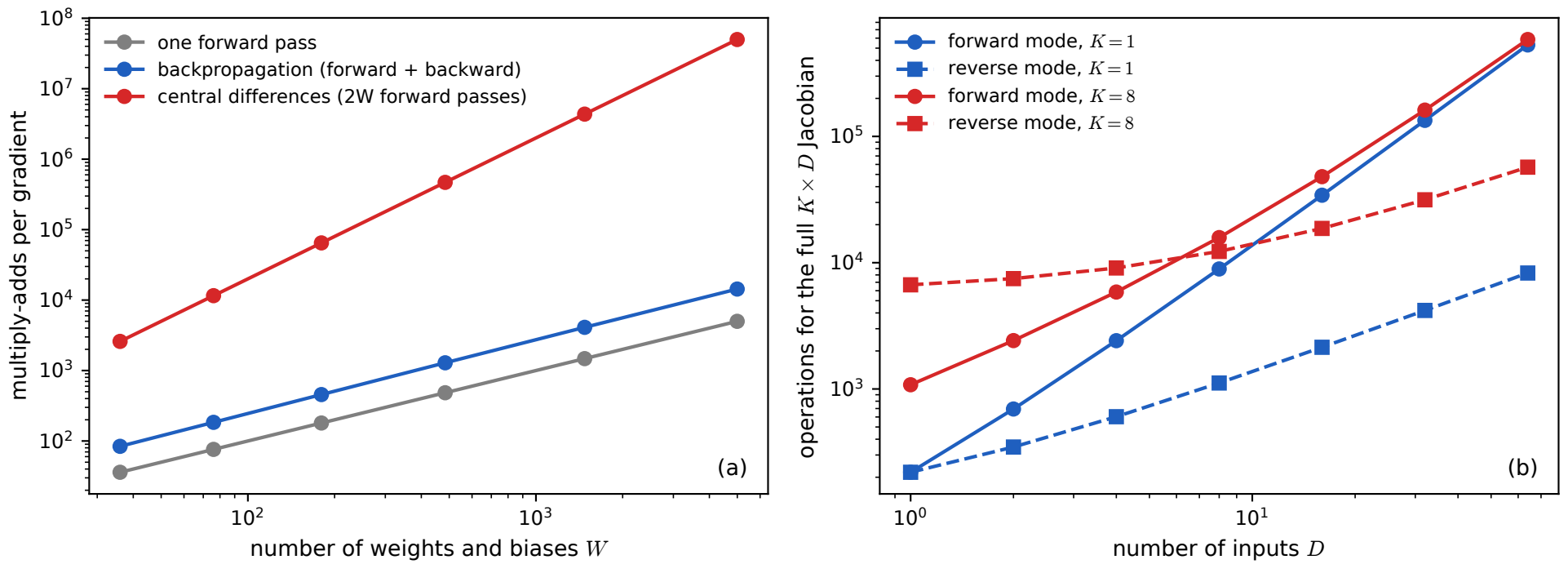

one forward pass
backpropagation (forward + backward)
central differences (2W forward passes)
multiply-adds per gradient
number of weights and biases W
(a)
forward mode, $K=1$
reverse mode, $K=1$
forward mode, $K=8$
reverse mode, $K=8$
operations for the full $K \times D$ Jacobian
number of inputs D
(b)