

„Natura non facit saltus“ (Die Natur macht keine Sprünge). Schon seit den alten Griechen gingen die Philosophen und Naturwissenschaftler davon aus, dass in der Natur beobachtbare Vorgänge und messbare Größen keine sprunghaften Änderungen aufweisen, sondern von Kontinuität und Stetigkeit geprägt sind. Bis zum Beginn des vergangenen Jahrhunderts hatte dieses Axiom unangefochtene Gültigkeit. Erst die relativistische Quantenmechanik stellte es in Frage (Quantensprung).

Allgemeines

Für die klassische Physik und Elektrotechnik sowie die von uns durch unsere Sinnesorgane erlebte Umwelt bewahrt sich das Stetigkeitsaxiom immer wieder: Temperaturen, Drücke, Wasserstände, Ladungen, Spannungen, Ströme ... – allgemein formuliert „Signale“ – verändern sich in endlichen Zeitabschnitten endlich. Diese Eigenart ihres Verhaltens macht es möglich, den zeitlichen Verlauf durch eine genügend dichte Folge von Einzelwerten zu beschreiben und die Zwischenwerte durch Interpolation zu gewinnen. Man erhält die Folgewerte durch regelmäßige Messung des Signals, was auch als „Abtasten“ bezeichnet wird. Der zeitliche Abstand der Folgewerte darf nur nicht so groß werden, dass relevante Signaländerungen zwischen zwei Folgewerten liegen und somit nicht erkannt werden können. Hier gilt das Prinzip „so selten wie möglich, so häufig wie nötig“.

Am Beispiel von Messungen des Wasserstandes (Pegel) von Gewässern wollen wir das verdeutlichen. Diese im Sekundenabstand vorzunehmen, ist unsinnig, weil sich der Pegel in dieser kurzen Zeit nur unwe-

sentlich ändert und riesige Messwertetabellen entstünden. Andererseits sind wöchentliche Messungen zu wenig, da zwischen ihnen die Überschwemmung ihren Lauf genommen haben könnte. Wir brauchen also eine an die Veränderungsgeschwindigkeit der Messgröße angepasste Messrate: Schnell veränderliche Größen müssen häufig gemessen werden, trägere weniger häufig, damit aus der Folge der Messwerte ein verlässliches Bild vom Verlauf der Messgröße zu gewinnen ist. Genau diese Erkenntnis schlägt sich im später erklärten Abtasttheorem der Nachrichtentechnik nieder.

Analoge Signale

Das elementarste Analog-Signal ist eine sinusförmige Funktion der Zeit. Deren Amplitude $s(t)$ hängt mit der Frequenz f , der Zeit t , dem Phasenwinkel φ und der Spitzenamplitude \hat{s} wie folgt zusammen:

$$s(t) = \hat{s} \cdot \sin(2\pi f t + \varphi) \quad (1)$$

Mit der Definition der Kreisfrequenz

$$\omega = 2\pi f \quad \text{Kreisfrequenz} \quad (2)$$

ergibt sich:

$$s(t) = \hat{s} \cdot \sin(\omega t + \varphi) \quad (3)$$

Phasenverschobene Sinusschwingung

Gleichung 3 entspricht dem in Abbildung 1 dargestellten Kurvenverlauf.

Wir merken uns an dieser Stelle, dass ein positiver Phasenwinkel die Kurve nach links zu früheren Zeiten verschiebt. Man sagt auch, die Kurve mit dem positiven Phasenwinkel im Argument eilt der ohne Phasenwinkel voraus (voreilend). Das ist leicht verständlich, wenn wir uns einen bestimmten Punkt der Kurve anschauen, z. B. den Nullpunkt vor der positiven Halbwelle. Hier ist wegen $\sin(0)=0$ das Argument $\omega t + \varphi = 0$, woraus $\omega t = -\varphi$ folgt. Negative Phasenwinkel verschieben die Kurve entsprechend nach rechts, d. h. zu späteren Zeitpunkten hin (nacheilend).

Die Frequenz f beschreibt, wie viele Schwingungen in einer Sekunde stattfinden. Ihre Dimension ist demnach $1/s$ oder Hz (Hertz). Die Dauer für eine ganze Schwingung ist der Kehrwert ihrer Frequenz und wird als Periodendauer T mit der Einheit s bezeichnet. Gleichung 4 beschreibt den Zusammenhang zwischen Periodendauer und Frequenz, Gleichung 5 zwischen Periodendauer und Kreisfrequenz:

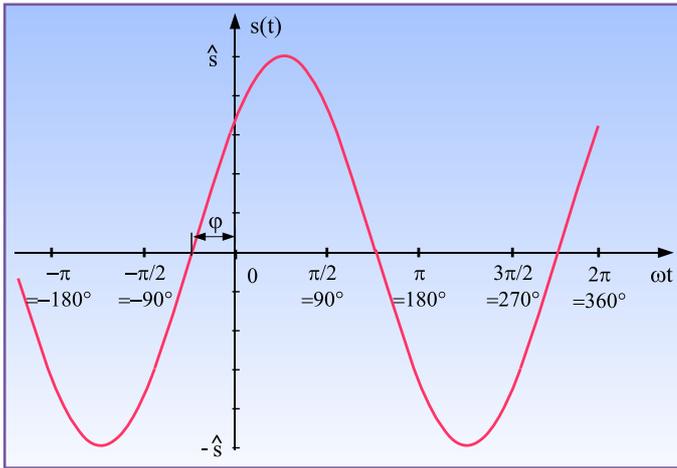


Bild 1: Die grafische Darstellung des zeitlichen Verlaufs einer um den Phasenwinkel φ voreilenden Sinusschwingung mit der Kreisfrequenz ω und der Amplitude \hat{s}

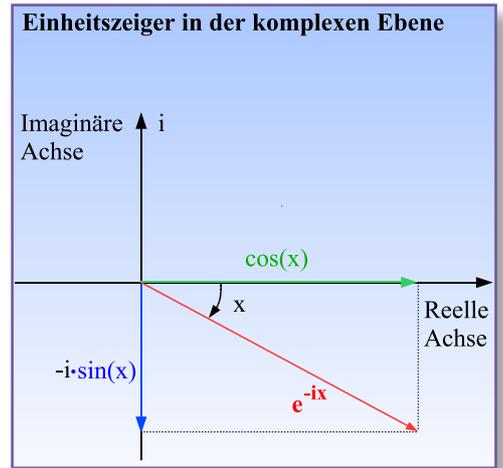


Bild 3: Die Darstellung aus Bild 2 mit umgekehrtem Vorzeichen des Exponenten

$$f = \frac{1}{T} \quad \text{Frequenz und Periodendauer} \quad (4)$$

bzw.

$$\omega = 2\pi f = \frac{2\pi}{T} \quad \text{Kreisfrequenz und Periodendauer} \quad (5)$$

Aus rechnerischen Gründen ist die komplexe Darstellung des Sinus oder Kosinus oft sehr vorteilhaft. Dazu verwendet man die Eulersche Identität:

$$e^{ix} = \cos x + i \sin x \quad \text{Eulersche Identität} \quad (6)$$

Ersetzt man x durch $-x$ folgt:

$$e^{-ix} = \cos x - i \sin x \quad (7)$$

i wird als imaginäre Einheit bezeichnet. Sie ist definiert als Quadratwurzel aus -1 . Dementsprechend gilt:

$$i = \sqrt{-1} \quad \text{Imaginäre Einheit} \quad (8)$$

und

$$i^2 = -1 \quad (9)$$

Man kann die Gleichungen 6 und 7 auch grafisch als Einheitszeiger mit der Länge 1 in der komplexen Ebene darstellen. Abbildung 2 ist die grafische Entsprechung von Gleichung 6.

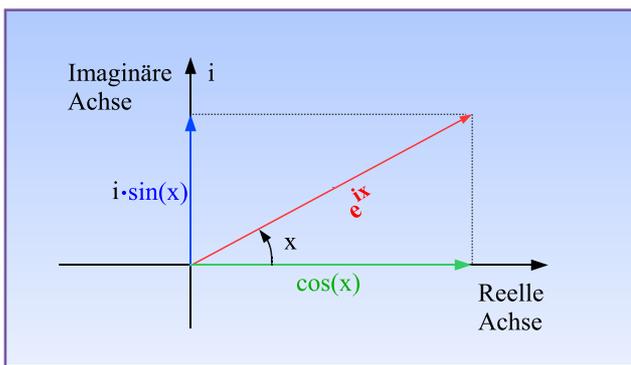
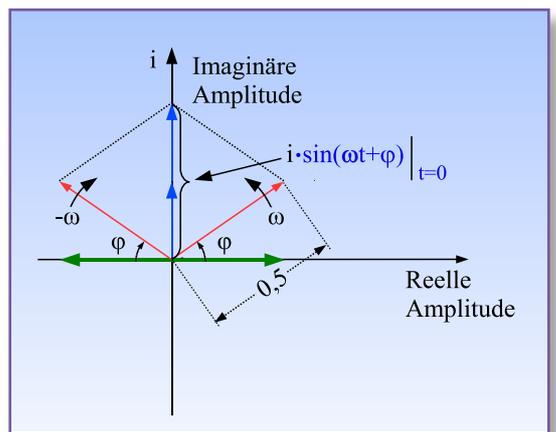


Bild 2: Die Eulersche Identität als Einheitszeiger in der komplexen Ebene

Bild 4: Die Sinusschwingung als Projektion zweier gegenläufiger, zur imaginären Achse symmetrischer Zeiger



Wegen des vertauschten Vorzeichens des Imaginärteils des zu (6) gehörigen Zeigers erscheint der mit Gleichung 7 korrespondierende Zeiger in Abbildung 3 an der reellen Achse gespiegelt. Auf $s(t)$ angewandt, folgt:

$$\hat{s} \cdot e^{i(\omega t + \varphi)} = \quad (10)$$

$$\hat{s} \cdot \cos(\omega t + \varphi) + i \hat{s} \cdot \sin(\omega t + \varphi) = \underbrace{s(t)}_{\text{Re} \left\{ \hat{s} \cdot e^{i(\omega t + \varphi)} \right\} + i \text{Im} \left\{ \hat{s} \cdot e^{i(\omega t + \varphi)} \right\}}$$

Man sieht, dass die reelle Zeitfunktion $s(t)$ gleich dem Imaginärteil des komplexen Ausdrucks ist.

Mit konjugierten komplexen Exponentialausdrücken (Exponenten mit umgekehrten Vorzeichen) kann man $s(t)$ auch wie folgt schreiben:

$$s(t) = \hat{s} \sin(\omega t + \varphi) = \frac{\hat{s} e^{i(\omega t + \varphi)} - \hat{s} e^{-i(\omega t + \varphi)}}{2i} \quad (11)$$

Hat $s(t)$ einen kosinusförmigen Verlauf, gilt:

$$s(t) = \hat{s} \cos(\omega t + \varphi) = \frac{\hat{s} e^{i(\omega t + \varphi)} + \hat{s} e^{-i(\omega t + \varphi)}}{2} \quad (12)$$

Zeigerdarstellung

An dieser Stelle sollte man sich die Darstellung einer phasenverschobenen Sinus- bzw. Kosinusschwingung mit der Spitzenamplitude 1 durch zwei in der komplexen Ebene gegenläufig mit der Frequenz ω rotierende Zeiger mit der halben Länge der Spitzenamplitude grafisch verdeutlichen. Abbildung 4 zeigt dies für die Sinusschwingung.

Zum Zeitpunkt $t = 0$ ist die Lage der beiden Zeiger der Länge 0,5 nur durch den Phasenwinkel φ bestimmt. Wir erkennen, dass sich die Projektionen auf die reelle Achse (\cos) aufheben und die auf die imaginäre Achse (\sin) addieren. Weil die rotierenden Zeiger die Länge 0,5 haben, entspricht die Summenlänge ihrer Projektionen (vektorielle Addition) auf die imaginäre Achse der Projektion eines einzelnen Zeigers der Länge 1. Die Kosinusschwingung illustriert Abbildung 5, ebenfalls zu $t = 0$. Hier heben sich die Projektionen auf die imaginäre Achse (\sin) auf und jene auf die reelle Achse (\cos) addieren sich.

Das Gesagte gilt natürlich auch für jeden beliebigen Zeitpunkt $t > 0$, wo dann das Argument der e -Funktion bzw. des Sinus/Kosinus entsprechend größer ist. Ein voller Umlauf der Zeiger um 2π (das Bogenmaß 2π entspricht 360° im Winkelmaß) beschreibt eine Periode der Schwingung. Die Kurve hat nach $t = 2\pi/\omega = T$ eine Pe-

Tabelle 1: Einige wichtige Eigenschaften der Fouriertransformation

Linearität	$a \cdot f(t) + b \cdot g(t) \iff a \cdot F(i\omega) + b \cdot G(i\omega)$
Zeitverschiebung	$f(t-t_0) \iff F(i\omega) \cdot e^{-i\omega t_0}$
Frequenzverschiebung	$f(t) \cdot e^{i\omega_0 t} \iff F(i\omega - i\omega_0)$
Skalierung	$f(at) \iff \frac{1}{ a } \cdot F\left(\frac{i\omega}{a}\right)$
Faltung im Zeitbereich	$f(t) * g(t) \iff F(i\omega) \cdot G(i\omega)$
Faltung im Frequenzbereich	$f(t) \cdot g(t) \iff F(i\omega) * G(i\omega)$
Differentiation im Zeitbereich	$\frac{d^n f(t)}{dt^n} \iff (i\omega)^n \cdot F(i\omega)$
Differentiation im Freq.bereich	$(-i \cdot t)^n \cdot f(t) \iff \frac{d^n F(i\omega)}{d\omega^n}$
Symmetrie	$f(t) \iff F(i\omega)$ $F(t) \iff 2\pi \cdot f(-i\omega)$

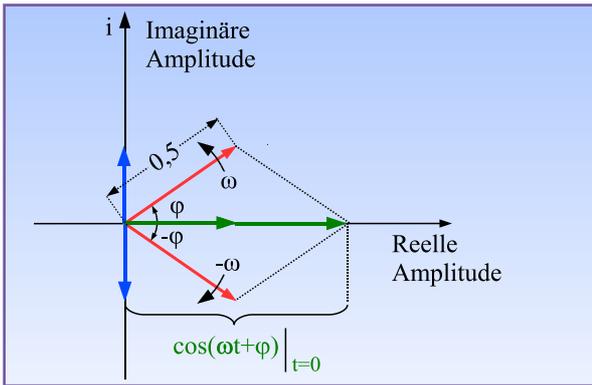


Bild 5: Die Kosinusschwingung als Projektion zweier gegenüberliegender, zur reellen Achse symmetrischer Zeiger

riode durchlaufen und den gleichen Wert wie nach $t = nT$ für alle ganzzahligen Werte von n . Man sagt, auch der Sinus sowie der Kosinus sind periodisch in 2π .

Eine allgemeingültige Formulierung für eine in T periodische Zeitfunktion lautet:

$$f(t) = f(t+T) \quad -\infty < t < \infty \quad (13)$$

Periodizität

Gerade und ungerade Funktionen

Es ist für die Rechenpraxis ungemein praktisch, die Symmetrieeigenschaften von Funktionen zu kennen. Wenn eine Funktion spiegelsymmetrisch zur Ordinatenachse ist, spricht man von einer geraden Funktion. In der Formelsprache ausgedrückt, ist eine Funktion gerade, wenn gilt:

$$f(x) = f(-x) \quad \text{Gerade Funktion} \quad (14)$$

Ungerade Funktionen sind punktsymmetrisch zum Ursprung. Für sie gilt:

$$f(x) = -f(-x) \quad \text{Ungerade Funktion} \quad (15)$$

Abbildung 6 demonstriert die Symmetrieeigenschaften.

Harmonische Analyse periodischer Signale

Periodische Signale bestehen aus gleichen Abschnitten, die sich alle T Sekunden wiederholen und dies von jeher bis in

alle Ewigkeit. Diese ewig währende Periodizität ist natürlich bei realen Signalen nie gegeben, da diese irgendwann einmal eingeschaltet und auch wieder ausgeschaltet werden. Aber je mehr Periodenlängen aneinander gereiht sind, umso mehr nähert man sich dem Idealzustand mit einer unendlichen Zahl von Perioden. Jedes periodische Signal lässt sich als unendliche Summe gewichteter sinusförmiger Schwingungen darstellen. Im „ELVjournal“ 4/06 wurde in dem Artikel „Oszilloskop und Spektrumanalyzer“ die Zerlegung eines periodischen Signals in seine Harmonischen beschrieben. Im Gegensatz zu der dort verwendeten Darstellung benutzen wir hier die komplexe Notation:

$$f(t) = \sum_{n=-\infty}^{+\infty} F_n \cdot e^{in\omega t} \quad (16)$$

Komplexe Fourierreihe

Die im Allgemeinen komplexen Fourierkoeffizienten F_n berechnet man über:

$$F_n = \frac{1}{T} \int_{t_0}^{t_0+T} f(t) \cdot e^{-in\omega t} dt \quad (17)$$

Komplexe Fourierkoeffizienten

Die Integration muss über eine Periode des Signals erfolgen, wobei t_0 als untere Integrationsgrenze beliebig gewählt werden kann, am besten so, dass sich die Rechnung möglichst einfach gestaltet. Wir wollen an dieser Stelle festhalten, dass jedes

periodische Signal aus einer Grundschwingung und unendlich vielen Harmonischen (Oberschwingungen mit Frequenzen, die Vielfache der Frequenz der Grundschwingung sind) besteht.

Harmonische Analyse nicht-periodischer Signale

Auch Signale, die nicht periodisch sind (diese nennt man auch aperiodisch), kann man aus Sinuskomponenten bilden. Dazu dienen die Fourierintegrale:

$$F(i\omega) = \int_{t=-\infty}^{+\infty} f(t) \cdot e^{-i\omega t} dt \quad (18)$$

Fourier-Transformierte

$$f(t) = \frac{1}{2\pi} \int_{\omega=-\infty}^{+\infty} F(i\omega) \cdot e^{i\omega t} d\omega \quad (19)$$

Fourier-Rücktransformierte

$F(i\omega)$ ist das komplexe Spektrum (oder genauer formuliert: die spektrale Dichtefunktion) der aperiodischen Zeitfunktion $f(t)$. Häufig ist die folgende symbolische Schreibweise für den Fourierintegralzusammenhang von Zeitfunktion und Spektrum anzutreffen:

$$f(t) \iff F(i\omega)$$

Die Auswertung der Integrale (18) und (19) zur Ermittlung des Frequenzspektrums einer Zeitfunktion bzw. umgekehrt kann schwierig sein. Deshalb benutzt man in der Praxis Korrespondenztabelle, in denen die meisten Transformationen bereits zusammengestellt sind.

Unter Berücksichtigung der wichtigsten Eigenschaften der Fouriertransformation (Tabelle 1) kann man auf diese Weise viele Fourierintegrale durch Analogieüberlegungen „lösen“.

Im nächsten Teil dieser Serie werden die Fourierkorrespondenzen wichtiger Funktionen und die wichtigsten allgemeinen Eigenschaften der Fouriertransformation

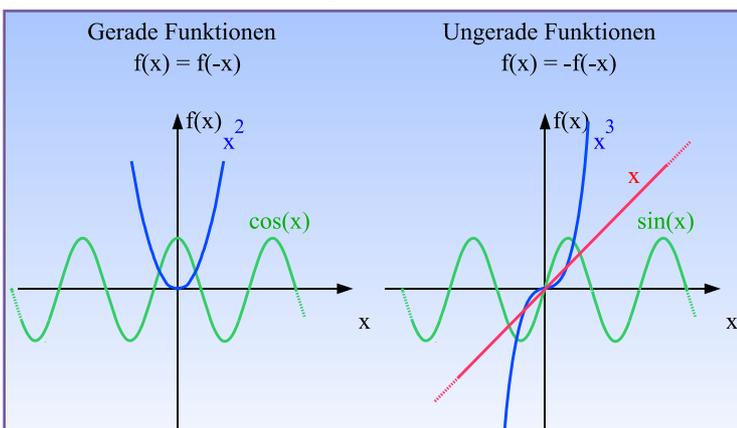
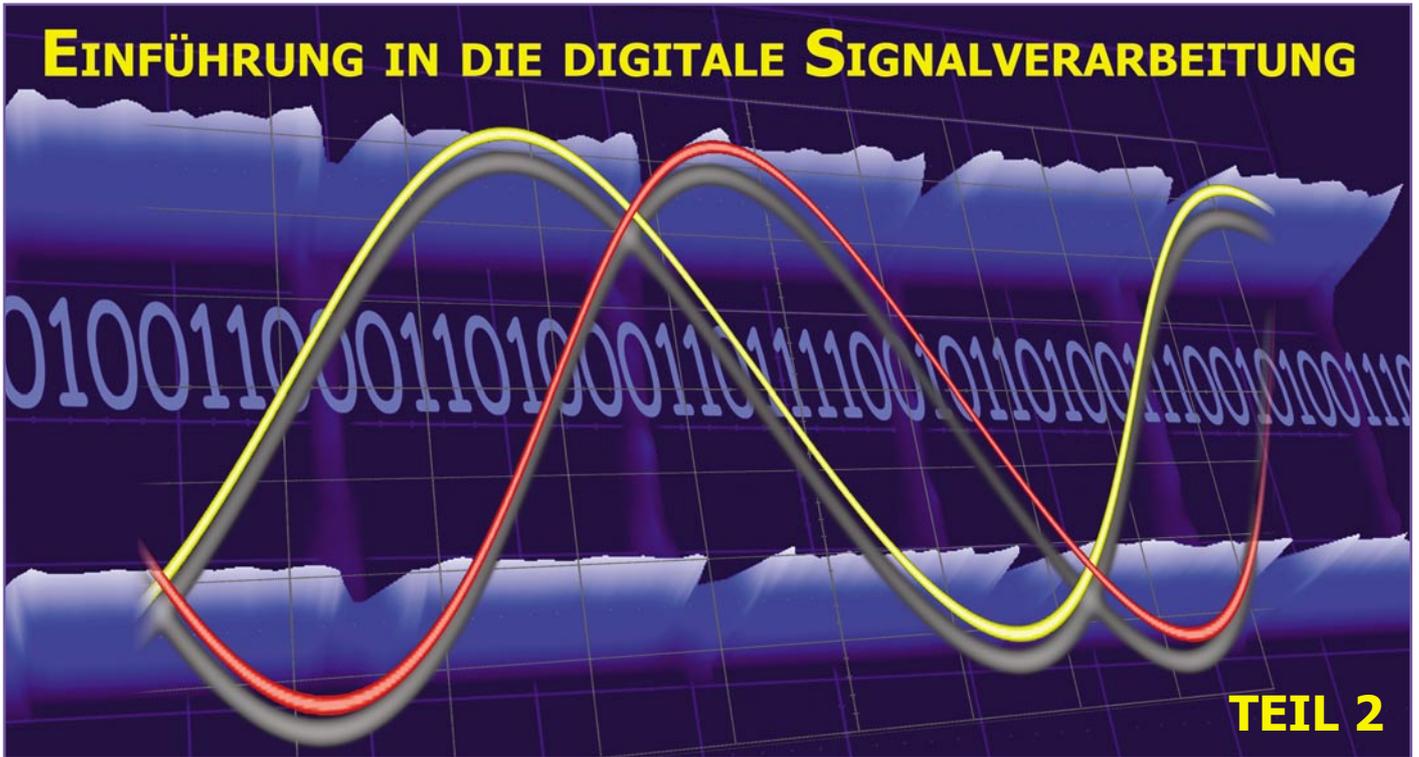


Bild 6: Gerade Funktionen sind spiegelsymmetrisch zur Ordinate (y-Achse), ungerade punktsymmetrisch zum Ursprung der xy-Ebene.



Wir beschäftigen uns eingehend mit der Fourieranalyse als mathematischem Hilfsmittel, um zwischen der Zeit- und Frequenzdarstellung (Korrespondenzen) allgemeiner Signale zu wechseln. Dabei werden die wichtigsten Eigenschaften der Fouriertransformation und der Zusammenhang zwischen periodischen und nicht-periodischen Signalen beleuchtet. Mit diesem Rüstzeug können wir in die Welt der digitalen Signale eintreten.

Fourierkorrespondenzen

Abbildung 7 zeigt drei elementare Fourierkorrespondenzen, die wir uns etwas genauer anschauen wollen.

Die erste nicht-periodische Zeitfunktion stellt den so genannten δ -Impuls (u. a. auch δ -Funktion, Dirac-Stoß oder δ -Distribution, Einheitsstoß genannt) dar. Er hat die Eigenschaft, nur an einer Stelle (dort, wo sein Argument null ist) mit einer

unendlichen Amplitude zu existieren.

$$f(t) = \delta(t) = \begin{cases} 0 & |t| > 0 \\ \infty & t = 0 \end{cases} \quad \delta\text{-Impuls} \quad (20)$$

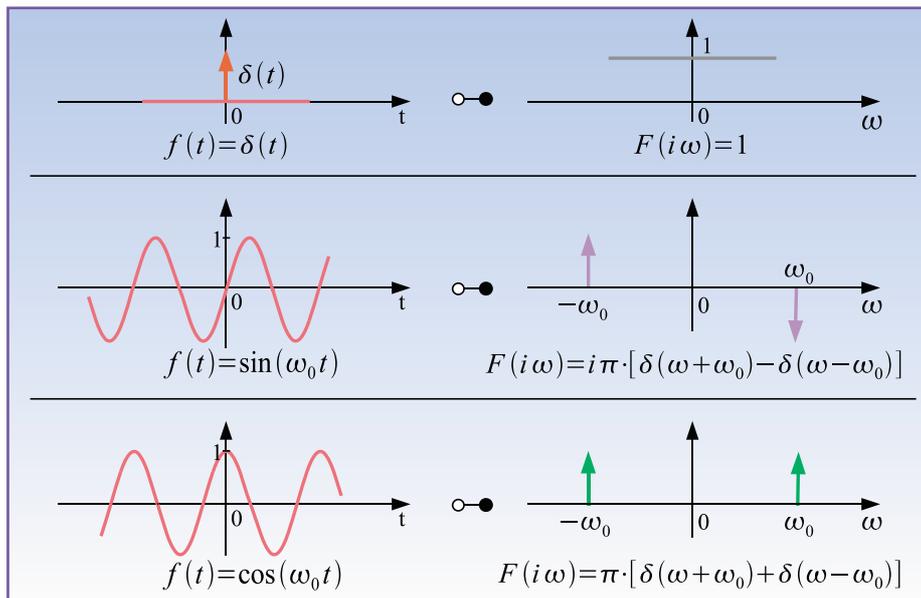


Bild 7: Drei elementare Fourierkorrespondenzen. Von oben: Dirac-Impuls, Sinusschwingung und Kosinusschwingung.

Eine solche unendlich schmale und unendlich hohe „klassische Funktion“ gibt es nicht. Der δ -Impuls ist nur als so genannte Distribution mathematisch sauber definierbar. Distributionen sind „verallgemeinerte Funktionen“, die vom französischen Mathematiker Laurent Schwartz (1915–2002) in der Mitte des vorigen Jahrhunderts theoretisch untermauert wurden (Distributionentheorie). Als Hilfsmittel finden Distributionen vielfältige Anwendungen, z. B. in der Theorie der gewöhnlichen und partiellen Differentialgleichungen, der Fourier-Analyse, der Physik oder der Elektrotechnik. Der δ -Impuls ist als Distribution über seine Ausblendeigenschaft gemäß Gleichung (22) definiert. Obwohl der δ -Impuls keine klassische Funktion ist, lässt er sich dennoch auch ohne tiefere Kenntnisse der Theorie der Distributionen äußerst fruchtbar in der Systemtheorie einsetzen.

Die Entstehung des δ -Impulses kann man sich als Grenzübergang eines Recht-

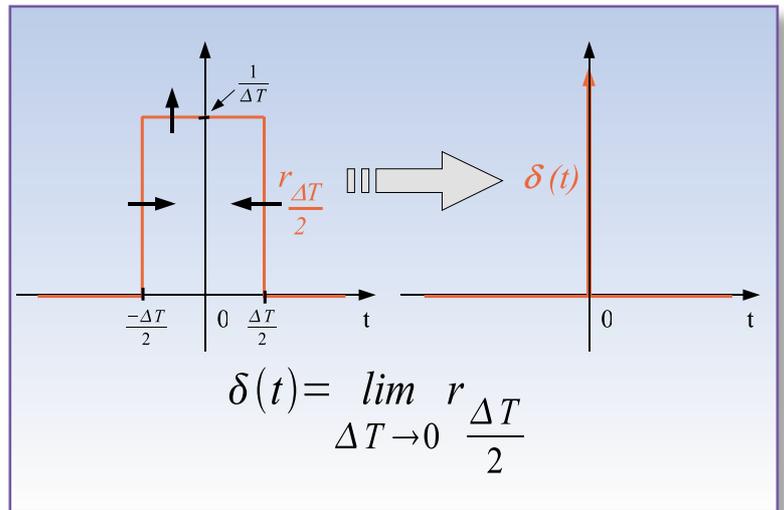
ecks der Breite ΔT und der Höhe $1/\Delta T$ für $\Delta T \rightarrow 0$ vorstellen (Abbildung 8). Je kleiner nun ΔT wird, umso schmaler, aber auch umso höher wird das Rechteck, behält also seinen Flächeninhalt 1 bei. Nach vollzogenem Grenzübergang ist der δ -Impuls unendlich schmal und unendlich hoch. Integriert man über ihn hinweg, erhält man seine Fläche zu 1. Es gilt also:

$$\int_{t=-\infty}^{+\infty} \delta(t) dt = 1 \quad (21)$$

Das Spektrum des δ -Impulses hat bei allen Frequenzen die Amplitude 1. Man kann sich das auch veranschaulichen, indem man die Fouriertransformierte eines Rechteckimpulses der Breite ΔT und der Höhe $1/\Delta T$ betrachtet. Sie ist eine si-Funktion mit den Nullstellen bei Vielfachen von $1/\Delta T$ und einer Amplitude der Hauptkeule von 1. Je schmaler das Rechteck wird, umso weiter wandern die Nulldurchgänge vom Ursprung der Frequenzachse weg. Im Grenzübergang wird die si-Funktion zu einer Parallelen im Abstand 1 zur ω -Achse gespreizt.

Mit anderen Worten: Ein unendlich kurzes Ereignis im Zeitbereich hat ein konstantes Spektrum im Frequenzbereich zur Folge. Allgemeiner formuliert: Je kürzer das Zeitereignis, umso breiter seine Entsprechung im Frequenzbereich. Dies ist nichts anderes als das Zeit-Bandbreite-Gesetz der Nachrichtentechnik. Etwas Vergleichbares können wir bei einem Schwingkreis beobachten. Ist er schmal-

Bild 8: Der Dirac-Impuls als Grenzübergang eines Rechteckimpulses mit der Fläche 1 (Breite ΔT x Höhe $1/\Delta T$) für $\Delta T \rightarrow 0$



bandig (verlustarm), wirkt eine Anregung als langsam abklingende Schwingung nach. Der breitbandige (verlustreiche) Schwingkreis schwingt dagegen nur kurz nach.

Eine besonders wichtige Eigenschaft des δ -Impulses für die Beschreibung eines nachrichtentechnischen Übertragungssystems ist seine Ausblendeigenschaft.

$$f(t) = \int_{\tau=-\infty}^{+\infty} f(\tau) \delta(t-\tau) d\tau \quad (22)$$

Ausblendeigenschaft

Die Erklärung: Weil $\delta(t-\tau)$ nur für das Argument $\tau = t$ ungleich null ist, lässt sich das Integral auch schreiben als:

$$f(t) = f(t) \cdot \int_{\tau=-\infty}^{+\infty} \delta(t-\tau) d\tau \quad (23)$$

Das Integral über den δ -Impuls ist definitionsgemäß 1, was (23) zur Identität macht und so die Gültigkeit von (22) bestätigt.

Die zweite Fourierkorrespondenz in Abbildung 7 betrifft die periodische Zeitfunktion einer immerwährenden Sinus-schwingung (sozusagen die Mutter aller periodischen Funktionen) mit der Frequenz ω_0 . Ihre Fouriertransformierte ist rein imaginär (violette Farbdarstellung) und besteht aus zwei δ -Funktionen, eine mit positivem Vorzeichen bei $-\omega_0$ und eine mit negativem Vorzeichen bei $+\omega_0$.

Die dritte Fourierkorrespondenz in Abbildung 7 zeigt das Transformationspaar einer periodischen Kosinusschwingung mit der Frequenz ω_0 . Deren Fouriertransformierte ist rein reell (grüne Farbdarstellung) und besteht ebenfalls aus zwei δ -Funktionen, beide mit positivem Vorzeichen bei $-\omega_0$ und $+\omega_0$.

Bereits jetzt erkennen wir zwei wichtige Eigenschaften der Fouriertransformation:

Satz 1: Gerade Zeitfunktionen haben ein reelles, gerades Spektrum.

Satz 2: Ungerade Zeitfunktionen haben ein imaginäres, ungerades Spektrum.

Zeitfunktionen, die weder gerade noch ungerade sind, haben ein komplexes Spektrum, das natürlich auch weder gerade noch ungerade ist. Das folgt unmittelbar aus der Tatsache, dass sich jede Zeitfunktion in eine Summe aus einem geraden und einem ungeraden Summanden zerlegen lässt, deren Spektren sich nach Satz 1 als reell und nach Satz 2 als imaginär ergeben. Beide ergänzen sich zu einem komplexen Spektrum. Das eben Gesagte wird in Abbildung 9 beispielhaft illustriert.

Wir wollen das mathematisch nachvollziehen. Die obere Zeitfunktion in Abbildung 9 ist eine abklingende Exponentialfunktion, deren Werte für negative Zeiten ($t < 0$) durch Multiplikation mit der so genannten Sprungfunktion $u(t)$ (0 für $t < 0$ und 1 für $t > 0$) auf null gesetzt werden.

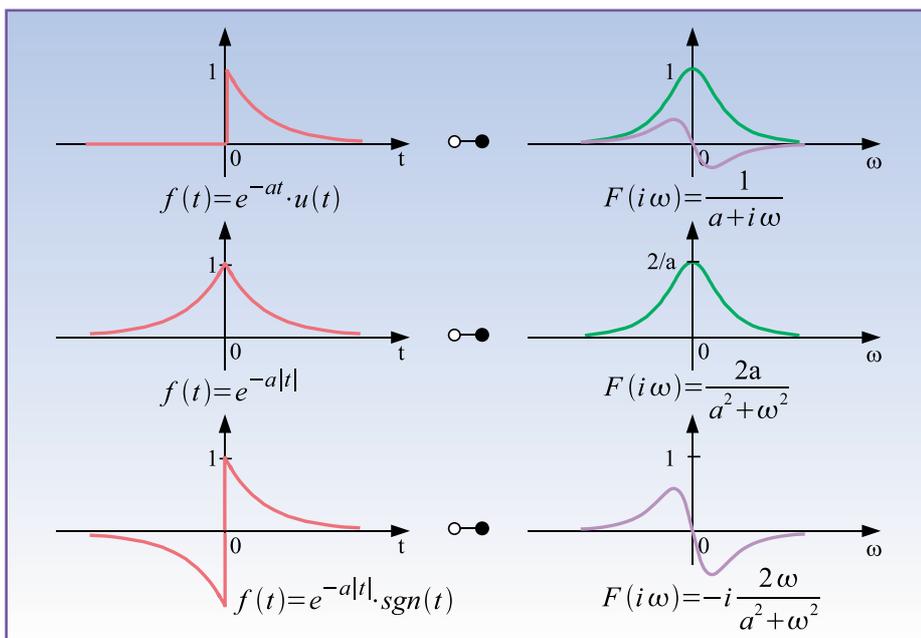


Bild 9: Die Fouriertransformierte einer beliebigen, in einen geraden und einen ungeraden Anteil zerlegbaren Funktion ist wegen der Linearitätseigenschaft der Fouriertransformation gleich der Summe der Fouriertransformierten dieser Anteile. Am Beispiel der abklingenden e-Funktion (oben), ihrem geraden Anteil (Mitte) und ihrem ungeraden Anteil (unten) kann man sich das verdeutlichen.

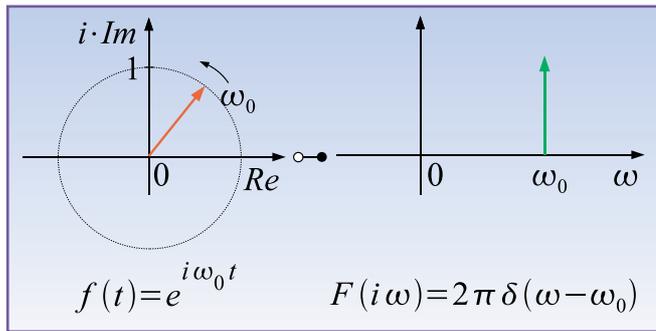


Bild 10: Der mit ω rotierende Einheitsdrehzeiger in der komplexen Ebene (links) hat nur eine Spektrallinie bei ω (rechts) als Korrespondenz im Frequenzbereich.

$$f(t) = e^{-at} \cdot u(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2} & t = 0 \\ e^{-at} & t > 0 \end{cases} \quad (24)$$

Sie hat die rechts in Abbildung 9 für $a = 1$ abgebildete komplexe Fouriertransformierte:

$$F(i\omega) = \frac{1}{a + i\omega} = \frac{a - i\omega}{(a + i\omega)(a - i\omega)} \quad (25)$$

$$= \frac{a}{a^2 + \omega^2} + i \frac{-\omega}{a^2 + \omega^2}$$

Realteil $F(i\omega)$ Imaginärteil $F(i\omega)$

Der Realteil ist der grüne gerade, der Imaginärteil der violette ungerade Kurvenzug. Nun kann man die obere Zeitfunktion (24) auch als halbierte Summe der mittleren und unteren Zeitfunktionen in Abbildung 9 darstellen, deren jeweilige Korrespondenzen rechts stehen. Die Summe dieser Korrespondenzen muss wegen der Linearitätseigenschaft der Fouriertransformation (siehe Tabelle 1 im Teil 1) also wieder (25) ergeben. Wir prüfen das nach:

$$F(i\omega) = \frac{1}{2} \left(\frac{2a}{a^2 + \omega^2} + i \frac{-2\omega}{a^2 + \omega^2} \right) \quad (26)$$

Offensichtlich führt die Zerlegung der Zeitfunktion in ihren geraden und ungeraden Anteil zu je einem achsen- und einem punktsymmetrischen Spektrum, deren Summe identisch mit dem der ursprünglichen Zeitfunktion ist.

Aber Zeitfunktionen müssen nicht immer

$$F(i\omega) = \int_{t=-\infty}^{+\infty} f(t) \cdot e^{-i\omega t} dt = \int_{t=0}^{+\infty} e^{-at} \cdot e^{-i\omega t} dt = \int_{t=0}^{+\infty} e^{-(a+i\omega)t} dt \quad (27)$$

$$F(i\omega) = \frac{1}{-(a+i\omega)} e^{-(a+i\omega)t} \Big|_{t=0}^{\infty} = 0 - \frac{1}{-(a+i\omega)}$$

$$F(i\omega) = \frac{1}{a+i\omega}$$

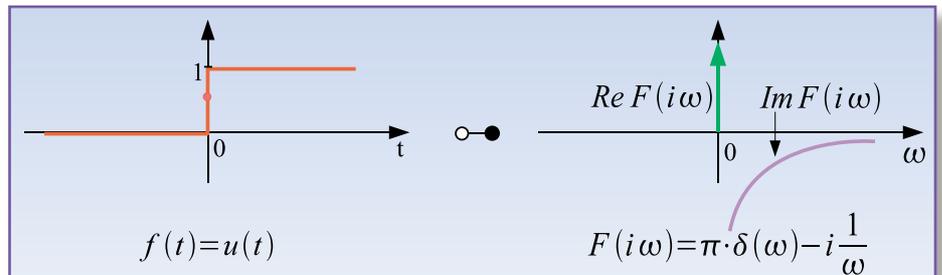


Bild 11: Die Sprungfunktion ist im Bereich negativer Zeiten null und springt bei $t = 0$ auf den Wert 1. Durch Multiplikation mit ihr kann man periodische Zeitfunktionen „einschalten“.

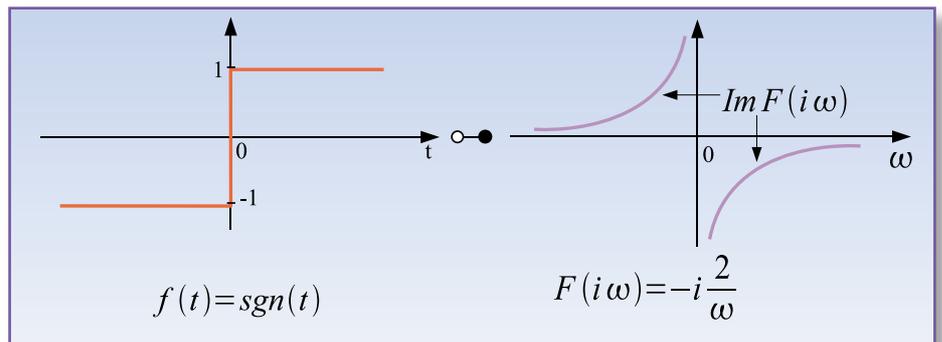


Bild 12: Die Signumfunktion springt bei $t = 0$ von -1 auf $+1$. Die Multiplikation einer Sinus- oder Kosinus-schwingung mit der Signumfunktion bewirkt an der Sprungstelle deren Phasendrehung um 180° .

reell sein. Ein Beispiel dafür ist der komplexe Drehzeiger aus (6) bzw. seine Entsprechung als Zeitfunktion und die zugehörige Korrespondenz in Abbildung 10.

Zwei Rechenbeispiele

Wenn auch nahezu alle denkbaren Transformationspaare Korrespondenztabelle entnommen werden können, wollen wir doch wenigstens an zwei Beispielen die Fouriertransformation „von Hand“ demonstrieren.

Beispiel 1

Zuerst betrachten wir die abklingende e-Funktion entsprechend (24) nach der Definitionsgleichung (18).

Noch einige Worte zum Gang der Rechnung in (27) für alle, die nicht (mehr?) ganz sattelfest beim Integrieren sind. Der Tatsache, dass es sich um eine bei $t = 0$ eingeschaltete abklingende e-Funktion handelt, trägt das Ersetzen der unteren Integralgrenze $t = -\infty$ durch $t = 0$ Rechnung.

Die Stammfunktion für $t = \infty$ ist null, für $t = 0$ ist der e-Term gleich 1 und es verbleibt das Ergebnis.

Beispiel 2

Nehmen wir an, wir wollen eine Kosinus-schwingung (Abbildung 14) – eingeschaltet bei $t = -T/2$ und ausgeschaltet bei $t = +T/2$ – aus dem Zeitbereich in den Frequenzbereich transformieren. Wir setzen die Kosinusfunktion in das Integral und berücksichtigen die Ein- und Ausschaltzeitpunkte in den Integralgrenzen (28). Wir verwenden die Identitäten in (29a) und (29b), setzen Sie in (28) ein und erhalten (30).

Der Integrand besteht jetzt nur noch aus einer Summe einfacher Sinus- und Kosinusfunktionen, die summandenweise integriert wird (31).

Die geraden Kosinusfunktionen heben sich nach dem Einsetzen der Integrationsgrenzen auf und es bleibt die Formel (32).

Damit wäre die in Abbildung 14 darge-

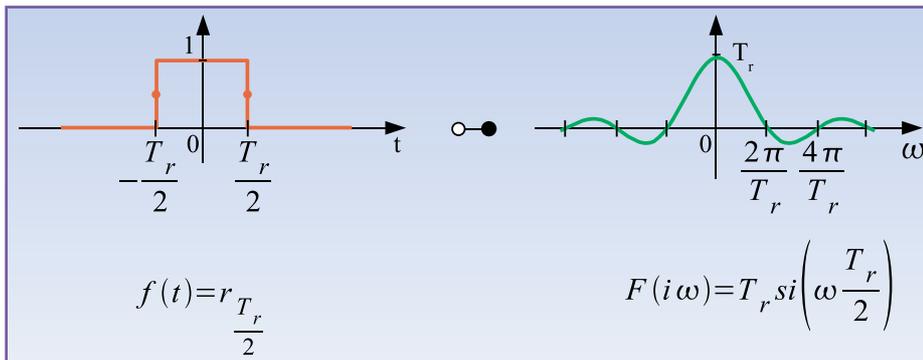


Bild 13: Die Rechteckfunktion kann zum Ausschneiden eines endlichen Abschnitts aus einer unendlich langen Funktion verwendet werden. Man spricht dann auch von einem „Rechteckfenster“.

Bild 14: Hier wurde mit einer Rechteckfunktion ein Abschnitt aus einer Kosinusschwingung erzeugt, sie wurde gewissermaßen ein- und dann ausgeschaltet. Die Dirac-Impulse, die eine unendlich lange Schwingung im Frequenzbereich charakterisieren, verändern sich dadurch zu si-Funktionen.

stellte Korrespondenz abgeleitet.

Es folgt eine kleine Auswahl weiterer Korrespondenzen.

Sprungfunktion

Die Sprungfunktion (Abbildung 11) – auch Einheitssprung genannt – ist in Formel 33 definiert.

Man sieht in Abbildung 11, dass die Fouriertransformierte des Einheitssprungs sowohl einen Real- als auch einen Imaginärteil aufweist. Differenziert man den Einheitssprung, erhält man den δ -Impuls. Bekanntlich liefert die Ableitung einer Funktion ihre Steigung, die beim Einheitssprung nur bei $t=0$ unendlich und überall sonst gleich null ist. Eben das war die Definition des δ -Impulses.

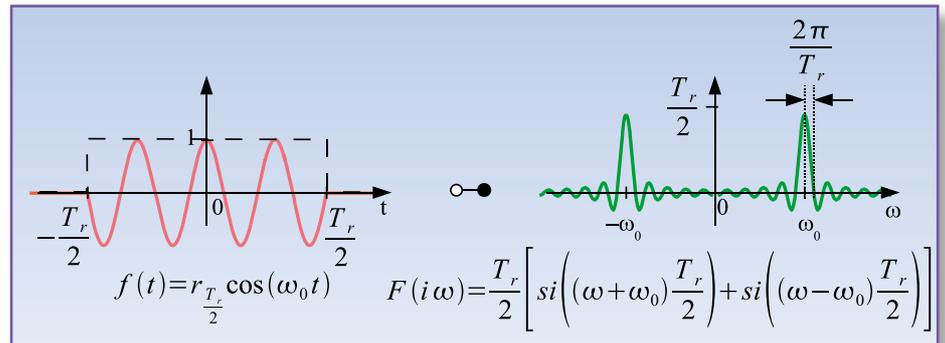
Signumfunktion

Wird von einem mit 2 multiplizierten Einheitssprung 1 abgezogen, erhält man die in Abbildung 12 gezeigte und in Formel (34) dargestellte Signumfunktion $\text{sgn}(t)$ (signum: Vorzeichen). Mit dem aus Abbildung 11 bekannten Ergebnis für den Einheitssprung ergibt sich durch summandenweise Transformation die Fourierkorrespondenz der Signumfunktion (35).

Rechteckfunktion

Von großer Wichtigkeit ist auch die Rechteckfunktion (Abbildung 13). Sie ist definiert als (36).

Auch hier werden wie bei $\delta(t)$ und $\text{sgn}(t)$ die Funktionswerte an den Sprungstellen als



$$F(i\omega) = \int_{t=-\frac{T_r}{2}}^{\frac{T_r}{2}} \cos(\omega_0 t) \cdot e^{-i\omega t} dt \quad (28)$$

$$= \int_{t=-\frac{T_r}{2}}^{\frac{T_r}{2}} [\cos(\omega_0 t) \cos(\omega t) - i \cos(\omega_0 t) \sin(\omega t)] dt$$

$$\cos \alpha \cdot \cos \beta = \frac{1}{2} [\cos(\alpha + \beta) + \cos(\alpha - \beta)] \quad (29a)$$

$$\cos \alpha \cdot \sin \beta = \frac{1}{2} [\sin(\alpha + \beta) - \sin(\alpha - \beta)] \quad (29b)$$

$$F(i\omega) = \frac{1}{2} \int_{t=-\frac{T_r}{2}}^{\frac{T_r}{2}} \left([\cos(\omega + \omega_0)t + \cos(\omega - \omega_0)t] - i [\sin(\omega + \omega_0)t - \sin(\omega - \omega_0)t] \right) dt \quad (30)$$

Mittelwerte der rechts- und linksseitigen Grenzwerte gesetzt. Mit der Rechteckfunktion lassen sich durch Multiplikation mit einer beliebigen anderen Funktion Abschnitte aus dieser erzeugen. Man nennt dies auch „Fenster“. Ein Beispiel für eine „gefensterte“ Funktion ist der ein- und ausgeschaltete Kosinus (Abbildung 14).

Die Fensterbreite T_r hat hier natürlich nichts mit der Periodendauer der geschalteten Kosinusschwingung zu tun. Um Verwechslungen zu vermeiden, hat sie den Index r erhalten.

$$F(i\omega) = \frac{1}{2} \left(\frac{\sin(\omega + \omega_0)t}{\omega + \omega_0} + \frac{\sin(\omega - \omega_0)t}{\omega - \omega_0} + i \frac{\cos(\omega + \omega_0)t}{\omega + \omega_0} - i \frac{\cos(\omega - \omega_0)t}{\omega - \omega_0} \right) \Bigg|_{t=-\frac{T_r}{2}}^{\frac{T_r}{2}} \quad (31)$$

$$F(i\omega) = \frac{1}{2} \left(\frac{\sin(\omega + \omega_0) \frac{T_r}{2}}{\omega + \omega_0} - \frac{\sin(\omega + \omega_0) (-\frac{T_r}{2})}{\omega + \omega_0} + \frac{\sin(\omega - \omega_0) \frac{T_r}{2}}{\omega - \omega_0} - \frac{\sin(\omega - \omega_0) (-\frac{T_r}{2})}{\omega - \omega_0} \right) \quad (32)$$

$$F(i\omega) = \frac{\sin(\omega + \omega_0) \frac{T_r}{2}}{\omega + \omega_0} + \frac{\sin(\omega - \omega_0) \frac{T_r}{2}}{\omega - \omega_0}$$

$$F(i\omega) = \frac{T_r}{2} \left[\text{si}\left(\left(\omega + \omega_0\right) \frac{T_r}{2}\right) + \text{si}\left(\left(\omega - \omega_0\right) \frac{T_r}{2}\right) \right]$$

Interessant ist eine genaue Betrachtung der aus zwei si-Funktionen um $\pm\omega$ bestehenden Transformaten des Kosinusabschnitts. Je schmaler der Abschnitt ist, umso breiter zerfließen die si-Spektren, wobei sie sich mit zunehmender Wirkung überlagern. Im Extremfall, wenn das Rechteckfenster nur noch die Spitze des Kosinus ausschneidet, die näherungsweise als konstant 1 betrachtet werden kann, überlagern sich die si-Funktionen zu der Transformaten eines schmalen Rechtecks. Vergrößert sich dagegen die Rechteckbreite T , werden also mehr Schwingungszüge aus dem periodischen Kosinus ausgeschnitten, nähern sich die Nullstellen der si-Funktionen weiter aneinander an und ihre Amplitude steigt.

Die si-Funktionen werden gewissermaßen „in die Höhe“ gestaucht. Dies ist auch anschaulich erklärbar. Schließlich steigt mit zunehmender Fensterbreite der Energieinhalt des Zeitsignals, was bei einem schmaleren Frequenzspektrum durch dessen entsprechend erhöhte Amplitude kompensiert werden muss.

Im Grenzübergang eines unendlich breiten Fensters gehen sie in die zwei δ -Funktionen des periodischen Kosinus über. Abbildung 15 demonstriert dies.

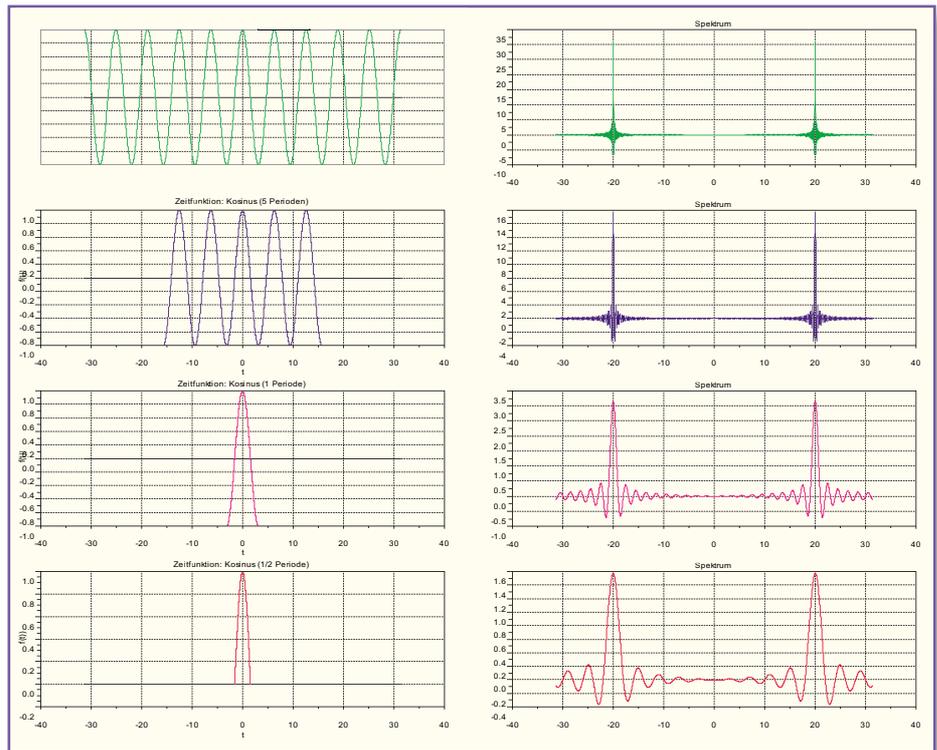


Bild 15: Hier kann man die Wirkung der Fensterbreite studieren. Je schmaler das Fenster im Zeitbereich, umso mehr „zerlaufen“ im Frequenzbereich die si-Funktionen. Je breiter das Fenster, desto mehr werden sie gestaucht.

$$f(t) = u(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2} & t = 0 \\ 1 & t > 0 \end{cases} \quad \text{Sprungfunktion} \quad (33)$$

$$f(t) = \text{sgn}(t) = \begin{cases} -1 & t < 0 \\ 0 & t = 0 \\ 1 & t > 0 \end{cases} \quad \text{Signumfunktion} \quad (34)$$

$$f(t) = 2 \cdot u(t) - 1 = \text{sgn}(t) \quad (35)$$

$$F(i\omega) = 2\pi \delta(\omega) - i \frac{2}{\omega} - 2\pi \delta(\omega) = -i \frac{2}{\omega}$$

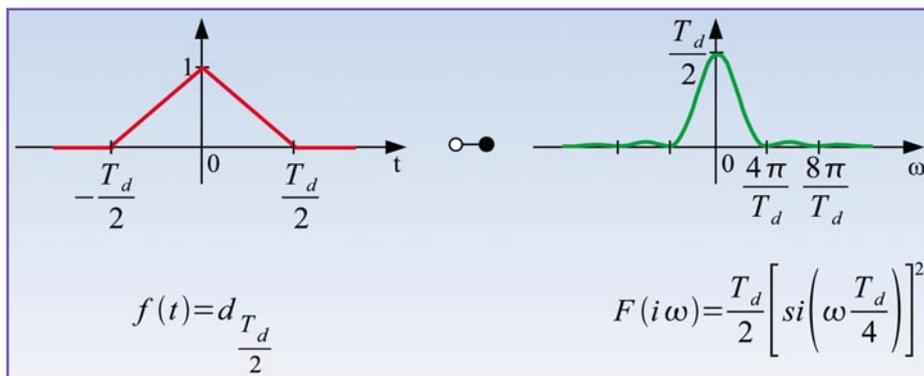
$$f(t) = r_{T_r} = \begin{cases} 0 & |t| > \frac{T_r}{2} \\ \frac{1}{2} & |t| = \frac{T_r}{2} \\ 1 & -\frac{T_r}{2} < t < \frac{T_r}{2} \end{cases} \quad \text{Rechteckfunktion} \quad (36)$$

$$f(t) = r_{T_r} \cdot \cos(\omega_0 t) = \begin{cases} 0 & |t| > \frac{T_r}{2} \\ \cos(\omega_0 t) & -\frac{T_r}{2} < t < \frac{T_r}{2} \end{cases} \quad (37)$$

$$f(t) = d_{T_d} = \begin{cases} 0 & |t| > \frac{T_d}{2} \\ -\frac{2}{T_d} \cdot t + 1 & |t| \leq \frac{T_d}{2} \end{cases} \quad \text{Dreieckfunktion} \quad (38)$$

Dreiecksfunktion

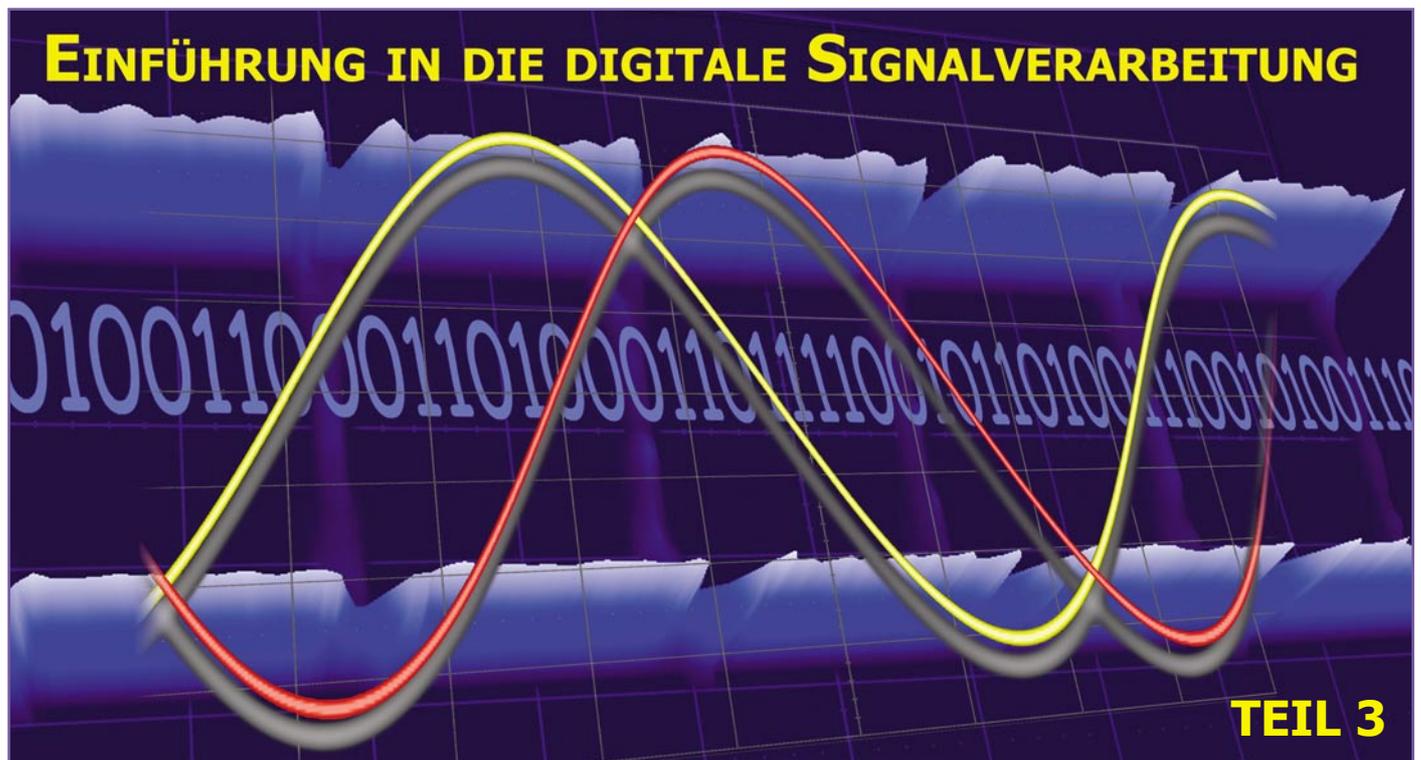
Als letztes Beispiel wollen wir die Dreiecksfunktion und ihr Spektrum betrachten (Abbildung 16). Die Dreiecksfunktion ist definiert als (38).



Auch hier erkennen wir, dass eine stetige Verbreiterung des Dreiecks sein Spektrum immer mehr staucht und dessen Amplitude zunehmen lässt. Im entgegengesetzten Fall „zerfließt“ das Spektrum.

In Teil 3 widmen wir uns noch dem Parseval'schen Theorem, um dann die Welt der digitalen Signale zu betreten. **ELV**

Bild 16: Auch die Dreiecksfunktion kann zum „Fenster“ verwendet werden.



Mit der Korrelation kann man kleinste Signale im Rauschen aufspüren und das Parseval'sche Theorem ableiten. Was die Multiplikation im Zeitbereich, ist die Faltung im Frequenzbereich und umgekehrt. Dieser elementare Zusammenhang hilft, tiefere Einblicke in die Signalverarbeitung zu gewinnen. So kann man damit begründen, dass ein abgetastetes Zeitsignal ein periodisches Spektrum hat.

Das Parseval'sche Theorem

Benannt nach dem französischen Mathematiker Marc-Antoine Parseval des Chênes (1755–1836), beschreibt es die Gleichheit der Gesamtenergie eines reellen Signals im Zeitbereich $x(t)$ und der desselben im Frequenzbereich $X(i\omega)$. Parseval erschien die Aussage seines Theorems so offensichtlich (schließlich handelt es sich hierbei ja nur um zwei verschiedene Sichtweisen ein und desselben Signals), dass er auf einen Beweis verzichtete.

Die Parseval'sche Formel lautet:

$$\int_{t=-\infty}^{+\infty} x(t)^2 dt = \frac{1}{2\pi} \int_{\omega=-\infty}^{+\infty} |X(i\omega)|^2 d\omega \quad (39)$$

Parsevalsche Formel

Während die linke Seite von (39) die Zeitintegration der Momentanleistung – eine Energie – darstellt, beschreibt die rechte Seite das Frequenzintegral über die Energiedichte – also ebenfalls eine Energie. Die rechte Seite von (39) leitet sich aus der Fouriertransformierten der so genannten Autokorrelation eines Energiesignals ab.

Das Wort Korrelation kommt aus dem Lateinischen und bedeutet dort so viel wie Wechselbeziehung. In der Systemtheorie der Nachrichtentechnik beschreibt man mit der Kreuzkorrelationsfunktion (KKF) das

Maß der Ähnlichkeit zweier Funktionen:

Die Kreuzkorrelierte (KKF) zweier Energie-Zeit-Funktionen $x(t)$ und $y(t)$ lautet:

$$k_{xy}(\tau) = \int_{t=-\infty}^{+\infty} x(t)y(t+\tau) dt \quad (40)$$

Kreuzkorrelierte

Mathematisch gesehen verschiebt man eine der Funktionen um einen gewissen zeitlichen Abstand τ und integriert das Produkt mit der anderen, unverschobenen Funktion über die Zeit. Bei Funktionen mit geringen Wechselbeziehungen ergeben sich kleine Kreuzkorrelierte. Stellt man sich vor, die beiden Zeitfunktionen würden sich immer ähnlicher (bis zur Gleichheit), geht die Kreuzkorrelierte in die Autokorrelierte über.

Die Autokorrelierte (AKF) einer Zeitfunktion $x(t)$ lautet also:

$$k_{xx}(\tau) = \int_{t=-\infty}^{+\infty} x(t)x(t+\tau) dt \quad (41)$$

Autokorrelierte

Auf den ersten Blick scheint es vielleicht merkwürdig, ein Signal mit seiner zeitverschobenen Kopie zu multiplizieren und das Produkt zu integrieren. Aber dieses Verfahren hat sich hervorragend bewährt, um schwache und stark verrauschte

elektromagnetische Signale aufzuspielen.

Zum Beispiel wird in einem Radarsystem das mit Rauschen überlagerte schwache und laufzeitverschobene Echo $x(t + \tau)$ mit dem Sendesignal $x(t)$ verglichen, indem man die Variable τ so lange verändert, bis die AKF maximal ist. Dann ist man einerseits sicher, das richtige Signal und kein Störecho zu empfangen, und kann aus der gefundenen Laufzeit τ den Abstand zum reflektierenden Objekt berechnen. Bei Funkwellen, die sich mit Lichtgeschwindigkeit 300.000 km/s ausbreiten, und unter Berücksichtigung der Tatsache, dass die Laufstrecke des Signals der doppelten Entfernung zum reflektierenden Objekt entspricht, führt $\tau = 200 \mu s$ zu einem Objektabstand von 30 km.

Ein weiteres Beispiel ist die Mustererkennung in einem Bild, wie sie bei der Überwachung und Steuerung industrieller Prozesse, des Straßenverkehrs, bei der Bildanalyse usw. stattfindet. Hier wird statt der eindimensionalen Zeitvariablen (t) aber mit den zweidimensionalen Ortsfrequenzvariablen (x, y) gearbeitet. Das führt zu erheblichem Rechenaufwand und schließt eine Echtzeitverarbeitung oft aus.

Über die AKF zur Parseval'schen Formel

Es gibt verschiedene Wege, die Parseval'sche Formel herzuleiten. Einer davon ist die Anwendung des Faltungssatzes, den

wir später noch ausführlich beleuchten.

Ein anderer Weg führt über die AKF. Setzt man die Autokorrelierte nach (41) in die Definitionsgleichung der Fouriertransformation (18) ein, ergibt sich (42). Im Gang der Rechnung haben wir vom Zeitverschiebungssatz aus Tabelle 1 Gebrauch gemacht und berücksichtigt, dass eine reelle Zeitfunktion ein Spektrum mit geradem Realteil und ungeradem Imaginärteil hat.

Der Zeitverschiebungssatz begründet die Anschrift:

$$\int_{\tau=-\infty}^{+\infty} x(\tau+t)e^{-i\omega\tau} d\tau = X(i\omega) \cdot e^{i\omega t}$$

Mit dem Satz über reelle Zeitfunktionen können wir herleiten, dass:

$$\int_{t=-\infty}^{+\infty} x(t) \cdot e^{+i\omega t} dt = \bar{X}(i\omega)$$

Offensichtlich wäre die linke Seite der Gleichung bei einem negativen Vorzeichen des Exponenten der e-Funktion definitionsgemäß gleich $X(i\omega)$. Wegen des positiven Exponenten erhalten wir aber $X(-i\omega)$. Weil $x(t)$ reell ist, gilt: $\text{Re}[X(-i\omega)] =$

$$K_{xx}(i\omega) = \int_{\tau=-\infty}^{+\infty} k_{xx}(\tau) e^{-i\omega\tau} d\tau \quad (42)$$

$$K_{xx}(i\omega) = \int_{\tau=-\infty}^{+\infty} \int_{t=-\infty}^{+\infty} x(t)x(t+\tau) e^{-i\omega\tau} dt d\tau$$

Jetzt vertauschen wir die Reihenfolge der Integration

$$K_{xx}(i\omega) = \int_{t=-\infty}^{+\infty} x(t) \underbrace{\int_{\tau=-\infty}^{+\infty} x(t+\tau) e^{-i\omega\tau} d\tau}_{X(i\omega) \cdot e^{i\omega t}} dt$$

$$K_{xx}(i\omega) = \int_{t=-\infty}^{+\infty} \underbrace{x(t) \cdot e^{i\omega t}}_{\bar{X}(i\omega)} dt \cdot X(i\omega)$$

$$K_{xx}(i\omega) = X(i\omega) \cdot \bar{X}(i\omega) = |X(i\omega)|^2 \quad \text{Energiedichtespektrum}$$

$$k_{xx}(0) = \int_{t=-\infty}^{+\infty} x(t)^2 dt \quad \text{Gesamtenergie} \quad (43)$$

$$k_{xx}(\tau) = \frac{1}{2\pi} \int_{\omega=-\infty}^{+\infty} K_{xx}(i\omega) e^{i\omega\tau} d\omega = \frac{1}{2\pi} \int_{\omega=-\infty}^{+\infty} |X(i\omega)|^2 e^{i\omega\tau} d\omega \quad (44)$$

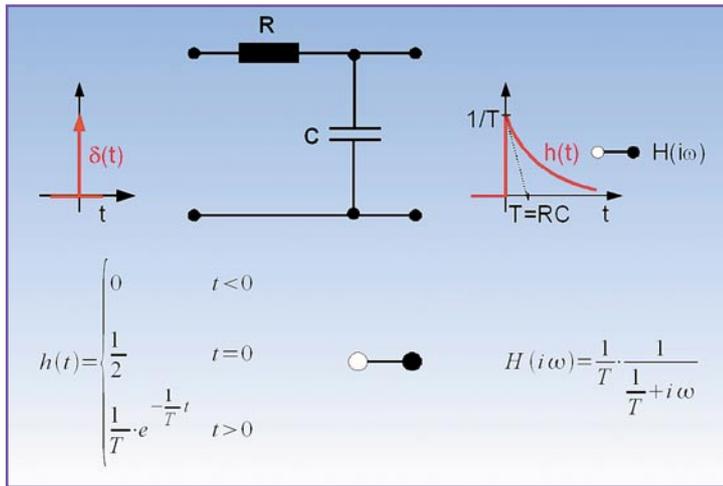


Bild 17: Der RC-Tiefpass ist ein Spannungsteiler, dessen unterer Teilerwiderstand ein Kondensator ist. Sein Widerstand nimmt mit zunehmender Frequenz ab und damit auch die Ausgangsspannung des Teilers.

$\text{Re}[X(i\omega)]$ und $\text{Im}[X(-i\omega)] = -\text{Im}[X(i\omega)]$, d. h. der Realteil hat sich nicht verändert, der Imaginärteil dagegen sein Vorzeichen umgekehrt, was zu zeigen war.

$K_{xx}(i\omega)$ stellt als Fouriertransformierte der Autokorrelierten des Zeitsignals $x(t)$ dessen Energiedichte dar, man nennt

$K_{xx}(i\omega)$ deshalb auch das Energiedichtespektrum von $x(t)$. Um zur Parseval'schen Gleichung zu gelangen, ist es nur noch ein kleiner Schritt. Die Gesamtenergie des Zeitsignals berechnet sich aus (41) als (43).

Um aus der Identität in der letzten Zeile von (42) wieder die erzeugende Zeitfunktio-

Signalenergie der Zeitfunktion:

$$E = \int_{t=-\infty}^{+\infty} h(t)^2 dt = \frac{1}{T^2} \int_0^{\infty} e^{-\frac{2}{T}t} dt = \frac{1}{T^2} \cdot \frac{-T}{2} e^{-\frac{2}{T}t} \Big|_{t=0}^{\infty} = \frac{1}{2T} \quad (45)$$

Signalenergie der Frequenzfunktion:

$$|H(i\omega)|^2 = \frac{1}{T^2} \left| \frac{1}{\frac{1}{T} + i\omega} \right|^2 = \frac{1}{T^2} \frac{1}{\frac{1}{T^2} + \omega^2} \quad (46)$$

on zu ermitteln, müssen wir diese als inverse Fouriertransformierte ausdrücken (44).

Gleichung (43) und (44) (mit $\tau = 0$) werden jetzt gleichgesetzt und es folgt die Parseval'sche Formel, wie in (39) angegeben.

Ein praktisches Beispiel zum Parseval'schen Theorem

Zur Vertiefung und als Rechenübung wollen wir das Parseval'sche Theorem anhand der Impulsantwort eines RC-Tiefpasses verifizieren. Die Beziehungen dazu sind in Abbildung 17 zusammengestellt. Die komplexe Fouriertransformierte $H(i\omega)$ der reellen Impulsantwort $h(t)$ wurde bereits in (27) abgeleitet und in Abbildung 9 oben dargestellt.

Zur Bestätigung des Parseval'schen Theorems soll nun gezeigt werden, dass $h(t)$ und $H(i\omega)$ den gleichen Signalenergieinhalt haben (s. Formel 45 und 46).

Zunächst berechnen wir das Betrags-

$$E = \frac{1}{2\pi T^2} \int_{\omega=-\infty}^{+\infty} \frac{1}{\frac{1}{T^2} + \omega^2} d\omega = \frac{1}{2\pi T^2} T \text{arc tan}(\omega T) \Big|_{\omega=-\infty}^{\infty} = \frac{1}{2\pi T} \left(\frac{\pi}{2} - \left(-\frac{\pi}{2}\right) \right) = \frac{1}{2T} \quad (47)$$

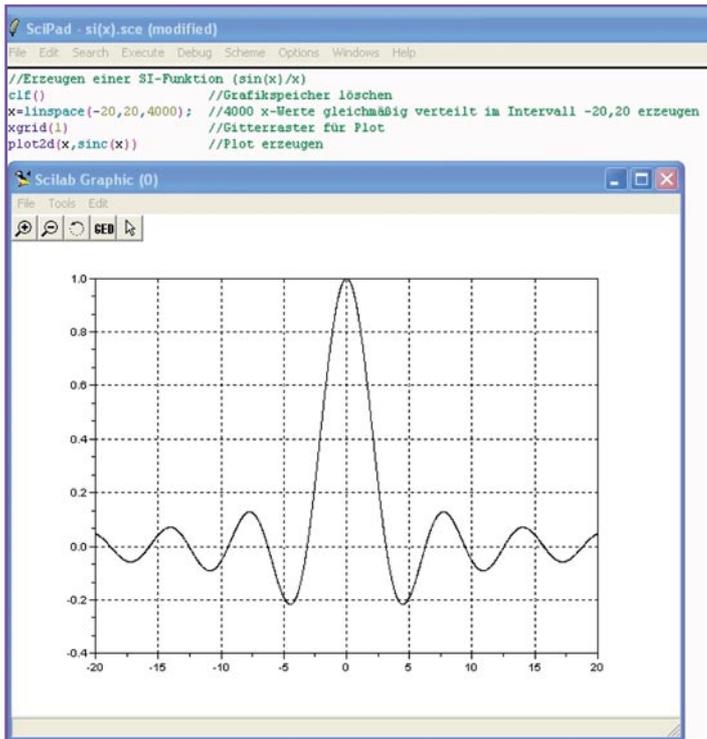


Bild 18: Mit wenigen Zeilen ist ein kleines Programm geschrieben, das die Funktion $\sin(x)/x$ als Grafik ausgibt.

quadrat von $H(i\omega)$ (46).

Das Betragsquadrat in die rechte Seite der Parseval'schen Formel eingesetzt, ergibt (47).

Damit sind die im Zeitbereich gemäß Gleichung (45) und im Frequenzbereich gemäß Gleichung (47) ermittelten Signalenergien identisch und die Parseval'sche Formel ist bestätigt.

Funktionsverläufe zeichnen

Liegt eine Funktion als Formel vor, kann man sich ein Bild von ihr verschaffen, indem man eine Tabelle erstellt, in der man die unabhängige Variable (t bei den Zeitfunktionen und ω bei den Spektren) eine gewisse Anzahl von Werten annehmen lässt und den dazugehörigen Funktionswert errechnet und daneben schreibt. Die so erhaltene Tabelle lässt sich zeichnerisch in Punkte in einem Koordinatensystem umsetzen. Verbindet man die Punkte, erhält man einen mehr oder weniger genauen Funktionsgraphen. Das ist bei komplizierteren Formeln aber abschreckend aufwändig.

Aber es gibt einen kostenlosen, mächtigen und eleganten Ausweg – SciLab. Seit Mitte Dezember 2006 in der Version 4.1, ist dieses hervorragende wissenschaftliche Softwarepaket unter www.scilab.org herunterladbar. Damit steht ein ausgezeichnetes Werkzeug für numerische Berechnungen und die grafische Präsentation der Ergebnisse zur Verfügung. Ja, es scheint so: „The best things in life are free.“

SciLab ist ein Interpreter, d. h., die geschriebenen Programme brauchen nicht übersetzt (kompiliert) zu werden, man kann sie sofort ausführen. Eine komplette

Einführung in SciLab würde den Rahmen dieser Artikelreihe bei weitem sprengen, aber wir wollen an einigen Beispielen ohne tiefere Erläuterungen demonstrieren, wie nützlich dieses Tool ist.

Angenommen, SciLab ist auf dem Rechner installiert. Nach dem Aufruf wechseln wir in den Editor SciPad. Dort erstellen wir das kleine Programm aus Abbildung 18. Um es zu testen, wählen wir über Execute den Menüpunkt „Load into Scilab“, und wenn wir alles richtig gemacht haben, poppt ein Fenster auf, das die gewünschte grafische Darstellung in Abbildung 18 zeigt.

Vom analogen zum digitalen Signal

Analoge Signale sind innerhalb ihres Wertebereichs durch eine unendlich feine Stufung bezüglich Zeit und Amplitude gekennzeichnet. Man sagt auch: „Analoge Signale sind zeit- und wertekontinuierlich“. Entnimmt man ihnen zu regelmäßigen Zeitpunkten Werteproben – diesen Vorgang bezeichnet man auch als Abtasten –, haben wir zunächst die Zeit diskretisiert, aber immer noch einen kontinuierlichen Wertebereich. Typische Vertreter solcher zeitdiskreten, aber wertekontinuierlichen Systeme sind SC-Filter, bei denen ohmsche Widerstände durch periodisch umgepolte Kondensatoren (SC: switched capacitor) ersetzt werden. Bei voll digital arbeitenden Schaltungen geht man noch einen Schritt weiter und ordnet den analogen Abtastwerten digitale Zahlenwerte zu. Diese werden durch Binärwörter mit endlicher Wortlänge repräsentiert, woraus eine gewisse „Granularität“ der Amplitudendarstellung resultiert. Jetzt haben wir

es also mit einer zeit- und wertediskreten Präsentation des Signals zu tun.

Das Abtasttheorem

Das Abtasttheorem beantwortet die Frage, in welchem zeitlichen Abstand einem Signal Proben entnommen werden müssen, um es aus denselben ohne Informationsverlust wieder vollständig rekonstruieren zu können. Das Abtasttheorem geht auf Arbeiten von Edmund Taylor Whittaker (1873–1956), Harry Nyquist (1889–1976), Wladimir Alexandrowitsch Kotelnikow (1908–2005) und Claude Elwood Shannon (1916–2001) (Abbildung 19) zurück und wird in der Literatur deshalb auch mit den Namen „Nyquist-Shannon-Abtasttheorem“ oder „WKS-Abtasttheorem“ (WKS: Whittaker Kotelnikow Shannon) belegt.



Bild 19: Claude Elwood Shannon hat die Nachrichtentechnik viel zu verdanken.

Mit den Kenntnissen aus der Fourieranalyse ist das Abtasttheorem einfach zu verstehen. Nach Fourier entsteht jede Funktion aus der Überlagerung elementarer harmonischer Schwingungen mit Vielfachen einer Grundfrequenz. Enthält das Signal nun eine Komponente mit der maximalen Frequenz f_{\max} , so sind dieser wenigstens zwei Abtastwerte pro Periode zu entnehmen, um sie wieder eindeutig rekonstruieren zu können. Mit anderen Worten: Wenn f_{\max} die höchste in einem Signal auftretende Frequenz ist, muss es mindestens mit der doppelten Maximalfrequenz $2 \cdot f_{\max}$ abgetastet werden, um Verluste bei der Rekonstruktion aus der Abtastwertefolge zu vermeiden.

Mit der Schreibweise f_s für die Abtastfrequenz (s steht für das englische Wort sample = Probe) zeigt Gleichung (48) die mathematische Formulierung des Abtasttheorems.

$$f_s \geq 2 \cdot f_{\max} \quad \text{Abtasttheorem} \quad (48)$$

Für die Einhaltung des Abtasttheorems ist also eine Begrenzung der Bandbreite des abzutastenden Signals (Bandbegrenzung) erforderlich. Dazu dient ein analoger Tiefpass, dessen Grenzfrequenz gleich oder kleiner als die halbe Abtastfrequenz sein muss. Damit ist sichergestellt, dass keine hohen Frequenzanteile im abzutastenden Signal enthalten sind, welche das Abtast-

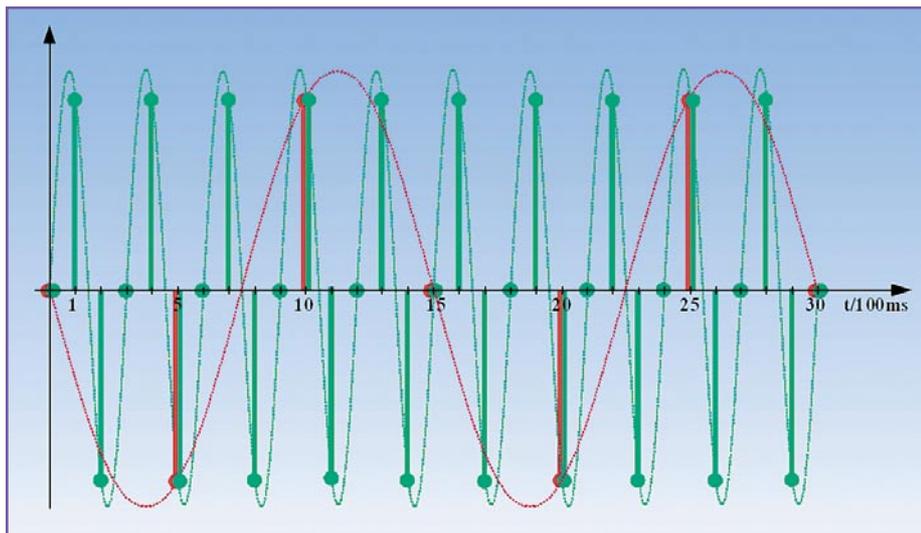


Bild 20: Wird das Abtasttheorem nicht eingehalten, entstehen Aliasfrequenzen.

theorem verletzen und so genannte Alias-Effekte hervorrufen würden. In der Praxis ist es wegen der endlichen Flankensteilheit des analogen Tiefpasses üblich, die Abtastfrequenz deutlich größer als die doppelte Maximalfrequenz im Signal zu wählen.

Am Beispiel in Abbildung 20 wird dieser Sachverhalt verdeutlicht. Die Zeitachse ist in ein 100-ms-Raster unterteilt. Eine Periode der grünen Sinuskurve dauert drei Rasterschritte, ihre Frequenz beträgt somit 3,33 Hz. Sie wird alle 100 ms, also mit einer Abtastfrequenz von 10 Hz abgetastet. Damit ist das Abtasttheorem erfüllt, welches ja mindestens eine Abtastfrequenz von 6,66 Hz verlangt, man spricht von Überabtastung. Zwischen den so gewonnenen grünen Abtastwerten lässt sich nur eine Sinuskurve – nämlich die grüne – einpassen. Dieser Vorgang ist die originalgetreue Rekonstruktion der ursprünglichen grünen abgetasteten Kurve. Anders sieht es aus, wenn wir die grüne Kurve nur alle 5 Takte, also mit 2 Hz, abtasten. Das Abtasttheorem wird verletzt (Unterabtastung) und die rekonstruierte rote Kurve ist nicht mehr identisch mit der grünen Ausgangskurve. Sie ist vielmehr ein niederfrequenteres Abbild (Alias), dessen Periode 15 Takte umfasst und dessen Frequenz somit 2/3 Hz beträgt.

Aliasfrequenzen entstehen nur bei Verletzung des Abtasttheorems, wenn also $f_s < 2 f_{\max}$ ist, also die höchste im Signal auftretende Harmonische nicht mindestens zweimal pro Periode abgetastet wird. Aliasfrequenzen berechnen sich dann nach der Beziehung Gleichung (49):

$$f_{\text{Alias}} = |n \cdot f_s - f| \quad n = 1, 2, 3, \dots \quad (49)$$

n ist als ganze Zahl so zu wählen, dass der Betrag von $n f_s - f$ den kleinstmöglichen Wert annimmt. Im Beispiel von Abbildung 20

$$x_s(t) = x(t) \cdot s_s(t) = x(t) \cdot \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) = \sum_{n=-\infty}^{+\infty} x(nT_s) \cdot \delta(t - nT_s) \quad (50)$$

$$y(t) = \int_{\tau=-\infty}^{+\infty} x_1(\tau) \cdot x_2(t - \tau) d\tau \quad \text{Faltung im Zeitbereich} \quad (51)$$

$$y(t) = x_1(t) * x_2(t) \quad \text{Symbolische Schreibweise mit Faltungsstern}$$

$$Y(i\omega) = X_1(i\omega) \cdot X_2(i\omega) \quad \text{Multiplikation im Frequenzbereich} \quad (52)$$

$$Y(i\omega) = \int_{\nu=-\infty}^{+\infty} X_1(i\nu) \cdot X_2(i\omega - i\nu) d\nu \quad \text{Faltung im Frequenzbereich} \quad (53)$$

$$Y(i\omega) = X_1(i\omega) * X_2(i\omega) \quad \text{Symbolische Schreibweise mit Faltungsstern}$$

$$y(t) = x_1(t) \cdot x_2(t) \quad \text{Multiplikation im Zeitbereich} \quad (54)$$

wäre demnach $n = 2$, woraus die Aliasfrequenz $(2 \cdot 2 - 3^{1/3}) \text{ Hz} = 2^{2/3} \text{ Hz}$ folgt. Ein weiteres Beispiel: Vier Frequenzen $f_1 = 25 \text{ Hz}$, $f_2 = 70 \text{ Hz}$, $f_3 = 160 \text{ Hz}$ und $f_4 = 510 \text{ Hz}$ werden mit $f_s = 100 \text{ Hz}$ abgetastet. Bei der Abtastung von f_2 , f_3 und f_4 wird das Abtasttheorem verletzt und es entstehen die Aliasfrequenzen $f_{2 \text{ Alias}} = \text{Betrag von } (100 - 70) \text{ Hz} = 30 \text{ Hz}$, $f_{3 \text{ Alias}} = \text{Betrag von } (200 - 160) \text{ Hz} = 40 \text{ Hz}$ sowie $f_{4 \text{ Alias}} = \text{Betrag von } (500 - 510) \text{ Hz} = 10 \text{ Hz}$.

Die Wirkung einer Unterabtastung ist auch in alten Westernfilmen in Gestalt sich scheinbar rückwärts drehender Wagenräder und als Stroboskopeffekt in der Diskothek beobachtbar.

Der Abtastvorgang wird mathematisch durch die Multiplikation einer kontinuierlichen Zeitfunktion $x(t)$ mit einer regelmäßigen Folge von Dirac-Impulsen $s_s(t)$ beschrieben. Der zeitliche Abstand der Dirac-Impulse entspricht der Abtastperiode T_s . Durch die Ausblendeigenschaft des Dirac-Impulses wird nur der Funktionswert von $x(t)$ selektiert, der zum Zeitpunkt des Dirac-Impulses existiert.

Gleichung (50) drückt dies aus.

Um aus dieser Gleichung weitere Erkenntnisse bequem ableiten zu können, werden wir uns jetzt ein wenig mit der Faltung – im Englischen Convolution genannt – beschäftigen.

Multiplizieren hier – Falten dort

Aus Tabelle 1 wissen wir, dass die Multiplikation zweier Zeitfunktionen der Faltung ihrer Spektren (auf Englisch: convolution) entspricht bzw. umgekehrt. Doch was ist nun eine Faltung zweier Funktionen? Mathematisch gesprochen ist die Faltung eine Vorschrift, nach der zwei Funktionen $x_1(t)$ und $x_2(t)$ auf eine dritte Funktion $y(t)$ abgebildet werden. Das hört sich komplizierter an, als es ist.

Im analogen Zeitbereich ist die Faltung

zweier Funktionen wie in Gleichung (51) definiert. Im Frequenzbereich ist die Fouriertransformierte von $y(t)$ gleich dem Produkt der Fouriertransformierten von $x_1(t)$ und $x_2(t)$, wie in Gleichung (52) dargestellt.

Werden zwei Frequenzfunktionen $X_1(\omega)$ und $X_2(\omega)$ miteinander gefaltet, entspricht das der Produktbildung ihrer Fourierrücktransformierten im Zeitbereich. Vergleiche Gleichungen (53) und (54).

Wir werden später genauer auf Anwendungen der Faltung eingehen. An dieser Stelle wollen wir als Merksatz knapp zusammenfassen:

Faltung im Zeitbereich entspricht Multiplikation im Frequenzbereich bzw. Multiplikation im Zeitbereich entspricht Faltung im Frequenzbereich.

Von der Faltungssumme zum Faltungsintegral

Wer mathematisch etwas vorgebildet ist, kann das Faltungsintegral aus der Gleichung auch anschaulich interpretieren.

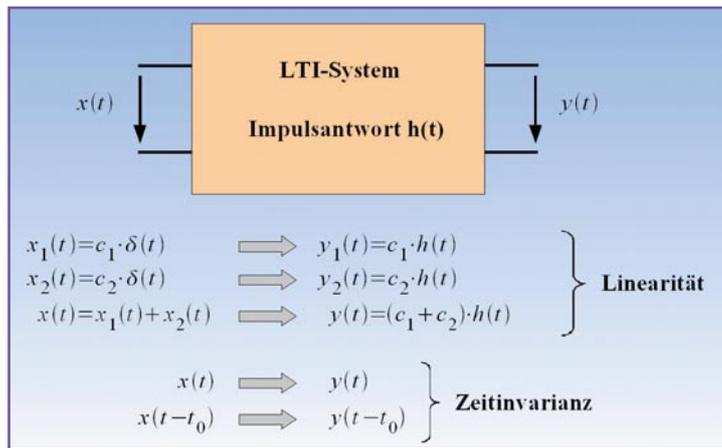


Bild 21: Die Eigenschaften eines LTI-Systems: linear und zeitunabhängig.

Rechteckdauer $\Delta\tau$ zum infinitesimal kurzen Zeitabschnitt $d\tau$, die Summe zum Integral und die diskreten Zeitpunkte $n\Delta\tau$ zur Laufvariablen τ wie in (56) und (57). Die Gleichung (56) ist nichts weiter als die bereits aus Gleichung (22) bekannte Ausblendeigenschaft des δ -Impulses. Gleichung (57) ist das Faltungsintegral aus dem Eingangssignal und der Impulsantwort des Systems. Wegen der LTI-Eigenschaften genügt es also, die Antwort $y(t) = h(t)$ des Systems auf einen Dirac-Impuls $\delta(t)$ als Eingangssignal zu kennen. Für ein beliebiges Eingangssignal $x(t)$ folgt dann aus Gleichung (57) die entsprechende Ausgangsspannung $y(t) = x(t)*h(t)$.

Vorausgesetzt wird ein lineares, zeitinvariantes System (LTI: Linear Time Invariant). Darunter versteht man ein System, das zu jedem Zeitpunkt auf die gleichen Eingangsanregungen mit den gleichen proportionalen Reaktionen antwortet. Abbildung 21 fasst diese Eigenschaften zusammen.

Zur Ableitung des Faltungsintegrals denken wir uns nun die Eingangsanregung $x(t)$ durch eine Treppenkurve angenähert (Abbildung 22). Sie besteht aus einer Folge von Rechtecken der Breite $\Delta\tau$ und einer Höhe, die dem Eingangssignalwert in der Mitte des Rechtecks entspricht. Am Beispiel eines RC-Tiefpasses sind die Antworten auf die einzelnen Rechtecke in deren Farben als ansteigende und abfallende e-Funktionen untereinander dargestellt. Wegen der LTI-Eigenschaften überlagern sie sich zu einer Summe entsprechend Gleichung (55).

Je feiner nun die Treppe gestuft wird, umso ähnlicher wird die Überlagerungssumme dem wahren Ausgangssignal des Systems. Im Grenzübergang für $\Delta\tau \rightarrow 0$ gehen die Rechtecke des treppenförmigen Eingangssignals $x_r(t)$ in eine unendliche

$$\tilde{y}(t) = \sum_{n=-\infty}^{\infty} x(n \cdot \Delta\tau) \cdot y_r(t - n \cdot \Delta\tau) \cdot \Delta\tau \quad \text{Überlagerungssumme} \quad (55)$$

$$x(t) = \lim_{\Delta\tau \rightarrow 0} \sum_{n=-\infty}^{\infty} x(n \cdot \Delta\tau) \cdot x_r(t - n \cdot \Delta\tau) \cdot \Delta\tau = \int_{-\infty}^{\infty} x(\tau) \cdot \delta(t - \tau) d\tau \quad (56)$$

$$y(t) = \lim_{\Delta\tau \rightarrow 0} \sum_{n=-\infty}^{\infty} x(n \cdot \Delta\tau) \cdot y_r(t - n \cdot \Delta\tau) \cdot \Delta\tau = \int_{-\infty}^{\infty} x(\tau) \cdot h(t - \tau) d\tau \quad (57)$$

$$x(t) * h(t) = h(t) * x(t) \quad \text{Die Faltung ist kommutativ} \quad (58)$$

$$[x(t) * h(t)] * g(t) = x(t) * [h(t) * g(t)] \quad \text{Die Faltung ist assoziativ} \quad (59)$$

$$x(t) * [h(t) + g(t)] = [x(t) * h(t)] + [x(t) * g(t)] \quad \text{Die Faltung ist distributiv} \quad (60)$$

$$x(t) * \delta(t) = x(t) \quad \text{Der Diracimpuls ist das neutrale Element der Faltung} \quad (61)$$

$$x(t) * \delta(t - t_0) = x(t - t_0) \quad \text{Verschiebungseigenschaft des Diracimpulses} \quad (62)$$

Folge gewichteter Dirac-Impulse über, die Rechteckantwort des Systems $y_r(t)$ wird zur Impulsantwort $h(t)$, die endliche

Eine erstaunliche Tatsache!
Es ist sehr nützlich, einige der wichtigsten in den Gleichungen (58–62) dargestellten Eigenschaften der Faltung zu kennen. Die meisten davon lassen sich unmittelbar aus dem Definitionsintegral nach Gleichung (51) oder (57) herleiten.

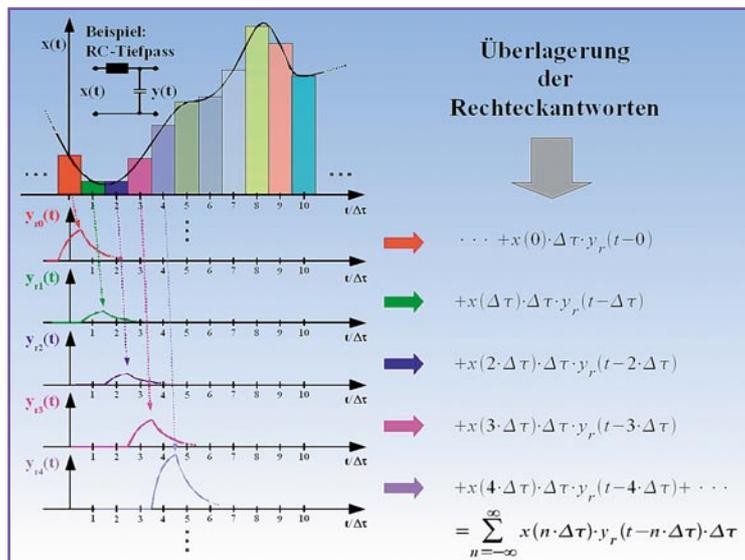


Bild 22: Die Überlagerung der Systemantworten auf eine durch eine Folge von Rechtecken nachgebildete Eingangsfunktion führt zur Faltungssumme als Ausgangsfunktion.

Grafische Veranschaulichung des Faltungsintegrals

Wir wollen anhand von Gleichung (63) grafisch die Wirkungsweise der Faltung nachvollziehen.

Als anregendes Signal $x(t)$ nehmen wir zwei aufeinanderfolgende Rechtecke – das erste negativ, das zweite positiv. Die Impulsantwort des Systems $h(t)$ ist eine abklingende e-Funktion. Abbildung 23 zeigt, wie $x(t)$ und $h(t)$ mit Hilfe des Faltungsintegrals Gleichung (63) in $y(t)$ überführt werden.

Abbildung 24 demonstriert grafisch, wie die Faktoren des Integranden, der Integrand selbst und das Integral zu aufeinanderfolgenden Zeitpunkten t_1, t_2, t_3 und t_4

$$y(t) = \int_{\tau=-\infty}^{\infty} x(\tau) \cdot h(t-\tau) d\tau = \int_{\tau=-\infty}^{\infty} x(t-\tau) \cdot h(\tau) d\tau \quad \text{Faltungsintegral} \quad (63)$$

$$X_s(i\omega) = \frac{1}{T_s} \sum_{k=-\infty}^{+\infty} X(i\omega - i k \omega_s) \quad (64)$$

Periodisches Spektrum einer abgetasteten Zeitfunktion

des analogen (unabgetasteten) Signals durch Tiefpassfilterung nicht mehr möglich. Somit kann auch das unabgetastete Signal nicht rekonstruiert werden.

Fassen wir also noch einmal zusammen: Durch das Abtasten eines bandbegrenzten Signals entstehen bei Vielfachen der Abtastfrequenz Kopien des ursprünglichen

aussehen. Mit zunehmender Zeit t schiebt sich das zeitgespiegelte Eingangssignal immer weiter über die Impulsantwort. Das Produkt aus beiden wird aufintegriert und ergibt den Momentanwert der Ausgangsspannung. Alle Momentanwerte zusammen repräsentieren den zeitlichen Verlauf des Ausgangssignals (gestrichelte Kurve).

Bei zeitlich begrenztem $x(t)$ und $h(t)$ entsteht erst dann ein Ausgangssignal, wenn sich die Faktoren des Integranden zu überlappen beginnen. Deshalb ist das Ausgangssignal $y(t)$ ebenfalls zeitlich begrenzt und so breit wie die Anregung $x(t)$ und die Impulsantwort $h(t)$ zusammen.

Wenn wir nun auf Gleichung (50) unser Wissen anwenden, dass der Faltung im Zeitbereich die Multiplikation im Frequenzbereich entspricht, führt dies unter Auslassung einiger Zwischenschritte bei der Berechnung zu Gleichung (64). Sie beschreibt nun nichts anderes als das mit der Abtastkreisfrequenz ω_s periodisch wiederholte Spektrum $X(i\omega)$ der nicht abgetasteten, bandbegrenzten Zeitfunktion $x(t)$. Mit anderen Worten: Wird eine Zeitfunktion $x(t)$, deren Spektrum $X(i\omega)$ ist, unter Einhaltung des Abtasttheorems regelmäßig im Raster $T_s = 2\pi/\omega_s$ abgetastet, so ist das Spektrum $X_s(i\omega)$ der abgetasteten Zeitfunktion eine in ω_s periodische Wiederholung von $X(i\omega)$. Wir merken uns deshalb:

Die Abtastung einer Zeitfunktion führt zur Periodisierung ihres Spektrums.

Wird das Abtasttheorem eingehalten, überlappen sich die periodischen Spektren nicht. Ist die Abtastfrequenz dagegen zu klein, überlagern sie sich. Deshalb lässt sich aus ihnen die ursprüngliche Zeitfunktion nicht mehr fehlerfrei zurückgewinnen. Abbildung 25 gibt dies wieder. Es ist jetzt

Bild 23: So werden das Eingangssignal $x(t)$ und die Impulsantwort des Systems zum Ausgangssignal „gefaltet“.

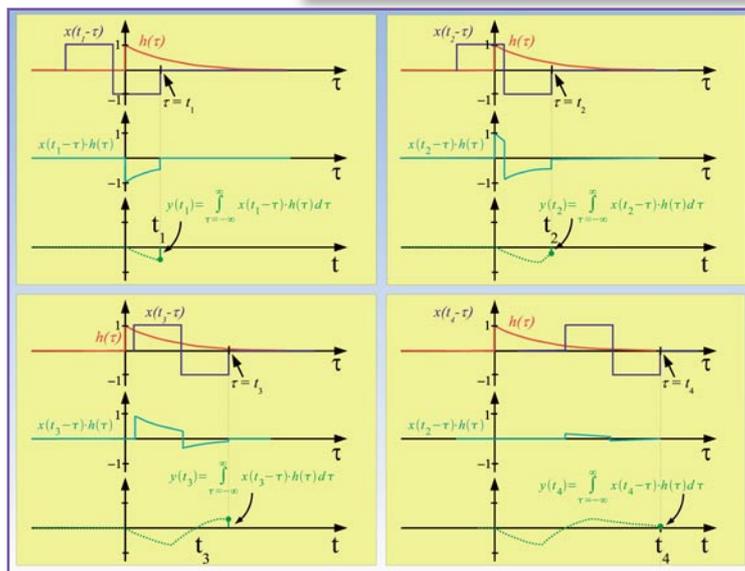
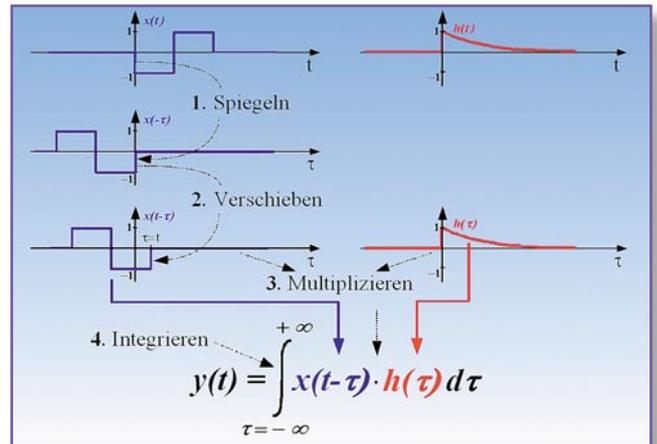


Bild 24: Der Faltungsprozess zu vier aufeinanderfolgenden Zeitpunkten

anschaulich zu erkennen, dass eine Verletzung des Abtasttheorems $\omega_s < 2 \omega_{max}$ eine Durchdringung des Basisbandspektrums mit seinen Nachbarn bewirkt. Damit ist die Wiedergewinnung des Basisbandspektrums

Spektrums des unabgetasteten Signals (Aliasspektren).

Die Betrachtung der Grenzfälle ist interessant. Lassen wir das Abtastintervall gegen null gehen, entnehmen wir also mit einer unendlich hohen Abtastfrequenz Signalproben, schieben sich die Aliasspektren ins Unendliche und es verbleibt das Basisbandspektrum des unabgetasteten Signals. Verringern wir die Abtastfrequenz dagegen immer mehr, rücken die periodischen Spektren immer weiter zusammen. Mit dem Verletzen des Abtasttheorems beginnen sie sich zu überlagern, bis sie im Grenzfall eines einzigen Abtastwertes den konstanten Summenwert 1 annehmen (65).

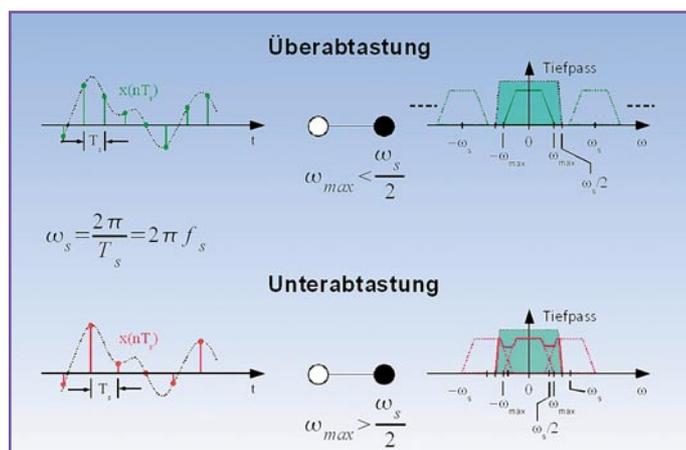
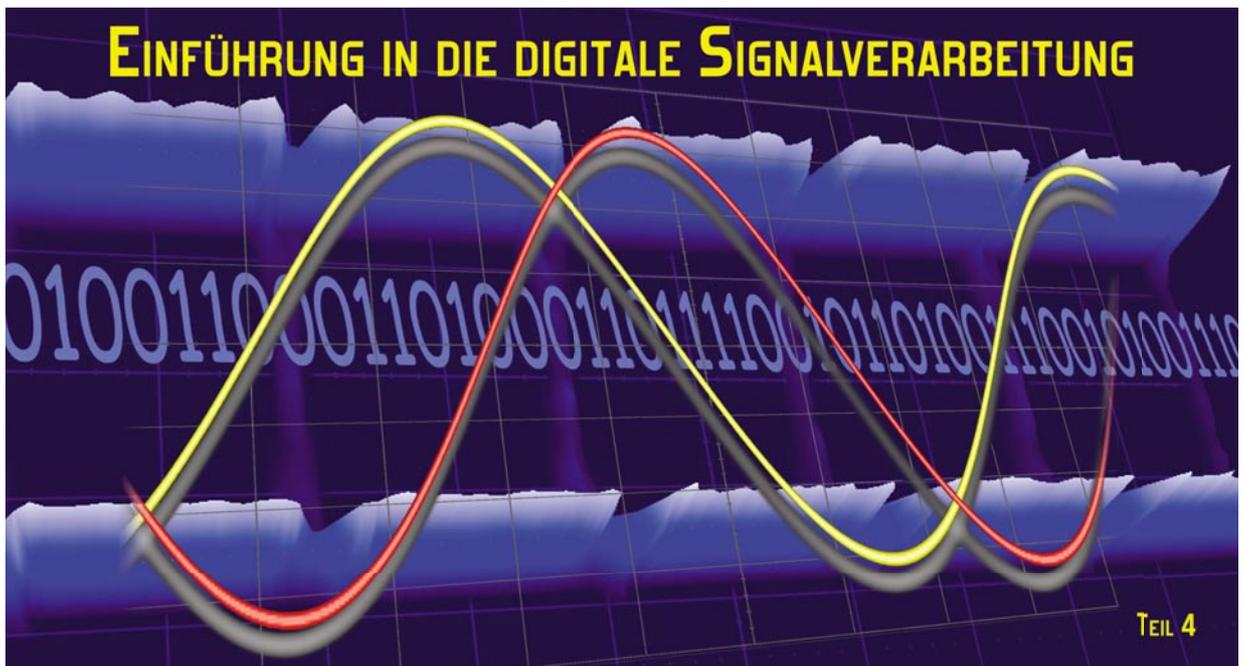


Bild 25: Wird das Abtasttheorem verletzt, überlagern sich die periodischen Spektren und das Spektrum des unabgetasteten Originalsignals lässt sich nicht wiedergewinnen. Es ist durch Aliasanteile verfälscht.

$$f(t) = \delta(t) \quad \circ \text{---} \bullet \quad F(i\omega) = 1 \quad (65)$$

In Teil 4 beschäftigen wir uns mit der Digitalisierung analoger Signale. **ELV**



Nachdem wir uns ausführlich mit den mathematischen Hintergründen der idealen Abtastung beschäftigt haben, wollen wir uns jetzt anschauen, wie die praktische Umsetzung des Abtastvorgangs aussieht.

Der Abtastvorgang in der Praxis

Abbildung 26 zeigt die wesentlichen Funktionsblöcke. Das analoge Eingangssignal wird durch einen analogen Tiefpass bandbegrenzt, um die Einhaltung des Abtasttheorems sicherzustellen.

Nun gibt es leider den idealen Abtaster, dessen Wirkungsweise durch Gleichung 50 beschrieben wird, in der Praxis nicht. Er müsste ja in unendlich kurzer Zeit den Signalwert zum Abtastzeitpunkt ermitteln, was technisch nicht möglich ist. Das Problem wird gelöst, indem man den Abtastwert am Beginn des Abtastintervalls über die ganze Dauer des Abtastintervalls „hält“ (Abbildung 27). Daher stammen auch die Begriffe Abtast-Halteglied oder Sample&Hold (S/H). Es entsteht so aus einer Folge bewerteter Dirac-Stöße eine Treppenkurve, wobei die Stufenwerte identisch sind mit den Funktionswerten zum Zeitpunkt des Abtaststoßes. In der Regel haben wir es deshalb mit einer Treppe mit ungleichmäßiger Stufenhöhe zu tun.

Weil das Abtast-Halteglied den Eingangswert über die Abtastperiode konstant hält, kann man es auch als Reihenschal-

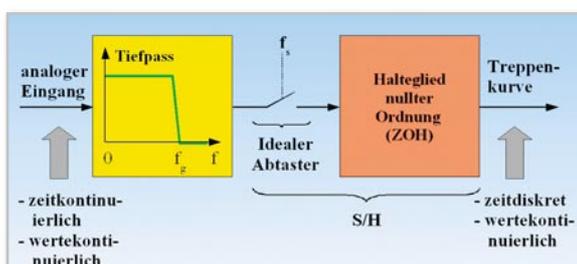


Bild 26: Der Name ist Programm: Sample&Hold (Abtasten und Halten)

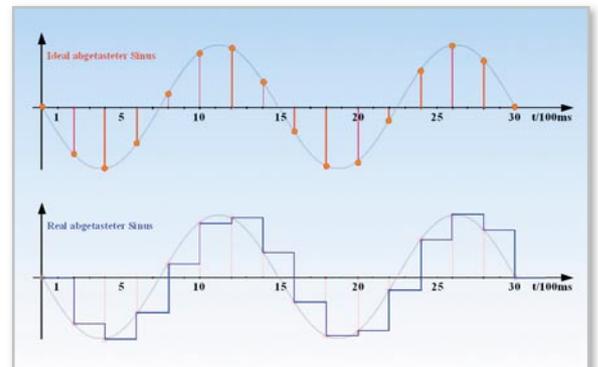


Bild 27: Der reale Abtaster hält den Abtastwert über ein Abtastintervall.

tingung aus einem idealen Abtaster mit einem Halteglied nullter Ordnung (Zero-Order-Hold: ZOH) beschreiben. Der Vorgang des Haltens eines idealen Abtastwerts hat natürlich seine Auswirkungen, insbesondere auf den Frequenzgang.

Impulsantwort und Frequenzgang eines ZOH

Abbildung 28 fasst die Wirkung eines ZOH im Zeitbereich zusammen. Der am Eingang zum Zeitpunkt $t = 0$ anliegende Impuls $\delta(t)$ wird über eine Abtastperiode T_s „gehalten“. Das ZOH reagiert auf den Eingangsimpuls $\delta(t)$ also durch eine rechteckförmige Impulsantwort.

Schauen wir uns jetzt einmal das Rechteck in Abbildung 13, Teil 2, an, erkennen wir, dass die Impulsantwort lediglich ein um eine halbe Rechteckbreite verzögertes Abbild davon ist. Mit der Zeitverschiebungseigenschaft aus Tabelle 1 in Teil 1 können wir deshalb sofort $H(j\omega)$ als Fouriertransformierte von

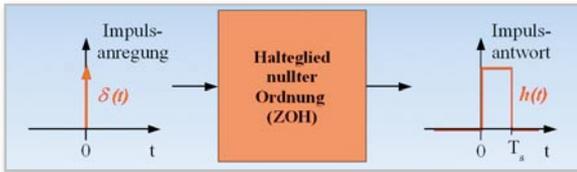


Bild 28: Die Impulsantwort eines Haltegliedes nullter Ordnung (ZOH) ist ein Rechteck von der Dauer eines Abtastintervalls.

$h(t)$ gemäß Gleichung 66 angeben:

$$H(i\omega) = T_s \operatorname{sinc}\left(\omega \frac{T_s}{2}\right) \cdot e^{-i\omega \frac{T_s}{2}} \quad (66)$$

Die Fouriertransformierte der Impulsantwort eines ZOH ist eine komplexe Funktion der Kreisfrequenz ω und hat somit einen Betrags- und einen Phasengang, die beide natürlich ebenfalls von ω abhängig sind (Abbildung 29).

Das ZOH hat also eine Tiefpasswirkung mit dem Betragsgang $|sinc(\omega T_s/2)|$. Der Vorfaktor T_s lässt sich ohne Beschränkung der Allgemeinheit auf 1 normieren. Man sieht, dass bei Vielfachen der Abtastfrequenz ω_s , also dort, wo die periodischen Aliasspektren liegen, die Nullstellen des Betragsgangsgangs auftreten. Dadurch werden die Aliasspektren deutlich gedämpft. Allerdings erfährt das Basisbandspektrum ebenfalls eine mit der Frequenz zunehmende Abschwächung, die jedoch durch einen inversen Frequenzgang des analogen Eingangstiefpasses kompensiert werden kann (Präemphase).

Bei kräftiger Überabtastung (wo sich das Basisbandspektrum im noch schwach abfallenden Bereich der sinc -Funktion befindet) ist der Fehler meist vernachlässigbar. Abbildung 30 demonstriert das für eine Sinusschwingung, die mit einer 10-mal so hohen Frequenz abgetastet wird. Dann wird die Spektrallinie bei ω_0 um 1,64 % durch den ZOH-Frequenzgang gedämpft, die Alias-Spektrallinie bei $\omega_s - \omega_0$ um 89,07 %, die bei $\omega_s + \omega_0$ um 91,06 %, die bei $2\omega_s - \omega_0$ um 94,82 % usw. Die Auswirkungen des ZOH im Zeitbereich lassen sich leicht an Abbildung 27 unten unter Auswertung von Gleichung (66) veranschaulichen (Abbildung 31). Wir können abzählen, dass eine Abtastperiode von 200 ms vorliegt, entsprechend 5 Hz Abtastfrequenz. Die Periodendauer der abgetasteten Sinusschwingung ist 1,5 s, entsprechend $2/3$ Hz. Es wird daher mit dem 7,5fachen der Sinusfrequenz abgetastet. Damit ist die Amplitude der durch einen idealen Tiefpass aus dem ZOH-Ausgangssignal ausgefilterten Grundwelle um 2,9 % verringert. Ihre Verzögerung beträgt $180^\circ/7,5 = 24^\circ$, was dem 15. Teil der Periode des Sinus bzw. der Hälfte der Abtastperiode, also 100 ms, entspricht.

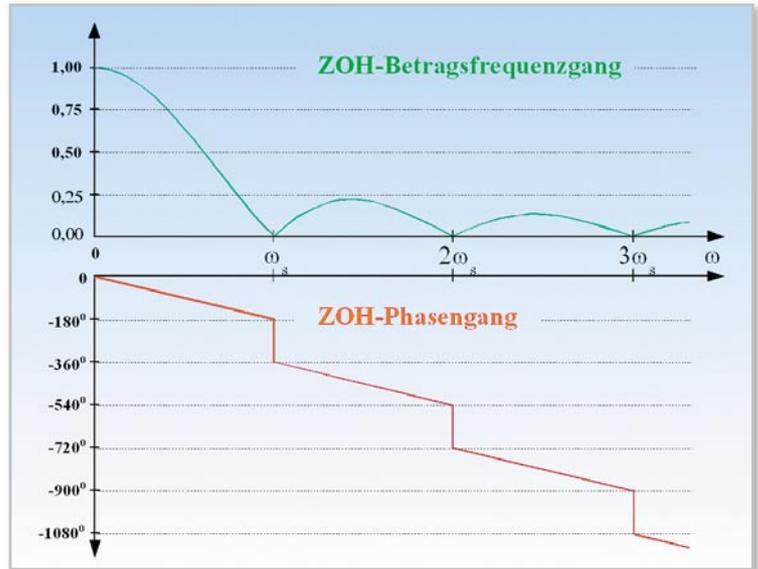


Bild 29: Der Frequenzgang eines ZOHs nach Betrag und Phase

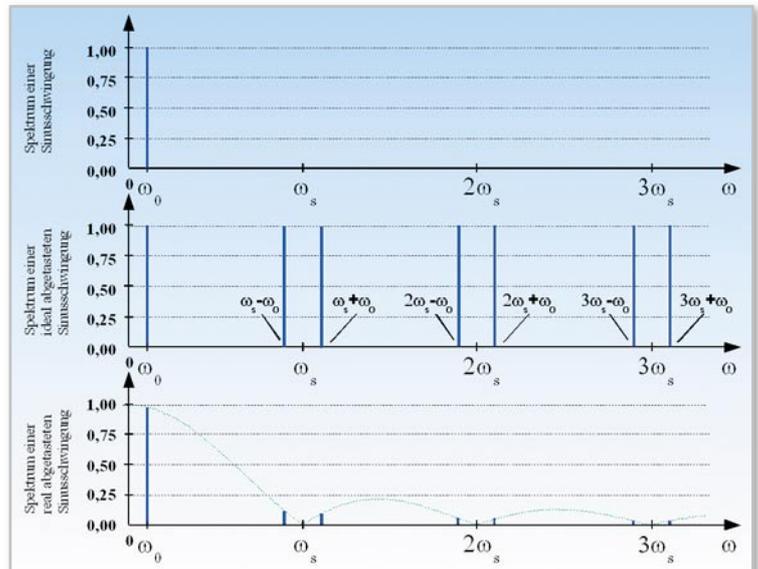


Bild 30: Die periodischen Spektren einer abgetasteten Funktion werden an den Nullstellen des ZOH-Frequenzgangs gestauch.

Von der Treppenkurve zum Digital-Signal

Die Treppenkurve hinter dem Abtast-Halteglied ist wertekontinuierlich. Die Stufenhöhe entspricht der Veränderung des Signalwerts von einem Abtastzeitpunkt zum nächsten. Dennoch ist die Treppenkurve kein getreues Abbild des Analog-Signals. Zu Beginn einer Stufe haben Treppe und Signal den gleichen Wert. Weil während des Abtastintervalls dieser Wert vom ZOH gehalten wird, bildet sich im Lauf des Ab-

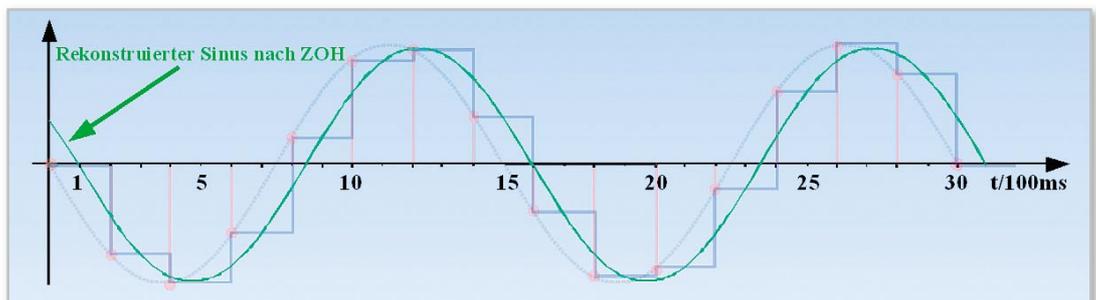


Bild 31: Nach dem ZOH ist der rekonstruierte Sinus leicht in der Amplitude reduziert und verzögert.

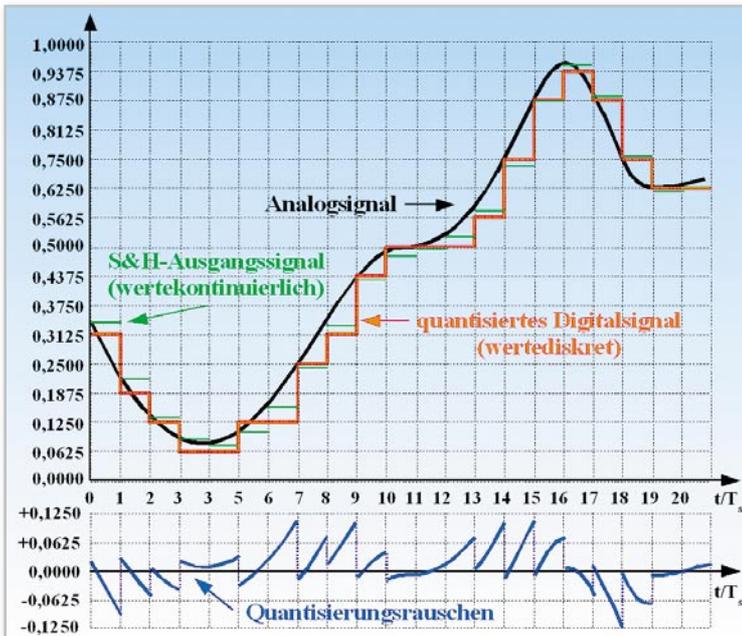


Bild 32: Durch das Darstellen des ZOH-Ausgangssignals mit endlicher Genauigkeit entsteht ein Quantisierungsfehler.

tastintervalls ein zunehmender Fehler. Dieser Fehler hat also seine Ursache in der Diskretisierung der Zeit.

In einem digitalen Signalverarbeitungssystem wird bekanntlich mit Binärzahlen (Folgen aus Nullen und Einsen = Wort) gerechnet. Deren Länge ist in der Praxis begrenzt, wodurch sich nur eine endliche Anzahl von Werten darstellen lässt. Dies ist eine Diskretisierung der Amplitudenwerte. Dazu muss die Stufenhöhe der Treppenkurve in ein Raster gebracht werden, dessen Feinheit von der Wortlänge vorgegeben wird. Diesen Vorgang nennt man Quantisierung. Abbildung 32 zeigt, was gemeint ist. Eine analoge Kurve wird real abgetastet. Hinter dem ZOH ergibt sich die grüne, wertekontinuierliche Treppenkurve mit uneinheitlicher Stufenhöhe. Um die Signalamplitude zwischen 0 und 1 mit einer 4-Bit-Zahl zu beschreiben, muss die Amplitude auf $2^4 = 16$ Stufenwerte beschränkt werden. Die Signalamplitude kann dann nur noch in Vielfachen von $1/16 = 0,0625$ beschrieben werden, sie ist quantisiert (rote Treppenkurve) und dadurch wertediskret.

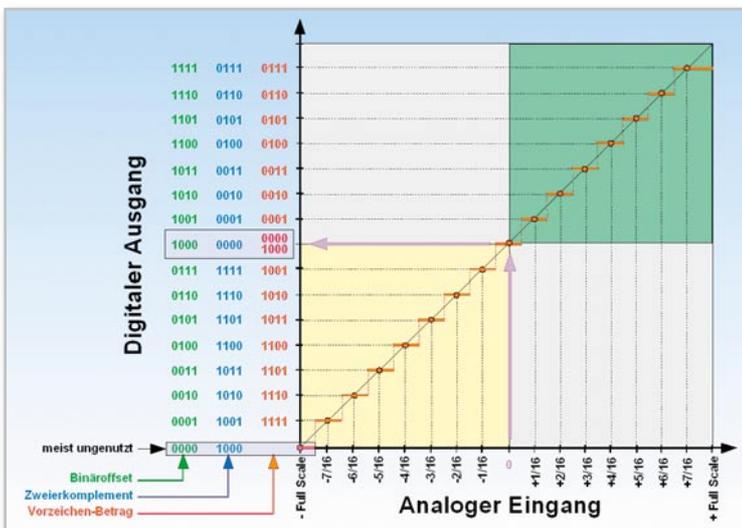


Bild 33: Die Kennlinie eines A/D-Wandlers beschreibt, wie analoge Eingangswerte in digitale Ausgangswörter umgesetzt werden.

Die Signalamplitude am Beginn des Abtastintervalls stimmt jetzt nicht mehr mit der Treppenhöhe überein. Die „Körnigkeit der digitalen Amplituden“, zusammen mit dem Fehler durch die Treppendarstellung, verursacht einen rauschartigen Effekt – das Quantisierungs- oder Rundungsrauschen (blaue Kurve). Es ist anschaulich, dass das Quantisierungsrauschen mit zunehmender Wortlänge und kürzeren Abtastintervallen abnimmt, weil die Approximation durch die feinere Treppenkurve genauer wird.

Analog-zu-digital-Wandler

In der Praxis übernimmt der Analog-Digital-Wandler (A/D-Wandler, ADC: Analogue to Digital Converter) die Funktion des Quantisierens. Sein analoger Eingang wird mit dem Ausgangssignal des S&H-Glieds gespeist (grüne Treppenkurve). Im Laufe der Abtastperiode muss das mit der Stufenhöhe am besten korrespondierende Binärwort berechnet werden. Damit wird die Folge der grünen Stufen in Abbildung 32 in eine Folge von Binärzahlen entsprechend den roten Stufen umgewandelt.

Die Kennlinie eines A/D-Wandlers ist in Abbildung 33 dargestellt.

Symmetrische Signale in einem Bereich von beispielsweise ± 800 mV können von diesem Wandler in 16 digitale Zahlenwerte gewandelt werden. Die Auflösung beträgt dadurch 100 mV.

Ein Beispiel aus der Praxis: Ein 12-Bit-ADC ist für einen Full-Scale-Messbereich von ± 10 V ausgelegt. Er kann also $2^{12} = 4096$ Amplitudenintervalle unterscheiden, die Intervallbreite beträgt demnach 4,88 mV. Bei einem Signalwert von beispielsweise +16 mV wird in der Vorzeichen-Betragsdarstellung der Wert 00000000011 ausgegeben, weil er im Intervall 14,64...19,52 mV liegt.

Digitale Zahlenformate

Computer arbeiten mit Zahlendarstellungen, die nur aus den Ziffern 0 und 1 bestehen und deshalb als zweiwertig, dual oder binär bezeichnet werden. Wir wollen uns nur knapp und übersichtsartig mit den wichtigsten Darstellungsformen beschäftigen. Das duale Zahlensystem gehört wie auch das Dezimal- oder das Hexadezimalsystem zur Klasse der polyadischen Zahlensysteme. Das bedeutet, dass jede endliche reelle Zahl als Summe von Potenzen einer ganzzahligen Basis B dargestellt werden kann. Gleichung (67) zeigt dies.

Am Beispiel für die dezimale Zahl 49,375 (Basis $B = 10$, Koeffizienten $0 \leq \xi \leq B - 1 = 9$) sei dies in Gleichung (68) erläutert. Die Darstellung als duale Zahl in Gleichung (69) erfolgt genau gleich, außer dass die Basis jetzt $B = 2$ ist und die Koeffizienten nur den Wert 0 oder 1 annehmen können.

Zur Vereinfachung kann man nun nur noch die Koeffizienten anschreiben und durch das Komma anzeigen, ab wo die Hochzahlen der Potenzen negativ sind. Aus Gleichung (69) wird dann Gleichung (70).

Bei ganzen Zahlen fallen die Potenzen mit negativen Hochzahlen weg, das Komma und was rechts davon steht, kann

$$Z = \sum_{v=-t}^{\tau} \xi_v \cdot B^v = \xi_{\tau} B^{\tau} + \xi_{\tau-1} B^{\tau-1} + \dots + \xi_0 B^0 + \xi_{-1} B^{-1} + \dots + \xi_{-t} B^{-t} \quad (67)$$

$$49,375 = 4 \cdot 10^1 + 9 \cdot 10^0 + 3 \cdot 10^{-1} + 7 \cdot 10^{-2} + 5 \cdot 10^{-3} \quad (68)$$

$$= 40 + 9 + 0,3 + 0,07 + 0,005 \quad \text{Dezimalsumme}$$

$$49,375_{10} = 1 \cdot 2^5 + 1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 + 0 \cdot 2^{-1} + 1 \cdot 2^{-2} + 1 \cdot 2^{-3} \quad (69)$$

$$= 32 + 16 + 0 + 0 + 0 + 1 + 0 + 0,25 + 0,125 \quad \text{Binärsumme}$$

$$49,375_{10} = 1100001,011_2 \quad (70)$$

gestrichen werden. Wir haben es dann mit einer vorzeichenlosen Ganzzahl (Unsigned Integer) oder Festkommazahl zu tun.

Zur Terminologie: Die Koeffizienten der Potenzen mit den möglichen Werten 0 oder 1 werden als Bits bezeichnet. Der Koeffizient der größten Potenz heißt MSB (Most Significant Bit = höchstwertiges Bit) und steht im Bitmuster ganz links. Der Koeffizient der kleinsten Potenz heißt LSB (Least Significant Bit = niedrigstwertiges Bit) und steht im Bitmuster ganz rechts. 8 Bits fasst man zu einem Byte (auch 8-Bits-Wort genannt) zusammen und ein Byte wiederum besteht aus zwei Halbbytes (4-Bits-Wort), den Nibbles.

Die höherwertigen linken vier Bits bilden das „upper nibble“, die rechten das „lower nibble“.

Weil sich alle Operationen mit Binärzahlen auf Additionen zurückführen lassen, genügt es, die elementare Rechenvorschrift für die Addition zweier Bits zu kennen: $0 + 0 = 0$, $0 + 1 = 1$, $1 + 0 = 1$, $1 + 1 = 10$. Man beachte, dass bei der Addition von zwei Einsen ein Überlauf in die höherwertige Stelle erfolgt. Einige binäre Darstellungsformen für Festkommazahlen zeigt Abbildung 34.

Vorzeichenlose Ganzzahl (Unsigned Integer)		Binärversatz (Offset Binary)		Vorzeichen-Betrag (Sign and Magnitude)		Zweierkomplement (Two's Complement)	
Dezimal	Bitmuster	Dezimal	Bitmuster	Dezimal	Bitmuster	Dezimal	Bitmuster
+15	1111	+8	1111	+7	0111	+7	0111
+14	1110	+7	1110	+6	0110	+6	0110
+13	1101	+6	1101	+5	0101	+5	0101
+12	1100	+5	1100	+4	0100	+4	0100
+11	1011	+4	1011	+3	0011	+3	0011
+10	1010	+3	1010	+2	0010	+2	0010
+9	1001	+2	1001	+1	0001	+1	0001
+8	1000	+1	1000	+0	0000	0	0000
+7	0111	0	0111	-0	1000	-1	1111
+6	0110	-1	0110	-1	1001	-2	1110
+5	0101	-2	0101	-2	1010	-3	1101
+4	0100	-3	0100	-3	1011	-4	1100
+3	0011	-4	0011	-4	1100	-5	1011
+2	0010	-5	0010	-5	1101	-6	1010
+1	0001	-6	0001	-6	1110	-7	1001
+0	0000	-7	0000	-7	1111	-8	1000
Wertebereiche: 8 bit: 0 ... 255 16 bit: 0 ... 65.535 32 bit: 0 ... 4.294.967.295	Wertebereiche: 8 bit: -127 ... 128 16 bit: -32.767 ... 32.768 32 bit: -2.147.483.647	Wertebereiche: 8 bit: -127 ... 127 16 bit: -32.767 ... 32.767 32 bit: -2.147.483.647	Wertebereiche: 8 bit: -128 ... 127 16 bit: -32.768 ... 32.767 32 bit: -2.147.483.648
	2.147.483.648	2.147.483.647	2.147.483.647				

Bild 34: Eine Übersicht über die wichtigsten Darstellungsformen für Binärzahlen

Zweierkomplement (Two's Complement)

Diese Darstellungsform ist in der digitalen Signalverarbeitung und in modernen Computern mit Abstand am beliebtesten. Das ist in der einfachen Art begründet, mit der sich Additionen und Subtraktionen hardwaremäßig ausführen lassen. Eine einfache Analogie ist der Walzenmechanismus eines Kilometerzählers in einem älteren Tachometer. Steht er auf null und man fährt in Vorwärtsrichtung los, wird er anzeigen „0000, 0001, 0002, 0003, ...“. Beim Rückwärtsfahren dagegen wird man ablesen „0000, 9999, 9998, 9997, ...“. In binärer Darstellung führt das bei den positiven Zahlen zu 0000, 0001, 0010, 0011, ...“ bzw. „0000, 1111, 1110, 1101, ...“. Diesen Zählmechanismus findet man in den aus Flipflops bestehenden Rechenregistern eines Computers wieder, was ihn so geeignet für die Maschine macht. Beim Rückwärtszählen aus dem positiven Bereich erfolgt der Übergang zu den negativen Zahlen einfach durch einen Überlauf $0000 \rightarrow 1111$. Die Interpretation der Zahl ist damit ganz einfach: Das MSB bestimmt die Polarität (MSB = 0: positiv, MSB = 1: negativ). Bei den positiven Zahlen wird der Rest des Bitmusters als Betrag bestimmt, bei den negativen Zahlen wird der Rest des Bitmusters invertiert ($0 \rightarrow 1$, $1 \rightarrow 0$) und 1 dazuaddiert, das Ergebnis wird (ohne MSB!) als Betrag der negativen Zahl interpretiert. Zum Beispiel $+7 \rightarrow -7$ entspricht $0111 \rightarrow 1000 + 0001 = 1001$. Addition und Subtraktion führen in Zweierkomplementdarstellung ohne weitere Korrektur

Vorzeichenlose Ganzzahl (Unsigned Integer)

Hier wird einfach von null ausgehend hochgezählt. Negative Zahlen können nicht dargestellt werden. Die Umwandlung in eine Dezimalzahl (oder eine Zahl mit einer anderen Basis als 10) ist einfach durch Anwendung von Gleichung (67) möglich.

Binärversatz (Offset Binary)

Wie Unsigned Integer, aber gegen die dezimalen Werte so verschoben, dass auch negative Zahlen darstellbar sind.

Vorzeichen-Betrag (Sign and Magnitude)

Hier wird ganz einfach das MSB als Vorzeichen (0: positiv, 1: negativ) und der rechts davon stehende Rest als Betrag interpretiert.

Dadurch „verschwendet“ man eine Ziffer, denn es entstehen eine positive Null (0000) und eine negative Null (1000).

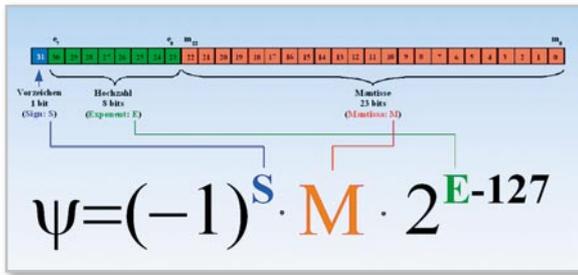


Bild 35: Gleitkommazahlen decken mit 32 Bit einen riesigen Wertebereich ab.

turen zum richtigen Ergebnis. Ein Beispiel: $+5 - 7 = +5 + (-7) = -2 \rightarrow 0101 + 1001 = 1110$

Gleitkommazahlen (Floating Point Numbers)

Das binäre Gleitkommazahlenformat lässt die Darstellung erheblich größerer Wertebereiche zu. Etwas Ähnliches kennen wir von dem wissenschaftlichen Zahlenformat eines Taschen-

rechners, der mit Dezimalzahlen rechnet. Wenn auf seinem Display z. B. die beiden Zahlen $-2,746$ (Mantisse) und 09 (Exponent) erscheinen, haben wir es mit dem Produkt aus der negativen Zahl $-2,746$ und der Zehnerpotenz $10^9 = 1.000.000.000$, d. h. mit $-2.746.000.000$ zu tun. Entsprechend ist die Darstellung einer Gleitkommazahl ψ im Binärformat mit einfacher Genauigkeit (Single Precision) nach dem Normungsstandard ANSI/IEEE Std. 754-1985 (Abbildung 35).

Der erste Term $(-1)^S$ in Gleichung (71) ist 1, wenn das Vorzeichen-Bit 0 ist, und -1 , wenn das Vorzeichen-Bit 1 ist, d. h. bei $S = 0$ ist die Zahl positiv, bei $S = 1$ negativ. Der zweite Term M ist die 23-stellige Mantisse, genauer gesagt ihr Nachkommateil, mit den Bitwerten $m_{22} \dots m_0$.

Sie ist so normiert, dass links vom Komma nur ein von null verschiedener Wert steht, im binären Zahlensystem ist dies immer die 1. Deshalb spart man sich dieses Bit und nutzt es zur Steigerung der Genauigkeit bei der Darstellung des Nachkommateils der Mantisse. Der dritte Term 2^{E-127} enthält die 8-stellige Hochzahl E mit den Bitwerten $e_7 \dots e_0$. Sie kann maximal den Wert 255 annehmen und minimal den Wert 0, was für den dritten Term einen Wertebereich von 2^{128} bis 2^{127} vorgibt. Dezimal ausgedrückt ist das die Spanne von $3,4028 \cdot 10^{38}$ bis $5,8775 \cdot 10^{-39}$.

Gleichung (72) stellt die Single-Precision-Gleitkommazahl als Binärzahl dar. Die zwei Beispiele in Gleichung (73) sollen das verdeutlichen.

$$\psi = (-1)^S \cdot M \cdot 2^{E-127} \quad \text{Gleitkommazahl mit einfacher Genauigkeit} \quad (71)$$

$$\psi = (-1)^S \cdot 1, m_{22} m_{21} m_{20} \dots m_3 m_2 m_1 m_0 \cdot 2^{e_7 e_6 e_5 e_4 e_3 e_2 e_0 - 1111111} \quad (72)$$

$$\text{Beispiel 1: } \underbrace{1 \ 00001001}_9 \ \underbrace{101000000000000000000000}_{0,625} \quad (73)$$

$$\text{daraus folgt } \psi = -1,625 \cdot 2^{9-127} = -4,890056 \cdot 10^{-36}$$

$$\text{Beispiel 2: } \underbrace{0 \ 10000011}_{131} \ \underbrace{001100000000000000000000}_{0,1875}$$

$$\text{daraus folgt } \psi = +1,1875 \cdot 2^{131-127} = -1,900000 \cdot 10^1 = -19$$

$$y = \sin(x) e^{-\frac{x}{5}} \quad (74)$$

$$\Delta = \frac{A}{2^w} \quad \text{Quantisierungsintervall} \quad (75)$$

$$\epsilon = x(nT_s) - x_q(nT_s) \quad -\frac{\Delta}{2} \leq \epsilon < \frac{+\Delta}{2} \quad \text{Quantisierungsfehler} \quad (76)$$

Prinzipschaltung A/D-Wandler mit S&H

Abbildung 36 zeigt das Prinzipschaltbild eines ADCs. Es besteht aus einem Eingangsverstärker, dessen Ausgang über einen Schalter S einen Haltekapazitor lädt, solange S geschlossen ist. Diesen Zeitabschnitt nennt man Sample-Phase (to sample: abtasten). Ist der Ausgangswiderstand des Verstärkers niedrig genug, folgt die Spannung am Kondensator der Eingangsspannung nahezu verzögerungsfrei. Die Hold-Phase (to hold: halten) beginnt mit dem Öffnen des Schalters. Wenn der auf den Kondensator folgende Impedanzwandler einen sehr hohen Eingangswiderstand hat, bleibt die Kondensatorspannung (und damit die Ausgangsspannung des Impedanzwandlers) bis zum Ende der Sample-Phase konstant. In diesem Zeitabschnitt muss der vom Impedanzwandler gespeiste Analog-zu-digital-Codierer den Sample-Wert in ein N -stelliges Binärwort umcodiert haben. Unmittelbar danach beginnt eine neue Abtastphase, gefolgt von einer Haltephase, in der wieder die Umcodierung erfolgt, usw. Die Spannung am Ausgang des Impedanzwandlers entspricht der roten Kurve. Die Anordnung aus Eingangsverstärker, Schalter und Impedanzwandler wird im Englischen meist als „Sample and Hold Amplifier“ (SHA) bezeichnet. Genauer wäre der Name „Track and Hold Amplifier“ (THA), weil ja die Spannung am Haltekapazitor bis zum Beginn der Haltephase der Eingangsspannung folgt.

Die Technik der A/D-Wandler ist eine eigene Artikelserie wert. So viel sei an dieser Stelle gesagt: Alle Wandlerfehler wie Jitter, Nonlinearitäten, Offsets, fehlende Codewörter (Missing Codes) usw. schlagen sich in der Abbildungsgenauigkeit nieder und verschlechtern die Ausgangssignalparameter.

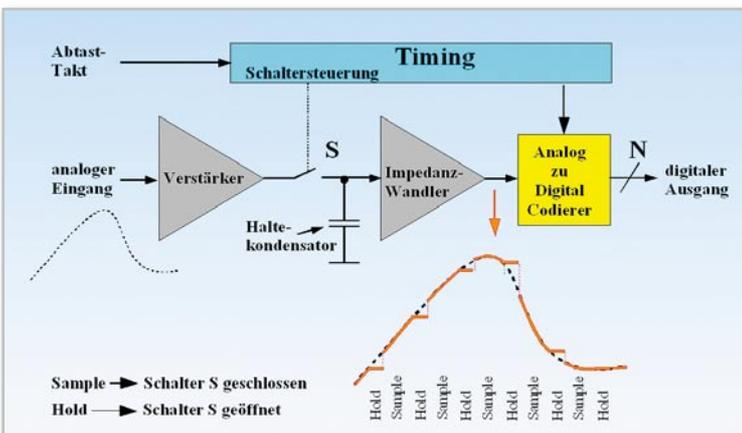


Bild 36: Vom Speicherkondensator wird der Signalwert bis zum Ende des Abtastintervalls gehalten.

Veranschaulichung der Diskretisierung von Zeit und Amplitude

Anhand der folgenden drei Abbildungen wollen wir die Auswirkungen der Rasterung (Diskretisierung) von Zeit (Abtastfrequenz) und Amplitude (Wertebereich) studieren. Dazu wurde in Scilab ein kleines Programm geschrieben, welches eine vorgebbare analoge, d. h. wert- und zeitkontinuierliche Zeitfunktion abtastet, den Abtastwert quantisiert und den quantisierten Abtastwert über die Abtastperiode konstant hält. Aus den Abweichungen zwischen der analogen kontinuierlichen Zeitfunktion und der treppenförmigen wert- und zeitgerasterten wird der Amplitudenfehler ermittelt.

Wir wählen als analoge Zeitfunktion eine exponentiell gedämpfte Sinusschwingung über 2,5 Perioden ($0 \dots 5\pi = 15,7$) gemäß Gleichung (74).

Die x -Werte werden als eine gewisse Zahl von Abtastwerten vorgegeben, die Stufenhöhe über die Länge des darstellenden Binärwortes gesteuert. Da wir einen y -Wertebereich von $-1 \leq y \leq +1$ voraussetzen, beträgt bei einer Wortlänge von w das Quantisierungsraster 2^{-w+1} . In Abbildung 37 wurde eine Wortlänge von $w = 5$ gewählt, deshalb ist die Stufenhöhe der Treppenkurve $2^{-5+1} = 2^{-4} = 0,0625$. Weil mit nur 50 Abtastwerten über das ganze Intervall $0 \dots 5\pi$ gearbeitet wird, ergibt sich eine recht grobe Kurve des Quantisierungsfehlers. In Abbildung 38 beträgt die Wortlänge $w = 8$, deshalb ist die minimale Stufenhöhe der Treppenkurve $2^{-7} = 0,0078125$. Das Funktionsintervall wird 100-mal abgetastet. Im Ergebnis ist der Quantisierungsfehler kleiner als in Abbildung 37. Wegen des relativ groben zeitlichen Abtastrasters ergeben sich aber dort, wo die analoge Funktion steil ist (d. h. ihre Werte in kurzer Zeit stark verändert), immer noch zeitbedingte große Fehler von einem Vielfachen des Quantisierungsintervalls.

Wenn man sowohl die Abtasthäufigkeit als auch die Wortlänge weiter steigert, gehen die damit verbundenen Abbildungsfehler weiter zurück.

Abbildung 39 zeigt die Verhältnisse bei 5000 Abtastwerten und 32 Bit Wortbreite. Durch die extrem feine Stufung der Treppenkurve ist diese praktisch deckungsgleich mit der analogen Kurve. Erst bei starker Vergrößerung zeigen sich die Abweichungen. Entsprechend klein sind auch die Quantisierungsfehler. Sie sind in diesem extremen Beispiel sicherlich kleiner als das jedem analogen Signal überlagerte mehr oder weniger starke Rauschen. An den flachen Teilen der Kurve lässt sich die hohe Amplitudenaufösung gut erkennen. Abbildung 40 zeigt dies an der Umgebung des dritten Maximums.

Zusammenfassend: Sowohl die wortbreitenbedingte Amplitudenaufösung als auch die zur Abtastfrequenz proportionale Zeitaufösung gehen im Zusammenwirken mit der Signalforn in den Quantisierungsfehler ein.

Quantisierungsrauschen

Bei linearer Quantisierung (gleiche Stufenhöhe für alle Signalwerte) wird der Aussteuerbereich A ($-$ Full Scale \dots + Full Scale) in 2^w gleich große Quantisierungsintervalle Δ zerlegt, wobei w die Wortlänge eines Abtastwerts in Bits beschreibt

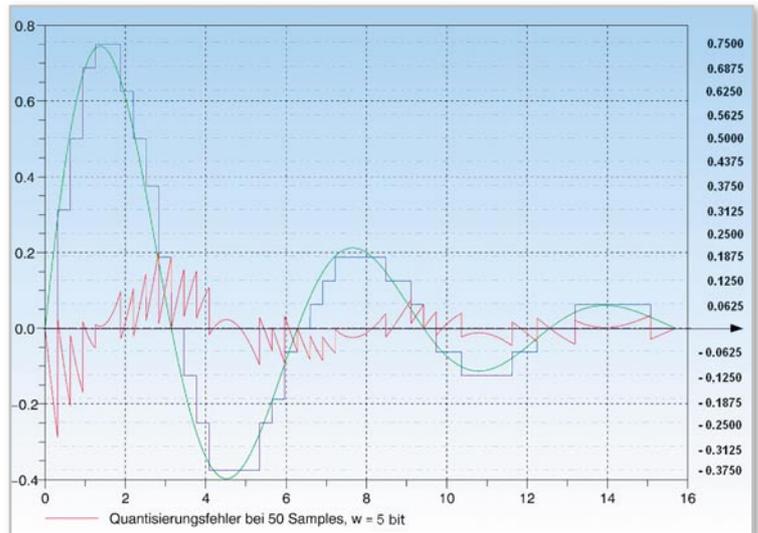


Bild 37: Bei 50 Abtastwerten und 4 Bit langen Binärwörtern sind die Fehler der digitalen Signaldarstellung erheblich.

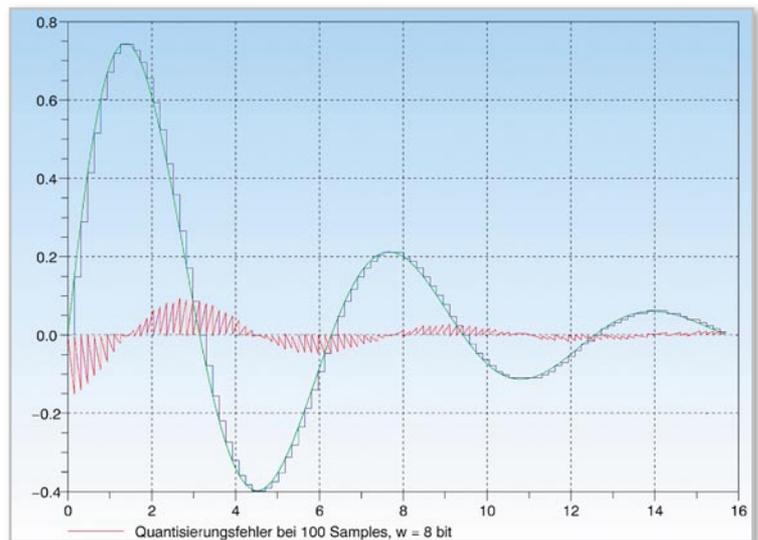


Bild 38: Bei 100 Abtastwerten und 8 Bit langen Binärwörtern kommen wir dem Originalsignal schon näher.

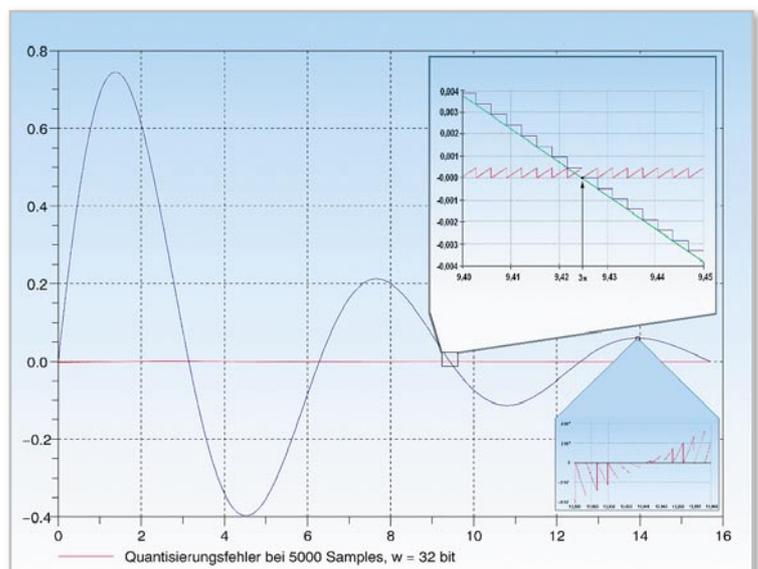


Bild 39: Derartig präzise Abtastung ist in der Praxis nicht erforderlich.

(Gleichung [75]).

Der Quantisierungs- bzw. Rundungsfehler ϵ in einem Abtast-

$$S/N = 10 \cdot \log_{10} \left(\frac{P_x}{P_\epsilon} \right) \quad (77)$$

$$P_\epsilon = \frac{1}{\Delta} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \epsilon^2 d\epsilon = \frac{1}{\Delta} \frac{1}{3} \epsilon^3 \Big|_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} = \frac{1}{3\Delta} \left[\frac{\Delta^3}{8} + \frac{\Delta^3}{8} \right] = \frac{\Delta^2}{12} \quad \text{Quantisierungsrauschleistung} \quad (78)$$

$$\frac{P_x}{P_\epsilon} = \frac{12 \cdot P_x}{A^2} \cdot 2^{2w} \quad \text{Signalleistung zu Quantisierungsrauschleistung} \quad (79)$$

$$S/N = 6,02 \cdot w + 10 \cdot \log_{10} \left(\frac{12 \cdot P_x}{A^2} \right) \text{ dB} \quad \text{Signal-Rauschabstand nach Quantisierung} \quad (80)$$

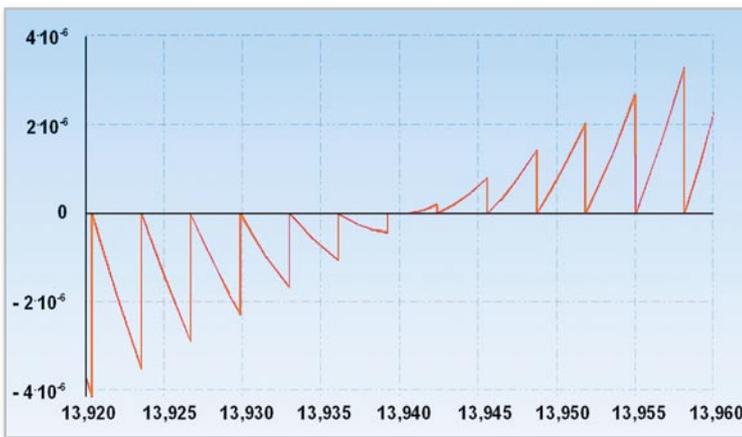


Bild 40: Am Beispiel des dritten Maximums wird deutlich, wie verschwindend gering der Quantisierungsfehler bei 5000 Abtastwerten und 32-Bit-Darstellung geworden ist.

intervall T_s ist definiert als der Unterschied zwischen wahren und quantisiertem Abtastwert (76).

Wegen des sprunghaften Verhaltens des Quantisierers an den Entscheidungsschwellen zählt er zu den nichtlinearen Elementen. Man kann ihn aber durch ein lineares Ersatzmodell nachbilden (Abbildung 41).

Das Quantisierungsrauschen wird über das Signal-Rausch-Verhältnis nach Gleichung (77) beschrieben:

Dabei stellen P_x die Signalleistung am Eingang des Quantisierers und P_ϵ die Leistung des Quantisierungsfehlers dar.

Setzt man voraus, dass die Quantisierungsfehler unkorrelierte, im Intervall $[-\Delta/2, \Delta/2]$ gleich verteilte Zufallsgrößen sind und die Quantisierungsfehler $\epsilon(nT_s)$ nicht von den wertekontinuierlichen Eingangsbabtastwerten $x(nT_s)$ abhängen, ergibt sich die Leistung des gleichspannungsfreien Quantisierungsfehlers $\epsilon(nT_s)$ gemäß Gleichung (78).

Mit der Definition des Quantisierungsintervalls nach Gleichung (75) wird aus Gleichung (78) die (79).

Gleichung (79) in Gleichung (77) eingesetzt, ergibt das gesuchte Signal-Rausch-Verhältnis in dB (80).

Es lohnt sich, Gleichung (80) etwas näher zu betrachten. Wir

sehen, dass der erste Summenterm auf der rechten Seite nur von der Wortbreite w (also der Länge der darstellenden Binärzahl) abhängt. Jedes Bit mehr erhöht den Signal-Rausch-Abstand um ca. 6 dB. Der zweite Summenterm hängt von der Leistung des Eingangssignals des Quantisierers ab. Je kleiner diese ist, desto kleiner wird der Rauschabstand. Andererseits bewirkt ein großer Aussteuerbereich A ebenfalls eine Abnahme von S/N , weil das Signal dann auf weniger Stufen quantisiert wird und die relativen Quantisierungsfehler entsprechend größer werden. Bei Vollaussteuerung mit einem Sinus ist der zweite Summenterm 1,76 dB, bei typischen digitalen Audio-Signalen $-8,5$ dB. Audio-CDs mit einer Wortbreite $w = 16$ Bit können deshalb maximal ein S/N von $96,3 - 8,5 = 87,8$ dB erreichen. Diese Betrachtungen setzen ein unverraushtes Eingangssignal voraus.

Für das Quantisierungs- oder Rundungsrauschen gut ausgesteuerter A/D-Wandler mit Wortlängen von mehr als 8 Bit merken wir uns als Faustregel: 6 dB S/N pro Bit Wortlänge.

In Teil 5 beschäftigen wir uns mit den diskreten Berechnungsmethoden zur Faltung, Fouriertransformation und ihrer schnellen Variante, der FFT.

ELV

Lineares Ersatzmodell eines Quantisierers

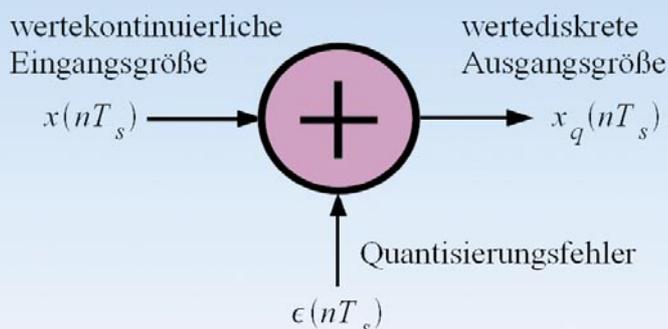


Bild 41: Das lineare Ersatzmodell eines Quantisierers fügt dem wertekontinuierlichen Signal nach dem ZOH eine rauschartige Störgröße hinzu, die den Quantisierungsfehler repräsentiert.

$$y(n) = \sum_{m=-\infty}^{\infty} x(m) \cdot h(n-m) = x(n) * h(n) \quad \text{Faltungssumme}$$

n	=	...	-2	-1	0	1	2	3	4	5	6	...	
x(n)	=	...	0	0	-1	0,5	-0,25	0	0	0	0	...	Nummer des Abtastintervalls
h(n)	=	...	0	0	1	0,75	0,5	0,25	0	0	0	...	Eingangsfolgewerte
h(n-1)	=	...	0	0	0	1	0,75	0,5	0,25	0	0	...	Impulsantwort
h(n-2)	=	...	0	0	0	0	1	0,75	0,5	0,25	0	...	Impulsantwort um ein Abtastintervall verzögert (m=1)
h(n-3)	=	...	0	0	0	0	0	1	0,75	0,5	0,25	...	Impulsantwort um zwei Abtastintervalle verzögert (m=2)
.
.
x(0) h(n)	=	...	0	0	-1	-0,75	-0,5	-0,25	0	0	0	...	Erster Summand der Faltungssumme
x(1) h(n-1)	=	...	0	0	0	0,5	0,38	0,25	0,125	0	0	...	Zweiter Summand der Faltungssumme
x(2) h(n-2)	=	...	0	0	0	0	-0,25	-0,1875	-0,125	-0,0625	0	...	Dritter Summand der Faltungssumme
x(3) h(n-3)	=	...	0	0	0	0	0	0	0	0	0	...	Vierter Summand der Faltungssumme
.
.
y(n)	=	...	0	0	-1	-0,25	-0,375	-0,1875	0	-0,0625	0	...	Gesamte Faltungssumme = Ausgangssignal

Tabelle 2: In der Tabelle sind alle Zahlenwerte für das Bestimmen einer Faltungssumme zum Nachvollziehen eingetragen.

Die in Abbildung 42 abgebildeten Folgewerte drückt Gleichung (82) aus.

Diskrete Faltung

Ebenso wie bei den kontinuierlichen Systemen ist auch ein LTI-Abtastsystem (LTI: linear, zeitinvariant) durch seine Impulsantwort $h(n)$ vollständig charakterisiert. Die Faltung eines beliebigen Eingangssignals $x(n)$ mit der Impulsantwort $h(n)$ ergibt deshalb das Ausgangssignal $y(n)$. Die Faltung ist kom-

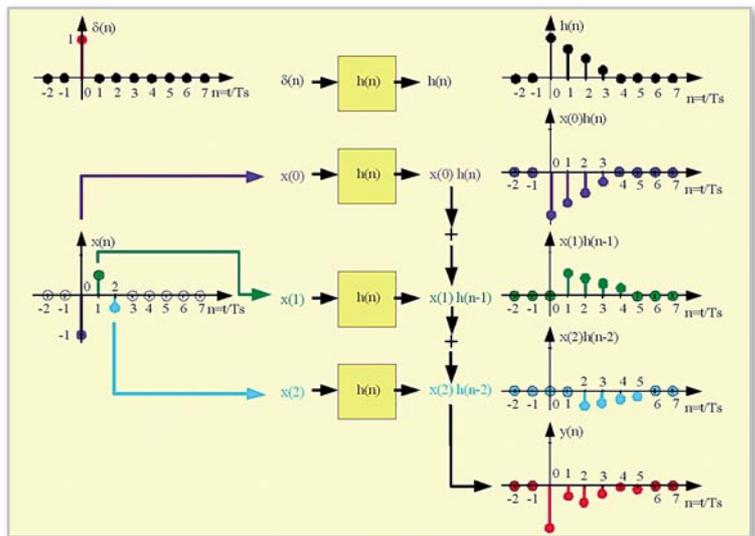


Bild 43: Jeder Abtastwert der Eingangsfolge erzeugt eine bewertete Impulsantwort, deren Überlagerung wiederum ist die Faltungssumme.

mutativ, was durch die zwei Anschriften der Faltungssumme in Gleichung (83) zum Ausdruck kommt.

An einem Beispiel wollen wir die Faltung noch einmal veranschaulichen. In Tabelle 2 wird von einem System ausgegangen, welches auf den Einheitsimpuls $\delta(n)$ an seinem Eingang mit der Impulsantwort $h(n)$ am Ausgang reagiert. Die Wertefolge von $\delta(n)$ hat nur bei $n = 0$ den Wert 1 und ist zu allen anderen Zeiten $n \neq 0$ null. Die Impulsantwort besteht aus vier Werten ungleich null, nämlich $h(0) = 1, h(1) = 0,75,$

```

scilab-4.1 (0)
File Edit Preferences Control Editor Applications ?
scilab-4.1
Copyright (c) 1989-2006
Consortium Scilab (INRIA, ENPC)

Startup execution:
loading initial environment
-->x=[-1 0.5 -.25]
x =
- 1.    0.5  - 0.25
-->h=[1 .75 .5 .25]
h =
  1.    0.75  0.5    0.25
-->y=convol(x, h)
y =
- 1.  - 0.25  - 0.375  - 0.1875  0.  - 0.0625
-->
    
```

Bild 44: Scilab als eine Art wissenschaftlicher Taschenrechner „faltet“ mit geringstem Aufwand.

$$\text{Analog: } y(t) \Rightarrow \text{Digital: } y(n \cdot T_s) = y(n) \quad (81)$$

Übergang von analoger zu digitaler Signaldarstellung

$$y(n) = \dots y(-2); y(-1); y(0); y(1); y(2); y(3); y(4); y(5); y(6); y(7); \dots \quad (82)$$

$$y(n) = \sum_{m=-\infty}^{\infty} x(m) \cdot h(n-m) = \sum_{m=-\infty}^{\infty} x(n-m) \cdot h(m) \quad \text{Faltungssumme} \quad (83)$$

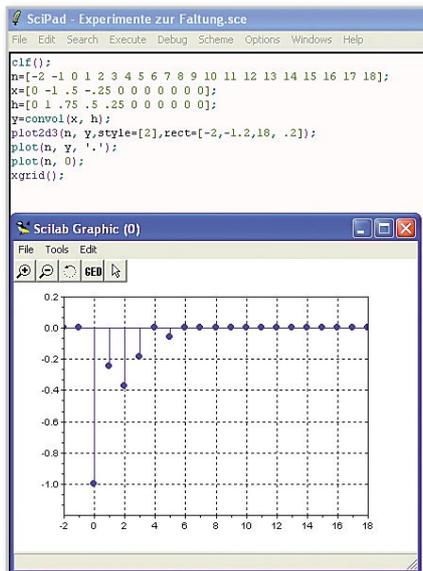


Bild 45: Auch die grafische Darstellung des Faltungsergebnisses erfordert nur wenige Zeilen Scilab-Programmcode.

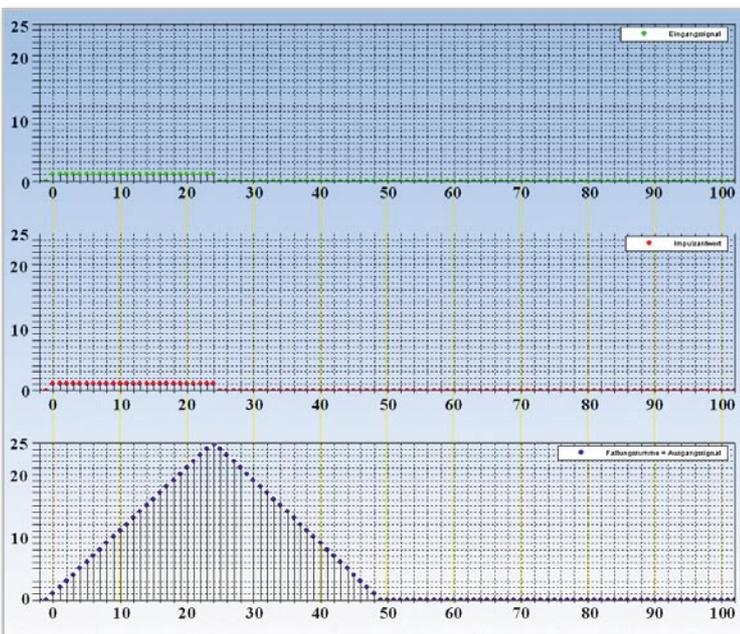


Bild 46: Die Faltung zweier Rechtecke ergibt ein Dreieck (Beispiel mit Scilab gerechnet).

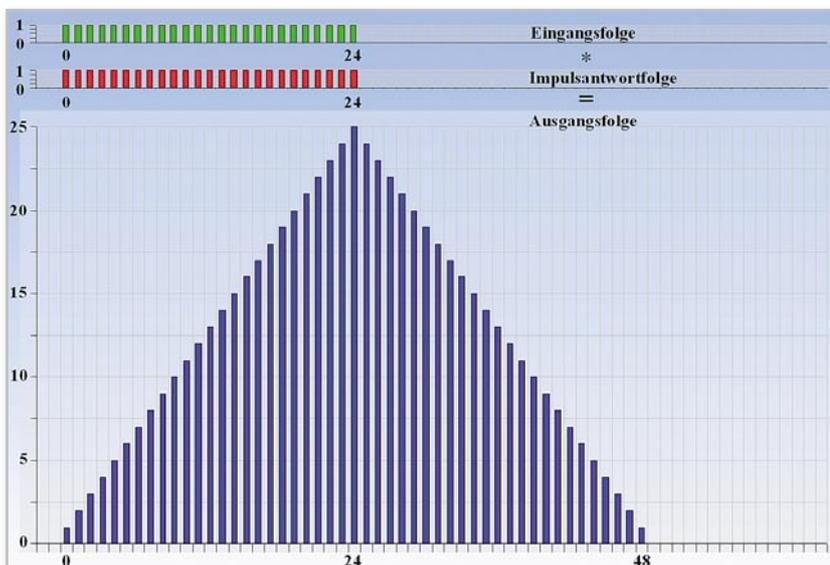


Bild 47: Mit OpenOffice Calc erhält man das gleiche Ergebnis wie in Bild 46.

$h(2) = 0,5$ und $h(3) = 0,25$. Nun geben wir ein Signal $x(t)$ auf den Eingang, welches aus drei Werten ungleich null besteht. Diese sind $x(0) = -1$, $x(1) = -0,5$, $x(2) = -0,25$. Auf jeden einzelnen dieser Eingangsimpulse reagiert das System gemäß seiner Impulsantwort und in der Überlagerung der Einzelimpulsantworten ergibt sich das Ausgangssignal als Reaktion auf die Eingangsfolge.

In Abbildung 43 ist die Aussage der Tabelle grafisch dargestellt. Wir sehen, dass die drei Eingangsimpulse zu drei verschobenen und bewerteten Impulsantworten (blau, grün und türkis) führen. Dabei beginnt keine der Impulsantworten vor dem anregenden Eingangsimpuls. Solche Systeme nennt man kausal, denn sie funktionieren nach dem Grundsatz: „keine Wirkung ohne und vor einer Ursache“. Die Summe ist das Ausgangssignal als Ergebnis der Faltung der Eingangsfolge mit der Impulsantwortfolge des Systems.

Die Faltungssumme kann mit einem Tabellenkalkulationsprogramm wie Microsoft Excel oder OpenOffice Calc leicht berechnet werden. So ist auch Tabelle 2 entstanden. Noch einfacher ist die Berechnung mit dem bereits erwähnten Programm Scilab. Hier braucht man nur die von null verschiedenen Impulsfolgenabschnitte von $x(n)$ und $h(n)$ und den Faltungsbehehl wie bei einem Taschenrechner einzugeben. Abbildung 44 zeigt den zugehörigen Screenshot.

Aber auch ein kleines Programm zur Durchführung der Faltung und grafischen Darstellung des Ergebnisses ist schnell geschrieben, wie Abbildung 45 beweist. Die Beispiele sollen Lust machen, sich doch mit dieser extrem nützlichen, kostenlosen Software etwas zu beschäftigen. Es lohnt sich auf jeden Fall!

Wichtig ist die Erkenntnis, dass bei der Faltung zweier begrenzter Folgen auch das Faltungsergebnis begrenzt ist. Wie Gleichung (84) zeigt, ist das Ausgangssignal immer so lang (L_y) wie die Summe der Längen von Impulsantwort (L_h) und Eingangssignal (L_x) minus 1.

In unserem Beispiel ist die Impulsantwort 4 Abtastwerte lang, das Eingangssignal 3 Abtastwerte, also muss die Ausgangsfolge aus $4 + 3 - 1 = 6$ Abtastwerten bestehen.

Das bedeutet natürlich im Umkehrschluss, dass wir mit Ausschnitten der Impulsantwort und der Eingangsfunktionen rechnen müssen, um keine unendlich lange Ausgangsfolge

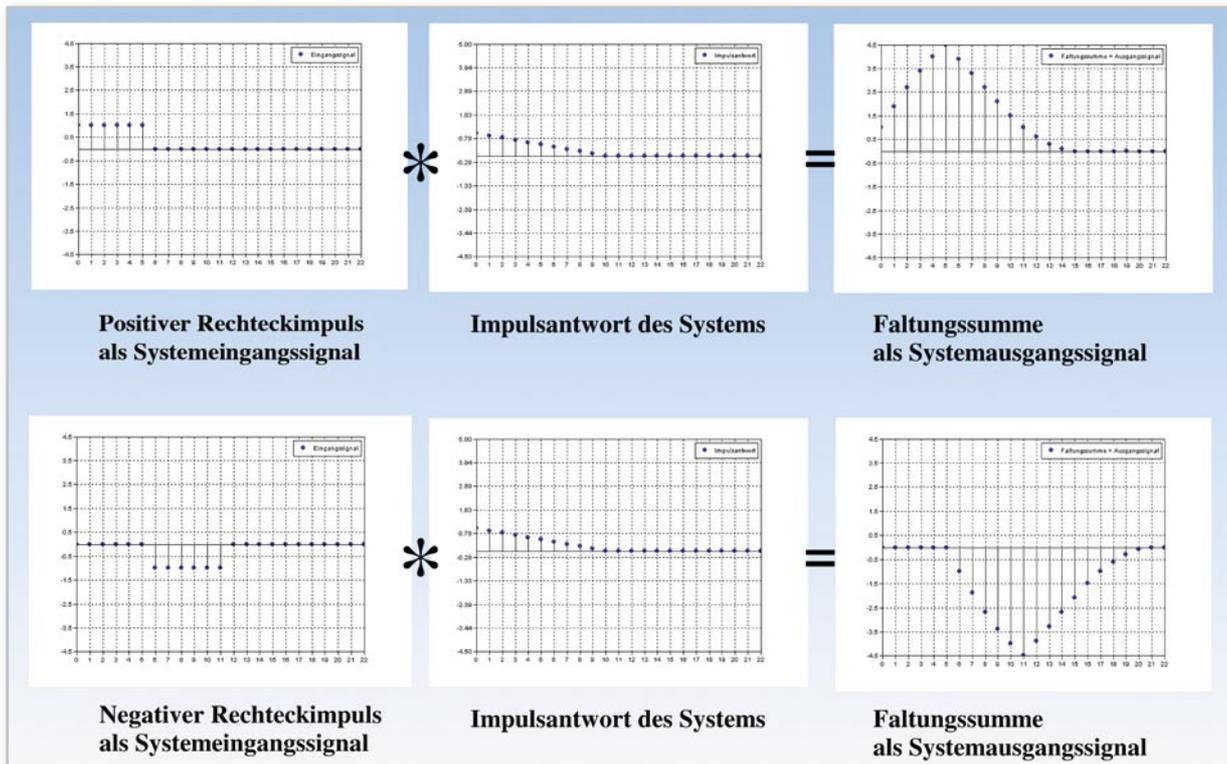


Bild 48: Zwei Eingangssignale führen zu je einem Ausgangssignal.

zu erhalten. Diesem Zweck dient die „Fensterung“ (Multiplikation mit einer Rechteckfunktion), mit deren Auswirkung wir uns später auseinandersetzen wollen. Die Kunst besteht darin, das Fenster so zu wählen, dass man mit angemessenem Rechenaufwand die Lösung mit der gewünschten Genauigkeit erhält.

Eine kleine Erkenntnis am Rande. Aus dem analogen Bereich wissen wir, dass die Fouriertransformierte einer Rechteck-

Zeitfunktion eine si-Funktion im Frequenzbereich ist. Eine Dreiecks-Zeitfunktion hat dagegen im Frequenzbereich das Quadrat zweier si-Funktionen zur Folge. Das Spektrum ist also immer positiv. Warum? Der Grund ist, dass die Dreiecksfunktion aus der Faltung zweier Rechtecke hervorgeht. Abbildung 46 demonstriert das über ein kleines Scilab-Programm, Abbildung 47 zeigt, dass eine ähnliche Darstellung auch mit OpenOffice Calc zu erhalten ist. Im Frequenzbereich werden

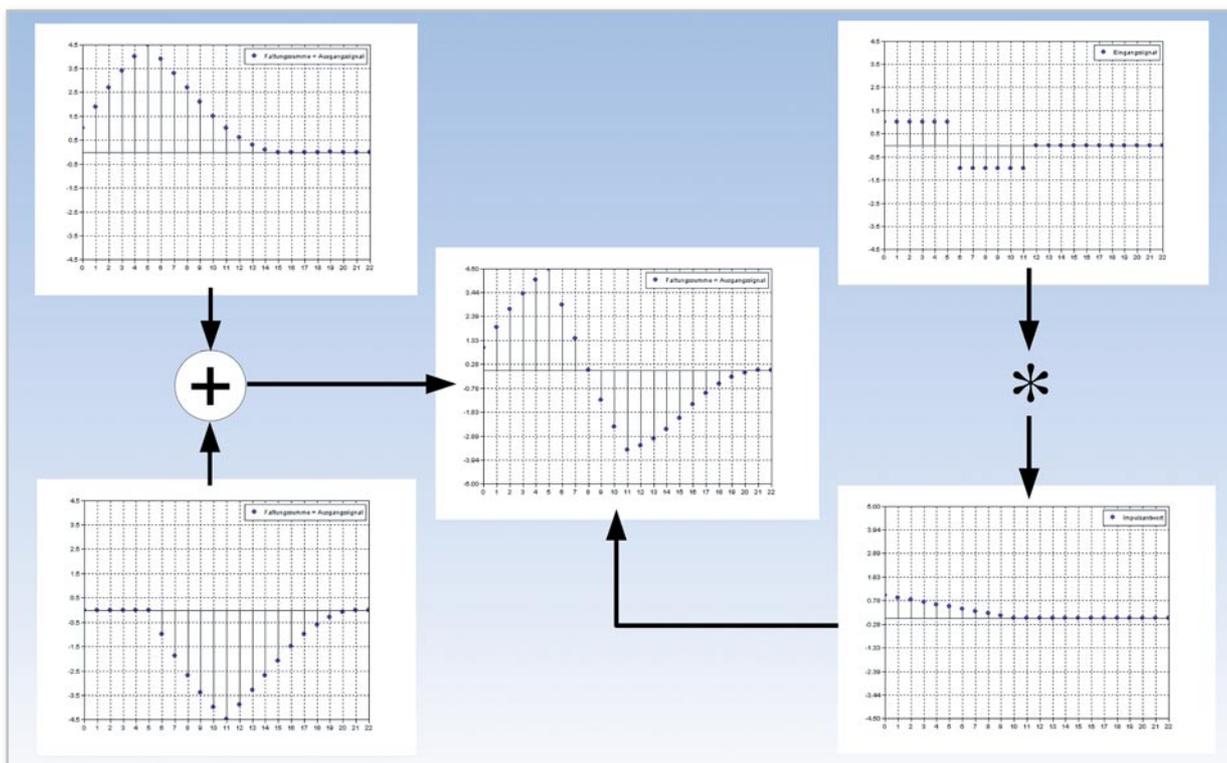


Bild 49: Die Überlagerung der zwei Eingangssignale aus Bild 48 führt zur Überlagerung der beiden Ausgangssignale. Das ist ein Charakteristikum von LTI-Systemen.

$$\underbrace{L_y}_{\text{Länge der Ausgangsfolge}} = \underbrace{L_x}_{\text{Länge der Eingangsfolge}} + \underbrace{L_h}_{\text{Länge der Impulsantwortfolge}} - 1 \quad (84)$$

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-i \frac{2\pi}{N} kn} = \sum_{n=0}^{N-1} x(n) W_N^{kn} \quad \text{mit } 0 \leq k \leq N-1 \quad (85)$$

Diskrete Fouriertransformierte (DFT)

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{i \frac{2\pi}{N} kn} = \frac{1}{N} \sum_{k=0}^{N-1} X(k) W_N^{-kn} \quad \text{mit } 0 \leq n \leq N-1 \quad (86)$$

Inverse Diskrete Fouriertransformierte (IDFT)

$$\Delta \omega = \frac{2\pi}{N} \quad \text{Abstand der Spektrallinien} = \text{spektrale Auflösung der } N\text{-Punkt-DFT} \quad (87)$$

$$W_N^{kn} = e^{-i \frac{2\pi}{N} kn} \quad \text{Drehfaktor} \quad (88)$$

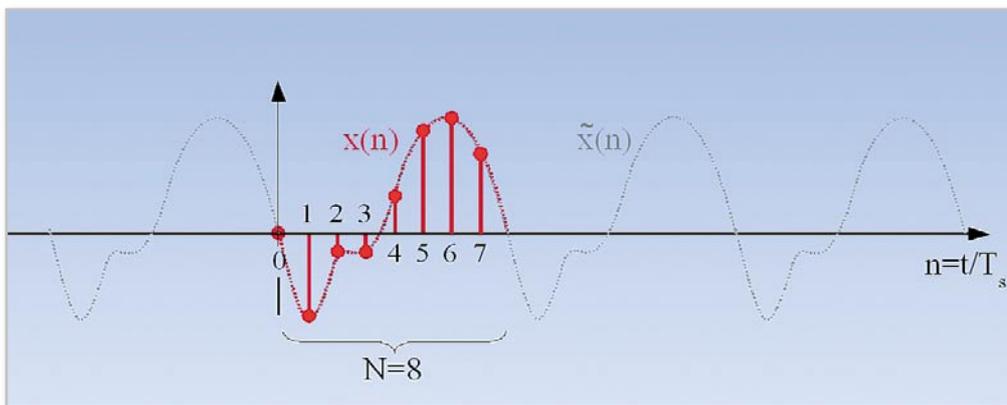


Bild 50: Wenn ein mit der DFT zu transformierender Signalabschnitt eine volle Periode des Signals darstellt, ist das Transformationsergebnis exakt.

die Spektren der gefalteten Rechtecke – zwei si-Funktionen – miteinander multipliziert. Ihre negativen Abschnitte sind deshalb nach dem Quadrieren positiv. Wir können das später mit der diskreten Fouriertransformation (DFT) am praktischen Beispiel nachvollziehen.

Noch ein Beispiel für die Nützlichkeit des Überlagerungsprinzips (Superpositionsprinzip), welches bei LTI-Systemen angewandt werden darf. Nehmen wir an, wir hätten ein System mit bekannter Impulsantwort und einem Eingangssignal, welches aus 6 Abtastwerten mit +1 (positives Rechteck) und darauf folgend 6 Abtastwerten mit -1 (verzögertes negatives Rechteck) besteht. Wenn wir die Systemantwort auf das positive Rechteck bereits kennen, ist uns auch die Systemantwort auf das negative, verzögerte Rechteck bekannt. Sie ist nichts weiter als die invertierte und um 6 Abtastintervalle verzögerte (nach rechts verschobene) Systemantwort auf das positive Rechteck. Diese „negative“ Systemantwort mit der „positiven“ überlagert, ergibt das gewünschte Resultat. Allgemein gilt: Wenn sich eine komplizierte Eingangsfunktion eines Systems in einfachere Bestandteile zerlegen lässt, kann

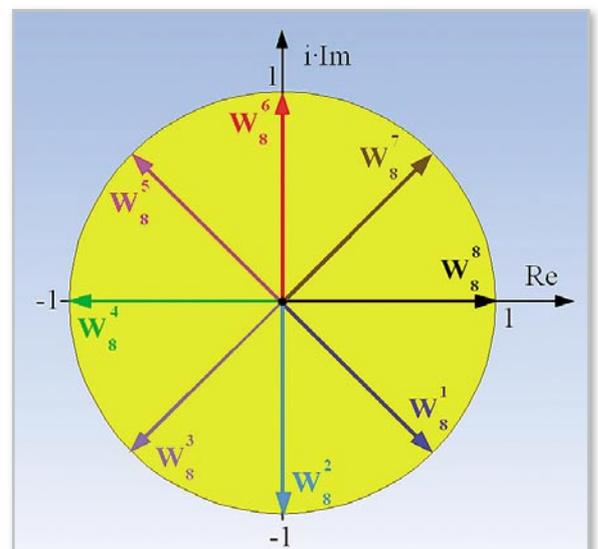


Bild 51: Der Drehfaktor ist im wahrsten Sinn des Wortes Dreh- und Angelpunkt der DFT und noch mehr ihrer schnellen Variante, der FFT.

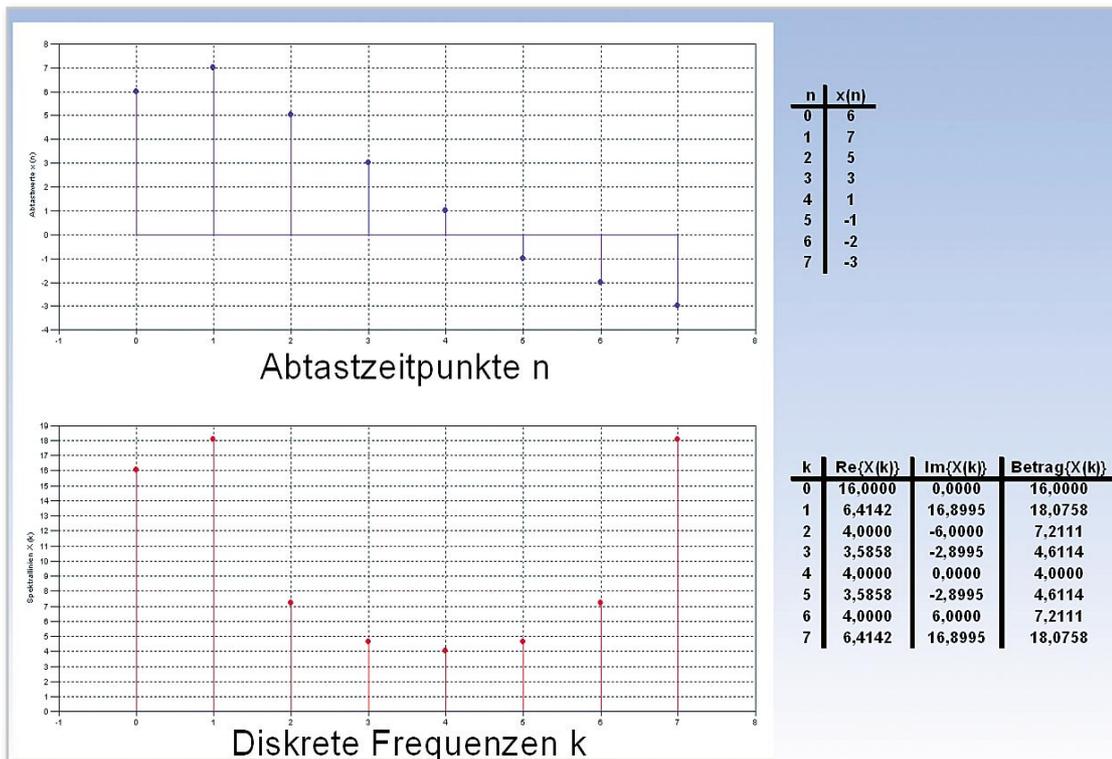


Bild 52: Die Berechnung der DFT nach Gleichung (85) von Hand ist stupide und mühsam, mit Scilab dagegen ganz einfach.

man die Antworten eines LTI-Systems darauf (Abbildung 48) ermitteln und zur Systemantwort auf die zerlegte, komplizierte Eingangsfunktion (Abbildung 49) überlagern. Das vereinfacht die Lösung eines Problems oft erheblich. Solche „divide and conquer“-Methoden (teile und herrsche) werden in vielen Bereichen der Mathematik und Naturwissenschaften angewandt.

Diskrete Fouriertransformation

Das aus der analogen Signaldarstellung bekannte Fourierintegral dient der Transformation einer Zeitfunktion in den Frequenzbereich (Zeitfunktion \rightarrow Spektrum), das inverse Fourierintegral der Rücktransformation in den Zeitbereich (Spektrum \rightarrow Zeitfunktion). Ohne Herleitung geben wir die aus den Gleichungen (18) für die Fouriertransformierte und (19) für die Fourierrücktransformierte von analogen Signalen die Beziehungen für zeitdiskrete (abgetastete) Signale an. Dabei wird noch vorausgesetzt, dass nur ein endlich langer Abschnitt $x(n)$ des Signals zu transformieren ist, den man als eine Periode eines periodischen Signals $\tilde{x}(n)$ auffasst (Abbildung 50). Dieses hätte – bedingt durch die Abtastung – ein periodisches Spektrum, von dem man aber ebenfalls nur eine Periode herausgreift. Die mathematische Formulierung dafür zeigen Gleichung (85) und Gleichung (86).

Wird das Zeitsignal N -mal abgetastet, ist der Abstand der Spektrallinien $2\pi/N$ (Gleichung (87)).

In der Anschrift der Gleichungen (85) und (86) wurde eine Abkürzung verwendet, die man Drehfaktor (in der englischen Literatur „twiddle factor“) nennt (Gleichung (88)).

Er ist von fundamentaler Bedeutung für die DFT, weil sei-

ne Symmetrieeigenschaften viele Vereinfachungen ermöglichen. Mathematisch gesehen ist der Drehfaktor ein komplexer Ursprungszeiger der Länge 1, der je nach Größe von N und kn verschiedene Richtungen annimmt. Am Beispiel $N = 8$ (8 Abtastwerte) und $kn = 1, 2, 3 \dots 8$ ergeben sich 8 Zeiger, alle höheren Werte für kn sind periodische Wiederholungen davon (Abbildung 51). Die Symmetrien bestehen darin, dass $W_8^8 = -W_8^4$, $W_8^7 = -W_8^3$, $W_8^6 = -W_8^2$ und $W_8^5 = -W_8^1$ ist, also jeder Zeiger ein am Ursprung gespiegeltes Gegenstück hat, dessen Real- und Imaginärteil ein inverses Vorzeichen aufweist.

Das Auswerten der Gleichung (85) von Hand mit Hilfe eines Taschenrechners ist nicht schwierig, aber mühselig und fehlerträchtig. Lediglich die erste Spektrallinie $X(k=0)$ ist einfach zu berechnen, weil dann der Drehfaktor 1 ist und sich der Spektrallinienwert einfach als Summe aller Abtastwerte ergibt. Entsprechendes gilt bei Gleichung (86). Mit einem kleinen Scilab-Programm kann man sich das Leben sehr erleichtern. Abbildung 52 zeigt, wie ein diskretes Zeitsignal bestehend aus 8 Abtastwerten gemäß Gleichung (85) in ein diskretes Spektrum aus ebenfalls 8 Spektrallinien umgerechnet wird. Weil die Spektrallinien im Allgemeinen komplex sind, wurde ihr Betrag berechnet und abgebildet. Phaseninformationen sind dabei verloren gegangen.

In Teil 6 studieren wir die Auswirkungen des Auffüllens von Folgen mit Nullwerten im Zeit- und Frequenzbereich („zero padding“), demonstrieren die diskrete Faltung mit Hilfe der DFT und beschäftigen uns mit dem Radix-2-Algorithmus der schnellen Fouriertransformation (FFT). **ELV**



Gemeinhin gilt: „Von nichts kommt nichts.“ Bei einer bestimmten Erweiterung von diskreten Signalfolgen im Zeit- und Frequenzbereich durch Nullwerte ist das anders. Wenn wir Letztere zwischen vorhandene Abtastwerte der Zeitsignalfolge oder in der Mitte der Frequenzsignalfolge einfügen, führt das zu interpolierten Werten in den korrespondierenden Folgen im jeweils anderen Bildbereich. Mit diesem „Zero Padding“ genannten Verfahren zur Auflösungsverbesserung beschäftigen wir uns in Teil 6 unserer Artikelserie.

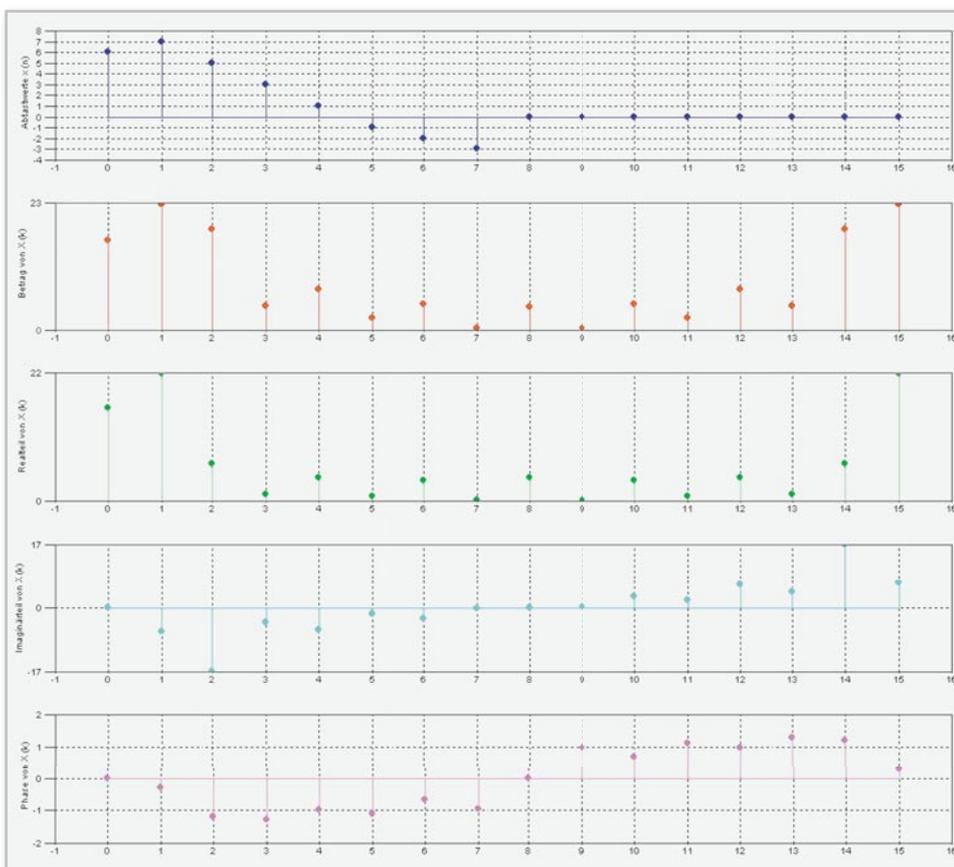


Bild 53: Das Erweitern der Zeitsignalfolge durch Anhängen von Nullen (hier 8) führt im Frequenzbereich zu interpolierten Spektralwerten.

Rückblick

Es wurde bereits gesagt, dass die diskrete Fouriertransformation eine gewisse Anzahl von Zeitwerten in ebenso viele Spektralwerte überführt. Bei der Wahl von N als der Anzahl von Abtastwerten sollte die Messdauer $N \cdot T_s$ mindestens gleich lang wie die Signaldauer sein, um dieses vollständig zu erfassen. Wenn wir N dann noch als Zweierpotenz (2, 4, 8, 16 ..., 1024 ...) wählen, ist das Verfahren der schnellen Fouriertransformation (Fast Fourier Transform: FFT) anwendbar. Weil die spektrale Auflösung umgekehrt proportional zur Anzahl der Abtastwerte ist (siehe Gleichung [87]), sollte N möglichst groß sein. Das kann durch eine häufigere Abtastung des zeitkontinuierlichen Signals geschehen oder durch das Anhängen von Nullen an die verfügbare Folge von Abtastwerten. Letzteres wird als Time Domain Zero Padding (TDZP) bezeichnet und bewirkt die Interpolation zusätzlicher Spektralwerte. Umgekehrt führt das Einfügen von Nullen (Frequency Domain Zero Padding: FDZP) in der Mitte der Spektralfolge zu interpolierten Zeitfolgenwerten.

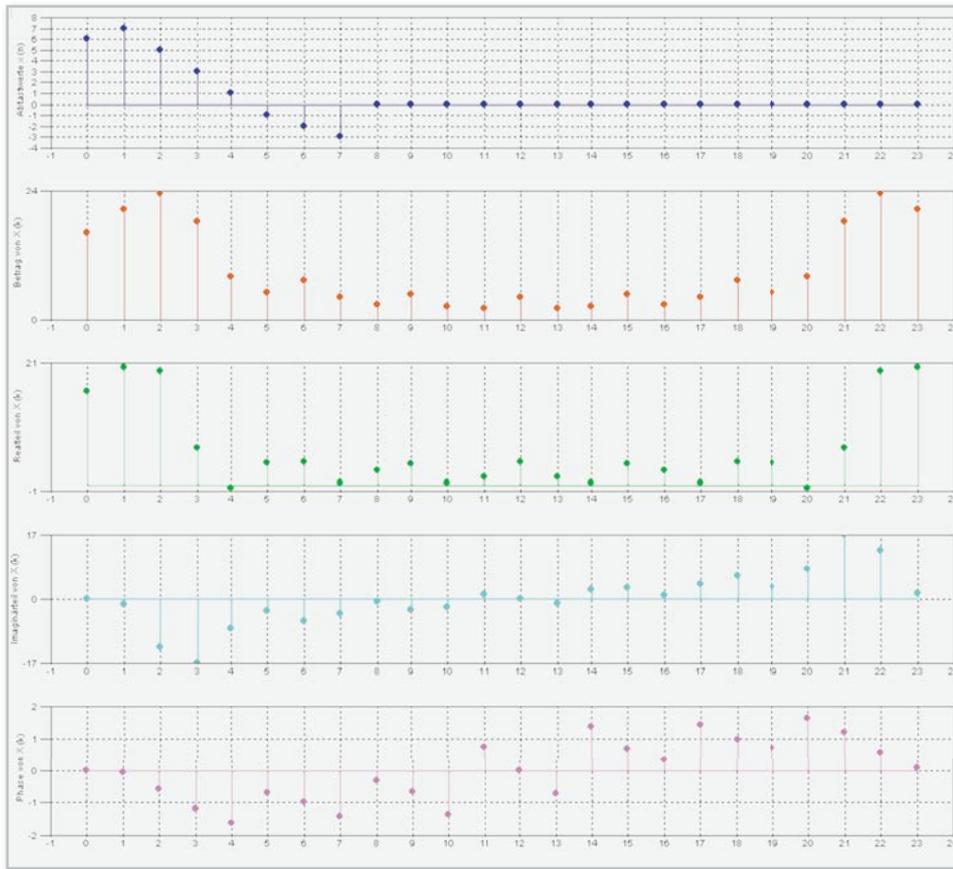


Bild 54: Zwei angehängte Nullen pro Original-Zeitreihenwert führen zu zwei interpolierten Werten zwischen den Werten der Original-Spektralfolge.

Time Domain Zero Padding

Schauen wir uns die Wirkung des Anhängens von Nullen an eine Signalfolge im Zeitbereich einmal genauer an. Im Englischen wird das als Time Domain Zero Padding (TDZP) bezeichnet, was in der Übersetzung so viel heißt wie: Auffüllen mit Nullen im Zeitbereich.

Die aufgefüllte Folge wird dann der DFT unterworfen, d. h., in den Frequenzbereich transformiert. Da die Zahl der Spektrallinien immer gleich der Anzahl der transformierten Zeit-

folgenwerte ist, erwarten wir also ein feiner gerastertes Linienspektrum, d. h. die spektrale Auflösung muss steigen. Abbildung 53 bestätigt das. Hier wurden an das Zeitsignal aus Abbildung 52 8 Nullen angehängt, so dass die Folge nun 16 Werte lang ist: $x(n) = (6, 7, 5, 3, 1, -1, -2, -3, 0, 0, 0, 0, 0, 0, 0, 0)$.

Bei genauer Betrachtung des roten Betragsspektrums in Abbildung 53 fällt auf, dass die Spektralwerte des nicht mit Nullen aufgefüllten Zeitsignals wieder auftreten, aber jeweils gefolgt von einem neuen, interpolierten Wert. Wir

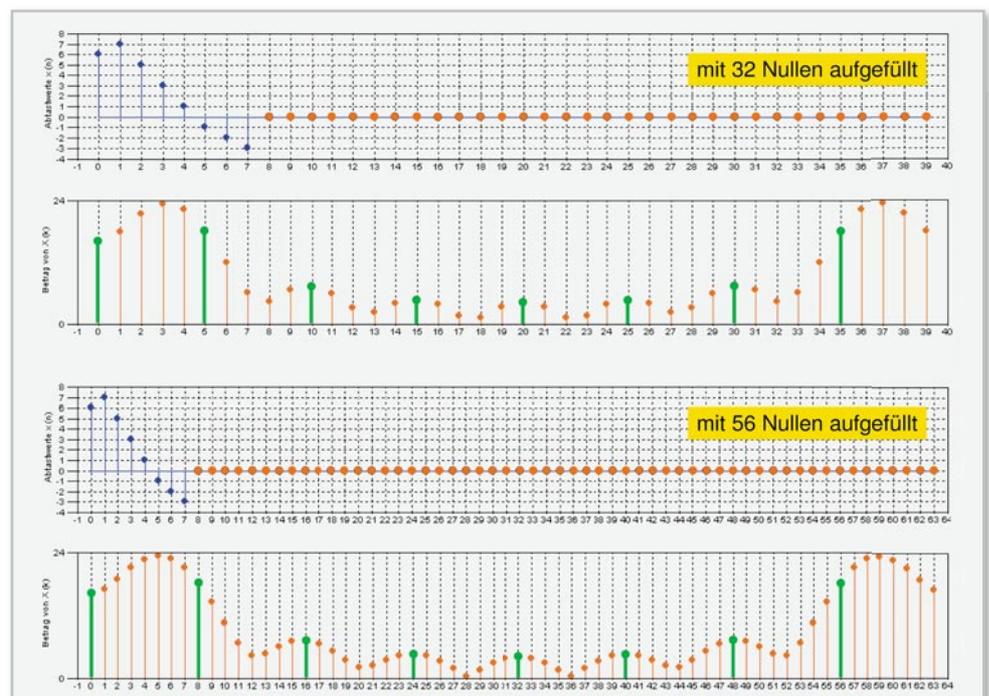


Bild 55: Zero Padding im Zeitbereich mit 8 Signalwerten und 32 bzw. 56 Nullwerten.

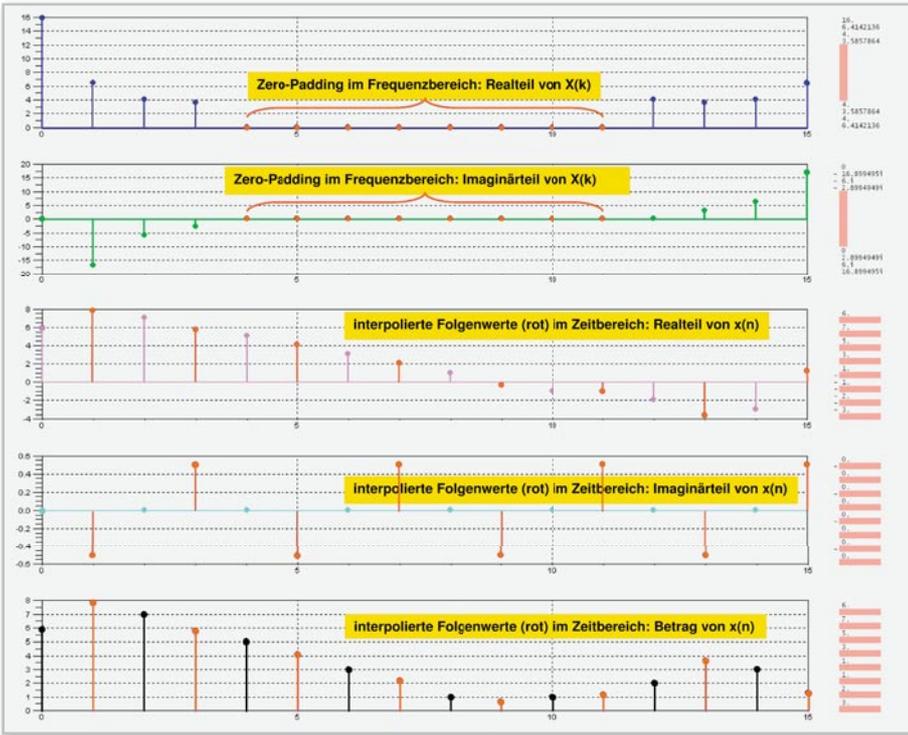
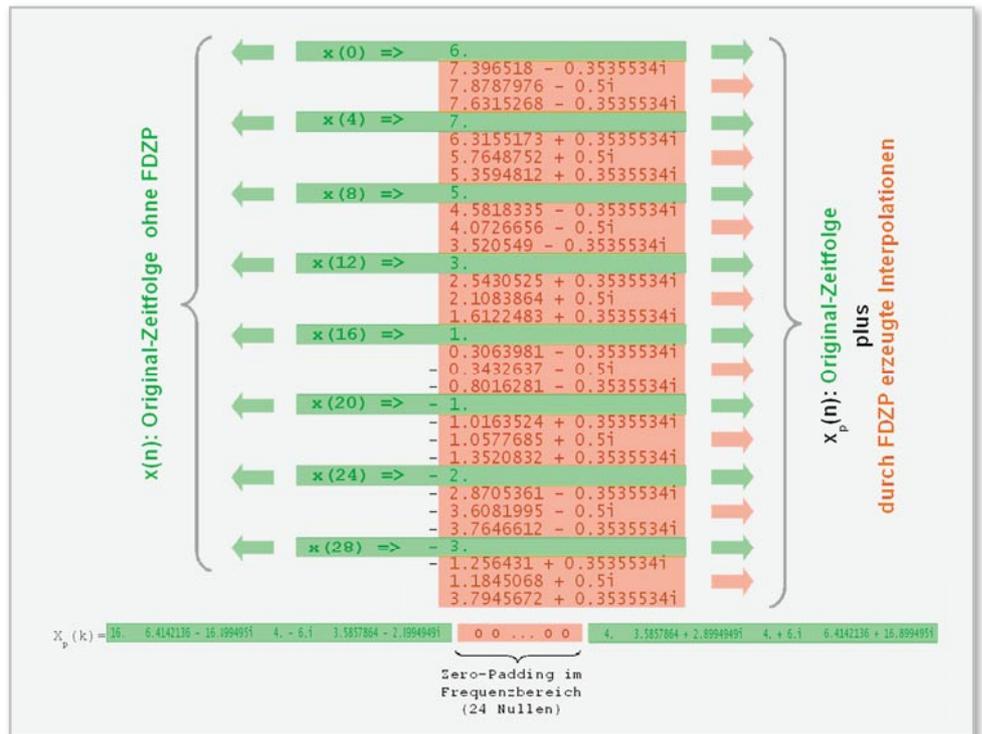


Bild 56: Zero Padding im Frequenzbereich erfolgt durch Einfügen einer Folge von Nullwerten in der Mitte der Original-Spektralfolge. Die Wirkung ist das Einschachteln von interpolierten Werten zwischen die Werte der Original-Zeitfolge.

Bild 57: Zero Padding im Frequenzbereich (FDZP) zum Nachvollziehen

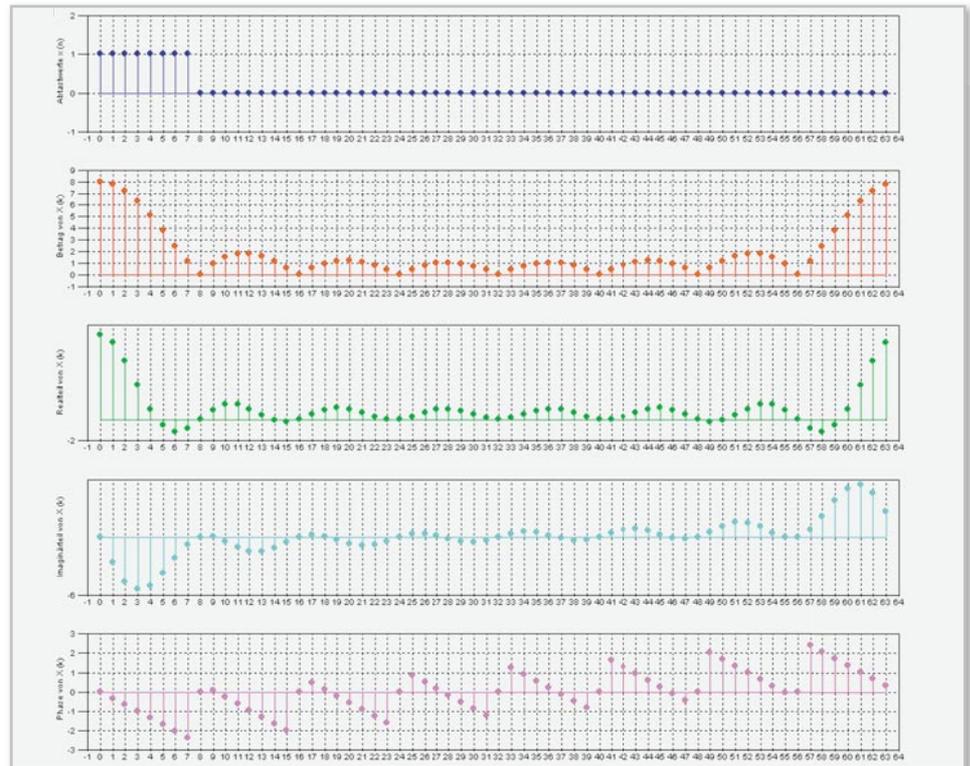


prüfen das an der gleichen Zeitfolge nach, der wir jetzt 16 Nullen anhängen (Abbildung 54). Tatsächlich sind jetzt zwei neue Werte zwischen die Spektralwerte der „ungepadeten“ Zeitfolge interpoliert worden. Ein dritter Versuch mit 32 bzw. 56 angehängten Nullen demonstriert die gestiegene Frequenzauflösung ganz augenscheinlich (Abbildung 55). Jetzt wurden 4 bzw. 7 neue Werte zwischen die Spektrallinien der Original-Ausgangszeitfolgen eingefügt. Die spektrale Auflösung hat sich vier- bzw. versiebenfach, allerdings ohne einen echten Zuwachs an Informationsgehalt, denn die interpolierten Zwischenwerte rühren ja nicht von einem feiner abgetasteten Zeitsignal her, sondern werden vom Zero Padding verursacht.

Frequency Domain Zero Padding

Entsprechendes können wir beobachten, wenn wir Nullen in der Mitte der Spektralfolge einfügen (FDZP: Frequency Domain Zero Padding). Hier wird – mit 8 Nullen – jeweils ein Wert zwischen den Original-Zeitfolgenwerten interpoliert (Abbildung 56), bei 16 Nullen zwei Werte usw. Man sieht, dass das Ergebnis der inversen DFT (IDFT) wieder die reellen Werte der Original-Zeitfolge erzeugt, aber zwischen ihnen komplexe Werte hinzufügt. Die gleiche Frequenzfolge mit 24 eingefügten Nullen führt zu den Zahlenwerten von Abbildung 57. Hier sind jeweils drei interpolierte Werte zwischen die Folgenwerte der Original-Zeitfolge verschachtelt.

Bild 58: Im Allgemeinen ist das Spektrum einer Zeitfolge durch einen Real- und einen Imaginärteil oder seinen Betrag und seine Phase gekennzeichnet. Das wird hier am Beispiel eines Rechtecks gezeigt.



In der Praxis tritt der Zero-Padding-Effekt immer dann auf, wenn mit einer festen Anzahl von Werten transformiert wird und die Signalfolge den Transformationsrahmen („frame“ oder „record“, man kann auch Analysefenster dazu sagen) nicht ausfüllt, also kürzer ist.

Den Zusammenhang zwischen Frequenzauflösung Δ , Frame-Länge T_0 und Samplingfrequenz $f_s = 1/T_s$ zeigt Gleichung 89. Durch das Zero Padding verkleinern sich die Amplituden der transformierten Folgewerte, weil ja die Signalenergie auf mehr Transformationsfolgewerte verteilt wird. Bei Einfügen gleich vieler Nullen wie Signalfolgewerte halbieren sich die

Amplituden der Transformationsfolge, bei doppelt so vielen Nullen gehen sie auf ein Drittel zurück usw. In unseren Beispielen wurde deshalb mit einem entsprechenden Faktor multipliziert, um die Originalwerte einfacher in der interpolierten Folge wiederzufinden.

$$\Delta f = \frac{f_s}{N} = \frac{1}{N \cdot T_s} = \frac{1}{T_0} \quad \text{Frequenzauflösung der DFT} \quad (89)$$

Ein weiteres Beispiel zeigt Abbildung 58. Hier wurde ein 64-Punkt-Rechteck-Zeitsignal aus 8 Eins-Werten und 56 Null-

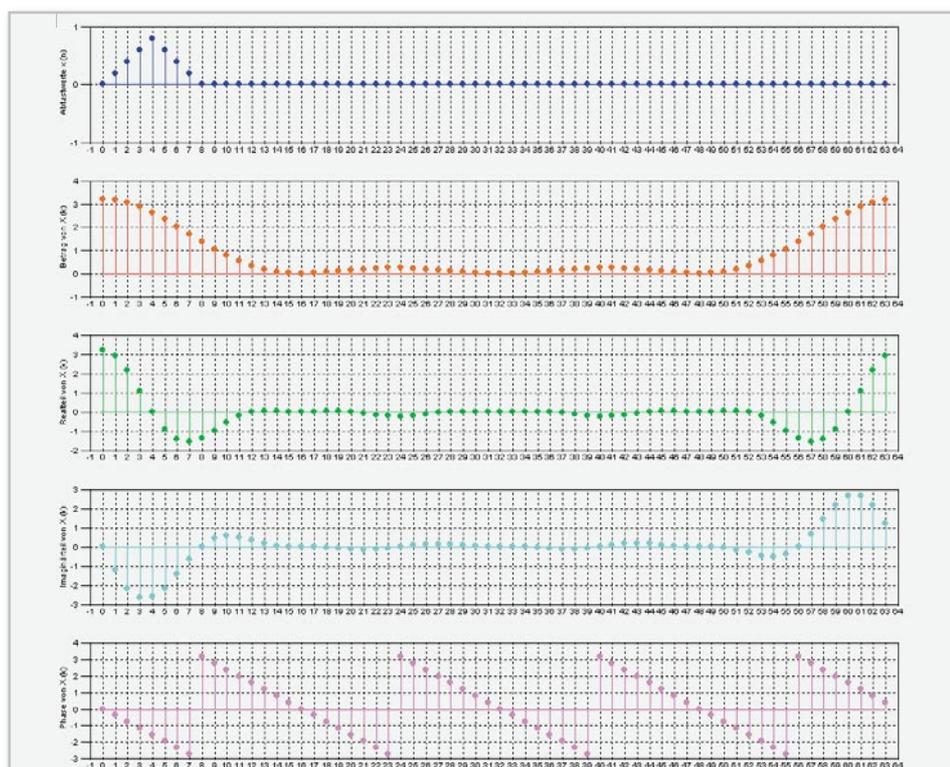


Bild 59: Die DFT eines Dreiecks

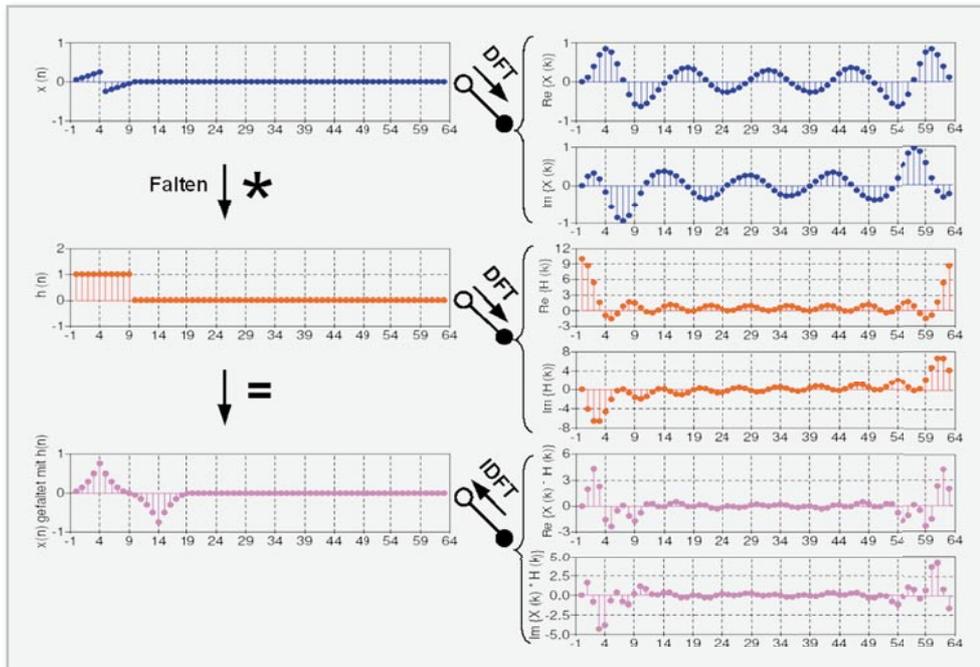


Bild 60: Man kann zwei Zeitfolgen auch falten, indem man ihre diskreten Fouriertransformaten miteinander multipliziert und das Produkt wieder in den Zeitbereich zurücktransformiert (Convolution by DFT).

Werten der DFT unterworfen. Es ergibt sich im Frequenzbereich der charakteristische $\sin(x)/x$ -Verlauf, wegen der Betragsbildung mit hochgeklappten negativen Anteilen. Eine Dreiecksfolge im Zeitbereich und im Frequenzbereich, deren Betrag, Real- und Imaginärteil sowie Phase zeigt Abbildung 59. Bezogen auf die Mitte des Spektralfensters ist der Realteil des Spektrums $\text{Re}\{X(i\omega)\}$ gerade, der Imaginärteil $\text{Im}\{X(i\omega)\}$ und die Phase sind dagegen ungerade.

Diskrete Faltung mit Hilfe der DFT

Gleichung 83 definierte die Faltung zweier Zeitsignalfolgen. Eine war das Eingangssignal und die andere die Impulsantwort eines LTI-Systems. Das Faltungsergebnis ist das Ausgangssignal als Systemantwort auf das Eingangssignal. Man kann das Ausgangssignal aber auch ohne Berechnung der Faltungssumme bestimmen. Dazu werden die endlichen Zeitfolgen, die $x(n)$ mit der Länge N_1 und $h(n)$ mit der Länge N_2 repräsentieren, mit so vielen Nullen aufgefüllt, dass ihre Länge $N \geq N_1 + N_2 - 1$ ist.

Dann werden die Folgen der DFT unterworfen, miteinander multipliziert und das Produkt wird mit Hilfe der IDFT (inversen DFT) wieder in den Zeitbereich zurücktransformiert.

Wir haben damit die Faltung als Operation im Zeitbereich über eine Multiplikation im Frequenzbereich ausgeführt.

Abbildung 60 fasst die Alternativen zusammen.

Auf der linken Hälfte werden $x(n)$ und $h(n)$ zum Ausgangssignal gefaltet.

Auf der rechten Seite werden die zugehörigen, per DFT ermittelten Spektren $X(k)$ und $H(k)$ miteinander multipliziert und über die inverse DFT zur Ausgangsfolge des Systems im Zeitbereich transformiert.

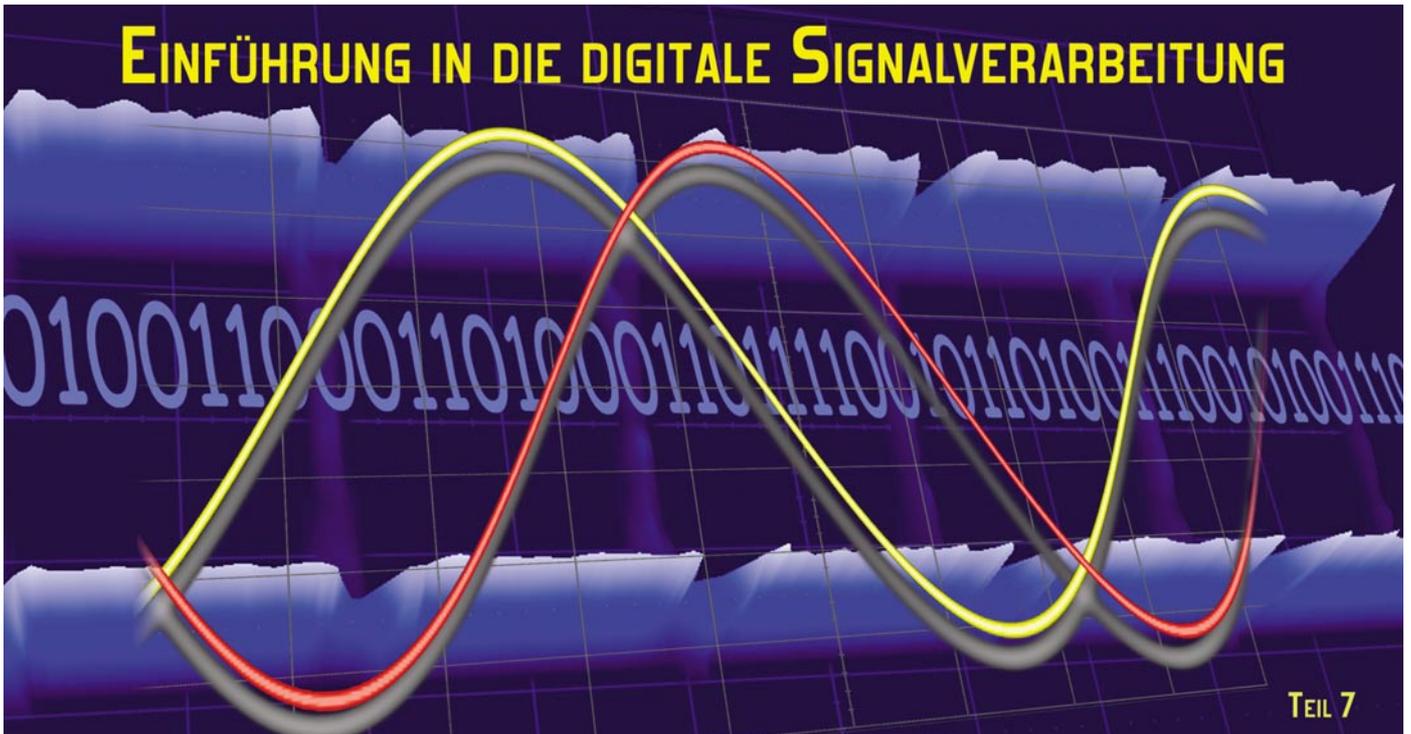
Auf den ersten Blick scheint der Weg über den Frequenzbereich aufwändiger. Aber bei langen Eingangsfolgen $x(n)$ dauert die Berechnung der Faltungssumme sehr lang. Mit den Methoden der schnellen Fouriertransformation (FFT: Fast Fourier Transform) und einer Segmentierung der Eingangs-

folge können beträchtliche Geschwindigkeitsgewinne erzielt werden.

Dazu werden die Faltungen der Teilfolgen mit der Impulsantwort durch Multiplikation ihrer Spektren berechnet und nach der Rücktransformation in den Zeitbereich wieder zusammengefügt. Bekannte Vertreter dieser segmentweisen Faltung (block convolution) sind die Overlap-add- und die Overlap-save-Methode, auf die wir hier aber nicht näher eingehen wollen.

In Teil 7 unserer Serie beschäftigen wir uns mit der „schnellen Fouriertransformation“ (FFT: Fast Fourier Transform), einem der wichtigsten Verfahren zur Analyse von Funktionen und Signalen in der digitalen Signalverarbeitung. Mit dem Einstieg in die z-Transformation schaffen wir uns dann das mathematische Rüstzeug, um Abtastsysteme effizient zu beschreiben.

ELV



In Teil 7 unserer Serie beschäftigen wir uns mit der „schnellen Fouriertransformation“ (FFT: Fast Fourier Transform), einem der wichtigsten Verfahren zur Analyse von Funktionen und Signalen in der digitalen Signalverarbeitung.

Die schnelle Fouriertransformation

Die schnelle Fouriertransformation (FFT: Fast Fourier Transform) ist ein Algorithmus, mit dem sich eine diskrete Fouriertransformation (DFT) erheblich schneller durchführen lässt. Am bekanntesten sind die Radix-2-FFT-Algorithmen, die der deutsche Mathematiker Carl Friedrich Gauß schon um 1805 für die Berechnung von Asteroidenbahnen entwarf. James W. Cooley and John W. Tukey stellten den Radix-2-Algorithmus 1965 in dem Aufsatz „An algorithm for the machine calculation of complex Fourier series“ in Mathematics of Computation 19, 297–301 (1965), für die „maschinelle Berechnung komplexer Fourierreihen“ vor. Dabei wird die zu transformierende Folge – deren Länge N eine Potenz zur Basis 2 ist – entweder im Zeitbereich (DIT: decimation in time) oder im Frequenzbereich (DIF: decimation in frequency) so lange zerlegt, bis elementare, einfachst

zu berechnende Strukturen – die Butterflies (Schmetterlinge) – entstehen. Bei der Herleitung der Algorithmen durch diese Teile-und-herrsche-Strategie spielen die Symmetrieeigenschaften des Drehfaktors W eine wichtige Rolle.

FFT durch Dezimierung im Zeitbereich

Wir gehen von der Definitionsgleichung 85 der DFT aus (s. Gleichung 90). Sie transformiert eine Folge von N Werten im Zeitbereich in N Frequenzwerte. Wenn die Voraussetzung erfüllt ist, dass die Anzahl der Zeitfolgenwerte $x(n)$ eine Potenz von zwei ist (also $N = 4, 8, 16, 32, 64 \dots$), dann ist eine Aufspaltung von $x(n)$ in zwei Teilfolgen $g(n)$ mit geraden und $u(n)$ mit ungeraden Indizes möglich (Gleichung 91). Damit lässt sich Gleichung 90 umschreiben als Gleichung 92.

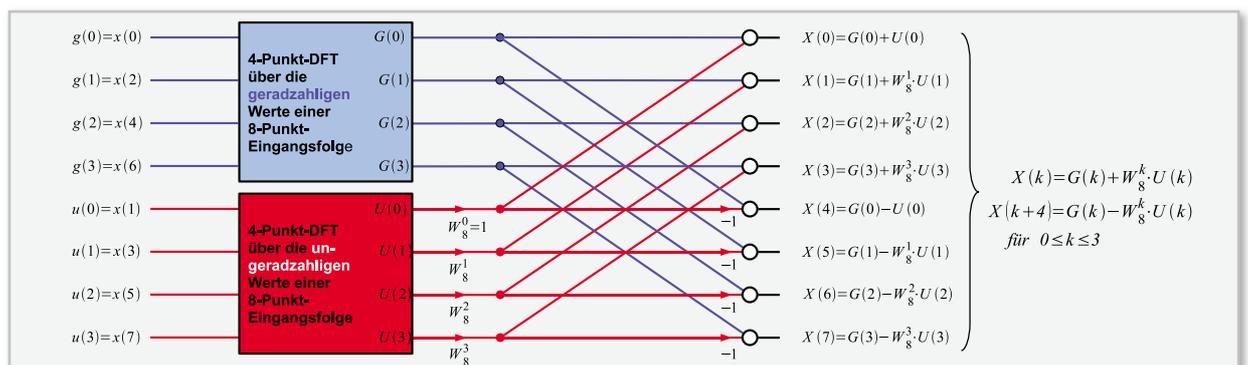


Bild 61: Von der DFT zur FFT: Anstatt eine 8 Werte lange Folge der DFT zu unterwerfen (8-Punkt-DFT), kann man auch die Werte der Eingangsfolge mit geradzahligem und ungeradzahligem Indizes transformieren (zwei 4-Punkt-DFTs) und die Ergebnisse verknüpfen.

Nun machen wir uns zunutze, dass Gleichung 93 gilt. Zum Verständnis von Gleichung 93 wollen wir den darin ausgedrückten Zusammenhang in Gleichung 94 einmal ableiten. Dabei nutzen wir die Definition des Drehfaktors aus Gleichung 88.

Nun setzen wir Gleichung 94 in Gleichung 92 ein und erhalten Gleichung 95. Die Summenterme sind nun nichts weiter als die Fouriertransformierten der $N/2$ langen Teilfolgen mit den geraden und ungeraden Indizes. Wenn wir die Fouriertransformierte von $g(n)$ als $G(k)$ bezeichnen und die von $u(n)$ als $U(k)$ geht Gleichung 95 in Gleichung 96 über. Nun könnte man die Richtigkeit von Gleichung 96 anzweifeln, weil $G(k)$ und $U(k)$ ja jeweils nur $N/2$ Punkte aufweisen. Aber wegen der Periodizität der komplexen Exponentialausdrücke kann man Gleichung 97 schreiben. Wegen Gleichung 97 gilt Gleichung 96 für alle k , also $0 \leq k \leq N-1$. Gleichung 96 ließe sich nun direkt in ein Blockschaltbild umsetzen. Zuvor wollen wir uns aber eine Vereinfachung zunutze machen, die sich direkt aus Gleichung 96 ableiten lässt. Betrachten wir $X(k)$ und $X(k + N/2)$ entsprechend Gleichung 98, die für $N = 8$ ihre direkte Entsprechung in Abbildung 61 findet.

Bei der verwendeten Graphendarstellung wurden die folgenden Vereinbarungen getroffen. Pfeile stellen Multiplizierer, voll ausgefüllte Punkte Verzweigungen und Kreise Addierer dar. Der Wert unter einem Pfeil ist mit dem Signal zu multiplizieren. Ist dieser Wert ein Drehfaktor, so handelt es sich um eine komplexe Multiplikation, ist er -1 , so muss der Signalwert nur invertiert werden.

Wir sehen in Gleichung 98, dass nur die aus den ungeraden Zeitfolgewerten entstandenen Spektralwerte $U(k)$ mit den Drehfaktoren (Twiddle Factors) multipliziert werden. Bei den Ergebnisspektralwerten $X(n)$ mit $0 \leq n \leq 3$ werden diese Produkte zu den von den geradzahigen Zeitfolgewerten herührenden Spektralwerten $G(k)$ addiert, für $X(n)$ mit $4 \leq n \leq 7$ wird dagegen die Differenz gebildet.

Weil wir anfangs voraussetzten, dass die Anzahl der Zeitfolgenwerte $x(n)$ eine Potenz von zwei ist (hier $N = 2^3 = 8$), können wir nun für jeden der 4-Punkt-DFT-Blöcke deren Eingangsfolge wieder aufspalten, was zu jeweils zwei 2-Punkt-

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn} \quad \text{mit } 0 \leq k \leq N-1 \quad (90)$$

$$g(n) = x(2n) \quad n=0, 1, 2, \dots, \frac{N}{2}-1 \quad \text{Teilfolge mit geraden Indizes} \quad (91)$$

$$u(n) = x(2n+1) \quad n=0, 1, 2, \dots, \frac{N}{2}-1 \quad \text{Teilfolge mit ungeraden Indizes}$$

$$X(k) = \sum_{l=0}^{\frac{N}{2}-1} g(l) W_N^{2lk} + \sum_{l=0}^{\frac{N}{2}-1} u(l) W_N^{(2l+1)k} \quad (92)$$

$$W_N^{2lk} = W_{\frac{N}{2}}^{lk} \quad \text{Symmetrieeigenschaft des Drehfaktors} \quad (93)$$

$$W_N^{2lk} = e^{-i \frac{2\pi}{N} 2lk} = e^{-i \frac{2\pi}{\frac{N}{2}} lk} = W_{\frac{N}{2}}^{lk} \quad (94)$$

$$X(k) = \sum_{l=0}^{\frac{N}{2}-1} g(l) W_{\frac{N}{2}}^{lk} + W_N^k \sum_{l=0}^{\frac{N}{2}-1} u(l) W_{\frac{N}{2}}^{lk} \quad (95)$$

$$X(k) = G(k) + W_N^k \cdot U(k) \quad (96)$$

$$G(k) = G\left(k + \frac{N}{2}\right) \quad (97)$$

$$U(k) = U\left(k + \frac{N}{2}\right)$$

$$X(k) = G(k) + W_N^k \cdot U(k) \quad (98)$$

$$X\left(k + \frac{N}{2}\right) = G\left(k + \frac{N}{2}\right) + W_N^{k + \frac{N}{2}} \cdot U\left(k + \frac{N}{2}\right) = G(k) + W_N^k \cdot \underbrace{W_N^{\frac{N}{2}}}_{-1} \cdot U(k) = G(k) - W_N^k \cdot U(k)$$

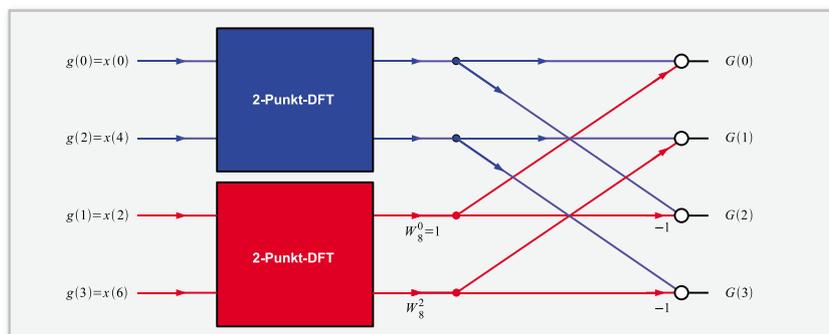
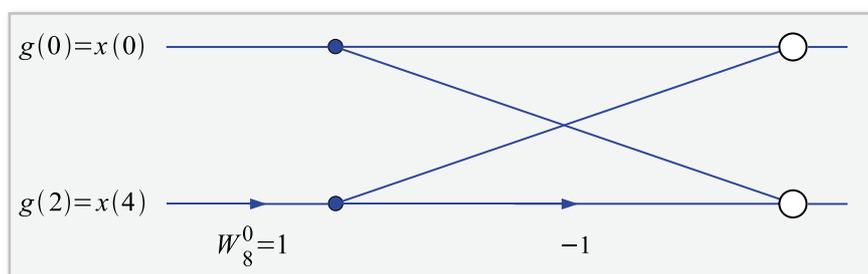


Bild 62: Eine 4-Punkt-DFT wird zu zwei 2-Punkt-DFTs.

Bild 63: Bei der 2-Punkt-DFT (elementarer Butterfly) ist keine weitere Zerlegung möglich.



$$G(k) = \sum_{l=0}^{\frac{N}{4}-1} g(2l) W_{\frac{N}{4}}^{lk} + W_{\frac{N}{2}}^k \sum_{l=0}^{\frac{N}{4}-1} g(2l+1) W_{\frac{N}{4}}^{lk} \quad (99)$$

$$G(k) = X(2k) = X(0) + X(2) + X(4) \dots + X(N-2) \quad k=0, 1, 2, \dots, \frac{N}{2}-1 \quad (100)$$

$$U(k) = X(2k+1) = X(1) + X(3) + X(5) \dots + X(N-1) \quad k=0, 1, 2, \dots, \frac{N}{2}-1$$

$$G(k) = \sum_{n=0}^{\frac{N}{2}-1} x(n) W_N^{2nk} + \sum_{n=\frac{N}{2}}^{N-1} x(n) W_N^{2nk} = \sum_{n=0}^{\frac{N}{2}-1} x(n) W_{\frac{N}{2}}^{nk} + \underbrace{\sum_{n=\frac{N}{2}}^{N-1} x(n) W_{\frac{N}{2}}^{nk}}_{\sum_{n=0}^{\frac{N}{2}-1} x(n+\frac{N}{2}) W_{\frac{N}{2}}^{(n+\frac{N}{2})k}} \quad (101)$$

$$= \sum_{n=0}^{\frac{N}{2}-1} x(n) W_{\frac{N}{2}}^{nk} + \sum_{n=0}^{\frac{N}{2}-1} x(n+\frac{N}{2}) \underbrace{W_{\frac{N}{2}}^{(n+\frac{N}{2})k}}_{W_{\frac{N}{2}}^{nk}}$$

$$G(k) = \sum_{n=0}^{\frac{N}{2}-1} \left[x(n) + x\left(n+\frac{N}{2}\right) \right] W_{\frac{N}{2}}^{nk}$$

$$U(k) = \sum_{n=0}^{\frac{N}{2}-1} x(n) W_N^{n(2k+1)} + \sum_{n=\frac{N}{2}}^{N-1} x(n) W_N^{n(2k+1)}$$

$$= \sum_{n=0}^{\frac{N}{2}-1} x(n) W_{\frac{N}{2}}^{nk} \cdot W_N^n + \underbrace{\sum_{n=\frac{N}{2}}^{N-1} x(n) W_{\frac{N}{2}}^{nk} \cdot W_N^n}_{\sum_{n=0}^{\frac{N}{2}-1} x(n+\frac{N}{2}) W_{\frac{N}{2}}^{(n+\frac{N}{2})k} \cdot W_N^{n+\frac{N}{2}}} \quad (102)$$

$$= \sum_{n=0}^{\frac{N}{2}-1} x(n) W_{\frac{N}{2}}^{nk} \cdot W_N^n + \sum_{n=0}^{\frac{N}{2}-1} x(n) \underbrace{W_{\frac{N}{2}}^{(n+\frac{N}{2})k}}_{W_{\frac{N}{2}}^{nk}} \cdot \underbrace{W_N^{n+\frac{N}{2}}}_{-W_N^n}$$

$$U(k) = \sum_{n=0}^{\frac{N}{2}-1} W_N^n \left[x(n) - x\left(n+\frac{N}{2}\right) \right] W_{\frac{N}{2}}^{nk}$$

DFTs führt. Für G(k) folgt daraus Gleichung 99, die wiederum nichts anderes ist als eine N/4-Punkt-DFT von g(n) für gerade n und eine mit dem Drehfaktor

$$W_{\frac{N}{2}}^k$$

bewertete N/4-Punkt-DFT von g(n) für ungerade n. Das Ergebnis für den oberen Block in Abbildung 61 unter Berücksichtigung von

$$W_4^0 = W_8^0, \quad W_4^1 = W_8^2, \quad W_4^2 = -W_8^0 \quad \text{und} \quad W_4^3 = -W_8^2$$

zeigt Abbildung 62. Die elementaren 2-Punkt-DFTs (im Englischen wegen der Gestalt ihres Graphen auch Butterflies = Schmetterlinge genannt) werden nun in völliger Analogie als dritter und letzter Dezimierungsschritt durchgeführt und in Abbildung 63 dargestellt.

Alle drei Dezimierungsschritte zusammen zeigt Abbildung 64. Die abgebildete Struktur benötigt allerdings die Eingangsfolgenwerte in einer umsortierten Reihenfolge.

Dabei wird der Index der Originalfolge als rückwärts gelesenes Bitmuster (bit reversed binary) zum Index in der umsortierten Folge.

Also z. B. $x(4) = x(100) \rightarrow x(001) = x(1)$, oder $x(3) = x(011) \rightarrow x(110) = x(6)$. Der erste, dritte, sechste und achte (letzte) Wert der Zeitfolge bleiben an ihren Plätzen, weil 000, 010, 101 und 111 vorwärts und rückwärts gelesen den gleichen Wert ergeben.

FFT durch Dezimierung im Frequenzbereich

Die fortlaufende Dezimierung der Ausgangs-Frequenzfolge (DIF: decimation in frequency) in immer kleinere Teilfolgen führt zu ähnlichen Algorithmen, wie sie die Zerlegung der Zeitfolgen liefert. Wir wollen die Herleitung nur kurz anreißen, sie geschieht in vollständiger formaler Analogie zur DIT.

Auch hier setzen wir voraus, dass die Länge N der zu zerlegenden Ausgangsfolge eine Potenz von zwei ist.

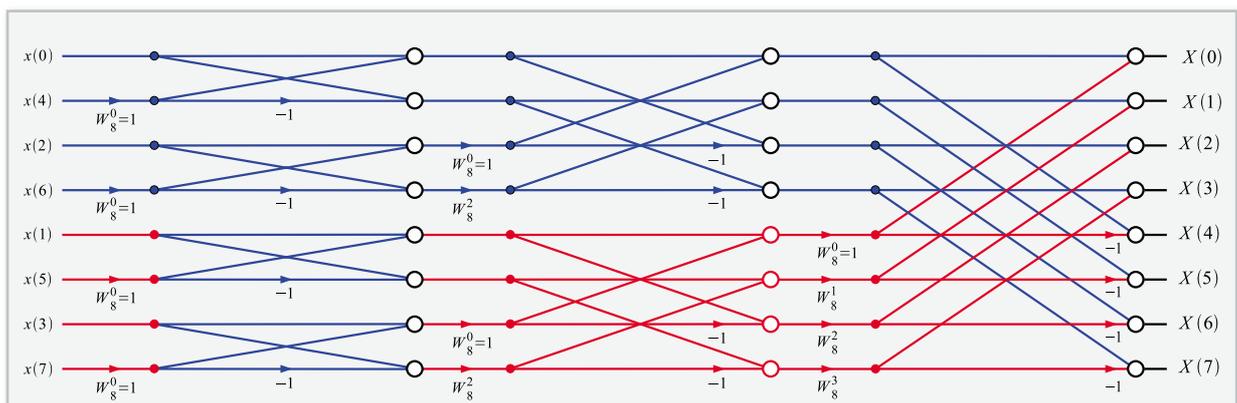


Bild 64: Der vollständig entwickelte DIT-FFT-Signalgraph für 8 Signale

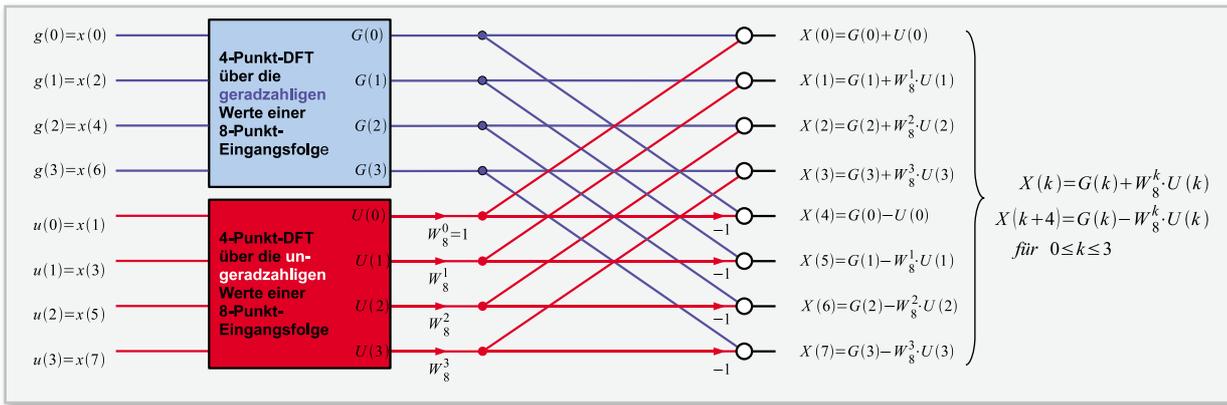


Bild 65: Der FFT-Algorithmus durch fortwährendes Zerlegen der Frequenzfolge (Decimation in Frequency: DIF) führt zu formal ähnlichen Signalgraphen, wie sie die Aufspaltung der Zeitfolge (Decimation in Time: DIT) ergibt.

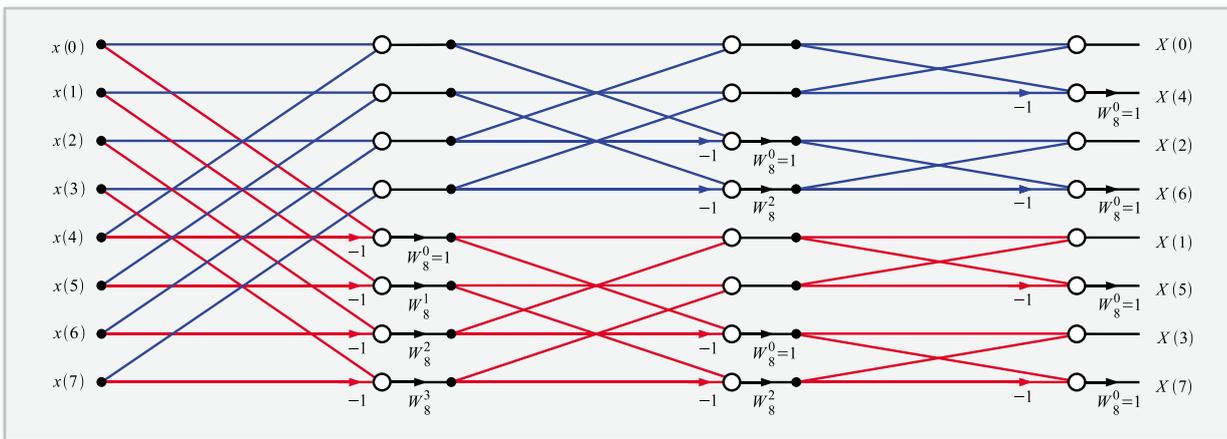


Bild 66: Der vollständig entwickelte DIF-FFT-Signalgraph für 8 Signalpunkte

Wir gehen von der Definitionsgleichung 85 der DFT aus und bilden die Teilfolgen $G(k)$ aus den geraden und $U(k)$ aus den ungeraden Werten von $X(k)$ (Gleichung 100).

Jetzt werden $G(k)$ und $U(k)$ jeweils in zwei Summen aufgeteilt, welche die ersten $N/2$ und letzten $N/2$ Punkte enthalten. Zunächst führen wir das in Gleichung 101 für $G(k)$ durch. $U(k)$ wird wie in Gleichung 102 in entsprechender Weise zerlegt.

Die Gleichungen 101 und 102 lassen sich für $N = 8$ unmittelbar in einen Graphen gemäß Abbildung 65 umsetzen.

Nach einer abermaligen Zerlegung der beiden Frequenzausgangsfolgen verbleibt als Ergebnis der Graph in Abbildung 66. Man sieht, dass jetzt die Eingangsfolgenwerte $x(n)$ in natürlicher Reihenfolge vorliegen, die Ausgangsfolgenwerte $X(k)$ dagegen in eine geänderte Reihenfolge durch Bitumkehr der Indizes (bit reversal) umzusortieren sind.

Bei genauer Betrachtung sowohl der DIT- als auch der DIF-Signalflussgraphen fällt auf, dass N Speicherplätze für die Variablen genügen, weil die N Eingangsgrößen des Graphen $x(0), x(1), \dots, x(n), \dots, x(N-1)$ nach der Verarbeitung in der ersten Berechnungsstufe für den weiteren Verlauf der Berechnung nicht mehr erforderlich sind.

Ihre Speicherplätze können also zur Zwischenablage der Ausgangsgrößen der ersten Berechnungsstufe überschrieben und als Eingangsgrößen für die zweite Berechnungsstufe verwendet werden usw., bis nach vollständiger Durchführung der FFT dort die Ergebniswerte $X(0), X(1), \dots, X(k), \dots, X(N-1)$ stehen. Diese Vorgehensweise wird als „in place operation“ bezeichnet.

Effizienzsteigerung der FFT gegenüber der DFT

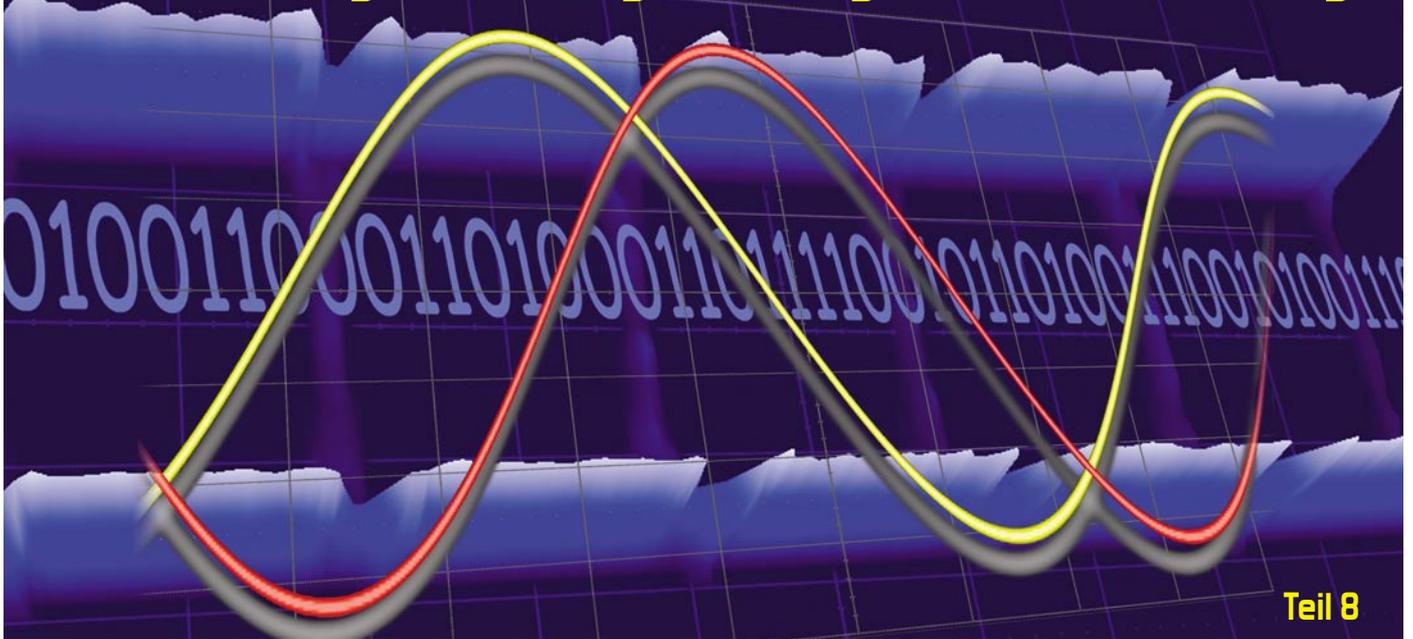
Bei der Berechnung der DFT nach dem Radix-2-Algorithmus ergeben sich erheblich weniger Berechnungen als bei der direkten Umsetzung der DFT-Formel. Bei einer Zeitfolgenlänge von $N = 2^n$ fallen insgesamt $0,5 \cdot N \cdot \text{ld}(N)$ komplexe Multiplikationen (ld : logarithmus dualis = Logarithmus zur Basis 2), $N \cdot \text{ld}(N)$ komplexe Additionen und $(N/2) \cdot \text{ld}(N)$ Inversionen an.

Betrachtet man nur die Zahl der komplexen Multiplikationen (die den größten Rechenaufwand darstellen), bei $n = 6$ (d. h. die Zeitfolge umfasst $N = 2^6 = 64$ Werte), erfordert die FFT $0,5 \cdot 64 \cdot 6 = 192$ komplexe Multiplikationen im Gegensatz zu $N^2 = 4096$ komplexen Multiplikationen bei der DFT. Der Multiplikationsaufwand der FFT beträgt also weniger als 5 % desjenigen bei Anwendung der DFT. Diese Effizienzsteigerung nimmt mit größeren Folgelängen immer weiter zu. Bei $N = 1024$ ist der FFT-Multiplikationsaufwand nur ca. 0,49 % im Vergleich zur DFT.

Sogar bei dem Beispiel in Abbildung 64 sind nur 12 Multiplikationen bei der FFT gegenüber 64 bei der DFT durchzuführen. In der Realität ist die Zahl der Multiplikationen sogar noch etwas geringer, da alle Drehfaktoren mit Exponent 0 den Wert 1 haben.

In Teil 8 dieser Folge beschäftigen wir uns mit den Analyseverfahren für digitale Abtastsysteme, insbesondere der Darstellung durch Differenzgleichungen im Zeitbereich und deren z-Transformation in den Frequenzbereich. **ELV**

Einführung in die digitale Signalverarbeitung



Teil 8

Die z-Transformation als diskrete Variante der Laplace-Transformation ist ein mächtiges Werkzeug zur Analyse und zum Verständnis eines Abtast-LTI-Systems. Sie stellt den Zusammenhang zwischen seiner Differenzgleichung im Zeitbereich und der korrespondierenden Übertragungsfunktion im Frequenzbereich her.

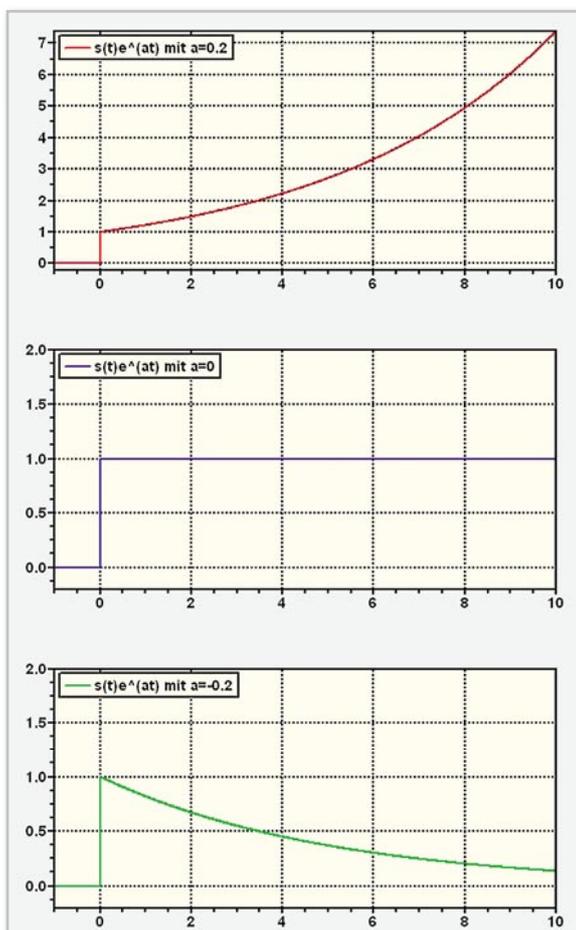


Bild 67: Die Funktion e^{at} klingt mit dem Dämpfungsfaktor $a > 0$ an, ist für $a = 0$ konstant und klingt für $a < 0$ ab.

Die z-Transformation

Zur Analyse digitaler Signale und Systeme (Abtastsysteme) im Frequenzbereich hat sich die z-Transformation bewährt. Vor ihrer Einführung wollen wir noch einmal einen Rückblick auf die Verhältnisse im analogen kontinuierlichen Frequenzbereich werfen. Gleichung (18) definierte die allgemeine Fouriertransformation eines beliebigen Zeitereignisses. Nun nehmen wir als untere Integrationsgrenze $t = 0$ an (kausale Zeitfunktion, die bei $t = 0$ beginnt) und ersetzen die imaginäre Frequenzvariable $i\omega$ durch die komplexe Frequenzvariable $s = \sigma + i\omega$. Die Erweiterung der imaginären Frequenzvariablen $i\omega$ um einen Realteil σ wird analytische Fortsetzung genannt. Damit geht das Fourierintegral über in das nach dem französischen Mathematiker Pierre-Simon Laplace (1749–1827) benannte einseitige Laplace-Integral als spektrale Dichtefunktion nach Gleichung (103).

Je nach Zeitfunktion $f(t)$ darf σ als Realteil von s einen gewissen Wert nicht über- oder unterschreiten, damit das Integral einen endlichen Wert annimmt (konvergiert). Durch das Einfügen des Dämpfungsterms $e^{-\sigma t}$ und den Beginn der Integration zum Zeitpunkt $t = 0$ (d. h. zum Einschaltzeitpunkt des Signals) lassen sich oft einseitige Laplace-Transformierte ermitteln, wo das Fourierintegral mangels Konvergenz nicht existiert.

Wir wollen das am einfachen Beispiel einer bei $t = 0$ beginnenden an- bzw. abklingenden e-Funktion verdeutlichen und setzen also $f(t) = e^{at}$. Für positives a klingt $f(t)$ auf unendlich an, für $a = 0$ geht $f(t)$ in die Sprungfunktion über (behält für alle Zeiten den Wert 1) und für negatives a klingt $f(t)$ auf null

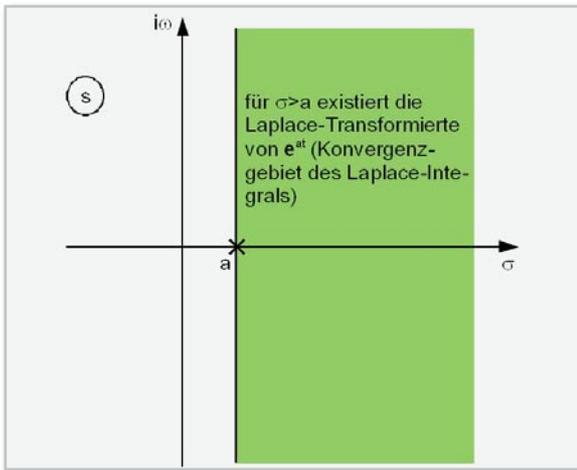


Bild 68: Im grün eingefärbten Gebiet konvergiert das Laplace-Integral von e^{at} .

ab (vgl. Abbildung 67). Aufgrund ihrer Wirkung wird a auch Dämpfungskonstante genannt.

Was bedeutet dies nun in der komplexen s -Ebene? Um das herauszufinden, setzen wir $f(t) = e^{at}$ (reellwertige e -Funktion) in Gleichung (103) ein und erhalten Gleichung (104).

Die Laplacetransformierte existiert nur, wenn $(a - \sigma) < 0$, d. h. $\sigma > a$ ist, weil nur dann das Einsetzen der oberen Grenze $t = \infty$ den Wert null ergibt und dadurch das Integral endlich ist. Man sagt auch, das Integral konvergiert für $\sigma > a$. Dieser Bereich wird deshalb auch Konvergenzgebiet genannt (im Englischen ROC: Region of Convergence).

Abbildung 68 gibt diesen Sachverhalt in Form eines PN-Diagramms (P = Pol, N = Nullstelle) von $L(s)$ wieder. Ein Pol ist eine Nullstelle des Nenners von $L(s)$ und wird mit einem Kreuz gekennzeichnet. Eine Nullstelle ist eine Nullstelle des Zählers von $L(s)$ und wird mit einem Kreis gekennzeichnet. Weil hier der Zähler von $L(s)$ die Konstante 1 ist, hat die Laplacetransformierte der reellwertigen e -Funktion keine Nullstelle. Wenn die imaginäre Achse ($\sigma = 0$) im Konvergenzgebiet liegt, geht auf ihr das Laplace-Integral $L(s)$ in das Fourierintegral $F(i\omega)$ über.

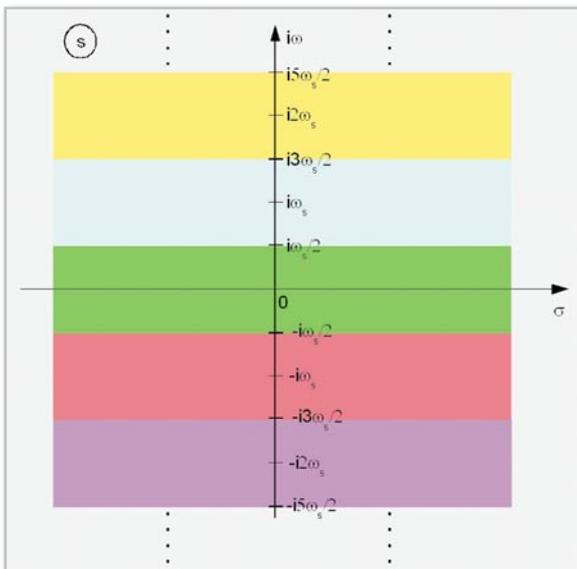


Bild 69: Der gesamte Informationsgehalt eines bandbegrenzten, abgetasteten Signals liegt in jedem einzelnen der abgebildeten Streifen in der Spektralebene. Die Streifen wiederholen sich periodisch in ω_s . Es genügt daher, den grünen Streifen um den Ursprung zu betrachten.

$$L(s) = \int_{t=0}^{\infty} f(t) \cdot e^{-st} dt = \int_{t=0}^{\infty} f(t) \cdot e^{-(\sigma+i\omega)t} dt \quad \text{Laplace-Integral} \quad (103)$$

mit $s = \sigma + i\omega$ komplexe Frequenzvariable

$$L(s) = \int_{t=0}^{\infty} e^{-at} \cdot e^{-st} dt = \int_{t=0}^{\infty} e^{-(a-s)t} dt \quad (104)$$

$$= \frac{1}{a-s} e^{(a-s)t} \Big|_{t=0}^{\infty} = \frac{1}{a-s} e^{\frac{(a-s)t}{<0}} \cdot e^{-i\omega t} \Big|_{t=0}^{\infty}$$

$$= \frac{1}{s-a}$$

$$X_s(s) = \int_{t=0}^{\infty} \sum_{n=0}^{\infty} x(t) \delta(t-nT_s) e^{-st} dt = \sum_{n=0}^{\infty} \int_{t=0}^{\infty} x(t) \delta(t-nT_s) e^{-st} dt \quad (105)$$

$$= \sum_{n=0}^{\infty} x(nT_s) e^{-snT_s} = \sum_{n=0}^{\infty} x(nT_s) e^{-\sigma nT_s} e^{-i\omega nT_s}$$

$$= \sum_{n=0}^{\infty} x(nT_s) e^{-\sigma nT_s} e^{-i2\pi n \frac{\omega}{\omega_s}}$$

$$e^{-i2\pi n \frac{\omega}{\omega_s}} = e^{-i2\pi n \frac{\omega \pm \nu \omega_s}{\omega_s}} = e^{-i2\pi n \frac{\omega}{\omega_s}} \cdot e^{\mp i2\pi n \nu} = e^{-i2\pi n \frac{\omega}{\omega_s}} \quad (106)$$

$= 1$ für allen ν

$$z = e^{sT} \quad (107)$$

$$X(z) = \sum_{n=0}^{\infty} x(n) \cdot z^{-n} \quad \text{einseitige } z\text{-Transformierte} \quad (108)$$

Zur Herleitung der z -Transformation wollen wir zunächst ein abgetastetes Zeitsignal gemäß Gleichung (50) Laplace-transformieren. Gleichung (105) zeigt den Rechengang:

Aus der letzten Zeile von Gleichung (105) wird ersichtlich, dass sich $X_s(s)$ in Vielfachen von $\pm\omega_s$ periodisch wiederholt. Mit anderen Worten: Es ergeben sich die gleichen Werte für $X(s)$ wenn man ω durch $\omega \pm \nu\omega_s$ ($\nu = 0, 1, 2, 3 \dots$) ersetzt. In Gleichung (106) wird das verdeutlicht.

Diese Periodizität der Laplacetransformierten eines abgetasteten (diskreten) Signals zeigt Abbildung 69.

Wir können das an einem Beispiel veranschaulichen. Angenommen, wir haben eine kontinuierliche Zeitfunktion $x(t)$, deren Laplacetransformierte $X(s)$ durch einen idealen Tiefpass mit der Grenzfrequenz $\omega_s/2$ bandbegrenzt wird. Abbildung 70 zeigt ihren Betrag. Wir sehen zwei Pole bei $\sigma < 0$ und $\pm i\omega_p$. Die Nullstellen liegen auf der imaginären Achse ($\sigma = 0$) bei $\pm i\omega_n$. Die zum bandbegrenzten Spektrum gehörende Zeitfunktion $x^*(t)$ können wir nun ohne Aliaseffekte mit ω_s abtasten. Wir erhalten die Wertefolge $x^*(n)$, die wir der Laplace-transformation unterwerfen. Das Ergebnis nach Betragsbildung $|H_p(s)|$ ist wie erwartet periodisch (Abbildung 71). Die Schnittkontur bei $\sigma = 0$ bildet den Betragsfrequenzgang für $s = i\omega$ (natürliche Frequenzen) ab.

Die Periodizität kann vermieden werden, indem man die s -Ebene durch Gleichung (107) auf die z -Ebene abbildet. Zur Schreibvereinfachung wurde in Gleichung (107) $T_s = T$ gesetzt (T_s : Sampling-Intervall). Die Abbildung $z = e^{sT}$ bewirkt, dass die Teile jeden Streifens mit $\sigma < 0$ in Abbildung 69 deckungsgleich auf das Innere eines Einheitskreises in der z -Ebene abgebildet wird. Für $\sigma > 0$ wird auf das Äußere des Einheits-

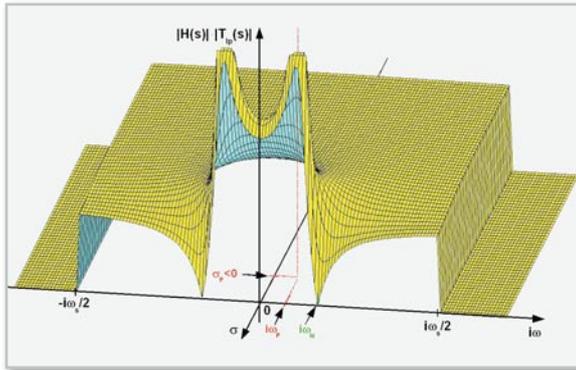


Bild 70: Das Spektrum einer biquadratischen kontinuierlichen Übertragungsfunktion muss bei $\pm i\omega_s/2$ bandbegrenzt werden, um Alias-Effekte zu vermeiden.

kreises abgebildet. Es genügt also, das von $\pm i\omega_s$ begrenzte Gebiet zu betrachten (Abbildung 72). Alle weiteren Streifen der s -Ebene legen sich als deckungsgleiche Blätter über die z -Ebene und enthalten keine weiteren Informationen.

Setzt man Gleichung (107) in Gleichung (105) ein, ergibt sich die Definitionsformel für die einseitige z -Transformierte (Gleichung [108]). Dabei wurden die Schreibvereinfachungen $x(nT_s) = x(n)$ und $X_s(z) = X(z)$ verwendet.

Die z -Transformierte einer Zeitfolge $x(n)$ existiert nur dann, wenn die Summe nach Gleichung (108) konvergiert, d. h. einen endlichen Summenwert aufweist.

Ein Beispiel für die Anwendung von Gleichung (108): Zu berechnen ist die z -Transformierte einer kausalen, reellen Exponentialfolge $x(n) = a^n$. Kausal bedeutet, dass alle Folgenwerte für $n < 0$ null sind. Für $a = 1,1$ ergeben sich also die Folgenwerte ... $x(-2) = 0$, $x(-1) = 0$, $x(0) = 1$, $x(1) = 1,1$, $x(2) = 1,21$, $x(3) = 1,33$, $x(4) = 1,46$, $x(5) = 1,61$, $x(6) = 1,77$, $x(7) = 1,95$ usw. Gleichung (109) zeigt den Rechengang.

Bei der Herleitung wurde die Summenformel für eine unendliche geometrische Reihe gemäß Gleichung (110) verwendet.

Für $a = 1$ geht die Exponentialfolge in eine Sprungfolge über, deren Werte ab dem Abtastindex 0 konstant den Wert 1 aufweisen, also $x(-2) = 0$, $x(-1) = 0$, $x(0) = 1$, $x(1) = 1$, $x(2) = 1$, $x(3) = 1$ usw. Die z -Transformierte $z/(z - 1)$ erhalten wir aus Gleichung (109), indem wir darin $a = 1$ setzen.

Nicht immer ist die Transformation so einfach wie in den obigen Beispielen. Aber es gibt umfangreiche Korrespondenztabelle, aus denen man Nutzen ziehen kann bei der Transformation und insbesondere bei der Rücktransformation aus dem z -Bereich in den diskreten Zeitbereich. Tabelle 3 zeigt einige wichtige Korrespondenzen für die Laplace- und die z -Transformation.

An einem weiteren Beispiel wollen wir die Auswirkungen der Transformation einer im s -Bereich gegebenen komplexen Übertragungsfunktion $H(s) = U_2(s)/U_1(s)$ in den z -Bereich studieren. Wir gehen davon aus, dass das zugrunde liegende System linear und zeitinvariant ist (LTI-System). Dann lässt sich seine Übertragungsfunktion als Quotient zweier Polynome in s (Zählergrad \leq Nennergrad) darstellen. Abbildung 70 stellt den Betrag einer solchen Übertragungsfunktion mit Bandbegrenzung bei $\pm i\omega_s/2$ dar. Die Übertragungsfunktion mit der Variablen s zeigt Gleichung (111).

- Die Nullstellen des Zählerpolynoms sind die Nullstellen der Übertragungsfunktion.
- Die Nullstellen des Nennerpolynoms bezeichnet man als Pole der Übertragungsfunktion, weil diese hier den Wert ∞ annimmt.

In der linken Hälfte von Abbildung 73 sehen wir oben links das dreidimensionale „Betragsgebirge“ und unten links eine zweidimensionale Darstellung der Lage der Nullstellen (dargestellt als o) und der Pole (dargestellt als x) in der s -Ebene. Die Betragsdarstellung wurde für $\sigma > 0$ abgeschnitten (wir erinnern uns: $s = \text{Re}[s] + i\text{Im}[s] = \sigma + i\omega$). Die Schnittkontur ist der Betragsfrequenzgang von $H(s)$ für $s = i\omega$.

Die dreidimensionale Darstellung lässt sich folgendermaßen veranschaulichen. Stellen wir uns ein elastisches Tuch vor, das flach auf der s -Ebene liegt. An den Nullstellen wird das Tuch an der s -Ebene angeheftet und an den Polen nach oben gedrückt. Dadurch modelliert das Tuch eine Oberfläche mit Bergen (Pole) und Tälern (Nullstellen). So wird verständlich, dass sich Pole und Nullstellen mit wachsendem Abstand zur imaginären Achse immer weniger auf den Betragsfrequenzgang auswirken. Dieser ergibt sich ja aus dem senkrechten Schnitt durch das Gebirge auf Höhe der imaginären Achse und spiegelt dann nur die Verhältnisse auf den Ausläufern der entfernten Pole wider. Umgekehrt wirken sich Pole und Nullstellen umso stärker aus, je näher sie an die imaginäre Achse heranrücken.

Die Transformation von $H(s)$ gemäß Gleichung (111) in $H(z)$ erfolgt durch Umrechnen der Pole und Nullstellen entsprechend der Definition von z in Gleichung (107). Es ergibt sich Gleichung (112).

Wir wollen das einmal in Gleichung (113) konkret vorführen. Dabei wird die Abtastperiode T gleich 1 gesetzt, was einer Normierung der komplexen Frequenz s entspricht.

Aus den Nullstellen und Polen im z -Bereich lassen sich durch die Anschrift in Gleichung (114) (Produkt der Linearfaktoren) das Zähler- und das Nennerpolynom und damit die Übertragungsfunktion in z berechnen.

Die dreidimensionale Darstellung von $H(z)$ rechts oben in Abbildung 73 erhält man, indem man für ein regelmäßiges Raster in der z -Ebene den Betrag von $H(z)$ nach Gleichung (114) berechnet. Ein anderer Weg führt zum Ziel, indem man s in $H(s)$ nach Gleichung (111) durch die nach s aufgelöste Definitionsgleichung (107) von z substituiert, d. h. durch $(\ln|z|)/T$ mit $T = 1$ (vergleiche Gleichung [115]), und den Betrag bildet.

Nun wäre noch zu klären, wie der Betragsfrequenzgang von $H(z)$ zu gewinnen ist. Wir

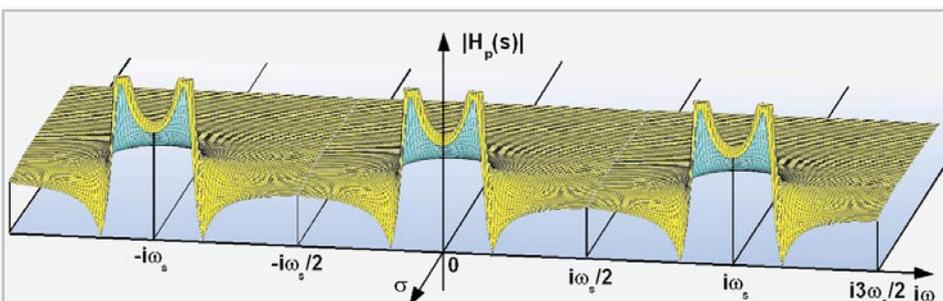


Bild 71: Durch die Abtastung mit ω_s wiederholt sich das Abtastspektrum periodisch in ω_s .

erinnern uns, dass die imaginäre Achse der komplexen s-Ebene zwischen $-\omega_s/2 < \omega < \omega_s/2$ auf den Einheitskreis in der z-Ebene abgebildet wird (vergl. Abbildung 72). Deshalb müssen wir in Abbildung 73 oben rechts die Werte finden, deren Koordinaten in der z-Ebene einen Abstand von 1 zum Ursprung aufweisen ($|z| = 1$). Am einfachsten kann man dies tun, indem wir alle Werte von $|H(z)|$ gleich null setzen, für die $|z| > 1$ gilt. So stanzt man gewissermaßen den Teil von $|H(z)|$ aus, der innerhalb des Einheitskreises liegt. Die Stanzfläche beschreibt den Betrag des Frequenzgangs im z-Bereich (Abbildung 74). Die untere Hälfte von Abbildung 73 zeigt in zwei PN-Diagrammen die genaue Lage von Polen (x) und Nullstellen (o) in der s- bzw. der z-Ebene.

Aus der Lage der Pole und Nullstellen in der s- bzw. z-Ebene kann man somit Rückschlüsse auf den Frequenzgang ziehen. Mit einem PN-Editor lässt sich durch gezieltes Anordnen der Pole und Nullstellen ein gewünschter Frequenzgang erzeugen.

Vertiefung

Bei jedem LTI-System ist das Verhältnis zwischen Ausgangsgröße U_2 (Reaktion) und Eingangsgröße U_1 (Anregung) als Quotient zweier Polynome in s oder z darstellbar. Man nennt diesen Quotienten die komplexe Übertragungsfunktion des LTI-Systems in der Gestalt einer gebrochen rationalen Funktion in s oder z. Die Nullstellen des Zählerpolynoms sind die Nullstellen der Übertragungsfunktion, die Nullstellen des Nennerpolynoms seine Pole.

Ein Polynom vom Grad n hat die allgemeine Gestalt gemäß Gleichung (116). Es hat n Nullstellen (Wurzeln), d. h. es gibt n Werte für x, an denen y den Wert null annimmt. Bei Polynomen mit reellen Koeffizienten hat jede komplexe Wurzel x_{01} ein konjugiert komplexes Gegenstück x_{02} , was Gleichung (117) vermitteln soll. Reelle und komplexe Wurzeln können auch jeweils in einer gewissen Vielfachheit auftreten. Weil sich jedes Polynom durch seine Wurzeln in Form eines Produkts von Linearfaktoren darstellen lässt, gilt: Gleichung (116) = Gleichung (118).

Beispiel 1:

Die reellen Wurzeln 1, 2, 3, 4 führen zu Gleichung (119).

Beispiel 2:

Die konjugiert komplexen Wurzeln $1 \pm i2$ und $3 \pm i4$ sowie die reelle Wurzel 5 führen zu Gleichung (120).

Beispiel 3:

Die doppelten konjugiert komplexen Wurzeln $1 \pm i2$ und die einfache konjugiert komplexe Wurzel $3 \pm i4$ sowie die reelle Wurzel 5 führen zu Gleichung (121).

Abbildung 75 stellt die Nullstellen von Gleichung (121) in der komplexen x-Ebene grafisch dar.

Die Beispiele bestätigen, dass konjugiert komplexe und reelle Wurzeln zu Polynomen mit reellen Koeffizienten gehören. Mit anderen Worten: Komplexe Pole und Nullstellen treten

Bild 73: Das Spektrum einer kontinuierlichen biquadratischen Übertragungsfunktion im s-Bereich (oben links) und ihr PN-Diagramm (unten links). Auf der rechten Seite die dreidimensionale Darstellung des Abtastspektrums im z-Bereich (oben rechts) und die zugehörige PN-Anordnung (unten rechts).

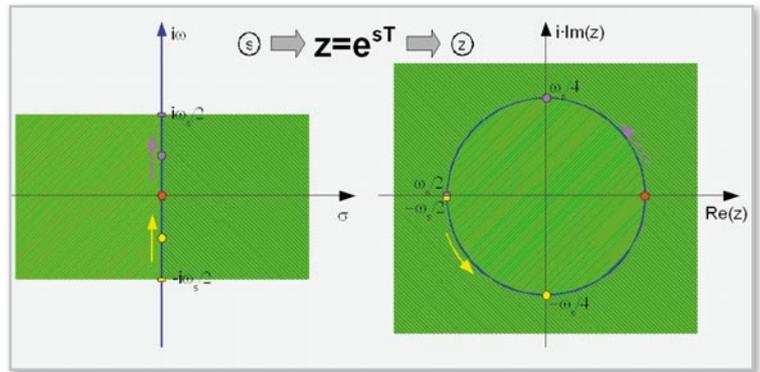
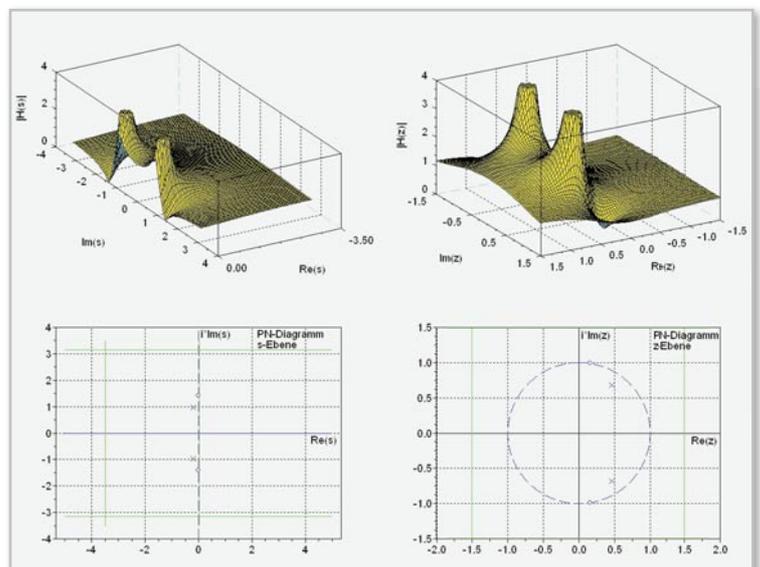


Bild 72: So bildet sich die s-Ebene über die Transformation $z = e^{sT}$ auf die z-Ebene ab.

Funktion	kontinuierliche Zeitfunktion	Laplace-Transformierte	diskrete Zeitfunktion	z-Transformierte
Dirac-Stoß	$\delta(t)$	1	$\delta(k)$	1
verzögerter Diracstoß	$\delta(t-t_0)$	$e^{-t_0 s}$	$\delta(k-k_0)$	z^{-k_0}
Einheitssprung	$\sigma(t)$	$\frac{1}{s}$	$\sigma(k)$	$\frac{z}{z-1}$
verzögerter Einheitssprung	$\sigma(t-t_0)$	$e^{-t_0 s} \cdot \frac{1}{s}$	$\sigma(k-k_0)$	$z^{-k_0} \cdot \frac{z}{z-1}$
Rampe	$\sigma(t) \cdot t$	$\frac{1}{s^2}$	$\sigma(k) \cdot k$	$\frac{z}{(z-1)^2}$
quadratische Parabel	$\sigma(t) \cdot t^2$	$\frac{2}{s^3}$	$\sigma(k) \cdot k^2$	$\frac{z(z+1)}{(z-1)^3}$
kubische Parabel	$\sigma(t) \cdot t^3$	$\frac{6}{s^4}$	$\sigma(k) \cdot k^3$	$\frac{z(z+1)(z+2)}{(z-1)^4}$
Parabel n-ten Grades	$\sigma(t) \cdot t^n$	$\frac{n!}{s^{n+1}}$	$\sigma(k) \cdot k^n$	$\frac{\prod_{v=0}^{n-1} (z+v)}{(z-1)^{n+1}}$
Exponentialfunktion	$\sigma(t) \cdot e^{-at}$	$\frac{1}{s+a}$	$\sigma(k) \cdot e^{-ak}$	$\frac{z}{z-e^{-a}}$
mit t multipliziert Exponentialfktn	$\sigma(t) \cdot t \cdot e^{-at}$	$\frac{1}{(s+a)^2}$	$\sigma(k) \cdot k \cdot e^{-ak}$	$\frac{\partial}{\partial a} \frac{z}{z-e^{-a}} = \frac{z \cdot e^{-a}}{(z-e^{-a})^2}$
mit t^n multipliziert Exponentialfktn	$\sigma(t) \cdot t^n \cdot e^{-at}$	$\frac{n!}{(s+a)^{n+1}}$	$\sigma(k) \cdot k^n \cdot e^{-ak}$	$\frac{\partial^n}{\partial a^n} \frac{z}{z-e^{-a}}$
Potenzfunktion	$\sigma(t) \cdot a^t$	$\frac{1}{s - \ln(a)}$	$\sigma(k) \cdot a^k$	$\frac{z}{z-a}$
Sinus	$\sigma(t) \cdot \sin(at)$	$\frac{a}{s^2 + a^2}$	$\sigma(k) \cdot \sin(ak)$	$\frac{z \cdot \sin(a)}{z^2 - 2 \cdot \cos(a) \cdot z + 1}$
exponentiell bedämpfter Sinus	$\sigma(t) \cdot e^{bt} \cdot \sin(at)$	$\frac{a}{(s-b)^2 + a^2}$	$\sigma(k) \cdot e^{bk} \cdot \sin(ak)$	$\frac{z \cdot e^b \cdot \sin(a)}{z^2 - 2 \cdot e^b \cdot \cos(a) \cdot z + e^{2b}}$
Kosinus	$\sigma(t) \cdot \cos(at)$	$\frac{s}{s^2 + a^2}$	$\sigma(k) \cdot \cos(ak)$	$\frac{z(z - \cos(a))}{z^2 - 2 \cdot \cos(a) \cdot z + 1}$
exponentiell bedämpfter Kosinus	$\sigma(t) \cdot e^{bt} \cdot \cos(at)$	$\frac{s-b}{(s-b)^2 + a^2}$	$\sigma(k) \cdot e^{bk} \cdot \cos(ak)$	$\frac{z(z - e^b \cdot \cos(a))}{z^2 - 2 \cdot e^b \cdot \cos(a) \cdot z + e^{2b}}$

Tabelle 3: Laplace- und z-Korrespondenzen



$$X(z) = \sum_{n=0}^{\infty} a^k z^{-k} = \sum_{n=0}^{\infty} (az^{-1})^k = \frac{1}{1-az^{-1}} = \frac{z}{z-a} \quad \text{Konvergenz für } a < |z| < \infty \quad (109)$$

$$\sum_{n=0}^{\infty} c^n = \frac{1}{1-c} \quad c < 1 \quad \text{Summe unendliche geometrische Reihe} \quad (110)$$

$$H(s) = \frac{s^2+2}{s^2+0.4s+1} \quad \text{biquadratische Übertragungsfunktion im } s\text{-Bereich} \quad (111)$$

$$z_{0,1,2} = e^{s_{0,1,2} T} \quad \text{Nullstellen im } z\text{-Bereich} \quad (112)$$

$$z_{\infty,1,2} = e^{s_{\infty,1,2} T} \quad \text{Pole im } z\text{-Bereich}$$

$$z_{0,1,2} = e^{\pm i \cdot 1,4142135624} = \cos(1,4142135624) \pm i \cdot \sin(1,4142135624) = 0,1559436948 \pm i \cdot 0,9877659460 \quad \text{Nullstellen im } z\text{-Bereich}$$

$$z_{\infty,1,2} = e^{-0,2 \pm i \cdot 0,9797958971} = e^{-0,2} [\cos(0,9797958971) \pm i \cdot \sin(0,9797958971)] = 0,4561902602 \pm i \cdot 0,6798606420 \quad \text{Pole im } z\text{-Bereich} \quad (113)$$

$$H(z) = \frac{(z-z_{0,1})(z-z_{0,2})}{(z-z_{\infty,1})(z-z_{\infty,2})} = \frac{z^2 - z(z_{0,1}+z_{0,2}) + z_{0,1} \cdot z_{0,2}}{z^2 - z(z_{\infty,1}+z_{\infty,2}) + z_{\infty,1} \cdot z_{\infty,2}}$$

$$= \frac{z^2 - 0,3118873895z + 1}{z^2 - 0,9123805204z + 0,6703200460} \quad \text{Übertragungsfunktion im } z\text{-Bereich} \quad (114)$$

$$z = e^{sT} \Leftrightarrow s = \frac{1}{T} \ln(z) \quad (115)$$

$$y = a_n \cdot x^n + a_{n-1} \cdot x^{n-1} + a_{n-2} \cdot x^{n-2} + \dots + a_2 \cdot x^2 + a_1 \cdot x + a_0 = \sum_{v=0}^n a_v \cdot x^v \quad (116)$$

$$x_{0,1} = \operatorname{Re}(x_0) + i \cdot \operatorname{Im}(x_0) \quad \text{hat zur Folge} \quad x_{0,2} = \operatorname{Re}(x_0) - i \cdot \operatorname{Im}(x_0)$$

$$x_{0,1/2} = \operatorname{Re}(x_0) \pm i \cdot \operatorname{Im}(x_0) \quad \text{Paarweise konjugiert komplexe Wurzeln} \quad (117)$$

$$y = (x-x_0) \cdot (x-x_1) \cdot (x-x_2) \cdot \dots \cdot (x-x_{n-1}) \cdot (x-x_n) = \prod_{v=0}^n (x-x_v) \quad (118)$$

$$y = (x-1) \cdot (x-2) \cdot (x-3) \cdot (x-4) = x^4 - 10x^3 + 35x^2 - 50x + 24 \quad (119)$$

$$y = [x-(1-i \cdot 2)] \cdot [x-(1+i \cdot 2)] \cdot [x-(3-i \cdot 4)] \cdot [x-(3+i \cdot 4)] \cdot [x-5]$$

$$= x^5 - 13x^4 + 82x^3 - 290x^2 + 525x - 625 \quad (120)$$

$$y = [x-(1-i \cdot 2)]^2 \cdot [x-(1+i \cdot 2)]^2 \cdot [x-(3-i \cdot 4)] \cdot [x-(3+i \cdot 4)] \cdot [x-5]$$

$$= x^7 - 15x^6 + 113x^5 - 519x^4 + 1515x^3 - 3125x^2 + 3875x - 3125 \quad (121)$$

$$H(s) = \frac{\sum_{\mu=0}^M b_{\mu} \cdot s^{\mu} \prod_{\nu=0}^M (s-s_{0\nu})}{\sum_{\nu=0}^N a_{\nu} \cdot s^{\nu} \prod_{\nu=0}^N (s-s_{\infty\nu})} \quad M \leq N \quad (122)$$

Übertragungsfunktion eines kontinuierlichen Systems in s

$$H(z) = \frac{\sum_{\mu=0}^M b_{\mu} \cdot z^{\mu} \prod_{\mu=0}^M (z-z_{0\mu})}{\sum_{\nu=0}^N a_{\nu} \cdot z^{\nu} \prod_{\nu=0}^N (z-z_{\infty\nu})} \quad M \leq N \quad (123)$$

Übertragungsfunktion eines diskreten Systems in z

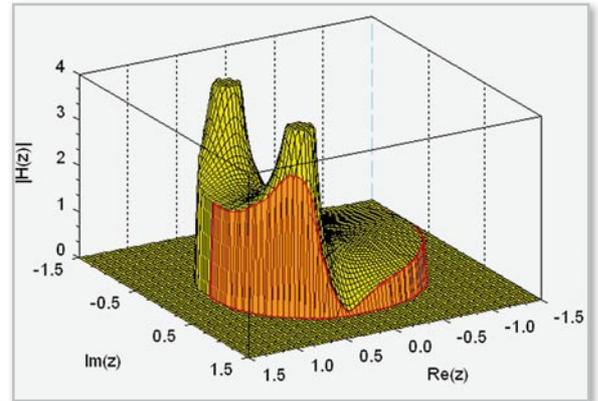


Bild 74: Stanzt man das dreidimensionale Betragsspektrum der Übertragungsfunktion in z im Abstand eins zum Ursprung der z-Ebene aus, lässt sich an der Stanzkontur der Frequenzgang ablesen.

immer spiegelbildlich zur reellen Achse in der s- oder z-Ebene auf, reelle Pole oder Nullstellen liegen stets auf der reellen Achse.

Bei einer LTI-Übertragungsfunktion sind sowohl Zähler als auch Nenner Polynome in s oder z. Aus Kausalitätsgründen (keine Wirkung vor ihrer Ursache!) darf der Grad des Zählerpolynoms (d. h. seine höchste Potenz in s oder z) nur kleiner oder gleich dem Grad des Nennerpolynoms sein. Damit bestimmt der Grad des Nennerpolynoms den Grad einer Übertragungsfunktion (dieser ist als der höhere Grad von Zähler- oder Nennerpolynom definiert), wenn diese ein lineares, zeitinvariantes und kausales System beschreibt.

Damit ist die allgemeine Anschrift der komplexen Übertragungsfunktion eines LTI-Systems in der s-Ebene durch Gleichung (122) plausibel. Die Nullstellen des Zählers tragen als Nullstellen der Übertragungsfunktion den Index 0, die des Nenners als Pole der Übertragungsfunktion den Index ∞ . Der zusätzliche Index s der Koeffizienten des Zähler- und des Nennerpolynoms in der Gleichung dient zur formalen Unterscheidung von den Koeffizienten der Übertragungsfunktion in z, die das Übertragungsverhalten des Abtastsystems beschreibt (Gleichung [123]). Später lassen wir diesen Unterindex weg, weil wir uns ausschließlich im z-Bereich bewegen. Im nächsten Teil dieser Reihe werden wir zeigen, wie sich die Differenzgleichungen im Zeitbereich und ihre Entsprechungen im z-Bereich in Schaltungen umsetzen lassen. **ELV**

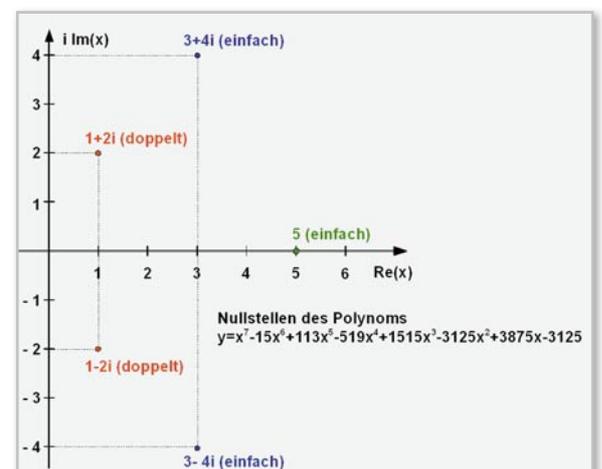


Bild 75: Ein Beispiel für den Zusammenhang zwischen der Koeffizientendarstellung eines Polynoms und dessen Nullstellen.

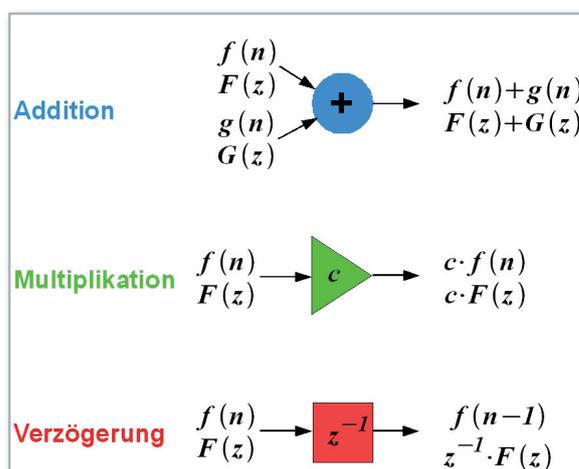


In dieser Folge werden wir sehen, wie sich Übertragungsfunktionen und Differenzgleichungen in eine Schaltungsstruktur umsetzen lassen.

Kanonische Strukturen

Lineare zeitinvariante Systeme (LTI-Systeme), bei denen die Operationen Addition, Multiplikation und Verzögerung gemäß Abbildung 76 auftreten, lassen sich durch Linearkombinationen bewerteter verzögerter Eingangswerte $x(n)$ und Ausgangswerte $y(n)$ in Form einer Differenzgleichung (124) beschreiben.

Die asymmetrische Anschrift von Gleichung 124 macht deutlich, dass das Ausgangssignal des Systems $y(n)$ aus einer Linearkombination verzögerter und bewerteter Eingangswerte (transversaler oder nichtrekursiver Teil) abzüglich einer Linearkombination verzögerter und bewerteter Ausgangswerte (rekursiver Teil) besteht. Zur Übertragungsfunktion in z^{-1} gelangen wir, indem wir die symmetrische Anschrift in Gleichung 124 summandenweise z -transformieren und $H(z) = Y(z)/X(z)$ bilden. In Gleichung 125 wird dies vorgeführt.



Wie schon mehrfach erwähnt, ist bei kausalen Systemen (keine Wirkung vor ihrer Ursache!) $M \leq N$. Man kann auch $M=N$ setzen und die Fälle mit $M < N$ durch Nullsetzen der Koeffizienten mit hohem Index berücksichtigen.

Die Gleichungen (124) und (125) lassen sich direkt in Signalflussgraphen überführen, die so genannten kanonischen Direktstrukturen (oder Direktformen). Um diese Umsetzung in die Gestalt eines Signalflussgraphen zu verdeutlichen, formen wir die asymmetrische Anschrift aus Gleichung (124) mit $a_0=1$ und $M=N$ geringfügig um und unterwerfen sie der z -Transformation (Gleichung 126).

Offensichtlich entspricht die Struktur in Abbildung 77 Gleichung (126). Sie wird als „Erste kanonische Direktform“ bezeichnet. In der Mathematik bedeutet das Wort kanonisch in etwa „sich aus der allgemeinen Struktur oder Definition ergebend“. Zudem wird in der Literatur meistens vorausgesetzt, dass eine minimale Anzahl von Zustandsspeichern (Ver-

Bild 76: Ein zeitdiskretes LTI-Abtastsystem besteht aus Addierern, Multiplizierern und Verzögerern (Speichern).

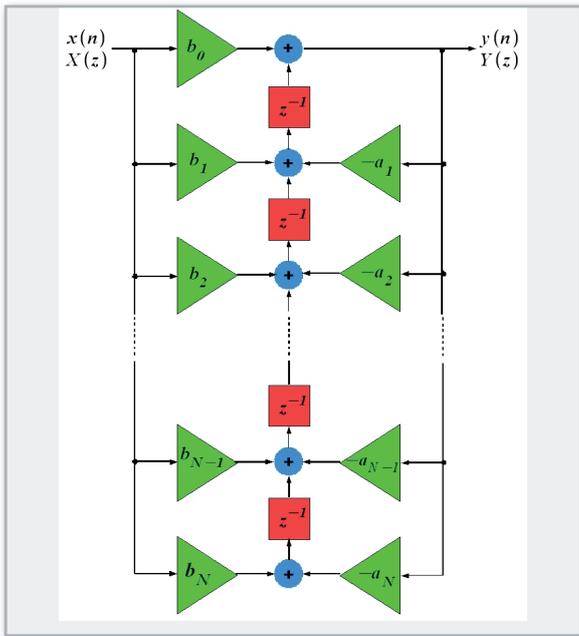


Bild 77: Die „Erste kanonische Direktstruktur“ zur Realisierung der gebrochen rationalen Übertragungsfunktion eines zeitdiskreten LTI-Systems vom Grad N.

zögerer) vorliegt, was hier gegeben ist (Grad N → N Zustandsspeicher). Damit erklärt sich der Name „Erste kanonische Direktform“.

Die rechten Multiplizierer bewerten das Ausgangssignal y(n] bzw. Y(z) mit den Faktoren -a_v (v=1 ... N). Ihr Ausgangssignal wird dann über je einen Addierer in die Verzögererkette rückgeführt (rekursiver Zweig). Entsprechend verfahren die Multiplizierer b_v (v=0 ... N) mit dem Eingangssignal x(n] bzw. X(z) (transversaler Zweig).

Ein Vergleich mit der Übertragungsfunktion H(z) in Gleichung 125 zeigt, dass das Zählerpolynom durch die Koeffizienten b_v und das Nennerpolynom durch die Koeffizienten a_v bestimmt wird. Weil die b_v das Eingangssignal x(n] vor der Einkopplung in die Verzögererkette bewerten, nennt man sie auch Einkoppelkoeffizienten. Entsprechend nennt man die a_v gelegentlich Rückkoppelkoeffizienten.

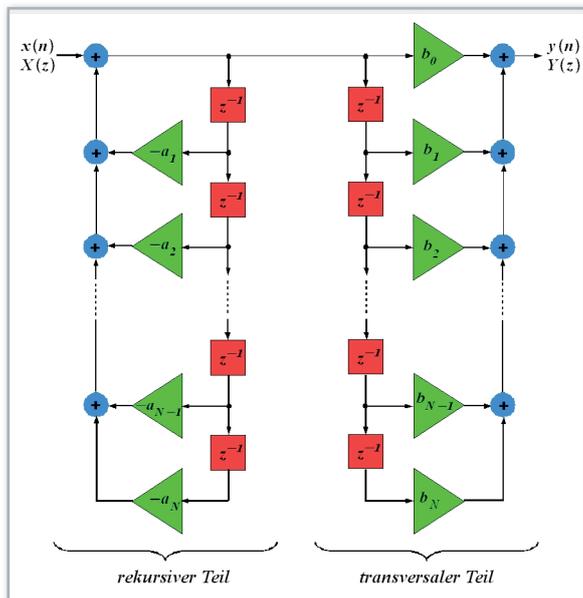


Bild 78: Ein Zwischenschritt auf dem Weg zur „Zweiten kanonischen Direktstruktur“.

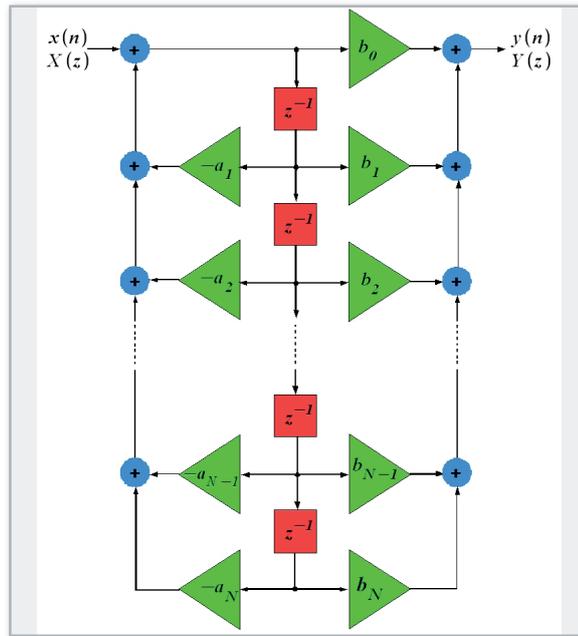


Bild 79: Die „Zweite kanonische Direktstruktur“.

Die „Zweite kanonische Struktur“ kann man aus der umgeordneten Gleichung 125 mit a_0=1 und M=N ableiten. H(z) nach Gleichung 127 wird in Abbildung 78 als Serienschaltung eines rekursiven und eines transversalen (nichtrekursi-

Differenzgleichung eines LTI – Systems (124)

$$b_0 x(n) + b_1 x(n-1) + \dots + b_M x(n-M) = a_0 y(n) + a_1 y(n-1) + \dots + a_N y(n-N)$$

$$\sum_{\mu=0}^M b_\mu \cdot x(n-\mu) = \sum_{\nu=0}^N a_\nu \cdot y(n-\nu)$$

symmetrische Anschrift

$$y(n) = \sum_{\mu=0}^M b_\mu \cdot x(n-\mu) - \sum_{\nu=1}^N a_\nu \cdot y(n-\nu)$$

asymmetrische Anschrift mit a_0=1

$$\sum_{\mu=0}^M b_\mu X(z) \cdot z^{-\mu} = \sum_{\nu=0}^N a_\nu Y(z) \cdot z^{-\nu} \quad z\text{-Transformierte der Diff.gl.} \quad (125)$$

$$X(z) \cdot \sum_{\nu=0}^M b_\nu z^{-\nu} = Y(z) \cdot \sum_{\mu=0}^N a_\mu z^{-\mu} \quad X(z) \text{ und } Y(z) \text{ vor die Summe ziehen}$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{\nu=0}^M b_\nu z^{-\nu}}{\sum_{\mu=0}^N a_\mu z^{-\mu}} = z^{N-M} \cdot \frac{\sum_{\nu=0}^M b_\nu z^{M-\nu}}{\sum_{\mu=0}^N a_\mu z^{N-\mu}}$$

neg. Exponenten von z *pos. Exponenten von z*

$$y(n) = b_0 \cdot x(n) + \sum_{\nu=1}^N [b_\nu \cdot x(n-\nu) - a_\nu \cdot y(n-\nu)] \quad (126) \quad \text{Differenzgleichung}$$

$$Y(z) = b_0 \cdot X(z) + \sum_{\nu=1}^N [b_\nu \cdot X(z) - a_\nu \cdot Y(z)] \cdot z^{-\nu} \quad z\text{-transformierte Diff.gl.}$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{\nu=0}^N b_\nu z^{-\nu}}{1 + \sum_{\nu=1}^N a_\nu z^{-\nu}} = \frac{1}{1 + \sum_{\nu=1}^N a_\nu z^{-\nu}} \cdot \underbrace{\sum_{\nu=0}^N b_\nu z^{-\nu}}_{\text{transversaler Strukturteil}} \quad (127)$$

rekursiver Strukturteil

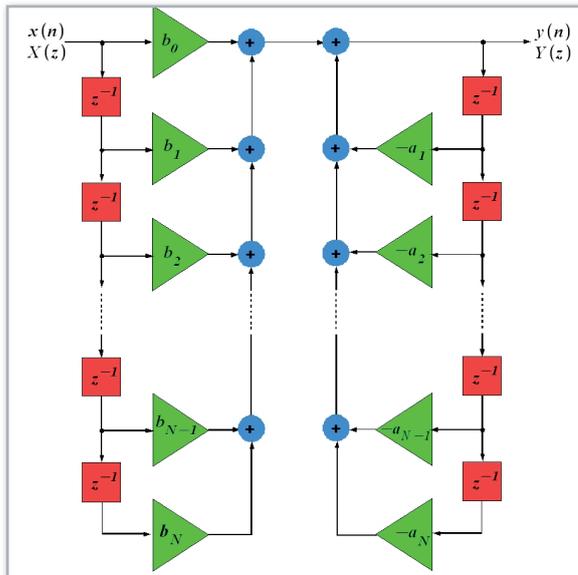


Bild 80: Serienschaltung von transversalem und rekursiven Strukturteil.

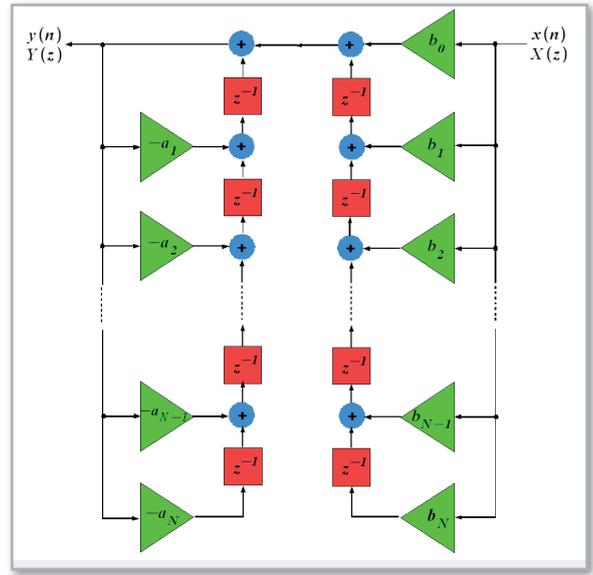


Bild 82: Vertauschung von transversalem und rekursiven Strukturteil aus Bild 81.

ven) Strukturteils interpretiert. Die darin enthaltenen doppelten Verzögererketten können zu einer einzigen zusammengezogen werden. Wir erhalten so die „Zweite kanonische Direktform“ nach Abbildung 79.

Die Zweite kanonische Direktform lässt sich aus der Ersten auch ableiten durch Anwendung eines allgemeinen Gesetzes der Graphentheorie. Demzufolge geht ein Graph durch Vertauschung von Eingang und Ausgang, Verzweigungsknoten gegen Summationsknoten und Umkehrung der Signalfflussrichtung in einen funktionsgleichen „transponierten“ Graphen über.

Die Reihenfolge der Faktoren in Gleichung 127 ist (wie bei jedem Produkt) beliebig. Deshalb kann man die Faktoren vertauschen und einen Strukturgraphen angeben, der aus der Serienschaltung des transversalen und des rekursiven Teils besteht (Abbildung 80). Wird dieser Strukturgraph transponiert, ergibt sich Abbildung 81, in der jetzt der Eingang rechts und der Ausgang links liegt. Das erneute Vertauschen der durch Transposition entstandenen rekursiven und transversalen Strukturteile führt zu Abbildung 82. Durch Spiegeln kommen Eingang und Ausgang wieder in die gewohnte Lage (Ab-

bildung 83). Wenn man jetzt noch die zwei Verzögererketten zusammenfasst, folgt daraus der bereits bekannte Strukturgraph gemäß Abbildung 77, der die „Erste kanonische Direktform“ beschreibt. Damit ist die Umwandlung der „Zweiten kanonischen Direktform“ in die „Erste kanonische Direktform“ vollzogen.

Aus einer Faktorisierung der Übertragungsfunktion $H(z)$ nach Gleichung 127 kann man Teilübertragungsfunktionen niedrigeren Grades ableiten. Werden diese – als erste oder zweite kanonische Direktformen realisiert - in Reihe geschaltet, weist die gesamte Kette wieder das von $H(z)$ vorgegebene Übertragungsverhalten in Form des Produktes aus den Teilübertragungsfunktionen der Kettenglieder auf und wird als „Dritte kanonische Form“ bezeichnet.

Es empfiehlt sich nur Teilübertragungsfunktionen zweiten Grades zu verwenden (biquadratische Blöcke). Ihre Eigenschaften sind überschaubar und die Zuordnung von Polen und Nullstellen auf die Teilblöcke ist sehr flexibel. Jeder

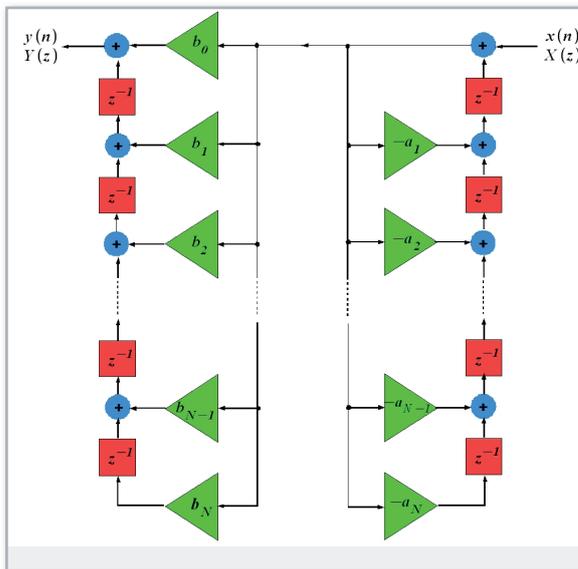


Bild 81: Die transponierte Struktur aus Bild 80.

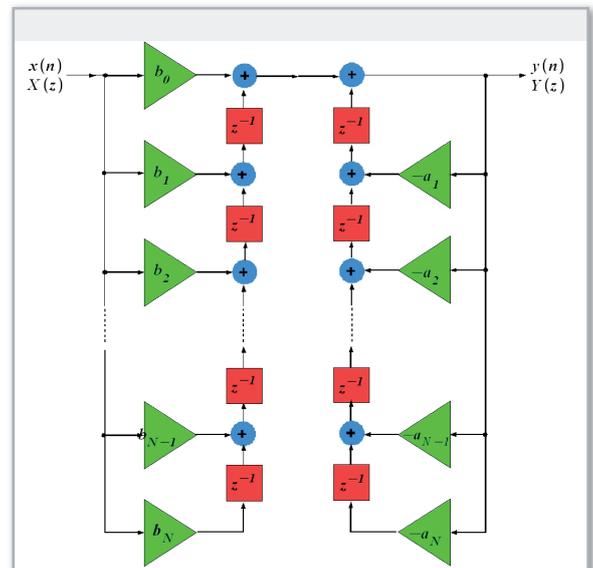


Bild 83: Gespiegelte Struktur aus Bild 82. Nach der Vereinigung der Verzögererketten erhält man wieder die „Erste kanonische Direktstruktur“ aus Bild 77.

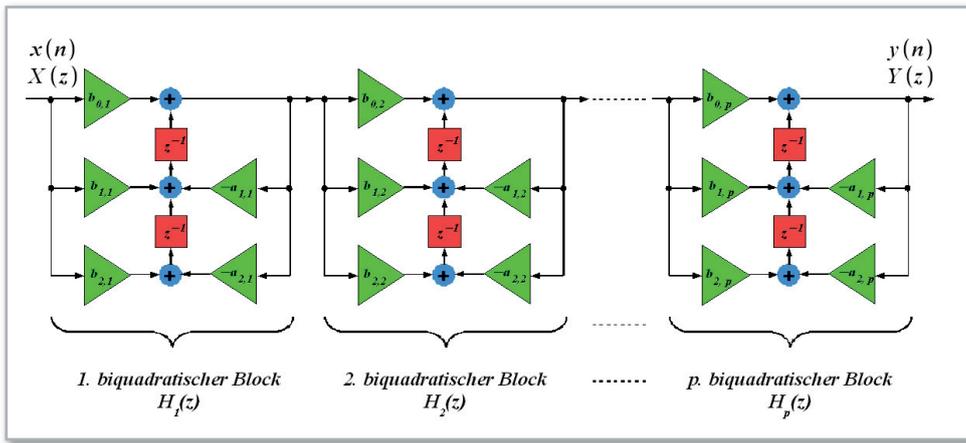


Bild 84: Die „Dritte kanonische Struktur“ besteht aus in Reihe geschalteten Strukturblöcken niedrigeren Grades.

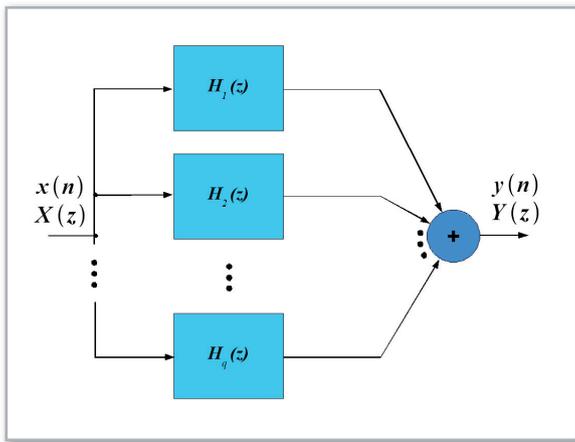


Bild 85: Die „Vierte kanonische Struktur“ besteht aus parallel geschalteten Strukturblöcken niedrigeren Grades.

Block realisiert eigenständig und ohne Wechselwirkung mit den anderen Blöcken – bei entsprechenden reellen Koeffizienten - bis zu zwei konjugiert komplexe oder reelle Pole und Nullstellen.

Die Gesamtübertragungsfunktion $H(z)$ und die eines beliebigen v -ten Blocks $H_v(z)$ zeigt Gleichung 128, den entsprechenden Strukturgraphen Abbildung 84.

Aus einer Partialbruchentwicklung von $H(z)$ lassen sich parallel arbeitende Teilstrukturen ableiten. Dazu wird die Übertragungsfunktion in eine Summe von q Brüchen aufgespalten, von denen jeder eine Teilübertragungsfunktion darstellt. Wenn die Teilübertragungssysteme wiederum kanonisch realisiert werden, entsteht die „Vierte kanonische Form“. Gleichung 129 und Abbildung 85 verdeutlichen dies. Auch hier ist die Zerlegung in Teilübertragungssysteme 1. und 2. Grades sinnvoll. Wegen der Parallelisierung hat diese Struktur mit zunehmender Ordnung der Gesamtübertragungsfunktion wachsende Geschwindigkeitsvorteile gegenüber seriellen.

Die Vorteile der dritten und vierten kanonischen Formen mit Teilübertragungsfunktionen zweiten und ersten Grades im Vergleich mit den 1. und 2. Kanonischen Direktformen in Kurzform:

- Pole und Nullstellen sind einfach zu realisieren
- Teilsysteme sind mehrfach verwendbar
- Ungenauigkeiten der Koeffizienten werden nur im betreffenden Teilsystem wirksam.

sen Umsetzung in eine elektronische Schaltung mit einer Übertragungsfunktion $H(s)$ oder $H(z)$ in Form eines Quotienten aus zwei Polynomen zweiten Grades (Zählerpolynom/ Nennerpolynom=gebrochen rationale Übertragungsfunktion) nennt man Biquad (biquadratische Struktur) oder SOS (Second Order Section). Biquads lassen sich mathematisch recht einfach analysieren und haben eine überschaubare Anzahl von Eigenschaften, aus denen man auf das physikalische Verhalten des durch sie realisierten elektronischen Systems schließen kann.

Übertragungsfunktionen höheren Grades werden vorteilhaft als Serien- oder Parallelschaltungen von Biquads gebildet. Wir wollen Biquads als Bausteine für komplexere Systeme

$$H(z) = \frac{X(z)}{Y(z)} = H_1(z) \cdot H_2(z) \cdot \dots \cdot H_p(z) = \prod_{v=1}^p H_v(z) \tag{128}$$

$$\text{mit } H_v(z) = \frac{b_{2,v}z^{-2} + b_{1,v}z^{-1} + b_{0,v}}{a_{2,v}z^{-2} + a_{1,v}z^{-1} + 1}$$

$$H(z) = \frac{X(z)}{Y(z)} = H_1(z) + H_2(z) + \dots + H_p(z) = \sum_{v=1}^q H_v(z) \tag{129}$$

$$H(z) = \frac{\sum_{v=0}^2 b_v z^{-v}}{1 + \sum_{v=1}^2 a_v z^{-v}} = \frac{b_2 z^{-2} + b_1 z^{-1} + b_0}{a_2 z^{-2} + a_1 z^{-1} + 1} \quad \text{Biquad-Übertragungsfunktion in } z^{-1} \tag{130}$$

$$= \frac{\sum_{v=0}^2 b_v z^{2-v}}{z^2 + \sum_{v=1}^2 a_v z^{2-v}} = \frac{b_0 z^2 + b_1 z + b_2}{z^2 + a_1 z + a_2} \quad \text{Biquad-Übertragungsfunktion in } z$$

Eigenschaften biquadratischer Strukturen

Einen Strukturgraphen gemäß Abbildung 86 und des-

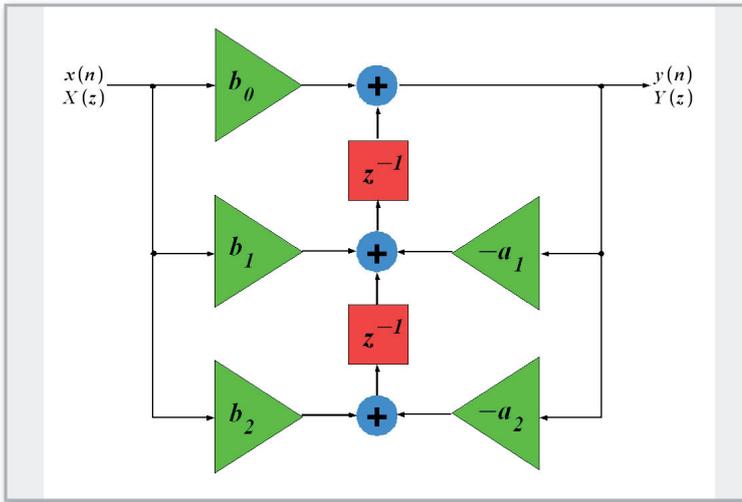


Bild 86: Der Biquad ist ein elementarer Baustein für komplexere LTI-Systeme.

anhand der Übertragungsfunktion (Gleichung 130) und der Differenzgleichung (Gleichung 131) näher untersuchen.

Nullstellen und Pole

Die Nullstellen der Übertragungsfunktion in der z-Ebene erhält man, indem man den Zähler von Gleichung 130 gleich Null setzt und nach z auflöst. Für die Pole muss man das gleiche mit dem Nenner tun.

Sind die Radikanden in den Gleichungen 132 und 133 positiv, ergeben sich zwei reelle Nullstellen bzw. Pole. Sind sie negativ, sind die Nullstellen bzw. Pole konjugiert komplex.

Stabilität

Für die Stabilität reellwertiger, kausaler Abtastsysteme, wie sie Gleichung 130 beschreibt, ist es erforderlich, dass ihre Pole innerhalb des Einheitskreises der komplexen z-Ebene liegen. Die tiefere Begründung dafür liegt in der Eigenschaft stabiler Systeme, auf eine begrenzte Eingangsanregung mit einer ebenfalls begrenzten Ausgangsreaktion zu antworten (BIBO-Stabilität, Bounded Input, Bounded Output). Im Falle konjugiert komplexer Pole muss also deren Betrag < 1 sein,

$$y(n) = b_0 \cdot x(n) + \sum_{v=1}^2 [b_v \cdot x(n-v) - a_v \cdot y(n-v)] \quad \text{Biquad - Differenzgleichung} \quad (131)$$

$$= b_0 \cdot x(n) + b_1 \cdot x(n-1) + b_2 \cdot x(n-2) - a_1 \cdot y(n-1) - a_2 \cdot y(n-2)$$

$$b_0 z^2 + b_1 z + b_2 = 0 \quad (132)$$

$$z_{0,1,2} = \frac{-b_1 \pm \sqrt{b_1^2 - 4b_0 b_2}}{2b_0} \quad \text{Biquad - Nullstellen}$$

$$z^2 + a_1 z + a_2 = 0 \quad (133)$$

$$z_{\infty,1,2} = -\frac{a_1}{2} \pm \sqrt{\frac{a_1^2}{4} - a_2} \quad \text{Biquad - Pole}$$

$$|z_{\infty,1,2}|^2 = \left(-\frac{a_1}{2} - i \sqrt{\frac{a_1^2}{4} - a_2} \right) \cdot \left(-\frac{a_1}{2} + i \sqrt{\frac{a_1^2}{4} - a_2} \right) < 1 \quad (134)$$

daraus folgt

$$a_2 < 1$$

Koeffizientenbedingung für $a_2 > \frac{a_1^2}{4}$ (konjugiert komplexe Pole)

$$-1 < -\frac{a_1}{2} \pm \sqrt{\frac{a_1^2}{4} - a_2} < 1 \quad (135)$$

daraus folgt

$$a_1 < -1 - a_2 \quad \text{und} \quad a_1 < 1 + a_2$$

Koeffizientenbedingung für $a_2 < \frac{a_1^2}{4}$ (reelle Pole)

$$H(z) = \frac{b_0 z^2 + b_1 z + b_2}{z^2 + a_1 z + a_2} \quad \text{mit} \quad b_0=1, \quad b_1=2, \quad b_2=1, \quad a_1=-1,8143, \quad a_2=0,8344 \quad (136)$$

$$H(z) = \frac{(z - z_{01})(z - z_{02})}{(z - z_{\infty 1})(z - z_{\infty 2})} \quad (137)$$

mit $z_{01} = -1$, $z_{02} = -1$, $z_{\infty 1} = 0,90715 + i \cdot 0,1071395$, $z_{\infty 2} = 0,90715 - i \cdot 0,1071395$

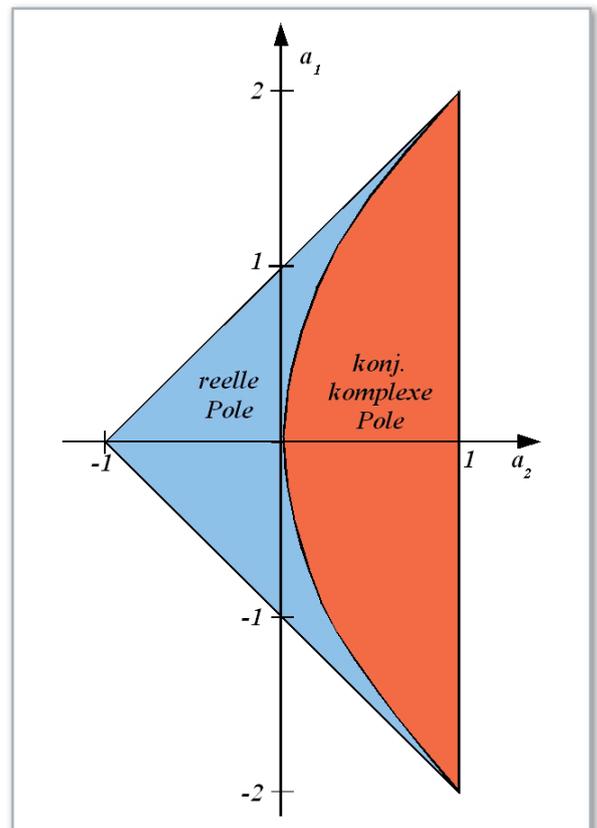


Bild 87: Die Koeffizienten a_1 und a_2 des Nenners einer stabilen Übertragungsfunktion liegen im dreieckigen Bereich.

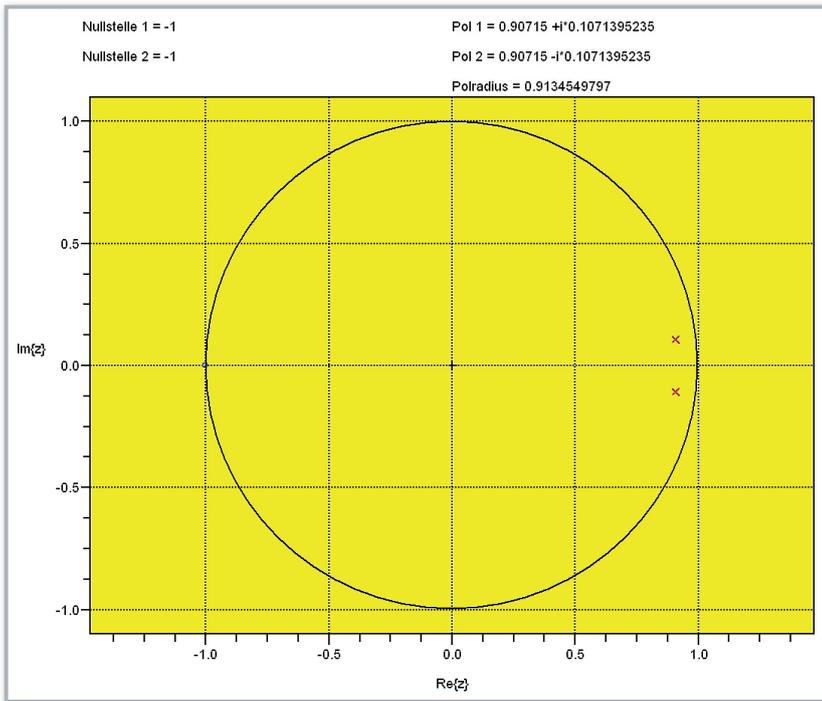


Bild 88: Pol-Nullstellen-Diagramm eines stabilen Biquads.

reelle Pole müssen auf der reellen z-Achse zwischen -1 und +1 liegen. Die sich daraus ergebenden Koeffizientenbedingungen werden in den Gleichungen 134 und 135 angegeben. Die grafische Darstellung der Bedingungen für die Nennerpolynomkoeffizienten einer stabilen, kausalen und reellen biquadratischen Übertragungsfunktion zeigt Abbildung 87. Ein Beispiel soll das verdeutlichen. Wir betrachten die durch Gleichung 136 gegebene Übertragungsfunktion $H(z)$ eines Biquads.

Die faktorisierte Darstellung des Zählers liefert die Nullstellen von $H(z)$, die des Nenners die Pole. Man kann sich überzeugen, dass die Koeffizientenbedingung für konjugiert komplexe Pole (Gleichung 134) erfüllt ist.

Abbildung 88 veranschaulicht das Gesagte. Die doppelte Nullstelle bei $\text{Re}(z)=-1$ markiert ein blauer Kreis. Die konjugiert komplexen Pole werden als zwei rote Kreuze dargestellt. Weil sie im Inneren des Einheitskreises liegen, ist dieser Biquad stabil. Die Impulsantwort im Zeitbereich bestätigt dies, wie Abbildung 89 zeigt. Eine Einheitsimpulsfolge am Eingang bewirkt eine schnell abklingende Impulsfolge am Ausgang. Es ist eine gute Übung, einige erste Werte $y(0)=1$, $y(1)=3,8143$, $y(2)=7,0859$, ... mit den gegebenen Koeffizienten und der Differenzgleichung (Gleichung 131) auszurechnen und mit der Impulsantwort in Abbildung 89 zu vergleichen.

In der nächsten Folge widmen wir uns der Analyse einiger elementarer Grundschaltungen. **ELV**

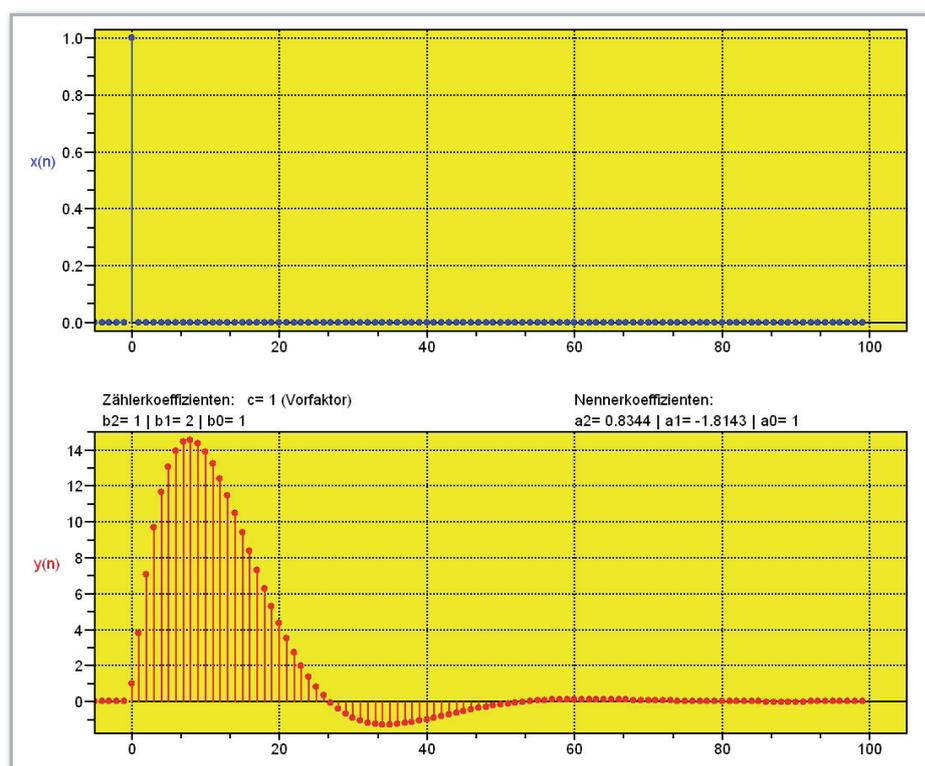


Bild 89: Impulsantwort des Biquads mit P-N-Diagramm aus Bild 88.

Einführung in die digitale Signalverarbeitung

Teil 10

In dieser Folge widmen wir uns der Analyse einiger elementarer Grundschaltungen.

FIR-Grundschaltungen

In Gleichung (127) haben wir gesehen, dass zur Realisierung eines allgemeinen LTI-Abtastsystems ein transversaler und ein rekursiver Strukturteil erforderlich sind. Sind die Koeffizienten des rekursiven Strukturteils alle Null, wird das Ausgangssignal nicht wieder in die Struktur rückgeführt. Es besteht vielmehr ausschließlich aus Werten, die aus dem verzögerten und bewerteten Eingangssignal aufsummiert werden. Ein Eingangsimpuls trägt deshalb solange zum Ausgangssignal bei, bis er alle Verzögerer durchlaufen hat. Dieses nimmt danach wieder den Wert Null an. Eine transversale Struktur weist also eine endliche Impulsantwort auf und wird deshalb auch als FIR-Struktur (Finite Impulse Response = endliche Impulsantwort) bezeichnet. Wir wollen einige solcher FIR-Strukturen untersuchen.

Der digitale Tiefpass 1. Ordnung

Der einfachste aller Tiefpässe ist der „Digitale Tiefpass 1. Ordnung“, wie ihn Abbildung 90 zeigt. Sein Ausgangssi-

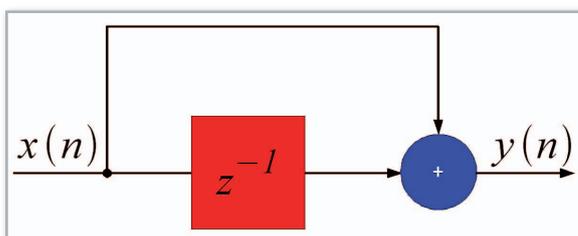


Bild 90: Einfacher geht's nicht. Ein Verzögerer und ein Addierer – fertig ist ein elementarer Tiefpass mit endlicher Impulsantwort.

gnal besteht aus der Summe des aktuellen Eingangsabtastwerts $x(n)$ und seines um eine Abtastperiode T verzögerten Vorgängers $x(n-1)$. Gleichung (138) beschreibt dies. Die dazugehörige z -Transformierte in Gleichung (139) können wir unmittelbar angeben und zur Übertragungsfunktion umordnen.

Wir stellen fest, dass die Übertragungsfunktion $H(z)$ eine Nullstelle bei $z=-1$ und einen Pol bei $z=0$ hat. Abbildung 91 zeigt $H(z)$ in dreidimensionaler Ansicht. Der Einheitskreis in der z -Ebene korrespondiert mit der reellen Frequenz ω . Setzen wir alle Werte von $H(z)$ für $|z|>1$ (also außerhalb des Einheitskreises der z -Ebene) zu Null, gibt uns die grüne Schnittfläche eine Vorstellung vom Frequenzgang des Betrages von $H(z)$ zwischen $0 < \omega < \omega_s/2$. Wir sehen, dass er von $\omega=0$ bis $\omega = \omega_s/2$ immer mehr abnimmt. Deshalb haben wir es mit

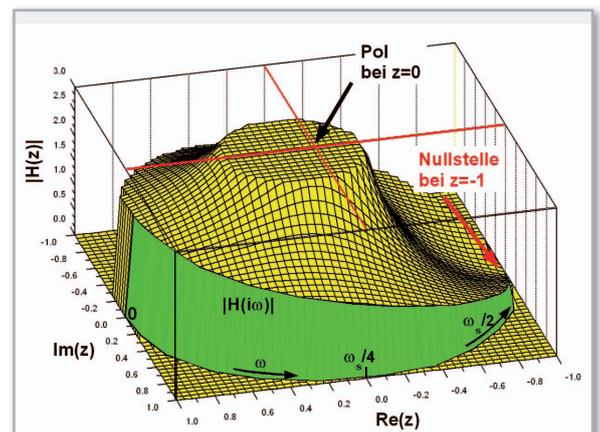


Bild 91: Über dem Einheitskreis der z -Ebene liegt die Übertragungsfunktion des FIR-Tiefpasses 1. Ordnung.

einem Tiefpassverhalten zu tun. Das wollen wir jetzt genauer untersuchen.

Um den Betrags- und Phasengang der Übertragungsfunktion zu berechnen setzen wir in Gleichung (139) die Definition von z ein und formen geringfügig um, wie es in Gleichung (140) gezeigt wird.

Die Aussage von Gleichung (140) wird in Abbildung 92 grafisch dargestellt. Der Betrag (rote Kurve) fällt von 0 bis zur halben Abtastfrequenz kosinusförmig von 2 auf Null ab, die Phase fällt im gleichen Intervall linear von 0 auf $-\pi/2$ ab.

Man kann dieses Ergebnis auch durch anschauliche Überlegungen zumindest qualitativ nachvollziehen.

Beim Betragsgang gehen wir zunächst davon aus, dass das Eingangssignal eine sehr niedrige Frequenz hat, die dementsprechend häufig in einer Periode abgetastet wird. Daraus folgt, dass ein Abtastwert und sein Vorgänger annähernd gleich groß sind und ihre Summe also etwa beim Zweifachen des Eingangssignals liegen wird. Je höher die Eingangsfrequenz wird, umso mehr unterscheiden sich die beiden benachbarten Eingangssignalwerte und die Summe wird immer kleiner. Das geht solange, bis dem Eingangssignal genau zwei Abtastwerte pro Periode (also eine pro Halbwelle) entnommen werden. Jetzt haben diese zwei Abtastwerte den gleichen Betrag, aber entgegengesetzte Vorzeichen. Es kommt zur Auslöschung. Ein weiteres Ansteigen der Eingangsfrequenz ist nicht zulässig, weil dadurch das Abtasttheorem verletzt würde.

Für die Betrachtung der Phase können wir ähnliche Überlegungen anstellen. Bei sehr niedrigen Frequenzen macht sich die Zeitverzögerung des einen Summanden im Ausgangssignals praktisch nicht bemerkbar. Der Grund: Das verzögerte Eingangssignal hat nur einen kleinen negativen Phasenwinkel und das Ausgangssignal nach dem Summierer hat die Hälfte dieses kleinen negativen Phasenwinkels. Im anderen Extremfall, wenn die Signalfrequenz den Wert der halben Abtastfrequenz annimmt, ist das Signal am Ausgang des Verzögerers um $-180^\circ = -\pi$ gegen das Eingangssignal verschoben. Der

$$y(n) = x(n) + x(n-1) \quad \text{Differenzgleichung eines diskreten Tiefpasses 1. Ordnung} \quad (138)$$

$$Y(z) = X(z) + z^{-1} \cdot X(z) \quad (139)$$

$$H(z) = \frac{Y(z)}{X(z)} = 1 + z^{-1} \quad \text{Übertragungsfunktion eines diskreten Tiefpasses 1. Ordnung}$$

$$= \frac{z+1}{z} \quad \text{Übertragungsfunktion in } z$$

$$H(z = e^{sT}) = 1 + e^{-sT} \quad \text{mit der komplexen Frequenz } s = \sigma + i\omega \quad (140)$$

Für reelle Frequenzen ist $\sigma = 0$, woraus folgt

$$H(z = e^{i\omega T}) = 1 + e^{-i\omega T} = \left(e^{i\frac{\omega T}{2}} + e^{-i\frac{\omega T}{2}} \right) e^{-i\frac{\omega T}{2}} = 2 \cos\left(\frac{\omega T}{2}\right) \cdot e^{-i\frac{\omega T}{2}}$$

mit $T = \frac{1}{f_s}$ und $2\pi f_s = \omega_s$ ergibt sich

$$|H(i\omega)| = 2 \left| \cos\left(\frac{\omega T}{2}\right) \right| = 2 \cos\left(\pi \frac{\omega}{\omega_s}\right) \quad \text{Betragsgang für } 0 \leq \omega < \frac{\omega_s}{2}$$

$$\arg(H(i\omega)) = -\frac{\omega T}{2} = -\pi \frac{\omega}{\omega_s} \quad \text{Phasengang für } 0 \leq \omega < \frac{\omega_s}{2}$$

$$y(n) = x(n) - x(n-1) \quad \text{Differenzgleichung eines diskreten Hochpasses 1. Ordnung} \quad (141)$$

$$Y(z) = X(z) - z^{-1} \cdot X(z)$$

$$H(z) = \frac{Y(z)}{X(z)} = 1 - z^{-1} \quad \text{Übertragungsfunktion eines diskreten Hochpasses 1. Ordnung}$$

$$= \frac{z-1}{z} \quad \text{Übertragungsfunktion in } z$$

Summierer addiert also ein Signal mit einem Phasenwinkel von Null Grad und ein zweites mit einem Phasenwinkel von $-\pi$. Der resultierende Phasenwinkel beträgt daher $-\pi/2$.

Der digitale Hochpass 1. Ordnung

Wenn wir in Abbildung 90 das Vorzeichen des verzögerten

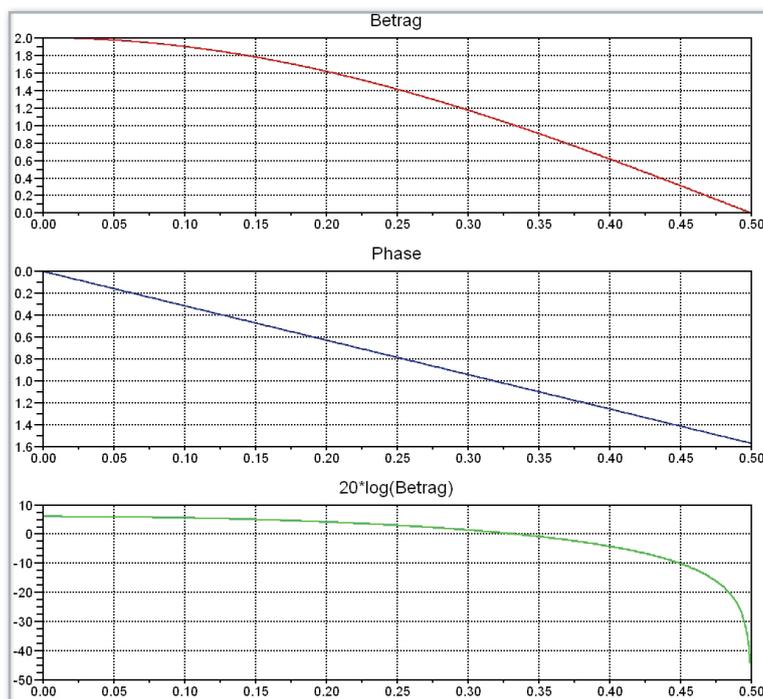


Bild 92: Der Betrag des digitalen Tiefpasses 1. Ordnung fällt mit zunehmender Frequenz von 2 für Gleichspannung auf 0 bei der halben Abtastfrequenz.

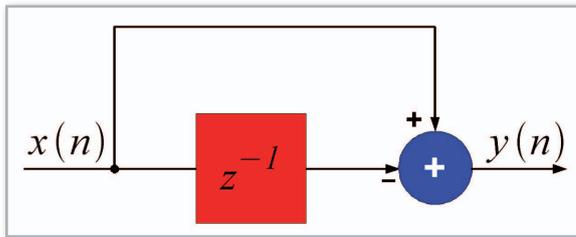


Bild 93: Beim FIR-JP 1. Grades wird der Ausgang des Verzögerers dem Addierer negiert zugeführt.

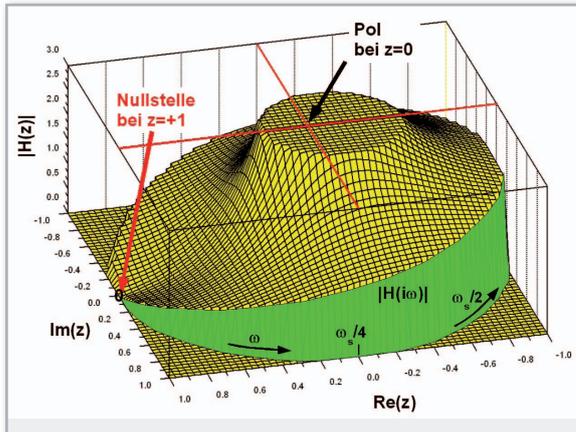


Bild 94: Die Übertragungsfunktion des FIR-Hochpasses 1. Grades über dem Einheitskreis.

Eingangssignals vor der Summation umkehren, erhalten wir die einfachste Form eines digitalen Hochpasses, nämlich den digitalen Hochpass 1. Ordnung (Abbildung 93). Bei einem Gleichspannungseingangssignal heben sich direkter und verzögerter Wert auf und das Ausgangssignal ist Null. Diese Kompensation nimmt mit zunehmender Frequenz ab und ist bei der halben Abtastfrequenz nicht mehr vorhanden. Dann addieren sich die Abtastwerte aus den beiden Halbwellen, die ja unterschiedliche Vorzeichen haben wegen der Vorzeichenumkehr am Summationsknoten und das Ausgangssig-

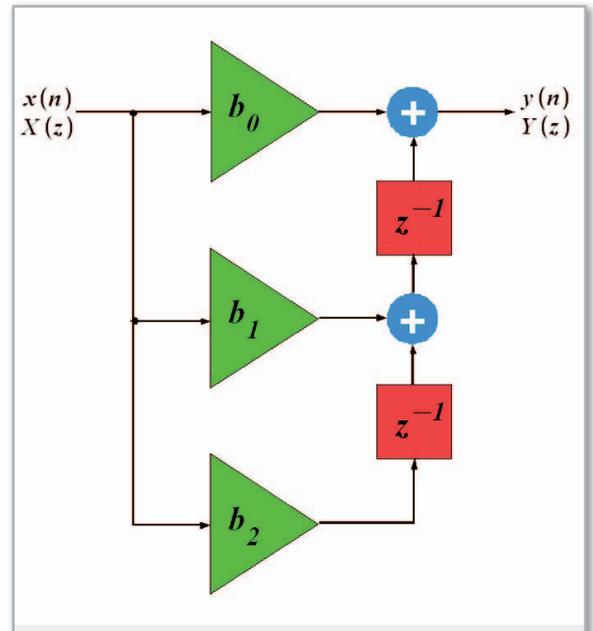


Bild 96: Diese FIR-Struktur 2. Grades ist der Transversalzweig der 1. kanonischen Direktform.

nal hat die doppelte Amplitude wie das Eingangssignal. Wir haben es also mit einem Hochpassverhalten zu tun.

Aus der Struktur in Abbildung 93 können wir wieder direkt deren Differenzgleichung ablesen und in den z -Bereich transformieren. In Gleichung (141) fassen wir dies zusammen. $H(z)$ hat einen Pol bei $z=0$, die Nullstelle liegt bei $z=+1$, wie in Abbildung 94 dargestellt. Wie beim Tiefpass berechnen wir wieder Betrag und Phase der Übertragungsfunktion (Gleichung 142).

Die Aussage von Gleichung (142) ist in Abbildung 95 grafisch dargestellt.

Der digitale Tiefpass 2. Ordnung

Die allgemeine Struktur 2. Ordnung mit endlicher Impulsantwort (FIR) lässt sich aus Abbildung 86 (Biquad in 1. kanonischer Realisierung) ableiten, indem wir den Rekursivzweig löschen (Abbildung 96). In ihrer Differenzgleichung Gleichung (131) und ihrer Übertragungsfunktion Gleichung (130) muss man daher a_1 und a_2 zu Null setzen und erhält die Gleichungen (143) und (144).

Einen digitalen Tiefpass 2. Ordnung erhält man zum Beispiel,

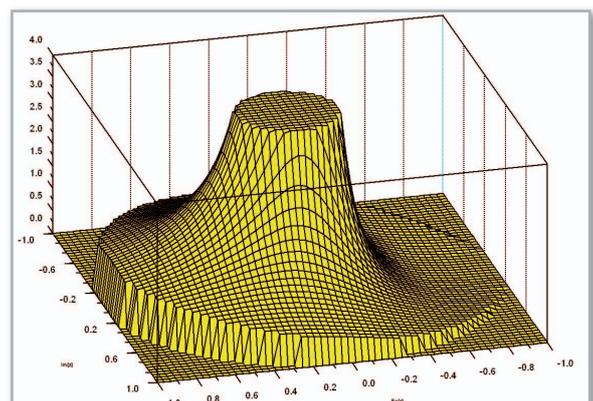


Bild 97: Dreidimensionale Darstellung des FIR-Tiefpasses 2. Grades.

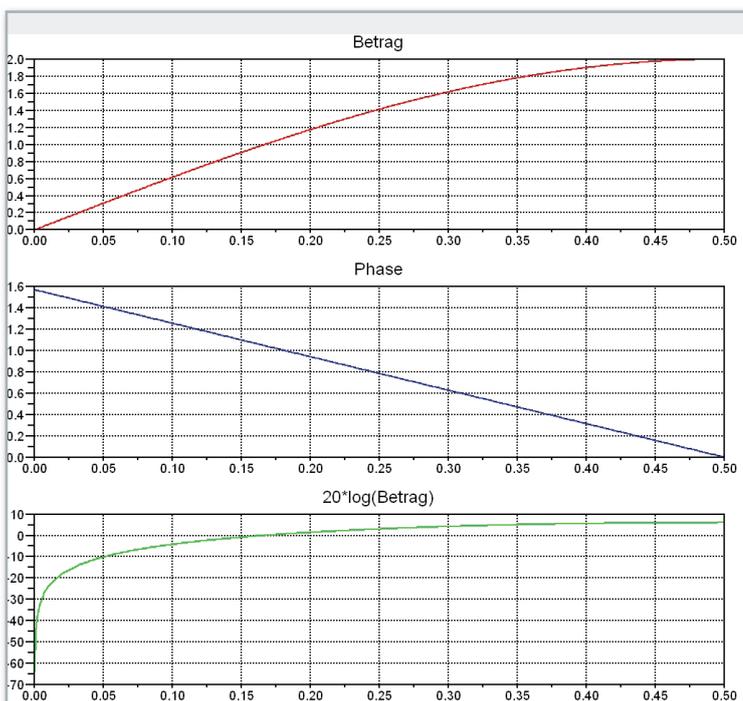
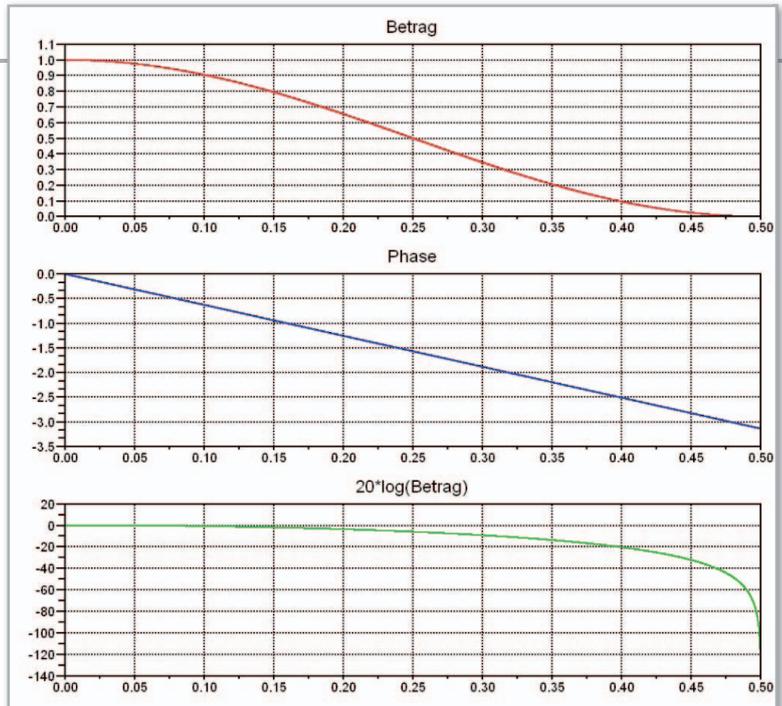


Bild 95: Betrag, Phase (im Bogenmaß) und der Betrag in dB.

Bild 98: Betrag und Phase des FIR-Tiefpasses 2. Ordnung.



indem man $b_0=b_2=0,25$ und $b_1=0,5$ setzt (Beispiele aus Tietze/Schenk Halbleiterschaltungstechnik, 10. Auflage, S. 811 ff). Nach Gleichung (144) hat das System zwei Nullstellen bei $z_0 1,2=-1$ und zwei Pole bei $z_\infty 1,2=0$.

Die Analyse ergibt die dreidimensionale Übertragungsfunktion in Abbildung 97 sowie den Betrags- und Phasengang in Abbildung 98. Die Berechnung soll an dieser Stelle nicht vorgeführt werden. Man kann sich das Leben durch die Beziehungen in den Gleichungen (147) und (148) sehr vereinfachen. Sie gelten allgemein für digitale Biquadschaltungen nach Gleichung (130) und vereinfachen sich bei FIR-Strukturen, weil für die Nennerkoeffizienten gilt: $a_1=a_1=0$.

Auch für die Verstärkung bei $\omega=0$ und $\omega=\omega_s/2$ können wir anschaulich einige Erkenntnisse ableiten:

1. Die Verstärkung bei $\omega=0$ (Gleichspannungsverstärkung) ergibt sich als die Summe aller Multipliziererkoeffizienten ($b_0+b_1+b_2=0,25+0,5+0,25=1$ im Beispiel) (Gleichung (145)).
2. Die Verstärkung bei $\omega=\omega_s/2$ ergibt sich als die Summe aller im Wechsel mit +1 und -1 gewichteten Multipliziererkoeffizienten ($b_0-b_1+b_2=0,25-0,5+0,25=1$ im Beispiel) (Gleichung (146)).
3. Ein Multiplizierer vor dem Eingang der FIR-Struktur kann in das Filter hineingezogen werden, indem man jeden Multipliziererkoeffizienten mit dem Faktor des Eingangsmultiplizierers malnimmt.
4. Die Antwort des FIR-Filters auf einen Einheitsimpuls $\delta(n)$ (Impulsantwort) besteht aus der Folge der Werte der Filterkoeffizienten. Bei einem Filter vom Grad N hat die Impulsantwort N+1 Folgewerte.

Die Ergebnisse für Hochpass, Bandpass und Bandsperre zweiten Grades werden ohne Rechnung in den Abbildungen 99, 100 und 101 zusammengefasst.

In der nächsten Folge untersuchen wir Biquad-Strukturen mit unendlich langer Impulsantwort (IIR: Infinite Impulse Response) und vergleichen die Ergebnisse mit den FIR-Strukturen.

$$H(z=e^{sT})=1-e^{-sT} \quad \text{mit der komplexen Frequenz } s=\sigma+i\omega \quad (142)$$

Für reelle Frequenzen ist $\sigma=0$, woraus folgt

$$H(z=e^{i\omega T})=1-e^{-i\omega T} = \left(e^{i\frac{\omega T}{2}} - e^{-i\frac{\omega T}{2}} \right) e^{-i\frac{\omega T}{2}} = 2i \sin\left(\frac{\omega T}{2}\right) \cdot e^{-i\frac{\omega T}{2}}$$

$$= 2 \sin\left(\frac{\omega T}{2}\right) e^{i\frac{\pi}{2}} e^{-i\frac{\omega T}{2}} = 2 \sin\left(\frac{\omega T}{2}\right) e^{i\left(\frac{\pi}{2}-\frac{\omega T}{2}\right)}$$

Phase

Betrag

mit $T=\frac{1}{f_s}$ und $2\pi f_s=\omega_s$ ergibt sich

$$|H(i\omega)|=2 \left| \sin\left(\frac{\omega T}{2}\right) \right| = 2 \sin\left(\pi \frac{\omega}{\omega_s}\right) \quad \text{Betragsgang für } 0 \leq \omega < \frac{\omega_s}{2}$$

$$\text{arc}(H(i\omega))=\frac{\pi}{2}-\frac{\omega T}{2}=\frac{\pi}{2}-\pi \frac{\omega}{\omega_s} \quad \text{Phasengang für } 0 \leq \omega < \frac{\omega_s}{2}$$

$$y(n)=b_0 \cdot x(n)+b_1 \cdot x(n-1)+b_2 \cdot x(n-2) \quad (143)$$

$$H(z)=b_2 z^{-2}+b_1 z^{-1}+b_0 \quad \text{Biquad-Übertragungsfunktion in } z^{-1} \quad (144)$$

$$= \frac{b_0 z^2+b_1 z+b_2}{z^2} \quad \text{Biquad-Übertragungsfunktion in } z$$

$$|H(i\omega=0)|=\sum_{k=0}^N b_k \quad \text{Verstärkung eines FIR-Filters bei der Frequenz Null (DC)} \quad (145)$$

$$\left| H(i\omega=i\frac{\omega_s}{2}) \right|=\sum_{k=0}^N (-1)^k \cdot b_k \quad \text{FIR-Filterverstärkung bei der halben Abtastfrequenz} \quad (146)$$

$$H(i\omega)=\sqrt{\frac{(b_0^2+b_1^2+b_2^2)+2b_1(b_0+b_2)\cos\left(2\pi\frac{\omega}{\omega_s}\right)+2b_0b_2\cos\left(4\pi\frac{\omega}{\omega_s}\right)}{(1+a_1^2+a_2^2)+2a_1(1+a_2)\cos\left(2\pi\frac{\omega}{\omega_s}\right)+2a_2\cos\left(4\pi\frac{\omega}{\omega_s}\right)}} \quad (147)$$

Betragsfrequenzgang eines diskreten Biquads

$$\varphi(i\omega)=\text{arc tan} \frac{(b_0+a_1b_1+a_2b_2)+(a_1b_0+b_1+a_2b_1+a_1b_2)\cos\left(2\pi\frac{\omega}{\omega_s}\right)+(b_0a_2+b_2)\cos\left(4\pi\frac{\omega}{\omega_s}\right)}{(a_1b_0-b_1+a_2b_1-a_1b_2)\sin\left(2\pi\frac{\omega}{\omega_s}\right)+(b_0a_2-b_2)\sin\left(4\pi\frac{\omega}{\omega_s}\right)} \quad (148)$$

Phasengang eines diskreten Biquads

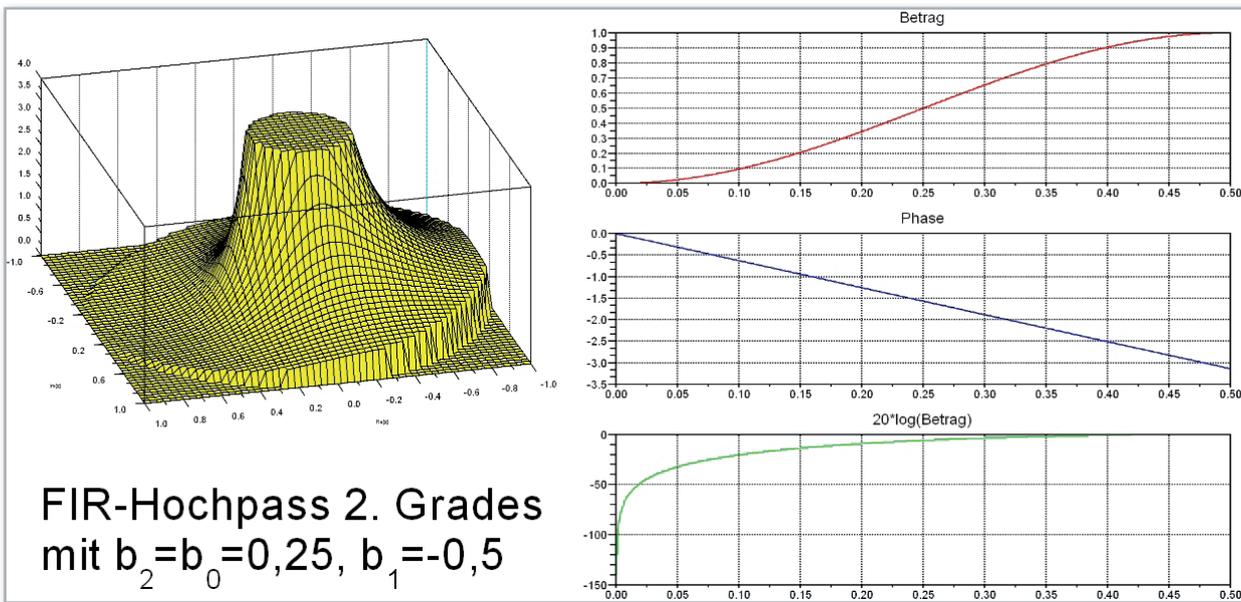


Bild 99: Betrag und Phase des FIR-Hochpasses 2. Ordnung.

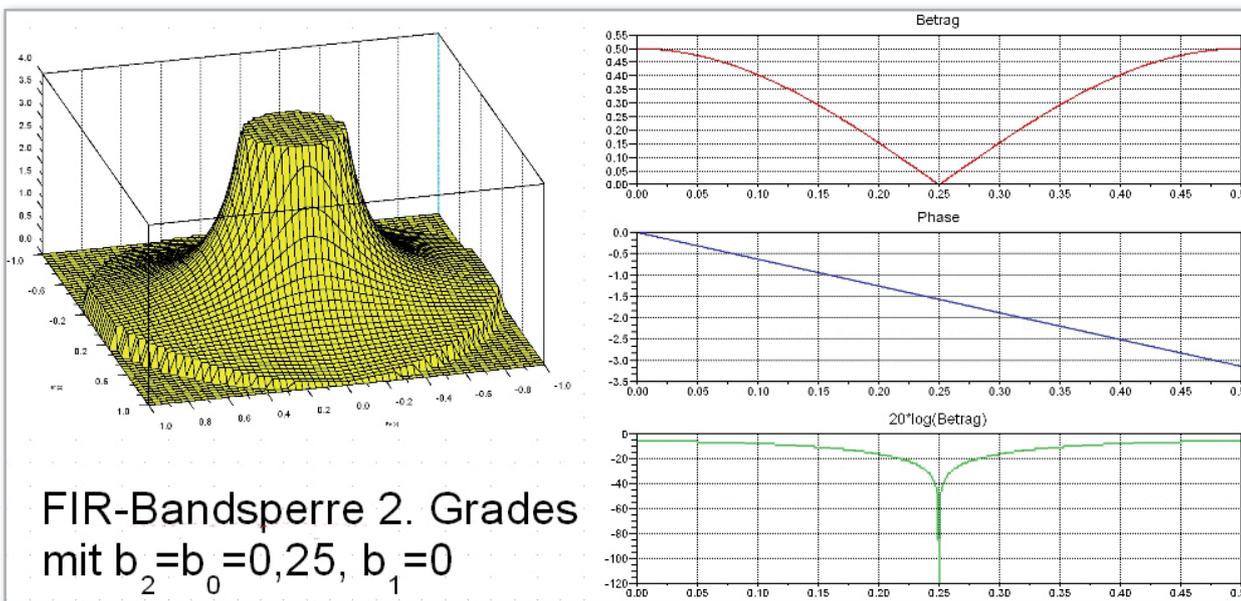


Bild 100: Betrag und Phase der FIR-Bandsperre 2. Ordnung.

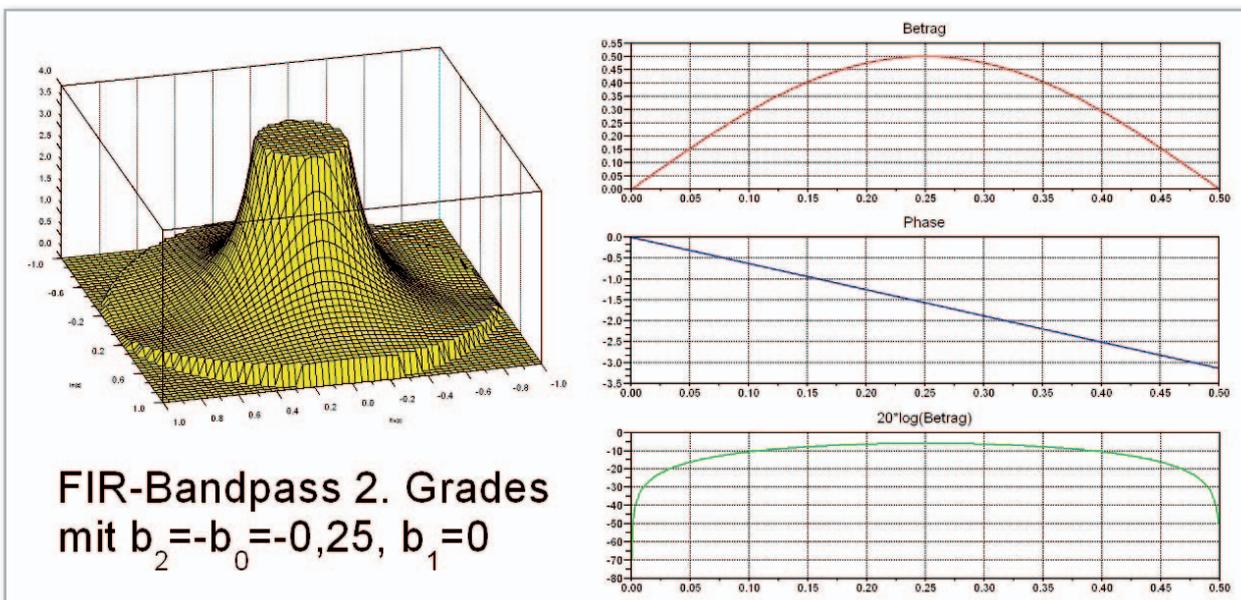


Bild 101: Betrag und Phase des FIR-Bandpasses 2. Ordnung.

Einführung in die digitale Signalverarbeitung

Teil 11

In dieser Folge untersuchen wir Biquad-Strukturen mit unendlich langer Impulsantwort (IIR: Infinite Impulse Response).

Allgemeines

Im letzten Teil haben wir uns mit einfachen Grundschaltungen - Tief-, Hoch- und Bandpässen 1. und 2. Ordnung - mit endlich langer Impulsantwort (FIR-Strukturen) beschäftigt. Ihnen ist gemeinsam, dass das Ausgangssignal nur aus dem bewerteten, verzögerten und aufsummierten Eingangssignal besteht.

In Abbildung 102 sind zwei FIR-Strukturen vom Grad N in einer in der Literatur üblichen Form dargestellt. Sie gehen wie in Folge 9 beschrieben durch Transponieren ineinander über (Vertauschung von Eingang und Ausgang, Verzweigungsknoten gegen Summationsknoten und Umkehrung der Signalflossrichtung) und entsprechen Abbildung 77 und Abbil-

dung 79, indem man alle Rückkoppelkoeffizienten a_v ($v=1 \dots N$) Null setzt. Die Beziehungen zwischen Eingangs- und Ausgangssignal in Gleichung (149) gelten also für beide Strukturen gleichermaßen. Wenn die Rückkoppelkoeffizienten ins Spiel kommen, ändert sich einiges.

IIR-Grundschaltungen

Eine Biquad-Übertragungsfunktion mit unendlich langer Impulsantwort (IIR: Infinite Impulse Response) nach Gleichung (130) setzt $a_1 \neq 0$ und/oder $a_2 \neq 0$ voraus. Damit wird das Ausgangssignal $y(n)$, $Y(z)$ mit a_1 und/oder a_2 bewertet in die Struktur zurückgeführt. Diese Rückkopplung bewirkt bei unendlicher Rechengenauigkeit, dass das Ausgangssignal nach

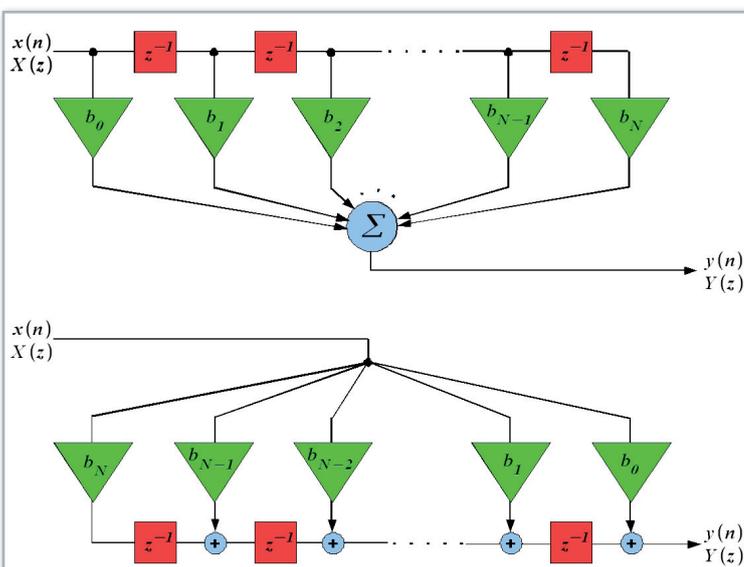
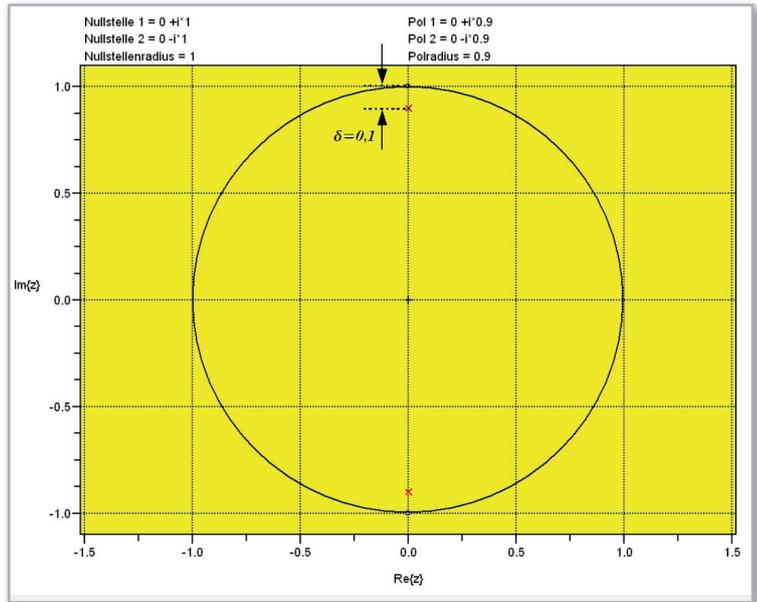


Bild 102: Die Form der Darstellung der Kanonischen Strukturen 1. und 2. Art ohne Rückkoppelkoeffizienten zeigt, dass ihr Ausgangssignal nichts weiter als eine Linearkombination der verzögerten Eingangsfolgenwerte ist. Bei der oberen Struktur spricht man auch von einer „angezapften“ Verzögerungskette (tapped delay line TDL).

Bild 104: Beim Notchfilter sind Pol und Nullstelle auf einer Ursprungsgerade angeordnet: Die Nullstellen auf dem Einheitskreis, die Pole in dessen Inneren.



einer zeitlich begrenzten Anregung durch das Eingangssignal (z. B. durch einen Einheitsimpuls) nicht mehr auf Null abklingt. Bei falscher Wahl der a_v ($v=1, 2$) liegen die Nullstellen des Nennerpolynoms, also die Pole der Biquad-Übertragungsfunktion außerhalb des Einheitskreises der z -Ebene, wodurch die Impulsantwort von Folgenwert zu Folgenwert anwächst. Das Filter ist dann instabil.

Betrags- und Phasengang

Im Gegensatz zum FIR-Filter, das nach Gleichung (127) mit Null gesetzten Rückkoppelkoeffizienten a_i ($i=1 \dots N$) seine Pole bei $z=0$ aufweist, können beim IIR-Filter die Pole durch Wahl der Rückkoppelkoeffizienten an beliebigen Stellen im Einheitskreis der z -Ebene angeordnet werden. Ein komplexer Pol $z_{\infty 1}$ hat dabei stets einen konjugiert komplexen Gegenpol mit $\text{Re}(z_{\infty 2})=\text{Re}(z_{\infty 1})$ und $\text{Im}(z_{\infty 2})=-\text{Im}(z_{\infty 1})$, wird also an der reellen Achse der z -Ebene gespiegelt. Um Verwechslungen zu vermeiden, verwenden wir im Folgenden den Buchstaben M für den Grad der Übertragungsfunktion, weil wir den Vektor zwischen einem Frequenzpunkt auf dem Einheitskreis der z -Ebene und einer Nullstelle mit N kennzeichnen.

Betrags- und Phasengang der Systemfunktion $H(z)$ lassen sich aus der Differenz N_v zwischen dem Frequenzpunkt (hier sind die Frequenzen reell!) und den Nullstellen z_{0i} ($i=1 \dots M$), sowie der Differenz P_v zwischen dem selben Frequenzpunkt und den Polen $z_{\infty i}$ ($i=1 \dots M$) angeben (Gleichung 150), wie es Abbildung 103 am Beispiel von drei Nullstellen und vier Polen veranschaulicht.

Die Konstante c ist für reellwertige Systemfunktionen (d. h. für reellwertige Koeffizienten von Zähler- und Nennerpolynom) ebenfalls reell und ihre Phasenwinkel damit Null. Auf Abbildung 103 übertragen, heißt das

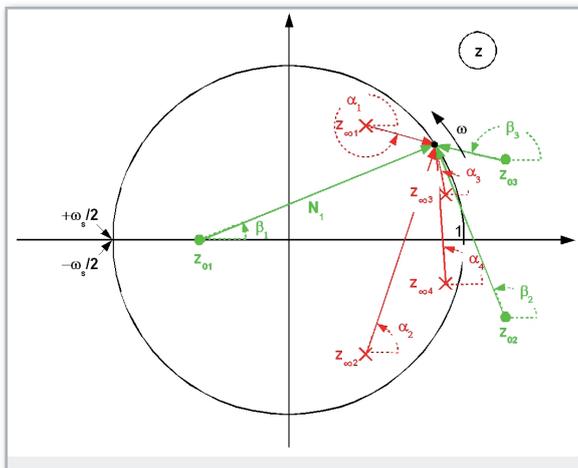


Bild 103: Aus der Lage der Pole und Nullstellen lassen sich Betrag und Phase der Systemfunktion ableiten.

- der Betrag der Systemfunktion ist das Produkt der grünen Pfeillängen, dividiert durch das Produkt der roten Pfeillängen (Zählerbetrag/Nennerbetrag),
- die Phase der Systemfunktion ist die Summe der grünen Phasenwinkel minus der Summe der roten Phasenwinkel (Zählerphase-Nennerphase).

$$y(n) = b_N x(n-N) + b_{N-1} x(n-N+1) + \dots + b_2 x(n-2) + b_1 x(n-1) + b_0 x(n) \quad (149)$$

$$= \sum_{v=0}^N b_v x(n-v)$$

Differenzgleichung eines FIR-Filters vom Grad N

$$\frac{Y(z)}{X(z)} = b_N z^{-N} + b_{N-1} z^{-N+1} + \dots + b_2 z^{-2} + b_1 z^{-1} + b_0$$

$$= \sum_{v=0}^N b_v z^{-v}$$

Übertragungsfunktion eines FIR-Filters vom Grad N

Die Folgenwerte der Impulsantwort entsprechen den Koeffizienten :
 $h(v) = b_v$ für $v=0 \dots N$

(150)

$$H(z = e^{j\omega T}) = c \cdot \frac{\prod_{v=1}^M (e^{j\omega T} - z_{0v})}{\prod_{v=1}^M (e^{j\omega T} - z_{\infty v})}$$

Systemfunktion aus PN-Darstellung

mit $(e^{j\omega T} - z_{0v}) = N_v \cdot e^{j\beta_v}$ und $(e^{j\omega T} - z_{\infty v}) = P_v \cdot e^{j\alpha_v}$ folgt:

$$A(i\omega) = |H(z = e^{j\omega T})| = c \cdot \frac{\prod_{v=1}^M N_v}{\prod_{v=1}^M P_v}$$

Amplitudengang aus PN-Diagramm

$$\varphi(i\omega) = \text{arc}(H(z = e^{j\omega T})) = \sum_{v=1}^M \beta_v - \sum_{v=1}^M \alpha_v$$

Phasengang aus PN-Diagramm

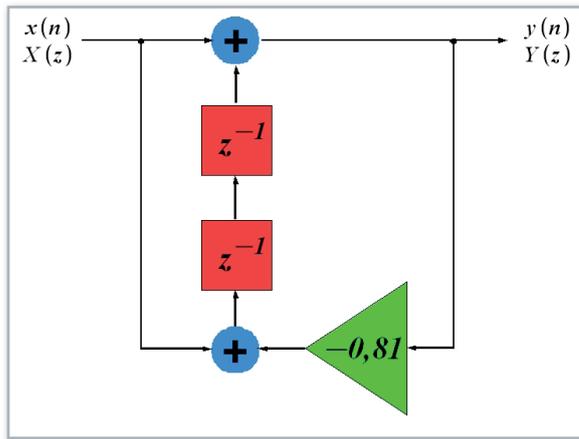


Bild 105: Notchfilter-Struktur mit Polen und Nullstellen nach Abbildung 104.

Beispiel Notchfilter

Ein Notchfilter zeichnet sich durch einen schmalen Einbruch des Betragsfrequenzgangs bei einer bestimmten Frequenz aus. Der Frequenzgang ist an dieser Stelle gewissermaßen „eingekerb“, deshalb auch der deutsche Name Kerbfilter. Zum Beispiel kann ein 50-Hz-Notchfilter den Netzbrumm aus einem Audiosignal ausfiltern. Ein Beispiel für ein IIR-Notchfilter 2. Grades gibt Gleichung (151). Dabei liegt die Nullstelle auf dem Schnittpunkt vom Einheitskreis mit der imaginären Achse der z -Ebene, also beim 0,25fachen der Abtastfrequenz. Der Pol liegt im Abstand δ ($0 < \delta < 1$) vom Einheitskreis entfernt in dessen Inneren auf der imaginären Achse. Pol und Nullstelle treten spiegelbildlich zur reellen Achse auf (Abbildung 104).

Aus der Linearfaktoranschrift für Zähler- und Nennerpolynom lässt sich, wie in Gleichung (151) gezeigt, die Systemübertragungsfunktion $H(z)$ bestimmen.

Durch Vergleich der Koeffizienten von Gleichung (151) mit denen der Systemübertragungsfunktion eines Biquads entsprechend Gleichung (130) ergeben sich $b_1=a_1=0$, $b_2=b_0=1$ und $a_2=0,81$. Damit können wir sofort die erzeugende Struktur in Abbildung 105 angeben, indem wir in Abbildung 86 die Koeffizienten entsprechend setzen. Den Betrag der Systemübertragungsfunktion innerhalb des Einheitskreises der z -Ebene zeigt Abbildung 106. Wir können der dreidimensionalen Darstellung anschaulich entnehmen, dass die „Kerbe“ umso schmaler wird, je näher der Pol an die Nullstelle auf dem Einheitskreis heranrückt.

Abbildung 107 stellt den Betrags- und Phasengang sowie die Verstärkung des Notchfilters dar. Es fällt der Phasensprung um π bei dem 0,25fachen der Abtastfrequenz auf. Er ist anschaulich erklärbar, wenn man die Verhältnisse um $\omega=\omega_s/4$ grafisch veranschaulicht. Dazu werden in Abbildung 108 die Phasenwinkel der Linearfaktoren mit Pol und Nullstelle des Notchfilters kurz bevor der Frequenzpunkt die Nullstelle auf dem Einheitskreis durchläuft und kurz danach dargestellt. Die Phase des Pol-Linearfaktors läuft dabei kontinuierlich über $270^\circ=3\pi/2$, die des Linearfaktors mit der Nullstelle macht

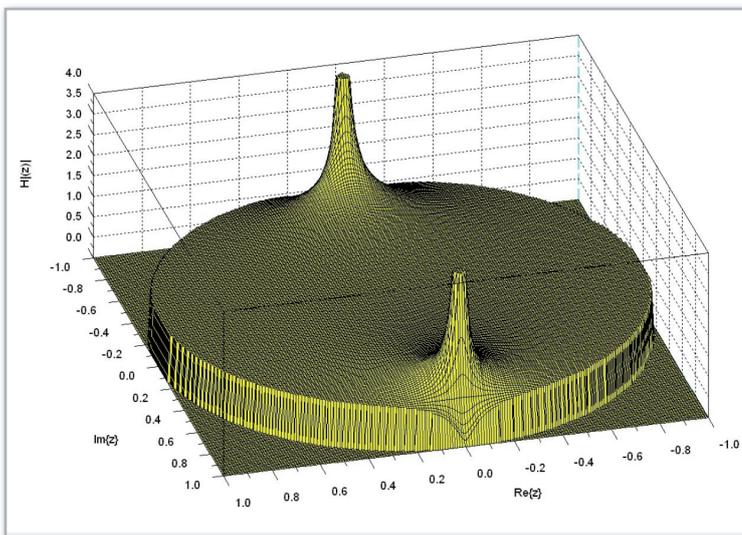


Bild 106: Dreidimensionale Betragsdarstellung des besprochenen Notchfilters 2. Grades.

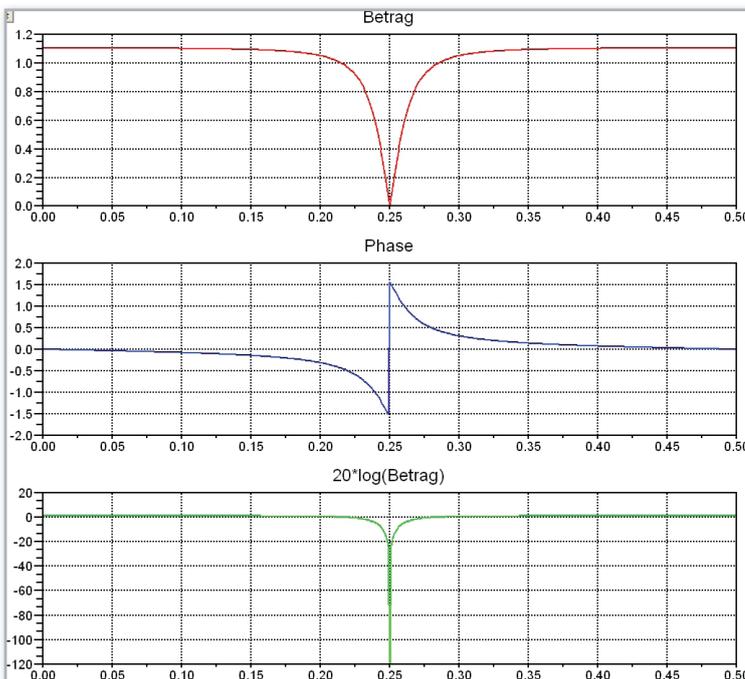
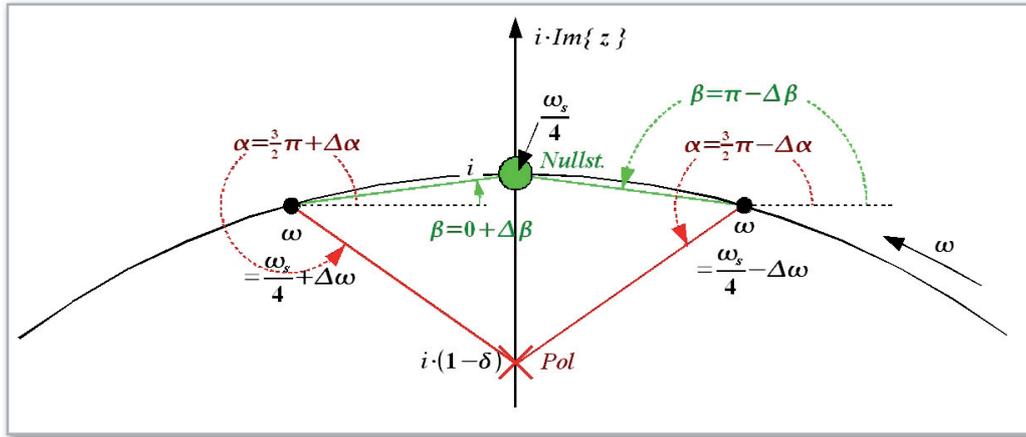


Bild 107: Betrag, Phase und Verstärkung des besprochenen Notchfilters 2. Grades.

Bild 108: Zur Erklärung des Phasensprungs in einer auf dem Einheitskreis liegenden Nullstelle.



dagegen einen Sprung um $180^\circ = \pi$, wie es der Phasengang in der mittleren Kurve von Abbildung 107 zeigt. Aus Abbildung 105 können wir sofort die Differenzgleichung des Notchfilters ablesen (Gleichung 152). Aus ihr lässt sich die Impulsantwort des Notchfilters, also die Reaktion auf einen Einheitsimpuls zum Zeitpunkt $n=0$ ermitteln. Weil zwischen den beiden Verzögerungsgliedern keine Signaleinspeisung stattfindet, ist jeder zweite Wert der Ausgangsfolge y gleich Null. Die Ausgangsfolgenwerte für $n=2, 4, 6, \dots$ ergeben sich durch den zwei Abtastintervalle zuvor anliegenden, mit $a_2=0,81$ multiplizierten Ausgangswert. Zu diesen Zeitpunkten spielt das Eingangssignal $x(n)$ keine Rolle mehr, weil es als Einheitsimpulsfolge definitionsgemäß nur bei $n=0$ den Wert Eins hat und überall sonst Null ist. Die Tabelle in Abbildung 109 zeigt die ersten 21 Werte der Impulsantwort y , Abbildung 110 stellt die Folge der ersten 100 Werte grafisch dar. Man erkennt, dass die Impulsantwortfolge nie vollständig auf Null abklingt, eine unendlich hohe Rechengenauigkeit

-2	0,0000000000
-1	0,0000000000
0	1,0000000000
1	0,0000000000
2	0,1900000000
3	0,0000000000
4	-0,1539000000
5	0,0000000000
6	0,1246590000
7	0,0000000000
8	-0,1009737900
9	0,0000000000
10	0,0817887699
11	0,0000000000
12	-0,0662489036
13	0,0000000000
14	0,0536616119
15	0,0000000000
16	-0,0434659057
17	0,0000000000
18	0,0352073836
19	0,0000000000
20	-0,0285179807

Bild 109: Tabelle der Werte der Impulsantwortfolge des besprochenen Notchfilters.

vorausgesetzt. Das Filter hat also eine unendlich lang andauernde Impulsantwort (IIR: Infinite Impulse Response). In der Praxis ist dies aber durch die endlichen Registerlängen der Multiplizierer, Addierer und Verzögerungsspeicher nicht gegeben, sodass wegen des daraus resultierenden Abschneideeffekts das Ausgangssignal doch irgendwann den Wert Null annimmt. In ungünstigen Fällen kann es aber auch zu einem stabilen, periodischen Oszillieren zwischen kleinen

$$H(z) = \frac{(z-i)(z+i)}{(z-(1-\delta)i)(z-(-1+\delta)i)} \quad (151)$$

$$= \frac{(z-i)(z+i)}{(z-0,9i)(z+0,9i)} = \frac{z^2+1}{z^2+0,81} = \frac{z^{-2}+1}{0,81z^{-2}+1} \quad \text{für } \delta=0,1$$

Systemübertragungsfunktion eines Notchfilters

$$y(n) = x(n) + x(n-2) - 0,81 y(n-2) \quad (152)$$

Differenzgleichung des Notchfilters

$$H(z) = \frac{0,14006 z^{-2} + 0,14006}{0,56076 z^{-2} - 1,24046 z^{-1} + 1} \quad (153)$$

$$= 0,14006 \frac{(z+i)(z-i)}{(z-[0,62023+i \cdot 0,41962])(z-[0,62023-i \cdot 0,41962])}$$

Der Nenner einer biquadratischen Übertragungsfunktion lässt sich darstellen als:

$$z^2 + a_1 z + a_2 = (z - z_\infty)(z - \bar{z}_\infty) = (z - r_\infty e^{i\alpha_\infty})(z - r_\infty e^{-i\alpha_\infty})$$

$$= z^2 - \underbrace{2r_\infty \cos(\alpha_\infty)}_{-2 \operatorname{Re}(z_\infty) = a_1} z + \underbrace{r_\infty^2}_{\operatorname{Re}^2(z_\infty) + \operatorname{Im}^2(z_\infty) = a_2}$$

Die durch Koeffizientenvergleich gewonnenen Beziehungen liefern:

$$\operatorname{Re}(z_\infty) = -\frac{a_1}{2} \quad \text{und} \quad \operatorname{Im}(z_\infty) = \pm \sqrt{\left(a_2 - \frac{a_1^2}{4}\right)}$$

Realteil und Imaginärteile der konjugiert komplexen Biquadpole

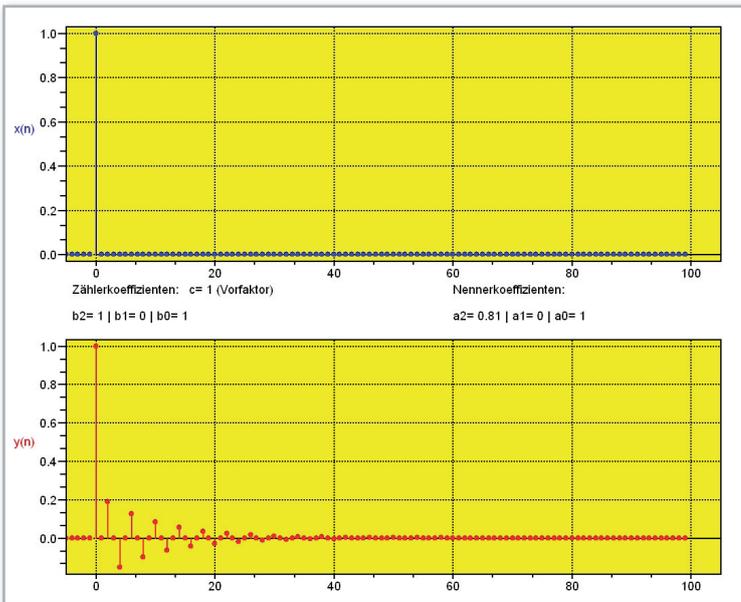
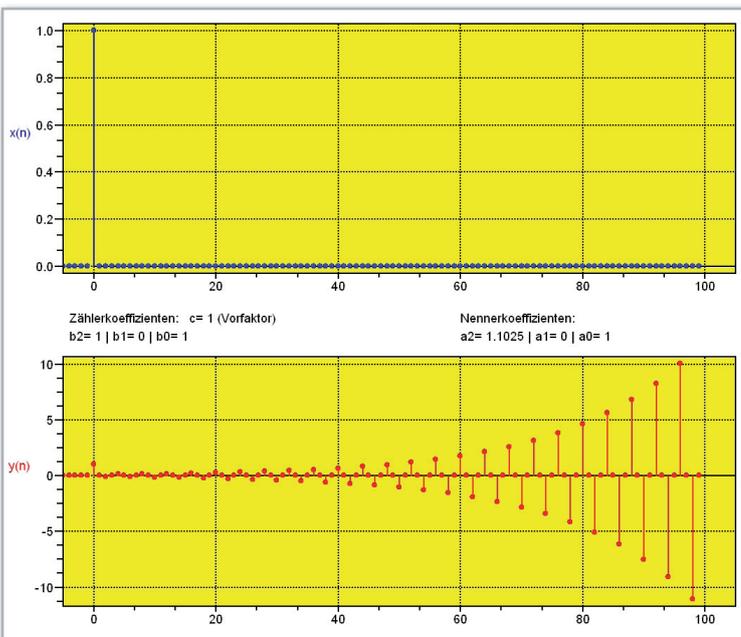


Bild 110: Die gegen Null konvergierende Impulsantwort des Notchfilters beweist seine Stabilität.



Grenzwerten kommen. Diese so genannten Grenzyklen (limit cycles) verursachen eine Art Grundrauschen, auch wenn kein Eingangssignal anliegt.

Wir hatten früher schon mehrfach festgehalten, dass die Pole einer stabilen Übertragungsfunktion innerhalb des Einheitskreises liegen müssen. Das war mit den Polen bei $\pm 0,9$ der Fall, wie die abklingende (gedämpfte) Impulsantwort in Abbildung 110 beweist. Schieben wir die Pole dagegen aus dem Einheitskreis z. B. zu $\pm 1,05i$ hinaus, klingt die Impulsantwort an (d. h. sie schaukelt sich auf), die Übertragungsfunktion ist somit instabil (Abbildung 111).

Beispiel Tiefpass 2. Grades

Wir wählen einen Tiefpassentwurf nach Cauer (Wilhelm Cauer, Begründer der linearen Schaltungssynthese, 1900 – 1945), der durch je zwei Pole und Nullstellen gekennzeichnet ist. Wegen der zur Approximation eines vorgegebenen Dämpfungsverlaufes verwendeten so genannten rationalen elliptischen

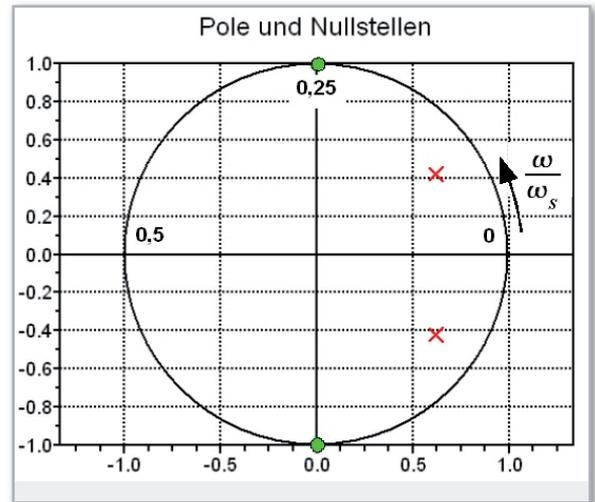


Bild 112: Pol-Nullstellen-Diagramm eines Cauer-Tiefpasses 2. Grades.

Bild 111: Liegen die Pole des besprochenen Notchfilters 2. Grades außerhalb des Einheitskreises schaukelt sich die Impulsantwort auf; das Filter ist instabil.

Funktionen, spricht man auch von einem elliptischen Tiefpass 2. Ordnung. Die Methoden des Entwurfs digitaler Filter werden in einer späteren Folge dieser Reihe besprochen.

Die Systemübertragungsfunktion beschreibt Gleichung (153).

Die Nullstellen und Pole können wir direkt Gleichung (153) entnehmen und in den PN-Plan gemäß Abbildung 112 umsetzen.

Eine qualitative Vorstellung vom Betrag des Frequenzgangs lässt sich ohne Rechnung entwickeln, indem man sich gedanklich auf dem Einheitskreis der z -Ebene von $z=1$ ausgehend über $z=i$ bis zu $z=-1$ bewegt. Das entspricht dem Durchlaufen des Frequenzbereichs $0 < \omega < 0,5 \omega_s$. Wir beginnen bei $\omega=0$ und steigen auf der Flanke des Pols aufwärts. Nachdem wir ihn passiert haben, geht es wieder abwärts bis auf die Höhe Null am Ort der Nullstelle bei $\omega=\omega_s/4$. Dahinter steigt der Betrag bis zur halben Abtastfrequenz $\omega_s/2$ nur noch mäßig

ßig an. Die dreidimensionale Darstellung der Übertragungsfunktion innerhalb des Einheitskreises in Abbildung 113 bestätigt diese Überlegungen. Abbildung 114 zeigt die exakten Verläufe des Betrags, der Phase und des mit 20 multiplizierten logarithmierten Betrags (Verstärkung in dB) über der Frequenz. Wieder sehen wir den Phasensprung in der Nullstelle, dessen Ursache in Abbildung 108 bereits beschrieben wurde. Zum Abschluss sei noch die Impulsantwort in Abbildung 115 gezeigt, die - wie aus der Lage der Pole zu erwarten - stabil ist, also abklingt.

Koeffizientenquantisierung

Wenn man die Übertragungsfunktion eines Biquads nach Gleichung 130 mit beliebig genauen Koeffizienten realisiert, lassen sich die Nullstellen und Pole ebenfalls beliebig genau in der z-Ebene plazieren. Eine endliche Wortlänge der Koeffizienten des Nenners bewirkt deren Quantisierung. Dadurch ist nur eine begrenzte Anzahl von Pollagen möglich. Komplexe Pole treten dabei stets als konjugiert komplexe Paare auf, die im Einheitskreis der z-Ebene liegen müssen, damit das System stabil ist. Um diese stabilen Pole zu ermitteln, drücken wir deren Real- und Imaginärteil durch die Koeffizienten a_1 und a_2 des Nennerpolynoms aus (Gleichung (154)). Lässt man nun a_1 alle Werte im Quantisierungsraster zwischen -2 und +2 annehmen ($-2 < a_1 < 2$) sowie a_2 alle Werte im Quantisierungsraster zwischen -1 und +1 ($-1 < a_2 < 1$) (vergleiche dazu Abbildung 87) ergeben sich alle möglichen Lagen der Pole im Einheitskreis der z-Ebene, also die stabilen Pole, die mit der gegebenen Koeffizientengenauigkeit realisierbar sind (Abbildung 116). Dabei zeigt sich eine geringe Dichte der möglichen Pollagen um $z=1$ und eine hohe um $z=i$. Tiefpässe mit einer niedrigen Grenzfrequenz werden also durch die kanonischen Realisierungen der Biquadschaltungen nur mit unverhältnismäßig genauen Koeffizienten, d. h. mit großen Wortlängen, eine annehmbare Genauigkeit erreichen.

In Teil 12 der Reihe wird es um Übertragungsfunktionen höherer Ordnung gehen. Es wird gezeigt, dass ihre Realisierung als Kettenschaltung von Biquad-Abschnitten (SOS cells) in vieler Hinsicht vorteilhaft ist.

ELV

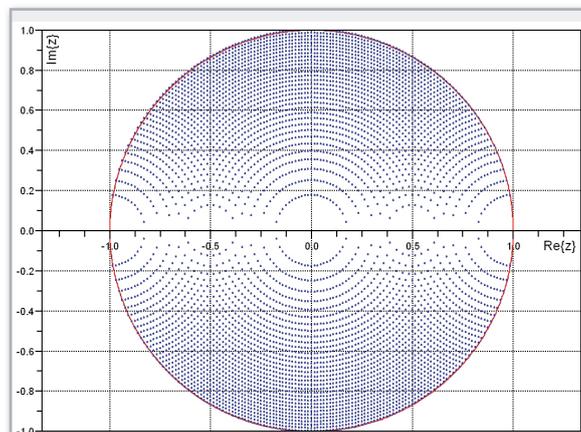


Bild 116: Nennerkoeffizienten einer Biquad-Systemfunktion mit einer Auflösung von $q=2-5$ erlauben nur die dargestellten Pollagen. Für eine feinere Justierung der Pole ist die Wortlänge der Koeffizienten zu erhöhen.

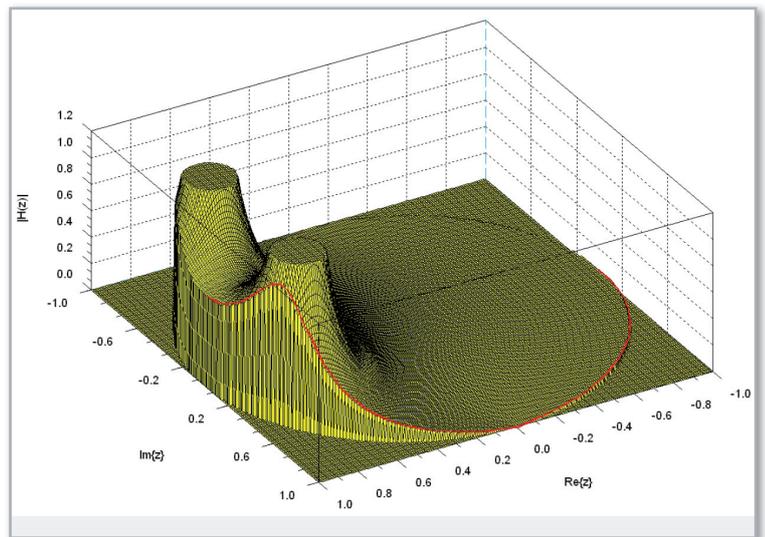


Bild 113: Dreidimensionale Darstellung des Betrags der Übertragungsfunktion eines Cauer-Tiefpasses 2. Ordnung.

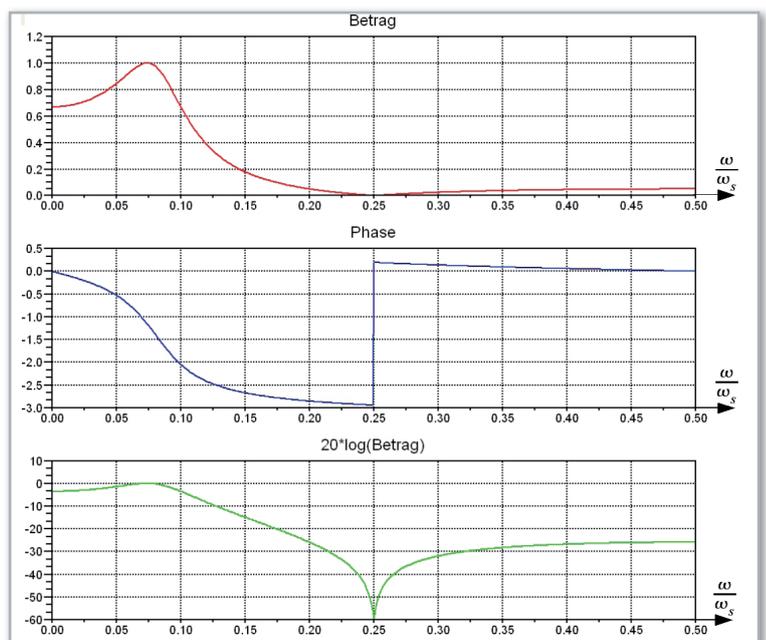


Bild 114: Betrag, Phase und Verstärkung des besprochenen Cauer-Tiefpasses 2. Grades.

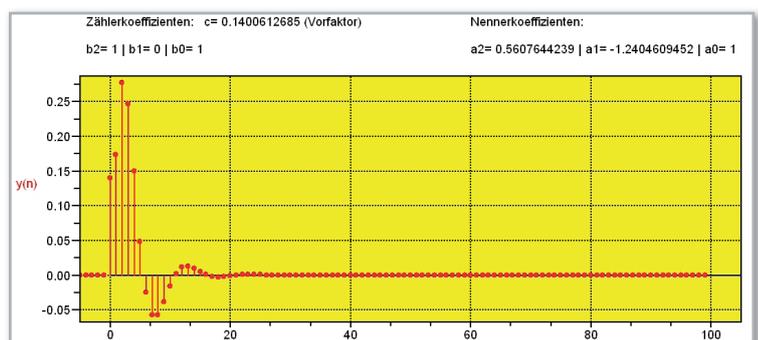
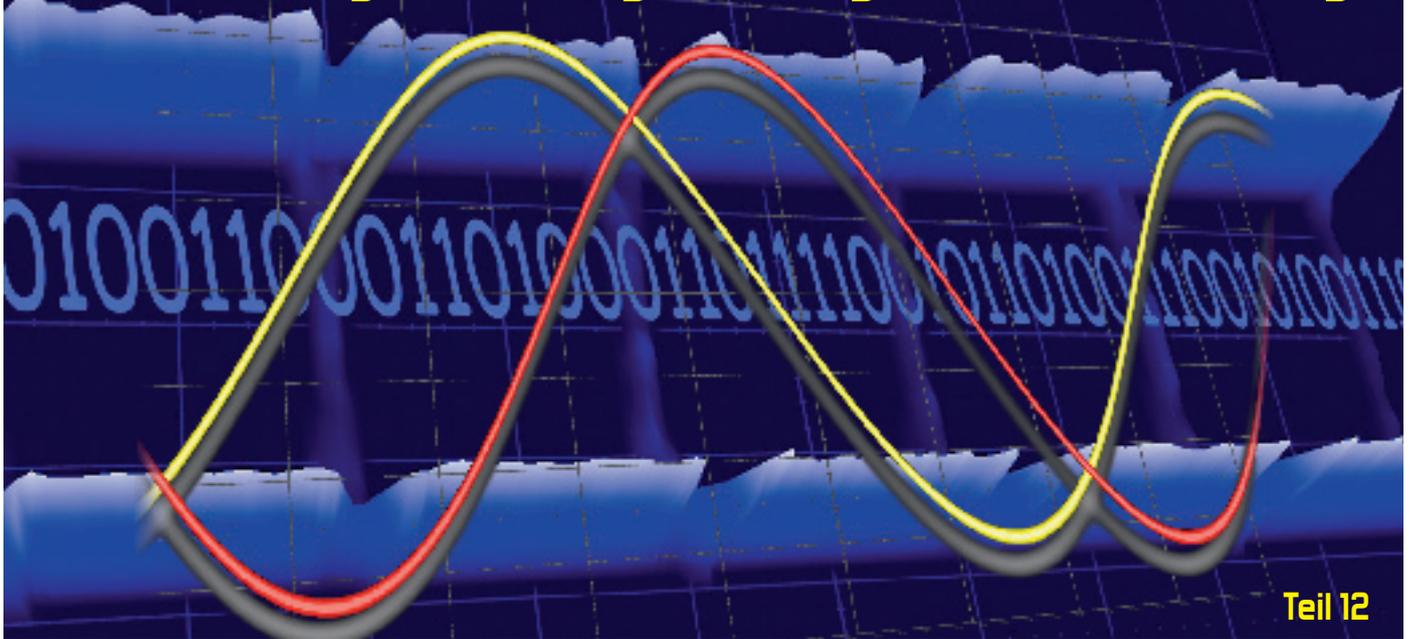


Bild 115: Die Impulsantwort des besprochenen Notchfilters 2. Grades.

Einführung in die digitale Signalverarbeitung



Teil 12

In diesem Teil unserer Folge beschäftigen wir uns mit der Partialbruchzerlegung einer Übertragungsfunktion höheren Grades in Teilübertragungsfunktionen niedrigeren Grades. Ziel ist es, Biquadschaltungen zu finden, die vom gleichen Eingangssignal gespeist, deren Ausgangssignale aber zum gewünschten Gesamtübertragungsverhalten aufsummiert werden.

Übertragungsfunktionen höheren Grades

In den Teilen 10 und 11 dieser Folge haben wir uns auf die Analyse von gebrochen rationalen Übertragungsfunktionen 2. Grades konzentriert. Damit sind Übertragungsfunktionen von Abtastsystemen gemeint, die sich als Quotient zweier quadratischer Polynome in z darstellen lassen. Sie sind die Bausteine zur Realisierung höhergradiger Übertragungsfunktionen. Diese kann man, wie früher gezeigt, in Teilübertragungsfunktionen 2. Grades (Biquad oder SOS) zerlegen, die so beschaffen sein müssen, dass ihre Reihenschaltung (Gleichung 128) oder ihre Parallelschaltung (Gleichung 129) zur gewünschten Gesamtübertragungsfunktion führt. Teilübertragungsfunktionen 1. und 0. Grades (eine Konstante) sind natürlich aus Biquads durch Nullsetzen der entsprechenden Koeffizienten zu gewinnen.

Parallelschaltung von Biquads

Die Berechnung von biquadratischen Übertragungsfunktionen derart, dass ihre Parallelschaltung zum gewünschten Gesamtübertragungsverhalten führt, beruht auf der Zerlegung in Partialbrüche.

Dabei entstehen im Fall einfacher Pole $z = \infty$ der zu zerlegenden Übertragungsfunktion höheren Grades Summanden in Bruchform, deren Nenner aus je einem Linearfaktor $z - z_{\infty}$ des Nenners der Ausgangsfunktion und deren Zähler aus einer Konstanten r_v (man nennt r_v auch das v -te Residuum) bestehen (Gleichung 155). Bringt man diese Summanden auf einen Hauptnenner, kann man die Residuen durch Vergleich mit den Koeffizienten des Zählers der Ausgangsfunktion bestimmen.

Der vollständige Formalismus der Partialbruchzerlegung, der auch mehrfache konjugiert komplexe und reelle Pole und den Fall des höheren Zählergrades einbezieht, soll hier nicht beschrieben werden. Er ist in jedem Lehrbuch der Mathematik oder Systemtheorie nachzulesen (z. B. Oppenheim, Schaffer: „Zeitdiskrete Signalverarbeitung“, 2. Auflage, Oldenbourg 1995, Seite 200 ff, oder Kiencke, Jäkel: „Signale und Systeme“, 3. Auflage, Oldenbourg 2005, S. 294 ff., oder Bronstein-Semendjajew: „Taschenbuch der Mathematik“, 6. Auflage, Verlag Harri Deutsch, 1971, Seite 287 ff). Stattdessen sorgen zwei Beispiele für einen ersten Eindruck (Gleichung 156, Gleichung 157).

Das bereits mehrfach erwähnte kostenlose Programm Scilab stellt eine Funktion (pfss(H)) zur Partialbruchzerlegung von rationalen Übertragungsfunktionen zur Verfügung, mit der auch

(155)

Partialbruchzerlegung für $m < n$ und einfache Nullstellen:

$$H(z) = \frac{b_0 z^m + b_1 z^{m-1} + b_2 z^{m-2} + \dots + b_{m-1} z + b_m}{a_0 z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_{n-1} z + a_n} \quad m < n$$

$$= \frac{b_0 z^m + b_1 z^{m-1} + b_2 z^{m-2} + \dots + b_{m-1} z + b_m}{(z - z_{\infty_1}) \cdot (z - z_{\infty_2}) \cdot \dots \cdot (z - z_{\infty_{n-1}}) \cdot (z - z_{\infty_n})}$$

$$= \frac{r_1}{z - z_{\infty_1}} + \frac{r_2}{z - z_{\infty_2}} + \dots + \frac{r_{n-1}}{z - z_{\infty_{n-1}}} + \frac{r_n}{z - z_{\infty_n}} \quad \text{Partialbruchzerlegung}$$

Beispiel 1 zur Partialbruchzerlegung:

(156)

$$H(z) = \frac{1}{z^3 - 3z^2 + 2z} = \frac{1}{z(z-1)(z-2)} = \frac{a_{11}}{z} + \frac{a_{21}}{z-1} + \frac{a_{31}}{z-2}$$

Partialbruchansatz
für einfache Pole

$$= \frac{(a_{11} + a_{21} + a_{31})z^2 + (-3a_{11} - 2a_{21} - a_{31})z + 2a_{11}}{z^3 - 3z^2 + 2z}$$

Koeffizientenvergleich ergibt:

$$\begin{aligned} a_{11} + a_{21} + a_{31} &= 0 \\ -3a_{11} - 2a_{21} - a_{31} &= 0 \\ 2a_{11} &= 1 \end{aligned}$$

Das ist ein lineares Gleichungssystem mit 3 Gleichungen für die 3 Unbekannten a_{11} , a_{21} , a_{31} . Seine Auflösung ergibt:

$$\begin{aligned} a_{11} &= 0,5 \\ a_{21} &= -1 \\ a_{31} &= 0,5 \end{aligned}$$

Damit ist die Partialbruchzerlegung vollzogen:

$$H(z) = \frac{1}{z^3 - 3z^2 + 2z} = \frac{0,5}{z} + \frac{-1}{z-1} + \frac{0,5}{z-2}$$

Beispiel 2 zur Partialbruchzerlegung:

(157)

$$H(z) = \frac{3z^2 + 2z - 10}{z^3 - 5z^2 + 8z - 4} = \frac{3z^2 + 2z - 10}{(z-1)(z-2)^2} = \frac{a_{11}}{z-1} + \frac{b_{11}z + c_{11}}{(z-2)^2}$$

Partialbruchansatz für
einen einfachen und
einen doppelten Pol

$$= \frac{(a_{11} + b_{11})z^2 + (-4a_{11} - b_{11} + c_{11})z + (4a_{11} - c_{11})}{(z-1)(z-2)^2}$$

Koeffizientenvergleich ergibt:

$$\begin{aligned} a_{11} + b_{11} &= 3 \\ -4a_{11} - b_{11} + c_{11} &= 2 \\ 4a_{11} - c_{11} &= -10 \end{aligned}$$

Das ist ein lineares Gleichungssystem mit 3 Gleichungen für die 3 Unbekannten a_{11} , b_{11} , c_{11} . Seine Auflösung ergibt:

$$\begin{aligned} a_{11} &= -5 \\ b_{11} &= 8 \\ c_{11} &= -10 \end{aligned}$$

Damit ist die Partialbruchzerlegung vollzogen:

$$H(z) = \frac{3z^2 + 2z - 10}{z^3 - 5z^2 + 8z - 4} = \frac{-5}{z-1} + \frac{8z - 10}{(z-2)^2} = \frac{-5}{z-1} + \frac{8z}{(z-2)^2} + \frac{-10}{(z-2)^2}$$

höhergradige Funktionen leicht in Partialbrüche aufzuspalten sind. Mit 5 Zeilen Programmcode lässt sich das Ergebnis des eben trotz einfach gewählter Zahlen relativ mühsam von Hand berechneten Beispiels auf dem Bildschirm ausgeben (Abbildung 117).

Wir wollen die Partialbruchzerlegung an einem Beispiel vertiefen. Dazu „basteln“ wir uns eine Übertragungsfunktion TF vom Grad 6 durch Vorgabe ihrer Pole und Nullstellen. Die zwei konjugiert komplexen Nullstellen sollen auf dem Einheitskreis der z-Ebene liegen und einen Imaginärteil von $\pm 0,75i$ bzw. $\pm 0,9i$ besitzen. Die Pole setzen wir auf $\pm 0,2i$, $\pm 0,4i$ und $\pm 0,6i$ und legen ihre Realteile so fest, dass sich ein ansprechender Tiefpass-Frequenzgang ergibt. Die Zahlenwerte von TF zeigt Gleichung (158). Damit die Gleichungen nicht zu breit werden, sind die Zahlenwerte auf 4 Nachkommastellen gerundet.

Nun wird TF in drei Partialübertragungsfunktionen 2. Grades H_1 , H_2 und H_3 zerlegt. Wenn man die Partialübertragungsfunktionen addiert und auf den Hauptnenner bringt, ergibt sich natürlich wieder TF. Die Ergebnisse der Rechnung fasst Abbildung 118 zusammen und Abbildung 119 veranschaulicht sie. Die Pole der Teilübertragungsfunktionen sind – dem Partialbruchansatz entsprechend – identisch mit denen der Gesamtübertragungsfunktion. Für die Nullstellen gilt das selbstverständlich nicht.

Bei einer Fahrt auf dem Einheitskreis von $\omega=0$ ($z=1$) über $\omega=\omega_s/4$ ($z=i$) bis zu $\omega=\omega_s/2$ ($z=-1$) können wir den Betragsfrequenzgang abschätzen. Zunächst halten die beiden nahe an der reellen z-Achse liegenden Pole den Frequenzgang hoch, bis der Einfluss der Nullstellen greift. In der logarithmierten Darstellung in Abbildung 119 unten rechts äußern sie sich in den beiden Dämpfungspolen bei $\omega/\omega_s=0,135$ bzw. $\omega/\omega_s=0,178$.

Noch plastischer wird das Dämpfungsverhalten durch das Betrachten des dreidimensionalen Betragsgebirges über der

z-Ebene in Abbildung 120. Oben links ist der Betrag unlogarithmiert dargestellt, was zu kaum erkennbaren Nullstellen führt. Erst in der extremen Ausschnittsvergrößerung oben rechts tauchen die Nullstellen auf. Deshalb ist die logarithmierte Darstellung sinnvoller. Unten links wird sie über dem ersten Quadranten der z-Ebene ($0 < \text{Re}\{z\} < 1$, $0 < \text{Im}\{z\} < 1$) gezeigt. Jetzt sind die Nullstellen klar zu erkennen. Unten rechts wird nun noch der Teil des logarithmierten Betragsgebirges ausgestanzt, der über dem Einheitskreis der z-Ebene liegt. Es ergibt sich als Stanzfläche der Amplitudengang in dB zwischen $0 < \omega < \omega_s/4$.

Nachteilig bei der Partialbruchzerlegung ist, dass die Null-

```

SciPad 6.129.BP2 - D:/Eigene Dateien ab Februar 2003/DOC/fachartikel/elw/Einführung in
File Edit Search Execute Debug Scheme Options Windows Help

1 s=poly(0,'s');
2 H=(3*s^2+2*s-10)/(s^3-5*s^2+8*s-4)
3 PB=pfss(H);
4 H1=PB(1)
5 H2=PB(2)
6

scilab-4.1.2 (0)
File Edit Preferences Control Editor Applications ?
--> H =
      -10 + 2s + 3s^2
      -----
      -4 + 8s - 5s^2 + s^3
H1 =
      -10 + 8s
      -----
      4 - 4s + s^2
H2 =
      -5
      -----
      -1 + s
-->

```

Bild 117: In Scilab steht die Funktion pfss zur Partialbruchzerlegung gebrochener rationaler Funktionen zur Verfügung.

Bild 118: Zusammenfassung der Zahlenwerte zum Partialbruch-Beispiel

$H1(z) = \frac{3.4983420514196 + 8.7702016410949z}{0.428129 - 1.246z + z^2}$	$\text{Zeros}(H1(z)) = -0.3988895802609$ $\text{Poles}(H1(z)) = +0.623 \pm 0.2i$
$H2(z) = \frac{-1.7050601772552 - 6.9982269010038z}{0.6784 - 1.44z + z^2}$	$\text{Zeros}(H2(z)) = -0.2436417397399$ $\text{Poles}(H2(z)) = +0.72 \pm 0.4i$
$H3(z) = \frac{0.0753091163654 - 1.7719747400911z}{0.595225 - 0.97z + z^2}$	$\text{Zeros}(H3(z)) = +0.0425001071751$ $\text{Poles}(H3(z)) = +0.485 \pm 0.6i$
$H_{ges}(z) = H1(z) + H2(z) + H3(z) = TF(z) = \frac{1 - 2.1946554442404z + \dots}{0.1728787642026 - 1.151824670628z + \dots}$ $\dots + 3.1532562594671z^2 - 2.1946554442404z^3 + z^4$ $\dots + 3.434991336825z^2 - 5.87431244z^3 + 6.101414z^4 - 3.656z^5 + z^6$	
$\text{Zeros}(H_{ges}(z)) = \begin{cases} +0.4358898943541 \pm 0.9i \\ +0.6614378277661 \pm 0.75i \end{cases}$ $\text{Poles}(H_{ges}(z)) = \begin{cases} +0.623 \pm 0.2i \\ +0.485 \pm 0.6i \\ +0.720 \pm 0.4i \end{cases}$	

stellen der Gesamtübertragungsfunktion von den Polen der Teilübertragungsfunktionen beeinflusst werden. Realisiert man diese durch kanonische Biquadschaltungen mit einem Polgitter nach Abbildung 116, führt die geringe Poldichte bei

$z=1$ dazu, dass die Pole nur mit geringerer Genauigkeit positioniert werden können. Die Fehler in der Pollage müssen sich dann in entsprechenden Verschiebungen der Nullstellen auswirken.

Wir überzeugen uns von dieser Erkenntnis, indem wir die Pole $z_{\infty 1,2}$ gemäß Gleichung (155) aus ihrer Solllage 0,623 um $\pm 0,01$ parallel zur reellen Achse verschieben und die Auswirkungen in Abbildung 121 betrachten. Schon die Veränderung einer einzelnen Pollage um wenige Promille hat starken Einfluss auf die Nullstellen der Gesamtübertragungsfunktion. Wir sehen, dass deren Nullstellen nicht mehr exakt auf dem Einheitskreis liegen, wodurch die Dämpfungspole an Tiefe einbüßen und frequenzmäßig verschoben werden. Man kann sich vorstellen, wie Quantisierungsfehler bei allen Polen der

Nullstellen von TF: $z_{0,1,2} = 0,4359 \pm 0,9i$; $z_{0,3,4} = 0,6614 \pm 0,75i$ (158)
 Pole von TF: $z_{\infty 1,2} = 0,623 \pm 0,2i$; $z_{\infty 3,4} = 0,7200 \pm 0,4i$; $z_{\infty 5,6} = 0,485 \pm 0,6i$

$$TF = \frac{(z-z_{0,1})(z-z_{0,2})(z-z_{0,3})(z-z_{0,4})}{(z-z_{\infty 1})(z-z_{\infty 2})(z-z_{\infty 3})(z-z_{\infty 4})(z-z_{\infty 5})(z-z_{\infty 6})}$$

$$= \frac{z^4 - 2,1947z^3 + 3,1533z^2 - 2,1947z + 1}{z^6 - 3,6560z^5 + 6,1014z^4 - 5,8743z^3 + 3,4350z^2 - 1,1518z + 0,1729}$$

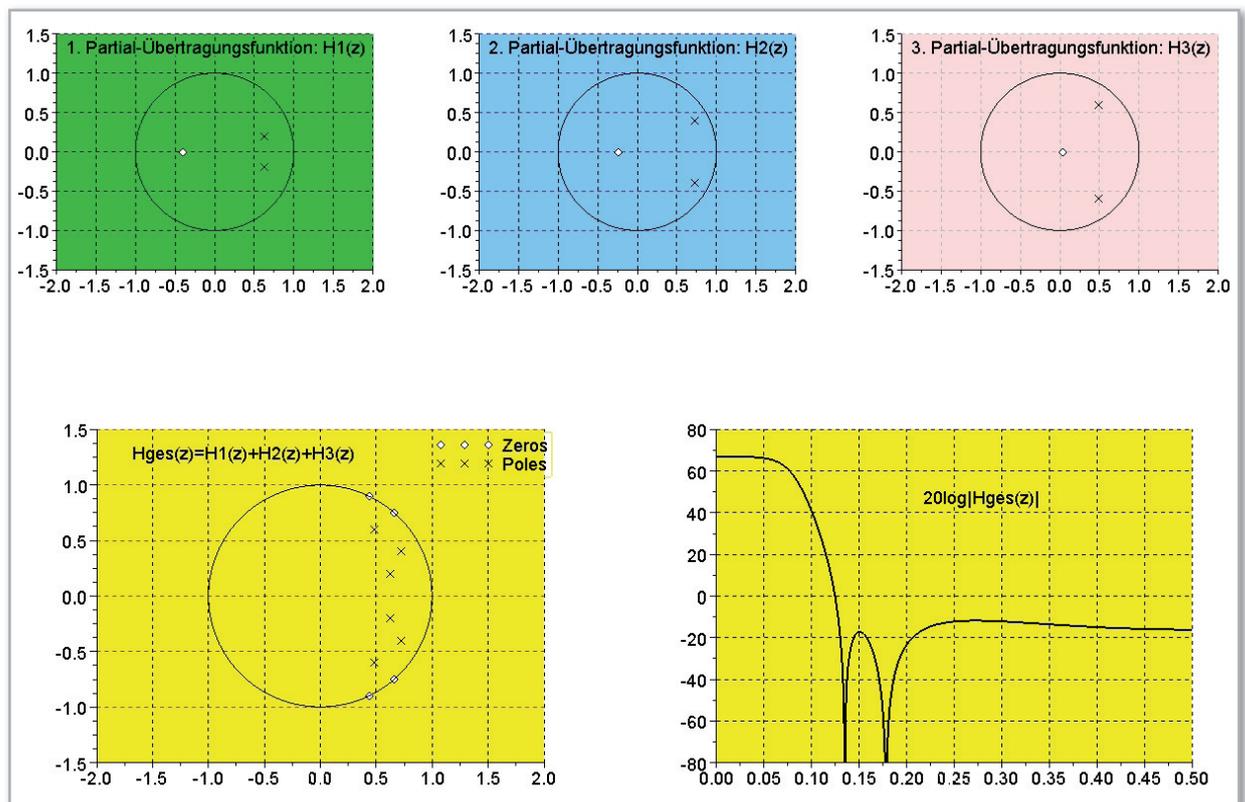


Bild 119: PN-Diagramme der Partialbrüche und der Gesamtübertragungsfunktion sowie deren logarithmierter Betrag (Gain).

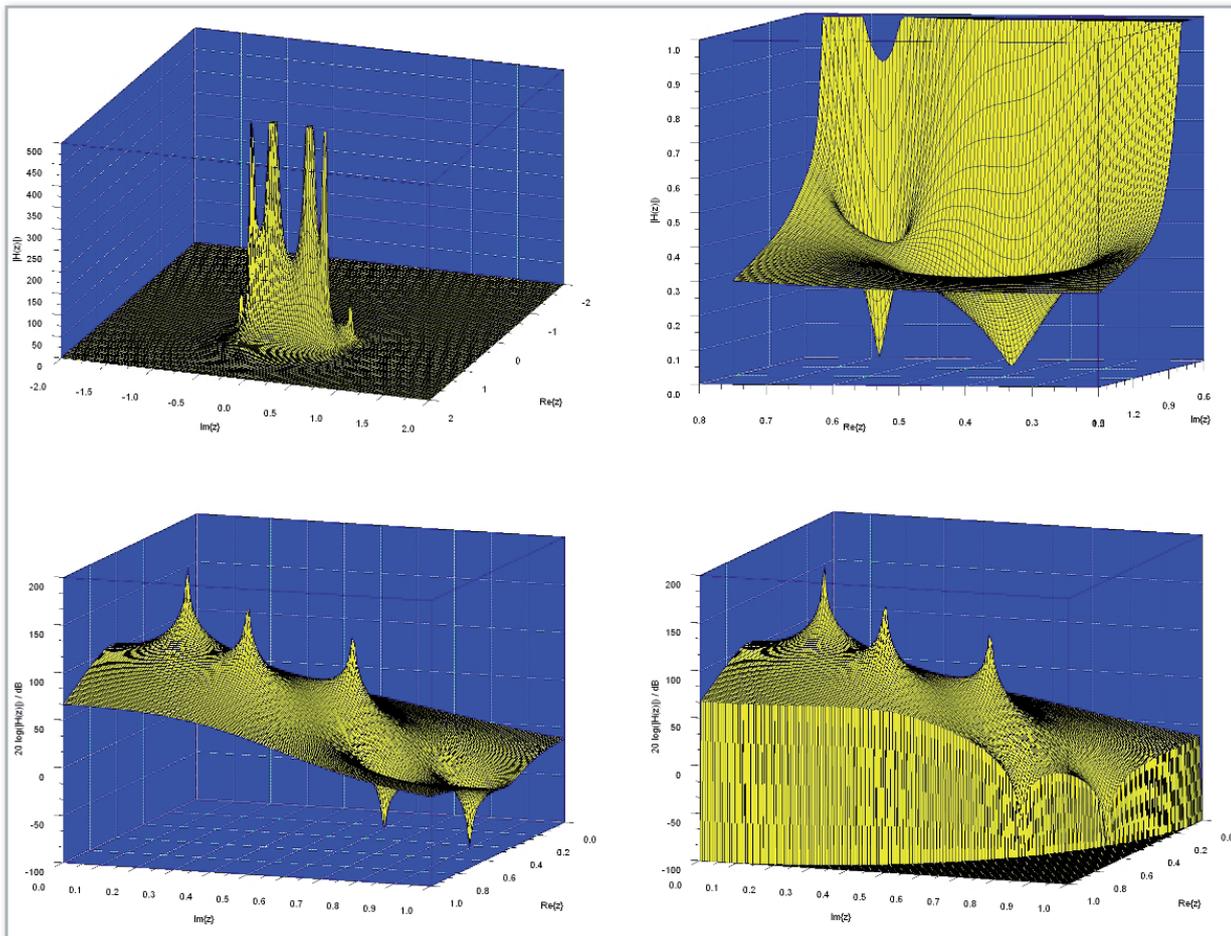


Bild 120: 3D-Darstellung von Polen und Nullstellen im ersten Quadranten der z-Ebene; oben linear unten logarithmisch

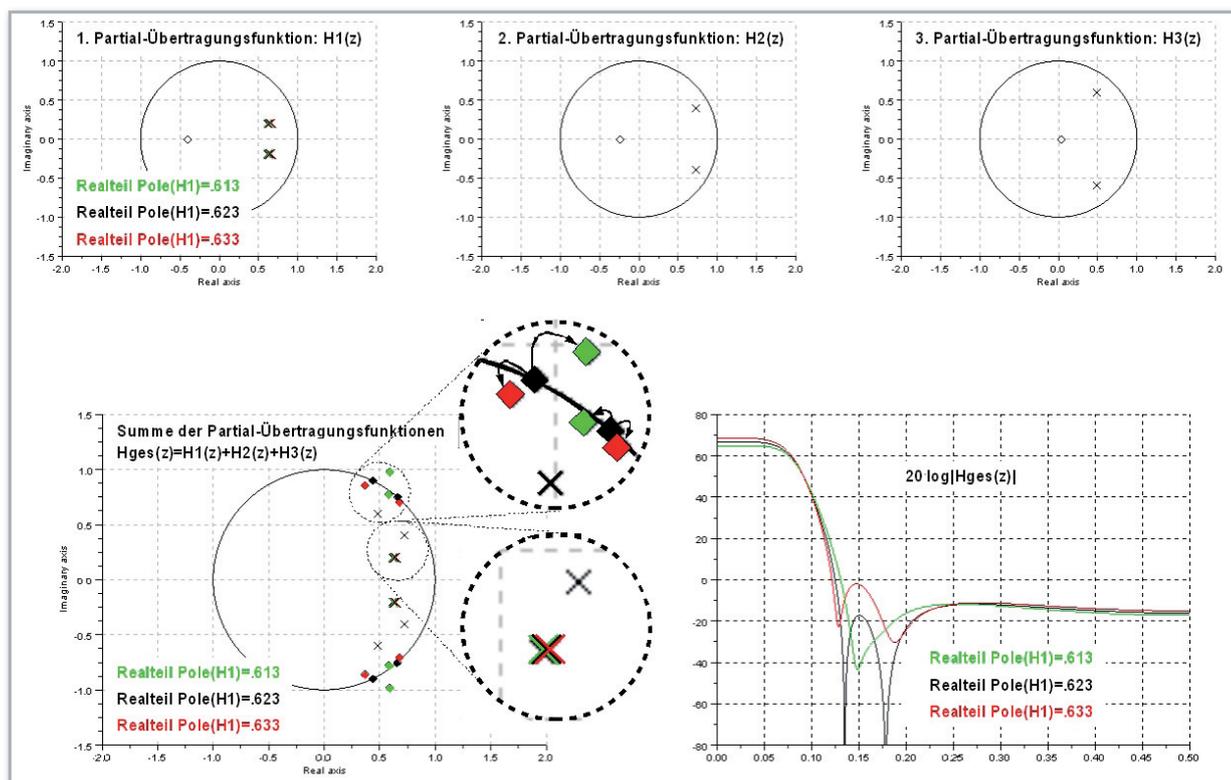


Bild 121: Auswirkungen einer leichten Polverschiebung in einer der Partialübertragungsfunktion auf die Nullstellen der Gesamtübertragungsfunktion

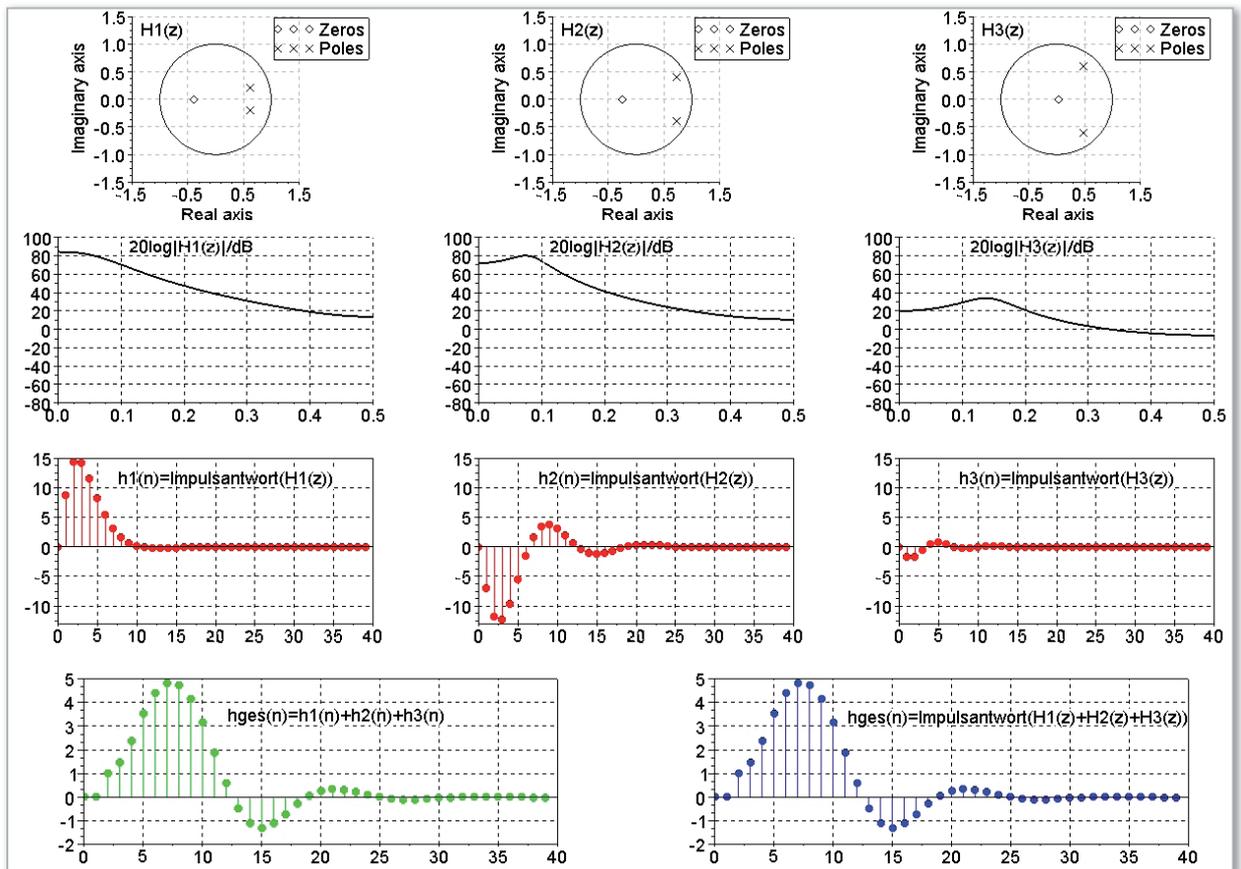


Bild 122: Die zu den Partialübertragungsfunktionen gehörigen Impulsantworten werden zu jedem Abtastzeitpunkt zu einem Summenwert aufaddiert. Die resultierende Impulsantwort entspricht der z-Rücktransformierten der Gesamtübertragungsfunktion.

Teilübertragungsfunktionen zu sehr schwer überschaubaren Auswirkungen auf den Frequenzgang des gesamten Filters führen. Das ist einer der Gründe warum in der Praxis die Kettenschaltung von Biquad-Abschnitten (SOS: Second Order Sections) vor der Parallelschaltung von Teilsystemen, die aus einer Partialbruchentwicklung gewonnen wurden, bevorzugt wird. Allerdings sind die Laufzeitverzögerungen des partiellen Gesamtsystems vom Grad der Übertragungsfunktion unabhängig, wogegen sie bei einer Kettenschaltung mit jedem Kettenglied zunehmen. In dieser Hinsicht sind die parallelen Strukturen, wie sie die Partialbruchzerlegung der Übertragungsfunktion liefert, besser für Echtzeitanwendungen geeignet.

Es ist lehrreich, sich einmal die Frequenzgänge der Partialübertragungsfunktionen zu betrachten. Ihre mit 20 multiplizierten logarithmierten Beträge sind in der zweiten Reihe von Abbildung 122 dargestellt. Offensichtlich ergibt deren Summe nicht den Gesamtbetragsfrequenzgang aus Abbildung 121 unten rechts. Das wird verständlich, wenn man bedenkt, dass die Ausgangsfolgen der Partialstrukturen, wie aus Abbildung 85 ersichtlich ist, zu einer Gesamtfolge Wert für Wert aufsummiert werden. Das zeigt Abbildung 122 am Beispiel der ersten 40 Folgenwerte der roten Impulsantworten $h_1(n)$, $h_2(n)$ und $h_3(n)$ im gleichen Maßstab in Reihe 3, die sich aus der inversen z-Transformation der Partialübertragungsfunktionen $H_1(z)$, $H_2(z)$ und $H_3(z)$ ergeben (wir erinnern uns, dass der Impulsantwort im diskreten Zeitbereich die Übertragungs-

funktion im Frequenzbereich über die z-Transformation zugeordnet ist und umgekehrt). Sie summieren sich zur grünen Gesamtimpulsantwort $h_{ges}(n)$ unten links auf. Eben diese Impulsantwort erhält man auch durch Rücktransformation der Gesamtübertragungsfunktion $H_{ges}(z)$ aus dem z-Bereich in den diskreten Zeitbereich, was zur blauen Kurve unten rechts führt. Die Tatsache, dass offensichtlich die Transformation der Summe $H_1(z) + H_2(z) + H_3(z)$ in den Zeitbereich, gleich der Summe der transformierten Summanden im Zeitbereich $h_1(n) + h_2(n) + h_3(n)$ ist, drückt die Linearitätseigenschaft des Systems aus.

Reihenschaltung von Biquads

Die Berechnung von biquadratischen Übertragungsfunktionen derart, dass ihre Reihenschaltung zum gewünschten Gesamtübertragungsverhalten führt, beruht auf der Faktorisierung in Brüche zweiten Grades (Biquads). Dabei kann man einfach jeweils ein konjugiert komplexes Nullstellenpaar mit einem konjugiert komplexen Polpaar kombinieren. Allerdings besteht eine große Vielfalt der Kombinationsmöglichkeiten bei der Zuordnung von Polen und Nullstellen und in der Reihenfolge der Biquadabschnitte.

Wir wollen uns damit eingehender in Teil 13 dieser Folge beschäftigen. 

Einführung in die digitale Signalverarbeitung

Teil 13

Die Berechnung von biquadratischen Übertragungsfunktionen derart, dass ihre Reihenschaltung zum gewünschten Gesamtübertragungsverhalten führt, beruht auf der Faktorisierung in Brüche zweiten Grades (Biquads). Allerdings besteht eine große Vielfalt der Kombinationsmöglichkeiten bei der Zuordnung von Polen und Nullstellen und in der Reihenfolge der Biquadabschnitte. Wir wollen uns damit eingehender in Teil 13 dieser Folge beschäftigen.

Reihenschaltung von Biquads

In Teil 12 demonstrierten wir die Aufspaltung von Übertragungsfunktionen höheren Grades durch Partialbruchzerlegung in eine Summe von Teilübertragungsfunktionen 1. und 2. Grades. Diese werden wie in Abbildung 85 und Gleichung (129) gezeigt, zu einer Parallelstruktur mit dem gewünschten Gesamtübertragungsverhalten zusammengefügt. Einfacher und gebräuchlicher ist die Reihenschaltung entsprechender Biquads gemäß Abbildung 84 sowie Abbildung 123 und Gleichung (128).

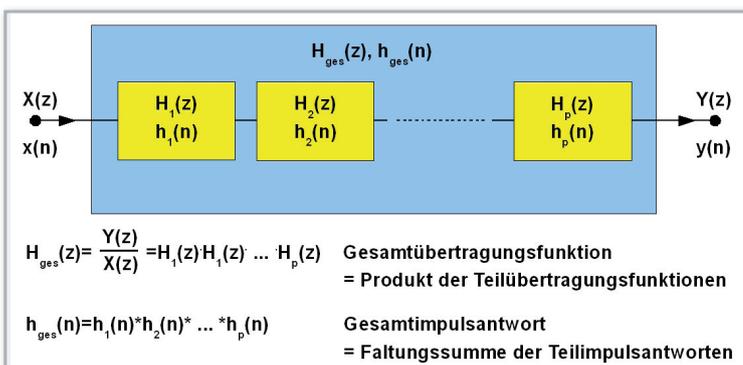


Bild 123: In einer Kaskade von Teilsystemen ergibt sich der Gesamtfrequenzgang durch Multiplikation der Frequenzgänge der Teilabschnitte. Entsprechend werden deren Impulsantworten zur Impulsantwort des Gesamtsystems gefaltet.

Die Übertragungsfunktion eines diskreten LTI-Systems (LTI: Linear Time Invariant) lässt sich bekanntlich als gebrochen rationale Funktion (d. h. als Quotient zweier Polynome in z , Zählergrad \leq Nennergrad) darstellen. Aus der Bestimmung der Nullstellen von Zähler- und Nennerpolynom (=Nullstellen und Pole der Übertragungsfunktion) ergibt sich die Linearfaktorschreibweise. Diesen Sachverhalt stellt Gleichung (159) dar. Jede dieser Polynomnullstellen hat – wenn sie komplex ist – ein konjugiert komplexes Gegenstück. Mit anderen Worten: Jeder Pol und jede Nullstelle der Übertragungsfunktion mit nicht verschwindendem Imaginärteil wird an der reellen Achse der z -Ebene gespiegelt.

Ein Biquad kann abhängig von den Werten der Zähler- und Nennerkoeffizienten eine Vielzahl von Nullstellen- und Polkonfigurationen realisieren. Die häufigsten sind:

- je ein konjugiert komplexes Nullstellen- und Polpaar ($b_{2,v}, b_{1,v}, a_{2,v}, a_{1,v} \neq 0$),
- nur ein konjugiert komplexes Polpaar ($b_{2,v}=b_{1,v}=0, a_{2,v}, a_{1,v} \neq 0$),
- eine reelle Nullstelle und ein konjugiert komplexes Polpaar ($b_{2,v}=0, b_{1,v}, a_{2,v}, a_{1,v} \neq 0$),
- eine reelle Nullstelle und einen reellen Pol ($b_{2,v}=a_{2,v}=0, b_{1,v}, a_{1,v} \neq 0$),
- nur einen reellen Pol ($b_{2,v}=b_{1,v}=a_{2,v}=0, a_{1,v} \neq 0$).

Für die Zuordnung der Nullstellen und Pole zu den Biquad-Übertragungsfunktionen, deren Positionierung in der Reihenschaltung und die Realisierung als 1. bzw. 2. kanonische

Struktur bestehen umso mehr Möglichkeiten, je länger die Biquad-Kette ist. Sind 5 Biquads in Reihe geschaltet, gibt es bereits 57600 mögliche Variationen.

Bei einer unendlich hohen Genauigkeit der Quantisierung, der Rechenarithmetik und Multiplizierer-Koeffizienten ergeben alle diese Varianten das gleiche Gesamtübertragungsverhalten. In der Praxis gilt dies nicht. Hier gibt es zahlreiche Imperfektionen unter dem Oberbegriff Wortlängeneffekte: beschränkte Zahlenbereiche, Multiplikationsfehler durch Abschneiden, Runden, Begrenzen, Überlauf der Ergebnisse usw. All das überlagert sich als rauschartige Störung (Quantisierungsrauschen) dem Nutzsignal und schränkt die Aussteuerbarkeit (Dynamikbereich) des Gesamtsystems ein. Für optimale Ergebnisse in dieser Hinsicht hat Leland B. Jackson (L. B. Jackson, „On the Interaction of Roundoff Noise and Dynamic Range in Digital Filters,“ Bell Sys. Tech. J., vol. 49 (1970 Feb.)) Faustregeln aufgestellt, die eine einfache Paarbildung von Polen und Nullstellen und die Reihenfolge der Biquads mit fast immer guten Resultaten ermöglichen und das aufwändige Suchen nach einer optimalen Lösung ersparen.

Die (vereinfachte) Regel besteht aus drei Vorschriften:

1. Der Pol mit dem geringsten Abstand zum Einheitskreis ist mit der ihm am nächsten liegenden Nullstelle zu paaren.
2. Vorschrift 1 ist solange anzuwenden, bis alle Pole und Nullstellen zu Paaren angeordnet sind.
3. Das zuerst gebildete Teilsystem ist am Ende der Kette anzuordnen, das zweite an vorletzter Stelle usw.

Dieses Vorgehen ist durchaus plausibel. Die Regeln zur Paarbildung bewirken, dass der Einfluss eines nahe am Einheitskreis liegenden Pols (starkes Ansteigen des Signalpegels) durch die naheliegendste Nullstelle am stärksten gemildert wird. Damit nimmt die Spitzenverstärkung in diesem Bereich ab, was die Gefahr von Überläufen verringert und zu einer geringeren Anhebung des Quantisierungsrauschens führt. Würde ein Biquad mit hoher Spitzenverstärkung am Eingang der Kette liegen, müsste sein Eingangssignal so stark abgeschwächt (skaliert) werden, dass weder er selbst noch der nachfolgende Biquad übersteuert werden. Das verschlech-

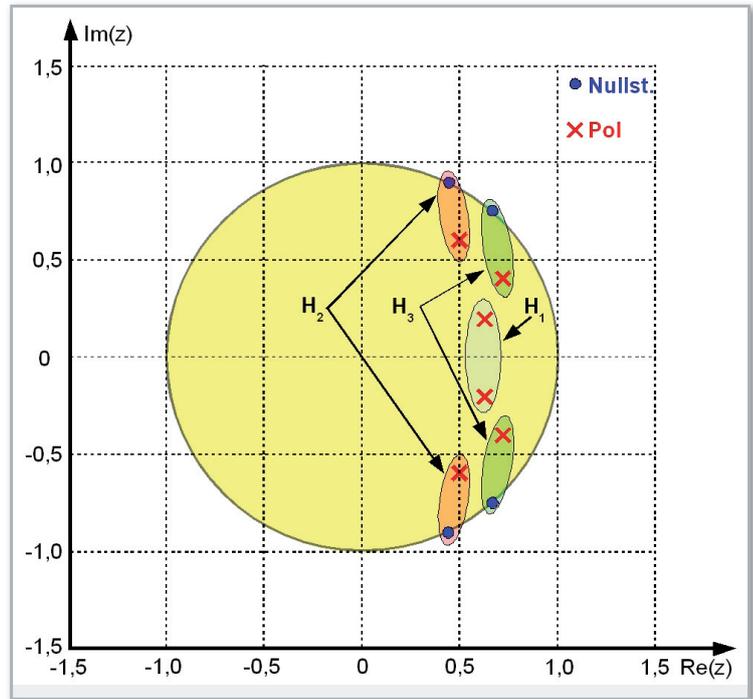


Bild 124: Bei der Zerlegung einer höhergradigen Übertragungsfunktion in praktisch realisierte Biquad-Übertragungsfunktionen müssen Pole und Nullstellen richtig gepaart werden. Jacksons Regel führt in den meisten Fällen zu guten Ergebnissen.

tert das Signal-Rausch-Verhältnis des Systems. Nicht zuletzt führt die Darstellungsgenauigkeit der Koeffizienten durch beschränkte Wortlängen zu Veränderungen der Pollagen und beeinflusst damit die Eigenschaften des Übertragungssystems, insbesondere die den Rauschpegel anhebende Spitzenverstärkung.

Die Beantwortung der Frage der Reihenfolge kann unter anderen Aspekten einschließlich denen der Gesamtausgangsrauschleistung und des Ausgangsrauschspektrums aber auch anders ausfallen, als oben beschrieben.

Abbildung 124 zeigt das Vorgehen bei der PN-Paarbildung am Beispiel von Gleichung (158), die wir bereits bei der Be-

(159)

$$\begin{aligned}
 H(z) &= \frac{Y(z)}{X(z)} = \frac{\sum_{v=0}^N b_v z^{N-v}}{z^N + \sum_{\mu=1}^N a_\mu z^{N-\mu}} = \frac{b_0 z^N + b_1 z^{N-1} + \dots + b_{N-1} z + b_N}{z^N + a_1 z^{N-1} + \dots + a_{N-1} z + a_N} \\
 &= b_0 \cdot \frac{\prod_{v=1}^N (z - z_{0,v})}{\prod_{\mu=1}^N (z - z_{\infty,\mu})} = b_0 \cdot \frac{(z - z_{0,1})(z - z_{0,2}) \dots (z - z_{0,N-1})(z - z_{0,N})}{(z - z_{\infty,1})(z - z_{\infty,2}) \dots (z - z_{\infty,N-1})(z - z_{\infty,N})} \\
 &= b_0 \cdot H_1(z) \cdot H_2(z) \cdot \dots \cdot H_p(z) = b_0 \cdot \prod_{v=1}^p H_v(z) \\
 \text{mit } H_v(z) &= \frac{b_{2,v} z^{-2} + b_{1,v} z^{-1} + 1}{a_{2,v} z^{-2} + a_{1,v} z^{-1} + 1} = \frac{z^2 + b_{1,v} z + b_{2,v}}{z^2 + a_{1,v} z + a_{2,v}} \\
 &\quad \text{Übertragungsfunktion des } v\text{-ten Biquads in } z^{-1} \text{ und } z
 \end{aligned}$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1}{(z - [0,623 + 0,2i])(z - [0,623 - 0,2i])} \quad \Rightarrow H_1(z)$$

$$\frac{(z - [0,4359 + 0,9i])(z - [0,4359 - 0,9i])}{(z - [0,485 + 0,6i])(z - [0,485 - 0,6i])} \quad \Rightarrow H_2(z)$$

$$\frac{(z - [0,6614 + 0,75i])(z - [0,6614 - 0,75i])}{(z - [0,720 + 0,4i])(z - [0,720 - 0,4i])} \quad \Rightarrow H_3(z)$$

$$H_1(z) = \frac{1}{z^2 - 1,246z + 0,428129}$$

$$H_2(z) = \frac{z^2 - 0,8718z + 1}{z^2 - 0,97z + 0,595225}$$

$$H_3(z) = \frac{z^2 - 1,3228z + 1}{z^2 - 1,44z + 0,6784}$$

trachtung der Partialbruchzerlegung zugrunde gelegt hatten. Jetzt werden die Übertragungsfunktionen der Teilsysteme nach Gleichung (160) berechnet. Ihr Produkt ergibt die Gesamtübertragungsfunktion. Das entspricht einer Reihenschaltung (Kaskade) der Teilsysteme.

In Abbildung 125 ist dies alles zusammengefasst. In der oberen Reihe sind die PN-Diagramme der drei Teilübertragungsdigramme $H_1(z)$, $H_2(z)$ und $H_3(z)$ wie sie Abbildung 124 vorgibt, zusammengestellt. Darunter sind zugehörigen mit 20 multiplizierten logarithmierten Betragsfrequenzgänge (Ver-

stärkung, Pegel in dB) abgebildet.

Unten links werden die Pegel von H_1 und H_2 aufaddiert, entsprechend $20 \cdot \log|H_1(z) \cdot H_2(z)| = 20 \cdot \log|H_1(z)| + 20 \cdot \log|H_2(z)|$. In der Mitte der unteren Reihe sehen wir die Summe der drei Verstärkungsfrequenzgänge aus der mittleren Reihe. Wie zu erwarten ist sie mit der aus Abbildung 119 identisch.

Unten rechts ist das zugehörige PN-Diagramm der Gesamtübertragungsfunktion zu sehen. Es entsteht durch Überlagerung der PN-Diagramme der Teilübertragungsfunktionen.

Abbildung 126 zeigt die Überlagerung der mit 20 multiplizierten logarithmierten Frequenzgänge dreidimensional über dem ersten Quadranten der z-Ebene ($0 \leq \omega \leq \omega_s/4$, bzw. normiert $0 \leq \Omega = \omega/\omega_s = f/f_s \leq 0,25$) zwischen -80 dB und +120 dB, wobei alle Werte außerhalb des Einheitskreises zu Null gesetzt werden. Die dabei entstehenden Konturflächen wurden zur besseren Erkennung der Frequenzgänge grün eingefärbt.

Man sieht deutlich, wie die Pole mit abnehmendem Abstand zum Einheitskreis immer spitzer werden. Läge ein Pol auf dem Einheitskreis, würde an dieser Stelle die Amplitude des Signals unendlich und damit das System instabil. Das deckt sich mit dem bereits besprochenen Kriterium: „Wenn ein digitales LTI-System stabil sein soll, müssen seine Pole im Inneren des Einheitskreises liegen.“

Wegen der Überlaufproblematik bei endlich genauen Zahlendarstellungen, ist es wichtig zu wissen, ob evtl. im Inneren des Filters größere Signalamplituden als die des Ausgangs-

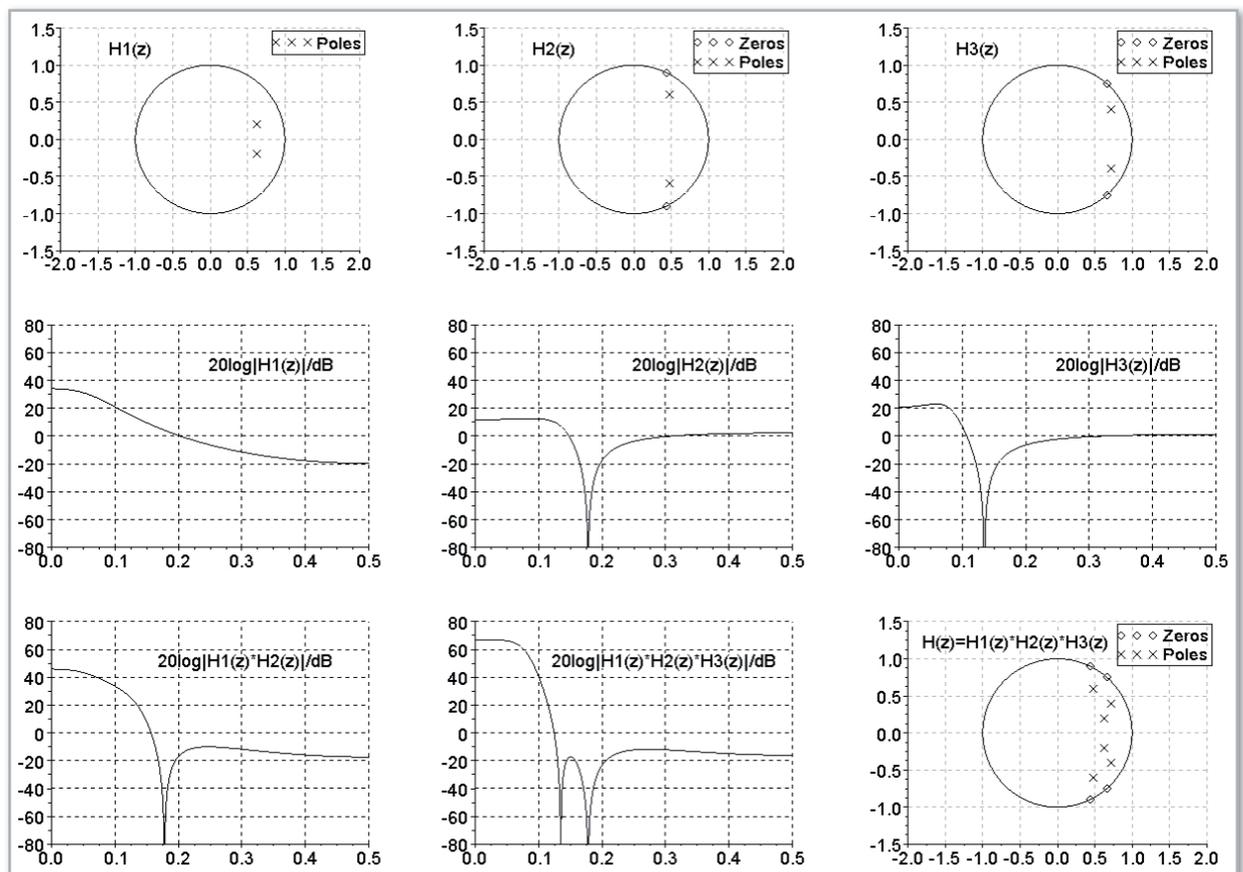


Bild 125: PN-Diagramme und Verstärkungen der Teilübertragungsabschnitte gemäß Gleichung (160) und deren Überlagerungen.

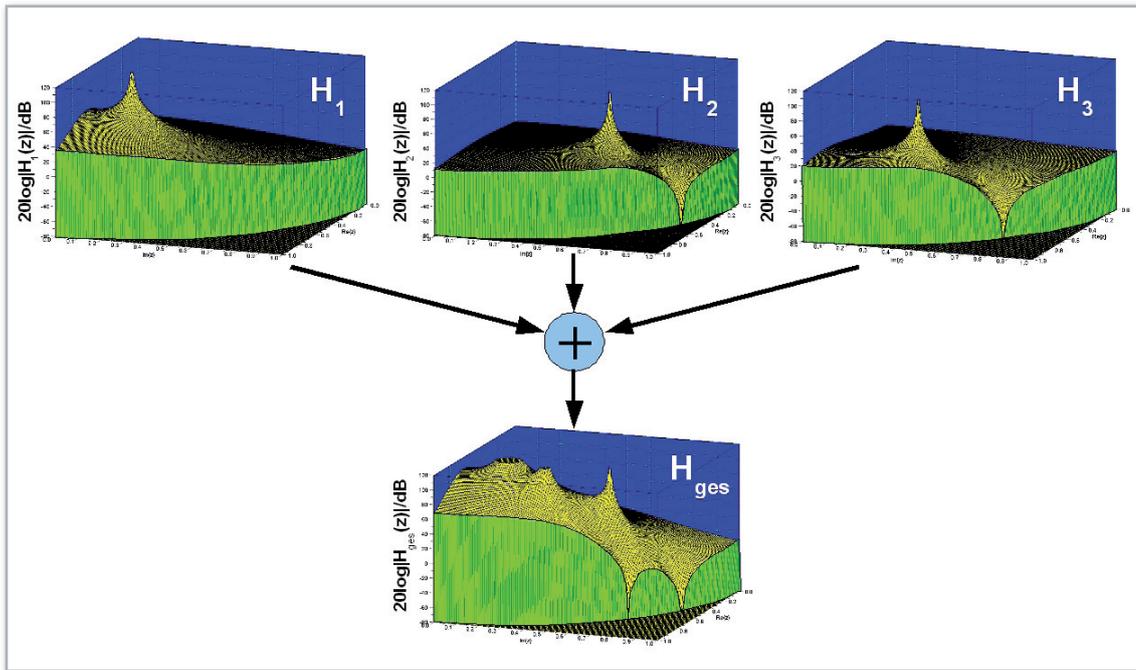


Bild 126: In der dreidimensionalen Darstellung wird deutlich, dass die Pole immer steiler ansteigen, je näher sie an den Einheitskreis heranrücken.

signals auftreten. Dann müsste das Eingangssignal so weit herabgesetzt (skaliert) werden, dass der höchste Pegel gerade noch nicht zu einer Überschreitung des darstellbaren Zahlenbereichs führt. Wir wollen das an einem Biquad in der 1. kanonischen Direktform untersuchen (Abbildung 127). Wir bezeichnen die inneren Knoten mit Knoten 1 und Knoten 2 und beziehen ihre Ausgangssignale auf das Eingangssignal $X(z)$. Damit haben wir die Übertragungsfunktionen $K_1(z)$ und $K_2(z)$ zu den inneren Knotenausgängen definiert. Knoten 0 bedarf keiner Betrachtung, weil sein Ausgangssignal das Ausgangssignal des Biquads ist.

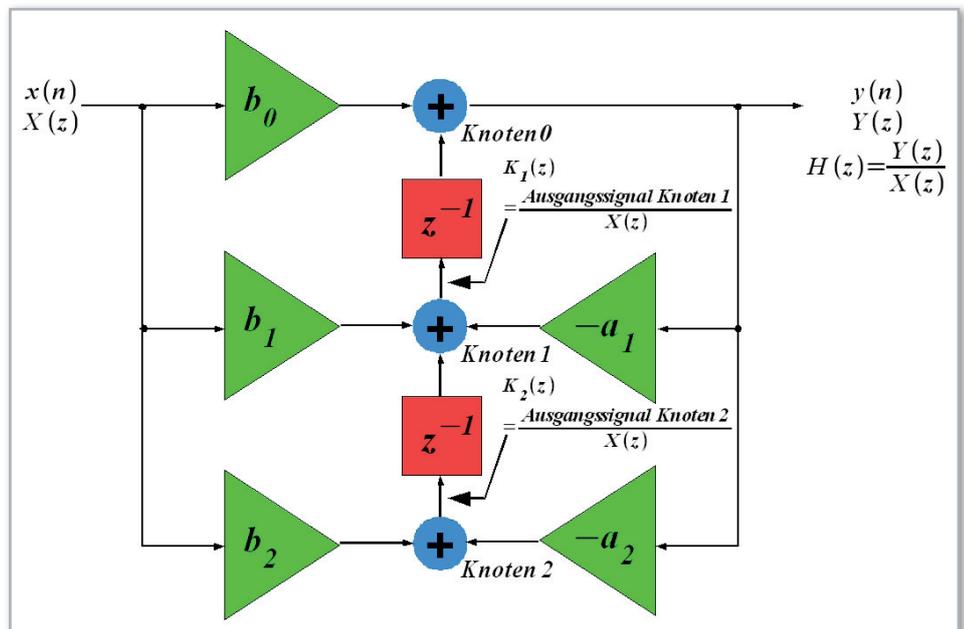
Die Berechnung von $K_1(z)$ und $K_2(z)$ wird in Gleichung (161) vorgeführt: Das Ergebnis ist in Abbildung 125 dreidimensional dargestellt. Auch hier wird nur der 1. Quadrant ($0 < \text{Re}(z) < 1, 0 < \text{Im}(z) < 1$) abgebildet. Dabei sind einige Nullstellen nicht zu erkennen. Man kann sie der Tabelle am Ende des Artikels entnehmen.

Die rote Linie in Abbildung 128 markiert den Pegel 0 dB. Man

erkennt, dass im „Inneren“ eines jeden Biquads an den Ausgängen der Summierknoten kein höherer Pegel als an seinem Eingang auftritt. Damit sind nur die Ausgänge der Kettenbiquads durch entsprechende Skalierung der Eingangssignale gegen Überlauf zu schützen.

Kommen wir nun zu den Impulsantworten. Sie hängen bei LTI-Systemen über die inverse z -Transformation mit den Übertragungsfunktionen zusammen. Betrachten wir dazu Abbildung 129, die mit Scilab erstellt wurde. In der oberen Zeile sind die Teilübertragungsfunktionen der in Reihe geschalteten Biquads abgebildet, darunter deren Impulsantworten in roter Farbe. Weil das Produkt der Teilübertragungsfunktionen die Gesamtübertragungsfunktion ist, werden die roten Impulsantworten zur grünen Gesamtimpulsantwort unten links gefaltet (Multiplikation im Frequenzbereich \neq Faltung im Zeitbereich!) Das gleiche Ergebnis erhält man, wenn die Gesamtübertragungsfunktion in den Zeitbereich transformiert wird (blaue Impulsantwort unten rechts). In der Beschriftung der durch

Bild 127: Zu den Summierknotenansgängen im Inneren der 1. Kanonischen Direktform bestehen die Übertragungsfunktionen $K_1(z)$ und $K_2(z)$.



Differenzgleichung am Ausgang von Knoten 1 aufstellen

$$k_1(n) = b_1 x(n) - a_1 y(n) + b_2 x(n-1) - a_2 y(n-1) \quad (161)$$

Differenzgleichung z -transformieren und durch $X(z)$ teilen

$$k_1(z) = b_1 X(z) - a_1 Y(z) + b_2 X(z)z^{-1} - a_2 Y(z)z^{-1}$$

$$K_1(z) = (b_1 + b_2 z^{-1}) - (a_1 + a_2 z^{-1})H(z) \quad \text{mit } H(z) = \frac{Y(z)}{X(z)}$$

$$H(z) = \frac{b_2 z^{-2} + b_1 z^{-1} + b_0}{a_2 z^{-2} + a_1 z^{-1} + 1} \quad \text{in } K_1(z) \text{ einsetzen und umformen ergibt}$$

$$K_1(z) = \frac{(b_2 - a_2 b_0)z^{-1} + (b_1 - a_1 b_0)}{a_2 z^{-2} + a_1 z^{-1} + 1} = \frac{(b_1 - a_1 b_0)z^2 + (b_2 - a_2 b_0)z}{z^2 + a_1 z + a_2}$$

ebenso ist mit der Differenzgleichung am Ausgang von Knoten 2 zu verfahren

$$k_2(n) = b_2 x(n) - a_2 y(n)$$

woraus in gleicher Weise folgt

$$K_2(z) = \frac{(a_1 b_2 - a_2 b_1)z^{-1} + (b_2 - a_2 b_0)}{a_2 z^{-2} + a_1 z^{-1} + 1} = \frac{(b_2 - a_2 b_0)z^2 + (a_1 b_2 - a_2 b_1)z}{z^2 + a_1 z + a_2}$$

Faltung entstandenen grünen Gesamtimpulsantwort wird der Doppelstern (**) als Symbol für die Faltungsoperation verwendet, weil in Scilab der einfache Stern (üblicherweise das Faltungssymbol) für die Multiplikation steht.

In Abbildung 121 wurde gezeigt, welchen großen Einfluss eine geringe Polverschiebung bei der Parallelschaltung der Partial-Übertragungsfunktionen auf die Lage der Nullstellen der Gesamtübertragungsfunktion hat. Wird bei der Serienschaltung der gleiche Pol in der ersten Teilübertragungsfunktion ($H_1(z)$) variiert, hat das keinerlei Auswirkungen auf Pole und Nullstellen der anderen Teilübertragungsfunktionen ($H_2(z)$, $H_3(z)$). Quantisierungsfehler bei der Darstellung der Koeffizienten eines Teilfilters wirken sich also nur auf dieses aus.

Auf jeden Fall ist für die praktische Realisierung digitaler Systeme die Berücksichtigung einer eingeschränkten Rechengenauigkeit unerlässlich. In seinem Buch „Digital Filters“ hat Richard Wesley Hamming (1915 - 1998), dem die Nachrichtentechnik viel zu verdanken hat (beispielsweise in der Codierungstheorie die

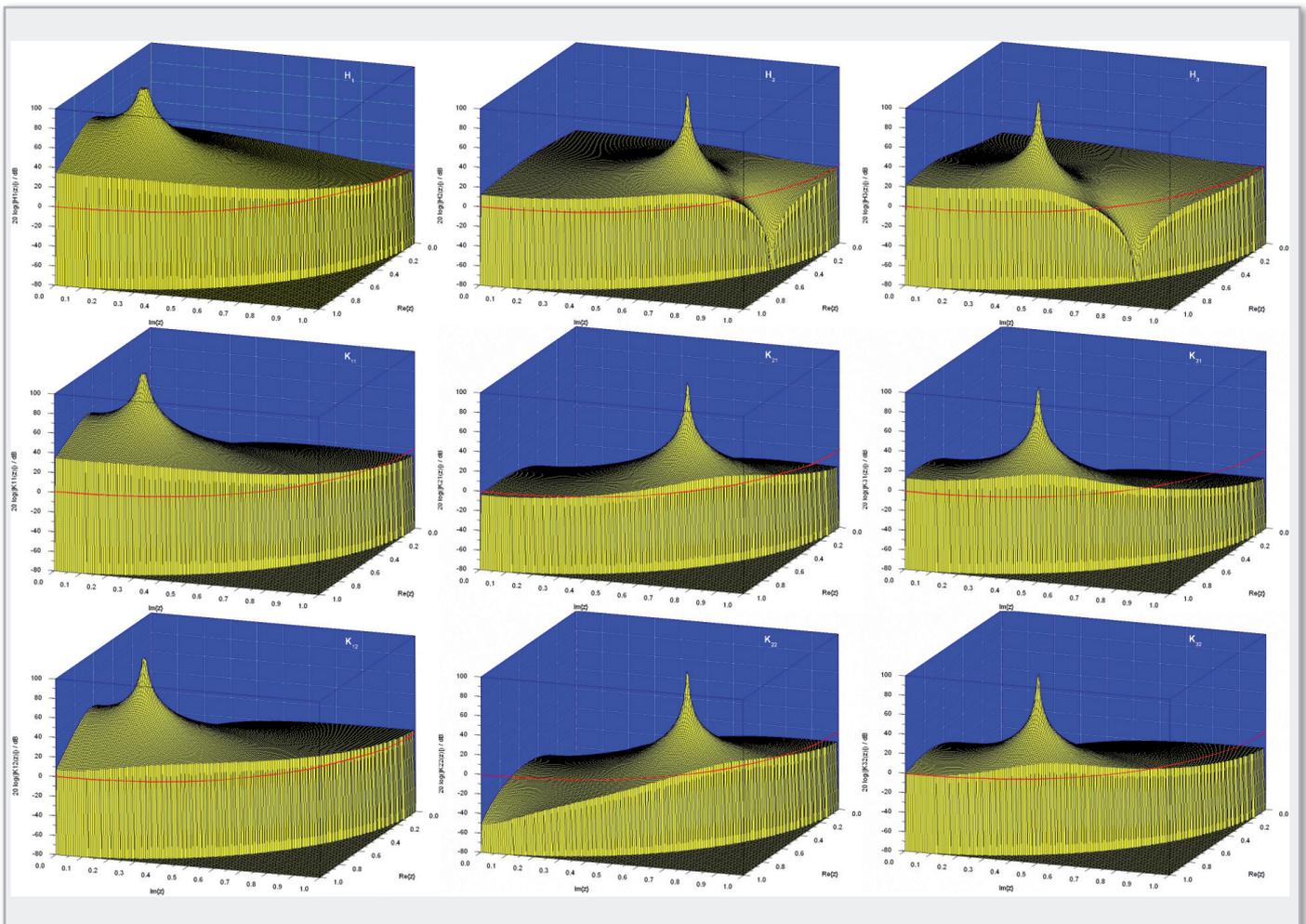


Bild 128: Das Übertragungsverhalten der Biquad-Kettenglieder an deren Ausgängen und an den inneren Summationsknoten.

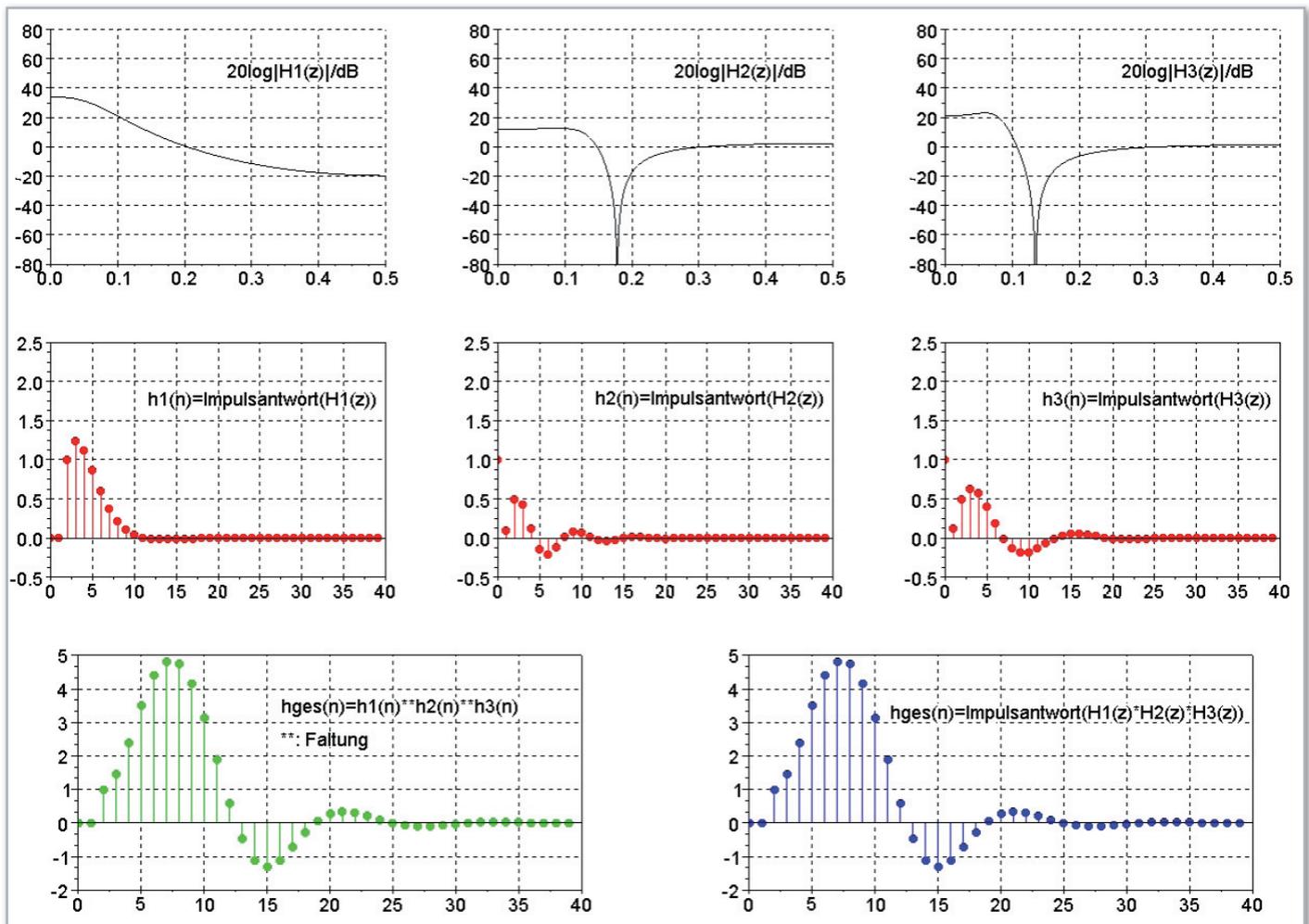


Bild 129: Durch Faltung der Impulsantworten der Biquad-Kettenglieder ergibt sich die Impulsantwort des Gesamtsystems.

„Hammingdistanz“, in der digitalen Signalverarbeitung das „Hammingfenster“), es so formuliert: „It is often convenient to think in terms of infinite precision arithmetic and perfect input data u_n ; but in the end we must deal with reality.“ Auf Deutsch heißt dies etwa: „Es ist oft bequem im Hinblick auf eine unendlich genaue Arithmetik und perfekte Eingangsdaten zu *denken*; aber letztendlich müssen wir mit der Realität fertig werden.“

Zusammenfassung

Kaskaden- und Parallelstrukturen von Biquads verhalten sich in Bezug auf Quantisierungseffekte günstiger als Direktstrukturen.

Anders als bei der Direktstruktur beeinflusst die Quantisierung der Filterkoeffizienten in einer Kaskade nicht alle Pole und Nullstellen gleichzeitig, sondern immer nur die PN-Paare des betreffenden Biquads und hat damit einen kleineren Einfluss auf den Frequenzgang.

Bei durch Partialbruchzerlegung gewonnenen Parallelstrukturen wirkt sich die Genauigkeit einer Pollage auch auf die Lage der Nullstellen aus. Liegen Nullstellen nicht mehr exakt auf dem Einheitskreis, hat das signifikante Auswirkungen auf den Frequenzgang. Diesem Nachteil steht als Vorteil gegenüber, dass das Quantisierungsrauschen eines Biquads nicht durch nachfolgende Biquads verstärkt werden kann und man sich

über die Paarung von Polen und Nullstellen sowie die Reihenfolge in einer Kette keine Gedanken zu machen braucht. Die Struktur ist vollständig durch die Partialbruchzerlegung der Übertragungsfunktion vorgegeben. Durch die Parallelisierung ist die Signallaufzeit zwischen Eingang und Ausgang kleiner als bei einer Serienschaltung, bei der sich die Gesamtlaufzeit als Summe der Laufzeiten der Kettenglieder ergibt.

Ausblick

In der nächsten Folge wollen wir uns mit den Eigenschaften binärer Zahlen (insbesondere des Zweierkomplements K_2), den Methoden zu deren Kürzung auf eine gewünschte Länge (abschneiden, runden) und den Effekten begrenzter Wortlängen eingehender beschäftigen. **ELV**

Biquad Nr. x =>	1	2	3
1. Nullstelle von K_{x1}	0	0	0
2. Nullstelle von K_{x1}	-	-4,121945	-2,774403
1. Nullstelle von K_{x2}	0	0	0
2. Nullstelle von K_{x2}	1,246000	1,114404	1,687228



In Teil 4 dieser Folge haben wir im Zusammenhang mit der Behandlung des A/D-Wandlers das Thema Zahlenformate, endliche Darstellungsgenauigkeit und die damit verbundenen rausch-ähnlichen Störungen gestreift. Für die genauere Analyse der Effekte greifen wir das Thema jetzt noch einmal auf.

Im Folgenden betrachten wir die Binärzahlendarstellung nach Vorzeichen und Betrag sowie als Zweierkomplement (K2). In einem Speicherregister mit $(B+1)$ Zellen sind die Bits zur Darstellung des Zahlenwertes abgelegt (Abbildung 130). An dem Registerinhalt allein kann man nicht erkennen, um welche Zahl es sich handelt. Das hängt allein von der Interpretation der Registerzellen ab.

Binärzahlendarstellung in Vorzeichen-Betrag-Darstellung

Eine auf den ersten Blick besonders einfache Darstellung von binären Zahlen scheint die Vorzeichen-Betrag-Darstellung (Sign-Magnitude Representation) zu sein. Negative ganze Zahlen x werden durch das Paar $(b_0, |x|)$ codiert. Dabei wird das ganz links im Binärwort stehende Bit b_0 als Vorzeichenbit gemäß Gleichung (162) interpretiert.

Bei negativen Binärwörtern hat b_0 den Wert 1, d. h. vor den positiven Rest des Binärworts (Betrag von x , $|x|$) wird ein Minuszeichen gesetzt. Bei positivem Vorzeichen ist $b_0 = 0$.



Bild 130: So werden die Bits für Vorzeichen und Betrag eines Binärwortes in einem Register mit $B+1$ Speicherzellen abgelegt. Erst die Interpretation dieser Speicherzellen führt zu einem konkreten Zahlenwert.

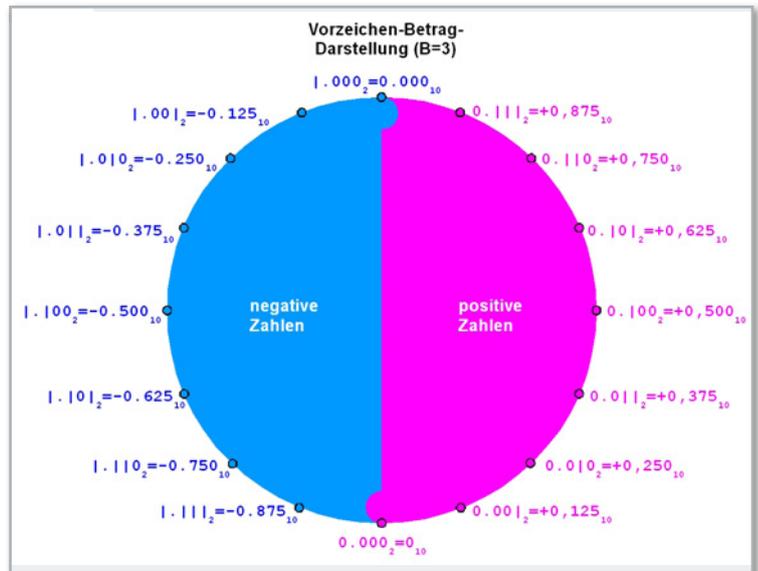
Zudem hat die Null zwei Repräsentationen, man verschwendet also einen Zahlenwert. Das Rechnen mit Binärzahlen ist technisch aufwändig, da viele Fallunterscheidungen aufgrund des Vorzeichens der Operanden und der doppelten Null notwendig sind. Im Zahlenring (Abbildung 131) wird dies anhand von dem mit $B = 3$ Bit dargestellten Zahlenbereich $-0,875 \leq x \leq +0,875$ (allgemein: $2^{-B-1} \leq x \leq -[2^{-B-1}]$) demonstriert. Der Punkt zwischen dem Vorzeichenbit und den Betragsbits heißt Binärpunkt. Er entspricht dem aus der Dezimalrechnung bekannten Dezimalkomma. Dass hier nur Bits rechts vom Binärpunkt angenommen werden, beschreibt, dass ausschließlich Stellenwertigkeiten $2^{-1} = 0,5$, $2^{-2} = 0,25$ und $2^{-3} = 0,125$ auftreten. In der Summe kann eine endliche Anzahl von Bits rechts vom Binärpunkt daher stets nur Werte <1 darstellen. Wegen der größeren Relevanz für die Binärarithmetik wollen wir es dabei belassen und uns eingehender mit dem Zweierkomplement beschäftigen.

Binärzahlendarstellung im Zweierkomplement

In der digitalen Signalverarbeitung ist die Darstellung von Fixpunktzahlen im Zweierkomplement K2 am gebräuchlichsten. Im K2 wird eine reelle Zahl im Intervall mit unendlicher Genauigkeit gemäß Gleichung (163) dargestellt:

X_m ist dabei ein beliebiger Skalierungsfaktor. Wird er zu 1 gewählt, beträgt der darstellbare Zahlenbereich $-1 \leq x < +1$. b_0 bezeichnet man als Vorzeichenbit und die b_v als Binärstel-

Bild 131: Zahlenringdarstellung eines 4-Bit-Wortes nach Vorzeichen und Betrag (VZB)



lenbits. Ist das Vorzeichenbit 0, ist die Zahl positiv (>0), andernfalls negativ (<0). Betrachten wir dazu das Zahlenbeispiel in Gleichung (164):

Wir wissen, dass der Kehrwert von 3 einen unendlichen (d. h. nicht abbrechenden) Dezimalbruch ergibt, nämlich 0,33333... In der Binärdarstellung äußert sich dies durch eine unendlich lange Folge von 01-Abschnitten rechts vom Binärpunkt. Die 10 nach dem Binärpunkt angegebenen Stellen entsprechen der Dezimalzahl 0,330078125, der unendlich lange Rest der 01-Folge hat den Wert 0,000325520833333... Zusammen haben wir 1/3 jetzt auf 15 Nachkommastellen genau angegeben, aber immer noch nicht exakt, was ja mit einer endlichen Zahlenlänge auch nicht möglich ist.

Es ist also offensichtlich erforderlich, einen Kompromiss zwischen Rechengenauigkeit und -aufwand bei der Länge der Zahlendarstellung einzugehen. Wir müssen dazu in Gleichung (163) als oberen Summationsindex statt ∞ den endlichen Wert B einsetzen (Gleichung 165).

Das Dach über den x-Werten zeigt an, dass durch Kürzen der Wortlänge die Genauigkeit der Darstellung abgenommen hat, weil die Zahlendarstellung jetzt quantisiert ist. Der kleinste Unterschied zwischen zwei Zahlen heißt Quantisierungsintervall (Gleichung [166]).

Rundung. Den nichtlinearen Zusammenhang zwischen exaktem und quantisierten x zeigt Abbildung 132 am Beispiel einer gerundeten Zweierkomplementzahl mit $B = 2$, d. h. einem Quantisierungsintervall von $2^{-B} = 1/4$. Weil in der unteren Hälfte eines Quantisierungsintervalls abgerundet und in der oberen aufgerundet wird, finden die Kennliniensprünge in der Intervallmitte statt. Das Bit links vom Binärpunkt b_0 wird als Vorzeichenbit bezeichnet. Bei positiven Zahlen ist es 0, bei negativen 1.

Überlauf. Entsprechend Gleichung (165) erhält man mit $B = 2$ die kleinste darstellbare Zahl mit dem Wert -1 (1.00₂). Die größte darstellbare Zahl ergibt sich für $b_0 = 0$ und $b_1 = b_2 = 1$ zu 0,75 (0.11₂). Ein Überlauf äußert sich im größtmöglichen Sprung des quantisierten Zahlenwerts von 0,75 (0.1₂) auf -1 (1.00₂) oder umgekehrt. In einer Reihe von Additionen und Subtraktionen dürfen daher die Zwischenergebnisse über-

laufen, wenn nur das Endergebnis im darstellbaren Zahlenbereich liegt. Das Negieren einer K2-Zahl erfolgt durch das Invertieren aller Bits (0→1, 1→0) und die Addition einer binären 1 an der niedrigstwertigen Stelle (LSB: Least Significant Bit). Deshalb lassen sich in der K2-Darstellung Subtraktionen einfachst auf Additionen zurückführen. Diese bemerkenswerten Eigenschaften des K2-Zahlenformats fasst Abbildung 133 zusammen.

$\hat{x} = X_m \cdot \left(\frac{(-1)^{b_0}}{\text{Vorzeichen}} \cdot \sum_{v=1}^B b_v \cdot 2^{-v} \right) \quad (162)$
<p>Endlich genaue Darstellung einer Zahl x nach Vorzeichen und Betrag mit B+1 Bits: $b_1, b_2, \dots, b_B \rightarrow \hat{x}_B$ und $b_0 = 0 \rightarrow +$ bzw. $b_0 = 1 \rightarrow -$ X_m: Skalierungsfaktor</p> <p>$-X_m \cdot (1 - 2^{-B}) \leq \hat{x} \leq X_m \cdot (1 - 2^{-B})$ Darstellbares Zahlenintervall</p> <p>$\hat{x}_B = b_0 \cdot b_1 b_2 b_3 \dots b_{B-1} b_B$ positionsbezogene Notation</p>
$x = X_m \left(-b_0 + \sum_{v=1}^{\infty} b_v \cdot 2^{-v} \right) \quad \text{Unendlich genaue Darstellung einer Zahl x im K2} \quad (163)$
<p>unendlicher Dezimalbruch: $\frac{1}{3} = 0,333333 \dots_{10} \hat{=} 0.0101010101 \dots_2$ $0.0101010101 \dots_2 = (1 \cdot 2^{-2} + 1 \cdot 2^{-4} + 1 \cdot 2^{-6} + 1 \cdot 2^{-8} + 1 \cdot 2^{-10} + \dots)_{10}$ $= \left(\frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} + \frac{1}{1024} + \dots \right)_{10} = 0,3330078125_{10} + \text{Rest}$ mit Rest = 0,000325520833333...₁₀</p>
$\hat{x} = X_m \left(-b_0 + \sum_{v=1}^B b_v \cdot 2^{-v} \right) \quad \text{Endlich genaue Darstellung einer Zahl x im K2} \quad (165)$
<p>mit $-X_m \leq \hat{x} < X_m$ Darstellbares Zahlenintervall</p> <p>und $\hat{x}_B = b_0 \cdot b_1 b_2 b_3 \dots b_{B-1} b_B$ positionsbezogene Notation</p>
$\Delta = X_m \cdot 2^{-B} \quad \text{Quantisierungsintervall im K2} \quad (166)$
$e = \hat{x} - x = X_m \cdot \hat{x}_B - x \quad \text{Quantisierungsfehler} \quad (167)$

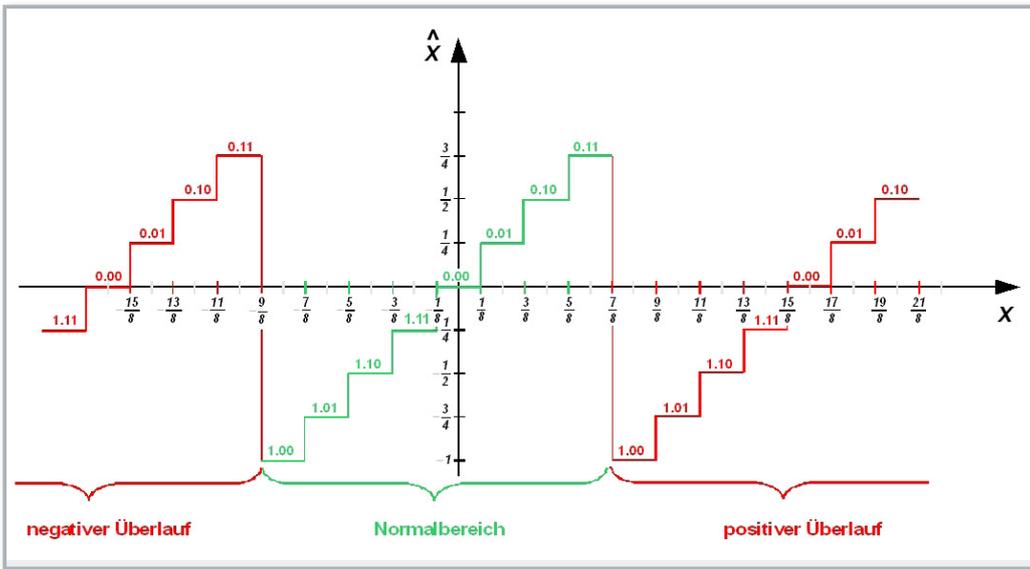


Bild 132: Die Kennlinie eines gerundeten 3-Bit-Binärwortes im Zweierkomplement (K2). Beim Überschreiten der Darstellungsgrenzen (Überlauf) ergibt sich der größtmögliche Fehler.

Die Darstellung der Zweierkomplementzahlen als Zahlenring (Abbildung 134) lässt (in Analogie zu einem mechanischen Zählwerk, wie es früher im Tonbandgerät eingesetzt wurde) anschaulich erkennen, warum Zwischenüberläufe beim Addieren und Subtrahieren ohne Belang sind.

Begrenzung. Wenn nicht davon ausgegangen werden kann, dass Zwischenüberläufe durch nachfolgende Additionen wieder kompensiert werden, indem das Endergebnis in den Normalbereich der K2-Darstellung zurückgeführt wird, muss man den Überlaufsprung vermeiden. Die Größe des mit einem Überlauf verbundenen Fehlers kann man limitieren, indem man den binären Zahlenwert konstant hält, solange der Signalwert die Darstellungsgrenzen über- oder unterschreitet (Abbildung 135). Dieses Verfahren ersetzt also den heftigen Überlaufsprung durch die darstellbaren Grenzwerte und wird auch als Sättigungsüberlauf oder Clipping bezeichnet. Es findet in der Regel bei A/D-Wandlern Anwendung, hebt aber die nützliche Eigenschaft der K2-Darstellung auf, nach der Zwischenüberläufe durch Additionen ohne Bedeutung sind,

wenn nur das Endergebnis im darstellbaren Zahlenbereich (Normalbereich) liegt.

Relativer Quantisierungsfehler. Im Zusammenhang mit A/D-Wandlern ist der relative Quantisierungsfehler als Abweichung der quantisierten von der unquantisierten Größe bezogen auf die unquantisierte Größe definiert. Bei gleichmäßiger Quantisierung (konstante Quantisierungsintervalle) ist er bei kleinen Signalwerten hoch und nimmt zu größeren ab. Bei digitalisierter Musik äußert sich das in einer „Rauheit“ des Klangs leiser Passagen. Dem kann man durch eine ungleichmäßige Quantisierungskennlinie entgegenwirken. Sie weist im Bereich kleiner Signalwerte eine feinere Stufung der Quantisierungsintervalle auf.

Durch entsprechende Formgebung der Quantisierungskennlinie (z. B. das logarithmische μ -law) lässt sich der Quantisierungsfehler (oder auch der Signal-Rausch-Abstand) über den gesamten Amplitudenbereich konstant halten. Natürlich benötigt der D/A-Wandler am Ausgang des Systems eine inverse Charakteristik.

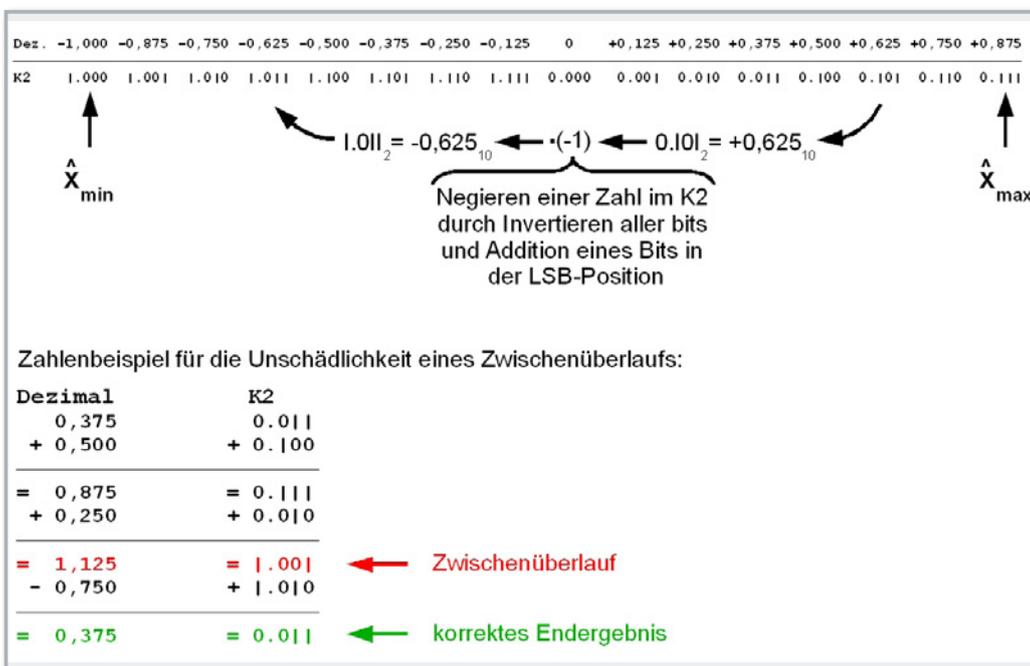
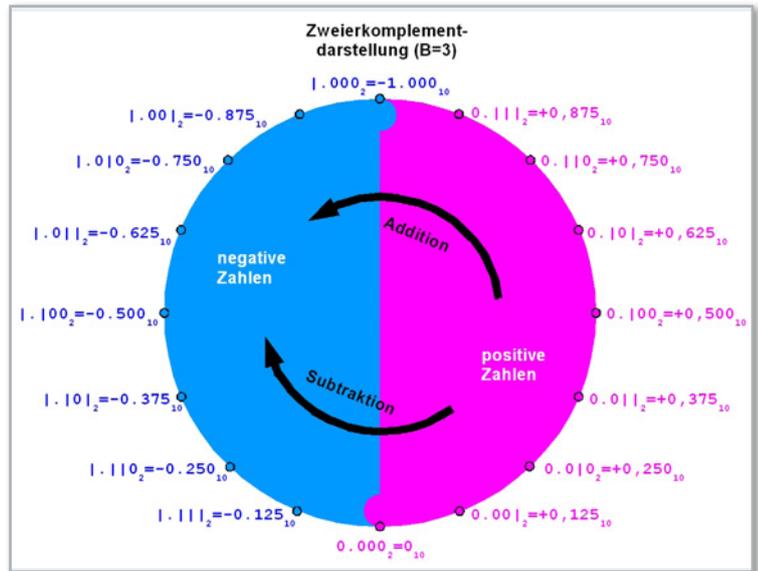


Bild 133: Im K2 wird das Vorzeichen einer Zahl ungekehrt, indem man jedes Bit invertiert und in der niedrigwertigsten Stelle (LSB) eine 1 addiert. Zwischenüberläufe sind im K2 unschädlich, solange das Endergebnis nur im Normalbereich liegt.

Bild 134: Zahlenringdarstellung eines 4-Bit-Wortes nach Vorzeichen und Betrag im K2



Weitere Quantisierungsmethoden. Wir sind bei der Darstellung der K2-Kennlinie in Abbildung 132 und Abbildung 135 davon ausgegangen, dass das niedrigstwertige Bit (LSB) der quantisierten Zahl als Rundungsergebnis die Größe des nicht dargestellten Restes widerspiegelt. Ist er größer oder gleich der Hälfte des LSB-Wertes, wird auf-, andernfalls abgerundet (round method). Eine andere Methode, zur quantisierten Zahl zu gelangen, besteht darin, den Zahlenrest einfach abzuschneiden (truncation method). Es ist offensichtlich, dass der dabei auftretende maximale Quantisierungsfehler doppelt so groß sein kann wie der Fehler bei Anwendung der Rundungsmethode. Dafür ist kein Aufwand für die Entscheidung zu treiben, ob auf- oder abgerundet werden soll. Abbildung 136 zeigt die zum Abschneiden gehörige Quantisierungskennlinie im K2.

Quantisierungsfehler. Man definiert den Quantisierungsfehler als Differenz zwischen dem durch Rundung oder Abschneiden gewonnenen und dem tatsächlichen Wert (Gleichung [167]).

In Abbildung 137 ist der Quantisierungsfehler gemäß Gleichung (167) als grüne Kurve am Beispiel einer mit drei Bit in Vorzeichen-Betrag-Darstellung quantisierten Sinusschwingung gezeigt. Er ist die Differenz zwischen der durch Auf-

oder Abrunden entstandenen Treppenkurve und dem tatsächlichen Verlauf des Sinus und wird deshalb als Rundungsfehler bezeichnet. In Abbildung 138 ist der Fehler blau dargestellt, der entsteht, wenn der Sinus durch Abschneiden in das Quantisierungsraster (Abschneidefehler) gebracht wird. Wie zu erwarten, ist der Fehler beim Runden (round) kleiner als beim Abschneiden.

Wird der Zahlenwert als Zweierkomplement dargestellt, verringern sich Rundungs- und Abschneidefehler im negativen Teil des Sinus, weil hier eine Quantisierungsstufe mehr zur Verfügung steht als bei der Vorzeichen-Betrag-Darstellung. Man muss auch die unterschiedliche Wirkung des Abschneidens bei negativen Zahlen beachten. In der Zweierkomplement-Darstellung bewirkt sie den Übergang zur nächstkleineren im Raster darstellbaren Zahl, in der Vorzeichen-Betrag-Darstellung zur nächstgrößeren (Abbildung 139, Abbildung 140).

Zusammenfassung. Bei der digitalen Zahlendarstellung führen zwei Effekte zu Fehlern: Quantisierung und Überläufe. Ohne die Zahl der Bits zu erhöhen, lässt sich Überläufen entgegenwirken, indem der Skalierungsfaktor X_m vergrößert wird. Dadurch nimmt aber auch die Größe des Quantisierungsfehlers proportional zu. Um den Aussteuerbereich zu

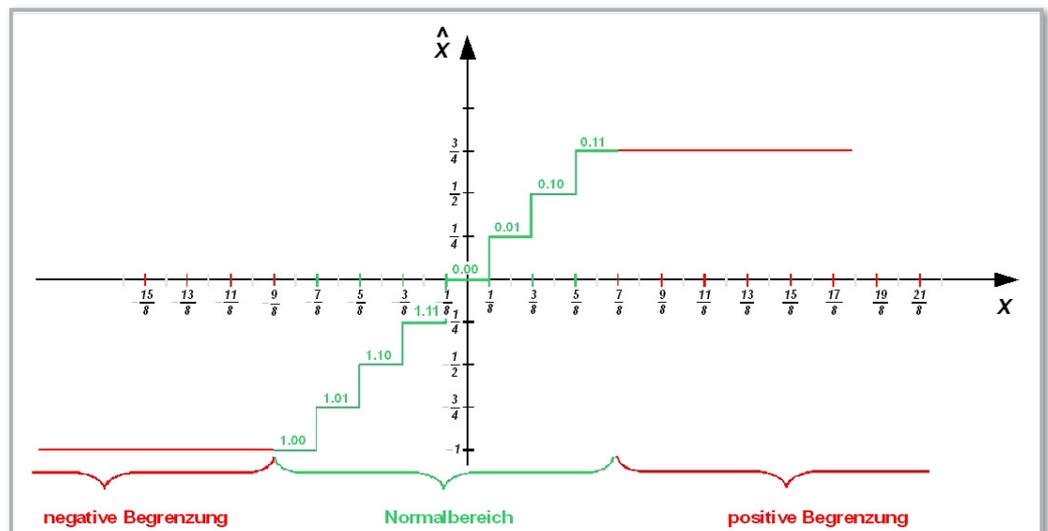


Bild 135: Im K2 wird der Sprung vom kleinstmöglichen auf den größtmöglichen Wert durch eine Sättigungskennlinie vermieden.

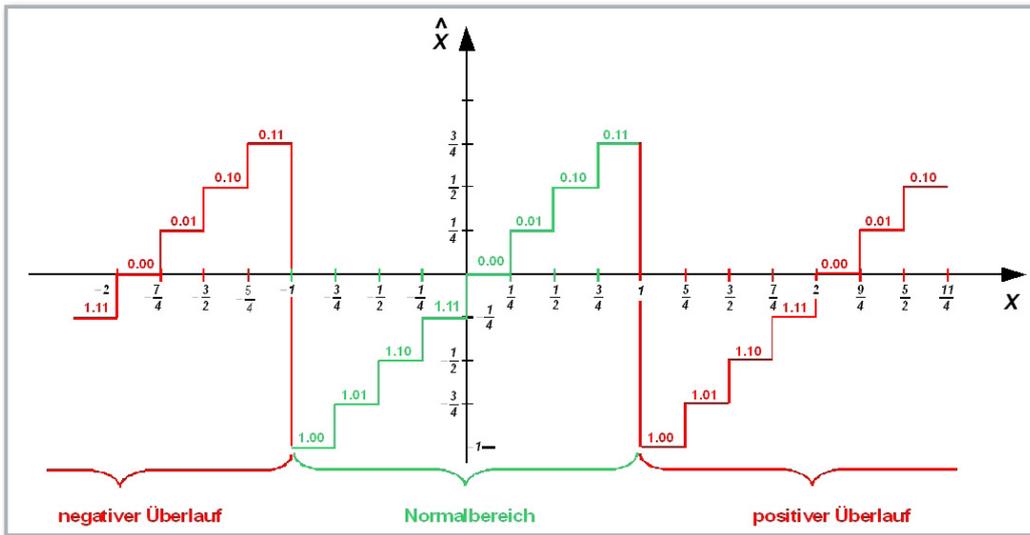


Bild 136: Kennlinie eines abgeschnittenen 3-Bit-Binärwortes im Zweierkomplement (K2)

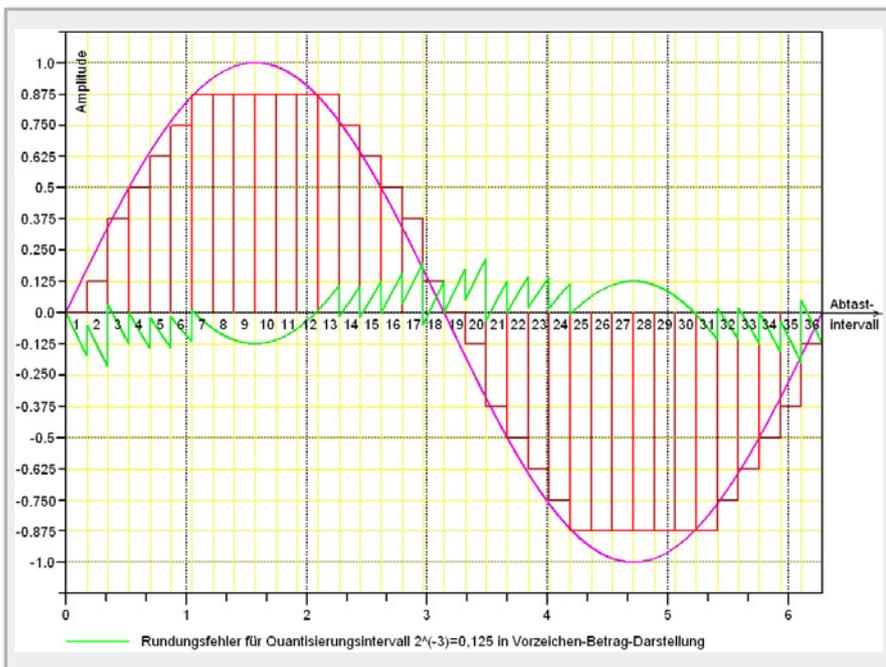


Bild 137: Der VZB-Rundungsfehler (grün) innerhalb einer Periode einer mit 3 Bit und 36 Abtastintervallen dargestellten Sinusschwingung

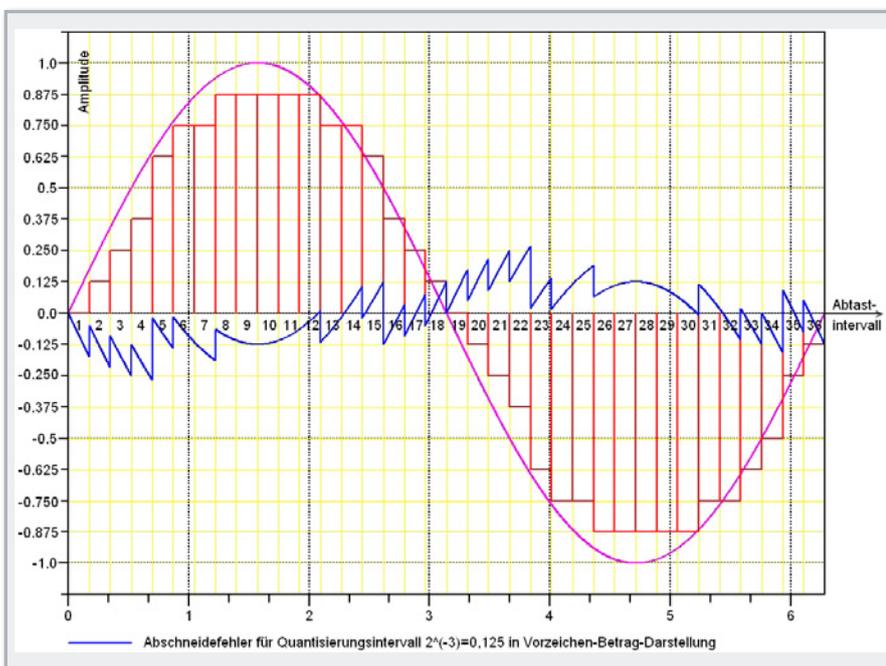


Bild 138: Der VZB-Abschneidefehler (blau) innerhalb einer Periode einer mit 3 Bit und 36 Abtastintervallen dargestellten Sinusschwingung

erweitern, ohne die Quantisierung zu vergrößern, muss daher die Zahl der Binärstellen (Wortlänge) erhöht werden. Bei gleicher Wortlänge ist der Fehler durch Runden am kleinsten.

Rechnen im Zweierkomplement

Die vier Grundrechenarten (Addition, Subtraktion, Multiplikation und Division) lassen sich auf Additionen, Shiften (Verschieben einer Binärzahl um eine oder mehrere Stellen nach links oder rechts) und bitweise Negation zurückführen. Wir wollen das nicht theoretisch herleiten, sondern anhand einiger Beispiele demonstrieren.

Addition im K2. In Abbildung 141 finden wir drei Beispiele zur Addition im Zweierkomplement.

Beispiel 1: Es werden zwei positive Dezimalzahlen $0,3125_{10}$ ($0,0101_2$) und $0,14453125$ ($0,0010010_2$) zu $0,45703125_{10}$ ($0,01110101_2$) addiert. Weil das Vorzeichenbit der Summe 0 ist, ist das Resultat positiv.

Merkregel: Das Ergebnis ist nur dann korrekt, wenn sein Vorzeichenbit 0 ist. Andernfalls wurde der Zahlenbereich überschritten.

Beispiel 2: Jetzt wird die erste Dezimalzahl von der zweiten abgezogen. Dazu wird sie vor der Addition zu $-0,3215_{10}$ ($1,1011_2$) negiert. Weil das Vorzeichenbit 1 ist, ist das Ergebnis negativ.

Merkregel: Bei Summanden mit unterschiedlichen Vorzeichen ist das Ergebnis auf jeden Fall korrekt. Sein Vorzeichen hängt davon ab, ob der Subtrahend oder der Minuend betragsmäßig größer ist.

Beispiel 3: Hier werden beide Dezimalzahlen negiert und

Bild 139: Der K2-Rundungsfehler (grün) innerhalb einer Periode einer mit 3 Bit und 36 Abtastintervallen dargestellten Sinusschwingung

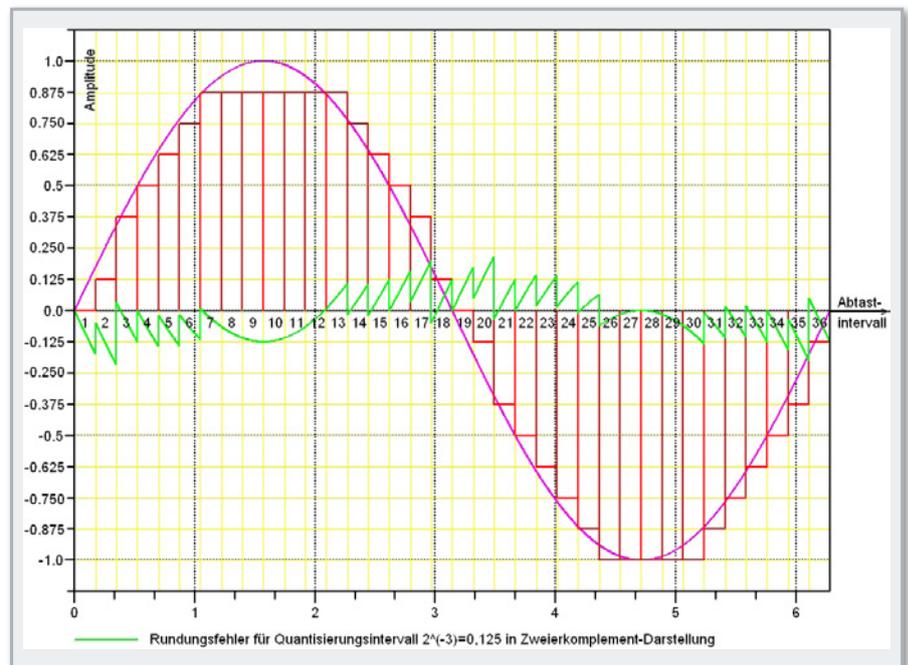
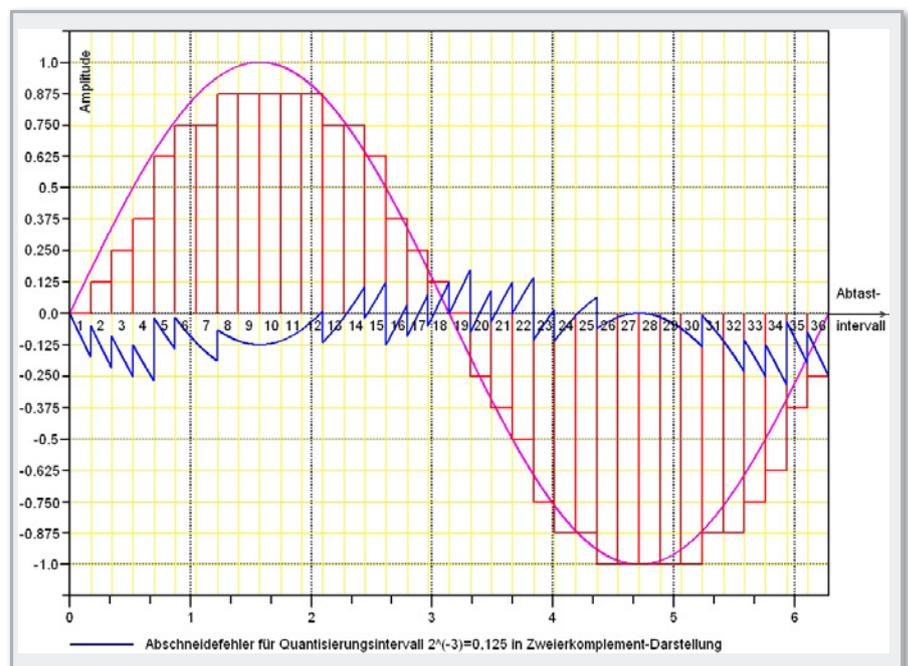


Bild 140: Der K2-Abschneidefehler (blau) innerhalb einer Periode einer mit 3 Bit und 36 Abtastintervallen dargestellten Sinusschwingung



dann addiert. Es ergibt sich ein Überlauf aus der Vorzeichenstelle, der vernachlässigt wird.

Merkregel: Das Ergebnis ist nur dann korrekt, wenn das Vorzeichenbit des Ergebnisses 1 ist und die Überträge in die Vorzeichenstelle hinein und aus ihr heraus gleich sind.

Multiplikation im K2

In den Beispielen 4 und 5 zur Multiplikation in Abbildung 142 werden die Beträge der Faktoren (erster Faktor = Multiplikator, zweiter Faktor = Multiplikand) miteinander multipliziert und das Ergebnis (Produkt) ggfs. negiert. Sind beide Vorzeichen positiv oder negativ, ist die Vorzeichenumkehr beim Produkt nicht notwendig, denn bekanntlich gilt: $+ * + = +$ und $- * - = +$. Sind die Vorzeichen der Faktoren unterschiedlich, muss das Ergebnis mit einem Minus versehen werden. Die Verknüpfung der Vorzeichen der Faktoren erfolgt über eine Exklusiv-Oder-Schaltung. Negiert wird im K2, indem jedes Bit der positiven Zahl invertiert ($0 \rightarrow 1$, $1 \rightarrow 0$) und zum LSB (Least Significant Bit) eine binäre 1 addiert wird.

Wie man den Beispielen zur Multiplikation entnehmen kann, ist die Länge des Produkts (12) gleich der Summe der Längen von Multiplikator (4) und Multiplikand (8). Weil voraussetzungsgemäß der Betrag der Faktoren ≤ 1 ist, kann der Betrag des Produkts nicht > 1 sein, es findet also kein Überlauf in die Vorzeichenstelle statt. Bei gleichen Registerlängen für die Faktoren und das Produkt stehen dagegen nicht immer genügend Speicherplätze für die niedrigwertigen Produktbits zur Verfügung, diese müssen unter Verursachung eines entsprechenden Rechenfehlers abgeschnitten werden.

Division im K2

Die Division (Dividend : Divisor) lässt sich ebenfalls auf die elementaren Operationen Addieren, Shiften und Invertieren zurückführen. Am einfachsten dürfte wohl die Vorgehensweise sein, die uns aus dem Schulunterricht für Dezimalzahlen vertraut ist. Zuvor vereinfachen wir uns das Leben, indem wir in beiden Faktoren den Binärpunkt so lange nach rechts schieben (das ist gleichbedeutend mit einer Verschiebung des Binärwortes nach links), bis beide ganzzahlig sind. Jede Verschiebung um 1 Bit entspricht einer Verdoppelung des Wertes. Im Ergebnis liefert die voll ausgeführte Division der ganzzahligen Werte das gleiche Resultat wie die Division der ursprünglich gebrochenen Werte. Wird die Division nicht bis zu ihrem Ende durchgeführt, entsteht ein Abschneidefehler. Aber weil ein Beispiel mehr als tausend Worte sagt, betrachten wir Beispiel 6 in Abbildung 143. Der Interessierte mag den Rechengang verfolgen: Wenn der Divisor nicht in den Dividenten passt, wird im Resultat eine 0 angeschrieben und eine Null aus den Nachkommastellen (die ja bei ganzen Zahlen stets 0 sind) geborgt, um den Dividenten zu verlängern. Ist der verlängerte Divident jetzt größer als der Divisor, wird eine 1 im Resultat angeschrieben und der Divisor vom verlängerten Dividenten abgezogen usw. Steht zur Aufnahme des Quotienten nur ein 16-Bit-Register zur Verfügung, müssen alle weiteren Stellen (rechts von der 15. Stelle nach dem Binärpunkt) abgeschnitten werden, was in Beispiel 5 zu einem Darstellungsfehler von 0,013 % führt. Eine Beschränkung auf 8 Bit erhöht den Fehler (weil mehr abgeschnitten wurde) auf 3,4 %.

Beispiel 1: Addition zweier positiver Zahlen im K2

Zweierkomplement:

$$\begin{array}{r} 0.0101 + 0.00100101 = 0.0101 \\ + 0.00100101 \\ \hline = 0.01101010 \\ = \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} \\ \text{Dezimal:} \\ 0,3125 + 0,14453125 = 0,45703125 \end{array}$$

Beispiel 2: Addition einer negativen und einer positiven Zahl im K2

Zweierkomplement:

$$\begin{array}{r} 1.1011 + 0.00100101 = 1.1011 \\ + 0.00100101 \\ \hline = 1.11011010 \\ = -1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} \\ \text{Dezimal:} \\ -0,3125 + 0,14453125 = -0,16796875 \end{array}$$

Beispiel 3: Addition zweier negativer Zahlen im K2

Zweierkomplement:

$$\begin{array}{r} 1.1011 + 1.11011011 = 1.1011 \\ + 1.11011011 \\ \hline = 1.10010101 \\ = -1 + \frac{1}{2} + \frac{1}{32} + \frac{1}{128} + \frac{1}{256} \\ \text{Dezimal:} \\ -0,3125 + (-0,14453125) = -0,45703125 \end{array}$$

Bild 141: Drei Beispiele für die Addition im K2

Ausblick

Alle bisher beschriebenen Effekte durch Runden und/oder Abschneiden können wir unter dem Begriff Quantisierungsfehler zusammenfassen. Sie führen im Allgemeinen zu Einschränkungen in der Genauigkeit der Darstellung eines Signalwertes. Man kann sich leicht vorstellen, dass die mittlere Größe der Fehler von der verfügbaren Wortlänge abhängt.

Bei realen Signalen (z. B. dem Verlauf der Luftdruckschwankungen bei einem Konzert) ist es jedoch völlig unmöglich, den exakten Fehlerverlauf vorherzusagen. Deshalb müssen wir die Methoden der Statistik heranziehen, um den zeitlichen Verlauf des Fehlers, der sich wie ein dem Originalsignal überlagertes Rauschen manifestiert, zu beschreiben.

Das wird Gegenstand von Teil 15 dieser Folge sein. **ELY**

Beispiel 4: Multiplikation zweier positiver Zahlen im K2

Zweierkomplement:
 $0.0|0| * 0.00|00|0| = 0.00000000$
 $+0.0000|00|0|$
 $+0.0000000000$
 $+0.000000|00|0|$

 $=0.0000|0||00|$
 $= \frac{1}{32} + \frac{1}{128} + \frac{1}{256} + \frac{1}{512} + \frac{1}{4096}$
 $= 0,045166015625$

Dezimal:
 $0,3125 * 0,14453125 = 0,045166015625$

Beispiel 5: Multiplikation einer negativen mit einer positiven Zahl im K2

Zweierkomplement:
 $1.1|0|| * 0.00|00|0| \Rightarrow 0.0|0| * 0.00|00|0| =$
 0.00000000
 $+0.0000|00|0|$
 $+0.0000000000$
 $+0.000000|00|0|$

 $=0.0000|0||00|$
 $\Rightarrow 1.1111|0|000||$
 $= -1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{64} + \frac{1}{1024} + \frac{1}{2048} + \frac{1}{4096}$
 $= -1 + 0,954833984375$
 $= -0,045166015625$

Dezimal:
 $-0,3125 * 0,14453125 = -0,045166015625$

Bild 142: Zwei Beispiele für die Multiplikation im K2

Beispiel 5: Division zweier positiver Zahlen im K2

Zweierkomplement: $0.00|00|0| : 0.0|0| = ?$ **Dezimal:** $0,14453125 : 0,3125 = 0,4625$

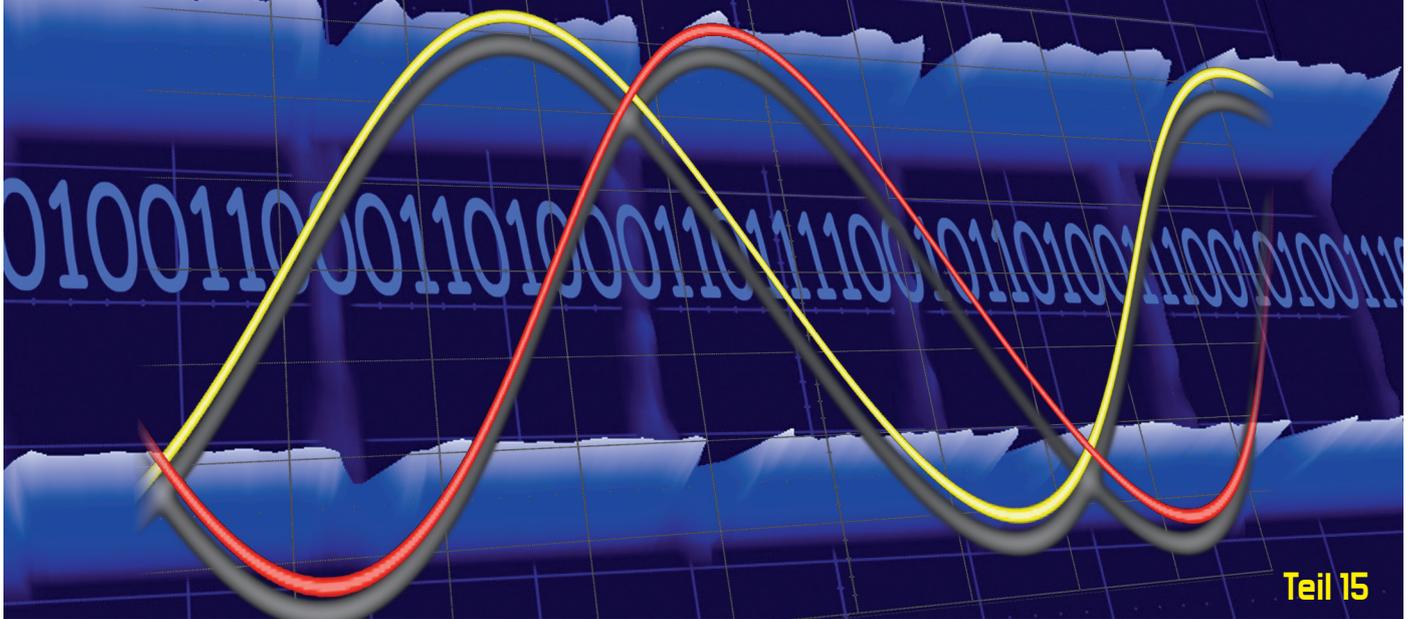
Ganzzahlig machen durch Verschieben des Binärpunkts um 8 Stellen nach rechts:
 Zweierkomplement: $000|00|0|. : 00|0|0000.$ **Dezimal:** $37 : 80 = 0,4625$

Rechengang:

$\begin{array}{r} 100 0 : 010000 = 0.011101100110011... \\ 100 0 \\ - 1010000 \\ \hline 1000 000 \\ - 1010000 \\ \hline 1110000 \\ - 1010000 \\ \hline 10000000 \\ - 1010000 \\ \hline 1100000 \\ - 1010000 \\ \hline 10000000 \\ - 1010000 \\ \hline 1100000 \\ - 1010000 \\ \hline 10000000 \\ - 1010000 \\ \hline 1100000 \\ \dots \\ \dots \\ \dots \end{array}$	\Rightarrow realer Quotient $0,462493896484375$ idealer Quotient $-0,462500000000000$ Fehler durch Beschränkung auf 16-Bit-Wort $-0,000006103515625$ ($\sim 0,013$ Promille) Fehler durch Beschränkung auf 8-Bit-Wort $-0,0015625$ ($\sim 3,4$ Promille)
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Bild 143: Ein Beispiel für die Division im K2

Einführung in die digitale Signalverarbeitung



Bevor wir in dieser 15. Folge der „Einführung in die digitale Signalverarbeitung“ das Phänomen des Rundungsrauschens, hervorgerufen durch eine endlich genaue digitale Arithmetik, näher betrachten, wollen wir deren Auswirkungen auf das Übertragungsverhalten an einem Beispiel etwas ausführlicher demonstrieren.

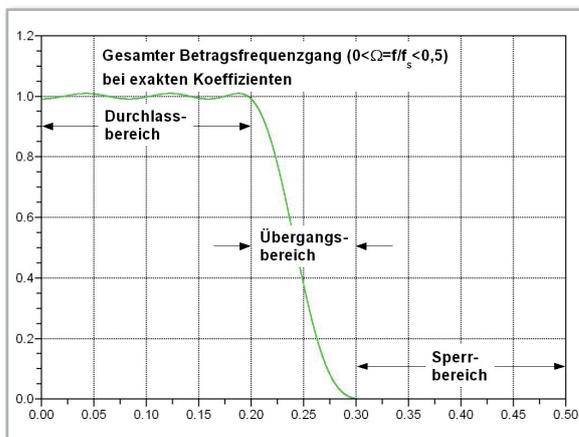


Abbildung 144: Gesamter Betragsfrequenzgang eines FIR-Tiefpasses mit 28 symmetrischen exakten Koeffizienten

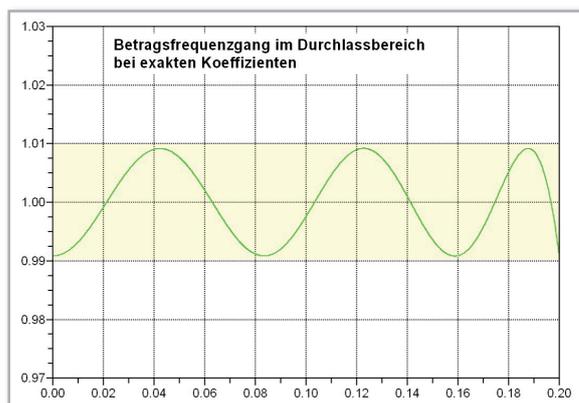


Abbildung 145: Betragsfrequenzgang eines FIR-Tiefpasses mit 28 symmetrischen exakten Koeffizienten im Durchlassbereich

Koeffizientenfehler und Übertragungsverhalten

In Anlehnung an ein Beispiel aus dem Standardwerk „Zeitdiskrete Signalverarbeitung“ von Alan V. Oppenheim et al. wird das Verhalten eines Tiefpassfilters mit endlicher Impulsantwort (FIR: Finite Impulse Response) betrachtet. Seine nicht-rekursive Struktur ergibt sich aus Abbildung 77 durch Weglassen des linken (rekursiven) Teils, seine Übertragungsfunktion aus Gleichung (127) durch Setzen aller $a_v=0$. Wir haben es dann mit einer Kette von Verzögerern zu tun, die zwischen den Verzögerungsgliedern sowie an Anfang und Ende „angepasst“ wird. Die abgegriffenen Signale werden über Multiplizierer bewertet und aufsummiert. Daher stammt auch der Name Tapped-Delay-Line-Filter (Filter mit angezapfter Verzögerer-Kette, siehe auch Abbildung 102 oben). Die Koeffizientenwerte sind identisch mit den Werten der Impulsantwort des Filters. Mathematisch gesehen werden die Folgenwerte des Eingangssignals mit der Impulsantwort im Zeitbereich zur Ausgangssignalfolge gefaltet. Im Frequenzbereich entspricht dies der Multiplikation des Eingangssignalspektrums mit der Übertragungsfunktion des Filters zum Ausgangssignalspektrum.

Der Frequenzgang weist im normierten Frequenzintervall $0 < \Omega = f/f_s < 0,5$ ($f_s = \text{sampling frequency} = \text{Abtastfrequenz}$) einen Durchlassbereich ($0 < \Omega < 0,2$), gefolgt von einem Übergangsbereich ($0,2 < \Omega < 0,3$) und abschließend dem Sperrbereich ($0,3 < \Omega < 0,5$). Das gesamte exakte Übertragungsverhalten zeigt Abbildung 144, der Betrag des Frequenzgangs im Durchlassbereich ist in Abbildung 145, der im

Sperrbereich in Abbildung 146 dargestellt. Die Lage der Nullstellen der Übertragungsfunktion lässt sich Abbildung 147 entnehmen. Man sieht deutlich, dass der Durchlassbereich im Wesentlichen von den nicht auf dem Einheitskreis der z-Ebene liegenden Nullstellen geprägt wird. Diese sind konjugiert reziprok, d. h. jeweils zwei liegen auf dem gleichen Strahl zum Ursprung und ihr Abstand zu diesem ist reziprok. Dies ist übrigens eine Bedingung für einen linearen Phasenverlauf im Durchgangsbereich. Der Sperrbereich wird dagegen fast ausschließlich von den Nullstellen auf dem Einheitskreis beeinflusst. Zur Veranschaulichung erinnern wir uns an das Modell eines elastischen Tuchs, das parallel zur z-Ebene aufgespannt und an den Nullstellen auf der z-Ebene „festgepinnt“ wird. Das Höhenprofil des so verformten Tuchs über dem Einheitskreis von $\Omega=0$ bis 0,5 ist der Betrag der Übertragungsfunktion ($|H(z)|$).

Koeffizientenwortlänge und Frequenzgang

Zur Realisierung sind die Filterkoeffizienten aus der Tabelle in Abbildung 148 erforderlich. Man gewinnt sie mit einem Entwurfsverfahren nach Parks-McLellan, auf das wir in einer späteren Folge noch eingehen. Realisiert man die exakten Filterkoeffizienten aus Spalte 2 in Vorzeichen-Betragsdarstellung mit 16, 14, 13 und 8 bit (1 bit für das Vorzeichen, der Rest zur Darstellung der abgeschnittenen Zahl), ergeben sich die Spalten 3, 4, 5 und 6. Die Koeffizienten lassen sich mit Scilab unter Verwendung der Funktionen „fmsg“ leicht in die zugehörige Übertragungsfunktion umrechnen. Mit „plot2d“ wird diese ebenso einfach abgebildet. Es ist nun interessant zu prüfen, welche Koeffizientenwortlängen erforderlich ist, damit die Übertragungsfunktion die Grenzen der gelben Toleranzschläuche nicht verletzt. Die Ergebnisse dieser Untersuchungen sind für 16, 13 und 8 bit lange Koeffizienten in den Abbildungen 149, 150 und 151 zusammengefasst. Bereits bei einer Koeffizientenwortlänge von 16 bit kommt es zu einer winzigen, aber noch vertretbaren Verletzung des Toleranzschlauchs im Sperrbereich, wie Abbildung 149 unten rechts zeigt. Bei 13 bit Wortlänge werden die Abweichungen vom exakten Sperrverlauf gravierender (Abbildung 150) und bei einer Wortlänge von 8 bit bricht der Betragsfrequenzgang heftig sowohl aus dem Sperr- als auch aus dem

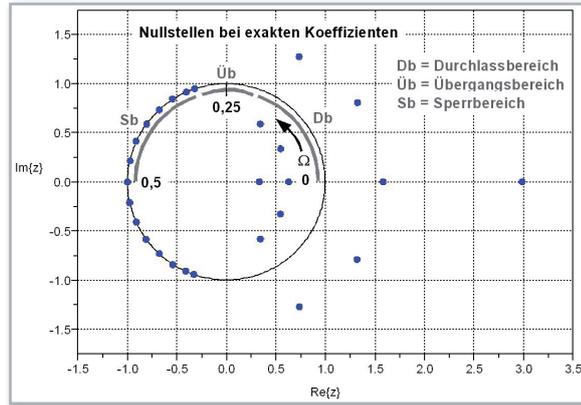


Abbildung 147: Nullstellendiagramm eines FIR-Tiefpasses mit 28 symmetrischen exakten Koeffizienten

Koeffizient	nicht quantisiert	16 bit	14 bit	13 bit	8 bit
$h(0)=h(27)$	0,001359657	0,0013427734375	0,0013427734375	0,001220703125	0,0
$h(1)=h(26)$	-0,001616993	-0,0015869140625	-0,0015869140625	-0,00146484375	0,0
$h(2)=h(25)$	-0,007738032	-0,007720947265625	-0,0076904296875	-0,007568359375	0,0
$h(3)=h(24)$	-0,002686841	-0,002685546875	-0,002685546875	-0,002685546875	0,0
$h(4)=h(23)$	0,01255246	0,012542724609375	0,012451171875	0,012451171875	0,0078125
$h(5)=h(22)$	0,00659153	0,006561279296875	0,0064697265625	0,00634765625	0,0
$h(6)=h(21)$	-0,02217952	-0,02215576171875	-0,0220947265625	-0,02197265625	-0,015625
$h(7)=h(20)$	-0,01524663	-0,015228271484375	-0,01513671875	-0,01513671875	-0,0078125
$h(8)=h(19)$	0,03720688	0,037200927734375	0,037109375	0,037109375	0,03125
$h(9)=h(18)$	0,03233332	0,032318115234375	0,0322265625	0,0322265625	0,03125
$h(10)=h(17)$	-0,06537057	-0,06536865234375	-0,0653076171875	-0,065385546875	-0,0625
$h(11)=h(16)$	-0,07528754	-0,075286865234375	-0,0751953125	-0,0751953125	-0,0703125
$h(12)=h(15)$	0,156097	0,15606689453125	0,156005859375	0,156005859375	0,1484375
$h(13)=h(14)$	0,4394094	0,43939208984375	0,4393310546875	0,439208984375	0,4375

Abbildung 148: Exakte und mit begrenzten Wortlängen dargestellte Koeffizienten des Beispiel-Tiefpasses

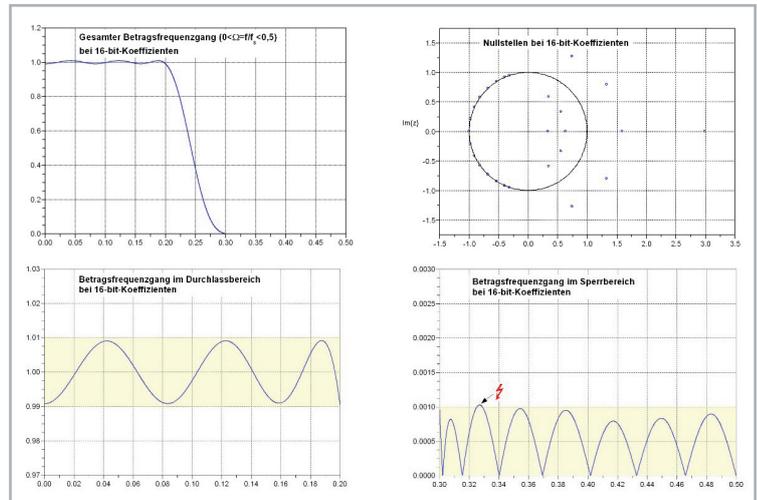


Abbildung 149: Betrag und Nullstellen des Beispiel-TP mit 16-bit-Koeffizienten

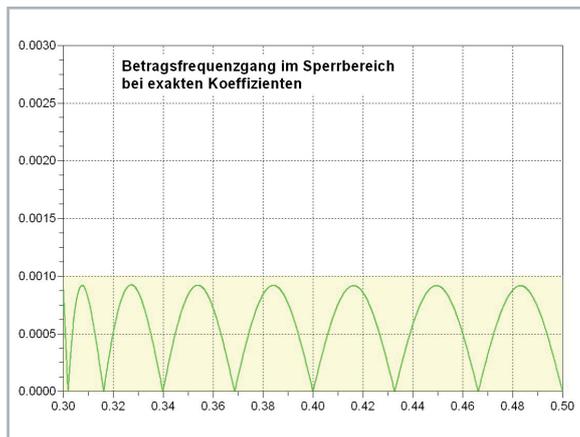


Abbildung 146: Betragsfrequenzgang eines FIR-Tiefpasses mit 28 symmetrischen exakten Koeffizienten im Sperrbereich

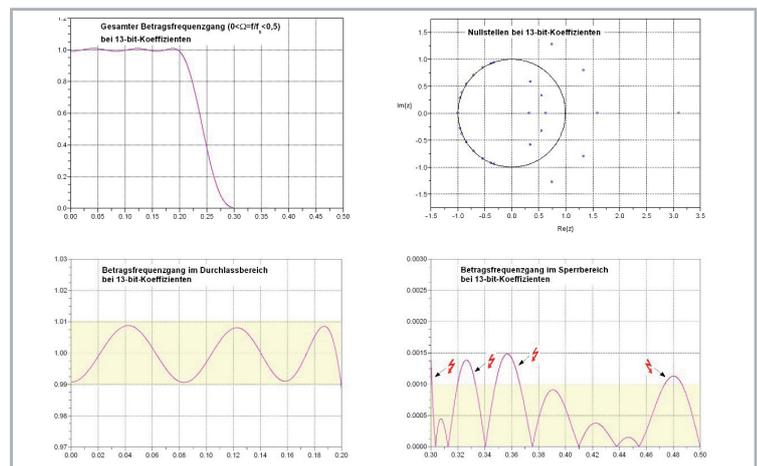


Abbildung 150: Betrag und Nullstellen des Beispiel-TP mit 13-bit-Koeffizienten

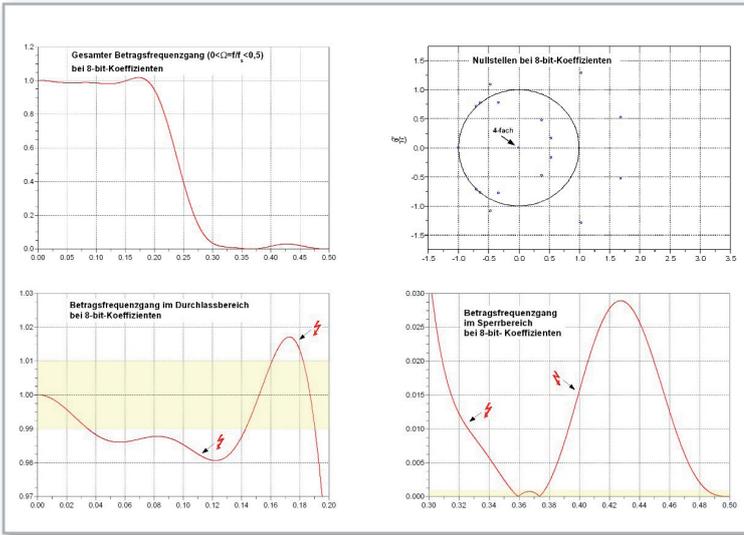


Abbildung 151: Betrag und Nullstellen des Beispiel-TP mit 8-bit-Koeffizienten

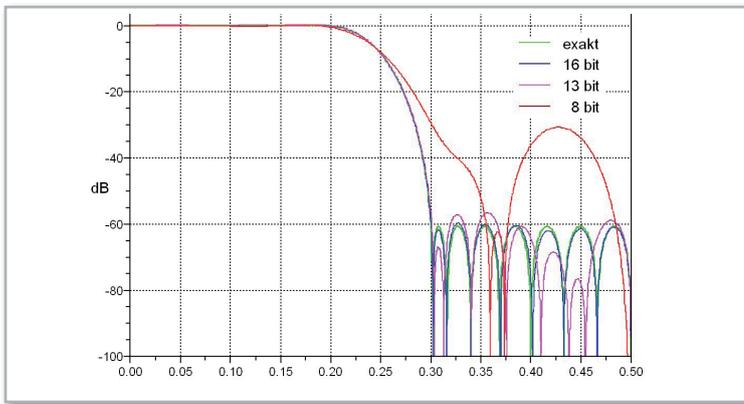


Abbildung 152: Betragsfrequenzgänge des FIR-Tiefpasses mit exakten sowie auf 16, 13 und 8 bit quantisierten Koeffizienten

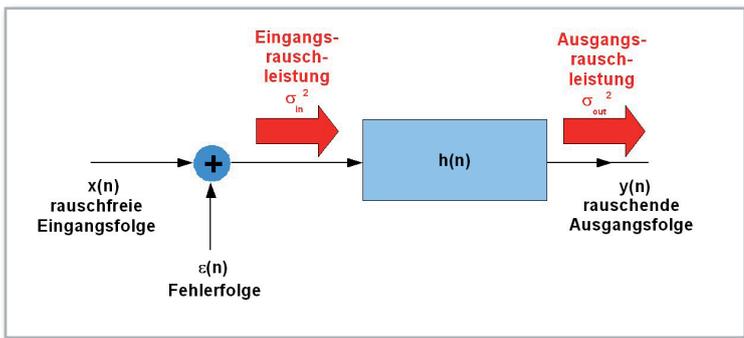


Abbildung 153: Eine Rauschfolge am Systemeingang erscheint mit der Impulsantwort gefaltet am Ausgang und überlagert das rauschfreie Ausgangssignal.

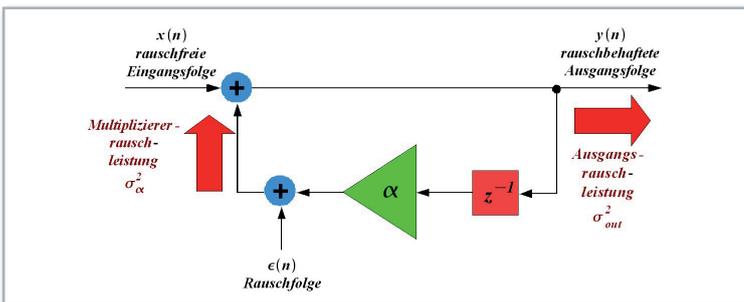


Abbildung 154: Der Rückkoppelkoeffizient α bestimmt die Stärke des Rauschens im Ausgangssignal.

Durchlassbereich aus (Abbildung 151). Dies äußert sich auch durch fehlende und verschobene Nullstellen im Nullstellenplan. Abbildung 152 fasst das Gesagte in Gestalt aller logarithmierten Beträge der Übertragungsfunktionen zusammen. Ganz offensichtlich ist für die Realisierung dieses FIR-Filters eine Koeffizientenwortlänge von 16 bit erforderlich.

Quantisierungsrauschen

Im 4. Teil dieser Reihe hatten wir den Fehler in der Signalrepräsentation nach einem A/D-Wandler infolge dessen endlicher Auflösung betrachtet. Unter der Voraussetzung, dass die Quantisierungsfehler $\epsilon(nT_s)$ unkorrelierte, im Intervall $[-\Delta/2, \Delta/2]$ gleichverteilte Zufallsgrößen sind und nicht von den wertekontinuierlichen Eingangsabstastwerten $x(nT_s)$ abhängen, ergibt sich für die Leistung des gleichspannungsfreien Quantisierungsfehlers $\epsilon(nT_s)$ nach Gleichung (168).

$$P_{\epsilon} = \sigma^2 = \frac{\Delta^2}{12} \quad P_{\epsilon}: \text{Quantisierungsrauschleistung} \quad (168)$$

σ^2 : Varianz der Fehlerfolge

Wir fassen noch einmal zusammen, dass unter den eben genannten Voraussetzungen die Varianz der Folgewerte aus der Differenz der tatsächlichen und der quantisierten Signalwerte nach Gleichung (76) (Fehlerfolge $\epsilon(n)$) die Rauschleistung bestimmt.

Ein solches rauschbehaftetes Eingangssignal kann wegen der vorausgesetzten Natur des Rundungsfehlers $\epsilon(n)$ als Summe aus der Fehlerfolge und der als amplitudenfehlerfrei angenommenen Eingangssignalfolge modelliert werden. Ein digitales System mit der Impulsantwort $h(n)$ entsprechend Abbildung 153 bewertet die an seinem Eingang anliegende Rauschleistung gemäß Gleichung (169).

$$\sigma_{out}^2 = \sigma_{in}^2 \cdot \sum_{m=0}^{\infty} h^2(m) \quad (169)$$

Diese Beziehung kann man auf die einfachste aller rekursiven Strukturen, wie sie Abbildung 154 zeigt (rückgekoppeltes System 1. Ordnung), anwenden. Bezüglich der Rauschleistung kann das Runden des Multiplikationsergebnisses in Analogie zum Rauschen eines A/D-Wandlers betrachtet werden. Deshalb ist ein Modell statthaft, in dem das Multipliziererauschen durch eine Rauschquelle am Multipliziererausgang repräsentiert wird. Diese „sieht“ wegen der Gleichwertigkeit der beiden Eingänge des Eingangssummierknotens die gleichen Übertragungseigenschaften wie ein Signal am Systemeingang. Für den Rauschanteil im multiplizierten Signal lautet deshalb ebenfalls die Impulsantwort $h(n)=\alpha^n$. Mit Gleichung (169) ergibt sich deshalb Gleichung (170). Das Modell funktioniert nicht mehr für den Grenzfall $\alpha=0$. Das Ausgangssignal des Multiplizierers ist dann konstant null und es entsteht somit auch kein Rundungsrauschen. Diesen Fall erfasst das Modell nicht korrekt. Die Struktur muss deshalb immer so weit angesteuert werden, dass am Multipliziererausgang Rundungsrauschen auftritt.

$$\sigma_{out}^2 = \sigma_{\alpha}^2 \cdot \sum_{m=0}^{\infty} h^2(m) = \sigma_{\alpha}^2 \cdot \sum_{m=0}^{\infty} \alpha^{2m} = \frac{\sigma_{\alpha}^2}{1-\alpha^2} \quad (170)$$

Nähert sich α dem Zahlenwert 1 an, wird die Rauschleistung am Ausgang immer größer. In der Darstellung des Frequenzübertragungsverhaltens des Systems in Abbildung 154 gemäß Gleichung (171) entspricht dies einem immer geringeren Abstand des Pols zum Einheitskreis der z-Ebene. Damit das Rauschen im Ausgangssignal einen bestimmten Wert nicht überschreitet, muss mit größeren Wortbreiten gearbeitet werden. Ganz allgemein kann gesagt werden, dass bei Annäherung des Pols an den Einheitskreis die Systemverstärkung zunimmt. Das erhöht das Quantisierungsrauschen am Ausgang und verschlechtert das Signal-Rausch-Verhältnis. Zudem kann es erforderlich werden, das Eingangssignal so weit abzuschwächen, dass kein Überlauf auftritt. Durch diese Skalierungsmaßnahme verschlechtert sich das Signal-Rausch-Verhältnis abermals.

$$y(n) = x(n) + \alpha \cdot y(n-1) \quad \text{Differenzgleichung zur Bestimmung der Ausgangsfolgewerte} \quad (171)$$

Differenzgleichung z-transformieren:
 $Y(z) = X(z) + \alpha \cdot Y(z) \cdot z^{-1}$
 daraus folgt:
 $H(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 - \alpha z^{-1}}$ Übertragungsfunktion

Ohne Zwischenrechnungen können wir das Rauschverhalten eines „Biquads“, wie er in Abbildung 86 dargestellt wurde (1. kanonische Direktstruktur), ermitteln, indem wir hinter jedem Multiplizierer eine additive Rauschquelle anordnen (Abbildung 155). Die Rauschleistungen im Eingangssignal (σ_x^2) und dieser fünf Quellen (σ_k^2 mit $k=1, \dots, 5$) ergeben am Ausgang eine Gesamtrauschleistung gemäß Gleichung (172). h_0 ist dabei die Impulsantwort vom Ein- zum Ausgang, die h_k gelten für die entsprechenden Rauschquellen an den Multipliziererausgängen.

$$\sigma_{out}^2 = \sigma_x^2 \cdot \sum_{m=0}^{\infty} h_0^2(m) + \sum_{k=1}^5 \left(\sigma_k^2 \cdot \sum_{l=0}^{\infty} h_k^2(l) \right) \quad (172)$$

An dieser Stelle sei erwähnt, dass rekursive Strukturen auch ohne Ansteuerung ($x(n)=0$) kleine periodische Ausgangssignale (Grenzyklen) abgeben können. Die Ursache liegt im nicht-linearen Verhalten des Systems, das infolge von Wortlängenbegrenzungen und Zahlenbereichsüberschreitungen entsteht. Anhand von Abbildung 156 wird die Entstehung eines Grenzyklus in dem rekursiven System 2. Ordnung nach Abbildung 154 beispielhaft demonstriert. Berücksichtigt man eine Rundung in Gleichung (171), kann dies zu einer zyklischen Störung am Ausgang führen. Um diese Behauptung zu beweisen, gehen wir von einer Zahlendarstellung nach Vorzeichen und Betrag mit einer Wortlänge von 4 bit gemäß Gleichung (162) aus. Damit repräsentieren die 3 bit rechts vom Binärpunkt den Zahlenbetrag (0, 1/8, 2/8, ... 7/8) und das bit links vom Binärpunkt das Vorzeichen.

In Gleichung (173) wird nun die Systemantwort auf einen Eingangsimpuls bei Runden des Multiplikationsergebnisses auf 4 Stellen ($R_4[\dots]$) abgeleitet.

$$y(n) = x(n) + \alpha \cdot y(n-1) \quad \text{Exakte Differenzgleichung}$$

$$\hat{y}(n) = x(n) + \alpha \cdot R_4[\hat{y}(n-1)] \quad \text{Differenzgleichung nach Runden auf 4 Stellen}$$

$$x(n) = \frac{7}{8} \cdot \delta(n) = 0.111 \cdot \delta(n) \quad \text{Eingangsimpuls zum Abtastzeitpunkt } n=0$$

Beispiel 1: $\alpha = \frac{1}{2} = 0.100$

$$\hat{y}(0) = +\frac{7}{8} = 0.111$$

$$\hat{y}(1) = R_4[\alpha \cdot \hat{y}(0)] = R_4[0.100 \cdot 0.111] = R_4[0.011100] = 0.100 = +\frac{1}{2}$$

$$\hat{y}(2) = R_4[\alpha \cdot \hat{y}(1)] = R_4[0.100 \cdot 0.100] = R_4[0.010000] = 0.010 = +\frac{1}{4}$$

$$\hat{y}(3) = R_4[\alpha \cdot \hat{y}(2)] = R_4[0.100 \cdot 0.010] = R_4[0.001000] = 0.001 = +\frac{1}{8}$$

$$\hat{y}(4) = R_4[\alpha \cdot \hat{y}(3)] = R_4[0.100 \cdot 0.001] = R_4[0.000100] = 0.001 = +\frac{1}{8} \quad \text{usw.}$$

Beispiel 2: $\alpha = -\frac{1}{2} = 1.100$

$$\hat{y}(0) = \frac{7}{8} = 0.111$$

$$\hat{y}(1) = R_4[\alpha \cdot \hat{y}(0)] = R_4[1.100 \cdot 0.111] = R_4[1.011100] = 1.100 = -\frac{1}{2}$$

$$\hat{y}(2) = R_4[\alpha \cdot \hat{y}(1)] = R_4[1.100 \cdot 1.100] = R_4[0.010000] = 0.010 = +\frac{1}{4}$$

$$\hat{y}(3) = R_4[\alpha \cdot \hat{y}(2)] = R_4[1.100 \cdot 0.010] = R_4[1.001000] = 1.001 = -\frac{1}{8}$$

$$\hat{y}(4) = R_4[\alpha \cdot \hat{y}(3)] = R_4[1.100 \cdot 1.001] = R_4[0.000100] = 0.001 = +\frac{1}{8}$$

$$\hat{y}(5) = R_4[\alpha \cdot \hat{y}(4)] = R_4[1.100 \cdot 0.001] = R_4[1.000100] = 1.001 = -\frac{1}{8} \quad \text{usw.}$$

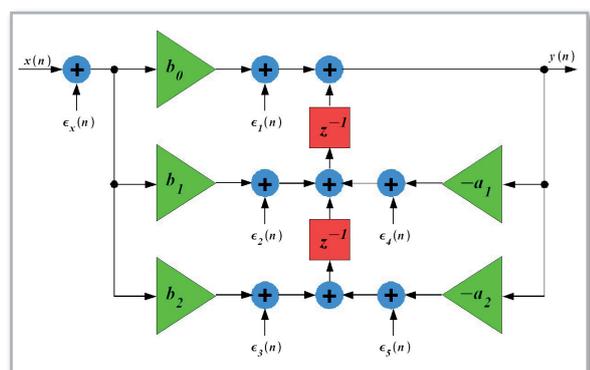


Abbildung 155: Bei Schaltungen mit mehreren Multiplizierern werden deren Rundungsrauschleistungen mit den Quadraten der jeweiligen Impulsantworten am Ausgang aufsummiert.

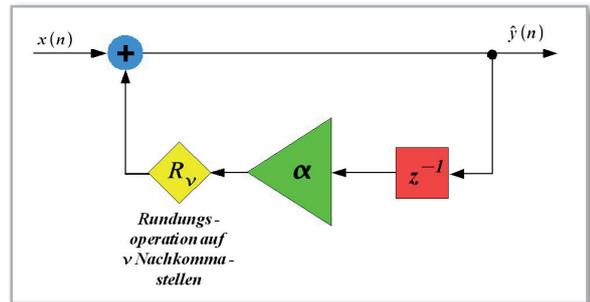


Abbildung 156: Rundungsoperationen können in rückgekoppelten Strukturen zu parasitären Schwingungen (Grenzyklen) führen.

Die Ausgangswerte $y(n)$ aus Gleichung (173), die sich für die Koeffizienten $\alpha=0,5$ und $\alpha=-0,5$ ergeben, sind in Abbildung 157 dargestellt. Man erkennt, dass für $\alpha=0,5$ die Antwort des Systems auf einen einzigen Eingangsimpuls der Höhe $7/8$ zum Zeitpunkt $n=0$ ab $n=2$ die Werte der Ausgangsfolge $y(n)$ stabil auf $1/8$ bleiben und nicht auf null abklingen. Für $\alpha=-0,5$ alternieren die Ausgangswerte ab $n=3$ zwischen $-1/8$ und $+1/8$. Diese nicht abklingenden Ausgangssignale in Abwesenheit eines Eingangssignals nennt man Grenzyklen. Sie hängen in diesem Beispiel offensichtlich von dem Wert des Rückkoppelkoeffizienten α ab und manifestieren sich als eine Art „Grundrauschen“.

Aber auch die Art, wie das Multiplikationsergebnis auf 4 Stellen zurückgeführt wird, spielt eine Rolle. Schneidet man in Gleichung (173) ab, statt zu runden, ergibt sich Gleichung (174). Abbildung 158 stellt die auf null abklingenden grenzyklusfreien Ausgangsfolgen dar. Das veränderte Verhalten ist auch plausibel, denn da, wo der Wert 0.0001 auf 0.001 aufgerundet wurde, also in den darstellbaren Zahlenbereich zurückkehrt, verbleibt jetzt beim Abschneiden 0.000.

zyklen. Nichtrekursive Strukturen sind wegen ihrer endlichen Impulsantwort prinzipiell grenzyklusfrei, weil alle Ausgangsfolgenwerte nach $M+1$ Takten null werden, nachdem die Eingangsfolgenwerte null sind und bleiben.

Im nächsten Teil dieser Reihe werden wir uns mit den Entwurfsverfahren für digitale Filter beschäftigen. **ELV**

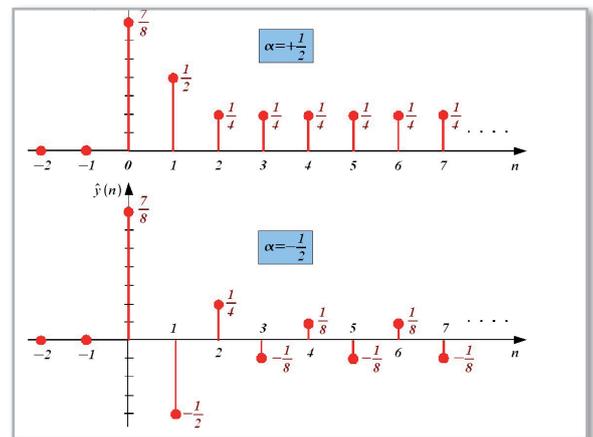


Abbildung 157: Beispiel für Grenzyklen bei Rundung des Multiplikationsergebnisses

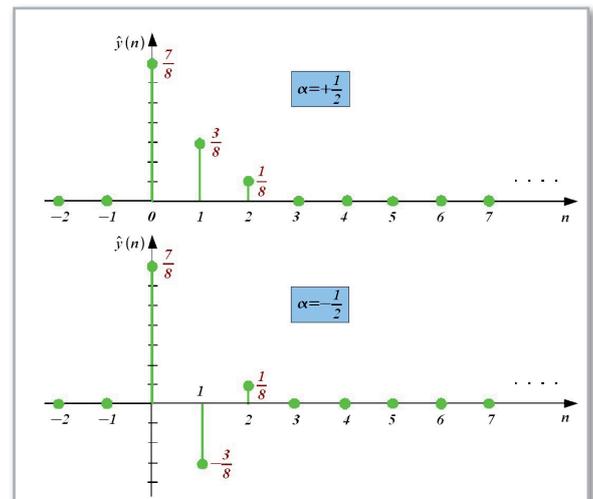


Abbildung 158: Die Grenzyklen verschwinden, wenn abgeschnitten statt gerundet wird.

(174)

$$\begin{aligned} y(n) &= x(n) + \alpha \cdot y(n-1) && \text{Exakte Differenzengleichung} \\ \hat{y}(n) &= x(n) + \alpha \cdot A_4[\hat{y}(n-1)] && \text{Differenzengleichung nach Abschneiden auf 4 Stellen} \\ x(n) &= \frac{7}{8} \cdot \delta(n) = 0.111 \cdot \delta(n) && \text{Eingangsimpuls zum Abtastzeitpunkt } n=0 \end{aligned}$$

Beispiel 1: $\alpha = \frac{1}{2} = 0.100$

$$\hat{y}(0) = +\frac{7}{8} = 0.111$$

$$\hat{y}(1) = A_4[\alpha \cdot \hat{y}(0)] = A_4[0.100 \cdot 0.111] = A_4[0.011100] = 0.011 = +\frac{3}{8}$$

$$\hat{y}(2) = A_4[\alpha \cdot \hat{y}(1)] = A_4[0.100 \cdot 0.011] = A_4[0.001100] = 0.001 = +\frac{1}{8}$$

$$\hat{y}(3) = R_4[\alpha \cdot \hat{y}(2)] = A_4[0.100 \cdot 0.001] = A_4[0.000100] = 0.000 = +0$$

$$\hat{y}(4) = R_4[\alpha \cdot \hat{y}(3)] = A_4[0.100 \cdot 0.000] = A_4[0.000000] = 0.000 = +0 \quad \text{usw.}$$

Beispiel 2: $\alpha = -\frac{1}{2} = 1.100$

$$\hat{y}(0) = \frac{7}{8} = 0.111$$

$$\hat{y}(1) = A_4[\alpha \cdot \hat{y}(0)] = A_4[1.100 \cdot 0.111] = A_4[1.011100] = 1.011 = -\frac{3}{8}$$

$$\hat{y}(2) = A_4[\alpha \cdot \hat{y}(1)] = A_4[1.100 \cdot 1.011] = A_4[1.001100] = 0.001 = +\frac{1}{8}$$

$$\hat{y}(3) = A_4[\alpha \cdot \hat{y}(2)] = A_4[1.100 \cdot 0.001] = A_4[1.000100] = 1.000 = -0$$

$$\hat{y}(4) = A_4[\alpha \cdot \hat{y}(3)] = A_4[1.100 \cdot 1.000] = A_4[1.000000] = 0.000 = +0 \quad \text{usw.}$$

Wie man sich leicht überzeugen kann, hat auch die Stärke der Eingangsanregung einen Einfluss auf das Grenzyklusverhalten. Setzt man zum Beispiel in Gleichung (173) als Eingangssignal $x(n)=0,25$ $\delta(n)=0,010$, sind die Ausgangswerte $y(n)$ bereits ab $n=2$ auf null zurückgegangen ($\hat{y}(0)=0,25$, $\hat{y}(1)=0,125$ bzw. $-0,125$, $\hat{y}(n>1)=0$).

Man kann an diesem einfachen Beispiel bereits erahnen, dass die exakte Analyse des Grenzyklusverhaltens komplexerer Systeme kompliziert ist. Allgemein gilt aber: Durch Vergrößerung der Wortlänge, d. h. Erhöhung der Rechengenauigkeit und Vermeidung von Addiererüberläufen, lassen sich Grenzyklen verkleinern oder vermeiden. Bei komplizierten Strukturen mit rekursiven Anteilen liefert nur die Simulation auf Bitebene sichere Erkenntnisse über mögliche Grenz-