**BERB review for “Considerations for neuromorphic supercomputing in semiconducting and superconducting optoelectronic hardware” by Advait Madhavan**

Summary:

Optimization aspects of a fully-dedicated opto-electronic brain-scale neuromorphic hardware system with 1010-11 neurons and an associated fan-in/fan-out of 104-5 synapses per neurons are considered.

Optical interconnect can provide binary, low energy, and low latency communication with minimal loss while the electronic circuits can provide complex neuronal and synaptic computational functionality.

Electronics split into room temperature and low temperature approaches with cooling costs accounted for.

Each section goes over important aspects of the system such as transmission, reception, computation and memory, followed by a conclusion that sets the goals for semiconducting and superconducting platforms, while describing the challenges associated with them.

Comments:

Overall, this paper was a good read and, in my opinion, lays a clear picture of what needs to be done if such an optimistic target is to be achieved. Describing the individual features of such a network such as reception, transmission, computation etc, and providing detailed analysis of opportunities and roadblocks makes this paper a good reference to assess the progress that has already been made and needs to be made.

I have arranged these comments by first providing a few general comments, associated with the vision and message of the paper and its tone, followed by some section specific comments. Please feel free ignore any/all of them.

1. Vision and tone:
   1. The current paper is very thorough and reads like a mixture of a roadmap and review concerning the challenges that face the design of the system described above and I believe that it is well referenced. Speculations regarding important features of such a system, such as transmission, reception and computation are made, based on assumptions of the technology used, resulting in certain estimates of energy and area. The paper concludes by describing the challenges that are going to be faced when building such a system.   
        
      Though this paper is a very detailed, fun read and provides a large amount of information, it is difficult to contextualize. Also, considering the width of the subject matter, I guess I was expecting a punchier punch line. How to understand this information, such as the key tradeoffs involved and simple first order estimations of the expected performance of such a system would be very beneficial. Also, how such a system would compare to state-of-the-art neuromorphic platforms/systems (similar scale as this work) such as Loihi, TrueNorth, Spinnaker etc. would in my opinion significantly strengthen the paper. I would suggest adding a table which would report optimistic estimates for semi- and/or super-conducting technologies, with some assumptions for memory and compute circuits. Estimated characteristics such as technology node, core area, learning rules, neuron and synapse density, synapse and neuron operation energies, and estimated speed/time constant would be useful. A similar but very detailed version of such a table is included in DOI: 10.1109/TBCAS.2018.2880425 (A 0.086-mm2 12.7-pJ/SOP 64k-Synapse 256-Neuron Online-Learning Digital Spiking Neuromorphic Processor in 28-nm CMOS, Frenkel et.al, Table 1).  
        
      The idea here is that if industry is going to want to pursue something of such a scale, 10 to 15 years down the line, the technology has to be competitive with existing systems. I suggest that you use the most optimistic estimates so the potential benefits to spending large amount of dollars and man-hours in this effort becomes clear. In my head, this would add a fixed (albeit speculative) target for such roadmap like paper, hence strengthening it.
   2. This paper describes a fully dedicated approach to implementing neuron models. This means that each synapse in the model will have a dedicated hardware designed for it. Section 5 describes a pessimistic case of L = 3.5 for a network of 106 neurons and k =100. For such a case, isn’t it true that “Which of the k synapses are going to be connected to each neuron?” is subject to change and plasticity from the learning algorithm. Or do you pre-determine the connectivity and hierarchy? This point is not clear from the system level perspective. Do you just build all 106 synapses and keep the most of them programmed to 0? Put simply I guess what I am asking is how does a fully dedicated approach as described in this work with a plastic hierarchy, deal with varying synaptic connectivities that may occur due to continuous on-line learning? Building all-to-all connectivity at the 106 neuron scale may be prohibitively expensive and redundant.
2. Section specific:
   1. *Section 2.1.1: Paragraph 3: “We assume that the detection of a single photon will be treated as the registering of a spiking event.”*  
      Is this a reasonable assumption? From what I understand, circuits such as SPADs (Single Photon Avalanche diodes) that are used to detect single photons, are prone to a lot of noise. In such circuits, there is a dark count rate which comprises of the noise floor of events that trigger avalanches in the diode, while the signal is governed by counts that are significantly larger than this. In this case, if each event is being treated as spike, wont this be very susceptible to noise? Spurious spikes can corrupt data as well as cost more unnecessary energy. Is there some noise analysis that can be added to this? For example, given the bandwidth of operation, the total noise in the frequency band in question can be integrated. This can provide estimates of signal magnitude required to overcome receiver noise.
   2. *Section 4. “Synaptic Memory”*Movement of data(Input data, weight data and gradient data) in such a large-scale system is what consumes a majority of the energy cost. How is the memory hierarchy of such a system designed? Is there buffering for local data storage? What is the energy cost of an always-on, on-line learning based memory. An estimation with assumptions of memory technology could be useful.
   3. *Section 5.2: “Generic Spatial Constraints”*This section does not account for a network structure of a large-scale neuromorphic system. In a fully connected approach, with 106 neurons per plane, and 104 planes, how will the all-to-all network be implemented? What is it’s hierarchical structure and what are its performance metrics? How much area on each electronic plane is going to be dedicated to the network? Most of the other large-scale neuromorphic systems such as Loihi, Spinnaker, BrainScales etc. explicitly define network topologies and spike routing schemes.