Auto calibration of incremental analog quadrature encoders

M.J.A. van Kuijk
DCT 2009.053
Preface

This report is the result of my graduation project for becoming a Master of Science. Before enrolling the Mechanical Engineering Master program at the Technical University of Eindhoven, I was a Bachelor student at the Technical College in ’s-Hertogenbosch. During this study I became very interested in the field of Mechatronics. After finishing this education with a graduation project at Océ Technologies I was only more intrigued, therefore I became a student at the TU/e and enrolled in the Control Systems Technologies group.

One obstacle in this research was especially difficult to tackle. On a Friday afternoon I was trying to overcome this problem when I finally had to conclude that a certain method didn’t work, since it wasn’t the first, I was very disappointed. To ease my mind I started to make a graphical presentation of the problem and trying some equations, not confident it would result in anything useful. After some missed attempts I finally came up with an idea that made me so enthusiastic I couldn’t hide it, which made my fellow students make fun of me. This was for me the highlight of my project.

I would like to thank René van de Molengraft for his direct response and supervision of this project. A great deal of thanks goes out to Roel Merry, for being ever enthusiastic, especially in times when I wasn’t. Also for his insights and help, which were always available on short notice.

Another word of thanks is for my family who fully supported my decision to do a second study.

A special word of gratitude goes to my girlfriend “Cecile”. Your patient support has been remarkable. Now my project is finished, I hope to join you in some of yours.
Summary

Incremental encoders are widely used in many applications to determine position. Most linear and rotary incremental encoders generate two quadrature signals. The common way to extract position information from these periodical signals is to count their zero crossings with the use of a counter card. The advantage of this method is that it is a very practical method, the limitation is that it does not use all the position information available in the quadrature signals. The quadrature outputs between the zero-crossings also contain position information. The position can be derived with a method, referred to as the arctangent method, at every sample of the quadrature signal.

The two quadratures are analog signals which contain four errors, referred to as narrow angle errors which reduce the position accuracy. The goal of this thesis is to develop an auto calibration method, which reduces these errors without the need of a reference encoder. The errors can be seen as the deviations from a sine and cosine, when noise is neglected. The errors are: amplitude distortion, mean offset, waveform distortion and phase offset. The first error is compensated with a very basic matlab routine. The second and third error are compensated with a novel method, the phase error is compensated with a truncated Heydemann model, which was found in literature. The compensation method or auto calibration method consist of two parts. The first is an offline error identification routine, which estimates the errors based on a data set of the quadrature signals, generated by the encoder. The second is an online compensation, which reduces the errors.

The novel compensation method is called waveform compensation, while it also compensates for the amplitude error. The mean offset is compensated prior to the waveform compensation. For low resolution encoders the quadrature signals closely resemble a sawtooth, for high resolution encoders they are more sinusoidal. A sawtooth signal can be expressed as a sum of higher spatial harmonics, which, with the aid of the Bromwich formula, can be expressed as a sum of higher order sines. An offline Fourier transformation on the quadrature signals, estimates the amplitudes of these spatial harmonics. These coefficients quantify the shape of the quadrature signal. With this information a model can be made with the quadrature signals as output and a sine as input. Online the Newton-Raphson method is implemented to do the opposite: estimate a perfect sine with the measured quadrature signals as input. The phase correction is done after the waveform correction and it compensate for the phase shift between the sine and cosine.

The result of this thesis is an auto calibration method, which estimates the narrow angle errors during a identification phase and reduces these errors online. The errors are reduced by approximately a factor 15 in rms value of the position error.
Incrementele encoder worden in veel toepassingen gebruikt voor het bepalen van positie. De meeste gebruikte encoders sturen twee analoge blokvormige signalen uit waarvan met behulp van een tellerkaart de nul doorgangen worden geteld en is daarmee een maat voor de positie. Dit is een zeer praktische methode, maar heeft als beperkingen dat het niet alle positie informatie gebruikt die voor handen is. Daarom zijn er ook encoders die twee analoge quadratuur signalen uitgeven. Dit zijn idealiter twee $90^\circ$ fase verschoven sinussen. De signaal inhoud tussen de nul doorgangen kan hier ook gebruikt worden voor de positie bepaling. Daarom heeft deze methode, afhankelijk van de sample frequentie, een hogere resolutie, wat een voordeel is van deze encoder.

Een beperkende factor voor de positie nauwkeurigheid in het gebruik van deze signalen is de aanwezigheid van de zogenaamde narrow angle errors. De doelstelling is om een auto-kalibratie methode te ontwikkelen die deze fouten reduceert zonder het gebruik van een referentie encoder. Als ruimte buiten beschouwing gelaten wordt, kunnen de fouten beschouwd worden als de afwijkingen van de quadratuur signalen op de ideale sinus en cosinus. De fouten zijn de afwijkingen in: gemiddelde waarde, amplitude, fase en de vorm van het signaal. De gemiddelde waarde wordt als eerste gereduceerd. Een nieuwe methode is ontwikkeld om de amplitude en vormfout te corrigeren. Als laatste wordt de fase fout verminderd door gebruik te maken van een gereduceerd Heydemann model, welke is gevonden in literatuur. Alle fouten worden eerst offline geidentificeerd, gebruik makend van de door de encoder zelf gegenereerde data.

De identificatie en compensatie van de amplitude en vormfout is gebaseerd op het feit dat de quadratuur signalen zijn opgebouwd uit een sommatie van oneven hogere orde sinussen. De eerste orde sinus is het ideale sinusvormige signaal en dus de gewenste output. Een Fast Fourier Transformatie is gebruikt om offline de amplitudes van de hogere orde sinussen te schatten. Met behulp van de Bromwich formule wordt het quadratuur signaal gemodelleerd. Online wordt een Newton-Raphson schatting toegepast om met behulp van het model, aan de hand van het quadratuur signaal, de ideale sinus en cosinus te schatten. De fase identificatie en compensatie is gedaan met behulp een gereduceerd Heydemann model. In deze compensatie wordt de geschatte sinus in fase verschoven zodanig dat deze in fase is met de cosinus, oftewel dat het resultaat twee $90^\circ$ fase verschoven sinussen zijn. De positie wordt berekend met de zogenaamde vier quadrant arctangens berekening.

Het resultaat van dit rapport is een auto-kalibratie methode die de narrow angle errors reduceert zonder het gebruik van een referentie encoder. De positie-fout reductie in rms waarde is een factor 15.
Contents

Prelude .................................................. i

Summary ................................................. iii

Summary Dutch .......................................... v

1 Introduction .......................................... 1
  1.1 Encoders ........................................... 1
  1.2 Research objective ................................. 2
  1.3 Approach ........................................... 3
  1.4 Report outline ..................................... 3

2 Incremental encoders ................................. 5
  2.1 4-Fold measuring principle ......................... 5
  2.2 Classical encoder output ............................ 8
  2.3 Using the analog signals ........................... 9
  2.4 Encoder errors .................................... 11

3 Literature study ....................................... 15
  3.1 Kalman filter ..................................... 15
  3.2 Neural network .................................... 15
  3.3 Heydemann method ................................ 16
  3.4 Look-up table .................................... 16
  3.5 Wavelets .......................................... 16
  3.6 Lock-in-Amplifiers ................................ 17
  3.7 Conclusion ........................................ 17

4 Auto calibration ....................................... 19
  4.1 Waveform correction ............................... 19
  4.2 Phase correction .................................. 25
  4.3 Position .......................................... 27

5 Implementation and results ......................... 29
  5.1 Implementation .................................... 29
  5.2 Results ............................................ 32
Chapter 1

Introduction

In many industries engineers are challenged to meet the ever increasing demands of the customer. Therefore, next generation products need to be smaller, faster and lighter. These market demands have their reflection on production machinery. So engineers try to push the envelope of operating speeds and accuracy. The highest feasible accuracy in production can be restricted by the quality of the measuring equipment. A very commonly used type of measuring equipment in industry is an encoder, which is used to determine position or speed. Pushing the envelope of encoder accuracy is therefore a very interesting subject.

1.1 Encoders

Encoders are applied in a wide variety of situations. That is why there is more than one type of encoder. Encoders can be distinguished by application and by type of output. There are encoders which produce an absolute position value and encoders which produce an incremental position value, only the latter is considered.

There are linear and rotary encoders. Both can be based on either the image scanning principle or interferential scanning principle. All of these encoders are commercially available with an analog quadrature output, which are ideally a sine and cosine with an amplitude of 0.5V, referred to as $1V_{pp}$, 1 Volt from peak to peak. The image scanning principle can be divided in a 4-fold scanning principle and in a single field scanning principle. which is more robust in the presence of contamination in open linear encoders. Both principles produce the same output. The interferential scanning principle is used for the high resolution encoders, which are also commercially available with a $1V_{pp}$ output. This report is about analog incremental encoders.

1.1.1 Analog incremental encoders

Rotary incremental encoders translate mechanical rotation of the axes into a voltage or current, which in turn is used to determine the position of the axis. This translation is done optically. Fig. 1.1 gives a schematic overview of this principle. This is a schematic overview of the 4-fold image scanning principle, which is the most used principle. The focus in this thesis is on the 4-fold principle.

The rotating encoder disk contains opaque and dark lines, which in combination with the light source result in a time-varying light intensity on the detectors. The detectors translate these
Chapter 1. Introduction

Figure 1.1: 4-fold measuring principle of a rotary encoder.

time-varying light intensities into two ideally sinusoidally shaped voltage signals. The most common way to determine position is to count the zero crossings of these analog outputs. This is a very practical method and therefore the most used method, see [4]. It is also a conservative method, in the sense that it can only determine position on the zero crossings. It only has one position value per line transition on the disk, all information in between the lines is neglected. In industry, counting zero crossings is often referred to as the "TTL" (Transistor-Transistor logic) method. To recap, only counting zero crossings is a practical but conservative position determination.

Instead of the TTL method, the analog signals can be measured by a data-acquisition system and then position information is available quasi-continuously. This results in a much higher position resolution only limited by the sample frequency of the data-acquisition equipment and operating speed. The accuracy of these signals is limited by the presence of errors. Reducing these errors is the main challenge in this research.

1.2 Research objective

The encoder errors can be divided into two groups by their frequency content, namely *wide angle errors* and *narrow angle errors*, which respectively have a low and high frequency content. The wide angle (WA) errors are related to the mechanical rotation frequency of the encoder, the narrow angle (NA) errors to the frequency of the quadrature signals which is equal to the mechanical frequency times the resolution of the encoder. The NA errors are identical for rotary and linear encoders while the (WA) errors are not. The narrow-angle errors are of main interest in this paper, the wide angle errors are not treated.

The objective of this study is to develop an auto calibration method, which minimizes the NA errors and determines the position with the use of the analog outputs and without the use of a reference encoder.
1.3 Approach

The sources of the narrow-angle errors are studied by means of literature, signal analysis and photography on microscopic level of the encoder lines. The effects on position are simulated. This is the basis for the development of the compensation method, which is a combined result of literature study and a complete novel idea. While the focus is on 4-fold measurement principle, this method is also applicable to the other analog quadrature encoders given that the NA errors are alike.

1.4 Report outline

Chapter 2 starts with a description of principle of operation of an incremental encoder. To understand the encoder errors, its sources are studied and revealed at the end of Chapter 2. Chapter 3 is all about the literature study. Several other error correction methods are compared as well as theories from other fields. With the use of the literature study and a novel idea, the narrow-angle encoder errors are corrected in Chapter 4. This chapter has a theoretical nature as apposed to Chapter 5 where the implementation and experimental results are discussed. This report ends with Chapter 6, giving conclusions and recommendations.
Chapter 2

Incremental encoders

The focus in this report is on 4-fold optical incremental encoders with analog output. The measurement principle and the sources of the errors are discussed. At the end of this chapter the effects of the errors on the position are shown by means of simulation results.

2.1 4-Fold measuring principle

The 4-fold measuring principle is simply named after the four scanning fields used in most rotary encoders, and most of the closed linear encoders. In the next section the layout of these encoders is explained.

2.1.1 Encoder layout

An incremental encoder translates mechanical rotation or translation into position information. This is done contact free with the use of light. A schematic overview of the principle is given in Fig. 2.1 in which L is a light source, usually a Light Emitting Diode or LED. The light is passed through a condensor lens K, which makes sure the light beam is homogeneous. The scanning reticle A is fixed on the encoder housing as are all the other parts with exception of the scale M, which is drawn as a linear scale. For a rotary encoder this disk is circular. The reticle and scale disk are typically made of glass and have a, usually chrome, coating where a regular pattern is etched out, resulting in a pattern of transparent and opaque radial lines.

The light beams which passes through the static reticle and moving disk illuminates the photo active cells P. The photo cells translate the received light into a micro current and by operational amplifiers into a voltage output signal. The disk M is (in)directly fixed to the moving object of interest. When the disk moves, the illuminated surface on the photo-active cells changes over time. Hence the output voltages of the photo cells change. This change in voltage is a measure for the movement.

2.1.2 Illumination principle

To distinguish between different positions, the reticle and disk have a grid of opaque and light emitting lines, see Fig. 2.3. The number of lines on the disk determines the resolution of the encoder. If the emitting and opaque lines on the disk and reticle all would have the same width, the encoder would have a triangular shaped output voltage in time as in Fig. 2.3. Perhaps this becomes more clear by an example: If in an apartment house everyone would close their curtains
simultaneously, the summed light intensity in that building would decay linearly in time. This would not be when some curtains were already slightly closed. The next section addresses this issue in more depth.

### 2.1.3 Moire effect

A mathematically more advantageous signal for position determination, is a sinusoid. To enforce a sinusoidal shape, the moire effect is used. For more information on the moire effect, see [23] or [2]. To incorporate this in the encoder, the width of the opaque and transparent lines on the disk and reticle is varied, see [10]. Every encoder can have a different combination of line widths and therefore, different output signals can vary in shape from a sawtooth to a sinusoid.

A sawtooth shaped signal \( q(t) \) can be expressed as a sum of sines, see (2.1), with speed \( \omega \) and time \( t \). The Fourier coefficients \( F_n \), are a quantification of the shape of the waveform. Remark that if, for \( n = 3, 5, 7, \ldots, \infty \), \( F_n = 0 \), \( q(t) = \sin(\omega t) \). Fig. 2.2 shows the influence of the Fourier coefficients (on the left) on the shape of the output signals \( q(t) \). Moire developed a model with the Fourier coefficients as output and the line widths and number of lines on the reticle as input, see [23] or [2]. It is beyond the scope of this thesis to explain the Moire effect. In practice can be stated that high resolution encoders have a more sinusoidal output and low resolution encoders a more triangular shaped output.

\[
q(t) = \sum_{n=1}^{\infty} F_n \sin((n)\omega t) \quad \text{for} \quad n = 1, 3, 5, \ldots, \infty \tag{2.1}
\]
2.1. 4-Fold measuring principle

\[ \log(F_n) \]

\[ q(t) \]

Figure 2.2: Influence of Fourier coefficients on the shape of the output signal.

Figure 2.3: Reticle and disk with identical ronchi rulings.
2.1.4 From 4-fold to two quadrature signals

The four photo cells supply four voltage signals, originally named: 4-fold signals. These are all positive signals. One electrical cycle or waveform corresponds to one set of lines of the scanning reticle, one dark and one opaque line. The four scanning reticles are, ideally, all radially interspaced half a line width. This results in the 90° phase shift between the 4-fold signals. To enforce symmetry around the horizontal (time) axis these four signals are paired into two voltage signals according to Fig. 2.4 [10]. See (2.2) for pairing of one quadrature signal $V_1$, with 4-fold voltages $V_{0°}$ and $V_{180°}$, offset $B$, amplitude $A$, velocity $\omega$ and time $t$. The resulting two voltage signals are called quadrature signals because they are shifted 90° degrees. The advantage of the zero mean offset quadrature signals is, that in case of led aging the signals only change in amplitude and not in mean offset.

\[
V_1 = V_{0°} - V_{180°} = b + a \cdot \sin(\omega t) - (b - a \cdot \sin(\omega t)) = 2a \cdot \sin(\omega t).
\]
\[
V_2 = V_{90°} - V_{270°} = d + c \cdot \sin(\omega t) - (d - c \cdot \sin(\omega t)) = 2c \cdot \sin(\omega t). \tag{2.2}
\]

The quadrature signals can be used to determine the position. Both the classical technique of counting the zero-crossings and a more direct arctangent method are explained below.

2.2 Classical encoder output

The most common way of extracting position from the two quadrature signals, is counting their zero-crossings. To do this, an analog network is incorporated in the encoder which translates the two quadrature signals into block signals, see Fig. 2.5. A counter card is needed to detect the edges of the block signal and count the pulses. For every encoder slit, four zero crossings are detected. To detect the direction of rotation, two signals are used. Figure 2.5 shows two block signals. At $t = 1.25$, the last counted pulse is from the zero crossing of $V_1$. If then the direction
would change, the next pulse would again be produced by \( V_1 \). Hence the last two pulses came from one quadrature signal and that is how change in direction is detected.

This method gives position information on four points per quadrature period while the analog signals contain position information in quasi-continuous time. For this reason, it is restrictive, hence interesting to look at a different method which is able to determine the position in between the zero-crossings. The number of samples per period of \( V_1 \) and \( V_2 \) are determined by \( \frac{F_s}{F_c} \), with sampling frequency \( F_s \) and \( F_c \) the frequency of \( V_1 \) and \( V_2 \).

2.3 Using the analog signals

This section explains the **arctangent computation** which calculates position from the analog outputs. The analog signals are fed to an analog digital converter and the position is calculated for every input sample of the quadrature signal.

2.3.1 Arctangent computation

Assuming that the quadrature signals are perfect sinusoids, the unwrapped four quadrant inverse tangent of these signals calculates the position faultless. See Fig. 2.6(b) for an example with a 10 Hz sine and cosine and position in line count. Eq. \( 2.3 \) calculates the electrical angle \( \theta \) and \( 2.4 \) the position \( x \). \( L \) is the number of lines of a rotary encoder disk and the position is given in revolutions. The unwrap function, is a built in matlab function which changes absolute jumps \( \geq \pi \) to their \( 2\pi \) complement.
2.3.2 Lissajous figure

The quadrature signals $V_1$, $V_2$ and arctangent in Fig. 2.6(a) and 2.6(b) can also be presented as in Fig. 2.7, called a Lissajous figure. The cosine is plotted on the x-axis and sine on the y-axis. The angle between the vector and the x-axis, is defined as $\theta$ and given in (2.3).

$$\theta = \tan^{-1} \left( \frac{V_2}{V_1} \right) \quad [\text{elec.rad}]$$  \hspace{1cm} (2.3)

$$x = \frac{\text{unwrap}(\theta)}{(2\pi L)} \quad [\text{rev}]$$  \hspace{1cm} (2.4)

The quadrature signals are not a perfect sine and cosine, as described in Section 2.1.3. They are also subjected to other errors, resulting in a shifted non-circular Lissajous figure and in a corrupted angle, hence corrupted position. How the encoder errors affect the position is described in the next section.
2.4 Encoder errors

The encoder errors can mainly be separated into two groups by their frequency content [15]. A group with their frequency content in the range of the mechanical rotation and are called wide angle errors and are only present in rotary encoders. The frequency band of the other group is near the frequency of the quadrature signals and are called narrow angle errors. The only error not be divided in a group is the error due to non-equidistant code slits, since this error has no fixed frequency content.

2.4.1 Non-equidistant code slits

Because of non perfect manufacturing of the code disk and reticle, the line widths may vary, which results in a deterministic deviation in position. For every travel of the encoder this error is repeatable and has a diverse frequency content.

2.4.2 Wide angle errors

This is a group of various mechanical imperfections resulting in errors with a frequency content of \( n f_m \) for \( n = 1,2 \). This group is specific for rotary encoders.

Eccentricity error

These errors are a result of imperfect manufacturing and imperfect mounting of the encoder, neither can be prevented. An example of imperfect manufacturing is eccentricity of the code disk, housing and shaft. While tolerances in manufacturing can be calibrated by the encoder supplier, the eccentricity due to mounting can only be calibrated after the final mounting. This would mean that after assembly a more accurate reference (encoder) should be used. The error in position has a typical frequency content of \( 1 f_m \) Hz, with \( f_m \) being the mechanical rotation frequency. The wide angle errors in the experimental setup are given in App. B.

Ellipticity of the code disk

This is a result of imperfect manufacturing. The code disk is oval instead of circular. A more detailed description is given in App. B. The error in position has a typical frequency content of \( 2 f_m \) Hz.

2.4.3 Narrow angle errors

These four errors describe the difference between the quadrature signals and two perfect by \( 90^\circ \) shifted sinusoids. The errors given in Fig. 2.8 are amplitude distortion \( \frac{l_1}{l_2} - 0.5 \), mean offset \( \frac{|l_1 - l_2|}{2} \), phase offset \( \frac{|\phi_1 + \phi_2|}{2} - 90^\circ \) and waveform distortion. The Lissajous figure is given in Fig. 2.7 and gives a clear graphical representation of how these errors effect the angle in the Lissajous figure and hence the position.

Below the four NA errors are described by means of a simulation with 10Hz quadrature signals.
Amplitude distortion

The output of the photo diodes is amplified by an analog circuit, see [11]. Typical output voltages are 0.6 to 1.2 Volt from peak to peak for the quadrature signals [11]. The amplitude ratio is constant in time and typically 0.8 to 1.25 and can be a result of electrical errors and deviations in the reticle. The circular shape in the Lissajous figure now becomes an ellipse as in Fig. 2.9(a). The error in position has a typical frequency content of $2nf_c$ for $n = 1, 2, 3, \ldots \infty$, with $f_c = \frac{f_m}{L}$ being the base frequency of the quadrature signals. The simulated error in position due to amplitude distortion is given by the Fast Fourier Transformation (FFT) in Fig. 2.9(b). The quadrature signals have a frequency of $f_c = 10$Hz. The error components are clearly visible at 20, 40, 60 and 80Hz. The y-axis gives the error in radials where $2\pi$ radials corresponds to one period of the waveform. This is also called the electrical angle or electrical period. The amplitude ratio $\frac{l_3}{p_3}$ is determinative for the error. For this simulation is arbitrarily chosen to only alter the amplitude of the sine.
2.4. Encoder errors

Mean offset

Due to differences in photo diodes and analog circuitry the 4-fold signals may have different mean values, therefore the quadrature signals may have non-zero mean values. Typical values are \(|V_0| \leq 0.065\) V.

The Lissajous figure is shifted due to the mean offset. The error in position has a typical frequency content of \(nf\) for \(n = 1, 2, 3 \ldots \infty\). Again simulation results are provided in Fig. 2.10(a) and Fig. 2.10(b) For this simulation is arbitrarily chosen to only give the cosine a non-zero mean offset. The error components are clearly visible at 10, 20, 30, ... Hz.

Phase offset

Due to imperfect interspacing of the reticles, see Section 2.1.4, the 4-fold signals may be shifted more or less than 90°. Therefore the quadrature signals are effected accordingly and typically phase shifted 90° ± 10°.

The ideal circular Lissajous figure is now distorted into an ellipse which is rotated around the origin, see Fig. 2.11(a).

The error in position has a typical frequency content of \(2nf\) for \(n = 1, 2, 3 \ldots \infty\) which is the same as for the amplitude distortion, see Fig. 2.11(b).

Waveform distortion

As described in Section 2.1.3, the waveforms can be sinusoids in case of a high resolution encoder, but most often they have a more triangular shape. This distortion in waveform gives the Lissajous figure a block shape, see Fig. 2.12(a). The error in position has a typical frequency content of \(nf\) for \(n = 4, 8, 12, 16 \ldots \infty\), see Fig. 2.12(b).
Figure 2.11: Phase distortion.

Figure 2.12: Waveform distortion
Chapter 3

Literature study

This chapter gives an overview of the most interesting methods for compensation of the narrow angle errors as described in Chapter 2.

3.1 Kalman filter

A standard Kalman filter can be used to reduce measurement error, see [13], for background information in English or [14] for Dutch. In [24], a Kalman filter is used to reduce deterministic and random measurement errors, with a deterministic error model. The misalignment of the masks is the deterministic error, which results in the phase error. The encoder is used to determine the radial position of a telescope. The inputs of the Kalman filter are the measured quadrature signals, the output is the radial position of the encoder.

A reduction in position deviation of 30% to 40% is claimed in [24]. Which is moderate, but considering only the error due to the misalignment of the masks is handled as a deterministic error, there is room for improvement. If the other errors can be modeled this could be used in the Kalman filter to improve position accuracy.

To conclude, the idea of using a Kalman filter is good in the sense that it reduces both deterministic and random errors. To apply a Kalman filter on the quadrature signals to reduce narrow angle errors and random noise, an error model should be made.

3.2 Neural network

The development of an adaptive online approach for the correction and interpolation of quadrature encoder signals, is described in [21]. It is based on the use of a two-stage double-layered radial basis function (RBF) neural network. The first RBF stage is used to correct for the imperfections in the encoder signals such as mean, phase offsets, amplitude deviation and waveform distortion. The second RBF stage serves as the inferencing machine to adaptively map the quadrature encoder signals to higher order sinusoids, thus, enabling intermediate positions to be derived.

The adaption algorithm to calculate the free parameters of the first RBF network has two sets of inputs. The first is the output of the RBF network, the corrected quadrature signals. The second set are the ideal quadrature signals. Since these two are subtracted they have to be synchronized. The compensation stage needs a training phase. During this phase the ideal quadrature signals have to be fed to the stage in order to train the parameters correctly. During this stage the quadrature signals can’t be synchronized. This problem can be compared with the problem of making a
look-up table without a reference encoder. The performance of the total RBF network is given by means of a tracking error of an experimental setup and therefore the improvement due to this method is not very transparent, since the tracking error without the correction stage is not given. The performance of the interpolation stage is given in a step response experiment, and position error is reduced approximately a factor five. Therefore this could be an interesting method. If this paper is to be used for further research, the addition is to be made in the direction of the synchronization. No significant other additions can be made in the compensation stage. A possible addition for the interpolation stage is to use the arctangent method as described in Section 2.3.1.

### 3.3 Heydemann method

Peter Heydemann developed a method in 1981 [12] to model and cancel errors in interferometers. The interferometer outputs are a sine and cosine and have three errors: mean offset, amplitude distortion and phase offset. The only difference with analog encoder signals is the waveform distortion. The Heydemann concept consists of an offline identification and an online compensation. The distorted quadrature signals are modeled and with the criterium or cost function, that the ideal Lissajous figure has a constant radius, the errors are identified offline. The inverse error model is used to correct the measured signals online. Simulations have shown the effectiveness of this method, see [C]. However the phase offset correction does not work for non-sinusoidal signals, hence before Heydemann correction the waveforms need to be corrected. Which has been a reason in previous research to drop this method, see [21] and [22] for some examples. Therefore addressing this issue is very interesting.

### 3.4 Look-up table

A more straightforward and well known method is the look-up table, as described in [21]. This method compensates for all four narrow angle errors. The table is constructed for one quadrature period, not for the entire rotation of the encoder. It contains the information to map the raw signals to the ideal signals. It also interpolates the ideal signals into higher order sinusoids and converts those into pulses. This is all done within the look-up table. The look-up table is constructed offline. A drawback of look-up tables is, that in order to make the mapping from measured to ideal signals a reference is needed. Constant speed control can not be done with the uncorrected quadrature signals of the encoder, since they contain errors. Open loop constant speed steering is not always an option in practice and constant speed can’t be guaranteed. The main drawback of a look-up table is, if it is generated without a reference encoder the disturbances in the setup will be put in the look-up table. This results in the encoder giving an inaccurate position. Therefore look-up tables are of no further interest.

### 3.5 Wavelets

Wavelets can be used to cancel the waveform distortion. For more information on wavelets and filter banks the reader is referred to [16] or [20]. As described in Section 2.1.3 the analog encoder signals can be described as a sum of sines. By making an online filter bank it is possible to extract
the first order sine and cosine from the quadrature signals. An online continuous wavelet transformacija as described in [8] is also tested. Simulations have shown that the bottleneck in this application is the delay which is typically half the length of the wavelet, see also [8]. Even for very short bell shaped wavelets the delay is unacceptable. Filter banks have a similar delay problem. Another idea, which was not found in literature in relation with this project, is to use a single Finite Impulse Response (FIR) band-pass filter to extract the base frequency of the quadrature signals. Even when it would be possible to adapt the frequency band online as the encoder accelerates, the phase drop is a big disadvantage in a feedback loop. Another problem would occur at very low speed, where the higher spatial harmonics would be very close to the first harmonic and impossible to separate.

3.6 Lock-in-Amplifiers

The operation of a lock-in amplifier relies on the orthogonality of sinusoidal functions. Specifically, when a sinusoidal function of frequency $x$ is multiplied by another sinusoidal function of frequency $y$ not equal to $x$ and integrated over a time much longer than the period of the two functions, the result is zero. In the case when $x$ is equal to $y$, and the two functions are in phase, the average value is equal to half of the product of the amplitudes. In essence, a lock-in amplifier takes the input signal, multiplies it by the reference signal (either provided from the internal oscillator or an external source), and integrates it over a specified time, usually on the order of milliseconds to a few seconds. The resulting signal is an essentially DC signal, where the contribution from any signal that is not at the same frequency as the reference signal is attenuated essentially to zero, as well as the out-of-phase component of the signal that has the same frequency as the reference signal (because sine functions are orthogonal to the cosine functions of the same frequency), and this is also why a lock-in is a phase sensitive detector. The problem with applying this method to the quadrature signals to extract the first order harmonic, is that the ideal frequency is needed as an input. This is possible with the use of a frequency estimator. The drawback is that especially for low frequencies the estimates become more inaccurate. When the wrong frequency is estimated, the error will summed over time resulting in a significant error. Therefore, this method is of no further interest.

3.7 Conclusion

The interesting methods are the Heydemann method, neural networks and a Kalman filter. The first two have the most promising results, while the Heydemann method gives the most insight. The problem of synchronization during learning phase for the neural networks remains a drawback. The "black-box" concept of correcting the errors without knowing them.

Another minor is that the author doesn’t give clear performance results of the error correction. The Heydemann method has no drawbacks, therefore it is chosen over the neural networks. The Kalman filter could possibly be used in combination with the Heydemann model. The surplus value with respect to the Heydemann method would be the reduction of noise. Which partially also can be achieved by using a low-pass filter. Therefore the Heydemann method is the basis for further research. The Heydemann method needs sinusoidal inputs. Creating these inputs from the raw quadrature signals is the first step and the main challenge in this report.
Chapter 4

Auto calibration

This chapter will reveal the theory behind the compensation of the narrow angle errors, which consists of the mean offset, waveform, amplitude and the phase errors. Fig. 4.1 gives a schematic overview of the order of correction. The 4-fold signals are paired into raw quadrature signals $Q(t)$. Mean offset correction gives $q(t)$. In a single routine the waveform and amplitude corrections are performed. The Heydemann correction [12], is used in a truncated form to only compensate for the phase error, hence called phase correction. The corrected quadrature signals $v(t)$ are used to determine the position by means of the arctangent compensation, as in 2.3.

The mean offset, amplitude and waveform correction are discussed in the first section, which is called waveform correction since it was originally developed to do just that. The second section explains the phase correction.

4.1 Waveform correction

The goal is to correct the waveform such that the outcome is a sinusoid. The amplitude of the output is 0.5 V, the same as the ideal input described by [11]. While the waveform correction is developed to reshape the waveform, it also reduces the amplitude error, which will be explained later. The mean offset needs to be corrected before the waveform correction in order to work properly, this will also be explained later.

This section starts with a conceptual description of the waveform correction. The correction consists of an offline and online part and are treated accordingly.

Concept

This section will explain the idea behind the waveform correction. The complete proof is given in the next two sections. This section explains the steps needed to correct the waveform. The correction is done per quadrature signal, therefore explained for a single signal.

A sawtooth $Z(t)$, with amplitude $A$, can be expressed as a sum of sines, as in $4.1^4$. The electrical angle $\theta$ is expressed in radians per slit (line on the encoder disk). The angle $\theta = \omega(t)t$, which is respectively the frequency multiplied with the time. Since $\theta$ is a more practical choice for encoder signals, it is proportional with position, this notation will be used. Eq. $4.1$ only holds for a pure sawtooth, not for the measured quadrature signals. The measured quadrature signals $q(t)$ can also be expressed as a sum of sines, but then as in $4.2$, where $f_n$ are the Fourier
coefficients. These coefficients are the result of the Fourier decomposition and the absolute values of the measured quadrature signals are shown in Fig. 4.2. These are different for every encoder. Encoders with a higher resolution will have lower $f_n$ for $n > 1$. In (4.2), $f_n = \frac{8A}{\pi^2} \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{(2p-1)^2}$ when $q(t)$ is a sawtooth. However this is not the case for the actual measurements, therefore (4.2) is used with the experimental signals.

$$Z(t) = \frac{8A}{\pi^2} \sum_{p=1}^{\infty} \frac{(-1)^{p+1}}{(2p-1)^2} \sin((2p-1)\theta) \quad \text{for } p = 1, 2, 3...\infty \quad (4.1)$$

$$q(t) = \sum_{n=1}^{\infty} f_n \sin(n\theta) \quad \text{for } n = 1, 3, 5, ..\infty \quad (4.2)$$

The ideal quadrature signal is $v(t) = \sin(\theta)$ instead of $\sin(n\theta)$ as in (4.2). The idea is to extract $\sin(\theta)$ from $q(t)$, since that is the desired output. To do this, the first step is to remove $n$ from the argument of, $\sin(n\theta)$. This can be done with the Bromwich formula (4.5), which expresses the higher spatial harmonics $\sin(n\theta)$ as a sum of higher order sines $\sin(\theta)^n$. This enables the step from (4.2) to (4.3).

The vector $F$ (4.4), contains the Fourier coefficients, with $F(n) = f_n$. Matrix $B$ (4.6) is derived from the Bromwich formula and will be explained later, see (4.6). The order of approximation of $q(t)$ is $m = \max(n)$.

$$q(t) \approx F \cdot B \cdot [\sin(\theta)^1 \ 0 \ \sin(\theta)^3 \ 0 \ ... \ \sin(\theta)^m]^T \quad (4.3)$$

$$F = [f_1 \ f_3 \ f_5 \ ... \ f_m] \quad (4.4)$$
4.1. Waveform correction

Figure 4.2: Fourier coefficients

Remark that (4.2) contains higher spatial harmonics, while (4.3) only contains higher order sines. The desired expression is the inverted form of (4.3), with \( \sin(\theta) \) as function of \( q(t) \). The reason for this is that, with input \( q(t) \) this inverted expression will give the desired ideal output \( w(t) = \sin(\theta) \). While there is no analytical solution for this problem, there are numerical approximations.

The next section explains in detail how the Fourier vector \( F \) and Bromwich matrix \( B \) are derived offline. The section after that explains how these are used to retrieve the ideal quadrature signals online.

Offline identification

The purpose of this section is to calculate a set of correction parameters which are used in the online correction.

To identify the shape of the waveform a FFT is done on the measured quadrature signals \( q(t) \). So \( f_n \) in (4.2) quantifies how much a higher spatial harmonic \( \sin(n\theta) \) is present in \( q(t) \). The vector \( F \) multiplied with matrix \( B \) quantifies how much each higher order \( \sin(\theta)^n \) is present in \( q(t) \). The step from (4.2) to (4.3) will now be explained in more detail.

Bromwich introduced a multiple angle formula in 1991 [5], to express the higher order spatial harmonics as a function of higher order sines as:

\[
\sin(n\theta) = nx - \frac{n(n^2 - 1^2)x^3}{3!} + \frac{n(n^2 - 1^2)(n^2 - 3^2)x^5}{5!} - \frac{n(n^2 - 1^2)(n^2 - 3^2)(n^2 - 5^2)x^7}{7!} \ldots \quad \text{for } n = 1, 3, 5, \ldots, x = \sin(\theta)
\] (4.5)
This formula can also be expressed in matrix form, with matrix $B$ and vector $S$, respectively (4.6) and (4.7). Then the left hand side of (4.5) is changed due to the matrix notation, see (4.8). If $m$ is defined as the order of approximation of the sawtooth, then left triangular matrix $B$ and vectors $S$ and $F$ are respectively truncated to $B \in \mathbb{R}^{(\frac{m+1}{2}) \times (\frac{m+1}{2})}$, $S \in \mathbb{R}^{(\frac{m+1}{2})}$ and $F \in \mathbb{R}^{(\frac{m+1}{2})}$, see (4.4).

$$\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
3 & \frac{-3(3^2-1^2)}{3!} & 0 & \cdots & 0 \\
5 & \frac{-5(5^2-1^2)}{5!} & \frac{5(5^2-1^2)(5^2-3^2)}{5!} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & \vdots \\
m & \frac{-m(m^2-1^2)}{m!} & \frac{m(m^2-1^2)(m^2-3^2)}{m!} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \pm \frac{m(m^2-1^2)\ldots(m^2-(m-2)^2)}{m!}
\end{bmatrix} \quad (4.6)$$

$$S = \begin{bmatrix}
\sin(\theta) \\
\sin^3(\theta) \\
\sin^5(\theta) \\
\vdots \\
\sin^m(\theta)
\end{bmatrix}^T \quad (4.7)$$

$$\begin{bmatrix}
\sin(\theta) \\
\sin(3\theta) \\
\sin(5\theta) \\
\vdots \\
\sin(m\theta)
\end{bmatrix} = BS \quad (4.8)$$

Vector $F \quad (4.4)$, multiplied with $BS \quad (4.8)$, results in $q(t) \approx F \cdot B \cdot S(\theta) \quad (4.9)$

Extracting $\sin(\theta)$ from (4.9) is impossible, due to the higher order sines in (4.7). Therefore a numerical approximation is used to estimate $\sin(\theta)$ based on the measured quadrature signals $q(t)$. This is the subject of the next section. The vector $F_B = FB$ is the result of the offline identification. The vector $F_B$ contains the compensation parameters which are used in the online compensation.

Online numerical approximation

The goal of this section is to correct the waveform online with the use of the waveform information contained in $F_B$.

The Newton-Raphson iteration method is a very suitable numerical approximation for this purpose because of its fast convergence and very low computational effort. Therefore no alternatives are considered. See [9] or [19] for a complete description and [17] or [18] for a Dutch or English internet reference. The Newton-Raphson method will not be explained in detail here.
4.1. Waveform correction

The objective function $Y(s)$, is a simple derivation from (4.9). In (4.10), $F$ and $B$ are substituted with vector $FB = FB$. The $\sin(\theta)$ terms in vector $S$ are substituted by constants $s$. When $Y(s)$ approximates zero then $s$ approximates $\sin(\theta)$. The idea is to iteratively approximate the zero of $Y(s)$ for every measured input. Fig. 4.3 gives a graphical representation of the iteration routine.

$$Y(s) = F \cdot B \cdot S(\theta) - q(\theta) \approx 0 \quad (4.10)$$

$$Y(s) = FB[s^1 \ s^3 \ ... s^m]^T - q(\theta) \approx 0 \quad (4.11)$$

The update law for the approximation of $s$ is given in (4.12). With the derivative function according to (4.13). Fig. 4.3 shows that $Y(s_{k+1})$ is a better estimation.

$$s_{(k+1)} = s_k - \frac{Y(s_k)}{dY(s_k)} \quad (4.12)$$

$$\frac{dY(s)}{ds} = FB[1s^0 \ 3s^2 \ 5s^4 \ ... \ ms^{m-1}]^T \quad (4.13)$$

For every input sample $q(i\Delta t)$, five iterations are done and an output $s(i\Delta t)$ is given, with sample number $i$. The fixed sampling interval is 1ms. Now $\sin(\theta)$ is approximated quasi-continuously by $s(i\Delta t)$. Simulation results have shown that five iterations are sufficient to reduce the estimation error to $< 1e^{-8}$ [V] with $m = 23$. As a final step the output value of $s$ is halved so that $s = 0.5 \sin(\theta)$. This is not necessary, but it is in accordance with the ideal encoder outputs.
Considerations

Local convergence
The Newton-Raphson method is locally convergent, see [9] or [19]. As an initial estimation for \( s_1 = q(\theta) \) is used, where \( A_q = \max(q(\theta)) \), meaning that the input quadrature signal with an amplitude of one is used as a starting point to estimate \( s_5 \approx \sin(\theta) \). Simulations and tests have indicated convergence for this initial estimation.

Newton Raphson derivative issues
If the derivative function is not continuous at a certain point then convergence may fail to occur at that point, see [18]. Since (4.13) is a continuous function this problem will not occur. In general the Newton-Raphson method has a quadratic convergence. When the derivative is zero then convergence is not quadratic, but still convergent. At the tops and bottoms of the quadrature signals the derivative is zero or very close to zero. While the convergence rate is slower the initial estimation is also very close to the output value, hence a lower convergence rate is no problem and five iterations are sufficient.

Mean offset
The quadrature signals which are fed to the waveform correction need to have zero mean offset, in order for the online approximation to work properly. Because \( \sin(\theta) \) is estimated and not \( \sin(\theta) + O \), with offset \( O \). The raw quadrature signals \( Q(t) \) are mean offset corrected online before the waveform correction by (4.14). Remark that in this report \( q(t) \) are also called raw quadrature signals. Experiments have shown that the mean offset correction, in the Heydemann model of [12], becomes redundant, see Section 4.2 Phase correction.

\[
q(t) = Q(t) - O
\]

(4.14)

Amplitude
The correctness of the estimated amplitudes of the output signal \( w(t) \) is mainly dependant on the quality of the Fourier coefficients. When the implementation requirements of Section 5.1.2 are fulfilled, the amplitude correction in the Heydemann model becomes redundant. This is validated by experiments.

Correction on 4-fold or on 2 quadrature signals
The waveform correction can be done on the 4-fold or on the two quadrature signals, which are a result of the pairing of the 180° shifted 4-fold signals as in 2.2. When two sawtooth non 180° phase shifted signals are paired, the resulting signal will have a flat top. See Fig. 4.4, where 4-folds \( A_+ \) and \( A_- \) are paired into quadrature \( CH_1 \). As a rule of thumb determined by simulation, for 4-fold signals with a phase shift larger then \( \pm 2^\circ \) it is better to do the waveform corrections on the 4-fold signals. Otherwise correction can be done on the quadrature signals because then the deviation in position is marginal due to the distortion at the tops. For the remainder the waveform correction is done on the two quadrature signals, because in the experimental setup, which will be explained later, the phase shift is within the \( \pm 2^\circ \) bound.

Other quadrature signals
The main assumption made for the waveform correction is the that the quadrature signals can be described as a sum of odd spatial sines as in (4.2). For the encoder used in this report, it is proven that the quadrature signals are a sum of odd spatial harmonics, with the use of an offline wavelet decomposition. When there are even spatial sines or (un)even spatial cosines the correction is incomplete and gives inaccurate output. Nevertheless the Bromwich formula is available for all
spatial sines and cosines, therefore it should be possible to extend this method. It is however highly unlikely that any quadrature signals will contain even spatial sines or spatial cosines. In order to proof this, a wide range of encoders should be tested, this is not done. For three reasons the assumption is very plausible:

1. Even block shaped signals and skewed sawtooth shaped signals can be expressed as a sum of odd spatial sines, see [4].

2. In the experiments, an encoder with very low resolution is used, which only contains odd spatial sines. A very high resolution encoder will give an almost perfect sinusoidal output, meaning only the first spatial sine is present.

3. With only varying the Fourier coefficients in (4.2), the shape of \( q(t) \) can be changed, from a sawtooth to a sine as in Fig. 2.2 which are most likely, the shapes of the quadrature signals for medium resolution encoders.

From low resolution to high resolution encoders the quadrature signals vary from approximately sawtooth signals towards sinusoidal signals. These facts make it very presumable that every encoder quadrature signal is suitable for the waveform correction, nevertheless it is only tested on a single encoder. Even when the encoder signals would contain even spatial sines or cosines, the Bromwich formula [5] makes it possible to extend this method.

### 4.2 Phase correction

In 1981, Peter Heydemann developed a method to identify and correct quadrature measurement errors in interferometers, see [12]. This method addresses the errors due to mean offset, amplitude distortion and phase offset. The input signals for the Heydemann correction are the waveform corrected quadratures of the previous section. The idea is to model the corrupted signals as function of the ideal signals and make an inverted model, which has the corrupted signals
as input and the ideal signals as output. The phase error is estimated offline and compensated online in the inverted model. The mean offset and amplitude distortion are already reduced before or by the waveform correction.

4.2.1 Model

The ideal quadrature signals are described by Eq. (4.15). In (4.16), the corrupted quadrature signals \( w_1 \) and \( w_2 \) are a function of the ideal quadrature signals \( v_1 \) and \( v_2 \). The phase error \( \alpha \) is modeled by the trigonometric relation given in (4.17) [3]. In (4.16), \( \sin(\theta) \) and \( \cos(\theta) \) of (4.17) are respectively substituted by \( v_2 \) and \( v_1 \).

\[
\begin{align*}
    v_1 &= \cos(\theta) \\
    v_2 &= \sin(\theta)
\end{align*}
\]  

(4.15)

\[
\begin{align*}
    w_1 &= v_1 \\
    w_2 &= v_2 \cos(\alpha) - v_1 \sin(\alpha)
\end{align*}
\]  

(4.16)

\[
\sin(\theta - \alpha) = \sin(\theta) \cos(\alpha) - \cos(\theta) \sin(\alpha)
\]

(4.17)

Offline least squares fit

The goal of this section is to estimate the phase error \( \alpha \). This is done offline on a batch of waveform corrected encoder data, \( w_1 \) and \( w_2 \). These signals describe an ellipse in the Lissajous figure due to the phase error, see Fig. 2.11(a). The ideal signals \( v_1(t) \) and \( v_2(t) \) can be expressed as a function of the phase uncorrected quadrature signals \( w_1(t) \) and \( w_2(t) \); see (4.22). On the basis of the constraint (4.18) that \( v_n \) describe a circle with constant radius \( R = 0.5 \) in the lissajous figure, parameters \( A \) through \( C \) (4.20) can be fitted. The Pythagoras relation of (4.18) states that the squared radius \( R^2 \) equals the sum of the squared (waveform corrected) quadrature signals \( w_1^2 \) and \( w_2^2 \). Simplifying (4.18) and putting it in a form useful for a least squares fit, results in (4.19). The least squares fit is performed on (4.19). The parameters \( A \) through \( C \) are fitted and given in (4.20). These parameters are needed since it is not possible to directly calculate the correction parameter \( \alpha \).

\[
R^2 = (w_1)^2 + \left( \frac{w_2 + w_1 \sin(\alpha)}{\cos(\alpha)} \right)^2,
\]

(4.18)

\[
[w_1^2 \ w_2^2 \ w_1w_2] \cdot \begin{bmatrix} A & B & C \end{bmatrix}^T = 1,
\]

(4.19)
4.3. Position

\[ \begin{align*}
A &= \frac{1 + \frac{\sin^2(\alpha)}{\cos^2(\alpha)}}{R^2}, \\
B &= \frac{1}{R^2 \cos^2(\alpha)}, \\
C &= \frac{2 \sin(\alpha)}{\cos^2(\alpha)}. 
\end{align*} \]  

(4.20)

From (4.20) \( \alpha \) can be derived as in (4.21). This is the result of the offline identification and used in the online correction.

\[ \alpha = \arcsin\left(\frac{C}{\sqrt{2BR^2}}\right) \]  

(4.21)

**Online inverted model correction**

The phase error compensation is done with the inverted model (4.22) and the waveform corrected quadrature signals \( w_1 \) and \( w_2 \) as inputs. Model (4.22) is derived from (4.16).

\[ \begin{align*}
v_1 &= w_1, \\
v_2 &= \frac{1}{\cos(\alpha)} (w_1 \sin(\alpha) + w_2).
\end{align*} \]  

(4.22)

The practical implementation and results of both correction methods are discussed in the next chapter. Below is described under what assumptions the auto-calibration is applicable. The last section of this report briefly describes the position determination.

4.3 Position

Now the error reduction is done, the position can be calculated. This is done by taking the unwrapped form (2.4) of the four quadrant arctangent computation (2.3) of the quadrature signals.
Chapter 5

Implementation and results

This chapter explains how the compensation methods of the previous chapter are implemented on an experimental setup. In Section 5.2, the test results are presented.

5.1 Implementation

5.1.1 Experimental setup

The tests are performed on a low resolution rotary analog encoder (ERN1080 from Heidenhain) with 100 periods per rotation. One period being a set of one dark and one opaque line. For reference, an analog encoder of 5000 periods per rotation is used (ERN480 from Heidenhain). The output of this encoder is fed to an interpolator (GEL214 of Lenord+Bauer) which gives $640e^3$ counts per rotation. They are mounted as in Fig. 5.1. The axis runs through the reference encoder making sure that the connection between the encoders is very stiff, enforcing equal mechanical input. The reference encoder has 50 times more periods per rotation, therefore the narrow angle errors have a 50 times higher frequency. Therefore the reference encoder is assumed to be a perfect reference near the frequency range of the narrow angle errors.

Data acquisition

The calculation of the compensation parameters is done in Matlab (v 7.4.0). These parameters are uploaded in Matlab Simulink, where a c-code is generated, which is used in a fully preemptive linux kernel, where the errors are compensated in real-time with a sampling frequency of 1kHz.

5.1.2 Offline identification

The identification is done on the two quadrature signals, depicted as $Q_1$ and $Q_2$ in Fig. 4.1. First the mean offsets are identified and corrected offline. With this new data the waveform correction parameters are identified offline. Then the mean offsets, waveform and amplitudes are corrected online, see (4.10) through (4.13). Another constant speed experiment is conducted and a second data set, $w(t)$ as in Fig. 4.11 is taken. The phase error is calculated offline. As a final step these parameters are uploaded online and the complete corrections are ready.

An alternative to creating the data set $w_1$ and $w_2$ online, is to perform the waveform correction offline and use that data for the estimation of the phase error. The small disadvantage of doing it
offline is the extra needed matlab code for correcting the waveforms offline and a small room for error if it is different from the online waveform correction code.

Constant speed

For the waveform correction it is necessary that the encoder data is gathered at constant speed, not for the Heydemann correction. Due to variations in speed the energy spreads over frequency in the FFT, causing the Fourier constants in (4.4) to decrease. Since it is undesirable to use a reference in practical applications, the feedback control is done with the uncorrected outputs of the low-res encoder. The identification of the phase error is unaffected by the speed variations as is the entire Heydemann identification, see [12]. Experiments with the feedback over the reference encoder give practically the same results as with feedback over the low-res encoder, which means the speed deviations due to the imperfections in the low-res encoder are not such that they significantly influence the estimation of the Fourier coefficients.

The upper bound for the rotation speed of the identification routine should satisfy (5.1). With sample frequency $F_s$, identification speed in waveforms per second $F_{id}$ and the order of approximation of the sawtooth $m$. A factor ten is chosen as a practical Nyquist frequency [4]. This results in a bound of 4Hz for a sample frequency of 1kHz and order $m$ of 23.

$$F_{id} \leq \frac{F_s}{10 \cdot m} \quad (5.1)$$

Length of data

The offline identification should be done for at least 100 seconds, in order to get a frequency resolution of $\frac{1}{100}$ Hz, which is essential for correct estimation of the Fourier coefficients. Simulations have shown that the Fourier coefficients are estimated with an absolute and relative error of respectively $\leq 3mV$ and $\leq 0.33\%$. The simulation is done on a pure sawtooth which coefficients are
Table 5.1: Online calculation times.

<table>
<thead>
<tr>
<th></th>
<th>max</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>waveform</td>
<td>$5e^{-3}$</td>
<td>$3e^{-3}$</td>
</tr>
<tr>
<td>phase</td>
<td>$5e^{-5}$</td>
<td>$2e^{-5}$</td>
</tr>
<tr>
<td>atan</td>
<td>$3e^{-5}$</td>
<td>$1.5e^{-5}$</td>
</tr>
</tbody>
</table>

describes analytically by (4.1). Experiments have shown that a higher frequency resolution will not decrease position error. The data should also be taken over the entire travel of the encoder, which can result in an active bound for the data length for high resolution encoders.

The result of the offline identification routine is a set of correction parameters. Which are a vector with coefficients for the waveform correction $F_B$ and respectively the amplitude, offset and phase correction parameters $A_1$, $A_2$, $o_1$, $o_2$ and $\alpha$.

### 5.1.3 Online compensation

Fig. 4.1 shows the online correction. The 4-fold outputs of the encoder are paired as described in (2.2). The mean offset is corrected before the waveform correction. The output of the waveform correction blocks, $w(t)$, are two quadrature signals as described by (4.16). Next the phase error is corrected by means of a truncated Heydemann model (4.22). Then the position is derived from the corrected signals, according to (2.3) in Section 2.3.1.

### Timing

The maximum and average online computation times of the blocks of Fig. 4.1 are given in Table 5.1. The maximum total computation time per sample is $< 18e^{-5}$s for the implementation as in Fig. 4.1. When the waveform correction is done on the 4-fold signals, the maximum computation time becomes $< 28e^{-5}$s. The calculations are done in the matlab simulink environment on a PC with a dual core 2.8GHz processor.

### Maximum speed

The maximum speed is determined by the unwrap function used in matlab to determine the position. This function needs a minimum of two samples per period of the input signal. So the maximum speed $f_{max}$ is a function of sampling frequency according to (5.2) and a upper bound on the frequency of the quadrature signals. This speed bound is coincidentally identical to the Nyquist frequency. For this setup this results in a maximum speed of 300 revolutions per minute. There are no other upper bounds or lower bounds for the speed. There are no bounds on the acceleration in order for the correction to work properly. The performance is independent of the speed or acceleration within this speed bound.

$$f_{max} \leq \frac{F_s}{2} \quad (5.2)$$
5.2 Results

This section shows the results of the online error compensation applied to the experimental setup. The quality of the correction is best expressed in position accuracy. The Lissajous figure remains a nice graphical representation of the improvement.

5.2.1 Waveform correction

Fig. 5.2 gives the Lissajous figure of the raw $Q(t)$, and corrected, $v(t)$, quadrature signals. The desired output is a circle with constant radius and centered on coordinates $(0, 0)$. Since the ideal encoder outputs have an amplitude of $0.5V$, so do the corrected signals. Graphically the result are very nice.

The quality of the waveform correction can be described as the absence of higher spatial harmonics in the quadrature signal. Fig. 5.3 gives the result of the FFT of an online corrected and uncorrected signal of 10Hz. The reduction of the higher spatial harmonics is a measure of the quality of the waveform correction.

5.2.2 Remaining errors in the quadrature signals

The raw and corrected signals are also given in the time domain in respectively Fig. 5.4(a) and 5.4(b). These figures do not give a clear indication of the errors, therefore the error values are given in Table 5.2. The worst case errors are given by the manufacturing specifications of [11]. The uncorrected errors are depicted in the second column. An estimation of the remainder of
the errors after correction are given in the last section. They are estimated by comparing the FFT’s of the corrected position errors with a simulated position error. A more detailed description of the error reduction and performance is given in the next section.

### 5.2.3 Position

The overall performance is best quantified by means of the error reduction in position. Fig. 5.5 shows the error in the time domain of the corrected and uncorrected signals with a frequency of 10Hz. The reduction in position error is clearly visible. The correction factor $C_f$ of (5.3) gives the fraction of the rms values of the error of the uncorrected and corrected position errors, respectively $\text{rms}_{\text{raw}}$ and $\text{rms}_{\text{cor}}$. The factor of improvement is approximately 15. Percentage wise, is 93% of the position error removed. Fig. 5.2.3 shows the FFT of the error. The error peak at 10Hz is due to the mean offset. The 20Hz peak is mainly due to the amplitude and phase offset. The 40,80 and 120Hz peaks are mainly due to the waveform distortion.

$$C_f = \frac{\text{rms}_{\text{raw}}}{\text{rms}_{\text{cor}}}$$  

(5.3)
Figure 5.4: The raw and corrected quadrature signals.

Figure 5.5: Position error.
5.2. Results

The error is reduced and not canceled, the remainder is visible in Fig. 5.5 and 5.2.3. In simulations comparable with the experiments the errors are corrected below noise levels of the experimental setup, see App. C. The presence of noise in the experiments only has a small contribution to the remaining 7% of the position error. The main difference in experiments and simulations are the variations in feedback control during the identification phase, the wide angle errors, see App. B, and the non equidistant lines on the reticle and mask. Since experiments with feedback over the reference encoder have given comparable results, the variations in the line widths and the effects of the wide angle errors on the narrow angle errors must be the cause. The variations in line widths are expected to cause small variations in the narrow angle errors. The conclusion is that the remaining 7% of the position error in rms value, is mostly due to these effects.
Chapter 6

Conclusion and recommendations

6.1 Conclusions

Both linear and rotary encoders inhibit errors. A group of high frequent errors, referred to as narrow angle errors, are identical for linear and rotary encoders. The quadrature signals contain four narrow angle errors namely, mean offset, amplitude distortion, waveform distortion and phase offset. A novel compensation method is developed for amplitude and waveform distortion. The phase offset is corrected with a truncated Heydemann model. The compensation method or auto calibration method consist of two parts. The first is an offline error identification routine, based on a data set of the quadrature signals generated by the encoder. The second is an online compensation, which reduces the errors. The main advantage of this method is that there is no need for a reference to calibrate the encoder, it is self calibrating. This makes this method very easy to implement. The position is determined with the corrected quadrature signals by the so called arctangent method.

The effectiveness of this method is shown by the improvement in position accuracy. The rms value of the error is reduced to 7% of its original value, a factor 15 improvement. While the offline identification routine may take up to 15 minutes due to size of the data set, the online computational burden is reduced to $18e^{-5}$s. If no other online calculations are done, this means the quadrature signals can be sampled and corrected up to $\approx 5$kHz. The maximum operating speed is expressed in waveforms per second and is equal to half the sample frequency.

6.2 Recommendations

Remaining error

The remaining 7% of the error is mainly due to the influences of the wide angle errors and the non-equidistant code slits, which makes the narrow angles vary slightly. To get a better understanding of the effects of non-equidistant code slits, an offline wavelet decomposition can be made on the quadrature signals and position (error) over the entire travel of the encoder. The deviations in the line widths are unique for every line, this makes the error in position deterministic over the entire travel of the actuator. So the error has a period equal to the mechanical rotation, in case of a rotary encoder, but within this period it contains some high frequent content. Now lets assume a wavelet decomposition with infinitely high frequency resolution and therefore an
infinite number of outputs. The wavelet decomposition might give a better view on the relations between the variations in line widths and the narrow angle errors than the FFTs.

**Waveform correction**

The waveform correction is used to correct a triangular shaped signal into a sine. On the basis that the input signal can be expressed as a sum of sines. The Bromwich formula is available for odd and even higher spatial sines and cosines, see [5]. Therefore the idea of the waveform correction can be used to correct other type of periodic signals. The current method then needs to be extended in a comparable fashion.

**Noise reduction**

This section explains why it is interesting to do a research of the possibility to extend the waveform compensation method to a disturbance reduction method. Its feasibility is not proven. Disturbance reduction might be achieved for applications where the measurand typically contains some higher periodic disturbances. The disturbance frequencies should be a fixed multiple of the measurand, just like the narrow angle errors with respect to the quadrature signals. A good testcase can be to cancel for periodic deviations in laser interferometers as in [7]. Another area of interest could be denoising of audio signals. A condition is that the measurand and the disturbance can be modeled as in (4.2). Another issue is that the method should be able to cope with a sum of phase shifted sines. While a sawtooth is sum of in-phase sines, this can not be expected from any signal. For the offline identification of the phases a wavelet decomposition might be useful, since it operates in the time-frequency domain as apposed to the Fourier decomposition which is a translation from one domain to the other. A skewed sawtooth $N$ is a sum of phase shifted sines and therefore a good starting point of how to model such a signal, [4].

\[
N = \frac{2A}{\alpha(\pi - \alpha)} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin(n\alpha) \sin(n\theta) \quad \text{for } n = 1, 2, 3...\infty
\]  

Figure 6.1: Skew sawtooth
The main difference in this equation compared to (4.1) is the extra sine term $\sin(n\alpha)$, with phase shift $\alpha$. This extra sine term can also be described by a $B$ and $S$ matrix as in (4.6) and (4.7). This calls for a multivariable online approximation routine, because of the extra $\alpha$ term. The Newton-Raphson method is also available in a multivariable form, see [9].
References

[1] Internet reference to the TUEDACS: www.tuedacs.nl. TUEDACS is a data Acquisition & Control System developed by the Technische Universiteit Eindhoven.


Appendix A

Reference encoder

This appendix describes the connection layout of the reference encoder to the interpolator used in the experimental setup.

A.1 Pinning layout interpolator

The output of the reference encoder is a combination of analog signals. These analog signals are mounted on 15pins connector. This connector is the input of the interpolator. The output of the interpolator is fed into the TUEDACS, a data-acquisition device. The TUEDACS has a 9pins digital encoder input. In order to connect the interpolator to the TUEDACS a "special" cable was made. Table A.1 shows the pin layout of the interpolator output to the TUEDACS input.

<table>
<thead>
<tr>
<th>Color</th>
<th>signal</th>
<th>15pins nr</th>
<th>9pins nr</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>null</td>
<td>1</td>
<td>1,2,3,4,5</td>
</tr>
<tr>
<td>Brown</td>
<td>+5V</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Green</td>
<td>index</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Yellow</td>
<td>A</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Gray</td>
<td>B</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>
Appendix B

Wide angle errors

B.1 Eccentricity errors

Several imperfections in manufacturing and mounting of the encoder result in one eccentricity error. In Fig. B.1 the eccentricity is presented as the vector from $S$ to $C$ which is the deviation from the center of the scale disk $S$ to the center of the actual center of rotation $C$. The center of the photo diodes are represented by $P$. For constant angular motor speed $\phi$, a circle can be drawn on the scale disk, where the speed is constant at point $P$. When the encoder lines $L$ pass the the photo diodes (and scanning reticle) they do this at a varying radius $R(\phi)$ on the disk. The radius varies from $R = |C\overrightarrow{P} - C\overrightarrow{S}|$ to $R = |C\overrightarrow{P} + C\overrightarrow{S}|$, where $C\overrightarrow{P}$ and $C\overrightarrow{S}$ are the vectors. The lines $L$ on the encoder are tapered. The widths of the lines on the disk passing the photo diode varies due to variations in $R$. The radius $R$ varies with the same period as the mechanical rotation $\omega = \frac{d\phi}{dt}$, as in (B.1). Therefore the position error due to eccentricity errors has the same frequency as the mechanical rotation frequency $\omega$.

$$R(\phi) = \sqrt{|C\overrightarrow{P}|^2 + |C\overrightarrow{S}|^2 - 2|C\overrightarrow{P}||C\overrightarrow{S}|\cos(\phi)} \quad (B.1)$$

B.2 Ellipticity of the code disk

An error which is comparable to the eccentricity error is the error due to ellipticity of the code disk. Fig. B.4 gives an exaggerated schematic view of an elliptical code disk. Photo diode $P$ is fixed on constant radius of the center of rotation. Comparable to the eccentricity error, the width of the lines at the photo diode vary during rotation. The elliptical code disk has maximum radius at two points, in Fig. B.4 these are pointing in the positive and negative vertical direction. A minimal radius is found in the positive horizontal direction and one in the negative horizontal direction.

The error in position now has a frequency of two times the mechanical rotation, since there are two maxima and two minima per rotation.

Effects in measured position

This section is about the effects of the wide angle errors in position. Both the reference encoder and the low resolution encoder embed eccentricity errors. To calculate the error in position the
Figure B.1: Eccentricity error

Figure B.2: Ellipticity error
B.3. Effects on narrow angle errors

The uncorrected position output of the low resolution encoder is subtracted from an illusive constant speed signal. The FFT is given in B.3(a). The same is done for the error, or difference, between the uncorrected position output and the output of the high resolution encoder. The FFT is given in B.3(b). The low frequent error is much lower for the latter, probably due to mounting error between motor and axle which is the same for both encoders. There are also some disturbances present in the setup, the most dominant are in the frequency interval of one and ten Hz. These are visible in B.3(a) and not in B.3(b).

B.3 Effects on narrow angle errors

The wide angle errors have an influence on the amplitude of the waveforms and the shape of the waveforms. Both are discussed.

B.4 Effect on amplitude

When the widths of the lines on the disk passing the reticle vary, the light passing through the reticle and disk on to the photo cell changes. See Fig. 2.3 for a schematic overview of the travel of the light. This changes the amplitudes of the 4-fold signals, hence the quadrature signals change in amplitude. Fig. B.4 gives a zoomed view of a quadrature signals, $Q_1$. The changes in amplitude are comparable for both quadrature signals. In ten seconds the encoder has completed one revolution. Besides the amplitude variations due to noise, the amplitudes vary with a period of ten seconds.

The conclusion is that the wide angle errors influence the amplitude error. The influences are small and most likely easy to compensate. This compensation is not done in this thesis.

B.4.1 Effect on waveform distortion

The variations of line widths has an influence on the Moire effect. See [23] or [2] for more information on Moire models and how variations in line widths influence the Moire effect. These variations alter the shape of the quadrature signals, which means the Fourier coefficients in (4.4) change with the same periods of the wide angle errors.

The conclusion is that the wide angle errors influence the waveform error. The influences are small. It depends on the variations in $F$, if they can be compensated. This compensation is not done in this thesis.
Appendix B. Wide angle errors

Figure B.3: FFT of position error.
Figure B.4: Amplitude deviation in $Q_1$ due to ellipticity error.
Appendix C

Simulations

C.1 Simulations with sawtooth input

This section gives the simulation results of the auto calibrations method. As input two sawtooth signals are used, with errors as in table 5.2. The sawtooth signals are generated with equation (4.1). The Lissajous figure C.1 shows the shapes of the input and corrected signals. Fig. C.1 gives the Fourier transformation of a raw and corrected quadrature signal, remark that the higher spatial harmonics are practically absent for the corrected quadrature. Fig. C.1 gives the position error and Fig. C.1 its Fourier transformation. The correction factor or factor of improvement is given by $C_{\text{sim}}$ in (C.2) and is defined as the rms value of the error of the uncorrected position divided by the rms value of the corrected error and is $\approx 1e3$. The rms value is defined as the sum of the squared position errors over 50s.

$$C_{\text{sim}} = \frac{\text{rms}_{\text{raw}}}{\text{rms}_{\text{cor}}}$$

$$= \frac{1.31e^{-2}}{1.23e^{-5}}$$

C.2 Different encoder resolution, different waveform

A low resolution encoder will have a more sawtooth like output while a high resolution will have a more sinusoidal output. The waveform correction works on the assumption the encoder output is a sum of odd higher spatial sines. This method can be extended to even higher spatial sines and all higher spatial cosines. This is however not the intention here. The goal of this section is to reproduce a series of quadrature signals with different waveforms, starting from a sawtooth and finishing with a sine. The signals are presented in the Lissajous figure C.2. In this thesis the assumption is made that all encoders outputs are shaped like the quadrature signals of Fig C.2 or something in between.

The signals are made by lowering the higher spatial harmonics in various ways, just some of them are shown. All of the quadrature signals are corrected, all results are comparable to those of the previous section and therefore not shown.
Appendix C. Simulations

Figure C.1: Lissajous figure

Figure C.2: Fft of quadrature signals
C.2. Different encoder resolution, different waveform

Figure C.3: Position error

Figure C.4: Fft of position error
Figure C.5: Lissajous figure of different simulated quadrature signals