PRINCETON UNIVERSITY

A Neural Implementation of the Kalman Filter

Robert C. Wilson and Leif H. Finkel

Penn

ABSTRACT

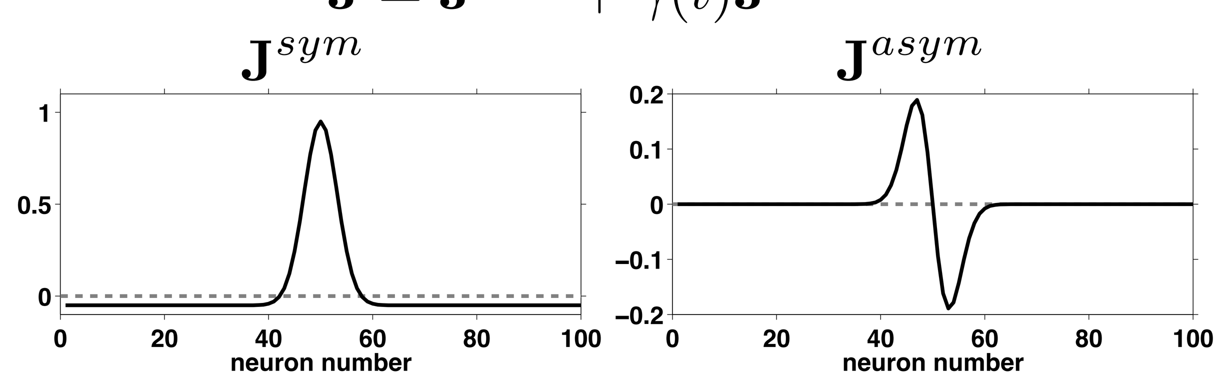
We introduce a novel neural network, based on a modification of Zhang's line attractor network for head direction cells [1], whose dynamics map exactly onto those of the one-dimensional Kalman filter when the prediction error is small. Crucially, our network is *not* itself a line attractor and hence does not suffer from the limitations of these networks such as the need for precisely tuned weights and sensitivity to noise. For large prediction errors the relationship to the Kalman filter breaks down and the network performs something akin to outlier and/or change-point detection.

ZHANG LINE ATTRACTOR MODEL [1]

Network architecture :

- Neurons on a ring with recurrent connections
- Weights consist of symmetric and asymmetric parts

$$\mathbf{J} = \mathbf{J}^{sym} + \gamma(t)\mathbf{J}^{asym}$$



Dynamics :

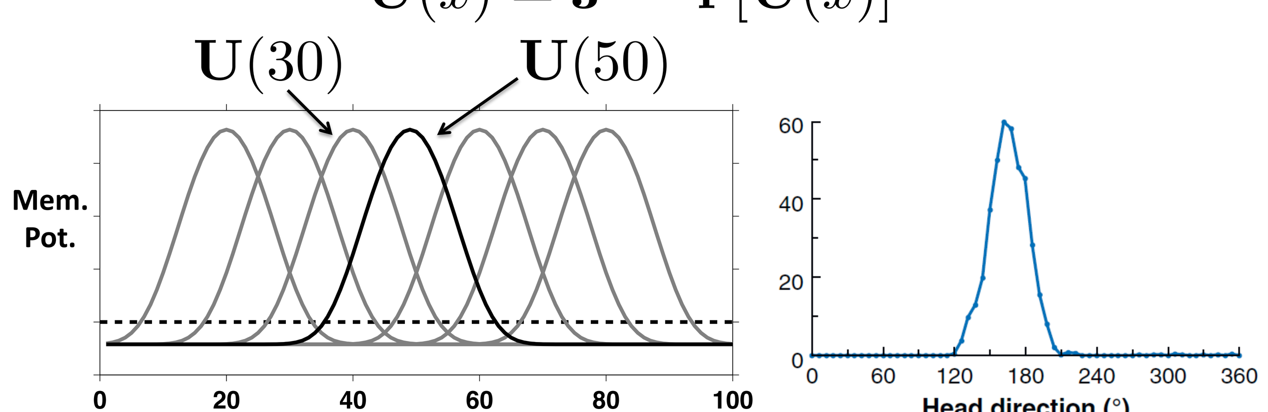
$$\mathbf{u}(t+1) = \mathbf{Jf}\left[\mathbf{u}(t)\right]$$

$\mathbf{u}(t)$: membrane potential
$\mathbf{f}[\cdot]$: activation rule

Fixed points :

- When $\gamma(t) = 0$ network is a continuous attractor
- Fixed points, $\mathbf{U}(x)$, can code and store position, x
- Fixed points satisfy :

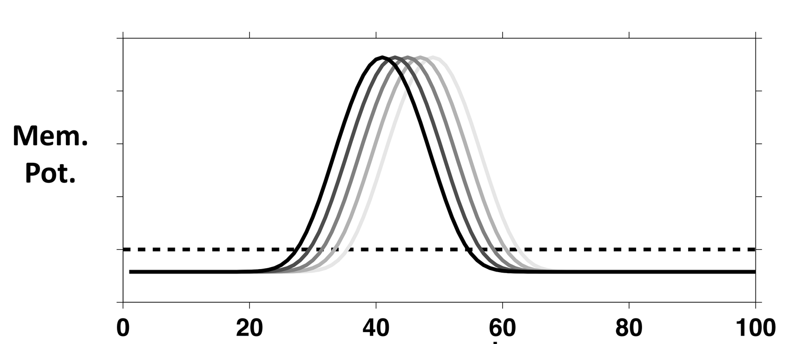
$$\mathbf{U}(x) = \mathbf{J}^{sym}\mathbf{f}\left[\mathbf{U}(x)\right]$$



(taken from [3])

Travelling waves :

- When $\gamma(t) \neq 0$ network exhibits travelling waves
- Waves move with speed $\gamma(t)$
- Can think of this as implementing **prediction**



Limitations :

- Not obvious how to handle inputs
- No notion of uncertainty
- Requires precisely tuned weights
- Not robust to noise

ONE-DIMENSIONAL KALMAN FILTER

Inference can be broken down into three stages :

1. Make prediction :

$$\bar{x}(t+1) = \hat{x}(t) + v(t)$$

2. Update position based on new data :

$$\hat{x}(t+1) = \bar{x}(t+1) + \frac{\hat{\sigma}_x(t+1)^2}{\sigma_z(t+1)^2}\left[z(t+1) - \bar{x}(t+1)\right]$$

3. Update uncertainty :

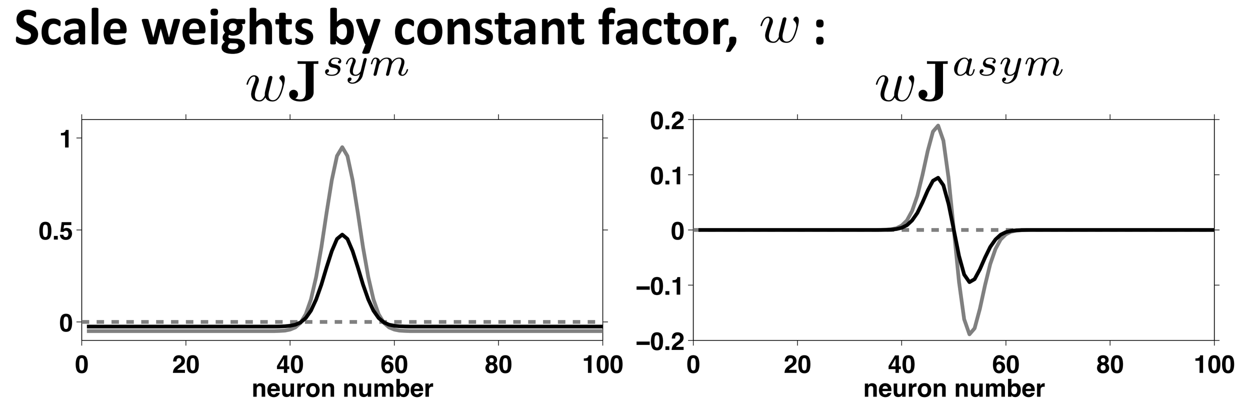
$$\frac{1}{\hat{\sigma}_x(t+1)^2} = \frac{1}{\hat{\sigma}_x(t)^2 + \sigma_v(t)^2} + \frac{1}{\sigma_z(t+1)^2}$$

MODIFICATIONS TO THE MODEL

Divisive normalization activation rule [2] :

$$f_i = \frac{[u_i]_+}{S + \mu\sum_j [u_j]_+}$$

Scale weights by constant factor, w **:**



Non-zero input :

- Limit form of input for purposes of analysis

$$\mathbf{I}(t) = A(t)\mathbf{U}(z(t))$$

Update equation is now :

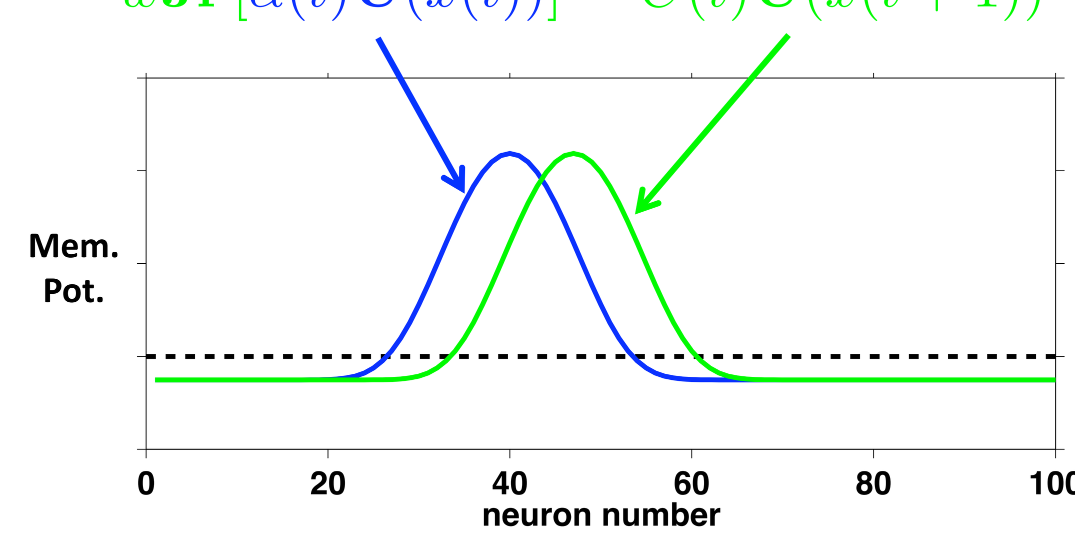
$$\mathbf{u}(t+1) = \underbrace{w\mathbf{Jf}\left[\mathbf{u}(t)\right]}_{\text{recurrent input}} + \underbrace{\mathbf{I}(t+1)}_{\text{external input}}$$

Propose *ansatz* of form :

$$\mathbf{u}(t) = \alpha(t)\mathbf{U}(\hat{x}(t))$$

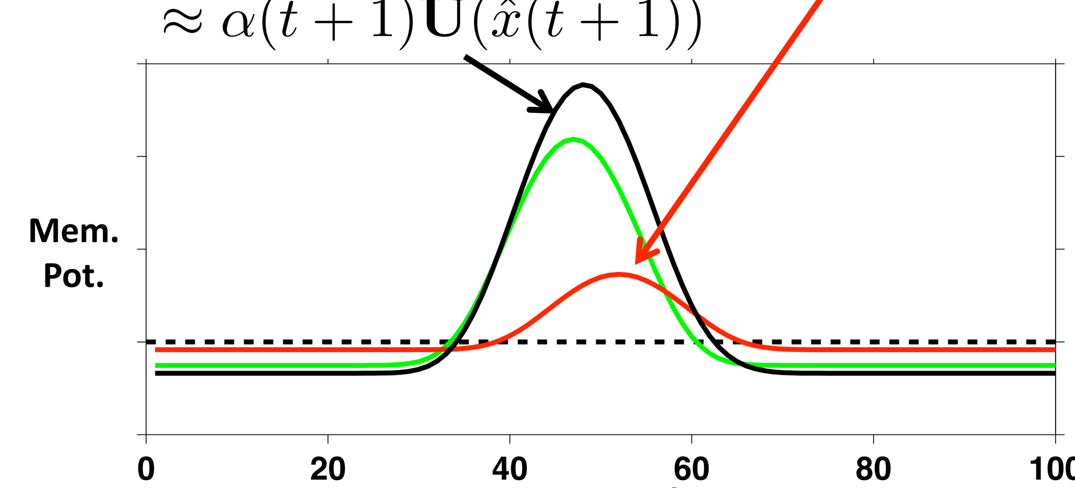
Recurrent input performs prediction (as before) :

$$w\mathbf{Jf}\left[\alpha(t)\mathbf{U}(\hat{x}(t))\right] = C(t)\mathbf{U}(\bar{x}(t+1))$$



External input introduces new observation :

$$\mathbf{u}(t+1) = C(t)\mathbf{U}(\bar{x}(t+1)) + A(t+1)\mathbf{U}(z(t+1))$$
$$\approx \alpha(t+1)\mathbf{U}(\hat{x}(t+1))$$



So long as weight scale factor is set to :

$$w = \frac{S}{S_0 + \mu_0\mathcal{I}}$$

Network dynamics update *ansatz* according to :

1. Make prediction :

$$\bar{x}(t+1) = \hat{x}(t) + \gamma(t)$$

2. Update position based on new data :

$$\hat{x}(t+1) = \bar{x}(t+1) + \frac{A(t+1)}{\alpha(t+1)}\left(z(t+1) - \bar{x}(t+1)\right)$$

3. Update scale factor :

$$\alpha(t+1) = \frac{1}{\frac{S}{w(S_0+\mu_0\mathcal{I})}\frac{1}{\alpha(t)} + \frac{\mu\mathcal{I}}{S}} + A(t+1)$$

These map directly onto the Kalman filter equations with the identifications :

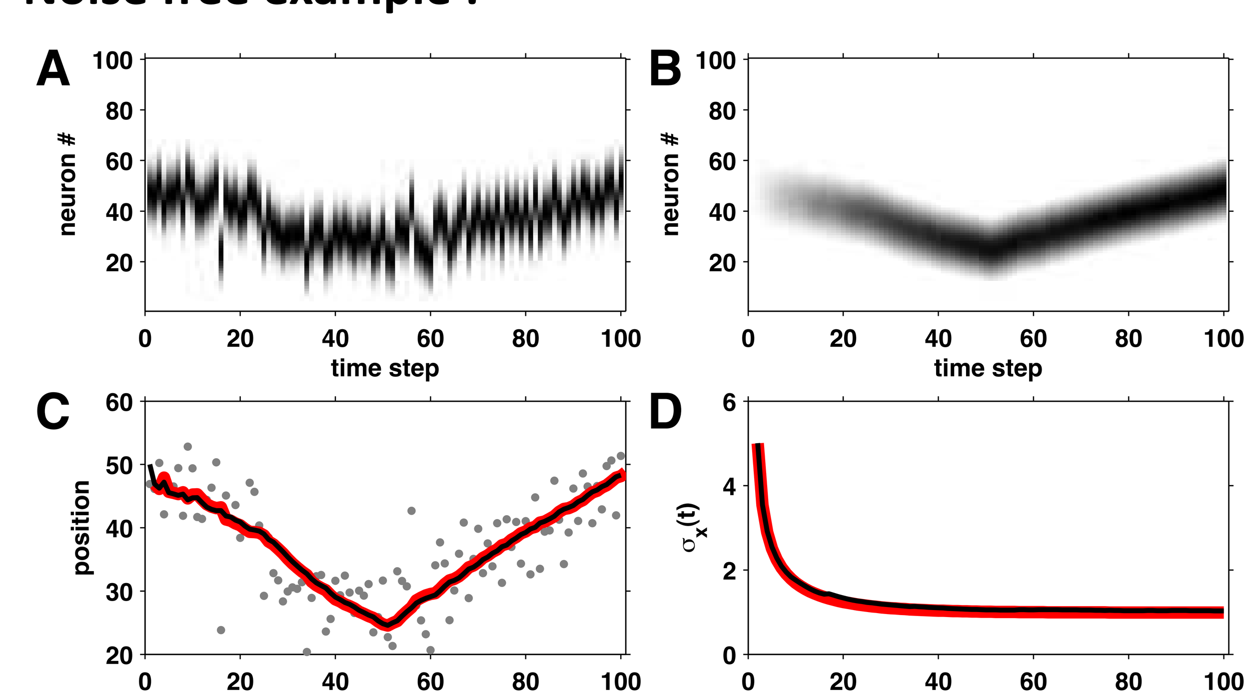
$A(t) \propto \frac{1}{\sigma_z(t)^2}$ Uncertainty in the input position is encoded in the scale factor of the input bump

$\alpha(t) \propto \frac{1}{\hat{\sigma}_x(t)^2}$ Uncertainty in the mean is encoded in the scale factor of the membrane potential profile

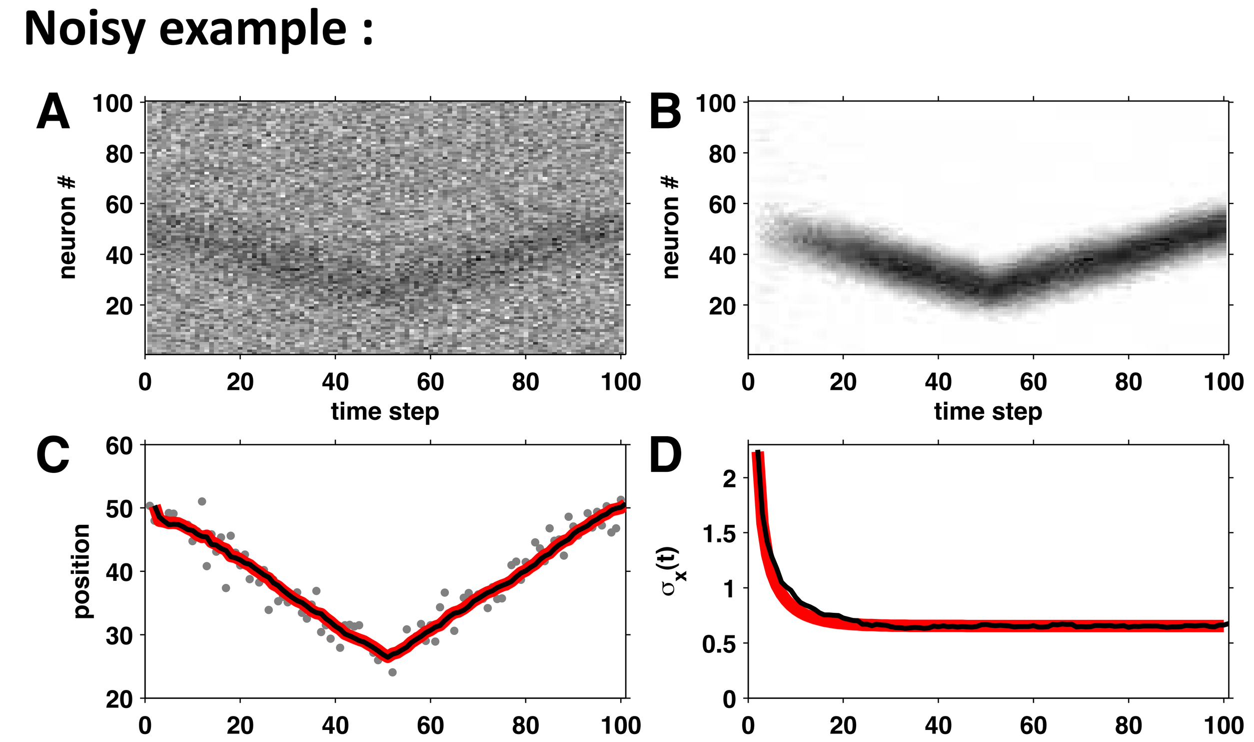
$\frac{\mu\mathcal{I}}{S} \propto \sigma_v(t)^2$ Uncertainty in velocity signal is encoded in the strength of the divisive normalization

EXAMPLES

Noise free example :



Noisy example :

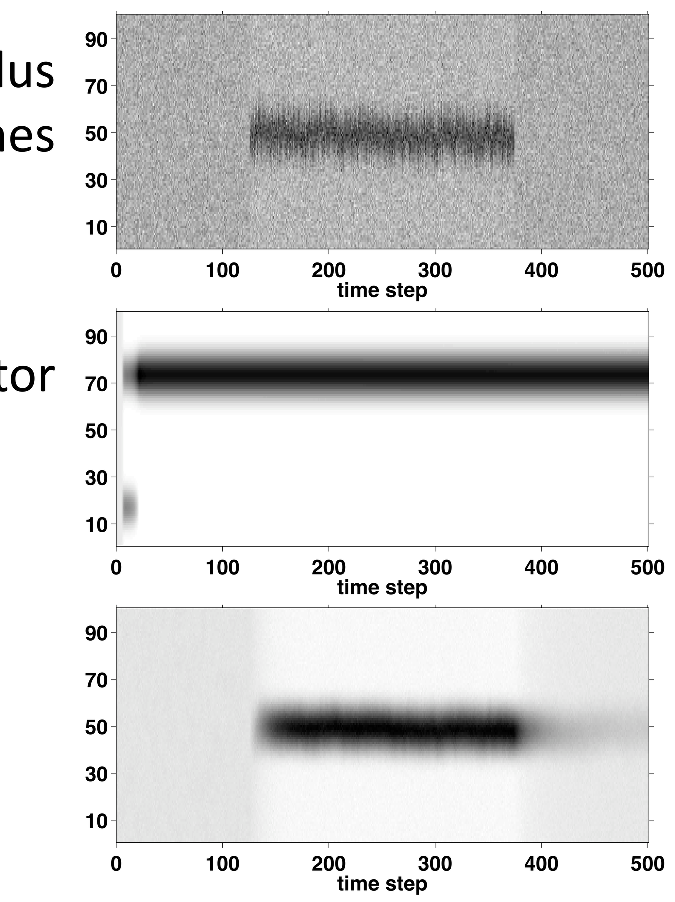


NETWORK IS *NOT* A LINE ATTRACTOR

Noisy input with stimulus present only between times 133 and 367

Response of line attractor where $w = 1$

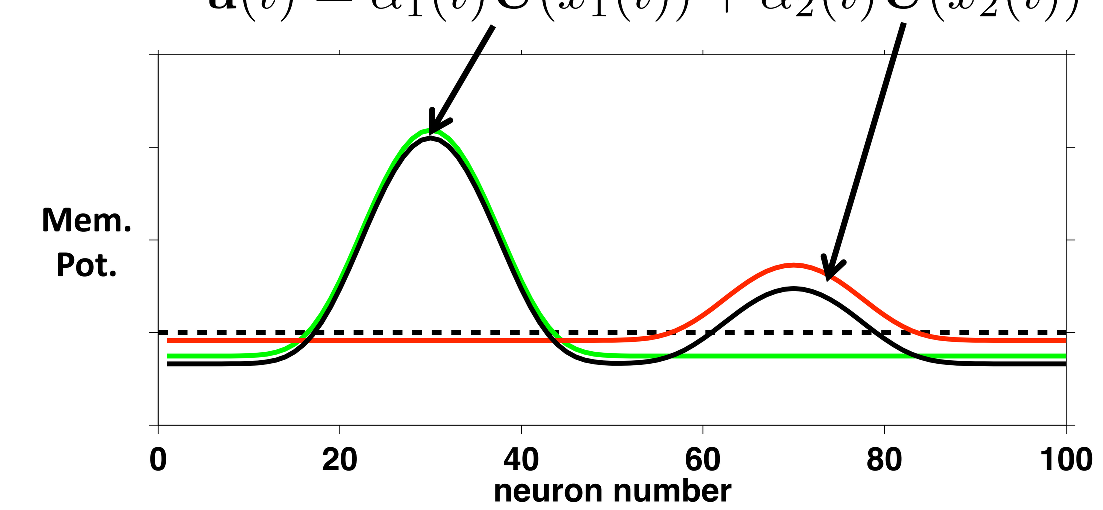
Response of our network $w = \frac{S}{S_0 + \mu_0\mathcal{I}}$



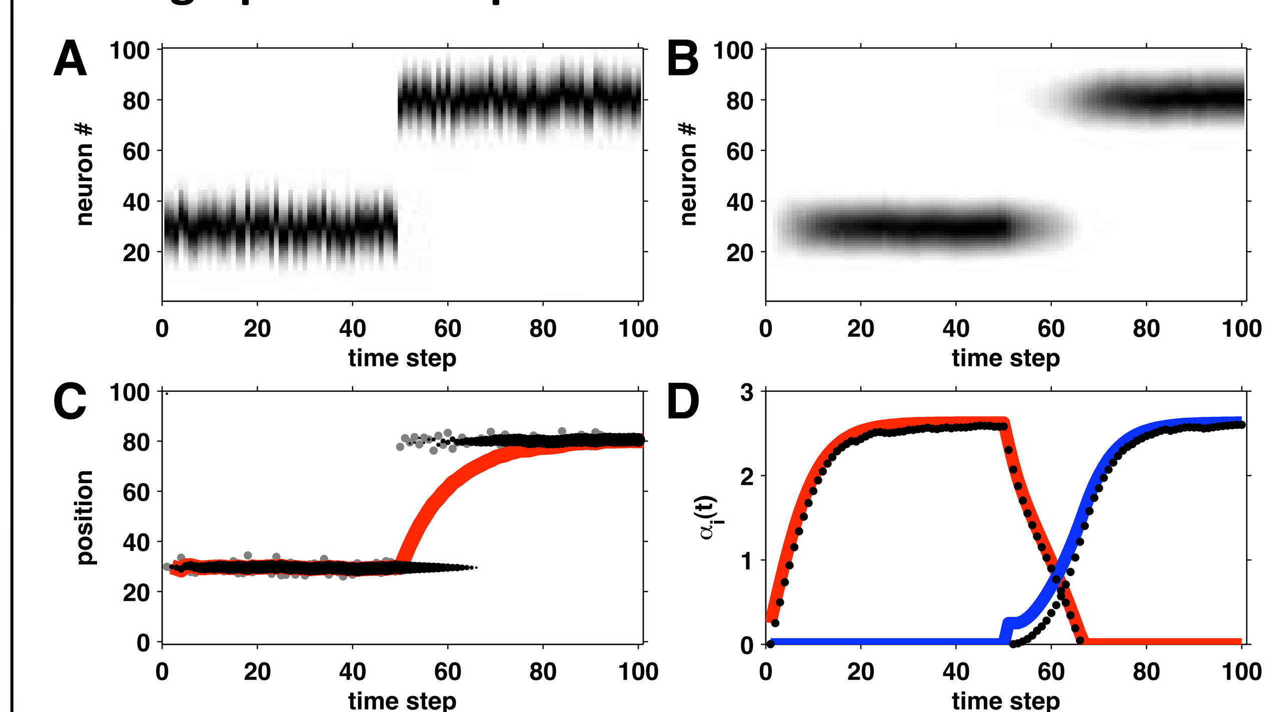
LARGE PREDICTION ERROR

Two bump ansatz :

$$\mathbf{u}(t) = \alpha_1(t)\mathbf{U}(\hat{x}_1(t)) + \alpha_2(t)\mathbf{U}(\hat{x}_2(t))$$



Change-point example :



CONCLUSIONS

We have introduced a novel neural network whose dynamics can implement a close approximation to optimal inference (i.e. a Kalman filter) when the prediction error is small. When the prediction error is large the model performs something akin to outlier and/or change-point detection. The model gives insights into how probability distributions can be encoded and manipulated in the brain.

REFERENCES

[1] K. Zhang. J. Neurosci., 16(6):2112–2126, 1996
[2] D. J. Heeger. Vis. Neurosci., 9:181–198, 1993
[3] J. S. Taube Annu. Rev. Neurosci. 2007. 30:181–207