

2. Grundlagen

2.1. Prinzip der Rastersondenmikroskopie

Hier soll die prinzipielle Arbeitsweise des Rastertunnelmikroskops, im Englischen Scanning Tunneling Microscope - STM genannt, erläutert werden, welches 1971 von YOUNG entwickelt und 1981 von BINNING und ROHRER zu einem funktionsfähigen System ausgebaut wurde ([Binn86]).

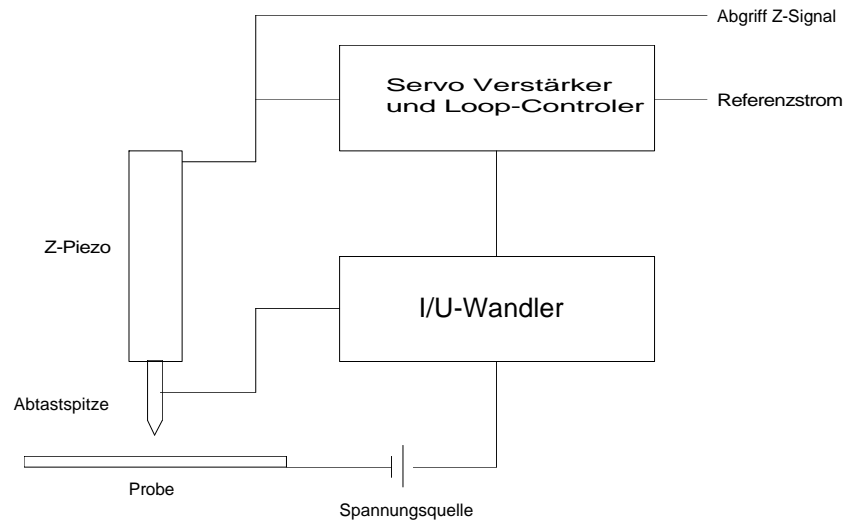


Abb. 1.: Arbeitsprinzip Rastertunnelmikroskop

Zwischen einer, auf einem piezoelektrischen Element (Z-Piezo) montierten Metallspitze und der zu untersuchenden (leitenden) Probe wird eine Spannung angelegt. Verringert man den Abstand zwischen Spitze und Probe auf etwa $10^{-9}m$, so ermöglicht der Tunneleffekt den Übertritt von Elektronen zwischen Spitze und Probe, es fließt ein *Tunnelstrom*. Der Tunnelstrom ist vom Abstand Spitze-Probe abhängig und läßt sich in Approximation in der Form

$$(2.1) \quad I \propto \frac{U}{s} e^{-A \cdot s \cdot \sqrt{\Phi}}$$

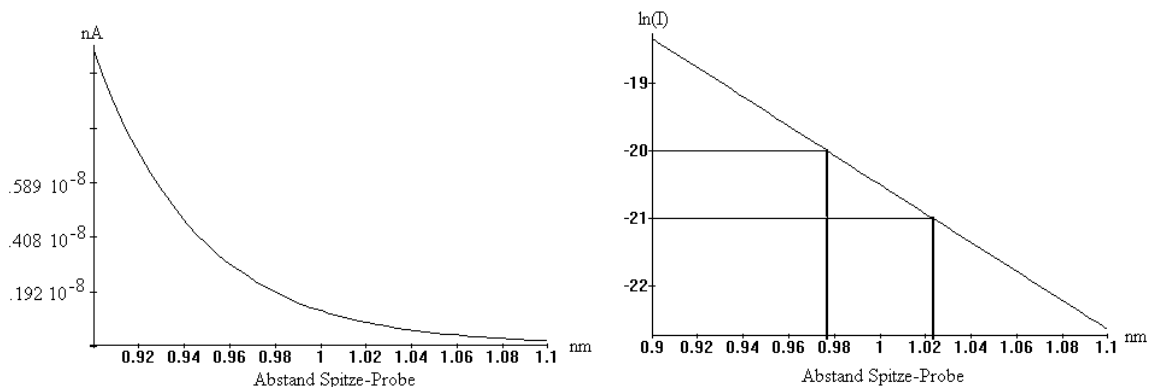


Abb. 2.: Tunnelstrom in Abhängigkeit vom Abstand Spitze-Probe (typisch ca. 1 nm)

darstellen ($A = \text{const.} = 10.25 \text{ eV}^{-1/2} \text{ nm}^{-1}$, Φ ist die mittlere Barrierenhöhe zwischen Spitze und Probe und liegt typischerweise bei etwa 4 eV). Der Tunnelstrom (zwischen 0.1 nA und 10 nA bei Tunnelspannungen U im Bereich von 1 mV bis 4 V ([Joerg93] und [Garn93]), siehe Abb 2.) wird verstärkt und anschließend einem Loop-Controller zugeführt. Dieser vergleicht ihn mit einem *Referenzstrom* und ermittelt aus der Abweichung ein Steuersignal für den Z-Piezo, welcher dann seinerseits den Abstand Spitze-Probe ändert. Ziel ist es, den Tunnelstrom konstant zu halten. Werden Spitze und Probe in X/Y-Richtung relativ zueinander bewegt und wird für diskrete X/Y-Werte der Wert des Z-Signals aufgezeichnet, so stellt die Menge der Tripel (x, y, z) die Topographie der Probe relativ zu einem Niveau $z = z_0$ dar. Der soeben beschriebene Meßmodus wird Constant Current Mode genannt.

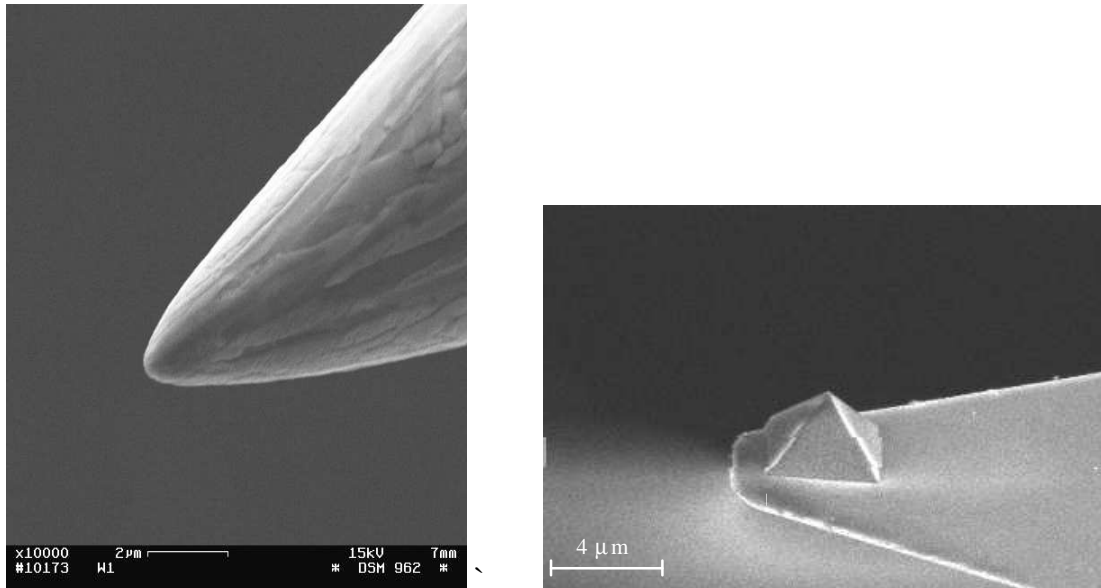


Abb. 3.: Abtastspitze eines Rastertunnel- und eines Rasterkraftmikroskopes

Es ist auch möglich im Constant Height Mode zu messen. Dabei wird das Z-Steuersignal konstant gehalten und der Tunnelstrom gemessen (der Vorteil dieses Modus besteht darin, daß der Loop-Controller nicht benutzt werden muß).

Weiterhin bekannt ist die Methode der Rasterkraftmikroskopie (RKM), bei der der Abstand zwischen Spitze und Oberfläche über die Kraftwechselwirkung geregelt wird ([Garn93]).

Häufig verwendet man für Rastersondenmikroskopische Methoden die englische Abkürzung SPM (Scanning Probe Microscopy), um die Gemeinsamkeiten der verschiedenen Systeme in einem Begriff zu erfassen.

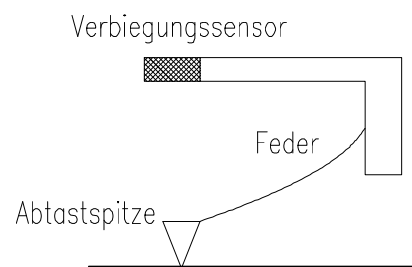


Abb. 4.: Prinzip des Rasterkraftmikroskopes

2.2. Definitionen und Begriffsbestimmungen

Gemessen wird im Rastersystem ein rechteckiger Bereich, der in seiner Lage durch zwei Ecken (a_x, a_y) und (b_x, b_y) gegeben ist. Zur Beschreibung des Meßvorganges führt man zunächst folgende Bezeichnungen ein:

- n_x, n_y : Anzahl Meßwerte pro Zeile bzw. Zeilen pro Bild
- d_x, d_y : physische Kantenlänge in x- und in y-Richtung
($dx := b_x - a_x$, $dy := b_y - a_y$, es sei o.B.d.A. $b_x > a_x$ und $b_y > a_y$)
- T_S : Gesamtdauer der Messung
- T_L : Dauer der Abtastung in einer Zeile

Horizontale Abtastfrequenz $f_u := n_x / T_L$, vertikale Abtastfrequenz $f_v := n_y / T_S$, maximale rekonstruierbare Frequenzkomponente in horizontaler Richtung $f_{u_{\max}} := f_u / 2$ und in vertikaler Richtung $f_{v_{\max}} := f_v / 2$ sind weitere Kenngrößen (siehe auch 2.4.1.). Da die Abtastung mit konstanter Geschwindigkeit erfolgen soll, kann man die Zeiten durch Wege ersetzen und gelangt damit zu den Definitionen $f_u := n_x / d_x$ und $f_v := n_y / d_y$. Die Zeit spielt erst bei der Untersuchung von Kriech- und Temperaturdrifteffekten eine Rolle. Die Ortskoordinaten (x, y) eines Rasterpunktes (i, j) lassen sich aus

$$x_j = a_x + dx \cdot \frac{j}{n_x - 1} \quad \text{und} \quad y_i = a_y + dy \cdot \frac{i}{n_y - 1}$$

bestimmen. Der Abgriff von Meßwerten erfolgt an allen, durch die Paare $(i, j) \in I_y \times I_x$; $I_y = [0, n_y - 1]$, $I_x = [0, n_x - 1]$ beschriebenen

Rasterpunkten von links oben nach rechts unten, das heißt in der Reihenfolge $(0, 0), \dots, (0, n_x - 1), (1, 0), \dots, (1, n_x - 1), \dots, (n_y - 1, n_x - 1)$. Das Paar $(0, 0)$ heißt linke obere Ecke, das Paar $(n_y - 1, n_x - 1)$ rechte untere Ecke. Die Menge aller Tripel $(i, j, z) \in I_x \times I_y \times M$ (M ist die Meßwertmenge) heißt Messung oder Bild der Messung.

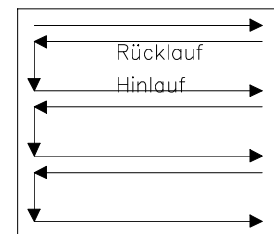


Abb. 5.: Spitzenbewegung relativ über der Probe

Nach [Jaro90] ist ein Bild ein zweidimensionales Signal $f: \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$, das zur visuellen Wahrnehmung durch den Menschen bestimmt ist. Für die Verwendung auf digitalen Rechenautomaten definiert man:

Def.: Ein Grauwertbild S ist die (n_x, n_y, n_z) Diskretisierung einer Abbildung $f: \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ über einem Intervall $[a_x, b_x] \times [a_y, b_y]$, mit:

$$f: \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}: (x, y) \alpha z \in \mathbf{R}, (x, y) \in [a_x, b_x] \times [a_y, b_y]$$

$$S: I_y \times I_x \rightarrow G: (i, j) \alpha \text{round}(f(x_j, y_i))$$

$$G = [0, n_z - 1] \subset N,$$

G heißt Grauwertmenge. Die Repräsentation einer Realisierung von S erfolgt als Matrix $s \in G^{n_y, n_x}$.

Bemerkung: Es werden grundsätzlich äquidistante Gitter zur Diskretisierung verwendet. Allerdings muß man beachten, daß durch Temperaturdrift oder Kriechen der X/Y-Piezos eines SPM-Gerätes Abweichungen auftreten können (siehe dazu 3.6.).

Für die von einem Rastersystem stammenden Bilddaten wird \mathbf{G} normalerweise auf die Meßwertmenge $\mathbf{M} = [m_{\min}, m_{\max}] \subseteq \mathbf{Z}$ erweitert, so daß eine Skalierung (siehe 3.2) notwendig ist, um S in ein Grauwertbild gemäß der Definition zu überführen.

2.3. Fouriertransformation und Diskrete Fouriertransformation

2.3.1. Fouriertransformation

Vielfach stellt sich bei der Analyse von Bilddaten die Frage nach dem Frequenzspektrum, zum Beispiel bei der Entwicklung von Filtern (siehe 3.4.1.).

Das Frequenzspektrum $\{(a_k, b_k)\}$ eines 1-dimensionalen periodischen Signals f erhält man, wenn es gelingt, die Funktion $f(x)$ in die FOURIERSche Reihe

$$(2.2) \quad f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cdot \cos(kx) + b_k \cdot \sin(kx))$$

zu entwickeln. Meist wird die Entwicklung, so sie existiert, nach endlich vielen Gliedern n abgebrochen. Dies ist auch in Hinblick auf die maximal rekonstruierbare Frequenzkomponente bei der Abtastung sinnvoll. Ist $f \in L^2([-\pi, \pi])$, so wird der mittlere quadratische Fehler

$$\delta^2 = \int_{-\pi}^{\pi} (f(x) - s_n(x))^2 dx$$

der Näherung

$$f(x) \approx s_n(x) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos(kx) + b_k \sin(kx))$$

minimal, wenn

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) dx, \quad k = 0, 1, 2, \dots$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) dx, \quad k = 1, 2, \dots$$

ist. Hat $f(x)$ die Periode $2l$ so wird kx durch $(k\pi x)/l$ ersetzt und es wird dann

$$a_k = \frac{1}{l} \int_{-l}^l f(x) \cos\left(\frac{k\pi x}{l}\right) dx, \quad k = 0, 1, 2, \dots$$

$$b_k = \frac{1}{l} \int_{-l}^l f(x) \sin\left(\frac{k\pi x}{l}\right) dx, \quad k = 1, 2, \dots$$

(2.2) läßt sich für $2l$ -periodische Funktionen auch in der Form

$$(2.3) \quad f(x) = \sum_{k=-\infty}^{\infty} c_k e^{i \frac{k\pi x}{l}}$$

mit

$$(2.4) \quad c_k = \frac{1}{2l} \int_{-l}^l f(x) e^{-i \frac{k\pi \cdot x}{l}} dx$$

schreiben. Für $2l_x, 2l_y$ -periodische Funktionen $f: \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ gilt dann analog:

$$(2.5) \quad f(x, y) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} c_{j,k} e^{i\pi \left(\frac{kx}{l_x} + \frac{jy}{l_y} \right)}$$

und

$$(2.6) \quad c_{j,k} = \frac{1}{4l_x l_y} \int_{-l_y}^{l_y} \int_{-l_x}^{l_x} f(x, y) e^{-i\pi \left(\frac{kx}{l_x} + \frac{jy}{l_y} \right)} dx dy$$

Was passiert, wenn man die Betrachtung auf $L^2(\mathbf{R})$ ausdehnt? Für diesen Fall erklärt man zunächst die Fouriertransformation.

Def.: Die Abbildung $\mathcal{F}: f \in L^2(\mathbf{D}) \rightarrow F \in L^2(\mathbf{\Omega})$ mit $F(u) = \int_{\mathbf{D}} f(x) \cdot e^{-i \cdot u^T x} dx$

heißt Fouriertransformation. $F(u)$ heißt Fouriertransformierte oder Spektrum oder Spektraldichte von $f(x)$. Für den 2-dimensionalen Fall wird

$$(2.7) \quad F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \cdot e^{-i \cdot (ux+vy)} dx dy.$$

Zum Vergleich der Eigenschaften der Fourierkoeffizienten mit denen der Fouriertransformierten seien jetzt f, g, h Abbildungen $\mathbf{R} \rightarrow \mathbf{R}$, mit eventuell eingeschränktem Definitionsbereich. f_k, g_k und h_k seien die Fourierkoeffizienten für den Fall $f, g, h \in L^2([-l, l])$ und F, G, H die Fouriertransformierten für den Fall $f, g, h \in L^2(\mathbf{R})$ (f, g, h periodisch fortgesetzt auf \mathbf{R}). Dann gilt:

- 1) Additionssatz: Sei $h = \alpha \cdot f + \beta \cdot g \Rightarrow h_k = \alpha \cdot f_k + \beta \cdot g_k$ bzw. $H = \alpha \cdot F + \beta \cdot G$
- 2) Verschiebungssatz: Sei $h(x) = f(x - \tau)$, dann gilt $h_k = f_k \cdot e^{-i \cdot k \cdot \tau}$ bzw. $H(u) = F(u) \cdot e^{-i \cdot u \cdot \tau}$

Für die Fouriertransformierte gilt weiterhin:

- 3) Ähnlichkeitssatz: Sei $h(x) = f(\alpha \cdot x) \Rightarrow H(u) = \frac{1}{|\alpha|} F\left(\frac{u}{\alpha}\right)$
- 4) Vertauschungssatz: $F(u) = F(-u)$
- 5) Differentiation-Integration: Sei $h(x) = \frac{d}{dt} f(x) \Rightarrow H(u) = -i \cdot u \cdot F(u)$.
- 6) Faltungssatz: Sei $h = f * g \Rightarrow H(u) = F(u) \cdot G(u)$.
- 7) Parsevalsche Beziehung: $\int |f(x)|^2 dx = \int |F(u)|^2 du$.

Diese Beziehungen gelten analog für den 2-dimensionalen Fall.

Zum Abschluß soll noch der Zusammenhang zwischen den Fourierkoeffizienten und der Fouriertransformierten angegeben werden, der in äquivalenter Form auch im 2-dimensionalen Fall Gültigkeit hat.

Satz: Sei $f(x) \in L^1(\mathbf{R})$, dann ist die Funktion $\varphi(x) = 2\pi \cdot \sum_{j=-\infty}^{\infty} f(x + 2\pi \cdot j)$ 2π -periodisch,

Element $L^1([-\pi, \pi])$ und es gilt: $\varphi_k = F(k)$. Im 2-dimensionalen Fall gilt

$$\varphi(x, y) = 4\pi^2 \cdot \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} f(x + 2\pi \cdot j, y + 2\pi \cdot i) \Rightarrow \varphi_{k,l} = F(k, l).$$

Beweis: Man erkläre $f_k(x) = f(x + 2\pi \cdot k)$; $x \in [-\pi, \pi]$, dann folgt wegen $f \in L^1(\mathbf{R})$

$$\|f\|_{L^1(\mathbf{R})} = \sum_{k=-\infty}^{\infty} \|f_k\|_{L^1(-\pi, \pi)}. \text{ Damit konvergiert die Reihe } \sum_{k=-\infty}^{\infty} f_k; f_k \in L^1(-\pi, \pi) \text{ absolut}$$

für fast alle x aus $[-\pi, \pi]$ gegen ein Element φ aus dem Banachraum $L^1(-\pi, \pi)$.

Es gilt weiter

$$\varphi(x) = 2\pi \cdot \sum_{k=-\infty}^{\infty} f(x + 2\pi \cdot k) = 2\pi \cdot \sum_{k \in \mathbb{Z}} f(x + 2\pi \cdot (k+1)) = \varphi(x + 2\pi).$$

$$\begin{aligned} \varphi_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(x) \cdot e^{-i \cdot k \cdot x} dx = \sum_{j=-\infty}^{\infty} \int_{-\pi}^{\pi} f(x + 2\pi \cdot j) \cdot e^{-i \cdot k \cdot x} dx \\ &= \int_{-\infty}^{\infty} f(x) \cdot e^{-i \cdot k \cdot x} dx = F(k) \end{aligned}$$

$\varphi_{k,l} = F(k, l)$ folgt analog. \square

2.3.2. Diskrete Fouriertransformation

Zunächst motiviert man die diskrete Fouriertransformation (im folgenden mit DFT abgekürzt) als Approximation der Fouriertransformation einer Funktion f , die der Bedingung $f(x) \equiv 0 \forall x: |x| > A/2$ genügt. Diese Bedingung ist für Bilder aus Rastertunnelsystemen als Abbildungen $\mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ a priori erfüllt (begrenzte Ausdehnung).

Die Fouriertransformation wird dann zu $F(u) = \int_{-\infty}^{\infty} f(x) \cdot e^{-i \cdot u \cdot x} dx = \int_{-A/2}^{A/2} f(x) \cdot e^{-i \cdot u \cdot x} dx$. Ist $f(x)$ an

den Stellen $x_k = -\frac{A}{2} + k \cdot \Delta x$; $\Delta x = \frac{A}{N}$ ($k=0, (1), N$; das sind $N+1$ Stützstellen) durch $f_k = f(x_k)$ gegeben, so kann man $F(u)$ mit Hilfe der Trapezregel approximieren:

$$(2.8) \quad F(u) \approx \frac{A}{N} \cdot \sum_{k=0}^{N-1} f_k \cdot e^{-i \cdot u \cdot x_k}.$$

(2.8) gilt nur, falls $f_0 \cdot e^{i \cdot u \cdot \frac{A}{2}} = f_N \cdot e^{-i \cdot u \cdot \frac{A}{2}}$ bzw. $f_0 \cdot e^{i \cdot u \cdot A} = f_N$. Es werden jetzt nur noch $N-1$ Stützstellen verwendet. Im nächsten Schritt schränkt man u unter Verwendung der Beziehungen

$$\Delta u = \frac{1}{A},$$

$$A \cdot \Omega = N$$

und

$$\Delta x \cdot \Delta u = \frac{1}{N}$$

ein.

Bemerkung: Die kleinste, bei der Diskretisierungsschrittweite Δx erkennbare Frequenz ist $1/A$. Vorausschauend auf die Implementation im Rechenautomaten werden die N Stützstellen mit $0..(N-1)$ numeriert.

Man erhält jetzt

$$\begin{aligned} F(u_l) &\approx \frac{A}{N} \cdot \sum_{k=0}^{N-1} f_k \cdot e^{-i \cdot u_l \cdot x_k} \\ &= \frac{A}{N} \cdot \sum_{k=0}^{N-1} f_k \cdot e^{-i \left(-\frac{N}{2A} + l \cdot \frac{N}{AN} \right) \left(-\frac{A}{2} + k \cdot \frac{A}{N} \right)} \\ &= \frac{A}{N} \cdot \sum_{k=0}^{N-1} f_k \cdot e^{-i \frac{(2l-N)(2k-N)}{4N}}. \end{aligned}$$

Wählt man N geradzahlig, so folgt

$$(2.9a) \quad F_l = \frac{1}{N} \sum_{k=-N/2+1}^{N/2} f_k \cdot e^{-i \frac{l \cdot k}{N}}$$

und in der alternativen Form

$$(2.9b) \quad F_l = \frac{1}{N} \cdot \sum_{k=0}^{N-1} f_k \cdot e^{-i \frac{l \cdot k}{N}}.$$

Die beiden Darstellungen liefern äquivalente Ergebnisse, unterscheiden sich aber dadurch, daß in (2.9b) die mit niedrigen Frequenzen korrespondierenden Stellen am Rand des Feldes F zu treffen sind, während sie bei (2.9a) zentriert liegen (siehe Abb. 6). (2.9a) wird deshalb auch zentrierte Darstellung genannt. Der Vorteil der zweiten Darstellung besteht darin, daß N nicht mehr geradzahlig sein muß. Es gilt weiter $F(u_l) \approx A \cdot F_l$. Verwendet man eine Grundfrequenz $\omega_0 = \frac{2\pi}{n}$ im Exponenten, so ist die Interpretation

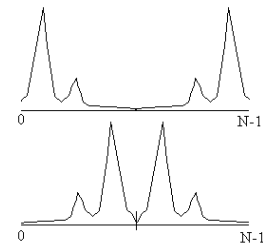


Abb. 6.: verschiedene Symmetrien bei DFT

der DFT als Annäherung an die Fourierkoeffizienten der A -periodisch fortgesetzten Funktion f sinnvoll. Interessant ist jetzt die Frage nach dem Fehler $\delta = |F_k - c_k|$ der diskreten Fouriertransformation bei der Approximation der Fourierkoeffizienten c_k einer Funktion f . Zunächst folgt für den 1-dimensionalen Fall mit

$$f_j = f(x_j) = \sum_{k=-\infty}^{\infty} c_k \cdot e^{i \cdot 2\pi \frac{j \cdot k}{n}}$$

und

$$F_k = \frac{1}{n} \sum_{j=-n/2+1}^{n/2} f_j \cdot e^{-i \cdot 2\pi \frac{k \cdot j}{n}} = \dots = \sum_{j=-\infty}^{\infty} c_{k+j \cdot n}$$

die diskrete Poisson-Summenformel

$$(2.10) \quad F_k = c_k + \sum_{j=1}^{\infty} (c_{k+jn} + c_{k-jn}).$$

Es ist jeweils $-n/2 + 1 \leq k \leq n/2$. Der aliasing-Effekt für ein nicht bandbegrenztetes Signal $f(x)$ läßt sich unmittelbar aus (2.10) ableiten:

$$(2.11) \quad \begin{aligned} \text{mit } \omega_k &= k \cdot \omega_0 \Rightarrow \\ e^{i\omega_k \cdot x_j} &= e^{i2\pi \frac{k \cdot j \cdot A}{A \cdot n}} = e^{i2\pi \frac{k \cdot j}{n}} \\ e^{i\omega_{k+mn} \cdot x_j} &= e^{i2\pi \frac{k+mn \cdot j \cdot A}{A \cdot n}} = e^{i2\pi \frac{k \cdot j}{n}} \cdot e^{i2\pi \cdot m \cdot j} = e^{i2\pi \frac{k \cdot j}{n}}, \end{aligned}$$

das heißt im Koeffizienten F_k der DFT sind die Koeffizienten c_{k+mn} der (exakten) Fouriertransformation enthalten. Oder anders: eine Unterscheidung der Koeffizienten $k+mn$ ist in der DFT nicht möglich.

Satz 1: Es seien $f(x)$ bis einschließlich $(p-1)$ -te Ableitung $f^{(p-1)}(x)$, $p \geq 1$ l -periodische Funktionen und stetig auf $[-l/2, l/2]$. $f^{(p)}(x)$ sei auf $[-l/2, l/2]$ beschränkt und stückweise monoton, dann gilt:

$$(2.12) \quad |c_k| \leq \frac{K}{|k|^{p+1}} \quad \forall k; K \in \mathbf{R} \text{ const.}$$

Womit unter gleichen Bedingungen der Fehler der diskreten Fouriertransformation zu

$$(2.13) \quad |F_k - c_k| \leq \frac{K}{n^{p+1}}$$

bestimmt werden kann.

Für den Fehler der DFT bei allgemeiner Eingabefunktion (keine Periodizität vorausgesetzt, nicht notwendig bandbegrenzt) gilt die folgende Aussage:

Satz 2: Die l -periodische Fortsetzung von f sei $(p-1)$ -mal stetig differenzierbar und $f^{(p)}$ sei beschränkt und stückweise monoton auf $[-l/2, l/2]$ für $p \geq 1$. Weiterhin sei vorausgesetzt, $f(x)=0$ für $|x| \geq l/2$. Falls die Funktion $f(x)$ mit n Stützstellen auf dem Intervall $[-l/2, l/2]$ gegeben ist, dann ist

$$(2.14) \quad |l \cdot F_k - F(\omega_k)| \leq \frac{Kl}{n^{p+1}}$$

der Fehler der diskreten n -Punkt Fouriertransformation an den Stellen $\omega_k = k/l$.
($k = -n/2 + 1 \dots n/2$, K von n und k unabhängige Konstante, $l = n\Delta x$)

Für die Beweisideen soll hier auf [Briggs95] verwiesen werden.
Zur Bestimmung des Fehlers bei der 2-dimensionalen DFT

$$(2.15a) \quad F_{\mu, \nu} = \frac{1}{m \cdot n} \sum_{k=-m/2+1}^{m/2} \sum_{l=-n/2+1}^{n/2} f_{k,l} \cdot e^{-i(\mu \cdot k + \nu \cdot l)}$$

beziehungsweise

$$(2.15b) \quad F_{\mu, \nu} = \frac{1}{M \cdot N} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} f_{k,l} \cdot e^{-i(\mu \cdot k + \nu \cdot l)}$$

verwendet man wieder die diskrete Poisson-Summenformel

$$(2.16) \quad \begin{aligned} F_{k,l} &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} c_{k+i, l+j} \\ &= c_{k,l} + \sum_{i=1}^{\infty} (c_{k+i, l} + c_{k-i, l}) + \sum_{j=1}^{\infty} (c_{k, l+j} + c_{k, l-j}) \\ &\quad + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (c_{k+i, l+j} + c_{k+i, l-j} + c_{k-i, l+j} + c_{k-i, l-j}). \end{aligned}$$

Lemma 1: $f(x, y)$ und $g(x, y)$ seien zwei Abbildungen $\mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ und jeweils in x bzw. y stetig differenzierbar, dann gilt

$$\iint f \cdot g_x \, dx dy + \iint f_x \cdot g \, dx dy = \int f \cdot g \, dy$$

und analog

$$\iint f \cdot g_y \, dy dx + \iint f_y \cdot g \, dy dx = \int f \cdot g \, dx$$

(partielle Integration).

Daraus folgt

$$\begin{aligned} 4 \cdot l_x l_y \cdot c_{k,l} &= \int_{-l_y}^{l_y} \int_{-l_x}^{l_x} f(x, y) \cdot e^{-i\pi \left(\frac{kx}{l_x} + \frac{ly}{l_y} \right)} dx dy \\ &= \int_{-l_y}^{l_y} \left[f(x, y) \cdot \frac{l_x}{-i \cdot \pi \cdot k} \cdot e^{-i\pi \left(\frac{kx}{l_x} + \frac{ly}{l_y} \right)} \right]_{-l_x}^{l_x} dy + \frac{l_x}{i \cdot \pi \cdot k} \int_{-l_y}^{l_y} \int_{-l_x}^{l_x} f_x(x, y) \cdot e^{-i\pi \left(\frac{kx}{l_x} + \frac{ly}{l_y} \right)} dx dy \end{aligned}$$

wobei der erste Term der rechten Seite zu 0 wird, falls f $2l_x$ -periodisch in x ist. f sei p -fach stetig differenzierbar in x und jeweils $2l_x$ -periodisch, so folgt letztlich

$$|c_{k,l}| \leq \frac{1}{4l_x l_y} \cdot \left(\frac{l_x}{\pi \cdot |k|} \right)^p \cdot C_l = \frac{K_l}{|k|^p}.$$

Es sei f weiterhin q -fach in y ableitbar, so gilt für alle Funktionen, die $2l_x, 2l_y$ -periodisch und p -fach in x sowie q -fach in y periodisch ableitbar sind, die Abschätzung der Fourierkoeffizienten

$$(2.17) \quad |c_{k,l}| \leq \frac{K}{|k|^p \cdot |l|^q}$$

und sogar

$$(2.18) \quad |c_{k,l}| \leq \frac{K}{|k|^{p+1} \cdot |l|^{q+1}}.$$

Die letzte Abschätzung gilt, da der Integrand der rechten Seite bei vorausgesetzter Beschränktheit und stückweiser Monotonie durch eine Konstante Ck^{-l} beschränkt ist. Damit ergibt sich nach (2.16)

$$(2.19) \quad |F_{k,l} - c_{k,l}| \leq K \cdot \frac{1 + n^{p+1} + m^{q+1}}{n^{p+1} m^{q+1}}$$

und der Fehler der diskreten (n,m) -Punkt Fouriertransformation lautet

$$(2.20) \quad |l_x l_y F_{k,l} - F(u_k, v_l)| \leq \frac{1 + n^{p+1} + m^{q+1}}{n^{p+1} m^{q+1}}.$$

Folgendes Beispiel in [Briggs95] motiviert die Einführung des Kürzels AVED (average value at endpoints and discontinuities). Man betrachte die nicht-periodische Funktion $f(x)=x$ auf dem Intervall $[-1/2, 1/2]$. Die exakten Fourierkoeffizienten ergeben sich zu

$$c_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} x \cdot e^{-i2\pi kx} dx = i \frac{\cos(\pi k)}{2\pi k}$$

und die der diskreten zu

$$F_k = i \frac{\cos(\pi k) \cdot \sin \theta_k}{4n \sin^2(\theta_k / 2)}; \quad \theta_k = \frac{2k\pi}{n}; \quad k = -n/2(1)n/2, \quad k \neq 0.$$

Für $n=8$ folgt dann (mit AVED: $f_4=0$):

k	f_k	DFT ohne AVED	mit AVED	diskr. Koeff.	Fourierkoeff.
-3	-0.3750	-0.0625+0.0259i	0.0000+0.0259i	0.0259i	0.0531i
-2	-0.2500	0.0625-0.0625i	0.0000-0.0625i	-0.0625i	-0.0796i
-1	-0.1250	-0.0625+0.1509i	0.0000+0.1509i	0.1509i	0.1592i
0	0.0000	0.0625+0.0000i	0.0000+0.0000i	0.0000i	0.0000i
1	0.1250	-0.0625-0.1509i	0.0000-0.1509i	-0.1509i	-0.1592i
2	0.2500	0.0625+0.0625i	0.0000+0.0625i	0.0625i	0.0796i
3	0.3750	-0.0625-0.0259i	0.0000-0.0259i	-0.0259i	-0.0531i
4	0.5000	0.0625+0.0000i	0.0000+0.0000i	-0.0000i	0.0398i

Man stellt fest, daß die approximierten Fourierkoeffizienten ohne AVED nicht mit den erwarteten übereinstimmen und für wachsendes k auch ein anderes Verhalten zeigen.

Dieser Effekt hat seine Ursache im Satz von Dirichlet.

Satz: $f(x)$ genüge in $(-\pi, \pi)$ den Dirichlet-Bedingungen:

- $(-\pi, \pi)$ läßt sich in endlich viele Teilintervalle zerlegen, in denen $f(x)$ stetig und monoton ist
- ist $f(x)$ unstetig in x_0 , so existieren $f(x_0+h)$ und $f(x_0-h)$, dann konvergiert die Fourierreihe von $f(x)$ und es gilt

$$\lim_{n \rightarrow \infty} \left[\frac{a_0}{2} + \sum_{k=1}^n (a_k \cos(kx) + b_k \sin(kx)) \right] = \left\{ \begin{array}{ll} f(x); & \text{falls } f \text{ stetig in } x \\ \frac{f(x+h) + f(x-h)}{2}; & \text{sonst} \end{array} \right\}.$$

Bei periodischer Fortsetzung außerhalb von $(-\pi, \pi)$ gilt dieser Satz für alle x . Im Beispiel tritt offensichtlich am Rand ein Problem auf, da die Fourierreihe gegen $(f(l/2-h)+f(l/2+h))/2 = (f(l/2-h)+f(-l/2+h))/2=0$ strebt. Um korrekt zu arbeiten, muß der nachfolgende Hilfssatz angewendet werden.

Lemma 2: (AVED) Damit das Verhalten der Koeffizienten der diskreten Fouriertransformation mit dem der Fourierkoeffizienten übereinstimmt, muß die Eingabe der DFT so gestaltet sein, daß für Unstetigkeitsstellen x_d , insbesondere an den Rändern gilt:

$$f(x_d) := (f(x_d - 0) + f(x_d + 0)) / 2.$$

Auf einen weiteren Effekt wird im folgenden aufmerksam gemacht.

In der Praxis kann nie sichergestellt werden, daß für die DFT immer ein ganzzahliges Vielfaches der Periode der Funktion $f(x)$ zur Verfügung steht.

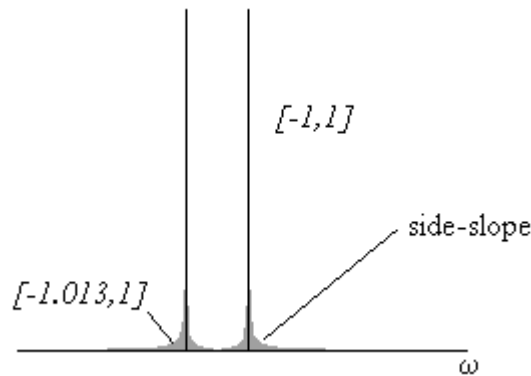


Abb. 7.: DFT der Funktion $\cos(20\pi x)$ über den Intervallen $[-1, 1]$ und $[-1.013, 1]$.

Würde die diskrete Fouriertransformation exakt arbeiten, würden nur die markierten Stellen in Abbildung 7 Werte ungleich 0 annehmen (Intervall $[-1, 1]$). Der sogenannte leakage-Effekt äußert sich darin, daß für Intervalle, die keine volle Periode umfassen, auch Koeffizienten der Fouriertransformierten, die nicht mit Frequenzen von $f(x)$ korrespondieren, ungleich 0 werden. Bisher wurde nur der 1-dimensionale Fall diskutiert. Für die Bildverarbeitung müssen aber Abbildungen $f: \mathbf{R} \times \mathbf{R} \rightarrow \mathbf{R}$ (bzw. $s: \mathbf{N} \times \mathbf{N} \rightarrow \mathbf{N}$ als Diskretisierung von f) betrachtet werden. Jedoch kann man sehr rasch eine Analogie zur schon behandelten Situation erkennen, womit der aliasing-Effekt leicht zu zeigen ist:

$$\begin{aligned} u_k &= k \cdot u_0 \quad \wedge \quad v_l = l \cdot v_0 \quad \Rightarrow \\ e^{\bar{\imath}(u_k \cdot x_j + v_l \cdot y_i)} &= e^{\bar{\imath} 2\pi \left(\frac{k \cdot j \cdot d_x}{d_x \cdot N} + \frac{l \cdot i \cdot d_y}{d_y \cdot M} \right)} = e^{\bar{\imath} 2\pi \cdot \left(\frac{k \cdot j}{N} + \frac{l \cdot i}{M} \right)} \\ (2.21) \quad e^{\bar{\imath}(u_{k+\mu \cdot N} \cdot x_j + v_{l+\nu \cdot M} \cdot y_i)} &= e^{\bar{\imath} 2\pi \left(\frac{k+\mu \cdot N}{d_x} \cdot \frac{j \cdot d_x}{N} + \frac{l+\nu \cdot M}{d_y} \cdot \frac{i \cdot d_y}{M} \right)} \\ &= e^{\bar{\imath} 2\pi \cdot (\mu \cdot j + \nu \cdot i)} \cdot e^{\bar{\imath} 2\pi \left(k \cdot \frac{j}{N} + l \cdot \frac{i}{M} \right)} = e^{\bar{\imath} 2\pi \cdot \left(k \cdot \frac{j}{N} + l \cdot \frac{i}{M} \right)}. \end{aligned}$$

Ebenso kann man ein Beispiel für den leakage-Effekt angeben (Abbildung 8).

Zum Abschluß dieses Kapitels sollen noch einige grundlegende Eigenschaften der DFT zusammengefaßt werden.

- Periodizität: Es gilt $F_{k+N} = F_k \quad \forall k \in \mathbf{Z}$.
- Linearität: Sei $h_k = \alpha \cdot f_k + \beta \cdot g_k \Rightarrow H_l = \alpha \cdot F_l + \beta \cdot G_l$.
- Verschiebung und Modulation: Sei $h_k = f_{n-j} \Rightarrow H_l = F_l \cdot \omega_N^{-j \cdot l}$; $\omega_N = N \cdot \omega_0$.
- hermitesche Symmetrie: Bezeichne \mathcal{D} die diskrete Fouriertransformation, dann ist

$$f_i = \overline{\mathcal{D}\{N \cdot \overline{F_k}\}_i}$$
 bzw. $F_k = \overline{\mathcal{D}^{-1}\left\{\frac{1}{N} \cdot \overline{f_i}\right\}_k}$

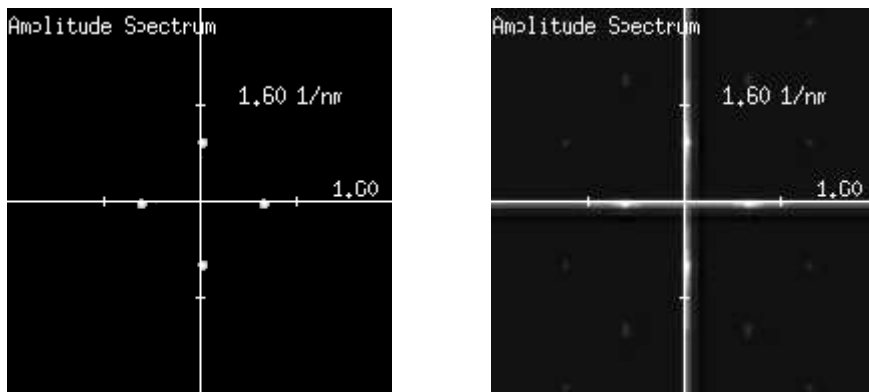


Abb. 8.: links DFT von $\cos(10\pi x) + \sin(10\pi x)$ im Intervall $[-10, 10] \times [-10, 10]$, rechts $[-10.13, 10.13] \times [-10.13, 10.13]$

Bemerkung: Die diskrete Fouriertransformierte wird meist über den Algorithmus der Schnellen Fouriertransformation (FFT) berechnet. Hierfür müssen $n=2^k$ Werte zur Verfügung stehen, im 2-dimensionalen Fall $n \cdot m = 2^{k_n} \cdot 2^{k_m}$ Werte. Die DFT wird dann mit dem Aufwand $O(n \cdot \log_2 n)$ bzw. $O(m \cdot n \cdot \log_2 n) + O(n \cdot m \cdot \log_2 m)$ berechnet ([Beth84]).

2.4. Einige Grundlagen der Nachrichtentechnik und Systemtheorie

Bei der Erläuterung der Arbeitsweise eines Rastertunnelmikroskopes wurde bereits der Begriff *Signal* verwendet. Dieser ist von fundamentaler Bedeutung für das Verständnis der technischen Hintergründe und unerlässlich für die mathematische Beschreibung von Vorgängen im (Rastertunnel-)System. Ein wesentlich tieferer Einblick in die Systemtheorie wird zum Beispiel in [Schue91], [Kress89] und [Lange75] vermittelt.

2.4.1. Signal und Abtasttheorem

Def.: Ein (zeitabhängiges) analoges Signal oder kurz Analogsignal ist eine Abbildung $f: \mathbf{R} \rightarrow \mathbf{W} \subseteq \mathbf{R}$. $D(f)$ repräsentiert die Zeit, $W(f)=\mathbf{W}$ die Werte einer physikalischen, zeitabhängigen Größe.

Beispiel.: Der Spannungsverlauf am z -Abgriff des Rastertunnelsystems.

Man unterscheidet zwischen deterministischen und zufälligen Signalen. Bei ersteren ist der Signalverlauf bekannt. Weiterhin werden periodische und nichtperiodische Signale unterschieden.

Ein Signal $s(t)$ kann mittels seiner Fourierreihenentwicklung im Frequenzbereich beschrieben werden. Betrachtet man zunächst ein periodisches Signal mit $s(t)=s(t+T)$, T heißt die Periode von s , dann stellt man fest, daß sich s aus der Grundschwingung mit der Frequenz $f_0=1/T$ und den Oberschwingungen mit Frequenzen $k \cdot f_0$; $k \in \mathbf{Z}$ zusammensetzen läßt. Das heißt mit $\omega_0 = 2\pi f_0$ gilt

$$\begin{aligned} s(t) &= s_0 + \sum_{k=1}^{\infty} S_k \cdot \cos(k \cdot \omega_0 \cdot t + \varphi_k) \\ &= s_0 + \sum_{k=1}^{\infty} [S_k \cdot \cos(\varphi_k) \cdot \cos(k \cdot \omega_0 \cdot t) + S_k \cdot \sin(\varphi_k) \cdot \sin(k \cdot \omega_0 \cdot t)]. \end{aligned}$$

s_0 wird Gleichanteil des Signals s genannt. Als weitere Kenngröße eines Signals werden der Klirrfaktor $k_n = S_n / S_1$, der Gesamtklirrfaktor $k = \frac{1}{S_1} \sqrt{S_2^2 + S_3^2 + \dots}$ und häufig auch der

Effektivwert $S_{\text{eff}} = \sqrt{s_0^2 + (S_1^2 + S_2^2 + \dots) / 2}$ verwendet.

Ein Beispiel für ein periodisches Signal ist der Rechteckimpuls mit der Periode T und dem Tastverhältnis T_I/T (Abb. 9). Die Fourierreihenentwicklung lautet

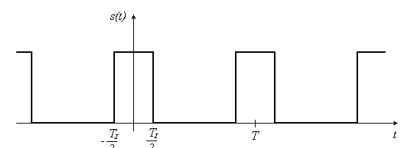


Abb. 9.: Rechteckimpuls mit Tastverhältnis T_I/T

$$s(t) = \frac{T_I}{T} + \sum_{k=1}^{\infty} \frac{2}{k\pi} \sin\left(k\pi \cdot \frac{T_I}{T}\right) \cos(k\omega_0 t).$$

Mit dem Grenzübergang $T \rightarrow \infty$ gelangt man zur Darstellung nichtperiodischer Signale im Frequenzbereich und gibt deren komplexe Spektraldichte $S(f) = \int_{-\infty}^{\infty} s(t) \cdot e^{2\pi i \cdot t \cdot f} dt$ an. Für den

Rechteckimpuls erhält man $S(f) = \frac{\sin(\pi \cdot f \cdot T_I)}{\pi \cdot f}$. Normiert man dieses Signal mit $1/T_I$ und läßt

T_I gegen 0 gehen, so erhält man den Einheits- oder Dirac-Impuls $\delta(t)$ (der Dirac-Impuls ist eine verallgemeinerte Funktion (Distribution)). Mittels des Einheits-Impulses kann die Abtastung eines zeitabhängigen Signals durch $s(t) \cdot \delta(t - t_0) = s(t_0) \cdot \delta(t - t_0)$ beschrieben werden. Dieses Verhalten nennt man auch die Ausblendeigenschaft des Dirac-Impulses. Ein

weiteres spezielles Signal ist der Einheitssprung, der durch $\sigma(t) = \{1; t \geq 0 \text{ und } 0; \text{sonst}\}$ dargestellt wird. Wichtig zu erwähnen ist der Zusammenhang zwischen Einheitssprung und Dirac-Impuls, es ist

$$(2.23) \quad \delta(t) = \frac{d}{dt} \sigma(t)$$

im Sinne der Ableitung von Distributionen. Nachdem der Begriff des Signals geklärt wurde, kann man sich mit der Abtastung desselben beschäftigen.

Bemerkung: Im weiteren werden Zeit und Ort als äquivalent angesehen. Dies wird durch die Abtastung mit konstanter Geschwindigkeit motiviert.

Satz: *Abtasttheorem von Shannon*

Sei $s(x)$ ein kontinuierliches und bandbegrenzt Signal ($S(f) = 0 \forall |f| > B$) auf ganz \mathbf{R} , dann kann $s(x)$ vollständig durch die Abtastwerte s_k rekonstruiert werden, wenn für den maximalen Abstand Δx zwischen zwei Abtastungen gilt

$$(2.24) \quad \Delta x \leq \frac{1}{2 \cdot B}.$$

Beweis: Sei $\Delta x = 1 / (2B)$ für alle Abstände, s auf ganz \mathbf{R} erklärt, (gleichmäßig) stetig und absolut integrierbar. Sei

$$\tilde{s}(x) = \sum_{k=-\infty}^{\infty} s(k \cdot \Delta x) \cdot \text{sinc}\left(2\pi \cdot B \cdot \left(x - \frac{k}{2B}\right)\right) \quad (*) ; \text{sinc}(x) := \frac{\sin(x)}{x}, \text{ dann folgt mit}$$

$$s(k \cdot \Delta x) = \int_{-\infty}^{\infty} s(x) \cdot \delta\left(x - \frac{k}{2B}\right) dx$$

die Gleichung

$$\sum_{k=-\infty}^{\infty} s(k \cdot \Delta x) \cdot \text{sinc}\left(2\pi \cdot B \cdot \left(x - \frac{k}{2B}\right)\right) = 2B \cdot \text{sinc}(2\pi Bx) * \left[\frac{1}{2B} \sum_{k=-\infty}^{\infty} s(k \cdot \Delta x) \delta\left(x - \frac{k}{2B}\right)\right].$$

Durch den Übergang in den Frequenzbereich erhält man

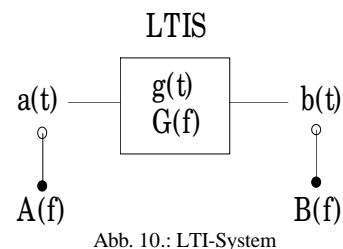
$$\mathcal{F}((*)) = \text{rect}\left(\frac{f+B}{2B}\right) \cdot \sum_{l=-\infty}^{\infty} S(f-2Bl); \text{rect}(x) = \begin{cases} 1; & x \in [0,1] \\ 0; & \text{sonst} \end{cases}.$$

Offensichtlich gilt $S(f-2Bl) = S(f)$ genau für $l=0$. Daraus folgt unmittelbar $\tilde{s}(x) = s(x)$ ([Jaro90]). Das Signal s ist also aus den Abtastwerten vollständig rekonstruierbar. \square

Eine Abtastung mit $\Delta x < 1 / (2B)$ heißt *oversampling*, mit $\Delta x > 1 / (2B)$ *undersampling*. Letzteres führt genau auf den aliasing-Effekt, das heißt höherfrequente Signalanteile sind in einer Alias-Frequenz enthalten und können nicht rekonstruiert werden. In den meisten Fällen betrachtet man das Signal s auf einem endlichen Intervall $[a, b]$. Damit kann keine Bandbegrenzung mehr garantiert werden und eine vollständige Rekonstruktion ist nicht möglich. Allerdings kann man für Signale mit $|S(\omega)| \leq \beta \cdot (1 + |\omega|)^{-(1+\mu)}$ (β und μ positive Konstanten) ein Δx so angeben, daß der Fehler bei der Interpolation mit *sinc* vernachlässigt werden kann ([Briggs95]). Der Fehler bei der Rekonstruktion mit endlich vielen Werten (truncation error) läßt sich ebenfalls angeben ([Bab87]).

2.4.2. Lineare Systeme

Ein lineares zeitinvariantes System (LTI-System) wird durch seine Impulsantwort oder seine Übertragungsfunktion beschrieben. Es wirkt auf ein Eingangssignal $a(t)$ vermöge $b(t) = a(t) * g(t)$ beziehungsweise $B(f) = A(f) \cdot G(f)$. Die Impulsantwort $g(t)$ ergibt sich aus $a(t) = \delta(t)$. Es folgt dann $b(t) = g(t)$. Die Bedingung für die ideale Übertragung durch ein LTI-System lautet $|G(f)| = 1$ und die für verzerrungsfreie Übertragung $|G(f)| = \text{const.}$



2.4.3. Zufällige Signale

Def.: Ein zufälliges Signal $s(t)$ ist die Realisierung einer zeit-/ortsabhängigen Zufallsgröße $S(t)$.

Zur Charakterisierung zufälliger Signale verwendet man die Mittelwerte

$$(2.25) \quad \bar{s} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} s(t) dt, \quad \overline{s^2} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} (s(t))^2 dt,$$

den Effektivwert $S_{\text{eff}} = \sqrt{\overline{s^2}}$, die Streuung $\delta^2 = \overline{s^2} - (\bar{s})^2$ sowie die Autokorrelationsfunktion

$$(2.26) \quad R_{ss}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T s(\tau) \cdot s(t - \tau) dt.$$

Bei stochastischen (=zufälligen) Signalen kann man, da der Zeitverlauf $s(t)$ nicht bekannt ist, keine Spektraldichte $S(f)$ angeben, jedoch kann man die spektrale Leistungsdichte

$$(2.27) \quad L_S(f) = \lim_{\Delta f \rightarrow \infty} \frac{\Delta P(f)}{\Delta f}$$

durch Messung der Leistung $\Delta P(f)$ des Signals im Frequenzbereich $[f, f + \Delta f]$ bestimmen. Der nachfolgende Satz beschreibt den Zusammenhang zwischen der spektralen Leistungsdichte und der Autokorrelationsfunktion eines zufälligen Signals.

Satz: Wiener-Tschintschin-Theorem

Sei $s(t)$ ein stochastisches Signal mit der spektralen Leistungsdichte L_S , dann gilt

$$(2.28) \quad L_S(f) = \mathcal{F}(R_{ss}(\tau)).$$

Die Wirkung eines LTI-Systems wird durch $L_b(f) = |G(f)|^2 \cdot L_a(f)$ beschrieben.

Beispiel: Weißes Rauschen ist ein stochastisches Signal, dessen Amplitudenverteilung unabhängig von der Frequenz ist. Die Amplitudendichte ist normalverteilt mit $f(A) = N(\mu, \sigma)$. Nach dem Wiener-Tschintschin-Theorem gilt bei konstanter Leistungsdichte: $R_{ss}(\tau) = 0 \quad \forall \tau \neq 0$.