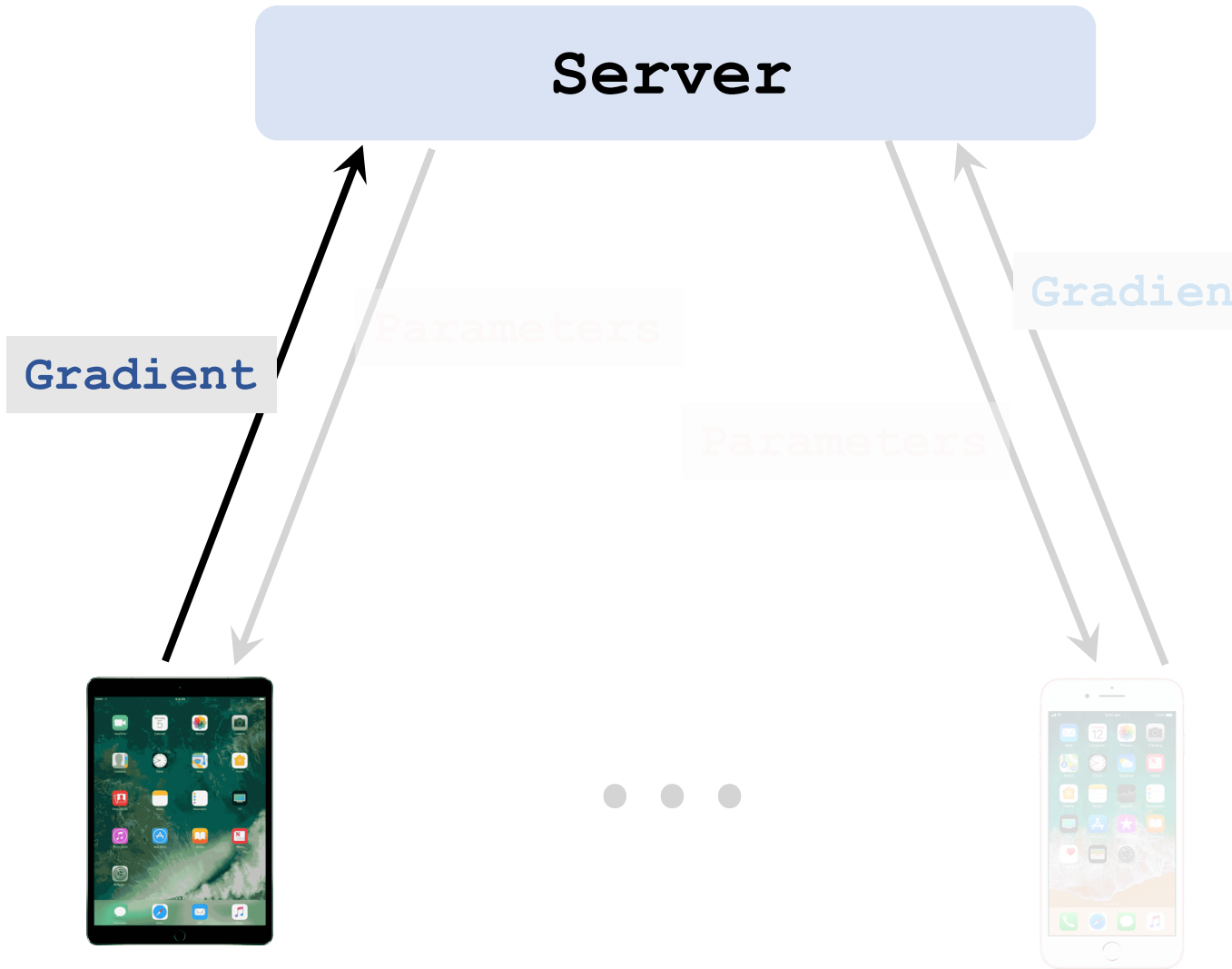


Let us recall **parallel gradient descent**



The server performs:

1. Receiving gradients $\mathbf{g}_1, \dots, \mathbf{g}_m$ from all the m workers.
2. Computing $\mathbf{g} = \mathbf{g}_1 + \dots + \mathbf{g}_m$.
3. Updating model parameters:

$$\mathbf{w} \leftarrow \mathbf{w} - \alpha \cdot \mathbf{g}.$$