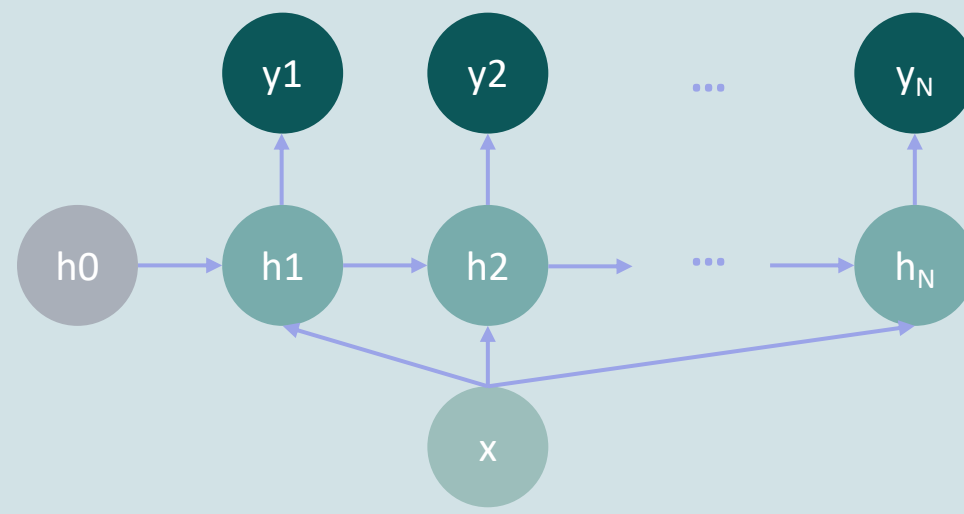


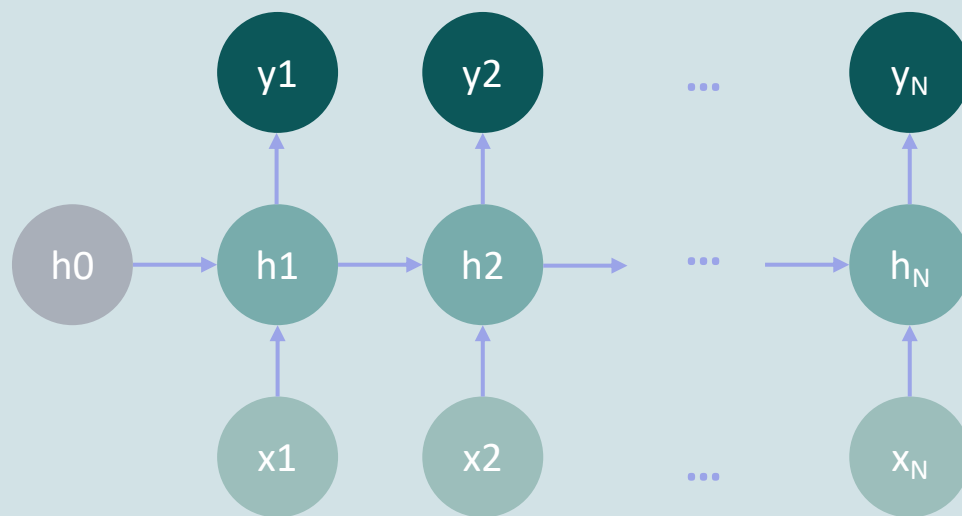
RNN模型的常见结构



N to 1
输入长度 输出长度

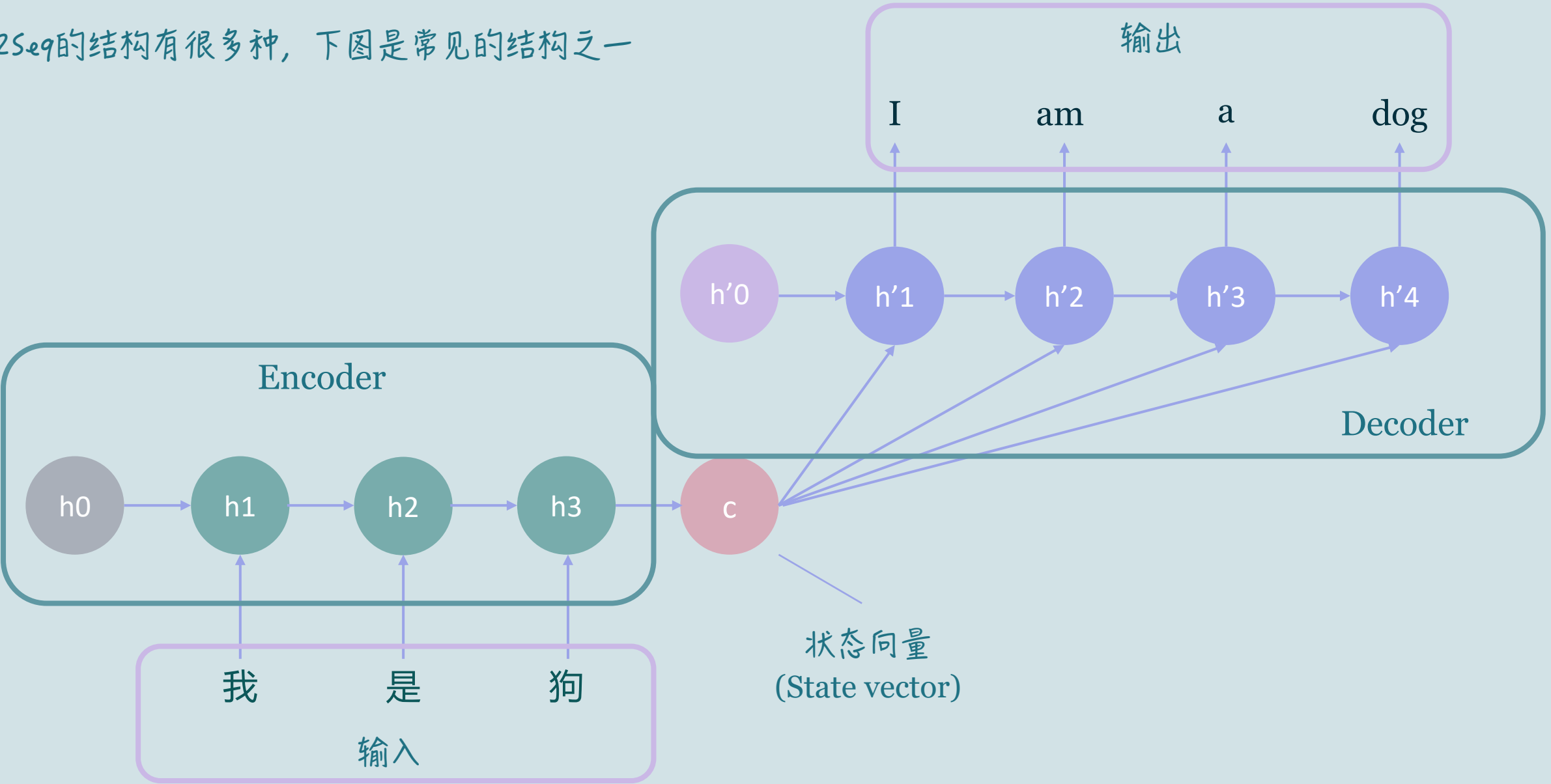


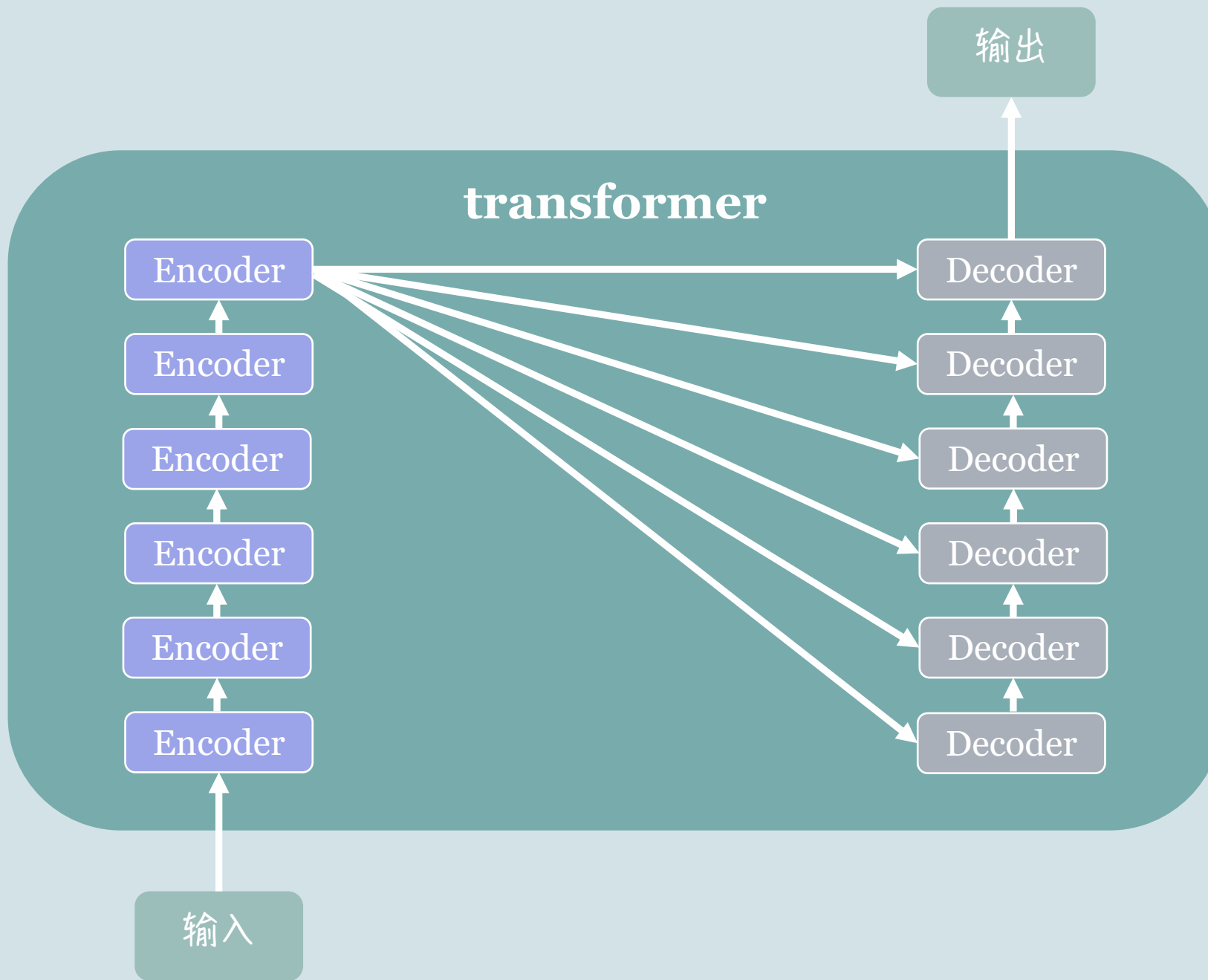
1 to N

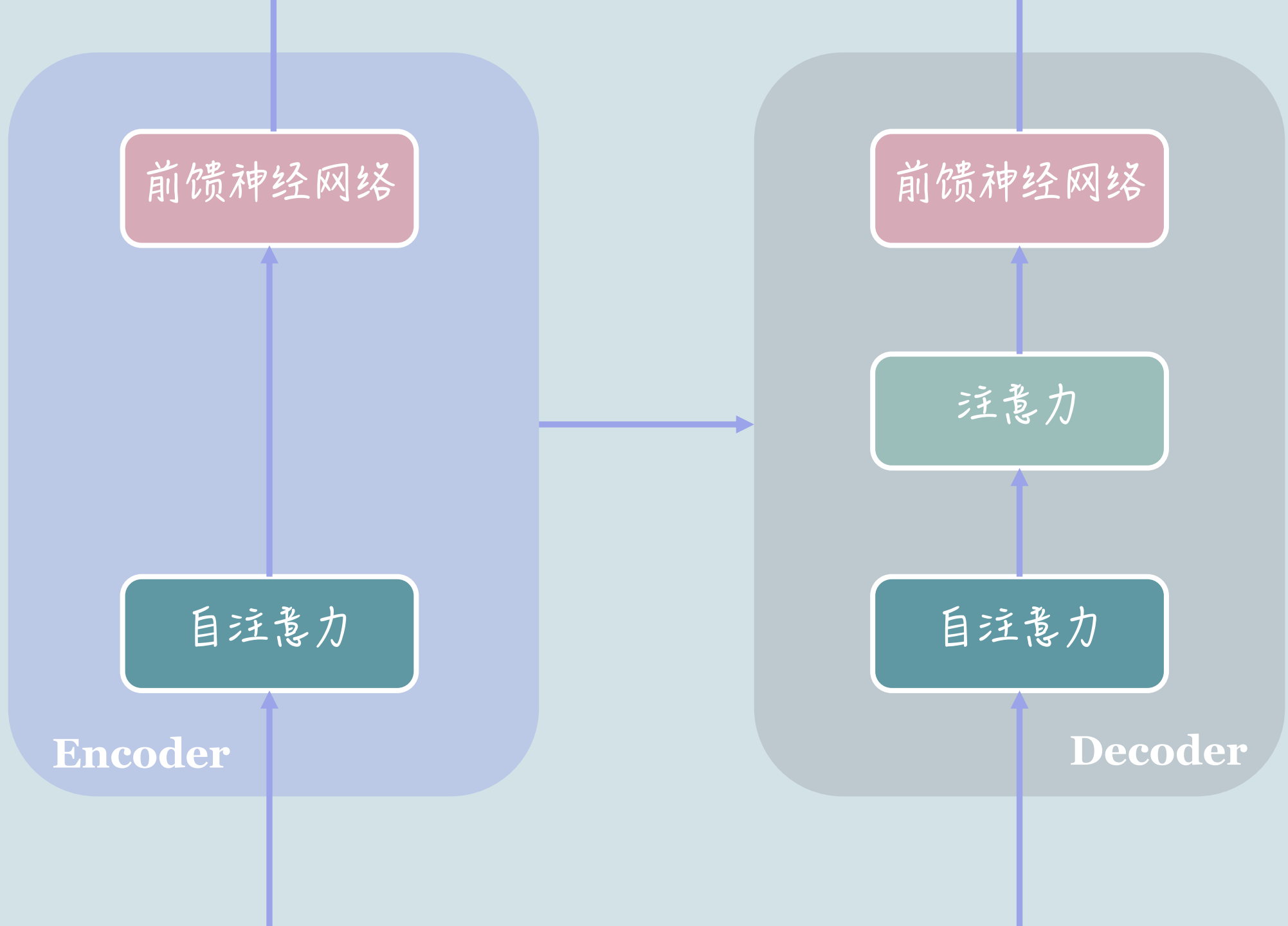


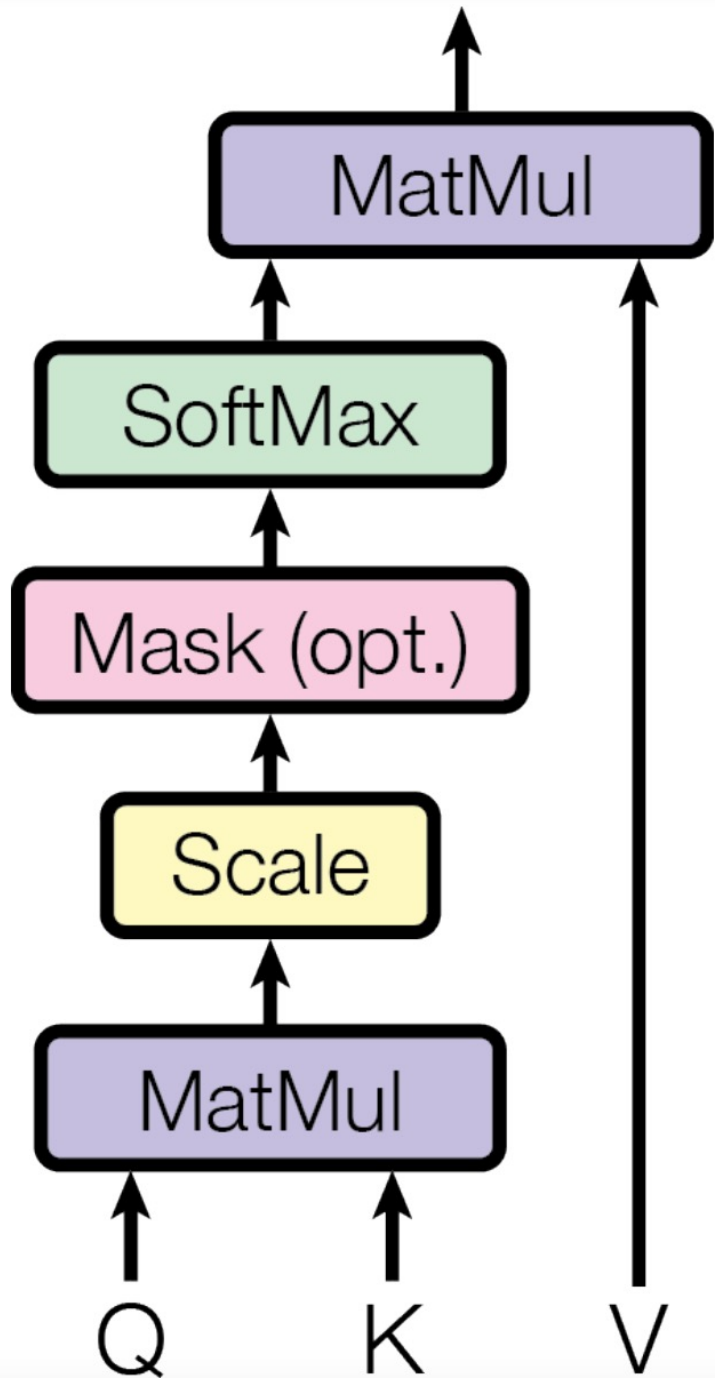
N to N

Seq2Seq的结构有很多种，下图是常见的结构之一









$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Query Key Value

$Q = W^Q X$ $K = W^K X$ $V = W^V X$

查询 索引 内容



渣男



备胎1



备胎2



备胎3

...



备胎N

择偶要求：Q
自身条件：K
被匹配到的备胎：V



渣男



备胎1



备胎2

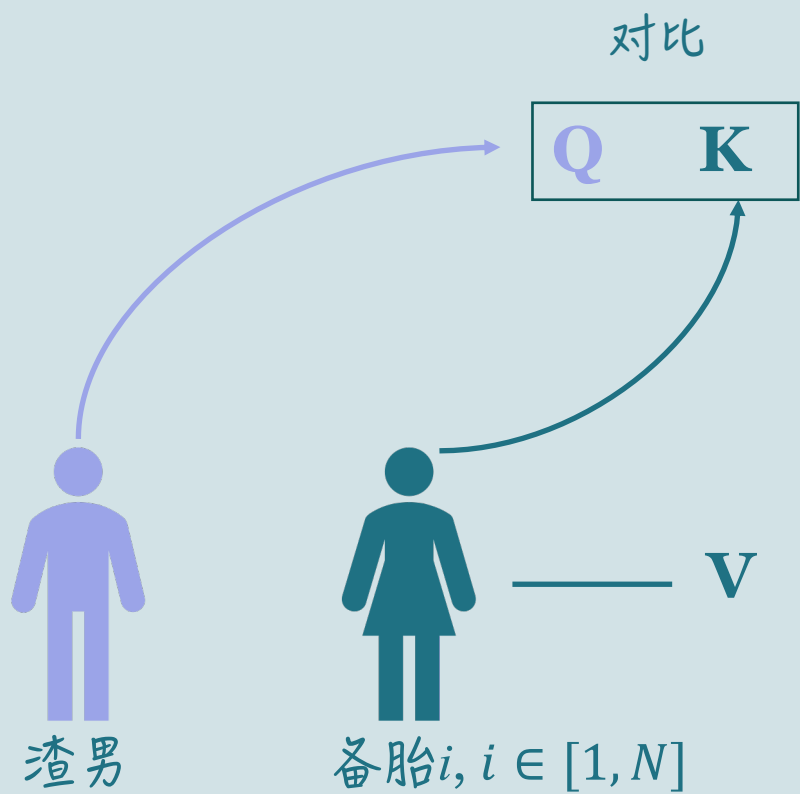


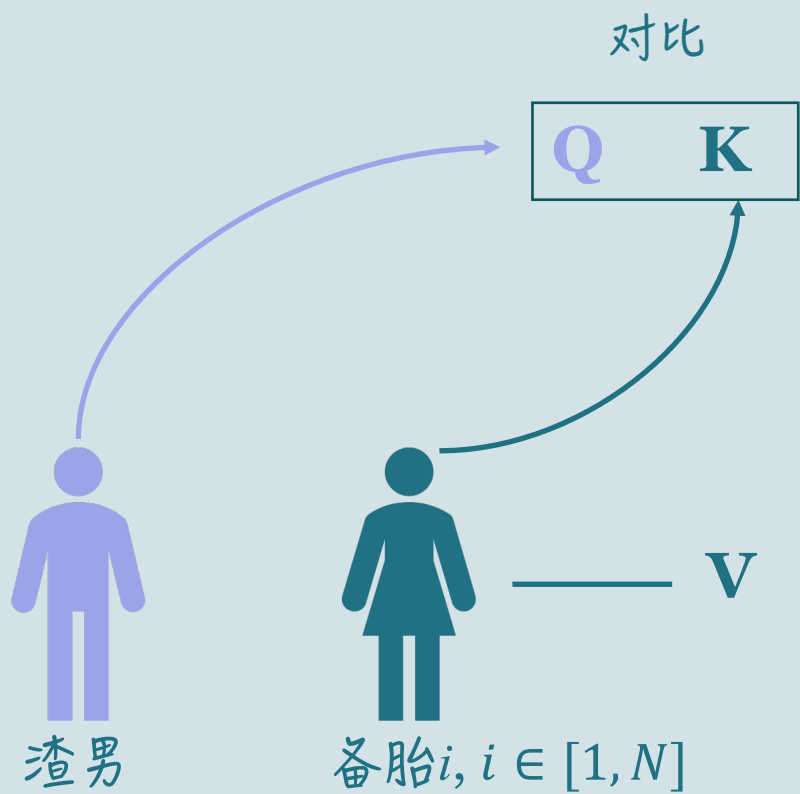
备胎3

...

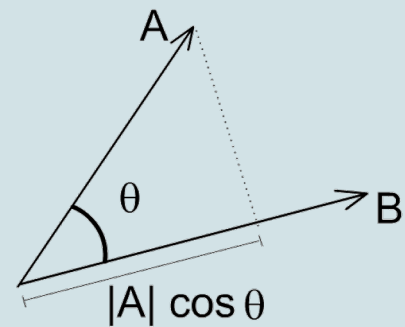
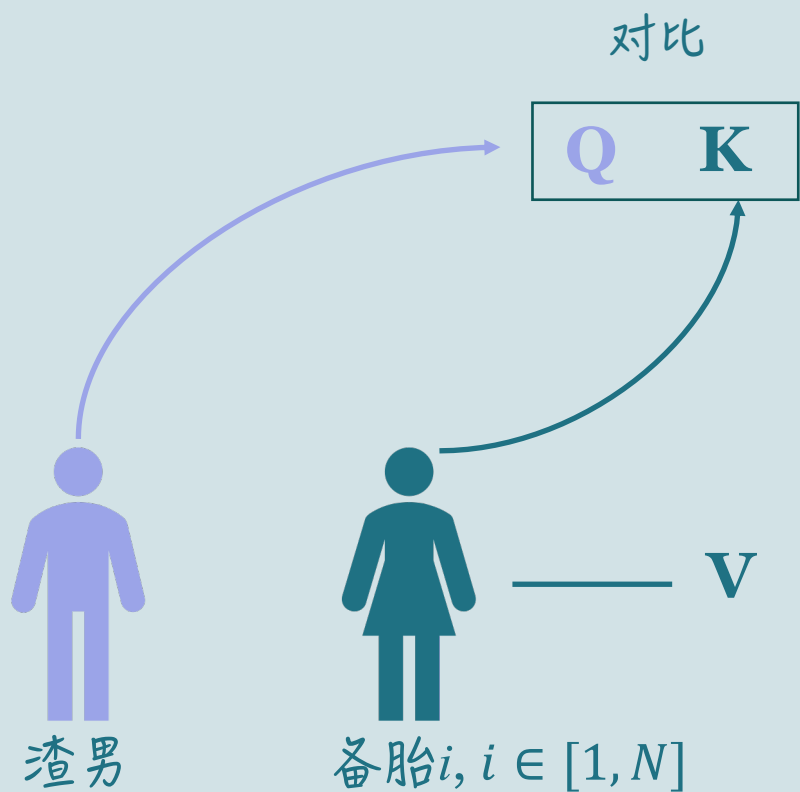


备胎N



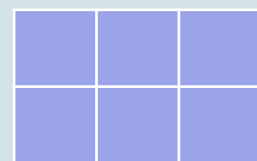


$$A \cdot B = |A||B| \cos \theta$$

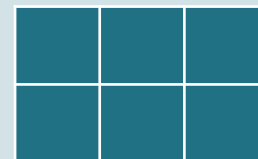


$$A \cdot B = |A||B| \cos \theta$$

$$QK^T = ?$$



Q



K

$$\text{softmax}(QK^T)$$

0.4

0.02

0.3

0.25

0.03



...



渣男自己

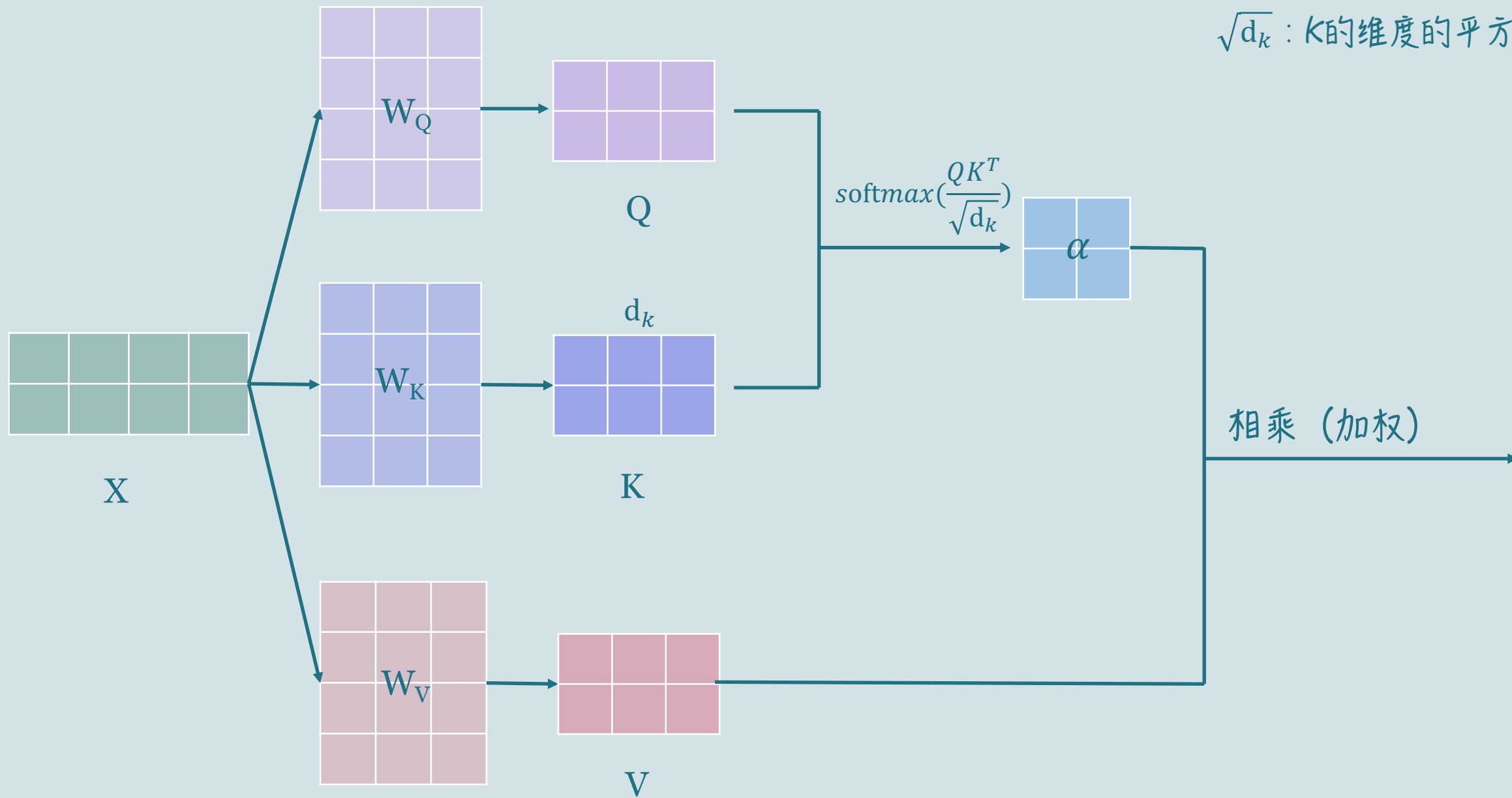
备胎1

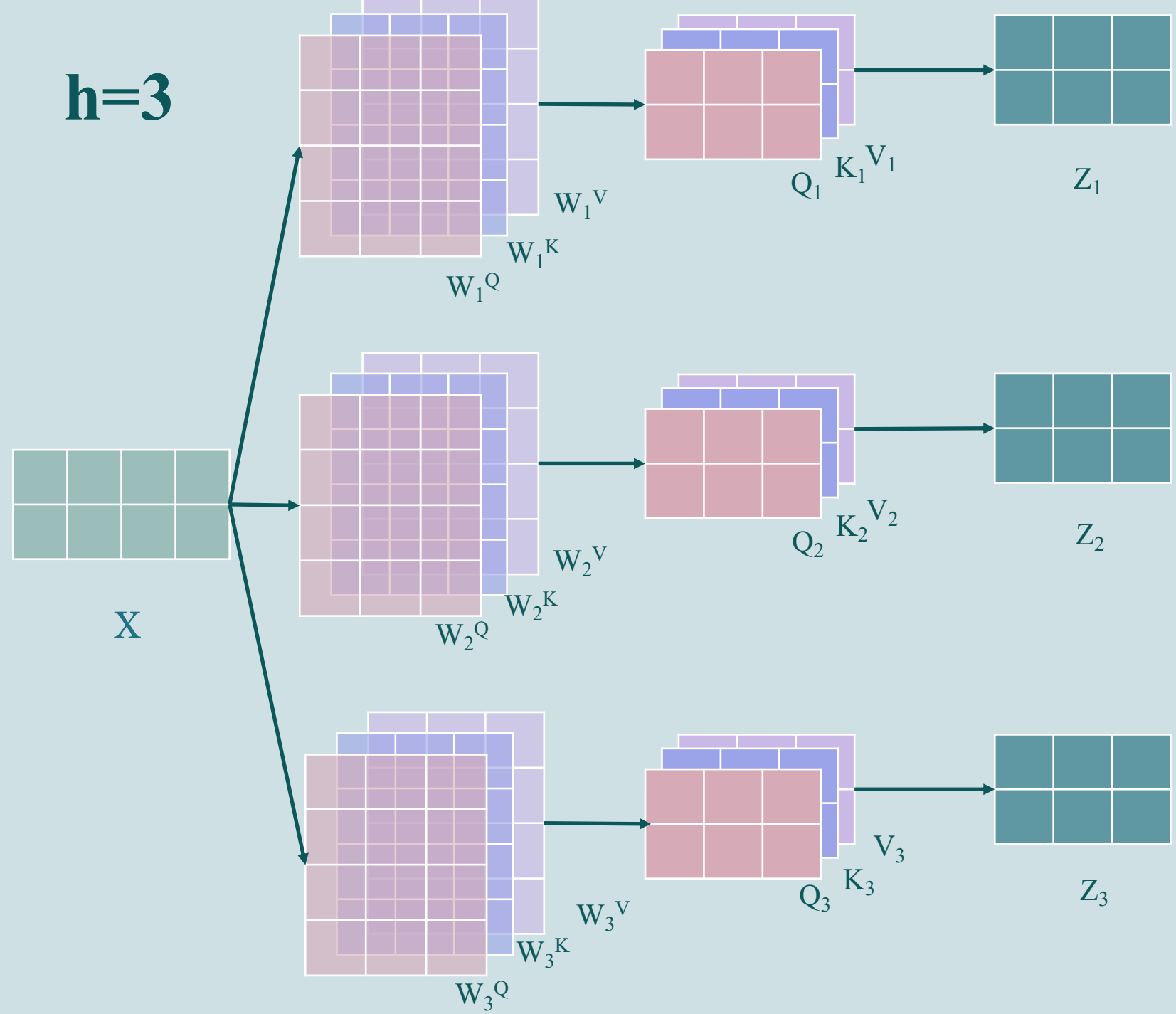
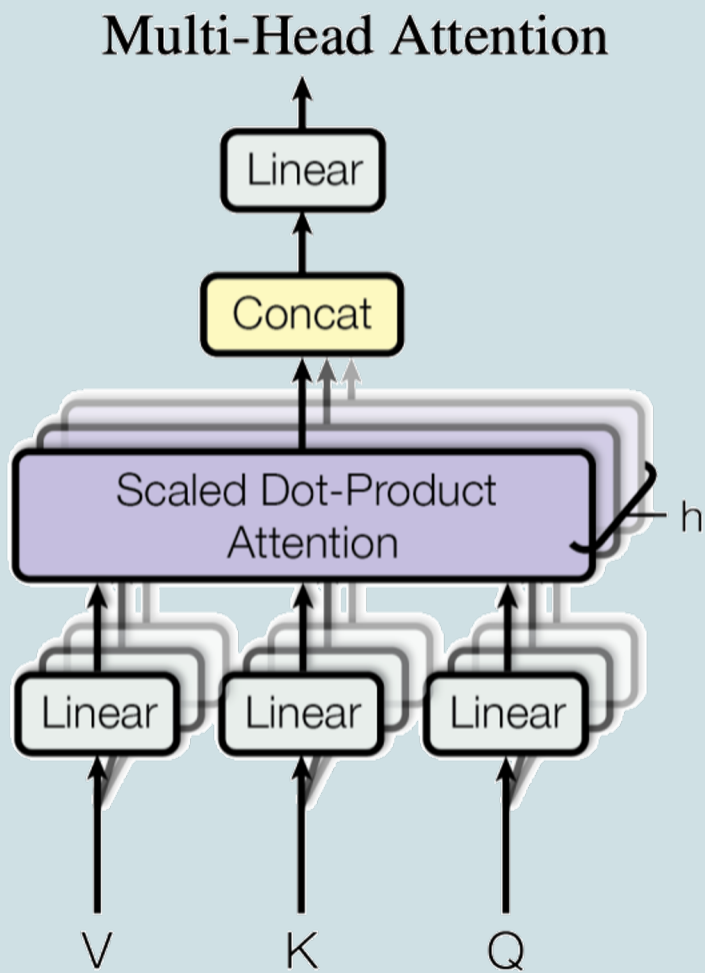
备胎2

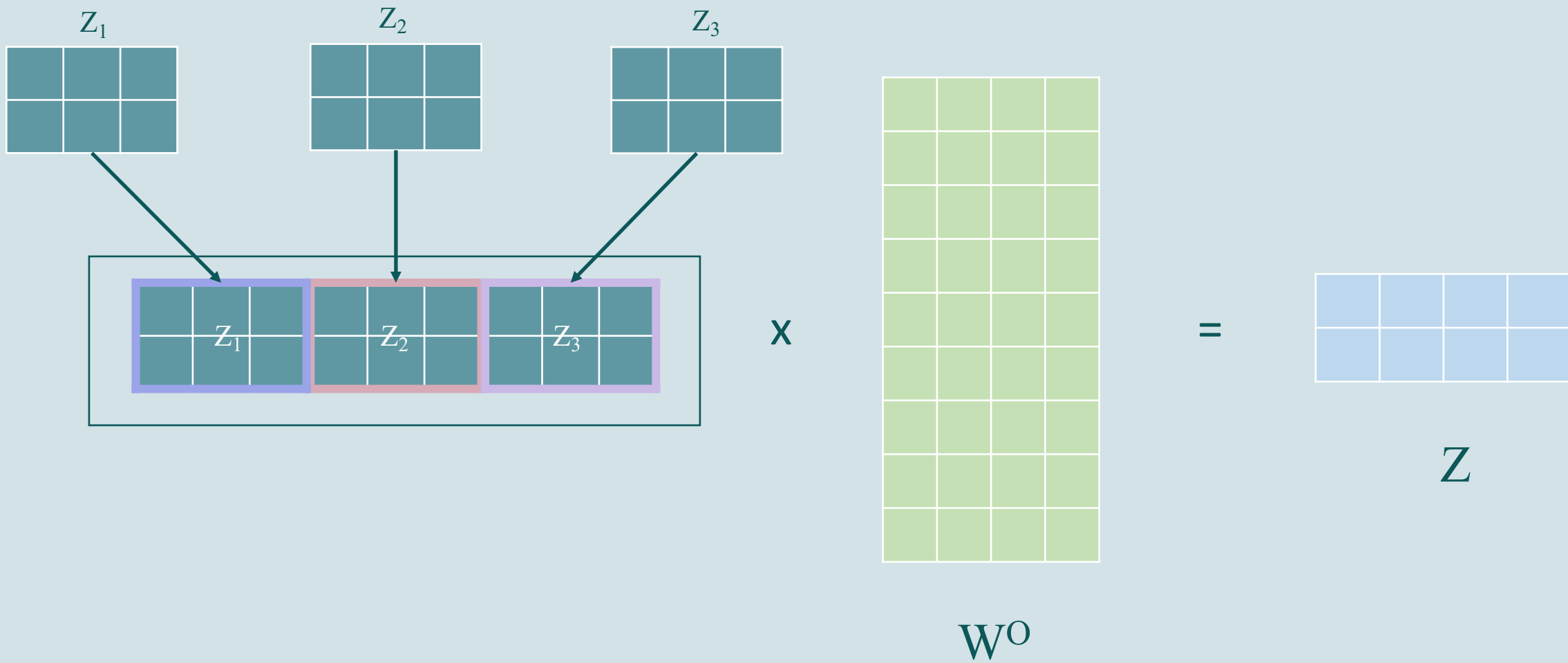
备胎3

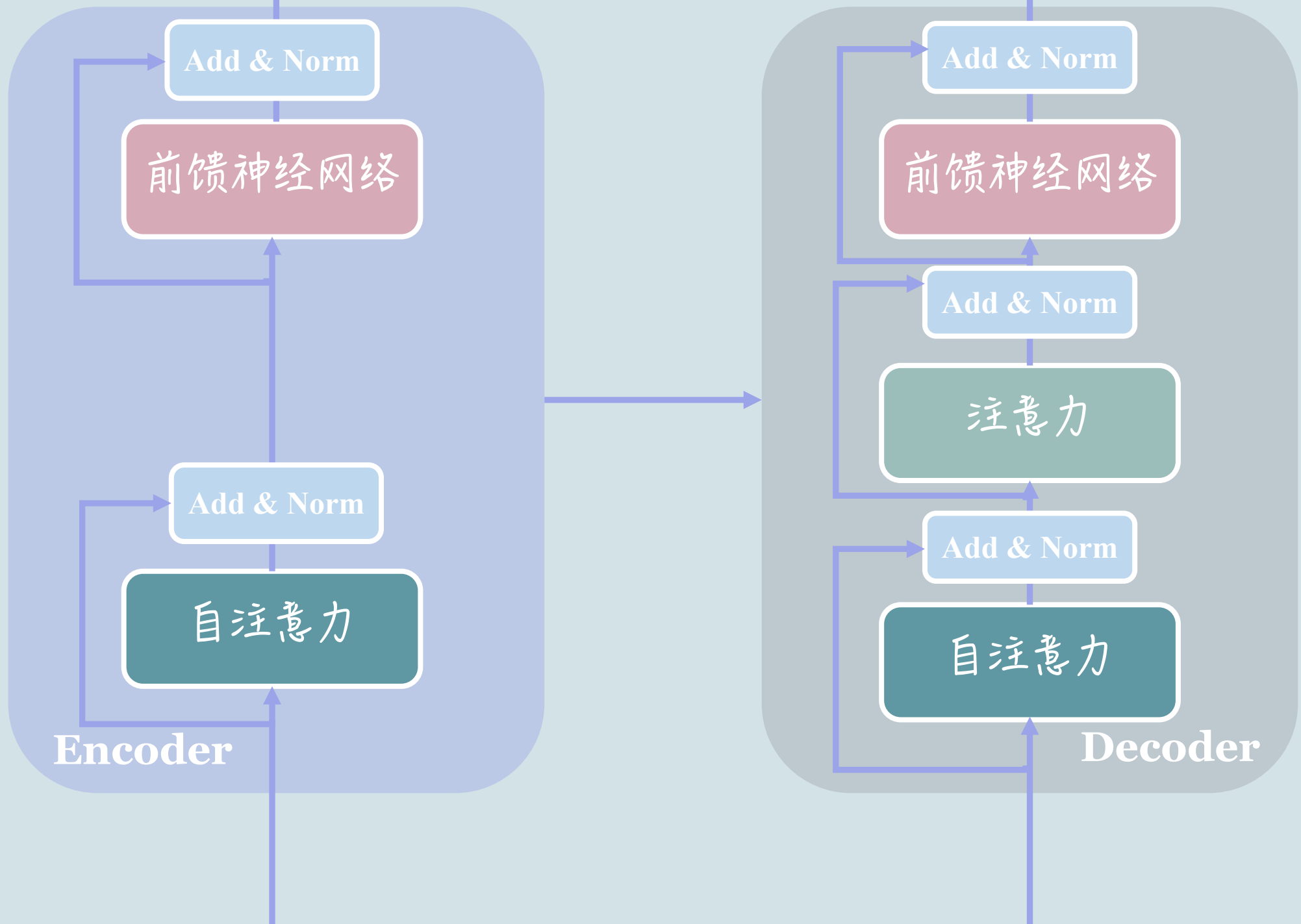
备胎N

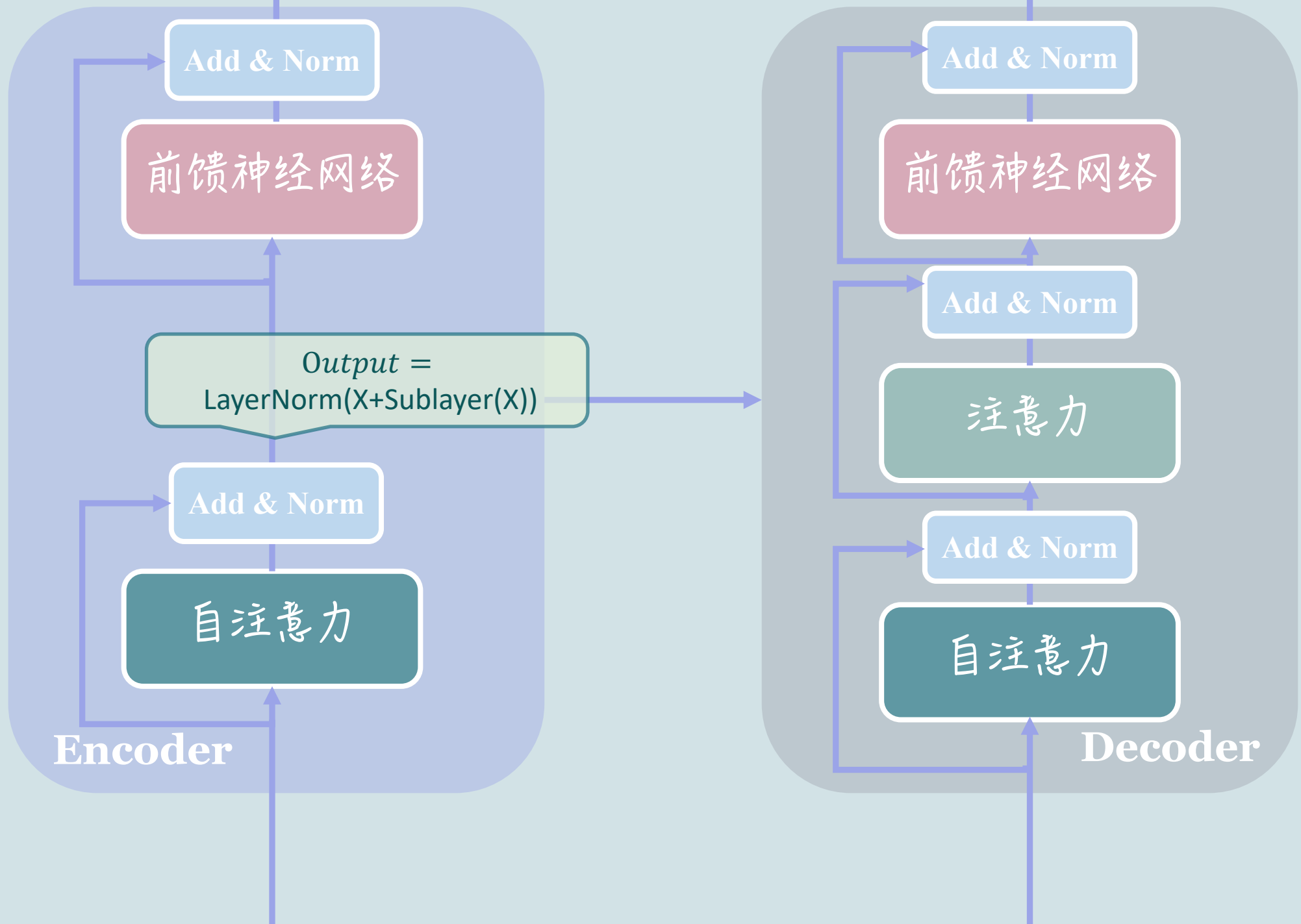
$\sqrt{d_k}$: k 的维度的平方根

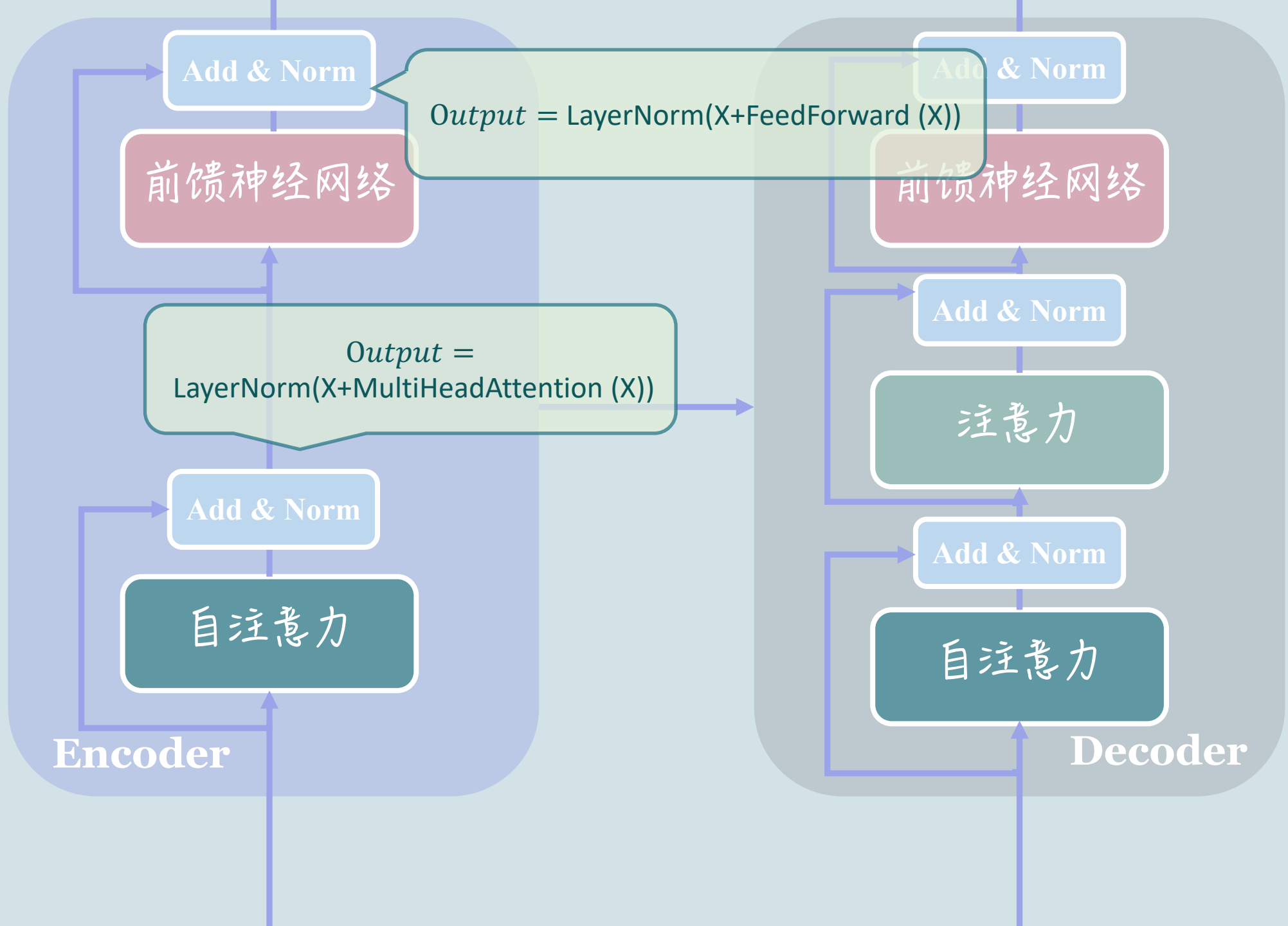


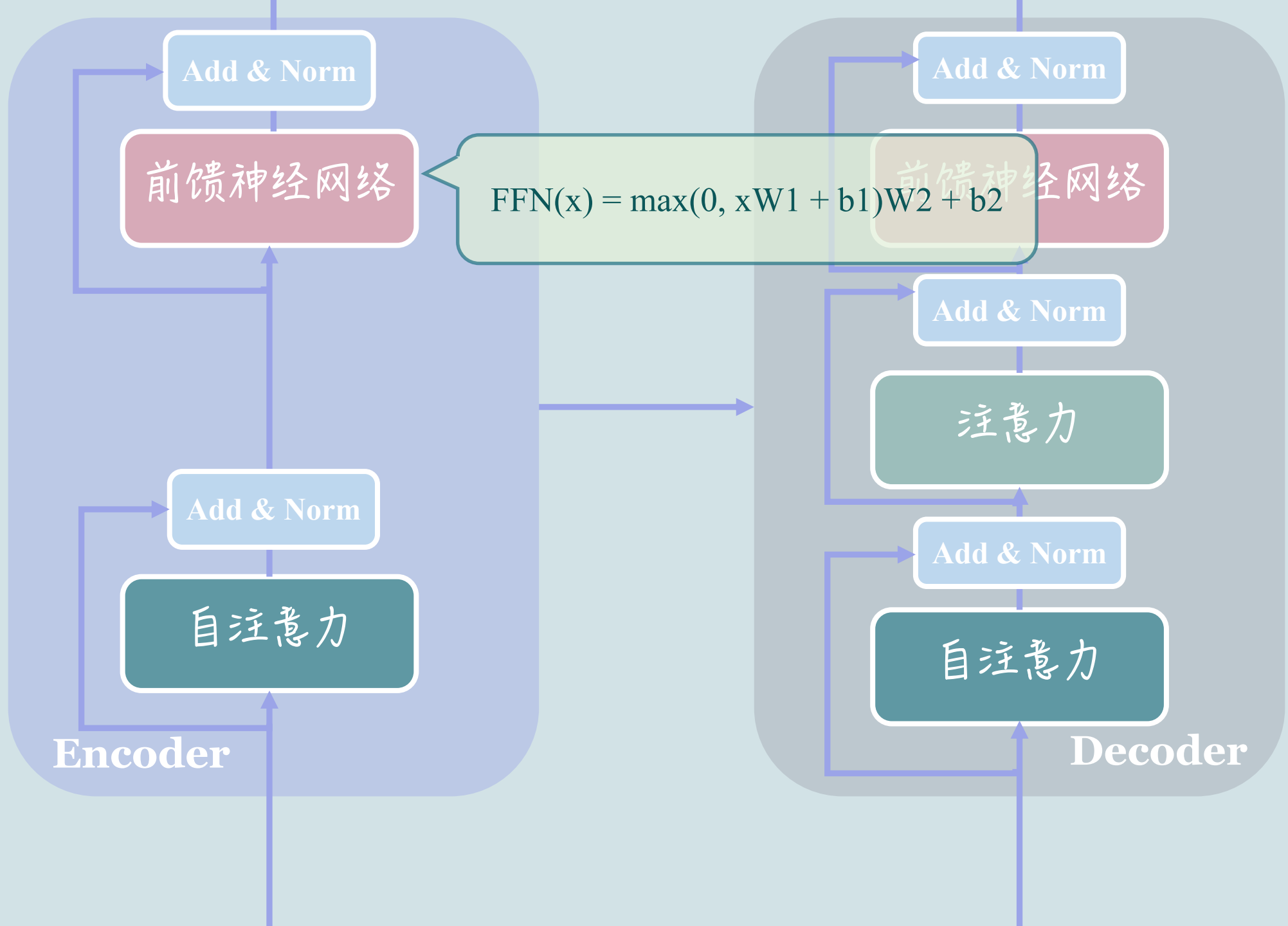


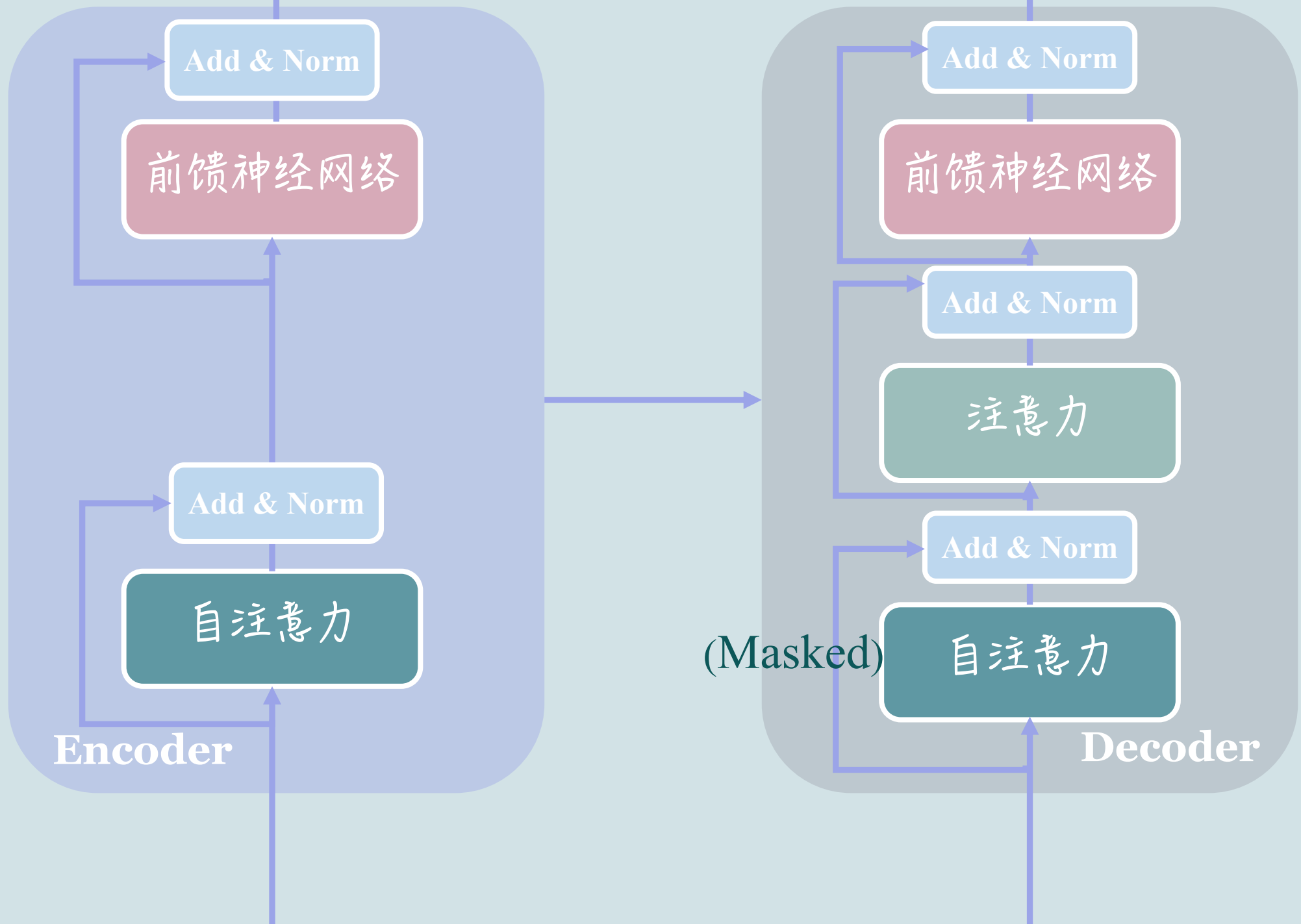


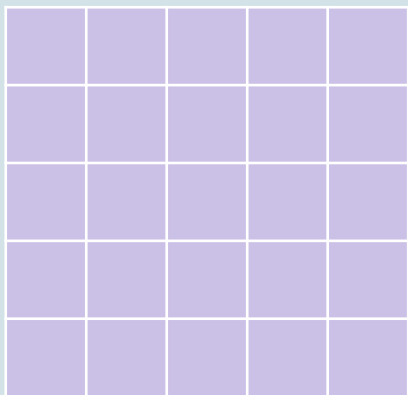












QK^T

| | | | | |
|--|--|--|--|--|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

QK^T

| | | | | |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 |

Mask

| | | | | |
|--|--|--|--|--|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

QK^T

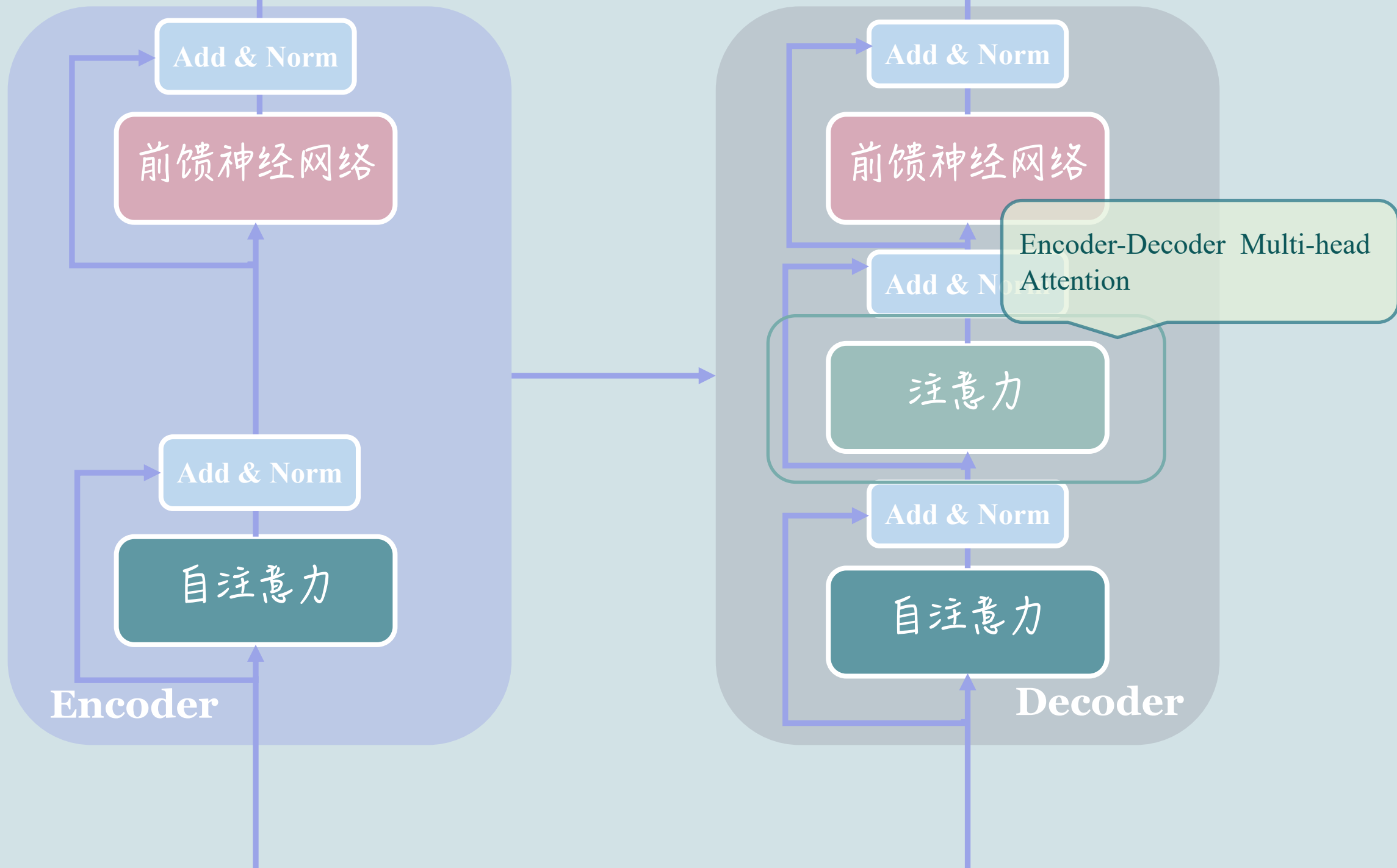
对位相乘

| | | | | |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 0 |
| 1 | 1 | 1 | 1 | 1 |

Mask

$-\infty$

| | | | | |
|--|--|--|--|--|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |



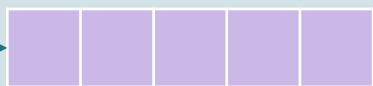


X

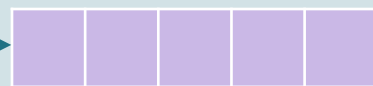
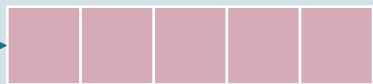


Positional
Encoding

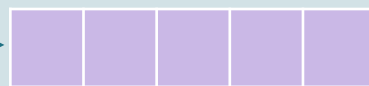
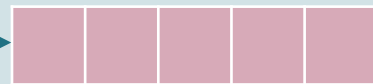
Embedding



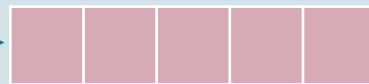
+



+



+



输入

<begin>

我

是

狗

<end>

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

<begin> 我 是 狗 <end>

$pos=0$ $pos=1$ $pos=2$ $pos=3$ $pos=4$

$$d_{\text{model}}/h = d_k = d_v$$

$$i \in [0, \frac{d_{\text{model}}}{2} - 1]$$

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

$$d_{\text{model}}/h = d_k = d_v$$

$$i \in [0, \frac{d_{\text{model}}}{2} - 1]$$

<begin> 我 是 狗 <end>

pos=0 pos=1 pos=2 pos=3 pos=4

$$\sin(A + B) = \sin A \cdot \cos B + \cos A \cdot \sin B$$

$$\cos(A + B) = \cos A \cdot \cos B - \sin A \cdot \sin B$$

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

$$d_{\text{model}}/h = d_k = d_v$$

$$i \in [0, \frac{d_{\text{model}}}{2} - 1]$$

<begin> 我 是 狗 <end>

$pos=0$ $pos=1$ $pos=2$ $pos=3$ $pos=4$

$$\sin(A + B) = \sin A \cdot \cos B + \cos A \cdot \sin B$$

$$\cos(A + B) = \cos A \cdot \cos B - \sin A \cdot \sin B$$



$$PE_{(pos+k,2i)} = PE_{(pos,2i)} \cdot PE_{(k,2i+1)} + PE_{(pos,2i+1)} \cdot PE_{(k,2i)}$$

$$PE_{(pos+k,2i+1)} = PE_{(pos,2i+1)} \cdot PE_{(k,2i+1)} - PE_{(pos,2i)} \cdot PE_{(k,2i)}$$