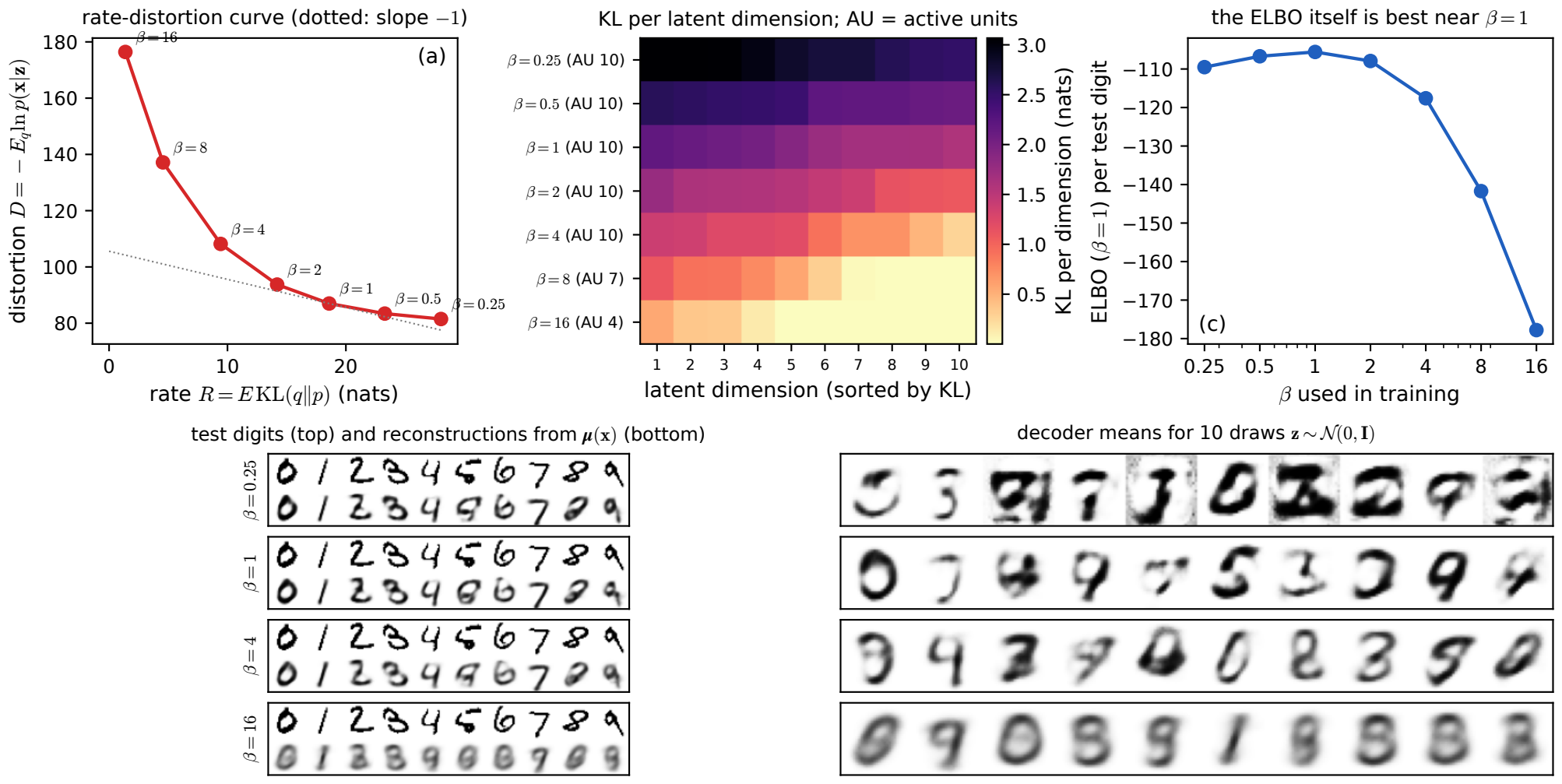

rate-distortion curve (dotted: slope −1)
(a)
β=16
β=8
β=4
β=2
β=1
β=0.5
β=0.25
distortion $D = -E_q \ln p(\mathbf{x}|\mathbf{z})$
rate $R = E\mathrm{KL}(q\|p)$ (nats)
KL per latent dimension; AU = active units
β=0.25 (AU 10)
β=0.5 (AU 10)
β=1 (AU 10)
β=2 (AU 10)
β=4 (AU 10)
β=8 (AU 7)
β=16 (AU 4)
KL per dimension (nats)
latent dimension (sorted by KL)
the ELBO itself is best near β=1
(c)
ELBO (β=1) per test digit
β used in training
test digits (top) and reconstructions from μ(x) (bottom)
β=0.25
β=1
β=4
β=16
decoder means for 10 draws z ~ N(0, I)