INAUGURAL — DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der
Ruprecht – Karls – Universität
Heidelberg

vorgelegt von
Dipl.-Ing. Stefan Trittler
aus Leonberg

Tag der mündlichen Prüfung:

Processing of

Interferometric Data

Gutachter: Prof. Dr. Fred A. Hamprecht

Abstract

In this thesis, fast and highly accurate interferometric metrology systems for both smooth and rough surfaces are presented. First, high-speed algorithms for white-light interferometry (WLI) and line scanning WLI are developed and their performance is compared. For large height differences, multiple wavelength interferometry is significantly faster, though, as in this approach the number of frames required for a surface estimate does not increase with surface height range.

A system based on a tunable diode laser is discussed in detail, and new sampling schemes and estimation algorithms for the device are derived. An approximation to the theoretically optimal sampling pattern is given and a corresponding fast estimation algorithm is presented. As a building block for that algorithm, accurate and fast phase and frequency estimation from a low number of samples is discussed, and a new approach based on an interpolated FFT is presented. The influence of laser speckle on rough surfaces is investigated. A robust, adaptive filtering algorithm is developed. It takes spatial relationships into account — without imposing strong smoothness constraints — and uses additional knowledge on the signal from the raw data to improve performance significantly, especially on rough surfaces.

Zusammenfassung

In dieser Dissertation werden interferometrische Messverfahren für glatte und raue Oberflächen vorgestellt. Zunächst werden Hochgeschwindigkeits-Auswerteverfahren für Weisslichtinterferometrie (WLI) und zeilenscannende WLI hergeleitet und miteinander verglichen. Bei größeren Höhenmessbereichen kann jedoch Mehrwellenlängeninterferometrie deutlich schneller sein, weil bei diesem Verfahren die Anzahl der erforderlichen Messpunkte unabhängig vom Höhenbereich ist.
Ein Messsystem mit einer durchstimmbaren Laserdiode wird im Detail analysiert, und neue Abtast- und Auswerteverfahren dafür werden hergeleitet. Es wird dabei eine Annäherung an die theoretische optimale Abtastung entwickelt und ein zugehöriges Auswerteverfahren vorgestellt. Dazu sind schnelle und hochgenaue Phasen- und Frequenzschätzverfahren auf Basis weniger Datenpunkte erforderlich. Ein neues Verfahren zur schnellen Frequenzschätzung mittels einer interpolierten FFT wird vorgestellt. Der Einfluss von Laser Speckle bei rauen Oberflächen wird untersucht und ein Verfahren zur robusten, adaptiven Filterung der Höhendaten wird gezeigt. Dieses verwendet räumliche Nachbarschaften — ohne dabei scharfe Anforderungen an die Glattheit der Oberfläche zu stellen — und zusätzliches Wissen über das Messsignal aus den Rohdaten, um die Ergebnisse insbesondere auf rauen Oberflächen deutlich zu verbessern.

Contents

THE KNACK OF FLYING IS LEARNING HOW TO THROW YOURSELF TO THE GROUND AND MISS.

(Douglas Adams, Hitchhiker's Guide to the Galaxy)

1. Introduction

1.1. Overview

Tolerances in industrial production get smaller and smaller. This in turn leads to a demand for more and more accurate metrology systems and more in-line process control. While in a laboratory environment measurement time is usually not that critical, in a production line a measurement system has to meet the line clock cycle, typically a few seconds only. For parts tolerances of $1\mu m$ a measurement repeatability better than $100nm$ is required. For larger fields of view this can only be achieved with interferometric systems. However, many precision parts have relatively rough surfaces or steps and therefore classic laser interferometry cannot be used. A description of available interferometric measurement systems and ways to optimize them for use in production environments, especially with respect to measurement time, constitutes the central part of this dissertation.

In the first chapter of this thesis, a brief review of available measurement principles is given, and the currently (April 2007) available commercial systems are summarized. Fundamental properties of rough surfaces and the speckle field caused by illumination with coherent light are introduced. Important basic concepts for data acquisition, data processing and parameter estimation are summarized.

In the second chapter, two different types of coherence based systems are discussed, a high-speed white-light interferometry system for surface measurements and a line-scanning white-light interferometry system. The focus of this work lies on the derivation of algorithms for hardware integration in order to deal with the tremendous amount of data that can be acquired with high speed cameras. Is is shown that good results can be obtained with algorithms that can easily be implemented on today's framegrabbers or intelligent cameras.

The third chapter discusses an alternative concept: multiple wavelength interferometry, in particular a system that uses a tunable laser to scan over a range of illumination frequencies. This system is analyzed in detail, both theoretically and in practice with measurements on both smooth and rough surfaces. In a first step, the hardware configuration and the resulting limitations on signal acquisition are discussed and theoretical limits on the accuracy are derived. An optimum sampling pattern for the system is presented. Signal processing in this case has to solve a frequency estimation problem. A fast algorithm for use with a nearly optimum sampling pattern is derived, and its performance is verified in simulations. As a building block for this algorithm, fast and

accurate frequency and phase estimation from a low number of samples is needed, and for that purpose a new, optimized interpolated FFT and a time-saving implementation of a linear least squares phase estimator are developed. These results can be applied to a wide variety of applications, some of which are presented in chapter four. The computational effort and the accuracy of the new algorithms are compared to alternative frequency estimation algorithms. Next two actual implementations of the algorithms for the given measurement system are presented, taking into account additional issues such as system calibration and offering insight into the trade-offs between processing time and accuracy. A new method for spatial filtering to deal with lower signal modulation, especially in the case of laser speckle, is shown. It uses additional information from the raw data and knowledge about the noise characteristics of the evaluation algorithm. Simulations and measurement results are presented that confirm the theoretical results above, and the influence of speckle in frequency scanning interferometry is discussed

In chapter four, further applications for the new algorithms that have been derived in the context of multiple wavelength interferometry are discussed. As an example, the use of phase and frequency estimation in the context of polarization imaging is presented. The results show that these algorithms can significantly reduce hardware requirements and increase the flexibility of polarization imaging systems.

In chapter five, the results are summarized, and the three systems from chapter two and three are compared with respect to their use in production environments.

1.2. Measurement Principles

There are several basic measurement principles for acquiring 3-D information [Schwartz et al., 1999]. The most important ones are discussed below with respect to their applicability for measuring rough surfaces quickly and accurately:

- Triangulation: This includes depth from focus, shape from shading, fringe projection, laser line, deflectometry and stereo camera systems.

- Interferometry: using coherent light. Single (classical laser interferometry), multiple (multiple wavelength interferometry) or a continuum of wavelengths (white-light interferometry) can be used. The height can be determined based on the relative phase, coherence or a combination of both [Häusler, 1999]. Holographic interferometry and electronic speckle pattern interferometry also belong to this category.

- Time of flight: measuring the time it takes for the light to travel to the object and back, based on the group velocity, typically incoherent.

Time of flight will not be discussed any further as the currently available systems are far too inaccurate (typical standard deviation is on the order of millimeters), and even though improvements are expected in the future, it is unlikely that sub-micron resolutions will be available any time soon, if at all.

1.2.1. Triangulation

The accuracy of triangulation depends on the triangulation angle θ. The physical limit on the z-accuracy for a laser line based triangulation system is given by the following equation:

$$\delta z = C\frac{\lambda}{2\pi \sin(u)\sin(\theta)} \tag{1.1}$$

with δz height resolution, λ wavelength, $\sin(u)$ aperture of the imaging system and θ triangulation angle. C is the speckle contrast, essentially the signal-to-noise ratio (SNR) for the pixel under investigation. In practice, a highly accurate triangulation system with a short working distance and a small field of view can reach a z-resolution of about 1 micron. For larger fields of view and larger working distances the resolution is lower, typically tens of microns. The same applies to fringe projection techniques.

Similar physical limits apply to depth from focus:

$$\delta z \propto \frac{\lambda}{\sin(u)^2} \tag{1.2}$$

There are a number of passive (i.e. not using a light source) triangulation based systems, including stereo cameras and photogrammetric approaches. At least for fast measurements these typically have a lower resolution as the SNR is worse.

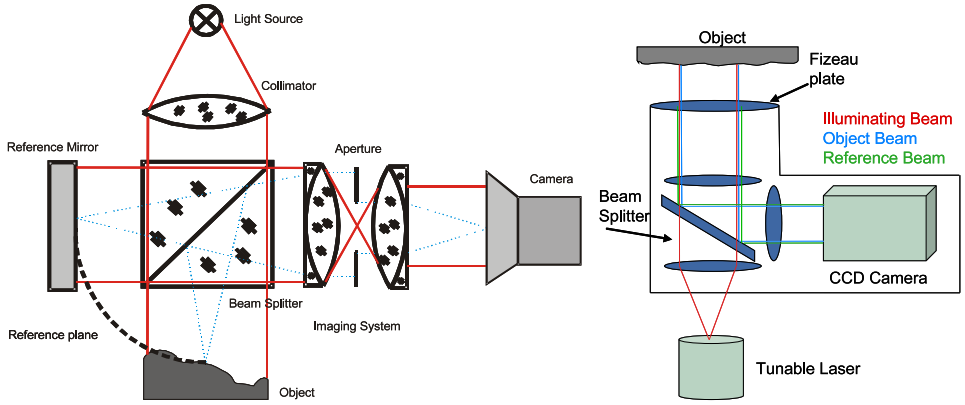
1.2.2. Interferometry

Light emitted by a laser is (at least approximately) monochromatic and polarized, and can be described as a plane electro-magnetic wave. In interferometry, this light is split into a reference and an object wave, and is later recombined and superimposed on a detector, e.g. a photo diode or a camera. Depending on the path difference there can be constructive or destructive interference, and based on this interference, one can in turn determine the path difference. As the signal is periodic and repeats with every height difference of $\lambda/2$, this method can only be used for smooth objects with very small height differences between neighbouring pixels, where a spatial unwrapping procedure (assuming that neighboring pixels are less than $\lambda/2$ apart) can be used to reconstruct the surface. There are multiple possible optical configurations, two of which are shown here (Figure 1.2.2 and Figure 1.2.2). Another possible setup is presented in section 2.2.

Light and every other electro-magnetic wave have to fulfill Maxwell's equations. This leads to the following equations (slightly simplified: In reality, the amplitude A might be complex and k is a vector perpendicular to A. This can be neglected as in typical interferometric configurations reference and object beam are parallel and both beams have the same polarization):

$$\begin{aligned} E_{ref}(t,x) &= A_{ref}e^{i(\omega t - kz_0)} \\ E_{obj}(t,x) &= A_{obj}e^{i(\omega t - kz_1)} \end{aligned} \tag{1.3}$$

with wave number $k = 2\pi f n/c$, where f is the laser frequency and n the index of refraction of the medium — in the following $n = 1$ will be assumed. The electric field



(a) Michelson Interferometer (b) Fizeau Interferometer

Figure 1.1.: Optical configurations for interferometric measurement systems

at the sensor is given by the superposition

$$E = E_{ref} + E_{obj}. \tag{1.4}$$

The sensor records the intensity only, the time average of the squared absolute value:

$$\begin{aligned} I(x) = \langle EE^* \rangle &= \left\langle \left(A_{ref} e^{i(\omega t - kz_0)} + A_{obj} e^{i(\omega t - kz_1)} \right) \right. \\ &\quad \left. \cdot \left(A_{ref} e^{-i(\omega t - kz_0)} + A_{obj} e^{-i(\omega t - kz_1)} \right) \right\rangle \\ &= A_{obj}^2 + A_{ref}^2 + A_{obj} A_{ref} \left(e^{ik(z_1 - z_0)} + e^{ik(z_1 - z_0)} \right) \\ &= A_{obj}^2 + A_{ref}^2 + 2A_{obj} A_{ref} \cos(k(z_1 - z_0)) \\ &= I_{obj} + I_{ref} + 2\sqrt{I_{obj} I_{ref}} \cos(k(z_1 - z_0)) \\ &= I_{obj} + I_{ref} + 2\sqrt{I_{obj} I_{ref}} \cos\left(\frac{2\pi f}{c} (z_1 - z_0) \right) \end{aligned} \tag{1.5}$$

As the path length difference $z_1 - z_0$ corresponds to twice the surface height difference Δh, one can rewrite the equation as follows with offset $C = I_{obj} + I_{ref}$ and modulation $A = 2\sqrt{I_{obj} I_{ref}}$:

$$I = A \cos\left(\frac{4\pi}{c} \cdot f \cdot \Delta h \right) + C \tag{1.6}$$

In this equation the relationship between offset and modulation is not directly visible any more, but this is usually not a problem as it is very hard to use that relationship in practice. It can only be used if there is no scattered or otherwise incoherent light present, and the camera has to be highly linear. Even then, the possible improvement is small [Wieler, 2006].

In phase shifting interferometry, the reference is moved in order to vary the phase difference. The signal phase is then determined from a low number of frames. There have been numerous discussions of optimum algorithms for processing of these signals

with respect to vibration and stability. The surface can then be reconstructed using spatial unwrapping.

The signal model above does not take into account different reflection properties of the surface. For reflection on a smooth reference mirror and object, the equation above is correct, but on a rough surface the phase of the reflected light might change. A more complete signal model is therefore:

$$I = A\cos\left(\frac{4\pi}{c}\cdot f\cdot\Delta h+\phi_0\right)+C \tag{1.7}$$

1.2.3. Coherence

For measurements of rough surfaces, classic interferometry cannot be used. Here the concept of coherence plays an important role. With its help the absolute distance of the object from the virtual reference plane can be determined. There are two main applications: Multiple wavelength interferometry and white-light interferometry. The interferometer types and the basic equations from the previous section also apply to coherence based methods; the advantages and disadvantages and the specific signal processing algorithms will be detailed in chapters 2.1, 2.2 and 3.

White-Light Interferometry

In white-light interferometry, a light source with a certain bandwidth is chosen, i.e. not a laser but a light emitting diode (LED) or a halogen lamp. There is direct relationship between the so called coherence length and the bandwidth of the light source:

$$l_c = c\cdot\frac{\lambda^2}{\Delta\lambda} \tag{1.8}$$

with bandwidth $\Delta\lambda$ and central wavelength λ. Definitions of the coherence length differ slightly in the literature. With the spectral full width for half maximum λ_H and l_c the length for which the modulation of the correlogram exceeds $1/e$, the constant is $c = 2\sqrt{\ln(2)/\pi}$ [Pavlíček, 1999].

The full signal model is more complex now than the equation for classical interferometry: The autocorrelation of the light enters into the modulation term. The coherence function, often called correlogram, is the autocorrelation of the light. If the spectrum of the light source is symmetric, the result can be split into an envelope function and a modulation with a carrier frequency. For the commonly used Gaussian spectrum, this leads to the following equation:

$$I(z) = I_{obj}+I_{ref}+2\sqrt{I_{obj}I_{ref}}\exp\left(\frac{4(z-z_0)^2}{l_c^2/2}\right)\cos\left(\frac{2\pi f}{c}(z-z_0)+\varphi_0\right) \tag{1.9}$$

The maximum of the envelope of the coherence function indicates the position of the virtual reference plane. In general, the accuracy increases with increasing bandwidth [Seiffert, 2007], but for rough surfaces there are limitations due to speckle: The coherence length must be larger than the surface height variation, otherwise the correlograms will be distorted. A commonly used threshold is

$$l_c \geq 2\pi\sqrt{2}\sigma_{obj} \approx 9\sigma_{obj}. \tag{1.10}$$

In case of a surface measurement, a scan in the z-direction is required. This usually requires a mechanical stage, and the accuracy of this stage plays an important role. There are three options for the scan: The object, the reference mirror or the whole measurement system can be moved. These are not quite equivalent from a theoretical point of view. If the measurement system or the object is moved, the speckle field changes during the scan; but it has the advantage that the system is always optimally focused at the true height. If the reference mirror is moved, the speckle field remains the same, but the scanning range is limited by the focal depth.

A high speed system for surface measurements is discussed in chapter 2.1.

It is possible to replace the z-scan with a spatial phase shift by reducing the sensor from 2-D to 1-D. Such a line sensor and its optical setup are described in chapter 2.2.

Multiple wavelength interferometry

Instead of using a continuous spectrum of light, one can also use multiple well-defined frequencies.

There are two possible system configurations:

- Multiple frequencies can be applied at the same time. They can either be separated optically, e.g. by using a grating, and therefore be detected by multiple detectors, or they can be detected together with a single sensor, when a time-dependent change in the detected intensity is introduced, e.g. by using an acousto-optic modulator in the reference beam.
- Alternatively, different frequencies can be applied consecutively, e.g. by using a tunable laser. This will be the focus of this thesis.

This increases the ambiguity interval from $\lambda/2$ in case of classical interferometry to $\Delta/2$, where Δ is the synthetic wavelength given by

$$\Delta = \left| \frac{1}{\frac{1}{\lambda_1} - \frac{1}{\lambda_2}} \right| \tag{1.11}$$

The same ambiguity also applies if more than two frequencies are used from a uniformly spaced grid of frequencies.

For a continuous change of the light source frequency f, a sinusoidal signal in f is obtained:

$$I(f) = I_{obj} + I_{ref} + 2\sqrt{I_{obj} I_{ref}} \cos\left(\frac{2\pi(z_1 - z_0)}{c} \cdot f + \varphi_0 \right) \tag{1.12}$$

1.3. State of the Art in Industrial Optical Metrology

As the focus of this thesis is on optical metrology systems for industrial applications, this brief overview is limited to a number of commercial system that are currently (April 2007) available to the author for measurements. Only systems with both high accuracy for height measurements while still offering large fields of view are mentioned here. This is by no means a complete list and only for illustrative purposes. Only systems that can be used for rough surfaces are considered.

- There are a large number of companies offering WLI systems, including Zygo, Veeco, 3D-Shape, Polytec, Mahr, ...

  To the knowledge of the author, the fastest commercially available system to date is a system from 3D-Shape (High-Speed Korad). This system, which was modified and presented internally at Bosch in early 2005, reaches more than $230\mu m/s$ at 512×512 resolution, without a significant increase in standard deviation (absolute numbers depend on surface roughness); up to $150\mu m/s$ are possible without subsampling. This was partly developed in the context of this thesis.

- Alicona offers a system called InfiniteFocus which is based on depth from focus. This system features nice color images even on rough and steep surfaces, but height resolution is only comparable to white-light interferometry or multiple wavelength interferometry for very small fields of view, on the order of $100\mu m$.

- Nanofocus offers a confocal microscope for surface measurements. Height resolution is good, but the largest available field of view is currently $1.6mm \times 1.6mm$.

- Siemens offers the SiScan system, another system based on confocal microscopy. This line sensor is essentially a point sensor replicated 64 times. It is very fast (up to 500,000 measurements per second), and its accuracy especially on "difficult" (i.e. high-contrast) surfaces is good. This system is a direct competitor to the line scanning white-light interferometer presented in chapter 2.2.

- A large number of fringe projection systems are available. They offer good performance, but are subject to the limitations discussed previously, i.e. the resolution for larger fields of view is lower than that of interferometric systems.

1.4. Laser Speckle

When measuring rough surfaces, laser speckle have to be considered.

Speckle are caused by the superposition of light from many scatterers on a rough surface, and lead to two main issues:

- The intensity of a speckle field follows a negative exponential distribution.

- The phase is uniformly distributed.

Properties of speckle are discussed in [Goodman, 1975]. There is no way around these effects, but there are several aspects to be taken into account in order to minimize their influence on measurement accuracy. Options include using a superposition of multiple independent speckle patterns to improve the intensity distribution as well as adjusting the pixel size of the camera such that it corresponds to the speckle size and no additional loss of contrast due to spatial integration occurs. In order to optimize measurement systems, the properties of the speckle field have to be known. For white-light interferometry, the issue has been discussed in detail by [Ettl, 2001] and [Pavlíček, 1999]. For multiple wavelength interferometry, [George & Jain, 1973] and [Salvadé,

1999] have analyzed the influence. This thesis adds some measurement results and discusses consequences for the algorithms in multiple wavelength interferometry.

1.5. Digital Signal Processing and Estimation Theory

Metrology today usually involves digitizing analog data and processing it on some kind of computer. The main objective of signal processing in metrology is reducing the dimensionality of the raw data to something that can be analyzed by a human or even automatically using classification methods, often leading to a binary decision: "good" or "bad" part. In case of interferometry, a series of frames taken by a camera has to be converted to a height map, which can later on be compared to e.g. CAD data. This is a parameter estimation problem. Both data acquisition and parameter estimation will be discussed next. Classification is outside the scope of this thesis.

1.5.1. Data Acquisition

The first step to digital signal processing is digitizing the data — which is done by analog-to-digital converters (ADCs). ADCs necessarily exhibit two types of noise, electronics noise and quantization noise [Seiffert, 2007]. In optical metrology for surface measurements, the sensor is usually a camera. Noise here includes electronics and readout noise (approximately Gaussian), and — as for all optical measurements — photon noise (which follows a Poisson distribution). Conversion to a digital signal introduces quantization noise (uniformly distributed).

The two main types of sensors currently used are CMOS and CCD imagers, with various subtypes. Several chapters on sensor concepts can be found in [Jähne et al., 1999] and will not be repeated here. Five main aspects which are relevant for interferometry will be highlighted here:

- Camera electronics and readout noise should be low. This noise is mainly thermal and can be reduced by cooling. This is most important for applications where the light intensity is low, which is generally not a problem in laser interferometry. It is very important to have a large dynamic range, and a low noise level obviously helps there, but more importantly, a large full well capacity is desirable. At the same dynamic range, a camera with higher noise and higher photon capacity is better because of lower relative photon noise.

- Camera speed is extremely important for white-light interferometry systems, but less important for other interferometric applications. CMOS sensors tend to be faster than CCD sensors overall, and a second speed advantage of some CMOS sensors is the ability to select arbitrary regions of interest, and thus reduce the data volume and increase speed if only a part of the field of view is needed. In case of CCD cameras, full camera lines have to be read.

- Camera spectral efficiency is also important for fast measurements when the available light intensity is limited. Both the active area (fill factor) and the quantum efficiency of the detection play a role. Quantum efficiency depends on the sensor and on the wavelength of the light. It can be very high at optimum

wavelength, more than 70% are common and some cameras reach more than 90%. Most CCD cameras use microlens arrays to increase the effective fill factor to close to 100%. CMOS cameras usually do not have microlenses and often only reach 30-40% fill factor.

- The camera spectral sensitivity must be a good match to the wavelength of the light source used (see quantum efficiency above). If the light source is in the near infrared, it is important that there are no filters used to limit the camera to the visible range (as done for consumer cameras to reach a natural looking image).

- Both CCD and CMOS cameras are available with different interfaces, including USB, Firewire, GigE (Gigabyte Ethernet) and Cameralink as well as several analog interfaces. For further processing on a PC, a digital interface is convenient. USB and Firewire offer reasonable speed at low cost, while Cameralink offers higher performance at a fairly high cost; GigE is a new standard which is just arriving in volume.

Based on these considerations, a fast CMOS camera with Cameralink interface was chosen for the white-light interferometry system discussed in chapter 2.1, a fast and highly linear CCD camera with Cameralink interface interface was chosen for the line scanning WLI system in chapter 2.2, and a CMOS camera with higher bit depth and USB interface was chosen for the frequency scanning system discussed in chapter 3.

1.5.2. Data Processing

Once the data is digitized, it has to be processed: For reasons of flexibility, most of the work in this thesis was performed on a standard PC. For high speed applications there are faster options though, especially if the problem can be parallelized.

Many framegrabbers and some cameras offer the opportunity to process data directly, using specialized signal processors of field programmable gate arrays (FPGAs). Most signal processors in intelligent cameras have been optimized for filtering individual images or for video compression and are of little use for the analysis of image sequences, as the memory access patterns and available memory bandwidth are insufficient for that task. FPGAs are freely configurable (though more difficult to program), and provided they have sufficient internal memory or sufficiently fast external memory banks, almost all current FPGAs are fast enough to (pre-)process camera data in real-time. In the case of both white-light and multiple wavelength interferometry, there is an individual time series for every pixel which has to be analyzed, and in an FPGA a large number of pixels can be processed in parallel (limited only by the number of logic elements).

The basic structure of an FPGA is shown in Figure 1.2. Several concepts for using an FPGA in white-light interferometry are discussed in detail in chapter 2.1.

A new alternative to using dedicated hardware is the use of PC graphics cards, which are essentially highly parallel signal processors, but with typically lower accuracy and some limitations on data structures and access patterns. Most of the limitations can be overcome with the newest generation of graphics cards with several

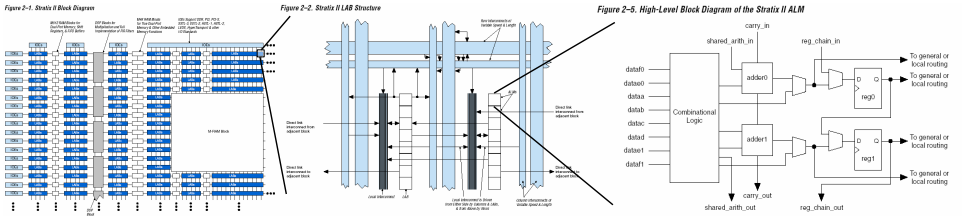


Figure 1.2.: Structure of a Xilinx Stratix II FPGA (taken from the manual). Left: Overall structure. Middle: Intermediate building blocks, so called LABs. Right: Basic building blocks, so called ALMs. An FPGA can have more than 100000 ALMs, several hundred dedicated multipliers for filter implementations and several megabytes of embedded memory.

hundred freely programmable shaders. Some compilers and mathematical libraries for these cards are already available.

1.5.3. Estimation Theory

Once the data has been acquired (and maybe filtered or pre-processed), algorithms are needed to extract an estimate of the desired parameters from the data. The following introduction to estimation theory gives a short summary of some important methods that will be used later on. A more detailed introduction can be found in many textbooks, including [Poor, 1994] and [Moon & Stirling, 2000]. The description and notation in this chapter closely follows [Poor, 1994]. In all cases considered in the following section, it is assumed that the data is discrete (i.e. sampled) and that the parameter to be estimated is continuous.

There are two general approaches to parameter estimation:

- Bayesian parameter estimation: The parameter is treated as a random variable statistically related to the observation.
- Non-random parameter estimation: The parameter is unknown, but no statistical properties are assumed explicitly.

Bayesian estimation determines the parameter estimate that minimizes the posterior cost given the distribution of the parameter, the distribution of the observations, a cost function defining the cost of estimation errors and an observation. This is described in detail in [Poor, 1994] on page 142.

There are multiple possible criteria and requirements for an estimator. In case of Bayesian estimation, commonly used criteria include:

- Minimum-Mean-Squared-Error (MMSE), also called conditional mean estimate.

$$C(a, \theta) = (a - \theta)^2 \tag{1.13}$$

- Minimum-Mean-Absolute-Error (MMAE), also called conditional median estimate.

$$C(a, \theta) = |a - \theta| \tag{1.14}$$

- Maximum A Posteriori Probability (MAP), also called conditional mode estimate (obtained for $\Delta \to 0$). This is not really a Bayesian estimate, but it fits into this framework.

$$C(a,\theta) = \begin{cases} 0 & \text{if } |a-\theta| \leq \Delta \\ 1 & \text{if } |a-\theta| > \Delta \end{cases} \tag{1.15}$$

These differ only in the cost function, the names and definitions are self-explanatory. A Bayesian approach is preferable if prior knowledge on the parameter distribution is available. It offers an intuitive method to introduce knowledge on surface properties into processing of the raw data by specifying a prior on the spatial distribution [Hissmann, 2005].

In nonrandom parameter estimation a prior distribution on θ is not needed. Without a prior, averaging can only be performed with respect to the conditional mean-squared error. There is no solution that minimizes the variance for all parameter values, and this would not make sense anyway — for any parameter value θ_0 one could choose the estimator $\hat{\theta} = \theta_0$ to get zero variance, but for all other parameter values such an estimator would obviously not be good. The restriction to unbiased estimators avoids this problem: The desired estimator is usually a minimum-variance unbiased estimator (MVUE). It has to be noted that it is often hard to find such an estimator, it does not always exist, and it is not always the best choice either, as for some examples a much lower variance can be obtained with a biased estimator.

Nevertheless, there are some important and useful results associated with this approach. The so-called information inequality gives a lower bound on the variance of an estimator:

$$\mathrm{Var}_\theta \geq \frac{\left[\frac{\partial}{\partial\theta} E_\theta(\hat{\theta}(Y))\right]^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log p_\theta(Y)\right)^2\right]} \tag{1.16}$$

The denominator is also called the Fisher information I_θ. For unbiased estimators this simplifies to

$$\mathrm{Var}_\theta \geq \frac{1}{I_\theta} = \frac{1}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log p_\theta(Y)\right)^2\right]} \tag{1.17}$$

This result is known as the Cramér-Rao Bound (CRB).

The Cramér-Rao bound as a lower bound on the variance of an unbiased estimator can often be determined even though it is very difficult to find a good estimator and might be very hard or impossible to find any of the estimators discussed above in a given estimation problem. This is a highly useful result if an available estimator is close to this bound, because it shows that the estimation cannot be improved much further, and therefore the current algorithm is a reasonable choice. The opposite does not hold true: The CRB is not necessarily a tight bound, and an algorithm might be good even though it is quite far from the CRB. The signal model might imply a threshold effect for larger noise, and therefore a bound based on looking at local

curvature might not be the best approach. There are a number of other bounds, for example [Bell et al., 1997]. For the frequency estimation problem in the presence of low noise, the CRB is a reasonably tight bound though, and the assumption of approximate unbiasedness of the estimator is valid.

Not every estimator can be found easily, and especially for a finite number of samples the problem can get highly complex for the Bayesian and nonrandom approaches discussed so far. A frequently used alternative with good asymptotic properties is Maximum Likelihood estimation. It can be seen as MAP estimation with a uniform prior, or as finding the value of θ that makes the observations most likely. An estimator that reaches the CRB is a maximum likelihood estimator; the inverse is not true, but for the number of observations $n \to \infty$ and independent and identically distributed noise, it reaches the CRB asymptotically.

For Gaussian noise with zero mean, maximum likelihood estimation is equivalent to classical least squares estimation:

$$\hat{\theta} = \arg\min_{\theta} \sum_{k=1}^{N} [Y_k - s_k(\theta)]^2 \tag{1.18}$$

The discussion above can be extended to the estimation of vector parameters as they occur in optical metrology. The notation in the following follows [Wieler, 2006]. We investigate performance bounds for an estimator $T(X)$ (where X denotes the data vector) of a (vector) parameter θ, given a probability density function $f(X, \theta)$.

The information inequality can be rewritten in vector form using this notation and with the estimator bias $b(T, \theta)$:

$$\mathrm{Var}(T, \theta) \geq \frac{1}{I(\theta)} \cdot \left(1 + \frac{\partial b(T, \theta)}{\partial \theta}\right)^2 \tag{1.19}$$

$$I(\theta) = E\left[\left(\frac{d}{d\theta} \log f(X, \theta)\right)^2\right] \tag{1.20}$$

In case of an unbiased parameter, this simplifies to

$$\mathrm{Var}(T, \theta) \geq \frac{1}{I(\theta)} \tag{1.21}$$

If we look at the signal model, we can write the actual detected data vector x as a sum of the "true" signal (characterized by the sampling points t and the true parameters θ) and noise:

$$x = y(t, \theta) + n \tag{1.22}$$

As given above, an element of the Fisher information matrix can be written as

$$I_{j,k}(y; t, \theta) = E\left[\left(\frac{\partial}{\partial \theta_j}(\log f(x; \theta)) \cdot (\frac{\partial}{\partial \theta_k}(\log f(x; \theta)))\right)\right] \tag{1.23}$$

In case of multivariate Gaussian noise with covariance matrix Σ, the probability density function is given by

$$f(x,\theta) = \frac{1}{\sqrt{(2\pi)^N \det(\Sigma)}} \cdot e^{-\frac{1}{2}[x-y(t,\theta)]^T \Sigma^{-1} [x-y(t;\theta)]} \tag{1.24}$$

Then the elements of the Fisher information matrix can be computed:

$$I_{j,k}(y;t,\theta) = E\left[\left(\frac{\partial}{\partial\theta_j}(\log f(x;\theta))\right) \cdot \left(\frac{\partial}{\partial\theta_k}(\log f(x;\theta))\right)\right] \tag{1.25}$$

$$\begin{aligned}
\frac{\partial}{\partial\theta_j}(\log f(x;\theta)) &= \frac{\partial}{\partial\theta_j}\left(-\frac{1}{2}\left[x-y(t,\theta)\right]^T \Sigma^{-1}\left[x-y(t;\theta)\right]\right) \\
&= -\frac{1}{2}\left(-\frac{\partial y}{\partial\theta_j}\right)^T \Sigma^{-1}\left[x-y(t;\theta)\right] - \frac{1}{2}\left[x-y(t;\theta)\right]^T \Sigma^{-1}\left(-\frac{\partial y}{\partial\theta_j}\right) \\
&= \left(\frac{\partial y}{\partial\theta_j}\right)^T \Sigma^{-1}\left[x-y(t;\theta)\right]
\end{aligned} \tag{1.26}$$

$$\begin{aligned}
I_{j,k}(y;t,\theta) &= E\left[\left(\frac{\partial y}{\partial\theta_j}\right)^T \Sigma^{-1}\left[x-y(t;\theta)\right]\left[x-y(t;\theta)\right]^T \Sigma^{-1}\left(\frac{\partial y}{\partial\theta_j}\right)\right] \\
&= \left(\frac{\partial y}{\partial\theta_j}\right)^T \Sigma^{-1} E\left[\left[x-y(t;\theta)\right]\left[x-y(t;\theta)\right]^T\right] \Sigma^{-1}\left(\frac{\partial y}{\partial\theta_j}\right) \\
&= \left(\frac{\partial y}{\partial\theta_j}\right)^T \Sigma^{-1} E\left[nn^T\right] \Sigma^{-1}\left(\frac{\partial y}{\partial\theta_j}\right) \\
&= \left(\frac{\partial y}{\partial\theta_j}\right)^T \Sigma^{-1}\Sigma\Sigma^{-1}\left(\frac{\partial y}{\partial\theta_j}\right) \\
&= \left(\frac{\partial y}{\partial\theta_j}\right)^T \Sigma^{-1}\left(\frac{\partial y}{\partial\theta_j}\right)
\end{aligned} \tag{1.27}$$

Up to now, this is a general approach with no constraints on the signal model or the number of parameters. The only limitation is the assumption that the noise is multivariate Gaussian.

In order to determine the CRB, the Fisher information matrix has to be inverted. If there is only a single parameter or if the parameters are independent, one can simply invert the diagonal elements directly. Usually, when there are multiple parameters a matrix inversion will be required. The computational effort can be reduced for nuisance parameters: As we are not interested in the full CRB matrix, but only in the entries corresponding to the θ_i we would like to estimate, a full matrix inversion can

be avoided. This is shown in [Wieler2006a] but will not be discussed in detail in this thesis.

Getting back to optical metrology, two estimation problems can be stated:

- In the first case, the available data is evaluated for each pixel individually. This type of problem lends itself to parallel processing. Usually an underlying signal model is available, but no informative prior on the distribution. This leads to nonrandom parameter estimation, and for tractable results maximum likelihood or least squares estimation are usually used.

- In case of surface measurements, reasonable assumptions for the prior are typically spatial relations, which can be used in Bayesian estimation approaches [Restle, 2003; Hissmann, 2005]. A signal model is required as discussed above. The additional prior depends on the measurement task, ranging from the assumption of a perfectly smooth surface to a weak preference for smoothness.

In practice both approaches are often combined by using the result from the first approach as an initial value for the second one, for example by filtering a height map to suppress outliers [Restle, 2003]. It is sometimes possible to directly use the second approach, thereby reaching a better result [Hissmann, 2005].

The best feasible approach for many applications is a maximum likelihood estimator. If not even that is possible, least squares estimation can be used. The Cramér-Rao bound is not a bound on the mean squared error but a bound on the minimum variance of an unbiased estimator. Unbiasedness is not an issue for the applications discussed in this thesis — most frequency estimators are quite good, and show very little bias if the SNR is high. All estimators in the following are designed to be (almost) unbiased. While they are not perfectly unbiased in practice, their performance is good enough that the bias term in the information inequality above is small and therefore the CRB alone yields useful results. The CRB can then, for example, be used to rate the performance of estimators obtained by heuristic methods. For stronger noise this is not true; estimators show threshold effects and performance decreases rapidly. These cases are not discussed here, as these noise levels are irrelevant for the metrology applications discussed in this thesis.

2. White-Light Interferometry

The first approach to interferometric measurements on rough surfaces discussed in this thesis is white-light interferometry. This is a relatively mature technology, and there are several commercial products available. Most of these are designed for laboratory use and not for in-line inspection, though. The key differences between these different fields of application are robustness and measurement speed. Ways to increase measurement speed are discussed in this chapter.

2.1. High-Speed White-Light Interferometry

2.1.1. Setup

Optical Setup

As mentioned in the introduction, there are multiple possible configurations for an interferometric measurement system. For the high-speed white-light interferometry system discussed next, a Michelson setup was chosen where the whole measurement system is moved relative to the measured object. This way, the area of the correlogram is always in focus, and a large working distance is possible. Additionally, the intensity of object and reference beam can be adjusted by simply inserting different filters into the reference beam. For measurements with a Piezo-driven stage, the reference mirror only instead of the whole measurement system was moved.

Light Source

Three different light sources have been investigated. The choice of light source has a direct influence on the signal processing, as the coherence width and the general shape of the correlogram correspond to the autocorrelation of the light source:

- a LED with approximately Gaussian spectrum,
- a fiber coupled high pressure Na-lamp and
- a halogen lamp.

For most applications a high intensity while keeping the spatial extent of the light source sufficiently small (to keep spatial coherence) is desirable. Signal processing is easiest if the envelope and thus the spectrum is a “nice” and smooth function, e.g. a Gaussian. The optimum coherence length depends on the surface roughness, in general a shorter coherence length yields more accurate results but it also makes subsampling difficult, causing longer measurement times.

The LED offers the highest luminance in this comparison, mainly due to its small size compared to the other sources. Its spectrum is smooth and its coherence length is quite long at more than ten microns.

The Na-lamp has an interesting spectrum with several strong lines, resulting in a very long coherence length and a spectrum that looks like a modulated sinusoid with a fairly complex envelope. This lamp has the highest overall intensity, but due to its lower luminance, the remaining intensity after fiber coupling and using a pinhole for spatial coherence was lower than that of the LED. This kind of spectrum offers some additional possibilities: Due to the continuously present sinusoidal signal, the position of the stage can be monitored from the camera signal (as in a laser interferometer). Finding the maximum of the envelope and thus the object height is more difficult, and might be the object of future research as in theory a high accuracy should be possible.

The halogen lamp also had a reasonably looking spectrum, but with a much shorter coherence length (on the order of two microns). While this is a good choice for highly accurate measurements, combined with the lowest intensity of the sources investigated here it slows down measurements as the signal has to be closely sampled and the exposure time has to be relatively long.

The LED was chosen for all further measurements as luminance is very important for a high-speed system, and because measurements with the high pressure Na-lamp spectrum turned out to be difficult to analyze. A more detailed analysis of light sources for interferometry can be found in [Höfer, 1994].

Motion Stage

Another important issue is the use of high precision stages for the mechanical scanning procedure. Any error in the stage position directly causes an error in the resulting height map. There are two main concepts: For small height ranges, a Piezo-driven stage can be used to move the reference mirror only. For larger height ranges, a linear stage (direct linear motor or spindle driven) can be used to move the whole system or the measurement object. Four stages have been analyzed in detail for this application:

- Newport XML-350 (linear motor, high-precision glass scale encoder, up to 300mm/s, 1nm resolution, bi-directional repeatability 50nm)
- PI M-511DG.K029 (spindle-driven, rotary encoder, gear, nominal resolution 6nm, up to 6mm/s, bi-directional repeatability 2 microns)
- PI M-511DD stage (“ActiveDrive”, glass scale, 100nm resolution, up to 100mm/s, bi-directional repeatability 0.1 microns)
- PI P-625.1CD, Piezo-driven stage with capacitive sensor

The characteristics above as given by the manufacturers are of little use for the desired application though: For high-speed white-light interferometry, it is not possible to move the stage to a position, acquire a frame, move it to the next position and so on. Instead, the stage has to be moved continuously while the images are being acquired. This is a totally different requirement on the behavior of the stage and its controller. The absolute difference between stage nominal and actual position is not a problem as

long as the actual position can be read whenever a frame is acquired. With the current software architecture, this is easily possible with the controller C-843 from PI, but more difficult with the XMS controller from Newport.

A number of measurements at different velocities were performed using a SIOS laser interferometer. The results and a more detailed discussion of the stages are given in appendix A.

Of the large stages, the Newport XML-350 stage offered the highest accuracy (comparable to the Piezo), but as it is also the most expensive stage (mainly due to the relatively expensive controller) and turned out to be more difficult to integrate into existing software, it was not used in the following.

The PI M-511DD stage turned out to be unsuitable for application in white-light interferometry. First of all, at 100nm the reported resolution of the encoder is not high enough to use it in the evaluation algorithms, and secondly, there was an issue with vibration (on the order of more than a micron). This may have been a calibration problem with that specific stage.

The PI M-511DG.K029 stage works reasonably well, but there is significant sampling jitter that gets worse with increased velocity. As the rotary encoder is at the motor, it cannot "see" errors caused by the transmission gear. These errors are clearly visible as frequency components in the analyzed data. Bidirectional repeatability is very poor, but for a white-light interferometry system unidirectional repeatability is more important, as returning to the same initial position before performing the next measurement is possible. These results were pretty good, and noise was much lower than for the table with the glass scale. For large distances on the order of centimeters this is not true any more, but for a white-light interferometry system a scanning range of millimeters is sufficient for most measurements. Especially results at 150 microns/s velocity were surprisingly good.

As expected, the piezo-driven stage P-625.1CD offers the best performance, but measurements showed that the velocity did not keep constant during the movement at first. This has since been fixed by PI. The capacitive sensor yields results that match the results of the laser interferometer very well (less than 15nm standard deviation).

The sampling jitter observed has a significant influence on the choice of the most appropriate algorithm. One option to reduce this error is using a high-speed-camera that can be triggered by an accurate position source, i.e. an integrated laser interferometer.

Camera

For a high speed system, the camera and its interface are important. As already mentioned in chapter 1.5.1 for that purpose a fast CMOS camera with Cameralink interface was chosen. The camera model Photonfocus MV-D1024-160-CL features a 160 MHz pixel clock, dual tap Cameralink connection (2×80 MHz) and a frame rate of 150fps at 1024×1024 with 8 bit resolution. As long as the available light intensity is sufficient, the frame rate can be increased almost linearly with a reduction in resolution, so that for most measurements a resolution of 512×512 at 480fps was used. An arbitrary field of view can be chosen without changes to the setup or the algorithms. An even lower resolution offers little benefit as the exposure time cannot be reduced any further for

rough metal surfaces with the LED light source described above, and therefore higher frame rates are not possible. For even higher frame rates, a stronger light source or a camera with higher quantum efficiency would be required (see Figure 2.1). Camera readout noise is about one gray level, but there is fairly strong fixed pattern noise. This is not a problem for this application as all pixels are analyzed individually and spatial relationships are not needed. The camera features an external trigger input and can thus be synchronized to the movement of a mechanical stage. Additionally, the camera characteristic can be modified between linear and logarithmic sensitivity, which is mainly useful for high contrast surfaces. Signal processing with non-linear camera characteristics is difficult, therefore whenever possible the linear setting should be used.

Synchronizing the camera to an external trigger input was implemented and tested, but while it did improve the correlogram shape, the cost in reduced scan velocity (as the speed has to be lowered such that the trigger interval is large enough even if two triggers closely follow each other due to jitter caused by the stage) was too high (about a factor of two in actual measurements). The implementation used an external "trigger box" that counted the increments of the step motor from the stage M511DG.K029 and used it to trigger the framegrabber (and therefore the camera) every N pulses. Using a free running camera, a significantly larger number of samples (about a factor of two) can be acquired in the same measurement time (i.e. closer spaced sampling or faster measurements), which more than compensates for the sampling jitter if the stage M511DG.K029 is used. Results might be different for other stages.

The data is transferred to a PC using a SiliconSoftware MicroEnable III CameraLink framegrabber. The fastest currently available PC chipsets at that time (2005) did not offer PCI-X interfaces (except for a few expensive and less compatible server mainboards), therefore a normal PCI slot was used. This limits throughput and thus camera pixel clock to approximately 118 MB/s (theoretical limit 133MB/s, but there is some bus overhead). Today, almost all mainboards feature PCIe interfaces and framegrabbers for PCIe are readily available, so this is not a problem any more. The PC itself was using a Pentium 4 processor running at 3.6GHz and had 2 GB of RAM.

2.1.2. Signal Processing

This chapter briefly summarizes a number of possible algorithms for finding the maximum of the envelope of the correlogram. The theoretical accuracy of the estimation has been discussed by [Seiffert, 2007], and some algorithms are described in detail in his thesis. In the following several additional algorithms are considered, and all algorithms are analyzed with respect to hardware acceleration.

The optimum algorithm strongly depends on the noise characteristics and on the signal properties; these in turn depend on the illumination, the stage, and the camera. Part of this analysis was performed together with Sébastien Wagner and is described in more detail in [Wagner, 2005].

For faster image acquisition, subsampling can be applied. This reduces the number of available samples on the correlogram and therefore increases relative noise. This is not discussed here as all algorithms remain applicable, they just "see" a different signal frequency and become more sensitive to noise and sampling jitter.

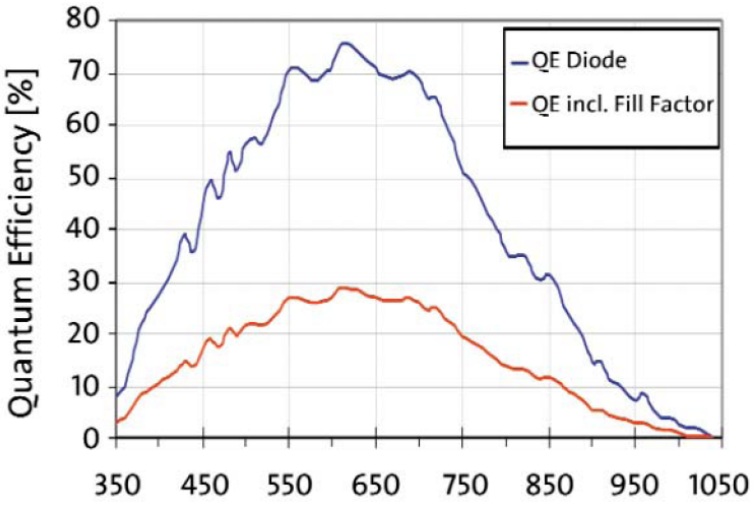


Figure 2.1.: Quantum efficiency of Photonfocus MV-D1024-160 CL vs. wavelength [Photonfocus, 2005]

The algorithms described below are usually not used individually, but in a combination of multiple steps:

1. First, the relevant part (i.e. the range of the correlogram) of the signal for every pixel has to be determined quickly, in order to reduce the amount of data. This is called pre-processing in the following.

2. Next, the envelope of the correlogram has to be reconstructed.

3. Then the maximum has to be found, and an interpolation can be performed to reach sub-pixel accuracy.

The first two steps can be combined if there is a fast way to get an accurate estimate of the envelope; the last step is usually separate as it requires knowledge on the position of the maximum and a local interpolation.

Pre-Processing

Several methods exist to find the range of the correlogram quickly. They are simple, and implementation is usually fast. Four of them are mentioned here:

- Minimum or maximum search in the raw data: Requires just one comparison per incoming pixel, and only three images are needed in memory at a time (current frame, maximum values so far, position of the maximum). Minimum search performs better than maximum search for two reasons: Additional incoherent light that might be temporarily present will always lead to an increase in intensity, only coherent light might decrease intensity. Second, photon noise is stronger for higher intensities.

- Contrast search: Finds the maximum difference between consecutive frames. Performance depends on the sampling interval, it is usually more robust than minimum or maximum search as incoherent light cancels out. Requires four images in memory and at least two comparisons.

- As the algorithms above are sensitive to noise, smoothing is helpful. A very good option is a sliding average of the contrast values, which needs two additional additions per frame. A much larger number of frames has to be kept in memory though (in a typical application approximately 30 frames are used, but that depends on sampling distance and coherence length), which imposes a significantly higher load on the memory subsystem.

Envelope Detection

There is a gradual transition from pre-processing to algorithms that reconstruct the true signal envelope: A sliding average reconstructs the envelope, and even the contrast method offers a coarse estimate. It is possible to do that much more accurately though - either directly in the time domain, possibly without additional pre-processing, or in the frequency domain. For longer measurements in a high speed application it is obviously not possible to perform an FFT on the whole set of data, but an FFT can be applied once the interesting range of the signal has been found by pre-processing.

Matched Filter and Correlation-Based Techniques

Correlation based techniques can offer higher accuracy than pure envelope based techniques, as they use the signal phase, too. A matched filter is the best choice here and is discussed in detail in [Seiffert, 2007]. The main disadvantage is its high computational complexity. However, if instead of a true matched filter the correlation is computed with a sine and cosine only (a good approximation for light sources with long coherence length), the effort can be reduced significantly. The implementation is then similar to the sliding average, just that now there are two running sums, one obtained by multiplying the incoming signal with a sinusoid and the other obtained by multiplying it with the same sinusoid shifted by $90°$. The envelope is then given by the squared sum of the values of the two running sums for every position. This requires just one additional value in memory and two additional multiplications (or N more values in memory and one additional multiplication) compared to the sliding average and will be described in more detail below and in Figure 2.6. Seiffert uses three sinusoids, but there is no reason for doing so as his formula can be simplified to the one given here (which is also used for example for demodulation in communication and radio systems). The performance of this algorithm is excellent if position jitter is low. For higher position jitter, the correlation length has to be shortened (at the cost of poorer noise suppression), but in this case N-bucket algorithms tend to be better (see below). If computational complexity is not an issue, [Seiffert, 2007] has shown methods to reconstruct the actual sampling positions and take them into account.

N-Bucket Algorithms

The so-called N-bucket algorithms are well known from phase shifting interferometry. Their properties have been studied extensively, and their primary purpose is the estimation of the phase of a signal sampled at known phase shifts. These algorithms can also be used in white-light interferometry, and they are very attractive due to their low memory requirements: A fairly accurate estimation of the envelope is possible using a small number of samples at a time (resulting in much lower memory requirements compared to FFT-based algorithms). This is an important aspect for hardware implementation.

This class of algorithms uses certain well-defined angles for the phase shifter in order to obtain simple closed form expressions for the phase, for example using $60°$ or $90°$ increments. This limits the possible measurement velocities. The simple and frequently used 3-frame algorithm is briefly derived next: it requires three images sampled every $90°$.

Let $I_1 = C + D\cos(\varphi - \alpha)$, $I_2 = C + D\cos(\varphi)$, $I_3 = C + D\cos(\varphi + \alpha)$. Then we obtain

$$\frac{I_3 - I_1}{I_1 - 2I_2 + I_3} = \frac{\cos(\varphi + \alpha) - \cos(\varphi - \alpha)}{\cos(\varphi - \alpha) - 2\cos(\varphi) + \cos(\varphi + \alpha)}. \tag{2.1}$$

For $\alpha = 90°$, this yields

$$\frac{-\sin(\varphi) - \sin(\varphi)}{\sin(\varphi) - 2\cos(\varphi) - \sin(\varphi)} = \frac{\sin(\varphi)}{\cos(\varphi)} = \tan(\varphi). \tag{2.2}$$

Therefore, in this case, the phase can be determined by

$$\varphi = \arctan\left(\frac{I_3 - I_1}{I_1 - 2I_2 + I_3}\right). \tag{2.3}$$

Similar tricks can be applied for other distances and larger numbers of samples.

A simple 4-bucket algorithm is given by [Wyant, 1982]:

$$\varphi = \arctan\left(\frac{I_2 - I_4}{I_3 - I_1}\right). \tag{2.4}$$

A 5-bucket algorithm is given by [Carre, 1966; Cheng & Wyant, 1985; Larkin, 1996]:

$$\varphi = \arctan\left(\frac{2(I_2 - I_4)}{-I_1 + 2I_3 - I_5}\right). \tag{2.5}$$

Instead of estimating the phase one can also use these algorithms to estimate the amplitude of the signal:

$$A_{2,\,3-Bucket} = (I_3 - I_1)^2 + (I_1 - 2I_2 + I_3)^2 \tag{2.6}$$

$$A_{2.5,\,4-Bucket} = (I_2 - I_4)^2 + (I_3 - I_1)^2 \tag{2.7}$$

$$A_{3,\ 5-Bucket} = 4(I_4 - I_2)^2 + (I_1 - 2I_3 + I_5)^2 \tag{2.8}$$

For three samples, there is a unique solution, but for more than three samples there are different options for making the resulting algorithms e.g. more robust to vibration [de Groot, 1995]. One modifications to make the 5-bucket algorithm more robust to sampling jitter is given by [Larkin, 1996]:

$$A_{3,\ 5-Bucket\ corrected} = |(I_2 - I_4)^2 + (I_1 - I_3)(I_3 - I_5)| \tag{2.9}$$

There are algorithms using more frames, in phase shifting interferometry (PSI) N is between 3 and 11 [Larkin, 1996; Hariharan et al., 1987] in most cases. Algorithms for much larger N can be derived (e.g. 101 bucket [de Groot, 1997]), which is interesting from a theoretical point of view as it shows a relationship to algorithms in the Fourier domain.

These algorithms can be implemented in hardware quite easily, therefore their implementation is described in more detail below and shown in Figure 2.5.

Single Side Band

Theoretically, the envelope of a sinusoidal signal can be obtained by applying the Hilbert transform, which creates a 90° shifted version of the original signal and can be used to obtain the instantaneous phase and frequency of a signal. This is also called "Single Side Band Transform" which will become clear when looking at the implementation in the Fourier domain. In the time domain the Hilbert transform requires convolution of the input values $f(z)$ with $\frac{1}{i\pi z}$:

$$H(z) = f(z) * \frac{1}{i\pi z}. \tag{2.10}$$

The so-called "analytic signal" is then given by

$$F(z) = f(z) + iH(z) = A(z) \cdot e^{i\varphi(z)} \tag{2.11}$$

with the signal phase

$$\varphi(z) = \tan^{-1}\left(\frac{H(z)}{f(z)}\right) \tag{2.12}$$

and signal amplitude

$$A(z) = \sqrt{f^2(z) + H^2(z)}. \tag{2.13}$$

A more efficient implementation is possible in the Fourier domain: The analytical signal can be obtained directly by performing a Fourier transform of the signal, setting the "negative frequencies" to zero and then performing an inverse Fourier transform. The results can be further improved by combining this with filtering in the Fourier domain, i.e. setting all frequencies to zero except those expected to be non-zero for

the given sampling distance and spectrum of the light source. This way noise can be suppresed. Instead of transforming the signal back, one can obtain the desired result directly by determining the slope of the phase in the Fourier domain — this algorithm (known as Frequency Domain Analysis, FDA) is not discussed further as it has been patented by P. de Groot from Zygo.

The computational complexity of these algorithms is relatively small as long as the number of points N used for the FFT is not too large. While the complexity of the other algorithms is linear in N, for the FFT it typically increases with $O(N \log N)$. For PCs there are very fast FFT implementations, but correlation or N-bucket algorithms are still simpler to implement, especially on signal processors or in hardware, even though the total number of multiplications might not be much different.

Results

The following results were obtained using the algorithms described above on two sets of simulated data. In the first case, only additive white noise was taken into account. In the second case, correlated position noise with the sampling positions according to the following equation was assumed:

$$\begin{aligned} s(t_i) &= s(t_{i-1}) + \delta_{pos,i} \\ \delta_{pos,i} &= c_1 \cdot delta_{pos,i-1} + n_i, \end{aligned} \tag{2.14}$$

with n AWGN (simulated for varying standard deviation) and c_1 a measure of correlation (chosen to be 0.9 for this simulation).

For the first case, the results according to Figure 2.2 have been obtained. The correlation based approach performs best, which is not surprising because it uses the actual sampling positions and the known modulation frequency. In case of correlated noise caused by sampling jitter, however, the result is totally different (Figure 2.3). Correlation based techniques fail as the assumptions on the sampling positions are not correct any more. It should be noted that both FFT and correlation based algorithms can be easily tuned and optimized for specific noise properties, for example by adjusting the length of the correlation or by filtering in the Fourier domain. This has not been done in this example, and it is not easily possible for N-Bucket algorithms.

Both of these models are not very realistic, as noise in practice will always be a combination of multiple sources. This is described in detail in [Seiffert, 2007]. The influence of laser speckle was neglected as well. However, the results obtained for the two cases above illustrate the fact that optimum estimation strongly depends on the correct choice of evaluation algorithm, and it shows that simple N-point algorithms that can easily be implemented in hardware offer fairly good performance in both cases. Using recorded actual sampling jitter as shown in appendix A shows similar results. A more detailed description of the software platform used and some additional simulations as well as results from measurements can be found in [Wagner, 2005].

2.1.3. Hardware Acceleration

For a significant improvement in performance, the PC as a bottleneck for data processing has to be removed. Parallel processing of the data is possible, but simply

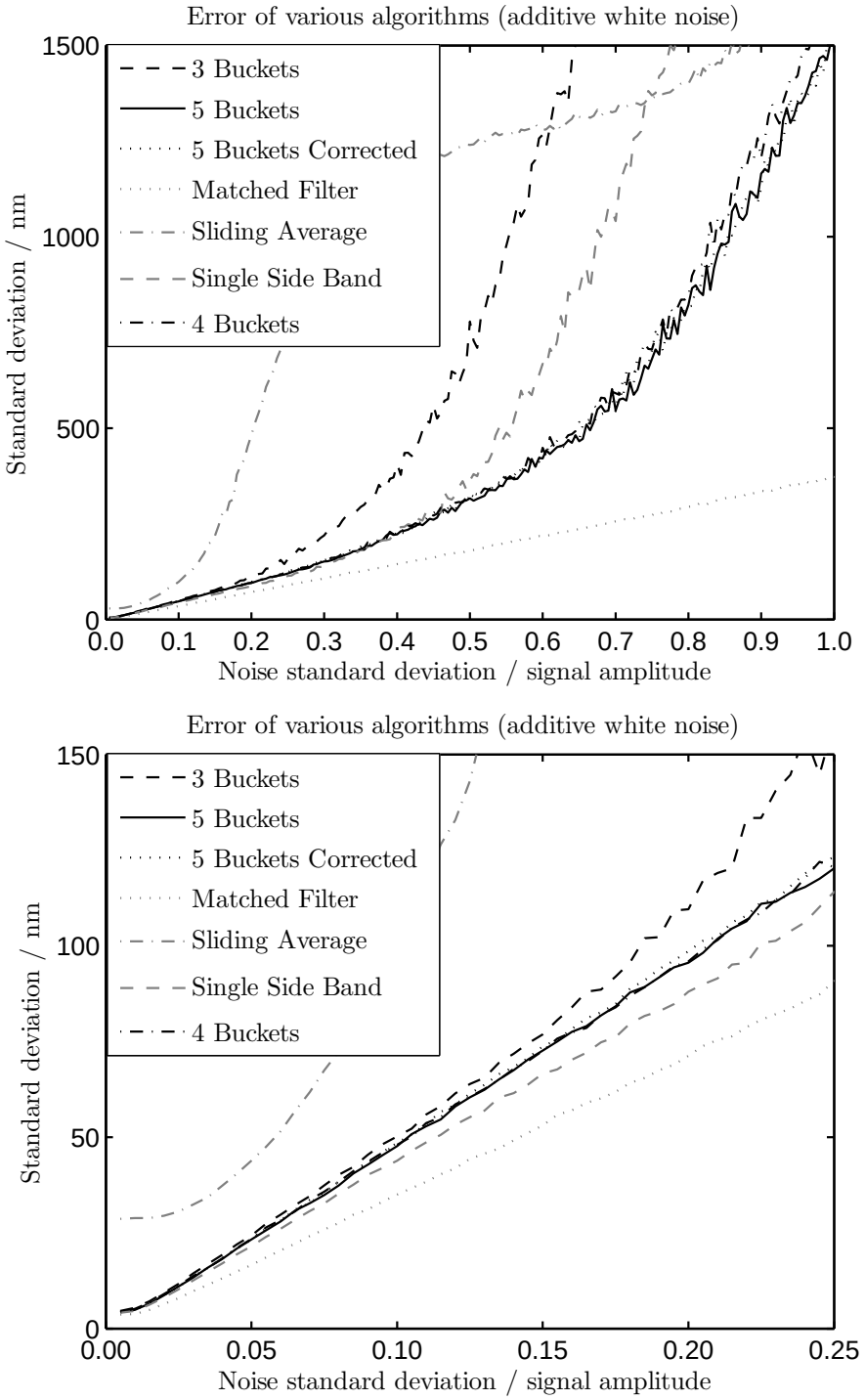


Figure 2.2.: Comparison of WLI algorithms in the presence of white noise, full view (top) and low noise only (bottom).

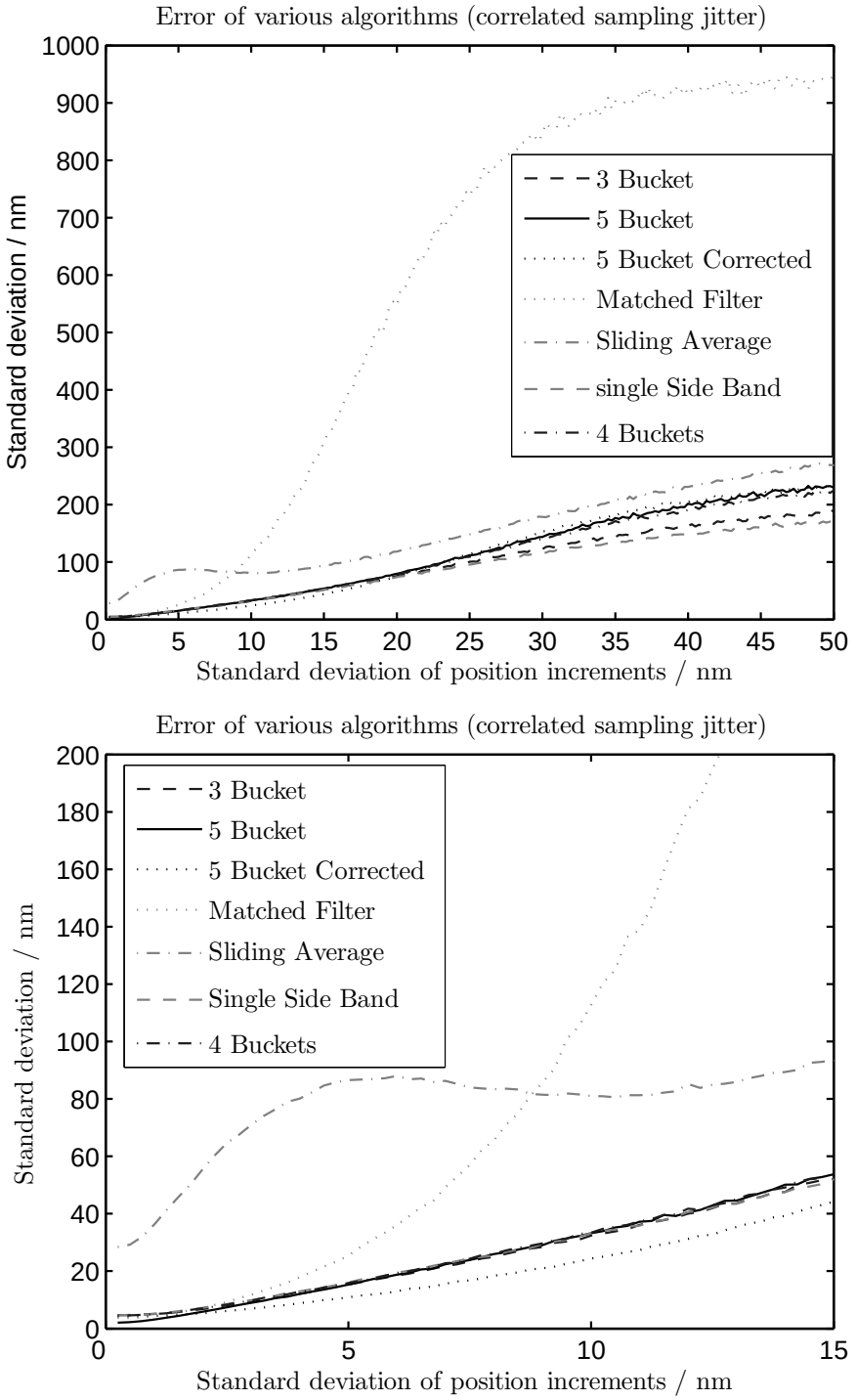


Figure 2.3.: Comparison of WLI algorithms in the presence of sampling jitter, modeled as correlated noise, full view (top) and low jitter only (bottom).

transferring the data to multiple PCs is difficult, as the amount of raw data is very large. Faster cameras (more than 600 million pixel per second) are readily available (though expensive), the same applies to faster framegrabbers. A higher camera speed can be used to increase lateral resolution significantly while keeping the exposure time constant. Alternatively, together with a higher intensity illumination (using new developments in LED technology), the exposure time can be reduced and thus the frame rate increased.

For that purpose, algorithms are analyzed with respect to their implementation on digital signal processors (DSPs), field programmable gate arrays (FPGAs), graphics processing units (GPUs) or multi-core systems, and their properties in the presence of sampling jitter and additive noise are investigated. In particular, several strategies for sharing the load between more specialized hardware and a standard PC are discussed. A true hardware implementation is not discussed here, as this is cost-prohibitive for a relative small number of measurement systems.

For hardware acceleration, all of the algorithms above have one thing in common: Each pixel is processed individually, which means that parallelization is trivial. This does not apply to spatial filtering, e.g. Bayesian methods [Hissmann, 2005], or to phase unwrapping (as used in classic phase shifting interferometry). These algorithms can still be parallelized quite well, but the pixels are not completely independent of each other any more. The focus in this thesis is on algorithms suitable for rough surfaces, and spatial relationships are typically not applicable there.

FPGA Implementation

As many framegrabbers and cameras have integrated FPGAs, the following discussion focuses on that aspect. Field programmable gate arrays provide very high computational power. Unlike a PC or a DSP, these are programmed using a hardware description language (e.g. VHDL), and a synthesis tool maps the logic to the elements physically available on the given FPGA. Large FPGAs today offer several hundred dedicated multipliers and hundreds of thousands of logic elements as well as several megabytes of fast embedded memory. Implementing a ring buffer as described above would only be possible using fast external memory, though — and while adding additional memory banks is possible, it increases system cost, and most current designs of cameras and framegrabers using FPGAs have limited memory bandwidth. This is by far the most parallel concept, thousands of operations can and will be performed in parallel.

There are significant differences between the algorithms discussed above:

For the minimum method, processing consists of one comparison per pixel and of finding the position of the smallest value. This only requires storage for two images (current minimum value for every pixel, and an “image” containing the index position of that minimum, which corresponds to the height value) and it is easy to handle for most memory architectures; on current processors cache sizes may be large enough to keep these values in the cache such that only the incoming pixels have to be loaded from system memory. For the contrast method, one more image has to be available and one additional subtraction is needed per pixel. For the sliding average, the number of operations increases only slightly (one more addition per pixel), but a much larger

number of images has to be kept in memory, and these images will definitely not all fit in the system cache, leading to at least twice the required memory bandwidth.

The accuracy of the methods described above is normally not sufficient, though. They are used for finding the approximate position of the optimum while still keeping the raw data in a ring buffer (in practice, a simple ring buffer is not sufficient, as there will be local maxima before the real global maximum occurs, so a somewhat more complex memory buffer structure using at least 1.5 times the number of desired elements is needed). This ring buffer is different for every pixel (i.e. memory access patterns depend on the position of the maximum of each correlogram). Such a complex structure is relatively well suited for processing on a PC, but very difficult to implement efficiently in hardware.

Typical FFT-based processing algorithms are not desirable either: They suffer from the large amount of memory accesses to data scattered all over the available memory (resulting from the ring buffers described above). It is obvious that FFT-based algorithms and in general all approaches that need to find and keep a range of pixels surrounding the optimum by pre-processing (as commonly used in PC-based system) are not a useful approach in an FPGA. First of all, the complexity is high and significantly increases development costs. Additionally, these algorithms will be memory limited, and most affordable framegrabbers have much less sophisticated memory management and smaller memory than a standard PC today. It is certainly possible to build a system with a very high memory bandwidth (there are enough I/O pins, and FPGA based systems are used for many applications requiring highly demanding I/O tasks), but the system cost is high in that case.

Therefore it is important to find alternative methods: If the pre-processing filter returns good estimates of the envelope itself, the ring buffer and the additional processing are not necessary, and only interpolation is needed to find the optimum position. Additionally, the data rate can be reduced significantly by lowering the number of frames used for the final interpolation. If the prefiltering uses a large number of frames, the individual estimates get fairly robust and are highly correlated, making subsampling of the filtered data possible and thus reducing the amount of data which has to be post-processed (and e.g. transferred to the PC).

Two possible algorithms are described in detail in the following: A structure for implementing N-bucket algorithms (Figure 2.5), and a concept for implementing correlation-based algorithm (Figure 2.7 and Figure 2.6).

The basic structure of both algorithms is shown in the next chapter, as they can be implemented much easier in line scanning white-light interferometry. All the data required for processing is easily available there, while it is spread over the whole measurement sequence in normal white-light interferometry, leading to memory bandwidth issues.

N-Bucket implementation

An N-bucket algorithm can be seen as filtering the input twice with an FIR filter, squaring the results and adding them up (Figure 2.5). But there is only one of the desired pixels in every camera frame, i.e. the pixels to be processed are spread far apart in the data stream, there may be a million unrelated pixels in between. The simple

approach of directly loading the relevant values from previous frames from memory must fail, as this would require a huge memory bandwidth: If the camera pixel clock is 500MHz (=500MB/s, if 8 bit/pixel are used) and a memory line has 128 bit (normally it is not possible — or at least not faster — to individually address 8 bits of data) and we would like to filter with 16 taps, we would need to read $500,000,000 \times 16 \times 15 = 120$GB/s. This does not take possible latency or addressing issues into account. And while this number is not completely impossible (as shown by current graphics cards), the author is not aware of any framegrabber getting even close to that.

It is very easy to reduce the effort, though: A number of N-bucket processing blocks can work in parallel, so that whole memory lines (e.g. 16 pixels, or preferably even larger blocks) can be processed at the same time. This might be limited by the number of multipliers on the FPGA, but for N-bucket algorithms almost no "real" multiplications are necessary. The "filter taps" can typically be implemented using additions and bit shifts, and only the squaring operations need two multipliers per pixel. If two or more memory banks are available, the frames can be written in a way that is alternating through the memory banks, making it possible to use the full memory bandwidth. It is also possible to re-order the data in memory while writing it by splitting the original frames and forming new frames that contain a range of spatial and temporal neighboring pixels in one block of memory. If these blocks contain all data from N consecutive frames for a certain number of pixels, only two such blocks have to be read for each filtering operation. This may be helpful to balance the complexity of read and write accesses, but its efficiency depends on the actual framegrabber hardware which cannot be discussed here.

This approach can be used to get accurate estimates of the envelope, and with some additional effort phase estimates can be obtained as well. The framegrabber does not perform all the processing, but is essentially a pre-processor, with the main difference being that the envelope detection is good and therefore the original data does not have to be preserved: no pixelwise ring-buffers are needed. The PC gets the resulting data at a significantly reduced data rate: As the filter output is highly correlated, downsampling is possible. Additionally, one can introduce arbitrary regions of interest, using a binary mask on the framegrabber to select interesting regions. The PC then searches for the maximum, implements the ring buffer, and finally performs e.g. a least squares curve fit to the data around the maximum. This is much faster than filtering all the data on the PC. It is also possible to search for the maximum directly on the frame grabber and directly return the height value, but this reduces accuracy.

The main disadvantage of such an approach is its lack of flexibility: It is not possible to change the number of taps or the sampling distance on the fly. A whole new compilation and place-and-route procedure is required as the N-bucket algorithms will have to be fixed (unless a very large number of multipliers is available on the FPGA, then it might be possible to keep some flexibility by changing the "filter taps" only).

Correlation implementation

Correlation based approaches are very attractive, too. Again, the straightforward implementation is shown in the chapter on line scanning WLI. Of the two concepts shown there, Figure 2.6 is usually more attractive due to the memory bandwidth issues in

white-light interferometry. In this case, a ring buffer containing the desired number of frames (equal to the correlation length) has to be implemented. For every pixel, there are two sliding averages which should be stored in internal memory of the FPGA as they have to be accessed very frequently. Six multiplications are needed for each pixel, but only one memory read (oldest frame in ring buffer) and one memory write (overwrite oldest frame with current frame data) is needed. The same values for sine and cosine are used for every pixel in a frame, but their value depends on the sampling pattern. They can e.g. be stored in an internal ring buffer (as the sampling pattern will usually be periodic). Again, multiple correlations can be computed in parallel, limited by the number of multipliers available. If a larger amount of internal memory and additional memory bandwidth is available, the number of multiplications can be reduced to four, at the cost of roughly four times the internal memory usage and twice the external memory bandwidth. In that case, two filtered values (with a higher bit depth) are stored instead of the raw pixel data as shown in Figure 2.7.

Further processing and additional options are similar to N-bucket algorithms. Correlation based approaches need more memory (as their correlation length is typically larger than the number of filter taps in N-bucket algorithms), but fewer memory accesses (two vs. N). They need less multipliers, but they need multiplications which cannot be easily replaced with bit shifts and additions (this is possible for certain sampling intervals, but this reduces flexibility and some of these intervals are known to perform poorly). Their main advantage is higher flexibility: Adapting the size of the ring buffer is easy, and the coefficients for the multiplication with sine and cosine can be changed easily.

Alternative Fast Implementations

There are several alternatives to FPGAs which are easier to program. First of all, multi-core systems become cheaper and more efficient every day; in 2004 almost all processors were single-core, and boards and processors with support for more than two processors were expensive. In 2007, some quad-core processors cost less than EUR 500, and systems using two quad-core processors are readily available. With the IBM "cell processor" (mainly known as the processor used by the Sony Playstation 3) an 8-core system with very good performance in many scientific applications is readily available [Williams et al., 2006]. Using multiple threads one can easily accelerate processing. Depending on the specific architecture, the system might be limited by memory bandwidth or PCI bus bandwidth (for the camera data) though. Nevertheless, a significant improvement at low cost can be expected from using such a system.

Dedicated signal processors usually offer very long instruction words and can perform the same operation on multiple blocks of data at once. Many are optimized for "multiply-accumulate" operations, can perform them in a single clock cycle and therefore perform very well at filtering. Several of these systems have been considered for this application, but there is one crucial problem: The order of the data and the required memory access create a bottleneck. If a whole series of measurements arrived in a single camera image, using a DSP would be easily possible, but with the current memory access patterns, there are no performance advantages over PCs, even though these are slower on raw multiply-accumulate operations. But the more flexible

Nominal depth in μm	Standard deviation in nm	Bias in nm
1	25	-12
5	28	-13
20	59	-90
50	56	95
200	88	566
600	105	-208
900	120	208

Table 2.1.: Standard deviation and bias obtained from 25 measurements of a height standard. Heights for each of the steps have been averaged first, and then the standard deviation of the height across all 25 measurements has been computed. The standard deviation of the pixels within each height step in a single measurement is on the order of 40nm. The bias has been computed using the calibration values from the PTB (not shown in this table).

memory access and better prefetch hardware help PCs.

A very special type of signal processor is highly interesting though: graphics processing units (GPUs). In 2004, their performance was high, but their flexibility was limited by short programs and low accuracy, and returning data from the GPU to the PC using AGP was slow. Today's (May 2007) graphics processors are able to perform all required filtering operations. They are typically integer only, but 32 bit accuracy is sufficient for WLI applications. They are highly parallel (320 shaders on ATI R600, 128 on nVidia G80), run at high clock frequencies (740 MHz ATI, 1450 MHz nVidia), have vast memory bandwidth (>100GB/s for both ATI and nVidia) and with PCIe x16 quick data transfer e.g. from a framegrabber to the graphics card and then to main memory is possible. Even multiple GPUs can be used in one system (called "SLI" by nVidia, "Crossfire" by ATI), and special systems for scientific applications are available.

2.1.4. Results

First a system using the components described above, but with the signal processing performed on a PC, was set up and optimized. Optimization included the choice of the best sampling intervals. These sampling intervals were chosen due to their high resistance to sampling jitter caused by the mechanical stage M511DG.K029, and have been found experimentally. A very large number of settings were used for repeatability measurements of a flatness standard, and the algorithms reaching the best accuracy were selected for further investigations. The influence of the stage is visible in repeatability measurements of a step height standard: In this case, the pixels on every height step were averaged to reduce other sources of error apart from the stage. Results are shown in Table 2.1.

Data analysis was performed by 3D-Shape's Korad software, which is able to preprocess more than 100MB/s of data on a 3Ghz Pentium 4. Final processing was performed using an SSB-based algorithm and least squares curve fitting for interpolation. In addition, measurements were performed using an implementation of the algorithms

Velocity in $\mu m/s$	Samples per signal period	Camera frame rate in fps	Standard deviation in nm
4.1	5	50	28.5
53.2	3.5	454	31.3
148.9	1.25	454	38.5
232.6	0.8	454	> 50.0

Table 2.2.: Standard deviation of flatness values based on repeated flatness measurements of 25 objects (absolute flatness values between 0.9 and 1.3 microns).

in Heurisko for comparison of the different possible algorithms. In this case, measurement speed had to be reduced by about a factor of 5, or — in order to analyze identical data — Korad was used to record videos which could then be analyzed off-line in Heurisko.

Subsampling reduces the number of data points and it reduces the signal contrast due to the camera integration time, but a faster measurement velocity can be reached with continuous stage movement and continuous illumination. The camera integration time has to be adjusted to avoid too much integration; there is a trade-off between reduced contrast due to integration and high relative noise due to camera characteristics and photon noise. This issue could be solved altogether by synchronizing the camera to the light source and using a high-intensity flashing illumination, but no such light source was available. The lower number of samples on the correlogram and the influence of the changing speckle field further reduce accuracy when subsampling, but for the values investigated sub-micron repeatability was still achieved.

The resulting system was tested extensively on various types of surfaces at measuring speeds of 4 microns per second (with 5 samples per signal period, 50 Hz camera), 53 microns per second (3.5 samples per signal period, 454 Hz camera) and 149 microns per second (1.25 samples per signal period and 454 Hz camera). At the slowest speed, the accuracy was comparable to the original system that used a CCD camera. Reducing the number of samples per signal period from 5 to 3.5 and increasing the camera frame rate to 454 Hz had no negative impact on the accuracy (in particular, performance with 3.5 samples per signal period was better than for e.g. 4 samples per signal period). Reducing the sampling rate further, to 1.25 samples per signal period turned out to be another good choice when looking at both speed and accuracy (values in between were significantly slower or did not perform better). The fastest setting with significantly lower but still reasonable accuracy was reached at 0.8 samples per signal period (230 micron/s with a 454 Hz camera). The system as discussed here is not ideal for further subsampling: Due to limited light intensity, the camera frame rate would have to be reduced significantly to avoid reduced contrast caused by signal integration. This was therefore not studied any further. For this analysis, the flatness of a flatness standard was measured. That way, the (maximum) error of a single pixel is found, and the influence of systematic errors caused by the stage is minimized as only a very small height difference is measured. This shows differences between various sampling strategies and algorithms. These results are summarized in Table 2.2.

The system as described above is not limited by any single component: The frame

rate is limited by the exposure time (caused by light source and camera quantum efficiency), transfer rate (limited by frame grabber interface) and processor speed for preprocessing of the data at the same time. Replacing a single component therefore cannot offer increased performance, and improving all components is significantly more expensive. The objective of optimizing all the components of the system by updating everything with fast but readily available components was therefore reached, and this part of the project finished.

In a second step, a closer look at hardware acceleration as described in the previous section showed that it is possible to adopt the algorithms for implementation on an FPGA. This way, a camera with higher pixel clock can be used. The performance of the algorithms has been demonstrated in a Heurisko implementation, and several concepts for hardware implementation have been developed in cooperation with framegrabber manufacturers. Their performance was simulated in detail (including aspects such as the influence of integer arithmetic, necessary bit depth in the individual processing steps etc.). No fundamental difficulties have been found.

While the expected performance benefit compared to other systems previously available was huge, the advantage in speed compared to the system developed in the first step was only about a factor of two, and it was obvious that this advantage would vanish rather soon when looking at alternative approaches (multi-core systems and GPUs), which were already on the horizon in 2005. While faster FPGAs will become available in the future as well, porting code to a new FPGA is much more difficult than porting code to a faster PC, which makes the FPGA option less flexible and therefore less attractive. In the end, the expected cost was too high for the anticipated temporary performance advantage (as mentioned above, almost all system components would have had to be changed), and therefore development was focused on other systems.

2.2. Line Scanning WLI

A line scanning approach has a number of advantages: As the measurement is now "single-shot", it is much more robust to vibration and can be used to measure dynamic processes. For a measurement of a surface a lateral scan is now needed though, which makes it less attractive for planar objects. If the measured object is cylindrical or if a number of measured lines on the object is sufficient, a line scanning system offers many advantages. All the algorithms and design options discussed for normal white-light interferometry can be applied to this system as well, therefore the following description will only highlight a number of important differences and a few results from optimization of the algorithms. For a more detailed description especially of the optical setup and laser speckle it is referred to [Hering, 2007].

2.2.1. Optical Setup

The optical setup for a line scanning system is more complicated; it is shown in Figure 2.4. This is a modified Mach-Zehnder interferometer. The optical aspects are discussed in much more detail by [Hering, 2007]; this analysis is not repeated here. The key idea of the setup is to introduce a spatial phase shift between object and reference path, perpendicular to the measured line (using cylinder lenses and a tilt of the

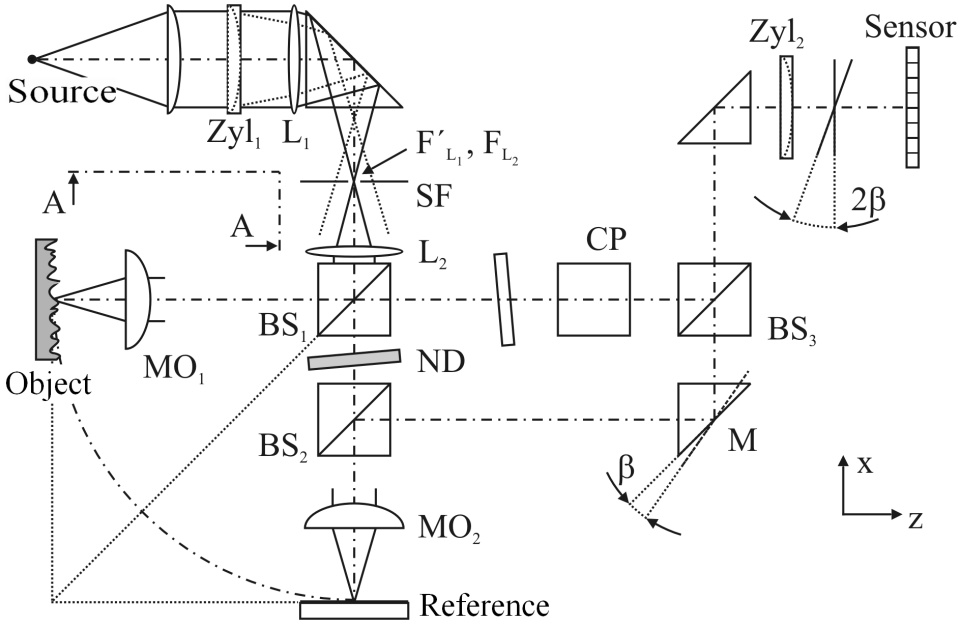


Figure 2.4.: Optical setup of a line scanning white light interferometer, using a spatial phase shift [Hering, 2007].

mirror M). The superposition is then recorded by a 2-D sensor array (camera), and therefore one obtains a correlogram for every measured pixel of the line directly, in a single shot.

As a sensor, the same CMOS camera as used in the previous chapter was used at first, which was later replaced by a CCD camera (SVCAM-svs2020U from VISTEK, using a Kodak KAI-2020 CCD chip) with more homogeneous pixels and higher resolution (1600×1200, yielding a larger measurement range). This camera is slower though, and a freely configurable field of view (to adjust the desired line width and height range) would be very nice in this application. The recently released successor to the original CMOS camera, the Photonfocus MV-D1024E-CL-160, looks very promising for that application (higher bit depth, lower noise, more homogeneous, camera read-out and exposure of the next frame in parallel), but its lower resolution (1024×1024) and thus measuring range might be a problem for some applications.

2.2.2. Sampling and Signal Properties

There are two key differences to the high-speed WLI system discussed in the previous sections:

- The system by design has no random sampling jitter. This makes an analysis of the data using correlation based algorithms easier.
- Now a single correlogram is recorded by multiple different pixels, which leads to much stronger requirements on the linearity of the camera across all pixels. This makes it necessary to calibrate the system.

In addition to that, there is significant additive noise and correlated noise due to laser speckle. The properties are different from a normal WLI system and are discussed in detail in [Hering, 2007].

2.2.3. Algorithms

Implementing signal processing algorithms is generally easier for that system than for "normal" white-light interferometry, as all the data that has to be processed is available at the same time in a single frame. This reduces memory requirements and makes processing right in the camera or on the framegrabber much easier. The approaches described for classic white light interferometry can all be used; the optimum algorithm again depends on the light source used and on the speckle field. In his work Hering looked at algorithms working in the Fourier domain for signal analysis: SSB and FDA (see section 2.1.2). These are both good and flexible, but difficult to implement in hardware - therefore work here focuses on the application of N-bucket or correlation based approaches instead.

2.2.4. Hardware Acceleration

In this case, hardware implementation is possible using off-the-shelf components: All data for the measured pixels is contained in a single camera frame, and some of the algorithms can be described as filters. Many cameras and frame grabbers offer the ability to directly filter the input data, which can be used to directly transfer an estimate of the signal envelope to the PC, possibly already subsampled.

The two examples from the previous chapter will be shown again, now adopted to line scanning white-light interferometry.

Implementation

N-bucket algorithms can be (partially) implemented using standard image processing filters (Figure 2.5). Using two Nx1 filter masks, one can simply convolve the input image with each of the filters and obtain two filtered images. This is a standard image processing operation and supported by many cameras and framegrabbers, even for relatively large filter sizes. Then each pixel in the two resulting images has to be squared and the two images can be added to obtain an image containing an estimate of the envelope in every line. While squaring a pixel is not a standard image processing operation, it is quite simple and can easily be implemented on an FPGA. The resulting image can be transferred to the PC where a maximum search and local interpolation (optionally using a least squares fitting procedure and/or Bayesian estimation) follow. If a freely configurable system is available, an implementation with a much lower computational effort is possible. N-bucket algorithms do not need real multiplications, but typically use small integers only to keep computational effort low. These "multiplications" can be implemented using bit shifts and additions.

The same approach can also be used for correlation-based algorithms, using two Nx1 filter masks with sinusoidal weights (shifted 90^{circ} relative to each other), where N now corresponds to the desired length of the correlation. As the correlation can be

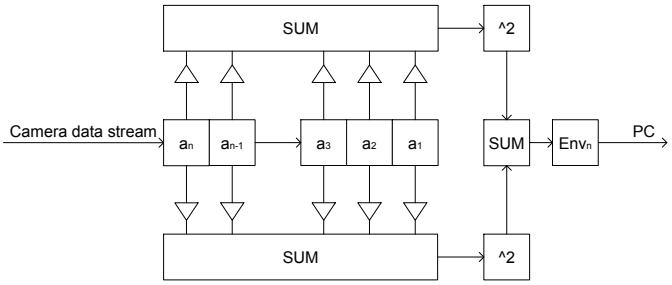


Figure 2.5.: N-bucket algorithm for LSWLI. The incoming data from the camera is placed in a shift register of length N. In every clock cycle, two FIR filter values are computed (typically hardcoded as no real multiplications but only bit-shifts and additions are needed). These values are then squared and added up. The resulting value Env_n will have very small latency and higher bit depth than the input values. In order to reduce computational effort and data rate to the PC, the shift register can sometimes be shifted without computing a new output value, returning a filter values every two or three pixels only. 2 "real" and $2N$ "simple"multiplications are required per pixel, storage requirement is a shift register of length N, some input and output buffers and storage for intermediate values in the computation.

implemented as a running sum, one can use that property to reduce the computational effort in the implementation. There are two possible solutions (Figure 2.7 and Figure 2.6). The first of these uses less memory, the other uses less multipliers. Both need to know the sampling intervals, which can be stored in a ring buffer.

Both approaches should be very fast and need very little memory (as long as all pixels of a correlogram arrive consecutively; some buffers might be needed in case of some multi-tap Cameralink cameras), so that they can be used with high-speed cameras. The diagrams only show the processing needed for one correlogram (essentially a pipelined structure that outputs one output for every input value); in an actual FPGA implementation it might be better to implement multiple such blocks and run them in parallel, but at a reduced clock frequency.

It is possible to go one step further: If a camera with an integrated FPGA is used, the algorithms described above as well as the maximum search can be integrated. Adding a ring buffer and a least squares fitting procedure around the maximum is difficult, though. However, if a relatively large N for N-bucket algorithms or a large correlation length is selected, an interpolation is already performed implicitly (though limited to the sampling grid), and the maximum of the obtained envelope is a good

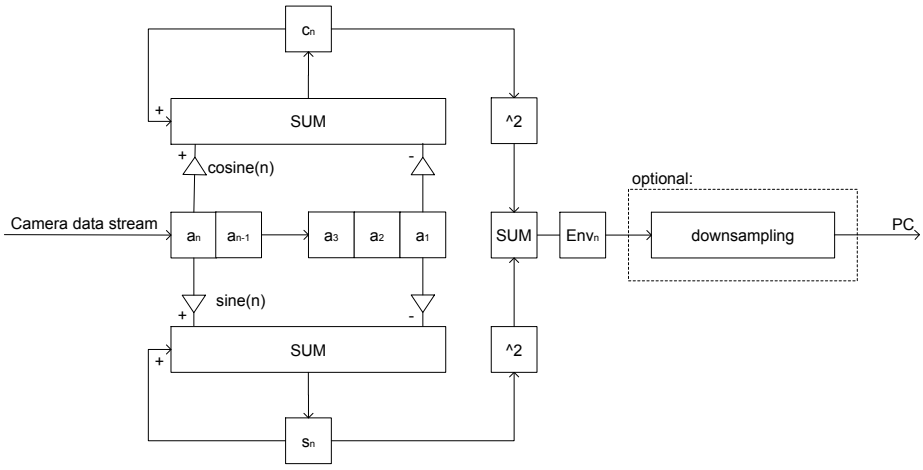


Figure 2.6.: Correlation-based algorithm for line scanning white-light interferometry. This implementation is very similar to the N-bucket algorithm described previously, but it uses the fact that the filter mask does not have to be shifted with the data in this case: A "sliding average" is sufficient. Again, the input data is put into a shift register, but now only the first and the last element have to be multiplied with a coefficient, which can be taken from an internal ring buffer. The entries of this buffer can be set according to the sampling pattern. This approach reduces the number of multiplications to six per pixel. In this algorithm, the computation has to be performed every clock cycle, but in order to reduce the data rate, downsampling at the output is possible. This algorithm is very attractive for standard white-light interferometry as well.

measure of the height. The disadvantage of that approach is that the resolution is limited to the spatial phase shift between neighboring pixels (on the order of 100nm). For measurements on smooth surfaces, this is not acceptable, but for measurements of rough surfaces this leads to a relatively small increase in measurement error (which is then dominated by laser speckle). Such a system constitutes a line camera that measures height values instead of intensities. It is therefore easy to integrate into an automatic inspection system and less prone to errors compared to a PC that requires an operating system, network communication and a lot of other overhead.

2.2.5. Results

Only a prototype system exists right now. Its properties (especially the bandwidth of its light source) are currently far from optimal for rough surfaces. A sample measurement taken from [Hering, 2007] is shown in Figure 2.8. FDA was used for signal processing.

In Figure 2.9 the raw data for a camera line is shown. The resulting envelope func-

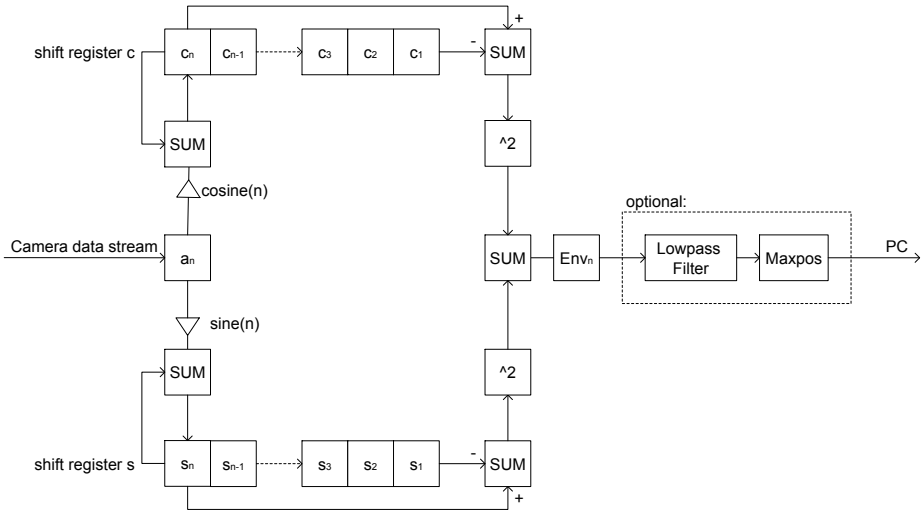


Figure 2.7.: Correlation-based algorithm for line scanning white-light interferometry. The algorithm described previously can be optimized further, but whether this solution is really better depends on the hardware used. The number of multipliers needed can be reduced to four by storing the cumulative sums instead of the raw data. This requires additional storage and higher bit depth in the additions, though. It is sufficient to choose a bit depth that ensures that the maximum value of the correlations can be accommodated, an overflow then simply leads to a "wrap around" that can be detected and corrected when the difference at the end is negative. This configuration makes it possible to implement more parallel processing elements on small FPGAs with a limited number of multipliers. At the output, an additional low-pass filter can be implemented (FIR or IIR), and then the position of the maximum can be recorded (which just needs a global frame counter and one additional register for the index of the maximum per pixel). Then this system is a compact, single-shot line camera that directly returns height information with interferometric accuracy (though limited to the sampling grid in this implementation).

tion using two different algorithms with two different settings each is also shown. SSB is used with a narrow filter and a wide filter in the Fourier domain; the simplified correlation as described above is computed using 100 and 27 pixels respectively. It is clearly visible that the results are slightly different, and the correlation based envelope is slightly noisier than SSB using a narrow filter. However, without ground truth it is not possible to tell which one is better. The coherence length of the light source used here is very large, which leads to a very broad envelope and decreases the accuracy of the results. Based on that data, it is impossible to predict which algorithm will perform better in a future improved system: The performance of the algorithms strongly depends on the signal properties, as described in the previous chapter. Using a better camera, better optics, adjusting the system for better speckle properties etc. will have

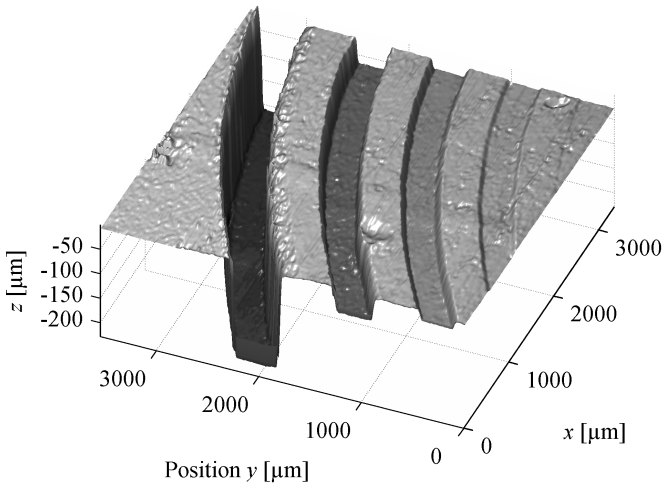


Figure 2.8.: Step height standard measured by line scanning WLI; the visible steps have calibration values of 0.97, 4.96, 19.90, 49.76 and 199.73 micron.

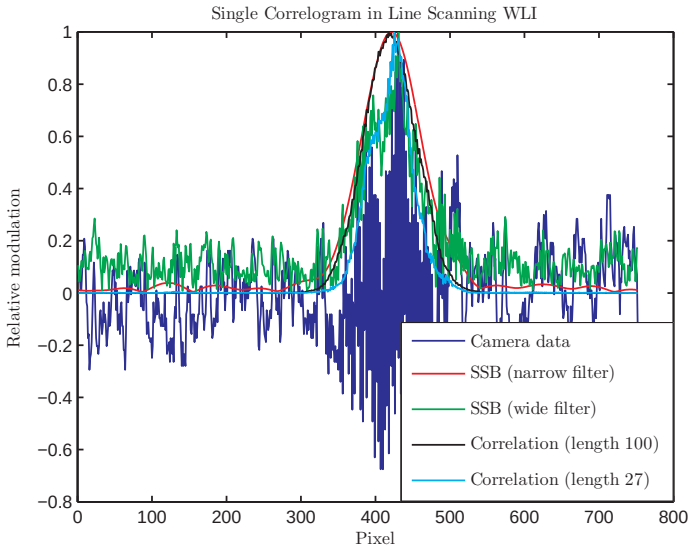


Figure 2.9.: Analysis of a single correlogram in line scanning WLI, using single side band and correlation based algorithms with different parameters.

a significant influence on that.

As this has not been pursued any further yet, no detailed quantitative analysis of the system and the performance of hardware accelerated algorithms can be given here.

3. Multiple Wavelength Interferometry

The main focus of this thesis is not on white-light interferometry but on fast interferometric measurement system for use in a production line. The principle of multiple wavelength interferometry, in particular frequency scanning multiple wavelength interferometry, is particularly interesting. A basic overview on key components and the signal model has already been given in the introduction.

In this chapter, several different aspects of the system will be discussed in detail. There are eleven sections describing all important aspects of modeling, simulation, implementation and results. In the first section, the special hardware required for a frequency scanning interferometer is described in more detail. Based on the information in the introduction and the description of the hardware, the signal model is derived in section two. This signal model is the basis for a derivation of the theoretical limits on accuracy in section three. Section four discusses the optimum sampling for the derived signal model, using the theoretical accuracy and ambiguity constraints as basis for the optimization. In section five, an approximation to the theoretically optimum sampling pattern is presented and a multi-step algorithm for frequency estimation based on that approximation is derived. In section six, very fast algorithms for frequency and phase estimation from short blocks of data are discussed, and new estimation algorithms developed. This is also an optimization problem, and the results are compared to the appropriate theoretical limits. In chapter seven, the aspect of monitoring the actual sampling positions (i.e. the actual wavelengths of the tunable laser in case of frequency scanning interferometry), is discussed and a method for calibration is presented. In chapter eight, approaches for using spatial information are discussed, and a fast, filter-based concept for improving the results is presented. At this point, all key components of the processing algorithm have been described, and therefore in chapter nine two specific implementations are described and compared. In chapter ten, simulations and measurement results obtained on various surfaces with different settings are described and the accuracy of the system is discussed. Based on these results, in chapter eleven the influence of speckle on the measurement of rough surfaces is discussed.

3.1. Hardware

In this chapter the key hardware components of a multiple wavelength interferometry system are described. The basic optical configuration of an interferometer has been discussed previously, therefore it is not repeated here. The other important component is the light source, in this case a tunable laser. For the camera the explanations from chapter 1.5.1 apply; the main difference is that for the multiple wavelength system a high camera speed is less relevant than a good dynamic range and a high SNR, as the number of frames needed is much lower.

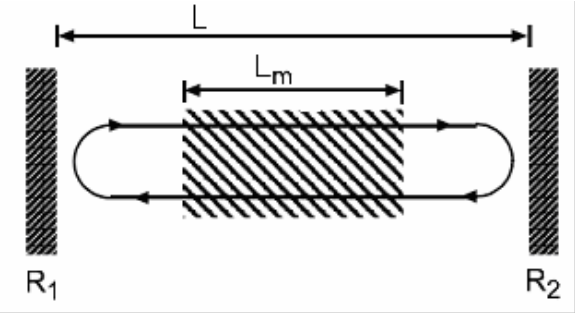


Figure 3.1.: Fabry-Pérot resonator

3.1.1. Tunable Lasers

For a multiple wavelength interferometry system, either a number of fixed wavelength laser sources or a tunable laser is required. The following analysis focuses on the latter, but in principle other configurations are possible. This subsection uses some material and graphs created by Francisca Klenke as part of her diploma thesis [Klenke, 2007].

Basic Laser Principles

The central part of a laser is the active laser medium where energy is added by a so called pumping process. Energy can be added e.g. by applying a voltage or using another light source, through which the electrons of the active laser medium reach a stimulated state. Passing photons with corresponding energy (i.e. the energy difference the electrons have to a lower energy state), can cause the electrons to return to a lower state and send out the surplus energy as a photon with the same frequency and phase as the passing one (stimulated emission). In addition to that, a resonator is needed which couples the light back into the active laser medium, such that the light can be amplified several times.

The simplest form of a resonator is the Fabry-Pérot resonator (Figure 3.1), which consists of two parallel mirrors surrounding the active laser medium.

For the laser to work, the amplification of a photon passing through the active laser medium must be more probable than its absorption (inversion). The reflection indices of the partially transmitting mirrors R_1 and R_2 have to be sufficiently high to keep enough light inside the resonator and the active laser medium. Additionally, the light that is coupled back into the laser must have the correct phase to avoid destructive interference. Because of this condition there is only a set of frequencies possible, which have an equidistant spacing between each other. Those frequencies are called modes and their frequencies are given by the following equation:

$$f_q = q\frac{c}{2L},\ q \in \mathbb{N} \tag{3.1}$$

with $L = L + L_m(n-1)$ and f_q distance between two adjacent frequencies, c speed of light, L length of the resonator, L_m length of the active laser medium and n index of refraction of the active laser medium.

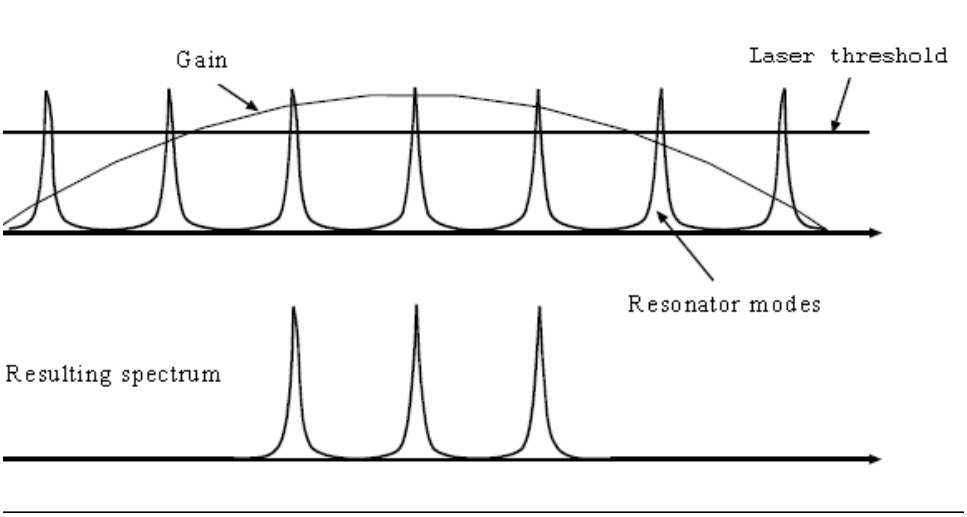


Figure 3.2.: Laser gain and possible laser modes

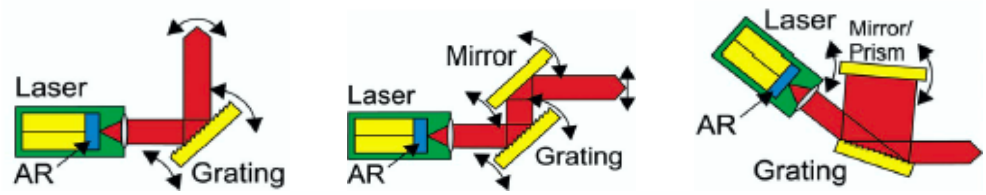
Theoretically, the number of the possible laser modes according to this equation is infinite. But the gain is only high enough in a certain frequency range (depending on properties of the active medium, especially the possible energy states of the stimulated electrons). Only if the gain is higher than the resonator losses (laser threshold), the laser works (Figure 3.2). The assumption that a laser sends out only one frequency (or mode) at a time is often incorrect, this depends on the configuration of the laser.

For application in an industrial measurement system, the tunable laser has to be compact and low cost. Laser power is not important as very little intensity is needed for imaging in this case. A laser diode as commonly used in CD writers is chosen in our case. The output power of such a laser is typically limited to about 100mW (only about 0.1mW are needed). The active laser medium is the pn-junction of the diode. Pumping is performed by applying a voltage and therefore causing a current through the diode. This laser diode does not emit a single mode, but several equally spaced modes (as explained above). The spacing $\Delta\lambda$ is given by

$$\Delta\lambda \approx \frac{c}{2nL} \tag{3.2}$$

with c speed of light, n index of refraction of the laser medium and L length of the resonator. This is a simplified model, more details and other aspects can be found in [Nagengast, 1995].

There are a number of disadvantages of laser diodes: They are very sensitive to temperature changes (for AlGaAS, a change in temperature of $\Delta T \approx 0.1K$ causes a frequency drift of $\Delta f \approx 3$GHz). The amplification also changes with temperature. Beam quality is generally not good, but can be improved with vertical cavity surface-emitting lasers (VCSELs), and diodes are sensitive to back-reflections - this is used for tuning. In addition to that, continuous tuning over a larger frequency range is not possible - but this can actually help for the measurement application as will be discussed later.



(a) Littrow configuration

(b) extended Littrow configuration

(c) Littman-Metcalf configuration

Figure 3.3.: Possible configurations for tunable lasers with external cavity and grating.

Laser Tuning

Standard laser diodes can be tuned by changing the temperature or the current through the diode. The modulation of the temperature offers a fairly large (several hundred GHz) hop-free tuning range, but it is quite slow. With modulation of the current the frequency range is much smaller (10kHz – MHz), but the modulation can be done faster. High currents or high temperatures decrease the lifetime of the diode, and aging might lead to gaps in the tuning range.

An alternative approach is using an external cavity. In that case, the reflection on one side of the laser diode is reduced (using anti-reflective coating). The diode is not a laser any more, as there is no resonator left. This is then introduced using an external cavity with a mirror, grating or prism. Changing the optical length of the resonator $L = nL_{LD} + L_0$ can be used to change the laser frequency. The distance between adjacent modes depends on the new resonator length.

In practice, the laser is usually tuned not by changing the resonator distance, but by coupling back light of a certain frequency using a grating. This is possible as the angle of the reflected light beam depends on the wavelength, and therefore the grating can be used to selectively couple light of a specific wavelength back into the diode. This can only work if the diode dimensions permit it, so in addition to that feedback the laser voltage and temperature have to be adjusted accordingly. With this configuration mode jumps occur. For a continuous tuning the angle of the grating and the length of the resonator would have to be changed at the same time, which requires more effort (and is not necessary for our application).

For tuning with an external grating there are several common configurations (Figure 3.3). Using the so called Littrow design is the easiest option, but the extended Littrow design and the Littman-Metcalf design help resolve the problem of a moving beam when the laser frequency is changed.

The Littrow configuration is inexpensive and easy to realize, but the beam angle changes when tuning. The extended Littrow configuration keeps the beam angle constant, but there is a parallel shift of the beam, which complicates e.g. fiber coupling. The Littman-Metcalf configuration keeps the beam angle and position constant. The additional reflections in this configuration increase resonator losses, though.

3.1.2. Monitor Cavity

As the laser diode is very sensitive to temperature changes, the actual frequency of the laser has to be monitored. This can be done by using a special type of interferometer. It consists of a glass plate which reflects a part of the laser beam at the top and another part at the bottom surface. The two reflected beams interfere, and a line along the diameter of the resulting circular interference pattern can be recorded by a line camera. This interference pattern changes with the laser frequency, and this change can be detected and analyzed. A very high sensitivity can be reached by using a relatively thick glass plate; in order to resolve the resulting closely spaced fringes the line camera can be tilted. Many options to monitor and stabilize the laser for use in a multiple wavelength system are presented by [Salvadé, 1999].

The analysis of this signal and the laser properties are discussed in detail in chapter 3.7.

3.1.3. Interferometer Setup

There are several types of interferometers, but this will not be discussed in detail here. Two important concepts have already been presented in the introduction (Fizeau and Michelson setup). The most important differences between these for a multiple wavelength system will be highlighted here.

A Fizeau configuration has a common path, and therefore the interferometer itself is highly robust to vibration: The same effects occur in both reference and object path, and therefore cancel out. This is very useful for a multiple wavelength system, as the data evaluation is highly sensitive to vibration. However there are two disadvantages: The reference plane is always in the same, fixed position right at the measurement head. This means that the working distance for many measurements is limited, as with increasing distance to the reference plane the processing becomes more and more sensitive to laser frequency variations.

In case of a Michelson setup, the virtual reference plane can be chosen to be in the middle of the object height range, with the working range available on both sides. Therefore a Michelson system can have a much larger working distance and measure objects of roughly twice the height. In addition to that, in a Michelson system it is easier to adjust the reference and object intensity by introducing a changeable filter into the reference path (and a clear glass plate to correct for dispersion effects in the object path). In a Fizeau setup, the Fizeau plate has to be changed, which is typically harder to do.

The optimum choice depends strongly on the laser properties and on the expected vibrations: For low vibrations and a less stable laser, the Michelson setup is preferable. For stronger vibrations or a highly stable laser, a Fizeau configuration is better.

3.1.4. Camera

The data is acquired by a camera. In contrast to white-light interferometry, the maximum frame rate of the camera is less important, as the number of frames is much lower. Spatial relations between neighboring pixels are not needed either. Sensitivity is less important than a good SNR, therefore a CMOS camera with a comparatively

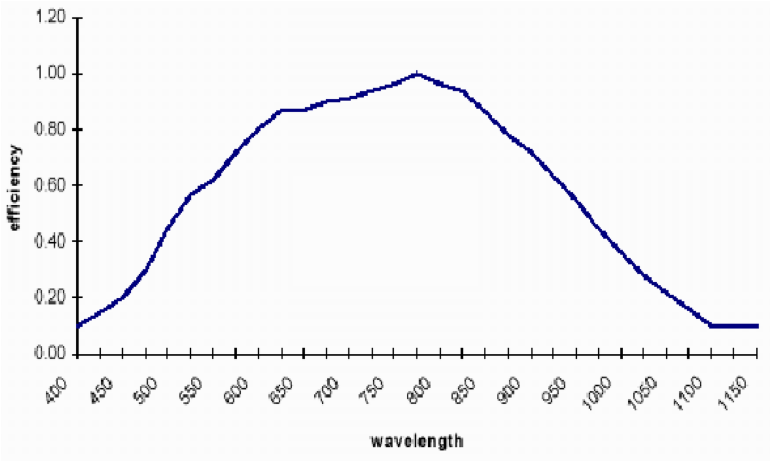


Figure 3.4.: Camera characteristic of the CMOS camera used for frequency scanning interferometry. This graph shows relative sensitivity, therefore the numbers cannot be compared directly to the ones in Figure 2.1.

large full well capacity is used. Its characteristic shows a very high sensitivity in the desired frequency range of about 790nm, which makes the system less sensitive to the influence of other light sources.

3.2. Signal Model

In this section, the signal model for a frequency scanning interferometer and its most important properties are derived. The relationship to the well known frequency estimation problem is shown. A theoretical limit on the accuracy based on the Cramér-Rao lower bound is given in the next section. This constitutes the basis for the optimization of the signal acquisition in chapter 3.4.2 and the reference the algorithms presented in chapters 3.5.1 and 3.6 are compared to.

Light emitted by a laser is (at least approximately) monochromatic and polarized, and can be described as a plane electro-magnetic wave. The camera records the intensity of the superposition of the waves coming from the object and the reference plane.

If the wavelength is changed linearly, the signal recorded by the camera is a sinusoid (cf. 1.7):

$$I(f) = A\cos\left(\frac{4\pi}{c}\cdot f\cdot\Delta h+\phi_0\right)+C \tag{3.3}$$

with modulation A, phase ϕ_0, offset C and frequency $\omega = \frac{4\pi}{c}\cdot\Delta h$. The signal frequency ω is proportional to the desired parameter, the true object height.

With the tunable diode laser discussed here, continuous frequency tuning is not possible. This can be taken into account by replacing f with $f_0 + n\Delta f$ with $n =$

$-(N-1)/2 \ldots (N-1)/2$:

$$I(n) = A\cos\left(\frac{4\pi}{c}\cdot(f_0 + n\Delta f)\cdot\Delta h + \phi_0\right) + C, \quad \text{with } n = -(N-1)/2\ldots(N-1)/2,\ N \in \mathbb{N} \tag{3.4}$$

f_0 is the mean laser frequency, Δf represents the frequency increments due to the laser mode jumps. Using that we obtain the following equation for the object height:

$$\Delta h = \frac{\omega c}{4\pi\Delta f} \tag{3.5}$$

There are a number of observations that are very important for the system design and signal analysis later on:

- If there is an unknown phase jump at the surface, there is an additional phase offset of the signal. This offset can be estimated from the signal, as will be discussed later. It is zero if reflection occurs at a smooth surface and the absolute signal frequency is known exactly.
- If the absolute laser frequency is not known exactly, there is a height-dependent phase variation. It is proportional to the object height range and the frequency error. This imposes requirements on the accuracy of the laser frequency monitoring.
- There is an ambiguity issue as the laser frequencies are on a fixed grid given by the laser mode jumps. There is no way to distinguish between ω and $2\pi - \omega$ or $2\pi + \omega$. This is well known from sampling theory: only frequencies up to half the sampling frequency (Nyquist frequency) can be sampled without aliasing. This leads to a corresponding ambiguity interval size d of

  $$d = \frac{c}{4\Delta f} \approx 1.6\text{mm} \tag{3.6}$$

  for the laser diode used here. If there is prior knowledge available on the part position, larger heights do not pose a problem as the height can be mapped to the correct interval. The change in the sign of ω above leads to the height map being inverted in every second interval, which has to be taken into account as well. There should be no measurements on the border of the intervals (i.e. at zero or Nyquist frequency), as in these areas measurement accuracy is very low (as shown below). The interval size can only be increased by using closer spaced frequencies. Theoretically, a single frequency not on the same grid would be sufficient, but this setup would be very sensitive to noise.
- A systematic error in the measurement of the frequency increments has no direct influence on the frequency estimation, but when the frequency is mapped to the absolute height, the result might be incorrect as the correct value of Δf is required for this conversion. This error is proportional to the absolute height and limits the working distance of the measurement system. It will also influence relative distances as the same error might lead to both increases and decreases in estimated height values, depending on the ambiguity interval (Nyquist frequency).

3.3. Theoretical Accuracy

The height estimation problem leads to a frequency estimation problem as has been shown in the previous chapter. Frequency estimation is a very well known problem, but the author has not found a derivation for the exact same problem (real-valued single tone frequency estimation with unknown phase, offset and amplitude). First, the derivation for the complex valued signal model according to [Rife & Boorstyn, 1974] is briefly repeated, and then the CRB for the real valued signal model as required in this application and some approximations to it are discussed.

If the noise is assumed to be independent and identically distributed and no offset C is present, the complex signal model leads to a very compact and simple result due to the rotational symmetry in the complex plane. As the CRB for the real signal model approaches these results asymptotically, this result is useful for all frequency estimation tasks. The following is summarized from [Rife & Boorstyn, 1974] with the notation adopted to match the other parts of this thesis.

$$\begin{aligned} x &= y(t;\theta) + n_x;\ y(t;\omega,\phi,A,C) = A\cdot\cos(\omega t-\phi) \\ v &= w(t;\theta) + n_v;\ w(t;\omega,\phi,A,C) = A\cdot\sin(\omega t-\phi) \end{aligned} \tag{3.7}$$

$$f(x,\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^N \cdot e^{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x-y(t;\theta))^2+(v-w(t;\theta))^2} \tag{3.8}$$

With equation 1.27 the following result for the first element of the Fisher information matrix is obtained:

$$\begin{aligned} I_{11} =& E\left[\left(-\frac{1}{2\sigma^2}\frac{\partial}{\partial\omega}\sum_{n=0}^{N-1}(x_n-y_n(t_n;\omega))^2+(v_n-w_n(t_n;\omega))^2\right)^2\right] \\ =& E\left[\left(-\frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x_n-y_n(\omega))\cdot(-t_nw_n(\omega))+(v_n-w_n(\omega))\cdot(t_ny_n(\omega))\right)^2\right] \\ =& \frac{1}{\sigma^4}\sum_{n=0}^{N-1}\Big(E\left[(x_n-y_n)^2\right]{t_n}^2{w_n}^2+E\left[(v_n-w_n)^2\right]{t_n}^2{y_n}^2 \\ & -2E\left[(x_n-y_n)(v_n-w_n)\right]{t_n}^2w_ny_n\Big) \\ & +\frac{1}{\sigma^4}\sum_{n=0}^{N-1}\sum_{\substack{m=0\\m\neq n}}^{N-1}E\left[(x_n-y_n)(v_m-w_m)\right]\cdot(\ldots) \end{aligned} \tag{3.9}$$

This derivation is only valid for uncorrelated noise, and the noise on the real and

imaginary parts has to be independent. Then we can simplify the result:

$$\begin{aligned} I_{11} =& \frac{1}{\sigma^4} \sum_{n=0}^{N-1} \left(\sigma^2 {t_n}^2 {w_n}^2 + \sigma^2 {t_n}^2 {y_n}^2 - 0 \right) + 0 \\ =& \frac{1}{\sigma^2} \sum_{n=0}^{N-1} {t_n}^2 \cdot ({w_n}^2 + {y_n}^2) \\ =& \frac{1}{\sigma^2} \sum_{n=0}^{N-1} {t_n}^2 \cdot \left(A^2 \sin^2(\omega t_n + \phi) + A^2 \cos^2(\omega t_n + \phi) \right) \\ =& \left(\frac{A}{\sigma} \right)^2 \cdot \sum_{n=0}^{N-1} {t_n}^2 \end{aligned} \tag{3.10}$$

The original paper [Rife & Boorstyn, 1974] is slightly misleading here: If n_x, n_v are Hilbert transforms of each other and only one of them is actually measured, the results above do not apply (the expectation over the mixed term does not yield zero). Independent measurement data of both the real and the imaginary part of the complex signal is needed for this relationship to hold. One cannot just take the real signal, perform a Hilbert transform and apply the equations above to the result.

The other elements of the Fisher information matrix can be computed similarly, and yield

$$I = \frac{1}{\sigma^2} \begin{bmatrix} A^2 \sum_{n=0}^{N-1} t_n^2 & 0 & A^2 \sum_{n=0}^{N-1} t_n \\ 0 & N & 0 \\ A^2 \sum_{n=0}^{N-1} t_n & 0 & A^2 N \end{bmatrix} \tag{3.11}$$

Based on that result, one can easily derive various bounds for frequency and phase estimation accuracy. These are given in [Rife & Boorstyn, 1974]. For frequency estimation with unknown phase and amplitude, the case most relevant to multiple wavelength interferometry, one obtains:

$$\begin{aligned} \mathrm{Var}(\omega) \geq & \frac{\sigma^2 A^2 N^2}{A^4 N^2 \sum_{n=0}^{N-1} {t_n}^2 - A^4 N \left(\sum_{n=0}^{N-1} t_n \right)^2} \\ = & \frac{\sigma^2}{A^2} \cdot \frac{1}{\sum_{n=0}^{N-1} {t_n}^2 - \left(\sum_{n=0}^{N-1} t_n \right)^2 / N} \end{aligned} \tag{3.12}$$

For uniform sampling, centered around zero, this leads to a simple and well known bound. Inserting

$$t_n = (n + n_0)T, \text{ with } n_0 = -\frac{N-1}{2}, \; n = 0 \ldots N-1 \tag{3.13}$$

in equation 3.12 yields:

$$\begin{aligned}
\mathrm{Var}(\omega) &\geq \frac{\sigma^2}{A^2T^2} \cdot \frac{1}{\sum_{n=0}^{N-1}(n+n_0)^2 - \frac{1}{N}\left(\sum_{n=0}^{N-1}(n+n_0)\right)^2} \\
&= \frac{\sigma^2}{A^2T^2} \cdot \frac{1}{\sum_{n=0}^{N-1} n^2 + 2n_0\sum_{n=0}^{N-1} n + Nn_0{}^2 - \frac{1}{N}\left(\sum_{n=0}^{N-1} n\right)^2 - 2n_0\sum_{n=0}^{N-1} n - Nn_0{}^2} \\
&= \frac{\sigma^2}{A^2T^2} \cdot \frac{1}{\frac{N(N-1)(2N-1)}{6} - \frac{N(N-1)}{2}\cdot\frac{N(N-1)}{2}\cdot\frac{1}{N}} \\
&= \frac{\sigma^2}{A^2T^2} \cdot \frac{12}{N(N-1)(2(2N-1)-3(N-1))} \\
&= \frac{\sigma^2}{A^2T^2} \cdot \frac{12}{N(N^2-1)}
\end{aligned} \tag{3.14}$$

If the signal model is a real-valued sinusoid, the solution gets more complicated. Therefore some approximations are derived as well.

Inserting equation 3.21 into equation 1.27 and assuming independent but not necessarily identical noise yields the following Fisher information matrix:

$$I = \begin{bmatrix}
A^2\sum_{n=0}^{N-1}\frac{{t_n}^2}{{\sigma_n}^2}\,\mathrm{s}^2 & -A^2\sum_{n=0}^{N-1}\frac{t_n}{{\sigma_n}^2}\,\mathrm{s}^2 & -A\sum_{n=0}^{N-1}\frac{t_n}{{\sigma_n}^2}\,\mathrm{s\,c} & -A\sum_{n=0}^{N-1}\frac{t_n}{{\sigma_n}^2}\,\mathrm{s} \\
-A^2\sum_{n=0}^{N-1}\frac{t_n}{{\sigma_n}^2}\,\mathrm{s}^2 & A^2\sum_{n=0}^{N-1}\frac{1}{{\sigma_n}^2}\,\mathrm{s}^2 & A\sum_{n=0}^{N-1}\frac{1}{{\sigma_n}^2}\,\mathrm{s\,c} & A\sum_{n=0}^{N-1}\frac{1}{{\sigma_n}^2}\,\mathrm{s} \\
-A\sum_{n=0}^{N-1}\frac{t_n}{{\sigma_n}^2}\,\mathrm{s\,c} & A\sum_{n=0}^{N-1}\frac{1}{{\sigma_n}^2}\,\mathrm{s\,c} & \sum_{n=0}^{N-1}\frac{1}{{\sigma_n}^2}\,\mathrm{c}^2 & \sum_{n=0}^{N-1}\frac{1}{{\sigma_n}^2}\,\mathrm{c} \\
-A\sum_{n=0}^{N-1}\frac{t_n}{{\sigma_n}^2}\,\mathrm{s} & A\sum_{n=0}^{N-1}\frac{1}{{\sigma_n}^2}\,\mathrm{s} & \sum_{n=0}^{N-1}\frac{1}{{\sigma_n}^2}\,\mathrm{c} & \sum_{n=0}^{N-1}\frac{1}{{\sigma_n}^2}
\end{bmatrix} \tag{3.15}$$

$$\mathrm{s} = \sin(\omega t_n - \phi) \tag{3.16}$$

$$\mathrm{c} = \cos(\omega t_n - \phi) \tag{3.17}$$

Inverting this matrix yields the Cramér-Rao bounds for estimation of the parameters (given by the diagonal elements). This looks rather ugly and is therefore not shown explicitly here, but there are two important results:

- The CRB for frequency estimation is proportional to $\frac{1}{A^2}$. The CRB for frequency estimation is also proportional to the noise level (scaling all σ_n with the same scaling factor c yields a change in CRB by c^2). For uniform noise σ, the CRB is directly proportional to $\left(\frac{\sigma}{A}\right)^2$, and all other remaining terms only describe the position of the sampling points relative to the signal. $\frac{A}{\sigma}$ will therefore be called signal-to-noise ratio in the following.

- The value of C has no influence on the result at all; it only matters whether there is an unknown offset present or not.

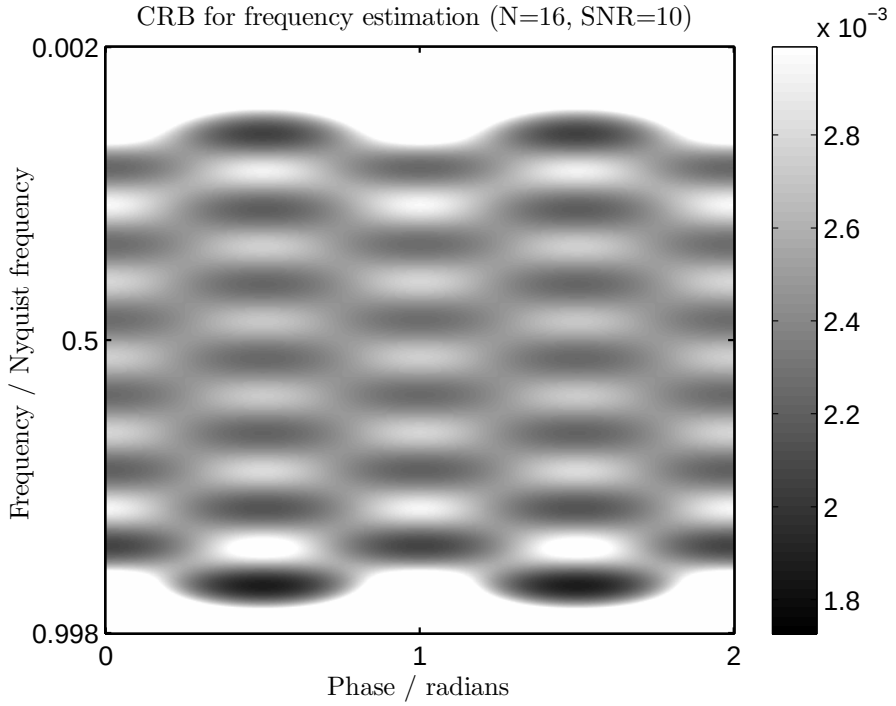


Figure 3.5.: Theoretical lower limit on the relative standard deviation of a frequency estimate for a single noisy tone (equation 1) using 16 equispaced samples and a SNR of 10, based on a numerical evaluation of the CRB. The plot underlines the fact that estimation becomes unreliable for frequencies close to zero or the Nyquist frequency, where the scale has been truncated.

If the signal is sampled uniformly and if the noise is independent and identically distributed, some simplifications are possible. These are not the main focus of this thesis, therefore only one especially important aspect is shown next. If there are no nuisance parameters present (i.e. ϕ, A, C known), the first diagonal element can be inverted directly. Together with the assumptions on sampling and noise this yields

$$\begin{aligned} I_{11} &= \frac{A^2}{\sigma^2} \sum_{n=0}^{N-1} {t_n}^2 \sin^2(\omega t_n - \phi) \\ &= \frac{A^2}{2\sigma^2} \left(\sum_{n=0}^{N-1} {t_n}^2 - \sum_{n=0}^{N-1} {t_n}^2 \cos(2\omega t_n - 2\phi) \right) \end{aligned} \tag{3.18}$$

Numerical evaluation yields Figure 3.5, showing the relative accuracy of the parameters for uniform sampling (16 samples, symmetric around 0) and $\sigma = 0.1A$. A and C are set to 1 (their values have no influence on the frequency estimation accuracy). The CRB is shown for all frequencies from zero to the Nyquist frequency and all phases.

This looks very different from the results given by [Rife & Boorstyn, 1974], but one can easily show that the results are closely related: The second sum is much smaller than the first one for almost all combinations of sampling points and frequencies,

and the more samples the bigger the difference gets. In the worst case, the second sum is equal to the first one and therefore the Fisher information becomes zero; in the best case the absolute value is identical, but the sign reversed, leading to twice the value. In this case the accuracy for the complex signal model is reached, even though only the real part of the signal was available (see below). On average across all possible signal phases, the second sum is zero; the first sum is the average Fisher information in this sense. A closed form expression can be given for this part, using $t_n = (n + n_0)T$, with $n_0 = -\frac{N-1}{2}$.

$$\begin{aligned} I_{11} &= \frac{A^2}{2\sigma^2} \sum_{n=0}^{N-1} {t_n}^2 \\ &= \frac{A^2T^2}{2\sigma^2} \sum_{n=0}^{N-1} \left(n - \frac{N-1}{2}\right)^2 \\ &= \frac{A^2T^2}{2\sigma^2} \left[\frac{N(N-1)(2N-1)}{6} - 2\frac{N-1}{2}\frac{N(N-1)}{2} + N\left(\frac{N-1}{2}\right)^2\right] \\ &= \frac{A^2T^2}{2\sigma^2} \frac{N(N^2-1)}{12} \end{aligned} \tag{3.19}$$

This can be used as a rough estimate for the CRB:

$$\mathrm{Var}(\omega) \geq I^{-1} \approx \frac{\sigma^2}{A^2T^2} \frac{24}{N(N^2-1)} \tag{3.20}$$

This is exactly twice the value compared to the complex valued case above, which is not surprising as the number of independent noisy measurements available is half that of the complex valued case.

3.4. Optimum Sampling

Fast and accurate frequency estimation for a noisy sinusoid is required not only in optical metrology (not limited to frequency scanning interferometry, there is also e.g. [Vanlanduit et al., 2004]), but is needed in many other applications as well, ranging from acoustic [Christensen & Jensen, 2006] to radar [Teague, 2002] signal processing. In the previous chapter, a lower bound on the theoretical accuracy of frequency estimation has been derived. The resulting accuracy for uniform sampling is well known, but it is an open question how the sampling should look like in order to reach the most accurate results in a frequency scanning interferometry system. In such a system, the chosen laser frequencies correspond to the sampling points the sinusoidal signal for each pixel is sampled at, and the desired height map corresponds to the frequency estimates for each one of these sinusoidal signals.

In most cases, uniform sampling is used as this is the easiest way to acquire data, for example if an electrical signal is recorded with an analog-to-digital-converter. Such systems have a constant sampling rate and acquire the signal for a given time period. Continuous broadband signals must be band-limited by the Nyquist frequency to allow for exact reconstruction from the samples and avoid aliasing [Oppenheim & Schafer,

1989]. In case of single tone frequency estimation, this aliasing is not necessarily a problem: It causes ambiguity, but the frequency estimates are still accurate, and the actual frequency can be determined if prior knowledge is available. The accuracy of such a system largely depends on the sampling time, i.e. on the number of samples and the noise level of these.

In some cases, however, the situation is different:

- For some applications (e.g. anti-aliasing in computer graphics [Cook, 1986]), better (in the case of computer graphics: visually more pleasing) results can be obtained with random sampling instead of uniform sampling. This is not discussed here.
- The sampling operation itself and the signal processing may be expensive, and therefore a low number of samples (but not necessarily uniform or close to each other) might be desirable. For instance, in frequency scanning interferometry, and more generally in all applications where the sinusoidal signal is explicitly sampled by choosing specific sampling points, there is a cost associated with the number of samples rather than with their spacing.

In the latter cases, sampling can be accelerated and costs reduced by carefully choosing the optimum sampling points. An optimum sampling scheme for a limited sampling range and sampling time with an arbitrary distribution of the sampling time across the samples has been introduced in [Wieler et al., 2006]. While the proposed sampling design is optimal given the above constraints, it does not, by itself, suggest an algorithm to efficiently estimate the frequency from the resulting data.

3.4.1. Optimization Criteria

There are three main objectives that define the optimality of the sampling pattern:

- Short measurement time: A low number of samples and a short measurement time per sample are desirable.
- Accurate results: The frequency and phase estimates should be as accurate as possible given the other constraints.
- Short processing time: A fast algorithm for obtaining the frequency estimates must exist.

Additionally, if possible there should be a continuous trade-off between accuracy and measurement time that can be easily adjusted depending on the actual measurement conditions.

The optimization problem is constrained by the properties of the camera (exposure time, frame rate, photon capacity, saturation) and the laser (tuning speed, intensity, bandwidth, and possible grid) as well as by the available hardware for processing. Optimizing these requirements together is a difficult problem as there are both theoretical and algorithmic aspects to consider. Therefore a three-stage approach is chosen: First, the theoretically optimum sampling pattern for highest accuracy at a given measurement effort is determined. Next the sampling pattern is modified such that a fast

algorithm is applicable, while trying to stay close to the theoretical boundary. In the third and last step, a fast algorithm is implemented and its performance is verified on simulated data. This way it is possible to obtain a quantitative measure for the quality of the final algorithm and compare it to the theoretical optimum.

Two different optimum sampling schemes will be discussed:

- The first sampling scheme allows assigning arbitrary weights to a fixed range of uniformly spaced frequencies. The assumption of a fixed range of uniformly spaced frequencies is dictated by the laser diode as described in chapter 3.1.1. Assigning arbitrary weights to samples (and therefore achieving a different SNR for different sampling points) is theoretically possible, but does not take camera constraints into account. The derivation in chapter 3.4.2 summarizes results from [Wieler et al., 2006] and mainly discusses the consequences for frequency scanning interferometry.

- The second sampling scheme additionally introduces the constraint that sampling should be done with uniform weights (i.e. the same SNR for all measurements), as this is much easier to implement in practice. In addition to that, algorithmic constraints are taken into account, leading to a multi-block approach for frequency estimation, cf. chapter 3.5.

Getting back to the signal model described in the previous section, we would like to accurately estimate, from as few samples as possible, the frequency of a noisy single tone

$$\begin{aligned} I(t) =& A \cdot \cos(\omega \cdot t_n + \varphi) + C + \epsilon_n, \\ & t_{min} < t_n < t_{max},\ n = 1, ..., N \end{aligned} \tag{3.21}$$

with amplitude A, offset C, frequency ω and ϵ_n independently and identically distributed Gaussian noise with zero mean and variance σ^2.

The sampling points t_n should be chosen such that the most accurate estimate for ω can be obtained and they need not be spaced equidistantly in time. “Most accurate” is here defined as minimizing the CRB while enforcing a minimum distance to secondary minima (i.e. keeping results unambiguous).

3.4.2. Theoretically Optimum Sampling Pattern

It has been shown in [Oliphant, 2006] that the sampling points near the boundary of the permissible sampling range are most important, and it has been shown in [Wieler et al., 2006] that for real-valued sinusoids — taking ambiguity issues into account — using several sampling points mainly at the borders yields optimum results. In that case, the optimum sampling pattern for an application depends on the assumed signal-to-noise ratio, as this ratio determines the ambiguity threshold that is required to keep the probability of outliers (cases where a secondary minimum is lower than the primary, correct solution) below a given level. Here the treatment is extended to signals with unknown offset.

There are two criteria for the optimum sampling pattern:

First, the (average) Cramér-Rao bound for frequency estimation is a measure of the (average) curvature along the parameter ω of the four-dimensional manifold spanned by the signal space in N-dimensional sampling space, where N is the number of sampling points, and A, C and ϕ are nuisance parameters. When the signal quickly changes with ω, ω can be determined accurately from the sampled noisy signal. Instead of minimizing the average CRB, one could also use other error criteria, i.e. minimize the maximum CRB or minimize the mean squared CRB. Additionally, one could maximize the Fisher information instead of minimizing the CRB.

An equation for the minimization of the CRB reads as follows

$$\begin{aligned} \boldsymbol{\lambda}_{\text{opt}}(\boldsymbol{t};\omega_r) &= \arg\min \int_{\omega_r}^{\omega_{\text{Ny}}-\omega_r} \int_{-\pi}^{\pi} \mathcal{I}^{-1}_{\omega;\phi,A}(\boldsymbol{t},\boldsymbol{\lambda};\omega,\phi)\, d\phi\, d\omega \\ \text{under the constraints} \quad & \begin{cases} \lambda_j \geq 0 & \forall\, j \\ \sum_j \lambda_j = \Lambda & \end{cases}, \end{aligned} \tag{3.22}$$

where $\mathcal{I}^{-1}_{\omega;\phi,A}$ is the first diagonal element of the inverse of the Fisher information matrix as described in 1.5.3. A slightly more general description is found in [Wieler et al., 2006]. The vector of possible sampling positions is denoted by $\boldsymbol{t}$, and the weights associated which every position are denoted by $\boldsymbol{\lambda}$.

In practice, the "relative weights" translate to "sampling effort", where it is assumed that the variance of a measured value is cut in half if the effort at that point is doubled. In the case of frequency scanning interferometry, sampling effort is measured by the acquisition time at a given point. The "total sampling effort" is given by Λ, and each element of the vector $\boldsymbol{\lambda}$ is defined by the noise level of the corresponding sample t_j: $\lambda(t_j) = 1/\sigma_j^2$. The signal amplitude is assumed to be $A = 1$, which is not a restriction as will be shown later. The offset C does not show up in this optimization, as has been discussed previously. According to the equation above, the sampling distribution turns out to be independent of the signal and the noise amplitude, and the "total sampling effort" has no influence on the sampling scheme either (everything is just scaled).

However, that does not yet account for the issue of ambiguity — there might be multiple minima in the error plane that are very close to each other in depth. This causes two problems: First of all, it becomes very hard to find the correct one of these algorithmically, and secondly, in the presence of noise the true minimum might not be lower than the secondary ones, leading to an incorrect result. Such a highly non-linear problem leads to a non-Gaussian error distribution: The small errors that might occur when the position of the minimum is not found exactly due to some noise and the finite curvature approximately follow a Gaussian distribution, but the errors that occur due to being in the wrong local minimum show up as outliers. As long as the noise on the signal is very small, the probability of such an outlier is close to zero, and the optimum sampling pattern can be derived by just looking at the local curvature (which dictates the width of the Gaussian). This leads to a sampling pattern with most of the time spent on the borders of the sampling range.

Looking at this sampling pattern in more detail shows that the secondary minimum for two blocks of samples with M samples each on a range of N samples in total is approximately at the position $\left(\frac{\pi}{N-M}; \frac{3\pi}{N-M}\right)$ for large $N - M$. All other minima

are "less deep" (assuming appropriately chosen A, C and ϕ). Doing a brute force comparison with all values of the signal would be too slow to implement in practice, but checking this secondary minimum is feasible.

The optimum sampling pattern now depends on the acceptable probability of outliers for a given SNR. This is a constraint on the "depth difference" between the primary minimum and secondary minima. The relative difference is given by:

$$\Delta_{min,rel} \geq \frac{8\left(\text{erf}^{-1}\left(1-2P_{false}\right)\right)^2}{\Lambda A^2}. \tag{3.23}$$

Arbitrary values for A are taken into account with the normalization by A^2. For uniformly weighted samples and noise, a more intuitive replacement for ΛA^2 is given by

$$\Lambda A^2 = M \cdot \left(\frac{A}{\sigma}\right)^2, \tag{3.24}$$

with number of samples M, signal amplitude A and additive Gaussian noise with standard deviation σ. P_{false}, A and σ are closely related and always occur in the same combination according to equation 3.23, yielding a one-dimensional field of possible sampling patterns depending on $\Delta_{min,rel}$. Without loss of generality, one can set $\Lambda = 1$ and does not have to take A or C into account in the optimization in 3.22.

Therefore this leads to just one additional constraint: $\Delta_{min,rel}$ must be smaller than the distance between primary and secondary minimum as described above. This condition must always be fulfilled, but only for small $\Delta \leq 0.1$ the simplified check of secondary minima as described above is possible, otherwise a full search for secondary minima might be required.

Two examples are shown in Figure 3.6. The first one uses $\Delta = 0.0068$, the second one uses $\Delta = 0.0151$, corresponding to a sampling effort of 3200 (i.e. 32 samples with SNR 10 each) and $P_{false} = 1\%$ and $P_{false} = 0.025\%$ respectively. The weights in the middle get more and stronger with increasing $\Delta_{min,rel}$, which leads to a reduction in outliers at the cost of a slight increase in CRB.

Figure 3.6 demonstrates that the optimum weight distribution focuses on the sampling points at the borders of the considered range, and few samples with very low weights can be found in-between.

3.5. Near-Optimum Sampling for Multiple Wavelength Interferometry

Now that the theoretically optimum sampling pattern has been determined, a practical approximation to this sampling pattern is needed. Arbitrary weights are hard to implement in practice: For instance, while there may be a theoretical benefit in varying the time dedicated to sampling at different frequencies in frequency scanning interferometry experiments, with a camera it is much easier to use the same camera exposure time for all samples, as close as possible to the limit dictated by the full well capacity of the sensor [Rhodes et al., 2004]. Additionally, optimum sampling might require a brute force approach to find the best frequency estimate. A uniform weight distribution is

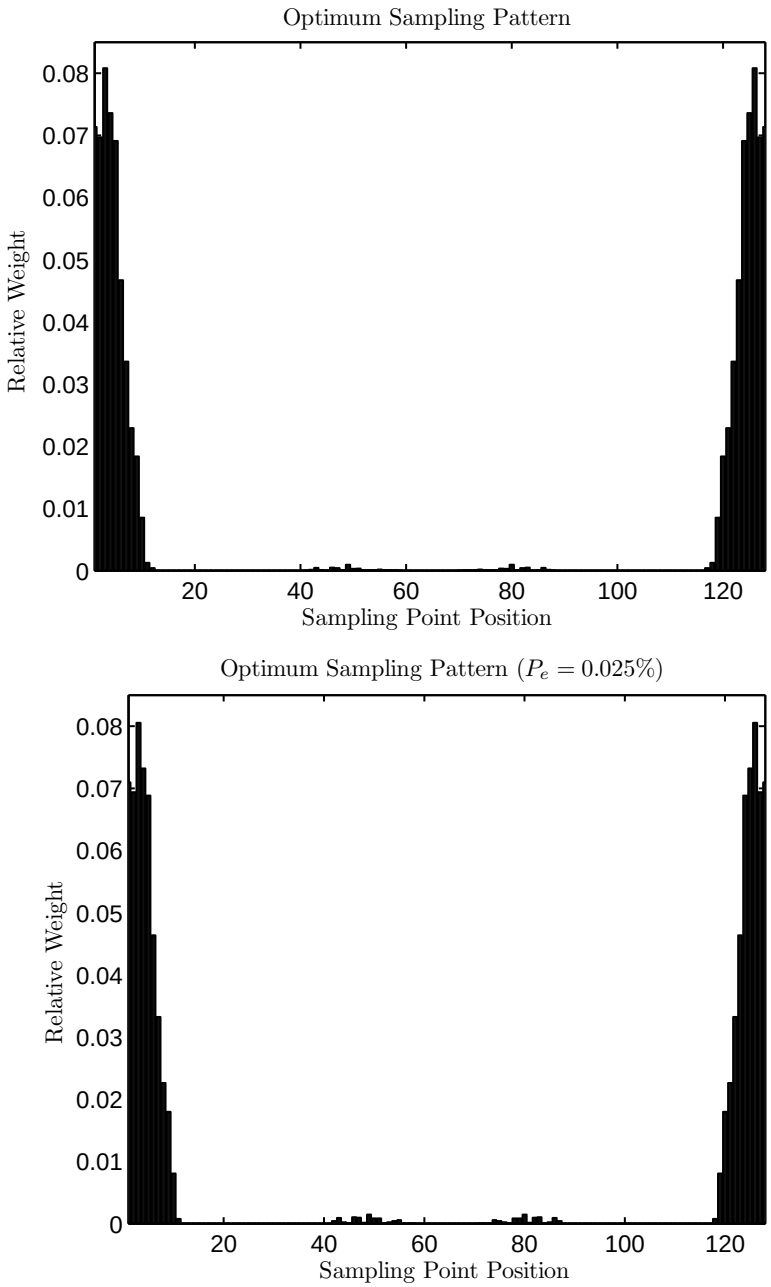


Figure 3.6.: Optimum sampling pattern for frequency estimation for $N = 128$ equidistant samples, probability of outliers $P_e = 1\%$ and $P_e = 0.025\%$. The weights indicate the effort that should be devoted to the acquisition of a measurement at each sampling point. The total sampling effort is equivalent to 32 samples of SNR 10.

far easier to implement, and the algorithm for frequency estimation should not require iterative optimization algorithms. This is true for many applications, and therefore in

the following description the problem is described in general, and only at the end the specific implementation for multiple wavelength interferometry is presented.

An obvious approximation to the pattern in Figure 3.6 is given by simply using two blocks of equally weighted samples at the borders of the range. The impact of the distance and the size of these blocks is analyzed in detail in the next section, and the results are compared to both uniform sampling and the optimum sampling pattern.

3.5.1. Derivation of a Fast Algorithm

For the reasons given above, an algorithm for evaluating experimental data that has been sampled in multiple blocks is derived and its properties are analyzed. For a sampling pattern that consists of multiple blocks of equally spaced and uniformly weighted samples, there is a straightforward procedure: First, the frequency and phase of the signal are determined for each block individually, and then the results are used to initialize a final estimate based on all observations. Fortunately, there is a very simple and highly accurate way of combining information from multiple blocks, as detailed below. The key to the following algorithm is the simple observation that, visually speaking, frequency is the slope of the phase. Considering the signal from eq. 3.21 sampled in two blocks centered at t_1 and t_2, ω and φ can be estimated separately for each block. Then, the following relationship holds as illustrated by Figure 3.7:

$$\varphi_1 + (t_2 - t_1) \cdot \omega = \varphi_2 + 2\pi k,\ k \in \mathbb{N} \tag{3.25}$$

As k is unknown, there is no unique solution for ω.

As a first guess for the frequency, the mean value of the frequency estimates from each of the blocks can be used (strictly speaking, only a frequency estimate from one block is required, but multiple blocks are needed for the phase estimation anyway):

$$\hat{\omega}_{init} = \frac{\hat{\omega}_1 + \ldots + \hat{\omega}_N}{N} \tag{3.26}$$

k is then chosen such that

$$\Delta = \hat{\varphi}_1 - \hat{\omega}_{init} t_1 - (\hat{\varphi}_2 - \hat{\omega}_{init} t_2) - 2\pi k,\ k \in \mathbb{N} \tag{3.27}$$

is minimized:

$$\hat{k}_{opt} = \arg\min_k (\hat{\varphi}_1 + \hat{\omega}_{init} \cdot (t_2 - t_1) - \hat{\varphi}_2 + 2\pi k) \tag{3.28}$$

Ambiguities are resolved correctly as long as the combined error caused by frequency and phase estimation errors as well as unknown sampling jitter does not exceed π.

Next, an improved frequency estimate can be computed. Its accuracy depends on the accuracy of the phase estimation only.

$$\hat{\omega}_{new} = \frac{\hat{\varphi}_2 - \hat{\varphi}_1 + 2\pi \hat{k}_{opt}}{t_2 - t_1} \tag{3.29}$$

The results of the phase estimation define a “ladder” of frequencies that are more or less compatible with the observations; the initial rough frequency estimate is used to

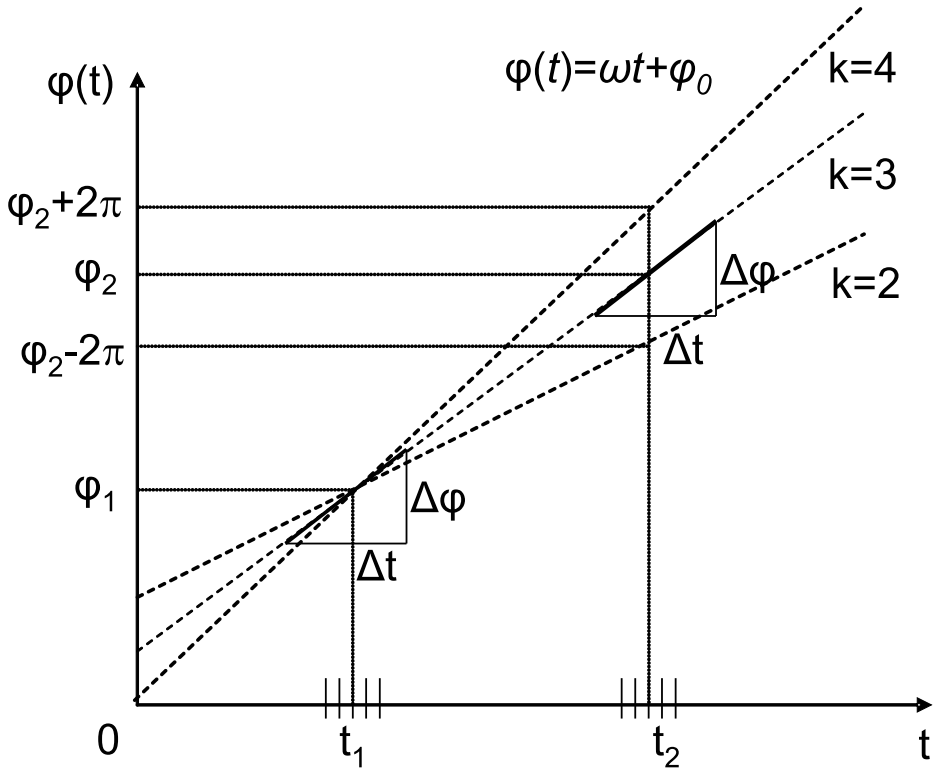


Figure 3.7.: Relationship between phase and frequency: Phase estimates from multiple blocks of data can be combined to obtain a more accurate frequency estimate. The phase values of the blocks define a grid of possible phase slopes (=frequencies), the correct one is chosen based on the frequency estimates from the individual blocks.

find k and thus identify the "right step on this ladder". Provided that k, the number of wavelengths between the sampling blocks, is correctly found, the accuracy of the final result depends only on the accuracy of the phase estimates, the accuracy of the block distance estimate (which might be influenced by sampling jitter) and the absolute distance of the blocks (a larger distance increases accuracy). The accuracy of the initial frequency estimate and the distance between the blocks determine the probability of outliers P_e, i.e. situations in which the estimate k is wrong (a smaller block distance reduces this probability).

Processing is very fast for this algorithm due to two factors:

1. The number of samples $2 \cdot M$ is much lower than N in case of uniform sampling, and all computationally expensive steps can be done per block individually, requiring only a very low number of samples and therefore little memory in every step.

2. The computational complexity is not higher than that of any other frequency estimation algorithm applicable to a low number of samples, i.e. $O(N \log N)$ in case of a typical FFT based implementation. For two blocks with $M < N/2$

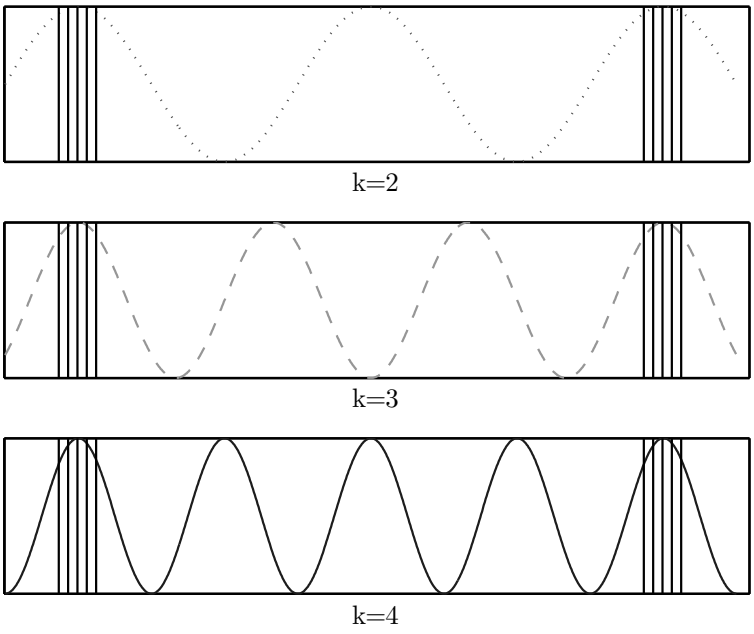


Figure 3.8.: Possible frequencies based on the phase estimates and the frequency estimate from the individual blocks. The signal at the sampling positions is very similar for all three cases depicted here, which can lead to a wrong k being chosen.

samples, the computational effort ($2 \cdot M \log M$) is lower than for the processing of uniformly sampled data ($N \log N$). Combining the results from multiple blocks needs a fixed low effort only.

In case of more than two blocks, the approach above can be applied iteratively, starting out with the two blocks with the smallest distance, and then consecutively choosing pairs of blocks with increasing distance, but using $\hat{\omega}_{new}$ obtained from the previous two blocks instead of $\hat{\omega}_{init}$. This procedure can be repeated until the two blocks with maximum distance are used, and therefore this leads to the same accuracy as if only the blocks with the largest distance were used, but with a lower probability of outliers (incorrect k).

If a certain number of outliers must not be exceeded, there are three ways to reach this goal, with different drawbacks:

- Increasing the number of sampling points per block increases measurement and processing time.
- Increasing the number of blocks also increases measurement and processing time.
- Reducing the distance between the blocks reduces accuracy.

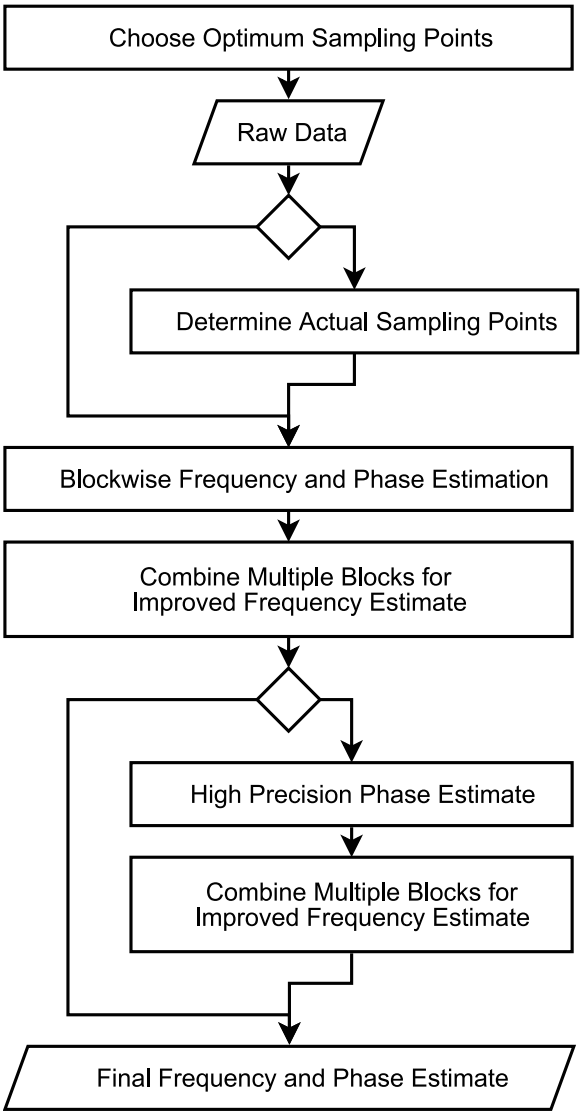


Figure 3.9.: Block diagram of the proposed algorithm for frequency estimation. Once the optimum sampling pattern for the application has been chosen, the raw data is acquired. Optionally the actual sampling positions might be determined for use in the algorithm. Then a frequency and phase estimate for each block of data is obtained, and a new frequency estimate is computed using the algorithm described in this paper. Optionally, a high precision phase estimate can be performed on the basis of the new frequency estimate, and can then be used for a new frequency estimate. Finally, the resulting frequency and phase estimates are returned. There are multiple extensions possible, including amplitude estimation, iterative approaches and using prior knowledge, but these are outside the scope of this paper.

3.5.2. Comparison to the Theoretical Bound

The performance of the proposed algorithm is compared to the theoretical lower bound (CRB) of the variance for both the theoretically optimum sampling pattern and the uniform sampling pattern. Data acquisition time, processing time and accuracy as well as robustness to outliers are discussed. For that purpose, an approximation for the probability of outliers P_e and for the accuracy of the algorithm given in section III is derived.

For the frequency estimation accuracy of the individual blocks an (approximate) lower bound is derived in [Rife & Boorstyn, 1974]. This bound applies to complex signals only, in the real-valued case the bound depends on the true frequency and phase of the signal. In addition, the bound in [Rife & Boorstyn, 1974] does not take an unknown signal offset into account. Asymptotically though, the variance of the real valued case with unknown offset approaches twice the variance of the complex valued case (which is intuitively clear as only half the number of independent measurements are assumed to be available). This is briefly shown in the appendix.

The relative standard deviation is then given by the square root of this approximate variance divided by π,

$$s_{\hat{\omega}} \geq \frac{2\sqrt{6}}{\pi} \cdot \frac{\sigma}{A} \cdot \frac{1}{M\sqrt{M\left(1-\frac{1}{M^2}\right)}} \tag{3.30}$$

Figure 3.10 (top) shows the lower bound on the standard deviation as a function of true frequency and phase when taking the unknown offset and the real-valuedness of the signal model into account.

For the phase estimation from a block of samples with known frequency or for the phase in the center of a block of samples with unknown frequency, using the same approximations as above, one obtains [Rife & Boorstyn, 1974] a relative standard deviation of

$$s_{\hat{\varphi}} \geq \frac{\sqrt{2}}{2\pi} \cdot \frac{\sigma}{A} \cdot \frac{1}{\sqrt{M}} \tag{3.31}$$

Again, eq. 3.31 does not take the unknown offset and the real signal model into account. The CRB can be computed exactly for the phase estimation, with an approach similar to the one for the frequency estimation, see Figure 3.10 (bottom).

Returning to the algorithm described in section III, the probability of outliers (incorrect $\hat{k}_{opt}$ in eq. 3.28) depends on the frequency and phase estimation accuracy as well as on the inter-block distance. In the following, we assume there are two blocks with M uniform samples each, and a total range (from one edge of one block to the other edge of the other block) of N (uniform) samples. Then we can compute an approximate variance for Δ defined in eq. 3.27, assuming independence of the frequency and phase estimates:

$$\mathrm{var}(\Delta) = \sigma^2 \approx 2 \cdot \left(2\pi s_{\hat{\phi}}\right)^2 + \left((N-M) \cdot \pi s_{\hat{\omega}}\right)^2 \tag{3.32}$$

Equation 3.32 is only an approximation, though; the real estimation accuracy depends on the true signal frequency and phase and on the algorithms used for frequency and

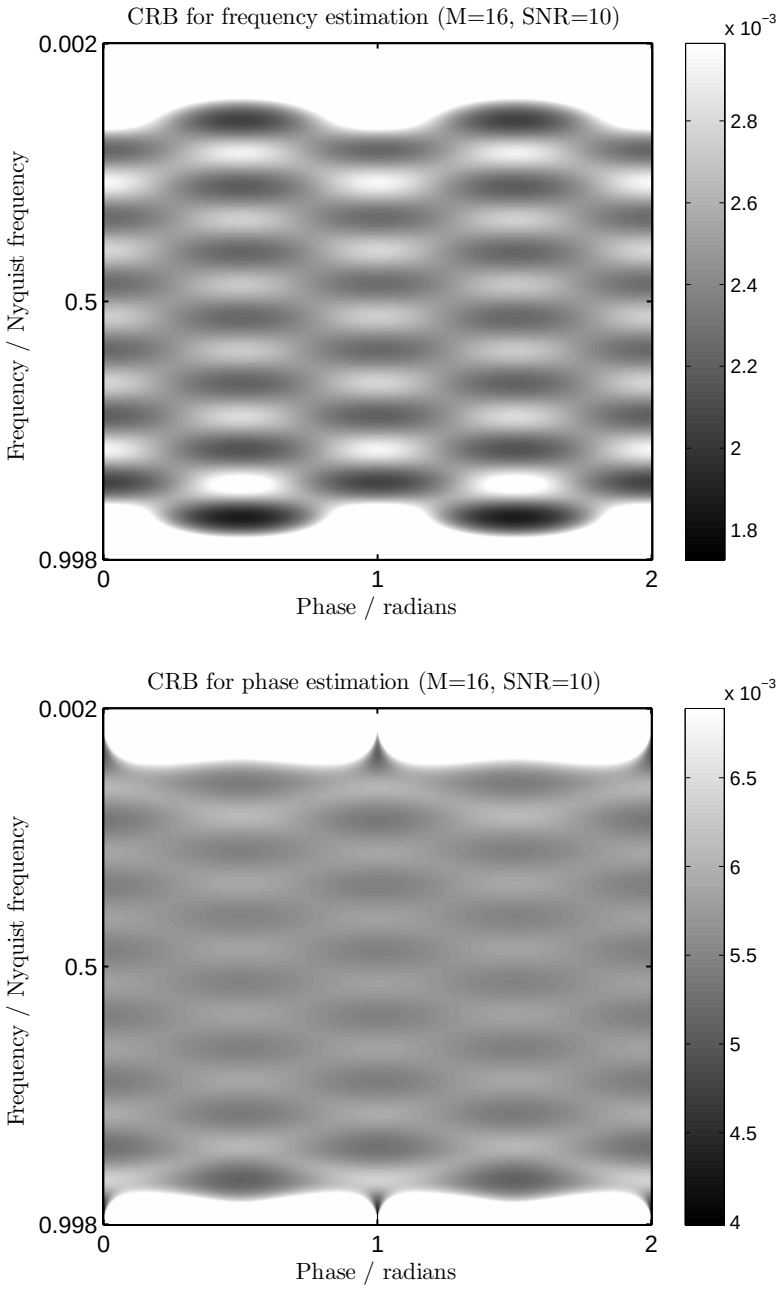


Figure 3.10.: Theoretical lower limit on the relative standard deviation of a frequency (top, previously shown in Figure 3.5) and a phase (bottom) estimate for a single noisy tone (eq. 1) using 16 equispaced samples and a SNR of 10, based on numerical evaluation of the CRB. The plot underlines the fact that estimation becomes unreliable for frequencies close to zero or the Nyquist frequency, where the color scale has been truncated.

phase estimation. In the optimum case (if the true phase and frequency lead to a minimum in the variance of the phase and frequency estimates), the variance is roughly half the one given above; in the worst case (i.e. when the signal frequency is close to the Nyquist frequency) it can be infinite. In addition, the phase and frequency estimates are not independent of each other, hence the above approach that assumes uncorrelated data is not exact. For frequencies far from zero or the Nyquist frequency (cf. Figure 3.10, the exact range depends on the SNR and the number of samples per block) the estimate above is good enough to show some general relations. If one assumes that the distribution of the parameter estimates is approximately Gaussian (which is a good approximation in case of low noise; for high noise the algorithm is not applicable as then a combined analysis of all sampled blocks instead of an analysis of the individual blocks is much better), one can easily compute the probability of outliers P_e: For Δ, a Gaussian distribution with zero mean and variance according to eq. 3.32 can be assumed. If the absolute value of Δ is larger than π, the phase coupling procedure fails. Thus the probability of outliers is approximated by

$$P_e = P(|\Delta| > \pi) = \operatorname{erfc}\left(\frac{\pi}{\sqrt{2}\sigma}\right) \tag{3.33}$$

This does not take into account outliers that are caused directly by the M-point frequency estimation, but if the SNR is high enough for the coupling of blocks to work, the probability of outliers occurring in the M-point frequency estimation step is negligible.

The standard deviation as given above is directly proportional to the noise level. For a given number of samples per block and a given SNR, one can compute the maximum (and therefore optimum) block distance N for a previously specified probability of outliers P_e as demonstrated in Figure 3.11.

This strategy is easy to implement even if analytical treatment becomes difficult in a practical application: One can simply implement the algorithm and look at a histogram of the phase differences Δ across all pixels in the image. It is then obvious when the algorithm fails (i.e. if the distribution becomes too broad) and very simple to adjust the parameters block size M and block distance $N - M$ empirically such that the desired performance and error probability for a given problem is reached. This strategy can therefore be applied even when the noise is correlated, multiplicative or a simple closed form solution does not exist for other reasons.

The accuracy of the result (disregarding outliers) is given by the accuracy of the phase estimate and block distance only.

$$s_{\omega_{new}} = \sqrt{2} \cdot \frac{2\pi s_\phi}{\pi(N-M)} = \frac{2}{\pi} \cdot \frac{\sigma}{A} \cdot \frac{1}{(N-M)\cdot\sqrt{M}} \tag{3.34}$$

A lower bound for the error based on the CRB can be computed as shown in Figure 3.12.

For the values used in Figure 3.12, the root mean squared value of the theoretical limit based on the CRB for the relative accuracy of the frequency estimation in the center frequency range from 0.125 to 0.875 is $1.39 \cdot 10^{-4}$. A numerical estimation on simulated data (using a linear least squares estimator for the phase) yields a standard deviation of approximately $1.40 \cdot 10^{-4}$, which shows that this accuracy can be reached

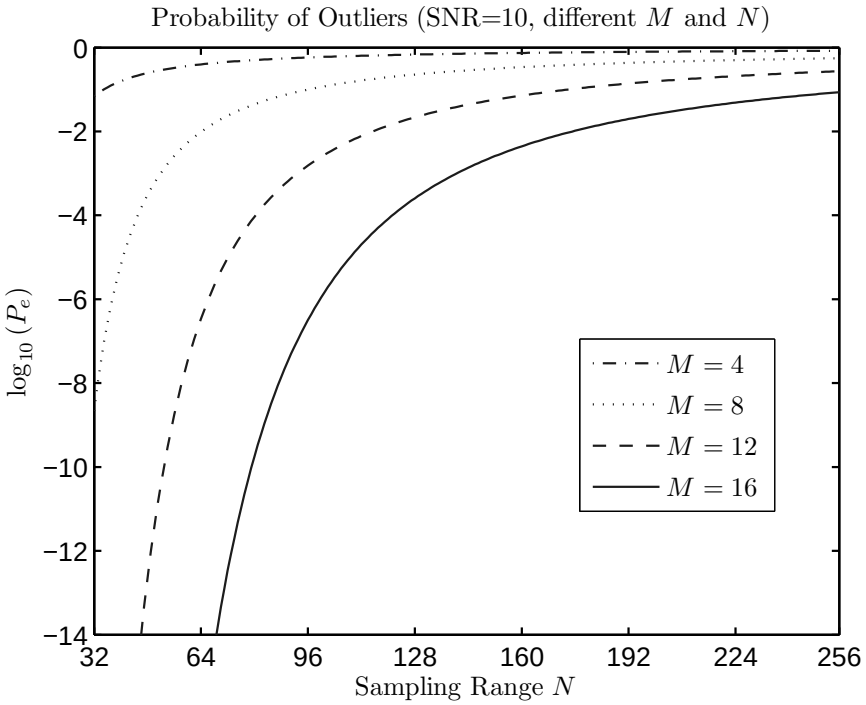


Figure 3.11.: Probability of outliers P_e as a function of sampling range N and block size M, logarithmic scale.

in practice. Both of these values are very close to the approximation in equation 3.34, which yields $1.42 \cdot 10^{-4}$.

Going back to the theoretically optimum sampling pattern as described in [Wieler et al., 2006], the following results are obtained: For $P_e = 2.5 \cdot 10^{-4}$ (approximately the same theoretical probability of outliers as in the case of two blocks with 16 samples each and a total range of 128 samples, according to Figure 3.11), a theoretical relative accuracy of $1.35 \cdot 10^{-4}$ is reached. The corresponding sampling pattern is shown in Figure 3.6 (top). This comparison might be unfair as we have not shown that there is an algorithm that can actually reach such a low probability of outliers, but the probability of outliers of very simple implementations can easily be shown to be far below 1%. With $P_e = 1\%$, as assumed for Figure 3.6 (bottom), the theoretical accuracy improves only slightly to about $1.34 \cdot 10^{-4}$.

In contrast, using the same number of samples ($2M = 32$) distributed uniformly across the measurement range N, the relative standard deviation is $2.15 \cdot 10^{-4}$ (in this case excluding the values at the border frequencies relative to the new Nyquist frequency, otherwise the results would be even worse).

This shows that the accuracy of the procedure described here is very close to the theoretical limit (to about 3% in this case): Even if arbitrary sampling weights are allowed, a significantly better frequency estimation is not possible as long as the constraints on sampling range and sampling effort are kept.

A more systematic comparison of various possible sampling strategies yields the

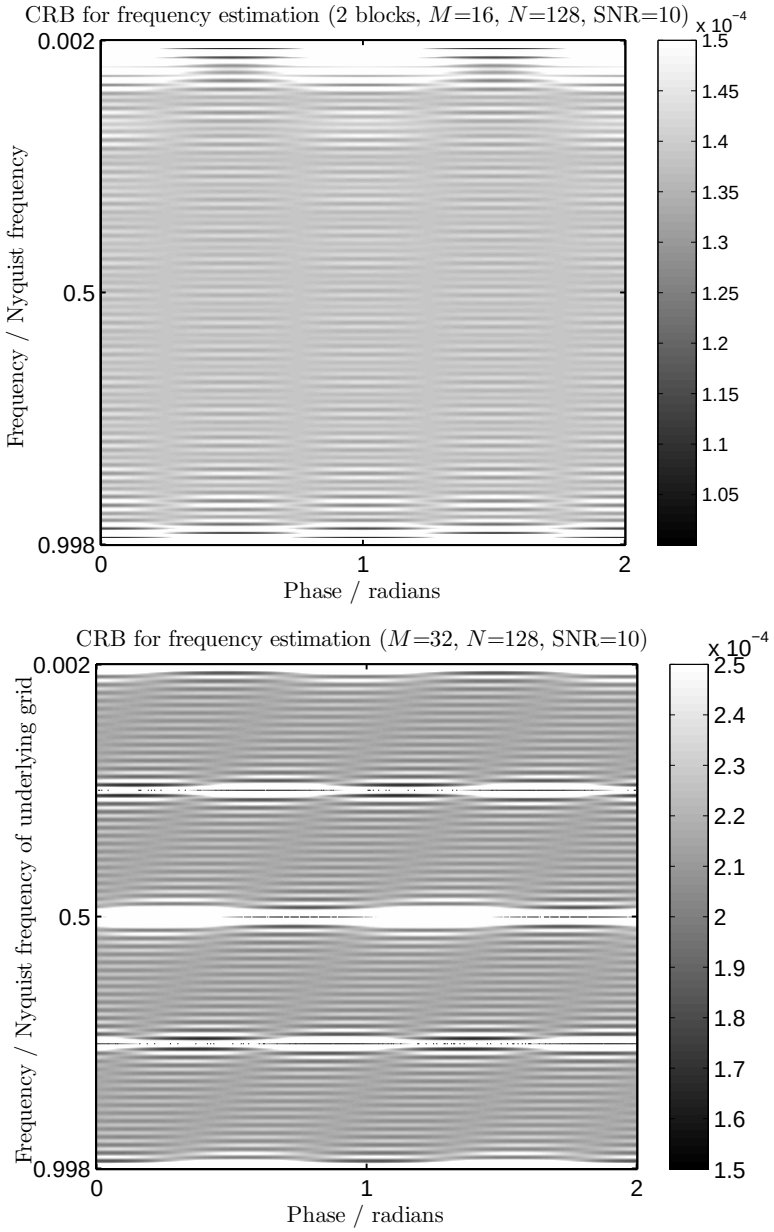


Figure 3.12.: Lower bound on the standard deviation of the frequency estimate from two blocks with 16 samples each from a range of 128 samples (top); for comparison the same bound is shown for the same total number of samples (32), but now uniformly spaced over the same range of 128 samples (bottom). The areas of zero and Nyquist frequency are clearly visible; and even in between the accuracy is more than 30% lower. The scales for both graphs are different, the scale is truncated in the white areas.

results shown in Figure 3.13. The following four cases are compared:

1. Uniform sampling with a fixed number of samples $2M$ over a fixed range of samples M:
$$t_i = i,\ 1 \leq i \leq 2M$$

2. Uniform sampling with a number of samples N, same sampling distance as above, N increasing. This would require significantly longer measurements:
$$t_i = i,\ 1 \leq i \leq N \geq 2M$$

3. Uniform sampling with $2M$ samples, but increasing distance of the samples such that the total range is identical to the case with N samples.[1] In practice this would cause ambiguity issues as the Nyquist frequency decreases.
$$t_i = \frac{i}{M} \cdot N,\ 1 \leq i \leq M$$

4. Sampling in two blocks with M samples each, the block distance increasing with N such that the total range is N.
$$t_i = \begin{cases} i, \text{ for } 1 \leq i \leq M \\ N - M + i, \text{ for } M + 1 \leq i \leq 2M \end{cases}$$
This is the main strategy proposed in this paper.

As the results are proportional to the noise level for sufficiently small noise, an SNR of 10:1 was chosen with little loss of generality. The results show that the proposed algorithm has a very good theoretical accuracy if there is a sufficiently large block distance (as long as the upper limit on the inter-block distance dictated by the acceptable probability of outliers is not exceeded). Performance is necessarily worse than using $N >> 2M$ samples. A more detailed comparison between strategy 2) and 4) is offered in Figure 3.14.

Sampling with 2×16 samples instead of 1×64 samples decreases measurement time by 50% and processing time (if an FFT based algorithm is used in both cases) by 75%, at the cost of a reduction in accuracy of less than 10%. Even at a quarter of the sampling time, the relative standard deviation increases by only 32% instead of the 100% one would expect from looking at the noise.

3.5.3. Extensions

Known sampling jitter

Phase estimation still works very well even if the samples are not equally spaced. If the jitter is large, it calls for more sophisticated algorithms for frequency estimation. If the jitter is not too large, it can simply be ignored in the frequency estimation step: For moderate block distances a sub-optimal frequency estimation does not lead to many outliers, and therefore the final result is still close to the theoretical limit.

[1] In addition, we investigated the case of N uniform samples, with lower measurement effort for each, such that the sum of relative weights amounts to $2M < N$. However the results are so similar to case 3) that they have been omitted in the graph of Figure 3.13.

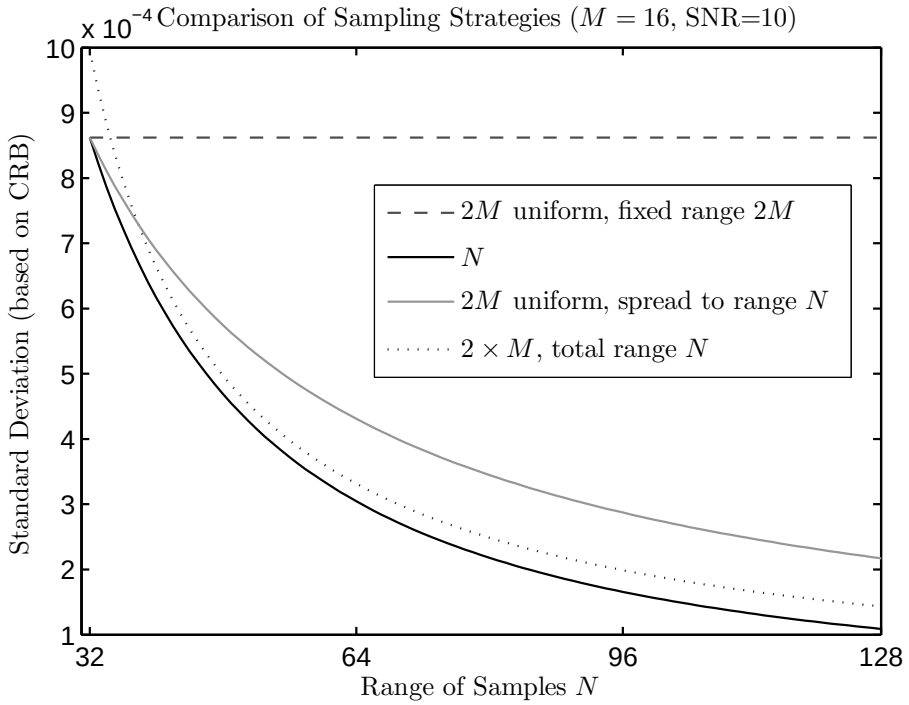


Figure 3.13.: Comparison of different sampling strategies: Three different ways for uniform sampling (fixed number of samples and distance; fixed number of samples and increasing distance; increasing number of samples) and sampling in two blocks with increasing distance are compared. The sampling strategy depicted with a black line uses $32, \ldots, 128$ samples, whereas all other strategies require 32 samples only. For a range of $N = 128$ and $2M = 32$ samples, the two-block strategy proposed here has a standard deviation which is about 34% lower than that of uniform sampling with the same number of samples and measurement range, at a slightly reduced computational cost and with a larger unambiguous range.

Multiple blocks of data

The algorithm can easily be extended to more than two blocks of data. This can be done by using two blocks of data at a time, starting with the blocks with the smallest distance, and then looking at increasing block distances. In this case the probability of outliers decreases, and the accuracy is determined by the largest available distance. Alternatively, a (weighted) least squares estimate could be obtained from all phase values simultaneously (which is especially relevant if for some reason the outer blocks do not offer good signal quality), but this is only applicable if the maximum distance of the blocks is small enough to avoid outliers.

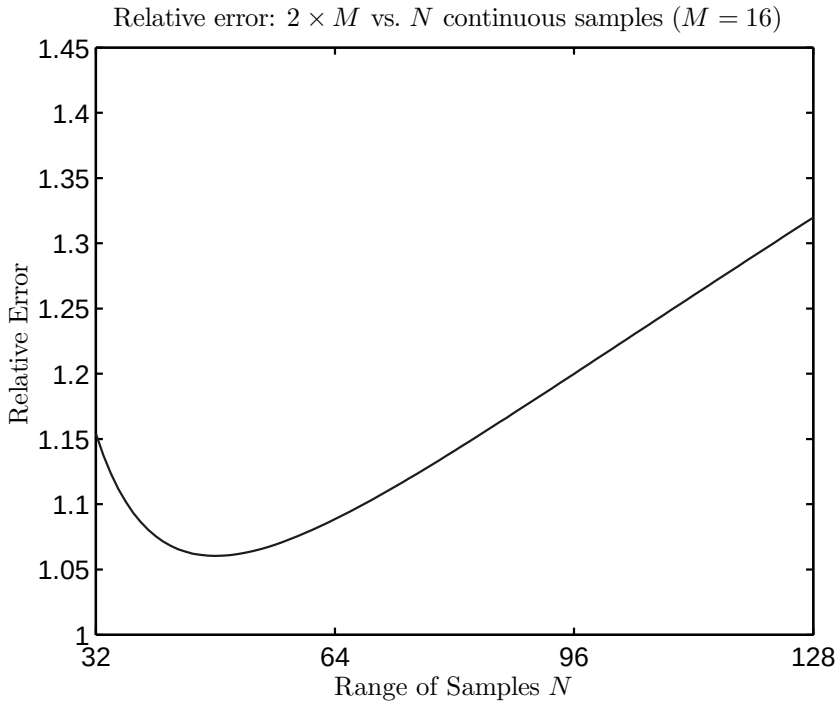


Figure 3.14.: Relative error of dual block method ($2M = 32$ samples in total, but increasing block distance $N - 2M$) compared to single block method (increasing range and number of samples N and therefore increasing total sampling time).

Tracking frequencies

The proposed algorithm can be modified for tracking quickly changing frequencies: In that case one can use the frequency and phase estimate from short blocks, and adjust the block distance used for the final frequency estimate depending on the input data. This can be done for example using a Kalman filter or a simple heuristic approach that increases the distance when the correction based on the phase estimate is small, and decreases it when the correction is larger. This is useful if the tracking is based on blockwise information, e.g. blocks of pilot symbols embedded in a signal [Noels et al., 2005].

Using prior knowledge on the signal phase

If prior knowledge on the signal phase is available for some t, a method for a more accurate frequency estimation can be derived. This can be applied in optical metrology, for example. For a smooth and continuous surface, the result of using this approach in a multiple wavelength interferometry system is identical to that obtained with spatial unwrapping, at a much lower computational cost. In addition to that, the phase estimation can also be used to obtain highly accurate measurement results for surfaces that are not continuous as long as the surface properties and therefore the signal phase

ϕ_0 remain the same.

For this approach, first the frequency ω is determined as accurately as possible with the algorithm described above. Next a single phase value is considered. This phase value can be obtained from only one block or from the whole set of samples. Once again the signal is given by eq. 3.21. Then in this special case we use our prior knowledge on the phase ϕ_0 and obtain

$$2\pi k = \varphi_0 - (\omega \cdot t + \varphi),\ k \in \mathbb{N}. \tag{3.35}$$

Again, there is no unique solution and therefore k is chosen such that the difference is minimized:

$$\hat{k}_{opt} = \arg\min_k(\hat{\varphi}_0 - (\hat{\omega} \cdot t + \hat{\varphi}) + 2\pi k) \tag{3.36}$$

Then an improved frequency estimate can be computed (again, the accuracy depends only on the phase estimate):

$$\hat{w}_{new} = \frac{\hat{\varphi} - \hat{\varphi}_0 + 2\pi\hat{k}_{opt}}{t} \tag{3.37}$$

This is very similar to the derivation above, but t is usually much larger than $t_2 - t_1$, and therefore the "ladder" of frequencies is very fine and the results are more accurate; but the probability of outliers increases.

If there are outliers, it is usually difficult to change the distance of the blocks (in case of optical metrology this distance is given by the laser frequency). This issue can only be resolved with a sufficiently accurate initial frequency estimate, which can be obtained by the algorithm described earlier, or by using prior knowledge on spatial relationships (e.g. smoothness constraints) to correct incorrect choices of k.

3.5.4. Summary and Conclusion

An efficient sampling scheme and algorithm for single tone frequency estimation has been presented. An implementation of the proposed algorithm consists of the following steps:

- Choice of sampling points (e.g. using two blocks of samples with maximum feasible distance for a desired probability of outliers)
- Frequency estimation for one or more blocks (section 3.6)
- Phase estimation for each block (section 3.6)
- Determination of the actual sampling points (section 3.7, optional, if available one can take sampling jitter into account)
- Improved frequency estimate by phase coupling
- Absolute phase estimation using prior knowledge (optional, if knowledge is available)

For a practical implementation, the block size and block distance have to be adjusted to reach the desired accuracy and probability of outliers of the frequency estimation.

The proposed algorithm has three key advantages:

- First of all, it is very fast. Processing time depends mainly on the algorithms used for phase and frequency estimation for the individual blocks, the rest of the algorithm takes much less than a second on a current PC. If fast approaches are used for phase and frequency estimation, a total processing time of less than 10s for 1 million frequency estimates using $M = 32$ frames can be achieved in Matlab on an Intel Core 2 Duo E6600 processor. Processing with dedicated hardware or more optimized software is expected to be significantly faster. Computational complexity is usually lower than that of uniform sampling, given the same number of sampling points. In particular, the algorithm is faster than taking the FFT on a single block of M uniformly sampled data points.

- The algorithm is also highly accurate: The performance of this algorithm by far exceeds that of uniform sampling with the same number of samples and gets very close to the theoretically optimum sampling scheme and theoretically best frequency estimation. On the one hand, the sampling pattern is close to the theoretically optimum sampling pattern, and on the other hand, the algorithm almost reaches the CRB (to less than 1% for an SNR better than 1) for this sampling pattern. Altogether, the result is within 3% of the theoretical limit on the accuracy for the theoretically optimum sampling pattern, i.e. any possible improvements are known to be very limited.

- The algorithm is highly flexible: It can easily be extended to take known sampling jitter or multiple sampling blocks into account, without extra computational effort. In addition, one can easily apply the method even if the noise is correlated or unknown by optimizing a histogram of phase differences as computed from eq. 3.27. The proposed method can therefore be generalized to a wide variety of applications.

3.5.5. Application to Multiple Wavelength Interferometry

The algorithm above is particularly well suited for multiple wavelength interferometry. Acquiring the data with laser frequencies on a roughly uniform grid is possible even if the laser exhibits longitudinal mode jumps, and in case of slight deviations from the optimum sampling pattern the actual laser frequencies can be taken into account in the phase estimation such that the accuracy does not decrease significantly. This will be shown in the following sections. A convenient side effect of the data acquisition in two blocks is that the exposure time can be adjusted such that intensity fluctuations of the laser across its bandwidth or fiber coupling issues are compensated by adjusting the exposure time per block.

The block distance is limited by the available laser bandwidth. For an application in a production line, measurement time is usually limited by the line cycle time, and therefore the number of frames is limited as well. A theoretical decision based on the SNR is usually not possible as the properties of the measurement object are often

not known precisely. Therefore ways to choose a near-optimum sampling pattern for multiple wavelength interferometry are discussed here.

For measuring smooth surfaces, a large block distance is desirable for high accuracy. For a given limit on the measurement time (and therefore a given block size), the bandwidth can be reduced (starting from the maximum laser bandwidth) until the number of outliers is acceptable. This can be determined by looking at the distribution of the difference in the phase coupling step, and is trivial to measure. If it works fine with the maximum laser bandwidth, then the number of samples can be reduced without significantly lowering measurement accuracy. For most applications this setting offers the best compromise between accuracy and measurement time: If the measurement time is the limiting factor, a smaller laser bandwidth can be chosen (resulting in less frames required, reduced measurement time and reduced accuracy). If the accuracy is not sufficient, the laser bandwidth should be increased. When this is not possible, increasing the number of frames in the center is not a good option. Repeating the measurement and averaging the results may lead to a larger increase in accuracy.

For measuring rough surfaces, the situation is different. Here the accuracy does not necessarily increase when the block distance increases, as the speckle field becomes decorrelated when the laser frequency changes. The same empirical method to determine the optimum settings can still be used, but it is possible that the accuracy decreases when the bandwidth increases, and the optimum block distance may be lower.

The optimization is complicated by the fact that a certain number of outliers might be acceptable, as they can be corrected by filters (see chapter 3.8), even 20% outliers might not be a problem. The optimum sampling pattern therefore depends on the requirements of the further processing steps.

A theoretical comparison of the Cramér-Rao bound for the optimum sampling pattern with uniform sampling and the proposed near-optimum sampling pattern, using typical values for a practical measurement system, yields a large improvement compared to uniform sampling: With the same ambiguity and measurement time, one can reach about 97% of the accuracy of the true optimum sampling scheme, or a seven-fold increase compared to uniform sampling!

This is still based on idealistic assumptions though — the derivation above assumes optimum frequency and phase estimation, and this is not possible in practice with limited processing time. However, it will be shown that even with sub-optimum signal processing this result can be reached in practice. The key point to note here is that the accuracy of the frequency estimation has no direct influence on the accuracy of the result; it just influences the number of outliers or the maximum block distance for a given number of outliers. For many tunable lasers the distance will be limited by the usable laser bandwidth, so that a sub-optimum algorithm has no detrimental effect. The accuracy of the phase estimation is very important for the result as seen above — but fortunately this is a relatively simple problem. A good implementation that uses knowledge on the actual sampling points to obtain the best possible results is discussed in the next section.

3.6. Frequency Estimation for a Low Number of Uniformly Spaced Samples

In order to use the sampling scheme and algorithm derived in the previous chapter, a large number of frequency and phase estimates from a very low number of samples are needed. For each camera pixel two phase and two frequency estimates are required, and with megapixel camera resolutions this quickly leads to millions of estimates. As the total processing time should be on the order of a few seconds, this requires very fast algorithms. This problem is discussed in this chapter in a more general context as the algorithms are not limited to multiple wavelength interferometry.

3.6.1. Introduction

The problem of estimating frequency from a sampled signal arises in many applications. There is a large number of different aspects to this problem: Signals may be stationary or time-varying, the signal may have only a single or multiple frequency components, it may be uniformly or non-uniformly sampled. It is therefore not surprising that a large number of techniques has been developed. Methods include autocorrelation based signal subspace techniques (Pisarenko's method [Pisarenko, 1973], MUSIC [Stoica & Soderstrom, 1991], ESPRIT [Lemma et al., 2003]), non-linear optimization techniques (iterative, in both time [Brown & Mao Wang, 2002] and frequency domain [Aboutanios & Mulgrew, 2005]), filter based techniques [Savaresi et al., 2003] and (windowed) FFT based approaches [Jain et al., 1979; Rife & Boorstyn, 1974; Rife & Vincent, 1970; Quinn, 1994]. Each of these readily available algorithms has advantages and disadvantages, and their (asymptotic) properties have been discussed in detail in many papers and textbooks, including [Quinn & Hannan, 2001; Moon & Stirling, 2000; Poor, 1994]. Some of them have been designed for single tone parameter estimation (e.g. the maximizer of the periodogram), others for multi tone parameter estimation (e.g. MUSIC).

However, without modifications these algorithms do not perform well for very few samples. Depending on the specific application, it may be possible to obtain the desired accuracy for the frequency estimate by taking a sufficient number of samples from the signal. In some cases, increasing the number of samples is not feasible though: There is a trade-off between noise suppression and good tracking of a quickly changing signal, and there are applications where significant cost is associated with sampling, such that a low number of samples is preferable. On the other hand, with a limited number of samples it is often possible to employ relatively complex, iterative approaches to estimate the frequency — but if a large number of frequency estimates is required in a very short period of time, this might also be too expensive.

In the following, the issue of very fast single tone frequency estimation from very few (typically 8–32) samples in the presence of additive noise is discussed; for even shorter signals the computational effort does not play such an important role any more, and for more samples the known algorithms perform quite well already. The following analysis is based on a signal model assuming a real-valued sinusoid with unknown frequency, phase, amplitude and offset; often the offset is not taken into account, but for a low number of samples its influence becomes important, and its inclusion does

not complicate the results much. Exactly uniform sampling is assumed, and the main objective besides accuracy is fast computation. The issues discussed here are an extension to the treatment in [Schoukens et al., 1992], made possible by the limitation to very short windows on the one hand and increases in available processing power on the other hand. The focus and therefore the visual presentation is different though — for a very low number of samples, systematic errors of the methods play a more important role. A more rigorous treatment of the asymptotic properties of interpolators can be found in [Quinn & Hannan, 2001], and it should be noted from the beginning that the new algorithms discussed in this paper are known to have poor asymptotic properties as they only use the absolute value of three Fourier coefficients. Nevertheless, the proposed strategy offers better performance under the given constraints than the methods described in [Quinn & Hannan, 2001].

First, the system model is described in detail and the design objectives are defined. Next, the new algorithm is derived. Then its performance is analyzed and compared to alternative approaches. Estimation of the remaining signal parameters phase, amplitude and offset is discussed, and finally the results are summarized.

3.6.2. Signal Model and Problem Description

Once again, the signal is assumed to be a real-valued noisy sinusoid

$$I(n) = A \cdot \cos(\omega \cdot n \cdot T + \phi) + C + \epsilon_n, \qquad n = 1 \dots N \tag{3.38}$$

with amplitude A, offset C, frequency ω and ϵ_n independently and identically distributed Gaussian noise with zero mean and variance σ^2 (repeated from eq. 3.21). For a continuous periodic signal (or an infinite number of samples), the maximizer of the periodogram is the maximum likelihood estimator. It can be approximated by the power spectral density (PSD), which can in turn be computed quickly by using the FFT. However, for a low number of samples there are several problems with this approach:

- The signal is implicitly windowed with (i.e. multiplied by) a rectangular window. The Fourier transform of that window is a sinc function. A real valued signal can be seen as the sum of two complex exponentials in the time domain or two Dirac pulses in the frequency domain. Convolving those with the sinc function yields two overlapping sinc functions in the frequency domain, and the side lobes of each can shift the maximum of the other. In addition, there is aliasing as the sinc function is not band-limited and the side lobes of the peaks can hence interfere with the main lobes due to aliasing.

- The FFT returns results on a discrete grid of frequencies and interpolation is simple, but either not very good or computationally expensive. There has been extensive treatment on how to interpolate as well as possible [Quinn & Hannan, 2001], however none of these interpolators works very well for a low number of samples. Quadratic or center of gravity interpolation between the Fourier coefficients is commonly used in practical applications, but suffers from systematic errors.

One approach to alleviate the aliasing problem is windowing: It can reduce the side lobes and therefore decreases phase dependent bias in the frequency estimation. Commonly used window shapes include the Hamming and Hanning windows. Unfortunately though, reducing the side lobes increases the width of the main lobe, and this makes the results more sensitive to additive noise.

The issue of correct interpolation in the Fourier domain can be solved by zero-padding before taking the FFT or by explicitly computing the Fourier transform for a number of frequencies close to the probable maximum. For band-limited signals this is the correct interpolation, but it is too slow for the application considered here.

3.6.3. Optimization of an Estimation Algorithm

In many applications a windowed FFT (using a Hamming or Hanning window) and interpolation in the Fourier domain is used if a fast algorithm is needed [Vanlanduit et al., 2004; Jain et al., 1979; Zhang et al., 2001]. Windowing reduces sidelobes and increases the width of the main lobe. The appropriate interpolating function depends on the shape of the main lobe; sometimes a closed-form solution exists. For a continuous signal and a Gaussian window, the main lobe in the frequency domain is also Gaussian, and the appropriate Gaussian interpolation could be implemented by taking the logarithm of the data and subsequent quadratic interpolation. Such an approach does not take aliasing or additive noise into account and is not applicable to a very low number of samples. More extensive treatment of this problem can be found in [Schoukens et al., 1992], [Quinn & Hannan, 2001] and [Rife & Boorstyn, 1974]. Most of the approaches in the literature focus either on optimization of the interpolation for non-windowed data, or on derivation of an optimum window. The new approach now combines these two optimization problems, and attempts to find an optimum solution for the frequency estimation problem. As it is impossible to consider all possible interpolation functions in an automated optimization procedure, a fixed set of simple interpolation functions has been selected and a look-up-table based bias correction has been added as a final step. A detailed description of the algorithm follows below. The "best" solution within the given set of constraints (basic structure, computational and memory constraints) is then compared to other algorithms and the theoretical optimum.

The modified algorithm is structured into five steps:

1. The signal offset is removed first. As this offset is not known, the easiest solution is to remove the sample mean from the data. This is not exact as a sampled sinusoid has a small, but non-zero mean for most sampling patterns. In addition to that, the discrete Fourier transform is only exact if there is an integer number of wavelengths on the uniformly sampled support. These two aspects are the reason that evaluating this signal with a Fourier transform introduces phase dependent errors due to the implicit rectangular windowing in the time domain. These errors are minimized in the following optimization, while at the same time keeping noise sensitivity low.

2. In order to reduce these phase dependent errors, the signal is multiplied with a more suitable window function in the time domain instead of the implicit rect-

angular window. The window shape is subject to optimization: Qualitatively, it has to reduce the side lobes of a truncated sinusoid considered in the frequency domain while keeping the width of the main peak as narrow as possible.

3. The FFT is taken, and a first estimate of the frequency based on the DFT coefficients is computed. As an initial step, the squared absolute value of the Fourier coefficients is computed and the position of the maximum is determined, as this is the area where most of the signal information can be found. Using only this data is a first step to reduce the dimensionality of the estimation problem, at the cost of a slightly reduced accuracy.

4. A more refined estimate of the true frequency is determined based on interpolation between the maximum Fourier coefficient and its two neighbors. For a fast computation, it is desirable to use only the Fourier coefficients and/or their squared absolute values as input to a simple interpolating function. This function is also subject to optimization.

5. Finally, a non-linear transform is applied to the end result in order to remove any remaining bias. When both interpolator and window have been chosen, a one-dimensional non-linear correction is defined by the frequency bias (averaged across all phase values) of the estimate (for a given signal-to-noise ratio). This step could be integrated into the interpolator: Using an arbitrary nonlinear transform as an interpolator is slightly more flexible and might offer slightly better performance, but it yields an optimization problem that is very hard to handle. The non-linear transform can be multidimensional (an extreme example would be a look-up table with N-dimensions for N samples); but more than one dimension is not desirable due to the memory requirements. Nevertheless, a two-dimensional example (using phase and frequency of the estimate) will be included for comparison.

The optimization of window and interpolating function is based on prior knowledge of the system and on the desired properties of the frequency estimate. Optimization parameters include:

- known frequency range and weighting of different frequencies,
- signal to noise ratio,
- desired error norm,
- sampling positions and sampling jitter (optional).

Commonly used windows include the Hamming, Hanning and Kaiser windows; often quadratic or center of gravity interpolation is used.

While steps 1, 3 and 5 are self-explanatory, there are several things to take into account in steps 2 and 4:

When optimizing the window function, the number of degrees of freedom is equal to the number of sampling points. However, the problem is symmetric as long as the

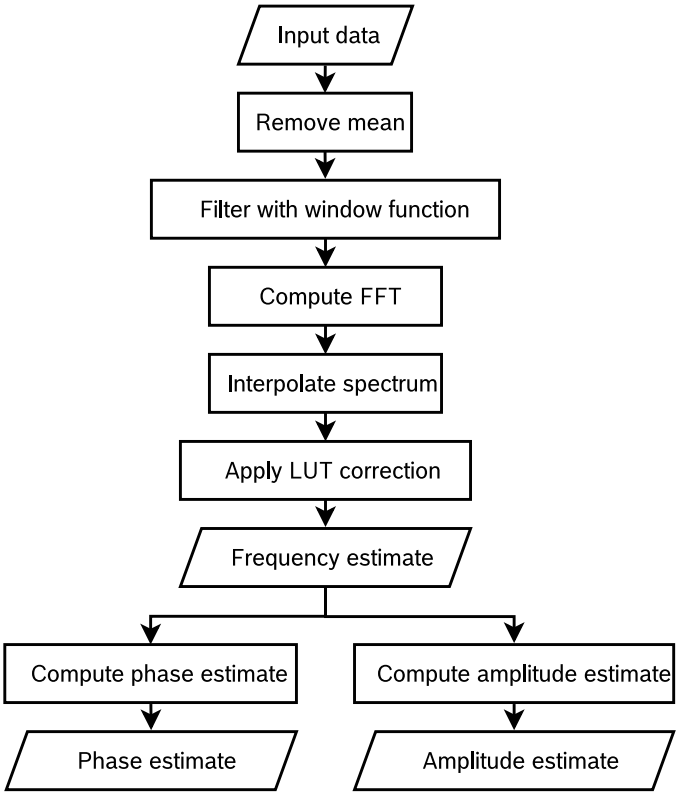


Figure 3.15.: Block diagram: Data flow of the proposed estimation algorithm.

phase is assumed to be unknown and uniformly distributed, which it is in many applications. Simulations have shown that convergence of the optimization algorithm is accelerated a lot when enforcing a symmetry constraint on the window. A further degree of freedom is eliminated by the normalization of the window. This results in a total of $N/2 - 1$ degrees of freedom. An iterative optimization algorithm based on a line search has been implemented, and convergence has been verified by comparing the results of the optimization for random initial values. For faster convergence a number of modifications including adaptive step sizes and initial smoothness constraints have been added as well; these do not influence the final result. The resulting optimum window depends on a number of system parameters and on the optimization criteria as illustrated below.

There are many degrees of freedom for the interpolation; arbitrary functions could be conceived. Known good interpolators are sometimes fairly complex and are based on the complex-valued Fourier coefficients, e.g. Quinn's second interpolator [Quinn & Hannan, 2001]. An analytical discussion of asymptotic properties of some interpolators can be found in [Quinn & Hannan, 2001]. A closed form expression for an optimum interpolator when an arbitrary window size is allowed does not seem feasible, though.

There are four requirements due to the signal model:

1. Interpolation has to be independent of the signal offset. This means that the zero

frequency component of the FFT cannot be used.

2. Interpolation should be as independent as possible of the signal phase; therefore coefficients where the main lobes of the peaks overlap are difficult to use. This can be taken into account in iterative algorithms, but is not desirable for a simple and fast algorithm.

3. Very few Fourier coefficients contribute almost all the information: In case of single tone frequency estimation, most of the signal energy is concentrated in a small range coefficients, and for white noise the noise energy is uniformly distributed across the spectrum. Therefore the signal to noise ratio of most coefficients is very poor, and it is possible to constrain the interpolation to a few coefficients next to the maximum.

4. Interpolation has to be independent of scaling as the signal amplitude is unknown. This limits interpolators to (possibly non-linear) functions of quotients of linear combinations of Fourier coefficients c_i, their real and imaginary parts or any norm. The coefficients can also be taken to the power of k. An arbitrary k can be used, but linear combinations of norms with other elements are limited to even k. l_i, m_i, n_i and $o_{i,p}$ are constants;

Hence we only investigate the following type of function for use in the interpolation:

$$\begin{aligned} f(g_1, g_2, ...), \text{ with } g_k &= \frac{h_{i,k}}{h_{j,k}} \\ h_{i,k} &= \sum_i \Big(l_i \cdot {c_i}^k + m_i \cdot \mathrm{Re}(c_i)^k \\ &\quad + n_i \cdot \mathrm{Im}(c_i)^k \Big) + \left(\sum_{p=1}^{\infty} \sum_i o_{i,p} \cdot |c_i|_p \right)^k \end{aligned} \tag{3.39}$$

This is still a fairly general approach, and only part of this space can be investigated. Using the squared absolute value (required for finding the maximum position) or the real or imaginary part of the Fourier coefficients is most attractive, as these values are available anyway at this point and do not have to be computed. The angle of the coefficients can be used as well; phase information might be useful later and helps if there is only one neighboring coefficient available due to the requirements 1. and 2. given above.

It is not necessary to consider all possible interpolation functions: If the maximum found by interpolation is a strictly monotonous function of the true signal frequency ω at least locally for every interval given by the respective maximum position, all systematic errors can be removed with a nonlinear transformation at the end, which could be implemented by a look-up table.

Using different interpolating functions has an influence on the optimum window shape, therefore the investigation has been performed (results not shown) with a large number of interpolating functions, including center of gravity, quadratic and Gaussian interpolation, applied to arbitrary powers of the absolute value of the Fourier coefficients.

Quadratic interpolation (i.e. fitting a parabola) to the squared absolute value of the Fourier coefficients turned out to perform best in these comparisons (although only slightly better than several others) and has the lowest computational effort. Therefore quadratic interpolation was used for the following performance comparisons. The optimum window shapes shown next depend directly on this interpolating function, results look differently if another interpolating function is chosen. A performance advantage over center of gravity interpolation was consistently present, but on the order of less than 5%.

3.6.4. Simulation Results

For the signal model according to eq. 2 and additive white Gaussian noise, the Cramér-Rao bound (CRB) has been computed: It gives a lower bound on the variance of any unbiased estimator [Poor, 1994]. This bound is applicable to the algorithm discussed above as it is by design almost unbiased [1]. Therefore the mean squared error can be directly compared to the CRB. The CRB is not necessarily a tight bound though, as it does not take threshold effects into account [Bell et al., 1997]. These occur in the highly non-linear frequency estimation problem in the presence of significant amounts of noise.

With the given real-valued signal model, both the CRB and the actual estimator performance depend on the true values of the parameters; in this model, though, the value of the offset C has no influence [2], and the amplitude A is taken into account indirectly by defining the SNR as A/σ. The resulting errors for a given SNR can therefore be plotted in the (ω, ϕ)-plane; and the most significant effects are in the ω direction. In all cases the square root of the variance or the mean squared error respectively are shown in the graphs due to the easy and direct connection to measurement accuracy in the optical metrology application this optimization was performed for.

In Figure 3.10 (top) the color scale is chosen such that the minimum corresponds to the CRB one would obtain for the complex signal model as described in [Rife & Boorstyn, 1974], and the maximum corresponds to three times this variance. In the white areas the CRB exceeds the scale, which shows that good results cannot be obtained for very low or very high frequencies, therefore one has to focus on a defined frequency range. In the following, the frequency range from 0.125 to 0.875 (relative to the Nyquist frequency), an SNR of 10, N=16 and $\phi = 0, \ldots, 2\pi$ are used, and an optimization according to Figure 3.16 is performed.

Figure 3.17 shows the phase averaged mean squared error vs. true frequency. It includes the CRB, the optimum window according to the algorithm described above, the best "standard" window (in this case a Hanning window) and an optimum window with a multi-dimensional look-up-table (using both the amplitude of the coefficients and their phase).

A more complete comparison of windows can be found in table 3.1, characterized by the mean squared error taken across all phases and frequencies. The main

[1] Due to the look-up table used at the end, frequency dependent bias is removed. There may be some phase dependent bias, but due to windowing this is very small.

[2] It is important whether there is an unknown offset or not, but the specific value of the offset is irrelevant.

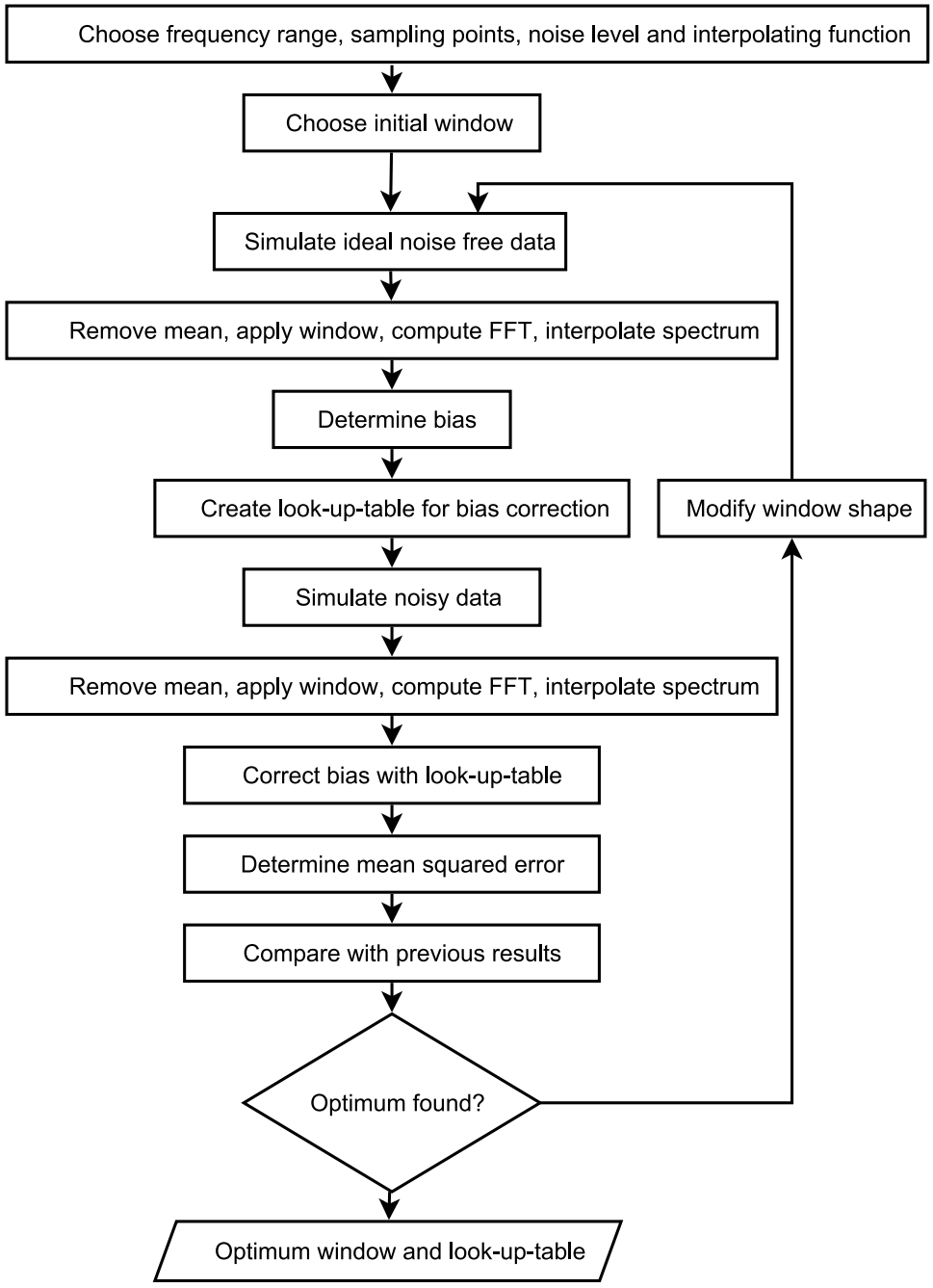


Figure 3.16.: Block diagram: Algorithm for finding the optimum window shape for a given set of criteria.

windows from [Schoukens et al., 1992], as well as several other windows in use in various applications have been investigated, but only a selection of well performing windows is included in the table. For windows characterized by continuous parameters, the parameter value was optimized. A much larger number of interpolators and

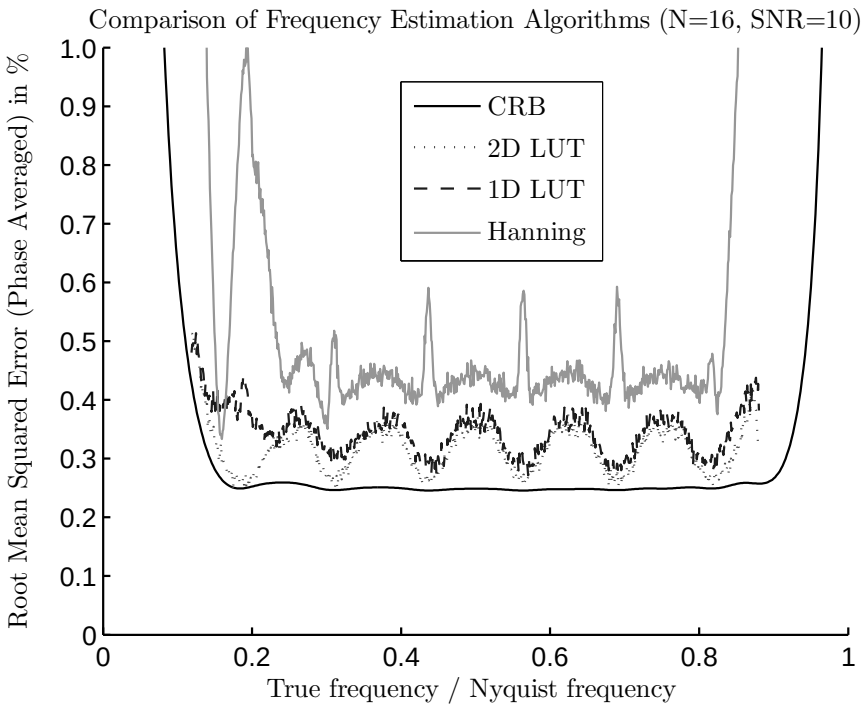


Figure 3.17.: Comparison of different algorithms (Hanning window, optimum window with 1D-LUT, optimum window with 2D-LUT) and the theoretical limit (CRB).

Window Interpolation	Hamming	Bartlett	Hanning	Kaiser	Optimum
Gravity	0.65%	0.64%	0.62%	0.45%	0.43%
Gravity+LUT	0.48%	0.47%	0.45%	0.39%	0.36%
Quadratic+LUT	0.46%	0.46%	0.43%	0.40%	0.35%

Table 3.1.: Relative RMS error of FFT based frequency estimates for various common window and interpolation functions and a look-up-table for bias correction, all using N=16 and SNR=10, frequency range 75%. The results for the Kaiser window were obtained using a 1-D optimization of the parameter β, the optimum window was freely optimized.

windows than shown in table 3.1 has been analyzed. This includes Gaussian interpolation with and without look-up-table bias correction and quadratic interpolation without bias correction. The Hann, Blackman, Blackmann-Harris, Nuttall, Parzen and Bohman windows and a number of windows with adjustable parameters, including Tchebychev and Gaussian windows, were considered. These performed worse than the Kaiser window included in the table above.

For N=16, SNR 10, and a frequency range of 75%, the "best" (using the L_2-norm) commonly used windowed FFT (Hanning window, 3-point center of gravity interpolation on squared absolute values of the Fourier coefficients) yields a relative root

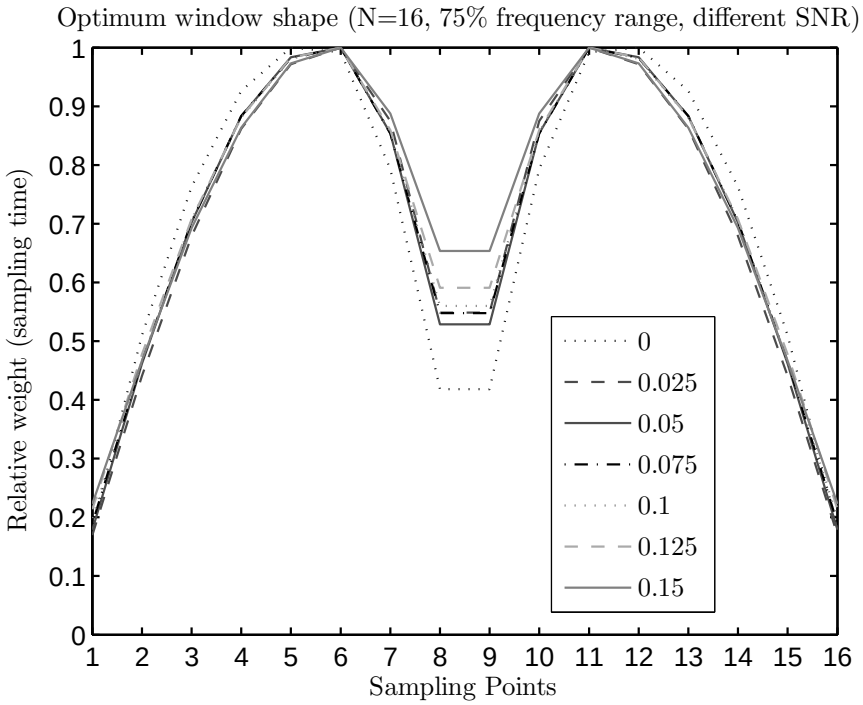


Figure 3.18.: Optimum window shape for N=16 and 75% frequency range, for various noise levels.

mean squared error for the frequency estimate of 0.62%. If the non-linear correction is enabled with this window, the error can be reduced to 0.45%. The window obtained from the optimization proposed in this paper has a root mean squared error of 0.35% (using quadratic interpolation in this case) and with considerably more difficult 2D-interpolation an error of 0.32% can be reached, while the CRB is at 0.25%.

The shape of the optimum window according to the algorithm above strongly depends on the SNR (Figure 3.18) and the desired frequency range (Figure 3.19). The higher the noise, the broader the window becomes, and the "dent" in the center gets smaller as well. Using a window that keeps more data improves the SNR of the estimate, at the cost of stronger systematic errors due to aliasing. There are two different basic window shapes, one with a single "dent" in the middle and the other with three "dents". Their performance is very similar, the second shape performs slightly better when the frequency range is smaller and ends at Fourier frequencies.

The resulting root mean squared error for each combination of SNR and frequency range is shown in Figure 3.20.

In addition to the parameters discussed above, the result of the optimization depends on the error norm. For implementation and comparison the mean squared error is very useful; for use in a measurement system a higher order norm might be preferable as outliers can be more critical than the standard deviation. The influence on the resulting window is quite small though and therefore this issue is not discussed any further.

The results as summarized in Figure 3.17 and Figure 3.20 show that the perfor-

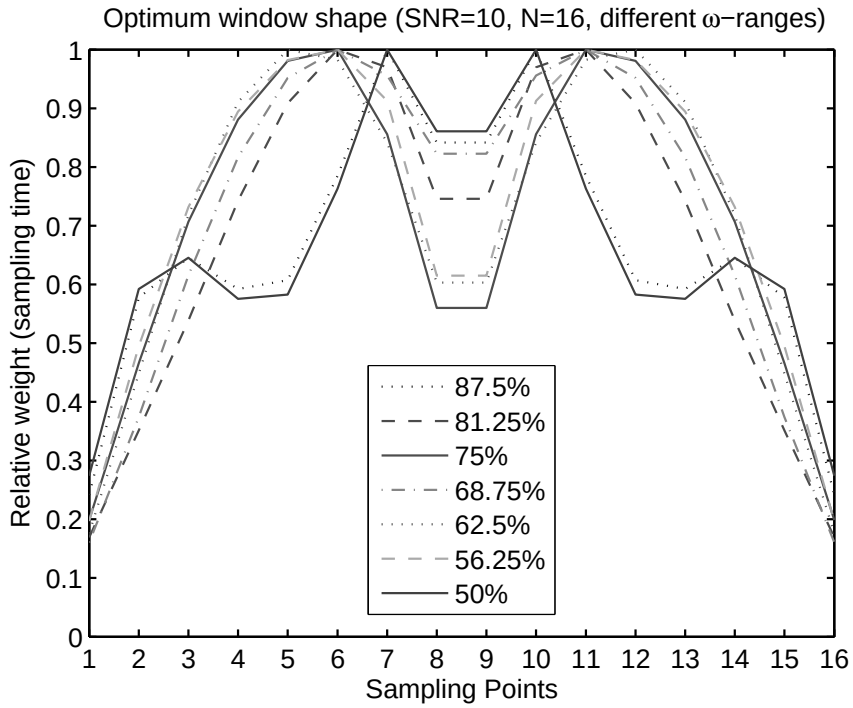


Figure 3.19.: Optimum window shape for N=16 and SNR=10, for various frequency ranges.

mance does not reach the CRB, but the discrepancy is not very large either. An advantage of the approach above is its high flexibility: For any set of parameters, one can quickly find a new optimum window, without any manual interaction. The more that is known about the system and noise parameters, the more the result can be improved. Compared to "normal" windows with the same computational effort for the frequency estimation itself, the improvement is significant. If the accuracy is still not sufficient, the frequency estimate can be used as an initial value for further optimization by other algorithms.

3.6.5. Estimation of Phase and Amplitude

The theoretical limit on the accuracy of phase estimation is shown in Figure 3.10 (bottom, the square root of the CRB is given, settings are chosen as for the frequency estimation in Figure 3.10 (top)). There is a problem with very low or very high frequencies, similar to the frequency estimation issues. Other than that, the theoretical accuracy is almost constant and does not depend much on the true signal frequency or phase.

It is possible to estimate the signal phase and amplitude directly based on the FFT coefficients. This is an option if processing time is very limited, but while it works reasonably well for the non-windowed case, performance is not very good with the optimum windows derived above. In any case, it is best to estimate the phase in the

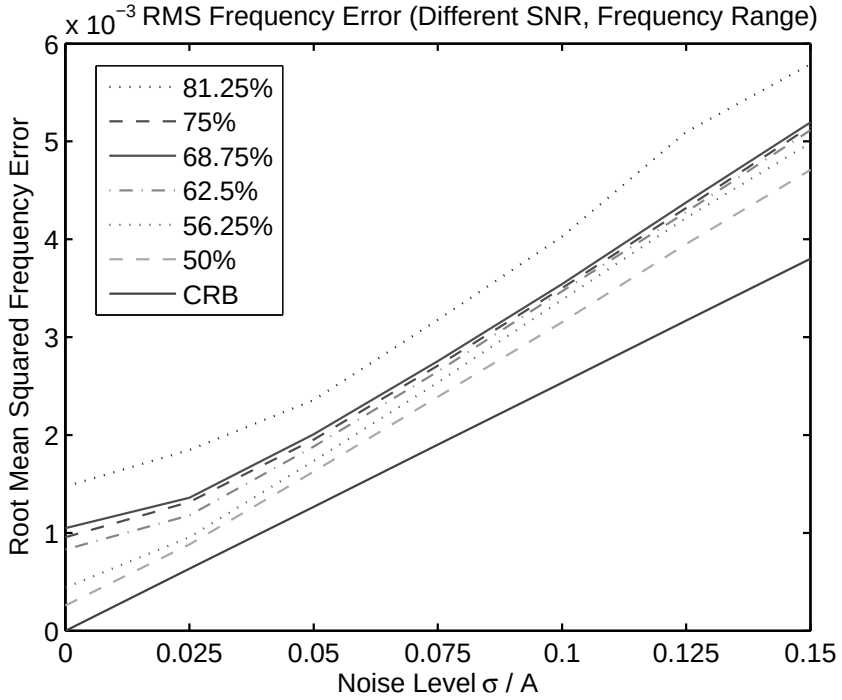


Figure 3.20.: Root Mean Squared (RMS) error of the optimum window for different SNR and frequency ranges. A lower limit based on the CRB for the estimation of real-valued sinusoids averaged across the frequency range is included for comparison. The difference in the average CRB for the various frequency ranges considered here is very small and therefore only the values for a frequency range of 75% are included in the graph.

center of the block of samples. If one uses the coefficients of the windowed FFT, the best (but still far from optimal) results can be obtained by using the phase of the coefficient with the largest absolute value only, as this peak has the best SNR. Interpolation between two peaks in the Fourier domain — as is frequently used elsewhere — is significantly more sensitive to noise.

It is also possible to use linear least squares estimation. Uniform sampling is not required in this case. The performance using this algorithm has recently been analyzed for the estimation of phase in [So, 2005]. A very fast and simple algorithm is derived below which further reduces complexity. It is designed for zero-mean data, as removing the mean is necessary for the frequency estimation above anyway. The signal model can be rewritten as follows:

$$\begin{aligned} I(t_n) = A_1 \cdot \left(\cos(\omega \cdot t_n) - \frac{1}{N} \sum_{k=1}^{N} \cos(\omega \cdot t_k) \right) \\ + A_2 \cdot \left(\sin(\omega \cdot t_n) - \frac{1}{N} \sum_{k=1}^{N} \sin(\omega \cdot t_k) \right) + D \end{aligned} \tag{3.40}$$

with amplitude $A = \sqrt{A_1^2 + A_2^2}$ and phase $\phi = \tan^{-1} \frac{A_1}{A_2}$.

If ω is known, then the equation above is linear in $A_{1,2}$ and D corresponds to the sample mean, which can be removed. Based on that we can define the linear least squares estimation problem:

$$\begin{aligned}
&A \cdot x = b, \\
&\text{with } b = \vec{I},\ x = [A_1\ A_2],\ \text{and} \\
&A = \begin{bmatrix} \mathrm{c}(t_1) & \mathrm{s}(t_1) \\ \mathrm{c}(t_2) & \mathrm{s}(t_2) \\ \cdots & \cdots \end{bmatrix} \\
&\mathrm{c}(t_n) = \cos(\omega \cdot t_n) - \frac{1}{N} \sum_{k=1}^{N} \cos(\omega \cdot t_k) \\
&\mathrm{s}(t_n) = \sin(\omega \cdot t_n) - \frac{1}{N} \sum_{k=1}^{N} \sin(\omega \cdot t_k)
\end{aligned} \tag{3.41}$$

The linear least squares solution $x = (A^T A)^{-1} A^T b$ requires inversion of a 2×2 matrix only, and due to the symmetry this yields:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}^{-1} = \frac{1}{ac - bb} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix} \tag{3.42}$$

For phase estimation, only the ratio of the two elements of x is needed, therefore the denominator does not have to be computed. $(A^T A)^{-1} A^T$ can be computed very quickly, and it can be stored in a table for the applicable frequency range (requiring 2N values per frequency). From this table, the best set of entries for phase estimation can be chosen based on the frequency estimate. The signal only needs to be multiplied with a matrix consisting of the corresponding 2N table entries; this yields two coefficients which can be used to compute phase and amplitude. Even with a very coarse precomputed grid (less than 100 frequencies), excellent results can be obtained (cf. Figure 3.21).

For high signal to noise ratios, results are within 1% of the theoretical limit, the Cramér-Rao bound. When the signal energy is equal to the noise energy, the error caused by least squares estimation reaches about 5%, for higher noise the difference increases further, as can be seen in Figure 3.22.

The amplitude of the signal can also be determined both in the Fourier domain or by linear least squares estimation. In the Fourier domain, the squared absolute value of all coefficients is needed for finding the position of the maximum anyway, and the signal energy can be found by summation of these values. However, not the signal energy but the amplitude of the sinusoidal part of the signal (also called modulation) is the desired quantity. Estimation based on linear least squares is also possible as described above ($A = \sqrt{A_1^2 + A_2^2}$) and more accurate.

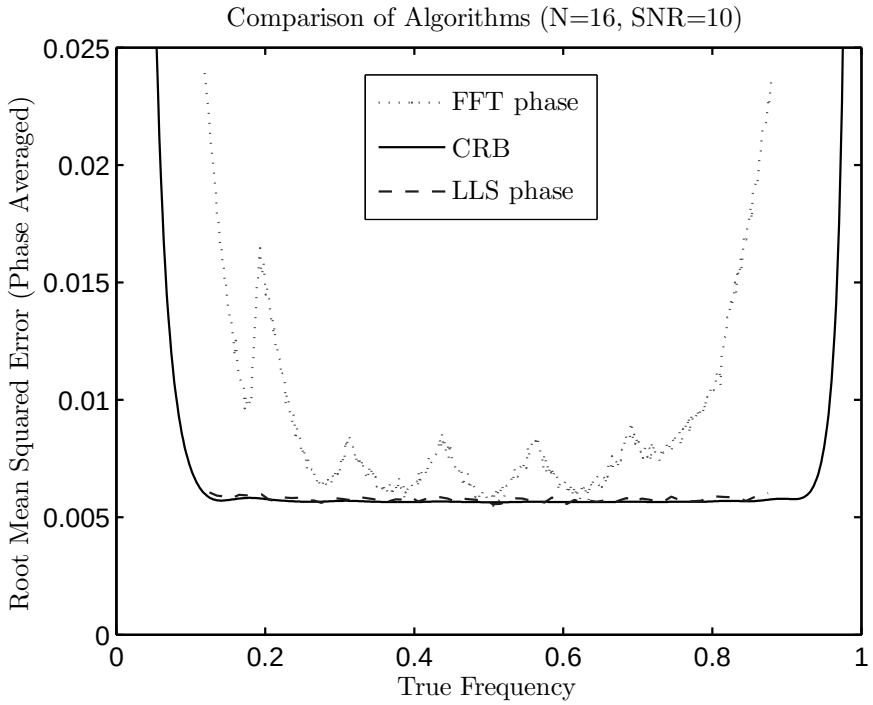


Figure 3.21.: Comparison of FFT based phase estimation and linear least squares (LLS) phase estimation with the theoretical limit, the CRB (for N=16, SNR=10).

3.6.6. Conclusion

An improved frequency estimation algorithm based on the well-known concept of an interpolated FFT has been discussed. This algorithm can easily be adapted to the specific conditions of a given estimation problem (number of samples, SNR, etc.); a simple procedure for that has been described. Some system parameter settings yield rather unexpected window shapes (Figure 3.19), but they consistently offer good performance and the results have been verified in extensive simulations (Figure 3.20 and Table 3.1). In all cases considered, the resulting estimation algorithm offers good performance for frequency estimation and often reduces the RMS error compared to standard approaches (such as a Hanning window with center of gravity interpolation in the frequency domain) by a factor of two while keeping the computational effort low. Optimizing a parametrized window such as a Kaiser window leads to a standard deviation of the resulting frequency estimates that is about 10% higher than with an optimum window.

The only drawback of the new approach is the required optimization procedure for adopting the window to the system parameters, but this is only needed once and can be stored for many possible parameter combinations. This optimization procedure has a key advantage though: It makes the algorithm highly adaptive. Depending on the application, the results can be optimized for different error criteria, different frequency

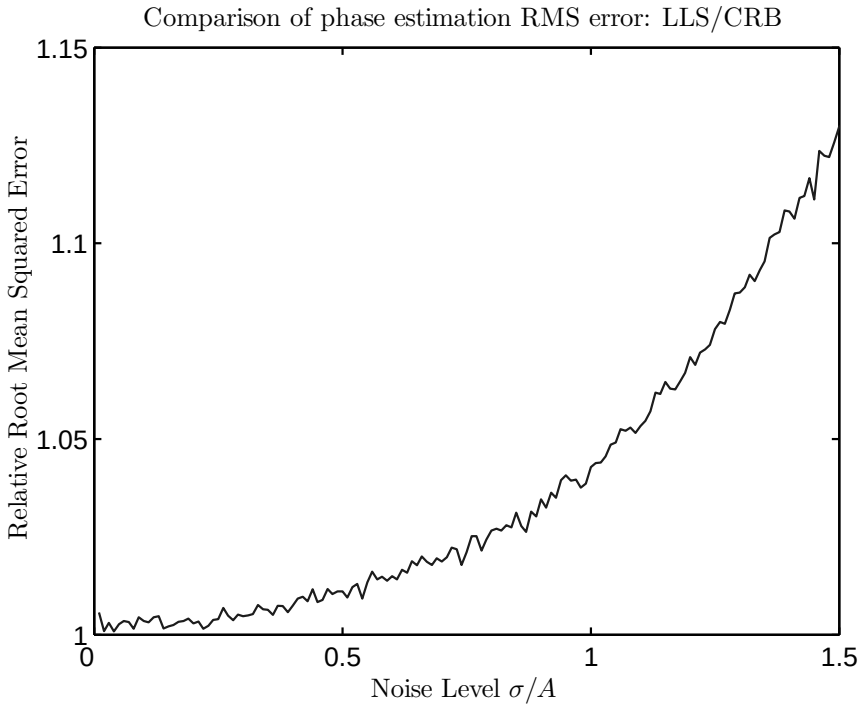


Figure 3.22.: RMS error of LLS phase estimation relative to the CRB (N=16, SNR variable).

ranges, different types of noise, or even for e.g. a low-precision fixed point hardware implementation. The computational effort of the estimation remains the same and the implementation of the estimation algorithm does not need to be changed at all, just the window and look-up-table have to be replaced. In addition, the optimization procedure can be modified to deliver a detailed characterization of the performance of the desired frequency estimator, including the distribution of errors (possibly resolved by frequency and phase) and the influence of various noise levels on estimation accuracy. This makes it easy to determine if the accuracy of this fast frequency estimation procedure is sufficient for a given application or if another algorithm should be preferred.

If both frequency and phase estimates are required, a very fast way to obtain accurate phase estimates has been shown which almost reaches the theoretical limit and can be integrated into any algorithm. Both approaches are drop-in replacements for existing frequency and phase estimators for short blocks of data, and can therefore be used in a large number of applications.

3.7. Laser Frequency Estimation

So far, in all discussions the laser frequencies (= sampling points) have been assumed to be known. There are several ways to measure the actual laser frequencies, e.g. by using a wavemeter, but it is not possible to record these once and for all: The laser frequencies change due to a variety of reasons, most importantly temperature fluctu-

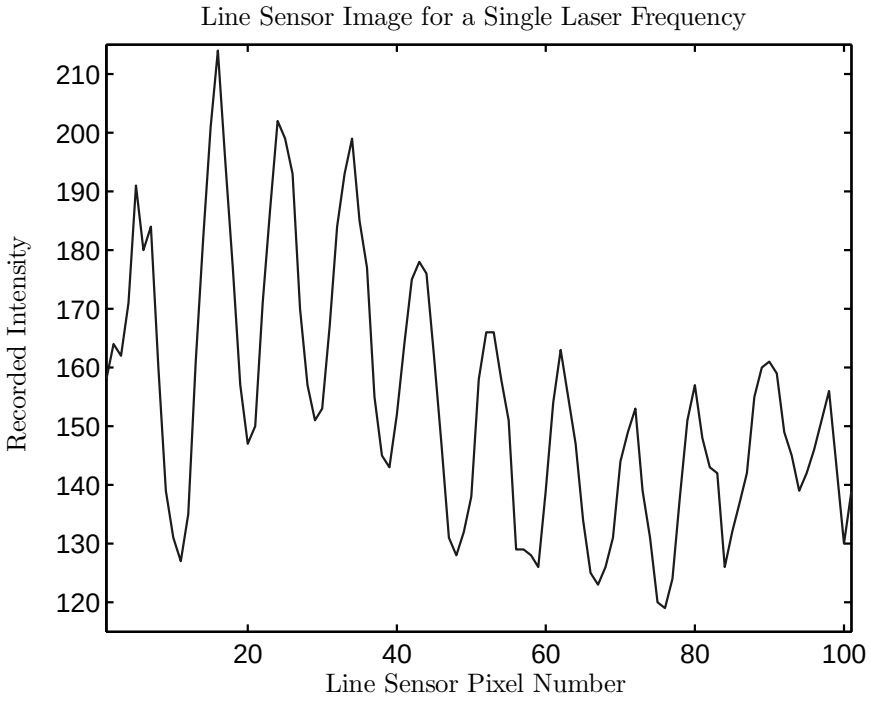


Figure 3.23.: Raw monitor cavity signal for a single laser frequency. The full line sensor signal before calibration is shown.

ations. Therefore a continuous monitoring of the laser frequencies is required. The physical setup of the monitor cavity has been discussed in chapter 3.1.2, in this chapter the focus is on the interpretation of this data.

The monitor cavity is a simple and compact interferometer. Theoretically, the signal is sinusoidal: The frequency of the signal changes slightly when the laser is tuned, and the phase of the signal changes significantly. For the current setup, a wavelength increment of 0.1nm corresponds to a change in phase of multiple signal periods, therefore this phase measurement is highly sensitive and offers a resolution on the order of a few MHz. In contrast to the analysis of the head data, in this case the computational effort is irrelevant as only a single measurement has to be analyzed, and due to the increased number of available samples the estimation accuracy can be much higher. However, the signal quality of the integrated line camera and its illumination are poor, and therefore calibration is an important issue and will be discussed next. While the principle is quite simple, accurate estimation is fairly difficult.

An example for the original recorded signal is shown in Figure 3.23. There is a large offset caused by incoherent light (which varies across the sensor), the modulation changes, and at the beginning and at the end additional artifacts are visible.

Correcting these distortions is possible when a large number of measurements with different phases are available, but this has to be done with care. It would obviously be pointless to force the shape of each measurement individually to be sinusoidal in order to improve the frequency and phase estimation later on, as for that one would

first need to estimate phase and frequency ...

The following calibration approach therefore only performs operations that are independent of signal frequency and phase. The algorithm then works as follows for the estimation of frequency and phase for a set of laser frequencies belonging to one measurement:

1. A range of samples from the line sensor is chosen (same for all laser frequencies). This eliminates areas with bad artifacts, especially at the beginning or end of the sensor array, and it forces the signal to have a (roughly) integer number of signal periods (not exact as the signal frequency is changing slightly).

2. The known offset (from calibration) due to incoherent light is subtracted for each pixel (same for all laser frequencies).

3. As the signal of the line sensor (for every laser frequency) should have zero mean, any remaining mean value is subtracted (different for different laser frequencies, same for all pixels fo the sensor).

4. The modulation of the signal is normalized (per pixel, from calibration, same for all laser frequencies).

5. The mean is removed once again (different for different laser frequencies, same for all pixels of the sensor).

6. The modulation is normalized per frequency (same for all pixels, different for different frequencies).

The result is shown in Figure 3.24.

The modulation per pixel and the offset per pixel are determined in a calibration procedure from a large number of measurements with all possible phase values. This step is more complex as it involves use of the Hilbert transform to obtain an envelope of the signal as well as a phase difference from sample to sample. The resulting estimates are averaged across many measurements and smoothed along the sensor array.

Additionally, the sampling positions known from calibration will be used for the phase estimation. These might or might not be uniformly spaced, depending on the position and alignment of the line sensor.

Now that the signal has been improved by this calibration procedure, frequency and phase estimation can be performed. Given the relatively large number of samples (about 100), the algorithm used to perform the estimation is less critical, the modeling error and the calibration procedure described above have a much larger influence on accuracy. Several methods have been implemented. One of them uses a windowed FFT (similar to the data processing for the measurement head). All the results of chapter 3.6 can be applied here as well, the implementation and optimization of the window is identical. A nonlinear least squares fitting routine has been implemented for comparison, as a FFT on the distorted signal might not yield a fit that minimizes the error in the least squares sense. In addition to that, an iterative FFT was implemented that corrects both its parameters and the sampling points in order to obtain the best fit.

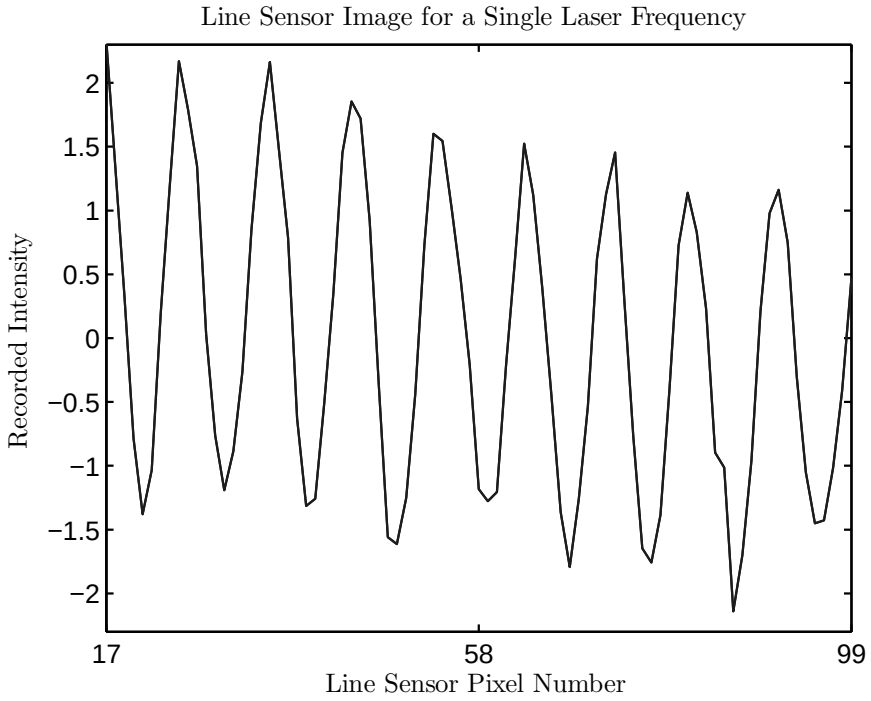


Figure 3.24.: Corrected monitor cavity signal for a single laser frequency. A selected range of the line sensor signal after calibration is shown.

It turns out that the differences in the results between these approaches are very small, on the order of a few MHz.

For further analysis of the system, the ratio of the detected sinusoidal signal to the residual was determined in addition to frequency and phase of the signal. The first result obtained is easy to explain: In the center of the frequency range this ratio is much higher, toward the borders a low frequency component shows up. This can be explained by the change in signal frequency: Only in the center region the choice of sampling points leads to an integer number of signal periods, and only then the mean removal is exact. Additionally, for some of the measurements outliers occur. These measurements also show outliers in the phase. During 200 measurements with 128 laser frequencies each (Figure 3.25), eleven outliers occurred. Visual inspection of the signal shows that these are not failures of the algorithm, but the signal is actually different there. Using a spectrum analyzer and manually playing with the laser parameters explains the reason: In these cases, the laser had multiple longitudinal modes simultaneously. The frequency is therefore not accurately estimated using a single tone model — but the single-tone method is sufficient to detect that there is a problem. Another issue that sometimes shows up is the laser not tuning at all, and therefore showing a completely different phase than expected. Therefore the cavity should be chosen such that the phases for different laser frequencies are different and a typical mode jump leads to a different phase value. This monitoring function is highly useful to decide whether the measurement results can be used.

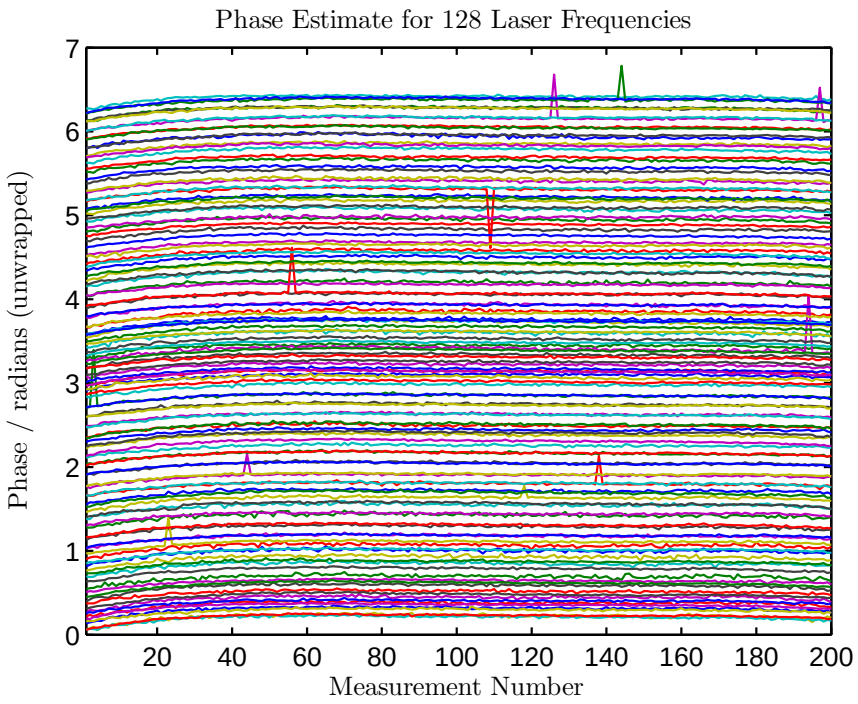


Figure 3.25.: Monitor cavity phase estimates for 200 measurements with 128 laser frequencies each. Both thermal drift (especially at the beginning) as well as occasional outliers can be seen.

In the measurement shown in Figure 3.25, one can see thermal effects when the laser is warming up causing changes in laser frequency (or monitor cavity dimensions). A statistical analysis (taking the physical properties of the laser into account) shows a standard deviation of about 0.2% of the phase for a given laser mode, which converts to about 37MHz (value based on the last 100 measurements); differences between the different algorithms discussed above are one order of magnitude smaller and therefore irrelevant. Looking at the phase differences, the frequency increments between different laser frequencies show significantly larger jitter, on the order of 150 MHz. These differences can be taken into account in the signal processing. Multimoding or tuning errors can be detected as well.

For actual height measurements and especially for using the absolute signal phase, the true absolute laser frequencies are needed. Obtaining them is theoretically simple, but difficult in practice. There is a simple relationship between cavity phase and frequency, but this requires knowledge of the physical dimensions of the cavity. These in turn can be found by measuring the laser frequency with a wavemeter and comparing the results. Unfortunately, no wavemeter of sufficient accuracy was available to analyze that in detail: Most wavemeters and spectrum analyzers are much less sensitive than the monitor cavity. Additionally, the cavity has to be very stable mechanically and must not change if the temperature changes.

3.8. Spatial Filtering

In every kind of measurement some noise is present. Reducing noise during the measurement is often impossible or time-consuming, and therefore it is desirable to obtain the best estimates of the desired parameters in the presence of noise by using all available knowledge. In case of surface measurements, the most obvious approach is using additional spatial knowledge in addition to the individual data for each pixel. This approach attempts to determine the posterior distribution based on prior knowledge (e.g. surface smoothness) and the data. The influence of the data depends on the quality of the data, which can, for example, be determined based on the distance between the signal model and the actual measured data. The general approach leads to Bayesian estimation.

In case of frequency estimation, there are two types of noise imposed on the estimates:

- For low noise on the data, the noise on the estimates is approximately proportional to the noise on the input data.
- For noise above a certain threshold, the estimation might fail "catastrophically": the estimation error suddenly increases rapidly.

This is a fundamental property of phase and frequency estimation, and has been discussed in chapter 3.4. With the algorithm presented in chapter 3.5 the same effect occurs, though somewhat earlier than theoretically necessary. On the other hand, the "catastrophic failure" now occurs in two well-defined steps and causes a very regular structure of the noise that can be exploited using spatial relationships: First, the phase coupling between the two (or more) sampled blocks fails, and (much) later the frequency estimation for the individual, uniformly sampled blocks fails as well. Using that, one can derive a simple filter procedure that performs comparable to Bayesian estimation for reasonable priors.

3.8.1. Filtering in Case of High Signal-to-Noise Ratio

If the signal quality is good, the noise can be roughly approximated as additive Gaussian noise, which is confirmed by measurements and simulations. Reducing this noise is easy and methods are well-known, for example by using a Gaussian lowpass filter. For discontinuous surfaces, edge-preserving filters can be used, including rank order filters and anisotropic diffusion filters, e.g Perona-Malik. All filters from standard image processing can be applied.

Results can be improved if the signal quality is taken into account by a weighting of the pixels within the filter mask. The signal quality can be approximated by the signal modulation: The noise level can be assumed to be uniform across the field of view, and then the signal modulation is proportional to the SNR for each individual pixel. While this is a good approximation, there are several additional properties of the noise that can be taken into account if desired: A more precise estimate can be obtained by taking the ratio between modulation and residual into account. Additionally, the SNR for two signals with the same modulation but different offset is usually better for the

signal with the lower offset (due to the properties of photon noise). Saturated pixels will yield very poor results, even though the modulation might be high.

Weighted Gaussian Filter

A more detailed discussion of optimum filters for noise suppression can be found in [Restle, 2003]. For the following measurements a very simple and very fast filter was used: A Gaussian filter (as the noise is approximately Gaussian) with the cut-off frequency adjusted to the measurement noise, but constant across the field of view. The quality of the pixels was taken into account by simple thresholding, i.e. low quality pixels were simply excluded. This filter is extremely fast, as it can be implemented as a linear filter: The invalid pixels can be set to zero, and the scale can be corrected by filtering the binary map of valid and invalid pixels with the same Gaussian filter mask, and dividing the two resulting images pixel by pixel. For regions with many invalid pixels, the result is obviously not very useful, therefore such areas were excluded and the pixels set to "not-a-number" (NaN). A more involved implementation can take the quality of the individual pixels within the filter mask into account by weighting the samples appropriately and by adjusting the filter mask dynamically.

Kriging

One way of taking the quality of individual pixels and a more complex surface model than just "smoothness" into account is kriging. The autocovariance function of the surface can be specified, and depending on its structure, this method will permit steps. The surface estimate will not necessarily go through the measurement points (nugget model), which is appropriate for smoothing. However, this is more suitable for the interpolation of values given a low number of measurements with large distances; on a closely sampled height map the advantages are small and the computational effort is high.

3.8.2. Filtering in Case of Low Signal-to-Noise Ratio

If the phase coupling between two blocks fails, the error consists of choosing an incorrect k in equations 3.28 or 3.36. If prior knowledge on spatial properties of the surface is available, one can detect and correct these errors. A simple approach could use a linear or a rank order filter as described above, but this would ignore the specific characteristics of the data and the noise on the estimates. Knowledge on the phase coupling error and the signal modulation for each pixel can be used to improve the results. The phase coupling error is a property specific to the algorithm described, and it is simply based on the observation that the approximately Gaussian noise gets either reduced to almost 0 or increased to (multiples of) 2π by an implicit modulo 2π operation limiting the phase to the interval from $-\pi$ to π. Most outliers therefore occur for pixels that are close to the borders of this interval. The error probability for a pixel close to the borders (simply based on looking at this distribution) is 50%, while it is much lower towards the center (exact values depend on the width of the distribution, but typically less than 1%). The signal modulation can be taken into account as well as it typically coincides with the SNR.

Adaptive Median Filter

For a fast filter implementation one can use a standard median filter. Its performance can be improved by adding adaptive thresholds based on the signal quality to determine which pixels will be changed at all and which pixels will be used as input values. The logic is fairly simple:

- If the input quality is larger than an upper threshold: output = input → done.
- Determine the median of the pixels within the filter mask, only using pixels above a defined minimum quality.
  - If the number of pixels above the minimum quality is sufficient: output = median → done.
  - If the number of pixels above the minimum quality is too low: output = NaN → done.

The first step ensures that pixels with high signal quality will not be modified at all. This preserves fine surface structures. The second step replaces pixels of poor signal quality with the median of the values in their neighborhood, if there are enough pixels in this neighborhood of sufficient signal quality. Otherwise these pixels are removed from the result, as there is not enough usable information and the probability of an error is therefore too high.

Remapping of Pixels

Based on the same general idea, but without performing median filtering, a filter can be implemented to specifically correct phase coupling errors. This slightly more complex version of the algorithm above removes almost all outliers from the measurement data without losing any details. This algorithm uses the special structure of the outliers as described above. Only the coupling coefficient k is modified by the filter, leading to very low requirements on spatial smoothness compared to other filters: At least half the pixels in the filter mask are assumed to lie in a region of ± 15 microns for the typical measurement settings. This is fulfilled by almost all surfaces where interferometric measurements are useful, even rough ones. At surface edges a decision might be impossible for some pixels (leading to NaN values if the signal quality is poor), but outliers are very unlikely. The algorithm has the following structure:

- If the weight of the input pixel is larger than a maximum weight: output = input → done.
- If the weight of the input pixel is lower than a minimum weight: output = NaN (The algorithm does not attempt to fill gaps or smooth data, this could be done in a second step with standard image processing.) → done.
- Determine median of the region around the input pixel, and cluster data into three bins: one bin centered around the median and the others centered on the values corresponding to errors in k of ± 1; values outside this range are ignored.

- If the number of pixels in the center bin is more than half the number of pixels within the filter mask or if at least 70% of the pixels in all three bins are in the center bin: Output = input*, mapped to the same interval as the median. This step is not strictly necessary and the requirements are heuristic, but it accelerates computation for the large number of cases where the result is totally obvious → done.
- If the number of pixels in all bins is less than a threshold: Output = NaN. This step is also not strictly necessary and only helpful for acceleration of the processing. → done.
- Compute weight (based on phase coupling error and signal modulation) for each pixel. This can be an arbitrary function, in the simplest case it could be the product of the signal modulation and the relative phase coupling match ("1" if the phase difference is 0, and "0" if the phase coupling difference is π).
- Determine a new median value based on the pixels in the three bins mapped to the same interval.
- Choose a new interval based on the total weight of the pixels within each of the bins.
  - If the total weight of the pixels in the three bins is smaller than a minimum weight: Output = NaN (insufficient data = no result) → done.
  - If the total weight of pixels in the three bins is larger than a minimum weight: Output = input (but mapped to the interval centered around the new median) → done.

The current implementation uses nested for-loops in Matlab and is much slower than a normal median filter, but with a more efficient implementation processing time can be reduced significantly. It should be possible to reach almost the same processing time (at least asymptotically for large filter sizes) as a normal median filter. The most expensive operation in any median filter is the sorting of the pixels in the filter mask, with a complexity of $O(N \log N)$, and for the new filter this can be implemented as in any other median filter. The additional binning procedure is linear in time with the number of pixels in the mask, as is the weight computation. The different thresholds and conditions introduce a fixed and very small additional computational offset. The algorithm does not require any global relationships and can therefore be partitioned for parallel execution without any loss in accuracy. In contrast to unwrapping algorithms it works perfectly well for discontinuous surfaces, and it still preserves all available detail except for very small (depending on signal quality and the filter size, on the order of less than five to fifty pixels) and at the same time very high (more than 15 microns) features with lower signal quality than the surrounding areas. Such features (e.g. a dust particle) remain visible, but they might get mapped to an incorrect interval.

In a second step, the resulting height map can be processed as described in the previous section in order to enforce spatial smoothness and reduce noise. Now that outliers have been removed, anisotropic filters or linear filters can be used, and no special robust filtering methods are needed any more.

Filtering in Case of Very Poor Signal Quality

When the frequency estimation from the uniformly sampled blocks of data fails completely, the filtering approaches above do not work any more. This effect occurs at a noise level that is much higher than the point of failure for the phase coupling procedure. It might still be possible to use neighborhoods to improve the final result, but this has to be done while the full raw data is available, analogous e.g. to [Hissmann, 2005] in case of white-light interferometry. For multiple wavelength interferometry such an approach is more difficult, as — unless uniform sampling is used, which would defeat the purpose — the probability distribution does not have a single mode, but a large number of modes (relatively narrow peaks) corresponding to each of the possible choices for the phase coupling. Additionally, once the frequency estimation for individual blocks of data breaks down, the noise is so high that there is little point in trying to improve the result using spatial relationships, as the accuracy will be very low anyway (as can be seen when looking at the CRB). It is therefore likely that other, better suited measurement principles are available and should be used instead (e.g. fringe projection or stereo camera systems).

Phase Unwrapping

If there is a very strong prior, accurate results are possible even in case of noisy measurement data. Under the assumption of a completely smooth surface (i.e. a height difference of less than $\lambda/2$ between neighboring pixels), spatial unwrapping can be used. Then only the phase of the signal is needed. If the object distance and therefore the signal frequency is not known, it can be determined by spatial averaging of the frequency estimates, preferably using robust estimation and weighting according to the estimated SNR of each pixel. Using that frequency, the phase can then be determined and any of the known algorithms for spatial unwrapping can be used.

3.8.3. Performance of Quality Measures

Before any kind of weighted filtering can be applied, the correlation between signal quality measures and estimation variance needs to be found. In theory, this is simple: The modulation should be inverse proportional to the standard deviation (chapter 3.3) if the noise is uniform and uncorrelated. However, the noise is not uniform, and it is highly correlated. The system model does not take errors such as dispersion or multiple reflections in the optical path into account.

The best method is therefore to measure this relationship directly. For the following graphs, pixels were sorted using bins of size 10% with respect to signal modulation or phase difference. For different values of the signal modulation the average RMS error of the pixels in the measurement of a flatness standard is shown in Figure 3.26. This result matches theoretical expectations very well. There seems to be a lower threshold at about 100nm RMS error though that might to be caused by systematic errors that are not reduced when the signal modulation increases. This may be caused by internal reflections or by camera nonlinearities.

The same analysis can be performed for the phase coupling error. The corresponding measurement results are shown in Figure 3.27. This relationship is slightly less

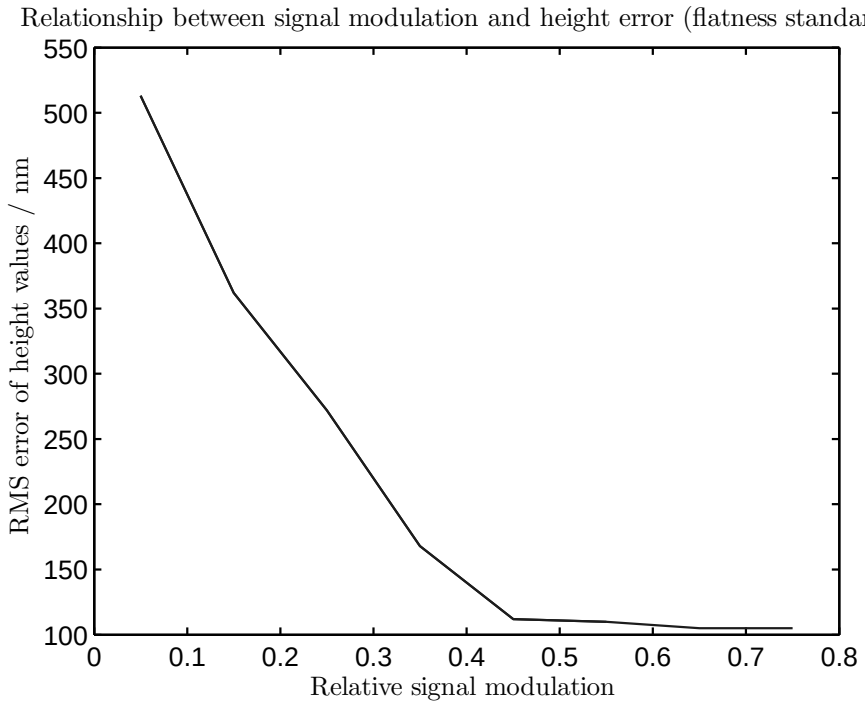


Figure 3.26.: This graph shows the relationship between signal modulation (determined by a least squares fit) and the RMS error of the height values (obtained by comparison with a reference measurement using spatial smoothness). N=16 samples per block, based on a measurement of a flatness standard.

pronounced, but it is still clearly visible. The standard deviation increases proportionally to the phase coupling error. The standard deviation is never zero, as there is always additive noise present which leads to errors in the frequency estimation, even though these errors might cancel when looking at the phase difference only.

When looking at the error for various signal modulations, it is important to know the histogram of the modulation for a typical measurement. This looks very different for rough surfaces as will be seen in section 3.11. For a smooth surface, it is shown in Figure 3.28. This result indicates that illumination is poor: As the surface has homogeneous reflectivity, this distribution is mainly caused by non-uniform illumination.

The same analysis can be performed for the phase coupling difference Figure 3.29. This shows that for 16 samples the phase distribution is quite narrow, and therefore a larger block distance or a smaller block size could be used.

3.8.4. Comparison of Results

Comparing the filters above on simulated data shows a significant reduction in noise and outliers. Some results for simulated smooth surfaces are shown in the next section in Tables 3.2 and 3.3.

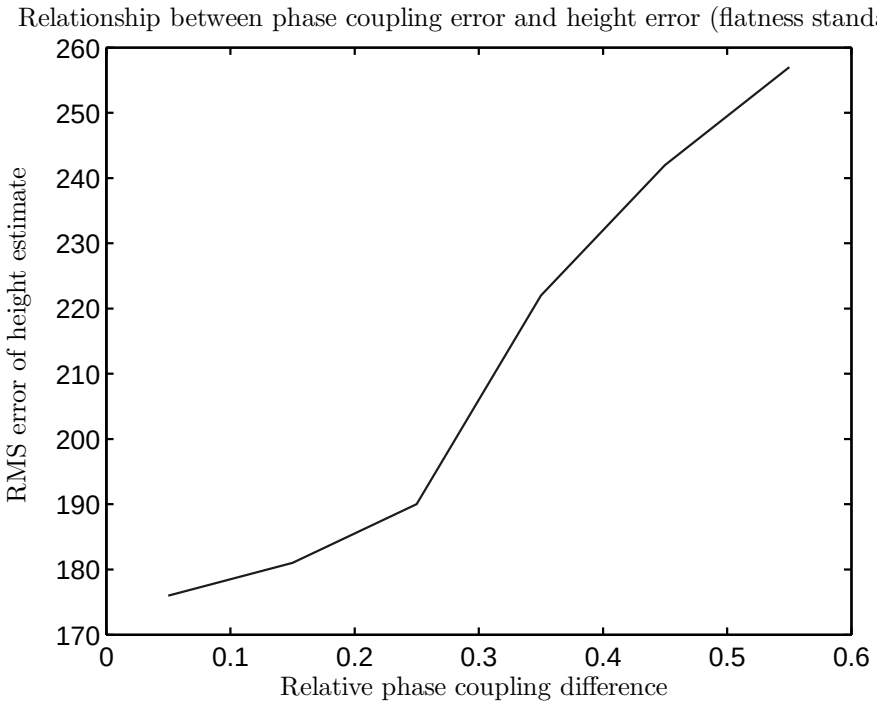


Figure 3.27.: This graph shows the relationship between the phase coupling error (using two blocks of data and least squares phase estimation) and the RMS error of the height values (obtained by comparison with a reference measurement using spatial smoothness). N=16, based on a measurement of a flatness standard.

In the following, the application of the filters to real measured data of a smooth object is shown. The different effects of the filters can be seen in Figure 3.31, Figure 3.32, Figure 3.33, and the unfiltered data is shown in Figure 3.30. A more detailed look at measurement accuracy will be presented in chapter 3.10.

These results show that only the median and the remapping filter perform well - which is not surprising as the noise mainly consists of outliers. The remapping procedure preserves much more detail though. The data without outliers can then be filtered with e.g. a Gaussian filter. This is not useful if the absolute phase has been used, though, as in these cases there is almost no noise present and the advantage of filtering will never outweigh the loss of detail. However, if absolute phase estimation is not used, e.g. on a (slightly) rough object, such filtering might be useful to reduce measurement noise.

Spatial filtering on a rough object in general is questionable, but in some cases it may be useful:

- If the roughness is small enough such that the weak smoothness constraint of the adaptive remapping filter is acceptable.
- If the surface microstructure is not interesting, but instead the global shape or

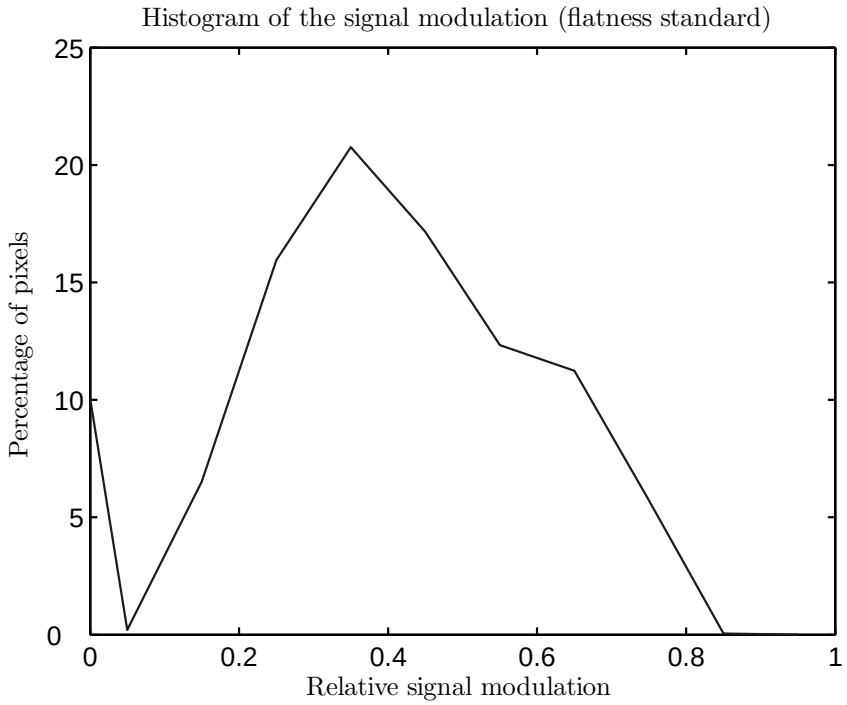


Figure 3.28.: This graph shows a histogram of the signal modulation for the measurement of a flatness standard. N=16, the value at 0 corresponds to pixels that have been excluded as they did not belong to the measurement object.

the distance between various surfaces has to be measured.

For the following analysis the same set of data from a smooth flatness standard will be used, but now only phase coupling between the two blocks is performed, without using the absolute phase. This is the processing that can be applied to rough surfaces or if the absolute laser frequency is not known exactly. For these results, the general measurement noise level can be higher, and — as the highly sensitive step of going to the absolute phase is not used — the probability of outliers (outliers now only consist of pixels where the phase coupling step between the two blocks failed) is much lower. Therefore much smaller filter kernels are sufficient to remove outliers (5×5 was used in the following; for real data of a rough object a larger filter size might be required due to the higher noise level for many pixels caused by speckle). The different effects of the filters can be seen in Figure 3.36, Figure 3.37, Figure 3.35 and the unfiltered data is shown in Figure 3.34.

Both median filter and adaptive remapping are able to remove all outliers, while the Gaussian filter performs poorly. The median filter obviously performs best for this data set as the measured object is actually smooth, but for a rough object it would not be appropriate. A combination of adaptive remapping and Gaussian filtering keeps slightly more details while also reducing the noise. This does not work very well in this example as the noise seems to be highly correlated and depends on the signal

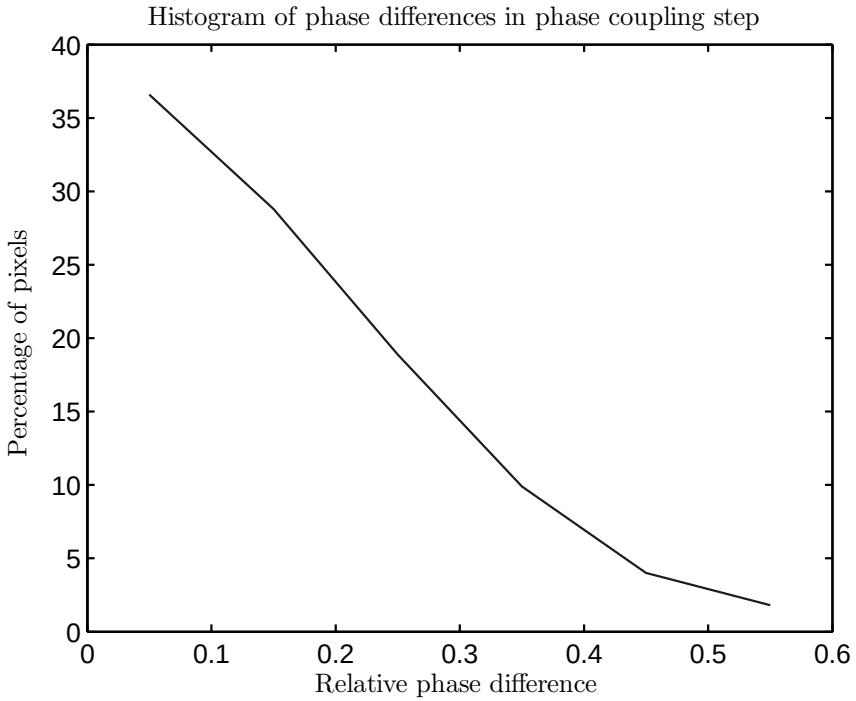


Figure 3.29.: This graph shows a histogram of the phase differences in the phase coupling step. N=16, based on a measurement of a flatness standard.

phase, but on a real rough object the noise should be less correlated. If there are no outliers (for example, due to a larger number of frames being acquired), the Gaussian filter can be applied directly.

These results are not directly applicable to measurements of rough objects due to speckle. A detailed analysis for rough objects is shown in chapter 3.11.

Appropriate filtering can improve measurement results significantly. The key to good results is the notion that outliers still contain useful information, there is just an incorrect mapping that can be corrected. Optimum filtering is an integral part of the system configuration, and using filters to correct outliers can be much more efficient than increasing the number of frames acquired. Often measurement time can be reduced by more than 50% without a noticeable impact on the accuracy of the results. A good indication of the necessity of using filters is the width of the distribution of the phase coupling differences. Whenever this distribution is not approximately zero at the borders, filtering to remap outliers is needed.

3.9. Implementation

All building blocks for an interferometric system have been presented in previous sections. There are multiple ways to combine them in an actual measurement system. Two implementations have been realized in the context of this thesis:

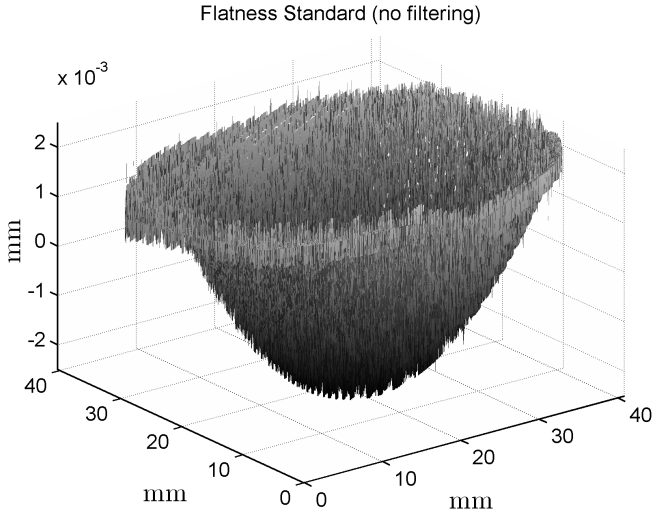


Figure 3.30.: Measurement of flatness standard, using 2x10 laser frequencies, absolute phase and no further processing. Approximately 50,000 outliers occur for the 220,000 pixels on the measurement object; about 1,000 outliers have been removed from the graph as they exceeded the graph range.

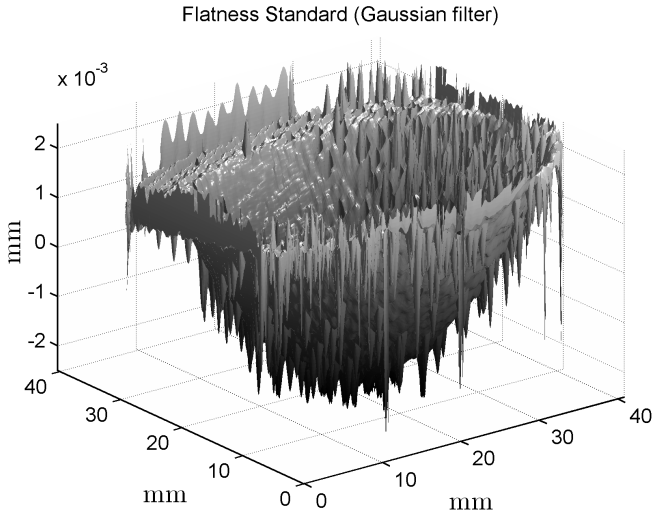


Figure 3.31.: Measurement of flatness standard, using 2x10 laser frequencies, absolute phase and Gaussian filtering (kernel size and cut-off frequency 9) to remove noise. This filter is not suitable for this kind of noise distribution; the graph has been truncated and does not show all outliers.

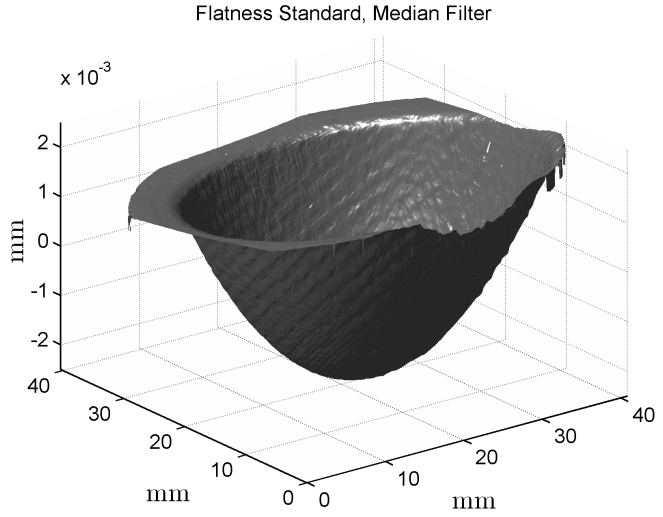


Figure 3.32.: Measurement of flatness standard, using 2x10 laser frequencies, absolute phase and median filtering (kernel size 9) to remove outliers. Most of the fine structures on the surface are gone, artifacts are visible, and a few outliers still remain. The graph shows all valid data points.

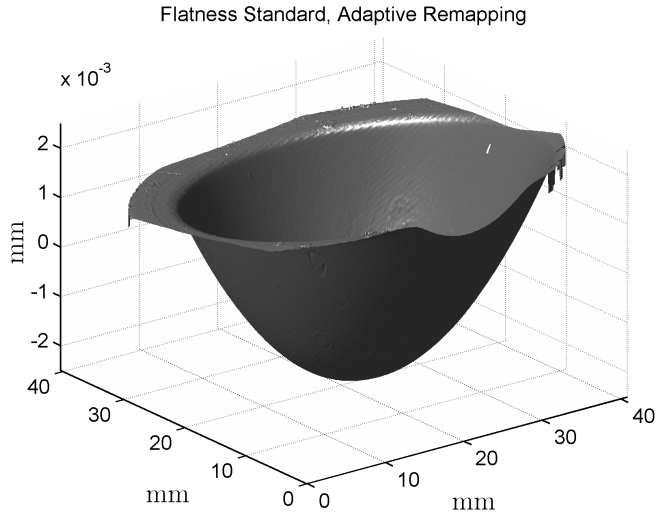


Figure 3.33.: Measurement of flatness standard, using 2x10 laser frequencies, absolute phase and adaptive remapping (kernel size 9). Fine structures on the surface are preserved, no visible artifacts. The graph shows all valid data points.

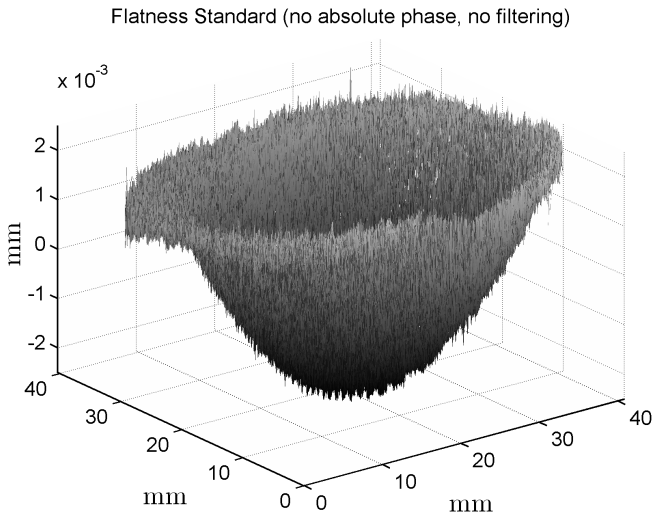


Figure 3.34.: Measurement of flatness standard, using 2x10 laser frequencies, no further processing. Approximately 800 outliers occur for 230,000 pixels on the measurement object; all outliers have been removed from the graph as they were far outside the graph range.

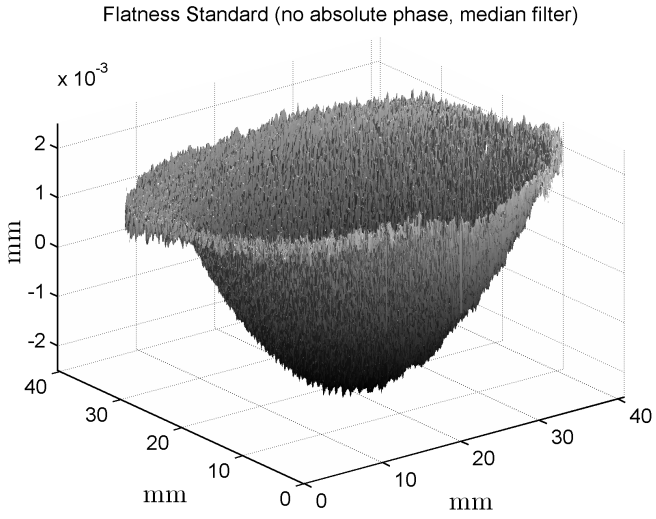


Figure 3.35.: Measurement of flatness standard, using 2x10 laser frequencies, median filtering (kernel size 3) to remove outliers. The signal is smoothed and all outliers have been removed. On this smooth object larger filter sizes would improve results; on a real rough object this would not be desirable.

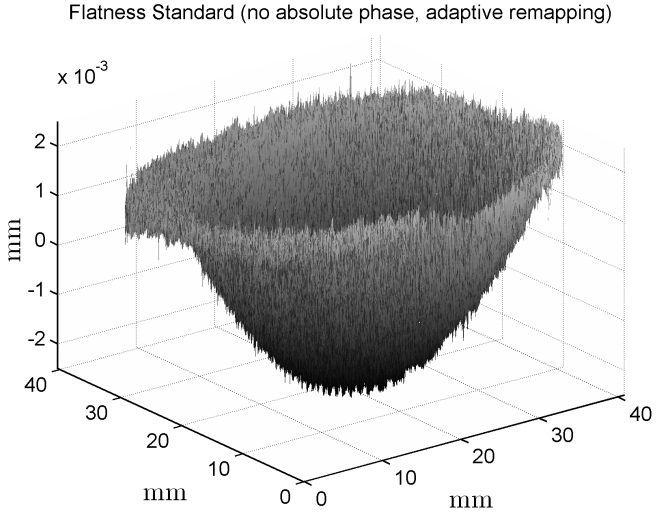


Figure 3.36.: Measurement of flatness standard, using 2x10 laser frequencies, adaptive remapping (kernel size 3). The height map is still very noisy, no spatial smoothing. For measuring a rough object this is desirable.

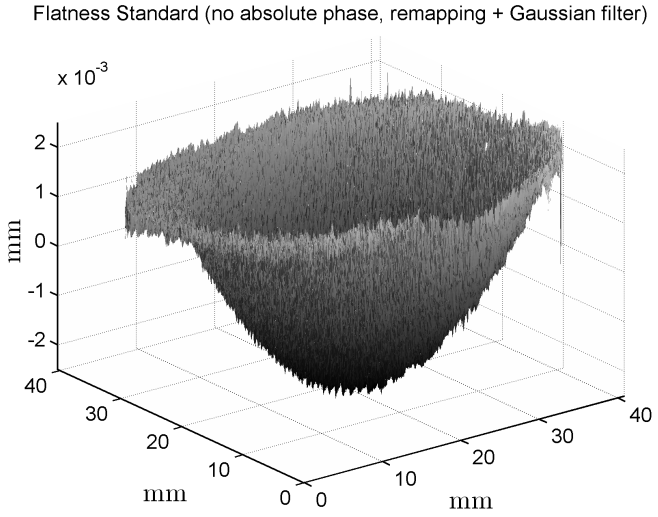


Figure 3.37.: Measurement of flatness standard, using 2x10 laser frequencies, using adaptive remapping (kernel size 3) and Gaussian filtering (kernel size 5 and cut-off frequency 3) to remove noise.

1. A simulation environment that is able to synthesize data, run various types of algorithms and can visualize the results (including a comparison to the CRB). This was used to optimize the algorithms and to obtain most of the results described and discussed in this thesis.

2. A bare-bone plug-in that only contains the parts most relevant for the estimation, optimized for speed and directly integrated into the measurement system software. Accuracy is identical to the full system, but flexibility is lower. It is faster and requires less memory. This was used for the repeatability measurements shown later, and is relevant for a practical implementation.

The structure and the components of both implementations are briefly described in this chapter.

3.9.1. Full Implementation

The full implementation consists of

1. a wrapper program for running benchmarks,

2. a tool for creating test data with various kinds of noise,

3. a program for computing the CRB and the expected areas of outliers,

4. a program for determining the cavity phase from reference cavity data,

5. a program for converting cavity phase to laser frequency,

6. an external tool for creating the cavity calibration data from a large set of measurements and

7. a main script that normalizes input data, calls various frequency and phase estimation algorithms, determines the confidence values, performs phase coupling, spatial filtering and offers a large number of visualizations and parameters.

The frequency estimation can be performed by

1. interpolated + windowed FFT (with various parameters which can be optimized according to many criteria),

2. iterative FFT and

3. non-linear optimization.

Phase estimation algorithms include

1. interpolated FFT,

2. linear least squares estimation and

3. non-linear optimization.

For spatial filtering,

1. an adaptive median filter,
2. a weighted Gaussian filter and
3. a filter for detecting and remapping incorrect height values

are available.

Phase and frequency estimation can be applied iteratively. Use of the absolute signal phase ϕ_0 is possible, and phase behavior at the surface can be given as prior knowledge or it can be estimated from the data. Prior knowledge on the working distance and phase can be used for improved accuracy of the phase estimation steps.

3.9.2. Plug-in Implementation

The plug-in implementation is optimized for speed and only contains the most important algorithms. There is a single data flow, but intermediate results can be returned and processing can be started later, i.e. not all steps have to be done sequentially. This enables use in a software framework from Corning Tropel and supports multiple applications with a single plug-in.

This includes the following steps

1. Get cavity phase
2. Convert phase to laser frequency
3. Interpolated FFT
4. Linear least squares phase estimate
5. Phase coupling
6. Confidence estimation
7. Spatial filtering by adaptive remapping

Prior knowledge on working distance and surface phase can be used in this algorithm as well.

Memory usage is significantly lower than for the full program as all parts have been optimized for low memory usage, use lower accuracy (single precision) wherever possible and discard all intermediate results as soon as possible. The results have been verified to be identical to the results of the full program.

3.10. Measurement Results

Individual aspects of the estimation algorithm and the signal acquisition have been discussed in the corresponding chapters. Now that all relevant parts of the estimation algorithm have been discussed, an overview on system performance is given. In the first section of this chapter, simulated data is used, while in the second section a set of measurements of a smooth flatness standard is used. This is by no means a complete analysis of all possible combinations of algorithms, only the most "interesting" settings (i.e. the ones with the best properties with respect to speed and accuracy) are shown.

3.10.1. Simulated Data

A simple sinusoidal signal model is used in this section. This is not a complete representation of the real signal, but it is a first approximation. Adding more aspects to the signal model and analyzing their influence is simple and would be useful in principle, but it has to be done based on knowledge of the actual system characteristics, otherwise there are simply too many degrees of freedom, and the results offer little insight.

A full system characterization is currently not possible: Measurement data suffers from strong noise due to poor and non-uniform illumination of the field of view, the absolute laser frequency is unknown, and issues with thermal stability and fiber coupling introduce additional errors. In addition, camera linearity is unknown and a spinning disk in the measurement head as well as internal reflections are further unknowns. The properties of the optical components (dispersion, aperture, focal depth, etc.) are also not known. Isolating individual influences has been (partially) possible for the laser intensity fluctuations and the frequency jitter. For now, all other effects are assumed to contribute to additive white Gaussian noise on the signal. An analysis of the residuals shows that the noise is not white, but a large part of the currently visible correlated noise can probably be removed with relatively low effort by additional calibration procedures (i.e. the camera could be calibrated before being mounted in the measurement head) or can and should be reduced by system modifications (i.e. the heating stability issues) before further characterizations are performed.

512×512 sinusoidal signals are simulated, their frequency uniformly distributed across the desired frequency range (typically $\frac{2}{N}$ to $1 - \frac{2}{N}$, as chosen for the optimization of the frequency estimation algorithm). The phase term ϕ_0 according to equation 1.6 is assumed to be constant for the simulation of smooth surfaces and is chosen randomly (uniformly distributed in $[0, 2\pi]$) for the analysis of algorithms for rough surfaces. The error metric used is the root mean squared (RMS) error of the estimates.

Processing steps cannot be treated individually as they depend on each other: Several of the possible options (e.g. a non-linear fitting procedure) offer a significant improvement when used in conjunction with a very basic estimation procedure, but they are actually a step backward when added after a phase coupling procedure using linear least squares phase estimates. In order to find out the most useful and most accurate combination of algorithms, simulations for a large number of possible building blocks have been performed. Each possible algorithm was run in conjunction with multiple

other algorithms, and the resulting performance was analyzed. If it improved performance, it was kept as part of the estimation program. Therefore it is unlikely that a particularly good optimization approach has been missed. This assumption is reinforced by the fact that the results are very close to the theoretical limit.

Then in a second part of the optimization, processing steps were removed in order to find the essential parts of the algorithm. The resulting core algorithm was optimized for speed and was implemented as a plug-in (see section 3.9). This procedure, especially the first step of finding good algorithms, involved a large number of trial runs, and a complete presentation would be much too long in this context. Therefore only the parts used in the final algorithm will be presented in detail, alternative algorithms will be briefly mentioned only. In addition to the influence on accuracy, the computational effort for the steps is considered.

The main observations are summarized below. Unless otherwise noted, all numbers are given for 16 samples in two blocks (on a grid of 128 possible laser frequencies, 47 GHz apart; center frequency 382.5 THz) and a SNR of 10. All algorithms have been run for a range of 8 to 24 samples and a SNR from 10 to 40: no qualitative differences in the behavior of the algorithms could be found.

- Results of the algorithm from section 3.5 are very close to the theoretical limit if the noise level is low and if linear least squares phase estimation is used (less than 5% difference). This confirms theoretical expectations.

- The influence of block size, block distance and the SNR on accuracy and probability of outliers have been discussed in detail in section 3.5. Extensive simulations and actual measurements (see below) confirm the relationships derived there.

- Fourier-based phase estimation according to the algorithm described in section 3.6 is less accurate compared to least squares phase estimation. Both the probability of outliers and the standard deviation increase significantly. Therefore least squares phase estimation will be used; its complexity is O(N) and the algorithm is simple.

- Fourier-based frequency estimation using an interpolated FFT as described in 3.6 is not optimal, but it has almost no influence on the accuracy as long as no outliers occur. The number of outliers is higher than theoretically necessary, though, especially if the distance between the two blocks is large or the noise level is high. This can be resolved by choosing a slightly smaller block distance or more samples for each block than theoretically necessary.

- Once an optimum block distance has been chosen such that the estimation algorithm performs well (this depends on the signal quality and on the acceptable level of spatial filtering), it performs better than almost all alternatives.

  – Slightly different estimation settings (e.g. using an optimum window designed for a SNR of 10 instead of 20) lead to a negligible increase in the standard deviation, but can cause a significant increase in outliers. Again,

this can be resolved by slightly reducing the block distance or by increasing the number of samples per block. This leads to an increase in standard deviation of about 1% in this case.

- Using a window that is significantly worse (e.g a Hamming window) increases the number of outliers dramatically, or — when compensated by reducing the block distance — typically increases the standard deviation by more than 10%.
- Some alternative methods for frequency estimation for the individual blocks can reduce the number of outliers, but these methods are much slower. An iterative approach in the Fourier domain and a fitting algorithm in the time domain have been implemented. Both are much slower, as each iteration requires roughly the same amount of time as the interpolated FFT. Typically three to five iterations are needed. The improvements are very small, and the algorithms seem to be more sensitive to noise caused e.g. by a slow intensity variation of the laser. Attempting to incorporate such effects into the signal model leads to even slower algorithms, and no noticeable performance improvement has been found.
- A number of well known general frequency estimation algorithms have been applied (including e.g. MUSIC). None of them offered good performance with reasonable computational effort for the given sampling pattern (see section 3.6).
- Non-linear optimization of the frequency estimation for all data points could theoretically offer better performance than an analysis of the individual blocks and then combining these results (this improvement must be small as indicated by the theoretical limit). Both a single step of a Gauss-Newton method and Matlab's "nlinfit" function (with its large number of tuning parameters and available algorithms) have been used. The author has been unable to find settings that improved the results: Performance degrades (slightly) compared to results obtained by the phase coupling procedure.

- Results can be improved by iteratively repeating both phase estimation and phase coupling to obtain a new frequency. This improvement in accuracy is small, less than 1% for the configuration given above. There is no change in the number of outliers. This is usually not an attractive option when looking at the increased computational effort. If the number of outliers is not a problem at all (i.e. if the block distance is small due to e.g. limited laser bandwidth), one can choose to perform phase coupling with an FFT-based phase estimate first, use the new frequency for a least squares phase estimate and repeat the phase coupling. This yields a slightly lower standard deviation than the previous method at a small increase in computational complexity.

- The algorithm is quite robust to several common sources of errors in a real system: Simulations with photon noise (Poisson distribution) as well as a linear change in signal amplitude with the laser frequency (20% change from first

N	SNR	CRB in nm	Number of invalid pixels	RMS error in nm	Number of outliers	RMS error in nm (filtered)	Number of outliers (filtered)
8	40	75.7	323	81.8	61	81.8	0
8	20	151.3	1387	159.6	5899	160.2	0
8	10	302.6	6106	312.1	53083	319.6	1436
16	40	57.1	340	59	0	59	0
16	20	114.2	789	116.2	0	116.2	0
16	10	228.3	6991	231.8	159	231.8	0

Table 3.2.: Simulation results for height estimation; the absolute signal phase is not used. This is the algorithm that can be used for rough surfaces; the influence of speckle is not modeled in the simulation though. For the filtered results, adaptive remapping with a 5×5 filter mask was used. The number of invalid pixels (which have been excluded from processing due to low modulation or inconsistencies) could be reduced for the filtered data, but that was not done in order to compare identical sets of data in the filtered and unfiltered case.

to last frequency in this simulation) had little influence on the accuracy of the results.

In Table 3.2 some example results for 8 and 16 samples are given. As the "true" height values are available in these simulations, results can be compared easily. The algorithms are rated by three criteria: The number of outliers, the number of invalid pixels and the standard deviation of the estimates (excluding outliers and invalid pixels). Invalid pixels have been detected to be too noisy based on confidence measures (as described in section 3.8). If spatial filtering is performed, the number of invalid pixels could be reduced at the cost of a somewhat higher standard deviation. In order to keep the standard deviations comparable, this was not done for the results shown here. "Outliers" are all values for which the error exceeds half the error caused by choosing an incorrect k in the phase coupling procedure. If outliers were not excluded, the results for the standard deviation would be much too sensitive to noise and have little meaning.

For illustration, the histogram of the phase differences for the case shown in the second row in Table 3.2 is shown in Figure 3.38, and a histogram of the estimation errors is shown in Figure 3.39 (no spatial filtering) and Figure 3.40 (after spatial filtering).

If the absolute phase is used (examples given in Table 3.3), one assumes that ϕ_0 is constant over (at least a part of) the field of view. This is not true for rough surfaces, but it can be used to improve results for smooth surfaces. No direct spatial relationship is required, therefore this also works for surfaces with steps. The resulting accuracy is much higher, as now the result has the same accuracy as the phase estimation, typically on the order of 1nm. This is much more accurate than required for most technical applications. The accuracy can be increased slightly by obtaining a phase estimate using both blocks of data, but this improvement is irrelevant compared to the error that is caused by an error in the absolute laser frequency (which can be more

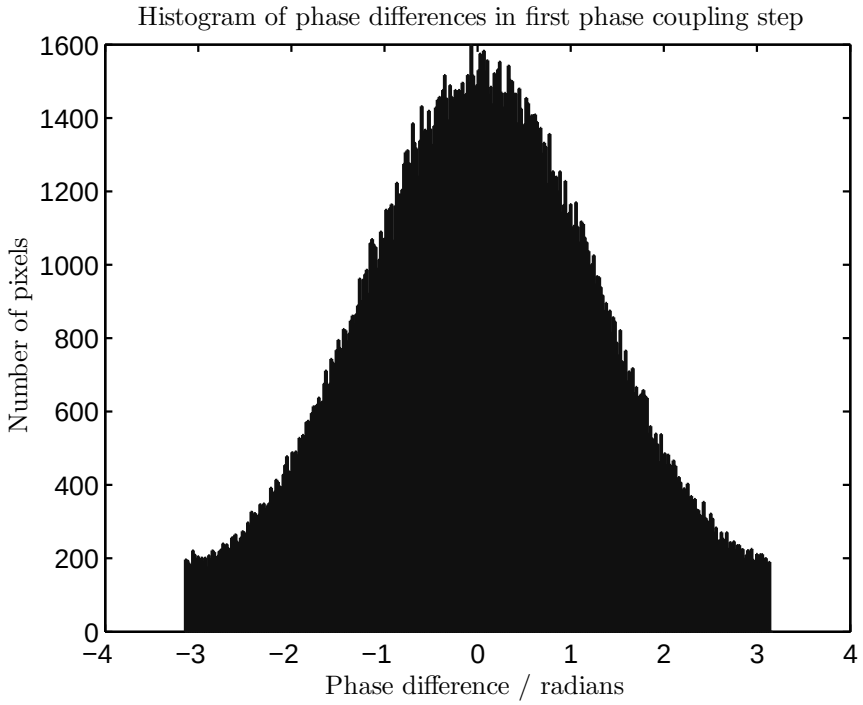


Figure 3.38.: Histogram of the phase differences in phase coupling (N=8, SNR: 20).

than an order of magnitude higher). Therefore it is more important to obtain accurate knowledge of the laser frequencies.

For this case, an example of the phase coupling steps (now there are two steps) for the fifth row in Table 3.3 is shown in Figure 3.41 and Figure 3.42, and a histogram of the estimation errors is shown in Figure 3.43 (no spatial filtering) and Figure 3.44 (after spatial filtering).

Additionally, the algorithms have been analyzed with respect to the influence of a particular type of error: sampling jitter. The tunable laser is not perfect, and therefore the laser frequency increments are not constant. This error is particularly interesting as there are several ways to deal with it. As a monitoring system is included, sampling jitter is known (with some uncertainty due to the measurement, see section 3.7). In the estimation algorithm discussed here, it is not taken into account in every step though. The error caused by sampling jitter increases with working distance and therefore limits the measurement range. Therefore it is interesting to see

- whether it is possible to improve the results by using the knowledge from the monitor cavity and
- what the limit on the working range is that is imposed by the laser frequency errors.

For the values in row five in Table 3.3, the influence of sampling jitter has been simulated. A sampling jitter of 150 MHz (as determined in section 3.7) normally

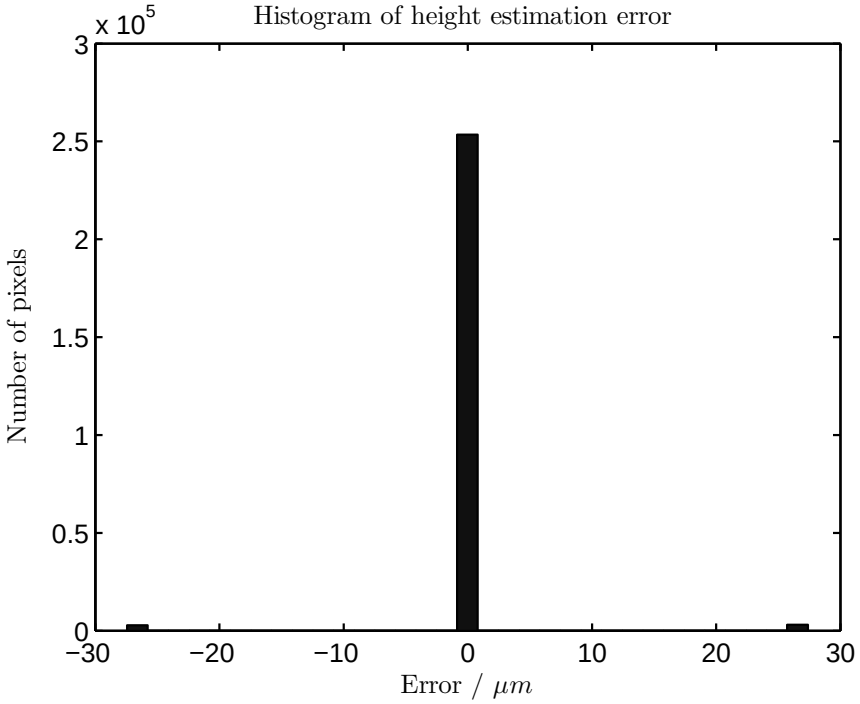


Figure 3.39.: Histogram of the estimation errors (N=8, SNR: 20, unfiltered). One main and two side peaks are visible. The side peaks correspond to pixels where the phase coupling step chooses the wrong k.

leads to a very small increase in the standard deviation (less than 10nm) as long as the measurement distance is assumed to be small. Results are slightly better if the laser frequency jitter is known. This knowledge can be taken into account in the linear least squares phase estimation (without additional effort), but it is not possible to use it directly in FFT based algorithms. Iterative frequency estimation algorithms might help, but these are much slower. When looking at the results for absolute phase measurements and when looking at absolute distances, some issues become visible: While the standard deviation is almost identical, systematic biases of the height measurements show up due to the sampling positions not being quite correct. For larger jitter or larger object distance, this bias can reach more than 100nm, but that depends on the specific sampling pattern. The current software framework is not well suited to perform simulations for different sampling jitter, therefore no precise values can be given here. In real measurement data this bias can be found as well (cf. Table 3.5). A possible solution to that problem would be a monitor cavity in the measurement head, so that the actual frequency as seen by the measurement head can be determined and used.

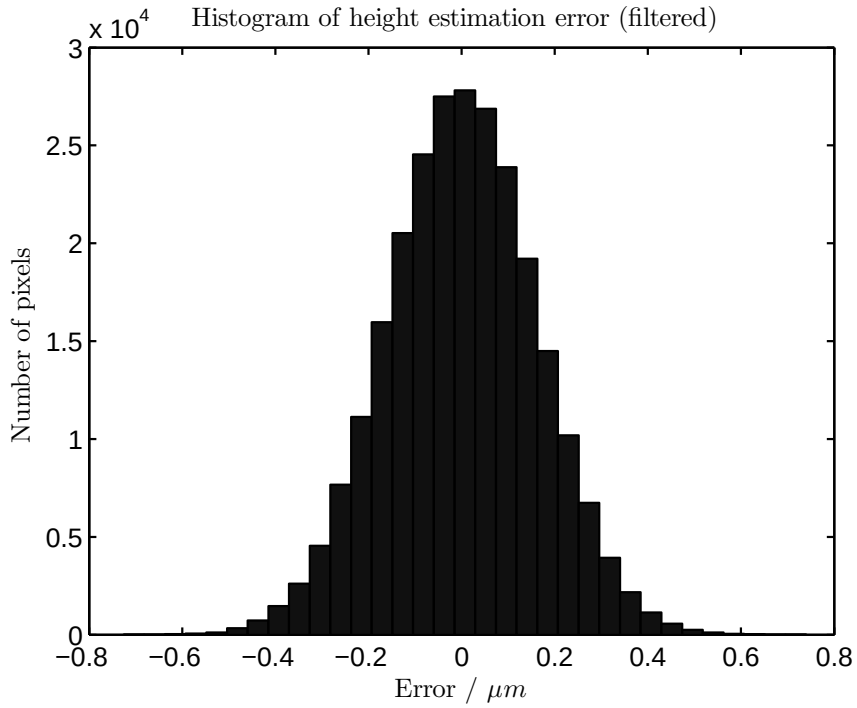


Figure 3.40.: Histogram of the estimation errors (N=8, SNR: 20, filtered). Filtering with an adaptive remapping filter with kernel size 5×5 removed the side peaks visible in 3.39, and now only the main peak remains.

3.10.2. Real Data

A series of measurements of a smooth flatness standard has been performed. As a reference measurement, the result obtained by adaptive remapping (which in this case corresponds to the result one would obtain by spatial unwrapping) is used.

There are four main objectives of this analysis:

- The results for the near-optimum sampling scheme have to be verified on real data. This is presented for various block sizes and block distances in this section.

- The performance of the optimized algorithms has to be compared on real data. This is important as the signal model in the theoretical optimizations is not complete, and therefore might not match reality. It has already been shown in simulations that the preferred algorithm is fairly robust, though. A comparison of different algorithms is easy, as the height maps can be compared to the reference, and the algorithm achieving a lower difference can be assumed to perform better. These experimental results match the theoretical and simulation results very well, but a quantitative statement is difficult as the noise characteristics of the system are not known. This comparison does not add additional insight, therefore the reader is referred to [Klenke, 2007] for a comparison of various types of FFT-based phase and frequency estimation methods and for a

N	S N R	Number of invalid pixels	RMS error in nm	Number of outliers	RMS error in nm (filtered)	Number of outliers (filtered)
8	40	230	0.82	4386	0.84	0
8	20	1184	1.37	61055	1.65	844
8	10	5852	2.39	168320	3.28	62627
16	40	206	0.57	235	0.57	0
16	20	788	1.03	23837	1.13	85
16	10	6895	1.75	112874	2.25	8359

Table 3.3.: Simulation results for height estimation; the absolute signal phase is used. This is an algorithm that can be used for smooth (but not necessarily continuous) surfaces only. The phase ϕ_0 is assumed to be constant for at least part of the field of view. For the filtered results, adaptive remapping with a 9×9 filter mask was used. The data set was created analogous to the one used in Table 3.2, and the results illustrate the huge performance improvement if the signal phase can be used. They also show that the adaptive remapping filter is very powerful and very useful if smooth surfaces have to be measured.

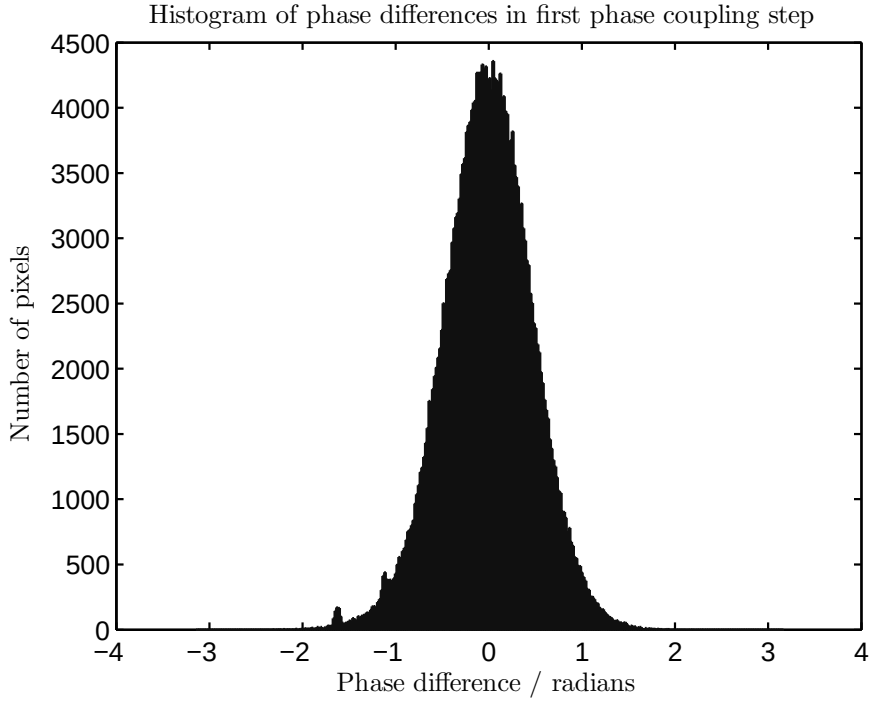


Figure 3.41.: Histogram of the phase differences in phase coupling (N=16, SNR: 20).

comparison of filters. A comparison of multiple other algorithms for frequency estimation, including MUSIC and a zero-padded FFT, can be found in [Pfaff,

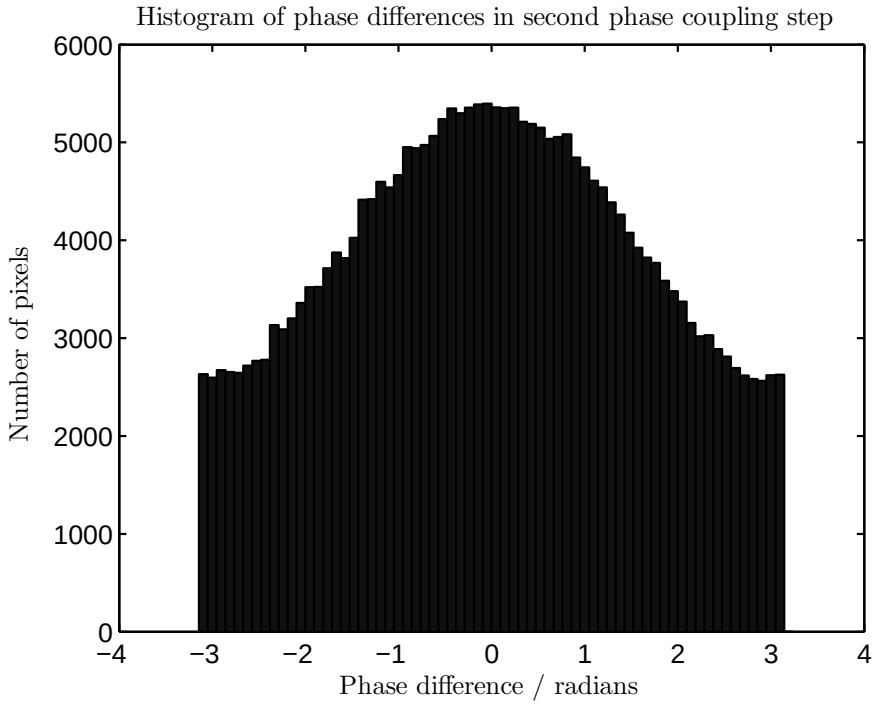


Figure 3.42.: Histogram of the phase differences in second phase coupling (N=16, SNR: 20). This is needed to use the absolute signal phase.

2005].

- Signal properties for filtering have to be determined based on real data. These results have already been presented in section 3.8.
- Long term stability and repeatability of the system have to be analyzed. This has been done under laboratory conditions; the system is currently too sensitive to vibration and changes in temperature for use in a production environment. Results for measurements of a step height artifact are shown below.
- The influence of rough surfaces has to be analyzed. This is discussed in section 3.11.

In Table 3.4, the results obtained from real measurements of a flatness standard with a depth of 4.7 microns for settings similar to the ones used in the simulations (Table 3.2) are given. The influence of varying block size is illustrated as well as the benefit of filtering the data. In this case, a measurement obtained using the absolute phase and adaptive remapping is used as a reference. The height map corresponding to that measurement is shown in Figure 3.33. Without using the absolute phase, the RMS error decreases with the number of samples used, but it does not go to zero. This is not a matter of the SNR only as can be seen in Figure 3.26. This is probably caused by systematic errors of the measurement system.

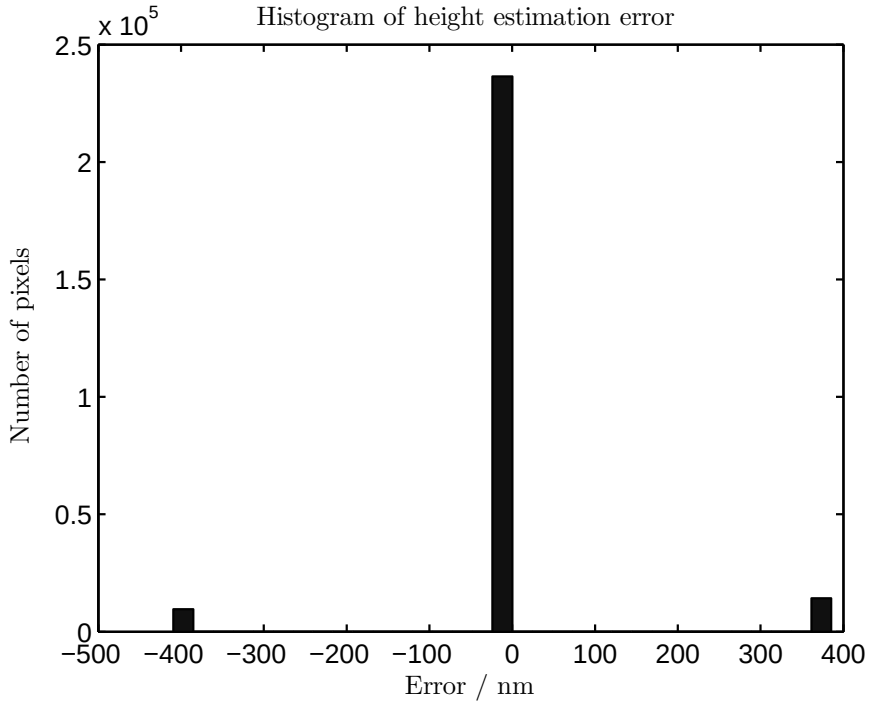


Figure 3.43.: Histogram of the estimation errors (N=16, SNR: 20, unfiltered). One main and two side peaks are visible. The side peaks correspond to pixels where the phase coupling step chose the wrong k.

This is confirmed by the fact that repeatability measurements (by performing multiple measurements without moving the object) show significantly lower variance. For the flatness standard that is discussed here, a standard deviation of the flatness of about 36nm was determined (nominal flatness value 4.7 microns). This still holds true for large measurement distances of 1cm and 2cm, which indicates that the laser tuning can be quite stable. This is less than half the error the error seen when looking at the differences to the true value, indicating that most of the error shown below is systematic and can probably be reduced somehow.

The phase coupling differences for one such measurement are shown (Figure 3.45 and Figure 3.46), as well as the error distribution with (Figure 3.48) and without (Figure 3.47) spatial filtering. These results match the results obtained from simulations quite well, and the analysis of the error for the different signal modulation levels in Figure 3.26 also confirms the relationships between the SNR and the expected measurement accuracy.

In order to investigate the long-term stability of the measurement system, more than 500 measurements of a step height artifact have been performed. Due to the large height differences, this is very sensitive to fluctuations in laser frequency. One measurement per minute was acquired. The results are shown in Figure 3.49. The standard deviation of the distance between two surfaces on the step height artifact is approximately 22 nm (absolute height difference is several mm). The results show

N	Invalid pixels	RMS error in nm	Out-liers	RMS error in nm (filtered)	Outliers (filtered)
8	19413	290	6904	231	0
10	19399	281	1010	219	0
12	19383	269	104	202	0
14	19380	261	45	190	0
16	19371	261	48	190	0

Table 3.4.: Measurement results of height estimation for a flatness standard; the absolute signal phase is not used (except for the reference measurement the results are compared to). For the filtered results, adaptive remapping with a 5×5 filter mask was used. This filter removed all outliers. The invalid pixels are mainly caused by data being unavailable in the corners of the field of view.

N	Invalid pixels	Standard deviation in nm (phase)	Bias in nm (phase)	Number of outliers (phase)	Number of outliers (phase, filtered)
8	19413	1.33	113	66436	2815
10	19399	0.96	83	53832	768
12	19383	0.81	60	36097	148
14	19380	0.61	18	25432	1
16	19371	-	-	24684	0

Table 3.5.: Measurement results of height estimation for a flatness standard; the absolute signal phase is used. This is an algorithm that can be used for smooth (but not necessarily continuous) surfaces only. For the filtered results, adaptive remapping with a 9×9 filter mask was used. Most outliers are removed by that filter. The standard deviation is obtained by comparison of the results to a reference obtained using the same method and therefore should be treated with care. It is clearly visible though that there is very little noise on the results (cf. Figure 3.33). The bias shows that unknown laser frequencies are problematic and that accurate monitoring is crucial.

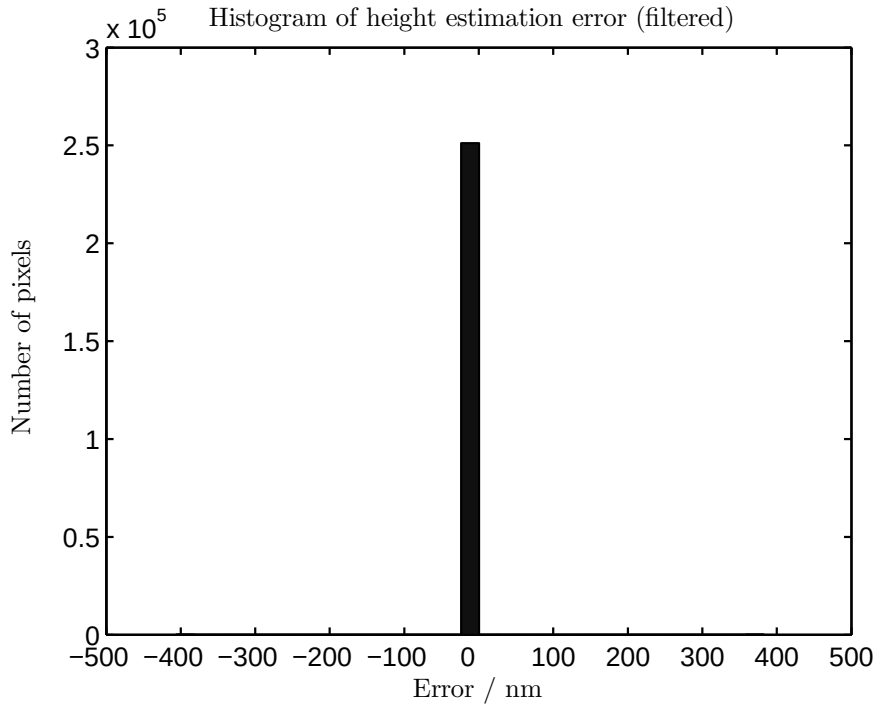


Figure 3.44.: Histogram of the estimation errors (N=16, SNR: 20, filtered). Filtering with an adaptive remapping filter with kernel size 9×9 reduces the side peaks (they are almost invisible in this graph), but it does not completely remove them. However, now the error is much lower, on the order of 1nm for the center peak as can be seen in Table 3.3.

some drift that is probably caused by temperature fluctuations.

Overall, these results show that the system is capable of accurate measurements of smooth surfaces with arbitrary geometry (i.e. step heights). The level of accuracy found here cannot be beaten easily with alternative measurement techniques (i.e. white-light interferometry), and measurement speed is expected to be significantly faster than competing methods. In the next section, the influence of rough surfaces on the measurement results is discussed.

3.11. Influence of Speckle

While it is important to know that the system works on smooth surfaces (with or without steps), the most interesting aspect of the system is its ability to measure rough surfaces.

Unfortunately, an analysis of the influence of surface roughness on the measurement results is difficult. There are three obvious ways to do that, but none of them is directly applicable in this case:

- The roughness of a surface can be determined based on a surface measure-

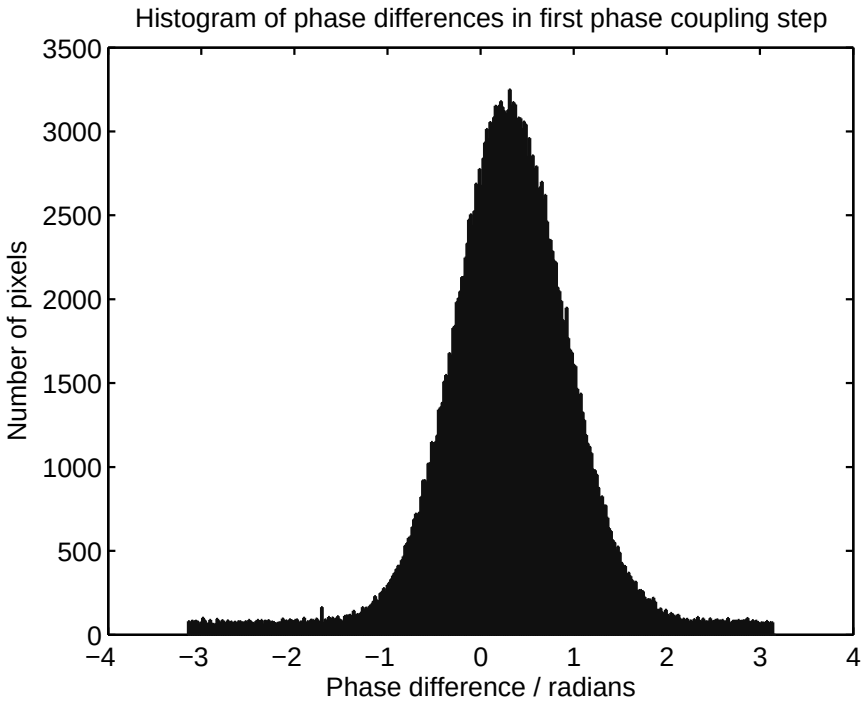


Figure 3.45.: Histogram of the phase differences in phase coupling for a measurement of a flatness standard with N=16.

ment and can then be compared to results from tactile reference measurements. While roughness standards and their parameters according to DIN are available, a comparison with results obtained by the new system fails due to different lateral resolutions. The new multiple wavelength system has a lateral resolution on the order of 40 microns; for roughness measurements according to DIN a measurement probe with 2 microns diameter has to be used. A comparison was attempted anyway, but the correlation was weak and the method is questionable.

- Results could be obtained by repeated measurements of a rough surface. But that does not yield an uncorrelated speckle field (if the object is not moved) or it introduces a movement of the object, which makes pixelwise repeatability measurements pointless as well. Due to system stability issues only a limited number of measurements is available, and there are plenty of other influences apart from speckle, including inhomogeneous illumination and vibration. Therefore there is no ground truth available and an isolation of the influence caused by speckle is not possible.

- Measurement results could be compared to theoretical predictions. However, most relevant optical parameters (e.g. aperture) and the surface parameters (for the given lateral resolution) are unknown; by adjusting these unknown values almost any simulation result can obtained, and therefore this approach cannot be used to analyze the influence of speckle.

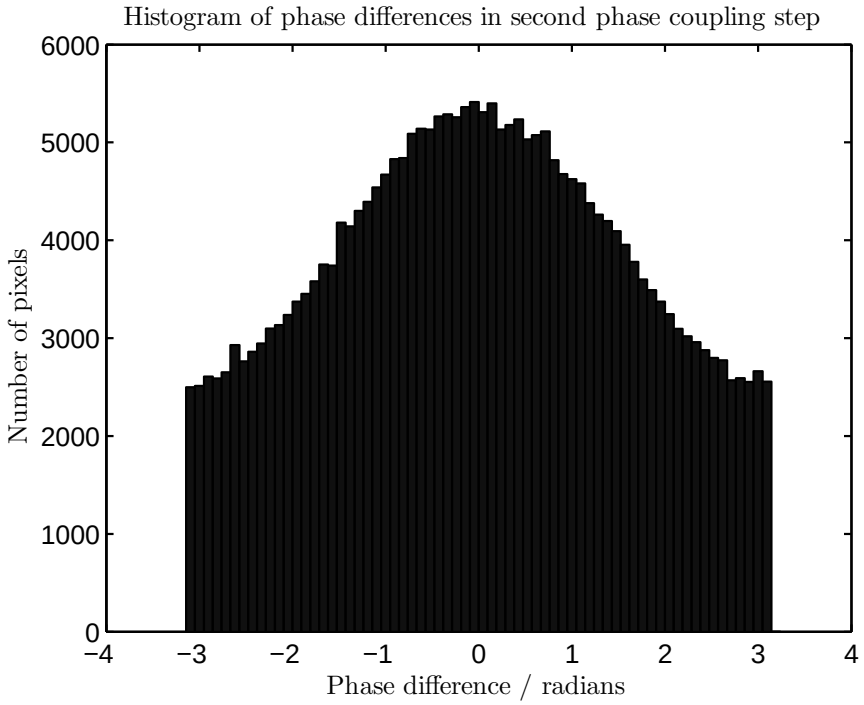


Figure 3.46.: Histogram of the phase differences in the second phase coupling step for a measurement of a flatness standard with N=16. This is needed to use the absolute signal phase.

For the reasons given above, the question of speckle in this system remains an open issue for future research. Nevertheless, an interesting approach for the analysis of speckle has been found and will be presented in this section. First, some basic properties of speckle in case of a change in wavelength as discussed in scientific literature are presented, and some results from related work in white-light interferometry are given. Next some qualitative results obtained from measurements are shown. An analysis based solely on intrinsic properties of the estimation algorithm is presented that could be used obtain some quantitative results for the influence of speckle on the accuracy of the results.

3.11.1. Theoretical Properties of Speckle

A derivation and a detailed analysis of the properties of laser speckle is outside the scope of this thesis. An extensive discussion can be found in Goodman's book [Goodman, 1975], and in a large number of papers. Some of these are mentioned and briefly summarized here, but this is not a complete review by any means. In [George & Jain, 1973] and [George et al., 1975], the issue of multiple wavelengths is discussed, and a threshold for decorrelation of the intensity is given. In [Pavlíček & Soubusta, 2003] the issue is discussed with respect to white-light interferometry, and [Hering, 2007] analyzes the influence of speckle in line scanning white-light interferometry, taking

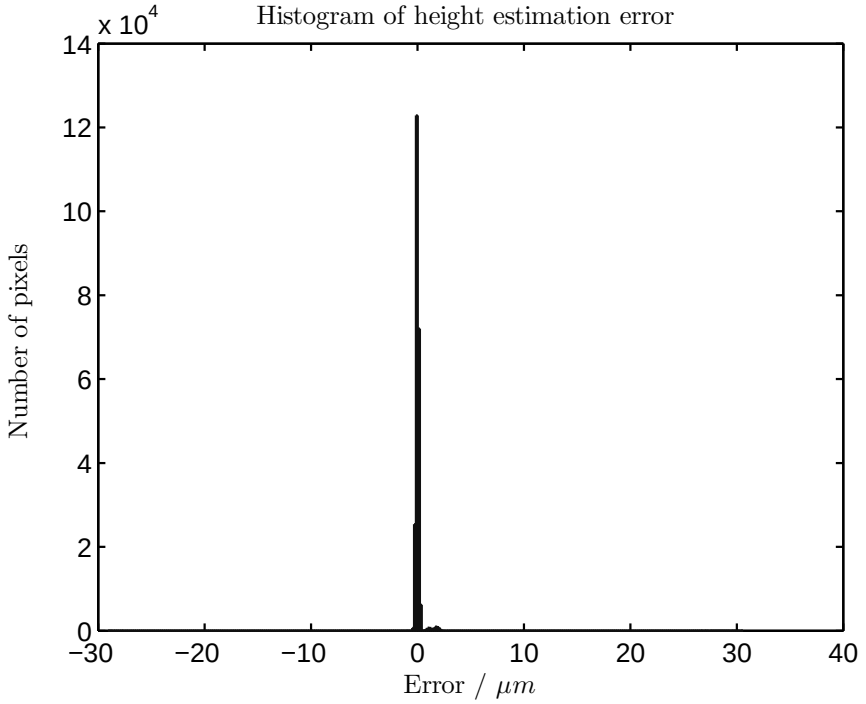


Figure 3.47.: Histogram of the estimation errors for a measurement of a flatness standard (N=16, unfiltered). One main and two side peaks are visible. The side peaks correspond to pixels where the phase coupling step chooses the wrong k. A measurement using strong spatial filtering and absolute phase evaluation was used as a reference.

second order statistics into account.

Only some simple results without proof or derivation are given here.

- In multiple wavelength interferometry, the statistical properties of the speckle field seen in every frame are relatively simple, as the light is monochromatic. The intensity distribution of a monochromatic speckle field can be described as a random walk in the complex plane and consequently follows a negative exponential distribution. This is confirmed by measurement results shown below, and illustrates the fact that the SNR for most pixels will be poor.

- An analysis based on first order statistics shows that the longitudinal uncertainty δ_z of the measurement results for a pixel i with modulation A_i in white-light interferometry is given by [Pavlíček & Soubusta, 2003]:

$$\delta_{z,i} = \frac{1}{\sqrt{2}} \sqrt{\frac{E_i[A_i]}{A_i}} \sigma_z \tag{3.43}$$

  The uncertainty is proportional to the RMS surface roughness σ_z and the relative modulation of the signal at pixel i and independent of the optical properties of

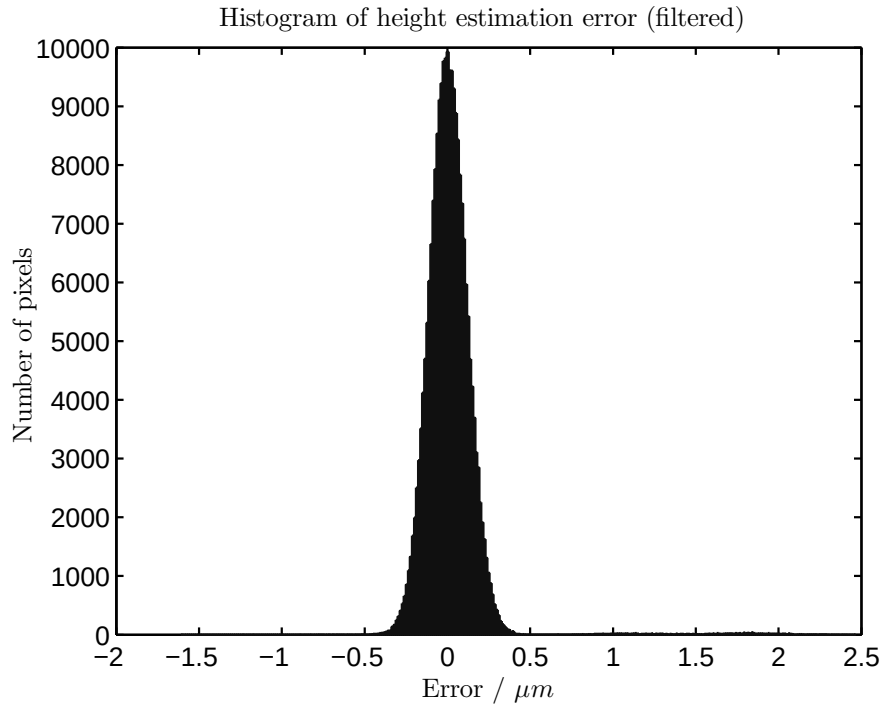


Figure 3.48.: Histogram of the estimation errors for a measurement of a flatness standard (N=16, filtered). Filtering with an adaptive remapping filter with kernel size 5×5 removed the side peaks.

the system (this will change if second order statistics are taken into account, cf. [Hering, 2007]).

- This assumption is only valid as long as

$$\sigma_z < \frac{l_c}{4}. \tag{3.44}$$

The coherence length l_c for a source with Gaussian spectrum and bandwidth $\Delta\nu$ is given by

$$l_c = \frac{c}{4\pi\Delta\nu}. \tag{3.45}$$

As this analysis is based on the phase slope, the same reasoning should be applicable to multiple wavelength interferometry.

- For multiple wavelengths, [George & Jain, 1973] derive the following threshold for decorrelation of the resulting speckle pattern:

$$\begin{aligned} \lambda_2 - \lambda_1 &\geq \frac{\lambda_0^2}{2\pi n_3 h_0}\sqrt{\frac{1-e^{-p^2h_0^2}}{1+(N-1)p^2h_0^2e^{-p^2h_0^2}}}, \\ \text{with } p &= \frac{2\pi n_3 h_0}{\lambda_0};\ \lambda_0 = \frac{\lambda_1+\lambda_2}{2} \end{aligned} \tag{3.46}$$

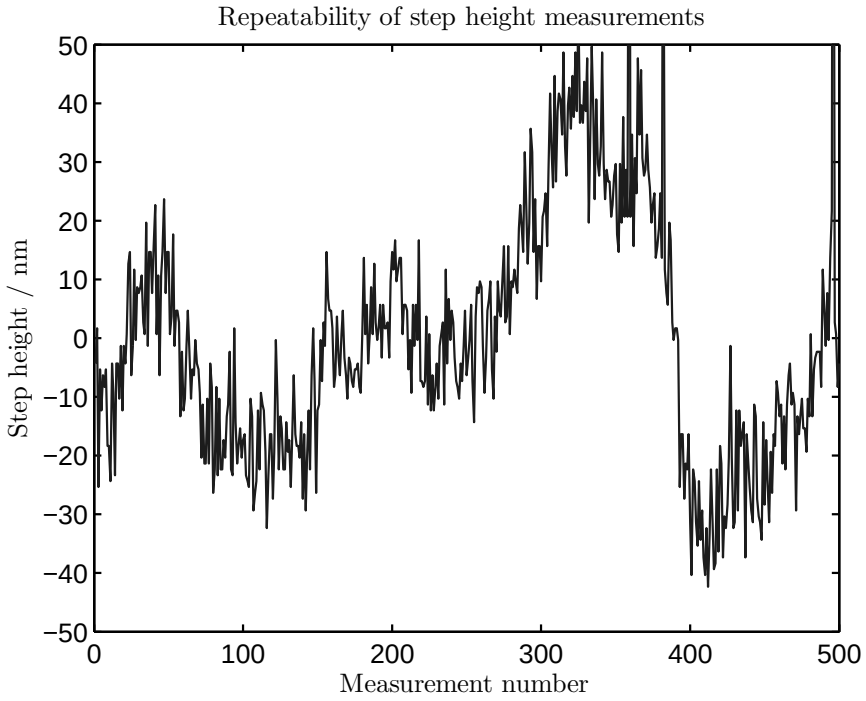


Figure 3.49.: Fluctuation in measured step distance for a step height artifact, based on 500 measurements. Average standard deviation is 22 nm, most of that probably caused by thermal drift.

This can be approximated asymptotically for a rough diffuser with $(ph_0)^2 >> 0$:

$$(\lambda_2 - \lambda_1) \geq \lambda_0^2/(2\pi n_3 h_0) \tag{3.47}$$

Applied to the multiple wavelength system discussed in this thesis, this indicates that within each of the blocks, the phase change in the speckle field is probably negligible, but that the frequency change between the blocks might lead to some issues.

The influence of surface roughness and speckle contrast on the accuracy is not surprising. In measurements using the algorithm described in section 3.5, a key property that can be measured is the expected change in the phase of the electro-magnetic field. Once this change exceeds π, there will be many outliers in the measurement results. This is therefore a natural criterion for decorrelation of the speckle field, and it is slightly different from the criteria used by Pavlicek and George. It may be possible to derive a closed form expression for that value, but this is subject to future research. In the next section, a method to determine this value experimentally is discussed.

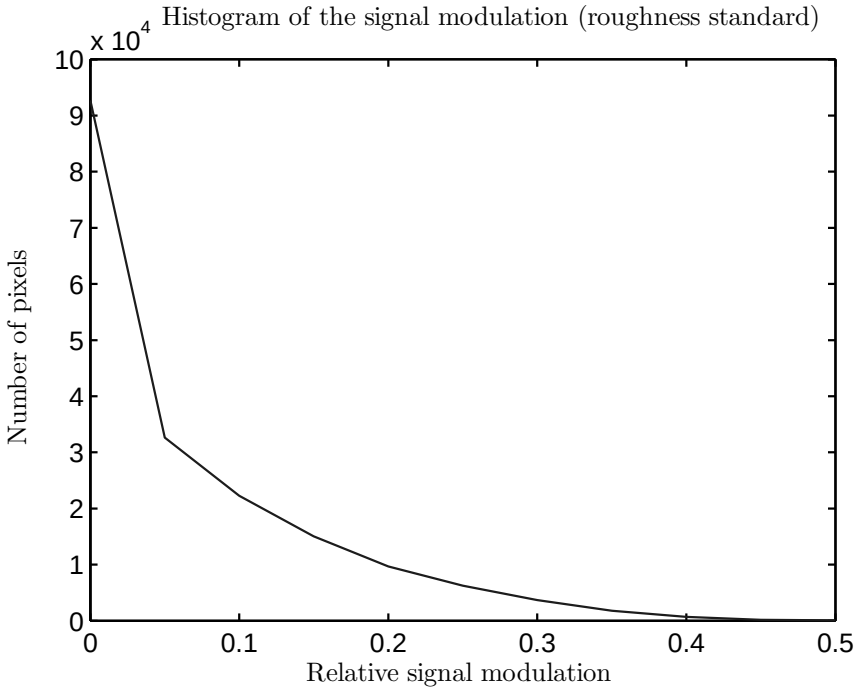


Figure 3.50.: This graph shows a histogram of the signal modulation for the measurement of a roughness standard, N=16 samples.

3.11.2. Influence of Laser Speckle on Phase Coupling

As described above, it is not possible to directly determine the influence of the speckle field by repeated measurements. However, the influence of speckle can be approximated based on a detailed analysis of an individual measurement, using theoretical results from previous sections.

First of all, the intensity distribution for measurements of rough surfaces follows the one expected from the theoretical analysis. This is shown in Figure 3.50. A corresponding result for smooth surfaces is shown in Figure 3.28.

The signal standard deviation from repeated measurements can be determined and the relationship between signal modulation and measurement error as well as between phase coupling error and measurement error can be shown (Figure 3.51 and Figure 3.52). It has to be noted that the meaning of this standard deviation is questionable — on the one hand, some of the differences are caused by a slight movement of the object, and on the other hand, the speckle field is not completely independent between the measurements.

These results show that the adaptive remapping method for filtering is promising in this case, too. This filtering can be applied as long as the surfaces roughness (peak to peak) within the size of the filter mask is smaller than the first step phase coupling error (in our case approximately 30 microns). This is true for many technical surfaces.

Additionally, the phase coupling histogram is an indication whether a surface can be

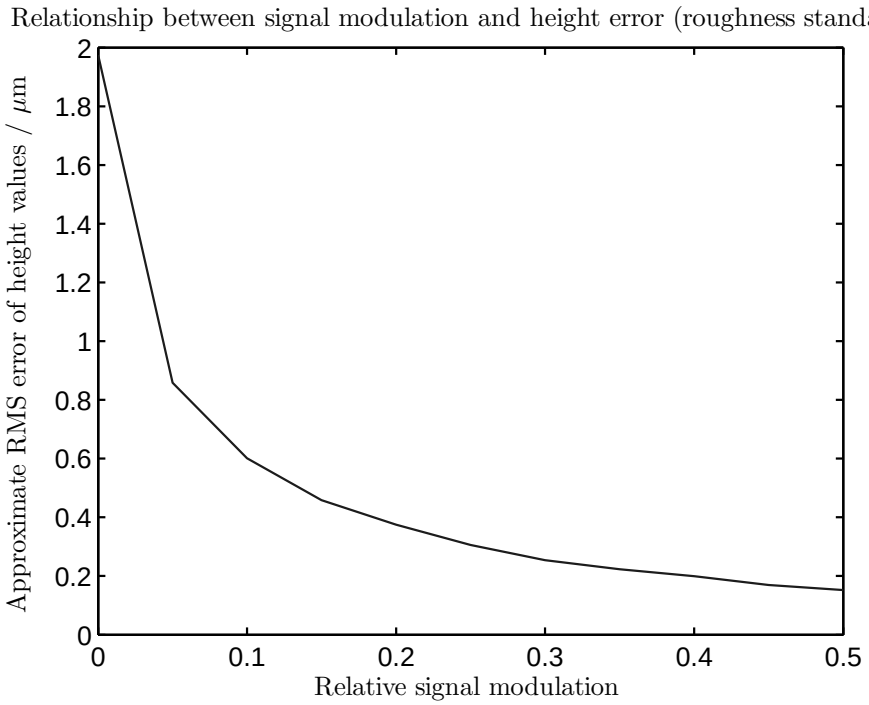


Figure 3.51.: This graph shows the relationship between signal modulation (determined by a least squares fit) and the RMS error of the height values (estimated by repeated measurements of a roughness standard), N=16 samples.

measured or not. As long as there is a peak visible in this histogram, the phase values for the two blocks are correlated. For the available roughness standards (R_a between 0.21 and 1.7; R_z between 1.2 and 10.3 according to tactile measurements) this was the case. The intensity distribution and the standard deviations above are given only for the smoothest one of these (with a mean modulation of less than 0.1), for the others the mean modulation was approximately 0.03 as the reflectivity was too low.

When using the two block algorithm presented in this thesis, the influence of speckle on the phase and frequency estimate of the individual blocks can be neglected. This follows from equation 3.47 above: This phase change should only play a role when the surface is so rough that interferometric measurements are not likely to be useful anyway, but not for technical surfaces with a roughness on the order of a wavelength.

Visual inspection of the residuals confirms this notion. Therefore one can expect the accuracy of the phase and frequency estimates of the individual blocks in the presence of speckle to be approximately identical to the accuracy obtained for these blocks when a smooth surface is measured, as long as the SNR is comparable (in case of speckle, the intensity distribution is very different, as shown above).

This in turn implies that the histograms of the phase differences in the coupling step should look similar. The phase difference consists of errors caused by additive noise and systematic errors of the frequency and phase estimation algorithms — and

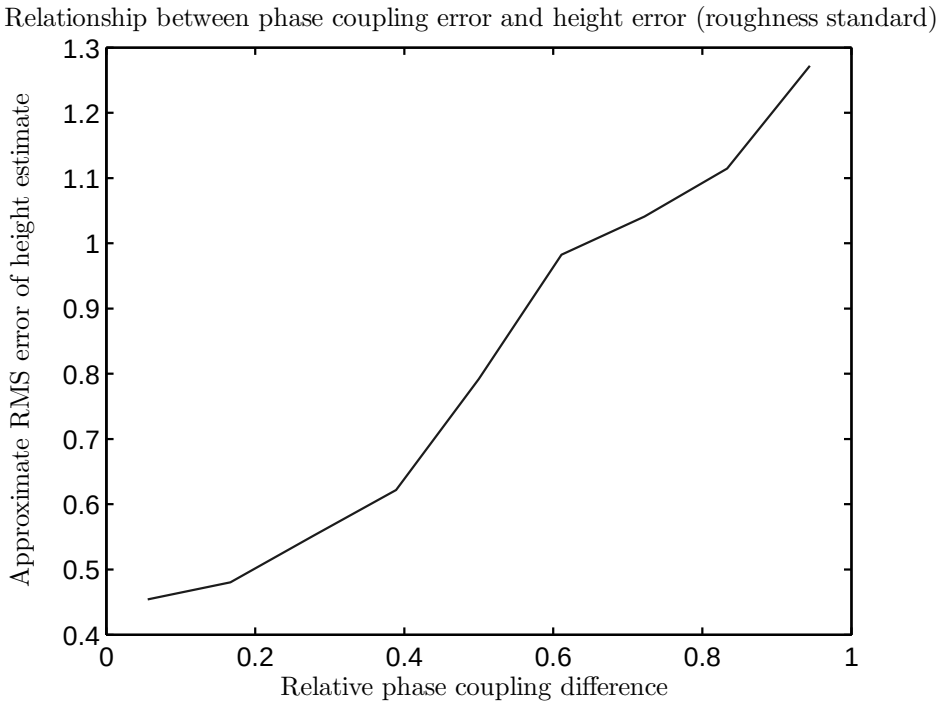


Figure 3.52.: This graph shows the relationship between the phase coupling error (using two blocks of data and least squares phase estimation) and the RMS error of the height values (estimated by repeated measurements of a roughness standard), N=16 samples.

an error caused by the phase change due to the variation of the speckle field when the laser frequency is changed.

The standard deviation of this phase coupling distribution can be determined for various values of the signal modulation, for both a smooth and a rough object. At the same signal modulation, the difference between the rough and the smooth object can then be attributed to a phase change: the increase in standard deviation corresponds to the average phase change of the speckle field when the laser frequency is changed from the first block to the second.

Unfortunately there is almost no overlap in signal modulation between the two measurements analyzed here, as their reflectivity is quite different and the system cannot be adjusted to that. If possible, such a measurement should be performed with both standards having similar reflectivity and occupying half the available field of view at approximately the same distance, such that errors in laser tuning are identical for both. The expected phase change in case of the rough object is fairly small, therefore the measurement noise needs to be very low. This is subject to further research.

Once the phase difference introduced by the speckle field is known, the influence on measurement accuracy when using the two block algorithm can be derived easily:

$$\Delta_h = \frac{\Delta_\phi}{\pi} \cdot \frac{s}{b} \tag{3.48}$$

with block distance b (in frequency increments) and ambiguity interval size s. This could be rewritten to only use the laser frequencies, but for most applications the equation above is more convenient.

For the laser parameters discussed so far, with a block distance of $b = 112$ and ambiguity interval size $s = 1.6$mm, a phase change of $\Delta_\phi = \pi/40$ (corresponding to a difference of about 10nm in phase unwrapping) leads to a height error of about 350nm, so this can quickly lead to a significant error.

These results illustrate three main points:

- The error can be broken down into an error due to lowered signal contrast and an error due to a phase change. Both can be determined. Lowered signal contrast can easily be taken into account by appropriate filtering; the phase error is more tricky, but outliers can be reduced by filtering quite well. It is clearly visible that the standard deviation — at the same noise level — is significantly higher for a rough surface than for a smooth surface. This is caused by changes in the speckle field.

- This analysis also shows that the algorithm is applicable as long as a peak is visible in the difference distribution, at least for pixels with higher confidence. These measurements proof that the measurement principle and the multi-block algorithm are applicable to measurements of rough surfaces. The optimum block distance might be smaller than for smooth surfaces though.

- The phase change combined with poorer accuracy of the estimates caused by the lower intensity (which follows a negative exponential distribution) leads to lower accuracy and therefore indicates that stronger filtering would be useful. Unfortunately, though, stronger spatial filtering automatically assumes surface smoothness, and is therefore not the correct approach for rough surfaces. Methods such as Gaussian filtering are not very helpful. The remapping procedure might still be applicable though, but this depends on the surface properties. The most useful approach in practice is adjusting the number and the distance of the wavelengths used such that the phase coupling histogram is narrow enough to keep the number of outliers low enough for the remapping procedure (at the maximum acceptable filter size, which in turn depends on the size of the surface areas where the height differences do not exceed the coupling error). These parameter values can be obtained by visual inspection of the results.

It has been shown that rough surfaces can be measured and that the influence of laser speckle can be determined. The total error due to laser speckle consists of a phase error and an increase in relative noise due to lower SNR. A more detailed investigation of that aspect of the system is a topic for further research.

4. Further Applications for the Derived Algorithms

There are multiple applications where the algorithms for phase and frequency estimation derived for multiple wavelength interferometry can be used.

The frequency estimation part alone can be used whenever fast frequency estimation from short blocks of data is required. Due to the optimization procedure involved, the properties of the resulting estimator will be known very well, which is useful for many applications. The simplification for phase estimation discussed in chapter 3.6.5 can be applied to a significant number of problems, too: Among others, it is currently being used for surface reconstruction by fringe reflection methods (which will not be discussed in this thesis though).

As a group member is working in the field of polarization imaging, the author applied these results in this context, and it turned out to be possible to use both frequency and phase estimation successfully in that field. Therefore this application is given as an example for use of the algorithms outside the field of frequency scanning interferometry. Optical setup and measurements for the following section were performed by Thomas Geiler, while the analysis of the data and the discussion of the algorithms has been done by the author. Several figures in this chapter have been used previously in this thesis, but will be repeated here (with captions tailored to polarization imaging) in order to keep this chapter easily readable.

4.1. Polarization Imaging

4.1.1. Introduction

Polarization imaging can be used for image and contrast enhancement and material distinction as well as surface reconstruction. It works particularly well for many diffuse scattering surfaces that are very difficult objects for conventional imaging. Therefore in recent years polarization has become an important approach to expand capabilities of computer vision systems.

In contrast to classic metrology techniques such as spectroscopic ellipsometry or single wavelength ellipsometry which are mainly used for film and material characterization, the methods in computer vision usually avoid the typical calibration procedures for linear polarizers by using just one of them.

It has been shown that illuminating the surface of dielectrics or metals with unpolarized light leads to partial polarization for specular and diffuse reflection [Wolff & Boult, 1991]. Phase shifts caused by reflection cannot be measured with a single polarizing element, but in most cases the differences of reflectance between parallel and

perpendicular polarization components can be used for a pixelwise estimation of the surface orientation or for material classification.

Existing interpretation concepts for polarization imaging need to determine three independent polarization parameters. Degree of polarization is a measure of the modulation of the intensity signal, the phase contains the angle of maximum intensity and finally the mean intensity is needed for a full signal description.

For the interpretation of phase, degree of polarization and intensity a stable reflection model is needed which is robust to changes in microstructure and surface roughness. Today there are approaches for different materials, using special illumination–camera configurations, for example shape-from-diffuse-polarization [Atkinson & Hancock, 2006], shape-from-specular-polarization [Rahmann & Canterakis, 2001] and surface reconstruction of transparent objects [Miyazaki & Ikeuchi, 2005].

Speed and accuracy of the image acquisition step are key aspects for optimizing the performance of polarization imaging. A mechanically rotating linear polarizer, a system using a beam splitter [Wolff, 1994] or a system using an electrically adjustable polarizing element (e.g PLZT or liquid crystal mounted cameras) can be used. Three different polarizer positions are, in principle, sufficient in order to determine the three unknowns phase, degree of polarization and mean intensity. But the resulting polarization images suffer from noise, and this leads to performance degradation especially for shape from polarization algorithms. Therefore a widely used approach to improve the quality is based on blurring the raw images in order to reduce noise and the influence of image shifts due to polarizer movement.

Here the focus lies on the comparison of different evaluation methods for the optimal interpretation of polarization series. Series acquisition conditions are analyzed concerning number, accuracy and position of angular sampling points for the polarizer.

There is potential for a significant speed-up and cost reduction of polarization imaging, which allows the use of this technique in a wider range of industrial and scientific applications. The speed advantage is reached by optimization of the algorithms and fast, unsynchronized image acquisition. Cost can be reduced because of dramatically reduced requirements on the accuracy of the stage moving the polarizer.

The measurement and processing time, accuracy and sensitivity to noise is presented and compared for a number of different setups. The results are applicable to the classical mechanical rotating polarizer setup as well as to the liquid crystal methods as there is no difference in the signal model. The polarizing element can be placed in front of the camera or in front of the light source.

4.1.2. Polarization and Reflection

Propagation and interaction of light can be described using Maxwell's equations. As there is no analytical solution for this set of coupled differential equations for the general case, light scattering has to be treated using numerical approaches. For reflection or transmission at a plane surface, Maxwell's equations reduce to boundary conditions. Fresnel's equations fulfill these conditions, and therefore light reflection and transmission can be treated easily in this case. Reflection coefficients for the amplitude of the

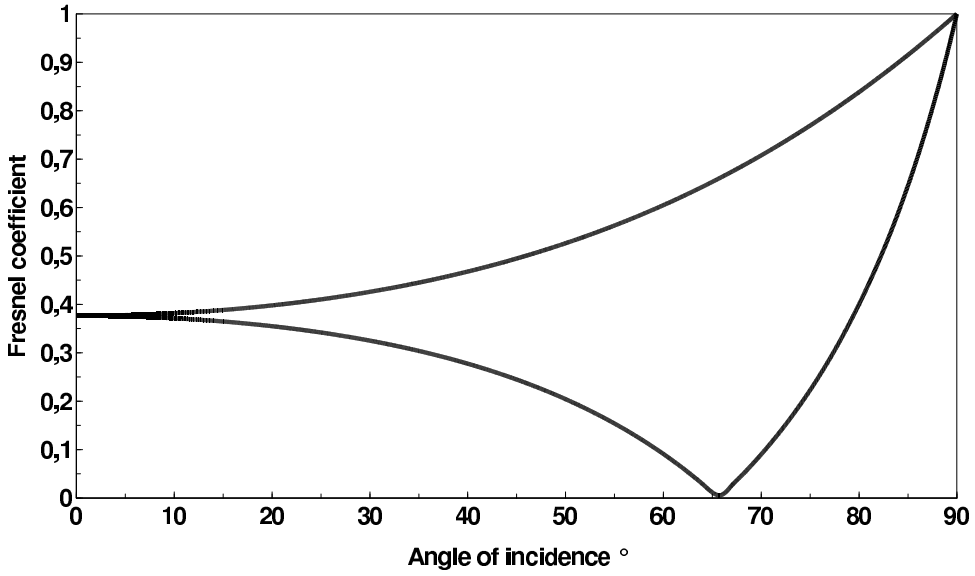


Figure 4.1.: Fresnel reflection coefficient for ZrO_2 [SOPRA, 2005] at $\lambda = 633nm$, $n = 2.21$.

electrical field are given by [Born & Wolf, 1999]:

$$r_{TM} = \frac{n_2 \cos(\theta_1) - n_1 \cos(\theta_2)}{n_1 \cos(\theta_2) + n_2 \cos(\theta_1)} \tag{4.1}$$

$$r_{TE} = \frac{n_1 \cos(\theta_1) - n_2 \cos(\theta_2)}{n_1 \cos(\theta_1) + n_2 \cos(\theta_2)} \tag{4.2}$$

$$t_{TM} = \frac{2n_1 \cos(\theta_1)}{n_1 \cos(\theta_2) + n_2 \cos(\theta_1)} \tag{4.3}$$

$$t_{TE} = \frac{2n_1 \cos(\theta_1)}{n_1 \cos(\theta_1) + n_2 \cos(\theta_2)} \tag{4.4}$$

with angle of incidence θ_1, transmission angle θ_2 given by Snell's law, n_1 and n_2 complex refractive index of medium 1 and 2, respectively.

The common nomenclature for the transversal electric (TE) and the transversal magnetic (TM) mode is used. For the TE component the electrical field is perpendicular to the plane of incidence (defined by the incoming and the reflected ray). Usually, reflection and transmission coefficients are different for both components (Figure 4.1). Therefore light becomes polarized upon reflection or transmission.

Atkinson and Hancock introduced a reflection model [Atkinson & Hancock, 2006] that can be used for surface reconstruction based on diffuse reflection on ceramic materials.

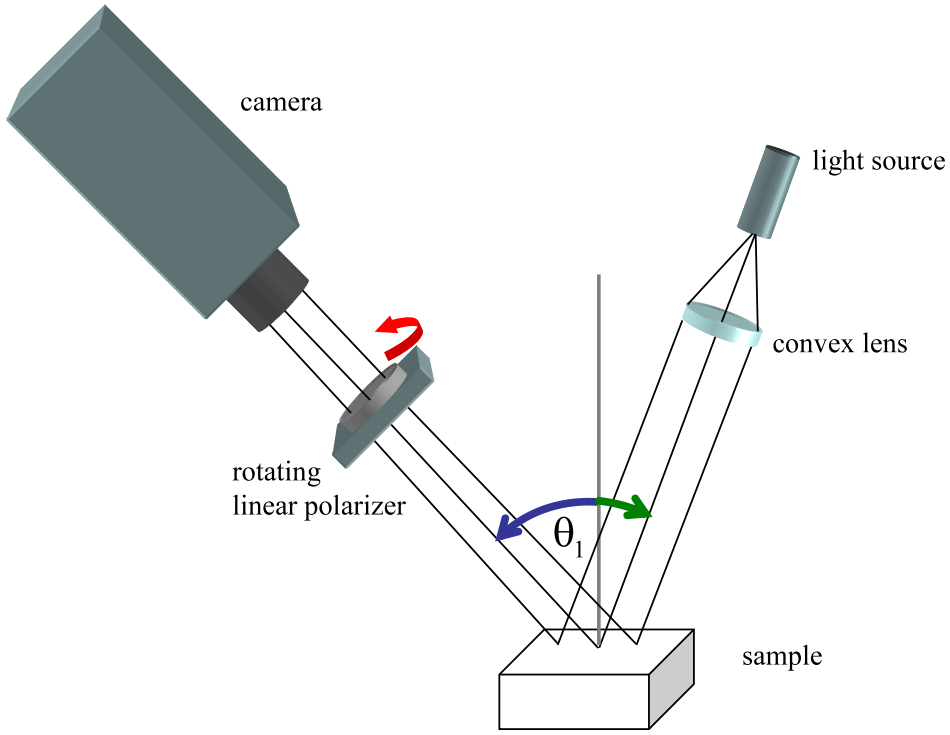


Figure 4.2.: Optical setup for shape-from-polarization

4.1.3. Polarization Imaging

Polarization parameters

Rotating a polarizer in front of a camera causes the light intensity to change if the light is at least partially polarized. The maximum and minimum observable intensities are denoted I_{max} and I_{min}. The degenerate degree of polarization ρ is defined as

$$\rho = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{4.5}$$

ρ is called "degenerate degree of polarization" because elliptical polarization cannot be treated with a setup using one linear polarizer only. For $\rho = 1$, the light is completely linearly polarized. The observed intensity changes when the polarizer is rotated, and the angle for which the highest intensity occurs can be determined. The degree and angle of polarization can be measured for each pixel. For a complete characterization the mean intensity is determined, too.

Optical Setup

For the development of the algorithms image sequences obtained with a shape-from-polarization setup as shown in Figure 4.2 are used. The algorithms can also be applied to other polarization techniques as there are no special restrictions.

The ceramic sample is illuminated by a halogen light source. A convex lens with a focal length of 80mm is used to shape the beam and obtain uniform illumination. The polarizer can be rotated using a high precision step motor. To improve the signal to noise ratio, one can change the shutter time of the camera and merge a series of radiometrically corrected images to obtain a highly dynamic, linear, low noise image.

Signal model

The interaction of light with the object surface and the linear polarizer can be described by the Stokes formalism [Walker, 1954; Azzam & Bashara, 1987]. The Stokes vector of the incident light has to be multiplied with the Mueller matrices of all elements in the optical path. The Mueller matrix for a plane surface is given by:

$$M = \begin{bmatrix} c_1 & c_2 & 0 & 0 \\ c_2 & c_1 & 0 & 0 \\ 0 & 0 & \mathrm{Re}(r_{TE} r_{TM}^*) & 0 \\ 0 & 0 & \mathrm{Im}(r_{TE} r_{TM}^*) & \mathrm{Re}(r_{TE} r_{TM}^*) \end{bmatrix} \tag{4.6}$$

$$c_1 = \frac{|r_{TE}|^2 + |r_{TM}|^2}{2}$$
$$c_2 = \frac{|r_{TE}|^2 - |r_{TM}|^2}{2}$$

The rotational matrix is given by:

$$T(2\varphi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(2\varphi) & \sin(2\varphi) & 0 \\ 0 & -\sin(2\varphi) & \cos(2\varphi) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.7}$$

The Mueller matrix of a rotated linear polarizer is given by:

$$P(\theta) = \frac{1}{2} \begin{bmatrix} 1 & \cos(2\theta) & \sin(2\theta) & 0 \\ \cos(2\theta) & \cos^2(2\theta) & \frac{1}{2}\sin(4\theta) & 0 \\ \sin(2\theta) & \frac{1}{2}\sin(4\theta) & \sin^2(2\theta) & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.8}$$

Using these results one obtains the intensity

$$S_{out} = P(\theta) T(-2\theta_0) M T(2\theta_0) \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.9}$$

And this yields the following signal intensity

$$I(\theta) = c_1 + c_2 \cos\left(2(\theta - \theta_0)\right) \tag{4.10}$$

Sampling	Uniform	Approximately uniform	Arbitrary
Known	Table 2	fast: see Table 2, accurate: LLS (C)	LLS (C), others (D)
Unknown	FFT (E) + Table 2	fast: FFT (E) + Table 2, accurate: FFT (E) + (F) + LLS (C)	(G)

Table 4.1.: Possible algorithms for various sampling patterns

Sampling	Special angles	Periodic	Any phase difference
Known + Uniform	N-bucket algorithms (A)	Fourier transform / QAM demodulation (B) or LLS (C)	LLS (C)

Table 4.2.: Possible algorithms for known, uniform sampling

The intensity can be rewritten to obtain the following equation that is well known from frequency and phase estimation theory. The samples t_i correspond to the chosen angular positions of the polarizer.

$$I(t_i) = A\cos(\omega t_i + \varphi) + C \tag{4.11}$$

Based on the measured intensity data the phase, modulation and mean have to be determined for further processing. As the signal model is a sinusoid, there are many possibilities for determining phase and amplitude. An overview of the possibilities discussed in this paper is given in Tables 1 and 2. Iterative curve fitting procedures have also been used sometimes [Brown & Mao Wang, 2002]. They offer more flexibility, but are significantly slower than the algorithms discussed below, and they do not necessarily offer better performance.

Much literature is available on the topics of phase and frequency estimation. This paper does not attempt to give a comprehensive review, but focuses on fast and robust estimation techniques and their application to polarization imaging. Most of the algorithms could be replaced by others, for a more in-depth discussion the reader is referred to section 3.6 and [Quinn & Hannan, 2001; Moon & Stirling, 2000; Oppenheim & Schafer, 1989; Poor, 1994], where the frequency and phase estimation problems are discussed in more detail. For reference a theoretical lower bound (the Cramér-Rao bound) is included in the graphs, and in many cases it shows that the algorithms derived in the following chapter are close to this theoretical limit.

4.1.4. Signal Processing Algorithms

The algorithms mentioned in Table 1 and 2 will be described briefly in this part:

N-bucket algorithms

This class of algorithms uses certain well-defined angles for the phase shifter in order to obtain simple closed form expressions for the phase, for example using $60°$ or $90°$

increments. The algorithms have been used for amplitude estimation in the context of white-light interferometry, but here in polarization vision they can be used for their original purpose, phase estimation. The following two algorithms have been used here:

$$\varphi = \arctan\left(\frac{I_3 - I_1}{I_1 - 2I_2 + I_3}\right) \tag{4.12}$$

and

$$\varphi = \arctan\left(\frac{2(I_4 - I_2)}{I_1 - 2I_3 + I_5}\right) \tag{4.13}$$

They are discussed in more detail in chapter 2.1.

Fourier transform / QAM demodulation

For a simple sinusoidal signal model with an integer number of wavelengths on the uniformly sampled support, one can compute the Fourier coefficient at the known signal frequency, and take its angle. This is identical to a demodulation by correlating with a sine and a cosine and determining the phase based on the quotient (QAM demodulation, frequently used in communications, e.g. [Rice et al., 2001]).

Linear least squares phase estimation

The phase estimation problem can be rewritten

$$I(t) = A_1 \cos(\omega t) + A_2 \sin(\omega t) + C \tag{4.14}$$

with the relation to eq. 4.11 given by

$$\varphi = \arctan\left(\frac{A_1}{A_2}\right) \text{ and } A = \sqrt{A_1^2 + A_2^2} \tag{4.15}$$

Estimating A_1, A_2 and C is obviously a linear problem if the frequency and the sampling points are known, and there is a simple solution that can be applied to uniform and non-uniform sampling. Written in vector form for a discrete number of samples the following set of equations results:

$$\begin{aligned} &Ax = b, \text{ with} \\ &A = \begin{bmatrix} \cos(\omega t_1) & \sin(\omega t_1) & 1 \\ \cos(\omega t_2) & \sin(\omega t_2) & 1 \\ \cdots & \cdots & \cdots \end{bmatrix}, \\ &x = [A_1\ A_2\ C]^T \text{ and } b = \vec{I}. \end{aligned} \tag{4.16}$$

In the presence of noise there is no exact solution to this system of equations, and one way of dealing with that problem is searching for the solution with the minimum mean squared error $||Ax - b||^2$. This solution is well known as linear least squares estimation:

$$A^T Ax = A^T b, \text{ or } x = (A^T A)^{-1} A^T b \tag{4.17}$$

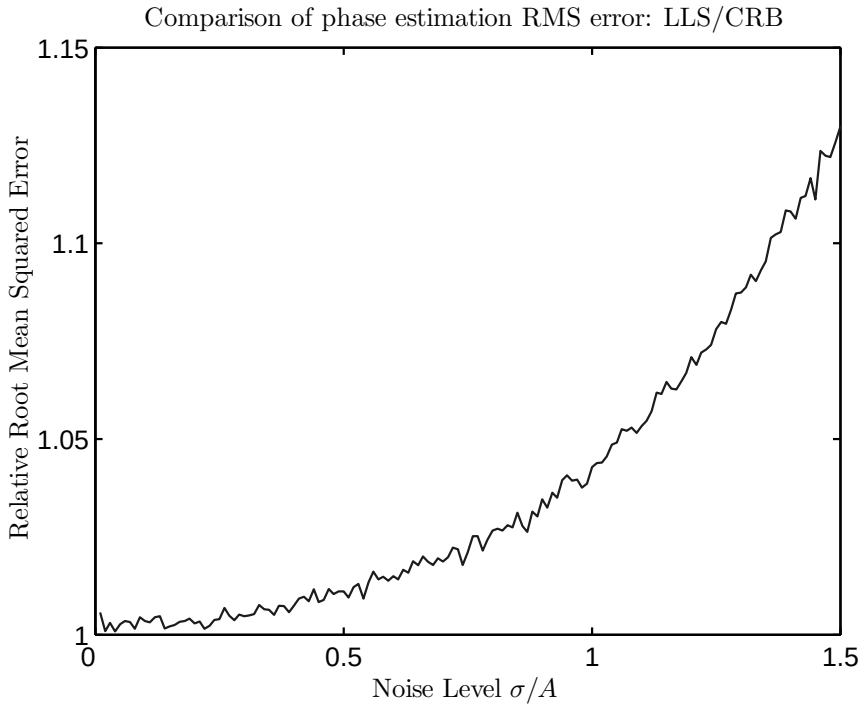


Figure 4.3.: Relative accuracy of linear least squares phase estimation, averaged across all phases and a large range of frequencies (excluding only frequencies around zero and the Nyquist frequency), compared to the Cramér-Rao bound. For the graph the frequency was assumed to be known, and N=16 samples were used. "Noise level" is the standard deviation of the Gaussian noise σ divided by the signal modulation A.

$M = (A^T A)^{-1} A^T$ needs to be computed only once, as the signal frequency is identical for all pixels. For phase estimation a multiplication of this $3 \times N$ matrix with the vector b for every pixel along N frames is needed, i.e. three multiplications for each pixel and frame. The properties of this algorithm for phase estimation have been discussed in detail in the literature [So, 2005]. This solution can be implemented quickly and robustly, and it shows very good performance when compared to the theoretical limit, the Cramér-Rao Lower Bound (CRB), as can be seen in Figure 4.3. If the signal amplitude is larger than the noise standard deviation, the variance of this estimate is less than 5% above the theoretical limit. Results are identical to the results from (B) if an integer number of signal periods is sampled (or if $N \to \infty$).

Other phase estimation algorithms

Other algorithms for phase estimation are available and many of these could be applied to polarization vision. A full review is outside the scope of this paper, though.

Frequency estimation

If the samples are known to be uniformly sampled, but the sampling distance (i.e. the velocity of the rotating stage for the polarizer) is unknown, one can reconstruct it from the data using the assumption that the signal is a sinusoid with identical frequency for all pixels. This is a frequency estimation problem, and here the algorithm described in section 3.6 is used. For a very low number of samples this is inaccurate, but it does work well in practice with eight or more samples as will be shown below. This still holds true if the sampling is only approximately uniform: One can still determine an average frequency as long as the difference does not exceed a few degrees or only affects a few of the samples. The sensitivity of the phase estimation to an error in the frequency estimate and limitations of this method are discussed below.

Sampling position estimation

When a rough estimate of the sampling positions is available e.g. from the frequency estimation above, one can obtain an estimate for the angle of the polarizer at every frame by first determining the phase for the whole time series for every pixel, and then looking at the phase difference of every single pixel from the estimated signal model. These differences can then be averaged across the whole frame in order to obtain a mean phase deviation for every frame and thus the real position of the polarizer. This only works well in the presence of low noise and fairly uniform polarization angle distribution across the field of view, is computationally more expensive and not very precise.

Direct phase difference estimation

Without estimating the signal frequency, one can try and obtain an estimate of the phase differences directly. This works by associating each intensity value with a phase value (based on a coarse modulation and offset estimate), computing the two possible phase differences (taking into account ambiguities caused by the unknown initial phase), and then looking at the resulting distribution of differences across the whole field of view. When using the mean of the estimated phase differences, the result requires a uniform distribution of the original phase — which is usually not the case. If the mode of the distribution is used instead, the result is independent of the distribution of the phase values, but becomes highly sensitive to noise. Figure 4.4 illustrates these properties. This has not been investigated in detail as its practical relevance is limited.

4.1.5. Experiments

Polarization measurements were performed using a ceramic object. An analysis of all algorithms and setups above has been performed. Two cases seem to be most relevant for practical applications as they are easy to implement in practice and offer the best accuracy:

- Known uniform sampling (i.e. using a synchronized camera and a step motor to rotate the polarization filter to known positions).

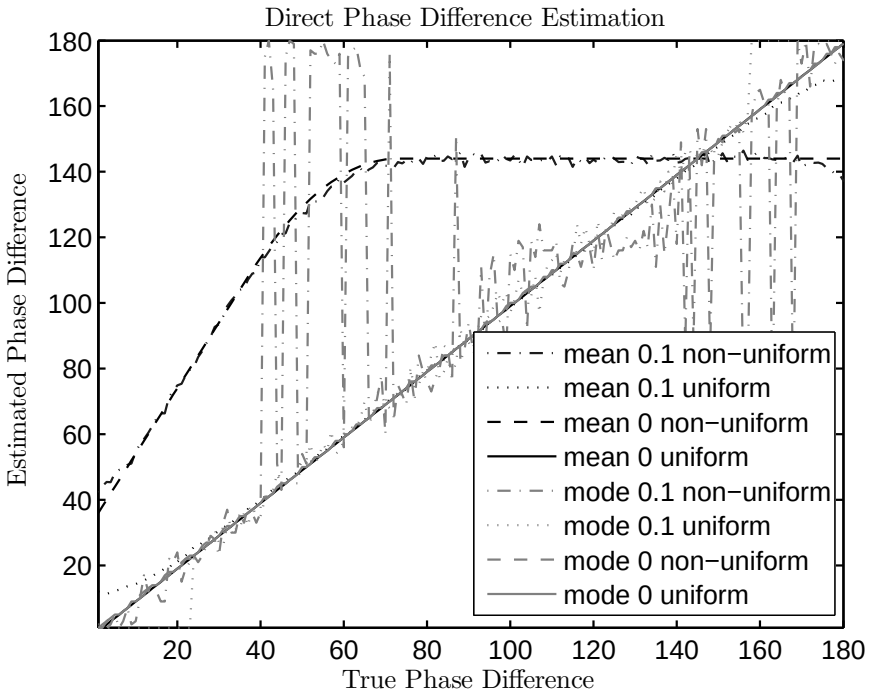


Figure 4.4.: Direct phase difference estimation accuracy for differences between $0°$ and $180°$. All possible combinations of uniform and non-uniform distribution of the phase across the field of view, with and without additive noise of relative standard deviation $\sigma/A = 0.1$, using mean or mode of the distribution as an estimate are shown. For the case of non-uniform phase distribution, the phase was chosen to be uniformly distributed between 0.3π and 0.4π.

- Unknown uniform sampling (i.e. using a rotating polarization filter and a free-running camera for reduced hardware cost and faster acquisition times).

In addition to these two configurations, a brief analysis will be presented for both known and unknown non-uniform sampling. These configurations are described in more detail in the subsections below. All results are compared to a reference measurement using 120 samples with $3°$ distance from each other. This reference data was analyzed using the least squares approach (C). In order to illustrate the robustness of the algorithms used, a "worst case" scenario for polarization image acquisition was included as well: A series of images was acquired using a hand-held power drill with an affixed polarization foil, combined with a free running camera. No camera or setup calibration was used, and no precautions at all were taken to reduce wobbling or other sources of errors. Therefore there is significant additive noise as well as sampling jitter, and the true signal frequency is unknown. No reference measurement for that part exists, therefore no quantitative results can be given.

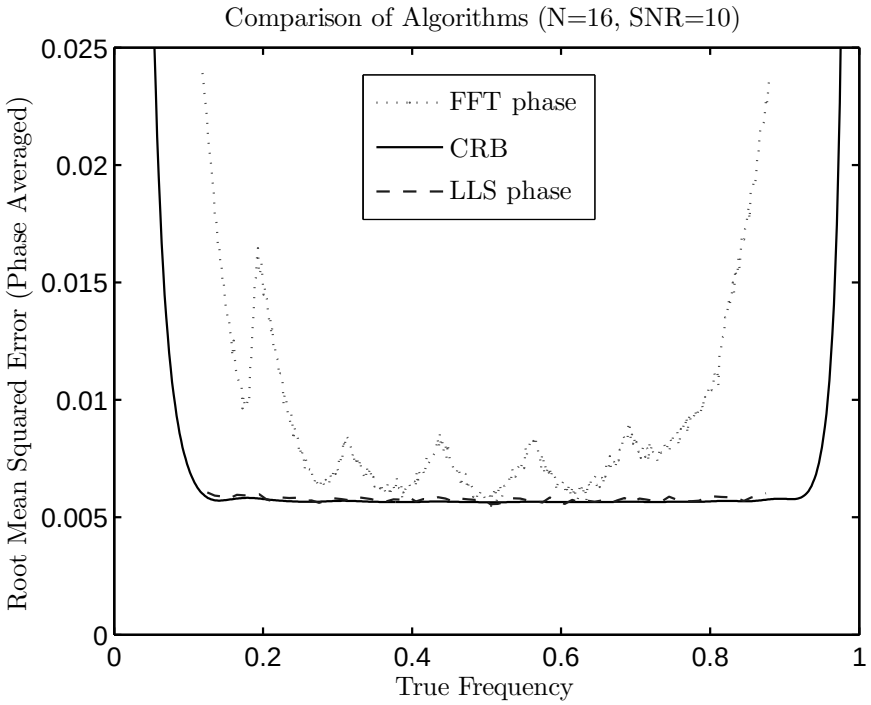


Figure 4.5.: Accuracy of phase estimation for various frequencies, averaged across all phases (N=16 samples, SNR 10). The LLS estimator almost reaches the theoretical limit, the CRB. An FFT based estimator shows much poorer performance in this case.

Known and unknown uniform sampling

The commonly used three frame algorithm is obviously the simplest and fastest option. A higher accuracy can be obtained by increasing the number of frames, though. For this purpose some selected N-bucket algorithms and linear least squares estimation are compared. For that purpose, the accuracy is normalized by the square root of the number of frames such that improvements that could be obtained by simply repeating the measurement are taken into account. The accuracy of phase estimation asymptotically reaches the following limit for the relative standard deviation [Rife & Boorstyn, 1974]:

$$s_\varphi = \frac{\sqrt{2}}{2\pi} \cdot \frac{\sigma}{A} \cdot \frac{1}{\sqrt{N}} \tag{4.18}$$

N corresponds to the number of images taken, σ is the noise level (additive white Gaussian noise assumed) and A the modulation of the signal. For a low number of images, this is not exact and the accuracy depends on both the signal frequency (which can be adjusted) and the true phase of the signal. This is shown in Figure 4.6 for 16 images and all possible phase and frequency values. The colormap is centered on the result from equation 4.18. The scale is truncated in the white areas where

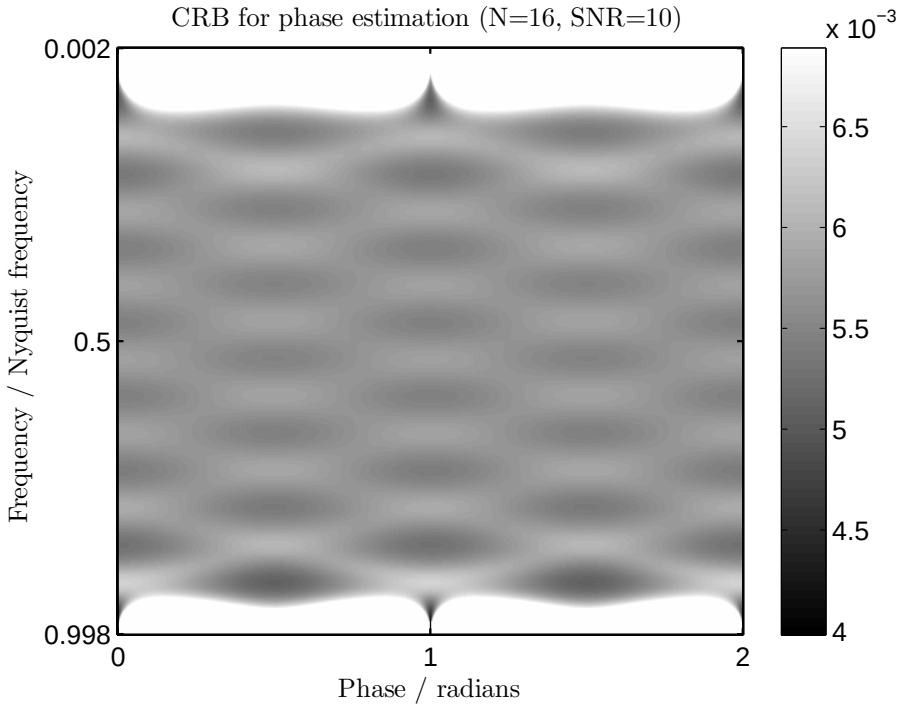


Figure 4.6.: Theoretical limit (CRB) on the phase estimation accuracy in case of N=16 samples, shown for all possible true frequencies and phases.

the accuracy becomes very low. Therefore it is important not to choose such a bad frequency (e.g. sampling only once per rotation), but apart from that the graph shows that a wide range of frequencies yields similar results. The estimation algorithm used reaches the theoretical bound, as long as the estimate is not too noisy (Figure 4.5).

The same data can also be used to analyze performance for the case when the true signal frequency is unknown. In this case, prior knowledge on the sampling positions and rotational speed is simply not used. Instead, the frequency of the signal is estimated using an interpolated FFT, and the resulting phase estimation is based on that result. In practice, unknown uniform sampling will occur if a rotating polarization filter is used without camera synchronization. Additionally, a continuously rotating filter leads to a decrease in signal modulation due to camera integration across a range of polarizer angles. This setup is significantly cheaper than using a step motor.

Figure 4.7 shows that the influence of camera integration is small. The upper line assumes that there is no offset to the original sinusoidal signal, and that the light source intensity or exposure time is adjusted such that the maximum intensity reaches camera saturation. The lower line assumes that no such adjustments are made at all. In practice, the influence of integration will be in-between. For a realistic integration angle of $20°$, the contrast is reduced between 1% and 2%, depending on the offset of the true signal.

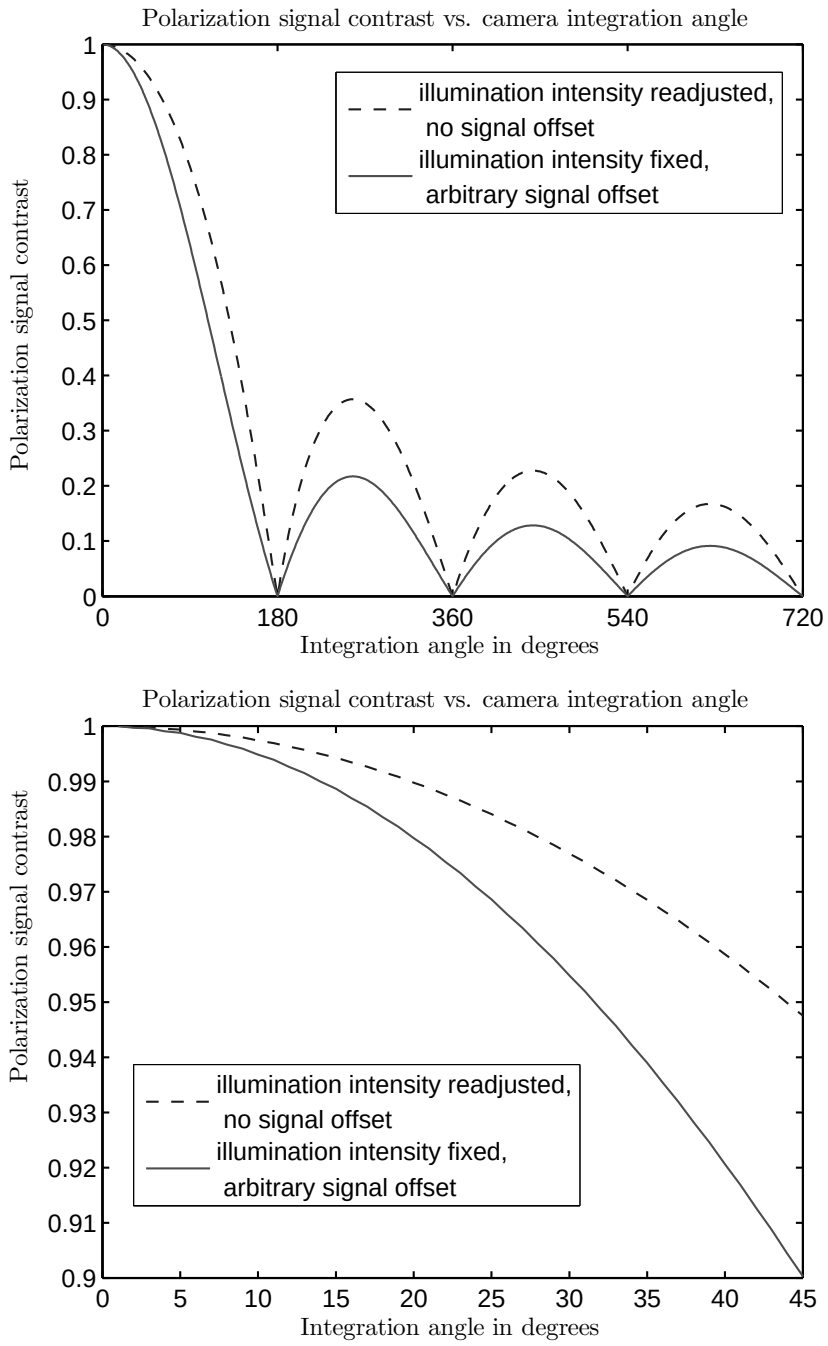


Figure 4.7.: Signal modulation (contrast) if a continuously spinning disk is used, both with fixed and readjusted illumination intensity, for an integration range of up to 720 degrees (top) and magnified for a range of up to 20 degrees (bottom).

Known and unknown non-uniform sampling

In this case the mechanical setup is identical to the previous section, but now a step motor is used to deliberately cause "sampling errors". Using the linear least squares algorithm the results for known non-uniform sampling are shown, and compared to the results for uniform sampling. Using a combination of frequency estimation and sampling position estimation, the same data is analyzed assuming unknown sampling positions. The accuracy of the results is compared to the case of known sampling.

4.1.6. Results

For the data set with uniform sampling, the number of frames N used for evaluation was varied between three and sixty. For each N all possible phase differences between the samples were used, and the resulting variance was averaged across all subsets of the data that fulfilled the conditions with respect to the number of samples and the sampling distance. The phase was estimated using linear least squares estimation. The difference to the reference measurement was computed and the variance of the difference across the field of view was determined. The resulting variance was multiplied with the number of samples used in order to take improvements that could be obtained by simply repeating the measurement into account as described above (Figure 4.9). An example polarization image is shown in Figure 4.8 (top). For comparison, a normal image of the exact same object, with the same illumination and angle but without polarization filter is shown in Figure 4.8 (bottom).

The results show that the performance of the linear least squares estimate improves if more samples are used. If only additive white Gaussian noise is present, one would expect a decrease in standard deviation with $1/\sqrt{N}$, or constant results after normalization in Figure 4.9. In practice, the improvement from repeated measurements is somewhat lower (which can be seen when looking at the case with sampling distance $60°$, which corresponds to repeating a 3-bucket measurement), and the results for other sampling patterns are slightly better. This is probably caused by systematic errors due to an incomplete signal model and correlated noise. There are many settings which perform close to each other, and therefore it is not very relevant how many samples are used. This number can be adjusted such that the desired signal quality is reached. As the computational effort increases linearly with the number of samples, performing repeated measurements or performing longer measurements makes no difference. It is important, though, that the samples are spaced such that they are spread evenly across at least one full signal period (i.e. $180°$ rotation of the polarizer). Measurements show, however, that performance is significantly better if full rotations of the polarizer ($360°$) are sampled, which is probably caused by errors introduced by the polarizer. It is obviously possible to obtain good results with samples spaced even further apart (by adding multiples of $180°$ to the sampling distance), but part of the slight additional improvements seen in the graph above might be caused by the reference measurement being based on a longer part of the same sequence of frames.

In addition to the evaluation above, two N-bucket algorithms were evaluated. The following results were obtained: 3-bucket (equation 4.12) with 0.0390 relative standard deviation of the phase estimate (normalized by the number of samples), and 5-bucket (equation 4.13) with 0.0320 relative standard deviation. This performance

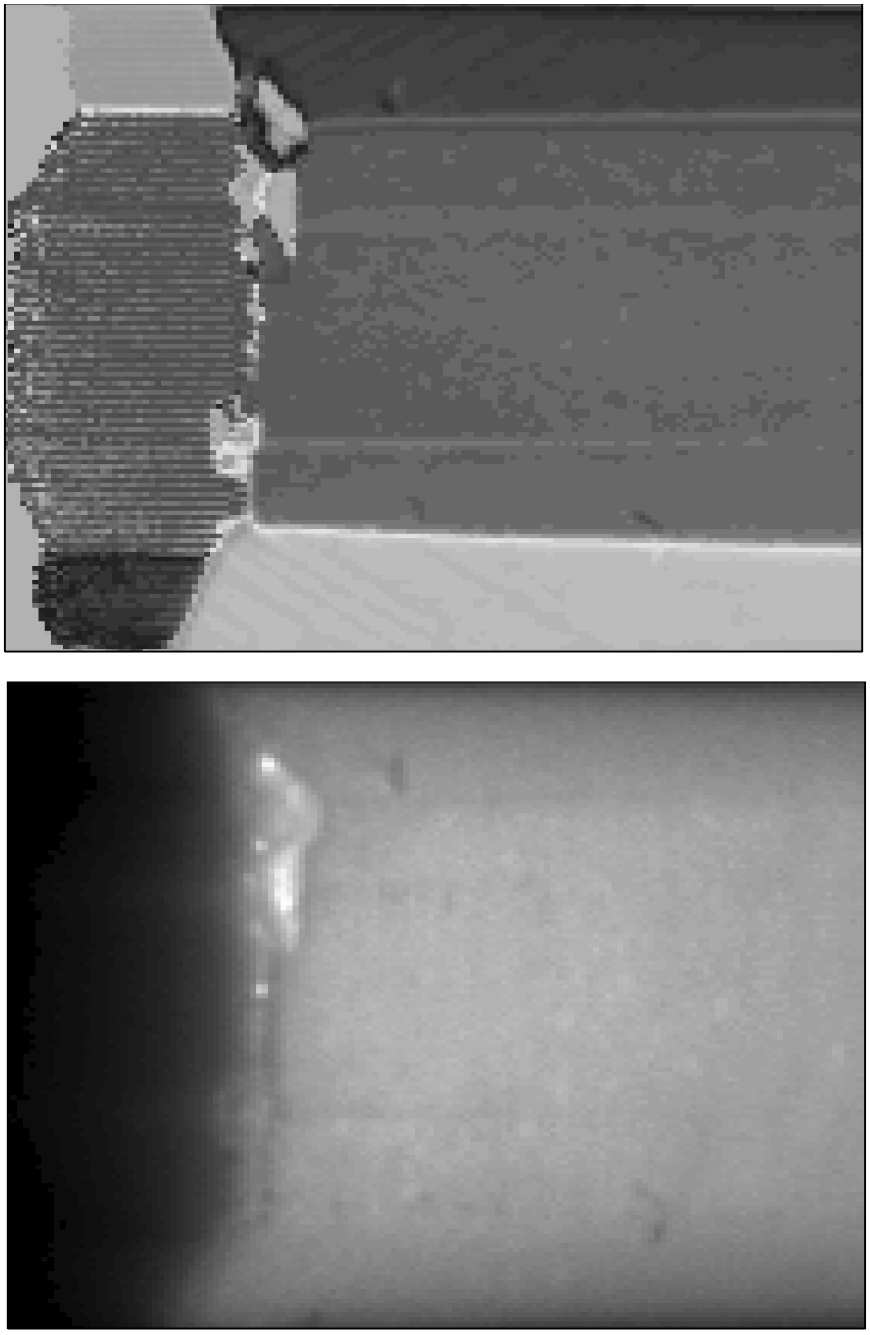


Figure 4.8.: Polarization image of a ceramic object, obtained using 17 frames with 21° distance (top) and normal image of the same object (bottom). Significantly less detail is visible.

(especially that of the 5-bucket algorithm) is slightly better than that of the linear least squares estimates, but N-bucket algorithms are less flexible in their implementation (they only work for a fixed interval), and are more difficult to derive for large num-

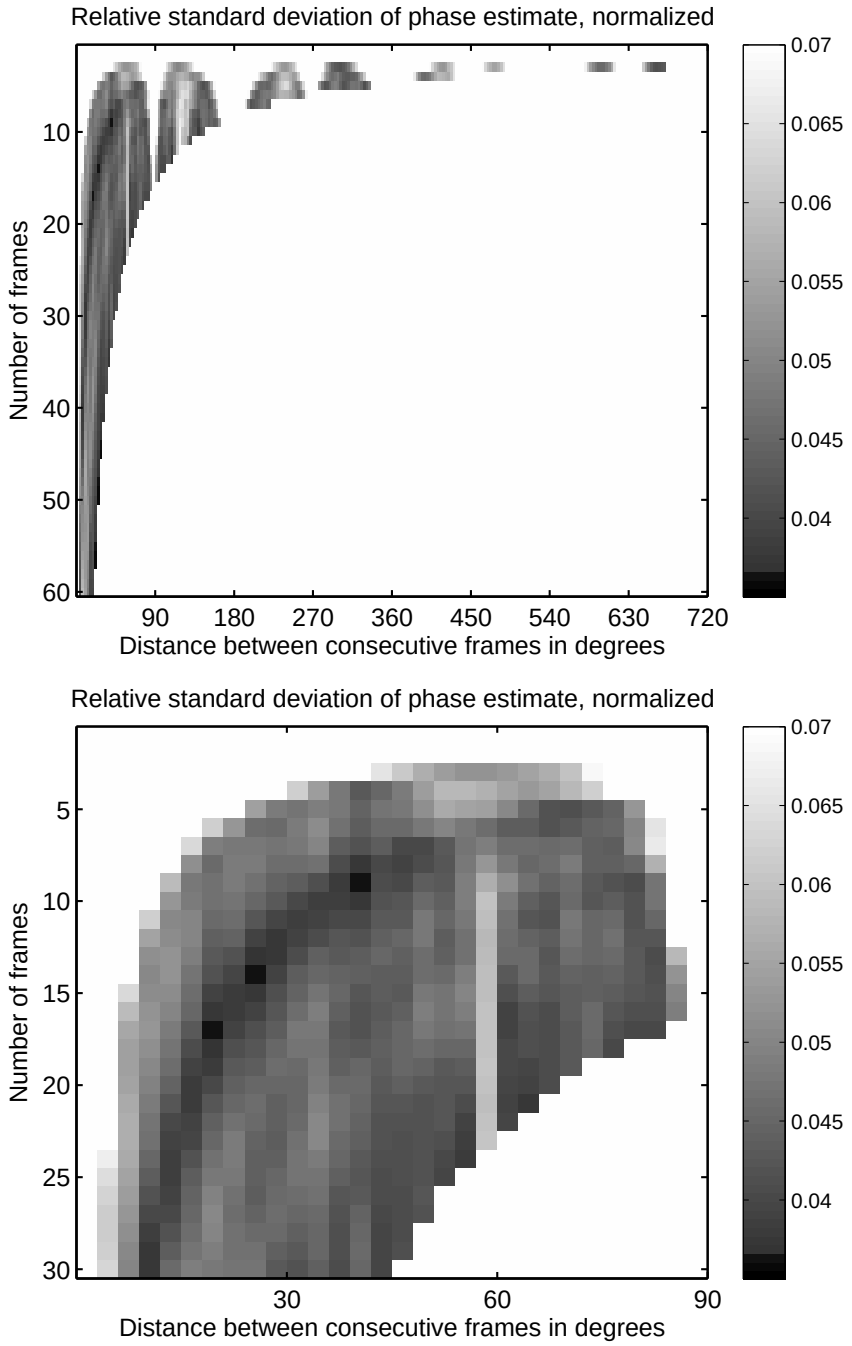


Figure 4.9.: Relative standard deviation of the phase estimate, normalized by the number of samples used (top), magnified to better show the results for a low number of samples (bottom).

bers of samples. The absolute performance is still significantly higher with algorithms using more frames, e.g. the 17 sample algorithm described next, and the linear least

Number of samples	Sampling distance	Known frequency	Unknown frequency
9	42°	0.0366	0.0372
14	27°	0.0365	0.0369
17	21°	0.0364	0.0367

Table 4.3.: Relative root mean squared error (not taking systematic offset into account) of the phase estimate for uniform sampling compared to the reference measurement.

Number of samples	Sampling distance	Known frequency	Unknown frequency
9	42°	0.0371	0.0376
14	27°	0.0367	0.0372
17	21°	0.0370	0.0372

Table 4.4.: Relative root mean squared error (not taking systematic offset into account) of the phase estimate for non-uniformly sampled data compared to the reference measurement.

squares approach is highly robust to errors as will be shown below. From a practical point of view, it is desirable to keep the rotation of the polarizer as slow as possible and the samples closely spaced: Then a free-running camera is possible, and integration time does not play a significant role. Settings that fulfill these conditions can be seen as dark squares in Figure 4.9 (bottom). For example, if we take 17 frames spaced by 21°, the result is highly accurate and the influence of integration time is less than 2% (cf. Figure 4.7), and therefore image acquisition can be performed with a continuously rotating polarization filter in front of the camera. In the next step, the signal frequency is assumed to be unknown, but the sampling is assumed to be uniform. This is the typical case if there is a rotating filter that is not synchronized to the camera. Three settings that performed well before were chosen (phase estimation using linear least squares estimation, as N-bucket is not applicable for arbitrary phase differences).

The difference in phase estimation performance between known and unknown frequency is very small, and it gets smaller for a larger number of samples. This has two main reasons: First of all, many time series are available for the frequency estimation, as the frequency must be identical for all pixels. The frequency estimates are not perfect, but for the measurement data available, the error is less than 1% of the Nyquist frequency (using an interpolated FFT according to section 3.6). Secondly, even if the frequency estimate is not quite correct, the resulting phase error is small: The phase is computed relative to the center of the sampling points, and the influence of the frequency estimates is very small at this position (it would be zero if there was no unknown offset).

For non-uniform sampling the results are shown in the next graph. Every third sample was moved by $\pm 3^\circ$ (alternating). 3° was chosen as this accuracy should be easy to reach in practice.

The results in Table 4.4 differ very little from the results for uniform sampling. If

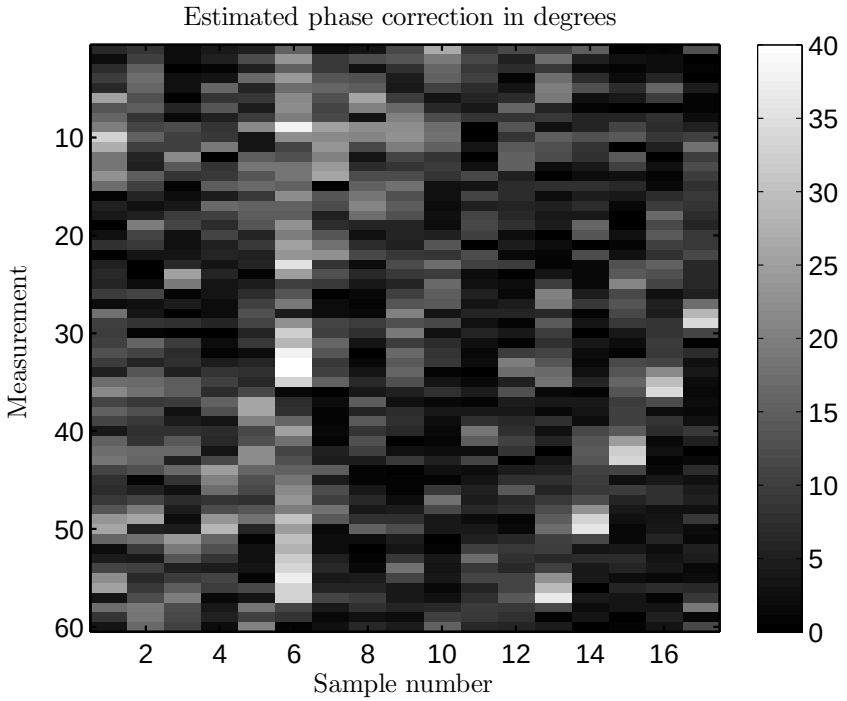


Figure 4.10.: Estimated phase correction angle (deviation from uniform sampling) in degrees. Based on a sampling pattern with 17 frames and distance $21°$; with sample six shifted by $18°$.

the jitter is small even in case of unknown sampling there is hardly any difference, and there is no need to modify the algorithms used. This shows that the described system is very robust. Reconstructing the actual sampling positions does not work very well, though. If the sixth sample only is moved by $18°$, the sampling position can be reconstructed well on simulated data if all phase values of the signal are equally likely. This is unrealistic in practical applications. For the data set used here, the following results are obtained (Figure 4.10): It is clearly visible that there is something wrong with sample six, but the sampling position estimates are quite noisy. Using this estimated sampling pattern for phase estimation does not improve the results.

As a last experiment, the worst case scenario using a free-running camera, a hand held power drill and some polarizing foil was used to acquire 17 frames. The resulting polarization image is shown in Figure 4.11. As there is no ground truth available, the noise level cannot be determined, but compared to a reference measurement obtained using 11 frames in increments of $18°$, the image looks correct and shows that with very little effort a polarization image with valuable phase information for every pixel can be obtained. It is very hard to tell which one of these images is better; at least for the noisy background, the one obtained with the power drill looks actually more accurate. The curvature of the object can be determined from the phase, which could possibly be used for shape from polarization algorithms.

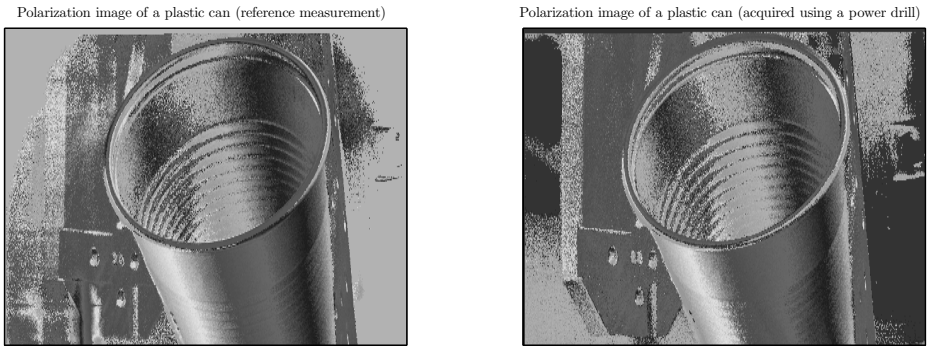


Figure 4.11.: Polarization phase images of a plastic can. The image on the left hand side was obtained from a standard high-accuracy measurement using 11 frames $18°$ apart, while the image on the right hand side has been acquired with a free running camera and a hand-held power drill. 17 frames were used in this case.

4.1.7. Computational complexity

For practical applications it is important that processing is fast. Computational complexity of the N-bucket algorithms is lowest, but the linear least squares solutions are very close and can still be computed very quickly on a standard PC (less than 0.5 seconds for 1 Million pixels and 8 frames on a Core 2 Duo at 2.4GHz). In a straightforward implementation, $3N$ multiplications, $3N - 3$ additions, one division and one table lookup are needed per pixel; for special sampling distances the effort is even lower with N-bucket algorithms. If the signal frequency is known, most of the processing can be performed frame-by-frame while the data is being acquired, and at the end only a division and computation of the inverse tangent (typically using a look-up-table) is needed. The algorithm for frequency estimation is also fast (about 5s for 1 million pixels, 8 frames), and in practice its performance is irrelevant as it does not have to be applied to all pixels: processing a few hundred good pixels (which can be chosen by looking at the modulation) is sufficient and takes a few milliseconds only.

4.1.8. Conclusion

Several algorithms known from other fields have been adapted for use in polarization measurements. There are four main results:

- Algorithms from other fields can easily be applied to polarization imaging, this includes both N-bucket algorithms and linear least squares estimation. These algorithms have been analyzed and their performance has been shown experimentally.

- The accuracy of polarization measurements has been improved significantly by increasing the number of raw images. The increase is directly proportional to

$\sqrt{N}$ if linear least squares estimation is used. The computational complexity remains low: The number of multiplications and additions increases linearly with the number of frames N, the computational cost for the rest of the algorithm is fixed.

- The proposed algorithms offers higher flexibility with respect to the number of samples and the sampling distance. The dependency of the accuracy on the signal-to-noise ratio and on the number of samples was shown quantitatively. This makes it possible to choose an optimum sampling pattern and algorithm for the desired application. The fast algorithms mentioned above can be applied to an arbitrary number of sampling points and arbitrary sampling distance.

- Furthermore, using the proposed algorithms for phase and frequency estimation it is possible to obtain accurate measurements without the need for synchronization of camera and polarizer rotation, significantly reducing the cost of such systems. A step motor is not needed; a continuously rotating polarizer is sufficient, and there are almost no requirements on the accuracy of the drive. The only difficult case of rapidly changing angular velocity can usually be fixed mechanically by simply increasing the weight of the rotating parts.

The algorithms presented are not limited to polarization vision and frequency scanning interferometry, but can be applied to other fields, including fringe projection, deflectometry, white-light interferometry, rotation sensors, radar signal processing and communication systems.

5. Summary

5.1. Comparison: WLI vs. FSI

Both white-light interferometry systems and frequency-scanning multiple wavelength interferometry systems are suitable for the measurement of optically rough and optically smooth surfaces, and both systems offer similar performance theoretically. However, there are huge differences with respect to the properties of the specific systems investigated here.

The high-speed scanning white-light interferometry system uses well known techniques, and there is a significant number of robust and stable (albeit slower) systems commercially available. Modifying hard- and software to reach higher measurement speeds is a challenging engineering task and can be expensive, but there are no fundamental difficulties as the implementation performed in the context of this thesis has shown. This type of system is ready for in-line production use. Its main disadvantage is measurement time: This time is proportional to the height range that is to be measured. Roughly speaking, one image has to be acquired per 100 nm, which quickly reaches huge numbers for large height differences (1cm: 100,000 frames). Even with high speed cameras, high intensity light sources and very fast processing, this is still a huge number and causes a measurement time of minutes instead of seconds. In some cases, this problem can be avoided:

- For relatively small height ranges and well-known part positions, a white-light interferometry system can be very fast. A scanning speed of more than 100 microns per second is possible, and this may be sufficient.

- Sometimes the object geometry has large height differences (e.g. the hull of a cylinder), but using optical components such as conical mirrors it can be made to appear plane. This has been investigated by the author, but it is not part of this thesis. An example is given in appendix C. Again, in these cases white-light interferometry is applicable right now.

- If there are multiple discontinuous surfaces that are far apart, one can easily "skip" the range in between while scanning — there is no data in between. This can be done as long as the stage is accurate enough (either using a glass scale or with an additional laser interferometer for absolute distance measurement). In these cases, only the much smaller effective height ranges where data is expected have to be scanned slowly. This method is also applicable right now.

The line scanning white-light interferometry system features a more complex optical setup, but its properties are now well understood [Hering, 2007]. There are two major advantages of that system: It is highly robust to vibration, and a line sensor is

very useful for scanning cylindric or free-form objects that would be difficult for another sensor. Measurement speed (in terms of height values per second) can be high if a high-speed camera is used, which also requires high-speed signal processing. Compared to 2-D scanning white-light interferometry, implementation of such algorithms on the camera or framegrabber is easier and requires less memory, but there is currently no such system commercially available. There is still significant development needed to build a compact and fast sensor for industrial applications, therefore it is currently not available for in-line application. A competing sensor with lower resolution, but better properties on high-contrast surfaces is the Siemens SiScan, which is commercially available.

The frequency scanning system analyzed in chapter 3 is less mature than the scanning WLI system and there are several open questions remaining, some of which will be discussed in the next section. The measurement principle is well known and well understood, but currently accurate laser tuning is still a serious issue. Additionally, the system is more sensitive to vibration and there is limited experience with respect to the long-term stability and reliability. The key advantage of this system is measurement time when larger height ranges are needed. The measurement time is on the order of a few seconds; it should be possible to reach a total measurement time of about 3s for the system described in this thesis. There is a limitation to about 1.2 mm continuous height range for the system discussed here, followed by 0.4 mm where the system is "blind", another 1.2 mm usable range, and so on. This is defined by the laser frequency increments in the tuning procedure. Measurements on both smooth and rough surfaces are possible, and the expected accuracy is comparable to white-light interferometry. For smooth surfaces there is a very elegant way of getting from using the signal frequency to using the signal phase. This could be particularly useful for thickness measurements, where the large measurement range can be used to automatically reference two measurement heads to each other with every single measurement (if the measurement object does not fill the whole field of view). Due to the much lower data rate, this system is interesting for high-resolution applications as the measurement time is mainly limited by exposure time and laser tuning. The system is currently not suitable for in-line application due to a number of stability issues, but these can probably be resolved.

5.2. Ideas for Further Development

Development for industrial use of the systems described here is not finished for any of the systems, but the systems are in very different states of development:

High-speed scanning white-light interferometry is ready for production use. A large number of algorithms with well-known advantages and disadvantages are available. Improvements are obviously possible and can and will occur when better components (faster cameras, higher dynamic range, better light sources, more accurate stages, faster computers etc.) become available. The impact of any such improvement can be predicted quite well, and there are no fundamentally different approaches expected. Apart from that there is a number of possible and useful improvements on a system level. These include

- Fully automated and robust self-calibration procedures (both lateral and in z-direction, automatic adjustment of the reference mirror, using depth from focus for finding the virtual reference plane, ...)

- Using special optics (e.g. conical mirrors) for measuring certain objects. Alignment and calibration are especially important for that.

- Flexible application and higher accuracy could by facilitated by more compact optical setups and by additional monitoring of the stage (i.e. using an integrated laser interferometer).

- Modular software frameworks that can be re-used for multiple sensors and offer image processing for data with a large dynamic range (32 bit is not sufficient for many height measurements, double precision might be required).

Some of solutions mentioned above have been developed by various manufacturers, but they are currently not well tested or readily available.

While all the issues mentioned above are also applicable to the line scanning WLI sensor and the multiple wavelength system, there is a large number of additional issues that have to be solved for these. For the line-scanning setup, they have been discussed by [Hering, 2007] and are not repeated here.

For the multiple wavelength system, a key issue is hardware stability:

- A fully automated system for tuning the laser is required. This includes automatic adjustment of the laser current and the grating positions.

- The hardware must be able to determine the absolute laser frequency and laser intensity fluctuations. A more accurate monitoring of the laser might be possible using part of the measurement head field of view, e.g. by placing a tilted step height artifact there. The relative change of the laser frequency can then be determined based on the phase of the sinusoidal signal on the slope for each individual frame, and the absolute frequency can be determined by looking at the step height resulting from a complete measurement.

- The system needs better thermal control, e.g. by using a Peltier element which can both heat and cool instead of the current simple heater with a binary on/off switch.

- The laser box is very sensitive to vibration. This can be tolerated if the box is stored in a protected environment (which is feasible as the fiber coupled laser light can reach other places easily), but a robust laser box would obviously reduce the overhead in implementing such a solution. A piezo instead of the currently used voice coil for tuning the laser might help.

- Illumination of the measurement head is poor. A more uniform illumination is required. This could be obtained with a better beam profile by using a different configuration for laser. As the absolute laser intensity is not critical, this should also be possible with optical changes in the measurement head.

- For some applications, a Michelson setup might have advantages (as discussed in chapter 3.1.3).

- The system needs some additional laser safety measures for easier usage. The required laser intensity is low, therefore it is possible to reach a class 1 laser designation without having to completely enclose the measurement head. The laser diode used is class 3B though, therefore precautions have to be taken to make sure that the emitted light does no exceed the limits of class 1. A concept for laser safety involving a monitoring diode at the laser head that detects both excessive intensity as well as a broken fiber has been implemented together with an appropriate electronics circuit for switching off the diode in case of errors, see appendix B.

Once the hardware improvements described above have been performed, both the software and the system characterization need to be completed:

- A more complete characterization of the system accuracy can be performed once the illumination is more homogeneous and the absolute laser frequencies are known. This includes a more detailed analysis of the influence of speckle, which would be most interesting if a system with a higher optical resolution was available.

- A simple user interface and user guide is required in order to facilitate use of the system. Several of the system properties are difficult to understand compared to other measurement systems, in particular the notion of an "ambiguity interval" and the alternating inverted height maps resulting from it.

- For filtering, the noise properties of the system need to be analyzed in more detail so that the optimum filter size and parameters can be detected automatically or at least with little user input.

- All algorithms described in this thesis need to be implemented in an appropriate programming language (using available libraries), the current Matlab implementation is less stable and consumes more memory than necessary.

Once all the improvements listed above have been completed, extensive testing will be needed before the system can be used for in-line applications; other issues might show up at that point.

5.3. Summary

In this thesis, three different measurement systems have been analyzed and optimized for in-line use in a production line. As the techniques used for this optimization include a variety of very different aspects, a detailed summary of the results has already been given in each of the individual sections. The key results for the three systems are briefly summarized here:

- A scanning white-light interferometry system has been built that is faster than any other system on the market at the time. Most of the work fell in the following three categories:
  - The hardware (in particular camera and framegrabber) was replaced with high-speed components and these were integrated into the system. Various components and measurement strategies were tested and compared (i.e. triggering the camera based on the stage position).
  - Various algorithms were analyzed for hardware-supported acceleration of the signal processing, and additionally a review of the properties of these algorithms in the presence of various types of noise was performed. In parallel, 3D-Shape GmbH worked on accelerating their processing software and added support for the new hardware components.
  - The system was extensively tested and its components were characterized. This included comparing various sampling strategies as well as different stages, cameras or light sources. The sensor was integrated into an automation framework and algorithms for automatic defect detection were developed.
- For the line scanning interferometry system, different algorithms were compared and concepts for a hardware-based implementation were developed. This way, a concept for a novel, highly integrated 3D-sensor has been created.
- The most extensive analysis and optimization has been performed for the frequency-scanning multiple wavelength interferometry system:
  - The hardware and the underlying physics were analyzed and a signal model was derived. Various sampling strategies, including some using non-uniform sampling, were analyzed for the first time for frequency scanning interferometry.
  - In the next step, an optimization problem for the estimation of the signal frequency was formulated. The theoretical aspect of that problem was analyzed in detail by Matthias Wieler, and its solution is given in [Wieler et al., 2006].
  - An approximation to the theoretically optimal sampling scheme was derived, and an estimation algorithm developed and analyzed. Performance was compared to other sampling schemes and algorithms. This is a general result, and can be applied to other fields as well.
  - As a building block for the new algorithm, a method to quickly and accurately estimate phase and frequency from short blocks of data was required. There was no readily available algorithm that satisfied the requirements on speed and accuracy. Therefore new approaches for very fast phase and frequency from a low number of samples have been derived. This is also a general result, and has been successfully applied to polarization imaging and other applications.

- The properties of the laser have been analyzed in detail, and algorithms to determine the laser frequency have been developed.
- Bayesian approaches to height map estimation have been investigated, and a simple spatial filtering approach ("adaptive remapping") using knowledge on modulation and phase coupling from the estimation algorithm has been implemented that is able to reach most of the benefits one would expect from Bayesian estimation at a fraction of the effort.
- The algorithms described above have been implemented and extensively tested and compared to other approaches (e.g. non-linear optimization). Plug-ins for a software framework of the hardware manufacturer have been written such that the algorithms could be used directly for measurements.
- Simulation and measurement results for both smooth and rough surfaces have been acquired and used to optimize and verify the algorithms.
- The influence of laser speckle on rough surfaces on the measurement accuracy has been discussed, and a method to measure the phase change of a speckle field with changing laser frequency has been proposed.

The main challenge in performing this analysis has been the wide range of techniques from different fields of science required to optimize these optical measurement systems. Relevant fields include estimation theory (theoretical limits, CRB), classical optics (imaging properties), speckle statistics, hardware and sensor technology (cameras, lasers), software architecture (automation, modular concepts for software), algorithms (FFT, filtering, image processing) and computer architecture (efficient and fast hardware-supported algorithms).

Work in this field is never going to be complete; there are a vast number of possible improvements with advances in hardware and computer technology. However, in a number of areas the improvements are going to be small: Given the assumptions on the frequency scanning interferometry system used in this thesis, the combination of the proposed new algorithms reaches a result that is less than 5% from the theoretical limit. For other results presented here, e.g. the performance benefit of hardware implementations of some algorithms, such a limit cannot be given. However, these results will probably still be useful in the future, even though the hardware might change significantly: Currently there is a strong trend towards parallelization (multi-core systems, GPUs), and such architectures impose requirements where both an analysis of the data flow and the issue of parallelization as discussed in this thesis are important.

Alternative measurement methods are currently under development, most notably electronic speckle pattern interferometry with multiple wavelengths. Only time will tell which one of these techniques succeeds in practice.

Acknowledgement

The author would like to thank all those who contributed to this thesis. First of all, I would like to thank Prof. Dr. Fred Hamprecht and Walter Happold who made this project possible. I'm grateful for the opportunity to work on both theoretical and practical issues and I think I have learned a lot more than what is written down in this thesis. I think this will greatly help me in the future.

In particular, many fruitful discussions with Prof. Dr. Fred Hamprecht brought this work ahead and helped me find new solutions. His knowledge as well as his attention to detail were highly important for me, and I'm deeply grateful that he supported me whenever possible.

I would also like to thank Dr. Ralf Zink for his valuable input and support, both in scientific and in administrative tasks.

Work on high-speed white-light interferometry would not have been possible without my colleagues Thomas Seiffert and Marco Hering. I would also like to thank Dr. Peter Ettl from 3D-Shape GmbH for his support.

Work on multiple wavelength interferometry would not have been possible without support from Joseph Marron, Tom Dunn, Mark Tronolone and many others at Corning Tropel. I really enjoyed working with them.

I would also like to thank all colleagues at the Interdisciplinary Center for Scientific Computing in Heidelberg and at the Robert Bosch GmbH in Schwieberdingen for their cooperation and a great work atmosphere. The author would also like to acknowledge financial support by the Robert Bosch GmbH.

I'm grateful to all interns and diploma students who contributed to this work, in particular Matthias Wieler. He did an excellent job working on an important aspect of the problem I assigned to him, and it helped me very much to cooperate closely with someone working on the same issues for a relatively long period of time. I'm glad he decided to pursue his PhD thesis in our group.

And last but not least, I'm grateful for the support of my parents who supported my education over many years. I'm also grateful to all others who helped me get to this point, first of all my friends and the professors in electrical engineering and information technology at Stuttgart university, in particular Prof. Dr. Kühn, as well as the professors in electrical and computer engineering at the Georgia institute of Technology in Atlanta, in particular Prof. Dr. Lanterman.

A. Properties of Linear Stages

For any scanning white-light interferometer, the accuracy is not only limited by noise and speckle field, but also by the accuracy of the stage. The following analysis does not offer fundamental new insight, but it illustrates the importance of an analysis of all components with respect to the specific issues in white-light interferometry. It shows that specifications alone do not offer all the required information, and may help the reader to perform a similar analysis for optimization of a white-light interferometry system.

In white-light interferometry, there are two main influences coming from the motion stage: Sampling jitter can lead to problems in analysis algorithms (particularly correlation based ones) and any position error of the stage directly leads to an error in the measured height. In order to deal with the first type of error, an appropriate algorithm for the expected sampling jitter has to be selected; there is no universal best solution. It is important to choose a stage with constant velocity and low vibration. The second type of error can only be reduced by obtaining more accurate position information from the stage or an additional measurement system. For high precision applications, the stage will always be equipped with a high precision encoder, and its signal is usually more accurate than the assumption of uniform movement. Following error can be eliminated that way — therefore it is helpful to record the encoder positions whenever a frame is acquired. Then later interpolating the maximum of the correlogram on the grid of recorded positions is possible. This obviously requires a controller that is able to provide the current position in sync with the camera, which can require relatively fast sampling (more than 10kHz for a high-speed system). In contrast to many other applications, it is not important that the stage is at a specific position at a given time, it is only important that the stage is moving with a constant velocity (low sampling jitter) and that the actual position at any given point in time is well known. This has been implemented by choosing control parameters to reach a very smooth movement, even though this might result in a larger deviation from the setpoint positions.

For the stages discussed next, the position has been recorded by both the integrated encoder as well as an external laser interferometer. The results presented next are mainly for illustrative purposes to show the influence of the stage type on the two sorts of errors described above. The results differ significantly from stage to stage; a detailed optimization of the stage or controller is outside the scope of this thesis.

A.1. M-511DG.K029

For this stage, the difference between the setpoint position and the actual position according to the rotary encoder is almost always close to zero during a movement. An encoder increment corresponds to 6nm, and the controller is able to keep to this position (except for the acceleration and deceleration phase) as there is a very small

and constant load, an equally low constant velocity, and — most importantly — the rotary encoder does not monitor the actual stage movement, but the motor position instead, so there is a very short closed loop. Unfortunately, this encoder information does not necessarily match the actual movement of the spindle driven stage, as can be seen in the following graphs. The maximum velocity of this stage is low, about 1mm/s. There are two types of noise: high frequency noise mainly due to vibration (on the order of 50nm peak-to-peak), and large position errors, mainly due to gears and spindle not being perfect. Bidirectional repeatability is poor (more than a micron difference), which indicates that the connection between the table and the motor is not completely rigid. Unidirectional repeatability is quite good though, and it seems to be possible to reduce errors by calibration. The resulting curves show a standard deviation of only about 60nm (400 nm peak-to-peak). The high frequency components cannot be reduced though, and there have been no investigations to verify whether the calibration remains valid over extended periods of time. These results only apply as long as the control is active. Finding the home position is not as accurate as desirable, so once the stage has been moved without closed-loop control, it might not be possible to find the exact same start position, and therefore the calibration data might become invalid. A series of measurement results for the velocities of $4\mu m/s$, $53\mu m/s$ and $149\mu m/s$ is shown in the following. In all cases, the absolute position error (Figure reffig:poserror), the remaining position error after calibration (Figure A.2) and the spectrum of the noise (Figure A.3) are shown. The presence of high-frequency noise prevents the use of correlation based algorithms and seriously degrades performance of most N-block and correlation based algorithms. FFT-based approaches perform best for that kind of signal due to the possibility to easily filter out unwanted frequency components.

A.2. M-511DD

Mechanically, the table of this stage is almost the same as the one described above, but there are two key differences: The actual position is measured by a glass scale of 20 micron length; yielding a nominal resolution of 100nm. It uses a different type of drive as well, and the maximum velocity is much higher than for the version above.

Measurements show that this stage has a much lower error for longer measurement ranges and has less high frequency jitter, but there are two problems: Misalignment of the encoder leads to periodic errors and system vibration, and the available resolution is too low, quite often two consecutive camera images are assigned to the same nominal position, leading to difficulties in interpolation. When using that stage, the signal turned out to be almost impossible to process - the vibration caused serious issues with all estimation algorithms. For this stage, the position error is shown in Figure A.4, the velocity error at $4\mu m/s$ is shown in Figure A.5 and the spectrum at $200\mu m/s$ is shown in Figure A.6. A comparison to M511DG.K029 is given in Figure A.7.

A.3. Newport XML-350

This stage offers a larger moving range (which is not needed for white-light interferometry, but no smaller version was available for testing) and is much faster than the stages described above (up to 300mm/s). It also uses a glass scale encoder, but a more accurate one (4 microns length) and internally the signal is subdivided by 32768, yielding an internal resolution of 0.122 nm. Externally, a resolution of 1nm is available. The stage accuracy is obviously lower than the encoder resolution, but it offers much better repeatability and accuracy than the stages described above. The data in Figure A.8 and Figure A.9 shows the stage following error at $10\mu m/s$ and is taken from measurement data supplied by Newport for the XMS-50 (a smaller version of the stage). Measurements performed by the author for the XML-350 roughly confirmed these results, but due to limited time with the system a full documentation was not possible.

A.4. PI P-625.1CD

Very accurate positioning is possible using piezo-driven stages. These have limited moving range (typically less than a millimeter), and can only carry relatively small loads. They cannot be used to move the whole measurement system, but they can move the reference mirror. Their accuracy is very high, measurement results are shown in Figure A.10 and Figure A.11. When moving the stage quickly, a problem with the stage velocity showed up as the control did not seem to attempt to keep the velocity constant. This has been fixed in the meantime. After correction of that error, the remaining noise standard deviation is about 13.6nm. A comparison of the integrated capacitive sensor with the laser interferometer shows that the difference between these two sensors (after performing a linear correction as the absolute values were slightly different) has a standard deviation of less than 10nm. Taking the measurement noise into account it is hard to tell which one of them is more accurate.

In addition to accuracy considerations, the lifetime of the stages has to be taken into account when using them for an application in a production line. Typically, in a production line such a stage will have to repeat the same short movement over and over again every few seconds. Most stages have not been designed for such short movements, therefore in order to increase the lifetime of the mechanical components (bearings etc.), one has to move the stage for a longer distance periodically to improve lubrication. Piezo-driven stages are more suitable for that kind of application, therefore if it is possible to use them they should be preferred.

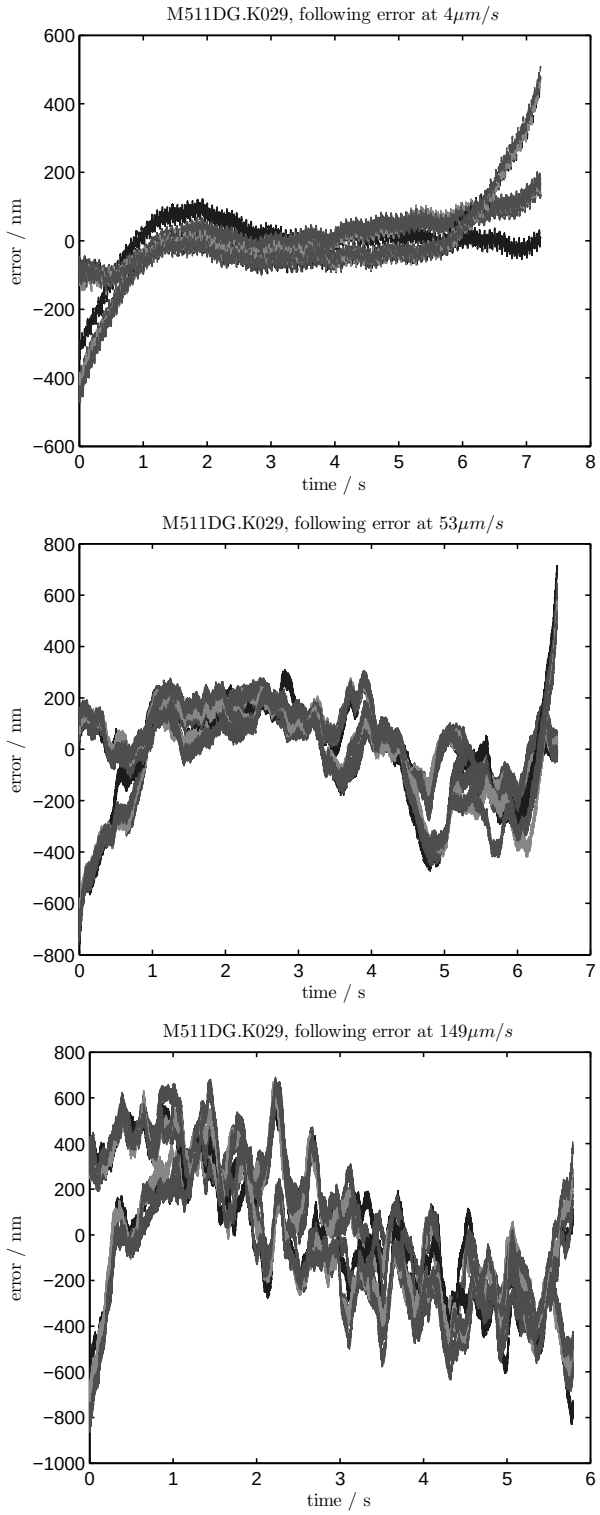


Figure A.1.: Position error of the stage M511DG.K029 at $4\mu m/s$, $53\mu m/s$ and $149\mu m/s$. Bidirectional movement (five measurements for each direction are shown in each graph).

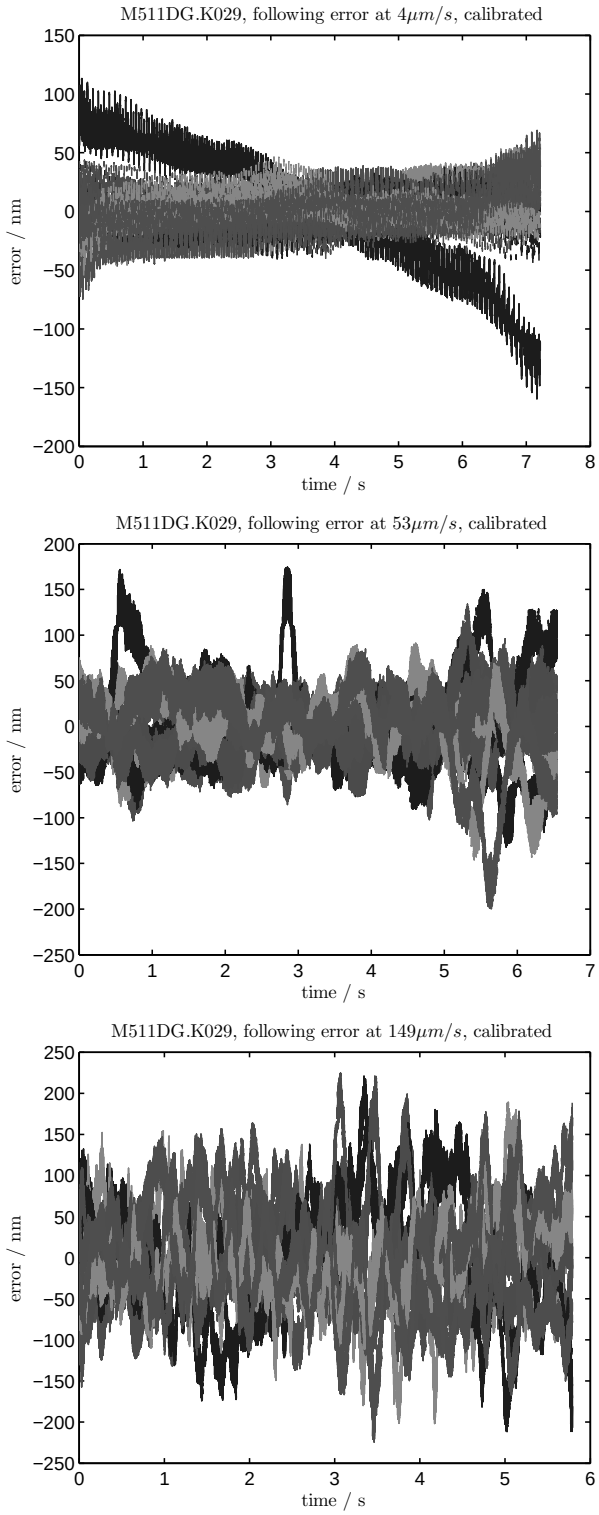


Figure A.2.: Position error of the stage M511DG.K029 at $4\mu m/s$, $53\mu m/s$ and $149\mu m/s$. Bidirectional movement (five measurements for each direction are shown in each graph). For each direction, an individual calibration has been applied. The calibration values were obtained by repeated measurements.

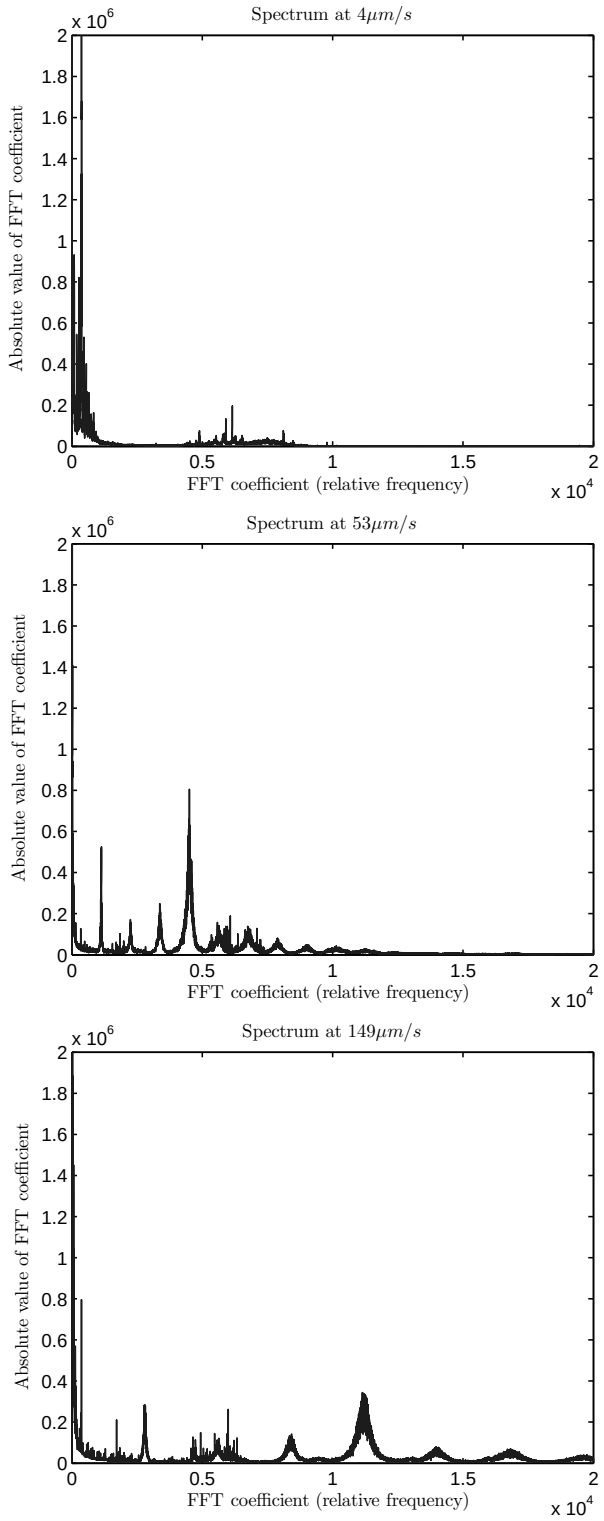


Figure A.3.: Spectrum of sampling jitter for M511DG.K029 at velocities of $4\mu m/s$, $53\mu m/s$ and $149\mu m/s$. The frequency of the jitter is inverse proportional to the speed, which indicates that it is linked to the gear ratio of the stage. A more detailed analysis confirms this guess.

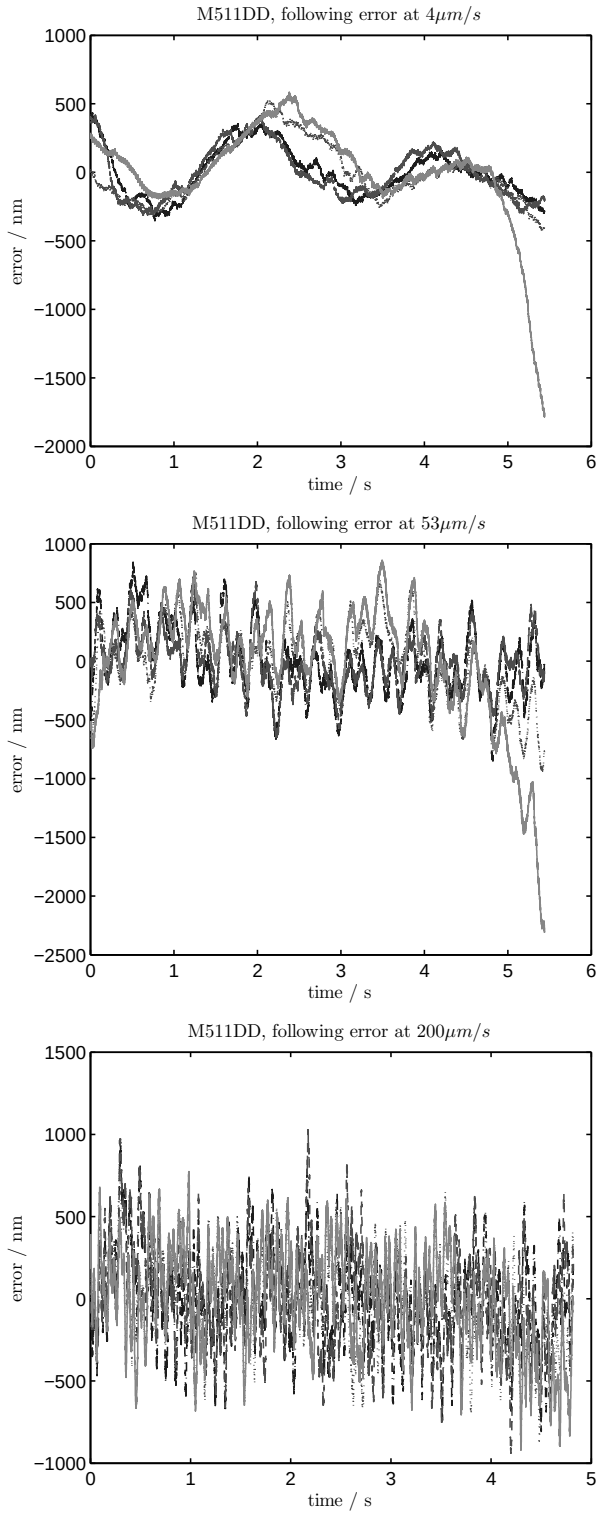


Figure A.4.: Position error of the stage M511DD at $4\mu m/s$, $53\mu m/s$ and $200\mu m/s$. Bi-directional movement (two measurements for each direction are shown in each graph).

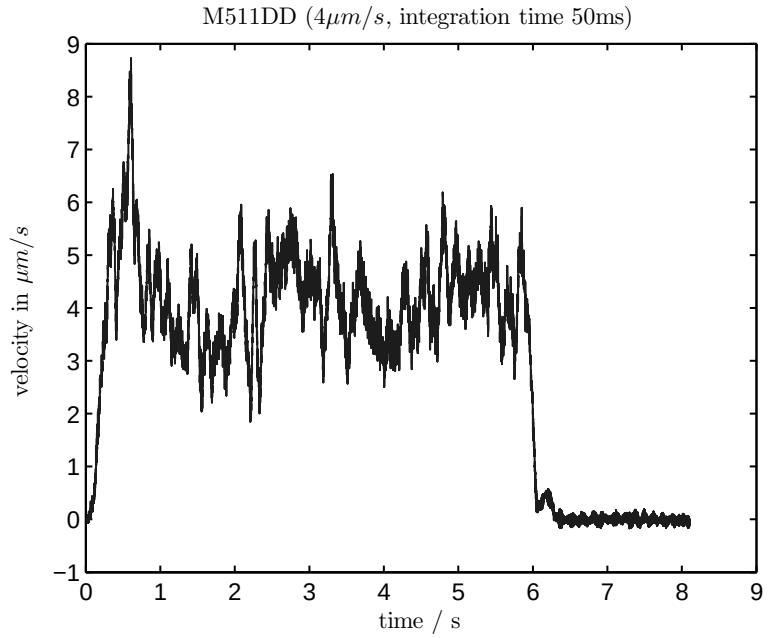


Figure A.5.: Actual velocity of the M511DD (nominal velocity $4\mu m/s$). The result is very poor.

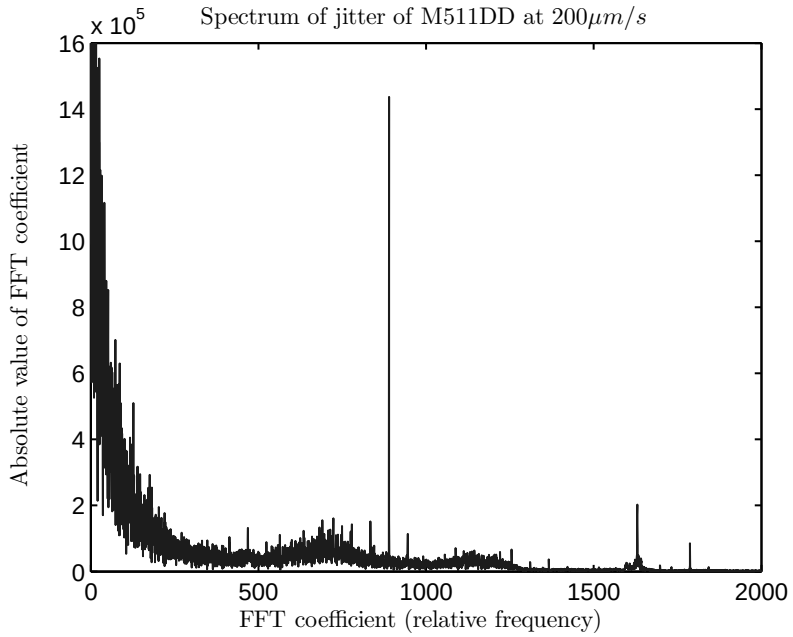


Figure A.6.: Spectrum of jitter for M511DD at $200\mu m/s$. Most of the noise is due to a single frequency component which corresponds to the length of the glass scale. This indicates misalignment of the glass scale.

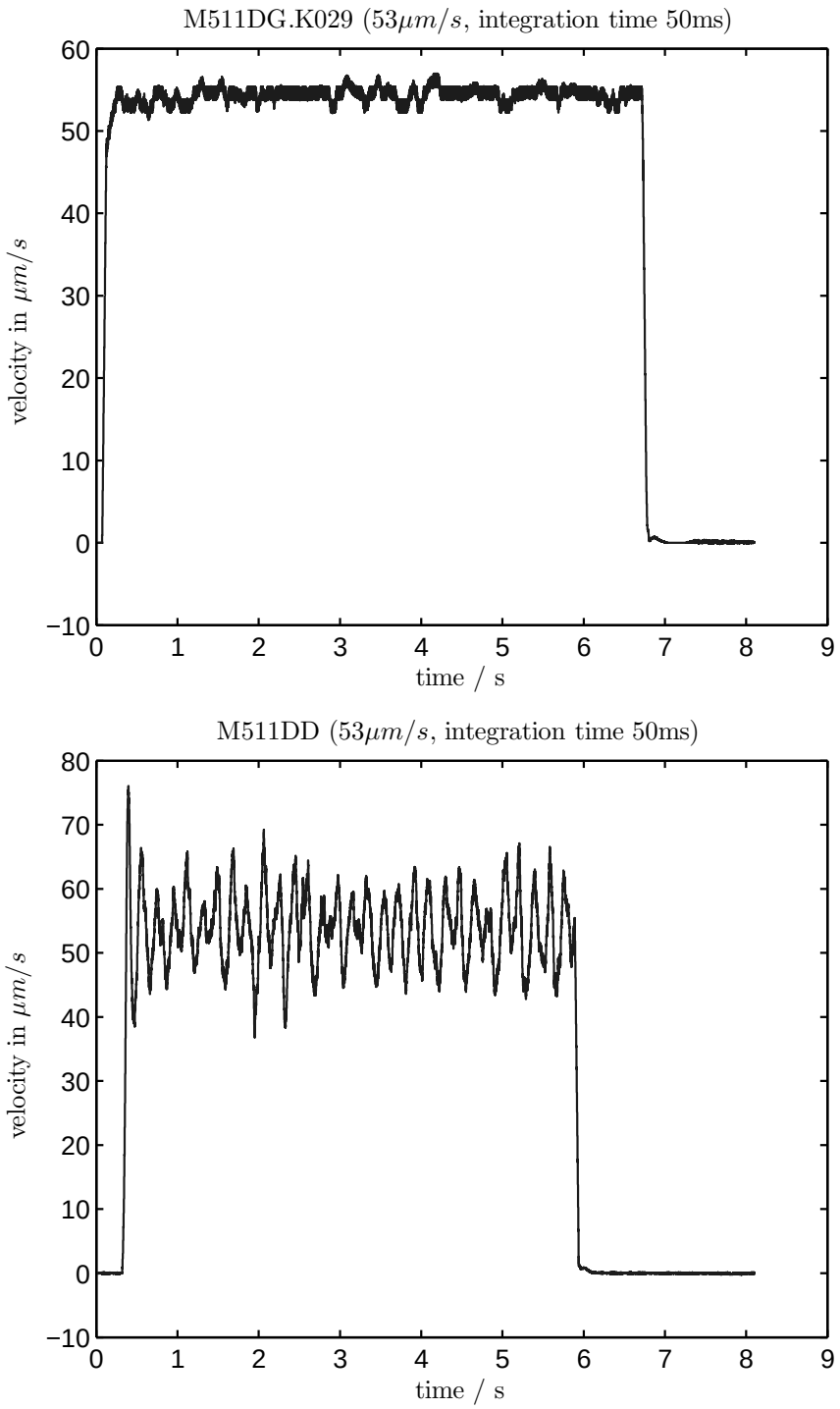


Figure A.7.: Actual velocity of the M511DD (bottom) and M511DG.K029 (top). Nominal velocity is $53\mu m/s$. There is clearly something wrong with the stage M511DD, as it performs much worse than M511DG and all other stages investigated.

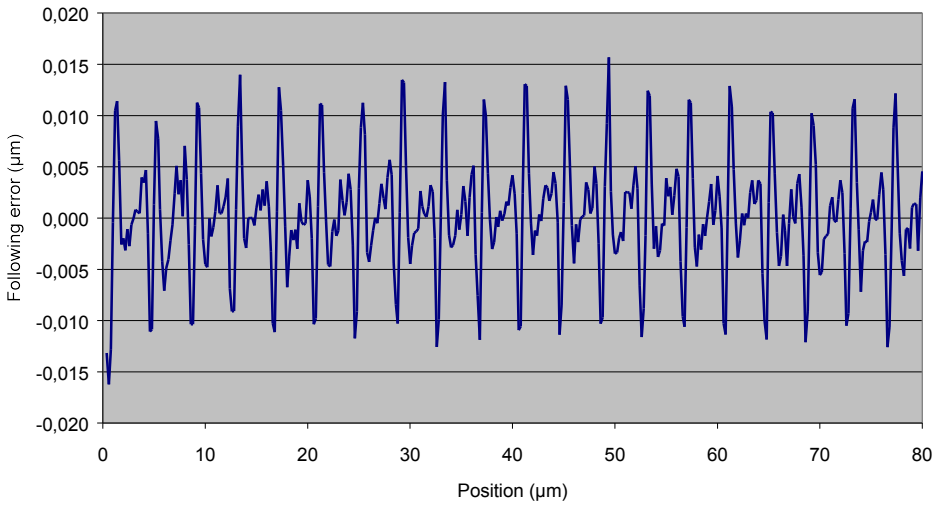


Figure A.8.: Position error of the stage XMS-50 at $10\mu m/s$. The error of the glass scale is clearly visible and the noise level is extremely low. Such a measurement is only possible under laboratory conditions, in a production environment errors will be significantly larger due to external influences. For larger distances the error increases (maximum about 500nm over a range of 5 cm), but calibration is probably possible similar to the other stages.

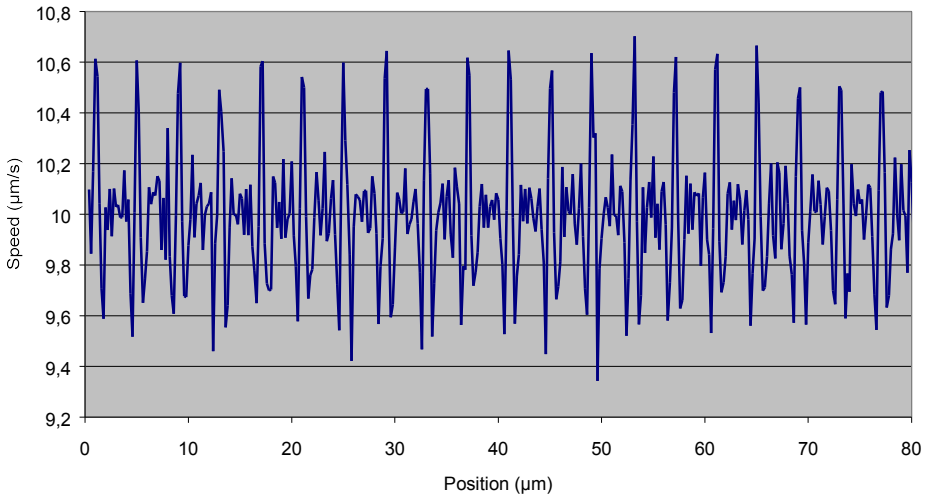


Figure A.9.: Actual velocity of the XMS-50 (nominal velocity $10\mu m/s$). The velocity variation is much smaller than for the other tables.

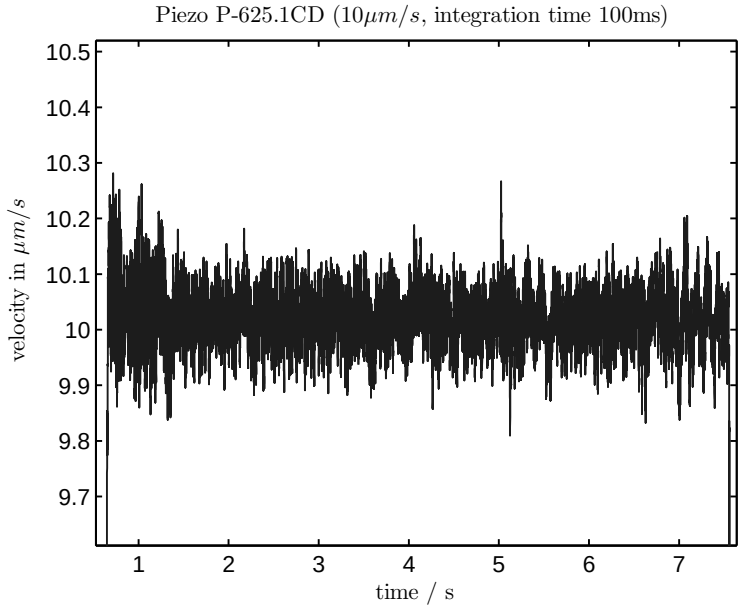


Figure A.10.: Actual velocity of the Pi Piezo P-625.1CD (nominal velocity $10\mu m/s$). The velocity is even more stable than in case of the Newport XMS-50, at a much lower load and looking at a smaller measurement range though.

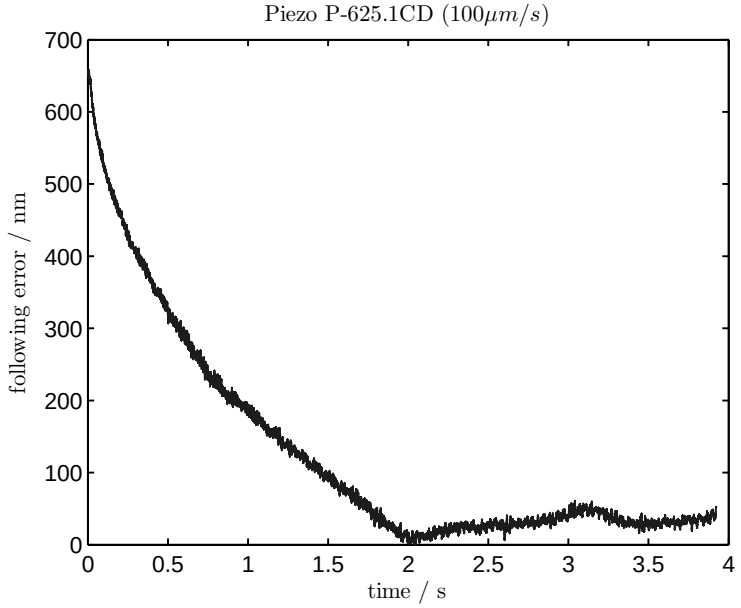


Figure A.11.: Following error of the Pi Piezo P-625.1CD (velocity $100\mu m/s$). There was something wrong with the controller as it did not keep the velocity constant throughout the measurement. This has been fixed in the meantime.

B. Laser Safety

Measurements were complicated by the fact that the laser diode used for the multiple wavelength system is a class 3 laser: Its power exceeds 1mW, and the frequency used is barely visible at about 785 nm. Therefore it is in general not safe to look into the laser beam, and protective measures have to be taken.

There are two possible approaches:

- First of all, the system could be put into an enclosure (which should not be a problem for use in a production line, but is difficult for manual measurements) such that its laser radiation is not accessible. Then the whole system can be classified as laser class 1, like a CD player.
- The alternative is to make sure that the radiation emitted from the measurement head is always below the limit for a class 1 laser system. This is possible as not much light is needed for the measurements.

The first approach is trivial to implement, but the second one is more desirable. As the measurement head is connected to the laser with a fiber, not only laser emissions from the measurement head but also the risk of a possible fiber break has to be considered. The easiest approach therefore is to determine the laser power in the fiber and make sure that this is low enough such that there is no risk.

Measurements show that the maximum power coupled into the fiber is approximately 1mW when using maximum intensity. For a typical measurement only a fraction of that power (on the order of $20\mu W$) is needed.

There are two kinds of limits to be taken into account: One is based on the energy of the light source and one is based on the irradiance as seen by the human eye. For that consideration the output at the measurement head is considered (if the fiber breaks, the resulting beam will be highly divergent and therefore only dangerous in a very short distance of the output). The maximum time the laser is looked at is also considered in the standards. A measurement will take a few seconds only, and the laser can be blocked between measurements. Additionally, there is no good reason for looking directly into the measurement head, so assuming an exposure time of 100s should be sufficient.

A calculation based on IEC-825 shows that with wavelength 785nm, emission duration 100s, assuming 1mW laser power and the properties of the measurement head, the laser energy and irradiance are far below the limits for laser class 1 (at 29% and 55% respectively). This shows that the system can easily work as a class 1 laser system as 1mW is far more than needed for measurements.

However, one still has to make sure that this level of radiation cannot be exceed in case of an error, and a system concept to realize that is presented here. There are two approaches to ensure that this power cannot be exceeded in case of an error:

- All possible faults can be discussed and it can be shown that the emission will still be below the limits. The system has been optimized and aligned to reach a high level of output, so it is unlikely that in case of a fault there will be more output power - it is far more probable that intensity will decrease or that there will be none at all. However, proofing that is almost impossible (as the diode itself has higher power, but a large amount of intensity is lost in tuning, fiber coupling, etc.).

- Alternatively, measures can be taken to ensure that no single fault can go undetected and will lead to an increase in radiation. This can be done with an additional electronics circuit which will be described in the following.

In order to detect excessive radiation or a break in the fiber, the signal at the measurement head has to be monitored. This can be done by coupling part of the light from the fiber into a photo diode at the measurement head. The logic works as follows:

- If the laser intensity exceeds a set limit (which can be set and verified experimentally and can typically be much lower than the 1mW used for the calculation above), an electronics circuit at the laser will detect that and block the laser beam immediately.

- In order to detect damage to the fiber, there must a lower limit as well - if the laser power is below that limit, the laser beam must be blocked as well.

- As this would make it impossible to turn the system on, the monitor can only be active when the laser blocker is open.

- A simple implementation like that would lead to oscillation, therefore a short time delay is needed to give the laser blocker time to open before closing it due to violation of the lower limit. Once an error is detected, this state must be kept when the laser blocker is closed.

- The time delay should not be there for the upper limit.

- Once an error occurs, the laser has to be blocked until a manual reset using a key switch is performed (this ensures that e.g. in case of a broken fiber the system cannot be turned on again without manual verification by the person responsible for operation of the system).

This has been implemented and should be sufficient for classification as a class 1 laser system.

C. Special Optics

For many objects, direct interferometric measurements are difficult or even impossible. This is true for some very rough objects or for objects with very high contrast; in these cases depth from focus might work better. But relatively smooth objects can also cause problems: If the slope of a smooth object is too steep, no light will be reflected back onto the detector. Therefore the measurement direction should be perpendicular to the surface. In case of a 1-D line sensor this is usually possible, but with a 2-D sensor it is impossible to be perpendicular to the whole surface if that surface e.g. has a strong curvature.

Sometimes it is possible to resolve this problem with an optimized optical setup. For example, a cylindrical surface can be imaged using a conical mirror, and then appears plane to the measurement system (Figure C.1 and Figure C.2). Then there is no problem with lack of light returned and the height range becomes very small (resulting in a short measurement time).

Key difficulties are the calibration of the mirror and the alignment of the system: Not only the part has to be aligned relative to the measurement system, but also the mirror has to be aligned correctly. This results in six additional degrees of freedom. There are methods to perform the alignment automatically without requiring full measurements and reconstruction of the height map. These are not discussed here.

Requirements on the mirror are very high, any surface roughness or deviation that cannot be removed by calibration will lead to additional measurement errors (as the light is reflected twice on the mirror, calibration is difficult and errors add up quickly).

To illustrate this principle, an example image of a cylindrical object measured by a white-light interferometer using a conical aluminum mirror is shown in Figure C.3.

It is important to notice that after backprojection the sampling grid is not uniform. Only a small part of the camera field of view can usually be used, therefore this method is probably not preferred if a very high lateral image resolution is required. For this application it is useful if the camera can be set to an arbitrary region of interest (e.g. a ring): Especially in case of white-light interferometry the data rate can be reduced significantly and thus the measurement can be accelerated. Such an optical setup can be used with any measurement system, including multiple wavelength interferometry, but it is most useful when white-light interferometry is to be used.

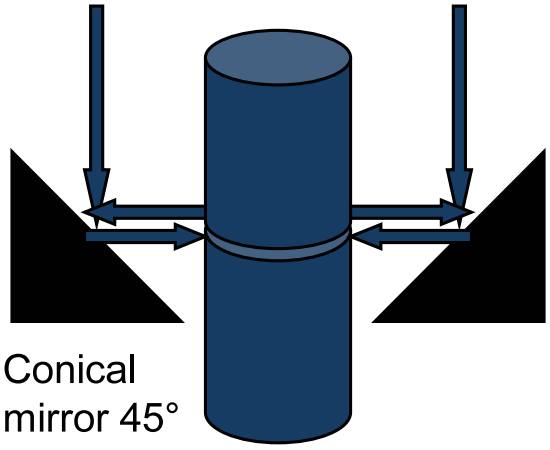


Figure C.1.: As shown in the diagram above, the surface of cylindrical objects can be measured with a white-light interferometry system by using a conical mirror. The object then looks flat to the measurement system.

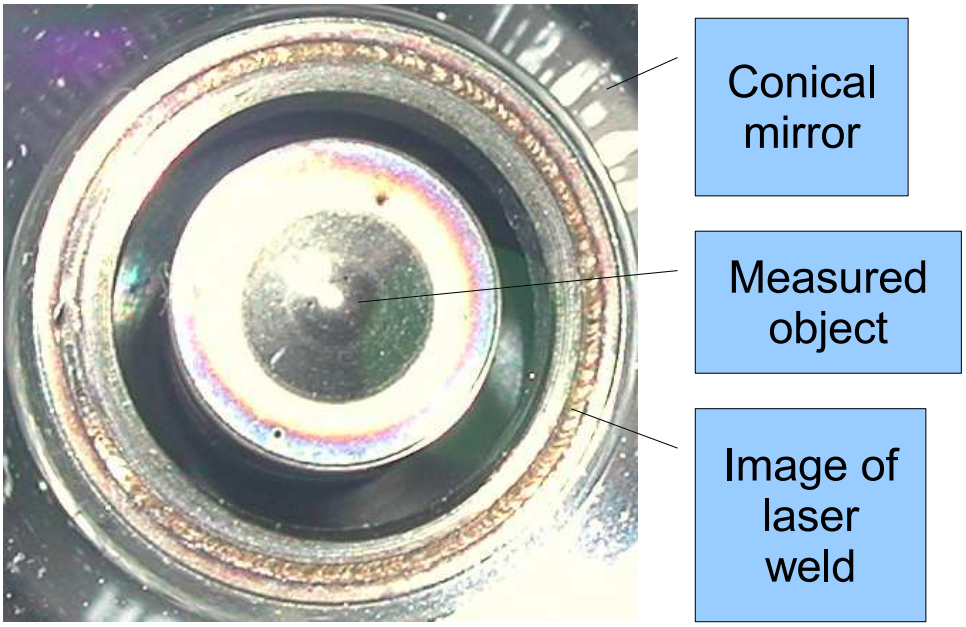


Figure C.2.: Measurement setup for a laser weld: Image of object and reference mirror taken by a normal camera. The image of the weld is clearly visible on the conical mirror.

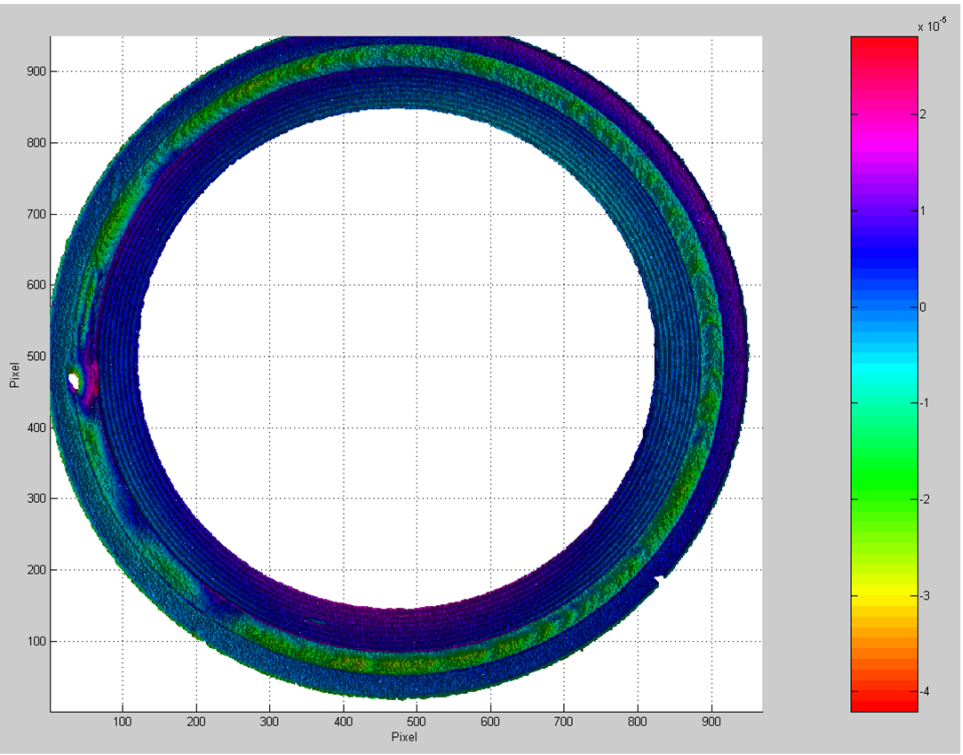


Figure C.3.: Measurement of the same laser weld as shown in Figure C.2, obtained by using a white-light interferometry system. The height is color-coded, the scale (in m) is given by the color bar. Height differences are clearly visible where a welding error occurred.

List of Figures

List of Tables

Bibliography

Aboutanios, E. & Mulgrew, B. (2005). Iterative frequency estimation by interpolation on Fourier coefficients. *IEEE Trans. Signal Process.*, *53*(4), 1237–1242.

Atkinson, G. & Hancock, E. R. (2006). Recovery of surface orientation from diffuse polarization. *IEEE Trans. Image Process.*, *15*(6), 1653–1664.

Azzam, R. M. A. & Bashara, N. M. (1987). *Ellipsometry and Polarized Light*. Amsterdam: North-Holland.

Bell, K. L., Steinberg, Y., Ephraim, Y., & Trees, H. L. V. (1997). Extended Ziv–Zakai lower bound for vector parameter estimation. *IEEE Trans. Inf. Theory*, *43*(2), 624–637.

Born, M. & Wolf, E. (1999). *Principles of Optics* (7th ed.). Cambridge: Cambridge University Press.

Brown, T. & Mao Wang, M. (2002). An iterative algorithm for single-frequency estimation. *IEEE Trans. Signal Process.*, *50*(11), 2671–2682.

Carre, P. (1966). Installation et utilisation du comparateur photoelectrique et interferentiel du bureau international des poids et mesures. *Metrologia*, *2*, 13–23.

Cheng, Y. Y. & Wyant, J. C. (1985). Phase shifter calibration in phase shifting interferometry. *Applied Optics*, *24*, 30–49.

Christensen, M. & Jensen, S. (2006). On perceptual distortion minimization and nonlinear least-squares frequency estimation. *IEEE Trans. Audio Speech Language Process.*, *14*(1), 99–109.

Cook, R. L. (1986). Stochastic sampling in computer graphics. *ACM Trans. Graphics*, *5*(1), 51–72.

de Groot, P. (1995). Vibration in phase-shifting interferometry. *J. Opt. Soc. Am.*, *12*, 354–365.

de Groot, P. (1997). 101-frame algorithm for phase shifting interferometry. In *Proc. SPIE Vol. 3098, Optical Inspection and Micromeasurements II*, (pp. 283–292).

Ettl, P. (2001). *Über die Signalenstehung bei Weißlichtinterferometrie*. PhD thesis, Friedrich-Alexander University Erlangen-Nuremberg.

George, N. & Jain, A. (1973). Speckle reduction using multiple tones of illumination. *Appl. Opt*, *12*(6), 1202–1212.

George, N., Jain, A., & Melville, R. (1975). Experiments on the space and wavelength dependence of speckle. *Applied Physics A: Materials Science & Processing*, *7*(3), 157–169.

Goodman, J. W. (1975). *Laser speckle and related phenomena*. Berlin: Springer.

Hariharan, P., Oreb, B. F., & Eiju, T. (1987). Digital phase-shifting interferometry: A simple error-compensating phase calculation algorithm. *Appl. Opt.*, *26*, 2504–2506.

Häusler, G. (1999). Three-dimensional sensors — potentials and limitations. In B. Jähne, H. Haussecker, & P. Geissler (Eds.), *Handbook of computer vision and applications*, volume 1 (pp. 485–506). San Diego: Academic Press.

Hering, M. (2007). *Angewandte statistische Optik in der Weißlicht-Interferometrie: Räumliches Phasenschieben und Einfluss technischer Oberflächen*. PhD thesis, University of Heidelberg.

Hissmann, M. (2005). *Bayesian Esitmation for White Light Interferometry*. PhD thesis, University of Heidelberg.

Höfer, V. (1994). Weißlichtinterferometrie mit dispersion: Ein aperturunabhängiges verfahren zur entfernungsmessung an optisch rauhen oberflächen. Master's thesis, Friedrich-Alexander University Erlangen-Nuremberg.

Jähne, B., Haußecker, H., & Geißler, P. (Eds.). (1999). *Handbook of Computer Vision and Applications*. Academic Press.

Jain, V. K., Collins, W. L., & Davis, D. C. (1979). High-accuracy analog measurements via interpolated FFT. *IEEE Trans. Instrum. Meas.*, *28*, 113–122.

Klenke, F. (2007). A multiple wavelength interferometer for inspection of rough surfaces: Measurements, evaluation, data processing algorithms. Master's thesis, University of Osnabrück.

Larkin, K. (1996). Efficient nonlinear algorithm for envelope detection in white light interferometry. *J. Opt. Soc. Am.*, *13*, 832–843.

Lemma, A., van der Veen, A.-J., & Deprettere, E. (2003). Analysis of joint angle-frequency estimation using ESPRIT. *IEEE Trans. Signal Process.*, *51*(5), 1264–1283.

Miyazaki, D. & Ikeuchi, K. (2005). Inverse polarization raytracing: Estimating surface shapes of transparent objects. In *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, (pp. 910–917)., San Diego.

Moon, T. K. & Stirling, W. C. (2000). *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, New Jersey, USA: Prentice Hall.

Nagengast, W. (1995). Entwicklung von durchstimmbaren externen Resonatoren hoher Stabilität für Halbleiterlaser. Master's thesis, Friedrich-Alexander University Erlangen-Nuremberg.

Noels, N., Steendam, H., Moeneclaey, M., & Bruneel, H. (2005). Carrier phase and frequency estimation for pilot-symbol assisted transmission: bounds and algorithms. *IEEE Trans. Signal Process.*, *53*(12), 4578–4587.

Oliphant, T. (2006). Optimal sampling for single tone frequency estimation. *unpublished*.

Oppenheim, A. & Schafer, R. (1989). *Discrete-time signal processing*. Upper Saddle River, NJ, USA: Prentice-Hall.

Pavlíček, P. (1999). *Das "Dispersionsradar" — ein hochgenauer faseroptischer Sensor zur Abstandsmessung*. PhD thesis, Friedrich-Alexander University Erlangen-Nuremberg.

Pavlíček, P. & Soubusta, J. (2003). Theoretical uncertainty of white-light interferometry on rough surfaces. *Applied Optics*, *42*, 1809–1813.

Pfaff, T. (2005). Frequency estimation algorithms for multiple wavelength interferometry. Internal report.

Photonfocus (2005). Photonfocus technical documentation. `http://www.photonfocus.com`.

Pisarenko, V. F. (1973). The retrieval of harmonics from a covariance function. *Geophys. J. Roy. Astr. Soc.*, *33*, 347–366.

Poor, H. V. (1994). *An Introduction to Signal Detection and Estimation*. New York, USA: Springer.

Quinn, B. G. (1994). Estimating frequency by interpolating using Fourier coefficients. *IEEE Trans. Signal Process.*, *42*, 1264.

Quinn, B. G. & Hannan, E. J. (2001). *The Estimation and Tracking of Frequency*. Cambridge, UK: Cambridge University Press.

Rahmann, S. & Canterakis, N. (2001). Reconstruction of specular surfaces using polarization imaging. In *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, (pp. 149–155)., Hawaii.

Restle, J. (2003). *Optimierung der Weisslichtinterferometrie für Applikationen der industriellen Qualitätskontrolle*. PhD thesis, University of Heidelberg.

Rhodes, H., Agranov, G., Hong, C., Boettiger, U., Mauritzson, R., Ladd, J., Karasev, I., McKee, J., Jenkins, E., Quinlin, W., et al. (2004). CMOS imager technology shrinks and image performance. In *Proc. IEEE Workshop Microelectronics and Electron Devices*, (pp. 7–18).

Rice, F., Cowley, B., Moran, B., & Rice, M. (2001). Cramér-Rao lower bounds for QAM phase and frequency estimation. *Communications, IEEE Transactions on*, *49*(9), 1582–1591.

Rife, D. & Boorstyn, R. (1974). Single tone parameter estimation from discrete-time observations. *IEEE Trans. Inf. Theory*, *20*(5), 591–598.

Rife, D. & Vincent, G. (1970). Use of the Discrete Fourier Transform in the measurement of frequencies and levels of tones. *Bell Syst. Tech. J*, *49*(2), 197–228.

Salvadé, Y. (1999). *Distance Measurement by Multiple-Wavelength Inteferometry*. PhD thesis, University of Neuchâtel.

Savaresi, S., Bittanti, S., & So, H. (2003). Closed-form unbiased frequency estimation of a noisy sinusoid using notch filters. *IEEE Trans. Autom. Control*, *48*(7), 1285–1292.

Schoukens, J., Pintelon, R., & Van Hamme, H. (1992). The interpolated Fast Fourier Transform: a comparative study. *IEEE Trans. Instrum. Meas.*, *41*(2), 226–232.

Schwartz, R., Heinol, H., Buxbaum, B., Ringbeck, T., Xu, Z., & Hartmann, K. (1999). Principles of three-dimensional imaging techniques. In B. Jähne, H. Haussecker, & P. Geissler (Eds.), *Handbook of computer vision and applications*, volume 1 (pp. 463–484). San Diego: Academic Press.

Seiffert, T. (2007). *Verfahren zur schnellen Signalaufnahme in der Weißlichtinterferometrie*. PhD thesis, Friedrich-Alexander University Erlangen-Nuremberg.

So, H.-C. (2005). On linear least squares approach for phase estimation of real sinusoidal signals. *IEICE Trans. Fundamentals*, *E88-A*(12), 3654–3657.

SOPRA (2005). SOPRA-NK database. `http://www.sopra-sa.com`.

Stoica, P. & Soderstrom, T. (1991). Statistical analysis of MUSIC and ESPRIT estimates of sinusoidal frequencies. In *Proc. Int. Conf. Acoustics, Speech and Signal Processing*, (pp. 3273–3276).

Teague, C. (2002). Root-MUSIC direction finding applied to multifrequency coastal radar. In *IEEE Int. Geoscience Remote Sensing Symp.*, volume 3, (pp. 1896–1898).

Vanlanduit, R., Vanherzeele, J., Guillaume, P., Cauberghe, B., & Verboven, P. (2004). Fourier fringe processing by use of an interpolated Fourier-transform technique. *Appl. Opt.*, *43*, 5206–5213.

Wagner, S. (2005). Simulation und Untersuchung von neuen Verfahren für die schnelle Berechnung von Höhenkarten basierend auf Weißlichtinterferometrie. Master's thesis, Esslingen University of Applied Sciences.

Walker, M. J. (1954). Matrix calculus and the Stokes parameter of polarized radiation. *Am. J. Phys.*, *22*, 170–174.

Wieler, M. (2006). Single tone frequency estimation from very few sampling points. Master's thesis, University of Heidelberg.

Wieler, M., Trittler, S., & Hamprecht, F. (2006). Optimal design for single tone frequency estimation. *unpublished*.

Williams, S., Shalf, J., Oliker, L., Kamil, S., Husands, P., & Yelick, K. (2006). The potential of the cell processor for scientific computing. In *Proceedings of the 3rd conference on computing frontiers*, (pp. 9–20)., New York. ACM Press.

Wolff, L. & Boult, T. (1991). Constraining object features using a polarization reflectance model. *IEEE Trans. Pattern Anal. Mach. Intell.*, *13*(7), 635–657.

Wolff, L. B. (1994). Polarization camera for computer vision with a beam splitter. *J. Opt. Soc. Am.*, *11*, 2935–2945.

Wyant, J. (1982). Interferometric optical metrology: basic principles and new systems. *Laser Focus*, *5*, 65–71.

Zhang, F., Geng, Z., & Yuan, W. (2001). The algorithm of interpolating windowed FFT for harmonic analysis of electric power system. *IEEE Trans. Power Del.*, *16*(2), 160–164.