Modellbildung und Simulation

Sommersemester 2011

2. Vorlesung

Klaus Kasper

Termine

Veranstaltung	Vorlesung Mo 18-21 (X) D14/104	Praktikum Mo 18-21 (Y) D15/107	Praktikum Fr 12-15 (X) D15/107
1. Termin	28.03.11	04.04.11	08.04.11
2. Termin	11.04.11	18.04.11	29.04.11
3. Termin	02.05.11	09.05.11	13.05.11
4. Termin	16.05.11	23.05.11	27.05.11
5. Termin	30.05.11	06.06.11	17.06.11
6. Termin	20.06.11	27.06.11	01.07.11

Organisation

- Moodle-Server: http://lernen.h-da.de/
- Kurs: Informatik/Informatik (M.Sc.)/Sommersemester 2011/Modellbildung und Simulation – Kasper - SS 2011
- Schlüssel: MuS_KAS_SS2011
- Vorlesungen, Aufgabenblätter und weitere Materialien werden auf dem Moodle-Server zur Verfügung gestellt.
- Fragen und Anmerkungen zur Veranstaltung bitte immer über das Diskussionsforum der Veranstaltung, so dass immer alle informiert sind.
- Wenn Sie interessante Informationen finden, stellen Sie diese bitte auch allen zur Verfügung.

Inhalt

- Wiederholung
  - Digitalisierung
- Parametrisierung
  - Fourier-Transformation
  - Fast Fourier-Transformation (FFT)
- Value-at-Risk
  - Einführung
  - statistische Kenngrößen
  - Normalverteilung

Digitalisierung

Glossar

Digitalisierung
Konvertierung analoger Signale in eine digitale Darstellung.
Beispiele: Tonsignale, Bildsignale, auch: Text

Quantisierung
Darstellung einer Größe in einem System, das nur diskrete Werte darstellen kann (vgl. Quantenpysik).

Abtastung
Aufzeichnung von Messwerten zu diskreten, meist äqui-distanten, Zeitpunkten.

Dimensionen eines analogen Signals

- Räumlich (Bild)
- Zeitlich (Sprache)
- Kombination (Film)

Werte eines analogen Signals

- Wertebereich und Genauigkeit hängen von der Sensorik ab.
- Genauigkeit ist auch durch die Möglichkeiten des Aus- bzw. Ablesens der Werte beschränkt.
- Werte können prinzipiell in beliebiger Auf-lösung erfasst werden.

Wie wird digitalisiert?

- räumliche und/oder zeitliche Abtastung des analogen Signals

- Quantisierung der Messwerte

Abtastung

- Das analoge Signal wird i.a. in äquidistanten Abständen abgetastet.
- Die auf diese Weise gefundene Repräsentation soll eine möglichst vollständige Darstellung des analogen Signals liefern.
- Bei einer Rekonstruktion des analogen Signals soll dies als transparent wahrgenommen werden.

Quantisierung

- Der Wertebereich des Signals wird in Intervalle oder Zellen eingeteilt, die durch einen Wert oder Vektor repräsentiert werden.
- Das Eingangsignal wird durch den Wert oder Vektor quantisiert, der das Intervall bzw. die Zelle repräsentiert, der das Eingangssignal zugeordnet werden kann.

Beispiel: Audiosignale

Glossar

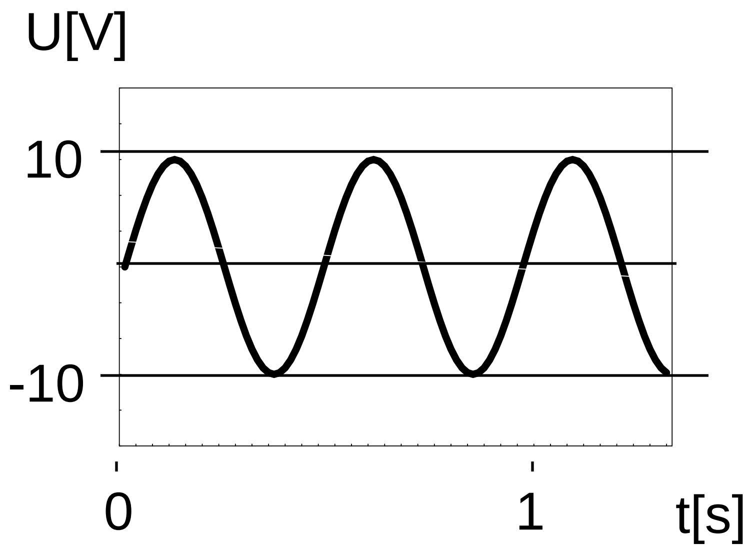
Amplitude
Bezeichnung für die maximale Auslenkung einer Schwingung bzw. einer Welle aus der Mittellage.

Frequenz
Bezeichnung für die Anzahl von Ereignissen innerhalb eines bestimmten Zeitraums (Hertz (Hz): Ereignisse/Sekunde).

Tiefpass
Filterung eines Signals, so dass Signalanteile mit Frequenzen unterhalb der Grenzfrequenz nicht verändert werden und Signalanteile mit Frequenzen oberhalb der Grenzfrequenz stark abgeschwächt werden.

Frequenz, Phase, Amplitude

Eine cos-Schwingung (oder sin-Schwingung) x(t) lässt sich durch drei Parameter vollständig beschreiben: Frequenz, Phase und Amplitude

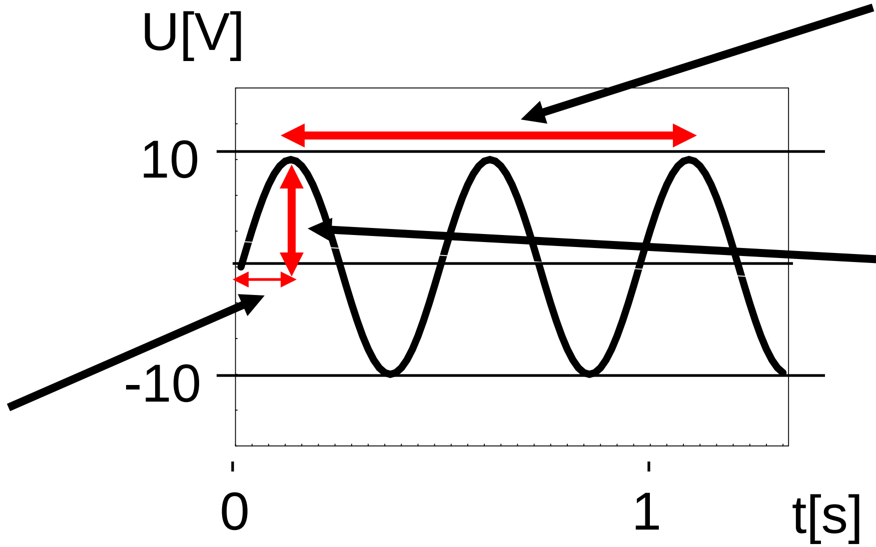


$$x(t) = A\cos(2\pi f t + \phi)$$

Frequenz, Phase, Amplitude

Eine cos-Schwingung (oder sin-Schwingung) x(t) lässt sich durch drei Parameter vollständig beschreiben: Frequenz, Phase und Amplitude

Frequenz f: 2 volle Schwingungen pro Sekunde, f = 2 Hz

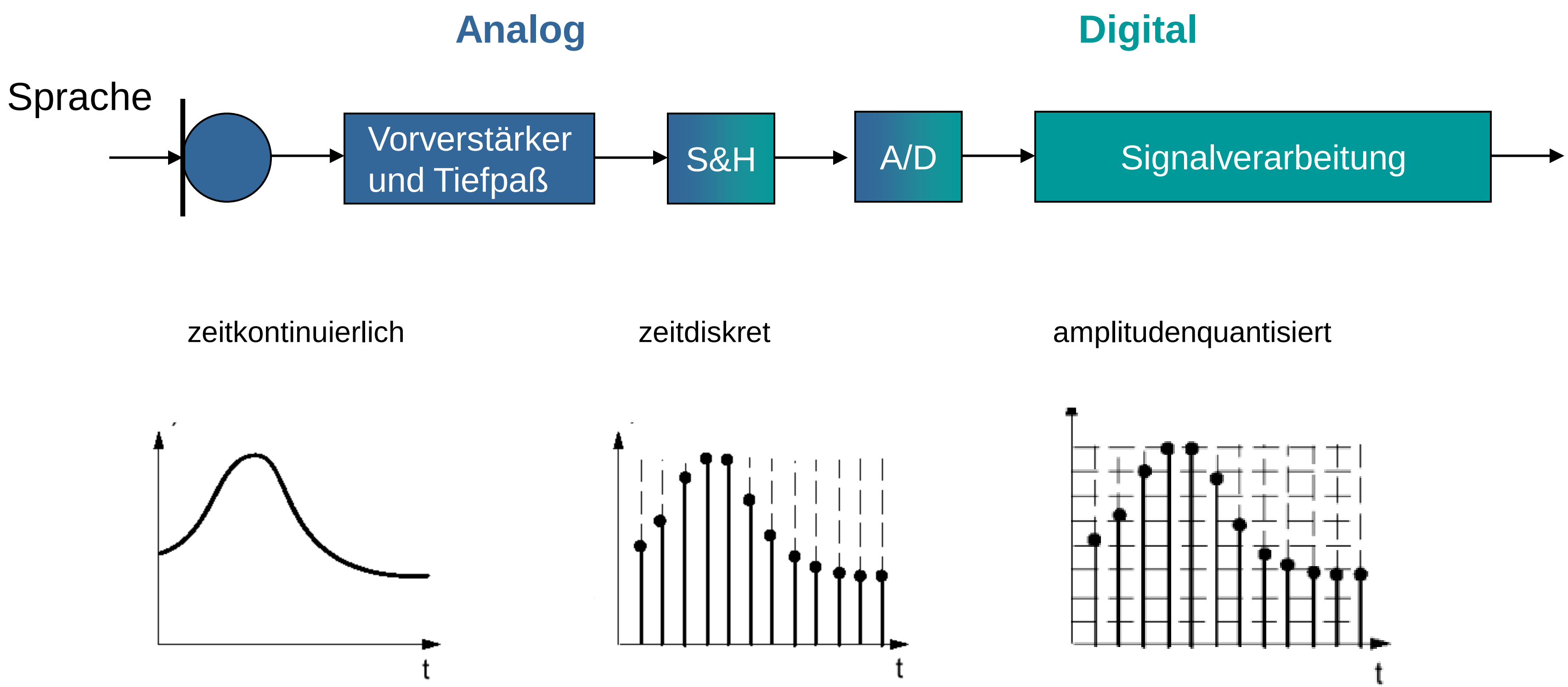


Phase ϕ: Verschiebung gegen null, ϕ = -1,6

Amplitude A: maximale Auslenkung, A = 10 V

$$x(t) = A\cos(2\pi f t + \phi)$$

Digitalisierung eines analogen Signals

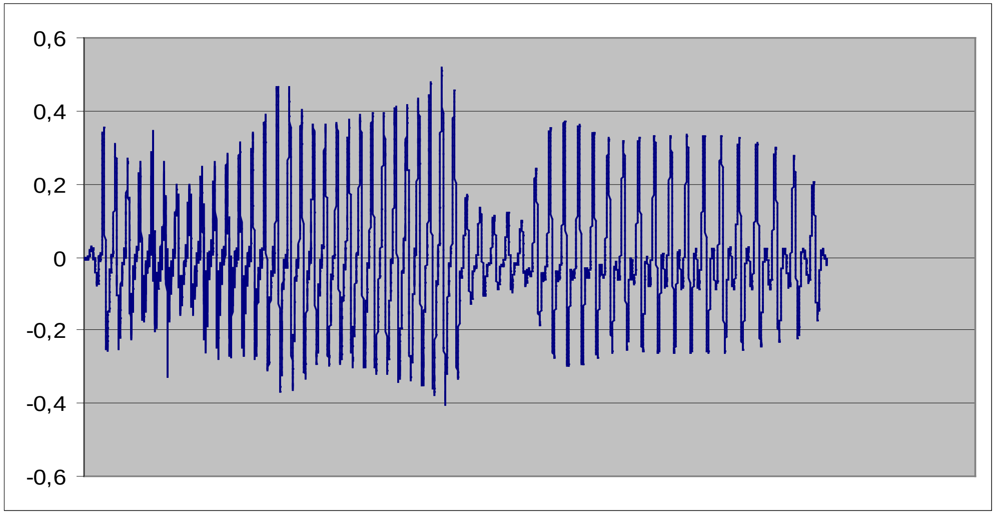


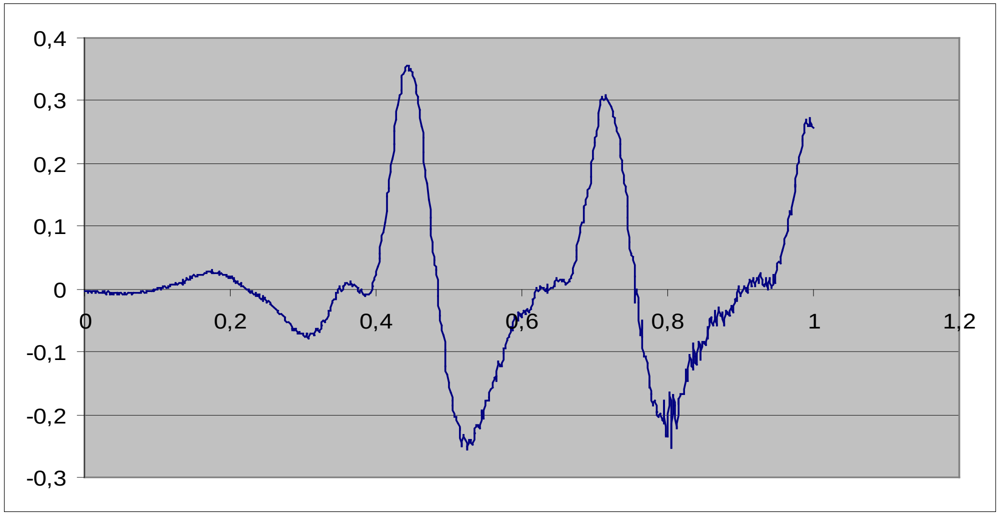
Analog-Digital (AD) - Wandlung

- analoges Mikrofonsignal
- Verstärker
- Tiefpassfilter
- Abtastung (Zeitdiskretisierung)
- Quantisierung der Amplituden

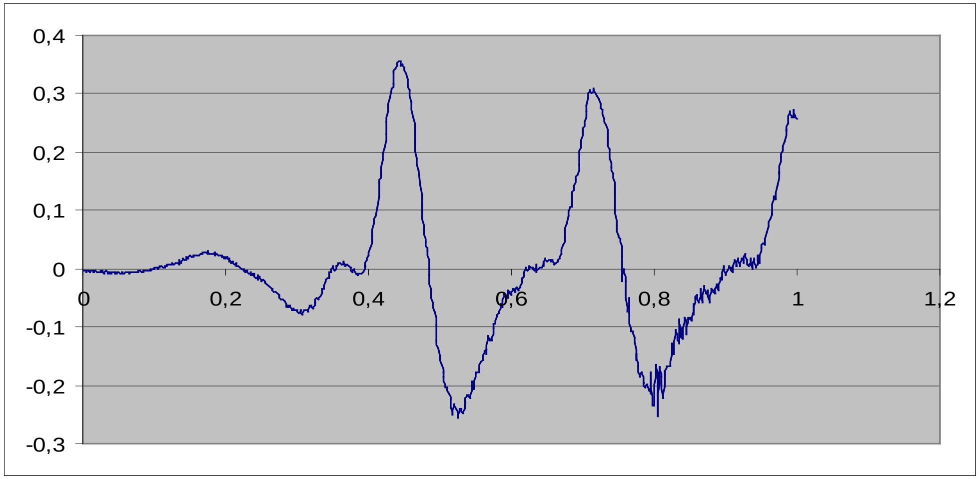
- Speicherung der digitalen Daten in einer Datei

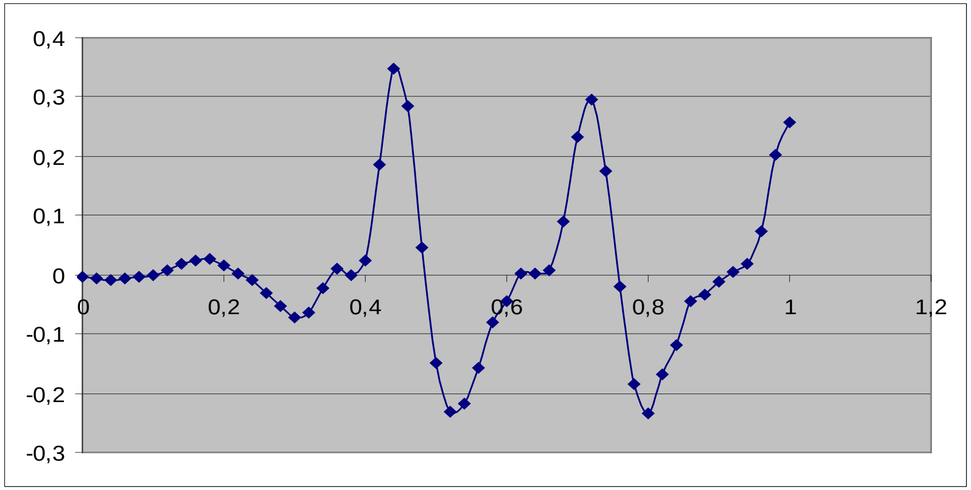
Analoges Signal



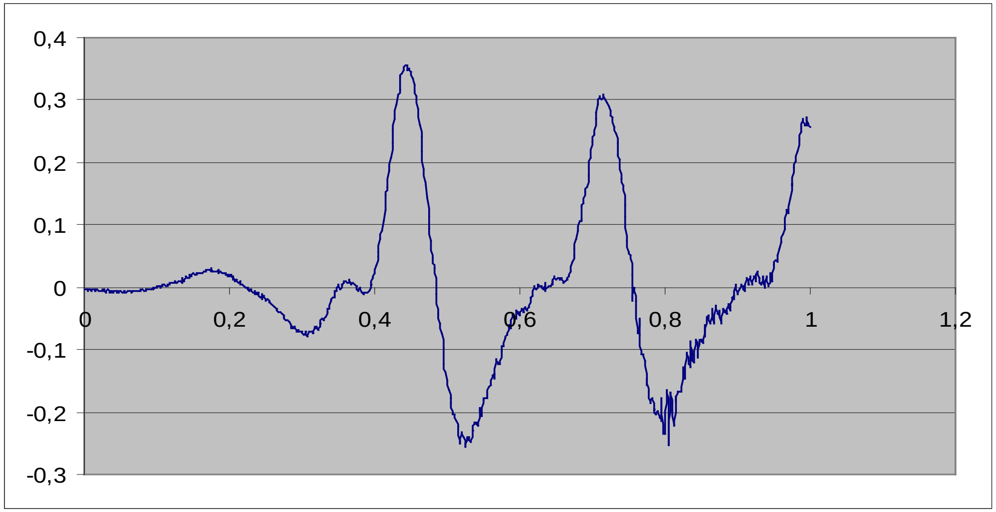


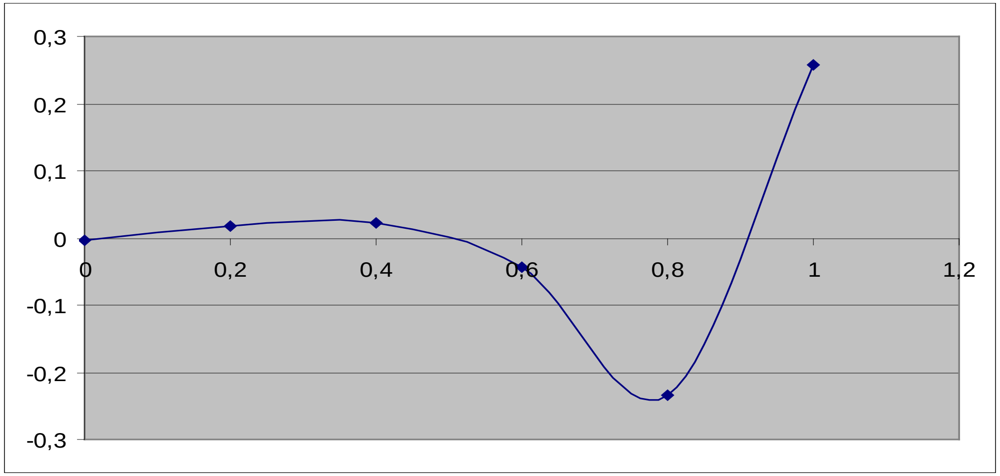
Abtastung des Signals





Abtastung des Signals





Abtasttheorem

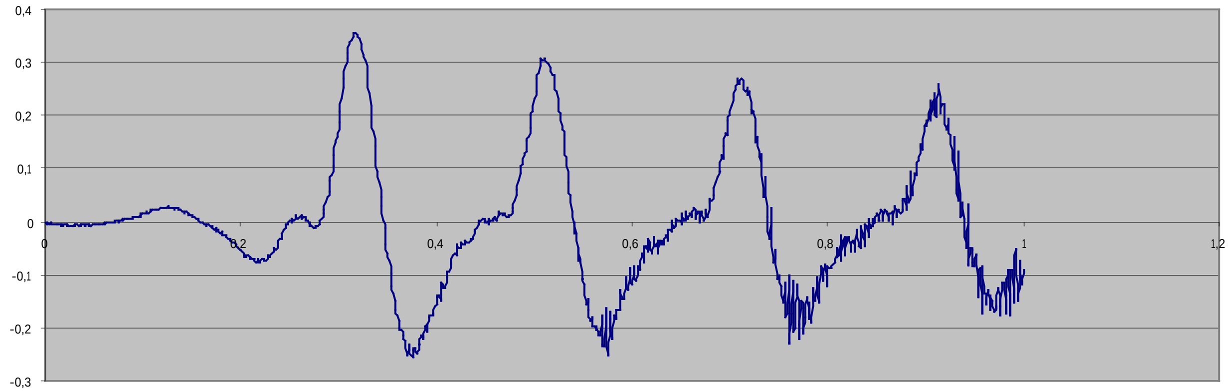
- Die Abtastfrequenz muss mindestens doppelt so hoch sein wie die maximale Frequenz, die im abgetasteten Signal enthalten ist (Shannon).
- Zur Gewährleistung dieser Bedingung muss das Eingangssignal entsprechend gefiltert werden (Tiefpass).

Typische Abtastraten

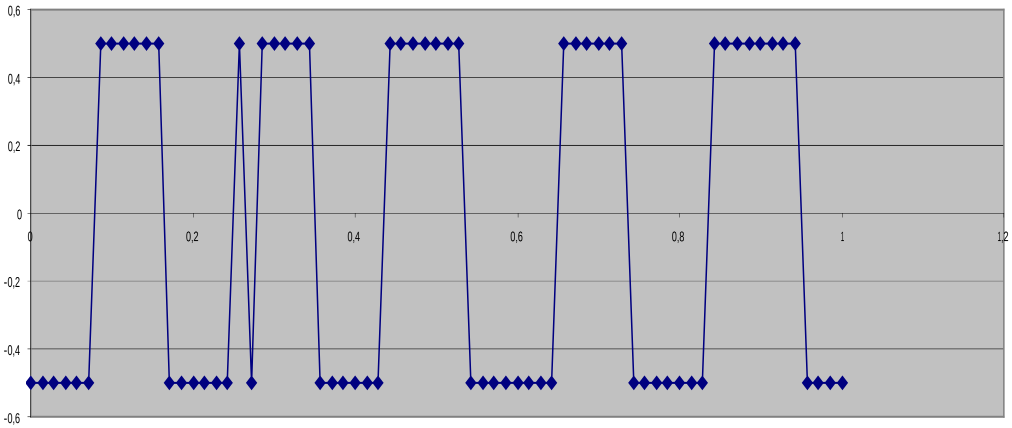
- DAT 48 kHz
- CD 44.1 kHz
- Breitbandsprache 16 kHz
- Diktiersysteme 11.025 kHz
- ISDN 8 kHz

Amplitudenquantisierung

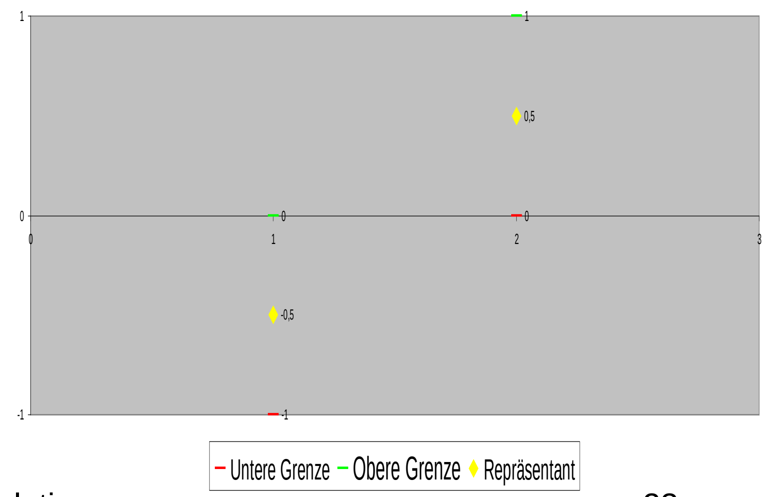
Original



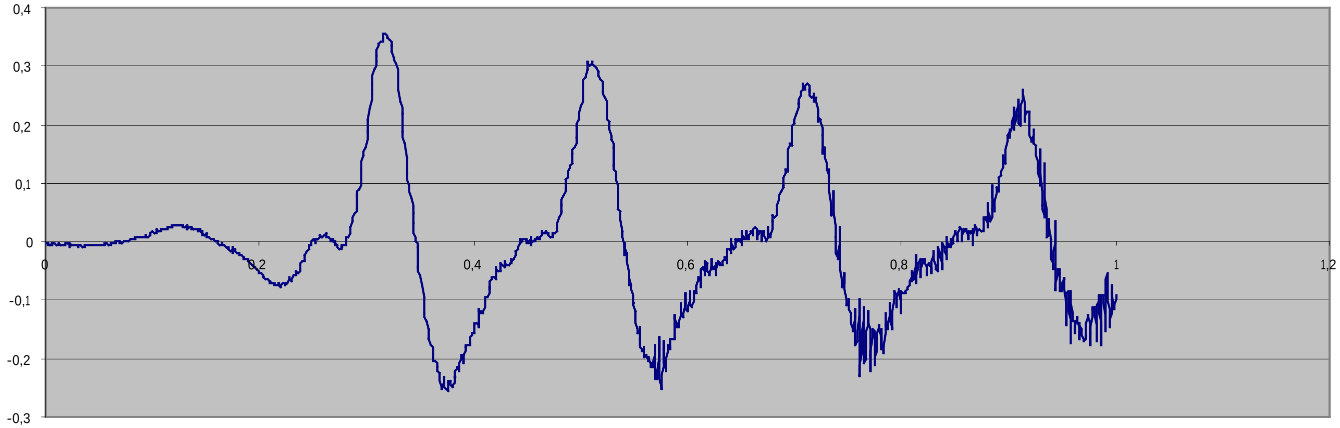
Quantisiert (1 bit)



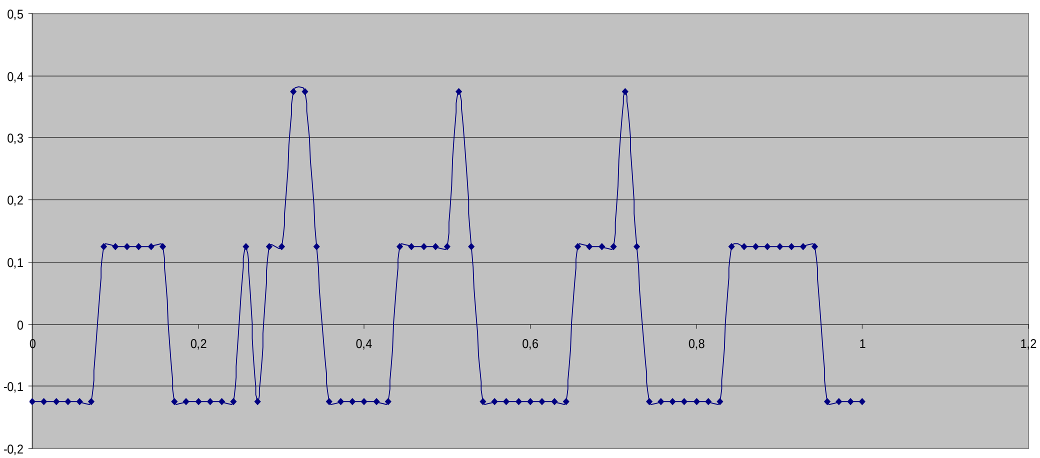
Intervalle



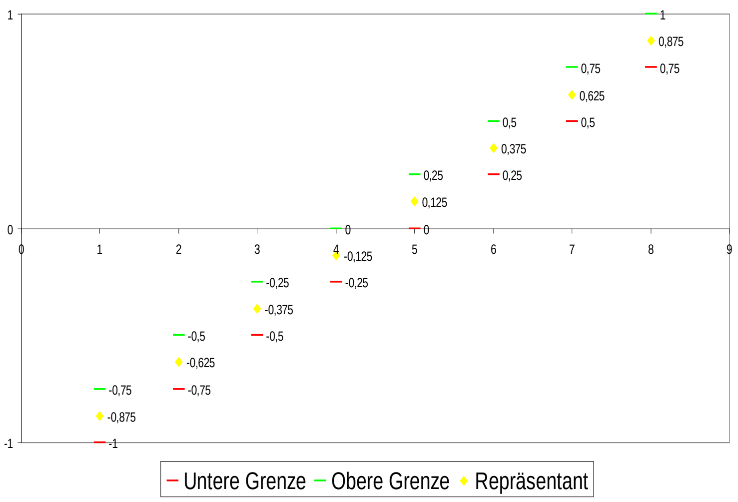
Original



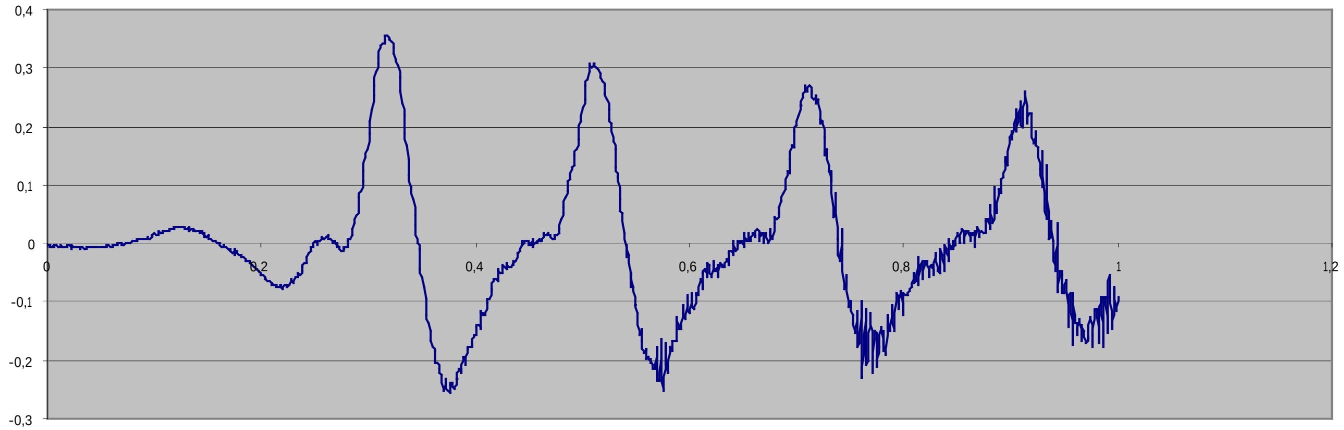
Quantisiert (3 bit)



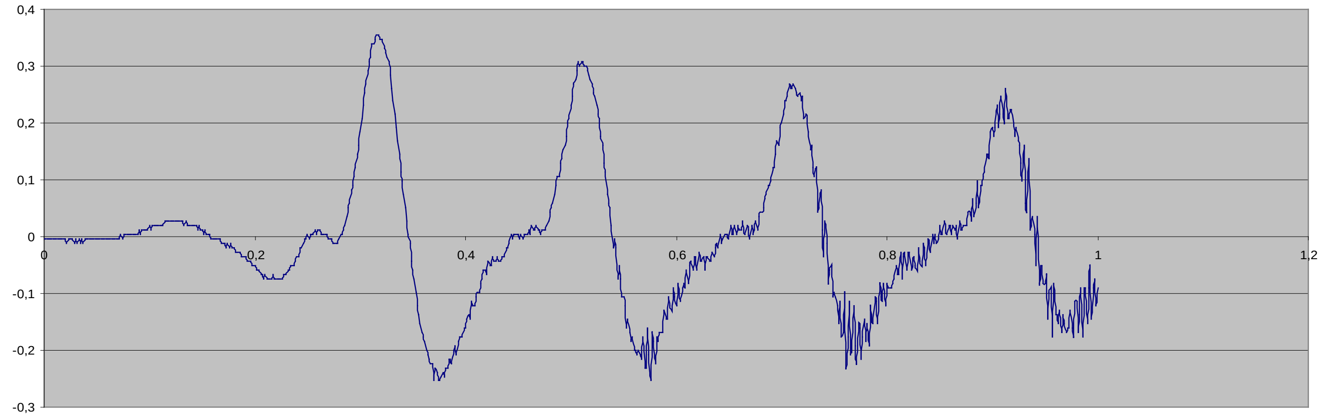
Intervalle (3 bit)



Original



Quantisiert (8 bit)



Optimale Quantisierung

- Die Intervalle bzw. Zellen werden so gewählt, dass bei gegebener Anzahl der Quantisierungsstufen bzgl. einer gegebenen Datenmenge der Quantisierungsfehler minimal wird.
- In diesen Prozess geht insbesondere die Verteilungsdichte der gegebenen Datenmenge ein.

Ziel

- Das Ziel ist die Realisierung einer digitalen Repräsentation eines analogen Signals, die für den Betrachter oder bzgl. des Analysesystems denselben Informationsgehalt hat wie das analoge Signal.
- Beispiel: MPEG

Optimierte Quantisierung für Sprache

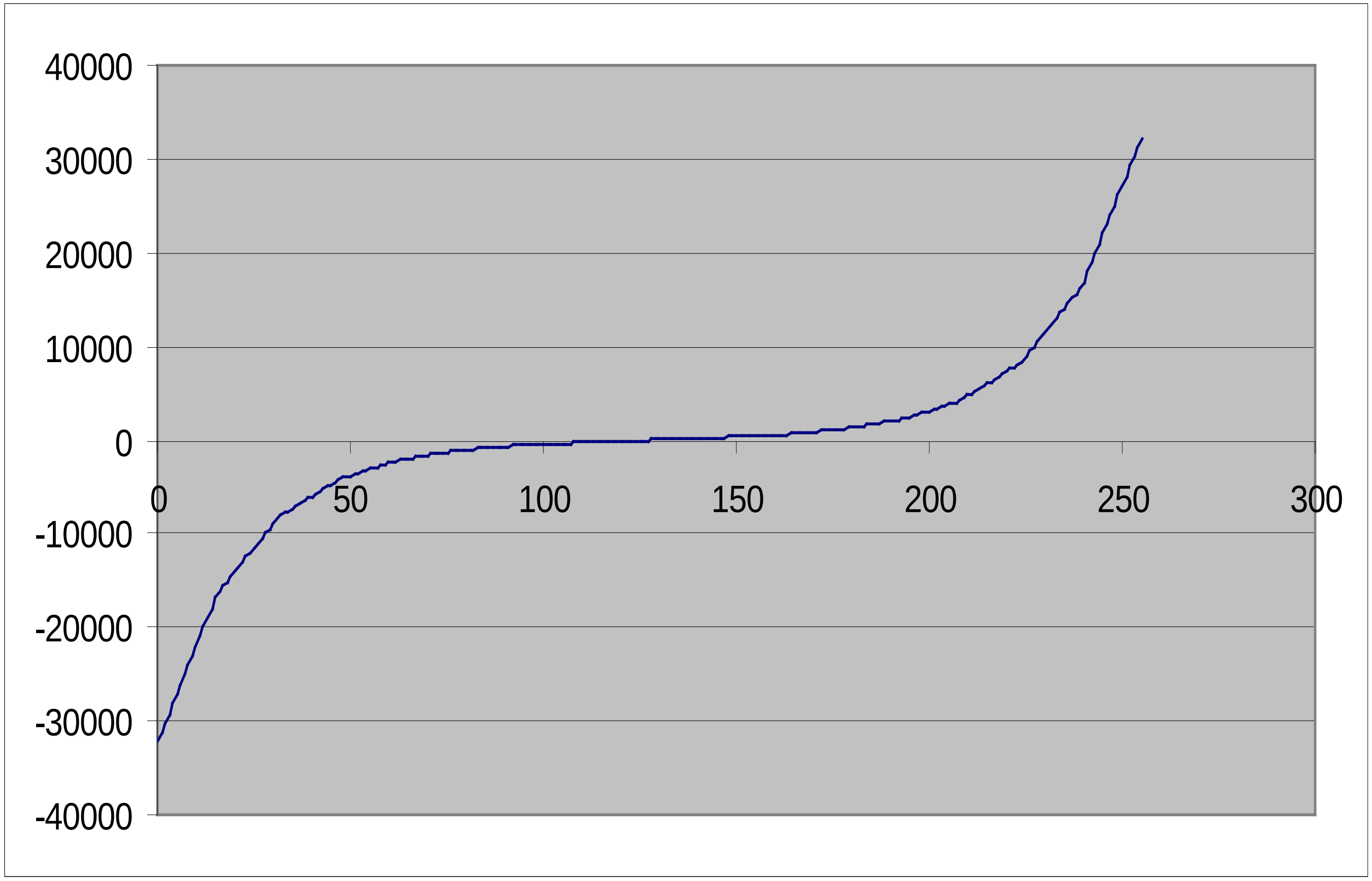
- Das menschliche Gehör nimmt Unterschiede bei leisen Signalen stärker wahr als bei lauten.
- Bei gleichförmiger Quantisierung (konstante Größe der Intervalle) werden Quantisierungsfehler daher bei leisen Signalen deutlicher wahrgenommen als bei lauten.
- Für eine optimierte Quantisierung werden die Intervallgrößen mit Hilfe von Segmenten unterschiedlicher Steigung gewählt.
- Mit 8 Bit gelingt eine Quantisierung, die der Qualität einer gleichförmigen Quantisierung mit 12 Bit entspricht.
- Amerika, Japan: mu-law (15 Segmente); Europa: a-law (13 Segmente)

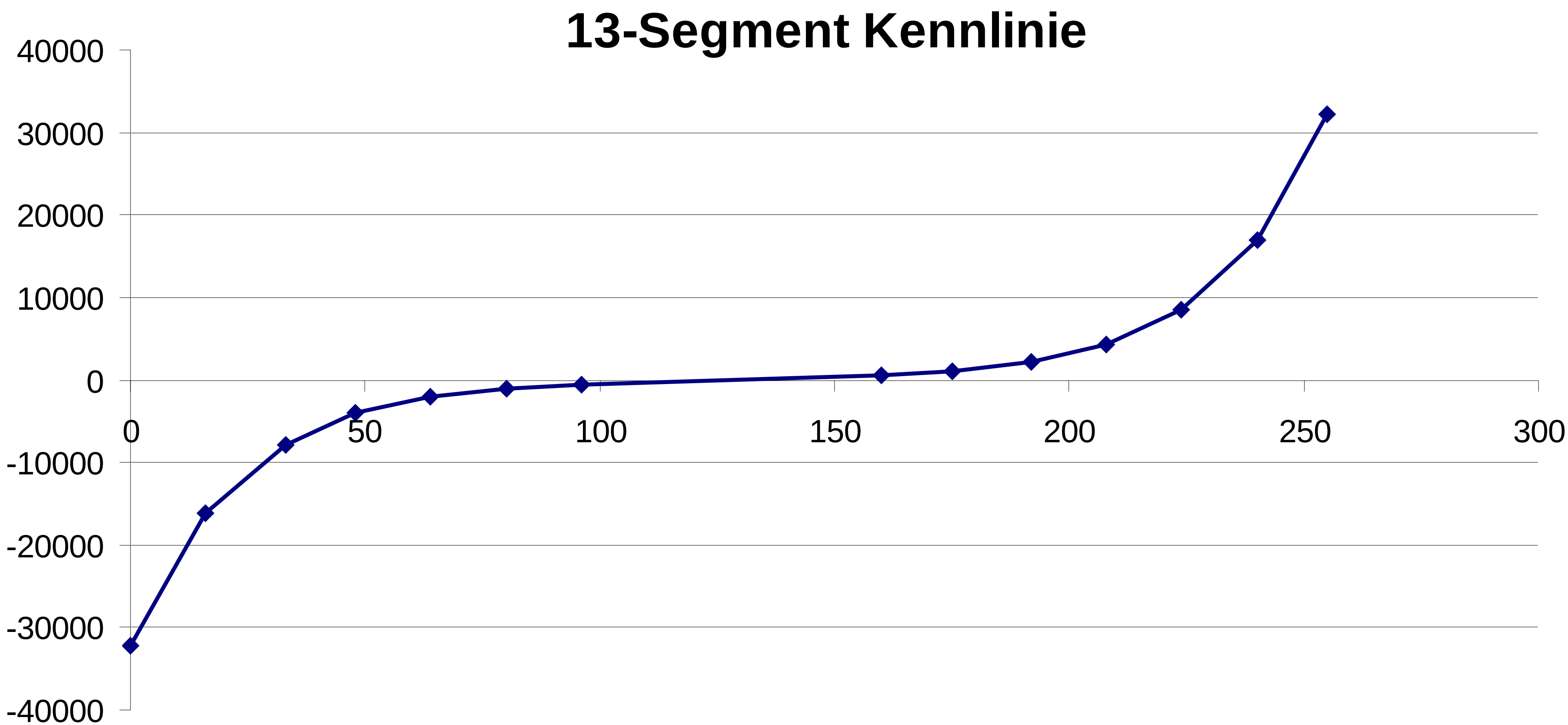
Tabelle a-law

a-law	short int
0	-32256
1	-31232
2	-30208
...	...
125	-40
126	-24
127	-8

a-law	short int
128	8
129	24
130	40
...	...
253	30208
254	31232
255	32256

a-law Codierung (8 bit)




13-Segment Kennlinie
40000
30000
20000
10000
0
-10000
-20000
-30000
-40000
0
50
100
150
200
250
300

Beschränkungen der Digitalisierung

- **Abtastung**

  bedingt eine Beschränkung der maximalen Frequenz des analogen Signals

- **Quantisierung**

  Fehler durch die Quantisierung der Amplitude kann durch den Signal-Rausch-Abstand (SNR) bewertet werden:

$$SNR = 10 * \log \frac{\sum x[n]^2}{\sum (\tilde{x}[n] - x[n])^2}$$

Dateiformate

- pcm, raw: ohne Header
- wav: Microsoft
- au: Sun Microsystems
- mp3: MPEG Layer 3

- beachte: Byte-Order

Header einer .wav-Datei (Struktur)

Length	**description**	**String**
4 Byte	< Magic Number RIFF >	“RIFF”
4 Byte	< size of file >	
4 Byte	< Magic Number WAVE >	“WAVE”

chunks

4 Byte	< ID of chunk >	f.e. “fmt “
4 Byte	< size of chunk >	
……… data		
4 Byte	< ID of chunk >	f.e. “fact”
4 Byte	< size of chunk >	
……… data		

… more chunks

Header einer .wav-Datei (Details)

Length	**description**	**name**
chunk: header information		
4 Byte	< ID of header >	“fmt “
4 Byte	< size of header segment (Byte) >	
2 Byte	< format tag >	
2 Byte	< number of channels >	
4 Byte	< number of samples per second >	
4 Byte	< number of bytes per second >	
……… more information		
chunk: audio data		
4 Byte	< ID of data >	“data”
4 Byte	< size of data segment (Byte) >	
……… audio data (f.e. 2 Byte/sample)		

Fourier-Transformation

Fourier-Transformation

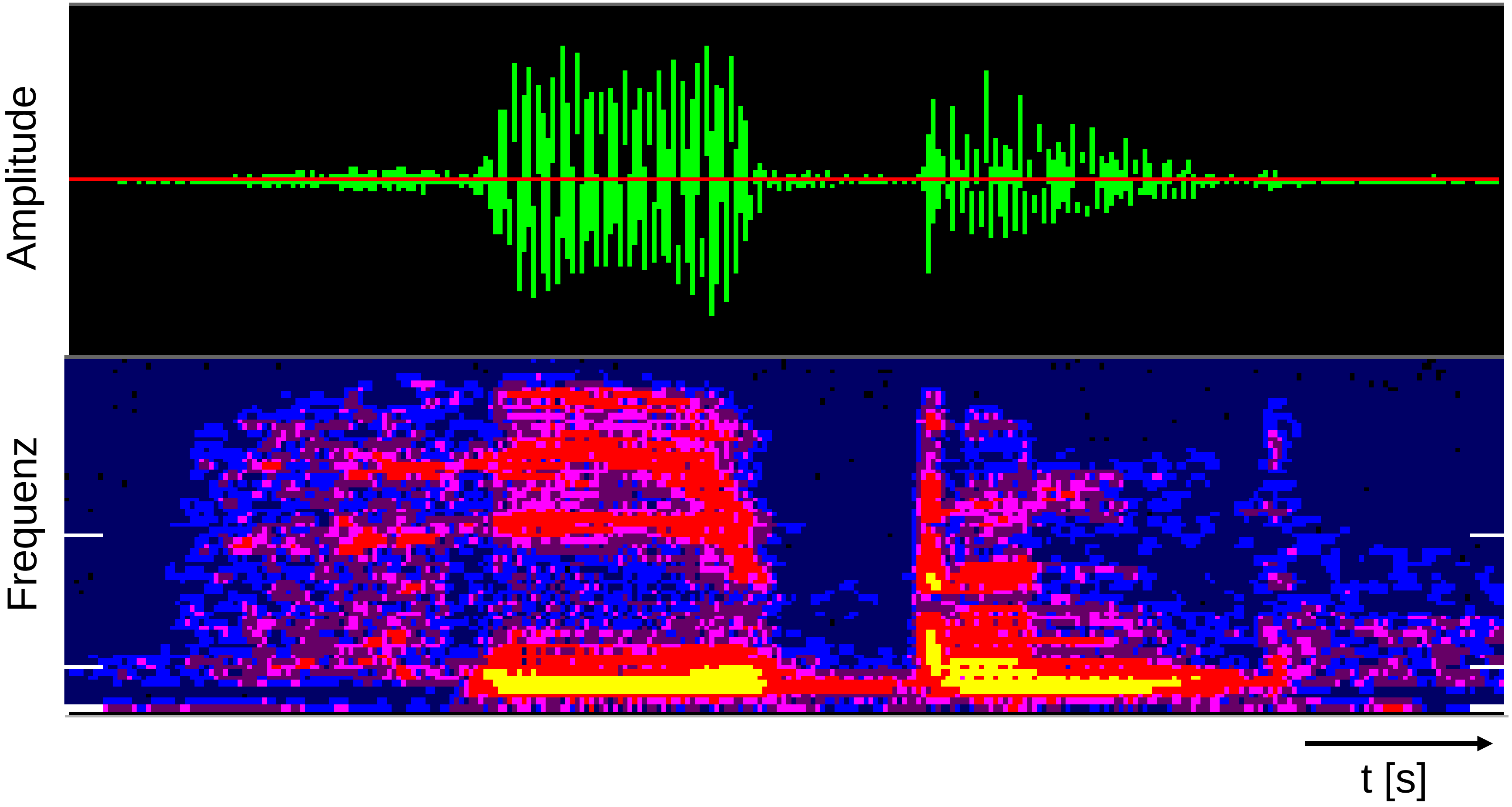
- Jean Baptiste Joseph Fourier (1768-1830)
- Französischer Mathematiker und Physiker
- Fourier (1807): „Ein beliebiges kontinuierliches und periodisches Signal kann durch adäquat gewählte Sinus- und Cosinus-Funktionen vollständig repräsentiert werden.“

Fourier-Transformation

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int e^{-i\omega t} f(t) dt$$

$$f(t) = \frac{1}{\sqrt{2\pi}} \int e^{i\omega t} F(\omega) d\omega$$

Fourier-Transformation



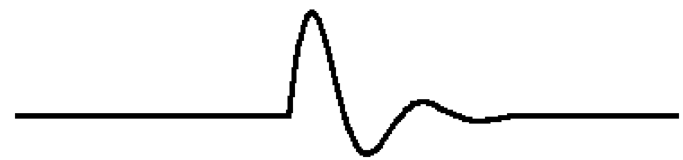
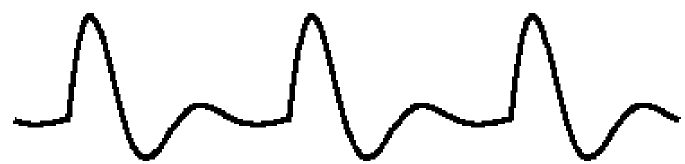
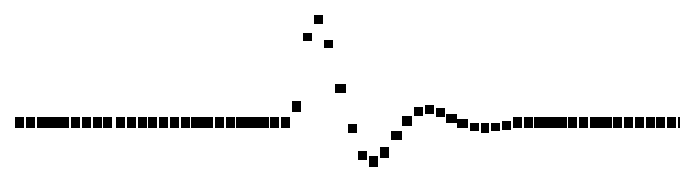
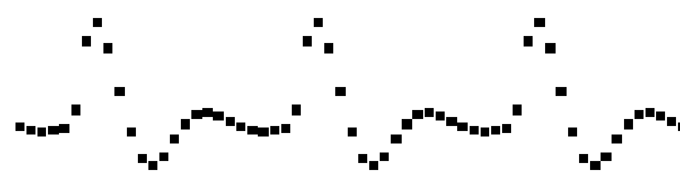
Interaktive Demonstration

[Applet zur Diskreten Fourier-Transformation]

Typisierung

- Fourier-Transformation
  (aperiodisch, kontinuierlich)
- Fourier-Reihe
  (periodisch, kontinuierlich)
- Zeitdiskrete Fourier-Transformation
  (aperiodisch, diskret)
- Diskrete Fourier-Transformation
  (periodisch, diskret)

Beispiele

Type of Transform	Example Signal
Fourier Transform *signals that are continious and aperiodic*	
Fourier Series *signals that are continious and periodic*	
Discrete Time Fourier Transform *signals that are discrete and aperiodic*	
Discrete Fourier Transform *signals that are discrete and periodic*	

Zeitdiskrete Fourier-Transformation

$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n] e^{-i\omega n}$$

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{i\omega n} d\omega$$

Diskrete Fourier-Transformation (DFT)

Für x[n] periodisch, d.h. x[n] = x[n+aN] gilt:

$$X[\nu] = \sum_{n=0}^{N-1} x[n] e^{-2\pi i \nu n / N}$$

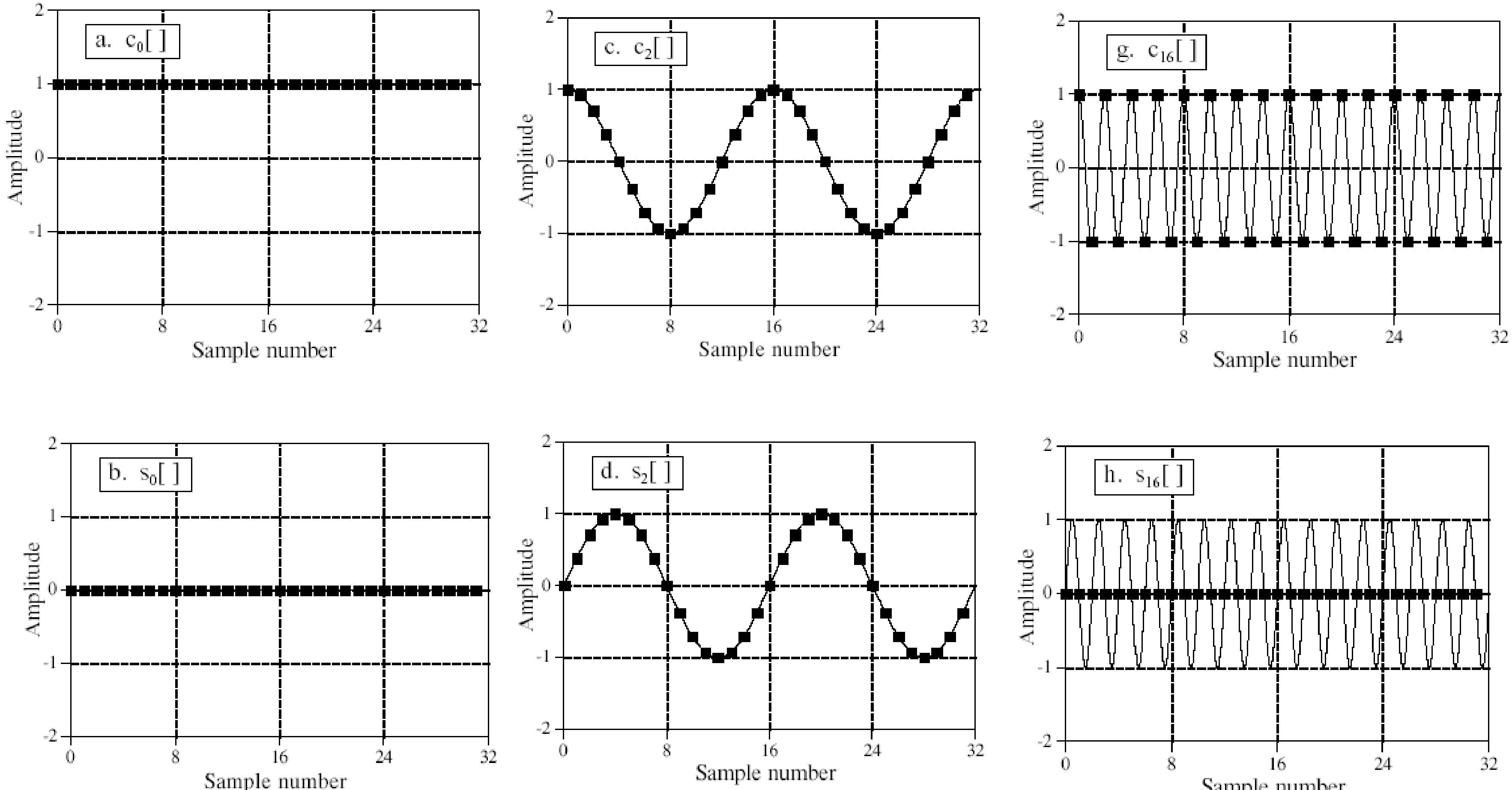
$$x[n] = \frac{1}{N} \sum_{\nu=0}^{N-1} X[\nu] e^{2\pi i \nu n / N}$$

Basisfunktionen der DFT

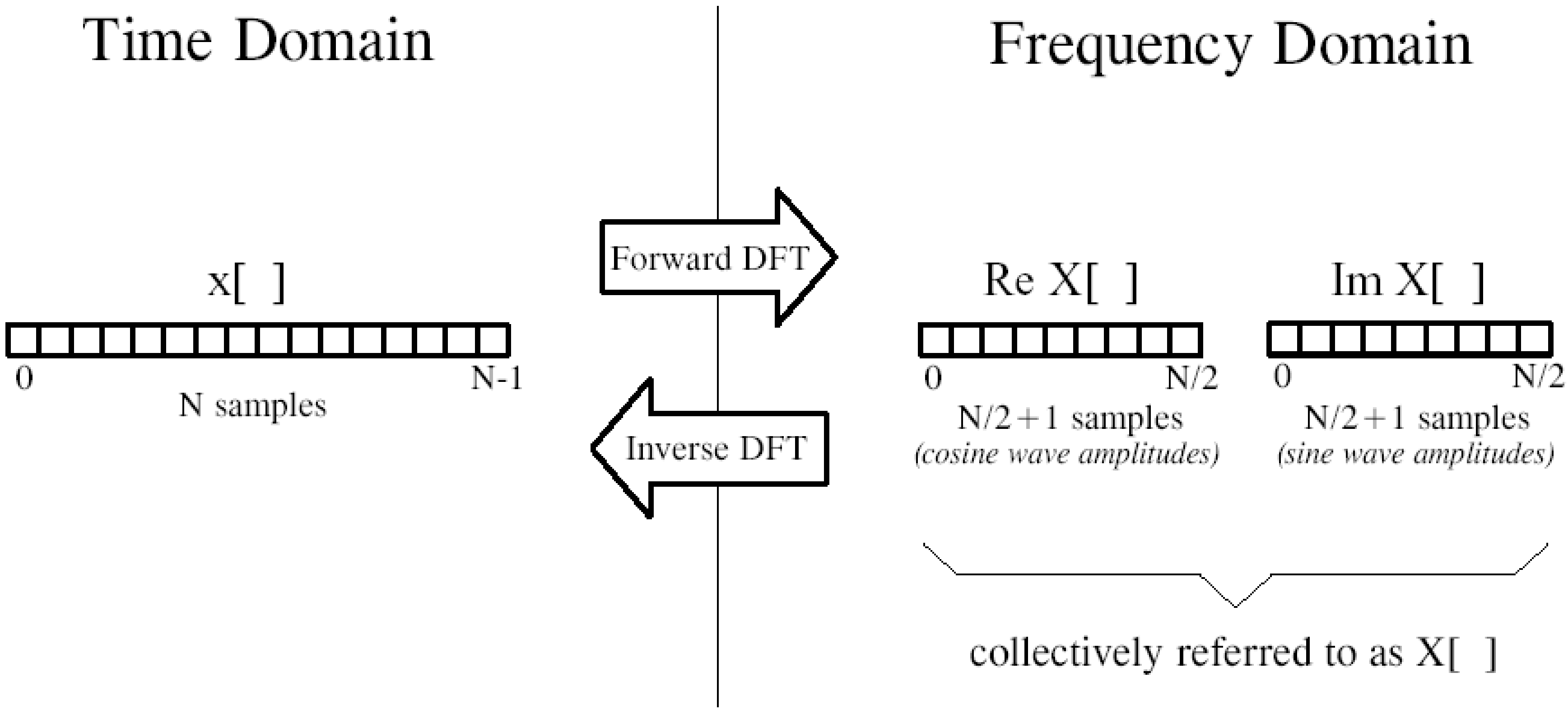
$$C_\nu[n] = \cos(2\pi n\nu / N)$$

$$S_\nu[n] = \sin(2\pi n\nu / N)$$

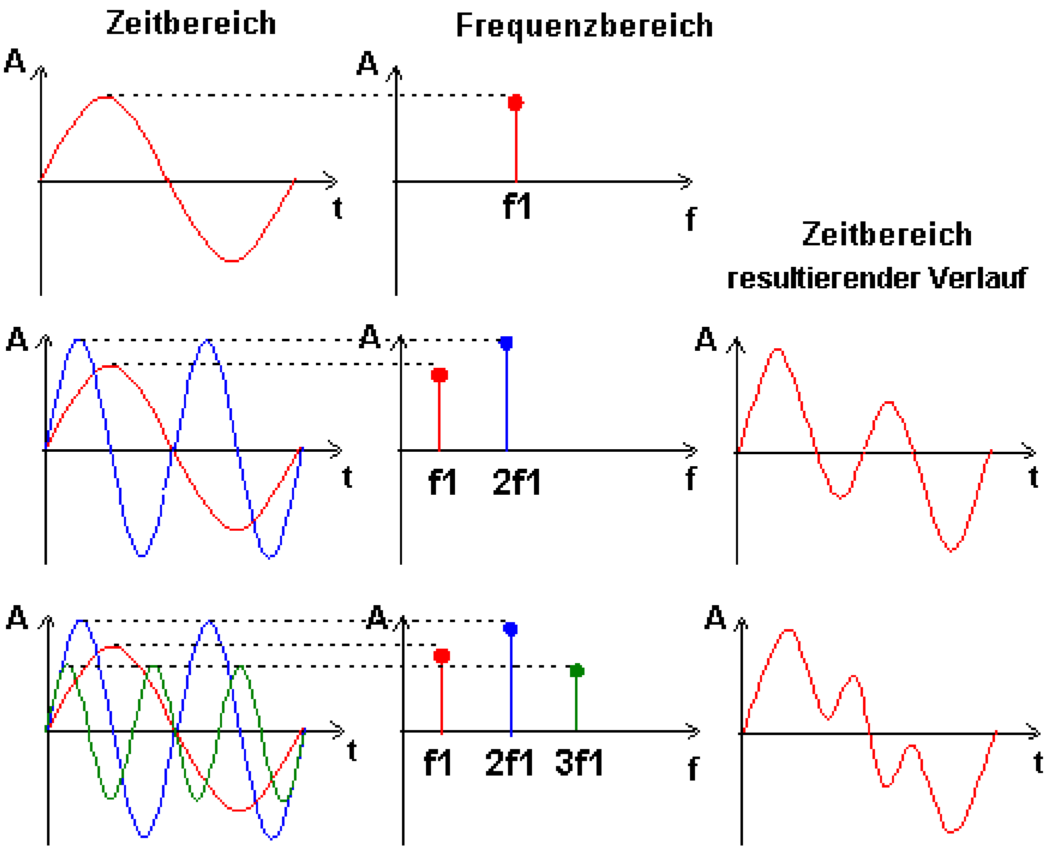
Ausgewählte Basisfunktionen



Schema der DFT



Beispiele



Zeitbereich
Frequenzbereich
A
t
f1
f
Zeitbereich
resultierender Verlauf
A
t
f1
2f1
f
A
t
f1
2f1
3f1
f

Berechnung der DFT über Korrelation

$$\operatorname{Re} X[\nu] = \sum_{n=0}^{N-1} x[n] \cos(2\pi n\nu / N)$$

$$\operatorname{Im} X[\nu] = -\sum_{n=0}^{N-1} x[n] \sin(2\pi n\nu / N)$$

$$\text{für } 0 \le \nu \le N/2$$

Berechnung

```
for(i=0,ReX[i]=0;i <= N/2;i++) // Frequenz
  for(n=0;n < N;n++) // Abtastwerte
    ReX[i] += x[n] * cos(2πni/N) // Korrelation
```

Inverse DFT

$$x[n] = \sum_{\nu=0}^{N/2} \mathrm{Re}\,\bar{X}[\nu] \cos(2\pi n\nu / N) + \sum_{\nu=0}^{N/2} \mathrm{Im}\,\bar{X}[\nu] \sin(2\pi n\nu / N)$$

wobei

$$\mathrm{Re}\,\bar{X}[\nu] = \frac{\mathrm{Re}\,X[\nu]}{N/2} \qquad \mathrm{Im}\,\bar{X}[\nu] = -\frac{\mathrm{Im}\,X[\nu]}{N/2}$$

aber

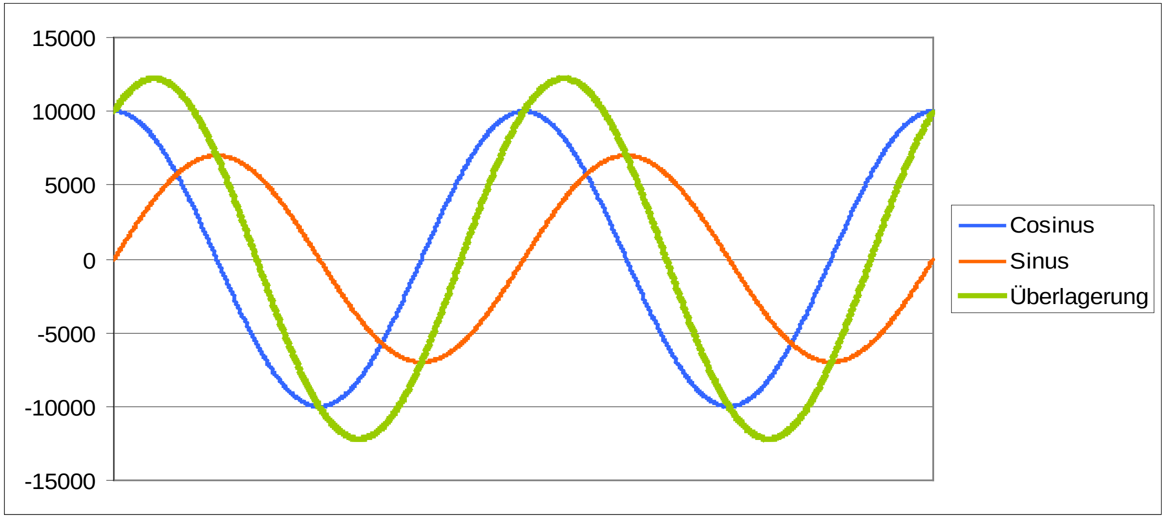
$$\mathrm{Re}\,\bar{X}[0] = \frac{\mathrm{Re}\,X[0]}{N} \qquad \mathrm{Re}\,\bar{X}[N/2] = \frac{\mathrm{Re}\,X[N/2]}{N}$$

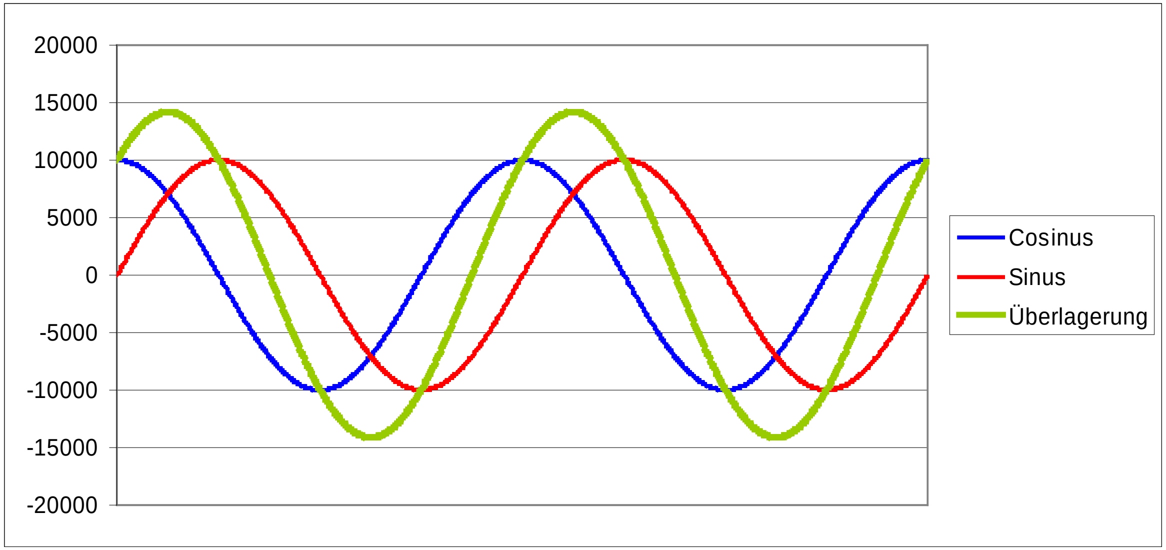
$$\text{für } 0 \leq n < N$$

Interaktive Demonstration

Applet zur Diskreten Fourier-Transformation

Überlagerung von *cos()* und *sin()* für gleichfrequente Schwingungen





Resultierende Amplitude für gleichfrequente Schwingungen

$$
\begin{aligned}
f(t) &= A_1 \sin(\omega t + \varphi_1) + A_2 \sin(\omega t + \varphi_2) \\
&= A \sin(\omega t + \varphi)
\end{aligned}
$$

$$A = \sqrt{A_1^2 + A_2^2 + 2 A_1 A_2 \cos(\varphi_2 - \varphi_1)}$$

für $\varphi_1 = 0,\ \varphi_2 = \pi / 2$ gilt:

$$A = \sqrt{A_1^2 + A_2^2}$$

Resultierende Phase für gleichfrequente Schwingungen

$$f(t) = A_1 \sin(\omega t + \varphi_1) + A_2 \sin(\omega t + \varphi_2)$$
$$= A \sin(\omega t + \varphi)$$

$$\tan(\varphi) = \frac{A_1 \sin(\varphi_1) + A_2 \sin(\varphi_2)}{A_1 \cos(\varphi_1) + A_2 \cos(\varphi_2)}$$

$$\text{für } \varphi_1 = 0,\ \varphi_2 = \pi / 2 \text{ gilt:}$$

$$\tan(\varphi) = \frac{A_2}{A_1} \Rightarrow \varphi = \arctan\left(\frac{A_2}{A_1}\right)$$

Betrags- und Phasenspektrum

- Betragsspektrum: Darstellung der resultierenden Amplitude in Abhängigkeit von der Frequenz.
- Phasenspektrum: Darstellung der resultierenden Phase in Abhängigkeit von der Frequenz.

- Die Darstellung mit Betrags- und Phasenspektrum wird als Polardarstellung oder spektrale Darstellung bezeichnet.
- Die Darstellung mit den Basisfunktionen *cos()* und *sin()* wird als komplexe Darstellung bezeichnet.

Interpretation der Frequenzen

- Das Ergebnis der DFT sind Amplituden der Frequenzen des Fourierspektrums, wobei die einzelnen Frequenzen durch Indizes gekennzeichnet werden.
- Eine Interpretation der Indizes durch sogenannte natürliche Frequenzen in Hz ist durch die Kenntnis der maximalen im Zeitbereich auftretenden Frequenz (diese wird durch das Abtasttheorem beschränkt) möglich.

Komplexität der Berechnung der DFT über Korrelation

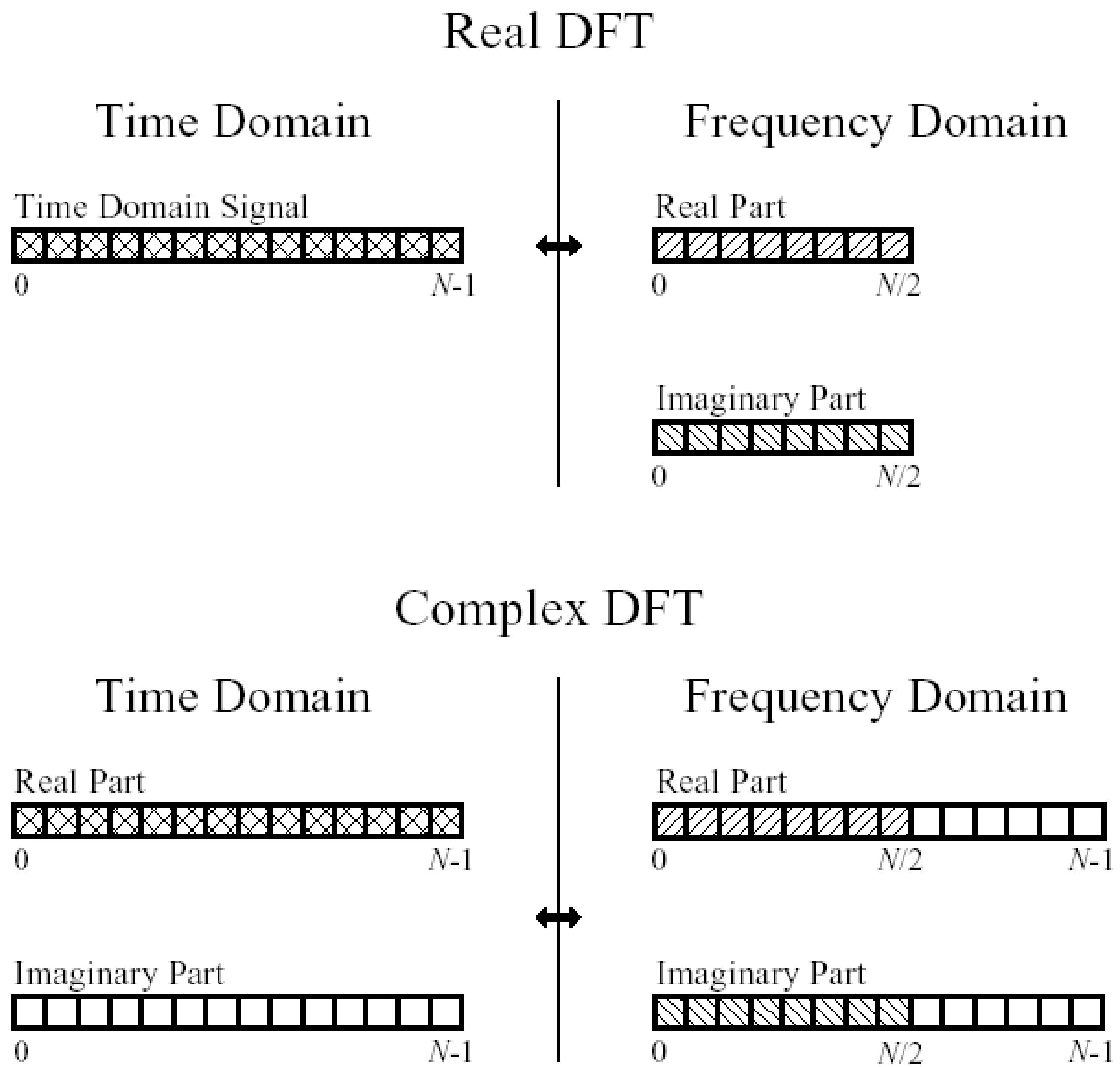
- Welche Komplexität hat die Berechnung?

- Für jeden Wert im Frequenzbereich müssen N Multiplikationen und Additionen ausgeführt werden.
- Komplexität: $\mathcal{O}(N^2)$

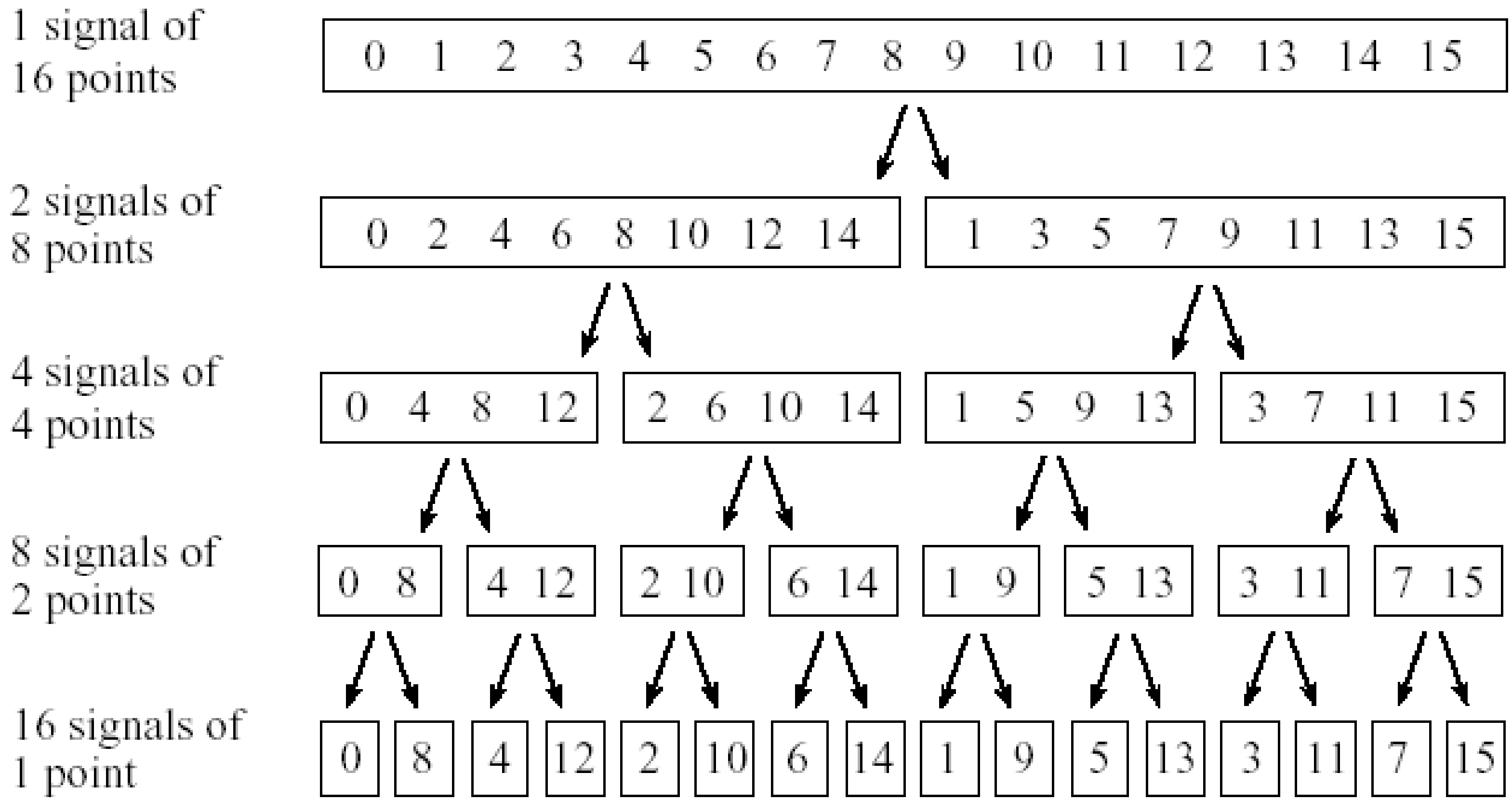
Fast Fourier-Transformation (FFT)

- Die Berechnung einer DFT hat die Komplexität N^2.
- Insbesondere für große N führt dies zu hohem Aufwand für die Berechnung der DFT.
- Durch eine geschickte Zerlegung zur Berechnung der Koeffizienten kann die Komplexität mit der FFT auf ld(N)*N reduziert werden.

Komplexe DFT



Zerlegung des Zeitsignals

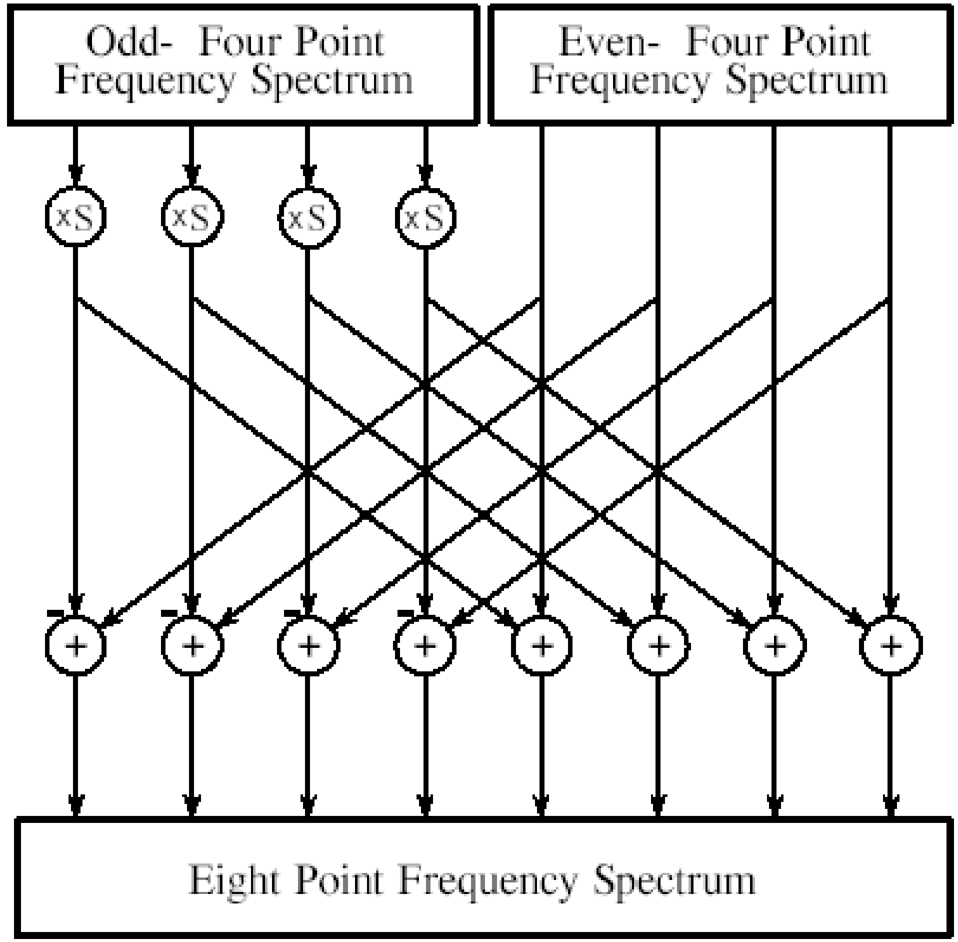


Bit Reversal

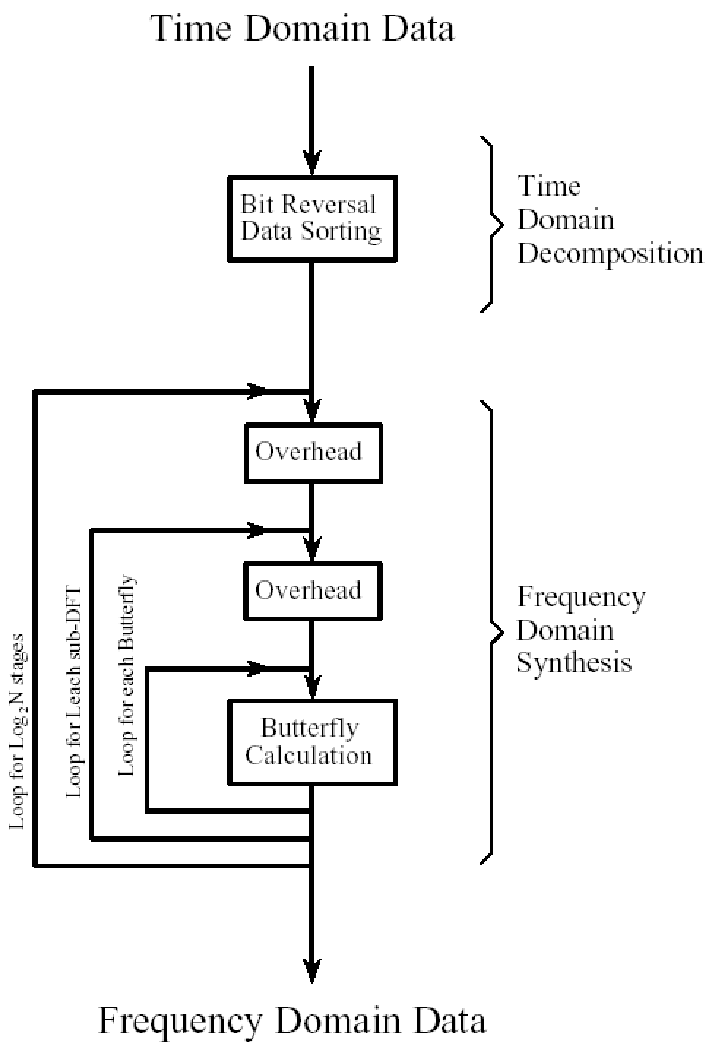
Decimal	*Binary*
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001
10	1010
11	1011
12	1100
13	1101
14	1110
15	1111

Decimal	*Binary*
0	0000
8	1000
4	0100
12	1100
2	0010
10	1010
6	0100
14	1110
1	0001
9	1001
5	0101
13	1101
3	0011
11	1011
7	0111
15	1111

Synthese im Frequenzbereich

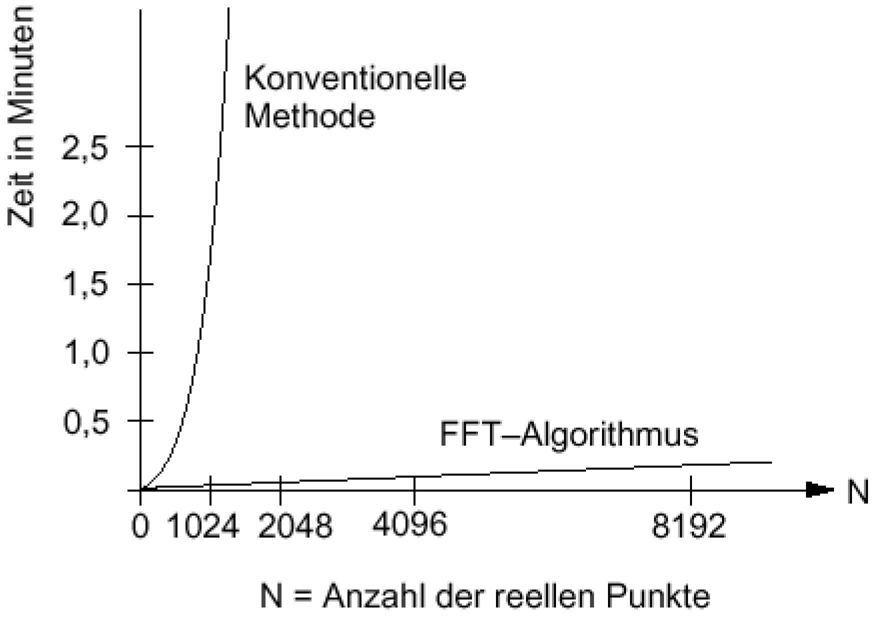


Schema der FFT



Fazit FFT

Schneller Algorithmus, der die Rechenzeit erheblich verkürzt. Er zerlegt eine Zahlenfolge der Länge N in periodische Teilfolgen, vorzugsweise auf der Basis 2.



Beispiel für $N = 2^{10} = 1024$

DFT: 1048567 Operationen $\sim N^2$

FFT: 10240 Operationen $\sim N \, ld(N)$

Value at Risk (VaR)

Modellbildung

- Modelle sind Hilfsmittel für den Umgang mit Realität.
- Die Realität wird nicht direkt abgebildet, sondern abstrahiert.
    - Nicht zentrale Details werden ignoriert.
    - Komplexität des Modells wird minimiert.

Glossar

Portfolio: In der Ökonomie bezeichnet der Begriff Portfolio ein Bündel von Investitionen, das im Besitz einer Institution oder eines Individuums ist.

Quantil: Quantile sind Punkte einer sortierten Messreihe. Der Wert des p-Quantils einer Verteilung gibt an, welcher Wert die unteren p-% der Datenwerte von den oberen (1 - p)-% trennt. Wenn man etwa das 5%-Quantil einer Verteilung angibt, so besagt dies, dass 5% aller Datenwerte kleiner sind als der betreffende Datenwert und 95% größer oder gleich diesem Wert.

Value **a**t **R**isk (**VaR**)

VaR bezeichnet ein Risikomaß, das den geschätzten, maximalen (Markt-wert-)Verlust eines Portfolios nach einer vorgegebenen Haltedauer angibt, der mit einer bestimmten Wahrscheinlichkeit unter üblichen Marktbeding-ungen nicht überschritten wird.

www.risknet.de/Value-at-Risk.116.0.html

Mit anderen Worten:
Der VaR ist als derjenige Portfolioverlust definiert, der innerhalb eines bestimmten Zeitraumes nur mit vorgegebener Wahrscheinlichkeit noch höher ausfallen kann.

Cremers, Heinz; Value at Risk-Konzepte für Marktrisiken; HfB

Entwicklung des Konzepts

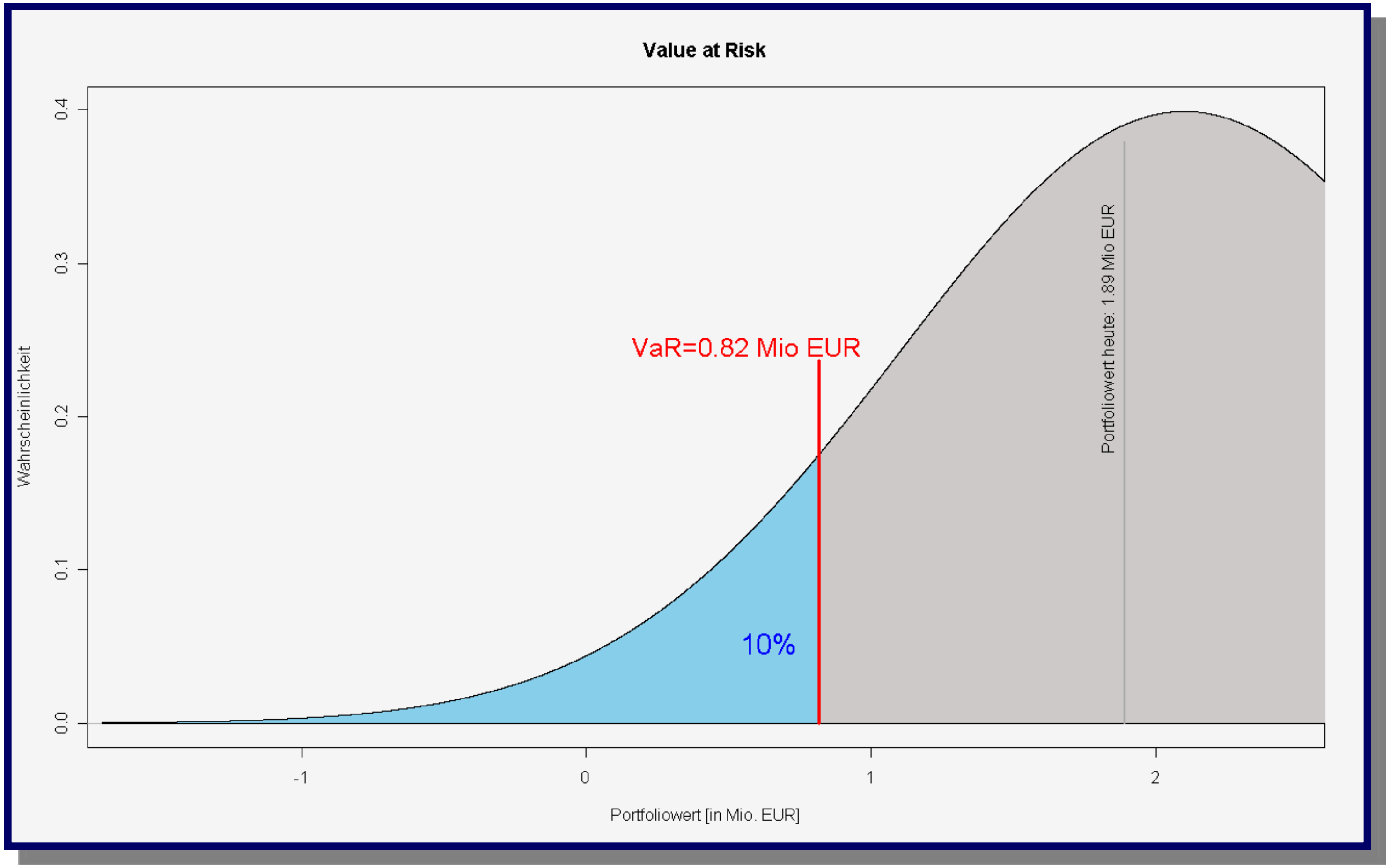
Dennis Weatherstone verlangte während seiner Zeit als Vorsitzender der US-amerikanischen Investmentbank J.P. Morgan täglich um 16.15 Uhr einen einseitigen Risiko-Bericht, in dem das gesamte Portfolio des Handelsbestandes der Bank sowie eine Schätzung der maximal möglichen Verluste in den folgenden 24 Stunden dargestellt waren. Weatherstone war es einfach leid, dass die Marktrisiken der verschiedenen Finanzinstrumente mit unterschiedlichen Methoden gemessen wurden. Sein bisheriger Report enthielt eine Unmenge von Beta-Faktoren und Volatilitäten, gearing factors, Deltas, Gammas und Thetas. Er wollte ein einheitliches Risiko-Maß für alle Finanzinstrumente haben. Dies war die Geburtsstunde des „Report 4.15“ (1994). Dieser Report basiert auf einer Methodik, die als Value at Risk bezeichnet wird.

Quelle: www.risknet.de/Value-at-Risk.116.0.html

Berechnung des VaR

- Sammlung und Analyse historischer Daten.
- Modellierung der Verteilung.
- Simulation von Kursentwicklungen gemäß der modellierten Verteilung (Monte Carlo Simulation).
- Große Zahl an Simulationen.
- Prognose von Kursentwicklungen.
- Berechnung des monetären Risikos für die Entwicklung des Portfolios.

Value at Risk



Glossar

Konfidenznineau gibt an, mit welcher Wahrscheinlichkeit die Schätzung eines statistischen Parameters auf der Basis einer Stichprobe für die Grundgesamtheit zutreffend ist.

Mark-to-Market: marktnahe Bewertung, d.h. Bewertung des Portfolio gemäß aktueller Marktpreise.

BaFin: Bundesanstalt für Finanzdienstleistungsaufsicht be-aufsichtigt und kontrolliert alle Bereiche des Finanzwesens in Deutschland (Gründung: 2002).
… eine Abteilung befasst sich mit Grundsatzfragen quantitativ-mathematischer Modellierung bei Markt-, Kredit-, Liquiditäts- und operationellen Risiken. Ihre Mitarbeiter prüfen die Modelle vor Ort …(link)

VaR im Einsatz (I)

Value-at-Risk ist ein quantitatives Maß für den potenziellen Wertverlust einer Handelsposition aufgrund von Marktschwankungen, der über einen vorgegebenen Zeitraum und mit einem bestimmten Konfidenzniveau nicht überschritten wird.

Unser Value-at-Risk für die Handelsgeschäfte erfolgt auf Basis unseres eigenen internen Value-at-Risk-Modells. Im Oktober 1998 hat das Bundesaufsichtsamt für das Kreditwesen, eines der Vorgängerinstitute der Bundesanstalt für Finanzdienstleistungsaufsicht, unser Value-at-Risk-Modell zur Berechnung des regulatorischen Kapitalbedarfs für das allgemeine und spezifische Marktrisiko genehmigt. Das Modell wurde seitdem regelmäßig überarbeitet und die Genehmigung aufrechterhalten.

Die Informationen sind Teil des Konzernabschlusses der Deutschen Bank zum 31.12.2009 (link).

VaR im Einsatz (II)

Wir berechnen den Value-at-Risk mit einem Konfidenzniveau von 99 %
Und einer Haltedauer von einem Tag. Das bedeutet, wir gehen von einer Wahrscheinlichkeit von 1 zu 100 aus, dass ein Mark-to-Market-Verlust aus unseren Handelspositionen mindestens so hoch sein wird wie der berichtete Value-at-Risk-Wert. Für aufsichtsrechtliche Meldezwecke beträgt die Haltedauer zehn Tage.

Wir verwenden historische Marktdaten, um den Value-at-Risk zu bestimmen. Als Basis dient eine historische Beobachtung über 261 gleich gewichtete Handelstage. Bei der Berechnung wird ein Monte Carlo-Simulationsverfahren angewandt, wobei wir davon ausgehen, dass Änderungen in den Risikofaktoren einer bestimmten Verteilung folgen, zum Beispiel der Normalverteilung oder logarithmischen Normalverteilung. Zur Berechnung des aggregierten Value-at-Risk benutzen wir über denselben 261-Tages-Zeitraum beobachtete Korrelationen zwischen den Risikofaktoren.

Lesen Sie den vollständigen Text auf der Seite der Deutschen Bank bitte sehr aufmerksam und stellen Sie etwaige Fragen bitte im Kontext der Wiederholung (link).

Berechnung des VaR

- Sammlung und Analyse historischer Daten.
- Modellierung der Verteilung.
- Simulation von Kursentwicklungen gemäß der modellierten Verteilung (Monte Carlo Simulation).
- Große Zahl an Simulationen.
- Prognose von Kursentwicklungen.
- Berechnung des monetären Risikos für die Entwicklung des Portfolios.

Monte Carlo Simulation

- In den 1940er Jahren im Kontext der Atomforschung für die Abschätzung der Wechselwirkung zwischen Neutronen und Materie entwickelt.
- Einsatz zur Lösung analytisch nicht lösbarer Probleme.
- Basis ist die Durchführung von Zufalls-experimenten.
- Der Name geht auf die berühmte Spielbank zurück. Prinzipiell könnten Zufallszahlen auch mit einem Roulette erzeugt werden.

Verteilung von Zufallsvariablen

- Roulette – wie sind die Zahlen des Roulette verteilt?

  Gleichverteilt - jede Zahl (sollte) gleich wahrscheinlich sein.

- Wie ist die Größe erwachsener Frauen eines Kulturkreises verteilt?

  Normalverteilt – eine mittlere Größe wird relativ häufig gefunden während Extremwerte sehr selten auftreten.

Mittelwert, Varianz, Standardabweichung

Mittelwert: $$\mu_x = \sum_{t=1}^{T} x(t)/T$$

Varianz: $$\sigma_x^2 = \sum_{t=1}^{T} (x(t) - \mu_x)^2 / T$$

Standardabweichung: $$\sigma_x = \sqrt{\sigma_x^2}$$

Statistische Kenngrößen einer Stichprobe (empirisch)

Erwartungswert: $$\widehat{\mu_x} = \sum_{i=1}^{N} x_i / N$$

Empirische Varianz: $$\widehat{\sigma_x^2} = \sum_{i=1}^{N} (x_i - \widehat{\mu_x})^2 / (N-1)$$

Empirische Standardabweichung: $$\widehat{\sigma_x} = \sqrt{\widehat{\sigma_x^2}}$$

Berechnung der empirischen Standardabweichung

$$\widehat{\sigma_x} = \sqrt{\frac{N * \sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2}{N * (N-1)}}$$

Besonderheit: Erwartungswert muss vorab nicht berechnet werden

Normalverteilung (Gauß-Verteilung)

- Johann Carl Friedrich Gauß (1777-1855)
- Die Normalverteilung (Gauß-Verteilung) ist die wichtigste kontinuierliche Wahrscheinlich-keitsverteilung. Ihre Wahrscheinlichkeitsdichte wird als Gaußdichte oder Glockenkurve bezeichnet.
- Der zentrale Grenzwertsatz besagt, dass eine Summe von n unabhängigen, gleichverteilten Zufallsvariablen in der Grenze *n gegen unendlich* normalverteilt ist.
- Eine Normalverteilung wird durch Mittelwert und Varianz vollständig definiert.

Gaußdichte

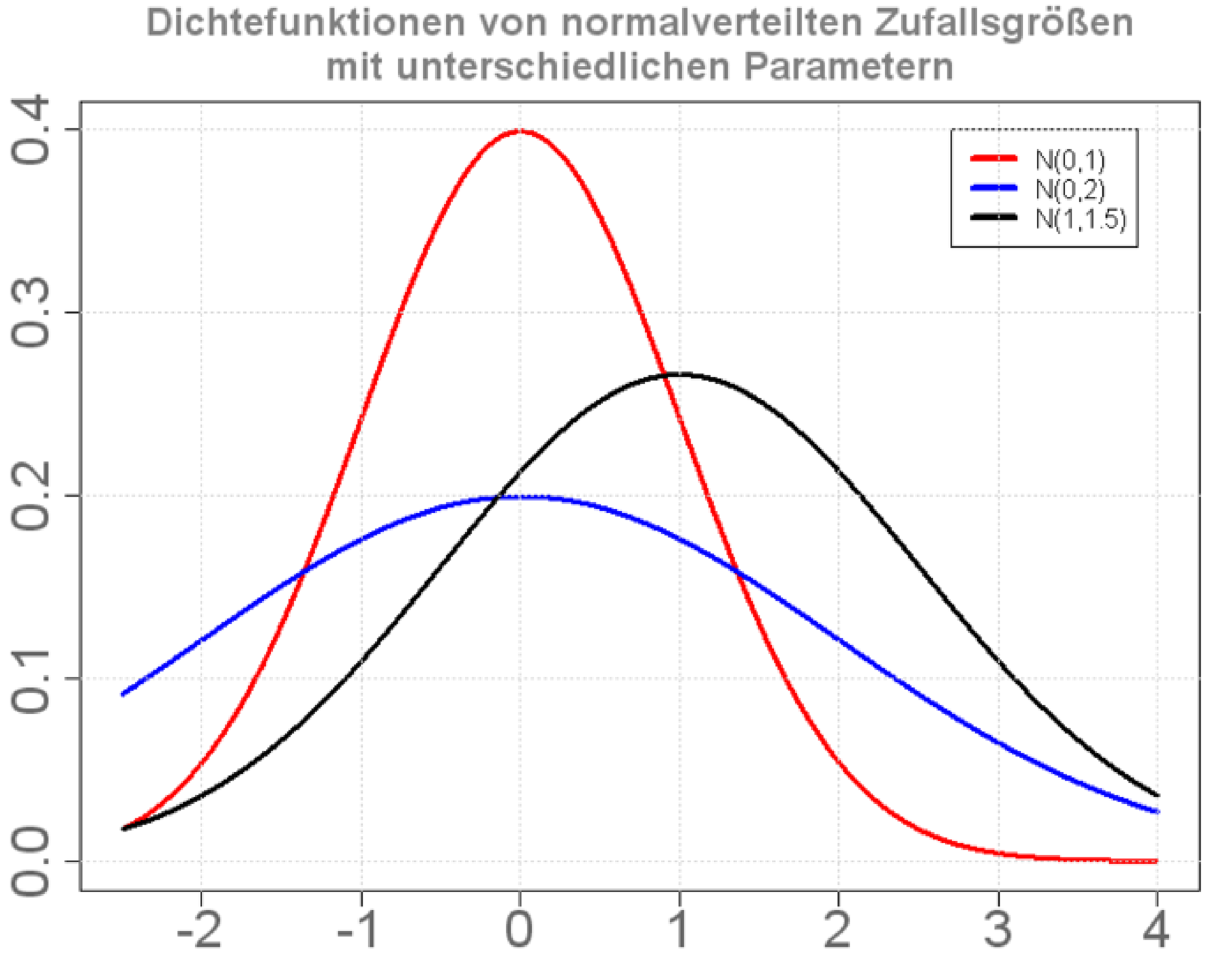
Gaußdichte der Normalverteilung:

$$f(x) = \frac{1}{\sqrt{2\pi \cdot \sigma_x}} e^{-(x-\mu_x)^2 / 2\sigma_x^2}$$

Gaußdichte der Standardnormalverteilung:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2 / 2}$$

Beispiele



Interaktive Demo:

http://campus.uni-muenster.de/fileadmin/einrichtung/imib/lehre/skripte/biomathe/bio/normn1.html

Charakteristische Wertebereiche

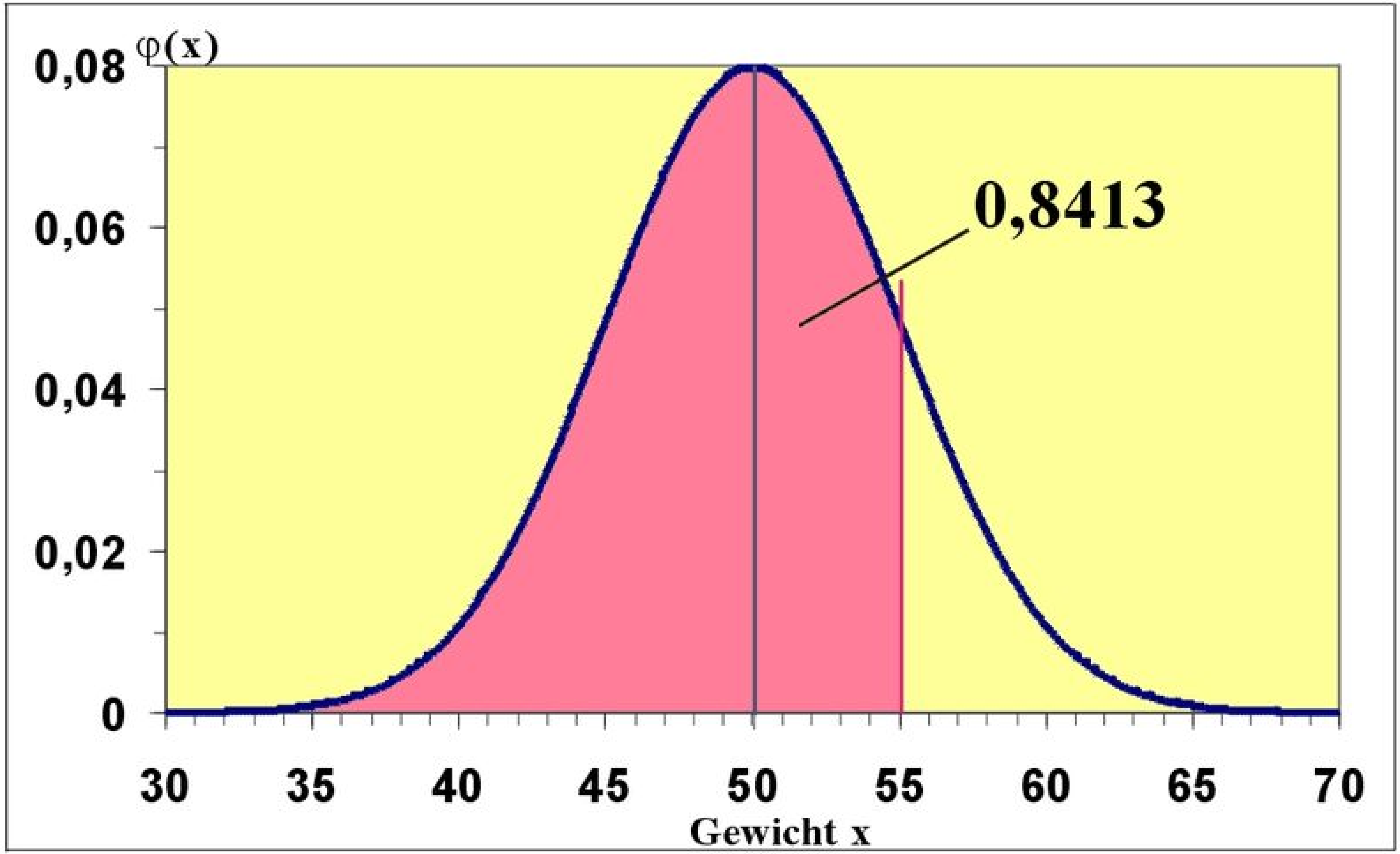
- 68,27% aller Werte einer normalverteilten Zufallsvariablen liegen innerhalb des Intervalls

$$\mu - \sigma \leq x \leq \mu + \sigma$$

- 95,45% aller Werte einer normalverteilten Zufallsvariablen liegen innerhalb des Intervalls

$$\mu - (2 * \sigma) \leq x \leq \mu + (2 * \sigma)$$

Beispiel



Quelle: http://de.wikibooks.org/wiki/Bild:NormDensEggs.jpg

Beispiel (IQ)

- Ein Intelligenztest sei so normiert, dass der mittlere IQ 100 beträgt und eine Varianz von 225 gemessen wird.
- Für einen Probanden wird ein IQ von 130 ermittelt.
- Schätzen Sie bitte den Anteil der Untersuchungs-gruppe ab, der einen geringeren IQ im Test erreichen wird.
- Ca. 97,7% der Probanden erreichen einen IQ unter 130.

Glossar

multivariat: Die gemeinsame Verteilung mehrerer Zufalls-variablen wird als multivariate Zufallsverteilung bezeichnet.

univariat: Die Verteilung einer einzelnen Zufallsvariablen wird als univariate Zufallsverteilung bezeichnet.

Erwartungsvektor und empirische Kovarianzmatrix

$$\widehat{\underline{\mu}_{\underline{x}}} = \sum_{n=1}^{N} \underline{x}(n) / N$$

$$\widehat{\underline{\underline{\mathrm{cov}}}_{\underline{x}}} = \begin{pmatrix} \mathrm{cov}(x_1, x_1) & \cdot & \cdot & \cdot & \mathrm{cov}(x_1, x_M) \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \mathrm{cov}(x_M, x_1) & \cdot & \cdot & \cdot & \mathrm{cov}(x_M, x_M) \end{pmatrix}$$

$$\mathrm{cov}(x_i, x_j) = \sum_{n=1}^{N} (x_i - \mu_i)(x_j - \mu_j) / (N-1) = \mathrm{cov}(x_j, x_i)$$

Korrelation

Die Korrelation ist ein Maß für die Beziehung zwischen zwei Zufallsvariablen. Hinweis: Eine starke Korrelation lässt keine Aussage über den kausalen Zusammenhang zwischen Zufallsvariablen zu.

Die Komponente $\operatorname{cov}(x_i, x_j)$ der Kovarianzmatrix der multivariaten Normalverteilung einer Modellgröße ist ein Maß für die Korrelation der *i*-ten Komponente der Modellgröße mit der *j*-ten Komponente der Modellgröße.

Scheinkausalität

Der Rückgang der Störche korreliert stark mit dem Rückgang der Geburtenrate. Daher sind Störche doch für unsere Fortpflanzung zuständig.

Diese Folgerung ist nicht zulässig. Erklärung: Die fortschreitende Industrialisierung hat sowohl zu einem Rückgang der Störche als auch zu einem Rückgang der Geburtenrate geführt.

Berechnung des Korrelationskoeffizienten

$$\rho(x_1, x_2) = \frac{\operatorname{cov}(x_1, x_2)}{\sigma(x_1) * \sigma(x_2)}$$

$\rho(x_1, x_2) = 0:$ Unabhängigkeit der beiden Variablen

$\rho(x_1, x_2) = 1:$ exakte positive lineare Abhängigkeit

$\rho(x_1, x_2) = -1:$ exakte negative lineare Abhängigkeit

Mehrdimensionale Gaußdichte

$$f(\underline{x}) = \frac{1}{\sqrt{(2\pi)^n \cdot \det(\underline{\underline{\mathrm{cov}}}_{\underline{x}})}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_x)^T (\underline{\underline{\mathrm{cov}}}_{\underline{x}})^{-1} (\underline{x}-\underline{\mu}_x)}$$