NEWS AND VIEWS

Neocortex: a lean mean memory storage machine

Beatriz E P Mizusaki, Armen Stepanyants, Dmitri B Chklovskii & P Jesper Sjöström

Connectivity patterns of neocortex exhibit several odd properties: for example, most neighboring excitatory neurons do not connect, which seems curiously wasteful. Brunel's elegant theoretical treatment reveals how optimal information storage can naturally impose these peculiar properties.

The connectivity patterns among neurons are a key determinant of brain computations. As a wealth of information on neocortical microcircuit connectivity patterns has accumulated over the past decade[1–4], several bewildering properties have emerged. First, connectivity among neighboring excitatory pyramidal cells is sparse, so that most pairs of nearby neurons are not actually connected. Second, bidirectionally connected pairs of neurons are generally overrepresented compared to what one would expect from a random network. Finally, reciprocally connected neurons are also coupled with particularly strong synapses. This arrangement seems mysteriously inefficient. For example, why do most neighboring excitatory neurons not talk directly to each other when they are physically capable of doing so[4]?

In this issue of *Nature Neuroscience*, Brunel[5] proposes a plausible explanation for these unexpected features. By building on his earlier work on cerebellar microcircuits[6], he argues that all these properties may arise solely from neocortical microcircuits being optimized for

Beatriz E.P. Mizusaki is at the Instituto de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil. Armen Stepanyants is in the Department of Physics and Center for Interdisciplinary Research on Complex Systems, Northeastern University, Boston, Massachusetts, USA. Dmitri B. Chklovskii is at the Simons Center for Data Analysis, Simons Foundation and New York University Neuroscience Institute, New York, New York, USA. P. Jesper Sjöström is at the Centre for Research in Neuroscience, Department of Neurology and Neurosurgery, Program for Brain Repair and Integrative Neuroscience, The Research Institute of the McGill University Health Centre, Montreal General Hospital, Montreal, Quebec, Canada.
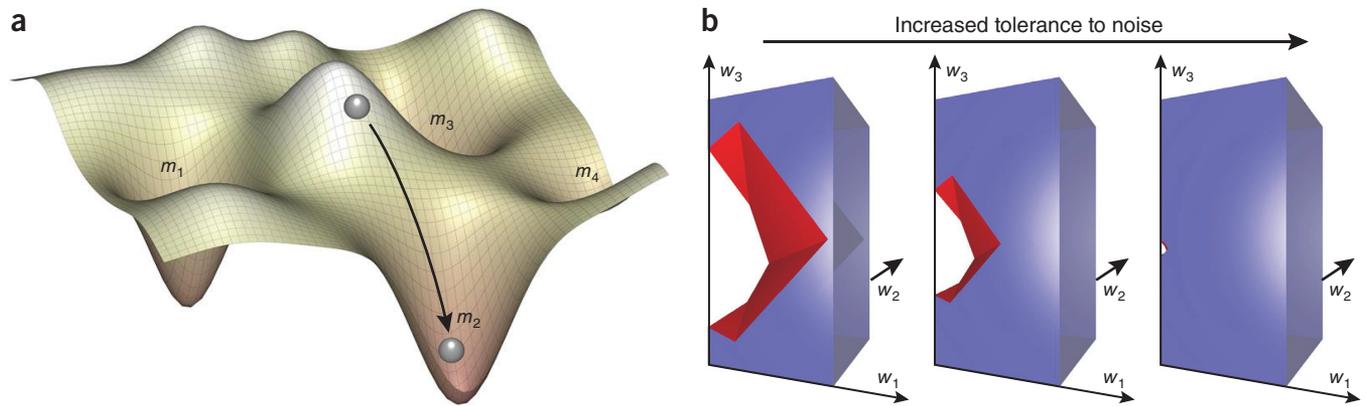e-mail: jesper.sjostrom@mcgill.ca

information storage. To show this, he used a combination of analytical mathematics and computer simulations with the assumption that the cortical microcircuit can be viewed as an attractor network that operates at maximum information storage capacity.

In symmetrically connected recurrent neural networks[7], stored memories can be thought of as stable neuronal activity patterns, known as attractor states. These patterns are stable because they correspond to local minima in an energy landscape, and the neural network is therefore attracted to them much like a ball rolling down a hill (**Fig. 1a**). The specific neuronal activity pattern of each attractor is defined by the weights of synapses linking neurons so that each memory is distributed across the network. The network also has to accommodate multiple patterns, which means the permissible weight space occupies a small volume in the total space of all possible weight values. For each new memory added, another constraint is imposed on the weights, so the permissible space shrinks as more information is stored. Even in a small recurrent network, there is a large number of these synaptic weights, so this space is of high dimensionality; to provide intuition, we illustrate only three dimensions (**Fig. 1b**).

In the presence of noise, the volume of permissible weights shrinks even further because the neural network then needs a safety factor to ensure correct recollection of its memories. It is unclear precisely how much noise to add in the theoretical treatment to emulate its neurobiological counterpart, since what is noise and what is signal in real neurons is subject to considerable debate[8]. What emerges clearly in Brunel's theory[5], however, is that as the volume of permissible weights shrinks with increasing safety factor, it will invariably end up at a point with most weights being zero

and only a few nonzero weights remaining (**Fig. 1b**). This means that in an optimal neural network that is operating at maximal capacity and with maximal tolerance to noise, most weights have to be zero for memory retrieval to function correctly.

Because zero-valued synaptic weights translate into ineffectual connections, this implies that most neighboring pairs of neurons should not be connected. This finding helps explain why many neighboring neurons do not connect with functional synapses even though they are so close that their axons and dendrites can touch[4].

This sparsity of connectivity has been shown before, for example, by Brunel himself for cerebellar Purkinje cells[6]. In the present study[5], Brunel extends his theoretical treatment to reveal that satisfying these conditions of optimality also leads to several other properties that have already been experimentally found in neocortical microcircuits[1–3]. These include an over-representation of reciprocally connected pairs of neurons and, furthermore, stronger synaptic strength between such bidirectional connections. These predicted properties closely match those found experimentally in young visual cortex layer 5 pyramidal cell connections[1], suggesting that the neocortex may be optimized for information storage.

Empirically, connectivity patterns differ between cortical areas[1,2]. Why might this be so, if these different networks are meeting the same optimality principles? One idea is that statistics of connectivity also depend on the functional nature of stored patterns[3]. For example, the over-representation of bidirectionally connected neurons could result from the persistency of neuronal representations[9,10]. Such persistency may vary among different cortices. To provide a specific example, because the autocorrelation

Figure 1 Memory storage in recurrently connected neural networks is optimal when most synaptic strengths are zero. (**a**) Each memory (m_1, m_2, m_3, . . .) can be thought of as the bottom of a well in an energy landscape[7]. Starting out at a high energy state, the neural network subsequently falls into the closest basin of attraction, in this example recalling memory m_2. (**b**) To store n memories, synaptic weights (w_1, w_2, w_3) must be in a permissible space (hollow region in blue box) constrained by n planes (red), but as the tolerance of the neural network to noise is increased, this space shrinks, eventually becoming a single point. Brunel shows[5] that this optimal point corresponds to most synaptic weights being zero (w_1 and w_2 here): that is, most neighboring neurons should not be connected at optimality, which is precisely what has been found in neocortical microcircuits[1–4].

time scale of visual stimuli may be greater than that of whisker stimulation, there is a longer persistency of visual representations. Taken together, this could explain why reciprocal connections between pyramidal cells are over-represented in visual[1] but not in barrel cortex[2]. To address the statistics of bidirectional connections in barrel cortex[2], Brunel analyzed a different scenario with activity patterns that changed sequentially in trajectories[7] and found that reciprocal connections were indeed no longer over-represented, in agreement with the experimental findings in barrel cortex[2].

However, Brunel ends the title of his paper—"Is cortical connectivity optimized for storing information?"—with a question mark, implying a degree of hesitation. There may be good reason for this hesitation because the match with the biological data is not perfect. For example, the synaptic weight distribution has a long tail[1] that Brunel's theory does not predict[5], yet other theories do while covering some of the other predictions as well[10,11]. It is, in other words, possible to predict different values for the same statistics with other starting assumptions about what should be optimal.

Of course, the neocortex may not be optimal for information storage. In fact, it was recently reported that, surprisingly, pyramidal cells in visual cortex of mature animals do not seem to interconnect at all, neither bidirectionally nor unidirectionally[12]. The biological data that Brunel relied on for validation of his theory were collected from juvenile animals[1], so there are developmental and other experimental caveats to consider here too, making it hard to pin down the ground truth with certainty.

Nevertheless, the history of optimization approaches in biology suggests that we are better off trying to answer "why" questions using optimization theories than not[13]. If one wants to make sense of the zoo of biological observations and construct a systematic theory of biological structure and function that also tells us why things should work the way they do, there are not many alternatives.

So where do we go next? One could increase the biological complexity of the theoretical models in an attempt to explain a larger set of experimental observations and to generate predictions of greater specificity. It would be interesting to see whether Brunel's theory could be extended to multiple classes of excitatory and inhibitory neurons and whether it could incorporate constraints arising from morphological differences and homeostatic regulation of connectivity[14,15]. In this way, theoretical predictions would be put to the test against the ever-increasing data on connectivity among different neuronal classes[1–4,12].

As scientists we derive satisfaction from explaining puzzling properties of living matter. Brunel's theory is elegant in its appealing conceptual simplicity, and the overall match with the biological data is intriguing. It is also quite pleasing to find concrete evidence that our brains are actually optimal.

1. Song, S., Sjöström, P.J., Reigl, M., Nelson, S. & Chklovskii, D.B. *PLoS Biol.* **3**, e68 (2005).
2. Lefort, S., Tomm, C., Floyd Sarria, J.C. & Petersen, C.C. *Neuron* **61**, 301–316 (2009).
3. Ko, H. *et al. Nature* **473**, 87–91 (2011).
4. Chklovskii, D.B., Mel, B.W. & Svoboda, K. *Nature* **431**, 782–788 (2004).
5. Brunel, N. *Nat. Neurosci.* **19**, 749–755 (2016).
6. Brunel, N., Hakim, V., Isope, P., Nadal, J.P. & Barbour, B. *Neuron* **43**, 745–757 (2004).
7. Hopfield, J.J. & Tank, D.W. *Science* **233**, 625–633 (1986).
8. Stein, R.B., Gossen, E.R. & Jones, K.E. *Nat. Rev. Neurosci.* **6**, 389–397 (2005).
9. Clopath, C., Büsing, L., Vasilaki, E. & Gerstner, W. *Nat. Neurosci.* **13**, 344–352 (2010).
10. Druckmann, S. & Chklovskii, D.B. *Curr. Biol.* **22**, 2095–2103 (2012).
11. Varshney, L.R., Sjöström, P.J. & Chklovskii, D.B. *Neuron* **52**, 409–423 (2006).
12. Jiang, X. *et al. Science* **350**, aac9462 (2015).
13. Parker, G.A. & Maynard Smith, J. *Nature* **348**, 27–33 (1990).
14. Chapeton, J., Gala, R. & Stepanyants, A. *Front. Comput. Neurosci.* **9**, 74 (2015).
15. Chapeton, J., Fares, T., LaSota, D. & Stepanyants, A. *Proc. Natl. Acad. Sci. USA* **109**, E3614–E3622 (2012).