Reference (Communication Efficiency)

1. McMahan, Moore, Ramage, Hampson, & Arcas. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*, 2017.
2. Stich. Local SGD converges fast and communicates little. In *ICLR*, 2018.
3. Li, Sahu, Talwalkar, & Smith. Federated optimization in heterogeneous networks. *arXiv*, 2018.
4. Wang & Joshi. Cooperative SGD: A unified framework for the design and analysis of communication-efficient SGD algorithms. *arXiv*, 2018.
5. Fan & Cong. On the convergence properties of a k-step averaging stochasticgradient descent algorithm for nonconvex optimization. In *IJCAI*, 2018.
6. Lin, Stich, and Jaggi. Don't use large mini-batches, use local SGD. arXiv, 2018.
7. Li, Huang, Yang, Wang, & Zhang. On the convergence of FedAvg on non-IID data. *arXiv*, 2019.
8. Khaled, Mishchenko, Richtárik. First analysis of local GD on heterogeneous data. *arXiv*, 2019.
9. Yu, Yang, Zhu. Parallel restarted SGD with faster convergence and less communication: Demystifying why model averaging works for deep learning. In *AAAI*, 2019.

…

(And many other work which I am unaware of.)