Digitale Wasserzeichen: Verteidigung von Authentizität und Urheberrecht in der Multimedia-Kommunikation

Prof. Dr.-Ing. Johannes Huber
Dipl.-Ing. Robert Bäuml, Dipl. Ing. Roman Tzschoppe
Lehrstuhl für Informationsübertragung, Technische Fakultät
Universität Erlangen–Nürnberg

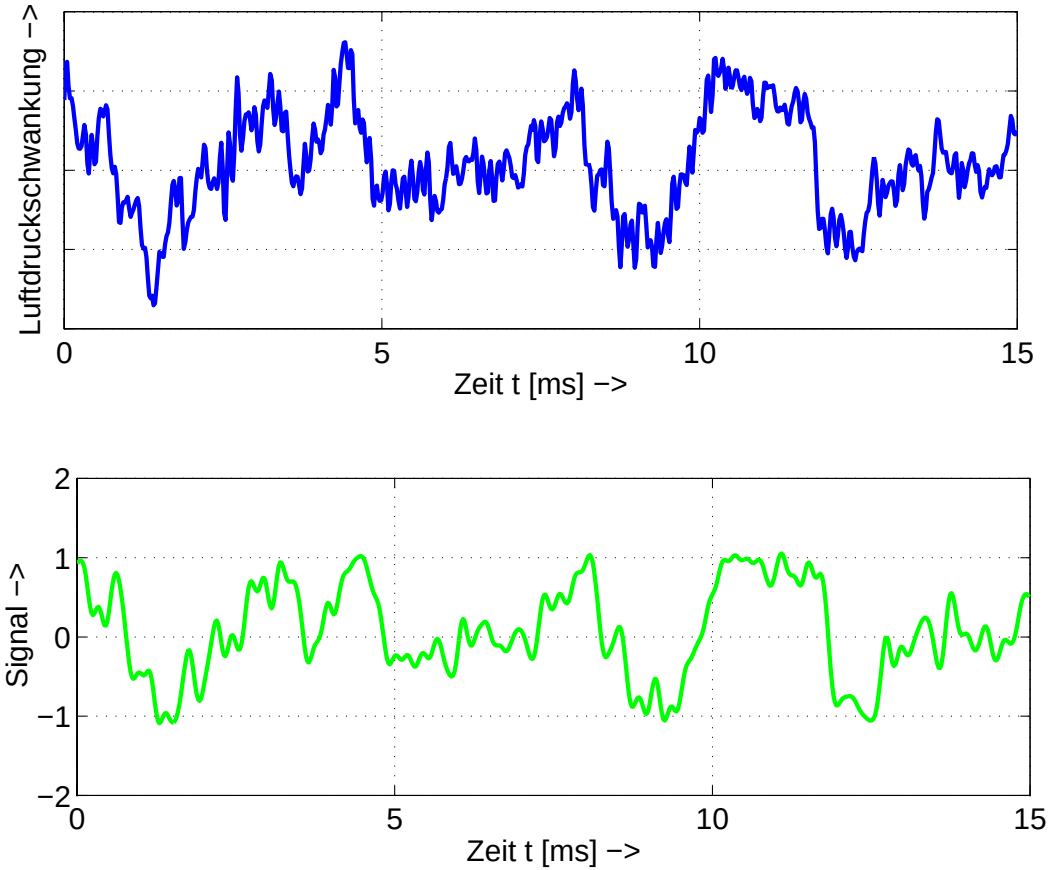
Arbeitskreis selbständiger Kultur-Institute e.V. AsKI
26. März, Frankfurt

Inhaltsübersicht

- Grundlagen der Multimedia-Kommunikation
- Digitale Wasserzeichen als Verteidigungsmaßnahme für Urheberrecht und Authentizität
- Klassifizierung von Verfahren
- Technik der digitalen Wasserzeichen
- Angriffe auf digitale Wasserzeichen
- Zusammenfassung

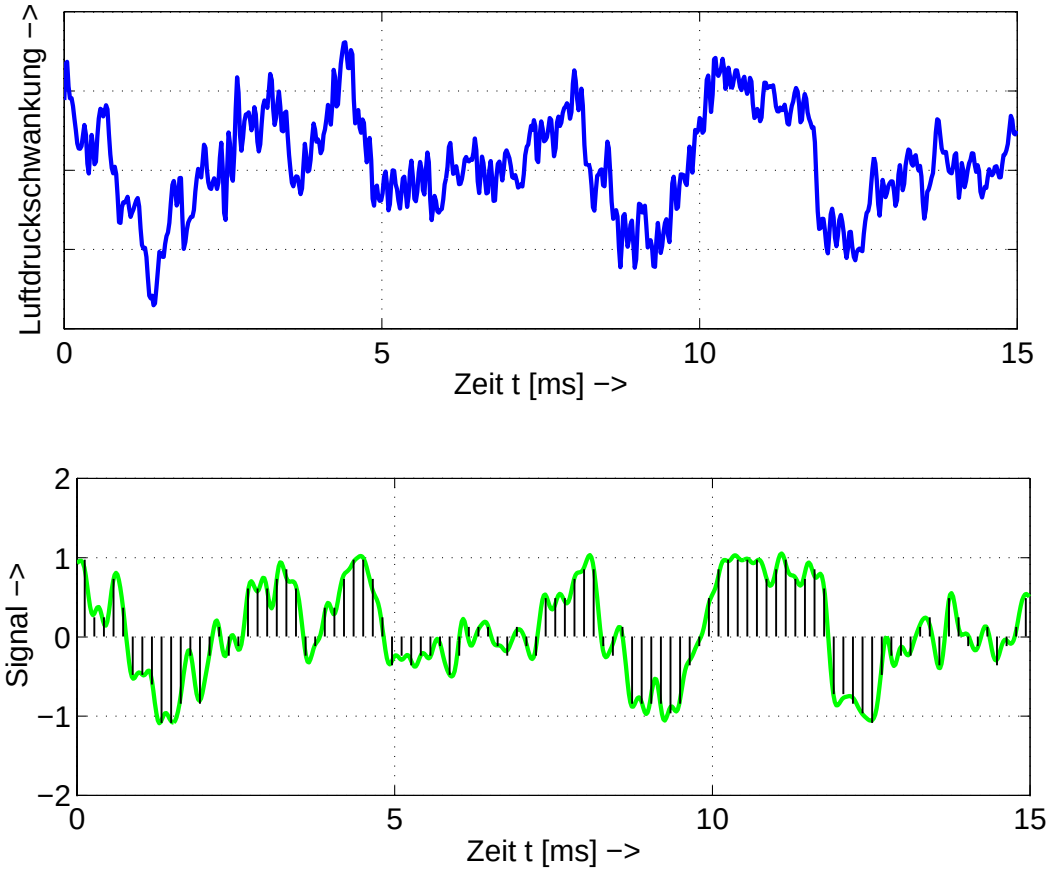
Digitalisierung, Quellencodierung, Kanalcodierung

Beispiel: Audiosignale

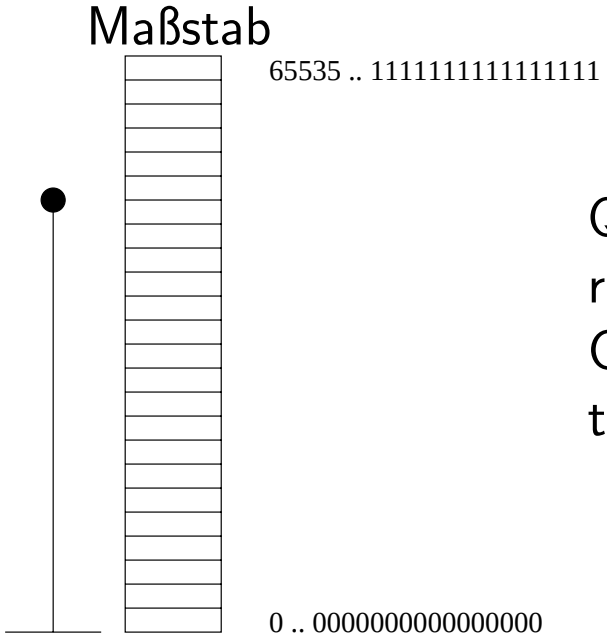


Digitalisierung, Quellencodierung, Kanalcodierung

Beispiel: Audiosignale



Quantisierung, Codierung



Quantisierung: irreversibler Genauigkeitsverlust, Verzerrung
Codierung: binäre Repräsentation der Nummer des Quantisierungsintervalls

Digitalisierung, Quellencodierung, Kanalcodierung

Audiosignal (CD): 44100 Abtastwerte/sec · 16 Bit/Abtastwert · 2 Kanäle = 1411200 Bit/sec ≈ 1,4 MBit/sec

Quellencodierung: Digitale Verarbeitung des Signals zur Reduktion von Redundanz und Irrelevanz

Redundanz: Nachricht die nicht unbedingt zur eindeutigen Informationsrepräsentation notwendig ist.

Irrelevanz: Information, die der Verbraucher zu verwerten nicht in der Lage ist.

z.B. MPEG Audio Codierung Layer 3 (MP3): ≈ 100 kBit/sec; Reduktionsfaktor 14

Videosignal: (625 Zeilen, PAL-Qualität): 13000000 Abtastwerte/sec · 10 Bit/Abtastwert = 130MBit/sec

z.B. MPEG2 ≈ 4 MBit/sec; Reduktionsfaktor 32

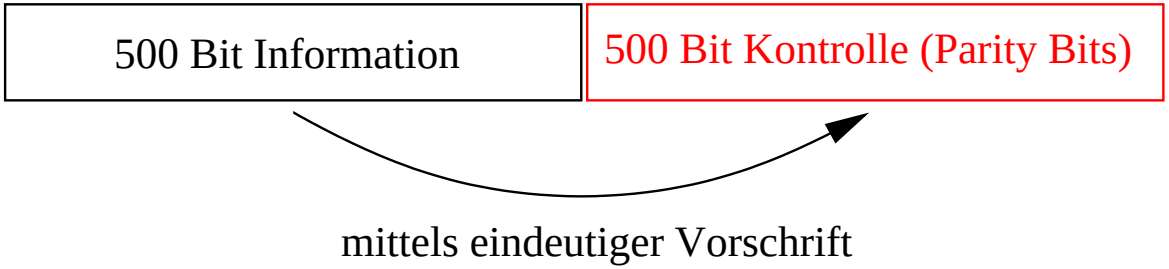
z.B. MPEG4 ≈ 2 MBit/sec; Reduktionsfaktor 64

Digitale Fotografie: 4000000 Bildpunkte · 12 Bit/Punkt · 3 Farben = 144 MBit/Bild = 18 MByte/Bild

z.B. JPEG ≈ 0,8 MByte/Bild; Reduktionsfaktor 22

Kanalcodierung

Sicherung von Daten vor Verfälschung durch gezieltes Einbringen von Redundanz, z.B. Blockcodierung mit Rate $1/2$ und Codewortlänge 1000



bis zu 57 verfälschte (invertierte) Bit sind korrigierbar!

→ Kanalcodierung erlaubt eine fehlerfreie Datenübertragung auch über gestörte Übertragungskanäle

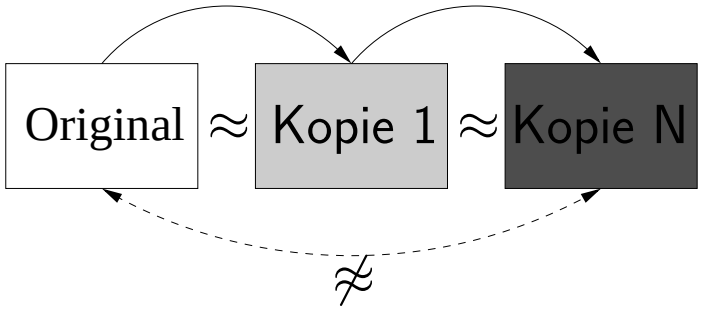
→ Bei genügend Redundanz sind Fehler prinzipiell vermeidbar bzw. korrigierbar

Basis der digitalen Kommunikation (Digitalisierung, Quellencodierung, Kanalcodierung, Kryptologie etc.): Informationstheorie (C.E. Shannon 1916-2001)

Motivation

Analoge Medien

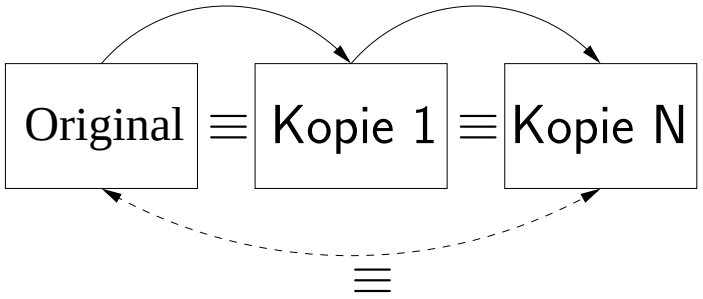
Audio Cassetten
Fotographien
VHS Bänder



- Vertrieb erfordert aufwändige Infrastrukturen
- Manipulationen sind , daher kostenintensiv und oft leicht erkennbar
- Qualitätsverlust von Kopie zu Kopie
- $\Rightarrow$ inhärenter Schutz gegen

Digitale Medien

CD, MP3 Audio
JPEG Bilder
DVD, MPEG Video

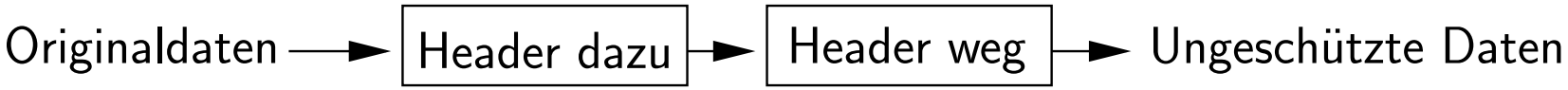


- Vertrieb über Datennetz
- Simple Manipulation durch extrem leistungsfähige Software
- Regeneration durch Kopieren, Bitfehler werden durch Kanalcodierung sogar wieder korrigiert
- Kein Schutz gegen

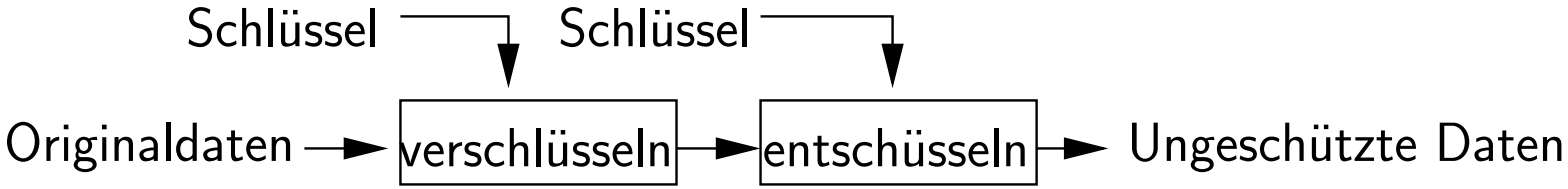
Kopieren, ungehemmt exponentielle Verbreitung, Manipulationen usw.

Traditionelle Schutzmechanismen

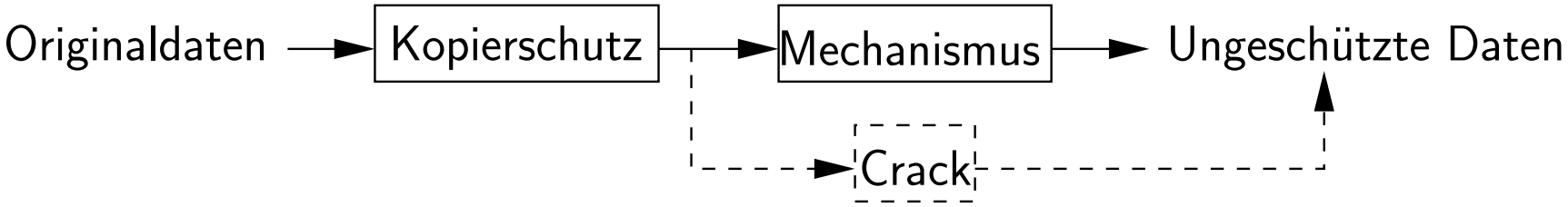
- Zugriffskontrolle durch Headerinformation: einfach zu ändern/entfernen



- Verschlüsselung: entschlüsselte Daten sind wieder ungeschützt

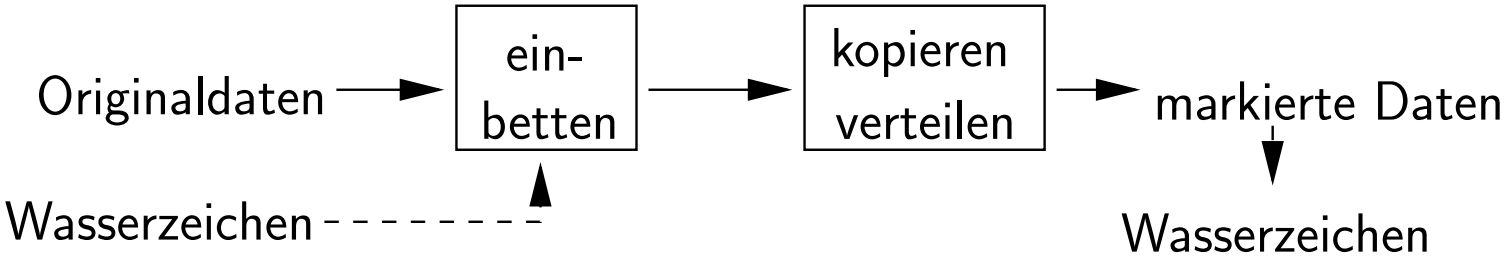


- Kopierschutz: Möglichkeit des Knackens



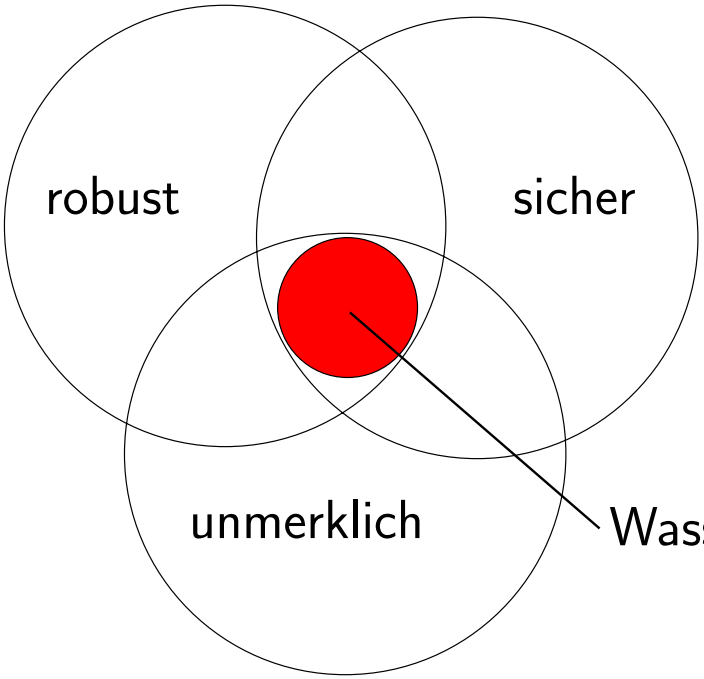
Wasserzeichen Einbettung

- $\Rightarrow$ Einbetten von Information die sich zusammen mit den Daten verbreitet

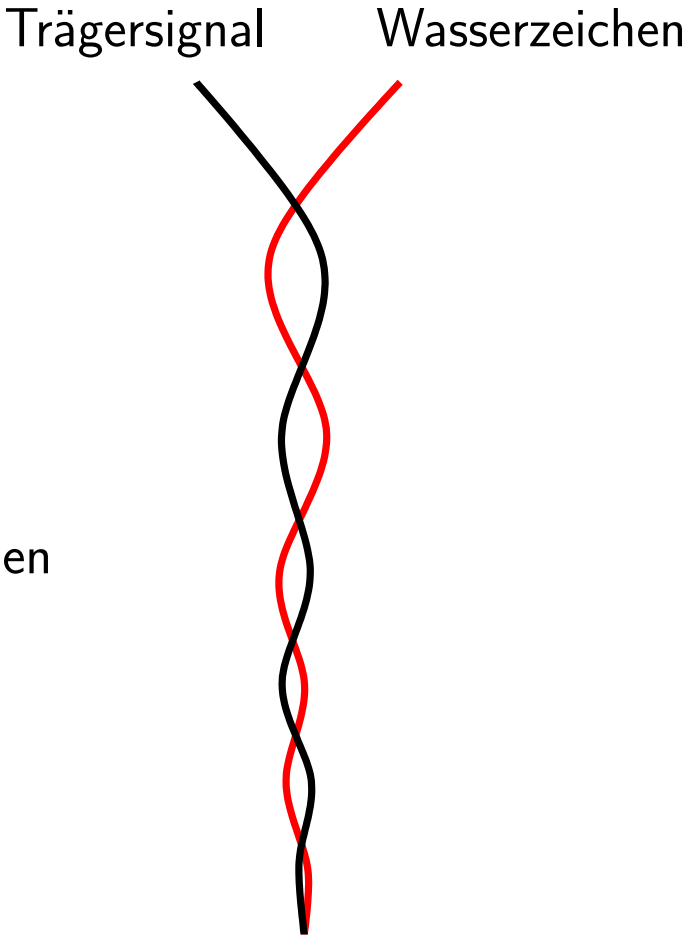


- Benötigt zusätzliche Rahmenbedingungen um Schutz zu bieten!
- "Last line of defense", gewöhnlich kombiniert mit traditionellen Methoden
- im weiteren Sinne analog zu Wasserzeichen in Papierdokumenten (z.B. Geldscheine)

Gewünschte Eigenschaften



Prinzipieller Austausch: eingebettete Informationsmenge gegen Einbettungsverzerrung und Robustheit gegen Angriffe



Anwendungen

- Zugriffskontrolle
  - Wiedergabe- und Kopierkontrolle
  - Urheberrechtsschutz, Eigentumsnachweis
- Verbreitungsverfolgung
  - digitaler Fingerabdruck
  - Identifizierung kollektiver Angriffe
- Urheberrechtsnachweis, Urheberidentifizierung
- Übertragungsüberwachung
- Authentifizierung von Medien (zerbrechliche Wasserzeichen)
- Steganographie
- Meta-Information
  - SmartImages von Digimarc Corp, ...
- u.v.a.m.

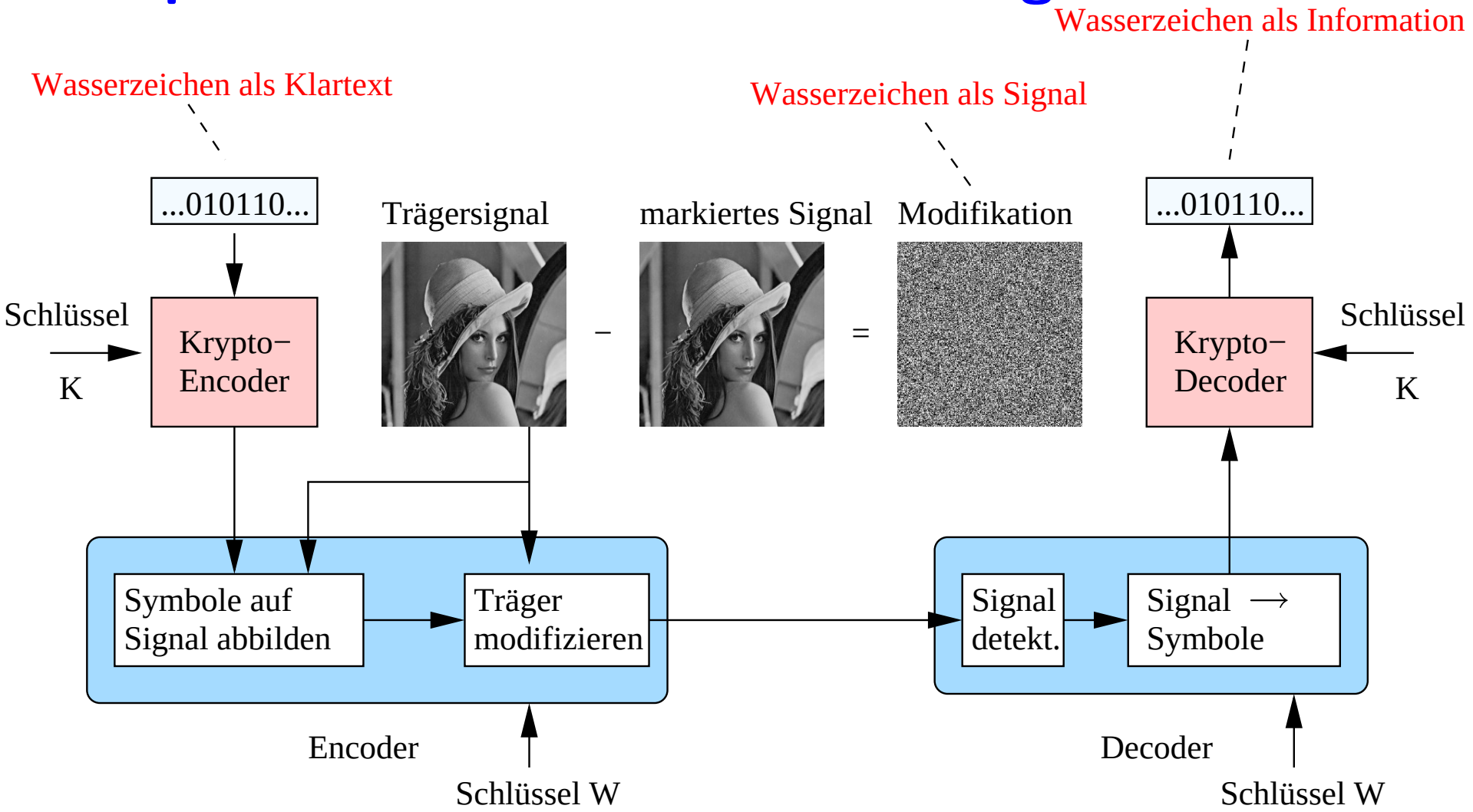
Einschränkungen

- Digitale Wasserzeichen verhindern nicht eine unberechtigte Verteilung von Dokumenten – aber eingebettete Information verbleibt (hoffentlich) in den Daten und erleichtert Verfolgung.
- Digitale Wasserzeichen: keine Komplettlösung für Zugriffs-/ Kopierkontrolle und/oder Urheberrechtsschutz!
- Digitale Wasserzeichen können nur einen Teil eines größeren Systems zum Schutz digitaler Information gegen unrechtmäßige Nutzung darstellen.

Zusätzliche Aspekte

- "Blinde" Wasserzeichen
  - Kein Zugriff des Decoders auf die ursprünglichen Daten möglich
  - Mögliche Interferenz mit dem Trägersignal
- Mehrfach-Wasserzeichen
  - Mehrere Datenströme innerhalb einer Kopie
  - Unterschiedliche Daten in unterschiedlichen Kopien
- Direkte Verarbeitung komprimierter Daten
  - Kombinierte Einbettung von Wasserzeichen mit Kompression
  - Ratenbegrenzung
- Implementierung
  - Geschwindigkeit, Komplexität, Größe, Kosten, ...

Prinzipielle Methoden der Einbettung



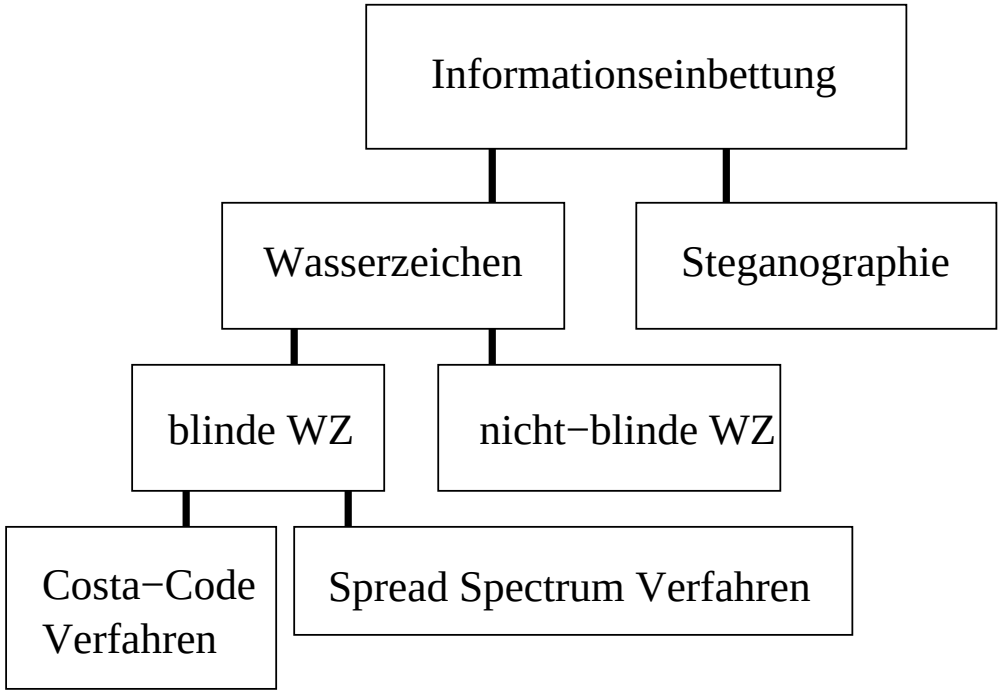
Technik digitaler Wasserzeichen

- Einbindung des Wasserzeichens auf der Ebene des vom Menschen verwertbaren Dokuments (Bild, Ton, etc.), nicht auf der Ebene des Datenstroms
  - Robustheit gegenüber unterschiedlichsten Arten der Repräsentation und Verarbeitung (Quellencodierverfahren, Kopierverfahren (Digital/Analog, optisch , elektroakustisch))
  - Kontrolle der Einbettungsverzerrung: Das Wasserzeichen soll zum einen nicht störend wirken, zum anderen soll für den Rechtsbrecher gar nicht erkennbar sein, dass der Schutz durch ein Wasserzeichen vorliegt, so dass er nicht motiviert wird das Wasserzeichen anzugreifen.
  - Forderung: Robustheit gegenüber Manipulation des Signals. Eine Manipulation, die ein digitales Wasserzeichen unlesbar macht, sollte auch das Trägersignal wertlos werden lassen!

Verbindung von digitalen Wasserzeichen mit Verfahren der Kryptologie

- auch wenn der Angreifer merkt, daß ein Wasserzeichen hinterlegt ist, soll dessen Bedeutung verborgen bleiben.
- alle gängigen Verfahren der Kryptologie mit geheimen Schlüsseln (Problem des Schlüsselaustausches!) oder öffentlichen Schlüsseln (public-key-Verfahren, z.B. RSA) können Anwendung finden.

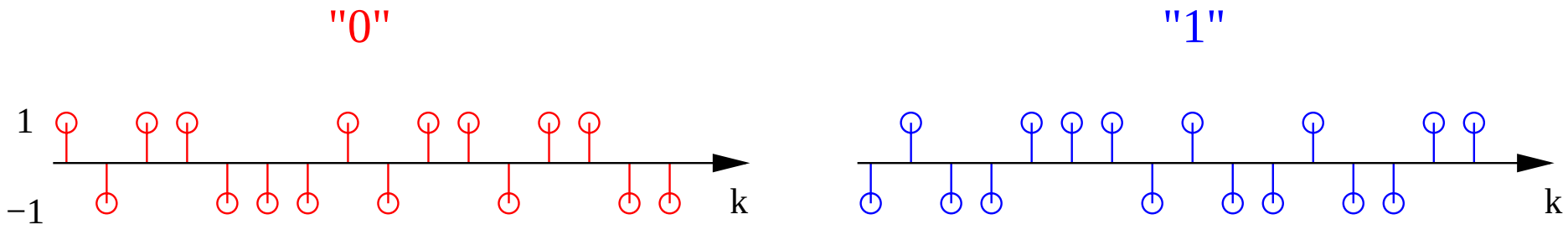
Technische Hierarchie



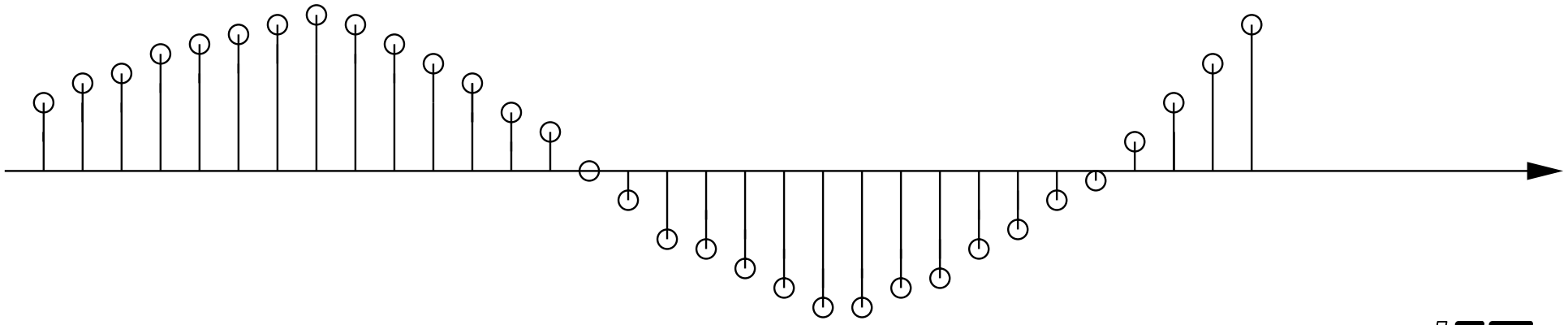
- WZ-Decodierung: Nutzbarmachung der WZ-Information
- WZ-Detektion: Nachweis der die Existenz eines Wasserzeichens
- Blinde Wasserzeichen: Keine Nutzung des originalen Trägersignals bei der Decodierung möglich

Spread Spectrum Technik

a) Repräsentanten der Binärsymbole durch viele (N) pseudozufällige Impulse ± 1
Beispiel: N=16

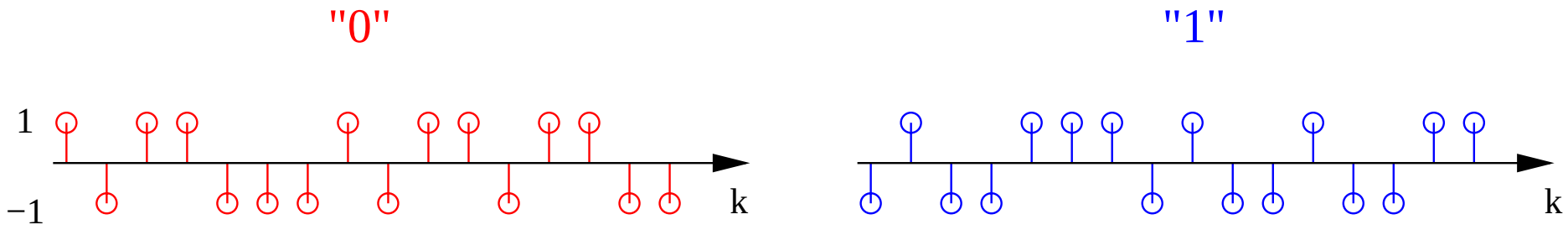


b) Überlagerung auf das Trägersignal mit sehr kleiner Leistung (wenig Einbettungsverzerrung)

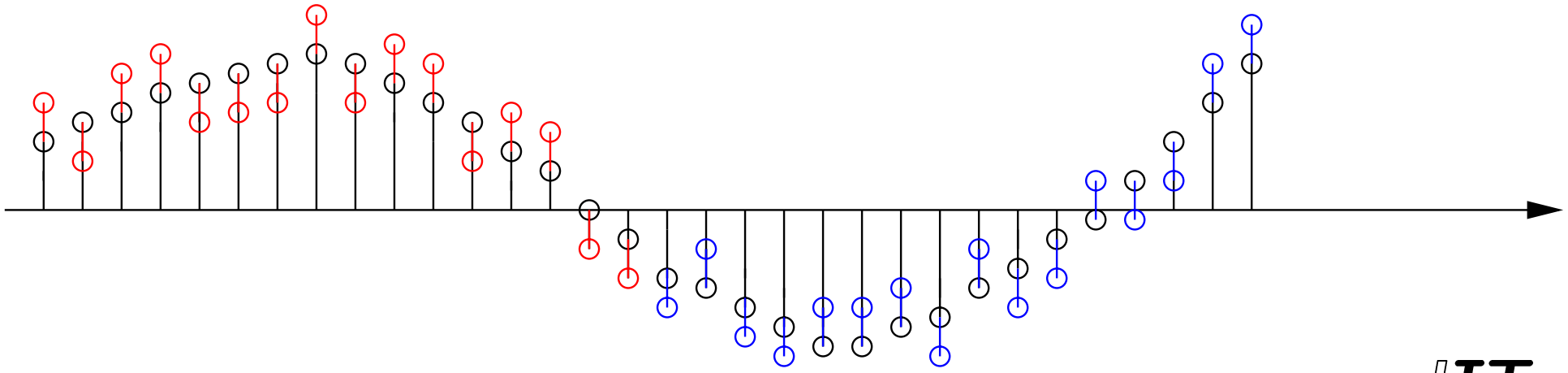


Spread Spectrum Technik

a) Repräsentanten der Binärsymbole durch viele (N) pseudozufällige Impulse ± 1
Beispiel: N=16

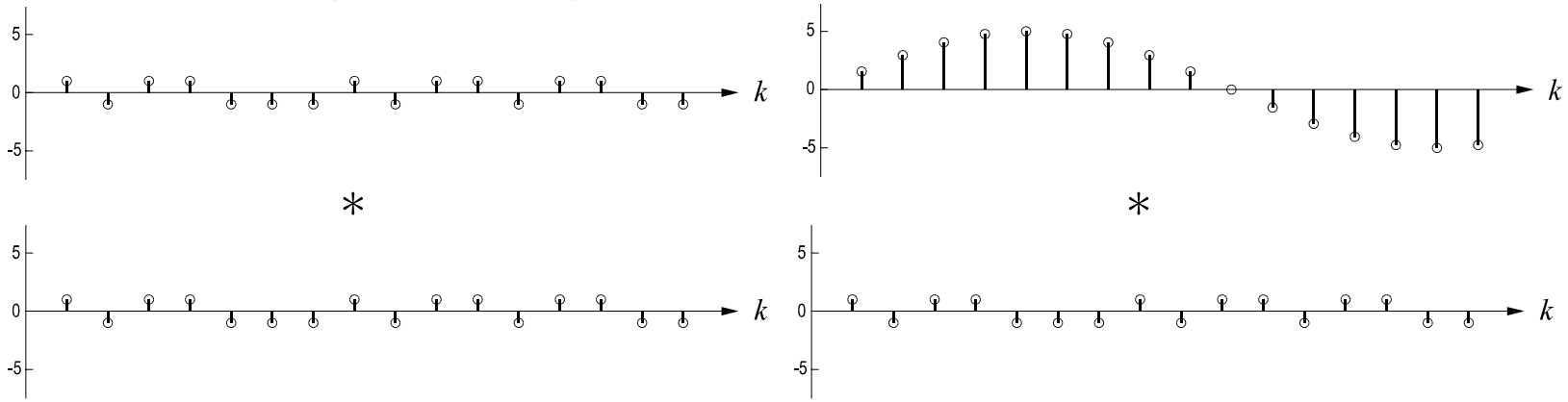


b) Überlagerung auf das Trägersignal mit sehr kleiner Leistung (wenig Einbettungsverzerrung)



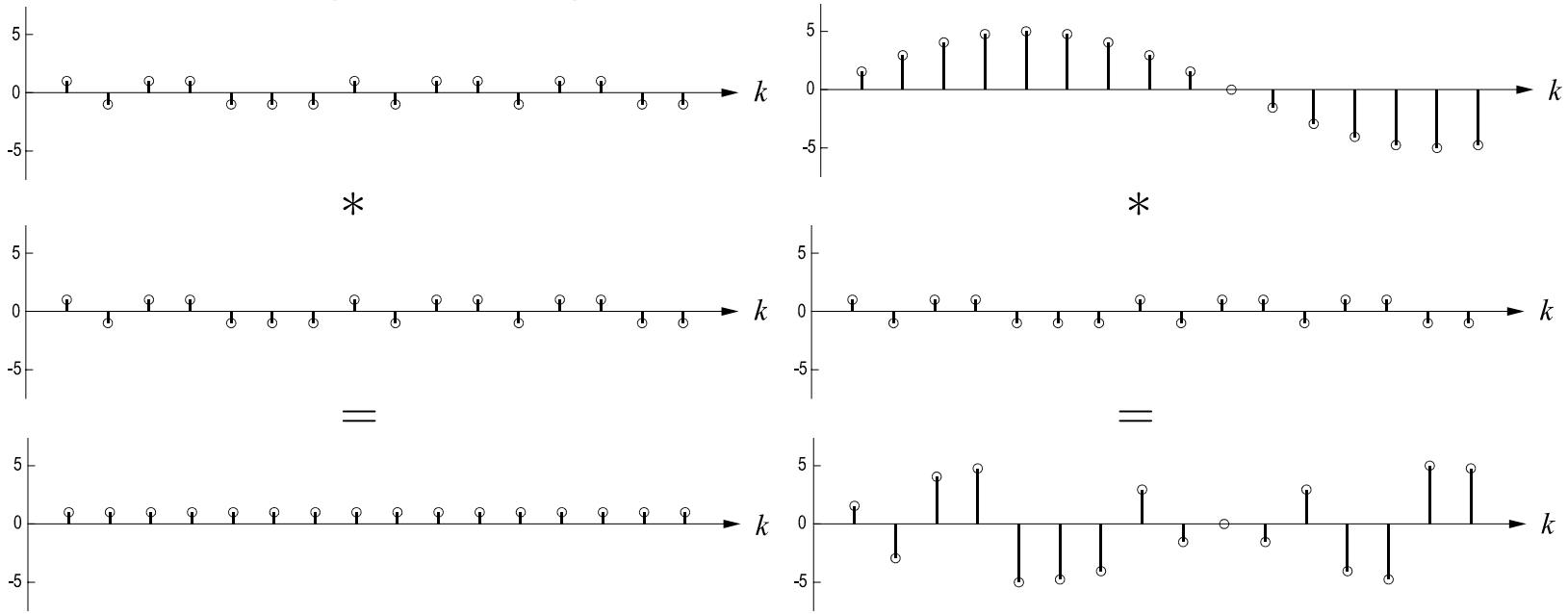
Spread Spectrum Technik

c) Ausfiltern des Wasserzeichensignals durch Multiplikation mit Pseudo-Zufallsfolge und Aufsummieren (Korrelation)



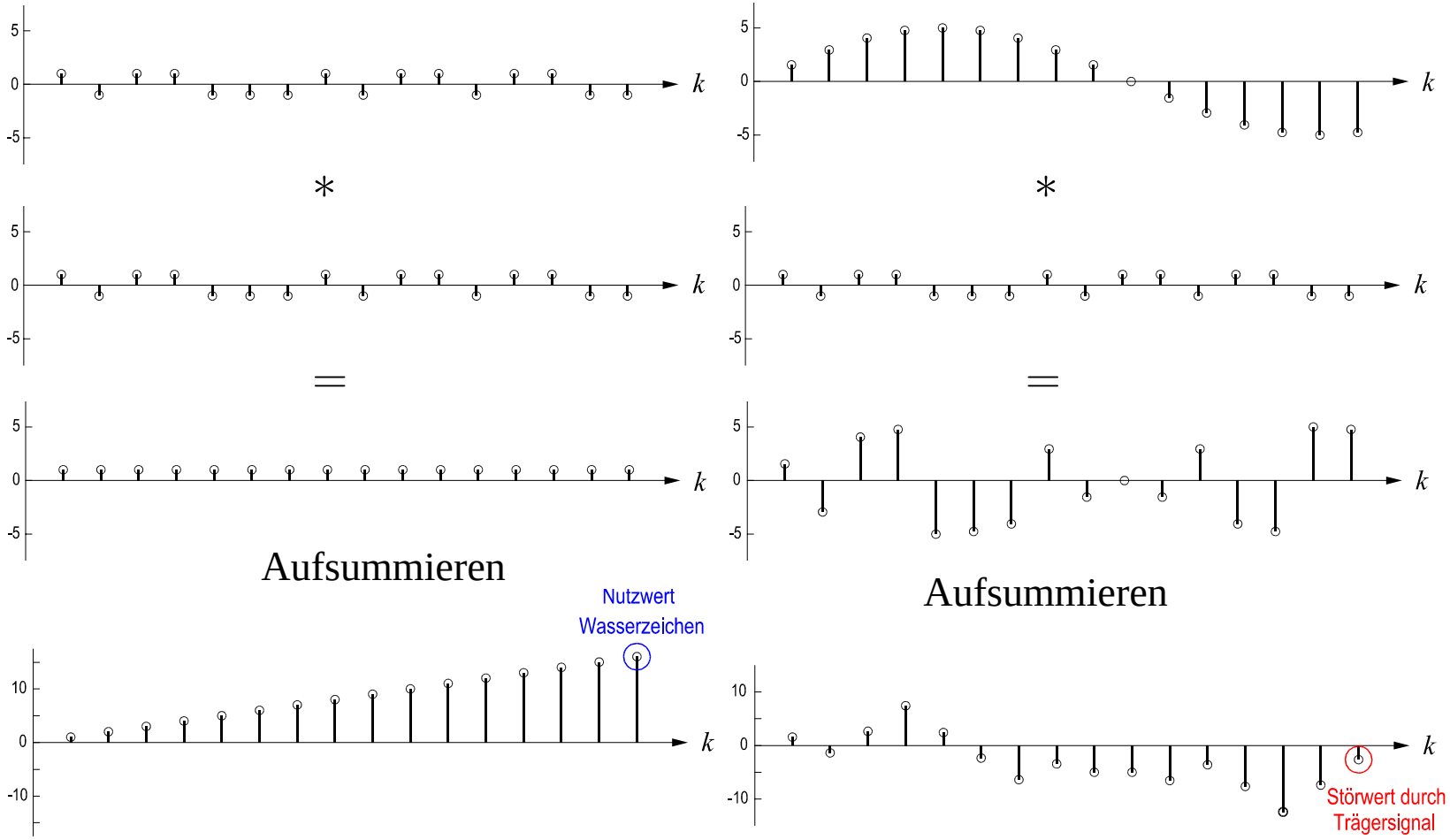
Spread Spectrum Technik

c) Ausfiltern des Wasserzeichensignals durch Multiplikation mit Pseudo-Zufallsfolge und Aufsummieren (Korrelation)

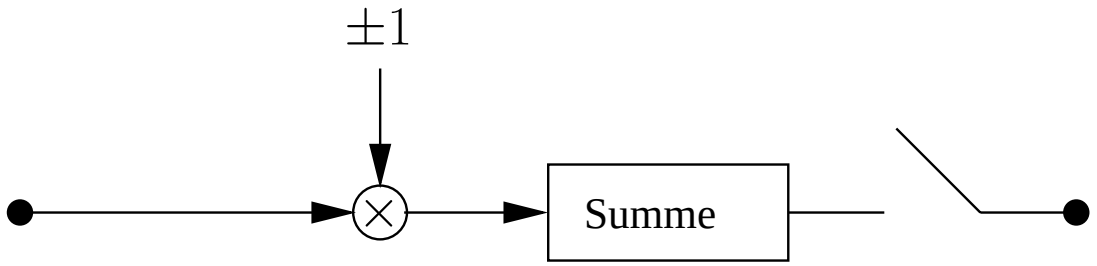


Spread Spectrum Technik

c) Ausfiltern des Wasserzeichensignals durch Multiplikation mit Pseudo-Zufallsfolge und Aufsummieren (Korrelation)



Spread Spectrum Technik



Verbesserung der Leistung des Wasserzeichensignals im Vergleich zum Trägersignal im Mittel um Faktor N

z.B. Wasserzeichen um Faktor 1000 schwächer als Trägersignal (30dB Einbettungsverzerrung)

Erforderliches Verhältnis

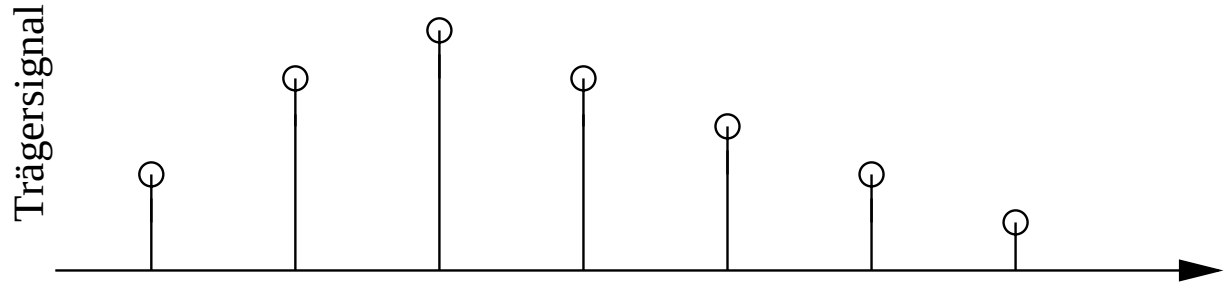
$$\frac{\text{Leistung des Wasserzeichensignals}}{\text{Störung durch Trägersignal}} \geq 10$$

$\rightarrow N = 10000$ Ein Bit Wasserzeicheninformation je 10000 Signalwerte des Trägersignals! (Beispiel CD: Audiosignal 4,41 Bit/Kanal/sec Wasserzeichenrate)

$\rightarrow$ Spread Spectrum Wasserzeichen sind nur für eine äußerst geringe Wasserzeicheninformationsmenge geeignet.

Verfahren beruhend auf Costa-Codes

Einführendes Beispiel: Quantization Index Modulation grob übertrieben dargestellt.
Die Abtastwerte aus dem Trägersignal liegen als Dualzahlen vor:



Dezimal	2	4	5	4	3	2	1
Dual	010	100	101	100	011	010	001

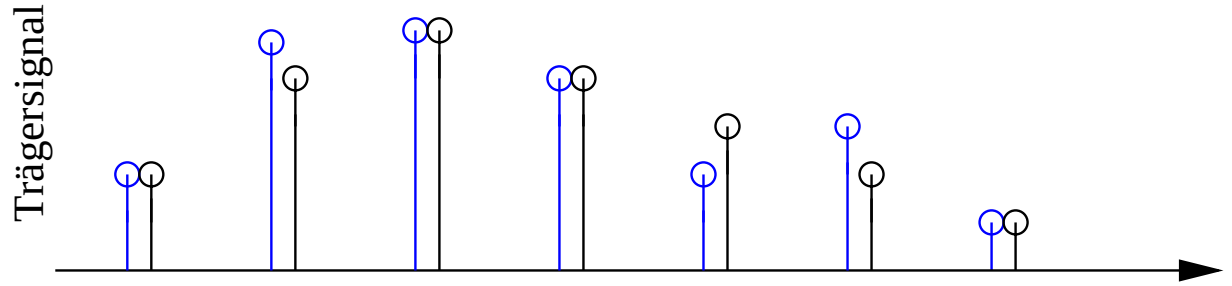
Die niederwertige Stelle wird für das Wasserzeichen geraubt
und durch die Wasserzeichenfolge

	0	1	1	0	0	1	1 ersetzt
Dual	010	101	101	100	010	011	001
Dezimal	2	5	5	4	2	3	1

→ Nach Quantisierung des Signals kann das Wasserzeichen direkt als niederwertige Stelle ausgelesen werden

Verfahren beruhend auf Costa-Codes

Einführendes Beispiel: Quantization Index Modulation grob übertrieben dargestellt.
Die Abtastwerte aus dem Trägersignal liegen als Dualzahlen vor:



Dezimal	2	4	5	4	3	2	1
Dual	010	100	101	100	011	010	001

Die niederwertige Stelle wird für das Wasserzeichen geraubt
und durch die Wasserzeichenfolge

	0	1	1	0	0	1	1 ersetzt
Dual	010	101	101	100	010	011	001
Dezimal	2	5	5	4	2	3	1

→ Nach Quantisierung des Signals kann das Wasserzeichen direkt als niederwertige Stelle ausgelesen werden

Verfahren beruhend auf Costa-Codes

$\longrightarrow$ 1 Bit Wasserzeicheninformation je Wert des Trägersignals

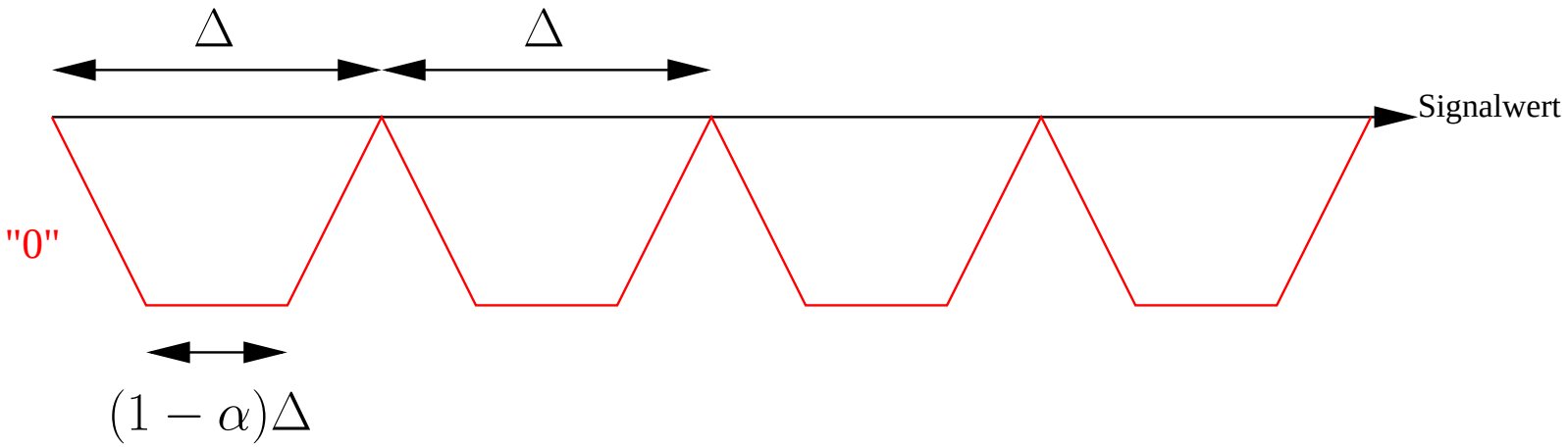
Problem: Extrem hohe Empfindlichkeit gegenüber Störungen und Angriffen

Abhilfe 1: Kanalcodierung

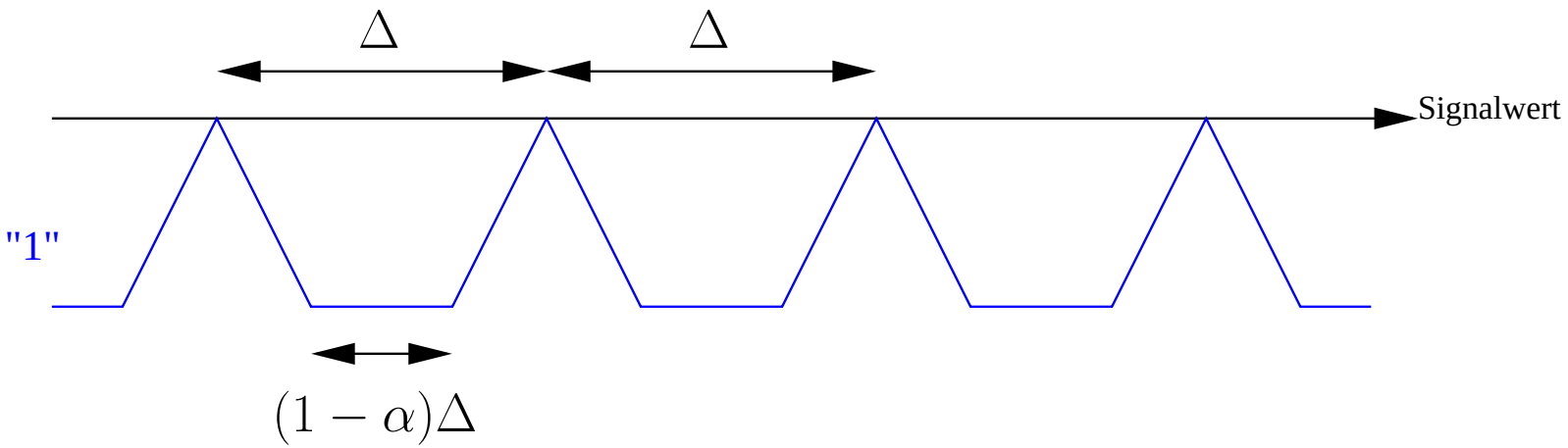
Anwendung eines niederratigen Kanalcodes $\longrightarrow$ Verteilung von wenigen Bit Wasserzeicheninformation auf sehr viele Werte des Trägersignals
z.B. Code der Rate 1/20 und Wortlänge 2000: 100 Bit Information, 1900 binäre Kontrollsymbole; sehr hohe Störresistenz erreichbar (Störung darf gegenüber Nutzen ca. um Faktor 20 überwiegen!)

Verfahren beruhend auf Costa-Codes

Abhilfe 2: Skalares Costa Schema (SCS-Watermarking)



Stauchung der Signalwerte innerhalb von Intervallen der Breite Δ auf Intervalle der Breite $(1-\alpha)\Delta$, $\alpha < 1$



Gestauchte Intervalle um $\Delta/2$ verschoben

Verfahren beruhend auf Costa-Codes

$\alpha < 1/2$: überlappende Bereiche – geringe Einbettungsverzerrung und dennoch robuste Wasserzeichendetektion infolge niederratiger Kanalcodierung möglich!!

Nachteil: Statistische Analysen verraten die Rasterung und Stauchung der Signalwerte

Abhilfe:
- Verschiebung der Stauchungsraster $i \cdot \Delta(+\Delta/2)$ um Pseudozufallszahlen $w[l]$ gleichverteilt im Intervall $0 \leq w[l] < \Delta$, Wahl von $w[l]$ statistisch unabhängig von den einzelnen Werten des Trägersignals.
- Werte $w[l]$ nur dem authorisierten Wasserzeichendetektor/decoder bekannt (Schlüssel!)
- Aufhebung der Konzentration der Signalwerte auf bestimmte Intervalle.
- Statistische Analysen liefern keinen Hinweis auf Wasserzeicheneinbettung.
- Keine Erhöhung der Einbettungsverzerrung

Informationstheorie: Costa-Verfahren bieten den günstigsten Austausch zwischen Einbettungsverzerrung, Wasserzeichenrate und Sicherheit gegen Angriffe durch additive Störung.

Authentizität

Beispiel zur Aufdeckung einer Bildmanipulation durch Zerstörung des Wasserzeichens

Authentizität

Beispiel zur Aufdeckung einer Bildmanipulation durch Zerstörung des Wasserzeichens

Authentizität

Beispiel zur Aufdeckung einer Bildmanipulation durch Zerstörung des Wasserzeichens

Authentizität

Beispiel zur Aufdeckung einer Bildmanipulation durch Zerstörung des Wasserzeichens

Angriffe auf eingebettete Wasserzeichen

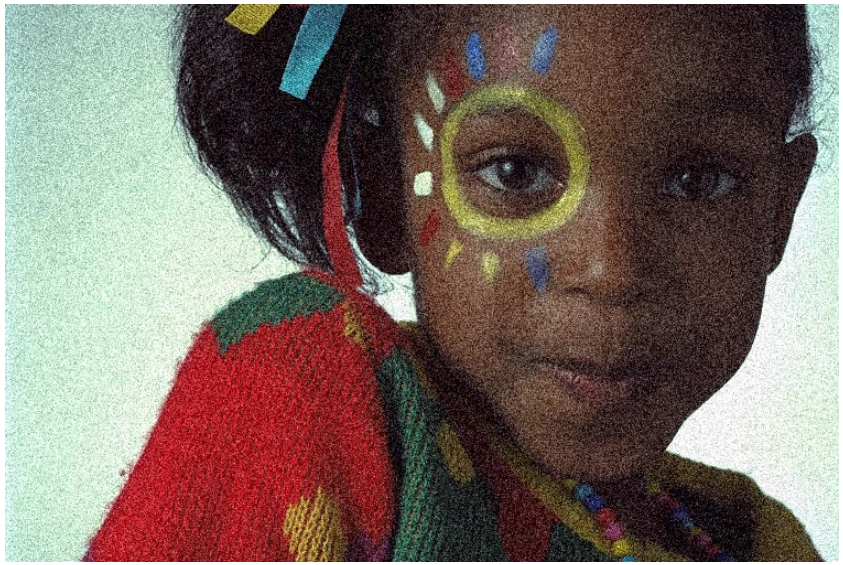
Zweck: Unleserliches Wasserzeichen verliert jegliche Beweiskraft

Problem des Angreifers: Zerstörung des Wasserzeichens ohne maßgebliche Beeinträchtigung des Trägersignals

Austausch: Angriffswirksamkeit $\leftrightarrow$ Angriffsverzerrung

einfaches Beispiel: Addition von Rauschen auf das Signal ($10 \cdot \log_{10}$ WNR $= -10$ dB)

$\rightarrow$



StirMark Angriff

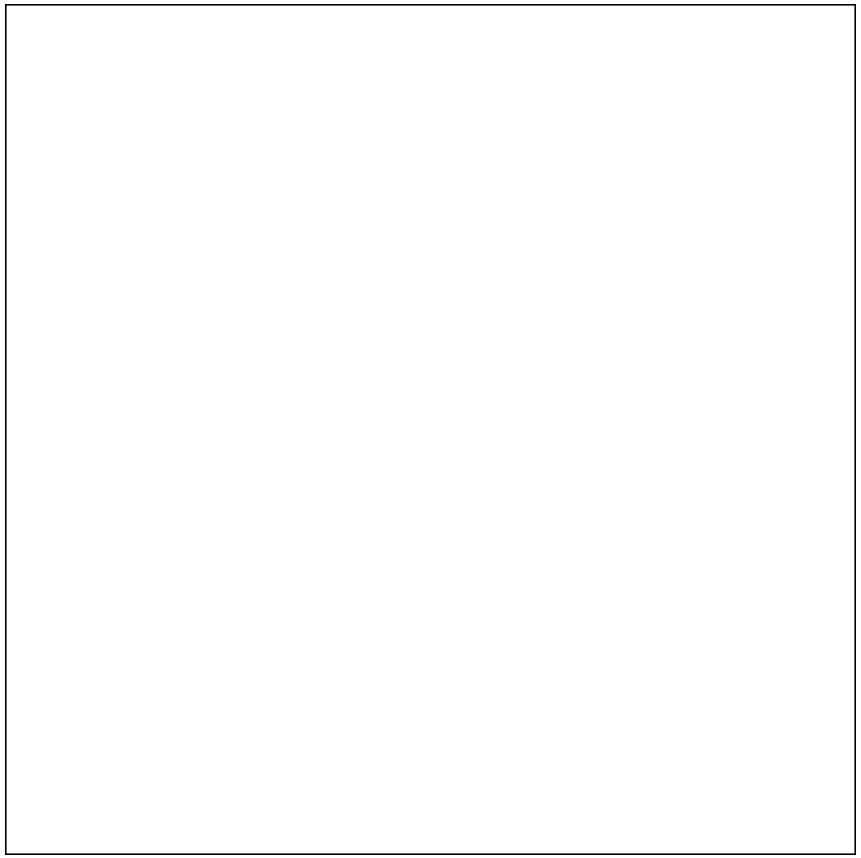
Original Lenna

Lenna nach Angriff

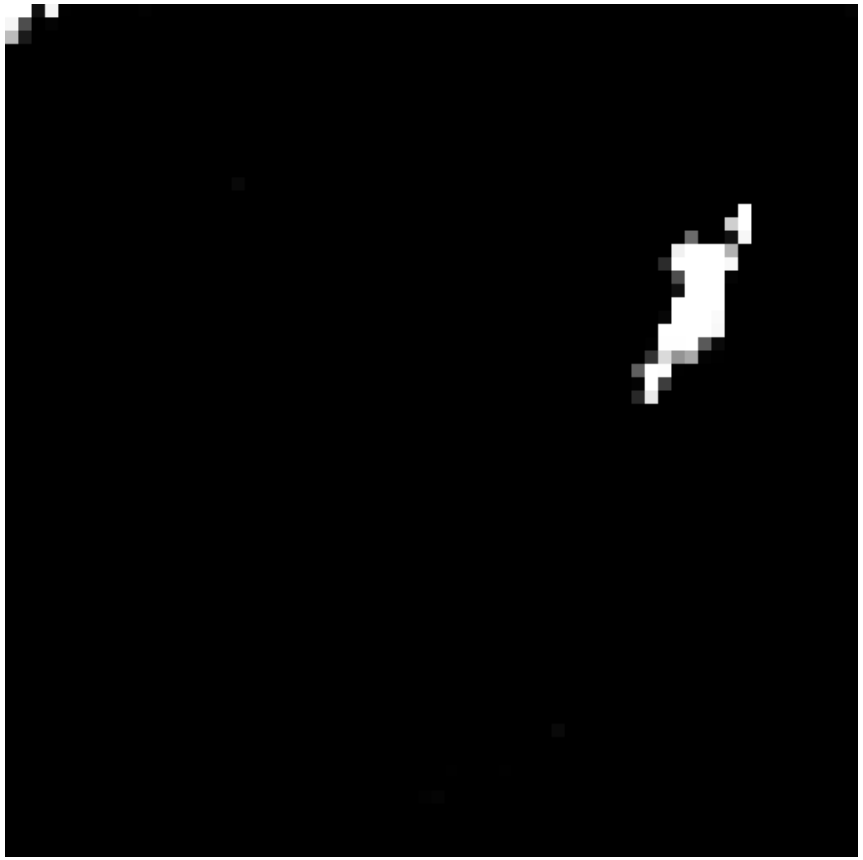
Gibt es wahrnehmbare Unterschiede?

Wasserzeichendetektionssicherheit

(weiss: sicher, schwarz: unsicher)

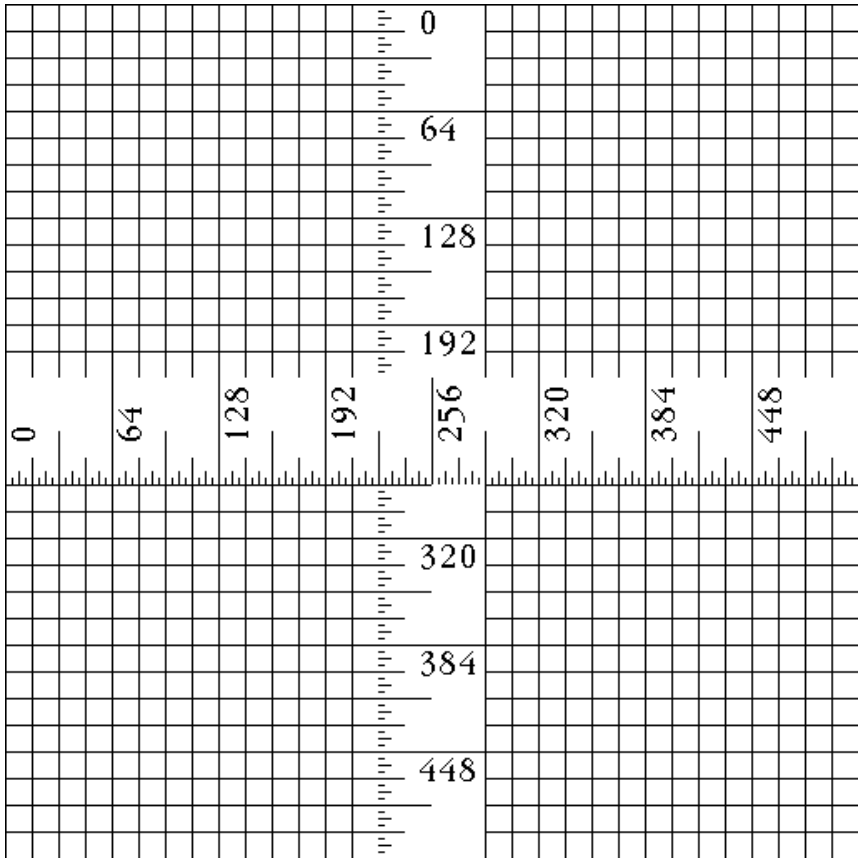


Original Lenna

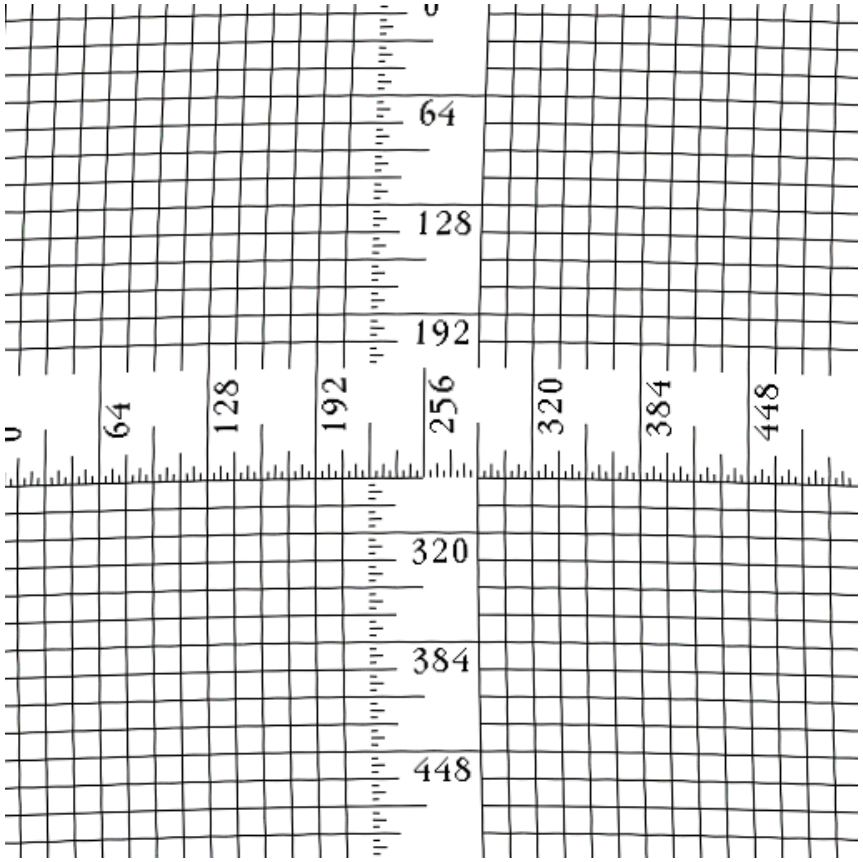


Lenna nach Angriff

Was geschieht bei diesem Angriff?

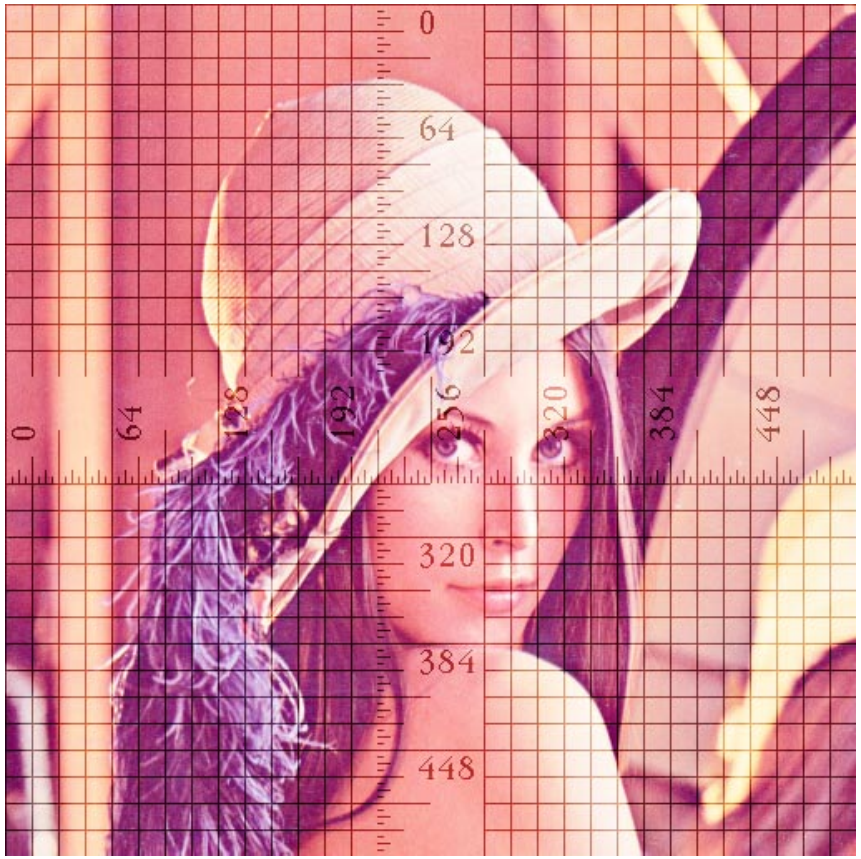


Original Raster

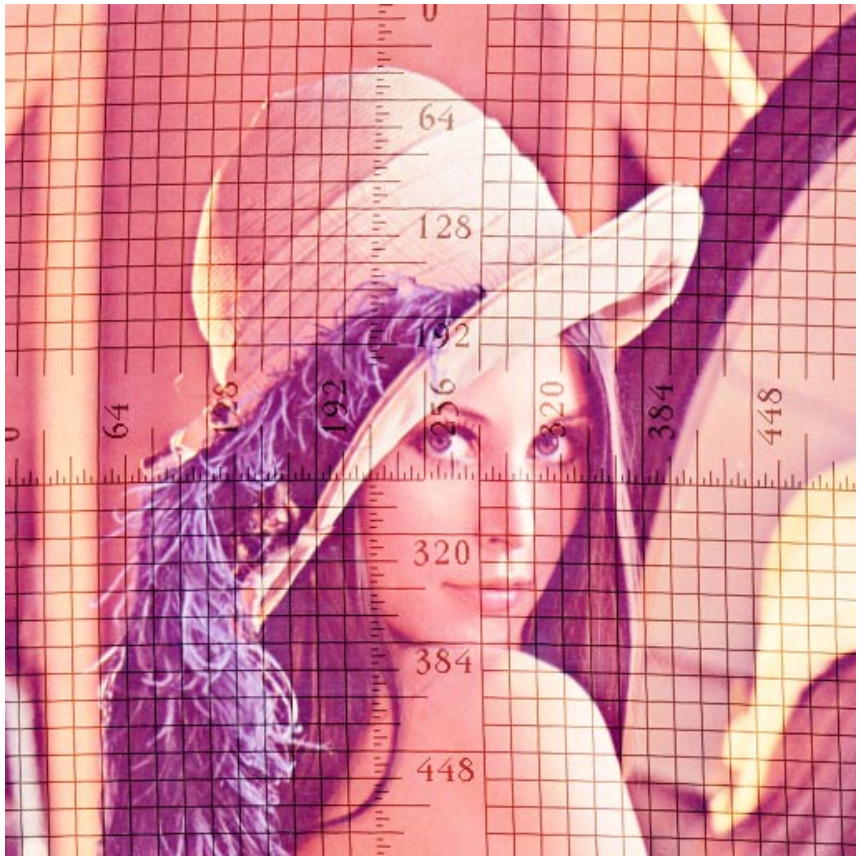


Raster nach Angriff

StirMark Angriff



Original Lenna mir Raster



Lenna nach Angriff mit Raster

Geschickte Desynchronisationsangriffe

- **sind nicht wahrnehmbar**
- **verhindern konventionelle Wasserzeichendetektion**

Zusammenfassung

- Wasserzeichen bauen auf grundlegende Methoden der digitalen Signalverarbeitung und Übertragung auf
- Wirkung nur in Verbindung mit traditionellen Methoden des Kopierschutzes und der Kryptologie
- Wasserzeichen bieten neben Kopierschutz viele weitere Möglichkeiten
- Spread Spectrum Wasserzeichen sind für niedrige Datenraten geeignet
- Costa-Codes liefern hohe Raten bei bekannten Randbedingungen
- Schlaue Angreifer versuchen aus den Randbedingungen des Einbetters auszubrechen
- Synchronisation auf das Wasserzeichensignal ist wohl das größte Problem bei Detektion / Decodierung