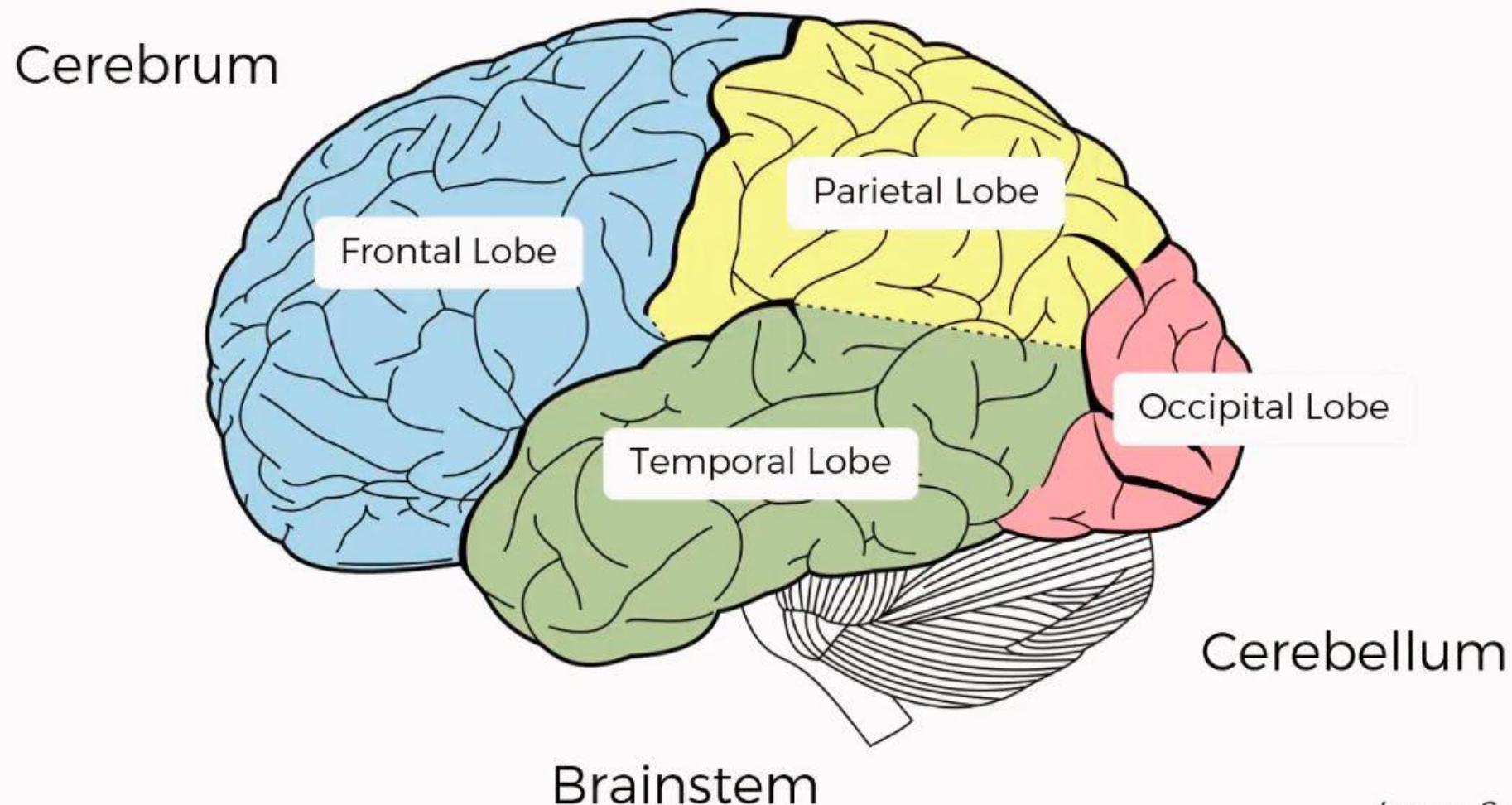
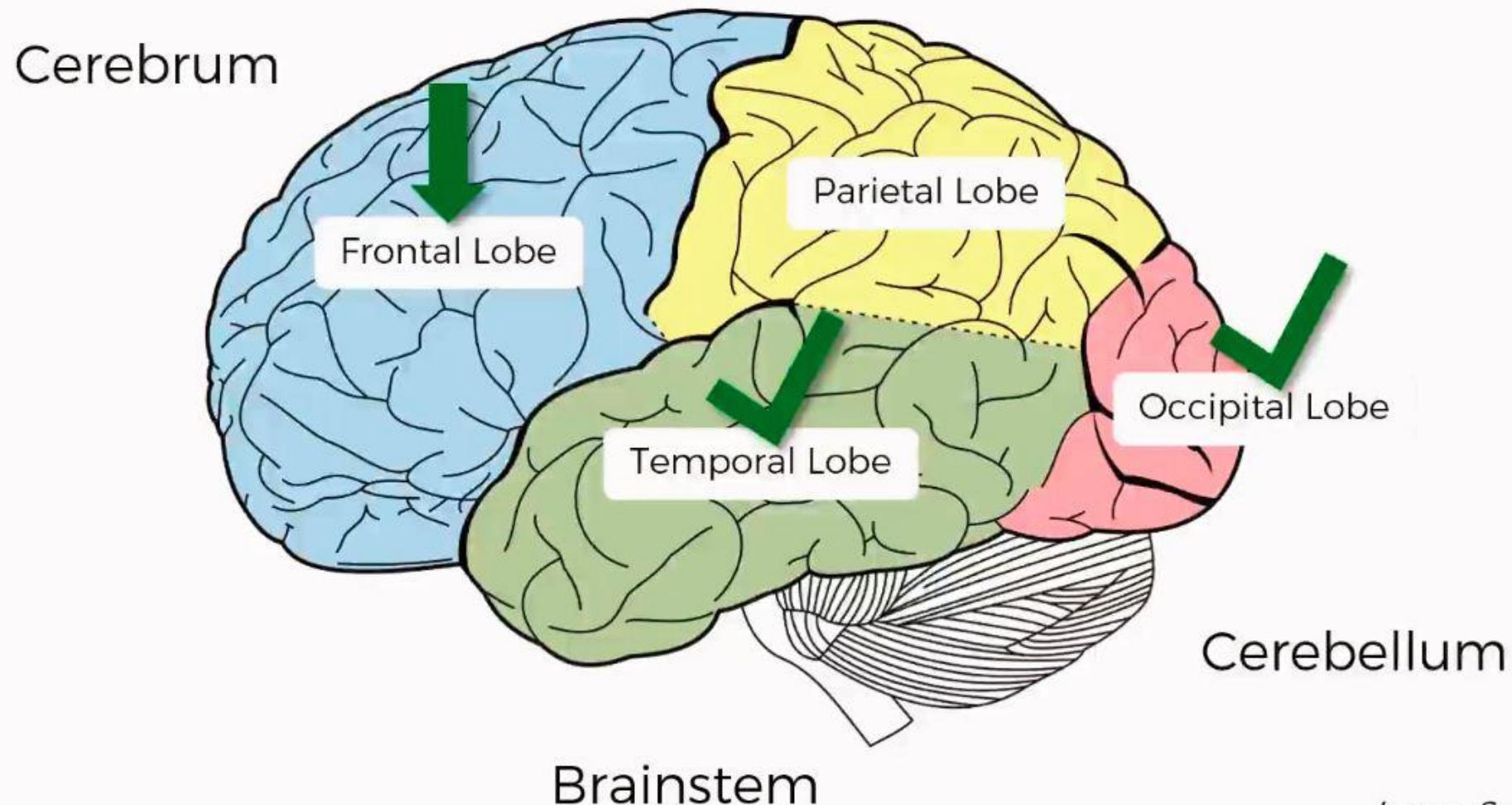


# **Recurrent Neural Networks**

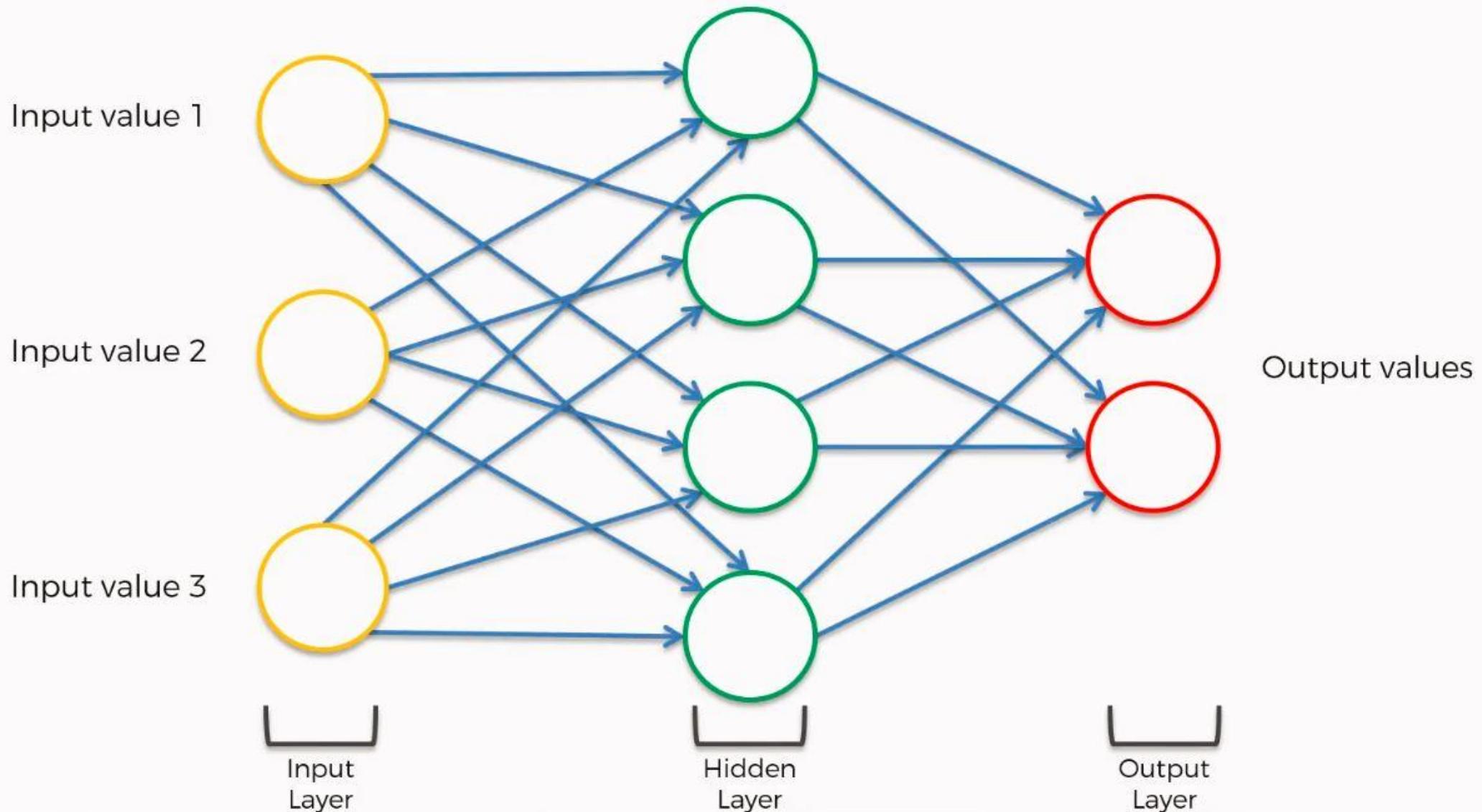
# Recurrent Neural Networks



# Recurrent Neural Networks



# Recurrent Neural Networks

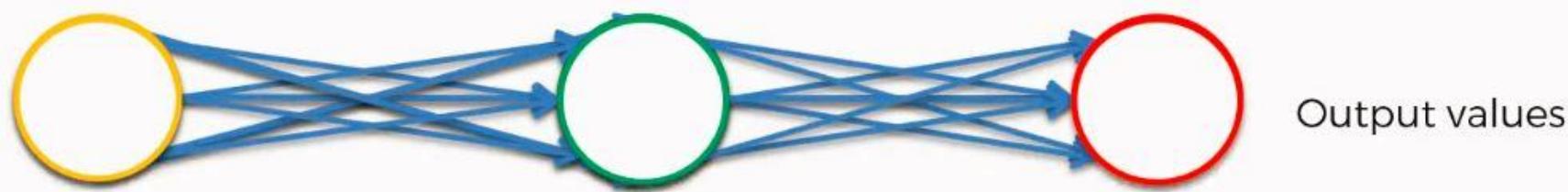


# Recurrent Neural Networks

Input value 1

Input value 2

Input value 3

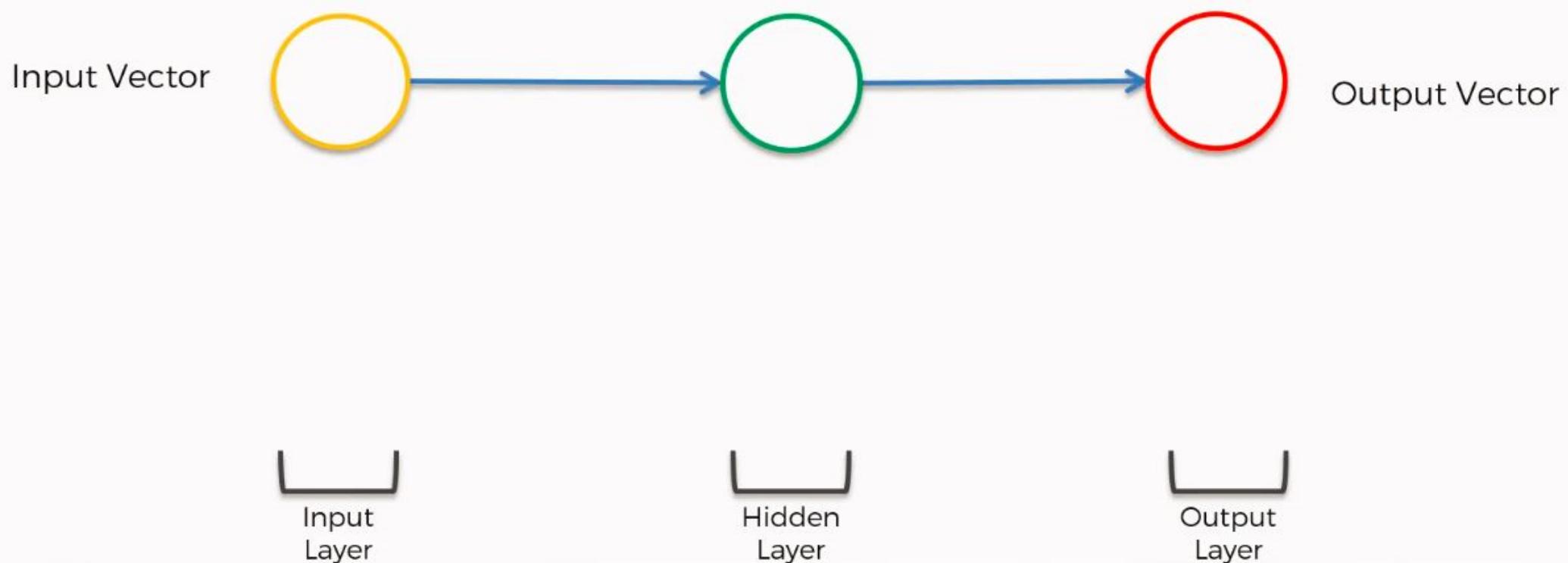


Input  
Layer

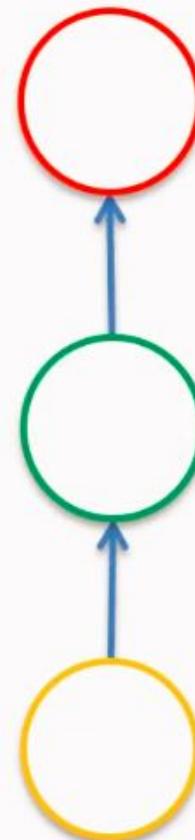
Hidden  
Layer

Output  
Layer

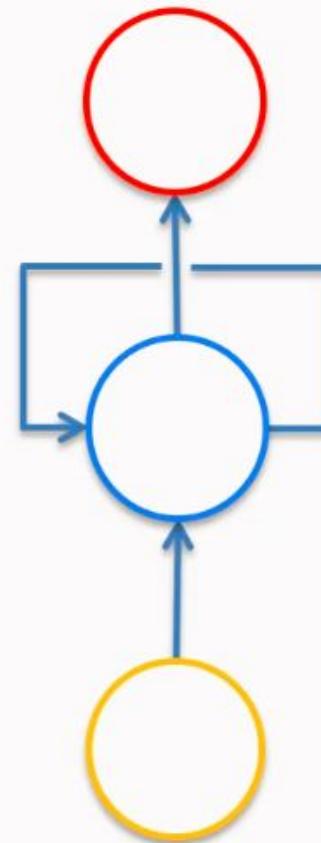
# Recurrent Neural Networks



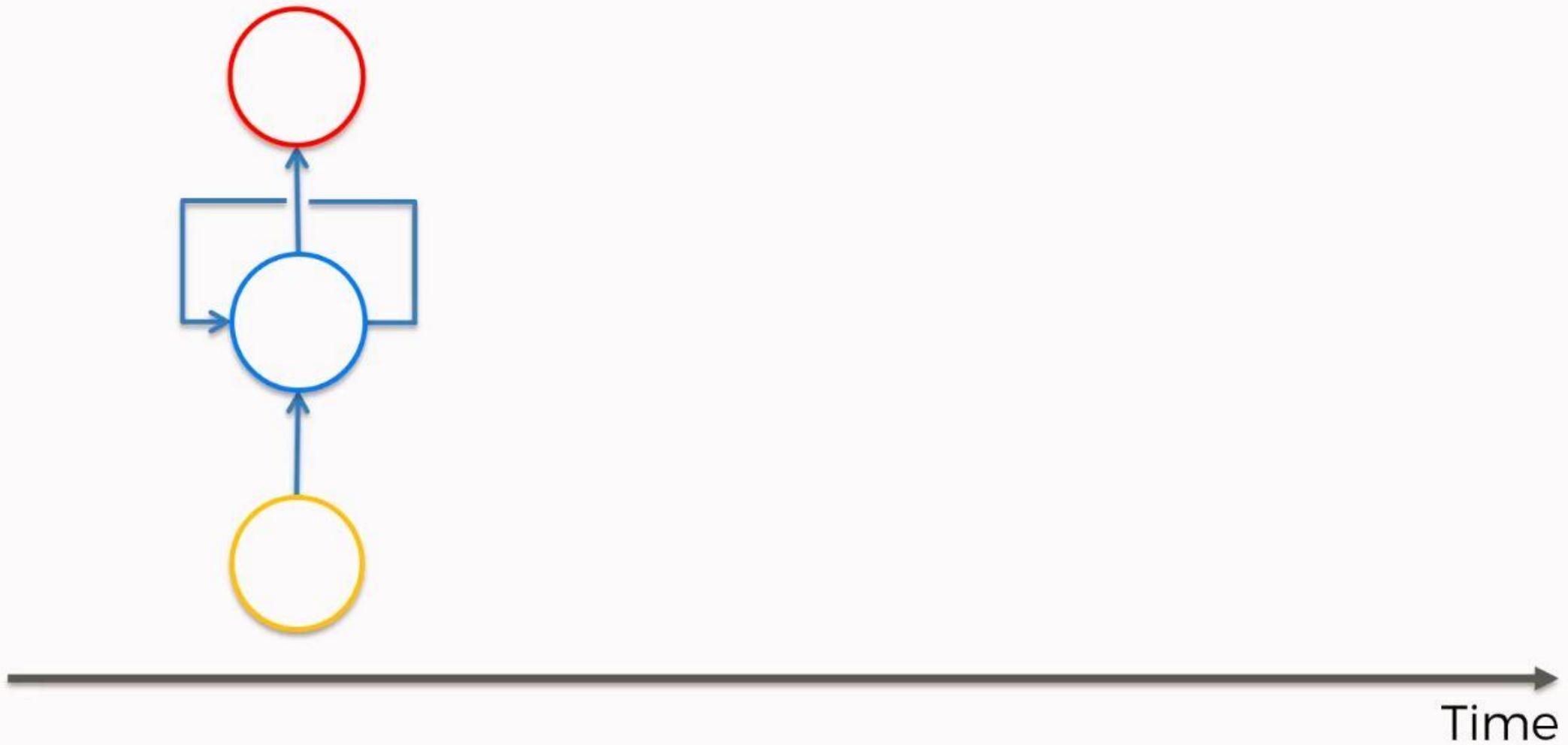
# Recurrent Neural Networks



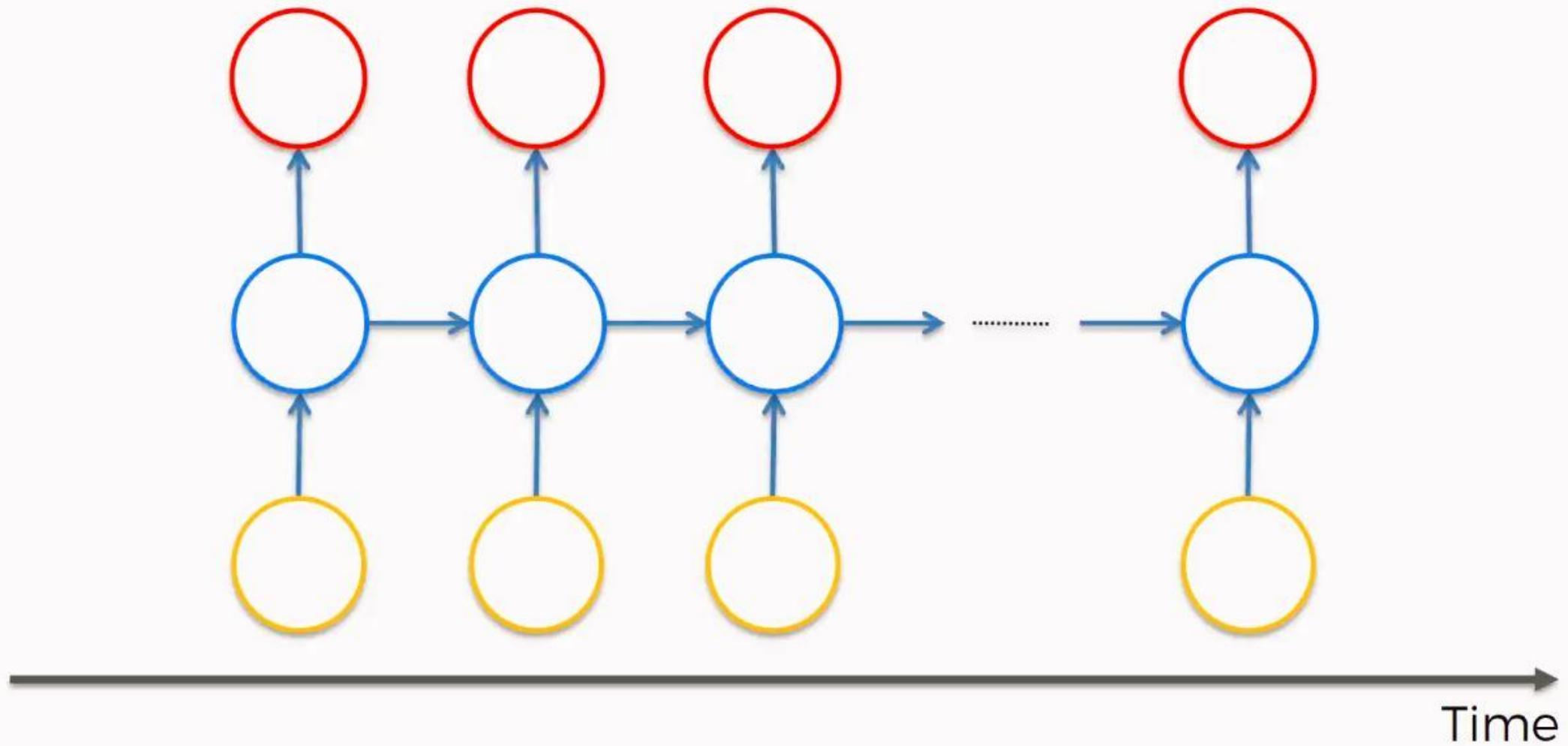
# Recurrent Neural Networks



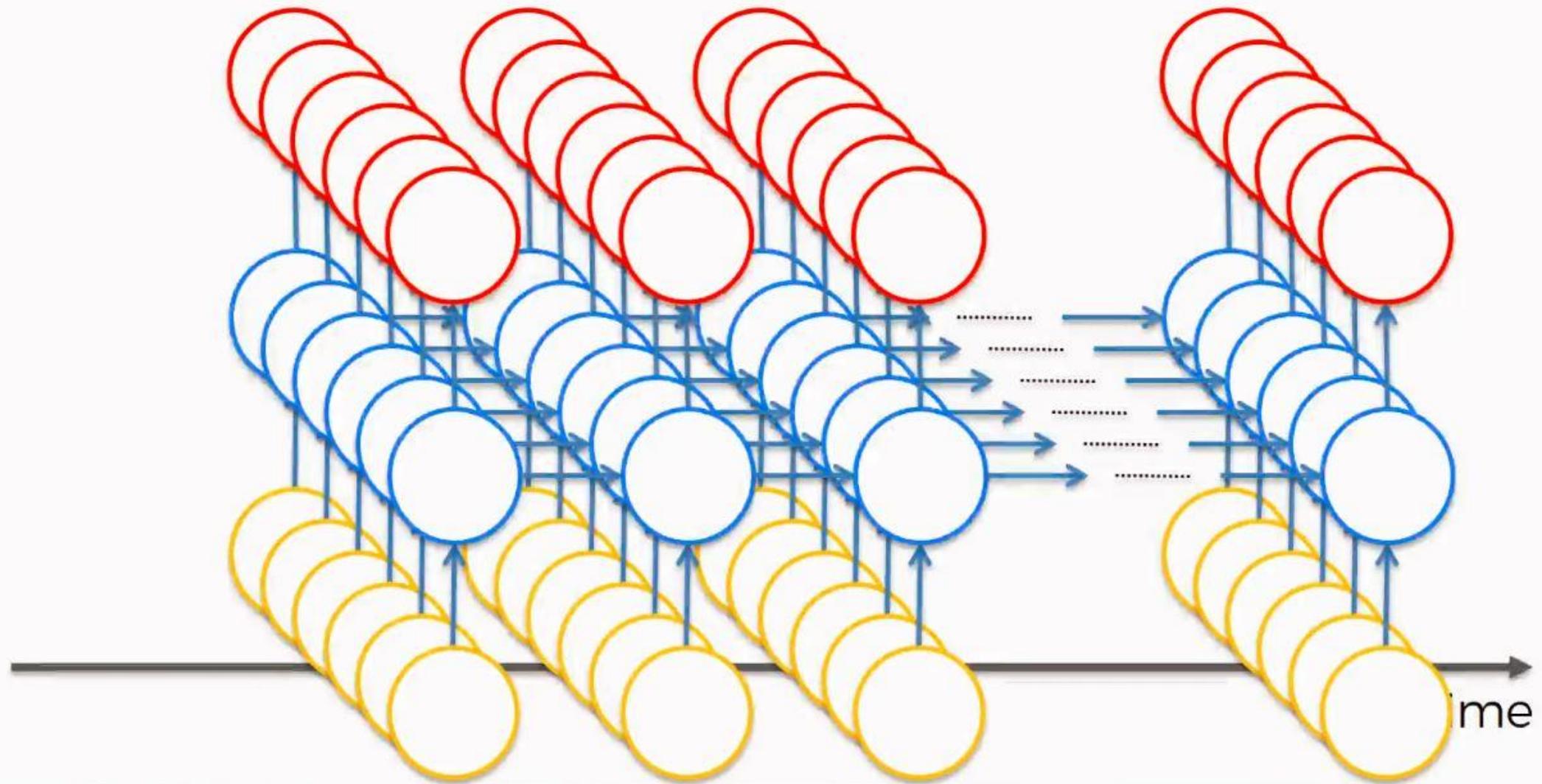
# Recurrent Neural Networks



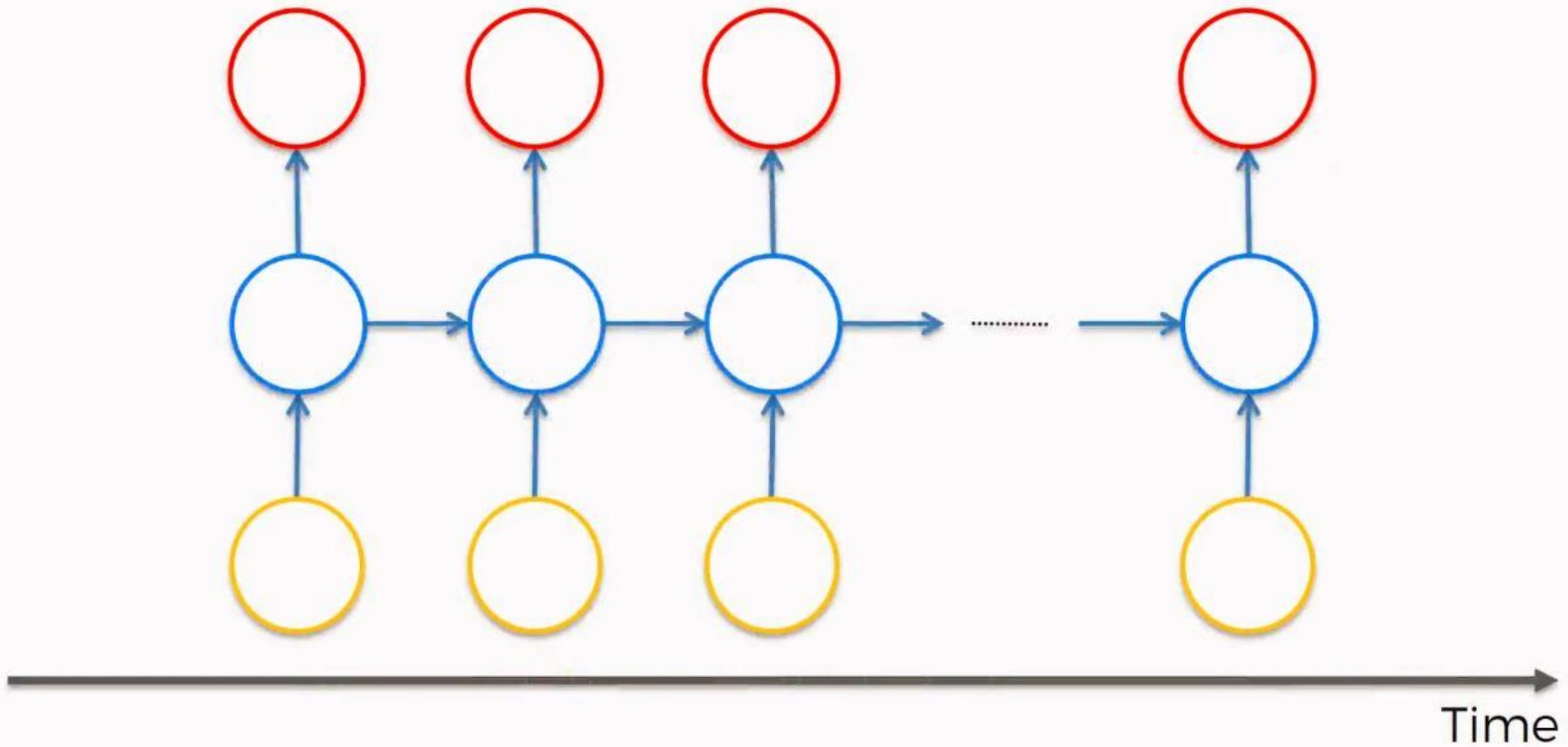
# Recurrent Neural Networks



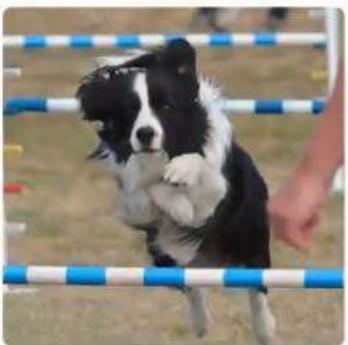
# Recurrent Neural Networks



# Recurrent Neural Networks

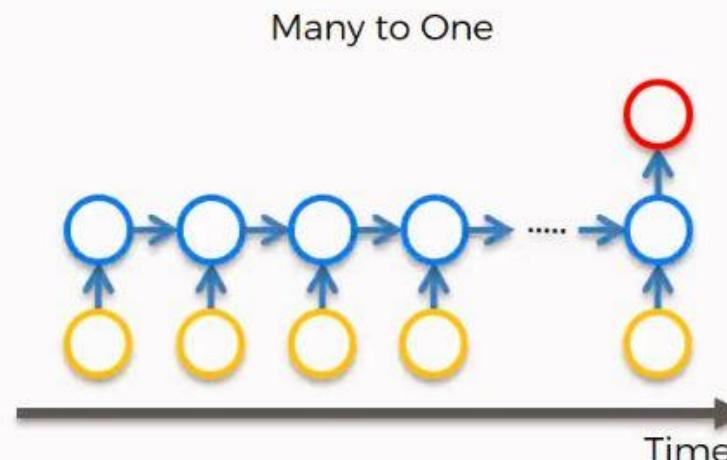
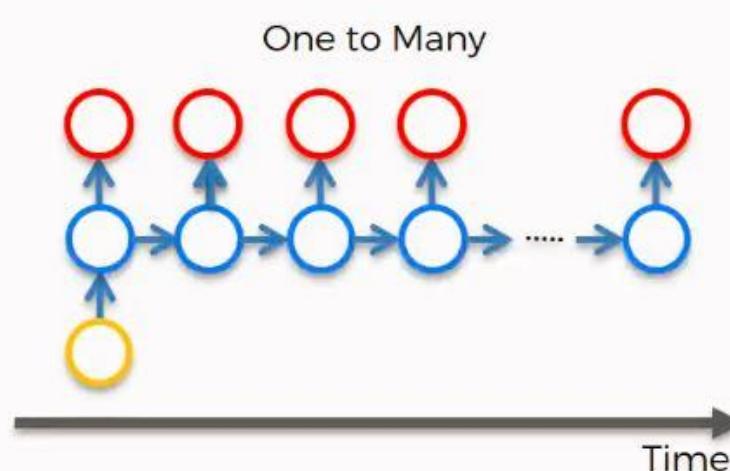


# Recurrent Neural Networks



"black and white  
dog jumps over  
bar."

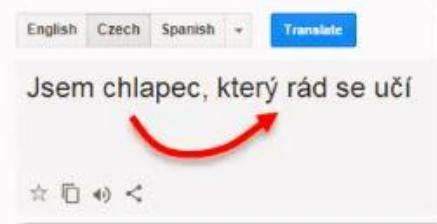
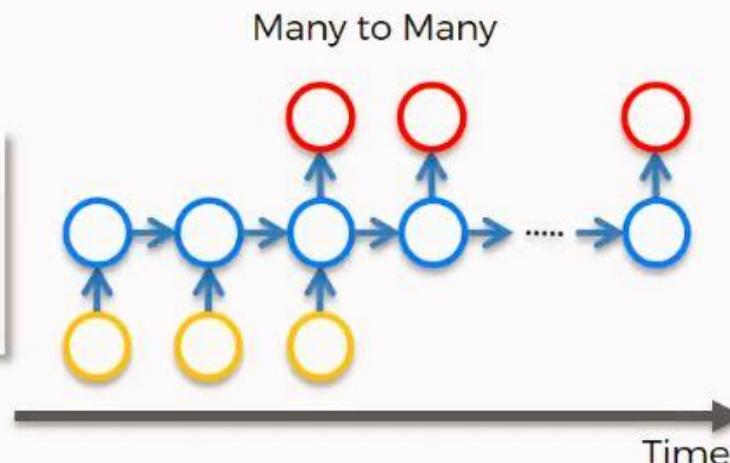
[karpathy.github.io](http://karpathy.github.io)



"Thanks for a great  
party at the  
weekend, we really  
enjoyed it!"

sentiment: positive  
score: 86%

[dev.havenondemand.com](http://dev.havenondemand.com)



Reference: [karpathy.github.io](http://karpathy.github.io)

# Recurrent Neural Networks



"black and white  
dog jumps over  
bar."

*karpathy.github.*

English Czech Spanish Translate  
Jsem chlapec, který rád

A screenshot of the Google Translate interface. The input text "I am a boy who likes to learn" is translated to "Jsem kluk, který rád učit". The interface includes language selection dropdowns (English, Czech, Spanish, Detect language), a "Translate" button, and a character counter (29/5000). Below the main window, there are links for "Google Translate for Business", "Translator Toolkit", "Website Translator", and "Global Market Finder". At the bottom, there are links for "About Google Translate", "Community", "Mobile", "G+", and "B".

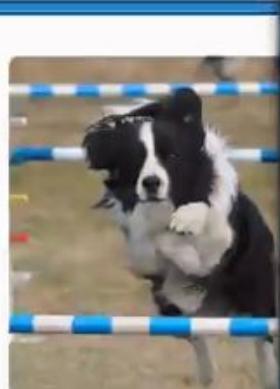
"Thanks for a great  
party at the  
weekend, we really  
enjoyed it!"

sentiment: positive  
score: 86%

dev.havenondemand.com

Reference: [karpathy.github.io](http://karpathy.github.io)

# Recurrent Neural Networks



"black and white  
dog jumps over  
bar."

*karpathy.github.*

English Czech Spanish Translate  
Jsem chlapec, který rád



Google Translate

Turn off instant translation

English Czech Spanish Detect language English Czech Spanish Translate

I am a girl who likes to learn Jsem holka, která ráda se učit

30/5000 Suggest an edit

Google Translate for Business: Translator Toolkit Website Translator Global Market Finder

About Google Translate Community Mobile G+ B

About Google Privacy & Terms Help Send feedback

"Thanks for a great  
party at the  
weekend, we really  
enjoyed it!"

sentiment: positive  
score: 86%

dev.havenondemand.com

Reference: [karpathy.github.io](http://karpathy.github.io)

Movie written by algont... X

Secure | https://arstechnica.com/the-multiverse/2016/06/an-ai-wrote-this-movie-and-its-strangely-moving/

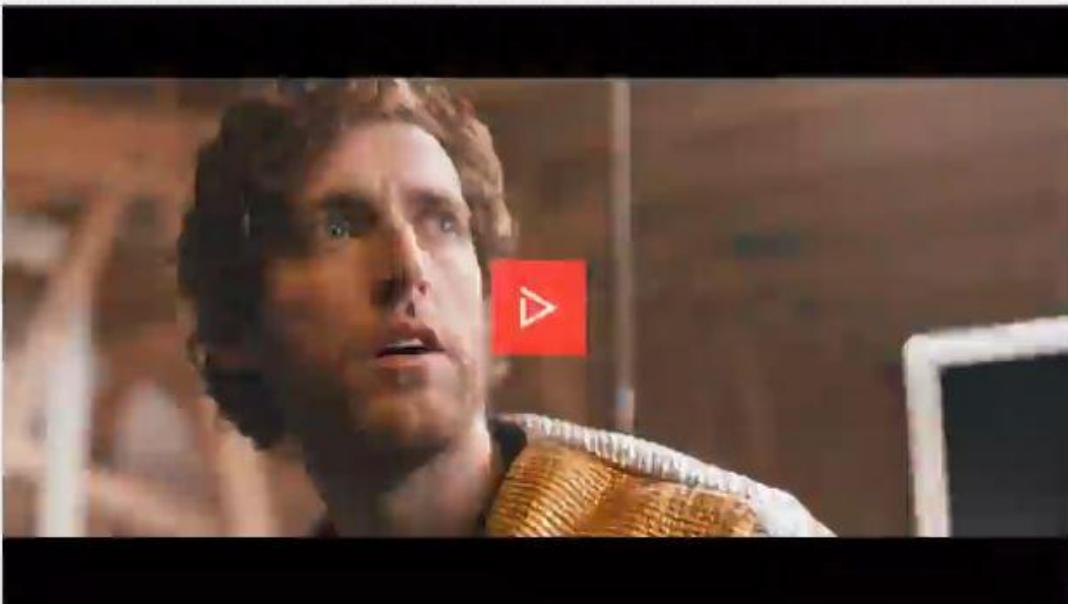
ars TECHNICA SEARCH BLOG POD SCIENCE POLICY CARS GAMING & CULTURE FORUMS SIGN IN AMERICAN FLAG

THE MULTIVERSE —

# Movie written by algorithm turns out to be hilarious and intense

For *Sunspring*'s exclusive debut on Ars, we talked to the filmmakers about collaborating with an AI.

ANNALEE NIENHUIS 6/9/2016, 8:30 PM

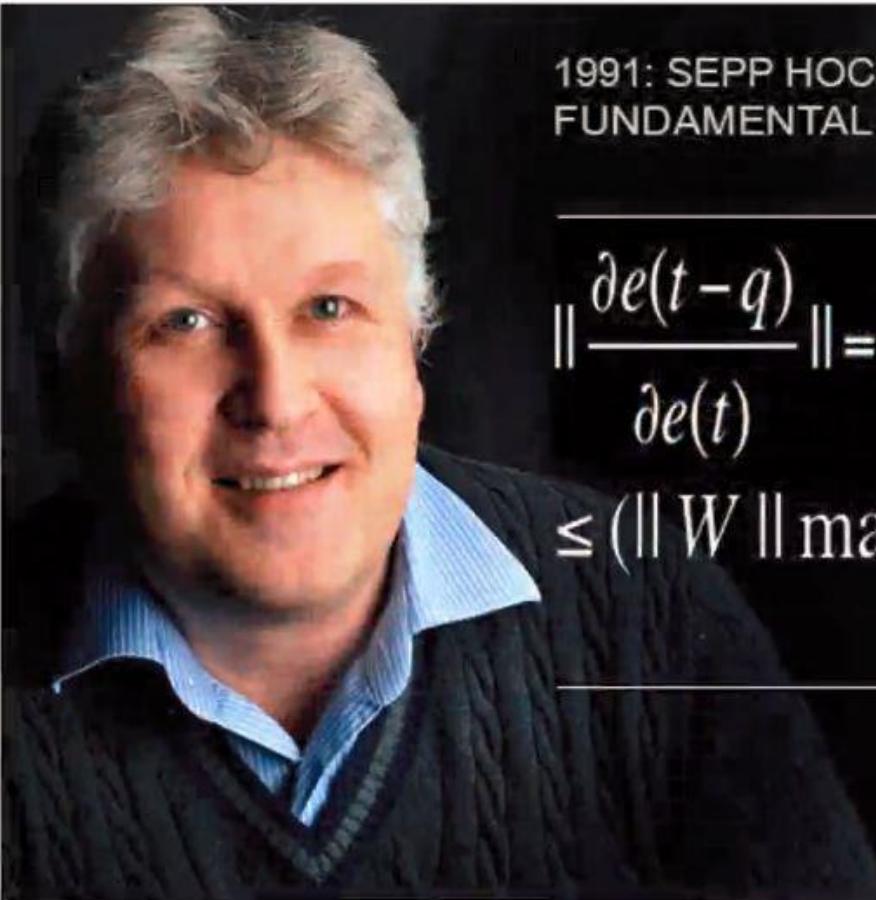


*Sunspring*, a short science fiction movie written entirely by AI, debuts exclusively on Ars today.

Ars is excited to be hosting this online debut of *Sunspring*, a short science fiction film that's not entirely what it seems. It's about three people living in a weird future, possibly on a space station, probably in a love triangle. You know it's the future because H (played with neurotic gravity by *Silicon Valley*'s Thomas Middleditch) is wearing a shiny gold jacket, H2 (Elisabeth Gray) is playing with computers, and C (Humphrey Ker) announces that he has to "go to the skull" before sticking

# Vanishing Gradient

# The Vanishing Gradient Problem

A portrait photograph of Sepp Hochreiter, a man with grey hair, wearing a dark sweater over a light blue collared shirt, smiling at the camera.

1991: SEPP HOCHREITER'S ANALYSIS OF THE FUNDAMENTAL DEEP LEARNING PROBLEM

$$\begin{aligned} \left\| \frac{\partial e(t-q)}{\partial e(t)} \right\| &= \left\| \prod_{m=1}^q W F'(\text{Net}(t-m)) \right\| \\ &\leq (\|W\| \max_{\text{Net}} \{\|F'(\text{Net})\|\})^q \end{aligned}$$

Image Source: [people.idsia.ch](http://people.idsia.ch)

# The Vanishing Gradient Problem

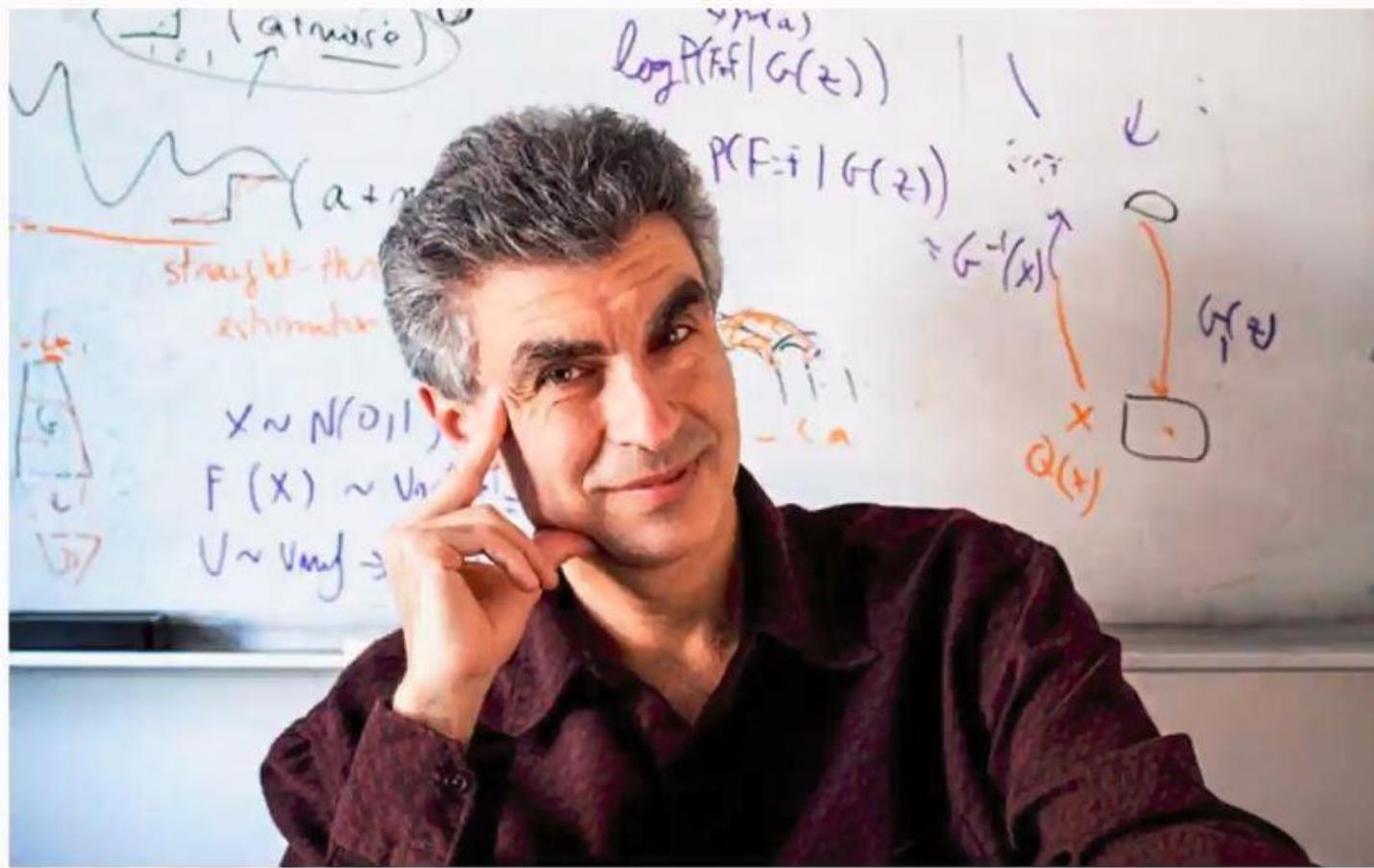


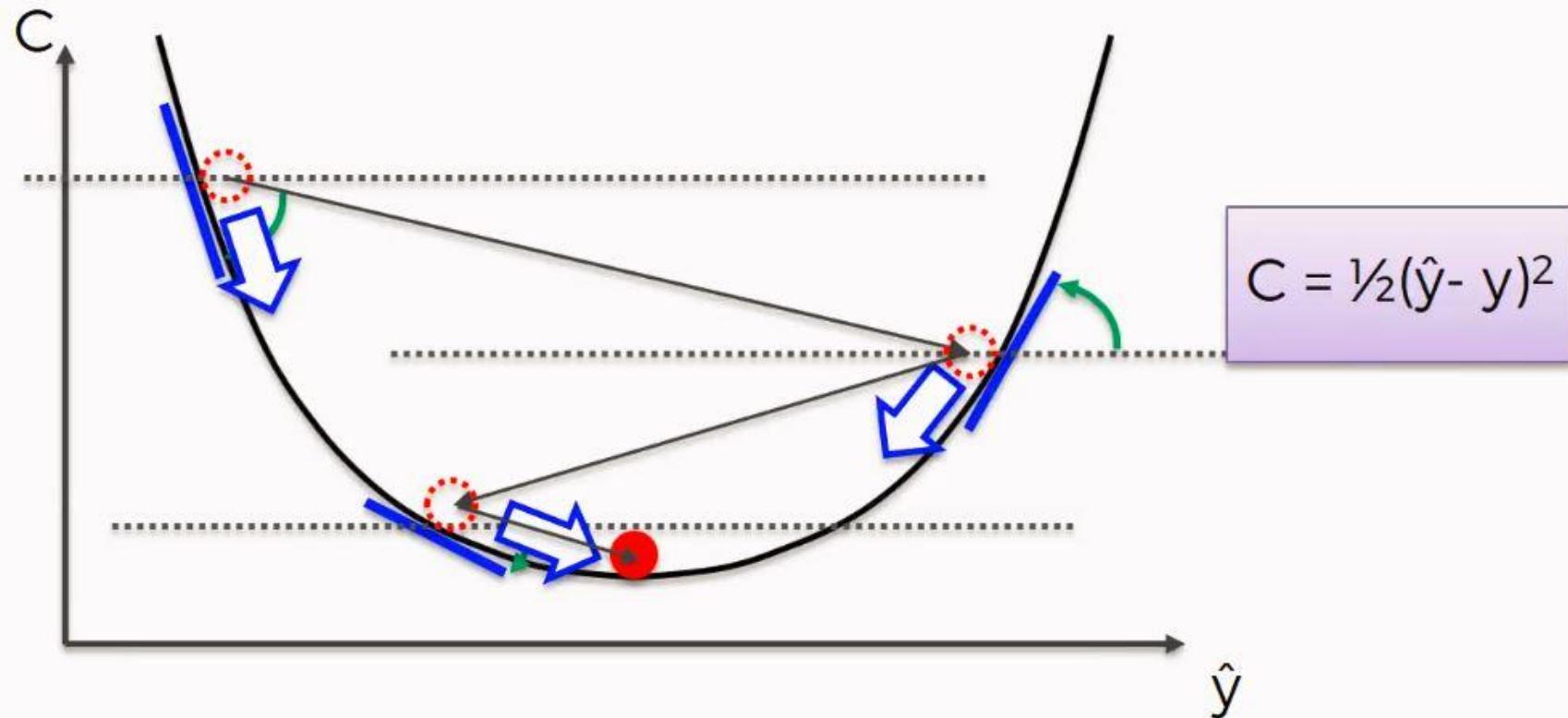
Image Source: Université Montréal

# The Vanishing Gradient Problem

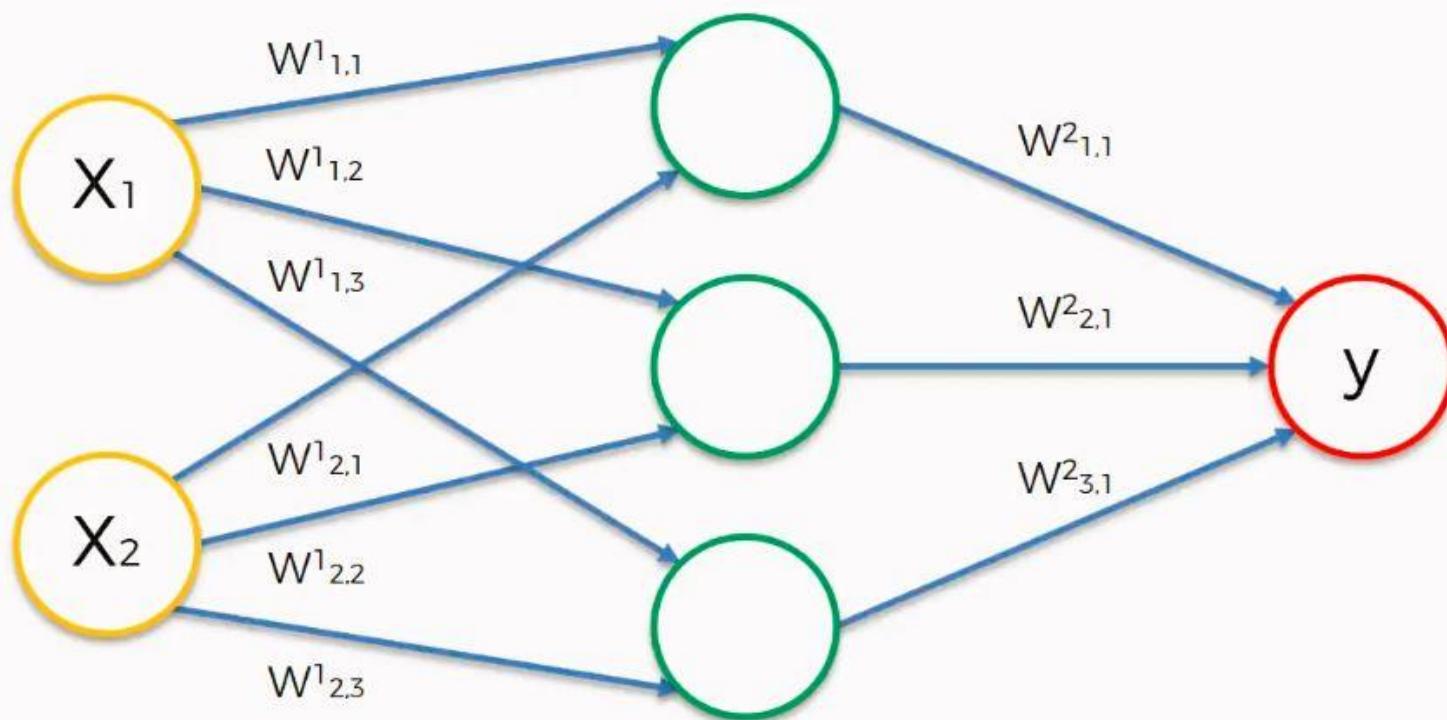


*Image Source: recode.net*

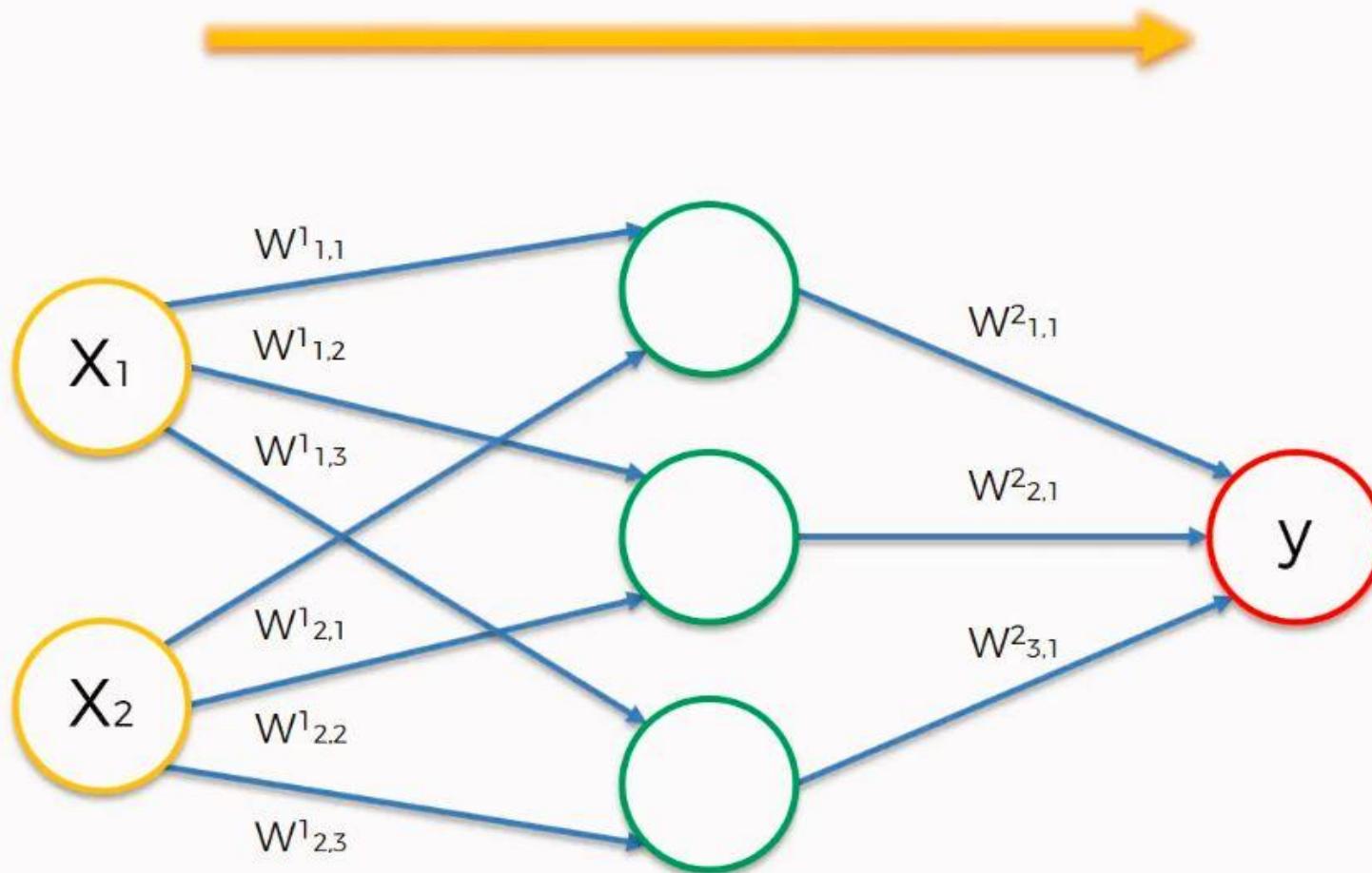
# The Vanishing Gradient Problem



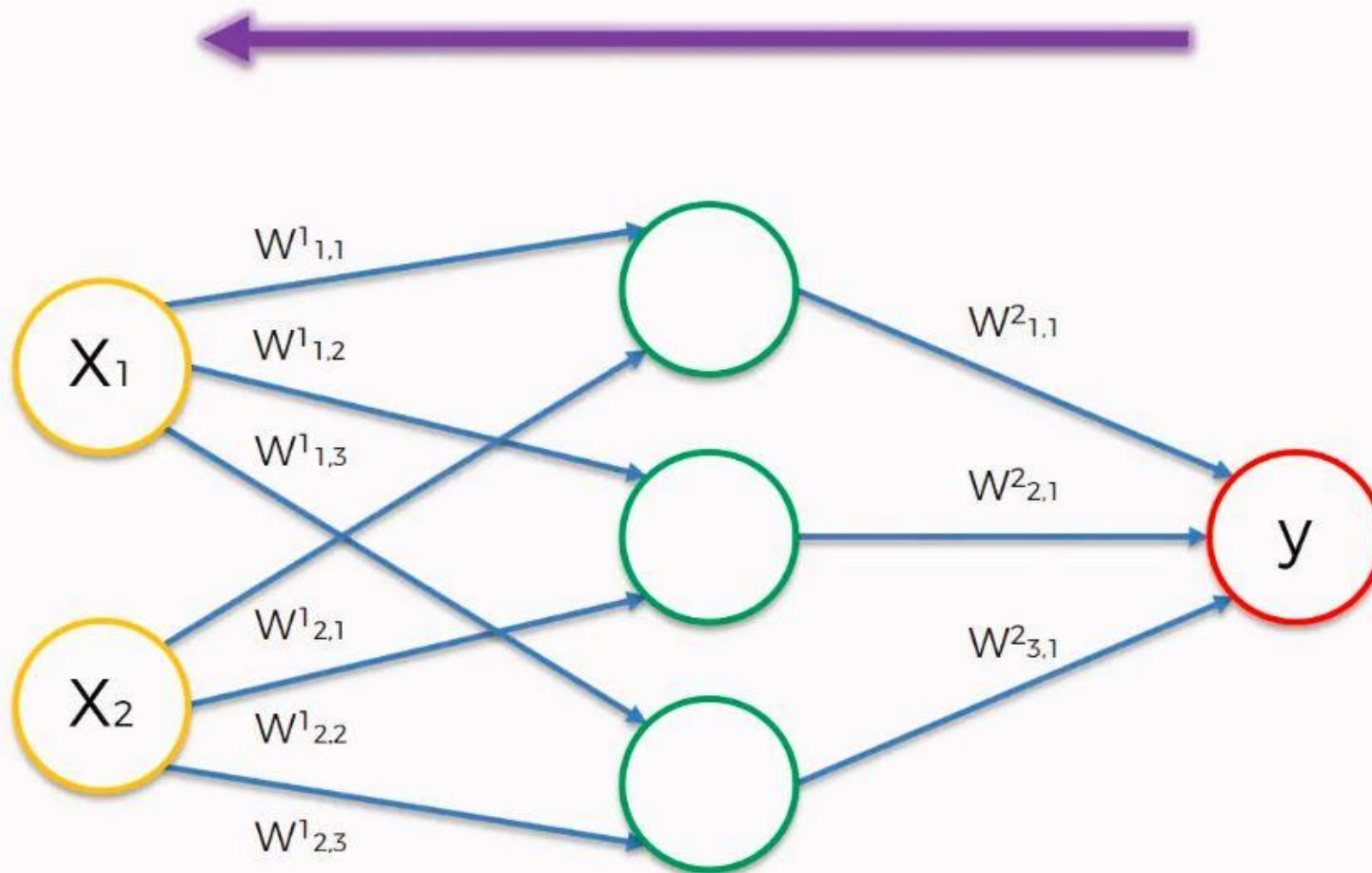
# The Vanishing Gradient Problem



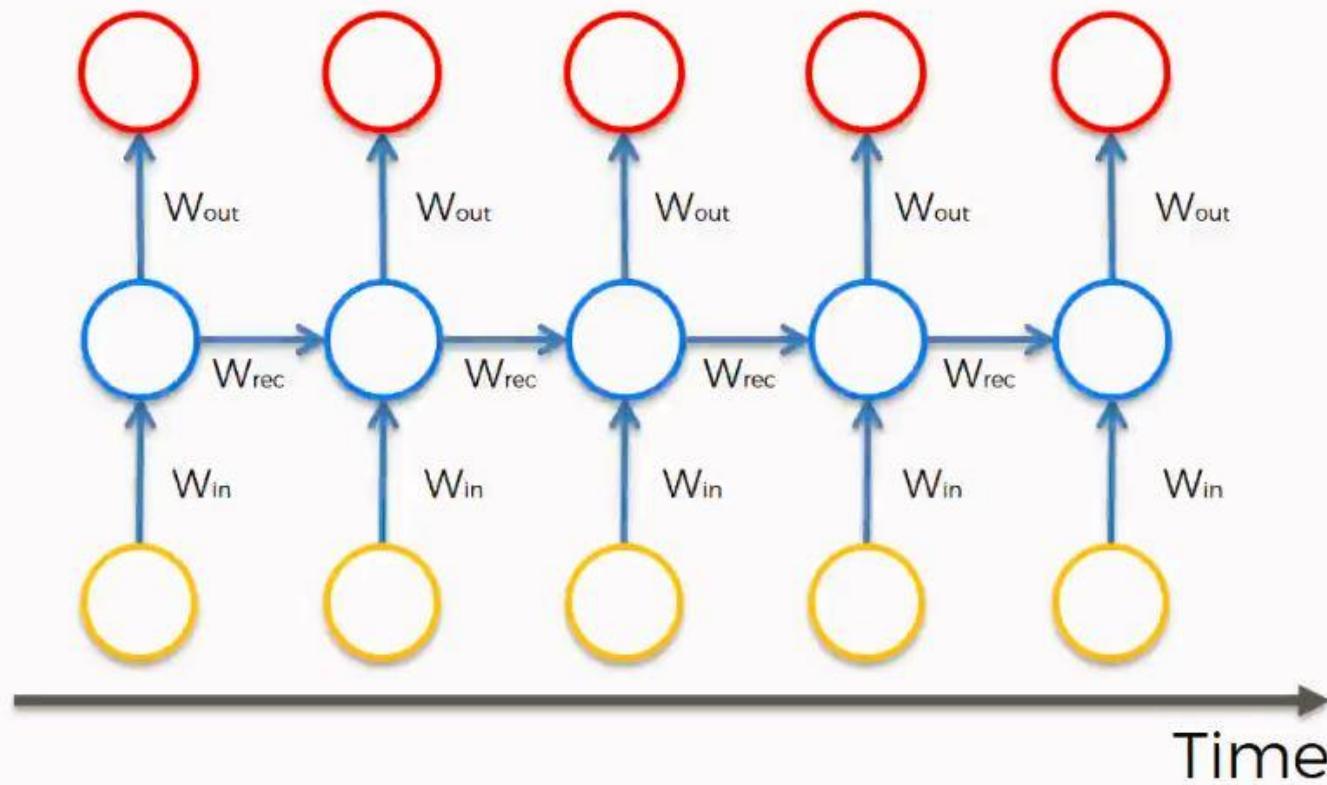
# The Vanishing Gradient Problem



# The Vanishing Gradient Problem

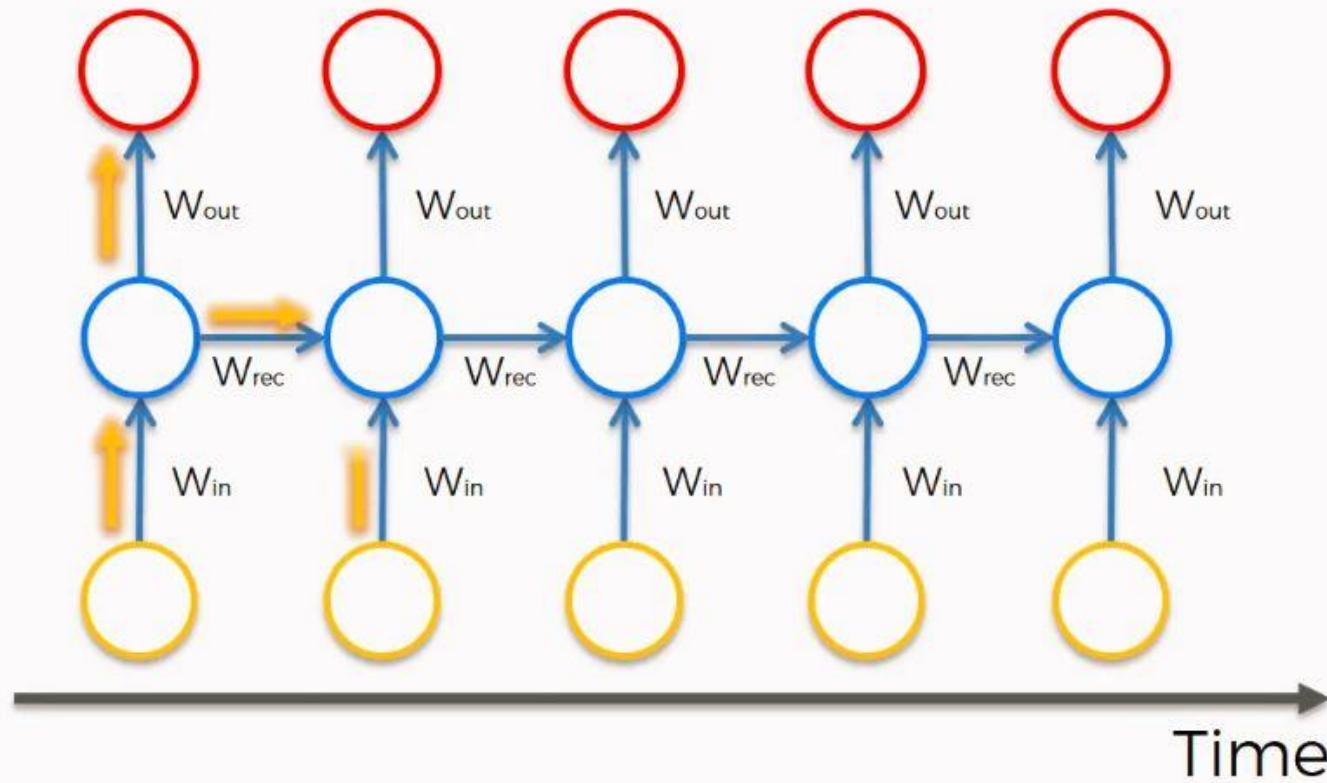


# The Vanishing Gradient Problem



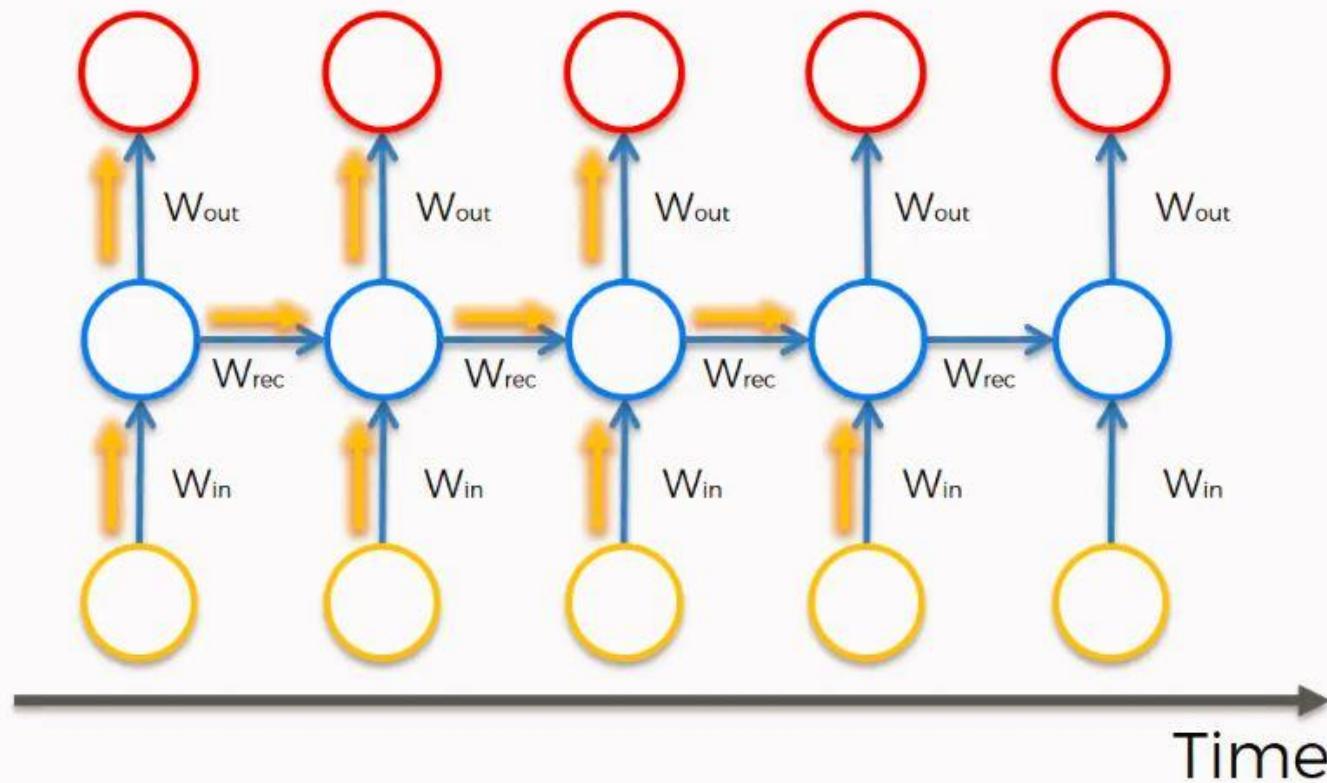
Formula Source: Razvan Pascanu et al. (2013)

# The Vanishing Gradient Problem



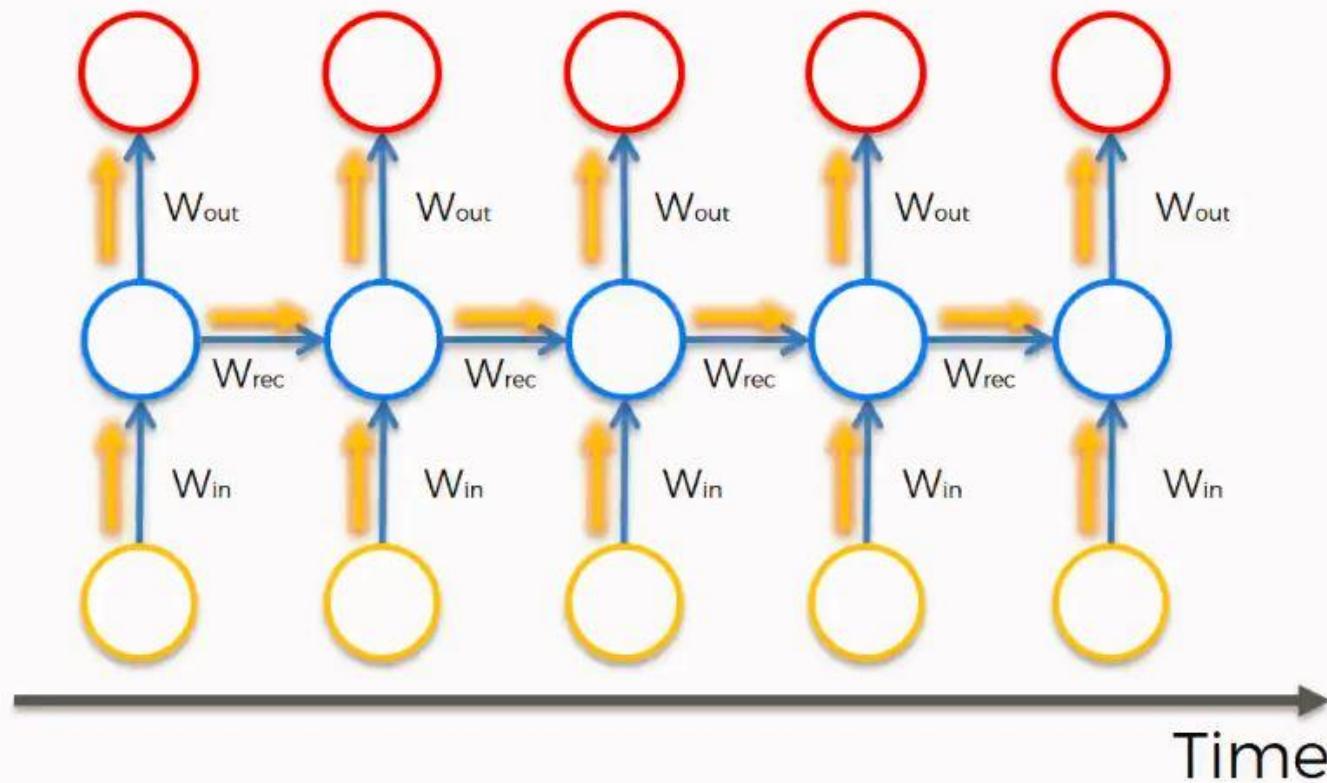
Formula Source: Razvan Pascanu et al. (2013)

# The Vanishing Gradient Problem



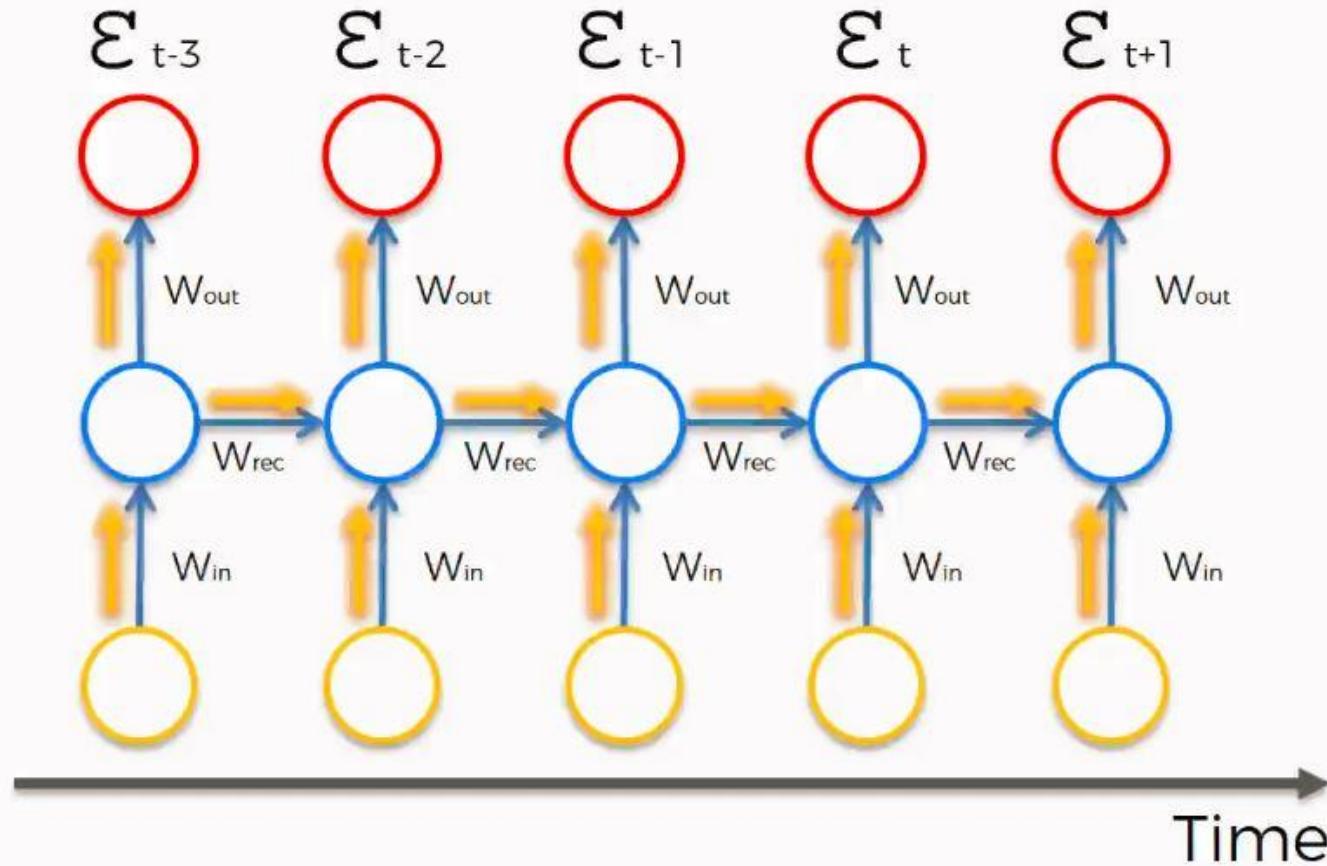
Formula Source: Razvan Pascanu et al. (2013)

# The Vanishing Gradient Problem



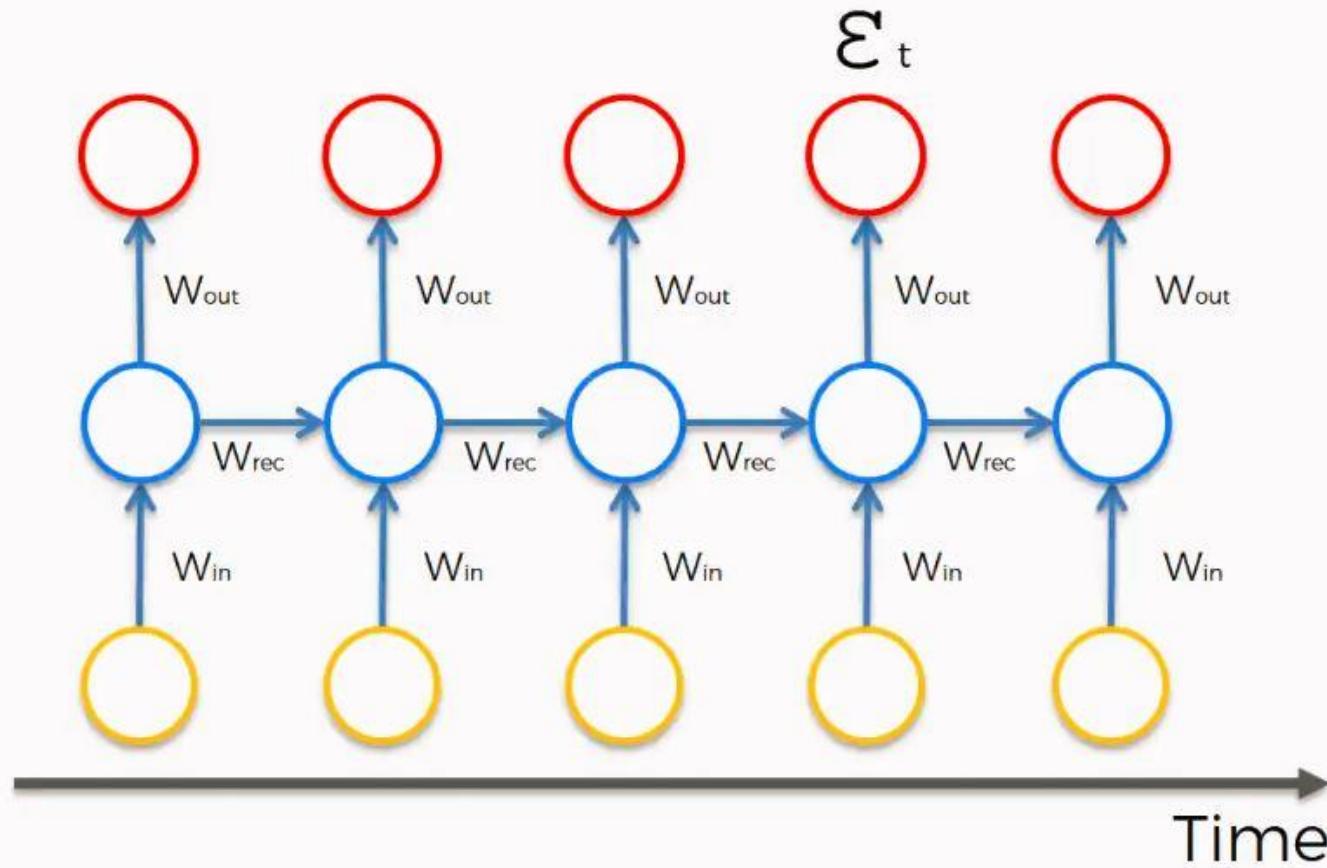
Formula Source: Razvan Pascanu et al. (2013)

# The Vanishing Gradient Problem



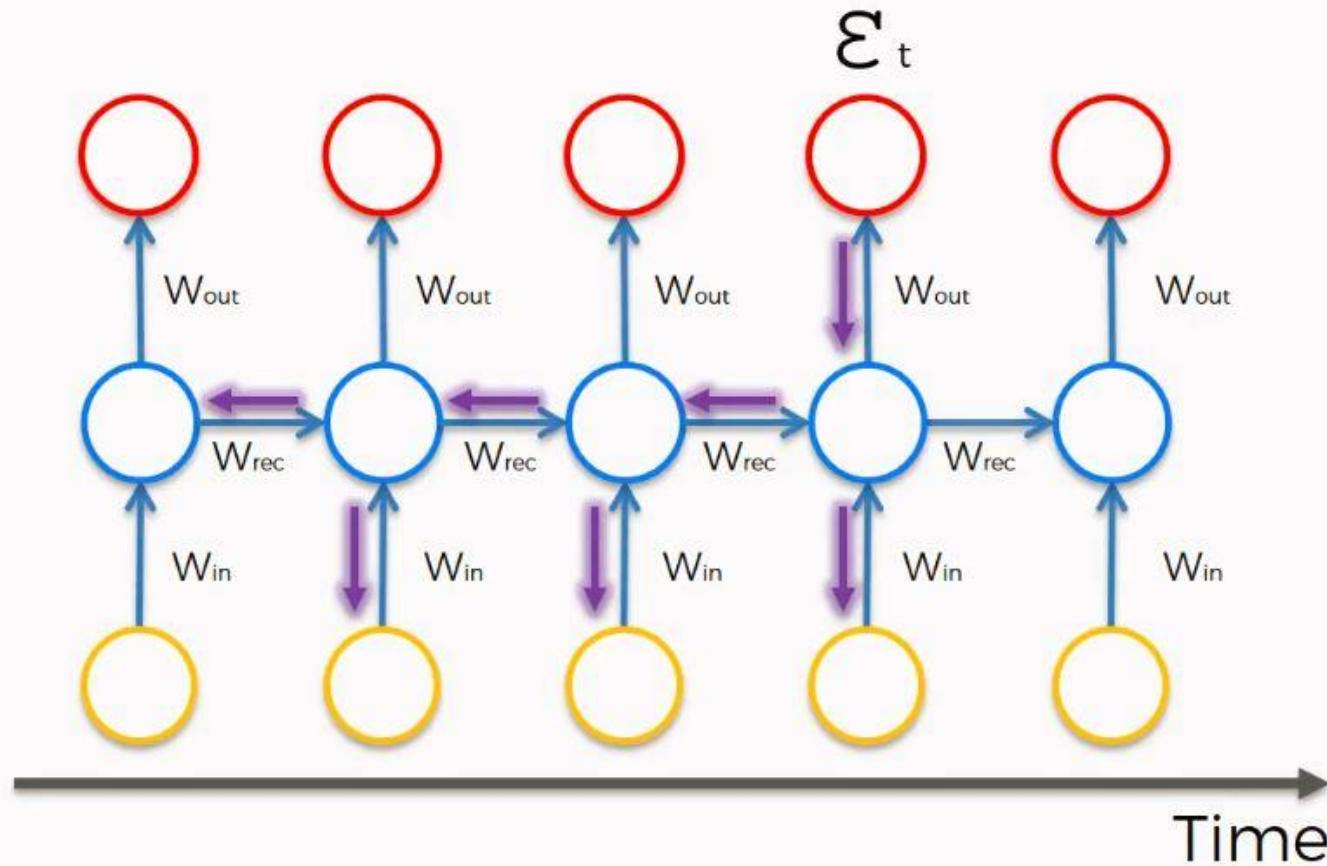
Formula Source: Razvan Pascanu et al. (2013)

# The Vanishing Gradient Problem



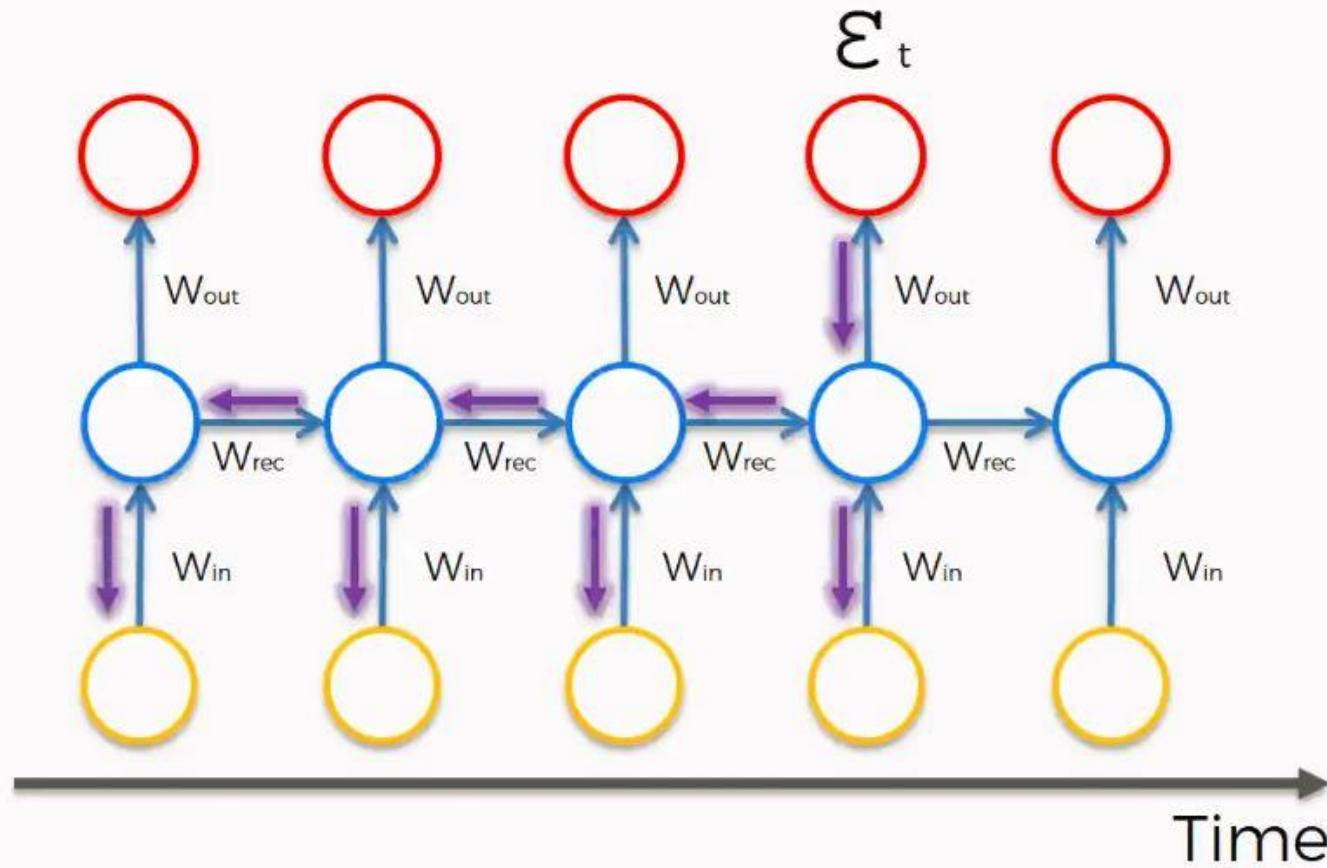
Formula Source: Razvan Pascanu et al. (2013)

# The Vanishing Gradient Problem



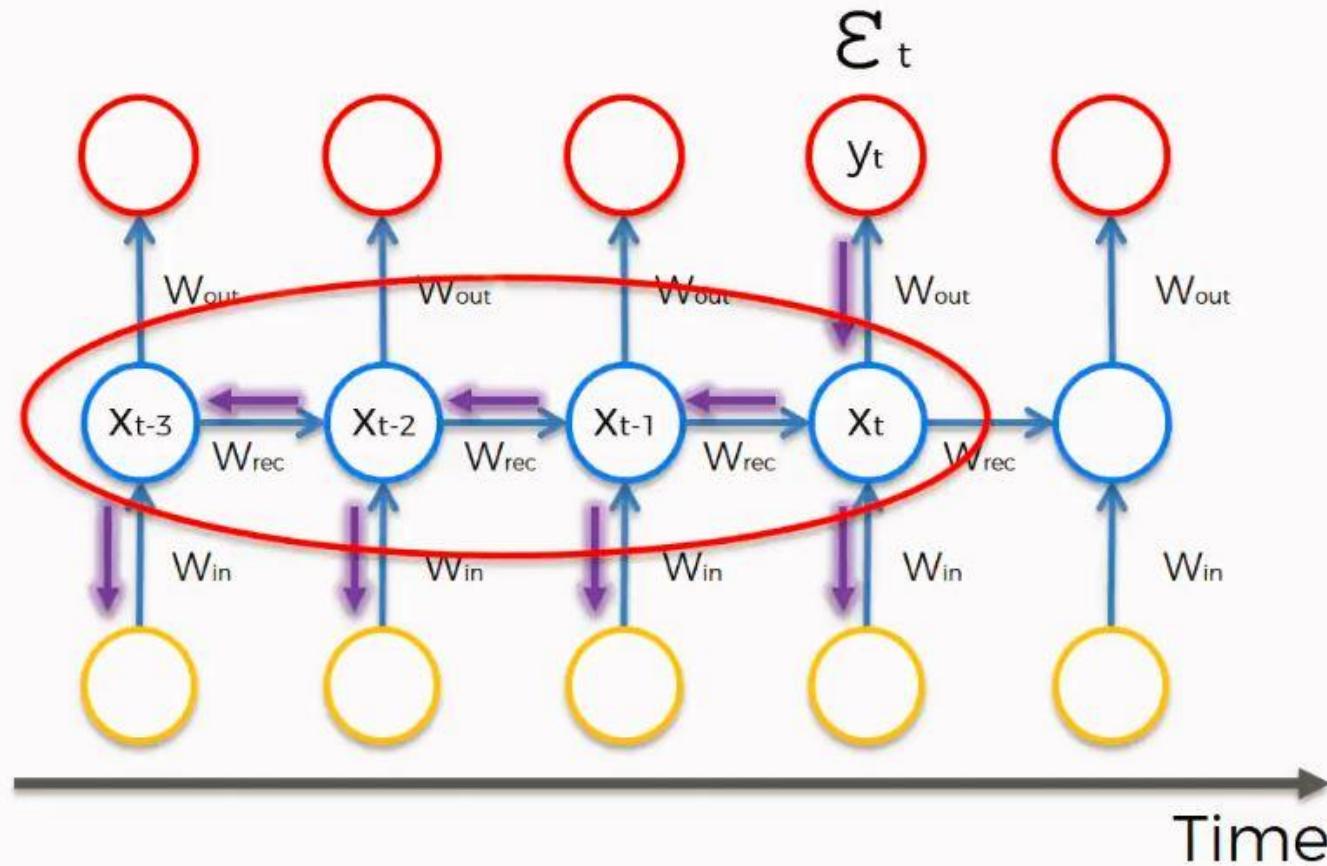
Formula Source: Razvan Pascanu et al. (2013)

# The Vanishing Gradient Problem



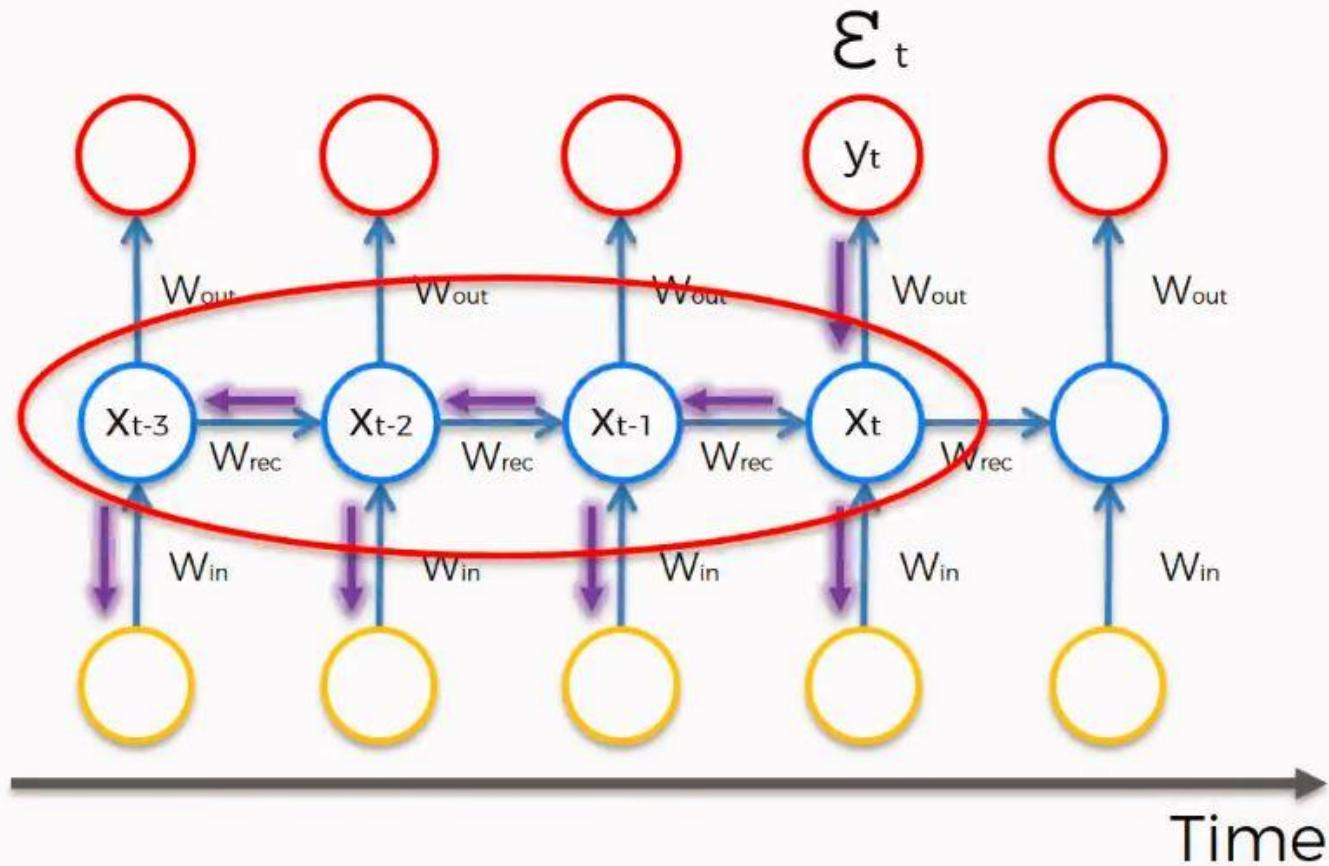
Formula Source: Razvan Pascanu et al. (2013)

# The Vanishing Gradient Problem



Formula Source: Razvan Pascanu et al. (2013)

# The Vanishing Gradient Problem



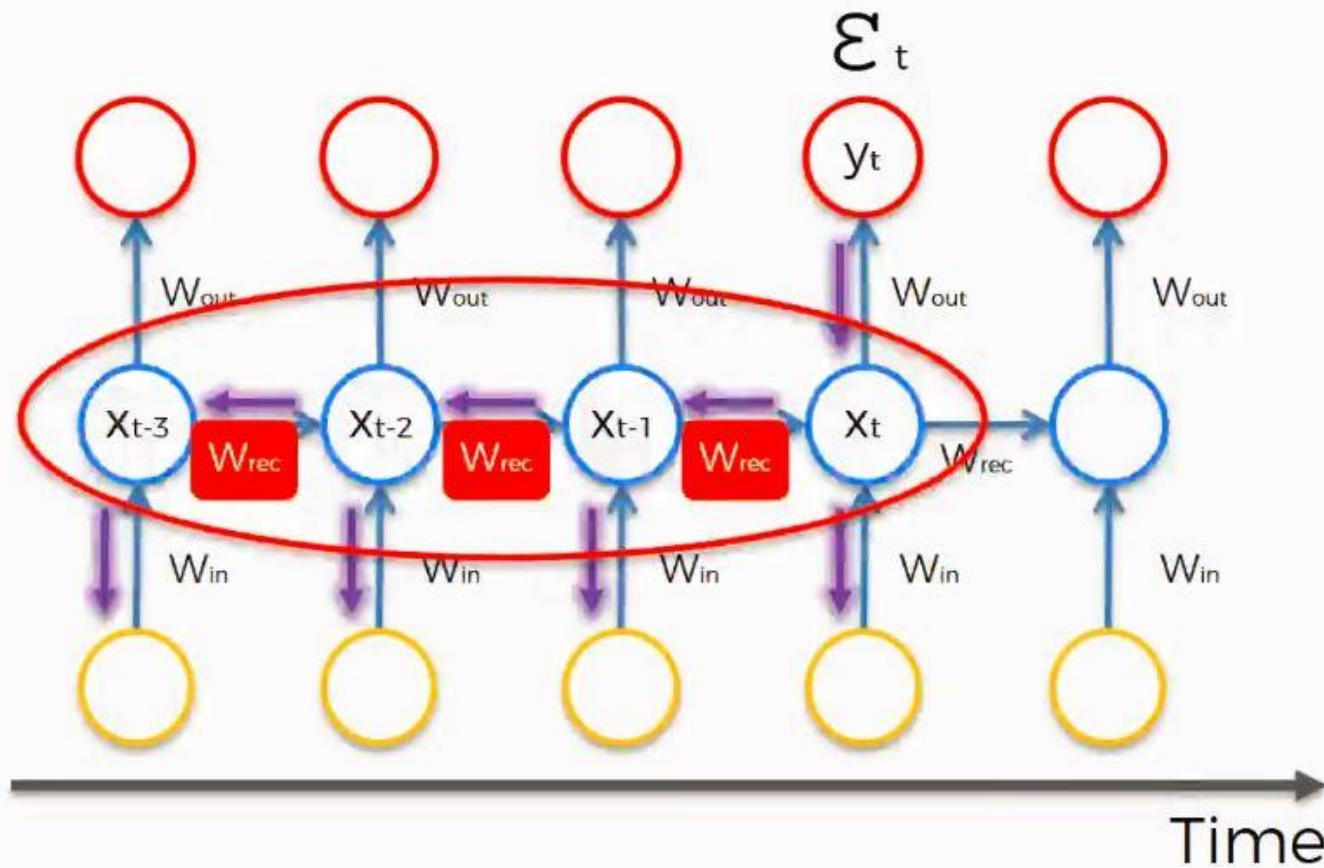
$$\frac{\partial \mathcal{E}}{\partial \theta} = \sum_{1 \leq t \leq T} \frac{\partial \mathcal{E}_t}{\partial \theta} \quad (3)$$

$$\frac{\partial \mathcal{E}_t}{\partial \theta} = \sum_{1 \leq k \leq t} \left( \frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} \frac{\partial^+ \mathbf{x}_k}{\partial \theta} \right) \quad (4)$$

$$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} = \prod_{t \geq i > k} \frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_{i-1}} = \prod_{t \geq i > k} \mathbf{W}_{rec}^T diag(\sigma'(\mathbf{x}_{i-1})) \quad (5)$$



# The Vanishing Gradient Problem



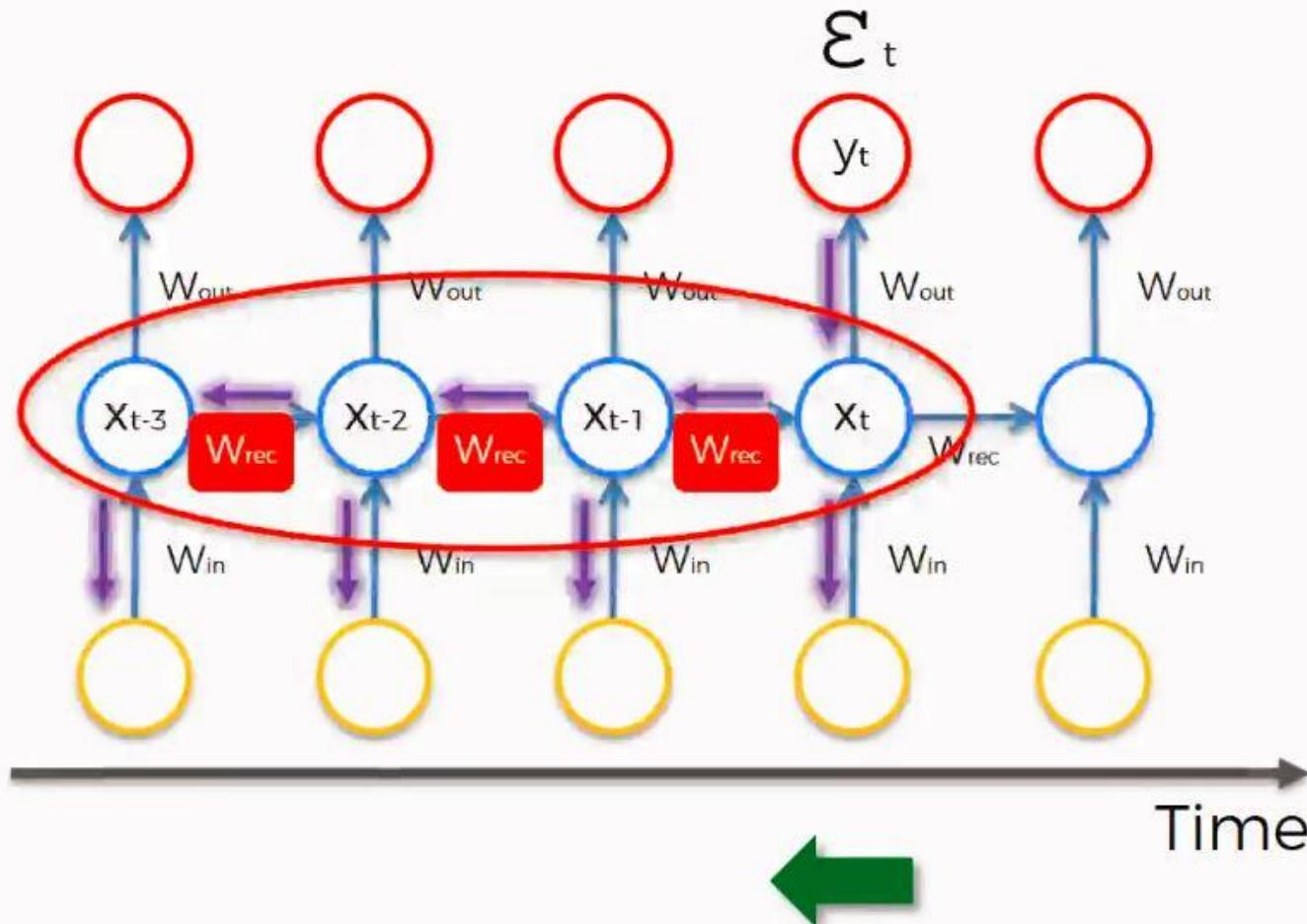
$$\frac{\partial \mathcal{E}}{\partial \theta} = \sum_{1 \leq t \leq T} \frac{\partial \mathcal{E}_t}{\partial \theta} \quad (3)$$

$$\frac{\partial \mathcal{E}_t}{\partial \theta} = \sum_{1 \leq k \leq t} \left( \frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} \frac{\partial^+ \mathbf{x}_k}{\partial \theta} \right) \quad (4)$$

$$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} = \prod_{i \geq i > k} \frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_{i-1}} = \prod_{i \geq i > k} \mathbf{W}_{rec}^T diag(\sigma'(\mathbf{x}_{i-1})) \quad (5)$$



# The Vanishing Gradient Problem



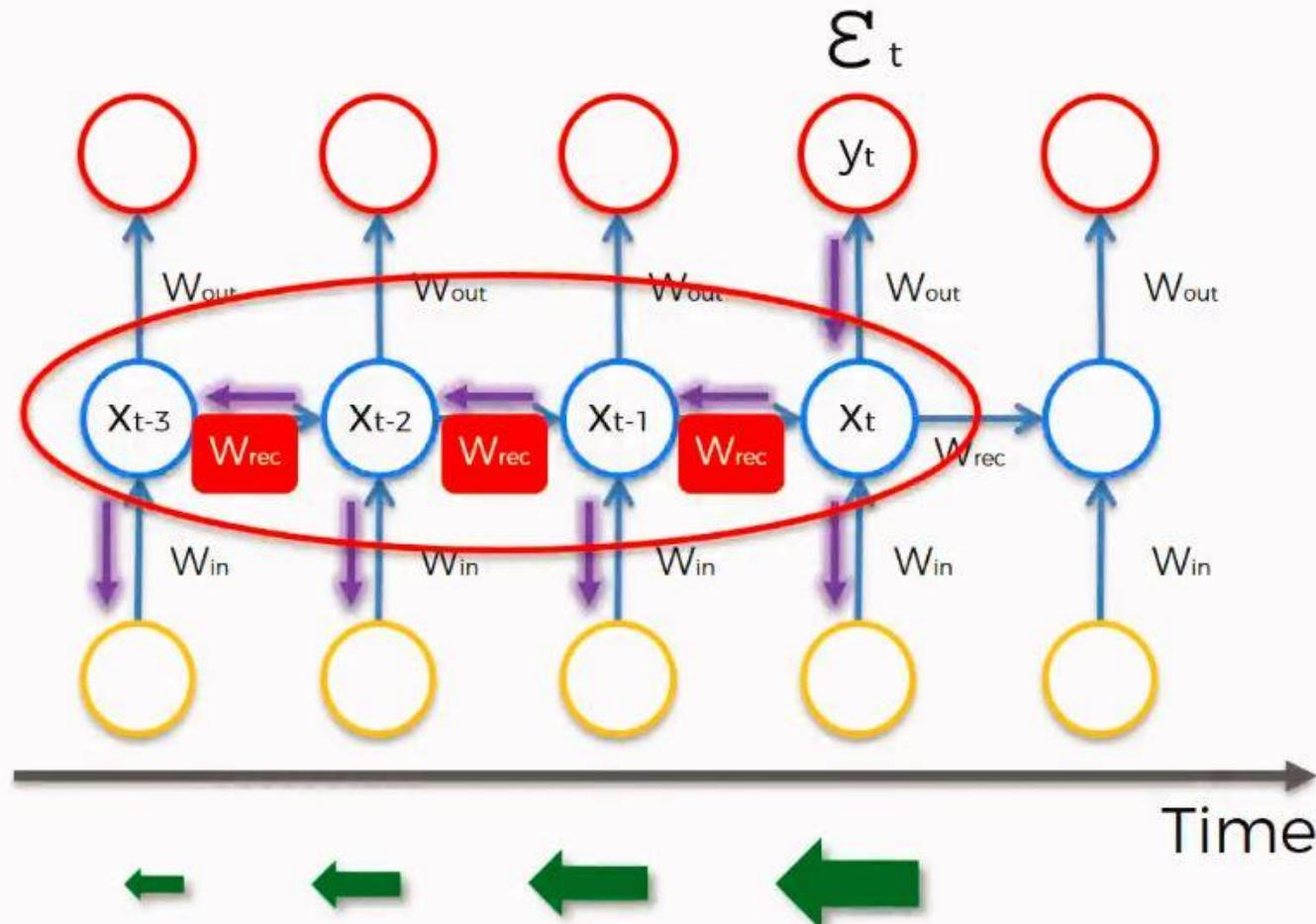
$$\frac{\partial \mathcal{E}}{\partial \theta} = \sum_{1 \leq t \leq T} \frac{\partial \mathcal{E}_t}{\partial \theta} \quad (3)$$

$$\frac{\partial \mathcal{E}_t}{\partial \theta} = \sum_{1 \leq k \leq t} \left( \frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} \frac{\partial^+ \mathbf{x}_k}{\partial \theta} \right) \quad (4)$$

$$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} = \prod_{i \geq i > k} \frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_{i-1}} = \prod_{i \geq i > k} \mathbf{W}_{rec}^T diag(\sigma'(\mathbf{x}_{i-1})) \quad (5)$$



# The Vanishing Gradient Problem



$$\frac{\partial \mathcal{E}}{\partial \theta} = \sum_{1 \leq t \leq T} \frac{\partial \mathcal{E}_t}{\partial \theta} \quad (3)$$

$$\frac{\partial \mathcal{E}_t}{\partial \theta} = \sum_{1 \leq k \leq t} \left( \frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} \frac{\partial^+ \mathbf{x}_k}{\partial \theta} \right) \quad (4)$$

$$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} = \prod_{i \geq i > k} \frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_{i-1}} = \prod_{i \geq i > k} \mathbf{W}_{rec}^T \text{diag}(\sigma'(\mathbf{x}_{i-1})) \quad (5)$$

$W_{rec} \sim \text{small}$  Vanishing  
 $W_{rec} \sim \text{large}$  Exploding

# The Vanishing Gradient Problem

Solutions:

## 1. Exploding Gradient

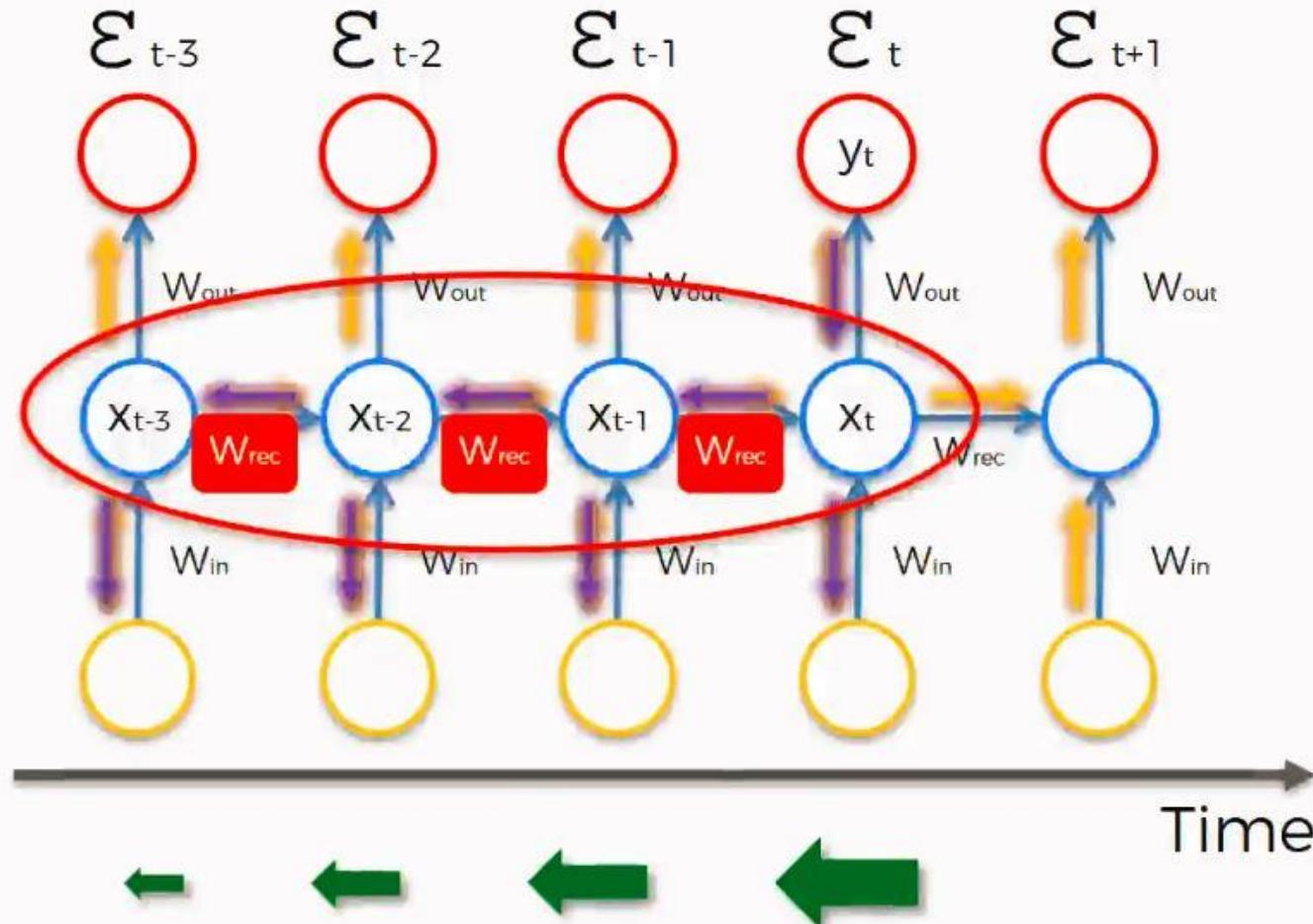
- Truncated Backpropagation
- Penalties
- Gradient Clipping

## 2. Vanishing Gradient

- Weight Initialization
- Echo State Networks
- Long Short-Term Memory Networks (LSTMs)



# Long Short-Term Memory



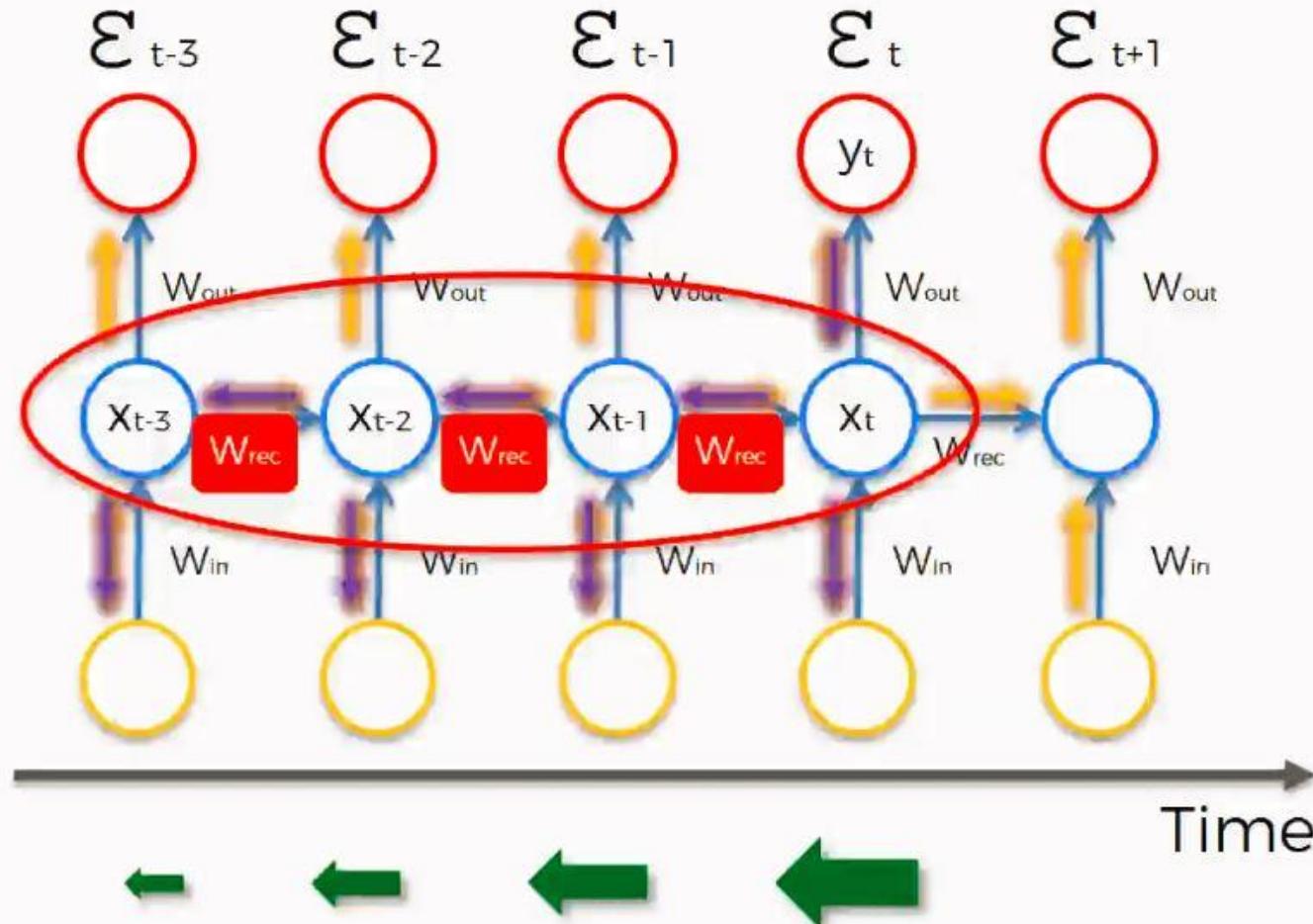
$$\frac{\partial \mathcal{E}}{\partial \theta} = \sum_{1 \leq t \leq T} \frac{\partial \mathcal{E}_t}{\partial \theta} \quad (3)$$

$$\frac{\partial \mathcal{E}_t}{\partial \theta} = \sum_{1 \leq k \leq t} \left( \frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} \frac{\partial^+ \mathbf{x}_k}{\partial \theta} \right) \quad (4)$$

$$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} = \prod_{i \geq i > k} \frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_{i-1}} = \prod_{i \geq i > k} \mathbf{W}_{rec}^T diag(\sigma'(\mathbf{x}_{i-1})) \quad (5)$$

$W_{rec} \sim \text{small}$  Vanishing  
 $W_{rec} \sim \text{large}$  Exploding

# Long Short-Term Memory



$$\frac{\partial \mathcal{E}}{\partial \theta} = \sum_{1 \leq t \leq T} \frac{\partial \mathcal{E}_t}{\partial \theta} \quad (3)$$

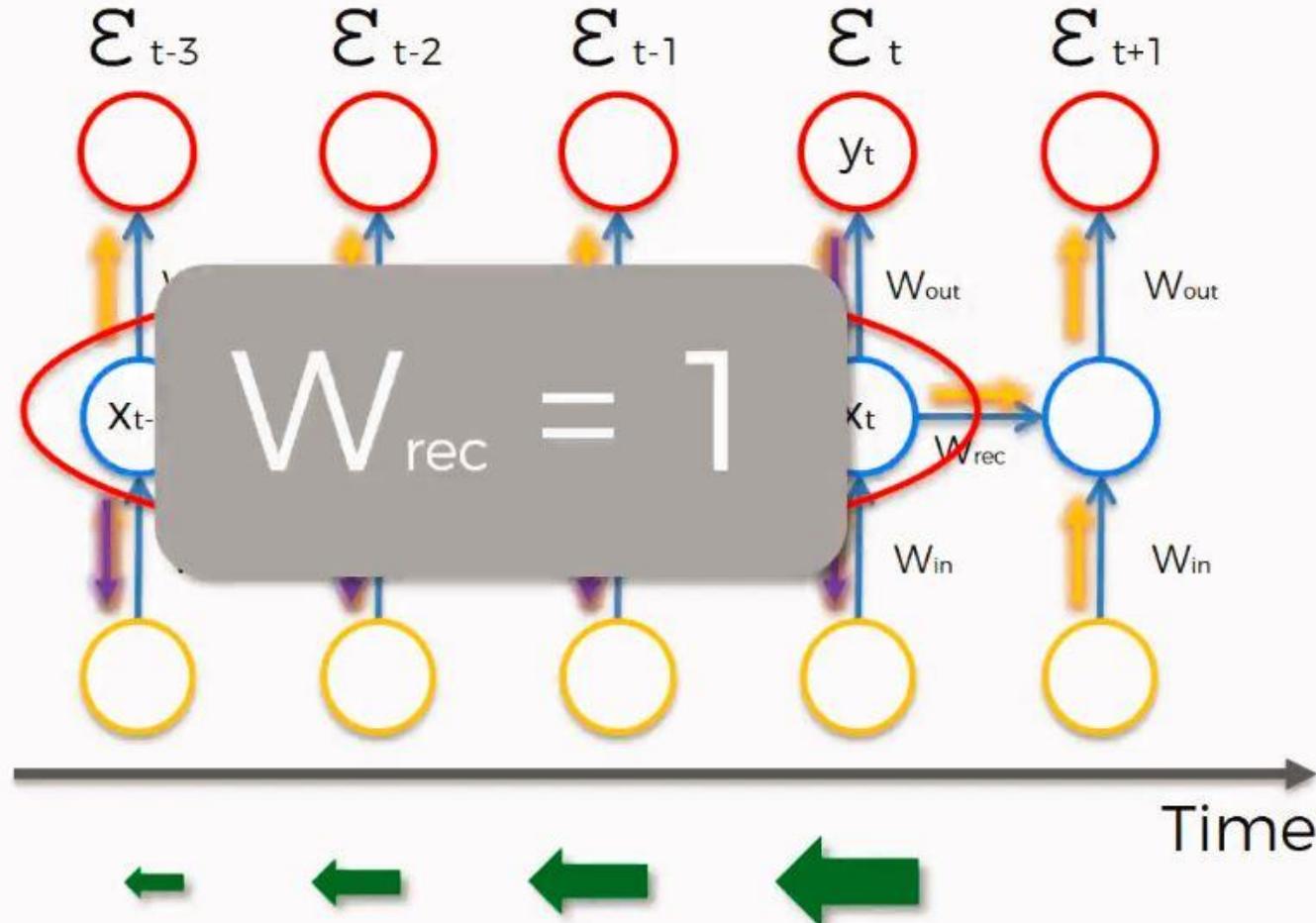
$$\frac{\partial \mathcal{E}_t}{\partial \theta} = \sum_{1 \leq k \leq t} \left( \frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} \frac{\partial^+ \mathbf{x}_k}{\partial \theta} \right) \quad (4)$$

$$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} = \prod_{i \geq i > k} \frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_{i-1}} = \prod_{i \geq i > k} \mathbf{W}_{rec}^T diag(\sigma'(\mathbf{x}_{i-1})) \quad (5)$$

$W_{rec} < 1$	→	Vanishing
$W_{rec} > 1$	→	Exploding

Formula Source: Razvan Pascanu et al. (2013)

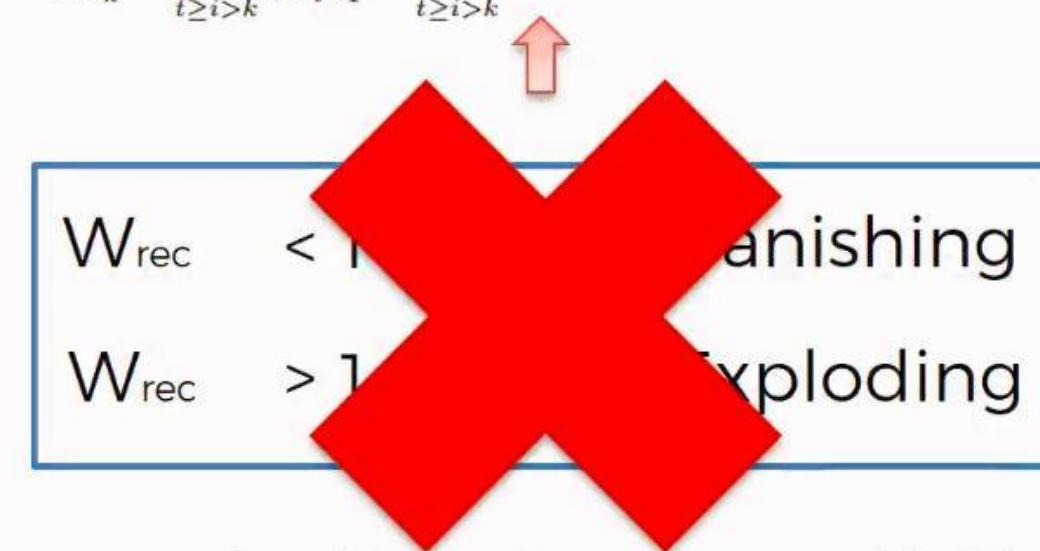
# Long Short-Term Memory



$$\frac{\partial \mathcal{E}}{\partial \theta} = \sum_{1 \leq t \leq T} \frac{\partial \mathcal{E}_t}{\partial \theta} \quad (3)$$

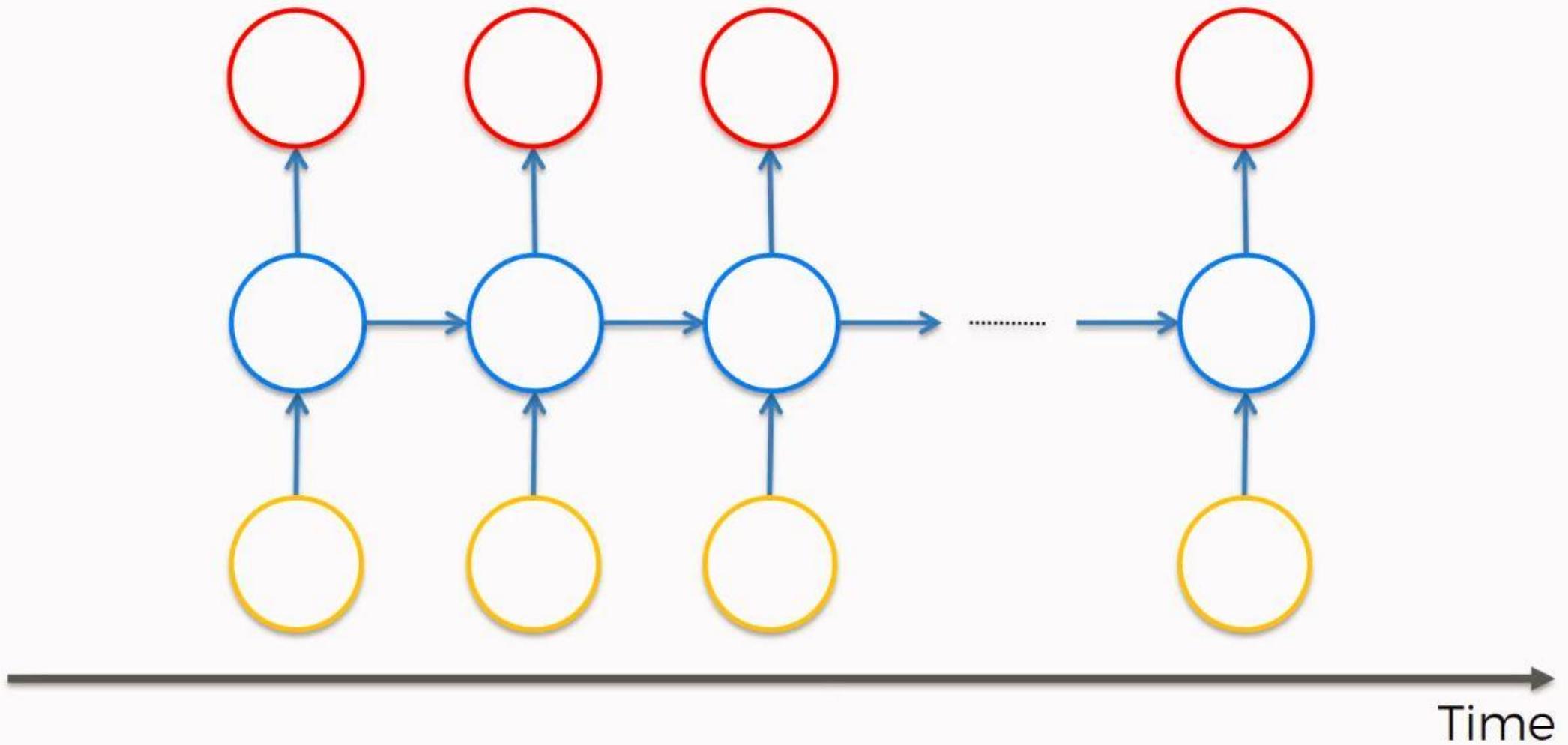
$$\frac{\partial \mathcal{E}_t}{\partial \theta} = \sum_{1 \leq k \leq t} \left( \frac{\partial \mathcal{E}_t}{\partial \mathbf{x}_t} \frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} \frac{\partial^+ \mathbf{x}_k}{\partial \theta} \right) \quad (4)$$

$$\frac{\partial \mathbf{x}_t}{\partial \mathbf{x}_k} = \prod_{t \geq i > k} \frac{\partial \mathbf{x}_i}{\partial \mathbf{x}_{i-1}} = \prod_{t \geq i > k} \mathbf{W}_{rec}^T diag(\sigma'(\mathbf{x}_{i-1})) \quad (5)$$

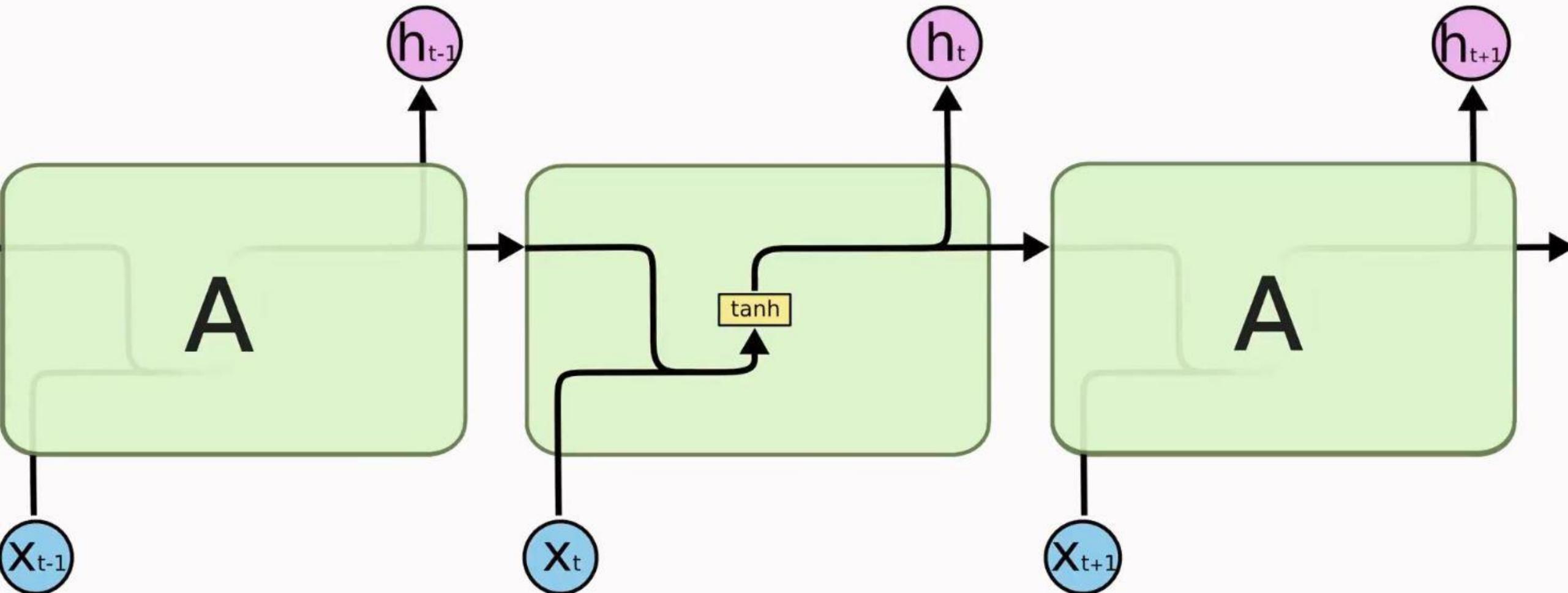


Formula Source: Razvan Pascanu et al. (2013)

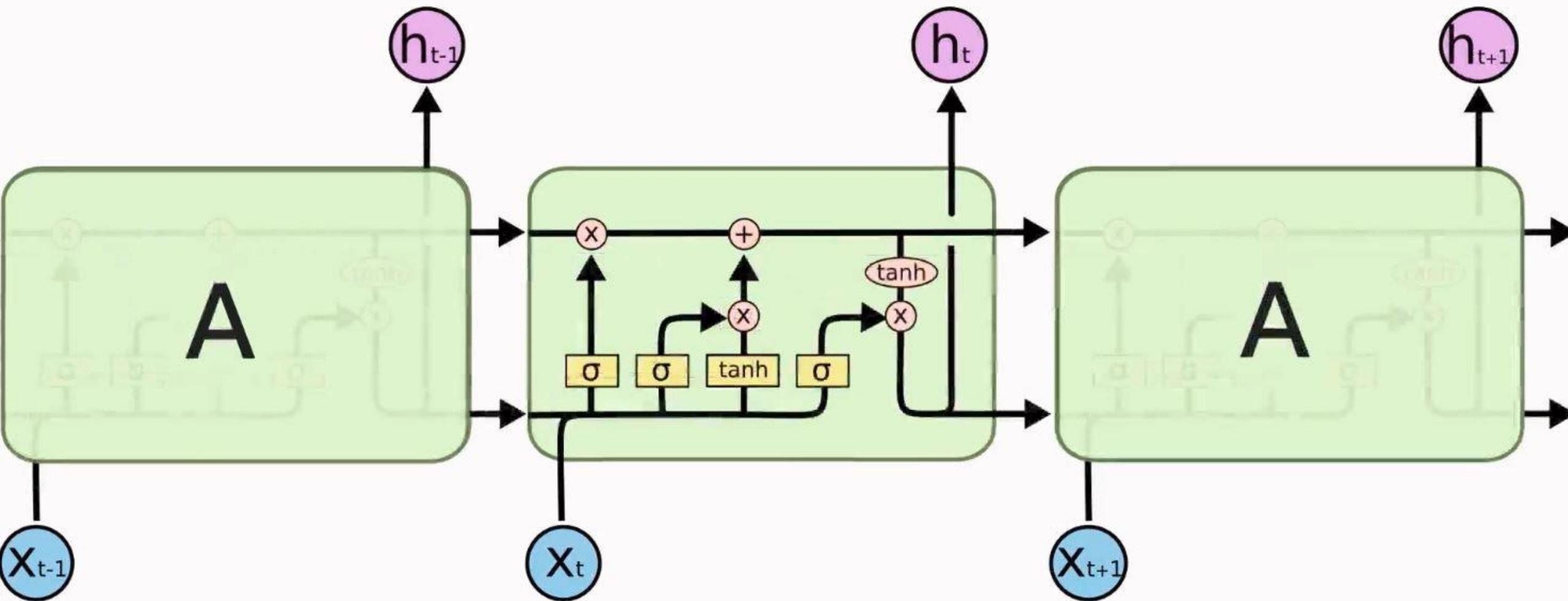
# Long Short-Term Memory



# Long Short-Term Memory



# Long Short-Term Memory



# Long Short-Term Memory

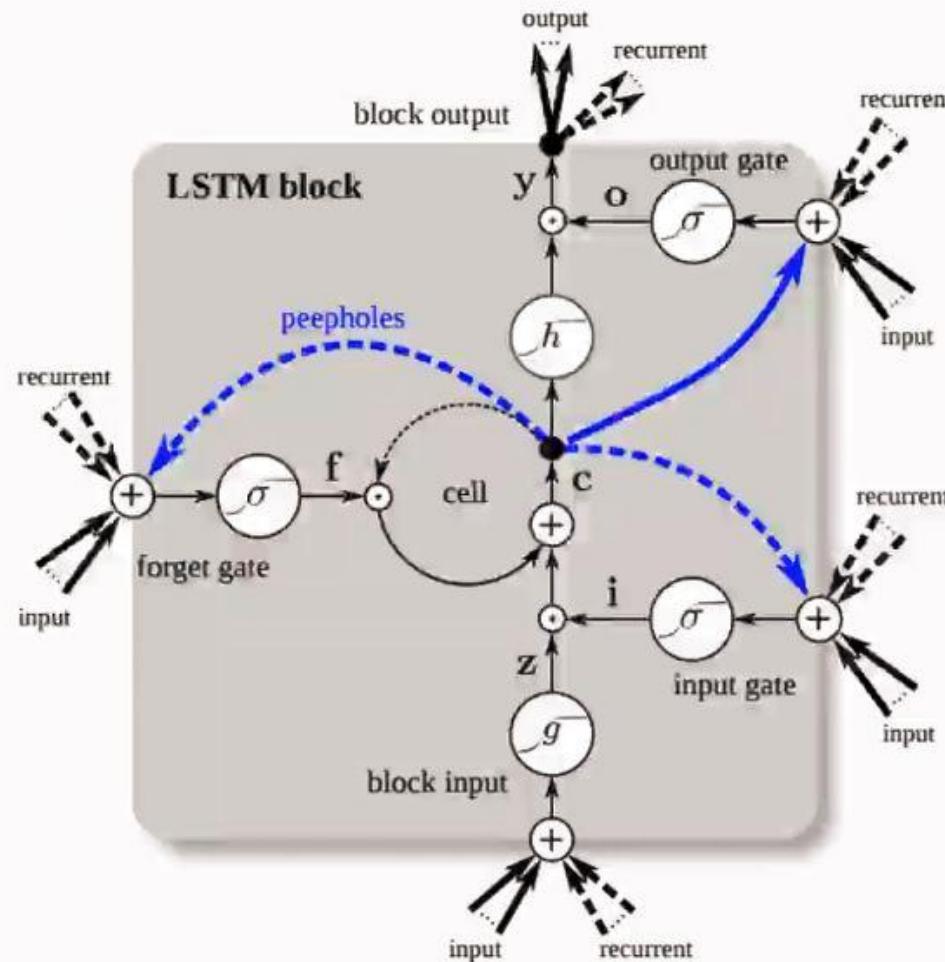
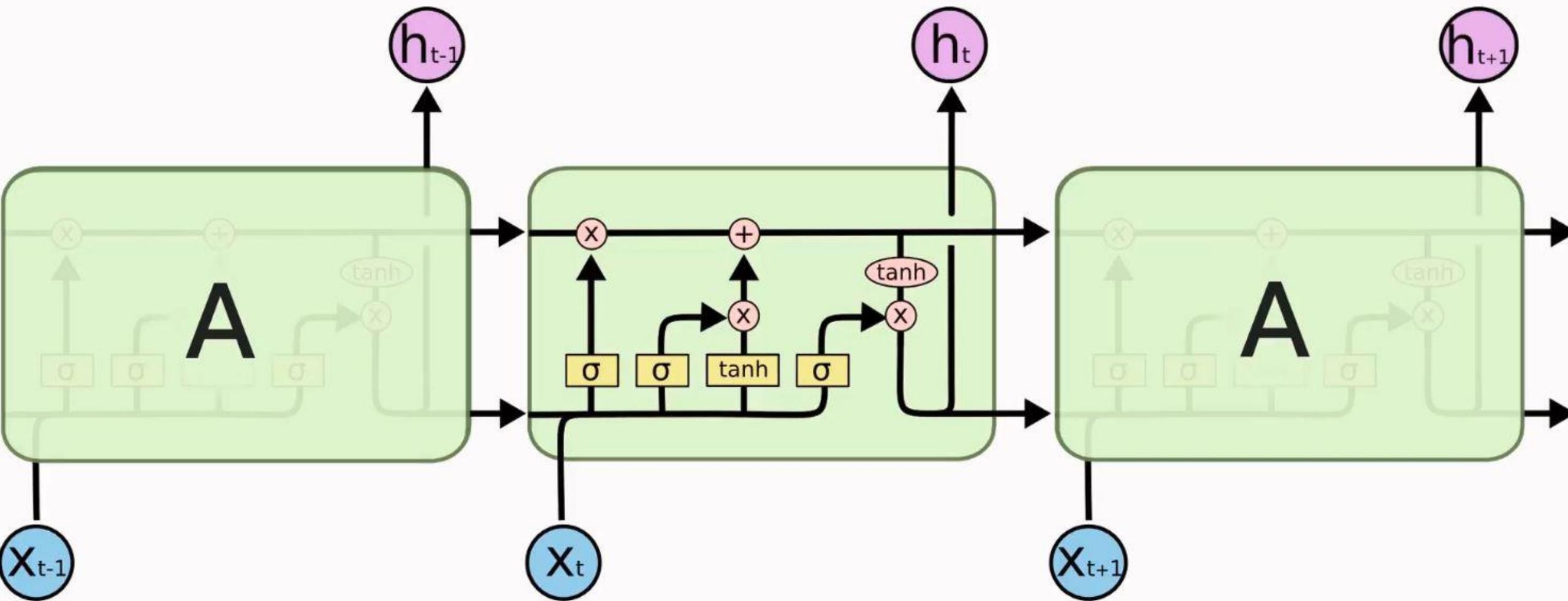


Image Source: arxiv.org/pdf/1503.04069.pdf

# Long Short-Term Memory



# Long Short-Term Memory

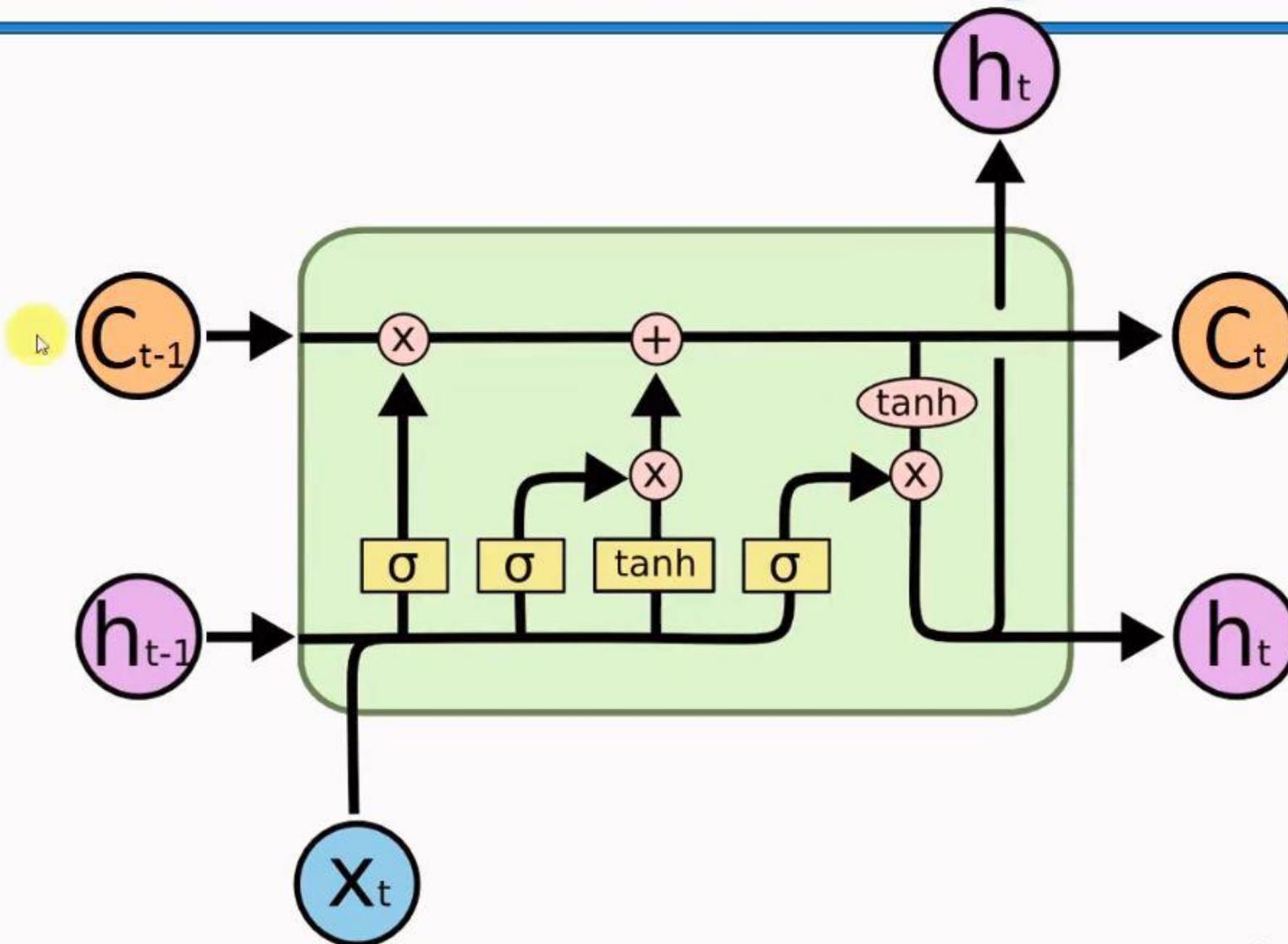


Image Source: colah.github.io

# Long Short-Term Memory

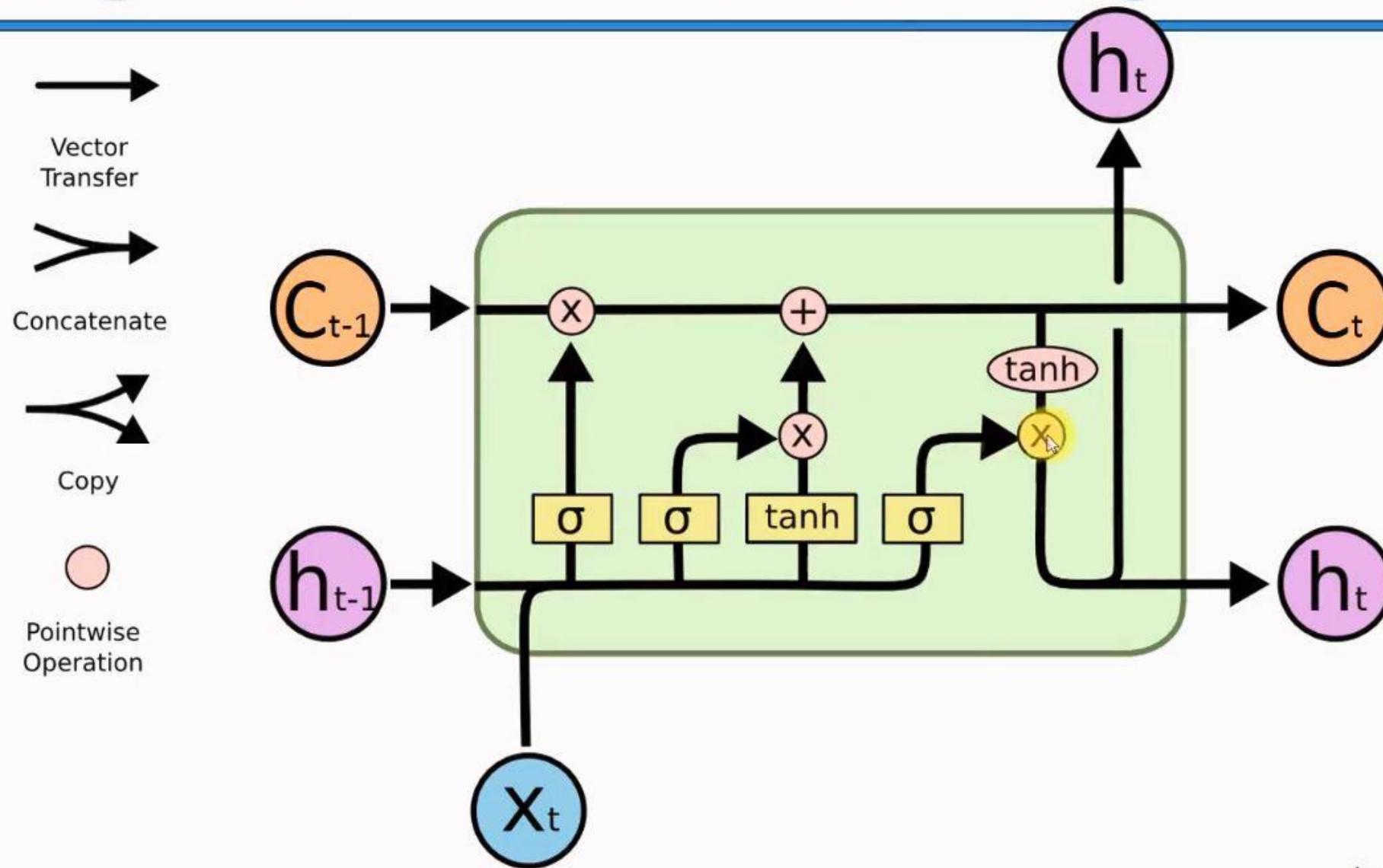


Image Source: colah.github.io

# Long Short-Term Memory



# Long Short-Term Memory

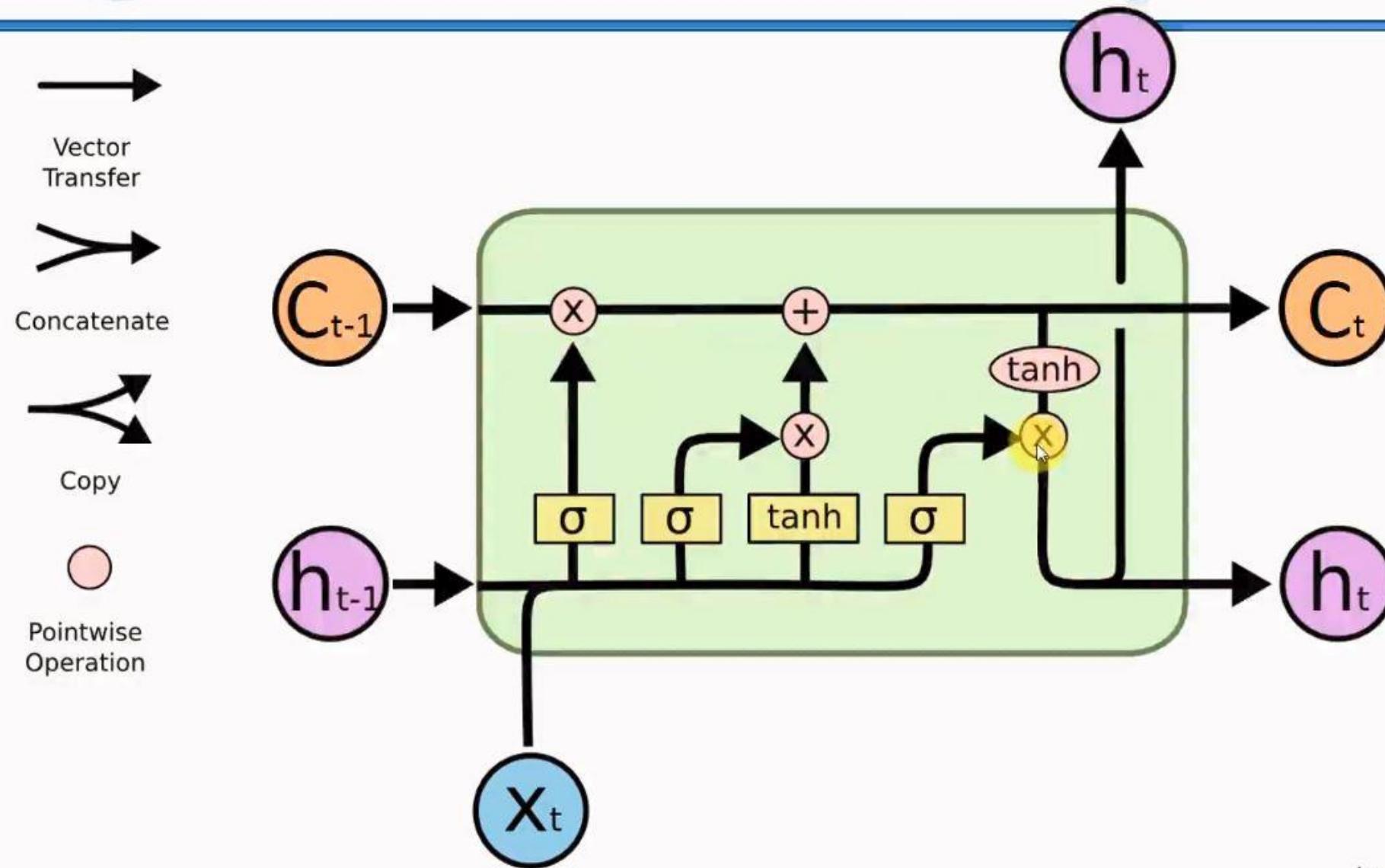


Image Source: [colah.github.io](https://colah.github.io)

# Long Short-Term Memory



# Long Short-Term Memory

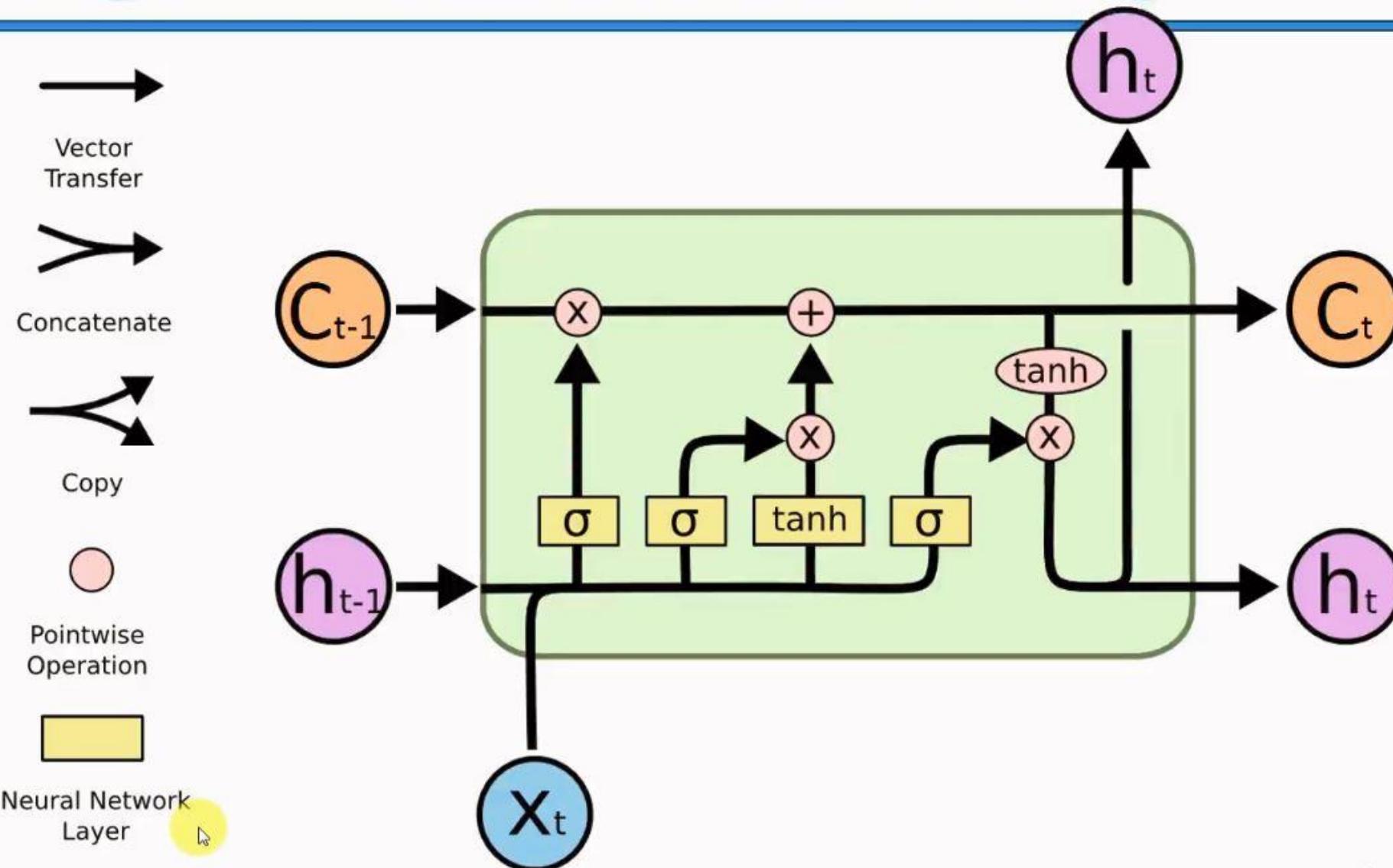


Image Source: colah.github.io

# Long Short-Term Memory

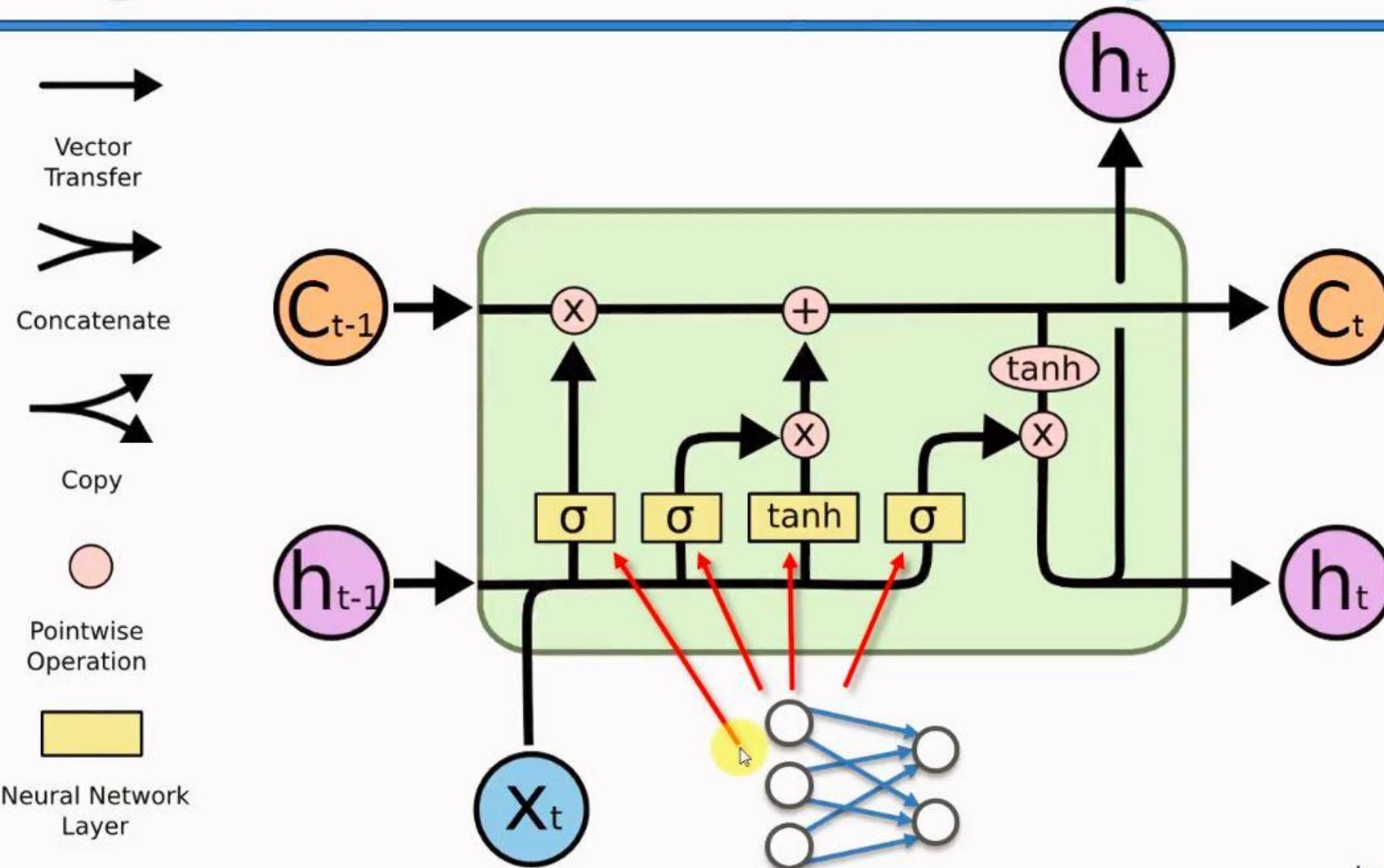


Image Source: colah.github.io

# Long Short-Term Memory

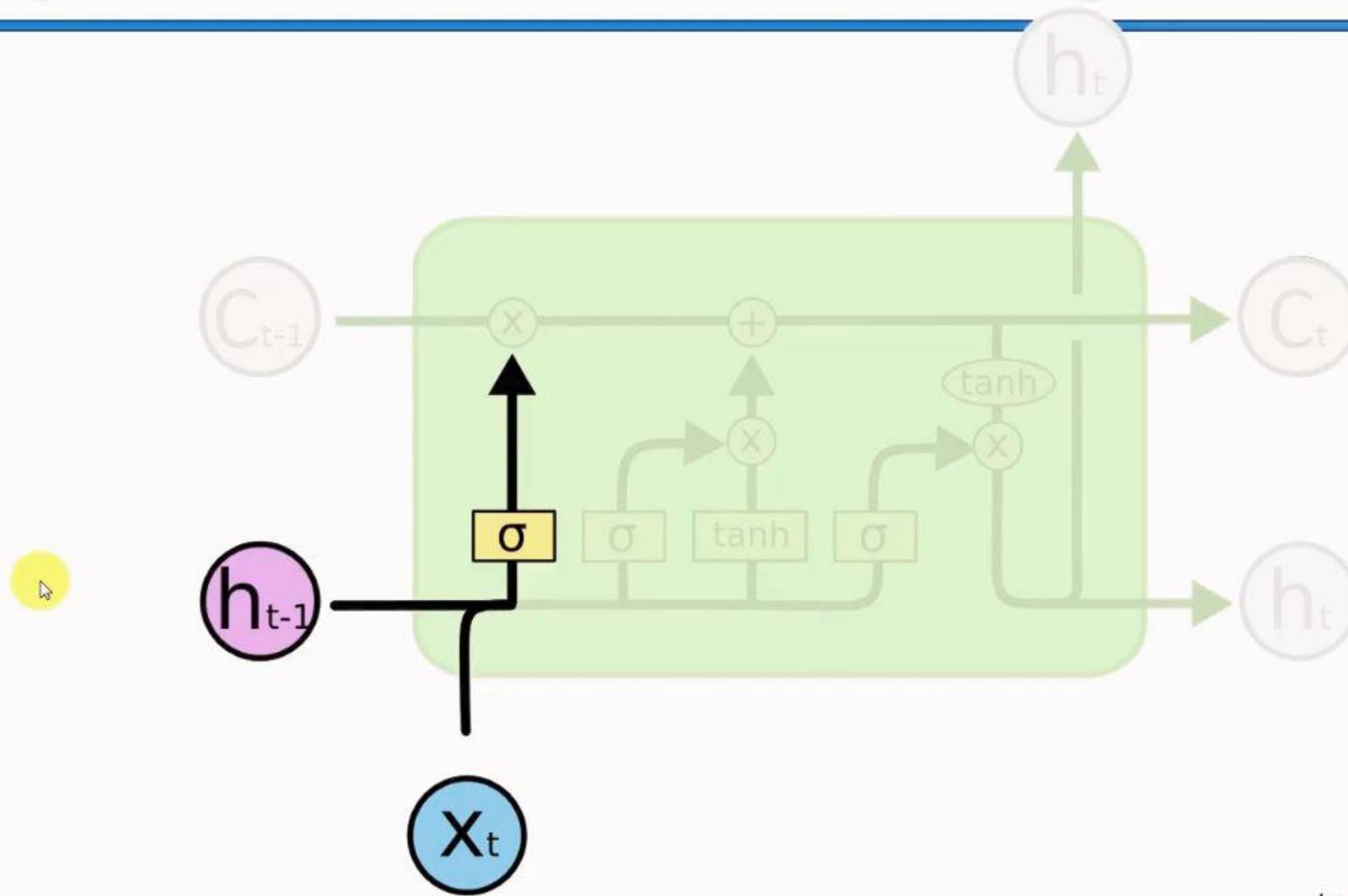


Image Source: colah.github.io

# Long Short-Term Memory

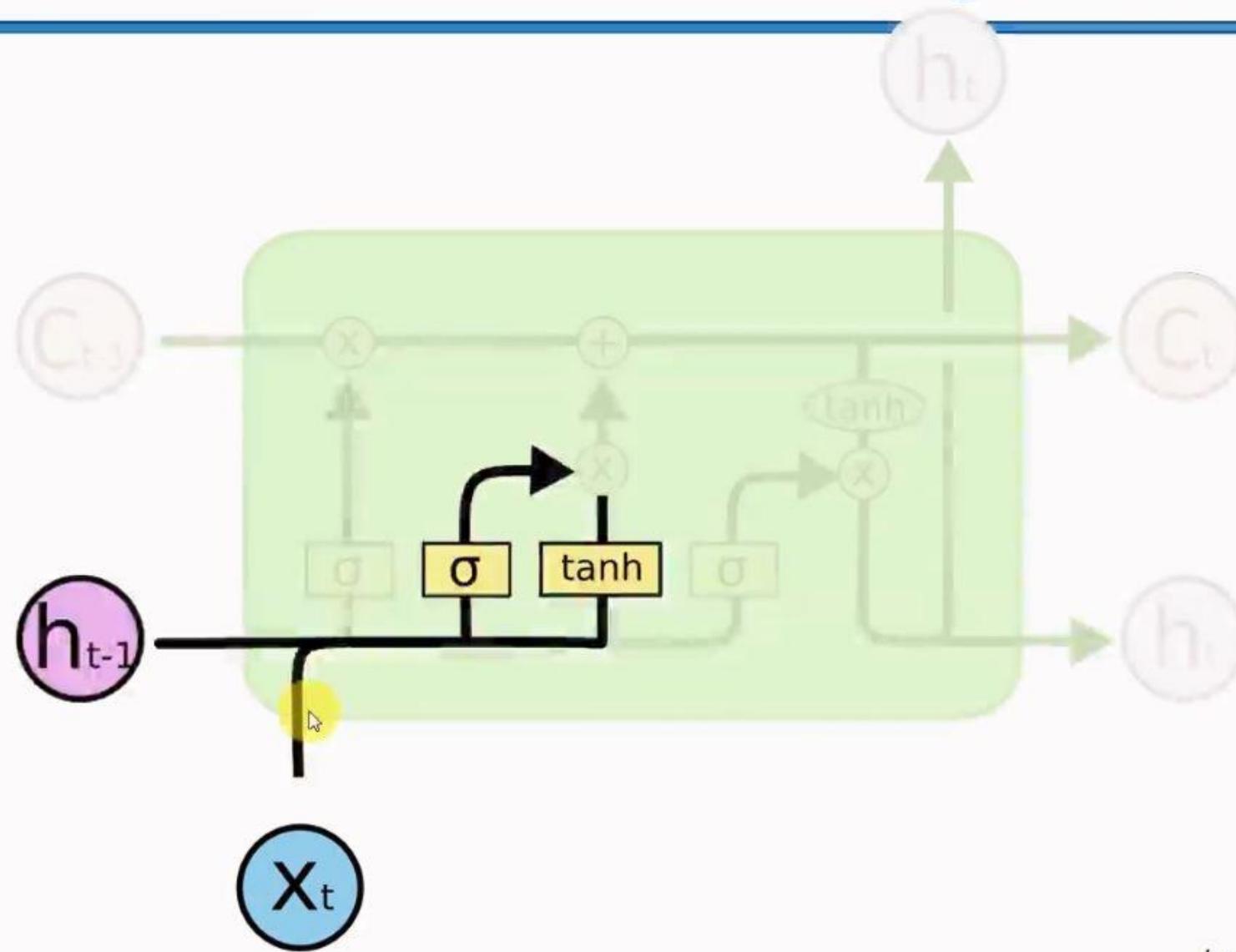


Image Source: colah.github.io

# Long Short-Term Memory

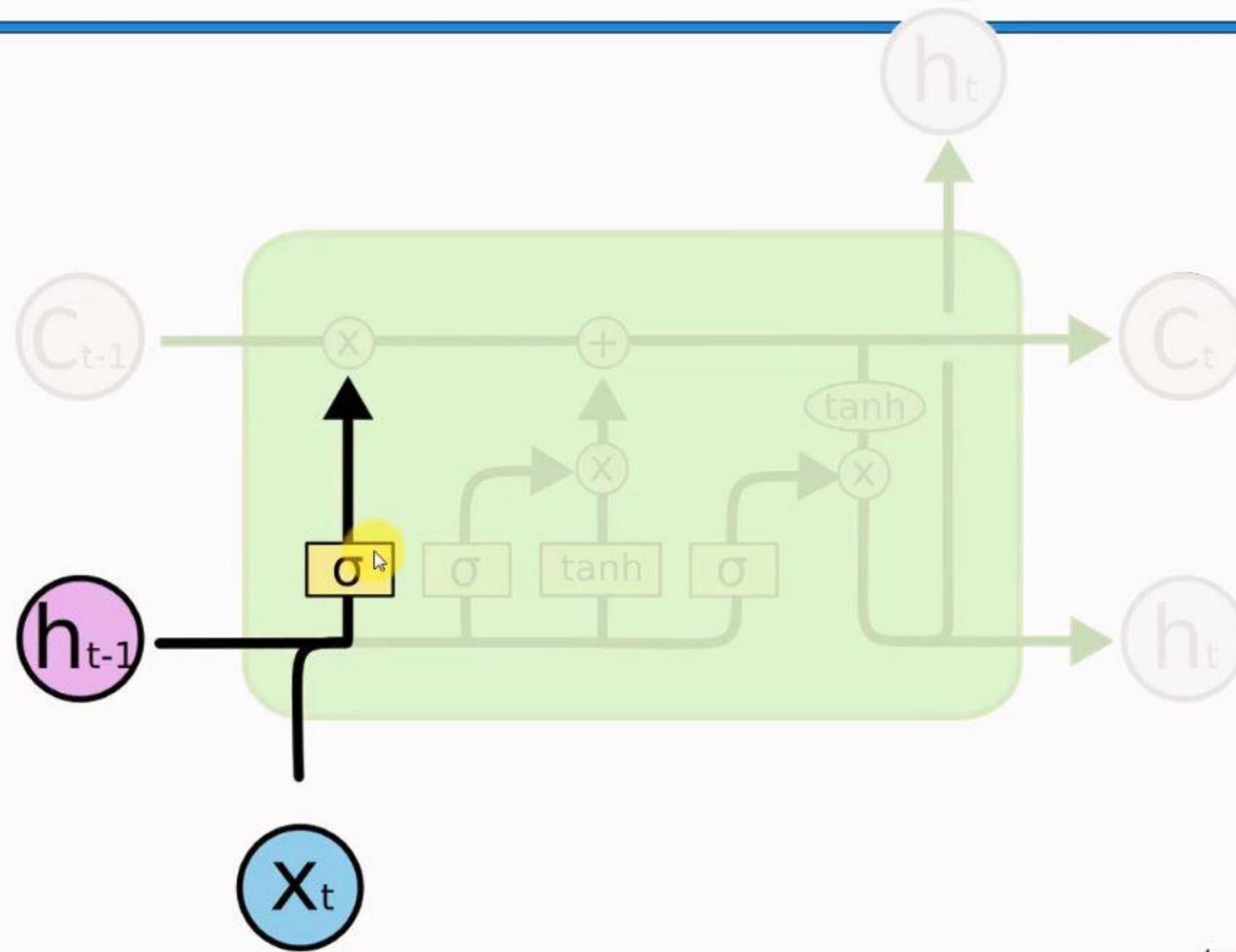


Image Source: colah.github.io

# Long Short-Term Memory

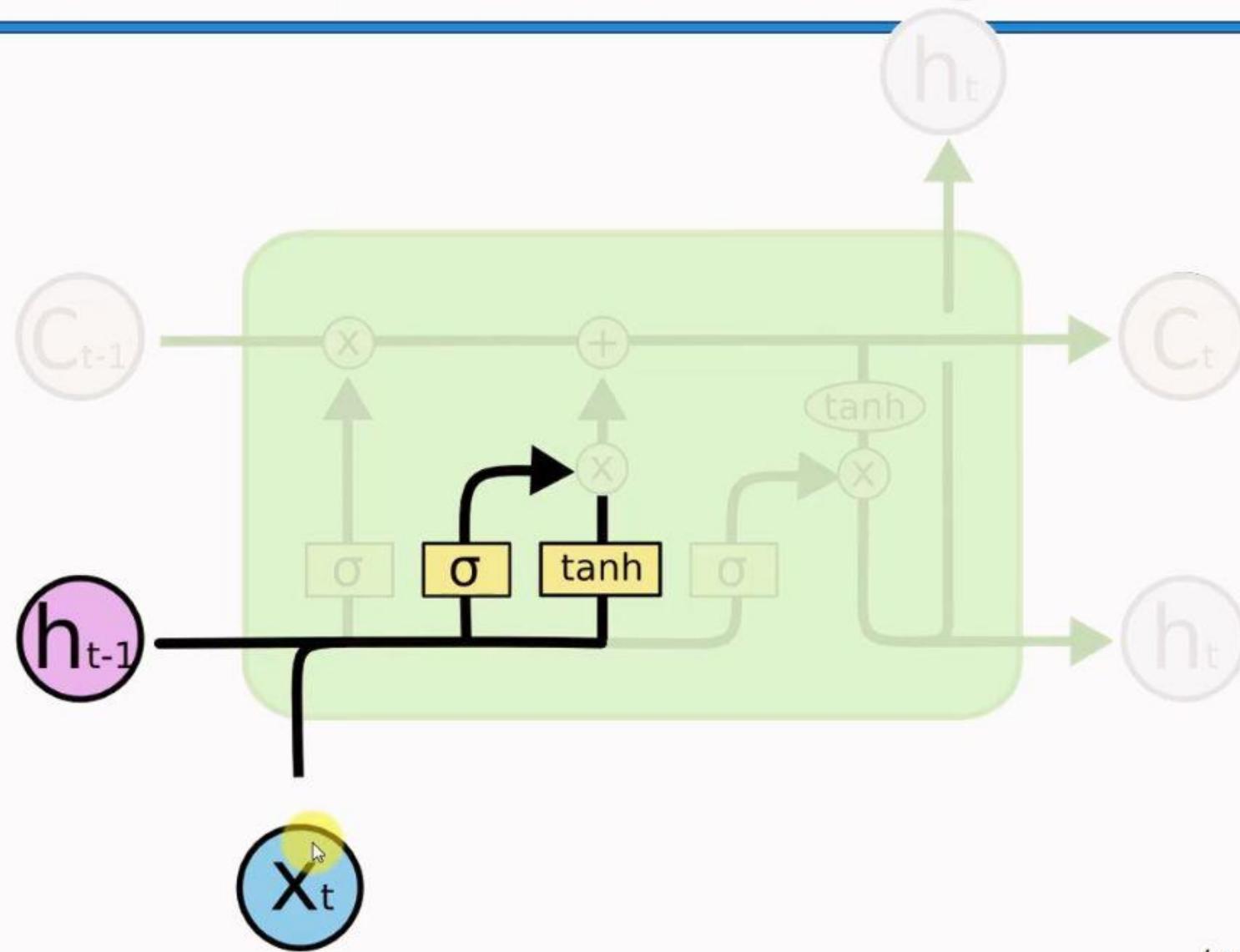


Image Source: colah.github.io

# Long Short-Term Memory

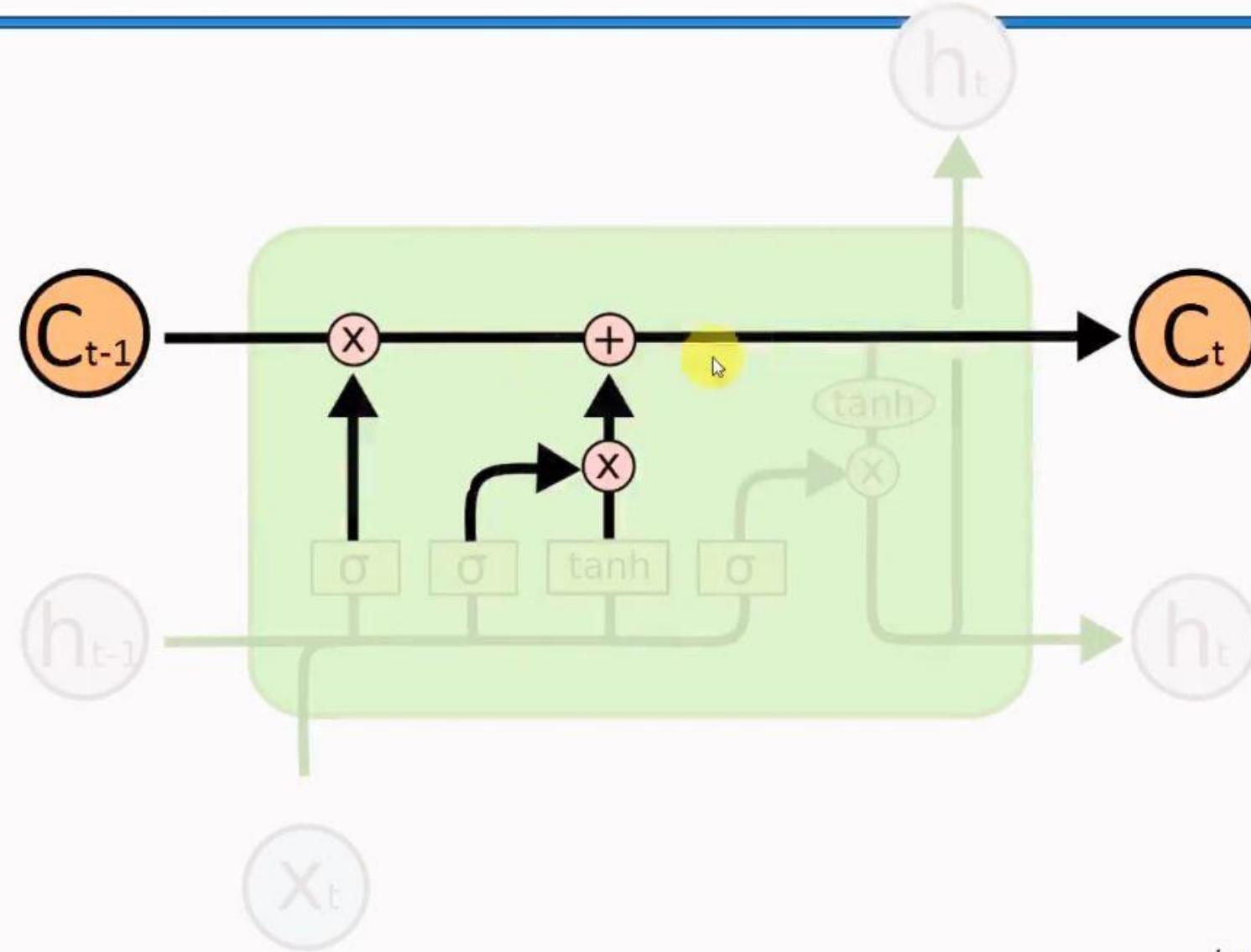


Image Source: [colah.github.io](https://colah.github.io)

# Long Short-Term Memory

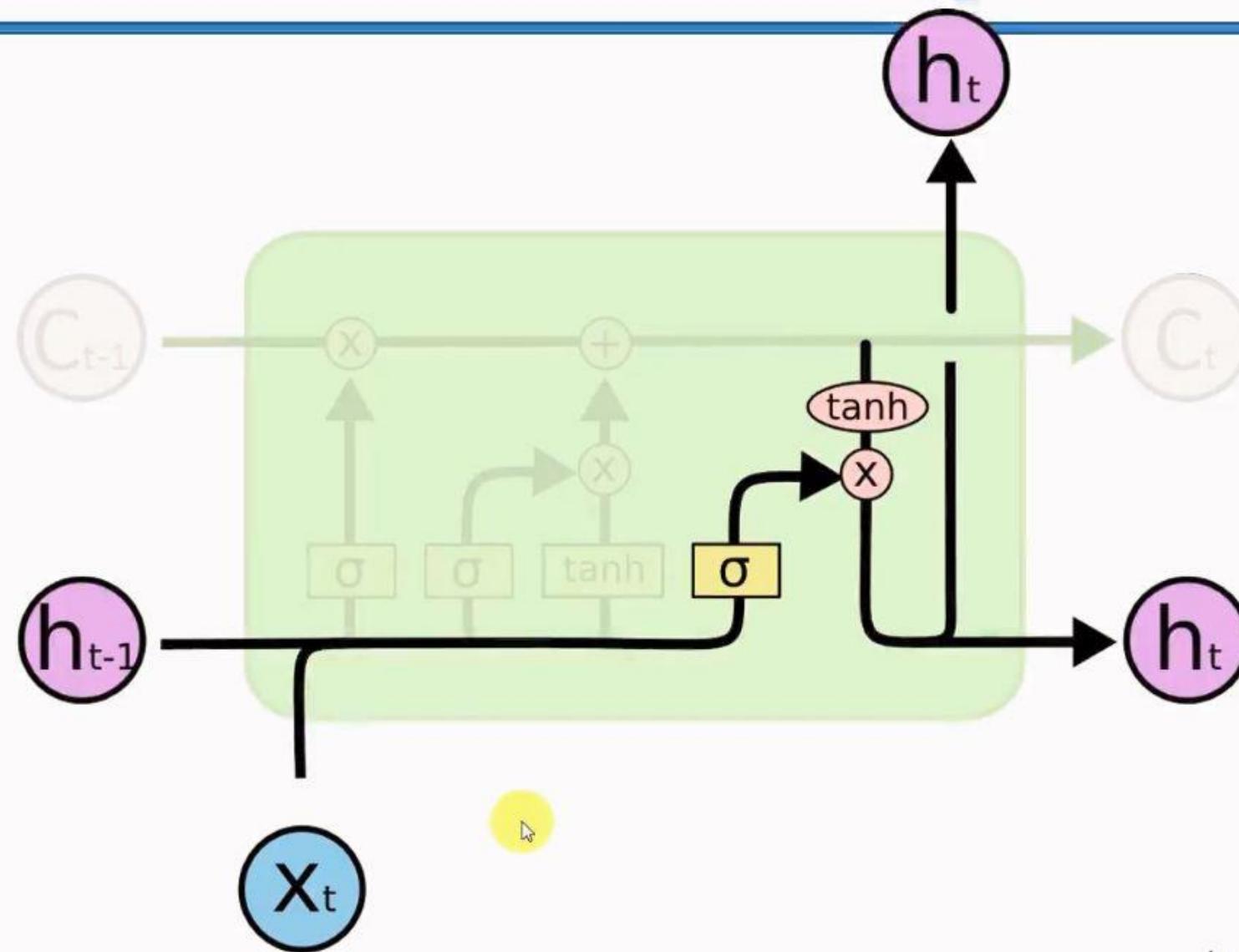


Image Source: colah.github.io

# Long Short-Term Memory

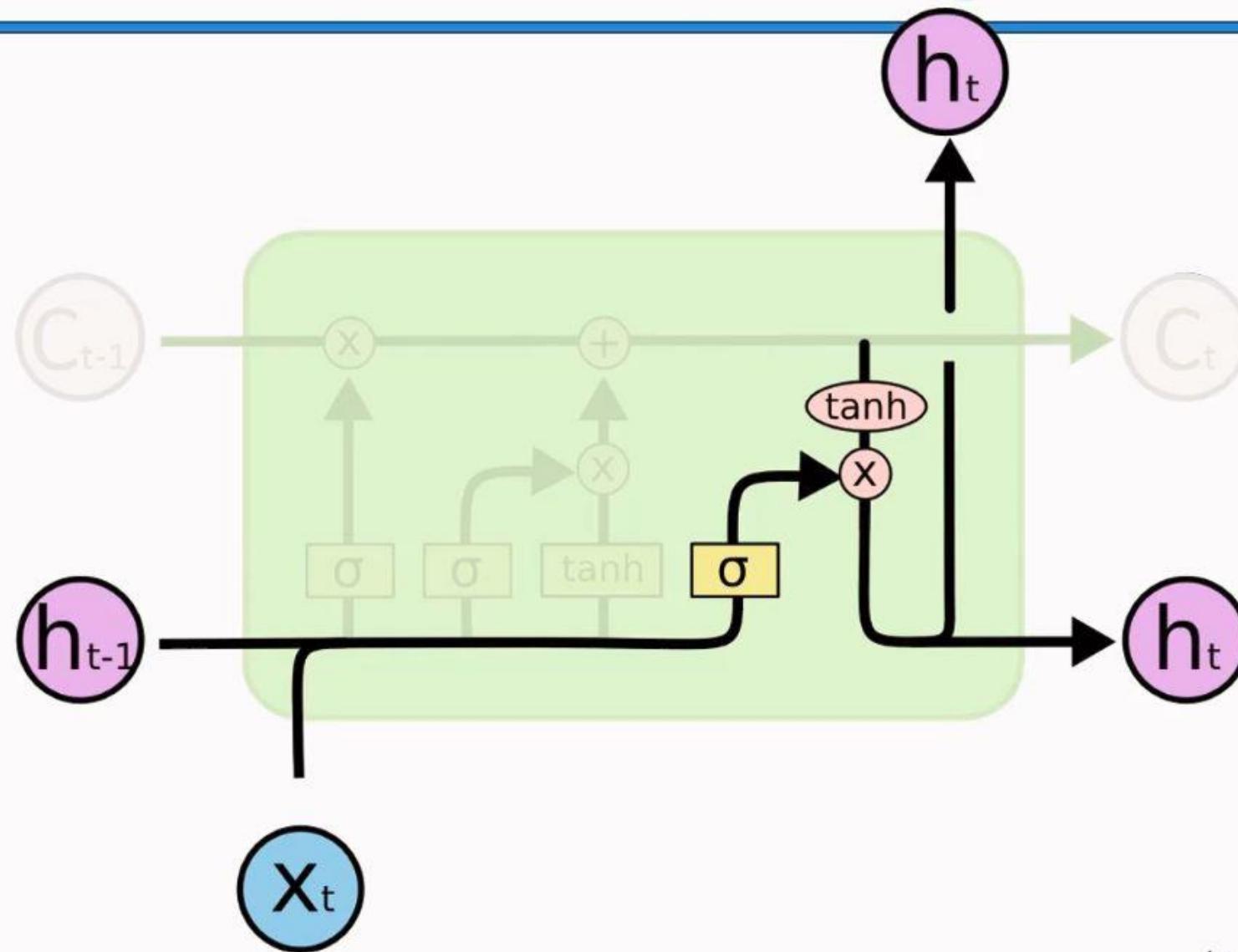


Image Source: colah.github.io

# Long Short-Term Memory

The image shows a screenshot of the Google Translate interface. At the top left is the Google logo. To its right are three icons: a grid of nine dots, a red circle with the number '1', and a small profile picture of a person wearing sunglasses. Below the logo, the word "Translate" is written in red. To its right is a link "Turn off instant translation" and a star icon. A horizontal line separates this from the main translation area. In the main area, there are two language selection bars. The first bar on the left has "English" selected and includes "Czech", "Spanish", "Detect language", and a dropdown arrow. The second bar on the right has "Czech" selected and includes "English", "Spanish", and a dropdown arrow. Between these bars is a "Translate" button with a blue gradient. Below the bars, the English sentence "I am a boy who likes to learn" is entered in a text input field, which also contains a microphone icon, a keyboard icon, and a character count "29/5000". To the right of the input field is the translated sentence in Czech: "Jsem kluk, který rád učit". Below the translated sentence are several small icons: a star, a square, a speaker, a left arrow, and a pencil.

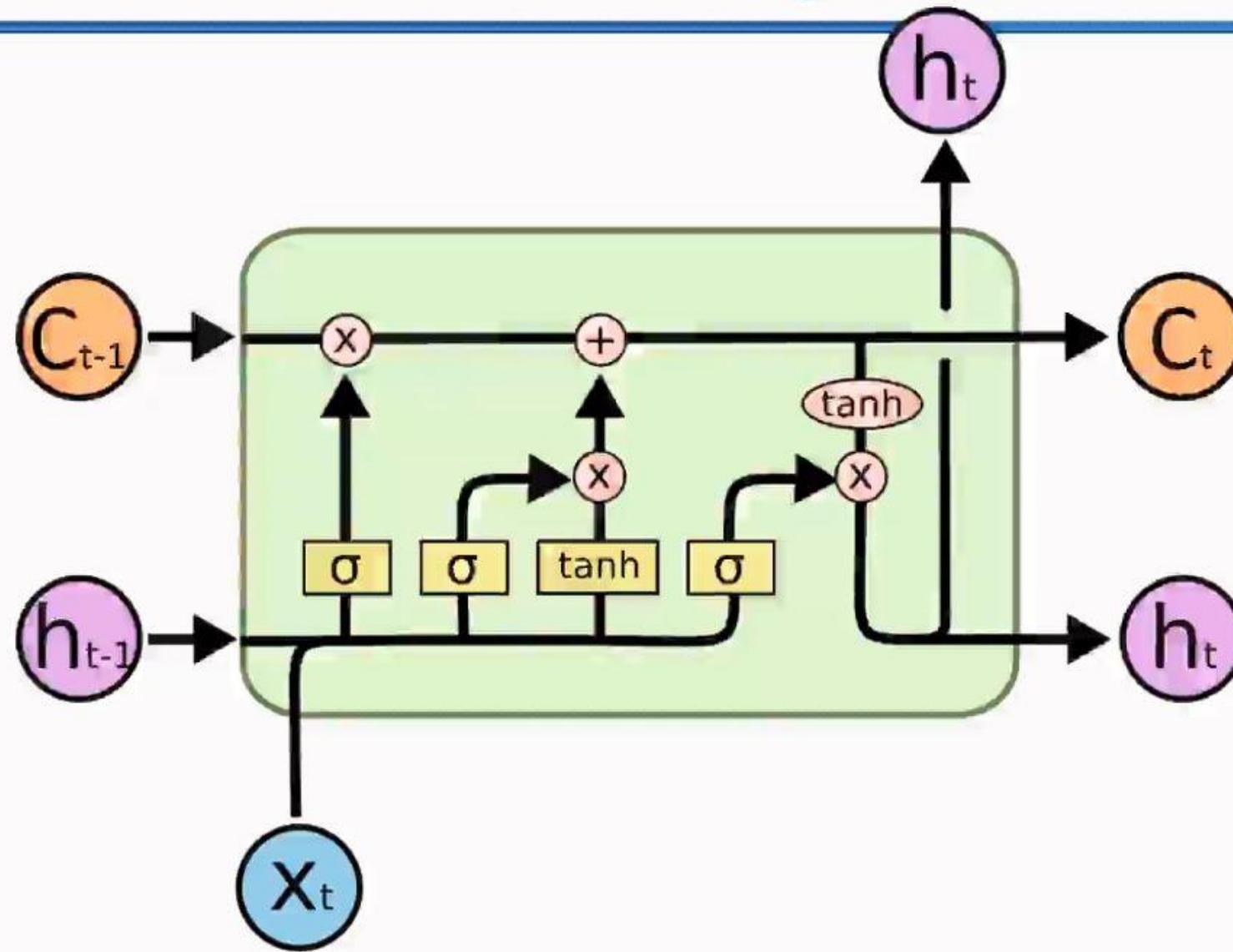
# Long Short-Term Memory

The image shows a screenshot of the Google Translate interface. At the top left is the Google logo. To its right are three small icons: a grid of nine dots, a red circle with the number '1', and a circular profile picture of a person wearing sunglasses. Below the logo, the word "Translate" is written in red. To the right of "Translate" are two buttons: "Turn off instant translation" and a star icon. The main area has two language selection bars. The first bar on the left has tabs for "English", "Czech", "Spanish", and "Detect language". The second bar on the right has tabs for "English", "Czech", "Spanish", and a blue "Translate" button. Between these bars is a double-headed arrow icon. Below the bars is a text input field containing the sentence "I am a boy who likes to learn". The word "boy" is underlined with a red line. To the right of the input field is the translated sentence in Czech: "Jsem kluk, který rád učit". A yellow circular cursor icon with a mouse arrow is positioned over the Czech text. At the bottom of the input field are icons for audio, microphone, keyboard, and character count (29/5000). At the bottom of the output field are icons for star, square, audio, and a pencil.

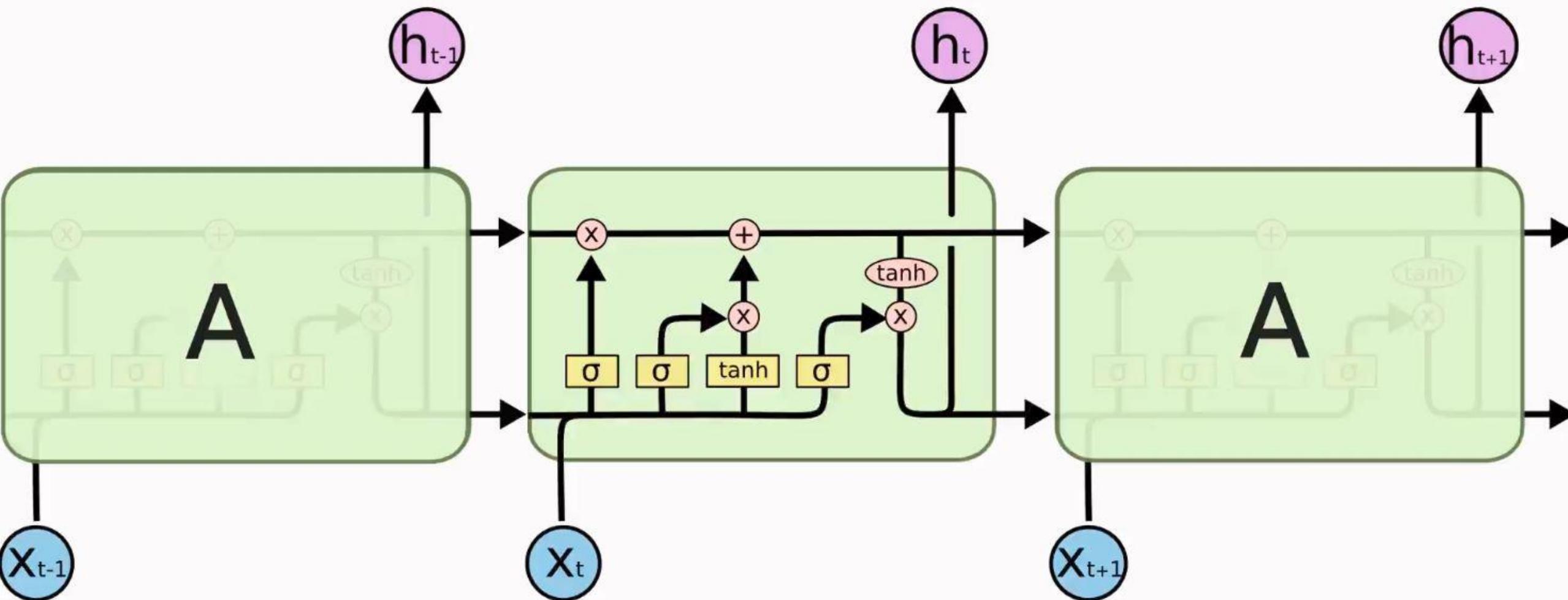
# Long Short-Term Memory

Google

Translate



# Long Short-Term Memory



# Long Short-Term Memory

Cell sensitive to position in line:

The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action--the one Kutuzov and the general mass of the army demanded--namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all--carried on by vis inertiae--pressed forward into boats and into the ice-covered water and did not, surrender.

Cell that turns on inside quotes:

"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."

# Long Short-Term Memory

Cell that robustly activates inside if statements:

```
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask,
                           siginfo_t *info)
{
    int sig = next_signal(pending, mask);
    if (sig) {
        if (current->notifier) {
            if (sigismember(current->notifier->mask, sig)) {
                if (!(current->notifier)(current->notifier->data)) {
                    clear_thread_flag(TIF_SIGPENDING);
                    return 0;
                }
            }
        }
        collect_signal(sig, pending, info);
    }
    return sig;
}
```

Cell that is sensitive to the depth of an expression:

```
#ifdef CONFIG_AUDITSYSCALL
static inline int audit_match_class_bits(int class, u32 *mask)
{
    int i;
    if (classes[class]) {
        for (i = 0; i < AUDIT_BITMASK_SIZE; i++)
            if (mask[i] & classes[class][i])
                return 0;
    }
    return 1;
}
```

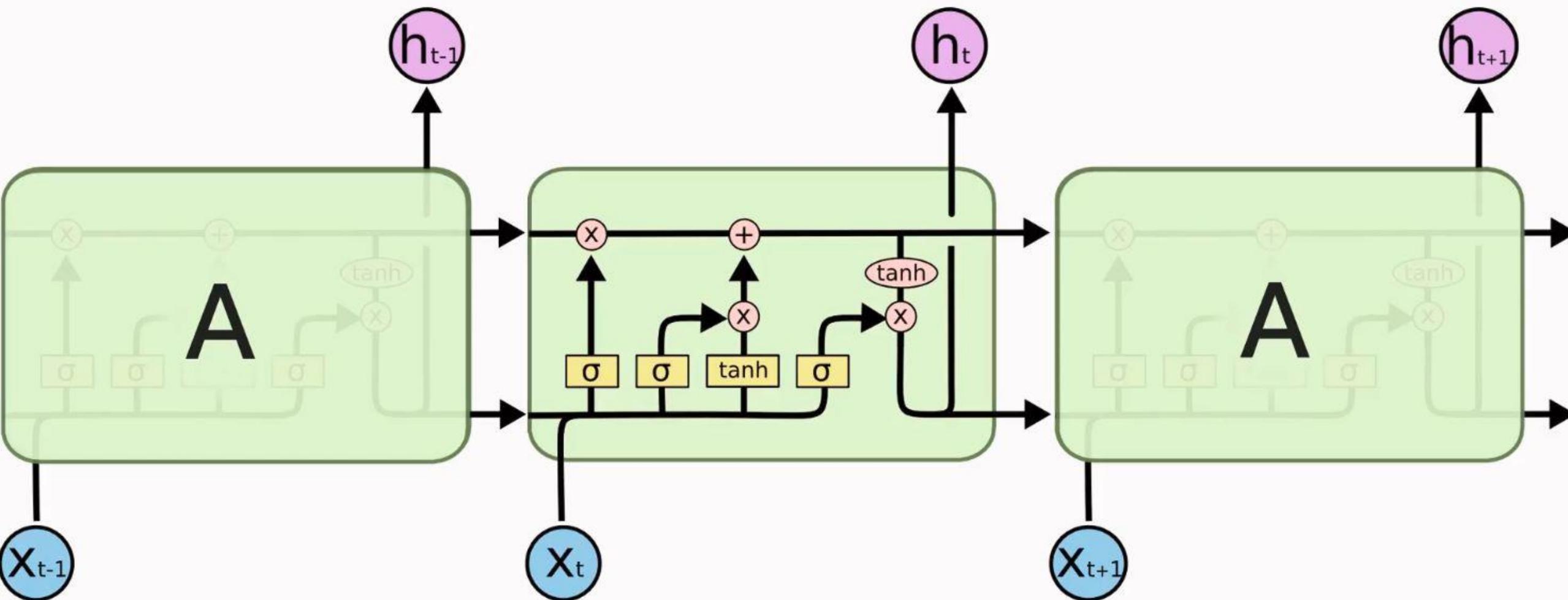
Image Source: [karpathy.github.io](https://karpathy.github.io)

# Long Short-Term Memory

A large portion of cells are not easily interpretable. Here is a typical example:

```
/* Unpack a filter field's string representation from user-space
 * buffer. */
char *audit_unpack_string(void **bufp, size_t *remain, size_t len)
{
    char *str;
    if (!*bufp || (len == 0) || (len > *remain))
        return ERR_PTR(-EINVAL);
    /* of the currently implemented string fields, PATH_MAX
     * defines the longest valid length.
    */
```

# Long Short-Term Memory



http://www.ynetnews.com/ ] English-language website of Isra  
el [ http://www.bacahets.com/ -xglish-languagesairsite of tsiae  
d : xne.waea..awatoa. s &ntiaca-sardeelh oan t bisanfanrei  
mw-2♦piiisoessiss./ern.c] (dceen epesaaiki ieel edh, irthraon  
dr. < ahb-nptwt.xi gh/ma) Tvdryzi couedlsu: tha-oo tu, stuif  
stp, tcoa2drulwoclensr] p. l l vaod, eytc-n dm-oibuvsv] bb imsul

gest newspaper ' '[ Yediot Ahronoth ] ' ' ' Hebrew-language  
newspaper [ Tel Aviv (feanemtii ) ' ' ' [ terrewslengua  
ir scoe ena iTThAoainnh Srmuw] ey s [ 'ineia'siwdde' h  
us..setlgors.asat Careeg'aClrisz] ie': , #: TAAAAT Baseeilo  
- tuaevrtid, tBAmSusyut] Asaoigs] ], . . : s MBolous: Toua-n: d  
a, d, iiuiticp.] ( ISvHvtusuiedDnoegano ., ] : { CCui bohe Cybksl s:

icals: ' ' ' ' [ [ Globes ] ] ' ' [ http://www.globes.co.il/ ] bus  
cal: ' ' ' ' [ Taaba ] ' ' ( [ http://www.buobal.comun/sA-yti  
stl' [ hAeoveltsahad:xge.waoir.rtoael.it.&ai  
tt' ' &[ &&mCoerone': , i'odw., niiisaue.eni/omlcc.(eft  
a'n:, C: & # \* : afDrusu] l, . omel p<, dha; deuoot/ihncsifS, urh  
nk i < ] : &11s TGuitrsi, : bacmr-xt pob-gresislerlnafad] losp

ily ' ' [ [ Haaretz | Ha'Aratz ] ] ' ' [ http://www.haaretz.co.il/ ]  
ly ' ' [ [ Terrdn Ferantah ] ] ' ' ( [ http://www.bonmdst.comun/s  
re ' ' hAiInntteHalsrcnol'sahad:xne.waamrtdheohol.c  
ki. : \* sCOSanlt hitim'li] e : , imcdw-2♦phi

http://www.ynetnews.com/ ] English-language website of Isra  
tp://www.bacahets.com/ -xglash linguagesairsite of tsiae  
d:xne.waea.awatoa.s &ntiacaca-sardeelh oan t bisanfanrei  
mw-2♦piiisoessiss.ern.c](dceen epesaaiki ieel edh, irthraon  
dr.<ahb-nptwt.xi gh/ma) Tvdryzi couedlsu:tha-oo tu, stuif  
stp,tcoa2drulwoclensr]p.llvaod,,eytc-n dm-oibuvsv]bb imsul

gest newspaper '' [[Yedidith Ahrenoth]]''' Hebrew-language  
e i awspaperso [[Tel t i(feane mti)]'' [errewslenguage  
ir scoe ena iTThAoainnh Srmuw] ey s [ineia'siwdde'h  
us..setlgors.asatCareeg'aClrisz]ie'::# Taaaaat Baseeilo  
- tuaevrtid,tBAmSusyut] Asaoigs]],.:s MBolous:Toua-n:d  
a,d,iiuiticp.](ISvHvtusuiedDnoegano.,,]:{ CCuibohCybksls:

icals:''' \*'[[Globes]]'' [http://www.globes.co.il/] bus  
cal :''' \*'[Taaba]'' ([http://www.buobal.comun/sA-yti  
stl' [hAeoveltsahaad:xge.waoir.rtoael.it.&ai  
tt' &[ &&mCoerone':,i'odw,:niiisaue.eni/omlcc.(eft  
a'n:,C:&#\*:afDrusu],.omel p<,dha;deuoot/ihncsifs,urh  
nk i <]:&11s TGuitrsi,:bacmr-xtpob-gresislerlnafad]Iosp

ily''' [[Haaretz|Ha'Aratz]]'' [[http://www.haaretz.co.il/]  
ly''' [[Terrdn Ferantah]]'' ([http://www.bonmdst.comun/s  
re' hAiInntteHalsrcnol'sahaad:xne.waamrtdheoh.oil.c  
ki.:\*sCOSanlt hitim'li]e:,,imcdw-2♦phi