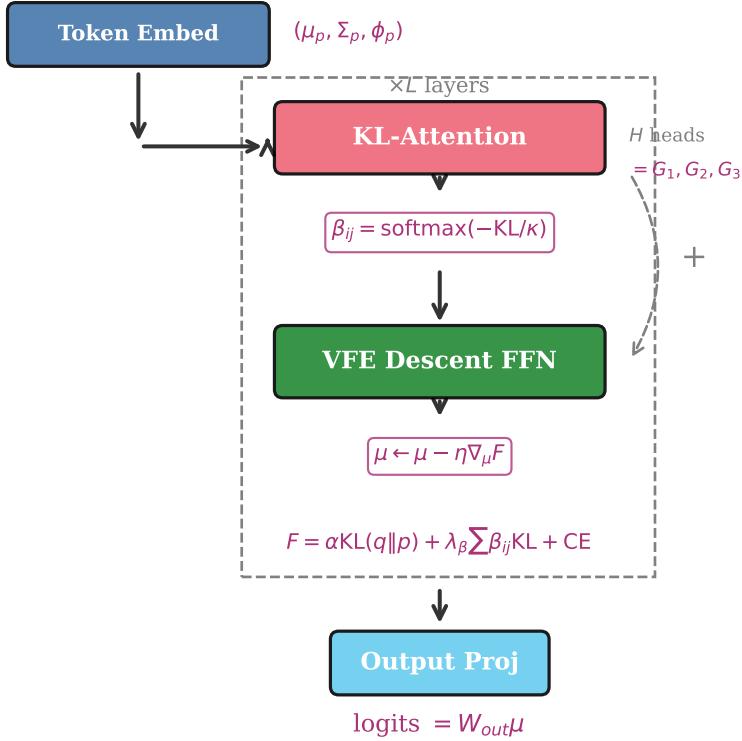
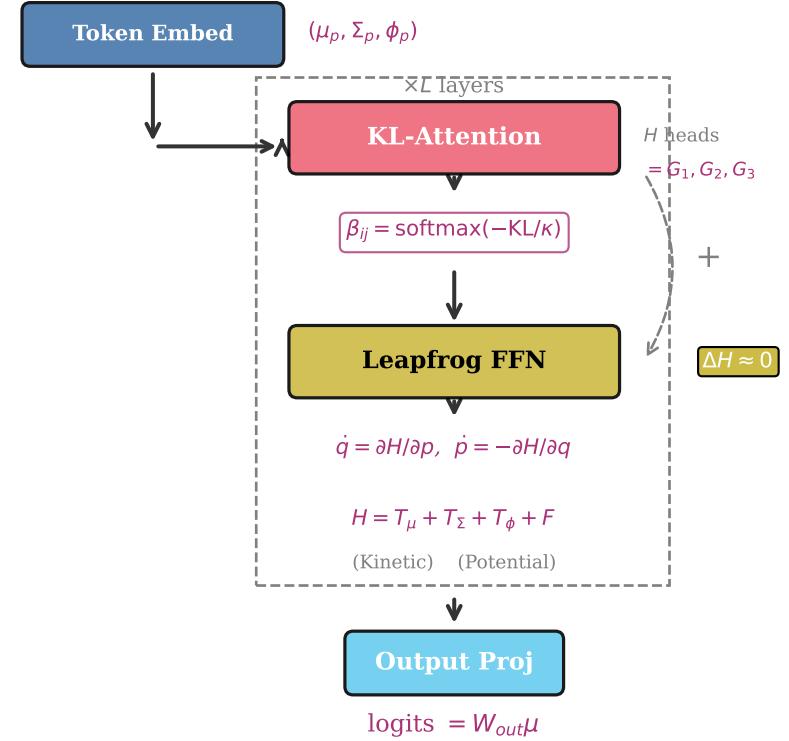


# Gauge Transformer Architectures: VFE vs Hamiltonian Dynamics

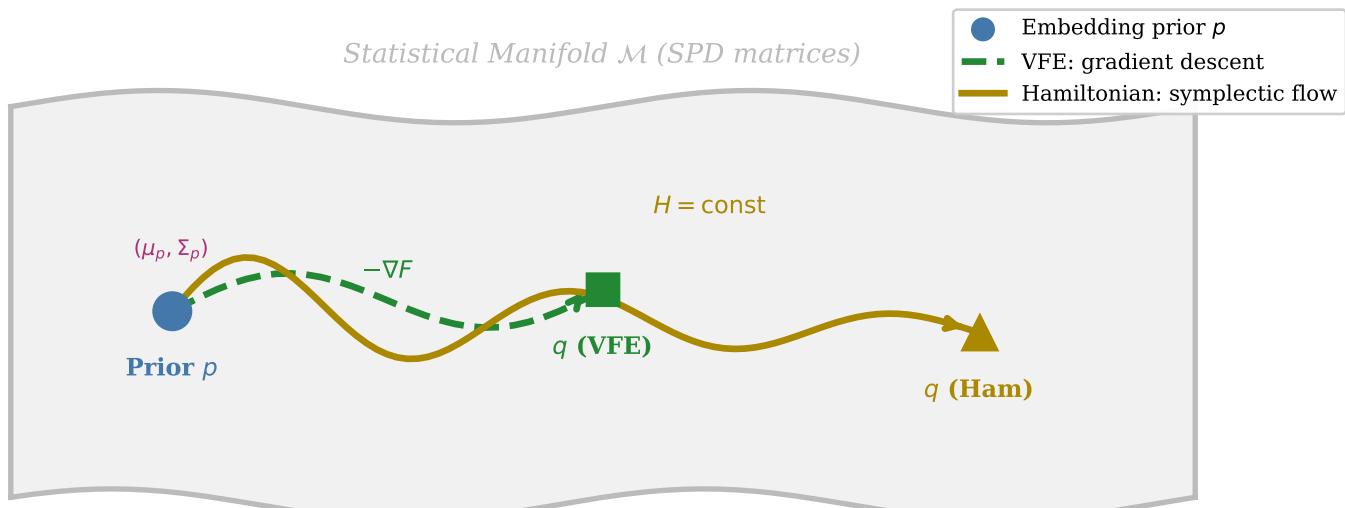
**(a) Gauge-VFE Transformer**



**(b) Gauge-Hamiltonian Transformer**



**(c) Belief Dynamics on Statistical Manifold**



Both architectures use KL-divergence based attention with gauge-equivariant parallel transport. VFE descent minimizes free energy via gradient flow. Hamiltonian dynamics conserves energy via symplectic integration. The self-consistency term  $\text{KL}(q \parallel p)$  anchors beliefs to embedding priors, enabling gradient flow to learn embeddings.