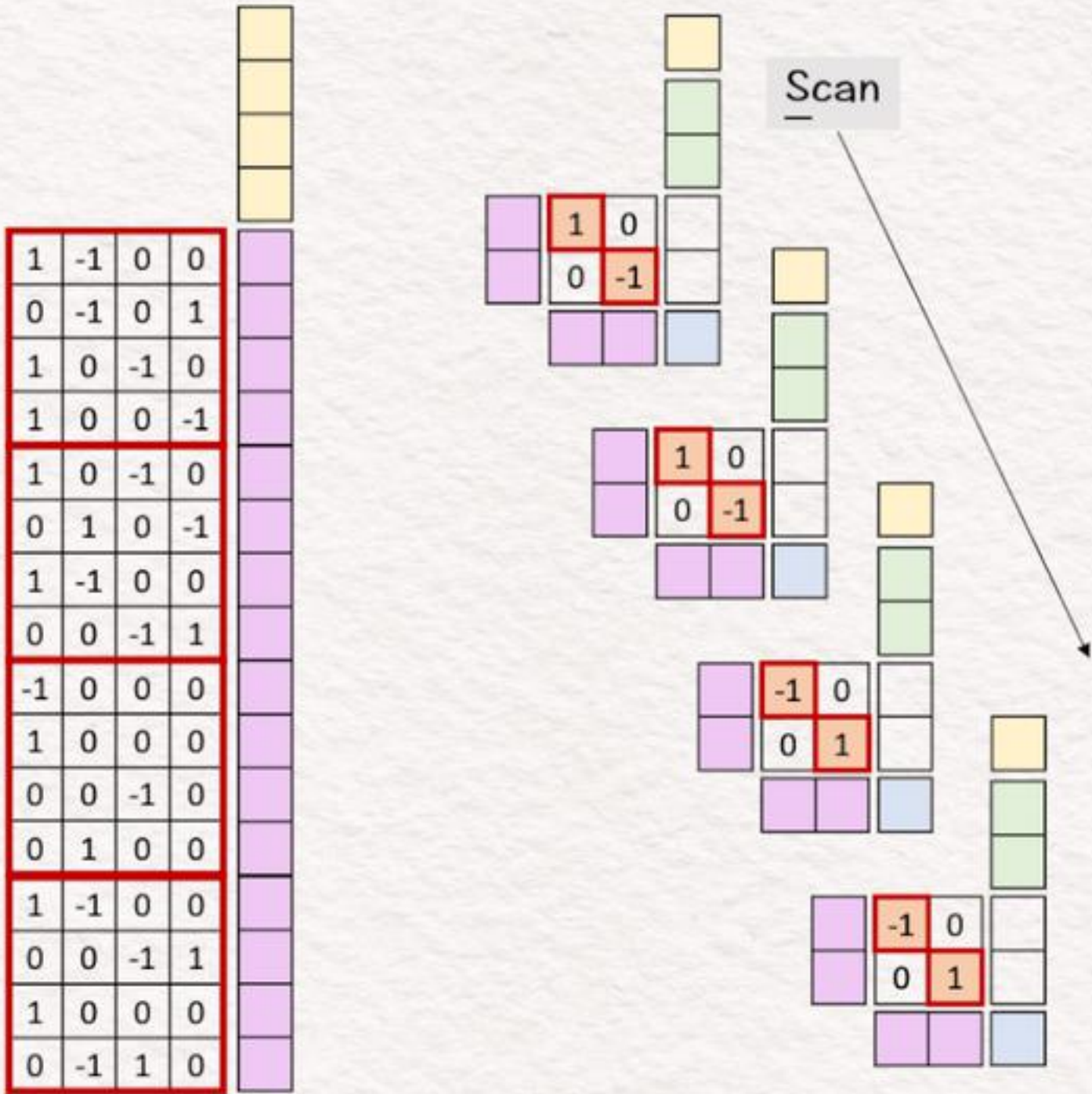
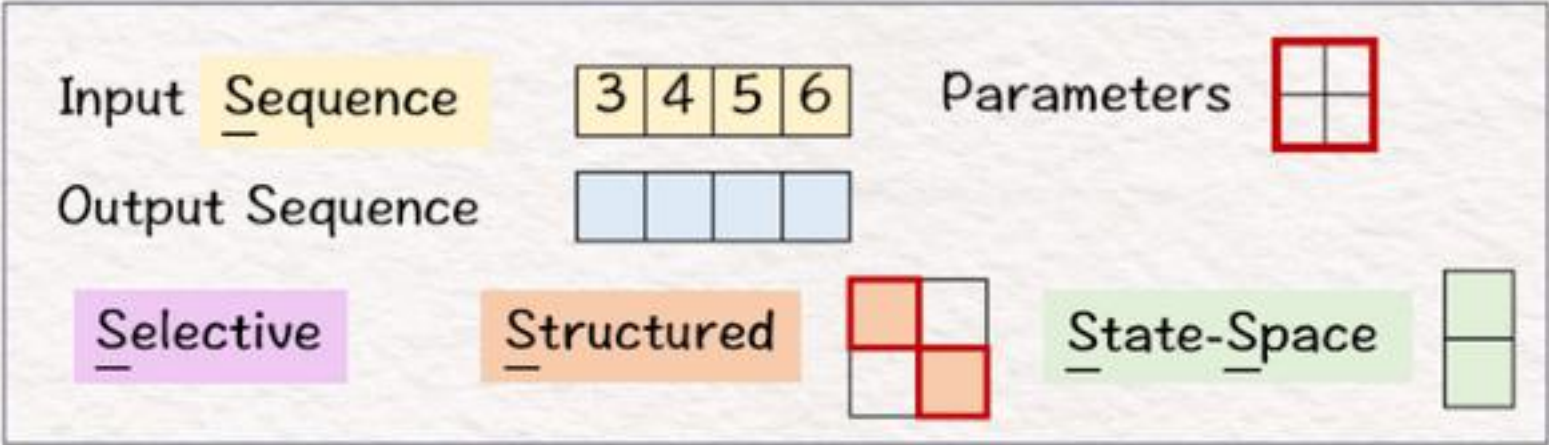


Mamba's S6 Model



© 2023 Tom Yeh

[1] 左侧：

输入序列中的所有四个token

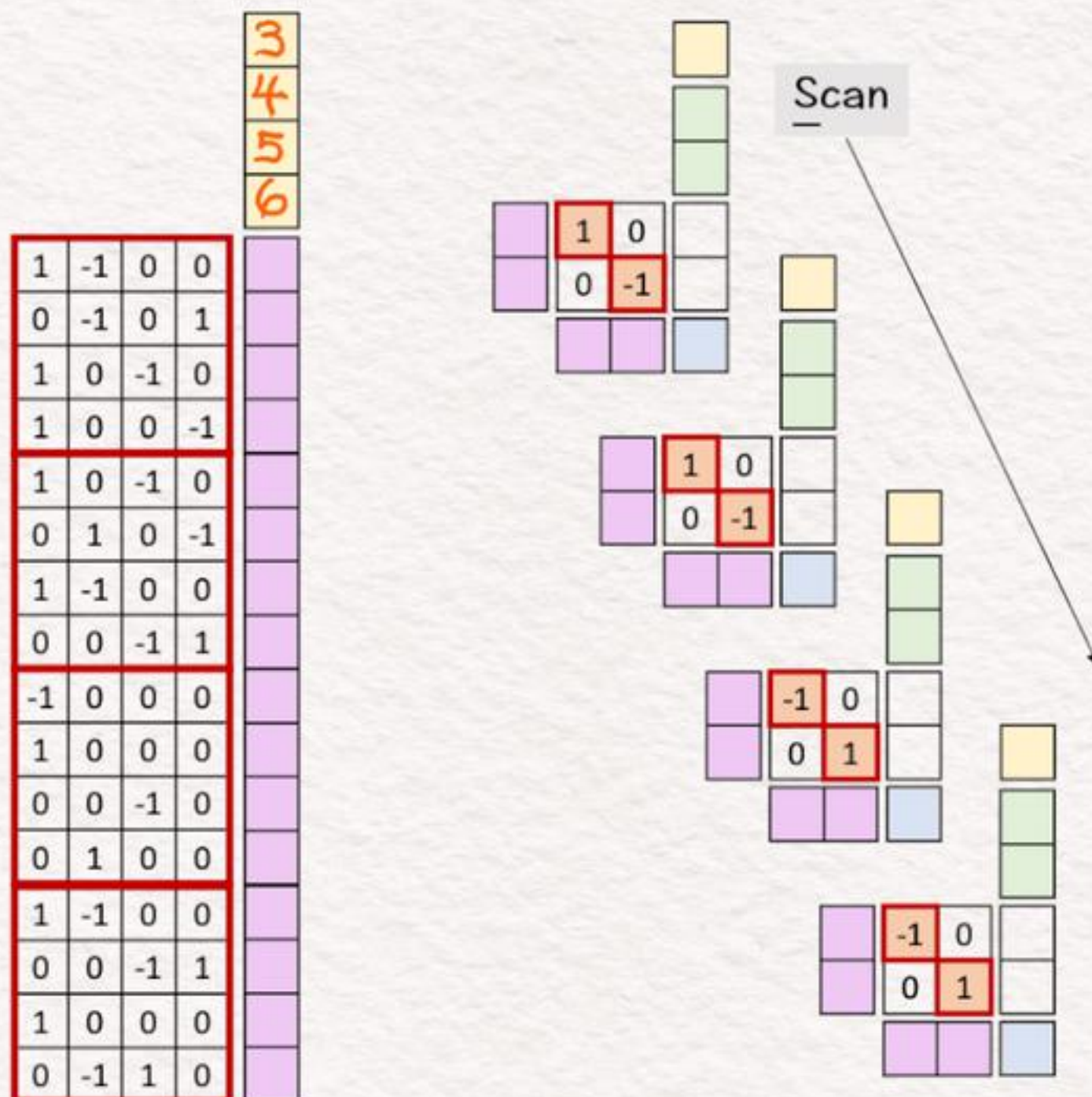
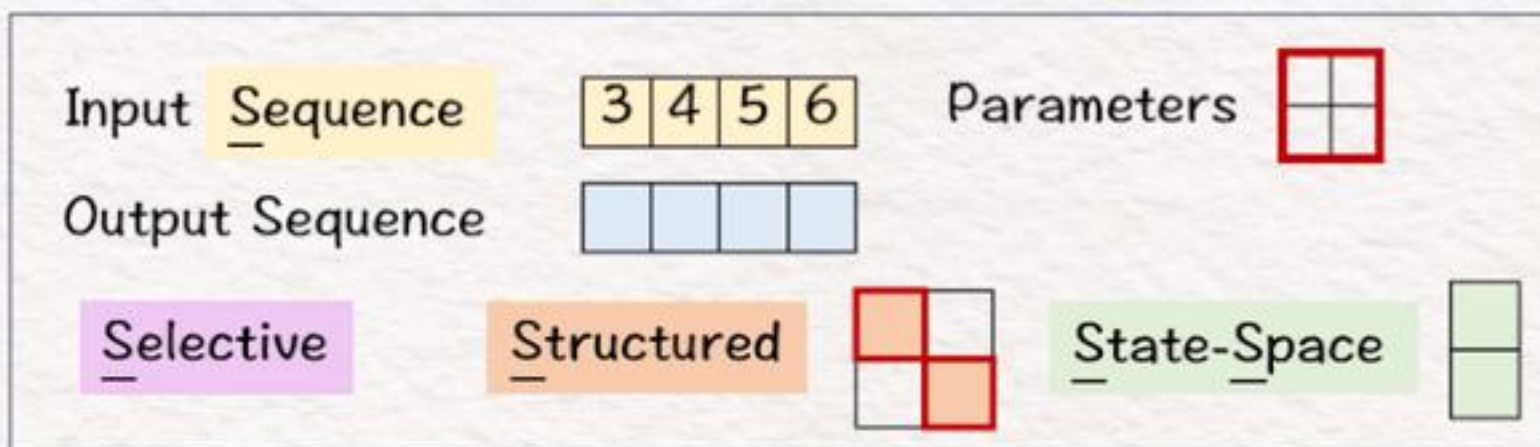
都通过一个线性层处理，以计算一组权重。

[1] Left:

All four tokens in the input sequence

are processed by a linear layer to calculate a set of weights.

Mamba's S6 Model



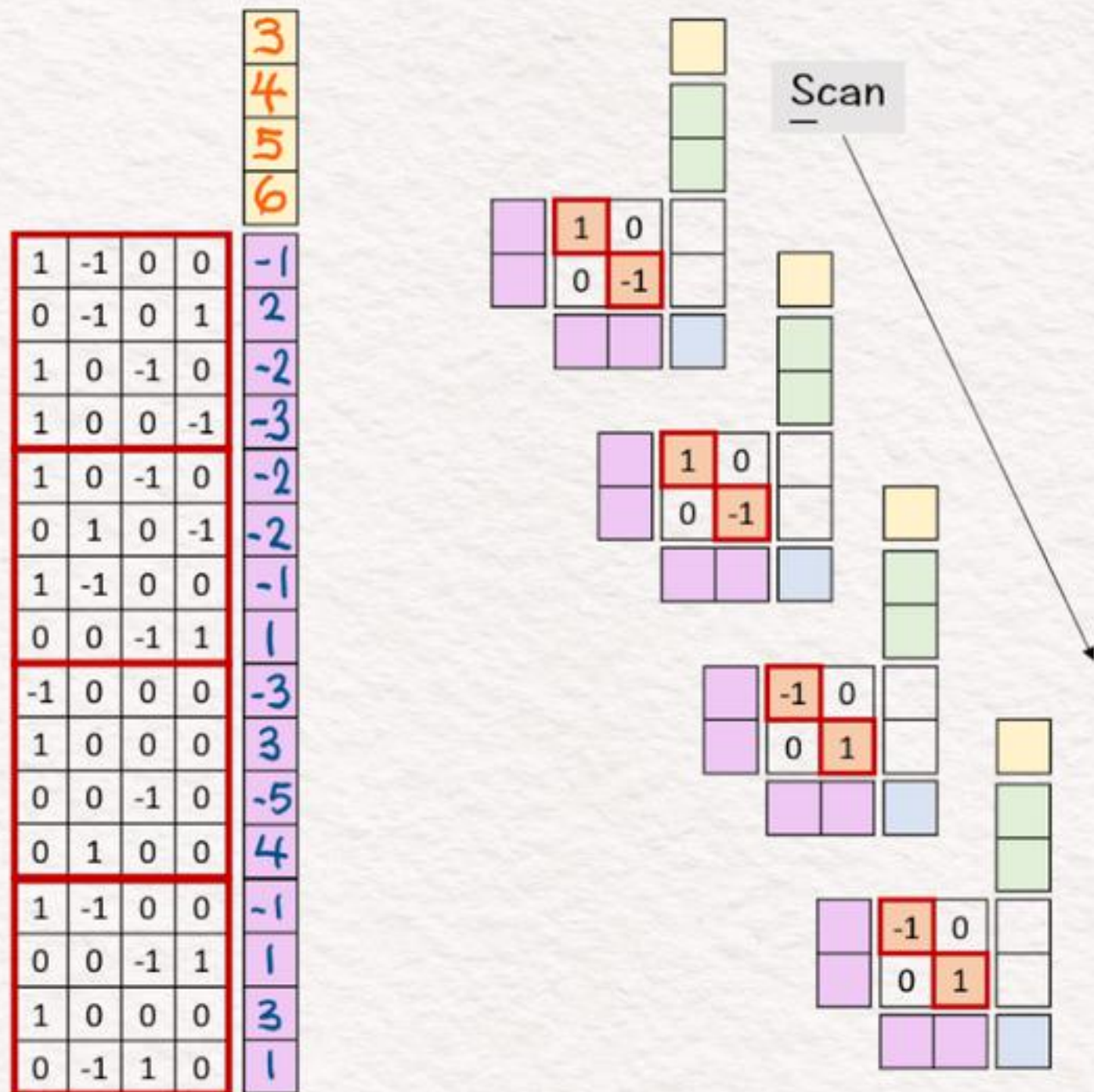
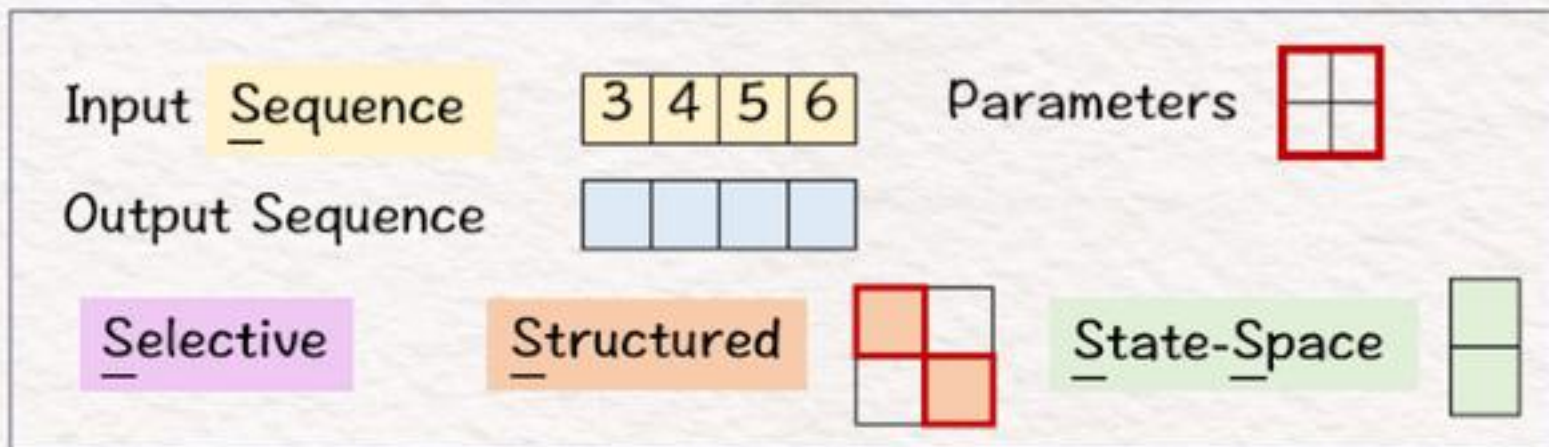
© 2023 Tom Yeh

[2] 右侧：

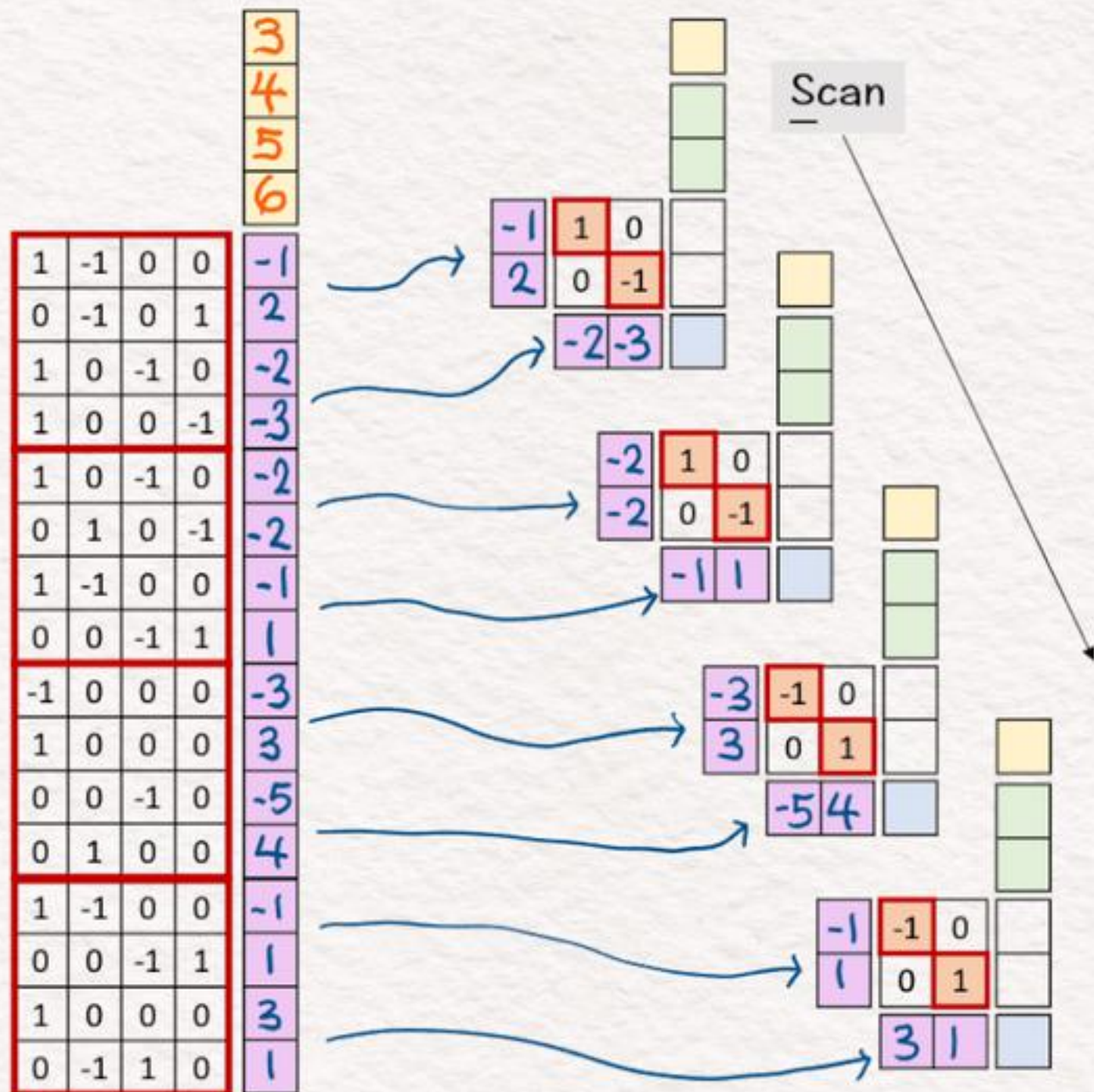
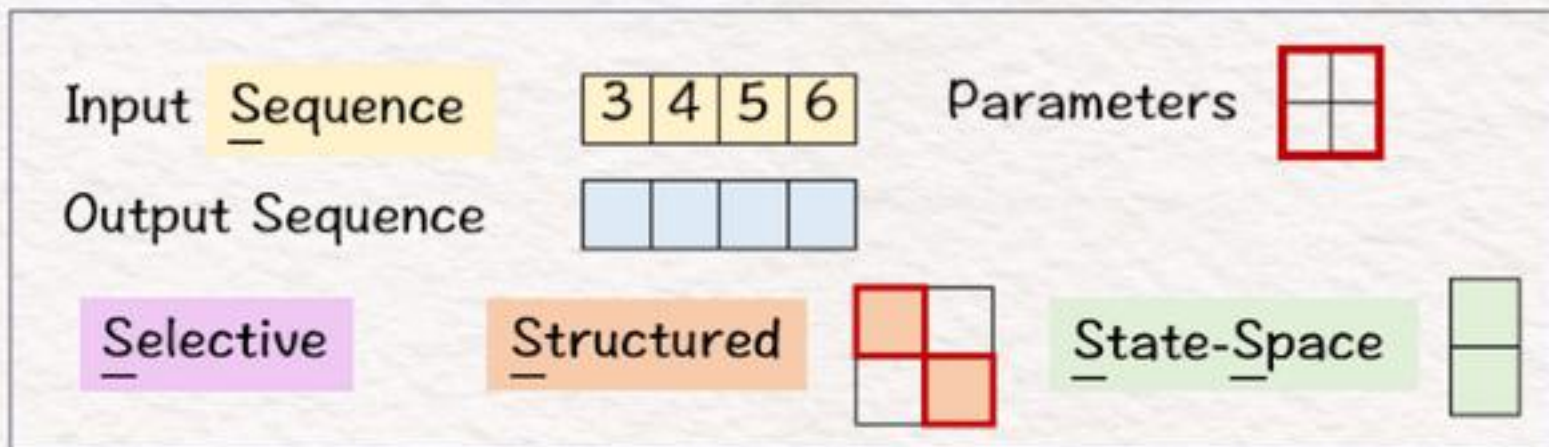
这些权重用于驱动一个类RNN网络。

[2] Right: These weights are used to drive an RNN-like network.

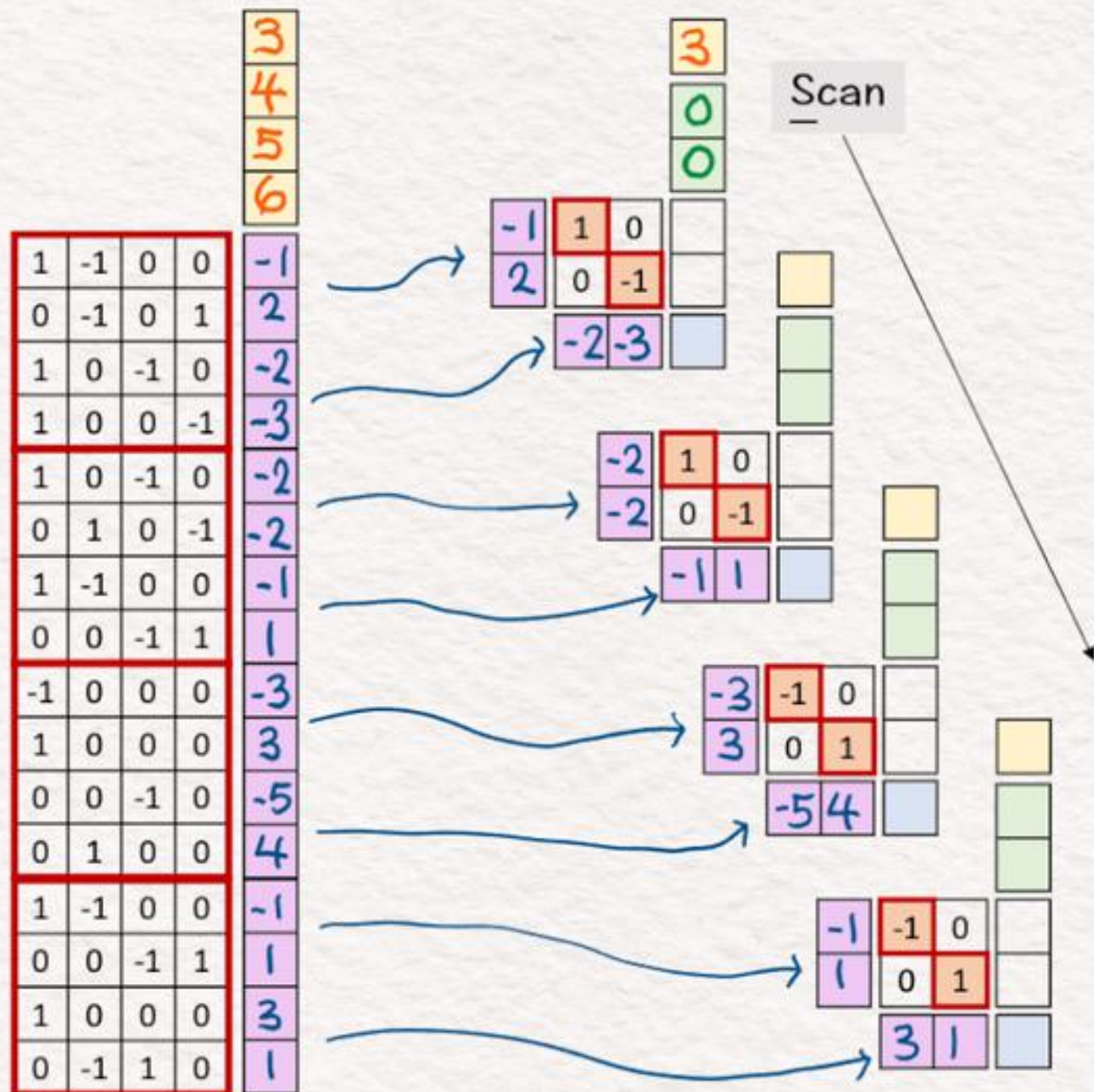
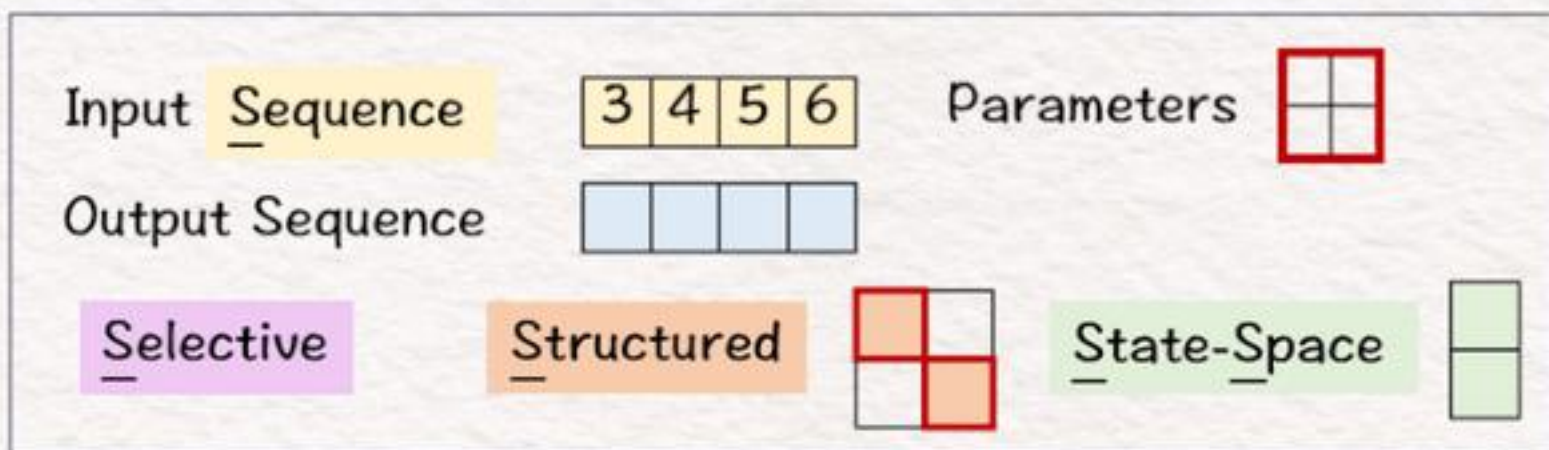
Mamba's S6 Model



Mamba's S6 Model



Mamba's S6 Model



© 2023 Tom Yeh

[3] 第一个输入[3]和隐藏状态[0, 0]

使用权重 $B = [-1; 2]$ 和 $A = [1, 0; 0, -1]$ 线性组合，

以计算新的隐藏状态 $\rightarrow [-3, 6]$ 。

注意，没有涉及非线性激活函数。

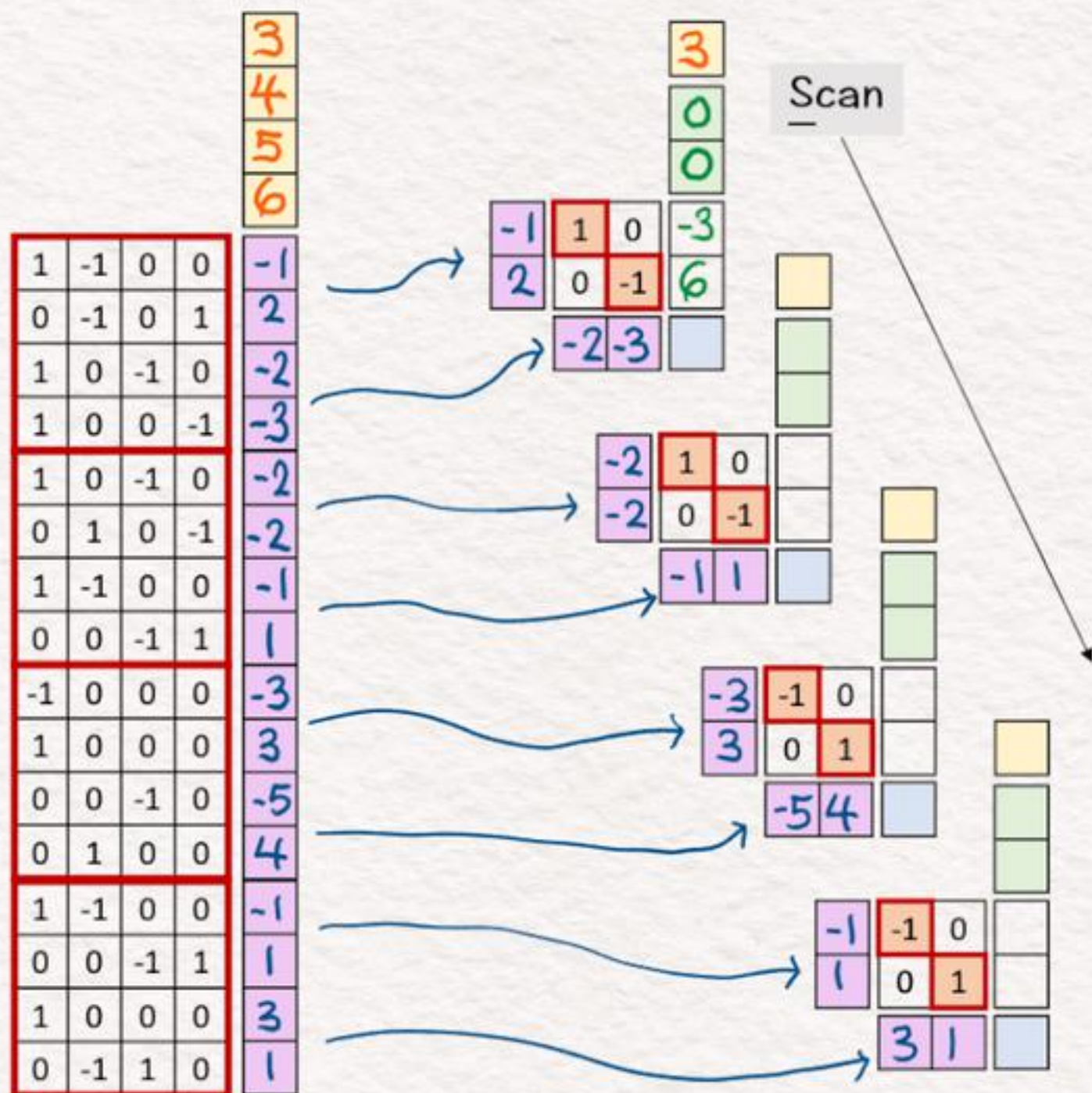
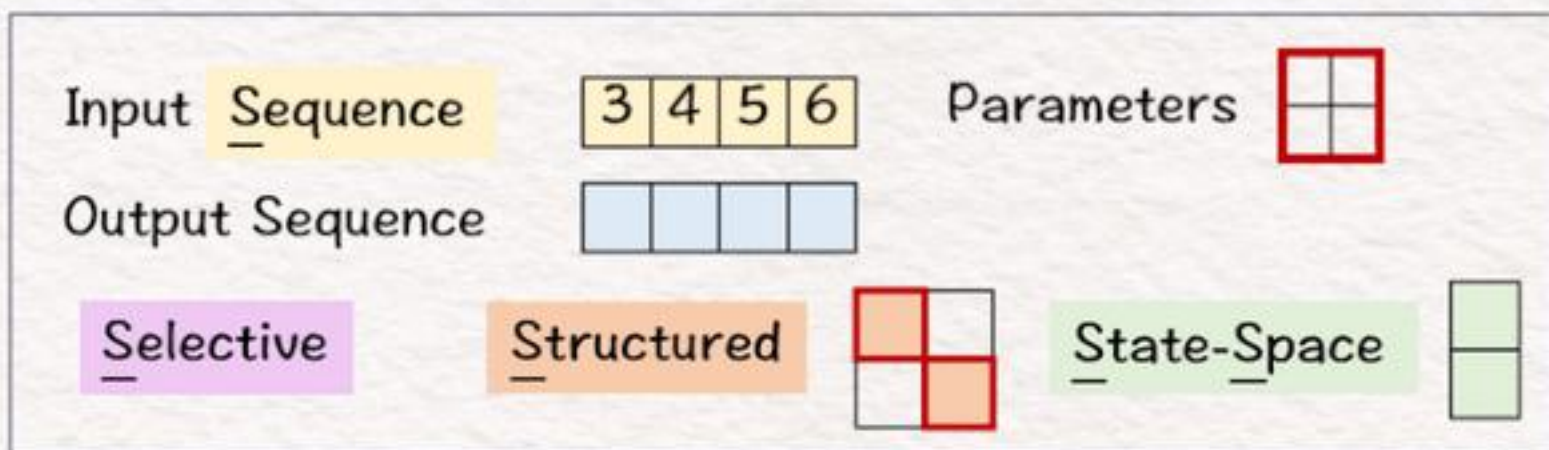
[3] The first input [3] and hidden states [0, 0]

are linearly combined using weights $B = [-1; 2]$ and

$A = [1, 0; 0, -1]$ to calculate new hidden states $\rightarrow [-3, 6]$.

Note that NO non-linear activation function is involved.

Mamba's S6 Model



© 2023 Tom Yeh

[3] 第一个输入[3]和隐藏状态[0, 0]

使用权重 $B=[-1; 2]$ 和 $A=[1, 0; 0, -1]$ 线性组合，

以计算新的隐藏状态 $\rightarrow [-3, 6]$ 。

注意，没有涉及非线性激活函数。

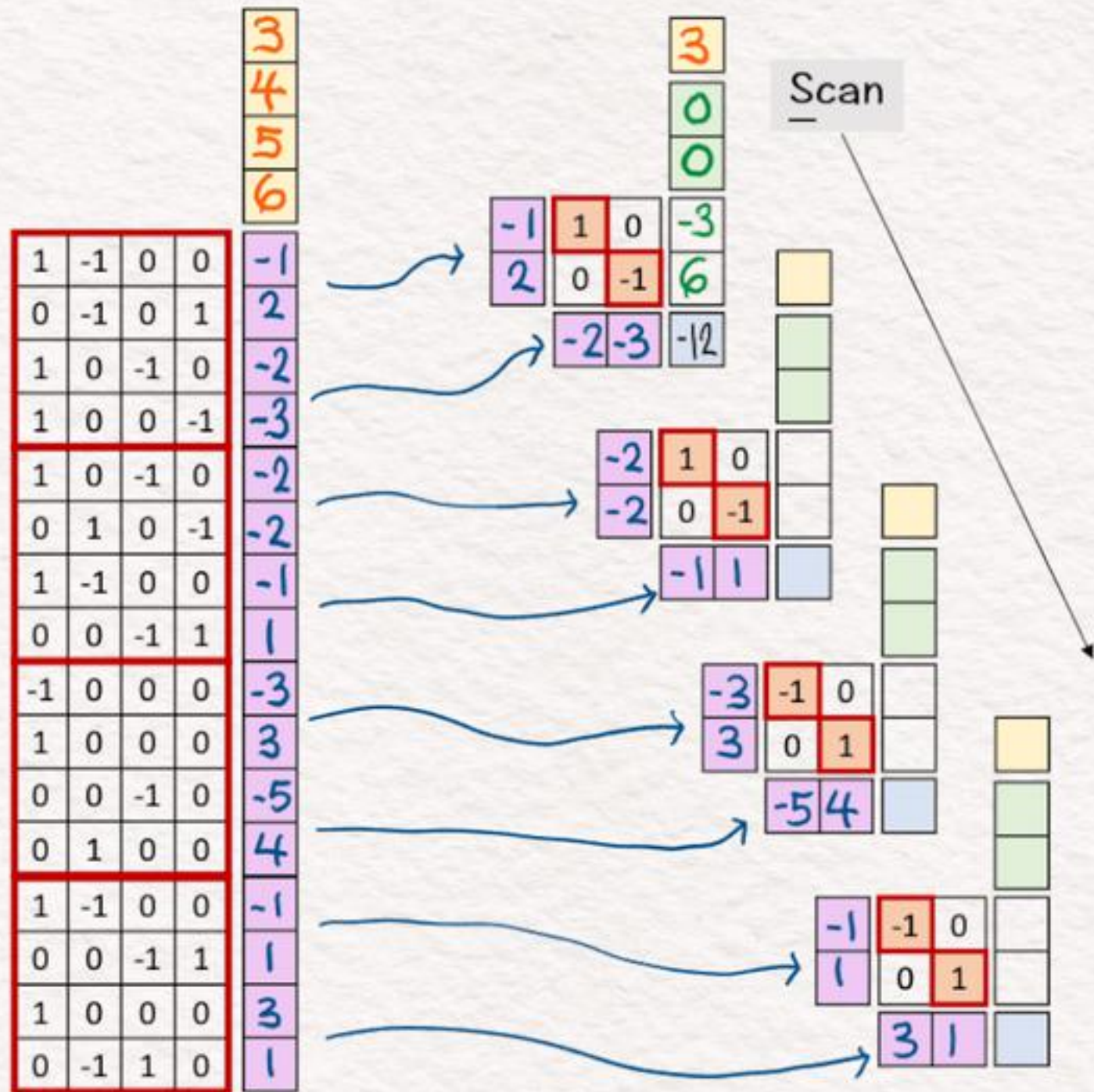
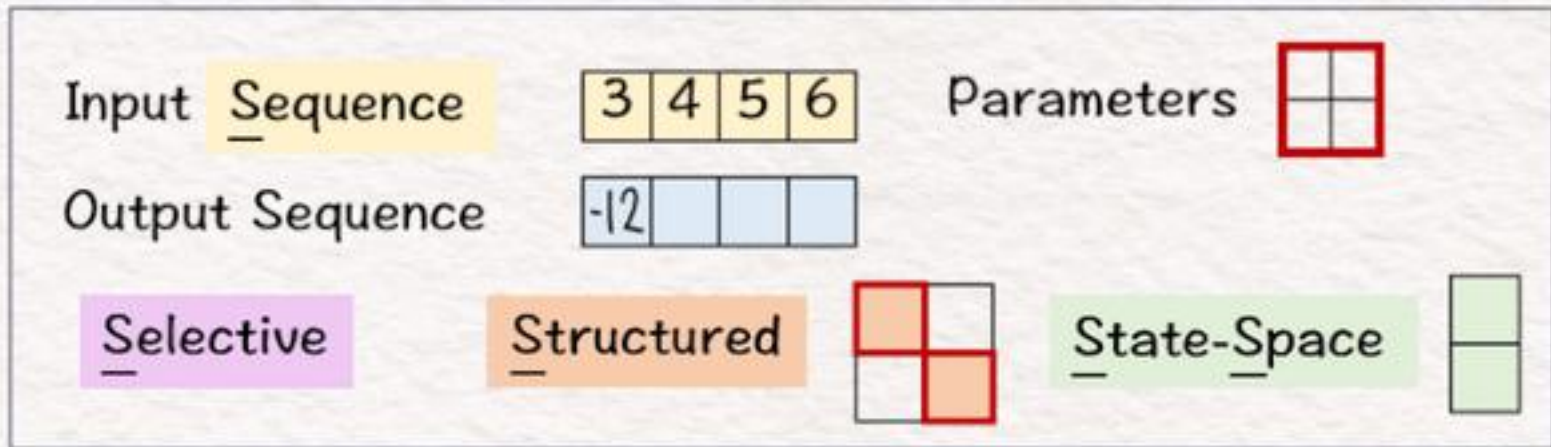
[3] The first input [3] and hidden states [0, 0]

are linearly combined using weights $B=[-1; 2]$ and

$A=[1, 0; 0, -1]$ to calculate new hidden states $\rightarrow [-3, 6]$.

Note that NO non-linear activation function is involved.

Mamba's S6 Model

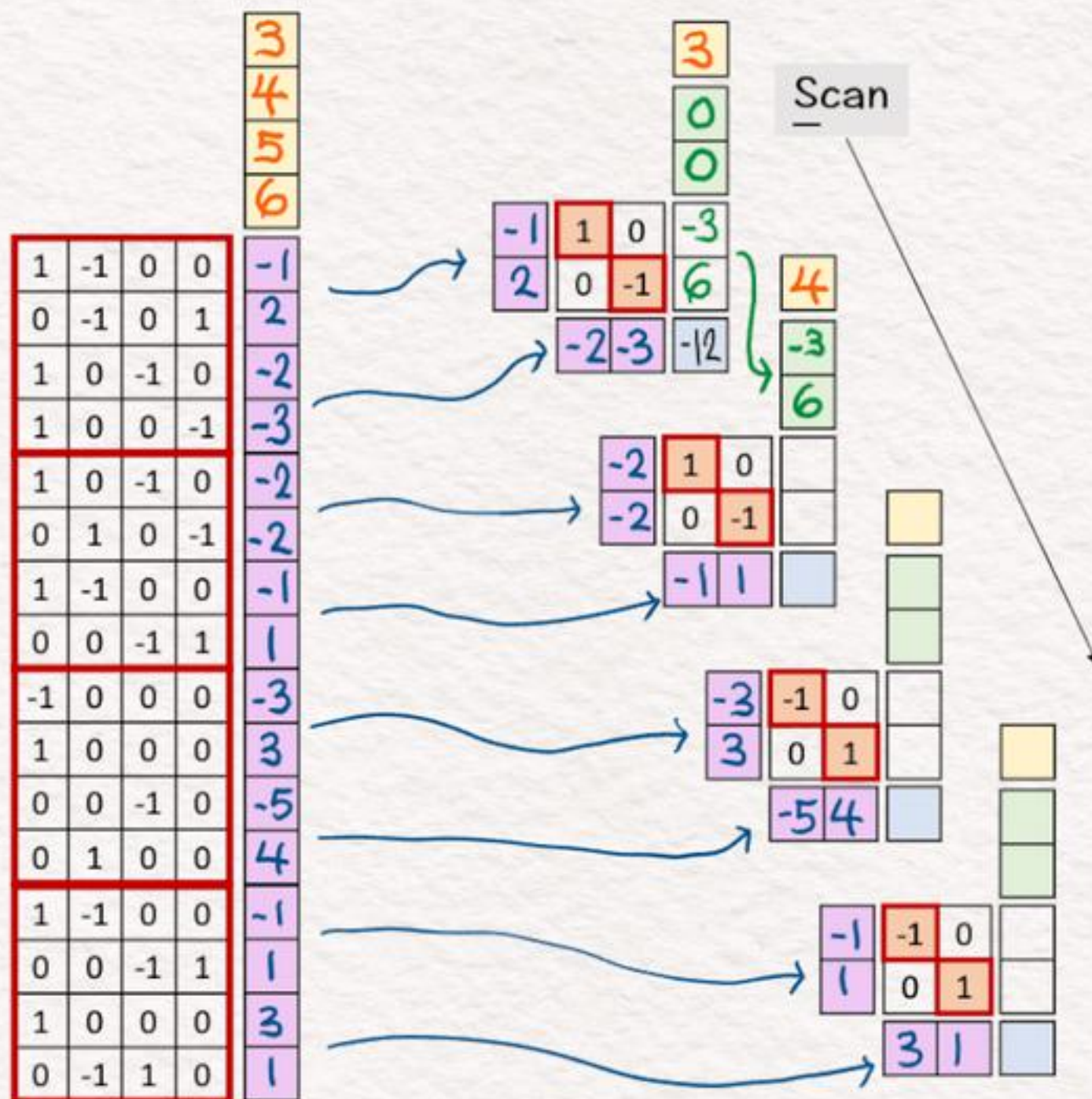
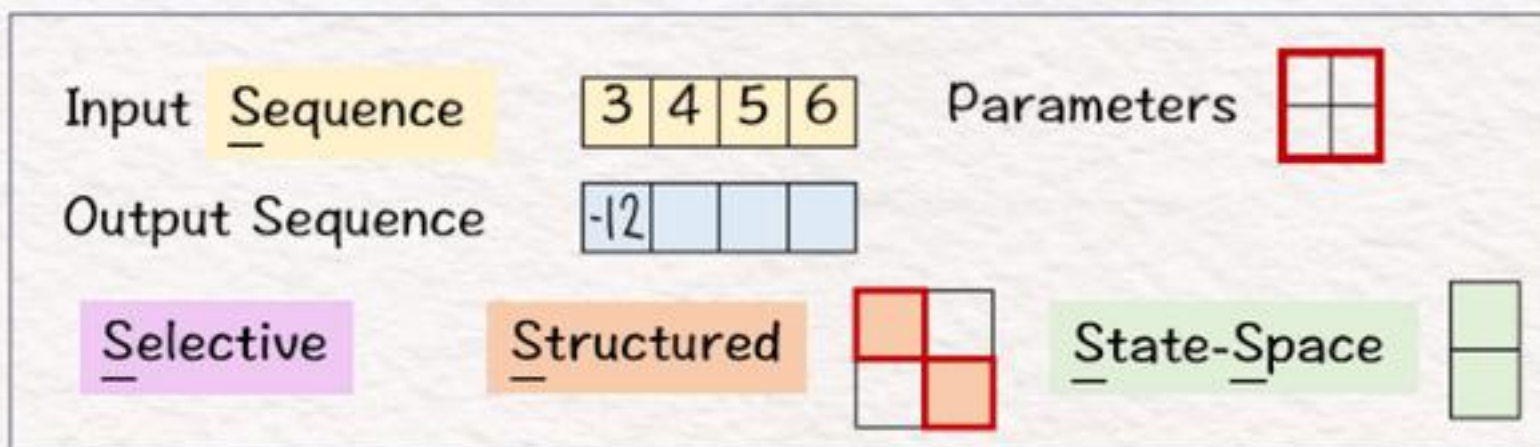


© 2023 Tom Yeh

[4] 隐藏状态[-3, 6]使用权重 $C = [-2, -3]$ 线性组合，以获得第一个输出[-12]。

[4] Hidden states [-3, 6] are linearly combined using weights $C = [-2, -3]$ to obtain the first output [-12].

Mamba's S6 Model

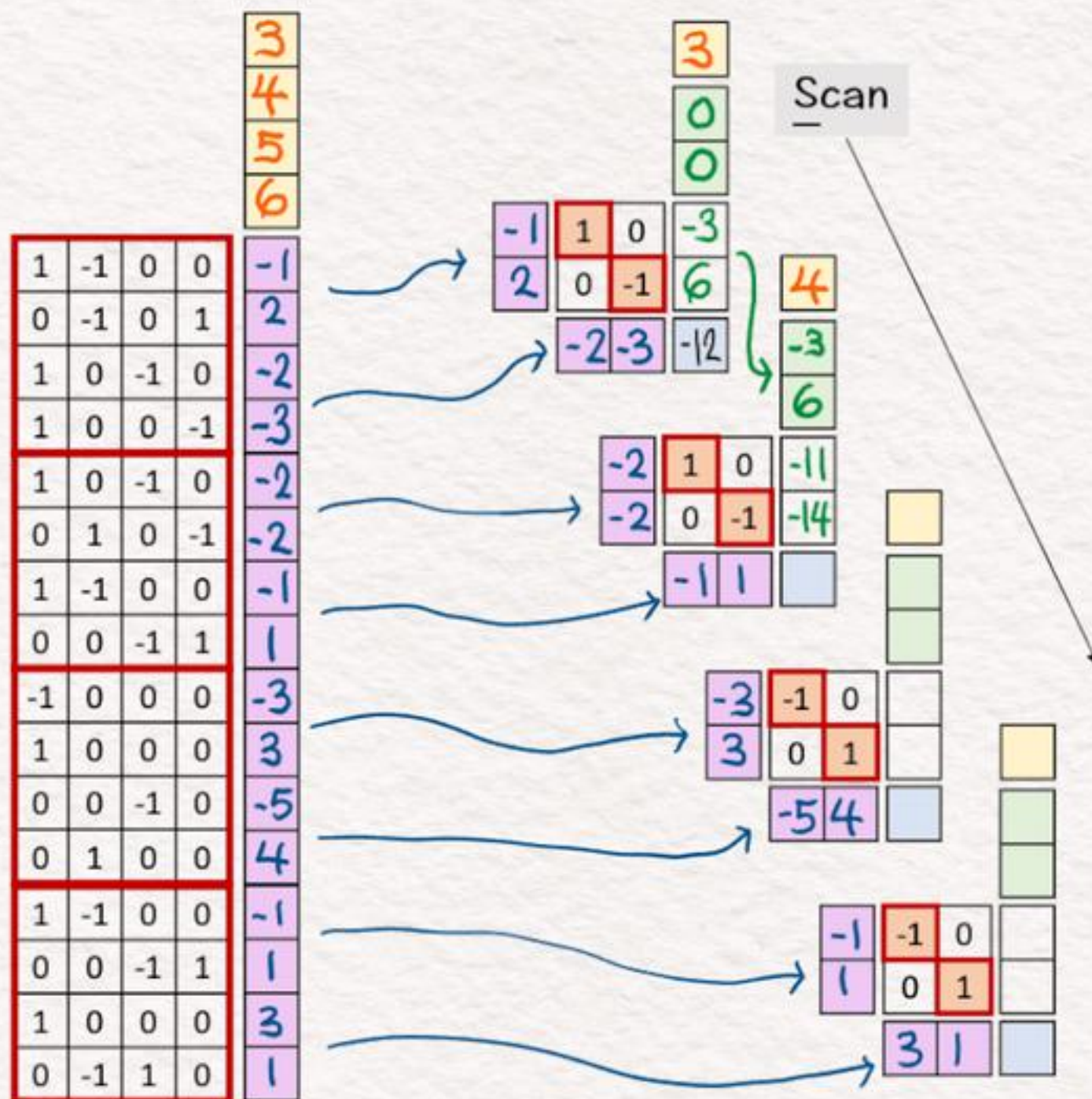
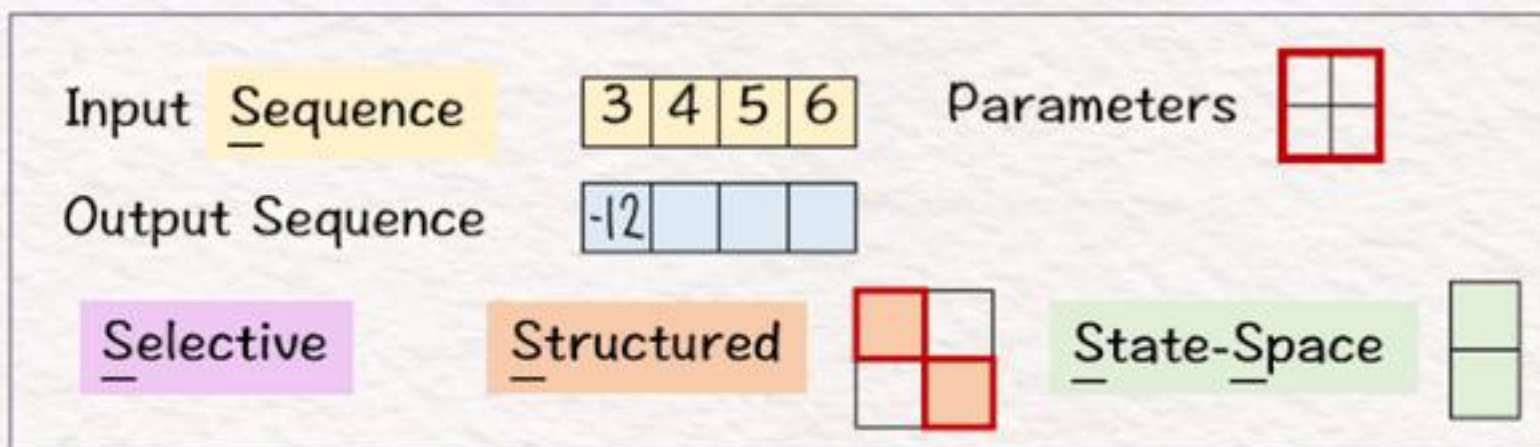


© 2023 Tom Yeh

[5] 使用不同的权重集A，B，C重复步骤3和4。

[5] Repeats steps 3 and 4 using different sets of weights A, B, C.

Mamba's S6 Model

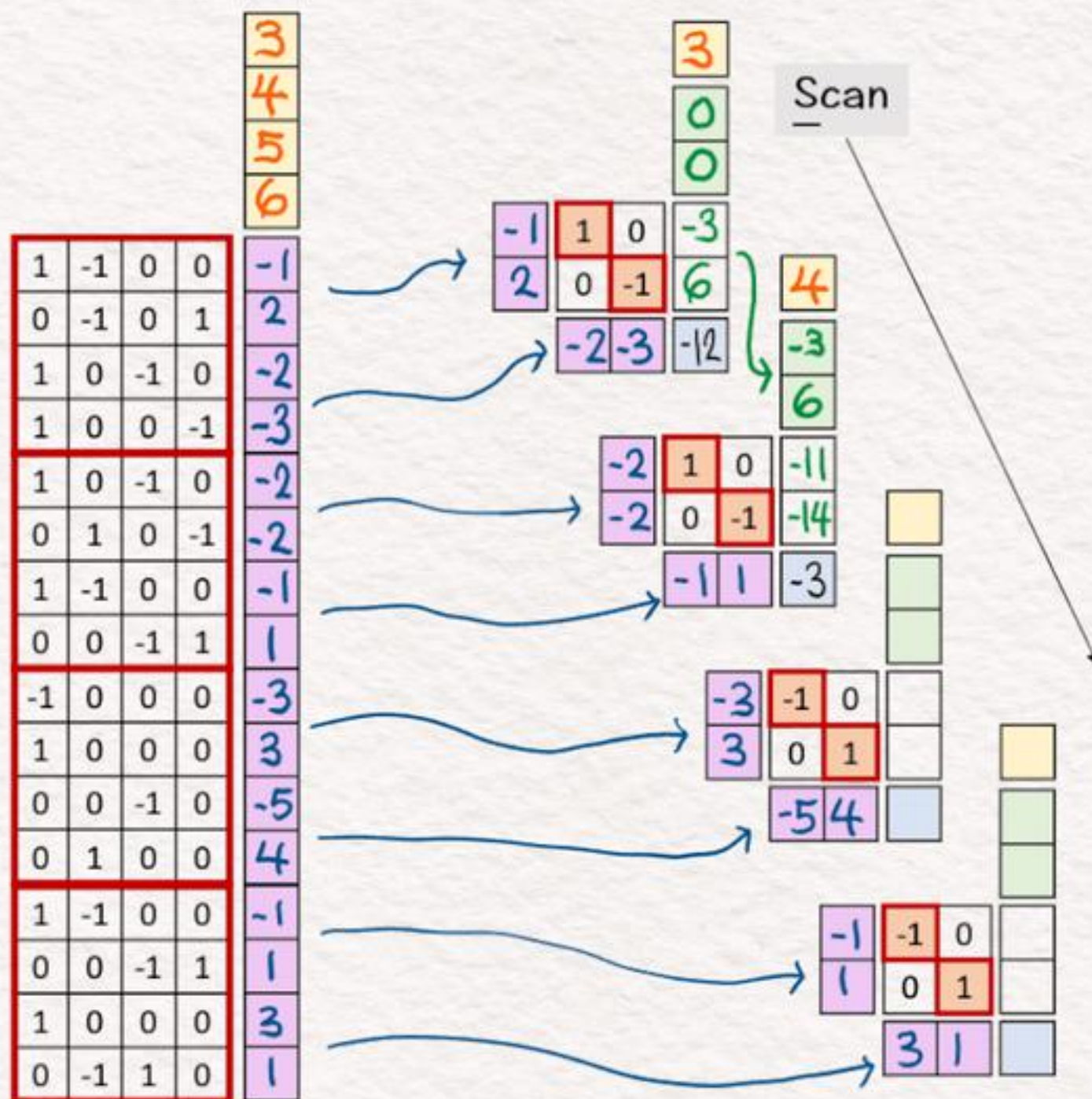
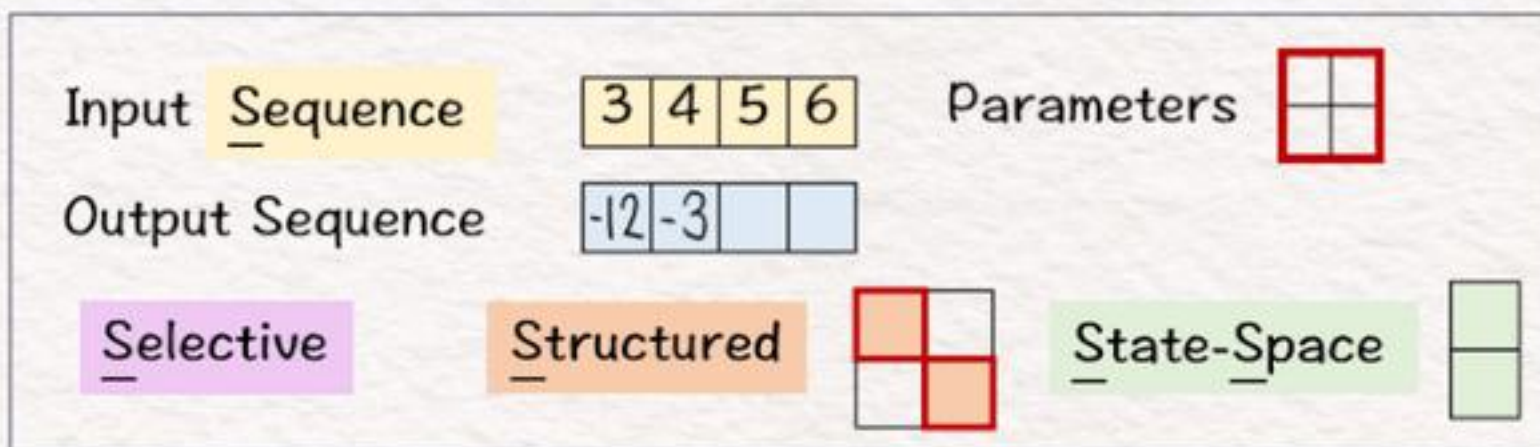


© 2023 Tom Yeh

[5] 使用不同的权重集A，B，C重复步骤3和4。

[5] Repeats steps 3 and 4 using different sets of weights A, B, C.

Mamba's S6 Model

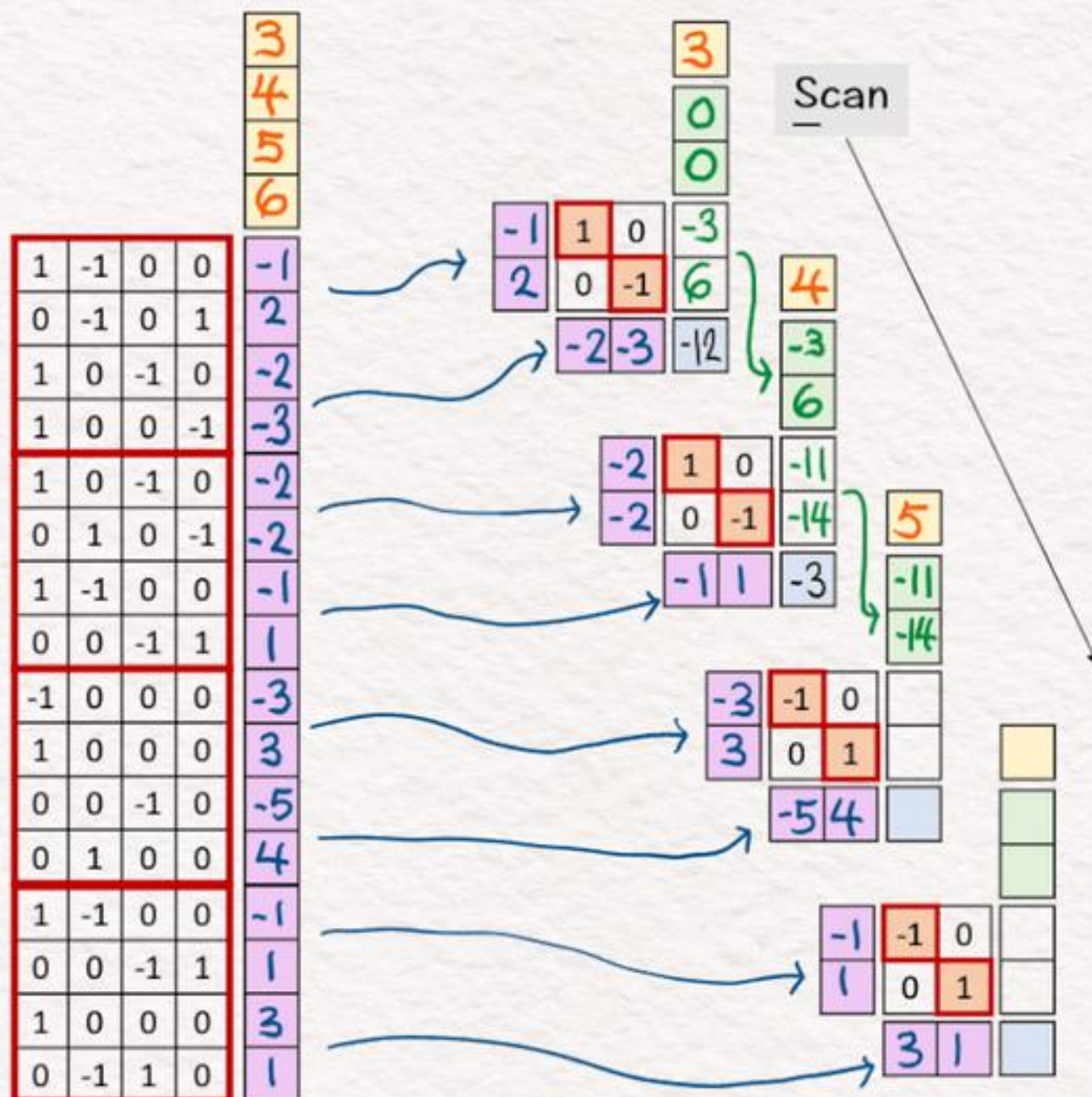
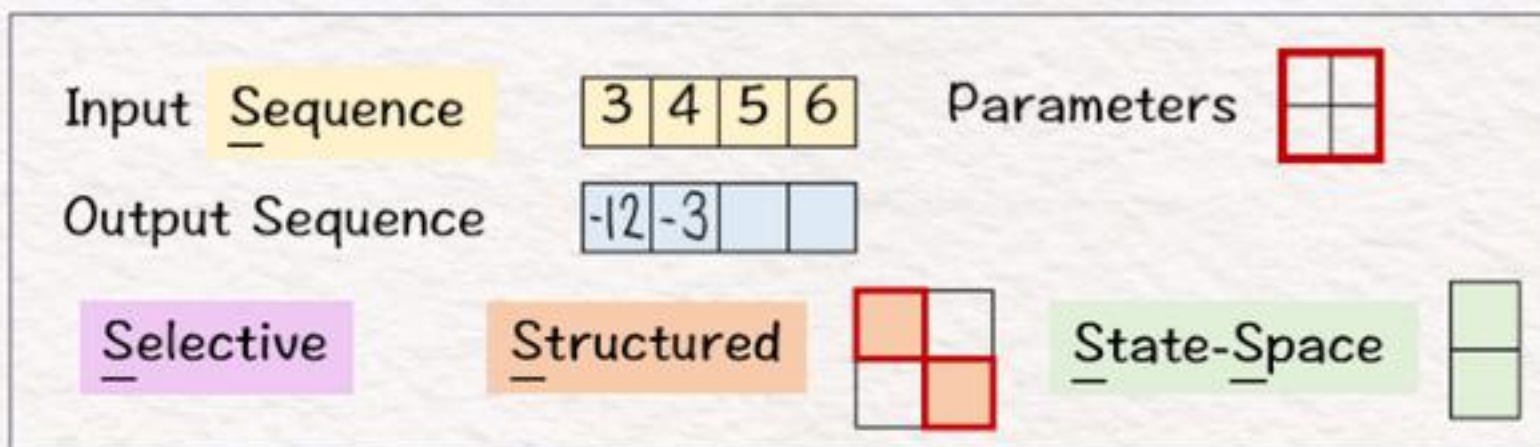


© 2023 Tom Yeh

[5] 使用不同的权重集A，B，C重复步骤3和4。

[5] Repeats steps 3 and 4 using different sets of weights A, B, C.

Mamba's S6 Model

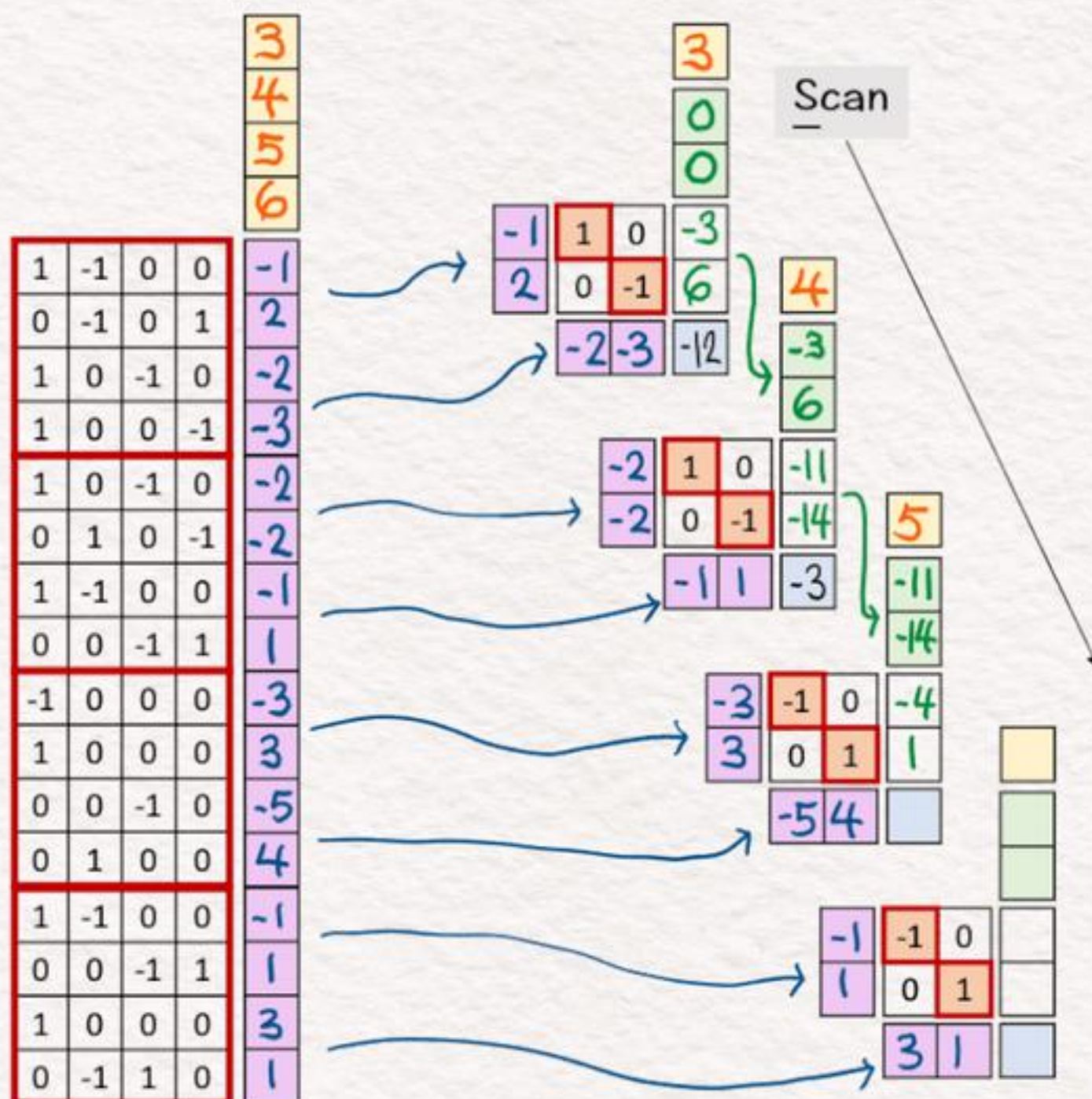
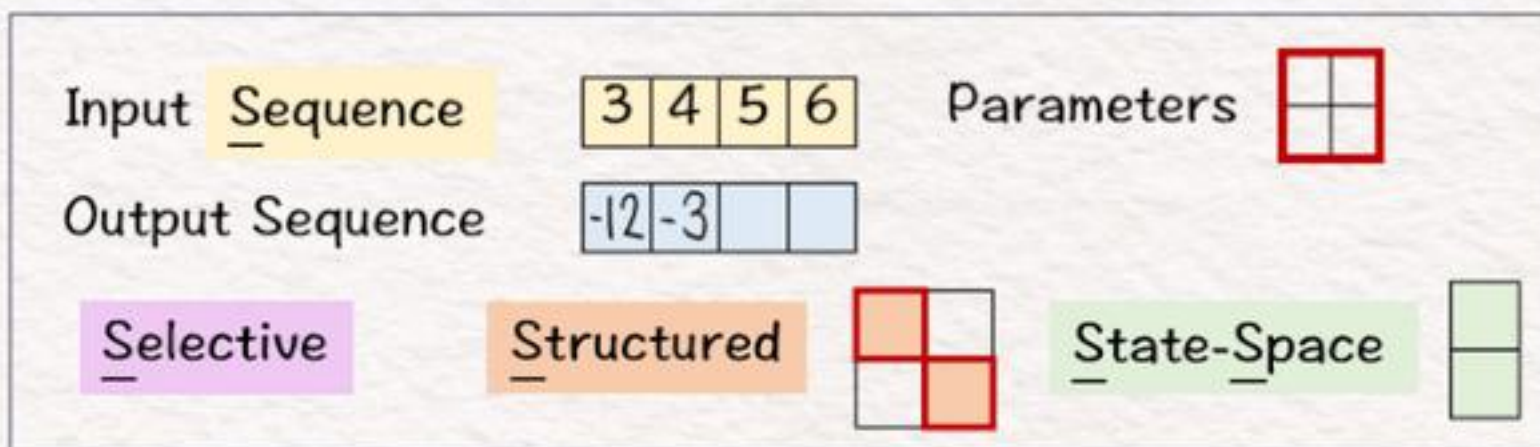


© 2023 Tom Yeh

[5] 使用不同的权重集A，B，C重复步骤3和4。

[5] Repeats steps 3 and 4 using different sets of weights A, B, C.

Mamba's S6 Model

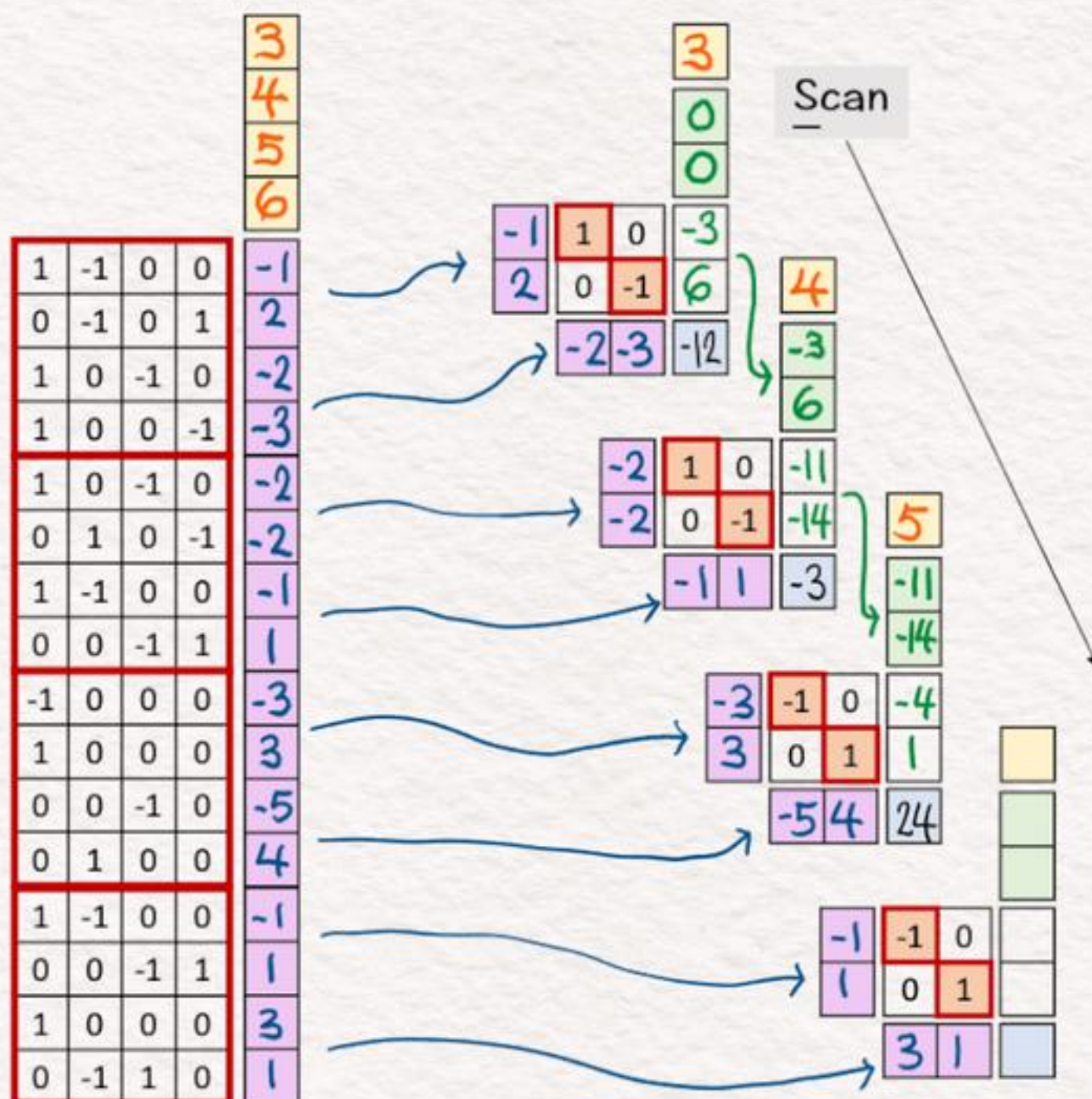
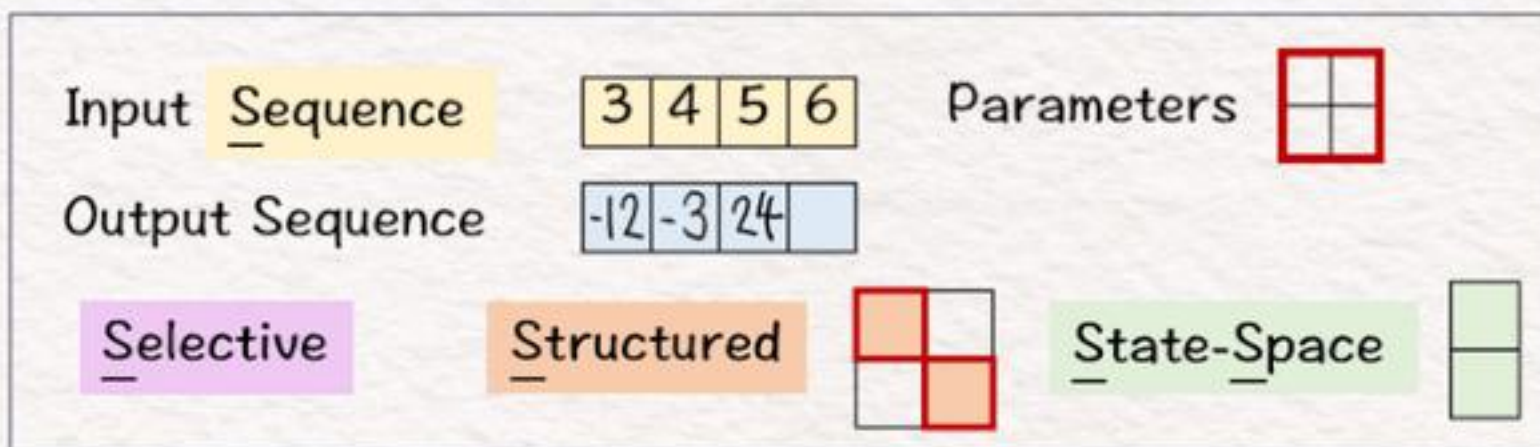


© 2023 Tom Yeh

[5] 使用不同的权重集A，B，C重复步骤3和4。

[5] Repeats steps 3 and 4 using different sets of weights A, B, C.

Mamba's S6 Model

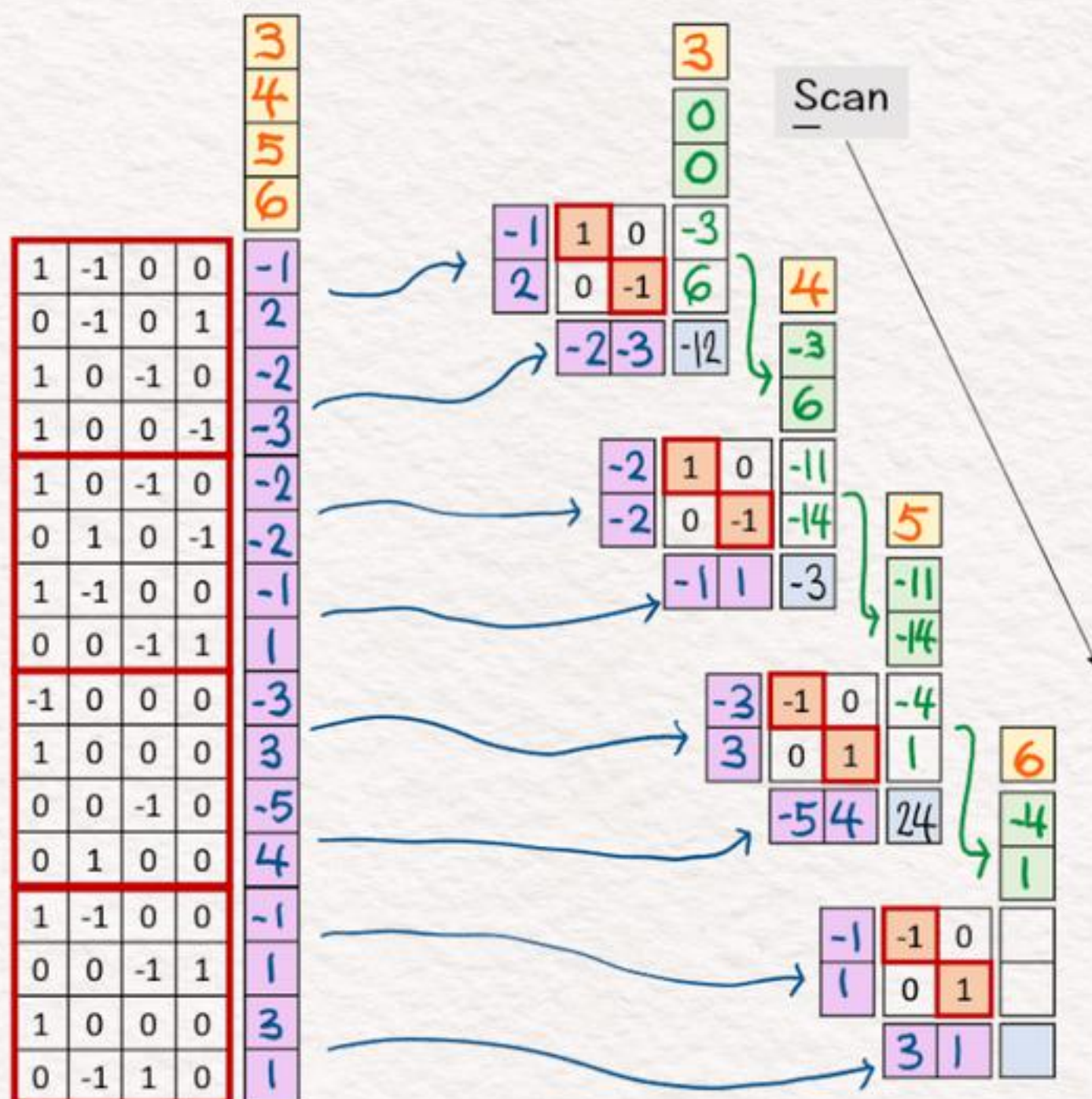
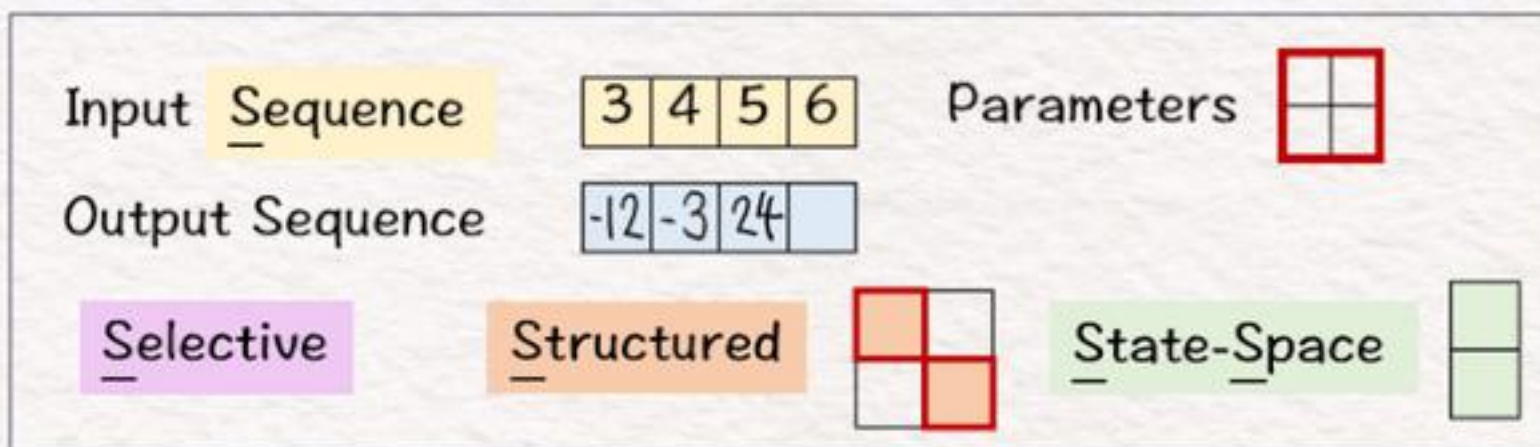


© 2023 Tom Yeh

[5] 使用不同的权重集A，B，C重复步骤3和4。

[5] Repeats steps 3 and 4 using different sets of weights A, B, C.

Mamba's S6 Model

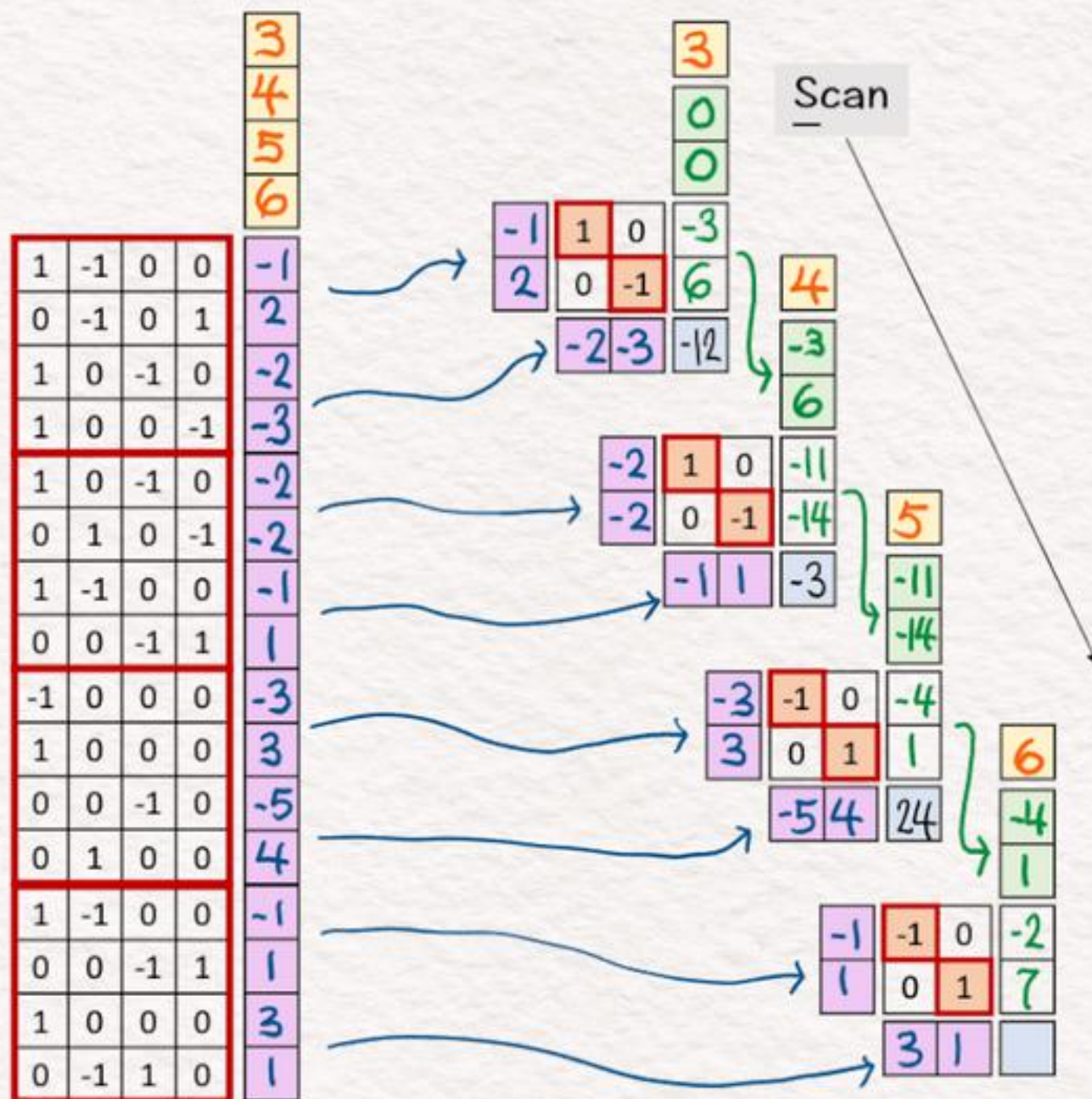
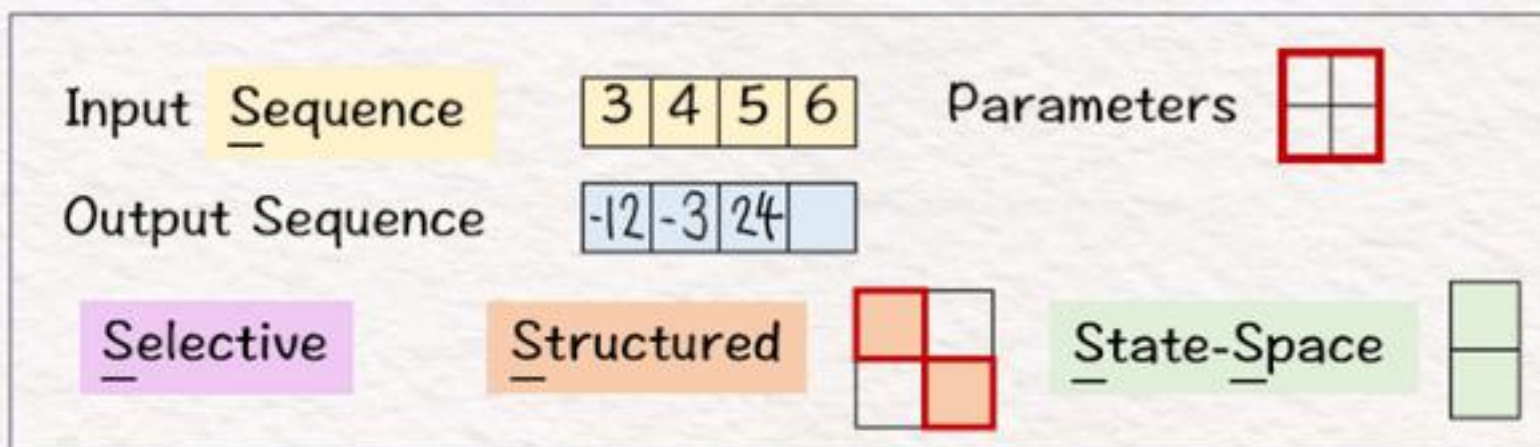


© 2023 Tom Yeh

[5] 使用不同的权重集A，B，C重复步骤3和4。

[5] Repeats steps 3 and 4 using different sets of weights A, B, C.

Mamba's S6 Model

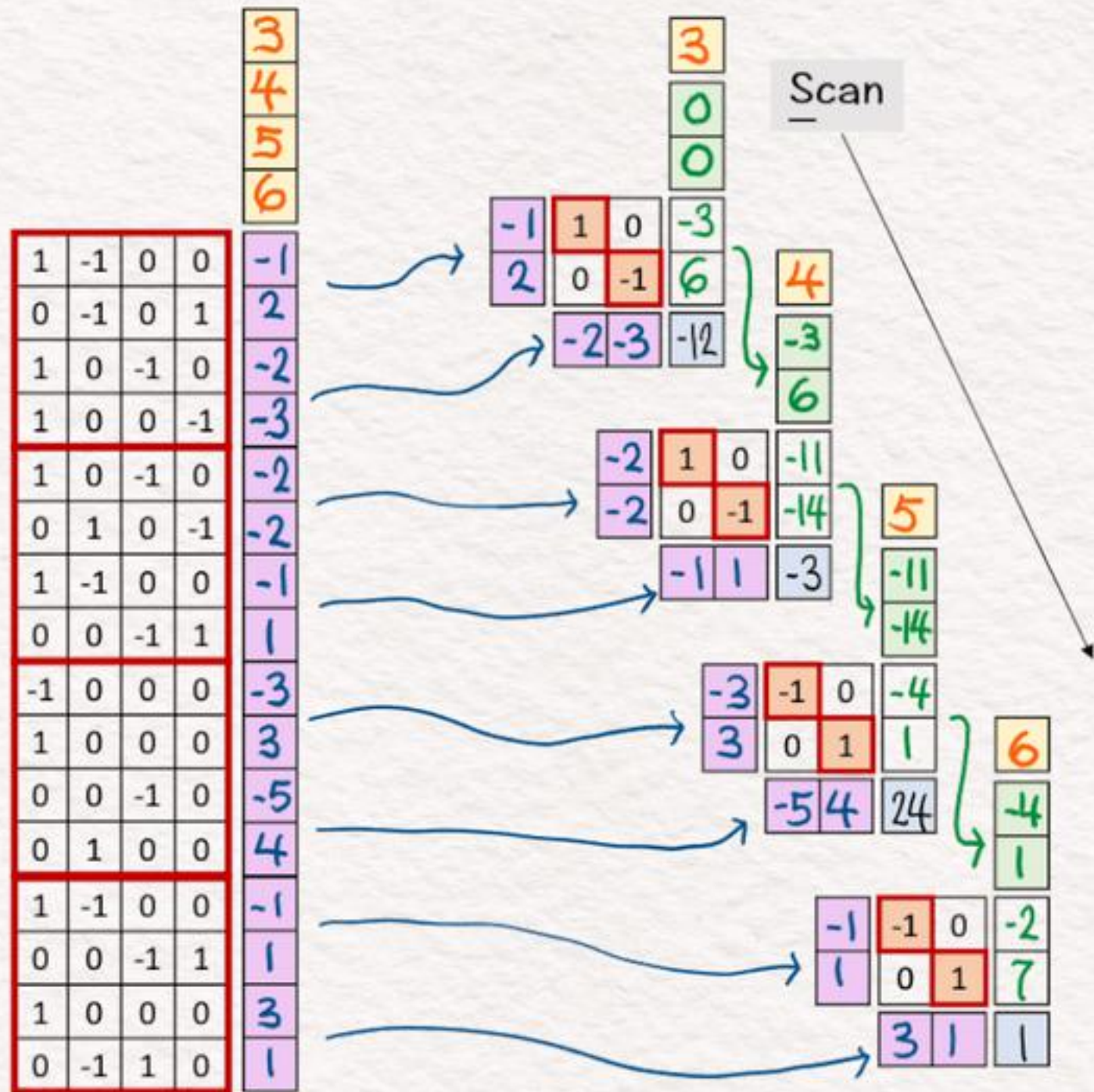
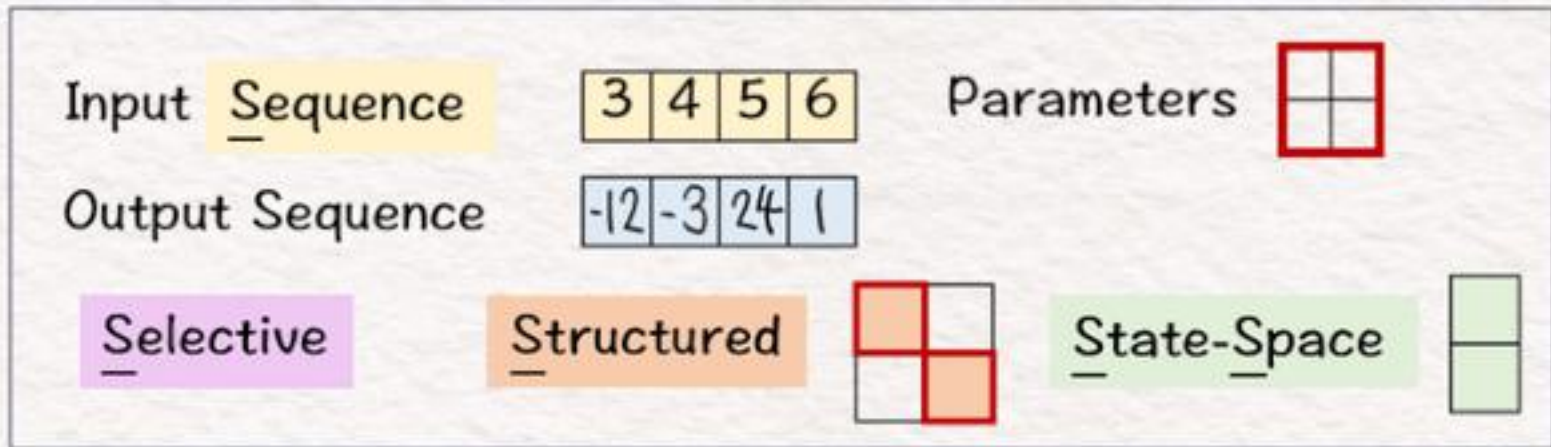


© 2023 Tom Yeh

[5] 使用不同的权重集A，B，C重复步骤3和4。

[5] Repeats steps 3 and 4 using different sets of weights A, B, C.

Mamba's S6 Model



© 2023 Tom Yeh

[5] 使用不同的权重集A，B，C重复步骤3和4。

[5] Repeats steps 3 and 4 using different sets of weights A, B, C.