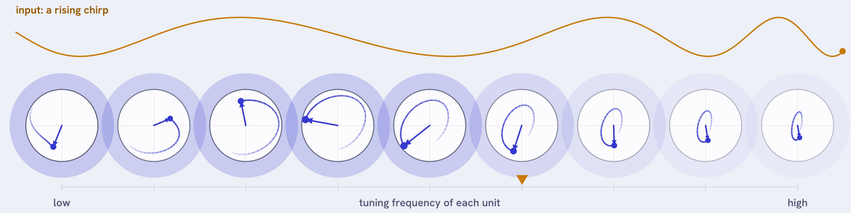
NeurIPS 2026 spotlight

Prospective Coding Improves Learning in Deep Continuous-Time Recurrent Networks

Shivang Rawat[1*], Mirko Morello[1], Flaviano Morone[1,2], David J. Heeger[1,2]

[1] Telepath [2] New York University [*] shivang@telepath.com

z bottom-up input y state of a unit $\chi_\tau\,\hat{y}$ look-ahead and prediction

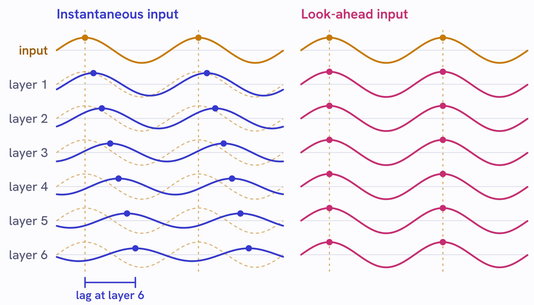


A layer of RQFs listening to a rising chirp. Each unit's complex state turns at its own frequency; the unit tuned to the input swings widest.

1 Deep stacks of slow neurons fall behind

A neuron integrates its input over a time constant τ. That gives it memory but also makes it late, and in a stack the delays add up. A teaching signal applied at the top then refers to activity that the early layers produced long ago.

$\tau\,\dot{y}_\ell = -y_\ell + y_{\ell-1}$ each layer is a leaky integrator driven by the layer below



111 ms behind the input at layer 6, with 51% of its amplitude

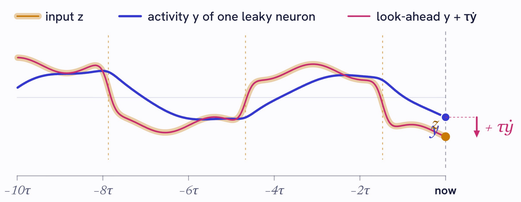
0 ms every layer in step with the input, at full amplitude

Six leaky integrators with τ = 20 ms, driven by a 4 Hz input. Dots mark each trace's peaks; dashed lines mark the input's peaks.

2 Neurons can look ahead

$\chi_\tau = 1 + \tau\,\frac{d}{dt}$ The look-ahead operator. It uses only the present activity and its slope.

Pyramidal cells in cortex, retinal ganglion cells, cerebellar Purkinje cells and muscle spindles all fire at a phase that leads their input. For a leaky integrator the look-ahead is the exact inverse, so applying it to every layer's input puts the whole stack back in step (right panel above).



3 From an energy principle to a band-pass filter

Each unit balances its bottom-up input against an internal prediction; γ sets the trade-off (Heeger, PNAS 2017).

$$E = \gamma\,|y - z|^2 + (1-\gamma)\,|y - \hat{y}|^2$$

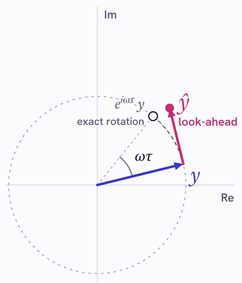
Descending the energy gives a leaky integrator that relaxes toward a blend of the two.

$$\tau\,\dot{y} = -y + \gamma\,z + (1-\gamma)\,\hat{y}$$

For the prediction, take the simplest model of an oscillation, $\dot{y} = i\omega y$, and look one time constant ahead.

$$\hat{y} = \chi_\tau(y)\big|_{\dot{y}=i\omega y} = (1 + i\omega\tau)\,y$$

$\hat{y}$ leads y in phase, so the state keeps chasing it and spirals.



Substituting gives the Recursive Quadrature Filter (RQF):

$$\tau\dot{y} = -y + \gamma z + (1-\gamma)(1 + i\omega\tau)\,y \;\Longrightarrow\; \boxed{\dot{y} = \lambda\,y + \alpha\,z}$$

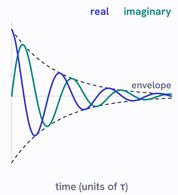
$\alpha = \gamma/\tau \qquad \omega_0 = (1-\gamma)\,\omega \qquad \lambda = -\alpha + i\,\omega_0$

stable for any γ in (0, 1) · two learnable numbers per unit: ω and γ

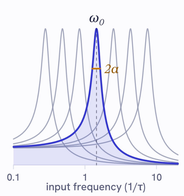
Free response: a spiral



Impulse response: a quadrature pair



A layer is a bank: ω tunes, γ sets the width

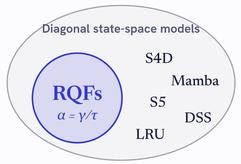


4 Every RQF layer is a diagonal SSM

RQF layer $\quad \dot{\mathbf{y}} = \mathrm{diag}(\boldsymbol{\lambda})\,\mathbf{y} + \mathrm{diag}(\boldsymbol{\alpha})\,W\,\mathbf{x}$

Diagonal SSM $\quad \dot{\tilde{\mathbf{y}}} = \Lambda\,\tilde{\mathbf{y}} + B\,\mathbf{x}$

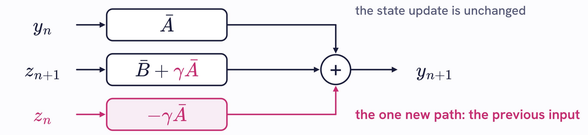
The only constraint: each channel's input gain is tied to its own decay rate, $\alpha = \gamma/\tau$. So RQFs inherit the SSM toolkit, parallel scans and FFT convolutions, and train on sequences of 16,384 steps.



5 Look-ahead on the input: two taps, zero parameters

Drive each layer with χ_τ of the layer below. Integrated exactly over each step, with the input's jump entering as an impulse, the look-ahead becomes one extra tap:

$$y_{n+1} = \bar{A}\,y_n + \left[\bar{B} + \gamma\bar{A}\right] z_{n+1} \;-\; \gamma\bar{A}\,z_n$$




```
- y = A_bar * y + B_bar * z[t]
+ y = A_bar * y + B_plus * z[t] + B_minus * z_prev
+ z_prev = z[t]
  # B_plus = B_bar + γ·A_bar,   B_minus = -γ·A_bar
```

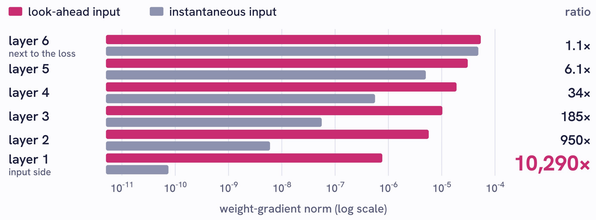
0 new parameters · same transition $\bar{A}$ · same parallel scan · any diagonal SSM

6 The error survives the trip down

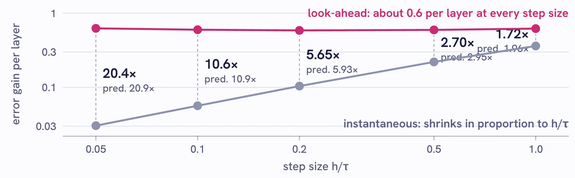
As a diagnostic, detach every earlier state (spatial-only backprop) so that errors travel only down through the layers. The step-size factor on each hop goes from $O(h/\tau)$ to $O(1)$:

$$\underbrace{\tfrac{h}{\tau}\,\gamma}_{\text{instantaneous}} \longrightarrow \underbrace{\left(1 + \tfrac{h}{\tau}\right)\gamma}_{\text{look-ahead}}$$

Measured weight-gradient norms, six-layer raw-audio RQF, h/τ = 0.2



Per-layer error gain across step sizes: measured ratio vs. exact-ZOH prediction



7 Small models, long sequences



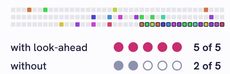
16,000 samples, one per step

96.1% **Raw-audio Speech Commands**
31.9k parameters, six layers, no residuals, gating or normalization.



128 × 128 image, 16,384 steps

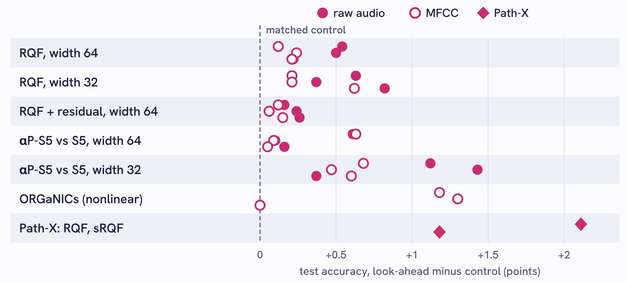
83.4% **Path-X, 16,384 steps**
Minimal six-layer RQF; 86.3% when the input gain is selective (sRQF).



16 tokens hidden in 4,096 steps

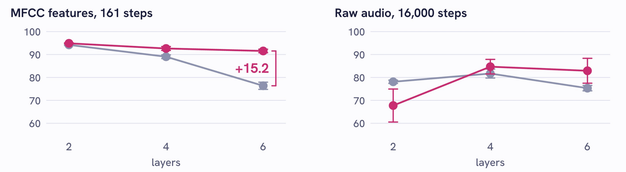
99.8% **Selective Copying**
Selective RQF (γ, ω and input gain depend on the input).

Under full BPTT, look-ahead matched or beat its control in all 35 comparisons



Each mark is one depth (2, 4, 6, top to bottom within a row): mean test accuracy over five seeds, look-ahead minus its matched control. Speech Commands unless noted. The one tie is ORGaNICs at six layers. The S5 rows compare the full αP-S5 construction with native S5.

With spatial credit only, the advantage grows with depth



Spatial-only backprop, width-64 RQF stacks; mean ± s.d. over five seeds.

Derived from cortex. Trained like an SSM.

Look-ahead input costs one extra tap and no parameters, and it applies to any diagonal SSM.

Paper · Code