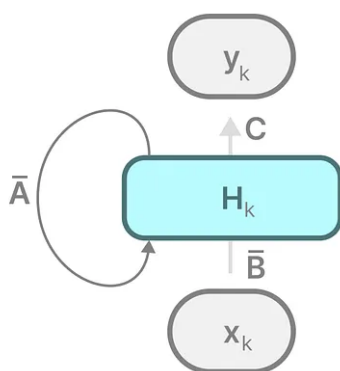
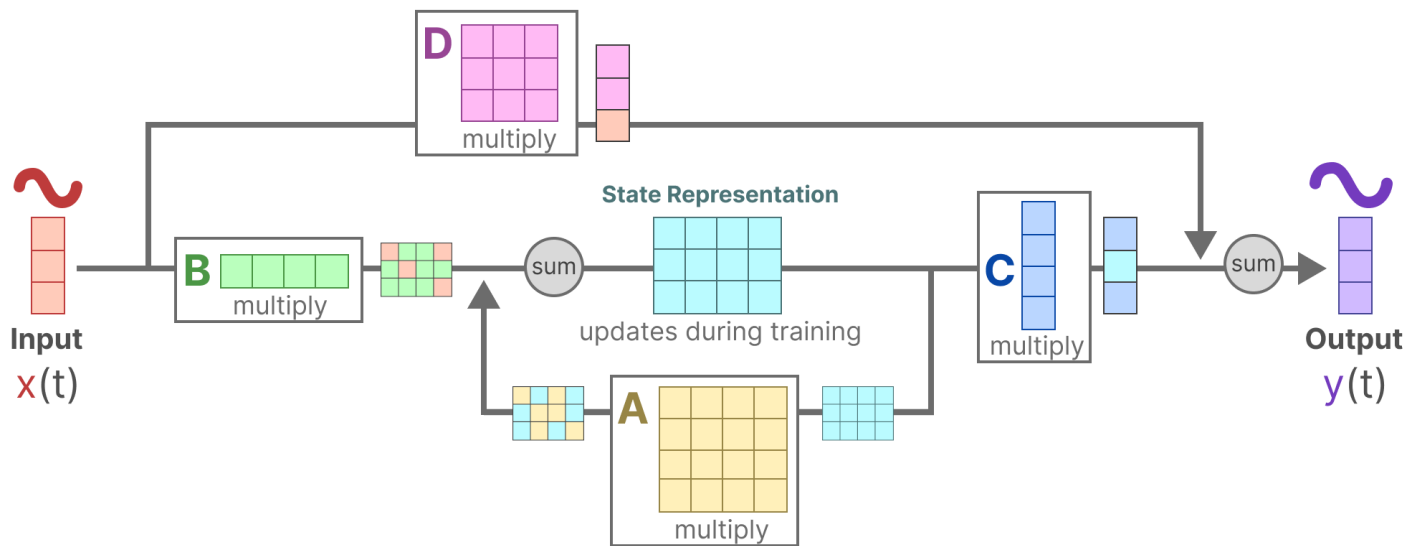


Mamba

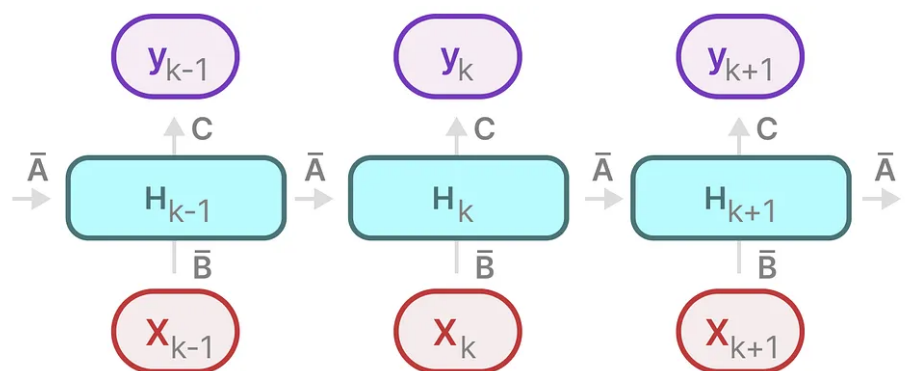
SSM

$$h'(t) = Ah(t) + Bx(t)$$

$$y(t) = Ch(t)$$



SSM
(Recurrent)



SSM
(Recurrent + Unfolded)

Mamba

1. SSM/S4



2. SSM/S4 attention /

3. Transformer



4.

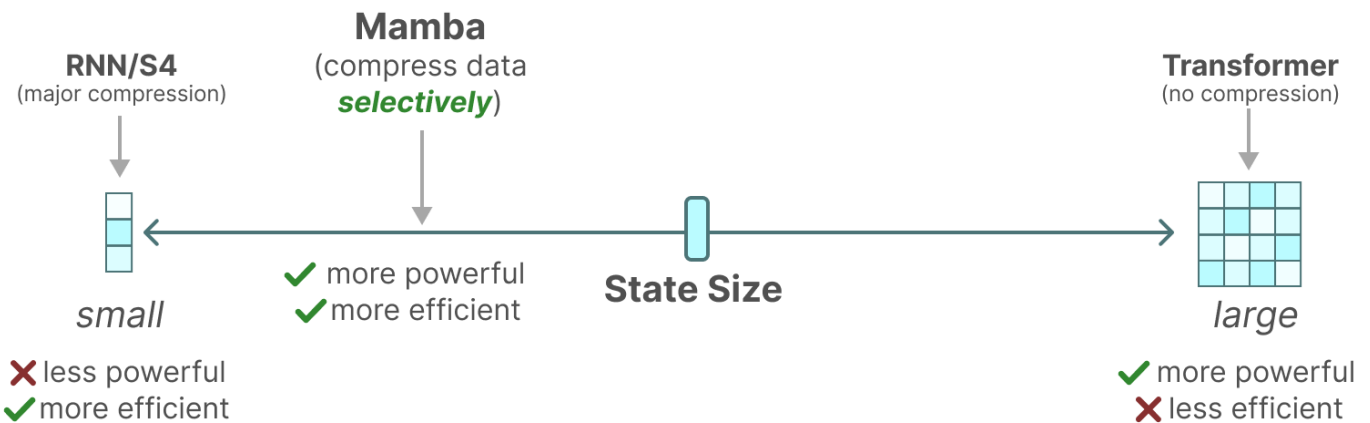
5. Manba Transformer

Manba

Transformer

KV cache

6. Manba



Matrix A

How the **current state** evolves over time

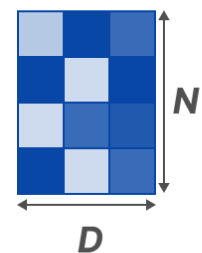
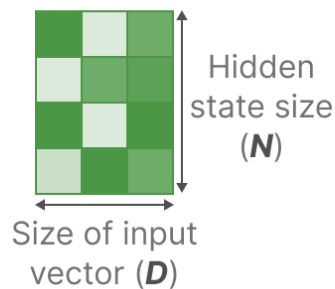
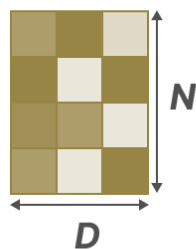
Matrix B

How the **input** influences the state

Matrix C

How the **current state** translates to the **output**

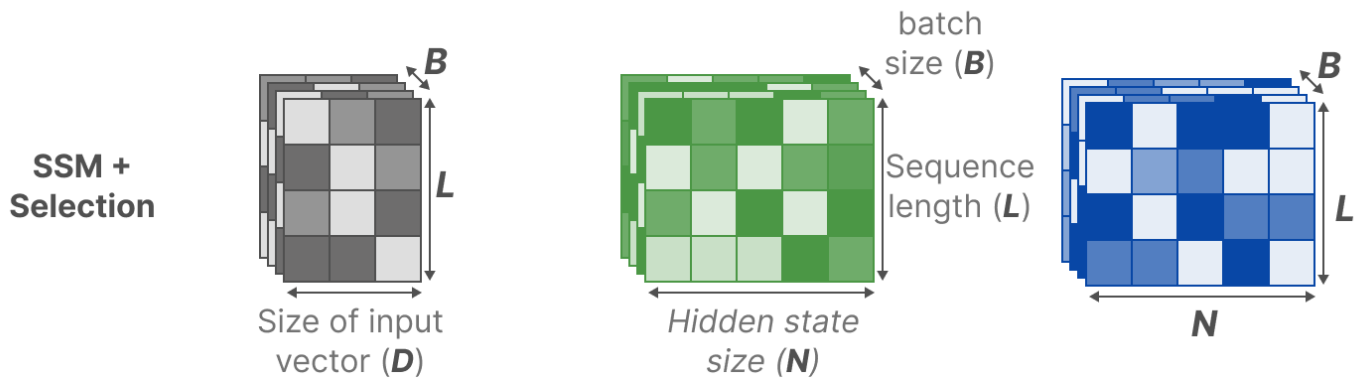
Structured State Space Model (S4)



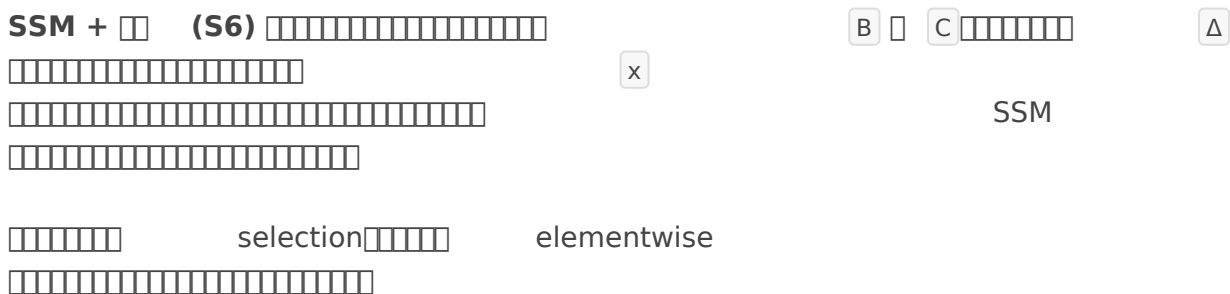
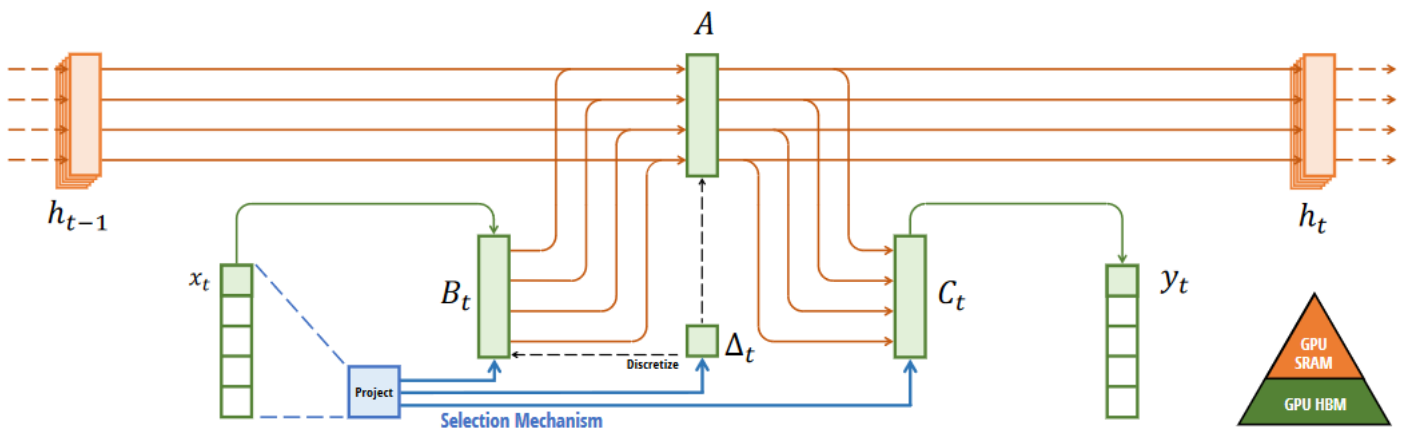
Step size (Δ)
Resolution of the **input**
(discretization parameter)

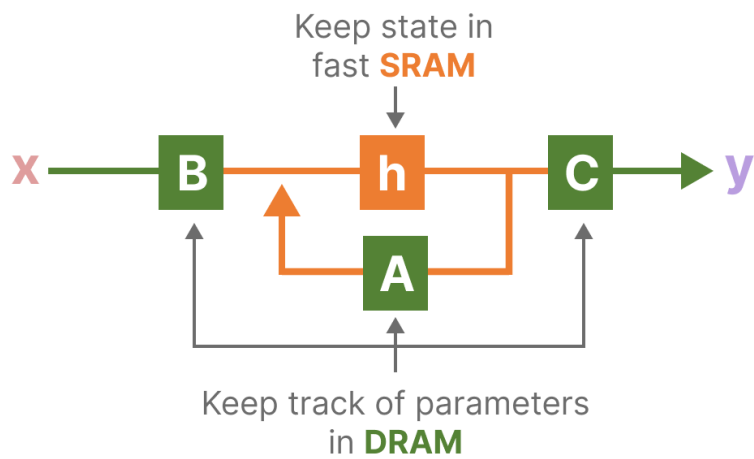
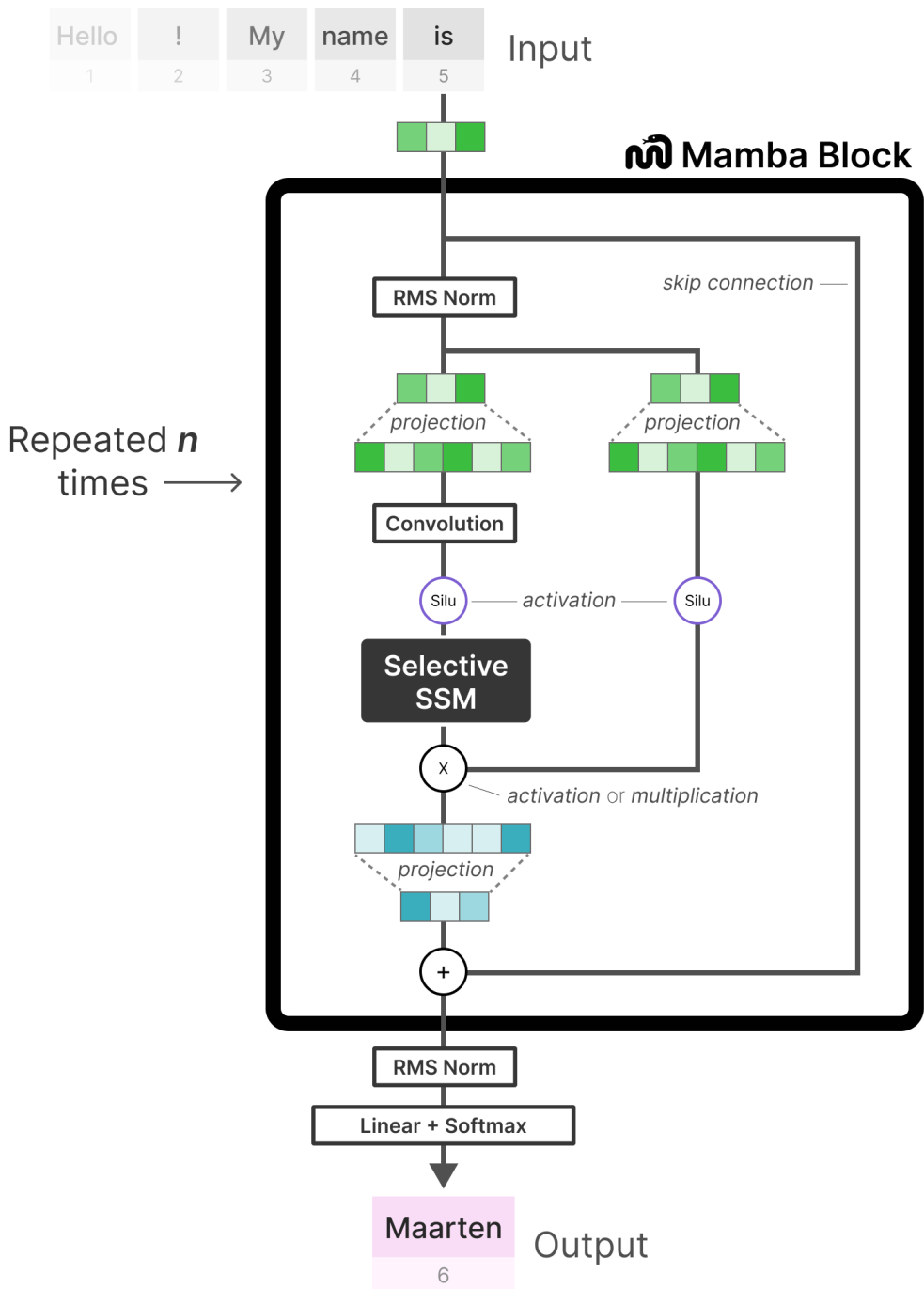
Matrix B
How the **input**
influences the state

Matrix C
How the **current state**
translates to the **output**



Selective State Space Model with Hardware-aware State Expansion





<https://newsletter.maartengrootendorst.com/p/a-visual-guide-to-mamba-and-state#%C2%A7what-is-a-state-space>

Vision Mamba

Revision #1

Created 11 January 2025 09:44:05 by Colin

Updated 11 January 2025 09:44:08 by Colin