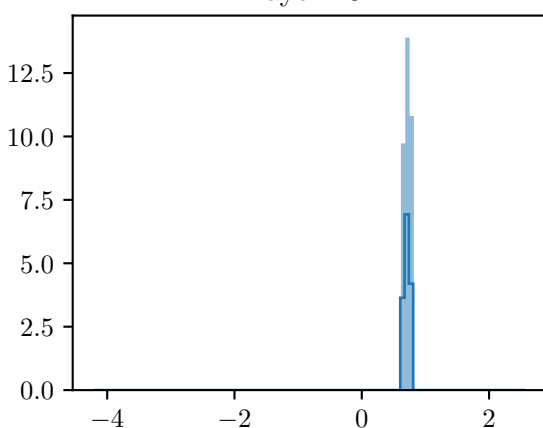
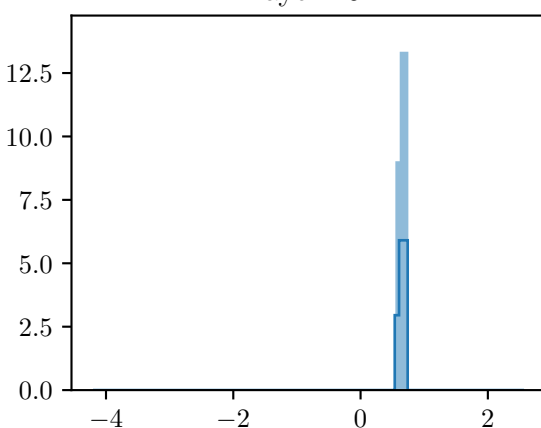


Histograms of attention logits $\frac{1}{\sqrt{d'}} \langle Qx_i, Kx_j \rangle$
(imdb_albert-xlarge-v2_params-random_dense-on_layers-48)

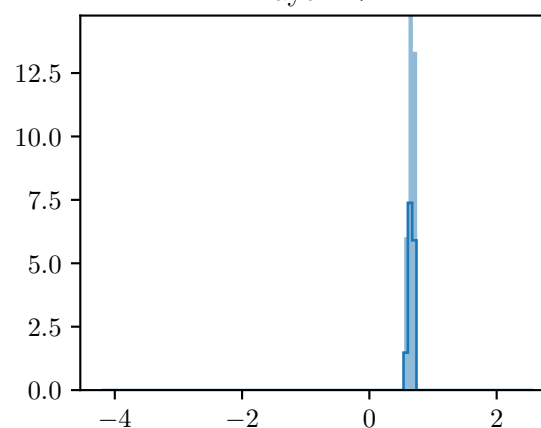
layer 25



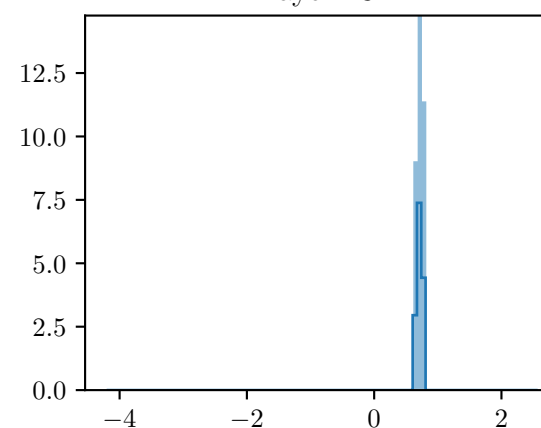
layer 26



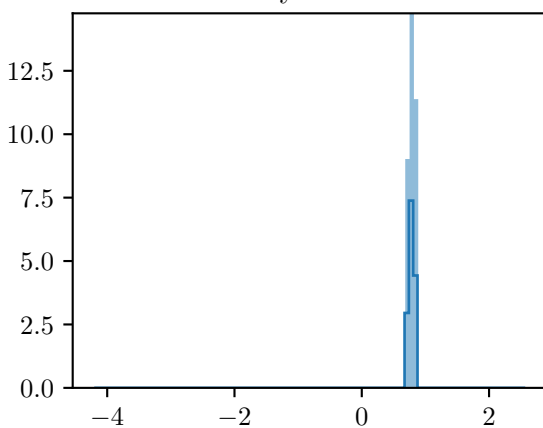
layer 27



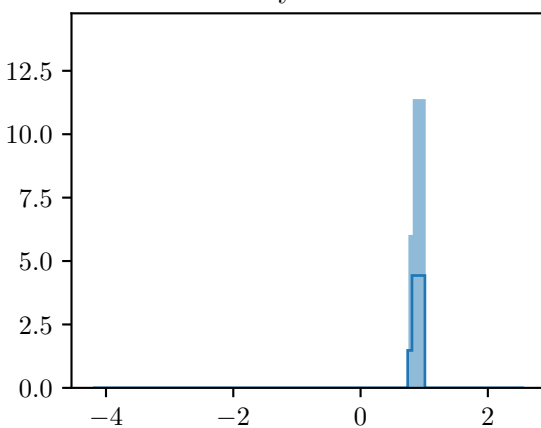
layer 28



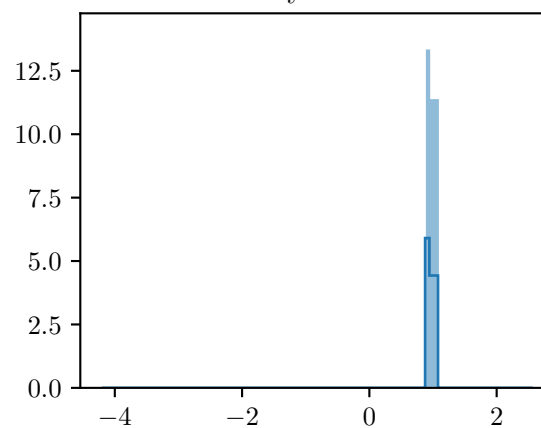
layer 29



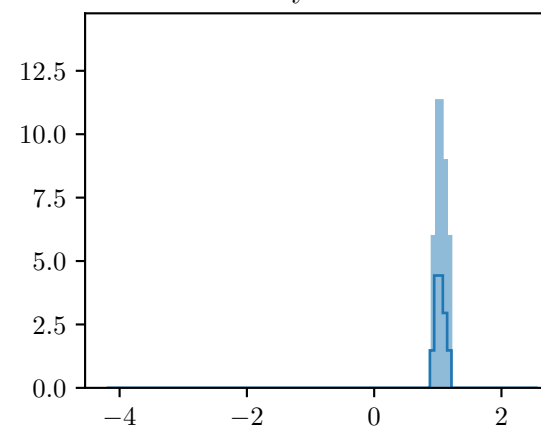
layer 30



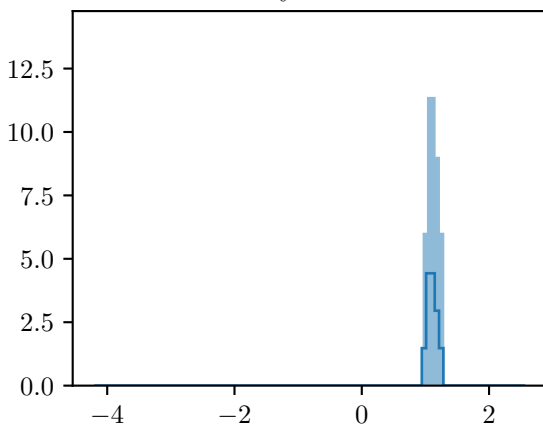
layer 31



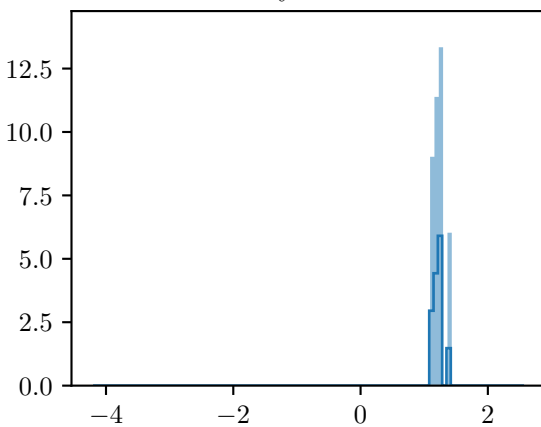
layer 32



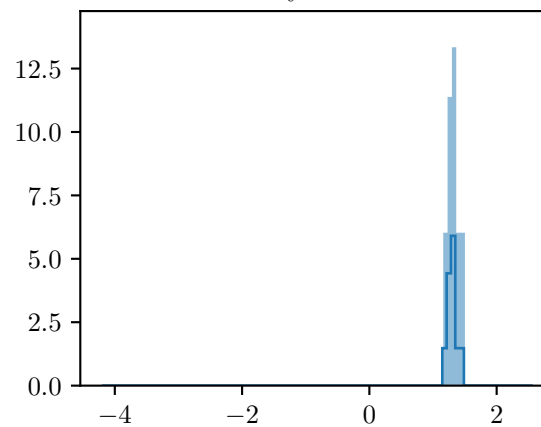
layer 33



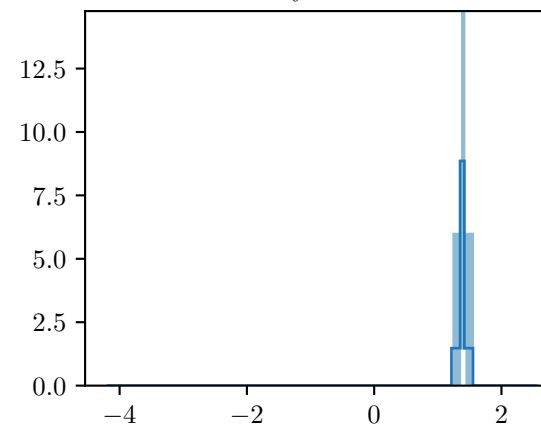
layer 34



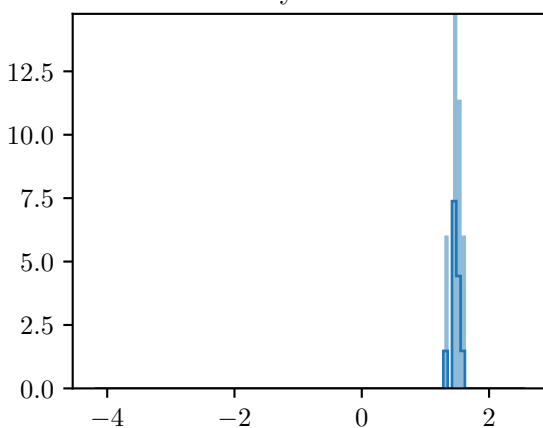
layer 35



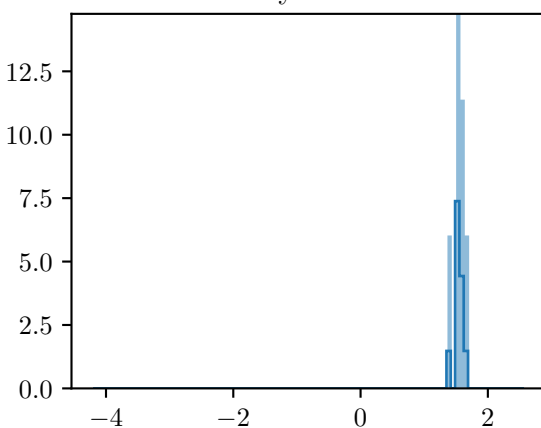
layer 36



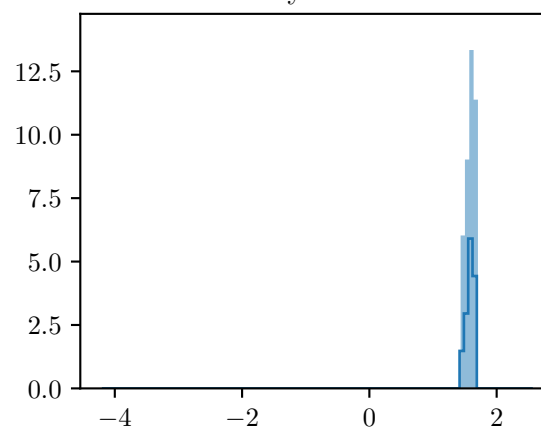
layer 37



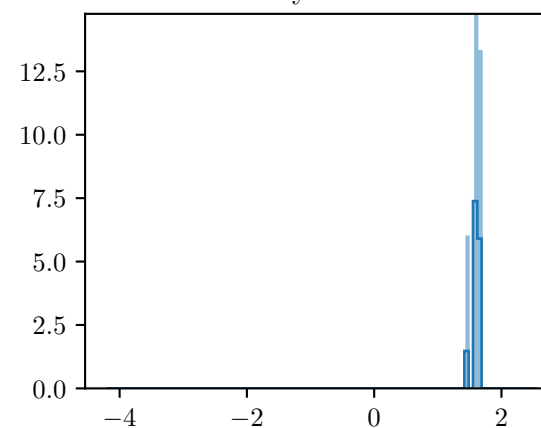
layer 38



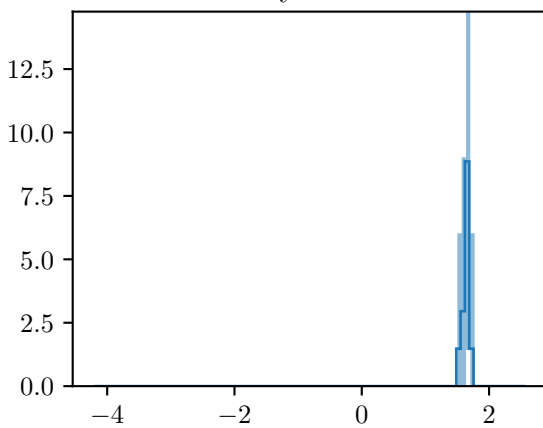
layer 39



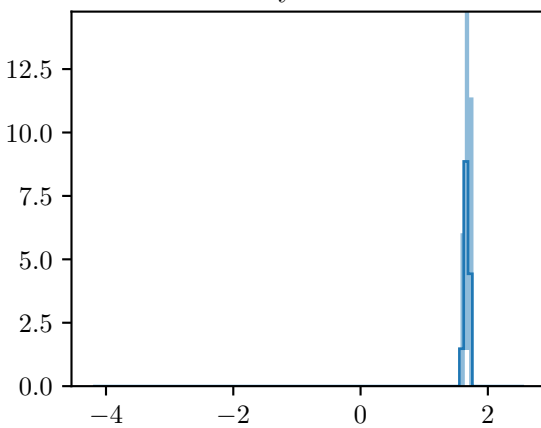
layer 40



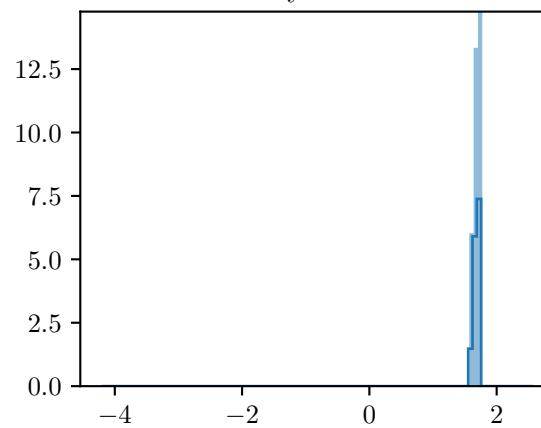
layer 41



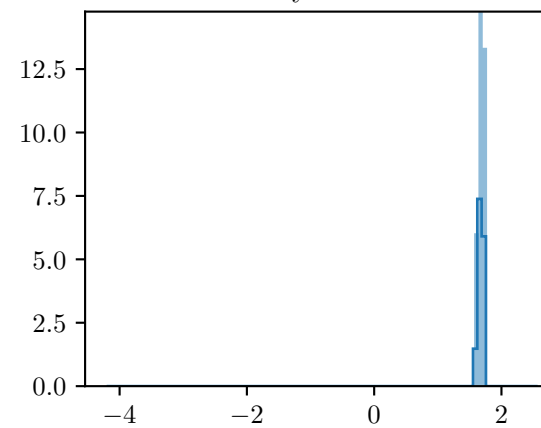
layer 42



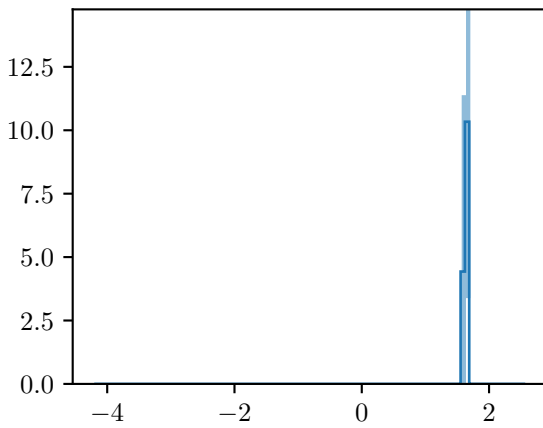
layer 43



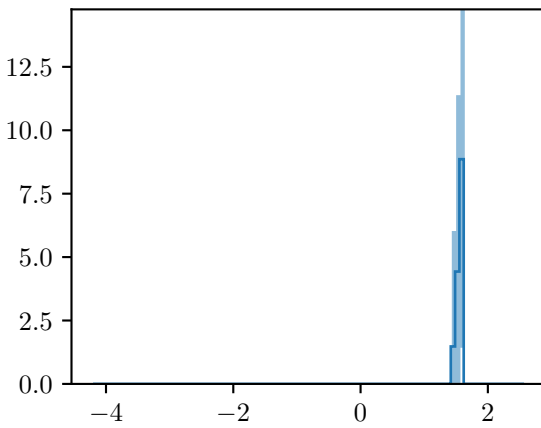
layer 44



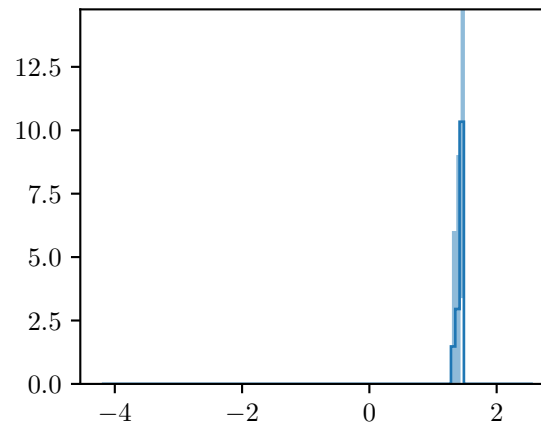
layer 45



layer 46



layer 47



layer 48

