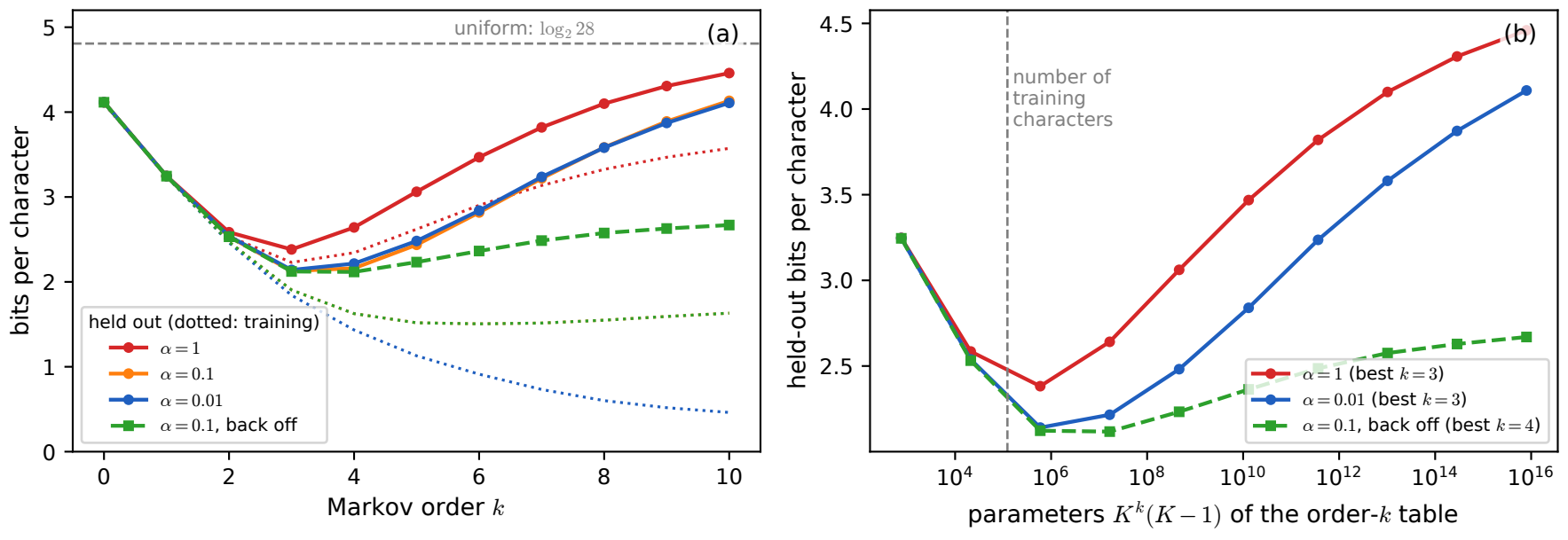

uniform: $\log_2 28$
(a)
bits per character
held out (dotted: training)
$\alpha = 1$
$\alpha = 0.1$
$\alpha = 0.01$
$\alpha = 0.1$, back off
Markov order k
number of training characters
(b)
held-out bits per character
$\alpha = 1$ (best $k = 3$)
$\alpha = 0.01$ (best $k = 3$)
$\alpha = 0.1$, back off (best $k = 4$)
parameters $K^k(K-1)$ of the order-k table