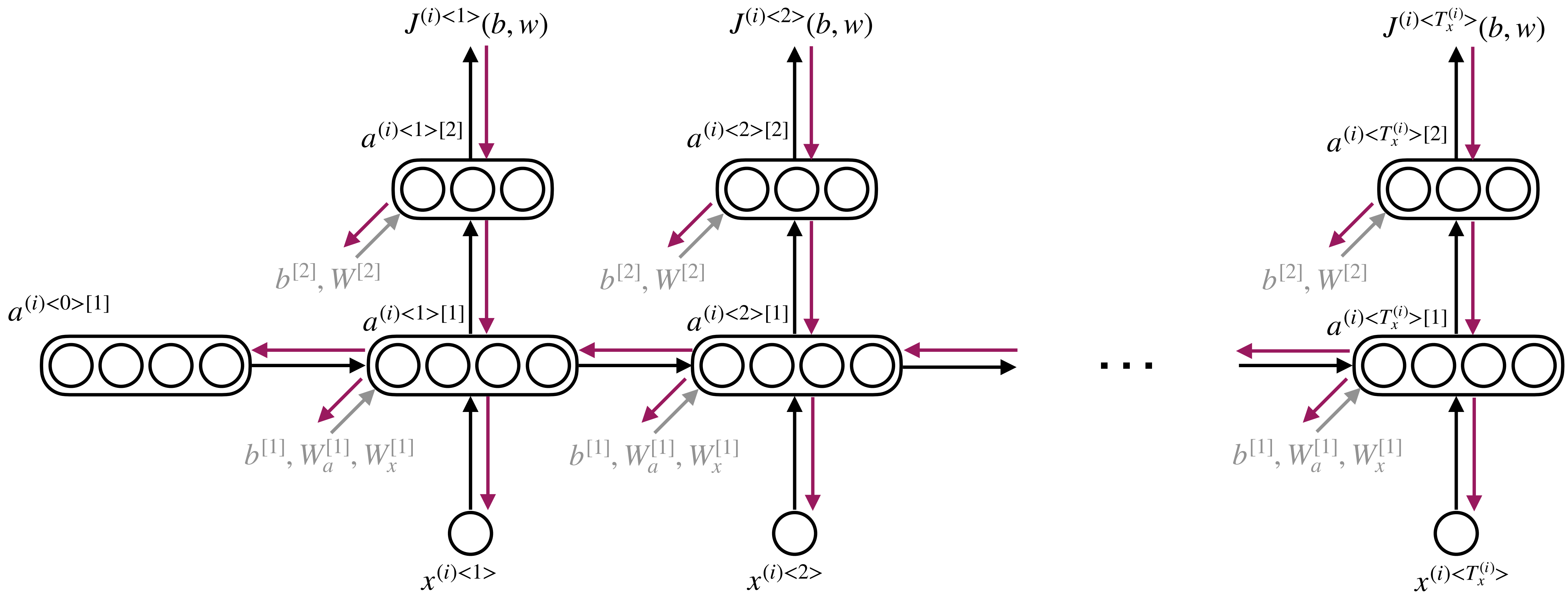

Forward Propagation: Start with $x^{(i)<1>}$ and $a^{(i)<0>[1]}$, work forwards in time

$$a^{(i)<t>[1]} = g^{[1]}(z^{(i)<t>[1]})$$

$$z^{(i)<t>[1]} = b^{[1]} + W_a^{[1]} a^{(i)<t-1>[1]} + W_x^{[1]} x^{(i)<t-1>}$$

Backward Propagation: Start with $J^{(i)<T_x^{(i)}>}(b,w)$, work backwards in time

For time t, $J^{(i)}(b,w) = \sum_{t=1}^{T_x^{(i)}} J^{(i)<t>}(b,w)$ depends on $a^{(i)<t>[1]}$ through $a^{(i)<t>[2]}$ (same time, next layer up) and $a^{(i)<t+1>[1]}$ (next time, same layer)

Therefore, $$\frac{\partial J^{(i)}}{\partial a^{(i)<t>[1]}} = \frac{\partial J^{(i)}}{\partial a^{(i)<t>[2]}} \frac{\partial a^{(i)<t>[2]}}{\partial a^{(i)<t>[1]}} + \frac{\partial J^{(i)}}{\partial a^{(i)<t+1>[1]}} \frac{\partial a^{(i)<t+1>[1]}}{\partial a^{(i)<t>[1]}}$$

$J^{(i)}(b,w)$ depends on $W_x^{[1]}$ through $a^{(i)<1>[1]}, \ldots, a^{(i)<T_x^{(i)}>[1]}$

Therefore, $$\frac{\partial J^{(i)}}{\partial W_x^{[1]}} = \sum_{t=1}^{T_x^{(1)}} \frac{\partial J^{(i)}}{\partial a^{(i)<t>[1]}} \frac{\partial a^{(i)<t>[1]}}{\partial W_x^{[1]}}$$

(a similar idea holds for $b^{[1]}$ and $W_a^{[1]}$)