

## Discussions with Dr. Adam Marblestone, 2019-2020

### Participants

- Dr. Adam Marblestone - Research Scientist, Google DeepMind
- Joseph Carlsmith - Research Analyst, Open Philanthropy

**Note:** These notes were compiled by Open Philanthropy and give an overview of the major points made by Dr. Marblestone. Some of these points were made in conversation, and some via electronic communication, at various points during 2019 and 2020. A few were made earlier than that. Dr. Marblestone also read a draft of Mr. Carlsmith's report on computation and the brain.

### Summary

Open Philanthropy reached out to Dr. Adam Marblestone of Google DeepMind as part of its investigation of what we can learn from the brain about the computational power ("compute") sufficient to match human-level task performance. The discussions focused on different factors relevant to evaluating compute estimates.

### FLOP/s

Dr. Marblestone thinks that it makes sense to think about the FLOP/s sufficient to run a task-functional model of the brain. And we can distinguish between the FLOP/s necessary to capture everything about what the brain does (including e.g. biological homeostasis), and the FLOP/s necessary to capture what's computationally relevant.

That said, FLOP/s is not the only relevant metric. Other relevant factors might include: memory access, fan in and fan out, local integration and storage processing, the complexity of the brain's three-dimensional wiring, parallelization, and coordination amongst local processors. Indeed, much of the brain's hardware might be devoted to distributed memory storage and look-up -- a possibility that could complicate FLOP/s estimates that attempt to bracket memory and memory bandwidth.

Relatedly, the architecture of a given computer (especially e.g. a standard von Neumann architecture) might create significant overhead. For example, the actual brain co-locates long-term memory and computing. If you had to store longer-term data in a conventional

RAM instead, many additional operations might be necessary in order to locate, address, and update relevant variables.

## **Breakdown of signaling mechanisms in the brain**

Mr. Carlsmith suggested the following categorization of signaling processes in the brain:

- canonical neuron signaling (e.g., mappings from synaptic inputs to firing decisions, and from pre-synaptic spiking to immediate impacts on the post-synaptic neuron),
- learning (e.g., experience-dependent changes to neurons and synapses over time),
- alternative signaling mechanisms (e.g. glia, ephaptic effects, and others),
- other unknowns.

This sounded like a reasonable classification to Dr. Marblestone, and one fairly similar to his own implicit ontology (though he would generally think of the first category under the heading of “fast, real-time computation”).

## **FLOPs per spike through synapse**

One approach to generating compute estimates is to budget one FLOP per spike through synapse. On this approach, each synapse has a weight, which determines the jump in post-synaptic voltage you get from a spike through that synapse. This jump is then added to the existing post-synaptic voltage. Because you’re adding two numbers, this is one FLOP.

However, Dr. Marblestone thinks that a larger number of FLOPs per spike through synapse might well be necessary. For example, instead of just adding two numbers, one could imagine the synapse “deciding” what impact to have on the post-synaptic cell based on a variety of factors, including: the presence of neuromodulators in different combinations, recent patterns of spiking at that synapse, the concentrations of different ions, the time since a backpropagating action potential reached the synapse from the apical tuft, and the particular combination of receptors and channels active at the synapse.

That is, a synapse could be a fairly complicated finite state machine, with on the order of dozens of states.

*Timescales*

Note, though, that for these synaptic decisions to imply an increase in required compute, it's not enough for various complex factors to influence the synaptic weight: the timescales of this influence matter. For example, it's possible while there is a lot of complex learning going on at synapses, adjustments to the synaptic weight are made fairly slowly, and hence the compute required per millisecond to implement these adjustments is small.

Compute increases are more likely to come from synaptic decisions that get computed on something like a per-spike basis. For example, you might need to do a lot of fast computation in order to set the synaptic "flag" variables involved in some neo-Hebbian three-factor learning rules, even if these variables take a long time to have effects.

It's also hard to rule out the possibility that even though relevant processes (e.g., neuropeptide signaling) are proceeding on slow timescales, there are so many of them, implicating sufficiently many possible states and sufficiently complex interactions, that a lot of compute is required regardless. Recent work from Prof. Stephen Smith and collaborators suggests that there are a very large number of neuropeptides in the cortex.

#### *Estimates in Sarpeshkar (2010)*

Prof. Rahul Sarpeshkar, in his 2010 book *Ultra Low Power Bioelectronics*, budgets 40 FLOPs per spike through synapse for second-order filter responses at synapses. Dr. Marblestone is unsure of the full details of how this number was generated, but he thinks it is grounded in something like the temporal shape of the response in the post-synaptic cell.

If you neglect this temporal shape, you'll get the wrong output: it matters that incoming spikes coincide and add up properly. However, Dr. Marblestone is not sure that you need the exact shape, or that it needs to be re-computed every time. Specialized hardware could also be helpful (though one can say this for everything). Overall, Dr. Marblestone expects it to be possible to either leave out or simplify this computation.

#### *Limitations on synaptic complexity*

There are limits to the degree of synaptic complexity that Dr. Marblestone finds plausible. For example, he thinks that thousands of FLOPs per spike through synapse would be very strange. This is partly because the number of computational mechanisms available inside a synapse is limited, and each will have a certain amount of stochasticity.

Dr. Marblestone does not think that synaptic activity could be at the level of sophistication involved in, to take an extreme example, re-orientating water molecules every picosecond

in order to form a lattice used to compute a cellular automata. Nor is a synapse doing anything that would require simulating full molecular dynamics.

## **Sources of intuition**

In general, Dr. Marblestone thinks of himself as somewhere in between “neuron computation is very simple, and our theoretical models of it are basically adequate” and “neurons could be performing very sophisticated forms of information storage and processing that we don’t yet understand.” His intuition is that better understanding of the brain’s biophysical mechanisms will yield important insight into how the brain structures and assembles itself, but that it won’t alter the picture of the brain’s computational capacity by orders of magnitude.

### *Stomatogastric ganglion*

One source of evidence informing Dr. Marblestone’s intuition in this respect comes from work by Prof. Eve Marder’s lab on the crustacean stomatogastric ganglion. These neurons create oscillations that can be very well modeled and understood using Hodgkin-Huxley type neuron models.

There are important molecular mechanisms at work, but these function to make the circuit robust. For example, across crabs, gene expression levels in equivalent stomatogastric neurons vary a lot, but they are correlated within a given crab, suggesting that there are many different gene expression solutions that can create the same functioning network, and that the cell’s mechanisms are set up to make sure the neurons find such a solution.

This system has many different possible states, which can be induced by different neuromodulators. But in any given one of those states, the real-time, fast computation is fairly understandable. Perhaps the whole brain is like that.

### *Evolution of neural mechanisms*

Some neural circuits, like ones in the spinal cord, are very simple. And one can imagine primitive synapses, involved in primitive computations like “if you get some dopamine, move this part of the jellyfish like so.” Genetic programs build these machines on the basis of relatively simple specifications, and you have to be able to reliably repurpose these machines without every molecule mattering.

Dr. Marblestone expects that evolution proceeded by reusing and recombining these relatively simple, reliable components.

## **Dendritic computation**

Mr. Carlsmith asked Dr. Marblestone's opinion of the argument that the tree-structure of dendrites will limit the additional computational complexity that dendritic computation can introduce. As Dr. Marblestone understands this argument, the idea is that while there may well be dendritic non-linearities, you should expect a tree-like structure of local interactions, and activity in one part of the tree can't exert fast, long-range influence on activity in another part. This rules out scenarios where, for example, any synapse can communicate with any other -- a scenario in which required compute could scale with the square of the number of synapses.

This argument is consistent with Dr. Marblestone's perspective, and he thinks it is very interesting, though it would be nice to formalize it more precisely.

*Beniaquev et al (2019)*

Dr. Marblestone is interested in work by Prof. Michael London and collaborators suggesting that you need a seven-layer deep neural network to replicate the input-output properties of a detailed biophysical model of a neuron -- results that may be in tension with the tree-structure argument just discussed. Dr. Marblestone would want to know more about how exhaustive the process was of searching for ways to perform the task using a smaller network. He thinks that this is a possible objection to arguments in favor of the adequacy of very simple models, but that given the caveats involved, it doesn't invalidate such models.

## **Learning**

Dr. Marblestone does not think that point neuron models are enough to fully explain why neural circuitry looks the way it does. He expects neural circuit design to have something to do with the biological implementation, via dendrites, of forms of learning akin to backpropagation.

Theories of backpropagation in the brain require that the same neuron is responsible both for the forward pass and the backwards pass. Current theories assign these tasks to different portions of the dendritic tree. The error signal received in the backwards pass

then needs to be able to update the processes involved in the forward pass. However, this would not be an extremely complicated or compute intensive process.

A lot of the learning models discussed in neuroscience are also significantly simpler than backpropagation: e.g., three-factor rules like “if the pre-synaptic neuron was active, and the post-synaptic neuron was active, and you had dopamine in the last ~3 seconds, then strengthen.”

Dr. Marblestone expects that both three-factor rules and backpropagation-type methods would imply compute burdens within an order of magnitude or two of estimates based on 1 FLOP per spike through synapse. He has not seen proposals for how second-order gradient methods of learning could be implemented in the brain.

There are also non-gradient methods of learning. For example, some people are interested in Bayesian belief propagation, though Dr. Marblestone is not aware of efforts to describe how this might be implemented at the level of e.g. dendrites. We shouldn't assume that the brain is doing some sort of gradient-based learning.

Online learning might be the hard thing that makes it necessary for the brain to have so many neurons.

## **Alternative signaling mechanisms**

### *Glia*

It's clear, at this point, that glia are playing a computational role in the brain. For example, a recent paper from the lab of Professor Misha Ahrens shows that glia, on fast time-scales, accumulate evidence that drives zebrafish behavior. But Dr. Marblestone does not see evidence that including glia would alter compute estimates by orders of magnitude.

### *Gap junctions*

Sometimes the coupling between neurons created by gap junctions is so fast that they are treated as one neuron for modeling purposes. Gap junctions are also often thought of as supporting some kind of oscillation or globally coherent behavior that might not require a lot of computation. Whether gap junctions could create more computationally-expensive, non-linear interactions between different parts of neurons is an interesting question.

### *Ephaptic effects*

Even if ephaptic effects are fast, to assess whether they could massively increase compute estimates, Dr. Marblestone would want to know whether they could support the type of combinatorics that synapses can.

## **Other unknowns**

Dr. Marblestone thinks that the probability that the field of neuroscience rests on some very fundamental paradigm mistake is very low. We're missing a unified explanation of behavior and intelligence, but the basic picture of neurons as modular elements with some sort of transfer function and some sort of (possibly complicated) learning rule, without some extreme amount of internal computation taking place inside the cell, seems fairly solid to Dr. Marblestone.

That said, Dr. Marblestone also assigns substantive probability to the possibility that his own compute estimates are based on some kind of basic confusion or mistake.

## **Low-end estimates**

One way you might need less than 1 FLOP per spike through synapse is if you don't need to model all of the neurons in the brain. For example, it might be that all of the neurons and synapses in the brain are there in order to make the brain more likely to converge on a solution while learning, but that once learning has taken place, the brain implements a function that can be adequately approximated using much less compute.

A large amount of neuroscience treats populations of neurons as redundant representations of high-level variables relevant to information-processing.

Many models in computational neuroscience focus on spike rates, as opposed to precise spike timings. However, Dr. Marblestone is not sure that this would imply much computational savings, relative to a  $\sim 10$ -100 FLOP per spike through synapse estimate.

## **Comparative success in peripheral sensory/motor system**

There are only so many things in neuroscience that have really been fully explained and understood. However, there are peripheral sensory and motor systems (e.g., the retina, the cochlea, the spinal cord) that people understand decently well.

## *The retina*

The retina looks like a pretty good success story for mechanistic neuroscience. We can predict its behavior fairly well both using mechanistic neuron models, and by training deep neural networks on the overall input/output function. These results do not suggest some mysterious and unknown computation mechanism inside neurons, or some deep problem in predicting neural behavior. Indeed, Dr. Marblestone finds recent efforts to predict retinal spike trains using deep neural networks fairly persuasive evidence that reduced, DNN-type models of the retina are feasible.

## *Selection effects*

Dr. Marblestone does not think that selection effects nullify the evidence provided by our understanding of peripheral sensory and motor systems. E.g., it's not that we did experiments on a bunch of systems, and some of them we couldn't figure out, and some of them we could. Rather, the distribution of neuroscientific success has more to do with our experimental access to peripheral sensory/motor systems, together with differences in the types of theories you would need to have in order to explain more architecturally-complex circuits deeper in the brain.

Similarly, Dr. Marblestone does not think that the fact that we can't simulate *C. elegans* is a good argument for any kind of special computation taking place within *C. elegans* neurons. Lots of other explanations are available: notably, that it's very difficult to figure out the right parameters. Dr. Marblestone also doesn't see a reason that *C. elegans* would need huge amounts of compute: all it really needs to do is swim around and follow the food gradient.

## **Comparisons with V1**

Dr. Marblestone does not think it obvious that the visual cortex should be thought of as doing something like object-detection. It could be, for example, making a more complicated transition model based on all of its multi-modal inputs, predicting future inputs and rewards, or doing some kind of iterative inference procedure. We just don't know quite how high-dimensional or complicated the task the visual system performs is. So any compute estimates based on comparisons between the visual system and current deep neural networks are highly uncertain.

## **Epistemic limitations**

Neuroscience is extremely limited by available tools. For example, we have the concept of a post-synaptic potential because we can patch-clamp the post-synaptic neuron and see a change in voltage. When we become able to see every individual dendritic spine, we might see that each has a different response; or when we become able to see molecules, we might see faster state transitions, more interesting spatial organization, or more complicated logic at the synapses. We don't really know, because we haven't been able to measure.

It's also possible that some theories in neuroscience emerge and persist primarily because (a) they are the type of simple ideas that humans are able to come up with, and (b) these theories explain some amount of data (though it's unclear how much). It's hard to formulate complicated ideas about how the brain works that can then be made testable.

## **Opinions in the field**

Neuroscience is a very broad field. People study very different things (e.g., Alzheimer's, place cells in bats, human cognition), using very different tools, and with very different levels of expertise about topics that are not their speciality (e.g., some very experimentally-focused people may not have a strong background in computational theory, and some very computationally-oriented people may not have a strong background in underlying biophysical mechanisms).

Partly for this reason, it is very unlikely that you'll be able to get a large number of neuroscientists to agree on a FLOP/s estimate for the brain. One possible view is that available data simply does not constrain what neurons could be doing, and that until we have a very detailed understanding of the brain's molecular processes, we simply can't say anything.

## **Overall best guesses**

Dr. Marblestone is fairly comfortable with one FLOP per spike through synapse as a low-end estimate, and ~100 FLOPs per spike through synapse (roughly comparable to the estimate offered by Prof. Rahul Sarpeshkar) as a high-end estimate. His best guess is 10-100 FLOPs per spike through synapse.

*All Open Philanthropy conversations are available at  
<http://www.openphilanthropy.org/research/conversations>*