Challenge 22

Neuromorphic computing at scale

- Koushik Seggari (984570188)

1. **Neuronal scalability is likely the biggest challenge**: The ability to support extremely large numbers of neurons – is very hard to achieve with current technology. Today, even reaching hundreds of millions of neurons requires connecting many neuromorphic chips together. Solving neuronal scalability would be transformative: it could allow brain-scale neural networks and even real-time full human brain simulations, enabling neuromorphic systems to tackle incredibly complex problems that are out of reach today

2. The article suggests that a modest neuromorphic system achieving a dramatic result – like how AlexNet's success on two GPUs set off the deep learning boom – could be the trigger. For example, a spiking neural network that outperforms conventional AI on a tough task (with far less power) would showcase neuromorphic computing's potential. This kind of breakthrough would inspire larger neuromorphic projects and open the door to new applications, from brain-scale simulations to solving complex problems (like large graph or optimization tasks) much more efficiently than traditional hardware

3. Build a common software layer across neuromorphic platforms. The paper highlights the need for a hardware abstraction layer that makes different neuromorphic chips interoperable. In practice, this means creating standard tools and interfaces so that a spiking neural network model can be "compiled" and run on any neuromorphic hardware, like how we can run neural networks on CPUs, GPUs or TPUs interchangeably. By developing such unified frameworks (and adding missing developer-friendly tools), we can bridge the hardware–software gap and ensure that neuromorphic systems work smoothly across different platforms

4. Use energy-centric metrics and standardize benchmarks. Beyond accuracy or throughput, metrics like energy efficiency (energy per operation or per synapse), power usage, latency, and robustness to noise give a fuller picture of neuromorphic performance. Often, multiple factors are combined – for example, measuring the energy-delay product (which balances speed and energy) or "energy per synaptic event" – to compare systems fairly. To standardize evaluations, the community is working on common benchmarks and metrics tailored to neuromorphic computing. Efforts like collaborative benchmarking suites aim to ensure that different neuromorphic chips and algorithms can be evaluated on equal footing, much as conventional computers are compared using agreed-upon benchmarks

5. Merging new memory tech with neuromorphic design unlocks in-memory computing. Emerging memory devices (such as memristors or phase-change memory) can both store data and perform computation in place, blurring the line between memory and processor in a way traditional systems can't. For example, an array of memristors can act like a neural network layer: it holds synaptic weights and carries out multiply-and-accumulate operations directly via electrical currents (Ohm's law), resulting in extremely compact and low-power computing. This approach enables massively parallel, brain-like processing and avoids the bottleneck of constantly moving data between separate CPU and memory units. Promising research directions include improving these memory technologies to meet practical requirements – ultra-low energy use, high endurance (many write cycles), nanoscale size, and easy integration with standard chips. By embedding such memory devices into neuromorphic architectures, we could create novel systems that learn and compute within the memory elements themselves, achieving forms of computing not possible on conventional digital platforms.