Product Graph Learning (PGL) layer with  $\theta_{PGL}$  parameters input  $\mathcal{L}_{PGL} = \sum_{r=0}^{F_{de}} \left[ \sum_{r=0}^{P} \left\| \mathbf{X}^{(0)}(r,:,f) - \mathbf{X}^{(0)}(s,:,f) \right\|_{2}^{2} \mathbf{W}_{P}(r,s) + \sum_{r=0}^{Q} \left\| \mathbf{X}^{(0)}(:,r',f) - \mathbf{X}^{(0)}(:,s',f) \right\|_{2}^{2} \mathbf{W}_{Q}(r',s') \right] + \lambda \left( \left\| \mathbf{W}_{P} \right\|_{F}^{2} + \left\| \mathbf{W}_{Q} \right\|_{F}^{2} \right)$ Spatial Attention with  $\theta_{SpAtt}$  parameters MGwAT with *H* heads &  $\theta_{GWAT}$  parameters  $\mathbf{X}^{(2)} \in \mathbb{R}^{P \times (QF')}$  $\mathcal{L}_{loss} = \mathcal{L}_{CE} + \mathcal{L}_{PGL}$  $\mathbf{X}^{(1)} \in \mathbb{R}^{P \times Q \times F'}$ PEROXO Flattening  $\mathbf{X}^{(0)} \in \mathbb{R}^{P \times Q \times F_{de}}$ AGC with F' kernels Nodes' features **BiGRU** with  $\boldsymbol{\beta}$  units &  $\theta_{AGC}$  params  $\mathbf{S}_{t+4}$ **BiGRU** with t-2**Sleep Stage Classification** BiGRU with  $\boldsymbol{\beta}$  units  $\mathbf{X}^{(3)} \in \mathbb{R}^{P \times \beta}$  $\theta_{BiGRU(\beta)}$  parameters