Discussions with Dr. Jess Riedel, Spring 2020

Participants

- Dr. Jess Riedel - Senior Research Scientist (Physics), NTT Research
- 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. Riedel. Some of these points were made in conversation, and some via electronic communication, at various points during spring of 2020 (Dr. Riedel also read drafts of different sections of Mr. Carlsmith's report on computation and the brain). A few were originally made during a conversation between Dr. Riedel and Asya Bergal (then a Fall Research Analyst at Open Philanthropy) in fall of 2018.

Summary

Open Philanthropy reached out to Dr. Jess Riedel of NTT Research 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 the application of Landauer's principle to the brain.

Landauer's principle

Landauer's principle states that erasing a bit of information requires a minimum energy expenditure -- specifically, kT ln2, where k is Boltzmann's constant, and T is the absolute temperature.

Landauer's principle follows almost trivially from basic principles of thermodynamics. Indeed, it can be understood simply as a rewriting of the definition of temperature. At a fundamental level, temperature is defined via the change in energy per unit change in entropy (up to a proportionality constant, Boltzmann's constant). The practical and folk definitions of temperature, which focus on the amount of energy in a system (e.g., the kinetic energy of vibrating atoms), can be recovered from this more fundamental definition in all but a small number of exceptional cases.

As the energy in a non-exceptional system increases, the number of states it can be in (and hence its maximum possible entropy) increases as well. If you have a system with a certain amount of energy, and you want to decrease its entropy, you need to put that entropy somewhere else, because total entropy is non-decreasing. Temperature gives us the exchange rate between energy and entropy. If you want to put some unit of entropy into a heat bath, you have to pay an energy cost, and the temperature of the bath *is* that cost.

A system like a brain or a computer contains non-information-bearing degrees of freedom that can absorb a finite amount of entropy. However, because the brain/computer is continuously processing and using energy, you can't keep dumping entropy into those degrees of freedom indefinitely. Eventually, you need to start pushing entropy into the environment.

If we assume that the states of the computer and the environment are not correlated (or at least, not in a way that we can realistically keep track of), then the total entropy will be the entropy of the computer plus the entropy of the environment. If the entropy of the computer goes down, the entropy of the environment must go up.

There is some dispute over Landauer's limit in the literature. Whether the basic assumptions it follows from apply in the real world is somewhat subtle.

In certain rare environments, you can decrease entropy by paying costs in conserved quantities other than energy (for example, you can pay costs in angular momentum). But this is not relevant in the context of the brain.

Landauer's principle and the brain

Mr. Carlsmith asked Dr. Riedel's opinion of the following type of upper bound on the compute required to replicate the brain's task performance. According to Landauer's principle, the brain, given its energy budget (~20 W) can be performing no more than ~1e22 bit-erasures per second. And if the brain is performing less than 1e22 bit-erasures per second, the number of FLOP/s required to replicate its task-performance is unlikely to exceed 1e22.

Dr. Riedel is very convinced by the claim that because of Landauer's principle, the brain can be implementing no more than ~1e22 bit-erasures per second. And he also thinks it very reasonable to infer from this that the brain's task performance can be replicated using less than 1e22 FLOP/s, conditional on the assumption that the brain's computation is well-characterized as digital and/or analog computation that can be simulated on a digital computer with modest overhead (he assigns some small probability to this assumption being false, though he would find its falsehood fairly shocking).

Indeed, Dr. Riedel expects the amount of computation performed by the brain to be much lower than the upper bound implied by Landauer's principle. This is partly because, from a basic physics perspective, the vast majority of what's going on in the brain (e.g., cell maintenance, other thermodynamic processes inside cells) generates entropy but has nothing to do with the computations that are happening.

Dissipation in biological systems

Presumably, we think we basically understand cases where the brain is sending very simple signals, like the signal to kick your leg. We know that the nerves involved in conveying these signals are operating in an irreversible way, and burning way more energy than the Landauer limit would say is necessary to communicate the number of bits needed to say e.g. how much to move the muscle. It seems this energy is required partly because the nerve is a big and complicated system, with many moving parts, so redundancy is necessary.

Biology may be very energy efficient in certain cases, but Dr. Riedel still thinks it very unlikely that the efficiency of the brain's computation is anywhere near Landauer's limit. There are also likely to be other examples in which biology is extremely inefficient relative to Landauer's principle, due to other constraints (for example, cases in which biological systems use chemical gradients involving billions of molecules to communicate ~5 bits of information).

Humans can, if necessary, create very special-purpose computational devices that get close to Landauer's limit (this is what "experimental tests" of Landauer's limit attempt to do), and our power plants, considered as thermodynamic heat engines, are very efficient (e.g., nearing thermodynamic bounds). However, our useful, scalable computers are not remotely close to the minimal energy dissipation required by Landauer's principle. This appears to be an extraordinarily hard engineering problem, and it's reasonable to guess that brains haven't solved it, even if they are very energy efficient elsewhere.

Relevant temperature

The temperature relevant to applying Landauer's limit to the brain is essentially that of the skull and blood. Even if the temperature outside the body is at a lower temperature, the brain will have to push entropy into its environment via those conduits. If there were some other cold reservoir inside the brain absorbing entropy (there isn't), it would quickly be expended.

In order for a different temperature to be relevant, the brain would need to be doing something very exotic, like coupling to cold baths outside the body using electromagnetic waves.

FLOP/s required per bit-erasure

Dr. Riedel thinks it very unlikely that you need more than a FLOP to replicate whatever the brain does per bit-erasure. In general, he expects that essentially all that the computations you want to do in the context of a biological system contain a very large number of irreversible steps (absent special steps to make them reversible), and therefore require a large number of bit erasures. This expectation is grounded in part in the following observations:

- When humans write software to accomplish human objectives, they use a lot of irreversible steps (though there are some non-atomic reversible intermediate computations, like Fourier transforms).
- When the world has some simple feature (e.g., the position and velocity of a rock heading towards your head), this feature is encoded in very complicated intermediate systems (e.g., the trillions of photons scattering from the rock and heading towards your eye). The brain has to distill an answer to a high-level question (e.g., "do I dodge left or right?") from the complicated intermediate system, and this involves throwing out a lot of entropy.
- Some useful computations, like factoring a composite number, involve exponentially more operations than the number of input bits, and hence would require very large numbers of bit-erasures (if you use irreversible operations).

Reversible computation

There is a simple algorithm for converting a computation that uses logically irreversible operations into an equivalent computation that uses logically reversible operations. This allows you to avoid almost all of the relevant logical bit-erasures.

For large computations, this conversion adds only a modest overhead in required time and memory. For example, the algorithm described in Charles Bennett's 1989 paper "Time/Space Trade-Offs for Reversible Computation" involves slow-downs of at worst a multiplicative factor, around 2-3X as slow.

However, if (as in current conventional computers) you're dissipating thousands of kT per operation, it isn't worth transitioning to logically reversible operations, because other forms of energy dissipation dominate the Landauer-mandated energy costs of logical irreversibility.

Dr. Riedel is skeptical of objections to the viability of reversible computing that appeal to the bit-erasures involved in receiving new inputs and writing new final outputs. It's true that reversible computing paradigms require bit-erasures for this, but for most interesting computations, the intermediate memory usage is much (often exponentially) larger than the input and output data.

Information processing in biological systems

In the context of a computational system, you can think of an "operation" as a small computation that can be treated as atomic, at least with respect to a particular architecture.

"Encode" generally implies a reversible mapping from inputs into outputs -- a mapping that could be computationally simple, or computationally complex. To "process" information generally means to perform a not-completely-trivial computation on it (where this computation could be reversible or irreversible).

One possible information-theoretic definition of signaling between cells (as opposed to other forms of functionally structured causal interaction) is that signaling occurs when the value of one cell's causal impact on a second cell arises in virtue of the correlation between the state of the first cell and its impact on the second cell (as opposed to, e.g., one cell sending the other one resources irrespective of the first cell's state). That is, signaling is the term you use when you want causal influence in some conditions, but not in others, just as the term "computation," understood as a mapping from sets of inputs to sets of outputs, generally implies that all inputs do not lead to the same output.

The information signaled is also generally independent of the specific physical medium (e.g., the cells would be just happy using one signaling molecule vs. another), and preserved via reversible manipulations.

From a computational perspective, electrical synapses lack gain -- the ability to amplify signals. Dr. Riedel recalls that gain is a key property of computational units like transistors.

In applying Landauer's principle to the brain, it may be more helpful to think about the number of simple logic operations necessary to replicate the brain's computation, as opposed to the number of FLOP/s. Logic operations are a more natural and basic computational unit, with a more direct connection to bit-erasures. However, because a

FLOP can be constructed out of simpler logic operations, one can translate between the two metrics fairly easily.

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