LEARNING TO ACCESS LATENT MEMORY: GENERALIZABLE ECG RESTORATION UNDER DISTRIBUTION SHIFT

Anonymous authors
Paper under double-blind review

ABSTRACT

Electrocardiogram (ECG) denoising is commonly learned through direct reconstruction from corrupted inputs to clean targets, which can couple restoration to training-specific signal–corruption relationships and lead to substantial degradation when signal or corruption characteristics shift at test time. To address this challenge, we treat clean ECGs not only as reconstruction targets, but also as a source of reusable latent structure that can be learned and retained independently of the corrupted-input mapping. We introduce a two-stage framework that first curates this clean structure in a key–value memory through clean-signal reconstruction, and then uses Memory Access Alignment (MAA) to align corrupted observations with the retained memory for reconstruction. During alignment, only the encoder bottleneck and memory router are adapted, while the curated memory and decoder remain frozen. Across multiple ECG datasets and both convolutional and Transformer-based backbones, the proposed framework consistently improves reconstruction under shifts in signal and corruption characteristics. Mechanistic analyses further show that closer agreement with the memory-access patterns induced by clean signals does not necessarily correspond to better reconstruction, while decoder-sensitive readout changes account for most of the improvement or deterioration induced by adaptation. These findings support learning effective access to retained clean structure as a strategy for robust restoration under distribution shift. Our code is available at `https://anonymous.4open.science/r/MAA_ECG_Denoising-A0D4/README.md`.

1 INTRODUCTION

Electrocardiography (ECG) provides a noninvasive measure of cardiac electrical activity (Jin et al. (2021); Li et al. (2024); Ma et al. (2023)), supporting cardiovascular screening and monitoring. However, ECG recordings are susceptible to baseline wander, muscle activity, and electrode motion (Zhang et al. (2025a)), which can distort the recorded waveforms. These distortions can obscure cardiac features or resemble abnormal activity Halvaei et al. (2021), making the underlying cardiac signal more difficult to interpret.

ECG denoising methods range from conventional filtering and signal decomposition (Van Alste & Schilder (1985); Kumar et al. (2021); Zhang et al. (2016); Ghafari et al. (2023); Kuang et al. (2020)) to deep learning approaches that learn restoration mappings from paired corrupted and clean recordings. Previous studies (Chiang et al. (2019); Chorney et al. (2023); Li et al. (2023); Zhang et al. (2025b)) have demonstrated effective reconstruction when training and evaluation conditions are closely matched. In practical deployment, however, recordings may come from different patient populations or acquisition settings and contain artifacts that differ from those represented during training. These shifts can degrade reconstruction performance when a model is applied to a new setting without additional training. For example, Figure 1 illustrates this degradation under a controlled clean-signal distribution shift: MITDB and QTDB exhibit distinct amplitude, spectral, and temporal characteristics, and, even under the same corruption source, both the CNN-based LUNet and Transformer-based ImputeECG reconstruct QTDB signals less accurately when trained on MITDB than when trained on QTDB. This degradation highlights the sensitivity of learned restoration mappings to the signal distributions represented during training. Therefore, maintaining performance

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.

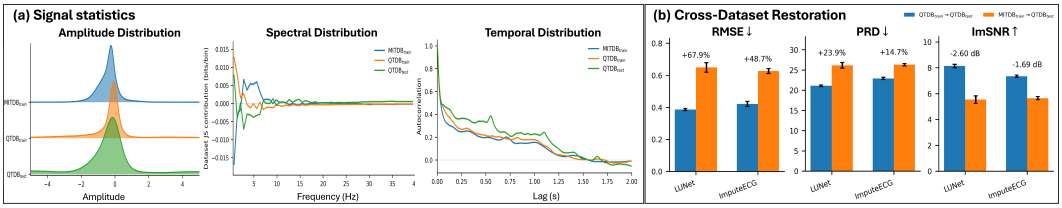

Figure 1: **ECG distribution differences and cross-dataset restoration under a shared noise source.** (a) Amplitude distributions, spectral differences, and temporal autocorrelation across the MITDB training, QTDB training, and QTDB test sets. (b) LUNet and ImputeECG evaluated on the same QTDB test set after training on either QTDB or MITDB. *Note:* Both datasets are corrupted using the same noise source.

under such zero-shot out-of-distribution conditions is an important requirement for practical ECG denoising, particularly when paired corrupted and clean recordings are unavailable for adaptation.

In supervised ECG denoising based on a direct reconstruction objective, clean recordings specify the desired output, while internal representations are learned through the mapping from corrupted inputs to clean targets. This objective does not explicitly require clean-signal structure to be learned and retained independently of that mapping. Clean recordings also provide direct observations of physiologically relevant structure, including heartbeat morphology and rhythm, that can be encoded into informative latent representations Li et al. (2026a); Jin et al. (2025). Recent ECG studies demonstrate that these representations can support reconstruction and synthesis. CardioRadar Lu et al. (2026) uses paired radar–ECG supervision to align radar observations with a learned ECG representation, whereas ECGEN Kranz et al. (2025) uses a learned ECG latent space for signal generation and completion. Beyond physiological signals, CodeFormer Zhou et al. (2022) demonstrates effective face restoration through a retained clean prior by adapting an encoder and learning discrete code prediction with clean-derived supervision, while keeping the codebook and decoder frozen. These advances establish the value of reusable clean representations and learned access to them. For ECG denoising, they motivate examining which objectives for learning access support restoration under distribution shift, and whether closer agreement with clean-derived representations reliably translates into better reconstruction.

We therefore introduce a two-stage framework that first curates clean ECG structure in a latent memory and then uses Memory Access Alignment (MAA) to learn how corrupted observations can effectively access this retained structure for reconstruction. Specifically, in the first stage, we encode clean signals, use their latent representations to access a key–value memory, and reconstruct the clean signals solely from the retrieved memory content. This reconstruction objective encourages the memory to capture reusable latent content that reflects clean ECG structure. In the second stage, MAA adapts the encoder bottleneck and memory router to enable corrupted signals to effectively access the memory curated from clean signals in Stage 1. This adaptation aligns the memory-access representations at the encoder bottleneck and the resulting memory readouts between corrupted and corresponding clean inputs, thereby guiding corrupted inputs toward memory retrieval that supports reconstruction of the corresponding clean signal. The memory and decoder remain frozen, restricting learning to the access mechanism rather than altering the curated memory or reconstruction pathway.

We evaluate the proposed framework across multiple ECG datasets and both convolutional and Transformer-based backbones, demonstrating improved restoration under shifts in signal and corruption characteristics. Our mechanistic analyses show that access adaptation changes the memory content retrieved from corrupted signals, and that changes along directions to which the frozen decoder is most sensitive account for most of the resulting improvement or deterioration in reconstruction. Our contributions can be summarized as follows:

- We formulate ECG denoising around retaining and reusing clean-signal structure, treating clean recordings not only as reconstruction targets but also as a source of latent reference representations for restoration under distribution shift.

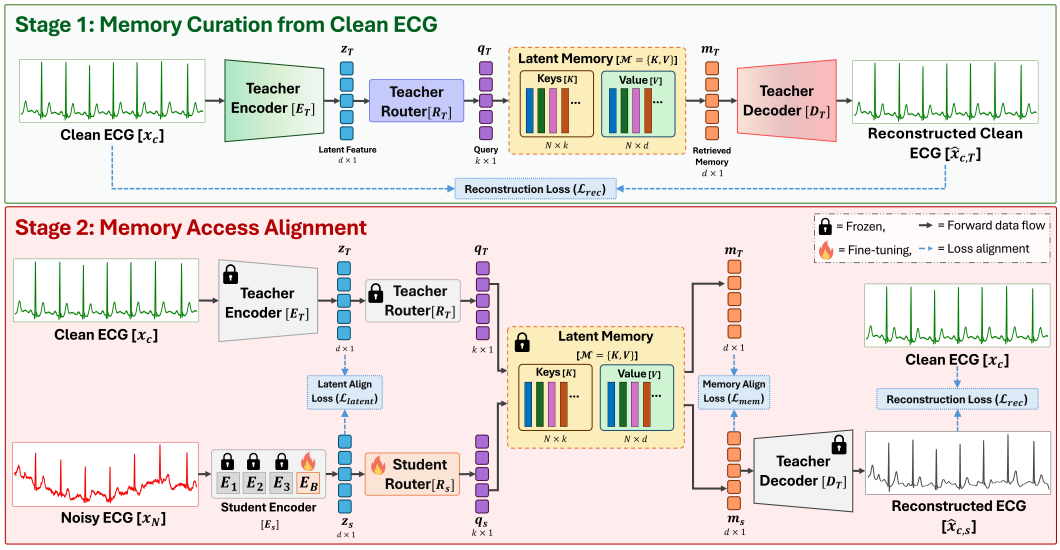

Figure 2: **Overview of clean-memory curation and memory access alignment. Stage 1:** Clean ECG signals guide the curation of a key–value memory through teacher reconstruction $\hat{x}_{c,T}$. **Stage 2:** Latent and retrieved-memory alignment transfers clean-teacher guidance to the noisy-input student. Only the student encoder bottleneck and router are adapted.

- We introduce a two-stage framework that first curates clean ECG structure in a key–value memory and then uses Memory Access Alignment to adapt how corrupted observations access this frozen memory, while keeping the memory and decoder unchanged.
- We demonstrate consistent restoration gains across the evaluated datasets and backbones, and show that clean-teacher access agreement alone does not determine reconstruction quality. Decoder-sensitive readout changes account for most adaptation-induced improvement or deterioration.

2 Learning to Access Clean Latent Memory

Given a corrupted ECG segment $x_N \in \mathbb{R}^{1\times T}$ and its corresponding clean reference $x_C \in \mathbb{R}^{1\times T}$, ECG denoising aims to learn a mapping $\hat{x}_C = f_\theta(x_N)$ that suppresses noise while preserving the underlying cardiac morphology. In this work, we introduce a key–value memory $\mathcal{M} = \{K, V\}$ to provide reusable latent content learned from clean ECG signals. As illustrated in Figure 2, our framework separates learning what the memory stores from learning how noisy observations access it. In **Stage 1**, clean-input reconstruction through the teacher pathway guides memory curation, establishing latent content for clean-signal reconstruction. In **Stage 2**, Memory Access Alignment (MAA) uses paired clean and noisy observations to align student representations and memory readouts with their clean-teacher references. The teacher, memory, and decoder remain frozen, allowing the student access pathway to adapt to corruption while preserving the curated content and its decoding function. At inference, the student accesses the fixed memory to reconstruct the ECG without clean-input guidance.

2.1 Memory Curation from the Clean ECG

We curate clean-signal latent content by requiring the model to reconstruct clean ECG signals exclusively from a shared, learnable key–value memory. Each clean signal is first encoded into a latent representation and projected into queries that are matched against the memory keys. These matches determine how the corresponding values are combined to form the retrieved representation, which serves as the sole input to the decoder for reconstructing the original clean signal. The memory, encoder, router, and decoder are jointly optimized, so reconstruction guides the values to encode useful latent content and the keys to make that content accessible through the clean-input queries.

The clean-input reconstruction model in this stage serves as the *teacher* for subsequent access alignment. We implement this model using LUNet and ImputeECG, instantiating the same memory-based reconstruction framework within CNN and Transformer architectures, respectively. For LUNet, we remove its encoder–decoder skip connections to prevent reconstruction from bypassing the memory retrieval. Let E_T, R_T, and D_T denote the teacher encoder, memory router, and decoder. For a clean ECG segment x_C, the encoder produces a latent representation $z_T = E_T(x_C)$. The router maps z_T to a sequence of queries $q_T = R_T(z_T)$ using a one-dimensional convolution with kernel size $w = 13$, incorporating local temporal context into each query for memory retrieval.

The queries access a shared, learnable key–value memory $\mathcal{M}$, which is defined as follows:

$$\mathcal{M} = \{(k_i, v_i)\}_{i=1}^{N}, \tag{1}$$

where N=48, $k_i \in \mathbb{R}^{32}$,and $v_i \in \mathbb{R}^{64}$. Each key is paired with a value and determines how strongly that value contributes to the readout for a given query. The keys therefore organize access to the memory, while the values store the latent content used for clean-signal reconstruction. For each query $q_{T,t}$ at latent temporal position t, we compute its cosine similarity with every memory key,

$$s_{t,i} = \frac{\langle q_{T,t}, k_i \rangle}{\|q_{T,t}\|_2 \|k_i\|_2}. \tag{2}$$

Retrieval is restricted to the $k = 8$ entries with the highest similarity scores, whose indices form the set $\mathcal{I}_t$. The selected scores are normalized using a softmax with temperature $\tau = 0.1$,

$$a_{t,i} = \frac{\exp(s_{t,i}/\tau)}{\sum_{j \in \mathcal{I}_t} \exp(s_{t,j}/\tau)}, \qquad i \in \mathcal{I}_t. \tag{3}$$

The corresponding values are combined to obtain the retrieved memory representation,

$$m_{T,t} = \sum_{i \in \mathcal{I}_t} a_{t,i} v_i. \tag{4}$$

Stacking the retrieved vectors $m_{T,t}$ in temporal order forms the memory readout sequence m_T. We denote the complete query-to-readout operation by $\mathcal{A}$, such that $m_T = \mathcal{A}(\mathcal{M}, q_T)$. The teacher decoder then maps this sequence to the reconstructed ECG, $\hat{x}_{C,T} = D_T(m_T)$.

We jointly optimize the teacher encoder, router, decoder, and memory using the clean ECG as the reconstruction target. The Stage 1 objective combines sample-wise reconstruction with normalized spectral-magnitude matching,

$$\mathcal{L}_{\text{rec}} = \text{MSE}(\hat{x}_{C,T}, x_C) + \text{L1}(\hat{x}_{C,T}, x_C) + 0.1\, \mathcal{L}_{\text{spectral}}(\hat{x}_{C,T}, x_C)\,. \tag{5}$$

Here, $\mathcal{L}_{\text{spectral}}$ measures the normalized mean absolute difference between the full-segment one-sided FFT magnitudes of the reconstructed and clean ECG signals.

After curation, the teacher and memory are frozen, providing the clean-input representations and readouts used to guide subsequent access alignment.

2.2 Memory Access Alignment (MAA)

Stage 2 learns how corrupted ECG observations access the clean latent memory curated in Stage 1. Corruption can shift the encoder representation and, consequently, the content retrieved from memory. MAA therefore uses paired clean-input teacher outputs to guide adaptation at two levels: latent alignment encourages z_S to approach z_T before memory access, while memory-readout alignment encourages m_S to approach m_T after retrieval. Throughout adaptation, the memory and decoder remain frozen, preserving the stored representations and their decoding function while the student learns to access them from corrupted observations.

For each paired noisy and clean segment (x_N, x_C), the frozen teacher processes x_C to provide the clean-input representation z_T and memory readout m_T defined in Stage 1. The student processes x_N using the same backbone architecture, with its encoder E_S and router R_S initialized from the corresponding Stage 1 teacher weights. Adaptation is restricted to the student encoder bottleneck E_B and router R_S. The remaining encoder blocks retain their teacher weights and remain frozen. The student forward pathway is

$$\begin{aligned} z_S &= E_S(x_N), & q_S &= R_S(z_S), \\ m_S &= \mathcal{A}(\mathcal{M}, q_S), & \hat{x}_{C,S} &= D_T(m_S), \end{aligned} \tag{6}$$

where $\mathcal{A}$ is the memory-access operation defined in Stage 1, $\mathcal{M}$ is the shared frozen memory, and D_T is the frozen teacher decoder.

Latent alignment. We align the corrupted-input student representation with its clean-input teacher reference:

$$\mathcal{L}_{\text{latent}} = \text{MSE}(z_S, z_T)\,. \tag{7}$$

This objective provides a direct training signal to E_B before query formation, encouraging the adapted encoder to produce representations consistent with those used to access the memory under clean input. Because it is evaluated at the encoder output, latent alignment constrains encoder adaptation without depending on the readout produced by the downstream router and memory-access operation.

Memory-readout alignment. Latent alignment alone does not determine the retrieved representation because the student router is also trainable. Even matching encoder outputs need not produce matching queries when the student and teacher routers differ. Query formation, key matching, top-k selection, and value aggregation can therefore yield a student readout that deviates from the clean-input reference. We additionally align the aggregated memory readouts:

$$\mathcal{L}_{\text{readout}} = \text{L1}(m_S, m_T)\,. \tag{8}$$

This objective supervises both E_B and R_S through the memory-access pathway, directly constraining the representation supplied to the frozen decoder. By targeting the aggregated readout, it does not explicitly require identical queries, routing weights, or selected memory slots. The two alignment objectives thus provide complementary supervision: latent alignment constrains the encoder representation before retrieval, whereas readout alignment constrains the result of encoder and router adaptation after retrieval.

Training objective. The clean-input teacher provides reference representations, but representation agreement is a surrogate for ECG restoration. We therefore also supervise $\hat{x}_{C,S}$ against the clean target x_C using the reconstruction criterion defined in Eq. 5, denoted by $\mathcal{L}^S_{\text{rec}}$. This term retains a direct denoising objective during access alignment by evaluating the student readout through the frozen decoder. The complete Stage 2 objective is

$$\mathcal{L}_{\text{Stage2}} = \mathcal{L}_{\text{latent}} + \mathcal{L}_{\text{readout}} + 0.1\,\mathcal{L}^S_{\text{rec}}. \tag{9}$$

Teacher outputs serve as alignment targets without gradient propagation through the teacher, and optimization updates only E_B and R_S. Thus, all three objectives act through the student access pathway without modifying the curated memory or its decoding function.

At inference, neither the clean reference nor the teacher encoder and router are required. The adapted student processes a corrupted ECG segment, retrieves a representation from the frozen memory, and passes this readout through D_T to obtain the restored signal $\hat{x}_{C,S}$.

3 Experiments

Dataset. We use MITDB (Moody & Mark (2001)) as the source domain, comprising 48 half-hour, two-lead recordings from 47 subjects sampled at 360 Hz, with a 70/15/15% record-level split for training, validation, and ID evaluation. QTDB (Laguna et al. (1997)) provides a cross-database OOD setting with 105 two-lead recordings sampled at 250 Hz. To ensure recording-level independence from MITDB, we exclude the 15 QTDB records originating from MITDB and evaluate on the remaining 90 records. SIMEMG (Atanasoski et al. (2023)) provides a complementary OOD setting with 110 paired clean and naturally EMG-contaminated single-lead recordings from 14 subjects at 500 Hz. MITDB and QTDB are corrupted with baseline-wander, muscle-artifact, and electrode-motion noise from NSTDB (Moody et al. (1984)), whereas SIMEMG is evaluated directly using its recorded clean/noisy pairs. Full preprocessing and corruption protocols are provided in Appendix A.1.

To assess transfer beyond ECG, we further evaluate EEG denoising using EEGDenoiseNet-EOG (Zhang et al. (2021)) for ID evaluation and the external Klados–Bamidis EEG/EOG dataset (Klados & Bamidis (2016)) for zero-shot OOD evaluation. EEGDenoiseNet-EOG provides 4,514 clean

EEG epochs and 3,400 EOG artifact epochs, from which 2-s, 512-sample segments at 256 Hz are constructed. Klados–Bamidis comprises 54 paired clean and semi-simulated EOG-contaminated, 19-channel EEG recordings from 27 subjects, each 30 s long and sampled at 200 Hz. Detailed EEG splitting, resampling, corruption, normalization, and window-generation procedures are provided in Appendix A.2.

Baselines. We compare MAA against six representative deep-learning baselines spanning major ECG denoising paradigms: the CNN-based LUNet (Hu et al. (2024)), transformer-based ImputeECG (Fang et al. (2026)), hybrid CNN–Transformer CS-TRANS (Peng et al. (2024)), GAN-based CAE-CGAN (Wang et al. (2022)), and diffusion-based EDDM (Li et al. (2025)) and TFCDiff (Li et al. (2026b)). To evaluate whether MAA generalizes across backbone architectures rather than depending on a particular denoiser, we instantiate it on both LUNet and ImputeECG.

Experimental Protocol. Unless otherwise specified, all experiments are conducted with three random seeds (42, 43, and 44), and results are reported as mean $\pm$ standard deviation across seeds. All methods are evaluated using identical data splits and corruption protocols to ensure controlled comparison. For each baseline, we follow the architecture and training configuration recommended in the original work and tune the relevant optimization hyperparameters on the validation set; the configuration achieving the best validation performance is then used for test evaluation. Throughout the experiments, MAA denotes the proposed memory-based approach, comprising clean-memory curation followed by memory access alignment.

3.1 Restoration Performance under Distribution Shift

We evaluate restoration fidelity using RMSE and PRD, waveform agreement using Pearson correlation (CC), and signal-quality improvement using ImSNR. Table 1 reports results on MITDB (ID), QTDB (cross-database OOD), and SIMEMG (external real-contamination OOD).

MAA improves both LUNet and ImputeECG, with larger relative RMSE reductions on the tested OOD datasets than on MITDB. Mean RMSE decreases by 1.5% and 5.4% on MITDB and by 13.2% and 10.8% on QTDB, respectively, with ImputeECG+MAA achieving the lowest mean QTDB RMSE among the evaluated methods. On SIMEMG, where contamination differs from the NSTDB-based training distribution, the corresponding RMSE reductions are 10.8% and 21.4%, accompanied by PRD reductions of 11.4% and 20.3% and ImSNR gains of 0.87 and 2.24 dB. LUNet+MAA achieves the best mean SIMEMG performance across all four metrics among the evaluated methods. These results show improved restoration under the tested distribution shifts while preserving ID performance. Representative qualitative denoising results on QTDB and SIMEMG are provided in Appendix A.4.

We further evaluate morphological fidelity on QTDB and transfer beyond ECG through EEG denoising (Appendix A.5). On QTDB, joint adaptation improves P-, QRS-, and T-wave RMSE and reduces QT and T-end bias relative to LUNet, although QRS correlation and QRS-width bias do not improve. For EEG, encoder-only adaptation improves all four metrics relative to LUNet under both ID and OOD evaluation. Mean OOD ImSNR nevertheless remains negative, indicating that these gains over LUNet do not yet yield an average signal-quality improvement over the noisy input.

3.2 Role of Fixed Memory in Reconstruction

We examine whether reconstruction depends on the intact memory pathway by changing or removing memory at inference while keeping all trained parameters unchanged. We compare intact (On), shuffled (Shuffle), zeroed (Zero), and removed (NO_MEM) memory, and include the clean-input teacher oracle as a privileged reference for the complete teacher–memory–decoder pathway.

Disrupting memory substantially degrades reconstruction across both backbones and all three datasets (Table 2). For LUNet, Shuffle increases mean RMSE from 0.1214 to 0.4058 on MITDB, 0.5631 to 1.1656 on QTDB, and 5.2883 to 33.0210 on SIMEMG, with CC falling close to zero. Zeroing or removing memory also substantially increases RMSE. The clean-input oracle achieves lower RMSE and higher CC than the noisy-input MAA model across all datasets, indicating that the curated pathway retains clean-input reconstruction capability under distribution shift. Furthermore, at the lowest tested data fraction, reducing memory-curation data causes greater reconstruc-

Table 1: **ECG restoration under distribution shift.** MITDB measures in-distribution (ID) performance, while QTDB and SIMEMG evaluate zero-shot out-of-distribution (OOD) restoration.

Method	RMSE↓	PRD↓	ImSNR↑	CC↑
MITDB (ID)				
CAE-CGAN	0.1321 ± 0.0029	31.3900 ± 0.4100	3.5500 ± 0.1500	0.8895 ± 0.0046
CS-TRANS	$\mathbf{0.1142 \pm 0.0014}$	$\mathbf{27.0300 \pm 0.3900}$	$\mathbf{5.0300 \pm 0.1100}$	$\mathbf{0.9088 \pm 0.0017}$
EDDM	0.1283 ± 0.0019	30.8400 ± 0.1600	3.8400 ± 0.0600	0.8866 ± 0.0016
TFCDiff	0.1429 ± 0.0022	34.1500 ± 0.6900	2.9200 ± 0.2400	0.8728 ± 0.0080
LUNet	0.1232 ± 0.0019	29.1600 ± 0.4200	4.2300 ± 0.1400	0.8990 ± 0.0035
LUNet (w/o skip control)	0.1265 ± 0.0005	30.3545 ± 0.0851	3.8099 ± 0.0384	0.8938 ± 0.0010
LUNet + MAA	0.1214 ± 0.0005	29.0900 ± 0.0500	4.2700 ± 0.0400	0.9012 ± 0.0010
ImputeECG	0.1295 ± 0.0014	30.5345 ± 0.3337	3.8013 ± 0.1225	0.8890 ± 0.0015
ImputeECG + MAA	0.1225 ± 0.0011	29.0562 ± 0.2599	4.3378 ± 0.1026	0.8936 ± 0.0024
QTDB (Zero-shot OOD)				
CAE-CGAN	0.6289 ± 0.0676	27.1800 ± 1.2800	5.2800 ± 0.5300	0.6698 ± 0.0135
CS-TRANS	0.6421 ± 0.0453	24.7400 ± 0.4700	$\mathbf{6.3700 \pm 0.2100}$	$\mathbf{0.6813 \pm 0.0033}$
EDDM	0.7820 ± 0.2019	27.4000 ± 2.7700	5.6200 ± 0.8300	0.6782 ± 0.0115
TFCDiff	0.6614 ± 0.0419	28.6500 ± 0.1200	5.0000 ± 0.0500	0.6500 ± 0.0135
LUNet	0.6490 ± 0.0296	26.1500 ± 0.7000	5.5400 ± 0.2900	0.6608 ± 0.0064
LUNet (w/o skip control)	0.6186 ± 0.0088	26.4155 ± 0.3355	5.4896 ± 0.1256	0.6570 ± 0.0032
LUNet + MAA	0.5631 ± 0.0120	24.6900 ± 0.3400	6.1200 ± 0.1700	0.6800 ± 0.0054
ImputeECG	0.6272 ± 0.0156	26.3006 ± 0.3006	5.6473 ± 0.1200	0.6672 ± 0.0010
ImputeECG + MAA	$\mathbf{0.5595 \pm 0.0124}$	$\mathbf{24.6639 \pm 0.1984}$	6.2472 ± 0.0799	0.6791 ± 0.0022
SIMEMG (Zero-shot OOD)				
CAE-CGAN	7.1461 ± 0.4676	32.0200 ± 2.1500	1.8100 ± 0.5500	0.9492 ± 0.0051
CS-TRANS	6.5217 ± 0.8799	29.4200 ± 4.2900	2.8700 ± 0.8700	0.9559 ± 0.0127
EDDM	7.7330 ± 0.3503	34.8800 ± 1.7000	1.3400 ± 0.3200	0.9474 ± 0.0011
TFCDiff	9.1415 ± 1.6028	41.4500 ± 7.6200	0.0900 ± 1.2900	0.9206 ± 0.0231
LUNet	5.9274 ± 0.2893	26.4500 ± 1.4100	3.3800 ± 0.4300	0.9658 ± 0.0047
LUNet (w/o skip control)	6.4769 ± 0.2121	28.4705 ± 0.8860	2.5228 ± 0.2592	0.9621 ± 0.0016
LUNet + MAA	$\mathbf{5.2883 \pm 0.2274}$	$\mathbf{23.4300 \pm 0.9500}$	$\mathbf{4.2500 \pm 0.3900}$	$\mathbf{0.9730 \pm 0.0022}$
ImputeECG	7.4371 ± 0.3203	33.1286 ± 1.3210	1.4942 ± 0.4037	0.9461 ± 0.0031
ImputeECG + MAA	5.8433 ± 0.3141	26.4093 ± 1.4555	3.7349 ± 0.4621	0.9602 ± 0.0050

Table 2: **Ablation study of memory content and availability.** RMSE and CC are reported for the clean-input teacher oracle and noisy-input memory controls. The oracle uses privileged clean input and is included only as a reference.

Backbone	Memory Condition	MITDB (ID) RMSE↓	MITDB (ID) CC↑	QTDB (OOD) RMSE↓	QTDB (OOD) CC↑	SIMEMG (OOD) RMSE↓	SIMEMG (OOD) CC↑
LUNet							
Teacher oracle (clean input)	ON	0.0315 ± 0.0004	0.9917 ± 0.0003	0.0516 ± 0.0015	0.9698 ± 0.0018	3.1250 ± 0.1866	0.9901 ± 0.0013
LUNet + MAA	ON	0.1214 ± 0.0005	0.9012 ± 0.0010	0.5631 ± 0.0120	0.6800 ± 0.0054	5.2883 ± 0.2274	0.9730 ± 0.0022
LUNet + MAA	SHUFFLE	0.4058 ± 0.0023	-0.0002 ± 0.0001	1.1656 ± 0.0057	0.0004 ± 0.0000	33.0210 ± 0.2826	0.0066 ± 0.0001
LUNet + MAA	ZERO	0.9116 ± 0.9258	0.0006 ± 0.0012	3.5648 ± 4.2681	0.0007 ± 0.0019	65.2771 ± 63.3286	-0.0006 ± 0.0020
LUNet + MAA	NO_MEM	8.1249 ± 1.7447	-0.0352 ± 0.1014	25.0105 ± 9.5640	0.0125 ± 0.0718	779.1420 ± 202.1085	0.0120 ± 0.0653
ImputeECG							
Teacher oracle (clean input)	ON	0.0105 ± 0.0049	0.9991 ± 0.0012	0.0164 ± 0.0120	0.9958 ± 0.0162	1.3161 ± 0.4059	0.9983 ± 0.0007
ImputeECG + MAA	ON	0.1225 ± 0.0011	0.8936 ± 0.0024	0.5595 ± 0.0124	0.6791 ± 0.0022	5.8433 ± 0.3141	0.9602 ± 0.0050
ImputeECG + MAA	SHUFFLE	0.4033 ± 0.0016	-0.0002 ± 0.0001	1.1829 ± 0.0061	0.0003 ± 0.0001	32.4871 ± 0.3588	0.0060 ± 0.0004
ImputeECG + MAA	ZERO	0.7058 ± 0.1444	0.0007 ± 0.0007	2.8384 ± 0.6730	-0.0005 ± 0.0005	50.2570 ± 9.8919	0.0010 ± 0.0006
ImputeECG + MAA	NO_MEM	0.7050 ± 0.1346	0.1760 ± 0.0769	2.9832 ± 0.6538	0.1383 ± 0.0809	51.4827 ± 8.8463	0.1423 ± 0.0645

tion degradation on SIMEMG than reducing access-alignment data, highlighting the importance of curation under this shift (Appendix A.12, Table A9).

Table 3: Component-wise Analysis of Memory Access Alignment Across Backbones. Evaluation on MITDB (ID) and zero-shot QTDB and SimEMG (OOD).

		MITDB (ID)		QTDB (OOD)		SIMEMG (OOD)	
Method	**Trainable**	**RMSE↓**	**CC↑**	**RMSE↓**	**CC↑**	**RMSE↓**	**CC↑**
LUNet backbone							
LUNet	–	0.1232 ± 0.0019	0.8990 ± 0.0035	0.6490 ± 0.0296	0.6608 ± 0.0064	5.9274 ± 0.2893	0.9658 ± 0.0047
Frozen access	–	0.2217 ± 0.0009	0.7889 ± 0.0008	1.4701 ± 0.0063	0.5579 ± 0.0003	4.5787 ± 0.1267	0.9785 ± 0.0009
Student-align	E_s	0.1229 ± 0.0025	0.8998 ± 0.0040	0.5684 ± 0.0210	$\mathbf{0.6863 \pm 0.0079}$	$\mathbf{4.9811 \pm 0.0380}$	$\mathbf{0.9767 \pm 0.0006}$
Student-align	R_s	0.1219 ± 0.0013	0.8998 ± 0.0022	0.5735 ± 0.0307	0.6764 ± 0.0075	5.1458 ± 0.1519	0.9745 ± 0.0015
Student-align	E_s, R_s	$\mathbf{0.1214 \pm 0.0005}$	$\mathbf{0.9012 \pm 0.0010}$	$\mathbf{0.5631 \pm 0.0120}$	0.6800 ± 0.0054	5.2883 ± 0.2274	0.9730 ± 0.0022
ImputeECG backbone							
ImputeECG	–	0.1295 ± 0.0014	0.8890 ± 0.0015	0.6272 ± 0.0156	0.6672 ± 0.0010	7.4371 ± 0.3203	0.9461 ± 0.0031
Frozen access	–	0.2285 ± 0.0002	0.7864 ± 0.0003	1.5262 ± 0.0010	0.5597 ± 0.0002	8.2904 ± 0.4913	0.9275 ± 0.0081
Student-align	E_s	0.1271 ± 0.0003	0.8841 ± 0.0013	0.5948 ± 0.0155	0.6594 ± 0.0006	6.5970 ± 0.2190	0.9480 ± 0.0058
Student-align	R_s	0.1341 ± 0.0013	0.8794 ± 0.0012	0.6603 ± 0.0082	0.6567 ± 0.0035	6.1101 ± 0.1115	0.9596 ± 0.0030
Student-align	E_s, R_s	$\mathbf{0.1225 \pm 0.0011}$	$\mathbf{0.8936 \pm 0.0024}$	$\mathbf{0.5595 \pm 0.0124}$	$\mathbf{0.6791 \pm 0.0022}$	$\mathbf{5.8433 \pm 0.3141}$	$\mathbf{0.9602 \pm 0.0050}$

3.3 Component Analysis of Memory Access Alignment under Distribution Shift

We investigate how to achieve reliable access to the fixed clean memory under distribution shift while preserving in-distribution restoration. We independently adapt the student encoder E_s, router R_s, or both while keeping the curated memory fixed. Specifically, E_s denotes bottleneck adaptation (E_{bot}), selected based on the encoder-depth analysis for its overall balance between ID preservation and OOD robustness (Appendix A.6). Table 3 compares these adaptation configurations across LUNet and ImputeECG.

Across both backbones, feature and router adaptation contribute to restoration, but their benefits are shift-dependent. For LUNet, adapting both E_s and R_s achieves the lowest RMSE on MITDB and QTDB, while adapting E_s alone performs best among the aligned variants on SIMEMG. For ImputeECG, adapting both E_s and R_s achieves the lowest RMSE across all three datasets. These results support joint adaptation as the default configuration, although the best configuration depends on the backbone and distribution shift. A complementary loss ablation shows that combining latent and memory-readout alignment losses with reconstruction supervision yields the lowest mean RMSE across both backbones and all three datasets (Appendix, Table A10).

3.4 From Memory-Access Alignment to Reconstruction

We investigate the reconstruction effects of memory-access adaptation from three complementary perspectives: agreement with the clean-input teacher, the interaction between the induced output change and the frozen reconstruction residual, and controlled interventions on the adaptation-induced change in the memory readout. Teacher agreement is evaluated at the levels of memory-access patterns (Appendix A.11) and retrieved representations (Appendix A.10) to determine whether closer agreement corresponds to better reconstruction. Output-level analysis examines how the direction and magnitude of the induced change relative to the frozen reconstruction residual account for both improvement and degradation (Appendix A.9). Controlled readout interventions (Table 4) test which components of the readout change reproduce the reconstruction effect and whether this effect depends on their direction when magnitude is preserved.

For the controlled interventions, we isolate the component of the adaptation-induced memory readout change to which the frozen decoder is most sensitive. For a corrupted input, let m_S^F denote the student readout before Stage 2 adaptation and m_S^A the readout after adaptation, giving $\delta m = m_S^A - m_S^F$. Using clean-teacher readouts from MITDB validation data, we estimate a projector P onto the smallest readout subspace accounting for at least 90% of the decoder's local sensitivity, yielding $r = 14$ dimensions for LUNet and $r = 3$ for ImputeECG (Appendix A.7). We compare the full adapted readout $m_S^F + \delta m$ with the projected intervention $m_S^F + P\delta m$, and with interventions obtained by randomly rotating $P\delta m$ within the same sensitive subspace while preserving its readout-space energy. We quantify the fraction of adaptation-induced readout energy retained by the sensitive component as $E_P = \|P\delta m\|^2/\|\delta m\|^2$, and the full adaptation effect relative to frozen access as $\Delta L_{\text{Full}} = L_{\text{Full}} - L_F$. The ratio $R_P = (L_P - L_F)/\Delta L_{\text{Full}}$ measures how closely the

Table 4: Reconstruction effects of memory-access adaptation.

	MITDB (ID)				QTDB (OOD)				SIMEMG (OOD)			
Adaptation	E_P	ΔL_{Full}	R_P	$\Delta L_{\text{Rot},P}$	E_P	ΔL_{Full}	R_P	$\Delta L_{\text{Rot},P}$	E_P	ΔL_{Full}	R_P	$\Delta L_{\text{Rot},P}$
LUNet ($r = 14, d = 64, r/d = 0.2188$)												
E_s	0.2784	−0.0604	0.9919	0.0107	0.2724	−7.3611	0.9846	0.9441	0.1940	2.4882	1.3148	14.8778
R_s	0.2749	−0.0609	0.9937	0.0102	0.2713	−7.3282	0.9918	0.8027	0.2119	4.7668	1.1574	14.5673
E_s, R_s	0.2810	−0.0613	0.9950	0.0113	0.2795	−7.4531	0.9917	0.9235	0.2125	6.7974	0.9960	22.8330
ImputeECG ($r = 3, d = 64, r/d = 0.0469$)												
E_s	0.4077	−0.0640	0.9934	0.0505	0.4086	−7.8181	0.9906	4.6102	0.4006	−30.3181	0.9741	130.6079
R_s	0.4055	−0.0623	0.9932	0.0427	0.4073	−7.5700	0.9898	3.6837	0.3944	−38.6663	0.9823	100.4993
E_s, R_s	0.4062	−0.0657	0.9934	0.0538	0.4074	−7.9812	0.9905	4.8259	0.4062	−43.1476	0.9859	144.8911

projected intervention reproduces the signed reconstruction effect of full adaptation, with values near one indicating similar effects. Finally, $\Delta L_{\text{Rot},P}$ measures the mean MSE change after energy-preserving random rotations, testing whether occupying the same sensitive subspace is sufficient or whether the specific direction selected by adaptation also matters. Negative MSE changes indicate improved reconstruction, with MSE evaluated on the restored physical scale.

On MITDB and QTDB, the decoder-sensitive component reproduces nearly the full mean MSE effect across both backbones and all three adaptation scopes, with mean R_P values of approximately 0.985–0.995. This occurs while retaining only 27–28% of the readout-change energy for LUNet and 40–41% for ImputeECG. Random rotations within the same subspace produce positive mean MSE changes in every reported setting, showing that preserving subspace membership and energy does not preserve the reconstruction benefit. On SIMEMG, the projected update reproduces the direction of both the beneficial adaptation effects for ImputeECG and the harmful effects for LUNet. Results are averaged across three training seeds, with rotations averaged within each seed; full results, including standard deviations and complementary-component effects, are provided in Appendix A.8.

The interventions show that a restricted decoder-sensitive component reproduces much of adaptation's MSE effect, while rotations demonstrate that subspace membership and readout-change energy alone do not preserve its benefit. Consistently, stronger teacher agreement in access patterns and smaller teacher–student readout errors do not reliably identify the best reconstruction (Appendices A.11 and A.10). The output decomposition clarifies this distinction: reconstruction improves when the output change opposes the frozen residual strongly enough to outweigh its own energy. For LUNet on SIMEMG, this corrective interaction is insufficient, explaining why adaptation increases MSE despite partially counteracting the existing error (Appendix A.9). Together, these findings distinguish matching a clean reference from retrieving a useful correction: the clean teacher guides memory access, but the reconstruction value of an access change depends on its corrective effect through the frozen decoder.

4 Conclusion

We presented a two-stage ECG denoising framework that separates learning clean latent structure from learning how corrupted observations access it. The framework first curates reusable clean ECG content in a key–value memory and then applies Memory Access Alignment (MAA) to adapt the encoder bottleneck and router while keeping the memory and decoder frozen. Across both convolutional and Transformer-based backbones, MAA improves restoration under the tested shifts in signal and corruption characteristics while preserving in-distribution performance. Our analyses show that closer clean-teacher access agreement does not necessarily yield better reconstruction. Controlled readout interventions further show that decoder-sensitive components of the readout change account for most of the improvement or deterioration induced by adaptation. Future work can extend memory-access learning to weakly supervised or unlabeled settings and investigate whether the retained clean structure supports tasks beyond restoration. These findings highlight the importance of evaluating memory-access adaptation through its effects on reconstruction, since teacher agreement alone does not reliably rank restoration outcomes.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing and manuscript editing, and to support preparation of the public code release, including code cleanup, removal of unused experimental branches and comments, and organization of the release package. All AI-assisted edits to the manuscript and code were manually reviewed by the authors before inclusion or release.

Generative AI tools were not used to originate the research idea, formulate the main contribution, generate datasets, or produce the experimental results reported in this paper. All reported results were obtained by executing and evaluating the implemented methods under the experimental protocol described in the paper. The authors take full responsibility for the final manuscript, released code, reported claims, and experimental results.

ETHICS STATEMENT

This study uses publicly available, previously released ECG and EEG datasets and does not involve the collection of new human-subject data. All recordings were used in accordance with the access conditions and usage terms specified by the original data providers. The experiments were conducted exclusively on previously published data, with no attempt to identify individual participants.

REPRODUCIBILITY STATEMENT

To support reproducibility, we provide implementation details, memory configurations, training hyperparameters, checkpoint selection criteria, and computational efficiency comparisons in Appendix A.3. Our code is available at `https://anonymous.4open.science/r/MAA_ECG_Denoising-A0D4/README.md`.

REFERENCES

Vladimir Atanasoski, Jovana Petrovic, Lana Popović Maneski, Marjan Miletić, Miloš Babić, Aleksandra Nikolić, Dorin Panescu, and Marija D Ivanović. A database of simultaneously recorded ecg signals with and without emg noise. *IEEE Open Journal of Engineering in Medicine and Biology*, 4:222–225, 2023.

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.

Wesley Chorney, Haifeng Wang, Lu He, Seunghan Lee, and Lir-Wan Fan. Convolutional block attention autoencoder for denoising electrocardiograms. *Biomedical Signal Processing and Control*, 86:105242, 2023.

Xiaocheng Fang, Haoyu Wang, Jieyi Cai, Qinghao Zhao, Jun Li, Shanwei Zhang, Guangkun Nie, Yujie Xiao, Shun Huang, Jiarui Jin, et al. Imputeecg: Deep learning reconstruction of complete 12-lead electrocardiograms from incomplete recordings for cardiac assessment. *arXiv preprint arXiv:2607.05009*, 2026.

Amir Ghafari, Niusha Pourjafari, and Ali Ghaffari. Vector-based postprocessing method for improving ecg denoising techniques by re-establishing lead relationships. *IEEE Transactions on Instrumentation and Measurement*, 73:1–9, 2023.

Hesam Halvaei, Emma Svennberg, Leif Sörnmo, and Martin Stridh. Identification of transient noise to reduce false detections in screening for atrial fibrillation. *Frontiers in physiology*, 12:672875, 2021.

Lei Hu, Wenjie Cai, Ziyang Chen, and Mingjie Wang. A lightweight u-net model for denoising and noise localization of ecg signals. *Biomedical Signal Processing and Control*, 88:105504, 2024.

Jiarui Jin, Haoyu Wang, Hongyan Li, Jun Li, Jiahui Pan, and Shenda Hong. Reading your heart: Learning ECG words and sentences via pre-training ECG language model. In *The Thirteenth International Conference on Learning Representations*, 2025. URL `https://openreview.net/forum?id=6Hz1Ko087B`.

Yanrui Jin, Chengjin Qin, Jinlei Liu, Zhiyuan Li, Haotian Shi, Ke Lin, Yunqing Liu, and Chengliang Liu. A novel incremental and interactive method for actual heartbeat classification with limited additional labeled samples. *IEEE Transactions on Instrumentation and Measurement*, 70:1–12, 2021.

Manousos A. Klados and Panagiotis D. Bamidis. A semi-simulated eeg/eog dataset for the comparison of eog artifact rejection techniques. *Data in Brief*, 8:1004–1006, 2016. ISSN 2352-3409. doi: https://doi.org/10.1016/j.dib.2016.06.032. URL https://www.sciencedirect.com/science/article/pii/S2352340916304000.

Dominik D Kranz, Jan F Krämer, Oruç Kahriman, Alexander Nelde, Niels Wessel, and Christian Meisel. Exploring latent diffusion models for ecg generation on the minute scale. *Computer Methods and Programs in Biomedicine*, pp. 109138, 2025.

Weichao Kuang, Shanjin Wang, Yingxin Lai, and Wing-Kuen Ling. Efficient and adaptive signal denoising based on multistage singular spectrum analysis. *IEEE Transactions on Instrumentation and Measurement*, 70:1–20, 2020.

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.

Pablo Laguna, Roger G Mark, A Goldberg, and George B Moody. A database for evaluation of algorithms for measurement of qt and other waveform intervals in the ecg. In *Computers in cardiology 1997*, pp. 673–676. IEEE, 1997.

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, 2023.

Huayu Li, ZhengXiao He, Xiwen Chen, Jingjing Wang, siyuan tian, Jinghao Wen, and Ao Li. Learning fingerprints for medical time series with redundancy-constrained information maximization. In *Forty-third International Conference on Machine Learning*, 2026a. URL https://openreview.net/forum?id=lrRBHFIgaK.

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, 2026b. doi: 10.1109/JBHI.2026.3732677.

Zhiyuan Li, Yanrui Jin, Yuanyuan Tian, Jinlei Liu, Mengxiao Wang, Xiaoyang Wei, Liqun Zhao, and Chengliang Liu. M-xaf: Medical explainable diagnosis system of atrial fibrillation based on medical knowledge and semantic representation fusion. *Engineering Applications of Artificial Intelligence*, 136:108890, 2024.

Zhiyuan Li, Yuanyuan Tian, Yanrui Jin, Xiaoyang Wei, Mengxiao Wang, Jinlei Liu, and Chengliang Liu. Eddm: A novel ecg denoising method using dual-path diffusion model. *IEEE Transactions on Instrumentation and Measurement*, 2025.

Zhi Lu, Jinbo Chen, Dongheng Zhang, Haoyu Wang, Pengcheng Huang, Fang Zhou, Yang Hu, Qibin Sun, Min Gao, and Yan Chen. Contactless 12-lead electrocardiogram via deep computational radar. *npj Biomedical Innovations*, 3(1):6, 2026.

Caiyun Ma, Xi Long, Weijie Sheng, Rik Vullings, Meicheng Yang, Lina Zhao, Ronald M Aarts, Jianqing Li, and Chengyu Liu. An atrial fibrillation detection strategy in dynamic ecgs with significant individual differences. *IEEE Transactions on Instrumentation and Measurement*, 73: 1–10, 2023.

George B Moody and Roger G Mark. The impact of the mit-bih arrhythmia database. *IEEE engineering in medicine and biology magazine*, 20(3):45–50, 2001.

George B Moody, WE Muldrow, and Roger G Mark. A noise stress test for arrhythmia detectors. *Computers in cardiology*, 11(3):381–384, 1984.

Huyang Peng, Xiaohan Chang, Zhenjie Yao, Donglin Shi, and Yongrui Chen. A deep learning framework for ecg denoising and classification. *Biomedical Signal Processing and Control*, 94: 106441, 2024.

Jan A Van Alste and TS Schilder. Removal of base-line wander and power-line interference from the ecg by an efficient fir filter with a reduced number of taps. *IEEE transactions on biomedical engineering*, (12):1052–1060, 1985.

Xiaoyu Wang, Bingchu Chen, Ming Zeng, Yuli Wang, Hui Liu, Ruixia Liu, Lan Tian, and Xiaoshan Lu. An ecg signal denoising method using conditional generative adversarial net. *IEEE Journal of Biomedical and Health Informatics*, 26(7):2929–2940, 2022.

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.

Jin Zhang, Jia-lun Lin, Xiao-ling Li, and Wei-quan Wang. Ecg signals denoising method based on improved wavelet threshold algorithm. In *2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pp. 1779–1784. IEEE, 2016.

Juya Zhang, Yu Guo, Xinming Dong, Tong Wang, Jinhai Wang, Xin Ma, and Huiquan Wang. Opportunities and challenges of noise interference suppression algorithms for dynamic ecg signals in wearable devices: A review. *Measurement*, pp. 117067, 2025a.

Yue Zhang, Chenyu Zhao, Wen Zhang, Jinliang Wang, Jun Guo, Wei Yang, and Qianjin Feng. Contrastive Disentanglement Learning Framework for Multi-lead Wearable ECG Denoising . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025*, volume LNCS 15962. Springer Nature Switzerland, September 2025b.

Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. *Advances in Neural Information Processing Systems*, 35:30599–30611, 2022.

A Appendix

A.1 ECG Data Preparation

Dataset splitting. MITDB is split at the record level into training, validation, and test sets using a 70/15/15% ratio with seed 42. Splitting is performed before window extraction, such that all segments originating from the same recording remain within a single subset. This prevents leakage through shared morphology, rhythm, acquisition characteristics, or temporal context.

QTDB is used only for zero-shot OOD evaluation. Because 15 QTDB records originate from MITDB, these recordings are removed before evaluation, leaving 90 non-overlapping QTDB records. We use all remaining records for OOD testing. This guarantees recording-level separation from the MITDB source domain; subject-level separation cannot be established reliably because consistent cross-database subject identifiers are unavailable.

SIMEMG is also used exclusively for zero-shot OOD evaluation. Its provided clean and EMG-contaminated recordings are treated as paired observations and are never used for training or model selection.

Resampling and segmentation. All ECG signals are represented at 360 Hz. MITDB and NSTDB are retained at their native sampling frequency. QTDB is resampled from 250 to 360 Hz using anti-aliased polyphase resampling with a rational factor of $36/25$, while SIMEMG is resampled from 500 to 360 Hz using a factor of $18/25$. For SIMEMG, clean and contaminated signals are resampled using identical parameters to preserve temporal alignment.

Recordings are subsequently divided into non-overlapping 10-s windows, corresponding to $T = 3600$ samples at 360 Hz. Only complete windows are retained.

Synthetic noise generation. For MITDB and QTDB, noisy ECG observations are generated using physiological artifacts from NSTDB: baseline wander (BW), muscle artifact (MA), and electrode motion (EM). Noise is added in the original amplitude domain before normalization. Given a clean ECG window $y \in \mathbb{R}^T$ and an NSTDB noise segment $\eta \in \mathbb{R}^T$, their mean-squared powers are

$$P_y = \frac{1}{T}\sum_{t=1}^{T} y(t)^2, \qquad P_\eta = \frac{1}{T}\sum_{t=1}^{T} \eta(t)^2. \tag{10}$$

For a target SNR of s dB, the noise scaling coefficient is

$$\gamma = \sqrt{\frac{P_y}{P_\eta, 10^{s/10}}}, \tag{11}$$

and the corrupted ECG is

$$x = y + \gamma\eta. \tag{12}$$

This yields

$$\mathrm{SNR}_{\mathrm{dB}} = 10\log_{10}\left(\frac{P_y}{P_{\gamma\eta}}\right). \tag{13}$$

Training uses both single-artifact and mixed-artifact corruption, whereas validation uses single-artifact corruption. Test samples are generated at fixed SNR levels of $0, 5, 10, 15$ dB using both single and mixed artifacts. For mixtures containing M artifact types, the target noise power is divided equally across the components, which are independently scaled and subsequently summed.

To reduce leakage through repeated noise realizations, NSTDB channel 0 is used for training and validation, whereas channel 1 is reserved for testing. Noise segments used during evaluation are therefore disjoint from those used for model development.

SIMEMG evaluation. Unlike MITDB and QTDB, SIMEMG requires no synthetic corruption. We directly use its simultaneously recorded EMG-contaminated ECG as input and the corresponding clean ECG as reference. Consequently, SIMEMG evaluates restoration under naturally acquired contamination that is not generated by the NSTDB mixing procedure used during training.

Normalization. Each window is standardized independently using statistics computed from the *noisy observation only*. Specifically, for noisy input x, we compute

$$\mu_x = \frac{1}{T}\sum_{t=1}^{T} x(t), \qquad \sigma_x = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x(t)-\mu_x)^2}, \tag{14}$$

and normalize the input and corresponding clean target as

$$\tilde{x} = \frac{x-\mu_x}{\sigma_x}, \qquad \tilde{y} = \frac{y-\mu_x}{\sigma_x}. \tag{15}$$

Using only statistics available from the observed noisy signal prevents oracle information from entering the preprocessing pipeline. Evaluation metrics are computed after transforming model predictions back to the original signal scale.

Table A1: Computational efficiency and reconstruction performance.

Method	Params (M) ↓	MACs (G) ↓	Peak Memory (MiB) ↓	RMSE MITDB ↓	RMSE QTDB ↓	RMSE SIMEMG ↓
CAE-CGAN	0.105	0.132	3.76	0.1321 ± 0.0029	0.6289 ± 0.0676	7.1461 ± 0.4676
CS-TRANS	0.091	0.065	2.83	0.1142 ± 0.0014	0.6421 ± 0.0453	6.5217 ± 0.8799
EDDM	1.373	150.823	23.03	0.1283 ± 0.0019	0.7820 ± 0.2019	7.7330 ± 0.3503
TFCDiff	3.598	650.623	69.97	0.1429 ± 0.0022	0.6614 ± 0.0419	9.1415 ± 1.6028
LUNet	0.023	0.014	2.19	0.1232 ± 0.0019	0.6490 ± 0.0296	5.9274 ± 0.2893
LUNet + MAA	0.060	0.012	10.89	0.1214 ± 0.0005	0.5631 ± 0.0120	5.2883 ± 0.2274
ImputeECG (2026)	110.688	55.255	466.00	0.1295 ± 0.0014	0.6272 ± 0.0156	7.4371 ± 0.3203
ImputeECG + MAA	111.230	55.419	511.73	0.1225 ± 0.0011	0.5595 ± 0.0124	5.8433 ± 0.3141

A.2 EEG Data Preparation

EEGDenoiseNet-EOG. We use the EOG subset of EEGDenoiseNet, which contains 4,514 clean EEG epochs and 3,400 EOG artifact epochs. Each example is represented as a 2-s segment (512 samples) at 256 Hz. The original training set is partitioned into training and validation subsets using an 80/20 split with seed 42, while the provided test set is retained unchanged for ID evaluation. Splitting is performed at the clean–artifact pair level: all ten noisy realizations of the same underlying pair, corresponding to SNRs from -7 to 2 dB, remain within the same subset to prevent leakage across corruption levels.

Klados–Bamidis EEG/EOG. For zero-shot OOD evaluation, we use the Klados–Bamidis semi-simulated EEG/EOG dataset, comprising 54 30-s recordings from 27 subjects, each with 19 EEG channels sampled at 200 Hz. We directly use the provided $Pure_{Data}$ and $Contaminated_{Data}$ signals as paired clean targets and contaminated inputs; no additional EOG corruption is introduced. Both signals are identically resampled from 200 to 256 Hz using polyphase resampling and segmented into non-overlapping 2-s windows (512 samples). Each channel is treated as an individual restoration example, yielding 15,390 OOD test windows. The dataset is used exclusively for testing and is never accessed during training, validation, hyperparameter tuning, or model selection.

For both EEG datasets, we apply the same noisy-input-derived z-score normalization protocol used for ECG (Appendix A.1).

A.3 Implementation Details and Computational Efficiency

All models are implemented in PyTorch and trained on an NVIDIA A100-SXM4-40GB GPU. For both LUNet and ImputeECG, the memory contains $N = 48$ slots with key dimension $d_k = 32$ and value dimension $d_v = 64$. Retrieval uses the top $K = 8$ memory entries, a router window of $w = 13$, and a softmax temperature of $\tau = 0.1$. Queries and keys are ℓ_2-normalized before cosine-similarity scores are computed.

In Stage 1, the teacher encoder, router, memory keys and values, and decoder are jointly trained from scratch using the clean-signal reconstruction objective defined in Section 2. In Stage 2, the student encoder and router are initialized from the corresponding Stage 1 teacher checkpoint. Optimization is restricted to the student encoder bottleneck E_B and student router R_S, while the remaining student encoder blocks, teacher pathway, curated memory, and decoder remain frozen.

Each stage is trained for 150 epochs using AdamW with an initial learning rate of 1.5×10^{-4} and a cosine annealing schedule. We apply global gradient clipping with a maximum norm of 1.0 and use FP16 automatic mixed precision. Batch sizes are 16 for LUNet and 4 for ImputeECG. For each stage, we select the checkpoint with the lowest validation RMSE measured on the original physical scale.

For computational efficiency, Table A1 compares the computational cost of MAA with that of the corresponding backbones. For LUNet, MAA increases the parameter count from 0.023M to 0.060M and peak GPU memory from 2.19 to 10.89 MiB. Although these increases are substantial relative to the small backbone, the resulting model remains lightweight in absolute size. The reported MAC estimates are of the same order, at 0.014G for LUNet and 0.012G for LUNet + MAA. For ImputeECG, the additional cost is small relative to the backbone: the parameter count increases from 110.688M to

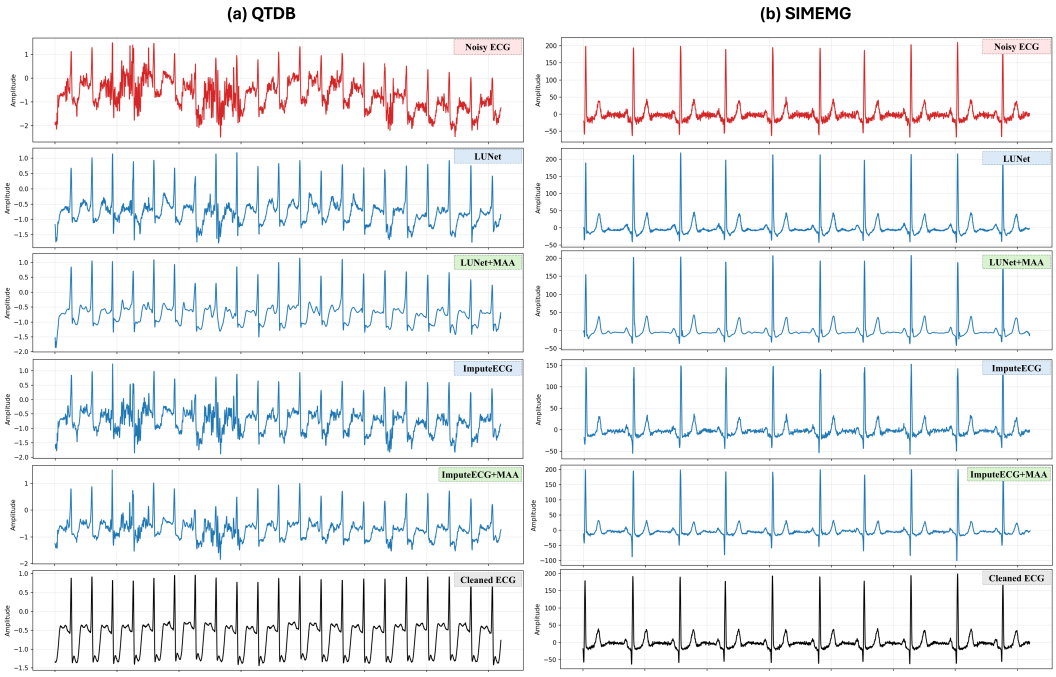

Figure A1: **Qualitative denoising results under distribution shift.** Representative examples from QTDB and SIMEMG comparing the noisy input, LUNet, LUNet + MAA, ImputeECG, and ImputeECG + MAA.

111.230M (0.49%), MACs from 55.255G to 55.419G (0.30%), and peak GPU memory from 466.00 to 511.73 MiB (9.81%). These measurements show that the memory mechanism introduces limited additional computation for the larger Transformer backbone, while its higher relative overhead for LUNet remains modest in absolute resource usage.

A.4 Qualitative Denoising Results

Figure A1 provides representative qualitative comparisons on the OOD test sets. We compare the noisy input with the original LUNet and ImputeECG backbones and their corresponding MAA variants. Across both architectures, the examples illustrate how MAA affects restoration under distribution shift, particularly in suppressing residual artifacts while preserving ECG morphology.

A.5 Morphological Fidelity and Cross-Signal Generality

We further test whether MAA preserves task-relevant signal structure and generalizes beyond ECG. On QTDB, we evaluate P-, QRS-, and T-wave reconstruction together with interval bias (Table A2). Joint alignment achieves the lowest RMSE across all three waveforms and substantially reduces QT and T-end bias, demonstrating that its OOD gains preserve physiologically meaningful morphology rather than only reducing pointwise error. We then instantiate the same memory-access formulation for EEG denoising without changing its core alignment mechanism and evaluate both ID and zero-shot OOD performance (Table A3). Feature alignment achieves the strongest adapted-student performance across both settings, demonstrating that the access-gap mechanism is not specific to ECG and can improve robustness across a distinct physiological modality and external distribution shift.

A.6 Effect of Encoder Adaptation Scope

We examine how the location and extent of encoder adaptation affect reconstruction and memory access across distribution shifts. Using the encoder-only alignment configuration (E_s) with LUNet,

Table A2: Morphological Preservation under Distribution Shift on QTDB. Waveform fidelity is evaluated for the P wave, QRS complex, and T wave, together with QRS-width, QT, and T-end interval bias. For interval bias, values closer to zero indicate better preservation.

	P wave		QRS		T wave		Interval bias (ms)		
Method	**RMSE↓**	**CC↑**	**RMSE↓**	**CC↑**	**RMSE↓**	**CC↑**	**QRS width**	**QT**	**T-end**
Noisy input	1.1907 ± 0.0000	0.5086 ± 0.0000	1.1725 ± 0.0000	0.8798 ± 0.0000	1.4717 ± 0.0000	0.5696 ± 0.0000	-27.74 ± 0.00	20.31 ± 0.00	36.79 ± 0.00
LUNet	0.4909 ± 0.0221	0.4889 ± 0.0367	0.5406 ± 0.0246	$\mathbf{0.8796 \pm 0.0031}$	0.6172 ± 0.0280	0.6233 ± 0.0052	$\mathbf{-30.25 \pm 0.18}$	11.61 ± 4.82	28.60 ± 4.68
Frozen	1.1198 ± 0.0040	$\mathbf{0.5437 \pm 0.0017}$	1.1413 ± 0.0045	0.8611 ± 0.0018	1.3885 ± 0.0055	0.6415 ± 0.0023	-30.75 ± 0.21	16.45 ± 1.38	33.58 ± 1.34
Feature E_s	0.4341 ± 0.0127	0.5393 ± 0.0079	0.4981 ± 0.0141	0.8683 ± 0.0021	0.5398 ± 0.0184	0.6543 ± 0.0045	-30.84 ± 0.22	14.52 ± 1.51	31.67 ± 1.41
Router R_s	0.4350 ± 0.0246	0.5332 ± 0.0059	0.5036 ± 0.0258	0.8664 ± 0.0008	0.5433 ± 0.0315	0.6524 ± 0.0048	-30.91 ± 0.36	9.15 ± 0.97	26.36 ± 0.75
Joint E_s, R_s	$\mathbf{0.4277 \pm 0.0090}$	0.5316 ± 0.0126	$\mathbf{0.4966 \pm 0.0058}$	0.8648 ± 0.0041	$\mathbf{0.5337 \pm 0.0113}$	$\mathbf{0.6548 \pm 0.0045}$	-30.58 ± 0.31	$\mathbf{9.15 \pm 1.83}$	$\mathbf{26.10 \pm 1.60}$

Table A3: **Generality of Memory Access Alignment Beyond ECG.** We instantiate the same memory-access framework for EEG denoising and evaluate in-distribution performance on EEGDenoiseNet-EOG and zero-shot generalization to the external Klados–Bamidis EEG/EOG dataset.

		EEGDenoiseNet-EOG (ID)				Klados–Bamidis EEG/EOG (Zero-shot OOD)			
Adaptation	**Trainable**	**RMSE↓**	**PRD↓**	**ImSNR↑**	**CC↑**	**RMSE↓**	**PRD↓**	**ImSNR↑**	**CC↑**
LUNet	–	0.2260 ± 0.0915	95.4315 ± 19.7076	5.5645 ± 4.9312	0.4823 ± 0.1483	0.0122 ± 0.0074	101.5875 ± 37.1750	-8.4037 ± 9.4211	0.4390 ± 0.1803
Teacher (noisy)	–	0.4147 ± 0.2426	176.8376 ± 84.7557	0.9700 ± 2.2173	0.3408 ± 0.1896	0.0116 ± 0.0076	102.0307 ± 65.5504	-7.7526 ± 7.9105	0.6013 ± 0.2101
Feature alignment	E_s	$\mathbf{0.2056 \pm 0.0819}$	$\mathbf{87.1136 \pm 16.4925}$	$\mathbf{6.3376 \pm 4.8918}$	$\mathbf{0.5559 \pm 0.1427}$	$\mathbf{0.0112 \pm 0.0068}$	$\mathbf{93.8939 \pm 35.7002}$	$\mathbf{-7.6793 \pm 9.2671}$	$\mathbf{0.5622 \pm 0.1349}$
Router alignment	R_s	0.2094 ± 0.0830	88.5401 ± 15.9031	6.1852 ± 4.8656	0.5401 ± 0.1452	0.0114 ± 0.0070	94.8420 ± 36.3906	-7.7596 ± 9.2293	0.5472 ± 0.1475
Joint alignment	E_s, R_s	0.2080 ± 0.0822	87.9929 ± 15.6502	6.2366 ± 4.8831	0.5410 ± 0.1472	0.0113 ± 0.0069	94.2611 ± 35.9502	-7.7098 ± 9.2225	0.5546 ± 0.1417

Table A4 compares bottleneck adaptation (E_B), adaptation of individual encoder stages (E_1–E_3), and full-encoder adaptation (E_{full}). Only the indicated encoder components are trainable. The memory router, curated memory, and decoder remain frozen. We evaluate reconstruction and agreement with clean-teacher memory access on MITDB and under zero-shot transfer to QTDB and SIMEMG.

On MITDB and QTDB, bottleneck adaptation recovers most of the reconstruction improvement achieved by full-encoder adaptation relative to frozen access. Full-encoder adaptation achieves the lowest mean RMSE, reducing error from 0.1229 to 0.1209 on MITDB and from 0.5684 to 0.5442 on QTDB compared with bottleneck adaptation. Adapting E_3 alone yields similar mean RMSE to bottleneck adaptation, whereas adapting E_2 or E_1 alone yields higher error on both datasets. The access diagnostics distinguish reconstruction performance from teacher-access agreement. On MITDB, bottleneck adaptation achieves the best mean query cosine, routing KL, and adjusted Top-K overlap, although full-encoder adaptation achieves lower mean RMSE. On QTDB, full-encoder adaptation produces the highest query cosine, while bottleneck adaptation retains the lowest routing KL and highest adjusted Top-K overlap. Thus, the additional reconstruction benefit of full-encoder adaptation is not accompanied by uniformly better agreement with clean-teacher access.

SIMEMG exhibits a different response to adaptation. Frozen access achieves the lowest mean RMSE at 4.5787, and none of the tested encoder-only adaptation scopes improves on this reference. Bottleneck adaptation achieves the lowest mean RMSE among adapted configurations at 4.9811, compared with 5.1413 for full-encoder adaptation and 6.1420 for adaptation of E_1 alone. Nevertheless, bottleneck adaptation improves routing KL and adjusted Top-K overlap relative to frozen access. This combination shows that closer agreement with teacher routing can coexist with poorer reconstruction under real contamination. Within the encoder-only adaptation setting, these results support the bottleneck as a restricted adaptation scope that retains most of the MITDB and QTDB reconstruction gains while yielding lower mean RMSE than full-encoder adaptation on SIMEMG.

A.7 Construction of the Decoder-Sensitive Subspace

We construct the decoder-sensitive projector P from the local sensitivity of the frozen decoder to clean-teacher memory readouts. This construction identifies readout-channel directions along which small perturbations produce large first-order changes in the decoder output, averaged over clean reference signals and readout positions. The projector is estimated exclusively from MITDB validation data and is subsequently kept fixed during the readout-intervention analysis.

Let $D_T(\cdot)$ denote the frozen teacher decoder and $m_{T,i} \in \mathbb{R}^{T_m \times d}$ the clean-teacher memory readout for validation sample i, where T_m is the number of readout positions and d is the readout-channel dimension. For each sample i and readout position t, we compute the Jacobian of the reconstructed

Table A4: **Effect of encoder adaptation scope.** The trainable encoder component is varied while the curated memory remains frozen. E_B denotes the bottleneck, E_1–E_3 denote individual encoder stages, and E_{full} denotes the full encoder. RMSE is reported with clean-teacher query cosine, routing KL, and chance-corrected Top-K overlap.

Adaptation	Trainable	RMSE↓	Query Cos.↑	Routing KL↓	Adj. Top-K↑
MITDB (ID)					
Frozen access	–	0.2217 ± 0.0009	0.8811 ± 0.0071	0.1479 ± 0.0200	0.8130 ± 0.0089
Encoder alignment	E_B	0.1229 ± 0.0025	$\mathbf{0.9286 \pm 0.0097}$	$\mathbf{0.1018 \pm 0.0118}$	$\mathbf{0.8410 \pm 0.0156}$
Encoder alignment	E_3	0.1231 ± 0.0010	0.9177 ± 0.0042	0.1198 ± 0.0110	0.8228 ± 0.0050
Encoder alignment	E_2	0.1329 ± 0.0005	0.9038 ± 0.0127	0.1526 ± 0.0239	0.5831 ± 0.3753
Encoder alignment	E_1	0.1569 ± 0.0049	0.8547 ± 0.0539	0.2679 ± 0.1271	0.7318 ± 0.0613
Encoder alignment	E_{full}	$\mathbf{0.1209 \pm 0.0001}$	0.9265 ± 0.0056	0.1066 ± 0.0080	0.8337 ± 0.0023
QTDB (OOD)					
Frozen access	–	1.4701 ± 0.0063	0.8081 ± 0.0117	0.2199 ± 0.0327	0.7389 ± 0.0133
Encoder alignment	E_B	0.5684 ± 0.0210	0.9098 ± 0.0108	$\mathbf{0.1365 \pm 0.0179}$	$\mathbf{0.7975 \pm 0.0279}$
Encoder alignment	E_3	0.5662 ± 0.0165	0.8996 ± 0.0053	0.1565 ± 0.0142	0.7775 ± 0.0081
Encoder alignment	E_2	0.6180 ± 0.0158	0.8736 ± 0.0191	0.2026 ± 0.0308	0.7416 ± 0.0103
Encoder alignment	E_1	0.7787 ± 0.0767	0.8124 ± 0.0690	0.3192 ± 0.1540	0.6745 ± 0.0616
Encoder alignment	E_{full}	$\mathbf{0.5442 \pm 0.0025}$	$\mathbf{0.9125 \pm 0.0060}$	0.1381 ± 0.0098	0.7933 ± 0.0053
SIMEMG (OOD)					
Frozen access	–	$\mathbf{4.5787 \pm 0.1267}$	$\mathbf{0.9683 \pm 0.0090}$	0.0885 ± 0.0318	0.8692 ± 0.0292
Encoder alignment	E_B	4.9811 ± 0.0380	0.9636 ± 0.0130	$\mathbf{0.0647 \pm 0.0200}$	$\mathbf{0.8811 \pm 0.0159}$
Encoder alignment	E_3	5.0316 ± 0.1954	0.9507 ± 0.0030	0.0821 ± 0.0114	0.8641 ± 0.0128
Encoder alignment	E_2	4.9898 ± 0.2405	0.9547 ± 0.0057	0.0826 ± 0.0147	0.8613 ± 0.0226
Encoder alignment	E_1	6.1420 ± 0.3685	0.9173 ± 0.0444	0.1751 ± 0.1178	0.7999 ± 0.0689
Encoder alignment	E_{full}	5.1413 ± 0.4240	0.9534 ± 0.0077	0.0737 ± 0.0131	0.8703 ± 0.0149

signal with respect to the corresponding readout vector,

$$J_{i,t} = \left. \frac{\partial D_T(m)}{\partial m_t} \right|_{m=m_{T,i}} \in \mathbb{R}^{L \times d}, \tag{16}$$

where L is the reconstructed signal length. Thus, $J_{i,t}$ characterizes the local effect of perturbing the d-dimensional readout representation at position t on the complete decoder output.

We aggregate these Jacobians across all $N_{\text{val}} = 1980$ MITDB validation samples and all T_m readout positions by vertically stacking them,

$$J_{\text{agg}} = \frac{1}{\sqrt{N_{\text{val}} T_m}} \begin{bmatrix} J_{1,1} \\ \vdots \\ J_{1,T_m} \\ \vdots \\ J_{N_{\text{val}},1} \\ \vdots \\ J_{N_{\text{val}},T_m} \end{bmatrix} \in \mathbb{R}^{(N_{\text{val}} T_m L) \times d}. \tag{17}$$

The corresponding channel-space sensitivity matrix is

$$G = J_{\text{agg}}^\top J_{\text{agg}} = \frac{1}{N_{\text{val}} T_m} \sum_{i=1}^{N_{\text{val}}} \sum_{t=1}^{T_m} J_{i,t}^\top J_{i,t} \in \mathbb{R}^{d \times d}. \tag{18}$$

This aggregation gives each validation sample and readout position equal weight and summarizes average local decoder sensitivity in the shared readout-channel space.

We obtain the decoder-sensitive directions through the singular value decomposition

$$J_{\text{agg}} = U \Sigma V^\top, \qquad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d, \tag{19}$$

where the columns of V define orthogonal directions in readout-channel space. Equivalently,

$$G = V\Sigma^2 V^\top, \tag{20}$$

so these directions are the eigenvectors of G, with eigenvalues $\lambda_j = \sigma_j^2$. Larger λ_j indicate directions for which local readout perturbations induce larger changes in the frozen decoder output on average.

We select the smallest rank r whose cumulative sensitivity accounts for at least 90% of the total sensitivity:

$$r = \min\left\{ k : \frac{\sum_{j=1}^{k} \sigma_j^2}{\sum_{j=1}^{d} \sigma_j^2} \geq 0.90 \right\}. \tag{21}$$

This criterion yields $r = 14$ for LUNet and $r = 3$ for ImputeECG. Let $V_r = [v_1, \ldots, v_r]$ contain the corresponding right singular vectors. We define the decoder-sensitive projector and its orthogonal complement as

$$P = V_r V_r^\top, \qquad Q = I - P. \tag{22}$$

Because $P \in \mathbb{R}^{d\times d}$ acts on the readout-channel dimension, the same projector is applied independently at every readout position.

For a corrupted input, let m_S^F denote the student memory readout before Stage 2 adaptation and m_S^A the readout after adaptation. The adaptation-induced readout change is

$$\delta m = m_S^A - m_S^F. \tag{23}$$

We decompose this change into its decoder-sensitive and complementary components,

$$\delta m_P = P\delta m, \qquad \delta m_Q = Q\delta m, \tag{24}$$

and construct the corresponding intervention readouts as

$$m_P = m_S^F + P\delta m, \qquad m_Q = m_S^F + Q\delta m. \tag{25}$$

The P-only intervention retains only the component of the adaptation-induced change within the decoder-sensitive subspace, whereas the Q-only intervention retains only its orthogonal complement. The memory and decoder remain frozen throughout these interventions.

Finally, we test whether occupying the decoder-sensitive subspace with the same update magnitude is sufficient to reproduce the effect of the original adaptation-induced change. For each projected update, we express its coordinates in the sensitive basis,

$$a = V_r^\top \delta m, \tag{26}$$

and apply a random orthogonal transformation $R \in \mathbb{R}^{r\times r}$, with $R^\top R = I$. The rotated update is

$$\widetilde{\delta m}_P = V_r R V_r^\top \delta m, \tag{27}$$

and the corresponding intervention readout is

$$\widetilde{m}_P = m_S^F + \widetilde{\delta m}_P. \tag{28}$$

Orthogonality preserves the update energy,

$$\left\| \widetilde{\delta m}_P \right\|_2^2 = \|P\delta m\|_2^2, \tag{29}$$

while changing its direction within the same decoder-sensitive subspace. We draw 100 independent random orthogonal rotations for each evaluated setting and average their reconstruction effects. This control separates sensitivity to the subspace itself from sensitivity to the specific direction selected by adaptation within that subspace.

Table A5: **Complete reconstruction effects of memory-access adaptation.** This table expands Table 4.

Dataset	Adaptation	E_P	ΔL_{Full}	ΔL_P	ΔL_Q	R_P	$\Delta L_{\text{Rot},P}$
LUNet	($r = 14, d = 64, r/d = 0.2188$)						
MITDB	E_s	0.2784 ± 0.0491	-0.0604 ± 0.0004	-0.0599 ± 0.0007	-0.0032 ± 0.0036	0.9919 ± 0.0060	0.0107 ± 0.0018
	R_s	0.2749 ± 0.0120	-0.0609 ± 0.0003	-0.0605 ± 0.0001	-0.0012 ± 0.0018	0.9937 ± 0.0023	0.0102 ± 0.0031
	E_s, R_s	0.2810 ± 0.0226	-0.0613 ± 0.0000	-0.0610 ± 0.0001	-0.0022 ± 0.0019	0.9950 ± 0.0030	0.0113 ± 0.0023
QTDB	E_s	0.2724 ± 0.0542	-7.3611 ± 0.0809	-7.2477 ± 0.1550	-0.5005 ± 0.5933	0.9846 ± 0.0103	0.9441 ± 0.2671
	R_s	0.2713 ± 0.0081	-7.3282 ± 0.1356	-7.2681 ± 0.1003	-0.1762 ± 0.3998	0.9918 ± 0.0047	0.8027 ± 0.2507
	E_s, R_s	0.2795 ± 0.0188	-7.4531 ± 0.0345	-7.3908 ± 0.0186	-0.2959 ± 0.4027	0.9917 ± 0.0071	0.9235 ± 0.1725
SIMEMG	E_s	0.1940 ± 0.0110	2.4882 ± 0.1252	3.2461 ± 0.8439	-0.3138 ± 0.3903	1.3148 ± 0.4053	14.8778 ± 3.0275
	R_s	0.2119 ± 0.0119	4.7668 ± 0.7424	5.5421 ± 1.1845	-0.5886 ± 0.3432	1.1574 ± 0.0682	14.5673 ± 4.5053
	E_s, R_s	0.2125 ± 0.0093	6.7974 ± 0.9752	6.7300 ± 0.4036	0.1595 ± 0.4743	0.9960 ± 0.0835	22.8330 ± 15.9924
ImputeECG	($r = 3, d = 64, r/d = 0.0469$)						
MITDB	E_s	0.4077 ± 0.1316	-0.0640 ± 0.0001	-0.0635 ± 0.0004	-0.0025 ± 0.0013	0.9934 ± 0.0047	0.0505 ± 0.0051
	R_s	0.4055 ± 0.1329	-0.0623 ± 0.0001	-0.0619 ± 0.0001	-0.0022 ± 0.0011	0.9932 ± 0.0045	0.0427 ± 0.0045
	E_s, R_s	0.4062 ± 0.1283	-0.0657 ± 0.0001	-0.0653 ± 0.0002	-0.0027 ± 0.0011	0.9934 ± 0.0044	0.0538 ± 0.0057
QTDB	E_s	0.4086 ± 0.1319	-7.8181 ± 0.1011	-7.7443 ± 0.0473	-0.3041 ± 0.1423	0.9906 ± 0.0068	4.6102 ± 0.4817
	R_s	0.4073 ± 0.1337	-7.5700 ± 0.0484	-7.4924 ± 0.0012	-0.2744 ± 0.1351	0.9898 ± 0.0065	3.6837 ± 0.3551
	E_s, R_s	0.4074 ± 0.1298	-7.9812 ± 0.0770	-7.9049 ± 0.0274	-0.3238 ± 0.1213	0.9905 ± 0.0062	4.8259 ± 0.4672
SIMEMG	E_s	0.4006 ± 0.1362	-30.3181 ± 4.2858	-29.6258 ± 5.5314	-2.8760 ± 3.2799	0.9741 ± 0.0448	130.6079 ± 84.7789
	R_s	0.3944 ± 0.1187	-38.6663 ± 1.3312	-37.9956 ± 2.2120	-1.9686 ± 1.8453	0.9823 ± 0.0234	100.4993 ± 59.6119
	E_s, R_s	0.4062 ± 0.1294	-43.1476 ± 3.2106	-42.5785 ± 4.2965	-2.8044 ± 2.8936	0.9859 ± 0.0262	144.8911 ± 95.9570

A.8 Complete Results for Readout Interventions

Table A5 expands the results summarized in Table 4. Both tables aggregate the same three training seeds: the main-paper table reports the means, while this table additionally reports their standard deviations and the individual effects of the P-only and complementary Q-only interventions. All effects are measured relative to frozen access, with negative MSE changes indicating improved reconstruction.

On MITDB and QTDB, the P-only intervention produces mean MSE reductions close to those of full adaptation for both backbones, whereas the Q-only intervention produces substantially smaller effects. On SIMEMG, the P-only intervention also reproduces the direction of the full adaptation effect: reconstruction improves for ImputeECG but worsens for LUNet. The sensitive component therefore reproduces much of the observed reconstruction effect, including degradation under distribution shift.

Energy-preserving random rotations within P produce positive mean MSE changes in every reported setting. Retaining the sensitive subspace and readout-update energy therefore does not preserve the original reconstruction effect; the direction of the update within that subspace also matters.

A.9 Output-Space Decomposition and Directional Controls

We next examine how the decoder-sensitive readout intervention defined in Appendix A.7 changes reconstruction after passing through the frozen decoder. Let m_P denote the P-only intervention and

$$\hat{x}_P = D(m_P)$$

its decoded output. To characterize whether this intervention improves or degrades reconstruction, let

$$e_F = \hat{x}_F - x_C, \qquad u_P = \hat{x}_P - \hat{x}_F, \tag{30}$$

where e_F is the frozen-access reconstruction residual and u_P is the output change induced by the decoder-sensitive readout component. The residual after intervention is therefore $e_F + u_P$, and its change in physical-scale MSE satisfies

$$\Delta L_P = L(\hat{x}_P, x_C) - L(\hat{x}_F, x_C) = 2\langle e_F, u_P \rangle + \|u_P\|_2^2, \tag{31}$$

where the inner product and squared norm use the same averaging as the MSE. The first term measures the interaction between the induced output change and the existing reconstruction error: a negative value indicates that the change opposes the frozen residual. The second term measures the magnitude of the induced output change and is nonnegative. Reconstruction improves only when the negative residual interaction exceeds this energy term.

Table A6: **Output-space decomposition of the P-only intervention.** The corresponding intervention effects are reported in Table A5.

Dataset	Adaptation	Residual interaction $2\langle e_F, u_P\rangle$	Output-change energy $\|u_P\|^2$	$\cos(e_F, u_P)$
LUNet				
MITDB	E_s	-0.1109 ± 0.0062	0.0510 ± 0.0054	-0.7378 ± 0.0086
	R_s	-0.1101 ± 0.0034	0.0496 ± 0.0033	-0.7421 ± 0.0018
	E_s, R_s	-0.1162 ± 0.0009	0.0551 ± 0.0006	-0.7440 ± 0.0015
QTDB	E_s	-11.5983 ± 0.7384	4.3506 ± 0.5834	-0.8165 ± 0.0058
	R_s	-11.3829 ± 0.4555	4.1148 ± 0.3552	-0.8169 ± 0.0022
	E_s, R_s	-11.9303 ± 0.0424	4.5396 ± 0.0238	-0.8207 ± 0.0011
SIMEMG	E_s	-7.5941 ± 0.0952	10.8402 ± 0.9390	-0.2094 ± 0.0165
	R_s	-8.1819 ± 2.6258	13.7240 ± 1.4412	-0.1844 ± 0.0674
	E_s, R_s	-9.4910 ± 2.0314	16.2210 ± 1.6278	-0.2034 ± 0.0455
ImputeECG				
MITDB	E_s	-0.1257 ± 0.0021	0.0622 ± 0.0024	-0.7518 ± 0.0010
	R_s	-0.1101 ± 0.0001	0.0482 ± 0.0003	-0.7318 ± 0.0005
	E_s, R_s	-0.1290 ± 0.0002	0.0637 ± 0.0001	-0.7633 ± 0.0037
QTDB	E_s	-13.4830 ± 0.4745	5.7387 ± 0.4272	-0.8147 ± 0.0001
	R_s	-11.5047 ± 0.0626	4.0123 ± 0.0636	-0.8040 ± 0.0001
	E_s, R_s	-13.6316 ± 0.0635	5.7268 ± 0.0361	-0.8261 ± 0.0013
SIMEMG	E_s	-84.3381 ± 5.7986	54.7123 ± 11.3300	-0.5675 ± 0.0239
	R_s	-72.9364 ± 2.0244	34.9408 ± 0.1876	-0.6002 ± 0.0190
	E_s, R_s	-94.4217 ± 4.6610	51.8432 ± 8.9573	-0.6563 ± 0.0281

Table A6 shows that this beneficial balance holds for all three adaptation scopes on MITDB and QTDB with both backbones. For example, LUNet joint adaptation on QTDB produces a residual interaction of -11.9303 and an output-change energy of 4.5396, yielding a mean MSE reduction. LUNet on SIMEMG exhibits the opposite balance. Under joint adaptation, the residual interaction remains negative at -9.4910, but the output-change energy reaches 16.2210, resulting in a mean MSE increase of 6.7300. Thus, the output change contains a corrective component but still worsens reconstruction overall. ImputeECG joint adaptation on SIMEMG retains the beneficial balance: its residual interaction of -94.4217 outweighs its output-change energy of 51.8432, reducing mean MSE by 42.5785.

We additionally test whether the direction of the readout update matters by randomly rotating its sensitive component within P while preserving its readout-space energy. The rotated intervention effects are reported in Table A5. Random rotations produce positive mean MSE changes in every reported setting. For ImputeECG joint adaptation on QTDB, for example, the mean rotated effect is 4.8259, compared with -7.9049 for the original P-only update. Preserving the sensitive subspace and readout-update energy therefore does not preserve the reconstruction benefit. Because these rotations can change both the direction and magnitude of the decoded output change, they test direction dependence in readout space rather than output orientation alone.

Together, these results show that a decoder-sensitive readout change is useful when its decoded effect counteracts the existing reconstruction error sufficiently relative to its magnitude. A change can oppose the residual yet still increase MSE, and an equally energetic change within the same sensitive subspace can have a different reconstruction effect. Effective adaptation therefore depends on the specific correction produced through the decoder, beyond the amount of readout change placed in sensitive directions.

A.10 Teacher-Reference Readout Errors and Reconstruction

We examine whether a student memory readout that is closer to the clean-teacher readout produces better reconstruction. For a paired corrupted and clean input, let m_S denote the student readout and

Table A7: **Teacher-reference readout errors and reconstruction.** Total and decoder-sensitive readout errors measure discrepancy from the paired clean-input teacher readout, before and after access adaptation. P is estimated from clean MITDB validation data; r and d denote sensitive and total channel dimensions.

Dataset	Adaptation	RMSE↓	ϵ_{total}↓	ϵ_{sens}↓
LUNet ($r = 14, d = 64, r/d = 0.2188$)				
MITDB (ID)	Frozen access	0.2217 ± 0.0009	0.0953 ± 0.0104	0.1031 ± 0.0186
	E_s	0.1229 ± 0.0025	$\mathbf{0.0804 \pm 0.0053}$	$\mathbf{0.0781 \pm 0.0165}$
	R_s	0.1219 ± 0.0013	0.0876 ± 0.0047	0.0817 ± 0.0162
	E_s, R_s	$\mathbf{0.1214 \pm 0.0005}$	0.0864 ± 0.0063	0.0816 ± 0.0167
QTDB (OOD)	Frozen access	1.4701 ± 0.0063	0.1190 ± 0.0136	0.1301 ± 0.0247
	E_s	0.5684 ± 0.0210	$\mathbf{0.0940 \pm 0.0076}$	$\mathbf{0.0892 \pm 0.0191}$
	R_s	0.5735 ± 0.0307	0.1026 ± 0.0071	0.0933 ± 0.0172
	E_s, R_s	$\mathbf{0.5631 \pm 0.0120}$	0.1006 ± 0.0078	0.0928 ± 0.0187
SIMEMG (OOD)	Frozen access	$\mathbf{4.5787 \pm 0.1267}$	0.0782 ± 0.0205	0.0686 ± 0.0144
	E_s	4.9811 ± 0.0380	$\mathbf{0.0667 \pm 0.0121}$	$\mathbf{0.0590 \pm 0.0161}$
	R_s	5.1458 ± 0.1519	0.0741 ± 0.0045	0.0639 ± 0.0115
	E_s, R_s	5.2883 ± 0.2274	0.0732 ± 0.0066	0.0624 ± 0.0093
ImputeECG ($r = 3, d = 64, r/d = 0.0469$)				
MITDB (ID)	Frozen access	0.2285 ± 0.0002	0.0524 ± 0.0135	0.3331 ± 0.0210
	E_s	0.1271 ± 0.0003	0.0325 ± 0.0075	0.2036 ± 0.0160
	R_s	0.1341 ± 0.0013	0.0339 ± 0.0077	0.2121 ± 0.0174
	E_s, R_s	$\mathbf{0.1225 \pm 0.0011}$	$\mathbf{0.0314 \pm 0.0072}$	$\mathbf{0.1958 \pm 0.0150}$
QTDB (OOD)	Frozen access	1.5262 ± 0.0010	0.0652 ± 0.0160	0.4379 ± 0.0346
	E_s	0.5948 ± 0.0155	0.0346 ± 0.0077	0.2290 ± 0.0224
	R_s	0.6603 ± 0.0082	0.0369 ± 0.0083	0.2424 ± 0.0233
	E_s, R_s	$\mathbf{0.5595 \pm 0.0124}$	$\mathbf{0.0330 \pm 0.0075}$	$\mathbf{0.2166 \pm 0.0192}$
SIMEMG (OOD)	Frozen access	8.2904 ± 0.4913	0.0674 ± 0.0321	0.3653 ± 0.0671
	E_s	6.5970 ± 0.2190	0.0377 ± 0.0151	0.2028 ± 0.0230
	R_s	6.1101 ± 0.1115	0.0405 ± 0.0159	0.2129 ± 0.0340
	E_s, R_s	$\mathbf{5.8433 \pm 0.3141}$	$\mathbf{0.0369 \pm 0.0159}$	$\mathbf{0.1907 \pm 0.0274}$

m_T the clean-teacher readout. Their difference, $e = m_S - m_T$, measures disagreement between the representations supplied to the decoder. We quantify this disagreement using the normalized total readout error,

$$\epsilon_{\text{total}} = \frac{\|m_S - m_T\|_2}{\|m_T\|_2}. \tag{32}$$

This quantity measures discrepancy in readout space, whereas reconstruction RMSE measures error between the decoded signal and the clean ECG.

Readout differences need not affect the decoder equally: a perturbation along one direction may substantially change the reconstructed signal, while a perturbation of the same magnitude along another direction may have little effect. We therefore also measure discrepancy within an estimated decoder-sensitive subspace, comprising readout directions to which the decoder output is most responsive. This subspace is estimated from the decoder's local responses around clean-teacher readouts on MITDB validation data. With P denoting the orthogonal projector onto this subspace, the normalized sensitive readout error is

$$\epsilon_{\text{sens}} = \frac{\|P(m_S - m_T)\|_2}{\|Pm_T\|_2}. \tag{33}$$

Table A7 compares both readout errors with reconstruction RMSE to examine whether teacher agreement within these sensitive directions corresponds more consistently to better reconstruction.

For LUNet, the configurations with the lowest mean readout errors do not consistently achieve the lowest mean RMSE. On MITDB and QTDB, encoder-only adaptation achieves the lowest mean total and sensitive readout errors, whereas joint adaptation achieves the lowest mean RMSE. On

Table A8: **Teacher–student agreement in memory access.** Restoration RMSE is reported alongside query cosine, routing KL, and chance-corrected Top-K overlap relative to the clean-input teacher.

Dataset	Adaptation	Trainable	RMSE↓	Query Cos.↑	Routing KL↓	Adj. Top-K↑
LUNet						
MITDB (ID)	Frozen access	–	0.2217 ± 0.0009	0.8811 ± 0.0071	0.1479 ± 0.0200	0.8130 ± 0.0089
	Encoder adaptation	E_s	0.1229 ± 0.0025	$\mathbf{0.9286 \pm 0.0097}$	$\mathbf{0.1018 \pm 0.0118}$	$\mathbf{0.8410 \pm 0.0156}$
	Router adaptation	R_s	0.1219 ± 0.0013	0.8906 ± 0.0175	0.1274 ± 0.0091	0.8109 ± 0.0090
	Joint adaptation	E_s, R_s	$\mathbf{0.1214 \pm 0.0005}$	0.8928 ± 0.0262	0.1311 ± 0.0162	0.8161 ± 0.0103
QTDB (OOD)	Frozen access	–	1.4701 ± 0.0063	0.8081 ± 0.0117	0.2199 ± 0.0327	0.7389 ± 0.0133
	Encoder adaptation	E_s	0.5684 ± 0.0210	$\mathbf{0.9098 \pm 0.0108}$	$\mathbf{0.1365 \pm 0.0179}$	$\mathbf{0.7975 \pm 0.0279}$
	Router adaptation	R_s	0.5735 ± 0.0307	0.8607 ± 0.0226	0.1674 ± 0.0181	0.7610 ± 0.0178
	Joint adaptation	E_s, R_s	$\mathbf{0.5631 \pm 0.0120}$	0.8697 ± 0.0304	0.1736 ± 0.0234	0.7719 ± 0.0172
SIMEMG (OOD)	Frozen access	–	$\mathbf{4.5787 \pm 0.1267}$	$\mathbf{0.9683 \pm 0.0090}$	0.0885 ± 0.0318	0.8692 ± 0.0292
	Encoder adaptation	E_s	4.9811 ± 0.0380	0.9636 ± 0.0130	$\mathbf{0.0647 \pm 0.0200}$	$\mathbf{0.8811 \pm 0.0159}$
	Router adaptation	R_s	5.1458 ± 0.1519	0.9426 ± 0.0136	0.0865 ± 0.0073	0.8582 ± 0.0041
	Joint adaptation	E_s, R_s	5.2883 ± 0.2274	0.9311 ± 0.0216	0.0908 ± 0.0098	0.8590 ± 0.0071
ImputeECG						
MITDB (ID)	Frozen access	–	0.2285 ± 0.0002	0.9781 ± 0.0197	0.0514 ± 0.0398	0.9365 ± 0.0554
	Encoder adaptation	E_s	0.1271 ± 0.0003	$\mathbf{0.9918 \pm 0.0071}$	$\mathbf{0.0182 \pm 0.0124}$	$\mathbf{0.9601 \pm 0.0400}$
	Router adaptation	R_s	0.1341 ± 0.0013	0.8853 ± 0.0674	0.0350 ± 0.0205	0.8904 ± 0.0885
	Joint adaptation	E_s, R_s	$\mathbf{0.1225 \pm 0.0011}$	0.9230 ± 0.0551	0.0225 ± 0.0133	0.9227 ± 0.0656
QTDB (OOD)	Frozen access	–	1.5262 ± 0.0010	0.9660 ± 0.0301	0.0824 ± 0.0644	0.9246 ± 0.0608
	Encoder adaptation	E_s	0.5948 ± 0.0155	$\mathbf{0.9907 \pm 0.0074}$	$\mathbf{0.0209 \pm 0.0130}$	$\mathbf{0.9613 \pm 0.0365}$
	Router adaptation	R_s	0.6603 ± 0.0082	0.8773 ± 0.0751	0.0419 ± 0.0267	0.8875 ± 0.0896
	Joint adaptation	E_s, R_s	$\mathbf{0.5595 \pm 0.0124}$	0.9202 ± 0.0583	0.0250 ± 0.0147	0.9230 ± 0.0638
SIMEMG (OOD)	Frozen access	–	8.2904 ± 0.4913	0.9838 ± 0.0116	0.0491 ± 0.0366	0.9328 ± 0.0479
	Encoder adaptation	E_s	6.5970 ± 0.2190	$\mathbf{0.9942 \pm 0.0041}$	$\mathbf{0.0161 \pm 0.0111}$	$\mathbf{0.9633 \pm 0.0363}$
	Router adaptation	R_s	6.1101 ± 0.1115	0.8998 ± 0.0561	0.0310 ± 0.0165	0.8854 ± 0.0795
	Joint adaptation	E_s, R_s	$\mathbf{5.8433 \pm 0.3141}$	0.9253 ± 0.0489	0.0209 ± 0.0125	0.9152 ± 0.0713

QTDB, for example, encoder-only adaptation yields a sensitive readout error of 0.0892, compared with 0.0928 under joint adaptation, while the corresponding RMSE values are 0.5684 and 0.5631. On SIMEMG, encoder-only adaptation reduces mean sensitive readout error from 0.0686 under frozen access to 0.0590, yet increases mean RMSE from 4.5787 to 4.9811. Thus, moving closer to the teacher within the estimated sensitive subspace can accompany worse reconstruction.

For ImputeECG, joint adaptation achieves the lowest mean total and sensitive readout errors together with the lowest mean RMSE on all three datasets. However, this correspondence does not hold for every comparison. On SIMEMG, router-only adaptation yields a higher mean sensitive readout error than encoder-only adaptation, 0.2129 versus 0.2028, but a lower mean RMSE, 6.1101 versus 6.5970. Teacher agreement therefore accompanies the best reconstruction in some comparisons but does not consistently rank the adaptation configurations by reconstruction quality.

These results show that closer agreement with the clean-teacher readout does not necessarily yield better ECG reconstruction, even when agreement is measured only within the estimated decoder-sensitive subspace. Teacher-reference readout error measures how closely adaptation reproduces the teacher's representation, but a lower value does not by itself indicate a more accurate restored signal.

A.11 Teacher–Student Agreement in Memory Access

We examine whether closer agreement with clean-teacher memory access corresponds to better ECG reconstruction. For each paired corrupted and clean input, we compare student and teacher access using three diagnostics. Query cosine measures the directional similarity of their query vectors, routing KL divergence measures the discrepancy between their distributions of weights over memory slots, and chance-corrected Top-K overlap measures agreement between their selected slots after accounting for overlap expected by chance. Higher query cosine and overlap, and lower routing KL, indicate closer teacher agreement. Table A8 reports these diagnostics alongside reconstruction RMSE for encoder-only, router-only, and joint adaptation with both backbones. Frozen access denotes the unadapted pathway evaluated on corrupted input. The curated memory and decoder remain frozen across all configurations.

On MITDB and QTDB, encoder-only adaptation achieves the strongest mean teacher agreement on all three diagnostics for both backbones, whereas joint adaptation achieves the lowest mean RMSE. For ImputeECG on QTDB, encoder-only adaptation achieves a mean query cosine of 0.9907, compared with 0.9202 under joint adaptation, and also achieves lower routing KL and higher adjusted Top-K overlap. Nevertheless, joint adaptation yields lower mean RMSE, 0.5595 versus 0.5948. LUNet exhibits the same ordering on both datasets. Thus, the additional reconstruction benefit of joint adaptation does not require closer agreement with the teacher on these access diagnostics.

SIMEMG reveals a further distinction between routing agreement and reconstruction. For LUNet, encoder-only adaptation reduces mean routing KL from 0.0885 under frozen access to 0.0647 and increases adjusted Top-K overlap from 0.8692 to 0.8811, while mean RMSE increases from 4.5787 to 4.9811. Better agreement in routing weights and selected slots therefore accompanies worse reconstruction in this comparison. For ImputeECG, encoder-only adaptation again achieves the strongest mean agreement on all three diagnostics, whereas joint adaptation achieves the lowest mean RMSE. The separation between the best teacher agreement and the best reconstruction therefore occurs across all three datasets for this backbone.

These results show that the adaptation configuration that most closely reproduces clean-teacher memory access does not consistently reconstruct the ECG most accurately. Query cosine, routing KL, and slot overlap describe how faithfully the student reproduces the teacher's access pattern, but their best mean values do not consistently identify the lowest reconstruction error.

A.12 Data Efficiency of Memory Curation and Memory-Access Alignment

We examine whether memory curation and memory-access alignment differ in their sensitivity to training-data availability. Using LUNet with joint encoder-bottleneck and router adaptation (E_s, R_s), we independently vary the fraction of MITDB training data used for Stage 1 memory curation and Stage 2 memory-access alignment, keeping the other stage at 100%. The curated memory and decoder remain frozen during Stage 2. Table A9 reports reconstruction and access diagnostics at 100%, 75%, 50%, and 25% data availability. This design evaluates reductions in each stage separately, rather than simultaneous reductions in both stages.

With memory curation fixed at 100%, reducing access-alignment data to 25% increases mean RMSE by 5.8% on MITDB, 8.4% on QTDB, and 3.6% on SIMEMG, relative to using 100% data in both stages. Conversely, with access alignment fixed at 100%, reducing memory-curation data to 25% increases mean RMSE by 7.2%, 8.7%, and 32.0%, respectively. At this reduction level, the two stages show similar reconstruction sensitivity on MITDB and QTDB, whereas SIMEMG is substantially more sensitive to reduced memory-curation data. The mean results are not strictly monotonic with data availability: using 75% of the memory-curation data yields the lowest mean QTDB RMSE of 0.5524.

These results indicate that joint memory-access alignment retains much of the full-data reconstruction performance with only 25% of the alignment data, provided that memory curation uses the full training set. The relative data requirements of the two stages nevertheless depend on the evaluation distribution: reducing memory-curation data produces a comparable RMSE increase on MITDB and QTDB but a substantially larger increase on SIMEMG. Thus, the greater sensitivity to memory-curation data is specific to the real-contamination shift in these experiments, rather than a uniform difference between the stages.

A.13 Disentangling Alignment and Reconstruction Objectives

We examine the contributions of reconstruction supervision and memory-access alignment during Stage 2. For each seed, all variants are initialized from the same Stage 1 checkpoint and update the same encoder bottleneck and student router, while the curated memory, decoder, and remaining encoder blocks remain frozen. We compare reconstruction-only training using $\mathcal{L}_{\mathrm{rec}}$, alignment-only training using $\mathcal{L}_{\mathrm{align}} = \mathcal{L}_{\mathrm{latent}} + \mathcal{L}_{\mathrm{readout}}$, and the full MAA objective combining both terms.

Table A10 shows that the full objective achieves the lowest mean RMSE for both backbones on all three datasets. Relative to reconstruction-only adaptation, full MAA reduces mean RMSE from 0.5755 to 0.5631 on QTDB and from 5.4509 to 5.2883 on SIMEMG for LUNet. The reduction is larger for ImputeECG on QTDB, where mean RMSE decreases from 0.6619 to 0.5595. Alignment-

Table A9: Data efficiency of memory curation and memory-access alignment.

Memory Curation	Access Alignment	RMSE↓	Query Cos.↑	Routing KL↓	Adj. Top-K↑
MITDB (ID)					
100%	100%	$\mathbf{0.1214 \pm 0.0005}$	0.8928 ± 0.0262	0.1311 ± 0.0162	$\mathbf{0.8161 \pm 0.0103}$
100%	75%	0.1230 ± 0.0006	0.8544 ± 0.0475	0.1505 ± 0.0205	0.7965 ± 0.0094
100%	50%	0.1261 ± 0.0028	0.9020 ± 0.0084	0.1368 ± 0.0129	0.8023 ± 0.0234
100%	25%	0.1284 ± 0.0027	$\mathbf{0.9035 \pm 0.0069}$	0.1373 ± 0.0133	0.8034 ± 0.0150
75%	100%	0.1216 ± 0.0012	0.8763 ± 0.0132	0.1357 ± 0.0163	0.7942 ± 0.0117
50%	100%	0.1242 ± 0.0007	0.8672 ± 0.0066	0.1305 ± 0.0233	0.7690 ± 0.0320
25%	100%	0.1302 ± 0.0012	0.8417 ± 0.0158	$\mathbf{0.1207 \pm 0.0135}$	0.7310 ± 0.0160
QTDB (OOD)					
100%	100%	0.5631 ± 0.0120	0.8697 ± 0.0304	0.1736 ± 0.0234	$\mathbf{0.7719 \pm 0.0172}$
100%	75%	0.5908 ± 0.0191	0.8346 ± 0.0384	0.1928 ± 0.0143	0.7550 ± 0.0161
100%	50%	0.6217 ± 0.0879	0.8760 ± 0.0081	0.1819 ± 0.0158	0.7476 ± 0.0340
100%	25%	0.6102 ± 0.1082	$\mathbf{0.8787 \pm 0.0020}$	0.1799 ± 0.0110	0.7527 ± 0.0231
75%	100%	$\mathbf{0.5524 \pm 0.0252}$	0.8460 ± 0.0094	0.1785 ± 0.0233	0.7607 ± 0.0143
50%	100%	0.5691 ± 0.0151	0.8316 ± 0.0100	0.1721 ± 0.0365	0.7285 ± 0.0417
25%	100%	0.6123 ± 0.0169	0.7929 ± 0.0163	$\mathbf{0.1677 \pm 0.0265}$	0.6787 ± 0.0384
SIMEMG (OOD)					
100%	100%	$\mathbf{5.2883 \pm 0.2274}$	0.9311 ± 0.0216	0.0908 ± 0.0098	$\mathbf{0.8590 \pm 0.0071}$
100%	75%	5.7023 ± 0.3736	0.8886 ± 0.0554	0.1167 ± 0.0260	0.8299 ± 0.0204
100%	50%	5.5185 ± 0.1727	0.9412 ± 0.0105	0.1010 ± 0.0201	0.8491 ± 0.0116
100%	25%	5.4799 ± 0.1614	$\mathbf{0.9471 \pm 0.0057}$	0.0848 ± 0.0031	0.8572 ± 0.0032
75%	100%	5.5243 ± 0.0543	0.9319 ± 0.0167	0.0788 ± 0.0197	0.8398 ± 0.0063
50%	100%	5.9789 ± 0.1981	0.9265 ± 0.0130	0.0670 ± 0.0151	0.8294 ± 0.0231
25%	100%	6.9809 ± 0.3916	0.9114 ± 0.0141	$\mathbf{0.0630 \pm 0.0092}$	0.7967 ± 0.0191

only training also achieves lower mean RMSE than reconstruction-only training in four of the six backbone–dataset settings, although the preferred individual objective depends on the backbone and dataset.

These results support combining output-space reconstruction supervision with alignment to the clean latent and memory-readout representations established during Stage 1. The combined objective achieves lower mean RMSE than either objective alone across all six evaluated settings, although some differences are small relative to the reported seed variability. Secondary metrics on SIMEMG do not uniformly favor the combined objective, so the evidence supports an additional benefit from alignment in mean RMSE rather than improvement across every restoration metric.

Table A10: **Alignment and reconstruction objectives in Stage 2.** We compare reconstruction-only, alignment-only, and combined training, where $\mathcal{L}_{\text{align}} = \mathcal{L}_{\text{latent}} + \mathcal{L}_{\text{readout}}$. Results are reported as mean ± standard deviation across three random seeds. Bold indicates the best mean within each backbone–dataset group.

Backbone	$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{align}}$	RMSE↓	PRD↓	ImSNR↑	CC↑
MITDB (ID)						
LUNet	✓		0.1218 ± 0.0012	29.1164 ± 0.2019	4.2527 ± 0.0851	0.8997 ± 0.0017
		✓	0.1248 ± 0.0025	29.2366 ± 0.1445	4.2028 ± 0.0339	0.8969 ± 0.0042
	✓	✓	$\mathbf{0.1214 \pm 0.0005}$	$\mathbf{29.0907 \pm 0.0514}$	$\mathbf{4.2679 \pm 0.0382}$	$\mathbf{0.9012 \pm 0.0010}$
ImputeECG	✓		0.1332 ± 0.0015	31.3379 ± 0.3234	3.5147 ± 0.0980	0.8809 ± 0.0018
		✓	0.1240 ± 0.0010	29.3607 ± 0.2201	4.2485 ± 0.0851	0.8902 ± 0.0025
	✓	✓	$\mathbf{0.1225 \pm 0.0011}$	$\mathbf{29.0562 \pm 0.2599}$	$\mathbf{4.3378 \pm 0.1026}$	$\mathbf{0.8936 \pm 0.0024}$
QTDB (Zero-shot OOD)						
LUNet	✓		0.5755 ± 0.0277	24.9467 ± 0.5977	5.9994 ± 0.2657	0.6739 ± 0.0046
		✓	0.5688 ± 0.0116	24.9550 ± 0.7880	6.0989 ± 0.1702	0.6726 ± 0.0100
	✓	✓	$\mathbf{0.5631 \pm 0.0120}$	$\mathbf{24.6944 \pm 0.3399}$	$\mathbf{6.1233 \pm 0.1653}$	$\mathbf{0.6800 \pm 0.0054}$
ImputeECG	✓		0.6619 ± 0.0173	27.9826 ± 0.4006	4.9047 ± 0.1355	0.6553 ± 0.0033
		✓	0.5717 ± 0.0159	25.0384 ± 0.2820	6.1126 ± 0.1098	0.6702 ± 0.0011
	✓	✓	$\mathbf{0.5595 \pm 0.0124}$	$\mathbf{24.6639 \pm 0.1984}$	$\mathbf{6.2472 \pm 0.0799}$	$\mathbf{0.6791 \pm 0.0022}$
SIMEMG (Zero-shot OOD)						
LUNet	✓		5.4509 ± 0.0968	23.4256 ± 0.8078	4.1895 ± 0.2372	0.9708 ± 0.0021
		✓	5.3325 ± 0.1084	$\mathbf{23.3473 \pm 0.5077}$	$\mathbf{4.2953 \pm 0.2983}$	$\mathbf{0.9743 \pm 0.0011}$
	✓	✓	$\mathbf{5.2883 \pm 0.2274}$	23.4317 ± 0.9488	4.2531 ± 0.3896	0.9725 ± 0.0015
ImputeECG	✓		5.9060 ± 0.1590	26.4841 ± 0.7988	3.5564 ± 0.2128	$\mathbf{0.9622 \pm 0.0039}$
		✓	6.0006 ± 0.3692	27.2739 ± 1.8015	3.6146 ± 0.4630	0.9567 ± 0.0094
	✓	✓	$\mathbf{5.8433 \pm 0.3141}$	$\mathbf{26.4093 \pm 1.4555}$	$\mathbf{3.7349 \pm 0.4621}$	0.9602 ± 0.0050