

Heterogeneity and Efficiency in the Brain

Vijay Balasubramanian

Abstract—The brain carries out enormously diverse and complex information processing operations to deal with a constantly varying world on a power budget of about 12W. I argue that this efficiency is achieved in part through the dedication of specialized circuit elements and architectures to specific computational tasks, in a hierarchy stretching from the scale of neurons to scale of the entire brain, in sharp contrast to the conventional von Neumann architectures. This article suggests that the heterogeneous computational repertoires of the brain are architectural memories of efficient computational procedures that are learned via evolutionary selection.

Index Terms—neuromorphic computation, visual processing, dynamic circuit reorganization, hierarchical information representation, circuit repertoires

I. INTRODUCTION

THE brain is an extremely complex computational engine. Its 100×10^9 neurons and 10^{14} synapses are densely connected so that a mm^3 of volume can contain 4 km of wire [1]–[3]. Energetically, the brain is the most expensive tissue in the body – it is 2% of body weight, but 20% of metabolic load, more expensive per gram than muscle when you are working out, suggesting that there will be evolutionary pressure towards computational efficiency [4]–[8]. On the other hand, the brain consumes a mere 20W of power – about the same as a refrigerator lightbulb [4], [5] – and uses this to (nearly) beat supercomputers at chess, produce art and music, store memories of a lifetime, experience emotions like love and anger, learn from experience, and build skyscrapers and nanoscale devices alike. How does it manage to do all this on such a meagre budget, with sloppy biochemical circuits, and a fraction of the component density that can be packed into a microchip?

I argue here that a key lies in the heterogeneity and diversity of circuit elements and architectures at every scale from neurons to the whole brain [1], [9]. An economic analogy is helpful – the specialization of occupations as human societies progress from hunter-gatherer to more complex forms is thought to squeeze out greater efficiencies as individuals become expert in specific tasks. In a like manner, I suggest, the brain achieves its efficiencies by adapting its circuit elements and architectures over evolutionary time, during development, and via learning, to the structure of the natural world, the underlying logic of computations and the cognitive tasks to be performed. In this view, the bewildering repertoire of neural types (more than 70 in the retina alone [10], [11]) and architectures is a memory of computations that have predictive value for behavior, learned over evolutionary time, encoded in the genome and developmental program, and then shaped

by experience-driven learning. These are memories of “procedures” rather than “facts”, but are memories nevertheless.

In what follows I will try to support these assertions by giving several examples. First, I will review how the diverse components of the nervous system are organized in a hierarchy of specialized modules, each dedicated to specific functions, which interact together to produce animal behavior. Then I will describe a general argument from information theory that suggests why distributing function in this way conserves resources. Next, I will consider in turn the functional logic of a sensory system (the sense of sight) and a cognitive system (the sense of place) in the brain. In each case, I will try to argue that the structure of circuits can be understood as an adaptation to efficiently process information under conditions of constraint.

In this article, I will focus primarily on efficiency in the sense of resource minimization. Of course, neural systems have many other desiderata – they should represent information in ways that are easy for the rest of the brain to process, they must be quick enough to support behavior in an uncertain and changing world, and they should be flexible, adaptable and evolvable. Engineered systems dealing with complex natural environments via adaptive responses and self-organizing architectures will face the same challenges in their functional organization. One could discuss how and to what extent heterogeneity of components also serves these goals, but given the space available I will focus largely on how component diversity reduces the resources required for computation. I will extrapolate from this argument in the conclusion, and suggest that engineered computational devices will need to escape from the hegemony of the von Neumann architecture to achieve the sort of efficiency that the brain shows us is possible.

II. HETEROGENEITY AND HIERARCHY IN THE BRAIN

Consider first the whole brain. In antiquity, the Egyptians observed that construction workers who experienced localized trauma to their skulls experienced specific sensory, motor or cognitive deficits suggesting that the brain was the seat of the mind (contrary to contemporary belief), and a localization of function within the brain (see description of the Smith Papyrus in [12]). In the modern era, the idea of localization came to the fore in the 19th century, with studies like those of Broca, who showed through patient studies that speech processing is specific to a particular small region of the cortex. Subsequent work, in the following one and a half centuries, has firmly established the idea of localization – in normal human brains the various sensory, cognitive and motor functions are heavily focused in specific brain areas. For example, visual stimuli are processed in the occipital lobe in the back of the head,

V. Balasubramanian is with the Department of Physics, David Rittenhouse Laboratory, University of Pennsylvania, Philadelphia, PA 19104, email: vijay@physics.upenn.edu

Digital Object Identifier: 10.1109/JPROC.2017.2447016

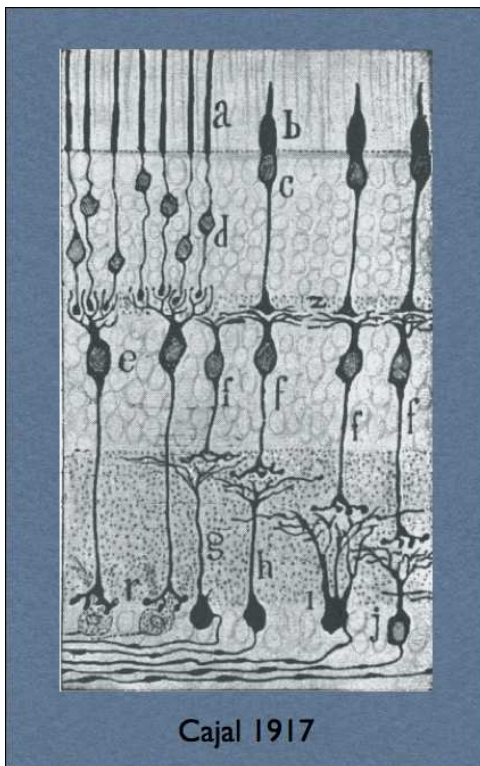


Fig. 1: Drawing of the feedforward circuitry of the retina by Ramón y Cajal. In humans the *photoreceptors* (rods and 3 types of cones) feed into ~ 10 types of *bipolar cells* which carry out diverse computations and communicate by analog signals. These feed into ~ 20 types of *ganglion cells* (the output cells of the retina that select visual features and communicate them to the central brain as digital signals sent over the optic nerve). This image leaves out the ~ 40 types of highly diverse inhibitory interneurons (*horizontal cells* and *amacrine cells*) that run laterally and synapse with bipolar cells and ganglion cells over many different spatial scales. The interneurons serve many different computational functions including carving out the features relevant for behavior (e.g. by making responses of certain ganglion cells direction selective), and adaptation to changing environmental conditions (e.g. by dynamically changing the responses of retinal circuits to cope with massive illumination differences between noon and dusk).

speech is processed by Broca's area, motor control is exerted by the motor cortex (a strip running down the middle of the head from the top towards the ear), the hippocampus in the mid-brain is critical for episodic memory, and planning and decision-making involve the pre-frontal cortex in the forehead [13]. Thus, the large-scale functions of the brain are organized in interacting, specialized modules.

If we go down a scale and look at the circuit structure within each area, we find further specialization. The visual cortex is anatomically and functionally segregated into areas V1 to V4, where V1 is mostly engaged in extracting low-level image features such as edges, which are pooled in some as yet unclear way to produce higher level features such as shape elements in V4. More broadly, the cortex divides different

kinds of visual information into separate streams – e.g., the dorsal stream up the back of the head processes motion, while the ventral stream down the base of the cortex analyzes shapes. Likewise, motor cortex is separated into regions responsible for controlling muscles in different parts of the body – the face is controlled by circuits mid-way down motor cortex, the knee is controlled by circuits at the top. This is why a localized stroke can damage control of a particular part of the body and leave others alone.

We can look more closely at each of the regions and examine the neurons and circuits that carry out the required computations. To be specific, consider the retina, a light-sensitive piece of the central brain that emerges from the neural tube during fetal development and makes its way to the front of the head. Starting with exquisite anatomical studies by the great Ramón y Cajal (Fig. 1), we have come to know that the retina achieves the basic task of converting light into signals that are interpreted by cortex by implementing an intricate, precisely wired three-layer circuit consisting of over 70 distinct types of neurons [10], [11]. First, the rods and cones (3 types in humans) transduce light into electrical signals. These feed forward to ~ 10 species of bipolar cells carrying out various analog computations (e.g. selection of bright spots vs. dark spots) that begin the process of feature extraction. These in turn feed into ~ 20 types of ganglion cells whose axons form the optic nerve, and which signal the local presence of visual features (e.g., bright/dark spots, color, and directed motion) in digital voltage signals (so-called “action potentials” or “spikes”). Running laterally between the first and second feed-forward stages are a plethora of “interneurons” – these circuit elements, generally inhibitory, carve away irrelevant parts of the sensory input to extract the visual features used for perception and to compute their strength. All told there are over 70 types of circuit elements, each carrying out specific computations, which can adapt to varying environmental conditions like the overall light level (which can change by 10 orders of magnitude between noon and dusk). This sort of ordered architecture with a heterogeneous repertoire of circuit elements is the norm in the brain [1], [14].

Thus, at each scale, from the whole brain to single cells, neural circuits are composed of highly heterogeneous components with specific functions, precisely coupled into circuits that aggregate hierarchically to produce animal behavior (Fig. 2). In fact, we could pursue this theme into the sub-cellular scale and discuss the varieties of neurotransmitters (i.e. chemical messengers between neurons), receptors (i.e. sensors of ions, neurotransmitters, hormones and other signals), and other molecular components. Experience of life modifies these circuits and rearranges connectivity at timescales from milliseconds to years. This learning has an adaptive purpose, and better enables the brain to carry out functions that benefit the organism in light of past history. Thus, circuits in the brain encode a procedural memory of the past, the better to predict the future and guide action.

But why is there such enormous diversity in the components? Why not have a small number of powerful computational units that take in a lot of data and flexibly compute many sorts of things as in the von Neumann architecture

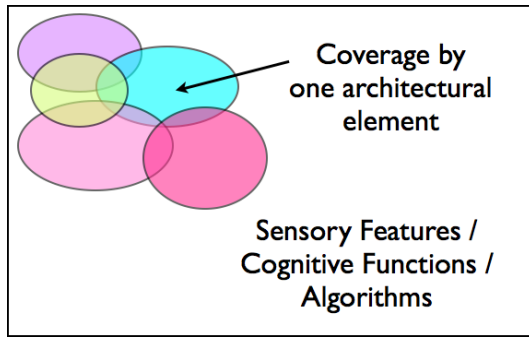


Fig. 2: At every scale of organization from neurons to the whole brain, neural circuits are composed of heterogeneous components (indicated here as colored ovals) specialized to carry out a limited portion of the overall task or computation. The elements of this diverse functional repertoire communicate with each other in precisely organized circuits laid out during the development of the brain, or organized by experience-dependent learning. In the sensory periphery (e.g. the retina), the components of the functional repertoire include diverse cells specialized to compute and communicate specific features of the sensory input. In terms of the whole brain one might similarly regard the different brain areas as functional modules specialized to carry out particular algorithms and computations efficiently, communicating the results across large-scale brain networks to implement animal behavior.

which has been dominant in computer design? Alternatively, why not have a gigantic, maximally connected network of homogeneous units that learn, and holistically produce, the brain’s functions in a grandly connectionist manner? Indeed, great neuroscientists such as Camilo Golgi at the turn of the 20th century famously argued for the latter picture in a long-running debate with Ramón y Cajal, who instead proposed a picture of heterogeneous and localized computation. There may be many factors driving the heterogeneity of components and architectures, not least the default explanation in biology for “why” something is as it is – evolution is history and living things are full of evolutionary accidents. But evolution is not just a random process – it involves natural selection of fitter organisms. While the “objective function” of this selection is not established, there is a case to be made that the heterogeneity of circuit architectures and cell types in the brain reduces the cost of computation.

To see this, consider the general structure of information transmission by channels with a power constraint [15], [16]. We imagine a noisy channel which transmits an input signal S as an output signal X . We imagine that X is transmitted as a sequence of symbols drawn from an alphabet (x_0, x_1, \dots, x_N) with energetic costs $E_0 < E_1 < \dots < E_N$. As a model of neurons as information channels, x_0 can represent silence (which has a baseline cost E_0 associated to protein synthesis and other cellular functions), while x_1 is a single action potential, x_2 is two action potentials and so on within a specified time bin. Let $I(S, X) = H(X) - H(X|S)$ be the mutual information between S and X with H being the entropy of the indicated quantities. On general grounds there is a law of diminishing

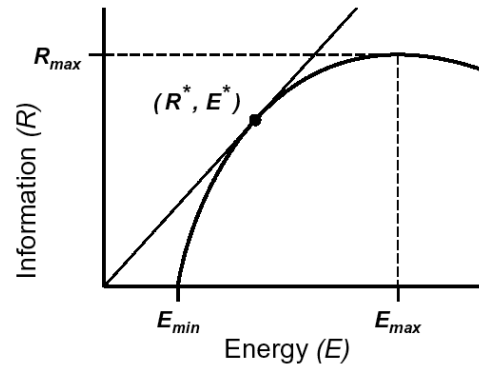


Fig. 3: Information Rate vs. Power for a channel where there is baseline cost E_{min} for operating the channel and communication symbols cost energy to use. The information curve is convex down indicating a law of diminishing returns – twice the energy gives less than twice the information [15], [16]. Bits/Energy at any point on the curve is given by the slope of a line from the origin to that point. Thus Information/Energy is maximized at an operating point (marked R^*, E^*) determined by the tangent from the origin. (Figure adapted from [15].)

returns – using twice the power increases the information rate by less than a factor of two [15], [16] (Fig. 3). Thus, the information/power ratio is maximized at an operating point determined by the tangent from the origin to a point on the Information vs. Power curve (Fig. 3). For neurons, power consumption is directly related to the firing rate, suggesting that it is beneficial from an information transmission point of view to break up information into types that generate information at a rate near this optimal operating point that will be determined by the energetics of neural biochemistry [1], [7]–[9], [15], [17]–[22], a point first emphasized in [17].

To illustrate this conclusion in a simple, concrete model, imagine that a particular region of the brain must transmit information at a rate of I bits/second in order to support downstream computations and behavior. Assume that this information is broken up into different “features” processed by k distinct functional elements which act as independent information channels. Assume that there is a law of diminishing returns relating information rate (I) and cost (C) so that $I = b + g(C)$ with $d^2g/dC^2 < 0$, as illustrated for an energy cost function in Fig. 3. We can argue generally for a law of diminishing returns as follows. Suppose you can transmit information at a maximum rate I_1 by paying a cost C_1 , and at a maximum rate $I_2 > I_1$ by paying a cost $C_2 > C_1$. Then if you are able to pay a cost $C_m = (C_1 + C_2)/2$, you can at least achieve an information rate $(I_1 + I_2)/2$ by transmitting with a cost C_1 half the time and with a cost C_2 half the time. This establishes that the maximum information rate at a cost C_m is at least as big as $(I_1 + I_2)/2$. Iterating this argument establishes a law of diminishing returns relating information and cost. We can write such a law of diminishing returns equivalently in terms of the cost per unit time for communication in any of these channels by writing $C(I) = a + f(I)$ where a is a baseline cost for maintaining the channel and $f(I)$ is a concave function so that $d^2f/dI^2 > 0$.

We want to find a partition into k channels that minimizes the total cost of transmitting information at a rate I . Since we are taking the channels to be independent for this simple demonstration, we have $I = \sum_i^k I_k$ and the total cost is the sum of the costs of the individual channels $C = \sum_{i=1}^k C_i$. Let us take the channels to be identical and neglect other considerations like ease of downstream decodability. Then the symmetry of the problem implies that the optimal solution will have the the same information rate and cost in each channel. So $I_i = I/k$ and $C = ka + kf(I/k)$. The optimum number of channels is determined by setting $dC/dk = a + f(I/k) - (I/k)f'(I/k) = 0$ where f' denotes the first derivative of f . (Here, for simplicity of exposition, we are treating k as a continuous variable which is adequate for our purpose if it turns out that $k \gg 1$.) We can write this equivalently as

$$\frac{I}{k} \left[\frac{d \ln C(x)}{dx} \right]_{x=I/k} = 1 \quad (1)$$

To get a sense for what this means we can examine the solution in specific examples.

Consider a power-law cost $C(x) = a + (bx)^\alpha$ for which the solution to (1) is $k = bI\sqrt{(\alpha - 1)/a}$. Note that: (a) a sharper growth of the cost with information rate (larger α) increases the optimum number of channels; (b) a larger baseline cost a drives down the optimal number of channels. Now consider an exponential cost, $C(x) = ae^{x/b}$. The optimal solution in this case is $k = I/b$. So, rapid growth of the cost as a function of the information rate (small b) increases the optimal number of channels. In this case, the baseline cost ($a = C(0)$) does not affect the optimum. As shown in these examples, a law of diminishing returns relating information and cost will drive the dispersion of information across a diversity of channels that each operates at a lower rate. This analysis can be generalized to include correlations between channels and the features they encode, inhomogeneities in the properties of channels encoding different features, and computational constraints where some partitions of the total information into easily decodable or computable chunks may also be desirable. But the broad point holds that minimizing cost produces a drive to partition the information stream and thus to diversity of function in the brain.

This idea is harder to formalize in terms of *computation* as opposed to information *transmission*. Nevertheless, the general point here is that specialization tends to increase efficiency, and this (partly) drives computational heterogeneity in the brain. To show the usefulness of this way of thinking in understanding the “whys” of neural circuit organization, I will now discuss examples drawn from two very different computational systems in the brain: the sense of sight and the sense of place.

III. THE SENSE OF SIGHT

The previous section described the architecture of the retina and the hierarchy of processing of the retinal output by the visual cortex. By now, it is well established that many features of this circuitry are adaptations – architectural memories, if you will – that increase the efficiency of visual information

processing. Famously, Atteneave, Barlow and Laughlin [23]–[25] argued that lateral connections in the retina are generally inhibitory because, given a limited number of cells, bounded bandwidth, and circuit noise, the retina faces pressure to remove redundancies in its input in order to maximize transmitted information. Indeed, natural visual scenes have long-range pairwise correlations of luminance and color [26]–[28], and one can remove these in the retinal output by appropriately structured lateral inhibition within retinal circuits [29], [30]. We can think of this powerfully in terms of resource minimization. The retina must convey a certain amount of information about light in the world to the brain. If this information is transmitted redundantly by neurons processing different parts of an image, the number of cells or their bandwidth (maximum spike firing rate) will have to increase with the redundancy, incurring costs including increased energy consumption. Thus, redundancy should be reduced to minimize cost, requiring a class of circuit elements (inhibitory interneurons) tasked with remodeling the information representation to remove redundancy.

A more refined version of this argument recognizes that neurons are noisy, and thus redundancy in the retina can support reliable decoding by downstream cortical processes. Quantitatively evaluating this tradeoff gives accurate predictions of the relative range over which the interneurons should provide inhibition relative to the angular resolution of the retinal output neurons [29], [30]. Likewise, the structure of the retinal ganglion cell mosaic and the relative spacing of cells is accurately reproduced [31], [32]. In the primary visual cortex, which receives and processes the retinal output, researchers have further predicted the functional properties of cells (e.g. edge detection) by using Independent Component Analysis to require that different circuit elements should be informationally independent (and not just pairwise decorrelated) [33]–[35]. These authors have emphasized that this sort of redundancy reduction through the carefully selection of circuit elements also leads to sparse information representations (i.e. fewer firing neurons, with lower spike rates) and hence to lower power consumption.

But what of the overall diversity of retinal circuit components? Can the particular functional repertoire of excitatory principal cells (the 3 cone types, the ~ 10 bipolar types in the second layer, and the ~ 20 ganglion cell types in the output layer) be explained in terms of circuit efficiency? Large survey studies have constructed an information budget showing that the bits in the retinal output are broadly distributed across ~ 20 neural types which all have similar firing rates when responding to natural movies (Fig. 4; [36], [37]) despite their very different functional roles (ranging from contrast detection to color detection to motion detection in the cardinal directions). This suggests that the visual input has been carved up into features chosen in such a way that all the output components lie at a similar (perhaps maximally efficient) operating point on the information vs. power curve for neurons (Fig. 3; [15], [16]). A challenge in really settling this point is that the visual features must also be selected to effectively support visual behaviors, and the high-level computations needed to support the repertoire of visual behaviors are not well-understood. But

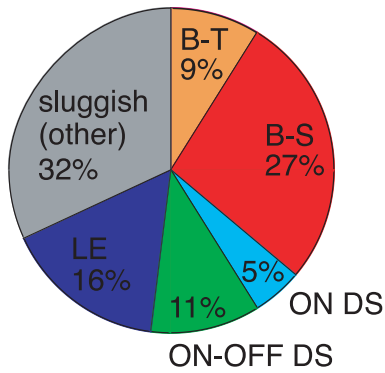


Fig. 4: Information traffic out of the retina is broadly distributed across the many types of output channels which are selective for different visual features shown here from a survey in guinea pig (adapted from [9], [36]). Types of BT = Brisk-transient (high peak rate, transient firing, ON and OFF types respond to onset and offset of spots of light); BS = Brisk-Sustained (higher peak rate, sustained firing, ON and OFF types respond to onset and offset of spots of light); ON DS = Direction Selective, responds to onset; ON-OFF DS = Direction Selective, responds to onset and offset, LE = Local Edge (responds to local edges in images); Sluggish = diverse lower peak firing rate types. Responding to natural images individual cells of each type send, on average, similar amounts of information to the brain (9-13 bits/s; 1.9-2.2 bits/spike) suggesting the selection of an efficient operating point given the neural device characteristics.

we can ask and answer a simpler question precisely: what is the best way to distribute resources over the experimentally measured computational repertoire of the retina (3 types of input channels (cones) for daylight vision and ~ 20 types of output channels) in order to maximize information or minimize cost?

This question can be answered precisely by measuring the statistics and information content of the visual input, and the properties of the circuit elements, and then asking how to maximize transmitted information given a fixed total number of components or total cost. Equivalently, we can fix the amount of information that is required and minimize the number of components. Suppose, for example, that we have a budget of N photoreceptors, and that these can be long (L), medium (M) or short (S) wavelength sensitive as in the human eye. What fraction of each kind should we have? Given the known cone spectral sensitivities and noise properties, and a measured distribution of light in different spectral bands of an ensemble of natural images [28], we can evaluate what sorts of cone mosaics will maximize chromatic information transmission. This maximization yields a surprise – it seems that the relative proportions of L and M cones are largely irrelevant over a 7-fold range, but the fraction of S cones should be less than 10% because there is less visual information in the blue wavelengths after filtering by the medium and pigments in the eye [38]. Satisfyingly, recent anatomical measurements have shown that S cones are rare (less than 6% in most mammals) and that there is massive

variability in the L/M cone ratios between humans with normal color vision [39].

The previous example was from the sensor layer of the visual system. What about circuits further downstream? Consider the retinal ganglion cells, the output cells of the retina, two synapses downstream from the photoreceptors. These cells complete the extraction of elementary visual features and send these to the central brain. Suppose we have a budget of N ganglion cells. How many should we invest in each ganglion cell type? To simplify, let us just consider two channels – ON cells, which detect bright spots, and OFF cells, which detect dark spots. Here “bright” and “dark” are defined relative to the near background. A good model of the cellular response is to image a center-surround difference-of-Gaussians filter of an image: $F(\vec{x}_0) = \int d\vec{x} L(\vec{x}) [G_c(\vec{x} - \vec{x}_0) - G_s(\vec{x} - \vec{x}_0)]$ where $L(\vec{x})$ is the luminance at point \vec{x} , and G_c and G_s are concentric, unit-normalized Gaussians with standard deviations $\sigma_c < \sigma_s$. Where $F(\vec{x}_0) > 0$, an ON cell responds, and where $F(\vec{x}_0) < 0$, and OFF cell responds, with a firing rate given by a sigmoidal function of the filter response (Fig. 5). If we have N cells, how many of them should be ON and how many of them should be OFF?

To answer this question we need to know two key facts about natural images: (1) the distribution of light is highly skewed, approximately log-normal, with a peak at low intensities and a heavy tail towards high intensities [40]; (2) there are long-range, scale-invariant correlations of luminance between pairs of points (the Fourier power spectrum scales as $1/|k|^2$ where k is the Fourier wavenumber [26], [27]). The skewed distribution implies that the mean luminance exceeds the median. Now, the average light intensity in a small central region will tend to be closer to the median, and hence will be lower than the average intensity in a larger surrounding region, which will be closer to the mean. The scale-invariant correlations imply that this difference between average central and average surrounding illumination will persist across all visual angles. So we can conclude that there are more dark spots (defined in the difference-of-Gaussians manner described above) in natural images. Given this fact, it is easy to see that an optimal retina should have more OFF cells. Consider the case $N = 1$, for example – if the budget only allows one cell, an OFF cell is a better investment as it is more likely to respond. A quantitative prediction can be made by constructing a physiologically realistic model of ganglion cell responses that approximately mimics the response range, response threshold, and noise properties of real ON and OFF cells. The analysis predicts that the typical vertebrate retina should have ~ 1.7 times as many OFF cells as ON cells, if it is efficient in its construction [41]. This prediction is confirmed by anatomy and physiology in many species (e.g., [42]).

Above, we discussed two examples in the visual periphery where circuit resources are seen to be committed to increase the efficiency of information processing. But what of the circuit repertoire of the visual cortex? Recent work has demonstrated a detailed relation between the perceptual salience of certain classes of visual textures (i.e. detection thresholds against a background of white noise), and the variance of the same textural patterns across patches of natural scenes [43],

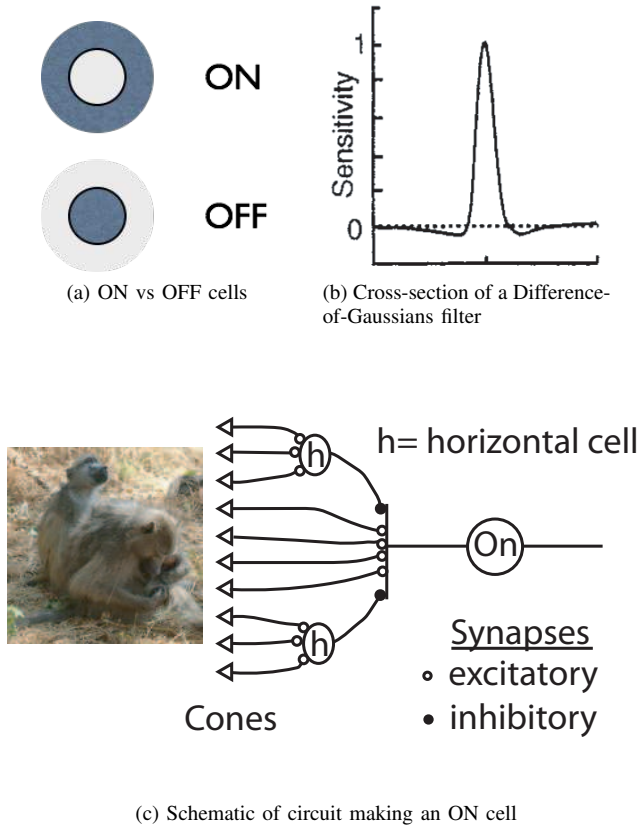


Fig. 5: A basic division in the visual system is between ON cells and OFF cells that respond to local contrast in images. ON cells respond to bright spots as compared to nearby surrounding region, while OFF cells respond to dark spots as compared to the nearby surround. ON and OFF cells can be simply modeled as Difference-of-Gaussian filters of images where ON cells respond when the filter response is positive and OFF cells respond when the filter response is negative. Schematically, an ON cell is built by feeding the responses of light sensors (cones) to the cell through an excitatory central pathway and through an inhibitory peripheral pathway. In the retina, the latter is built by specialized inhibitory circuit elements, the horizontal cells. The image is taken from the UPenn Natural Image Database [28].

[44]. Specifically, the higher the natural variance, the lower the perceptual threshold. This research addressed visual behavior rather than neural circuits, but a natural implementation of the variable threshold for different textures is to have more neurons and circuit mechanisms invested in processing the textural features of natural scenes that are more variable between image patches, and hence more informative about them. This is precisely what would be predicted by the efficient distribution of circuit resources in the cortical context where sampling limitations in each retinotopic patch lead to significant detection noise for complex patterns [44]. The basic intuition for this is that when input noise is limiting, signals with larger variance can be more reliably detected, and so circuit resources should be preferentially devoted to them

[44]–[47].

All of these examples strongly suggest tuning of visual circuit architecture to the structure of natural images and the resource constraints of neural computation. Can the distribution of information traffic over all the ~ 20 retinal output channels be understood in this way? This is a subject of ongoing research. There is also evidence that circuits in the retina dynamically reorganize to perform different functions depending on environmental circumstances – for example, the rod-driven receptor pathway that is active at low light uses specialized interneurons to transfer information over into circuits that are normally driven by cones during daylight, thus re-using already committed circuitry [48]. Stepping back, the conventional view of high level visual processing describes a hierarchy of processing levels localized in distinct brain areas that extract ever more abstract visual features. For example, primary visual cortex (V1) extract edges that are composed into corners and other elementary shape features in V4, and then further composed into shape detectors in the inferior temporal cortex (IT). In fact, in primates, a sub-region of IT, the fusiform face area, even has specialized cells that respond to individual faces. There is an emerging view that this hierarchy exists to efficiently exploit the inherently object-based nature of visual images – i.e. they are built as occluding compositions of physical things that have continuity over time. In this view, the invariant objects (which are the things of behavioral interest) can be efficiently computed and sparsely represented via hierarchical composition of features that are appropriately adapted at each representational scale. In the context of the brain, this requires commitment of a heterogeneous repertoire of specific cell types and circuits at each level of the visual hierarchy. This diversity enables sparsity in the representation and will thus allow lower activity levels and power consumption [7], [15], [17]–[19], [36]. Furthermore, following intuitions coming from the study of Support Vector Machines, one might expect that sparse, high-dimensional representations will allow higher-order computations of invariant visual percepts (e.g. “grandmother”) to be achieved more easily by simple (and thus less expensive) linear operations [49].

IV. THE SENSE OF PLACE

The idea of efficient computation has a long history in the sensory periphery, but what about circuits supporting more complex cognitive processes? Recent findings suggest that some of the underlying circuits might also be organized to efficiently exploit neural resources. For example, there is evidence that neurons can accumulate the log likelihood of events that might have happened in the world in their firing rates, with decisions ensuing when the likelihood reaches some threshold [50]. Following on this, psychologists and cognitive scientists have now presented significant evidence that human behavior can be Bayes-optimal given priors derived from experience [51]. These ideas and results concern the outcomes of neurally-based computations. Are the circuit architectures themselves efficient in the sense of conserving resources?

Cognitive circuits have been less studied than sensory ones because of the experimental difficulties of accessing and

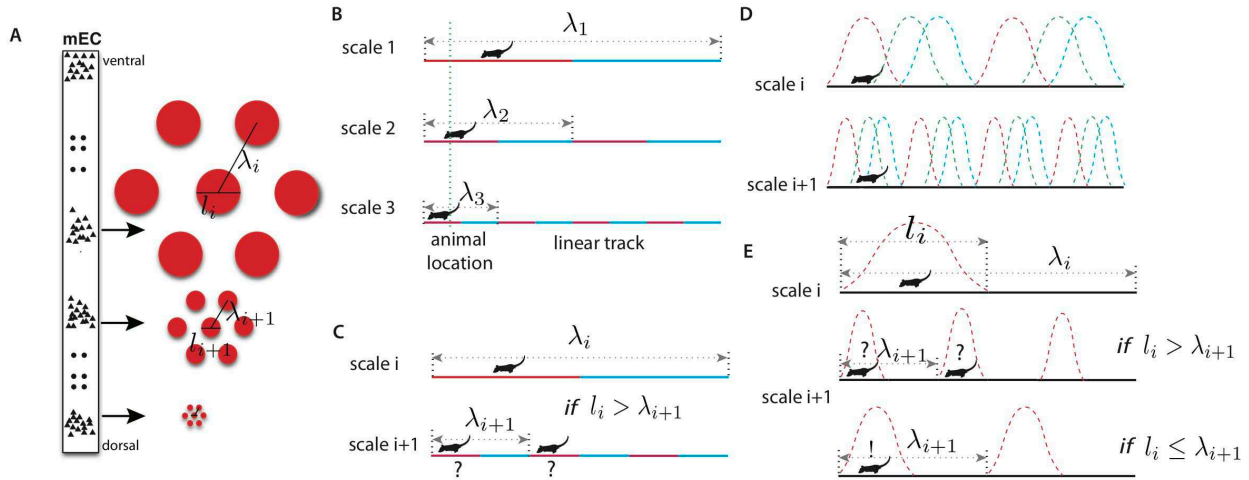


Fig. 6: Representing place in the grid system (adapted from [52]). (A) Grid cells (small triangles) in the medial entorhinal cortex (MEC) respond when the animal is in a triangular lattice of physical locations (red circles) [53], [54]. The scale of periodicity (the “grid scale”, λ_i) and the size of the regions evoking a response above a noise threshold (the “grid field width”, l_i) vary modularly along the dorso-ventral axis of the MEC [54]. Grid cells within a module vary in the phase of their spatial response, but share the same period and grid orientation (in two dimensions) [55]. (B) A simplified binary grid scheme for encoding location along a linear track. At each scale (λ_i) there are two grid cells (red vs. blue firing fields). The periodicity and grid field widths are halved at each successive scale. (C) The binary scheme in (B) is ambiguous if the grid field width at scale i exceeds the grid periodicity at scale $i + 1$. E.g., if the grid fields marked in red respond at scales i and $i + 1$, the animal might be in either of the two marked locations. (D) The grid system is composed of discrete modules, each of which contains neurons with periodic “tuning curves”, and varying phase, in space. Dashed lines of a given color represent the firing rate of a given grid cell when the animal at a given location. (E) For a simple Winner-Take-All decoder which decodes location based on the most active cell in each module of the grids in panel D, decoded position will be ambiguous unless $l_i \leq \lambda_{i+1}$, analogously to panel C (see text). Variants of this limitation occur in other decoding schemes.

measuring them. However, considerable recent effort has been invested in uncovering the neural basis of the “sense of place”, i.e. an animal’s ability to know where it is within a familiar environment. “Knowing” where one is implies the existence of an internal map of location. How could a population of neurons internally encode the physical location of an animal? Consider an animal living in an 8m linear track and requiring spatial precision of 1m to support its behavior. The animal could achieve the required resolution in a *unary place coding* scheme by having 8 neurons tuned to respond when the animal is in 1m wide, non-overlapping regions. Consider an alternative, the hierarchical *binary grid coding* scheme in Fig. 6B where neurons are tuned to respond if the animal is in one of a periodic array of locations. Here the two neurons in the largest module have periods $\lambda_1 = 8m$ and tuning curves of width $l_1 = 4m$ so that their responses just indicate the left and right halves of the track. The pairs of neurons in the next two modules have periods that are successively halved $\lambda_2 = 4m$ and $\lambda_3 = 2m$ and grid proportionally smaller field widths $l_2 = 2m$ and $l_3 = 1m$ respectively. These pairs successively localize the animal into 2m and 1m bins. An animal’s location is unambiguously reported by the firing of three neurons, one at each scale of representation. All told only 6 neurons are required in the binary grid coding scheme, less than in the unary scheme. This suggests that grid schemes that integrate multiple scales of representation can encode space more efficiently, i.e. with fewer neural resources.

Of course many animals move in two dimensional environments, and some (like bats, birds and fish) move in three dimensions. We can generalize the simple grid coding scheme described above to d-dimensions by imagining that each neuron responds if the animal is in a location that lies within a rectangular lattice of bins. For example, consider a two-dimensional $8m \times 8m$ arena where we must resolve location in $1m \times 1m$ bins. In a *unary place coding* scheme we dedicate one neuron to respond when the animal is in each distinct bin. In an alternative *binary grid coding* scheme, at the largest scale, we can imagine four neurons each encoding a quadrant (i.e. neurons with a response periodicity of $8m$ in the x and y directions, tuned to respond in $4m \times 4m$ bins, thus giving a single response region in an $8m \times 8m$ arena). At the next scale, we imagine four more neurons that subdivide each quadrant into four (i.e. they respond in $2m \times 2m$ bins with a period of $4m$ in each cardinal direction). Finally, we consider a third scale with $1m \times 1m$ bins and a $2m$ period (Fig. 6). As in our one-dimensional example, the firing of neurons at each of these three scales of representation is needed to resolve position unambiguously, but the binary grid scheme achieves the same resolution with fewer neurons than the unary scheme (12 vs. 64 in this example), again suggesting that a grid-like multi-scale representation of position would be more efficient for the brain [52], [56]–[59]. In general a grid scheme simply needs diverse tuning curves distributed over modules with different periodicities defined on some lattice,

and need not have any particular relation imposed between the different modules.

A priori the abstraction and seeming complexity of grid schemes make it seem unlikely that the brain would implement a grid-like system for encoding location, even if it is more efficient in consumption of resources. Remarkably, however, a potential locus of such a multi-scale cognitive map of location was recently discovered in a structure known as the medial entorhinal cortex (MEC) [53], [54]. When rats freely explore a two dimensional open environment, individual “grid cells” in the MEC display spatial firing fields that form a periodic triangular grid which tiles space (Fig. 6A). The scale of grid fields varies systematically along the dorso-ventral axis of the MEC (Fig. 6A) [54]. The grids are partly formed on the basis of path integration inside the animal’s brain, but are anchored to sensory cues from the environment [60]. They are maintained even in darkness and, if the environment is slowly deformed, the grid pattern deforms with it [54], [61]. It was shown that grid cells are organized in discrete modules within which the cells share the same lattice orientation and periodicity, but vary randomly in phase [54], [55].

Is the grid coding scheme implemented by the entorhinal cortex efficient in the use of neural resources? To test this, we can follow [52] to formalize the problem as follows (Fig. 6A): consider a hierarchy of modules where all the neurons in module i have the same period λ_i ($\lambda_1 > \lambda_2 \dots \lambda_n$). In each module, the grid firing fields (i.e. the connected spatial regions that evoke firing) are compact (with a diameter denoted l_i) after thresholding for activity above the noise level. Within any module, grid cells have a variety of spatial phases so that at least one cell will respond at any physical location (Fig. 6B,D). Grid modules with smaller field widths l_i provide more local spatial information than those with larger scales. However, this increased spatial precision comes at a cost: the correspondingly smaller periodicity λ_i of these modules leads to increased ambiguity since there are more grid periods within a given spatial region. Thus, there will be a tradeoff between precision and ambiguity. Finally, consider Fig. 6C where the cells with the grid fields marked in red respond at scales i and $i + 1$. Then the animal might be in either of the two marked locations. Avoiding ambiguity at each scale of representation requires that λ_{i+1} , the period at scale $i + 1$, must exceed l_i , the grid field width at scale i .¹

How should a grid system be organized to minimize the resources required to represent location unambiguously with a given resolution? Consider a simple one dimensional grid system. If d cells respond above the noise threshold at each point, the number of grid cells n_i in module i will be $n_i = d\lambda_i/l_i$ and the total number of grid cells is $N = \sum_{i=1}^m d\frac{\lambda_i}{l_i}$ where m is the number of grid modules. Now imagine a decoder which considers the animal as localized within the grid fields of the most responsive cell in each module. The smallest interval that can be resolved in this way will be l_m . We therefore quantify the resolution of the grid system (the number of

spatial bins that can be resolved) as the ratio of the largest to the smallest scale, $R_1 = \lambda_1/l_m$, which we assume to be large and fixed by the animal’s behavior. In terms of the period ratios $r_i = \lambda_i/\lambda_{i+1}$, the resolution is $R_1 = \prod_{i=1}^m r_i$, where we also defined $r_m = \lambda_m/l_m$. Unambiguous decoding requires that $l_i \leq \lambda_{i+1}$ (Fig. 1C,E), or, equivalently, $\frac{\lambda_i}{l_i} \geq r_i$. To minimize $N = d \sum_i \lambda_i/l_i$, all the $\frac{\lambda_i}{l_i}$ should be as small as possible; so this fixes $\frac{\lambda_i}{l_i} = r_i$. Thus we are reduced to minimizing the sum $N = d \sum_{i=1}^m r_i$ over the parameters r_i , while fixing the product $R_1 = \prod_i r_i$. Because this problem is symmetric under permutation of the indices i , the optimal r_i turn out to all be equal, allowing us to set $r_i = r$. This is our first prediction: **(1)** the ratios between adjacent periods will be constant. The constraint on resolution then gives $m = \log_r R$, so that we seek to minimize $N(r) = dr \log_r R_1$ with respect to r : the solution is $r = e$ (Fig. 2E; Supplementary Information). This gives a second prediction: **(2)** the ratio of adjacent grid periods should be close to $r = e$. Repeating this analysis in two dimensions (where resolution will be set by a ratio of areas λ_1^2/l_m^2), predicts a constant period ratio of \sqrt{e} between adjacent modules, each arranged in a triangular lattice, for a grid system that minimizes the number of neurons required to achieve a given resolution [52]. The analysis above made various simplifying assumptions, but the result is robust to relaxing these conditions [52]. Amazingly, this prediction is precisely confirmed by experiments [55], and ongoing experiments will test the predictions for the three dimensional grid system (a scale ratio of $e^{1/3}$) in bats [64].

The purpose of this detailed discussion was to demonstrate the intricacy of the considerations involved in determining the most efficient computational architectures to solve even a simple problem like representation of an animal’s location. Here the most parsimonious solution in terms of neuron number involved a heterogeneous population of neurons organized in modules with very different tuning properties to spatial location. These diverse modules then jointly represent spatial location. The fact that evolution has produced circuits that implement these architectures suggests a selective pressure for efficiency which reduces, where possible, the number of circuit elements involved in a computation. Doing this requires specializing the architectures to the specific task that must be performed (as in the grid system illustrated above) and then organizing interactions between such special-purpose circuits.

Of course, there may be constraints on a neural code that preclude an organization which would nominally be more efficient. For example, in the sensory periphery light is sensed by single molecules that capture photons. Likewise the pressure waves forming sound are naturally sensed by devices (hair cells) that resonate at particular frequencies. In both these cases the biophysics of sensing plays an important role in determining the structure of circuits and the sensory code. Furthermore, while an animal can only be in one place at one time, natural sounds involve simultaneous excitations of different amplitudes in all frequencies, while natural images have light of different intensities at all locations in an image. (Correlations between sound frequencies or image locations reduce the dimensionality somewhat, but it is still very high.)

¹Theoretically, one could resolve the ambiguity in Fig. 6C by combining the responses of many grid modules with incommensurate periods [56], [57], but this requires a complex readout that examines all the grid modules at the same time that is anatomically disfavored [62], [63].

As such, auditory and visual stimuli have a much higher intrinsic dimension than the spatial location of an animal. Similar considerations as above would predict that in an optimal grid representation in d -dimensions the periods of modules would scale by a factor of $e^{1/d}$. For large d this is very close to 1. There is no plausible biological mechanism for implementation of a such a finely scaling grid given the noisiness of neural responses and cellular developmental mechanisms. A sub-optimal grid will have exponentially more neurons than the optimal one, and, in view of this, an economical strategy for the auditory and visual systems might be to first reduce the dimension of the stimuli by extracting informative features. This is thought to happen in the auditory and visual cortices. Perhaps there are later, as yet undiscovered, stages of processing where a sufficiently low-dimensional subset of these sensory modalities is manipulated in a grid-like way.

Readers familiar with spatial representation in the brain may at this stage also wonder about the status of the so-called “place cell” system in the hippocampus which seems to resemble the much less efficient *unary place code* discussed above. If the grid system is so efficient, why also have an apparently less efficient “place system” implemented in a different region of the brain, and what does this imply for the general thesis of this paper that heterogeneous, special-purpose architectures are the route to efficiency in computation? We will turn to this and other related questions in the conclusion.

V. CONCLUSION: EFFICIENT CIRCUITS IN THE BRAIN

I have explained above that the brain has a hierarchical organization from the scale of individual neurons to the scale of “brain areas” tasked with specific functions like speech production or motor control. At each level of this hierarchy the brain implements a heterogeneous collection of functional units (these might be individual neurons or they might be circuit motifs) which are adapted to specific sub-computations that must be performed. There are extensive feedforward and feedback interactions between many levels of the hierarchy and the different functional types at each scale. These interactions sometimes lead theorists to treat the brain as an abstract interconnected network of stereotyped units resembling some of the early thinking in the field of neuroscience a century ago. But, contrary to this view, it seems clear that separations of timescales between levels of the computational hierarchy and the differences of connectivity between the functional units at each level imply that we should think of the brain as a highly heterogeneous computational architecture, composed of specialized components adapted to the procedures that must be performed at each scale of computational analysis. Schematically, at a given computational scale we might represent this as in Fig. 2 (where interactions between the functional units and between levels of the computational hierarchy have been left out).

I have argued that this heterogeneous architecture composed of highly diverse and specialized components is partly responsible for the enormous computational efficiency of the brain. The argument is essentially economic – specialization of function enables more efficient procedures that consume

fewer resources. Information theory also supports the view that any given physical channel will have an optimal operating point for maximizing bits/energy, and thus it would pay to break up information into components that can be processed at this optimal point (Fig. 3). In order to support this view I gave extensive examples from the circuits that support sensory processing in the visual system, and from the circuits that support the cognitive sense of place. In each case we saw a remarkable correspondence between predictions for circuit architecture based on resource minimization and the actual structure in the brain.

In this article we did not have space to ask how these well-adapted and efficient circuits are built in brain. In some cases (e.g. the retina) the wiring is established during development of the brain and thus is encoded in circuit layout mechanisms that are remembered in the genome. In other cases (e.g. the grid system for spatial cognition that we discussed) a dynamical mechanism of self-organization is at play, since animals develop a new map in each new environment after ~ 30 minutes of exploration. This self-organized map is remembered, and can be rapidly reloaded when an animal returns to a familiar environment. In still other examples, experience-dependent learning and synaptic plasticity is involved. Examples abound of structural changes in the brain that follow learning and repeated practice – new synapses are made and circuits can adapt and change their functions. Indeed, partial recovery of functions after the damage caused by strokes depends on such dynamical reconfigurability that can reconstruct appropriately specialized circuits to perform specific tasks. Of course, none of this resembles a conventional von Neumann architecture with a centralized, highly general CPU separated from remembered data.

The examples that I discussed in this article were in computational domains that have an inherently low dimensionality. For example, although natural images have very many pixels and change constantly, the light coming to the eye from different points and moments is highly correlated. Ultimately this is because the visual world has an underlying structure in terms of occluding objects that are in turn made of pieces that move as wholes, and which in turn are identifiable in terms of their edges and corners. This sort of structure makes a relatively low-level feature-based representation possible and it seems that the brain exploits this. Likewise, in the example of the “sense of place”, physical location is limited by the dimensionality of the world, the behavioral range of animal, and the spatial resolution they require because of their size and the size of the objects in the world that they interact with.

But there are also situations where the inherent dimensionality of a computational problem vastly exceeds the resources available, at least naively. Consider the challenge faced by the olfactory system. There are very many kinds of volatile molecules (perhaps 10^5 of them) that can be mixed in different concentrations to make odors. But the typical mammal has only $O(1000)$ odor receptors, and a fly only has $O(100)$, leading to a massive mismatch between the chemotopic dimensionality of odor space and the dimension of the receptor space. How can animals possibly deal with this situation? A key insight is that a complex odor produced by an animal

or plant will typically consist of only $O(100)$ odorants. While this allows $O(10^{500})$ possible odors that could be relevant to an animal's behavior, it also implies that signals drawn from the natural odor space have a particular kind of sparseness – if we think of an odor as a vector in a 10^5 dimensional chemical space, then a natural odor will have only $k \approx 100$ non-zero entries. Because such “k-sparse” signals do not lie in a fixed linear subspace, conventional techniques of dimensionality reduction such as Principal Component Analysis will not work. However, recent theorems in the mathematics of *compressive sensing* say that a random projection of the high-dimensional data vector into an $O(k)$ dimensional space will preserve all the information in the original signal in a manner that preserves distances between vectors and hence their classifiability [65], [66]. This suggests that rather than computing and representing “odor features”, efficiency would drive the olfactory system to randomize its sensing. Perhaps this explains why odor receptors sense molecules diffusely (each receptor binds broadly to many odorants, and each odorant binds to many receptors [67]). Perhaps this is also why the projections to the olfactory cortex from the sensory periphery, appear to all intents and purposes to be random [68], in stark contrast to the highly organized representations in the visual system that we discussed above.

Another example of randomization as a route to efficiency might occur in the hippocampus, where neurons seem to respond to a very high dimensional space of behavioral contexts [69], [70]. Contexts are also k-sparse in the sense discussed above – although the space of possible contextual influences is enormous, in any given situation a relatively small number appear. A particularly prominent context is spatial location, and hence cells in the hippocampus often show “place fields” – i.e. once an environment becomes familiar, particular hippocampal cells will respond when the animal is in a particular spatial location and this fixed map is remembered and reloaded whenever the animal re-enters the environment. Thus, until the discovery of grid cells (discussed above), the hippocampus was thought to be the locus of the “sense of place”. However, the place cells in this structure also respond to many other kinds of context, e.g. sounds, smells and other cues (see, e.g., [71]). Interestingly, like the projections to the olfactory cortex, the axonal projections from other areas to the hippocampus may also be disordered [69], [70]. In general, animals or computers dealing with open environments face this sort of situation because the space of possible contexts for behavior is very large indeed, and randomization may provide a route for efficiently representing information.

This special issue was dedicated to the role of memories in information processing, particularly with reference to the efficiency of computation. In this article, I have not discussed memory in the conventional sense of remembering “facts”. There are many sorts of memory of this kind in the brain. For example, there is *working memory* where Miller's Law says that a human can hold about seven numbers in his or her head for a short duration in order to perform a task [72]. Then there is *short-term memory* where subjects retain facts over hours or days. *Long-term memories* are also stored in the brain, sometimes associatively. Memories can be *episodic* (about

sequences) or *declarative* (about “facts”). Consistently with the themes of this article, these different sorts of memory seem to be implemented by distinct computational mechanisms and circuits in the brain. Here, I have discussed circuit level architectures as “memories” of computations that provide an efficient procedural basis for the diverse goals faced by an animal. In fact, such structural memories are also related to “factual memories” – there are cells in the fusiform face area of the primate brain that recognize and respond to specific faces, and there are single cells in the temporal lobe that can respond selectively to complex concepts (e.g. just to the actress Jennifer Aniston, her voice, her name and other attributes [73]). In any case, the main message here is that the brain implements an enormously heterogeneous repertoire of computational elements at each level of processing, and that this likely has a bearing on the efficiency of the brain (~ 12 W of power consumption) even while performing complex tasks that are difficult for powerful conventional computers. Perhaps this provides a guide for the design of efficient computational devices dealing with open environments.

ACKNOWLEDGMENT

Vijay Balasubramanian was supported by NSF grants PHY-1058202, EF-0928048 and PHY-1066293 and thanks Ann Hermundstad, Kamesh Krishnamurthy, Jason Prentice, Peter Sterling and Xuexin Wei for their many insights in the collaborations upon which this work is based.

REFERENCES

- [1] P. Sterling and S. Laughlin, *Principles of Neural Design*. Cambridge, MA: MIT Press, 2014.
- [2] V. Braitenberg and A. Schütz, “Anatomy of the cortex,” *Spring-Verlag, Berlin*, 1991.
- [3] S. Herculano-Houzel, “The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost,” *Proceedings of the National Academy of Sciences*, vol. 109, no. Supplement 1, pp. 10 661–10 668, 2012.
- [4] L. Sokoloff, “The metabolism of the central nervous system in vivo,” *Handbook of physiology-neurophysiology*, vol. 3, pp. 1843–1864, 1960.
- [5] D. Rolfe and G. C. Brown, “Cellular energy utilization and molecular origin of standard metabolic rate in mammals,” *Physiological reviews*, vol. 77, no. 3, pp. 731–758, 1997.
- [6] B. K. Siesjo, *Brain energy metabolism*. John Wiley and Sons Ltd, 1978.
- [7] D. Attwell and S. B. Laughlin, “An energy budget for signaling in the grey matter of the brain,” *Journal of Cerebral Blood Flow & Metabolism*, vol. 21, no. 10, pp. 1133–1145, 2001.
- [8] P. Lennie, “The cost of cortical computation,” *Current biology*, vol. 13, no. 6, pp. 493–497, 2003.
- [9] V. Balasubramanian and P. Sterling, “Receptive fields and functional architecture in the retina,” *The Journal of physiology*, vol. 587, no. 12, pp. 2753–2767, 2009.
- [10] R. H. Masland, “The fundamental plan of the retina,” *Nature neuroscience*, vol. 4, no. 9, pp. 877–886, 2001.
- [11] M. A. MacNeil and R. H. Masland, “Extreme diversity among amacrine cells: implications for function,” *Neuron*, vol. 20, no. 5, pp. 971–982, 1998.
- [12] L. M. Zimmerman and I. Veith, *Great ideas in the history of surgery*. Williams & Wilkins Baltimore, 1961.
- [13] E. R. Kandel, J. H. Schwartz, T. M. Jessell, et al., *Principles of neural science*. McGraw-Hill New York, 2000, vol. 4.
- [14] G. M. Shepherd et al., *The synaptic organization of the brain*. Oxford University Press New York, 2004, vol. 3.
- [15] V. Balasubramanian, D. Kimber, and M. J. Berry II, “Metabolically efficient information processing,” *Neural Computation*, vol. 13, no. 4, pp. 799–815, 2001.

- [16] R. E. Blahut, *Principles and practice of information theory*. Addison-Wesley Longman Publishing Co., Inc., 1987.
- [17] W. B. Levy and R. A. Baxter, "Energy efficient neural codes," *Neural computation*, vol. 8, no. 3, pp. 531–543, 1996.
- [18] —, "Energy-efficient neuronal computation via quantal synaptic failures," *The Journal of Neuroscience*, vol. 22, no. 11, pp. 4746–4755, 2002.
- [19] R. Sarpeshkar, "Analog versus digital: extrapolating from electronics to neurobiology," *Neural computation*, vol. 10, no. 7, pp. 1601–1638, 1998.
- [20] J. A. Perge, K. Koch, R. Miller, P. Sterling, and V. Balasubramanian, "How the optic nerve allocates space, energy capacity, and information," *The Journal of neuroscience*, vol. 29, no. 24, pp. 7917–7928, 2009.
- [21] J. A. Perge, J. E. Niven, E. Mugnaini, V. Balasubramanian, and P. Sterling, "Why do axons differ in caliber?" *The Journal of Neuroscience*, vol. 32, no. 2, pp. 626–638, 2012.
- [22] S. B. Laughlin and T. J. Sejnowski, "Communication in neuronal networks," *Science*, vol. 301, no. 5641, pp. 1870–1874, 2003.
- [23] F. Attneave, "Some informational aspects of visual perception," *Psychological review*, vol. 61, no. 3, p. 183, 1954.
- [24] H. B. Barlow, "Possible principles underlying the transformation of sensory messages," *Sensory communication*, pp. 217–234, 1961.
- [25] M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: a fresh view of inhibition in the retina," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 216, no. 1205, pp. 427–459, 1982.
- [26] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *JOSA A*, vol. 4, no. 12, pp. 2379–2394, 1987.
- [27] v. A. van der Schaaf and J. v. van Hateren, "Modelling the power spectra of natural images: statistics and information," *Vision research*, vol. 36, no. 17, pp. 2759–2770, 1996.
- [28] G. Tkačik, P. Garrigan, C. Ratliff, G. Milčinski, J. M. Klein, L. H. Seyfarth, P. Sterling, D. H. Brainard, and V. Balasubramanian, "Natural images from the birthplace of the human eye," *PLoS One*, vol. 6, no. 6, p. e20409, 2011.
- [29] J. J. Atick and A. N. Redlich, "Towards a theory of early visual processing," *Neural Computation*, vol. 2, no. 3, pp. 308–320, 1990.
- [30] —, "What does the retina know about natural scenes?" *Neural computation*, vol. 4, no. 2, pp. 196–210, 1992.
- [31] B. G. Borghuis, C. P. Ratliff, R. G. Smith, P. Sterling, and V. Balasubramanian, "Design of a neuronal array," *The Journal of Neuroscience*, vol. 28, no. 12, pp. 3178–3189, 2008.
- [32] Y. S. Liu, C. F. Stevens, and T. O. Sharpee, "Predictable irregularities in retinal receptive fields," *Proceedings of the National Academy of Sciences*, vol. 106, no. 38, pp. 16 499–16 504, 2009.
- [33] A. J. Bell and T. J. Sejnowski, "The independent components of natural scenes are edge filters," *Vision research*, vol. 37, no. 23, pp. 3327–3338, 1997.
- [34] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.
- [35] J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 265, no. 1394, pp. 359–366, 1998.
- [36] K. Koch, J. McLean, R. Segev, M. A. Freed, M. J. Berry II, V. Balasubramanian, and P. Sterling, "How_i i_z much_i/i_z the eye tells the brain," *Current Biology*, vol. 16, no. 14, pp. 1428–1434, 2006.
- [37] K. Koch, J. McLean, M. Berry, P. Sterling, V. Balasubramanian, and M. A. Freed, "Efficiency of information transmission by retinal ganglion cells," *Current Biology*, vol. 14, no. 17, pp. 1523–1530, 2004.
- [38] P. Garrigan, C. P. Ratliff, J. M. Klein, P. Sterling, D. H. Brainard, and V. Balasubramanian, "Design of a trichromatic cone array," *PLoS computational biology*, vol. 6, no. 2, p. e1000677, 2010.
- [39] H. Hofer, J. Carroll, J. Neitz, M. Neitz, and D. R. Williams, "Organization of the human trichromatic cone mosaic," *The Journal of Neuroscience*, vol. 25, no. 42, pp. 9669–9679, 2005.
- [40] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual review of neuroscience*, vol. 24, no. 1, pp. 1193–1216, 2001.
- [41] C. P. Ratliff, B. G. Borghuis, Y.-H. Kao, P. Sterling, and V. Balasubramanian, "Retina is structured to process an excess of darkness in natural scenes," *Proceedings of the National Academy of Sciences*, vol. 107, no. 40, pp. 17 368–17 373, 2010.
- [42] E. Chichilnisky and R. S. Kalmar, "Functional asymmetries in on and off ganglion cells of primate retina," *The Journal of Neuroscience*, vol. 22, no. 7, pp. 2737–2747, 2002.
- [43] G. Tkačik, J. S. Prentice, J. D. Victor, and V. Balasubramanian, "Local statistics in natural scenes predict the saliency of synthetic textures," *Proceedings of the National Academy of Sciences*, vol. 107, no. 42, pp. 18 149–18 154, 2010.
- [44] A. M. Hermundstad, J. J. Bruglio, M. M. Conte, J. D. Victor, V. Balasubramanian, and G. Tkačik, "Variance predicts salience in central sensory processing," *eLife*, vol. 3, p. e03722, 2014.
- [45] J. Van Hateren, "A theory of maximizing sensory information," *Biological cybernetics*, vol. 68, no. 1, pp. 23–29, 1992.
- [46] E. Doi and M. S. Lewicki, "A simple model of optimal population coding for sensory systems," *PLoS computational biology*, vol. 10, no. 8, p. e1003761, 2014.
- [47] —, "Characterization of minimum error linear coding with sensory and neural noise," *Neural computation*, vol. 23, no. 10, pp. 2498–2510, 2011.
- [48] R. G. Smith, M. A. Freed, and P. Sterling, "Microcircuitry of the dark-adapted cat retina: functional architecture of the rod-cone network," *The Journal of neuroscience*, vol. 6, no. 12, pp. 3505–3517, 1986.
- [49] J. J. DiCarlo and D. D. Cox, "Untangling invariant object recognition," *Trends in cognitive sciences*, vol. 11, no. 8, pp. 333–341, 2007.
- [50] J. I. Gold and M. N. Shadlen, "The neural basis of decision making," *Annu. Rev. Neurosci.*, vol. 30, pp. 535–574, 2007.
- [51] N. Chater, J. B. Tenenbaum, and A. Yuille, "Probabilistic models of cognition: Conceptual foundations," *Trends in cognitive sciences*, vol. 10, no. 7, pp. 287–291, 2006.
- [52] X. Wei, J. Prentice, and V. Balasubramanian, "The sense of place: grid cells in the brain and the transcendental number e," *Submitted*, 2014.
- [53] M. Fyhn, S. Molden, M. P. Witter, E. I. Moser, and M.-B. Moser, "Spatial representation in the entorhinal cortex," *Science*, vol. 305, no. 5688, pp. 1258–1264, 2004.
- [54] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, "Microstructure of a spatial map in the entorhinal cortex," *Nature*, vol. 436, no. 7052, pp. 801–806, 2005.
- [55] H. Stensola, T. Stensola, T. Solstad, K. Frøland, M.-B. Moser, and E. I. Moser, "The entorhinal grid map is discretized," *Nature*, vol. 492, no. 7427, pp. 72–78, 2012.
- [56] I. R. Fiete, Y. Burak, and T. Brookings, "What grid cells convey about rat location," *The Journal of Neuroscience*, vol. 28, no. 27, pp. 6858–6871, 2008.
- [57] S. Sreenivasan and I. Fiete, "Grid cells generate an analog error-correcting code for singularly precise neural computation," *Nature neuroscience*, vol. 14, no. 10, pp. 1330–1337, 2011.
- [58] A. Mathis, A. V. Herz, and M. Stemmler, "Optimal population codes for space: Grid cells outperform place cells," *Neural Computation*, vol. 24, no. 9, pp. 2280–2317, 2012.
- [59] A. Mathis, A. V. Herz, and M. B. Stemmler, "Resolution of nested neuronal representations can be exponential in the number of neurons," *Physical review letters*, vol. 109, no. 1, p. 018103, 2012.
- [60] E. I. Moser, E. Kropff, and M.-B. Moser, "Place cells, grid cells, and the brain's spatial representation system," *Annu. Rev. Neurosci.*, vol. 31, pp. 69–89, 2008.
- [61] C. Barry, R. Hayman, N. Burgess, and K. J. Jeffery, "Experience-dependent rescaling of entorhinal grids," *Nature neuroscience*, vol. 10, no. 6, pp. 682–684, 2007.
- [62] N. Van Strien, N. Cappaert, and M. Witter, "The anatomy of memory: an interactive overview of the parahippocampal–hippocampal network," *Nature Reviews Neuroscience*, vol. 10, no. 4, pp. 272–282, 2009.
- [63] T. Solstad, E. I. Moser, and G. T. Einevoll, "From grid cells to place cells: a mathematical model," *Hippocampus*, vol. 16, no. 12, pp. 1026–1031, 2006.
- [64] M. M. Yartsev and N. Ulanovsky, "Representation of three-dimensional space in the hippocampus of flying bats," *Science*, vol. 340, no. 6130, pp. 367–372, 2013.
- [65] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.
- [66] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [67] E. A. Hallem and J. R. Carlson, "Coding of odors by a receptor repertoire," *Cell*, vol. 125, no. 1, pp. 143–160, 2006.
- [68] D. D. Stettler and R. Axel, "Representations of odor in the piriform cortex," *Neuron*, vol. 63, no. 6, pp. 854–864, 2009.
- [69] W. B. Levy, "A computational approach to hippocampal function," *Computational models of learning in simple neural systems*, vol. 23, pp. 243–305, 1989.

- [70] W. B. Levy, A. B. Hocking, and X. Wu, "Interpreting hippocampal function as recoding and forecasting," *Neural Networks*, vol. 18, no. 9, pp. 1242–1264, 2005.
- [71] K. J. Jeffery, "Integration of the sensory inputs to place cells: what, where, why, and how?" *Hippocampus*, vol. 17, no. 9, pp. 775–785, 2007.
- [72] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information." *Psychological review*, vol. 63, no. 2, p. 81, 1956.
- [73] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, "Invariant visual representation by single neurons in the human brain," *Nature*, vol. 435, no. 7045, pp. 1102–1107, 2005.



Vijay Balasubramanian received B.Sc. degrees in Physics and Computer Science from MIT and an M.Sc. in Computer Science from MIT. Following a Ph.D. in Theoretical Physics at Princeton University, he was a Junior Fellow at the Harvard Society of Fellows. He then joined the physics faculty at the University of Pennsylvania where he conducts research in theoretical physics and in neuroscience. He seeks to understand the principles that organize information processing by living systems, and especially by neural circuits. In theoretical physics he has

written about the statistical inference of models, on the apparent destruction of information by black holes, on the transformation of information between microscopic and coarse-grained descriptions of the world. Broadly, he seeks to understand how information is created, stored, transformed and used by physical systems,