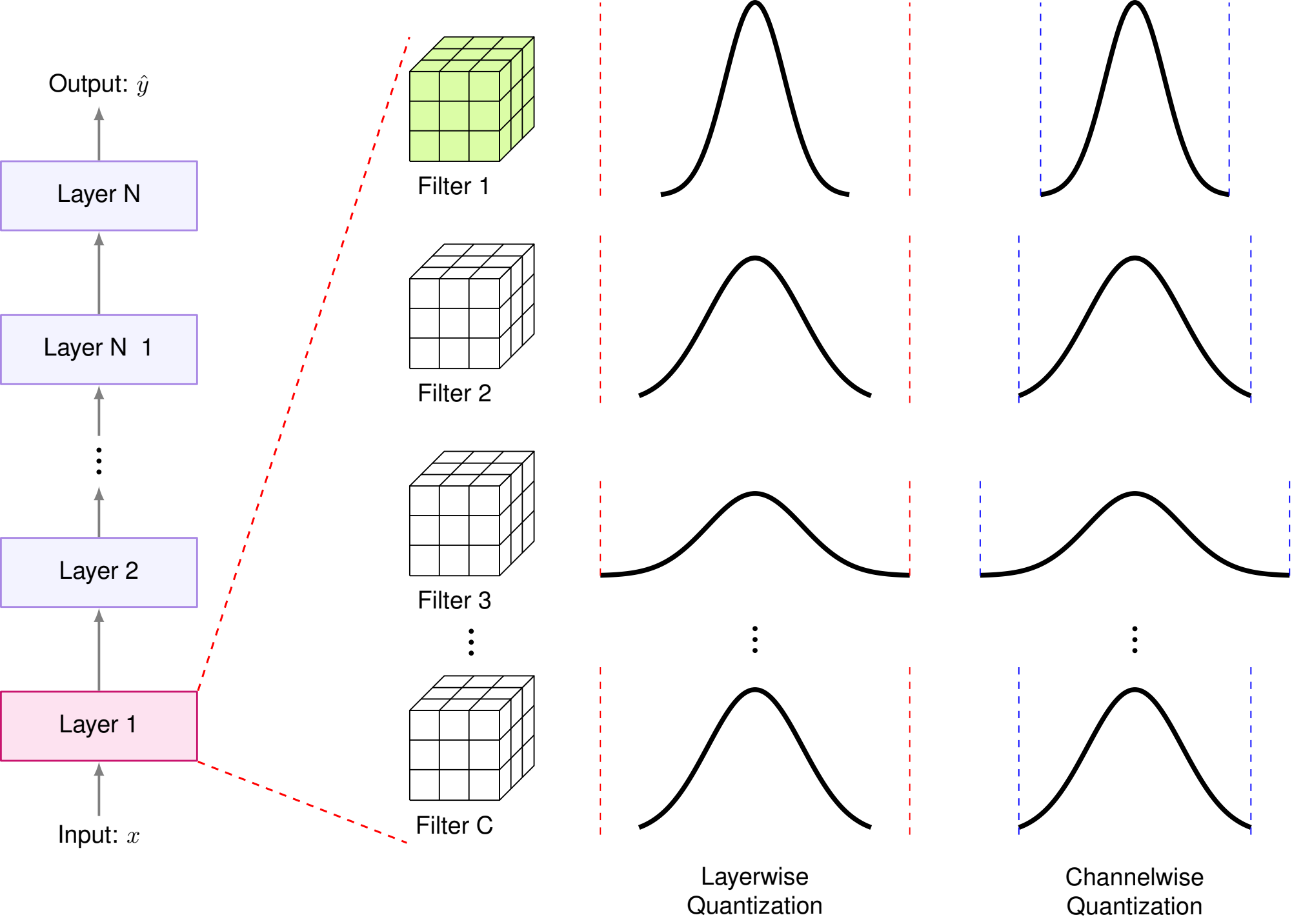
Output: $\hat{y}$
Layer N
Layer N 1
Layer 2
Layer 1
Input: x
Filter 1
Filter 2
Filter 3
Filter C
Layerwise Quantization
Channelwise Quantization