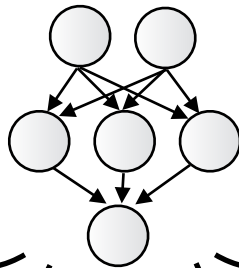
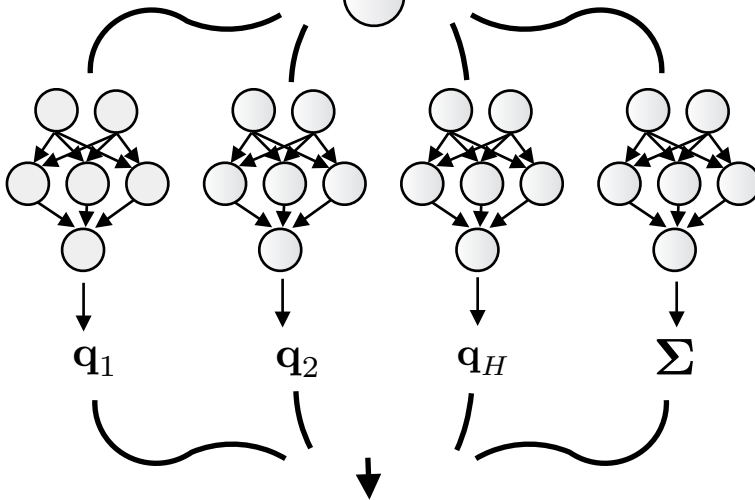


input



'body'

'heads'



1. Evaluate loss for H heads:  $\mathcal{L} = \sum_{h=1}^H \mathcal{L}_{\text{NLL}}(\mathbf{q}_h, \mathbf{q}_t, \Sigma)$

$$\mathcal{L}_{\text{NLL}}(\mathbf{q}, \mathbf{q}_t, \Sigma_a) = \frac{1}{2} \phi^\top \Sigma_a^{-1} \phi + \frac{1}{2} \log \det(\Sigma_a),$$

$$\phi = \text{Log}(\mathbf{q} \otimes \mathbf{q}_t^{-1})$$