- We introduce a cyclostationarity index that quantifies cycle structure from clean data and gives a training-free indication of when phase conditioning is likely to help.
- We achieve state-of-the-art restoration and improve downstream task performance across studied benchmarks in four modalities, with a sampling-efficient solution incorporating antithetic coupling.

2 Related work

Restoration for physiological signals. Medical recordings are noisy. Artifacts obscure the details that clinical interpretation relies on, and corrupted segments are otherwise discarded, from intensive care to ambulatory and consumer recordings (Chatterjee et al., 2020; Ismail et al., 2021). Classical restoration pipelines built on Kalman filtering, empirical mode decomposition, adaptive filtering and wavelet shrinkage remain standard (Sayadi & Shamsollahi, 2008; Blanco-Velasco et al., 2008; Gao et al., 2010; Lee & Zhang, 2003; Ismail et al., 2021; Jiang et al., 2019; Boyer et al., 2023). However, these methods assume fixed signal and noise spectra (Ruipérez-Campillo et al., 2024; 2026b), and they degrade under non-stationary artifacts and suppress the low-amplitude morphology that carries diagnostic content (Chatterjee et al., 2020; Watanabe et al., 2025). Learned deterministic restorers, primarily convolutional, recurrent and attention-based autoencoders, outperform them (Chiang et al., 2019; Antczak, 2018; Chen et al., 2024; Lai et al., 2025; Cui et al., 2024; Wang et al., 2023), but learn a single deterministic map and model the conditional mean implicitly, with no control over estimator variance. Generative restoration (Ho et al., 2020; Song et al., 2021b; Saharia et al., 2022) addresses this for different time-series modalities, i.e. DeScoD-ECG (Li et al., 2024), SDEMG (Liu et al., 2024)

and EEGDfus (Huang et al., 2025). Moreover, general-purpose restoration samplers (Kawar et al., 2022; Chung et al., 2023) focus on the inverse problem, while accelerated solvers (Song et al., 2021a; Lu et al., 2022) reduce the number of sampling steps; neither directly targets estimator variance at fixed Number of Function Evaluations (NFE). Further per-modality details are provided in Appendix A.

Transformed-domain diffusion. Diffusion-based time-series models typically operate in the native time domain, but recent work has explored whether alternative representations provide more suitable inductive biases. Crabbé et al. (2024) argue that diffusing in a spectral basis suits signals with localized spectral structure, and FIDE (Galib et al., 2024) prevents high-frequency components from vanishing in the reverse process. Global bases sacrifice temporal localization, motivating time–frequency representations (Li, 2022): MECG-E uses the STFT (Hung et al., 2024) and TFCDiff the DCT (Li et al., 2026). Common to these, and made explicit here, is that the choice of frame, i.e., the signal representation in which diffusion operates, determines which symmetries are built into the denoiser. In this work, we adopt the Stationary Wavelet Transform (SWT) for its shift-covariance rather than its spectral properties (Nason & Silverman, 1995; Coifman & Donoho, 1995).

Periodicity and variance reduction. Periodic structure is routinely exploited in forecasting architectures (Wu et al., 2023; 2021). In physiological modeling, PulseImpute (Xu et al., 2022) and PulseDiff (Jenkins et al., 2023) exploit cardiac repetition for imputation, while latent-trajectory models represent cycles geometrically (Ryser et al., 2022). Furthermore, oscillatory and burst structure is documented for EEG and EMG (Brenner & Schaul, 1990; Khorrami Chokami et al., 2021). These works establish that periodicity can be useful, but not how much a given signal should benefit from it, which is the gap addressed in Section 3.4. Antithetic variates are a variance-reduction technique (Hammersley & Morton, 1956; Owen, 2013; Choi & Park, 2026) and were introduced into conditional diffusion sampling in physiological time series by Ruipérez-Campillo et al. (2026a). We adopt this coupling, and derive its effect at matched NFE in the restoration setting.

3 METHOD: PTSDIFF

We propose PTSDIFF[1], a method to incorporate the cyclostationary structure (Section 3.1) into diffusion through a shift-covariant representation and explicit cycle-phase conditioning. We then define the training objective, a training-free predictor of phase utility, and an antithetic inference estimator.

3.1 PRELIMINARIES

Cyclostationary restoration. Let $x_0, y \in \mathbb{R}^L$ denote a clean signal and its corrupted observation, respectively, each of length L and with sample indices $n, n' \in \{1, \dots, L\}$; restoration seeks to recover x_0 given y. We assume x_0 is drawn from a process that is cyclostationary in the *wide sense*: there exists a latent phase $\varphi_n \in [0, 1)$, advancing at a locally varying rate, with

$$\mathbb{E}[x_0[n] \,|\, \varphi_n = \varphi] = \mu(\varphi), \qquad \mathrm{Cov}(x_0[n], x_0[n'] \,|\, \varphi_n, \varphi_{n'}) = C(\varphi_n, \varphi_{n'}), \tag{1}$$

so first- and second-order statistics depend on phase, not on absolute time. This is weaker than periodicity, since successive cycles differ in length and morphology, and stronger than stationarity. Different physiological systems may satisfy this to different degrees.

Conditional diffusion. We use an ϵ-parameterized DDPM (Ho et al., 2020). For a generic representation u_0 with variance schedule $\{\beta_t\}_{t=1}^T$, $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$. The forward process is

$$u_t = \sqrt{\bar{\alpha}_t}\, u_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{2}$$

A denoising network ϵ_θ is trained to predict the injected noise ϵ at each diffusion step. At inference, the reverse process generates samples starting from noise $u_T \sim \mathcal{N}(0, I)$. Sections 3.2–3.3 specify the representation u_0 and the conditioning supplied to the denoiser.

3.2 DIFFUSION IN A SHIFT-COVARIANT FRAME

The first inductive bias removes sensitivity to the arbitrary position at which a signal window begins. Under the cyclostationary model, absolute position is not informative; what matters is position

[1]The code is publicly available in: `https://anonymous.4open.science/r/PTSDiff/`.

relative to the underlying cycle. We therefore perform diffusion in a shift-covariant representation that preserves temporal alignment.

Shift-covariant frame. We instantiate the shift-covariant representation with the stationary wavelet transform (SWT) (Nason & Silverman, 1995). Let $\mathcal{W}$ denote its analysis operator and $\mathcal{W}^{-1}$ its inverse, with $x_0^{\mathrm{F}} = \mathcal{W}(x_0)$ and $y^{\mathrm{F}} = \mathcal{W}(y)$ denoting the corresponding frame coefficients of the clean signal x_0 and its corrupted observation y, respectively. For an input a_0, the level-j decomposition is

$$a_j[n] = \sum_k h_j[k]\, a_{j-1}[n-k], \qquad d_j[n] = \sum_k g_j[k]\, a_{j-1}[n-k], \tag{3}$$

with h_j and g_j the upsampled low- and high-pass filters at level j. Unlike the standard discrete wavelet transform (DWT), the SWT omits *decimation*, i.e. the downsampling applied after each level. As a result, $\mathcal{W} : \mathbb{R}^L \to \mathbb{R}^{(J+1)\times L}$ preserves temporal resolution and is exactly shift-covariant, or equivalently translation-equivariant. Writing $(S_\tau x)[n] = x[n-\tau]$ for the cyclic shift by τ samples,

$$\mathcal{W}(S_\tau x) = S_\tau\, \mathcal{W}(x) \qquad \forall \tau \in \mathbb{Z}, \tag{4}$$

where S_τ acts identically on every channel. The decimated DWT does not satisfy this property for arbitrary shifts: downsampling by 2^j at level j restricts equivariance at that level to shifts that are multiples of 2^j, and therefore to multiples of 2^J for the full J-level transform. This loss of shift covariance motivates the undecimated construction used in translation-invariant wavelet denoising (Nason & Silverman, 1995; Coifman & Donoho, 1995). Because $\mathcal{W}$ maps L samples to $(J+1)L$ coefficients, it is redundant rather than a basis. More precisely, it forms an *overcomplete frame*: a redundant representation that nevertheless admits stable reconstruction. By definition (Daubechies, 1992), there exist constants $0 < A \le B < \infty$ such that

$$A\|x\|_2^2 \;\le\; \|\mathcal{W}x\|_2^2 \;\le\; B\|x\|_2^2. \tag{5}$$

For the orthogonal wavelet families used here, the undecimated transform is a tight frame with $A = B = J+1$, so $\mathcal{W}^{-1} = A^{-1}\mathcal{W}^\top$ and reconstruction is non-expansive up to the factor $A^{-1/2}$. Two consequences are useful for our model. First, the subbands remain aligned on the original sample grid, allowing them to be stacked as channels of a length-preserving backbone. Second, the time-domain phase field introduced in Section 3.3 can be transformed by $\mathcal{W}$ and concatenated channel-wise without resampling. We use $J = 4$ levels, yielding detail channels $(d_1, \ldots, d_4)$ and one approximation channel a_4. For sampling rate f_s, the channel d_j corresponds approximately to $[f_s 2^{-(j+1)}, f_s 2^{-j}]$, while a_J contains the residual frequencies below $f_s 2^{-(J+1)}$. The resulting per-modality band mappings and the choice of wavelet family are reported in Appendix C.2; the choice of J is evaluated in Appendix D.9.

Diffusion in the frame domain. We apply the process of Section 3.1 to $x_0^{\mathrm{F}} = \mathcal{W}(x_0)$ rather than x_0, so the latent is $x_t^{\mathrm{F}} \in \mathbb{R}^{(J+1)\times L}$ and the time-domain output is $\hat{x}_0 = \mathcal{W}^{-1}(\hat{x}_0^{\mathrm{F}})$. We use $T = 50$ and the quadratic noise schedule of Li et al. (2026).

3.3 Phase conditioning

The second inductive bias makes position relative to the underlying physiological cycle explicit to the denoiser. Let $\mathcal{T}(x) = \{e_1 < \cdots < e_M\}$ denote the detected cycle-event indices of a signal (e.g., R-peaks[2] in an ECG). For a sample n satisfying $e_k \le n < e_{k+1}$, we define its cycle phase, local cycle rate, and soft event mask as

$$\phi_n = \frac{n - e_k}{e_{k+1} - e_k} \in [0, 1), \qquad r_n = \frac{f_s}{e_{k+1} - e_k}, \qquad m_n = \max_j \exp\!\left(-\frac{(n - e_j)^2}{2s^2}\right), \tag{6}$$

where f_s denotes the sampling rate and s the event width. These quantities form the dense phase field

$$\Phi = \big(m,\ \phi,\ \sin 2\pi\phi,\ \cos 2\pi\phi,\ r\big) \in \mathbb{R}^{5\times L}. \tag{7}$$

Here, m localizes cycle events, ϕ gives the relative position within each cycle, and r captures local variation in cycle rate. The sine and cosine channels embed phase on the circle, avoiding the discontinuity between $\phi = 1^-$ and $\phi = 0$, which would otherwise force the network to learn that $\phi = 0.99$ and $\phi = 0.01$ are adjacent.

[2]The R peak is the prominent spike in an ECG and serves as a reference point for heartbeat timing.

Phase encoder. At inference, the clean signal x_0 is unavailable, so its phase field must be estimated from the corrupted observation y. We use a learned, shift-equivariant convolutional phase encoder $\mathcal{E}_{ph}$ and set $\hat{\Phi} = \mathcal{E}_{ph}(y)$. As illustrated in Figure 1, the predicted field is transformed into the frame domain, applying $\mathcal{W}$ independently to each channel, as $\hat{\Phi}^{\mathrm{F}} = \mathcal{W}(\hat{\Phi})$ and concatenated channel-wise with x_t^{F} and y^{F}, providing the denoiser with dense per-sample phase information. Since both $\mathcal{E}_{ph}$ and $\mathcal{W}$ are shift-equivariant, it holds that $\mathcal{W}\mathcal{E}_{ph}(S_\tau y) = S_\tau \mathcal{W}\mathcal{E}_{ph}(y)$. Thus, the dense phase conditioning shifts consistently with the input. The phase encoder $\mathcal{E}_{ph}$ can be trained through the restoration objective alone. For ECG, where a reliable cycle-event detector is available, we additionally warm-start it using a target phase field $\Phi^\star = \mathcal{D}(x_0)$ computed from the clean signal. Here, $\mathcal{D}$ denotes the analytical construction of Equations 6–7. The analytical detector is used only for initialization and not at inference. For the remaining modalities, $\mathcal{E}_{ph}$ is trained without warm-starting. Full warm-start details are given in Appendix C.2.

Context encoder. In parallel, a context encoder $\mathcal{E}_{\mathrm{ctx}}$ maps y to pooled, shift-invariant context embeddings $c_y = (c_y^{(0)}, c_y^{(1)}, c_y^{(2)}, c_y^{(g)})$, with one embedding for each U-Net resolution and one global embedding. The resolution-specific embeddings modulate the corresponding U-Net features through FiLM (Perez et al., 2018), while $c_y^{(g)}$ is combined with the diffusion-step embedding. Unlike the phase field, this pathway provides global information about the corrupted observation rather than sample-wise conditioning. Because the phase conditioning is shift-equivariant and the pooled context conditioning is shift-invariant, neither introduces dependence on absolute window position. Architectural details are given in Appendix C.2.

3.4 Measuring cycle structure

The shift-covariant representation and phase conditioning both exploit the cyclostationary structure of Equation 1. Their usefulness should therefore depend on how strongly cycle phase structures the signal. We want to quantify this dependence from clean data before committing to model training.

Definition 1 (Cyclostationarity index)**.** *For a process with latent phase φ,*

$$\Pi \;=\; \frac{\mathrm{Var}_\varphi\big(\mathbb{E}[x \mid \varphi]\big)}{\mathrm{Var}(x)} \;\in\; [0,1] \tag{8}$$

is the fraction of signal variance explained by cycle phase.

Π is estimated from clean data by phase-aligned averaging: detect cycle events, resample each cycle to a common phase grid, average to obtain $\hat{\mu}(\varphi)$, and take the ratio of its variance to the total. The estimator, its finite-sample bias and a cheaper autocorrelation proxy are given in Appendix B.3. Where no reliable cycle event detector exists, we use the autocorrelation proxy instead.

Proposition 1 (Phase-explained variance)**.** *Let x be cyclostationary in the sense of Equation 1. Then*

$$\mathrm{Var}(x) - \mathbb{E}_\varphi\big[\,\mathrm{Var}(x \mid \varphi)\big] \;=\; \mathrm{Var}_\varphi\big(\mathbb{E}[x \mid \varphi]\big) \;=\; \Pi \cdot \mathrm{Var}(x), \tag{9}$$

so in the absence of an observation, conditioning on phase reduces the minimum mean-squared error exactly by $\Pi\,\mathrm{Var}(x)$.

Proof in Appendix B.1. The identity establishes that the variance phase explains is computable from clean data before any model exists. It does not extend to the case where y is observed (Appendix B.1), so we treat Π as a measure of how strongly phase structures a signal rather than as a certificate.

3.5 Learning objective

The denoiser predicts the injected noise as $\hat{\epsilon}_t = \epsilon_\theta(x_t^{\mathrm{F}}, y^{\mathrm{F}}, \hat{\Phi}^{\mathrm{F}}, c_y, t)$ and is trained with the standard noise-prediction loss (Ho et al., 2020) on frame coefficients:

$$\mathcal{L}_{\mathrm{diff}} = \mathbb{E}_{(x_0,y),\,t,\,\epsilon}\left[\|\epsilon - \hat{\epsilon}_t\|_2^2\right], \qquad t \sim \mathcal{U}\{1,\dots,T\}. \tag{10}$$

However, low noise-prediction error does not by itself control time-domain reconstruction error, and the relation between the two depends on t. Substituting Equation 2 into the clean estimate $\hat{x}_0^{\mathrm{F}} = (x_t^{\mathrm{F}} - \sqrt{1-\bar{\alpha}_t}\,\hat{\epsilon}_t)/\sqrt{\bar{\alpha}_t}$ yields

$$\hat{x}_0^{\mathrm{F}} - x_0^{\mathrm{F}} = \sqrt{\tfrac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\,(\epsilon - \hat{\epsilon}_t), \qquad \big\|\hat{x}_0 - x_0\big\|_2 \;\le\; A^{-1/2}\sqrt{\tfrac{1-\bar{\alpha}_t}{\bar{\alpha}_t}}\,\big\|\epsilon - \hat{\epsilon}_t\big\|_2, \tag{11}$$

where the bound uses $\hat{x}_0 - x_0 = \mathcal{W}^{-1}(\hat{x}_0^{\mathrm{F}} - x_0^{\mathrm{F}})$ together with $\|\mathcal{W}^{-1}\|_2 = A^{-1/2}$, which holds because $\mathcal{W}$ is tight with $A = B = J + 1$ (Equation 5). The first identity in Equation 11 states that $\mathcal{L}_{\mathrm{diff}}$ is a signal-space error weighted by the signal-to-noise ratio $\bar{\alpha}_t/(1-\bar{\alpha}_t)$: uniform weight in noise space is sharply decaying weight in signal space. Under the fixed quadratic schedule of Li et al. (2026), the amplification factor $\sqrt{(1-\bar{\alpha}_t)/\bar{\alpha}_t}$ spans four orders of magnitude. Thus, sampling t uniformly does not weight reconstruction error uniformly: $\mathcal{L}_{\mathrm{diff}}$ is least sensitive to coefficient error at exactly the steps with large t, where that error is amplified most. We counteract this with two terms applied to $\hat{x}_0 = \mathcal{W}^{-1}(\hat{x}_0^{\mathrm{F}})$ at the same sampled t:

$$\mathcal{L}_{\mathrm{rec}} = \mathbb{E}\|\hat{x}_0 - x_0\|_1,\ \mathcal{L}_{\mathrm{morph}} = \mathbb{E}\,\bar{\alpha}_t\|\nabla\hat{x}_0 - \nabla x_0\|_1,\ \mathcal{L} = \mathcal{L}_{\mathrm{diff}} + \lambda_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{morph}}\mathcal{L}_{\mathrm{morph}}, \quad (12)$$

with ∇ the first-order temporal difference, which penalizes slope error and encourages sharp transients; $\bar{\alpha}_t$ restricts it to low noise, where slopes are informative. We use ℓ_1 rather than ℓ_2 for robustness to the large residuals that $\hat{x}_0$ exhibits at high t. We set $\lambda_{\mathrm{rec}} = 0.3$ and $\lambda_{\mathrm{morph}} = 0.1$; Appendix D.9 reports a sweep over both weights.

3.6 Antithetic reverse estimation

The preceding components shape $p_\theta(x_0 \mid y)$; we now fix it and change only how it is sampled. We target the model's conditional mean $\bar{\mu}_\theta(y) = \mathbb{E}_{p_\theta}[x_0 \mid y]$, the squared-error-optimal estimate under p_θ, and compare two estimators with the same budget of KT network-function evaluations (NFEs). MC-K averages K independent trajectories, whereas AV-K averages $K/2$ antithetic pairs (K even):

$$\hat{x}_{0,\text{MC-}K} = \frac{1}{K}\sum_{k=1}^{K}\hat{x}_0^{(k)}, \qquad \hat{x}_{0,\text{AV-}K} = \frac{1}{K}\sum_{j=1}^{K/2}\big(\hat{x}_{0,+}^{(j)} + \hat{x}_{0,-}^{(j)}\big). \quad (13)$$

Both estimators are unbiased for $\bar{\mu}_\theta(y)$ and differ only in variance. Each antithetic pair starts from $x_{T,+}^{\mathrm{F}} \sim \mathcal{N}(0, I)$ and $x_{T,-}^{\mathrm{F}} = -x_{T,+}^{\mathrm{F}}$. Since $\mathcal{N}(0, I)$ is symmetric, its two members can receive opposite noise at every reverse step without changing their marginal $p_\theta(x_0 \mid y)$:

$$x_{t-1,\pm}^{\mathrm{F}} = \mu_\theta\Big(x_{t,\pm}^{\mathrm{F}}, y^{\mathrm{F}}, \hat{\Phi}^{\mathrm{F}}, c_y, t\Big) \pm \sigma_t z_t, \qquad z_t \sim \mathcal{N}(0, I), \quad \sigma_t = \sqrt{\tilde{\beta}_t}. \quad (14)$$

Proposition 2 (Error decomposition at matched NFE)**.** *Fix y and write $\bar{\mu}_\theta(y) = \mathbb{E}_{p_\theta}[x_0 \mid y]$. Let $\sigma^2(\tau)$ be the per-trajectory variance of the reconstruction at sample index τ and $\rho(\tau)$ the correlation between the two members of an antithetic pair. Then the K-trajectory estimator satisfies*

$$\mathrm{Var}[\hat{x}_{0,\text{AV-}K}(\tau)] = \frac{\sigma^2(\tau)}{K}\big(1 + \rho(\tau)\big), \quad (15)$$

and its expected squared error decomposes as

$$\mathbb{E}\big\|\hat{x}_{0,K} - x_0\big\|_2^2 = \underbrace{\big\|\bar{\mu}_\theta(y) - x_0\big\|_2^2}_{\textit{model bias}} + \frac{1+\bar{\rho}}{K}\underbrace{\textstyle\sum_\tau \sigma^2(\tau)}_{\textit{trajectory variance}}, \qquad \bar{\rho} = \frac{\sum_\tau \rho(\tau)\sigma^2(\tau)}{\sum_\tau \sigma^2(\tau)}, \quad (16)$$

with $\bar{\rho} = 0$ for independent trajectories. Coupling reduces the compute-controllable term by the factor $1 + \bar{\rho}$ at equal NFE whenever $\bar{\rho} < 0$, and is no worse than independent sampling to $O(\bar{\rho})$ when $\bar{\rho} \approx 0$.

Proof in Appendix B.2. At matched NFE, antithetic coupling reduces sampling variance when $\bar{\rho} < 0$; Appendix D.11 evaluates this condition and the quality–compute trade-off.

4 Experimental Setup

Physiological signals and datasets. We evaluate datasets of four physiological modalities with different structures. First, for PPG we use PPG-DaLiA (Reiss et al., 2019), wrist blood-volume pulse (BVP) at 64 Hz corrupted with Noise Stress Test Database (NSTDB)(Moody et al., 1992; Goldberger et al., 2000) motion and baseline artifacts. Second, ECG uses PTB-XL (Wagner et al., 2022; 2020) for training and in-domain test, with MIT-BIH Arrhythmia (Moody & Mark, 2001) and QTDB (Laguna et al., 1997) held out to evaluate distribution shift, corrupted from the MIT-BIH NSTDB. Next, EMG uses NinaPro DB2 (Atzori et al., 2014) contaminated with ECG interference from

	Model	ΔSNR ↑	PRD [%] ↓	CC ↑
PPG	DDAE	8.97 ± 0.14	36.25 ± 0.38	0.91 ± 0.00
	PTSDIFF-MC-10	$\underline{12.99 \pm 0.11}$	$\underline{24.64 \pm 0.30}$	$\underline{0.96 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{13.23 \pm 0.10}$	$\mathbf{24.03 \pm 0.27}$	$\mathbf{0.96 \pm 0.00}$
ECG	TCDAE	12.49 ± 0.65	33.24 ± 2.08	0.95 ± 0.00
	TFCDiff-10	13.35 ± 0.13	30.27 ± 0.45	0.94 ± 0.00
	DeScoD-10	13.85 ± 0.04	28.76 ± 0.15	0.95 ± 0.00
	PTSDIFF-MC-10	$\underline{15.58 \pm 0.12}$	$\underline{23.81 \pm 0.27}$	$\underline{0.96 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{15.87 \pm 0.12}$	$\mathbf{23.08 \pm 0.25}$	$\mathbf{0.97 \pm 0.00}$
EMG	FCN	19.28 ± 0.20	26.51 ± 0.53	0.96 ± 0.00
	SDEMG	21.04 ± 0.30	21.91 ± 0.75	0.97 ± 0.00
	PTSDIFF-MC-10	$\underline{22.41 \pm 0.20}$	$\underline{18.80 \pm 0.57}$	$\underline{0.98 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{22.70 \pm 0.19}$	$\mathbf{18.23 \pm 0.55}$	$\mathbf{0.98 \pm 0.00}$
EEG	EEGIFNet	15.05 ± 0.04	$\mathbf{19.96 \pm 0.10}$	$\mathbf{0.97 \pm 0.00}$
	EEGDfus	15.03 ± 0.21	22.18 ± 0.79	0.96 ± 0.00
	PTSDIFF-MC-10	$\underline{15.31 \pm 0.07}$	20.84 ± 0.19	0.97 ± 0.00
	PTSDIFF-AV-10	$\mathbf{15.39 \pm 0.07}$	$\underline{20.72 \pm 0.19}$	$\underline{0.97 \pm 0.00}$

Table 1: Restoration across four modalities against each modality's own baselines (3 training seeds). Values are mean ± standard deviation. **Bold** marks best, underline second best, within each modality. Modalities are ordered by phase-shuffle margin.

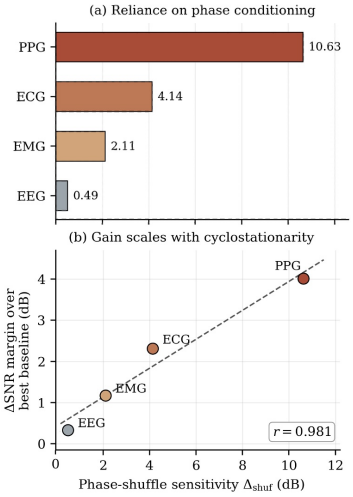


Figure 2: **(a)** ΔSNR drop under phase shuffling. **(b)** Phase-shuffle sensitivity versus the margin over the best baseline.

Normal Sinus Rhythm Database (NSRDB) following the SDEMG protocol (Liu et al., 2024). Finally, EEG uses EEGdenoiseNet (Zhang et al., 2021) in the EEG–EOG setting following the EEGDfus protocol (Huang et al., 2025). To test distribution shift, the PPG, EMG and EEG models are also evaluated on one held-out dataset each: pulse-oximeter PPG from critically ill patients in BIDMC (Pimentel et al., 2017), sEMG from trans-radial amputees in NinaPro DB3 (Atzori et al., 2014), and EEG from the PhysioNet Motor Movement/Imagery Database (EEGMMIDB) (Schalk et al., 2004), each corrupted with the test protocol of its training dataset. Preprocessing follows each benchmark's established protocol; details in Appendix C.1.

Baselines. Each modality is compared against its strongest published restoration methods, retrained on identical splits and preprocessing: DDAE (Lai et al., 2025) for PPG; TCDAE (Chen et al., 2024), DeScoD-ECG (Li et al., 2024) and TFCDiff (Li et al., 2026) for ECG; SDEMG (Liu et al., 2024) and FCN (Wang et al., 2023) for EMG; EEGDfus (Huang et al., 2025) and EEGIFNet (Cui et al., 2024) for EEG. Other common baselines (FIR, SWT shrinkage) are reported in Appendix D.1. Stochastic baselines use the sampling procedure specified in their own publications. Every baseline figure was produced under our protocol. Appendix C.3 provides implementation and adaptation details.

Metrics and protocol. We report ΔSNR (dB), percentage root-mean-square difference (PRD, %) and Pearson correlation (CC) to evaluate restoration performance, computed per signal and averaged, with 95% bootstrap confidence intervals (CI) over 1,000 resamples. All results use a single training seed, except Table 1, which is trained and evaluated across three seeds and reported as mean ± standard deviation (exact seeds in Appendix C.2). Benchmark-specific metrics are given in Appendix C.4. Paired comparisons against each baseline use one-sided Wilcoxon signed-rank tests with Holm–Bonferroni correction (Appendix C.5). Finally, sampling variants are labeled MC-K for Monte Carlo averaging over K independent trajectories and AV-K for $K/2$ antithetic pairs; both use K reverse trajectories and hence KT NFEs. Additional experimental details are included in Appendix C.

5 RESULTS

We evaluate PTSDIFF across four physiological modalities and answer: *(i) Restoration,* does it improve the quality and preserve downstream utility? *(ii) Phase conditioning,* when does phase help, and can we predict its benefit? *(iii) Components,* which parts drive the gains? *(iv) Efficiency,* can antithetic sampling improve the inference quality–compute trade-off? Further results in Appendix D.

Restoration performance. Table 1 reports successful restoration performance for all four modalities against modality-specific baselines, and Figures 14–16 (Appendix D.14) show representative

reconstructions. PTSDIFF-AV-10 achieves the best $\Delta\mathrm{SNR}$ on all four modalities and the best PRD and CC on three. On EEG, EEGIFNet retains the best PRD and CC: these metrics are averaged on a linear rather than a decibel scale, so poorly restored segments weigh more heavily, and EEGIFNet's quality is more uniform across segments. On ECG, PTSDIFF matches or exceeds every baseline, classical filters included, on the metrics of prior work (Table 7); all paired comparisons are significant after Holm–Bonferroni correction (Table 12), and the improvement shifts the whole per-sample distribution (Figure 12). Its advantage there widens at higher noise levels (Tables 8–10) and holds on held-out datasets without retraining (Figure 8), highlighting strong generalization capabilities. On MIT-BIH, which is rich in arrhythmias, the other diffusion models lose their edge over the deterministic TCDAE while PTSDIFF keeps a clear lead, consistent with a phase field defined beat by beat rather than by a fixed rhythm. Figures 6 and 7 stratify the other modalities by noise level, with the main metrics and with each benchmark's own (Table 6), respectively. Under dataset shift without retraining, PTSDIFF keeps its lead over the strongest baseline on PPG, ECG and EMG but falls behind EEGIFNet on EEG, the modality where its in-domain margin is also smallest (Table 11, Appendix D.3).

Table 2: Downstream utility across modalities. Extended results in Table 16.

	Task	Baseline	PTSDIFF
PPG	HR MAE ↓	5.947	**5.155**
ECG	PTB-XL AUROC ↑	0.840	**0.848**
ECG	MIT-BIH F1 ↑	0.543	**0.630**
EMG	Activation bal. acc. ↑	0.749	**0.750**
EEG	Band-power error ↓	4.56	**3.31**

The improvements also carry over downstream: PTSDIFF matches or exceeds the strongest restoration baseline on every task across the four modalities (Table 2), and on ECG it recovers most of the macro-AUROC and all of the abnormal-beat sensitivity lost to noise, with the best F1 among restoration methods (Tables 13–15). Across modalities, the restoration margin over the strongest baseline is largest on PPG and smallest on EEG, an ordering we relate to cycle structure next.

Phase conditioning and cyclostationarity are connected. Phase conditioning improves restoration more when cycle structure is stronger. To measure how much the denoiser relies on phase, we permute the phase fields across the samples of each test batch while keeping the inputs unchanged. This lowers $\Delta\mathrm{SNR}$ in every modality, most on PPG and least on EEG (Figures 2a and 9), and the drop tracks the margin of PTSDIFF over the strongest baseline, with $r = 0.981$ across the four modalities (Figure 2b). A controlled synthetic study, which varies the cyclic share of variance and the cycle-length jitter with all else fixed, supports this: the $\Delta\mathrm{SNR}$ improvement from phase conditioning increases with the cyclostationarity index Π of Definition 1 ($r = 0.85$; Figure 3, Appendix D.8). Table 18 reports the proxy Π_{AC} for all nine datasets (Appendix D.7), ordering them as the ablations (Table 17) suggest, in which removing phase conditioning altogether costs most on the two cardiac modalities, PPG and ECG, and little or nothing on EMG and EEG. Unlike the shuffle (Appendix D.10), which probes how much a trained model relies on phase, this ablation retrains the model without it (Appendix D.6). Learning the phase is key. On ECG, the learned encoder improves every metric over the phase-free model, whereas the analytical QRS detector of Table 4 is not enough: applied to the noisy input, it performs on par with omitting phase (Table 3). The detector's errors propagate into the phase field, while the encoder learns to infer the clean signal's phase from the same input.

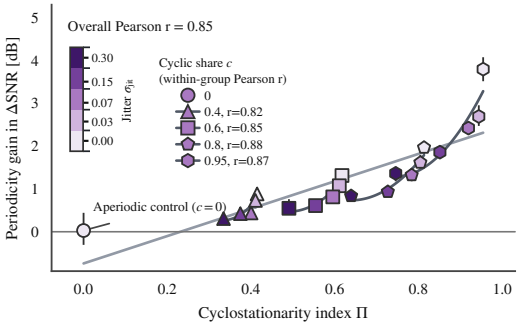


Figure 3: Synthetic sweep: $\Delta\mathrm{SNR}$ gain over the no-periodicity ablation versus the cyclostationarity index Π. Further details in Appendix D.8.

Frame (SWT)	Phase	Context	Analytical phase	Auxiliary loss	ΔSNR ↑	PRD [%] ↓	CC ↑
✓	–	–	–	✓	15.05 ± 0.21	25.51 ± 0.60	0.96 ± 0.00
✓	–	✓	–	✓	15.36 ± 0.21	24.54 ± 0.55	0.96 ± 0.00
✓	✓	–	–	✓	15.28 ± 0.20	25.35 ± 0.65	0.96 ± 0.00
✓	✓	✓	✓	✓	15.34 ± 0.21	24.50 ± 0.54	0.96 ± 0.00
✓	✓	✓	–	–	$\underline{15.52 \pm 0.21}$	24.01 ± 0.53	0.96 ± 0.00
–	✓	✓	–	✓	15.48 ± 0.21	$\underline{23.97 \pm 0.53}$	$\underline{0.96 \pm 0.00}$
✓	✓	✓	–	✓	$\mathbf{16.00 \pm 0.21}$	$\mathbf{22.68 \pm 0.50}$	$\mathbf{0.97 \pm 0.00}$

Table 3: Component ablation on ECG (PTB-XL, SNR -2.0 dB, one antithetic pair, single seed; mean $\pm$ 95% CI). Phase from a classical detector performs similarly to no phase; the learned encoder improves on both.

Further Analyses. In Table 17, we remove one component at a time in each modality; Table 3 details the ECG case, where the full model is best on every metric. Each component improves ΔSNR in three of the four modalities. Phase and context conditioning are complementary: on ECG, each improves the model more when the other is present. The time-domain auxiliary losses (Equation 12) help, as the weighting argument of Section 3.5 predicts, and most of all on EEG, which uses its own wavelet and loss weight (Table 5). The shift-covariant frame contributes comparably on ECG, EMG and EEG, whereas on PPG, whose restoration relies most on phase, removing the frame or the context slightly improves ΔSNR. On ECG, the choice of wavelet and decomposition level matters more than the loss weights, and four levels (Figure 13) are best for every wavelet tested (Tables 19 and 20).

Sampling efficiency. Antithetic coupling helps at no extra cost: at equal NFE, PTSDIFF-AV-10 improves ΔSNR and PRD over PTSDIFF-MC-10 in every modality (Table 1), consistent with Proposition 2 for negatively correlated pairs (measured directly in Figure 10). On ECG, a single antithetic pair already outperforms ten independent trajectories with five times fewer NFEs (Figure 4), and further pairs add little because the remaining error is then dominated by the model-bias term of Equation 16, which does not shrink with K (Figure 11). The benefit transfers to other diffusion restorers: applied to DeScoD-ECG and TFCDiff without retraining, one antithetic pair again matches or exceeds ten independent trajectories at a fifth of the wall-clock time, and PTSDIFF is the fastest of the three at equal NFE (Table 21). Because averaging approximates the conditional mean, these improvements concern distortion rather than the realism of individual samples (Appendix B.2).

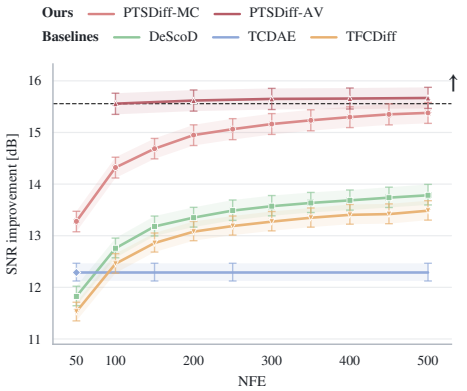


Figure 4: Sampling efficiency on ECG. A single antithetic pair (AV-2) outperforms ten independent trajectories (MC-10) with $5\times$ fewer NFEs. Wall-clock time decreases similarly. Extended results in Figure 11.

6 CONCLUSION

We introduced PTSDIFF, a diffusion-based framework for restoring cyclostationary physiological signals, leveraging the additional structure present in the data that prior work disregards. The framework combines a shift-covariant frame representation with a learned dense phase field, a training-free cyclostationarity index for estimating when phase conditioning is likely to be useful, and antithetic reverse sampling for improved efficient point estimation. Our experiments show that the benefit of phase conditioning increases with the cyclostationary structure of the signal: the cyclostationarity index separates the cardiac modalities from others of less cyclic nature, and its autocorrelation proxy, averaged over datasets, orders the modalities as the ablation gains do. Ablations further demonstrate that this benefit depends on learning a noise-robust phase representation as analytical phase estimated from the corrupted input provides no measurable improvement, whereas a learned encoder captures the required information to act as a directed bias. Finally, antithetic sampling improves distortion-based restoration metrics by better approximating the posterior mean, without modifying the architecture or diffusion schedule in an efficient manner. We show how the results of this cyclostationary-informed approach outperforms prior restoration baselines across four real world modalities and datasets.

Limitations and future work. The relationship between cyclostationarity and restoration benefit should be further tested across broader datasets and naturally occurring artifacts. Similarly, the insights focus on single-channel settings and extending the current framework to multi-lead ECG and multi-channel EEG could exploit correlations between channels. Like prior diffusion-based methods, our approach remains more expensive than single-pass restoration despite antithetic sampling. More broadly, this work opens a direction for designing time-series models around measurable cyclostationary structure. The proposed index provides a way to identify when explicit phase information is likely to be useful, while learned phase representations offer a mechanism for exploiting this structure when it is not directly observable from corrupted data. Extending these ideas to multichannel, irregular, and more weakly cyclic signals could enable adaptive inductive biases beyond physiological restoration, including further time-series reconstruction and generative modeling settings.

REPRODUCIBILITY STATEMENT

The method is specified in Section 3, with architectural details, training hyperparameters, random seeds and the compute environment documented in Appendix C.2 and Table 5. All clean-signal datasets (Wagner et al., 2022; Moody & Mark, 2001; Laguna et al., 1997; Reiss et al., 2019; Atzori et al., 2014; Zhang et al., 2021; Pimentel et al., 2017; Schalk et al., 2004) and noise sources (Moody et al., 1992; Goldberger et al., 2000) are public; the corruption protocols, splits and preprocessing needed to regenerate every evaluation set are given in Appendix C.1, including the subject- and record-level disjointness constraints. Baselines were retrained under identical data protocols rather than quoted from their publications (Section 4); porting, reimplementation and adaptation details are in Appendix C.3. Metric definitions are in Appendix C.4, the statistical procedure in Appendix C.5, and the downstream task setups in Appendix D.5. Proofs of Propositions 1 and 2 are in Appendices B.1 and B.2; the estimator for Π is in Appendix B.3 and the synthetic sweep protocol in Appendix D.8. Anonymized source code is provided at `https://anonymous.4open.science/r/PTSDiff/`.

ETHICS STATEMENT

This work uses only publicly available, de-identified physiological recordings released for research under their respective licenses. No new human-subject data was collected and no re-identification was attempted; as a secondary analysis of public, de-identified data, the work did not require additional ethical review. The principal risk is clinical: a restoration model can produce a plausible waveform that is diagnostically wrong, and a clean-looking output may command more confidence than a visibly noisy one. Two aspects of our results bear on this. First, distortion metrics do not establish diagnostic fidelity, which is why we evaluate downstream tasks (Table 2, Appendix D.5). Second, averaging reverse trajectories approximates the conditional mean, which improves distortion but not the realism of individual samples (Appendix B.2), and can therefore suppress rare morphology that may carry diagnostic weight. Restoration of this kind should not be deployed in a diagnostic pathway without task-level validation on the population of interest, and outputs should not be presented to clinicians as recordings. Benchmark corruption is also synthetic and drawn from a small set of artifact sources, so robustness to real ambulatory artifacts across diverse populations and devices remains unverified.

AI USE STATEMENT

In this work, we used generative AI tools to help implement our methods and experiments in code. We did not use them to generate synthetic datasets, clean or reformat datasets, help develop theoretical models or conceptual frameworks, formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, or interpret results. Translation and qualitative or thematic data analysis do not apply to this work. Additionally, we used generative AI tools to draft and edit parts of the paper text, including condensing captions and laying out tables and figures in LaTeX, but not to search for, identify, or summarize related literature. We reviewed all AI-assisted output: code was checked and tested before any experiment used it, and text was verified against our methods and results. We take full responsibility for everything in this paper, including any text, claims, and artifacts created with generative AI assistance.

REFERENCES

Norbert Ádám, Dávid Val'ko, Zoltán Balogh, Branislav Madoš, and Ján Hurtuk. Comparative evaluation of filtration techniques for ECG signal denoising with emphasis on stationary wavelet transform. *Scientific Reports*, 15(1):42514, 2025. doi: 10.1038/s41598-025-26476-1.

Paul S Addison. Wavelet transforms and the ECG: a review. *Physiological Measurement*, 26(5): R155–R199, 2005. doi: 10.1088/0967-3334/26/5/R01.

Karol Antczak. Deep recurrent neural networks for ECG signal denoising. *arXiv preprint arXiv:1807.11551*, 2018.

Manfredo Atzori, Arjan Gijsberts, Claudio Castellini, Barbara Caputo, Anne-Gabrielle Mittaz Hager, Simone Elsig, Giorgio Giatsidis, Franco Bassetto, and Henning Müller. Electromyography data for non-invasive naturally-controlled robotic hand prostheses. *Scientific Data*, 1(1):140053, 2014. doi: 10.1038/sdata.2014.53.

Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*, 2018.

Giulio Basso, Xi Long, Reinder Haakma, and Rik Vullings. Reduction of motion artifacts from photoplethysmography signals using learned convolutional sparse coding. *Physiological Measurement*, 47(2):025001, 2026. doi: 10.1088/1361-6579/ae35cb.

Lisa Bedin, Gabriel Cardoso, Josselin Duchateau, Remi Dubois, and Eric Moulines. Leveraging an ECG beat diffusion model for morphological reconstruction from indirect signals. In *Advances in Neural Information Processing Systems*, volume 37, pp. 84409–84446, 2024. doi: 10.52202/079017-2682.

Manuel Blanco-Velasco, Binwei Weng, and Kenneth E. Barner. ECG signal denoising and baseline wander correction based on the empirical mode decomposition. *Computers in Biology and Medicine*, 38(1):1–13, 2008. doi: 10.1016/j.compbiomed.2007.06.003.

Marianne Boyer, Laurent Bouyer, Jean-Sébastien Roy, and Alexandre Campeau-Lecours. Reducing noise, artifacts and interference in single-channel EMG signals: A review. *Sensors*, 23(6):2927, 2023. doi: 10.3390/s23062927.

Richard P. Brenner and Neil Schaul. Periodic EEG patterns: Classification, clinical correlation, and pathophysiology. *Journal of Clinical Neurophysiology*, 7(2):249–268, 1990. doi: 10.1097/00004691-199004000-00007.

Shubhojeet Chatterjee, Rini Smita Thakur, Ram Narayan Yadav, Lalita Gupta, and Deepak Kumar Raghuvanshi. Review of noise removal techniques in ECG signals. *IET Signal Processing*, 14(9): 569–590, 2020. doi: 10.1049/iet-spr.2020.0104.

Meng Chen, Yongjian Li, Liting Zhang, Lei Liu, Baokun Han, Wenzhuo Shi, and Shoushui Wei. Elimination of random mixed noise in ECG using convolutional denoising autoencoder with transformer encoder. *IEEE Journal of Biomedical and Health Informatics*, 28(4):1993–2004, 2024. doi: 10.1109/JBHI.2024.3355960.

Hsin-Tien Chiang, Yi-Yen Hsieh, Szu-Wei Fu, Kuo-Hsuan Hung, Yu Tsao, and Shao-Yi Chien. Noise reduction in ECG signals using fully convolutional denoising autoencoders. *IEEE Access*, 7: 60806–60813, 2019. doi: 10.1109/ACCESS.2019.2912036.

Kangjoon Choi and In-Cheol Park. Enhanced-throughput antithetic sample generation architecture for Monte Carlo simulation. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 34(5):1682–1686, 2026. doi: 10.1109/TVLSI.2026.3671706.

Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023.

Ronald R Coifman and David L Donoho. Translation-invariant de-noising. In *Wavelets and statistics*, pp. 125–150. Springer, 1995. doi: 10.1007/978-1-4612-2544-7_9.

Jonathan Crabbé, Nicolas Huynh, Jan Pawel Stanczuk, and Mihaela Van Der Schaar. Time series diffusion in the frequency domain. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 9407–9438. PMLR, 2024. URL https://proceedings.mlr.press/v235/crabbe24a.html.

Heng Cui, Chang Li, Aiping Liu, Ruobing Qian, and Xun Chen. A dual-branch interactive fusion network to remove artifacts from single-channel EEG. *IEEE Transactions on Instrumentation and Measurement*, 73:1–12, 2024. doi: 10.1109/TIM.2023.3342863.

Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992. doi: 10.1137/1.9781611970104.

Philip de Chazal, M. O'Dwyer, and R.B. Reilly. Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Transactions on Biomedical Engineering*, 51(7): 1196–1206, 2004. doi: 10.1109/TBME.2004.827359.

Ergun Erçelebi. Electrocardiogram signals de-noising using lifting-based discrete wavelet transform. *Computers in Biology and Medicine*, 34(6):479–493, 2004. doi: 10.1016/S0010-4825(03)00090-8.

Asadullah Hill Galib, Pang-Ning Tan, and Lifeng Luo. FIDE: Frequency-inflated conditional diffusion model for extreme-aware time series generation. In *Advances in Neural Information Processing Systems*, volume 37, pp. 114434–114457, 2024. doi: 10.52202/079017-3634.

Jianbo Gao, Hussain Sultan, Jing Hu, and Wen-Wen Tung. Denoising nonlinear time series by adaptive filtering and wavelet shrinkage: A comparison. *IEEE Signal Processing Letters*, 17(3): 237–240, 2010. doi: 10.1109/LSP.2009.2037773.

William A Gardner. Statistically inferred time warping: Extending the cyclostationarity paradigm from regular to irregular statistical cyclicity in scientific data. *EURASIP Journal on Advances in Signal Processing*, 2018(1):59, 2018. doi: 10.1186/s13634-018-0564-6.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23):e215–e220, 2000. doi: 10.1161/01.CIR.101.23.e215.

J. M. Hammersley and K. W. Morton. A new Monte Carlo technique: Antithetic variates. *Mathematical Proceedings of the Cambridge Philosophical Society*, 52(3):449–475, 1956. doi: 10.1017/S0305004100031455.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Xiaoyang Huang, Chang Li, Aiping Liu, Ruobing Qian, and Xun Chen. EEGDfus: A conditional diffusion model for fine-grained EEG denoising. *IEEE Journal of Biomedical and Health Informatics*, 29(4):2557–2569, 2025. doi: 10.1109/JBHI.2024.3504716.

Kuo-Hsuan Hung, Kuan-Chen Wang, Kai-Chun Liu, Wei-Lun Chen, Xugang Lu, Yu Tsao, and Chii-Wann Lin. MECG-E: Mamba-based ECG enhancer for baseline wander removal. In Wei Ding, Chang-Tien Lu, Fusheng Wang, Liping Di, Kesheng Wu, Jun Huan, Raghu Nambiar, Jundong Li, Filip Ilievski, Ricardo Baeza-Yates, and Xiaohua Hu (eds.), *IEEE International Conference on Big Data, BigData 2024, Washington, DC, USA, December 15-18, 2024*, pp. 6469–6475. IEEE, 2024. doi: 10.1109/BIGDATA62323.2024.10825253.

Shahid Ismail, Usman Akram, and Imran Siddiqi. Heart rate tracking in photoplethysmography signals affected by motion artifacts: a review. *EURASIP Journal on Advances in Signal Processing*, 2021(1):5, 2021. doi: 10.1186/s13634-020-00714-2.

Alexander Jenkins, Zehua Chen, Fu Siong Ng, and Danilo Mandic. Improving diffusion models for ECG imputation with an augmented template prior. *arXiv preprint arXiv:2310.15742*, 2023.

Xiao Jiang, Gui-Bin Bian, and Zean Tian. Removal of artifacts from EEG signals: A review. *Sensors*, 19(5):987, 2019. doi: 10.3390/s19050987.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 23593–23606, 2022.

Amir Khorrami Chokami, Mauro Gasparini, and Roberto Merletti. Identification of periodic bursts in surface EMG: Applications to the erector spinae muscles of sitting violin players. *Biomedical Signal Processing and Control*, 65:102369, 2021. doi: 10.1016/j.bspc.2020.102369.

Ashish Kumar, Harshit Tomar, Virender Kumar Mehla, Rama Komaragiri, and Manjeet Kumar. Stationary wavelet transform based ECG signal denoising method. *ISA Transactions*, 114:251–262, 2021. doi: 10.1016/j.isatra.2020.12.029.

Pablo Laguna, Roger G Mark, Ary L Goldberger, and George B Moody. The QT Database, 1997. URL `https://physionet.org/content/qtdb/`.

Salim Lahmiri. Comparative study of ECG signal denoising by wavelet thresholding in empirical and variational mode decomposition domains. *Healthcare Technology Letters*, 1(3):104–109, 2014. doi: 10.1049/htl.2014.0073.

Shin-Chi Lai, Yao-Feng Liang, Yi-Chang Zhu, Li-Chuan Hsu, Shang-Sian Wu, and Szu-Ting Wang. Enhanced noise reduction in photoplethysmography signals using a denoising autoencoder. *Sensors and Materials*, 37(11):5123, 2025. doi: 10.18494/sam5927.

C.M. Lee and Y.T. Zhang. Reduction of motion artifacts from photoplethysmographic recordings using a wavelet denoising approach. In *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, 2003.*, pp. 194–195, 2003. doi: 10.1109/APBME.2003.1302650.

Han-Wook Lee, Ju-Won Lee, Won-Geun Jung, and Gun-Ki Lee. The periodic moving average filter for removing motion artifacts from PPG signals. *International Journal of Control, Automation, and Systems*, 5(6):701–706, 2007. URL `https://koreascience.kr/article/JAKO200704503664957.page`.

Chunyan Li. Short-time Fourier transform and introduction to wavelet analysis. In *Time Series Data Analysis in Oceanography: Applications using MATLAB*, pp. 415–426. Cambridge University Press, Cambridge, 2022. doi: 10.1017/9781108697101.024.

Huayu Li, Gregory Ditzler, Janet Roveda, and Ao Li. DeScoD-ECG: Deep score-based diffusion model for ECG baseline wander and noise removal. *IEEE Journal of Biomedical and Health Informatics*, 28(9):5081–5091, 2024. doi: 10.1109/JBHI.2023.3237712. arXiv:2208.00542 [eess.SP].

Pengxin Li, Yimin Zhou, Jie Min, Yirong Wang, Wei Liang, Qingling Xia, and Wang Li. TFCDiff: Robust ECG denoising via time-frequency complementary diffusion. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–13, 2026. doi: 10.1109/JBHI.2026.3732677.

Yu-Tung Liu, Kuan-Chen Wang, Kai-Chun Liu, Sheng-Yu Peng, and Yu Tsao. SDEMG: Score-based diffusion model for surface electromyographic signal denoising. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1736–1740, 2024. doi: 10.1109/ICASSP48485.2024.10446154. arXiv:2402.03808 [eess.SP].

Juan Miguel Lopez Alcaraz and Nils Strodthoff. Diffusion-based time series imputation and forecasting with structured state space models. *Transactions on Machine Learning Research*, 2023. URL `https://openreview.net/forum?id=hHiIbk7ApW`.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, volume 35, pp. 5775–5787, 2022.

G.B. Moody and R.G. Mark. The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001. doi: 10.1109/51.932724.

George B Moody, WE Muldrow, and Roger G Mark. The MIT-BIH Noise Stress Test Database, 1992. URL https://physionet.org/content/nstdb/.

Antonio Napolitano and William A Gardner. Discovering and modeling hidden periodicities in science data. *EURASIP Journal on Advances in Signal Processing*, 2025(1):13, 2025. doi: 10.1186/s13634-025-01207-w.

G. P. Nason and B. W. Silverman. The stationary wavelet transform and some statistical applications. In Anestis Antoniadis and Georges Oppenheim (eds.), *Wavelets and Statistics*, pp. 281–299. Springer, New York, NY, 1995. ISBN 978-1-4612-2544-7. doi: 10.1007/978-1-4612-2544-7_17.

Mang Ning, Mingxiao Li, Jianlin Su, Jia Haozhe, Lanmiao Liu, Martin Benes, Wenshuo Chen, Albert Ali Salah, and Itir Onal Ertugrul. DCTdiff: Intriguing properties of image generative modeling in the DCT space. In Aarti Singh, Maryam Fazel, Daniel Hsu, Simon Lacoste-Julien, Felix Berkenkamp, Tegan Maharaj, Kiri Wagstaff, and Jerry Zhu (eds.), *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 46498–46524. PMLR, 2025. URL https://proceedings.mlr.press/v267/ning25c.html.

Meir Nitzan and Zehava Ovadia-Blechman. Physical and physiological interpretations of the PPG signal. In John Allen and Panicos Kyriacou (eds.), *Photoplethysmography*, pp. 319–340. Academic Press, 2022. ISBN 978-0-12-823374-0. doi: 10.1016/B978-0-12-823374-0.00009-8.

Art B. Owen. *Monte Carlo Theory, Methods and Examples*. Self-published, 2013. URL https://artowen.su.domains/mc/.

Jiapu Pan and Willis J. Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, BME-32(3):230–236, 1985. doi: 10.1109/TBME.1985.325532.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. doi: 10.1609/aaai.v32i1.11671.

Marco A. F. Pimentel, Alistair E. W. Johnson, Peter H. Charlton, Drew Birrenkott, Peter J. Watkinson, Lionel Tarassenko, and David A. Clifton. Toward a robust estimation of respiratory rate from pulse oximeters. *IEEE Transactions on Biomedical Engineering*, 64(8):1914–1923, 2017. doi: 10.1109/TBME.2016.2613124.

Huiling Qin, Yuanxun Li, and Weijia Jia. WaveDiST: A wavelet diffusion transformer for spatio-temporal estimation on unobserved locations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 15635–15643, 2026. doi: 10.1609/aaai.v40i18.38593.

Attila Reiss, Ina Indlekofer, and Philip Schmidt. PPG-DaLiA, 2019. URL https://archive.ics.uci.edu/dataset/495.

Francisco Perdigon Romero, Liset Vazquez Romaguera, Carlos Román Vázquez-Seisdedos, Cícero Ferreira Fernandes Costa Filho, Marly Guimarães Fernandes Costa, and João Evangelista Neto. Baseline wander removal methods for ECG signals: A comparative study. *arXiv preprint arXiv:1807.11359*, 2018.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

Samuel Ruipérez-Campillo, Alain Ryser, Thomas M Sutter, Ruibin Feng, Prasanth Ganesan, Brototo Deb, Kelly A Brennan, Maarten ZH Kolk, Fleur VY Tjong, Albert J Rogers, et al. Can generative ai learn physiological waveform morphologies? a study on denoising intracardiac signals in ischemic cardiomyopathy. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–4. IEEE, 2024.

Samuel Ruipérez-Campillo, Moritz Rau, Prasanth Ganesan, Kelly A Brennan, Ruibin Feng, Sabyasachi Bandyopadhyay, Albert J Rogers, Sanjiv M Narayan, and Julia E Vogt. Physics-inspired diffusion probabilistic models for improved denoising in intracardiac time series. In *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1–5. IEEE, 2025.

Samuel Ruipérez-Campillo, Pablo Blasco-Fernández, Moritz Rau, Prasanth Ganesan, Sabyasachi Bandyopadhyay, Charles Sillett, Lukas P. Arts, Esteban Peralta, Albert J. Rogers, Fleur V. Y. Tjong, Sanjiv M. Narayan, and Julia E. Vogt. Antithetic sampling enhanced probabilistic diffusion for denoising cardiac time series. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–14, 2026a. doi: 10.1109/JBHI.2026.3705976.

Samuel Ruipérez-Campillo, Alain Ryser, Thomas M. Sutter, Brototo Deb, Ruibin Feng, Prasanth Ganesan, Kelly A. Brennan, Albert J. Rogers, Maarten Z. H. Kolk, Fleur V. Y. Tjong, Sanjiv M. Narayan, and Julia E. Vogt. Reducing diverse sources of noise in ventricular electrical signals using variational autoencoders. *Expert Systems with Applications*, 300:130185, 2026b. doi: 10.1016/j.eswa.2025.130185.

Alain Ryser, Laura Manduchi, Fabian Laumer, Holger Michel, Sven Wellmann, and Julia E. Vogt. Anomaly detection in echocardiograms with dynamic variational trajectory models. In *Proceedings of the 7th Machine Learning for Healthcare Conference*, volume 182, pp. 425–458. PMLR, 2022. URL https://proceedings.mlr.press/v182/ryser22a.html.

Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, SIGGRAPH '22, pp. 1–10, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450393379. doi: 10.1145/3528233.3530757.

Omid Sayadi and Mohammad Bagher Shamsollahi. ECG denoising and compression using a modified extended Kalman filter structure. *IEEE Transactions on Biomedical Engineering*, 55(9):2240–2248, 2008. doi: 10.1109/TBME.2008.921150.

Gerwin Schalk, Dennis J. McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R. Wolpaw. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004. doi: 10.1109/TBME.2004.827072.

Alex Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404:132306, 2020. doi: 10.1016/j.physd.2019.132306.

Brij N. Singh and Arvind K. Tiwari. Optimal selection of wavelet basis function applied to ECG signal denoising. *Digital Signal Processing*, 16(3):275–287, 2006. doi: 10.1016/j.dsp.2005.12.003.

Pratik Singh and Gayadhar Pradhan. A new ECG denoising framework using generative adversarial network. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(2):759–764, 2021. doi: 10.1109/TCBB.2020.2976981.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *9th International Conference on Learning Representations*, 2021a.

Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021b. URL https://openreview.net/forum?id=PxTIG12RRHS.

Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, and Wojciech Samek. Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1519–1528, 2021. doi: 10.1109/JBHI.2020.3022989.

Mohammad Sadman Tahsin, Haitham Y. Adarbah, and Afzel Noore. A lightweight channel-attention CNN for robust beat-level arrhythmia detection. *IEEE Access*, 14:62803–62817, 2026. doi: 10.1109/ACCESS.2026.3686810.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Dieter Kreiseler, Fatima I. Lunze, Wojciech Samek, and Tobias Schaeffter. PTB-XL, a large publicly available electrocardiography dataset. *Scientific Data*, 7(1):154, 2020. doi: 10.1038/s41597-020-0495-6.

Patrick Wagner, Nils Strodthoff, Ralf-Dieter Bousseljot, Wojciech Samek, and Tobias Schaeffter. PTB-XL, a large publicly available electrocardiography dataset. *PhysioNet*, 2022. doi: 10.13026/kfzx-aw45. Version 1.0.3.

Kuan-Chen Wang, Kai-Chun Liu, Sheng-Yu Peng, and Yu Tsao. ECG artifact removal from single-channel surface EMG using fully convolutional networks. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2023. doi: 10.1109/ICASSP49357.2023.10096409. ISSN: 2379-190X.

Yuna Watanabe, Natasha Yamane, Aarti Sathyanarayana, Varun Mishra, and Matthew S. Goodwin. Beyond motion artifacts: Optimizing PPG preprocessing for accurate pulse rate variability estimation. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp Companion '25, pp. 1176–1182, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400714771. doi: 10.1145/3714394.3756241.

Peter D. Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15(2):70–73, 1967. doi: 10.1109/TAU.1967.1161901.

R. F. Woolson. Wilcoxon signed-rank test. In *Wiley Encyclopedia of Clinical Trials*, pp. 1–3. John Wiley & Sons, Ltd, 2008. ISBN 978-0-471-46242-2. doi: 10.1002/9780471462422.eoct979.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems*, volume 34, pp. 22419–22430, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. TimesNet: Temporal 2D-variation modeling for general time series analysis. In *The Eleventh International Conference on Learning Representations*, 2023.

Liping Xie, Zilong Li, Yihan Zhou, Yiliu He, and Jiaxin Zhu. Computational diagnostic techniques for electrocardiogram signal analysis. *Sensors*, 20(21):6318, 2020. doi: 10.3390/s20216318.

Maxwell Xu, Alexander Moreno, Supriya Nagesh, Varol Aydemir, David Wetter, Santosh Kumar, and James Rehg. PulseImpute: A novel benchmark task for pulsative physiological signal imputation. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 26874–26888. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/ac01e21bb14609416760f790dd8966ae-Paper-Datasets_and_Benchmarks.pdf.

Haoming Zhang, Mingqi Zhao, Chen Wei, Dante Mantini, Zherui Li, and Quanying Liu. EEGdenoiseNet: a benchmark dataset for deep learning solutions of EEG denoising. *Journal of Neural Engineering*, 18(5):056057, 2021. doi: 10.1088/1741-2552/ac2bf8.

Qiang Zhu, Xin Tian, Chau-Wai Wong, and Min Wu. Learning your heart actions from pulse: ECG waveform reconstruction from PPG. *IEEE Internet of Things Journal*, 8(23):16734–16748, 2021. doi: 10.1109/JIOT.2021.3097946.

APPENDIX CONTENTS

A Extended Related Work . . . 18

B Extended Methods . . . 19

B.1 Proof of Proposition 1 (phase-explained variance) . . . 19

B.2 Proof of Proposition 2 (matched-NFE variance) . . . 19

B.3 Estimating the Cyclostationarity Index . . . 20

C Extended Experimental Setup . . . 21

C.1 Per-Modality Preprocessing . . . 21

C.2 Architecture and Training Details . . . 22

C.3 Baseline Implementation Details . . . 23

C.4 Evaluation Metrics . . . 24

C.5 Statistical Evaluation Procedure . . . 25

D Extended Results . . . 25

D.1 Full ECG results across datasets . . . 25

D.2 Results stratified by corruption level . . . 25

D.3 Cross-dataset generalization . . . 31

D.4 Paired statistical comparison . . . 31

D.5 Extended downstream evaluation . . . 31

D.6 Component ablation across modalities . . . 34

D.7 Cyclostationarity index across datasets . . . 34

D.8 Synthetic Cyclostationarity Sweep . . . 35

D.9 Frame and objective selection . . . 35

D.10 Phase-shuffle experiment . . . 35

D.11 Antithetic sampling . . . 37

D.12 Per-sample distribution . . . 39

D.13 The frame decomposition . . . 39

D.14 Qualitative results . . . 39

A Extended Related Work

Section 2 organizes prior work by idea. This appendix gives the per-modality detail compressed out of the main text.

ECG. Classical restoration spans Kalman filtering (Sayadi & Shamsollahi, 2008), empirical mode decomposition (Blanco-Velasco et al., 2008), adaptive filtering and wavelet shrinkage (Gao et al., 2010), with dedicated treatments of baseline wander (Romero et al., 2018) and powerline interference (Chatterjee et al., 2020). Learned approaches progress from convolutional autoencoders (Chiang et al., 2019) through recurrent models (Antczak, 2018; Sherstinsky, 2020) to transformer autoencoders (Chen et al., 2024; Vaswani et al., 2017) and adversarial formulations (Singh & Pradhan, 2021). Generative restoration begins with DeScoD-ECG (Li et al., 2024); time–frequency variants include MECG-E (Hung et al., 2024) and TFCDiff (Li et al., 2026). Wavelet choice for ECG has its own literature (Addison, 2005; Singh & Tiwari, 2006; Kumar et al., 2021; Erçelebi, 2004; Lahmiri, 2014), and the frequency content justifying our band assignment is documented by Xie et al. (2020).

PPG. Motion artifacts and baseline drift dominate (Ismail et al., 2021), addressed classically by wavelet thresholding (Lee & Zhang, 2003), periodic moving-average filtering (Lee et al., 2007) and EMD, and more recently by dilated denoising autoencoders (Lai et al., 2025). The cardiac origin of PPG pulse structure (Nitzan & Ovadia-Blechman, 2022) places it at the high-Π end of the spectrum, and distortion of the pulse contour is the principal failure mode of aggressive filtering (Watanabe et al., 2025).

EEG. Preprocessing conventionally removes ocular, muscular, cardiac and instrumental artifacts by regression, ICA, PCA or EMD (Jiang et al., 2019). Learned approaches include dual-branch convolutional–recurrent fusion (Cui et al., 2024) and conditional diffusion (Huang et al., 2025). EEG carries oscillatory rather than event-locked structure (Brenner & Schaul, 1990), giving it low Π.

EMG. Classical work targets motion artifacts, cross-talk and high-frequency noise via filtering, regression and wavelet-ICA/EMD (Boyer et al., 2023). Fully convolutional models remove ECG interference effectively (Wang et al., 2023), and SDEMG (Liu et al., 2024) applies score-based diffusion with explicit burst modeling. Motor-unit firing gives EMG burst-locked rather than waveform-locked structure (Khorrami Chokami et al., 2021), placing it between ECG and EEG.

Related restoration and representation work. Broader context includes DCT-domain diffusion (Ning et al., 2025), learning-based reconstruction under structured corruption (Zhu et al., 2021), imputation for physiological streams (Lopez Alcaraz & Strodthoff, 2023; Xu et al., 2022; Jenkins et al., 2023), and comparative evaluations of restoration pipelines (Ádám et al., 2025; Basso et al., 2026; Bedin et al., 2024).

B Extended Methods

B.1 Proof of Proposition 1 (phase-explained variance)

By the law of total variance applied to the conditioning on φ,

$$\mathrm{Var}(x) = \underbrace{\mathrm{Var}_{\varphi}\big(\mathbb{E}[x \mid \varphi]\big)}_{=\Pi\ \mathrm{Var}(x)} + \mathbb{E}_{\varphi}\big[\,\mathrm{Var}(x \mid \varphi)\big], \tag{17}$$

which rearranges to Equation 9. Since $\mathrm{MMSE}(x) = \mathrm{Var}(x)$ and $\mathrm{MMSE}(x \mid \varphi) = \mathbb{E}_{\varphi}[\mathrm{Var}(x \mid \varphi)]$, the identity states that knowing φ reduces the minimum mean-squared error by exactly $\Pi\,\mathrm{Var}(x)$: Π is the share of total variance that phase explains, and $(1-\Pi)\,\mathrm{Var}(x)$ is the share it cannot. □

Scope. Equation 9 concerns the unconditional problem and does not extend to the case where y is observed. Applying the law of total variance conditionally on y gives

$$\mathrm{MMSE}(x \mid y) - \mathrm{MMSE}(x \mid y, \varphi) = \mathbb{E}_{y}\big[\,\mathrm{Var}_{\varphi|y}\big(\mathbb{E}[x \mid y, \varphi]\big)\big], \tag{18}$$

which $\Pi\,\mathrm{Var}(x)$ does not bound. Equation 1 permits $C(\varphi_n, \varphi_m)$ to depend on phase, whereas Π captures only the first-moment dependence, so a process whose phase modulates conditional variance but not conditional mean has $\Pi = 0$ while Equation 18 is strictly positive: knowing φ tells the estimator which conditional variance to assume. A phase prior can therefore help more than Π suggests. We use Π as a measure of how strongly phase structures a signal, and the phase-conditioning analysis in Section 5 tests empirically whether realized gains track it.

B.2 Proof of Proposition 2 (matched-NFE variance)

Fix the observation y and a sample index τ; all expectations are over the reverse-process noise with y held fixed. Let $\hat{x}_{0,+}$ and $\hat{x}_{0,-}$ be the two reconstructions of an antithetic pair generated by Equation 14. Because z_t and $-z_t$ are identically distributed, and the two chains apply the same deterministic map μ_θ with the same conditioning, the two members are exchangeable:

$$\mathbb{E}[\hat{x}_{0,+}(\tau)] = \mathbb{E}[\hat{x}_{0,-}(\tau)] = \mu(\tau), \qquad \mathrm{Var}[\hat{x}_{0,+}(\tau)] = \mathrm{Var}[\hat{x}_{0,-}(\tau)] = \sigma^2(\tau). \tag{19}$$

The pair estimator $\hat{x}_{0,\mathrm{AV}} = \frac{1}{2}(\hat{x}_{0,+} + \hat{x}_{0,-})$ is therefore unbiased for $\mu(\tau)$, the same quantity estimated by independent averaging. Writing $\rho(\tau)$ for the correlation between the pair members,

$$\mathrm{Var}[\hat{x}_{0,\mathrm{AV}}(\tau)] = \tfrac{1}{4}\big(\sigma^2(\tau) + \sigma^2(\tau) + 2\rho(\tau)\sigma^2(\tau)\big) = \tfrac{1}{2}\sigma^2(\tau)\big(1 + \rho(\tau)\big). \tag{20}$$

The K-trajectory estimator averages $K/2$ pairs drawn independently, so

$$\mathrm{Var}\left[\hat{x}_{0,\text{AV-}K}(\tau)\right] = \frac{2}{K} \cdot \frac{\sigma^2(\tau)\big(1+\rho(\tau)\big)}{2} = \frac{\sigma^2(\tau)}{K}\big(1+\rho(\tau)\big), \tag{21}$$

establishing Equation 15. Each pair costs two reverse chains, so K antithetic trajectories and K independent trajectories both require KT network evaluations; the independent estimator has variance $\sigma^2(\tau)/K$. The ratio is exactly $1 + \rho(\tau)$, which is < 1 iff $\rho(\tau) < 0$.

For Equation 16, the estimator is unbiased for $\bar{\mu}_\theta(y)$ at every index, so the standard bias–variance split applies coordinatewise,

$$\mathbb{E}\big\|\hat{x}_{0,K} - x_0\big\|_2^2 = \sum_{\tau} \Big(\big(\bar{\mu}_\theta(y)(\tau) - x_0(\tau)\big)^2 + \mathrm{Var}\left[\hat{x}_{0,K}(\tau)\right] \Big). \tag{22}$$

Substituting Equation 15 into the second term and factoring out K^{-1} gives $K^{-1}\sum_\tau \sigma^2(\tau)(1 + \rho(\tau)) = K^{-1}(1+\bar{\rho})\sum_\tau \sigma^2(\tau)$ by the definition of the variance-weighted mean correlation $\bar{\rho}$, which is Equation 16. Setting $\rho \equiv 0$ recovers the independent case. □

Equation 16 separates the error of the learned posterior mean from the sampling variance that additional trajectories can reduce. At matched NFE, antithetic coupling improves over independent averaging whenever $\bar{\rho} < 0$, with gains decaying as $1/K$ and eventually saturating at the model-bias term. These gains improve the estimate of the conditional mean, and therefore distortion, but do not imply a better learned distribution or better individual samples.

When $\rho < 0$ holds. For a function f monotone in a single uniform variable, antithetic negative correlation is classical (Owen, 2013, Ch. 8.2). No such argument applies here, since the map from $\{z_t\}$ to $\hat{x}_0$ passes through a learned nonlinear network and ρ could in principle be positive, in which case Equation 15 predicts a variance increase. In the conditional-diffusion setting Ruipérez-Campillo et al. (2026a) report pair correlations between -0.98 and -0.88 through most of the reverse trajectory, relaxing near $t = 0$. Appendix D.11 measures the variance-weighted correlation of PTSDIFF directly and finds it negative in every modality: it never rises above -0.89 along the reverse trajectory and relaxes only near $t = 0$, in close agreement with Ruipérez-Campillo et al. (2026a).

B.3 Estimating the Cyclostationarity Index

Definition 1 is estimated from clean data by phase-aligned averaging. Given clean signals $\{x^{(i)}\}$:

1. **Detect cycle events.** Apply the analytical ECG detector $\mathcal{D}$ (Table 4) to the *clean* signal, giving event times $\{t_k\}$. This is the same detector used for phase-encoder warm-start, so no additional machinery is introduced.
2. **Resample to a phase grid.** For each cycle $[t_k, t_{k+1})$, linearly resample the segment onto a fixed grid of G phase bins ($G = 100$ throughout), giving $s_k \in \mathbb{R}^G$. Cycles whose length falls outside the physiological range $[1/f_{\max}, 1/f_{\min}]$ are discarded.
3. **Average and normalize.** Set $\hat{\mu}(\varphi) = N^{-1} \sum_k s_k[\varphi]$ over the N retained cycles and

$$\hat{\Pi} = \frac{\mathrm{Var}_\varphi\left(\hat{\mu}(\varphi)\right)}{G^{-1} \sum_\varphi \left(N^{-1} \sum_k \mathrm{Var}(s_k[\varphi])\right) + \mathrm{Var}_\varphi\left(\hat{\mu}(\varphi)\right)}, \tag{23}$$

   amplitude-normalizing each cycle beforehand so that $\hat{\Pi}$ measures shape consistency rather than amplitude stability.

Finite-sample bias. $\mathrm{Var}_\varphi(\hat{\mu})$ is upward-biased at small N, since averaging few cycles leaves residual noise in $\hat{\mu}$ that inflates its variance. The bias is $O(1/N)$ and we correct it by subtracting the within-bin variance divided by N; with $N \geq 200$ cycles per dataset the correction changes $\hat{\Pi}$ by less than 0.01 in all cases we checked.

A cheaper proxy. Where no reliable detector exists, the peak normalized autocorrelation within a plausible cycle-lag window,

$$\Pi_{\mathrm{AC}} = \frac{1}{K} \sum_{k=1}^{K} \max_{\ell \in [L_{\min}, L_{\max}]} \frac{\hat{R}_k(\ell)}{\hat{R}_k(0)}, \qquad \hat{R}_k(\ell) = \frac{1}{W - \ell} \sum_{t=1}^{W-\ell} x_{k,t}\, x_{k,t+\ell}, \tag{24}$$

requires no event detection and is computable in one pass. It is computed on K mean-removed windows x_k of $W = 4f_s$ samples (or the whole segment if shorter) with 50% overlap, with $L_{\min} = f_s/f_{\max}$ and $L_{\max} = \min(f_s/f_{\min}, W/2)$; the factor $1/(W - \ell)$ removes the downward bias at long lags, and the short windows track slow drift in cycle length. This is a weaker quantity, measuring self-similarity at a single lag instead of the variance decomposition of Definition 1, and is not the quantity Proposition 1 bounds, but it works as a screen. We report both for ECG and the proxy for the other modalities in Table 18 (Appendix D.7).

C Extended Experimental Setup

C.1 Per-Modality Preprocessing

This section details the datasets, splits, held-out test sets and corruption protocols summarized in Section 4. All datasets follow their established benchmark protocols. For ECG, EMG and EEG, corruption noise is always drawn from held-out records or record halves, disjoint from training noise in both channel and segment, whereas for PPG training and testing noise are sampled from the same segments.

ECG (PTB-XL, MIT-BIH, QTDB). PTB-XL (Wagner et al., 2022; 2020) contains 21,799 ten-second 12-lead records from 18,869 patients; we use the 500 Hz records, lead II, and the official stratified folds (1–8 train, 9 validation, 10 test). Following Li et al. (2026), signals are converted to fixed 10 s single-lead windows at 360 Hz, band-passed with a zero-phase fifth-order Butterworth filter (0.5–40 Hz), smoothed with a short moving average, and polyphase-resampled. QTDB (Laguna et al., 1997; Goldberger et al., 2000) uses the same pipeline, excluding records sourced from MIT-BIH to avoid overlap. MIT-BIH (Moody & Mark, 2001) is natively 360 Hz and is filtered, baseline-corrected and split into non-overlapping 10 s windows without resampling; lead MLII, split DS1/DS2 following de Chazal et al. (2004). Corruption mixes baseline-wander, electrode-motion and muscle-artifact records from NSTDB (Moody et al., 1992) as $y = x_0 + \lambda \frac{\mathrm{ptp}(x_0)}{\mathrm{ptp}(n)} n$ with $\lambda \in \{0.20, \ldots, 2.00\}$; training noise comes from the first half of channel 1, evaluation noise from the second half of channel 2.

PPG (PPG-DaLiA, BIDMC). Wrist BVP recorded with an Empatica E4 at 64 Hz (Reiss et al., 2019). No reference implementation of the DDAE baseline or its preprocessing was available, so both were reimplemented following Lai et al. (2025); the resulting dataset is frozen and shared identically between PTSDiff and the baseline, so differences are attributable to the model, not to data handling. Signals are split into non-overlapping 512-sample (≈8 s) windows and screened by signal-quality index: segments are discarded if nearly flat, if they lack two pulse peaks separated by at least 20 samples, or if skewness falls below 0.3; the smoothness test requires $\frac{1}{L-1}\sum_n |s[n+1] - s[n]| \geq 0.002\,\mathrm{ptp}(s)$. Accepted segments are min–max normalized. NSTDB noise is resampled to 64 Hz and mixed with weights $(\beta_{\mathrm{MA}}, \beta_{\mathrm{BW}}, \beta_{\mathrm{EM}}) \in \{(0.33, 0.33, 0.34), (0.60, 0.20, 0.20), (0.20, 0.60, 0.20), (0.20, 0.20, 0.60)\}$, covering balanced, muscle-dominant, baseline-dominant and motion-dominant mixtures. Subject-level 80/10/10 split; all 4×6 noise-configuration/SNR combinations are materialized for test. The held-out BIDMC set (Pimentel et al., 2017) contains 53 eight-minute pulse-oximeter recordings at 125 Hz; the PLETH channel is polyphase-resampled to 64 Hz and then windowed, screened, normalized and corrupted exactly as the PPG-DaLiA test split, which keeps 2,317 of 3,180 windows from 52 records (55,608 test segments).

EMG (NinaPro DB2, DB3). Following the SDEMG protocol (Liu et al., 2024), clean sEMG comes from NinaPro DB2 (Atzori et al., 2014) at 2000 Hz and ECG interference from NSRDB (Goldberger et al., 2000). EMG channels are band-passed (fourth-order zero-phase Butterworth, 20–500 Hz), downsampled to 1000 Hz, normalized per recording by maximum absolute amplitude, and split into non-overlapping 10 s windows ($L = 10{,}000$). Training uses channel 2, exercise 1, subjects S11–S40; validation channel 2, exercise 3, subjects S1–S10; test channel 9, exercise 2, subjects S1–S10. ECG interference is resampled from 128 Hz to 1000 Hz and band-limited (third-order zero-phase, 10–200 Hz), with records split disjointly across partitions (test records include 16420, 16539, 16786). Mixing uses $y = x_0 + \alpha w$ with $\alpha = \sqrt{\sum_n x_0[n]^2 / (10^{\lambda/10} \sum_n w[n]^2)}$; training and validation SNR grid $\{-5, -7, -9, -11, -13, -15\}$ dB, test grid $\{0, -2, \ldots, -14\}$ dB, with 10 contaminated copies per training segment and 3 for validation and test. The held-out NinaPro DB3 (Atzori et al., 2014) records 11 trans-radial amputees with the same 12-electrode, 2 kHz setup as DB2. We apply the DB2 test configuration unchanged (exercise 2, channel 9, same filtering, normalization, windowing, SNR grid and NSRDB test records, which are disjoint from the training and validation interference), excluding subjects 6 and 7, whose channel 9 is flat. Each of the 1,065 windows is mixed once with each test record at every SNR (25,560 test segments).

Table 4: Analytical ECG cycle-event detector used to construct the warm-start target $\Phi^\star = \mathcal{D}(x_0)$. Its parameters are fixed and the detector is not used at inference.

Modality	Detector mode	Refractory [ms]	Smoothing [ms]	Event width [ms]	Threshold scale	$f_{\min}$ [Hz]	$f_{\max}$ [Hz]
ECG	QRS peaks	220	80	35	0.5	0.5	3.0

EEG (EEGdenoiseNet, EEGMMIDB). EEG–EOG setting, following the EEGDfus protocol (Huang et al., 2025) on EEGdenoiseNet (Zhang et al., 2021). Epochs have $L = 512$; clean EEG and EOG artifact epochs are randomly paired after truncating the EEG array to the number of available EOG epochs, and each signal is standardized before mixing. An 80/10/10 split is used, with 11 random recombinations of clean and artifact sources to enlarge the training set. Training SNR is drawn uniformly from $[-5, 5]$ dB; all integer levels in $\{-5, \dots, 5\}$ dB are materialized for test. Mixing follows the same α formula as EMG. The held-out EEGMMIDB (Schalk et al., 2004; Goldberger et al., 2000) contains 64-channel recordings of 109 subjects. Channel Oz is band-passed (1 Hz to 80HZ), notch-filtered at 60 Hz, resampled to 256 Hz and split into non-overlapping 512-sample windows. As these recordings are not curated to be artifact-free, windows with blinks, excess high-frequency (30–80 Hz) power, or outlying kurtosis or amplitude are rejected; the kurtosis and high-frequency thresholds are the 99th percentiles of the EEGdenoiseNet clean epochs. Up to ten windows per subject (1,066 from 108 subjects) are standardized and mixed with held-out EEGdenoiseNet EOG epochs at every SNR in $\{-5, \dots, 5\}$ dB (11,726 test segments).

C.2 Architecture and Training Details

Architecture. The denoising network ϵ_θ is a one-dimensional U-Net (Ronneberger et al., 2015; Ho et al., 2020) with three resolution levels, two residual blocks per level, stride-2 downsampling, nearest-neighbor upsampling with skip connections, and self-attention at the bottleneck and deepest encoder and decoder blocks. Its dense input is the channel-wise concatenation of x_t^{F}, y^{F}, and $\hat{\Phi}^{\mathrm{F}}$. Since x_t^{F} and y^{F} each contain $J + 1$ channels and $\hat{\Phi}$ contains five channels, the SWT-transformed phase field contributes $5(J + 1)$ channels, giving $7(J + 1) = 35$ input channels for $J = 4$. The phase encoder $\mathcal{E}_{ph}$ consists of a convolutional stem followed by six dilated residual blocks with dilations $1, 2, 4, 8, 16, 32$ (Bai et al., 2018). Its receptive field spans multiple local cycles, allowing it to estimate both cycle position and the local cycle rate r_n of Equation 6. The context encoder $\mathcal{E}_{ctx}$ produces the pooled conditioning $c_y = \{c_y^{(\ell)}\}$ used in the main text: one embedding for each U-Net resolution and one global embedding. The resolution-specific embeddings modulate the corresponding U-Net features through FiLM (Perez et al., 2018), while the global embedding is combined with the diffusion-step embedding. Figure 5 shows the block-level architecture.

Band mapping across modalities. For a signal sampled at f_s with J decomposition levels, detail channel d_j covers approximately $[f_s 2^{-(j+1)}, f_s 2^{-j}]$ and the approximation channel a_J the residual below $f_s 2^{-(J+1)}$. With $J = 4$: at 360 Hz (ECG) the channels are 90–180, 45–90, 22.5–45, 11.25–22.5 and 0–11.25 Hz, so the physiological band (Xie et al., 2020) concentrates in a_4, d_4, d_3 while d_1, d_2 carry predominantly artifact energy. At 64 Hz (PPG) the same structure lands at 16–32, 8–16, 4–8, 2–4 and 0–2 Hz; at 1000 Hz (EMG) at 250–500, 125–250, 62.5–125, 31.25–62.5 and 0–31.25 Hz. Finally, at 256 Hz, the bands are 64–128, 32–64, 16–32, 8–16 and 0–8 Hz. The assignment is derived from f_s, not tuned per modality.

Training hyperparameters and compute environment. Table 5 summarizes the modality-specific training settings; all unlisted architectural and optimization settings are shared across modalities. Unless otherwise noted, all results use a single training seed (67); the main restoration table (Table 1) is the exception, trained and evaluated with three seeds (42, 67, 123) and reported as mean $\pm$ standard deviation. Experiments were run as single-GPU jobs on a Slurm cluster equipped primarily with NVIDIA RTX 4090 GPUs, with some runs performed on RTX 3090 GPUs. For PTSDiff, training used four CPU data-loading workers, while validation used synchronous loading.

Phase-encoder warm-start. For ECG, we warm-start $\mathcal{E}_{ph}$ against the target phase field $\Phi^\star = \mathcal{D}(x_0)$ computed from the clean signal, where $\mathcal{D}$ applies a Pan–Tompkins-style QRS detector (Pan & Tompkins, 1985) with the fixed parameters of Table 4, followed by the phase-field construction

Table 5: Training hyperparameters across modalities. Frame, backbone, diffusion schedule, objective form and sampler are shared; only the specified entries differ. EEG uses a different wavelet and morphology weight inherited from an earlier configuration that continues to perform better for that modality; a full ablation matching Tables 19 and 20 has not been run for EEG.

Hyperparameter	ECG	PPG	EEG	EMG
Learning rate	1×10^{-4}	1×10^{-3}	5×10^{-4}	1×10^{-4}
Scheduler	MultiStepLR	ReduceLROnPlateau	MultiStepLR	ReduceLROnPlateau
Scheduler settings	milestones $(0.4, 0.7, 0.9)$, $\gamma = 0.1$	factor 0.5, patience 10	milestones (0.95), $\gamma = 0.1$	factor 0.5, patience 5
Phase warm-start	Yes	No	No	No
Wavelet	Sym4	Sym4	Db4	Sym4
Decomposition levels J	4	4	4	4
Objective	$\mathcal{L}_{\text{diff}} + 0.1\mathcal{L}_{\text{morph}} + 0.3\mathcal{L}_{\text{rec}}$	same as ECG	$\mathcal{L}_{\text{diff}} + 0.05\mathcal{L}_{\text{morph}} + 0.3\mathcal{L}_{\text{rec}}$	same as ECG
Batch size	32	32	32	16
Gradient accumulation	1	1	1	2
Effective batch size	32	32	32	32
Epochs	500	300	420	60
Base channels	64	64	128	64
EMA	disabled	disabled	disabled	enabled, decay 0.995

Shared: AdamW (weight decay 10^{-4}), gradient clipping at norm 1.0; $T = 50$ diffusion steps with a quadratic schedule, β_t from 10^{-4} to 0.5; conditioning dropout 0.3; per-sample input normalization by the standard deviation of the corrupted input; checkpoint selected by validation ΔSNR.

of Equations 6–7. The warm-start objective combines a dense field loss $\mathcal{L}_{\text{feat}} = \|\hat{\Phi} - \Phi^\star\|_2^2$, a Huber loss on the local rate estimate $\mathcal{L}_f = H(\hat{f} - f^\star)$, and a circular phase-alignment loss $\mathcal{L}_\phi = (\sin 2\pi\hat{\phi}_0 - \sin 2\pi\phi_0^\star)^2 + (\cos 2\pi\hat{\phi}_0 - \cos 2\pi\phi_0^\star)^2$, summed with unit weights. We pretrain the phase encoder for 60 epochs, freeze it for the first 200 epochs of joint restoration training, and subsequently optimize it through the restoration objective. The analytical construction $\mathcal{D}$ is used only for this ECG initialization and is never used at inference; the phase encoders for PPG, EEG, and EMG are trained directly through the restoration objective.

C.3 Baseline Implementation Details

This section details the baselines listed in Section 4, whose results appear in Table 1 and Appendix D. All learned baselines are retrained on the same prepared tensors, splits, and normalization as PTSDiff, using seed 67 and checkpoint selection by validation loss. We preserve the published architectures and objectives unless noted below. Deterministic methods produce one reconstruction; diffusion baselines use the original publications' sampling procedure.

ECG. The classical ECG baselines are a zero-phase, 721-tap Hamming-window FIR filter with a 0.5–40 Hz passband and an eight-level `sym6` SWT denoiser using band pruning and soft BayesShrink thresholding. TCDAE (Chen et al., 2024) is a deterministic convolutional–Transformer autoencoder that combines local feature extraction with global temporal attention. We port its architecture and frequency-weighted reconstruction loss from Keras to PyTorch and cap training at 400 epochs with early stopping. DeScoD-ECG (Li et al., 2024) is a conditional time-domain DDPM with a hierarchical multi-kernel denoiser, while TFCDiff (Li et al., 2026) performs conditional diffusion on the first 1,000 DCT coefficients. Both use their published 50-step quadratic schedules and are trained for at most 400 epochs. The PTB-XL checkpoints are evaluated on MIT-BIH and QTDB without retraining.

PPG. DDAE (Lai et al., 2025) is a deterministic dilated residual autoencoder designed for PPG motion-artifact removal. Because no reference implementation was available, we reimplemented the published architecture and its equally weighted MSE–ℓ_1 objective. We use the reported 250 epochs, AdamW optimizer, and batch size 32, but disable additional noise augmentation because the shared training tensors are already corrupted.

EEG. EEGIFNet (Cui et al., 2024) separately estimates clean EEG and artifact components before combining them through a learned mask; we use its ocular-artifact variant. EEGDfus (Huang et al., 2025) is a dual-branch conditional diffusion model using a 500-step linear schedule. For both methods, we retain the published model and training settings but replace the runtime data split with the fixed split shared by all methods.

EMG. FCN (Wang et al., 2023) is a deterministic fully convolutional encoder–decoder. We transfer it to EMG by adapting only the input length and retraining it on NinaPro with its ℓ_1 reconstruction

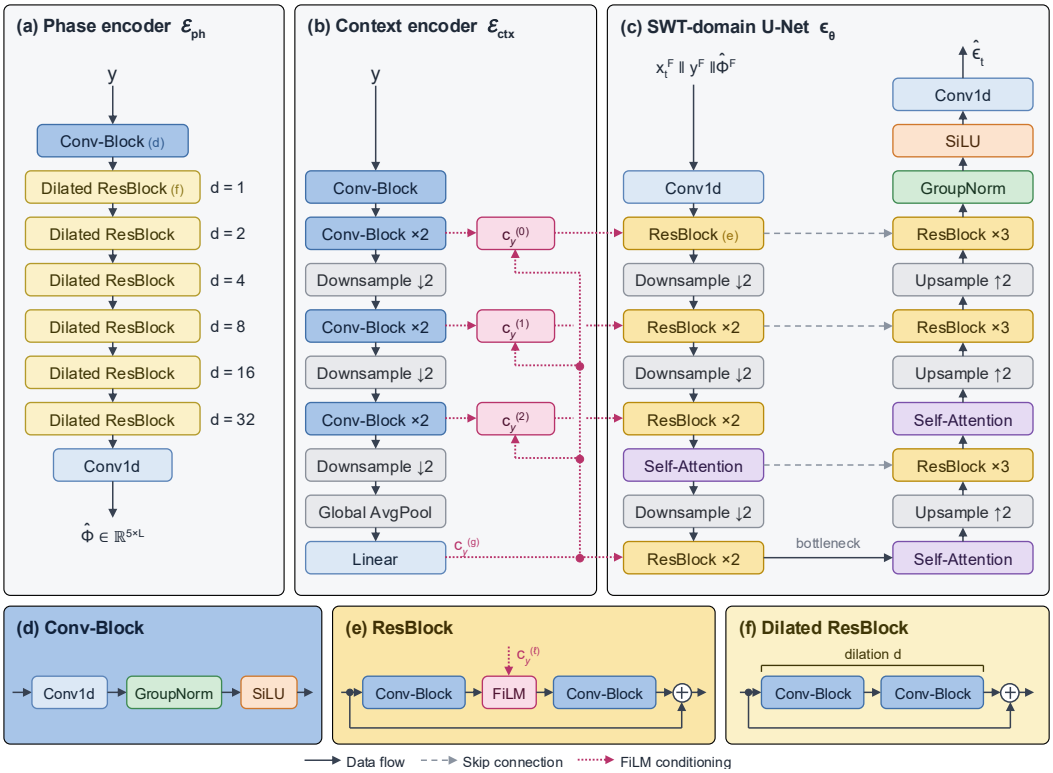


Figure 5: Block-level architecture of PTSDIFF, complementing the overview in Figure 1. **(a)** Phase encoder $\mathcal{E}_{ph}$: a convolutional stem, six dilated residual blocks ($d = 1, 2, 4, 8, 16, 32$) whose receptive field spans several cycles, and a projection to the five channels of the predicted phase field $\hat{\Phi}$, which is then frame-transformed as $\hat{\Phi}^{\mathrm{F}} = \mathcal{W}(\hat{\Phi})$. **(b)** Context encoder $\mathcal{E}_{ctx}$: three strided stages followed by global average pooling, emitting one pooled embedding $c_y^{(\ell)}$ per U-Net resolution and a global embedding $c_y^{(g)}$. **(c)** SWT-domain denoising U-Net ϵ_θ, taking the channel-wise concatenation $x_t^{\mathrm{F}} \,\|\, y^{\mathrm{F}} \,\|\, \hat{\Phi}^{\mathrm{F}}$ ($7(J{+}1) = 35$ channels at $J = 4$) and predicting $\hat{\epsilon}_t$; dashed arrows are skip connections. **(d–f)** Internal structure of the Conv-Block, the FiLM-modulated ResBlock and the dilated residual block. Each detail panel carries the colour of the block it expands, and the letter shown at a block type's first occurrence points to its panel. $\times k$ denotes k stacked blocks. Decoder FiLM connections and the diffusion-step embedding are omitted for clarity.

objective. SDEMG (Liu et al., 2024) is a conditional diffusion model specialized for removing ECG contamination from surface EMG; we retain its 50-step cosine schedule and ℓ_2 noise-prediction objective, but we adapt its batch size from 64 to 16 and increase gradient accumulation from 1 to 4 batches due to memory constraints. We also adapt its data interface to the shared tensors.

C.4 Evaluation Metrics

Table 6 defines the three metrics reported in the main text (Section 4) and the benchmark-specific metrics used in Table 7 and Figure 7. All metrics are computed per test sample and averaged. Let x_0 be the clean signal, y the corrupted input, $\hat{x}_0$ the estimate, $e = \hat{x}_0 - x_0$, and $\mathrm{SNR}(x_0, u) = 10\log_{10}(\sum_n x_0[n]^2 / \sum_n (u[n] - x_0[n])^2)$. For the EEG spectral metric, $\mathrm{PSD}(x) = |\mathrm{FFT}_{400}(x)|^2/400$. For EMG, $a_k(x) = |I_k|^{-1} \sum_{n\in I_k} |x[n]|$ is the mean rectified value on the k-th 1000-sample window and $\mathrm{MF}_m(x) = \sum_i f_i S_x[i,m] / \sum_i S_x[i,m]$ the mean frequency of frame m, with S_x the STFT magnitude over 10–500 Hz and A the active-stimulus frames (Wang et al., 2023).

Table 6: Evaluation metrics. The first block is reported for every modality; the remainder are the benchmark-specific metrics conventionally used in each literature.

Setting	Metric	Direction	Definition
All modalities	ΔSNR	$\uparrow$	$\mathrm{SNR}(x_0, \hat{x}_0) - \mathrm{SNR}(x_0, y)$
All modalities	PRD	$\downarrow$	$100\sqrt{\sum_n e[n]^2 / \sum_n (x_0[n] - \bar{x}_0)^2}$
All modalities	CC	$\uparrow$	$\frac{\sum_n (x_0[n]-\bar{x}_0)(\hat{x}_0[n]-\bar{\hat{x}}_0)}{\sqrt{\sum_n (x_0[n]-\bar{x}_0)^2 \sum_n (\hat{x}_0[n]-\bar{\hat{x}}_0)^2}}$
ECG	SSD	$\downarrow$	$\sum_n e[n]^2$
ECG	MAD	$\downarrow$	$\max_n \lvert e[n] \rvert$
ECG	CosSim	$\uparrow$	$\langle x_0, \hat{x}_0 \rangle / (\lVert x_0 \rVert_2 \lVert \hat{x}_0 \rVert_2)$
PPG	RMSE	$\downarrow$	$\sqrt{L^{-1} \sum_n e[n]^2}$
PPG	$\mathrm{PRD}_{\mathrm{PPG}}$	$\downarrow$	$100\sqrt{\sum_n e[n]^2 / \sum_n x_0[n]^2}$
EEG	RRMSE_t	$\downarrow$	$\sqrt{\mathrm{mean}(e^2)} / \sqrt{\mathrm{mean}(x_0^2)}$
EEG	RRMSE_s	$\downarrow$	Relative RMSE between $\mathrm{PSD}(\hat{x}_0)$ and $\mathrm{PSD}(x_0)$
EMG	RMSE	$\downarrow$	$\sqrt{L^{-1} \sum_n e[n]^2}$
EMG	ARV	$\downarrow$	$\sqrt{K^{-1} \sum_k (a_k(x_0) - a_k(\hat{x}_0))^2}$
EMG	MF	$\downarrow$	$\sqrt{\lvert A \rvert^{-1} \sum_{m \in A} (\mathrm{MF}_m(x_0) - \mathrm{MF}_m(\hat{x}_0))^2}$
Classification	AUROC	$\uparrow$	One-vs-rest AUC; macro: $C^{-1} \sum_c \mathrm{AUROC}_c$

C.5 Statistical Evaluation Procedure

Unless stated otherwise, reported values carry 95% bootstrap confidence intervals over 1,000 resamples (Strodthoff et al., 2021); Table 1 instead reports mean $\pm$ standard deviation over three training seeds (Section 4). Paired comparisons pair samples by test-set index, so each comparison uses the difference between PTSDiff and a baseline on the same signal. For metric m the paired advantage is $\Delta_i = m_i^{\mathrm{PTSDiff}} - m_i^{\mathrm{baseline}}$ where higher is better and the negation otherwise, so $\Delta_i > 0$ always favours PTSDiff. We test $\Delta_i > 0$ with a one-sided Wilcoxon signed-rank test (Woolson, 2008), chosen because it does not assume normally distributed differences, excluding zero differences by the standard procedure. Multiple PTSDiff–baseline comparisons are corrected by Holm–Bonferroni at $\alpha = 0.05$. Table 12 reports mean paired advantage and adjusted p-values; all comparisons favor PTSDiff and all remain significant after correction.

D Extended Results

D.1 Full ECG results across datasets

Table 7 gives all reported metrics on PTB-XL, QTDB and MIT-BIH, including the classical baselines omitted from the main text and the SSD/MAD/CosSim metrics used by prior ECG work (Li et al., 2024; 2026). PTSDiff is best on every metric and dataset except MAD on QTDB, where it ties TFCDiff.

D.2 Results stratified by corruption level

Section 5 summarizes the finding that PTSDiff's advantage increases under heavier corruption. Tables 8–10 stratify ECG performance by the corruption scaling factor λ and report both our main metrics and the metrics used by prior work. PTSDiff is best on every metric in every bin on every dataset, except MAD in the 0.2–0.6 and 0.6–1.0 bins. Figure 6 gives the corresponding stratification for the other three modalities. Figure 7 repeats this stratification with each benchmark's own metrics (Table 6): the ordering follows the main metrics on PPG and EEG, while ARV and mean-frequency error on EMG are mixed under heavy corruption.

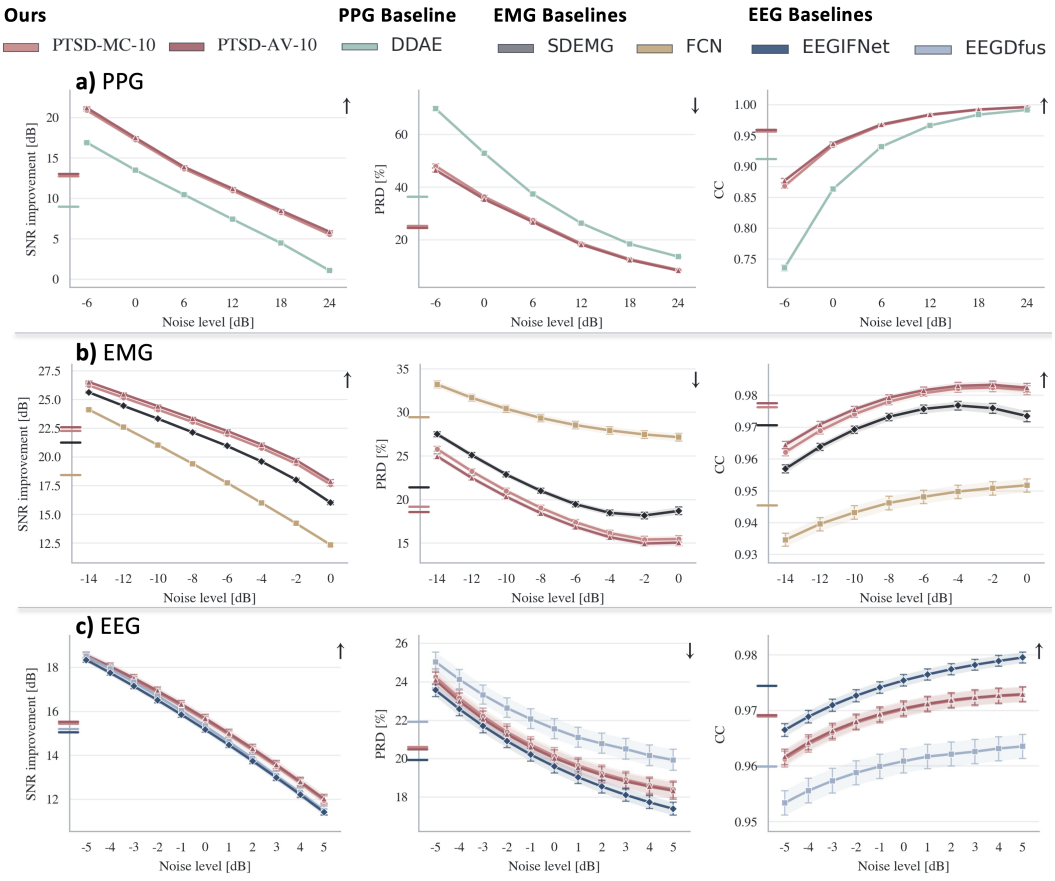


Figure 6: Restoration across corruption levels for PPG-DaLiA (a), NinaPro (b) and EEGdenoiseNet (c), grouped by each modality's corruption grid. The y-intercept marks the average per model and metric; arrows indicate the better direction.

Table 7: Denoising comparison across ECG datasets (PTB-XL, QTDB, MIT-BIH). Values are mean ± 95% CI. ↑ higher is better, ↓ lower is better. Bold and underlined values mark the best and second-best denoising method within each dataset and metric.

Dataset	Model	ΔSNR ↑	PRD [%] ↓	CC ↑	SSD ↓	MAD ↓	CosSim ↑
PTB-XL	Noisy ($SNR_{in} = -2.0$ dB)	0.00	152.86 ± 3.67	0.61 ± 0.01	210.66 ± 17.44	0.60 ± 0.02	0.60 ± 0.01
	FIR filter	7.44 ± 0.14	64.13 ± 1.44	0.83 ± 0.01	35.52 ± 2.61	0.36 ± 0.01	0.83 ± 0.01
	SWT denoising	6.22 ± 0.16	68.16 ± 1.19	0.76 ± 0.01	35.02 ± 2.16	0.35 ± 0.01	0.76 ± 0.01
	TCDAE	12.29 ± 0.17	34.03 ± 0.65	0.95 ± 0.00	8.71 ± 0.76	0.25 ± 0.01	0.95 ± 0.00
	DeScoD-1	11.82 ± 0.19	36.11 ± 0.70	0.92 ± 0.00	10.14 ± 0.97	0.26 ± 0.01	0.92 ± 0.00
	DeScoD-3	13.18 ± 0.19	31.04 ± 0.61	0.94 ± 0.00	7.73 ± 0.89	0.22 ± 0.01	0.94 ± 0.00
	DeScoD-5	13.49 ± 0.20	29.99 ± 0.59	0.94 ± 0.00	7.34 ± 1.01	0.21 ± 0.01	0.94 ± 0.00
	DeScoD-10	13.78 ± 0.20	29.05 ± 0.57	0.95 ± 0.00	6.91 ± 0.93	0.20 ± 0.01	0.95 ± 0.00
	TFCDiff-1	11.53 ± 0.18	36.95 ± 0.71	0.91 ± 0.00	11.38 ± 1.57	0.24 ± 0.01	0.91 ± 0.00
	TFCDiff-3	12.86 ± 0.18	31.98 ± 0.65	0.93 ± 0.00	9.16 ± 1.38	0.22 ± 0.01	0.93 ± 0.00
	TFCDiff-5	13.19 ± 0.19	30.86 ± 0.64	0.94 ± 0.00	8.61 ± 1.31	0.21 ± 0.01	0.94 ± 0.00
	TFCDiff-10	13.47 ± 0.19	29.91 ± 0.61	0.94 ± 0.00	8.30 ± 1.38	0.21 ± 0.01	0.94 ± 0.00
	PTSDIFF-MC-10	$\underline{15.79 \pm 0.21}$	$\underline{23.17 \pm 0.51}$	$\underline{0.97 \pm 0.00}$	$\underline{4.26 \pm 0.51}$	$\underline{0.15 \pm 0.01}$	$\underline{0.97 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{16.09 \pm 0.21}$	$\mathbf{22.43 \pm 0.50}$	$\mathbf{0.97 \pm 0.00}$	$\mathbf{4.00 \pm 0.50}$	$\mathbf{0.15 \pm 0.01}$	$\mathbf{0.97 \pm 0.00}$
QTDB	Noisy ($SNR_{in} = -0.8$ dB)	0.00	133.58 ± 0.89	0.65 ± 0.00	413.54 ± 10.84	0.80 ± 0.01	0.64 ± 0.00
	FIR filter	6.48 ± 0.04	60.04 ± 0.36	0.85 ± 0.00	79.42 ± 2.06	0.52 ± 0.01	0.85 ± 0.00
	SWT denoising	6.25 ± 0.04	59.43 ± 0.31	0.82 ± 0.00	76.15 ± 1.90	0.47 ± 0.00	0.81 ± 0.00
	TCDAE	10.33 ± 0.05	36.22 ± 0.16	0.95 ± 0.00	29.45 ± 1.09	0.40 ± 0.00	0.94 ± 0.00
	DeScoD-10	11.07 ± 0.06	33.91 ± 0.17	0.94 ± 0.00	42.21 ± 2.73	0.38 ± 0.00	0.93 ± 0.00
	TFCDiff-10	11.71 ± 0.06	31.85 ± 0.17	0.94 ± 0.00	46.48 ± 3.25	$\mathbf{0.33 \pm 0.00}$	0.94 ± 0.00
	PTSDIFF-MC-10	$\underline{13.05 \pm 0.06}$	$\underline{27.01 \pm 0.14}$	$\underline{0.96 \pm 0.00}$	$\underline{15.38 \pm 0.50}$	0.33 ± 0.00	$\underline{0.96 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{13.23 \pm 0.06}$	$\mathbf{26.51 \pm 0.14}$	$\mathbf{0.96 \pm 0.00}$	$\mathbf{14.84 \pm 0.49}$	$\underline{0.33 \pm 0.00}$	$\mathbf{0.96 \pm 0.00}$
MIT-BIH	Noisy ($SNR_{in} = -1.7$ dB)	0.00	148.07 ± 2.67	0.62 ± 0.01	921.71 ± 48.68	1.27 ± 0.03	0.61 ± 0.01
	FIR filter	6.82 ± 0.11	64.72 ± 1.05	0.83 ± 0.00	172.43 ± 9.05	0.81 ± 0.02	0.83 ± 0.00
	SWT denoising	6.48 ± 0.12	64.42 ± 0.92	0.78 ± 0.00	170.17 ± 8.78	0.75 ± 0.02	0.78 ± 0.00
	TCDAE	11.21 ± 0.13	37.10 ± 0.52	0.93 ± 0.00	58.07 ± 3.42	0.65 ± 0.02	0.93 ± 0.00
	DeScoD-10	11.25 ± 0.15	37.75 ± 0.55	0.91 ± 0.00	73.22 ± 4.55	0.64 ± 0.02	0.91 ± 0.00
	TFCDiff-10	11.21 ± 0.15	37.85 ± 0.58	0.91 ± 0.00	79.97 ± 5.19	0.60 ± 0.02	0.91 ± 0.00
	PTSDIFF-MC-10	$\underline{13.68 \pm 0.16}$	$\underline{28.53 \pm 0.47}$	$\underline{0.95 \pm 0.00}$	$\underline{37.22 \pm 3.08}$	$\underline{0.55 \pm 0.02}$	$\underline{0.95 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{13.88 \pm 0.16}$	$\mathbf{27.93 \pm 0.47}$	$\mathbf{0.95 \pm 0.00}$	$\mathbf{35.86 \pm 3.00}$	$\mathbf{0.55 \pm 0.02}$	$\mathbf{0.95 \pm 0.00}$

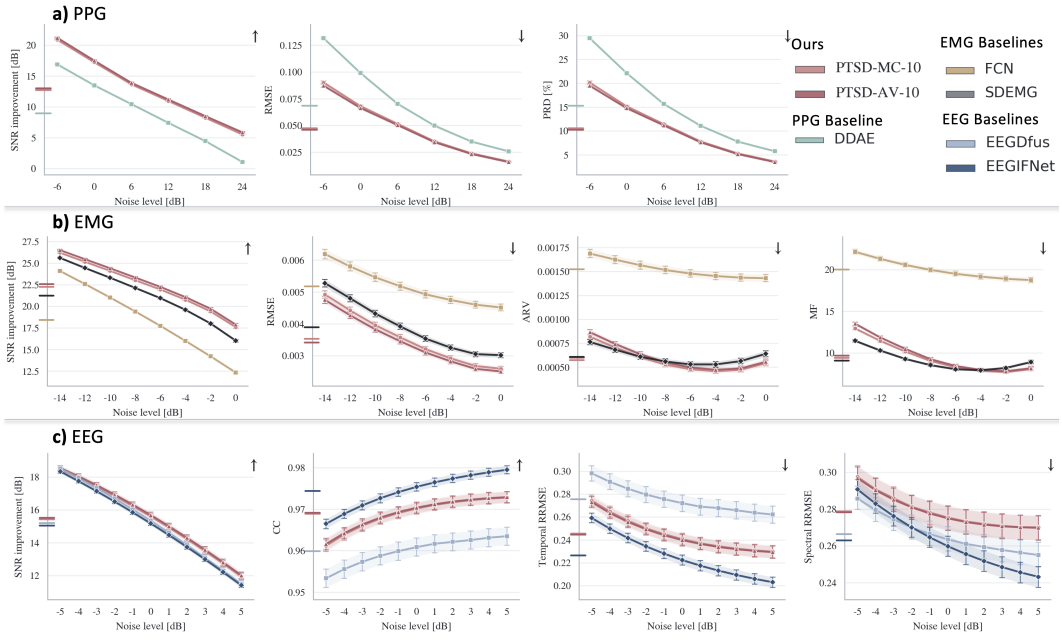


Figure 7: Restoration across corruption levels using each benchmark's own metrics. **(a)** PPG-DaLiA (ΔSNR, RMSE, PRD), following Lai et al. (2025). **(b)** NinaPro (ΔSNR, RMSE, ARV, MF), following Liu et al. (2024). **(c)** EEGdenoiseNet (ΔSNR, CC, temporal and spectral RRMSE), following Huang et al. (2025). On PPG and EEG the ordering follows the main metrics; on EMG, ARV and mean-frequency error are more mixed under heavy corruption, consistent with trajectory averaging mildly smoothing amplitude and spectral summaries (Appendix B.2).

Table 8: Denoising comparison by noise level on PTB-XL. Values are mean ± 95% CI. ↑ higher is better, ↓ lower is better. Bold marks the best and underline the second best within each bin and metric.

Noise level	Model	ΔSNR ↑	PRD [%] ↓	CC ↑	SSD ↓	MAD ↓	CosSim ↑
0.2-0.6	Noisy	0.00	55.45 ± 2.05	0.87 ± 0.01	22.48 ± 2.70	0.21 ± 0.01	0.87 ± 0.01
	FIR filter	6.79 ± 0.28	25.28 ± 0.92	0.96 ± 0.00	4.68 ± 0.63	0.13 ± 0.01	0.96 ± 0.00
	SWT denoising	3.04 ± 0.38	38.20 ± 1.12	0.92 ± 0.01	9.88 ± 1.57	0.17 ± 0.01	0.92 ± 0.01
	TCDAE	8.14 ± 0.34	21.25 ± 0.67	0.98 ± 0.00	2.66 ± 0.32	0.13 ± 0.01	0.98 ± 0.00
	DeScoD-1	7.49 ± 0.35	23.23 ± 0.88	0.97 ± 0.00	3.73 ± 0.60	0.12 ± 0.01	0.97 ± 0.00
	DeScoD-3	8.83 ± 0.37	20.08 ± 0.80	0.98 ± 0.00	3.00 ± 0.59	0.11 ± 0.01	0.98 ± 0.00
	DeScoD-5	9.13 ± 0.38	19.45 ± 0.80	0.98 ± 0.00	2.75 ± 0.53	0.10 ± 0.01	0.98 ± 0.00
	DeScoD-10	9.39 ± 0.38	18.94 ± 0.79	0.98 ± 0.00	2.63 ± 0.51	0.10 ± 0.01	0.98 ± 0.00
	TFCDiff-1	7.25 ± 0.32	23.76 ± 0.89	0.97 ± 0.00	4.20 ± 0.74	0.13 ± 0.01	0.97 ± 0.00
	TFCDiff-3	8.62 ± 0.34	20.55 ± 0.88	0.97 ± 0.00	3.56 ± 0.79	0.12 ± 0.01	0.97 ± 0.00
	TFCDiff-5	8.99 ± 0.34	19.75 ± 0.85	0.98 ± 0.00	3.33 ± 0.71	0.12 ± 0.01	0.98 ± 0.00
	TFCDiff-10	9.30 ± 0.34	19.09 ± 0.83	0.98 ± 0.00	3.03 ± 0.64	0.11 ± 0.01	0.98 ± 0.00
	PTSDIFF-MC-10	$\underline{10.99 \pm 0.37}$	$\underline{15.67 \pm 0.63}$	$\underline{0.99 \pm 0.00}$	$\underline{1.55 \pm 0.19}$	$\underline{0.08 \pm 0.00}$	$\underline{0.99 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{11.29 \pm 0.37}$	$\mathbf{15.17 \pm 0.60}$	$\mathbf{0.99 \pm 0.00}$	$\mathbf{1.44 \pm 0.18}$	$\mathbf{0.07 \pm 0.00}$	$\mathbf{0.99 \pm 0.00}$
0.6-1.0	Noisy	0.00	113.84 ± 3.56	0.68 ± 0.01	100.30 ± 14.05	0.44 ± 0.02	0.67 ± 0.01
	FIR filter	7.64 ± 0.30	47.50 ± 1.48	0.90 ± 0.01	17.41 ± 2.44	0.26 ± 0.02	0.90 ± 0.01
	SWT denoising	6.14 ± 0.30	55.08 ± 1.31	0.83 ± 0.01	20.16 ± 2.15	0.27 ± 0.01	0.83 ± 0.01
	TCDAE	12.07 ± 0.31	28.29 ± 0.90	0.97 ± 0.00	5.23 ± 0.91	0.19 ± 0.01	0.97 ± 0.00
	DeScoD-1	11.50 ± 0.31	30.46 ± 1.01	0.95 ± 0.00	6.56 ± 1.08	0.19 ± 0.02	0.95 ± 0.00
	DeScoD-3	12.89 ± 0.33	26.11 ± 0.90	0.96 ± 0.00	5.00 ± 0.90	0.17 ± 0.01	0.96 ± 0.00
	DeScoD-5	13.15 ± 0.33	25.36 ± 0.90	0.96 ± 0.00	4.63 ± 0.79	0.16 ± 0.01	0.96 ± 0.00
	DeScoD-10	13.47 ± 0.34	24.48 ± 0.86	0.97 ± 0.00	4.37 ± 0.77	0.16 ± 0.01	0.97 ± 0.00
	TFCDiff-1	10.86 ± 0.29	32.63 ± 1.06	0.94 ± 0.00	7.70 ± 1.24	0.20 ± 0.01	0.94 ± 0.00
	TFCDiff-3	12.17 ± 0.29	28.22 ± 0.97	0.95 ± 0.00	6.11 ± 1.12	0.19 ± 0.01	0.95 ± 0.00
	TFCDiff-5	12.51 ± 0.31	27.18 ± 0.96	0.96 ± 0.00	5.81 ± 1.14	0.18 ± 0.01	0.96 ± 0.00
	TFCDiff-10	12.78 ± 0.31	26.38 ± 0.94	0.96 ± 0.00	5.39 ± 0.99	0.18 ± 0.01	0.96 ± 0.00
	PTSDIFF-MC-10	$\underline{15.15 \pm 0.37}$	$\underline{20.30 \pm 0.76}$	$\underline{0.98 \pm 0.00}$	$\underline{2.86 \pm 0.53}$	$\underline{0.12 \pm 0.01}$	$\underline{0.98 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{15.40 \pm 0.37}$	$\mathbf{19.79 \pm 0.75}$	$\mathbf{0.98 \pm 0.00}$	$\mathbf{2.75 \pm 0.52}$	$\mathbf{0.11 \pm 0.01}$	$\mathbf{0.98 \pm 0.00}$
1.0-1.5	Noisy	0.00	172.88 ± 4.28	0.53 ± 0.01	224.67 ± 25.78	0.68 ± 0.03	0.52 ± 0.01
	FIR filter	7.63 ± 0.25	71.84 ± 1.76	0.81 ± 0.01	39.18 ± 4.34	0.40 ± 0.02	0.81 ± 0.01
	SWT denoising	7.21 ± 0.23	74.04 ± 1.46	0.73 ± 0.01	38.98 ± 3.84	0.38 ± 0.01	0.73 ± 0.01
	TCDAE	13.78 ± 0.25	35.68 ± 0.94	0.95 ± 0.00	8.65 ± 1.03	0.26 ± 0.01	0.95 ± 0.00
	DeScoD-1	13.27 ± 0.28	38.29 ± 1.12	0.91 ± 0.01	10.77 ± 1.39	0.28 ± 0.02	0.91 ± 0.01
	DeScoD-3	14.64 ± 0.29	32.82 ± 0.98	0.94 ± 0.00	7.99 ± 1.00	0.24 ± 0.01	0.94 ± 0.00
	DeScoD-5	14.94 ± 0.29	31.73 ± 0.96	0.94 ± 0.00	7.54 ± 0.98	0.23 ± 0.01	0.94 ± 0.00
	DeScoD-10	15.27 ± 0.29	30.63 ± 0.95	0.94 ± 0.00	7.11 ± 0.97	0.22 ± 0.01	0.94 ± 0.00
	TFCDiff-1	12.94 ± 0.26	39.12 ± 1.07	0.91 ± 0.01	10.84 ± 1.31	0.25 ± 0.01	0.91 ± 0.01
	TFCDiff-3	14.27 ± 0.27	33.80 ± 0.97	0.93 ± 0.01	8.83 ± 1.29	0.23 ± 0.01	0.93 ± 0.01
	TFCDiff-5	14.62 ± 0.26	32.48 ± 0.93	0.94 ± 0.00	8.03 ± 1.15	0.22 ± 0.01	0.94 ± 0.00
	TFCDiff-10	14.86 ± 0.27	31.64 ± 0.93	0.94 ± 0.00	7.78 ± 1.12	0.22 ± 0.01	0.94 ± 0.00
	PTSDIFF-MC-10	$\underline{17.46 \pm 0.30}$	$\underline{24.12 \pm 0.86}$	$\underline{0.96 \pm 0.00}$	$\underline{4.51 \pm 0.73}$	$\underline{0.16 \pm 0.01}$	$\underline{0.96 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{17.79 \pm 0.31}$	$\mathbf{23.29 \pm 0.83}$	$\mathbf{0.97 \pm 0.00}$	$\mathbf{4.16 \pm 0.63}$	$\mathbf{0.16 \pm 0.01}$	$\mathbf{0.97 \pm 0.00}$
1.5-2.0	Noisy	0.00	240.37 ± 6.07	0.42 ± 0.01	434.06 ± 50.37	0.96 ± 0.04	0.41 ± 0.01
	FIR filter	7.56 ± 0.28	100.72 ± 2.61	0.70 ± 0.01	71.79 ± 6.82	0.57 ± 0.02	0.70 ± 0.01
	SWT denoising	7.80 ± 0.26	96.59 ± 2.13	0.60 ± 0.01	63.55 ± 5.65	0.52 ± 0.02	0.60 ± 0.01
	TCDAE	14.20 ± 0.23	47.22 ± 1.32	0.92 ± 0.00	16.61 ± 2.44	0.38 ± 0.02	0.92 ± 0.00
	DeScoD-1	13.99 ± 0.29	48.96 ± 1.40	0.86 ± 0.01	17.82 ± 2.83	0.40 ± 0.02	0.86 ± 0.01
	DeScoD-3	15.30 ± 0.30	42.24 ± 1.25	0.89 ± 0.01	13.73 ± 2.79	0.34 ± 0.02	0.89 ± 0.01
	DeScoD-5	15.65 ± 0.32	40.63 ± 1.26	0.90 ± 0.01	13.21 ± 3.18	0.33 ± 0.02	0.90 ± 0.01
	DeScoD-10	15.93 ± 0.32	39.43 ± 1.23	0.91 ± 0.01	12.42 ± 2.95	0.32 ± 0.02	0.91 ± 0.01
	TFCDiff-1	13.98 ± 0.28	48.91 ± 1.54	0.85 ± 0.01	20.84 ± 5.08	0.36 ± 0.02	0.85 ± 0.01
	TFCDiff-3	15.28 ± 0.29	42.49 ± 1.38	0.89 ± 0.01	16.69 ± 4.57	0.32 ± 0.02	0.89 ± 0.01
	TFCDiff-5	15.55 ± 0.29	41.18 ± 1.34	0.90 ± 0.01	15.88 ± 4.38	0.31 ± 0.02	0.90 ± 0.01
	TFCDiff-10	15.85 ± 0.30	39.81 ± 1.30	0.90 ± 0.01	15.54 ± 4.73	0.30 ± 0.02	0.90 ± 0.01
	PTSDIFF-MC-10	$\underline{18.31 \pm 0.32}$	$\underline{30.67 \pm 1.23}$	$\underline{0.94 \pm 0.00}$	$\underline{7.48 \pm 1.62}$	$\underline{0.23 \pm 0.02}$	$\underline{0.94 \pm 0.00}$
	PTSDIFF-AV-10	$\mathbf{18.63 \pm 0.33}$	$\mathbf{29.65 \pm 1.21}$	$\mathbf{0.95 \pm 0.00}$	$\mathbf{7.07 \pm 1.62}$	$\mathbf{0.23 \pm 0.02}$	$\mathbf{0.95 \pm 0.00}$

Table 9: Denoising comparison by noise level on QTDB. Values are mean ± 95% CI. ↑ higher is better, ↓ lower is better. Bold marks the best and underline the second best within each bin and metric.

Noise level	Model	ΔSNR ↑	PRD [%] ↓	CC ↑	SSD ↓	MAD ↓	CosSim ↑
0.2-0.6	Noisy	0.00	48.26 ± 0.49	0.90 ± 0.00	48.54 ± 2.29	0.29 ± 0.00	0.90 ± 0.00
	FIR filter	4.52 ± 0.09	27.97 ± 0.26	0.96 ± 0.00	15.92 ± 0.79	0.27 ± 0.01	0.96 ± 0.00
	SWT denoising	3.05 ± 0.10	32.84 ± 0.27	0.95 ± 0.00	23.19 ± 1.52	0.24 ± 0.00	0.94 ± 0.00
	TCDAE	5.02 ± 0.10	26.09 ± 0.21	0.98 ± 0.00	16.57 ± 1.51	0.28 ± 0.01	0.97 ± 0.00
	DeScoD-10	6.08 ± 0.11	23.66 ± 0.24	0.97 ± 0.00	24.91 ± 4.52	0.24 ± 0.01	0.97 ± 0.00
	TFCDiff-10	6.81 ± 0.11	22.32 ± 0.27	0.98 ± 0.00	32.81 ± 6.16	**0.21 ± 0.01**	0.97 ± 0.00
	PTSDIFF-MC-10	7.47 ± 0.11	20.05 ± 0.19	0.98 ± 0.00	8.04 ± 0.41	0.27 ± 0.01	0.98 ± 0.00
	PTSDIFF-AV-10	**7.62 ± 0.11**	**19.76 ± 0.19**	**0.98 ± 0.00**	**7.84 ± 0.40**	0.27 ± 0.01	**0.98 ± 0.00**
0.6-1.0	Noisy	0.00	97.02 ± 0.77	0.73 ± 0.00	189.20 ± 8.74	0.59 ± 0.01	0.73 ± 0.00
	FIR filter	6.51 ± 0.08	45.43 ± 0.32	0.91 ± 0.00	41.12 ± 1.88	0.40 ± 0.01	0.90 ± 0.00
	SWT denoising	6.14 ± 0.08	47.07 ± 0.31	0.89 ± 0.00	45.60 ± 2.35	0.37 ± 0.01	0.88 ± 0.00
	TCDAE	9.90 ± 0.08	30.94 ± 0.24	0.96 ± 0.00	22.39 ± 1.95	0.33 ± 0.01	0.96 ± 0.00
	DeScoD-10	10.54 ± 0.09	29.28 ± 0.27	0.96 ± 0.00	37.99 ± 6.11	0.31 ± 0.01	0.95 ± 0.00
	TFCDiff-10	11.05 ± 0.10	28.02 ± 0.31	0.96 ± 0.00	46.69 ± 8.23	**0.29 ± 0.01**	0.95 ± 0.00
	PTSDIFF-MC-10	12.26 ± 0.09	24.06 ± 0.23	0.97 ± 0.00	11.82 ± 0.63	0.30 ± 0.01	0.97 ± 0.00
	PTSDIFF-AV-10	**12.43 ± 0.10**	**23.63 ± 0.23**	**0.97 ± 0.00**	**11.45 ± 0.61**	0.30 ± 0.01	**0.97 ± 0.00**
1.0-1.5	Noisy	0.00	152.18 ± 1.06	0.58 ± 0.00	449.84 ± 17.38	0.91 ± 0.01	0.57 ± 0.00
	FIR filter	7.14 ± 0.07	66.49 ± 0.41	0.83 ± 0.00	84.66 ± 3.11	0.57 ± 0.01	0.82 ± 0.00
	SWT denoising	7.30 ± 0.06	64.76 ± 0.38	0.79 ± 0.00	81.28 ± 3.05	0.52 ± 0.01	0.79 ± 0.00
	TCDAE	12.14 ± 0.06	37.57 ± 0.25	0.94 ± 0.00	29.46 ± 1.94	0.41 ± 0.01	0.94 ± 0.00
	DeScoD-10	12.79 ± 0.08	35.51 ± 0.28	0.93 ± 0.00	43.71 ± 5.26	0.39 ± 0.01	0.93 ± 0.00
	TFCDiff-10	13.32 ± 0.08	33.64 ± 0.29	0.94 ± 0.00	46.71 ± 6.06	0.35 ± 0.01	0.94 ± 0.00
	PTSDIFF-MC-10	14.88 ± 0.08	28.13 ± 0.24	0.96 ± 0.00	15.45 ± 0.71	0.34 ± 0.01	0.95 ± 0.00
	PTSDIFF-AV-10	**15.05 ± 0.08**	**27.62 ± 0.24**	**0.96 ± 0.00**	**14.93 ± 0.70**	**0.34 ± 0.01**	**0.96 ± 0.00**
1.5-2.0	Noisy	0.00	213.44 ± 1.36	0.45 ± 0.00	856.33 ± 31.74	1.27 ± 0.02	0.45 ± 0.00
	FIR filter	7.37 ± 0.06	91.35 ± 0.58	0.74 ± 0.00	157.12 ± 6.07	0.77 ± 0.01	0.73 ± 0.00
	SWT denoising	7.87 ± 0.06	85.62 ± 0.51	0.69 ± 0.00	138.88 ± 5.34	0.69 ± 0.01	0.68 ± 0.00
	TCDAE	13.15 ± 0.06	47.31 ± 0.33	0.91 ± 0.00	45.53 ± 2.73	0.55 ± 0.01	0.90 ± 0.00
	DeScoD-10	13.82 ± 0.08	44.33 ± 0.34	0.89 ± 0.00	57.87 ± 5.62	0.53 ± 0.01	0.88 ± 0.00
	TFCDiff-10	14.61 ± 0.08	40.85 ± 0.36	0.91 ± 0.00	56.65 ± 5.84	0.44 ± 0.01	0.90 ± 0.00
	PTSDIFF-MC-10	16.38 ± 0.08	33.89 ± 0.33	0.94 ± 0.00	24.19 ± 1.55	0.40 ± 0.01	0.93 ± 0.00
	PTSDIFF-AV-10	**16.58 ± 0.08**	**33.16 ± 0.32**	**0.94 ± 0.00**	**23.22 ± 1.51**	**0.40 ± 0.01**	**0.93 ± 0.00**

Table 10: Denoising comparison by noise level on MIT-BIH. Values are mean ± 95% CI. ↑ higher is better, ↓ lower is better. Bold marks the best and underline the second best within each bin and metric.

Noise level	Model	ΔSNR ↑	PRD [%] ↓	CC ↑	SSD ↓	MAD ↓	CosSim ↑
0.2-0.6	Noisy	0.00	52.72 ± 1.51	0.88 ± 0.00	98.31 ± 9.28	0.44 ± 0.02	0.88 ± 0.01
	FIR filter	5.15 ± 0.26	28.56 ± 0.79	0.96 ± 0.00	27.96 ± 2.82	0.37 ± 0.02	0.96 ± 0.00
	SWT denoising	3.28 ± 0.29	34.95 ± 0.86	0.93 ± 0.00	49.65 ± 6.14	$\mathbf{0.36 \pm 0.02}$	0.93 ± 0.00
	TCDAE	6.63 ± 0.29	23.87 ± 0.60	0.97 ± 0.00	22.28 ± 3.97	0.41 ± 0.02	0.97 ± 0.00
	DeScoD-10	6.58 ± 0.31	24.96 ± 0.77	0.96 ± 0.00	26.46 ± 3.73	0.38 ± 0.02	0.96 ± 0.00
	TFCDiff-10	6.27 ± 0.33	26.59 ± 0.91	0.96 ± 0.00	40.82 ± 6.68	0.39 ± 0.03	0.96 ± 0.00
	PTSDIFF-MC-10	8.30 ± 0.30	20.22 ± 0.61	0.98 ± 0.00	14.20 ± 1.97	0.42 ± 0.03	0.98 ± 0.00
	PTSDIFF-AV-10	$\mathbf{8.50 \pm 0.30}$	$\mathbf{19.82 \pm 0.61}$	$\mathbf{0.98 \pm 0.00}$	$\mathbf{13.71 \pm 1.92}$	0.42 ± 0.03	$\mathbf{0.98 \pm 0.00}$
0.6-1.0	Noisy	0.00	108.02 ± 2.33	0.70 ± 0.01	400.45 ± 33.69	0.93 ± 0.03	0.69 ± 0.01
	FIR filter	6.90 ± 0.21	48.62 ± 0.95	0.89 ± 0.00	83.29 ± 7.07	0.63 ± 0.03	0.89 ± 0.00
	SWT denoising	6.44 ± 0.21	50.63 ± 0.93	0.86 ± 0.00	92.60 ± 7.87	0.58 ± 0.02	0.86 ± 0.00
	TCDAE	11.26 ± 0.22	29.96 ± 0.80	0.95 ± 0.00	37.98 ± 6.81	0.52 ± 0.03	0.95 ± 0.00
	DeScoD-10	10.81 ± 0.27	32.62 ± 1.02	0.94 ± 0.00	53.45 ± 8.82	0.53 ± 0.03	0.94 ± 0.00
	TFCDiff-10	10.61 ± 0.28	33.34 ± 1.06	0.94 ± 0.00	60.49 ± 10.16	0.51 ± 0.03	0.94 ± 0.00
	PTSDIFF-MC-10	13.17 ± 0.26	24.78 ± 0.79	0.96 ± 0.00	26.31 ± 5.83	0.47 ± 0.03	0.96 ± 0.00
	PTSDIFF-AV-10	$\mathbf{13.38 \pm 0.27}$	$\mathbf{24.25 \pm 0.78}$	$\mathbf{0.96 \pm 0.00}$	$\mathbf{25.45 \pm 5.80}$	$\mathbf{0.47 \pm 0.03}$	$\mathbf{0.96 \pm 0.00}$
1.0-1.5	Noisy	0.00	170.77 ± 3.19	0.54 ± 0.01	1029.43 ± 68.65	1.48 ± 0.04	0.53 ± 0.01
	FIR filter	7.38 ± 0.19	72.60 ± 1.42	0.80 ± 0.01	189.45 ± 13.19	0.91 ± 0.03	0.80 ± 0.01
	SWT denoising	7.54 ± 0.18	70.71 ± 1.22	0.75 ± 0.01	187.36 ± 13.55	0.84 ± 0.03	0.75 ± 0.01
	TCDAE	12.65 ± 0.17	40.04 ± 0.85	0.92 ± 0.00	62.99 ± 5.26	0.70 ± 0.03	0.92 ± 0.00
	DeScoD-10	12.71 ± 0.22	40.83 ± 0.95	0.90 ± 0.01	83.84 ± 8.08	0.71 ± 0.03	0.90 ± 0.01
	TFCDiff-10	12.73 ± 0.22	40.63 ± 0.97	0.91 ± 0.01	86.36 ± 9.15	0.66 ± 0.03	0.91 ± 0.01
	PTSDIFF-MC-10	15.38 ± 0.22	30.29 ± 0.82	0.94 ± 0.00	38.33 ± 3.98	0.58 ± 0.03	0.94 ± 0.00
	PTSDIFF-AV-10	$\mathbf{15.58 \pm 0.22}$	$\mathbf{29.67 \pm 0.81}$	$\mathbf{0.95 \pm 0.00}$	$\mathbf{36.92 \pm 3.94}$	$\mathbf{0.57 \pm 0.03}$	$\mathbf{0.94 \pm 0.00}$
1.5-2.0	Noisy	0.00	235.06 ± 4.32	0.42 ± 0.01	1898.28 ± 137.32	1.99 ± 0.06	0.41 ± 0.01
	FIR filter	7.51 ± 0.21	99.41 ± 1.88	0.71 ± 0.01	345.47 ± 24.24	1.20 ± 0.04	0.71 ± 0.01
	SWT denoising	8.00 ± 0.19	93.34 ± 1.66	0.64 ± 0.01	312.92 ± 23.36	1.10 ± 0.04	0.64 ± 0.01
	TCDAE	13.38 ± 0.16	50.73 ± 1.05	0.89 ± 0.01	98.09 ± 9.47	0.89 ± 0.03	0.88 ± 0.01
	DeScoD-10	13.88 ± 0.25	49.16 ± 1.12	0.85 ± 0.01	115.16 ± 11.91	0.87 ± 0.04	0.85 ± 0.01
	TFCDiff-10	14.11 ± 0.23	47.84 ± 1.16	0.86 ± 0.01	119.92 ± 13.99	0.78 ± 0.04	0.86 ± 0.01
	PTSDIFF-MC-10	16.67 ± 0.24	36.67 ± 1.11	0.92 ± 0.01	63.68 ± 9.14	0.70 ± 0.03	0.92 ± 0.01
	PTSDIFF-AV-10	$\mathbf{16.87 \pm 0.24}$	$\mathbf{35.88 \pm 1.10}$	$\mathbf{0.92 \pm 0.01}$	$\mathbf{61.27 \pm 8.72}$	$\mathbf{0.69 \pm 0.03}$	$\mathbf{0.92 \pm 0.01}$

D.3 Cross-dataset generalization

Figure 8 evaluates ECG models on MIT-BIH and QTDB with no dataset-specific retraining. PTSDiff-AV-10 achieves the highest ΔSNR and lowest PRD on both. On MIT-BIH, the diffusion baselines lose their advantage over the deterministic TCDAE while PTSDiff does not; we read this as evidence for beat-local rather than global rhythm conditioning, since arrhythmic beats violate a global period but leave a local phase field well-defined. Table 11 extends the comparison to the other modalities. On PPG (PPG-DaLiA to BIDMC) and EMG (NinaPro DB2 to DB3), PTSDiff-AV-10 improves on the strongest baseline in all three metrics, by 2.08 and 1.49 dB ΔSNR over DDAE and SDEMG, respectively. On EEG (EEGdenoiseNet to EEGMMIDB), EEGIFNet stays ahead on all three metrics, by 0.37 dB ΔSNR, consistent with EEG being the modality with the smallest in-domain margin (Table 1). PTSDiff does however outperform the EEGDfus baseline on all three metrics, by 2.02 dB ΔSNR, 13.54 PRD points and 0.105 CC. On this dataset, all restorers nonetheless perform markedly worse than in-domain, because they remove much of the clean EEG below 4 Hz, the band occupied by the EOG artefacts seen in training.

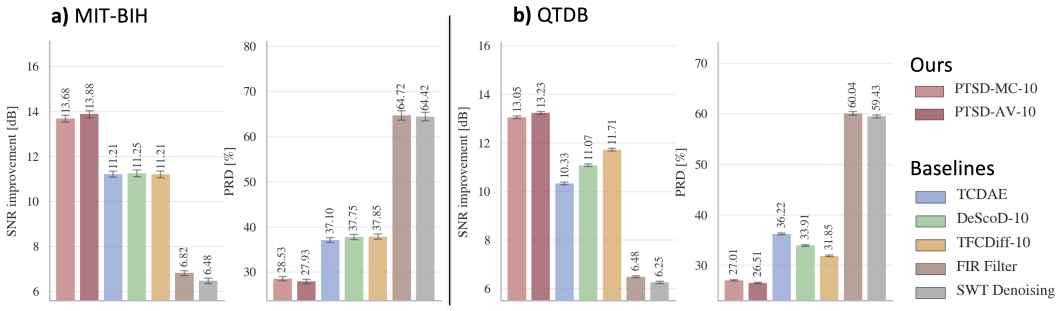


Figure 8: ECG Cross-dataset generalization to MIT-BIH (a) and QTDB (b) without retraining. Error bars are 95% CI.

Table 11: Dataset shift for all modalities: train on one dataset, evaluate on a held-out dataset with no retraining, against the strongest restoration baseline on the held-out set (named per row). Δ columns give the margin of PTSDiff-AV-10 over that baseline (positive = PTSDiff better) on each of the three main metrics.

Modality	Train	Held-out	Best baseline	ΔSNR ↑	Δ	PRD [%] ↓	Δ	CC ↑	Δ
PPG	PPG-DaLiA	BIDMC	DDAE	16.41	+2.08	32.06	+8.09	0.945	+0.048
ECG	PTB-XL	MIT-BIH	DeScoD-10	13.88	+2.63	27.93	+9.83	0.949	+0.039
ECG	PTB-XL	QTDB	TFCDiff-10	13.23	+1.52	26.51	+5.35	0.963	+0.019
EMG	NinaPro DB2	NinaPro DB3	SDEMG	23.45	+1.49	16.16	+3.20	0.985	+0.007
EEG	EEGdenoiseNet	EEGMMIDB	EEGIFNet	5.35	−0.37	56.18	−2.75	0.808	−0.029

D.4 Paired statistical comparison

Table 12 reports the paired tests of Appendix C.5 behind the ECG significance claims of Section 5. PTSDiff-AV-10 is significantly better than every learned and classical baseline on all three main metrics and all three ECG datasets after Holm–Bonferroni correction.

D.5 Extended downstream evaluation

Tables 13 and 14 give the per-class AUROC and the binary beat-classification results summarized in Section 5; Table 15 adds sensitivity, specificity and F1. The threshold-dependent metrics show that the AUROC gains are not a pure ranking effect, but also that PTSDiff is not uniformly best on sensitivity and specificity, so the downstream benefit depends on the operating point. The PTB-XL classifier is an Inception1d model (Strodthoff et al., 2021) trained on folds 1–8; MIT-BIH uses a CNN adapted from Tahsin et al. (2026) on ±144-sample windows around annotated R-peaks, collapsed to normal versus abnormal. Because the classifier is single-lead, its absolute performance is below 12-lead references (Strodthoff et al., 2021); the comparison across restoration methods is unaffected.

Table 12: Paired statistical comparison of PTSDIFF-AV-10 against denoising baselines across the main metrics. Each cell reports the mean paired difference (PTSDIFF-AV-10 − baseline) in the metric's native units, with the Holm–Bonferroni-adjusted p-value below. ΔSNR (dB) and CC are higher-is-better, PRD (percentage points) is lower-is-better; stars mark where PTSDIFF-AV-10 is significantly better.

Baseline	Metric	MIT-BIH	PTB-XL	QTDB
TFCDiff-10	ΔSNR	$+2.68^{***}$ $p_{adj} =< 1e-300$	$+2.62^{***}$ $p_{adj} =< 1e-300$	$+1.52^{***}$ $p_{adj} =< 1e-300$
	PRD [%]	-9.92^{***} $p_{adj} =< 1e-300$	-7.48^{***} $p_{adj} =< 1e-300$	-5.35^{***} $p_{adj} =< 1e-300$
	CC	$+0.035^{***}$ $p_{adj} =< 1e-300$	$+0.024^{***}$ $p_{adj} = 3.4e-290$	$+0.019^{***}$ $p_{adj} =< 1e-300$
TCDAE	ΔSNR	$+2.67^{***}$ $p_{adj} =< 1e-300$	$+3.80^{***}$ $p_{adj} =< 1e-300$	$+2.90^{***}$ $p_{adj} =< 1e-300$
	PRD [%]	-9.17^{***} $p_{adj} =< 1e-300$	-11.60^{***} $p_{adj} =< 1e-300$	-9.71^{***} $p_{adj} =< 1e-300$
	CC	$+0.019^{***}$ $p_{adj} =< 1e-300$	$+0.013^{***}$ $p_{adj} = 1.3e-263$	$+0.018^{***}$ $p_{adj} =< 1e-300$
DeScoD-10	ΔSNR	$+2.63^{***}$ $p_{adj} =< 1e-300$	$+2.30^{***}$ $p_{adj} =< 1e-300$	$+2.16^{***}$ $p_{adj} =< 1e-300$
	PRD [%]	-9.83^{***} $p_{adj} =< 1e-300$	-6.62^{***} $p_{adj} =< 1e-300$	-7.40^{***} $p_{adj} =< 1e-300$
	CC	$+0.039^{***}$ $p_{adj} =< 1e-300$	$+0.021^{***}$ $p_{adj} =< 1e-300$	$+0.028^{***}$ $p_{adj} =< 1e-300$
FIR Filter	ΔSNR	$+7.06^{***}$ $p_{adj} =< 1e-300$	$+8.65^{***}$ $p_{adj} =< 1e-300$	$+6.75^{***}$ $p_{adj} =< 1e-300$
	PRD [%]	-36.80^{***} $p_{adj} =< 1e-300$	-41.70^{***} $p_{adj} =< 1e-300$	-33.53^{***} $p_{adj} =< 1e-300$
	CC	$+0.12^{***}$ $p_{adj} =< 1e-300$	$+0.13^{***}$ $p_{adj} =< 1e-300$	$+0.11^{***}$ $p_{adj} =< 1e-300$
SWT Denoising	ΔSNR	$+7.40^{***}$ $p_{adj} =< 1e-300$	$+9.87^{***}$ $p_{adj} =< 1e-300$	$+6.98^{***}$ $p_{adj} =< 1e-300$
	PRD [%]	-36.50^{***} $p_{adj} =< 1e-300$	-45.73^{***} $p_{adj} =< 1e-300$	-32.92^{***} $p_{adj} =< 1e-300$
	CC	$+0.16^{***}$ $p_{adj} =< 1e-300$	$+0.21^{***}$ $p_{adj} =< 1e-300$	$+0.14^{***}$ $p_{adj} =< 1e-300$

$^{***}p_{\text{adj}} < 0.001$, $^{**}p_{\text{adj}} < 0.01$, $^{*}p_{\text{adj}} < 0.05$ after Holm–Bonferroni correction.

Table 13: Per-superdiagnostic-class and macro-averaged AUROC on PTB-XL. Values are point estimates with 95% bootstrap confidence intervals. Bold/underlined: best/second-best (excluding clean/noisy).

	Per-class AUROC					Macro
Model	CD	HYP	MI	NORM	STTC	AUROC
clean	.871 ± .019	.782 ± .031	.845 ± .020	.916 ± .012	.878 ± .017	.859 ± .011
noisy	.830 ± .022	.722 ± .034	.753 ± .024	.866 ± .015	.822 ± .020	.799 ± .012
FIR Filter	.828 ± .023	.729 ± .034	.758 ± .025	.855 ± .015	.818 ± .020	.798 ± .012
SWT Denoising	.788 ± .024	.716 ± .032	.658 ± .028	.798 ± .019	.772 ± .024	.746 ± .013
TCDAE	.855 ± .020	.760 ± .033	.829 ± .022	.902 ± .013	.853 ± .018	.840 ± .012
DeScoD	.854 ± .020	.769 ± .034	.826 ± .021	.901 ± .013	.853 ± .018	.840 ± .012
TFCDiff	.851 ± .021	.760 ± .034	.825 ± .022	.895 ± .013	.850 ± .019	.836 ± .012
PTSDIFF-MC-10	**.864 ± .020**	.774 ± .032	.836 ± .021	**.906 ± .013**	**.860 ± .018**	**.848 ± .011**
PTSDIFF-AV-10	.863 ± .020	**.775 ± .032**	**.836 ± .020**	.905 ± .012	.858 ± .018	.848 ± .011

Table 14: Downstream classification on MIT-BIH binary beat classification. Values are point estimates with 95% bootstrap confidence intervals. Bold/underlined: best/second-best (excluding clean/noisy).

Model	AUROC	Specificity	Sensitivity	F1
clean	.926 ± .004	.991 ± .001	.580 ± .014	.703 ± .012
noisy	.718 ± .008	.989 ± .001	.151 ± .010	.243 ± .014
FIR Filter	.734 ± .008	**.990 ± .001**	.157 ± .010	.254 ± .014
SWT Denoising	.722 ± .008	.971 ± .002	.206 ± .011	.286 ± .014
TCDAE	.839 ± .007	.979 ± .001	.436 ± .013	.543 ± .013
DeScoD	.822 ± .007	.919 ± .003	.541 ± .014	.490 ± .011
TFCDiff	.803 ± .007	.935 ± .002	.460 ± .013	.463 ± .012
PTSDIFF-MC-10	.880 ± .006	.965 ± .002	.579 ± .014	.622 ± .012
PTSDIFF-AV-10	**.883 ± .006**	.962 ± .002	**.602 ± .013**	**.630 ± .011**

The extended Table 16, introduced in Table 2 in Section 5, widens the comparison to PPG, EMG and EEG, with clean and corrupted references. As for ECG, the task models are trained on clean signals and applied frozen to clean, corrupted and restored test signals, with bootstrap intervals clustered by clean segment since each appears once per corruption. For PPG, an Inception1d regressor (Strodthoff et al., 2021) predicts the PPG-DaLiA reference heart rate (from the chest ECG, aligned to each 8 s segment) on the restoration subject split; we report MAE with Pearson r and R^2. For EMG, a small MLP on fixed amplitude features detects muscle activation, labelled from NinaPro's movement cue, in 1,024-sample windows of the DB2 test recordings (trained on subjects 1–8, tested on subject 10). Every input is divided by the root-mean-square of the clean training windows, and we report balanced accuracy since 71% of windows are active. For EEG, we compute α (8–13 Hz) and β (13–30 Hz) band power per epoch with Welch's method (Welch, 1967) and report the relative error to the clean epoch, averaged over both bands.

Table 15: Per-superdiagnostic-class and macro-averaged sensitivity, specificity, and F1 on PTB-XL. Values are point estimates with 95% bootstrap confidence intervals. Bold/underlined: best/second-best per row (excluding clean/noisy).

Class	clean	noisy	FIR Filter	SWT Denoising	TCDAE	DeScoD	TFCDiff	PTSDIFF-MC-10	PTSDIFF-AV-10
Sensitivity									
CD	.695 ± .041	.592 ± .043	.727 ± .040	**.775 ± .035**	.641 ± .043	.647 ± .041	.683 ± .041	.691 ± .039	.679 ± .041
HYP	.520 ± .066	.143 ± .046	.161 ± .050	.184 ± .053	.534 ± .066	**.592 ± .067**	.583 ± .064	.538 ± .065	.543 ± .065
MI	.662 ± .044	.827 ± .036	.880 ± .031	**.954 ± .021**	.600 ± .048	.585 ± .045	.624 ± .042	.621 ± .045	.624 ± .045
NORM	.892 ± .019	.446 ± .033	.425 ± .031	.166 ± .024	.893 ± .019	.880 ± .020	.860 ± .022	**.905 ± .019**	**.905 ± .018**
STTC	.669 ± .044	.862 ± .031	**.675 ± .043**	.669 ± .041	.614 ± .047	.650 ± .045	.638 ± .044	.610 ± .046	.606 ± .045
Macro	.688 ± .022	.574 ± .018	.574 ± .018	.550 ± .016	.656 ± .024	.671 ± .023	**.677 ± .023**	.673 ± .022	.671 ± .022
Specificity									
CD	.886 ± .015	.905 ± .014	.786 ± .019	.618 ± .024	**.899 ± .015**	.887 ± .015	.863 ± .017	.882 ± .015	.881 ± .015
HYP	.838 ± .017	.959 ± .009	**.962 ± .009**	.945 ± .011	.824 ± .018	.791 ± .019	.790 ± .019	.810 ± .018	.800 ± .019
MI	.827 ± .018	.499 ± .023	.438 ± .022	.189 ± .020	**.858 ± .017**	.855 ± .017	.838 ± .018	.852 ± .017	.852 ± .017
NORM	.789 ± .024	.939 ± .014	.935 ± .014	**.990 ± .006**	.755 ± .026	.783 ± .025	.781 ± .025	.765 ± .025	.760 ± .025
STTC	.894 ± .015	.571 ± .024	.787 ± .020	.731 ± .022	.880 ± .016	.867 ± .016	.862 ± .017	.895 ± .015	**.900 ± .015**
Macro	.847 ± .007	.775 ± .008	.781 ± .008	.695 ± .008	**.843 ± .007**	.837 ± .008	.827 ± .008	.841 ± .008	.838 ± .008
F1									
CD	.677 ± .035	.627 ± .036	.607 ± .035	.522 ± .031	.654 ± .034	.647 ± .035	.647 ± .034	**.671 ± .034**	.661 ± .033
HYP	.365 ± .048	.193 ± .058	.220 ± .061	.225 ± .058	.358 ± .048	**.358 ± .046**	.352 ± .044	.347 ± .047	.340 ± .045
MI	.566 ± .037	.436 ± .030	.430 ± .030	.371 ± .026	.556 ± .038	.543 ± .037	.551 ± .037	.564 ± .038	**.565 ± .039**
NORM	.836 ± .018	.588 ± .031	.567 ± .030	.283 ± .035	.821 ± .018	.826 ± .018	.814 ± .019	**.832 ± .018**	.830 ± .017
STTC	.672 ± .034	.544 ± .029	.581 ± .033	.538 ± .034	.621 ± .038	.633 ± .035	.620 ± .037	.633 ± .036	**.635 ± .036**
Macro	.623 ± .016	.478 ± .018	.481 ± .019	.388 ± .018	.602 ± .018	.601 ± .018	.597 ± .016	**.609 ± .018**	.606 ± .017

Table 16: Downstream utility across all four modalities.

Modality	Task (metric)	Clean	Corrupted	Best baseline	PTSDIFF-AV-10
PPG	HR estimation (MAE, bpm) ↓	4.772	8.483	5.947	**5.155**
ECG	PTB-XL superdiagnostic (macro AUROC) ↑	.859	.799	.840	**.848**
ECG	MIT-BIH beat classification (F1) ↑	.703	.243	.543	**.630**
EMG	NinaPro DB2 muscle activation (balanced acc.) ↑	.755	.501	.749	**.750**
EEG	Band-power preservation (α, β rel. error, %) ↓	0.00	4.75	4.56	**3.31**

D.6 Component ablation across modalities

Table 17 replicates the ECG ablation of Table 3 in every modality, retraining without one component at a time (Section 5). Removing phase costs most on PPG (-1.82 dB) and ECG (-0.64 dB) and little on EEG and EMG, whereas removing the frame costs 0.5–0.7 dB on ECG, EMG and EEG but slightly helps PPG.

Table 17: Component ablation replicated across modalities. Cells report the change in ΔSNR relative to the full model on the same dataset as mean $\pm$ 95% bootstrap CI. The analytical phase row is reported only for ECG, where the detector comparison is part of the main mechanism analysis. Values use one antithetic-variates pair.

Ablation (Δ from full model, dB)	PPG	ECG	EMG	EEG
− Phase conditioning	-1.82 ± 0.03	-0.64 ± 0.05	$+0.05 \pm 0.01$	-0.24 ± 0.01
− Temporal context	$+0.28 \pm 0.02$	-0.72 ± 0.07	-0.35 ± 0.01	-0.27 ± 0.01
− Shift-covariant frame	$+0.37 \pm 0.03$	-0.51 ± 0.05	-0.51 ± 0.01	-0.74 ± 0.02
− Auxiliary losses	-0.59 ± 0.02	-0.48 ± 0.05	$+0.02 \pm 0.01$	-2.80 ± 0.04
Analytical phase (vs. learned)	–	-0.66 ± 0.05	–	–

D.7 Cyclostationarity index across datasets

Section 5 relates the gain from phase conditioning to how strongly each signal is organized by its cycle. Table 18 measures this on the clean test sets, using the autocorrelation proxy Π_{AC} for all nine datasets and the detector-based $\hat{\Pi}$ for ECG (Appendix B.3). Averaged per modality, Π_{AC} orders the modalities PPG (0.74) > ECG (0.71) > EEG (0.37) > EMG (0.10), the same order as the cost of removing phase conditioning in Table 17.

Modality	Dataset	AC window [Hz]	Π_{AC} [95% CI]	$\hat{\Pi}$ [95% CI]
PPG	PPG-DaLiA	0.5–3	0.62 [0.60, 0.63]	–
	BIDMC†	0.5–3	0.85 [0.81, 0.89]	–
	Mean		0.74	–
ECG	PTB-XL	0.5–3	0.71 [0.70, 0.72]	0.87 [0.86, 0.87]
	QTDB†	0.5–3	0.81 [0.78, 0.83]	0.94 [0.92, 0.95]
	MIT-BIH†	0.5–3	0.61 [0.56, 0.66]	0.80 [0.70, 0.87]
	Mean		0.71	0.87
EEG	EEGdenoiseNet	1–30	0.34 [0.33, 0.35]	–
	EEGMMIDB†	1–30	0.40 [0.38, 0.42]	–
	Mean		0.37	–
EMG	NinaPro DB2	5–50	0.11 [0.10, 0.12]	–
	NinaPro DB3†	5–50	0.09 [0.09, 0.10]	–
	Mean		0.10	–

Table 18: Cyclostationarity of the clean test sets, ordered by the mean Π_{AC}, which is the autocorrelation proxy of Appendix B.3, computed identically for every dataset in a per-modality lag window fixed before analysis. $\hat{\Pi}$ (Definition 1) requires a reliable cycle detector and is reported for ECG only (Table 4). †Held-out datasets, evaluated without retraining (Table 11); *Mean* rows average the datasets of each modality. Because the proxy takes its peak at a single lag within each window, beat-to-beat rate variability lowers it, and it reads below $\hat{\Pi}$ on ECG. Brackets are 95% cluster-bootstrap CIs.

D.8 Synthetic Cyclostationarity Sweep

We isolate cyclostationarity from modality-specific confounders using synthetic signals composed of a fixed cyclic template and band-limited aperiodic background:

$$x_0 = \mathcal{N}\left(\sqrt{c}\, x_{\mathrm{cyc}} + \sqrt{1-c}\, x_{\mathrm{ap}}\right), \tag{25}$$

where $c \in \{0, 0.4, 0.6, 0.8, 0.95\}$ is the nominal cyclic variance share. Cycle intervals have mean P and coefficient of variation $\sigma_{\mathrm{jit}} \in \{0, 0.03, 0.07, 0.15, 0.30\}$. We train the full model and no-periodicity ablation at each of the 21 distinct conditions using three seeds. The measured gain is $\Delta\mathrm{SNR}_{\mathrm{full}} - \Delta\mathrm{SNR}_{\text{no-period}}$. Figure 3 shows that stronger measured cyclostationarity predicts a larger benefit from periodicity conditioning ($r = 0.85$ overall). The relationship also holds within fixed-share groups ($r = 0.82$–0.88), indicating that it is not driven solely by changing c. These results support the central prediction that periodicity conditioning becomes more useful as phase explains a larger fraction of signal structure. Curves are descriptive guides only and do not imply a particular functional relationship.

D.9 Frame and Objective Selection

Table 19 sweeps wavelet family and decomposition level; Table 20 sweeps the auxiliary loss weights. The frame choice matters more than the loss weighting: ΔSNR ranges from 14.58 (coif3, $J = 8$) to 15.98 (sym4, $J = 4$) across the frame sweep, against a narrower band across loss weights. Performance degrades consistently for $J > 4$ at every wavelet family, as deeper decompositions push the approximation band below the signal's fundamental and the coarsest channel stops carrying cycle-scale morphology. These sweeps support the frame of Section 3.2 ($J = 4$, `sym4`) and the loss weights of Equation 12.

D.10 Phase-Shuffle Experiment

Figure 9 gives the raw phase-shuffle measurements underlying Figure 2a. At evaluation time the phase-encoder outputs are permuted across samples within a batch while corrupted inputs are left intact, so the model receives a structurally valid but mismatched phase field. All four modalities show a positive drop, confirming that the conditioning channel is used rather than ignored.

The shuffled condition is not equivalent to removing phase conditioning. A mismatched field is actively misleading whereas an absent one is merely uninformative, so Δ_{shuf} upper-bounds the

Table 19: PTSDIFF SWT ablation on PTB-XL: ΔSNR (dB, higher better) and PRD (%, lower better) per (wavelet, decomposition level J), single antithetic-variate (AV) pair. Cells are mean ± 95% CI; best per metric in **bold**.

		Metric	Daubechies			Symlets			Coiflets		
			db_4	db_6	db_8	sym_4	sym_6	sym_8	$coif_3$	$coif_4$	$coif_5$
J	2	ΔSNR↑	15.57 ± 0.21	15.46 ± 0.20	15.16 ± 0.21	15.13 ± 0.21	15.57 ± 0.21	15.47 ± 0.20	15.21 ± 0.20	15.39 ± 0.21	15.38 ± 0.21
		PRD↓	24.21 ± 0.62	24.49 ± 0.58	25.80 ± 0.69	25.13 ± 0.57	24.01 ± 0.58	24.43 ± 0.60	24.79 ± 0.56	24.91 ± 0.64	24.54 ± 0.58
	4	ΔSNR↑	15.67 ± 0.21	15.50 ± 0.21	15.35 ± 0.21	$\mathbf{15.98 \pm 0.21}$	15.76 ± 0.21	15.60 ± 0.21	15.49 ± 0.21	15.43 ± 0.21	15.41 ± 0.21
		PRD↓	23.62 ± 0.56	24.09 ± 0.54	24.54 ± 0.56	$\mathbf{22.73 \pm 0.50}$	23.33 ± 0.52	24.14 ± 0.62	24.28 ± 0.58	24.33 ± 0.57	24.43 ± 0.56
	6	ΔSNR↑	15.07 ± 0.20	14.86 ± 0.21	14.91 ± 0.21	15.24 ± 0.21	14.93 ± 0.21	15.07 ± 0.20	14.86 ± 0.20	14.89 ± 0.20	14.95 ± 0.20
		PRD↓	25.20 ± 0.55	25.72 ± 0.55	25.52 ± 0.54	24.66 ± 0.54	25.70 ± 0.57	25.14 ± 0.54	25.77 ± 0.55	25.60 ± 0.55	25.36 ± 0.54
	8	ΔSNR↑	14.88 ± 0.20	14.71 ± 0.21	14.65 ± 0.21	14.85 ± 0.20	14.87 ± 0.20	14.80 ± 0.20	14.58 ± 0.21	14.67 ± 0.20	14.78 ± 0.20
		PRD↓	25.59 ± 0.52	26.10 ± 0.54	26.32 ± 0.56	25.65 ± 0.52	25.56 ± 0.53	25.78 ± 0.53	26.49 ± 0.54	26.19 ± 0.55	25.73 ± 0.52

Table 20: PTSDIFF auxiliary-loss weighting ablation on PTB-XL: ΔSNR (dB, higher better) and PRD (%, lower better) per (morphology, reconstruction) loss weight, single antithetic-variate (AV) pair. Cells are mean ± 95% CI; best per metric in **bold**. Ablations use the full model with a db4 wavelet and 4 decomposition levels.

		Metric	$\lambda_{\text{reconstruction}}$					
			0.05	0.1	0.2	0.3	0.5	1
$\lambda_{\text{morphology}}$	0.05	ΔSNR↑	15.61 ± 0.20	15.65 ± 0.21	15.57 ± 0.21	15.66 ± 0.20	15.74 ± 0.21	15.57 ± 0.21
		PRD↓	23.74 ± 0.53	23.70 ± 0.56	24.16 ± 0.62	23.78 ± 0.57	23.48 ± 0.54	24.03 ± 0.58
	0.1	ΔSNR↑	15.54 ± 0.21	15.57 ± 0.21	15.30 ± 0.21	$\mathbf{15.79 \pm 0.19}$	15.59 ± 0.21	15.00 ± 0.21
		PRD↓	23.98 ± 0.55	23.86 ± 0.54	24.54 ± 0.54	$\mathbf{23.29 \pm 0.54}$	23.84 ± 0.54	25.25 ± 0.54
	0.2	ΔSNR↑	15.47 ± 0.21	15.61 ± 0.22	15.60 ± 0.21	15.65 ± 0.21	15.66 ± 0.21	15.65 ± 0.21
		PRD↓	24.14 ± 0.52	23.79 ± 0.54	23.89 ± 0.56	23.87 ± 0.59	23.92 ± 0.60	23.68 ± 0.55
	0.3	ΔSNR↑	15.55 ± 0.21	15.66 ± 0.20	15.61 ± 0.21	15.13 ± 0.21	15.52 ± 0.20	15.70 ± 0.20
		PRD↓	23.96 ± 0.53	23.66 ± 0.55	23.87 ± 0.56	24.92 ± 0.54	24.11 ± 0.57	23.68 ± 0.57
	0.5	ΔSNR↑	15.55 ± 0.21	15.79 ± 0.20	15.64 ± 0.21	15.55 ± 0.20	15.65 ± 0.21	15.53 ± 0.20
		PRD↓	23.95 ± 0.55	23.30 ± 0.55	23.80 ± 0.56	24.05 ± 0.57	23.95 ± 0.60	24.18 ± 0.58
	1	ΔSNR↑	15.57 ± 0.21	15.72 ± 0.21	15.67 ± 0.21	15.61 ± 0.21	15.65 ± 0.21	15.62 ± 0.21
		PRD↓	23.91 ± 0.54	23.35 ± 0.52	23.65 ± 0.54	23.82 ± 0.56	23.77 ± 0.57	23.94 ± 0.58

ablation gap: $+4.14$ dB shuffled against 0.64 dB ablated on ECG (Table 3). It therefore measures reliance on the phase channel, the quantity plotted in Figure 2, and is available for all four modalities; Appendix D.6 instead shows the retrained ablation.

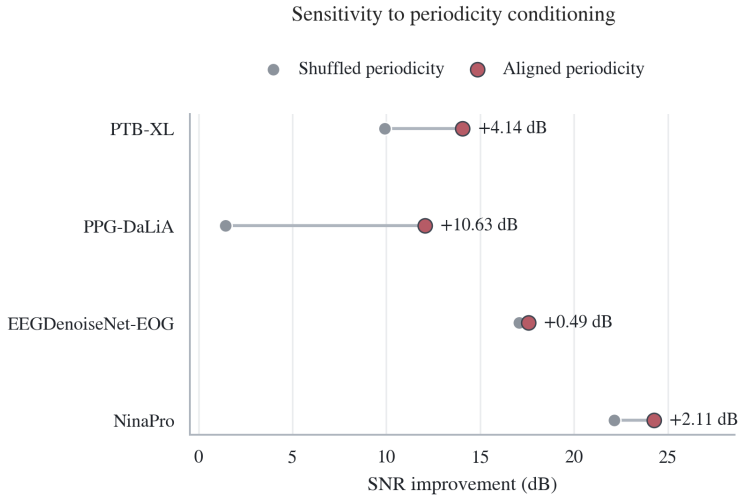


Figure 9: Change in restoration performance when phase-encoder features are shuffled across samples within a batch. Values are the ΔSNR difference between correctly matched and shuffled phase fields.

D.11 Antithetic sampling

Proposition 2 (proof in Appendix B.2) guarantees a variance reduction at matched NFE only when the pair correlation $\bar{\rho}$ is negative, which a learned nonlinear reverse process does not ensure, so we measure it directly. For each modality, we hold 32 corrupted observations fixed and draw 128 antithetic pairs per observation. At every reverse state, the covariance and marginal variance of the pair members are estimated across pairs separately for each observation, then aggregated over signal coordinates and observations with the variance weighting of Equation 16, giving $\widehat{\bar{\rho}}_t$. Confidence intervals come from a hierarchical bootstrap over observations and, within each, independent blocks of 16 pairs. The correlation remains negative throughout the reverse process and at $t = 0$ for every modality and fixed observation evaluated (Figure 10), so the condition of Proposition 2 holds. Figure 11 shows the resulting quality–compute frontier on ECG: a single antithetic pair already exceeds ten independent trajectories, and further pairs add little once the model-bias term of Equation 16 dominates. Table 21 applies the same sampler swap to DeScoD-ECG and TFCDiff without retraining: one antithetic pair exceeds ten independent trajectories for both at a fifth of the wall-clock time, so the gain comes from the sampler rather than from PTSDiff.

Table 21: Antithetic coupling applied to competing diffusion restorers without retraining (sampler swap only), with matched-NFE compute accounting. Establishes whether Proposition 2 is a property of the sampler or of PTSDiff.

Model / sampler	NFE	Params [M]	Wall-clock [ms]	ΔSNR $\uparrow$	Δ vs. MC
DeScoD-ECG, MC-10	500	1.93	361	13.78	—
DeScoD-ECG, AV-2	100	1.93	72	13.86	+0.08
TFCDiff, MC-10	500	4.12	803	13.47	—
TFCDiff, AV-2	100	4.12	161	13.67	+0.20
PTSDiff, MC-10	500	3.44	267	15.79	—
PTSDiff, AV-2	100	3.44	53	16.00	+0.21

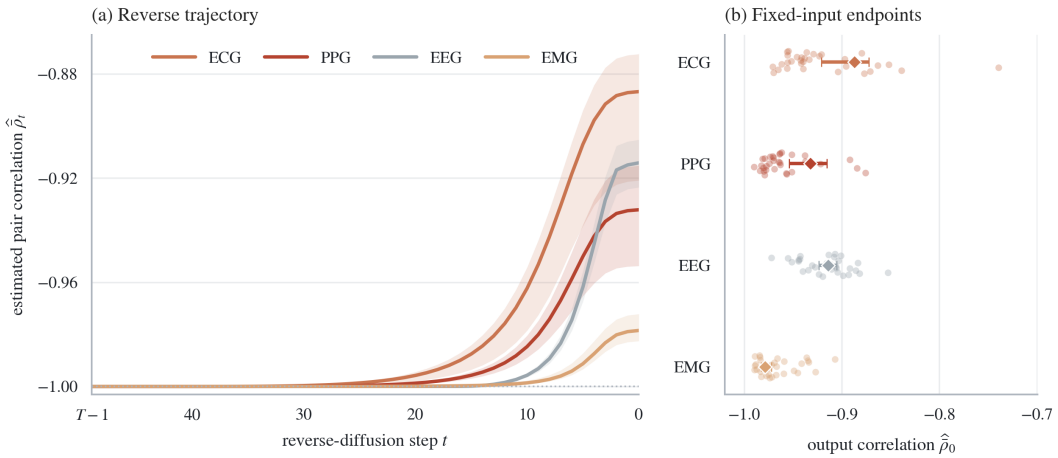


Figure 10: Empirical antithetic-pair correlation. **(a)** Variance-weighted correlation $\widehat{\rho}_t$ over the 50 post-update states $t = T-1, \ldots, 0$, with 95% confidence intervals. **(b)** Output correlation $\widehat{\rho}_0$: transparent points are single fixed observations, diamonds are modality aggregates with 95% confidence intervals.

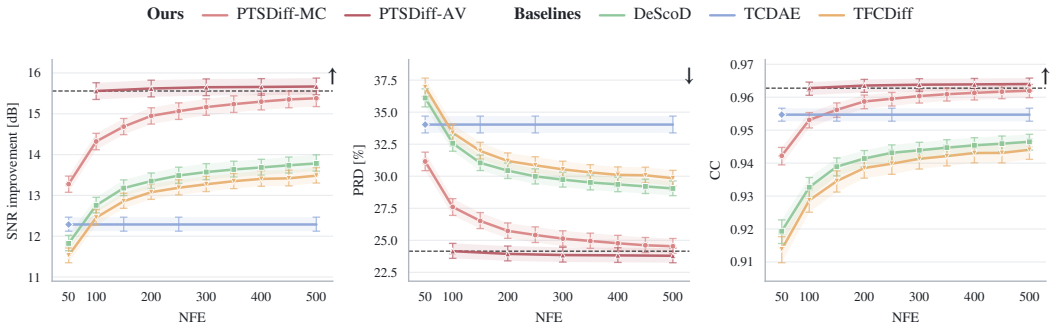


Figure 11: Quality–compute frontier on ECG. PRD, CC and ΔSNR against NFE, for independent (MC) and antithetically coupled (AV) sampling. The dotted line marks a single antithetic pair, which already exceeds ten independent trajectories. TCDAE is deterministic and therefore flat. Arrows indicate the better direction.

D.12 PER-SAMPLE DISTRIBUTION

Figure 12 complements the PTB-XL means of Table 7 with the full per-sample ΔSNR distribution, overall and stratified by corruption level, confirming that the mean improvements reflect a shifted distribution rather than a heavy tail.

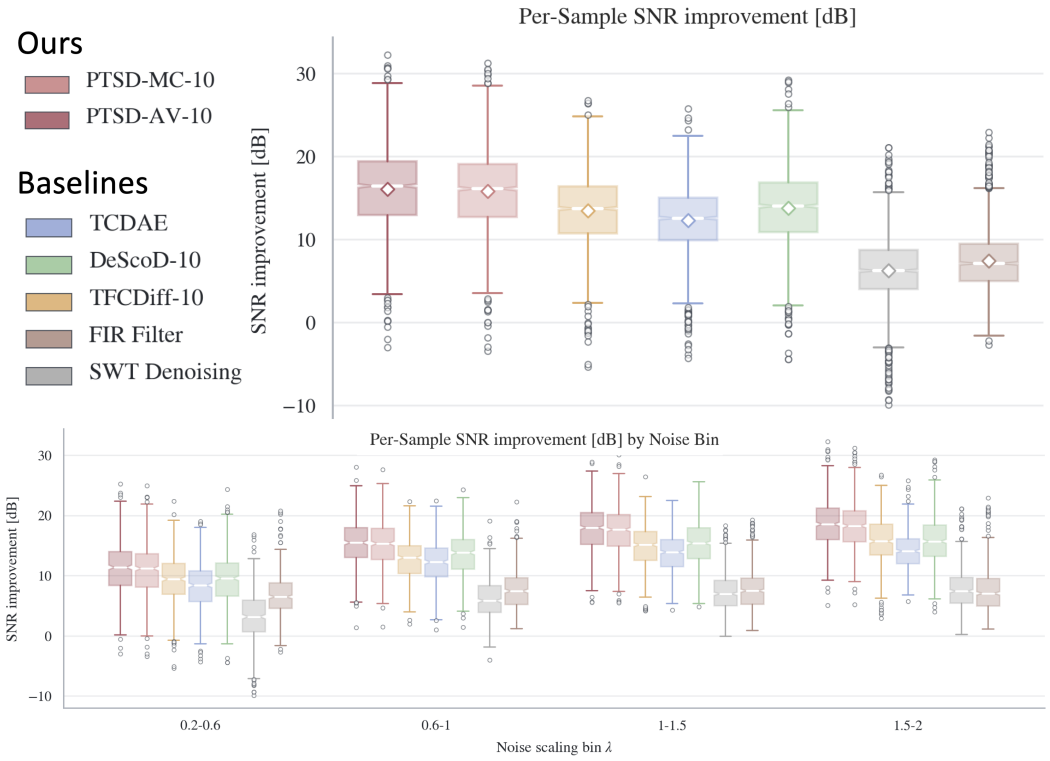


Figure 12: Per-sample ΔSNR on the PTB-XL test set. Notched box plots: box is the IQR, white line the median, notch its 95% CI, whiskers span 1.5×IQR, open circles are outliers. **Top:** overall (white diamond = mean). **Bottom:** stratified by corruption scaling factor λ.

D.13 THE FRAME DECOMPOSITION

Figure 13 illustrates the frame $\mathcal{W}$ of Section 3.2 on a clean and a corrupted ECG at 360 Hz. Baseline wander concentrates in the approximation channel a_4, while electrode motion and muscle artifacts spread across the detail channels (Chatterjee et al., 2020), which is why d_1 and d_2 carry predominantly artifact energy.

D.14 QUALITATIVE RESULTS

Figures 14–16 show representative reconstructions. These are single samples and should be read as illustration rather than evidence.

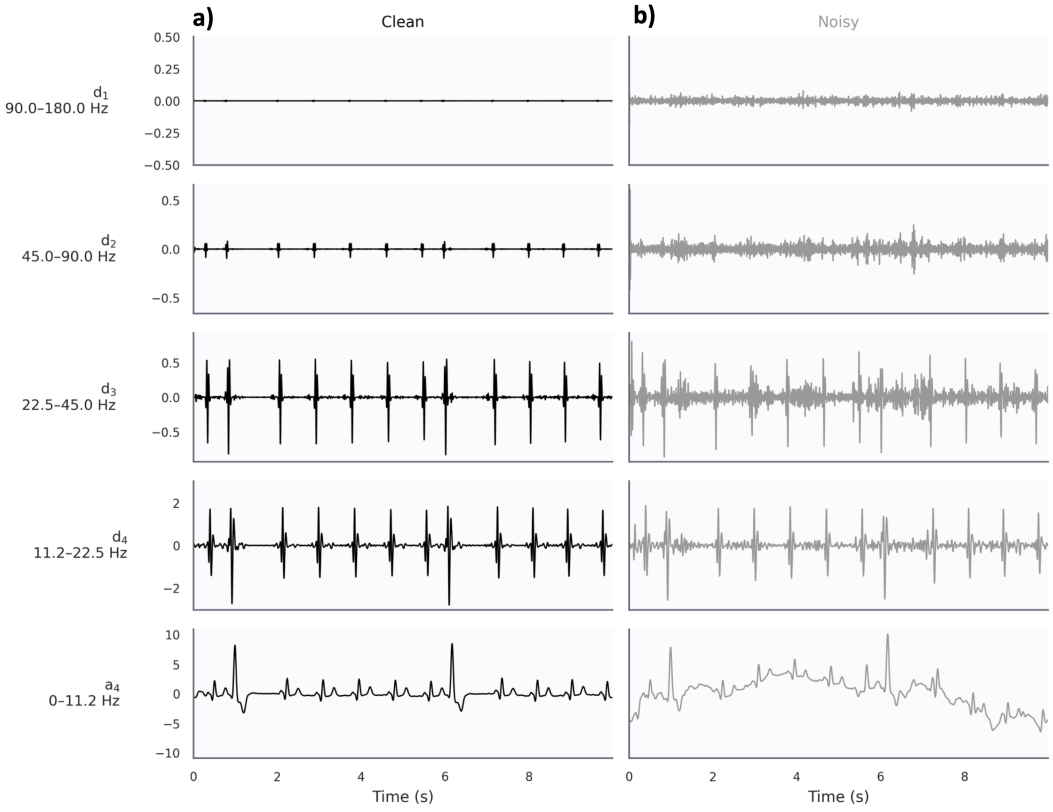


Figure 13: Four-level stationary wavelet decomposition with the `sym4` wavelet: one approximation and four detail channels, all at the input length. **(a)** Clean signal. **(b)** Corrupted signal.

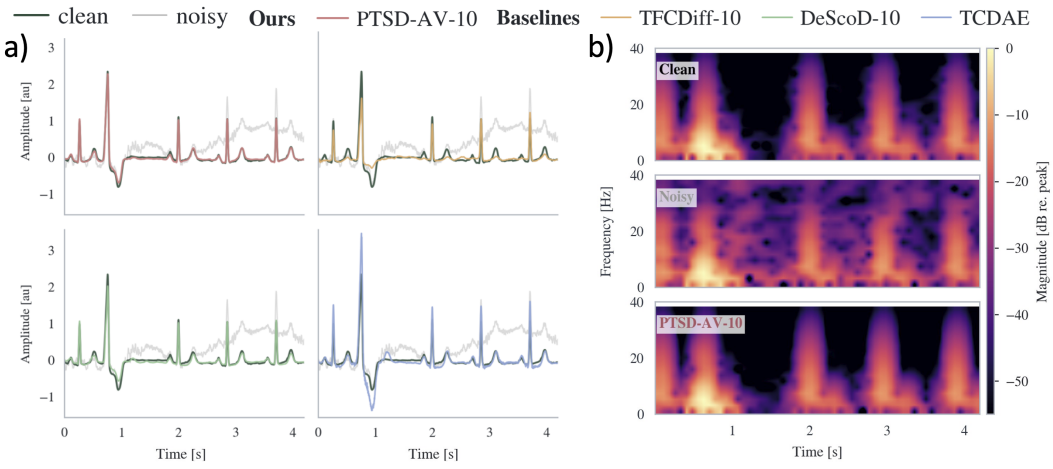


Figure 14: ECG under heavy corruption. **Left:** clean, corrupted, and reconstructions from PTSDIFF-AV-10 and the strongest baselines on a representative PTB-XL test sample. **Right:** time–frequency representations of the clean (top), corrupted (middle) and PTSDIFF-AV-10 reconstructed (bottom) signal.

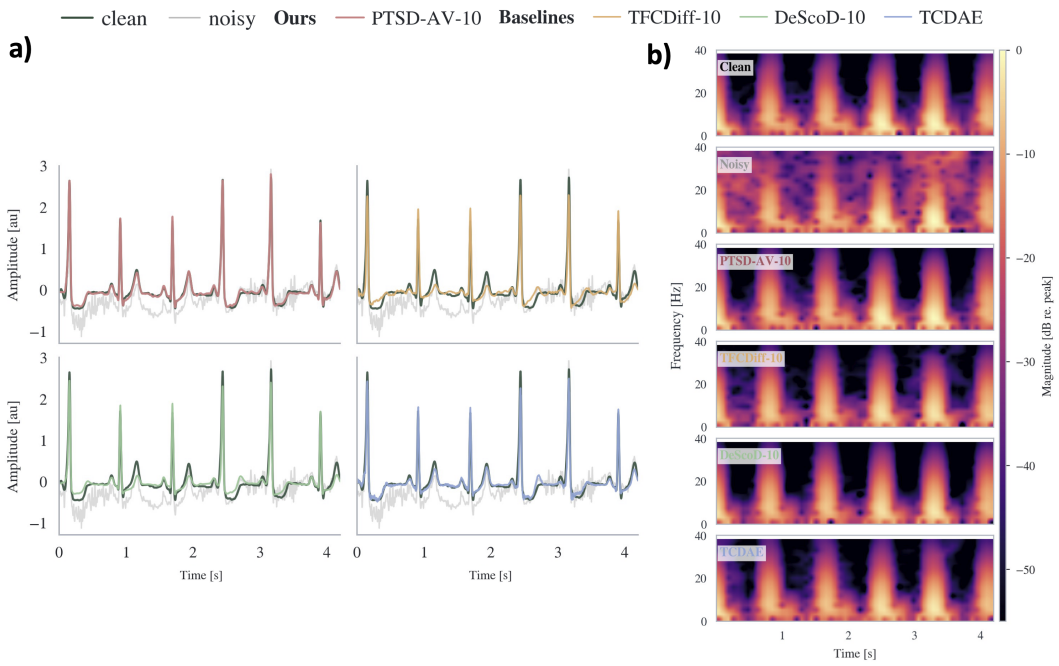


Figure 15: ECG under light corruption ($\lambda \in [0.2, 0.6]$). **(a)** Clean, corrupted, and reconstructions. **(b)** Corresponding spectrograms.

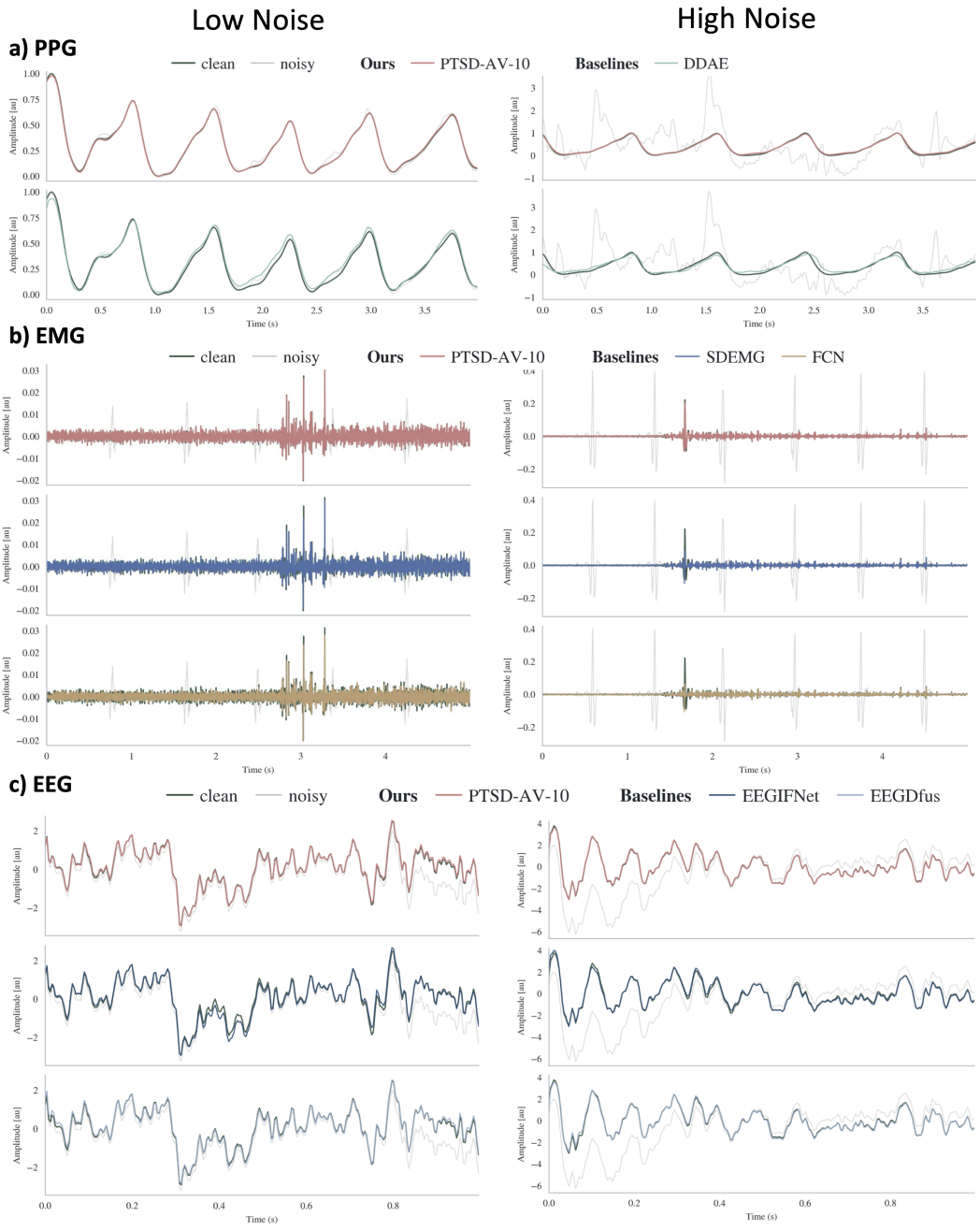


Figure 16: Reconstructions for the non-ECG modalities at a light and a heavy corruption level: **(a)** PPG (24 and -6 dB), **(b)** EMG (0 and -14 dB), **(c)** EEG (5 and -5 dB).