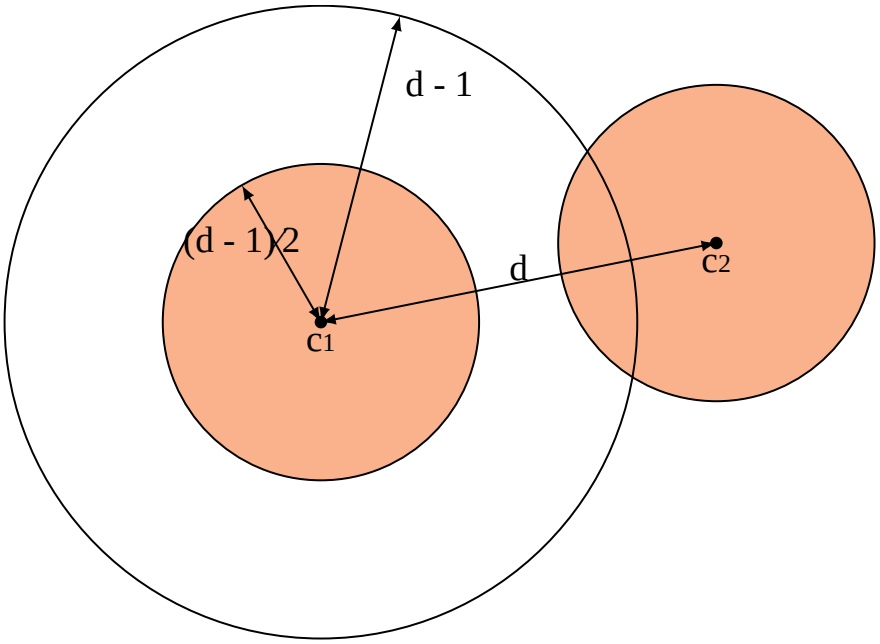
Error Detection/Correction and the Hamming Bound

In general, a code with minimal distance d can ***detect*** up to $d-1$ bit flip errors: in $d-1$ bit flip 'steps', those bit flips can never result in another valid codeword. The code can accurately ***correct*** up to $\frac{d-1}{2}$ bit flip errors. Look at the diagram below to understand why: if two codewords c_1 and c_2 are guaranteed to be at least d bit flips apart, then the set of strings that result from $\frac{d-1}{2}$ bit flips on c_1 (the orange circle on the left) never overlaps with the set of strings that result from $\frac{d-1}{2}$ bit flips on c_2 (the orange circle on the right).



Because every codeword is guaranteed to have this neighborhood around it that it does not share with any other codeword, we can upper bound the total number of codewords in terms of the distance.

Proposition: Hamming bound

If C is a *binary* code of block length n and minimum distance 3, then $|C| \leq \frac{2^n}{n+1}$.

Proof

For each $c \in C$, define the neighborhood of c to be $N(c) := \{y \in \{0,1\}^n \mid d(x,y) \leq 1\}$. Every such neighborhood contains

exactly $n+1$ elements. Since $d=3$, $N(c) \cap N(c') = \emptyset$ whenever $c \neq c'$. Hence,

$$2^n \geq |\bigcup_{c \in C} N(c)| = \sum_{c \in C} |N(c)| = |C| \cdot (n+1),$$

and the result follows.

The $[7,4]$ Hamming code is optimal in the sense that it achieves this Hamming bound: it is a code with block length 7 and minimal distance 3, so an upper bound to the codebook size is $\frac{2^7}{7+1} = 2^4$. The Hamming code achieves exactly this codebook size.