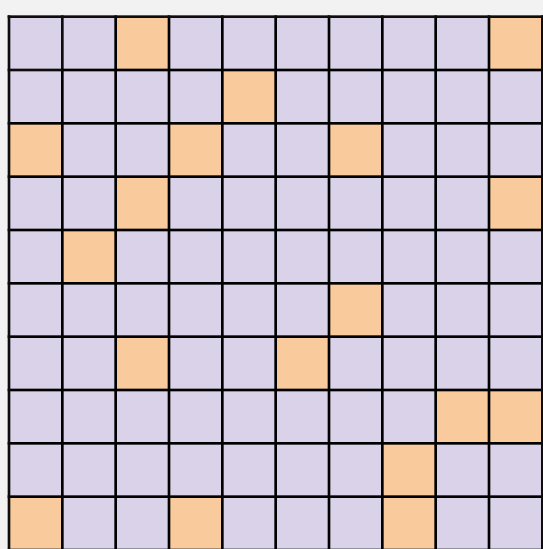


Graph Attention

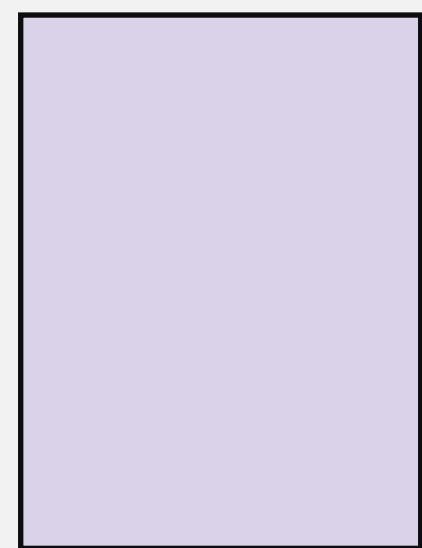
A Walkthrough



@



=



Inputs

```
# hyperparameters
```

```
n_nodes          = 10
```

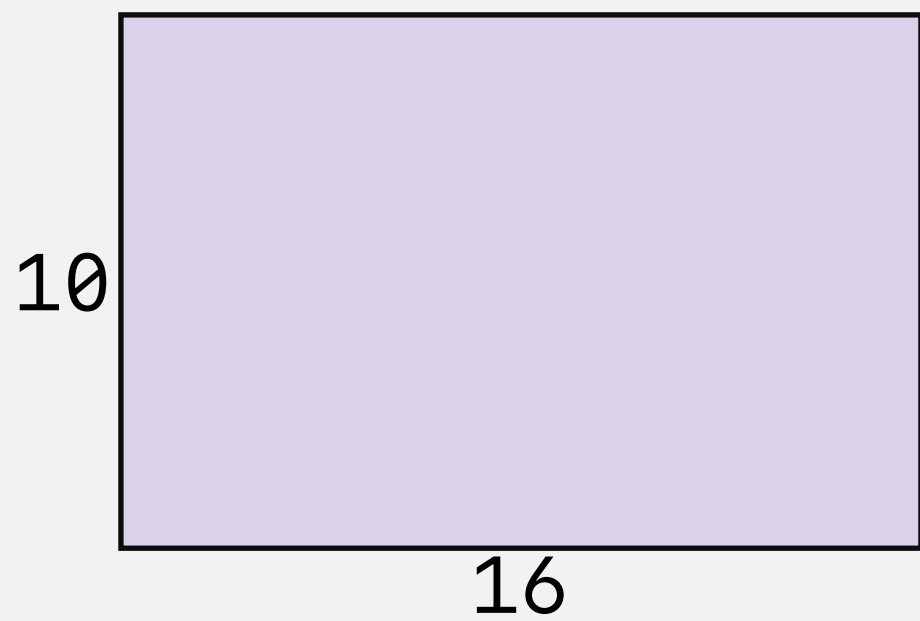
```
in_features      = 16
```

```
hidden_features  = 8
```

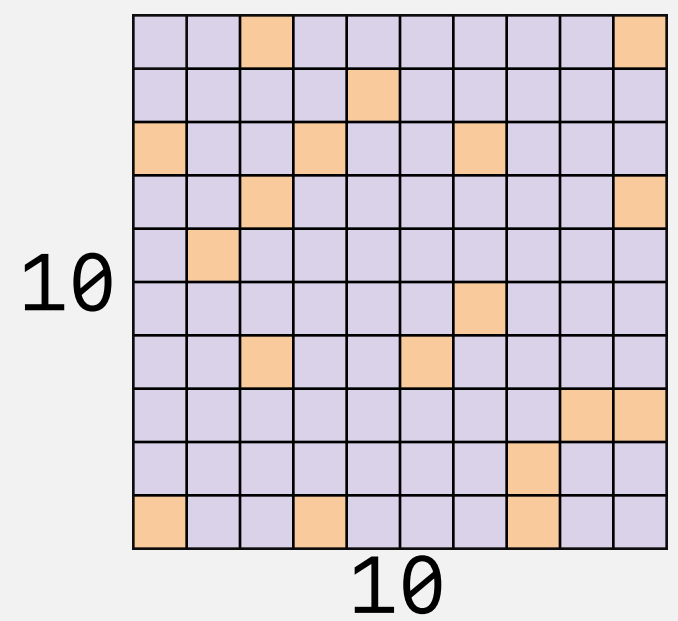
```
n_classes       = 7
```

```
n_heads         = 4
```

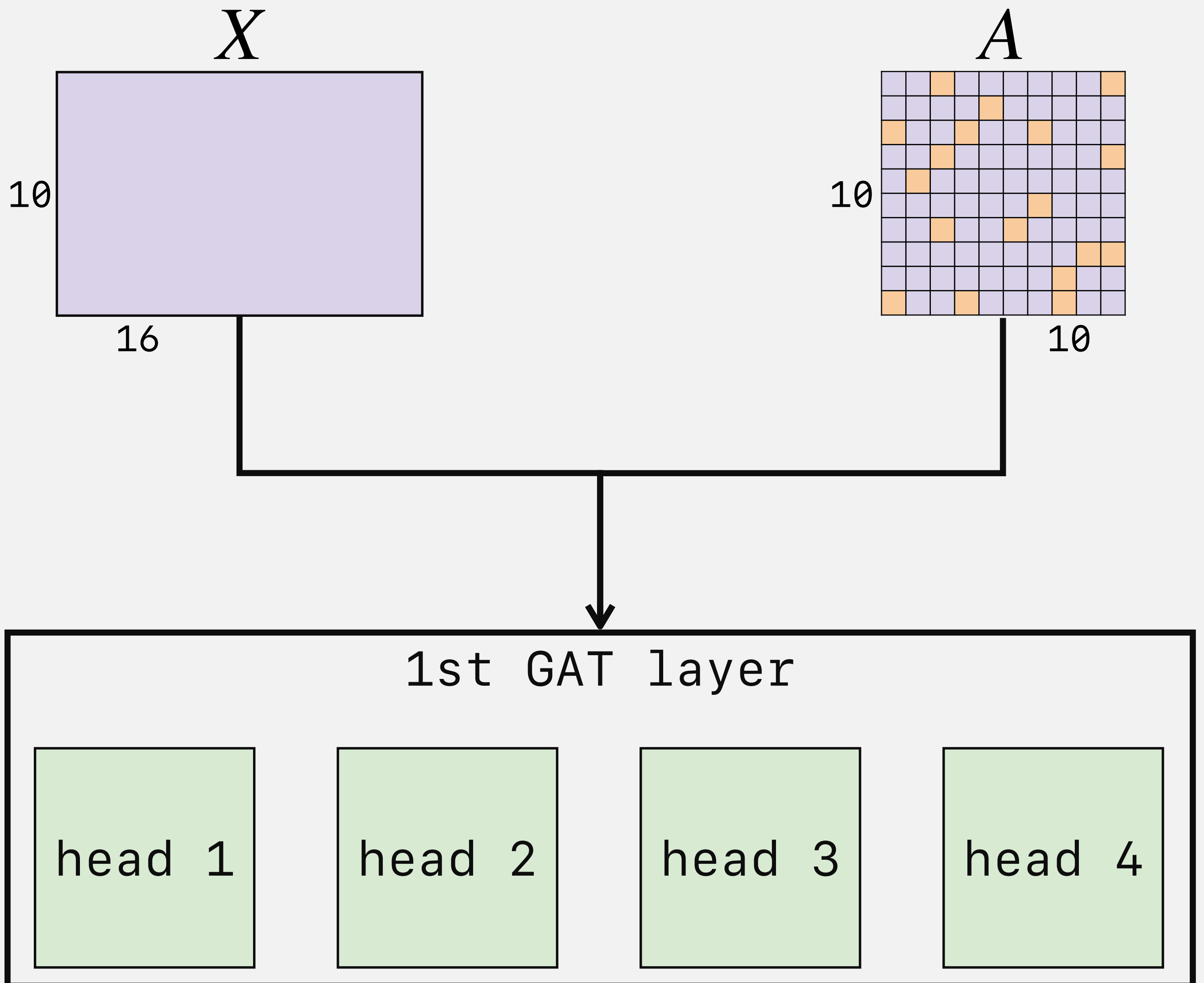
X



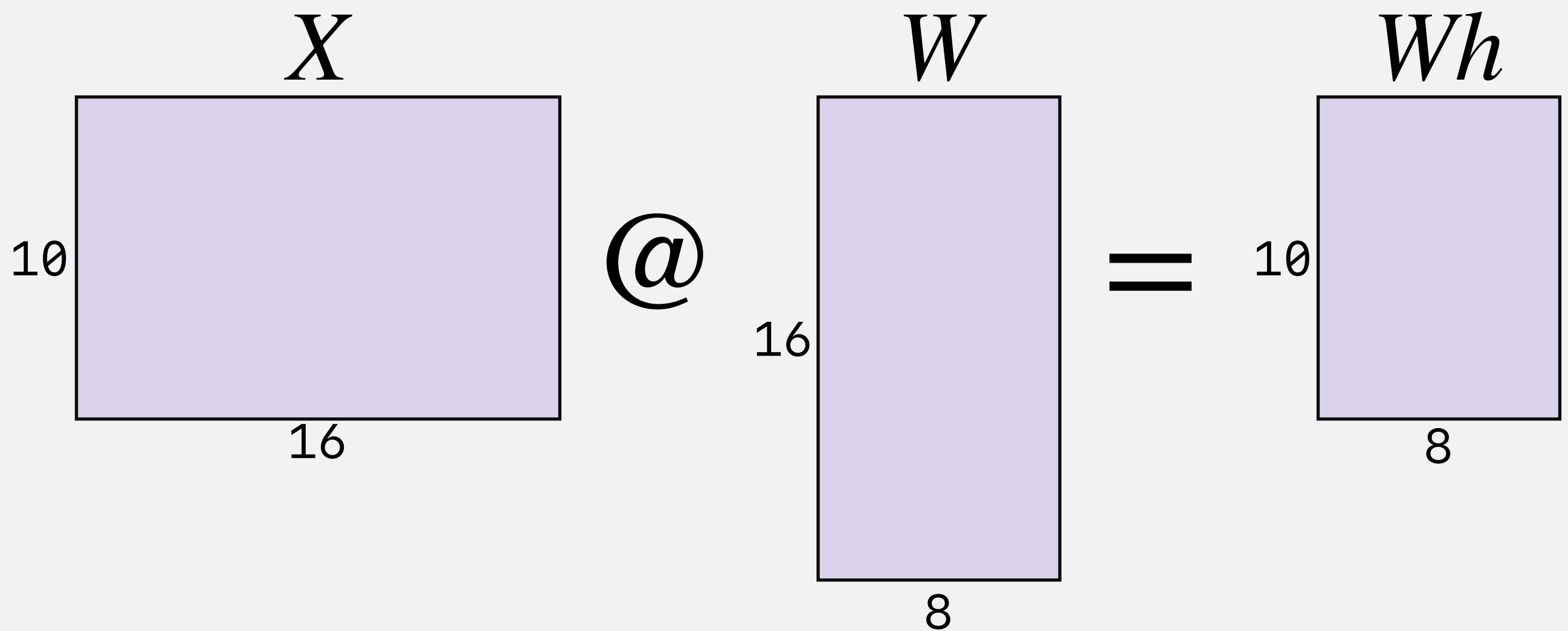
A



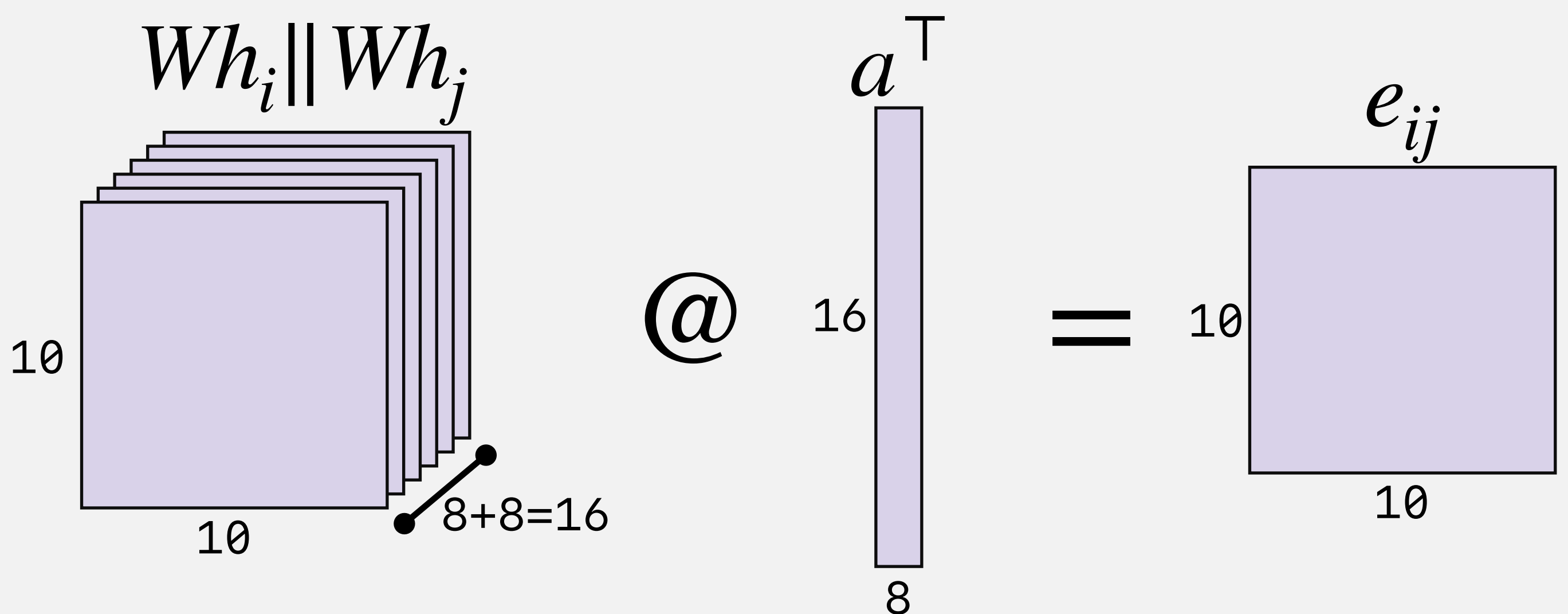
1st GAT layer, input



1st GAT layer, head 1



After applying the weight matrix W , the resulting node representations are denoted as Wh . For any two nodes i and j , the concatenation $Wh_i || Wh_j$ combines their feature vectors along the feature dimension.



1st GAT layer, head 1

$$\text{LeakyReLU} \left(\begin{array}{c} e_{ij} \\ \text{10} \times \text{10} \end{array} \right)$$

```
# masked attention  
torch.where(adj > 0, e, zero_vec)
```

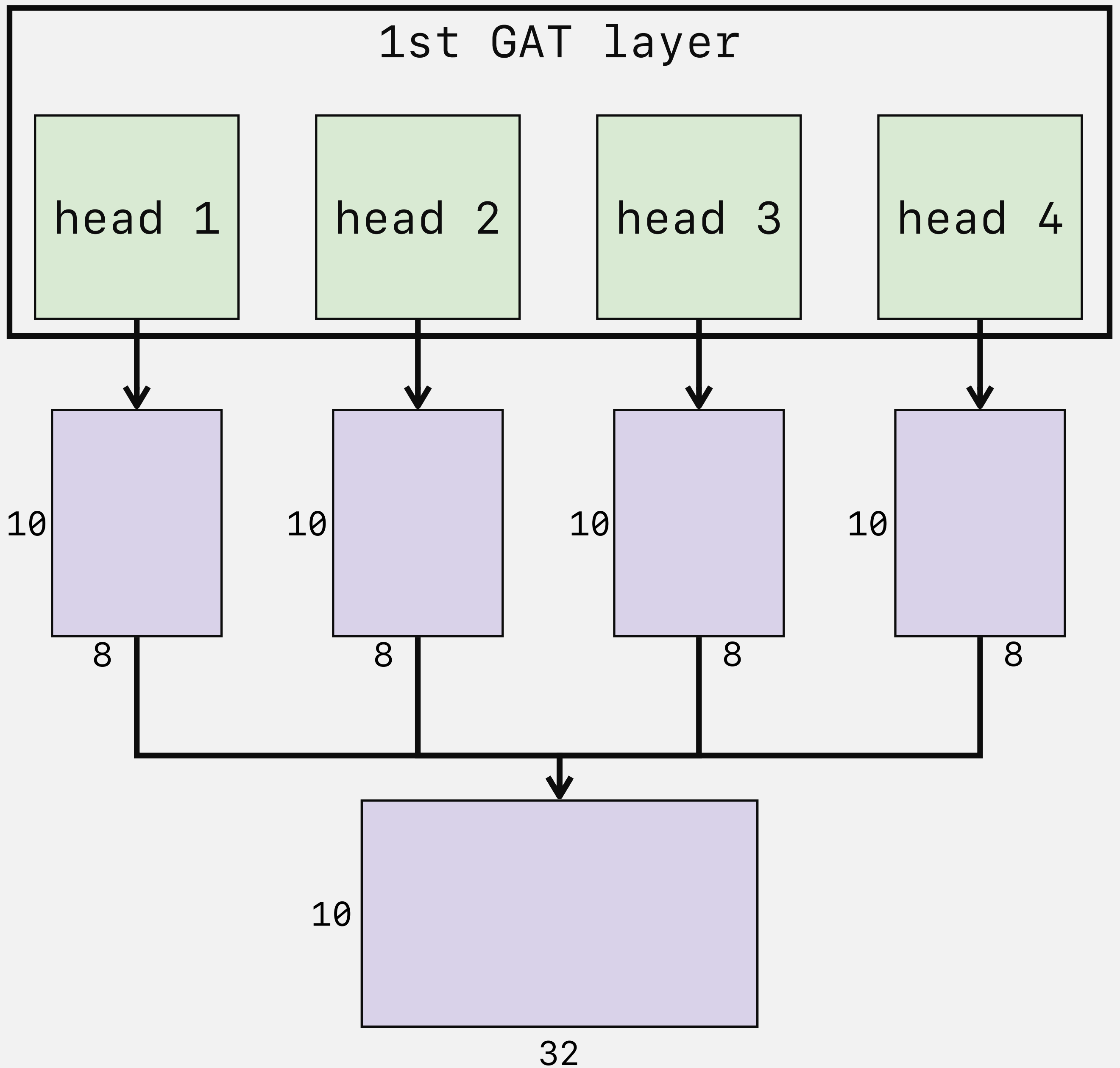
Masked attention must be applied before the softmax function. Reversing this order leads to softmax being calculated over incorrect values.

1st GAT layer, head 1

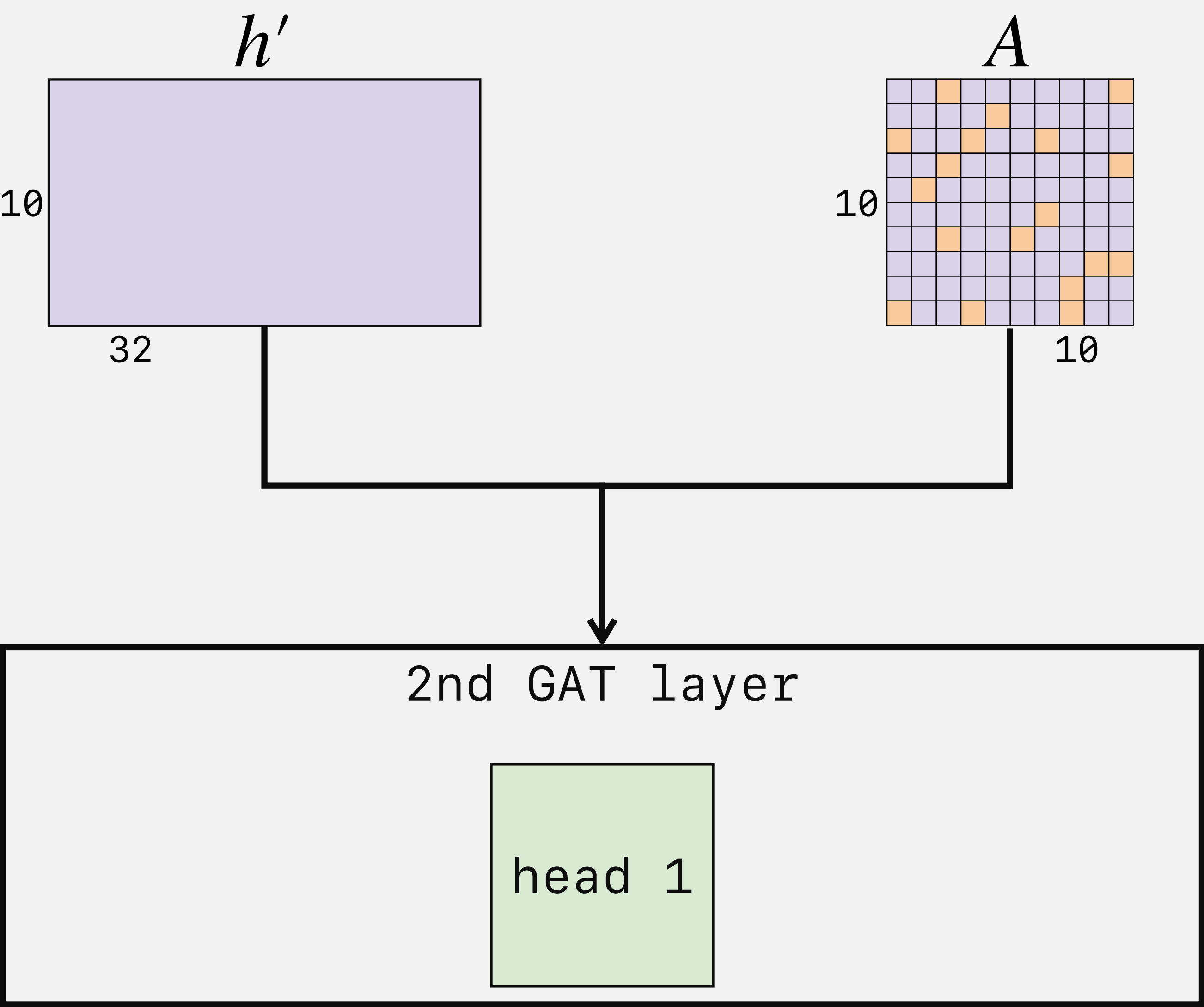
$$\text{SoftMax} \left(\begin{matrix} & & e_{ij} & & \\ & & \begin{matrix} \text{10} \times \text{10} \text{ grid} \end{matrix} & & \\ & & & & \end{matrix} \right) = \begin{matrix} & & \alpha_{ij} & & \\ & & \begin{matrix} \text{10} \times \text{10} \text{ grid} \end{matrix} & & \\ & & & & \end{matrix}$$

$$\begin{matrix} & & \alpha_{ij} & & \\ & & \begin{matrix} \text{10} \times \text{10} \text{ grid} \end{matrix} & & \\ & & & & \end{matrix} @ \begin{matrix} & & Wh & & \\ & & \begin{matrix} \text{10} \times 8 \text{ grid} \end{matrix} & & \\ & & & & \end{matrix} = \begin{matrix} & & h' & & \\ & & \begin{matrix} \text{10} \times 8 \text{ grid} \end{matrix} & & \\ & & & & \end{matrix}$$

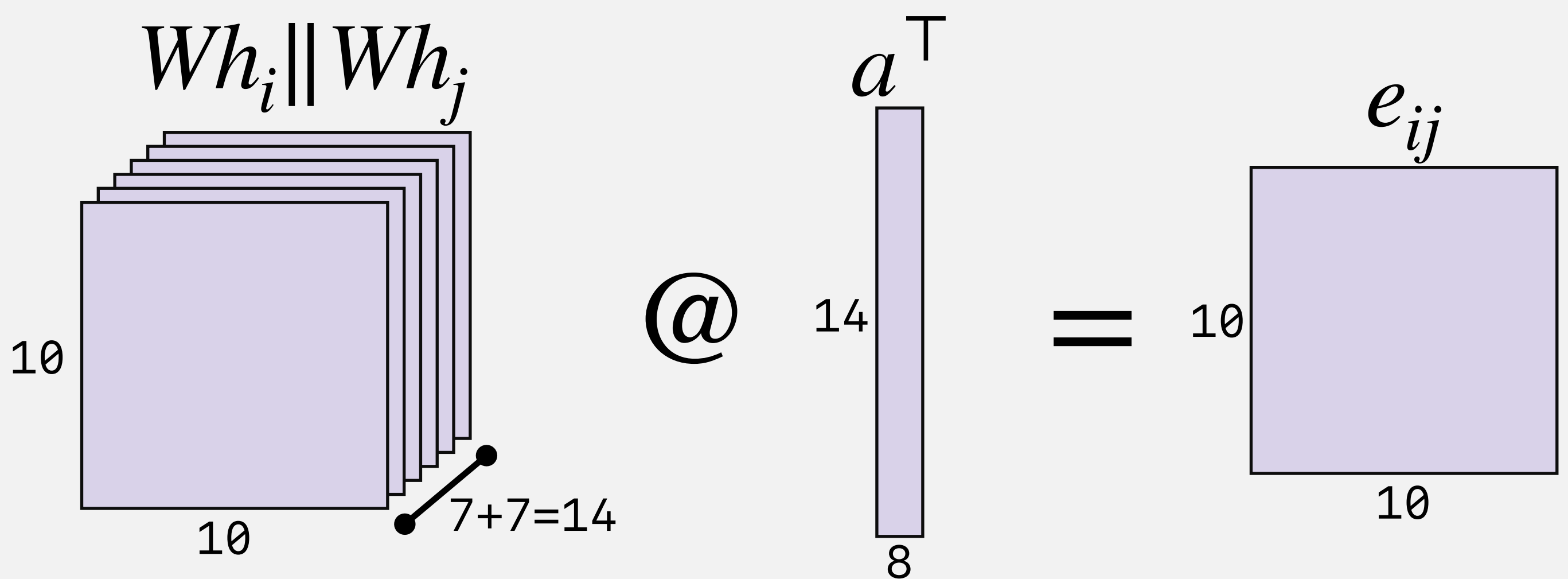
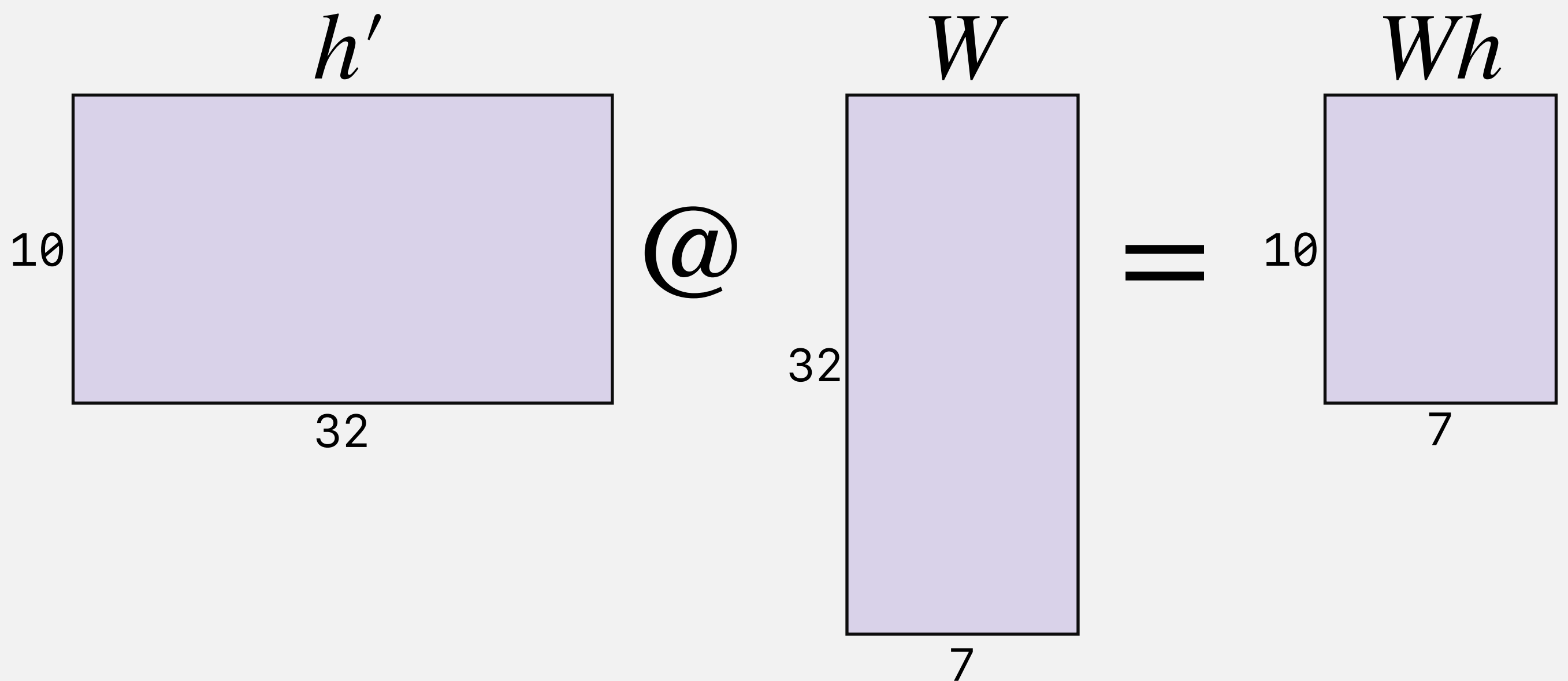
1st GAT layer, output



2nd GAT layer, input



2nd GAT layer, head 1



2nd GAT layer, head 1

$$\text{LeakyReLU} \left(\begin{array}{c} e_{ij} \\ \text{10} \times \text{10} \end{array} \right)$$

```
# masked attention  
torch.where(adj > 0, e, zero_vec)
```

2nd GAT layer, head 1

$$\text{SoftMax} \left(\begin{matrix} & e_{ij} \\ 10 & \begin{matrix} \text{10x10 grid with orange squares} \end{matrix} \\ & 10 \end{matrix} \right) = \begin{matrix} & \alpha_{ij} \\ 10 & \begin{matrix} \text{10x10 grid with orange squares} \end{matrix} \\ & 10 \end{matrix}$$

$$\begin{matrix} & \alpha_{ij} \\ 10 & \begin{matrix} \text{10x10 grid with orange squares} \end{matrix} \\ & 10 \end{matrix} @ \begin{matrix} & Wh \\ 10 & \begin{matrix} \text{10x7 purple rectangle} \end{matrix} \\ & 7 \end{matrix} = \begin{matrix} & h' \\ 10 & \begin{matrix} \text{10x7 purple rectangle} \end{matrix} \\ & 7 \end{matrix}$$

number of classes ←