Prinzipien zur Erzeugung von Zufallszahlen in der Informatik

Vortragender: Dipl.-Inform. Thomas Hinze

Motivation

Zufallszahlen

- besitzen mannigfaltige Anwendungen in der Informatik
- ermöglichen den effizienten computergestützten Einsatz und
- beschleunigte Bearbeitung zahlreicher Aufgabenstellungen

wichtige Anwendungsgebiete und -beispiele in der Informatik

- **probabilistische Algorithmen**
  - z.B. Monte-Carlo- und Las-Vegas-Methoden
- **Simulation**
  - Generierung zufälliger Ereignisse entsprechend statistischer Vorgaben
- **Kryptographie / Kryptoanalyse**
  - z.B. Schlüssel- und Parametererzeugung, Verschlüsselungsverfahren, Verfahren zur digitalen Signatur, Dummy Traffic, Angriffe auf kryptographische Verfahren
- **Test**
  - Generierung von Testsätzen

Grundlegende Begriffe und Zusammenhänge

zufälliges Ereignis

- **Ereignis**, dessen Eintrittszeitpunkt und Wirkung **nicht** mit absoluter Sicherheit **vorhergesagt** werden kann
  - setzt eine unvollständige Beschreibung des Systems, in dem das zufällige Ereignis stattfinden kann, voraus
  - Zufall $\longrightarrow$ Ausdruck von **Unwissenheit**

Zufallsgröße

- meßbare / analysierbare Repräsentation der Gesamtheit aller zufälligen Ereignisse der zugrundeliegenden Systemkomponente
- **Wertebereich, Wertevorrat**

diskret $\longrightarrow$ **endlicher** oder **abzählbar unendlicher** Wertevorrat

Zufallszahl

- **Wert**, den eine Zufallsgröße bei ihrer Bestimmung annimmt

Zufallszahlenfolge

- Folge voneinander möglichst **unabhängiger** Zufallszahlen, die einer gegebenen **statistischen Verteilung** genügen

Vereinbarungen und Forderungen in der Informatik

$\rightarrow$ **Verarbeitung** der Zufallszahlen (ZZ) in jedem Fall durch Computer
$\rightarrow$ **Erzeugung** der ZZ entweder durch Computer oder extern

Grundanforderungen an Zufallszahlengeneratoren

- **Uniformität** – Gleichvert. der ZZ innerhalb jeder erzeugten Folge
- **Unabhängigkeit** – keine Autokorrelation zwischen den ZZ innerhalb jeder erzeugten Folge

zusätzliche Wünsche/Erfordernisse für Anwendungen der Informatik

- **diskrete** Zufallsgrößen
- anwendungsabhängig: **Reproduzierbarkeit/Irreproduzierbarkeit** der Zufallszahlenfolgen
- Gleichvert. der Zufallszahlen oft im **Intervall** $[0, 1)$, $\{0, 1\}$, $\{0, ..., 9\}$
- **Schnelligkeit, Effizienz, Unvorhersagbarkeit** des Generators

Bemerkungen

- Transformation Gleichverteilung $\rightarrow$ andere bekannte Verteilungen mathematisch herleitbar und berechenbar
- Intervall $[0, 1)$ gequantelt entspr. computerinterner Zahlendarst.

Definitionen aus der mathematischen Statistik

diskret gleichverteilte Zufallsgröße

Eine diskrete Zufallsgröße X heißt **gleichverteilt** auf den Ereignissen $a_1, \ldots, a_n$, wenn für $i = 1, \ldots, n$ gilt:

- Gleichwahrscheinlichkeit der Ereignisse: $P(X = a_i) = \frac{1}{n}$
- Erwartungswert: $EX = \frac{1}{n} \sum_{i=1}^{n} a_i$
- Streuung: $D^2X = \frac{1}{n} \sum_{i=1}^{n} (a_i - EX)^2$

Korrelation

- Grad des Zusammenhangs zwischen Zufallsgrößen

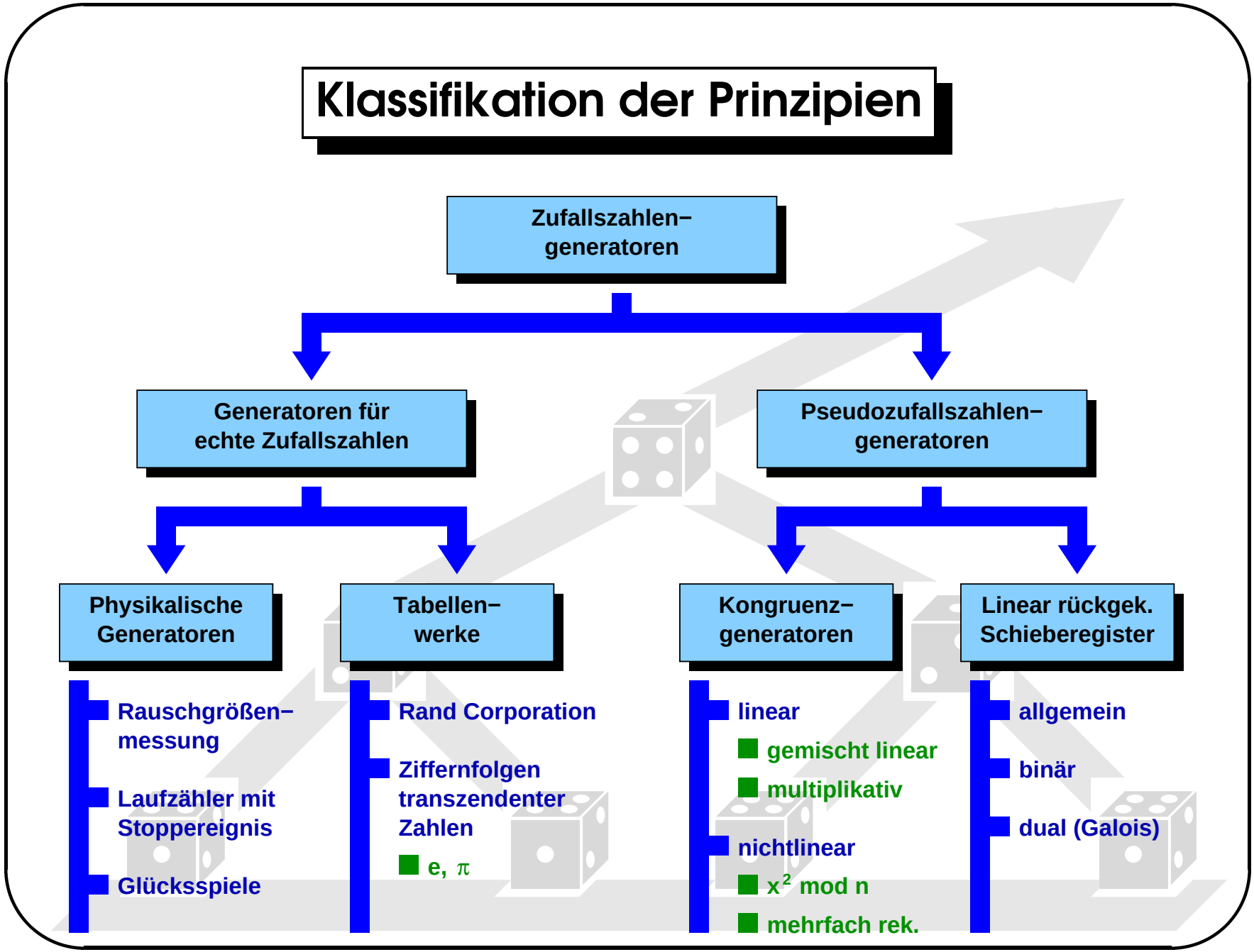
Korrelationskoeffizient

- zweier Zufallsgrößen X und Y: $\rho(X,Y) := \frac{E(X \cdot Y) - EX \cdot EY}{\sqrt{D^2X \cdot D^2Y}}$
- Es gilt: $\rho(X,Y) = \begin{cases} = 0 & \text{oder } \operatorname{cov}(X,Y) = 0 \longrightarrow X, Y \text{ unkorreliert} \\ \neq 0 & \longrightarrow X, Y \text{ voneinander abhängig} \end{cases}$

Autokorrelation

- Korrelation einer Zufallsgröße X mit sich selbst $(\rho(X,X))$

Klassifikation der Prinzipien



Rauschgrößenmessung

Prinzip

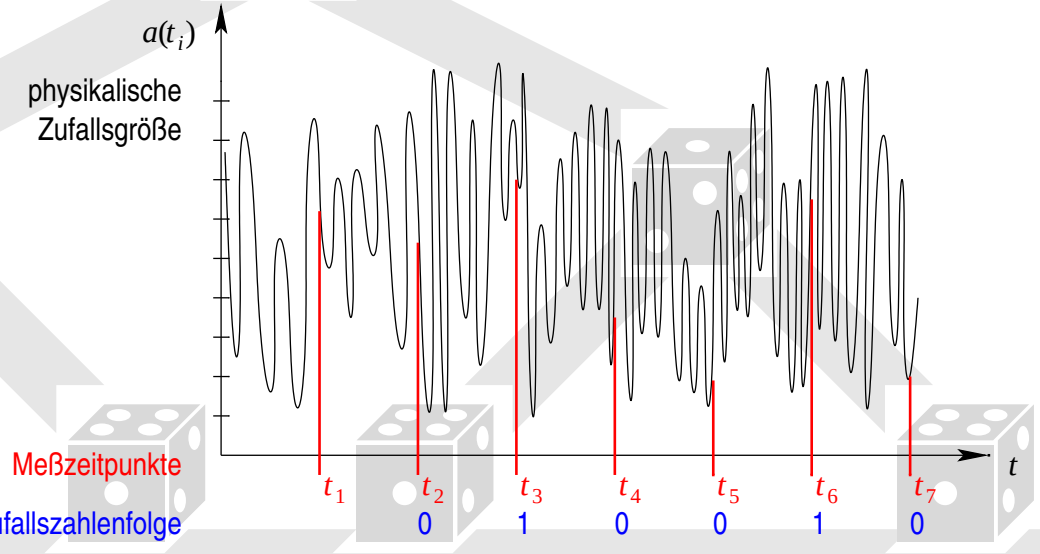
- zeitquantisierte Erfassung einer zugrundeliegenden physikalischen Zufallsgröße und Transformation in Zufallszahlenfolge
- Transformation z.B.: $a(t_i) \geq a(t_{i-1}) \longrightarrow$ Ausgabe $y(t_i) = 1$
  $a(t_i) < a(t_{i-1}) \longrightarrow$ Ausgabe $y(t_i) = 0$

Vorteile

- optimale statistische Eigenschaften
- ohne Speich. keine Reproduzierbarkeit

Nachteile

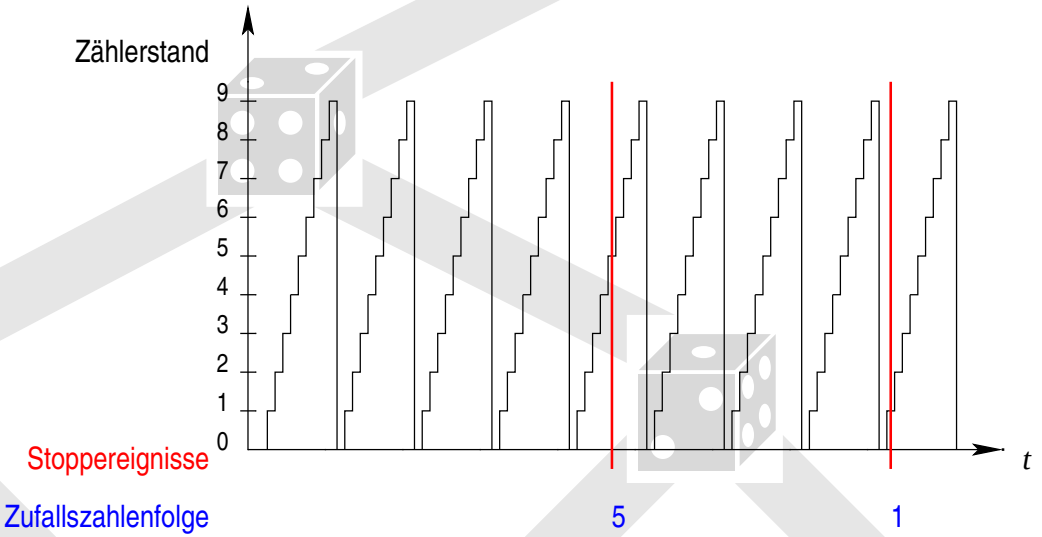
- Auswirkungen von Meßfehlern
- Verfügbarkeit und max. Abtastrate abhängig von phys. Zufallsgröße



Laufzähler mit Stoppereignis

Prinzip

- Zähler modulo n, der in schneller Folge fortlaufend von 0 bis $n-1$ zählt und jeweils bei Eintreten des zufälligen zählerunabhängigen Stoppereignisses sofort den aktuellen Zählerwert als Zufallszahl bereitstellt



Vorteile

- optimale statistische Eigenschaften
- ohne Speich. keine Reproduzierbarkeit

Nachteile

- Organisation der Stoppereignisse
- Stoppereignisse nicht zeitlich äquidistant
- mehrfaches vollständiges Durchzählen zwischen aufeinanderfolgenden Stoppereignissen sicherzustellen

Rand Corporation

Prinzip

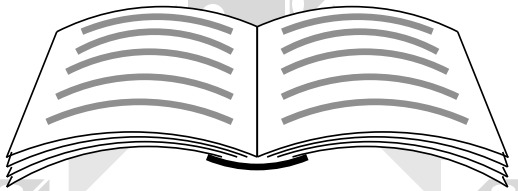
- Tabellenwerk (Buch) von Zufallszahlen, entstanden 1955
- enthält 1.000.000 Zufallszahlen
  - 400 Seiten zu 50 Zeilen mit je 50 Zufallszahlen (Dezimalziffern)
  - Buch zufällig aufschlagen, Folge wählen und
  - als genutzt kennzeichnen
- Zufallszahlen mittels physikalischem Generator erzeugt

Vorteile

- leichte Verfügbarkeit
- optimale statistische Eigenschaften
- Reproduzierbarkeit/Irreproduzierbarkeit
  leicht vom Anwender steuerbar

Nachteile

- endlicher Vorrat an Zufallszahlen
- Übertragung Buch $\longrightarrow$ Computer
- Bekanntheit des Buches



Ziffernfolgen transzendenter Zahlen

Definition transzendente Zahlen

Zahlen, die nicht algebraisch sind, d.h. die nicht als Nullstellen beliebiger Polynome $a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$ mit $a_i \in \mathbb{Q}$ und $i \in \{0, \ldots, n\}$, $n \in \mathbb{N} \setminus \{0\}$ vorkommen können. (z.B. e, π)

Prinzip

- Berechnung und Tabellierung der gewünschten transzendenten Zahl auf hinreichend viele zuverlässige Nachkommastellen (z.B. mittels Spigotalgorithmen)
- zufällige Auswahl einer Ziffernfolge, Nutzung als Zufallszahlenfolge und Kennzeichnung als bereits genutzt

3, 141 592 653
2, 718 281 828 ...
589 793 238 ...

Vorteile

- leichte rechnergestützte Realisierbarkeit
- leichte Vorabgen. für spätere Nutzung

Nachteile

- keine gesicherten statistischen Eigenschaften, nur Annahmen
- hoher Bekanntheitsgrad der Ziffernfolgen

Pseudozufallszahlengeneratoren

- Gewinnung der Zahlenfolgen durch **deterministische Berechnung**
- Zufall $\longrightarrow$ Pseudozufall: Wegfall der generellen Unvorhersagbarkeit

gemeinsames Prinzip

- ausgehend von einem zufällig gewählten **Startwert** und **geeigneten Parameterbelegungen** erfolgt Bestimmung der Pseudozufallszahlenfolge durch spezifische **rekursive Berechnungsvorschrift**

Forderung für perfekten Pseudozufallszahlengenerator

- Es soll **keinen effizienten** (polynomiellen) **Algorithmus** geben, der eine **Pseudozufallszahlenfolge** ohne Kenntnis der Berechnungsvorschrift, des Startwertes und der Parameterbelegungen signifikant von einer **echten Zufallszahlenfolge unterscheiden** kann.

gemeinsame Eigenschaft

- **Periodizität:** zyklische Wiederholung der Pseudozufallszahlenfolge (PZZ-Folge) nach endlicher Länge
- **Ziel:** Maximierung der Periodenlänge

Multiplikativer Kongruenzgenerator

Prinzip

Parameter: $a, m \in \mathbb{N}$ mit $a \geq 2, a < m, m > 1$
Startwert: $z_0 \in \{1, \ldots, m-1\}$
Rekursion: $z_i = (a \cdot z_{i-1}) \mod m$
PZZ-Folge: $r_i = z_i / m$ Es gilt: $r_i \in [0, 1)$

maximale Periodenlänge

- m unter folgenden Bedingungen (Kobayashi):
  - $m = p^l$ oder $m = 2p^l$ mit p ungerade Primzahl, $l \in \mathbb{N} \setminus \{0\}$
  - $a^{m-1} \equiv 1 \mod m$
  - $\mathrm{ggT}(z_0, m) = 1$
- Minimierung der Autokorrelation: $a \approx \sqrt{m}$

Vorteile

- Schnelligkeit des Verfahrens und leichte Implementierbarkeit

Nachteile

- keine optimalen statistischen Eigenschaften, kleine Periodenlänge
- leichte Vorhersagbarkeit der PZZ-Folge (nicht perfekt)

Gemischt linearer Kongruenzgenerator

Prinzip

Parameter: $a, c, m \in \mathbb{N}$ mit $a < m, a \geq 2, c < m, c > 0, m > 1$

Startwert: $z_0 \in \{1, \ldots, m-1\}$

Rekursion: $z_i = (a \cdot z_{i-1} + c) \mod m$

PZZ-Folge: $r_i = z_i / m$ Es gilt: $r_i \in [0, 1)$

maximale Periodenlänge

- m unter folgenden Bedingungen (Lehmer):
  - $\mathrm{ggT}(c, m) = 1$
  - $a \equiv 1 \mod q$ für jeden Elementarteiler q von m
  - $a \equiv 1 \mod 4$ falls 4 Teiler von m ist
- Minimierung der Autokorr. (Fishman/Greenb.): $a \approx \sqrt{m - \frac{6c}{m}(1 - \frac{c}{m})}$

Vorteile

- Schnelligkeit des Verfahrens und leichte Implementierbarkeit

Nachteile

- keine optimalen statistischen Eigenschaften, kleine Periodenlänge
- leichte Vorhersagbarkeit der PZZ-Folge (nicht perfekt)

Mehrfach rekursiver Kongruenzgenerator

Prinzip

Parameter: $r, m, a_1, \ldots, a_r \in \mathbb{N}$ mit $r \geq 1$, $a_i \in \{0, \ldots, m-1\}$, $m > 1$

Startwerte: $z_1, \ldots, z_r \in \{0, \ldots, m-1\}$ mit $\exists j \in \{1, \ldots, r\} \,.\, z_j \neq 0$

Rekursion: $$z_i = \left(\sum_{k=1}^{r} a_k \cdot z_{i-k} \right) \mod m$$

PZZ-Folge: $r_i = z_i / m$ Es gilt: $r_i \in [0, 1)$

- Spezialfall: Fibonacci-Generator

maximale Periodenlänge

- $m^r - 1$ Belegungsvorschrift der Parameter für max. Periodenlänge ändert sich mit jedem Vorgabewert für r, keine geschl. Darst.

Vorteile

- größere maximale Periodenlänge, schwierigere Vorhersagbarkeit
- Schnelligkeit des Verfahrens und leichte Implementierbarkeit

Nachteile

- keine optimalen statistischen Eigenschaften, nicht perfekt
- hoher Aufwand zum Finden geeigneter Parameterbelegungen

$x^2 \bmod n$ Generator

- Verfahren nach Blum, Shub

Prinzip

Parameter: $s, p, q \in \mathbb{N}$ mit p, q prim, $p \approx q$, $p, q \equiv 3 \mod 4$,
$0 < s < pq$, $\mathrm{ggT}(s, pq) = 1$

Startwert: $z_0 = s^2 \mod (pq)$

Rekursion: $z_i = z_{i-1}^2 \mod (pq)$

PZZ-Folge: $r_i = z_i \mod 2$ Es gilt: $r_i \in \{0, 1\}$

maximale Periodenlänge

- pq

Vorteile

- perfekter Pseudozufallszahlengenerator
- leichte Implementierbarkeit
- leichte Wahl geeigneter Parameterbelegungen

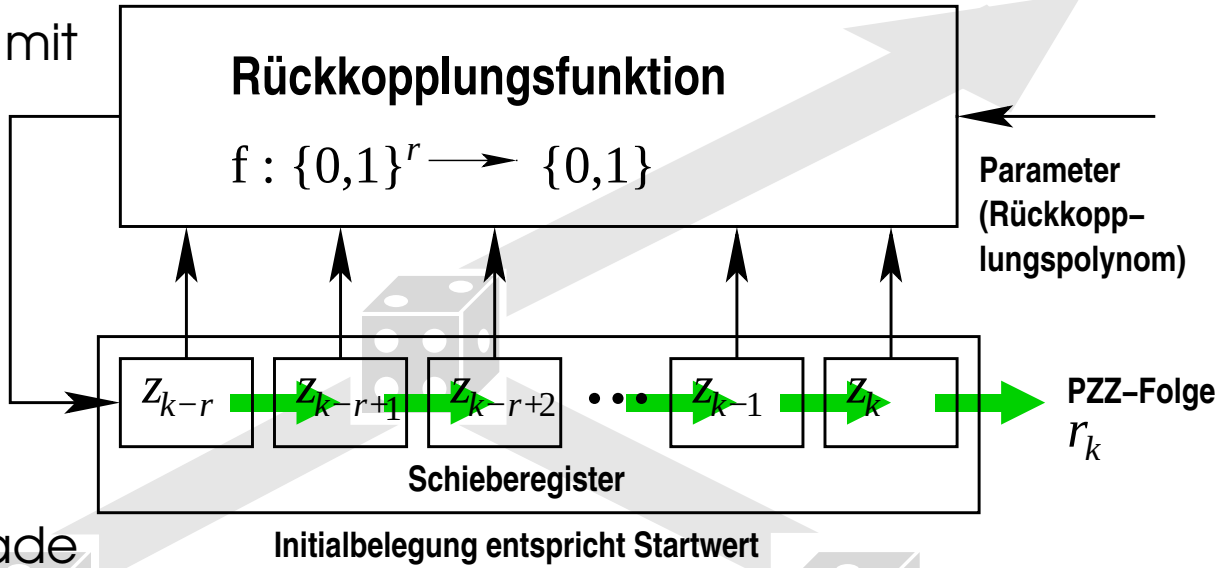
Nachteile

- nur ein Bit pro Rekursionsschritt $\longrightarrow$ langsam
- kleine Periodenlänge

Linear rückgekoppeltes Schieberegister

Prinzip

- Schieberegister mit Bitfolge $\neq 0^r$ initialisiert
- jede Zelle speichert ein Bit für einen Takt (D1)
- Bits schieben sich taktweise durch die Kaskade



- Bit der letzten Zelle als Pseudozufallsbit taktweise ausgegeben
- Bit der ersten Zelle mittels Rückkopplungsfunktion berechnet
- Rückkopplungsfunktion und ihre Parameter bestimmen wesentlich die Qualität der statistischen Eigenschaften der PZZ-Folge
- gleichwertige Notation auch als Rekursionsschema möglich

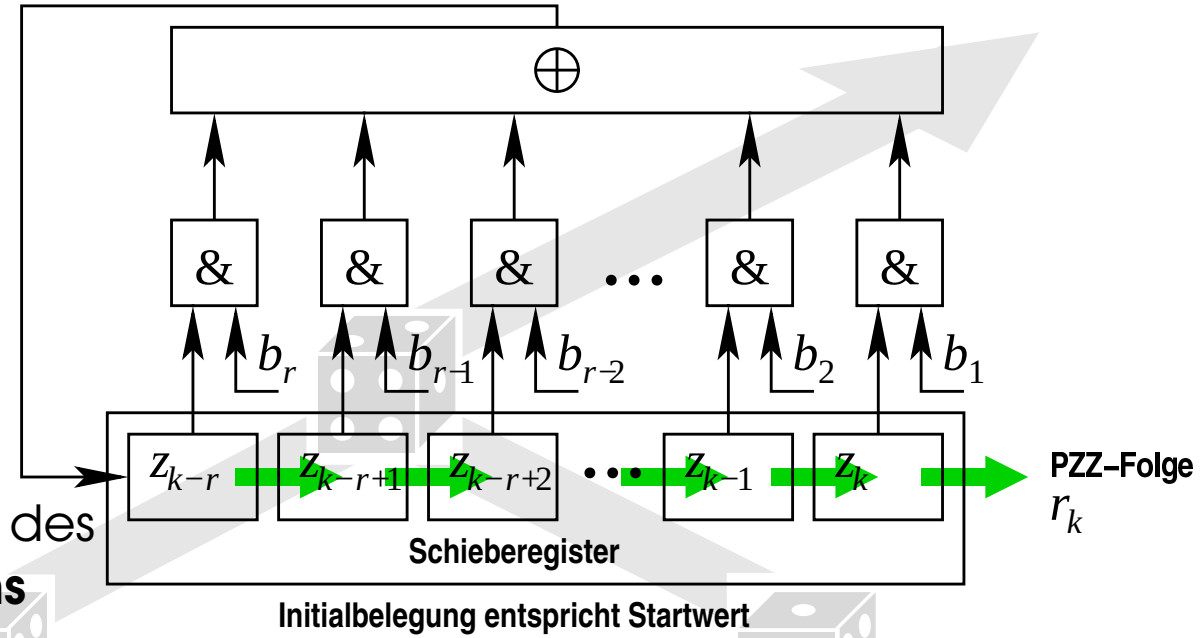
maximale Periodenlänge

- $2^r - 1$ Bit, unabhängig von der gewählten zulässigen Initialisierung

Binäres linear rückgekoppeltes Schieberegister

Prinzip

- Rückkopplungsfunktion ist mit den Parametern $b_i \in \{0,1\}$ mit $i \in \{1,\ldots,r\}$ behaftet
- Parameter b_i bilden den Koeffizienten des **charakt. Polynoms**
  $\mathrm{p} : \{0,1\} \rightarrow \{0,1\}$
  $\mathrm{p}(x) = b_r x^{r-1} \oplus \ldots \oplus b_3 x^2 \oplus b_2 x \oplus b_1$
- zum Erreichen der maximalen Periodenlänge $2^r - 1$ muß charakteristisches Polynom **irreduzibel** sein



Vorteile

- leicht in Hardware implementierbar
- sehr schnell, einfache Parameterwahl

Nachteile

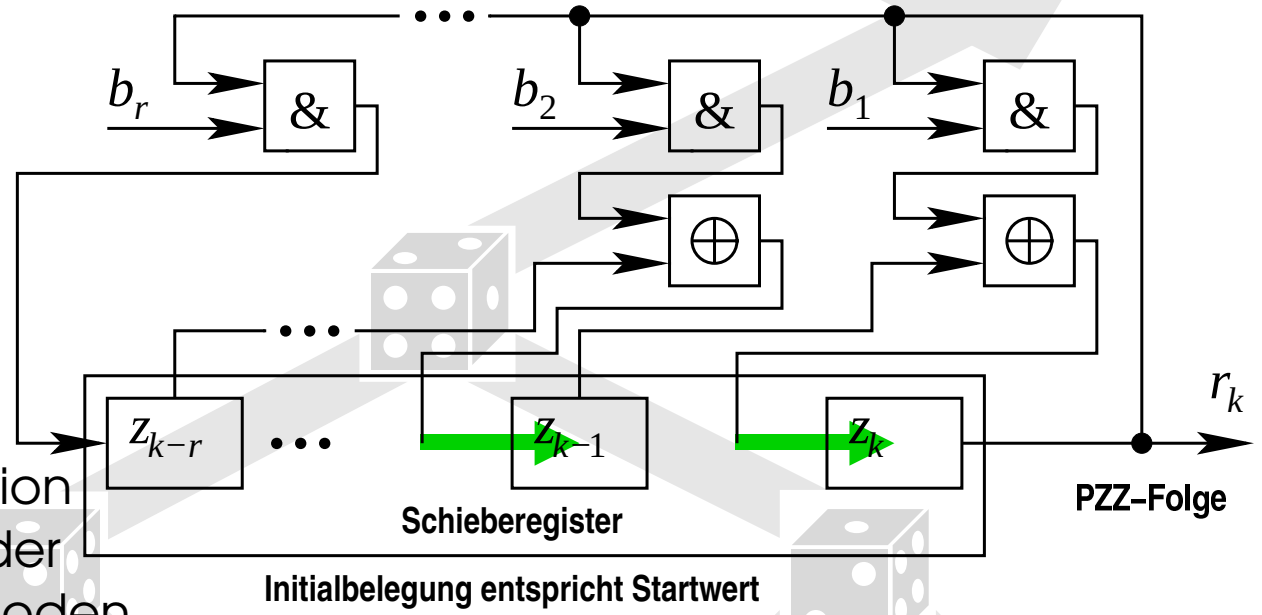
- nicht perfekt

Duales linear rückgekoppeltes Schieberegister

Galois-Schieberegister

Prinzip

- entspricht dem binären linear rückgekoppelten Schieberegister (SR)
- Einsparung einer $\oplus$-Operation
- zum Erreichen der maximalen Periodenlänge $2^r - 1$ muß charakteristisches Polynom **irreduzibel** sein
- Transformation duales $\leftrightarrow$ binäres linear rückgek. SR möglich



Vorteile

- leicht in Hardware implementierbar
- sehr schnell, einfache Parameterwahl

Nachteile

- nicht perfekt

Statistische Tests zur Bewertung der Prinzipien

- basieren auf konkreten erzeugten (Pseudo)Zufallszahlenfolgen
- Generatoren selbst nicht herangezogen
- Aufschluß darüber, **„wie gut“** Unabhängigkeit und Uniformität eingehalten werden
- **Hypothese** (Vorliegen Gleichverteilung, keine Autokorrelation) auf definiertem **Signifikanzniveau** (z.B. 5%, 1% Fehlertoleranz) akzeptiert oder verworfen
- Tests als Algorithmen der mathematischen Statistik notiert und angewendet, z.B.:

χ^2-Test (u.a. auf Gleichverteilung)

- Standardverfahren der mathematischen Statistik
- für lange (Pseudo)Zufallszahlenfolgen ($\geq$ 50 Zahlen) empfohlen

Kolmogorov-Smirnov-Test (auf Gleichverteilung)

- Standardverfahren der mathematischen Statistik
- i.a. genauer als χ^2-Test, aber aufwendiger, auch für kurze (Pseudo)Zufallszahlenfolgen

Gap-Test (auf Autokorrelation)

- Überprüfung der Intervallängen bis zum Wiedererscheinen der gleichen Zahl unter
- Anwendung des Kolmogorov-Smirnov-Tests

Poker-Test (auf Autokorrelation)

- analysiert die Häufigkeit, mit der sich die Ziffern in der (Pseudo)ZZ-Folge wiederholen

Mustersuche (auf Autokorrelation)

- graphische Darstellung und Auswertung der (Pseudo)Zufallszahlenfolge
- Gibt es wiederkehrende regelmäßige Muster oder zyklische Variationen?

Zusammenfassung

Resümee

- vorgestellte Prinzipien i.a. seit mehreren Jahrzehnten **bekannt** und
- im praktischen Einsatz **bewährt**
- Prinzipien decken ein **großes Spektrum** nutzbarer und eigenständiger **Strategien** ab
- Repertoire umfaßt Prinzipien, die bevorzugt **hardware-** oder **softwarebasiert** implementiert werden können
- Pseudozufallszahlengeneratoren, die üblicherweise in **Programmiersprachen** oder von **Computer-Algebra-Systemen** bereitgestellt werden, sind vorgestellt worden und spezialisieren sich durch ihre Parameterbelegungen

Ausblick

- weiterführende aufbauende Arbeiten beschäftigen sich u.a. mit der Untersuchung
  - von **Zufallsfunktionen** (Orakeln)
  - von **chaotischen Systemen**

Ausgewählte Literatur

L. Blum, M. Blum, M. Shub. *A Simple Unpredictable Pseudo-Random Number Generator.* SIAM J. on Computing **15(2)**:364–383, 1986

W. Göhler. *Höhere Mathematik.* Deutscher Verlag für Grundstoffindustrie Leipzig, 1986

A. Grube. *Moderne Erzeugung von Zufallszahlen.* S.-Toeche-Mittler-Verlag Darmstadt, 1975

D. Knuth. *The Art of Computer Programming.* Vol. 2: Seminumerical Algorithms. Addison-Wesley Ontario, 1998

N. Schmitz, F. Lehmann. *Monte-Carlo-Methoden I.* Verlag Anton Hain Meisenheim, 1976

B. Schneier. *Applied Cryptography.* John Wiley and Sons Inc. New York, 1994

R. Zielinski. *Erzeugung von Zufallszahlen.* Verlag Harri Deutsch Frankfurt/M., 1978