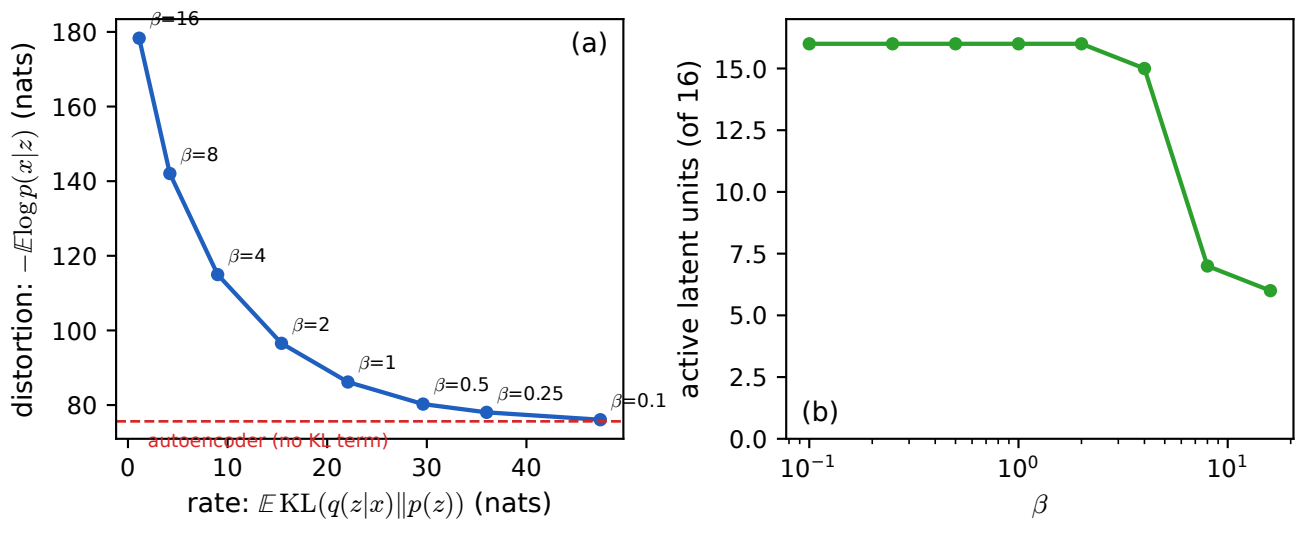

(a)
β=16
β=8
β=4
β=2
β=1
β=0.5
β=0.25
β=0.1
autoencoder (no KL term)
distortion: $-\mathbb{E}\log p(x|z)$ (nats)
rate: $\mathbb{E}\,\mathrm{KL}(q(z|x)\|p(z))$ (nats)
(b)
active latent units (of 16)
β