STRIDEDIFFUSION: ACCELERATING DIFFUSION MODELS FOR TIME-SERIES GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

StrideDiffusion accelerates time-series diffusion without retraining by using spectral activity to choose denoising times. The sampler tracks band power and changes in power and phase, uses a history correction on eligible jumps, and returns to fine steps near the clean endpoint. We test whether its gains come from grid placement rather than the update rule alone. On six datasets, we select configurations on validation data and compare them on separate data using three independently trained models per dataset. A uniform-grid replay matches every batch's denoiser call count and the choice between a history correction and a deterministic DDIM update at each update index. The spectral grid has lower mean Context-FID, a feature-based distribution distance, on five datasets, with ETTh as the exception. The complete sampler also has lower mean Context-FID than DPM-Solver++ on five datasets and UniPC on all six at approximately matched call budgets, and is faster than both in synchronized sampling measurements. These gains do not extend to every quality metric. Our analysis separates DDIM displacement from the history residual and gives a local sufficient condition for DDIM. The implemented activity gate does not certify that condition. Experiments on additional backbones test spectral scheduling in other parameterizations. The results support spectral grid selection with a history update, while leaving the necessity of online adaptation open. Code repository: `https://anonymous.4open.science/r/stridediff-ts`.

1 INTRODUCTION

Diffusion models generate time series by repeatedly refining a noisy sequence. They support unconditional synthesis and conditional tasks such as imputation and forecasting (Yuan & Qiao, 2024; Tashiro et al., 2021; Rasul et al., 2021). Their inference cost comes from repeated neural-network evaluations. Reducing this cost without retraining is useful when a pretrained generator must serve different latency budgets.

Fast solvers reduce the number of evaluations through larger integration steps (Song et al., 2021; Lu et al., 2022; 2025; Zhao et al., 2023). Their performance depends on both the update formula and the placement of the time points. For time series, the spectrum provides an additional source of information: trends, periodic components, and short-scale variation need not change equally along a denoising trajectory. This suggests choosing the next time point from the current signal rather than fixing the entire grid in advance.

We propose StrideDiffusion, a training-free sampler that uses spectral activity to choose its next denoising time. Figure 1 illustrates the idea. The method monitors band power, power drift, and phase change between visited states. It chooses among three strides, uses a history correction on eligible jumps, and returns to fine steps near the clean endpoint. The controller and update rule together define the sampler. We compare this complete method with fast solvers and test grid selection separately under fixed update rules.

Bandwise changes motivate this controller, but we do not assume a universal order in which frequency bands converge. Diffusion-TS itself includes frequency-aware modeling and a Fourier loss (Yuan & Qiao, 2024). These choices can affect its reverse trajectories. We consequently test transfers to other backbones and distinguish successful sampling from a claim about a shared spectral mechanism.

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.

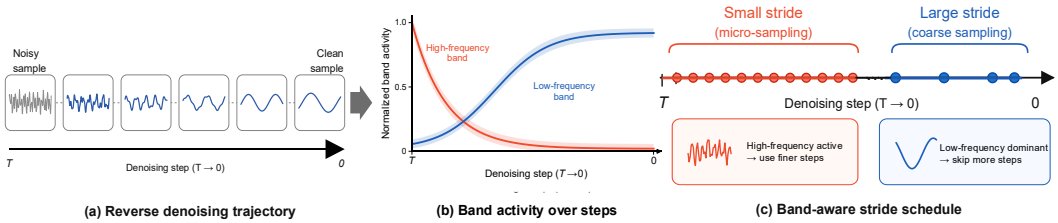

Figure 1: Concept of band-aware sampling. (a) Reverse denoising from noise to a clean sequence. (b) Different bands can change at different rates. (c) Finer steps resolve active high-frequency components, while larger steps skip quieter parts of the trajectory. The curves and sample positions are schematic. The implemented controller also enforces a final fine-step window.

Our analysis expresses each band's single-step displacement through the current state, the model's clean prediction, and an explicit history residual. This decomposition gives a local sufficient condition for DDIM and shows what must also be controlled for a history-corrected jump. The practical gate measures changes between visited states, which alone do not bound these terms. We therefore use the analysis to explain the quantities involved in a jump and experiments to assess the implemented controller.

Our main comparison separates configuration selection from evaluation. For each of six datasets, we select spectral and uniform-history settings on a validation model and compare the fixed choices on three independently trained models. The spectral grid has lower mean Context-FID on five datasets when uniform replay matches both call counts and the update choice at every index. The complete sampler improves mean Context-FID over DPM-Solver++ on five datasets and UniPC on six at approximately matched call budgets. The matched-replay comparison identifies a benefit of grid placement for the implemented history rule, not a solver-independent benefit. Separate experiments test grids from other noise draws, threshold reuse, high-frequency power, additional backbones, and longer sequences.

2 Related Work

Time-series diffusion. Diffusion-TS uses temporal decomposition and a frequency-aware objective (Yuan & Qiao, 2024). CSDI and DiffTime study conditional or constrained generation (Tashiro et al., 2021; Coletta et al., 2023), while DiffWave provides a convolutional diffusion architecture (Kong et al., 2021). ImagenTime uses image transforms for short and long sequences (Naiman et al., 2024). ImagenFew studies few-shot adaptation of pretrained generators (Gonen et al., 2025). Latent-to-data cascaded diffusion separates latent generation from data refinement (Shen et al., 2026). We use Diffusion-TS for controlled sampling experiments and the latter three models to test compatibility. Stochastic-process diffusion, log-signature models, and Latent Laplace Diffusion address continuous or irregular time representations (Biloš et al., 2023; Barancikova et al., 2025; You et al., 2026). Our FFT controller assumes a regular sampling grid.

Solvers and spectral scheduling. DDIM, DPM-Solver++, and UniPC reduce diffusion evaluations without retraining (Song et al., 2021; Lu et al., 2025; Zhao et al., 2023). Spectral analyses motivate frequency-aware sampling (Rissanen et al., 2023; Yang et al., 2023; Lee et al., 2025). Zheng et al. (2026) optimize a fixed schedule using a frequency-weighted objective and a precomputed training-data spectrum. We instead use the evolving sample to choose the next time, coupled to a particular update rule. Our controls test this coupling rather than assuming that spectral schedules help every solver.

Caching and other acceleration routes. SeaCache filters input features in the spectral domain and uses accumulated feature change to decide when cached outputs can be reused (Chung et al., 2026). It saves denoiser computation while retaining the underlying sampler grid. StrideDiffusion changes the next diffusion time and does not cache denoiser outputs for reuse at skipped times. SeaCache incurs cache storage and reuse error. Our method incurs FFT overhead and a risk of inaccurate large

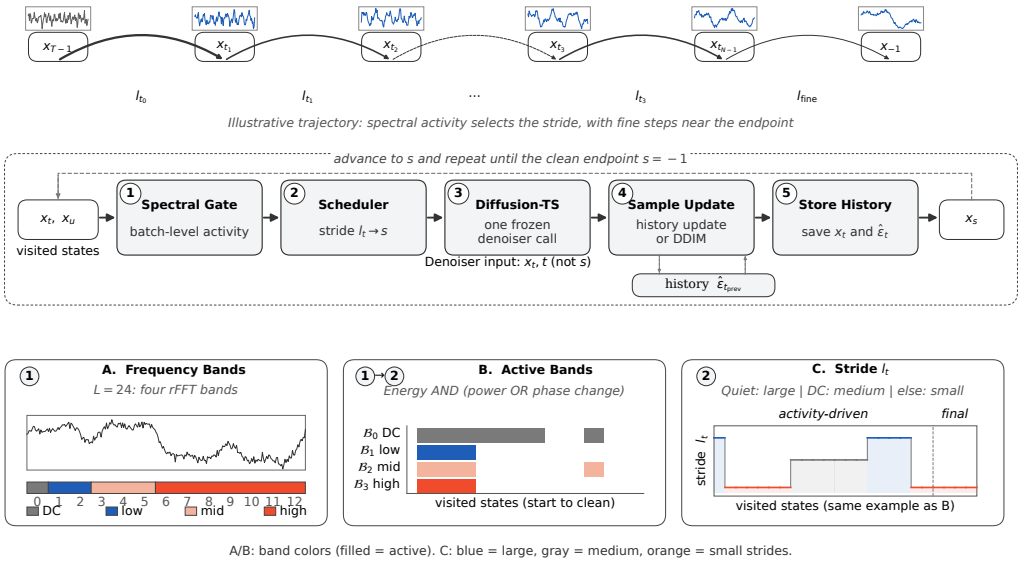

Figure 2: StrideDiffusion sampling framework. Top: an illustrative reverse trajectory. Middle: consecutive visited states determine the stride, followed by one frozen denoiser call, a history update or DDIM fallback, and history storage. Bottom: the $L = 24$ band partition, a schematic binary activity pattern, and its corresponding stride choices. Filled cells denote active bands. The final window overrides activity with fine steps. Waveforms and activity patterns are schematic, not measured results. The history branch requires previous noise, $t - s > 1$, and $s \geq 0$. The first and terminal transitions use DDIM. No spectral projection is applied.

jumps. Both are training-free, but their spectral signals and decisions differ. We make no empirical claim of superiority over SeaCache. Distillation trains a new model (Salimans & Ho, 2022), while quantization changes arithmetic precision (He et al., 2024). These address different parts of the inference cost and are not included in our fixed-checkpoint sampler comparisons.

3 StrideDiffusion

Figure 2 shows the implemented sampling loop.

3.1 Pretrained denoiser and update rule

Let $x_t \in \mathbb{R}^{L\times d}$ be a noisy sequence, with $\rho_t = \bar{\alpha}_t$ and t decreasing towards the data endpoint. We use $0 < \rho_{T-1} < \cdots < \rho_0 < 1$ and the clean endpoint $\rho_{-1} = 1$. The forward marginal is $x_t = \sqrt{\rho_t}x_0 + \sqrt{1-\rho_t}\epsilon$. Diffusion-TS predicts the clean signal. Our implementation clips that prediction to $[-1, 1]$ and reconstructs the noise estimate as

$$\hat{\epsilon}_t = \frac{x_t - \sqrt{\rho_t}\hat{x}_{0,t}}{\sqrt{1-\rho_t}}. \tag{1}$$

For a selected next time $s < t$, deterministic DDIM gives

$$D_{t\to s}(x_t) = \sqrt{\rho_s}\hat{x}_{0,t} + \sqrt{1-\rho_s}\hat{\epsilon}_t. \tag{2}$$

We retain the implemented history correction so that all component comparisons evaluate the same complete method. Write $\lambda_t = \frac{1}{2}\log(\rho_t/(1-\rho_t))$, $h = \lambda_s - \lambda_t$, and let $\hat{\epsilon}_{\text{prev}}$ be the estimate at the preceding visited time. The correction is

$$\tilde{\epsilon}_t = \hat{\epsilon}_t + \frac{h}{2h_0}(\hat{\epsilon}_t - \hat{\epsilon}_{\text{prev}}), \qquad h_0 = 0.1, \tag{3}$$

$$H_{t\to s}(x_t) = \sqrt{\rho_s/\rho_t}\, x_t - \sqrt{1-\rho_s}(e^h - 1)\tilde{\epsilon}_t. \tag{4}$$

The fixed h_0 is a coefficient in this implementation, not the measured previous log-SNR interval. This is an approximate history update, not the standard variable-step DPM-Solver-2 or DPM-Solver++ formula. We do not claim second-order convergence. We use H when history is available, $t - s > 1$, and $s \geq 0$. Otherwise we use D, including the terminal transition to $s = -1$ with $\rho_{-1} = 1$. Changing the grid can therefore change which branch is invoked, even when the update rule is held fixed.

3.2 Spectral controller

Different bands can require different temporal resolution. We use changes in relative band power and phase to choose the next visited time. This is a design motivation, not an assumption that high frequencies always stabilize first.

At length 24, we partition the rFFT bins into $\{0\}$, $\{1, 2\}$, $\{3, 4, 5\}$, and $\{6, \ldots, 12\}$. The controller uses an unnormalized forward rFFT (`norm='backward'`), with float32 states and complex64 Fourier coefficients in the main experiments. For each sample, let $X_t[f, j]$ denote these coefficients and define mean band power and relative power by

$$P_t[B] = \frac{1}{|B|} \sum_{f \in B} \sum_j |X_t[f, j]|^2, \qquad \pi_t[B] = \frac{P_t[B]}{\sum_{B'} P_t[B'] + \varepsilon}. \tag{5}$$

Band-averaged power is a controller statistic. It is not the orthogonal-projection energy used in Section 4. With $u > t$ the previous visited time, the other two statistics are

$$\delta_t[B] = |\log(P_t[B] + \varepsilon) - \log(P_u[B] + \varepsilon)|, \tag{6}$$

$$v_t[B] = \sqrt{\frac{\sum_{f \in B} (p_t[f] + \varepsilon)\, \bar{\phi}_t[f]^2}{\sum_{f \in B} (p_t[f] + \varepsilon) + \varepsilon}}, \tag{7}$$

where $p_t[f] = \sum_j |X_t[f, j]|^2$ and $\bar{\phi}_t[f] = d^{-1} \sum_j \arg(X_t[f, j]\overline{X_u[f, j]})$. The implementation adds ε to each bin weight and once more to their sum in the denominator. Phase uses numerical $\mathrm{atan2}$ with the zero conventions in Appendix C.1. It measures change, not a time derivative. We average π, δ, v across the batch. A band is active when

$$\bar{\pi}_t[B] \geq \tau_E \quad \text{and} \quad \big(\bar{\delta}_t[B] \geq \tau_P \text{ or } c_B \bar{v}_t[B] \geq \tau_\phi(t)\big), \tag{8}$$

where $c_B = 1$ for DC and 1.5 otherwise, and $\tau_\phi(t) = \tau_\phi[1 - 0.5(1 - t/T)]$. We use $\varepsilon = 10^{-12}$. Initially, the previous state is the same noise sample, giving zero log-power change. The phase score follows the numerical convention above.

3.3 Stride selection and computational cost

Let $\mathcal{A}_t$ contain the active bands. The implementation chooses a large stride if $\mathcal{A}_t$ is empty, a medium stride if only band 0 (DC) is active, and a small stride otherwise. We set the small stride to one. A final window $t \leq K$ overrides activity with this unit stride. Thus the controller uses consecutive *visited* states, which need not be adjacent diffusion indices. Larger jumps are capped at t, preserving a visit to zero before the terminal step. Appendix D gives the pseudocode and dataset settings.

The three thresholds describe activity, while the stride set and final window specify the compute budget. Given a partition of the rFFT bins into a fixed number of nonempty bands, for batch size $n \geq 1$, channels $d \geq 1$, and length $L \geq 2$, the gate costs $O(ndL \log L)$ and history storage is $O(ndL)$. These bounds use fixed-precision, unit-cost FFT arithmetic, with constants independent of n, d, L. Each visited time makes one denoiser call. Batch size is part of the protocol because the gate aggregates across samples. Skipping times saves calls but adds controller overhead per visit.

4 What the Bandwise Analysis Supports

We separate an exact displacement identity from the conditions under which that displacement is small. Let $\mathcal{P}_B$ be an orthogonal Fourier projector onto a conjugate-symmetric band. Fix the current state and its clean prediction, and write $\rho' = \rho + \Delta$ for a cleaner noise level. DDIM can be expressed as

$$D_{\rho \to \rho'}(x) = ax + b\hat{x}_0, \qquad a = \sqrt{\frac{1 - \rho'}{1 - \rho}}, \quad b = \sqrt{\rho'} - a\sqrt{\rho}. \tag{9}$$

Proposition 1 (Band displacement with an explicit history residual). *Let $x, \hat{x}_0, R \in \mathbb{R}^{L\times d}$, $0 \le \rho < 1$, and $\rho \le \rho' \le 1$, with $\hat{\epsilon}$ reconstructed from the fixed pair $(x, \hat{x}_0)$ by Equation* (1)*. For $U = D_{\rho\to\rho'}(x) + R$,*

$$\mathcal{P}_B(U - x) = (a-1)\mathcal{P}_B x + b\mathcal{P}_B\hat{x}_0 + \mathcal{P}_B R, \tag{10}$$

$$\|\mathcal{P}_B(U-x)\|_F \le |a-1|\,\|\mathcal{P}_B x\|_F + |b|\,\|\mathcal{P}_B\hat{x}_0\|_F + \|\mathcal{P}_B R\|_F\,. \tag{11}$$

For DDIM, $R = 0$, including $\rho' = 1$. For the history update in Equation (4)*, take $(\rho, \rho') = (\rho_t, \rho_s)$ with $0 < \rho_t < \rho_s < 1$, $\hat{\epsilon}_{\text{prev}} \in \mathbb{R}^{L\times d}$, and $h_0 = 0.1 > 0$. Then*

$$R = -\sqrt{1-\rho'}(e^h - 1)\frac{h}{2h_0}(\hat{\epsilon}_t - \hat{\epsilon}_{\text{prev}}). \tag{12}$$

The identity follows from linearity. Equation (12) follows because the uncorrected exponential update equals DDIM. Appendix C gives the algebra and endpoint qualifications.

Corollary 1 (Local DDIM sufficient condition). *Fix positive $\rho_{\min}, \gamma$ with $0 < \rho_{\min} \le \rho \le \rho' \le 1 - \gamma < 1$, and set $\Delta = \rho' - \rho \ge 0$. If both $\|\mathcal{P}_B x\|_F$ and $\|\mathcal{P}_B\hat{x}_0\|_F$ are at most $\epsilon_B \ge 0$, then a DDIM step satisfies*

$$\|\mathcal{P}_B(D-x)\|_F \le C\Delta\epsilon_B, \qquad C = \frac{1}{\gamma} + \frac{1}{2\sqrt{\rho_{\min}}}. \tag{13}$$

This bounds a band's displacement during one step with the current prediction fixed. It does not bound the difference from many recomputed fine steps, cumulative sampling error, or final distributional quality. For a history update the residual term in Equation (11) remains. Consecutive-state inactivity does not bound that residual or either absolute band norm. Likewise, a relative-energy test can be small while absolute power is large.

The analysis motivates measuring spectral activity, but does not certify Equation (8). The final fine-step window is an empirical safeguard near the endpoint, where the local constants grow. We test the gate against a stricter current-sample/clean-prediction energy check in Appendix E.3. That diagnostic also uses relative power and is not an exact implementation of Corollary 1.

5 Experiments

5.1 Validation selection and independent confirmation

For each of six length-24 datasets, we select sampler settings with one validation Diffusion-TS model and evaluate them with three separately trained confirmation models. All samplers share the checkpoint, initial noise, reference subset and feature encoder within each comparison. Training budgets are fixed, and scaling is fitted only to training data.

Each of the spectral and uniform-history families has six predeclared candidates. We select each family's lowest mean validation C-FID over three generation seeds, subject to at most 96 mean denoiser calls per batch. Configurations and the shared, validation-fitted TS2Vec encoder are then frozen. These benchmarks informed earlier development, so this split separates selection from confirmation within the present experiment, not from all prior exposure to the datasets.

The six methods are the selected complete sampler, independently selected uniform history, official DPM-Solver++ (2M), UniPC-bh2, full DDIM and a matched uniform replay. The replay changes diffusion times while matching every batch's call count and history/DDIM choice at each update index. It needs a source spectral run and diagnoses grid placement rather than standalone sampling speed.

Each confirmation model uses three generation seeds and up to 2,048 reference and generated sequences per draw. We average within models, then report means and sample SDs across the three models, not nine independent training runs. We make no significance claim. Appendix A gives all five metrics, paired model-level differences, selection records and sampling costs, including replay's separate source-grid acquisition.

Table 1: Confirmation C-FID on six length-24 datasets. Means and sample SDs use three independently trained checkpoints, each averaged over three generation seeds. Methods share noise, checkpoint and validation-fitted encoder within each paired block. Uniform replay matches per-batch NFE and update positions but requires a source grid. Uniform history is selected separately. Official-solver budgets approximately match ours and retain official endpoint handling. Historical-protocol scores are not comparable. Blue shading identifies our method or spectral grid. Bold and underline mark the lowest and second-lowest distinct displayed means in each column across the listed budgets, including ties, not statistical significance.

Method	Sines	Stocks	ETTh	MuJoCo	Energy	fMRI
StrideDiffusion	**0.0179** ±0.0025	**86.7702** ±0.4429	3.0902 ±0.0264	**0.0648** ±0.0019	**16.7116** ±0.1175	<u>0.5728</u> ±0.0236
Matched uniform replay	0.0218 ±0.0012	87.1267 ±0.1251	<u>3.0567</u> ±0.0708	0.0760 ±0.0046	17.2315 ±0.1310	0.6491 ±0.0283
Selected uniform history	0.0215 ±0.0002	<u>86.8932</u> ±0.1795	**2.9559** ±0.0451	<u>0.0699</u> ±0.0053	17.3620 ±0.1235	0.7254 ±0.0415
DPM-Solver++ (2M)	0.0238 ±0.0006	87.0356 ±0.1089	3.1069 ±0.0106	0.0747 ±0.0029	<u>17.1169</u> ±0.1529	**0.5547** ±0.0171
UniPC (bh2)	0.0224 ±0.0009	87.2824 ±0.1354	3.2817 ±0.0018	0.0763 ±0.0034	17.3382 ±0.1272	0.6182 ±0.0245
Full DDIM	<u>0.0208</u> ±0.0009	86.9812 ±0.1381	3.2028 ±0.0265	0.0702 ±0.0039	17.2631 ±0.1188	0.5941 ±0.0251

Complete-method comparison. Table 1 compares the six methods under one confirmation protocol. StrideDiffusion has lower mean Context-FID than DPM-Solver++ on five datasets and UniPC on all six. DPM-Solver++ is better on fMRI. The official solvers use the nearest integer to our measured mean calls per batch, so these are approximately matched call budgets. The independently selected uniform-history control is better on ETTh and worse on the other five datasets. Its validation-selected budget can differ from ours. Full DDIM uses the full diffusion horizon and has higher mean Context-FID on all six datasets. Neither comparison should be described as exact matched-NFE evidence.

Measured sampling time. StrideDiffusion is faster than both official solvers on all six datasets. Their mean sampling times divided by ours range from 1.16 to 1.30 for DPM-Solver++ and from 1.20 to 1.37 for UniPC. Relative to full DDIM, the corresponding factors range from 11.84 to 29.80. These are ratios of dataset-level mean sampling times, not means of paired ratios or end-to-end speedups. The selected uniform-history control is faster on fMRI but has higher mean Context-FID there. Each draw uses the median of three synchronized sampling repeats on an exclusive H200 GPU. We average three draws within each checkpoint and then average the three checkpoint means. Loading, warm-up, sample transfers, writes, scoring and the separate source-grid acquisition pass are outside this timing interval.

5.2 Attributing the gains

Grid placement with matched updates. Uniform replay uses the same checkpoint and initial noise as the spectral run. For every batch, it matches the exact call count and the history/DDIM choice at each update index while replacing the diffusion times with uniform placement. The spectral grid has lower mean Context-FID on five datasets. The reductions are 17.73% on Sines, 0.41% on Stocks, 14.69% on MuJoCo, 3.02% on Energy and 11.75% on fMRI. ETTh has 1.10% higher error. These percentages compare dataset means. All three checkpoint-level differences favor the spectral grid on Sines, MuJoCo, Energy and fMRI. Two favor it on Stocks, where the mean gap is small, and one on ETTh. With three training runs, we report these paired differences without a significance claim.

This control attributes the difference to grid placement under the fixed history rule and matched branch positions. It does not show that the spectral grid improves every solver. Replay also needs a source spectral run, whose calls and time are reported separately and added to the replay cost in the appendix. Replay-only time is not a standalone deployment cost.

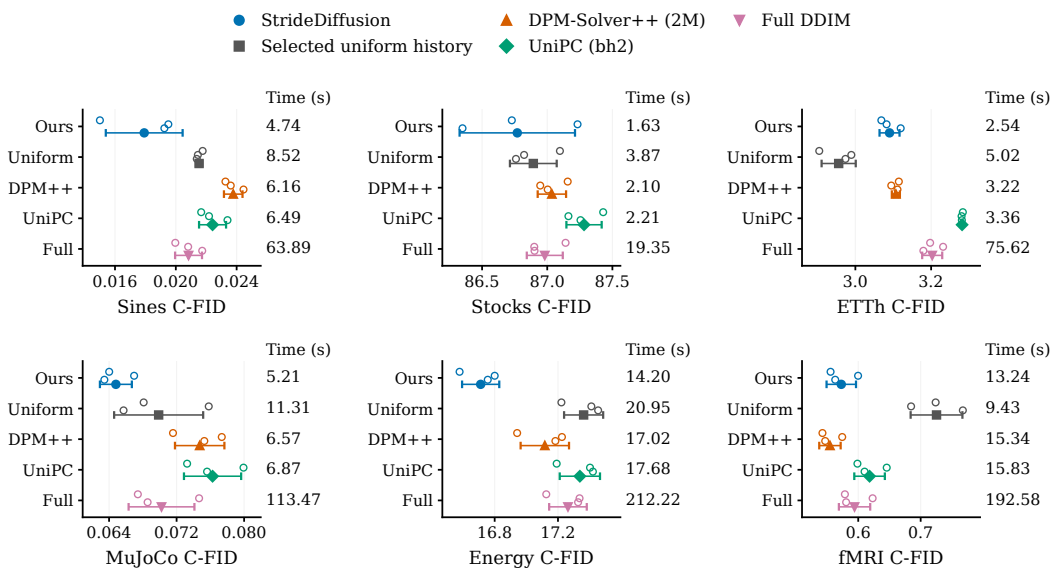

Figure 3: Quality and sampling time for six datasets and five deployable samplers, shown on separate method rows to avoid overlap. Open markers are checkpoint means over three generation seeds. Filled markers and horizontal bars show the mean and sample SD across three trained checkpoints, not confidence intervals. Right-hand columns give mean synchronized sampling time in seconds, not end-to-end latency. Lower C-FID and time are better within a panel. Sample counts and scales differ. Budgets follow Table 1 and need not match across methods. Appendix A retains exact scores, timing variation and source-dependent replay costs.

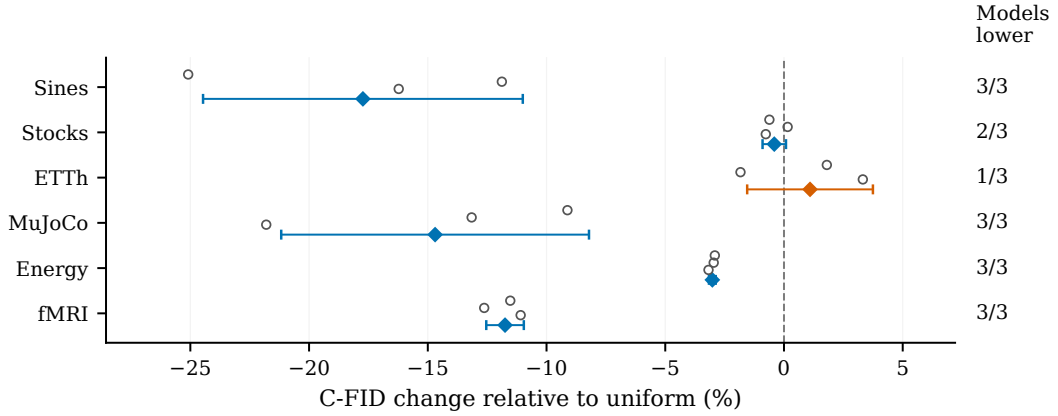

Figure 4: Grid placement with identical per-batch NFE and history/DDIM update positions. Points show three trained checkpoints, each averaged over three generation seeds. Diamonds and bars are the mean and sample SD of paired C-FID differences, divided by the dataset's mean uniform-replay C-FID and expressed as percentages. Diamonds are ratios of means minus one, not means of paired ratios. Bars are not confidence intervals. Negative favors the spectral grid. Counts give checkpoints with lower C-FID. Replay requires a source spectral run and is not a standalone sampler.

Supporting fixed-checkpoint controls. Earlier fixed-checkpoint studies test same-grid DDIM, online Spectral-DDIM and grids recorded from other noise draws. Their separate reference windows and encoders prevent ranking absolute Context-FID against the confirmation results. Gains depend on the update rule, and recorded grids recover much of the benefit, leaving the need for per-batch online adaptation open. Appendices G, E.2 and F retain the complete comparisons and seed protocols.

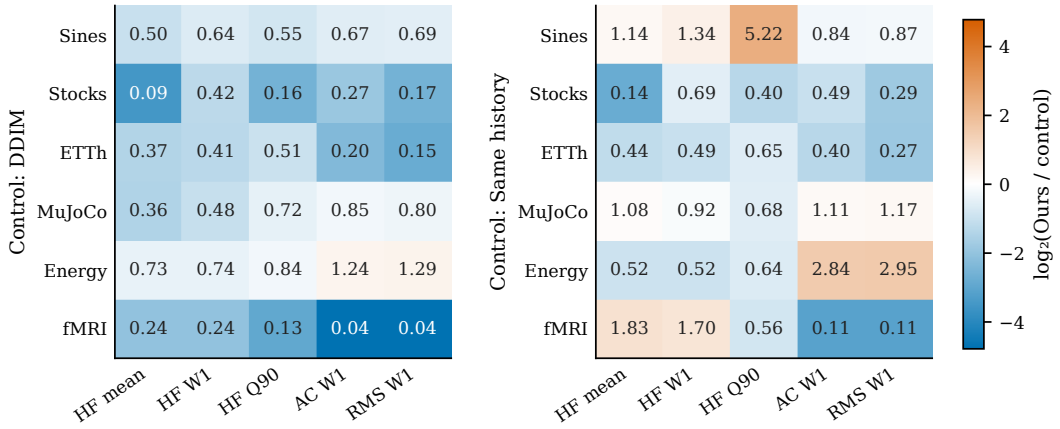

Figure 5: Absolute-power error ratios in the earlier fixed-checkpoint study. Cells divide Ours' three-generation-seed mean error by the control's. Blue is below one, orange above one. Colors use a symmetric $\log_2$ scale, while numbers are ordinary ratios. The two controls are not pooled. HF, AC and RMS denote high-frequency power, total non-DC power and centered RMS. W1 is reference-normalized Wasserstein distance and Q90 is relative 90th-percentile error. Appendix Tables 19–20 give exact means and SDs.

5.3 Frequency fidelity, transfer, and additional backbones

The following power-distribution study uses the earlier fixed-checkpoint protocol. Its Parseval-weighted diagnostics differ from the raw high-frequency power error in the confirmation tables. In the confirmation comparison, lower Context-FID does not imply lower correlation or high-frequency error. MuJoCo, for example, has lower mean Context-FID than matched uniform replay but higher errors on both diagnostics.

Does skipping suppress high-frequency power? We compare per-sample absolute high-frequency power, total non-DC power, and centered RMS using a Parseval-weighted FFT. Error scales come only from the reference data. This avoids masking amplitude loss by dividing each generated signal by its own power. Against uniform DDIM, StrideDiffusion has lower high-frequency mean error, Wasserstein distance, and 90th-percentile error on all six datasets. Under the same conditional history rule, those counts are 3/6, 4/6, and 5/6. Figure 5 shows all five diagnostics under both controls. The full tables in Appendix E.4 support improved power-distribution matching for the complete method, not preservation of every sample's details or phase.

Can settings be reused? Table 2 tests Sines' three activity thresholds across all six datasets while retaining each target's strides and final window. The largest mean increase in scheduler-only DDIM Context-FID is .0004. This earlier fixed-checkpoint study supports threshold reuse, not transfer of the entire compute schedule or a statistical equivalence claim. Its absolute scores are not comparable to Table 1. Appendix I gives the full-schedule stress test and model-only calibration study. No weights are updated.

Other parameterizations. On ImagenTime and ImagenFew, we implement an EDM-specific spectral grid with a two-step history update and selected Heun corrections, denoted Spectral PC2. Uniform PC2 matches its NFE, transition count, and number of corrections. Correction placement can differ, and the length-64 ImagenFew port also uses a different history coefficient. These are matched-budget port comparisons, not uniformly grid-only controls. Configuration selection uses seed 12345, followed by evaluation of frozen candidates on seeds 2032, 2033, and 2034. Figure 6 shows lower mean Context-FID than both the native full sampler and uniform PC2 in four settings. The spectral controller remains slower than the matched PC2 control. Correlation and spectral errors do not all improve. Appendix I.3 specifies the update choices and reports the other evaluated candidates.

Table 2: Frozen-threshold transfer for scheduler-only DDIM. Values are mean ± sample standard deviation over three generation seeds. T/NFE measures the reduction in denoiser calls relative to full-step sampling. Both quality columns use our spectral gate with DDIM. Blue marks the transferred thresholds, not a distinct backbone. Bold marks the lower displayed C-FID, and underline the other value when distinct. Ties share bold.

Data	Target-tuned C-FID	**Transferred threshold C-FID**	NFE	T/NFE
Sines	**.0163 ±.0032**	**.0163 ±.0032**	35.0	14.29
Stocks	**.2080 ±.0150**	**.2080 ±.0150**	31.2	16.02
ETTh	**.3882 ±.0572**	.3883 ±.0540	21.9	22.84
MuJoCo	**.0392 ±.0051**	.0396 ±.0052	100.3	9.97
Energy	**.1128 ±.0112**	**.1128 ±.0112**	59.0	16.95
fMRI	.1973 ±.0129	**.1933 ±.0130**	107.0	9.35

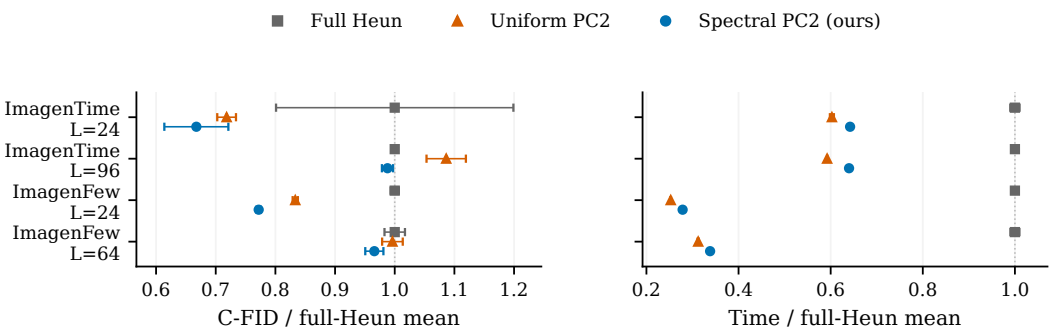

Figure 6: Quality and latency of the PC2 ports on Sine. Each group uses one frozen backbone. Means and sample SDs over three generation seeds are divided by that group's full-Heun mean as a fixed display scale, not uncertainty of a paired ratio. Lower is better. Uniform and spectral PC2 match NFE and correction counts, but other update choices can differ. Full/PC2 NFE are 35/21, 35/21, 71/18 and 71/22 from top to bottom. Appendix Table 37 retains all absolute values.

The unchanged ImagenTime configuration gives .06796 versus .06850 Context-FID at length 96 with 21 versus 35 NFE, a small quality difference. Appendix I.2 gives auxiliary metrics and our paper-based L2D-Diff implementation, without author code or checkpoints. Longer-sequence tests use three independently trained Diffusion-TS checkpoints per dataset–length pair (18 total). Gains vary with the dataset and update at lengths 24/96/192, so length transfer still needs validation (Appendix H). These ports are empirical tests, not extensions of the DDIM condition to other updates.

6 Discussion and Conclusion

StrideDiffusion couples spectral grid selection with a history update without retraining. With validation-selected settings and three independently trained models per dataset, it lowers mean Context-FID against DPM-Solver++ on five datasets and UniPC on six at approximately matched call budgets. Exact call-count and update-position matching favors the spectral grid on five datasets, with ETTh as the exception. Other metrics do not all improve.

Recorded grids recover much of the gain, leaving the need for online adaptation open. The practical gate does not bound the current-state, clean-prediction and history terms in our analysis. Controlling the history residual is a useful next step. Compute settings remain task-specific and do not reliably transfer unchanged across lengths. The benchmarks informed earlier development, despite the present validation/confirmation split. Irregular-time sampling and wider backbone retraining remain untested.

Reproducibility Statement

Section 3 and Appendix D specify the update equations, gate and endpoint behavior. Appendix C proves the local displacement bound. The main comparison separates validation selection from confirmation and uses three independently trained models per dataset. We average three generation seeds within each model before reporting variation across the three models. Appendix A gives the data splits, training budgets, all validation candidates, five metric definitions and sampling-cost accounting. It also reports model-level paired differences. Earlier fixed-checkpoint studies use separate protocols in Appendices E, F and G. Their absolute C-FID values are not ranked against the confirmation results. Appendix H reports independently trained longer-sequence models. Appendix J distinguishes the adapted baselines' training-seed variation from the historical Ours results' generation-seed variation at one checkpoint. The local research workspace retains source snapshots, result files and provenance hashes. The anonymous manuscript ZIP contains paper build inputs, not raw training or sampling records. Historical material in Appendix K has its own protocol, and the examples in Appendix L retain explicit provenance limits. The anonymous repository is linked in the abstract.

Ethics Statement

Synthetic time series can reproduce biases or private patterns from their training data. Fewer denoising evaluations do not provide privacy guarantees. Distributional scores do not establish fitness for clinical, financial, or other safety-critical uses. The fMRI benchmark used here is a realistic BOLD simulation rather than a study with newly recruited participants.

AI Use Statement

Generative AI tools assisted with experiment design and implementation, including sampling and evaluation code, synthetic-data generation scripts, and data preprocessing. They also assisted with mathematical statements and proofs, result interpretation and checks, figure code, literature lookup, translation, and manuscript drafting and editing. Saved artifacts and executable checks were used to examine the code and reported results. These checks do not replace author review. The authors are responsible for the final claims, citations, code, and submission.

References

Barbora Barancikova, Zhuoyue Huang, and Cristopher Salvi. Sigdiffusions: Score-based diffusion models for time series via log-signature embeddings. In *International Conference on Learning Representations*, 2025.

Marin Biloš, Kashif Rasul, Anderson Schneider, Yuriy Nevmyvaka, and Stephan Günnemann. Modeling temporal data as continuous functions with stochastic process diffusion. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 2452–2470. PMLR, 2023. URL https://proceedings.mlr.press/v202/bilos23a.html.

Jiwoo Chung, Sangeek Hyun, MinKyu Lee, Byeongju Han, Geonho Cha, Dongyoon Wee, Youngjun Hong, and Jae-Pil Heo. Seacache: Spectral-evolution-aware cache for accelerating diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14283–14294, June 2026.

Andrea Coletta, Sriram Gopalakrishnan, Daniel Borrajo, and Svitlana Vyetrenko. On the constrained time-series generation problem. *Advances in Neural Information Processing Systems*, 36:61048–61059, 2023.

Tal Gonen, Itai Pemper, Ilan Naiman, Nimrod Berman, and Omri Azencot. Time series generation under data scarcity: A unified generative modeling approach. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 172751–172790. Curran Associates, Inc., 2025. doi: 10.52202/085713-5750. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/fc1767dae81344d5b9290abfdceed417-Paper-Conference.pdf.

Yefei He, Jing Liu, Weijia Wu, Hong Zhou, and Bohan Zhuang. EfficientDM: Efficient quantization-aware fine-tuning of low-bit diffusion models. In *International Conference on Learning Representations*, 2024.

Zhifeng Kong, Wei Ping, Jiaji Huang, Kexin Zhao, and Bryan Catanzaro. Diffwave: A versatile diffusion model for audio synthesis. In *International Conference on Learning Representations*, 2021.

Haeil Lee, Hansang Lee, Seoyeon Gye, and Junmo Kim. Beta sampling is all you need: Efficient image generation strategy for diffusion models using stepwise spectral analysis. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4215–4224. IEEE, 2025.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ODE solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems*, volume 35, pp. 5775–5787, 2022. doi: 10.52202/068431-0418. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/260a14acce2a89dad36adc8eefe7c59e-Paper-Conference.pdf.

Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models. *Machine Intelligence Research*, 22 (4):730–751, 2025. doi: 10.1007/s11633-025-1562-4.

Ilan Naiman, Nimrod Berman, Itai Pemper, Idan Arbiv, Gal Fadlon, and Omri Azencot. Utilizing image transforms and diffusion models for generative modeling of short and long time series. In *Advances in Neural Information Processing Systems*, volume 37, pp. 121699–121730. Curran Associates, Inc., 2024. doi: 10.52202/079017-3868. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/dc6748383752138af7f00b3185a0a404-Paper-Conference.pdf.

Kashif Rasul, Calvin Seward, Ingmar Schuster, and Roland Vollgraf. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8857–8868. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/rasul21a.html.

Severi Rissanen, Markus Heinonen, and Arno Solin. Generative modelling with inverse heat dissipation. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=4PJUBT9f2OI.

Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. In *International Conference on Learning Representations*, 2022.

Lifeng Shen, Kai Syun Hou, Weiyu Chen, and James Kwok. Latent-to-data cascaded diffusion models for unconditional time series generation. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=nAyeE7cAS0.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=St1giarCHLP.

Yusuke Tashiro, Jiaming Song, Yang Song, and Stefano Ermon. CSDI: Conditional score-based diffusion models for probabilistic time series imputation. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 24804–24816. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper/2021/file/cfe8504bda37b575c70ee1a8276f3486-Paper.pdf.

Xingyi Yang, Daquan Zhou, Jiashi Feng, and Xinchao Wang. Diffusion probabilistic model made slim. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pp. 22552–22562, 2023.

Zinuo You, Jin Zheng, and John Cartlidge. Latent laplace diffusion for irregular multivariate time series. *arXiv preprint arXiv:2605.19805*, 2026.

Xinyu Yuan and Yan Qiao. Diffusion-ts: Interpretable diffusion for general time series generation. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 41582–41610, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/b5b66077d016c037576cc56a82f97f66-Paper-Conference.pdf.

Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, and Jiwen Lu. UniPC: A unified predictor-corrector framework for fast sampling of diffusion models. In *Advances in Neural Information Processing Systems*, volume 36, pp. 49842–49869, 2023. doi: 10.52202/075280-2170.

Yue Zheng, Xianfeng Li, and Ying Zhan. Frequency-aware timestep optimization for efficient sampling in diffusion models. *Journal of King Saud University Computer and Information Sciences*, 38:467, 2026. doi: 10.1007/s44443-026-00886-1. URL https://doi.org/10.1007/s44443-026-00886-1.

GUIDE TO THE APPENDICES

Appendix A gives the protocol and full results for the main validation-selected comparison, including all five metrics, sampling costs and paired differences across the three trained models. It supports the main table and grid-attribution figure. The fixed-checkpoint studies use separate reference sets and encoders, so their absolute C-FID scores are not comparable to the main confirmation results. Their grid controls remain in Appendices F and G.

Table highlighting. Blue shading identifies Ours or a spectral variant. Unless a caption states otherwise, bold and underlining mark the best and second-best distinct displayed means within the stated comparison group, including ties. Lower errors and sampling times are better. NFE counts are not ranked. These marks do not imply statistical significance.

Topic	Appendix	Page
Validation selection and independent confirmation	A	13
Proofs and scope of the bandwise displacement bound	C	24
Sampling loop, settings, and budget matching	D	25
Earlier metrics, timing, frequency fidelity, and caching	E	27
Fixed-budget controls with matched update positions	F	37
Native-grid replay and update-rule controls	G	40
Independent training and longer sequences	H	45
Threshold reuse and additional backbones	I	50
Additional trained DiffWave and DiffTime generators	J	54
Historical reference results and component diagnostics	K	58
Archived imputation and forecasting examples	L	59

A SELECTION AND CONFIRMATION PROTOCOL

Data separation. All sequences have length 24. Stocks, ETTh and Energy are split by file row into 70% training, 15% validation and 15% confirmation partitions before windowing. Windows never cross a partition boundary. ETTh has a checked date order. The Stocks and Energy files have no date column, so we describe their split as file-order rather than independently verified chronological separation. Sines uses independently generated sets of 10,000, 2,048 and 2,048 sequences with data seeds 9101, 9102 and 9103. The fMRI benchmark is simulated BOLD data, split by subject into 35/7/8 subjects with seed 9021. Windows remain within each subject. MuJoCo uses 10,000 previously saved trajectories split into 7,000/1,500/1,500 with seed 9022, not newly simulated trajectories.

Channel-wise minima and maxima are fitted only to training data. Reference values outside the training range remain unclipped. Generated samples follow the sampler's mapping from $[-1, 1]$ to $[0, 1]$ and clipping convention. We record out-of-range reference counts and generated clipping counts rather than clipping the reference to hide a range mismatch.

Training and random seeds. Training seed 711 is used only for validation selection. Seeds 721, 732 and 743 train the confirmation models. The fixed update budgets for Sines, Stocks, ETTh, MuJoCo, Energy and fMRI are respectively 12,000, 10,000, 18,000, 14,000, 25,000 and 15,000. Each update accumulates two batches of 128 examples, sampled uniformly with replacement. This is not the historical shuffled-epoch data loader. We retain the frozen backbone architecture, learning-rate schedule, Adam coefficients $(0.9, 0.96)$ and EMA decay 0.995 with updates every ten steps. Gradients are clipped at norm 1. The final fixed-budget checkpoint is used without a generation-quality-based training-length choice.

Validation candidates. The six spectral profiles use the following values, in the order (large stride, medium stride, small stride, final window, energy threshold, log-power threshold, phase threshold):

Profile	Large	Medium	Small	Window	Energy	Log-power	Phase
Sines	50	20	1	20	.10	.01	.04
Stocks	50	10	1	20	.10	.01	.04
ETTh	50	10	1	12	.70	.02	.08
MuJoCo	20	5	1	20	.50	.02	.04
Energy	40	5	1	12	.10	.01	.06
fMRI	20	5	1	20	.50	.01	.06

Names identify the historical source of a profile, not a restriction on where it can be selected. Every dataset evaluates all six profiles. Strides remain integer diffusion steps without horizon rescaling. The uniform-history family tests budgets 16, 24, 32, 48, 64 and 96. Both families use the stated approximate history update with $h_0 = 0.1$ and DDIM bootstrap and terminal updates.

Each candidate uses generation seeds 72101, 72102 and 72103, batch size 64, and up to 512 validation references. Reference-index seed 54321 and paired initial noise are common to all candidates. Each family minimizes mean validation C-FID among candidates with mean batch NFE at most 96. Exact ties prefer lower mean NFE, then candidate ID. A numerically failed candidate is retained in the record but cannot win. If no candidate meets the budget, a minimum-NFE diagnostic choice is flagged as a budget failure, not a successful budget-constrained selection. If all candidates fail, the dataset and the complete experiment remain incomplete. No confirmation score enters selection.

Shared feature encoder. One TS2Vec encoder per dataset is fitted only to the fixed validation reference subset with seed 972101, batch size 8, learning rate 0.001 and representation dimension 320. Training uses 200 updates when that subset contains at most 100,000 scalar elements and 600 otherwise. The encoder is frozen for all candidates and confirmation models. Consequently, absolute C-FID values are not ranked against historical tables that use different fitted encoders or reference sets.

Confirmation and metrics. Each of the three confirmation models uses generation seeds 83101, 83102 and 83103, up to 2,048 references and batch size 64. All six methods use the same saved

reference subset and paired noise. C-FID compares feature means and covariances with the original matrix square-root formula. We add no covariance regularizer or score rescaling. The square root must have maximum absolute imaginary component at most 10^{-6} and real-part relative Frobenius residual at most 10^{-6}. Failures are retained. Only final negative roundoff in $[-10^{-8}, 0)$ is set to zero, with its raw value also recorded.

Correlation error is the lower-triangular lag-zero correlation-matrix L_1 difference divided by ten. High-frequency error is the absolute difference in mean unnormalized rFFT power summed over bins 6–12. It has no generated-signal-dependent normalization and is distinct from the Parseval-weighted distribution diagnostics in the historical appendix. Discriminative score is the absolute difference between real/fake GRU classification accuracy and chance. Real and generated sequences are independently shuffled and split into 80% training and 20% held-out evaluation sets. Predictive score trains a GRU on generated sequences to predict the next value of the last channel from the other channels. It reports MAE over all real evaluation sequences. Both GRUs have hidden width $\max(\lfloor d/2 \rfloor, 1)$, where d is the channel count, and batches capped at 128 sequences. Discriminative and predictive evaluators use respectively 2,000 and 5,000 training updates and seeds 7101, 7102 and 7103. All five metrics use the same saved draws.

Calls and timing. We report mean calls per batch, sample-weighted mean NFE and total batched denoiser calls. These differ when the final batch is smaller. Official-solver budgets are the nearest integer to the complete method's mean batch NFE, rounded half up. This does not promise exactly matched sample-weighted compute or FLOPs. The official DPM-Solver++ cosine-schedule clipping retains 498 of 500 or 996 of 1,000 horizon steps. Other controls retain the full horizon.

Formal timings use one task per H200 GPU and three complete sampling repeats in randomized method order. Initial noise is identical across repeats and saved outputs must match. Synchronized sampling time includes the controller, update arithmetic and in-memory traces. Loading, three warm-up forwards, sample transfers, writes and scoring are excluded. Loading, warm-up, source-grid acquisition and primary-metric computation have separate aggregate timing records. Sample transfers and file writes are not individually timed. The sum of acquisition and replay seconds is a sum of two measurements, not a measured end-to-end pipeline latency. Validation selection and encoder-fitting costs are reported separately from deployment.

A.1 Complete confirmation results

Differences across quality metrics. Lower C-FID does not mean that every metric improves. Compared with matched uniform replay, the spectral grid has higher high-frequency error and discriminative score on Sines. On MuJoCo, it has lower C-FID and discriminative score but higher correlation and high-frequency errors. Predictive error is higher on ETTh and fMRI. Energy favors the spectral grid on all five mean scores, though its discriminative scores remain close to 0.5 for every method. Stocks also has discriminative scores close to 0.5 across all methods. Under the definition $|\mathrm{accuracy} - 0.5|$, these values indicate that the evaluator can distinguish real and generated samples. Small differences near this ceiling should not be taken as evidence that the distributions match. We therefore report all five metrics and restrict the main comparative claim to C-FID, the metric used for validation selection.

Reading the tables. All five metrics use the same saved samples. Means and sample SDs are over three independent checkpoint means, each averaging three generation seeds. Corr. is lag-zero lower-triangular correlation L1 error divided by ten. HF is absolute unnormalized high-frequency power error. Disc./Pred. average three GRU evaluator seeds. Lower quality scores are better. Batch NFE and sample-weighted NFE are separate estimands. Total calls sum all batches per draw. Seconds are synchronized sampling medians over three repeats, aggregated in the same hierarchy. They exclude loading, warmup, noise transfer, writes, metrics and source-grid acquisition. Uniform replay times exclude the required grid-acquisition pass, which is reported separately. Replay + grid pass adds their calls and separately measured seconds, not end-to-end latency. Differences are StrideDiffusion minus the named control, with one column per training seed. Blue identifies Ours. Bold and underline mark the best and second-best distinct displayed means in each quality column, including ties. Costs and signed differences are not ranked. No significance test is implied.

Table 3: Sines confirmation results: 2048 samples in 32 batches per draw, with batch size 64. Values are means ± sample SD across three independently trained checkpoints, each averaging three generation seeds. The metrics, cost accounting, paired differences and highlighting follow Appendix A.1. Lower quality scores are better. Replay costs exclude grid acquisition, reported separately.

Method	C-FID	Corr.	HF	Disc.	Pred.
StrideDiffusion	**0.0179** ±0.0025	**0.0178** ±1.872×10⁻⁴	0.0044 ±0.0023	0.0361 ±0.0032	**0.0932** ±1.091×10⁻⁴
Matched uniform replay	0.0218 ±0.0012	0.0257 ±0.0028	**0.0016** ±0.0015	0.0294 ±0.0081	0.0936 ±1.308×10⁻⁴
Selected uniform history	0.0215 ±2.050×10⁻⁴	0.0232 ±0.0033	0.0028 ±0.0021	**0.0292** ±0.0066	0.0938 ±4.311×10⁻⁴
DPM-Solver++ (2M)	0.0238 ±6.104×10⁻⁴	0.0255 ±0.0019	0.0093 ±0.0028	**0.0292** ±0.0062	0.0936 ±1.764×10⁻⁴
UniPC (bh2)	0.0224 ±8.881×10⁻⁴	0.0198 ±0.0012	0.0109 ±0.0025	0.0361 ±0.0054	0.0939 ±1.767×10⁻⁴
Full DDIM	0.0208 ±8.852×10⁻⁴	0.0206 ±0.0017	0.0045 ±0.0017	0.0366 ±0.0056	0.0939 ±3.291×10⁻⁴

Method	Batch NFE	Sample NFE	Total calls	Seconds
StrideDiffusion	29.8542 ±0.3528	29.8542 ±0.3528	955.3333 ±11.2891	4.7358 ±0.0441
Matched uniform replay	29.8542 ±0.3528	29.8542 ±0.3528	955.3333 ±11.2891	3.9351 ±0.0387
Selected uniform history	64.0000 ±0.0000	64.0000 ±0.0000	2048.0000 ±0.0000	8.5192 ±0.0341
DPM-Solver++ (2M)	30.0000 ±0.3333	30.0000 ±0.3333	960.0000 ±10.6667	6.1622 ±0.0758
UniPC (bh2)	30.0000 ±0.3333	30.0000 ±0.3333	960.0000 ±10.6667	6.4896 ±0.0631
Full DDIM	500.0000 ±0.0000	500.0000 ±0.0000	1.600×10⁴ ±0.0000	63.8857 ±0.1845
Grid acquisition	29.8542 ±0.3528	29.8542 ±0.3528	955.3333 ±11.2891	4.8727 ±0.0400
Replay + grid pass	59.7083 ±0.7056	59.7083 ±0.7056	1910.6667 ±22.5783	8.8078 ±0.0785

C-FID control	Mean Δ	Seed721	Seed732	Seed743
Matched uniform replay	−0.0039 ±0.0015	−0.0055	−0.0026	−0.0035
Selected uniform history	−0.0036 ±0.0027	−0.0068	−0.0019	−0.0021
DPM-Solver++ (2M)	−0.0059 ±0.0022	−0.0083	−0.0041	−0.0052
UniPC (bh2)	−0.0045 ±0.0020	−0.0067	−0.0027	−0.0042
Full DDIM	−0.0029 ±0.0019	−0.0050	−0.0013	−0.0025

Table 4: Stocks confirmation results: 530 samples in 9 batches per draw, with batch size 64. Values are means ± sample SD across three independently trained checkpoints, each averaging three generation seeds. The metrics, cost accounting, paired differences and highlighting follow Appendix A.1. Lower quality scores are better. Replay costs exclude grid acquisition, reported separately.

Method	C-FID	Corr.	HF	Disc.	Pred.
StrideDiffusion	**86.7702** ±0.4429	**0.2701** ±0.0041	0.0149 ±0.0032	0.5000 ±0.0000	0.0094 $\pm 7.230\times10^{-4}$
Matched uniform replay	87.1267 ±0.1251	0.2973 ±0.0037	0.0313 $\pm 4.529\times10^{-4}$	**0.4981** ±0.0025	0.0097 ±0.0015
Selected uniform history	86.8932 ±0.1795	0.2863 ±0.0064	0.0158 ±0.0048	0.4993 $\pm 3.026\times10^{-4}$	**0.0085** $\pm 4.186\times10^{-4}$
DPM-Solver++ (2M)	87.0356 ±0.1089	0.2771 ±0.0025	**0.0076** $\pm 3.217\times10^{-4}$	0.4998 $\pm 3.026\times10^{-4}$	0.0100 $\pm 9.257\times10^{-4}$
UniPC (bh2)	87.2824 ±0.1354	0.2876 ±0.0019	0.0210 $\pm 2.629\times10^{-4}$	0.4997 $\pm 6.052\times10^{-4}$	0.0103 $\pm 5.795\times10^{-4}$
Full DDIM	86.9812 ±0.1381	0.2911 ±0.0044	0.0244 ±0.0019	0.4995 $\pm 5.241\times10^{-4}$	0.0092 $\pm 8.184\times10^{-4}$

Method	Batch NFE	Sample NFE	Total calls	Seconds
StrideDiffusion	34.6049 ±0.1711	34.5706 ±0.1859	311.4444 ±1.5396	1.6347 ±0.0085
Matched uniform replay	34.6049 ±0.1711	34.5706 ±0.1859	311.4444 ±1.5396	1.3680 ±0.0042
Selected uniform history	96.0000 ±0.0000	96.0000 ±0.0000	864.0000 ±0.0000	3.8663 ±0.0179
DPM-Solver++ (2M)	34.5556 ±0.1925	34.5556 ±0.1925	311.0000 ±1.7321	2.0971 ±0.0119
UniPC (bh2)	34.5556 ±0.1925	34.5556 ±0.1925	311.0000 ±1.7321	2.2076 ±0.0220
Full DDIM	500.0000 ±0.0000	500.0000 ±0.0000	4500.0000 ±0.0000	19.3503 ±0.0873
Grid acquisition	34.6049 ±0.1711	34.5706 ±0.1859	311.4444 ±1.5396	1.7001 ±0.0088
Replay + grid pass	69.2099 ±0.3421	69.1413 ±0.3718	622.8889 ±3.0792	3.0681 ±0.0130

C-FID control	Mean Δ	Seed721	Seed732	Seed743
Matched uniform replay	−0.3566 ±0.4298	−0.5352	0.1338	−0.6682
Selected uniform history	−0.1230 ±0.4622	−0.3691	0.4102	−0.4101
DPM-Solver++ (2M)	−0.2654 ±0.4909	−0.4287	0.2863	−0.6539
UniPC (bh2)	−0.5122 ±0.5144	−0.7010	0.0699	−0.9055
Full DDIM	−0.2110 ±0.4756	−0.4124	0.3322	−0.5527

Table 5: ETTh confirmation results: 2048 samples in 32 batches per draw, with batch size 64. Values are means ± sample SD across three independently trained checkpoints, each averaging three generation seeds. The metrics, cost accounting, paired differences and highlighting follow Appendix A.1. Lower quality scores are better. Replay costs exclude grid acquisition, reported separately.

Method	C-FID	Corr.	HF	Disc.	Pred.
StrideDiffusion	3.0902 ±0.0264	0.4167 ±0.0055	<u>0.0655</u> ±0.0013	**0.4390** ±0.0014	0.2203 ±0.0033
Matched uniform replay	<u>3.0567</u> ±0.0708	**0.3817** ±0.0136	0.1875 ±0.0113	<u>0.4396</u> ±0.0066	**0.2072** ±0.0119
Selected uniform history	**2.9559** ±0.0451	<u>0.4092</u> ±0.0057	**0.0307** ±0.0050	0.4493 ±0.0019	<u>0.2185</u> ±0.0042
DPM-Solver++ (2M)	3.1069 ±0.0106	0.4242 ±0.0018	0.1130 $\pm 8.524\times10^{-4}$	0.4445 ±0.0011	0.2210 ±0.0017
UniPC (bh2)	3.2817 ±0.0018	0.4363 $\pm 8.324\times10^{-4}$	0.1882 $\pm 8.484\times10^{-4}$	0.4499 ±0.0016	0.2236 ±0.0014
Full DDIM	3.2028 ±0.0265	0.4335 ±0.0029	0.1565 ±0.0015	0.4482 ±0.0028	0.2236 ±0.0028

Method	Batch NFE	Sample NFE	Total calls	Seconds
StrideDiffusion	13.8125 ±0.0376	13.8125 ±0.0376	442.0000 ±1.2019	2.5379 ±0.0203
Matched uniform replay	13.8125 ±0.0376	13.8125 ±0.0376	442.0000 ±1.2019	2.1706 ±0.0223
Selected uniform history	32.0000 ±0.0000	32.0000 ±0.0000	1024.0000 ±0.0000	5.0171 ±0.0460
DPM-Solver++ (2M)	14.0000 ±0.0000	14.0000 ±0.0000	448.0000 ±0.0000	3.2194 ±0.0270
UniPC (bh2)	14.0000 ±0.0000	14.0000 ±0.0000	448.0000 ±0.0000	3.3581 ±0.0186
Full DDIM	500.0000 ±0.0000	500.0000 ±0.0000	1.600×10^{4} ±0.0000	75.6178 ±0.5905
Grid acquisition	13.8125 ±0.0376	13.8125 ±0.0376	442.0000 ±1.2019	2.6089 ±0.0205
Replay + grid pass	27.6250 ±0.0751	27.6250 ±0.0751	884.0000 ±2.4037	4.7795 ±0.0424

C-FID control	Mean Δ	Seed721	Seed732	Seed743
Matched uniform replay	0.0335 ±0.0809	0.0552	−0.0560	0.1015
Selected uniform history	0.1343 ±0.0368	0.1639	0.0930	0.1458
DPM-Solver++ (2M)	−0.0167 ±0.0276	−0.0462	−0.0124	0.0085
UniPC (bh2)	−0.1915 ±0.0274	−0.2151	−0.1980	−0.1615
Full DDIM	−0.1126 ±0.0468	−0.1288	−0.1492	−0.0599

Table 6: MuJoCo confirmation results: 1500 samples in 24 batches per draw, with batch size 64. Values are means ± sample SD across three independently trained checkpoints, each averaging three generation seeds. The metrics, cost accounting, paired differences and highlighting follow Appendix A.1. Lower quality scores are better. Replay costs exclude grid acquisition, reported separately.

Method	C-FID	Corr.	HF	Disc.	Pred.
StrideDiffusion	**0.0648** ±0.0019	0.2230 ±0.0067	0.0070 ±0.0016	0.0286 ±0.0039	**0.0073** $\pm 3.938\times10^{-4}$
Matched uniform replay	0.0760 ±0.0046	0.2191 ±0.0124	0.0025 $\pm 5.488\times10^{-4}$	0.0349 ±0.0083	0.0074 $\pm 2.018\times10^{-4}$
Selected uniform history	0.0699 ±0.0053	0.2199 ±0.0028	0.0029 $\pm 3.139\times10^{-4}$	0.0335 ±0.0089	**0.0073** $\pm 2.557\times10^{-4}$
DPM-Solver++ (2M)	0.0747 ±0.0029	**0.2070** ±0.0024	**0.0020** $\pm 3.421\times10^{-4}$	0.0306 ±0.0048	0.0075 $\pm 2.012\times10^{-4}$
UniPC (bh2)	0.0763 ±0.0034	0.2088 ±0.0023	0.0022 $\pm 2.548\times10^{-4}$	0.0275 ±0.0022	0.0077 $\pm 3.449\times10^{-4}$
Full DDIM	0.0702 ±0.0039	0.2117 ±0.0041	0.0023 $\pm 1.020\times10^{-4}$	**0.0244** ±0.0041	**0.0073** $\pm 2.934\times10^{-4}$

Method	Batch NFE	Sample NFE	Total calls	Seconds
StrideDiffusion	38.0046 ±0.0080	38.0047 ±0.0082	912.1111 ±0.1925	5.2081 ±0.0475
Matched uniform replay	38.0046 ±0.0080	38.0047 ±0.0082	912.1111 ±0.1925	4.4318 ±0.0454
Selected uniform history	96.0000 ±0.0000	96.0000 ±0.0000	2304.0000 ±0.0000	11.3138 ±0.1083
DPM-Solver++ (2M)	38.0000 ±0.0000	38.0000 ±0.0000	912.0000 ±0.0000	6.5656 ±0.0616
UniPC (bh2)	38.0000 ±0.0000	38.0000 ±0.0000	912.0000 ±0.0000	6.8716 ±0.0583
Full DDIM	1000.0000 ±0.0000	1000.0000 ±0.0000	2.400×10^4 ±0.0000	113.4715 ±1.0845
Grid acquisition	38.0046 ±0.0080	38.0047 ±0.0082	912.1111 ±0.1925	5.3112 ±0.0462
Replay + grid pass	76.0093 ±0.0160	76.0095 ±0.0164	1824.2222 ±0.3849	9.7430 ±0.0909

C-FID control	Mean Δ	Seed721	Seed732	Seed743
Matched uniform replay	−0.0112 ±0.0049	−0.0069	−0.0100	−0.0166
Selected uniform history	−0.0051 ±0.0034	−0.0041	−0.0088	−0.0023
DPM-Solver++ (2M)	−0.0099 ±0.0022	−0.0076	−0.0104	−0.0118
UniPC (bh2)	−0.0115 ±0.0020	−0.0092	−0.0130	−0.0122
Full DDIM	−0.0054 ±0.0022	−0.0034	−0.0077	−0.0051

Table 7: Energy confirmation results: 2048 samples in 32 batches per draw, with batch size 64. Values are means ± sample SD across three independently trained checkpoints, each averaging three generation seeds. The metrics, cost accounting, paired differences and highlighting follow Appendix A.1. Lower quality scores are better. Replay costs exclude grid acquisition, reported separately.

Method	C-FID	Corr.	HF	Disc.	Pred.
StrideDiffusion	**16.7116** ±0.1175	**6.4202** ±0.0097	**0.0658** ±0.0091	0.4998 ±1.565×10^{-4}	**0.2558** ±4.968×10^{-4}
Matched uniform replay	17.2315 ±0.1310	6.4352 ±0.0214	0.0848 ±0.0045	0.4998 ±2.070×10^{-4}	0.2573 ±2.805×10^{-4}
Selected uniform history	17.3620 ±0.1235	6.4305 ±0.0231	0.0973 ±0.0137	0.4998 ±1.565×10^{-4}	0.2568 ±6.325×10^{-4}
DPM-Solver++ (2M)	17.1169 ±0.1529	6.4324 ±0.0221	0.1110 ±0.0045	**0.4993** ±9.485×10^{-4}	0.2563 ±0.0014
UniPC (bh2)	17.3382 ±0.1272	6.4396 ±0.0232	0.1080 ±0.0032	0.4996 ±3.585×10^{-4}	0.2576 ±0.0016
Full DDIM	17.2631 ±0.1188	6.4324 ±0.0204	0.0931 ±0.0074	0.4997 ±2.347×10^{-4}	0.2568 ±0.0020

Method	Batch NFE	Sample NFE	Total calls	Seconds
StrideDiffusion	58.4167 ±0.3254	58.4167 ±0.3254	1869.3333 ±10.4137	14.2017 ±0.1390
Matched uniform replay	58.4167 ±0.3254	58.4167 ±0.3254	1869.3333 ±10.4137	12.6272 ±0.1031
Selected uniform history	96.0000 ±0.0000	96.0000 ±0.0000	3072.0000 ±0.0000	20.9518 ±0.0903
DPM-Solver++ (2M)	58.5556 ±0.5092	58.5556 ±0.5092	1873.7778 ±16.2936	17.0185 ±0.2079
UniPC (bh2)	58.5556 ±0.5092	58.5556 ±0.5092	1873.7778 ±16.2936	17.6822 ±0.2218
Full DDIM	1000.0000 ±0.0000	1000.0000 ±0.0000	3.200×10^4 ±0.0000	212.2191 ±0.9857
Grid acquisition	58.4167 ±0.3254	58.4167 ±0.3254	1869.3333 ±10.4137	14.3314 ±0.1240
Replay + grid pass	116.8333 ±0.6509	116.8333 ±0.6509	3738.6667 ±20.8273	26.9587 ±0.2226

C-FID control	Mean Δ	Seed721	Seed732	Seed743
Matched uniform replay	−0.5199 ±0.0245	−0.5019	−0.5101	−0.5478
Selected uniform history	−0.6504 ±0.0438	−0.6433	−0.6107	−0.6974
DPM-Solver++ (2M)	−0.4053 ±0.0361	−0.3636	−0.4249	−0.4274
UniPC (bh2)	−0.6266 ±0.0331	−0.6134	−0.6023	−0.6643
Full DDIM	−0.5515 ±0.0176	−0.5477	−0.5361	−0.5707

Table 8: fMRI confirmation results: 1416 samples in 23 batches per draw, with batch size 64. Values are means ± sample SD across three independently trained checkpoints, each averaging three generation seeds. The metrics, cost accounting, paired differences and highlighting follow Appendix A.1. Lower quality scores are better. Replay costs exclude grid acquisition, reported separately.

Method	C-FID	Corr.	HF	Disc.	Pred.
StrideDiffusion	0.5728 ±0.0236	3.7332 ±0.0695	0.1741 ±0.0381	0.4499 ±0.0044	0.1016 ±7.097×10^{-4}
Matched uniform replay	0.6491 ±0.0283	3.8393 ±0.0656	0.3087 ±0.0215	0.4554 ±0.0045	0.0998 ±5.538×10^{-4}
Selected uniform history	0.7254 ±0.0415	**3.5033** ±0.0384	0.6962 ±0.0376	0.4671 ±0.0042	0.1068 ±8.699×10^{-4}
DPM-Solver++ (2M)	**0.5547** ±0.0171	3.5219 ±0.0510	**0.0797** ±0.0245	**0.4433** ±0.0144	0.0997 ±6.912×10^{-4}
UniPC (bh2)	0.6182 ±0.0245	3.6287 ±0.0618	0.2154 ±0.0226	0.4487 ±0.0178	**0.0986** ±4.276×10^{-4}
Full DDIM	0.5941 ±0.0251	3.8145 ±0.0545	0.2338 ±0.0273	0.4553 ±0.0081	0.1015 ±6.155×10^{-4}

Method	Batch NFE	Sample NFE	Total calls	Seconds
StrideDiffusion	61.4106 ±0.5290	61.0797 ±0.5744	1412.4444 ±12.1670	13.2373 ±0.1176
Matched uniform replay	61.4106 ±0.5290	61.0797 ±0.5744	1412.4444 ±12.1670	12.0437 ±0.1071
Selected uniform history	48.0000 ±0.0000	48.0000 ±0.0000	1104.0000 ±0.0000	9.4252 ±0.0348
DPM-Solver++ (2M)	61.5556 ±0.3849	61.5556 ±0.3849	1415.7778 ±8.8527	15.3389 ±0.1067
UniPC (bh2)	61.5556 ±0.3849	61.5556 ±0.3849	1415.7778 ±8.8527	15.8274 ±0.0672
Full DDIM	1000.0000 ±0.0000	1000.0000 ±0.0000	2.300×10^4 ±0.0000	192.5813 ±0.9633
Grid acquisition	61.4106 ±0.5290	61.0797 ±0.5744	1412.4444 ±12.1670	13.3503 ±0.1201
Replay + grid pass	122.8213 ±1.0580	122.1594 ±1.1488	2824.8889 ±24.3341	25.3940 ±0.2176

C-FID control	Mean Δ	Seed721	Seed732	Seed743
Matched uniform replay	−0.0762 ±0.0051	−0.0748	−0.0819	−0.0720
Selected uniform history	−0.1526 ±0.0613	−0.1683	−0.0849	−0.2044
DPM-Solver++ (2M)	0.0182 ±0.0065	0.0130	0.0254	0.0161
UniPC (bh2)	−0.0453 ±0.0021	−0.0430	−0.0459	−0.0470
Full DDIM	−0.0212 ±0.0033	−0.0230	−0.0233	−0.0175

A.2 Validation candidates and selection cost

Figure 7 shows all candidates rather than only the selected settings. Selection minimizes validation C-FID separately within each family under the same 96-call ceiling. The selected spectral profiles for Sines, Stocks, ETTh, MuJoCo, Energy and fMRI are respectively Stocks, Sines, ETTh, Stocks, Energy and fMRI. Their independently selected uniform-history budgets are 64, 96, 32, 96, 96 and 48. Table 9 reports all recorded selection costs.

Table 9: Validation selection costs in seconds. Encoder fitting is a component, not an extra cost to add again to stage time. Stage time sums every recorded attempt. Completed-attempt timings exclude the first selection-lock verification, shown separately. The Attempts column marks any timeout bounds. Each uses the whole PBS allocation and may include prior partial lock verification, so it is not a pure candidate-stage duration. Do not add overlapping lock-check time to these bounds. These are one-time selection costs, not sampling latencies.

Dataset	Attempts	Encoder fit	Selection stage	Lock check
Sines	1	6.0285	73.8767	4.1618
Stocks	1	6.1127	75.9758	4.3070
ETTh	1	5.8881	75.6084	4.0656
MuJoCo	1	10.0874	92.5234	4.1808
Energy	1	8.9227	130.7495	4.2095
fMRI	1	9.8950	136.4716	4.2934

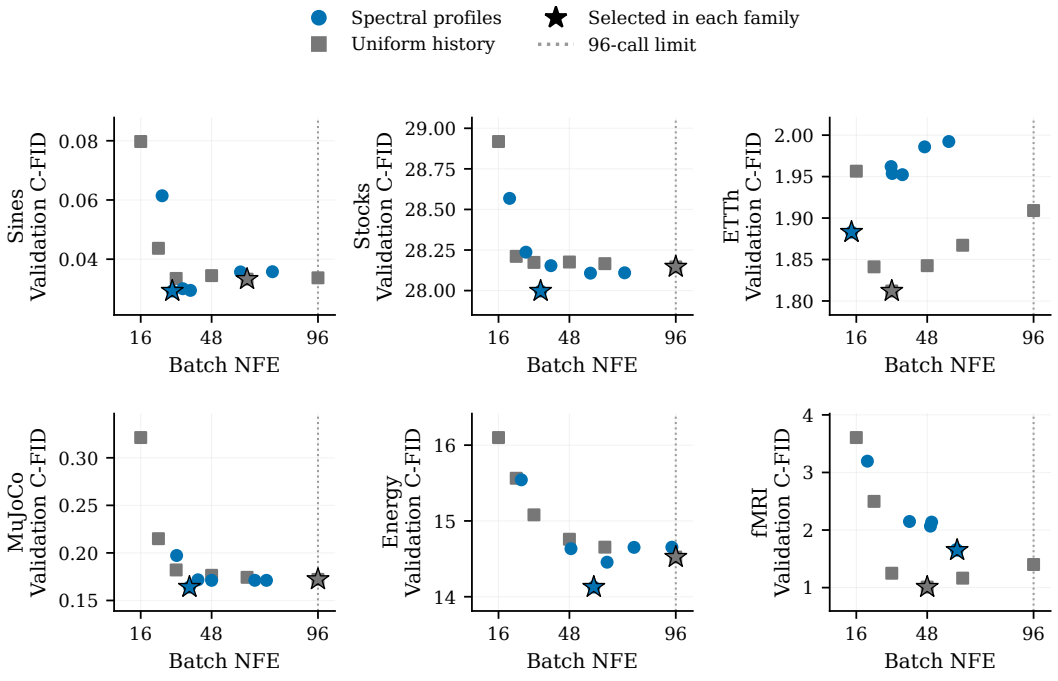

Figure 7: All 72 validation candidates: six spectral profiles and six uniform-history budgets per dataset. Points are means over three generation seeds on one development checkpoint, not independent training replicates. Stars mark the frozen selection within each family and the dotted line marks the 96-call ceiling. All candidates completed and are budget-feasible. Compare points only within a panel, whose C-FID scale differs from the others. No confirmation scores enter selection. Exact candidate identities, C-FID, NFE and status are retained in `plot_data/core_validation_candidates.csv`.

B EARLIER FIXED-CHECKPOINT COMPARISONS

B.1 PROTOCOL AND COMPLETE-METHOD COMPARISON

Table captions define the colors and ranks. Highlighting identifies methods and descriptive mean comparisons, not statistical significance.

We use one frozen Diffusion-TS checkpoint per dataset, with $L = 24$. Stocks, ETTh, and Energy are observational datasets. Sines and MuJoCo are simulated, and fMRI uses the benchmark's realistic BOLD simulation. The diffusion horizon is 500 for Sines, Stocks, and ETTh and 1,000 for the other three. Every method shares its initial noise and sample count within a generation seed. We use seeds 12345, 2026, and 2027.

NFE counts actual denoiser forwards. Fixed-grid DDIM, DPM-Solver++(2M), and UniPC-bh2 use the nearest integer to StrideDiffusion's measured batch-average NFE for each dataset and seed. Same-grid DDIM replays every per-batch time grid chosen by StrideDiffusion, while online Spectral-DDIM generates its own grid and therefore has its own measured NFE. DPM-Solver++ and UniPC use their official implementations with data-prediction adapters. No weights are updated.

Context-FID uses up to 5,000 real and generated sequences and a shared TS2Vec encoder trained only on reference windows. Correlational measures lag-zero correlation-matrix mismatch. Discriminative measures the deviation of a trained real/fake classifier from chance. Predictive measures real-data one-step MAE after training a predictor on generated data. For the last three metrics we first average five evaluation seeds within each generation seed, then report the mean and sample standard deviation across the three generation seeds. Appendix E defines the evaluators and gives the complete 36-row table. Reference windows follow the frozen benchmark preprocessing, so these scores measure distribution matching to that reference, not out-of-distribution generalization.

Table 10: Matched-budget comparison on Diffusion-TS. C-FID (mean ± sample standard deviation over three generation seeds), lower is better. Ours uses 35.0, 32.2, 14.0, 59.3, 59.0, and 67.6 mean NFE in dataset order. Fixed controls match rounded per-seed budgets. Online Spectral-DDIM has its own measured NFE, given in Appendix E. Underline marks the second-lowest distinct displayed mean within the same comparison groups as bold, including ties. Neither mark indicates significance.

Data	**Stride Diffusion**	Same-grid DDIM	Spectral-DDIM	Uniform DDIM	DPM++(2M)	UniPC-bh2
Sines	**0.0120 ±0.0014**	0.0163 ±0.0031	0.0163 ±0.0031	0.0160 ±0.0022	0.0193 ±0.0012	0.0192 ±0.0010
Stocks	**0.1170 ±0.0039**	0.2117 ±0.0132	0.2117 ±0.0132	0.2250 ±0.0228	0.1797 ±0.0169	0.2229 ±0.0194
ETTh	**0.1502 ±0.0265**	0.3832 ±0.0563	0.3814 ±0.0572	0.3553 ±0.0484	0.1845 ±0.0349	0.2300 ±0.0323
MuJoCo	0.0442 ±0.0028	0.0393 ±0.0051	**0.0392 ±0.0051**	0.0408 ±0.0043	0.0447 ±0.0032	0.0479 ±0.0031
Energy	0.0978 ±0.0099	0.1128 ±0.0112	0.1128 ±0.0112	0.1045 ±0.0094	**0.0776 ±0.0083**	0.0965 ±0.0101
fMRI	**0.1184 ±0.0079**	0.1965 ±0.0130	0.1976 ±0.0127	0.2149 ±0.0128	0.1243 ±0.0082	0.1545 ±0.0095

Table 10 shows that the complete sampler has lower mean Context-FID than each of DPM-Solver++ and UniPC on five datasets. Energy favors the strong solvers. Compared with uniform DDIM, StrideDiffusion improves five datasets and loses on MuJoCo. Across all four metrics, StrideDiffusion is numerically lower than uniform DDIM in 20 of 24 dataset–metric cells. These are descriptive comparisons over three generation seeds, not significance tests or independent retrainings.

Measured latency. Figure 8 shows quality and sampling time together. StrideDiffusion is faster than DPM-Solver++ and UniPC in all six cases, with lower mean Context-FID on five datasets. Uniform DDIM is faster, but has higher mean Context-FID on five datasets. Thus neither NFE nor time alone describes the tradeoff. All timings use the same NVIDIA H200 GPU model, exclude three warm-up forwards, and synchronize CUDA around each batch. They include the denoiser, grid construction, update arithmetic, in-loop randomness, and FFT controller. They exclude model/data loading, initial-noise construction and transfer, saving, and scoring. Appendix E.1 gives the exact times and separately identifies timing-only architectures. Newly trained DiffWave/DiffTime adaptations appear in Appendix J.

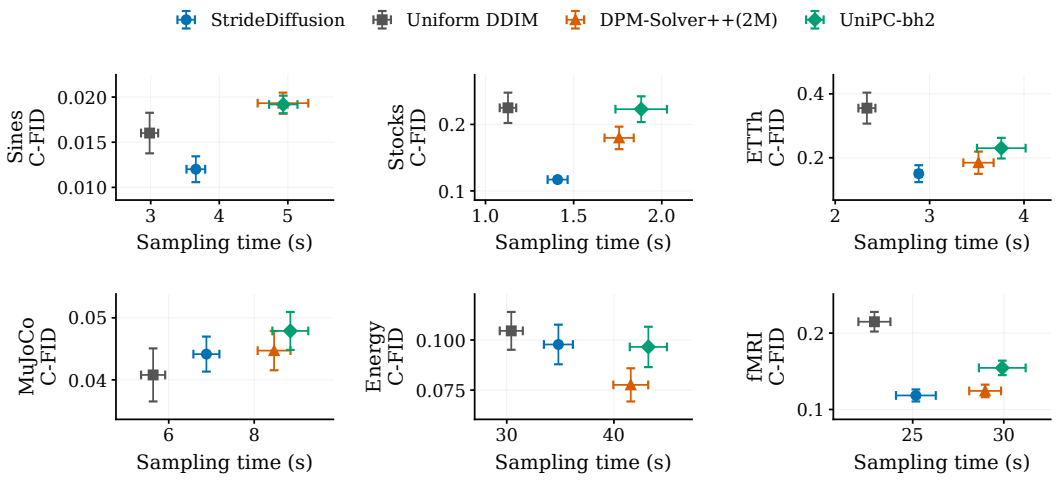

Figure 8: Quality and sampling latency for StrideDiffusion and three fixed-grid baselines on the same frozen Diffusion-TS checkpoints. Component controls from Table 10 are not plotted here. Lower and further left is better. Points and error bars are mean ± sample SD over three generation seeds (12345, 2026, 2027), not independent retrainings. Fixed-grid baselines round our mean NFE to the nearest integer. Times cover each dataset's complete sample count. Compare methods within a panel, as sample counts and axis ranges differ. Tables 10 and 15 give the exact values.

How does the controller allocate calls? Figure 9 revisits the original per-batch traces. Sines and Energy use nearly fixed call counts, while Stocks, MuJoCo, and fMRI vary more across batches. Similar call counts can still correspond to different visited times. Appendix Figure 18 also shows pooled gate decisions over normalized iteration position. These historical traces illustrate controller behavior and are separate from the three-seed quality and timing comparison.

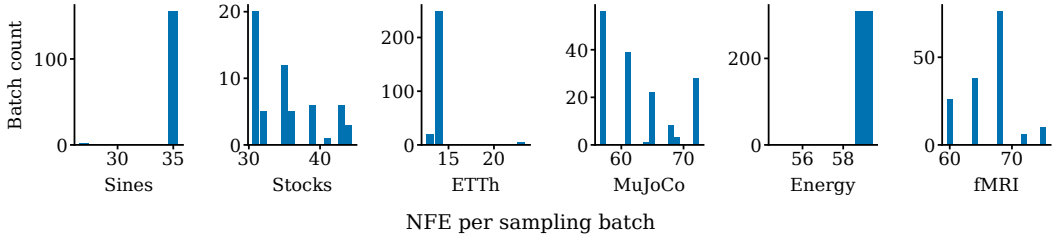

Figure 9: Call allocation in the original controller traces at seed 12345, replotted from the batch logs underlying Figure 18. Each bar counts batches at an integer NFE, with one shared grid per batch. From left to right, the six datasets contain 157, 58, 272, 157, 309, and 156 batches. Axis scales differ across panels. These are within-run distributions, not independent-seed uncertainty or a timing comparison.

Three interventions separate the components. First, DDIM replay on StrideDiffusion's selected grid removes the history term while fixing times and NFE. The complete method has lower Context-FID in five of six cases. MuJoCo favors same-grid DDIM. Second, online Spectral-DDIM retains the controller but uses DDIM throughout. It beats uniform DDIM on only three datasets. Third, a uniform grid under the same conditional history-update rule tests the effect of grid selection without replacing that rule. This control favors StrideDiffusion on five datasets (Table 26), with Energy again the exception.

The last control, reported in Appendix G, fixes the formula and branching condition, not the realized sequence of branches. The grid affects whether a transition is a one-step DDIM fallback or a history-corrected jump. Our stricter replay control matches the exact NFE and the history/DDIM choice at each update index in each batch. The diffusion times differ between grids. Figure 10 compares the spectrum-derived grid with this uniform replay. Both use the same approximate-history rule,

checkpoint and initial noise. Their C-FID encoder is shared within each run and fitted separately from Table 10, so absolute scores across these experiments should not be ranked.

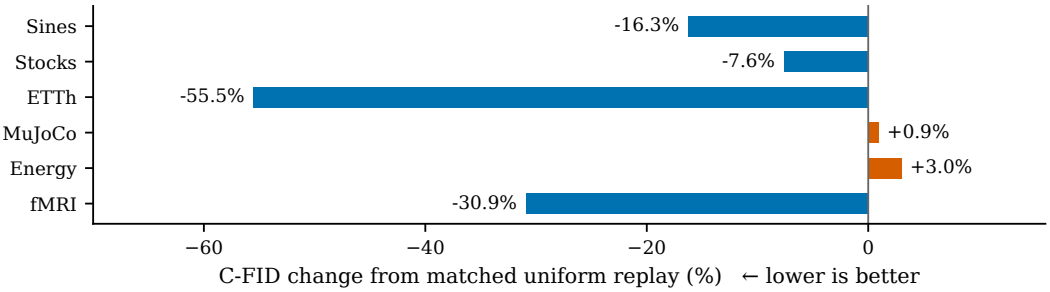

Figure 10: Grid attribution with identical NFE and history/DDIM update positions. Bars show $100(\overline{\text{C-FID}}_{\text{spectral}}/\overline{\text{C-FID}}_{\text{uniform}} - 1)$ over three generation/evaluator seeds and one frozen checkpoint per dataset. Left of zero favors the spectral grid. This is a ratio of means, not a significance test. The exact means, SDs and NFE appear in Appendix Table 25.

Under this stricter control, the spectral grid improves mean C-FID on 4/6 datasets, with Energy and MuJoCo as exceptions. The gains depend on the update rule, and grids recorded from independent noise also recover the main improvements. The evidence therefore supports choosing a spectral grid together with its history update. It does not establish that the same grid helps every solver or must be recomputed for each batch. Appendices G, E.2, and F report the complete donor-grid, solver, and fixed-budget controls.

Historical absolute-power diagnostics. Figure 5 in the main text summarizes this earlier fixed-checkpoint comparison. The definitions and complete means and SDs remain in Appendix E.4. These diagnostics are separate from the raw high-frequency power error in the confirmation experiment.

C Bandwise Displacement: Proof and Scope

C.1 Fourier convention

Let $\mathcal{F}$ be the unitary discrete Fourier transform along time. For a conjugate-symmetric, binary diagonal mask M_B, set $\mathcal{P}_B = \mathcal{F}^{-1} M_B \mathcal{F}$. Then $M_B = M_B^* = M_B^2$, so $\mathcal{P}_B$ is linear and orthogonal, and the energy $\|\mathcal{P}_B x\|_F^2$ is the sum of two-sided spectral powers in the band. A real one-sided FFT with the same unitary normalization (`norm='ortho'`) gives the same energy when the interior positive-frequency powers receive weight two and DC and the even-length Nyquist bin receive weight one. With an unnormalized forward FFT, the weighted sum must instead be divided by L. The gate's mean-bin power is a different statistic and need not have these projector properties.

Numerical phase convention. The controller computes $\arg z$ with `torch.angle`, using the runtime's numerical angle convention in $[-\pi, \pi]$. Our CPU regression test with PyTorch 2.4.1+cu118 gives $\operatorname{atan2}(\pm 0, +0) = \pm 0$ and $\operatorname{atan2}(\pm 0, -0) = \pm \pi$. It also gives zero phase for an all-zero FFT multiplied by its conjugate. These are tested runtime behaviors, not a guarantee about signed zeros on every backend. Zero-power bins still receive the ε regularizer in the phase statistic. The bandwise proofs below do not use this statistic.

C.2 Proof of Proposition 1

Substituting Equation (1) into Equation (2) gives $D = ax + b\hat{x}_0$. Applying the linear projector to $U - x = (a-1)x + b\hat{x}_0 + R$ gives Equation (10). The triangle inequality gives Equation (11).

For an interior history transition, $0 < \rho < \rho' < 1$. Since

$$e^h = \sqrt{\frac{\rho'(1-\rho)}{\rho(1-\rho')}},$$

the uncorrected update satisfies

$$\sqrt{\rho'/\rho}\,x - \sqrt{1-\rho'}(e^h-1)\hat{\epsilon} = \sqrt{\rho'}\hat{x}_0 + \sqrt{1-\rho'}\hat{\epsilon} \\ = D_{\rho\to\rho'}(x).$$

The additional part of Equation (3) is exactly Equation (12). This proves the result for the implemented approximate update without substituting a standard solver for it. The terminal transition uses DDIM, so $R = 0$ there and no infinite log-SNR is evaluated.

C.3 Proof of Corollary 1

For fixed ρ and $r \in [\rho, \rho']$, the derivatives of the affine coefficients obey

$$\left|\frac{\partial a(\rho,r)}{\partial r}\right| = \frac{1}{2\sqrt{1-\rho}\sqrt{1-r}} \le \frac{1}{2\gamma}, \qquad \left|\frac{\partial b(\rho,r)}{\partial r}\right| \le \frac{1}{2\sqrt{\rho_{\min}}} + \frac{1}{2\gamma}.$$

At $r = \rho$, $a = 1$ and $b = 0$. Integrating these derivative bounds over $[\rho, \rho']$ and using both band norms at most ϵ_B proves the stated constant C. The argument fixes the model prediction at the current state. It does not differentiate the network or compare with a trajectory that repeatedly recomputes predictions.

C.4 Why the gate is not the sufficient condition

A band can be constant across two visited states while its amplitude is arbitrarily large. For example, $\mathcal{P}_B x_t = \mathcal{P}_B x_u = Av$ with $\|v\|_F = 1$ gives zero change for every A. Relative power also fails to bound absolute power, even with the regularizer: set $P[B] = A$ and another band's power to A^2. Then $\pi[B] = A/(A + A^2 + \varepsilon) \to 0$ while $P[B] \to \infty$. Neither condition checks $\mathcal{P}_B\hat{x}_0$ or the previous prediction needed by R.

If one separately bounded $\|\mathcal{P}_B(\hat{\epsilon}_t - \hat{\epsilon}_{\text{prev}})\|_F$ by η_B, Equation (11) would add

$$\sqrt{1-\rho'}|e^h - 1|\frac{|h|}{2h_0}\eta_B.$$

This is an explicit residual requirement, not a bound furnished by the controller. In particular, labeling a history update "second order" would not establish this requirement. The theorem-level claim in this paper is restricted to the displayed local algebra and sufficient condition. The numerical benefit of the complete rule is assessed by the controls in Section 5.2.

D Implementation and experimental settings

D.1 Exact sampling loop

Algorithm 1 specifies the StrideDiffusion sampling loop on Diffusion-TS. Appendix A gives the validation-selected settings for the main comparison. The settings and reference checks below describe the earlier fixed-checkpoint studies. The denoiser predicts a clean sample at every visited state. Storing the preceding noise prediction does not require another forward. This history is updated after every transition, including DDIM fallbacks.

All configurations considered here use $k_s = 1$. The stored noise schedules in all six original checkpoints satisfy the strict interior ordering required above. For $t > 0$, every branch gives $0 \le s < t$. At $t = 0$, the unit stride gives exactly $s = -1$. Thus every history call has $0 < \rho_t < \rho_s < 1$, and only DDIM reaches $\rho_{-1} = 1$. The terminal clean prediction is counted as a forward. Matching nominal grid points without this endpoint convention can produce unequal NFE. We count calls in the execution loop rather than infer them from a solver's step argument. For the earlier fixed-checkpoint studies, the controlled implementation was checked against the original sampler on one two-sample batch for each of six datasets and three generation seeds, using identical initial noise. All 18 checks recorded the same time grid and NFE, with maximum absolute output difference zero. The checker's acceptance tolerance was 10^{-6}. These finite regression checks do not prove equivalence for all inputs. The experiment runner rejects non-finite final samples, but the frozen gate has no internal check for non-finite inputs or intermediate statistics.

Algorithm 1 StrideDiffusion on Diffusion-TS

Require: Frozen denoiser, schedule ρ, strides (k_b, k_m, k_s), final window K, thresholds, batch size $n \geq 1$
Require: $T, k_b, k_m \in \mathbb{N}_{>0}$, $k_s = 1$, $K \in \mathbb{N}_0$, $0 < \rho_{T-1} < \cdots < \rho_0 < 1$
Draw a Gaussian batch $x \in \mathbb{R}^{n \times L \times d}$, set $x_{\text{prev}} \leftarrow x$, $e_{\text{prev}} \leftarrow \varnothing$, $t \leftarrow T - 1$
while $t \geq 0$ **do**
    Predict and clip $\hat{x}_0$, reconstruct e using Eq. (1)
    Compute $\mathcal{A}$ from batch-averaged statistics of x, x_{prev} using Eq. (8)
    if $t \leq K$ **then**
        $k \leftarrow k_s$
    else if $\mathcal{A} = \varnothing$ **then**
        $k \leftarrow \min(k_b, t)$
    else if $\mathcal{A} \subseteq \{0\}$ **then**
        $k \leftarrow \min(k_m, t)$
    else
        $k \leftarrow \min(k_s, t)$
    end if
    $k \leftarrow \max(1, k)$, $s \leftarrow t - k$
    if $e_{\text{prev}} \neq \varnothing$ and $k > 1$ and $s \geq 0$ **then**
        $x_{\text{next}} \leftarrow H_{t \to s}(x)$ using Eq. (4)
    else
        $x_{\text{next}} \leftarrow D_{t \to s}(x)$ using Eq. (2), with $\rho_{-1} = 1$
    end if
    $(x_{\text{prev}}, e_{\text{prev}}, x, t) \leftarrow (x, e, x_{\text{next}}, s)$
end while
return the batch of n samples x

Table 11: Settings for the earlier fixed-checkpoint Diffusion-TS studies, not the validation-selected main comparison. All sequences have length 24. k_b, k_m, k_s are the large, medium, and small strides. K is the final small-step window.

Data	T	Samples	Batch	(k_b, k_m, k_s)	K	τ_E	(τ_P, τ_ϕ)
Sines	500	10,000	512	$(50, 20, 1)$	20	.10	$(.01, .04)$
Stocks	500	3,662	512	$(50, 10, 1)$	20	.10	$(.01, .04)$
ETTh	500	17,397	512	$(50, 10, 1)$	12	.70	$(.02, .08)$
MuJoCo	1000	10,000	512	$(20, 5, 1)$	20	.50	$(.02, .04)$
Energy	1000	19,712	256	$(40, 5, 1)$	12	.10	$(.01, .06)$
fMRI	1000	9,977	256	$(20, 5, 1)$	20	.50	$(.01, .06)$

D.2 Earlier samplers and matching

The earlier fixed-checkpoint studies use the following controls. Same-grid DDIM uses recorded per-batch grids from the complete method. Spectral-DDIM reruns the controller on its own DDIM trajectory. Uniform DDIM spaces visited indices uniformly. DPM-Solver++(2M) uses the released multistep data-prediction solver. UniPC uses its released bh2 variant. These last three controls match the rounded mean NFE of the complete method separately for each dataset and generation seed. Thus their batch-averaged NFE can differ slightly from ours, even with a common target budget. The endpoint-matched DPM++ experiment in Appendix E.2 is a separate pair with shared terminal treatment.

D.3 Dataset and model scope

The earlier fixed-checkpoint studies keep each Diffusion-TS checkpoint fixed, including its spectral architectural and loss choices. They test samplers rather than differences between training procedures. Sines and MuJoCo are synthetic benchmarks. The fMRI arrays come from the benchmark's realistic BOLD simulator and should not be interpreted as independently collected clinical scans. For these earlier studies, the reference arrays are normalized training windows from the benchmark

preprocessing, and sampling-seed variation does not measure variation across independently trained diffusion models. The main confirmation experiment instead uses separate data partitions and three independently trained models per dataset, as specified in Appendix A.

E EARLIER CONTROLLED RESULTS AND EVALUATION DEFINITIONS

This appendix retains the earlier fixed-checkpoint comparisons and their separately specified diagnostics. Their reference sets, encoder fits, generation seeds and sampling measurements are separate from the main confirmation experiment in Appendix A. Absolute C-FID values should not be ranked across the two protocols.

Context-FID. We fit one TS2Vec encoder to reference sequences within each dataset and generation seed, then use that encoder for every sampler in the comparison. The encoder uses 320 output dimensions, batch size 8, learning rate 10^{-3}, and full-series pooling. We select at most 5,000 reference windows without replacement and take the corresponding number of generated windows. For representation means μ_r, μ_g and covariances Σ_r, Σ_g, the metric is

$$\text{C-FID} = \|\mu_r - \mu_g\|_2^2 + \operatorname{tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r^{1/2}\Sigma_g\Sigma_r^{1/2})^{1/2}\right). \tag{14}$$

This is a time-series embedding metric, not an image Inception score. A separate experiment can fit a separate encoder, so small differences between tables with different evaluator fits are not paired method effects.

Correlational. The benchmark evaluator standardizes channels over samples and time, computes lag-zero channel products, and averages over samples and time. It reports one tenth of the sum of absolute differences over the lower triangle, including its diagonal. This is a cross-channel statistic at lag zero, not a full temporal autocorrelation score. The auxiliary correlation-signature metric below instead averages the absolute difference and is not interchangeable with this score.

Discriminative and Predictive. Discriminative is $|a - 0.5|$, where a is the test accuracy of a one-layer GRU classifier trained for 2,000 updates to distinguish reference from generated sequences. Predictive trains a one-layer GRU for 5,000 updates on generated sequences to predict the next value of the last variable from the other variables. It evaluates MAE on real sequences under the benchmark's normalized scale. Both GRUs have hidden width $\max(\lfloor d/2 \rfloor, 1)$. We average five evaluator seeds within each generation seed and report the mean and sample standard deviation across generation seeds 12345, 2026, and 2027. These nested evaluator runs are not 15 independent generator retrainings.

Table 12: Complete four-metric results, part 1. Mean ± sample standard deviation across three generation seeds. All four metrics are lower-is-better. Five evaluator seeds are averaged within each generation seed for Correlational, Discriminative, and Predictive. Underline marks the second-lowest distinct displayed mean within the same comparison groups as bold, including ties. Neither mark indicates significance.

Data	Method	NFE	C-FID	Correlational	Discriminative	Predictive
Sines	**StrideDiffusion**	35.0	**0.0120 ±0.0014**	**0.0144 ±0.0017**	0.0237 ±0.0068	**0.09370 ±0.00022**
Sines	Same-grid DDIM	35.0	0.0163 ±0.0031	0.0206 ±0.0025	0.0363 ±0.0042	0.09414 ±0.00013
Sines	Spectral-DDIM	35.0	0.0163 ±0.0031	0.0206 ±0.0025	0.0329 ±0.0033	0.09399 ±0.00019
Sines	Uniform DDIM	35.0	0.0160 ±0.0022	0.0150 ±0.0015	0.0278 ±0.0007	0.09456 ±0.00011
Sines	DPM++(2M)	35.0	0.0193 ±0.0012	0.0238 ±0.0024	**0.0120 ±0.0034**	0.09405 ±0.00015
Sines	UniPC-bh2	35.0	0.0192 ±0.0010	0.0185 ±0.0018	0.0273 ±0.0102	0.09463 ±0.00018
Stocks	**StrideDiffusion**	32.2	**0.1170 ±0.0039**	0.0357 ±0.0017	**0.0737 ±0.0046**	**0.03718 ±0.00024**
Stocks	Same-grid DDIM	32.2	0.2117 ±0.0132	0.0123 ±0.0030	0.1076 ±0.0042	0.03754 ±0.00001
Stocks	Spectral-DDIM	32.2	0.2117 ±0.0132	0.0123 ±0.0030	0.1080 ±0.0079	0.03754 ±0.00004
Stocks	Uniform DDIM	32.0	0.2250 ±0.0228	**0.0087 ±0.0023**	0.1169 ±0.0056	0.03747 ±0.00015
Stocks	DPM++(2M)	32.0	0.1797 ±0.0169	0.0215 ±0.0030	0.1059 ±0.0083	0.03724 ±0.00011
Stocks	UniPC-bh2	32.0	0.2229 ±0.0194	0.0152 ±0.0033	0.1181 ±0.0064	0.03745 ±0.00014
ETTh	**StrideDiffusion**	14.0	**0.1502 ±0.0265**	0.0819 ±0.0033	**0.0825 ±0.0085**	0.12468 ±0.00022
ETTh	Same-grid DDIM	14.0	0.3832 ±0.0563	0.0888 ±0.0012	0.1263 ±0.0010	0.12592 ±0.00163
ETTh	Spectral-DDIM	15.1	0.3814 ±0.0572	0.0888 ±0.0012	0.1255 ±0.0034	0.12403 ±0.00038
ETTh	Uniform DDIM	14.0	0.3553 ±0.0484	0.0826 ±0.0024	0.1220 ±0.0009	0.12480 ±0.00034
ETTh	DPM++(2M)	14.0	0.1845 ±0.0349	0.0797 ±0.0022	0.0984 ±0.0022	0.12422 ±0.00211
ETTh	UniPC-bh2	14.0	0.2300 ±0.0323	**0.0743 ±0.0041**	0.0992 ±0.0014	**0.12370 ±0.00128**

Table 13: Complete four-metric results, part 2. Mean ± sample standard deviation across three generation seeds. All four metrics are lower-is-better. Five evaluator seeds are averaged within each generation seed for Correlational, Discriminative, and Predictive. Underline marks the second-lowest distinct displayed mean within the same comparison groups as bold, including ties. Neither mark indicates significance.

Data	Method	NFE	C-FID	Correlational	Discriminative	Predictive
MuJoCo	**StrideDiffusion**	59.3	0.0442 ±0.0028	0.1900 ±0.0095	0.0355 ±0.0037	**0.00765 ±0.00025**
MuJoCo	Same-grid DDIM	59.3	0.0393 ±0.0051	0.1930 ±0.0100	0.0341 ±0.0007	0.00786 ±0.00044
MuJoCo	Spectral-DDIM	60.8	**0.0392 ±0.0051**	0.1930 ±0.0100	**0.0328 ±0.0019**	0.00792 ±0.00035
MuJoCo	Uniform DDIM	59.0	0.0408 ±0.0043	0.1925 ±0.0105	0.0329 ±0.0037	0.00800 ±0.00056
MuJoCo	DPM++(2M)	59.0	0.0447 ±0.0032	**0.1893 ±0.0098**	0.0372 ±0.0029	0.00790 ±0.00042
MuJoCo	UniPC-bh2	59.0	0.0479 ±0.0031	0.1909 ±0.0108	0.0373 ±0.0069	0.00768 ±0.00021
Energy	**StrideDiffusion**	59.0	0.0978 ±0.0099	**0.8265 ±0.0321**	0.1301 ±0.0045	**0.25075 ±0.00003**
Energy	Same-grid DDIM	59.0	0.1128 ±0.0112	0.8430 ±0.0396	**0.1299 ±0.0041**	**0.25075 ±0.00002**
Energy	Spectral-DDIM	59.0	0.1128 ±0.0112	0.8430 ±0.0396	0.1315 ±0.0053	0.25078 ±0.00002
Energy	Uniform DDIM	59.0	0.1045 ±0.0094	0.8582 ±0.0272	0.1448 ±0.0025	0.25082 ±0.00011
Energy	DPM++(2M)	59.0	**0.0776 ±0.0083**	0.8381 ±0.0269	0.1499 ±0.0004	0.25080 ±0.00004
Energy	UniPC-bh2	59.0	0.0965 ±0.0101	0.8779 ±0.0285	0.1548 ±0.0013	0.25085 ±0.00003
fMRI	**StrideDiffusion**	67.6	**0.1184 ±0.0079**	0.9988 ±0.0123	0.0947 ±0.0195	0.09944 ±0.00007
fMRI	Same-grid DDIM	67.6	0.1965 ±0.0130	1.4122 ±0.0236	0.1978 ±0.0089	0.09939 ±0.00013
fMRI	Spectral-DDIM	66.9	0.1976 ±0.0127	1.4137 ±0.0234	0.2035 ±0.0114	0.09942 ±0.00009
fMRI	Uniform DDIM	67.7	0.2149 ±0.0128	1.3454 ±0.0187	0.1763 ±0.0090	0.09928 ±0.00004
fMRI	DPM++(2M)	67.7	0.1243 ±0.0082	**0.9883 ±0.0121**	**0.0915 ±0.0078**	0.09928 ±0.00011
fMRI	UniPC-bh2	67.7	0.1545 ±0.0095	1.1614 ±0.0192	0.1123 ±0.0196	**0.09924 ±0.00016**

Table 14: Four-metric comparison at matched budgets. Each pair is StrideDiffusion / uniform DDIM, shown as the mean over three generation seeds. The complete six-method results with uncertainty and NFE are in Tables 12 and 13. Lower is better. Blue backgrounds identify the Ours value in each pair. Bold and underline mark the lower and higher distinct displayed means, respectively, with ties sharing bold. These are descriptive comparisons, not significance tests.

Data	C-FID	Correlational	Discriminative	Predictive
Sines	**0.0120** / 0.0160	**0.0144** / 0.0150	**0.0237** / 0.0278	**0.09370** / 0.09456
Stocks	**0.1170** / 0.2250	0.0357 / **0.0087**	**0.0737** / 0.1169	**0.03718** / 0.03747
ETTh	**0.1502** / 0.3553	**0.0819** / 0.0826	**0.0825** / 0.1220	**0.12468** / 0.12480
MuJoCo	0.0442 / **0.0408**	**0.1900** / 0.1925	0.0355 / **0.0329**	**0.00765** / 0.00800
Energy	**0.0978** / 0.1045	**0.8265** / 0.8582	**0.1301** / 0.1448	**0.25075** / 0.25082
fMRI	**0.1184** / 0.2149	**0.9988** / 1.3454	**0.0947** / 0.1763	0.09944 / **0.09928**

E.1 TIMING BOUNDARY AND BASELINE PROVENANCE

The controlled H200 runs use the sample counts and batches in Table 11. We synchronize CUDA before and after each batch and exclude three warm-up forwards. The timer includes timestep construction, every denoiser call, update arithmetic, in-loop random draws, and the spectral controller where applicable. It excludes model construction, checkpoint and dataset loading, initial-noise generation, transfers outside the sampling loop, serialization, and all metric evaluation. This measures sampling latency after initialization, not end-to-end application latency.

Table 15 gives the exact times plotted in Figure 8. Table 14 summarizes our comparison with uniform DDIM, while the preceding complete tables retain every evaluated method and metric.

Table 15: Sampling time in seconds for the full sample counts in Appendix D. Same H200 GPU model and timing boundary, with FFT gating included. Mean ± sample standard deviation over three generation seeds. Underline marks the second-lowest distinct displayed mean within the same comparison groups as bold, including ties. Neither mark indicates significance.

Data	StrideDiffusion	Uniform DDIM	DPM++(2M)	UniPC-bh2
Sines	3.658 ±0.136	**2.984 ±0.125**	4.929 ±0.369	4.934 ±0.207
Stocks	1.409 ±0.057	**1.127 ±0.047**	1.758 ±0.083	1.883 ±0.146
ETTh	2.885 ±0.032	**2.336 ±0.090**	3.517 ±0.160	3.760 ±0.258
MuJoCo	6.881 ±0.304	**5.635 ±0.282**	8.465 ±0.384	8.845 ±0.420
Energy	34.818 ±1.352	**30.421 ±1.076**	41.570 ±1.625	43.212 ±1.734
fMRI	25.177 ±1.091	**22.901 ±0.878**	28.976 ±0.873	29.916 ±1.282

For completeness we timed local implementations of DiffWave and CSDI-style DiffTime with full ancestral sampling. These models use initialized weights and implement the published execution graphs for timing only. Their times in Table 16 are not paired with the published quality scores in Appendix K, so they cannot establish a quality–speed improvement over those trained baselines. Appendix J separately reports newly trained adaptations, with quality and timing from the same checkpoints. The cited benchmark does not report DiffWave or DiffTime sampling times. Their omission is not evidence of a timeout or prohibitive cost.

Table 16: Timing-only architecture reproductions on H200. Mean ± sample standard deviation over three seeds, in seconds. Full ancestral sampling uses one call per diffusion level.

Data	Steps / NFE	DiffWave	DiffTime
Sines	500	5.00 ± .07	156.44 ± .06
Stocks	500	2.44 ± .02	61.01 ± .23
ETTh	500	10.21 ± .07	308.57 ± .15
MuJoCo	1000	20.49 ± .12	528.62 ± .00
Energy	1000	130.97 ± .71	1732.68 ± 1.76
fMRI	1000	105.95 ± 1.12	1418.49 ± .37

E.2 SOLVER-SPECIFIC GRID CONTROL

Using the same spectral gate with an official DPM-Solver++ update does not reproduce the complete method's gains. The endpoint-matched comparison in Table 17 is approximately tied on MuJoCo and favors uniform DPM++ elsewhere. Table 18 includes the controller overhead. Together with the scheduler-only DDIM rows, this is evidence that the useful grid depends on the update rule and its trajectory.

Table 17: Endpoint-matched DPM-Solver++(2M) grid control. Entries are mean NFE / Context-FID (mean ± sample standard deviation over three generation seeds). Blue identifies our spectral grid used with the DPM++ update, not the complete StrideDiffusion sampler. Bold marks the lower displayed C-FID mean in each row and underline the other value when distinct. Ties share bold. These marks do not indicate significance. NFE is not ranked.

Data	**Spectral DPM++**	uniform DPM++
Sines	35.0/.0191±.0011	35.0/**.0176±.0014**
Stocks	34.2/.3443±.0400	34.3/**.1788±.0185**
ETTh	13.0/.2440±.0342	13.0/**.1833±.0340**
MuJoCo	59.5/**.0434±.0034**	59.7/.0435±.0037
Energy	59.0/.0753±.0065	59.0/**.0727±.0091**
fMRI	68.2/.1309±.0089	68.0/**.1221±.0081**

Table 18: Endpoint-matched DPM++ sampling time in seconds (mean over three generation seeds). The spectral timing includes FFT-grid management. Blue identifies our spectral grid used with the DPM++ update, not the complete StrideDiffusion sampler. Bold marks the lower displayed mean in each row and underline the other value when distinct. Ties share bold. These marks do not indicate significance. Lower sampling time is better.

Data	**Spectral DPM++**	uniform DPM++
Sines	5.609	**4.696**
Stocks	2.162	**1.819**
ETTh	3.933	**3.325**
MuJoCo	9.869	**8.443**
Energy	47.586	**41.616**
fMRI	31.283	**27.872**

E.3 Current-sample and predicted-clean gate diagnostic

We compare consecutive-state gating with a stricter diagnostic that checks relative band energy in both the current sample and the clean prediction. Both experiments use DDIM throughout, at generation seed 12345. An energy-check flag marks a coarse decision where the original gate permits a jump but this separate check finds an active non-low band. All 1,319 flags originate in the current sample under this relative criterion. None originate in the predicted-clean estimate. The stricter gate removes these flags at the cost of 3.2 to 11.0 times more NFE, with C-FID improving on three datasets and worsening on three. This diagnostic checks relative energy, so removing its flags does not certify the norm condition in Corollary 1.

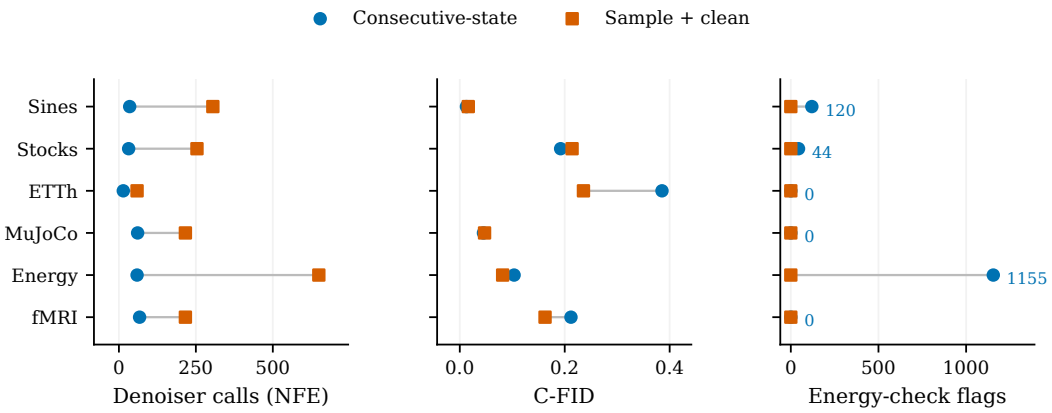

Figure 11: Consecutive-state gating versus the sample-and-clean diagnostic on all six datasets, using DDIM and generation seed 12345. Connected points show NFE, C-FID and energy-check flags. Lower is better in each panel. All strict-check flag counts are zero. This is one run per setting, with no uncertainty estimate. Removing relative-energy flags does not certify the theorem's norm condition. Exact values are in `plot_data/gate_tradeoff.csv`.

E.4 Absolute spectral power and distributional errors

For each sequence and channel we subtract the temporal mean, then take an orthonormal rFFT. Let $w_f = 2$ for interior positive-frequency bins and $w_f = 1$ for DC and Nyquist. At length 24, the per-sequence high-frequency power is

$$q_{\mathrm{HF}}(x) = \frac{1}{Ld} \sum_{j=1}^{d} \sum_{f=6}^{12} w_f |X[f, j]|^2. \tag{15}$$

Total AC power uses all positive bins. Centered RMS is the square root of total AC power. We compare these scalar distributions without assigning arbitrary real/generated sample pairs. Mean error is $|\mathbb{E}q_g - \mathbb{E}q_r|/\mathbb{E}q_r$. W1/ref is the one-dimensional Wasserstein distance divided by $\mathbb{E}q_r$. Q90 error is the absolute gap between 90th percentiles divided by the reference percentile. These denominators use only reference windows, not generated-output energy. Tables 19 and 20 report the complete results.

Table 19: Absolute power-distribution errors against uniform DDIM. Ours and control are shown on separate rows, with mean ± sample standard deviation across three generation seeds. All errors are normalized by reference statistics only. Underline marks the second-lowest distinct displayed mean within the same comparison groups as bold, including ties. Neither mark indicates significance.

Data	Method	HF mean error	HF W1/ref	HF Q90 error	AC W1/ref	RMS W1/ref
Sines	**Ours**	**0.016 ±0.002**	**0.021 ±0.001**	**0.008 ±0.001**	**0.037 ±0.001**	**0.020 ±0.001**
Sines	Control	0.033 ±0.001	0.033 ±0.001	0.014 ±0.002	0.055 ±0.001	0.029 ±0.001
Stocks	**Ours**	**0.055 ±0.027**	**0.263 ±0.034**	**0.104 ±0.086**	**0.146 ±0.038**	**0.059 ±0.010**
Stocks	Control	0.620 ±0.029	0.620 ±0.029	0.639 ±0.031	0.536 ±0.027	0.356 ±0.013
ETTh	**Ours**	**0.247 ±0.007**	**0.273 ±0.006**	**0.305 ±0.009**	**0.072 ±0.002**	**0.034 ±0.001**
ETTh	Control	0.662 ±0.002	0.662 ±0.002	0.601 ±0.005	0.366 ±0.003	0.231 ±0.002
MuJoCo	**Ours**	**0.008 ±0.008**	**0.012 ±0.007**	**0.008 ±0.006**	**0.017 ±0.005**	**0.009 ±0.003**
MuJoCo	Control	0.023 ±0.011	0.025 ±0.010	0.011 ±0.007	0.020 ±0.010	0.011 ±0.005
Energy	**Ours**	**0.053 ±0.002**	**0.053 ±0.002**	**0.078 ±0.003**	0.043 ±0.001	0.022 ±0.000
Energy	Control	0.072 ±0.002	0.072 ±0.002	0.093 ±0.003	**0.035 ±0.000**	**0.017 ±0.000**
fMRI	**Ours**	**0.038 ±0.001**	**0.038 ±0.001**	**0.023 ±0.000**	**0.005 ±0.001**	**0.003 ±0.000**
fMRI	Control	0.156 ±0.001	0.156 ±0.001	0.171 ±0.001	0.142 ±0.001	0.073 ±0.000

Table 20: Absolute power-distribution errors against uniform grid with the same conditional history rule. Ours and control are shown on separate rows, with mean ± sample standard deviation across three generation seeds. All errors are normalized by reference statistics only. Underline marks the second-lowest distinct displayed mean within the same comparison groups as bold, including ties. Neither mark indicates significance.

Data	Method	HF mean error	HF W1/ref	HF Q90 error	AC W1/ref	RMS W1/ref
Sines	**Ours**	0.016 ±0.002	0.021 ±0.001	0.008 ±0.001	**0.037 ±0.001**	**0.020 ±0.001**
Sines	Control	**0.014 ±0.002**	**0.016 ±0.001**	**0.001 ±0.001**	0.044 ±0.001	0.023 ±0.001
Stocks	**Ours**	**0.055 ±0.027**	**0.263 ±0.034**	**0.104 ±0.086**	**0.146 ±0.038**	**0.059 ±0.010**
Stocks	Control	0.380 ±0.030	0.382 ±0.027	0.260 ±0.031	0.298 ±0.021	0.206 ±0.011
ETTh	**Ours**	**0.247 ±0.007**	**0.273 ±0.006**	**0.305 ±0.009**	**0.072 ±0.002**	**0.034 ±0.001**
ETTh	Control	0.556 ±0.004	0.556 ±0.004	0.472 ±0.007	0.181 ±0.003	0.123 ±0.002
MuJoCo	**Ours**	**0.008 ±0.008**	**0.012 ±0.007**	**0.008 ±0.006**	0.017 ±0.005	0.009 ±0.003
MuJoCo	Control	**0.008 ±0.007**	0.013 ±0.004	0.011 ±0.010	**0.016 ±0.003**	**0.008 ±0.001**
Energy	**Ours**	**0.053 ±0.002**	**0.053 ±0.002**	**0.078 ±0.003**	0.043 ±0.001	0.022 ±0.000
Energy	Control	0.102 ±0.002	0.102 ±0.002	0.123 ±0.003	**0.015 ±0.000**	**0.007 ±0.000**
fMRI	**Ours**	0.038 ±0.001	0.038 ±0.001	**0.023 ±0.000**	**0.005 ±0.001**	**0.003 ±0.000**
fMRI	Control	**0.021 ±0.002**	**0.022 ±0.001**	0.040 ±0.002	0.046 ±0.001	0.023 ±0.000

Figure 5 summarizes the two preceding tables without dropping any metric or dataset. The source package also includes these means, SDs and ratios in `plot_data/absolute_power_matrix.csv`.

We also report relative spectral and temporal statistics and MSE between real and generated mean and standard-deviation profiles in Figure 12. Profile MSE compares distributions in normalized $[0, 1]$ space. Unconditional samples have no natural one-to-one target, so a samplewise MSE would depend on an arbitrary pairing. Ordinary paired MSE remains appropriate for conditional imputation and forecasting.

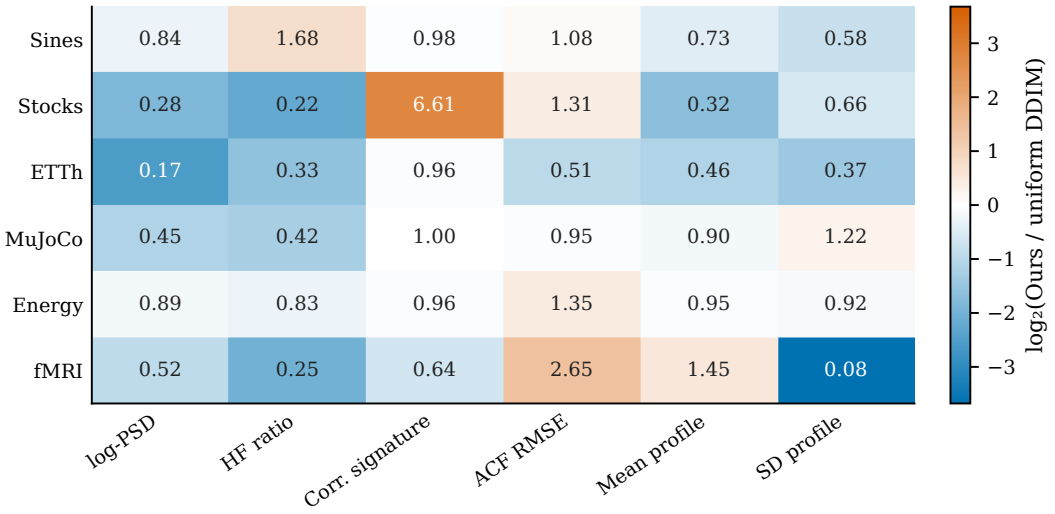

Figure 12: Spectral, temporal and profile-error ratios for all six datasets and six diagnostics. Each number is Ours/matched-NFE uniform DDIM from the three-generation-seed means. Mean and SD profile columns are profile MSE, not uncertainty estimates. Blue is lower error and orange higher error. Color uses a symmetric $\log_2$ ratio scale with no clipping. These are descriptive ratios of the archived displayed means, not significance tests. The underlying absolute values are in `plot_data/fidelity_matrix.csv`.

E.5 Relation to caching, distillation, and quantization

SeaCache uses filtered input features to decide when a scheduled transition can reuse a cached output (Chung et al., 2026). Table 21 contrasts this with changing the visited time grid. We also implement a SeaCache-style port for Diffusion-TS and compare it on the same frozen checkpoints and reference data. This port caches the decoder-output residual and auxiliary mean, trend, and seasonal components. A cache hit skips the encoder and decoder blocks but still runs input/time conditioning, modulation, and output heads. We therefore report total model-entry calls, full evaluations, partial evaluations, and measured time separately.

Table 21: Mechanism comparison. Neither method requires retraining.

Aspect	SeaCache	**StrideDiffusion**
Decision	Reuse an output at a scheduled time	Choose the next time to visit
Signal	Distance of SEA-filtered input features	Sample band power, power change, phase change
Time grid	Preset sampler grid	State-dependent grid
Saved work	Skip full blocks, retain partial work	Remove skipped transitions
Extra work	Filter and cached-output storage	FFT gate and history storage
Failure mode	Stale output	Missed changes across coarse jumps

Tables 22 and 23 include all planned cache settings: 64 or the full diffusion schedule, each at thresholds $\delta \in \{0, 0.1, 0.3, 0.6\}$. The controls are uniform DDIM at 16, 32, and 64 calls and the complete StrideDiffusion sampler. Each run generates 2,048 samples with a shared reference-only TS2Vec encoder. We report three generation/evaluator seeds, 31415, 27182, and 16180. Sampling times include the gate, update arithmetic, and trace construction on an exclusive H200 GPU. Each seed's time is the median of three repeated samplings with the same initial noise.

The port and StrideDiffusion offer different quality and time tradeoffs. For example, on Sines, the 64-step cache at $\delta = 0.6$ is faster than StrideDiffusion (2.11 versus 2.81 seconds), but has higher C-FID (0.03285 versus 0.02649). On ETTh, StrideDiffusion is both faster and lower in C-FID than that setting. The full tables retain settings that favor either method. These measurements concern our port, not an unchanged official SeaCache implementation. Combining caching and stride selection would require checking both stale-output error and missed changes across coarse jumps.

Table 22: SeaCache-style port and sampling controls, part 1. All planned operating points are shown, without matching call budgets across rows. Total calls include both full backbone evaluations and cached partial evaluations. A partial evaluation still runs input/time conditioning and output heads, so it is not free. Counts are batch means averaged over three generation seeds. C-FID and H200 sampling time (seconds) are mean ± sample standard deviation over those seeds. Timing includes cache decisions and FFT gating where used. This is our port to Diffusion-TS, not the unchanged official SeaCache implementation. Blue shading identifies the spectral method or grid, not an overall winner across unequal budgets.

Data	Method	Total	Full	Partial	C-FID	Seconds
Sines	**StrideDiffusion**	35.0	35.0	0.0	0.0265 ±0.0094	2.81 ±0.04
Sines	Uniform DDIM 16	16.0	16.0	0.0	0.0316 ±0.0070	1.06 ±0.02
Sines	Uniform DDIM 32	32.0	32.0	0.0	0.0299 ±0.0092	2.12 ±0.04
Sines	Uniform DDIM 64	64.0	64.0	0.0	0.0303 ±0.0103	4.19 ±0.06
Sines	Cache 64, $\delta = 0.0$	64.0	64.0	0.0	0.0303 ±0.0103	4.70 ±0.06
Sines	Cache 64, $\delta = 0.1$	64.0	64.0	0.0	0.0303 ±0.0103	4.69 ±0.06
Sines	Cache 64, $\delta = 0.3$	64.0	34.0	30.0	0.0298 ±0.0089	2.93 ±0.05
Sines	Cache 64, $\delta = 0.6$	64.0	20.0	44.0	0.0328 ±0.0071	2.11 ±0.06
Sines	Cache 500, $\delta = 0.0$	500.0	500.0	0.0	0.0322 ±0.0114	36.70 ±0.60
Sines	Cache 500, $\delta = 0.1$	500.0	257.7	242.3	0.0319 ±0.0112	22.43 ±0.47
Sines	Cache 500, $\delta = 0.3$	500.0	105.0	395.0	0.0311 ±0.0106	13.35 ±0.20
Sines	Cache 500, $\delta = 0.6$	500.0	56.5	443.5	0.0303 ±0.0097	10.50 ±0.20
Stocks	**StrideDiffusion**	33.2	33.2	0.0	0.1079 ±0.0231	2.89 ±0.08
Stocks	Uniform DDIM 16	16.0	16.0	0.0	0.1980 ±0.0301	1.15 ±0.01
Stocks	Uniform DDIM 32	32.0	32.0	0.0	0.1975 ±0.0274	2.30 ±0.01
Stocks	Uniform DDIM 64	64.0	64.0	0.0	0.1933 ±0.0260	4.54 ±0.05
Stocks	Cache 64, $\delta = 0.0$	64.0	64.0	0.0	0.1933 ±0.0260	5.10 ±0.01
Stocks	Cache 64, $\delta = 0.1$	64.0	64.0	0.0	0.1933 ±0.0260	5.04 ±0.04
Stocks	Cache 64, $\delta = 0.3$	64.0	37.0	27.0	0.2023 ±0.0264	3.32 ±0.03
Stocks	Cache 64, $\delta = 0.6$	64.0	24.9	39.1	0.2207 ±0.0281	2.53 ±0.03
Stocks	Cache 500, $\delta = 0.0$	500.0	500.0	0.0	0.1855 ±0.0251	39.46 ±0.02
Stocks	Cache 500, $\delta = 0.1$	500.0	299.6	200.4	0.1871 ±0.0253	26.57 ±0.06
Stocks	Cache 500, $\delta = 0.3$	500.0	136.9	363.1	0.1898 ±0.0256	16.05 ±0.03
Stocks	Cache 500, $\delta = 0.6$	500.0	74.9	425.1	0.1944 ±0.0260	12.02 ±0.12
ETTh	**StrideDiffusion**	14.1	14.1	0.0	0.1729 ±0.0246	1.34 ±0.04
ETTh	Uniform DDIM 16	16.0	16.0	0.0	0.3680 ±0.0229	1.25 ±0.02
ETTh	Uniform DDIM 32	32.0	32.0	0.0	0.2865 ±0.0164	2.50 ±0.05
ETTh	Uniform DDIM 64	64.0	64.0	0.0	0.2584 ±0.0143	5.01 ±0.08
ETTh	Cache 64, $\delta = 0.0$	64.0	64.0	0.0	0.2584 ±0.0143	5.51 ±0.08
ETTh	Cache 64, $\delta = 0.1$	64.0	64.0	0.0	0.2584 ±0.0143	5.50 ±0.08
ETTh	Cache 64, $\delta = 0.3$	64.0	33.0	31.0	0.2985 ±0.0166	3.29 ±0.07
ETTh	Cache 64, $\delta = 0.6$	64.0	19.5	44.5	0.3460 ±0.0195	2.31 ±0.03
ETTh	Cache 500, $\delta = 0.0$	500.0	500.0	0.0	0.2333 ±0.0134	43.04 ±0.58
ETTh	Cache 500, $\delta = 0.1$	500.0	249.9	250.1	0.2368 ±0.0136	25.10 ±0.33
ETTh	Cache 500, $\delta = 0.3$	500.0	98.1	401.9	0.2477 ±0.0144	14.28 ±0.35
ETTh	Cache 500, $\delta = 0.6$	500.0	52.9	447.1	0.2654 ±0.0144	10.93 ±0.18

Table 23: SeaCache-style port and sampling controls, part 2. All planned operating points are shown, without matching call budgets across rows. Total calls include both full backbone evaluations and cached partial evaluations. A partial evaluation still runs input/time conditioning and output heads, so it is not free. Counts are batch means averaged over three generation seeds. C-FID and H200 sampling time (seconds) are mean ± sample standard deviation over those seeds. Timing includes cache decisions and FFT gating where used. This is our port to Diffusion-TS, not the unchanged official SeaCache implementation. Blue shading identifies the spectral method or grid, not an overall winner across unequal budgets.

Data	Method	Total	Full	Partial	C-FID	Seconds
MuJoCo	**StrideDiffusion**	60.1	60.1	0.0	0.0675 ±0.0258	5.59 ±0.06
MuJoCo	Uniform DDIM 16	16.0	16.0	0.0	0.0674 ±0.0253	1.26 ±0.04
MuJoCo	Uniform DDIM 32	32.0	32.0	0.0	0.0633 ±0.0252	2.48 ±0.06
MuJoCo	Uniform DDIM 64	64.0	64.0	0.0	0.0642 ±0.0251	4.94 ±0.08
MuJoCo	Cache 64, $\delta = 0.0$	64.0	64.0	0.0	0.0642 ±0.0251	5.48 ±0.12
MuJoCo	Cache 64, $\delta = 0.1$	64.0	64.0	0.0	0.0642 ±0.0251	5.45 ±0.08
MuJoCo	Cache 64, $\delta = 0.3$	64.0	28.0	36.0	0.0632 ±0.0246	2.92 ±0.06
MuJoCo	Cache 64, $\delta = 0.6$	64.0	17.0	47.0	0.0681 ±0.0243	2.10 ±0.06
MuJoCo	Cache 1000, $\delta = 0.0$	1000.0	1000.0	0.0	0.0675 ±0.0253	85.61 ±1.84
MuJoCo	Cache 1000, $\delta = 0.1$	1000.0	367.0	633.0	0.0671 ±0.0253	40.47 ±1.13
MuJoCo	Cache 1000, $\delta = 0.3$	1000.0	142.0	858.0	0.0661 ±0.0253	24.26 ±0.70
MuJoCo	Cache 1000, $\delta = 0.6$	1000.0	75.0	925.0	0.0651 ±0.0253	19.35 ±0.56
Energy	**StrideDiffusion**	58.8	58.8	0.0	0.1051 ±0.0132	7.33 ±0.14
Energy	Uniform DDIM 16	16.0	16.0	0.0	0.1572 ±0.0043	1.77 ±0.04
Energy	Uniform DDIM 32	32.0	32.0	0.0	0.1186 ±0.0131	3.52 ±0.07
Energy	Uniform DDIM 64	64.0	64.0	0.0	0.1062 ±0.0181	7.05 ±0.09
Energy	Cache 64, $\delta = 0.0$	64.0	64.0	0.0	0.1062 ±0.0181	7.54 ±0.12
Energy	Cache 64, $\delta = 0.1$	64.0	64.0	0.0	0.1062 ±0.0181	7.55 ±0.12
Energy	Cache 64, $\delta = 0.3$	64.0	30.0	34.0	0.1236 ±0.0162	4.01 ±0.09
Energy	Cache 64, $\delta = 0.6$	64.0	18.8	45.2	0.1464 ±0.0107	2.85 ±0.06
Energy	Cache 1000, $\delta = 0.0$	1000.0	1000.0	0.0	0.0973 ±0.0215	117.60 ±2.23
Energy	Cache 1000, $\delta = 0.1$	1000.0	411.7	588.3	0.0981 ±0.0218	56.79 ±1.02
Energy	Cache 1000, $\delta = 0.3$	1000.0	159.0	841.0	0.0993 ±0.0212	30.49 ±0.63
Energy	Cache 1000, $\delta = 0.6$	1000.0	84.0	916.0	0.1028 ±0.0207	22.60 ±0.44
fMRI	**StrideDiffusion**	67.0	67.0	0.0	0.1467 ±0.0029	10.29 ±0.36
fMRI	Uniform DDIM 16	16.0	16.0	0.0	0.5680 ±0.0139	2.21 ±0.03
fMRI	Uniform DDIM 32	32.0	32.0	0.0	0.3135 ±0.0095	4.43 ±0.05
fMRI	Uniform DDIM 64	64.0	64.0	0.0	0.2262 ±0.0078	8.85 ±0.11
fMRI	Cache 64, $\delta = 0.0$	64.0	64.0	0.0	0.2262 ±0.0078	9.29 ±0.09
fMRI	Cache 64, $\delta = 0.1$	64.0	52.0	12.0	0.2836 ±0.0103	7.75 ±0.11
fMRI	Cache 64, $\delta = 0.3$	64.0	24.0	40.0	0.6343 ±0.0204	4.11 ±0.07
fMRI	Cache 64, $\delta = 0.6$	64.0	14.0	50.0	1.2935 ±0.0452	2.76 ±0.03
fMRI	Cache 1000, $\delta = 0.0$	1000.0	1000.0	0.0	0.1678 ±0.0055	145.69 ±2.32
fMRI	Cache 1000, $\delta = 0.1$	1000.0	271.0	729.0	0.1813 ±0.0063	50.26 ±0.68
fMRI	Cache 1000, $\delta = 0.3$	1000.0	102.0	898.0	0.2218 ±0.0082	28.07 ±0.45
fMRI	Cache 1000, $\delta = 0.6$	1000.0	54.0	946.0	0.2968 ±0.0110	21.67 ±0.25

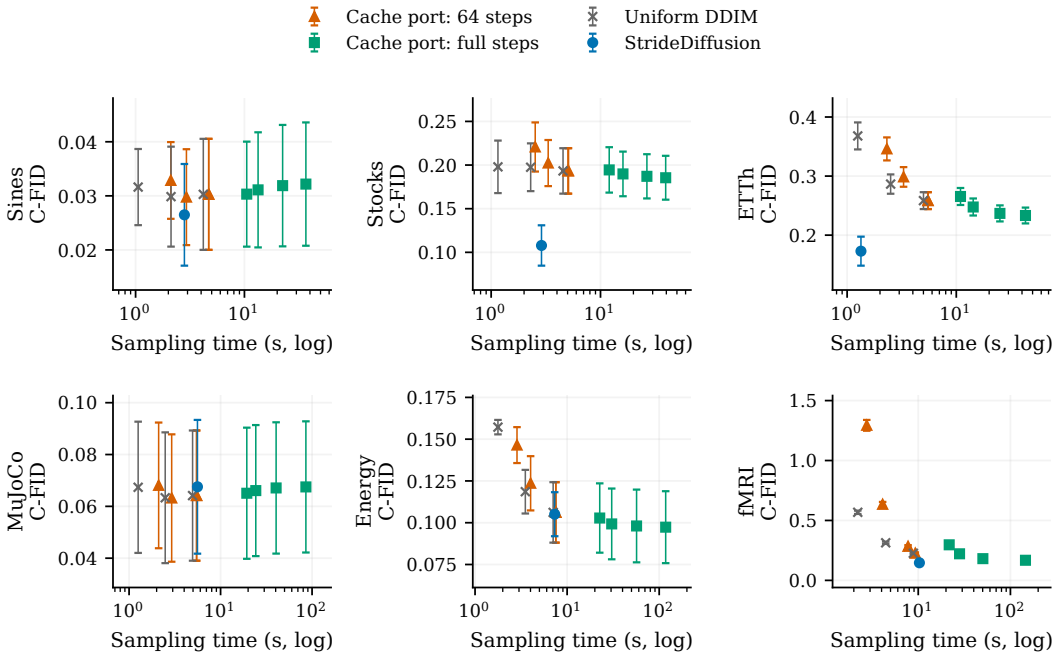

Figure 13: Quality–time comparison including every evaluated cache-port setting: 64/full schedules with four thresholds each, three uniform-DDIM budgets, and StrideDiffusion. All 12 settings per dataset are shown, including overlapping points. Points are three-generation/evaluator-seed means, with C-FID sample-SD bars. Time is on a logarithmic axis, and panel ranges differ. Lower and further left is better. These are one-checkpoint comparisons on an exclusive H200, not the official unchanged SeaCache implementation. Exact times, time SDs, full/partial calls and all metrics remain in Tables 22–23 and `plot_data/cache_frontier.csv`.

Distillation changes model weights through additional training. Quantization changes numerical precision and needs separate calibration and hardware checks. Our controlled empirical claim is restricted to the evaluated training-free solvers on unchanged checkpoints. The results do not show superiority over all acceleration families.

F Controls with matched update positions

These experiments test whether changing the time grid helps when the update positions are fixed. They use a separate evaluation protocol from the main result table: generation/evaluator seeds 31415, 27182, and 16180, 2,048 Diffusion-TS samples or 2,000 ImagenFew samples, and a shared TS2Vec encoder fitted on up to 2,000 reference sequences within each run. Diffusion-TS uses a fixed subset of the historical training-reference pool, not unseen held-out data. ImagenFew uses a fixed synthetic reference drawn with seed 54321. The same initial noise is used for every method in a run. The standard deviations describe generation and evaluator variation, not independent model training. Separately fitted encoders make absolute C-FID values across tables unsuitable for direct ranking.

Synchronized sampling time includes the gate, update arithmetic, and trace construction. It excludes model/data loading, saving, and metric evaluation. Each generation seed uses the median of three repeated samplings with the same noise. Spectral timings include FFT gating. A single job occupies each H200 GPU during these measurements.

F.1 Fixed-budget Diffusion-TS controls

For a budget of N calls, the uniform control visits rounded, equally spaced diffusion indices from $T-1$ to 0, followed by the final clean prediction. The spectral control instead chooses a stride from the remaining distance divided by the remaining number of transitions. It multiplies this nominal stride by 2 when no band is active, by 1 for DC-only activity, and by $1/2$ otherwise. It reserves one

integer index for every remaining call. The penultimate call moves to 0, so both methods use exactly N calls.

This budget-constrained rule is a diagnostic variant. It differs from the original sampler's fixed candidate strides and fine-step tail. Its last nonterminal jump can be large when earlier decisions defer progress. Figures 14 and 15 report all three budgets on all six datasets. The DDIM comparison uses DDIM at every position. The history comparison uses the same approximate history formula at every interior position, including unit strides, with DDIM at initialization and at the terminal prediction. Both grids therefore use the same branch positions.

The spectral variant has lower mean C-FID in 6 of 18 DDIM comparisons and 8 of 18 history comparisons. Its runtime is higher in every matched-budget comparison. These outcomes show why a grid-only claim needs both a specified controller and a specified update rule. They do not rank the complete original sampler, which uses a different schedule and branch policy.

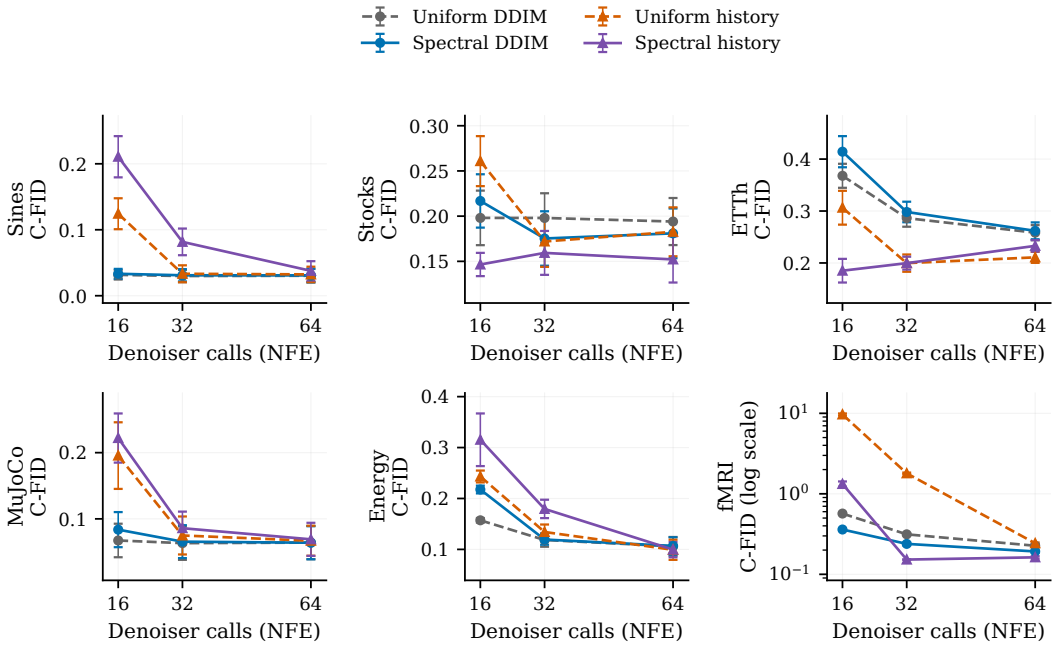

Figure 14: Fixed-budget grid diagnostic on all six datasets. Solid curves use the spectral grid and dashed curves the uniform grid, with DDIM or the same approximate-history update. Points and bars are C-FID mean ± sample SD over three generation/evaluator seeds and one frozen checkpoint per dataset. Lower is better. Both grids have identical NFE and branch positions. The fMRI vertical axis is logarithmic to show the full range. Other panels use linear axes, with dataset-specific ranges. These are diagnostic controllers, not the complete original sampler. The source package includes all underlying values in `plot_data/appendix_budget_quality.csv`.

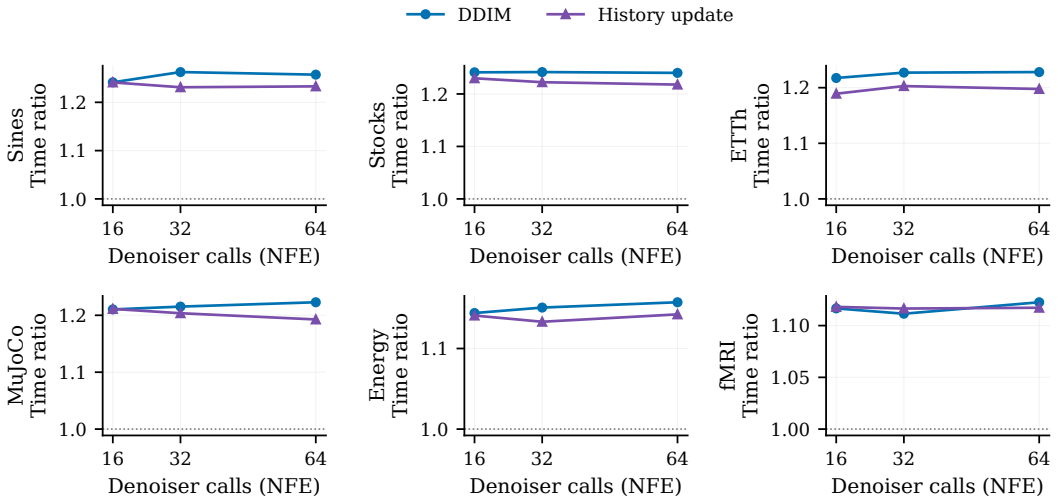

Figure 15: Sampling overhead for the same 36 grid pairs as Figure 14. Each value is mean spectral time divided by mean uniform time at identical NFE. Values above the dotted line at one are slower. The timer includes FFT gating on an exclusive H200. These are ratios of three-seed mean times, not confidence intervals. Panel ranges differ. Exact mean times and ratios are included in `plot_data/appendix_budget_cost.csv`.

F.2 Matched correction positions on ImagenFew

The ImagenFew comparisons in Table 24 hold the Heun-correction positions and history scale fixed as well as the call budget. We use the same frozen, sine-fine-tuned checkpoint at each length for every method. The correction positions are spaced evenly among eligible transitions and checked during sampling. Both grids use either 16 transitions with 2 or 6 corrections, or 20 transitions with 6 corrections. Each setting is evaluated with history scale 0.75 and 1. The actual history coefficient also depends on adjacent step lengths, so it can differ between grids.

Spectral PC2 has lower mean C-FID in 8 of the 12 settings, while taking more time in every pair. Thus a grid effect remains in some settings after matching the correction positions, but it is not uniform across budgets or lengths. Both lengths use synthetic sine data and one fine-tuned checkpoint each. This experiment tests the EDM port rather than the exact Diffusion-TS update.

Table 24: Strict ImagenFew grid controls on synthetic sine sequences. T/C lists transitions and Heun corrections. Uniform and spectral PC2 share the correction positions, history scale h, exact NFE and initial noise. C-FID is mean $\pm$ sample standard deviation over three generation/metric seeds using one fine-tuned checkpoint per length. Time ratio is spectral/uniform. The 12 settings are all reported. These are grid controls for the EDM port, not independent training replicates or transfer of the Diffusion-TS update formula. Blue identifies our spectral grid. Bold marks the lower displayed mean in each row and underline the other value when distinct. Ties share bold. These marks do not indicate significance.

Length	T/C	h	NFE	Uniform C-FID	**Spectral C-FID**	Time ratio
24	16/2	0.75	18	2.0494 $\pm$0.4023	**1.9138 $\pm$0.3382**	1.09
24	16/2	1	18	2.1055 $\pm$0.4236	**1.9513 $\pm$0.3519**	1.10
24	16/6	0.75	22	2.4561 $\pm$0.5764	**2.3314 $\pm$0.5226**	1.08
24	16/6	1	22	2.4826 $\pm$0.5860	**2.3601 $\pm$0.5335**	1.07
24	20/6	0.75	26	**2.1727 $\pm$0.4669**	2.3531 $\pm$0.5297	1.08
24	20/6	1	26	**2.2039 $\pm$0.4782**	2.3870 $\pm$0.5421	1.08
64	16/2	0.75	18	**0.9387 $\pm$0.0256**	0.9505 $\pm$0.0263	1.10
64	16/2	1	18	**0.9328 $\pm$0.0249**	0.9437 $\pm$0.0255	1.10
64	16/6	0.75	22	0.9312 $\pm$0.0235	**0.9291 $\pm$0.0257**	1.08
64	16/6	1	22	0.9301 $\pm$0.0230	**0.9260 $\pm$0.0247**	1.08
64	20/6	0.75	26	0.9528 $\pm$0.0239	**0.9349 $\pm$0.0244**	1.09
64	20/6	1	26	0.9512 $\pm$0.0232	**0.9332 $\pm$0.0238**	1.08

G NATIVE-GRID REPLAY AND UPDATE-RULE CONTROLS

Table 25 gives the exact values underlying Figure 10.

Table 25: Grid replay with identical NFE and the same history/DDIM choice at each update index, but different diffusion times. Both grids use the same approximate-history rule and initial noise. C-FID is mean $\pm$ sample SD over three generation/metric seeds (31415, 27182, 16180) and one frozen checkpoint per dataset. Each run generates 2,048 windows. C-FID uses a fixed 2,000-window real/fake subset shared across methods. Its encoder is fitted on the real subset and shared within each run, separately from Table 10. NFE is averaged over batches and seeds. Δ is the percentage change from uniform to spectral, so negative is better. Blue identifies our grid, and bold marks the lower displayed mean, not significance. Underline marks the second-lowest distinct displayed mean within the same comparison groups as bold, including ties. Neither mark indicates significance.

Data	NFE	Uniform C-FID	**Spectral C-FID**	Δ
Sines	35.00	0.0316 $\pm$0.0107	**0.0265 $\pm$0.0094**	-16.3%
Stocks	33.25	0.1179 $\pm$0.0247	**0.1090 $\pm$0.0241**	-7.6%
ETTh	14.15	0.3885 $\pm$0.0557	**0.1729 $\pm$0.0245**	-55.5%
MuJoCo	60.06	**0.0669 $\pm$0.0258**	0.0675 $\pm$0.0258	+0.9%
Energy	58.77	**0.1020 $\pm$0.0212**	0.1050 $\pm$0.0133	+3.0%
fMRI	67.02	0.2124 $\pm$0.0070	**0.1467 $\pm$0.0034**	-30.9%

Conditional-rule control. Table 26 uses the original comparison's generation seeds 12345, 2026 and 2027. It keeps the conditional update formula fixed, but changing the grid can change where history updates and DDIM fallbacks occur. It is a separate protocol from the replay study below, which matches exact NFE and the branch choice at each update index. The paired encoder is shared within each run and fitted separately from the main solver comparison and replay study.

Replay protocol. The budget-constrained controller in Appendix F changes the original stride policy. Here we instead record the original sampler's per-batch time grid and replay it with controlled alternatives. The complete factorial contains six datasets, three generation/metric seeds (31415, 27182, 16180), and 19 methods per seed, for 342 method results. Each dataset uses one frozen

Table 26: Grid comparison under the same conditional history-update rule. Three generation seeds. The grid can change which branch of that rule is used. The paired C-FID evaluator is shared within this experiment. Blue identifies our spectral grid. Bold marks the lower displayed mean in each row and underline the other value when distinct. Ties share bold. These marks do not indicate significance.

Data	NFE (ours / uniform)	**Spectral grid C-FID**	Uniform grid C-FID
Sines	35.0 / 35.0	**0.0120 ±0.0014**	0.0199 ±0.0004
Stocks	32.2 / 32.0	**0.1150 ±0.0065**	0.2583 ±0.0203
ETTh	14.0 / 14.0	**0.1501 ±0.0263**	0.1780 ±0.0243
MuJoCo	59.3 / 59.0	**0.0442 ±0.0028**	0.0459 ±0.0030
Energy	59.0 / 59.0	0.0978 ±0.0099	**0.0938 ±0.0094**
fMRI	67.6 / 67.7	**0.1184 ±0.0079**	0.1301 ±0.0083

Diffusion-TS checkpoint and 2,048 historical benchmark training-reference windows, with 2,048 generated windows per method. C-FID uses a fixed subset of 2,000 real and generated windows shared across methods within each run. The TS2Vec encoder is fitted on that real subset and shared within the run. These are generation-seed comparisons, not independent retrainings or unseen-test results.

Grid controls. For each original grid, uniform replays exactly the same number of model calls. Uniform plus same tail also preserves the original fine-step suffix. Shuffled plus same tail permutes the remaining strides within update-branch strata, with seed 73411, while preserving the initial and terminal treatment. All three controls preserve the recorded history/DDIM choices for approximate-history replay. For example, if update k uses history on the original grid, it also uses history on the control grid, although its diffusion times can differ. An independent donor grid is recorded using a different initial-noise seed (85973). Its own uniform control matches its budget. Native and donor grids can have different NFE, so their direct comparison is descriptive rather than budget-matched.

Update controls. Every grid is replayed using DDIM, the original approximate-history update, and a DPM-Solver++ 2M formula using the actual adjacent log-SNR intervals. DDIM replay uses DDIM at every transition. The 2M replay uses DDIM for the first and terminal transitions and 2M for all others, rather than applying the recorded history/DDIM choices. It implements the variable-step formula within our discrete-time endpoint wrapper, not the complete official sampler used in Table 10. All grids were acquired using the original approximate-history sampler. DDIM and 2M therefore test those fixed grids, not grids selected online from their own evolving trajectories.

Results and interpretation. Against exact-NFE uniform replay, the spectrum-derived grid improves mean C-FID on 4/6 datasets with the approximate-history rule. ETTh, fMRI, Sines and Stocks improve by 55.5%, 30.9%, 16.3% and 7.6%, respectively, with all three generation seeds improving. Energy and MuJoCo worsen by 3.0% and 0.9%. Keeping the same fine tail gives improvements on 5/6 datasets, although the fMRI and MuJoCo differences are small. DDIM improves on 4/6 datasets but regresses on Energy and ETTh. Actual-span 2M improves only on Stocks. Thus the favorable results depend on the update rule and dataset.

Independent donor grids reproduce the main history-update gains, while native-versus-donor differences are small. This supports a useful interaction between grid placement and the history rule, but does not show that choosing a new grid for each sample batch is necessary. Tables 27–29 report every replay setting.

Timing boundary. All runs use an exclusive H200 GPU, three synchronized timing repeats, and median sampling time within each generation seed. Replay includes model calls, updates and trace construction, but excludes acquiring the grid. Online timings additionally include spectral gating. Model loading, saving and metric evaluation are excluded in both. Table 30 compares online sampling with replay of its identical grid and update. Online selection adds 9.4–24.4% to replay time. Replay latency must not be presented as end-to-end online latency.

Table 27: Complete native-grid factorial with DDIM, under Appendix G. Values are mean ± sample SD over three generation/metric seeds on one frozen checkpoint. The first four grids match native NFE, and the last two match donor NFE. Blue identifies the spectrum-derived grid. Bold/underline rank C-FID within each of these two groups per dataset, including ties. Replay times exclude grid acquisition and are not ranked. No significance claim is made.

Data	Grid	NFE	C-FID	Replay seconds
Sines	Uniform	35.00	0.0297 ±0.0096	2.4210 ±0.0541
Sines	Uniform + same tail	35.00	**0.0291 ±0.0090**	2.4196 ±0.1028
Sines	Shuffled + same tail	35.00	0.0299 ±0.0076	2.4385 ±0.0478
Sines	**Spectrum-derived**	35.00	0.0295 ±0.0071	2.4246 ±0.0393
Sines	Donor-budget uniform	35.00	0.0297 ±0.0096	2.3968 ±0.0362
Sines	Independent donor	35.00	**0.0295 ±0.0071**	2.4058 ±0.0443
Stocks	Uniform	33.25	0.1936 ±0.0277	2.4249 ±0.1249
Stocks	Uniform + same tail	33.25	0.1914 ±0.0261	2.4138 ±0.0510
Stocks	Shuffled + same tail	33.25	**0.1858 ±0.0288**	2.4703 ±0.1373
Stocks	**Spectrum-derived**	33.25	0.1910 ±0.0293	2.4313 ±0.1382
Stocks	Donor-budget uniform	33.44	0.1941 ±0.0274	2.4227 ±0.0783
Stocks	Independent donor	33.44	**0.1912 ±0.0286**	2.4628 ±0.1066
ETTh	Uniform	14.15	**0.3912 ±0.0286**	1.1676 ±0.0125
ETTh	Uniform + same tail	14.15	0.3920 ±0.0273	1.1752 ±0.0243
ETTh	Shuffled + same tail	14.15	0.4369 ±0.0269	1.1685 ±0.0154
ETTh	**Spectrum-derived**	14.15	0.4201 ±0.0271	1.1983 ±0.0119
ETTh	Donor-budget uniform	13.94	**0.3928 ±0.0253**	1.1658 ±0.0230
ETTh	Independent donor	13.94	0.4208 ±0.0261	1.1585 ±0.0085
MuJoCo	Uniform	60.06	0.0642 ±0.0253	4.7919 ±0.1739
MuJoCo	Uniform + same tail	60.06	0.0643 ±0.0252	4.9382 ±0.2369
MuJoCo	Shuffled + same tail	60.06	**0.0638 ±0.0251**	4.8655 ±0.2496
MuJoCo	**Spectrum-derived**	60.06	**0.0638 ±0.0250**	4.8709 ±0.1472
MuJoCo	Donor-budget uniform	61.00	0.0641 ±0.0252	4.8087 ±0.0726
MuJoCo	Independent donor	61.00	**0.0635 ±0.0247**	4.8934 ±0.1061
Energy	Uniform	58.77	0.1082 ±0.0175	6.7893 ±0.2209
Energy	Uniform + same tail	58.77	**0.1071 ±0.0165**	6.8004 ±0.1907
Energy	Shuffled + same tail	58.77	0.1229 ±0.0133	6.8352 ±0.2446
Energy	**Spectrum-derived**	58.77	0.1206 ±0.0126	6.7650 ±0.1412
Energy	Donor-budget uniform	59.00	**0.1082 ±0.0175**	6.7945 ±0.1651
Energy	Independent donor	59.00	0.1206 ±0.0126	6.8024 ±0.2609
fMRI	Uniform	67.02	0.2234 ±0.0089	9.8266 ±0.3849
fMRI	Uniform + same tail	67.02	**0.2158 ±0.0092**	9.7657 ±0.8471
fMRI	Shuffled + same tail	67.02	0.2195 ±0.0098	9.8486 ±0.7791
fMRI	**Spectrum-derived**	67.02	0.2176 ±0.0092	10.0328 ±0.4234
fMRI	Donor-budget uniform	66.75	0.2235 ±0.0083	9.9498 ±0.2092
fMRI	Independent donor	66.75	**0.2178 ±0.0086**	9.9400 ±0.1696

Table 28: Complete native-grid factorial with the approximate-history update, under Appendix G. Values are mean ± sample SD over three generation/metric seeds on one frozen checkpoint. The first four grids match native NFE, and the last two match donor NFE. Blue identifies the spectrum-derived grid. Bold/underline rank C-FID within each of these two groups per dataset, including ties. Replay times exclude grid acquisition and are not ranked. No significance claim is made.

Data	Grid	NFE	C-FID	Replay seconds
Sines	Uniform	35.00	0.0316 ±0.0107	2.3854 ±0.0155
Sines	Uniform + same tail	35.00	0.0391 ±0.0141	2.4928 ±0.0926
Sines	Shuffled + same tail	35.00	0.0475 ±0.0159	2.4040 ±0.0586
Sines	**Spectrum-derived**	35.00	**0.0265 ±0.0094**	2.4354 ±0.0203
Sines	Donor-budget uniform	35.00	0.0316 ±0.0107	2.5031 ±0.0306
Sines	Independent donor	35.00	**0.0265 ±0.0094**	2.4220 ±0.0537
Stocks	Uniform	33.25	0.1179 ±0.0247	2.4667 ±0.0631
Stocks	Uniform + same tail	33.25	0.1167 ±0.0243	2.4836 ±0.1039
Stocks	Shuffled + same tail	33.25	**0.1082 ±0.0176**	2.4564 ±0.0574
Stocks	**Spectrum-derived**	33.25	0.1090 ±0.0241	2.4955 ±0.0735
Stocks	Donor-budget uniform	33.44	0.1198 ±0.0253	2.5420 ±0.1093
Stocks	Independent donor	33.44	**0.1092 ±0.0232**	2.5207 ±0.1413
ETTh	Uniform	14.15	0.3885 ±0.0557	1.2131 ±0.0068
ETTh	Uniform + same tail	14.15	0.3928 ±0.0538	1.2033 ±0.0178
ETTh	Shuffled + same tail	14.15	0.2850 ±0.0421	1.2168 ±0.0195
ETTh	**Spectrum-derived**	14.15	**0.1729 ±0.0245**	1.2245 ±0.0204
ETTh	Donor-budget uniform	13.94	0.3999 ±0.0504	1.1934 ±0.0615
ETTh	Independent donor	13.94	**0.1736 ±0.0245**	1.2111 ±0.0178
MuJoCo	Uniform	60.06	**0.0669 ±0.0258**	4.9476 ±0.1516
MuJoCo	Uniform + same tail	60.06	0.0685 ±0.0262	5.0264 ±0.2684
MuJoCo	Shuffled + same tail	60.06	0.0675 ±0.0259	4.9251 ±0.2113
MuJoCo	**Spectrum-derived**	60.06	0.0675 ±0.0258	4.9016 ±0.1912
MuJoCo	Donor-budget uniform	61.00	**0.0667 ±0.0255**	5.1814 ±0.3211
MuJoCo	Independent donor	61.00	0.0670 ±0.0252	5.0440 ±0.1857
Energy	Uniform	58.77	0.1020 ±0.0212	6.8734 ±0.1609
Energy	Uniform + same tail	58.77	**0.0977 ±0.0221**	6.8424 ±0.1554
Energy	Shuffled + same tail	58.77	0.0997 ±0.0251	6.8739 ±0.1718
Energy	**Spectrum-derived**	58.77	0.1050 ±0.0133	6.6026 ±0.0804
Energy	Donor-budget uniform	59.00	**0.1028 ±0.0214**	6.9387 ±0.2252
Energy	Independent donor	59.00	0.1051 ±0.0133	6.9786 ±0.2978
fMRI	Uniform	67.02	0.2124 ±0.0070	9.5066 ±0.4397
fMRI	Uniform + same tail	67.02	0.1479 ±0.0035	9.7822 ±0.4537
fMRI	Shuffled + same tail	67.02	0.1479 ±0.0036	9.9621 ±0.5679
fMRI	**Spectrum-derived**	67.02	**0.1467 ±0.0034**	10.1998 ±0.2811
fMRI	Donor-budget uniform	66.75	0.2151 ±0.0079	10.1019 ±0.1886
fMRI	Independent donor	66.75	**0.1465 ±0.0027**	9.9916 ±0.3936

Table 29: Complete native-grid factorial with actual-span DPM-Solver++ 2M, under Appendix G. Values are mean ± sample SD over three generation/metric seeds on one frozen checkpoint. The first four grids match native NFE, and the last two match donor NFE. Blue identifies the spectrum-derived grid. Bold/underline rank C-FID within each of these two groups per dataset, including ties. Replay times exclude grid acquisition and are not ranked. No significance claim is made.

Data	Grid	NFE	C-FID	Replay seconds
Sines	Uniform	35.00	**0.0318 ±0.0116**	2.4697 ±0.0415
Sines	Uniform + same tail	35.00	0.0331 ±0.0121	2.4254 ±0.0077
Sines	Shuffled + same tail	35.00	0.0353 ±0.0126	2.4949 ±0.0713
Sines	**Spectrum-derived**	35.00	0.0334 ±0.0151	2.5010 ±0.0782
Sines	Donor-budget uniform	35.00	**0.0318 ±0.0116**	2.4970 ±0.1090
Sines	Independent donor	35.00	0.0334 ±0.0151	2.5339 ±0.1025
Stocks	Uniform	33.25	0.1515 ±0.0232	2.4917 ±0.1083
Stocks	Uniform + same tail	33.25	0.1502 ±0.0224	2.4623 ±0.0642
Stocks	Shuffled + same tail	33.25	0.1176 ±0.0210	2.5416 ±0.1136
Stocks	**Spectrum-derived**	33.25	**0.1033 ±0.0295**	2.4868 ±0.0598
Stocks	Donor-budget uniform	33.44	0.1523 ±0.0234	2.5275 ±0.1040
Stocks	Independent donor	33.44	**0.1007 ±0.0263**	2.5256 ±0.1072
ETTh	Uniform	14.15	**0.2186 ±0.0162**	1.1958 ±0.0250
ETTh	Uniform + same tail	14.15	0.2187 ±0.0160	1.2146 ±0.0351
ETTh	Shuffled + same tail	14.15	0.2438 ±0.0190	1.2147 ±0.0220
ETTh	**Spectrum-derived**	14.15	0.2451 ±0.0167	1.1959 ±0.0133
ETTh	Donor-budget uniform	13.94	**0.2192 ±0.0173**	1.1953 ±0.0327
ETTh	Independent donor	13.94	0.2457 ±0.0169	1.1750 ±0.0409
MuJoCo	Uniform	60.06	**0.0674 ±0.0254**	4.9453 ±0.1536
MuJoCo	Uniform + same tail	60.06	0.0680 ±0.0259	4.9592 ±0.1397
MuJoCo	Shuffled + same tail	60.06	0.0675 ±0.0257	4.9688 ±0.2844
MuJoCo	**Spectrum-derived**	60.06	0.0675 ±0.0257	5.0587 ±0.2327
MuJoCo	Donor-budget uniform	61.00	0.0672 ±0.0252	5.1681 ±0.2720
MuJoCo	Independent donor	61.00	**0.0670 ±0.0252**	5.1381 ±0.2488
Energy	Uniform	58.77	0.0962 ±0.0213	6.8978 ±0.2304
Energy	Uniform + same tail	58.77	**0.0955 ±0.0212**	6.8799 ±0.3013
Energy	Shuffled + same tail	58.77	0.0997 ±0.0164	6.7912 ±0.2319
Energy	**Spectrum-derived**	58.77	0.0964 ±0.0133	6.8675 ±0.1475
Energy	Donor-budget uniform	59.00	**0.0962 ±0.0213**	6.9102 ±0.1423
Energy	Independent donor	59.00	0.0963 ±0.0133	6.8823 ±0.1733
fMRI	Uniform	67.02	**0.1493 ±0.0039**	10.1705 ±0.3232
fMRI	Uniform + same tail	67.02	0.1566 ±0.0048	9.7030 ±0.6823
fMRI	Shuffled + same tail	67.02	0.1564 ±0.0049	10.0081 ±0.6489
fMRI	**Spectrum-derived**	67.02	0.1574 ±0.0049	9.8118 ±0.6107
fMRI	Donor-budget uniform	66.75	**0.1492 ±0.0040**	9.8685 ±0.4842
fMRI	Independent donor	66.75	0.1570 ±0.0049	10.1082 ±0.1356

Table 30: Online native sampling versus replay of the same selected grid with the same approximate-history update. Replayed outputs are checked against online outputs. Means and sample SD use three generation/metric seeds. The online/replay time ratio measures the additional online-grid cost, not acceleration over a full diffusion trajectory. Blue marks online StrideDiffusion. Replay uses the same grid but excludes its online acquisition cost, so the two timing columns are not ranked.

Data	**Online C-FID**	**Online seconds**	Replay seconds	Time ratio
Sines	0.0265 ± 0.0094	3.0296 ± 0.0853	2.4354 ± 0.0203	1.24
Stocks	0.1090 ± 0.0241	3.0066 ± 0.1487	2.4955 ± 0.0735	1.20
ETTh	0.1729 ± 0.0245	1.4378 ± 0.0277	1.2245 ± 0.0204	1.17
MuJoCo	0.0675 ± 0.0258	5.8817 ± 0.2765	4.9016 ± 0.1912	1.20
Energy	0.1050 ± 0.0133	7.8147 ± 0.2760	6.6026 ± 0.0804	1.18
fMRI	0.1467 ± 0.0034	11.1629 ± 0.4178	10.1998 ± 0.2811	1.09

H INDEPENDENT TRAINING AND LONGER SEQUENCES

Training and held-out data. We train Diffusion-TS separately for Sines and ETTh at lengths 24, 96, and 192, using training seeds 411, 422, and 433. All 18 runs use fixed final EMA checkpoints, with 12,000 optimizer steps for Sines and 18,000 for ETTh. The EMA decay is 0.995. We do not choose checkpoints using sampler quality. Sines uses 10,000 training sequences and independent validation/test draws, with data seeds 123, 456, and 789. ETTh uses chronological 70/15/15 row splits, and windows never cross a split boundary. Min–max scaling is fitted on training rows only. Test values are not clipped. All methods share the same output clipping to $[0, 1]$, which cannot reproduce the small fraction of out-of-range ETTh test values.

Each frozen checkpoint is evaluated with generation/metric seeds 31415, 27182, and 16180. Each run uses 2,048 generated and held-out reference sequences, selected with seed 54321 where subsampling is needed. Its TS2Vec encoder is fitted on 2,000 reference sequences and shared across all methods within that run. We first average the three generation-seed results within a checkpoint, then report the mean and sample standard deviation across the three training checkpoints. Thus the independent training count is three, not nine. Training seeds vary initialization and optimization randomness on the same training data, not the data split.

Methods and budgets. All 20 methods use the same checkpoint and initial noise within a run. Uniform and spectral DDIM/history use the fixed-budget controller from Appendix F, at 16, 32, and 64 calls. They fix update-branch positions but differ from the original stride policy, including its fine-step tail. Their frequency cutoffs scale with length. DPM-Solver++ and UniPC use the official wrappers at the same three budgets. Full DDIM uses every diffusion step. The original complete sampler is also included. At lengths 96 and 192 it deliberately retains its original unscaled bands and is labeled accordingly. This tests zero-change transfer, not a length-scaled redesign of the full method.

Metrics and timing. C-FID follows the shared reference-only encoder protocol above. Corr. sums absolute differences in the lower triangle, including the diagonal, of channel-correlation matrices and divides by 10. Correlations pool every sample and temporal position. For PSD, we average unweighted rFFT power across samples and channels, normalize each dataset's spectrum to sum to one, and average absolute log-spectrum differences with a 10^{-12} floor. HF is the absolute difference in normalized power summed over bins $k \geq 6$. This legacy bin threshold changes its physical frequency with sequence length, so HF is compared only within a fixed length and is not the absolute-power metric in Appendix E.4. These diagnostics do not substitute for the discriminative and predictive evaluations in the main results.

Timing uses an exclusive H200 GPU and three synchronized repeats per generation seed. It includes sampling, gating, update arithmetic and trace construction, but excludes training, checkpoint loading, metric evaluation and saving. We take the median repeated time within each generation seed,

then average within and across checkpoints. NFE counts actual denoiser calls, including terminal predictions.

Results. Spectral DDIM improves mean C-FID in 10/18 length–budget settings. ETTh improves in all nine settings, with all three training checkpoints improving in each. Sines improves in only one of nine. Spectral history improves in six settings, all on Sines at lengths 96 and 192, and loses all nine ETTh settings. Sines length 96 improves across all three checkpoints at each budget, whereas length 192 has two checkpoint wins at each budget. These repeated effects support specific dataset–length–update combinations, not a uniform benefit from spectral scheduling.

The complete original sampler improves over full DDIM in three of six dataset–length combinations. A listed lower-NFE DPM-Solver++ setting has lower C-FID in five combinations. The remaining case, ETTh length 24, uses fewer than 16 calls for the original method but also favors the 16-call DPM-Solver++ control in C-FID. Unchanged original bands therefore do not give reliable length transfer. This does not rule out a separately validated length-scaled method, but such a method must be assessed as a new configuration.

Figure 16 shows every paired setting with checkpoint-level variation. Tables 31–36 retain all methods and auxiliary metrics. Each within-run paired comparison shares its encoder. We aggregate these runs hierarchically, but do not compare absolute C-FID across dataset–length cells or against the historical-reference tables.

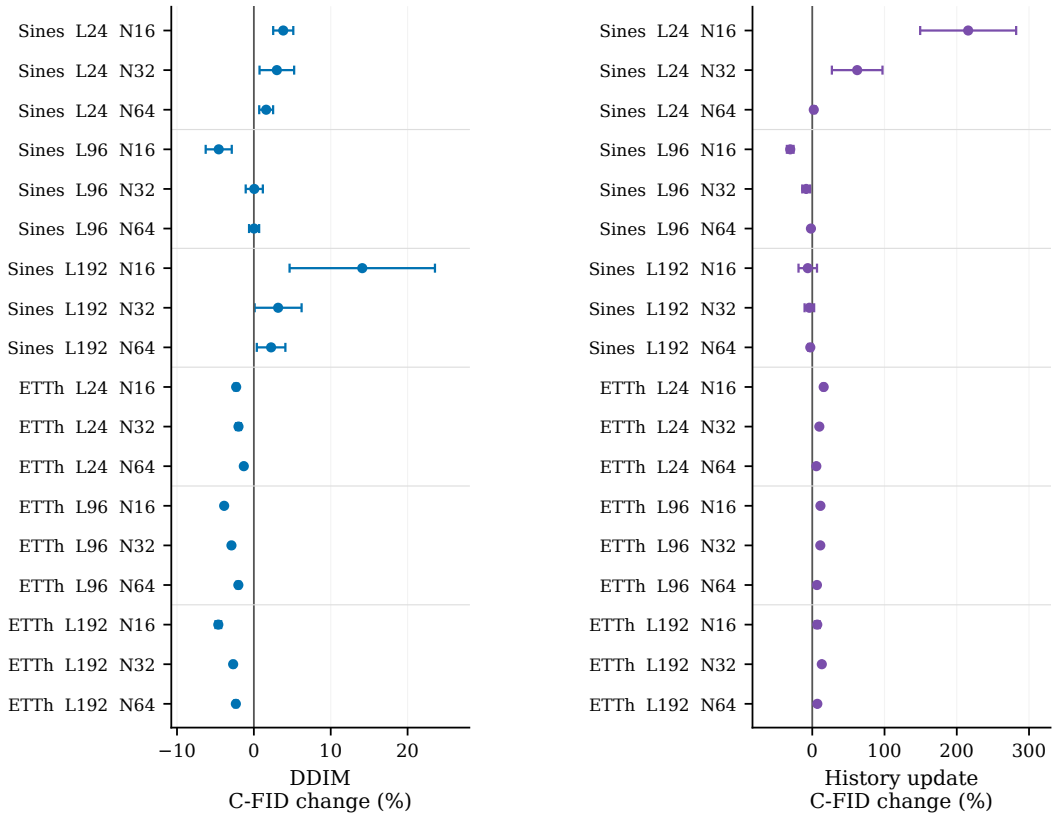

Figure 16: Paired grid effects across all 36 dataset–length–budget–update settings. L is sequence length and N is NFE. For each independently trained checkpoint, we average three generation/evaluator seeds, then calculate $100(\text{C-FID}_{\text{spectral}}/\text{C-FID}_{\text{uniform}} - 1)$. Points and bars show the mean and sample SD across three checkpoints. Left of zero favors the spectral grid. The two horizontal axes use different ranges. Both grids have the same NFE and update positions. This is the remaining-budget diagnostic controller, not unchanged transfer of the full sampler. The source package includes all means, SDs, checkpoint effects and win counts in `plot_data/appendix_long_grid_effects.csv`.

Table 31: Sines, length 24: all 20 methods. C-FID is mean ± sample SD across three trained checkpoints, each averaging three generation seeds. Other entries are means. Protocol and highlighting follow Appendix H. Ranks are within each 16-, 32- or 64-NFE group only.

Method	NFE	C-FID	Corr.	PSD	HF	Seconds
Full DDIM	500.00	0.0301 ±0.0041	0.0118	0.0164	1.35e-05	65.86
Ours	34.85	0.0313 ±0.0057	0.0198	0.0203	1.89e-05	5.66
Uniform DDIM	16.00	0.0359 ±0.0055	0.0280	0.0322	3.37e-05	**2.12**
Spectral DDIM	16.00	0.0373 ±0.0062	0.0285	0.0364	3.93e-05	2.67
Uniform history	16.00	0.0792 ±0.0020	0.0657	0.0556	6.56e-05	2.20
Spectral history	16.00	0.2509 ±0.0590	0.1460	0.1833	1.95e-04	2.73
DPM++(2M)	16.00	0.0322 ±0.0033	0.0219	**0.0128**	**1.12e-05**	3.37
UniPC-bh2	16.00	**0.0305 ±0.0038**	**0.0118**	0.0178	1.81e-05	3.53
Uniform DDIM	32.00	0.0305 ±0.0049	0.0155	0.0239	2.32e-05	**4.22**
Spectral DDIM	32.00	0.0314 ±0.0045	0.0157	0.0255	2.54e-05	5.32
Uniform history	32.00	0.0337 ±0.0033	0.0187	**0.0098**	**6.03e-06**	4.40
Spectral history	32.00	0.0540 ±0.0068	0.0467	0.0411	4.92e-05	5.44
DPM++(2M)	32.00	**0.0297 ±0.0035**	0.0147	0.0260	2.92e-05	6.75
UniPC-bh2	32.00	0.0305 ±0.0037	**0.0116**	0.0233	2.50e-05	7.09
Uniform DDIM	64.00	0.0297 ±0.0046	0.0124	0.0199	1.80e-05	**8.45**
Spectral DDIM	64.00	0.0301 ±0.0044	0.0125	0.0204	1.87e-05	10.62
Uniform history	64.00	0.0296 ±0.0040	0.0129	0.0161	1.29e-05	8.79
Spectral history	64.00	0.0302 ±0.0038	0.0153	**0.0103**	**6.54e-06**	10.94
DPM++(2M)	64.00	**0.0283 ±0.0030**	0.0133	0.0204	2.11e-05	13.49
UniPC-bh2	64.00	0.0302 ±0.0036	**0.0115**	0.0235	2.48e-05	14.19

Table 32: Sines, length 96: all 20 methods. C-FID is mean ± sample SD across three trained checkpoints, each averaging three generation seeds. Other entries are means. Protocol and highlighting follow Appendix H. Ranks are within each 16-, 32- or 64-NFE group only.

Method	NFE	C-FID	Corr.	PSD	HF	Seconds
Full DDIM	500.00	0.1157 ±0.0009	0.0214	0.0163	9.20e-06	67.94
Ours, original bands	27.70	0.1052 ±0.0028	0.0209	0.0165	9.14e-06	4.57
Uniform DDIM	16.00	0.1048 ±0.0016	0.0207	0.0179	**9.15e-06**	**2.17**
Spectral DDIM	16.00	0.1000 ±0.0029	0.0211	0.0185	9.52e-06	2.72
Uniform history	16.00	0.1349 ±0.0017	0.0221	0.0241	1.60e-05	2.25
Spectral history	16.00	**0.0937 ±0.0051**	**0.0204**	**0.0139**	9.16e-06	2.77
DPM++(2M)	16.00	0.1032 ±0.0006	0.0208	0.0394	2.93e-05	3.43
UniPC-bh2	16.00	0.1133 ±0.0014	0.0216	0.0392	2.87e-05	3.58
Uniform DDIM	32.00	**0.1096 ±0.0009**	0.0212	0.0167	8.50e-06	**4.35**
Spectral DDIM	32.00	**0.1096 ±0.0005**	0.0210	**0.0152**	**8.03e-06**	5.43
Uniform history	32.00	0.1240 ±0.0022	0.0212	0.0159	1.12e-05	4.52
Spectral history	32.00	0.1136 ±0.0064	0.0251	0.0192	1.45e-05	5.56
DPM++(2M)	32.00	0.1098 ±0.0007	**0.0209**	0.0644	4.31e-05	6.85
UniPC-bh2	32.00	0.1171 ±0.0011	0.0216	0.0479	3.34e-05	7.16
Uniform DDIM	64.00	0.1125 ±0.0007	0.0213	0.0162	8.50e-06	**8.70**
Spectral DDIM	64.00	0.1125 ±0.0001	0.0211	**0.0155**	**8.24e-06**	10.85
Uniform history	64.00	0.1180 ±0.0013	0.0214	0.0183	1.06e-05	9.00
Spectral history	64.00	0.1158 ±0.0012	**0.0209**	0.0175	1.01e-05	11.11
DPM++(2M)	64.00	**0.1113 ±0.0008**	**0.0209**	0.0383	2.57e-05	13.68
UniPC-bh2	64.00	0.1169 ±0.0007	0.0215	0.0428	2.89e-05	14.38

Table 33: Sines, length 192: all 20 methods. C-FID is mean ± sample SD across three trained checkpoints, each averaging three generation seeds. Other entries are means. Protocol and highlighting follow Appendix H. Ranks are within each 16-, 32- or 64-NFE group only.

Method	NFE	C-FID	Corr.	PSD	HF	Seconds
Full DDIM	500.00	0.1459 ±0.0116	0.0168	0.0608	6.31e-05	71.49
Ours, original bands	86.25	0.1447 ±0.0232	0.0164	0.0684	8.09e-05	14.74
Uniform DDIM	16.00	0.1483 ±0.0236	0.0162	0.0763	1.01e-04	**2.30**
Spectral DDIM	16.00	0.1704 ±0.0372	0.0161	0.0921	1.36e-04	2.85
Uniform history	16.00	0.1530 ±0.0118	0.0180	**0.0494**	**2.46e-05**	2.37
Spectral history	16.00	0.1433 ±0.0217	**0.0156**	0.0597	3.26e-05	2.92
DPM++(2M)	16.00	**0.1357 ±0.0164**	0.0166	0.0666	3.60e-05	3.55
UniPC-bh2	16.00	0.1470 ±0.0160	0.0160	0.0593	8.02e-05	3.73
Uniform DDIM	32.00	0.1443 ±0.0164	0.0165	0.0688	8.29e-05	**4.59**
Spectral DDIM	32.00	0.1492 ±0.0213	**0.0160**	0.0807	1.09e-04	5.68
Uniform history	32.00	0.1460 ±0.0080	0.0167	0.0516	3.81e-05	4.72
Spectral history	32.00	**0.1404 ±0.0166**	0.0173	**0.0488**	**1.04e-05**	5.78
DPM++(2M)	32.00	0.1406 ±0.0140	0.0163	0.0990	3.37e-05	7.11
UniPC-bh2	32.00	0.1486 ±0.0139	0.0165	0.0675	6.70e-05	7.44
Uniform DDIM	64.00	0.1446 ±0.0138	0.0167	0.0650	7.37e-05	**9.17**
Spectral DDIM	64.00	0.1479 ±0.0163	**0.0164**	0.0758	9.89e-05	11.35
Uniform history	64.00	0.1472 ±0.0104	0.0170	0.0534	4.25e-05	9.49
Spectral history	64.00	**0.1431 ±0.0087**	0.0167	**0.0512**	**3.56e-05**	11.67
DPM++(2M)	64.00	**0.1431 ±0.0126**	**0.0164**	0.0607	4.98e-05	14.23
UniPC-bh2	64.00	0.1480 ±0.0126	0.0167	0.0635	6.21e-05	14.93

Table 34: ETTh, length 24: all 20 methods. C-FID is mean ± sample SD across three trained checkpoints, each averaging three generation seeds. Other entries are means. Protocol and highlighting follow Appendix H. Ranks are within each 16-, 32- or 64-NFE group only.

Method	NFE	C-FID	Corr.	PSD	HF	Seconds
Full DDIM	500.00	4.2523 ±0.0129	0.4253	0.8412	9.27e-04	76.42
Ours	13.79	4.2177 ±0.0340	0.4055	0.3809	4.31e-04	2.56
Uniform DDIM	16.00	4.4217 ±0.0262	0.4296	1.2258	1.18e-03	**2.46**
Spectral DDIM	16.00	4.3198 ±0.0259	0.4257	1.0532	1.07e-03	3.00
Uniform history	16.00	**3.6431 ±0.0580**	**0.3809**	0.4095	9.48e-04	2.53
Spectral history	16.00	4.2177 ±0.0033	0.4075	**0.3888**	**4.21e-04**	3.05
DPM++(2M)	16.00	4.0774 ±0.0182	0.4165	0.6004	7.04e-04	3.71
UniPC-bh2	16.00	4.3709 ±0.0099	0.4283	1.0754	1.11e-03	3.86
Uniform DDIM	32.00	4.3416 ±0.0232	0.4276	1.0402	1.08e-03	**4.91**
Spectral DDIM	32.00	4.2547 ±0.0171	0.4248	0.9069	9.72e-04	5.99
Uniform history	32.00	**3.8944 ±0.0480**	**0.4055**	**0.2426**	**1.71e-04**	5.07
Spectral history	32.00	4.2744 ±0.0045	0.4210	0.7193	8.15e-04	6.11
DPM++(2M)	32.00	4.1870 ±0.0186	0.4256	0.6201	7.07e-04	7.40
UniPC-bh2	32.00	4.3290 ±0.0107	0.4282	0.8947	9.76e-04	7.72
Uniform DDIM	64.00	4.2966 ±0.0151	0.4265	0.9354	1.00e-03	**9.82**
Spectral DDIM	64.00	4.2404 ±0.0163	0.4246	0.8604	9.39e-04	11.94
Uniform history	64.00	**4.0437 ±0.0066**	**0.4178**	**0.4202**	**4.14e-04**	10.15
Spectral history	64.00	4.2706 ±0.0122	0.4241	0.8231	9.10e-04	12.24
DPM++(2M)	64.00	4.2053 ±0.0038	0.4273	0.6807	7.69e-04	14.78
UniPC-bh2	64.00	4.2894 ±0.0076	0.4277	0.8036	8.94e-04	15.46

Table 35: ETTh, length 96: all 20 methods. C-FID is mean ± sample SD across three trained checkpoints, each averaging three generation seeds. Other entries are means. Protocol and highlighting follow Appendix H. Ranks are within each 16-, 32- or 64-NFE group only.

Method	NFE	C-FID	Corr.	PSD	HF	Seconds
Full DDIM	500.00	8.9378 ±0.2516	0.4235	0.8440	4.25e-03	81.55
Ours, original bands	25.86	9.3536 ±0.1280	0.4045	0.6830	3.62e-03	5.02
Uniform DDIM	16.00	9.1195 ±0.2169	0.4315	1.2916	4.92e-03	**2.62**
Spectral DDIM	16.00	8.7671 ±0.2111	0.4261	1.1405	4.81e-03	3.17
Uniform history	16.00	**7.4230 ±0.2641**	0.3833	0.4761	**1.70e-04**	2.69
Spectral history	16.00	8.2591 ±0.2249	**0.3821**	**0.3830**	7.30e-04	3.22
DPM++(2M)	16.00	8.2504 ±0.2130	0.4216	0.5763	3.67e-03	3.89
UniPC-bh2	16.00	9.2068 ±0.2020	0.4311	1.1319	4.63e-03	4.03
Uniform DDIM	32.00	9.0393 ±0.2152	0.4288	1.0709	4.61e-03	**5.23**
Spectral DDIM	32.00	8.7759 ±0.2227	0.4253	0.9709	4.52e-03	6.27
Uniform history	32.00	**8.1298 ±0.1556**	**0.3999**	**0.3649**	**1.57e-03**	5.42
Spectral history	32.00	9.0338 ±0.2598	0.4089	0.5017	3.53e-03	6.46
DPM++(2M)	32.00	8.6697 ±0.1764	0.4259	0.5546	3.72e-03	7.73
UniPC-bh2	32.00	9.1530 ±0.1923	0.4292	0.9026	4.36e-03	8.06
Uniform DDIM	64.00	8.9986 ±0.2306	0.4264	0.9481	4.43e-03	**10.46**
Spectral DDIM	64.00	8.8173 ±0.2454	0.4240	0.8719	4.33e-03	12.59
Uniform history	64.00	**8.5530 ±0.2347**	**0.4128**	**0.3553**	**3.26e-03**	10.82
Spectral history	64.00	9.1050 ±0.2477	0.4201	0.8147	4.19e-03	12.86
DPM++(2M)	64.00	8.8276 ±0.2030	0.4265	0.6361	3.92e-03	15.53
UniPC-bh2	64.00	9.0639 ±0.1987	0.4277	0.7866	4.19e-03	16.17

Table 36: ETTh, length 192: all 20 methods. C-FID is mean ± sample SD across three trained checkpoints, each averaging three generation seeds. Other entries are means. Protocol and highlighting follow Appendix H. Ranks are within each 16-, 32- or 64-NFE group only.

Method	NFE	C-FID	Corr.	PSD	HF	Seconds
Full DDIM	500.00	12.5901 ±1.0836	0.4369	0.7803	1.98e-02	85.61
Ours, original bands	28.27	13.2244 ±0.7717	0.4174	0.6090	1.84e-02	5.72
Uniform DDIM	16.00	12.7655 ±1.2590	0.4471	1.1907	2.08e-02	**2.72**
Spectral DDIM	16.00	12.1751 ±1.2030	0.4437	1.0577	2.08e-02	3.26
Uniform history	16.00	**10.4703 ±0.2712**	0.3906	0.4159	**1.46e-02**	2.81
Spectral history	16.00	11.1963 ±0.5598	**0.3840**	**0.3941**	1.47e-02	3.34
DPM++(2M)	16.00	11.5072 ±0.8600	0.4340	0.5138	1.89e-02	4.01
UniPC-bh2	16.00	12.9526 ±1.1951	0.4450	1.0505	2.03e-02	4.19
Uniform DDIM	32.00	12.6777 ±1.1541	0.4423	0.9796	2.03e-02	**5.46**
Spectral DDIM	32.00	12.3371 ±1.1404	0.4402	0.9190	2.03e-02	6.55
Uniform history	32.00	**11.3221 ±0.8347**	**0.4082**	**0.3359**	**1.69e-02**	5.57
Spectral history	32.00	12.8167 ±0.9715	0.4176	0.4280	1.88e-02	6.66
DPM++(2M)	32.00	12.1410 ±1.0493	0.4381	0.4879	1.91e-02	7.95
UniPC-bh2	32.00	12.9082 ±1.1725	0.4426	0.8260	2.00e-02	8.34
Uniform DDIM	64.00	12.6669 ±1.1345	0.4395	0.8701	2.01e-02	**10.88**
Spectral DDIM	64.00	12.3687 ±1.0814	0.4377	0.8107	2.00e-02	13.00
Uniform history	64.00	**12.0392 ±1.0032**	**0.4267**	**0.3526**	**1.88e-02**	11.16
Spectral history	64.00	12.8816 ±1.0721	0.4314	0.7400	1.97e-02	13.35
DPM++(2M)	64.00	12.4159 ±1.1023	0.4405	0.5718	1.94e-02	15.96
UniPC-bh2	64.00	12.8063 ±1.1516	0.4415	0.7172	1.98e-02	16.77

I THRESHOLD REUSE AND ADDITIONAL BACKBONES

I.1 THRESHOLDS VERSUS THE FULL COMPUTE SCHEDULE

The three Sines thresholds $(.10, .01, .04)$ can be copied to the other datasets while keeping each target's stride set and final window. Table 2 reports this scheduler-only DDIM experiment. Its endpoint convention and measured budgets differ from the exact-sampler comparisons in the main text, so it should be read as a paired transfer test within this table. The largest mean C-FID increase is .0004, compared with a seed standard deviation of .0052 in that cell. This is descriptive evidence of threshold robustness, not a statistical equivalence test.

Copying thresholds, strides, and the final window together is a stronger intervention. When scaled only by the diffusion horizon, those settings can leave fMRI with 10 to 19 evaluations and unstable quality. We therefore separate threshold reuse from selection of a compute budget. A fixed rule that targets 80 NFE offers a further test: a 64-trajectory model-only pilot chooses a threshold multiplier without consulting reference data or a quality metric. Figure 17 shows that this controls the call budget reasonably well but does not consistently match target-tuned quality. The pilot time is additional setup cost, not included in sampling-only latency. We do not use this pilot as the default method or as evidence of universal calibration.

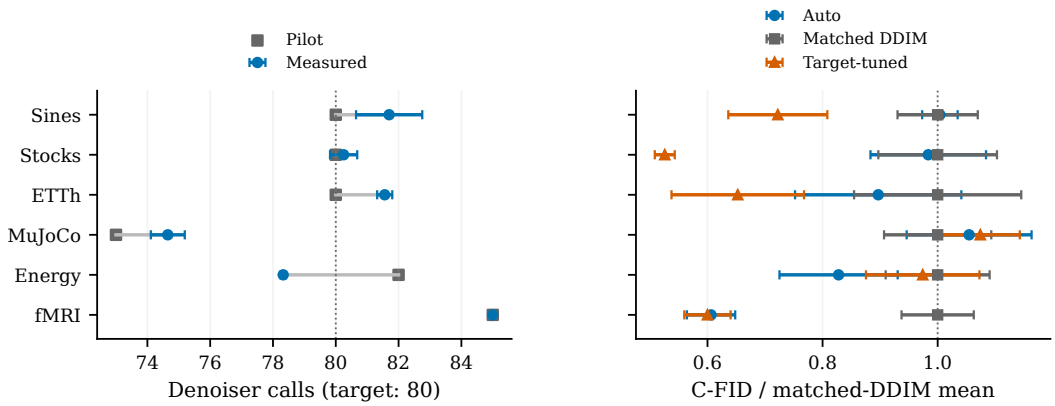

Figure 17: Model-only calibration on all six datasets. Left: pilot NFE and measured NFE (mean ± sample SD over three generation seeds), with the target at 80. Right: three-seed C-FID means and sample SDs divided by the matched-DDIM mean as a fixed display scale. This is not a paired-ratio confidence interval. Target-tuned settings use different NFE budgets, not a matched-budget control. Pilot setup time is additional. Exact budgets, quality, uncertainty and pilot cost are in `plot_data/calibration_tradeoff.csv`.

I.2 EDM AND CASCADED DIFFUSION PORTS

Table 37 gives the absolute values in Figure 6. ImagenTime (Naiman et al., 2024) and Imagen-Few (Gonen et al., 2025) generate an image representation of time series with an EDM parameterization. We decode the visited state to a time series for spectral monitoring and adapt the time grid in noise scale. Their native full samplers use Heun updates. Our port, named *Ours: Spectral PC2 (StrideDiffusion port)*, combines an EDM-specific spectral grid with a two-step history update and selected Heun corrections. It is a parameterization-specific port, not an unchanged call to Algorithm 1. Uniform PC2 matches the transition count, NFE, and Heun-correction count. The next subsection states which update choices differ.

ImagenTime uses the official code with a separately trained Sine checkpoint at each length. The length-96 setting extends the Sine experiment and is not an official long-series benchmark configuration. ImagenFew starts from the released pretrained checkpoint and uses its 50-example, all-parameter adaptation on Sine. Training or adaptation builds the fixed backbone for each experiment. The sampler comparison itself does not update those weights. Each row generates 2,000 sequences per

Table 37: Spectral PC2 ports on Sine. Full Heun is the native sampler. Uniform PC2 matches NFE and correction count, but not every update choice (Appendix I.3). Mean ± sample standard deviation over seeds 2032, 2033, and 2034. Samplers share a frozen checkpoint within each backbone/length setting. Bold marks the lowest mean in each group across the listed NFE budgets. Underline marks the second-lowest distinct displayed mean within the same comparison groups as bold, including ties. Neither mark indicates significance.

Backbone / length	Sampler	NFE	Time (s)	C-FID
ImagenTime 24	Full Heun	35	6.662 ±.067	.02740 ±.00545
ImagenTime 24	Uniform PC2	21	**4.015 ±.029**	.01968 ±.00043
ImagenTime 24	**Spectral PC2 (ours)**	21	4.278 ±.003	**.01829 ±.00147**
ImagenTime 96	Full Heun	35	8.295 ±.050	.06850 ±.00034
ImagenTime 96	Uniform PC2	21	**4.913 ±.001**	.07441 ±.00226
ImagenTime 96	**Spectral PC2 (ours)**	21	5.308 ±.016	**.06767 ±.00064**
ImagenFew 24	Full Heun	71	11.731 ±.097	2.82171 ±.01939
ImagenFew 24	Uniform PC2	18	**2.965 ±.006**	2.35131 ±.01236
ImagenFew 24	**Spectral PC2 (ours)**	18	3.268 ±.009	**2.17834 ±.01021**
ImagenFew 64	Full Heun	71	11.903 ±.116	1.07573 ±.01845
ImagenFew 64	Uniform PC2	22	**3.718 ±.016**	1.07162 ±.01859
ImagenFew 64	**Spectral PC2 (ours)**	22	4.026 ±.011	**1.03898 ±.01635**

seed. C-FID encoders are shared within each backbone/length comparison. Metric scales across different backbones need not be comparable.

Configuration selection. We searched stride budgets, spectral profiles, history coefficients, and numbers of Heun corrections on generation seed 12345. We selected configurations using quality and runtime, then evaluated frozen settings on seeds 2032, 2033, and 2034. Several frozen candidates were evaluated in this confirmation stage. The table presents selected operating points rather than an independent test of a single predeclared default. The uniform PC2 controls share the selected numerical-update budget, but we have not performed an equally large grid-family search for every baseline. Table 38 reports all auxiliary metrics alongside C-FID rather than selecting a different configuration for each metric. In particular, ImagenFew's correlation error can worsen even when C-FID improves.

The ImagenTime length-24 setting also transfers unchanged to length 96: 21 NFE and C-FID $.06796 \pm .00024$, versus 35 NFE and $.06850 \pm .00034$ for full Heun. It has lower C-FID on two of the three seeds. The small improvement supports reuse of that setting across these two lengths, rather than a broad claim about all long sequences.

I.3 PC2 UPDATE AND CONFIRMATION CANDIDATES

For an EDM state z_i at scale σ_i, define $d_i = (z_i - \hat{z}_{0,i})/\sigma_i$ and $s_i = \sigma_{i+1} - \sigma_i$. A history step uses

$$z_{i+1} = z_i + s_i \left[d_i + \frac{\eta}{2} \frac{s_i}{s_{i-1}} (d_i - d_{i-1}) \right]. \tag{16}$$

The first step uses Euler if no correction is selected. A Heun correction makes one additional denoiser call at the Euler proposal and averages the current and proposed derivatives. A fixed correction count contributes to NFE on top of the transition count. The spectral controller selects indices from a denser noise-scale grid subject to a remaining-transition budget. Thus both budget constraints and spectral activity can determine the next index.

For ImagenFew at length 64, the stronger C-FID row therefore changes more than the grid. An already evaluated alternative with evenly spaced corrections and $\eta = 1$ yields C-FID $1.06747 \pm .01322$, versus $1.07162 \pm .01859$ for uniform PC2 at the same 22 NFE. This smaller difference gives a more restricted comparison. We retain both outcomes rather than attributing the entire 1.03898 result to spectral placement alone. Tables 40 and 41 list every evaluated candidate from these confirmation runs. Their codes identify the numerical configuration, not distinct backbones.

Table 38: Additional-backbone results on additional diffusion implementations. Values are mean ± sample standard deviation over three generation seeds. "Win" counts seeds with lower sampling time and Context-FID than the corresponding full method. "Baseline" denotes the native full sampler, "Control" a non-spectral matched-budget sampler, and "Ours" a StrideDiffusion port. Uniform and Spectral rows match NFE and correction counts. Correction placement and history coefficients can differ, as specified in Appendix I.3. Bold marks the lowest mean in each group across the listed NFE budgets. Underline marks the second-lowest distinct displayed mean within the same comparison groups as bold, including ties. Neither mark indicates significance.

Backbone	Length	Method	NFE	Time (s)	Context-FID	Corr. error	log-PSD error	Win
ImagenTime	24	Full Heun	35	6.662 ±.067	.02740 ±.00545	.03058 ±.00418	.01992 ±.00335	–
ImagenTime	24	Uniform PC2	21	**4.015 ±.029**	.01968 ±.00043	.01470 ±.00189	**.00575 ±.00217**	3/3
ImagenTime	24	**Spectral PC2 (ours)**	21	4.278 ±.003	**.01829 ±.00147**	**.01285 ±.00123**	.00720 ±.00181	3/3
ImagenTime	96	Full Heun	35	8.295 ±.050	.06850 ±.00034	**.01148 ±.00074**	.01651 ±.00717	–
ImagenTime	96	Uniform PC2	21	**4.913 ±.001**	.07441 ±.00226	.01487 ±.00143	**.00748 ±.00348**	0/3
ImagenTime	96	**Spectral PC2 (ours)**	21	5.308 ±.016	**.06767 ±.00064**	.01189 ±.00120	.01203 ±.00666	3/3
ImagenFew	24	Full Heun	71	11.731 ±.097	2.82171 ±.01939	**.03068 ±.00063**	.38813 ±.00078	–
ImagenFew	24	Uniform PC2	18	**2.965 ±.006**	2.35131 ±.01236	.03632 ±.00105	.40144 ±.00033	3/3
ImagenFew	24	**Spectral PC2 (ours)**	18	3.268 ±.009	**2.17834 ±.01021**	.05373 ±.00168	**.35148 ±.00043**	3/3
ImagenFew	64	Full Heun	71	11.903 ±.116	1.07573 ±.01845	**.06947 ±.00173**	.31669 ±.00671	–
ImagenFew	64	Uniform PC2	22	**3.718 ±.016**	1.07162 ±.01859	.06981 ±.00171	**.27834 ±.00719**	3/3
ImagenFew	64	**Spectral PC2 (ours)**	22	4.026 ±.011	**1.03898 ±.01635**	.07123 ±.00143	.40982 ±.00609	3/3
L2D-Diff (our impl.)	24	Full cascade	200	8.335 ±.003	.52602 ±.02619	.01986 ±.00193	1.45977 ±.00679	–
L2D-Diff (our impl.)	24	Uniform history	90	**4.958 ±.001**	.51575 ±.02617	**.01869 ±.00198**	**1.45777 ±.00701**	3/3
L2D-Diff (our impl.)	24	**Spectral history (ours)**	90	5.467 ±.001	**.51361 ±.02612**	.01881 ±.00197	1.46020 ±.00700	3/3

Table 39: Update choices for the reported PC2 ports. Uniform PC2 uses evenly spaced correction positions and $\eta = 1$ in every setting. Both methods have the transition and correction counts shown here.

Backbone	Length	Transitions	Corrections	**Ours η**	**Ours correction trigger**
ImagenTime	24	15	6	1.00	Evenly spaced
ImagenTime	96	15	6	1.00	Evenly spaced
ImagenFew	24	16	2	1.00	Spectral activity
ImagenFew	64	16	6	0.75	Relative derivative change

Table 40: All evaluated confirmation-stage candidates for ImagenTime. Candidate codes encode transitions (t), Heun corrections (c), correction policy (p), threshold multiplier (q), and history scale (h). Policies e, s, and v mean evenly spaced, spectral, and derivative-change based. The suffix h075 means 0.75 and h100 means 1.0. These are the same three generation seeds used in the main backbone table. Blue rows and bold method names identify our spectral PC2 candidates. No numerical rank is assigned across different budgets and update choices.

Length	Candidate	NFE	Time (s)	C-FID
24	Full Heun n35	35	6.662 ±0.067	0.02740 ±0.00545
24	Uniform Heun n21	21	3.983 ±0.025	0.08587 ±0.01106
24	Uniform PC2 t15 c06 n21	21	4.015 ±0.029	0.01968 ±0.00043
24	**Spectral PC2** t15 c06 pe q101010 h100	21	4.315 ±0.033	0.02081 ±0.00421
24	**Spectral PC2** t15 c06 pe q050505 h100	21	4.278 ±0.003	0.01829 ±0.00147
96	Full Heun n35	35	8.295 ±0.050	0.06850 ±0.00034
96	Uniform Heun n21	21	4.871 ±0.009	0.07086 ±0.00096
96	Uniform PC2 t15 c06 n21	21	4.913 ±0.001	0.07441 ±0.00226
96	**Spectral PC2** t15 c06 pe q101010 h100	21	5.308 ±0.016	0.06767 ±0.00064
96	**Spectral PC2** t15 c06 pe q050505 h100	21	5.292 ±0.008	0.06796 ±0.00024

Table 41: All evaluated confirmation-stage candidates for ImagenFew. Candidate codes encode transitions (t), Heun corrections (c), correction policy (p), threshold multiplier (q), and history scale (h). Policies e, s, and v mean evenly spaced, spectral, and derivative-change based. The suffix h075 means 0.75 and h100 means 1.0. These are the same three generation seeds used in the main backbone table. Blue rows and bold method names identify our spectral PC2 candidates. No numerical rank is assigned across different budgets and update choices.

Length	Candidate	NFE	Time (s)	C-FID
24	Full Heun n71	71	11.731 ±0.097	2.82171 ±0.01939
24	Uniform Heun n18	18	2.964 ±0.033	3.74746 ±0.02255
24	Uniform Heun n20	20	3.286 ±0.010	3.36119 ±0.02122
24	Uniform PC2 t18 c02 n20	20	3.301 ±0.013	2.40351 ±0.01260
24	Uniform PC2 t16 c02 n18	18	2.965 ±0.006	2.35131 ±0.01236
24	**Spectral PC2** t18 c02 pe q050505 h075	20	3.637 ±0.032	2.11308 ±0.00951
24	**Spectral PC2** t16 c02 ps q101010 h100	18	3.268 ±0.009	2.17834 ±0.01021
24	**Spectral PC2** t18 c02 pe q101010 h075	20	3.609 ±0.004	2.11308 ±0.00951
64	Full Heun n71	71	11.903 ±0.116	1.07573 ±0.01845
64	Uniform Heun n22	22	3.683 ±0.016	1.04374 ±0.01849
64	Uniform PC2 t16 c06 n22	22	3.718 ±0.016	1.07162 ±0.01859
64	**Spectral PC2** t16 c06 pv q101010 h075	22	4.041 ±0.039	1.03898 ±0.01635
64	**Spectral PC2** t16 c06 pe q050505 h100	22	4.020 ±0.002	1.06747 ±0.01322
64	**Spectral PC2** t16 c06 pv q050505 h075	22	4.026 ±0.011	1.03898 ±0.01635

L2D-Diff implementation. We built a local implementation based on L2D-Diff (Shen et al., 2026) and trained it on Stock. These runs use our code and checkpoints, not the authors' implementation. The results are labeled "our impl." throughout. The full cascade uses 200 NFE across its latent and data branches. Both reduced variants use 30 latent-branch calls and 60 data-branch calls. The spectral data grid gives C-FID $.51361 \pm .02612$, compared with $.51575 \pm .02617$ for the matched uniform grid and $.52602 \pm .02619$ for the full cascade. The 0.42% mean improvement over the matched control is small. These are supporting experiments on a local reimplementation, not an official L2D-Diff reproduction or proof that the DDIM result covers a cascade.

J ADDITIONAL TRAINED GENERATORS

Training and sampling. We train a multivariate DiffWave adaptation and a local, unconstrained DiffTime implementation on each of the six datasets at length 24. DiffWave retains the original residual blocks with multivariate input and output projections. DiffTime follows the paper's model design rather than using an author-provided checkpoint. Each baseline uses training seeds 411, 422, and 433, 12,000 Adam updates, batch size 128, and gradient clipping at norm 1. We use the final checkpoint without selecting it by generation quality. DiffWave learns noise with an L1 loss and learning rate 2×10^{-4}. DiffTime uses squared error and learning rate 10^{-4}. Both use 50-step ancestral sampling. DiffWave uses a linear noise schedule from 10^{-4} to 0.05 and clips intermediate states. DiffTime uses a quadratic schedule from 10^{-6} to 0.5 without intermediate clipping. These are new adapted-model experiments, separate from both the initialized timing models in Appendix E.1 and the original benchmark rows.

Evaluation and replication. Each model generates 2,048 sequences for each of seeds 31415, 27182, and 16180. The reference windows come from the training-reference pool, so these scores assess distribution matching on that pool. Within each dataset, training-label and generation-seed group, all methods share a reference-fitted TS2Vec encoder. Three evaluator seeds are averaged before the aggregation described in the table captions. Baseline uncertainty is measured across three trained models, after averaging generation seeds within each model. Ours uses one existing EMA checkpoint per dataset. Table 42 reports its generation variation, not independent retraining uncertainty. The actual Ours training steps are listed in that caption. Training histories and sampling budgets are unequal.

Quality and cost. At these training budgets, Ours has lower mean C-FID than both adapted baselines on all six datasets. DiffTime has lower predictive error on Sines and lower Subset Corr. on Stocks. Ours is faster than both baselines on ETTh, but slower on the other five datasets. Table 43 reports actual NFE, sampling time and final clipping. All quality scores use the final clipped samples. Table 44 computes differences within groups before aggregation, keeping the shared encoder fixed for each comparison. These results compare complete generators. They neither establish that the baselines have converged nor isolate the spectral scheduler. The same-checkpoint controls in Section 5.2 address that separate question.

Table 42: Generation quality at length 24, using 2,048 training-reference windows. DiffWave is a multivariate adaptation and DiffTime is our paper-based implementation. Only these two baselines use the 12,000-update training budget. Ours uses frozen EMA checkpoints at Sines 12,000, Stocks 10,000, ETTh 16,200, MuJoCo 14,000, Energy 25,000, fMRI 15,000 updates. Training batch sizes and EMA histories also differ. Baselines average three generation seeds per model, then report mean and sample SD across three trained models. Ours reports three generation seeds paired with baseline label 411, from one frozen model per dataset. Its SD measures generation variation, not retraining uncertainty. All quality metrics score the final samples after $x \mapsto \mathrm{clip}((x+1)/2, 0, 1)$. Subset Corr. is the L1 difference of lag-zero lower-triangular correlation signatures divided by 10. Each of three evaluator seeds draws 682 of 2,048 real and generated windows with replacement. Disc. and Pred. also average three evaluator seeds before the generation/training aggregation. Each dataset/training-label/generation block fits TS2Vec on 2,000 real windows and shares it across all three methods. Lower is better. This compares complete generators with unequal training histories, not convergence, unseen-data generalization or scheduler attribution. Differences are descriptive, without significance testing. Blue shading identifies Ours. Bold and underline mark the lowest and second-lowest distinct displayed means within each dataset, including ties. These are descriptive ranks under unequal training histories, not significance tests.

Data	Method	Models	C-FID	Subset Corr.	Disc.	Pred.
Sines	**StrideDiffusion**	1	**0.0265 ±0.0094**	**0.0325 ±0.0057**	**0.0263 ±0.0037**	0.0974 ±0.0004
Sines	DiffWave (adapted)	3	2.1079 ±0.2674	0.0949 ±0.0291	0.4311 ±0.0115	0.1664 ±0.0169
Sines	DiffTime (local)	3	0.0315 ±0.0253	0.0501 ±0.0169	0.0339 ±0.0344	**0.0970 ±0.0014**
Stocks	**StrideDiffusion**	1	**0.1066 ±0.0207**	0.0341 ±0.0032	**0.0797 ±0.0104**	**0.0375 ±0.0001**
Stocks	DiffWave (adapted)	3	5.0945 ±0.4088	0.0666 ±0.0090	0.3810 ±0.0158	0.0419 ±0.0007
Stocks	DiffTime (local)	3	1.2521 ±1.2115	**0.0161 ±0.0022**	0.1980 ±0.0591	0.0398 ±0.0031
ETTh	**StrideDiffusion**	1	**0.1741 ±0.0250**	**0.0807 ±0.0070**	**0.0678 ±0.0059**	**0.1207 ±0.0020**
ETTh	DiffWave (adapted)	3	2.0856 ±0.3756	0.2332 ±0.0496	0.2865 ±0.0180	0.1475 ±0.0077
ETTh	DiffTime (local)	3	3.4698 ±0.1983	0.3991 ±0.0153	0.4386 ±0.0049	0.1320 ±0.0004
MuJoCo	**StrideDiffusion**	1	**0.0672 ±0.0253**	**0.3851 ±0.0092**	**0.0280 ±0.0075**	**0.0067 ±0.0002**
MuJoCo	DiffWave (adapted)	3	2.5560 ±0.3607	0.7823 ±0.1255	0.3743 ±0.0289	0.0128 ±0.0005
MuJoCo	DiffTime (local)	3	2.8433 ±0.0164	1.0291 ±0.0014	0.4719 ±0.0034	0.0406 ±0.0003
Energy	**StrideDiffusion**	1	**0.1044 ±0.0129**	**1.3657 ±0.1254**	**0.1119 ±0.0088**	**0.2522 ±0.0004**
Energy	DiffWave (adapted)	3	4.9638 ±0.3706	5.9301 ±0.4320	0.4888 ±0.0074	0.2528 ±0.0001
Energy	DiffTime (local)	3	8.6419 ±0.1411	10.1629 ±0.0195	0.4930 ±0.0012	0.2535 ±0.0008
fMRI	**StrideDiffusion**	1	**0.1466 ±0.0028**	**1.5569 ±0.0097**	**0.0791 ±0.0086**	**0.1003 ±0.0003**
fMRI	DiffWave (adapted)	3	3.7188 ±0.1278	2.8907 ±0.2516	0.4869 ±0.0212	0.1018 ±0.0016
fMRI	DiffTime (local)	3	1.2474 ±0.0359	5.6415 ±0.0285	0.4899 ±0.0163	0.1036 ±0.0001

Table 43: Sampling cost and clipping for Table 42. Baselines average three generation seeds per model, then report mean and sample SD across three trained models. Ours reports three generation seeds paired with baseline label 411, from one frozen model per dataset. Its SD measures generation variation, not retraining uncertainty. NFE is the mean number of sequential denoiser calls per batch trajectory, not total calls for all samples. Baselines use 50-step ancestral sampling. Ours is adaptive, so budgets are not matched. Seconds cover all 2,048 samples at batch size 64. Each run takes the median of three synchronized repeats on an exclusive H200. Timing includes denoising, updates and FFT control. It excludes training, loading, three denoiser-forward warmups, initial-noise preparation, host/device transfers, the final transform, saving and metrics. Final clip is the percentage of values outside [-1,1] before the shared final transform and scoring. It does not count DiffWave intermediate-step clipping. Blue shading identifies Ours. Bold and underline mark the lowest and second-lowest distinct displayed means within each dataset, including ties. These are descriptive ranks under unequal training histories, not significance tests. Only sampling time is ranked. NFE and clipping are not quality rankings.

Data	Method	NFE	Seconds	Final clip (%)
Sines	**StrideDiffusion**	34.8854 ±0.1985	5.6117 ±0.0414	0.0000 ±0.0000
Sines	DiffWave (adapted)	50.0000 ±0.0000	5.2166 ±0.0504	0.0000 ±0.0000
Sines	DiffTime (local)	50.0000 ±0.0000	**4.2069 ±0.0230**	1.6419 ±1.2193
Stocks	**StrideDiffusion**	34.6042 ±0.0955	5.9852 ±0.0684	0.0000 ±0.0000
Stocks	DiffWave (adapted)	50.0000 ±0.0000	5.1743 ±0.0622	0.0000 ±0.0000
Stocks	DiffTime (local)	50.0000 ±0.0000	**4.4092 ±0.0287**	2.8246 ±3.4177
ETTh	**StrideDiffusion**	14.1771 ±0.4964	**2.6947 ±0.0811**	0.0000 ±0.0000
ETTh	DiffWave (adapted)	50.0000 ±0.0000	5.2012 ±0.0559	0.0000 ±0.0000
ETTh	DiffTime (local)	50.0000 ±0.0000	4.6747 ±0.0338	0.0208 ±0.0124
MuJoCo	**StrideDiffusion**	63.9479 ±0.2658	11.9615 ±0.1968	0.0000 ±0.0000
MuJoCo	DiffWave (adapted)	50.0000 ±0.0000	**5.1609 ±0.0966**	0.0000 ±0.0000
MuJoCo	DiffTime (local)	50.0000 ±0.0000	6.2815 ±0.0068	0.0163 ±0.0019
Energy	**StrideDiffusion**	58.8854 ±0.1985	15.0950 ±0.1034	0.0000 ±0.0000
Energy	DiffWave (adapted)	50.0000 ±0.0000	**5.1759 ±0.0499**	0.0000 ±0.0000
Energy	DiffTime (local)	50.0000 ±0.0000	9.8137 ±0.0373	0.7371 ±0.5416
fMRI	**StrideDiffusion**	66.7292 ±0.7142	19.8953 ±0.1844	0.0000 ±0.0000
fMRI	DiffWave (adapted)	50.0000 ±0.0000	**5.1768 ±0.0862**	0.0000 ±0.0000
fMRI	DiffTime (local)	50.0000 ±0.0000	15.6002 ±0.0181	0.0038 ±0.0011

Table 44: Paired comparisons for the complete generators and training budgets in Table 42. Metric differences are Ours minus baseline, so negative values favor Ours. Time ratio is baseline time divided by Ours time, so values above 1 favor Ours. Within each baseline training group, we average three paired generation seeds, then report mean and sample SD across three groups. All groups share one Ours checkpoint per dataset, so SD is not Ours retraining uncertainty. All quality scores use final $[0, 1]$-clipped samples. Subset Corr. draws 682 of 2,048 windows with replacement per evaluator seed. Subset Corr., Disc. and Pred. average three evaluator seeds. Each dataset/training-label/generation block shares one TS2Vec encoder fitted on 2,000 real reference windows. Training costs and NFE differ. These paired differences do not isolate the scheduler. Blue cells favor Ours. Orange cells favor the baseline. Colors follow the mean only, not statistical significance. Quality columns have a zero reference, while the time-ratio reference is one. Signed differences and time ratios are not ranked across datasets or controls.

A. Ours versus DiffWave (adapted)

	Quality difference (↓, reference 0)				
Data	Δ C-FID	Δ Corr.	Δ Disc.	Δ Pred.	Time ratio (↑, reference 1)
Sines	-2.0814 ± 0.2674	-0.0623 ± 0.0291	-0.4001 ± 0.0015	-0.0691 ± 0.0169	0.9160 ± 0.0229
Stocks	-4.9879 ± 0.4088	-0.0326 ± 0.0090	-0.3049 ± 0.0190	-0.0044 ± 0.0007	0.8586 ± 0.0037
ETTh	-1.9114 ± 0.3756	-0.1525 ± 0.0496	-0.2097 ± 0.0201	-0.0276 ± 0.0084	1.9446 ± 0.0322
MuJoCo	-2.4888 ± 0.3607	-0.3971 ± 0.1255	-0.3477 ± 0.0315	-0.0056 ± 0.0001	0.4253 ± 0.0080
Energy	-4.8594 ± 0.3706	-4.5644 ± 0.4320	-0.3748 ± 0.0087	-0.0006 ± 0.0002	0.3439 ± 0.0084
fMRI	-3.5720 ± 0.1279	-1.3338 ± 0.2516	-0.4089 ± 0.0239	-0.0015 ± 0.0016	0.2536 ± 0.0025

B. Ours versus DiffTime (local)

	Quality difference (↓, reference 0)				
Data	Δ C-FID	Δ Corr.	Δ Disc.	Δ Pred.	Time ratio (↑, reference 1)
Sines	-0.0050 ± 0.0253	-0.0175 ± 0.0169	-0.0030 ± 0.0409	0.0003 ± 0.0014	0.7386 ± 0.0082
Stocks	-1.1456 ± 1.2115	0.0179 ± 0.0022	-0.1219 ± 0.0634	-0.0023 ± 0.0031	0.7317 ± 0.0101
ETTh	-3.2956 ± 0.1983	-0.3185 ± 0.0153	-0.3617 ± 0.0101	-0.0122 ± 0.0011	1.7477 ± 0.0239
MuJoCo	-2.7761 ± 0.0164	-0.6440 ± 0.0014	-0.4453 ± 0.0025	-0.0335 ± 0.0008	0.5177 ± 0.0072
Energy	-8.5375 ± 0.1411	-8.7972 ± 0.0195	-0.3790 ± 0.0026	-0.0012 ± 0.0009	0.6522 ± 0.0182
fMRI	-1.1006 ± 0.0358	-4.0846 ± 0.0285	-0.4119 ± 0.0242	-0.0033 ± 0.0001	0.7646 ± 0.0180

K HISTORICAL REFERENCE RESULTS AND COMPONENT DIAGNOSTICS

The results in this section use separate historical protocols. They are not pooled with the controlled H200 experiments. Table 45 reports the Diffusion-TS baseline records and published DiffWave/DiffTime values without a cross-protocol ranking. Full Diffusion-TS uses ancestral DDPM with T calls, Fast uses deterministic DDIM with T calls, and Fast-200 uses 200 calls. Diffusion-TS quality aggregates five runs with seeds 12345, 2026, 2027, 2028, and 2029. Each seed is used for training and generation, with a separate checkpoint for that seed. We report the mean and sample standard deviation of the per-run metric means, so this variation includes both training and generation randomness. We omit historical sampling times because their provenance could not be verified. DiffWave and DiffTime quality values reproduce Table 1 of Yuan & Qiao (2024). They are published reference values, not local reruns under the other rows' protocol.

Table 45: Historical quality values under separate protocols. Diffusion-TS reports mean and sample standard deviation across five runs, each with its own training seed and checkpoint and the same seed reused for generation. DiffWave and DiffTime reproduce Table 1 of Yuan & Qiao (2024). These entries do not form a matched-protocol ranking. Historical timing is omitted because the available records do not establish its provenance. No StrideDiffusion result is included in this historical reference table. We do not highlight cross-protocol winners. Gray highlights source labels only, not better scores.

Metric	Baseline / source	Sines	Stocks	ETTh	MuJoCo	Energy	fMRI
Context-FID Score (↓)	Diffusion-TS	0.011 ±.002	0.182 ±.026	0.136 ±.009	0.016 ±.002	0.102 ±.011	0.109 ±.005
	Diffusion-TS (Fast)	0.016 ±.001	0.195 ±.032	0.227 ±.005	0.031 ±.003	0.091 ±.006	0.114 ±.003
	Diffusion-TS (Fast-200)	0.016 ±.001	0.195 ±.025	0.234 ±.011	0.030 ±.002	0.100 ±.014	0.132 ±.002
	DiffWave (published)	0.014 ±.002	0.232 ±.032	0.873 ±.061	0.393 ±.041	1.031 ±.131	0.244 ±.018
	DiffTime (published)	0.006 ±.001	0.236 ±.074	0.299 ±.044	0.188 ±.028	0.279 ±.045	0.340 ±.015
Correlational Score (↓)	Diffusion-TS	0.018 ±.004	0.009 ±.003	0.054 ±.003	0.209 ±.016	0.943 ±.114	1.283 ±.096
	Diffusion-TS (Fast)	0.019 ±.003	0.011 ±.005	0.068 ±.003	0.199 ±.015	0.921 ±.076	1.107 ±.055
	Diffusion-TS (Fast-200)	0.017 ±.003	0.011 ±.005	0.072 ±.006	0.204 ±.012	0.909 ±.046	1.099 ±.066
	DiffWave (published)	0.022 ±.005	0.030 ±.020	0.175 ±.006	0.579 ±.018	5.001 ±.154	3.927 ±.049
	DiffTime (published)	0.017 ±.004	0.006 ±.002	0.067 ±.005	0.218 ±.031	1.158 ±.095	1.501 ±.048
Discriminative Score (↓)	Diffusion-TS	0.011 ±.008	0.080 ±.018	0.066 ±.006	0.012 ±.004	0.122 ±.006	0.153 ±.012
	Diffusion-TS (Fast)	0.018 ±.006	0.109 ±.014	0.095 ±.006	0.027 ±.006	0.142 ±.007	0.116 ±.012
	Diffusion-TS (Fast-200)	0.017 ±.007	0.109 ±.016	0.094 ±.006	0.027 ±.005	0.147 ±.003	0.128 ±.010
	DiffWave (published)	0.017 ±.008	0.232 ±.061	0.190 ±.008	0.203 ±.096	0.493 ±.004	0.402 ±.029
	DiffTime (published)	0.013 ±.006	0.097 ±.016	0.100 ±.007	0.154 ±.045	0.445 ±.004	0.245 ±.051
Predictive Score (↓)	Diffusion-TS	0.093 ±.000	0.037 ±.000	0.121 ±.002	0.008 ±.000	0.251 ±.000	0.100 ±.000
	Diffusion-TS (Fast)	0.094 ±.001	0.037 ±.000	0.122 ±.002	0.007 ±.000	0.251 ±.000	0.099 ±.000
	Diffusion-TS (Fast-200)	0.094 ±.000	0.038 ±.000	0.122 ±.002	0.007 ±.000	0.251 ±.000	0.099 ±.000
	DiffWave (published)	0.093 ±.000	0.047 ±.000	0.130 ±.001	0.013 ±.000	0.251 ±.000	0.101 ±.000
	DiffTime (published)	0.093 ±.000	0.038 ±.001	0.121 ±.004	0.010 ±.001	0.252 ±.000	0.100 ±.000

K.1 COMPONENT ABLATIONS

Table 46 reports the original component ablations. The recovered records provide the displayed metrics but do not identify the GPU. Removing a component can change NFE and the sampled

trajectory, so these are descriptive ablations, not matched-NFE attribution tests. Their timing is not combined with the controlled H200 results.

Table 46: Original component ablations under a separate protocol. The recovered records do not identify the GPU. Time is sampling latency in seconds. Speedup is relative to the full DDPM row within each dataset. Lower C-FID and Discriminative score are better. These are not matched-NFE comparisons. Blue identifies the complete method. Bold marks the best displayed value within each row across the listed, unequal budgets, not statistical significance. Underline marks the second-best distinct displayed value in each row, including ties. Speedup is higher-is-better, all other ranked rows are lower-is-better. No significance claim is implied.

Dataset	Metric	**Ours**	Full DDPM	No gate	No final fine steps	w/o τ_{energy}
SINES	Time (s) ↓	9.30	130.32	**3.14**	6.17	38.76
	Speedup ↑	14.0×	1.0×	**41.5×**	21.1×	3.4×
	C-FID ↓	**0.0086**	0.0161	0.3263	0.0099	0.0103
	Disc ↓	0.0129	0.0110	0.1182	**0.0104**	0.0365
STOCKS	Time (s) ↓	5.94	78.53	**2.63**	5.90	26.58
	Speedup ↑	13.2×	1.0×	**29.9×**	13.3×	3.0×
	C-FID ↓	0.1275	0.2398	0.4064	**0.1270**	0.1959
	Disc ↓	0.0888	0.1112	0.1108	**0.0760**	0.0903
ENERGY	Time (s) ↓	56.46	1178.48	**28.01**	49.45	115.24
	Speedup ↑	20.9×	1.0×	**42.1×**	23.8×	10.2×
	C-FID ↓	0.0883	0.0817	0.1231	**0.0813**	0.0942
	Disc ↓	0.1233	0.1421	0.2629	0.1362	**0.1226**
FMRI	Time (s) ↓	44.41	638.29	47.28	**39.16**	68.58
	Speedup ↑	14.4×	1.0×	13.5×	**16.3×**	9.3×
	C-FID ↓	**0.0937**	0.1154	**0.0937**	0.1039	0.1548
	Disc ↓	0.1021	0.1632	**0.0981**	0.1636	0.1874

K.2 Controller decisions

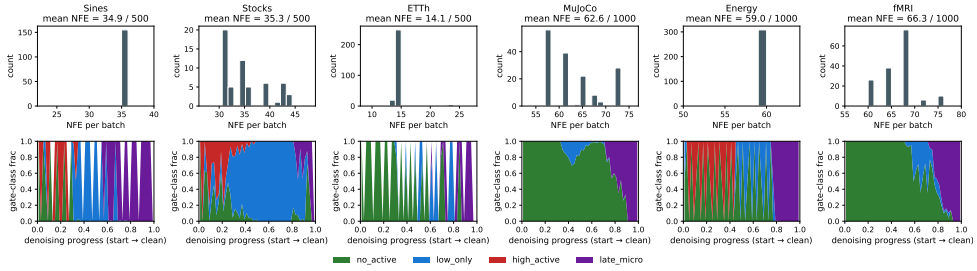

Figure 18: Original spectral-controller behavior. Top: NFE per sampling batch, with a shared grid within each batch. Sines, Stocks, ETTh, MuJoCo, Energy, and fMRI contain 157, 58, 272, 157, 309, and 156 batches, respectively. Bottom: gate-class fractions pooled into 50 bins of normalized iteration position i/N, not diffusion time. This is not an additional timing comparison.

L Archived conditional examples

Figures 19 and 20 retain the original selected conditional-generation examples at length 48. The original plots and legends are unchanged, with the banded sampler in red. These are separate from the controlled unconditional experiments.

The original prediction arrays and exact checkpoint/export lineage have not been recovered. The surviving plotting code selects cases using prediction errors, but its version at export is unknown.

Random selection has not been established. These are archived illustrations, not a reproducible comparison of current checkpoints. We do not infer aggregate quality or speed gains from them.

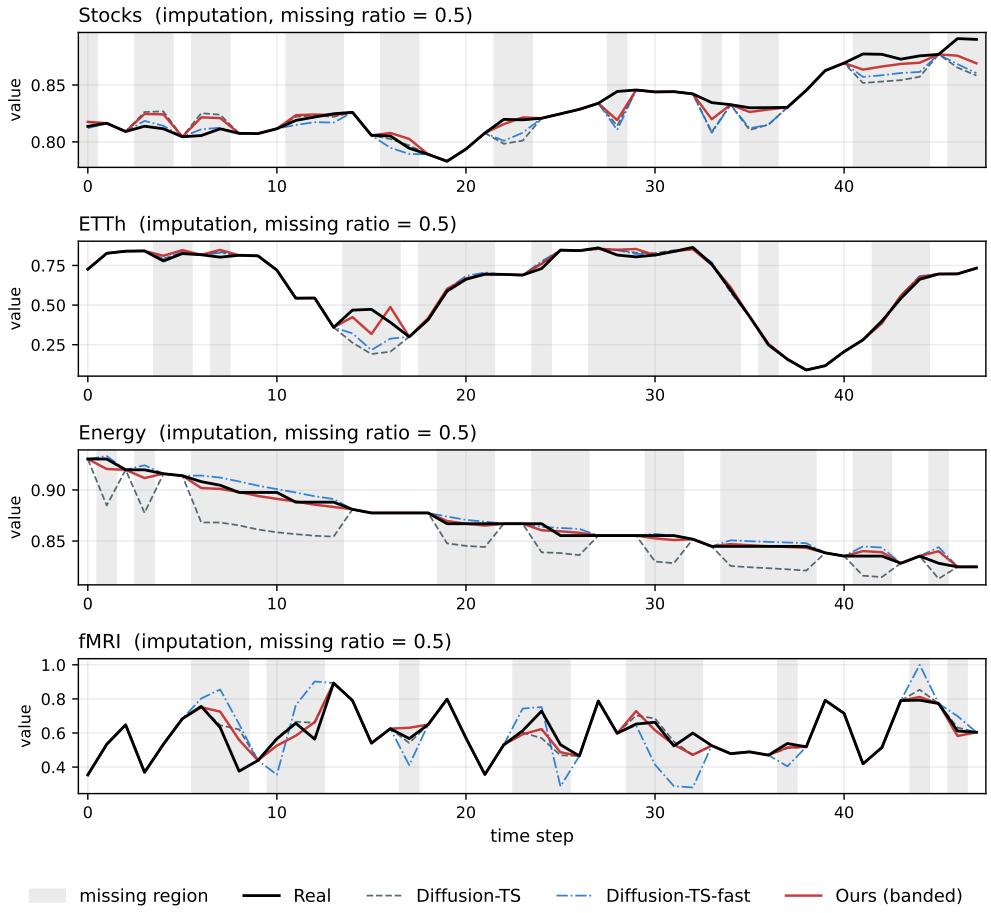

Figure 19: Archived imputation examples with nominal missing ratio 0.5 and sequence length 48. Rows show Stocks, ETTh, Energy, and fMRI. Black is the full reference, gray dashed is Diffusion-TS, blue dash–dot is Diffusion-TS-fast, and red is the original banded sampler labeled Ours. Shading marks held-out positions (mask 0). Unshaded reference values are observed inputs. The original axes and legends are retained. These selected single-channel trajectories have no uncertainty estimates and are not used for quantitative rankings because the original prediction/checkpoint lineage is incomplete.

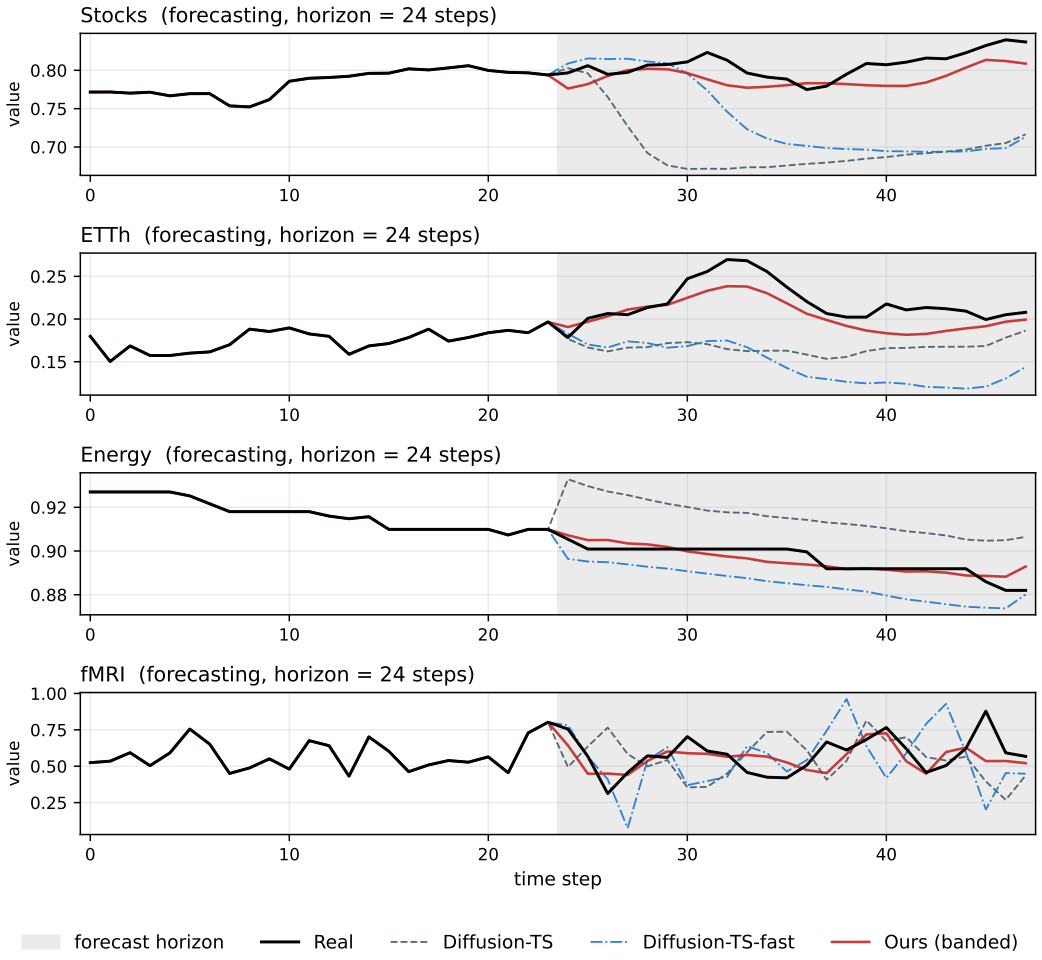

Figure 20: Archived forecasting examples with 24 observed and 24 future steps. Rows and method colors follow Figure 19. Shading marks the forecast horizon. These plots illustrate individual trajectories, including departures from the reference. They do not establish average forecasting accuracy or transfer of the current method. The same incomplete prediction/checkpoint lineage applies.