

Codierung (und Kompression)

LaufLängenCodierung (Run-Length-Encoding, RLE)

mehrere gleiche Zeichen/Symbole/Zahlen hintereinander

→ Übertragung des Zeichens einmal + dessen Anzahl

z.B.

AAAAA → A5

- + einfach
- schnell
- verlustfrei

- vergleichsweise ineffizient
- nur einsetzbar, wenn gleiches Zeichen oft hintereinander vorkommt

Blockcode
→ Codewörter haben gleiche Länge

Huffman-Codierung

- häufiger vorkommende Buchstaben/Zeichen erhalten einen kürzeren Code → auf einen bestimmten Fall optimiert

→ Vorkommenswahrscheinlichkeiten müssen bekannt sein

Vorgehen:

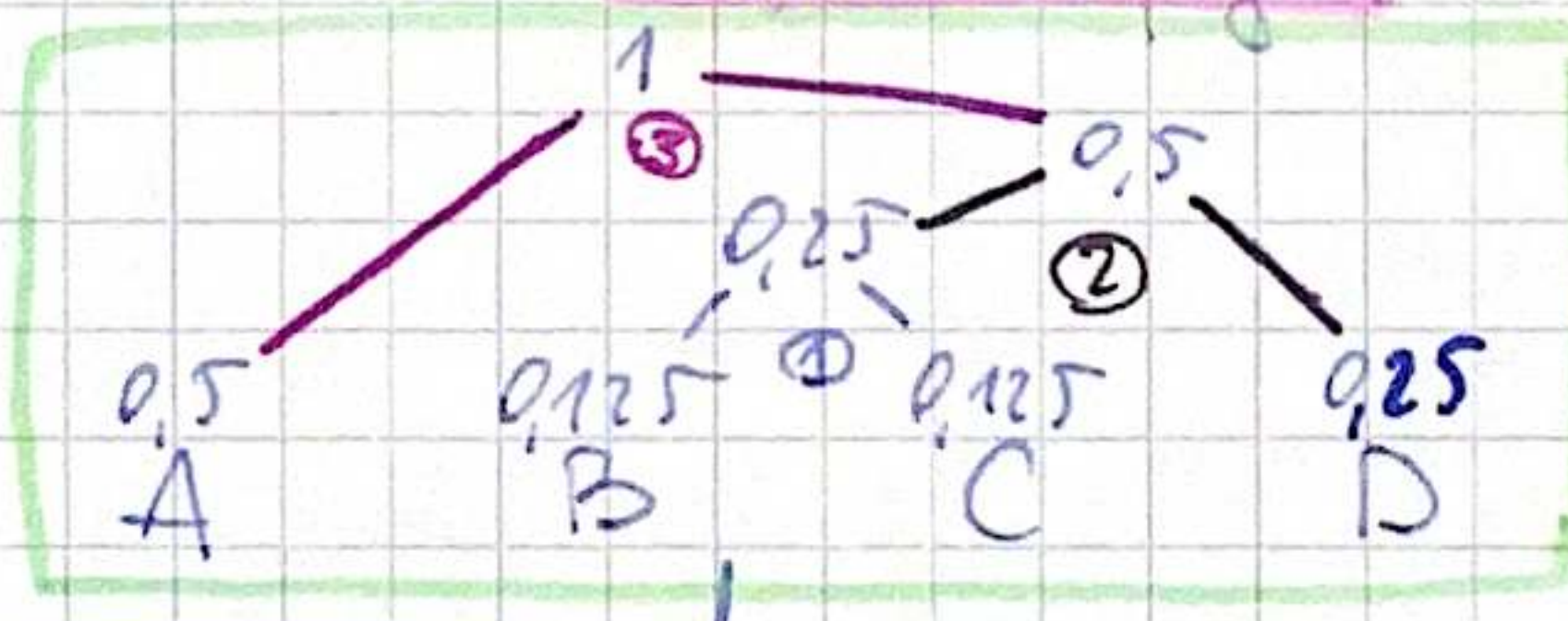
1. "AAABBCADD"

1. Wahrscheinlichkeiten ermitteln

A → 0,5
B → 0,125
C → 0,125
D → 0,25

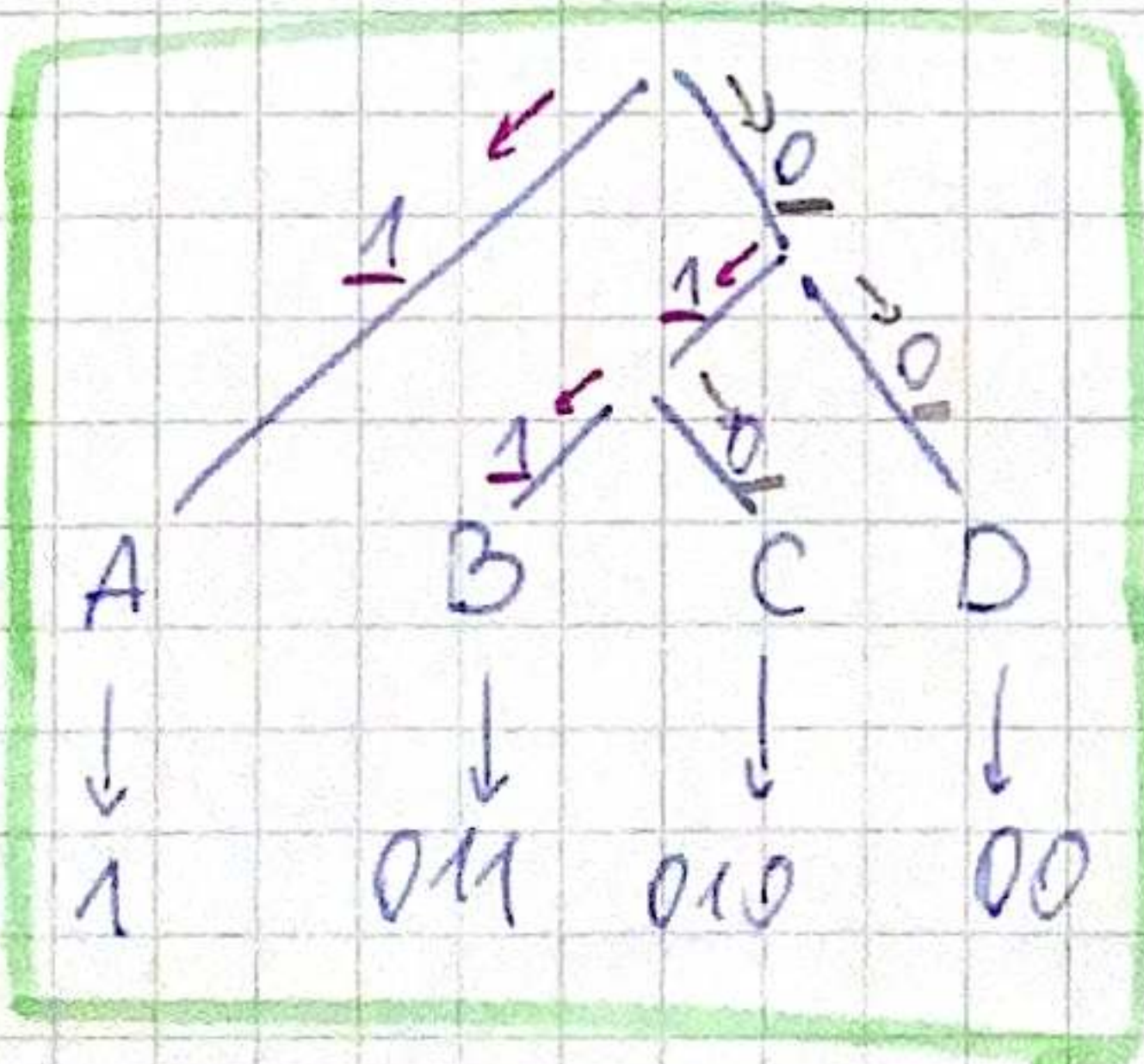
2. Huffman-Baum erstellen

→ zwei kleinste Wahrscheinlichkeiten immer zusammenfügen



3. Zweige beschriften

→ rechts 0, links 1 {oder umgekehrt, aber konsistent!}



4. Codes ablesen

A → 1, B → 011, C → 010, D → 00

Kompressionsrate:

$\frac{\text{unkomprimiert}}{\text{komprimiert}} \approx$

Kompressionsfaktor

$\frac{\text{komprimiert}}{\text{unkomprimiert}}$

- Codewort: Block aus verfügbaren Zeichen (Zeichenverat)
- hat eine Wortlänge
- Codebaum { alle Codewörter (= Code)

Präfixcode:

kein Codewort ist Anfang eines anderen Codeworts
→ immer eindeutig decodierbar

- + verlustfrei
- kürzester präfixfreier Code