Selective-scan mechanics (shrinking deep dips vs. rare micro-ridges)

Each MambaBlock wraps the official mamba_ssm.Mamba with (d_state,d_conv,expand)=(16,4,2); FiLM-conditioned layer norms gate the scan (src/models/mamba_uv.py:39-62,95-175). At wavelength i:

$$s_i = A(x_i) \odot s_{i-1} + B(x_i) \odot x_i, \qquad \hat{y}_i = C(x_i) \odot s_i.$$

Gate logits A,B,C and the decay Δ stem from the FiLM-scaled token (geometry head always enabled during training; scripts/run_cross_validation.py:177-229). Off-manifold tokens—e.g., an unusually deep absorption—drive Δ large, so A(x_i)→0 and B(x_i) shrinks; the state carries smoother history and the dip relaxes toward the cohort average (“shallower dip 6”).

In the convolutional view, the cumulative product of (1−A) shifts the adaptive kernel’s mass to neighboring wavelengths, partially refilling an extreme trough. The opposite regime (structured noise mimicking a band) keeps the decay small for a few steps, allowing B(x_i) to inject energy and C(x_i) to read it out before DipAwareLoss penalties (src/losses.py:197-274) suppress the excursion—observed as rare micro-ridges.