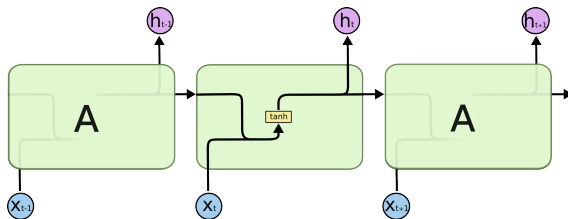


Vanilla RNN



$$h_t = \tanh(W[h_{t-1}; x_t] + b) \quad (3)$$

$[h_{t-1}; x_t]$ is concatenation of h_{t-1} and x_t

- Vanishing gradient problem.
- Non-linear transformation always applied.
 \Rightarrow Type theory: h_t and h_{t-1} live in different vector spaces.