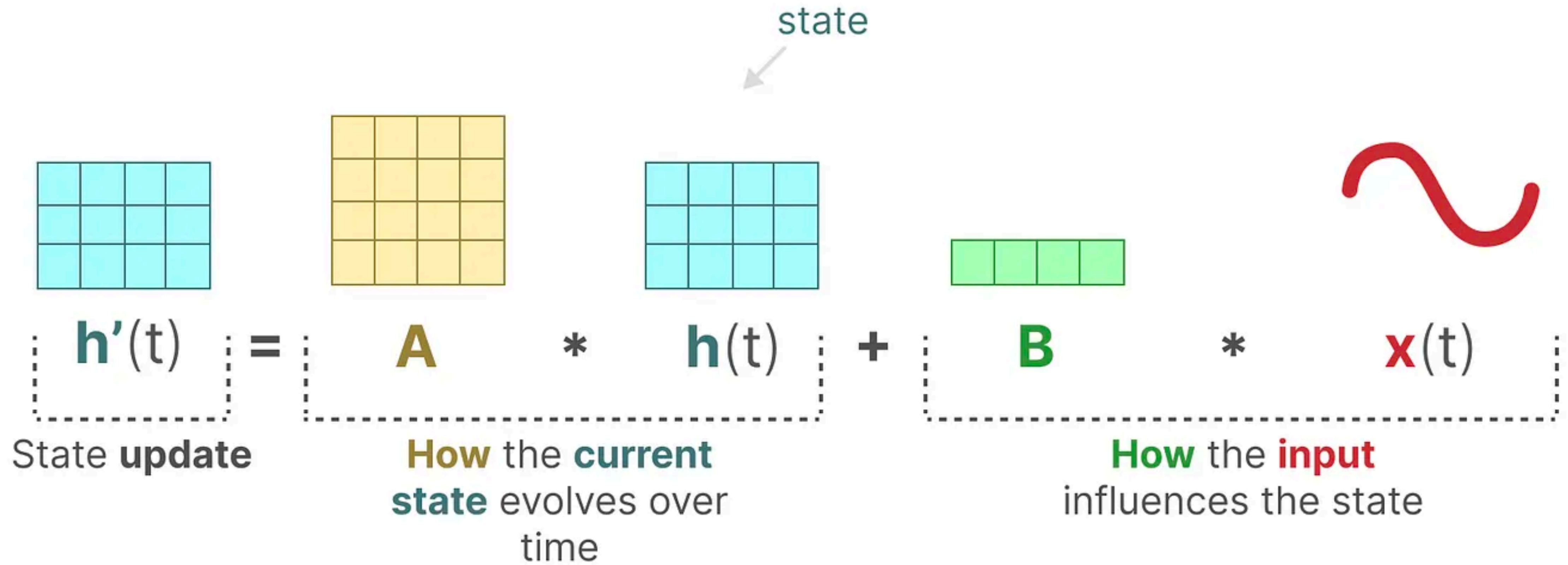


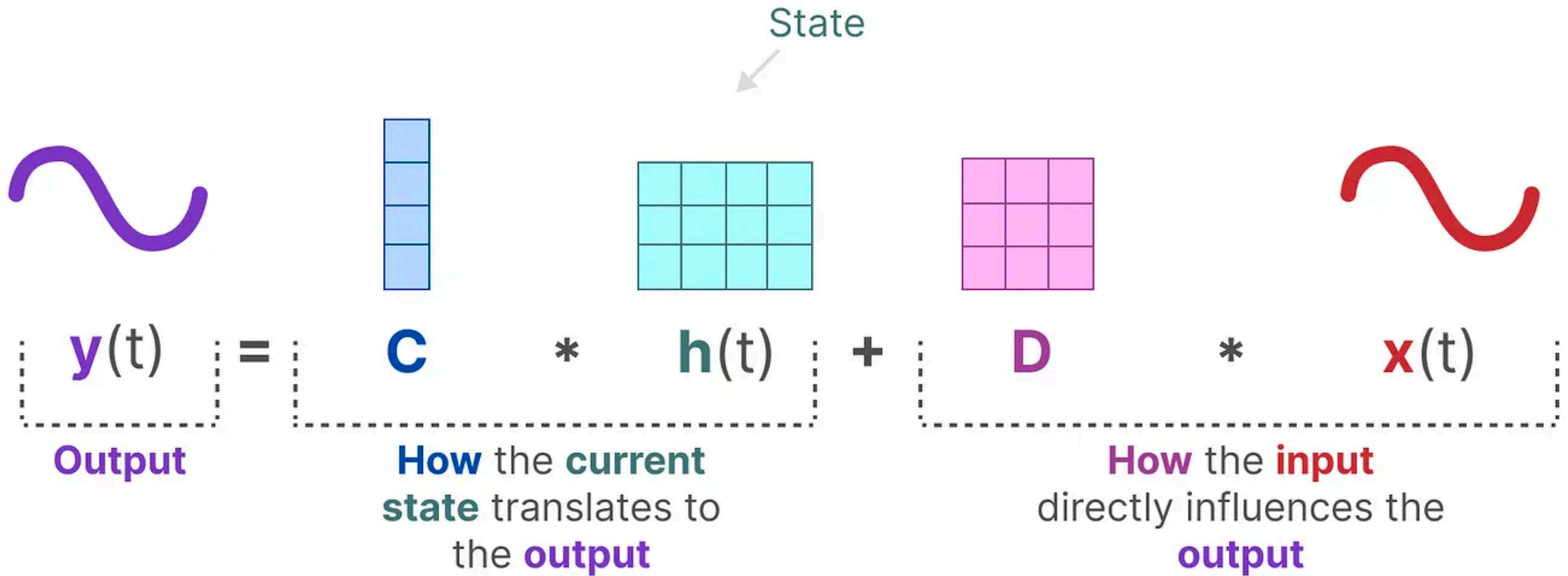
# Mamba

~~САЙТ ЗАКОНСТВ~~

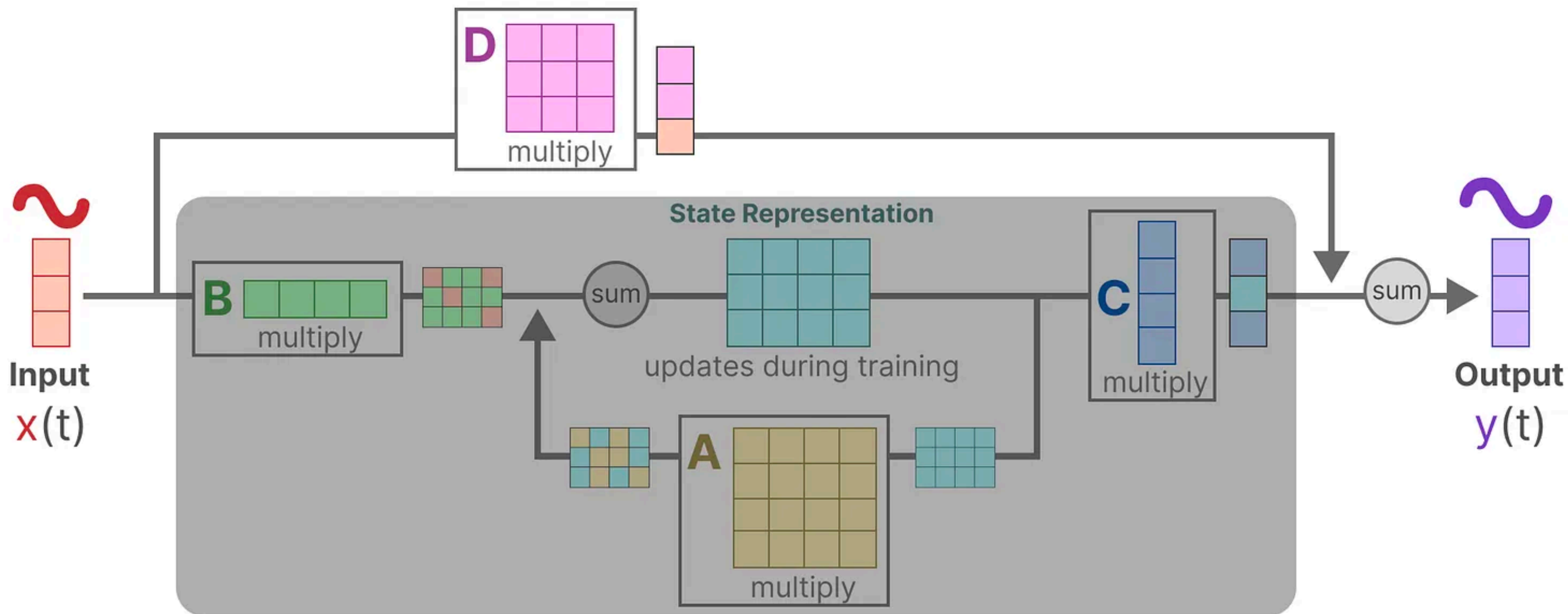
# State Space Models



# State Space Models



# State Space Models



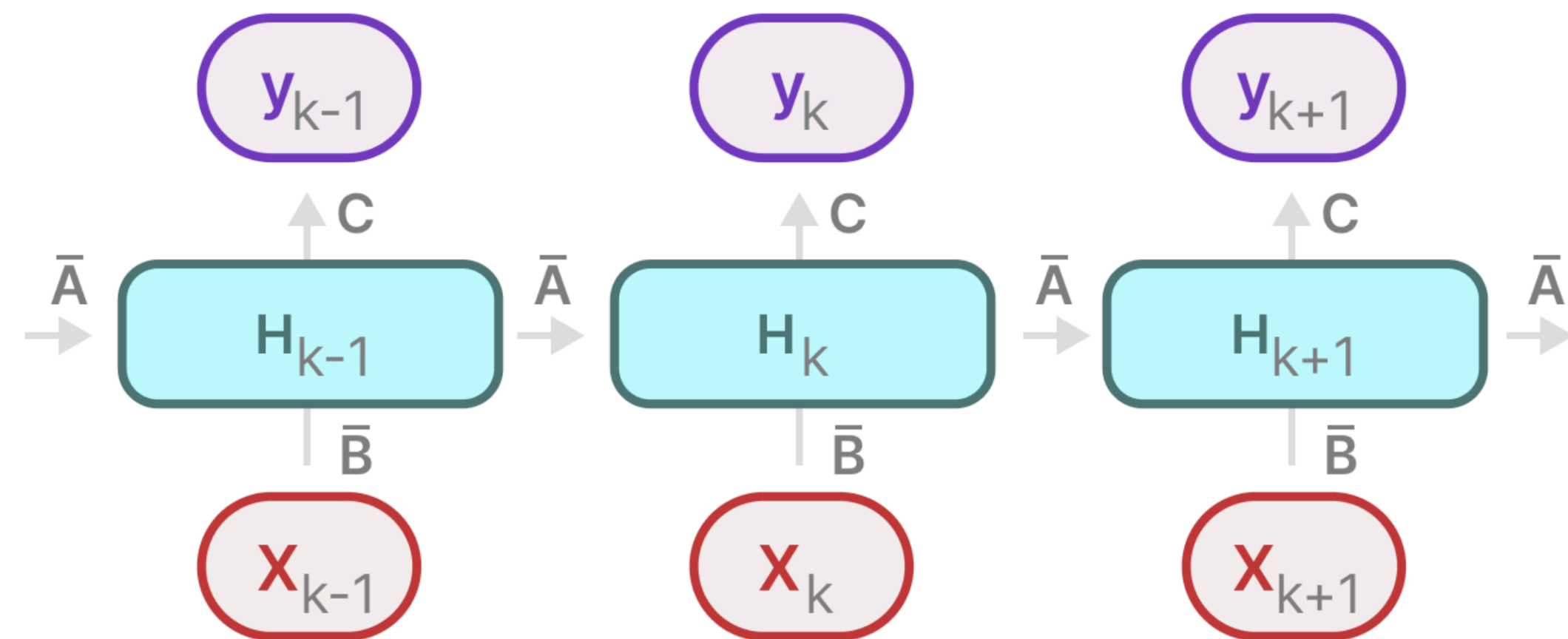
State Space Model

# State Space Models

## Discretisation

Discretized matrix  $\mathbf{A}$        $\bar{\mathbf{A}} = \exp(\Delta \mathbf{A})$

Discretized matrix  $\mathbf{B}$        $\bar{\mathbf{B}} = (\Delta \mathbf{A})^{-1} (\exp(\Delta \mathbf{A}) - \mathbf{I}) \cdot \Delta \mathbf{B}$



**SSM**  
(Recurrent + Unfolded)

# State Space Models



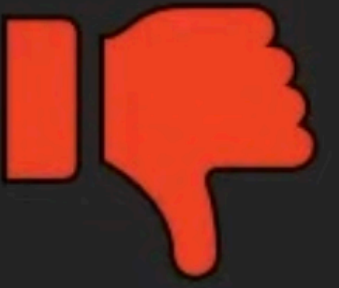


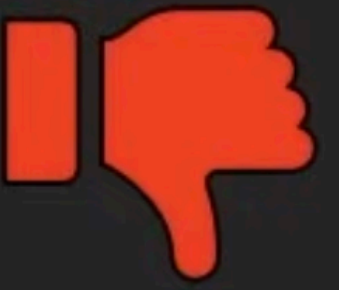



$$\text{kernel} \longrightarrow \bar{\mathbf{K}} = (\mathbf{C}\bar{\mathbf{B}}, \mathbf{C}\bar{\mathbf{A}}\bar{\mathbf{B}}, \dots, \mathbf{C}\bar{\mathbf{A}}^{k-1}\bar{\mathbf{B}}, \dots)$$

$$\mathbf{y} = \mathbf{x} * \bar{\mathbf{K}}$$

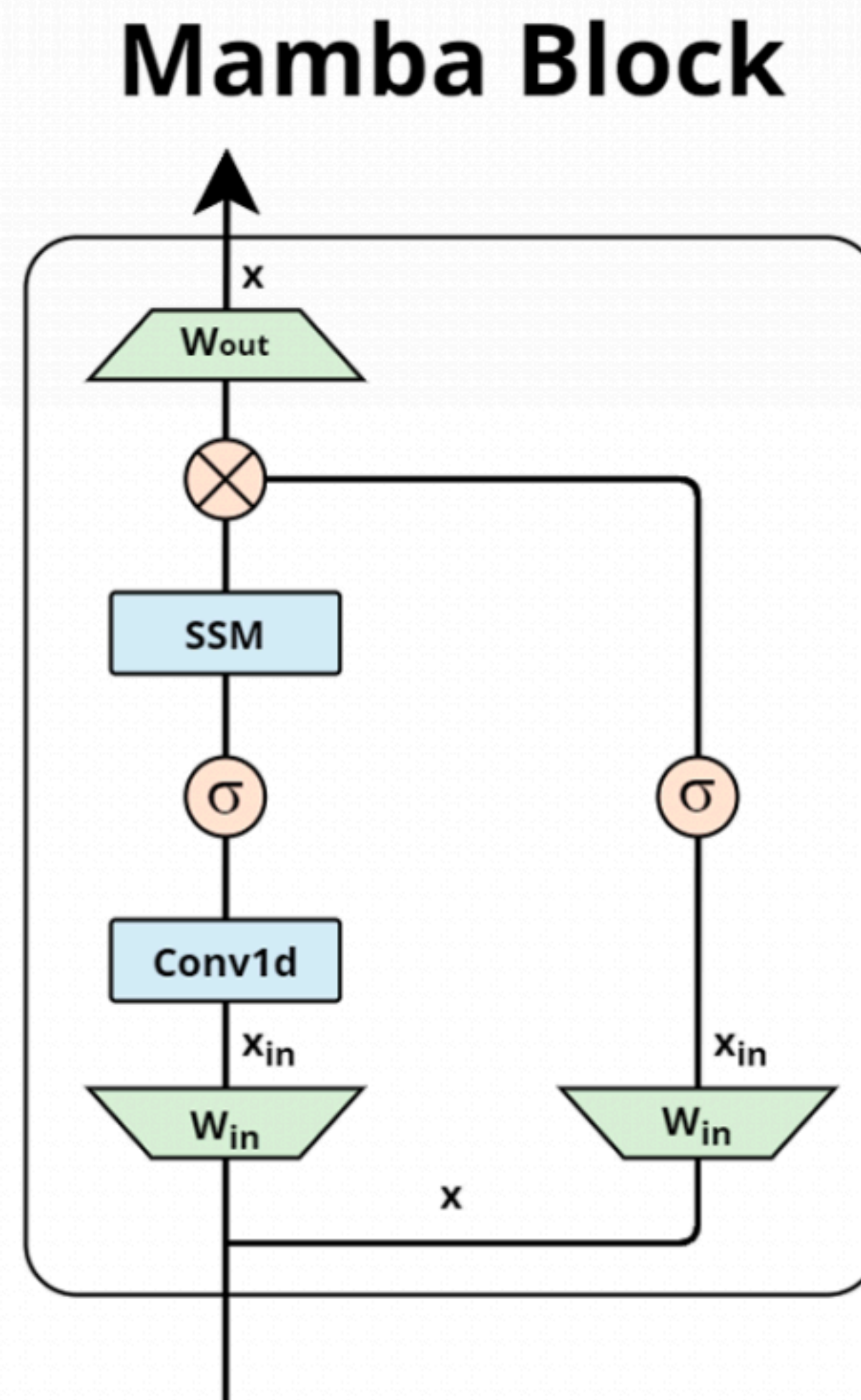
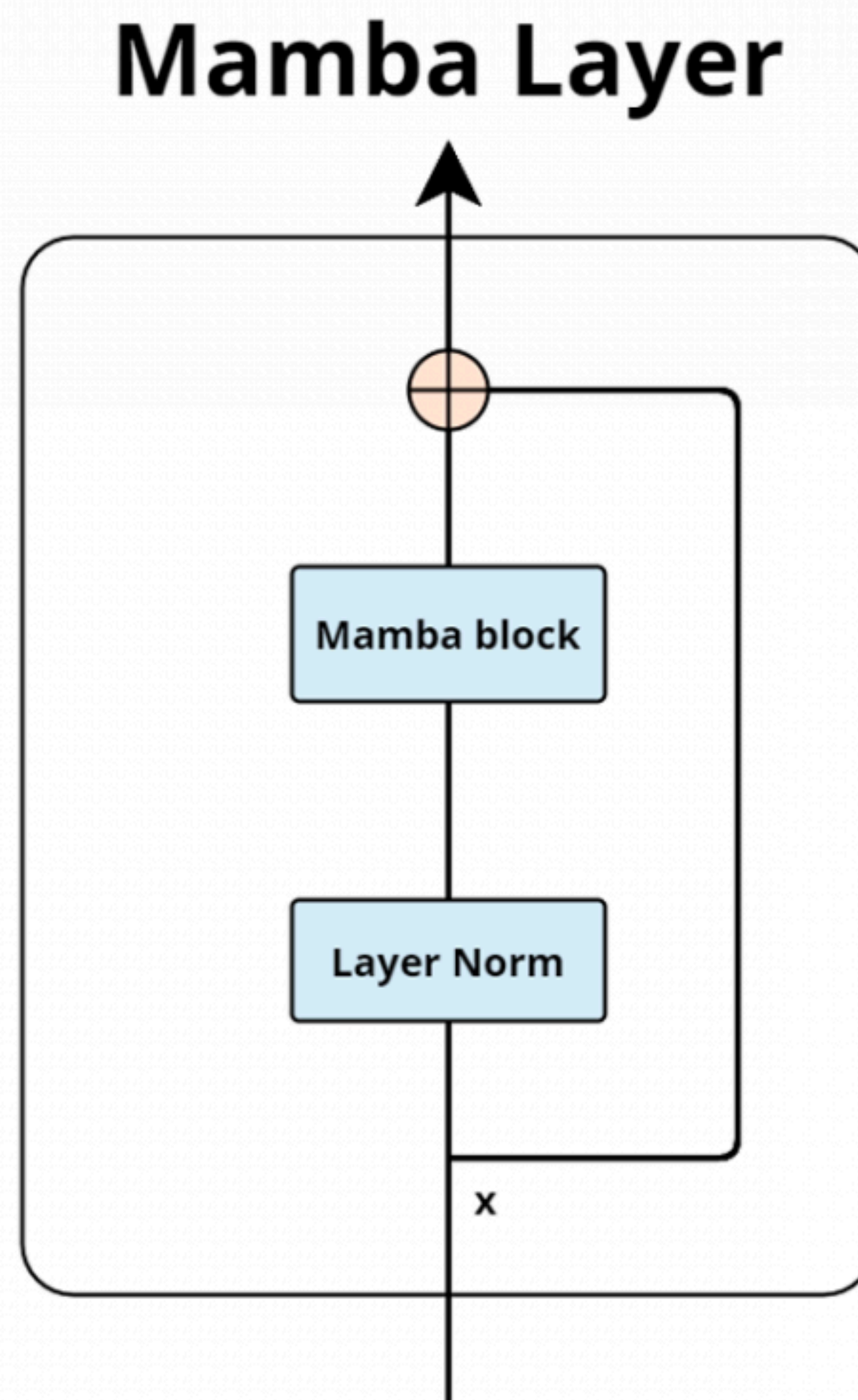
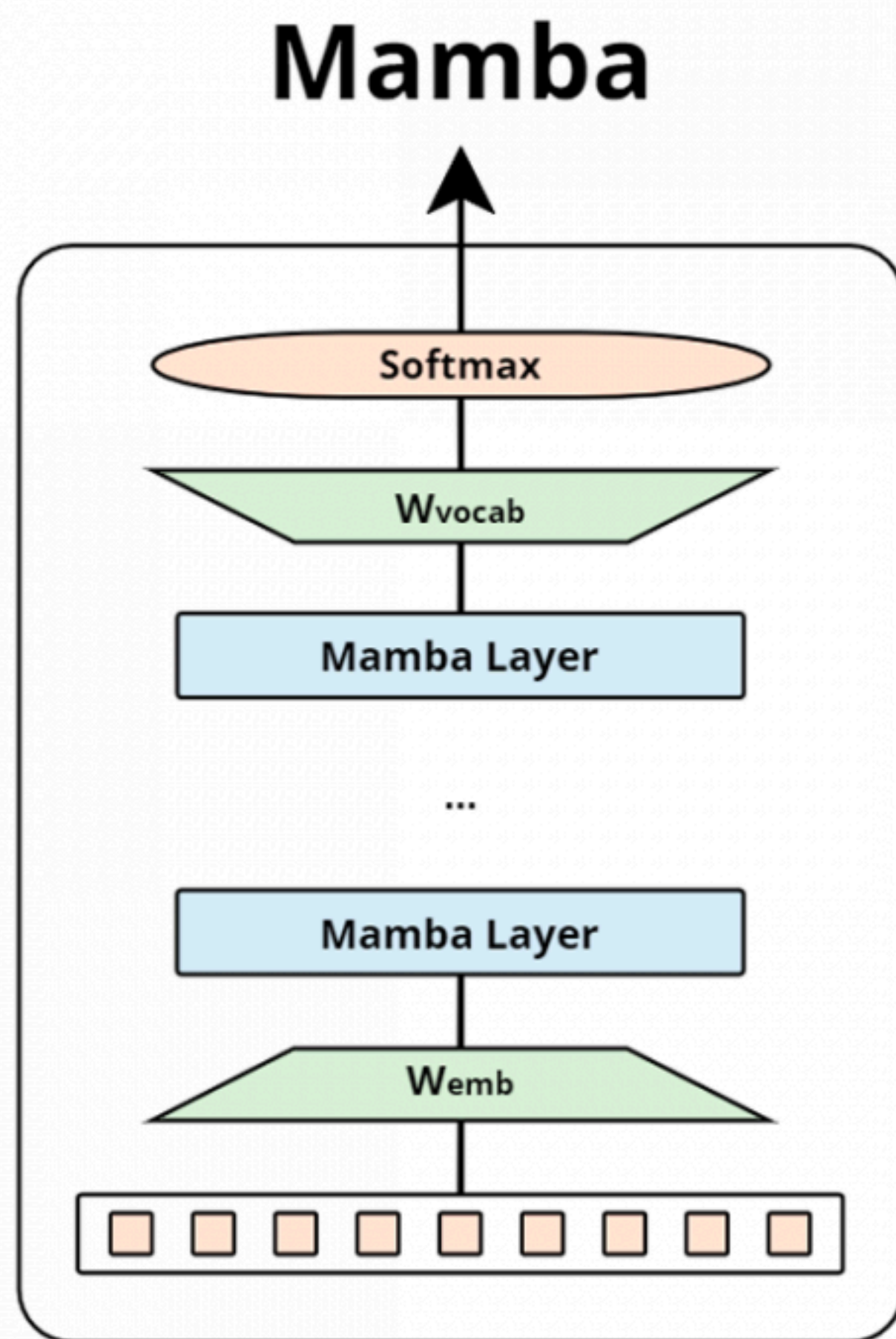
output      input      kernel



# Space Models

	Selective SSMs Mamba	State Space Models (SSMs)	Transformers
speed			
memory			
performance (e.g., accuracy)			

# Mamba





# In-context-learning

- Mamba

---

```
mom - maman; dad - papa; cat - chat; dog - chien; hello - bonjour; man - homme; woman - femme; god - dieu; sun - soleil; moon - lune; tree -  
arbre; red - rouge; blue - bleu; green -  
green; black - noir;
```

---

```
boy - girl; buy - sell; expensive - cheap; child - adult; clean - dirty; cold - hot; discomfort - comfort; correct - wrong; create - destroy;  
good -  
bad; happy - sad; healthy - sick; hot - cold; light - dark; love - -; -
```

- GPT-2

---

```
[{'generated_text': 'mom - maman; dad - papa; cat - chat; dog - chien; hello - bonjour; man - homme; woman - femme; god - dieu; sun - soleil;  
l; moon - lune; tree - arbre; red - rouge; blue - bleu; green - vert; black - ernest;'},
```

---

```
Setting pad_token_id to eos_token_id 150256 for open end generation.  
[{'generated_text': 'boy - girl; buy - sell; expensive - cheap; child - adult; clean - dirty; cold - hot; discomfort - comfort; correct - wrong;  
ong; create - destroy; good - _____ - perfect.\n\nDictionaries\n\nFrench - dictionnaire\n\nEnglish - standard English word list for co  
mputer programs and other texts\n\nEnglish - International'},
```

# Metrics

## Lambada:

BooksCorpus where a missing target word is predicted in the last sentence of each passage

	small gpt2	small mamba	big gpt2	big mamba
perplexity	40.0554	16.0659	10.6341	5.0425
accuracy	0.3256	0.4425	0.5121	0.6495

## WinoGrande

WSC tasks (aka yes/no questions)

	small gpt2	small mamba	big gpt2	big mamba
accuracy	0.5162	0.5209	0.5833	0.6140

## ARC

multiple-choice question-answering dataset

	small gpt2	small mamba	big gpt2	big mamba
arc_easy	0.4381	0.4798	0.5829	0.6553
arc_challenge	0.1903	0.1971	0.2500	0.2978

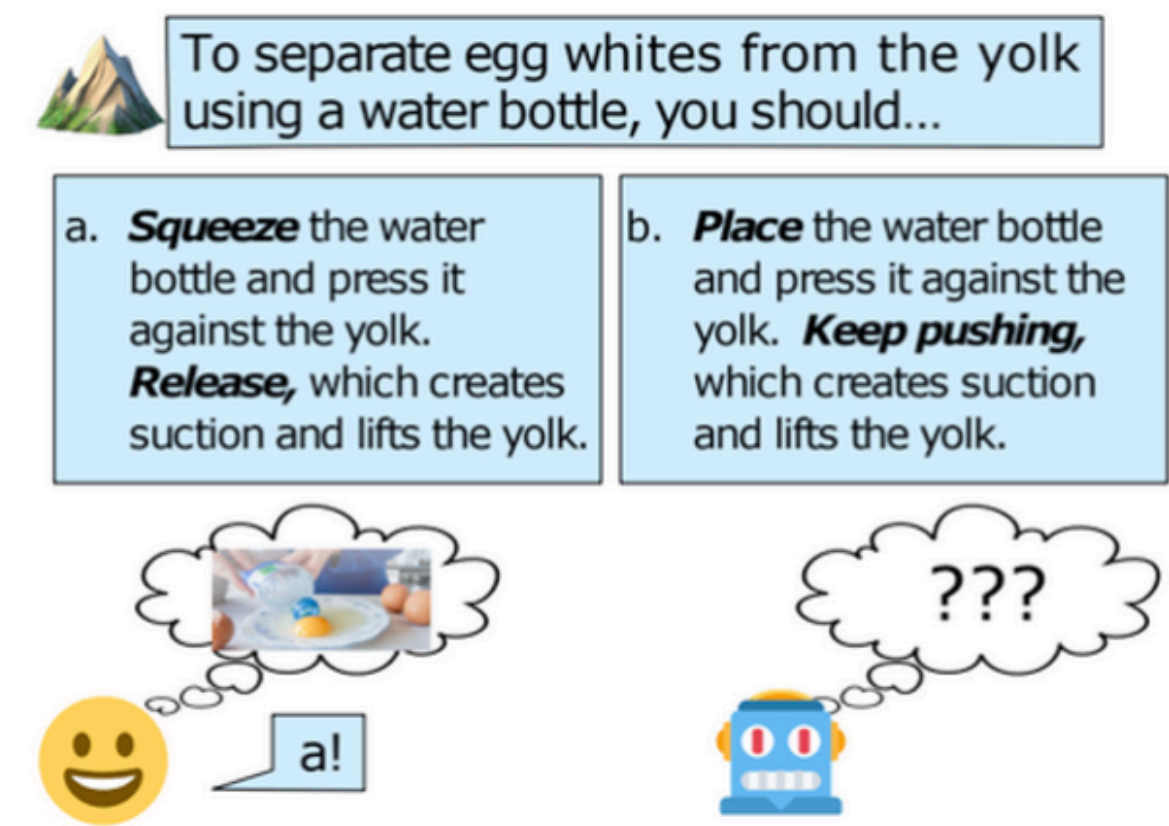
# BHW GAN

Mamba models: small(130m) and big(1.4b)

GPT models: small(137m) and big(1.6b)

## PIQA (Physical Interaction: Question Answering)

PIQA is a dataset for commonsense reasoning, and was created to investigate the physical knowledge of existing models in NLP.



	small gpt2	small mamba	big gpt2	big mamba
piqa	0.6289	0.6447	0.7084	0.7416