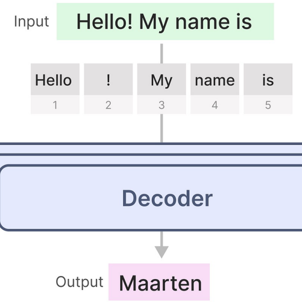
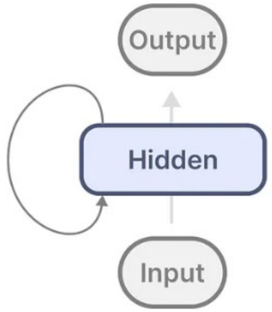
Mamba: Linear-Time Sequence Modeling with Selective State Spaces

State Space Model

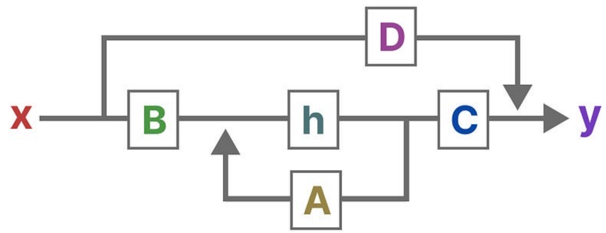
What is state space?



Transformer

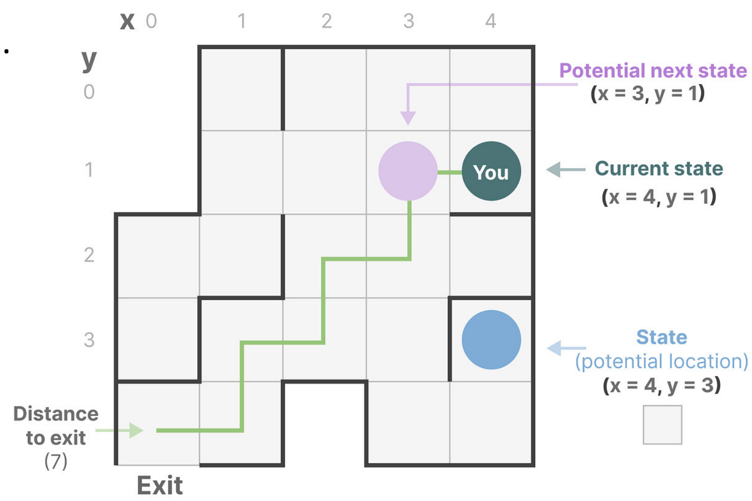


RNN



State Space Model

State Space: the minimal number of states to describe a system.



State:

position: (x_i, y_j),

Potential next position: (x_i', y_j')

Action: (left , right)

Dist: distance to exit

What is a state space model?

SSM: describe these state and make predictions of what their next state could be depending on some input.

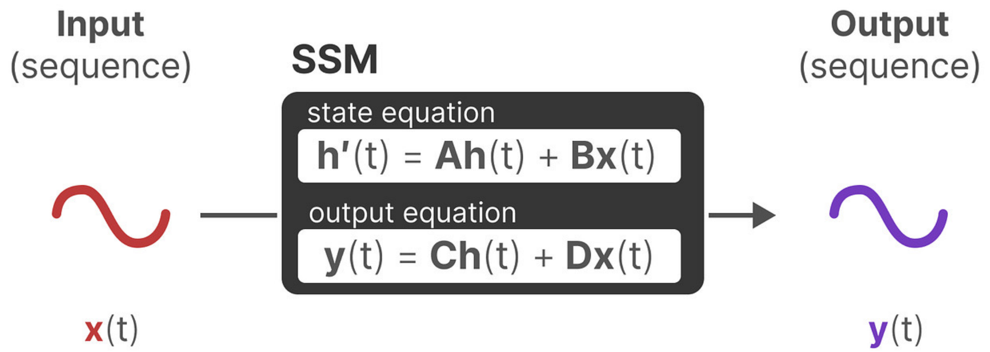


Fig A SSM model

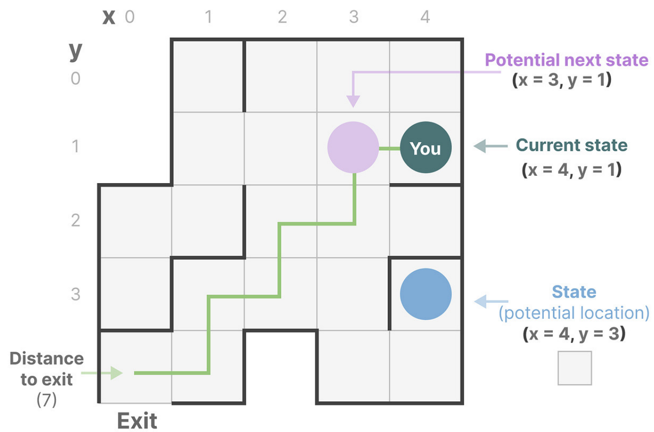
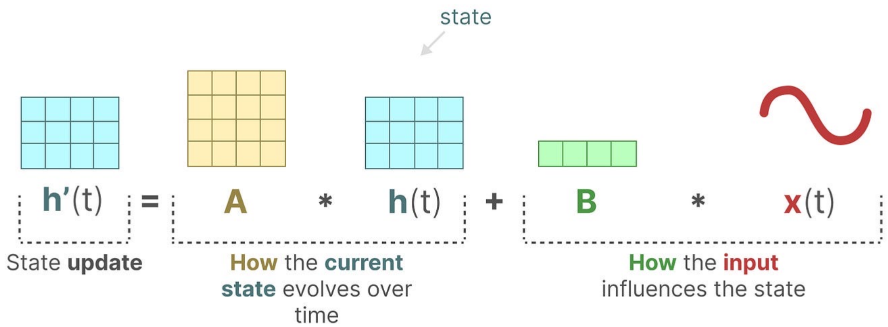


Fig Maze Model

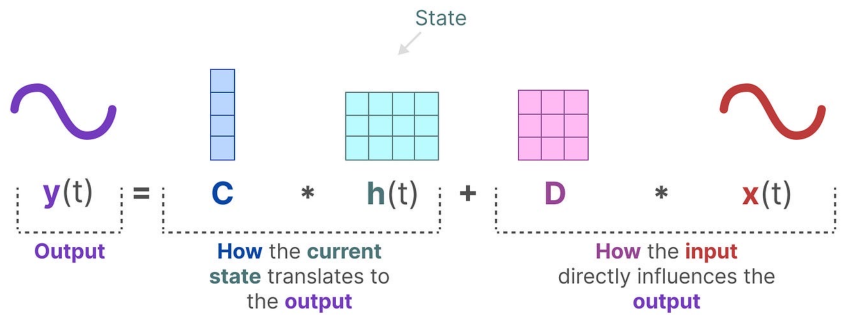
(At time t)	SSM	Maze	LLM(SSM-based)
Input sequence	$x(t)$	Move left	Sequence tokens
Hidden states	$h(t)$	Position + Distance to exit	Memory information
A predicted output	$y(t)$	Move down	Next token

What is a state space model?

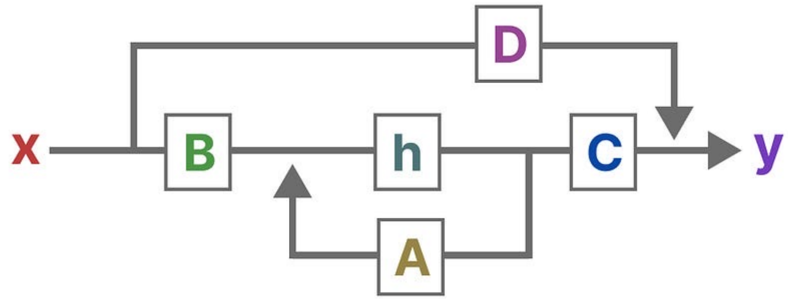
SSM: describe these state and make predictions of what their next state could be depending on some input.



Hidden states update



Prediction



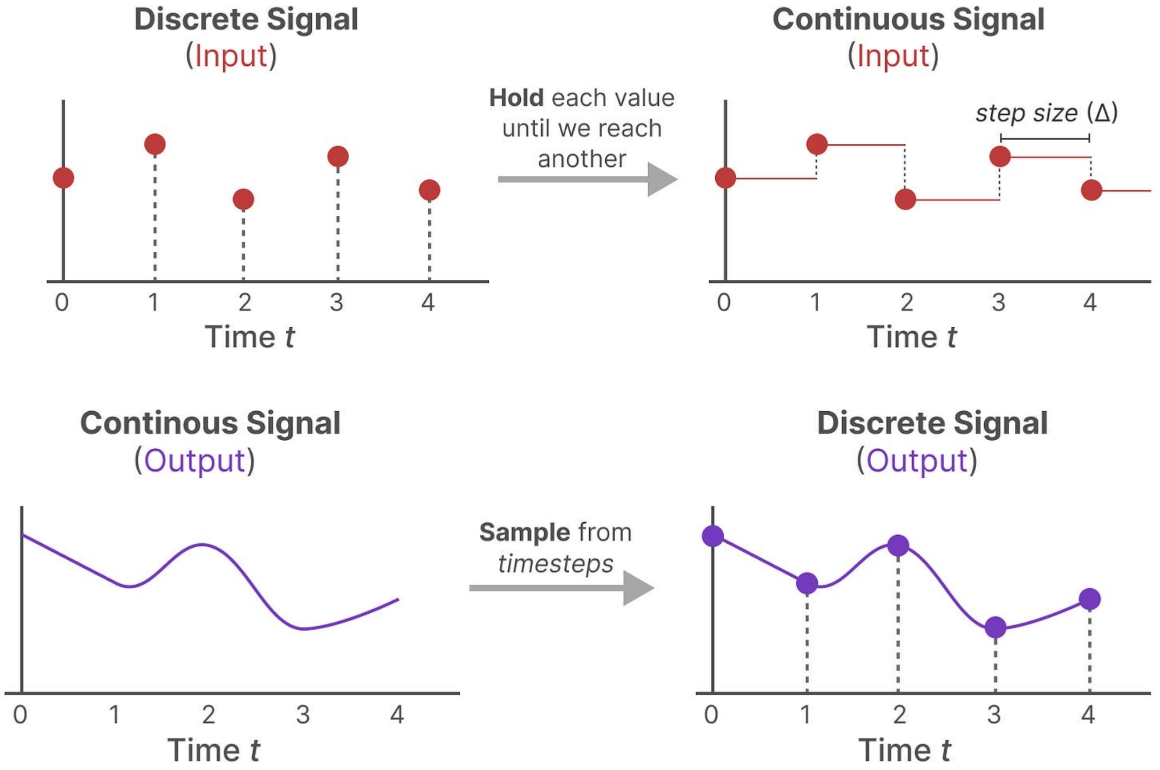
SSM pipeline
continuous-time representation

Continuous system：

1. **Too difficult to find an initial hidden state h(t)**
2. **Computing is always discrete**
3. **We need the input as a discrete type (text sequence)**

From a continuous to a discrete signal

How? -> Answer: Sample + ZoH (zero-order hold)



Every time we receive a discrete signal, we hold its value until we receive a new discrete signal.

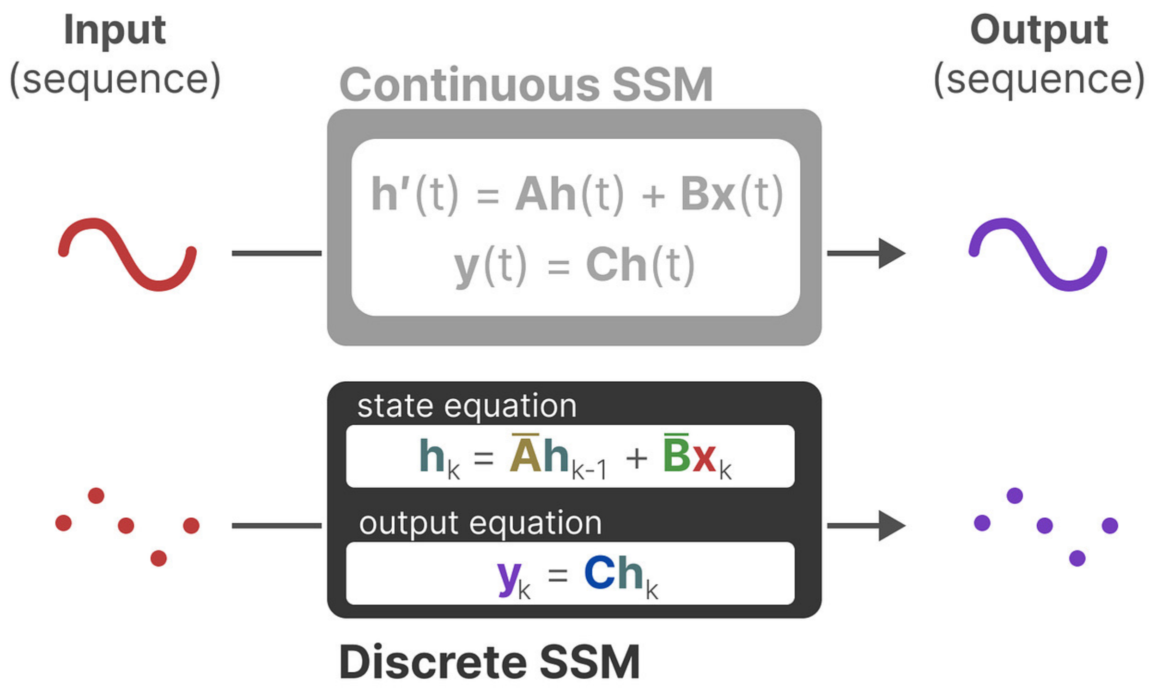
How long we hold the value is represented by a new learnable parameter, called the *step size* **Δ**.

Discrete format of Matrix A and B

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

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

Summary



D can be treated as a skip-connection and we can ignore it in SSM block