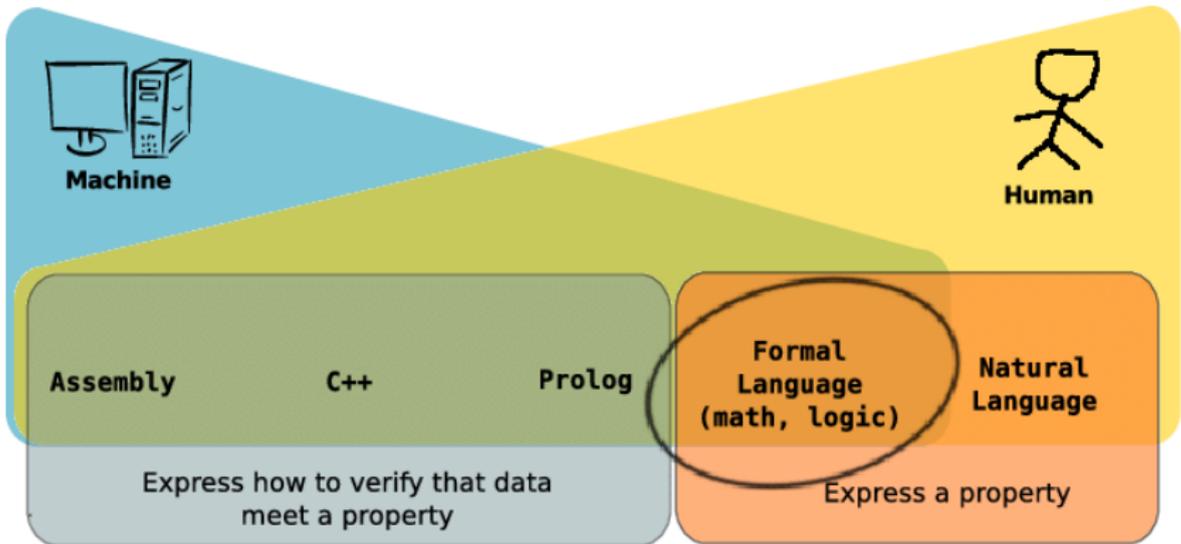**What is...Hamming distance?**

Or: Cubes!

Formal languages

- ▶ Formal language = interpolation between what a machine likes and what a human likes
- ▶ For this video we take an alphabet A strings in A are called words
- ▶ Hamming space H = words of a fixed length

Distance on formal languages

- "**karolin**" and "**kathrin**" is 3.

- "**karolin**" and "**kerstin**" is 3.

- "**kathrin**" and "**kerstin**" is 4.

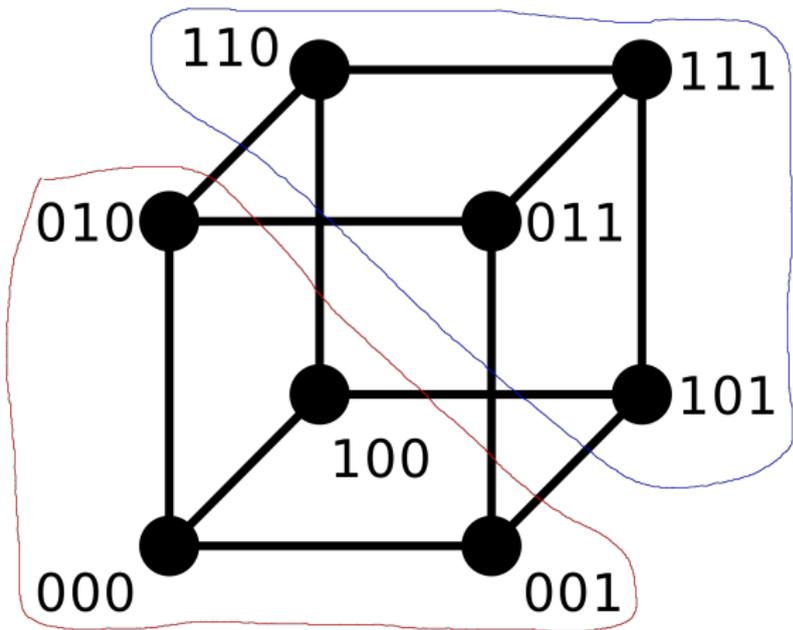- **0000** and **1111** is 4.

- **2173896** and **2233796** is 3.

▶ Hamming distance on $H =$ number of different symbols

▶ This is used as an estimate of error

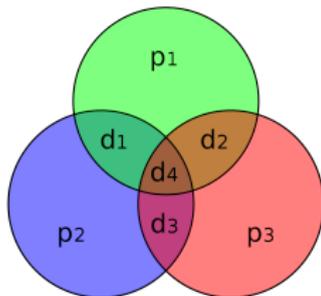▶ Plays a huge role in error detection and error correction

The binary example

- ▶ Codewords = subset of H (H is a $\mathbb{Z}/2\mathbb{Z}$ vector space)

- ▶ Linear code = linear subspace of codewords

- ▶ $0 \mapsto 000$, $1 \mapsto 111$ and transmitting 000 and 111 has a better correction rate than transmitting 0 and 1
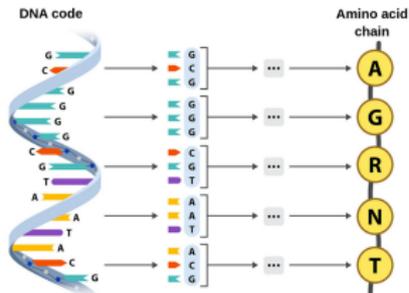
Hamming distance is a metric on H

▶ Hamming distance is used in error detection via Hamming code (more next)

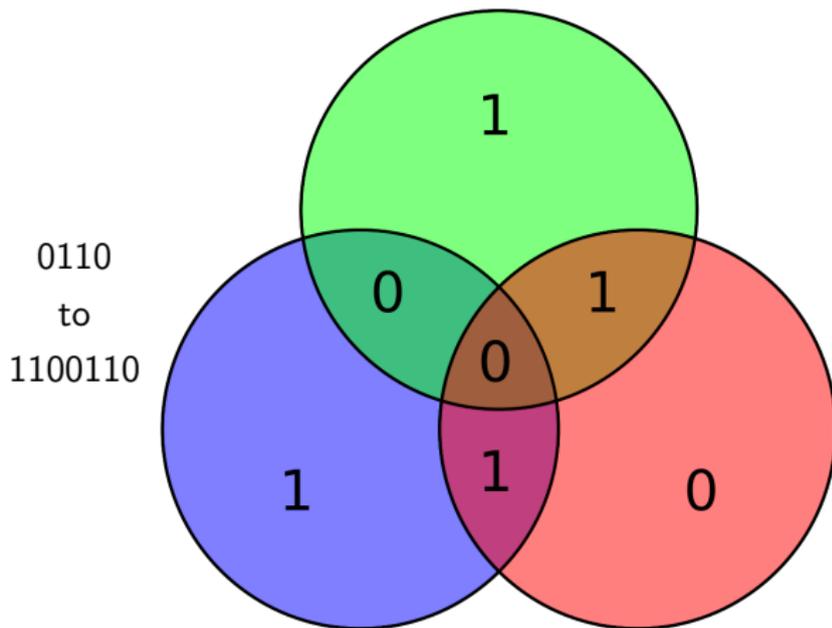▶ Hamming distance is also used as a measure of genetic distance

0110
to
1100110

- Hamming (7,4) = make 4 bits into 7 bits

- Up to 1 swapped digit can be detected in this way

Thank you for your attention!

I hope that was of some help.