**Back Propagation Outline**

1. Why back propagation: scheduling algorithm assume G is dag, which only models the forward propagation. Back propagation consumes a significant amount of computation time and hence needs to be accounted for.
2. How to Schedule to back propagation: schedule gradient on the same machine as the operation in forward pass.
3. Experiment : ratio between gradient computation time and forward pass operation computation time is approximately constant. [insert graph]
4. Reason for not considering applied gradient associated with variable nodes: it does not have the same dependency graph and can be done at any time in the back propagation process after it is computed
5. Makespan guarantee from proof: total makespan of 1 iteration (C0 + 1)T. T is the original scheduling makespan. C0 is the max ratio between gradient computation and original operation computation.
6. Approximation ratio does not change guarantee from proof (constant case).
7. Since communication does not increase proportionally, need new proof for OPT guarantee