SUPPORT IS FREE, BANDWIDTH IS NOT: BASIS-PARAMETERISED CONVOLUTIONS FOR GENERATIVE EEG

Anonymous authors
Paper under double-blind review

ABSTRACT

Functional convolutions write a kernel as q coefficients over a fixed basis sampled on an R-point grid, so their parameter count does not depend on the support R. We show that the ratio R/q buying that free support is also a *band-limit*: a Fourier-basis kernel cannot represent frequencies above $f_s/(2R/q)$, and Legendre and B-spline kernels follow the same R/q scaling (slope 1.00–1.04, $R^2 \geq 0.997$) with basis-specific constants. We test what this costs on raw four-channel consumer EEG, with a flow-matching model whose velocity predictor is built from these kernels and held at a 64–128 Hz cutoff against a 40 Hz signal band. Across six checkpoints per model its samples carry 1.21 ± 0.60 times real beta power (13–30 Hz) and 1.63 ± 0.95 times real gamma (30–40 Hz), against 0.052 and 0.020 for a functional VAE on the same substrate, with no overlap in either band ($p = 0.00108$). A discrete-kernel twin of the velocity predictor is not distinguishable from it in either band over three paired seeds, so this fidelity is not what basis parameterisation buys. What it buys is an analysable band-limit, and spending that headroom on longer support raises held-out loss monotonically at every basis budget while parameter count explains none of it ($R^2 = 0.004$). A linear aggregate 0–40 Hz spectral error ranks the VAE better on the same twelve checkpoints and does not separate the two models ($p = 0.91$), because two bins below 0.5 Hz account for at least 94% of it. Over 30 seeds, a gaze regressor trained only on the flow model's samples at $5\times$ real volume shows no confirmed degradation against one trained on real data, while the VAE's samples drive it to zero correlation. A 1 Hz high-pass shows that this task reads sub-1 Hz ocular signal. Support is free in parameters; it is paid for in bandwidth.

1 INTRODUCTION

A convolutional kernel's receptive field and its parameter count are, in the standard parameterisation, the same number: a layer with support R over C_{in} input and C_{out} output channels stores $R \cdot C_{\text{in}} \cdot C_{\text{out}}$ free weights, so context is bought at a fixed exchange rate. Basis-parameterised functional convolutions (Heinrichs et al., 2023) remove that rate: writing the kernel as $w(t) = \sum_{i<q} w^i \varphi_i(t)$ over q fixed analytic basis functions evaluated on an R-point grid makes the learned tensor $(q, C_{\text{in}}, C_{\text{out}})$, in which R does not appear. Formally $\partial P/\partial R = 0$: support becomes free with respect to parameter count, and the affordable ratio is exactly R/q.

That is an arithmetic fact. This paper asks what the decoupling costs when it is spent, and what a generative model built from these kernels does when it is not. We build a flow-matching generative model for raw EEG whose velocity predictor is assembled entirely from these kernels. Its samples keep the top of the spectrum, where power is smallest and where a variational autoencoder on the same substrate and the variational autoencoding neural operator (VANO) of Seidman et al. (2023) over-smooth. The ratio R/q that makes support cheap is also a **band-limit** on the kernel, with a cutoff at $f_s/(2R/q)$ for the Fourier basis we use and at a basis-specific constant times that ratio elsewhere. Our configuration keeps that cutoff clear of the signal band, a necessary condition for the velocity field to represent the high frequencies the data carries. Spending the same headroom on cheap receptive field removes them from the reachable set, where no schedule or initialisation recovers them.

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.

Why raw EEG is the setting that decides this. The domain makes the trade quantitative. Raw consumer EEG is sampled at 256 Hz with its information below roughly 40 Hz, so the headroom between the kernel cutoff and the signal band is finite and computable: $f_s/(2R/q)$ drops beneath the band edge as soon as $R/q > 3.2$. Change the sampling rate or the band and the arithmetic moves that threshold, not the conclusion. The domain also supplies a second model to measure against, the *same* substrate under a variational autoencoder, and the pair shows that a linear aggregate spectral error ranks the two models backwards.

Related work. Decoupling kernel extent from parameter count is the premise of several lines of work: continuous kernels parameterised by small networks (Romero et al., 2022b;a), state-space kernels (Gu et al., 2022), and filters expanded over a truncated fixed basis (Qiu et al., 2018; Heinrichs et al., 2023). That truncation limits bandwidth is also known: the Fourier neural operator caps frequency by its number of retained modes (Li et al., 2021), and FlexConv controls kernel bandwidth analytically to avoid aliasing at unseen resolutions (Romero et al., 2022a). For a Fourier basis the cutoff $f_s/(2R/q)$ follows from the harmonics a q-term expansion contains. What we add is its scaling across non-harmonic bases and a controlled test of what spending it costs in a trained generative model. Function-space generative models include VANO (Seidman et al., 2023), functional diffusion processes (FDP; Franzese et al., 2023) and functional flow matching (Kerrigan et al., 2024); kernels supported on a fixed number of grid points, as here, are not discretisation-invariant in the neural-operator sense (Kovachki et al., 2024; Liu-Schiaffini et al., 2024). EEG synthesis has used GANs (Hartmann et al., 2018) and diffusion (Tosato et al., 2023), mostly for augmentation (Rommel et al., 2022), and gaze decoding from EEG has its own benchmark (Kastrati et al., 2021).

What we do, and what we find. We build two generative models on one basis-parameterised substrate, then sweep kernel support R and basis count q independently at a fixed training budget, with the unconstrained-kernel comparator *measured* rather than extrapolated. Four findings follow. *First*, across six independently trained checkpoints per model, the flow-matching model keeps the beta and gamma power that a functional VAE on the same substrate loses. The aggregate 0–40 Hz linear spectral error ranks the two the other way and cannot separate them, because the two bins below 0.5 Hz decide at least 94% of it (§3.1). A discrete-kernel twin of the velocity predictor is not distinguishable from the functional one in either band, so this fidelity comes from the flow model rather than from basis parameterisation (§3.2). *Second*, basis parameterisation is a band-limit: a kernel written as q coefficients on an R-point grid is confined to a q-dimensional subspace whatever the optimiser does, with a cutoff at $f_s/(2R/q)$ that projection alone reveals and that reappears in the velocity field a trained network fits (§3.4). *Third*, the decoupling is real arithmetically and unprofitable here: parameter count stays fixed as support grows while the unconstrained twin grows by an order of magnitude, and held-out loss degrades monotonically, governed by R/q and not by parameter count (§3.5). *Fourth*, the trained functional model runs at window lengths and sampling rates absent from training (§3.6). We show this qualitatively, and we have not tested whether the discrete twin, which is also fully convolutional, can do the same.

How the two results connect. The quantity that buys support is the quantity that buys bandwidth, so the configuration that keeps high frequencies representable and the configuration that is parameter-efficient are the same dial at opposite ends. Scope: single-corpus results on four-channel consumer EEG at a fixed 20,000-step budget for the sweep, with sampler instability we quantify (§4).

2 METHODS

2.1 DATA AND PREPROCESSING

All experiments use a publicly available EEG-eye-tracking corpus collected with the Muse headband (Afonso & Heinrichs, 2025): four channels at 256 Hz with a synchronised gaze trace. Generative models consume 4-second windows of shape $(N, T{=}1024, C{=}4)$, with the conditioning stimulus at $(N, 1024, 2)$ and gaze in $[-1, 1]$ after temporal pre-alignment. Preprocessing is intentionally light, 50/60 Hz notch filtering and global z-scoring with no bandpass, which preserves the broadband statistics that make the top of the spectrum available to be lost. We use the corpus's smooth-pursuit recordings with its subject-level split: 12 held-out test subjects give 380 non-overlapping test windows.

2.2 SHARED FUNCTIONAL SUBSTRATE

The substrate rests on three pieces: a functional convolution, a residual block built from it, and a resolution-invariant resampling operator.

Basis-parameterised functional convolution. Following Heinrichs et al. (2023), a functional convolutional layer maps an input function $H_j^{\ell-1}(t)$ for input channel j to an output function $H_k^{\ell}(t)$ for output channel k by shift-invariant convolution against a learned kernel function,

$$H_k^{\ell}(t) = \sigma\left(\sum_{j=1}^{J_{\ell-1}} \int_0^{R/f_s} w_{j,k}^{\ell}(s)\, H_j^{\ell-1}(t-s)\,\mathrm{d}s\right), \tag{1}$$

where the integration runs over the kernel's compact support of R grid points, i.e. R/f_s seconds at sampling rate f_s, and the kernel $w_{j,k}^{\ell}$ is expanded over a fixed basis $\{\varphi_i\}_{i=0}^{q-1}$ with scalar weights $w_{j,k}^{\ell,i}$:

$$w_{j,k}^{\ell}(t) \;=\; \sum_{i=0}^{q-1} w_{j,k}^{\ell,i}\, \varphi_i(t). \tag{2}$$

The canonical models use a Fourier basis, whose first element is the constant $\varphi_0 \equiv 1$; this supplies the per-filter offset, so the layer carries no separate bias term and the q coefficients of (2) are the layer's entire parameter budget. With only discrete samples available, (1) is evaluated by quadrature on the R-point kernel grid, making it a same-padded R-tap convolution whose taps are $w_{j,k}^{\ell}$ sampled on that grid; the basis being fixed and analytic, the basis matrix and quadrature weights are pre-computed once per resolution and the layer stays differentiable in the coefficients. The functional convolution is the only learned operator in the substrate.

FuncResBlock and spectral resampling. The FuncResBlock instantiates the standard ResNet shape (He et al., 2016) entirely in the functional domain: two functional convolutions with non-linearity in between, plus a residual skip. All intermediate activations remain function-valued; no operation collapses to a discrete vector. Resolution changes are performed in Fourier space (Li et al., 2021): downsampling truncates high-frequency modes, upsampling pads them with zeros and inverts (equivalent to sinc interpolation in the time domain). Because the basis can be sampled on any grid, a model trained at $T{=}1024$ can be evaluated at a different sampling rate or window length without retraining (§3.6). Per-layer diagrams of this block and of both models are in supplementary §B.

Parameter cost, and the decoupling this paper is about. Write q for the number of basis functions and R for the *support*: the number of grid points on which the kernel is evaluated, i.e. the receptive field of one layer in samples. A discrete convolution stores one free weight per tap, so a layer with C_{in} input and C_{out} output channels costs $P_{\text{disc}} = R\,C_{\text{in}}C_{\text{out}}$, and its receptive field and its parameter count are, up to the constant $C_{\text{in}}C_{\text{out}}$, *the same number*. Under (2) the learned quantity is instead the coefficient tensor $w_{j,k}^{\ell,i}$ of shape $(q, C_{\text{in}}, C_{\text{out}})$, so $P_{\text{func}} = q\,C_{\text{in}}C_{\text{out}}$ and $P_{\text{disc}}/P_{\text{func}} = R/q$. The support R does not appear in P_{func}. Formally $\partial P_{\text{func}}/\partial R = 0$: **kernel support is free with respect to parameter count**, and the affordable ratio is exactly R/q.

What the decoupling costs. Two things. *Expressivity*: the kernel is confined to the q-dimensional subspace $\mathrm{span}\{\varphi_0, \ldots, \varphi_{q-1}\} \subset \mathbb{R}^R$, so raising R at fixed q does not merely lengthen the kernel, it *dilates* it, becoming longer and smoother together (§3.4). *Compute*: a convolution costs $\mathcal{O}(TRC_{\text{in}}C_{\text{out}})$ however its kernel is parameterised, so the saving is in parameter count and not arithmetic, at an exchange rate of again R/q. One consequence matters for reading the results: in the released FuncFlow configuration R/q is 2 in the narrow outer layers and exactly 1 in the wide deep ones where the parameters are, so the whole network is only $1.13\times$ smaller than its unconstrained twin and 91.1% of its parameters sit in uncompressed kernels. The canonical model therefore says little about basis parameterisation, which §3.4–§3.5 vary directly (supplementary §D).

2.3 TWO GENERATIVE MODELS ON ONE SUBSTRATE

We instantiate the substrate twice, under different generative frameworks. The pair is the evidence for §3.1: holding the substrate fixed and changing the framework moves high-band fidelity by more than

an order of magnitude, which is what makes the spectrum, rather than the substrate alone, the thing to measure.

FuncFlow. FuncFlow places flow matching (Kerrigan et al., 2024) on top of the substrate. The reference distribution is a Gaussian process with the Matérn-$\frac{1}{2}$ (Ornstein–Uhlenbeck) kernel, $k(t,t') = \sigma^2 \exp(-|t-t'|/\ell)$ with $\ell = 0.01$, $\sigma^2 = 0.1$, chosen so that samples $x_0 \sim \mathrm{GP}(0,k)$ resemble the local correlation structure of raw EEG, and source–target pairs are joined by optimal-transport coupling. The velocity predictor v_θ is trained with the standard flow-matching loss $\mathcal{L}_{\mathrm{FM}} = \mathbb{E}_{t,x_0,x_1}\|v_\theta(x_t,t) - (x_1 - x_0)\|^2$ on the linear interpolant $x_t = (1-t)x_0 + tx_1$, and generation integrates $\mathrm{d}x_t/\mathrm{d}t = v_\theta(x_t,t)$ from $x_0 \sim \mathrm{GP}$ by Heun's method with $N = 50$ steps. v_θ is a U-Net (Ronneberger et al., 2015) assembled entirely from FuncResBlocks, with skips between matching resolution levels and Fourier-space resampling between them, and no attention layers. The gaze stimulus is injected at each level through a conditioning encoder built from the same functional layers, with 10% conditioning dropout (supplementary §B). The model has $3{,}809{,}488$ parameters and trains for at most $100{,}000$ steps, stopping early (between steps $42{,}000$ and $62{,}000$ for the six seeds) when a generation metric computed on the test windows stops improving. We score the final weights, so no checkpoint is chosen on test data, although the stopping step is.

FuncVAE. FuncVAE puts the same FuncResBlock backbone under a variational autoencoder, encoding $(1024, 4) \to (64, 256)$ by spectral downsampling to a functional bottleneck $z(t) \in \mathbb{R}^{64\times 8}$ regularised against a GP-Matérn prior with $\ell = 1.0$, with a free-bits floor against posterior collapse. The decoder takes the same gaze trace through the same FuncCondEncoder and dropout as FuncFlow, and generation draws z from the GP prior and decodes it *against a real stimulus trace*, so synthetic windows carry the gaze target they were generated for. It trains on the same split, preprocessing and optimiser settings, stops early on test loss between steps $60{,}000$ and $83{,}500$, and is likewise scored at its final weights (supplementary §B).

What this comparison does and does not isolate. Substrate, conditioning, data, schedule and optimiser are shared, but the models are not otherwise matched and we do not call this a clean framework ablation. Three structural differences push the same way as the objective, the largest being the bottleneck: 64 points over a $4\,\mathrm{s}$ window imposes an $8\,\mathrm{Hz}$ grid Nyquist on the latent path, which is most of the band §3.1 measures. §3.3 tests it by widening the bottleneck.

2.4 Evaluation

Spectral measurement. We score generated EEG by its spectral content on a log scale, which normalises for the large differences in absolute power between bands; this is the most faithful scoring we found. Model comparisons use the **band power ratio**, generated over real mean power in delta (0.5–4), theta (4–8), alpha (8–13), beta (13–30) and gamma (30–40 Hz), and the scale-free **log-spectral error**, the mean $|\Delta \log_{10} P|$ over each band's bins; they use $1{,}024$ generated windows per checkpoint against the 380 real test windows, for **six independently trained checkpoints per model**. Sweep cells are summarised by a single **log-spectral distance**, the mean squared $\log_{10}$ difference over the 0–40 Hz bins of the *median* spectrum across generated windows, which resists the occasional runaway sample; squared relative error on the *mean* PSD cannot here rank a trained model against an untrained one (supplementary §A). Every quality number comes with the **divergence rate**, the fraction of generated windows whose standard deviation exceeds five times the median real one. We also report the **flow loss** on the held-out split, which the evaluation's fixed RNG makes *paired* and by an order of magnitude our lowest-variance signal.

Downstream substitution. The downstream task is gaze regression from a four-channel EEG window with the spatial-filter CNN regressor of Afonso & Heinrichs (2025), framed as a **data-quality test**: the head is trained from scratch on a single source of EEG, real *or* synthetic and never mixed, then evaluated on the real held-out subjects. A synthetic set at $k\times$ volume is k independent generator draws *per real training window*, each decoded against that window's own gaze trace, so the stimulus distribution is held fixed across ratios. Synthetic sets come from each model's seed-0 run at the checkpoint its trainer kept as best on the test-split criterion above; we did not regenerate them at the final weights. Statistics: 30 head-training seeds per condition, paired by seed with the real baseline; one-sided paired t-tests for inferiority, Holm–Bonferroni at $\alpha = 0.025$ over $k = 9$ comparisons per model (3 ratios $\times$ 3 metrics: mean Euclidean gaze error (MED) in mm, and the Pearson correlations

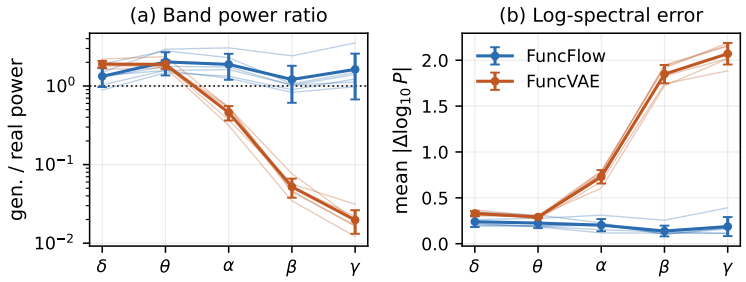

Figure 1: Where each model agrees with real EEG, across **six independently trained checkpoints per model** at their final weights (mean ± s.d., every seed overplotted faintly). **(a)** Generated over real band power; 1.0 is exact, below 1.0 is over-smoothing. **(b)** Log-spectral error: FuncFlow's stays below 0.4 in every band while the VAE's rises sevenfold from theta to gamma.

r_x, r_y), with the composite r_{mean} descriptive and outside the hypothesis family (supplementary §F, §E).

Efficiency protocol. Every cell of the study in §3.4–§3.5 retrains the FuncFlow architecture from scratch for exactly 20,000 steps with early stopping disabled and scores the final checkpoint, so that quality-per-parameter is not confounded with quality-per-step, and passes through a single driver script, so only the named factor varies. Kernel support is clamped per level to that level's time-axis length and every clamp is recorded. The two cells carrying a headline claim use **five seeds per arm** and sweep cells use three; single-seed cells are marked in supplementary Table A5 and we rest no claim on them.

3 Results

3.1 The flow model keeps the high band, and an aggregate metric cannot see it

Built entirely from basis-parameterised convolutions, FuncFlow generates raw EEG that keeps the top of the spectrum, the part of the band the other function-space generators we tested either lose or overshoot. §3.2 adds a discrete-kernel baseline. We judge on generated samples throughout and do not rank these models by reconstruction, which measures a different quantity for a bottleneck model than for a flow (supplementary §B).

FuncFlow keeps the high band; the VAE on the same substrate loses it (Figure 1). Generated-to-real band power for FuncFlow is 1.21 ± 0.60 in beta (13–30 Hz) and 1.63 ± 0.95 in gamma (30–40 Hz), against 0.052 ± 0.014 and 0.020 ± 0.007 for FuncVAE, which keeps one part in fifty of real gamma power. Separation across the twelve checkpoints is perfect in both bands: FuncFlow's lowest seed (beta 0.83, gamma 0.98) still exceeds the VAE's highest (beta 0.077, gamma 0.031), an exact one-sided $p = 0.00108$ and the floor of a six-against-six design. FuncFlow errs by excess rather than deficit, over-producing gamma by up to $3.5\times$ in one seed and theta and alpha by about $2\times$ on average (supplementary Table A3). How much of the VAE's loss belongs to the objective and how much to the 8 Hz Nyquist ceiling of its 64-point bottleneck is what §3.3 answers. The log-spectral error agrees: FuncFlow's stays below 0.4 in every band of every seed while the VAE's reaches 1.85 ± 0.10 in beta and 2.07 ± 0.12 in gamma.

Both prior function-space generators fail here, in opposite directions. On the same instrument, VANO over-smooths further than FuncVAE does (0.031 ± 0.004 of real beta, 0.008 ± 0.001 of gamma over six seeds) while FDP, with a single trained seed, overshoots at $352\times$ real beta power and a divergence rate of 1.000 (supplementary §B).

A linear aggregate metric cannot make this distinction. The natural single-number summary of spectral fidelity is a squared error aggregated over the 0–40 Hz bins. On the same twelve checkpoints it separates nothing: FuncVAE's median is *lower* (apparently better) at 6.1×10^7 against FuncFlow's 7.3×10^7, and the distributions overlap almost completely (exact one-sided $p = 0.91$).

More seeds would not resolve this, for an arithmetic reason. The aggregate is an unweighted mean over frequency bins, so it decomposes exactly into per-band contributions, and **the two bins below $0.5\,\mathrm{Hz}$ supply at least 94% of it on every one of the twelve checkpoints** (medians 98.3% for FuncFlow, 96.8% for FuncVAE). In unfiltered EEG those bins carry drift rather than neural activity; beta and gamma together contribute 0.012% and 0.067%. A score in which the bands under discussion are one part in ten thousand cannot rank models by their behaviour there. Dropping the two bins moves the median ranking to FuncFlow (1.5×10^6 against 1.9×10^6) but only weakly ($p = 0.047$), since the linear metric is still decided by whichever band carries the most absolute power. A log scale, which normalises for those power differences, separates the two clearly: on mean $|\Delta \log_{10} P|$ over 0.5–40 Hz FuncFlow leads by $9.7\times$ (0.154 against 1.50, medians). This is why we score on log power throughout (§2.4).

A downstream task agrees with the ordering, at the other end of the spectrum. The sharper test is whether synthetic data can *stand in* for real data. Training a gaze-regression head from scratch on a single source of EEG and evaluating on the real held-out subjects (30 seeds, one-sided paired t-tests, Holm–Bonferroni at $\alpha = 0.025$, $k = 9$), FuncFlow synthetic data at $5\times$ volume shows no confirmed degradation against real data on any metric: MED $96.29 \pm 1.37\,\mathrm{mm}$ against the real baseline's $96.07 \pm 2.31\,\mathrm{mm}$, and $r_x = 0.383 \pm 0.024$ against 0.379 ± 0.110. FuncVAE's output is Holm-confirmed inferior at every ratio ($r_x = -0.020 \pm 0.073$ at $5\times$). This is a failure to reject rather than a demonstration of equivalence, the gap being $0.22\,\mathrm{mm}$ against a baseline s.d. of $2.31\,\mathrm{mm}$, and at matched $1\times$ volume the synthetic data is confirmed inferior on every metric (supplementary §F).

What the downstream task actually reads. A four-channel frontopolar and temporoparietal montage is not where cortical gaze correlates live; it is where the corneoretinal dipole is largest (Plöchl et al., 2012). Retraining the same head on real data high-passed at increasing cutoffs removes the task entirely: in a separate three-seed run, r_x falls from 0.384 ± 0.028 unfiltered to 0.016 ± 0.072 at 1 Hz and stays at chance above it, with MED pinned at the near-constant-gaze value of $100.0\,\mathrm{mm}$. Eighty-one per cent of the corpus's power sits below 1 Hz, and so does all of the gaze-predictive signal. The substitution result therefore shows that FuncFlow's samples carry recoverable sub-1 Hz ocular structure and that the VAE's do not. It is *not* independent confirmation of the beta and gamma finding, since the two measurements read opposite ends of the spectrum.

3.2 Is high-band loss general? A discrete-kernel baseline

Every model in §3.1 is continuous-time or operator-based, so the comparison cannot distinguish "the top of the spectrum is hard" from "the top of the spectrum is hard *for this family*". We therefore trained the tightest available discrete control: the same U-Net velocity predictor with unconstrained `nn.Conv` kernels and `AvgPool`/nearest-neighbour resampling in place of basis expansions and Fourier truncation, with every other component, and the objective, data, split, schedule and optimiser, identical.

It holds the high band. Running both arms through the same driver at matched seeds and scoring final checkpoints against the test split as loaded by the training pipeline (755 windows), the discrete twin reproduces 0.797 of real beta and 0.837 of real gamma against the functional arm's 0.742 and 0.784 (three paired seeds, divergence 0.000 throughout). The paired differences favour the discrete arm in two seeds and reverse in the third, so at this sample size the two are not distinguishable. High-band loss is therefore not what generative models of raw EEG do in general; on this evidence it is what the continuous-time models we tested do.

This bounds the capability claim: the substrate does not buy high-band fidelity a discrete architecture of the same shape cannot reach. What it adds is the analysable band-limit that makes the design rule of §3.4 statable (supplementary §J), and the off-grid evaluation of §3.6, which we have not tested on the twin.

3.3 Is it the objective or the bottleneck? Widening the bottleneck

FuncVAE routes everything through a 64-point functional bottleneck, an $8\,\mathrm{Hz}$ grid Nyquist over a $4\,\mathrm{s}$ window, so most of the band under discussion leaves the latent path before the objective can act. Widening it to 512 points lifts that ceiling to $64\,\mathrm{Hz}$, entirely above the signal band, and moves

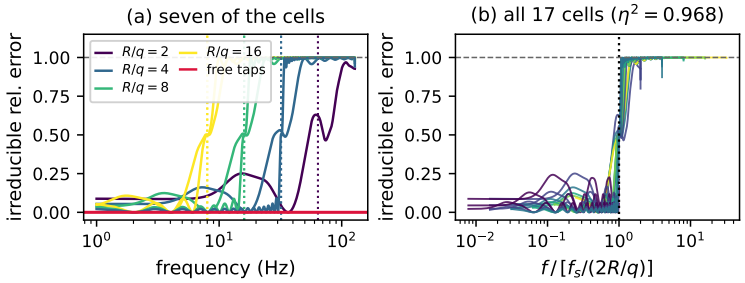

Figure 2: The band-limit with nothing else present: the irreducible error of the basis subspace, computed by projection. **(a)** Seven of the 17 configurations, each climbing from nil to unity at its own predicted cutoff (dotted, matched colour); curves are coloured by R/q, so q=4, R=64 and q=8, R=128 share a colour and a cutoff despite differing in both kernel length and basis count. Free taps (red) have no cutoff. **(b)** Rescaling frequency by each configuration's own cutoff collapses all 17 onto a single step.

beta from 0.063 to 0.156 and gamma from 0.031 to 0.136 (one seed per rung; the 64-point rung is the canonical model rescored under this protocol): **real in direction and an order of magnitude too small to account for the gap**, leaving the model at about one seventh of real high-band power against FuncFlow's 1.21 and 1.63. The operative variable is the prior, whose length scale is specified in *latent grid steps*, so leaving it at the default while widening the grid also sharpens it in physical time; running the ladder both ways turns that into the control. The unscaled ladder moves gamma by 300× against the time-matched ladder's 4.5×, and at 512 points, where only the prior differs, gamma is 0.136 against 9.186. For a functional VAE the output bandwidth is set by the correlation time of the latent prior, the same species of constraint as the kernel band-limit of §3.4 acting elsewhere (supplementary §K).

3.4 Basis parameterisation is a band-limit

The canonical model runs at $R/q \in \{1, 2\}$, cutoffs of 128 and 64 Hz against a 40 Hz band, so its kernels suppress nothing the data carries. That headroom is what the efficiency argument proposes to spend, and we measure its worth first where nothing else is present.

A basis-parameterised kernel is $w = Bc$, with B the (R, q) matrix of basis functions sampled on the kernel grid, so whatever the optimiser does w is confined to $\mathrm{col}(B)$ and the best any such layer can do at reproducing a target kernel is its orthogonal projection P onto that subspace: a statement about the *reachable set*, which no schedule, initialisation or learning rate enlarges. Probing it with sinusoidal targets and recording the irreducible error $\mathrm{rel}(f) = \|k_f - Pk_f\|^2/\|k_f\|^2$ therefore measures the parameterisation itself, with no model, data or optimisation present.

Figure 2 reports the 17 configurations with $q \le R$ drawn from $q \in \{4, 8, 16, 32\}$ and $R \in \{8, \dots, 128\}$. In the 14 whose cutoff lies below Nyquist, the error rises from nil to unity at the predicted cutoff $f_s/(2R/q)$, and rescaling frequency by that cutoff collapses every curve onto one. The correlation ratio η^2, used because $\mathrm{rel}(f)$ is step-like, is 0.968 for normalised frequency against 0.393 for absolute frequency (probed over 1–128 Hz), 0.058 for kernel length and 0.056 for basis count. Parameter count explains none of it by construction, since $q\, C_{\mathrm{in}} C_{\mathrm{out}}$ does not contain R. For the Fourier basis the cutoff also follows analytically from the harmonics a q-term expansion contains, so the projection confirms it rather than discovers it; it matters for the bases below, which have no such closed form.

The cutoff is **asymptotic in** q: predicted and measured agree to within 0.26% at q=32 but only 15.6% at q=4, R=8, where a four-dimensional subspace makes a cutoff a coarse notion.

The R/q scaling is a subspace property; the constant is the basis's. Running the same projection on three other dictionaries, the scaling transfers and the constant does not: log-log slopes of 0.99, 1.00 and 1.04 for Fourier, Legendre and B-splines at $R^2 \ge 0.997$, with prefactors 1.00, 0.78 and 0.75. The band-limit is therefore $c_{\mathcal{B}}\, f_s/(2R/q)$ with $c_{\mathcal{B}} = 1$ only for Fourier, whose edge is also the

sharpest, so elsewhere the formula locates the centre of a roll-off rather than a wall. A db4 wavelet dictionary, which as implemented stretches rather than translates its functions, shows no cutoff at all (supplementary §C).

The same signature appears in a trained network. On the velocity field the network actually fits, relative error saturates at unity above each configuration's own cutoff (0.807 to 0.979 as R/q runs 4 to 32) while *falling* below it (0.685 $\rightarrow$ 0.521), and a free-tap control at matched support shows neither; R/q alone reaches $R^2 = 0.906$ there. Generated-sample spectra show nothing, because integrating that field against correlated noise launders the frequency structure (supplementary §G–§H).

3.5 What the band-limit costs: the support sweep

§3.4 predicts where the trade turns: at $f_s = 256$ Hz against a 40 Hz band the cutoff crosses the band edge at $R/q = 3.2$, so the q=8 sweep straddles it and quality should fall as support grows. Holding $q = 8$ and sweeping $R \in \{8, 16, 32, 64, 128\}$, the functional network holds **exactly** 3,653,840 **parameters at every point on the axis** while the unconstrained-kernel model of identical topology and support grows to 42,819,792, an 11.7$\times$ spread, both counts agreeing bit-for-bit with the analytic predictions. **Quality does not follow.** The best support for the functional arm is the *shortest* tested, and the paired flow loss degrades monotonically from 0.1932 at R=8 to 0.2388 at R=128, roughly five standard errors on a metric whose per-cell standard error is 0.009.

The ratio is the dominant variable, and not the only one. An *optimisation* account, in which long kernels are slow to fit, makes the governing variable R; the *representational* account of §3.4 makes it R/q. Neither is pure. The unconstrained arm says so before any regression: carrying no basis constraint at all, it still degrades from 0.1969 to 0.2135, which is 0.0166 of the functional arm's 0.0456, so about a third of the penalty is a length effect the basis has nothing to do with. Varying q and R independently separates them: the paired flow loss regresses on both as $0.0133 \log_2(R/q) + 0.0053 \log_2 q$ at $R^2 = 0.957$, while parameter count alone reaches $R^2 = 0.004$ across cells spanning 2.07 to 13.16 M and the ratio alone reaches 0.869, against 0.760 for support. The collapse is not exact, since the $\log_2 q$ coefficient sits at 40% of the ratio coefficient rather than vanishing, so a residual kernel-length penalty survives at matched ratio (supplementary Figure A5).

At matched parameters and matched support, the basis buys a better fit and nothing else. At q=R=8 the compression ratio is 1.0, so no band-limit applies inside the signal band and both arms hold identical parameter counts (3,653,840), support and resampling. On the flow-matching objective at five seeds per arm, every basis seed beats every free-tap seed (exact one-sided $p = 0.0040$); on generated-sample spectra the advantage vanishes, does not survive a change of initialisation at R=64, and does not return in a scarce-data regime (supplementary §L–§M).

The penalty is not the schedule, and it is not visible in sample power. Two objections remain: that a 20,000-step budget does not give a long kernel time to fit, and that held-out loss is not what a reader cares about. Retraining q=8, $R \in \{8, 64, 128\}$ for the canonical 100,000 steps answers the first: over three seeds the mean gap between shortest and longest support is 0.0486, against 0.0480 at the sweep budget, so **the penalty does not shrink with a fivefold longer schedule** where an optimisation account predicts it closes toward zero. Sample power answers the second in an unexpected direction: generated-to-real gamma rises monotonically with the ratio (0.96, 7.26, 16.8 as R/q runs $1, 8, 16$), as do beta and the divergence rate, so as the kernel cutoff descends through the signal band the high band is not *lost* but increasingly *over*-produced by an increasingly unstable sampler. The band-limit constrains the velocity field, and integrating that field against correlated noise does not carry the constraint through to sample power (supplementary §I).

3.6 Evaluation off the training grid

Because the basis functions can be sampled on any grid and resolution changes happen in Fourier space, one trained model runs on grids absent from training: at a fixed 4 s window, weights trained at 256 Hz generate coherent samples at 512 Hz, and at a fixed rate the same weights generate 1 to 8 s windows. Support R is fixed in *samples*, so raising f_s raises the cutoff $f_s/(2R/q)$ with it, while lowering f_s would lower it; we therefore claim upward rate transfer and window-length transfer, not downward rate transfer (supplementary §E). We show this by inspecting generated samples and spectra rather than by a quantitative test. The discrete twin is also fully convolutional and could

plausibly run at other window lengths; we have not tested it, so we do not claim the capability is unique to the functional substrate.

4 DISCUSSION

One dial, two ends. The capability and the negative result are one measurement read in opposite directions: the ratio that makes support cheap fixes a cutoff at $f_s/(2R/q)$, a property of the reachable set no schedule or initialisation relaxes. Our configuration keeps that cutoff clear of the signal band; the efficiency argument asks for exactly that headroom, and spending it does not pay. Support is free in parameters, not in compute (the exchange rate is R/q either way), and not in *bandwidth*; parameter count, the quantity the efficiency argument is stated in, predicts nothing ($R^2 = 0.004$).

The design rule this yields is usable outside our setting: choose $q \geq R \cdot 2f_{\max}/(c_{\mathcal{B}} f_s)$, with $c_{\mathcal{B}} = 1$ for a Fourier basis and smaller for the others we measured, or lose the reachable set above the cutoff. Three qualifications belong with it. The dependence is not *purely* on the ratio, since a residual kernel-length penalty survives at matched ratio and appears in the unconstrained arm too. The constant is basis-specific even though the scaling is not (§3.4). And the claim is about what the kernel can represent rather than what leaves the sampler: the band-limit is plain in the velocity field the network fits and invisible in generated-sample spectra, where integrating against correlated noise launders the frequency structure (supplementary §H). We state the rule as a constraint on capacity, not as a prediction about sample power.

The measurement problem is the paper's second result. Two models on one substrate gave us a control we did not expect to need. Band-resolved fidelity and a 30-seed downstream substitution test separate them cleanly and in the same direction, though the second reads the sub-1 Hz end of the spectrum rather than the band the first measures. A linear aggregate 0–40 Hz spectral error, the natural single-number summary, ranks them the other way on its median and cannot separate them across twelve checkpoints. More seeds will not fix that, because the aggregate is not mostly a measurement of the EEG band at all: at least 94% of it is decided by the two bins below 0.5 Hz (§3.1). Scoring on a log scale, which normalises for the power differences between bands, is what recovers the ordering, and it is the scoring we use throughout.

Seed-level instability compounds it: three seeds of one sweep cell span a factor of 40 at divergence rates from 0% to 83.3%, so any comparison here reporting a single seed, or a mean without the divergence rate, is reporting noise (supplementary §A).

Scope. The band-limit is a property of the parameterisation and its R/q scaling holds in every basis we measured, though the constant does not; everything else here is one corpus of four-channel consumer EEG with a Fourier basis, and the prior at issue is smoothness, so the construction suits continuously sampled data. The $20{,}000$-step sweep budget still bounds the ablation. On dry frontal and temporal electrodes without a bandpass, power above 30 Hz likely includes muscle activity, so gamma fidelity here means fidelity to the recorded signal rather than to cortical gamma. The downstream synthetic sets come from checkpoints the trainers selected on a test-split criterion. Stimulus conditioning is small in energy but not absent: the real gaze trace rather than a null one moves the output by only 0.70–$1.66\,\mu$V per channel, yet a head trained purely on synthetic EEG recovers $r_x = 0.383$ on *real* held-out recordings where the identically conditioned VAE returns $r_x \approx 0$.

5 CONCLUSION

A generative model assembled entirely from basis-parameterised convolutions keeps beta and gamma power in raw EEG, erring by excess, where a functional VAE on the same substrate keeps a twentieth and a fiftieth of it. At three seeds it is not distinguishable in either band from a discrete twin with unconstrained kernels, so that fidelity is not what the basis buys. What the basis adds is an analysable cutoff at $f_s/(2R/q)$, which must stay above the signal band: the efficiency on offer is not free capacity but that headroom spent differently. The trained model also runs at window lengths and sampling rates absent from training, a capability we show qualitatively and have not tested on the discrete twin.

AI USE STATEMENT

Generative AI tools (Anthropic Claude, accessed through an agentic coding interface) were used in this work for software implementation, experiment orchestration, analysis scripting, and drafting and editing of prose. They were not used to generate research ideas, to select the hypotheses under test, or to choose the statistical procedures; the study's central hypothesis, its pre-registered decision rule, and the decision to widen the seed count that ultimately withdrew one of the paper's claims were author decisions.

All AI-assisted work has been reviewed. Every number in the paper is traceable to a JSON sidecar emitted by the run that produced it, and the efficiency results (parameter counts, flow losses, the support and ratio regressions, the matched-parameter ablation and the sample-efficiency test) are additionally checked against those sidecars by a verification script released with the experiment code, which is re-runnable so that a later re-evaluation cannot leave a stale number in the prose. That script does not yet cover the band-resolved, downstream or conditioning numbers, which we checked against their sidecars by hand; one figure that failed that check has been corrected. AI-assisted analysis code was verified against the training loop's own instrumentation wherever a discrepancy would have changed a conclusion. In particular, the frequency-resolved velocity measurement reproduces the evaluation loss to within 0.1% on every cell, a check that would have failed had the measurement been wrong. Three intermediate results that did not survive this checking are reported as withdrawn in the supplementary material rather than removed.

We take responsibility for the final content of this work, including text, claims and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The comparator in every controlled comparison was trained under matched conditions rather than extrapolated from a parameter count, and the decision rule for the sample-efficiency test was fixed before its deciding runs; the version-controlled record of that ordering will be released with the code. The optimiser settings, training budget and arm definitions are given in the supplementary, together with complete per-cell tables including seed-level values and sampler divergence rates, and the statistical procedures: exact one-sided tests, the family definitions, and the Holm corrections applied to each. Driver scripts for all reported experiments, the analysis scripts, and the figure-generation scripts with their numeric sidecars will be released with the camera-ready version; they are withheld here to preserve anonymity. Superseded sidecars are retained rather than overwritten, so every revision of a reported number, including the move of the band-resolved comparison of §3.1 from test-selected to final checkpoints, is auditable.

ETHICS STATEMENT

All experiments use a publicly released EEG and eye-tracking corpus (Afonso & Heinrichs, 2025). According to its documentation, participants gave informed consent to participation and to data sharing, the recordings are anonymised, and no ethics approval was required under the guidelines of the German Research Foundation for non-invasive EEG recordings of this kind. We collected no new data. Synthetic EEG from the models studied here is not a substitute for real recordings in any clinical or diagnostic setting.

REFERENCES

Tiago Vasconcelos Afonso and Florian Heinrichs. Consumer-grade EEG-based eye tracking. arXiv:2503.14322, 2025.

Giulio Franzese, Giulio Corallo, Simone Rossi, Markus Heinonen, Maurizio Filippone, and Pietro Michiardi. Continuous-time functional diffusion processes. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Albert Gu, Karan Goel, and Christopher Ré. Efficiently modeling long sequences with structured state spaces. In *International Conference on Learning Representations*, 2022. arXiv:2111.00396.

Kay Gregor Hartmann, Robin Tibor Schirrmeister, and Tonio Ball. Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals, 2018. URL https://arxiv.org/abs/1806.01875.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

Florian Heinrichs, Mavin Heim, and Corinna Weber. Functional neural networks: Shift invariant models for functional data with applications to EEG classification. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 12862–12882. PMLR, 2023.

Ard Kastrati, Martyna Beata Plomecka, Damián Pascual, Lukas Wolf, Victor Gillioz, Roger Wattenhofer, and Nicolas Langer. EEGEyeNet: a simultaneous electroencephalography and eye-tracking dataset and benchmark for eye movement prediction. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, 2021. arXiv:2111.05100.

Gavin Kerrigan, Giosue Migliorini, and Padhraic Smyth. Functional flow matching. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 3934–3942. PMLR, 2024.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2024.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Miguel Liu-Schiaffini, Julius Berner, Boris Bonev, Thorsten Kurth, Kamyar Azizzadenesheli, and Anima Anandkumar. Neural operators with localized integral and differential kernels. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*. PMLR, 2024.

Michael Plöchl, José P Ossandón, and Peter König. Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, 6:278, 2012.

Qiang Qiu, Xiuyuan Cheng, Robert Calderbank, and Guillermo Sapiro. DCFNet: Deep neural network with decomposed convolutional filters. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*. PMLR, 2018.

David W. Romero, Robert-Jan Bruintjes, Jakub M. Tomczak, Erik J. Bekkers, Mark Hoogendoorn, and Jan C. van Gemert. FlexConv: Continuous kernel convolutions with differentiable kernel sizes. In *International Conference on Learning Representations*, 2022a.

David W. Romero, Anna Kuzina, Erik J. Bekkers, Jakub M. Tomczak, and Mark Hoogendoorn. CKConv: Continuous kernel convolution for sequential data. In *International Conference on Learning Representations*, 2022b.

Cédric Rommel, Joseph Paillard, Thomas Moreau, and Alexandre Gramfort. Data augmentation for learning predictive models on EEG: a systematic comparison. *Journal of Neural Engineering*, 2022. arXiv:2206.14483.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241. Springer, 2015.

Jacob H. Seidman, Georgios Kissas, George J. Pappas, and Paris Perdikaris. Variational autoencoding neural operators. arXiv:2302.10351, 2023.

Giulio Tosato, Cesare M. Dalbagno, and Francesco Fumagalli. EEG synthetic data generation using probabilistic diffusion models. arXiv:2303.06068, 2023.