Efficient neural computation in the Laplace domain

Marc W. Howard, Karthik H. Shankar, and Zoran Tiganj
Department of Psychological and Brain Sciences
Boston University
{marc777,shankark,zorant}@bu.edu

Abstract

Cognitive computation ought to be fast, efficient and flexible, reusing the same neural mechanisms to operate on many different forms of information. In order to develop neural models for cognitive computation we need to develop neurally-plausible implementations of fundamental operations. If the operations can be applied across sensory modalities, this requires a common form of neural coding. Weber-Fechner scaling is a general representational motif that is exploited by the brain not only in vision and audition, but also for efficient representations of time, space and numerosity. That is, for these variables, the brain appears to support functions $f(x)$ by placing receptors at locations x_i such that $x_i - x_{i-1} \propto x_i$. The existence of a common form of neural representation suggests the possibility of a common form of cognitive computation across information domains. Efficient Weber-Fechner representations of time, space and number can be constructed using the Laplace transform, which can be inverted using a neurally-plausible matrix operation. Access to the Laplace domain allows for a range of efficient computations that can be performed on Weber-Fechner scaled representations. For instance, translation of a function $f(x)$ by an amount δ to give $f(x+\delta)$ can be readily accomplished in the Laplace domain. We have worked out a neurally-plausible mapping hypothesis between translation and theta oscillations. Other operations, such as convolution and cross-correlation are extremely efficient in the Laplace domain, enabling the computation of addition and subtraction of neural representations. Implementation of neural circuits for these elemental computations would allow hybrid neural-symbolic architectures that exhibit properties such as compositionality and productivity.

1   Introduction

Cognitive computation in the brain is fast, efficient and flexible. Emulating this ability would result in extremely important technological advances. A general computational framework should be able to operate on a wide range of content without learning each exemplar. Such a framework should generalize across not only different specific operands but also across sensory domains, providing a general computational language for cortical computation. Mathematical operations are an important aspect of symbolic processing. Because of the combinatorics of these problems, learning each set of operands and the appropriate outcome is not feasible.

This paper argues that

1. The brain represents functions of many quantities, including time, using a common form of coding that we refer to as Weber-Fechner scaling.

2. Some of these quantities can be efficiently computed using the Laplace domain and a neurally-plausible mechanism for approximating the inverse Laplace transform.

3. Computational operations, including translation, convolution, and an analog of cross-correlation, can be efficiently computed in a neurally-plausible way with access to the Laplace domain.

This suggests the hypothesis that the brain uses the Laplace domain as a common computational currency across modalities, enabling reuse of the same neural mechanisms for flexible computations on a range of different kinds of information.

1.1 Weber-Fechner scaling of one-dimensional functions in the brain

In this paper we restrict our attention to one-dimensional quantities defined over the positive real line from zero (or some relatively small value) to some large (effectively unbounded) value. We argue that the brain represents functions over variables with these properties using Weber-Fechner scaling. If the ith receptor has a receptive fields centered at x_i, then we define *Weber-Fechner* scaling to mean that

1. the spacing of adjacent receptors is such that $x_i - x_{i-1} \propto x_i$.

2. the width of the receptive field of the unit at x_i should be proportional to x_i.

These two constraints imply a logarithmic scale internal scale for x, which we label $\overset{*}{x}$ to avoid confusion between external physical variables and internal representations. We refer to this coding scheme as a Weber-Fechner scale because it can readily implement the behavioral Weber-Fechner law [2].

There is good evidence that Weber-Fechner scaling is obeyed in the brain in coding extrafoveal retinal position [7, 21]. In the case of vision, Weber-Fechner scaling can be attributed to the structure of the retinal circuitry. However, Weber-Fechner scaling appears to be more general. Neural evidence [15] suggests that Weber-Fechner scaling applies to neural representations of numerosity. For instance, [14] observed approximately Weber-Fechner coding for numerosity in the activity of PFC neurons during the delay period of a working memory task. Different neurons had different preferred numerosities. The width of the tuning curves went up linearly with the cell's preferred numerosity.[1] Weber-Fechner scaling for numerosity cannot be attributed to a property of a physical receptor.

Behavioral [1] and theoretical work [16] suggests that this Weber-Fechner coding scheme should extend also to functions of remembered time. Indeed, a growing body of neurophysiological evidence suggests that representations of time also obey Weber-Fechner scaling. Figure 1 shows evidence illustrating evidence suggesting that the neural representation of time may obey Weber-Fechner scaling. First, the observation of *time cells* suggests that the brain supports functions of past time [12]. During the delay of a memory task, different time cells fire sequentially at circumscribed periods of the delay. At any moment, observation of the set of active cells provides an estimate of the time in the past at which the delay began. Time cells have now been observed in a variety of brain regions with similar qualitative properties that suggest Weber-Fechner coding. It is known that the width of the receptive fields of time cells increases with delay in the hippocampus [5, 11], medial entorhinal cortex [10], mPFC [19] and striatum [13]. Moreover, the density of preferred times decreases with the delay [11, 10, 13, 19]. Collaborative work to quantitatively assess Weber-Fechner coding in a large dataset of hippocampal time cells is ongoing.

If the neural representation of time obeys Weber-Fechner scaling this is a non-trivial computational challenge. A representation of a timeline must update itself in real time. Because the spacing between t_i and t_{i+1} is different than the spacing between t_{i+1} and t_{i+2}, information would have to flow at different rates for different values of t. This seems neurally implausible. We have proposed a solution to this challenge—updating the Laplace transform of history rather than history itself—that we argue also lends itself readily to efficient and flexible computation.

---

[1]Although they did not assess the spacing in a quantitative way, the number of neurons did go down with preferred numerosity.
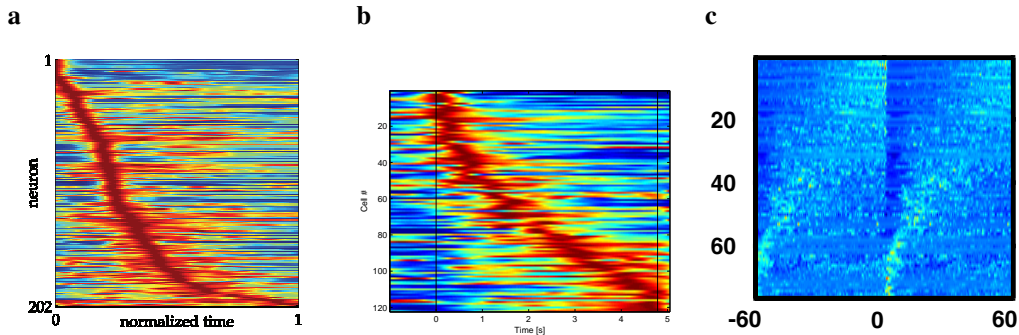
Figure 1: A neural Weber-Fechner scaling for time? In each plot, false color firing rate maps are shown as a function of time for a number of extracellularly-recorded neurons. The neurons are ordered according to their time of peak firing. If the neurons showed linear spacing, these plots would appear as a straight "ridge" of constant width. To the extent the ridges are curved, that implies a decreasing number density of neurons with preferred time of firing, consistent with Property 1. To the extent that the ridges get wider during the delay, this implies an increase in receptive field with preferred time of firing, consistent with Property 2. **a.** Neurons in medial entorhinal cortex during running on a treadmill [10]. The duration of the delay was on average 16 s (see [10] for details). **b.** Neurons in medial PFC during the delay of a temporal discrimination task [19]. **c.** Neurons in the striatum during the 60 s prior to (left) and following reward in a fixed interval procedure [13]. Note that the ordering is such that neurons with time fields earlier in the delay are at the bottom of the figure rather than the top.

2 Constructing Weber-Fechner scale functions of "hidden" variables using the Laplace transform

We have developed a formal mechanism for efficiently representing a Weber-Fechner timeline. The key insight is that while the timeline itself cannot be evolved self-sufficiently in time, the Laplace transform of the timeline can be [16]. The model can be understood as a two-layer feedforward architecture (Fig 2). At each moment a single input node $f(t)$ projects to a set of units $F(s)$ that store the Laplace transform up to the current moment; s indexes the different units. Through a local set of feed forward connections (represented by an operator $\mathbf{L}_k^{-1}$), the second layer approximately inverts the encoded Laplace transform to represent a fuzzy reconstruction of the actual stimulus history itself, $\tilde{f}(\overset{*}{\tau})$. The operator $\mathbf{L}_k^{-1}$ implements the Post inversion formula keeping k terms and can be readily implemented with simple feedforward projections [16].

This simple computational scheme for representing a Weber-Fechner timeline is sufficient to account for canonical behavioral effects in a variety of learning and memory paradigms across species [6]; the long functional time constants necessary to encode $F(s)$ could be computed using known neurophysiological mechanisms [18]. This mechanism can be straightforwardly generalized to represent one-dimensional spatial position, numerosity or any other variable whose time derivative is available at each moment [5]. More precisely, by modulating the differential equations governing the Laplace transform by $\alpha(\tau) = dx/dt$ we can obtain the Laplace transform with respect to x rather than t. This mechanism is sufficient to account for a variety of neurophysiological findings regarding place cells and time cells in the hippocampus [5] and can be generalized to numerosity. For instance, if we initialize a representation with $f(\tau = 0)$ set to a single delta function input, then let it evolve with $\alpha(\tau)$ set to the rate of change of some variable x during an interval T, then at the end of the interval $\tilde{f}(\overset{*}{x}, T)$ will give a scale-invariant estimate of the net quantity x accumulated from time 0 to time T. When $\alpha(\tau)$ is set to zero, the estimate of $\tilde{f}$ stops changing so that α can also be used as a control signal to maintain information in working memory.

As with all path integration models, this approach is subject to cumulative error. That is if $\alpha(\tau) = dx/dt + \eta$, the estimate of $\tilde{f}(\overset{*}{x})$ will grow more imprecise over time. However, note that in the absence of noise, the "blur" in the representation of time, place, and number does not reflect stochastic variability. Rather, the blur is more analogous to a tuning curve with non-zero width.
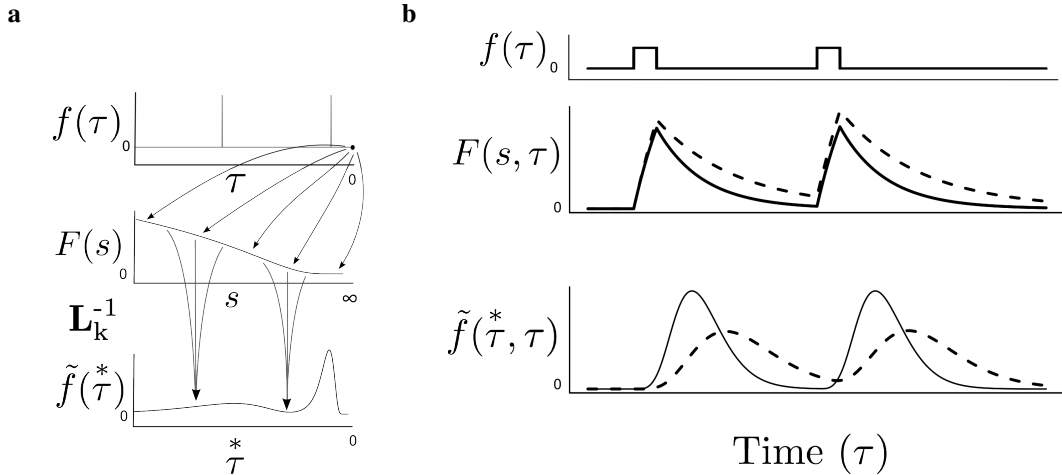
Figure 2: **a.** Schematic of the model for encoding a temporal history $f(\tau)$. At each time step, input from a single node provides input to a set of nodes $F(s)$. Each node of F is indexed by a different value of s which can be identified with the real Laplace variable. Nodes in $F(s)$ project locally to another set of nodes in $\tilde{f}(\overset{*}{\tau})$ *via* an operator $\mathbf{L}_k^{-1}$. The nodes in $\tilde{f}$ approximate the original function $f(\tau)$. The error in $\tilde{f}(\overset{*}{\tau})$ is scale invariant. We choose the distribution of nodes across s and thus also $\overset{*}{\tau}$ to implement Weber-Fechner spacing (not shown). **b.** Nodes in $\tilde{f}$ behave like neural time cells. In this plot the input $f(\tau)$ and the activity of two nodes in $F(s)$ with different values of s and two corresponding nodes in $\tilde{f}(\overset{*}{\tau})$ are shown evolving in time. Note that the units in $F(s)$ behave like charging and discharging capacitors with different rate constants (controlled by their value of s). The units in $\tilde{f}(\overset{*}{\tau})$ behave like neural time cells, responding a characteristic time after the input. The time at which each unit's activity peaks is controlled by $\overset{*}{\tau} = k/s$.

3 Flexible computations in the Laplace domain

If time, space, and number, as well as sensory representations share a common coding scheme, then mechanisms for computing with representations of that form could be reused across many types of information. Here we sketch neurally implementable mechanisms for three operations in the Laplace domain, translation, convolution, and a close analog of cross-correlation. Of these three, translation is the most thoroughly worked out, with a detailed mapping hypothesis onto neurophysiological mechanisms related to theta oscillations [17]. Translation of functions of time can be used to anticipate the future to inform decision-making in the present; translation of functions of other variables can be used to imagine alternative states of the world to inform decision-making in the world in its current state. Convolution and cross-correlation can be used for the addition and subtraction of functions, respectively (among other uses). Because the Post inversion formula is not well-defined for $-s$, we describe an analog of cross-correlation that can be implemented within the neural framework we have developed.

3.1 A neural mechanism for translation *via* hippocampal theta oscillations.

Access to the Laplace domain facilitates flexible translation of one-dimensional representations. A function $f(x)$ can be translated to obtain $f(x + \delta)$ in the Laplace domain *via* a simple point-wise multiplication with the function $\exp(-s\delta)$ where s is the Laplace domain variable. This can be understood in the context of the two layer network as modulation of the synaptic weights in $\mathbf{L}_k^{-1}$ between F and $\tilde{f}$ [22]. Consideration of the computational requirements for translation in the Laplace domain coupled with the hypothesis that hippocampal theta phase precession implements translation leads to several results [17].

The resulting neural model accomplishes translation across scales and at the same time explains and organizes a broad variety of neurophysiological findings related to hippocampal theta oscillations. The hypothesis is that theta oscillations implement translation from zero to some large value within each theta cycle. This successive translation of the present into the past enables prediction of the fu-

a   **b**

$$f(\tau)$$
$$\alpha(\tau) \longrightarrow \Downarrow$$
$$F(s)$$
$$\delta(\tau) \longrightarrow \mathbf{L_k^{-1}} \Downarrow$$
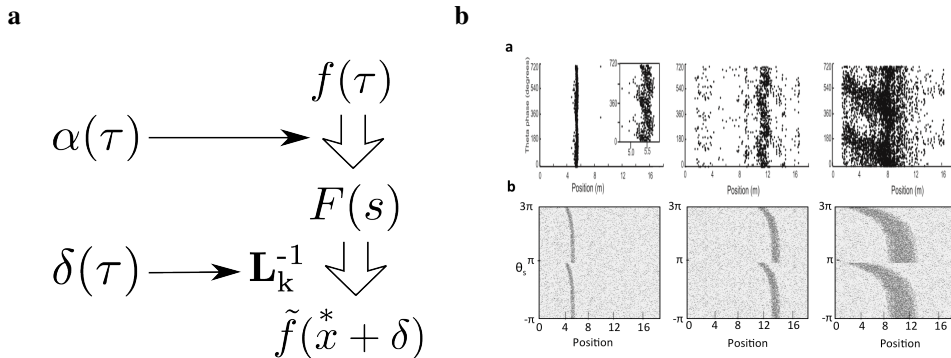$$\tilde{f}(\overset{*}{x} + \delta)$$

Figure 3: A neurophysiological mechanism for translation of one-dimensional functions exploiting theta oscillations. **a.** A generalized circuit for computing and translating scale-invariant functions of one-dimensional variables. $\alpha(\tau)$ enables the same circuit to represent functions of time or any other one-dimensional quantity for which the time derivative is available. Thus, if $f(\tau)$ can be rewritten as $f(x(\tau))$ and $\alpha(\tau) = dx/d\tau$, then the reconstruction is with respect to x rather than τ and we write $\overset{*}{x}$. δ provides a mechanism to translate the function. **b.** Theta phase precession shows properties resembling translation to different future points of the trajectory within a theta cycle. Top: neurophysiological data from [9]. Place cells from different positions along the dorsoventral axis of the hippocampus have place cells of different size. However, cells at all scales still precess over the same range of phases. Bottom: model predictions show the same qualitative patterns [17].

ture at successively more distant points. This model accounts for the finding that all scales (different values of s) phase precess through the same range of local theta phases (Fig. 3). Moreover, coherent translation *requires* that both past time (controlled by the values of s) and future time (controlled by the rate at which δ changes within a theta cycle) obey Weber-Fechner scaling. Finally, the model predicts that cells coding for predicted events should ramp up their firing from the time at which the prediction becomes available to the time at which the predicted stimulus is obtained, phase precessing through at most one theta cycle. This prediction is analogous to findings for neurons in the ventral striatum [20].

We found good evidence aligning translation of functions of space and time from 0 to some large value of δ to neurophysiological findings during hippocampal theta oscillations. However, translations with other properties could be implemented during other neurophysiological events. For instance, translation by negative values would correspond to search through memory for the past; translation to a single non-zero value of δ (rather than sweeping through a range of values) would facilitate retrieval of a memories at a specific range of past times [4]. In two spatial dimensions, one can imagine a series of translations tracing out an imagined path in a navigation task [8]. In the visual modality, translation could be used to simulate planned (or imagined) eye movements or motion of objects in the world. Although these translations could have somewhat different neurophysiological signatures, they are all computationally related to one another. And in all cases, the translation facilitates decision-making and behavior in the present by enabling examination of imagined states of the world.

3.2 Arithmetic operations on functions through parallel computations

Access to the Laplace domain facilitates operations other than translation. In the same way that point-wise multiplications in the Laplace domain can be achieved in a parallel fashion to implement translation of any function, it is also possible to perform *addition* and *subtraction* operations on any two functions by point-wise parallel computations with similar efficiency in the Laplace domain. For this, we start with a definition of the operations addition and subtraction on numbers represented by distribution functions.

Let $f(x)$ and $g(x)$ be functions representing two distributions of possible values for the number x in the range 0 to x_{max}. Outside this range, the functions are assumed to vanish. We shall define the operation of 'addition' of these two distributions to be $[f + g](x)$ to be the convolution of the two

functions.

$$[f + g](x) \equiv \int_0^\infty f(x')g(x - x')\,dx'$$

The justification for this definition is rather straightforward. By considering the two functions to be Dirac delta functions at two different positions, x_1 and x_2, note that $[f + g]$ is a Dirac delta function at $x_1 + x_2$. Moreover, the addition operation is bilinear with respect to the two functions, and hence the above generalized definition for addition is justified. Importantly, since we have access to the Laplace transform of the functions, namely $F(s)$ and $G(s)$, the addition operation can be performed in the Laplace domain. The Laplace transform of $[f + g]$ is simply the point wise multiplication of $F(s)$ and $G(s)$, which can be computed in a parallel fashion, independently for each s value. Finally, the $\mathbf{L}_k^{-1}$ operator can be employed to invert the Laplace transform of $[f + g]$ and obtain a fuzzy addition operation.

It is easy to convince oneself that subtraction operation can similarly be defined to be[2]

$$[f - g](x) \equiv \int_0^\infty f(x')g(x' + x)\,dx'$$

By defining a reflected function $g_r(x) = g(x_{max} - x)$, it can be seen that the Laplace transform of $[f - g]$ is simply the point wise multiplication of the Laplace transform of $f(x)$ and $g_r(x)$. A point of subtlety here is that for the subtraction operation, we have to consider both positive and negative values of x although the two functions are assumed to be non vanishing only for $x > 0$. However, noting that $[f - g](x) = [g - f](-x)$ for positive x, we can perform the subtraction operation for negative x values also. In this entire process, only positive values of s are utilized, and hence the inverse Laplace operator $\mathbf{L}_k^{-1}$ is always well defined and the entire process can be performed in parallel.

We have not yet carefully considered the neurophysiological substrates that could support these arithmetic operations. However, the computational efficiency of performing these operations in the Laplace domain is considerable. Given these considerations, it may be reasonable for the brain to encode the Laplace transform even for variables that are provided along a Weber-Fechner scale due to the property of the sensory receptors.

4   Discussion

We suggest that the brain uses a common form of coding, Weber-Fechner scaling, to represent unbounded one-dimensional quantities. It can be shown that Weber-Fechner scaling is an optimal response to signals that are long-range correlated, which are found throughout the natural world [16]. Weber-Fechner scaling allows for representation of exponential scales with linear resources.

Representation of variables such as time, space and numerosity is greatly facilitated by access to the Laplace transform. Many computations can be efficiently performed in the Laplace domain. For instance, translation of representations of space and time toward the past can be used to estimate the future. Recent work has developed a detailed mapping between a translation operator and hippocampal theta oscillations. We sketched implementable operations for addition and subtraction of functions on a Weber-Fechner scale. These operations could be used for combining functions, or for comparing one function to another. Because the outcome of a specific operation does not need be learned, but can be computed on-line, the existence of these operations provides an important step towards endowing neural systems with the properties of productivity and compositionality that are taken to be essential aspects of symbolic computation and cognition more broadly [3]. For instance, it is clear that arithmetic obeys the properties of compositionality and productivity (modulo edge effects). If the result of an addition operation is a function with the same neural code as the addends, then one can in principle represent an effectively infinite number of possible problems. For instance, given only two input functions f and g one could compute $f + g$, or $(f + g) + g$, or $(f + f) + g$, etc.

There are several design considerations that are important in developing this into a general framework for cognitive computation. The first consideration is whether computation for different information should be performed in a central location, as in a von Neumann architecture or performed

---

[2]The challenge of this approach is that the Post inversion formula does not work when the transform is growing exponentially as with $-s$. If that were not the case, cross-correlation would suffice to implement subtraction.

locally. The answer may depend on the form of operation. Consider Fig. 3a. Different settings for $\alpha(\tau)$ and different settings for $f(\tau)$ can give rise to a very broad range of representations, corresponding to a broad taxonomy of cells in the hippocampus and related structures [5]. All of these representations can be translated by modulating the same weights used to construct the representation (modulation by δ). Here the control signal for translation is a scalar per representation and the output of the computation can be written to the same cells that are used to hold the representation itself.[3] This means that the cost of local implementation of translation is small per translatable function. In contrast, addition and subtraction operators require additional resources to hold the output of the computation. The storage cost of implementing this operation locally would be relatively substantial. Moreover, because there are many pairwise combinations of representations that might need to be combined, there is in addition a considerable wiring cost associated with local processing. For these reasons addition and subtraction of functions ought not to be performed locally.

Acknowledgments

References

[1] BALSAM, P. D., AND GALLISTEL, C. R. Temporal maps and informativeness in associative learning. *Trends in Neuroscience 32*, 2 (2009), 73–78.

[2] FECHNER, G. *Elements of psychophysics. Vol. I.* Houghton Mifflin, 1860/1912.

[3] FODOR, J. A., AND PYLYSHYN, Z. W. Connectionism and cognitive architecture: A critical analysis. *Cognition 28*, 1 (1988), 3–71.

[4] FOSTER, D. J., AND WILSON, M. A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature 440*, 7084 (2006), 680–3.

[5] HOWARD, M. W., MACDONALD, C. J., TIGANJ, Z., SHANKAR, K. H., DU, Q., HASSELMO, M. E., AND EICHENBAUM, H. A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *Journal of Neuroscience 34*, 13 (2014), 4692–707.

[6] HOWARD, M. W., SHANKAR, K. H., AUE, W., AND CRISS, A. H. A distributed representation of internal time. *Psychological Review 122*, 1 (2015), 24–53.

[7] HUBEL, D. H., AND WIESEL, T. N. Uniformity of monkey striate cortex: a parallel relationship between field size, scatter, and magnification factor. *Journal of Comparative Neurology 158*, 3 (1974), 295–305.

[8] JOHNSON, A., AND REDISH, A. D. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience 27*, 45 (2007), 12176–89.

[9] KJELSTRUP, K. B., SOLSTAD, T., BRUN, V. H., HAFTING, T., LEUTGEB, S., WITTER, M. P., MOSER, E. I., AND MOSER, M. B. Finite scale of spatial representation in the hippocampus. *Science 321*, 5885 (2008), 140–3.

[10] KRAUS, B. J. *Time and distance coding by the hippocamus and medial entorhinal cortex.* PhD thesis, Boston University, 2012.

[11] KRAUS, B. J., ROBINSON, 2ND, R. J., WHITE, J. A., EICHENBAUM, H., AND HASSELMO, M. E. Hippocampal "time cells": time versus path integration. *Neuron 78*, 6 (2013), 1090–101.

[12] MACDONALD, C. J., LEPAGE, K. Q., EDEN, U. T., AND EICHENBAUM, H. Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron 71*, 4 (2011), 737–749.

[13] MELLO, G. B., SOARES, S., AND PATON, J. J. A scalable population code for time in the striatum. *Current Biology 25*, 9 (2015), 1113–1122.

[14] NIEDER, A., AND MERTEN, K. A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *Journal of Neuroscience 27*, 22 (2007), 5986–93.

---

[3] This is possible because the original untranslated function can be recovered simply by setting $\delta = 0$.

[15] NIEDER, A., AND MILLER, E. K. Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron 37*, 1 (2003), 149–57.

[16] SHANKAR, K. H., AND HOWARD, M. W. Optimally fuzzy temporal memory. *Journal of Machine Learning Research 14* (2013), 3753–3780.

[17] SHANKAR, K. H., SINGH, I., AND HOWARD, M. W. Neural mechanism to simulate a scale-invariant future. *arXiv preprint arXiv:1503.03322* (2015).

[18] TIGANJ, Z., HASSELMO, M. E., AND HOWARD, M. W. A simple biophysically plausible model for long time constants in single neurons. *Hippocampus 25*, 1 (2015), 27–37.

[19] TIGANJ, Z., KIM, J., JUNG, M. W., AND HOWARD, M. W. Temporal coding across scales in the rodent mPFC. *Cerebral Cortex* (In revision).

[20] VAN DER MEER, M. A. A., AND REDISH, A. D. Theta phase precession in rat ventral striatum links place and reward information. *Journal of Neuroscience 31*, 8 (2011), 2843–54.

[21] VAN ESSEN, D. C., NEWSOME, W. T., AND MAUNSELL, J. H. The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Research 24*, 5 (1984), 429–48.

[22] WYBLE, B. P., LINSTER, C., AND HASSELMO, M. E. Size of CA1-evoked synaptic potentials is related to theta rhythm phase in rat hippocampus. *Journal of Neurophysiology 83*, 4 (2000), 2138–44.