

Empfang und Auswertung intensitätsmodulierter optischer Signale mittels Photonic-Mixer-Device (PMD) in Applikationen der Messtechnik und Kommunikation

Vom Fachbereich Elektrotechnik und Informatik
der Universität Siegen
zur Erlangung des akademischen Grades

**Doktor der Ingenieurwissenschaften
(Dr. Ing.)**

genehmigte Dissertation

vorgelegt von
Dipl.-Ing. Holger Heß
geb. am 03.01.1972 in Halle/S.

- | | |
|---|--------------------------------|
| 1. Gutachter: | Prof. Dr.-Ing. Rudolf Schwarte |
| 2. Gutachter: | Prof. Dr.-Ing. Elmar Griese |
| 3. Vorsitzender der Prüfungskommission: | Prof. Dr.-Ing. Hubert Roth |

Tag der mündlichen Prüfung: 20. Dezember 2006

Inhalt	I
Abstract	V
1 Einleitung	1
1.1 Motivation	1
1.2 Zielsetzung	2
1.3 Gliederung	3
2 Systemtheorie der multidimensionalen Sensorik	5
2.1 Hintergründe zur Ausbreitung von Raum-Zeit-Signalen	5
2.2 Informationsbegriff der modernen Sensorik	10
2.3 Information in der Nachrichtentechnik	14
2.4 Grundprinzipien der Signaltheorie	15
2.2.1 Signale und das Fourier-Prinzip	15
2.2.2 Signale und das Prinzip der Unschärfe	16
2.2.3 Signale und das Prinzip der Symmetrie	19
2.5 Grundlagen der Systemtheorie	20
2.5.1 Eindimensionale Systeme	20
2.5.2 Erweiterung zur mehrdimensionalen Systemtheorie	30
3 Systemkomponenten der optischen Sensorik und Signalübertragung	35
3.1 Elektrooptischer Effekt	35
3.2 Prozess der Photonenemission	40
3.3 Prozess der Photonenabsorption	42
3.4 Modulierbare Halbleitersendeelemente	43
3.4.1 Lichtemitterdioden	43
3.4.2 Laserdioden	44
3.4.3 VCSELs	49
3.4.4 Superlumineszenzdioden	51
3.5 Halbleiterempfangselemente	51
3.5.1 Halbleiterphotodioden	51
3.5.2 Avalanche-Photodioden	57
3.5.3 Photogate-PMD-Strukturen als quasioptische Mischer	60

3.5.4 MSM-PMD-Strukturen als quasioptische Mischer	79
3.6 Optischer Korrelationsempfang	82
4 Phasenauswertung intensitätsmodulierter Signale der optischen Messtechnik mittels Phase-Shift-Interferometrie	87
4.1 Prinzip der Phase-Shift-Interferometrie	88
4.2 Auswertung von Phase-Shift-Interferogrammen	90
4.2.1 Phasenbestimmung durch Phase-Shift-Interferometrie	90
4.2.2 Systematische Approximationsfehler in Phase-Shift-Verfahren	92
4.2.3 Statistische Approximationsfehler in Phase-Shift-Verfahren	94
4.3 Auswahl geeigneter Phase-Shift-Algorithmen	97
4.3.1 Korrektur eines systematischen Phasenoffsets durch die Wahl geeigneter Gewichte	97
4.3.2 Minimierung des Phasenfehlers für bekannte Oberwellen	99
4.3.3 Minimierung des Einflusses verrauschter Messwerte	103
4.4 Mehrdimensionale Phase-Shift-Interferometrie	107
4.4.1 Laterale Abtastung zweidimensionaler Bildaufnehmer	107
4.4.2 Interferogramme im Ortsfrequenzbereich	117
4.5 Probleme optischer Phasenmessverfahren	121
4.5.1 Modenstabilität in Halbleiterlaserdioden	121
4.5.2 Impedanzstabilität des quasioptischen Mischers	136
4.6 Praktische Anwendung der optisch intensitätsmodulierten Phase-Shift-Interferometrie	140
4.6.1 Eindimensionale TOF-Entfernungsmessung	140
4.6.2 Zweidimensionale TOF-Entfernungsmessung und deren Informationsgewinn	148
4.6.3 Objekterkennung in der interferometrischen 3D-Messtechnik	156
4.6.4 Zeitaufgelöste Spektroskopie	158
4.7 In-Phase-Quadrature-Mischer als Konkurrenz zur klassischen Phase-Shift-Interferometrie	166
4.7.1 Grundprinzip der Phasenbestimmung mittels IQ-Demodulation	166
4.7.2 IQ-Demodulation eines optischen Signals mittels PMD	171

4.8 Fazit	174
5 Phasenauswertung intensitätsmodulierter Chirp-Signale in der optischen Kommunikationstechnik	177
5.1 Bandspreizung und der Begriff des Systemgewinns	179
5.2 Spread-Spectrum-Verfahren in der Kommunikationstechnik	180
5.3 Systemtheoretische Betrachtung eines Chirp-Signals	187
5.4 Dynamische Kommunikationsverfahren mittels variabler Chirp-Signale	190
5.4.1 Bedeutung der Phasenmodulation	190
5.4.2 Bedeutung der Amplitudenmodulation	195
5.4.3 Bedeutung der Frequenzmodulation	196
5.4.4 Systemtheoretische Betrachtung des chirpbasierten Mehrfachzugriffs	197
5.4.5 Bedeutung der mehrdimensionalen Parametervariation	200
5.5 Optimierte Chirp-Signale	205
5.6 Korrelationsempfang optischer Chirp-Signale mittels PMD	211
5.7 Fazit	216
6 Zusammenfassung	219
7 Anhang	223
8 Verzeichnis der verwendeten Formelzeichen	232

Abstract

The human thinking is based on three successive fundamental steps: perception, abstraction and interpretation. One of the most interesting subjects in the field of human perception and the scientific world is definitively the spatial sense, which is illustrated by many papers, from the ancient world till this day in all different sciences. In nature two different principles are known: ultrasonic distance measurement and stereo vision. The first measures the time of a travelling signal, starting from the emitting animal to the reflecting objects in his environment back to animal's receiver. The second uses the triangulation effect between two spatial different sensors, which allows an estimation of the distance to the environmental objects.

After a general introduction (Chapter 2 – 3) the first main part of this dissertation is engaged in the comprehensive theme of the technical realisation of depth perception (Chapter 4). The second main part (Chapter 5) works on another interesting field in research and technology: the optical spread spectrum communication. Both topics seem to have little relation, but they are merged in their multidimensional approach and their underlying principle of correlation reception.

A more detailed overview is given in the following:

After the introduction to the both subjects in Chapter 1, the fundamental terms of space and time are discussed at the beginning of Chapter 2, comparing the opinions and interpretations of the most important scientists in history. Afterwards some elementary principles of modern signal and system theory are illustrated to provide the understanding of the presented work. Chapter 3 gives a survey of the components which are utilisable in optical time of flight measurement systems. Particular attention is drawn to the novel photonic mixer device, which fuses the receiving and processing of optical signals into a smart device. Therein a mixing of the optical signal and a reference takes place during the electro-optical generation of electrons. Based on this unique feature the principle of a correlation reception is declared. Chapter 4 gives detailed facts of the presented distance measurement system, which is known as optical time of flight measurement. Therein a lot of technical problems in electro-optical correlation reception are discussed in detail, starting from lasers as one of the most

effective light sources over the novel unique optical receiver named PMD to the optimisation of distance calculation algorithms. Based on the one-dimensional TOF-distance measurement a new 3D-camera is presented and discussed in comparison to approaches using the similarities in several conventional 2D-pictures. Finally a second application similar to the distance measurement is presented shortly: the fluorescent lifetime measurement. After that the suggested spread-spectrum communication is explained in Chapter 5. This multiple access system bases on the use of chirp signals, which allows an effective management of the resource bandwidth. A discussion of all chirp parameters in the time and frequency domain demonstrates their efficiency. After that an experimental setup, also using the novel optical receiver PMD, is presented with first results. At least a summery of the main points of this work and future aspects of the PMD technology in 3D-camera systems and optical communication is given in Chapter 6.

1 Einleitung

1.1 Motivation

Der Sehapparat des Menschen ist in der Lage, sein Umfeld mit relativ hoher Sicherheit dreidimensional zu erfassen. In der Technik jedoch steht trotz intensiver Entwicklungsarbeit eine solche Sensorik noch nicht zur Verfügung. Erste Versuche zur Realisierung orientierten sich sehr stark an diesem menschlichen Sehen, wobei hier jedoch eine wesentliche Tatsache vernachlässigt wurde: Das sichere Orientieren, Navigieren und Interagieren des Menschen in seiner dreidimensionalen Umwelt basiert nicht allein auf dem 3D-Sehen mittels des stereoskopischen Sehapparates, sondern ist ein Ergebnis eines komplexen Abgleichs der optischen Umfeldwahrnehmung mit anderen bereits gesammelten und gespeicherten Sinneserfahrungen. Neben den Augen bildet hier beispielsweise auch der Gleichgewichtssinn eine erhebliche Rolle, da er das Koordinatensystem des Umfelds mit dem des im Gehirn erstellten Abbildes ständig abgleicht. Eine solche Verschmelzung mehrerer Sinne bezeichnet man im technischen Bereich allgemein als Sensorfusion, die ein sehr anspruchsvolles Aufgabengebiet der Ingenieurwissenschaften darstellt.

Wie in der vorliegenden Ausarbeitung noch deutlich erläutert wird, kranken die am menschlichen Sehen orientierten Verfahren, die unter dem Begriff Stereoskopie zusammengefasst sind, an deutlichen Nachteilen. Einen völlig neuartigen Ansatz zur dreidimensionalen Umfelderkennung stellen die Laufzeitverfahren (engl. time of flight – TOF) dar, die ebenso ihr Pendant in der Natur finden. So orientieren sich Fledermäuse an der Laufzeit eines ausgesandten Ultraschallsignals, das direkt vom gesamten Signalweg abhängig ist, vom Sender über das reflektierende Objekt bis hin zum Empfänger. Ein solches Verfahren ist natürlich auch mit optischen Signalen denkbar, jedoch gestaltete sich deren technische Umsetzung bisher sehr schwierig. Mit der Entwicklung der Photomischdetektoren (engl. Photonic Mixer Device - PMD) steht nun jedoch ein robuster optischer Empfänger zur Verfügung, der die einfache und preiswerte Realisierung einer optischen 3D-TOF-Kamera erlaubt.

Mit dem Potential der PMD-Technologie besteht nun eine hohe Motivation, eine serienreife 3D-TOF-Kamera aufzubauen. Dabei sind jedoch noch viele technische Probleme, die sich aus der Vielfalt der unterschiedlichen Umgebungsbedingungen ergeben, zu hinterfragen und zu lösen. Nur so kann ein wirklich robustes Sensorsystem realisiert werden, das den unterschiedlichsten Wünschen der Anwender gewachsen ist.

Hierzu ist eine ausführliche systemtheoretische Betrachtung notwendig, um auftretende Effekte in ihrer Ursache einer Systemgruppe oder gar einem einzelnen Bauteil zuordnen zu können. Eine zweidimensionale Anordnung mehrerer Photomischdetektoren als Bildaufnehmer mit integrierter Signalverarbeitung bringt jedoch auch eine enorme Informationsdichte mit sich, die sich mit der klassischen eindimensionalen Systemtheorie nur noch unzureichend bearbeiten lässt. Somit besteht hier ebenso eine große Dringlichkeit, systemtheoretische Grundlagen für diese Technologie zu schaffen.

1.2 Zielsetzung

Der rein funktionelle Nachweis einer neuen Technologie reicht noch nicht, um sie dauerhaft in den unterschiedlichen Anwendungen zu etablieren. Aufbauend auf ihrer prinzipiellen Funktion müssen weitere Grundlagen geschaffen werden, die den störungsfreien Betrieb im alltäglichen Einsatz innerhalb einer bestimmten Applikation garantieren. Ziel dieser Arbeit ist es deshalb, den robusten Betrieb eines einzelnen PMD-Pixels in verschiedenen Applikationen sicher zu stellen. Die hierbei auftretenden Probleme sollen derart beschrieben werden, dass darauf aufbauend eine zuverlässige Implementierung einer PMD-Matrix in den verschiedensten Anwendungen möglichst risikoarm durchführbar ist. Hierbei muss deutlich darauf hingewiesen werden, dass - aufbauend auf den bisher geschaffenen Grundlagen der PMD-Technologie, die im wesentlichen aus halbleitertechnologischen Betrachtungen bestehen - nun ein umfassendes Verständnis für das Gesamtsystem entwickelt werden muss. Deshalb soll hier ein besonderes Augenmerk auf eine umfassende systemtheoretische Beschreibung gelegt werden. Innerhalb dieser Arbeit sind deshalb für das verwendete interferometrische Entfernungsmessverfahren erste Ansätze einer mehrdimensionalen Systemtheorie darzustellen. Dabei sind im Weiteren zwei Anwendungsgebiete zu betrachten, die in der Optoelektronik zurzeit hochaktuell sind, die Phasenmessung eines optischen Signals zur Laufzeitmessung einerseits und die optische Kommunikation andererseits.

1.3 Gliederung

Ausgehend von der im Abschnitt 1.2 formulierten Aufgabenstellung werden im Kapitel 2 theoretische Grundlagen formuliert, die als Basis für die weiteren wissenschaftlichen Betrachtungen unabdingbar sind. Da in den beiden zu bearbeitenden Anwendungsgebieten der optischen Laufzeitmessung und Kommunikation die Signalausbreitung in Raum und Zeit im Zentrum aller Betrachtungen steht, werden zuerst diese beiden physikalischen Grundkategorien hinreichend diskutiert. Darauf aufbauend wird der Begriff der Information im nachrichtentechnischem Sinne aufbereitet. Daran schließen sich als weitere theoretische und methodologische Grundlagen dieser Arbeit die wesentlichen Grundprinzipien der Signaltheorie an, das Fourier-Prinzip, das Prinzip der Unschärfe und das der Symmetrie. Für die weitere Betrachtung der PMD-Technologie innerhalb nachrichtentechnischer Systeme sind noch systemtheoretische Grundlagen notwendig, die dieses Kapitel abrunden.

Kapitel 3 dient nun der Betrachtung der wichtigsten Systemkomponenten, wie sie in PMD-basierten Messsystemen zum Einsatz kommen. Hierzu werden zuerst kurz die Halbleiterphysikalischen Grundlagen des elektrooptischen Effekts erläutert, mit dessen Hilfe sich die Prozesse der Photonenemission und -absorption erklären lassen. Mit diesem Verständnis werden nun verschiedene modulierbare optische Emittoren betrachtet. Die anschließende Analyse optischer Empfänger beschränkt sich absichtlich nicht allein auf die PMD-Technologie, sondern vergleichend werden hier ebenso Halbleiter- und Avalanche-Photodioden einbezogen. Abschließend wird das Hauptmerkmal einer PMD-Struktur, das der inhärenten quasioptischen Signalverarbeitung, beschrieben, wobei hier das weitere Augenmerk auf dem optischen Korellationsempfang liegt.

Das umfangreiche Kapitel 4 widmet sich einem der beiden Hauptthemen dieser Arbeit, der Phasenauswertung eines intensitätsmodulierten optischen Signals in der Messtechnik. Hierbei werden zuerst die Grundlagen und ausführliche Fehlerbetrachtungen der Phaseshift-Interferometrie präsentiert, so dass darauf aufbauend für jede gewünschte Applikation der optimale Algorithmus gewählt werden kann. Danach erfolgt eine Erweiterung der theoretischen Grundlagen für die mehrdimensionale Phaseshift-Interferometrie und eine ausführliche Beschreibung zweier wesentlicher Fehlerquellen, die Modenstabilität eines optischen Senders auf Basis eines Halbleiterlasers und die Impedanzstabilität eines PMD-Empfangelements. Anschließend werden verschiedene messtechnische Applikationen der PMD-Technik vorgestellt, die die Anwendungsvielfalt der PMD-Technologie unterstreichen sollen.

Abschließend wird zur bisher verwendeten klassischen Phaseshift-Interferometrie ein weiteres Signalverarbeitungskonzept angeboten, das Konzept der In-Phase-Quadrature-(IQ-) Demodulation.

Das zweite Hauptthema dieser Arbeit, die Phasenauswertung eines intensitätsmodulierten optischen Signals in der Kommunikationstechnik, steht im Zentrum des Kapitels 5. Hierbei wird nach einer Gegenüberstellung der grundsätzlichen Bandspreizverfahren ein Ansatz beschrieben, der auf der Verwendung sogenannter Chirp-Signale basiert. Nach einer systemtheoretischen Betrachtung geht dieses Kapitel auf die drei wesentlichen Variationsmöglichkeiten eines Chirp-Signals innerhalb des vorgeschlagenen Verfahrens ein, auf die Phasen-, Frequenz- und Amplitudenmodulation. Hierbei wird die Bedeutung der mehrdimensionalen Variation dieser Parameter noch weiter vertieft. Darüber hinaus werden so genannte optimierte Chirp-Signale systemtheoretisch diskutiert und abschließend der realisierte Versuchsaufbau und erste Messergebnisse des beschriebenen Kommunikationsverfahrens präsentiert.

Eine abschließende Zusammenfassung und Bewertung dieser Arbeit erfolgt im Kapitel 6. Dabei wird ein kurzer Ausblick auf zukünftige Arbeiten gegeben.

2 Systemtheorie der multidimensionalen Sensorik

2.1 Hintergründe zur Ausbreitung von Raum-Zeit-Signalen

Die thematische Ansiedlung dieser Arbeit im Bereich der dreidimensionalen Umwelterfassung einerseits und der optischen Freiraumkommunikation andererseits verlangt eine Auseinandersetzung mit ihren elementaren Größen, den Begriffen von Raum und Zeit. Je nach wissenschaftlichem Fachgebiet ist das Verständnis ihrer Eigenschaften durchaus völlig verschieden, was in einem anwendungsorientierten und fachübergreifenden Bereich wie der Ingenieurwissenschaft zu einigen Problemen führen kann. Die im Ingenieursbereich meist verankerte klassische Betrachtung der Begriffe von Raum und Zeit stößt z.B. schon bei der Signalübertragung über große Entfernungen an ihre Grenzen. Betrachtet man beispielsweise das Global Positioning System (GPS), dann unterliegt die Signalstrecke der in diesem Fall gegenläufigen Effekte der Zeitdilatation durch den Gravitationseinfluss einerseits und durch die relative Bewegung von Sender und Empfänger zueinander andererseits. Diese relativistischen Effekte lassen sich mit der klassischen Definition der Begriffe Raum und Zeit nicht mehr vereinen.

Die klassische Beschreibung der räumlichen Ausbreitung eines Signals durch die Signallaufzeit t_L ergibt sich sehr einfach aus deren Proportionalität zu der mit Lichtgeschwindigkeit c zurückgelegten Wegstrecke s :

$$t_L = \frac{s}{c}. \quad \text{Gl. 2.1}$$

Wie zuvor schon angedeutet ist diese Betrachtung nur eingeschränkt anwendbar, denn sie bedingt, dass innerhalb der Wegstrecke s ein konstantes Gravitationsfeld vorliegt. Diese Annahme gilt jedoch nur näherungsweise auf relativ kurzen Strecken, da bei ihnen die Änderung des Gravitationsfeldes als verschwindend gering betrachtet werden kann. Bei Signalwegen über große Distanzen jedoch sind die Änderungen im Gravitationsfeld mit zu berücksichtigen. Mit einem vorhandenen Gravitationspotential ϕ ergibt sich so eine Zeitdilatation von

$$\Delta t = \Delta t_0 \cdot \left(1 + \frac{\phi}{c^2}\right), \quad \text{Gl. 2.2}$$

womit die Zeit in einem zu einer gegebenen Masse entfernten ruhenden Punkt schneller verläuft als in deren Nähe. Der Einfluss der Zeitdilatation aufgrund einer relativen Bewegung von Sender und Empfänger ergibt sich aus der Bewegungsgeschwindigkeit v :

$$\Delta t = \Delta t_0 \cdot \sqrt{1 - \left(\frac{v^2}{c^2}\right)}. \quad \text{Gl. 2.3}$$

Für das bewegte Objekt vergeht also in Relation zum Beobachter die Zeit langsamer. Aber auch die relativistischen Betrachtungen von Raum und Zeit werden in der anwendungsorientierten Ingenieursarbeit nicht genügen, da die Signaldetektion oft quantenmechanischen Prozessen gegenübersteht. Ein Beispiel ist hierfür die hochgenaue Vermessung des Abstandes zwischen Erde und Mond. Die klassische Bestimmung dieser Entfernung R

$$R = \frac{t_l \cdot c}{2} \quad \text{Gl. 2.4}$$

nach dem Prinzip der Echolaufzeit mittels ausgesendeter Laserpulse führt zu einer Differenz zwischen tatsächlichem und berechnetem Entfernungswert, da hier eine Krümmung des Signalweges durch die verschiedenen Gravitationspotentiale von Erde, Mond und Sonne nicht berücksichtigt werden kann. Für die exakte Berechnung ist also eine relativistische Betrachtung der Raumzeit unabdingbar. Bleibt man bei der Betrachtung dieses Beispiels, dann steht man bei der Signaldetektion des reflektierten Laserimpulses dem quantenmechanischen Prozess des elektrooptischen Effekts gegenüber, welcher sich nur noch statistisch beschreiben lässt. Die ingenieurwissenschaftliche Betrachtung eines Mess- bzw. Kommunikationssystems auf Basis optischer Signale macht also ein Verständnis für die Begriffe Raum und Zeit im klassischen, relativistischen und quantentheoretischen Sinne notwendig.

Die klassischen Begriffsdefinitionen wurden im Wesentlichen durch die wissenschaftlichen Arbeiten von Isaac Newton (1643-1727) geprägt. Seine Entdeckung des Gravitationsgesetzes und der drei Bewegungsgesetze zur Trägheit, zur Proportionalität von Kraft und Beschleunigung und zur Gleichheit von Wirkung und Gegenwirkung hatten erheblichen Einfluß auf die Wahrnehmung und Vorhersehbarkeit

unserer menschlichen Umwelt. So entwickelt Newton selbst schon in seinen Ausführungen Vorstellungen über den Charakter von Zeit und Raum.¹ Nach Newton stellen Raum und Zeit etwas für sich eigenständiges dar, was nichts anderes bedeutet, als dass Raum und Zeit existieren könnten, auch wenn es nichts Materielles in ihnen gäbe. Für ihn floss die Zeit gleichförmig und linear und ohne Rücksicht auf irgendwelche Dinge. In Newtons Physik waren Zeit und Raum damit völlig voneinander getrennte und damit unabhängige Kategorien. Die von ihm geführte Schlussfolgerung über den damit ebenso absoluten Charakter der Bewegung bedeutet jedoch nicht, dass eine Beschreibung relativ zu anderen Körpern unmöglich ist. Diese Vorstellung fand natürlich seine Kritiker, wie z.B. durch Immanuel Kant (1724-1804).² Seine Argumentation basiert auf der Existenz des Menschen als beobachtendes Subjekt. Verzichtet man nun auf dieses Subjekt als Bedingung der Anschauung, so bedeutet die Vorstellung vom Raume gar nichts. Nur durch den Raum als Anschauungsform ist es möglich, dass Dinge für uns äußere Gegenstände sind. Darauf aufbauend ist für Kant die Zeit an sich ebenso nicht existent, sondern reine Anschauung, weil sich die Verhältnisse der Zeit nur an einer äußeren Anschauung ausdrücken lassen. Sie ist also ebenso eine subjektive Bedingung unserer menschlichen Anschauung. Gottfried Wilhelm Freiherr von Leibniz (1646 – 1716), der ebenso als Kritiker Newtons zu nennen ist, betrachtet den Raum als eine Anordnung des Existierenden und Zeit ist die Anordnung des nacheinander Existierenden. Damit bedingen sich Raum und Zeit einander und unterliegen einem relationalem Zusammenhang, was sich mit dem absoluten Verständnis Newtons nicht vereinen lässt.

Die am Anfang des 20. Jahrhunderts aufgekommene relativistische Betrachtung führte zu einem entscheidenden Brückenschlag zwischen den Begriffen von Raum und Zeit. Die Untersuchungen zur Ausbreitung des Lichtes führte Albert Einstein (1879-1955) im Jahr 1905 zu seiner Entwicklung der speziellen Relativitätstheorie. Seine Überlegungen baut er auf der Darstellung von so genannten Inertialsystemen auf. Hierbei wird ein Bezugssystem als Inertialsystem verstanden, wenn die Bahnen von unbeeinflussten punktuellen Probekörpern in diesem System geradlinig sind. Basis seiner Ausführung bildet die Annahme, dass alle Inertialsysteme gleichwertig sind und sich somit relativ

¹ vgl. Isaac Newton; *Philosophiae naturalis principia mathematica*; in deutscher Ausgabe: *Mathematische Grundlagen der Naturphilosophie*, Felix Meiner Verlag, Hamburg 1988

² vgl. Immanuel Kant; *Kritik der reinen Vernunft*

zueinander mit konstanter Geschwindigkeit bewegen. Des Weiteren legt er das Prinzip der Konstanz der Lichtgeschwindigkeit zugrunde, was bedeutet, dass sich das Licht in jedem Inertialsystem in gleicher Weise ausbreitet. Hier findet sich das eigentlich Innovative, insofern Einstein die Beziehung zwischen Raum, Zeit und Bewegung ohne die Vorstellung einer absoluten Zeit formulieren kann. Er zeigte, dass sich Anordnungen in einem Inertialsystem nur dann forminvariant verhalten, wenn sowohl das Längenmaß als auch der Schritt der Zeitmessung von der Relativgeschwindigkeit zwischen den betrachteten Systemen abhängen. Die Verknüpfung des Bewegungszustandes mit Raum- und Zeitmaßen führt zu einer vierdimensionalen Raumzeit mit gewissem absoluten Charakter, wogegen der dreidimensionale Raum und die eindimensionale Zeit nur noch relativ zu begreifen sind.

Entgegen ihrer herausragenden Bedeutung erkannte Einstein in seiner speziellen Relativitätstheorie eine gewisse Unvollkommenheit, da sie sich ausschließlich auf die beschriebenen Inertialsysteme beschränkt und damit beliebig beschleunigte Nicht-Inertialsysteme ausschließt. Ebenso können mit ihr gravitative Vorgänge nicht erfasst werden. Eine Auflösung dieser Schranken gelang Einstein 1916 mit der Formulierung der Allgemeinen Relativitätstheorie. Sie beruht auf der Ansicht, dass einerseits alle Körper gleichmäßig schnell und unabhängig von ihren Eigenschaften, wie z.B. der Masse, in einem gegebenen Gravitationsfeld fallen, und dass andererseits in einem konstanten Gravitationsfeld für einen frei fallenden Beobachter die identischen physikalischen Gesetzmäßigkeiten gelten, wie für einen ruhenden. Diese Annahmen führen zu dem essentiellen Einsteinschen Äquivalenzprinzip:

„An jedem Punkt der Raumzeit mit einem beliebigen Gravitationsfeld ist es stets möglich, ein lokales Inertialsystem so zu wählen, dass in einer hinreichend kleinen Umgebung des betrachteten Punktes (innerhalb welcher die Gravitation als konstant betrachtet werden kann) die Naturgesetze dieselbe Form annehmen, wie in einem unbeschleunigten kartesischen Koordinatensystem ohne Gravitationsfeld.“³

Das Materielle bzw. die Energie strukturiert also die Raumzeit, denn Masse krümmt die Raumzeit. Entgegen der Vorstellung der Gravitation als Kraftwirkung zwischen Materiellem besteht diese in der Krümmung der Raumzeit. Krümmung und Gravitation

³ nach S. Weinberg; Gravitation and Cosmology: Principles and Application of the General Theory of Relativity, J. Wiley & Sons, New York 1972, S. 68

sind somit gleichzusetzen. Obwohl die Allgemeine Relativitätstheorie eine Beziehung zwischen Materie und Raumzeit im makrophysikalischen Bereich aufstellt, ist die grundlegende Konzeption des Verhältnisses von Raum, Zeit und Materie weiterhin ungeklärt, denn auch sie stößt in der mikrophysikalischen Betrachtung von Teilchen an ihre Grenzen.

Die quantentheoretische Diskussion der Begriffe Raum und Zeit wurde entscheidend durch Nils Bohr (1885-1962) und Werner von Heisenberg (1901-1976) und der von ihnen entwickelten Quantenmechanik geprägt. Entgegen der klassisch physikalischen Sichtweise absoluter und damit funktional unabhängiger Elementargrößen, in der eine solche Größe keinerlei Einschränkung erfährt, löst die Quantenphysik diese Betrachtungsweise auf, indem man beliebige Werte funktional unabhängiger Eigenschaften nicht zulässt. Hierfür steht der Begriff der inkompatiblen Eigenschaften. Das bekannteste Beispiel hierfür stammt aus der Quantenmechanik anhand der Betrachtung von Ort und Impuls. In der Quantenmechanik gibt es keinen Zustand, in dem Ort und Impuls einen bestimmten Wert besitzen. Nähert sich der Wert der einen Eigenschaft dem definiten Wert an, ist der Wert der anderen so unbestimmter. Es gibt keinen Zustand, in dem man beliebig genau lokalisiert und ebenso genau den Wert des Impulses hat. Es ist eine fundamentale Einsicht der Quantentheorie, dass jede Messung ihren Gegenstand beeinflusst, nicht im Sinne einer Störung eines mikroskopischen Systems durch eine makroskopische Messung, sondern aus der Konsequenz inkompatibler Eigenschaften.

„Der wichtigste Unterschied zur klassischen Theorie besteht darin, dass bei der Beobachtung irgendeiner physikalischen Größe die Störung wesentlich in Betracht gezogen werden muss, die das zur Beobachtung ausgeführte Experiment am zu messenden System hervorruft.“⁴

Gemessen werden immer relationale Eigenschaften, die in einer Wechselbeziehung zwischen Quantensystem und einer Messapparatur bestehen. Damit ist nicht vorhersehbar, welchen Wert eine Eigenschaft im Falle einer Messung besitzt, nur Wahrscheinlichkeiten sind berechenbar. Im Sinne der Kausalität stört das beobachtende Subjekt immer das Objekt und die Grenzen zwischen diesen verschwimmen. Die

⁴ W. Heisenberg; Naturwissenschaften, 1929; in: Paul Feyerabend; Probleme des Empirismus, Braunschweig 1981, S. 439

Heisenbergsche Unschärferelation kann somit so interpretiert werden, dass in der Teilchenwelt die Materie nicht an festen Orten existiert. Sie ist immer in einem Bewegungszustand, es hängt nur von der Messmethode ab, welche Erscheinungsform der atomaren Wirklichkeit wir erkennen. Diese Betrachtung wird in der Welle-Teilchen-Dualität ersichtlich. Die Beobachtung eines Elektrons führt unter bestimmten Beobachtungsbedingungen zu einer punktförmigen Ausbildung an einem bestimmten Ort; unter anderen Bedingungen erweist es sich jedoch als ausgedehnte Welle.

„Wenn ein Strom von Elektronen durch eine dünne Metallfolie hindurchgetrieben wird, so verhalten sich die Elektronen beim Durchgang wie Wellen: sie interferieren und kehren nach dem Durchgang wieder ihren korpuskularen Charakter hervor. Wenn man nun ein bestimmtes Elektron vor dem Durchgang mit einem bestimmten einzelnen Elektron nach dem Durchgang vergleicht und fragt, ob sie dasselbe Elektron sind, so kennt die Physik kein Mittel, die Frage zu entscheiden. Ja, im Gedankensystem der Wellenmechanik verliert es jeden Sinn, eine solche Frage aufzuwerfen, weil diese Theorie eine solche Struktur besitzt, dass sie prinzipiell nur Fragen über das wahrscheinliche Auftreten von Elektronen zulässt.“⁵

Somit wird eine räumliche Beschreibung des Elektrons unmöglich, was zur Folge hat, dass die menschliche Kategorie Raum ihren Sinn verliert. Diese Betrachtung führt zu dem Schluss, dass in der Quantenphysik die menschliche Abstraktion an ihre Grenzen stößt und die vom Menschen geschaffenen Begriffe Raum und Zeit lösen sich auf.

2.2 Informationsbegriff der modernen Sensorik

Die Unterschiedlichkeit dessen, was unter dem Begriff der Information (v. lat.: *informare* = bilden, eine Form geben) verstanden wird, resultiert aus der Vielzahl unterschiedlicher wissenschaftlicher Modelle, Sichtweisen und Theorien. Hierbei werden viele verschiedene Begriffe wie Abstraktion, Relevanz, Redundanz, Syntax oder Semantik in unterschiedlicher Ausprägung verwendet. Sicher erscheint eine umfassende

⁵ vgl. Friedrich Waismann; Logik-Sprache-Philosophie, Stuttgart 1985, S.294

Ausarbeitung einer einheitlichen Definition des Begriffs der Information sinnvoll, jedoch führt dieses Anliegen unmittelbar zu der Frage, ob diese allgemeine Theorie der Information überhaupt möglich ist, solange sich die unterschiedlichen Wissenschaftsmodelle nicht vereinen lassen. Aufgrund der Komplexität einer solchen Aufgabe sollen hier nur die im technischen Sinne relevanten Aspekte des Begriffs der Information innerhalb der Sensorik diskutiert werden.

Einen solch wichtigen Aspekt in der naturwissenschaftlichen Betrachtung der Information stellt der Begriff der Abstraktion (v. lat. *abstractus* = abgezogen, entfernt, getrennt) dar. Dieser bezeichnet den Prozess einer Veranschaulichung von Objekten, Situationen, Zuständen usw. in der Art, dass auf individuelle oder gar zufällige Einzelheiten zugunsten des Wesentlichen verzichtet wird. Jeder Begriffsbildung liegt somit Abstraktion zugrunde, wodurch eine klare Unterscheidung zwischen zum Begriff gehörendem und außerhalb befindlichem möglich ist. Das Prinzip der Abstraktion generiert somit die Grundlage einer jeden wissenschaftlichen Betrachtung, da mit ihr erst wesentliche Zusammenhänge durch die Verwendung bestimmter Begriffe für Zustände, Objekte, Eigenschaften usw. dargestellt werden können. Jede Aussage, die in einem wirklichkeitstreuem Zusammenhang steht und somit keinen zufälligen Charakter besitzt, basiert auf dem Prozess der Abstraktion und kann als die Gewinnung einer Information betrachtet werden. Aufgrund dieses Umstands besteht eine enger Bezug zwischen dem Begriff der Information und dem der Abstraktion.

Die mit der Abstraktion gegebene Reduktion auf das Wesentliche führt zwangsläufig zu der Frage nach der Relevanz. Ebenso wie der Informationsbegriff wird die Relevanz in den unterschiedlichen Bereichen sehr unterschiedlich definiert. Beispielsweise handelt es sich in der Wissenschaftstheorie um einen relativen Ausdruck, der sich aus dem Verhältnis zu einem bestimmten Kontext ergibt. In der Informationstheorie jedoch ist sie eine quantitative Größe, die sich so objektiv bewerten lässt. Auch innerhalb der Ingenieurwissenschaften besitzt die Relevanz einer Information eine recht unterschiedliche Bedeutung. Während in der Mess-, Steuer- und Regelungstechnik bestimmte Informationen, welche auch als Messgrößen bezeichnet werden, eine enorme Relevanz besitzen, hat dieser Begriff für den Nachrichtentechniker keine Bedeutung, solange er sich allein um die Übertragung einer Information bemüht. Beschäftigt er sich aber beispielsweise mit der Kodierung einer Information, also dem Übertragen der Information in eine alternative Darstellung, dann begegnet auch ihm der Begriff der Redundanz (v. lat. *redundare* = im Überfluss vorhanden sein), der mitunter als die

Umkehrung des Begriffs Relevanz verstanden wird. Diese Begriffe sollte man jedoch klar voneinander trennen, da beispielsweise die mehrfache Übertragung ein und derselben Information eine gewisse Redundanz mit sich bringt, ihre Relevanz davon jedoch unberührt bleibt.

Die Abbildung einer Information, also auch einer Messgröße, auf einen bestimmten Informationsträger, sei es eine Sprache oder eine elektrische Größe, muss natürlich in einem nachvollziehbaren Zusammenhang stehen, also bestimmten Regeln folgen, was als Syntax (aus d. Griechischen für: die Zusammenstellung) bezeichnet werden kann. Im einfachsten Fall eines linearen analogen Wandlers, der die Messgröße beispielsweise in einen Spannungswert wandelt, stellt die zugehörige lineare Gleichung diese Syntax dar. Die Bedingung der Eindeutigkeit für eine verwendete Syntax führt zwangsläufig zu einem Kausalzusammenhang, also der Anerkennung des Ursache-Wirkungs-Prinzips. Des Weiteren sollte eine Syntax einer gewissen Sinnhaftigkeit folgen, wofür im Allgemeinen der Begriff der Semantik verwendet wird. Auch dieser Begriff wird in den unterschiedlichsten Wissenschaftsdisziplinen unterschiedlich verstanden. So bezeichnet die Semantik in der soziologischen Systemtheorie den gesamten Wissensvorrat, der dauerhaft, wieder verwendbar und gesellschaftlich übergreifend verfügbar ist. Dagegen ist die Semantik der Informationstheorie allein die Bedeutung einer Zeichenfolge. Eine reine Zufallsfolge, wie sie in technischen Systemen durch auftretende Rauschprozesse verursacht werden, besitzt keine Semantik, jedoch einen sehr hohen Informationsgehalt, da jedes Zeichen für sich eine Information darstellt, aber ihnen über die zugehörige Syntax kein Sinn zugeordnet werden kann.

Begrenzt man den Informationsbegriff auf den Bereich der Sensorik, dann handelt es sich hierbei um eine Abstraktion im Sinne einer Unterscheidung, die auf einer gewissen Sensibilität gegenüber einem zu beobachtenden Objekt beruht. Ein Sensor ist also das Mittel zur Unterscheidung von Zuständen eines oder mehrerer Objekte indem er diese in eine zumeist physikalische Größe transformiert. Mit der zugehörigen Transformationsregel, welche hier als Syntax begriffen werden kann, gelingt eine eindeutige Aussage über verschiedene Zustände. Eine Information ist innerhalb der Sensorik also immer dann gegeben, wenn aufgrund einer Sensitivität und einer Syntax eindeutige Unterscheidungen von Zuständen möglich sind. Der Begriff der Relevanz spielt innerhalb der Sensorik nur eine untergeordnete Rolle, da diese wesentlich durch denjenigen gegeben wird der auf dieser Information bestimmte Entscheidungen aufbaut. Die Sensorik interessiert sich jedoch nicht für diese Entscheidungen, sondern nur dafür,

dass verschiedene Situationen reproduzierbar in eine Größe abgebildet werden. Damit ist der Informationsbegriff der Sensorik stark mit dem Begriff Syntax verbunden. Ist ein solcher nicht bekannt, ist es unmöglich einen Bezug zwischen der Ausgangsgröße und der gerade tatsächlich vorhandenen Messsituation herzustellen.

Die moderne Sensorik beschäftigt sich heute jedoch nicht mehr allein mit einer einzelnen Messaufgabe, sondern verbindet mehrere Einzelsensoren zu komplexen mehrdimensionalen Systemen. Mit dieser Komplexität wächst auch die Anforderung an die verwendete Syntax, da es nun gilt den betrachteten Zustand mehrdimensional abzubilden. Hier liegt nun die eigentliche Aufgabe der modernen Sensorik, die Einzelinformationen aller Sensoren in eine adäquate mehrdimensionale Information umzusetzen.

Eine Grenze des hier formulierten sensorischen Informationsbegriffs findet sich beispielsweise im Bereich der Optosensorik. Da hier quantenmechanische Prozesse betrachtet werden, ist es nicht vorhersehbar, welchen Wert eine Eigenschaft im Falle einer Messung besitzt, nur Wahrscheinlichkeiten sind berechenbar. Dies steht der Kausalität jedoch entgegen. In der Quantentheorie kann der Zustand eines einzelnen Teilchens nicht beschrieben werden, sondern nur seine Zugehörigkeit und Beziehung zum beobachtenden Subjekt. Um jedoch Kausalität im strengen Sinn anzuwenden, muss der Zustand eines physikalischen Objekts genau feststellbar sein, um anhand eines bekannten Kausalgesetzes den späteren Zustand exakt vorherzusagen. Die absolute Isolierung, eine klassische wissenschaftlich Methode, greift nicht mehr. Kausalität gilt somit immer nur unter bestimmten Bedingungen und ist ein Instrument zur technischen und praktischen Beherrschung der Natur. Andererseits finden in der Optosensorik auch klassische Wissenschaftstheorien ihre Anwendung, wie der Wellencharakter des Lichts. Der Widerspruch jedoch, dass Licht objektiv aus Wellen und Teilchen besteht, ist nicht in den Dingen selbst begründet, sondern besteht in der menschlichen Abstraktion und unserer auf Kausalität aufbauenden Logik. Auch Quanten sind eine Abstraktion bzw. theoretische Vereinheitlichung mit dem Sinn Beobachtung und Vorhersage zu ermöglichen. Wo der Mensch etwas beobachtet setzt er der Wirklichkeit die voneinander getrennten Kategorien Raum und Zeit auf und es lässt sich behaupten, dass diese nicht an sich existieren, sondern nur innerhalb des menschlichen Bewusstseins:

„Die Welt, die wir uns aus unseren Empfindungen und Wahrnehmungen konstruieren und die wir uns bequemerweise immer so denken, als sei sie an und für sich einfach

schlechthin vorhanden, ist also nicht schon durch das bloße Vorhandensein auch wirklich manifest, sondern dazu bedarf es der Gehirnfunktionen.“⁶

2.3 Information in der Nachrichtentechnik

Mit der zunehmenden Komplexität in allen Bereichen der menschlichen Gesellschaft, sei es im administrativen, wissenschaftlichen oder rein technischem Sinne, entsteht zwangsläufig ein zunehmender Bedarf an Informationen. Das Einschätzen und Werten einer bestimmten Situation anhand der vorhandenen Informationen ist eine anspruchsvolle Aufgabe der Informationsverarbeitung, welche jedoch in ihrer Qualität im starken Maße von der beschafften Informationsmenge, ihrer Aktualität und Genauigkeit abhängt. Voraussetzung dafür ist deshalb grundsätzlich eine Informationsgewinnung in ausreichendem Maße, da das Fehlen an Information die Möglichkeit von Fehleinschätzungen in sich birgt. Dies schließt die eigentliche Informationsgewinnung und die sich anschließende Informationsübermittlung ein. Trotz gesteigener und weiter steigender Komplexität bleibt die Aufgabe der Informationsübermittlung gleich: Die ausgesandten Signale sollen über weite Entfernungen ohne Verlust an Information übertragen werden, entgegen allgegenwärtiger Störungen.

In den letzten Jahrzehnten kamen jedoch weitere Anforderungen hinzu, wobei hier neben dem Aspekt der Sicherheit, also dem Schutz der Information vor dem Zugriff Unbefugter, der enorme Anstieg der Informationsmenge zu nennen ist. Dieser Anstieg führt zu einer besonders drängenden Aufgabe, nämlich die begrenzte Kapazität eines Übertragungsweges, im technischen Sinn Übertragungskanal genannt, möglichst effektiv zu nutzen. Ein ideales Übertragungsverfahren im nachrichtentechnischen Sinn sollte deshalb nur ein Minimum an Bandbreite nutzen und gleichzeitig störsicher sein.

Technisch gesehen ist die Informationsübertragung an einen Träger gebunden, dem die entsprechende Information nach einem bestimmten Muster aufgeprägt wird. Grundsätzlich sind also zur Informationsübertragung folgende zwei Aufgaben zu unterscheiden: die Vermittlungsaufgabe und die Transportaufgabe. Die Aufgabe der Informationsvermittlung kommt im nachrichtentechnischen Sinn dem

⁶ Erwin Schrödinger; Geist und Materie, Zürich 1989, S. 92

Übertragungsverfahren zu, da dieses die Art und Weise beschreibt, wie eine Information auf einen Träger abgebildet und wieder gelesen wird. Diesem Informationsträger wiederum kommt die eigentliche Transportaufgabe zu.

Die Darstellung einer Information (Vermittlungsaufgabe) durch eine physikalische Größe heißt in der Nachrichtentechnik Signal. Die Weitergabe einer Information bedingt eine Änderung dieser physikalischen Größe, womit sich die Signalübertragung durch eine Funktion dieser Größe nach Raum bzw. Zeit definieren lässt. Eine vollständige Beschreibung einer Informationsübermittlung kann also nur durch eine nachrichtentechnische Beschreibung des Übertragungsverfahrens und durch eine physikalische Beschreibung des Trägers erfolgen.

2.4 Grundprinzipien der Signaltheorie

2.4.1 Signale und das Fourier-Prinzip

Da jede mikrophysikalische Größe einen unvorhersehbaren und damit zufälligen Charakter besitzt, ergeben sich unendlich viele Möglichkeiten an Zustandsänderungen, die durch einen Empfänger beobachtet werden können. Damit sind also unendlich viele Signale möglich, deren mathematische Beschreibung unmöglich erscheint. Beschränkt man sich jedoch auf Elementarsignale, wie z.B. die Sinusfunktion, dann ist diese Beschreibung sehr einfach möglich. Eine solche periodische Funktion wiederholt sich immer wieder nach einer bestimmten Zeitdauer auf die gleiche Art. Idealerweise betrachtet ist sie dadurch unendlich weit in ihrer Vergangenheit und Zukunft bekannt und somit exakt vorhersehbar. Eine interessante Idee liegt nun darin, komplexe Signale auf die einfache Beschreibung solcher Elementarsignale zurückzuführen. Dem französischen Mathematiker, Naturwissenschaftler und Berater Napoleons Jean Baptist Fourier (1768-1830) gelang dies, indem er beliebige periodische Signale aus mehreren Sinusschwingungen unterschiedlicher Frequenz und Amplitude zusammensetzte. Die Möglichkeit der Darstellung eines beliebigen Signals durch das Zusammensetzen aus mehreren Sinusschwingungen wird als Fourier-Prinzip bezeichnet und kann auch auf nichtperiodische oder einmalige Signale angewendet werden. Die besondere Bedeutung dieses Prinzips liegt für die Nachrichtentechnik in seiner Umkehrung: Ist bekannt, wie

ein beliebiges lineares zeitinvariantes System auf Sinusschwingungen aller Frequenzen reagiert, dann ist ebenso klar, wie es auf jedes beliebige Signale reagiert. Schließlich sind all diese aus bestimmten Sinusschwingungen zusammengesetzt.⁷

2.4.2 Signale und das Prinzip der Unschärfe

Eine der wichtigsten Erkenntnis der modernen Physik ist die Tatsache, dass bestimmte Größen nicht unabhängig voneinander gemessen werden können. Solche Größen bezeichnet man als komplementär⁸. Die Messung solcher Größen unterliegt einem Naturgesetz, welches als Heisenbergsche Unschärfe bezeichnet wird. Es basiert auf dem Welle-Teilchen-Dualismus, mit dem sich nach de Broglie jedem Teilchen mit dem Impuls p ein Wellenpaket der Länge λ zuordnen lässt. Die Lösung der Wellengleichung in ihrer einfachsten eindimensionalen Form lautet dabei:

$$s(x, t) = a \cdot \cos(\omega \cdot t - k \cdot x) \text{ mit } k = \frac{2\pi}{\lambda}. \quad \text{Gl. 2.5}$$

Betrachtet man diese für einen festen Zeitpunkt, dann erhält man eine Momentaufnahme, und der Parameter t kann entfallen. Mit der Beziehung $\cos(x) = \cos(-x)$ ist nur noch folgender Ausdruck von Interesse:

$$s(x) = a \cdot \cos(k \cdot x). \quad \text{Gl. 2.6}$$

Nun steht der Impuls mit dem Wellenvektor k in einem festen Zusammenhang:

$$p = \frac{h}{2 \cdot \pi} \cdot k. \quad \text{Gl. 2.7}$$

⁷ vgl. U. Karrenberg; Signale, Prozesse, Systeme; Springer Verlag Berlin, 2002, S. 39

⁸ aus dem Lateinischen für „sich ergänzend“

Ist nun die Kenntnis des Impulses absolut scharf, dann erhält man ein unendlich ausgedehntes Wellenpaket (**Abb. 2.1**). Ist jedoch der Impuls mit einer gewissen Unschärfe belastet, dann ist das Wellenpaket deutlich erkennbar (**Abb. 2.2**). Hier wird nun deutlich, dass der unscharfe Impuls eine Ortsbestimmung des Wellenpaketes bzw. des Teilchens ermöglicht. Aus umgekehrter Richtung betrachtet, ist mit einer scharfen Ortskenntnis jedoch keine exakte Bestimmung der Wellenlänge bzw. des Impulses möglich. Die Betrachtung der Unschärfeproblematik in dieser einfachen Form lässt einen Zusammenhang zum Fourier-Prinzip erahnen. Tatsächlich kann das örtlich begrenzte Wellenpaket aus einer Summe vieler unendlicher Cosinusfunktionen zusammengesetzt werden, wenn deren Werte für k (und damit deren Wellenlängen) um den vorgegebenen Wert herum schwanken. Dieser Zusammenhang soll nun durch die Betrachtung eines Elektrons veranschaulicht werden. Sein Wellenpaket lässt sich wieder unter Vernachlässigung der Zeit folgendermaßen beschreiben:

$$\Psi(k) = \begin{cases} \Psi_0 & \text{für } k_0 - \Delta k / 2 \leq k \leq k_0 + \Delta k / 2 \\ 0 & \text{sonst.} \end{cases} \quad \text{Gl. 2.8}$$

Durch die Anwendung der Fourier-Transformation erhält man nun die Darstellung dieses Wellenpakets im Ortsbereich bzw. in Abhängigkeit von x :

$$\begin{aligned} \Psi(x) &= \frac{C}{2 \cdot \pi} \int_{-\infty}^{\infty} \Psi(k) \cdot e^{jkx} dk = \frac{C}{2 \cdot \pi} \int_{k_0 - \Delta k / 2}^{k_0 + \Delta k / 2} \Psi_0 \cdot e^{jkx} dk \\ &= \underbrace{\frac{C}{2 \cdot \pi} \Psi_0 \cdot \Delta k \cdot e^{jkx}}_{\text{Grundschwingung}} \cdot \underbrace{\frac{\sin(\Delta kx / 2)}{\Delta kx / 2}}_{\text{Einhüllende}} \end{aligned} \quad \text{Gl. 2.9}$$

Deutlich ist nun hier der Zusammenhang zwischen Grundschwingung und Einhüllender des Wellenpakets zu erkennen.

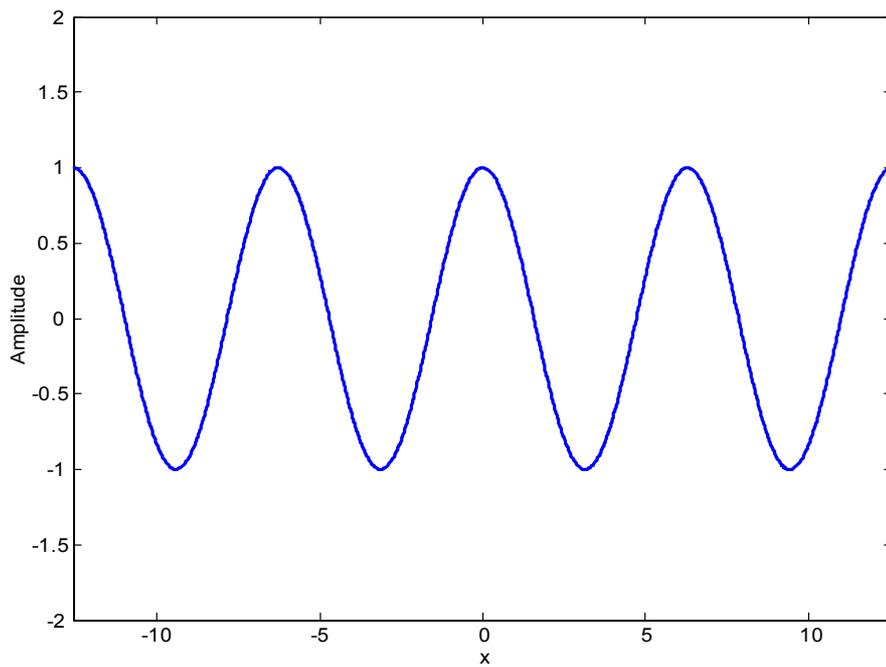


Abb. 2.1: Unendlich ausgedehntes Wellenpaket aufgrund eines scharfen Impulses nach Gl. 2.9

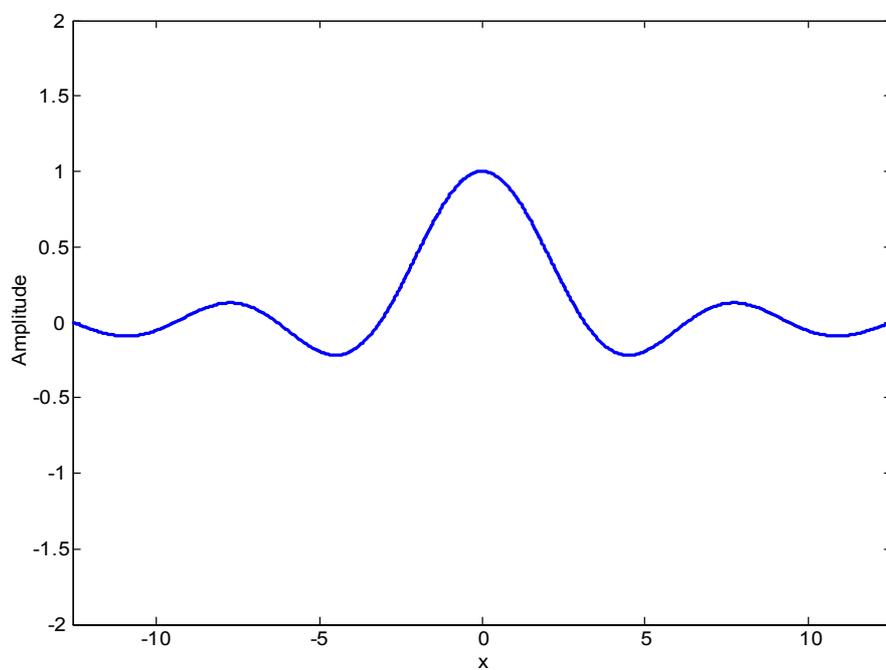


Abb. 2.2: Örtlich begrenztes Wellenpaket aufgrund eines unscharfen Impulses nach Gl. 2.9

Für die Systemtheorie hat nun die Heisenbergsche Unschärfe eine sehr entscheidende Bedeutung, denn sie stellt nicht nur das Prinzip der Kausalität in Frage (**Kap. 2.1.1**), sondern begrenzt ihre Leistungsfähigkeit in der technischen Anwendung. In der Sensorik wird dies sehr schnell deutlich, wenn man versucht unterschiedliche Systeme hinsichtlich ihrer Genauigkeit zu vergleichen. Heisenbergs Unschärferelation lehrt uns, dass für einen objektiven Vergleich die Messrate eines Sensors mit einzuschließen ist, denn selbst der qualitativ schlechteste Sensor wird mit einem ausreichend großem Zeitfenster die notwendige Auflösung erreichen, um konkurrierende zu übertreffen.

2.4.3 Signale und das Prinzip der Symmetrie

Der Begriff der Symmetrie (altgriechisch: *symmetria* = Ebenmaß) stellt ein wichtiges Strukturmerkmal in den Naturwissenschaften dar, der mit dem Begriff der Abstraktion in engem Zusammenhang steht. Die Abstraktion beschreibt den Vorgang der Abgrenzung und ermöglicht es, Objekte, Zustände und Änderungen zu unterscheiden. Das Unterscheiden beruht immer auf zwei gegensätzlichen Eigenschaften wie z.B. groß und klein, schnell und langsam, oben und unten. Ein einzelner Begriff allein ist nutzlos, erst sein Gegenstück erfüllt ihn mit einem Sinn. Eine wesentliche Frage in der Naturwissenschaft ist nun, ob mit der Abstraktion immer eine Symmetrie einhergeht. Dies würde zu Abstraktion immer ein spiegelbildliches Gegenstück erfordern. In der Mathematik ist dieses Prinzip absolut üblich, so dass ein penibler Mathematiker für eine gegebene quadratische Fläche A die Seitenlänge nicht nur mit $s = \sqrt{A}$ angibt, sondern ebenso die Lösung $s = -\sqrt{A}$. So war die Existenz von Antimaterie für den Mathematiker Paul Dirac selbstverständlich möglich, und tatsächlich konnte diese am 2. August 1932 durch Carl Anderson auf der Sternwarte von Mount Wilson nachgewiesen werden. Anderson führte Untersuchungen zur kosmischen Strahlung durch, wobei er in einer Nebelkammer, die von Bleiplatten durchtrennt war, und einer Kamera eintreffende Teilchen detektierte. Ein parallel liegendes magnetisches Feld krümmt hierbei seine Flugbahn. Diese Krümmung hängt nun aber von der Masse des Teilchens, seiner Geschwindigkeit und der Polarität seiner Ladung ab. Bei einer aufgenommenen Spur handelte es sich deutlich um eine positive Ladung, und mit dem damaligen Wissen würde man schließen, dass es sich um ein Proton handelte. Allerdings durchdringt die

Spur eine Bleiplatte, so dass das Teilchen eine sehr hohe Geschwindigkeit gehabt haben musste. Ein Proton dieser Geschwindigkeit würde jedoch einen viel größeren Krümmungsradius aufweisen. Mittels weiterer Untersuchungen der Spur vor und hinter der Bleiplatte konnte die Geschwindigkeit und Masse des Teilchens bestimmt werden. Das Ergebnis ergab eine Masse, die gleich groß der eines Elektrons ist. Anderson hatte somit ein positives Elektron entdeckt. Dieses Positron war das erste Anti-Materie-Teilchen, das je in der menschlichen Geschichte dokumentiert wurde.

Weitet man nun dieses Symmetrieprinzip auf die Betrachtung von Signalen aus, dann sollten negative Frequenzen existieren. Im Zeitbereich ist die Symmetrie für ein periodisches Signal gegeben, da die Vergangenheit für $t < 0$ und die Zukunft für $t > 0$ darstellbar ist. Für die Existenz negativer Frequenzen lässt sich relativ leicht argumentieren, wenn man sich von einer zeitlichen Abhängigkeit löst. Betrachtet man beispielsweise ein örtlich abhängiges Signal (wie z.B. in der Bildverarbeitung), dann steht es einem völlig frei, dieses in der einen oder anderen (und damit entgegengesetzten) Richtung zu beschreiben. Je nach Richtung der Betrachtung kann man von einer positiven oder negativen Frequenz sprechen, wobei jede der beiden gleichwertig zu betrachten ist. Überträgt man diese Betrachtung auf ein zeitabhängiges Signal, dann ist die Messung der Periodendauer bzw. Frequenz $f = 1/T$ in positiver Koordinatenrichtung ebenso willkürlich. Eine Messung in negativer Richtung führt zu einer negativen Periodendauer und resultiert in einer negativen Frequenz.

2.5 Grundlagen der Systemtheorie

2.5.1 Eindimensionale Systeme

a) Der systemtheoretische Begriff des Systems

Eine Informationsübertragung ist im Allgemeinen ein sehr komplexer Vorgang, welcher sich nur schwer in seiner Gesamtheit vollständig beschreiben lässt. Eine Analyse aller ihr anhaftenden Eigenschaften, z.B. mittels Differentialgleichungen, ist relativ mühsam und unanschaulich. Zur Vereinfachung kann man den gesamten Übertragungsvorgang in einzelne Teilsysteme aufteilen. Jedes Teilsystem für sich lässt sich nun einzeln und unabhängig von den anderen beschreiben, wenn sie sich rückwirkungsfrei verhalten.

Eine weitere Vereinfachung erfolgt dadurch, dass man ein solches Teilsystem nur durch normierte Eingangsgrößen und Ausgangsgrößen beschreibt, ohne direktes Wissen über den inneren Aufbau. Entsprechend dem Fourier-Prinzip ist ein System trotz dieser Abstraktion vollständig beschrieben, da das Wissen über die Reaktion auf alle möglichen Sinusschwingungen eine Beschreibung für alle anderen Signalformen erlaubt. Die feste Zuordnung eines idealisierten Eingangssignals zu einem ebenfalls idealisierten Ausgangssignal führt zu der sogenannten Systemtheorie. Ein System ist somit durch die Transformation $Tr\{\dots\}$ eines beliebigen Eingangssignal $s(t)$ zu einem eindeutigen Ausgangssignal $g(t)$ definiert:

$$g(t) = Tr\{s(t)\}. \quad \text{Gl. 2.10}$$

Der wesentliche Vorteil einer systemtheoretischen Beschreibung findet sich darin, dass die Betrachtung eines Gesamtsystems so auf die Analyse unabhängiger Teilsysteme zurückgeführt werden kann.

b) Lineare zeitinvariante (LTI)-⁹ Systeme:

Eine bedeutende Gruppe aller möglichen Systeme kann durch eine lineare Differentialgleichung mit konstanten Koeffizienten beschrieben werden. Ihre Transformationsgleichungen sind somit besonders einfach, wobei zwei entscheidende Eigenschaften mit ihnen verbunden und von besonderer Bedeutung sind. Die Eigenschaft der Linearität ist für ein System gegeben, wenn jede Linearkombination von Eingangssignalen zu einer entsprechenden Linearkombination von Ausgangssignalen führt. Ihre Transformationsgleichung muss also dem Superpositionssatz genügen:

$$g(t) = \sum_i Tr\{a_i \cdot s_i(t)\} = \sum_i a_i \cdot Tr\{s_i(t)\} = \sum_i a_i \cdot g_i(t). \quad \text{Gl. 2.11}$$

Diese elementare Eigenschaft linear zeitinvariante Systeme ist eine Folge des Fourier-Prinzips, welches sich nur durch die Regel der Superposition erfüllen lässt.

⁹ aus dem Englischen: Linear Time-Invariant

Der Eigenschaft der Zeitinvarianz unterliegen Systeme, deren Form des Ausgangssignals vom Zeitpunkt bzw. einer Verschiebung des Eingangssignals unabhängig ist. Anders formuliert reagiert ein zeitinvariantes System auf eine bestimmte Anregung, welche zeitlich verzögert ist, mit der gleichen Ausgangsreaktion, nur ebenso zeitverzögert:

$$\text{Tr}\{s(t-t_0)\} = g(t-t_0). \quad \text{Gl. 2.12}$$

c) Das Faltungstheorem

Die Beschreibung eines LTI-Systems kann näherungsweise dadurch erfolgen, dass man die Reaktion $g_0(t)$ auf einen Rechteckimpuls $s_0(t)$ der Dauer T_0 und der Höhe $1/T_0$ bestimmt. Dann kann jede beliebige Eingangsfunktion $s(t)$ durch eine Reihe aufeinander folgender Rechteckimpulse angenähert werden, um so die Reaktion $g(t)$ ebenfalls näherungsweise zu bestimmen (**Abb. 2.3**). Die Amplituden der einzelnen Rechteckimpulse der Treppenfunktion werden hierbei mit dem gegebenen Eingangssignal $s(t)$ gewichtet. Dazu wird der Rechteckimpuls $s_0(t)$ mit dem Wert $s(nT_0)T_0$ multipliziert. Die approximierende Treppenfunktion lautet so:

$$s(t) \approx s_a(t) = \sum_{n=-\infty}^{\infty} s(nT_0) \cdot s_0(t-nT_0) \cdot T_0. \quad \text{Gl. 2.13}$$

Folgt man dem Fourier-Prinzip bzw. dem Superpositionssatz, dann reagiert ein LTI-System mit:

$$g_a(t) = \sum_{n=-\infty}^{\infty} s(nT_0) \cdot g_0(t-nT_0) \cdot T_0 \approx g(t). \quad \text{Gl. 2.14}$$

Die Approximation wird umso genauer, je kürzer die Dauer T_0 des Einzelrechteckimpulses gewählt wird. Für den Grenzübergang $T_0 \rightarrow 0$ gehen die Summen in Integrale über.

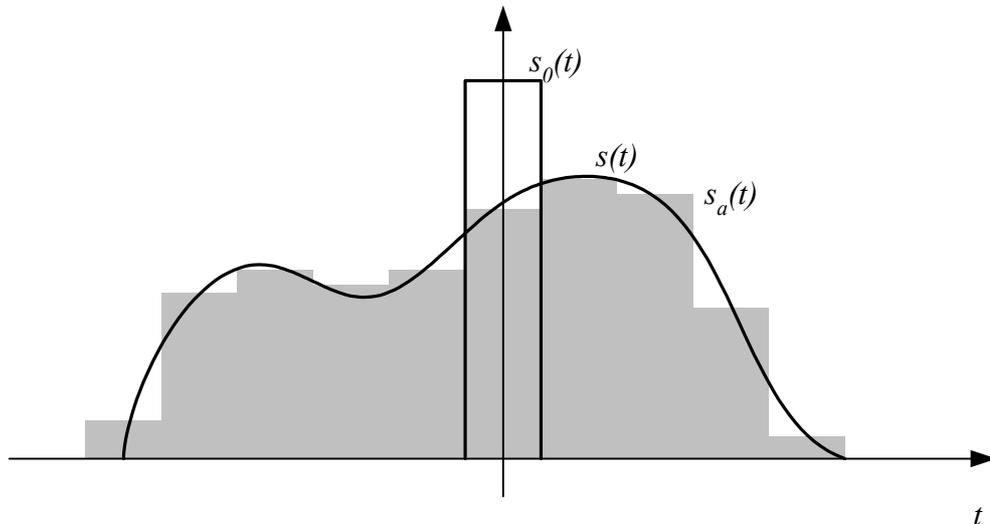


Abb. 2.3: Approximation der Funktion $s(t)$ durch eine Treppenfunktion $s_a(t)$

Der ursprüngliche Rechteckimpuls $s_0(t)$ wird nun durch einen Dirac-Stoß $\delta(t)$ und die Antwort $g_0(t)$ als Stoßantwort $h(t)$ beschrieben. Die so entstandenen Beziehungen werden als Faltungsintegrale bezeichnet:

$$s(t) = \int_{-\infty}^{\infty} s(\tau) \cdot \delta(t - \tau) d\tau = s(t) * \delta(t). \quad \text{Gl. 2.15}$$

Dieser Ausdruck veranschaulicht eine unendliche feine gestufte Treppenfunktion als eine unendliche Reihe von Dirac-Stößen. Dagegen stellt

$$g(t) = \int_{-\infty}^{\infty} s(\tau) \cdot h(t - \tau) d\tau = s(t) * h(t) \quad \text{Gl. 2.16}$$

eine exakte Antwort $g(t)$ eines LTI-Systems auf ein beliebiges Eingangssignal $s(t)$ dar und erlaubt somit die gewünschte systemtheoretisch abstrahierende Beschreibung.

d) Die Fourier-Transformation

Der Informationstransport und die damit verbundene zeitliche Abhängigkeit einer physikalischen Größe erfordert eine Betrachtung im sogenannten Zeitbereich, während das Fourier-Prinzip die Betrachtung auf den Frequenzbereich ausweitet. Alle Signale können aus diesen zwei Perspektiven betrachtet werden, denn alle Signalinformationen sind jeweils in ihnen enthalten. Je nach Anwendungsfall erscheint die eine wie die andere als vorteilhaft. So gibt es nun Funktionen, deren Übertragung über ein LTI-System allein von ihrer Amplitude abhängen. Diese werden als Eigenfunktionen bezeichnet, für die gilt:

$$g(t) = s_E(t) * h(t) = H \cdot s_E(t). \quad \text{Gl. 2.17}$$

Eine solche Eigenfunktion wird mathematisch folgendermaßen in komplexer und exponentieller Form ausgedrückt:

$$s_E(t) = \cos(2\pi ft) + j \sin(2\pi ft) = e^{j2\pi ft}. \quad \text{Gl. 2.18}$$

Setzt man den exponentiellen Ausdruck in das Faltungsintegral (Gl. 2.16) ein, dann so ergibt sich für beliebige Frequenzen:

$$\begin{aligned} s_E(t) * h(t) &= e^{j2\pi ft} * h(t) \\ &= \int_{-\infty}^{\infty} h(\tau) \cdot e^{j2\pi f(t-\tau)} d\tau \\ &= e^{j2\pi ft} \cdot \int_{-\infty}^{\infty} h(\tau) \cdot e^{-j2\pi f\tau} d\tau \\ &= e^{j2\pi ft} \cdot H(f). \end{aligned} \quad \text{Gl. 2.19}$$

Der so genannte Eigenwert $H(f)$ stellt in der Systemtheorie die Übertragungsfunktion des LTI-Systems im Frequenzbereich dar, welcher im direkten Zusammenhang zur Stoßantwort $h(t)$ im Zeitbereich steht. Entsprechend dem Fourier-Prinzip kann jede beliebige Zeitfunktion aus einer Summe verschieden gewichteter Sinus- bzw. Cosinusschwingungen zusammengesetzt werden, welche jede für sich eine einzelne

Eigenfunktion darstellt. Daher kann der Ausdruck in **Gl. 2.19** ganz allgemein auf alle Signale ausgeweitet werden:

$$S(f) = \int_{-\infty}^{\infty} s(t) \cdot e^{-j2\pi ft} dt. \quad \text{Gl. 2.20}$$

Dieser Ausdruck wird als Fourier-Transformation bezeichnet, wobei die Integrationsvariable τ durch t ersetzt wurde. Die entsprechende Rücktransformation nennt sich inverse Fourier-Transformation:

$$s(t) = \int_{-\infty}^{\infty} S(f) \cdot e^{j2\pi ft} df. \quad \text{Gl. 2.21}$$

Hat man zwei aufeinander folgende Systeme mit den Stoßantworten h_1 und h_2 , dann ergibt sich aus der Faltungsalgebra:

$$\begin{aligned} g(t) &= [s_E(t) * h_1(t)] * h_2(t) \\ &= H_1(f) \cdot H_2(f) \cdot s_E(t). \end{aligned} \quad \text{Gl. 2.22}$$

Wie aus dieser Gleichung ersichtlich wird, geht die Faltung im Zeitbereich durch die Fourier-Transformation in eine Multiplikation im Frequenzbereich über. Führt man rein formal eine Faltung von Fourier-Spektren durch, so geht diese durch die inverse Fourier-Transformation in eine Multiplikation im Zeitbereich über.

e) Das Bandpasssignal im äquivalenten Tiefpassbereich

Ein beliebiges Bandpasssystem lässt sich durch ein sogenanntes äquivalentes Tiefpasssystem beschreiben. Hierbei ergibt sich die Übertragungsfunktion $H_{BP}(f)$ eines idealen Bandpasses im Frequenzbereich durch zwei gegenüber dem Ursprung symmetrische Rechteckfunktionen, die sich folgendermaßen darstellen lassen:

$$H_{BP}(f) = \text{rect}\left(\frac{f + f_0}{f_\Delta}\right) + \text{rect}\left(\frac{f - f_0}{f_\Delta}\right). \quad \text{Gl. 2.23}$$

Mit Hilfe des Faltungsproduktes kann der Zusammenhang zu der Übertragungsfunktion eines idealen Tiefpasses hergeleitet werden:

$$H_{BP}(f) = \text{rect}\left(\frac{f}{f_\Delta}\right) * [\delta(f - f_0) + \delta(f + f_0)]. \quad \text{Gl. 2.24}$$

Der generelle Zusammenhang zwischen Bandpassspektrum $S(f)$ und äquivalenten Tiefpassspektrum $S_T(f)$ lautet:

$$S(f) = \frac{1}{2} [S_T(f - f_0) + S_T^*(-f - f_0)] \quad \text{Gl. 2.25}$$

bzw.

$$S_T(f) = 2 \cdot S(f + f_0) \cdot \text{rect}\left(\frac{f}{f_\Delta}\right). \quad \text{Gl. 2.26}$$

Im Zeitbereich lässt sich dieser Zusammenhang für das Bandpasssignal $s(t)$ und das äquivalente Tiefpasssignal $s_T(t)$ folgendermaßen darstellen:

$$s(t) = \text{Re}\{s_T(t) \cdot e^{j2\pi f_0 t}\} \quad \text{Gl. 2.27}$$

bzw.

$$s_T(t) = 2 \cdot s(t) \cdot e^{-j2\pi f_0 t}. \quad \text{Gl. 2.28}$$

f) Der Vergleich zwischen Faltungs- und Korrelationsprodukt

Die direkte Gegenüberstellung der Integraldarstellung des Faltungsproduktes

$$s(\tau) * g(\tau) = \int_{-\infty}^{+\infty} s(t) \cdot g(-t + \tau) dt \quad \text{Gl. 2.29}$$

und des Korrelationsproduktes

$$s(\tau) \otimes g(\tau) = \int_{-\infty}^{+\infty} s(t) \cdot g(t + \tau) dt = \varphi_{sg}(\tau) \quad \text{Gl. 2.30}$$

verdeutlicht eine scheinbare Übereinstimmung dieser zwei Ausdrücke. Eine Substitution $t = -u$ ergibt in $\varphi_{sg}(\tau)$:

$$\begin{aligned} \varphi_{sg}(\tau) &= \int_{-\infty}^{+\infty} s(-u) \cdot g(\tau - u) \cdot (-du) \\ &= \int_{-\infty}^{+\infty} s(-u) \cdot g(\tau - u) \cdot du \quad . \end{aligned} \quad \text{Gl. 2.31}$$

Damit besteht ein formaler Zusammenhang, wobei verallgemeinert gilt:

$$s(\tau) \otimes g(\tau) = s(-\tau) * g(\tau) \quad \text{Gl. 2.32}$$

bzw.:

$$s(\tau) * g(\tau) = s(-\tau) \otimes g(\tau). \quad \text{Gl. 2.33}$$

Das ineinander Überführen von Korrelation und Faltung hat aber seine Grenzen, da sich nicht alle Rechengesetze der Faltung (wie Assoziativ- und Kommutativgesetz) auf die Korrelation anwenden lassen. Da sich die Faltung auf die Betrachtung von LTI-Systemen bezieht, wird ein Vertauschen der beiden Funktionen s und g keine Änderung des Ergebnisses bewirken. Der Grund hierfür findet sich in der für LTI-Systeme geforderten Kausalität, also der Betrachtung von Ursache und Wirkung. Ein

Vertauschen von s und g ändert im Faltungsprodukt nicht die Blickrichtung entlang der Zeitachse. Ein Vertauschen im Korrelationsprodukt hingegen führt zu einer solchen Änderung der Blickrichtung, womit das unterschiedliche Verhalten begründet werden kann. Im komplexen Fourier-Bereich lässt sich dieser Zusammenhang folgendermaßen darstellen:

$$s * g \begin{array}{c} \xrightarrow{F} \\ \xleftarrow{F^{-1}} \end{array} S \cdot G \quad \text{Gl. 2.34}$$

$$s \otimes g \begin{array}{c} \xrightarrow{F} \\ \xleftarrow{F^{-1}} \end{array} S \cdot G^* \quad \text{Gl. 2.35}$$

Deutlich ist hier zu erkennen, dass ein Vertauschen von s und g (bzw. S und G) in der Korrelation zu einer Richtungsänderung führt und so dem Prinzip der Symmetrie (**Kap. 2.2.3**) widerspricht. Folglich muss das Vertauschen in der Korrelation unterschiedliche Ergebnisse liefern.

Diese Diskussion lässt sich aber auch auf einem weiteren Weg führen, indem man auf das Superpositionsprinzip bei der Herleitung des Faltungstheorems näher eingeht (**Gl. 2.14**). Hiernach kann der multiplikative Ausdruck innerhalb des Faltungsintegrals als ein „unechtes Produkt“ verstanden werden, da das resultierende Ausgangssignal als eine Summe vieler unterschiedlich gewichteter Stoßantworten verstanden werden kann. Diese eindeutige Gewichtung ist eine direkte Folge des angelegten Eingangssignals, womit einerseits das lineare Verhalten und andererseits die geforderte Kausalität in LTI-Systemen zum Tragen kommen. Hierbei ist die Kausalität klar und deutlich innerhalb des Faltungsintegrals im gegensätzlichen Vorzeichen des Parameters t in den beiden Funktionen s und g verankert. Auf die Korrelation hingegen lässt sich das Prinzip der Superposition nicht anwenden, da zu einem gegebenen Eingangssignal s sich kein eindeutiges Ausgangssignal zuordnen lässt. Damit stellt ein Korrelator im Sinne der Systemtheorie ein nichtlineares System dar. Vielmehr kann die Funktion g nicht mehr als eine vollständige Beschreibung eines Systems verstanden werden. Vielmehr handelt es sich um ein weiteres Eingangssignal, welches mit s multiplikativ verknüpft ist. Man spricht hier also von einem „echten Produkt“, womit sich der Kausalzusammenhang nicht mehr herstellen lässt, da das Ausgangssignal nun von zwei unbekanntem

Funktionen s und g abhängt. Aus einem geänderten Eingangssignal lässt sich somit die Änderung am Eingang nicht mehr eindeutig ableiten. So hat sich für g gänzlich die Blickrichtung gedreht, was sich im gleichartigen Vorzeichen des Parameters t von s und g innerhalb des Korrelationsintegrals manifestiert.

Zusammenfassend lässt sich also sagen, dass:

1. ein formaler mathematischer Ausdruck zu absolut unterschiedlichen Ergebnissen führen kann, wenn man sie in verschiedenen Theoriemodellen anwendet, da ihnen unterschiedliche Annahmen, Vereinfachungen, Abstraktionen usw. zugrunde liegen;
2. der scheinbare Zusammenhang zwischen Faltung und Korrelation nicht besteht und eine klare Trennung beider Begriffe notwendig und vernünftig ist.

g) Der Zusammenhang zwischen Matched Filter und Korrelationsempfang

Eine in vielen Anwendungen verwendete Art der Signalverarbeitung ist die des „matched filtering“ (engl. matched - angepasst). Dieses oftmals auch als Optimalfilter oder Optimalempfänger¹⁰ bezeichnete Prinzip ist durch eine Pulsantwort charakterisiert, die dem Eingangssignal gleich und zeitgespiegelt ist:

$$h(t) = s(-t). \quad \text{Gl. 2.36}$$

Dieser Ausdruck stellt durch das entgegengesetzte Vorzeichen im Parameter t eindeutig einen nichtkausalen Zusammenhang dar, wodurch der Begriff des Filters hier leicht missverstanden kann. Es ist deshalb sicherer den Begriff des Korrelationsempfängers zu verwenden.

Der Grund, weshalb der Begriff der Korrelation hier zutrifft, findet sich in der Faltung von Stoßantwort $h(t)$ und empfangenem Signal $r(t)$. Da das empfangene Signal $r(t)$ dem zeitverzögerten Sendesignal $s(t-\tau)$ entspricht, gilt:

¹⁰ vgl. H. D. Lüke; Signalübertragung: Grundlagen der digitalen und analogen Nachrichtenübertragungssysteme; Springer Verlag Berlin 1995, S. 136

$$r(t) = s(t - \tau) = s(t) * \delta(t - \tau). \quad \text{Gl. 2.37}$$

Die Faltung dieses Signals führt damit zu:

$$g(t) = r(t) * h(t) = s(t) * \delta(t - \tau) * s(-t). \quad \text{Gl. 2.38}$$

Mit der in **Gl. 2.32** dargestellten Beziehung zwischen Faltung und Korrelation folgt:

$$g(t) = \varphi_{ss} * \delta(t - \tau) \quad \text{Gl. 2.39}$$

oder in exponentieller Form:

$$g(t) = \varphi_{ss} \cdot e^{i\varphi}. \quad \text{Gl. 2.40}$$

Aufgrund dieses Zusammenhangs ist beim Korrelationsempfänger zu beachten, dass hier auf die Impulsautokorrelationsfunktion zurückgegriffen wird. Da sich dadurch wieder die Blickrichtung entlang der Zeitachse ändert (**Gl. 2.38**), ist hier darauf zu achten, dass nicht alle Rechengesetze der Faltungsalgebra angewendet werden dürfen.

2.5.2 Erweiterung zur mehrdimensionalen Systemtheorie

Die eindimensionale Betrachtung stößt sehr schnell an ihre Grenzen, wenn ein System beispielsweise von mehreren Eingangsgrößen abhängt. Mit jeder zusätzlichen Eingangsgröße hat die systemtheoretische Betrachtung mit einer zusätzlichen Dimension umzugehen. Während ein eindimensionales Messsystem für einen Messwert genau einen Vorgänger und einen Nachfolger besitzt, also durch eine Anzahl N direkter Nachbarn von nur $N=2$ charakterisiert ist, erhöht sich diese Zahl mit jeder weiteren Dimension erheblich. Für eine beliebige Anzahl D an Dimensionen ergibt sich folgende Anzahl an Nachbarwerten:

$$N = 3^D - 1 \quad \text{Gl. 2.41}$$

Für ein Bild aus einer herkömmlichen 2D-Kamera ergeben sich für jeden einzelnen Farbwert schon acht Nachbarwerte, mit der Betrachtung einer zeitlichen Bildfolge sogar schon 26. Eine PMD-basierte 3D-Kamera stellt nun ein 5-dimensionales Problem dar, da neben den Raumkoordinaten x , y und z auch der Grauwert G und der Verlauf über die Zeit t betrachtet werden können. Hier existieren nun für jeden Messwert 242 Nachbarn, was die enorme Komplexität der Beschreibung einer solchen Kamera verdeutlicht. Für die Betrachtung solcher mehrdimensionaler Systeme ist nun die Erweiterung der eindimensionalen Systemtheorie zwingend erforderlich.

Ausgehend von einem eindimensionalen System stellt die Sprungantwort $h(t)$ im Zeitbereich oder der Übertragungsfunktion $H(f)$ im Frequenzbereich eine ausreichende Systembeschreibung dar. Für ein mehrdimensionales Signal gilt dagegen in vektorieller Schreibweise:

$$s(\underline{t}) = s(t_1, t_2, \dots, t_n) \quad \text{mit: } \underline{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}. \quad \text{Gl. 2.42}$$

Daraus kann man wieder mehrere eindimensionale Signale gewinnen, indem man jeweils alle Dimensionen bis auf eine konstant hält. Die Stoßantwort eines mehrdimensionalen Systems ist ebenso eine mehrdimensionale Funktion, welche sich ebenso vektoriell darstellen lässt:

$$h(\underline{t}) = h(t_1, t_2, \dots, t_n) \quad \text{mit: } \underline{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_n \end{bmatrix}. \quad \text{Gl. 2.43}$$

a) Das mehrdimensionale Faltungstheorem

Formell kann die Faltung mehrdimensionaler Systeme in vektorieller Form dargestellt werden als:

$$g(\underline{t}) = s(\underline{t}) * h(\underline{t}). \quad \text{Gl. 2.44}$$

Die mehrdimensionale Faltung in Integralform lautet:

$$g(\underline{t}) = \int_{-\infty}^{\infty} s(\underline{\tau}) \cdot h(\underline{t} - \underline{\tau}) d\underline{\tau}$$

Gl. 2.45

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} s(\tau_1, \tau_2, \dots, \tau_n) \cdot h(t_1 - \tau_1, t_2 - \tau_2, \dots, t_n - \tau_n) \cdot d\tau_1 \cdot d\tau_2 \cdot \dots \cdot d\tau_n .$$

Zu beachten ist hierbei, dass sich die Einzelintegrale hintereinander lösen lassen, unabhängig von ihrer Reihenfolge.

b) Die mehrdimensionale Fourier-Transformation / Hankel-Transformation

Durch die Einführung der vektoriellen Schreibweise ergibt sich für die mehrdimensionale Fourier-Transformation:

$$S(\underline{f}) = \int_{-\infty}^{\infty} s(\underline{t}) \cdot e^{-j2\pi \underline{f}^T \underline{t}} d\underline{t}$$

Gl. 2.46

$$= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} s(t_1, t_2, \dots, t_n) \cdot e^{-j2\pi f_1 t_1} \cdot e^{-j2\pi f_2 t_2} \dots \cdot e^{-j2\pi f_n t_n} dt_1 dt_2 \dots dt_n .$$

Die auftretenden Einzelintegrale der mehrdimensionalen Fourier-Transformation können generell schrittweise gelöst werden, wobei die Reihenfolge keinen Einfluss auf das Gesamtergebnis besitzt.

Die Lösung vieler mehrdimensionaler Problemstellungen lässt sich jedoch noch durch das gezielte Ausnutzen von Symmetrieeigenschaften vereinfachen. Beispielsweise besitzt eine radialsymmetrische Funktion der Form

$$s(x, y) \equiv s(r) \quad \text{mit } r = \sqrt{x^2 + y^2}$$

Gl. 2.47

eine Fourier-Transformierte, die ebenfalls radialsymmetrisch ist:¹¹

$$S(u, v) \equiv S(q) \quad \text{mit } q = \sqrt{u^2 + v^2}. \quad \text{Gl. 2.48}$$

Dies wird ersichtlich, wenn man die zweidimensionale Fourier-Transformation des kartesischen Koordinatensystems

$$S(u, v) = \int_{-\infty-\infty}^{+\infty+\infty} \int_{-\infty-\infty}^{+\infty+\infty} s(x, y) \cdot e^{-j2\pi(ux+vy)} dx dy \quad \text{Gl. 2.49}$$

in das Polarkoordinatensystem überführt:

$$S(q) = \int_0^{+\infty} \int_0^{+2\pi} s(r, \theta) \cdot e^{-j2\pi q r \cos(\theta)} r dr d\theta. \quad \text{Gl. 2.50}$$

Aufgrund der radialen Symmetrie der Funktion s_r ist dieser Ausdruck nun unabhängig von der Variablen θ , so dass er sich vereinfacht zu:

$$\begin{aligned} S(q) &= \int_0^{+\infty} s(r) \left[\int_0^{2\pi} e^{-j2\pi q r \cos(\theta)} d\theta \right] r dr \\ &= 2\pi \int_0^{+\infty} s(r) J_0(2\pi q r) r dr \end{aligned} \quad \text{Gl. 2.51}$$

Hierbei ist J_0 eine gewöhnliche Besselfunktion der ersten Art und der Ordnung 0:

$$J_0(z) = \frac{1}{2\pi} \int_0^{2\pi} e^{-jz \cos(\beta)} d\beta. \quad \text{Gl. 2.52}$$

Die Gleichung 2.51 wird als Fourier-Bessel-Transformation oder auch Hankel-Transformation bezeichnet, deren Inverse wiederum lautet:

¹¹ R. Bracewell; „The fourier transform and its applications“, Second edition, revised, McGraw-Hill, New York, 1986

$$s(r) = 2\pi \int_0^{+\infty} S(q) J_0(2\pi qr) q dq .$$

Gl. 2.53

Durch diesen Zusammenhang reduziert sich die zweidimensionale Fourier-Transformation einer radialsymmetrischen Funktion auf eine eindimensionale Hankel-Transformation. Ein typischer Anwendungsbereich dieser Vereinfachung liegt im Bereich optisch abbildender Systeme. So lassen sich mit ihrer Hilfe die rotationssymmetrischen Abbildungsfehler einer Kameraoptik recht einfach korrigieren.

3 Systemkomponenten der optischen Sensorik und Signalübertragung

3.1 Elektrooptischer Effekt

Die Grundlage jeder optischen Signalübertragung ist die Ausnutzung des elektrooptischen Effekts, sei es um ein elektrisches Signal auf die Intensität eines Lichtsenders abzubilden oder umgekehrt ein optisches Signal zu detektieren und in ein elektrisches Signal zu wandeln. Der als „elektrooptischer Effekt“ bezeichnete Prozess lässt sich auf atomarer Ebene durch die Wechselwirkung zwischen Elektronen und Photonen beschreiben. Diese Wechselwirkung wurde 1887 durch Heinrich Hertz erstmals bei seinen Untersuchungen zur elektrischen Entladung beobachtet und beschrieben.¹ Nach weiteren Experimenten durch Phillip Eduard Anton von Lenard gelang die Bestimmung des Ladungs-Massen-Verhältnisses der durch Lichteinstrahlung freigesetzten Teilchen. Die durch das ultraviolette Licht einsetzende Funkenverstärkung war eine Folge der durch die Einstrahlung zusätzlich freigesetzten Elektronen, wobei hier erstmals der Begriff der Photoelektronen geprägt wurde. Das Ladungs-Massen-Verhältnis lässt sich bestimmen, indem man den emittierenden Photoelektronen eine zusätzliche Kraft durch ein elektrisches Feld entgegenstellt. Die minimale Gegenspannung, die zum Aufbau des elektrischen Feldes benötigt wird und ein Erreichen der Detektorfläche durch die Photoelektronen gerade noch verhindert, wird als Grenzspannung V_0 bezeichnet. Mit ihr gelang eine energetische Beschreibung der emittierten Teilchen:

$$\frac{1}{2} \cdot m_0 \cdot v_{\max}^2 = q_e \cdot V_0. \quad \text{Gl. 3.1}$$

Ein weiteres Ergebnis der Untersuchungen zum elektrooptischen Effekt stellte die Wissenschaftswelt vor ein ernstes Problem. Lenard zeigte, dass diese Grenzspannung unabhängig von der Strahlungsintensität des einfallenden Lichts ist, was im Widerspruch zur klassischen elektromagnetischen Theorie steht. Daraus konnte man lediglich schließen, dass sich die maximale Energie nur mit der Lichtwellenlänge ändert und in keinem Zusammenhang mit der einfallenden Gesamtenergie steht. Eine Lösung

¹ vgl. Heinrich Hertz; Gesammelte Werke, Bd. II, Über einen Einfluss des ultravioletten Lichtes auf die elektrische Entladung; Leipzig, 1914, S. 69

für dieses Problem war Max Plancks Ansatz, wonach sich die Energie des Feldes nur mit ganzzahligen Vielfachen von $h\nu$ ändern kann. Einstein ging mit dieser Idee der Quantisierung noch kühner vor, indem er das Strahlungsfeld selbst quantisierte und ihm einen gewissen Teilchencharakter zugestand. Somit quantisierte er nicht nur die Energie des Feldes, sondern gab diesem auch einen örtlichen Charakter, was später zu dem Begriff Photon führte. Mit dieser Abstraktion wurde der elektrooptische Effekt erklärbar. Ein emittiertes Elektron ist in seiner kinetischen Energie durch die Differenz aus absorbiertes Lichtquantenergie $h\nu$ des Photons und materialabhängiger Austrittsarbeit Φ genau festgelegt:

$$\frac{1}{2} \cdot m_0 \cdot v^2 = h \cdot \nu - \Phi . \quad \text{Gl. 3.2}$$

Wenn nun die Austrittsarbeit materialabhängig diskret festliegt, kann eine Änderung der eingestrahlten Lichtintensität nur zu einer Änderung der emittierten Elektronenanzahl führen. Die oben eingeführte Grenzspannung V_0 bleibt immer gleich. Diese Überlegungen weisen darauf hin, dass die noch im Material gebundenen Elektronen sich in ebenso diskreten Energiezuständen befinden müssen. Der hier andeutungsweise dargestellte Effekt der Wechselwirkung zwischen Licht und Material bildet die Basis für das Verständnis optischer Systeme. Für die systemtheoretische Beschreibung der grundlegenden Komponenten optischer Systeme ist die physikalische Betrachtung so genannter Halbleiter unabdingbar, da die wichtigsten optischen Sende- und Empfangselemente auf dieser Technologie beruhen. Hierbei liegt jedoch nicht wie bei den frühen Experimenten zum elektrooptischen Effekt das Hauptaugenmerk auf dem Austritt von zuvor materiell gebundenen Elektronen in das umgebende Vakuum. Vielmehr werden hier Zustandsänderungen innerhalb des Materials betrachtet, welche auf den Wechselwirkungen zwischen Licht in Form der Photonen und den Ladungsträgern beruhen, was bedeutet, dass hier die materialspezifische Austrittsarbeit Φ an Bedeutung verliert. Deshalb spricht man hier auch vom inneren elektrooptischen Effekt. Von großer Bedeutung sind hier die diskreten Energiezustände der atomar gebundenen Elektronen innerhalb eines Halbleiters. Ein Bild solcher Energiezustände gewinnt man durch die modellhafte Betrachtung zweier Wasserstoffatome, welche anfänglich sehr weit voneinander entfernt sind, so dass keinerlei Wechselwirkungen zwischen ihnen stattfinden. Die Ladungsdichte eines solchen isolierten

Wasserstoffatoms (**Abb. 3.1a**) unterliegt einer einfach zu beschreibenden sphärischen Symmetrie. Jedes Atom in diesem wechselwirkungsfreien Zustand besitzt eine bestimmte Energie E_0 , so dass die Gesamtenergie des Systems $2E_0$ ist. Wird der Abstand zwischen beiden Atomen soweit verringert, dass die eintretenden Wechselwirkungen auf die Ladungsverteilung Einfluss nehmen, so ergibt sich als Gesamtenergie des Systems:

$$E = 2 \cdot E_0 + u(R). \quad \text{Gl. 3.3}$$

Die Größe $u(R)$ beschreibt hier die mittlere Energie der Coloumb-Wechselwirkungen zwischen den einzelnen Ladungen. In der klassischen Betrachtung ist sie immer positiv und kann nur zu einer Abstoßung beider Wasserstoffatome führen (**Abb. 3.1b**). Die Existenz eines H_2 -Moleküls kann so nicht erklärt werden, jedoch gelingt dies durch eine quantenmechanische Betrachtung. Hier ergibt sich die mittlere Energie der Wechselwirkungen aus der Summation zweier Terme, worin einer der beiden der klassischen Form entspricht und somit immer positiv ist. Der zweite als Austauschenergie bezeichnete Term kann unterschiedliche Vorzeichen besitzen und ist abhängig von der relativen Orientierung der Elektronenspins. Sind die Spins zweier Elektronen antiparallel, hat die Größe $u(R)$ ein Minimum, wodurch die Bildung eines stabilen H_2 -Moleküls möglich wird (**Abb. 3.2**).

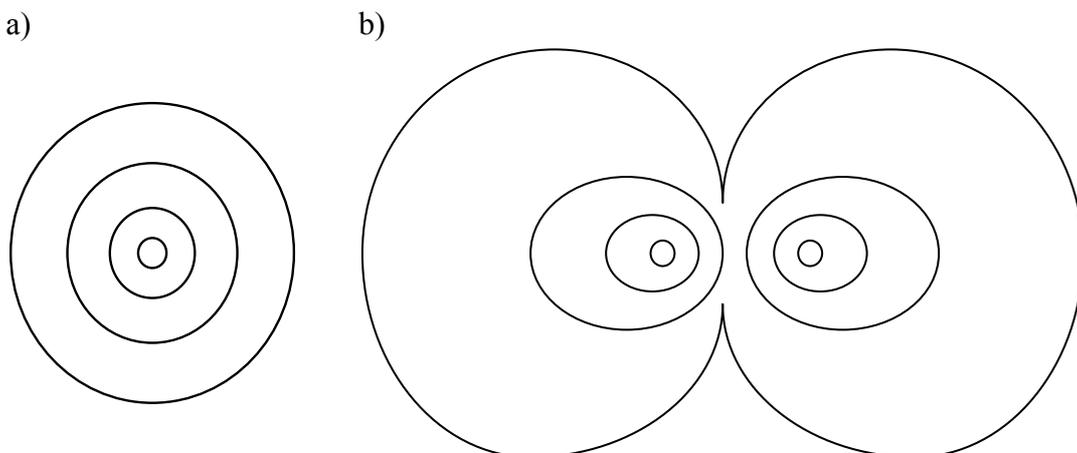


Abb. 3.1: Schematische Darstellung der Ladungsträgerdichte a) eines einzelnen Wasserstoffatoms und b) zweier im klassischen Modell wechselwirkender Wasserstoffatome

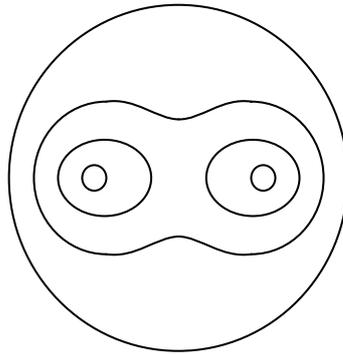


Abb. 3.2: Schematische Darstellung der Ladungsträgerdichte eines H_2 -Wasserstoffmoleküls

Eine wesentliche Besonderheit dieser so genannten homöopolaren Bindung ist ihre räumliche Ausrichtung, die sich durch die gemeinsame Elektronenwolke mit asphärischem Charakter ergibt. Mit dieser einfachen Betrachtung zweier Wasserstoffatome kann die räumliche Abhängigkeit der Ladungsträgerenergie veranschaulicht werden, jedoch gelingt damit nicht die Erklärung diskreter Energiezustände. Diese findet sich in de Broglies Hypothese, dass jedes Teilchen ebenso Wellencharakter besitzt. Sein Vorschlag, jedem Teilchen mit dem Impuls p durch

$$\lambda = \frac{h}{p} \quad \text{Gl. 3.4}$$

eine Wellenlänge zuzuordnen, fand 1927 durch Davisson und Germer seine Bestätigung. Betrachtet man damit das um den Wasserstoffkern kreisende Elektron als Welle, so ergeben sich Radien bzw. so genannte Orbitale, auf denen sich diese Welle durch destruktive Interferenz selbst zerstört. Ein Elektron kann also auf einem solchen Orbit nicht existieren. Da der Impuls und damit auch die Wellenlänge den Coulombkräften unterliegen, ergeben sich für Anordnungen aus mehreren Atomen verbotene Energiebereiche durch die gegebene Kristallstruktur. Orientiert man sich an der Kristallstruktur des Halbleiters und definiert den sogenannten Wellenvektor k als

$$k = \frac{p}{\hbar} \quad \text{mit} \quad \hbar = \frac{h}{2\pi}, \quad \text{Gl. 3.5}$$

dann erhält man indirekt eine Abhängigkeit der Energie vom Ladungsträgerimpuls. Orientiert man sich mit dem Wellenvektor an den kristallografischen Richtungen des Halbleiters, dann ergeben sich die als Bandstruktur bezeichneten Graphen (**Abb. 3.3**). Für die Erklärung der elektrooptischen Wechselwirkungen wie Photonenemission und -absorption ist die Bandstruktur der verschiedenen Halbleiter von entscheidender Bedeutung. Von weiterer Bedeutung sind hierbei die Begriffe Valenz- und Leitungsband. Das Leitungsband befindet sich in der Bandstruktur oberhalb der verbotenen Zone und enthält alle Elektronen, die zur elektrischen Leitung beitragen können. Das unterhalb der verbotenen Zone befindliche Valenzband enthält die an das Kristallgitter gebundenen Elektronen, die aufgrund ihrer Bindung nicht zu einem elektrischem Strom beitragen können.

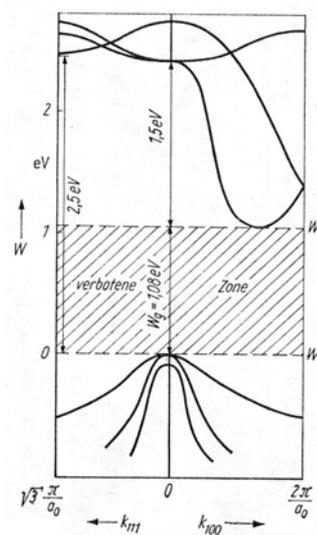


Abb. 3.3: Bandstruktur von Silizium²

² A. Möschwitzer, „Halbleiterelektronik“, Hüthig Verlag, Heidelberg 1988, S. 25

3.2 Prozess der Photonenemission

Da ein Elektron mit hohem Impuls statistisch selten ist, sammeln sich diese in der Nähe der Minima der Bandstruktur. In den Maxima der Bandverläufe ergeben sich somit Zustände, die von Elektronen besetzt werden könnten und als Löcher bezeichnet werden. Diese besitzen eine entsprechend entgegengesetzte positive Ladung. In einem Halbleiter im sogenannten thermodynamischen Gleichgewicht ergibt sich ein Gleichgewicht zwischen Ladungsträgererzeugung (Generation) und -vernichtung (Rekombination). Dieses Gleichgewicht zeichnet sich durch konstante Ladungsträgerdichten n_0 und p_0 aus und die Raten der ständig stattfindenden Generation und Rekombination sind gleich. Trifft nun ein Elektron mit einem Loch zusammen, so kann es sich nur durch Energieabgabe W_r mit ihm vereinigen, was einem Überwinden der verbotenen Zone gleichkommt. Die Breite der verbotenen Zone legt hierbei die Größe der abzugebenden Energie fest. Liegen das Maximum des Valenzbandes und das Minimum des Leitungsbandes direkt gegenüber (**Abb. 3.4**), dann handelt es sich um sogenannte direkte Halbleiter und eine Rekombination erfolgt ohne eine wesentliche Änderung des Impulses, da sich der Wellenvektor k für beide Zustände gleich. Die Energieabgabe erfolgt wiederum durch eine elektromagnetische Welle, der nach de Broglie wieder ein Teilchen zugeordnet wird.

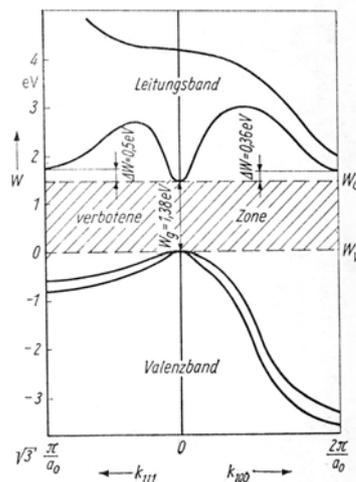


Abb. 3.4: Bandstruktur eines direkten Halbleiters (GaAs)³

³ A. Möschwitzer u. K. Lunze, „Halbleiterelektronik“, Hüthig Verlag Heidelberg 1988, S. 25

Ein solches Teilchen muss aufgrund des direkten Überganges einen extrem kleinen Impuls besitzen. Der Impuls eines solchen als Photon bezeichneten Teilchens beträgt

$$p_{\text{Photon}} = \frac{h}{\lambda} \quad \text{Gl. 3.6}$$

und steht mit der vom Elektron abgegebenen Energie in folgender Beziehung:

$$E_{\text{Photon}} = W_r = c \cdot p. \quad \text{Gl. 3.7}$$

In einem indirekten Halbleiter mit zueinander versetzten Minimum und Maximum in der Bandstruktur (**Abb. 3.3**) muss eine solche Rekombination auch unter einer erheblichen Impulsänderung erfolgen. Diese Impulsabgabe kann nicht durch ein Photon erfolgen, sondern muss einem weiteren Teilchen zugeordnet werden. Diesem als Phonon bezeichneten Teilchen ist ebenso eine elektromagnetische Welle zugeordnet, die als Gitterschwingungen betrachtet wird und sich als Wärmestrahlung äußert. Ihre sehr geringe Energie ergibt sich aus der Gitterkonstanten a_0 und der Schallgeschwindigkeit c_S :

$$E_{\text{Phonon}} = \frac{h \cdot c_S}{2 \cdot a_0}. \quad \text{Gl. 3.8}$$

Die Impulsänderung hingegen ist:

$$\Delta p_{\text{Elektron}} = p_{\text{Photon}} = \hbar \cdot \Delta k = \frac{h}{2 \cdot a_0}. \quad \text{Gl. 3.9}$$

Diese indirekte Rekombination mit Phononenwechselwirkung muss über Zwischenniveaus erfolgen, welche im verbotenen Bereich der Bandstruktur liegen und als Rekombinationszentren bezeichnet werden. Da der bisher beschriebene Rekombinationsprozess statistischen Charakter besitzt, wird er auch als spontane Photonenemission bezeichnet. Dem gegenüber steht die stimulierte oder auch induzierte Emission. Hierbei ist der Wellencharakter des Photons von entscheidender Bedeutung,

denn die Betrachtung als elektromagnetische Welle lässt wieder Interferenzeffekte zu. Dies beinhaltet neben der Auslöschung auch die Verstärkung von Wellenzügen gleicher Phasenlage. Die stimulierte Emission ist also eine durch eine elektromagnetische Welle erzwungene Rekombination. Auf diesen Spezialfall wird später bei der Betrachtung von Halbleiterlasern noch konkreter eingegangen.

3.3 Prozess der Photonenabsorption

Dem Fall der Rekombination steht die Ladungsträgergeneration gegenüber. Hierbei ist dem Halbleitermaterial in der Art Energie zuzuführen, dass Valenzelektronen in das Leitungsband wechseln können. Dies kann auf unterschiedlichen Wegen erfolgen, wie z.B. durch thermische Effekte oder durch das Anlegen eines ausreichend starken elektrischen Feldes. Für optische Systeme entscheidend ist hierbei jedoch der Einfluss einer Belichtung. Hier erfolgt die für den Bandwechsel notwendige Energiezufuhr durch eine Photoneinwirkung. Dabei handelt sich hierbei grundsätzlich um die Umkehrung des inneren elektrooptischen Effekts.

Generell hängt die Generation freier Ladungsträger von vielen Faktoren ab. Aufgrund des Quantencharakters des Lichts besteht auch hier der Zusammenhang zwischen verbotener Zone und notwendiger Lichtwellenlänge. Die Generationsrate hängt darüber hinaus auch von der Anzahl einfallender Lichtquanten, der Photonenflussdichte Φ ab. Jedoch wird nicht jedes Photon mit ausreichend Energie Ladungsträger freisetzen, da bei der Anhebung eines Elektrons Impuls- und Energieerhaltungssatz erfüllt sein müssen. Trifft auf eine Halbleiteroberfläche ein Photonenstrom der Dichte Φ , dann unterliegt ein Teil dem Prozess der Reflexion R , was den in das Halbleitermaterial tatsächlich eindringenden Photonenstrom reduziert. Die Generationsrate als Funktion der Materialtiefe x lautet damit:

$$G(x) = \alpha \cdot \Phi \cdot (1 - R) \cdot e^{-\alpha x} . \quad \text{Gl. 3.10}$$

Hierbei ist α der Absorptionskoeffizient, der wiederum eine von der Lichtwellenlänge abhängige Größe darstellt.

3.4 Modulierbare Halbleitersendeelemente

3.4.1 Lichtemitterdioden

Der Prozess der Photonenemission kann in einem Halbleitermaterial einfach dadurch in Gang gesetzt werden, dass in einem in Flussrichtung gepolten pn^+ -Übergang eines direkten Halbleiters Elektronen von der n^+ -Seite in den p-Bereich diffundieren. Solche einfachen Strukturen nutzen nur die rein spontan strahlende Rekombination, welche in der direkten Umgebung des pn^+ -Übergangs erfolgt (**Abb. 3.5**). Die aus dem n-Bereich kommenden Elektronen treffen erst hier auf eine ausreichend große Anzahl an Löchern. Diese wiederum bewegen sich vom p-Bereich in Richtung des n-Bereichs. Hier treffen sie im Bereich des Übergangs auf die gegenläufigen Elektronen. Die während der Rekombination stattfindende Energieänderung entspricht der Differenz zwischen Valenz- und Leitungsbandkante:

$$E_{\text{photon}} = W_C - W_V = e \cdot U = h \cdot f. \quad \text{Gl. 3.11}$$

Grundsätzlich zeigt eine Lichtemitterdiode das Verhalten einer einfachen Diode. Liegt an ihr eine äußere Spannung an und erreicht diese die charakteristische Flussspannung, dann wird der Potentialwall für Elektronen und Löcher gleichermaßen abgebaut. Sie können jeweils in das benachbarte Gebiet als Minoritätsträger diffundieren und rekombinieren. Jedoch ist die Energieabgabe während der Rekombination auch in Form eines Phonons möglich.

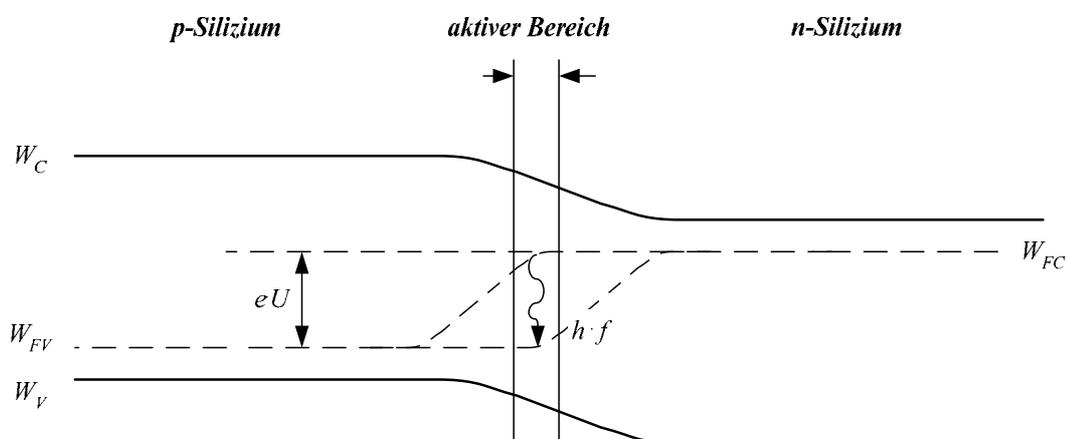


Abb. 3.5: Energieniveauschema einer Lumineszenzdiode; pn^+ -Übergang in Durchlassrichtung

Welche Art überwiegt, entscheidet der Impulserhaltungssatz, genau wie im umgekehrten Prozess der Photonenabsorption (**Kap. 3.2**). Hierzu muss die Differenz der Impulse der beiden Rekombinationsteilnehmer Elektron und Loch gleich dem des emittierten Photons sein. Da Photonen nur einen sehr kleinen Impuls besitzen, kann bei der Emission eines Photons der Impulserhaltungssatz nur erfüllt werden, wenn ein direkter Bandübergang gelingt. Hierzu muss in der Darstellung der Bandstruktur ein Minimum im Leitungsband einem Maximum im Valenzband direkt gegenüberliegen (**Abb. 3.4**). Liegt dagegen kein direkter Halbleiter vor, wie z.B. bei Silizium (**Abb. 3.3**), dann kann die Energieabgabe nur in Verbindung mit großen Impulsänderungen erfolgen, was ausschließlich zu einer Emission von Phononen bzw. Wärme führt. In einem direkten Halbleiter sind jedoch beide Möglichkeiten gegeben, so dass hier Photonen und Phononen abgegeben werden können. Interessant ist hierbei jedoch das Verhalten bei sehr hohen Flussspannungen bzw. Strömen. Hier erreichen nicht alle Ladungsträger den direkten Übergang, da dieser noch durch einen Vorgänger besetzt ist. Für diese Ladungsträger verbleibt so nur die Möglichkeit indirekt zu rekombinieren, was zwangsläufig mit einer großen Impulsänderung bzw. Phononenabgabe verbunden ist. Aufgrund dieses Zusammenhangs ist die erreichbare optische Leistung einer LED-Struktur begrenzt. Während bei geringen Strömen noch lineares Verhalten zwischen Flusstrom und abgestrahlter Lichtintensität besteht, nähert sich diese bei großen Strömen einem Maximum. Jede weitere Stromerhöhung kann hier nun nicht mehr zu einer Photonenemission führen. Stattdessen wird immer mehr Leistung thermisch umgesetzt, was den Aufbau leistungsstarker LED-Sender stark eingrenzt.

3.4.2 Laserdioden

Wie in den vorangegangenen Betrachtungen angedeutet ist neben der spontanen Emission ebenso eine stimulierte bzw. induzierte Emission möglich. Diese Möglichkeit wurde bereits 1917 von Einstein beschrieben.⁴ Grundvoraussetzung hierfür ist, dass das stimulierende Photon auf einen emissionsbereiten Rekombinationszustand trifft, was nur mit einer gewissen Wahrscheinlichkeit eintritt. Das nun zusätzlich emittierende Photon ist in Energie, Ausbreitungsrichtung, Polarisation und Phase vollständig

⁴ Albert Einstein; Phys. Z. 18 (1917)6, Zur Quantentheorie der Strahlung; S. 121

identisch. Das stimulierende Photon erfährt hierbei keine Änderung und bleibt damit von der Induktion unbeeinflusst. Die Stimulation eines weiteren Photons lässt sich mit dem Begriff der Verstärkung beschreiben, die sich im englischen Begriff LASER (light amplification by stimulated emission of light) wieder findet. Dieser beschreibt eine technische Anordnung zur gezielten Ausnutzung dieses Effekts. Hier wird mittels zweier gegenüberliegender Spiegel ein optischer Resonator gebildet, so dass die am Spiegel reflektierte Welle sich mit der einfallenden Welle phasengleich und damit verstärkend überlagert. Mittels der zwei gegenüberliegenden Spiegel wird die Gesamtenergie der durch Mehrfachüberlagerung erzeugten Welle im Resonator derart erhöht, dass induzierte Emission möglich wird. Detaillierter lässt sich dieser Vorgang mit der allgemeinen Wellengleichung in der Form

$$\nabla^2 \varphi + \gamma^2 \psi = 0 \quad \text{Gl. 3.12}$$

beschreiben. Dieser verlustbehaftete Vorgang wird durch die komplexe Ausbreitungskonstante $\gamma = \alpha + i\beta$ charakterisiert. Die Lösung dieser Gleichung für das elektrische Feld entlang einer Koordinate führt mit

$$E_z(x, y, z) = X_x \cdot Y_y \cdot Z_z \quad \text{Gl. 3.13}$$

zu folgendem Differentialgleichungssystem:

$$X'' + X\gamma_x^2 = Y'' + Y\gamma_y^2 = Z'' + Z\gamma_z^2 = 0. \quad \text{Gl. 3.14}$$

Hierbei handelt es sich um eine Differentialgleichung zweiter Ordnung, deren Lösung in Richtung der z-Koordinate lautet:

$$Z(z) = A_z \cdot e^{(-i\gamma_z z)}. \quad \text{Gl. 3.15}$$

Berücksichtigt man nun die Zeitabhängigkeit, dann gilt:

$$Z(z, t) = A_z \cdot e^{(\beta z)} \cdot e^{[i(\alpha z - \omega t)]}. \quad \text{Gl. 3.16}$$

Grundsätzlich kann diese Beschreibung der Ausbreitung einer elektromagnetischen Welle innerhalb eines verlustbehafteten Mediums auch auf Laser angewendet werden. Dazu ist jedoch seine verstärkende Wirkung durch einen Vorzeichenwechsel bei der Ausbreitungskonstante γ zu berücksichtigen:

$$Z(z, t) = A_z \cdot e^{(-\beta z)} \cdot e^{[i(\omega t - \alpha_z z)]} \quad \text{Gl. 3.17}$$

Betrachtet man nun zwei sich verstärkende gegenläufige Wellen A und B innerhalb eines Laseroszillators, dann überlagern sich beide entsprechend dem Superpositionsprinzip. Die Einhüllenden

$$A_z \cdot e^{(-\beta z)} \quad \text{Gl. 3.18}$$

und

$$B_z \cdot e^{(\beta z)} \quad \text{Gl. 3.19}$$

der nach rechts laufenden Welle $A(z)$ und der nach links laufenden Welle $B(z)$ sind qualitativ in **Abb. 3.6** dargestellt. Anhand dieser vereinfachten Annahme einer konstanten Verstärkung wird deutlich, dass die Feldstärke der stehenden Welle sich nicht konstant entlang der z -Koordinate ausbilden kann. In den Gebieten mit hoher Feldstärke werden entsprechend mehr Photonen stimuliert. Diese Gebiete unterliegen somit einer geringeren Inversion, was wiederum die Verstärkung in diesen Gebieten vermindert. Des Weiteren berücksichtig diese klassische wellentheoretische Betrachtung keine statistischen Prozesse. Die stimulierte Emission innerhalb eines Laserresonators unterliegt jedoch verschiedenen Rauschprozessen, was einen erheblichen Einfluss auf die resultierende Welle im Resonator ausübt und das symmetrische Verhalten stören wird. Die Betrachtung solcher statistischer Prozesse und die Kopplung über die Inversion soll später im Zusammenhang mit der Modenstabilität und deren Auswirkungen auf eine Phasenmessung optischer Signale beschrieben werden (**Kap. 4.6.1**).

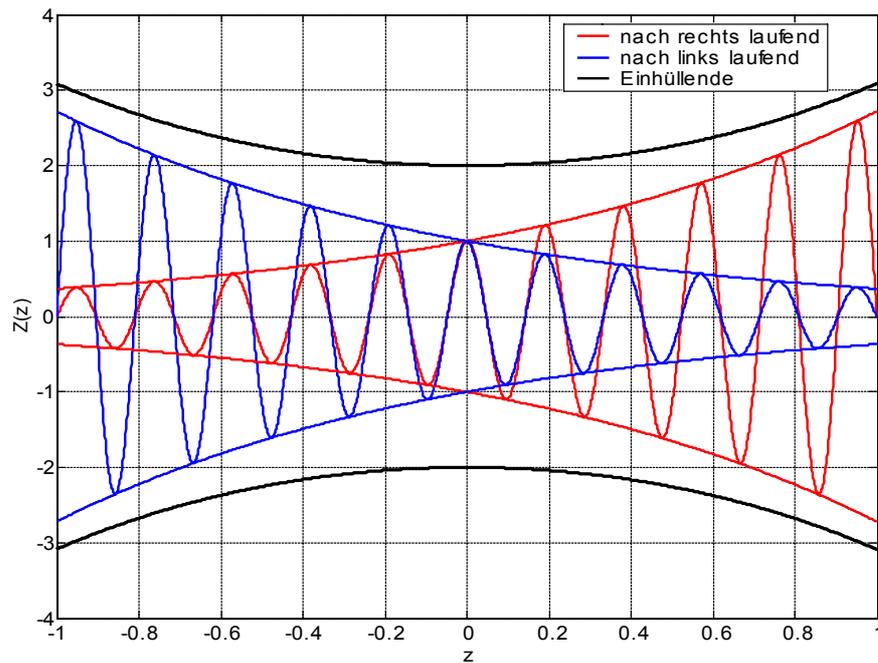


Abb. 3.6: Qualitative Darstellung der vor- und rücklaufenden Welle in einem Laserresonator und der resultierenden Einhüllenden

Ein in der optischen Kommunikations- und Messtechnik weit verbreiteter Laser ist die Halbleiterlaserdiode. Diese auch unter dem Begriff der Injektionslaser zusammengefassten Halbleiterbauelemente werden durch einen pn-Übergang gebildet, der im Design derart gestaltet ist, dass der Prozess der Emission (spontan und stimuliert) dem der Absorption überwiegt. In einer Halbleiterstruktur im thermodynamischen Gleichgewicht halten sich die Prozesse Emission und Absorption im Gleichgewicht. Hier liegt das Quasiferminiveau des Leitungsbandes W_{FL} auf gleicher Höhe zum Quasiferminiveau des Valenzbandes W_{FV} . Wird dieses Gleichgewicht jedoch durch ein äußeres Feld gestört (**Abb. 3.7**), dann kann die stimulierte Emission der Absorption überwiegen und der Halbleiter wirkt verstärkend. Hierzu muss die Bernard-Douraforg-Inversionsbedingung⁵ erfüllt sein:

$$E_{ph} = e \cdot U = h \cdot f < (W_{FL} - W_{FV}). \quad \text{Gl. 3.20}$$

⁵ B. Mroziewicz, M. Bugajski, W. Nakwaski; „Physics of Semiconductor Lasers“, North-Holland 1991, S. 35

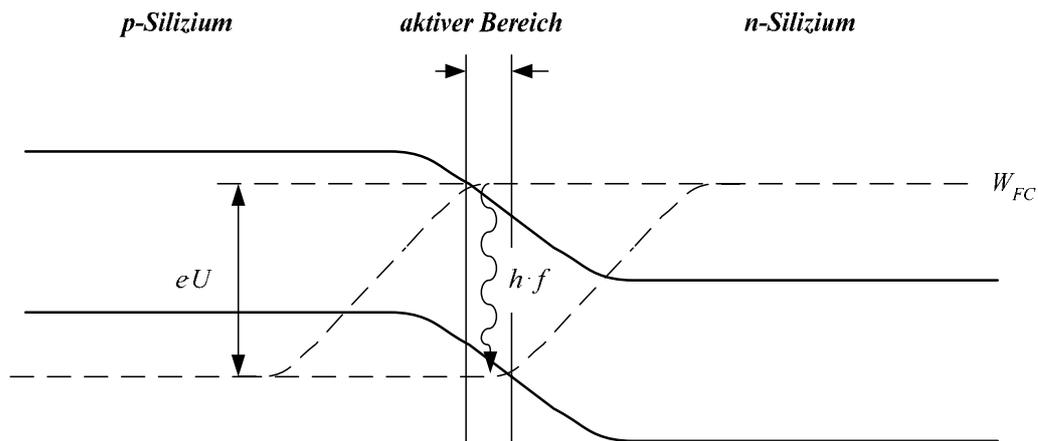


Abb. 3.7: Energieniveauschema eines Halbleiterlasers; entarteter pn-Übergang in Durchlassrichtung

Voraussetzung für eine stimulierte Emission sind Bandübergänge überwiegend strahlender Natur, also direkte Übergänge mit kleinen Impulsänderungen. Dies setzt, wie bei den zuvor beschriebenen LED (**Kap. 3.3.1**), die Verwendung direkter Halbleiter voraus. Befindet sich ein solcher Halbleiter im thermodynamischen Gleichgewicht, dann wird sein absorbierendes Verhalten durch den Absorptionskoeffizienten $\alpha(h \cdot f)$ beschrieben. Wird dieses Gleichgewicht durch ein äußeres elektrisches Feld gestört, was im Allgemeinen auch als Pumpen bezeichnet wird, dann setzt allmählich Inversion ein. Dabei entsteht am pn-Übergang ein aktiver Bereich, der sowohl Elektronen als auch Löcher enthält und Rekombination ermöglicht. Mit zunehmender Feldstärke nimmt somit der Absorptionskoeffizient α ab und wird schließlich sogar negativ. Der Halbleiter wirkt dann verstärkend mit dem optischen Gewinn:

$$g(h \cdot f) = -\alpha(h \cdot f). \quad \text{Gl. 3.21}$$

Dieser Gewinn als Funktion der Strahlungsenergie E und der Inversionsdichte n (**Abb. 3.8**) wird später für die systemtheoretische Betrachtung einer Halbleiterlaserdiode noch von entscheidender Bedeutung sein.

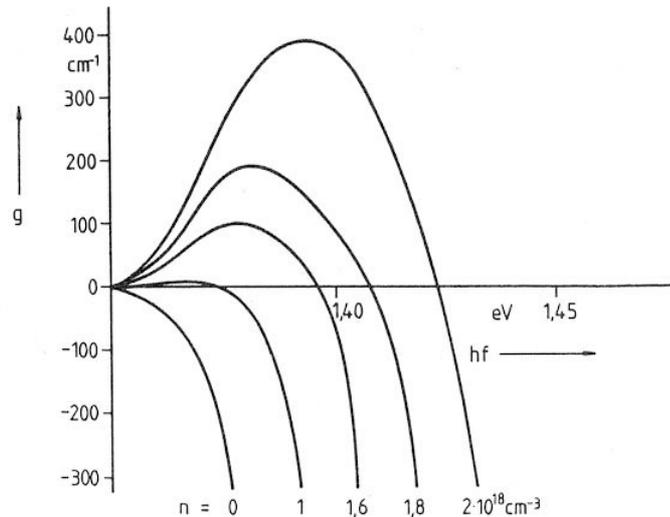


Abb. 3.8: Optischer Gewinn g eines GaAs-Halbleiterlasers als Funktion der Strahlungsenergie $E = h \cdot f$ und der Inversionsdichte n .⁶

3.4.3 VCSELs

Eine Variante der herkömmlichen Laserdioden stellen die Vertical Cavity Surface Emitting Lasers (VCSEL) dar. Der wesentliche Unterschied besteht in der Auskopplung des Laserlichts. Während die herkömmlichen Kantenemitter die optische Strahlung quer zum Stromfluss auskoppeln, erfolgt dieser Schritt bei VCSEL-Strukturen in der Stromrichtung. Zusätzlich ist die Länge der verstärkenden Zone sehr kurz und von mehreren Lagen $\frac{1}{4} \cdot \lambda$ -dicker dielektrischer Spiegel umgeben (**Abb. 3.9**). Die Geometrie des dielektrischen Spiegels selektiert eine bestimmte Lichtwellenlänge durch ihre Schichtdicken d_1 , d_2 und Brechungsindizes n_1 , n_2 :

$$\frac{1}{2} \lambda = n_1 \cdot d_1 + n_2 \cdot d_2. \quad \text{Gl. 3.22}$$

Eine derartige Struktur erlaubt eine konstruktive Interferenz der teilweise reflektierten Wellen und wirkt durch die periodische Variation des Brechungsindizes ähnlich einem Gitter. Ein Nachteil dieser Struktur findet sich in der sehr kurzen Verstärkerlänge.

⁶ W. Harth u. H. Grothe, „Sende- und Empfangsdioden für die Optische Nachrichtentechnik“, B.G. Teubner Stuttgart 1984, S. 65

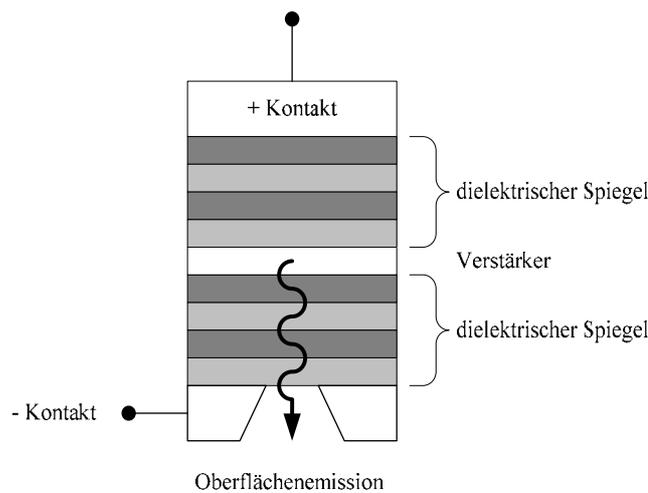


Abb. 3.9: Grundlegende Struktur eines Vertical Cavity Surface Emitting Lasers (VCSEL)

Da hiermit für ein Photon die Wahrscheinlichkeit sehr gering ist, ein weiteres mitzureißen, besitzen VCSEL relativ geringe Ausgangsleistungen. Ein entscheidender Vorteil findet sich aber im Design der Austrittsfläche. Entgegen herkömmlichen Kantenemittern besteht hier die Möglichkeit, radialsymmetrische Austrittsöffnungen vorzusehen. Dadurch besitzt der austretende Laserstrahl ebenso radialsymmetrischen Charakter, was eine exakte optische Abbildung wesentlich vereinfacht. Noch ein weiterer Punkt spricht für die Verwendung eines VCSEL. Aufgrund der sehr geringen Resonatorlänge kommt es nur zur Ausbildung eines longitudinalen Modes, was die Problematik der Modenstabilität zu entschärfen scheint. Jedoch ist hier anzumerken, dass für Systeme, die auf eine stabile Polarisierung angewiesen sind, solche Laser nicht anwendbar sind. Die Rotationssymmetrie der VCSEL bevorzugt keine feste Polarisationsrichtung, so dass diese Strukturen hier unter starken Schwankungen der Laserpolarisation leiden. Das Verhalten dieser Schwankungen lässt sich ähnlich der Modenstabilität von Kantenemittern (**Kap. 4.6.1**) statistisch über die Kopplung von Inversion, spontaner und oszillierender Emission beschreiben.⁷

⁷ M.P. van Exter, M.B. Willemsen, J.P. Woerdman; „Effect of mode-partition noise on intensity squeezing in a two-mode laser“, Journal of Optics B: Quantum and Semiclassical Optics Vol. 1 No. 6 (1999), S. 637-645

3.4.4 Superlumineszenzdioden

Die Verstärkung einer Lichtwelle durch stimulierte Emission ist kein Effekt, der sich nur in der Lasertechnik beobachten und nutzen lässt. Ersetzt man z.B. bei einer Laserdiodenstruktur die Grenzfläche auf einer der beiden Austrittsflächen durch einen absorbierenden Bereich W (**Abb. 3.10**), dann kann keine Oszillation anschwingen.

Die stimulierte Emission ist jedoch weiterhin möglich, jedoch nun nicht mehr nur für eine spezifische Wellenlänge. Mit zunehmender Breite des Absorptionsbereichs W entfernt man sich zunehmend von der eigentlichen Laserstruktur, so dass das Emissionsspektrum einen immer breiteren Bereich umfasst. Quantenmechanisch betrachtet sinkt zusätzlich jedoch auch die Wahrscheinlichkeit, ein Photon gleicher Energie, Ausbreitungsrichtung, Polarisation und Phase mitreißen zu können. Damit nähert sich das Verhalten einer Superlumineszenzdiode immer weiter der LED an und zeigt dann ebenso die typischen Sättigungserscheinungen (**Abb. 3.11**).

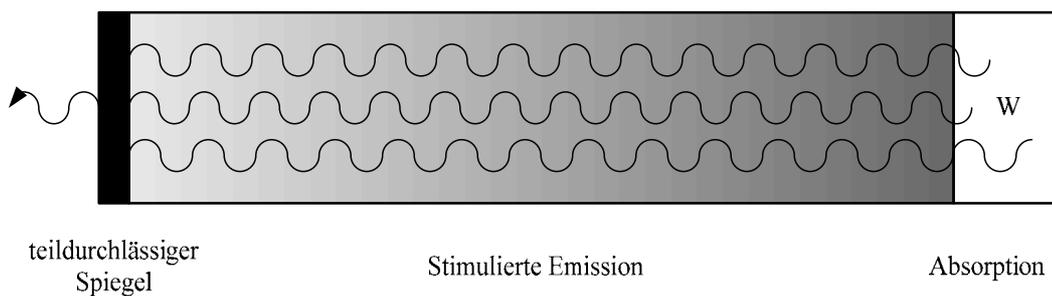


Abb. 3.10: Prinzipieller Aufbau einer Superlumineszenzdiode

3.5 Halbleiterempfangselemente

3.5.1 Halbleiterphotodioden

a) Arbeitsweise

Halbleiterphotodioden sind zumeist in einer PIN-Struktur aufgebaut. Hierbei wird auf einem hoch n-leitenden Substrat eine schwach n-leitende oder sogar nur eigenleitende (intrinsische) Schicht aufgebracht, in die eine flache p-leitende Zone eindiffundiert wird (**Abb. 3.12**). Der am meisten verbreitete Betriebsfall erfolgt in Sperrrichtung der

Diode, da die zumeist geringen Photoströme sich in Durchlassrichtung nur schwer vom Betriebsstrom unterscheiden lassen. In Sperrrichtung muss sich der Photostrom nur noch deutlich genug vom Dunkelstrom abheben.

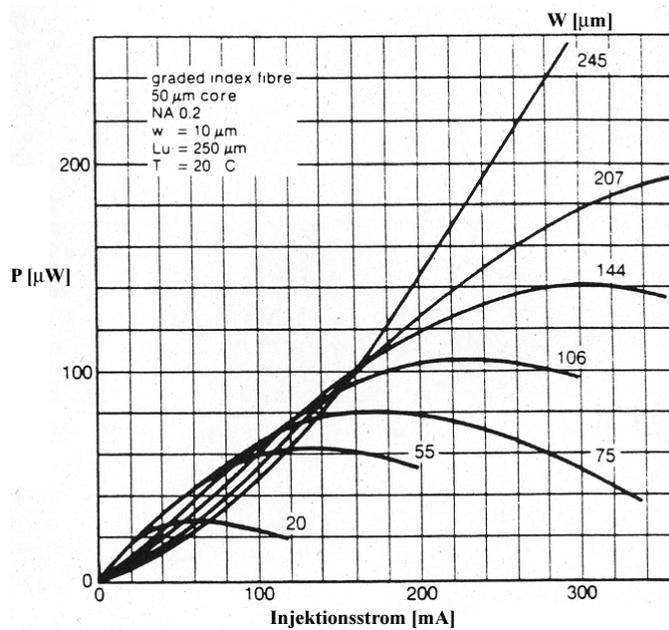


Abb. 3.11: Optische Leistung einer Superlumineszenzdiode in Abhängigkeit des Injektionsstroms I und der Absorptionsbreite W .⁸

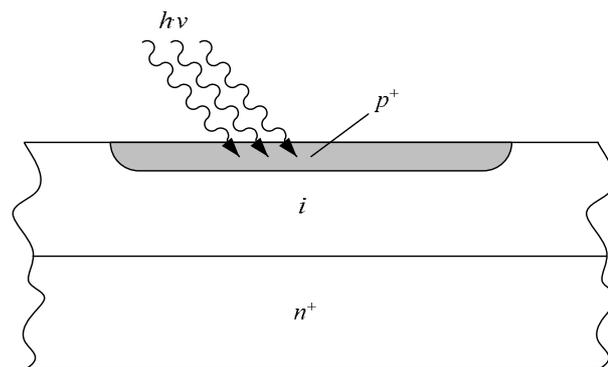


Abb. 3.12: Struktur einer planaren PIN-Photodiode

⁸ D. Rittich, H. Storm, T. Wiesmann, H.-G. Zielinski; „Physikalische Grundlagen der optischen Nachrichtentechnik“; ANT- Nachrichtentechnische Berichte Heft 3 1986

Eine wesentliche Begrenzung der Empfindlichkeit von Photodioden liegt in der Reflexion des einfallenden Lichtes. Mit einem material- und wellenlängenabhängigen Reflexionskoeffizienten r kann nur der nicht reflektierende Anteil zu einem Photostrom beitragen:

$$P = (1 - |r|^2) \cdot P_0. \quad \text{Gl. 3.23}$$

Mit Hilfe eines aufgebrachtten Antireflexbelags kann dieser Effekt nahezu beseitigt werden. Hierzu wird der optische Wellenwiderstand des Halbleiters an den freien Raum angepasst. Hierzu gilt für die Brechzahl des Belages

$$n_b = \sqrt{n_d} \quad \text{Gl. 3.24}$$

und seine Dicke

$$d_b = \frac{\lambda}{4 \cdot n_b}, \quad \text{Gl. 3.25}$$

wobei n_d die Brechzahl der Diodenoberfläche bezeichnet. Eine Ladungsträgergeneration kann grundsätzlich in allen drei Schichten der PIN-Struktur erfolgen. In der gezeigten planaren Struktur erfolgt der Lichteinfall durch die p-Zone hindurch. Schon hier kann eine Absorption bzw. Ladungsträgergeneration erfolgen, abhängig von der wellenlängenabhängigen Eindringtiefe der Photonen. In zunehmend tieferen Gebieten folgt die einfallende und durch die Reflexion r gedämpfte optische Leistung einem mit dem Dämpfungskoeffizienten α exponentiellen abklingenden Verlauf:

$$P = (1 - |r|^2) \cdot P_0 \cdot e^{-2\alpha x}. \quad \text{Gl. 3.26}$$

Dieser Ausdruck vereinfacht die Verhältnisse innerhalb einer Photodiode, da längst nicht alle generierten Ladungsträger zum Photostrom beitragen, denn es besteht auch die Möglichkeit der Rekombination. Aus systemtheoretischer Sicht ist eine einfache

messtechnisch zu gewinnende Größe von entscheidender Bedeutung, die der Ansprechempfindlichkeit R (engl. responsivity)

$$R = \frac{I_{ph}}{P_0} \quad \text{Gl. 3.27}$$

entspricht und all derartige Effekte mit einschließt. Sie steht mit der Quantenausbeute η in folgender Beziehung:

$$R = \frac{q \cdot \eta}{h \cdot \nu} \quad \text{Gl. 3.28}$$

Können alle Störeinflüsse vernachlässigt werden, ergibt sich die Quantenausbeute aus dem Absorptionskoeffizienten α :

$$\eta = (1 - |r|^2) \cdot (1 - e^{-2\alpha x}) \quad \text{Gl. 3.29}$$

Dieser ideale Wert kann zur Beurteilung einer vermessenen realen Struktur herangezogen werden.

b) Frequenzverhalten

Eine weitere aus systemtheoretischer Sicht entscheidende Eigenschaft ist der Frequenzgang bzw. Amplitudengang. Eine sinusförmig schwankende Strahlungsintensität P ändert den generierten Photostrom mit gleicher Modulationskreisfrequenz ω . Mit zunehmender Frequenz werden jedoch begrenzte Driftgeschwindigkeit der Ladungsträger die Stromschwankungen gegenüber denen des Lichtes verzögern. Werden eine homogene Elektronen-Loch-Paar-Erzeugung (schwache Absorption) und eine sich gleichende Sättigungsgeschwindigkeit v_S für

Elektronen und Löcher (bei ausreichend großer Feldstärke) angenommen, dann ergibt sich aus den Kontinuitätsgleichungen eine Grenzfrequenz f_g von:⁹

$$f_g = 0,59 \cdot \frac{v_s}{d}. \quad \text{Gl. 3.30}$$

Die Variable d bezeichnet hierbei die Dicke der intrinsischen Schicht der PIN-Struktur. Für den Fall der starken Absorption ergibt sich eine etwas reduzierte Grenzfrequenz:¹⁰

$$f_g = 0,44 \cdot \frac{v_s}{d}. \quad \text{Gl. 3.31}$$

Das Frequenzverhalten einer beschalteten PIN-Photodiode kann mit einem einfachen Ersatzschaltbild beschrieben werden. Hierbei liegt eine ideale Stromquelle I_{ph} parallel zur Kapazität der Verarmungszone C_d . Mit einem zu dieser Anordnung in Serie liegenden Widerstand R_S ist die Photodiode grundsätzlich charakterisiert, jedoch hat der Lastwiderstand R_L eines angeschlossenen Verstärkers ebenso entscheidenden Einfluss auf das Verhalten des Detektors. Die Zeit konstante τ errechnet sich somit zu:

$$\tau = C_d \cdot (R_S + R_L). \quad \text{Gl. 3.32}$$

Da der Serienwiderstand R_S bei hoher Dotierung vernachlässigbar sein kann, muss für ein schnelles Ansprechen der Photodiode die Kapazität C_d sehr klein sein. Das bedeutet, dass für eine bestimmte Dicke d der intrinsischen Schicht die Kapazität nur mit sehr kleinen Diodenflächen A gering gehalten werden kann:

$$C_d = \varepsilon \cdot \frac{A}{d}. \quad \text{Gl. 3.33}$$

⁹ W. Harth, H. Grothe; Send- und Empfangsdioden für die optische Nachrichtentechnik; B.G. Teubner, Stuttgart 1984, S. 164

¹⁰ W. W. Gaertner; Depletion-layer photoeffects in semiconductors; Phys. Rev. 116, 1959 S.84-87

Setzt man die Laufzeit $t = d/v_S$ zum Durchqueren der intrinsischen Schicht mit der Zeitkonstante τ des Ersatzschaltbildes gleich, dann ergibt sich eine optimale Dicke:

$$d_{opt} = \sqrt{\varepsilon \cdot A \cdot v_S \cdot R_L} . \quad \text{Gl. 3.34}$$

c) Rauschcharakteristik

Neben dem nahezu rein kapazitiven Frequenzverhalten einer PIN-Diode ist aus systemtheoretischer Sicht ihre Rauschcharakteristik von großer Bedeutung. In erster Linie unterliegt die Photogeneration durch die Quantisierung der eingestrahltten Energie der sogenannten Photonenstatistik bzw. dem Photonenrauschen. Da die Absorption von zwei Photonen gleichzeitig wie auch zueinander unendlich verzögert stattfinden kann, handelt es sich hierbei um weißes Rauschen. Da in realen Systemen jedoch nur eine endliche Zeit beobachtet werden kann, wird diese durch die vom System vorgegebene Messbandbreite B begrenzt. Der Beitrag des Photonen- bzw. Schrotrauschens lautet somit:

$$\overline{i_S^2} = 2 \cdot e \cdot I_{ph} \cdot B . \quad \text{Gl. 3.35}$$

Neben der Photogeneration erfolgt ebenso eine thermische Ladungsgeneration. Deren Rauschanteil ergibt sich über einen beliebigen Lastwiderstand R_L mit:

$$\overline{i_{th}^2} = \frac{4 \cdot k \cdot T \cdot B}{R_L} . \quad \text{Gl. 3.36}$$

Mittels des Effektivwertes i des Signalstromes I_{ph} kann das Signal-zu-Rauschverhältnis einer Photodiode bestimmt werden:

$$\frac{S}{N} = \frac{i^2}{\overline{i_S^2} + \overline{i_{th}^2}} . \quad \text{Gl. 3.37}$$

Für die minimal detektierbare Leistung ($S/N = 1$) ist aufgrund der thermischen Ladungsträgergeneration die Diodenfläche A ebenso von entscheidender Bedeutung, da neben einer aktiven Kühlung nur eine Verkleinerung der Fläche deren Rauschanteil minimiert.

3.5.2 Avalanche-Photodioden

a) Arbeitsweise

Avalanche-Photodioden bzw. Lawinenphotodioden ähneln prinzipiell der PIN-Photodiode. Ebenso in Sperrichtung betrieben werden jedoch so große Feldstärken erzeugt, dass innerhalb der Sperrschicht der Prozess der Lawinenverstärkung einsetzt. Hierbei wird ein photonisch erzeugter Ladungsträger derart beschleunigt, dass seine Energie ausreicht, bei entsprechender Wechselwirkung (Stoßionisation) einen weiteren Ladungsträger mitzureißen. Je nach Größe der Sperrschicht und anliegender Feldstärke kann sich dieser Prozess mehrfach wiederholen, so dass diese Photodioden eine enorme innere Verstärkung M besitzen:

$$M = \frac{I}{I_{ph}} = \frac{1}{1 - \left(\frac{U}{U_B}\right)^{\frac{\kappa+1}{2}}}. \quad \text{Gl. 3.38}$$

Die Parameter U und U_B bezeichnen die anliegende Sperr- und die materialabhängige Durchbruchspannung. Die Potenznäherung der Ionisierungskoeffizienten $\alpha_{n,p}$ liefert den Exponenten κ , welcher experimentell ermittelt werden kann und für Silizium typischerweise 6 ist. Bei starken Photoströmen wird die Sperrspannung der Avalanche-Diode vermindert, was sich durch einen Spannungsabfall an einem Serienwiderstand R_S darstellen lässt:

$$M = \frac{I}{I_{ph}} = \frac{1}{1 - \left(\frac{U - I \cdot R_S}{U_B} \right)^{\frac{\kappa+1}{2}}}. \quad \text{Gl. 3.39}$$

Große Bedeutung für den Einsatz von Avalanche-Dioden hat ihre Temperaturabhängigkeit, da sich die Durchbruchspannung U_B mit steigender Temperatur erhöht. Für eine konstante Verstärkung ist somit die anliegende Sperrspannung nachzuregeln.

b) Frequenzverhalten

Die zuvor beschriebene Auswirkung des Photostromes auf die interne Verstärkung M resultiert nicht nur in einem nichtlinearen Verhalten, sondern beeinflusst auch deren Zeitverhalten. Die Lawinenansprechzeit τ_L für geringe Photoströme hängt nur von der sogenannten mittleren Stoßzeit τ_S ab, welche die Laufzeit von Ladungsträgern zwischen den Stoßprozessen repräsentiert:

$$\tau_L = (M - 1) \cdot \frac{\tau_S}{2}. \quad \text{Gl. 3.40}$$

Mit zunehmender Photointensität kann die Beeinflussung der Sperrspannung nicht mehr vernachlässigt werden, was zu einer Lawinenansprechzeit von

$$\tau_L = \tau_S \cdot \sqrt{\frac{U_B}{2 \cdot (\kappa + 1) \cdot R_S \cdot I_{ph}}}. \quad \text{Gl. 3.41}$$

führt. Auch dem Abklingen eines kurzen optischen Impulses wird nur allmählich gefolgt. Dadurch kann eine sinusförmige Intensitätsmodulation nur mit der vollen Gleichstromverstärkung M detektiert werden, wenn ihre Modulationsfrequenz genügend gering ist. Ähnlich einer PIN-Photodiode kann eine Avalanche-Diode als Tiefpass erster Ordnung betrachtet werden. Für geringe Verstärkungsfaktoren ($M \sim 1$) ergibt sich die

Grenzfrequenz, ähnlich zur PIN-Photodiode, allein durch die Laufzeit der Ladungsträger durch die Driftzone d :

$$f_g = 0,44 \cdot \frac{v_S}{d}. \quad \text{Gl. 3.42}$$

Bei großen Verstärkungen ($M \gg 1$) und relativ kurzer Driftzone überwiegt jedoch die Dynamik des Lawinenprozesses. Dafür ergibt sich die Grenzfrequenz aus der Lawinengleichung durch die Lawinenzone d_a :

$$f_g = \frac{v_S}{\pi \cdot M \cdot d_a}. \quad \text{Gl. 3.43}$$

c) Rauschcharakteristik

Die Beschreibung der Rauschprozesse innerhalb einer Avalanche-Diode kann von den PIN-Photodioden grundsätzlich übernommen werden, jedoch unterliegen diese wie der Signalstrom ebenso der internen Verstärkung M . Das Schrotrauschen des Photostromes lautet somit:

$$\overline{i_s^2} = 2 \cdot e \cdot I_{ph} \cdot \overline{M^2} \cdot B. \quad \text{Gl. 3.44}$$

Das thermische Rauschen einer Avalanche-Diode ist natürlich auch dieser internen Verstärkung M unterworfen:

$$\overline{i_{th}^2} = \frac{4 \cdot k \cdot T \cdot \overline{M^2} \cdot B}{R_L}. \quad \text{Gl. 3.45}$$

Wichtig ist bei dieser Betrachtung, dass die Verstärkung mit der sogenannten Rauschzahl F zu gewichten ist. Das mittlere Schwankungsquadrat des internen Gewinns lautet: ¹¹

$$\overline{M^2} = F(M) \cdot M^2. \quad \text{Gl. 3.46}$$

Diese beschreibt den Anstieg des Schrotrauschens durch den Zufallscharakter der Lawinenvervielfachung. Diese ist somit keine feste Größe, sondern schwankt regellos. Das Signal-zu-Rausch-Verhältnis S/N lautet für einen Modulationsgrad des optischen Signals von 100%: ¹²

$$\frac{S}{N} = \frac{I_{ph}^2}{4 \cdot e \cdot B \cdot \left(I_{ph} \cdot F(M_0) + \frac{2 \cdot k \cdot T}{M_0^2 \cdot e \cdot R_L} \right)}. \quad \text{Gl. 3.47}$$

Damit wird mit zunehmender Verstärkung der Beitrag des thermischen Rauschens gegenüber dem Schrotrauschen des empfangenen optischen Signals reduziert, da der Lastwiderstand R_L um den Faktor M_0^2 vergrößert ist.

3.5.3 Photogate-PMD-Strukturen als quasioptische Mischer

Die Technologie der Photonic Mixer Devices (PMD) beruht auf einem einfachen aber sehr wirkungsvollen Prinzip. In ihnen ist die Möglichkeit der Mischung des detektierten optischen Signals s_{opt} mit einem zusätzlichen elektrischen Signal s_{el} derart implementiert, dass der Prozess der Signalmischung direkt während der Ladungsträgergeneration erfolgt. Neben diesem als inhärente oder quasioptische Mischung bezeichneten Prozess sind PMD-Strukturen zusätzlich zweikanalig realisiert, um so die Vorteile eines Gegentaktbetriebs mit differentiellem Ausgang auszunutzen.

¹¹ R. C. Mc Intyre; Multiplication noise in uniform avalanche diodes, IEEE Transaction Electron. Dev. ED-13, 1966, S. 164-168

¹² J. Müller; Photodiodes for optical communication, Advances in Electronics and Electron Physics No. 55, 1981, S. 183-307

Da das Hauptaugenmerk dieser Arbeit auf der optischen Signalverarbeitung liegt, soll im Weiteren auf zwei der wichtigsten PMD-Strukturen etwas genauer eingegangen werden, auf das Photogate- und das MSM-PMD. An ihrem Beispiel werden die Vorteile dieser Technologie detaillierter herausgearbeitet.

a) Arbeitsweise des Photogate-PMD

Das Photogate-PMD basiert auf einer der am weitesten verbreiteten und preiswertesten Halbleitertechnologien, dem Complementary-Metal-Oxid-Silicon (CMOS) – Prozess. Die Grundstruktur eines solchen PMD ähnelt stark einem sogenannten n-Kanal-Anreicherungs-MOS-Transistor (**Abb. 3.13**). Der Unterschied beider Strukturen findet sich in folgenden zwei Merkmalen:

- Der abgebildete Wechsel des Gate-Materials von Metall zu Polysilizium schafft die zur Lichtdetektion notwendige Transparenz.
- Die Teilung des Gates in die sogenannten Modulationsgates A und B ermöglicht den zweikanaligen Gegentaktbetrieb.

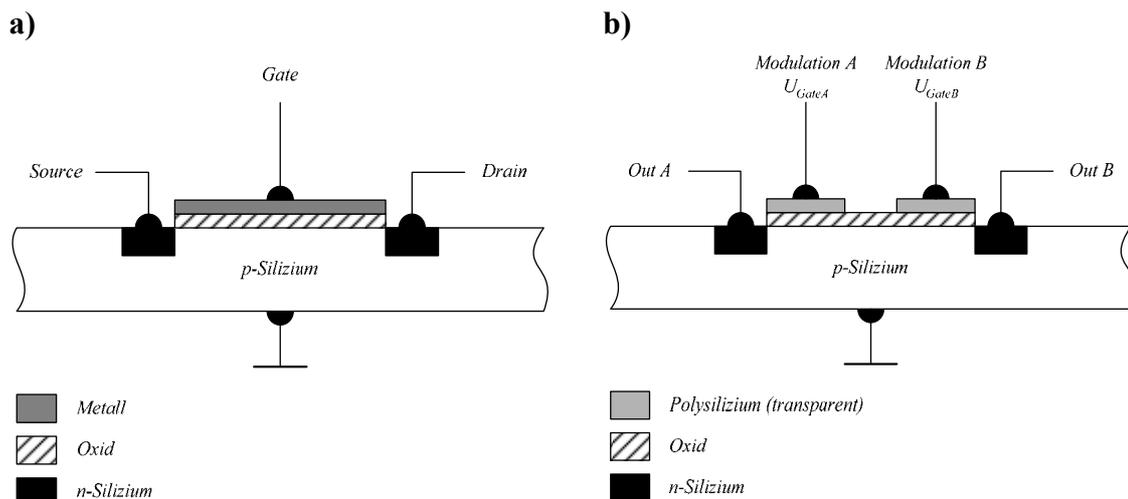


Abb. 3.13: Vereinfachte Darstellung einer a) n-Kanal-MOS-Transistor-Struktur und eines b) Photogate-PMD

Trifft nun ein optisches Signal die Oberfläche einer solchen Photogate-Struktur, durchdringt es aufgrund der geschaffenen Transparenz das Gate und das als Isolator fungierende Oxid. Im Bereich des p-Siliziums, das mitunter auch als Substrat bezeichnet wird, erfolgt nun eine Absorption der Photonen. Damit verbunden ist die Generation freier Ladungsträger, was unter Einwirkung eines elektrischen Feldes einen Photostrom zur Folge hat. Die räumliche Ausprägung eines solchen elektrischen Feldes kann über die beiden Modulationsgates beeinflusst werden (**Abb. 3.14**).

Da freie Ladungsträger in Richtung des Gradienten des elektrischen Feldes beschleunigt werden, kann der entstehende Stromfluss in seiner Flussrichtung derart beeinflusst werden, dass er im Idealfall nur zu einer der beiden Auslesedioden fließt. Vertauscht man die Werte der beiden anliegenden Gate-Spannungen $U_{Gate A}$ und $U_{Gate B}$, dann erfolgt der Stromfluss zu der entgegen gesetzten Auslesediode (**Abb. 3.15**).

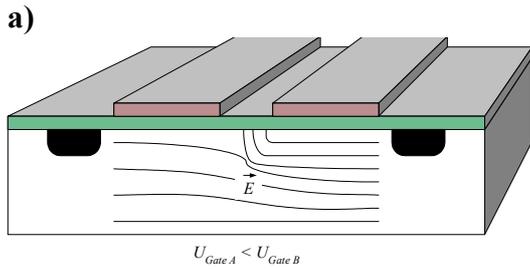
Prinzipiell stellt das PMD einen elektrischen Mischer dar, der ein empfangenes optisches Signal mit einem weiteren elektrischen multiplikativ verknüpft. Hierbei kann der Begriff *quasioptisch* benutzt werden, da der Mischprozess direkt während der Detektion des optischen Signals erfolgt.

b) Frequenzverhalten des Photogate-PMD

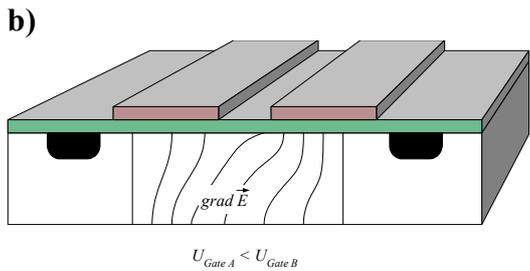
Für den Frequenzgang des Mischverhaltens ist die Kapazität der Polysilizium-Oxid-Substrat-Struktur maßgeblich. Orientiert man sich an der aus der CMOS-Welt bekannt MIS-Struktur (Metall-Insulator-Semiconductor), dann ergibt sich die Kapazität eines solchen Photogates aus einer Reihenschaltung der Oxidkapazität C_{oxid} mit der Kapazität C_{dep} der Verarmungszone (engl. depletion layer) im Halbleiter:

$$\frac{1}{C} = \frac{1}{C_{oxid}} + \frac{1}{C_{dep}}. \quad \text{Gl. 3.48}$$

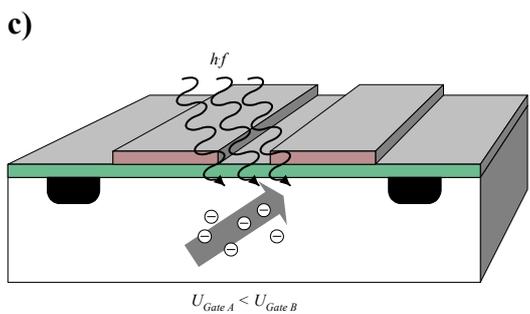
Die Kapazität C_{oxid} der Oxidschicht bestimmt sich aus der Oxiddicke d_{oxid} , der Fläche des Gate-Kontakts A_{gate} und der spezifischen Dielektrizitätskonstante ϵ_{oxid} :



Potentialverlauf des elektrischen Feldes \vec{E} unterhalb der Photogates bei unterschiedlichen Gate-Spannungen



Resultierende Krafrichtung $grad\vec{E}$



Elektrooptische Generation freier Ladungsträger mit der sich einstellenden Bewegungstendenz

Abb. 3.14: Beeinflussung der Driftbewegung elektrooptisch generierter Ladungsträger durch die gezielte Steuerung des Potentialverlaufs durch die Gate-Spannungen $U_{Gate A}$ und $U_{Gate B}$.

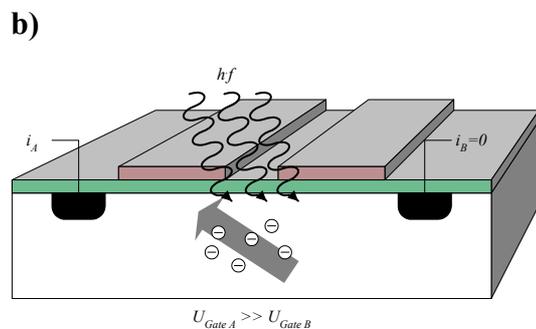
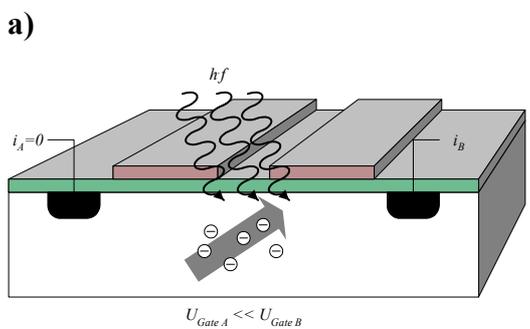


Abb. 3.15: Quasioptische Gegentaktmischung eines optischen Signals in der Photogate-Struktur: a) $U_{Gate A} \ll U_{Gate B}$; b) $U_{Gate A} \gg U_{Gate B}$

$$C_{\text{oxid}} = \epsilon_{\text{oxid}} \cdot \epsilon_0 \cdot \frac{A_{\text{gate}}}{d_{\text{oxid}}}. \quad \text{Gl. 3.49}$$

Diese Kapazität kann als konstant angesehen werden, da ihre relevanten Parameter ϵ_{oxid} , d_{oxid} und A_{gate} von Arbeitspunkt und Frequenzgang unabhängige Größen darstellen. Die Kapazität der Sperrschicht bzw. Verarmungszone ergibt sich aus ihrer Tiefe d_{dep} , der spezifischen Dielektrizitätskonstante des Siliziumsubstrats $\epsilon_{\text{silicon}}$ und der Gate-Fläche A_{gate} :

$$C_{\text{dep}} = \epsilon_{\text{silicon}} \cdot \epsilon_0 \cdot \frac{A_{\text{gate}}}{d_{\text{dep}}}. \quad \text{Gl. 3.50}$$

Die Tiefe der Verarmungszone ist von der Stärke des einwirkenden elektrischen Feldes abhängig. Während in einem herkömmlichen Plattenkondensator die Feldstärke örtlich einen konstanten Verlauf besitzt, stellen sich in einem Halbleiter völlig andere Verhältnisse ein. Hier werden die vorhandenen freien Ladungsträger (Majoritätsladungsträger) durch das Feld beschleunigt, während die dadurch ionisierten Akzeptoren weiterhin ortsfest sind.

In der MIS-Struktur eines p-Halbleiters werden die Majoritäten (Löcher) an der Oxidgrenzfläche durch die Feldeinwirkung verdrängt und es bleiben die ortsfesten negativen Akzeptorladungen übrig (**Abb. 3.16 b**). Entsprechend dem Superpositionsprinzip kann sich aus der Überlagerung des äußeren Feldes, der einzelnen Felder der Akzeptoren und der Felder der verdrängten Löcher keine konstante Feldstärke mehr einstellen. Durch eine Erhöhung der Gate-Spannung findet die Verarmung in immer tieferen Regionen statt. Die Ausdehnung der Raumladungszone nimmt weiter zu und in der Folge sinkt ihre Kapazität. Ein weiteres Anwachsen der Gate-Spannung führt schließlich zum Effekt der Inversion. Hier wird durch das Einwirken des äußeren Feldes ein Überschuss der ursprünglichen Minoritäten (Elektronen) erreicht, die aufgrund des jetzt bestehenden Akzeptormangels nicht mehr ortsfest gebunden sein können. Sie werden stattdessen zur Oxidgrenzfläche hin beschleunigt werden (**Abb. 3.16 c**). Dadurch verringert sich wieder zunehmend die Ausdehnung der Raumladungszone und die Kapazität steigt.

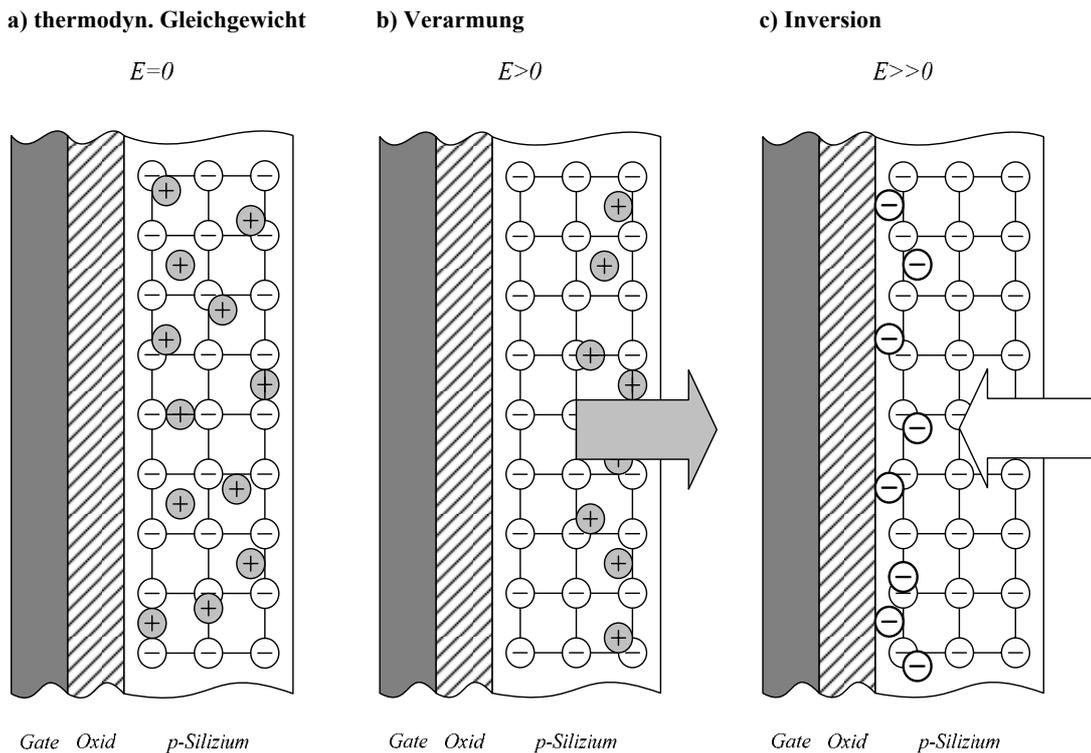


Abb. 3.16: Einfluß eines elektrischen Feldes auf die freien Ladungsträger innerhalb eines p-Halbleiters

Im bisher betrachteten statischen Fall wird sich immer ein Ladungsgleichgewicht einstellen, welches für die Verarmung bestimmt wird durch:

$$Q_{Gate} = Q_A - Q_p \quad \text{Gl. 3.51}$$

Hier bezeichnet Q_{Gate} die insgesamt auf das Gate aufgebraute Ladung. Demgegenüber stehen die Gesamtladung Q_A aller Akzeptoren und die Ladung Q_p aller Majoritätsträger, die aufgrund ihrer Polarität mit entgegengesetztem Vorzeichen eingeht. Betrachtet man die Ladungen bezogen auf ein Volumenelement dV , dann lässt sich die Gesamtladung der Akzeptoren aus der dotierungsabhängigen Akzeptorkonzentration N_A bestimmen:

$$\begin{aligned} Q_A &= \int q \cdot N_A \cdot dV \\ &= \int q \cdot N_A \cdot dx \cdot dy \cdot dz \end{aligned} \quad \text{Gl. 3.52}$$

Bei einer ausreichend breiten Gate-Struktur sind die Änderungen in ihrem Zentrum entlang einem parallelen Flächenelement $dA=dx \cdot dy$ vernachlässigbar, wodurch das Ladungsträgergleichgewicht als eindimensionaler Fall betrachtet werden kann:

$$\begin{aligned} \frac{Q_A}{dA} &= \frac{Q_A}{dx \cdot dy} \\ &= \int q \cdot N_A \cdot dz. \end{aligned} \quad \text{Gl. 3.53}$$

Über die Substratdicke d_S ergibt sich für eine homogene Dotierung:

$$\frac{Q_A}{dA} = q \cdot N_A \cdot d_S. \quad \text{Gl. 3.54}$$

Im thermodynamischen Gleichgewicht (**Abb. 3.17 a**) entspricht die Majoritätsträgerkonzentration p der Akzeptorkonzentration N_A , so dass gilt:

$$\frac{Q_p}{dA} = -q \cdot N_A \cdot d_S. \quad \text{Gl. 3.55}$$

Bildet sich jedoch durch ein äußeres Feld eine Verarmungszone der Tiefe d_{dep} aus, dann gilt für die Majoritätsträgerkonzentration:

$$\frac{Q_p}{dA} = -q \cdot N_A \cdot (d_S - d_{dep}). \quad \text{Gl. 3.56}$$

Für das Ladungsträgergleichgewicht der Verarmung (**Abb. 3.17 b**) gilt nun:

$$\begin{aligned} \frac{Q_{Gate}}{dA} &= q \cdot N_A \cdot d_S - q \cdot N_A \cdot (d_S - d_{dep}) \\ &= q \cdot N_A \cdot d_{dep}. \end{aligned} \quad \text{Gl. 3.57}$$

Wird die Feldstärke so weit erhöht, dann befinden sich keine Majoritätsladungen mehr innerhalb des Substrats. Die Raumladungszone breitet sich nun über die ganze Substratdicke aus. Erfolgt nun eine weitere Erhöhung der Feldstärke, dann kommt es zum Effekt der Inversion. Der damit verbundene Überschuss an Minoritäten ist nicht ortsfest gebunden, da alle vorhandenen Akzeptoren bereits besetzt sind. Sie orientieren sich zur Oxidschicht hin und bilden dort ein Inversionsgebiet der Tiefe d_i aus. Das Ladungsträgergleichgewicht (Abb. 3.17 c) lautet nun mit der entsprechenden Minoritätenkonzentration n :

$$\frac{Q_{Gate}}{dA} = q \cdot N_A \cdot d_S + q \cdot n \cdot d_i. \quad \text{Gl. 3.58}$$

Da die Inversionsladungen relativ langsamen Generations- und Rekombinationsraten unterliegen, können sie schnellen Feldänderungen nicht folgen. Damit wird die Ausbreitung der Inversionszone eine stark frequenzabhängige Größe, was sich im Kapazitäts-Spannungsverlauf einer MIS-Struktur entscheidend auswirkt (Abb. 3.18). Mit sehr hohen Frequenzen wird durch die Unterdrückung der Inversion ein Ausbreiten der Raumladungszone auf die gesamte Substratdicke d_S erreicht. Dieser Zustand wird als tiefe Verarmung bezeichnet. Wendet man die Betrachtungen der MIS-Kapazität auf die Modulationsgates einer Photogate-Struktur an, so ist aus systemtheoretischer Sicht mit einem frequenzabhängigen Verhalten des Mischprozesses zu rechnen.

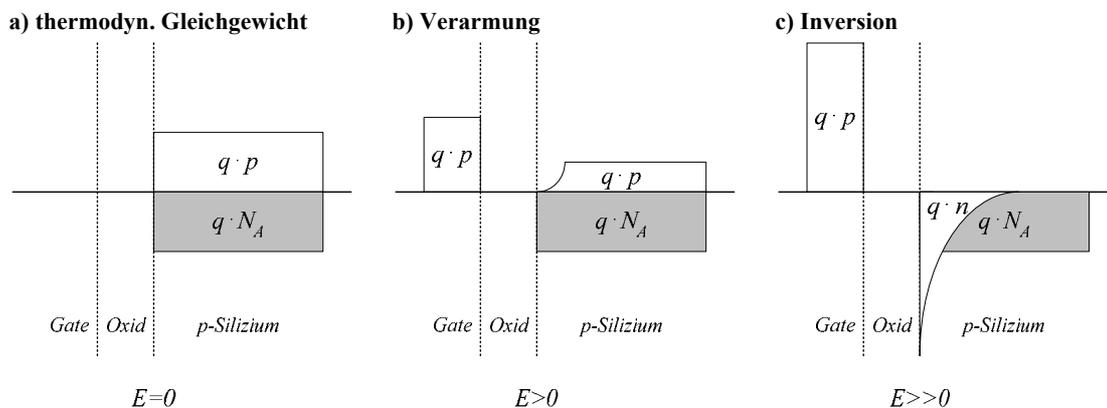


Abb. 3.17: Einfluss eines elektrischen Feldes E auf die Ladungsverteilung im statischen Gleichgewicht

Die Messung der Kapazität eines einzelnen Photogates einer typischen PMD-Struktur zeigt das erwartete Verhalten (**Abb. 3.19**). Deutlich ist die Verringerung der Inversion mit steigender Frequenz zu erkennen. Die Kapazitäten eines typischen PMD-Photogates kann in Anlehnung an die MIS-Struktur berechnet werden. Die Kapazität C_i der Oxidschicht ergibt sich hierbei aus:

$$C_i = \frac{\epsilon_i \cdot A}{d} = \frac{3,9 \cdot 8,85E-12F/m \cdot 2,970E-9m^2}{14,5E-9m} \approx 7pF . \tag{Gl. 3.59}$$

Die Kapazität der Raumladungszone steht im direkten Zusammenhang mit ihrer Tiefe W_m . Die resultierende Gesamtkapazität C'_{min} lautet:

$$C_{min} \cong \frac{\epsilon_i \cdot A}{d + \frac{\epsilon_i}{\epsilon_s} \cdot W_m} . \tag{Gl. 3.60}$$

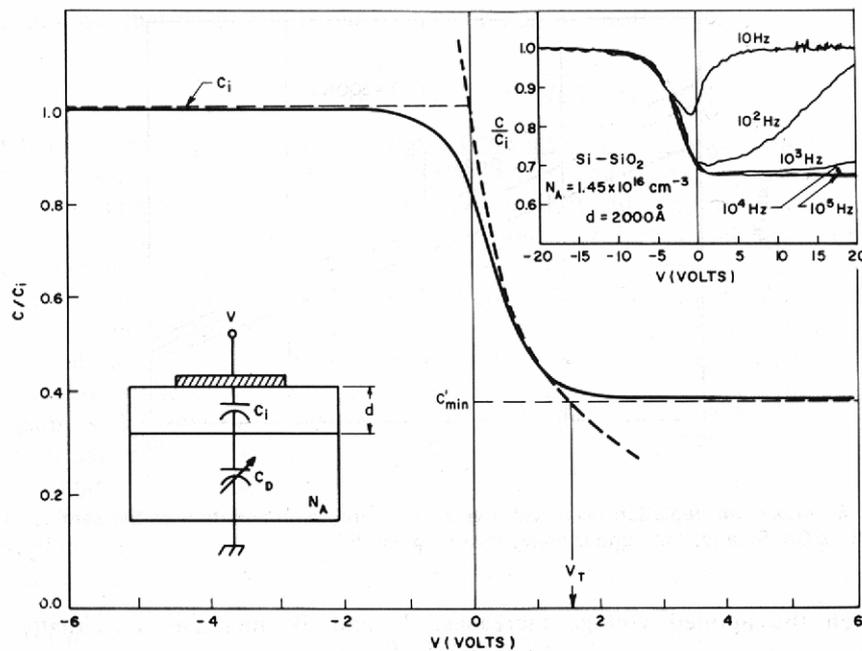


Abb. 3.18: Spannungsverlauf einer MIS-Struktur in Abhängigkeit der Frequenz ¹³

¹³ aus S.M.Sze; Physics of Semiconductor Devices, John Wiley & Sons, New York 1981, S. 374

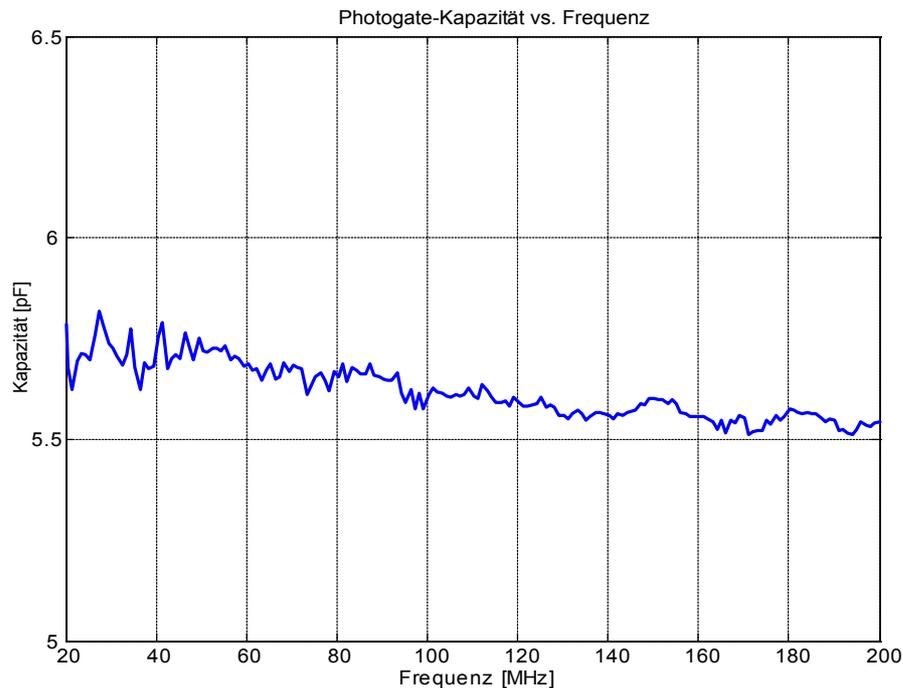


Abb. 3.19: Kapazitiver Anteil der Impedanzmessung einer Photogate-Struktur

Für die gegebene Gate-Struktur und eine Raumladungszone von $0,1\mu\text{m}$ ergibt sich eine resultierende Gesamtkapazität von ca. 2pF . Diese theoretischen Werte stimmen sehr gut mit der Messung überein, insbesondere dann, wenn man eine Kapazität für Bondpad und Zuleitung von weiteren 3pF berücksichtigt. Für den theoretischen Wert von 10pF über das Oxid und einer gemessenen Gesamtkapazität von $5,5\text{pF}$ ergibt sich eine Tiefe der Raumladungszone von etwa $0,02\mu\text{m}$. Für das Übertragungs- bzw. Mischverhalten einer Photogate-Struktur ist neben dem kapazitiven ebenso der Ohmsche Anteil maßgeblich. Dieser resultiert direkt aus dem Ohmschen Verhalten des verwendeten Gate-Materials. In einer Photogate-Struktur ergibt sich für das verwendete Polysilizium bei konstanter Schichtdicke ein Flächenwiderstand $33\text{ Ohm}/\mu\text{m}^2$. Für die vermessene Struktur berechnet sich damit ein Widerstand von 65 Ohm , was sich durch die Messung von etwa 100 Ohm bestätigte (**Abb. 3.20**). Bei der Messung ist zu berücksichtigen, dass hier die gesamte Verbindungstechnik wie Bond-Pad, -Draht usw. mit eingeschlossen wurde und somit insgesamt über dem rechnerischen Wert liegen muss.

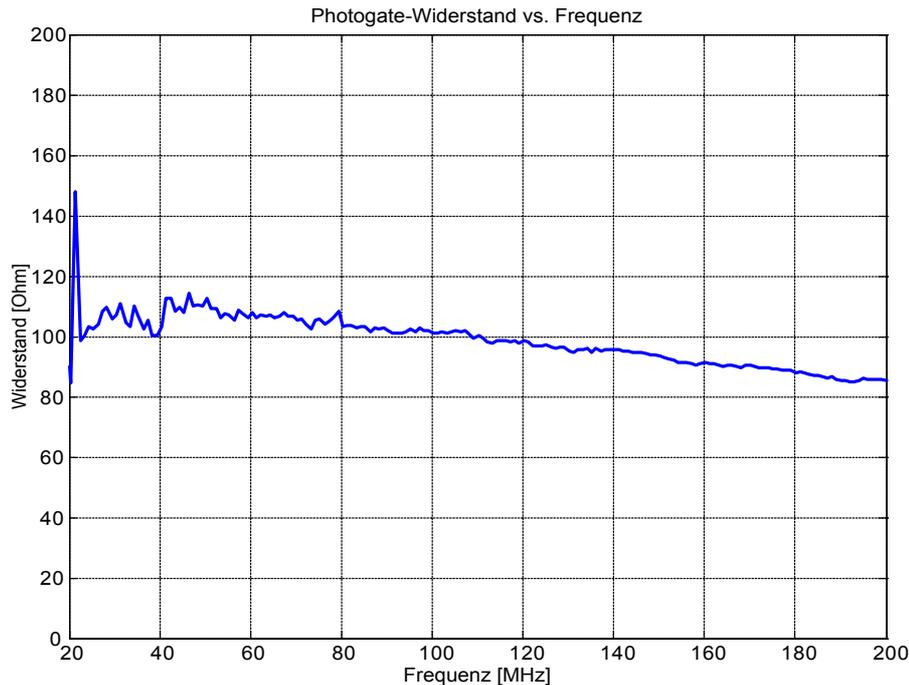


Abb. 3.20: Ohmscher Anteil der Impedanzmessung einer Photogate-Struktur

c) Rauschcharakteristik des Photogate-PMD

Für eine systemtheoretische Betrachtung eines Photogate-PMD kommt neben dem Impedanzgang dem Phänomen des Rauschens eine große Bedeutung zu, da dadurch die Qualität eines Sensorsystems hinsichtlich seiner Messgenauigkeit begrenzt wird. Das bedeutet, dass ein Signal nur dann detektierbar ist, wenn es sich deutlich genug von der Summe aller Rauschvorgänge abhebt. Im optischen Bereich tritt eine wesentliche Rauschquelle als physikalische Grenze in den Vordergrund, welche sich aus der Quantisierung der elektromagnetischen Strahlung ergibt. Für die Optimierung eines optischen Sensorsystems heißt es, sich dieser Grenze des idealen optischen Empfangs so weit wie möglich zu nähern. Die wichtigsten Rauschquellen eines realen optischen Empfängers sind das Schrotrauschen bzw. Shot Noise

$$\overline{i_s^2} = 2 \cdot q \cdot (I_{ph} + I_D) \cdot |\Gamma(\omega)|^2 \cdot B \quad \text{Gl. 3.61}$$

und das interne thermische Rauschen bzw. Johnson Noise

$$\overline{i_i^2} = 4 \cdot k \cdot T \cdot \frac{1}{R_i} \cdot B. \quad \text{Gl. 3.62}$$

Dieser Ausdruck gilt jedoch nur für die entkoppelte Betrachtung eines Empfängers, so dass für ein vollständiges System die nachfolgende Signalkette durch einen zum internen Widerstand R_i parallel liegenden Lastwiderstand R_L mit einzubeziehen ist:¹⁴

$$\overline{i_{th}^2} = 4 \cdot k \cdot T \cdot \left(\frac{1}{R_i} + \frac{1}{R_L} \right) \cdot B. \quad \text{Gl. 3.63}$$

Diese zwei Rauschquellen werden zur Beschreibung einer Signaldetektion mit einer idealen Photodiode mit anschließendem Verstärker angewendet. Jedoch für ein reales Photogate-PMD ergeben sich noch weitere Rauschquellen und systematische Unterschiede, die im Folgenden beschrieben werden sollen.

Am Anfang der Signalkette eines jeden optischen Empfängers steht die Photodetektion bzw. die photonische Ladungsträgergeneration. Aus der Quantisierung der elektromagnetischen Strahlung des Lichtes und deren Signalumsetzung in einen Signalstrom, der ebenfalls als Energiequanten zu betrachten ist, folgt ein statistischer Prozess, der als Schrotrauschen bzw. Shot Noise bezeichnet wird. Das Schrotrauschen des Photostroms eines Photogate-PMD lässt sich als Rauschspannung über die Integrationskapazität C und die zugehörige Integrationszeit t_i berechnen:

$$v_{ph} = \frac{\sqrt{i_{ph} \cdot t_i \cdot q}}{C}. \quad \text{Gl. 3.64}$$

Es lässt sich aber ebenso durch eine Anzahl an verrauschten Elektronen ausdrücken:

$$n_{ph} = \sqrt{\frac{i_{ph} \cdot t_i}{q}}. \quad \text{Gl. 3.65}$$

¹⁴ H. Melchior, Demodulation and Photodetection techniques, in laser handbook, vol. 1, North-Holland Publishing Company, Amsterdam 1972, S. 740

Dem Rauschanteil des Photostroms ist der des Dunkelstroms zusätzlich überlagert. Dessen Schrotrauschen ergibt eine Rauschspannung von:

$$v_D = \frac{\sqrt{i_D \cdot t_i \cdot q}}{C} \quad \text{Gl. 3.66}$$

bzw. als Elektronenanzahl:

$$n_D = \sqrt{\frac{i_D \cdot t_i}{q}} \quad \text{Gl. 3.67}$$

Mit einem gemessenen Dunkelstrom $i_D = 2,8 \text{ pA}$ ergibt sich bei einer Integrationszeit $t_i = 10 \text{ ms}$ eine Rauschspannung von:

$$v_D \approx 54 \mu\text{V},$$

bzw. als Elektronenanzahl:

$$n_D \approx 420 \cdot e^-.$$

Auf dem Weg zur Auslesediode unterliegt jedes Elektron, egal ob zum Dunkel- oder Photostrom gehörig, Wechselwirkungen mit dem umgebenden Medium. Hierbei wird der kontinuierlichen Bewegung eine zufällige überlagert. Das Durchqueren eines solchen Gebietes kann somit nicht zeitlich genau vorhergesagt werden. Dieser statistische Prozess erzeugt einen Rauschanteil, der im Wesentlichen vom Leitwert des Mediums abhängt. Da ebenso eine Temperaturabhängigkeit besteht, wird dieser Prozess oft als Thermisches Rauschen bezeichnet. Hier soll aber der Begriff Johnson-Rauschen benutzt werden, um sich vom Rauschen des ebenfalls thermisch abhängigen Dunkelstroms abzugrenzen.

Die Komponente des Johnson-Rauschens ist für eine PMD-Struktur am schwierigsten abzuschätzen. Zwar kennt man die Geometrie der PMD-Struktur, jedoch ist die Wegstrecke und damit die Anzahl der Wechselwirkungen für jede photogenerierte Ladung unterschiedlich (**Abb. 3.21**). Selbst für eine bekannte Wegstrecke ist dieser

Rauscheffekt so lange nicht zu fassen, wie der Leitwert und damit der zu diesen Wechselwirkungen äquivalente Widerstand R_i (entsprechend Gl. 3.62) unbekannt ist. Nimmt man formal einen Widerstand von 1Ohm an, dann ergibt sich eine Rauschspannung von:

$$v_{th} = \frac{4 \cdot k \cdot T \cdot q}{R_i \cdot t_i \cdot C} \approx 1,3\mu V. \quad \text{Gl. 3.68}$$

Eine Erhöhung dieses Widerstandes wird diesen Wert weiter verringern und sich gegenüber dem Schrotrauschen des Fotostroms verschwindend auswirken. Aus diesem Grund wird dieser Rauschanteil nicht weiter berücksichtigt werden.

Bei der Integration des Photo- und Dunkelstroms ergibt sich die resultierende Rauschspannung aus der Summe ihrer Rauschanteile. Ein weiterer Effekt darf jedoch nicht unberücksichtigt bleiben, da er einen nicht unerheblichen Einfluss auf das Integrationsergebnis besitzt. Dieses hängt natürlich direkt vom Startwert der Integration ab, was im technischen Sinn der Anzahl schon vorhandener Ladungen auf der Kapazität entspricht. Das Rauschen dieses Startwertes wird als kTC -Rauschen bezeichnet, da es in einer Abhängigkeit zur Temperatur T und zur Integrationskapazität C steht:

$$v_{kTC} = \sqrt{\frac{k \cdot T}{C}} \quad \text{Gl. 3.69}$$

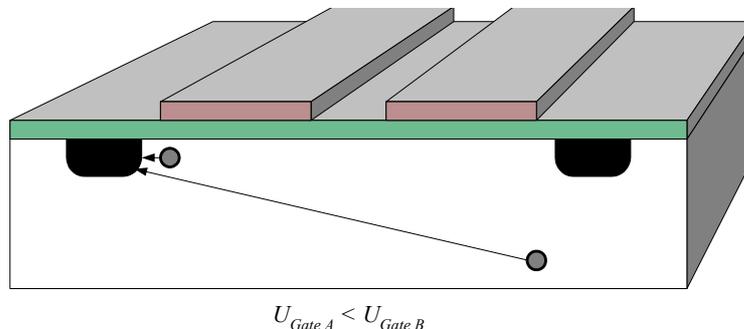


Abb. 3.21: Mögliche Diffusionswege einer photogenerierten Ladung innerhalb einer Photogate-PMD-Struktur.

bzw.

$$n_{kTC} = \frac{\sqrt{k \cdot T \cdot C}}{q} . \quad \text{Gl. 3.70}$$

Da der Startwert der Integration technisch durch einen Reset-Transistor realisiert wird, existiert für diesen Prozess der Begriff des Reset-Rauschens. Für eine typische Photogate-Struktur mit einer Integrationskapazität $C=1,25pF$ berechnet sich das Reset-Rauschen bei einer Temperatur $T = 300K$ zu:

$$v_{kTC} \approx 58\mu V$$

bzw.

$$n_{kTC} \approx 450 \cdot e^- .$$

Diese Rauschwerte eines PMD-Kanals sind zueinander unkorreliert und die resultierende Rauschspannung berechnet sich zu:

$$v_{PMD} = \sqrt{v_{ph}^2 + v_D^2 + v_{kTC}^2} . \quad \text{Gl. 3.71}$$

In diesem Gesamtrauschen eines PMD-Kanals wird das Photonenrauschen als physikalische Grenze erst dann dominant, wenn sich dieses bei einem ausreichenden Photostrom aus der Summe der restlichen Stromquellen deutlich erhebt (**Abb. 3.22**). Bedeutsam wird das Photonenrauschen erst oberhalb eines Photostroms von etwa $10pA$. Die anderen Rauschquellen können aber erst ab der nächsten Dekade, was einem Photostrom von $100pA$ entspricht, vernachlässigt werden.

Neben den bisher besprochenen PMD-bedingten Rauschquellen kommt durch die Schaltungstechnik der Pixelauslese eine weitere hinzu. Auf eine Beschreibung der CMOS-typischen Rauschquellen soll hier verzichtet werden. Stattdessen wurde für eine typische PMD-Pixel-Struktur das Rauschen simulativ ermittelt. Hierbei wurde 1/f- und das thermische Rauschen berücksichtigt. Da dieses Rauschen schaltungsbedingt tiefpassbeschränkt ist, ist der Frequenzgang der Ausleseelektronik (**Abb. 3.23**) hierbei

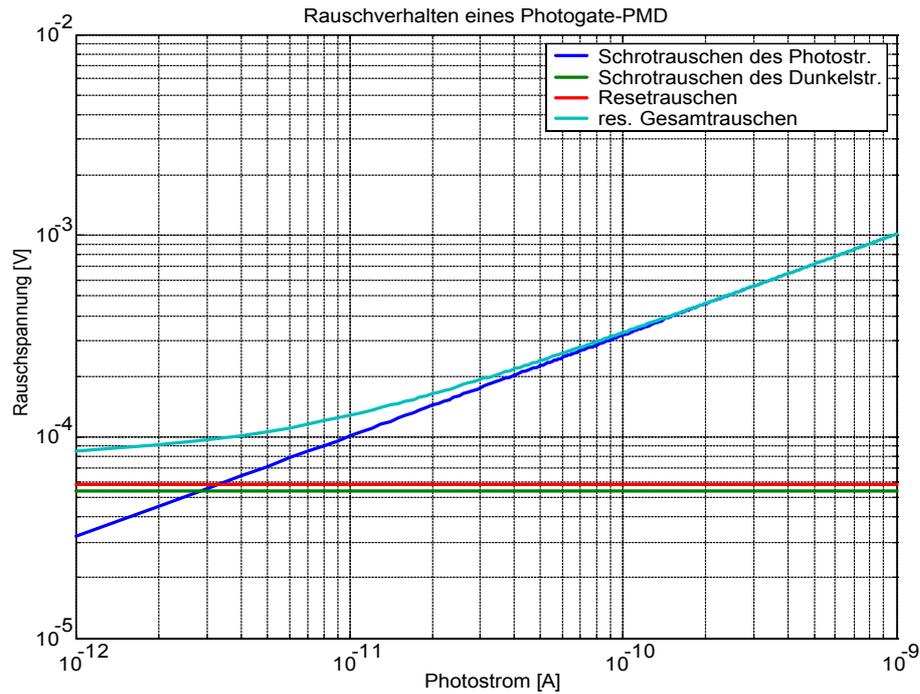


Abb. 3.22: Berechnete Rauschspannung eines Kanals einer typischen Photogate-PMD-Struktur

mit zu berücksichtigen. Des Weiteren ist zu beachten, dass alle vorangestellten bzw. in die Auslese eingespeisten Rauschquellen mit dem Verstärkungsfaktor der Auslese zu bewerten sind, um so das Gesamtrauschen eines PMD-Pixels mit Ausleseelektronik bestimmen zu können. Wie am Frequenzgang des Beispiels zu erkennen ist, besitzt dieses einen Verstärkungsfaktor von $k = 0,786$.

Aus der Simulation (**Abb. 3.24**) der spektralen Rauschspannungsdichte der Ausleseelektronik ist zu erkennen, dass ein starker niederfrequenter Anteil dominiert, der als $1/f$ -Rauschen interpretiert werden kann. Die Integration über die gesamte Verstärkerbandbreite ergibt eine mittlere Rauschspannung von:

$$v_V \approx 70 \mu V.$$

Das veränderte Rauschverhalten aufgrund des Eigenrauschens der Ausleseelektronik und ihrer Verstärkung zeigt **Abb. 3.25**. Aufgrund des starken niederfrequenten Rauschanteils sollte der Einfluss der höheren Frequenzen nicht unterschätzt werden. Durch einen nachgeschalteten Tiefpass mit einer Grenzfrequenz von $f_g = 100 \text{ kHz}$ verringert sich die mittlere Rauschspannung auf:

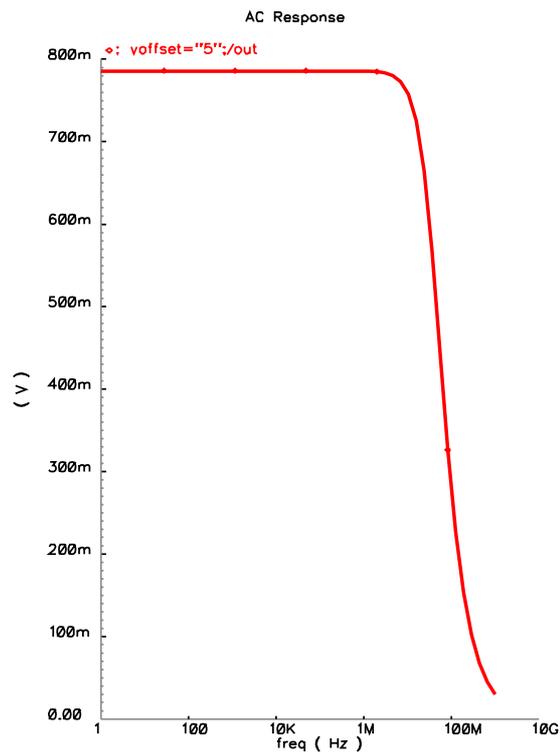


Abb. 3.23: Simulation der Tiefpassbegrenzung einer typischen Photogate-PMD-Auslese

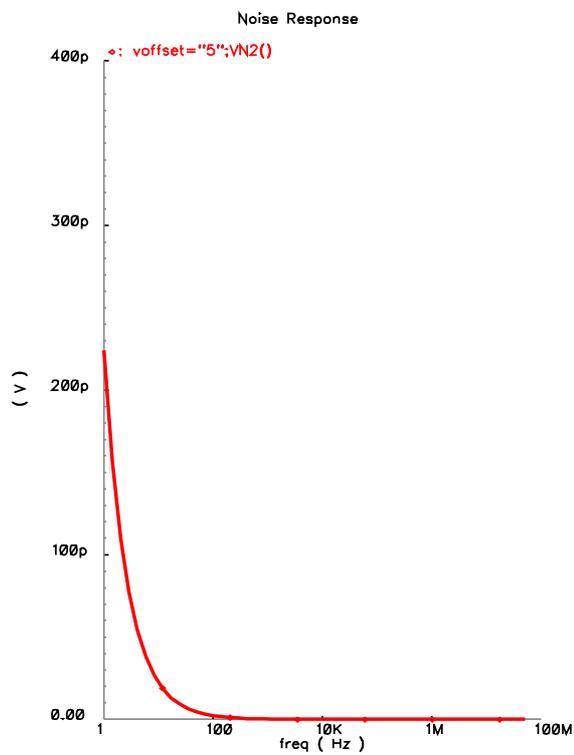


Abb. 3.24: Simuliertes Rauschspektrum einer typischen Photogate-PMD-Auslese

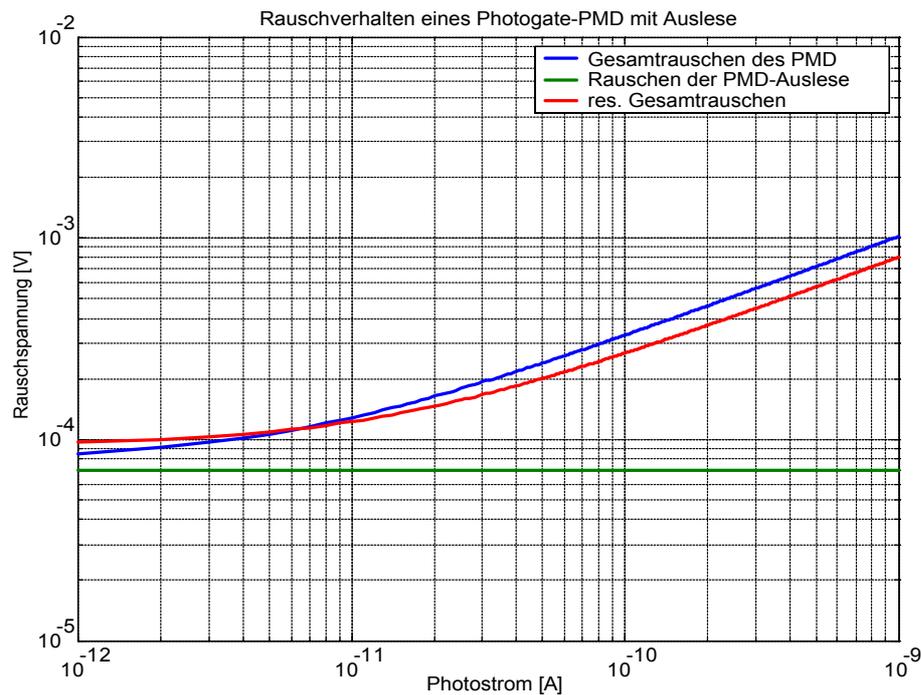


Abb. 3.25: Berechnete Rauschspannung eines Kanals einer typischen Photogate-PMD-Struktur mit Ausleseelektronik

$$v_V \approx 50 \mu V (0 \dots 100 \text{kHz}).$$

Das entspricht immerhin einer Reduzierung von fast 30%. Die Wahl der Grenzfrequenz kann jedoch nicht beliebig klein gewählt werden. Sie ist vielmehr mit der gewünschten Messrate eines PMD-basierten Sensors abzustimmen. In einem solchen Sensorsystem ist neben der Ausleseelektronik auch das Quantisierungsrauschen eines nachgeschalteten Analog-Digital-Wandlers zu berücksichtigen:

$$v_{ADC} = \frac{V_{LSB}}{\sqrt{12}}. \quad \text{Gl. 3.72}$$

Für eine zwölf Bit breite Auflösung eines Spannungshubes von $3V$ ergibt sich ein Quantisierungsrauschen von etwa $211 \mu V$. Das Gesamtrauschen (**Abb. 3.26**) eines PMD-Kanals mit Ausleseelektronik und AD-Wandler ergibt sich somit zu:

$$v = \sqrt{k^2 \cdot (v_{ph}^2 + v_D^2 + v_{kTC}^2) + v_V^2 + v_{ADC}^2}. \quad \text{Gl. 3.73}$$

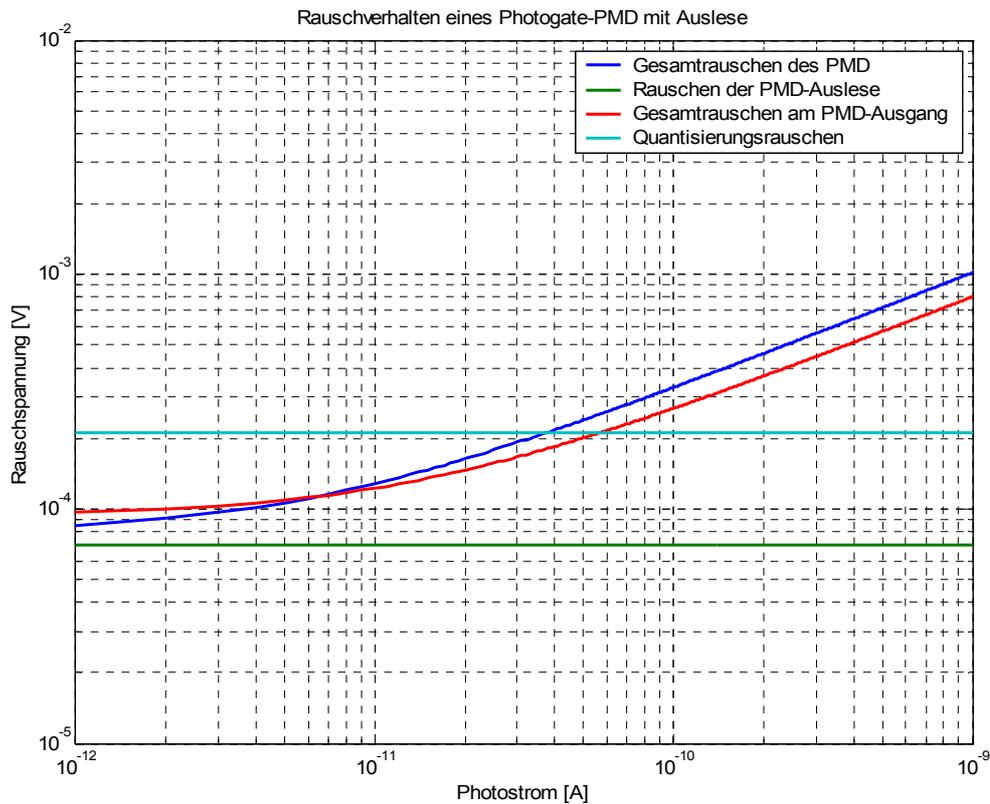


Abb. 3.26: Berechnete Rauschspannung eines Kanals einer typischen Photogate-PMD-Struktur mit Ausleseelektronik im Vergleich zum Quantisierungsrauschen eines 12Bit-AD-Wandlers

Sehr deutlich wird an dieser Stelle der Einfluss der Quantisierung bzw. des AD-Wandlers. Eine Erhöhung der Auflösung scheint angesichts der Größenordnung der anderen Rauschquellen angebracht. Jedoch zeigt die Praxis, dass in der technischen Umsetzung systembedingte Störquellen wie Spannungsversorgungsschaltungen, Quartzoszillatoren oder Logikbausteine zu noch höheren Rauschpegeln führen. Mit einer Erhöhung würde somit nur das Rauschen dieser dominanten Quellen genauer abgetastet werden.

3.5.4 MSM-PMD-Strukturen als quasioptische Mischer

Mit der Entwicklung der Lasertechnik wurde es möglich, sehr kurze Lichtpulse hoher Intensität zu generieren. Das Gegenstück hierzu, der konventionelle optische Empfang über Photodioden, verhält sich hierzu jedoch vergleichsweise langsam. Problematisch erweist sich hier der Zusammenhang zwischen optischer Empfangsfläche und Kapazität des pn-Übergangs (**Kap. 3.4.1**). Ein neuer Ansatz bildet hier die Metal-Semiconductor-Metal-Struktur, dessen Dynamik nur durch die Ladungsträgergeneration selbst begrenzt ist. Dieser Zusammenhang lässt sich sehr einfach an der in **Abb. 3.27** dargestellten Struktur erläutern. Hier befindet auf der Unterseite des Halbleitermaterials eine durchgehende Metallschicht, wogegen auf der Oberseite ein Metallstreifen mit einer kleinen Unterbrechung aufgebracht ist. Diese stellt ein optisches Fenster dar, durch das Lichtwellen in das Halbleitermaterial eindringen können. Die Abhängigkeit ihrer Eindringtiefe von der verwendeten Lichtwellenlänge führt in dieser Struktur zu völlig unterschiedlichem Verhalten. Mit einer kurzwelligen λ_1 Bestrahlung erfolgt die photonische Ladungsträgergeneration sehr nahe der Oberfläche. Die Metalllücke wird so gebrückt und ein elektrisches Signal kann auf den Ausgang durchgeschaltet werden. Eine langwellige Bestrahlung λ_2 jedoch generiert Ladungsträger auch in wesentlich tieferen Schichten, welche nun einen Stromfluss zur Metallschicht auf der Unterseite ermöglichen. Dadurch wird das am Eingang anliegende Signal kurzgeschlossen und kann auch nicht kapazitiv zum Ausgang überkoppeln. Vereinfachend könnte man sagen,

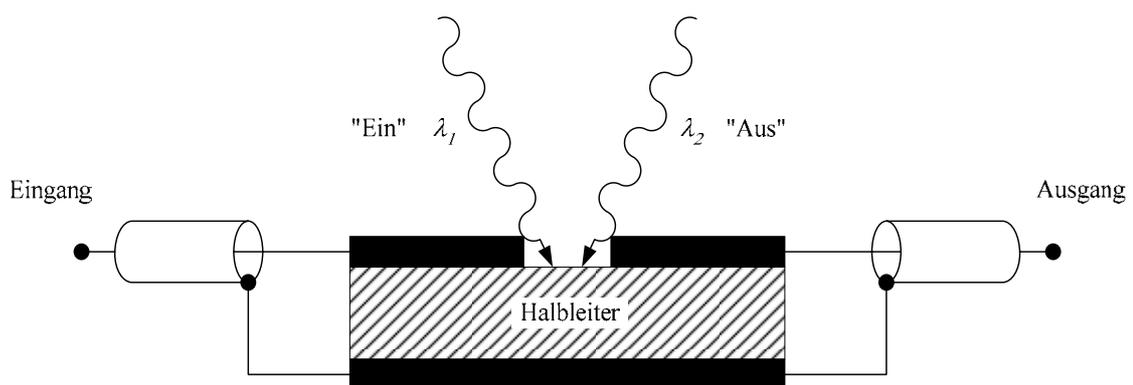


Abb. 3.27: Optoelektronischer Schalter auf Basis eines MSM-Übergangs¹⁵

¹⁵ D.H. Auston; „Picosecond optoelectronic switching and gating in silicon“, Applied Physics Letters 1975 Vol. 26, Issue 3, S. 101-103

dass das Schaltverhalten dieser Struktur nur von der Breite der Metallücke und der Ladungsträgergeneration abhängt. Im Detail können jedoch auch in einem Metall-Halbleiterkontakt Potentialbarrieren bestehen, für deren Überwindung zusätzliche Arbeit durch einen Ladungsträger zu verrichten ist. In Abhängigkeit von der Dotierung des Halbleitermaterials können hierbei verschieden Fälle unterschieden werden.¹⁶ Wesentlich hierbei ist jedoch nur, dass sowohl Ohmsche als auch gleichrichtende Kontakte hergestellt werden können. Die gleichrichtende Eigenschaft eines solchen Kontakts wurde durch Schottky erstmals 1938 beschrieben und durch eine Potentialbarriere erklärt.¹⁷ Diese sogenannte Schottkybarriere führt natürlich zur Ausprägung einer Raumladungszone, welcher wie bei den PIN-Photodioden eine Kapazität zuzuordnen ist.

Das gleichrichtende Verhalten eines einzelnen Metall-Halbleiterkontakts führt zu einer Diodenkennlinie, vergleichbar zu PN-Dioden. MSM-Dioden bestehen jedoch aus zwei entgegengesetzt gepolten Dioden. Theoretisch würde für jede Polarisationsrichtung immer eine der beiden Dioden sperren und einen Stromfluss verhindern. In der jeweils gesperrten Diode kommt es jedoch zu der Ausbildung einer Raumladungszone. Ähnlich einer herkömmlichen in Sperrrichtung betriebenen PIN-Photodiode kann nun trotzdem ein Stromfluss zustande kommen, wenn durch Temperatur oder Lichteinstrahlung freie Ladungsträger innerhalb der Raumladungszone generiert werden. Da die Ausbreitung der Raumladungszone abhängig von der anliegenden Sperrspannung ist, kann mitunter ein Zusammenhang zwischen Photostrom und Spannung beobachtet werden. Die Ursache hierfür liegt darin, dass bei kleinen Spannungen die Raumladungszone kleiner ist als die wellenlängenabhängige Eindringtiefe. Ein gewisser Teil der Ladungsträgergeneration erfolgt dann außerhalb und kann nicht zum Photostrom beitragen. Die resultierende Übertragungskennlinie einer solchen belichteten MSM-Struktur zeigt **Abb. 3.28**.

Im dynamischen Betrieb zeigt die MSM-Struktur ein Verhalten, das dem einer Photogate-PMD-Struktur sehr ähnlich ist, jedoch auch noch zusätzliche Vorteile aufweist. Der funktionelle Zusammenhang wird in der **Abb. 3.29** dargestellt. Je nach Polarität der Diode fließt hier der in der Raumladungszone generierte Photostrom über die Diode A oder B ab. Ist die eingestrahlte Lichtleistung gleichgroß, dann wird in

¹⁶ E.H. Roderick u. R.H. Williams; „Metal-Semiconductor Contacts“, 2nd ed., Clarendon Press, Oxford (1988)

¹⁷ W. Schottky; Naturwissenschaften 26 (1938), S. 843

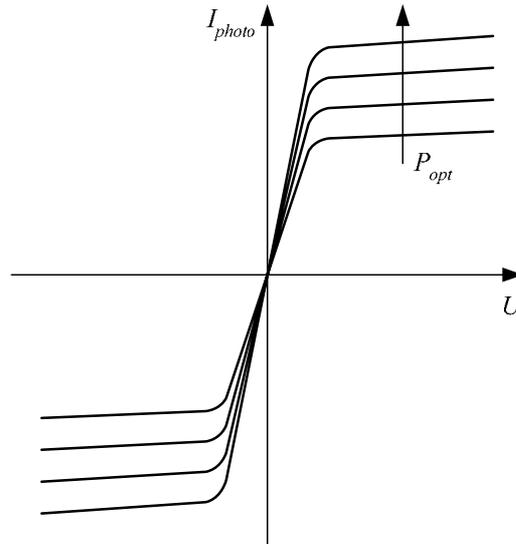


Abb. 3.28: DC-Kennlinie einer MSM-Struktur

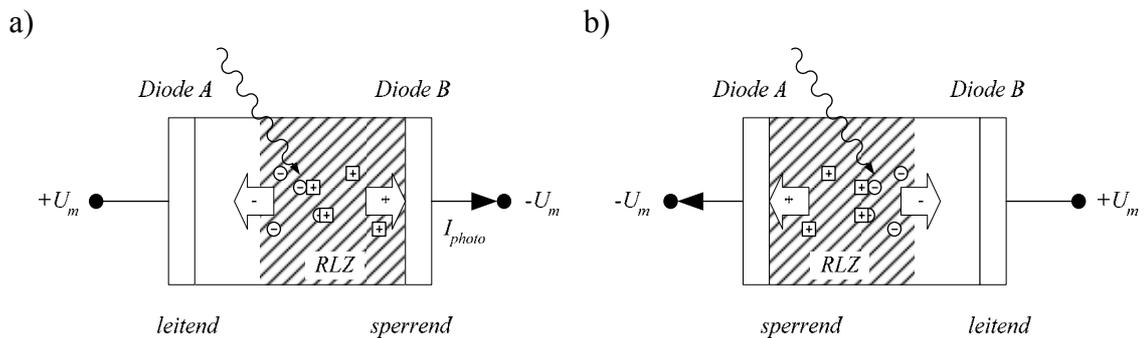


Abb. 3.29: Dynamische Gegentaktsteuerung einer MSM-Photodiode mittels einer Modulationsspannung U_m

beiden Fällen der gleiche Photostrom generiert, der sich nur im Vorzeichen unterscheidet. Ist jedoch das optische Signal intensitätsmoduliert, dann werden sich die fließenden Ströme auch in ihrer Größe unterscheiden. Damit wird ersichtlich, dass die MSM-Diode ähnlich dem Photogate-PMD einen quasioptischen Mischer darstellt. Zwischen beiden findet sich jedoch ein entscheidender Unterschied im Umgang mit unmodulierter Lichteinstrahlung. Ein solches Gleichlicht fließt im Photogate-PMD über beide Auslesedioden ab und sammelt sich auf den Kapazitäten der Integratoren. Bei einer MSM-Diode hingegen fließen im ersten Moment die generierten Ladungen

beispielsweise von der Diode A über die Diode B ab. Die Integratoren beider Seiten unterliegen diesem Stromfluss und werden zwar in gleicher Größe aber mit unterschiedlichem Vorzeichen reagieren. Im nachfolgenden Zyklus mit umgekehrter Polarität wird das Gleichlicht Ladungen gleicher Anzahl generieren. Der resultierende Photostrom fließt in der gleichen Größe nun mit umgekehrten Vorzeichen über die Dioden A und B. Die Integratoren beider Seiten werden nun ebenso entgegengesetzt reagieren und die zuvor gesammelten Ladungen nun wieder ausgleichen. Damit wird klar ersichtlich, dass MSM-Dioden im Gegensatz zum Photogate-PMD eine inhärente Hintergrundunterdrückung besitzen, da ihre Dioden das Rückfließen von Ladungen nicht verhindern. Wichtig hierbei ist jedoch, dass ein Zurückfließen nur in der Größenordnung des gerade generierten Photostroms möglich ist. Intensitätsmodulierte Lichteinstrahlung wird somit immer eine Spannungsdifferenz an den beiden Integratoren erzeugen.

3.6 Optischer Korrelationsempfang

Die Detektion eines optischen Signals stellt vordergründig betrachtet kein Problem dar, jedoch ist man in vielen Applikationen sehr stark in der empfangenen Lichtleistung begrenzt. Es ist daher sinnvoll, die verwendeten optischen Detektoren mit zusätzlichen Integratoren zu kombinieren. Dadurch können über einen festen Zeitraum photonisch generierte Ladungen gesammelt werden. Diese Verbindung von Detektion und Integration ist ein sehr altes Prinzip, welches schon seit langem in der Fotografie angewendet wird. Dort wird das in herkömmlichen Fotoapparaten verwendete Filmmaterial über ein bestimmte Zeit belichtet, wobei hier die Integrationszeit mit der sogenannten Belichtungszeit festgelegt wird. Das Filmmaterial selbst bildet hierbei den Integrator. Der Signalgewinn durch die Integration der Empfangsleistung ist natürlich mit einem Nachteil behaftet. Schnelle zeitliche Änderungen können so nicht mehr beobachtet werden. Besitzt man ein gewisses Vorwissen über das empfangene Signal, kann trotz der langen Integrationszeiten sein dynamisches Verhalten ermittelt werden, indem man es mit einem ähnlichen Signal vergleicht. Der Prozess des Vergleichens wird in Mathematik und Systemtheorie als Korrelation bezeichnet. Der allgemeine analytische Ausdruck hierfür lautet:

$$\rho_{12}(v) = \overline{a_1(u) \cdot a_2(u+v)} \quad \text{Gl. 3.74}$$

$$= \lim_{U \rightarrow \infty} \frac{1}{2U} \int_{-U}^U a_1(u) \cdot a_2(u+v) du .$$

Wählt man beispielsweise zwei um 90° phasenverschobene harmonische Schwingungen $a_1 = \cos(\omega \cdot t)$ und $a_2 = \sin(\omega \cdot t)$, dann erhält man für $u=t$ und $v=\tau$:

$$\rho_{12}(\tau) = \frac{1}{2} \sin(\omega \cdot \tau) . \quad \text{Gl. 3.75}$$

Bei dieser Korrelationsfunktion gilt für $\tau=0$:

$$\rho_{12}(0) = 0 . \quad \text{Gl. 3.76}$$

Mit diesem Ergebnis bezeichnet man die beiden Funktionen a_1 und a_2 als zueinander orthogonal, da beide harmonische Schwingungen in einem Phasenwinkel von 90° zueinander stehen. In diesem Zustand sind sie beide maximal unähnlich. Wird der Phasenwinkel kleiner 90° , dann erhöht sich die Ähnlichkeit solange, bis der Phasenwinkel 0° erreicht. Für diesen Fall ist das Korrelationsergebnis eins. Es besteht also maximale Ähnlichkeit. Verändert sich der Phasenwinkel in entgegen gesetzter Richtung, dann steigt die Ähnlichkeit ebenso an, ist aber mit einem negativen Vorzeichen behaftet. Für einen Phasenwinkel von 180° wird somit eine maximale negative Ähnlichkeit erreicht, so dass die Korrelation als Ergebnis minus eins liefert.

Will man eine Korrelation technisch realisieren, dann geht aus Gl. 3.74 deutlich hervor, dass hierzu eine Multiplikation bzw. ein Mischen beider Signale erforderlich ist, dessen Ergebnis anschließend über die Zeit zu integrieren ist (**Abb. 3.30**). Hierbei stellt sich das Problem der nach Gl. 3.74 unendlichen Integrationszeiten. Reale Korrelatoren können nur zeitlich begrenzt arbeiten, was mitunter in der systemtheoretischen Beschreibung durch das Einbringen eines weiteren Rechteckimpulses zu berücksichtigen ist. Ist die Integrationszeit gegenüber der Periodendauer der zu korrelierenden Signale wesentlich größer, dann kann dieser Umstand vernachlässigt werden. Für den Korrelationsempfang optischer Signale reicht die empfangene

Signalleistung oft nicht aus, so dass diese nach der Signaldetektion oftmals noch breitbandig zu verstärken ist (**Abb. 3.31**). Realisiert man aber einen Korrelationsempfang mit einer Photogate- oder MSM-PMD-Struktur, dann entfällt dieser Verstärker und der Mischprozess erfolgt direkt während des Detektionsprozesses. Man bezeichnet ihn deshalb als quasioptisch.

Des Weiteren bieten PMD-Strukturen den Vorteil des zweikanaligen Aufbaus (**Abb. 3.32**). Da der Empfang eines intensitätsmodulierten optischen Signals naturgemäß nur positive Leistungswerte annehmen kann, kann in einkanaligen Aufbauten bei ungünstigen Phasenlagen die empfangene Signalleistung vollständig unterdrückt werden. So kann z.B. bei $\varphi = 90^\circ$ das Mischen so verstanden werden, dass die maximale Signalamplitude mathematisch mit Null multipliziert wird. Im zweikanaligen Aufbau hingegen wird das empfangene Signal im Gegenteil gemischt, so dass im ungünstigsten Fall in einem Zweig zwar gar keine Signalleistung erscheint, diese jedoch dann vollständig im anderen auftritt. Mit einem zweikanaligen Aufbau ist es somit möglich, unabhängig von der Phasenlage immer die volle detektierte Signalleistung auszuwerten.

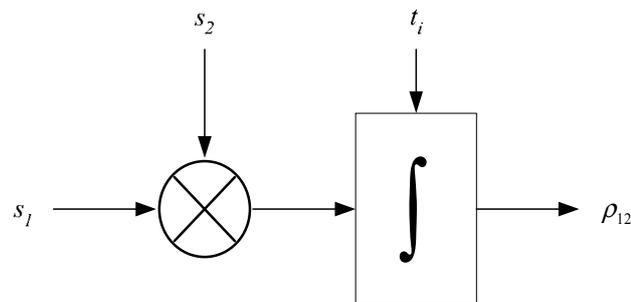


Abb. 3.30: Blockschaltbild eines allgemeinen technischen Korrelators

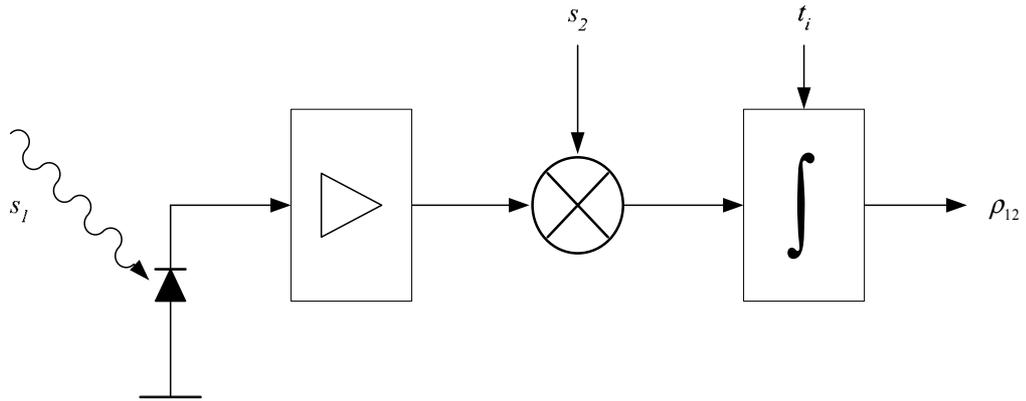


Abb. 3.31: Blockschaltbild des optischen Korrelationsempfangs

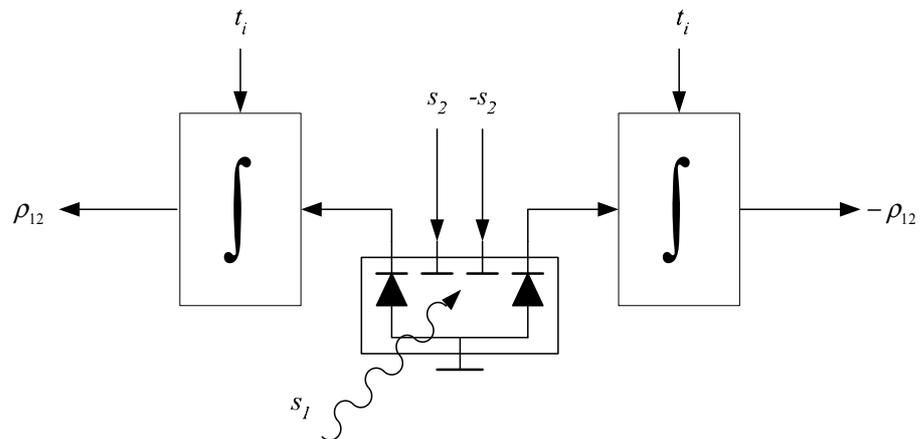


Abb. 3.32: Blockschaltbild des zweikanaligen optischen Korrelationsempfangs mittels einer Photogate-PMD-Struktur

4 Phasenauswertung intensitätsmodulierter Signale der optischen Messtechnik mittels Phase-Shift-Interferometrie

Die Phaseninformation eines Signals ist in vielen technischen Anwendungen von großer Bedeutung, da sie oft ein Abbild wichtiger messtechnischer Größen darstellt, wie z.B. der Entfernung, Beschleunigung oder Geschwindigkeit. So beruhen viele Verfahren in Geodäsie, physikalischer Chemie oder Elektrotechnik auf der Bestimmung der Phasenlage zweier Signale zueinander.

Aber auch in der Informationsübertragung hat der Begriff der Phase seine Bedeutung. Schon in ihren Anfängen stand man in der Funktechnik dem Problem gegenüber, dass über weite Übertragungstrecken hinweg keine sichere Aussage über die Signallaufzeit und damit die Signalphase getroffen werden konnte. In solch klassischen Übertragungsverfahren, wie z.B. dem Radioempfang, wird dieses Problem durch einen Phase-Locked-Loop (PLL) gelöst. Hierbei wird die Phase eines frequenzvariablen lokalen Oszillators zum Empfangssignal synchronisiert, was natürlich ein Messen der Phase zwischen beiden voraussetzt. Geeignete Verfahren und Algorithmen zur Phasenmessung gewinnen noch zusätzlich an Bedeutung, da sich die Problematik unterschiedlicher Signallaufzeiten in den letzten Jahren durch die gestiegenen Übertragungsgeschwindigkeiten und damit erhöhten Taktraten noch zusätzlich verstärkte.

Phasenmessverfahren rücken aber auch im Bereich der klassischen Messtechnik immer stärker ins Blickfeld, da eine Vielzahl berührungsloser Messverfahren auf der Auswertung von Frequenz, Amplitude oder Phase einer ausgestrahlten elektromagnetischen Welle beruhen. Ein in vielen optischen Anwendungen geeignetes Verfahren stellt die Phase-Shift-Interferometrie dar, da sie die Möglichkeit bietet, mit relativ wenigen Signalabtastungen den räumlichen oder zeitlichen Kohärenzgrad (also einen hochgenauen Phasen- oder Amplitudenwert) zu ermitteln.

In diesem Kapitel wird, ausgehend von der historischen Entwicklung, das Grundprinzip der Phase-Shift-Interferometrie erläutert, um darauf aufbauend deren Algorithmen auf den optischen Korrelationsempfang anzuwenden, zu optimieren und auf mehrdimensionale Systeme auszuweiten. Abschließend werden die gewonnenen Erkenntnisse an Applikationsbeispielen näher erläutert, wobei auch deren spezifische Probleme nicht außer Acht gelassen werden.

4.1 Prinzip der Phase-Shift-Interferometrie

Klassische interferometrische Verfahren beruhen auf der Überlagerung zweier Lichtwellen, was im Allgemeinen als Interferenz bezeichnet wird. Mathematisch kann dieser Vorgang als eine Summation bzw. Superposition zweier Sinusschwingungen betrachtet werden. Betrachtet man die Feldstärken zweier sich überlagernten elektromagnetischen Wellen, dann gilt:

$$E(x, t) = E_{01} \cdot \cos\left[\omega_1 \cdot \left(t - \frac{x}{c}\right) + \varphi_1\right] + E_{02} \cdot \cos\left[\omega_2 \cdot \left(t - \frac{x}{c}\right) + \varphi_2\right]. \quad \text{Gl. 4. 1}$$

Betrachtet man dies Überlagerung an einem festen Ort, dann reduziert sich dieser Ausdruck auf:

$$\begin{aligned} E(t) &= E_{01} \cdot \cos[\omega_1 \cdot t + \varphi_1] + E_{02} \cdot \cos[\omega_2 \cdot t + \varphi_2] \\ &= E_{01} \cdot e^{j(\omega_1 t + \varphi_1)} + E_{02} \cdot e^{j(\omega_2 t + \varphi_2)} \end{aligned} \quad \text{Gl. 4. 2}$$

Sind die Wellenlängen ω_1 und ω_2 gleich, dann lässt sich dieser Ausdruck allein auf die Phasen φ_1 und φ_2 reduzieren:

$$E(t) = E_{01} \cdot e^{j(\varphi_1)} + E_{02} \cdot e^{j(\varphi_2)}. \quad \text{Gl. 4. 3}$$

Für die Beobachtung der Interferenz zweier elektromagnetischer Wellen ergibt sich nach dem Superpositionsprinzip:

$$\begin{aligned} \left| \vec{E}_{01} + \vec{E}_{02} \right|^2 &= \left(\vec{E}_{01} + \vec{E}_{02} \right) \cdot \left(\vec{E}_{01} + \vec{E}_{02} \right)^* \\ &= \left(E_{01} \cdot e^{i\varphi_1} + E_{02} \cdot e^{i\varphi_2} \right) \cdot \left(E_{01} \cdot e^{-i\varphi_1} + E_{02} \cdot e^{-i\varphi_2} \right) \\ &= E_{01}^2 + E_{02}^2 + 2 \cdot E_{01} \cdot E_{02} \cdot \cos(\varphi_2 - \varphi_1) \end{aligned} \quad \text{Gl. 4. 4}$$

Hierbei wird der Ausdruck $2 \cdot E_{01} \cdot E_{02} \cdot \cos(\varphi_2 - \varphi_1)$ im Allgemeinen als Interferenzterm bezeichnet. Die Phasendifferenz $\Delta\varphi = \varphi_2 - \varphi_1$ ergibt sich in solch klassischen Verfahren aus der Wellenzahl k , der Lichtwellenlänge λ , dem Brechungsindex n des Mediums und der Längendifferenz Δs beider Ausbreitungswege:

$$\Delta\varphi = k \cdot n \cdot \Delta s = \frac{2 \cdot \pi}{\lambda} \cdot n \cdot \Delta s . \quad \text{Gl. 4. 5}$$

Überlagern sich jedoch zwei Wellen unterschiedlicher Längen ω_1 und ω_2 , dann ergibt sich bei gleichen Amplituden $E_{01} = E_{02}$:

$$\begin{aligned} E(x,t) &= E_{01} \cdot \cos(k_1 \cdot x - \omega_1 \cdot t) + E_{02} \cdot \cos(k_2 \cdot x - \omega_2 \cdot t) \\ &= E_{01} [\cos(k_1 \cdot x - \omega_1 \cdot t) + \cos(k_2 \cdot x - \omega_2 \cdot t)] \end{aligned} \quad \text{Gl. 4. 6}$$

Mit folgenden Definitionen

$$\bar{\omega} = \frac{1}{2}(\omega_1 + \omega_2), \quad \text{Gl. 4. 7}$$

$$\omega_m = \frac{1}{2}(\omega_1 - \omega_2), \quad \text{Gl. 4. 8}$$

$$\bar{k} = \frac{1}{2}(k_1 + k_2) \quad \text{Gl. 4. 9}$$

und

$$k_m = \frac{1}{2}(k_1 - k_2) \quad \text{Gl. 4. 10}$$

kann **Gl. 4. 6** folgendermaßen umformuliert werden:

$$E(x, t) = 2 \cdot E_{01} \cdot \cos(k_m \cdot x - \omega_m \cdot t) \cdot \cos(\bar{k} \cdot x - \bar{\omega} \cdot t). \quad \text{Gl. 4. 11}$$

Deutlich ist zu erkennen, dass das resultierende Interferenzsignal aus der Überlagerung der zwei Frequenzen ω_m und $\bar{\omega}$ besteht. Man spricht hierbei von einer sich einstellenden Schwebung.

Die Aufnahme eines Interferogramms (also des zeitlichen oder örtlichen Verlaufs mehrerer sich überlagernder Wellen) soll hier nicht im Detail erläutert werden. Es sei hier nur darauf hingewiesen, dass man grundsätzlich zwischen sequentieller und räumlicher Interferometrie unterscheidet. Für die weitere Betrachtung ist es jedoch nur entscheidend, dass alle klassischen Interferometrie-Verfahren sich auf die Wellenlänge λ des verwendeten Lichtes beziehen. Ihre Algorithmen jedoch dienen im Wesentlichen nur dazu, aus den gewonnenen Messwerten die Phasendifferenz $\Delta\varphi$ zu ermitteln, unabhängig davon wie diese Daten gewonnen wurden. Es ist also nahe liegend, solche Algorithmen auch in der intensitätsmodulierten Optosensorik anzuwenden, wobei hier die Lichtwellenlänge durch die Modulationswellenlänge zu ersetzen ist. So eignen sie sich auch im Bereich des hier verwendeten Korrelationsempfangers hervorragend zur Signalauswertung und ermöglichen so eine genaue Erfassung stark leistungsbegrenzter intensitätsmodulierter optischer Signale.

4.2 Auswertung von Phase-Shift-Interferogrammen

4.2.1 Phasenbestimmung durch Phase-Shift-Interferometrie

Alle Parameter eines gemessenen Interferogramms können näherungsweise dadurch bestimmt werden, dass eine geeignete Funktion mit größtmöglicher Ähnlichkeit in die Messwerte approximiert wird. Die hier verwendete Approximation beruht auf der Methode der kleinsten Fehlerquadrate. Dabei wird eine Zielfunktion f so angepasst, dass die Summe aller quadrierten Abweichungen zwischen realen Messwerten und Zielfunktion minimal wird. Besitzen die Messwerte einen periodischen Charakter, dann können zunächst die Überlegungen auf die Grundwelle einer periodischen Funktion beschränkt werden. Die Zielfunktion lautet dann:

$$f(t) = \frac{a_0}{2} + A \cdot \sin(\omega \cdot t + \varphi). \quad \text{Gl. 4. 12}$$

Nimmt man für eine Messreihe linearen Charakter an, dann sind die einzelnen Messwerte mit gleich bleibendem Abstand über den gesamten Messbereich verteilt. Zur Phasenbestimmung reicht grundsätzlich die Vermessung einer Periode aus, aber auch über mehrere Perioden verteilte Messungen können grundsätzlich ausgewertet werden. Unter der Annahme, dass die Messreihe aus N Messwerten linear über eine Periode verteilt ist, gilt für den Phasenversatz φ_S zwischen zwei Messwerten:

$$\varphi_S = \frac{2\pi}{N}. \quad \text{Gl. 4. 13}$$

Die Aufnahme jedes einzelnen Messwertes Y_n erfolgt dabei zum Zeitpunkt

$$\omega \cdot t_n = \varphi_S \cdot n. \quad \text{Gl. 4. 14}$$

Die Differenz zwischen dem tatsächlichen Messwert Y_n und dem approximierten Wert $f(t_n)$ stellt den Fehler s dar, den es zu minimieren gilt. Für die Approximation nach der Methode der kleinsten Fehlerquadrate lautet die so genannte quadratische Fehlerfunktion:

$$s^2 = \sum_{n=1}^N \left[Y_n - \frac{a_0}{2} - A \cdot \sin(\varphi_S \cdot n + \varphi) \right]^2. \quad \text{Gl. 4. 15}$$

Das Ziel der minimalen Abweichung des Phasenwertes zwischen gemessenem (bzw. abgetastetem) Funktionsverlauf und der approximierten Funktion erfordert eine Minimierung der quadratischen Fehlerfunktion E^2 nach der Phase φ . Mathematisch bedeutet dies die Ableitung der quadratischen Fehlerfunktion E^2 nach φ , welche anschließend zu Null zu setzen ist:

$$\frac{ds^2}{d\varphi} = 2 \sum_{n=1}^N \left[Y_n - \frac{a_0}{2} - A \cdot \sin(\varphi_S \cdot n + \varphi) \right] \cdot (-A \cdot \cos(\varphi_S \cdot n + \varphi)) = 0. \quad \text{Gl. 4. 16}$$

Die Vereinfachung dieses Ausdrucks, entsprechend der im **Anhang A.1** dokumentierten ausführlichen Herleitung, ergibt die allgemeine Formel des Phase-Shift-Algorithmus für eine beliebige Anzahl N an Messwerten Y_n innerhalb eine Periode:

$$\tan \varphi = \frac{\sum_{n=1}^N Y_n \cdot \cos(\varphi_S \cdot n)}{\sum_{n=1}^N Y_n \cdot \sin(\varphi_S \cdot n)}. \quad \text{Gl. 4. 17}$$

Dieser allgemeine Ausdruck wurde in einer ähnlichen Form erstmals durch C.J. Morgan für ein Twyman-Green-Interferometer beschrieben.¹ Hierin erscheinen die Ausdrücke im Zähler und Nenner vertauscht, was auf die unterschiedliche Zielfunktion (Cosinus) im Ansatz seiner Herleitung zurückzuführen ist.

4.2.2 Systematische Approximationsfehler in Phase-Shift-Verfahren

Die Genauigkeit eines approximierenden Verfahrens wird im ersten Schritt durch die Wahl einer mehr oder weniger geeigneten Zielfunktion wesentlich mitbestimmt. Für die Phase-Shift-Verfahren nach **Gl. 4. 17** ist die Zielfunktion durch eine Sinusschwingung gegeben. Je weniger die abgetastete Funktion dieser Zielfunktion gleicht, desto fehlerhafter ist deren Approximation. Die Abweichung Δy_n eines Abtastwertes zur Zielfunktion ergibt sich aus:

$$Y_n = A_1 \cdot \sin\left(\frac{2\pi}{N} \cdot n + \varphi\right) + \Delta y_n. \quad \text{Gl. 4. 18}$$

Für die Phasenbestimmung gilt damit:

¹ C. J. Morgan; Optics Letters No. 7, 1982, S. 368-70

$$\tan(\varphi + \Delta\varphi) = \frac{\sum_{n=1}^N \left[A_1 \cdot \sin\left(\frac{2\pi}{N}n + \varphi\right) + \Delta y_n \right] \cdot \cos(\varphi_S \cdot n)}{\sum_{n=1}^N \left[A_1 \cdot \sin\left(\frac{2\pi}{N}n + \varphi\right) + \Delta y_n \right] \cdot \sin(\varphi_S \cdot n)}. \quad \text{Gl. 4. 19}$$

Die Abweichungen im Zähler ΔZ und Nenner ΔN dieses Ausdrucks kann folgendermaßen zusammengefasst werden:

$$\Delta Z = \sum_{n=1}^N [\cos(\varphi_S \cdot n) \cdot \Delta y_n], \quad \text{Gl. 4. 20}$$

$$\Delta N = \sum_{n=1}^N [\sin(\varphi_S \cdot n) \cdot \Delta y_n]. \quad \text{Gl. 4. 21}$$

Normiert man diesen Fehler auf die Amplitude A_1 der Grundwelle, mit

$$z = \frac{\Delta Z}{A_1} \quad \text{Gl. 4. 22}$$

und

$$n = \frac{\Delta N}{A_1} \quad \text{Gl. 4. 23}$$

dann lässt sich der Fehler $\Delta\varphi$ über Additionstheoreme bestimmen zu:²

$$\tan(\Delta\varphi) = \frac{z \cdot \cos \varphi - n \cdot \sin \varphi}{1 + n \cdot \cos \varphi + z \cdot \sin \varphi}. \quad \text{Gl. 4. 24}$$

² K.G. Larkin, "Propagation of errors in different phase-shifting algorithms: a special property of the arctangent function", SPIE Conference on Interferometry: Techniques and Analysis, SPIE Proceeding Vol. 1755, San Diego, 1992

Mit dieser grundlegenden Beziehung lässt sich für unterschiedlichste Fälle der resultierende Phasenfehler quantitativ bestimmen.

4.2.3 Statistische Approximationsfehler in Phase-Shift-Verfahren

Neben den zuvor beschriebenen systematischen Einflüssen unterliegen Phase-Shift-Verfahren wie alle realen Systeme Rauschprozessen, die sich ebenso auf die Phasenbestimmung auswirken und nur statistisch beschreiben lassen. Da jedoch die statistische Angabe der Standardabweichung kein Vorzeichen einschließt und ein Vorzeichenwechsel in **Gl. 4. 24** zu unterschiedlichen Ergebnissen führt, kann diese zur Berechnung des statistischen Phasenfehlers nicht herangezogen werden. Mit Hilfe des Gaußschen Fehlerfortpflanzungsgesetzes kann dieser Fehler über die partiellen Ableitungen bestimmt werden. Nimmt man einen beliebigen Phase-Shift-Algorithmus mit

$$\varphi = \arctan \frac{\sum_{n=1}^N G_n^N \cdot Y_n}{\sum_{n=1}^N G_n^Z \cdot Y_n}, \quad \text{Gl. 4. 25}$$

wobei G_n^N und G_n^Z die Gewichtung der Messwerte Y_n jeweils im Zähler und Nenner darstellen, dann gilt:

$$\sigma_\varphi^2 = \sqrt{\sum_{n=1}^N \left\{ \left| \frac{\partial \varphi}{\partial Y_n} \right|^2 \cdot \sigma_{Y_n}^2 \right\}}. \quad \text{Gl. 4. 26}$$

Mit der Ableitung der Arcustangensfunktion

$$\frac{d}{dz} \arctan(z) = \frac{1}{1+z^2} \quad \text{Gl. 4. 27}$$

und der Ableitung seines Arguments z nach einem einzelnen Messwert Y_n

$$\frac{dz}{dY_m} = \frac{G_m^Z \cdot \sum_{n=1}^N (G_n^N \cdot Y_n) - G_m^N \cdot \sum_{n=1}^N (G_n^Z \cdot Y_n)}{\left(\sum_{n=1}^N (G_n^N \cdot Y_n) \right)^2} \quad \text{Gl. 4. 28}$$

lautet der Phasenfehler:

$$\sigma_\varphi^2 = \sum_{m=1}^N \left[\frac{G_m^Z \cdot \sum_{n=1}^N (G_n^N \cdot Y_n) - G_m^N \cdot \sum_{n=1}^N (G_n^Z \cdot Y_n)}{\left\{ \sum_{n=1}^N (G_n^N \cdot Y_n) \right\}^2 + \left\{ \sum_{n=1}^N (G_n^Z \cdot Y_n) \right\}^2} \right]^2 \cdot \sigma_{Y_m}^2 \quad \text{Gl. 4. 29}$$

Normiert man die Messwerte Y_n zusätzlich auf die Amplitude A , dann vereinfacht sich der Phasenfehler für beliebige Abtastungen N zu:

$$\sigma_\varphi^2 = \frac{1}{A^2} \cdot \sum_{m=1}^N \left[\frac{G_m^Z \cdot \sum_{n=1}^N (G_n^N \cdot y_n) - G_m^N \cdot \sum_{n=1}^N (G_n^Z \cdot y_n)}{\left\{ \sum_{n=1}^N (G_n^N \cdot y_n) \right\}^2 + \left\{ \sum_{n=1}^N (G_n^Z \cdot y_n) \right\}^2} \right]^2 \cdot \sigma_{Y_m}^2 \quad \text{Gl. 4. 30}$$

Unter der Voraussetzung, dass alle Messwerte Y_n gleichmäßig über das Intervall $\{0; 2\pi\}$ verteilt sind bzw. der Phaseshift $\varphi_S = 2\pi/N$ beträgt, gilt:

$$\left\{ \sum_{n=1}^N (G_n^N \cdot y_n) \right\}^2 + \left\{ \sum_{n=1}^N (G_n^Z \cdot y_n) \right\}^2 = \frac{N}{2} \quad \text{Gl. 4. 31}$$

Nimmt man für alle Abtastwerte N ein konstantes Rauschen δ_y an, dann gilt zusätzlich:

$$\sum_{n=1}^N \left[G_m^Z \cdot \sum_{n=1}^N (G_n^N \cdot y_n) - G_m^N \cdot \sum_{n=1}^N (G_n^Z \cdot y_n) \right]^2 = \frac{N}{2} \quad \text{Gl. 4. 32}$$

Das durch ein Rauschen in der Abtastung verursachte Phasenrauschen lässt sich somit berechnen aus:

$$\sigma_{\varphi} = \frac{1}{A} \cdot \sqrt{\frac{2}{N}} \cdot \sigma_Y. \quad \text{Gl. 4. 33}$$

Die Richtigkeit dieses Ausdrucks konnte simulativ geprüft werden (**Abb. 4.1**). Seine Genauigkeit wird im Gegensatz zur Literatur³ mit zunehmender Rauschleistung nicht schlechter, da im dargestellten Ansatz keine Näherungen (wie $\sin(x)=x$ und $\cos(x)=1$ für kleine x) verwendet wurden.

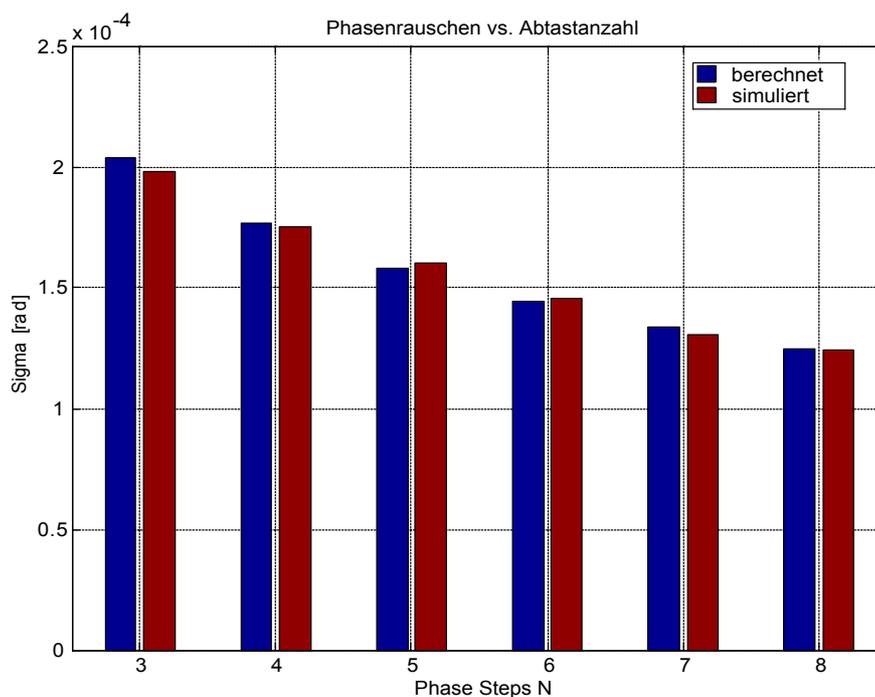


Abb. 4.1: Gegenüberstellung der berechneten und simulativ ermittelten Phasenunsicherheit für verschiedene Abtastzahlen N (Phase Steps).

³ H. Heinol, „Entwicklung und Untersuchung von modulationslaufzeitbasierten 3D-Sichtsystemen“, Dissertation, Institut für Nachrichtenverarbeitung, Universität Siegen, Siegen 2001, S. 108

Aufgrund des allgemeinen Charakters lässt sich die hergeleitete Beziehung natürlich auch auf Photodetektoren anwenden, wobei sich hier die Signalamplitude A aus spektraler Empfindlichkeit R_s , Integrationszeit t_i , Integrationskapazität C_i , mittlerer optischer Empfangsleistung P_{opt} sowie dem Modulationskontrast von Sender K_S und Empfänger K_E ergibt:

$$A = \frac{R_s \cdot P_{opt} \cdot t_i}{C_i} \cdot K_S \cdot K_E. \quad \text{Gl. 4. 34}$$

Damit lautet die Standardabweichung der Phasenbestimmung mittels Phase-Shift-Verfahren für den optischen Empfang und beliebige Rauschamplituden:

$$\sigma_\varphi = \frac{1}{K_S \cdot K_E} \cdot \frac{C_i}{R_s \cdot P_{opt} \cdot t_i} \cdot \sqrt{\frac{2}{N}} \cdot \sigma_Y. \quad \text{Gl. 4. 35}$$

Neben dem Rauschen des eigentlichen Empfängers können hierbei auch systematische Rauschquellen wie z.B. das Quantisierungsrauschen der Abtastung oder eine verrauschte Versorgungsspannung berücksichtigt werden. Die physikalische Grenze ergibt sich jedoch allein durch das Photonenrauschen und lautet:

$$\sigma_{\varphi \min} = \frac{1}{K_S \cdot K_E} \cdot \sqrt{\frac{2 \cdot e}{N \cdot R_s \cdot P_{opt} \cdot t_i}}. \quad \text{Gl. 4. 36}$$

4.3 Auswahl geeigneter Phase-Shift-Algorithmen

4.3.1 Korrektur eines systematischen Phasenoffsets durch die Wahl geeigneter Gewichte

Durch die grundsätzliche Möglichkeit auch einen phasenverschobenen Sinus als Zielfunktion einzusetzen liegt die Vermutung nahe, dass durch eine geeignete Wahl von Gewichten ein bekannter systembedingter Phasenoffset direkt durch den Algorithmus

eliminiert werden kann. Mathematisch ist der Ansatz zur Approximation nach der Methode der kleinsten Quadrate um einen solchen Phasenoffset $\varphi_{\text{Syst.}}$ zu erweitern:

$$E^2 = \sum_{n=1}^N \left[Y_n - \frac{a_0}{2} - A \cdot \sin(\varphi_S \cdot n + \varphi_{\text{Syst.}} + \varphi) \right]^2. \quad \text{Gl. 4. 37}$$

Entsprechend der im **Anhang A.2** dargestellten Herleitung ergibt sich folgender Phase-Shift-Algorithmus:

$$\tan \varphi = \frac{\sum_{n=1}^N Y_n \cdot \cos(\varphi_S \cdot n + \varphi_{\text{Syst.}})}{\sum_{n=1}^N Y_n \cdot \sin(\varphi_S \cdot n + \varphi_{\text{Syst.}})}. \quad \text{Gl. 4. 38}$$

Das Ergebnis dieser Herleitung zeigt deutlich, dass ein systembedingter Phasen-Offset durch eine Variation der Gewichte eliminiert werden kann. Die Möglichkeit einer solchen Korrektur wurde simulativ mit Matlab nachgewiesen. In diesem Beispiel wurden die Gewichte G_n^Z für den Zähler und G_n^N für den Nenner eines 4-Phasen-Algorithmus für einen Phasenoffset $\varphi_{\text{Syst.}}$ von 10° entworfen:

$$G_n^Z = \cos(\varphi_S \cdot n + \varphi_{\text{Syst.}}) = \cos(90^\circ \cdot n + 10^\circ), \quad \text{Gl. 4. 39}$$

$$G_n^N = \sin(\varphi_S \cdot n + \varphi_{\text{Syst.}}) = \sin(90^\circ \cdot n + 10^\circ) \quad \text{Gl. 4.33}$$

Der entsprechende Phase-Shift-Algorithmus lautet für dieses Beispiel:

$$\tan \varphi = \frac{\sum_{n=1}^N Y_n \cdot G_n^Z}{\sum_{n=1}^N Y_n \cdot G_n^N} = \frac{\sum_{n=1}^N Y_n \cdot \cos(90^\circ \cdot n + 10^\circ)}{\sum_{n=1}^N Y_n \cdot \sin(90 \cdot n + 10^\circ)}, \quad \text{Gl. 4. 40}$$

$$\tan \varphi = \frac{Y_1 \cdot (-0,173) + Y_2 \cdot (-0,985) + Y_3 \cdot (0,173) + Y_4 \cdot (0,985)}{Y_1 \cdot (0,985) + Y_2 \cdot (0,173) + Y_3 \cdot (-0,985) + Y_4 \cdot (-0,173)}. \quad \text{Gl. 4. 41}$$

Das Ergebnis der Simulation zeigt deutlich die erwünschte Verschiebung von 10° (Abb. 4.2) und bestätigt die Möglichkeit einer Phasenoffsetkorrektur durch den Algorithmus selbst.

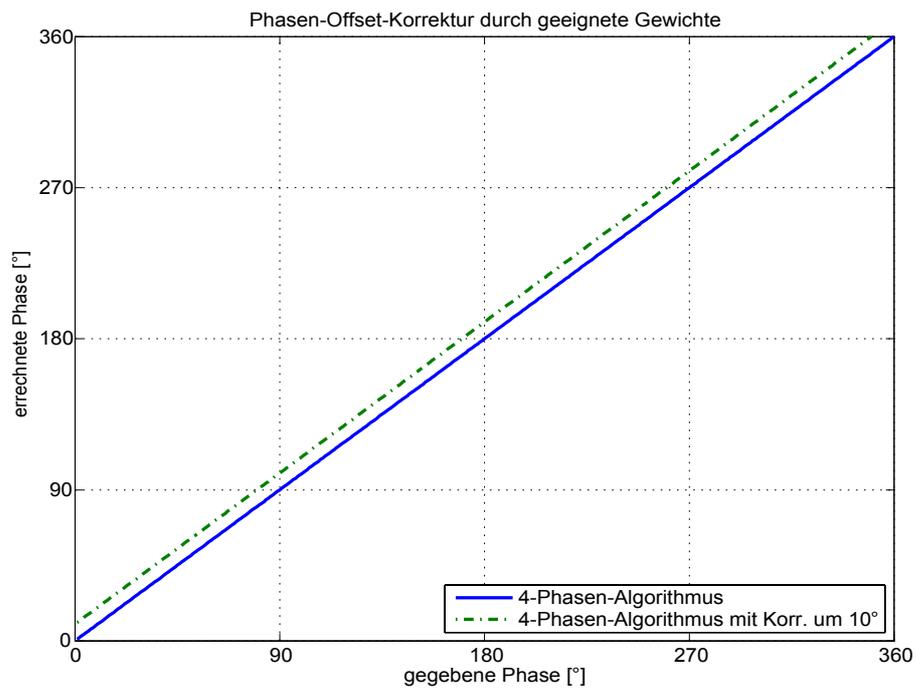


Abb. 4.2: Korrektur eines systembedingten Phasenoffsets durch eine geeignete Wahl der Gewichte im Phase-Shift-Algorithmus

4.3.2 Minimierung des Phasenfehlers für bekannte Oberwellen

Die Abweichung des approximierten vom tatsächlichen Phasenwert findet seine Ursache in der Differenz zwischen den Werten der Zielfunktion und denen der Messung. Für eine solche Abweichung kann es viele Ursachen geben, welche sich jedoch grundsätzlich in zwei Gruppen einteilen lassen. Entgegen den statistischen Prozessen führen alle Nichtlinearitäten in der Signalkette zu einer Verformung der übertragenen Funktion. Da beim Phase-Shift-Verfahren eine Sinusfunktion im Mittelpunkt der Betrachtung liegt, führt

eine Nichtlinearität in der Signalkette zu einer Überlagerung mit entsprechenden Oberwellen. Aus Gründen einer einfacheren technischen Realisierung wird sogar in vielen Systemen auf eine Rechteckmodulation zurückgegriffen, was sich ebenfalls durch Oberwellen beschreiben lässt. Eine solche beliebige periodische Funktion lässt sich mit ihren Koeffizienten A_n beschreiben mit:

$$f(t) = A_1 \cdot \sin(\omega_0 \cdot t + \varphi) + \sum_{k=2}^K A_k \cdot \sin(k \cdot \omega_0 \cdot t + k \cdot \varphi). \quad \text{Gl. 4. 42}$$

Die Grundwelle mit der Amplitude A_1 entspricht hierbei der Zielfunktion, so dass alle anderen Summanden die Abweichung Δy darstellen:

$$\Delta y = \sum_{k=2}^K A_k \cdot \sin(k \cdot \omega_0 \cdot t + k \cdot \varphi). \quad \text{Gl. 4. 43}$$

Die Abweichung des Messwertes Y_n von der Zielfunktion lautet damit für den n -ten Wert eines N -Phasen-Algorithmus:

$$\Delta y_n = \sum_{k=2}^K A_k \cdot \sin\left(k \cdot \frac{2\pi}{N} \cdot n + k \cdot \varphi\right). \quad \text{Gl. 4. 44}$$

Berücksichtigt man diesen Ausdruck in der Phasenbestimmung entsprechend **Gl. 4. 19**, dann lautet der entsprechende Phase-Shift-Algorithmus:

$$\tan(\varphi + \Delta\varphi) = \frac{\sum_{n=1}^N \left[A_1 \cdot \sin\left(\frac{2\pi}{N} n + \varphi\right) + \sum_{k=2}^K A_k \cdot \sin\left(k \frac{2\pi}{N} n + k \cdot \varphi\right) \right] \cdot \cos(\varphi_S \cdot n)}{\sum_{n=1}^N \left[A_1 \cdot \sin\left(\frac{2\pi}{N} n + \varphi\right) + \sum_{k=2}^K A_k \cdot \sin\left(k \frac{2\pi}{N} n + k \cdot \varphi\right) \right] \cdot \sin(\varphi_S \cdot n)}. \quad \text{Gl. 4. 45}$$

Die Abweichungen Δy_n unterliegen durch den Algorithmus ebenso den Gewichten $\cos(\varphi_S \cdot n)$ bzw. $\sin(\varphi_S \cdot n)$, so dass sich der Fehler in Zähler und Nenner berechnet zu:

$$\Delta Z = \sum_{n=1}^N \left[\cos(\varphi_S \cdot n) \cdot \sum_{k=2}^K A_k \cdot \sin \left(k \cdot \frac{2\pi}{N} \cdot n + k \cdot \varphi \right) \right] \quad \text{Gl. 4. 46}$$

bzw.

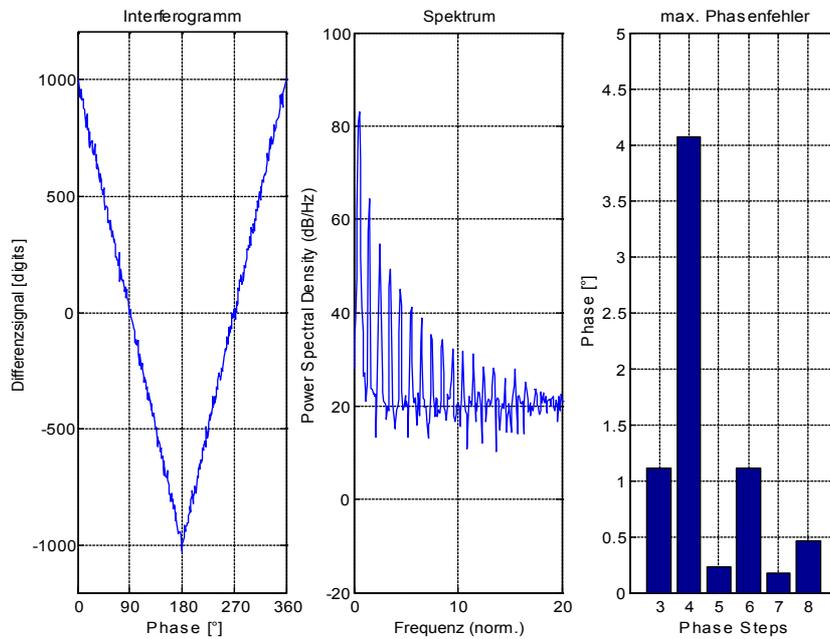
$$\Delta N = \sum_{n=1}^N \left[\sin(\varphi_S \cdot n) \cdot \sum_{k=2}^K A_k \cdot \sin \left(k \cdot \frac{2\pi}{N} \cdot n + k \cdot \varphi \right) \right]. \quad \text{Gl. 4. 47}$$

Will man den Phasenfehler analytisch bestimmen, steht man dem Problem gegenüber, dass neben der Phase auch die Amplituden aller Oberwellen bekannt sein müssen. Die gewonnenen Beziehungen sind jedoch nicht nutzlos, sondern verdeutlichen den grundsätzlichen Zusammenhang zwischen Oberwellen und Phasenfehler. So kann die Multiplikation der abgetasteten Werte Y_n mit den Gewichten $G_n^Z = \cos(\varphi_S \cdot n)$ bzw. $G_n^N = \sin(\varphi_S \cdot n)$ als ein Mischprozess betrachtet werden, dessen Ergebnis anschließend abgetastet wird. Hierin wird ersichtlich, dass alle spektralen Anteile des Mischergebnisses zum Phasenfehler beitragen werden. Aber auch die Abtastung trägt zur Ausprägung eines Phasenfehlers bei. Wesentlich ist hierbei die Einhaltung des Nyquistkriteriums. Beinhalten die Abtastwerte Y_n nur Anteile 2. Ordnung, dann genügt ein 4-Phasen-Algorithmus gerade dem Nyquist-Kriterium. Oberwellen höherer Ordnung erfüllen jedoch dieses Kriterium nicht mehr und unterliegen damit einer Unterabtastung.

Für die Wahl einer geeigneten Anzahl an Abtastungen N ist das Signalspektrum sehr entscheidend. Es erscheint daher sinnvoll, das abzutastende Signal zuvor spektral zu untersuchen. Mit den hier gewonnen Amplituden A_n der einzelnen Frequenzanteile lässt sich nun nach **Gl. 4. 24** der Phasenfehler bestimmen. Vergleicht man die Ergebnisse der in den **Abb. 4.3** dargestellten Beispiele, dann wird der Einfluss der Oberwellen auf den resultierenden Phasenfehler besonders deutlich. Grundsätzlich wird wie erwartet der maximale Phasenfehler mit zunehmender Anzahl an Abtastungen geringer, da immer weniger Oberwellen einer Unterabtastung unterliegen. Bei genauerer Betrachtung wird jedoch deutlich, dass der durch Oberwellen ungeradzahligter Ordnung verursachte Phasenfehler durch eine ebenfalls ungeradzahlige Abtastung geringer ausfällt (**Abb. 4.3 Beispiel b**). Grundsätzlich lässt sich sagen, dass geradzahlige Phase-Shift-Algorithmen geeigneter erscheinen, wenn die Oberwellen geradzahligter Ordnung überwiegen.

Ungeradzahlige Algorithmen sind dagegen für Signale mit überwiegend ungeradzahligen Oberwellen geeignet (**Abb. 4.3 Beispiel a**).

Beispiel a)



Beispiel b)

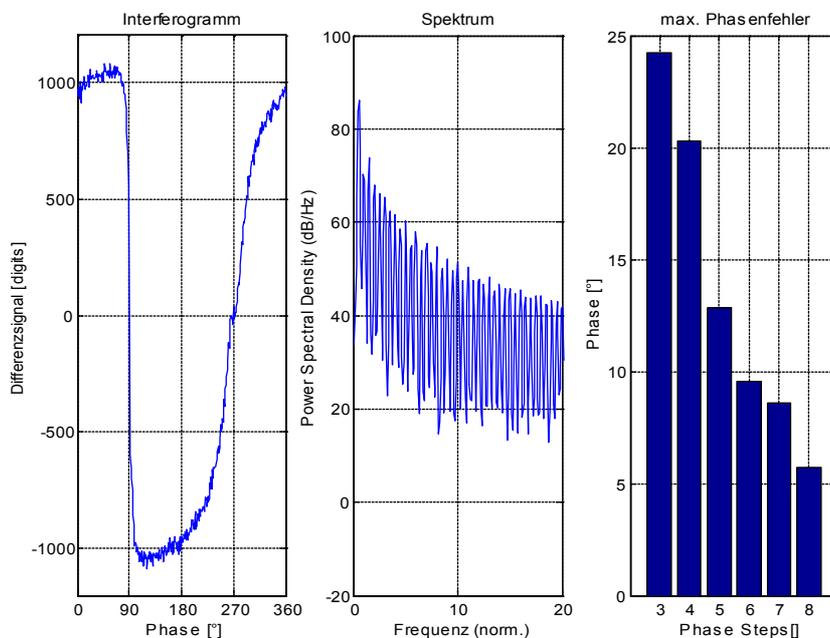


Abb. 4.3: Beispiele verschiedener Interferogramme, ihrer spektralen Anteile und der davon abhängige maximale Phasenfehler in Abhängigkeit der Abtastanzahl N .

4.3.3 Minimierung des Einflusses verrauschter Messwerte

Neben der physikalischen Grenze des Photonengeräusches unterliegen reale optische Systeme vielen weiteren Rauschprozessen. Neben den systembedingten Ursachen wie verrauschte Versorgungsspannungen, AD-Wandler- oder Verstärkerrauschen finden sich allein für ein photosensitives Bauelement schon genug charakteristische Rauschquellen wie z.B. kTC- oder Johnson-Rauschen. Neben verschiedenen schaltungstechnischen Maßnahmen können auch verschiedene Auswertelgorithmen zur Rauschunterdrückung beitragen. Mit Phase-Shift-Verfahren lässt sich das Rauschen allein durch die Anzahl N an Abtastungen reduzieren (Gl. 4. 33), wobei hier natürlich die Messrate durch die zusätzlichen Messungen leidet. Interessanter sind also Lösungsvorschläge, die bei vergleichbarer Messrate das Rauschen der berechneten Phase verbessern. Einziger Weg bleibt, hier das Rauschen der in den Phase-Shift-Algorithmus eingebrachten Messwerte zu verringern, was sich nur durch eine mehrfache Abtastung der Messwerte realisieren lässt (Abb. 4.4).

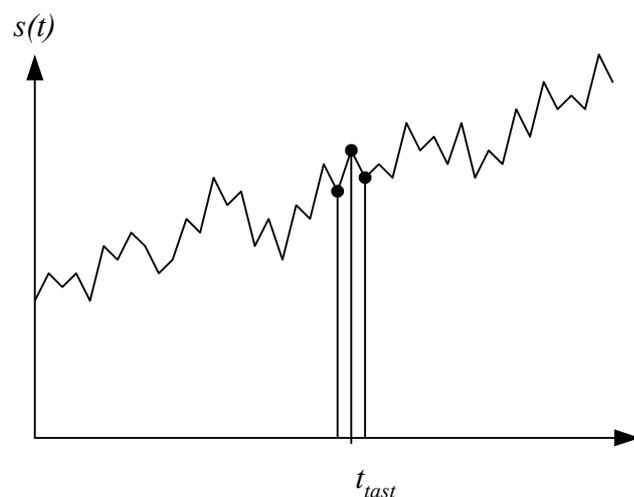


Abb. 4.4: Prinzip der einfachen Mehrfachabtastung

Eine solche Mehrfachabtastung besitzt den Vorteil der einfachen Berechnung, da hier nur die Summe der Einzelabtastungen gebildet und durch die Abtastanzahl geteilt werden muss. Im Falle der Phase-Shift-Verfahren nach Gl. 4. 17 besteht nun ein einzelner in den Algorithmus eingebrachter Wert Y_n aus mehreren Einzelmesswerten y_k :

$$Y_n = \frac{\sum_{i=1}^k y_k}{k} . \quad \text{Gl. 4. 48}$$

Ist die Anzahl k an Mehrfachabtastungen für alle Messwerte Y_n konstant, dann kann sogar auf die Division verzichtet werden. Das Rauschen σ_{Y_n} des Einzelwertes Y_n reduziert sich durch die Mehrfachabtastung mit deren Anzahl k :

$$\sigma_{Y_n} = \sigma_{y_k} \cdot \frac{1}{\sqrt{k}} . \quad \text{Gl. 4. 49}$$

Diese Art der Rauschunterdrückung besitzt neben dem Vorteil der sehr einfachen Berechnung den Nachteil, dass nur unkorreliertes Rauschen unterdrückt wird. Ein korrelierter Anteil wie das kTC-Rauschen bleibt voll erhalten, da er als Offset alle Abtastwerte y_k gleichermaßen betrifft (**Abb. 4.5**). Die Integration des Photostromes auf der Auslesekapazität kann aber in vielen Fällen auch komplett abgetastet und aufsummiert werden (**Abb. 4.6**). Dieses Aufsummieren aller abgetasteten Werte kommt nun einer doppelten Integration gleich. Hierbei ist jedoch darauf zu achten, dass nur der lineare Bereich genutzt wird. Für diesen Fall gilt:

$$Y_n = \Delta t \cdot y_1 + \Delta t \cdot y_2 + \dots + \Delta t \cdot y_k = \Delta t \cdot \sum_{i=1}^k y_k \quad \text{Gl. 4. 50}$$

$$= \int_0^{t_i} at^2 = \frac{a}{2} t^2 .$$

Ist für alle Werte Y_n die Integrationszeit t_i gleich, dann hängen diese nur noch linear vom Anstieg a der zugehörigen Geradengleichung des linearen Bereichs ab. Die aufsummierten bzw. integrierten Werte y_k können deshalb direkt in den Phasenalgorithmus (**Gl. 4. 17**) eingebracht werden.

Eine weitere Möglichkeit der Rauschunterdrückung durch Mehrfachabtastung findet sich in der Approximation einer Geradengleichung in die Integrationskurve des Photostroms (**Abb. 4.7**).

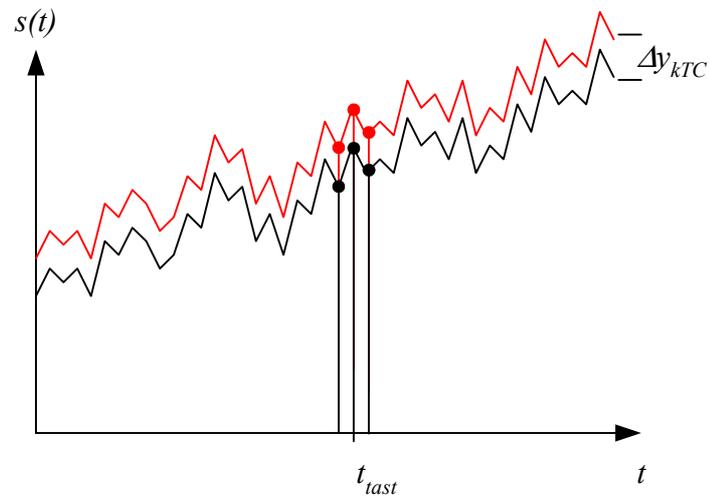


Abb. 4.5: Einfluss des korrelierten Rauschens (wie z.B. kTC-Rauschen) auf die Mehrfachabtastung. Die Änderung im Offset bleibt im Ergebnis der Mehrfachabtastung erhalten.

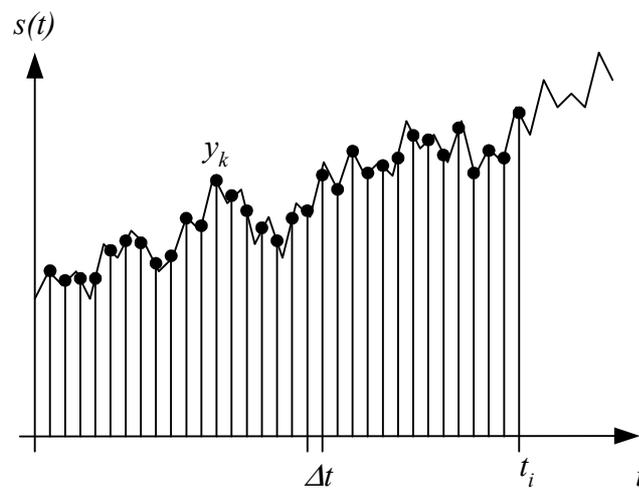


Abb. 4.6: Prinzip der Summation bzw. Integration der Mehrfachabtastung

Hierzu kann ein Least-Squares-Ansatz gewählt werden, wobei hier die gewählte Zielfunktion f der allgemeinen Geradengleichung entspricht:

$$f(t) = a + b \cdot t.$$

Gl. 4. 51

Entsprechend der Herleitung im **Anhang A.3** wird der Anstieg b berechnet aus den Abtastungen y_k , deren Anzahl n und der zugehörigen Abtastzeitpunkte t_k :

$$b = \frac{n \cdot \sum_{k=1}^n y_k \cdot t_k - \sum_{k=1}^n y_k \cdot \sum_{k=1}^n t_k}{n \cdot \sum_{k=1}^n (t_k)^2 - \left(\sum_{k=1}^n t_k \right)^2} . \tag{Gl. 4. 52}$$

Der Vorteil eines solchen Geraden-Fits liegt in der Unterdrückung der korrelierten Rauschteile, welche den Gleichanteil a beeinflussen. Unabhängig von diesem wird mit der Geraden-Approximation der gleiche Anstieg b berechnet (**Abb. 4.8**).

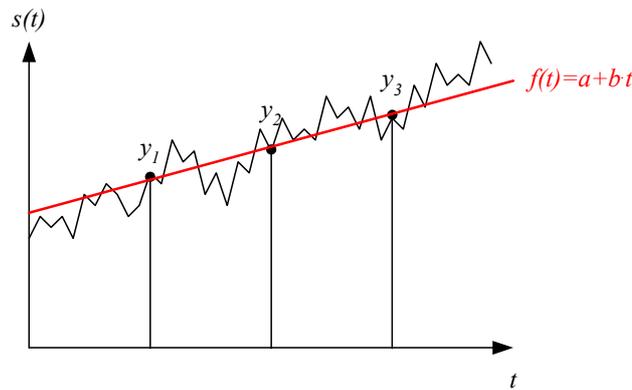


Abb. 4.7: Mehrfachabtastung mit Geraden-Approximation

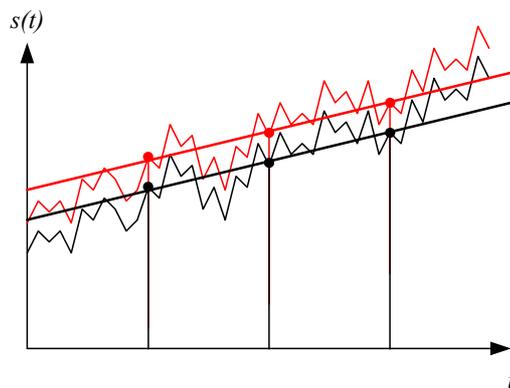


Abb. 4.8: Einfluss des korrelierten Rauschens (wie z.B. kTC-Rauschen) auf die Mehrfachabtastung mit Geraden-Approximation. Der Anstieg der approximierten Gerade ist vom Offset unbeeinflusst.

Der Nachteil dieser Art der Rauschunterdrückung ist in dem erhöhten Rechenaufwand zu finden, der gerade bei der Realisierung von Low-Cost-Geräten die Rechenleistung von preiswerten Mikroprozessoren übersteigen kann.

4.4 Mehrdimensionale Phase-Shift-Interferometrie

4.4.1 Laterale Abtastung zweidimensionaler Bildaufnehmer

Grundlage einer vollständigen systemtheoretischen Beschreibung einer Signalübertragung ist die Übertragungsfunktion eines jeden Teilsystems. Eine mehrdimensionale Anordnung von PMD-Pixeln und der zugehörigen Empfangsoptik lässt sich jedoch nicht mehr allein durch eine zeitliche Übertragungsfunktion beschreiben. Hierzu ist weiterhin räumliche Übertragungsfunktion notwendig. In Analogie zur Systembeschreibung im Zeit- bzw. Frequenzbereich spricht man hierbei von Orts- und Ortsfrequenzbereich. Im Folgenden soll nun die mehrdimensionale Übertragungsfunktion einer PMD-Matrix betrachtet werden, da sich mit ihrer Hilfe viele Bildverarbeitungsschritte im Ortsfrequenzbereich schnell umsetzen lassen.

Die gebräuchlichsten Bildaufnehmer bestehen aus einer zweidimensionalen Anordnung mehrerer Einzelpixel, auf die der Intensitätsverlauf einer gegebenen Szenerie über eine entsprechende Optik abgebildet wird (**Abb. 4.9 a**). Betrachtet man ein einzelnes Pixel, welches meist eine quadratische bzw. rechteckige Geometrie besitzt, dann erfasst es nur einen Teilbereich der Bildszene. Der in diesem Bereich vorhandene Intensitätsverlauf wird somit mathematisch mit dem Faktor 1 gewichtet, wogegen alle Intensitäten außerhalb mit null multipliziert werden (**Abb. 9 b**). Die räumliche Übertragungsfunktion eines einzelnen Pixels lässt sich somit als eine zweidimensionale Rechteckfunktion beschreiben:

$$h(x, y) = \text{rect}\left[\frac{x}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right]. \quad \text{Gl. 4. 53}$$

Die Übertragungsfunktion eines quadratischen Pixels zeigt **Abb. 4.10**, wogegen die Übertragungsfunktion eines rechteckigen Pixels in **Abb. 4.11** dargestellt ist.

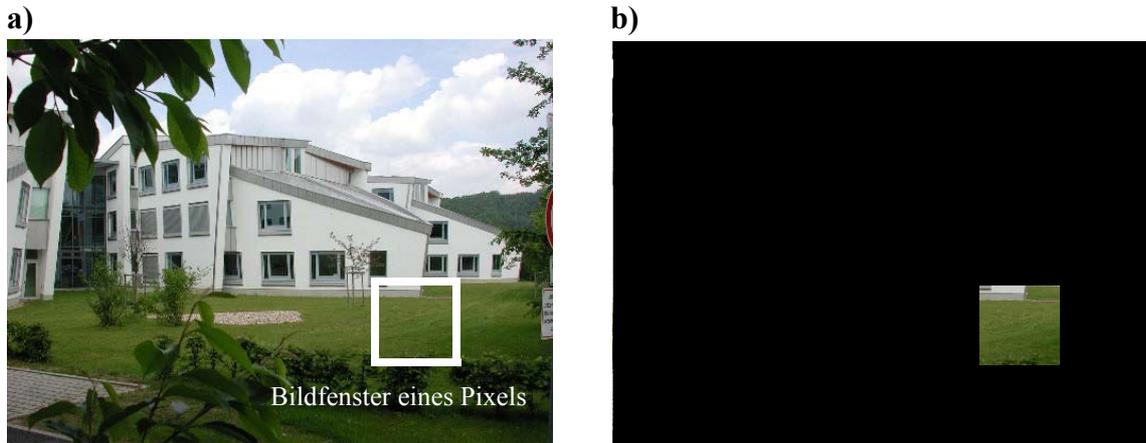


Abb. 4.9: Abbildung einer Bildszene auf einen Bildaufnehmer a) gegebenes Intensitätsbild b) Abbildung auf ein quadratisches Einzelpixel

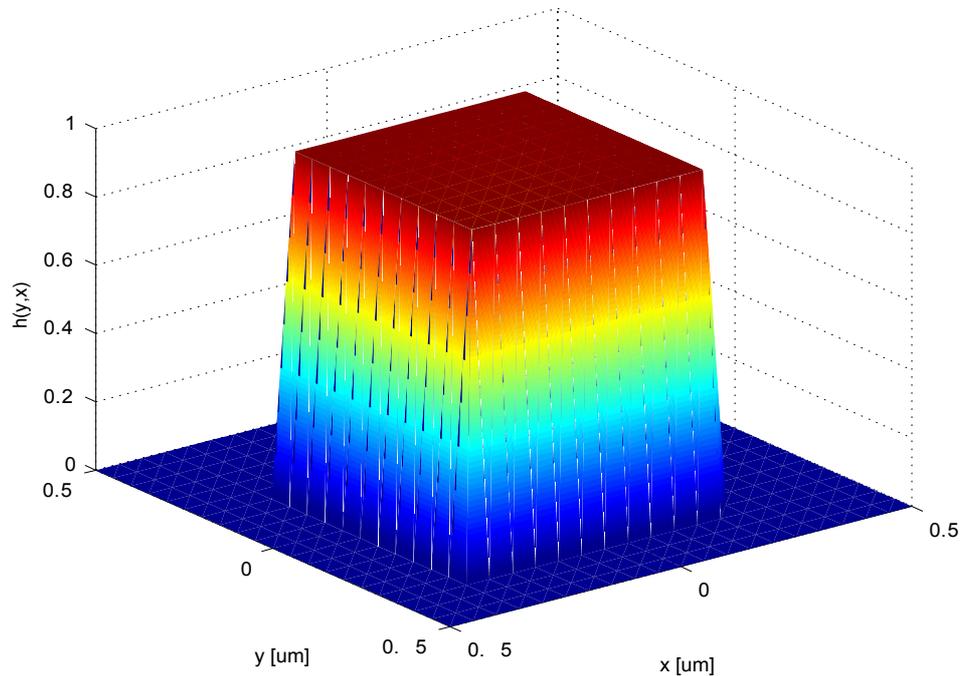


Abb. 4.10: Zweidimensionale Übertragungsfunktion $h(x,y)$ eines einzelnen quadratischen Bildaufnehmers

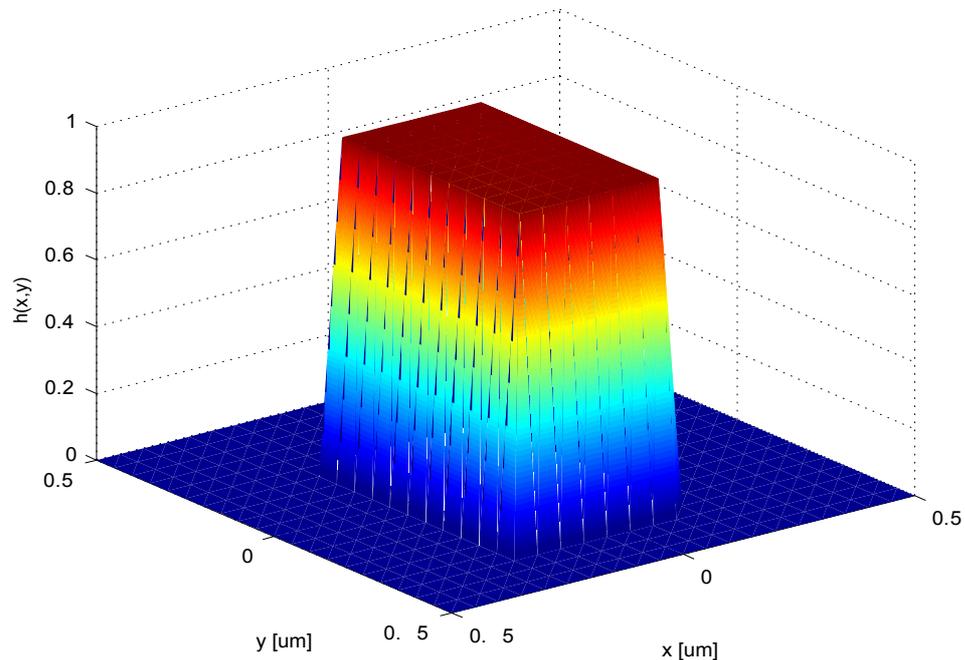


Abb. 4.11: Zweidimensionale Übertragungsfunktion $h(x,y)$ eines einzelnen rechteckigen Bildaufnehmers

Die Übertragung eines Signals kann in der klassischen Systemtheorie als ein einfaches algebraisches Produkt beschrieben werden, wenn das Signal $s(t)$ und die Übertragungsfunktion $h(t)$ zuvor der Fourier-Transformation unterzogen werden. Diese Vereinfachung kann ebenso in einem mehrdimensionalen System genutzt werden, wobei hier jedoch die mehrdimensionale Fourier-Transformation anzuwenden ist. Diese lässt sich in der vektoriellen Form beschreiben durch:

$$S(f_x, f_y) = \int_{-\infty}^{\infty} h(\underline{l}) \cdot e^{-j2\pi \underline{f}^T \underline{l}} dt. \quad \text{Gl. 4. 54}$$

Der Vektor \underline{l} beinhaltet hierbei die Variablen eines beliebigen mehrdimensionalen Systems, wobei er in der Bildverarbeitung durch den Ortsvektor $\underline{s} = x + y$ zu ersetzen ist. Für einen rechteckigen Einzelpixel mit der in **Gl. 4. 53** gegebenen Übertragungsfunktion lautet nun die zweidimensionale Fourier-Transformation:

$$H(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{-j2\pi f_y y} dx dy. \quad \text{Gl. 4. 55}$$

Die Lösung des Fourier-Integrals lautet:

$$\begin{aligned} H(f_x, f_y) &= \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot e^{-j2\pi f_x x} dx \cdot \int_{-\infty}^{\infty} \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_y y} dy \\ &= X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \end{aligned} \quad \text{Gl. 4. 56}$$

Die gewonnene Übertragungsfunktion $H(f)$ im Frequenzbereich für ein quadratisches Pixel zeigt **Abb. 4.12**. Deutlich ist hier zu erkennen, wie sich der symmetrische Charakter der Übertragungsfunktion in den Fourier-Bereich überträgt. Da ein rechteckiger Pixel diese Symmetrie nun nicht mehr aufweist, ist seine Übertragungsfunktion $H(f)$ im Frequenzbereich ebenso nicht mehr symmetrisch. Sein Verhalten wird also einen richtungsabhängigen Charakter besitzen, was direkt am Beispiel einer Kantendetektion deutlich wird.

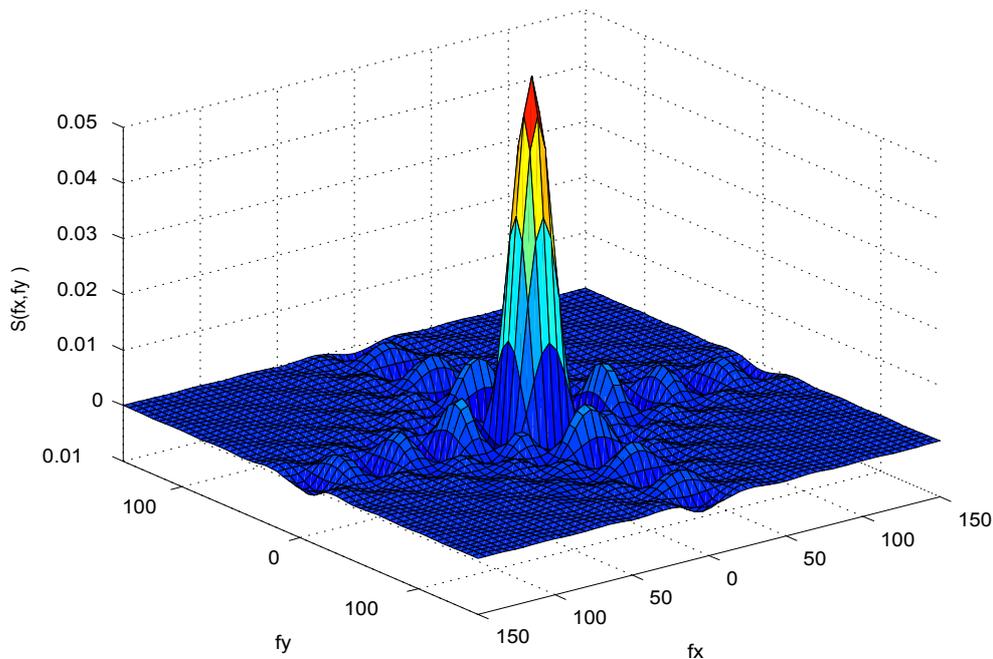


Abb. 4.12: Darstellung der Übertragungsfunktion $H(f_x, f_y)$ eines quadratischen Bildaufnehmers ($250\mu\text{m} \times 250\mu\text{m}$)

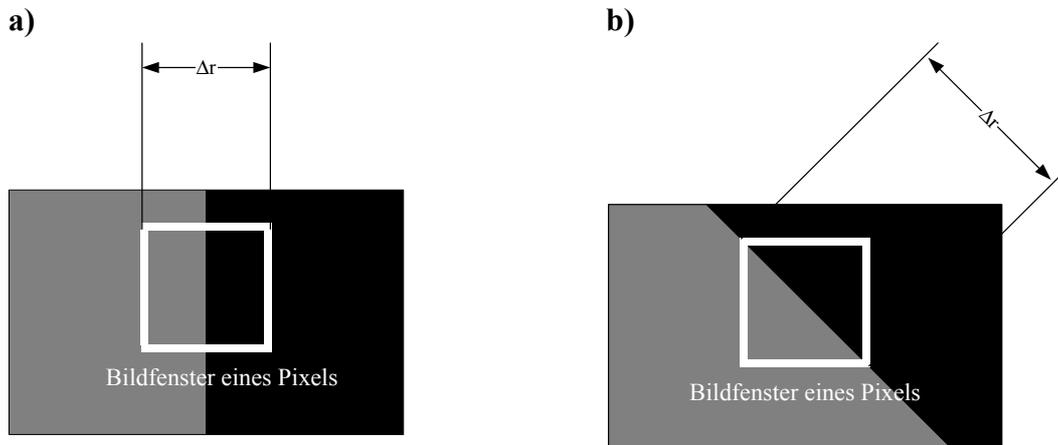


Abb. 4.13: Richtungsabhängige Ortsauflösung eines quadratischen Pixels; a) entlang der Pixelkante x; b) entlang der Diagonalen

Liegt im Intensitätsbild eine Sprungfunktion parallel zu einer Pixelkante vor, dann entspricht der Bereich, der nicht genau einer der beiden Intensitäten I_1 und I_2 zugeordnet werden kann, der Pixelbreite x . Für diesen Fall ist also die Ortsauflösung Δr nur durch die Pixelbreite x gegeben (**Abb. 4.13 a**). Im Gegensatz dazu können diagonale Kanten wesentlich schlechter aufgelöst werden, da hier die Ortsauflösung Δr durch die Pixeldiagonale gegeben ist (**Abb. 4.13 b**). Besitzt nun ein Pixel eine rechteckige Geometrie, dann fehlt in der zweidimensionalen Fourier-Transformation der symmetrische Charakter und sie ist somit in x - und y -Richtung unterschiedlich ausgeprägt (**Abb. 4.14**). Realisiert man eine lineare Anordnung mehrerer Pixel, dann ist dieser Effekt noch deutlicher ausgeprägt, denn diese Pixelanordnung kann näherungsweise auch als ein lang gezogenes Einzelnes betrachten werden:

$$h(x, y) = \text{rect}\left[\frac{x - x_0}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right] + \text{rect}\left[\frac{x + x_0}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right]. \tag{Gl. 4. 57}$$

Die Fourier-Transformierte hierfür lautet damit:

$$S(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y} dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{+j2\pi f_x x_0} \cdot e^{-j2\pi f_y y} dx dy. \tag{Gl. 4. 58}$$

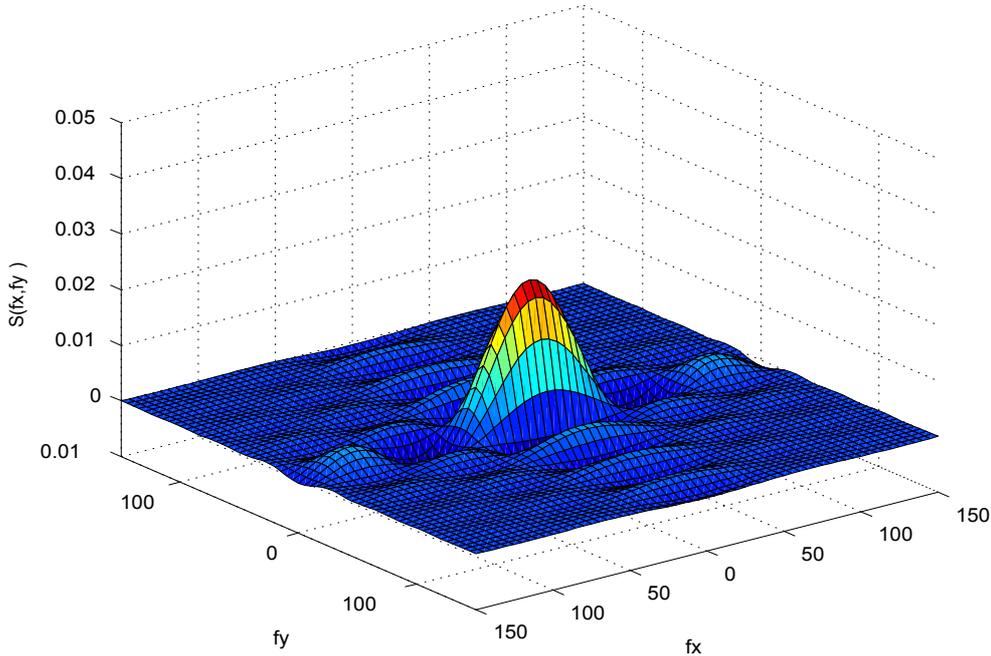


Abb. 4.14: Darstellung der Übertragungsfunktion $H(f_x, f_y)$ eines rechteckigen Bildaufnehmers (100 μm x 250 μm)

Dieser Ausdruck kann folgendermaßen vereinfacht werden:

$$\begin{aligned}
 S(f_x, f_y) &= e^{-j2\pi f_x x_0} \cdot \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot e^{-j2\pi f_x x} dx \cdot \int_{-\infty}^{\infty} \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_y y} dy \\
 &+ e^{j2\pi f_x x_0} \cdot \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot e^{-j2\pi f_x x} dx \cdot \int_{-\infty}^{\infty} \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_y y} dy
 \end{aligned}
 \tag{Gl. 4. 59}$$

bzw.

$$\begin{aligned}
 S(f_x, f_y) &= (e^{j2\pi f_x x_0} + e^{-j2\pi f_x x_0}) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\
 &= 2 \cos(2\pi f_x x_0) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y)
 \end{aligned}
 \tag{Gl. 4. 60}$$

Für eine beliebige geradzahlige Anzahl n linear angeordneter Pixel gilt nun folgender Ausdruck:

$$S(f_x, f_y) = X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \cdot \sum_1^n 2 \cos(2\pi f_x x_0) \quad \text{mit: } \frac{n}{2} \in N. \quad \text{Gl. 4. 61}$$

Eine gute Übersicht der Auswirkung linearer Anordnung verschiedener Pixelanzahlen zeigt **Abb. 4.15**. Da sich mit steigender Pixelanzahl natürlich die insgesamt detektierte Leistung erhöht, zeigen sich in der Darstellung des Fourier-Bereichs alle Frequenzanteile mit ausgeprägteren Amplituden. Da jedoch auch die Anzahl der blinden Bereiche zwischen den Pixeln steigt, werden auch ihre charakteristischen Raumfrequenzen immer deutlicher sichtbar. Dieser Effekt ist noch einmal deutlich zu betonen, denn die räumliche Übertragungsfunktion wird nicht nur durch die Form und Abmessung des einzelnen Pixels bestimmt, sondern ebenso durch ihre Abstände. Für die 3D-Bewegtbildverarbeitung sind jedoch zweidimensionale Pixelanordnungen von größerem Interesse. Für die einfachste aller Matrizenanordnungen, einer 2x2-Matrix lautet die zweidimensionale Übertragungsfunktion:

$$\begin{aligned} h(x, y) = & \text{rect}\left[\frac{x-x_0}{X}\right] \cdot \text{rect}\left[\frac{y+y_0}{Y}\right] \\ & + \text{rect}\left[\frac{x+x_0}{X}\right] \cdot \text{rect}\left[\frac{y+y_0}{Y}\right] \\ & + \text{rect}\left[\frac{x-x_0}{X}\right] \cdot \text{rect}\left[\frac{y-y_0}{Y}\right] \\ & + \text{rect}\left[\frac{x+x_0}{X}\right] \cdot \text{rect}\left[\frac{y-y_0}{Y}\right]. \end{aligned} \quad \text{Gl. 4. 62}$$

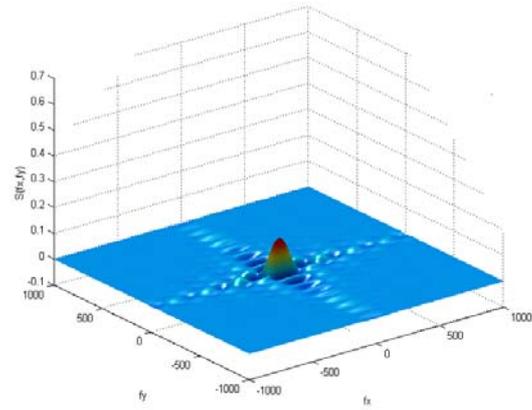
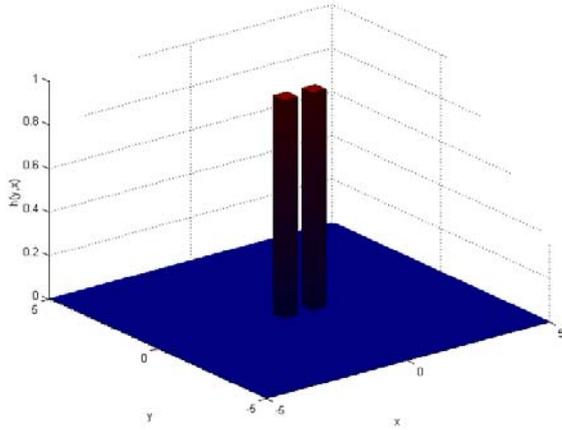
Die Transformation in den Ortsfrequenzbereich lautet dafür:

$$\begin{aligned} S(f_x, f_y) = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y} \cdot e^{-j2\pi f_y y_0} dx dy \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{+j2\pi f_x x_0} \cdot e^{-j2\pi f_y y} \cdot e^{-j2\pi f_y y_0} dx dy \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y} \cdot e^{+j2\pi f_y y_0} dx dy \\ & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{+j2\pi f_x x_0} \cdot e^{-j2\pi f_y y} \cdot e^{+j2\pi f_y y_0} dx dy. \end{aligned} \quad \text{Gl. 4. 63}$$

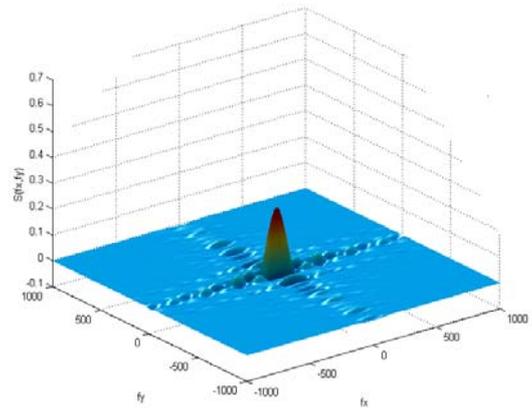
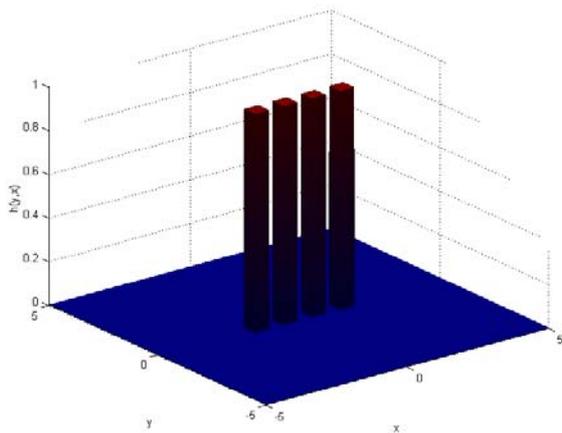
Ortsbereich

Ortsfrequenzbereich

a) 1 x 2



b) 1 x 4



c) 1 x 10

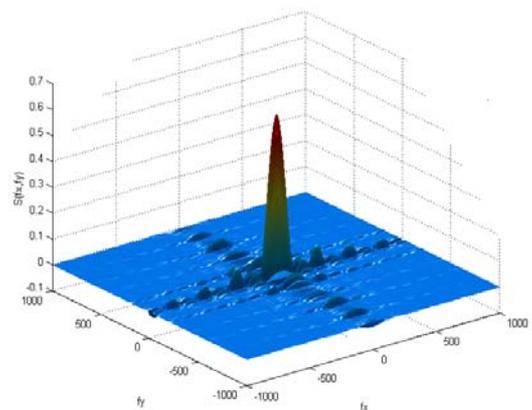
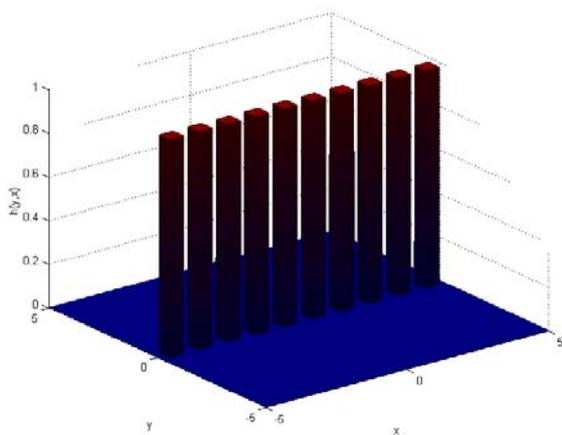


Abb. 4.15: Die Übertragungsfunktion im Orts- und Ortsfrequenzbereich unterschiedlicher Anzahlen linear angeordneter Pixel

was sich entsprechend der Herleitung im **Anhang A.4** zu folgendem Ausdruck vereinfachen lässt:

$$S(f_x, f_y) = 2 \cos(2\pi(f_x x_0 + f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\ + 2 \cos(2\pi(f_x x_0 - f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y). \quad \text{Gl. 4. 64}$$

In Anlehnung an diese Herleitung lässt sich die Fourier-Transformierte der Übertragungsfunktion für jede zweidimensionale Pixelanordnung berechnen. So lautet diese beispielsweise für eine 4x4-Matrix:

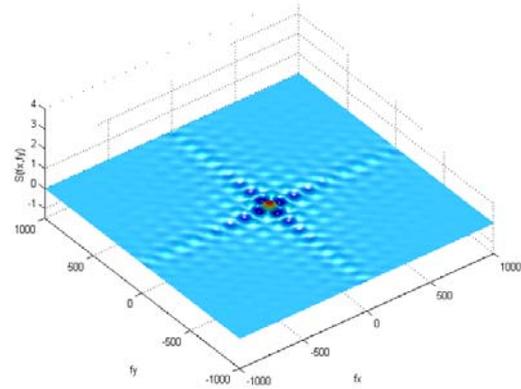
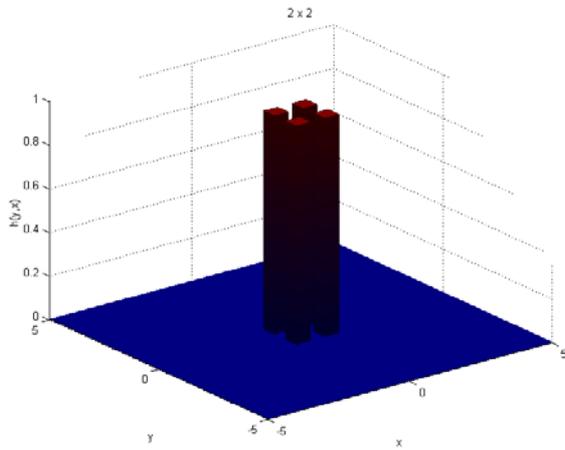
$$S(f_x, f_y) = 2 \cos(2\pi(f_x x_0 + f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\ + 2 \cos(4\pi(f_x x_0 + f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\ + 2 \cos(2\pi(f_x x_0 - f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\ + 2 \cos(4\pi(f_x x_0 - f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\ + 2 \cos(2\pi(f_x x_0 + 2f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\ + 2 \cos(2\pi(f_x x_0 - 2f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\ + 2 \cos(2\pi(2f_x x_0 + f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\ + 2 \cos(2\pi(2f_x x_0 - f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y). \quad \text{Gl. 4. 65}$$

Vergleicht man nun die verschiedenen Fourier-Transformierten für zweidimensionale Pixelanordnungen (**Abb. 4.16**), dann steigt mit zunehmender Pixelanzahl der Maximalwert an. Dies kann wieder einfach durch die mit jedem weiteren Pixel zusätzlich detektierte Signalleistung erklärt werden. Würde man ein Einzelpixel, das in der Größe vergleichbar zu einer Matrix ist, dieser gegenüberstellen, wäre dieser Leistungszusammenhang vergleichbar oder zumindest ähnlich. Jedoch fällt mit zunehmender Pixelanzahl wieder ein immer deutlicher werdender Effekt ins Gewicht, die Überlagerung mit den Frequenzanteilen, die durch den Abstand zwischen den Pixeln gegeben sind. Hier entstehen in den Randbereichen der Übertragungsfunktion im Ortsfrequenzbereich zunehmend komplexere Strukturen.

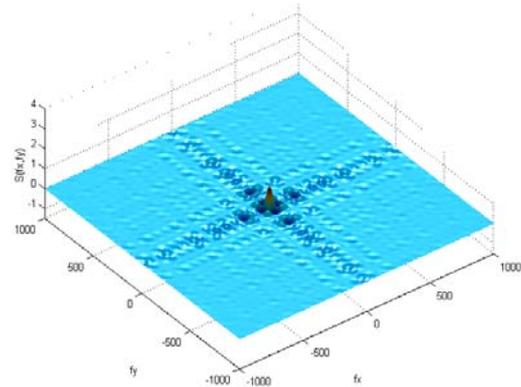
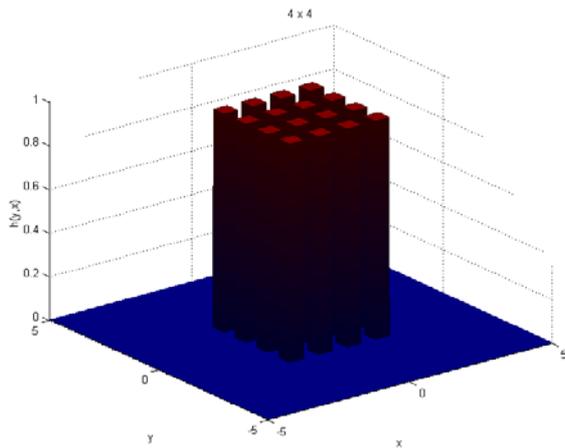
Ortsbereich

Ortsfrequenzbereich

a) 2 x 2



b) 4 x 4



d) 8 x 8

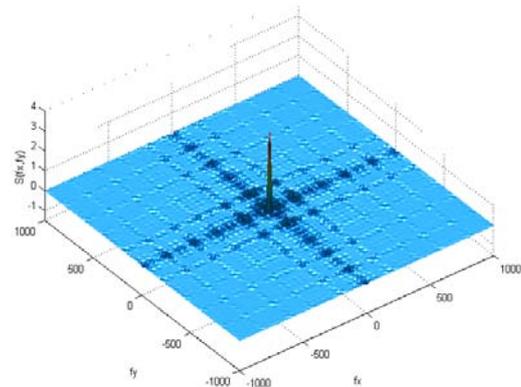
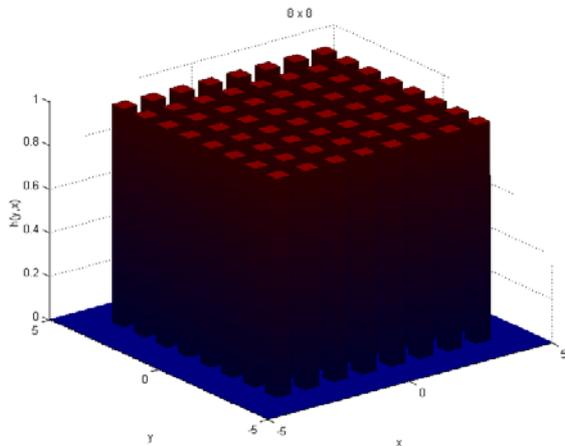


Abb. 4.16: Die Übertragungsfunktion im Orts- und Ortsfrequenzbereich unterschiedlicher Anzahlen zweidimensional angeordneter Pixel

4.4.2 Interferogramme im Ortsfrequenzbereich

Die zweidimensionale Phase-Shift-Interferometrie mittels Matrizenanordnungen basiert genau wie die eindimensionale auf der Auswertung phasenabhängiger Intensitätsbilder. Für einen N -Phasen-Algorithmus ergeben sich somit N verschiedenen Intensitätsbilder und ebenso N verschiedene Fourier-Transformierte:

$$I_n(x, y, z, \varphi + n \cdot \varphi_s) = h(x, y, z) \cdot P(x, y, z, \varphi + n \cdot \varphi_s). \quad \text{Gl. 4. 66}$$

Will man die Fourier-Transformierten aller N Intensitätsbilder bildlich gegenüberstellen, ist zu beachten, dass hier die Darstellung des Betragsspektrums des Ortsfrequenzbereichs nicht ausreicht. Vielmehr ist hier ebenfalls dessen Phasengang zu berücksichtigen. Dieser Zusammenhang kann leicht anhand einer 2×2 -Matrix veranschaulicht werden. Hierbei decken jeweils zwei Pixel Bereiche mit unterschiedlichen Phasenverschiebungen φ ab. Für die Applikation der Entfernungsmessung bedeutet dies, dass zwei Pixel ein Objekt in der Entfernung z_1 und die anderen zwei ein Objekt in der Entfernung z_2 beobachten (**Abb. 4.17**).

Mit der von gesuchter Phase φ und zusätzlichem Phase-Shift φ_s abhängigen Empfangsleistung $P(x, y, \varphi + n \cdot \varphi_s)$ ergibt sich im Fourier-Bereich folgendes Intensitätsbild:

$$I(f_x, f_y, f_z) = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} \text{rect}\left[\frac{x}{X}\right] \cdot \text{rect}\left[\frac{y}{Y}\right] dx dy \cdot e^{-j2\pi f_x x} \cdot e^{-j2\pi f_y y} \cdot \left. \begin{array}{l} P(-x_0, -y_0, \varphi + n \cdot \varphi) \cdot e^{j2\pi f_x x_0} \cdot e^{j2\pi f_y y_0} \\ + P(-x_0, y_0, \varphi + n \cdot \varphi) \cdot e^{j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \\ + P(x_0, -y_0, \varphi + n \cdot \varphi) \cdot e^{-j2\pi f_x x_0} \cdot e^{j2\pi f_y y_0} \\ + P(x_0, y_0, \varphi + n \cdot \varphi) \cdot e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \end{array} \right\} \cdot \quad \text{Gl. 4. 67}$$

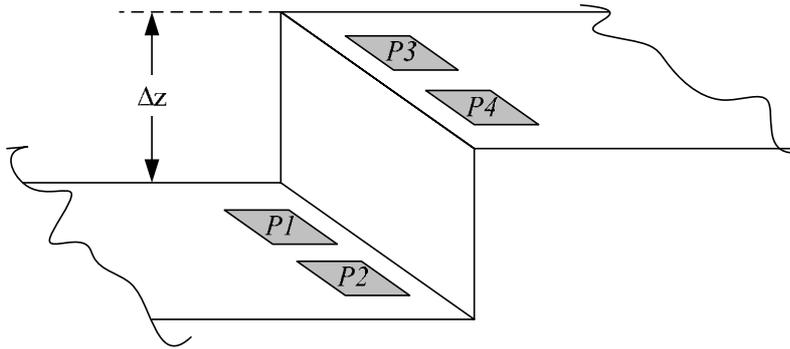


Abb. 4.17: Blick einer 2x2 Matrix auf eine Kante aus unterschiedlichen Entfernungen

Entspricht der räumliche Versatz Δz zwischen beiden Bereichen einem Phasenversatz $\Delta\varphi$ von genau $\pi/2$, dann ergeben sich für das erste Intensitätsbild ($n=1$) eines 4-Phasen-Algorithmus folgende Empfangsleistungen $P(x,y)$ (**Abb. 4.18 a**):

$$P(-x_0, -y_0) = P(-x_0, y_0) = 0, \tag{Gl. 4. 68}$$

$$P(x_0, -y_0) = P(x_0, y_0) = 1. \tag{Gl. 4. 69}$$

Das transformierte Intensitätsbild lautet hierfür:

$$I_1(f_x, f_y, f_z) = X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \cdot \left. \begin{array}{l} 0 \\ + 0 \\ + e^{-j2\pi f_x x_0} \cdot e^{j2\pi f_y y_0} \\ + e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \end{array} \right\}. \tag{Gl. 4. 70}$$

Für $n=2$ empfängt die 2x2-Matrix folgende Leistungen $P(x,y)$ (**Abb. 4.18 b**):

$$P(-x_0, -y_0) = P(-x_0, y_0) = 0.5. \tag{Gl. 4. 71}$$

$$P(x_0, -y_0) = P(x_0, y_0) = 0.5. \tag{Gl. 4. 72}$$

Das transformierte Intensitätsbild lautet nun hierfür:

$$I_2(f_x, f_y, f_z) = X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \cdot \left. \begin{array}{l} 0,5 \cdot e^{j2\pi f_x x_0} \cdot e^{j2\pi f_y y_0} \\ + 0,5 \cdot e^{j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \\ + 0,5 \cdot e^{-j2\pi f_x x_0} \cdot e^{j2\pi f_y y_0} \\ + 0,5 \cdot e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \end{array} \right\} \quad \text{Gl. 4. 73}$$

Analog lauten die noch fehlenden zwei Fourier-Transformierten ($n=3$; $n=4$) mit den entsprechenden Empfangsleistungen $P(x,y)$ (**Abb. 4.18 c, d**):

$$I_3(f_x, f_y, f_z) = X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \cdot \left. \begin{array}{l} e^{j2\pi f_x x_0} \cdot e^{j2\pi f_y y_0} \\ + e^{j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \\ + 0 \\ + 0 \end{array} \right\} \quad \text{Gl. 4. 74}$$

und

$$I_4(f_x, f_y, f_z) = X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \cdot \left. \begin{array}{l} 0,5 \cdot e^{j2\pi f_x x_0} \cdot e^{j2\pi f_y y_0} \\ + 0,5 \cdot e^{j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \\ + 0,5 \cdot e^{-j2\pi f_x x_0} \cdot e^{j2\pi f_y y_0} \\ + 0,5 \cdot e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \end{array} \right\} \quad \text{Gl. 4. 75}$$

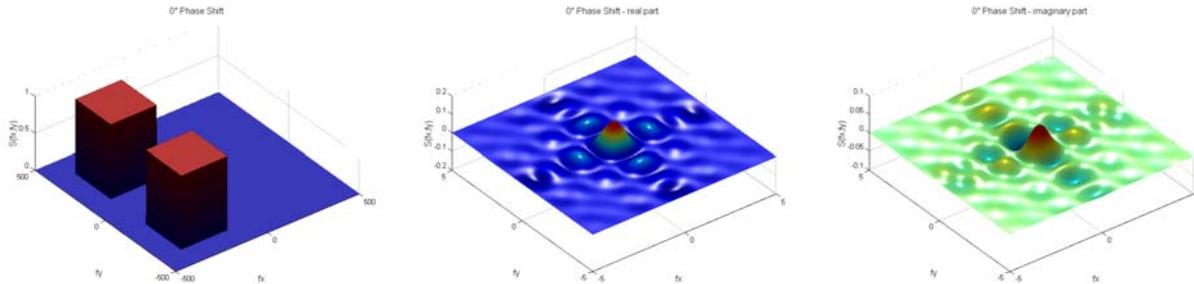
In der Visualisierung der Fourier-Transformierten (**Abb. 4.18**) wird deutlich, dass sich der Realteil durch einen Phase-Shift nicht ändert. Nur im Imaginärteil wird dieser abgebildet. Neben dieser wichtigen Aussage wird jedoch noch eines deutlich, die Entfernungsmessung mittels Phase-Shift-Interferometrie liefert die Entfernungswerte schon im Ortsfrequenzbereich, denn jeder dieser Werte ist charakterisiert durch Frequenz und Phase. Erst durch eine Umrechnung der Phase und der Modulationsfrequenz erhält man auch Entfernungswerte im Ortsbereich. Damit ist für eine vollständige Darstellung eines durch Phase-Shift-Verfahren ermittelten 3D-Bildes im Ortsfrequenzbereich nur jeweils eine Transformation in x- und y- Richtung durchzuführen. Eine vollständige Darstellung im Ortsbereich dagegen erfordert nur eine Umrechnung der Phasenwerte in eine z-Komponente. Von praktischem Interesse ist die mehrdimensionale Fourier-Transformation für eine der Phase-Shift-Algorithmik folgenden 3D-Bild- bzw. 3D-Bewegtbildverarbeitung.

Intensitätsbild im Ortsraum

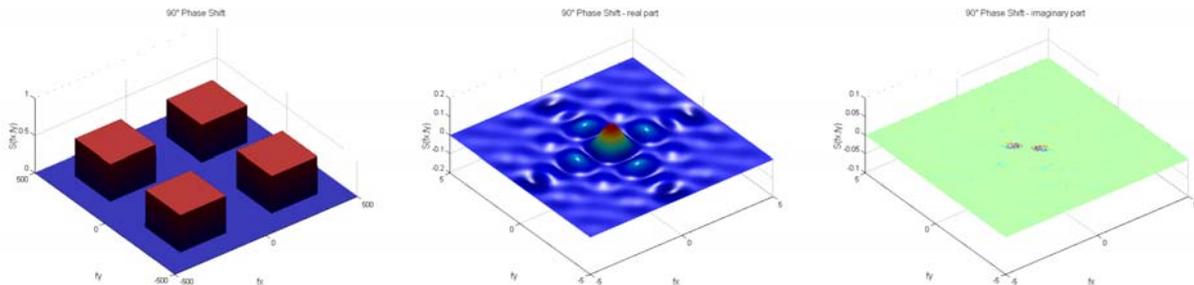
Realteil der Fourier-Transformierten im Ortsfrequenzraum

Imaginärteil der Fourier-Transformierten im Ortsfrequenzraum

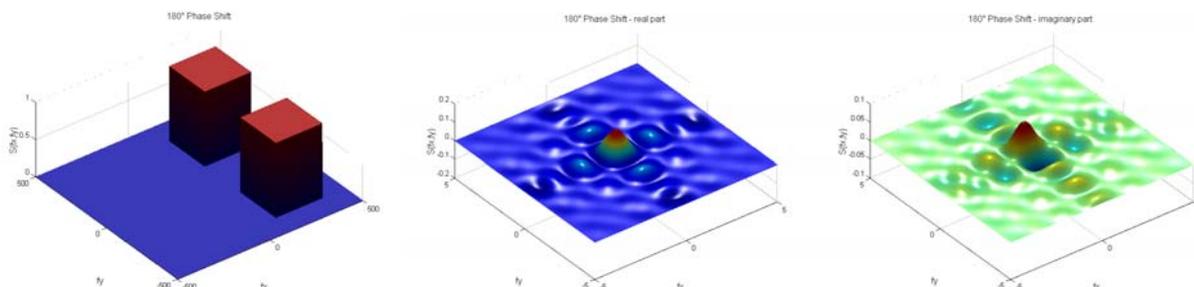
a) $\varphi_S=0^\circ$



b) $\varphi_S=90^\circ$



c) $\varphi_S=180^\circ$



d) $\varphi_S=270^\circ$

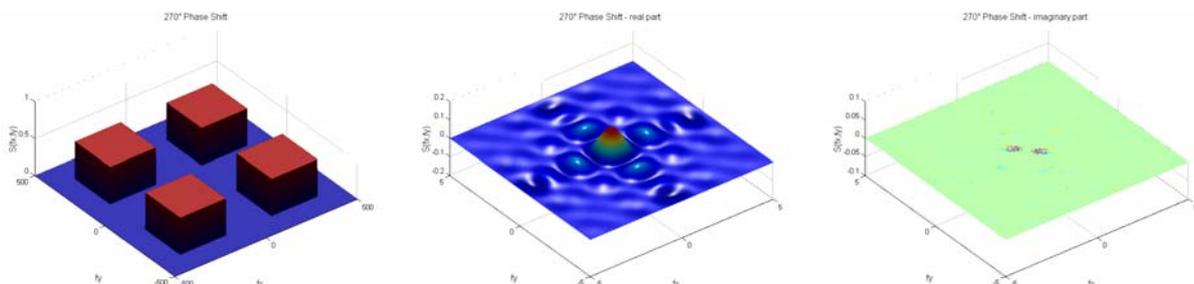


Abb. 4.18: Intensitätsbilder einer 4-Phasen-Messung mit einer 2x2-Matrix auf einer Kante und deren Fourier-Transformierte

Hier vereinfacht sich die Erkennung und Ortsbestimmung eines Objektes durch eine Korrelation im Ortsraum in eine einfache Multiplikation zweier Datensätze im Orstfrequenzraum. Beispielhaft wird die Einfachheit eines solchen Verfahrens noch im **Kap. 4.6.3** dargestellt.

4.5 Probleme optischer Phasennessverfahren

Die enormen Anforderungen an die Phasennessung optischer Signale lassen sich auf die enorme Ausbreitungsgeschwindigkeit des Lichts zurückführen. Im Bereich der Laufzeitentfernungsmessung erfordert eine Entfernungsauflösung Δs von 1mm eine Zeitauflösung Δt von:

$$\Delta t = \frac{\Delta s}{c} = \frac{1 \cdot 10^{-3} \text{ m}}{3 \cdot 10^8 \text{ m/s}} = 3,33 \text{ ps} . \quad \text{Gl. 4. 76}$$

Eine solche hohe Zeitauflösung stellt natürlich eine ebenso hohe Anforderung an das Messverfahren und die verwendete Systemtechnik. Sie bedingt damit natürlich eine enorm hohe Signalstabilität innerhalb des gesamten Systems, angefangen vom Sender über Signalstrecke, Empfänger und Auswertalgorithmik. Hierzu sollen im Folgenden die Laserdiode als das am schwierigsten zu beherrschende Sendemodul und der PMD-Empfänger genauer betrachtet werden.

4.5.1 Modenstabilität in Halbleiterlaserdioden

Aus dem Bereich der optisch kohärenten Messverfahren ist das Phänomen des instabilen Modenzustands von Halbleiterlaserdioden hinreichend bekannt. Für diese Instabilität

werden in der Fachliteratur die unterschiedlichsten Begriffe wie Mode-Competition⁴, Mode-Hopping⁵ oder Mode-Partition⁶ verwendet. Da diese Begriffe sich nicht klar nach Ursache oder Wirkung unterscheiden lassen, werden diese hier bewusst nicht verwendet. Vielmehr werden hier Modeninstabilitäten entsprechend ihrer physikalischen Ursache bezeichnet.

a) Moden- und Intensitätswechsel aufgrund thermischer Effekte

Ein Modenwechsel durch thermische Effekte lässt sich durch die verschiedenen temperaturabhängigen Größen erklären. So ist die durch den Bandabstand W_g gegebene emittierte Wellenlänge λ einer Laserdiode eine temperaturabhängige Größe. Dies resultiert in einer Verschiebung der optischen Gewinnkurve:

$$\left(\frac{d\lambda}{dT}\right)_g = \frac{d\lambda}{dW_g} \cdot \frac{dW_g}{dT} = -\frac{h \cdot c}{W_g^2} \cdot \frac{dW_g}{dT}. \quad \text{Gl. 4. 77}$$

Der Temperaturgang des Bandabstands W_g führt indirekt noch zu einer weiteren Temperaturabhängigkeit, der des Brechungsindex \bar{n} an den teilreflektierenden Austrittsflächen der Laserdiode:

$$\left(\frac{d\lambda}{dT}\right)_{FP} = \frac{2 \cdot L}{m} \cdot \frac{d\bar{n}}{dT} \quad (m \in N). \quad \text{Gl. 4. 78}$$

⁴ H.-G. Unger; Optische Nachrichtentechnik, Teil 2: Komponenten, Systeme, Messtechnik, Hüthig Verlag Heidelberg 1985, S. 347

⁵ W. Harth, H.Grothe; Sende- und Empfangsdioden für die Optische Nachrichtentechnik, B. G. Teubner Stuttgart, S. 103

⁶ M.P. van Exter, M.B. Willemsen, J.P. Woerdman; „Effect of mode-partition noise on intensity squeezing in a two-mode laser”, Journal of Optics B: Quantum and Semiclassical Optics Vol. 1 No. 6 (1999), S. 637-645

Neben der Änderung von Brechungsindex und Bandabstand kommt es mit steigender Temperatur zu einer Ausdehnung des Halbleiterkristalls. Diese führt wiederum zu einem Anwachsen der Resonatorlänge.

$$\left(\frac{d\lambda}{dT}\right)_L = \frac{2 \cdot \bar{n}}{m} \cdot \frac{dL}{dT} \quad (m \in N). \tag{Gl. 4. 79}$$

Diese unterschiedlichen Temperaturkoeffizienten konkurrieren miteinander bei der Ausprägung der unterschiedlichen Moden. Dieser Sachverhalt wird an der graphischen Darstellung des optischen Gewinns und der Fabry-Perot-Resonanzen deutlich (**Abb. 4.19**). Mit steigender Temperatur wachsen Resonatorlänge und Brechzahl. Dadurch verschieben sich die einzelnen Moden zu höheren Wellenlängen. Dieser Verschiebung ist eine gleichgerichtete Bewegung des optischen Gewinns überlagert, die jedoch einem größeren Temperaturkoeffizienten folgt. Damit verschiebt sich das Maximum der Gewinnkurve zu einem benachbarten Mode und entfacht diesen.

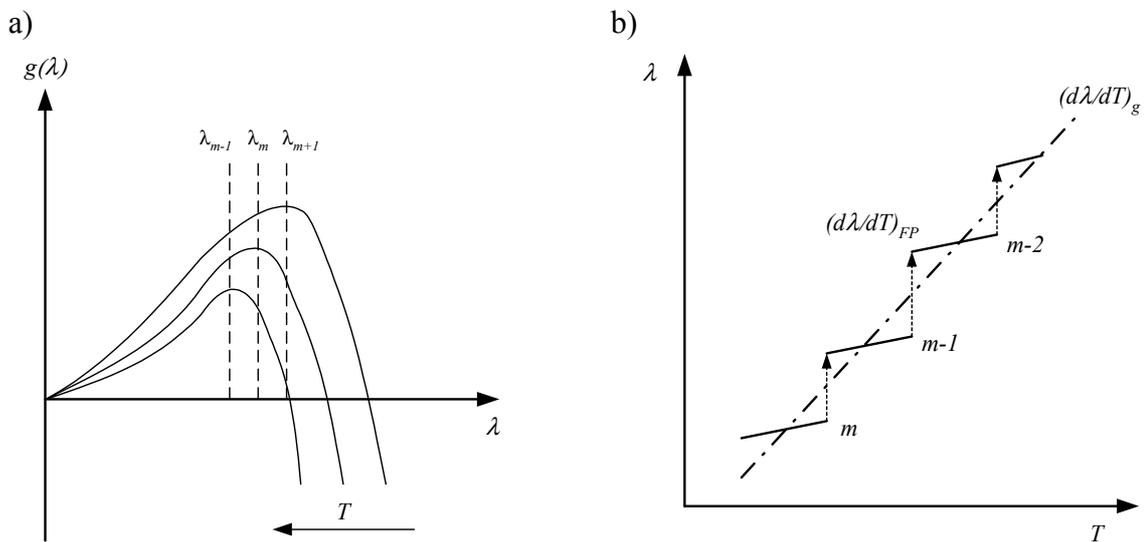


Abb. 4.19: Modensprung eines Halbleiterlasers aufgrund von Temperatur-schwankungen; a) temperaturabhängige Verschiebung der Gewinnkurve $g(\lambda)$ über die Fabry-Perot-Resonanzen λ_n ; b) resultierender Temperaturverlauf der dominanten Emissionswellenlänge λ

Für GaAs-Strukturen ($\lambda=850nm$) betragen beispielsweise die genannten Temperaturkoeffizienten:

$$\left(\frac{d\lambda}{dT}\right)_{FP} = 0,1 \text{ nm} / K \quad \text{Gl. 4. 80}$$

und

$$\left(\frac{d\lambda}{dT}\right)_g = 0,27 \text{ nm} / K . \quad \text{Gl. 4. 81}$$

Mit dem wellenlängenabhängigen Brechungsindex $\bar{n}_e = 4,5$ und einer Resonatorlänge von $L = 400\mu m$ ergibt sich ein Modenabstand von ⁷

$$\begin{aligned} \Delta\lambda_{FP} &= \frac{\lambda^2}{2 \cdot \bar{n}_e \cdot L} \\ &= \frac{(850nm)^2}{2 \cdot 4,5 \cdot 400\mu m} \approx 0,2nm . \end{aligned} \quad \text{Gl. 4. 82}$$

Um einen benachbarten Lasermode durch Erwärmung in das Maximum der optischen Gewinnkurve zu verschieben, ist Temperaturänderung ΔT von

$$\Delta T = \frac{\Delta\lambda_{FP}}{\left(\frac{d\lambda}{dT}\right)_g - \left(\frac{d\lambda}{dT}\right)_{FP}} = \frac{0,2 \text{ nm}}{0,27 \text{ nm} / K - 0,1 \text{ nm} / K} \approx 1,18 \text{ K} \quad \text{Gl. 4. 83}$$

nötig. Da sich ein Mode jedoch nicht zwangsläufig im Maximum befinden muss, können schon wesentlich kleinere Temperaturschwankungen zu einem solchen Modenwechsel führen. In experimentellen Versuchen konnten aber trotz sehr konstanter

⁷ W. Harth, H. Grothe; Sende und Empfangsdioden für die optische Nachrichtentechnik, Teubner 1984, Stuttgart, S. 68

Temperaturbedingungen von kleiner $0.3 \times 10^{-3} \text{ K}$ noch sehr deutliche Modensprünge nachgewiesen werden.⁸ Diese können nicht durch die Temperatur sondern nur über so genannte Modenkopplungseffekte erklärt werden.

b) Moden- und Intensitätswechsel durch Modenkopplungseffekte

Das Problem der instabilen Lasermoden ist aus dem Bereich der kohärenten Messverfahren schon länger bekannt. Anfänglich waren die Leistungsschwankungen eines einzelnen Moden von Interesse, welche 1984 von Henry durch ein statistisches Modell erklärt wurden.⁹ In diesem wurden die Schwankungen der Leistung I eines Moden einer Single-Longitudinal-Laserdiode durch Fluktuationen in der spontanen Emission J erklärt. Diese Modell beruht also auf einer Kopplung zwischen spontaner Emission und Laseroszillation. Hierbei lautet das elektrische Feld des oszillierenden Moden:

$$E = \beta \cdot e^{-i\omega t} + \beta^* \cdot e^{i\omega t}. \quad \text{Gl. 4. 84}$$

Die Amplituden β und β^* sind über folgende stochastische Differentialgleichung definiert:

$$\frac{d\beta}{dt} = -\frac{1}{2} \cdot v_g \cdot (\gamma - g) \cdot \beta + F_\beta(t) \quad \text{Gl. 4. 85}$$

und

$$\frac{d\beta^*}{dt} = -\frac{1}{2} \cdot v_g \cdot (\gamma - g) \cdot \beta^* + F_{\beta^*}(t). \quad \text{Gl. 4. 86}$$

In diesen Formulierungen bezeichnet g die Verstärkung des Moden pro Längeneinheit, γ den durchschnittlichen Verlust pro Längeneinheit. $F_\beta(t)$ bzw. $F_{\beta^*}(t)$ stehen für eine durch

⁸ Motoichi Ohtsu; Analyses of Mode-Hopping Phenomena in an AlGaAs Laser, IEEE Journal of Quantum Electronics, vol. 22 no. 4, S. 535-543, 1986

⁹ C. Henry et al.; Partitions Fluctuations in Nearly Single-Longitudinal-Mode Lasers, J. of Lightwave Technology, Vol. LT-2 No. 3, June 1984

spontane Emission verursachte Langevin-Rauschquelle.¹⁰ Die zeitliche Änderung bzw. Ableitung der Amplituden β und β^* des elektrischen Feldes sind hier wieder eine Funktion ihrer selbst. Die Ausprägung des Lasermoden innerhalb einer Laserdiode ist proportional zur Differenz zwischen Verlust γ und Verstärkung g . Im statischen Zustand (ohne Rauschen) ist diese Differenz konstant, so dass folgende Zeitkonstante definiert werden kann:

$$\tau = \frac{1}{v_g \cdot (\gamma - g_1)}. \quad \text{Gl. 4. 87}$$

Damit vereinfachen sich die stochastischen Differentialgleichungen zu:

$$\frac{d\beta}{dt} = -\frac{\beta}{2 \cdot \tau} + F_{\beta}(t) \quad \text{Gl. 4. 88}$$

und

$$\frac{d\beta^*}{dt} = -\frac{\beta^*}{2 \cdot \tau} + F_{\beta^*}(t). \quad \text{Gl. 4. 89}$$

Die Lichtintensität (bzw. Anzahl der erzeugten Photonen) des Lasermoden entspricht dem Betrag der **Gl. 4. 84**:

$$J = \beta^* \cdot \beta. \quad \text{Gl. 4. 90}$$

Die Ableitung dieser Gleichung und das Einsetzen von **Gl. 4. 89** führen zu der Lichtintensität der nicht oszillierenden und damit spontanen Emission:

¹⁰ Diese sind nach P. Langevin beschrieben, der 1908 Diffusionsvorgänge durch solche stochastische Differentialgleichung beschrieb. Entscheidend Annahme Langevin's war es hierbei, die einem diffundierenden Teilchen entgegengebrachte Reibungskraft proportional zu seiner Geschwindigkeit zu setzen. Siehe: P. Langevin; Comt. Rend. 146, 1908, S. 530

$$\frac{dJ}{dt} = -\frac{J}{\tau} + R + F_J(t) \quad \text{Gl. 4. 91}$$

mit

$$R + F_J(t) = \beta^* \cdot F_\beta(t) + \beta \cdot F_{\beta^*}(t). \quad \text{Gl. 4. 92}$$

Hierbei bezeichnet R den zeitlichen Mittelwert des Rauschens der beiden komplexen Summanden und kann als die Rate der spontanen Emission interpretiert werden. Auf Basis dieses Modells lässt sich die Wahrscheinlichkeit für den Fall berechnen, dass die Intensität der nicht oszillierenden spontanen Emission J das Produkt aus Frequenz und mittlerer Intensität des Lasermodes I_0 übersteigt:

$$p_f = \exp\left(-f \cdot \frac{I_0}{J_0}\right). \quad \text{Gl. 4. 93}$$

Diese spektrale Analyse zeigt deutlich den niederfrequenten Charakter der beschriebenen Kopplung (**Abb. 4.20**).

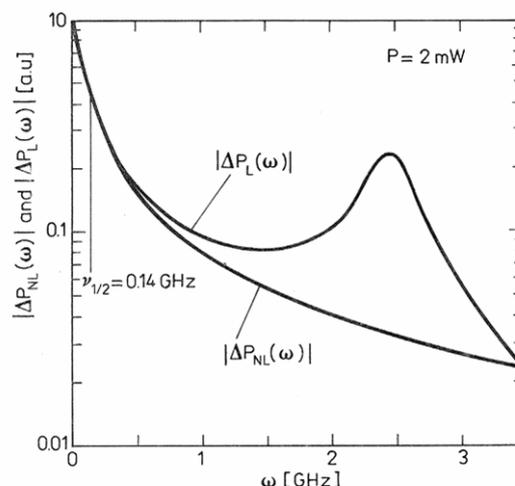


Abb. 4.20: Spektrum der Intensitätsschwankungen der spontanen Emission P_{NL} und der zugehörigen Antwort im Lasermode P_L ¹¹.

¹¹ B. Mroziwicz, M. Bugajski, W. Nakwaski; Physics of Semiconductor Lasers, Polish Scientific Publishers 1991, S. 381

Für die direkte Kopplung zweier konkurrierender Moden kann ein ähnliches Modell formuliert werden. Hierbei kann jedem möglichen Mode i innerhalb einer Laserdiode eine Photonengeneration P_i zugeordnet werden:

$$\frac{dP_i}{dt} = (G_i - \gamma_i)P_i + R_i + F_i(t). \quad \text{Gl. 4. 94}$$

Hierin bezeichnet G_i die Verstärkungsfaktoren der Moden, γ_i die durch Absorption bedingten Verluste und R_i die Raten der spontanen Emission. Des Weiteren herrscht innerhalb der Laserdiode eine bestimmte Inversionsrate I , die die Anzahl der angeregten Elektronen im Leitungsband beschreibt:

$$\frac{dI}{dt} = \Lambda - \gamma \cdot I - \sum G_i P_i + F_I(t). \quad \text{Gl. 4. 95}$$

Sie wird bestimmt aus der so genannten Pumprate Λ , die das Bereitstellen von immer neuen angeregten Elektronen beschreibt, und einem Inversionsverlust γ . Die Größe F_i in der Photonengeneration und der Inversion beschreiben wieder die Langevin-Rauschquellen. Beschränkt man sich für eine theoretische Betrachtung auf zwei ausgeprägte Moden P_1 und P_2 , dann gilt für die Inversion:

$$\frac{dI}{dt} = \Lambda - \gamma \cdot I - (G_1 P_1 + G_2 P_2) + F_I(t). \quad \text{Gl. 4. 96}$$

Der Verstärkungsfaktor G_i eines Modes i kann in einen inversionsabhängigen Term $G(I)$ und einen inversionsunabhängigen Teil separiert werden. Für die begrenzte Betrachtung zweier Moden wird dieser inversionsunabhängige Teil durch die Eigendämpfung α_i und die Dämpfung β_i des konkurrierenden Modes beschrieben. Die Verstärkungsfaktoren der Moden lauten dadurch:

$$G_1 = G(I) - \alpha_1 P_1 - \beta_1 P_2 \quad \text{Gl. 4. 97}$$

und

$$G_2 = G(I) - \alpha_2 P_2 - \beta_2 P_1. \quad \text{Gl. 4. 98}$$

Für eine weitere Vereinfachung wird eine symmetrische Ausbildung der beiden Moden angenommen. Einzig die Absorptionsverluste γ_i der einzelnen Moden in Gleichung 4.90 unterscheiden sich. Damit gilt für das beschriebene Gleichungssystem:

$$\gamma_1 \neq \gamma_2, \quad \text{Gl. 4. 99}$$

$$G_1(I) = G_2(I) = G(I), \quad \text{Gl. 4. 100}$$

$$R_1 = R_2 = R, \quad \text{Gl. 4. 101}$$

$$\alpha_1 = \alpha_2 = \alpha, \quad \text{Gl. 4. 102}$$

$$\beta_1 = \beta_2 = \beta. \quad \text{Gl. 4. 103}$$

Die Autokorrelationen der Langevin-Rauschquellen F_i und F_I lauten mit dieser Vereinfachung:

$$F_i(t_1) * F_i(t_2) = 2 \cdot R \cdot P_i \cdot \delta(t_1 - t_2) \quad \text{Gl. 4. 104}$$

und

$$F_I(t_1) * F_I(t_2) = [r \cdot \Lambda + \gamma \cdot I + 2 \cdot R \cdot (P_1 + P_2)] \cdot \delta(t_1 - t_2). \quad \text{Gl. 4. 105}$$

Die entsprechenden Kreuzkorrelationen ergeben sich zu:

$$F_1(t_1) * F_2(t_2) = 0 \quad \text{Gl. 4. 106}$$

und

$$F_i(t_1) * F_i(t_2) = -2 \cdot R \cdot P_i \cdot \delta(t_1 - t_2). \quad \text{Gl. 4. 107}$$

Damit steht das Inversionsrauschen in einer Abhängigkeit zu Änderungen der Anregung (bzw. Pumprate), der spontanen Emission und den stimulierten Emissionen in den Moden 1 und 2. Im zeitlichen Mittelwert gilt für einen Betrieb der Laserdiode unter stabilen Bedingungen:

$$\frac{dP_i}{dt} = 0 = (\bar{G}_i - \gamma_i) \bar{P}_i + R \quad \text{Gl. 4. 108}$$

bzw.

$$R = (\gamma_1 - \bar{G}_1) \bar{P}_1 = (\gamma_2 - \bar{G}_2) \bar{P}_2 \quad \text{Gl. 4. 109}$$

bzw.

$$\frac{\bar{P}_1}{\bar{P}_2} = \frac{(\gamma_2 - \bar{G}_2)}{(\gamma_1 - \bar{G}_1)}. \quad \text{Gl. 4. 110}$$

Hierin ist klar ersichtlich, dass schon kleinste Änderungen in den Absorptionsverlusten γ_i oder den Verstärkungen G_i zu großen Änderungen in den Photonengenerationen bzw. Lichtintensitäten der einzelnen Moden führen. Die Lösung eines solchen Differentialgleichungssystems führt zu der in **Abb. 4.21** dargestellten spektralen Abhängigkeit des Modenrauschens. Als wesentliches Ergebnis bleibt festzuhalten, dass es sich auch bei zwei konkurrierenden Moden um einen niederfrequenten Effekt handelt.

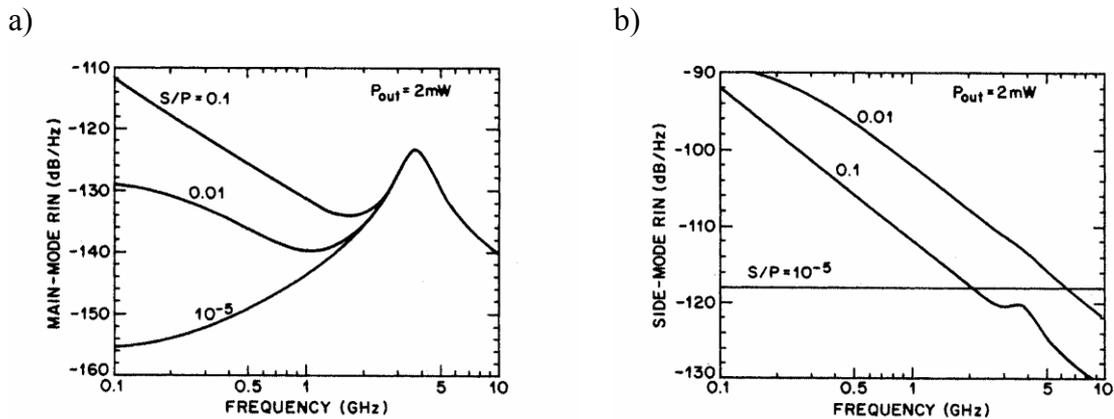


Abb. 4.21: Spektren des Intensitätsrauschens (RIN – Relative-Intensity-Noise) für a) Hauptmode P und b) Nebenmode S einer Singlemode Laserdiode für unterschiedliche Verhältnisse von S/P ¹².

c) Auswirkungen auf das Systemverhalten

Mit den beschriebenen thermischen Effekten einerseits und den Kopplungseffekten andererseits können Schwankungen in Leistung und Wellenlänge erklärt werden. Eine wesentliche Frage bleibt jedoch die Auswirkung auf die systemtheoretische Beschreibung eines laserbasierten optischen Senders.

Die beschriebene Temperaturabhängigkeit des Laserbetriebs hat direkten Einfluss auf das Modulationsverhalten und somit auf die Phase des optischen Signals. Die Verzugszeit zwischen einem Stromimpuls und dem Einsetzen der stimulierten Emission eines Lasers ist abhängig von der Schwellenstromdichte J_S . Diese ist wiederum eine Funktion der Temperatur:

$$J = J_S \cdot \exp\left(\frac{eU}{n \cdot k \cdot T}\right). \quad \text{Gl. 4. 111}$$

Liegen für zwei unterschiedliche Temperaturen T_1 und T_2 an der Laserstruktur die gleichen Stromdichten J an, dann gilt:

¹² G. P. Agrawal; Mode-partition noise and intensity correlation in a two-mode semiconductor laser, Physical Review A, vol. 37, no. 7, S. 2488 – 2494, 1988

$$J(T_1) = J(T_2) \quad \text{Gl. 4. 112}$$

bzw.

$$J_{S1} \cdot \exp\left(\frac{1}{T_1}\right) = J_{S2} \cdot \exp\left(\frac{1}{T_2}\right). \quad \text{Gl. 4. 113}$$

Mit den angenommenen Werten $d=0,1\mu\text{m}$, $n_S=2\cdot 10^{18}\text{cm}^{-3}$ und $\tau=3\text{ns}$ ergibt sich für eine GaAs-Laserdiode eine Schwellenstromdichte J_S von:¹³

$$J_S = \frac{e \cdot n_S \cdot d}{\tau} \approx 1\text{kA/cm}^2. \quad \text{Gl. 4. 114}$$

Erhöht sich die Temperatur T von 293K auf 294K, dann ergibt sich eine ebenfalls erhöhte Schwellenstromdichte von:

$$J_S(294\text{K}) = J_S(293\text{K}) \cdot \exp\left(\frac{1}{293\text{K}} - \frac{1}{294\text{K}}\right) \approx 1,00012\text{kAcm}^2. \quad \text{Gl. 4. 115}$$

Für eine feste Schwellenstromdichte J von $1,1\text{kA/cm}^2$ ändert sich somit die Verzugszeit um:

$$\Delta t_d = \tau \cdot \ln\left(\frac{J - J_S(293\text{K})}{J - J_S(294\text{K})}\right) \approx 3,6\text{ps}. \quad \text{Gl. 4. 116}$$

Diese geringfügige Änderung der Temperatur von 1K führt schon zu einem Laufzeitfehler eines aufmodulierten Rechtecksignals, dass von der Größenordnung ebenso sehr gering erscheint. Jedoch führt schon dieser marginale Fehler zu einer Abweichung in der optisch-inkohärenten Laufzeitmessung von etwa 1mm . Hier ist klar ersichtlich, dass beim Verwenden von Halbleiterlaserdioden entweder für eine konstante Temperatur oder eine

¹³ W. Harth, H. Grothe; Sende und Empfangsdioden für die optische Nachrichtentechnik, Teubner 1984, Stuttgart, S. 75

entsprechende Kompensation gesorgt werden muss. Um weiterhin unabhängig von der Verzugszeit t_d zu werden, ist ein Betrieb der Laserdioden innerhalb der Inversion zu gewährleisten. Das bedeutet, dass die Stromdichte aufgrund der Modulation die Schwellenstromdichte nicht unterschreiten darf. Dies führt zu einem Signal-Offset und schränkt natürlich den Modulationskontrast im gesendeten Signal ein. Wird ein solcher Betrieb sichergestellt, dann kann jedoch eine Änderung der Schwellenstromdichte trotzdem zu einer Änderung des Modulationsverhaltens führen. Ein Ansteigen der Schwellenstromdichte führt bei einem konstanten Strom natürlich zu einer Erhöhung der spontanen und einer Reduzierung der stimulierten Emission. Über Kopplung der Inversion mit der spontanen und stimulierten Emission ist neben einem Modenwechsel, egal ob thermisch oder statistisch bedingt, ebenso mit einer Änderung des Übertragungsverhaltens zu rechnen. Eine Zunahme der Inversion wird zu einem Ansteigen der Leitfähigkeit innerhalb des aktiven Bereichs der Laserdiode führen.

Betrachtet man nun das Ersatzschaltbild einer Laserdiode als eine einfache Kombination aus ohmscher und kapazitiver Last, dann kann eine solche Änderung näherungsweise systemtheoretisch beschrieben werden (**Abb. 4.22**). Ein Betrieb oberhalb der Schwellenspannung U_S besitzt als ohmsche Last den differentiellen Widerstand r , der sich aus der Dioden(I,U)-Kennlinie bestimmen lässt. Dieser bewegt sich typischerweise in einem Bereich von etwa 10 bis 20 Ohm.

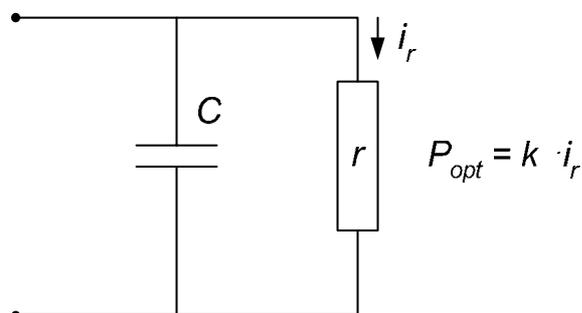


Abb. 4.22: RC-Ersatzschaltbild einer Laserdiode

Die Ersatzkapazität ergibt sich aus dem Frequenzverhalten einer Laserdiode mit:

$$\frac{0,34}{f_g} = \tau = r \cdot C . \quad \text{Gl. 4. 117}$$

Mit einer angenommenen Grenzfrequenz von 3GHz und einem differentiellen Widerstand von 150Ω errechnet sich eine Ersatzkapazität von $7,5\text{pF}$. Der Frequenzgang einer solchen Ersatzschaltung mit einer Variation des differentiellen Widerstandes von $\pm 50\Omega$ zeigt **Abb. 4.23**. Für eine Modulation mit 20MHz ist der Phasenversatz scheinbar marginal, bekommt jedoch für die Applikation der Entfernungsmessung eine entscheidende Bedeutung. Die Änderung des Phasenwinkels für unterschiedliche r berechnet sich zu:

$$\Delta\varphi = \arctan(2\pi \cdot f \cdot r_1 \cdot C) - \arctan(2\pi \cdot f \cdot r_2 \cdot C) . \quad \text{Gl. 4. 118}$$

Im gegebenen Beispiel beträgt der Phasenversatz für eine geringfügige Änderung des differentiellen Widerstandes um 50Ω etwa $0,27^\circ$. Diese geringe Änderung verursacht in der Entfernungsmessung direkt einen Messfehler von $5,6\text{mm}$. Diese einfache Betrachtung mittels eines RC-Tiefpassverhaltens spiegelt nur annähernd das Modulationsverhalten einer Laserdiode wieder. Eine genauere Betrachtung gelingt auf analytischem Weg über die Kopplung der Ladungsträgerdichte $n(x,t)$ und der Photonendichte $\psi(x,t)$:

$$\frac{\partial n}{\partial t} = D \frac{\partial^2 n}{\partial x^2} + J(x,t) - J_s(x,t) - \frac{n}{\tau_s} \quad \text{Gl. 4. 119}$$

Hierfür ergibt sich das Übertragungsverhalten eines Verzögerungsgliedes zweiter Ordnung als Laplace-Transformierte: ¹⁴

$$M(s) = \frac{\omega_0^2 / c}{s^2 + s\gamma + \frac{c_1}{c} \omega_0^2} . \quad \text{Gl. 4. 120}$$

¹⁴ Willie W. Ng and E. A. Sovero; An Analytic Model for the Modulation Response of Buried Heterostructure Lasers, IEEE Journal of Quantum Electronics, Vol. 20, No. 9, 1984, S.1008 - 1015

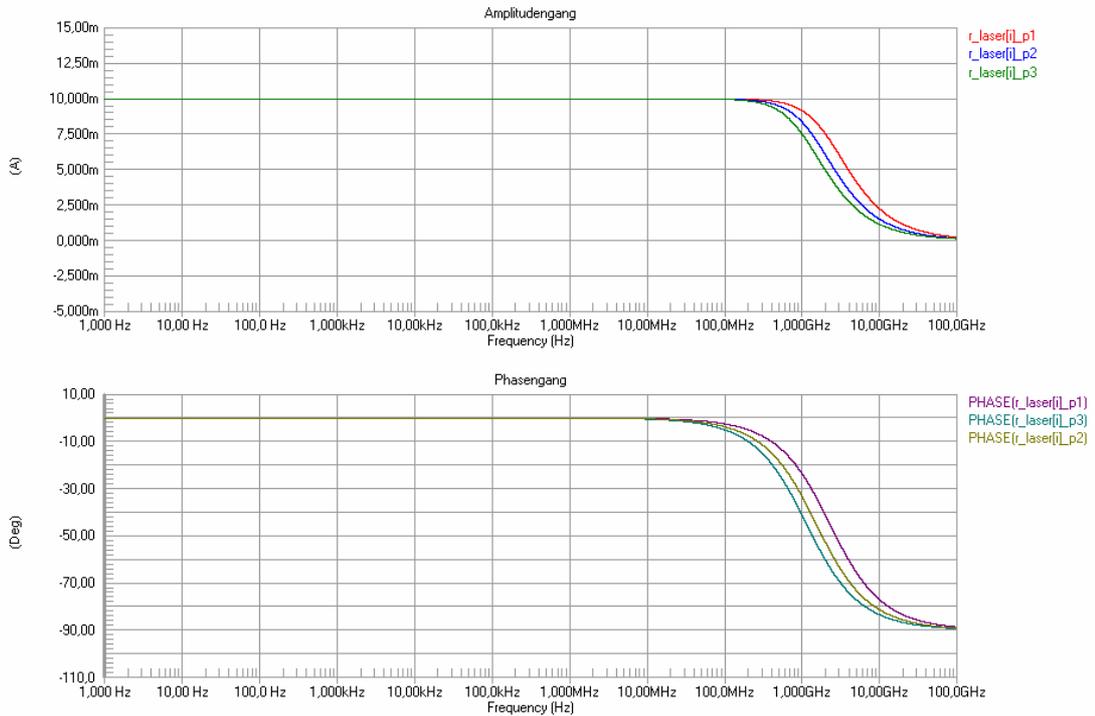


Abb. 4.23: Simulierter Frequenzgang des RC-Ersatzschaltbildes einer Laserstruktur ($R = 10/15/20\text{ Ohm}$; $C = 7,5\text{pF}$)

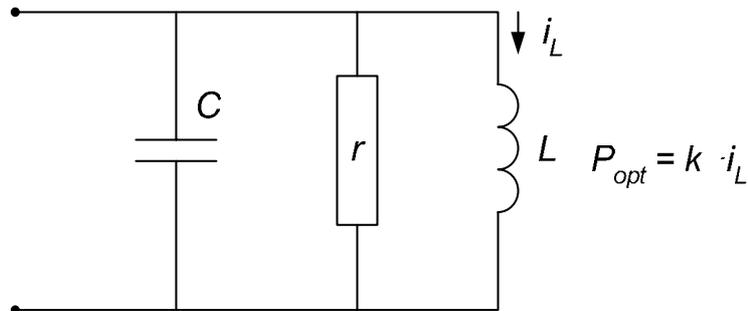


Abb. 4.24: RCL-Ersatzschaltbild einer Laserdiode

Das Modulationsverhalten einer Laserdiode kann somit durch ein entsprechendes RCL-Ersatzschaltbild beschrieben werden (**Abb. 4.24**). Die Größe $M(s)$ beschreibt hier direkt das optische Ausgangssignal in Abhängigkeit von der Frequenz. Durch eine SPICE-Simulationen (**Abb. 4.25**) konnte für ein solches Tiefpassverhalten bei einer Änderung des differentiellen Widerstands r um 5Ohm eine Änderung des Phasenwinkels $\Delta\varphi$ von etwa $0,44^\circ$ ermittelt werden, was in der Entfernungsmessung einem Messwertfehler von $9,2\text{mm}$

entspricht. Eine Bewertung dieser Simulationen muss berücksichtigen, dass die verwendete Modulationsquelle einer idealen Stromquelle entspricht und parasitäre Einflüsse einer Verbindungstechnik vernachlässigt wurden.

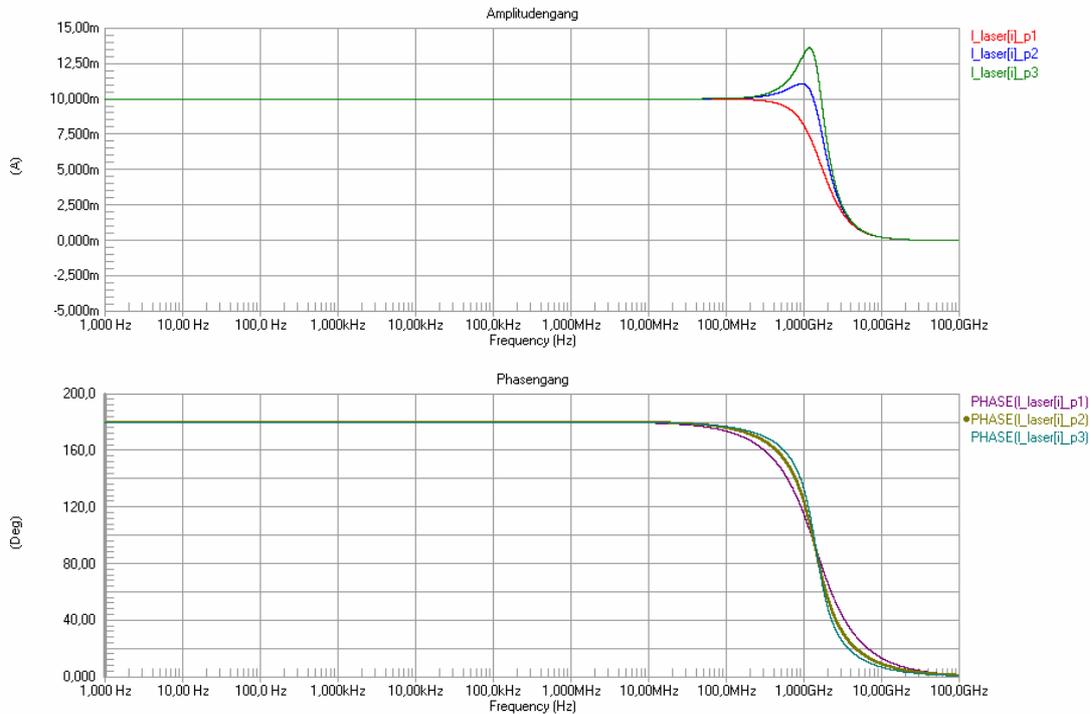


Abb. 4.25: Simulierter Frequenzgang des RCL-Ersatzschaltbildes einer Laserstruktur ($R = 10/15/20 \text{ Ohm}$; $C = 7,5 \text{ pF}$; $H = 1,8 \text{ nH}$)

4.5.2 Impedanzstabilität des quasioptischen Mischers

Für die systemtheoretische Beschreibung eines optischen Empfängers ist es zweckmäßig, diesen durch ein Ersatzschaltbild zu veranschaulichen. Herkömmliche Empfänger wie PIN- oder Avalanche-Photodioden sind in der Literatur ausführlich beschrieben (**siehe Kap. 3.3.1 und 3.3.2**). Diese besitzen in erster Näherung ein rein kapazitives Verhalten. Sicher hat die Intensität des eingestrahelten Lichts und die damit verbundene Ladungsträgergeneration einen Einfluss auf deren Kapazität, da in diesem Modell Ladungsträger innerhalb des Dielektrikums eingebracht werden. Betrachtet man den

generierten Photostrom im Gegensatz zum klassischen Modell nicht als ideale parallele Stromquelle, sondern berücksichtigt einen zusätzlichen hochohmigen und variablen Parallelwiderstand, dann wird sich für diese Anordnung eine intensitätsabhängige Zeitkonstante einstellen (**Abb. 4.26**). Für rein kapazitives Verhalten ergibt sich die Zeitkonstante aus der Diodenkapazität und der ohmschen Last R_L :

$$\tau = R_L \cdot C_D . \quad \text{Gl. 4. 121}$$

Die Berücksichtigung des variablen Parallelwiderstands r führt zu:

$$\tau = [R_L \parallel r(P_{opt})] \cdot C_D . \quad \text{Gl. 4. 122}$$

Mit der intensitätsabhängigen Zeitkonstante stellt sich bei der Detektion intensitätsmodulierter optischer Signale ein mehr oder weniger zeitverzögerter Signalstrom ein. Dies hat jedoch keinen weiteren direkten Einfluss auf die nachgeschaltete Signalverarbeitung. So wird einem angeschlossenen Mischer zwar ein mehr oder weniger verzögertes Signal übergeben, aber sein eigenes Verhalten hängt nicht zwangsläufig von der empfangenen Lichtintensität ab. Hier liegt ein entscheidender Unterschied zu quasioptischen Mischern wie dem Photogate- oder MSM-PMD. Da zusätzlich zur Photogeneration auch der Mischprozess innerhalb der Diodenkapazität erfolgt, ist das Mischverhalten ebenso intensitätsabhängig wie die Signaldetektion.

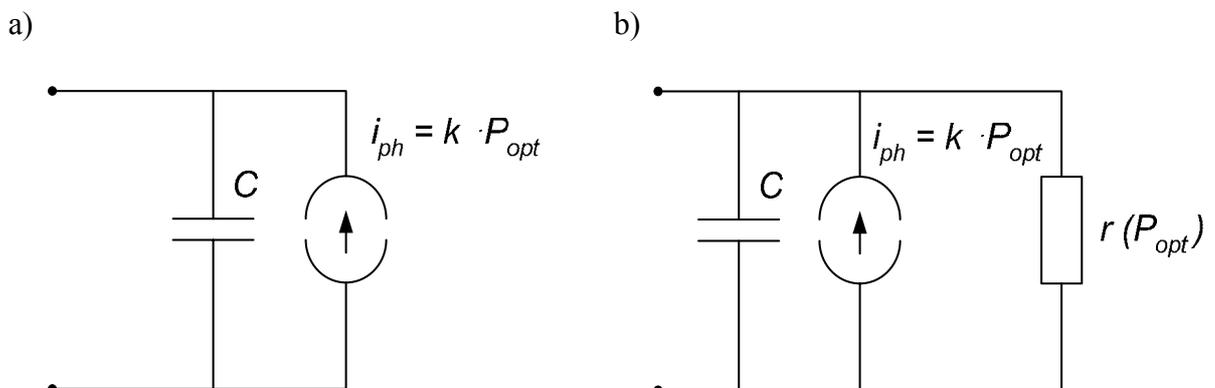


Abb. 4.26: Unterschiedliche Ersatzschaltbilder einer PIN-Photodiode a) rein kapazitiv b) mit parallelem variablen Widerstand

Da die Prozesse der Detektion und des Mischens sich örtlich wie auch zeitlich nicht unterscheiden lassen, können keine separaten Zeitkonstanten bestimmt werden. Aus systemtheoretischer Sicht besitzt eine solche Unterscheidung keine weitere Bedeutung, sondern nur das intensitätsabhängige Gesamtverhalten ist von Interesse. Die Impedanzmessung des Photogate-PMD (siehe **Kap. 3.4.4**) unter kontrollierter Intensitätseinstrahlung zeigt deutlich deren Einfluss (**Abb. 4.27** u. **Abb. 4.28**). Die durch Lichteinstrahlung generierten Photoelektronen stören erheblich den Zustand der Verarmung. Mit zunehmender Ladungsträgergeneration zeigt sich insbesondere für höhere Frequenzen eine deutliche Änderung der Photogate-Kapazität. Der ohmsche Anteil bleibt jedoch insbesondere im hochfrequenten Bereich nahezu unbeeinflusst. Nach **Gl. 4. 122** stellt sich mit zunehmender Lichteinstrahlung eine Erhöhung der Zeitkonstante ein, wobei jedoch in realen Systemen die Impedanz der Modulationsquelle zu Berechnung des resultierenden Phasenfehlers mit zu berücksichtigen ist. Mit einer Frequenz von beispielsweise 40MHz ändert sich die Zeitkonstante durch die Einstrahlung von $50\mu\text{W}$ (bei $\lambda=650\text{nm}$) von 0.5692 auf 0.7316ns . Neben der photonischen kann ebenso eine thermische Generation von Ladungsträgern unterhalb der Photogates stattfinden. Auf den Zustand der Verarmung unterhalb der Photogates wird dieser Effekt ebenso Einfluss besitzen wie die Lichteinstrahlung.

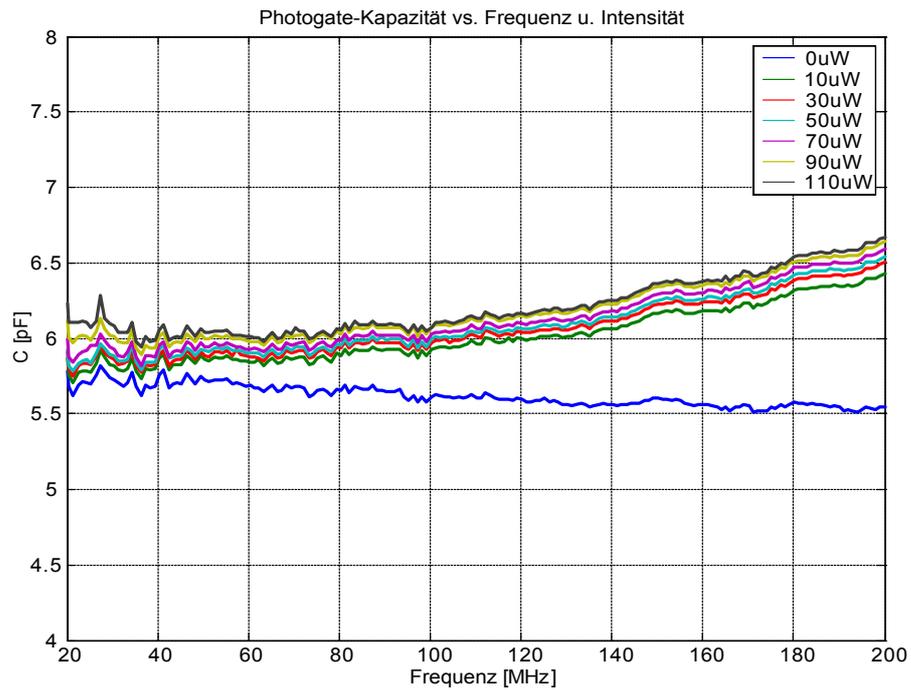


Abb. 4.27: Änderung des kapazitiven Anteils der Impedanzmessung einer Photogate-Struktur in Abhängigkeit der eingestrahelten optischen Leistung

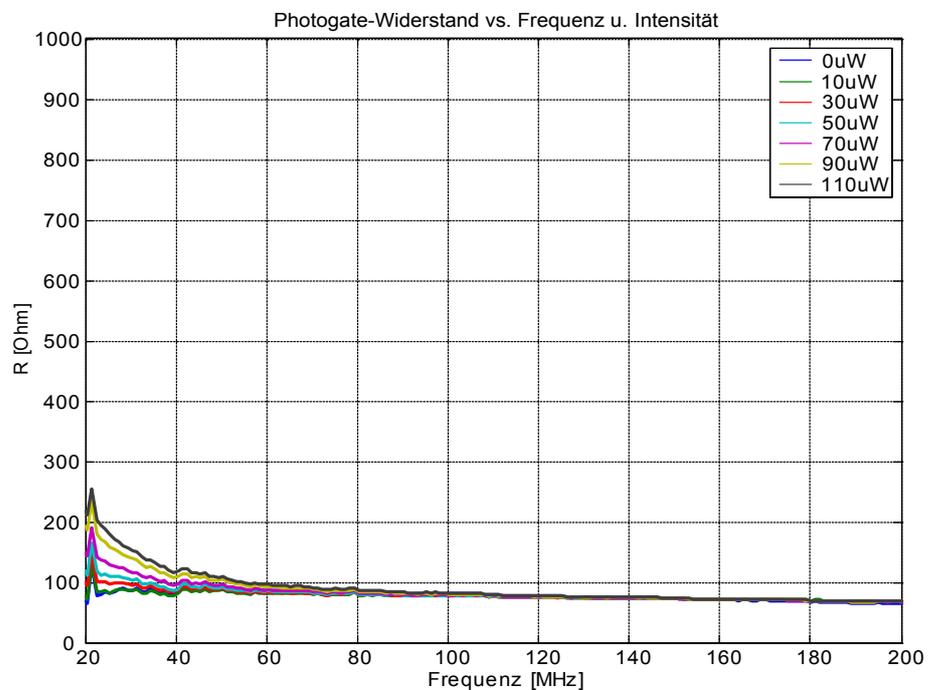


Abb. 4.28: Änderung des ohmschen Anteils der Impedanzmessung einer Photogate-Struktur in Abhängigkeit der eingestrahelten optischen Leistung

4.6 Praktische Anwendung der optisch intensitätsmodulierten Phase-Shift-Interferometrie mittels PMD

4.6.1 Eindimensionale TOF-Entfernungsmessung

Auf Basis der PMD-Technologie wurde neben verschiedenen PMD-Zeilen und -Matrizen auch ein Einzelpixeldemonstrator realisiert, der hinsichtlich Messrate und Genauigkeit den höchsten Ansprüchen unterliegt. Ein solcher Demonstrator im monokularen Aufbau zeigt **Abb. 4.29**. Aufgrund seiner geringen Empfangsapertur, der geringen Sendeleistung von ca. $1mW$ und der geforderten Messrate von $50Hz$ ergeben sich sehr harte Anforderungen an die Schaltungstechnik, Optik und Auswertalgorithmik. Das Ergebnis einer ersten Realisierung unterlag einer erheblichen Langzeitdrift von ca. $15cm$ (**Abb. 4.30**). Neben relativ langsamen Änderungen konnten hier ebenso sprunghafte Abweichungen vom tatsächlichen Entfernungswert festgestellt werden. Zur Verminderung der zuvor beschriebenen thermischen Effekte im Laserdiodensender wurde zur Stabilisierung der Lasertemperatur eine thermische Kopplung an einen Kühlkörper mit möglichst großer Wärmekapazität vorgenommen, was zu einer Verbesserung des Messfehlers um den Faktor 4 führte (**Abb. 4.31**).



Abb. 4.29: Technologiedemonstrator eines PMD-basierten eindimensionalen Laserentfernungsmessgerätes für den industriellen Einsatz

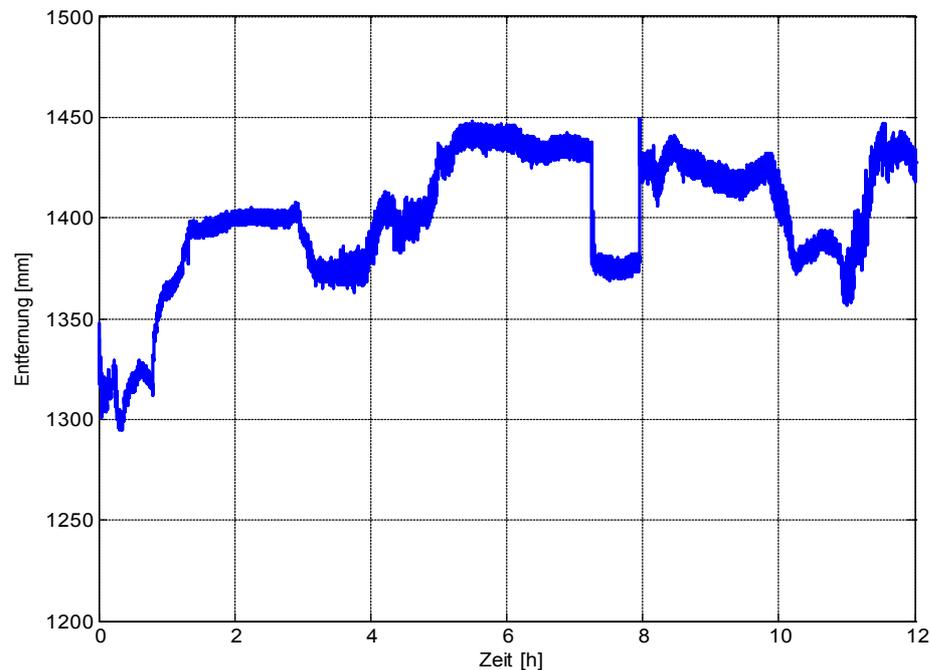


Abb. 4.30: Langzeitmessung einer intensitätsmodulierten Laser-TOF-Entfernungsmessung ohne stabilisierende Massnahmen ($f_{mod} = 16\text{MHz}$, $P_{opt} = 1\text{mW}$, $f_{mess} = 12,5\text{Hz}$)

Die langsamen Drifterscheinungen konnten deutlich reduziert werden, jedoch ergeben sich immer noch sehr deutliche sprunghafte Änderungen von mehreren Zentimetern. Neben der gesunkenen Amplitude (von ehemals 6cm auf etwa 3cm) verschob sich das Zeitverhalten hin zu deutlich zu höheren Frequenzen.

Insgesamt konnte jedoch noch kein zufrieden stellendes Ergebnis erzielt werden. Als Ursache dieses Verhaltens wurden Instabilitäten im Lasersender aufgrund der in **Kap. 4.5.1** beschriebenen Kopplungseffekte vermutet. Eine Untersuchung des spektralen Verhaltens der verwendeten Laserdiode betätigte diesen Verdacht. Im gemessenen Spektrum einer Single-Mode-Laserdiode ist grundsätzlich ein Monomodebetrieb zu erkennen, jedoch erscheinen spontan deutliche Nebenmoden (**Abb. 4.33**). Durch das Einschalten der Modulation von 16MHz zeigten sich noch wesentlich deutlichere Nebenmoden (**Abb. 4.34**). Die Ursache hierfür findet sich in der optischen Gewinnkurve $g(h \cdot f)$ (**Abb. 4.34**), die über die Abhängigkeit von der Inversionsdichte ebenfalls moduliert wird. Damit liegen temporär auch Nebenmoden im Maximum der Gewinnkurve, schwingen an und können so zur Emission beitragen.

Ein weiterer Effekt der Intensitätsmodulation ist die Modulation der Lichtwellenlänge, da der Brechungsindex neben der Temperatur und der Wellenlänge selbst auch von der Inversionsdichte n abhängt.¹⁵ Für eine systemtheoretische Betrachtung des modulierten Lasers ist das Auftreten mehrerer Lasermoden von entscheidender Bedeutung. Der sonst spontane und niederfrequente Wechsel in einen anderen Mode wird nun sehr schnell erzwungen. Die örtlichen Inhomogenitäten in der Inversionsdichte des statischen Betriebs werden reduziert (**Abb. 4.35**), wodurch sich bei einer ausreichend hohen Modulationsfrequenz und Modulationstiefe eine konstante Impedanz einstellen wird. Ein örtlicher Stau angeregter Ladungsträger innerhalb der Laserdiodenstruktur kann nun immer durch einen benachbarten Mode abgeräumt werden.

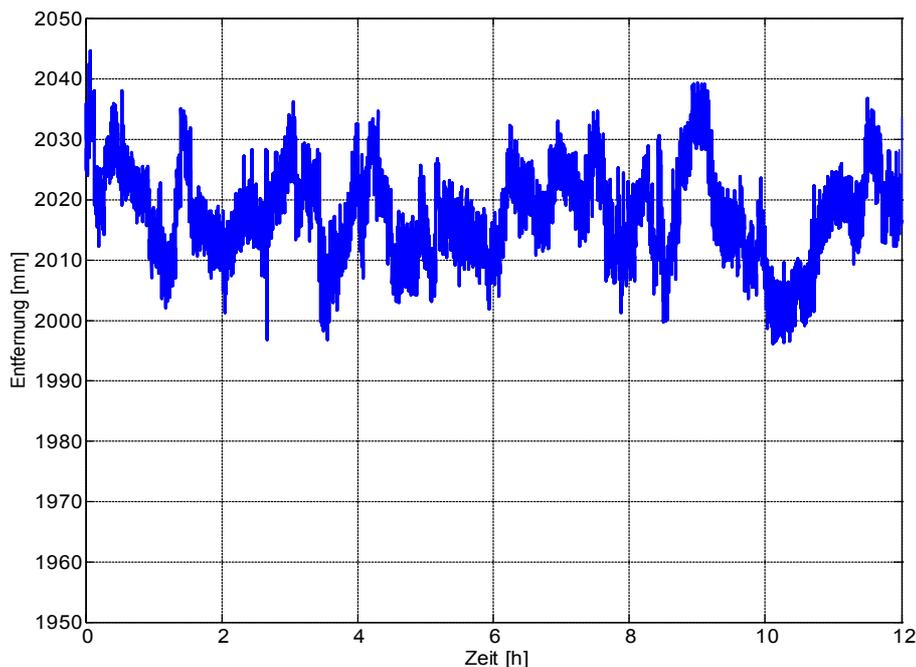


Abb. 4.31: Langzeitmessung einer intensitätsmodulierten Laser-TOF-Entfernungsmessung mit stabilisierter Lasertemperatur ($f_{mod} = 16\text{MHz}$, $P_{opt} = 1\text{mW}$, $f_{mess} = 12,5\text{Hz}$)

¹⁵ K. Kishino et al.; Wavelength variation of 1.6 μm wavelength buried heterostructure GaInAsP/InP lasers due to direct modulation, IEEE Journal of Quantum Electronic, Vol. 18, 1982, S. 343-351

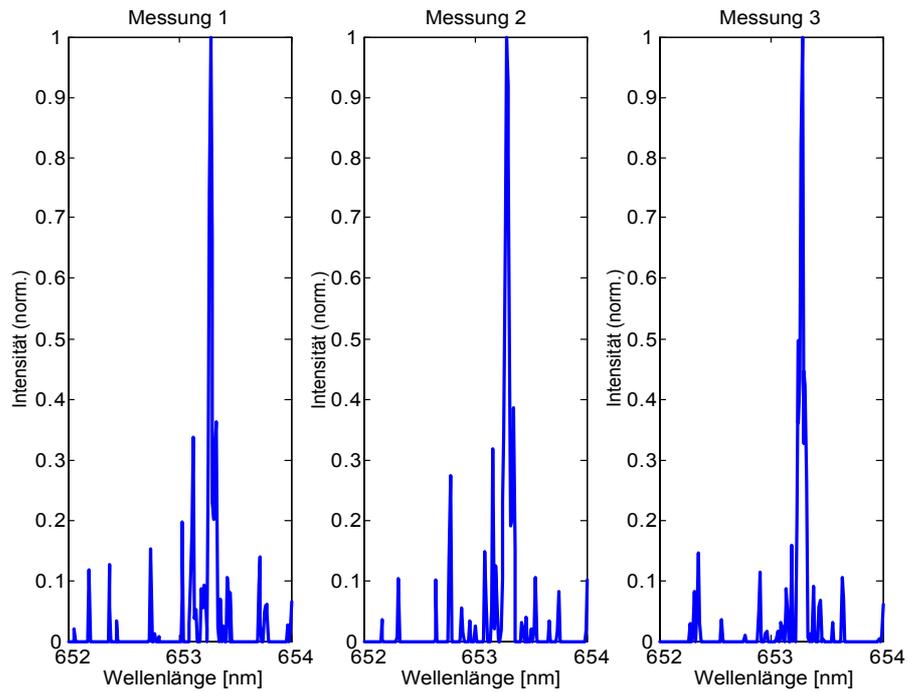


Abb. 4.32: Gemessenes Spektrum der unmodulierten Single-Mode-Laserdiode

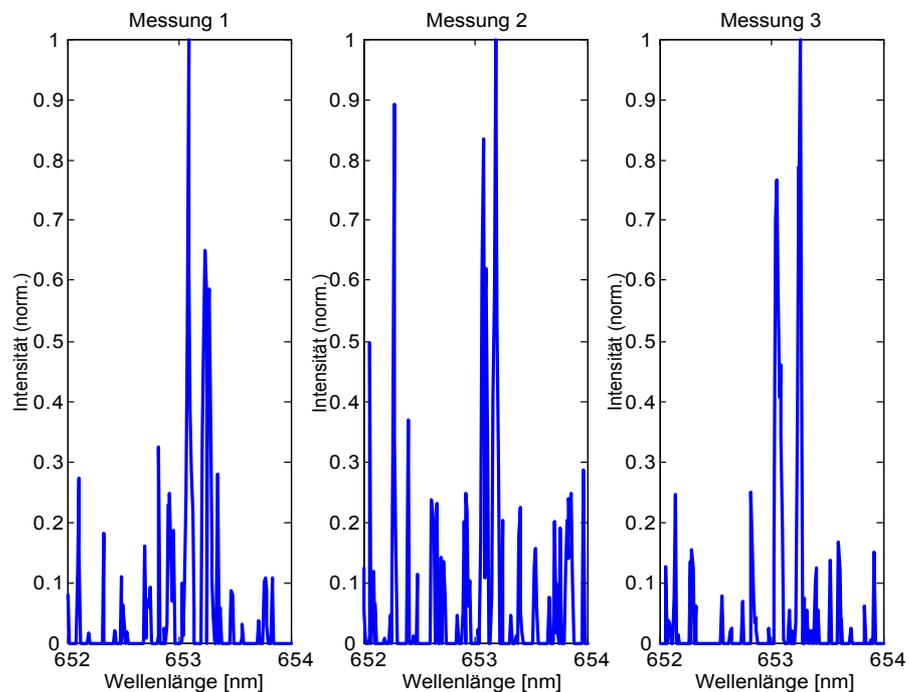


Abb. 4.33: Gemessenes Spektrum der mit 16MHz modulierten Single-Mode-Laserdiode

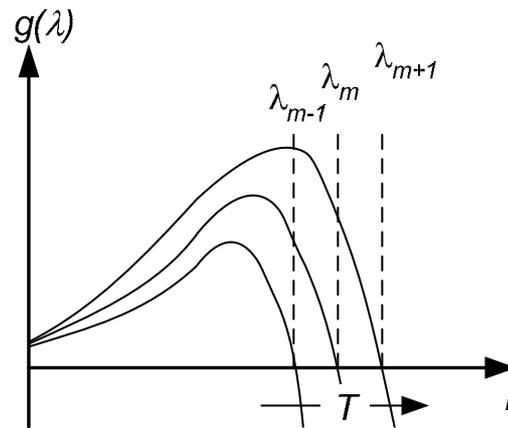


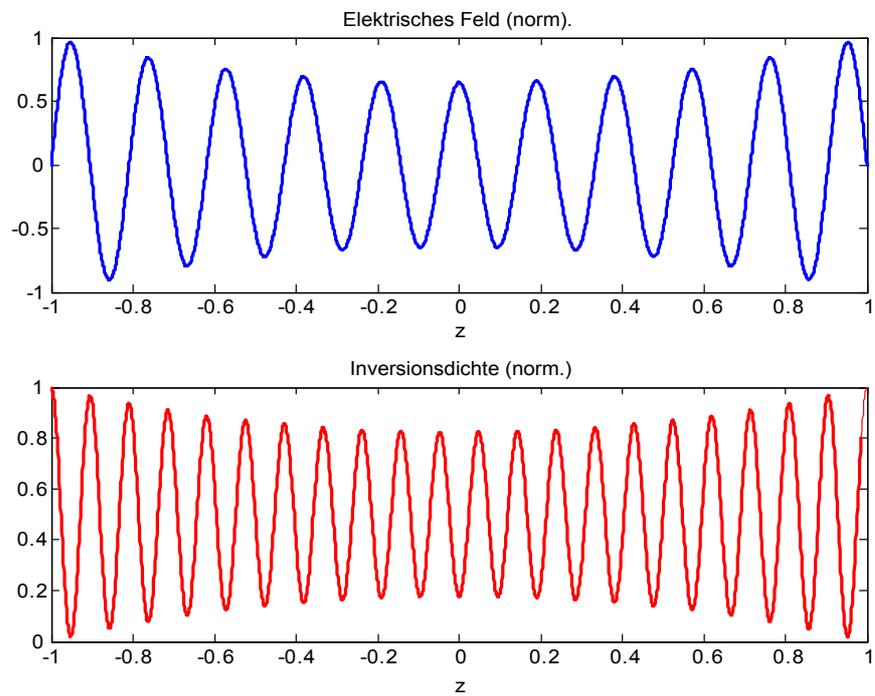
Abb. 4.34: Abhängigkeit des optischen Gewinns $g(\lambda)$ einer Laserdiode von der Inversionsdichte n .

Die Auswirkung auf das spektrale Verhalten einer solchen HF-Modulation zeigt **Abb. 4.36**. Hier wurde das zur Entfernungsauflösung benötigte Signal von 15MHz mit zusätzlichen 460MHz überlagert. Die Ausprägung der einzelnen Moden folgt nun einer Gausverteilung, so dass hier von einem Quasi-Multimode-Betrieb gesprochen werden kann. Deutlich ist zu erkennen, dass sich neben dem Hauptmode Nebenmoden mit hoher Amplitude ausgeprägt haben und ihre Schwankungen über mehrere aufeinander folgende Messungen relativ gering ausfallen. Dieses Verhalten konnte durch eine deutliche Reduzierung der Langzeitdrift in der Entfernungsmessung bestätigt werden (**Abb. 4.37**). Der nun konstante Entfernungswert verdeutlicht, dass sich neben der Aufweitung des Laserspektrums die Impedanz des Lasers stabilisiert haben muss. Die Stabilisierung der Phasenlage des gesendeten optischen Signals ist eine Problematik aller laserbasierten Laufzeitentfernungsmessverfahren. Der Einsatz der PMD-Technologie auf der Empfängerseite bietet jedoch einen entscheidenden Vorteil, der im zweikanaligen differentiellen Aufbau zu finden ist. Dieser führt dazu, dass (wie in **Kap. 3.5** beschrieben) die gesamte empfangene Signalleistung zum Mischergebnis beiträgt. In Abhängigkeit zur Phasenlage wird ein einkanaliger Aufbau immer einen Teil des Photostrom unterdrücken. Dies lässt sich sehr einfach dadurch veranschaulichen, dass man den Prozess des Mischens als eine mathematische Multiplikation ausdrückt:

$$g(t) = s_1(t) \cdot s_2(t).$$

Gl. 4. 123

a)



b)

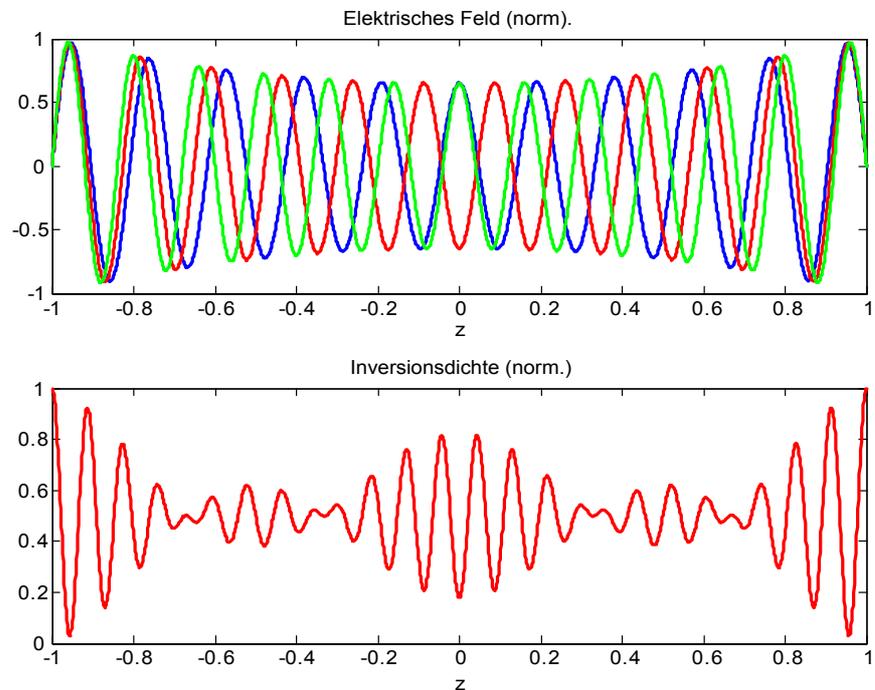


Abb. 4.35: Simulation einer stehenden Welle und der Inversionsdichte für eine konstante Stromdichte J in a) einem Singlemode-Laser und b) im Drei-Mode-Betrieb

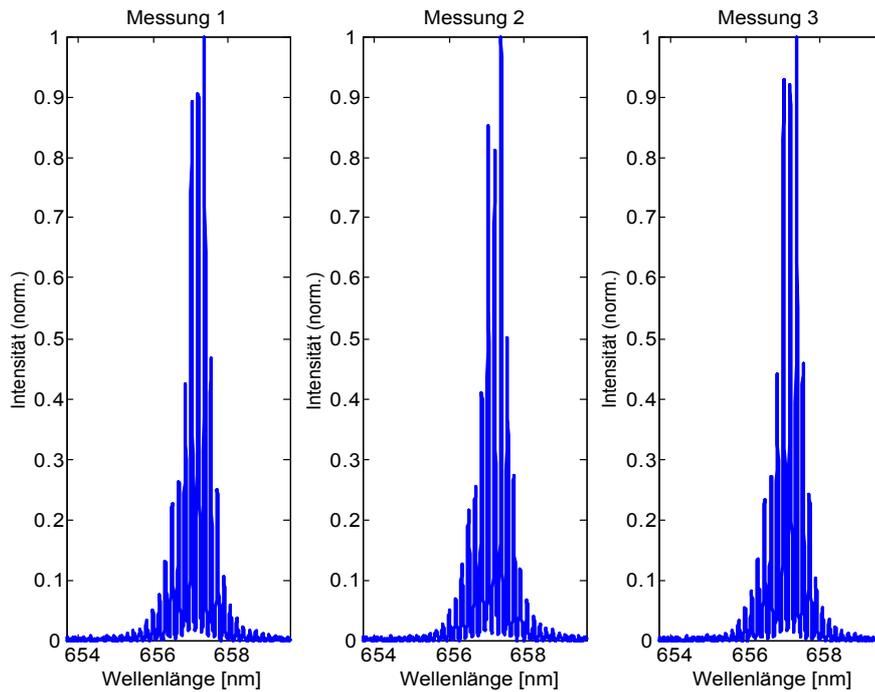


Abb. 4.36: Gemessenes Spektrum einer intensitätsmodulierten Singlemodelaserdiode bei einer Überlagerung von 15 und 460MHz

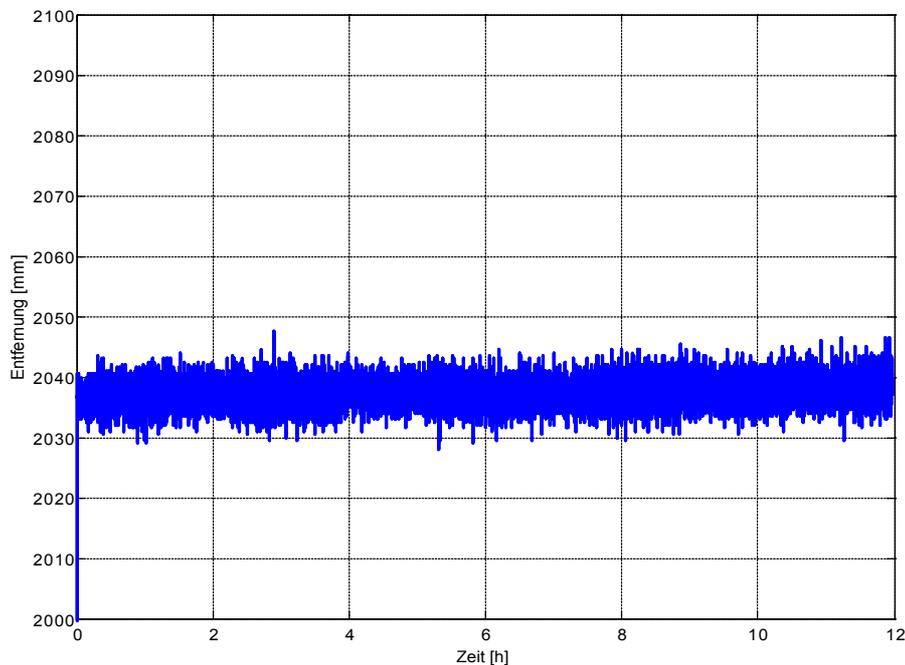


Abb. 4.37: Langzeitmessung einer intensitätsmodulierten Laser-TOF-Entfernungsmessung mit zusätzlicher HF-Modulation ($f_{mod} = 16\text{MHz}/460\text{MHz}$, $P_{opt} = 1\text{mW}$, $f_{mess} = 12,5\text{Hz}$)

Nimmt man für die miteinander zu mischenden Signale s_1 und s_2 jeweils eine Sinusschwingung an, dann gilt für das Mischergebnis $g(t)$ unter Berücksichtigung einer Phasenverschiebung τ .

$$g(t) = \sin(\omega(t + \tau)) \cdot \sin(\omega t). \quad \text{Gl. 4. 124}$$

Das Ergebnis $g(t)$ ist immer dann null, wenn einer der beiden Multiplikatoren zu null wird. Im schlechtesten Fall kann die Phasenverschiebung τ dazu führen, dass genau im Maximum des empfangenen Signals s_1 eine Multiplikation mit Null erfolgt. Das führt somit zu einer Reduktion der empfangenen Signalleistung und verschlechtert die erreichbare Messgenauigkeit. Durch den zweikanaligen Aufbau der PMD-Strukturen hingegen erfolgt für den Fall der Multiplikation des empfangenen Signals mit Null gleichzeitig eine Multiplikation mit Eins im Nachbarkanal. So wird unabhängig von der Phasenverschiebung τ immer die komplette Empfangsleistung durch die nachfolgende Signalverarbeitung verwertet, worin deutlich eine Stärke der PMD-Technologie zu sehen ist.

Eine weitere anspruchsvolle Aufgabe bei der Realisierung einer 1D-Entfernungsmessung ist das Erreichen einer möglichst geringen Messunsicherheit. Hierzu wurden neben verschiedenen schaltungstechnischen Maßnahmen die in **Kap. 4.3.3** vorgestellten Subsampling-Methoden umgesetzt und getestet. Eine Gegenüberstellung der Rauschunterdrückung mittels Mehrfachabtastung mit einfacher Summenbildung und Geradenapproximation zeigt **Abb. 4.38**. Deutlich ist zu erkennen, dass die Geradenapproximation gegenüber der einfachen Summenbildung geringere Messunsicherheiten erreicht. Die Ursache liegt in der unterschiedlichen Unterdrückung korrelierter Rauschanteile, die beispielsweise durch das kTC-Rauschen der Integrationskapazität, durch Änderungen in der Modulationsfrequenz oder durch verrauschte Spannungsquellen hervorgerufen werden.

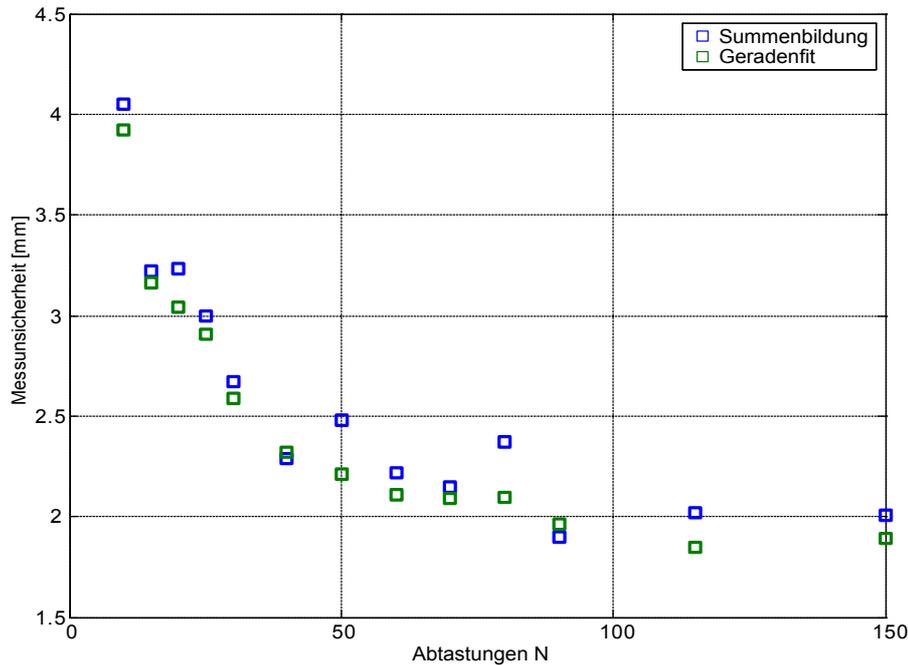


Abb. 4.38: Verringerung der Messunsicherheit der Entfernungsmessung durch verschiedene Subsampling-Methoden und Anzahlen N an Abtastungen

4.6.2 Zweidimensionale TOF-Entfernungsmessung und deren Informationsgewinn

Die vollständige Aufnahme einer realen dreidimensionalen Szene ist trotz der enormen Entwicklungen im Bereich derameratechnik immer noch eine der anspruchsvollsten Aufgaben. Die Ursache findet man darin, dass durch die optische Abbildung der realen Situation auf einen ebenen Bildaufnehmer eine wesentliche Information, die der Tiefe, verloren geht. Einzige verbleibende Information ist der zweidimensionale Intensitätsverlauf der beobachteten Szenerie. In vielen Situationen ist diese Information völlig ungenügend, da sich einige Objekte nicht entscheidend in ihrer Intensität von der Umgebung abheben. Dies stellt die heutige 2D-Bilddatenverarbeitung immer noch vor erhebliche Probleme, so dass ein erheblicher Aufwand in die Ausleuchtung einer Bildszene investiert werden muss. Ein universeller Einsatz eines solchen Systems in allen Umgebungen ist jedoch in keinem

Falle realisierbar. Würde jedoch die verloren gegangene Information der Tiefe noch zur Verfügung stehen, dann ließen sich auch kontrastarme Bereiche der beobachteten Szene wirklichkeitsgetreu beschreiben.

Es existieren viele unterschiedliche Ansätze dem Verlust der Tiefeninformation zu entgehen, wobei sich diese heute prinzipiell in folgende zwei Gruppen einteilen lassen:

- **Korrespondenzsysteme:** Hier werden korrespondierenden Merkmale in mehreren 2D-Bildern gesucht und derart ausgewertet, dass sich Entfernungswerte berechnen lassen. Diese Gruppe beinhaltet also alle Systeme, welche auf herkömmlichen 2D-Bildaufnehmern beruhen, wie z.B. Stereovision (Auswertung eines 2D-Bildpaares) und Optischer Fluss (Auswertung einer 2D-Bildfolge). Aber auch Lichtschnittverfahren können hierzu gerechnet werden, da hier letztlich Korrespondenzen zwischen der bekannten Form des Lichtgitters und dessen Abbildung auf die Bildszene ausgewertet werden.
- **Time-of-Fligh-Systeme:** Diese Verfahren unterscheiden sich ganz wesentlich dadurch, dass sie jeden Bildpunkt in der Tiefe tatsächlich messen. Hierzu wird die Laufzeit eines gesendeten Signals bzw. dessen Reflektion vom beobachteten Objekt ermittelt, welche die Entfernung zwischen Kamera und Beobachtungspunkt repräsentiert.

Die Diskussion dieser zwei grundlegenden Ansätze besitzt sehr interessante Aspekte, da sich hier schon ganz markante Vor- und Nachteile formulieren lassen, ohne zu sehr auf die einzelnen technischen Herausforderungen eines jeden konkreten Ansatzes einzugehen.

a) Informationsverarbeitung eines Korrespondenzansatzes

Wie zuvor beschrieben leiden alle auf herkömmlichen 2D-Bildaufnehmern basierende Verfahren am Verlust der Tiefeninformation. Dieser Abbildungsproblematik unterliegen so auch alle Korrespondenzansätze, da sie auf dieser weit verbreiteten Technik aufsetzen. Als Informationsbasis steht ihnen somit nur der aufgenommene Intensitätsverlauf zur Verfügung, welcher sie nur noch bezüglich bestimmter Merkmale interpretieren lässt.

Mit der optischen Abbildung geht jedoch noch eine weitere Einschränkung einher, die sich durch die Blendenwirkung einer jeden Optik erklärt. Die so genannte Blendenproblematik führt dazu, dass mitunter Objekte nicht vollständig erkannt werden sondern nur Objektausschnitte. Die Auswirkung wird besonders deutlich, wenn der Bildausschnitt nur noch den Blick auf eine Kante des beobachteten Objektes zulässt. Dann lässt sich die Bewegung des Objektes nicht mehr eindeutig bestimmen. Das lässt sich dadurch begründen, dass nur noch Veränderungen orthogonal zur Kante detektiert werden. Eine Bewegung entlang der beobachteten Kante ist aufgrund fehlender Intensitätsänderungen nicht mehr zu erkennen. Jedoch nicht nur die optische Abbildung auf den Bildaufnehmer begrenzt die Möglichkeiten der vollständigen Beobachtung der tatsächlichen 3D-Bewegtbildszene. Auch die Abtastungsproblematik des Bildaufnehmers in lateraler und zeitlicher Dimension stellt eine Beschränkung dar, welche zu kleine Objektstrukturen und zu schnelle Bewegungen nicht mehr erkennen lässt.

Alle Korrespondenzansätze stellen nun den Versuch dar, durch geeignete Algorithmen Bildpaare oder -folgen so zu interpretieren, dass über die Suche vorhandener Übereinstimmungen die verlorene Tiefe rekonstruiert wird. Der nun zur Verfügung stehende Intensitätsfluss repräsentiert jedoch den beobachteten Bewegungsfluss der Objekte nur sehr unzureichend. Besonders deutlich wird diese Divergenz, wenn man bedenkt, dass eine Änderung im Intensitätsbild nicht zwangsläufig seine Ursache in einem bewegten Objekt haben muss, sondern hierfür ebenso einfache Änderungen in der Beleuchtungssituation in Frage kommen. Deutlicher formuliert kann nicht jede Intensitätsänderung auf ein Objekt und seine Bewegung zurückgeführt werden. Aber auch der umgekehrte Fall lässt keine Eindeutigkeit zu, denn nicht jedes Objekt und seine Bewegung führt zwangsläufig zu Änderungen im Intensitätsbild. Alle Korrespondenzsysteme unterliegen dieser Divergenz zwischen dem tatsächlichen Bewegungsfluss der Objekte und dem Intensitätsfluss der aufgenommenen Bilder. Sie stellt also das Hauptproblem des Korrespondenzansatzes dar und lässt eine sichere Detektion eines Objektes in allen Raumdimensionen zumindest als sehr fraglich erscheinen. Der ungünstigste Fall eines schwarzen Objektes vor schwarzem Hintergrund kann von einem solchen System nicht bewältigt werden, aber auch ein Objekt beliebiger Intensität kann vor einem gleichfarbigen Hintergrund nur noch über den eventuell sichtbaren Schattenwurf erkannt werden. Ein Algorithmus zur Auswertung von Korrespondenzen muss also die vorhandenen Intensitätsänderungen durch unterschiedliche Beleuchtungssituationen, durch

den Schattenwurf der beobachteten Objekte und durch die Objekte selbst sicher auswerten, was zwangsläufig einen erheblichen Rechenaufwand mit sich bringt.

Aus der Abtastungsproblematik des optischen Empfängers resultiert jedoch noch eine weitere Schwierigkeit, da sich auch bei ähnlichen Objekten Korrespondenzen finden lassen, die eine falsche Interpretation der Orts- oder Bewegungsrichtung zulassen. Existiert beispielsweise ein periodischer Grauwertverlauf, dann ist dessen Verschiebung um eine einzelne Periodendauer nicht von deren Vielfachen unterscheidbar. Aber auch Bewegungen eines Objektes die nicht nur linearen Charakter sondern ebenso eine rotierende Komponente besitzen, erschweren die Suche nach Korrespondenzen, da in zwei Bildern zwar das selbe Objekt beobachtet wird, jedoch aus unterschiedlichen Richtungen. All diese Schwierigkeiten kann man unter dem Begriff Korrespondenzproblem zusammenfassen, welches ebenso wie das Divergenzproblem eine sichere Rekonstruktion des tatsächlichen Bewegungsflusses in Frage stellt.

Da die Rekonstruktion des tatsächlichen Bewegungsflusses auf die Auswertung von Intensitätsunterschieden im Bildfluss angewiesen ist, können Entfernungs- oder Bewegungsinformationen nur in Bildbereichen mit entsprechenden Kontrastunterschieden ermittelt werden. Umgekehrt betrachtet lassen sich in Gebieten mit schwachen Intensitätsunterschieden keine oder nur wenig Entfernungs- bzw. Bewegungsinformationen berechnen. Bei großen Intensitätsunterschieden verhält es sich exakt umgekehrt. Hier können viele Informationen über das tatsächliche Bewegungsbild rekonstruiert werden. Letztlich hat man somit in allen Korrespondenzansätzen das Problem, dass man keine gleichmäßig über die beobachtete Szene verteilte Informationsdichte an rekonstruierter Tiefeninformation erhält. Diese Verteilungsproblematik ist eine weitere Unzulänglichkeit des Korrespondenzansatzes. Eine schematische Übersicht des beschriebenen Informationsflusses eines Korrespondenzansatzes mit allen auftretenden Problemen zeigt **Abb. 4.39**.

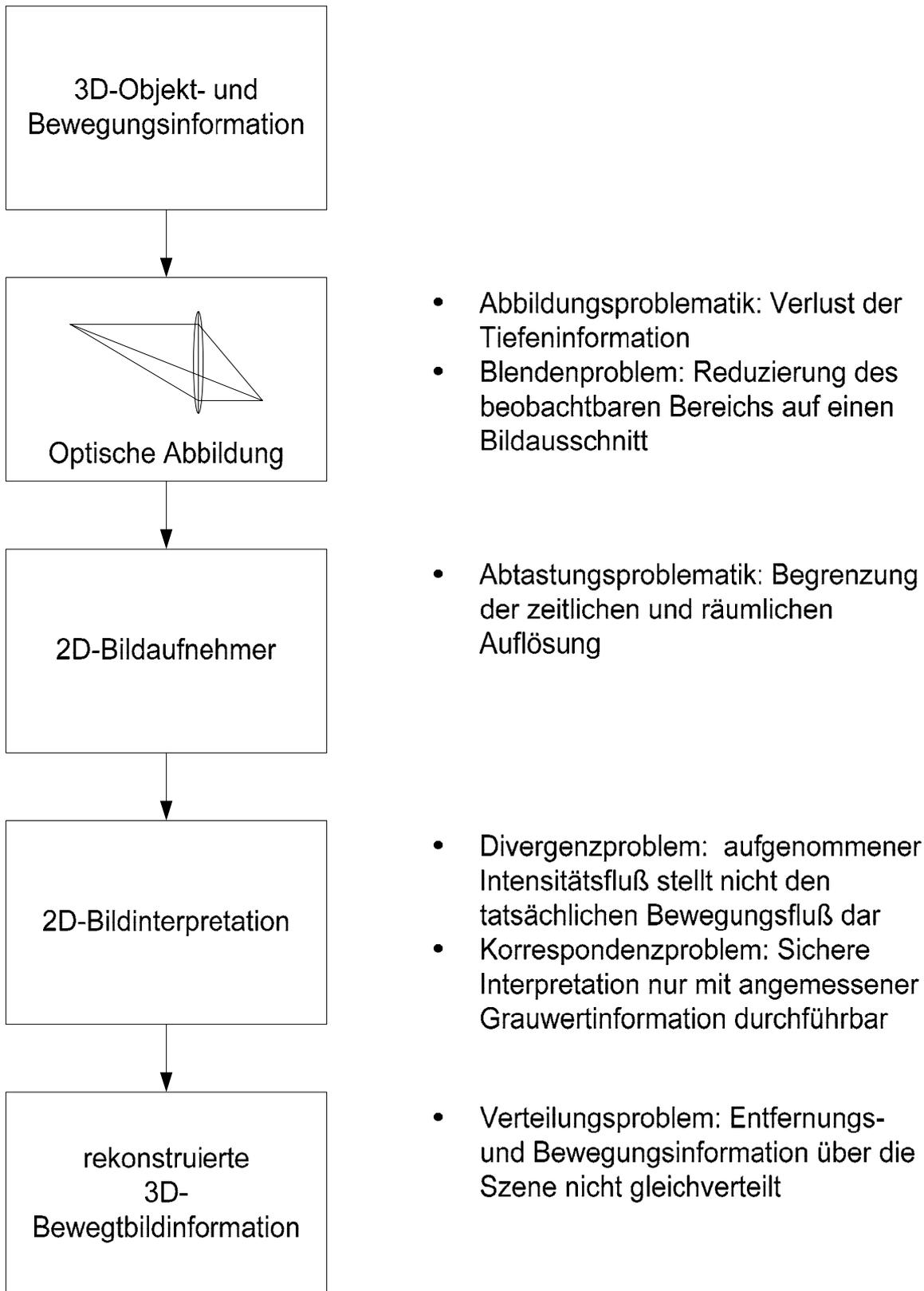


Abb. 4.39: Schematische Übersicht der Informationsverarbeitung eines Korrespondenzansatzes zur 3D-Bewegtbilderfassung

b) Informationsverarbeitung eines TOF-Ansatzes

Wie bei den Korrespondenzverfahren wird bei TOF-Ansätzen eine Abbildungsoptik verwendet, so dass der Schluss nahe liegt, dass hier die vergleichbaren Probleme auftreten. Die Einschränkung des beobachtbaren Bereichs durch die Blendenproblematik bleibt natürlich auch beim TOF-Ansatz erhalten, jedoch geht mit diesen Verfahren die Tiefeninformation durch die Laufzeitmessung nicht verloren. Der zur Detektion notwendige Bildaufnehmer begrenzt in Verbindung mit der Empfangsoptik erwartungsgemäß auch den TOF-Ansatz in der räumlichen und zeitlichen Auflösung, so dass diese Systeme ebenso der Abtastungsproblematik unterliegen. Da jedoch aufgrund der TOF-Messung die Tiefeninformation erhalten bleibt, entfallen hier die Divergenz-, Korrespondenz- und Verteilungsproblematik, denn eine Bildinterpretation bzw. Auswertung von Bildkorrespondenzen ist hier nicht mehr notwendig. Eine schematische Übersicht der Informationsverarbeitung eines TOF-Ansatzes zeigt **Abb. 4.40**.

c) Vergleich der Informationsverarbeitung beider Ansätze

Vergleicht man beide Schemata, dann wird sehr schnell deutlich, dass ein TOF-Ansatz deutlich weniger Problemen unterliegt als alle Korrespondenzansätze. Im Moment kann man noch insoweit für einen Korrespondenzansatz argumentieren, dass hierfür deutlich höher auflösende Bildaufnehmer erhältlich sind. Jedoch ist hierbei zu bedenken, dass durch die Auswertung von vorhandenen Korrespondenzen keine Entfernungs- oder Bewegungsinformationen in dieser Auflösung rekonstruiert werden können. Anders ausgedrückt besitzt eine dreidimensionale Beschreibung der beobachteten Szene durch ein Korrespondenzverfahren eine deutlich geringere Auflösung als der des verwendeten Bildaufnehmers. Dies relativiert den Vorteil einer hohen Bildauflösung deutlich, da TOF-Ansätze die dreidimensionale Information exakt in der Auflösung des Bildaufnehmers liefern. Des Weiteren kann man davon ausgehen, dass die Auflösung von TOF-Bildaufnehmern zukünftig noch anwachsen wird, da ihre Entwicklung noch in den Anfängen steht und alle technologischen Möglichkeiten sicher noch nicht genutzt wurden.

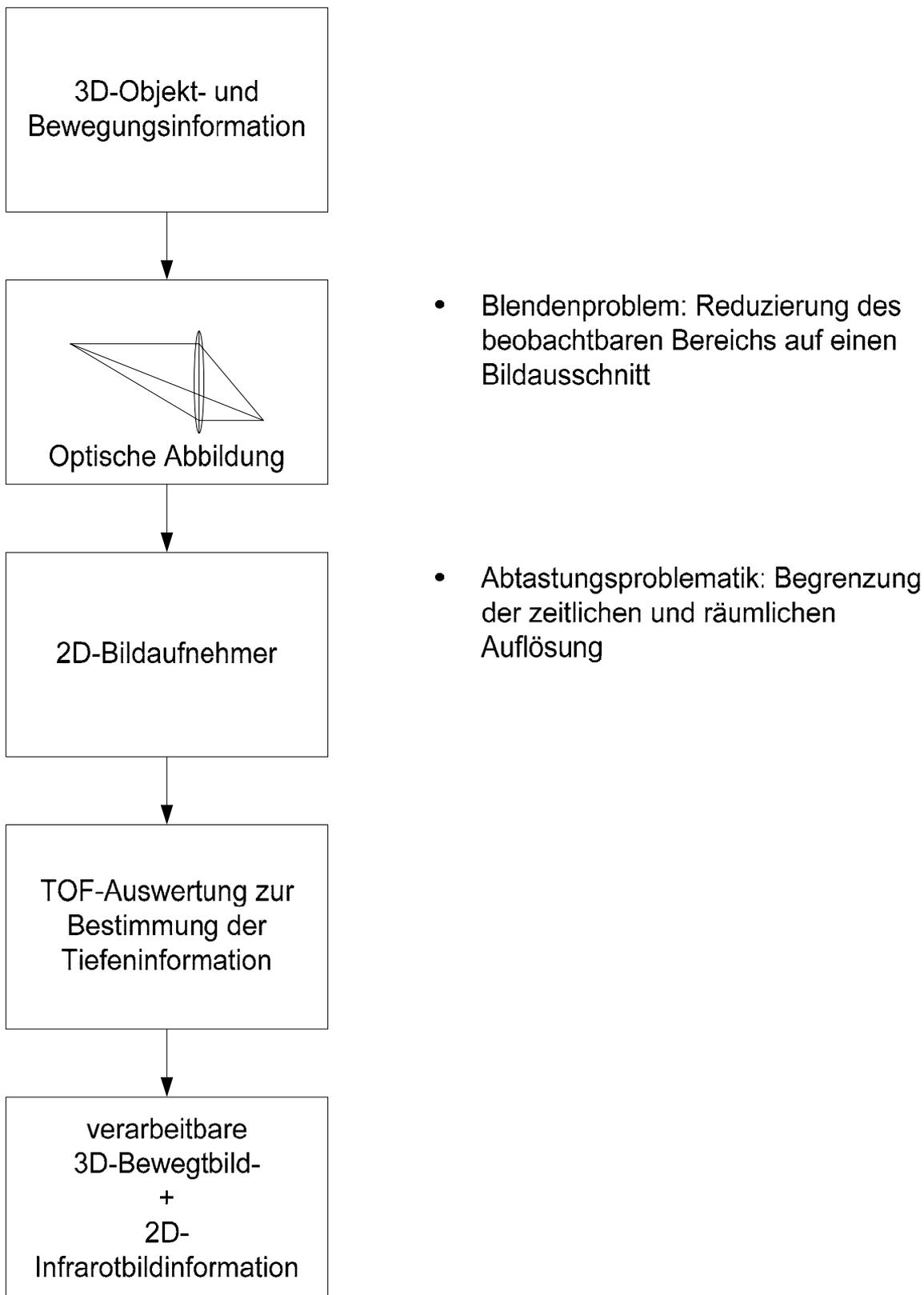


Abb. 4.40: Schematische Übersicht der Informationsverarbeitung eines TOF-Ansatzes zur 3D-Bewegtbilderfassung

Mit der angestoßenen Entwicklung kostengünstiger TOF-Bildaufnehmer und der damit verbundenen Entwicklung von 3D-Kamerasystemen entwickeln sich nun auch wesentlich komplexere Systeme, welche sich nicht mehr durch die eindimensionale Betrachtung der klassischen Systemtheorie vollständig beschreiben lassen. Schon die klassische 2D-Bewegtbildverarbeitung stellt ein dreidimensionales System dar (x, y, t) . Mit der Verwendung einer 3D-TOF-Kamera jedoch bildet sich nun sogar ein vierdimensionales System heraus (x, y, z, t) , was erhöhte Anforderungen an eine optimale Datenverarbeitung bzw. Algorithmik stellt.

Abb. 4.41 zeigt einen Technologiedemonstrator einer PMD-basierten 3D-TOF-Kamera für den industriellen Einsatz. Mit einer lateralen Auflösung von 64x48 Pixeln und einer Entfernungunsicherheit von wenigen Zentimetern können hiermit schon zahlreiche Applikationen angegangen werden. Im nachfolgenden Teil dieser Arbeit sollen auf Basis dieses Demonstrators nun grundsätzliche systemtheoretische Betrachtungen zur 3D-Bild- und 3D-Bewegtbildverarbeitung angestellt werden.

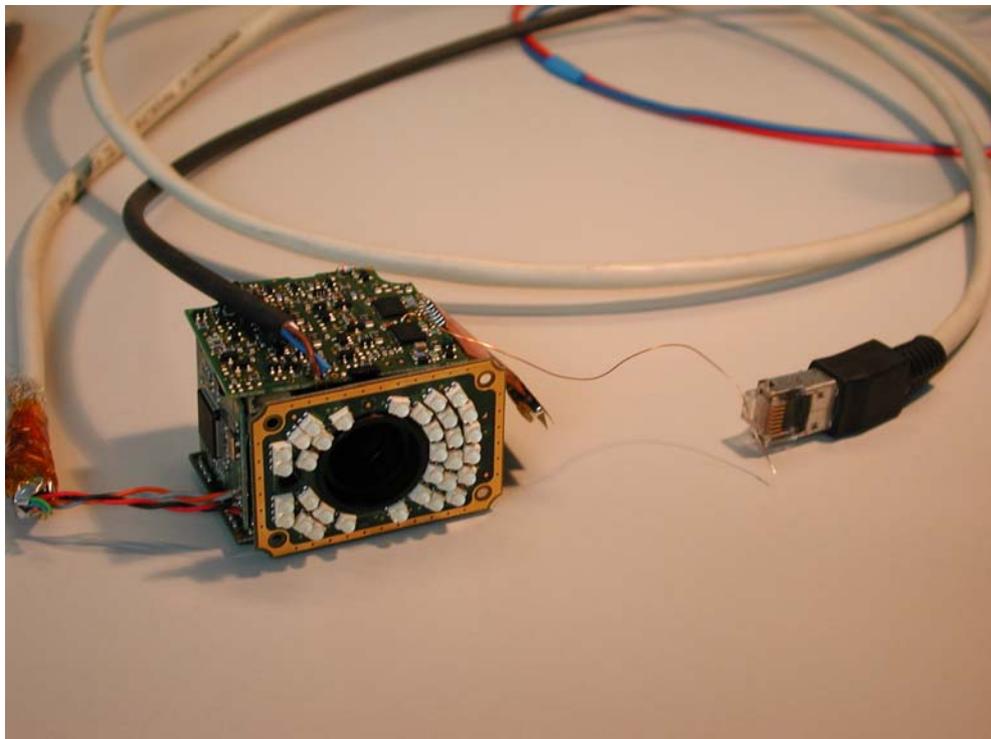


Abb. 4.41: Technologiedemonstrator einer PMD-basierten 3D-TOF-Kamera (64x48 Pixel) für den industriellen Einsatz

4.6.3 Objekterkennung in der interferometrischen 3D-Messtechnik

Neben den zuvor beschriebenen eindimensionalen Systemen bietet die PMD-Technologie auch die Möglichkeit, mehrere Pixel als Zeile oder Matrix anzuordnen. Mit einem solchen mehrdimensionalen System können mehrere optische Signale parallel detektieren und ausgewertet werden. Benutzt man eine solche PMD-Matrix im Bereich der optischen Entfernungsmessung, dann erhält man eine dreidimensional messende Kamera, mit der sich Objekte geometrisch erfassen und vermessen lassen. Die Qualität einer optischen Erkennung von Objekten lässt sich mittels interferometrischer 3D-Messtechnik stark verbessern, was dadurch erklärt werden kann, dass gegenüber der herkömmlichen 2D-Bilder eine 3D-Bildinformation einen wesentlich höheren Informationsgehalt aufweist. Da der 2D-Bildverarbeitung nur Farbinformationen in x- und y-Koordinaten zur Verfügung stehen, ist sie für eine Objekterkennung auf möglichst starke Kontraste angewiesen. Der 3D-Bildverarbeitung steht jedoch zusätzlich die Entfernungsinformation zur Verfügung, so dass eine sichere Erkennung auch gänzlich ohne Kontrast möglich ist. Hierzu kann ähnlich zur 2D-Bildverarbeitung die Erkennung und Bestimmung der Position eines Objektes durch Korrelation erfolgen. Hierzu ist der Entfernungsdatensatz $z(x,y)$ mittels zweidimensionaler Fourier-Transformation in den Ortsfrequenzraum abzubilden. Bildet man die Kontur $z_K(x,y)$ des gesuchten Objektes ebenso in den Ortsfrequenzraum ab, dann vereinfacht sich das Erkennen mittels Korrelation im Ortsraum zu einer einfachen punktweisen Multiplikation dieser beiden Abbildungen:

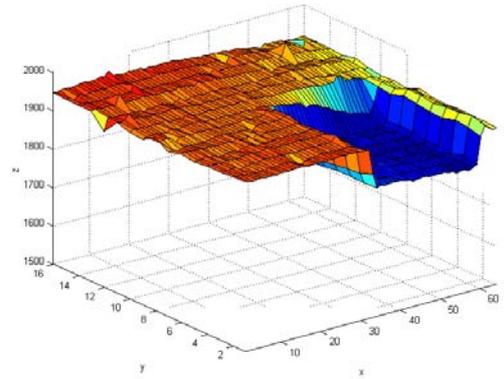
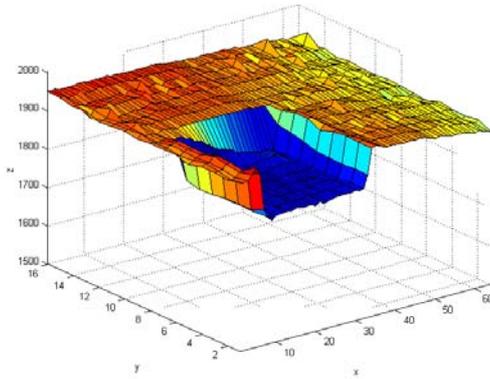
$$\rho(x,y) = F^{-1} \left\{ F(z(x,y)) \cdot F(z_K(x,y))^* \right\}. \quad \text{Gl. 4. 125}$$

Erreicht die Korrelation einen bestimmten Schwellwert, ist das gesuchte Objekt gefunden. Die Lage des Maximums innerhalb des Korrelogramms liefert direkt dessen Position. Ändert sich der Entfernungsdatensatz $z(x,y)$, dann verschiebt sich auch das Maximum innerhalb des Korrelogramms $\rho(x,y)$. **Abb. 4.42** verdeutlicht diesen Zusammenhang für einen Versuch zur Detektion eines Quaders innerhalb eines 3D-Datensatzes. Deutlich hebt sich hier in den berechneten Korrelogrammen das Zentrum des Würfels heraus.

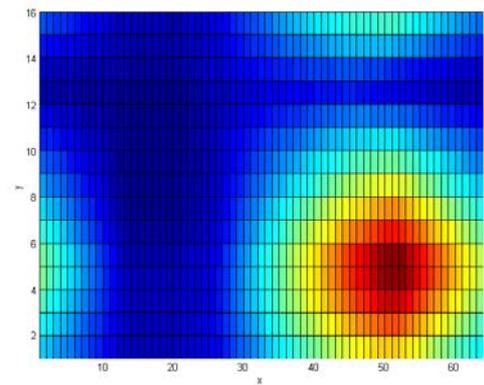
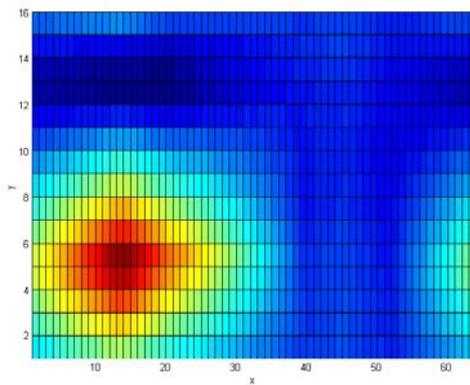
Position links

Position rechts

a) Entfernungswerte



b) Korrelogramm



c) Grauwerte

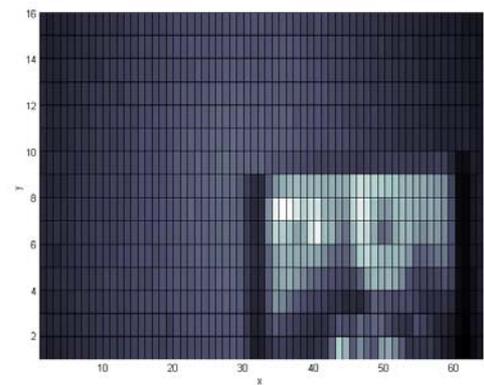
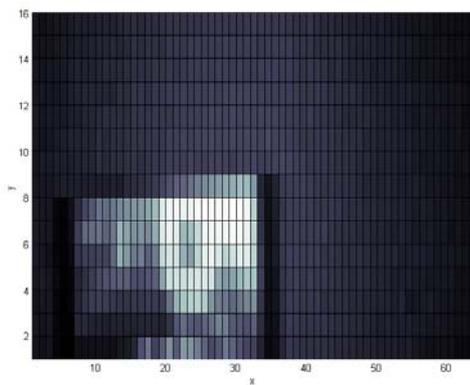


Abb. 4.42: Positionsbestimmung eines Würfels mittels zweidimensionaler Fourier-Transformation

Die mehrdimensionale Fourier-Transformation bietet im Bereich der 3D-Kameratechnik jedoch noch wesentlich mehr Möglichkeiten. Weitet man die Transformation neben den Entfernungswerten auch auf die Grauwerte $g(x,y,z)$ aus, dann lassen sich nicht nur Objekte einer bestimmten Geometrie detektieren bzw. von der Umgebung unterscheiden. Es besteht zusätzlich die Möglichkeit, auch Objekte gleicher Geometrie über ihre Farbe zu trennen. Rechnerisch ist hierzu die Fourier-Transformation auf drei Dimensionen aufzuweiten:

$$\rho(x, y, z) = F^{-1} \left\{ F(g(x, y, z)) \cdot F(g_K(x, y, z))^* \right\}. \quad \text{Gl. 4. 126}$$

Dadurch kann mit einer einzigen Multiplikation auch eine gesuchte Geometrie mit einem bestimmten Farbverlauf von der Umgebung getrennt werden.

Neben der beschriebenen Objekterkennung und Positionsbestimmung lassen sich mit der mehrdimensionalen Fourier-Transformation aber auch Objekte verschiedener Geschwindigkeiten unterscheiden. Hierzu sind die Fourier-Transformierten einer Bildfolge auszuwerten, was eine Erweiterung um die Dimension der Zeit zur Folge hat.¹⁶

4.6.4 Zeitaufgelöste Spektroskopie

Eine weitere interessante Applikation der Phasenmessung optischer inkohärenter Signale findet sich im Bereich der zeitaufgelösten Spektroskopie. Sie bietet neben der spektralen Vermessung eines optischen Signals die Möglichkeit, zugehörige Abklingzeiten zu bestimmen. Nutzt man den Prozess der laserinduzierten Fluoreszenz, dann lassen sich das Vorhandensein von fluoreszierenden Stoffen bis hin zu einzelnen Molekülen nachweisen. Der im Folgenden vorgestellte Messaufbau dient jedoch nicht dieser anspruchsvollen Aufgabe des Einzelmolekülnachweises, sondern soll die Eignung der PMD-Technologie belegen, solche Messungen auch zeitlich aufzulösen.

Der Vorgang der Fluoreszenz lässt sich im Allgemeinen als zweistufiger Prozess beschreiben, indem durch die Absorption einer definierten Lichteinstrahlung eine

¹⁶ Jähne, B. (1993b); Spatio-Temporal Image Processing; Lecture Notes in Computer Science; Springer-Verlag, Heidelberg. No. 751, S. 92 f.

Anregung erfolgt, welche nach einer gewissen Zeit wieder durch spontane Emission abgebaut wird. Diese spontane Emission wird auch als Fluoreszenz bezeichnet. Sie hängt von verschiedenen Parametern wie der Besetzung des Grundzustandes, dem Druck, der Temperatur usw. ab. Die Wechselwirkungen zwischen fluoreszierenden Molekülen und ihrer Umgebung beeinflussen die spektrale Verteilung als auch die Abklingdauer und Polarisations-eigenschaften der Fluoreszenz. Dieser Mechanismus wird in vielen messtechnischen Aufgaben der Grundlagenforschung wie in der Genetik oder Quantendynamik genutzt. Aber auch in den praxisorientierten Wissenschaften wie Medizin-, Halbleiter- und Umwelttechnik hat die Fluoreszenz ihre Bedeutung. Die Beschreibung der Fluoreszenz im sogenannten Zwei-Niveau-Modell beruht auf einer quantenmechanischen Betrachtung unter der vereinfachenden Annahme, dass das betrachtete System in nur zwei Zuständen vorliegen kann (**Abb. 4.43**).

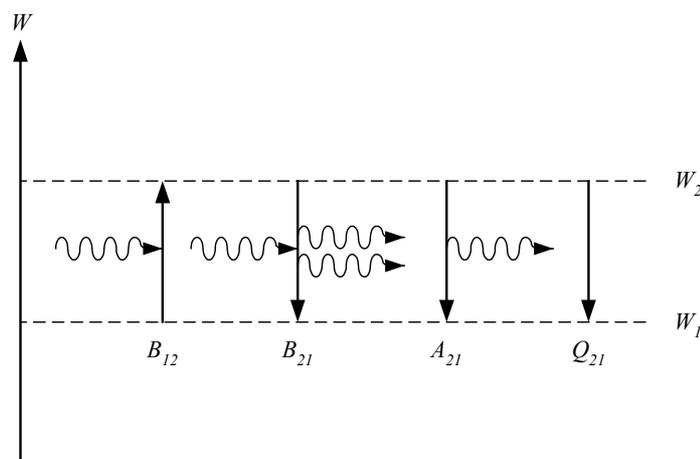


Abb. 4.43: Mögliche Wechselwirkungen im Zwei-Niveau-Modell

Werden in ein solches System Laserphotonen mit einer konstanten Intensität eingestrahlt, dann können sich folgende Prozesse einstellen:

- Absorption eines einfallenden Photons. Ein Elektron geht hierbei vom Grund- in den angeregten Zustand über (B_{12}).
- Stimulierte Emission eines Photons durch ein einfallendes Photon. Hier wechselt ein Elektron vom angeregten in den Grundzustand (B_{21}).

- Spontane Emission (Fluoreszenz) eines Photons: Ein Elektron wechselt mit einer gewissen Wahrscheinlichkeit spontan vom angeregten in den Grundzustand. (A_{21})
- Stoßlöschung/Quenching: Über Stoßprozesse verliert das angeregte Elektron seine Energie und wechselt so strahlungslos in den Grundzustand (Q_{21}). (Ein isoliertes Teilchen im idealen Vakuum kann keinen Quenchprozess eingehen.)

Des Weiteren kann ein angeregtes Elektron aufgrund chemischer Reaktionen dem Zwei-Niveau-Modell völlig entzogen werden:

- Prädissoziation: Zerfall der molekularen Struktur durch einen Übergang aus dem angeregten in einen nicht bindenden Zustand.
- Photoionisation: Durch erneute Anregung eines bereits angeregten Elektrons löst sich dieses vom gesamten Molekül, welches nun ionisiert erscheint.

Eine quantitative Aussage über das Fluoreszenzverhalten gelingt über die Geschwindigkeitskonstanten bzw. Raten der verschiedenen Übergänge. Vernachlässigt man die Prädissoziation und Photoionisation, dann gilt für die Fluoreszenzleistung S :

$$S \propto N^0 \cdot f_B(T) \cdot B_{12} \cdot I_L \cdot \frac{A_{21}}{A_{21} + Q_{21}} \quad \text{Gl. 4. 127}$$

Hier bezeichnet B_{12} die Rate der Absorption bzw. Anregung, A_{12} die Rate der spontanen Emission bzw. Fluoreszenz und Q_{21} die Stoßlösch- bzw. Quenchrage. Die Intensität der Laserstrahlung ist durch I_L gegeben. Die Teilchendichte N^0 der fluoreszierenden Probe geht ebenso wie ein temperaturabhängiger Boltzmann-Term $f_B(T)$ in diese Gleichung ein. Damit erlaubt die laserinduzierte Fluoreszenzmessung eine Ermittlung von Zustandsbesetzungen, den Teilchenkonzentrationen und der Temperatur.

Ein weiterer interessanter Aspekt ist die zeitliche Betrachtung der spontanen Emission im Sinne der Fluoreszenz. Wird ein System durch einen zeitlich definierten Impuls angeregt, dann liefert die zeitverzögerte fluoreszierende Antwort eine Aussage über die Ladungsträgerdynamik bzw. Energierelaxation innerhalb einer Probe. Mit einer solchen zeitaufgelösten Spektroskopie ist ein Zeitbereich von mehreren hundert Nanosekunden bis hinunter in den Femtosekundenbereich aufzulösen. In der nachfolgenden Beschreibung soll ein sehr einfacher Aufbau beschrieben werden, der die Eignung der PMD-Technologie für

diese Applikation präsentiert. Im Gegensatz zu der Mehrheit von lasergepulsten Verfahren wird hier mit einem kontinuierlich intensitätsmodulierten Laser gearbeitet. In Anlehnung an die zuvor beschriebene Entfernungsmessung wird hier ebenso mit einer Phaseshift-Technik gearbeitet, jedoch mit einer so feinen Phasenauflösung, dass die gesamte Kreuzkorrelation von Fluoreszenz und Laserstimulation ermittelt wird. Ein Vorteil dieser Technik ist der Verzicht auf die sonst üblichen Verstärkerschaltungen zur Detektion einer einzelnen Fluoreszenzantwort. Die Korrelationstechnik verzichtet hierauf und gewinnt dagegen eine ausreichende Signalintensität durch eine Integration mehrerer solcher Ereignisse. Selbst kleinste Signale (die sich vom Dunkelstrom abheben) können so durch eine Verlängerung der Integrationszeit detektiert werden. Die Verwendung eines PMD als quasi-optischen Korrelator besitzt weiterhin den entscheidenden Vorteil der differentiellen Auswertung einer Gegentaktmischung. Unabhängig von der Phasenlage zwischen Laserstimulanz und Fluoreszenz wird so die gesamte empfangene Signalleistung ausgewertet.

Der realisierte Demonstrationsaufbau (**Abb. 4.44**) beruht auf einem einzelnen PMD-Pixel, da hier nur der Nachweis der Machbarkeit einer Fluoreszenzlebensdauerermessung geführt werden soll. Eine gleichzeitige spektrale Messung der Fluoreszenz ist natürlich denkbar und könnte mittels einer PMD-Zeilenanordnung und einem Gitter auf klassischem Wege realisiert werden.

Eine vereinfachte schematische Darstellung des Versuchsaufbaus zeigt **Abb. 4.46**. Zentrales Element ist das PMD, welches sein Gegentaktsignal von einem DDS (Direct Digital Synthesizer) bezieht. Diese Synthesizer sind in der Lage, über eine sehr genaue Wertetabelle hochgenaue Frequenzen und Phasenverschiebungen zu generieren. In dieser Wertetabelle ist eine Sinusfunktion abgelegt, deren Werte nacheinander mit dem Systemtakt an einen internen Digital-Analog-Wandler geführt werden. Mit einem Tiefpassfilter, der dessen Analogausgang nachgeschaltet ist, kann die Qualität der Sinusschwingung noch weiter erhöht werden. Will man die Frequenz verändern, ist keine Änderung des Systemtakts nötig. Durch gezieltes Überspringen von Werten kann die Frequenz beispielsweise erhöht werden.

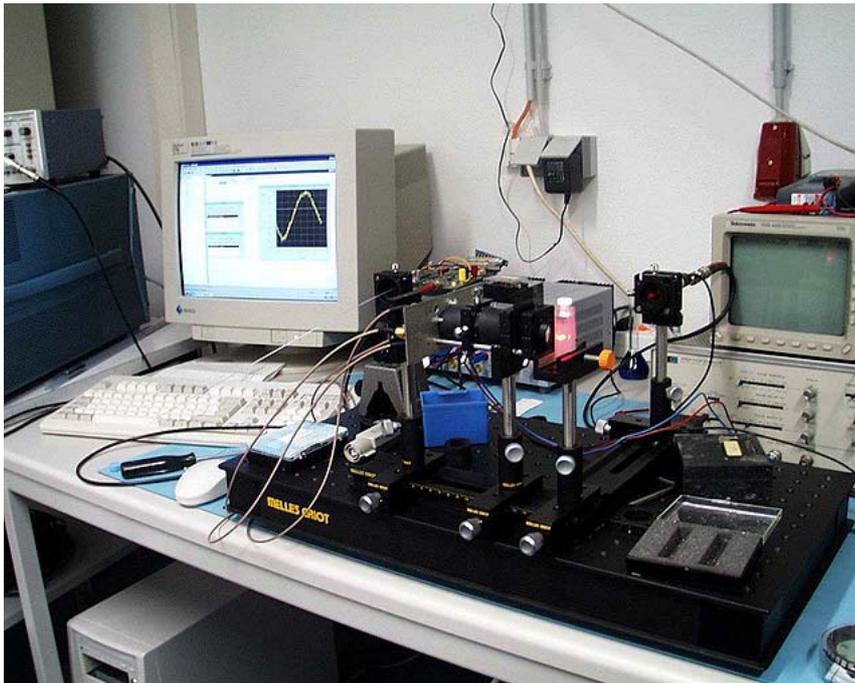


Abb. 4.44: Realisierter Versuchsaufbau zur zeitaufgelösten Spektroskopie mit Photogate-PMD, Halbleiterlaser ($\lambda=650nm$, 1mW), abbildender Optik, CPU, DDS und Ausgabe der gemessenen Korrelationsfunktion.

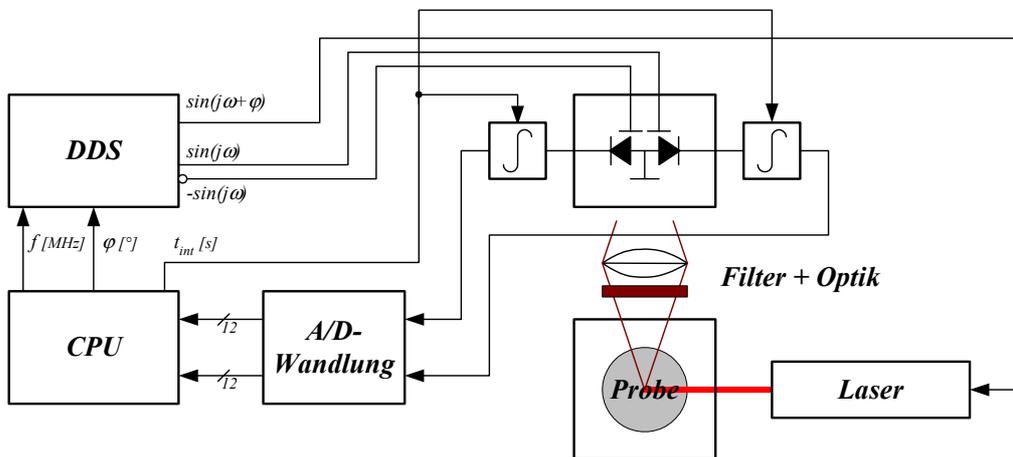


Abb. 4.45: Schematischer Aufbau des Versuchsaufbaus zur zeitaufgelösten Spektroskopie.

Die Phase zweier solcher Kanäle kann ebenso variiert werden. Hier wird eine gemeinsame Tabelle durch beide Kanäle versetzt abgearbeitet. Mit einer 24Bit-Darstellung gelingt so

eine Phasenauflösung von $2 \cdot 10^{-5}^\circ$. Das DDS des Demonstrationsaufbaus verschiebt wie beschrieben das Modulationssignal des Lasers

$$s_L(t) = \sin(\omega \cdot t) \quad \text{Gl. 4. 128}$$

gezielt zu dem am PMD anliegenden Gegentaktsignalen

$$s_{PMD,A}(t) = \sin(\omega \cdot t + \varphi) \quad \text{Gl. 4. 129}$$

und

$$s_{PMD,B}(t) = -\sin(\omega \cdot t + \varphi). \quad \text{Gl. 4. 130}$$

Das PMD mischt so die durch den Laser stimulierte Fluoreszenz $g(t-\tau)$ und liefert durch die nachfolgende Integration die Korrelation mit beiden Kanälen für eine Integrationszeit t_i .

$$\varphi_{sg,A}(\tau) = \int_0^{t_i} s_{PMD,A}(t) \cdot g(t + \tau) dt \quad \text{Gl. 4. 131}$$

bzw.

$$\varphi_{sg,B}(\tau) = \int_0^{t_i} s_{PMD,B}(t) \cdot g(t + \tau) dt. \quad \text{Gl. 4. 132}$$

Mit einer vollständigen Variation der vom DDS generierten zusätzlichen Phase φ von 0 bis 360° kann die komplette Korrelationsfunktion ermittelt werden. Vergleicht man diese mit der Korrelation einer nicht fluoreszierenden Referenzprobe, dann kann leicht die Fluoreszenzlebensdauer bestimmt werden (**Abb. 4.46**). Hierzu kann in der Nähe des Nulldurchgangs eine Gerade in die Messwerte approximiert werden, wozu sich ein Ansatz nach der Methode der kleinsten Quadrate anbietet. Mit dieser gelingt eine hochgenaue Bestimmung des Nulldurchgangs, welcher im Vergleich mit der Referenz einen genauen Messwert für die Fluoreszenzlebensdauer der Probe repräsentiert. Die Genauigkeit kann

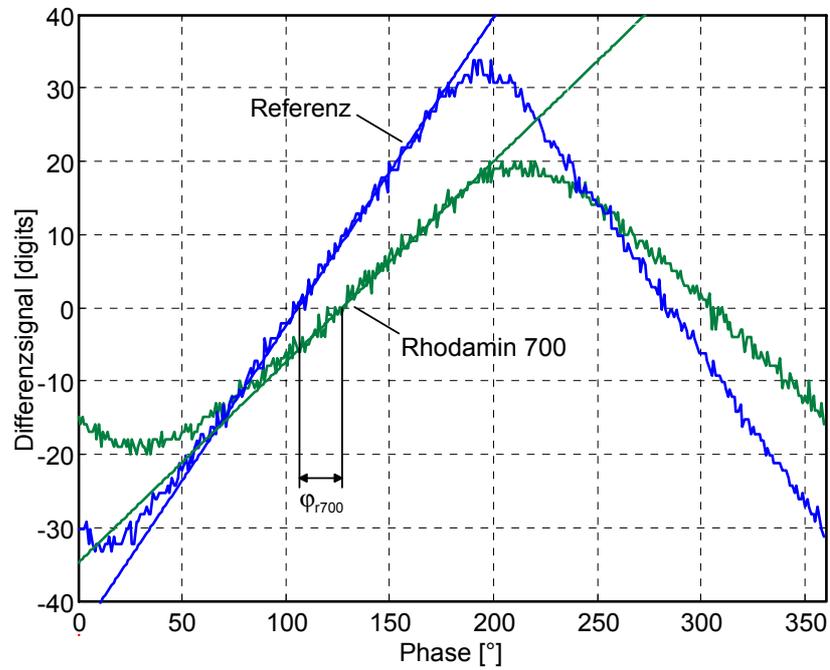
hier noch weiter erhöht werden, indem keine reinen Sinussignale, sondern Rechtecksignale miteinander korreliert werden. Die Korrelationsfunktion zweier idealer Rechtecksignale entspricht einer Dreiecksfunktion, welche jedoch durch die Tiefpassbegrenzung des Messaufbaus und der Proben in ihrem Maximum bzw. Minimum verschliffen werden. Der noch völlig ausreichend ausgeprägte lineare Bereich kann komplett in die Geradenapproximation mit einbezogen werden.

In der hier beschriebenen Versuchsanordnung konnte für die beiden Proben eine Fluoreszenzlebensdauer von $3,054ns$ (Rhodamin 700) und $2,289ns$ (Rhodamin 800) bestimmt werden. Literaturangaben liefern hierzu für verschiedene Rhodamine adäquate Werte zwischen $3ns$ und $8ns$.¹⁷ Diese Werte bestätigen die Eignung der PMD-Technologie für spektrometrische Anwendungen, insbesondere wenn der inhärente Mischer zu einer zusätzlichen zeitlichen Auflösung genutzt wird.

Die theoretische Auflösungsgrenze liegt wie bei allen optischen Verfahren im Photonenrauschen und kann über eine optische Leistungsbilanz ähnlich der Anwendung der optischen Entfernungsmessung (**Kap. 4.6.1**) bestimmt werden. Der wesentliche Unterschied findet sich in der empfangenen Leistung, die hier nicht von der Reflektivität eines Objektes, sondern von der Intensität der Fluoreszenz der Probe abhängt.

¹⁷ E. Thiel; Eigenschaften angeregter Rhodamin-Farbstoffe und deren Wirkung im Farblaser, Habilitation, Universität Siegen, Siegen 1997

a)



b)

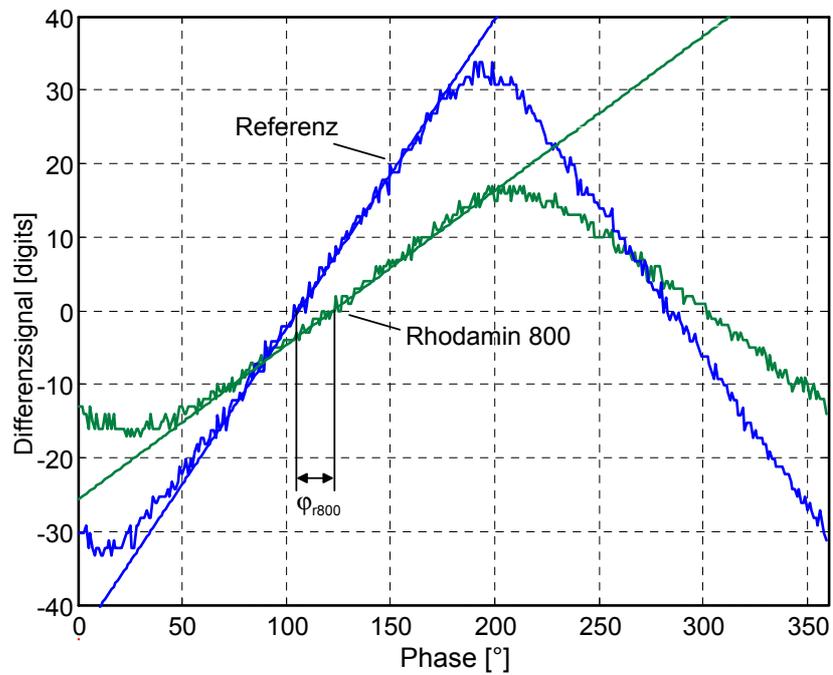


Abb. 4.46: Gemessene Korrelationsfunktionen für a) Rhodamin 700, b) Rhodamin 800 jeweils im Vergleich zu einer Referenzprobe.

4.7 In-Phase-Quadrature-Mischer als Konkurrenz zur klassischen Phase-Shift-Interferometrie

4.7.1 Grundprinzip der Phasenbestimmung mittels IQ-Demodulation

Ein wesentlicher Nachteil der Phase-Shift-Interferometrie findet sich in der sequentiellen Abtastung des Interferogramms. Dies erfordert neben mehreren Einzelmessungen immer einen Phasenschieber. Im klassischen nachrichtentechnischem Sinn handelt es sich hierbei meist um ein Zeitmultiplexverfahren, wobei jedoch auch ein Räummultiplexverfahren grundsätzlich realisierbar ist. Hierfür ist jedoch eine parallele Mischung des Empfangssignals mit mehreren Referenzsignalen notwendig, was einen nicht unerheblichen Schaltungsaufwand bedeutet. Aufgrund der Nachteile dieser beiden Verfahren wäre jedoch eine Lösung von großem Interesse, die den gewünschten Phasenwert des empfangenen Signals direkt mit nur einer einzigen Messung liefert. Eine solche Möglichkeit bietet die sogenannte IQ-Mischung (engl. „In-phase-Quadrature“), welche prinzipiell auf der gleichzeitigen Mischung mit nur zwei Referenzsignalen beruht, die um 90° zueinander phasenverschoben sind. In Kombination mit einer geeigneten Filterschaltung und Integration erhält man zwei Korrelationsergebnisse, die jeweils dem Imaginär- und Realteil des empfangenen Signals entsprechen (**Abb. 4.47**).

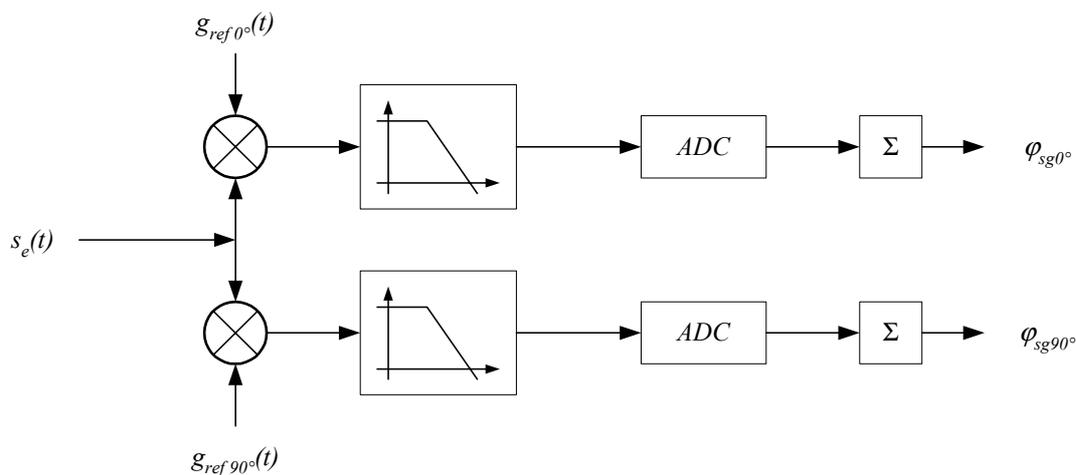


Abb. 4.47: Prinzip einer Phasenbestimmung mittels IQ-Mischung und anschließender Integration

Der Phasenwert lässt sich daraus sehr einfach mittels der Arcustangensfunktion ermitteln:

$$\begin{aligned}\varphi_s &= \arctan\left(\frac{\text{Im}}{\text{Re}}\right) \\ &= \arctan\left(\frac{\varphi_{sg90^\circ}}{\varphi_{sg0^\circ}}\right).\end{aligned}\tag{Gl. 4. 133}$$

Die Signalverläufe einer Phasenbestimmung mittels IQ-Demodulation zeigt **Abb. 4.48**. Neben dem Empfangssignal s , dessen Phase gesucht ist, werden hier die um 90° zueinander verschobenen Demodulationssignale $g_{demod\ 0^\circ}$ und $g_{demod\ 90^\circ}$ dargestellt. Das Ergebnis der Demodulation von s ergibt sich aus:

$$\begin{aligned}s \cdot g_{demod\ 0^\circ} &= \sin(\omega \cdot t + \varphi) \cdot \sin(\omega \cdot t) \\ &= \frac{1}{2}(\cos(\omega \cdot t + \varphi - \omega \cdot t) - \cos(\omega \cdot t + \varphi + \omega \cdot t)) \\ &= \frac{1}{2}(\cos(\varphi) - \cos(2\omega \cdot t + \varphi))\end{aligned}\tag{Gl. 4. 134}$$

bzw.

$$\begin{aligned}s \cdot g_{demod\ 90^\circ} &= \sin(\omega \cdot t + \varphi) \cdot \cos(\omega \cdot t) \\ &= \frac{1}{2}(\sin(\omega \cdot t + \varphi - \omega \cdot t) + \sin(\omega \cdot t + \varphi + \omega \cdot t)) \\ &= \frac{1}{2}(\sin(\varphi) - \sin(2\omega \cdot t + \varphi))\end{aligned}\tag{Gl. 4. 135}$$

Für beide Kanäle sind neben einem nur von der Phase φ abhängigen Term noch Anteile der doppelten Frequenz vorhanden. Diese lassen sich durch eine geeignete Filterschaltung eliminieren, wodurch sich Ausgangssignale reduzieren auf:

$$s \cdot g_{demod\ 0^\circ} = \frac{1}{2} \cos(\varphi)\tag{Gl. 4. 136}$$

bzw.

$$s \cdot g_{\text{demod } 90^\circ} = \frac{1}{2} \sin(\varphi). \quad \text{Gl. 4. 137}$$

Solchermaßen gefilterte Ausgangssignale für unterschiedliche Phasenwerte des Empfangssignals s zeigt **Abb. 4.49**. Entsprechend **Gl. 4. 133** lässt sich aus den Ausgangssignalen beider Filter die Signalphase direkt berechnen. Zu beachten ist hierbei jedoch, dass die Filtercharakteristik ebenso Einfluss auf den rein phasenabhängigen Term nehmen kann und so durchaus einen zusätzlichen systematischen Phasen-Offset verursacht. Da die Signalleistung an beiden Filterausgängen nicht nur von der Leistung des Empfangssignals s , sondern ebenso von der Signalphase φ abhängt, ist mit einem phasenwertabhängigen Rauschverhalten zu rechnen. Der Grund findet sich im Verlauf der Arcustangensfunktion, der für verschiedene Werte unterschiedliche Anstiege besitzt.

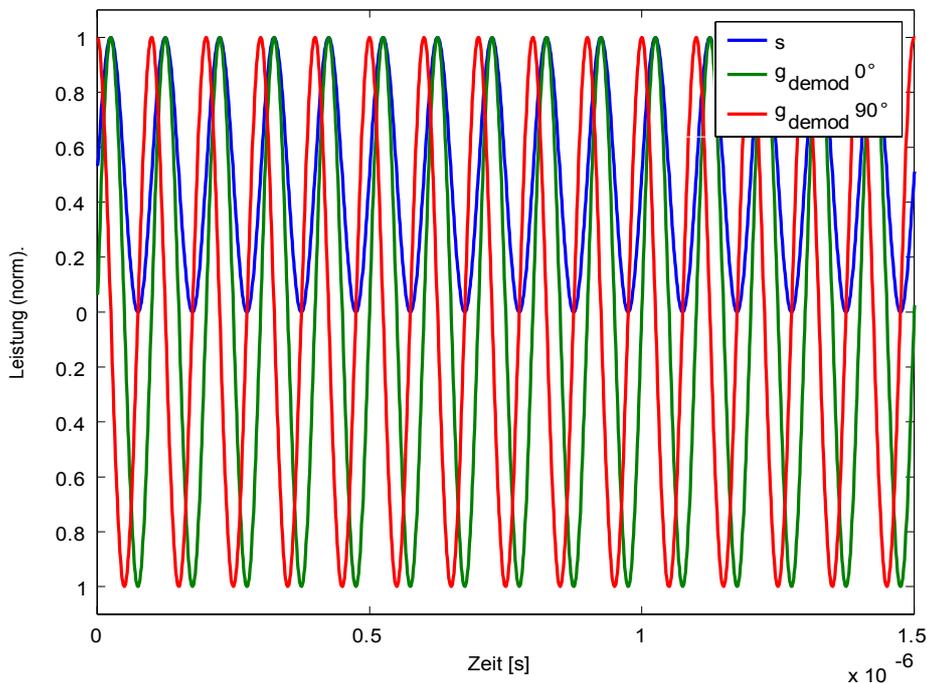


Abb. 4.48: Darstellung der IQ-Demodulation ($g_{\text{demod}0^\circ}$, $g_{\text{demod}90^\circ}$) eines empfangenen Signals s im Zeitbereich ($f=10\text{MHz}$).

Damit wird ein Rauschsignal in Abhängigkeit der Phase verschieden stark auf das Ergebnis durchschlagen. Eine simulative Untersuchung dieses Sachverhaltes bestätigt diese Vermutung (**Abb. 4.50**). Wie hierin ebenso deutlich wird, steht die Phasenunsicherheit in direktem Zusammenhang zum Filterdesign, da sich die vorhandene Rauschleistung durch eine Tiefpassbegrenzung reduzieren lässt. Somit hat also die Auslegung der Filterschaltung einen entscheidenden Einfluss auf das Gesamtverhalten einer Phasenbestimmung mittels IQ-Demodulation.

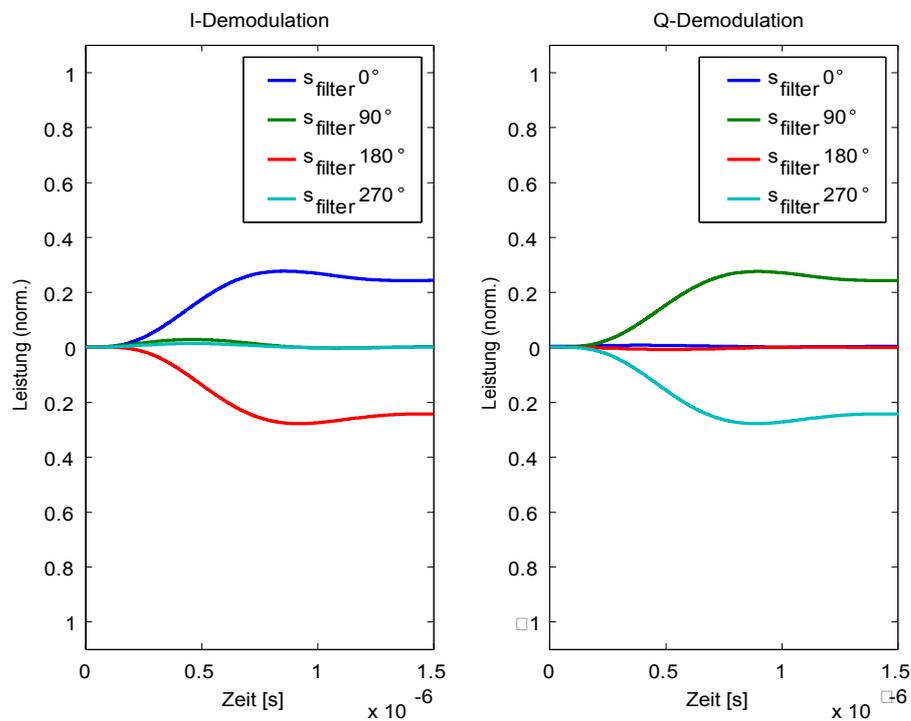
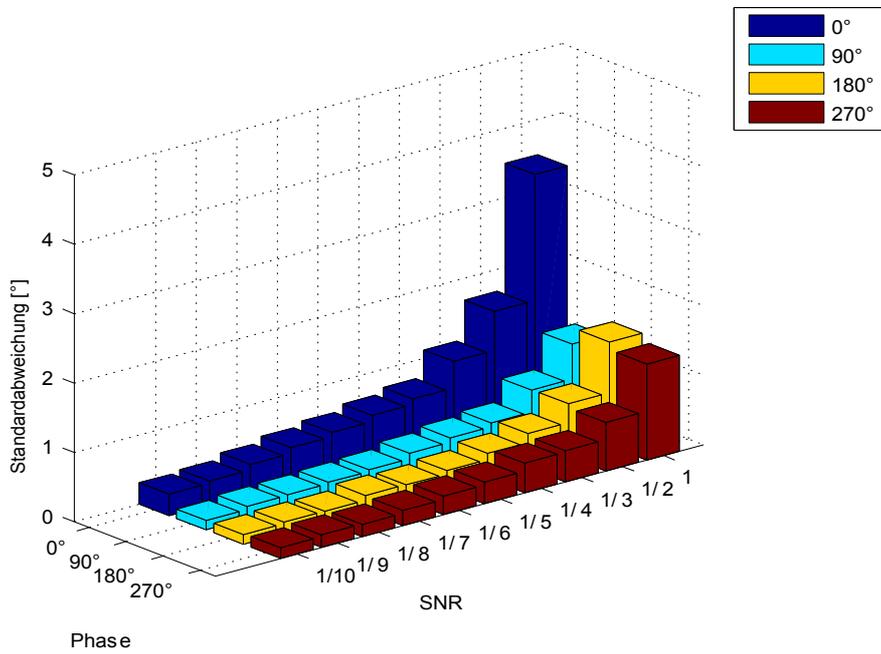


Abb. 4.49: Ausgangssignal einer der IQ-Demodulation nachfolgenden Tiefpassfilterung für unterschiedliche Phasenwerte 0° , 90° , 180° , 270° ($f_{demod}=10MHz$; $f_g=6MHz$).

a)



b)

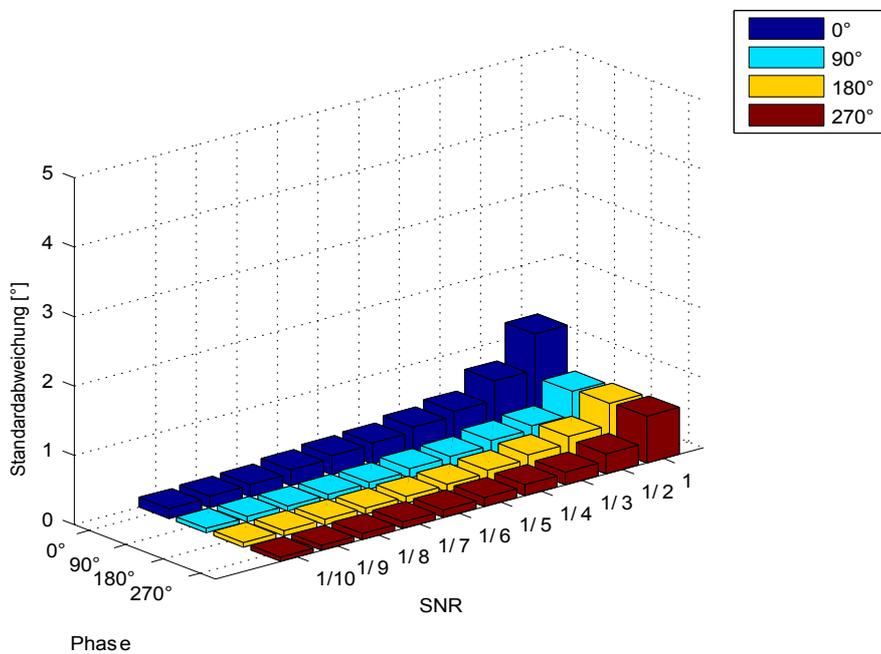


Abb. 4.50: Phasenunsicherheit der IQ-Demodulation ($f_{demod}=10\text{MHz}$) in Abhängigkeit zum Signal-Rausch-Verhältnis für unterschiedliche Filtercharakteristiken a) $f_g=5\text{MHz}$ b) $f_g=1\text{MHz}$

4.7.2 IQ-Demodulation eines optischen Signals mittels PMD

Eine Kombination aus den einzigartigen Eigenschaften der PMD-Technologie und denen der IQ-Demodulation bietet einen vielversprechenden Lösungsansatz für preiswerte und schnelle Lösungen im Bereich der optischen Phasenmessverfahren. Da bei allen optischen Mess- bzw. Übertragungsverfahren die Problematik des Hintergrundlichtes eine wesentliche Einschränkung darstellt, ist hierbei ein Ansatz gesucht, der hintergrundunabhängig nur das aktive Empfangssignal auswertet. **Abb. 4.51** zeigt einen solchen Vorschlag. Hier wird mittels einer Gegentaktmodulation der Kreisfrequenz ω_2 das empfangene optische Signal mit ω_1 demoduliert und auf die Differenzfrequenz $\Delta\omega = \omega_1 - \omega_2$ verschoben. Mit einem nachfolgenden Differenzverstärker werden beide Ausgangssignale vom Gleichanteil befreit und eine angeschlossene Filterschaltung eliminiert die vorhandenen Oberwellen. Anzumerken ist hierbei, dass durch die Differenzbildung nur noch ein Filter vorzusehen ist und somit eventuelle Fehler durch Toleranzen einer zweikanaligen Lösung vermieden werden. Die zugehörigen Signalverläufe dieses Mischens in ein Zwischenband zeigt **Abb. 4.52**.

Die Wirkung des sich anschließenden Filters zeigt **Abb. 4.53**. Deutlich ist hierin zu erkennen, wie die Phaseninformation des Empfangssignals s aus dem hochfrequenten in das Zwischenband abgebildet wird. Mit diesem durch die Differenzbildung offsetbefreiten Signal auf der niederfrequenten Zwischenfrequenz $\Delta\omega$ kann nun die IQ-Demodulation durchgeführt werden, wobei hier eine rein digitale Lösung vorgeschlagen wird. Mittels eines Analog-Digital-Wandlers und einer Recheneinheit wird das digitalisierte Signal über eine einfache Multiplikation demoduliert. Entscheidend ist hierbei, dass die Zwischenfrequenz $\Delta\omega$ der IQ-Demodulation gleich und zu ihr in einem festen Phasenzusammenhang steht. Die abschließende Integration der beiden IQ-Demodulationsausgänge erfolgt ebenso digital und dient dem Sammeln einer ausreichend großen Signalenergie. Aufgrund der zwei wesentlichen Prozesse Multiplikation (Mischen) und Integration handelt es sich also trotz des Zwischenschrittes über die Differenzfrequenz $\Delta\omega$ ebenso um ein korrelatives Messverfahren, welches sich durch eine hohe Gleichtaktunterdrückung auszeichnet.

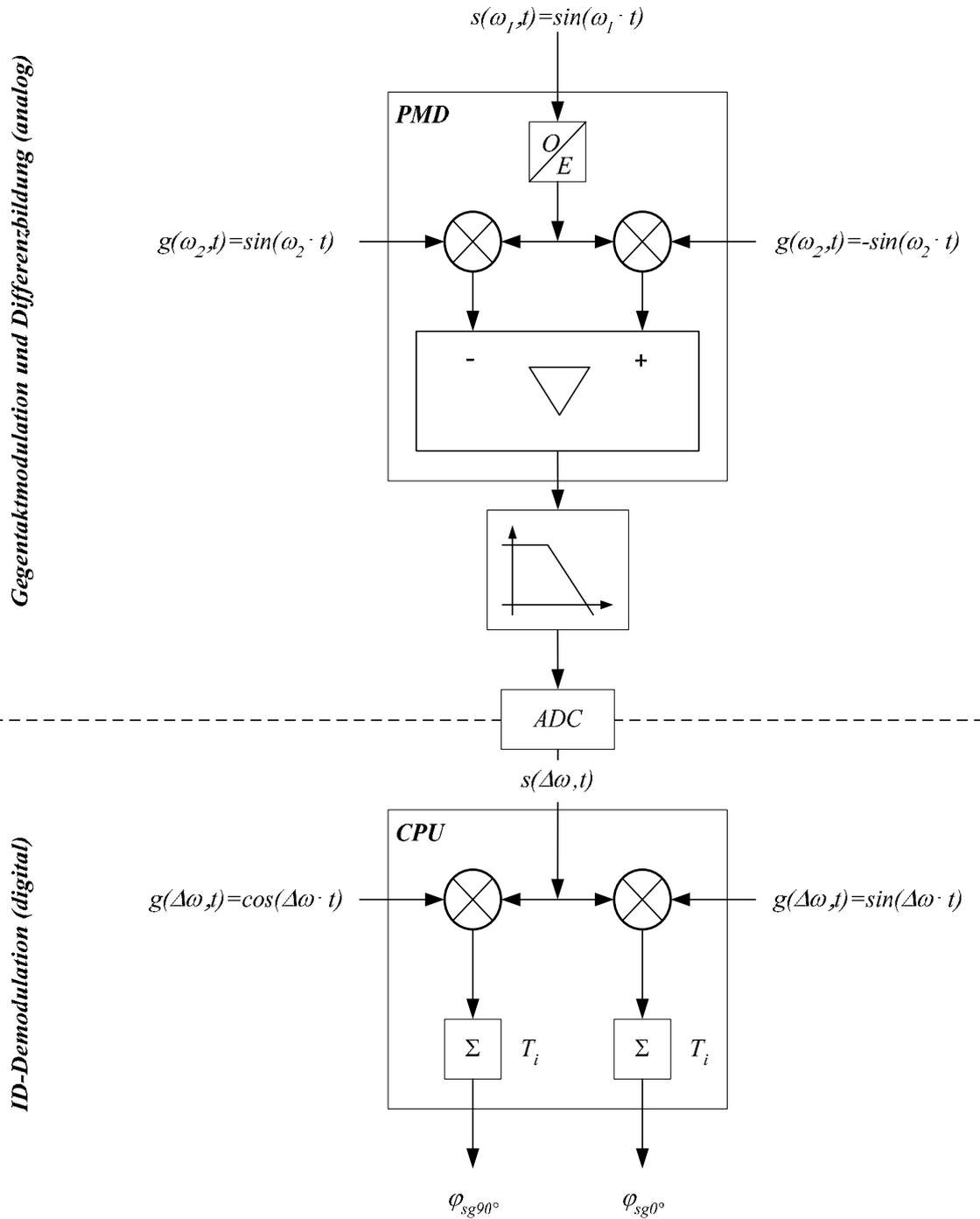


Abb. 4.51: Blockschaltbild einer hintergrundunabhängigen IQ-PMD-Lösung

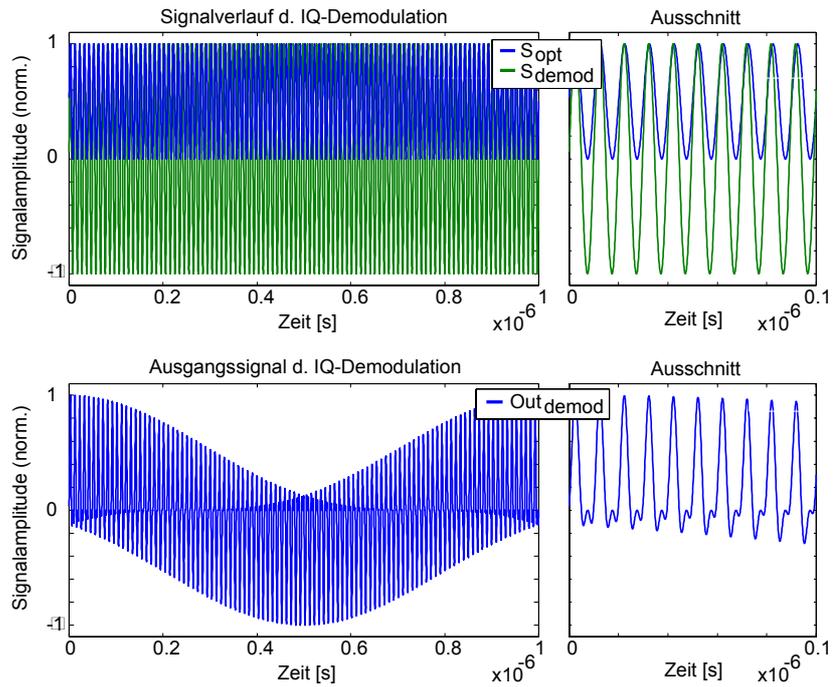


Abb. 4.52: Signalverlauf der IQ-Demodulation bzw. des Mischens in ein niederfrequentes Zwischenband ($f_1=10MHz, f_2=10.1MHz$) mit resultierendem Ausgangssignal

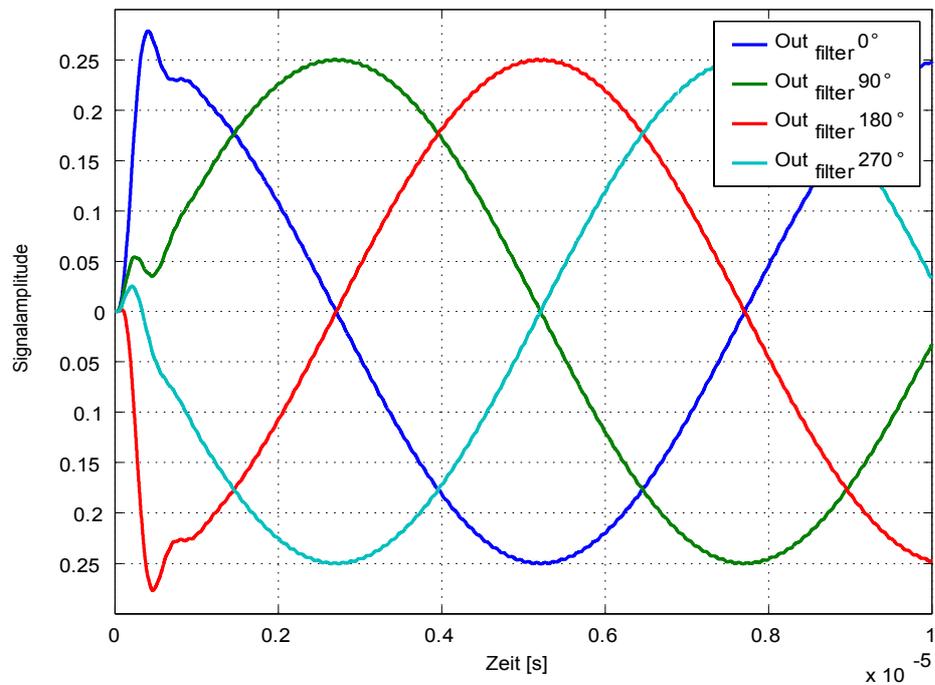


Abb. 4.53: Ausgangssignal der IQ-Demodulation für verschiedene Signalphasen nach der Tiefpassfilterung

4.8 Fazit

Die Auswertung eines intensitätsmodulierten optischen Signals ist eine anspruchsvolle messtechnische Aufgabe. Mit der PMD-Technologie steht jedoch ein Werkzeug zur Verfügung, das diese Problematik deutlich vereinfacht. So kann die Bestimmung der Phaseninformation des modulierten optischen Signals mit PMD-Empfängern sehr unkompliziert realisiert werden, da hier beispielsweise auf aufwendige Breitbandverstärker verzichtet werden kann und sich die Signalkette deutlich verkürzt.

Kombiniert man den PMD-Korrelationsempfang mit den Algorithmen der klassischen Phase-Shift-Interferometrie, dann lassen sich sehr preiswerte und trotzdem hochgenaue Meßsysteme aufbauen, da sich diese Algorithmen flexibel an das aufgebaute System anpassen lassen. Sie geben so die Möglichkeit dessen Schwächen zumindest teilweise zu kompensieren. Hierbei ist anzumerken, dass der Korrelationsempfang mit PMD schon sehr gute Voraussetzungen für ein solches Meßsystem bietet und Schwächen zumeist auf der Senderseite zu suchen sind. So konnte in der Entfernungsmessung mit der erreichten Messgenauigkeit das Problem der Modeninstabilität eines Halbleiterlasers sehr gut beobachtet werden. Mit den theoretischen Grundlagen der Halbleiterlaser, ihrer umfassenden statistischen Beschreibung und einer systemtheoretischen Betrachtung konnte der Zusammenhang zwischen Modenwechsel und der sprunghaften Änderung des gemessenen Entfernungswertes hergestellt werden. Der Erfolg der darauf aufbauenden Maßnahme der zusätzlichen HF-Modulation bestätigt dies zusätzlich.

Neben der hervorragenden Eignung der PMD-Technologie für die optische Entfernungsmesstechnik bietet sie aber noch wesentlich mehr Anwendungsmöglichkeiten, wie hier mit der zeitaufgelösten Spektroskopie gezeigt wurde. Insbesondere die Anordnung mehrerer PMD-Pixel zu einer Zeile oder Matrix bietet enormes Potential und lässt die Fülle an Möglichkeiten als unüberschaubar erscheinen. Selbst für den nahe liegenden Bereich der 3D-Kameratechnik wird in der Zukunft noch eine Menge neuer faszinierender Anwendungen zu erwarten sein, da die 3D-Bewegtbildverarbeitung noch am Anfang ihrer Entwicklung steht. Hier ist noch anzumerken, dass es sich hierbei um ein sehr komplexes mehrdimensionales System handelt, dessen Möglichkeiten es noch mit intelligenten Algorithmen auszureizen gilt. Ganz deutlich konnte der grundlegende Vorteil eines PMD-basierten 3D-TOF-Kamerasystems gegenüber den Korrespondenzansätzen herausgearbeitet werden. Insbesondere das Entfallen der Divergenz- und Korrespondenzproblematik ist hier hervorzuheben, was zukünftig die Entwicklung deutlich sicherer Systeme erwarten lässt.

Neben den hinreichend diskutierten Messverfahren und Algorithmen der Phase-Shift-Interferometrie wurde ein neuartiges Verfahren vorgeschlagen, das den schaltungstechnischen Aufwand deutlich reduziert und die Einfachheit einer IQ-Demodulation mit den Vorteilen der PMD-Technologie kombiniert. Erste Simulationen deuten das Potenzial dieser viel versprechenden Kombination an und lassen interessante Anwendungen im Bereich der optischen Phasenmesstechnik erwarten.

5 Phasenauswertung intensitätsmodulierter Chirp-Signale in der optischen Kommunikationstechnik

Moderne Kommunikationsverfahren unterliegen hohen Anforderungen, denn sie sollen neben einem hohen Datenaufkommen ebenso den Zugriff durch mehrere Nutzer erlauben (engl. Multiple Access). Demgegenüber stehen aber technische Grenzen, die sich beispielsweise aus maximaler Sendeleistung oder einem begrenzten Frequenzband ergeben. Dieser Antagonismus fordert neue intelligente Übertragungsverfahren, die ihre begrenzten Ressourcen dynamisch verwalten und somit ökonomischer nutzen. Ziel eines solchen dynamischen Verhaltens ist es, die als Gesamtkanalkapazität bezeichnete Übertragungsgrenze so aufzuteilen, dass sie einerseits jedem Nutzer in einem ausreichenden Maß zur Verfügung steht und andererseits nicht vergeudet wird. Dies erfordert ein intelligentes Management dieser eingeschränkten Ressource, zumal sich nicht nur sehr unterschiedliche Anforderungen an Übertragungsgeschwindigkeit und -sicherheit durch den einzelnen Nutzer ergeben, sondern ebenso die physikalischen Eigenschaften eines Übertragungsweges sehr vielfältig sein können. Die unterschiedlichen Anforderungen durch den einzelnen Nutzer lassen sich sehr einfach an den verschiedenen nutzbaren Diensten moderner Kommunikationsdienste verdeutlichen. Während einerseits die Übertragung eines Sprachsignals durch ihre Redundanz relativ hohe Fehlerraten verkraftet und die Übertragung der Information immer noch gesichert ist, kann damit andererseits eine Übertragung sicherheitsrelevanter Daten, wie sie z.B. im elektronischen Geldtransfer vorkommen, nicht mehr sichergestellt werden.

Wie unterschiedlich die Eigenschaften eines Übertragungsweges sein können, wird am Beispiel des „Global System for Mobile Communication“ (GSM) deutlich. Hier kann die Länge einer Übertragungstrecke von einigen Metern bis zu mehreren Kilometern variieren. Zusätzlich verlangt auch der nicht zu vermeidende Effekt der Mehrwegeausbreitung (auch als Mehrfachreflexion bezeichnet) eine entsprechende Anpassung, welche sogar einer gewissen Dynamik unterliegt, wenn es sich um mobile Nutzer handelt.

Generell stehen für die Abbildung einer Information auf einen Informationsträger nur wenige Möglichkeiten zur Verfügung. Beschränkt man sich auf elementare Signale, wie z.B. das Sinussignal, dann ergeben sich nur folgende drei grundsätzlichen Möglichkeiten: Amplituden- (AM), Frequenz- (FM) und Phasenmodulation (PM). Zu

beachten ist hierbei, dass diese Modulationsarten nur die Aufgabe der Informationsvermittlung erfüllen, also das Abbilden der Information auf einen Informationsträger. Die eigentliche Transportaufgabe übernimmt der Informationsträger, beispielsweise elektromagnetische Wellen wie Funk oder Licht. Auf diesen Informationsträger wirken nun die unterschiedlichsten Störungen ein. Sind diese bekannt und kalkulierbar, kann ein Verfahren derart gewählt werden, dass trotzdem ein sicherer Transport möglich ist. Hierbei legt man sich auf eine der drei Grundmodulationsarten AM, FM, PM oder einer Kombination aus ihnen fest. Ändern sich jedoch unerwartet die Umgebungsverhältnisse, d.h. eine weitere oder eine schon vorhandene Störung wirkt verstärkt auf den Übertragungskanal, dann kann die Informationsübermittlung fehlerhaft sein oder sogar unmöglich werden, da die festen Parameter des gewählten Übertragungsverfahrens keine Reaktion auf die neuen Bedingungen erlauben.

Der Zugriff durch mehrere Nutzer verlangt eine Aufteilung der Gesamtkanalkapazität, wobei im Wesentlichen drei Prinzipien dominieren: die Code-, Frequenz- und Zeitmultiplex-Verfahren, im Englischen mit Code Division Multiple Access (CDMA), Frequency Division Multiple Access (FDMA) und Time Division Multiple Access (TDMA) bezeichnet. Während CDMA- und FDMA-Verfahren einen parallelen Mehrfachzugriff erlauben, arbeiten TDMA-Verfahren sequentiell, indem sie Daten verschiedener Nutzer zeitlich gestaffelt übertragen. TDMA besitzt dadurch die Möglichkeit, je nach Anforderung jedem Nutzer sehr flexibel unterschiedliche Zeitfenster zuzuweisen. Somit wird die Gesamtkanalkapazität asymmetrisch aufgeteilt und jedem Nutzer eine spezifische Datenrate zur Verfügung gestellt. Damit erfüllen die TDMA-Verfahren die Forderung nach einer möglichst ökonomischen Aufteilung der Kanalkapazität. Dem steht jedoch ein gravierender Nachteil gegenüber. Einen Systemgewinn durch Bandspreizung besitzen sie nicht. Damit leiden sie mit steigender Reichweite und Datenrate zunehmend unter Störungen, sei es durch Mehrwegeausbreitung oder zu geringer Empfangsleistung. Wie im Folgenden dargestellt, haben CDMA-Verfahren dagegen einen solchen Systemgewinn, jedoch sind sie unflexibel und hinsichtlich einer ökonomischen Aufteilung der Gesamtkapazität auf möglichst viele Nutzer mit unterschiedlichen Anforderungen begrenzt. Der Einsatz sogenannter Chirp-Signale soll nun die Vorteile beider Verfahren verbinden.

5.1 Bandspreizung und der Begriff des Systemgewinns

Wie zuvor dargestellt, wird die eigentliche Aufgabe des Informationstransports durch den Informationsträger erfüllt. Weit verbreitet ist hierbei die Annahme, dass die Frage, ob eine Information den Empfänger erreicht oder nicht, allein von der Leistung seines Trägers abhängt. Die Reichweite eines Signalträgers ist jedoch vielmehr eine Funktion der Energie. Da die Energie dem Integral der Leistung über der Zeit entspricht, kann selbst ein sehr leistungsschwacher Signalträger über ein entsprechend langes Zeitintervall sicher detektiert werden. Hierbei ist die entsprechende Information ebenso lange auf dem Informationsträger abzubilden. Damit sinkt zwar die erreichbare Datenrate, aber eine sichere Übertragung ist trotzdem möglich. Diese Betrachtungsweise führt zwangsweise zur Verwendung eines Korrelationsempfängers, wie er in **Kap. 2.3.1** bzw. **3.6** beschrieben wurde. Ein ideales Übertragungsverfahren zeichnet sich weiterhin durch eine absolute Immunität gegenüber Störsignalen aus. Sich dem Einfluss einer Störung zu entziehen, gelingt durch einen gezielten Wechsel auf ein störungsfreies Frequenzband. Um auf jedwede Störung flexibel reagieren zu können, sollte ein Übertragungsverfahren in einem möglichst breiten Frequenzband arbeiten können. Da sich gerade in Kommunikationssystemen mit mobilen Teilnehmern Störeinflüsse ständig ändern, liegt die Idee nahe, das Signal direkt über ein breites Frequenzband zu verteilen. Hierfür kommen eine Vielzahl unterschiedlicher Methoden in Betracht, welche sich alle unter dem Begriff der Bandspreizung bzw. Spread Spectrum zusammenfassen lassen. Der Vorteil der Bandspreizung ist aus **Abb. 5.1** leicht ersichtlich. Liegt eine Störung beispielsweise direkt auf einem herkömmlichen amplitudenmodulierten Signal, so hat diese direkt einen starken Einfluss auf das empfangene Signal, da sie ebenso durch den Empfänger detektiert wird (**Abb. 5.1 a**). Übertragungsverfahren auf Basis der Bandspreizung verwenden bei der Übertragung eine Bandbreite, die wesentlich größer ist als die minimal erforderliche:

„Die Spread-Spectrum Modulation verteilt ein relativ niederdimensionales Informationssignal in einen hochdimensionalen Signalraum. Das Signal wird im Signalraum versteckt.“¹

¹ Goiser, Alois M. J.: Handbuch der Spread-Spectrum-Technik, Springer Verlag, Wien [u.a.] 1998, S.

Demnach beruht eine Spread-Spectrum-Übertragung auf einer Aufweitung des Datenspektrums mit anschließender Komprimierung auf die ursprüngliche Datenbandbreite im Empfänger. Liegt nun im Übertragungskanal über dem aufgeweiteten Signal eine Störung, so hat diese nur einen Einfluss auf den Signalanteil, der während der Demodulation den gleichen Frequenzbereich überstreicht (**Abb. 5.1 b**). Empfängt man das im Kanal gestörte Signal und entspreizt es wieder ins Basisband, dann hat das Störsignal wesentlich weniger Einfluss genommen als bei konventionellen Übertragungsverfahren. Diese Reduktion des Störeinflusses bezeichnet man im Allgemeinen als Systemgewinn.

5.2 Spread-Spectrum-Verfahren in der Kommunikationstechnik

a) Frequency Hopping

Eine der ältesten Methoden der Spread-Spectrum-Technik ist das Frequency Hopping (FH oder FHSS). Im Unterschied zur klassischen Schmalbandkommunikation wird beim FH zwischen mehreren Bändern/Frequenzen nach einem bestimmten Schema, dem Hopping-Plan, gewechselt (**Abb. 5.2**). Dieses Schema muss natürlich nicht nur dem Sender, sondern ebenso dem Empfänger bekannt sein.

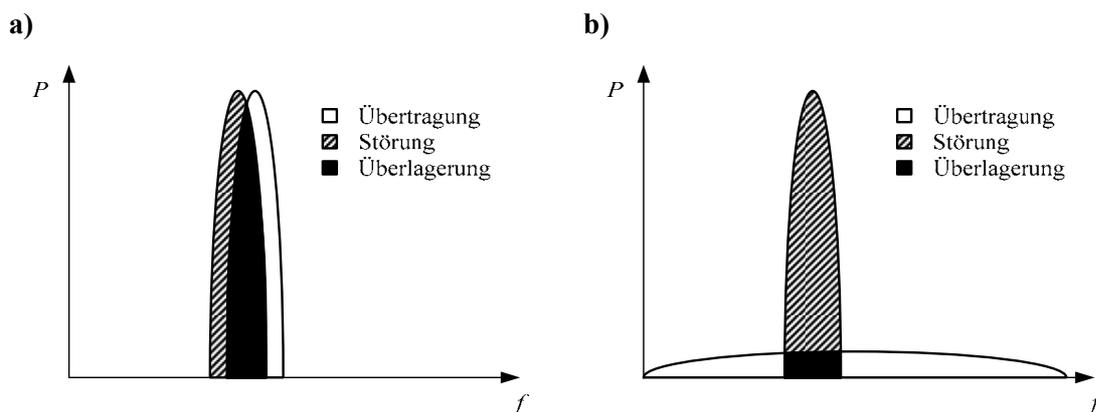


Abb. 5.1: Grundprinzip der Spread-Spectrum-Technik: a) Störung eines konventionellen Übertragungsverfahrens und b) Systemgewinn durch Bandspreizung

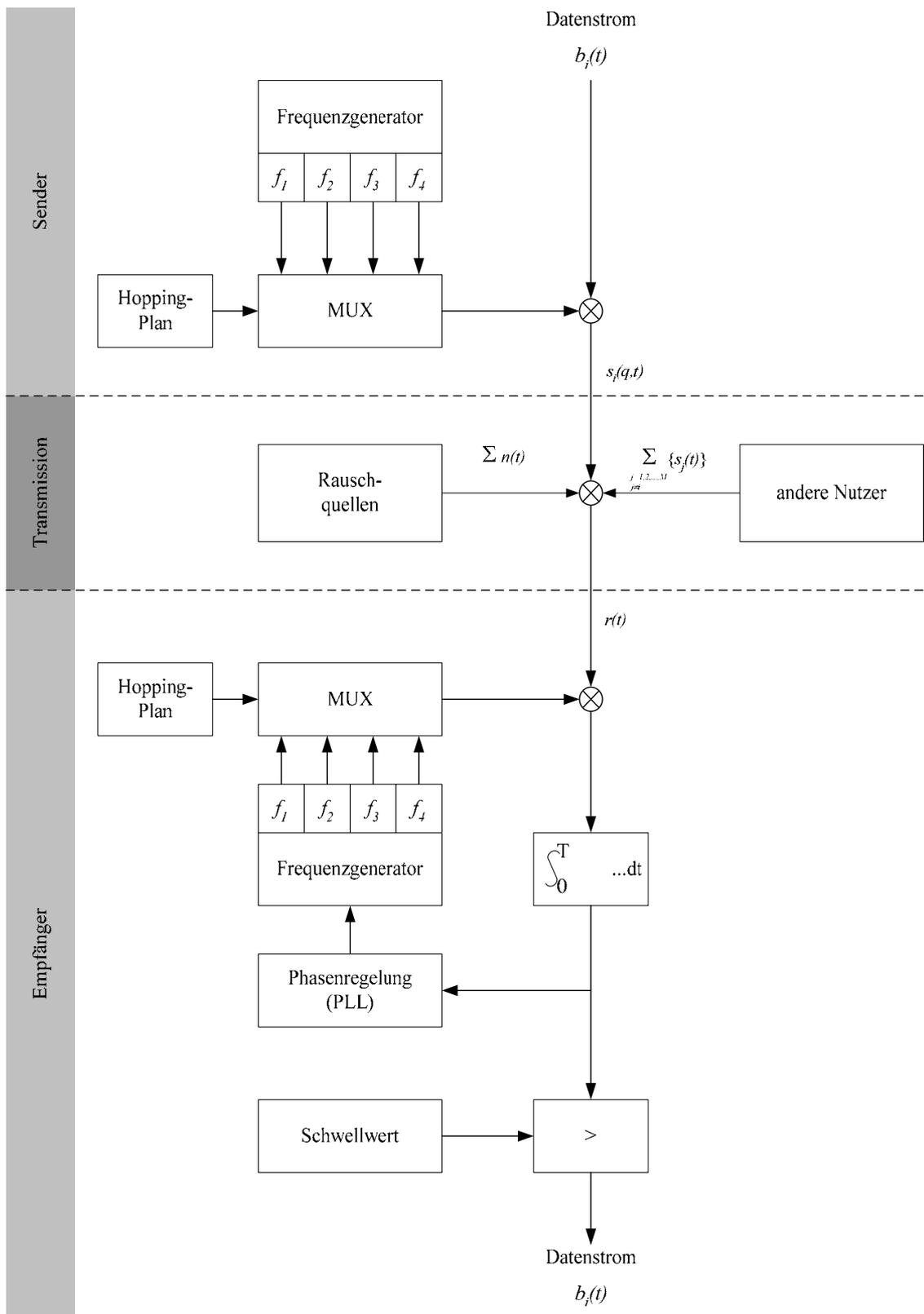


Abb. 5.2: Schematische Darstellung einer Datenübertragung mittels Frequency Hopping Spread Spectrum (FHSS)

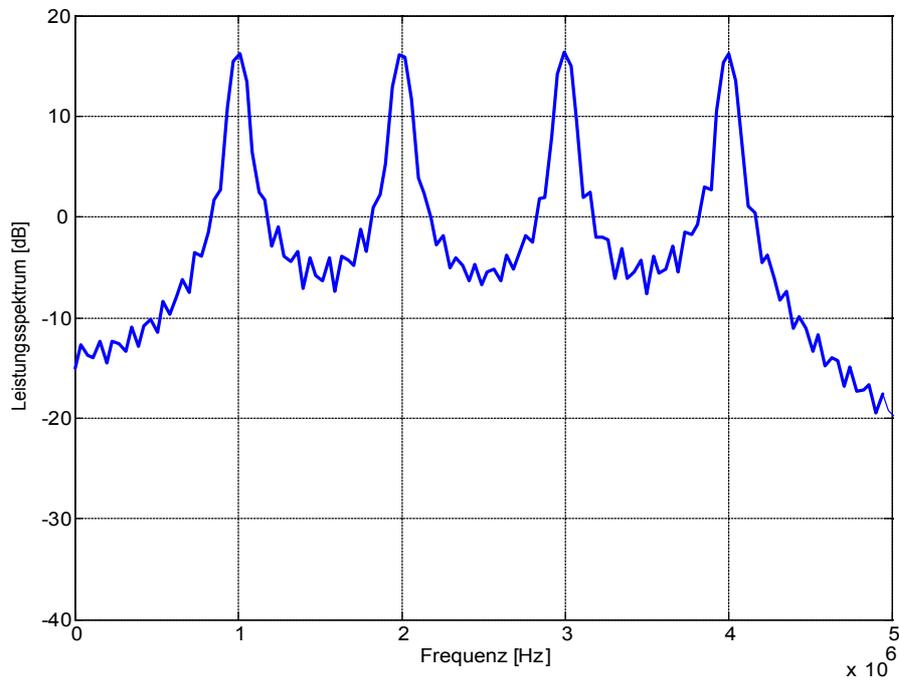
Da sich, wie in **Kap. 3.6** dargestellt, der Einsatz eines Korrelationsempfängers bei geringen Signalleistungen besonders eignet, ist bei seiner Verwendung das Hopping zwischen den einzelnen Frequenzen im empfangenen Signal mit der Demodulation zu synchronisieren, was einen zusätzlichen Aufwand an Schaltungstechnik bedeutet.

Betrachtet man nun die spektrale Verteilung des Frequency Hoppings (**Abb. 5.3 a**), dann wird die Problematik eines solchen Übertragungsverfahrens relativ schnell deutlich. Trifft nämlich eine Störung auf eine der verwendeten Frequenzen und verhindert eine sichere Verbindung (**Abb. 5.3 b**), dann ist die Kommunikation mitunter vollständig gestört, denn die verlorengegangene Information kann die über alle Bänder übertragene Gesamtinformation nutzlos werden lassen. Zur Vermeidung von solchen Übertragungsverlusten bzw. zur Verbesserung der Übertragungsleistung sind dynamische Hopping-Pläne zu realisieren, welche die gestörten Frequenzen bewusst seltener verwenden. Ganz vermeiden sollte man diese Bänder jedoch nicht, da zu einem späteren Zeitpunkt die Störung verschwinden und eine sichere Kommunikation wieder möglich sein könnte.

b) Direct Sequence Spread Spectrum

Direct Sequence Spread Spectrum (DS oder DSSS) bezeichnet die Verfahren, die eine zu übertragende Information direkt auf einen Signalträger aufmodulieren, der schon in gespreizter Form vorliegt. Das bekannteste Verfahren stellt hierbei der Code-Division-Multiple-Access dar, in dem ein binäres Datensignal mit einer zusätzlichen Sequenz verknüpft wird. Geeignete Sequenzen stellen hierbei Pseudo-Random- (PR) bzw. Pseudo-Noise-Folgen (PN) dar, da sie annähernd die statistischen Eigenschaften eines idealen Rauschens besitzen und so für eine gleichmäßige Bandspreizung sorgen (**Abb. 5.4**). Wichtig bei der Betrachtung solcher CDMA-Verfahren ist die Tatsache, dass nur ein beschränkter Umfang an PN-Signalen zur Verfügung steht. Damit erscheint die Implementierung einer intelligenten dynamischen Ressourcenverwaltung von vornherein schwierig.

a)



b)

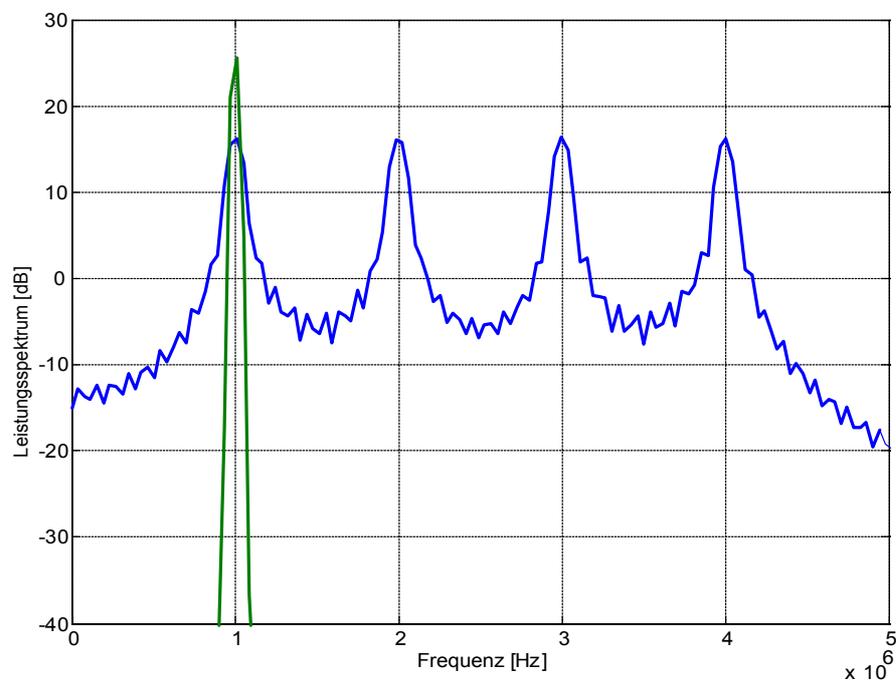


Abb. 5.3: Leistungsdichtespektrum eines Frequency-Hopping-Systems: a) ungestört und b) Störung auf Band 1

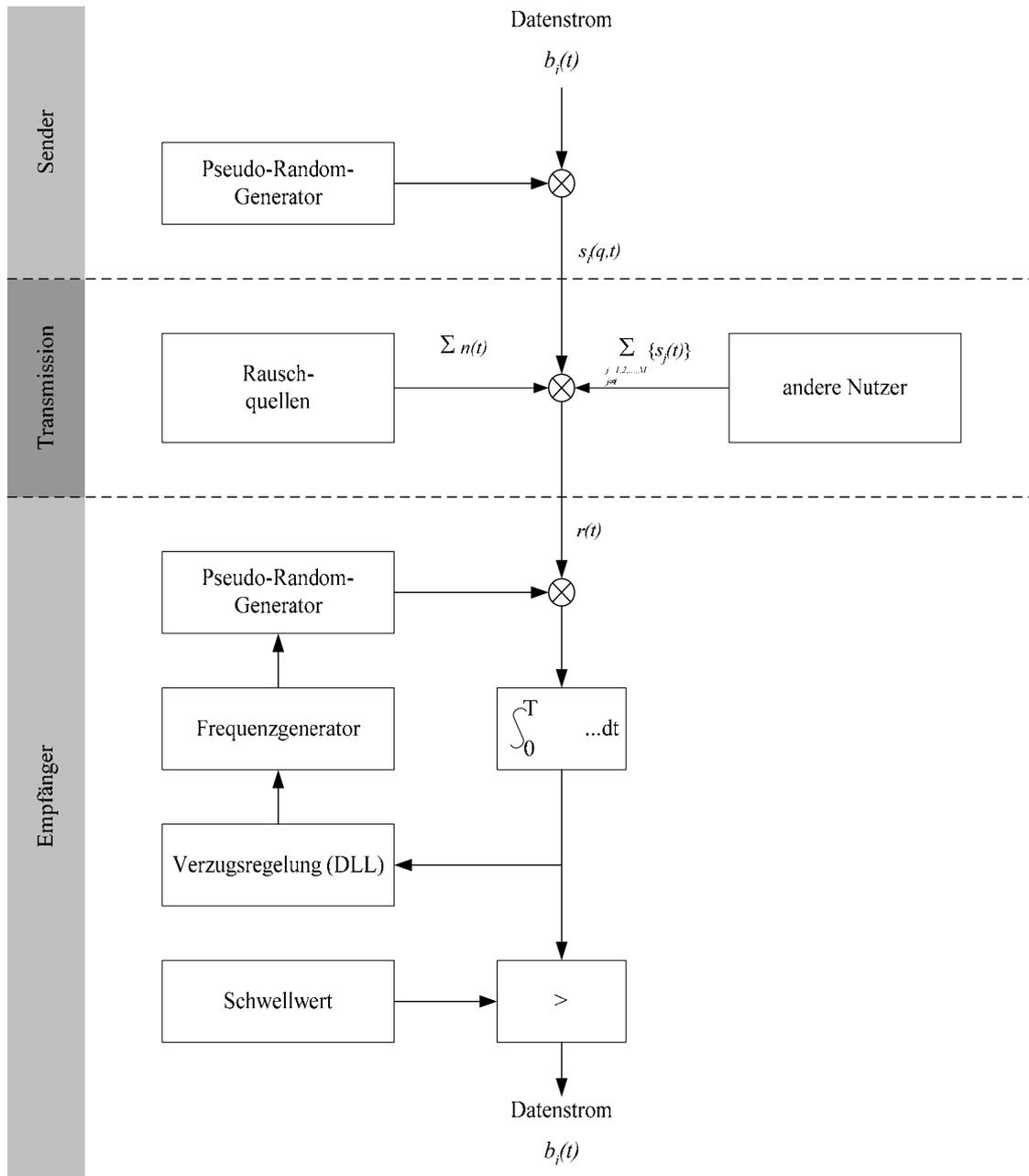


Abb. 5.4: Schematische Darstellung einer Datenübertragung mittels Direct Sequence Spread Spectrum (DSSS)

c) Frequency Chirping

Der Begriff des Frequency Chirping (FC oder FCSS) leitet sich namentlich aus dem englischen „to chirp“ (zwitschern, zirpen) ab. Das Funktionsprinzip ähnelt dem des FH, jedoch werden hier nicht feste Frequenzen für bestimmte Zeiträume angesprungen, sondern ein fester Frequenzbereich in einer festen Zeitspanne meist linear durchlaufen (**Abb. 5.5**). Dieser Zusammenhang lässt sich nicht nur als eine kontinuierliche Frequenzmodulation betrachten, sondern ebenso als Phasenmodulation, wobei hier die Phase eine quadratische Funktion der Zeit darstellt:

$$s(t) = \cos(2\pi f_0 t + \pi k_r t^2). \quad \text{Gl. 5.1}$$

Der Phasenwinkel eines Chirps zu einem festen Augenblick lautet somit:

$$\varphi(t) = \pi k_r t^2. \quad \text{Gl. 5.2}$$

Der zugehörige Frequenzwert ergibt sich somit der mathematischen Differentiation der Phasenfunktion nach der Zeit:

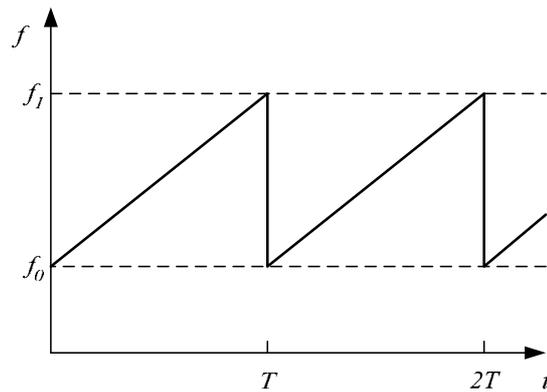
$$f(t) = \frac{1}{2\pi} \cdot \frac{d\varphi(t)}{dt} = k_r \cdot t. \quad \text{Gl. 5.3}$$

Alle charakteristischen Parameter eines Chirps wie Startfrequenz f_1 , Stopfrequenz f_2 und Chirp-Rate $k_r = \Delta f / \Delta t$ müssen sowohl dem Sender als auch dem Empfänger bekannt sein. Liegt nun eine Störung in diesem Frequenzbereich, dann wird die Verbindung erst dann entscheidend gestört, wenn die empfangene Störenergie gegenüber der gesamten Energie eines Chirps dominiert.

Das Blockschaltbild eines Übertragungsverfahrens mittels Chirpsignalen zeigt **Abb. 5.6**. Auf die wesentlichen Eigenschaften von Chirpsignalen soll im Folgenden

detailliert eingegangen werden, insbesondere hinsichtlich ihrer Verwendung in der Kommunikationstechnik.

a)



b)

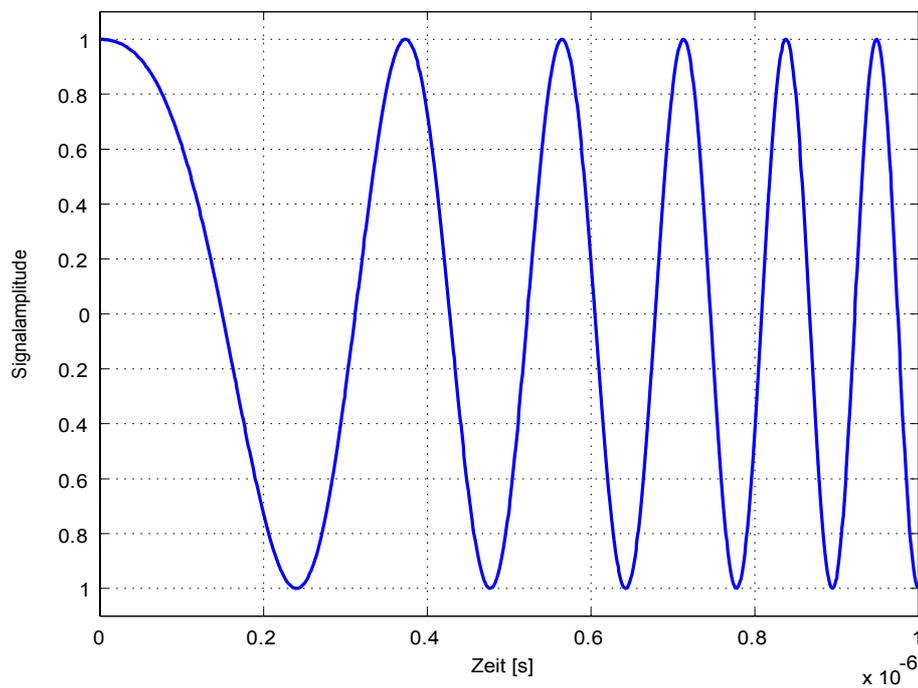


Abb. 5.5: Darstellung eines linearen Chirpsignals von 1 bis 10MHz: a) lineare Änderung der Frequenz innerhalb eines Zeitintervalls T ; b) Darstellung im Zeitbereich

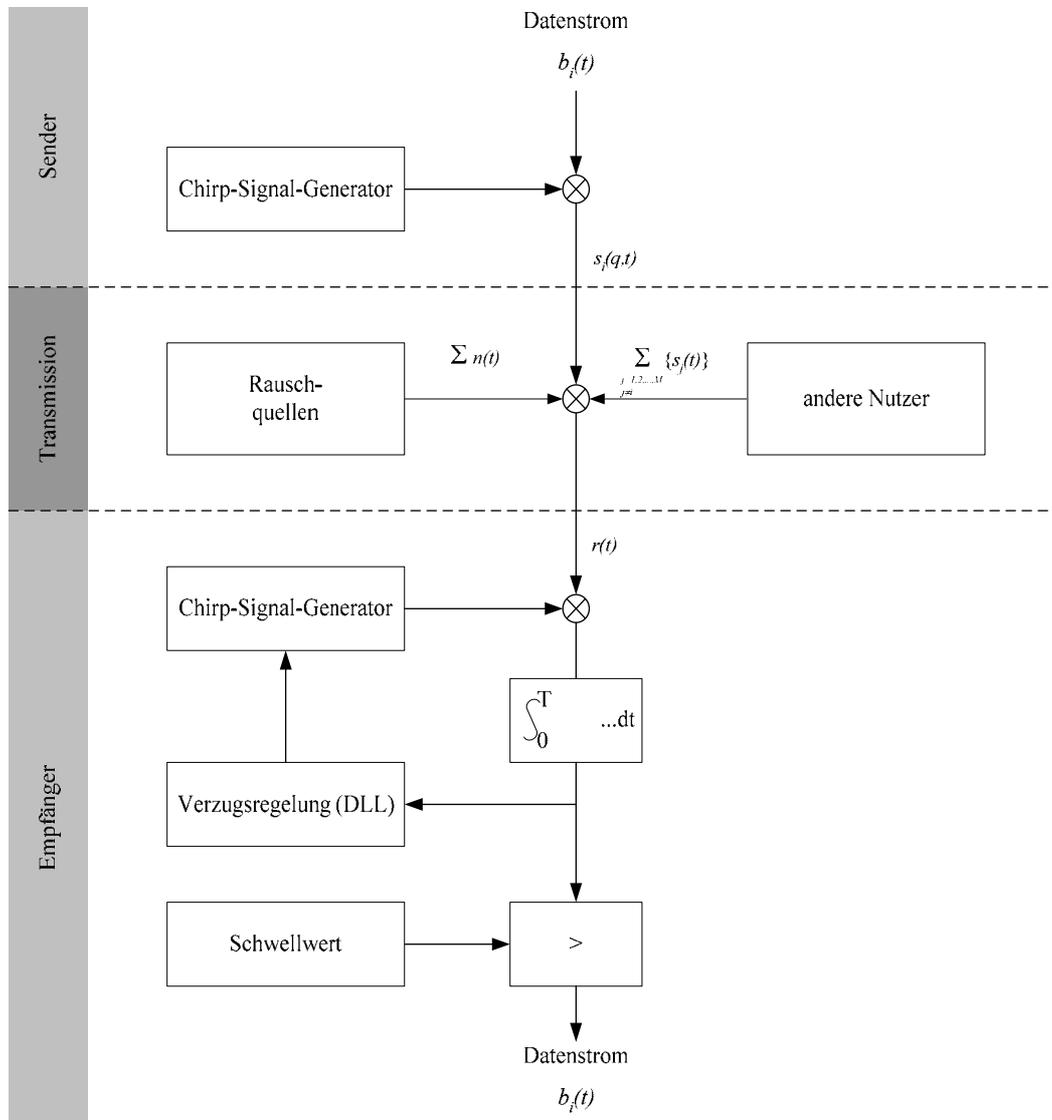


Abb. 5.6: Schematische Darstellung einer Datenübertragung mittels Chirped Spread Spectrum (CSS)

5.3 Systemtheoretische Betrachtung eines Chirp-Signals

Chirp-Signale basieren auf einer linearen Frequenzmodulation (engl. Linear Frequency Modulation - LFM) bzw. dem „chirping“ und lassen sich mathematisch nach **Gl. 5.1** als eine quadratische Funktion der Phase über der Zeit beschreiben. Hierin stellt f_0 die

Startfrequenz dar, während die kontinuierliche Änderung der Frequenz durch die Konstante k_r , auch Chirp-Rate genannt, ausgedrückt wird. Sie ergibt sich aus der gewünschten Bandbreite B des Spreizspektrums B und dem zum Durchlaufen von B benötigten Zeitfenster T :

$$k_r = \frac{B}{T}. \quad \text{Gl.5.7}$$

Reale Chirp-Signale sind somit durch ein Intervall T zeitbegrenzt und lauten somit in ihrer komplexen Form:

$$s(t) = \text{rect}\left[\frac{t}{T}\right] \cdot e^{j(2\pi f_0 t + \pi k_r t^2)}. \quad \text{Gl.5.8}$$

Die Transformation in ein äquivalentes Tiefpasssignal nach **Gl. 2.24** resultiert in folgendem Ausdruck:

$$s_T(t) = \text{rect}\left(\frac{t}{T}\right) \cdot e^{j\pi k_r t^2}. \quad \text{Gl.5.9}$$

Ein empfangenes Chirp-Signal $s(t)$ lässt sich von einem gesendeten durch die Laufzeit τ unterscheiden, welche sich aus der Entfernung von Sender zu Empfänger und der Übertragungsgeschwindigkeit innerhalb des genutzten Mediums ergibt:

$$r(t) = r_0 \cdot s(t - \tau) = s(t) * \delta(t - \tau). \quad \text{Gl.5.10}$$

Durch den in Bandspreizverfahren üblichen Korrelationsempfang ist nur die Autokorrelationsfunktion $\varphi(\tau)$ von weiterem Interesse, da sich das Empfangssignal $r(t)$ auf das originale Chirp-Signal $s(t)$ zurückführen lässt.

$$r(t) = r_0 \cdot \text{rect}\left[\frac{t - \tau}{T}\right] \cdot e^{j(2\pi f_0 (t - \tau) + \pi k_r (t - \tau)^2)}. \quad \text{Gl.5.11}$$

Analog zur äquivalenten Tiefpassdarstellung eines Bandpasssignals lässt sich die Autokorrelationsfunktion ebenso in den Tiefpassbereich transformieren:

$$\varphi_{ss}(t) = \operatorname{Re}\left\{\varphi_{ss_T}(t) \cdot e^{j(2\pi f_0 t)}\right\}. \quad \text{Gl.5.12}$$

Entsprechend der Darstellung von Korrelation und Faltung (**Kap. 2.3**) lässt sich diese ebenso formell als Faltungsintegral darstellen:

$$\begin{aligned} \varphi_{ss_T}(t) &= \frac{1}{2} s_T^*(t) \otimes s_T(t) = \frac{1}{2} s_T(t) * s_T^*(-t) \\ &= \frac{1}{2} \int_{-\infty}^{\infty} s_T(\tau) \cdot s_T^*(-(t-\tau)) \, d\tau \end{aligned} \quad \text{Gl.5.13}$$

Angewandt auf den hergeleiteten analytischen Ausdruck eines Chirpsignals ergibt sich so:

$$\begin{aligned} \varphi_{ss_T}(t) &= \frac{1}{2} \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{\tau}{T}\right] \cdot e^{j\pi k_r \tau^2} \cdot \operatorname{rect}\left[-\frac{\tau-t}{T}\right] \cdot e^{-j\pi k_r (\tau-t)^2} \cdot d\tau \\ &= \frac{1}{2} \cdot e^{j\pi k_r t^2} \cdot \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{\tau}{T}\right] \cdot \operatorname{rect}\left[\frac{\tau-t}{T}\right] \cdot e^{-j2\pi k_r t \tau} \cdot d\tau \quad \text{Gl. 5.14} \\ &= \begin{cases} 0; & -T \geq t \geq T \\ \frac{T}{2} \cdot \Lambda\left[\frac{t}{T}\right] \cdot \operatorname{si}\left(\pi k_r T \cdot t \cdot \Lambda\left[\frac{t}{T}\right]\right); & |t| \leq T \end{cases} \end{aligned}$$

Dieser allgemeine Ausdruck der Autokorrelationsfunktion eines beliebigen Chirp-Signals ist in **Abb. 5.7** graphisch dargestellt. Anhand dieser Darstellung wird der Effekt des „Despreizens“ mittels Korrelationsempfang deutlich. Die gesamte Leistung des empfangenen Signals, die im Übertragungskanal über das gesamte Frequenzband verteilt ist, wird durch die Autokorrelation (Demodulation und Integration) gebündelt.

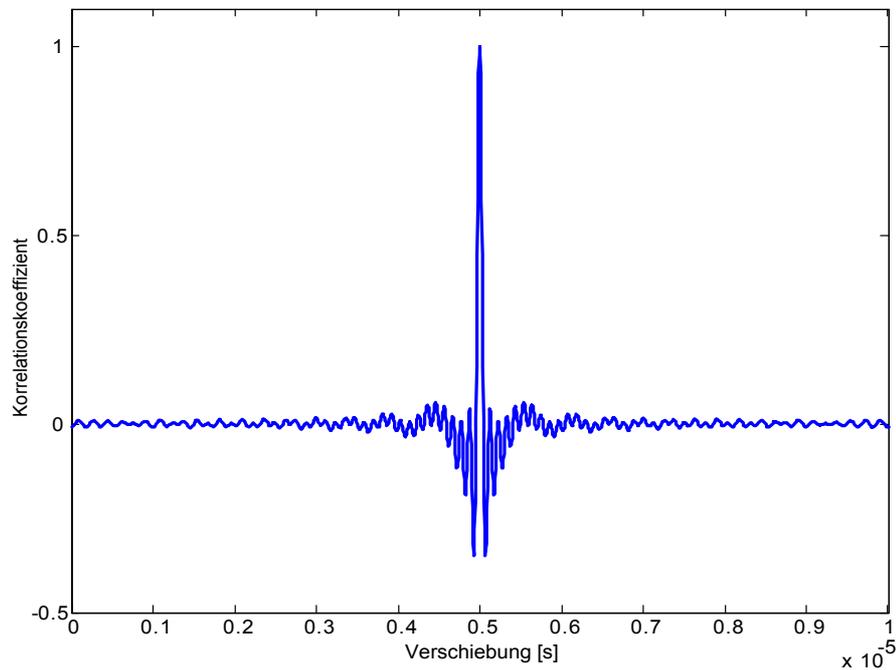


Abb. 5.7: Simulierte Autokorrelationsfunktion eines Chirp-Signals ($f=1-10\text{MHz}$, $T=10\mu\text{s}$)

5.4 Dynamische Kommunikationsverfahren mittels variabler Chirp-Signale

5.4.1 Bedeutung der Phasenmodulation

Mit den bisher angestellten Überlegungen ist nur das grundsätzliche Prinzip der Bandspreizung mittels Chirp-Signalen dargestellt. Die eigentliche Aufgabe der Informationsvermittlung jedoch, also der Abbildung der Information auf einen einzelnen Chirp, wurde bisher nicht betrachtet. Ausgehend von den drei grundsätzlichen Modulationsarten kann zur Informationsabbildung nur noch die Phasenmodulation genutzt werden. Frequenz- und Amplitudenmodulation erscheinen deshalb nicht sinnvoll, weil sie direkten Einfluss auf das Signalspektrum und damit auf

den Prozess der Bandspreizung nehmen. Die Abbildung der zu übertragenden Information auf ein Chirp-Signal mittels Phasenmodulation kann sehr einfach durch eine Variation der Initialphase θ erfolgen. Handelt es sich hierbei um ein binäres Datensignal $b(t)$, so ergibt sich für das zu sendende Chirp-Signal $s(t)$ folgender mathematischer Ausdruck:

$$s(t) = \cos(2\pi f_0 t + \pi k_r t^2 + b(t) \cdot \theta); \quad b \in (0;1) \quad \text{Gl. 5.15}$$

$$= \begin{cases} \cos(2\pi f_0 t + \pi k_r t^2); & b = 0 \\ \cos(2\pi f_0 t + \pi k_r t^2 + \theta); & b = 1 \end{cases}$$

Ist $\theta = \pi$, dann ändert der Datenstrom $b(t)$ nur das Vorzeichen des Chirp-Signals $s(t)$:

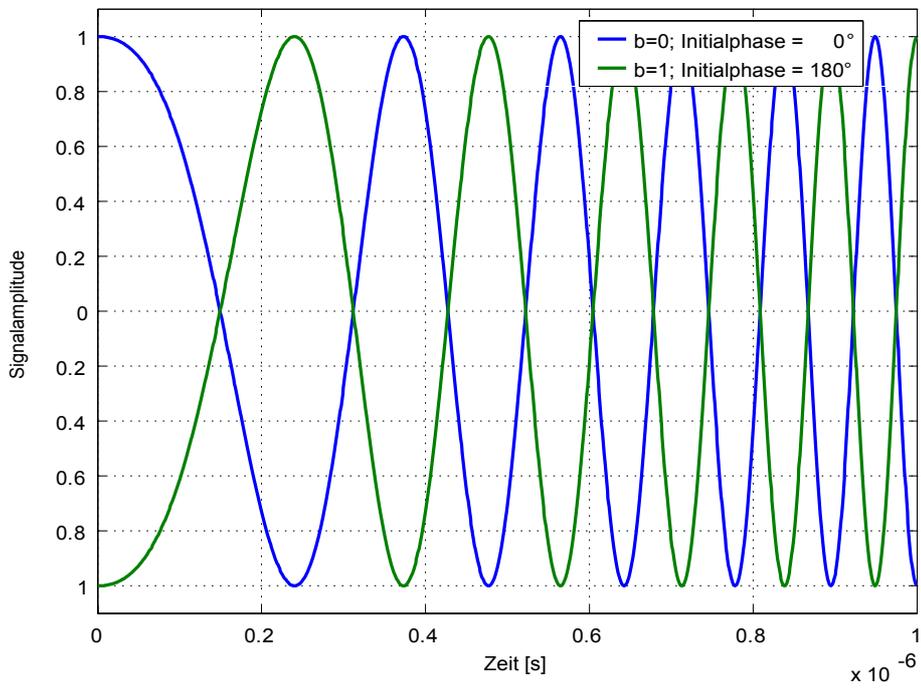
$$s(t) = \begin{cases} \cos(2\pi f_0 t + \pi k_r t^2); & b = 0 \\ -\cos(2\pi f_0 t + \pi k_r t^2); & b = 1 \end{cases} \quad \text{Gl. 5.16}$$

Eine derartige Variation der Startphase ist technisch sehr einfach durch den Einsatz eines Inverters realisierbar. Wendet man den durch den Datenstrom $b(t)$ verursachten Vorzeichenwechsel auf die Autokorrelation eines Chirp-Signals an, dann ergibt sich darin ebenso ein Vorzeichenwechsel:

$$\varphi_{ss_T}(t) = \begin{cases} 0; & -T \geq t \geq T \\ +\frac{T}{2} \cdot \Lambda\left[\frac{t}{T}\right] \cdot \text{si}\left(\pi k_r T \cdot t \cdot \Lambda\left[\frac{t}{T}\right]\right); & |t| \leq T, \quad b = 0 \\ -\frac{T}{2} \cdot \Lambda\left[\frac{t}{T}\right] \cdot \text{si}\left(\pi k_r T \cdot t \cdot \Lambda\left[\frac{t}{T}\right]\right); & t \geq T, \quad b = 1 \end{cases} \quad \text{Gl. 5.17}$$

Der Einfluss unterschiedlicher Initialphasen θ auf die Signalform eines Chirps und die damit einhergehenden Auswirkungen auf die Kreuzkorrelationsfunktion zeigt **Abb. 5.8**.

a)



b)

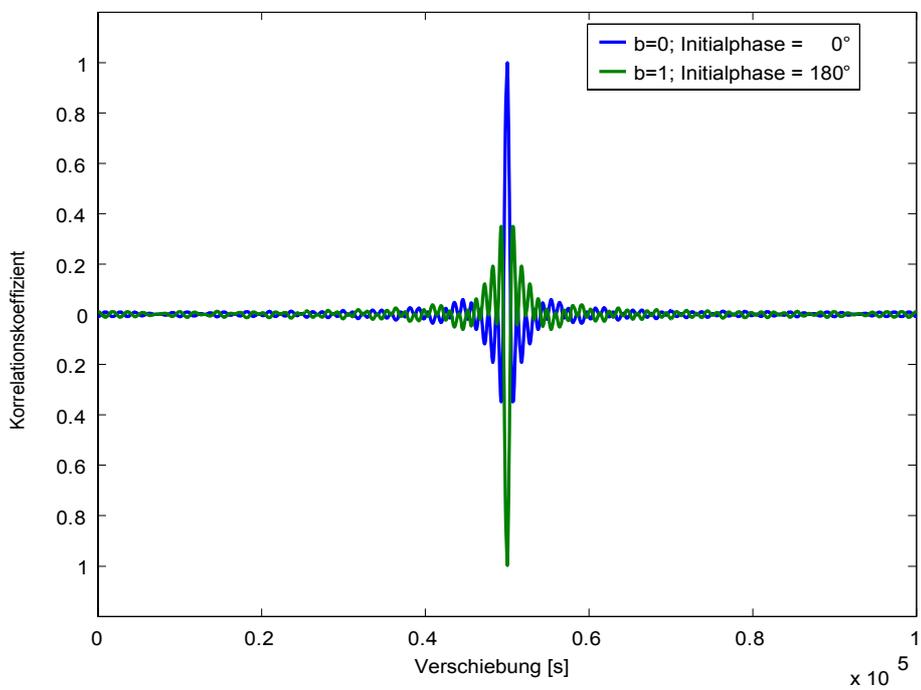


Abb. 5.8: Phasenkodierung eines Chirp-Signals ($\theta = 0; \pi$) im a) Zeitbereich und b) ihr Effekt auf die Kreuzkorrelationsfunktion

Eine Initialphase von 180° führt, wie analytisch hergeleitet, zu einer horizontalen Spiegelung der Korrelation bzw. zu einem technisch leicht zu unterscheidenden Vorzeichenwechsel. Für eine analoge Signalübertragung ist das Übertragungsverhalten des Korrelators bezüglich der Initialphase θ von großem Interesse. Für die analytische Bestimmung ist das gesendete Signal $s(t)$ um die Initialphase θ zu erweitern:

$$s(t) = \text{rect}\left[\frac{t}{T}\right] \cdot e^{j(2\pi f_0 t + \pi k_r t^2 + \theta)} . \quad \text{Gl. 5.18}$$

Das äquivalente Tiefpasssignal lautet hierfür:

$$s_T(t) = \text{rect}\left(\frac{t}{T}\right) \cdot e^{j\pi k_r t^2 + \theta} . \quad \text{Gl. 5.19}$$

Das Verhalten des Korrelationsempfängers lässt sich wieder durch eine Korrelationsfunktion beschreiben. Jedoch geht die in der bisherigen Betrachtung beschriebene Autokorrelation in eine Kreuzkorrelation über, da sich jetzt das Empfangssignal $r(t)$ von der Demodulation $s(t)$ durch die Initialphase θ unterscheidet:

$$\begin{aligned} \varphi_{rc_T}(t) &= \frac{1}{2} \int_{-\infty}^{\infty} \text{rect}\left[\frac{\tau}{T}\right] \cdot e^{j(\pi k_r \tau^2 + \theta)} \cdot \text{rect}\left[-\frac{t-\tau}{T}\right] \cdot e^{-j\pi k_r (\tau-t)^2} \cdot d\tau \\ &= \frac{1}{2} \cdot e^{j\theta} \cdot e^{-j\pi k_r t^2} \cdot \int_{-\infty}^{\infty} \text{rect}\left[\frac{\tau}{T}\right] \cdot \text{rect}\left[\frac{\tau-t}{T}\right] \cdot e^{j2\pi k_r t \tau} \cdot d\tau . \end{aligned} \quad \text{Gl. 5.20}$$

An diesem Ausdruck wird deutlich, dass die Einhüllende der Korrelationsfunktion unverändert bleibt, jedoch die Grundfrequenz ist mit einem zusätzlichen Phasenoffset belegt (**Abb. 5.9**). In der Korrelationsfunktion eines Chirp-Signals prägt sich nach Gl. 5.12 nur der Realteil des Phasenoffsets aus, so dass das Übertragungsverhalten der Initialphasenmodulation nichtlinearen Charakter besitzt und dem Verlauf einer Cosinus-Schwingung entspricht (**Abb. 5.10**).

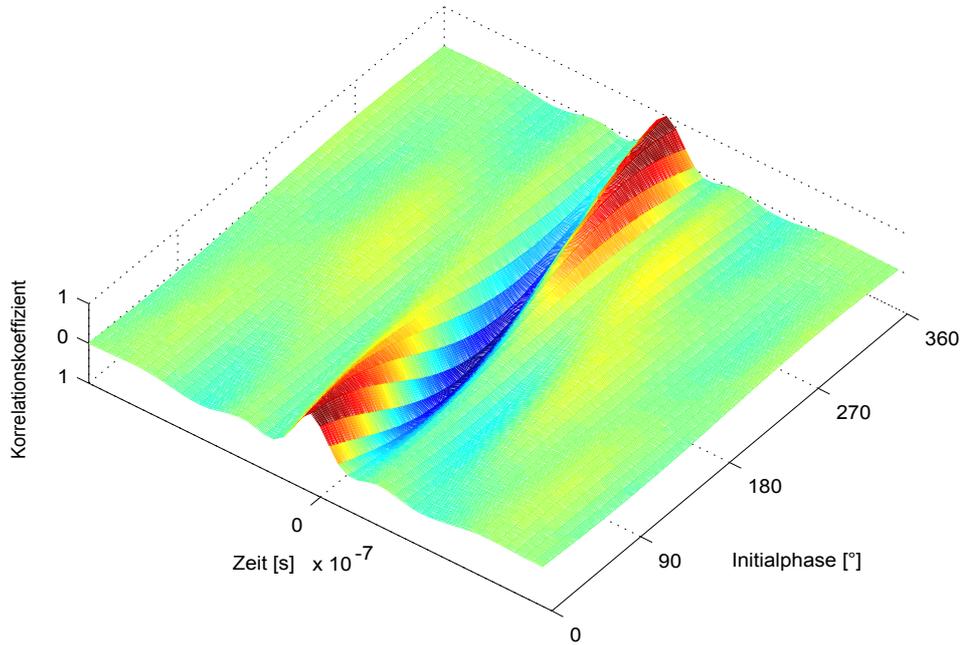


Abb. 5.9: Kreuzkorrelationsfunktionen von empfangenem Chirpsignal und Referenzchirp für unterschiedliche Initialphasen θ

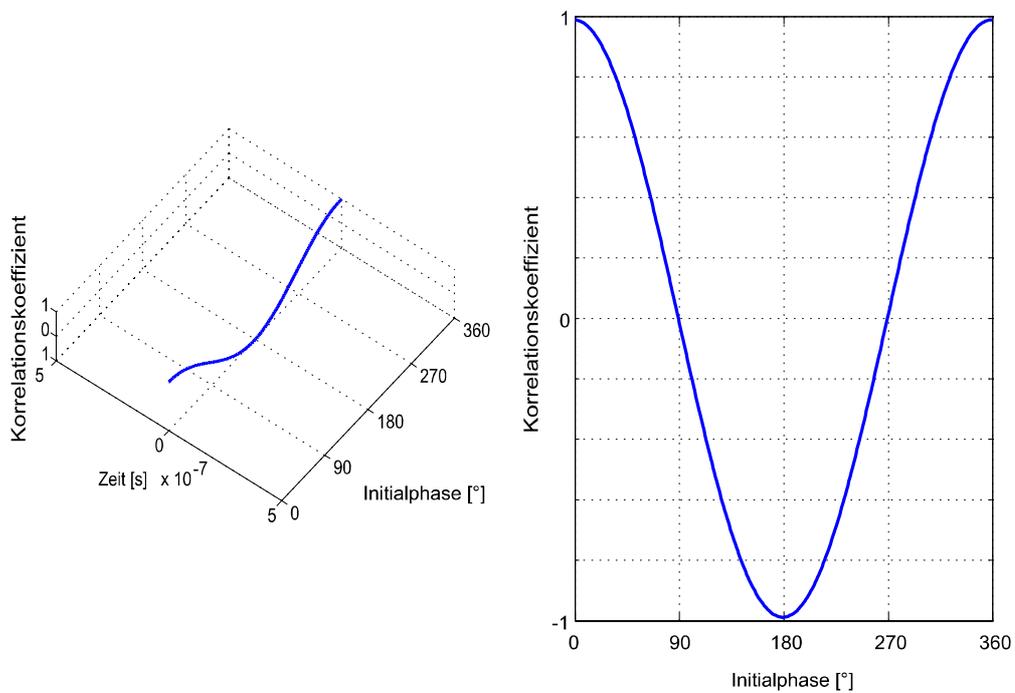


Abb. 5.10: Simulierte Übertragungsfunktion eines Korrelationsempfängers für initialphasenmodulierte Chirp-Signale

Diese Übertragungsfunktion lässt sich sehr einfach als Schnitt der Kreuzkorrelationsfunktion entlang der Koordinatenachse der Initialphase ($\theta = 0^\circ$) verstehen. Die Initialphasenmodulation erscheint somit für die Übertragung analoger Signale weniger geeignet, da das nichtlineare Verhalten durch ein weiteres Linearisierungsglied zu korrigieren wäre. Wie zuvor dargestellt ist diese Modulationsart jedoch hervorragend zur Übertragung binärer Datenströme geeignet.

5.4.2 Bedeutung der Amplitudenmodulation

Für die Betrachtung einer Amplitudenmodulation ist der allgemeine Ausdruck eines Chirp-Signals um den Faktor A zu erweitern:

$$s_i(t) = A \cdot \left[\cos(\theta + 2\pi \cdot f_0 \cdot t + 2\pi \cdot k_i \cdot t^2) \right]. \quad \text{Gl. 5.21}$$

Dieser repräsentiert die Energie des ausgesendeten Signals $s(t)$:

$$A = \sqrt{\frac{2 \cdot E}{T}}. \quad \text{Gl. 5.22}$$

Reicht die durch den Korrelationsempfang detektierte Signalenergie nicht aus, um eine geforderte Bitfehlerrate zu erreichen, dann besteht mit dieser Größe eine Möglichkeit, das Gesamtsystem an die Verluste im Übertragungskanal durch eine Erhöhung der Sendeleistung anzupassen. Anderenfalls kann aber auch die Sendeleistung reduziert werden, wenn die Kanalverluste gering ausfallen. Die reduzierte Sendeleistung verringert einerseits die Interferenzen mit den Einzelübertragungskanälen anderer Nutzer. Andererseits kann auch die gewonnene Leistungsreserve anderen Nutzern zur Verfügung gestellt werden. Die Amplitudenmodulation stellt somit eine Möglichkeit dar, die Übertragungskapazität des Gesamtkanals an die momentanen Verhältnisse dynamisch anzupassen und so der Forderung einer ökonomischen Ressourcenverwaltung nachzukommen.

5.4.3 Bedeutung der Frequenzmodulation

Die Modulation der Frequenz übernimmt bei Chirp-Signalen, wie in Kap. 5.2 dargestellt, die Funktion der Bandspreizung. In Kommunikationssystemen kommt ihr noch eine weitere Aufgabe zu, der der Kanaltrennung. Bei linearen Chirp-Signalen können allein zwei verschiedene Kanäle dadurch unterschieden werden, dass die Chirp-Rate k_r (Gl. 5.1) mit unterschiedlichen Vorzeichen behaftet wird (Abb. 5.11 u. Abb. 5.12).

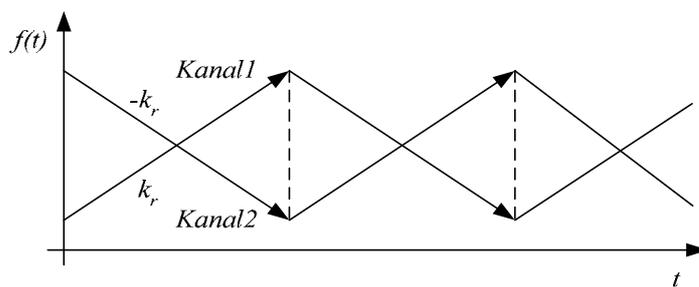


Abb. 5.11: Kanaltrennung durch entgegengesetzte Chirp-Raten k_r (Frequenz-Zeit-Verlauf)

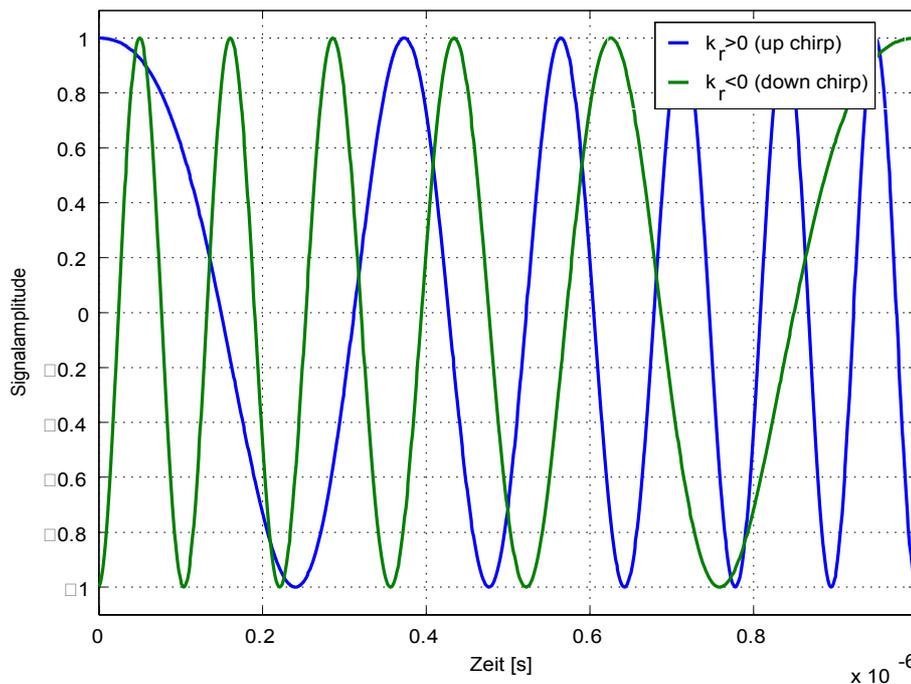


Abb. 5.12: Chirp-Signale mit entgegen gesetzten Chirp-Raten k_r im Zeitbereich

Verwendet man zwei Chirp-Raten k_{r1} und k_{r2} , die sich nicht im Vorzeichen, sondern nur in ihrem Betrag unterscheiden, dann kann auch hierüber eine Kanaltrennung erfolgen (**Abb. 5.13**). Diese erfolgt hier durch unterschiedliche konstante Chirp-Raten k_r . Noch wesentlich mehr Flexibilität lässt sich durch den Einsatz variabler Chirp-Raten erreichen. Hierbei stellt die Chirp-Rate k_r selbst eine Funktion der Zeit dar:

$$k_r = f(t). \quad \text{Gl. 5.23}$$

Hierbei kann, wie in **Abb. 5.14** angedeutet, ein gestörter Bereich des Übertragungsbandes zusätzlich unterdrückt werden. Grundsätzlich ist bei der Verwendung variabler Chirp-Raten zu beachten, dass die Korrelationsfunktion ebenso beeinflusst und damit im Vergleich zu einer mit konstanten k_r verformt wird.

5.4.4 Systemtheoretische Betrachtung des chirpbasierten Mehrfachzugriffs

Die Qualität eines Kommunikationsverfahrens mit der Möglichkeit des Mehrfachzugriffs lässt sich nur durch eine Untersuchung der verschiedenen Störeinflüsse beurteilen, sei es zwischen den Subkanälen untereinander oder mit beliebigen Rauschquellen.

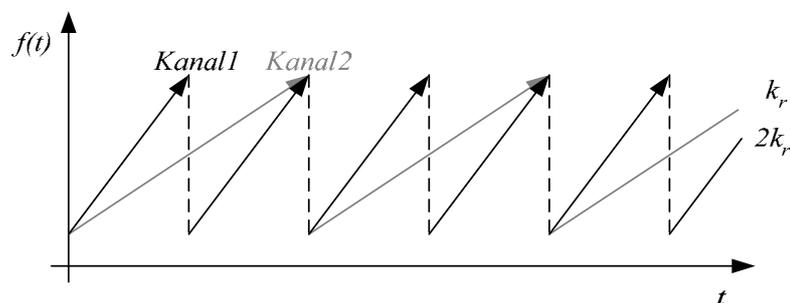
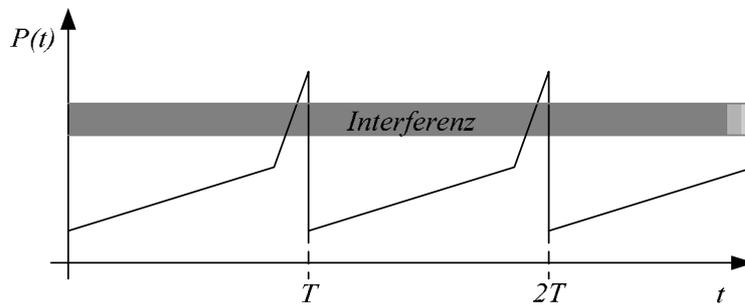


Abb. 5.13: Kanaltrennung durch unterschiedlich große Chirp-Raten k_r bzw. $k_1=2 \cdot k_2$ (Frequenz-Zeit-Verlauf)

a)



b)

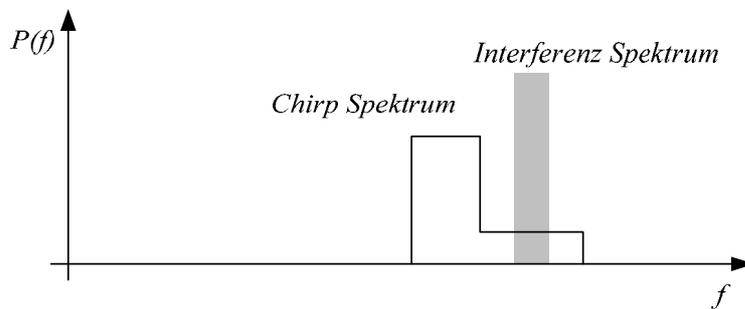


Abb. 5.14: Zusätzliche Störunterdrückung mittels variabler Chirp-Raten k_r im a) Zeit- und b) Frequenzbereich

Betrachtet man die Übertragung eines in der Initialphase binär kodierten Chirp-Signals s_i innerhalb eines Subkanals, dann erscheint dieses am Empfänger um die Laufzeit τ verschoben:

$$s_i(t, b) = \sqrt{\frac{2 \cdot E}{T}} \cdot \left[\cos \left(b \cdot \theta + 2\pi \cdot f_0 \cdot (t - \tau_i) + 2\pi \cdot k_r \cdot (t - \tau_i)^2 \right) \right] ; b \in \{0;1\}. \quad \text{Gl. 5.24}$$

Der Empfänger detektiert jedoch ebenso alle Signale der anderen Teilnehmer bzw. Subkanäle. Berücksichtigt man nun noch beliebige Rauschquellen durch zusätzliches weißes Gauss-Rauschen $n(t)$, dann setzt sich das insgesamt detektierte Signal entsprechend additiv zusammen:

$$r(t) = \sum_{i=1}^M s_i(t) + n(t). \quad \text{Gl. 5.25}$$

Die Korrelation des Empfangssignals $r(t)$ mit dem Referenzsignal $c_j(t)$ des Kanals j erfolgt durch eine Mischung bzw. Demodulation und Integration über das Zeitintervall $\{t_0; t_1\}$:

$$u_j(\tau) = \frac{1}{\sqrt{E}} \int_{t_0}^{t_1} c_j(t) \cdot r(t - \tau) dt . \quad \text{Gl.5.26}$$

Das Ergebnis der Korrelation für den Empfänger j ergibt sich aus:

$$u_j(\tau) = \frac{1}{\sqrt{E}} \int_{t_0}^{t_1} c_j(t) \cdot \left(\sum_{i=1}^M s_i(t - \tau) + n(t) \right) dt . \quad \text{Gl. 5.27}$$

Das Ergebnis ist die Summe der einzelnen Korrelationen des Demodulationssignals mit dem Sendesignal des eigenen Kanals j , den Sendesignalen s_i ($i \neq j$) aller anderen Subkanäle und dem Rauschprozess $n(t)$:

$$\begin{aligned} u_j(\tau) = & \frac{2\sqrt{E}}{T_j} \int_{t_j}^{t_j+T_j} \cos(2\pi f_0 t + 2\pi k_j t^2) \cdot \cos(2\pi f_0 (t - \tau_j) + 2\pi k_j (t - \tau_j)^2 + b_j \theta) dt \\ & + \frac{2\sqrt{E}}{T_j} \int_{t_j}^{t_j+T_j} \cos(2\pi f_0 t + 2\pi k_j t^2) \cdot \sum_{\substack{i=0 \\ i \neq j}}^M c_i(b_i, t - \tau_j) dt \\ & + \frac{2\sqrt{E}}{T_j} \int_{t_j}^{t_j+T_j} \cos(2\pi f_0 t + 2\pi k_j t^2) \cdot n(t) dt \end{aligned} \quad \text{Gl. 5.28}$$

Die Autokorrelationsfunktion innerhalb eines Subkanals entspricht der in Kap. 5.3 hergeleiteten Gleichung 5.14. Die Kreuzkorrelationen zwischen allen Subkanälen hängen sehr stark von der gewählten Art der Kanaltrennung ab. Im einfachsten Fall kann diese durch ein Zeitmultiplexverfahren erfolgen, in dem jeder Subkanal die gleichen Start- und Stopfrequenzen wie auch Chirp-Raten verwendet. Das Verhältnis von Signalleistung eines Kanals zu den Störungen durch die anderen Subkanäle ergibt sich aus der Überlagerung aller Korrelationsfunktionen (**Abb. 5.15**). Diese einfache Möglichkeit beschränkt jedoch wesentlich die Flexibilität des Übertragungsverfahrens.

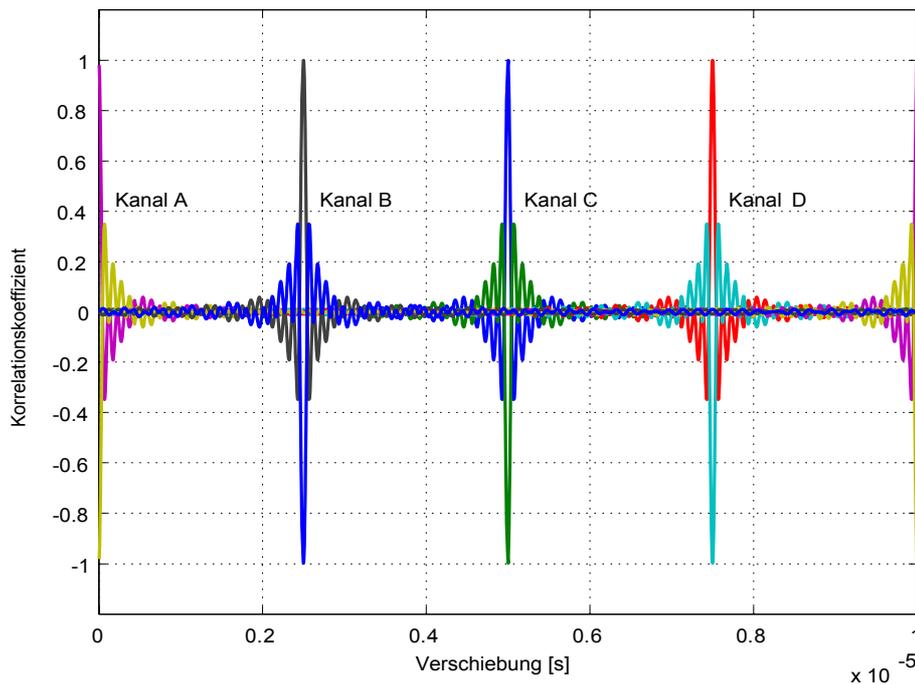


Abb. 5.15: Kanaltrennung mittels Zeitmultiplex von 4 initialphasenkodierten Chirpsignalen

Eine flexible Lösung erfordert aber einen sehr umfangreichen Rechenaufwand, der die Kreuzkorrelationen aller gerade aktiven Subkanäle mit ihren aktuellen Parametern (Start-/Stopfrequenz u. Chirp-Rate) berücksichtigt.

5.4.5 Bedeutung der mehrdimensionalen Parametervariation

Das Potential eines chirpbasierten Übertragungsverfahrens lässt sich sehr anschaulich anhand des Informationswürfels nach Shannon erklären (**Abb. 5.16**). Deutlich ist hier das Zeit-Bandbreiten-Produkt als Grundfläche des Würfels zu erkennen. Die gesamte Signalenergie entspricht hierbei seiner rückseitigen Fläche. Da Spread-Spektrum-Verfahren nun ihre Signalenergie über ein Frequenzintervall $\{f_1; f_2\}$ spreizen, entsteht ein Quader, dessen Seitenfläche in Analogie zum Zeit-Bandbreiten-Produkt mit Frequenz-Leistungs-Produkt bezeichnet werden kann.

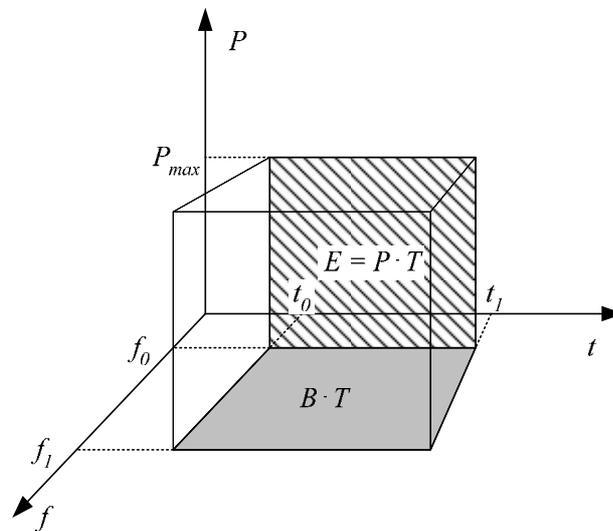


Abb. 5.16: Informationswürfel nach Shannon für ein Spread-Spectrum-Verfahren

Ein interessanter Aspekt dieser Betrachtung findet sich in einer weiteren Analogie. Wenn die Bandbreite B die Frequenzspreizung repräsentiert, dann lässt sich ebenso für das Zeitintervall T der Begriff Zeitspreizung formulieren. Die Verwendung dieses Begriffs erscheint dahingehend sinnvoll, dass eine Variation der Chirp-Rate k_r nur einen Einfluss auf das Zeitintervall T nimmt und nicht auf die Bandbreite B . Trotzdem wird die Übertragungscharakteristik entscheidend gestört, da das Signal dadurch über ein größeres Zeitintervall verteilt (in dem Sinne also zeitlich gespreizt) wird.

Eine wichtige Größe zur Beschreibung eines Übertragungsverfahrens ist die Kanalkapazität C . Für ein binär kodiertes System ergibt sie sich aus dem Verhältnis von Signal- und Rauschleistung P_S/P_N :

$$C = B \cdot \lg \left(1 + \frac{P_S}{P_N} \right) \quad [\text{bit} / \text{s}]. \quad \text{Gl. 5.29}$$

Anzumerken ist hierbei, dass dieser Ausdruck einen rauschenden Hintergrund berücksichtigt, der über das gesamte Spektrum gleichverteilt ist. Sollen Rauschquellen berücksichtigt werden, die nicht dieser Voraussetzung genügen, dann sind deren Signalenergien ($E_N = P_N \cdot T_N$ und $E_S = P_S \cdot T_S$) zu betrachten:

$$C = B \cdot \text{ld} \left(1 + \frac{E_S}{E_N} \right) \quad [\text{bit} / \text{s}]. \quad \text{Gl. 5.30}$$

Die Informationsmenge eines solchen Übertragungsverfahrens ergibt sich direkt aus dem Produkt von Kanalkapazität C mit dem Intervall der Zeitspreizung T :

$$I = \int_{t_1}^{t_2} C \cdot dt = B \cdot (t_2 - t_1) \cdot \text{ld} \left(1 + \frac{E_S}{E_N} \right) \quad [\text{bit}]. \quad \text{Gl. 5.31}$$

Der Informationswürfel für ein solches durch Rauschen begrenztes Modell erhält somit das in **Abb. 5.17** dargestellte Aussehen. Das verbleibende Volumen des Informationswürfels stellt nun die gesamte zur Verfügung stehende Ressource des Übertragungsverfahrens dar, die es durch die Variation aller Signalparameter ökonomisch zu verteilen gilt. Die Vielfältigkeit dieser Aufgabe soll nun anhand eines einfachen Beispiels erläutert werden, der den Mehrfachzugriff mittels Zeitmultiplex ermöglicht und zusätzlich für jeden Kanal die Signalamplitude variiert. Ausgehend von 4 Nutzern ergibt sich entsprechend dem klassischen Zeitmultiplex die schon in **Abb. 5.15** gezeigte Kanalaufteilung.

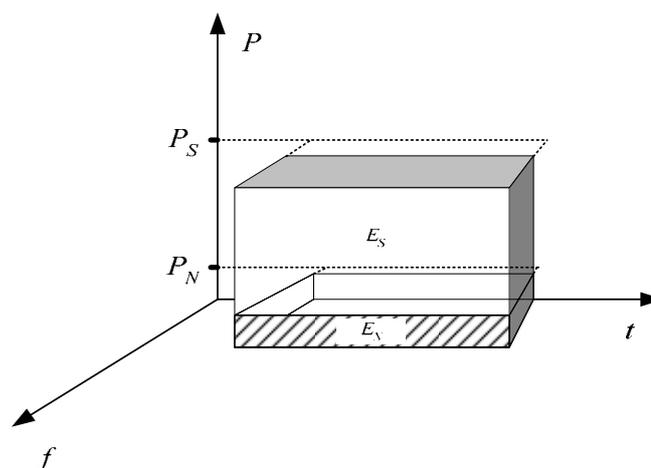


Abb. 5.17: Informationswürfel eines rauschbegrenzten Spread-Spectrum-Übertragungsverfahrens

Wie diese Darstellung zeigt, interferieren die zueinander verschobenen Chirp-Signale miteinander, da die Korrelation zweier gleichartiger Chirp-Signale trotz einer großen zeitlichen Verschiebung nicht ganz Null ist. Nun ist jedoch der Übertragungskanal zum Nutzer 2 (z.B. aufgrund einer größeren Entfernung) gestört, wobei dieser zusätzlich aber eine hohe Übertragungssicherheit benötigt. Die restlichen Kanäle dagegen unterliegen keinen Störeinflüssen und benötigen nur eine geringe Übertragungssicherheit (wie z.B. für die Übertragung eines Sprachsignals). In diesem Fall ergeben sich zwei Möglichkeiten zur dynamischen Anpassung des gesamten Systems. Einerseits kann die Signalleistung für Nutzer 2 soweit erhöht werden, wie sie sich für die restlichen reduzieren lässt (**Abb. 5.18**). Andererseits können aber auch die Zeitverschiebungen zwischen den einzelnen Kanälen derart gewählt werden, dass die Interferenz mit Nutzer 2 verringert wird. Hierzu sollte der Zeitabstand zu allen anderen Kanälen maximal werden (**Abb. 5.19**). Eine Kombination aus beiden Varianten ist natürlich ebenso möglich (**Abb. 5.20**). Neben diesem einfachen Beispiel ergeben sich noch wesentlich mehr Möglichkeiten an Parametervariationen. So müssen nicht zwangsläufig einfache lineare Chirp-Signale verwendet werden. Mittels variabler Chirp-Raten können auch ohne Zeitmultiplex mehrere Subkanäle sicher unterschieden werden. Dieses Verfahren besitzt den Vorteil, dass es ohne definierte Zeitfenster auskommt und alle Subkanäle wirklich gleichzeitig bzw. parallel bedient.² Da eine solch variable Chirp-Rate natürlich die Autokorrelationsfunktion beeinflusst, soll im Folgenden auf optimierte Chirp-Signale eingegangen werden.

² S. K. Machineni et. al; „On the use of wideband time- varying signalling for multiuser wireless communications”, 4th IEEE Workshop on Signal Processing, Advances in Wireless Communications, S.60-64, 2003

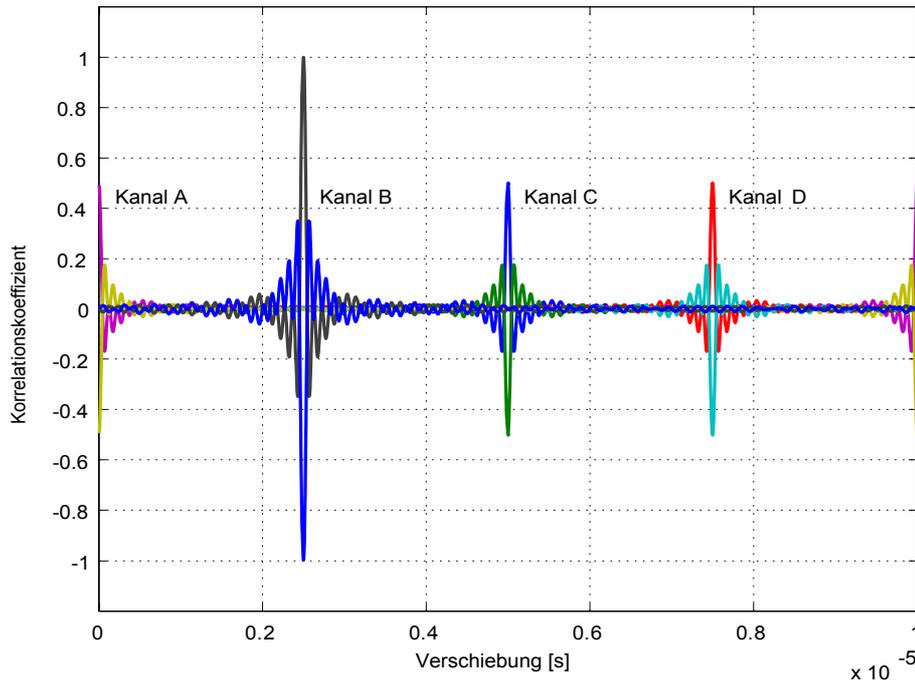


Abb. 5.18: Dynamische Anpassung des Übertragungsverfahrens mittels Variation der Signalleistung P_i

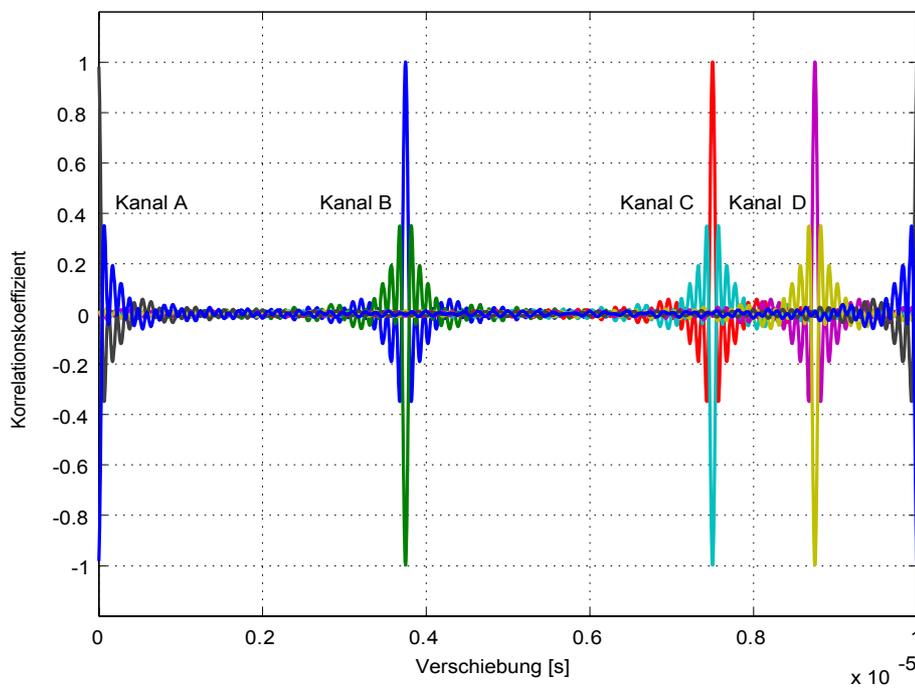


Abb. 5.19: Dynamische Anpassung des Übertragungsverfahrens mittels Variation der Zeitfenster T_i

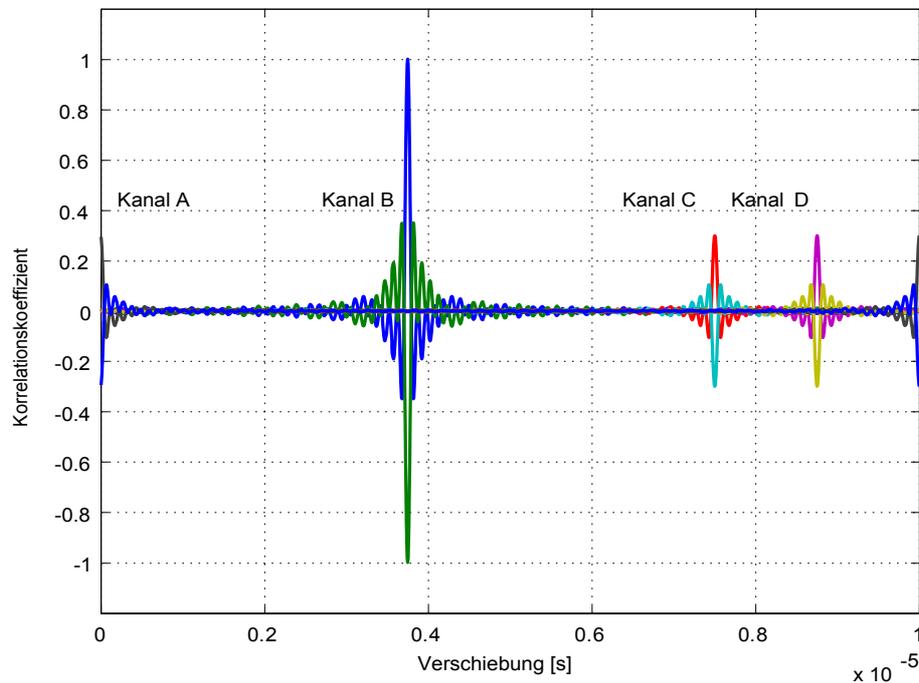


Abb. 5.20: Dynamische Anpassung des Übertragungsverfahrens mittels Variation von Signalleistung P_i und Zeitfenster T_i

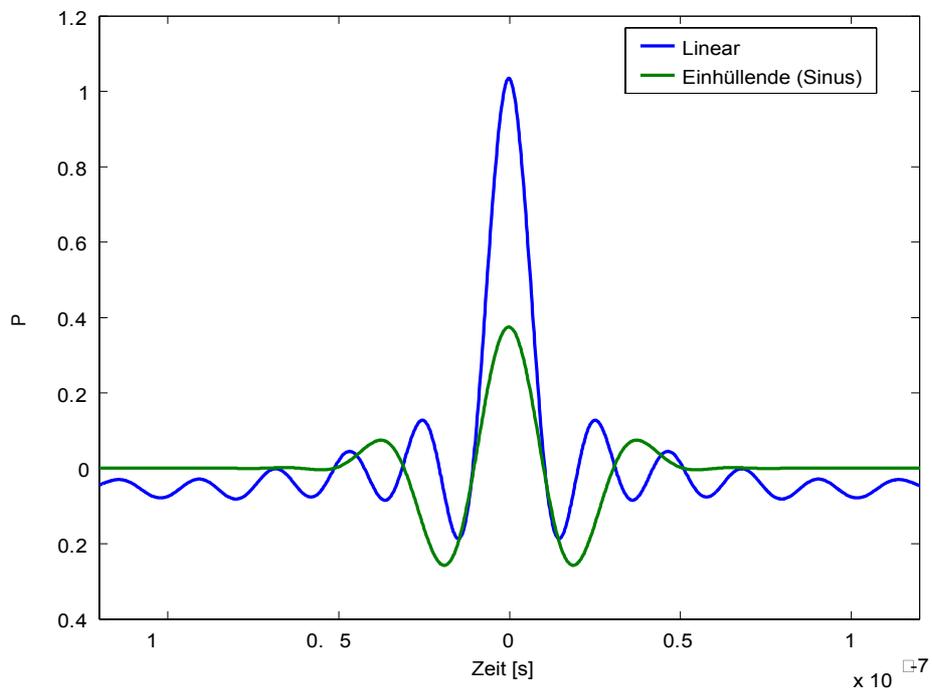
5.5 Optimierte Chirp-Signale

Für die vielfältigen Möglichkeiten des Ressourcen-Managements in einem Übertragungsverfahren mit Mehrfachzugriff kann es sinnvoll sein, von der bisher beschriebenen einfachen Form eines Chirp-Signals nach **Gl.5.1** abzuweichen. Die Idee hierfür entstand in Anlehnung an die in **Kap. 5.4.3** dargestellte Ausblendung eines gestörten Bandbereichs durch verschiedene Chirp-Raten. Dabei werden natürlich die spektrale Verteilung und die damit zusammenhängende Autokorrelationsfunktion des verwendeten Chirp-Signals mit beeinflusst. Zur Vermeidung von Interferenzen zwischen mehreren Subkanälen ist jede Variation interessant, die die Autokorrelationsfunktion näher an einen idealen Dirac-Stoß bringt. Dies kann man dadurch erreichen, dass der Chirp beispielsweise im Zeitbereich mit einer Einhüllenden versehen wird.

In einer systemtheoretischen Beschreibung eines realen Systems wird ein idealer Dirac-Stoß durch eine Si-Funktion gleichgesetzt, da ein reales System nie die notwendige unendliche Bandbreite besitzen kann, die ein idealer Dirac-Stoß erfordert. Grenzt man jedoch das zu dieser Si-Funktion gehörende rechteckförmige Spektrum mit einer Sinuswelle ein (**Abb. 5.21 a**), dann erhält man einen realen Impuls, der nun näher am Ideal des Dirac-Stoßes liegt (**Abb. 5.21 b**). Zwar ist in der Zeitdarstellung die Amplitude des Impulses etwas geringer, und es ergeben sich zusätzlich größere Überschwinger in direkter Nachbarschaft, jedoch zeigt sich der restliche Impuls deutlich näher am Ideal der Nulllinie.

Da nach dem Wiener-Kintche-Theorem ebenso ein Zusammenhang zwischen Korrelation und Spektrum besteht, liegt der Gedanke nahe, Signalspektren so zu optimieren, dass sich ein gleicher oder ähnlicher Effekt in der Autokorrelation ergibt. Würde dies gelingen, dann hätte man die Interferenzeffekte zwischen den Subkanälen im zuvor beschriebenen chirpbasierten TDMA-Verfahren deutlich reduziert. Wendet man nun diesen Sinus-Shape direkt im Zeitbereich auf ein Chirp-Signal an (**Abb. 5.22 a**), dann ergibt sich die in **Abb. 5.22 b** dargestellte Korrelationsfunktion.

a)



b)

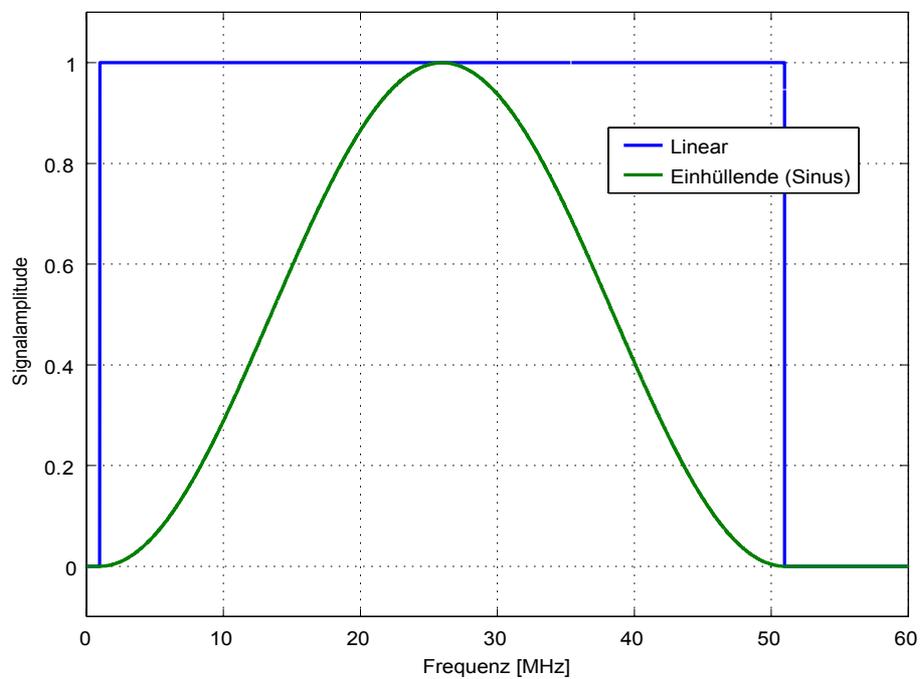
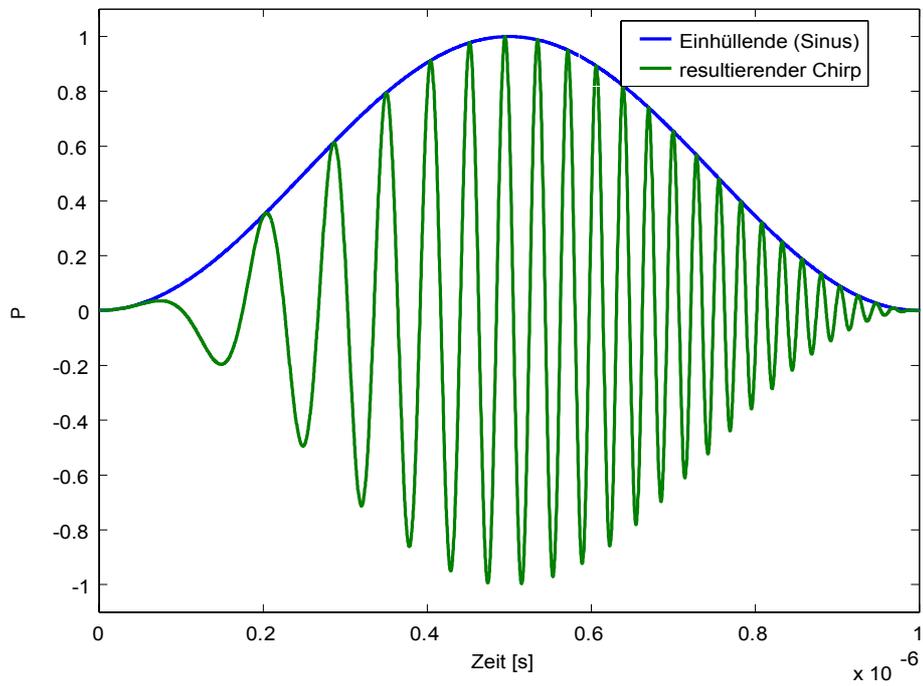


Abb. 5.21: Gegenüberstellung eines linearen Chirps bzw. Si-Impulses und eines Sinus-Shaped-Spectrum-Impulses im a) Zeit- und b) Frequenzbereich

a)



b)

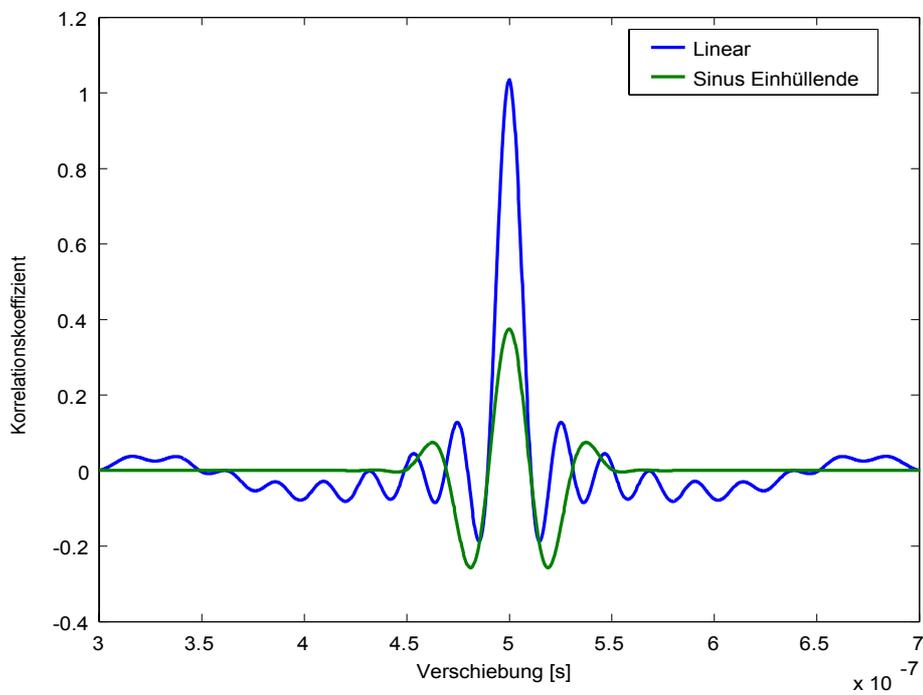
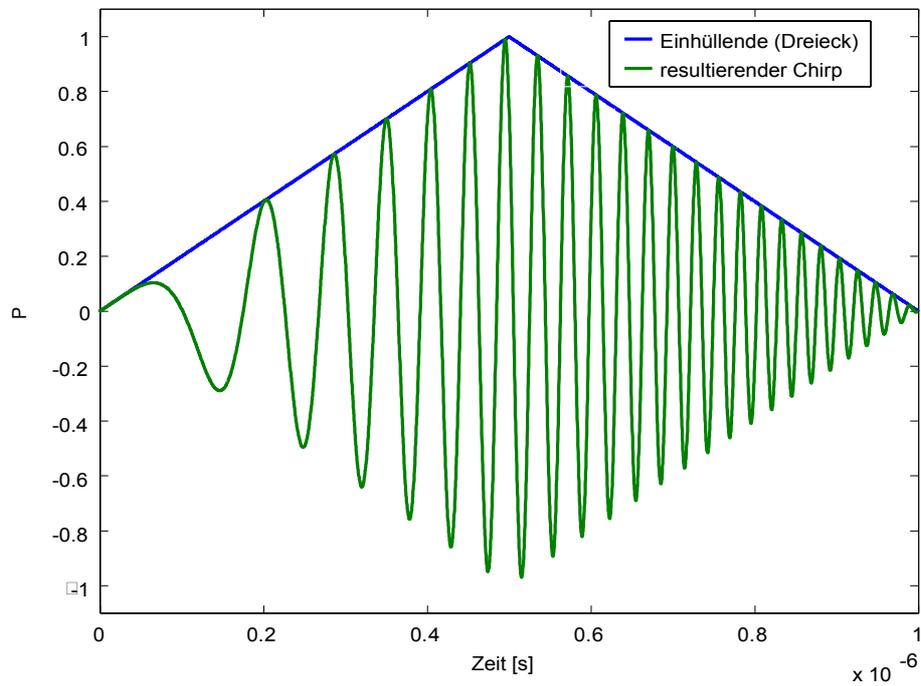


Abb. 5.22: Sinus-Shaped-Chirp-Signal im a) Zeitbereich und b) zugehörige Autokorrelationsfunktion

a)



b)

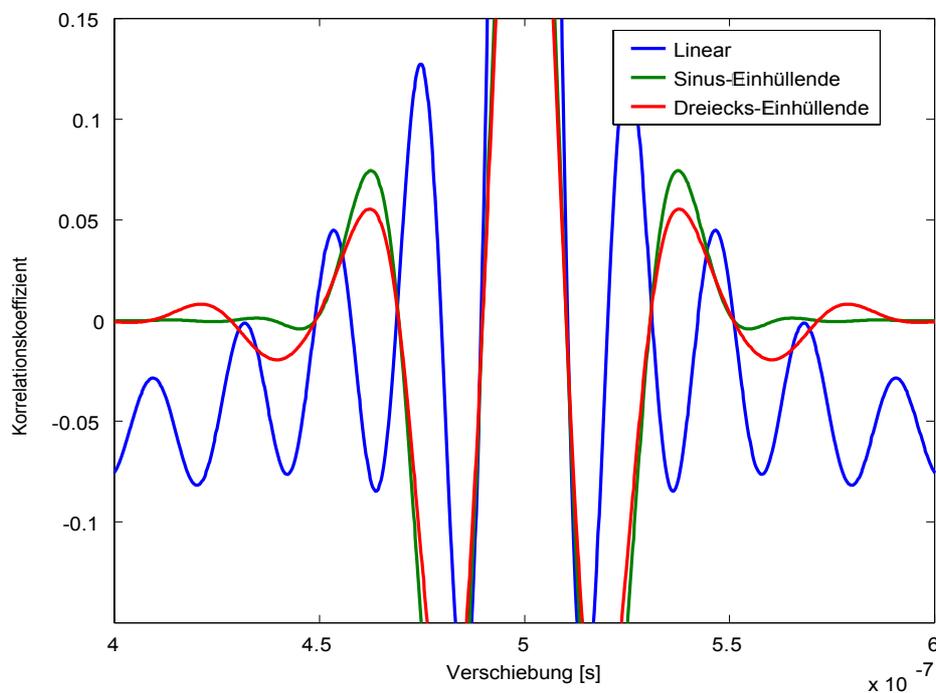


Abb. 5.23: Die Dreiecksfunktion als Einhüllende eines Chirp-Signals; a) Signalverlauf und b) Auswirkung auf die Impulsformung in der Autokorrelation gegenüber dem idealen Chirp und dem Sinus-Shaped-Chirp

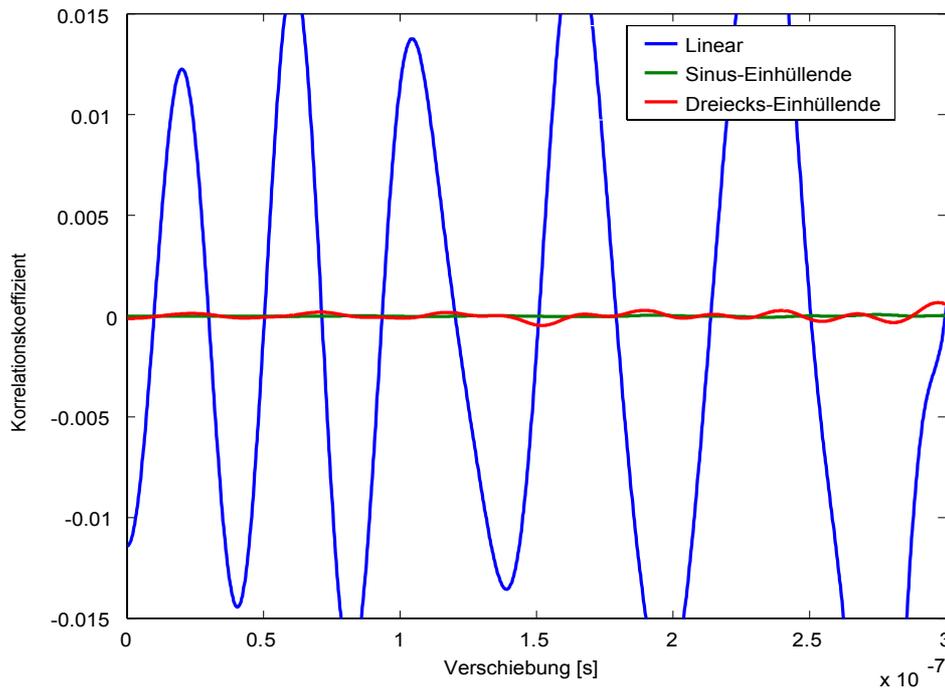


Abb. 5.24: Restwelligkeit der Autokorrelation außerhalb des Impulsmaximums für vergleichbare Signale (ohne Einhüllende/Linear, sinusförmige Einhüllende, dreieckförmige Einhüllende)

Zwar sinkt darin die Amplitude des Impulses in der Autokorrelation auf 36%, aber im danebenliegenden Bereich verringert sich die Abweichung von der Nulllinie ungleich mehr auf etwa 8% (**Abb. 5.24**). Damit lassen sich nun die Interferenzen zwischen den Subkanälen eines chirpbasierten TDMA-Verfahrens deutlich reduzieren. Weitere Varianten an Shaped-Chirp-Signalen ergeben sich unter Verwendung unterschiedlicher Einhüllenden. Beispielphaft soll hier noch kurz auf die Dreiecksfunktion eingegangen werden (**Abb. 5.23 a**), welche noch etwas günstigere Eigenschaften ausweist. Zwar ist hier die Amplitude des Impulses noch geringfügig stärker gedämpft (auf 32%), nähert sich jedoch wesentlich schneller der Nulllinie an (**Abb. 5.23 b**). Insgesamt bleibt im Vergleich zum Cosinus-Shape jedoch außerhalb des eigentlichen Impulses eine höhere Grundwelligkeit erhalten (**Abb. 5.24**). Anhand dieser Beispiele wird deutlich, mit welcher Vielfalt Einfluss auf die Charakteristik eines Chirp-Signals genommen werden kann, um in einem chirp-basierten Übertragungsverfahren dynamisch auf Störungen im Übertragungskanal reagieren zu können.

5.6 Korrelationsempfang optischer Chirp-Signale mittels PMD

Für den Korrelationsempfang eines optischen Signals kann auf das unter Kap. 5.2 beschriebene Prinzip zurückgegriffen werden. Will man hierbei jedoch mit Chirp-Signalen arbeiten, dann ist die Systemtechnik mit entsprechenden Signalgeneratoren auszustatten. In der Funktechnik werden hierzu sogenannte DDL (Dispersive Delay Line) eingesetzt, wie z.B. Oberflächenwellenfilter (SAW - Surface Acoustic Wave Filter).³ In dem hier beschriebenen Testaufbau wurden jedoch Direct Digital Synthesizers (DDS) eingesetzt, mit denen sehr variabel verschiedenste Signalformen erzeugt werden können. Gerade für Testaufbauten stellen diese sehr flexible Generatoren dar, die in Frequenz und Phase hochgenaue Signale zur Verfügung stellen. Mit ihnen lassen sich alle Parameter eines Chirp-Signals frei wählen bzw. programmieren, so dass hiermit ebenso die direkte Abbildung der Information $b_i(t)$ auf die Initialphase des Chirp-Signals sehr einfach möglich ist. Eine funktionelle Übersicht der realisierten Kommunikation auf der Basis optischer Chirp-Signale und eines PMD-Empfängers zeigt **Abb. 5.25**.

Entscheidender Vorteil einer Kommunikation mittels PMD-basierten Korrelationsempfangs liegt in der einfachen Bitentscheidung. Hierzu wird nur ein Vergleich der analogen Ausgänge beider Integratoren benötigt, wodurch unabhängig vom Fremdlicht wieder ein binäres Datensignal zurück gewonnen wird. Anzumerken ist hierbei, dass der empfängerseitige Chirp-Generator nicht in seiner Initialphase variiert wird. Hier wird diese Möglichkeit nicht benötigt, da hier die Dateninformation durch eine Korrelation mit einem festen Referenz-Chirp $c_i(t)$ ($\theta = const.$) ausgewertet wird. Die Entscheidung, LED im vorgestellten Aufbau als Sendeelement einzusetzen, liegt darin begründet, dass ihre Emissionsfläche gegenüber Laserdioden eine größere Ausdehnung besitzen und dadurch weniger der Problematik der Augensicherheit unterliegen. Des Weiteren liefern LED eine relativ breite Abstrahlkeule und unterliegen aufgrund der inkohärenten Abstrahlung keinen Interferenzeffekten. Nachteilig ist jedoch ihre geringere Strahlungseffizienz, wobei hier jedoch anzumerken ist, dass aufgrund der Entwicklungen im Bereich der kommerziellen Infrarotkommunikation (z.B. IRDA) die Entwicklung effizienterer LEDs einen enormen Schub bekommen hat.

³ Wolfgang Hascher, „nanoNET: sichere Verbindung für Sensor-/Aktor-Netzwerke der Zukunft“, Elektronik 22/2002, Weka Fachzeitschriftenverlag 2004, S. XX

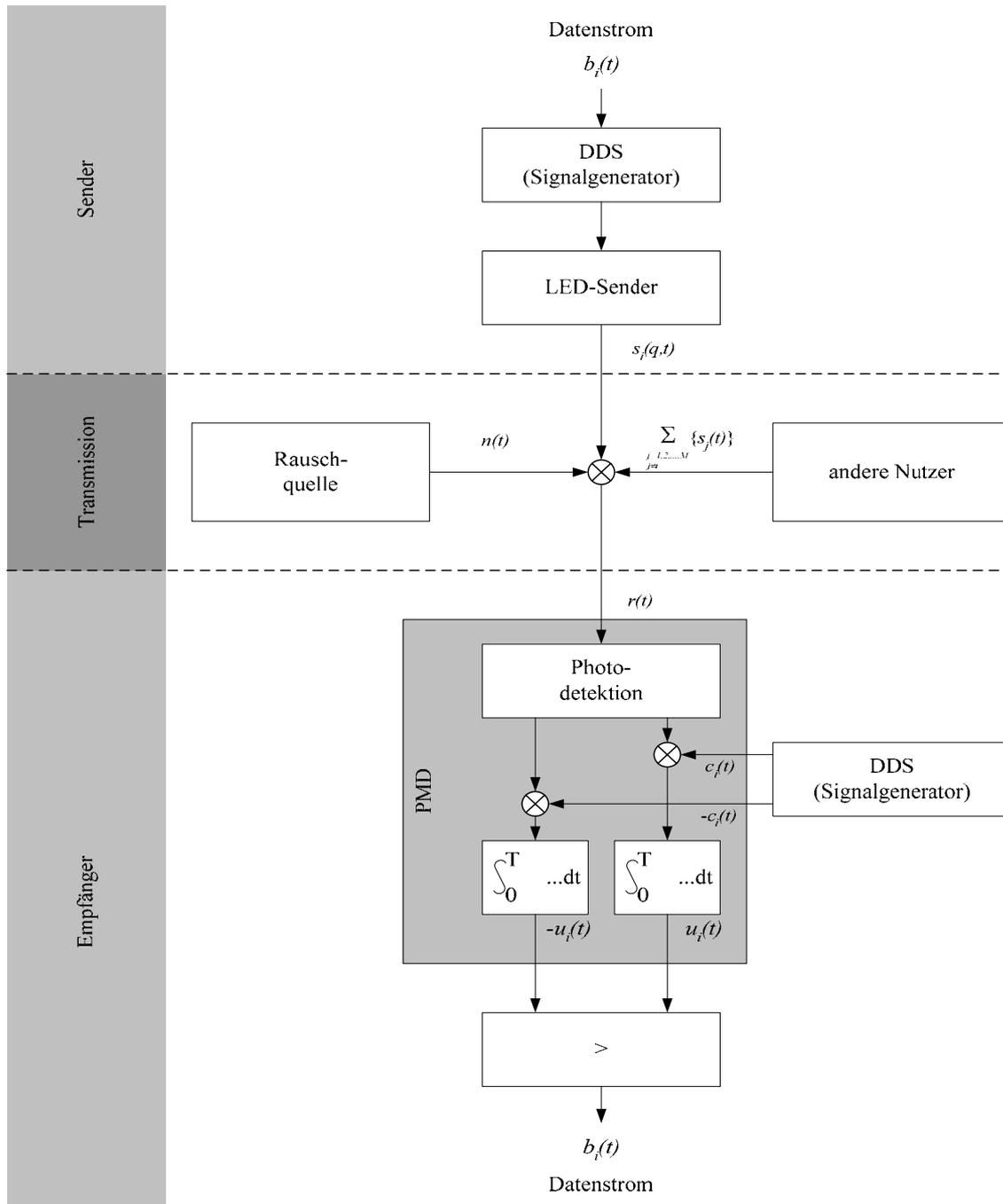
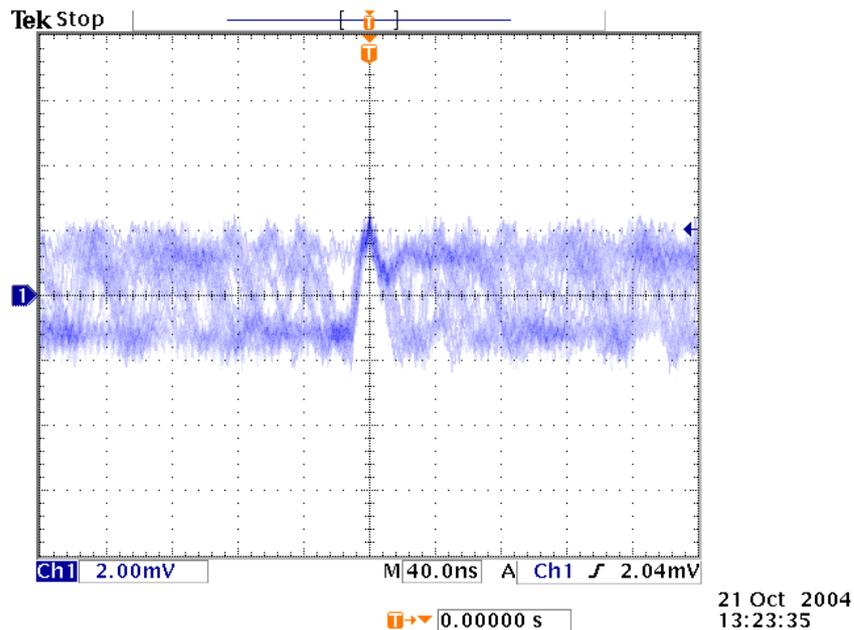


Abb. 5.25: Blockdiagramm der realisierten Kommunikationstrecke auf Basis optischer Chirp-Signale und eines PMD-Empfängers

Das im Aufbau realisierte optische Chirp-Signal zeigt **Abb. 5.26 a**, wobei dieses mit einem breitbandigem optischen Detektor aufgenommen wurde (**Abb. 5.26 b**).

a)



b)

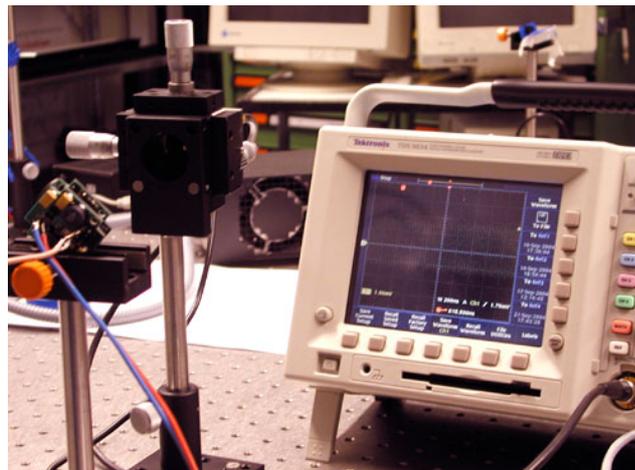
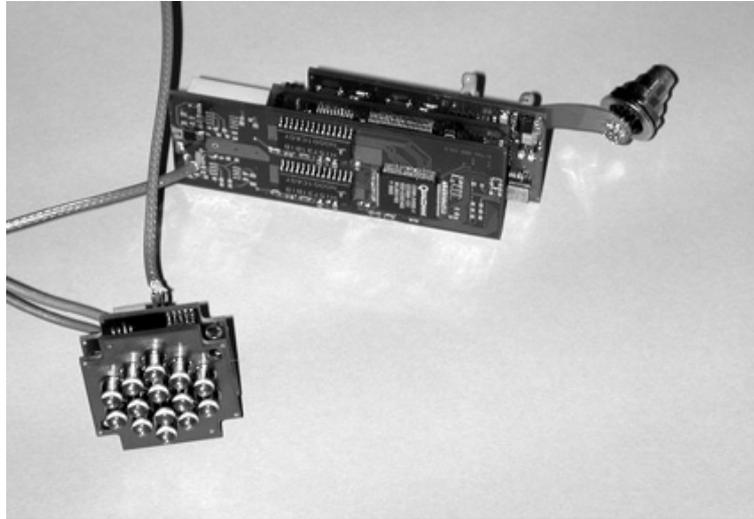


Abb. 5.26: Realisierter LED-Sender a) optisches Chirp-Signal b) Vermessung des optischen Signals mit breitbandigem Photodetektor

Die gesamte realisierte Hardware ist für die Sender- wie auch Empfängerseite in **Abb. 5.27** dargestellt. Der durch das empfangene optische Chirpsignal generierte Signalhub am den beiden Ausgängen des PMDs zeigt **Abb. 5.28** und **5.29** für unterschiedliche Bitkodierungen $b(t)=0$ und $b(t)=1$. Deutlich ist zur Erkennen wie in Abhängigkeit der Bitinformation das Vorzeichen des Differenzsignals aus beiden PMD-Ausgängen wechselt. Durch die Bandspreizung erweist sich dieser Aufbau neben seiner Einfachheit als ebenso robust gegenüber Störsignalen.

a)



b)



Abb. 5.27: Hardware des realisierten Demonstrators eines chirpbasierten Übertragungsverfahrens mittels PMD a) Emitter b) Receiver, jeweils mit Chirp-Generator zur Modulation bzw. Demodulation

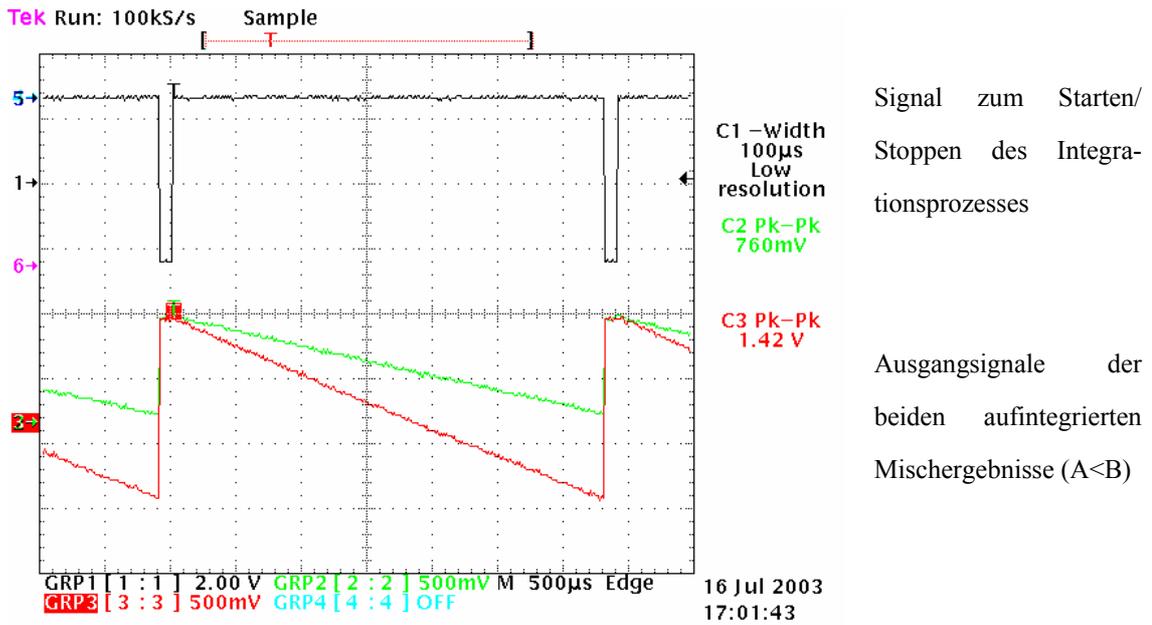


Abb. 5.28: Messung des Integrationsprozesses der Kreuzkorrelation von empfangenem Chirpsignal mit der Dateninformation $b(t)=1$ und Referenzchirp mittels eines Photogate-PMD's

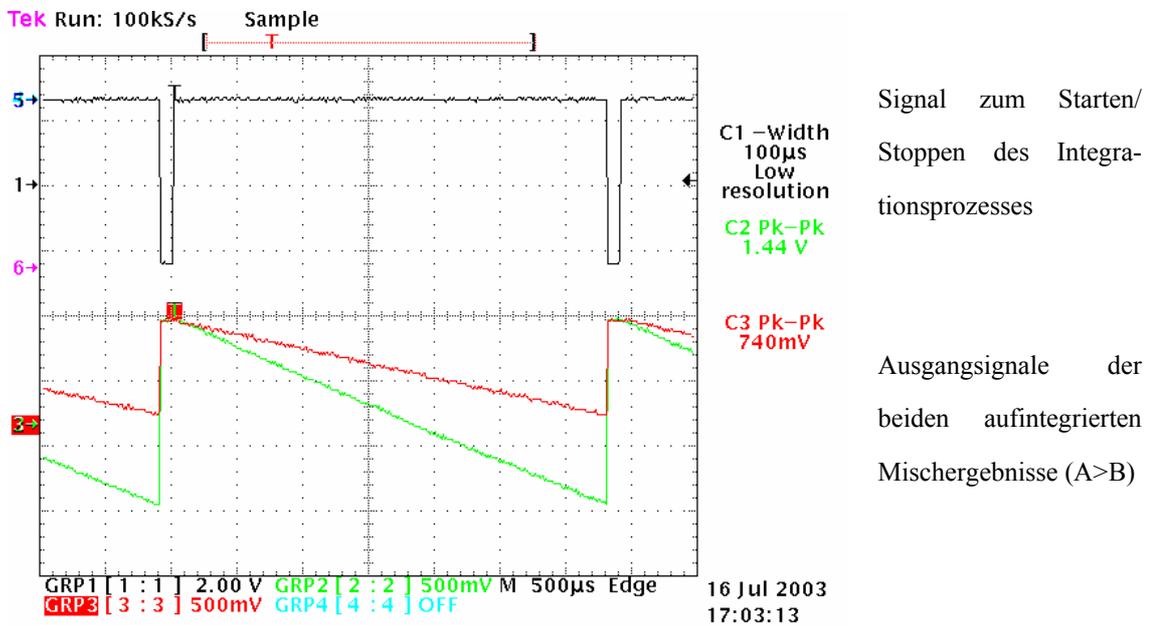


Abb. 5.29: Messung des Integrationsprozesses der Kreuzkorrelation zweier Chirp-Signale mit der Dateninformation $b(t)=0$

5.7 Fazit

Wie bisher ausführlich erläutert wurde, besitzen Chirp-Signale wesentliche Eigenschaften, die sie für zukünftige Kommunikationssysteme besonders geeignet erscheinen lassen. Stellt man sie den bisherigen Bandspreizverfahren gegenüber, dann ergeben sich die in **Tab. 5.1** aufgeführten Vor- und Nachteile. Aus der systemtheoretischen Beschreibung eines Chirp-Signals und der Bedeutung der darin verankerten grundsätzlichen Modulationsarten in Phase, Amplitude und Frequenz wird der Antagonismus zwischen der gewünschten Flexibilität in der Aufteilung der Gesamtbandbreite im Mehrbenutzerzugriff und dem damit verbundenen Steuerungsaufwand deutlich. Schließlich stellt jeder Subkanal schon ein dreidimensionales System dar, so dass mit jedem weiteren Teilnehmer dieses gesamte Übertragungsverfahren ein noch wesentlich komplexeres System darstellt, welches kontinuierlich gesteuert bzw. optimiert werden muss. Mit der enormen Rechenleistung zukünftiger Steuerungssysteme kann man erwarten, dass sich zukünftig auch diese notwendigerweise sehr komplexen Steuerungsstrategien in Echtzeit umsetzen lassen. Weiterhin können mit den so genannten optimierten Chirp-Signalen die Interferenzen zwischen den einzelnen Subkanälen reduziert werden, was die Attraktivität von chirp-basierten Übertragungsverfahren für zukünftige Kommunikationssystemen weiter erhöht. Wendet man diese Erkenntnisse auf ein optisches Kommunikationssystem an, dann ist man auf einen intelligenten optischen Empfänger angewiesen, der als Korrelationsempfänger einsetzbar ist. Diese Anforderung wird durch ein PMD hervorragend erfüllt, wie an dem hier präsentierten Versuchsaufbau deutlich geworden ist.

	Vorteil	Nachteil
Frequency Hopping (FS)	<ul style="list-style-type: none"> - relativ kurze Einstellzeit/ Synchronisation - kein Korrelationsempfänger notwendig 	<ul style="list-style-type: none"> - benötigt mehrere Frequenzgeneratoren - leicht im Spektrum zu erkennen - Datenverluste auf gestörter Frequenz möglich
Direct Sequence (DS)	<ul style="list-style-type: none"> - sehr gute Unterscheidung zwischen Mehrwegesignalen - sehr gute Stör- und Rauschunterdrückung - sehr schwierig zu entdecken und zu entschlüsseln 	<ul style="list-style-type: none"> - relativ lange Einstellzeit/ Synchronisation - benötigt zusätzlichen Codegenerator - breiter Frequenzbereich wird benötigt - Korrelationsempfänger notwendig - nur diskrete Parametervariationen möglich - begrenzte Anzahl an PN-Folgen beschränkt eine dynamische Verwaltung der Kanalkapazität
Frequency Chirping (FC)	<ul style="list-style-type: none"> - sehr gute Unterscheidung zwischen Mehrwegesignalen - sehr gute Stör- und Rauschunterdrückung - sehr schwierig zu entdecken - alle Parameter (Frequenz, Amplitude, Phase) können frei variiert werden 	<ul style="list-style-type: none"> - relativ lange Einstellzeit/ Synchronisation - benötigt zusätzlichen Chirp-Generator - breiter Frequenzbereich wird benötigt - Korrelationsempfänger notwendig

Tab. 5.1: Gegenüberstellung der wesentlichen Eigenschaften der verschiedenen Bandspreiztechniken

6 Zusammenfassung und Ausblick

Dieser Arbeit lag die Aufgabe zu Grunde, den praktischen Einsatz der PMD-Technologie zu prüfen und auftretende Probleme systemtheoretisch zu beschreiben. Mit den anfänglich diskutierten systemtheoretischen Grundlagen wurde die Basis für die durchgeführten Untersuchungen geschaffen, zumal die Grenzen einer solchen Beschreibung klar formuliert werden konnten. Besonders deutlich wurden diese Grenzen innerhalb der Diskussion des Unterschiedes zwischen Faltungstheorem und Korrelationsintegral. Hier konnte aufgezeigt werden, dass sich beide zwar in ihrem mathematischen Ausdruck vergleichen lassen, jedoch in ihren Auswirkungen und den resultierenden Rückschlüssen völlig unterschiedlich sind. Dies liegt darin begründet, dass sich die Faltung in einem Theoriemodell bewegt, das auf bestimmten Voraussetzungen beruht und sich somit nur auf einen speziellen Bereich der Wirklichkeit beschränkt. Erst die in der Systemtheorie geforderten Kausalität, Linearität und Zeitinvarianz eines Systems erlauben die Vorhersage eines Signalverlaufs durch die Anwendung der Faltung. Realisiert man jedoch einen Korrelationsempfänger, dann bewegt man sich außerhalb dieses abgegrenzten Bereiches und kann das Ausgangssignal nicht so ohne weiteres vorhersagen. Eine Korrelation kann schließlich nur die Ähnlichkeit zwischen zwei Funktionen, Messreihen, Signalverläufen oder ähnlichem bestimmen; eine Zuordnung von Ursache und Wirkung jedoch ist nicht möglich. Aber auch darüber hinaus bleibt die Betrachtung eines jeden optoelektronischen Systems eine anspruchsvolle Aufgabe, die ein fundiertes Wissen in vielen Wissenschaftsbereichen voraussetzt. Beispielsweise erlaubt erst eine quantentheoretische Betrachtung aufgrund der Teilchennatur des Lichtes eine Rauschanalyse in der Optoelektronik. Eine systemtheoretische Betrachtung allein würde hier nicht genügen.

Für die Signalphasenmessung wurde in dieser Arbeit auf das in der klassischen Interferometrie vielfach eingesetzte Phase-Shift-Verfahren zurückgegriffen. Neben den zu diesem Verfahren gehörigen Grundlagen wurde eine ausführliche Beschreibung der Auswirkungen der Fehler systematischer Natur oder statistischer Ursache innerhalb dieses Approximationsverfahrens erarbeitet. Darauf aufbauend kann nun für eine gegebene Applikation der hinsichtlich der Messunsicherheit und Nichtlinearität optimale Phase-Shift-Algorithmus gewählt werden.

Die Betrachtung der PMD-Technologie innerhalb einer konkreten Applikation verlangt ebenso eine gründliche Analyse aller verwendeten Systemkomponenten. Auch hierbei ist eine systemtheoretische Betrachtung allein nicht ausreichend, wie am Phänomen der Modeninstabilität einer Halbleiterlaserdiode deutlich wurde. Erst eine statistische Betrachtung der halbleiterelektronischen Zusammenhänge innerhalb dieses Elements machte die auftretenden Sprünge im Messwert einer eindimensionalen TOF-Entfernungsmessung erklärbar.

Aber noch ein weiteres Problem der PMD-Technologie zeichnete sich in dieser Applikation ab, nämlich das der Abhängigkeit der Impedanz der Modulation von der eingestrahlten Lichtintensität. Dieses in allen Halbleiterempfangselementen bekannte Problem kann jedoch aufgrund des zweikanaligen Aufbaus einer PMD-Struktur gemessen, beurteilt und korrigiert werden. Unsere Überlegungen zur Applikationsproblematik der eindimensionalen TOF-Entfernungsmessung erlaubten erstmalig den Einsatz der PMD-Technologie in einem industrietauglichen Sender, der nicht nur in Preis und Baugröße seinesgleichen sucht, sondern im Gegensatz zu herkömmlichen Ansätzen ohne eine kontinuierliche parallele Referenzierung auskommt. Dies kann im Wesentlichen auf die Möglichkeit der inhärenten quasioptischen Mischung innerhalb der optoelektronischen Signalwandlung zurückgeführt werden, da sich hierdurch die Signalkette gegenüber herkömmlichen Konzepten deutlich verkürzt und die damit einhergehenden Möglichkeiten zur Signaldrift deutlich eingeschränkt werden. Zukünftig ist durchaus noch mit einer weiteren Performance-Steigerung zu rechnen, wenn die hier in Ansätzen beschriebene IQ-Demodulation umgesetzt wird. Die Chance dieses Ansatzes liegt darin, dass hier der zur Phasenbestimmung notwendige Real- und Imaginärteil kontinuierlich zur Verfügung steht und nicht erst, wie im bisherigen Phase-Shift-Verfahren notwendig, durch mehrere sequentielle Einzelmessungen bestimmt werden muss.

Die in der eindimensionalen TOF-Entfernungsmessung gemachten Erfahrungen bildeten die Grundlage für den nächsten Schritt, den Aufbau einer für den industriellen Einsatz fähigen 3D-TOF-Kamera. Innerhalb dieser Arbeit konnte sehr deutlich der Gewinn einer solchen Kamera gegenüber herkömmlichen 3D-Ansätzen, wie Stereoskopie und Optischer Fluss, herausgearbeitet werden. Zusätzlich wurde eine einfache 3D-Bildverarbeitung zur Lagebestimmung eines Objekts realisiert, die aufgrund einer Transformation der gewonnenen Tiefeninformation in den

dreidimensionalen Fourier-Raum mit hoher Geschwindigkeit arbeiten kann, da darin nur noch auf eine einzelne Multiplikation zurückgegriffen werden muss. Hier wird der Anspruch an die Anwendung der mehrdimensionalen Systemtheorie sehr deutlich, was gleichzeitig das enorme Forschungs- und Entwicklungspotenzial in diesem Bereich aufzeigt.

Auf der gleichen Basis unserer Erfahrungen in der eindimensionalen TOF-Entfernungsmessung legt diese Dissertation einen Messaufbau zur zeitaufgelösten Vermessung von Fluoreszenzeffekten vor. Die Eignung eines PMD-Empfängers für diese Applikation konnte sehr deutlich nachgewiesen werden. Die Kenntnisse über die Modeninstabilität einer Halbleiterdiode und über die Intensitätsabhängigkeit der Impedanz der Modulationsgates einer PMD-Struktur waren hier gleichermaßen umsetzbar und erfolgreich. Damit steht dem Einsatz einer PMD-Matrix zur zweidimensionalen Vermessung der Fluoreszenzeffekte einer Probe nichts mehr im Wege.

Im Bereich der optischen Kommunikation wurde in dieser Arbeit erstmals die Eignung der PMD-Technologie zur optischen chirp-basierten Informationsübertragung beschrieben und nachgewiesen. Dies bildet die Grundlage für ein flexibles wie auch robustes Kommunikationssystem mit Mehrfachzugriff. Darüber hinaus konnten erste Optimierungsmöglichkeiten für ein solches System simulativ untersucht werden, womit sicher das Entwicklungspotenzial des beschriebenen Verfahrens noch nicht ausgereizt sein wird. Vielmehr bietet sich hier noch ein großes Feld für zukünftige Forschungs- und Entwicklungsaufgaben. Der Schlüssel für die Umsetzung eines solchen chirp-basierten Kommunikationsverfahrens, egal ob im optischen Bereich oder Funk, liegt im Management der zur Verfügung stehenden Gesamtbandbreite, was aufgrund der Variationsmöglichkeiten in den Parametern Amplitude, Frequenz und Phase im Subkanal eines jeden Benutzers eine sehr komplexe mehrdimensionale Aufgabe darstellt.

Insgesamt ist mit dieser Arbeit ein großer Schritt in Richtung Produktreife der PMD-Technologie gelungen, was sich nicht nur an dem ersten kommerziell erhältlichen PMD-basierten Sensor für den industriellen Bereich ablesen lässt, sondern ebenso an der Vielzahl möglicher weiterer Applikationen, die aufgrund der Grenzen einer solchen Arbeit hier nicht vollständig genannt werden können.

A1 Herleitung der Phasenberechnung mittels Phase-Shift-Interferometrie

a) Bestimmung der Phase φ

Zielfunktion:

$$f(t) = \frac{a_0}{2} + A \cdot \sin(\omega \cdot t + \varphi) \quad \text{mit:} \quad \omega \cdot t = \varphi_S \cdot n = \frac{2\pi}{N} \cdot n \quad \text{Gl. A1.1}$$

$$f(t) = \frac{a_0}{2} + A \cdot \sin(\varphi_S \cdot n + \varphi) \quad \text{Gl. A1.2}$$

Quadratische Fehlerfunktion:

$$S = E^2 = \sum_{n=1}^N \left[\left(Y_n - \frac{a_0}{2} - A \cdot \sin(\varphi_S \cdot n + \varphi) \right)^2 \right] \quad \text{Gl. A1.3}$$

Ableitung nach φ :

$$\frac{dS}{d\varphi} = 2 \sum_{n=1}^N \left[\left(Y_n - \frac{a_0}{2} - A \cdot \sin(\varphi_S \cdot n + \varphi) \right) \cdot (-A \cdot \cos(\varphi_S \cdot n + \varphi)) \right] = 0 \quad \text{Gl. A1.4}$$

$$\frac{dS}{d\varphi} = 2 \sum_{n=1}^N \left[\left(Y_n - \frac{a_0}{2} \right) \cdot (-A \cdot \cos(\varphi_S \cdot n + \varphi)) \right] + 2 \sum_{n=1}^N \left[A^2 \cdot \sin(\varphi_S \cdot n + \varphi) \cdot \cos(\varphi_S \cdot n + \varphi) \right] = 0 \quad \text{Gl. A1.5}$$

Mit dem Additionstheorem

$$\sin(x) \cdot \cos(y) = \frac{1}{2} (\sin(x-y) + \sin(x+y)) \quad \text{Gl. A1.6}$$

und der Bedingung $N > 1$ gilt für die zweite Summenfolge der **Gl. A1.5**:

$$2 \sum_{n=1}^N \left[A^2 \cdot \sin(\varphi_S \cdot n + \varphi) \cdot \cos(\varphi_S \cdot n + \varphi) \right] = 0 \quad \text{Gl. A1.7}$$

Gl. A1.5 vereinfacht sich somit zu:

$$\frac{dS}{d\varphi} = 2 \sum_{n=1}^N \left[\left(Y_n - \frac{a_0}{2} \right) \cdot (-A \cdot \cos(\varphi_S \cdot n + \varphi)) \right] = 0 \quad \text{Gl. A1.8}$$

Mit dem Additionstheorem $\cos(\alpha \pm \beta) = \cos(\alpha) \cdot \cos(\beta) \mp \sin(\alpha) \cdot \sin(\beta)$ gilt:

$$\frac{dS}{d\varphi} = \sum_{n=1}^N \left[\left(Y_n - \frac{a_0}{2} \right) \cdot A \cdot (-\cos(\varphi_S \cdot n) \cdot \cos(\varphi) + \sin(\varphi_S \cdot n) \cdot \sin(\varphi)) \right] = 0 \quad \text{Gl. A1.9}$$

bzw.:

$$\sum_{n=1}^N \left(Y_n - \frac{a_0}{2} \right) \cdot \cos(\varphi_S \cdot n) \cdot \cos(\varphi) = \sum_{n=1}^N \left(Y_n - \frac{a_0}{2} \right) \cdot \sin(\varphi_S \cdot n) \cdot \sin(\varphi) \quad \text{Gl. A1.10}$$

$$\frac{\sin \varphi}{\cos \varphi} = \frac{\sum_{n=1}^N \left(Y_n - \frac{a_0}{2} \right) \cdot \cos(\varphi_S \cdot n)}{\sum_{n=1}^N \left(Y_n - \frac{a_0}{2} \right) \cdot \sin(\varphi_S \cdot n)} \quad \text{Gl. A1.11}$$

Mit der Bedingung $N > 1$ gilt:

$$\frac{a_0}{2} \cdot \sum_{n=1}^N \sin(\varphi_S \cdot n) = \frac{a_0}{2} \cdot \sum_{n=1}^N \cos(\varphi_S \cdot n) = 0 \quad \text{Gl. A1.12}$$

Damit vereinfacht sich **Gl. A1.11** zu der allgemeinen Least-Squares-Approximation:

$$\tan \varphi = \frac{\sum_{n=1}^N Y_n \cdot \cos(\varphi_S \cdot n)}{\sum_{n=1}^N Y_n \cdot \sin(\varphi_S \cdot n)} \quad \text{Gl. A1.13}$$

b) Bestimmung der Amplitude A

Zielfunktion:

$$f(t) = \frac{a_0}{2} + A \cdot \sin(\omega \cdot t + \varphi) \quad \text{mit:} \quad \omega \cdot t = \varphi_S \cdot n = \frac{2\pi}{N} \cdot n \quad \text{Gl. A1.14}$$

$$f(t) = \frac{a_0}{2} + A \cdot \sin(\varphi_S \cdot n + \varphi) = \frac{a_0}{2} + a \cdot \sin(\varphi_S \cdot n) + b \cdot \cos(\varphi_S \cdot n) \quad \text{Gl. A1.15}$$

mit

$$A = \sqrt{a^2 + b^2} \quad \text{Gl. A1.16}$$

und

$$\tan \varphi = \frac{b}{a} \quad \text{Gl. A1.17}$$

Die orthogonale Beziehung (**Gl. A1.15**) erlaubt zwei getrennte Ansätze für die Einzelamplituden der Sinus- und der Cosinusfunktion. Die Lösungen lassen sich mittels **Gl. A1.16** wieder zu der Gesamtamplitude zusammenführen. Die Quadratische Fehlerfunktion für den Ansatz des Sinusanteils lautet:

$$S = E^2 = \sum_{n=1}^N \left[\left(Y_n - \frac{a_0}{2} - a \cdot \sin(\varphi_S \cdot n) \right)^2 \right] \quad \text{Gl. A1.18}$$

Die Ableitung der quadratischen Fehlerfunktion nach a lautet:

$$\frac{dS}{da} = 2 \sum_{n=1}^N \left[\left(Y_n - \frac{a_0}{2} - a \cdot \sin(\varphi_S \cdot n) \right) \cdot (-\sin(\varphi_S \cdot n)) \right] = 0 \quad \text{Gl. A1.19}$$

$$\frac{dS}{da} = -2 \sum_{n=1}^N [Y_n \cdot \sin(\varphi_S \cdot n)] + 2 \sum_{n=1}^N \left[\frac{a_0}{2} \cdot \sin(\varphi_S \cdot n) \right] + 2 \sum_{n=1}^N [a \cdot \sin^2(\varphi_S \cdot n)] = 0 \quad \text{Gl. A1.20}$$

Unter der Bedingung $N > 1$ gilt:

$$\sum_{n=1}^N \left[\frac{a_0}{2} \cdot \sin(\varphi_S \cdot n) \right] = 0 \quad \text{Gl. A1.21}$$

$$0 = -2 \sum_{n=1}^N [Y_n \cdot \sin(\varphi_S \cdot n)] + 2 \sum_{n=1}^N [a \cdot \sin^2(\varphi_S \cdot n)] \quad \text{Gl. A1.22}$$

Aufgrund der Orthogonalität gilt:

$$\sum_{n=1}^N [\sin(\varphi_S \cdot n) \cdot \sin(\varphi_S \cdot n)] = \frac{N}{2} \quad \text{Gl. A1.23}$$

Damit vereinfacht sich **Gl. A1.22**:

$$0 = -2 \sum_{n=1}^N [Y_n \cdot \sin(\varphi_S \cdot n)] + a \cdot N \quad \text{Gl. A1.24}$$

Die Amplitude des Sinusanteils berechnet sich aus:

$$a = 2 \cdot \frac{\sum_{n=1}^N [Y_n \cdot \sin(\varphi_S \cdot n)]}{N} \quad \text{Gl. A1.25}$$

Die Berechnung der Amplitude des Cosinusanteils kann analog zu der des Sinusanteils hergeleitet werden. Die Quadratische Fehlerfunktion des Cosinusanteils lautet:

$$S = E^2 = \sum_{n=1}^N \left[\left(Y_n - \frac{a_0}{2} - b \cdot \cos(\varphi_S \cdot n) \right)^2 \right] \quad \text{Gl. A1.26}$$

Die Amplitude des Cosinusanteils ergibt sich damit aus:

$$b = 2 \cdot \frac{\sum_{n=1}^N [Y_n \cdot \cos(\varphi_S \cdot n)]}{N} \quad \text{Gl. A1.27}$$

Die Gesamtamplitude A kann nun aus diesen beiden Teilergebnissen mit **Gl. A1.16** ermittelt werden. Die Amplitude des abgetasteten Signals berechnet sich nach der Methode der kleinsten Fehlerquadrate aus:

$$A = \frac{2}{N} \sqrt{\left[\sum_{n=1}^N [Y_n \cdot \sin(\varphi_S \cdot n)] \right]^2 + \left[\sum_{n=1}^N [Y_n \cdot \cos(\varphi_S \cdot n)] \right]^2} \quad \text{Gl. A1.28}$$

A2 Herleitung der Phasenberechnung mittels Phase-Shift-Interferometrie mit einer Kompensation eines zusätzlichen Phasenoffsets

Zielfunktion:

$$f(t) = \frac{a_0}{2} + A \cdot \sin(\omega \cdot t + \varphi_{\text{Syst.}} + \varphi) \quad \text{mit:} \quad \omega \cdot t = \varphi_S \cdot n = \frac{2\pi}{N} \cdot n \quad \text{Gl. A2.1}$$

Quadratische Fehlerfunktion:

$$S = E^2 = \sum_{n=1}^N \left[Y_n - \frac{a_0}{2} - A \cdot \sin(\varphi_S \cdot n + \varphi_{\text{Syst.}} + \varphi) \right]^2 \quad \text{Gl. A2.2}$$

Analog zur Herleitung in **Anhang A1** folgt:

$$\tan \varphi = \frac{\sum_{n=1}^N Y_n \cdot \cos(\varphi_S \cdot n + \varphi_{\text{Syst.}})}{\sum_{n=1}^N Y_n \cdot \sin(\varphi_S \cdot n + \varphi_{\text{Syst.}})} \quad \text{Gl. A2.3}$$

A3 Rauschunterdrückung mittels Mehrfachabtastung und Geradenapproximation

Zielfunktion:

$$f(t) = a + b \cdot t \quad \text{Gl. A3.1}$$

Quadratische Fehlerfunktion:

$$S = E^2 = \sum_{n=1}^N [Y_n - a + b \cdot t_n]^2 \quad \text{Gl. A3.2}$$

Die Ableitung der quadratischen Fehlerfunktion nach a lautet:

$$\frac{dS}{da} = -2 \cdot \sum_{n=1}^N [Y_n - a + b \cdot t_n] = 0 \quad \text{Gl. A3.3}$$

Die Ableitung der quadratischen Fehlerfunktion nach b lautet:

$$\frac{dS}{db} = -2 \cdot \sum_{n=1}^N [Y_n - a + b \cdot t_n] \cdot t_n = 0 \quad \text{Gl. A3.4}$$

Dies führt zu folgenden vereinfachten Ausdrücken:

$$n \cdot a + b \cdot \sum_{n=1}^N [t_n] = \sum_{n=1}^N Y_n \quad \text{Gl. A3.5}$$

und

$$a \cdot \sum_{n=1}^N t_n + b \cdot \sum_{n=1}^N t_n^2 = \sum_{n=1}^N [Y_n \cdot t_n] \quad \text{Gl. A3.6}$$

In Form einer Matrix ergibt sich folgendes Gleichungssystem:

$$\begin{bmatrix} n & \sum_{n=1}^N t_n \\ \sum_{n=1}^N t_n & \sum_{n=1}^N t_n^2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^N [Y_n] \\ \sum_{n=1}^N [Y_n \cdot t_n] \end{bmatrix} \quad \text{Gl. A3.7}$$

bzw.

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} n & \sum_{n=1}^N t_n \\ \sum_{n=1}^N t_n & \sum_{n=1}^N t_n^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{n=1}^N [Y_n] \\ \sum_{n=1}^N [Y_n \cdot t_n] \end{bmatrix} \quad \text{Gl. A3.8}$$

Die Lösung des Gleichungssystems ergibt:

$$a = \frac{\sum_{n=1}^N [Y_n] \cdot \sum_{n=1}^N t_n^2 - \sum_{n=1}^N t_n \cdot \sum_{n=1}^N [Y_n \cdot t_n]}{N \cdot \sum_{n=1}^N t_n^2 - \left(\sum_{n=1}^N t_n \right)^2} \quad \text{Gl. A3.9}$$

und

$$b = \frac{N \cdot \sum_{n=1}^N [Y_n \cdot t_n] - \sum_{n=1}^N t_n \cdot \sum_{n=1}^N [Y_n]}{N \cdot \sum_{n=1}^N t_n^2 - \left(\sum_{n=1}^N t_n \right)^2} \quad \text{Gl. A3.10}$$

A4 Herleitung der zweidimensionalen Übertragungsfunktion im Ortsfrequenzbereich einer 2x2-Matrix

Die Übertragungsfunktion im Ortsraum lautet:

$$\begin{aligned}
 h(x, y) = & \operatorname{rect}\left[\frac{x-x_0}{X}\right] \cdot \operatorname{rect}\left[\frac{y+y_0}{Y}\right] \\
 & + \operatorname{rect}\left[\frac{x+x_0}{X}\right] \cdot \operatorname{rect}\left[\frac{y+y_0}{Y}\right] \\
 & + \operatorname{rect}\left[\frac{x-x_0}{X}\right] \cdot \operatorname{rect}\left[\frac{y-y_0}{Y}\right] \\
 & + \operatorname{rect}\left[\frac{x+x_0}{X}\right] \cdot \operatorname{rect}\left[\frac{y-y_0}{Y}\right]
 \end{aligned}$$

Gl. A4.1

Die Transformation in den Ortsfrequenzbereich lautet dafür:

$$\begin{aligned}
 S(f_x, f_y) = & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{x}{X}\right] \cdot \operatorname{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{-j2\pi f_y y} \cdot e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} dx dy \\
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{x}{X}\right] \cdot \operatorname{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{-j2\pi f_y y} \cdot e^{+j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} dx dy \\
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{x}{X}\right] \cdot \operatorname{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{-j2\pi f_y y} \cdot e^{-j2\pi f_x x_0} \cdot e^{+j2\pi f_y y_0} dx dy \\
 & + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{x}{X}\right] \cdot \operatorname{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} \cdot e^{-j2\pi f_y y} \cdot e^{+j2\pi f_x x_0} \cdot e^{+j2\pi f_y y_0} dx dy
 \end{aligned}$$

Gl. A4.2

welche sich umstellen lässt zu:

$$\begin{aligned}
 S(f_x, f_y) = & e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \cdot \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} dx \cdot \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{x}{X}\right] \cdot e^{-j2\pi f_y y} dy \\
 & + e^{+j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \cdot \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} dx \cdot \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{x}{X}\right] \cdot e^{-j2\pi f_y y} dy \\
 & + e^{-j2\pi f_x x_0} \cdot e^{+j2\pi f_y y_0} \cdot \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} dx \cdot \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{x}{X}\right] \cdot e^{-j2\pi f_y y} dy \\
 & + e^{+j2\pi f_x x_0} \cdot e^{+j2\pi f_y y_0} \cdot \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{y}{Y}\right] \cdot e^{-j2\pi f_x x} dx \cdot \int_{-\infty}^{\infty} \operatorname{rect}\left[\frac{x}{X}\right] \cdot e^{-j2\pi f_y y} dy
 \end{aligned}$$

Gl. A4.3

bzw.

$$\begin{aligned}
S(f_x, f_y) &= \left(e^{-j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \right) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\
&\quad + \left(e^{+j2\pi f_x x_0} \cdot e^{-j2\pi f_y y_0} \right) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\
&\quad + \left(e^{-j2\pi f_x x_0} \cdot e^{+j2\pi f_y y_0} \right) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\
&\quad + \left(e^{+j2\pi f_x x_0} \cdot e^{+j2\pi f_y y_0} \right) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\
&= \left(e^{j2\pi(f_x x_0 + f_y y_0)} + e^{-j2\pi(f_x x_0 + f_y y_0)} \right) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\
&\quad + \left(e^{j2\pi(f_x x_0 - f_y y_0)} + e^{-j2\pi(f_x x_0 - f_y y_0)} \right) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y)
\end{aligned}$$

Gl. A4.4

Unter Verwendung trigonometrischer Funktionen ergibt sich:

$$\begin{aligned}
S(f_x, f_y) &= 2 \cos(2\pi(f_x x_0 + f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y) \\
&\quad + 2 \cos(2\pi(f_x x_0 - f_y y_0)) \cdot X \cdot \text{si}(\pi \cdot f_x \cdot X) \cdot Y \cdot \text{si}(\pi \cdot f_y \cdot Y)
\end{aligned}$$

Gl. A4.5

Verzeichnis der verwendeten Formelzeichen

a	Amplitude oder allg. Konstante
b	Geradenanstieg
$b_i(t)$	binärer Datenstrom
c	Lichtgeschwindigkeit
d	Dicke
e	Elementarladung
f	Frequenz
f_g	Grenzfrequenz
g	Verstärkung innerhalb d. Laser-Resonators
$g(t)$	Zeitsignal
$g(\lambda)$	längenwellenabhängiger optischer Gewinn
h	Planck'sches Wirkungsquantum
$h(t)$	Stoßantwort
$h(x,y)$	zweidimensionale Übertragungsfunktion
i	Effektivwert der Stromstärke
k	Wellenvektor oder Boltzmannkonstante
m_0	Ruhemasse
n	Brechungsindex
n_D	Anzahl verrauschter Dunkelstromelektronen
n_{KTC}	Anzahl verrauschter Resetelektronen
n_{ph}	Anzahl verrauschter Photostromelektronen
p	Impuls
q_e	Elementarladung
s	Wegstrecke o. Fehler
$s(t)$	Zeitsignal
t	Zeitvariable
t_i	Integrationszeit
$u(R)$	Mittlere Energie der Coloumb-Wechselwirkung
v	Geschwindigkeit
v_{ADC}	Rauschspannung der Digital-Analog-Wandlung
v_D	Rauschspannung des Dunkelstroms
v_{KTC}	Rauschspannung des Resets
v_{ph}	Rauschspannung des Photostroms
v_{th}	Rauschspannung des thermischen Stroms
v_s	Sättigungsgeschwindigkeit
x	Raumkoordinate
y	Raumkoordinate
y_k	Abtastwert
z	Raumkoordinate
A	Fläche
A_{12}	Fluoreszenzrate
B	Bandbreite
B_{12}	Absorptionsrate
C	elektrische Kapazität o. Kanalkapazität
C_i	Integrationskapazität
E	Energie o. elektrische Feldstärke

F	Rauschzahl
$F_{\beta}(t)$	Langewin-Rauschquelle
E_0	Ruheenergie
E_{ph}	Photonenenergie
G	Generationsrate oder Wichtung
G(f)	Signal Im Frequenzbereich
H(f)	Übertragungsfunktion im Frequenzbereich
I	elektrische Stromstärke, optische Intensität, Inversionsrate o.
Im	Informationsmenge
Im	Imaginärteil
J	Emissionsintensität o. elektrische Stromdichte
J_S	Schwellenstromdichte
K	Signalkontrast
K_S	Signalkontrast des Senders
K_E	Signalkontrast des Empfängers
M	Verstärkung
N	Rauschleistung o. Stützstellenanzahl
P	Leistung
P_i	Photonengeneration
P_{opt}	optische Leistung
Q	Gesamtladung
Q_{12}	Quenchrage
R	Empfindlichkeit o. elektrischer Widerstand
R_i	Emissionsrate
Re	Realteil
S	Signalleistung o. Fluoreszenzleistung
S(f)	Signal im Frequenzbereich
S(x,y)	Signal im Ortsfrequenzraum
T	Zeitintervall oder Temperatur
U	Spannung
V_0	Grenzspannung
W	Arbeit
Y	Messwert
α	Absorptionskoeffizient
β	Amplitude des oszillierenden elektrischen Feldes
γ	Verlust innerhalb d. Laser-Resonators
$\delta(t)$	Dirac-Stoß
ϵ	Dielektrizitätskonstante
η	Quantenausbeute
λ	Wellenlänge
ρ	Korrelation
σ	Standardabweichung
τ	Zeitverschiebung
φ	Phasenwinkel
φ_S	Phasen-Shift
φ_{SS}	Autokorrelation
ω	Kreisfrequenz

Δ	allg. Änderung
Λ	Pumprate bzw. Dreiecksfunktion
Σ	Summe
Φ	Austrittsarbeit