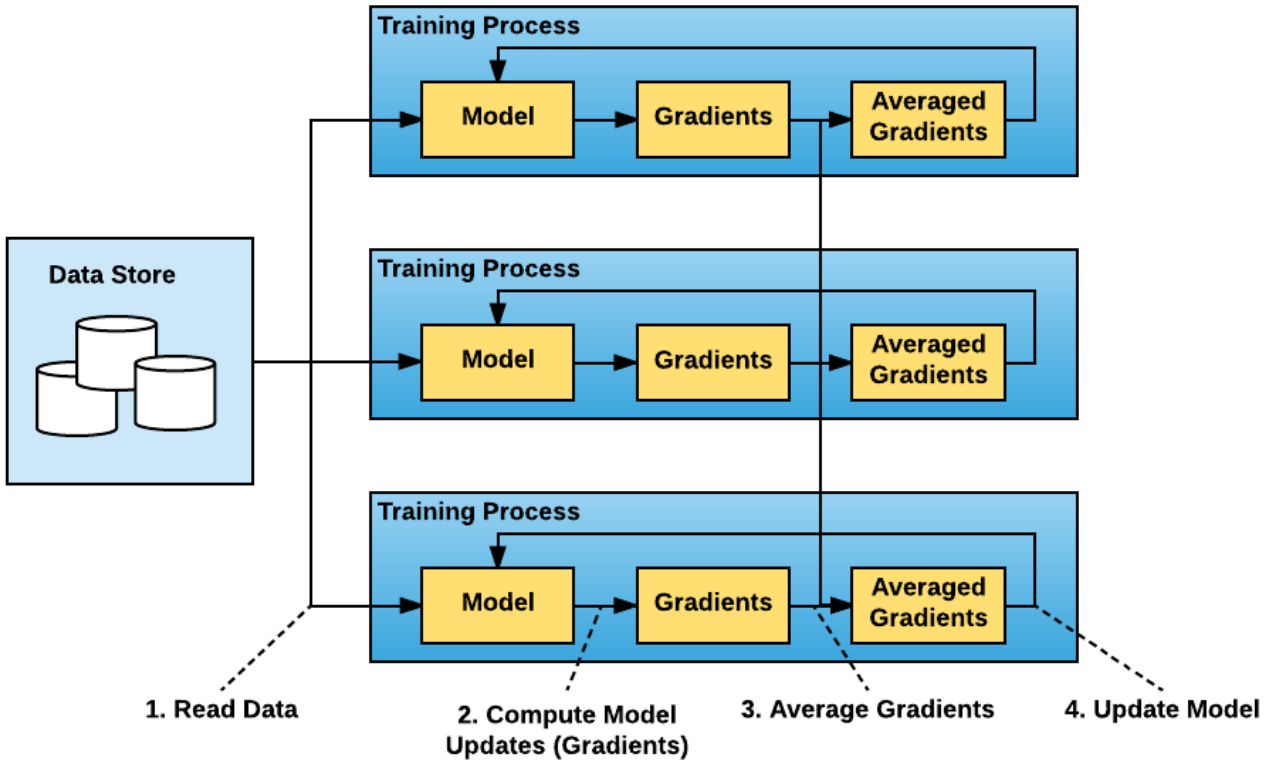
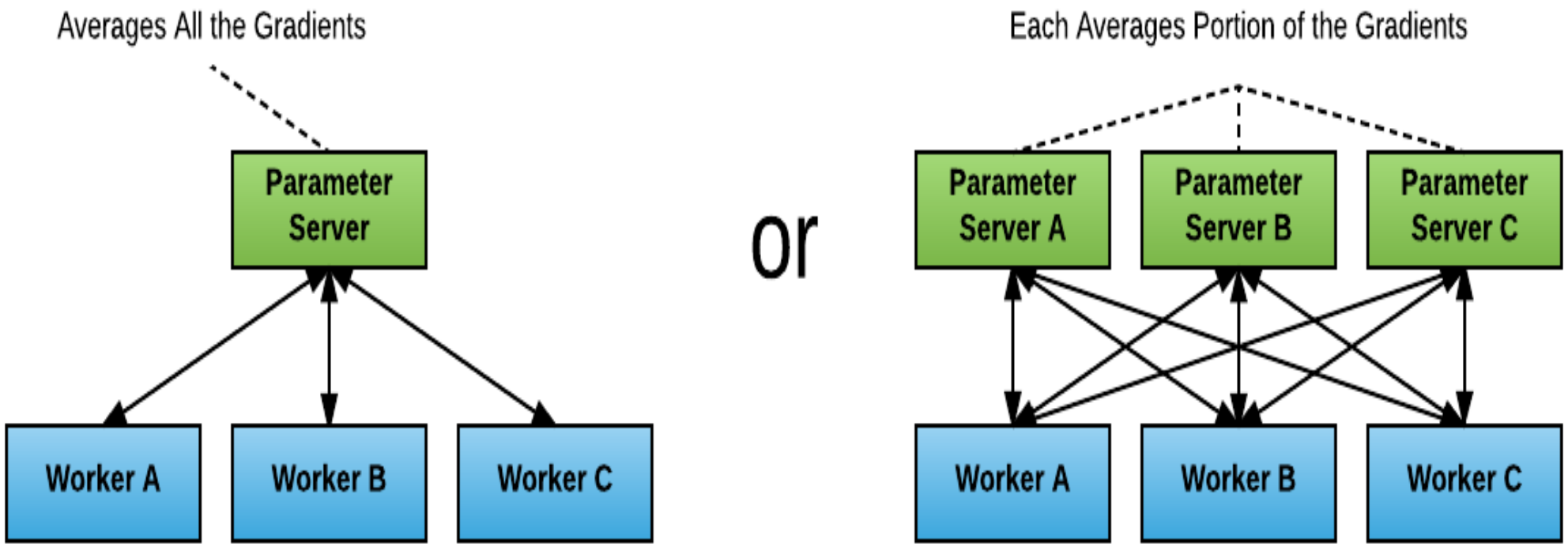
Multi-Node Training

Multi-GPU / Multi-Node Training



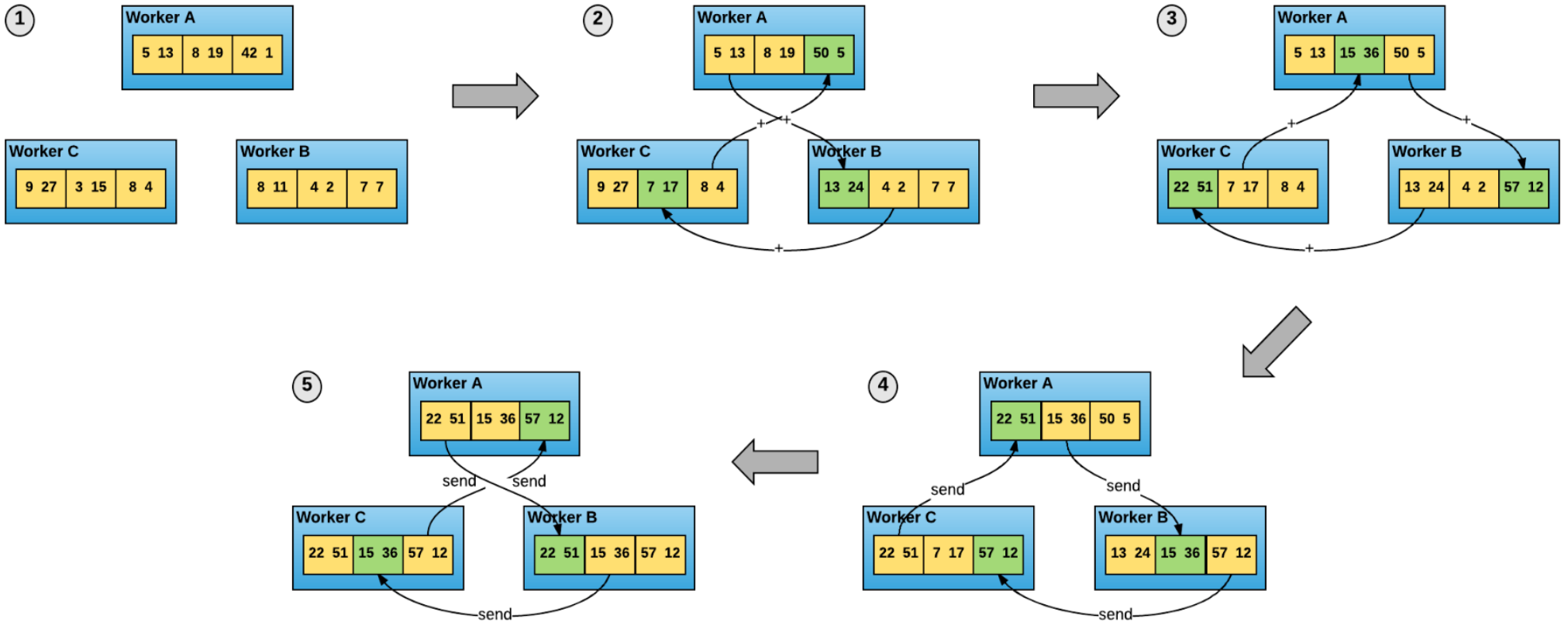
(source: Horovod)

Parameter Servers



(source: Horovod)

Direct Parameter Exchanges



(source: Horovod)

Different Kinds of "Multi-Node Jobs"

- model fits into single GPU
- model needs multiple GPUs but fits on single node
- model needs multiple nodes
- model is replicated across multiple GPUs on a single node
- model is replicated across multiple nodes

How do you distribute?

- node / VM deployment
- Kubernetes
- MPI / NCCL / Gloo
- Horovod

Mixed Communications

- PCI bus for CPU, CPU$\leftrightarrow$GPU (computations involving single model)
- RDMA to CPU memory (data)
- GPUdirect for I/O to GPU, RDMA (data loading, inter-node parameters)
- NVLINK between GPUs (multi-GPU SGD within a node)
- standard Ethernet to CPU memory (control messages, maybe data)

Bottlenecks of Single Node Jobs

- limited number of GPUs per node
- limited PCI bus bandwidth
- limited CPU power
- limited local storage