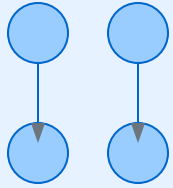


# Mathematical Framework for Bridge Neural Networks

## Modified Transformer Equations



Standard Transformer Flow:

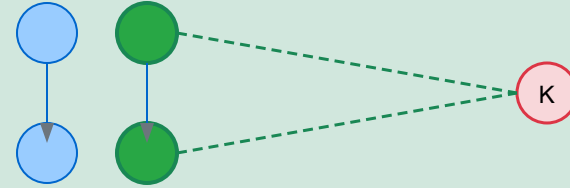
$$H^{(l+1)} = \text{TransformerLayer}_l(H^{(l)})$$

Bridge Neural Network Flow:

$$H^{(l+1)} = \text{TransformerLayer}_l(H^{(l)}) + \mathbb{1}_{\{B^{(l)}\}} \cdot l(K(Q(a_{\{B^{(l)}\}})))$$

## Bridge Activation Function

Activation Indicator:



$$\mathbb{1}_{\{B^{(l)}\}} = \begin{cases} 1, & \text{if } D(a_{\{B^{(l)}\}}) > \tau \\ 0, & \text{otherwise} \end{cases}$$

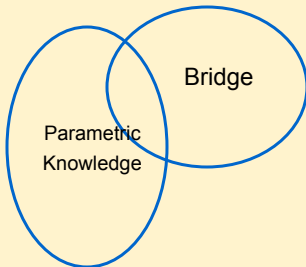
Activation Probability:

$$P(B^{(l)} | X) = \sigma(W_d \cdot a_{\{B^{(l)}\}} + b_d)$$

## Information Theoretic Perspective

Information Utility Function:

$$U(X, B, K) = I(Y; X, B, K) - \lambda C(B, K)$$



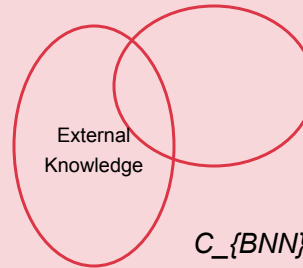
Activation Condition:

$$\nabla_B U(X, B, K) > 0$$

Decision Boundary

$$D(a_{\{B^{(l)}\}}) = \tau$$

## Capacity Analysis



Optimal Bridge Allocation:

$$|B|^* = (1/\beta)((\alpha\beta/\lambda) - 1)$$

Effective Capacity:

$$C_{\{BNN\}} = C_P + \sum_{l=1}^L \mathbb{E}[\mathbb{1}_{\{B^{(l)}\}}] \cdot C_N$$