

COMMENTARY

Brain power

Vijay Balasubramanian^{a,b,1}

The human brain is just 2% of the body's weight, but 20% of its metabolic load (1–3), and 10 times more expensive per gram than muscle. On the other hand, the brain manages to produce poetry, design spacecraft, and create art on an energy budget of ~20 W, a paltry sum given that the computer on which this article is being typed requires 80 W. So where in the brain is power consumed, what is it used for, why is it so expensive relative to other costs of living, and how does it achieve its power efficiency relative to engineered silicon? Many classic papers have studied these questions. Attwell and Laughlin (4) developed detailed biophysical estimates suggesting that neural signaling and the postsynaptic effects of neurotransmitter release combined to account for 80% of the brain's adenosine triphosphate (ATP) consumption, conclusions that are also supported by the overall physiology and anatomy of neural circuits (5, 6). Numerous studies explored the structural and functional consequences of this expenditure for limiting brain size (7) and scaling (8), efficient wiring patterns (9), analog (graded potential) vs. digital (spiking) signaling (10), distributed neural codes (11–13), the distribution of information traffic along nerve tracts and their size distribution (14–16), and computational heterogeneity and efficiency (17). Many of these ideas have been synthesized by Sterling and Laughlin (18) into a set of principles governing the design of brains. Now, in PNAS, Levy and Calvert (19) propose a functional accounting of the power budget of the mammalian brain, suggesting that communication is vastly more expensive than computation, and exploring the functional consequences for neural circuit organization.

Levy and Calvert (19) build on the earlier literature by focusing primarily on the relative power committed to different modes of information processing in the human brain, rather than on different aspects of cellular function. They make a key distinction between communication and computation. "Communication" refers, in general, to the transport of information, perhaps encoded in some way, from one site to another without

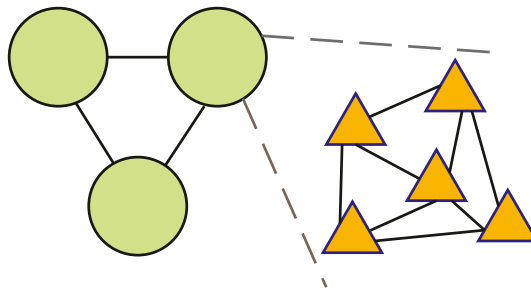


Fig. 1. Computation in the brain is distributed among specialized components communicating in networks. Green circles denote a network of cortical areas, each executing different functions. Orange triangles denote a network of pyramidal cells coordinating to compute the function of one cortical area. Levy and Calvert (19) argue that the cost of computation by the elementary units is dwarfed by the cost of communication between them.

transformation of the representation, extraction of salient features, or mapping to decisions or outcomes. An example in the brain is the transport of visual information, unchanged, along the optic nerve from the eye to the central brain. "Computation," a more subtle concept, is generally understood in terms of an input–output transformation. Levy and Calvert, building on previous work of Levy, view each neuron as performing a "microscopic estimation or prediction" of latent variables in its input and encoding this output in interpulse intervals (IPIs), that is, in the relative timing of the action potentials used in signaling by most neurons. Estimation and prediction are sufficiently general frameworks to subsume other views of neural computation (e.g., neurons as dynamical systems or logic gates), and also to encompass the role played by single neurons within networks charged with carrying out a computational function. Likewise, information coding in IPIs includes other possibilities like rate codes and pattern codes as specific cases. Thus, an example of a "computation" by a neuron in the brain could be

^aComputational Neuroscience Initiative, Richards Laboratories, University of Pennsylvania, Philadelphia, PA 19104; and ^bDepartment of Physics, David Rittenhouse Laboratories, University of Pennsylvania, Philadelphia, PA 19104

Author contributions: V.B. wrote the paper.

The author declares no competing interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

See companion article, "Communication consumes 35 times more energy than computation in the human cortex, but both costs are needed to predict synapse number," [10.1073/pnas.2008173118](https://doi.org/10.1073/pnas.2008173118).

¹Email: vijay@physics.upenn.edu.

Published August 2, 2021.

the signaling by a simple cell in V1 of the presence of a horizontal bar in the visual input.

Employing this perspective, the authors conclude that the biophysical processes supporting communication consume a startling 35 times more power (ATP molecules per second in biologically relevant units, or joules per second in physical units) than the processes supporting computation (19). They come to this conclusion by summing up the ATP cost to recover ionic gradients from the excitatory currents per IPI involved in computation vs. costs such as axonal resting potentials, action potentials, and vesicle recycling involved in communication. This vast difference formalizes findings implicit in ref. 4, whose authors showed that action potentials and their postsynaptic effects dominate power consumption in the brain. In another supporting line of evidence from ref. 15, mitochondrial distributions in neurons track firing rates and synaptic transmission so that the thickness of axons may be largely determined by the need to supply synaptic terminals whose use consumes 65% of the energy budget of the mammalian brain (20). These expensive processes from refs. 4 and 15 are describing communication, not computation, in the framework of ref. 19. Interestingly, Levy and Calvert also estimate that 27% of the cortical power expenditure is spent on costs associated with synaptogenesis, such as growth via actin polymerization, membrane synthesis and incorporation, and associated intracellular transport. This interesting refinement of previous energy budgets suggests that more than a quarter of the energy cost of owning a brain is to facilitate ongoing learning, consistent with our qualitative impression of the purpose of this organ.

In physical terms, Levy and Calvert (19) calculate that cortical gray matter consumes about 3 W of power out of the 17 W to 20 W derived from glucose uptake in the brain (1, 2), 10% of which remains unused under normal operation. About 3 times as much (~9 W) is lost to heat, and the remainder powers the rest of the brain. The 3-W estimate for the power consumption of gray matter is consistent with earlier work of Lennie (21), who, using data available two decades ago, found a different partitioning of the energy budget suggesting that 50% of the power is devoted to processes that are independent of neural electrical activity (i.e., computation and communication). The differences arise partly from the primary biophysical data, including estimates of the number of synapses and their success rate. The new estimates also suggest that, on a per-neuron basis, human gray matter uses about $2 \cdot 10^{-9}$ W per neuron, consistent with the work of Herculano-Houzel (22).

Levy and Calvert (19) use their results to estimate the bits of information computed per joule, measured in terms of the mutual information between a noisy neural integrator's input and output. In this definition, they find that neurons consume 10^8 times more power per bit of computation than the idealized thermodynamic bound of Landauer (23), and suggest that the large difference arises from the cost of communication (which Landauer neglected) and the biological imperative to compute sufficiently quickly to be relevant for behavior. While these two constraints are certainly

relevant, future researchers will want to carefully examine assumptions about what constitutes computation, and the biophysical constraints imposed by computation with living cells. In Levy and Calvert's analysis, the energy budget and the bits produced per joule consumed both depend on a parameter: the average number of input synapses to a cortical pyramidal neuron times their success rate. They find their bits per joule estimate is maximized when this parameter is about 2,000, a number close to the value they derive from data. This interesting result recalls earlier

In PNAS, Levy and Calvert propose a functional accounting of the power budget of the mammalian brain, suggesting that communication is vastly more expensive than computation, and exploring the functional consequences for neural circuit organization.

efforts to, for example, use energy efficiency to understand how neural firing should be organized (11, 13), and how information should be partitioned across cells and structures in the brain (14, 17). The new work opens the door to more refined analyses of the architecture of circuits in the brain and of how they partition their computational tasks.

Levy and Calvert's (19) results are pertinent for the interpretation of functional MRI measurements of regional brain metabolism in terms of ongoing computation and communication. The new data are also pertinent for studies of the evolution of the brain, specifically, the notion that brain is expensive tissue (7) and that metabolism has been a physiological constraint on brain architecture (8, 22). The results will also interest engineers who seek to design the next generation of low-power, intelligent computational devices. The proposed power budget of 3 W for computation and communication by gray matter is remarkably low, and, even after accounting for 9 W apparently lost to heat, the brain outperforms the typical laptop computer by nearly an order of magnitude. Previous work has suggested that ubiquitous laws of diminishing returns in the relation between information rates and power will drive efficient brain-inspired computers toward heterogeneous architectures where, at each scale, the overall function will occur through coordination between many specialized units, each processing information at a maximally efficient rate determined by the physical substrate (17). The paper by Levy and Calvert suggests that this drive to heterogeneity will have to be balanced against the cost of network learning and communication between the diverse components, because, although the actual computations by each element will be relatively cheap, the communication between them will be expensive (Fig. 1).

Acknowledgments

V.B. was supported, in part, by the Simons Foundation through Mathematical Modeling of Living Systems Grant 400425.

- 1 L. Sokoloff, "The metabolism of the central nervous system in vivo" in *Neurophysiology*, J. Field, H. W. Magoun, V. E. Hall, Eds. (Handbook of Physiology, American Physiological Society, 1960), sec. 1, vol. 3, pp. 1843–1864.
- 2 B. K. Siesjö, *Brain Energy Metabolism* (John Wiley, 1978).
- 3 D. F. Rolfé, G. C. Brown, Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol. Rev.* **77**, 731–758 (1997).
- 4 D. Attwell, S. B. Laughlin, An energy budget for signaling in the grey matter of the brain. *J. Cerebr. Blood Flow Metab.* **21**, 1133–1145 (2001).
- 5 A. Ames 3rd, CNS energy metabolism as related to function. *Brain Res. Brain Res. Rev.* **34**, 42–68 (2000).
- 6 V. Braitenberg, A. Schüz, *Anatomy of the Cortex: Statistics and Geometry* (Springer Science & Business Media, 2013), vol. 18.

- 7 L. C. Aiello, P. Wheeler, The expensive-tissue hypothesis: The brain and the digestive system in human and primate evolution. *Curr. Anthropol.* **36**, 199–221 (1995).
- 8 K. Fonseca-Azevedo, S. Herculano-Houzel, Metabolic constraint imposes tradeoff between body size and number of brain neurons in human evolution. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18571–18576 (2012).
- 9 D. B. Chklovskii, A. A. Koulakov, Maps in the brain: What can we learn from them? *Annu. Rev. Neurosci.* **27**, 369–392 (2004).
- 10 R. Sarpeshkar, Analog versus digital: Extrapolating from electronics to neurobiology. *Neural Comput.* **10**, 1601–1638 (1998).
- 11 W. B. Levy, R. A. Baxter, Energy efficient neural codes. *Neural Comput.* **8**, 531–543 (1996).
- 12 B. T. Vincent, R. J. Baddeley, T. Troscianko, I. D. Gilchrist, Is the early visual system optimised to be energy efficient? *Network* **16**, 175–190 (2005).
- 13 V. Balasubramanian, D. Kimber, M. J. Berry, 2nd, Metabolically efficient information processing. *Neural Comput.* **13**, 799–815 (2001).
- 14 K. Koch *et al.*, How much the eye tells the brain. *Curr. Biol.* **16**, 1428–1434 (2006).
- 15 J. A. Perge, J. E. Niven, E. Mugnaini, V. Balasubramanian, P. Sterling, Why do axons differ in caliber? *J. Neurosci.* **32**, 626–638 (2012).
- 16 V. Balasubramanian, P. Sterling, Receptive fields and functional architecture in the retina. *J. Physiol.* **587**, 2753–2767 (2009).
- 17 V. Balasubramanian, Heterogeneity and efficiency in the brain. *Proc. IEEE* **103**, 1346–1358 (2015).
- 18 P. Sterling, S. Laughlin, *Principles of Neural Design* (MIT Press, 2015).
- 19 W. B. Levy, V. G. Calvert, Communication consumes 35 times more energy than computation in the human cortex, but both costs are needed to predict synapse number. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2008173118 (2021).
- 20 B. Sengupta, M. Stemmler, S. B. Laughlin, J. E. Niven, Action potential energy efficiency varies among neuron types in vertebrates and invertebrates. *PLOS Comput. Biol.* **6**, e1000840 (2010).
- 21 P. Lennie, The cost of cortical computation. *Curr. Biol.* **13**, 493–497 (2003).
- 22 S. Herculano-Houzel, Scaling of brain metabolism with a fixed energy budget per neuron: Implications for neuronal activity, plasticity and evolution. *PLoS One* **6**, e17514 (2011).
- 23 R. Landauer, Irreversibility and heat generation in the computing process. *IBM J. Res. Dev.* **5**, 183–191 (1961).