All-Optical Graphical Models for Probabilistic Inference

Pierre-Alexandre Blanche, Madeleine Glick, John Wissinger, Khanh Kieu, Masoud Babaeian, Houman Rastegarfar, Veysi Demir, Mehmetcan Akbulut, Patrick Keiffer, Robert A. Norwood, Nasser Peyghambarian, and Mark Neifeld.

College of Optical Sciences, The University of Arizona, Tucson AZ 85721

Considering that high performance electronic computation has become extremely efficient, for an optical hardware accelerator to be relevant, it must solve a type or a set of problems where its electronic counterpart is still struggling in term of size, energy, or time. We have identified one such challenge as the minimization of large scale Ising Hamiltonians when the number of particles is on the order of a million. Here we discuss an algorithmic approach based on probabilistic inference using graphical model and message passing.

A graphical model is a probabilistic network where the state of the nodes depends on a structural component (the node connections or edges), and a parametric component (the initial conditions). It is a powerful method that can be adapted to solve a large variety of problems and has found application in numerous cases. The complexity of the graph scales with the number edges, and the probability vector size. We found that dedicated electronic systems can easily solve densely connected graphs with tens of thousand of nodes. Therefore, the optical solutions under consideration must be able to scale above that lower bound, which puts size and power consumption constraints on the type of hardware used in the implementation.

Among the different algorithms used in graphical model, we selected message passing which is schematically represented in figure 1 for the belief update of node n. This algorithm involves vector matrix multiplications (VMM) that address the interaction of node n with every other node, followed by a vector dot product (.Π) that aggregates all the vectors, and finally a normalization function that rescales the vector into the probability domain.

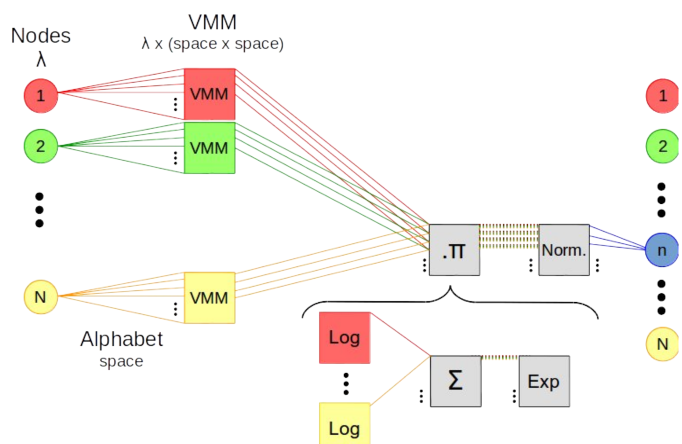

Figure 1: Schematic representation of the message passing algorithm to update the probability vector of node n. When nodes are not connected, the matrix of the VMM connecting them is null.

In the implementation we have selected for its most promising scalability, the nodes are encoded as different wavelengths (represented by different colors in figure 1), and the probability vector is distributed in space. This can be understood by the fact that there are many more nodes than elements in the vector, and wavelength multiplexing is a powerful way to compress the information.

Considering optical VMM has already been described in the literature, we are focusing our effort on the dot product and normalization which are operation that are not easily computed in the optical domain. We replaced the dot product function by its log-sum-exponent mathematical equivalent, which can be achieved by two photon absorption (TPA) followed by a fan in, and a saturable absorber

(SA) respectively. The advantage of these mechanisms is that they are passive and don't require additional power.

However, TPA and SA are not perfect logarithm and exponential functions, but close approximations:

$$TPA: \; I_{out} = \frac{I_{\text{in}}}{1 + C_{TPA} I_{\text{in}}} \approx \log\left(I_{\text{in}}\right) \tag{1}$$

$$SA: \; I_{out} = I_{in} \exp\left[-\alpha \frac{L}{1 + I_{in}/I_{sat}}\right] \approx \exp\left(I_{\text{in}}\right) \tag{2}$$

If we include the fan in operation for summation, we obtain the following equation for the dot product:

$$\prod_{Node\,1}^{N} I_{\text{in}_i} \simeq A \times SA\left[B \times \sum_{Node\,1}^{N} TPA\left(C \times I_{\text{in}_i}\right)\right], \tag{3}$$

where A, B and C are coefficients that reflect either gain or attenuation in the system.

We have simulated the effect of the TPA and SA discrepancies on the convergence of the graph model using 100 nodes, and have demonstrated that the TPA and SA functions can fulfill the requirements of the message passing algorithm. The redundancy in densely connected graph actually helps to ensure the convergence.

In figure 2 we see that, when some of the edges are dropped at random from a fully connected graph (density 100%), the algorithm starts having trouble converging only when the density drop below 20%, even for perfect logarithm and exponential functions (0% error, blue line). The other lines have been simulated with increasing error injected in the computation of the logarithm and exponential as it could be expected from SA and TPA devices. For 10% error (green line), the algorithm starts converging to the "wrong" ground state when the connection density is beneath 30%.

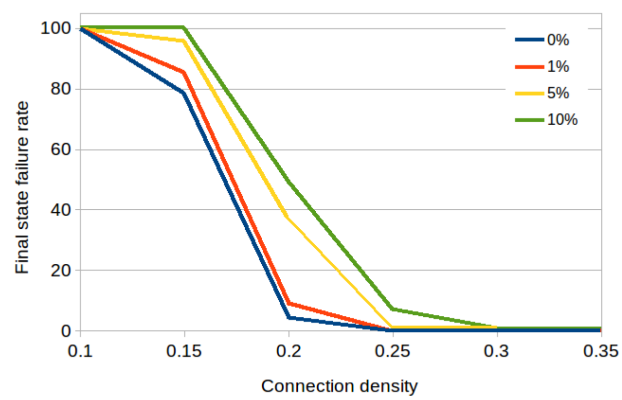

Figure 2: Algorithm failure rate according to connection density for a 100 node graph. Different curves has been calculated with increasing amount of error in the exp and log functions.

Our first physical implementation of the multiplication function between two optical sources used silicon waveguides for TPA and carbon nanotubes in a fiber taper for SA. An EDFA amplifier was used to provide gain before the SA. This initial result has shown that, although we can achieve an RMS error below 1% for the value of the product, the dynamic range was limited to 0.41dB. For our scaled up system we will use nano-scale low power devices. These devices must also fulfill the requirements of equation 3 without requiring excessive gain in the system.

Authors acknowledge the support of the Office of Naval Research under grant #N00014-14-1-0505, as well the NSF ERC CIAN under grant #EEC-0812072.