Back-propagation updates the model parameters of each layer

Input from a minibatch at node i

Forward pass produces output and calculates loss

$\frac{1}{n}\sum\nabla g_k^{i,2}$ $\nabla g_k^{i,2}$ $\frac{1}{n}\sum\nabla g_k^{i,3}$ $\nabla g_k^{i,3}$ ... $\frac{1}{n}\sum\nabla g_k^{i,L-1}$ $\nabla g_k^{i,L-1}$ $\frac{1}{n}\sum\nabla g_k^{i,L}$ $\nabla g_k^{i,L}$
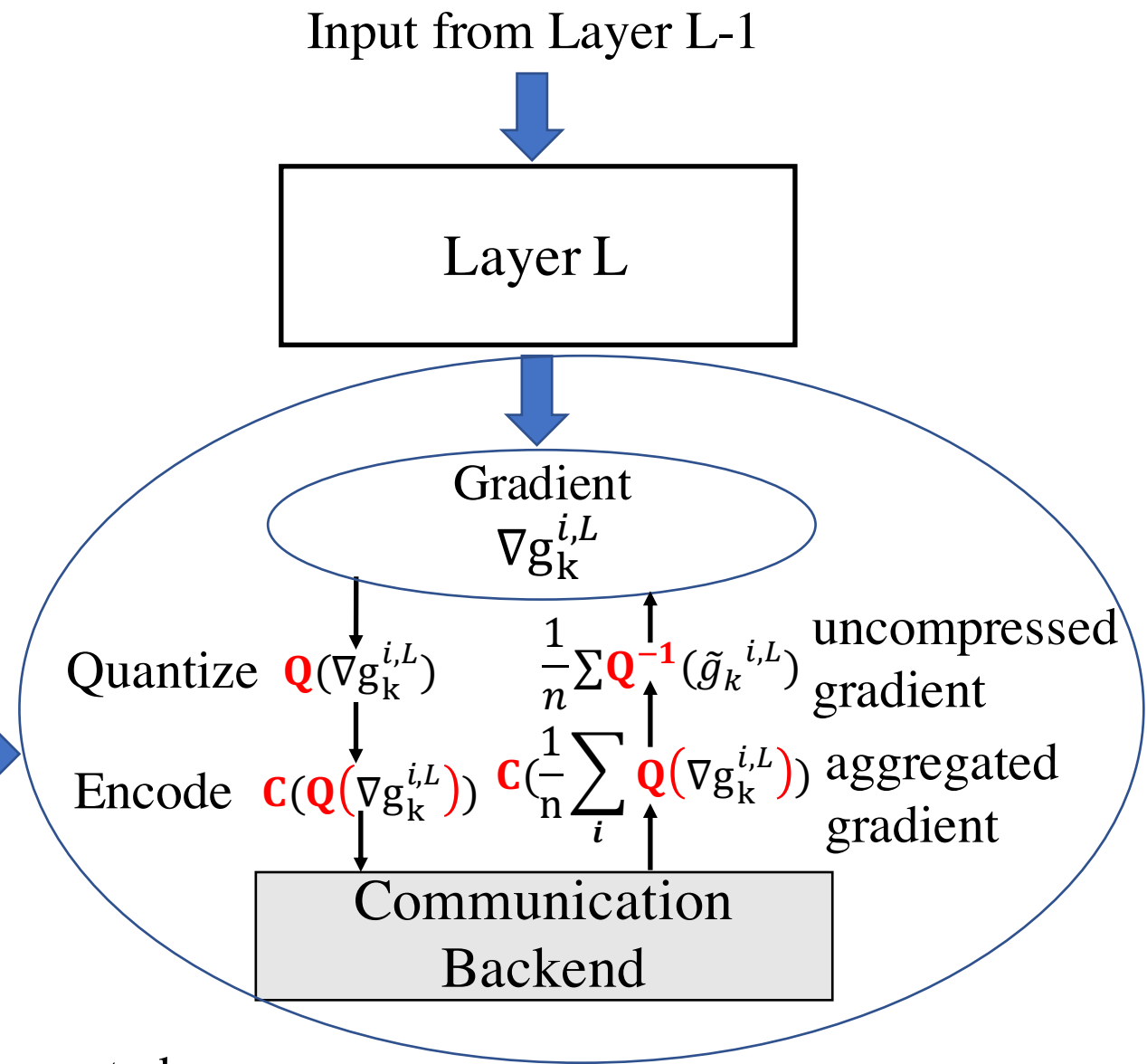
Communication Backend

Each layer's stochastic gradient $\nabla g_k^{i,j}$ is communicated and aggregated
$\frac{1}{n}\sum_i \nabla g_k^{i,j}$ is used to update the parameters of each layer

(a)

Input from Layer L-1

Layer L

Gradient
$\nabla g_k^{i,L}$

Quantize $\mathbf{Q}(\nabla g_k^{i,L})$ $\frac{1}{n}\sum \mathbf{Q^{-1}}(\tilde{g}_k^{i,L})$ uncompressed gradient

Encode $\mathbf{C}(\mathbf{Q}(\nabla g_k^{i,L}))$ $\mathbf{C}(\frac{1}{n}\sum_i \mathbf{Q}(\nabla g_k^{i,L}))$ aggregated gradient

Communication Backend

(b)