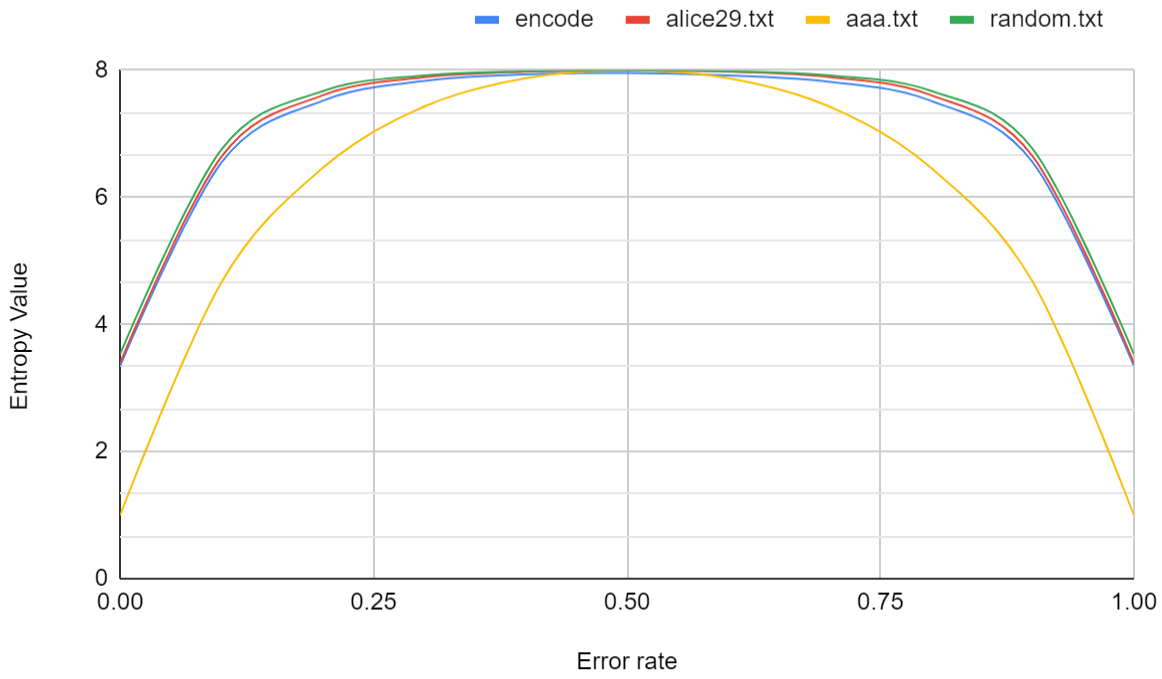

What causes entropy high or low?
More random/uncertain events inside the text, then more entropy.

Analysis based on the graph result:
Based on the above graph results we can clearly see that the graph is a curve shape. Meaning that entropy goes up as error rate increases until the estimated 50 percent error rate, then entropy will go down after 50 percent error rate and eventually has the same entropy value(0 percent error rate entropy = 100 percent error rate entropy).

Why the result is curve shape:
For example:

Error rate	0 percent	25 percent	50 percent	75 percent	100 percent
Text	aabc	azbc	azxc	azxm	zzxm
differences	0 character	1 character	2 characters	3 characters	4 characters

Entropy is the same during 0 error rate and 100 error rate because 100 error rate will change every character in the text, in the example above aabc and zzxm have the same entropy, because they both have the same randomness pattern. Obviously, entropy will go up as error rates go up because of the increase of random events. But when error rate passes 50 percent, the text will start having the same randomness pattern as previous, so entropy will go down after 50 percent error rate.

Why do some texts have higher/lower entropy overall?
Because uncertain events, randomness and lengths are different. More uncertain events = more entropy. For example, aaa.txt only has a in the text, so uncertain event is low, because low uncertain event. In contrast, random.txt has the highest entropy because it's completely random.

***Why is the entropy of a text before encode always higher than after encode?**
It’s not showing in the graph, but all of the texts have the same situation: the text before the encode has higher entropy than the same text after encoding. This is because encodes take a character(8 bits), separate to 4 bits and turn each 4 bits to 8 bits after multiplying to matrix G(turn it to hamming code). So the text after encode has longer (2x) length than the text before encode and hamming codes should have less uncertain bits and less randomness than character. So entropy of encoded text should have less entropy.