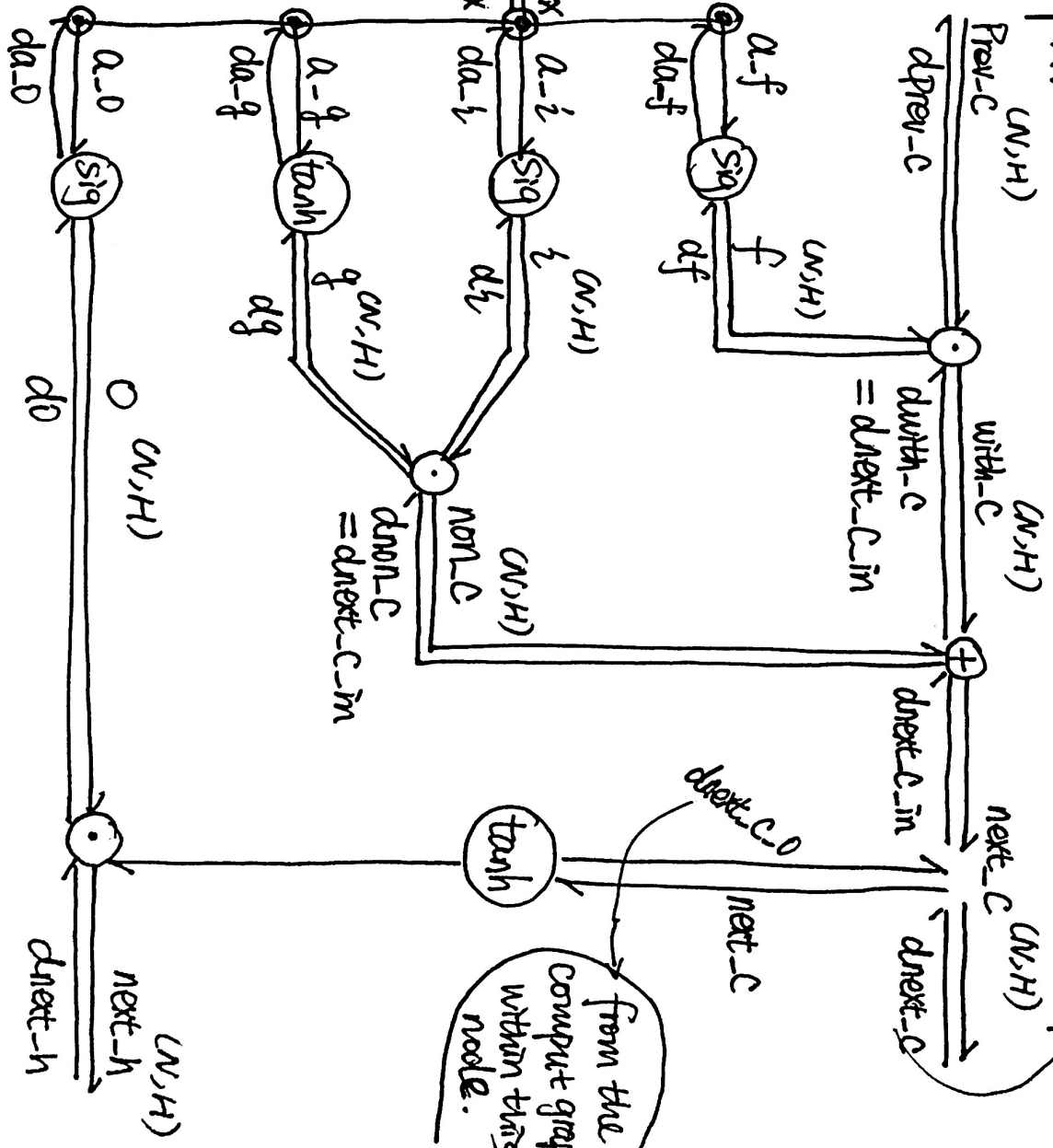
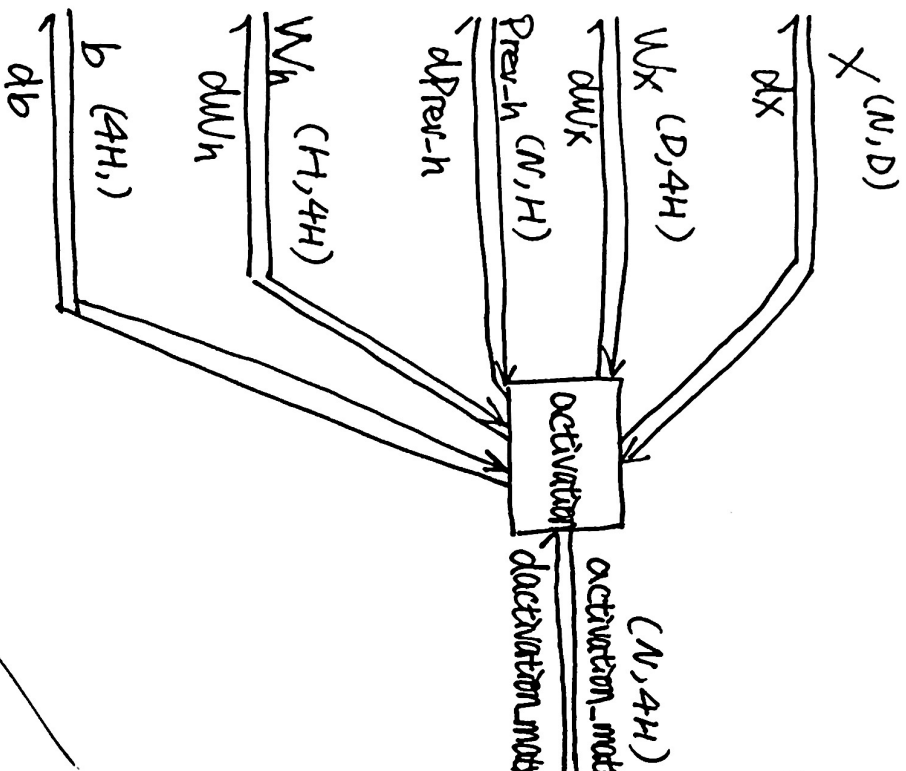


# Long Short Term Memory (LSTM)

## Single node computational graph.

activation-matrix  
 $= \text{Prev-h} \cdot \text{dot}(W_h) + x \cdot \text{dot}(W_k) + b$



direct-c: from next node of this RNN network, take as input for this backprop process

from the compute graph within this node.

here: reorganize  $da-f, da-i, da-g, da-o$  back into activation-matrix,  $(N, 4H)$

direct-c-in = direct-c + direct-c-o  
 For direct-c, its derivatives are from two parts.