Universität Bielefeld
Technische Fakultät

R|V|S
Rechnernetze und
Verteilte Systeme

Technische Informatik I

Vorlesung 12: Kodierungstheorie II

Joachim Schmidt
jschmidt@techfak.uni-bielefeld.de

Rückblick

- Übertragung durch gestörte Kanäle
- Redundanz
- Entropie
- Hammingabstand
- Prüfziffern
- Beispiele
  - Raumfahrt
  - Audio-CD

Vorgehen

- Zur Datenübertragung:
  - Keine Fehlerkorrektur
  - Lediglich Fehlererkennung
  - Im Fehlerfall nochmalige Übertragung der Daten (falls dies möglich ist, sonst s.u.)
- Zur Datenspeicherung:
  - Fehlerkorrektur (Codierung) angepaßt zur Fehleranfälligkeit des Kanals, dazu Kenntnis der Statistischen Eigenschaften notwendig
  - Sicherung der Nachricht durch gezieltes Hinzufügen von Redundanz

CRC (Cyclic Redundancy Check)

- Aufteilung der Nachricht in
  - m Nachrichtenstellen (gegeben)
  - k Kontrollstellen (gesucht)
  - n = m + k Gesamtstellen (Code)
- Wahl eines Generatormusters / ~polynoms
  - $G(u) = g_k u^{k-1} + ... + g_1 u^1 + g_0 u^0$
  - Polynom vom Grad k-1 über Feld **2** = < {1,0},+,x,0,1 >
  - Interpretation der Polynomkoeffizienten g als Zahlenfolge (Bitfolge)
  - Beispiel: $u^3 + u + 1$ → 1 0 1 1

CRC (Cyclic Redundancy Check)

- Interpretiere Code ebenso als Polynom
  - $X(u) = x_n u^{n-1} + ... + x_2 u^1 + x_1 u^0$
  - Vom Grad n-1
  - Beispiel: 1 0 0 1 0 0 0 → $u^6 + u^3$
  - Beachte: nur die „oberen“ m Nachrichtenstellen sind belegt (hier m=4)
- Somit rein formaler Übergang auf Polynome

CRC (Cyclic Redundancy Check)

- Berechnung des Codes X_{crc} durch Polynomdivision
  - X_{crc} = X(u) – X(u) mod G(u)
  - Beispiel: 1 0 0 1 0 0 0 / 1 0 1 1 gibt Rest 1 1 0
  - X_{crc} = 1 0 0 1 0 0 0 - 1 1 0 = 1 0 0 1 1 1 0

- Decodierung wieder durch Polynomdivision
  - Err = X(u) mod G(u)
  - Fehler, falls Err ≠ 0
  - Beispiel: 1 0 0 1 1 1 0 / 1 0 1 1 gibt Rest 0

CRC (Cyclic Redundancy Check)

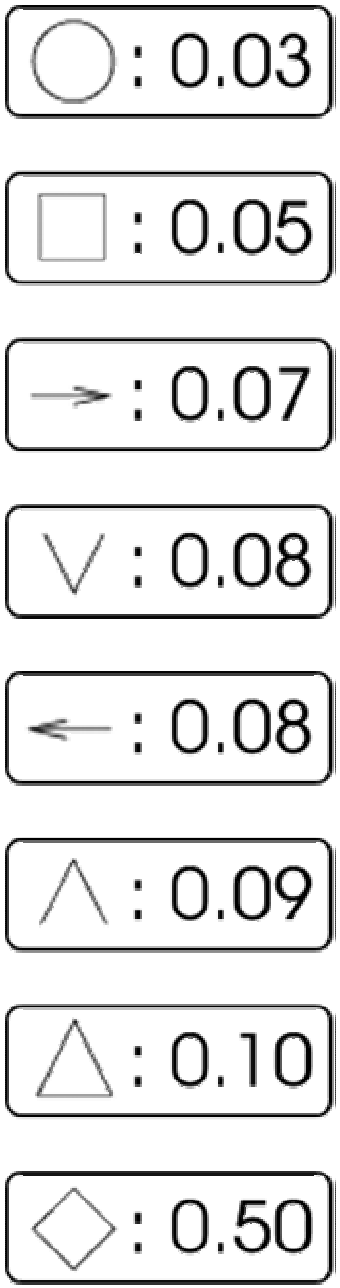
- CRC-Codes
  - Linearer Zyklischer Code
  - Sehr einfach zu realisierende Codierung
  - Erzeugung von mehreren Prüfbits aus vorgegebenen Nachrichtenbits durch das
  - Generatorpolynom
  - Dadurch sehr kompakt
  - Fehlererkennung und Korrektur möglich
  - Sehr verbreitet bei Datenträgern (Harddisk, Floppy) und in Netzwerken
  - Einfache Umsetzung in Hardware
  - Beispiel: CRC-12: $X^{12}+X^{11}+X^3+X^2+X+1$

Huffman-Kodierung (1952)

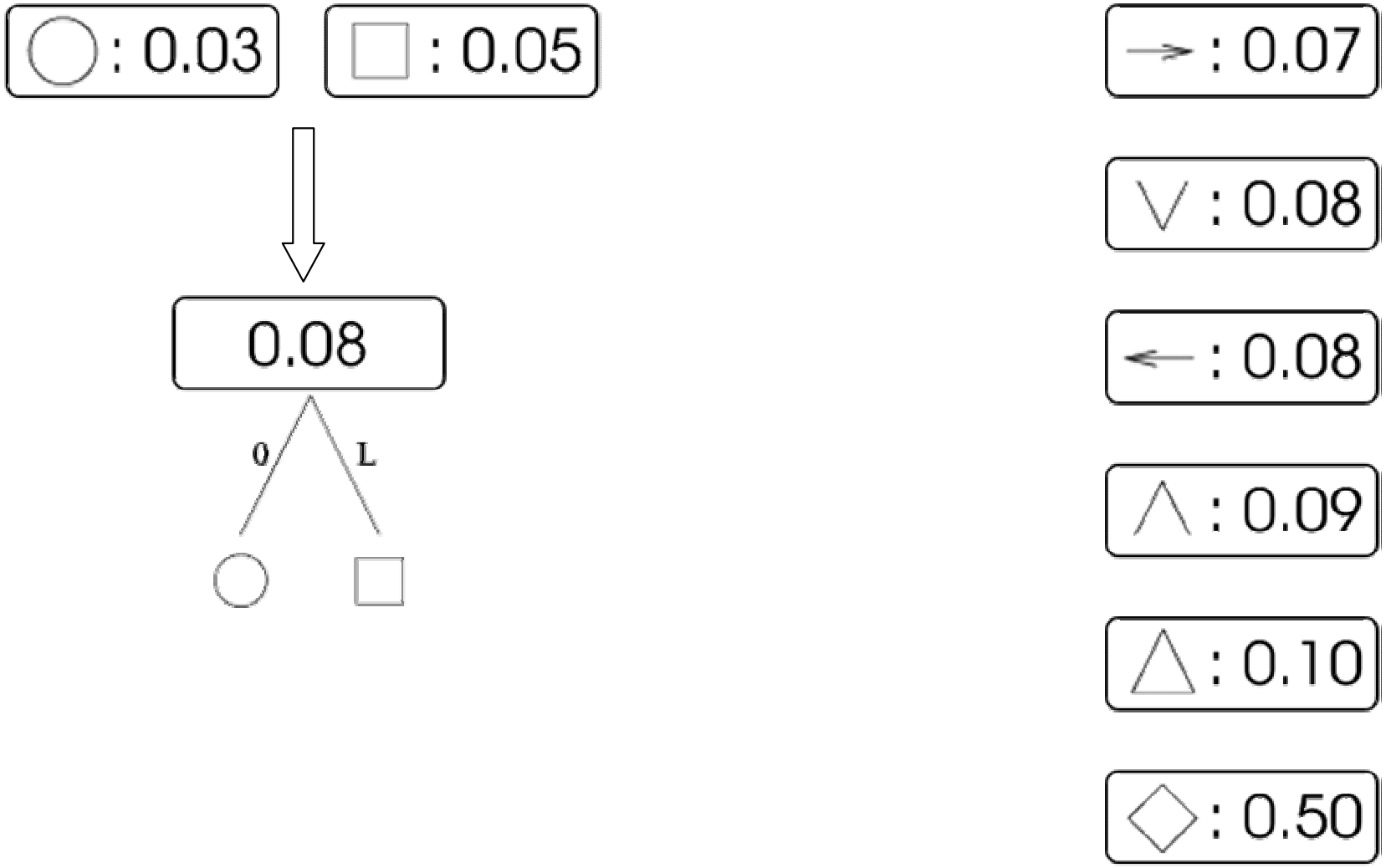
- Idee:
  - Häufig vorkommende Zeichen können durch ein kurzes Codewort (CW) dargestellt werden
  - Seltenere Zeichen bekommen ein längeres CW
  - Länge der codierten Nachricht ist geringer als uncodiert
  - Berücksichtigung der Auftrittswahrscheinlichkeit

Beispiel Huffman

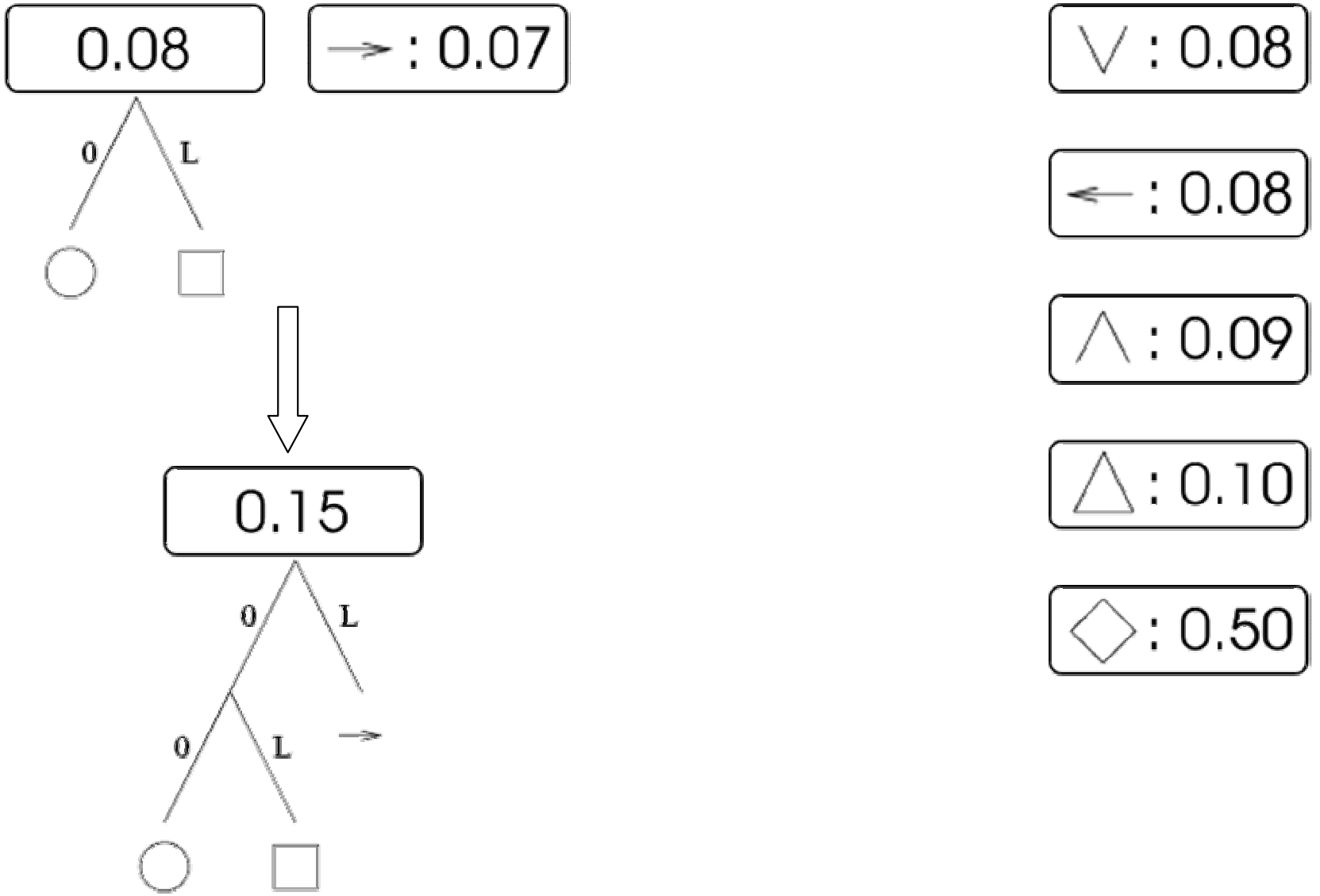
- Beispiel:
  - 8 zu codierende Zeichen mit
  - Unterschiedlichen Auftrittswahrscheinlichkeiten
- Vorgehen:
  - Erstelle eine Liste aller Zeichen
  - Sortiere alle Einträge
  - Fasse die zwei kleinsten Wahrscheinlichkeiten in einem B-Baum zusammen
  - Wiederhole die letzten 2 Schritte, bis alle Einträge zusammengefaßt sind



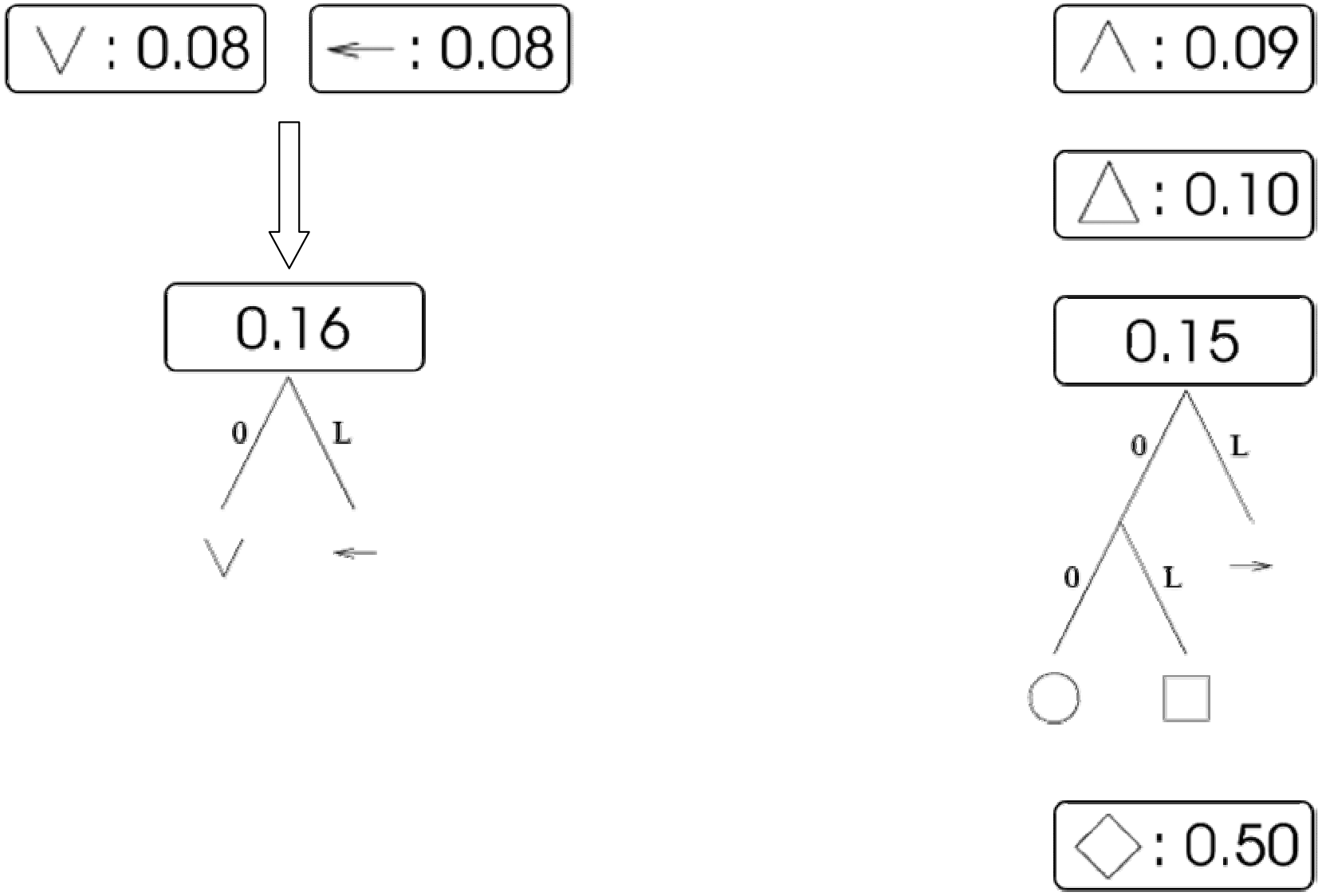
Beispiel Huffman



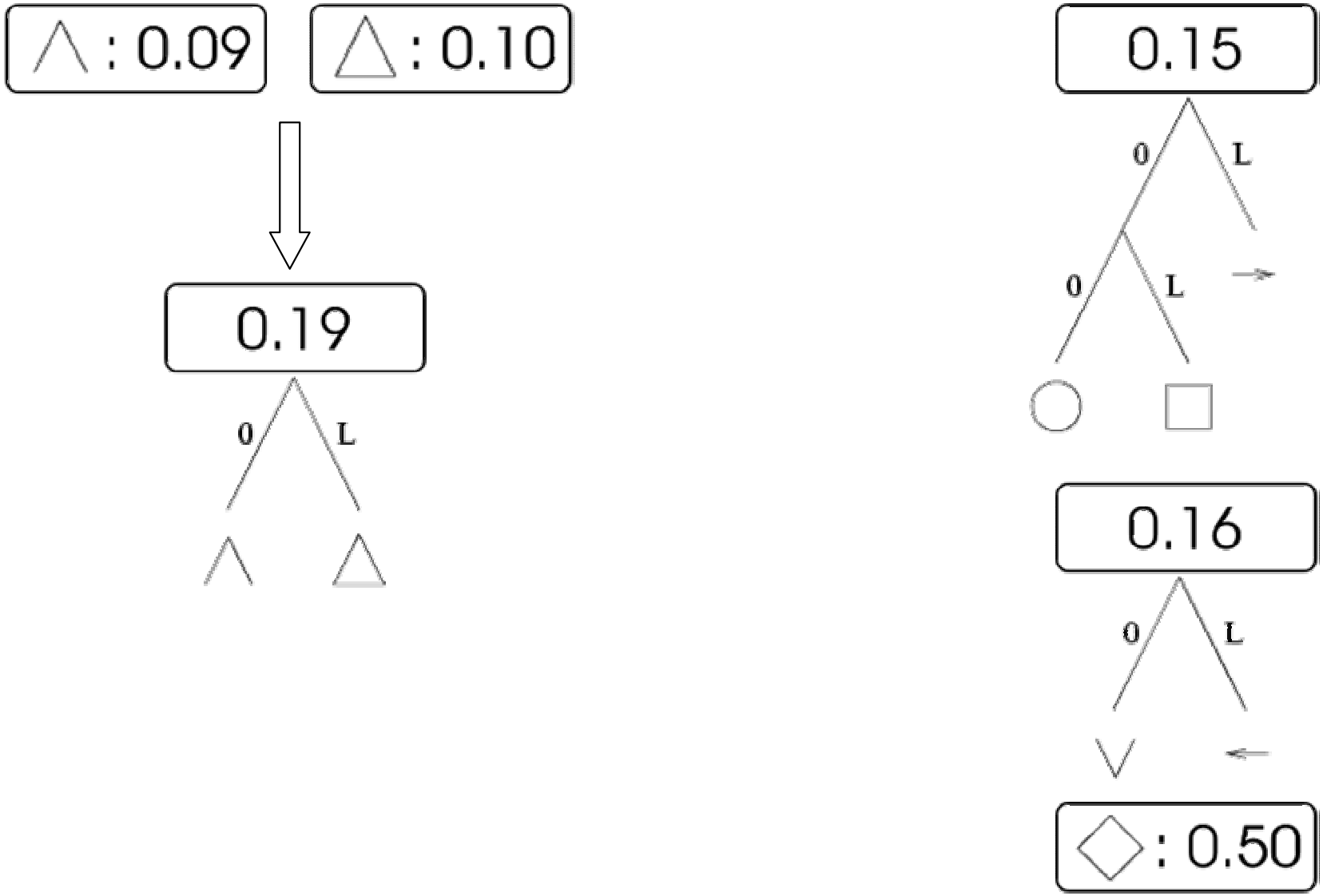
Beispiel Huffman



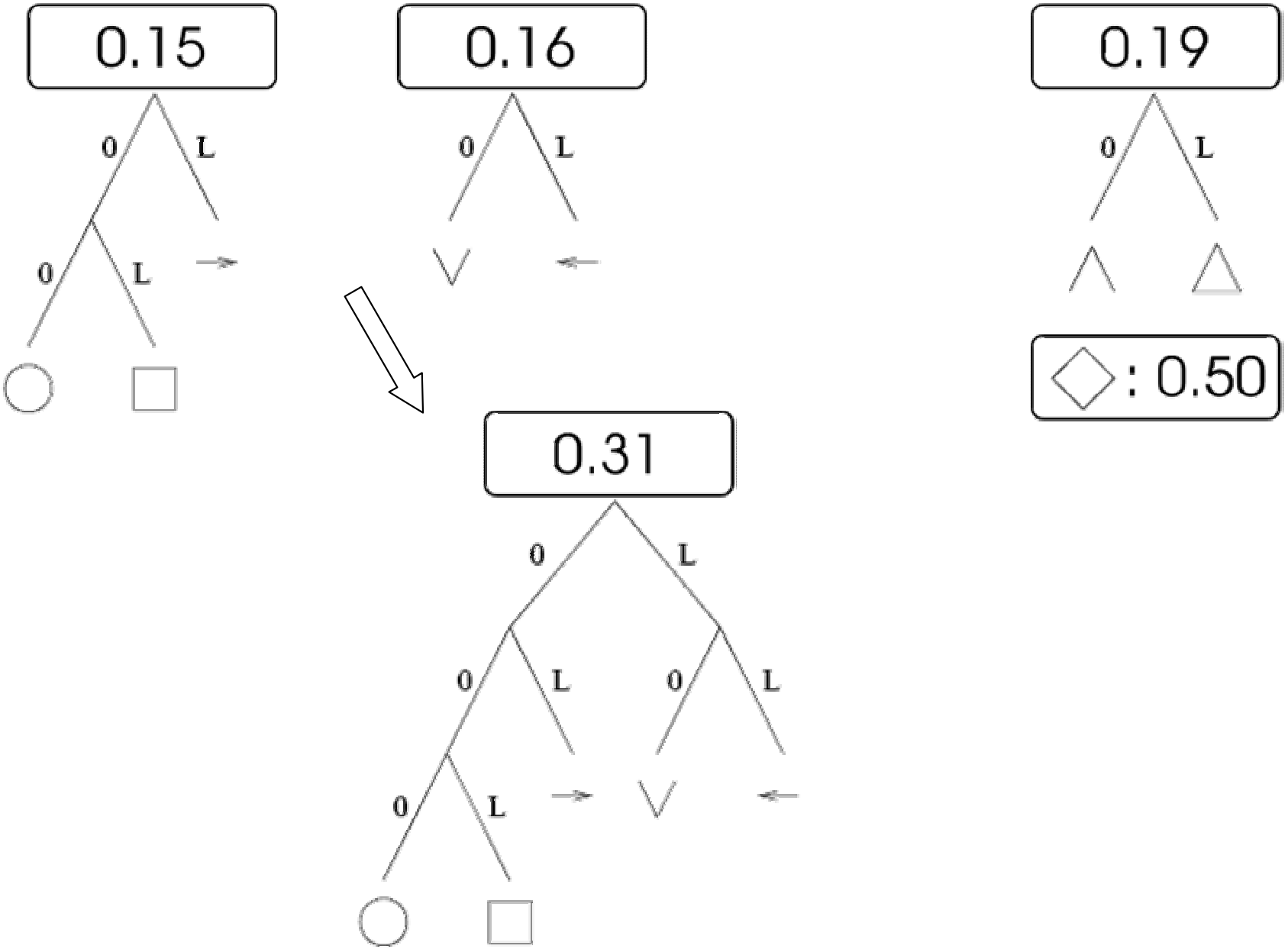
Beispiel Huffman



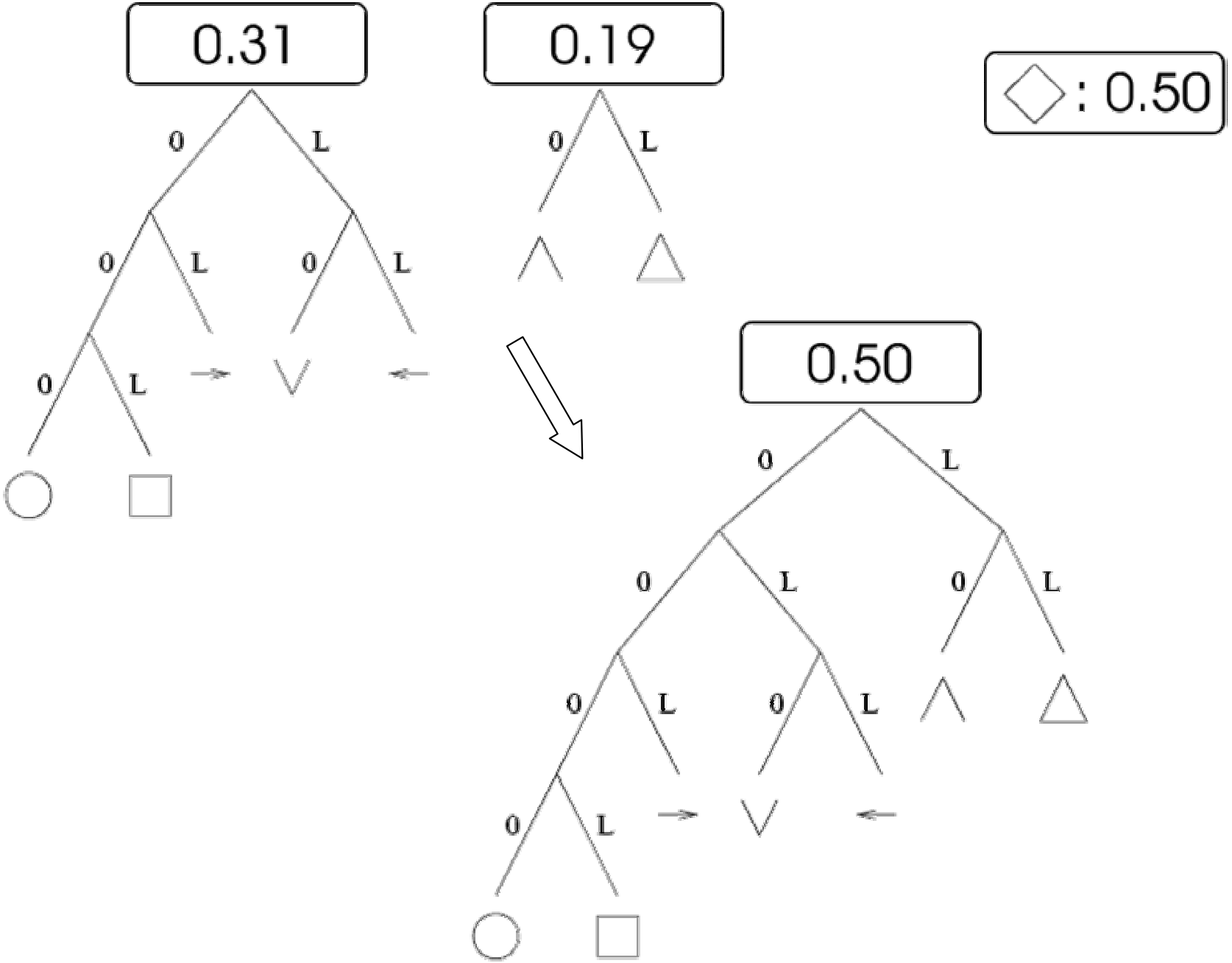
Beispiel Huffman



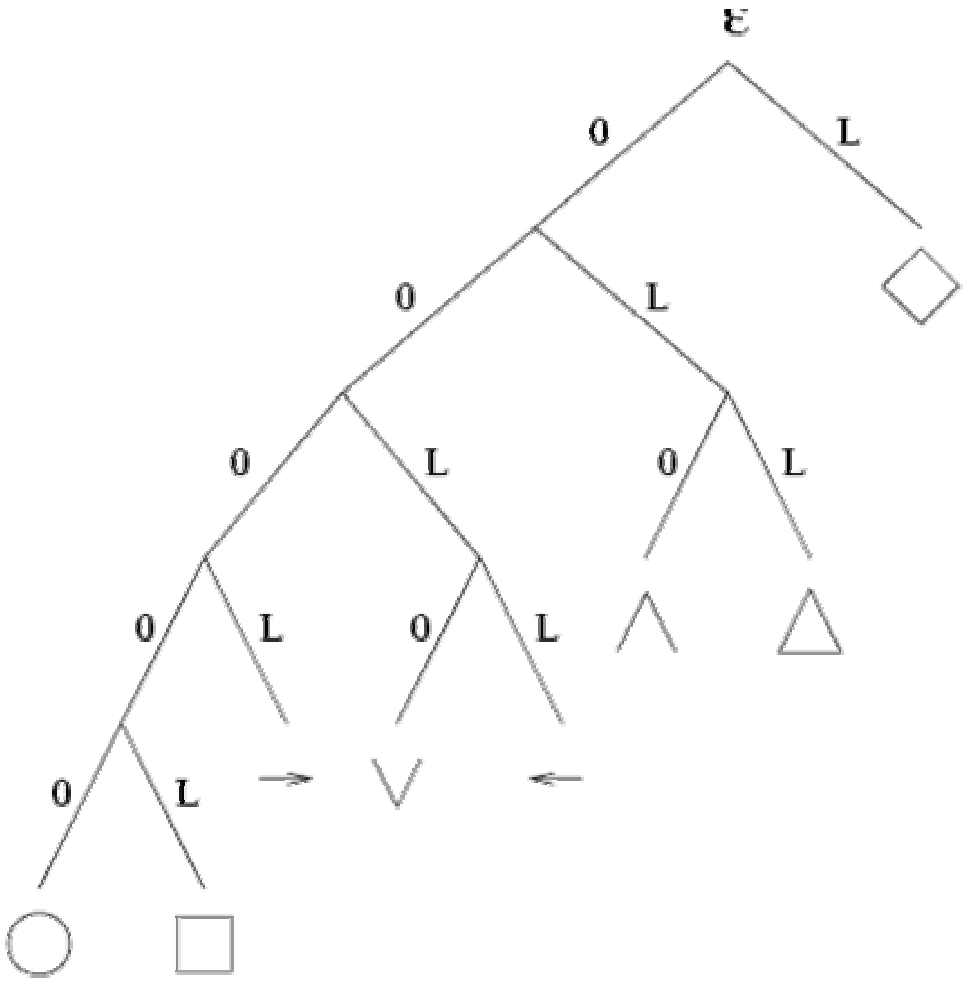
Beispiel Huffman



Beispiel Huffman



Beispiel Huffman



Beispiel Huffman

- Damit ergibt sich die Codetabelle:

Zeichen	p_i	CW	$p_i \cdot L_i$	$p_i \cdot \mathrm{ld}\ p_i$
◇	0,5	L	0,5	−0,5
△	0,1	$0LL$	0,3	−0,332
□	0,05	$0000L$	0,25	−0,216
○	0,03	00000	0,15	−0,152
∧	0,09	$0L0$	0,27	−0,313
∨	0,08	$00L0$	0,32	−0,292
→	0,07	$000L$	0,28	−0,269
←	0,08	$00LL$	0,32	−0,292
			$\sum = 2,39$	$\sum = 2,364$

Durchschnittliche Anzahl Bit/Zeichen

Entropie der Quelle

Huffman

- Verfahren erzeugt bei gegebenen Auftrittswahrscheinlichkeiten Optimum bezüglich der Länge der Codierung
- Problem: genaue Bestimmung der Auftrittswahrscheinlichkeiten
- Vielfache technische Anwendung

Lempel-Ziv-Kodierung (1977)

- Ersetzung von Zeichengruppen durch einen Zeiger auf einen Codebuch-Eintrag
- Welche Zeichengruppen kommen ins Lexikon?
- Wie wird die Nachricht in diese Zeichengruppen aufgeteilt?
  1. Codierer sucht längsten Codebucheintrag, der mit den nächsten Zeichen übereinstimmt
  2. Codierer sucht den längsten Substring, der mit einem Eintrag im Codebuch übereinstimmt
     LFF (Longest Fragment First)

Lempel-Ziv-Kodierung

- Beispiel:
  - Text ‚aaabbbabaabaaaabab‘
  - Lineare Parsierung von links

	$\underbrace{a}$	$\underbrace{aa}$	$\underbrace{b}$	$\underbrace{ba}$	$\underbrace{baa}$	$\underbrace{baaa}$	$\underbrace{bab}$
Gruppen-Nr.:	1	2	3	4	5	6	7
Ergebnis:	$(\perp, a)$	$(1, a)$	$(\perp, b)$	$(3, a)$	$(4, a)$	$(5. a)$	$(4, b)$

  - Original: 8 Bit pro Zeichen * 16 Zeichen = 128 Bit
  - LZ-Code: (3 Bit für Zeiger + 8 Bit für Zeichen) * 7 Gruppen = 77 Bit

Lempel-Ziv-Kodierung

- Hohe Kompression bei großer Gruppenlänge
- Problem: Lexikon benötigt immer mehr Platz

- Umsetzungen:
  - Abraham Lempel und Jacob Ziv (1977)
  - Die bekannteste ist sicherlich die LZW-Kodierung von Terry A. Welch (1984)

Shannonsches Theorem

- Sei:
  - $H = \sum_a p_a \, ld(1/p_a)$ mit $a \in A$: Zeichen aus A

    die Entropie der Nachrichtenquelle
  - $L = \sum_a p_a \, c(a)$ mit c(a) = Wortlänge

    die mittlere Wortlänge
- Dann gilt:
  1. $H \leq L$
  2. L – H kann durch geeignete Codierung beliebig klein gemacht werden

Shannonsches Theorem

- Die Entropie ist die untere Grenze für die mittlere Wortlänge
- Die Differenz L – H heißt Coderedundanz
- 1 – H/L heißt relative Coderedundanz

Abhängigkeit von der Datenquelle

- Bei Kenntnis der Datenquelle differenzierte Algorithmen, z.B:
- Klartext sehr gut komprimierbar:
  - eingeschränkter Zeichenvorrat
  - Linguistisch-Statistische Analyse liefert durchschnittliche Auftrittswahrscheinlichkeiten für Buchstaben/Wörtern in Texten
- Audioinformationen mäßig komprimierbar:
  - Bei Stereo: Korrelation der Kanäle ausnutzen
  - Lineare Vorhersage des nächsten Samples
  - Nur Abspeicherung des Vorhersagefehlers

Literatur

- Codierungstheorie

  Dieter Jungnickel, Spektrum, 1995
- Informatik, Band 1
  - Manfred Broy, Springer, 1998
- Structured Computer Organization

  Andrew S. Tanenbaum, Prentice Hall, 1999
- elearn.rvs.uni-bielefeld.de
  - Diese Folien
  - Weblinks mit Beispielen