ULTRA-PRECISE ON-AXIS ENCODER SELF-CALIBRATION FOR FAST ROTARY PLATFORMS

by

Darya Amin-Shahidi

B.A.Sc., University of British Columbia, 2006

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

The Faculty of Graduate Studies

(Mechanical Engineering)

The University of British Columbia
(Vancouver)

April 2009

Abstract

This thesis presents the design, implementation, and experimental testing of a rotary platform and its use for encoder calibration research. So far, the experimental results have demonstrated a calibration repeatability of better than 200 nano-radians (0.04 arc-second), which is much better than the state of the art precision angular encoder error of about 1 arc-second. The major contributions of this thesis include achieving ultra-accurate encoder calibration results as well as the design of an ultra-precise rotary table and a set of high-speed electronics.

Others have developed many encoder calibration techniques. This thesis aims to identify, improve, and experimentally demonstrate the accuracy limit of an encoder self-calibration method developed at UBC. The work can be categorized into the following three parts:

1) Ultra Precise Rotary Platform Design, Manufacturing, and Assembly:
A precision rotary platform was designed for encoder calibration and other spindle metrology research. The rotary platform uses a precision airbearing. Two different rotary encoders are included for rotation measurement. Three displacement probes monitor the motion of a target ball to identify and compensate for rotor vibrations. The platform is driven by an ultra-low cogging torque motor for minimum velocity ripple. The setup's versatility makes it an excellent spindle metrology research tool.

2) High-Speed Electronics for Signal Processing:
Fast processing is required for encoder calibration and precise stage control. Since no commercial controller met our fast signal detection and processing requirements, a set of electronics was designed: an FPGA-based high-speed digital controller (Avalanche) and a precision analog processing board (NanoRAD). Avalanche incorporates a fast-Virtex4 FPGA chip, a novel 1.2-GHz timer, five fast ADCs, four fast DACs, Gigabit Ethernet, High-Speed USB, and DDR RAM. NanoRAD incorporates fast comparators and precision buffers for analog processing and filtering.

3) Self-Calibration Algorithm Development and Experiments:
After integration of the electronics with the mechanical assembly, the encoder calibration technique was examined. The method was adapted for the new setup, the damping estimation portion of the algorithm was improved, and the whole method was tested experimentally and in simulation. The encoder accuracy has been experimentally enhanced by 25-times from 5,000 (1 arc-second) to 200 (0.04 arc-second) nano-radians.

Table of Contents

List of Tables

List of Figures

Acknowledgements

First, I would like to thank my thesis supervisor, Professor Xiaodong Lu, for introducing me to the area of precision engineering and creating the opportunity for me to work on very exciting research. He helped me learn many interesting topics, such as circuit design, control, and mechanical design. More than anything else, I am grateful to him for being such a great inspiration and helping me develop and believe in my abilities. He trusted me with tasks which seemed next to impossible to me at first; however, he was patient and followed and directed my progress. I find his analytical skills, work standard, and personality inspirational. It has been a privilege to work with Professor Lu. It is an honor for me to be his first student.

I am thankful to Professor Yusuf Altintas and Professor Juri Jastkevich for being in my thesis committee. They have always been very supportive and I have learned so much from their courses in the UBC mechatronics program.

KLA Tencor financially supported this research project. Collaboration with KLA Tencor was an excellent professional experience for me. I would like to thank Marek Zywno, Upendra Ummethala, Joshua Clyne, and Tao Zhang. Their feedback and critical questions were very helpful to deriving my research in the right direction.

I have been very lucky to have lab mates, who are great friends and colleagues. I have enjoyed working with them and have had great times together inside and outside the lab. Their help has been very important to my research. Kris Smeds designed the software for interfacing to my electronics. Richard Graetz and I worked very closely on the encoder calibration testing and development as well as the daughterboard design. Arash Jamalian helped me with the mechanical drawings and the daughterboard design. Thank you for all your help.

I am indebted to my parents; with their support, they have made many of the opportunities in my life possible. I want to thank my sister, Yalda, for helping me with her photography skills. I would like to thank my Mashad FC team mates, particularly Arash, for the championship and the exciting times we had together. I am grateful to Sophiya for making my life so much more beautiful. I feel very confident having a dependable friend like her.

For my grandmas,

CHAPTER 1

Introduction

Rotary encoders are widely used for rotation angle sensing. High precision encoders are desired for better rotation angle and velocity measurement. Among other areas, semiconductor industry, precision manufacturing, and astronomy are some sectors that can benefit from high-precision encoders. Encoder calibration is commonly used to enhance the encoder accuracy. In this thesis, a precision rotary table and enabling high-speed electronics are implemented and are used to improve and experimentally demonstrate better encoder calibration results.

The encoder calibration method originally developed in [1] is used for this research. So far, an encoder calibration repeatability of 150 nano-radians[1] has been experimentally demonstrated which is significantly better than the previous results achieved using this method. Our simulations and analysis shows that this method can potentially reach an accuracy level of 10 nano-radians.

Next, we discuss in details an example application of precision angular encoders in the semiconductor industry.

[1] This thesis uses nano-radians as angular measurement unit (1nano-radian = 0.0002 arc-second).

1.1 High Accuracy Angular Encoder Applications in Semiconductor Industry

The semiconductor industry manufactures electronics chips which are the basic blocks of all the electronic devices. The Semiconductor Industry Association (SIA) estimates an annual worth of about $257-billion in 2008 and predicts that 234,000 people are employed by this industry in United States alone in 2005 [2]. This industry is very important to economic growth as it supplies the core resource of many important sectors such as telecommunications, internet, and media.

The semiconductor industry's fast rate of advancement differentiates it from other industries. Smaller semiconductor devices with more performance and less cost are being offered every year. In 1965, Gordon E. Moore predicted that the number of transistors in an electronic device was doubling every two years [3]. The industry has kept up with this prediction since then. This industry's fast advancing nature has created a very competitive atmosphere and a great need for new and innovative technologies to push the manufacturing limits and to enable creating smaller and faster devices.

The International Technology Roadmap for Semiconductors (ITRS) is an association which helps the advancement of the semiconductor industry by directing the efforts of the industry, national, and university research labs working in this field. ITRS publishes a roadmap for advancement of the semiconductor industry and outlines the challenges and difficulties to be addressed. The roadmap is created based on the opinion of more than 800 experts working in the field around the world. To create a unified practical advancement criterion, ITRS defines the roadmap based on the distance between

the metal lines connecting to the DRAM bit cells. The distance is termed as half-pitch and refers to the average of the metal line width and the distance between two adjacent lines. The roadmap for the DRAM devices generated by ITRS is shown in Table 1-1. The expected simultaneous reduction in size and cost, shown in the ITRS estimate, provides a strong incentive for advancement of the semiconductor industry.

Table 1-1: ITRS roadmap for expected reduction in DRAM ½ pitch size and cost/bit [3]

	Year								
	2007	2008	2009	2010	2011	2012	2013	2014	2015
DRAM 1/2 Pitch (nm)	65	57	50	45	40	36	32	28	25
Cost/Bit @ Production (micro-cents)	0.96	0.68	0.48	0.34	0.24	0.17	0.12	0.08	0.06

Until now, the industry has used 193-nm optical lithography for mass manufacturing of the semiconductor devices; however, this method has reached its optical diffraction limit. In the optical lithography process, a light source with 193-nm wavelength is used to replicate the patterns on a master mask onto a wafer coated with photo-sensitive materials. The wavelength of the light source limits the minimum feature size which can be clearly transferred and focused from the mask onto the wafer. ITRS finds the optical lithography process to be limited to 65-nm DRAM half pitch.

To continue the trend of decreasing half pitch other lithography methods such as Immersion, Double Patterning, Extreme Ultra Violet (EUVL), Mask-less (ML), and Imprint lithography methods are being researched and developed [4]. Immersion lithography operates in the water medium where due to high refraction index of water the

wavelengths become smaller. Double patterning uses two masks, each with twice the critical dimension, creating a single image. Extreme ultra violet light with only 14-nm wavelength is used in EUVL to perform lithography. Mask-less methods use an electron beam to etch features into the wafer. The imprint lithography method replicates a master pattern using means similar to mechanical molding.

ITRS outlines the potential manufacturing processes for different generations of DRAM half pitch. ITRS's predictions are summarized in the chart in Figure 1-1, where potential solutions at each DRAM half pitch stage are listed in the same order as their probability of being the dominant solution. ITRS finds 193-nm optical lithography to be limited to 65-nm DRAM half pitch. ITRS predicts that water immersion lithography can be used to extent optical lithography to 32-nm and double patterning has the potential to extent immersion optical lithography to below 32-nm DRAM half pitch. ITRS terms EUVL, ML, and Imprint as post-optical alternatives and expects them to be the dominant manufacturing solution at DRAM half pitch of 32-nm and below. Currently, ML methods have a limited throughput, and hence, are used for prototyping and low volume manufacturing only. ITRS finds EUVL to be the most probable post-optical alternative; however, it mentions that breakthroughs with ML processes can be very significant and can create paradigm shifts. If adapted for mass manufacturing, ML methods will eliminate the need for master-masks, and hence, will reduce cost and cycle times.

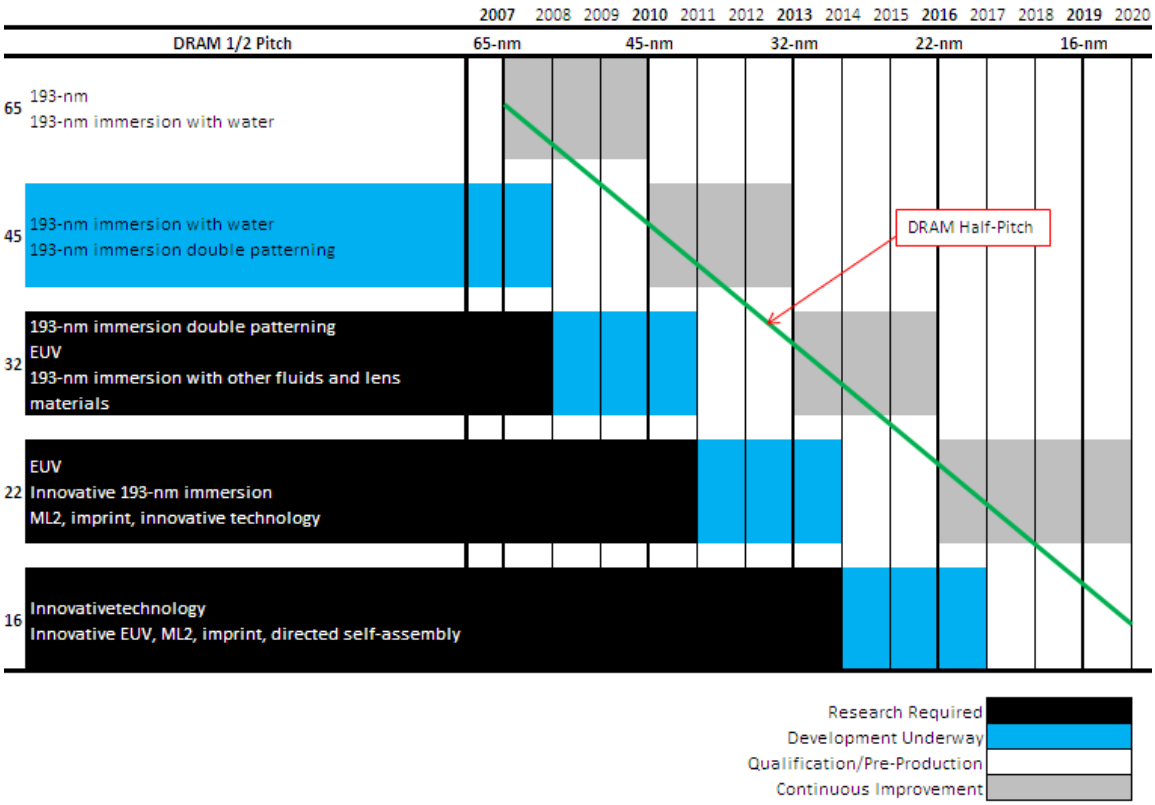

Figure 1-1: Potential semiconductor manufacturing solutions at each stage of ITRS roadmap [4]

KLA Tencor, a leader in semiconductor manufacturing and inspection equipment development, is developing a mask-less electron-beam lithography technique which can significantly increase the throughput of the ML methods. Such a system will enable cheaper and more efficient ML production and will bring the semiconductor industry a step closer to mass manufacturing using ML techniques. A key component to increasing the throughput of their e-beam lithography machine is an angular encoder with accuracy of 10 nano-radians at speeds up to 600-rpm. Currently, there are no rotation measurement solutions which have approached such high accuracy requirements at such high speeds.

The state of the art encoder technology is only accurate to 2000 nano-radians at speeds only as high as 60-rpm. The available encoder calibration methods have enabled

enhancing an encoder's effective accuracy to levels as low as 200 nano-radians at low speeds. Dr. Xiaodong Lu, at the University of British Columbia, has invented a novel encoder calibration method which can potentially meet the accuracy requirements of KLA Tencor's next generation e-beam machines. The research in this thesis is aimed to improve and experimentally demonstrated the accuracy limit of the calibration technique. An overview of the previous encoder calibration work is provided in the next section.

1.2 Prior Art Rotary Metrology Calibration Methods

Angle encoders are widely used for large rotation angle sensing. The state-of-the-art rotary encoder technology can only reach accuracy levels up to 2,000 nano-radians due to error in the scale's grating lines spacing, mounting misalignments, and other imperfections[5]. Attempts have been made to calibrate rotary encoders and improve their accuracy. An ideal calibration method would identify and compensate the repeatable portion of the encoder error. So far, the best stable calibration repeatability is 50 nano-radians which has been achieved at *National Institute of Advanced Industrial and Science Technology (AIST)* [6]and *Physikalisch-Technische Bundesanstalt (PTB)* [7]. All developed encoder calibration methods can be divided into four categories: comparison-calibration-methods, cross-calibration-methods, direct-calibration-methods, and temporal self-calibration methods. Some of the major prior art calibration methods are discussed under these categories in the next four sub-sections.

1.2.1 Comparison Calibration Methods

The calibration methods which rely on a more accurate measurement reference to calibrate a less accurate sensor are termed as comparison-methods. Precision polygons along with photo-electric autocollimator can be used as a calibration reference. As an alternative, Renishaw has integrated a precision indexed rotary table along with a laser interferometer to create an angle reference accurate to 4000 nano-radians [8]. Ring lasers have also been used as an accurate angle reference in [9-11], where accuracies as low as 500 nano-radians have been demonstrated. Finally, other calibration methods have been used to develop accurate angle references which can now be used for comparison

calibration of rotary sensors. The work done by AIST in [6] and PTB in [7] are the most accurate examples of such work where 50 nano-radians reference angle repeatability has been achieved.

The comparison-calibration methods are limited to their reference angle's accuracy. Most angle references only cover a small portion of a single rotation, so the calibration needs to be performed in multi steps. As a result, the process will be time-consuming and less accurate. Furthermore, the comparison-calibration methods require connecting the reference sensor and the sensor being calibrated on a single axis. Large uncertainties are introduced due to misalignments associated with this connection. Finally, recalibration of the encoder in the final product is not possible using this method, or in other words, this is an off-axis calibration method. As a result, the comparison calibration methods cannot capture the variations due to changes in mounting and operating conditions, such as speed and temperature.

1.2.2 Cross-Calibration Methods

In the cross-calibration methods, a second artefact using an auxiliary non-calibrated rotation sensor is introduced. Opposed to the comparison-calibration methods, the auxiliary sensor does not require a higher accuracy as this method calibrates both sensors together. In this method, the difference in measurements of the primary sensor and the auxiliary sensor is termed as relative error and can be measured. The primary sensor is fixed at different relative angular positions to the secondary sensor, and at each relative position, the relative error is measured along a full rotation. The measured relative error combined with circular closure equations of each sensor can provide the error map for both of the sensors. [7, 12] are examples of this technique. PTB has attempted cross-

calibration using their setup in [7] and has demonstrated 50 nano-radians long-term and 10 nano-radians short-term repeatability.

The cross-calibration method can produce higher accuracy. However, it is very time consuming and difficult as the relative error needs to be measured for different relative angular positions of the primary and the secondary sensors. This method can be applied at very low rotation speed so it cannot capture the variations in the error resulting from the changes in speed. Also, it is not practical to recalibrate an encoder during a machine's operation using this method, and as a result, the variations due to changes in temperature and other operating conditions cannot be identified. In addition, for a limited number of relative positions of the sensors, some harmonics of the error will not be calibrated. As a result, the effectiveness of the method depends on how many relative positions are used. Furthermore, the estimate of the higher harmonic error components is very sensitive to the relative angle between the two sensors. Finally, this method would require an additional rotation sensor plus extra features for precise relative positioning of the two sensors; therefore, it is impractical to implement this method on every setup. However, the results can be used as a reference for comparison-calibration.

1.2.3 Direct Calibration Methods

Direct calibration methods can be seen as a modification of the cross-calibration method where no manual adjustment of the setup is required during the calibration process. In such systems, multiple read heads are utilized to provide the secondary artefact at different relative angular positions. As a result, there is no need for manual movement of the auxiliary sensor relative to the primary sensor, and hence, the sensor can be automatically calibrated. Such methods are termed as direct methods because, opposed to

cross-calibration, the calibration is performed directly on a single system without the need for manual adjustments to the system. [6, 7, 13] have applied the direct-calibration method. AIST [6] has demonstrated 0.01 arc-second uncertainty, aside from the shaft eccentricity, using this method. PTB has performed both cross-calibration and direct-calibration using their setup in [7], but has obtained more accurate results using the cross-calibration method.

This type of calibration method can provide reasonably good calibration results and is much easier to apply compared to the cross-calibration methods. However, the method still requires multiple read-heads that are accurately spaced. The effectiveness of the method depends on the number of the heads and the accuracy of their spacing. For a limited number of read-heads, certain components of the error will remain uncalibrated. The accuracy of the spacing of the heads will affect the precision of the high frequency components of the generated error-map. Finally, because the read head is varying for each artefact, the calibration method fails to capture the error specific to the read-heads.

1.2.4 Temporal Calibration Methods

Temporal calibration methods rely on time measurements of the signals to identify the encoder error map. Such methods assume a specific relation between the angular position and time, which allows them to convert time measurement to an angular position reference for encoder calibration. Some examples of temporal calibration methods are [1, 14, 15]. The main limitation of this method is that it requires good and consistent system dynamics; however, this is the case for most precision setups. If implemented effectively, temporal calibration methods will have numerous advantages: they will require no additional components for referencing to, such as a second encoder or multiple read-

heads; they can perform automatically; they can calibrate the encoder on the application axis, and hence, will eliminate all error including mounting and read-head error; and, they can calibrate all the components of the error down to the single count level.

In [14], a hyperbolic function is fitted, using the least square method, to the trail of the encoder readings which are captured as a function of time. The identified encoder is compensated using a delay circuitry and as result is only effective for constant speeds. In this work, the encoder error is not separated before applying the fit. Finally, the assumed hyperbolic relation may not be the most suitable expression of the system dynamics.

In [15], the angular velocity of the rotor, which has a fly-wheel attached to it, is assumed to be constant. Then, time measurements are directly transferred to a position reference using a constant gain. The platform velocity was estimated as the average velocity during one revolution. Due to the unrealistic constant velocity assumption, the results of this method have limited accuracy.

In [1], a dynamic model with first order damping is assumed. The dynamic model is used to create a set of spatially discrete equations for one revolution which can be solved to obtain an inertial reference frame for encoder calibration. The method finds the platform's initial velocity by combining circular closure equation and the set of dynamic equations. The dynamic model parameters, which include the damping parameters and the inertia, are estimated using the least square method. To prevent encoder error from biasing the least square fit, a novel dynamic reversal method is used to eliminate the error from the least square estimates by looking at the difference of two revolutions rather than

a single revolution. Lu in [1], has experimentally demonstrated a repeatability of 5000 nano-radians which is 30 times smaller than the original encoder error. The accuracy obtained in [1] was limited due to the precision of the encoder, the mechanical setup, and the processing electronics.

Our research identifies the temporal calibration methods to be the most practical method with a potential accuracy beyond any other method. The method developed in [1] is found to be the most complete among temporal calibration methods. As a result, the work in thesis is based on the temporal calibration method developed by Lu in [1]. More details on the workings of this method and the modifications made to it are provided in chapter-4.

1.3 Thesis Overview

The thesis goal is to enhance the state of the art rotation metrology technology to meet electron beam lithography requirement of 10 nano-radians RMS accuracy at 180-RPM. Since the state of the art encoder technology is only accurate to 2000 nano-radians, encoder calibration is required to enhance accuracy. Although two of the available encoder calibration techniques [6, 7] have been able to achieve an angle reference accuracy of 50 nano-radians, their methods work only at speeds below 10-RPM. As a result, they fail to capture the variations due to the changes in speed, and hence, their calibration results in operation are only valid to 200 radians or less. The temporal self-calibration method developed by Lu [1] can work at high-speeds and can capture thermal effects by recalibration during the operation of the device.

We attempt to improve and experimentally demonstrate the accuracy limit of the on-axis encoder self-calibration method developed by Lu[1] To enable achieving the best calibration results, an ultra-precise rotary table and a set of high-speed electronics have been developed in this thesis. The developed setup is then used to carry out encoder calibration research. This thesis is organized into three chapters: ultra precise rotary table design, high-speed electronic design, and encoder calibration algorithm development and experiment. The following sub-sections provide a brief overview of each sub-section.

1.3.1 Precise Rotary Table Design

An ultra precise rotary table, shown in Figure 1-2, is designed, manufactured, and assembled as a part of this thesis. For effective encoder calibration, the platform requires to have a repeatable motion and linear dynamics. An airbearing is utilized within the

design for its linear dynamics, high stiffness, and repeatable motion. We have designed the rotary table to include two types of rotary encoders, a 3-D displacement measurement system consisting of a precision ball and three capacitive probes, and a low cogging torque motor. These features are incorporated into the design to enable spindle metrology and control research. More detailed description of the design is provided in Chapter 2, *Mechanical Design*.

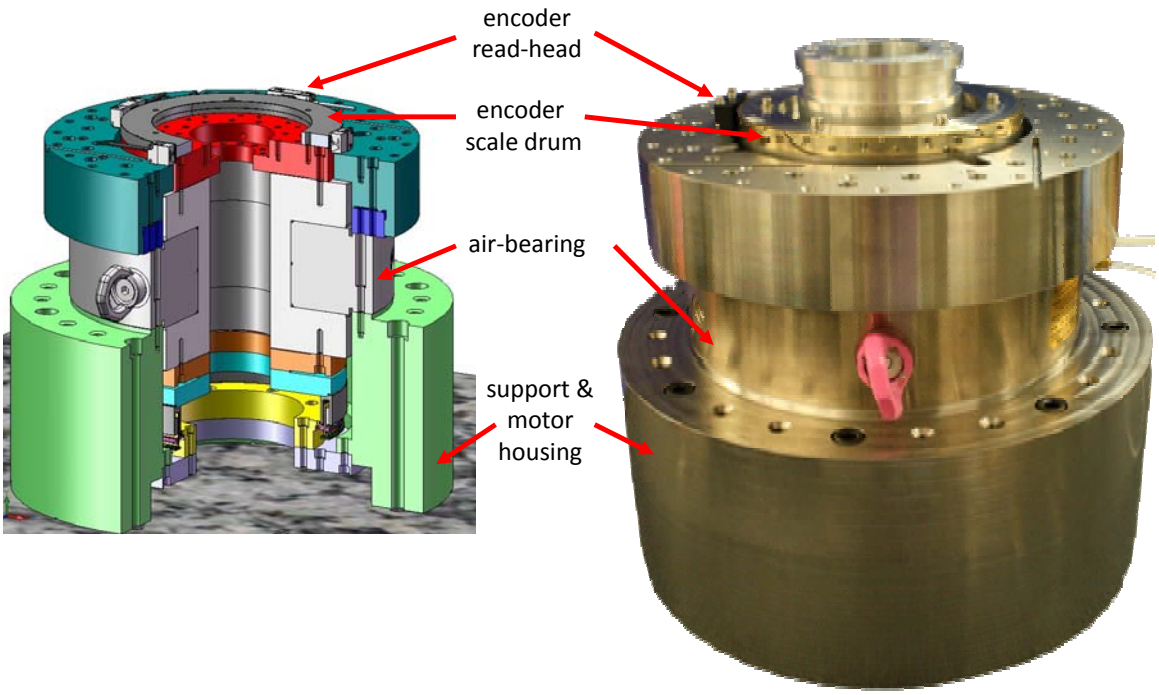

Figure 1-2: Ultra precise rotary table designed for spindle metrology and control research

1.3.2 High-Speed Electronic Design

To achieve high metrology and control accuracy, fast real-time signal processing is required. Since no commercial controller met our requirements, I designed and developed our own set of high-speed electronics. A fast FPGA-based motherboard, *Avalanche*, with 300-MHz clocking frequency was built for fast digital processing. A daughterboard, *NanoRAD*, was built to perform application specific analog signal processing, buffering, and routing. Figure 1-3 shows the motherboard and the daughterboard. More details on the electronics design is provided in Chapter 3, *High-Speed Electronics*.

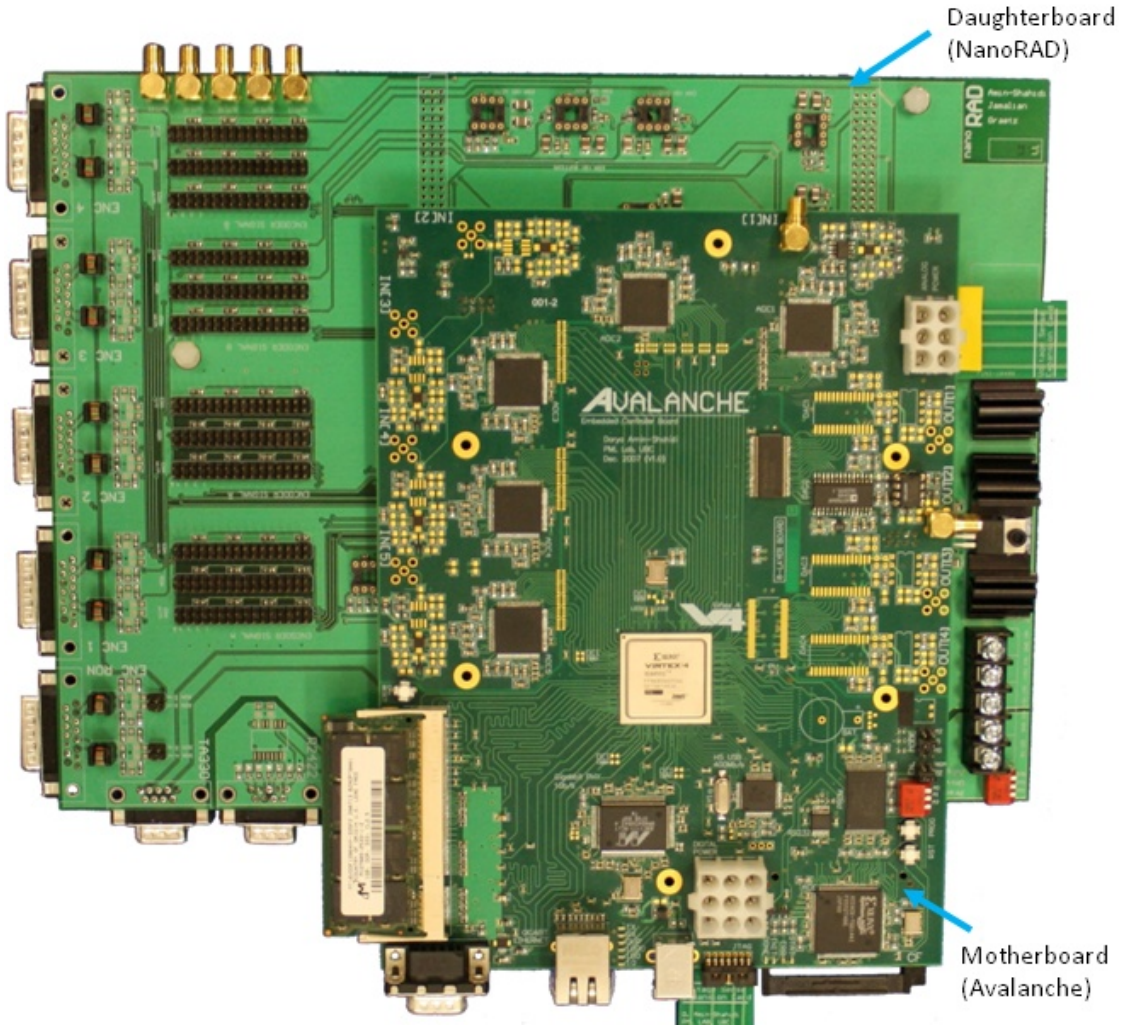

Figure 1-3: Ultra-fast motherboard (Avalanche) and daughterboard (NanoRAD) stack-up

1.3.3 Encoder Calibration

To achieve angular measurement accuracy much better than the angular encoder's accuracy, the encoder included in the setup must be calibrated. The encoder self-calibration method developed by Lu [1] is used as the basis in this thesis. We try to improve this method to achieve better rotation measurement accuracy. The experimental setup consisting of the mechanical and the electrical design is used to carry out encoder calibration research. The research aims to improve and experimentally demonstrate the accuracy limit of the method in [1]. One encoder error-map obtained experimentally for the ERA4282C encoder utilized in our rotary table is shown in Figure 1-4. More details on the encoder calibration work are provided in Chapter 4, *Encoder Calibration*.

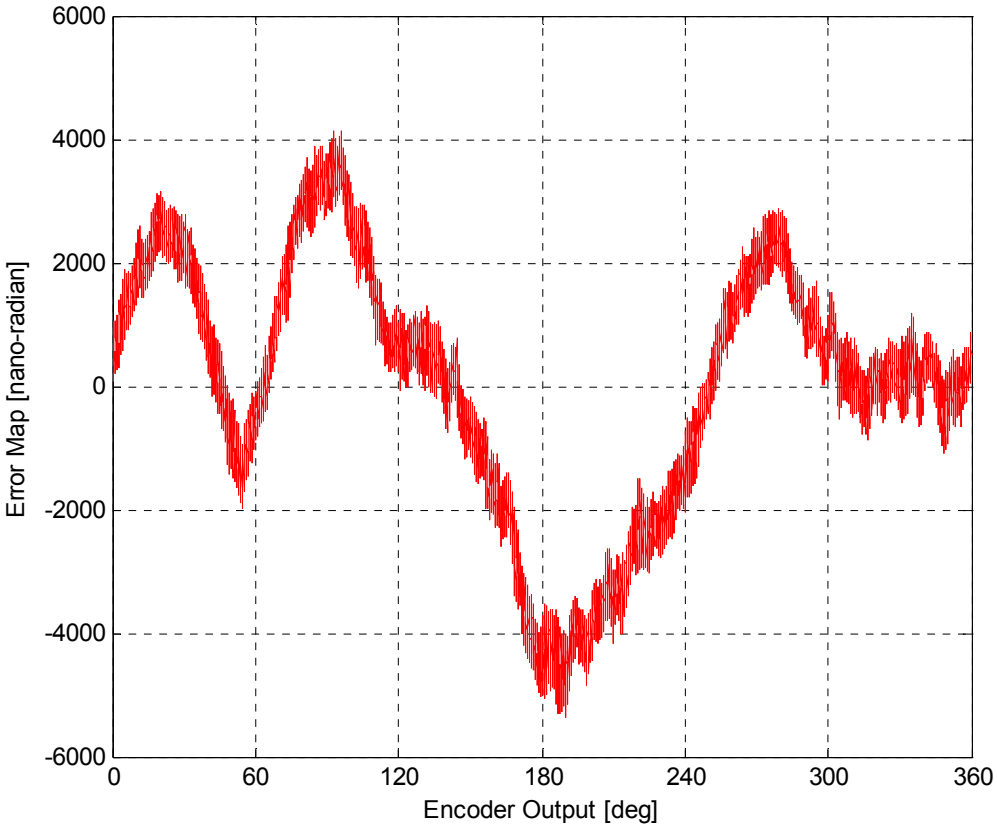

Figure 1-4: Error map obtained experimentally at 220 RPM Using a 200MHz Timer for the Heidenhain ERA4282C (serial# 23-502-234) encoder used in our rotary table design.

CHAPTER 2

Precise Rotary Table

This chapter presents the mechanical design of the precise rotary table (PRT) which is a motorized airbearing-based rotary platform designed to achieve repeatable motion and linear dynamics, both of which are required for effective encoder calibration. The rotary table is equipped with two rotary encoders and a 3D displacement measurement system. The platform is used for testing and development of rotation measurement and control methods.

2.1 Design Objective

The main design objective of the mechanical setup is to demonstrate repeatable motion and linear dynamics because the achievable rotation measurement accuracy depends on these two qualities. Encoder calibration will be used to enhance the accuracy of our encoders and achieve better rotation measurement accuracy. Calibration is only effective for removing synchronous (repeatable) error and does not eliminate asynchronous (random) error. Therefore, the mechanical setup is designed to minimize any possible random effect, or in other words, the mechanical setup is designed for repeatability of motion. Moreover, the calibration method finds the error by establishing a linear model

of the system dynamics. Non-linear dynamics should be minimized as they cannot be represented by the calibration algorithm's linear model. The PRT is a research tool which will be used for exploring and improving the rotation measurement and control techniques. So, the design includes features, such as two encoders with multiple read-heads and a error motion identification system, which enable investigating different research scenarios.

2.1 Configurations

The PRT is designed around a Professional Instruments -10R Block-Head® airbearing which has low error motion and high stiffness as indicated in table 2-1. The rotary table is motorized using a brushless ThinGap TG8263® motor which has minimal torque ripple due to the moving iron-core design. Two different rotary encoders, referred to as Encoder-1 and Encoder-2 from now on, are utilized. Encoder-1 works at all speed ranges and is intended as the final rotation measurement device. Encoder-2 has a stable error-map and is incorporated to compare our results to rotation measurement standards available elsewhere. Also, a displacement measurement system consisting of a precision machined ball and three capacitive probes is incorporated to monitor the spindle error motion. The whole setup is mounted to a granite table top, which is supported by vibration isolation mounts.

Table 2-1: Specifications for Professional Instrument's 10R Block-Head ® airbearing with 150-psi operating pressure [16]

		Ultimate	Working
Load Capacity	Radial	1780 N	890 N
	Axial	10700 N	5350 N
	Tilt	680 N-m	340 N-m
Stiffness	Radial	350 N/micrometer	
	Axial	1750 N/micrometer	
	Tilt	11.3 N-m/microradian	
Error Motion	Radial	< 25 nanometers	
	Axial	< 25 nanometers	
	Tilt	< 0.1 microradian	
Maximum Speed		1800 rpm	

Given the design of Encoder-2 and the displacement measurement system (DMS), they cannot be mounted onto the same side of the airbearing at same time. One way to overcome this issue is to move one of the two to the bottom side. Therefore, a configuration design, shown in Figure 2-1, where Encoder-2 is mounted onto the bottom side was considered. Moving Encoder-2 to the bottom side creates a long mechanical path, as indicated by *L* in figure 2-1, and introduces Abby error relative to the Encoder-1 rotation measurement. Because it is unnecessary to operate both Encoder-2 and DMS at the same time, another configuration was selected where the user can select which one of the two to use by dismounting one and mounting the other one. The two settings are shown in Figure 2-2. The advantage of the latter configuration design is that it uses the airbearing to isolate the sensitive sensors from any unbalanced force disturbance from the motor. In addition, the sensors are more accessible and can be more easily and better aligned.

The latter configuration where the sensors are mounted on top is selected as the overall configuration of the mechanical setup. The configuration design in Figure 2-2 only shows the general relative configuration of the major parts of the design: sensors, motor, and airbearing. Typical to all the operating modes, an airbearing in the center separates the motor mounted on the bottom side from the sensors mounted on top side. However as mentioned before, the mechanical setup can be adjusted to operate in several modes, each customized for a certain investigation. For example, the system can incorporate two encoders at the same time or can have one encoder and the DMS. The details of the operating modes will be discussed in the subsections to follow.

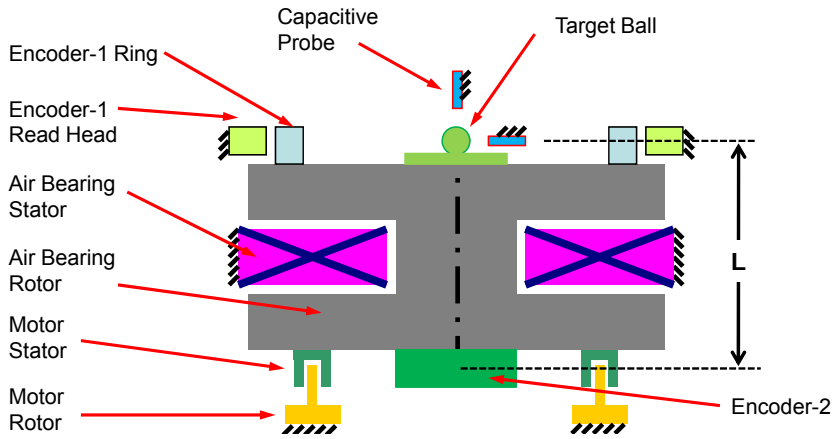

Figure 2-1: Schematic view of mechanical configuration design - sensors on two sides

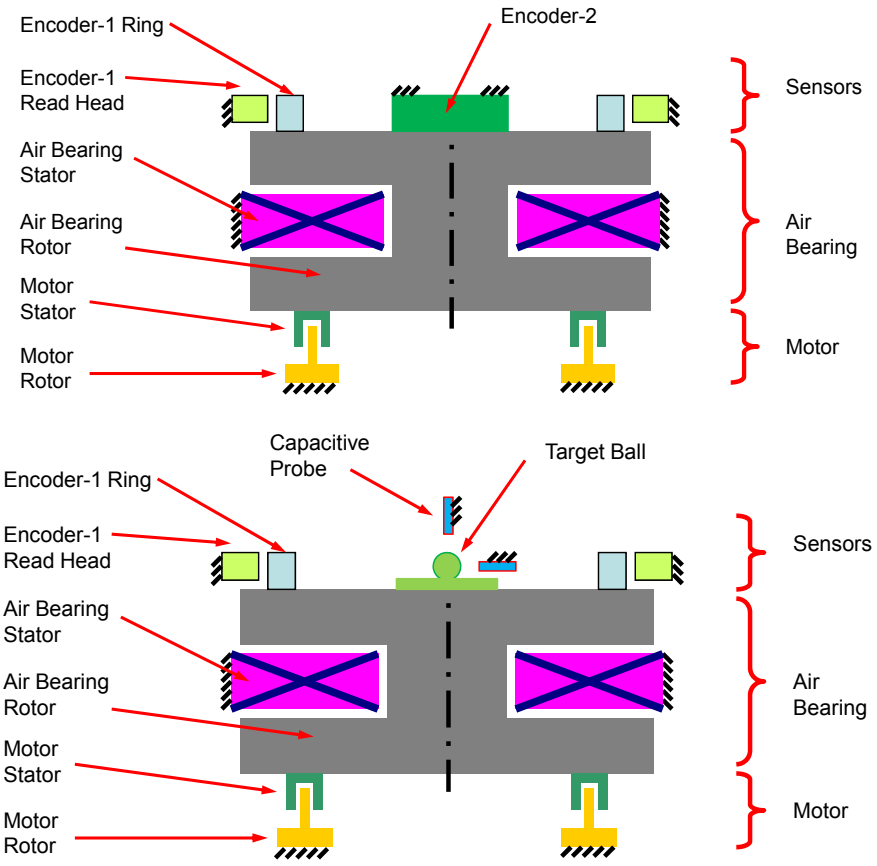

Figure 2-2: Schematic view of mechanical configuration design - sensors on top
(top: with two encoders, bottom: one encoder and DMS)

2.1.1 Two-Encoder Mode

In this operating mode the setup is positioned vertically and incorporates two different encoders at once. This operating mode allows for simultaneous use and calibration of two different encoders under the same conditions. The Encoder-1 is selected to serve as the final rotation measurement device, and Encoder-2, whose error-map is relatively stable, is used to compare our calibration results with encoders calibrated by other methods.

The cross-section view of the whole assembly is shown in Figure 2-3. In this mode, both encoders are mounted on the top side and measure the rotation of the airbearing rotor relative to the bearing's stator at the same time. Typical to all the operating modes, the motor is installed on the bottom side and the whole structure is attached to a granite table. The role and features of each encoder and is discussed in more details in section 4.3.4.

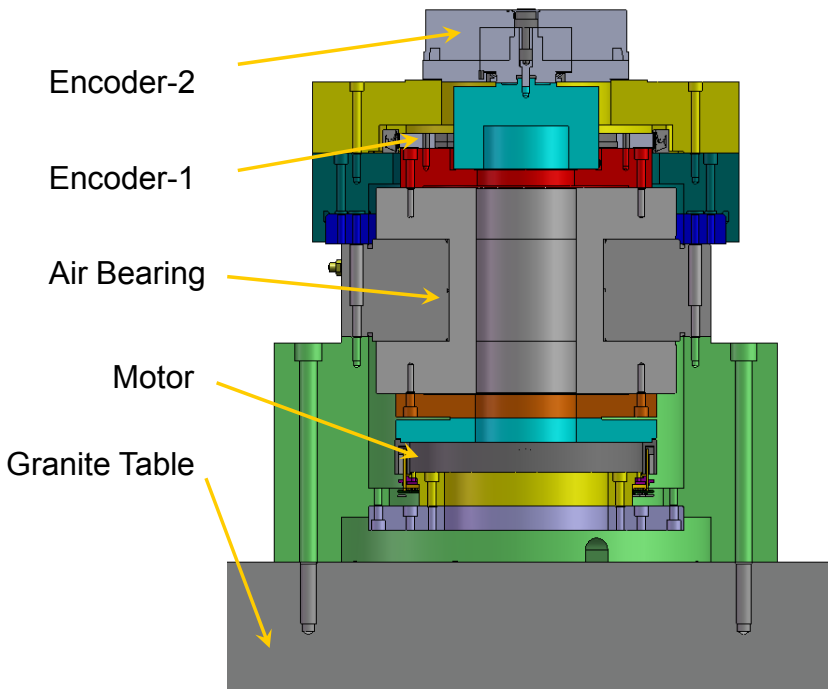

Figure 2-3: Cross-section view of the two-encoder operating mode

2.1.2 Encoder and 3D Displacement Sensor Mode

In this operating mode the setup is positioned vertically and utilizes one encoder and a 3D displacement measurement system (DMS). Figure 2-4 shows the cross-section view of the Single Encoder plus Displacement Sensors. Encoder-1 is left unchanged on the top side; however, Encoder-2 is replaced with the 3-D displacement measurement system.

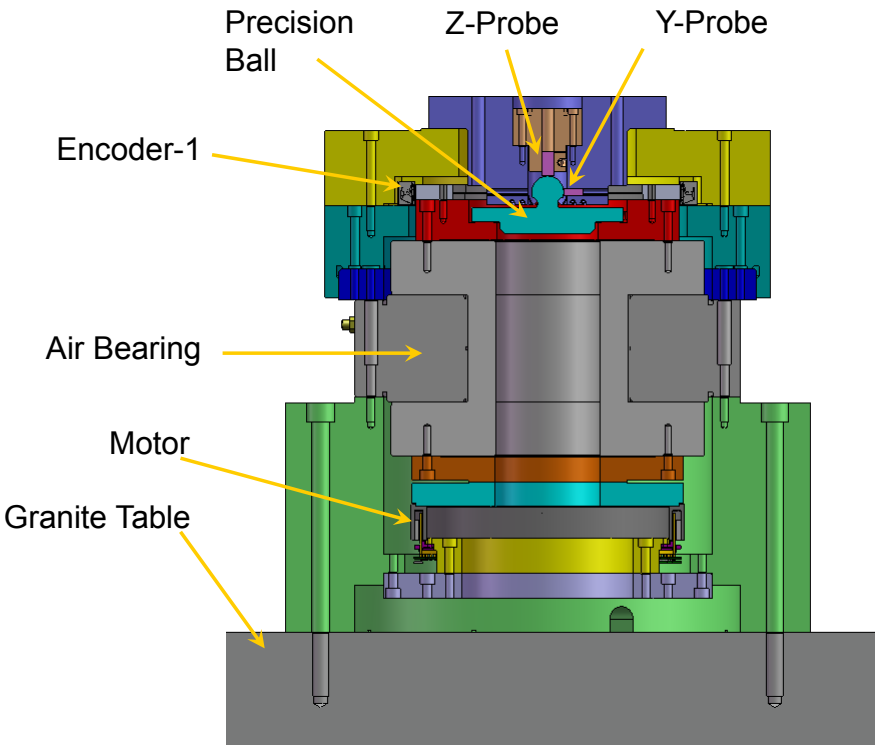

Figure 2-4: Cross-section view of the encoder and 3D displacement sensor mode

In this operating mode, Encoder-1 measures rotation of the airbearing while the motion of the center of the encoder ring is monitored in the three translational degrees of freedom (DOF) using the 3D displacement sensing system. An ideal bearing would perfectly constrain all DOFs except for one rotational DOF. However, no bearing is ideal and there

is unavoidable error motion associated with the bearing. In this operating mode, the error motion is identified, and its effect on rotation measurement can be studied and compensated. In the mechanical design, the ball's center and the midpoint of the encoder drum axis coincide. This ensures that the motion measured by the DMS is the translation of the encoder drum and is not coupled to any other rotational degrees of freedom.

2.1.3 Horizontal Mounting Mode

The two previous operating modes are arranged vertically. It is possible to change the orientation of the whole setup to be horizontal. This configuration is referred to as Horizontal Mounting Mode and its cross-section view is shown in Figure 2-5.

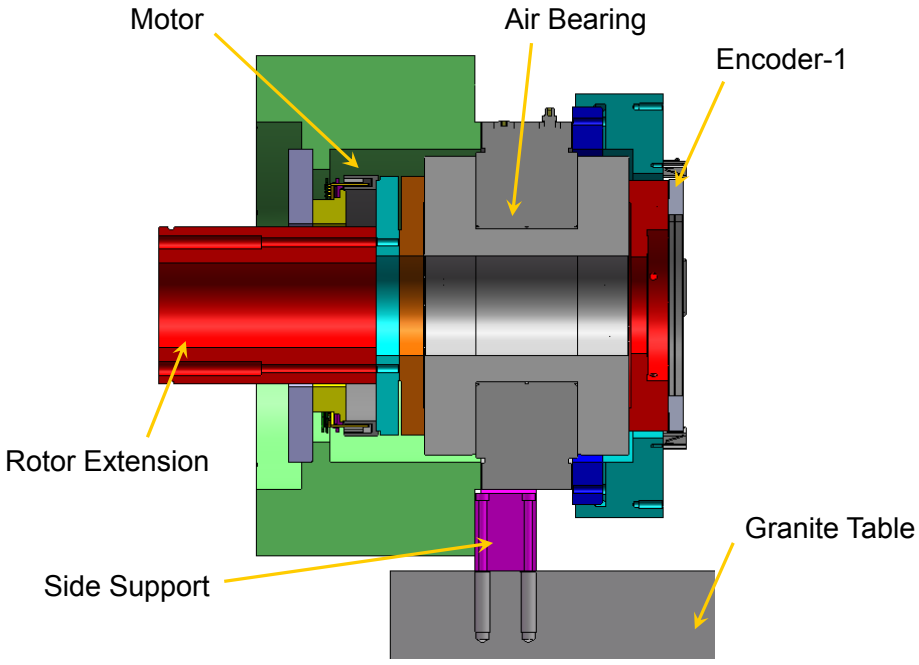

Figure 2-5: Cross-section view of the horizontal mounting mode

The setup is fastened to the granite table using an additional side support piece which uses a V-section to hold the outer diameter of the airbearing's stator. The rotor

Extension fixes to the spindle rotor and extends the rotor. This arrangement creates access to the motor as well as the spindle rotor. This operating mode is designed mainly for motor torque constant identification. We propose to use a hanging mass wrapped around the rotor extension to provide a constant torque; then, monitor the current required to act against that constant torque to find the variations in the torque constant as a function of rotation angle.

2.2 Encoders Assembly

The encoders are utilized for rotation measurement of the rotary table. The encoders are the central part of the design because the achievable angular motion accuracy is directly dependent on the rotation measurement accuracy from the encoders. To add to the versatility o f this setup as a research tool, two different encoders are incorporated into the design: one encoder without an integral bearing, Encoder-1 in the configuration diagrams, and an encoder with its own built in bearing system, Encoder-2 in the configuration diagrams. Generally speaking, angle encoders with integrated bearings compared to the the encoders without integral bearings have measurements which are less sensitive to mounting variations. On the other hand, they have lower operating speed range, and their integrated bearing can influence and degrade the system dynamics. In PRT, we have integrated both to benefit from the features of each type. This section presents the selected encoders and their mounting design.

2.2.1 Encoder without Integral Bearing

Heidenhain's ERA4282C encoder with 32768-lines is selected as one of the encoders that are used within the setup. The ERA4282C encoder is shown in Figure 2-6.

The encoder consists of a line carrier drum which will be mounted and centred to the air bearing rotor, and a read-head to scan the lines, which will be fixed to the stator.

Figure 2-6: ERA4282C encoder from Heidenhain [courtesy of Heidenhain] [17]

The ERA4282C does not have an integral bearing, meaning that the read-head and the drum are separate parts and need to be constrained with respect to each other using an independent bearing, such as the airbearing. To improve rotation measurement, four read-heads, which are equally spaced around the encoder drum, are used in the PRT design to obtain four simultaneous angle measurements based on a single rotating encoder drum. Figure 2-7 shows the read-heads' arrangement around the encoder ring. The cross-section view in Figure 2-8 shows how the encoder is mounted onto the air bearing. The red piece is used to attach the drum to the airbearing rotor and the green and the blue pieces are used together to fix the read-heads to the airbearing stator and position them around the drum.

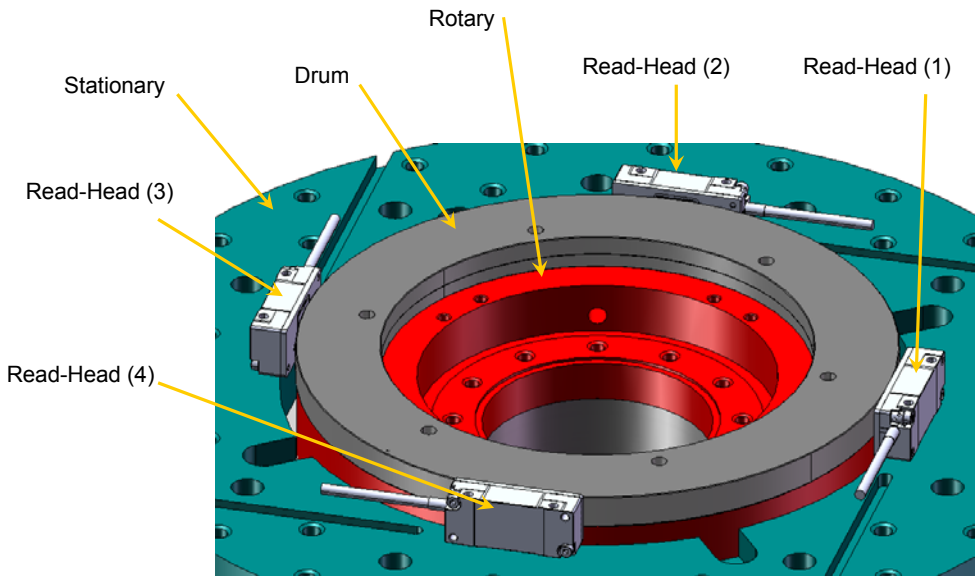

Figure 2-7: ERA4282C read heads arrangement in the PRT design

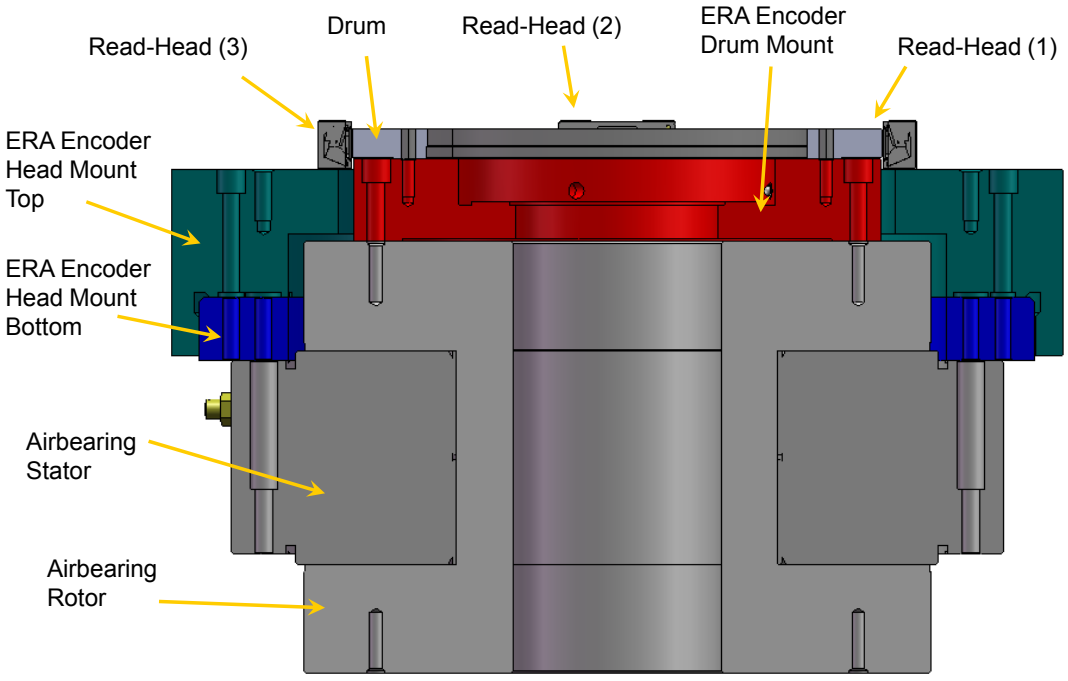

Figure 2-8: Section-view of the ERA4282C encoder's mounting onto the airbearing

Centering and alignment of the encoders in the setup is very important because it directly influences the achievable rotation measurement accuracy. Any eccentricity of the

drum with respect to the rotor or any improper alignment of the read-heads to the drum would introduce rotation measurement error. As a result, mechanisms are included in the design to ensure centering and alignment accuracy to less than 0.1 μm. Figure 2-9 shows the centering mechanism used with the ERA encoder Drum. A capacitive displacement sensor looking at the outside of the drum is fixed by two V-groove fixtures. As the rotor rotates, the capacitive sensor measures the eccentricity of the drum. The measured eccentricity is corrected by adjusting the four finely threaded set screws, which push on the inside of the encoder drum through a flexible element. Each pair of opposing set screws constrains the drum in one translation degree of freedom. The flexible element enables finely adjusting the position by tightening one setscrew without the need for loosening its opposing setscrew. The two pairs together can adjust the center of the drum in the horizontal plane. The fine 100 threads/inch setscrews enable adjustment resolution of better than 700-nm per degree.

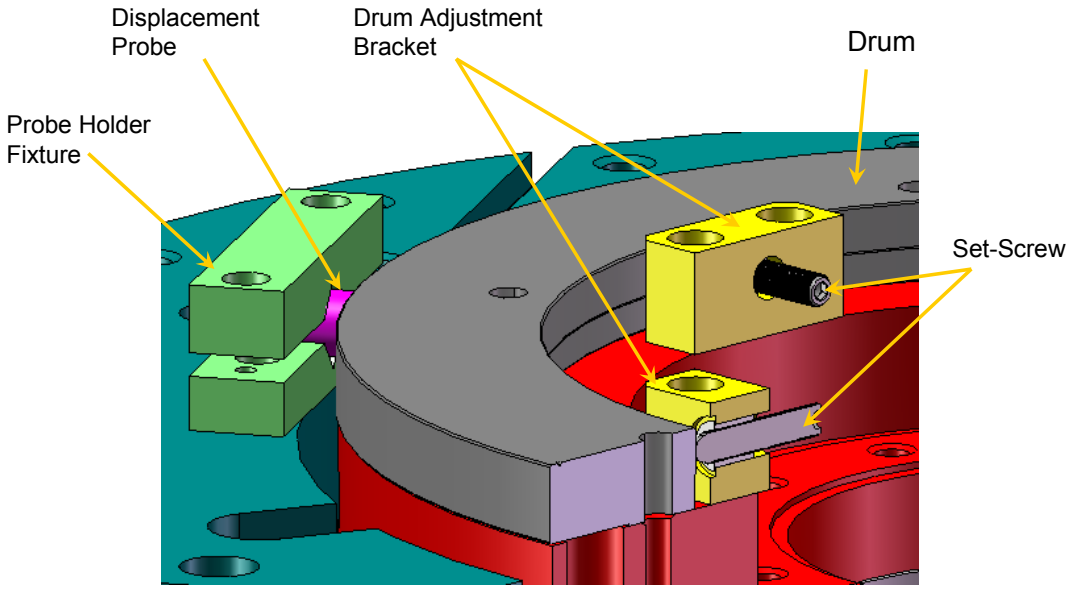

Figure 2-9: Section-view of the encoder drum adjustment mechanism

However, the experimental centering of the drum showed that tapping could be as effective as using the fine adjustment setscrews. An eccentricity of better than 300 nm was achieved using both the setscrew mechanism and tapping. The achievable centering was found to be limited by the roundness of the encoder drum's outer surface which was used as the centering target.

The read-head, shown in Figure 2-10, must be aligned properly to the encoder drum. The two ball features on the read-head are used with a shim to position the read-head at the right radius to the drum as shown in Figure 2-11.

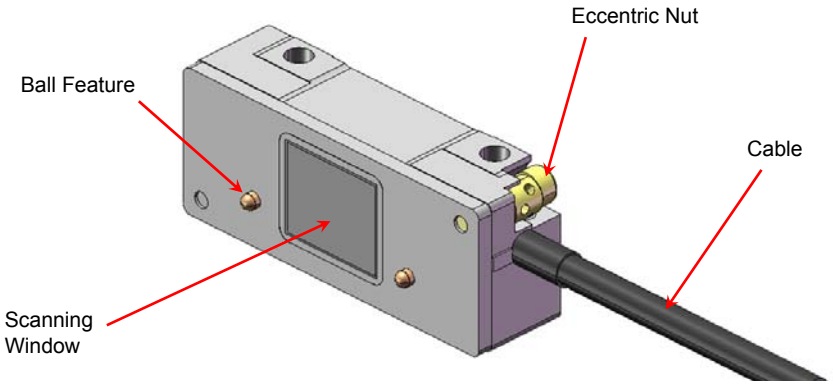

Figure 2-10: ERA4282C read-head

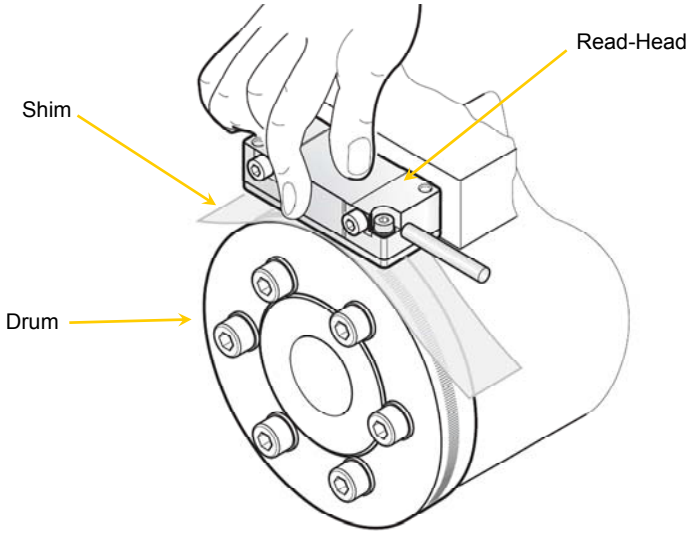

Figure 2-11: Radial adjustment of the read-head for the ERA4282C encoder [17]

Still, the read-head needs to be adjusted tangent to the drum to match the phase of the output signals from the four read- heads. One method considered for tangential adjustment was to use a flexural bearing along with setscrews as shown in Figure 2-12. The two radial flexure leaves constrain five degrees of freedom of the read-head while allowing free tangential motion. Two setscrews from left and one setscrew from right help adjust and constrain the bearing supporting the read-head. The flexure method inevitably lowers the stiffness of the read-head's support. This is against the design goal to achieve a consistent and precise setup. As an alternative, a temporary fixture can be used to adjust the read-head tangent to the drum while constraining other movements. As shown in Figure 2-13, the fixture uses four set screws which push against the read-head. Using this fixture, the head can be moved tangentially while constrained in other degrees of freedom. To enable smoother adjustment, flexible rubber pads can be used in front of the setscrews and the head can be moved by preloading and compressing those pads. This will create a continuous and levered motion.

Finally, the eccentric nut built into the encoder read-head is used to adjust the sensing window of the read-head. Rotating the eccentric nut will rotate the sensing window around its normal axis and will move it in its plane at the same time. Throughout the adjustment process, the encoder signal is used as the feedback for aligning the read-head. The signal is viewed using an oscilloscope. The alignment adjustment pursues the following goals:

1. Maximize the individual read-head signals
2. Align the individual read-head incremental signal with its reference signal
3. Align the phase of the four read-heads' signals

Due to time constraints, so far, we have aligned one read-head using the two ball-features only. Inspecting the results, we found the alignment to be satisfactory.

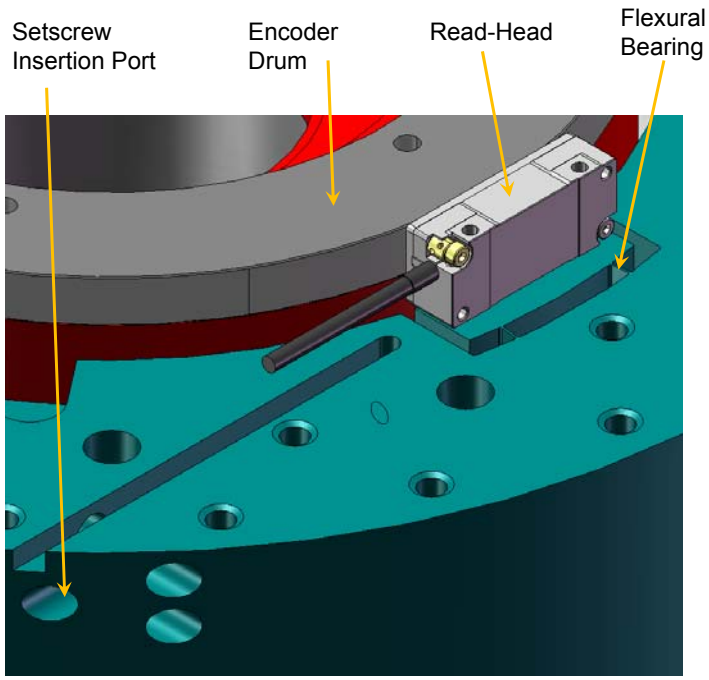

Figure 2-12: Tangential adjustment of the read-head for the ERA4282C encoder – flexural bearing method

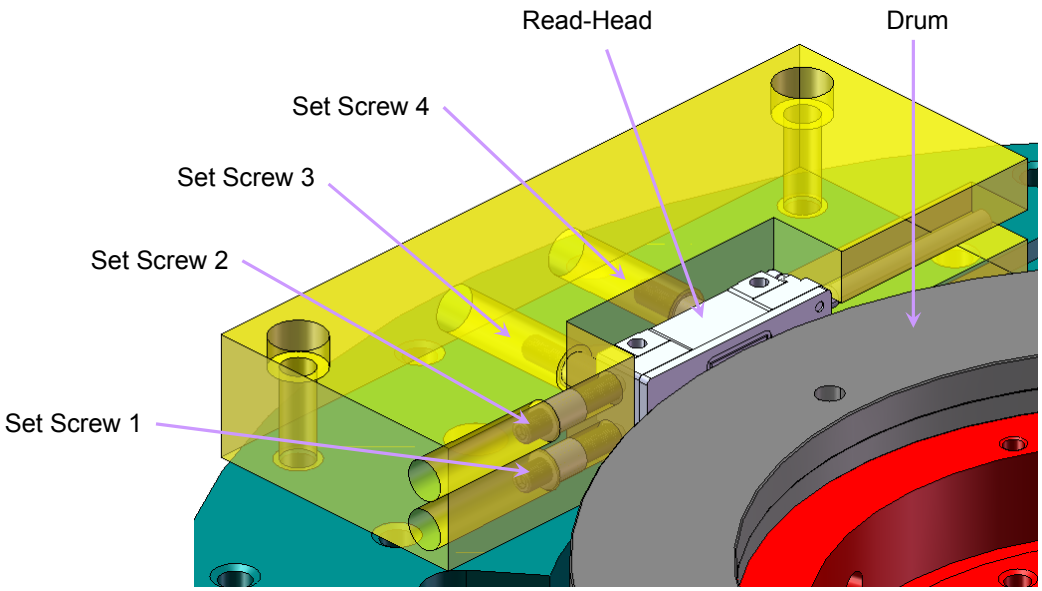

Figure 2-13: Tangential adjustment of the read-head for the ERA4282C encoder – rigid removable fixture method

2.2.2 Encoder with Integral Bearing

The second encoder selected for this setup is RON905 from Heidenhain which is shown in Figure 2-14. The RON encoder has an integral bearing which constrains the grating carrier to the scanning read-head; therefore, it appears as a single part with rotary and stationary sub-parts. The rotor is centered and fixed to the airbearing rotor, and the stator is centered and fixed to the airbearing stator. Internal flexures of the encoder allow its read-head to move relative to the housing to avoid over-constraint. The cross section view in Figure 2-15 shows the mounting design for the RON905 encoder. The RON Rotor Mount piece extends the rotor further to connect to the rotary part of the RON905 encoder. The Ron Stator Mount fixes the stationary part of the RON905 encoder to the stator trough the ERA encoder head mount pieces.

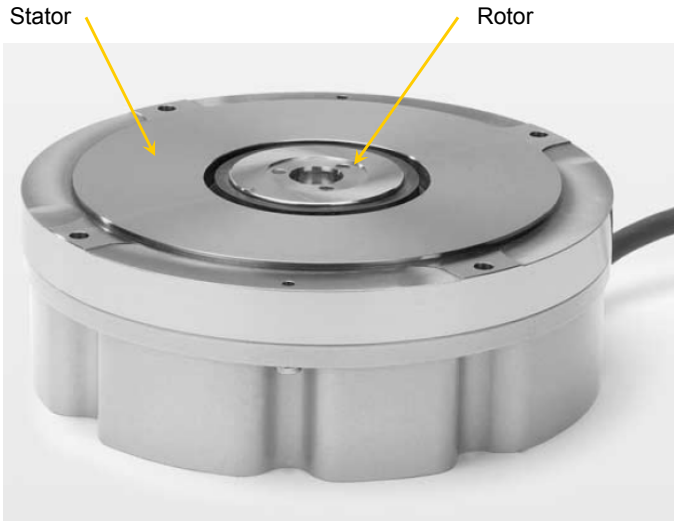

Figure 2-14: RON905 encoder from Heidenhain [courtesy of Heidenhain] [17]

Similar to the ERA4282C encoder, the RON905 has alignment requirements that need to be ensured. In the PRT setup, the mounting pieces are centered to the rotor using a dial indicator. Once the mounting parts are centered to 10μm, clearance positional fits are used to center the RON encoder to the mounting pieces and hence to the airbearing

rotor. Figure 2-16 zooms on the section view to show the fits used for positioning the encoder.

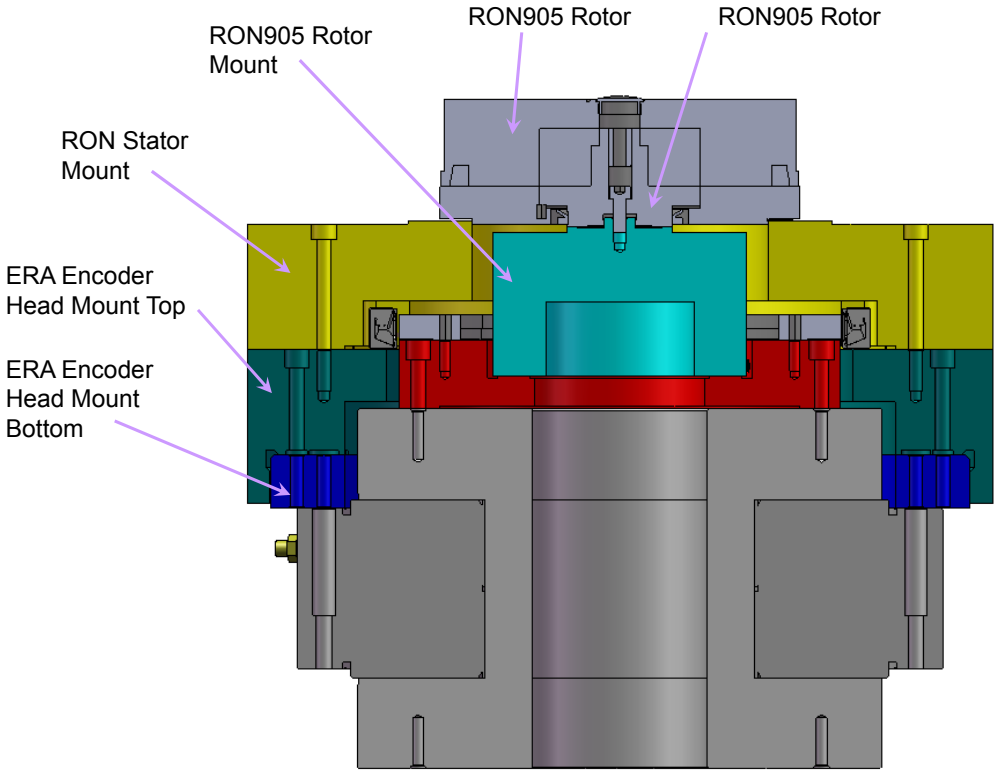

Figure 2-15: Section-view showing the mounting of RON605

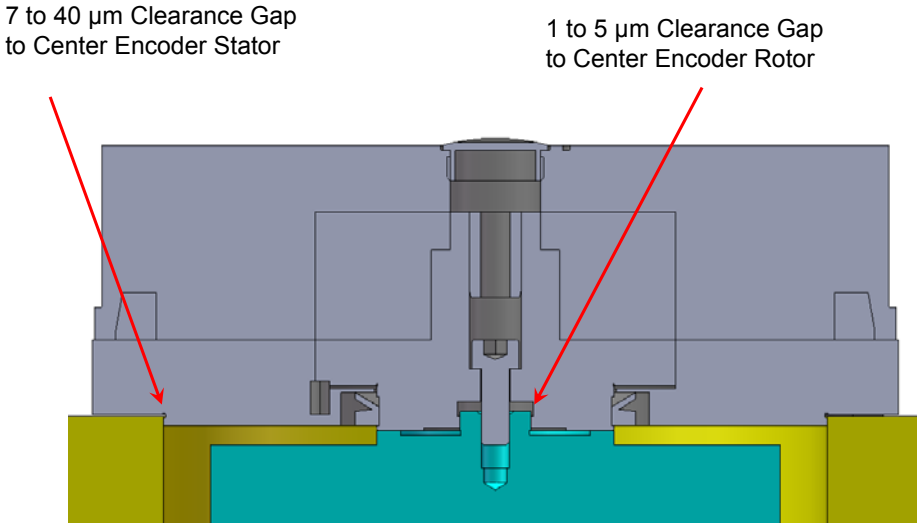

Figure 2-16: Section-view showing the fits used for centering the RON905 encoder

2.3 Ball-Probe Assembly

The ball-probe setup can be used to identify the spindle error motion in three translational degrees of freedom (DOF) at the center of the ERA4282C encoder drum. The error motion is significant because it influences the angular measurement accuracy. When a single encoder read-head is used, the airbearing's 25-nm radial error motion translates to 250 nano-radians at the encoder drum read-out. The ball probe setup consists of a precision machined target ball with a roundness of about 10-nm, which is centered onto the rotor, and three displacement sensors which look at the ball in the three orthogonal directions. Figure 2-17 shows the cross section view of the ball probe setup.

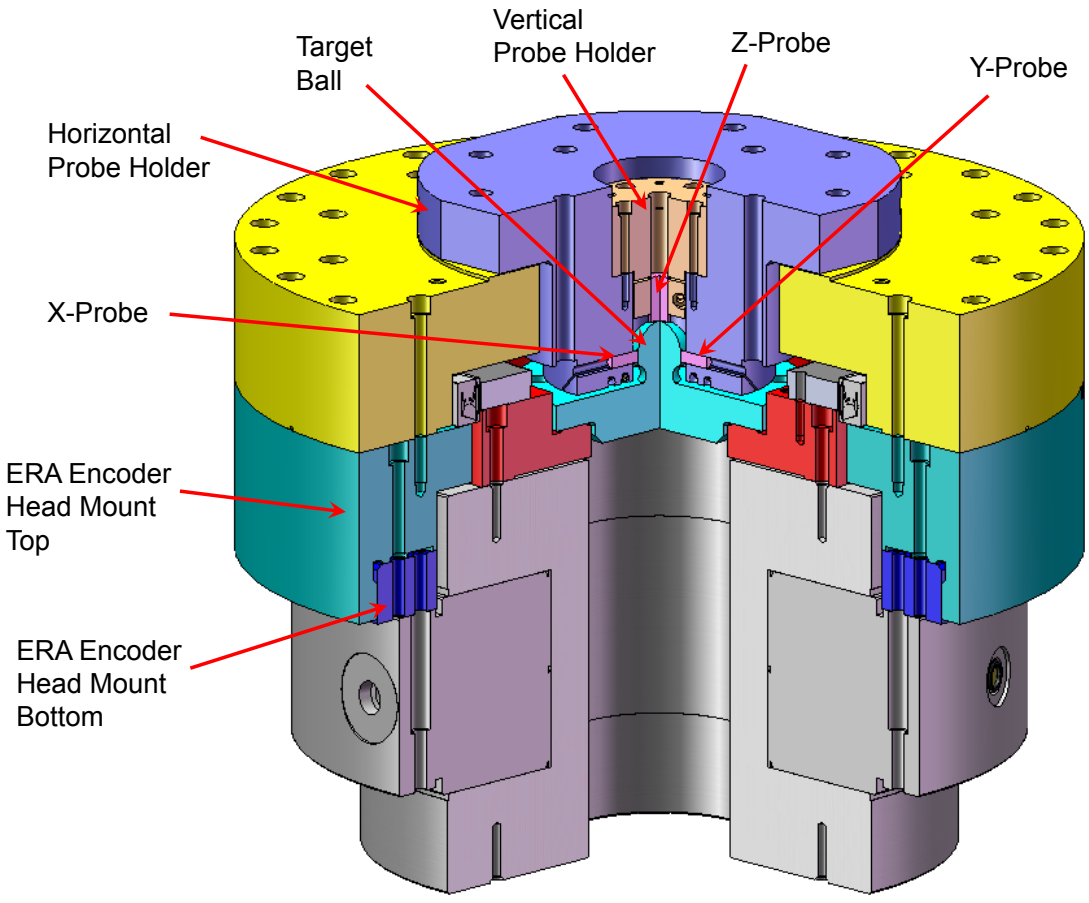

Figure 2-17: Section-view of the ball-probe setup

2.3.1 Capacitive Probes

The capacitive displacement sensors are used for measuring the motion of the target ball. The cylindrical shaped probes are shown in Figure 2-18. The vertical error motion does not affect the rotation measurements directly, and hence, has lower precision requirements compared to the horizontal probes. As a result a probe with a larger range but lower resolution is selected for the vertical direction. The vertical probe is slightly larger than the two horizontal probes. The probe holder pieces fix the probes to look at the target ball. Each probe is secured inside a cylindrical clamp built into the probe holder pieces. The probes need to be aligned to look at the equator of the ball. Also, the probes are designed to work at specific distance from the target; therefore, the gap between the probes and the ball needs to be adjusted. Figure 2-19 shows the adjustment mechanism for the probes.

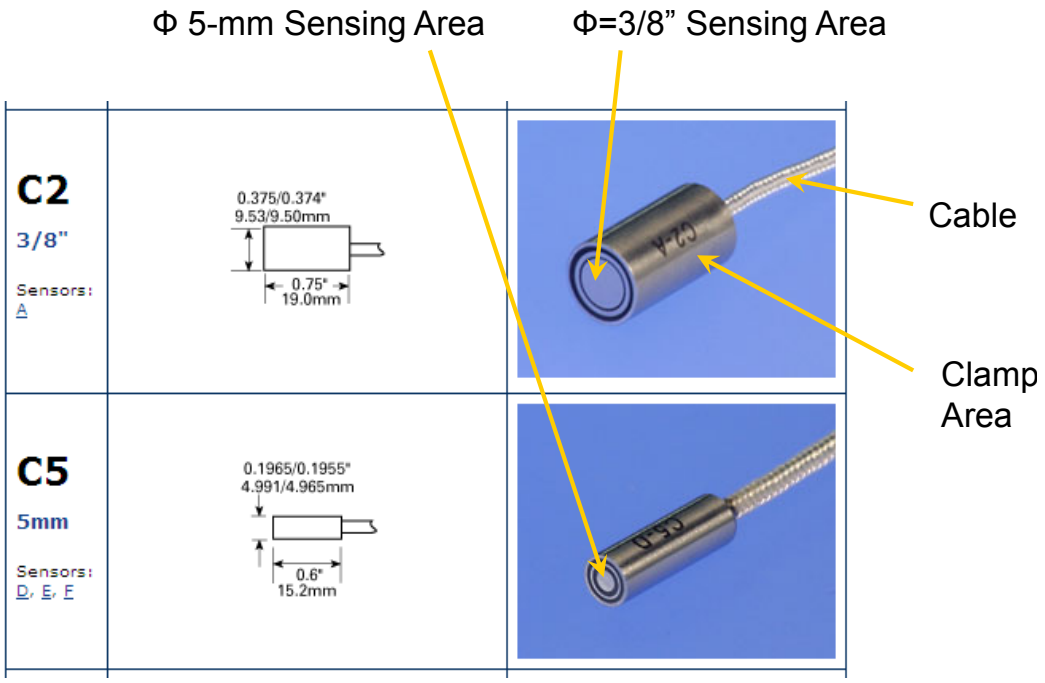

Figure 2-18: Vertical (top) and horizontal (bottom) Lion Precision capacitive probes

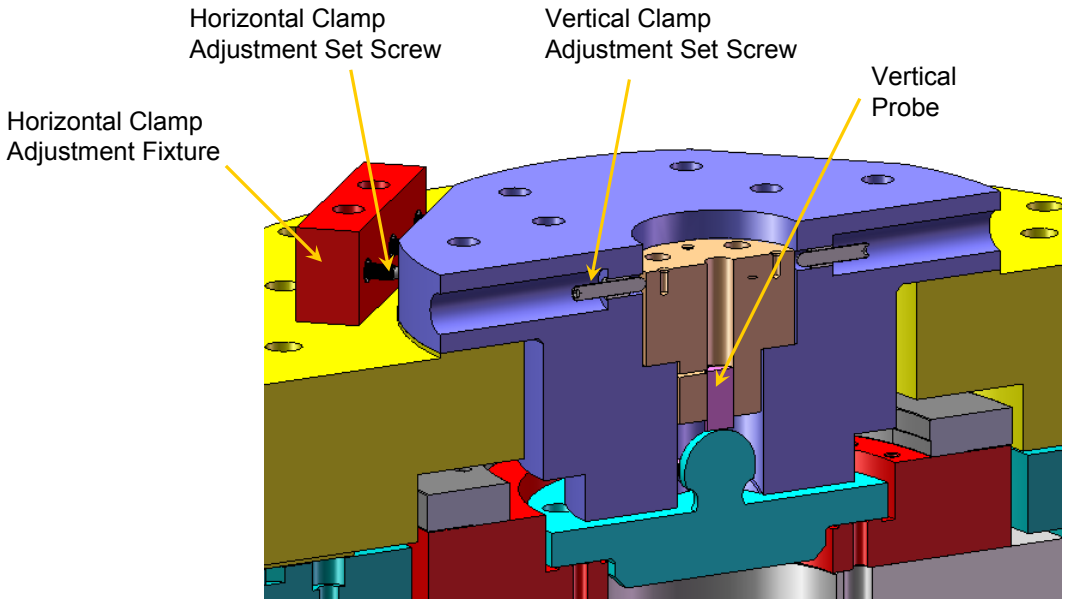

Figure 2-19: Section view showing the probe alignment mechanisms

To adjust the gap spacing of the horizontal probes, four Horizontal Clamp Adjustment Fixtures are temporarily added to the setup. Each adjustment fixture will have three set screws which push on the sides of the horizontal clamp. Each two opposing adjustment fixtures enable adjustment and firm constraint of one horizontal probe. The gap spacing of the vertical probe is adjusted by moving the probe inside the vertical clamp. For aligning the horizontal probes to the equator the design relies on the vertical tolerances of the parts and can use precision metal shims between the purple *horizontal probe holder* and the yellow *RON stator mount* as a backup plan. Aligning the vertical probe to the ball center is done by using the four set screws placed every 90-degree around the vertical clamp. The four set screws can be used to move and constrain the probe to point at the equator of the target ball.

2.3.2 Precision Target Ball

The precision machined ball is used as a target to monitor the error motion of the spindle. The target ball, manufactured by Professional Instruments, is 1-inch in diameter and has a nominal roundness of 10-nm or better if used with capacitive probe sensors. It is necessary to center the ball because any eccentricity of the ball relative to the rotor will show up as a rotation synchronous component on the error motion measurements. Figure 2-20 shows the centering mechanism. A capacitive probe sensor, held in place by a probe holder fixture and V-shaped clamp, is used to identify eccentricity of the ball. The identified eccentricity is corrected by using the set screws which push against the base of the target ball. The four setscrews in the design enable adjustment of the ball center in the horizontal plane. The pockets cut into the top ERA Encoder Head Mount enable adjustment of the setscrews without having to remove the ERA Encoder Head Mount.

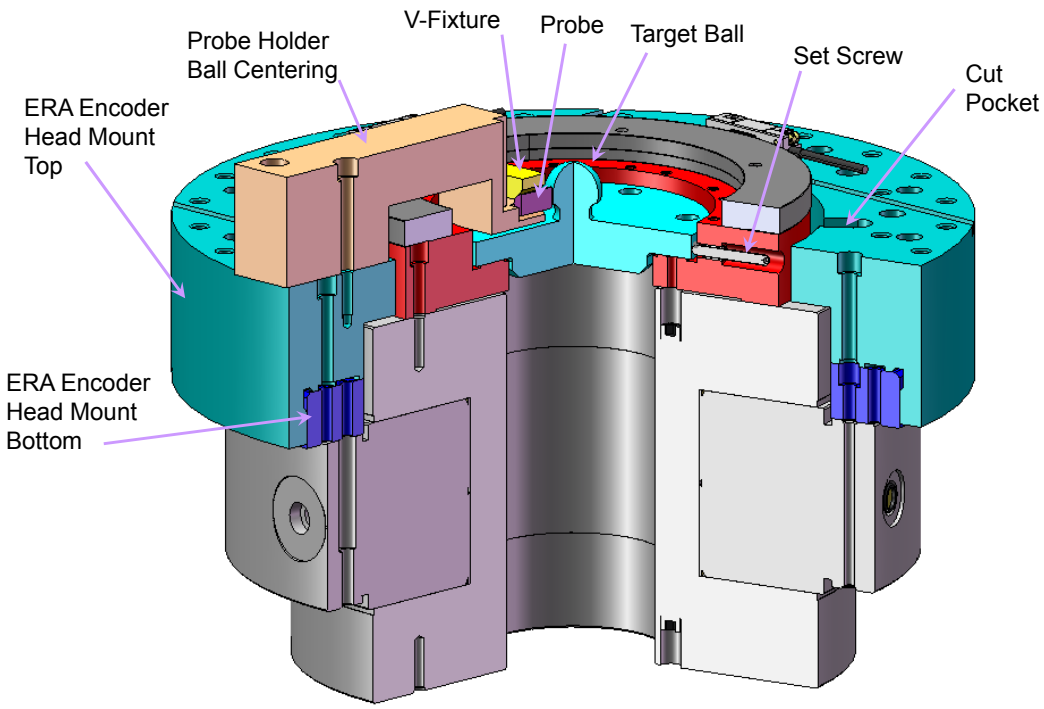

Figure 2-20: Section-view showing the centering mechanism for the target ball

As discussed in the previous section, the capacitive displacement sensors can achieve sub nanometer resolution. However, the target ball's 10-nm roundness limits motion detection resolution to ±10 nm. The ball roundness error can be identified using reversal methods. The generated roundness calibration map can be used to compensated for the roundness error of the ball and improve the displacement measurement resolution. To perform the reversal technique the probes need to be adjusted precisely in rotation with respect to the airbearing stator. While moving the probes relative to the stator, it is important that everything else in the setup stays the same. Therefore, it is required to move the probes in the assembled setup. To avoid changing the probe to ball gap, which could introduce error due to non-linearity of the probes, and to avaoid damaging the target ball, we have incorporated a sliding bearing into the design which enables us to rotate the probes without moving them off the spindle center.

The sliding bearing is created by dividing the ERA encoder head mount piece into two parts: top and bottom ERA encoder head mounts. The top ERA head mount fits and slides on the bottom ERA head mount. During the adjustment process, the airbearing rotor will be fixed to its stator. The ERA head mount's top part will be released from the bottom part and will rotate together with the probes. As the probe rotate, the ERA encoder will measure the rotation angle because the read heads will rotate around the fixed encoder drum. The precise rotation measurement from the ERA encoder would enable us to adjust the ball's relative angular position with very high accuracy.

To safely rotate the probes and to minimize the effect of probes' non-linearity, the sliding bearing needs to have ±1μm bearing precision. As a result, the bottom and top parts' surfaces need to match very closely. To achieve best results without extensive cost,

we use surface replication techniques to match the surfaces very closely. The bottom part will be machined with tighter tolerances and will be used as the reference surface. The bottom part will have 5μm cylindricity as well as 10μm flatness and parallelism. The top part will be designed with a 0.1-mm to 0.2-mm clearance with the bottom part. The bottom part will be coated with mold release and the two parts will be put together with epoxy in between. Once the epoxy solidifies, the bottom part can be removed, and the top part's surface will match the surface on the bottom part.

2.4 Motor Assembly

A direct drive rotary motor is used to actuate the rotor. The TG8263 three-phase brushless motor from ThinGap is used for this setup. This motor does not have its own bearing; instead, the rotor and stator of the motor are attached to the airbearing's rotor and stator respectively. Figure 2-21 shows the cross-section of the motor.

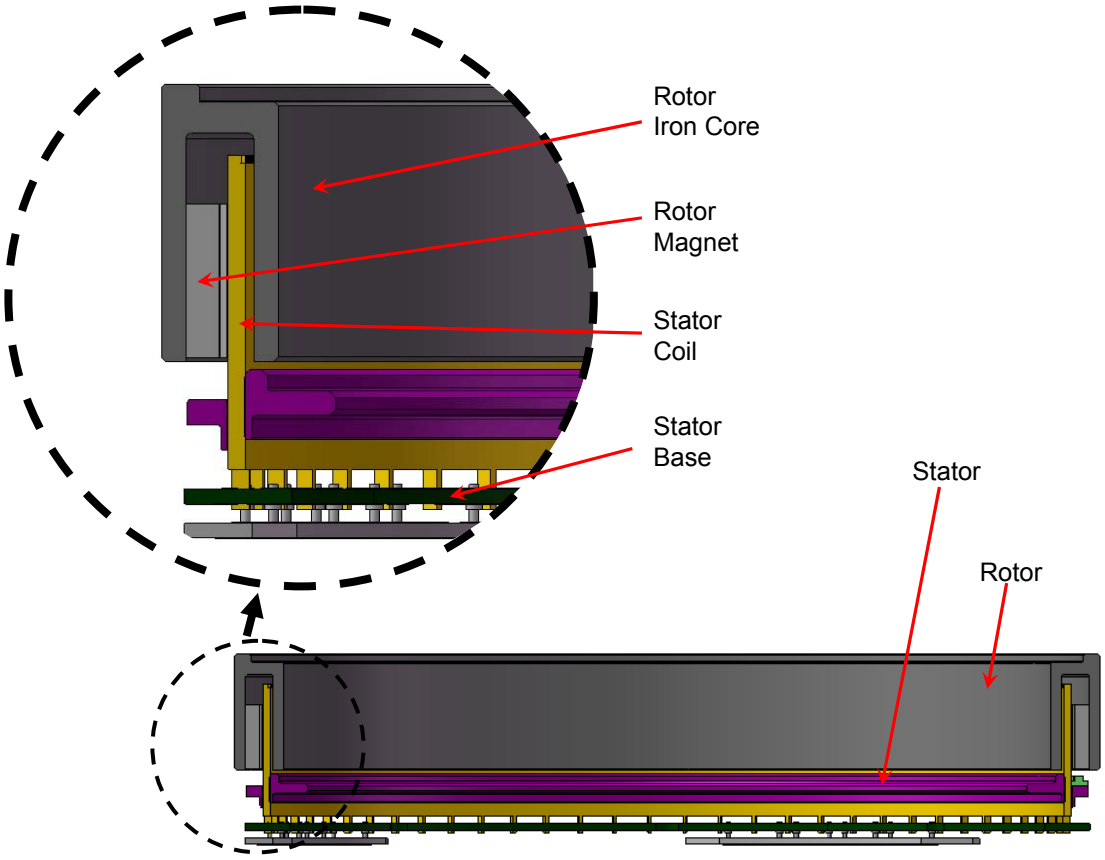

Figure 2-21: Section-view of the TG8263 brushless motor

The precision machined stator coil is placed within the inside gap of the rotor. The precisely machined coil is very effective in attenuating the torque ripples. The rotor consists of a revolved C-shaped iron core and a circular array of permanent magnets fixed to the core. Because the stator is core-less and the magnets do not move relative to the stator core, there will be no cogging torques from surface imperfection and no drag from

eddy-current losses [18]. As a result, the ThinGap motor is able to achieve very low cogging torque. ThingGap claims a maximum torque constant variation limit of ±0.045% [19].

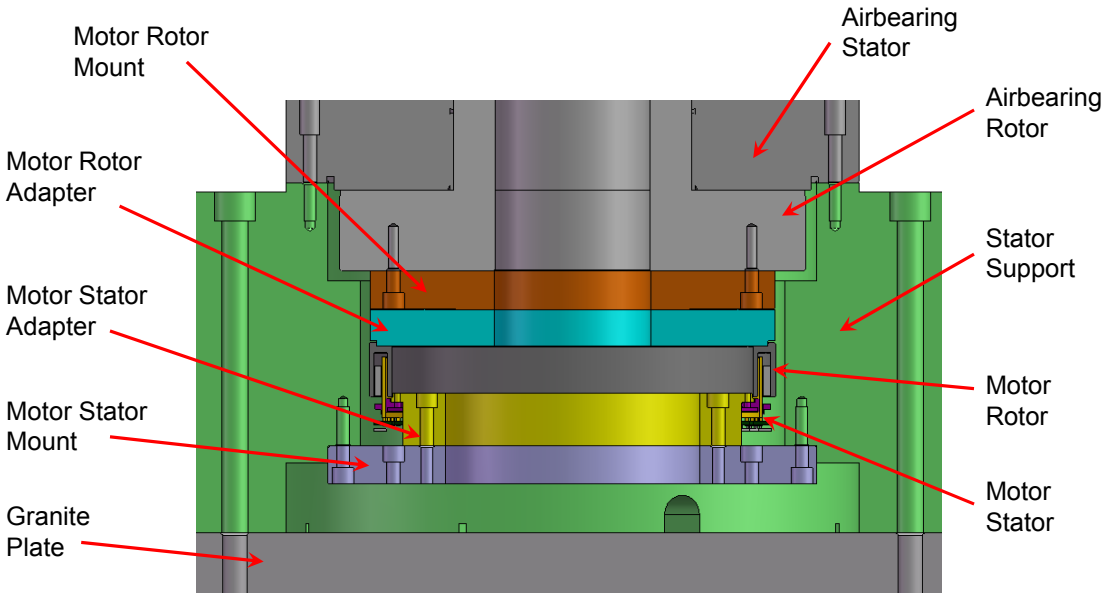

Figure 2-22: Section-view showing the mounting of the TG8263 motor

The threaded holes on top of the motor rotor are used to attach it to the airbearing rotor. The holes on the green stator base are used to fix it to the stator. The mounting of the motor onto the airbearing is shown in Figure 2-22. Because the bolt circle on the motor rotor and the airbearing rotor interfere, the Motor Rotor Adapter piece is used to translate the bolts of the motor rotor to a lower diameter circle. The motor rotor mount is then used to mount the motor rotor to the airbearing rotor. Due to space limitations in the thing gap motor design, the motor stator can be fastened only from the bottom of the stator base. Using a single mounting piece will not allow room for adding and tightening the bolts from the bottom; therefore, a Motor Stator Adapter part is used to enable a two step mounting process where the fastener can be easily accessed in each step.

The gap between the stator coil and the permanent magnets is very small, only 0.48mm. If the motor and the stator are not centered on the airbearing, there will be rotation dependent motor characteristic changes and there may even be mechanical interference between the rotor and the stator; therefore, the motor needs to be centered. The challenge in the stator installation is that the direct measurement of the gap between the motor stator and rotor is available not available during the installation. Several methods have been considered to ensure centered motor mounting:

1. Use clearance positional fits between motor mounting parts:

   The advantage of this method is that it does not require any adjustments during the assembly process. However, it is difficult to achieve good centering using this method given the long chain of part parts positioning the rotor and stator together.

2. Use a dial indicator to center parts one by one:

   Because parts are centered independently good centering can be achieved. The drawback of this method is that it requires round surfaces for dialing, and also it is not always easy to access the parts being centered

3. Use a hybrid combination of methods 1 and 2:

   This method uses dialing in, where possible, to achieve excellent centering and uses fits where access to the parts is limited. The combination of the two makes a good and feasible centering procedure.

4. Find the center based on the maximum ends [KT]:

   In this method suggested by KLA Tencor [KT], the motor stator is placed inside the motor rotor and is moved to both maximum ends. Based on the displacement from one end to the other end, the centered position is identified as the middle

point. The advantage of this method is its flexibility. This method is relatively difficult to implement and it is difficult to achieve good results without any visual feedback.

The third method is selected to be used for this setup. All the parts are centered using a dial indicator, except for the motor stator where a close fit is used. The fit between Motor Stator Mount and Stator Support is a loose fit and is not intended for immediate centering; it is to facilitate the blind insertion of the stator into the rotor. The fit between the Motor Rotor and Motor Rotor Adapter is a loose fits and is for mechanical strength not centering.

To center the parts using a dial indicator, a round surface on the part is required to be probed using the dial indicator. The fasteners fixing the part need to be accessible as well so that the part can be loosened to be adjusted. Motor Stator Adapter and Motor Rotor are fixed to other parts and are mounted onto the airbearing; therefore, their fasteners are not accessible once they are installed onto the airbearing. To solve this problem, these two parts are first centered to their mating parts: Motor Rotor is centered to Motor Rotor Adapter and Motor Stator Adapter is centered to Motor Rotor Mount. Then each pair of relatively centered parts can be treated as a single concentric part. They will be installed into the setup, and there, only their mating part needs to be centered.

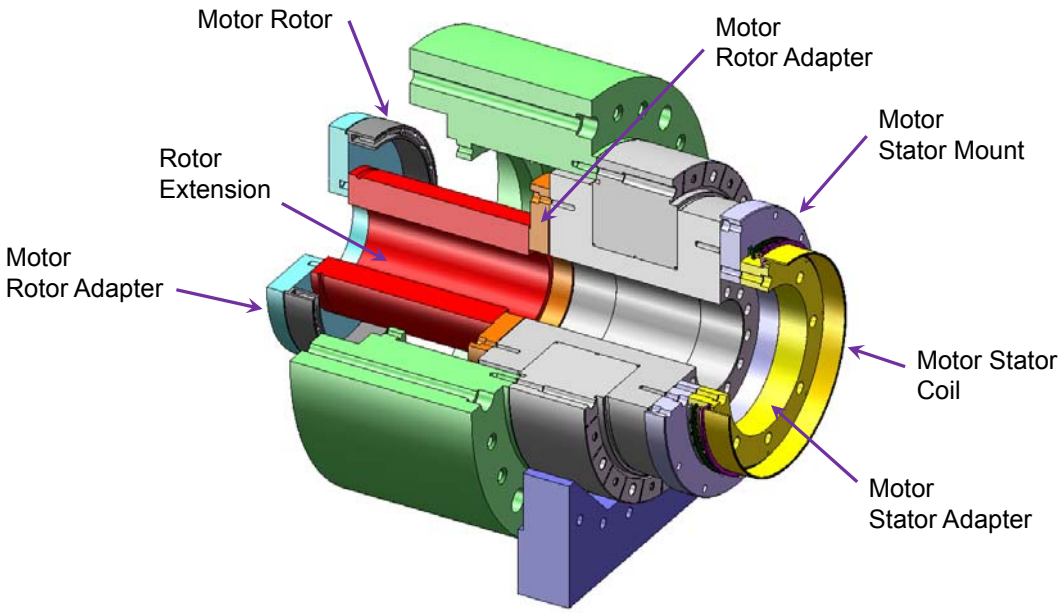

Figure 2-23: Section view showing initial centering of motor parts

The initial relative centering of the parts is shown in Figure 2-23. To relatively center two parts, each of them is centered to the airbearing rotor; this ensures that they are centered with respect to each other as well. On the left of the airbearing, the Motor Rotor is centered to the Motor Rotor Adapter; the rotor extension is used just as a spacer here. On the right side of the airbearing, the Motor Stator Adapter is centered to the Motor Stator Mount; the hole-pattern is designed such that the face of the air bearing can be used as a mounting plate. The following summarizes the motor assembly procedure:

1. Rotor Mount is installed and centered
2. Motor Rotor and Rotor Adapter are relatively centered
3. Stator Adapter and Stator Mount are relatively centered
4. Motor Rotor and Rotor Adapter pair is installed and centered
5. Stator Adapter and Stator Mount pair is installed and centered

2.5 Summary

This chapter presented the mechanical design of a motorized airbearing rotary table. The general configuration of the setup was discussed. The different operating modes of the setup were pointed out and their configuration designs were overviewed. The chapter continued by covering the mechanical design and installation procedure of the major components of the setup: encoders, ball-probe, and motor assemblies.

The setup consists of a direct torque drive brushless TG8263 motor from ThinGap at the bottom side of the air bearing and sensors on the top side of the airbearing. The selected sensors include an ERA4282C encoder with no integral bearing from Heidenhain, a RON905 encoder with integral bearing, and a 3D ball – capacitive probe setup. The setup can be configured into three operating modes:

1) Vertically oriented with two rotary encoders for cross-comparison
2) Vertically oriented with the ERA rotary encoder and a ball-probe setup for radial motion monitoring
3) Horizontal orientation of mode 1 or 2

CHAPTER 3

High-Speed Electronics

The performance of the electronics, in time and voltage domain, is a key to ensuring the required accuracy of our encoder calibration and motion control research. This chapter presents the design of custom electronics used to achieve speed and high quality signal processing required by the encoder calibration, which is not commercially available. A brief overview of software developed for the custom electronics is provided. Finally, a novel timer architecture, which has been developed as a part of this thesis, with up to 4-times improvement in timing accuracy is introduced in this chapter.

The custom electronics consist of three boards: motherboard, daughterboard, and powerboard. The motherboard is the main part of the custom electronics. It is designed around the Xilinx Virtex-4 FPGA and is used for high-speed digital signal manipulation and processing. The Virtex-4 FPGA incorporates a PowerPC405 processor core, so it can process the signals using both the processor and the FPGA fabric and logic circuits. The motherboard has 512 MB of Double Data Rate Read Access Memory (DDR RAM). It incorporates five 16-bit 100MHz ADCs and four 40-MHz DACs. It also utilizes Gigabit Ethernet and High-Speed USB for communication with other devices. The daughterboard is used for application specific analog signal processing. The daughterboard includes high-speed precision buffers and fast comparators for encoder signal digitization. The

powerboard consists of high quality power supplies for the motherboard. The powerboard generates linear power levels of ±15, ±5, and 3.3 volts for the analog components and switching power levels of 1.2, 1.8, 2.5, 3.3, and 5 volts for the digital components of the motherboard. Figure 3-1 shows the motherboard and the daughterboard connected together.

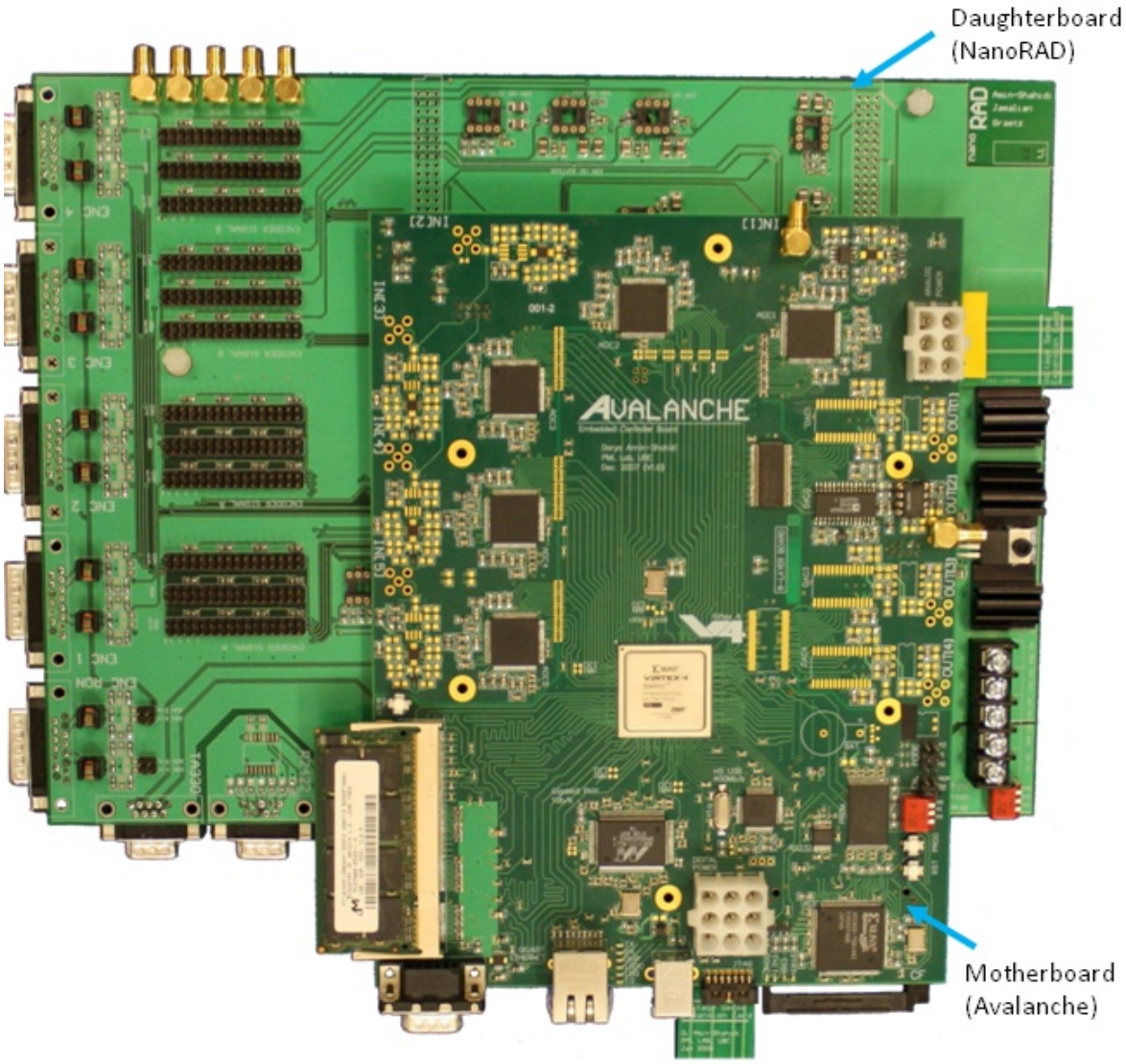

Figure 3-1: High-speed electronics – motherboard attached onto the daughterboard

3.1 Design Objective

The broad goal of the designing custom electronics is to achieve the speed and the quality required for this research. Each portion of the research has specific requirements which lead to certain electronic design goals. These requirements and the design goals will be discussed in this section.

3.1.1 Time Resolution Requirement

High timing resolution is needed for accurate encoder calibration. As discussed in chapter-4, the time resolution directly limits the accuracy of the encoder calibration map at a specific rotation speed. As a result, better timing resolution, or in other words, faster timing speed is desired.

For this research, a Xilinx Virtex-4 FPGA will be used to measure signal pulse width at 300-MHz. In section 3.6, a novel method that I have developed will be presented which enables timing the signal at 1.2GHz. As discussed in section 4.2.1, a counting frequency of 1.2GHz at 180-rpm translates to an uncertainty of +/-7.8 nano-radians.

3.1.2 Voltage Precision Requirement

Throughout the research information, such as rotation angle, motor current, or position, is communicated as voltage signals. The electronics need to be able to receive and interpret the signals with minimum noise. Within the scope of this research, the most sensitive signal is the encoder signal. As discussed in section 4.2.2, for achieving the target +/-10 nano-radian accuracy target, at least a 3500:1 signal to noise ratio is required on the

encoder. The electronics will also be monitoring the very high quality capacitive probe signals, which are expected to have a signal-to-noise-ratio of about 9500:1.

For this design, a signal-to-noise-ratio of at least 10000:1 on the electronics side is targeted which translates to 13 to 14 effective bits binary representation. Precision buffers with low noise and thermal drift specifications are used to minimize the noise. In addition, 16-bit Analog-Digital-Converters and 16-bit Digital-Analog-Converters are utilized to ensure a 14 effective bits.

3.1.3 Data Acquisition Speed

The incorporated analog-digital-converters (ADC) allow monitoring analog signals digitally and the digital-analog-converters (DAC) enable generating analog command signals digitally. The conversion speed of ADC and DAC need to be well above the bandwidth of the signal are being monitored or generated. This would prevent aliasing, provide a good picture of the signal, and reduce the delay on the signal. Within the scope of this research, the encoder signal and the current feedback are the two fastest changing signals. The encoder signal can reach 300-kHz frequency and the current controller can be designed for a bandwidth of 100-kHz. For this design, ADCs with 100-MHz and DACs with 40-MHz sampling frequencies are selected. The high sampling frequency minimizes the time discretization effects.

3.1.4 Algorithm Execution Speed

The custom electronics are used as a controller as well. Although the fast ADC and DAC enable high speed data acquisition, the controller algorithm asserting the command based on the input signal needs to be processed very fast as well. Within the scope of this

research, the highest bandwidth system to be controlled is the digital current controller with 100-kHz bandwidth. To properly control such a high bandwidth system, the controller needs to have 1-MHz or faster execution frequency.

For this design, the Virtex-4 FPGA is selected. The FPGA's fabric and the embedded PowerPC core can work together to achieve faster control execution time. The FPGA fabric can execute relatively simple but frequent parts and processor core can run the more complex but slower processes. For simplified control algorithm, this configuration is expected to achieve execution speeds above 1-MHz.

3.1.5 Communication Data Rate

Encoder calibration generates large amount of data very quickly. This data needs to be transferred to a computer to be logged and further analyzed. Also, if the electronics are to be used as a data acquisition card, issuing commands to the high speed ADCs and DACs will involve large sums of data. Finally, the graphical user interface will be developed on a computer and will connect to the board through the communication standard. A fast communication standard is required to reduce data transfer times and minimize the delays in communications between the PC and the custom electronics.

For this design, two of the fast available communication standards are included: Gigabit Ethernet transferring data at 1Gb/s and High-Speed USB transferring data at 480Mb/s. the choice of two standards adds flexibility to the communication method.

3.2 Architecture Design

The overall layout of the custom electronics is shown in Figure 3-2. The motherboard is the central part of the custom electronics. The motherboard converts signals between analog and digital domains, processes the signals, and communicates with a graphical user interface (GUI) running on a host computer. The Daughterboard is application specific and functions as a bridge between the motherboard and the parts connecting to the motherboard. The daughterboard buffers and adjust the input and output signals to levels compatible with the motherboard. It can also perform application specific operations on signals; as an example, in this design it digitizes the analog encoder signals, so they can be interpreted by the digital circuitry on the motherboard. The following subsections discuss the architecture design of each of the boards used within the overall design.

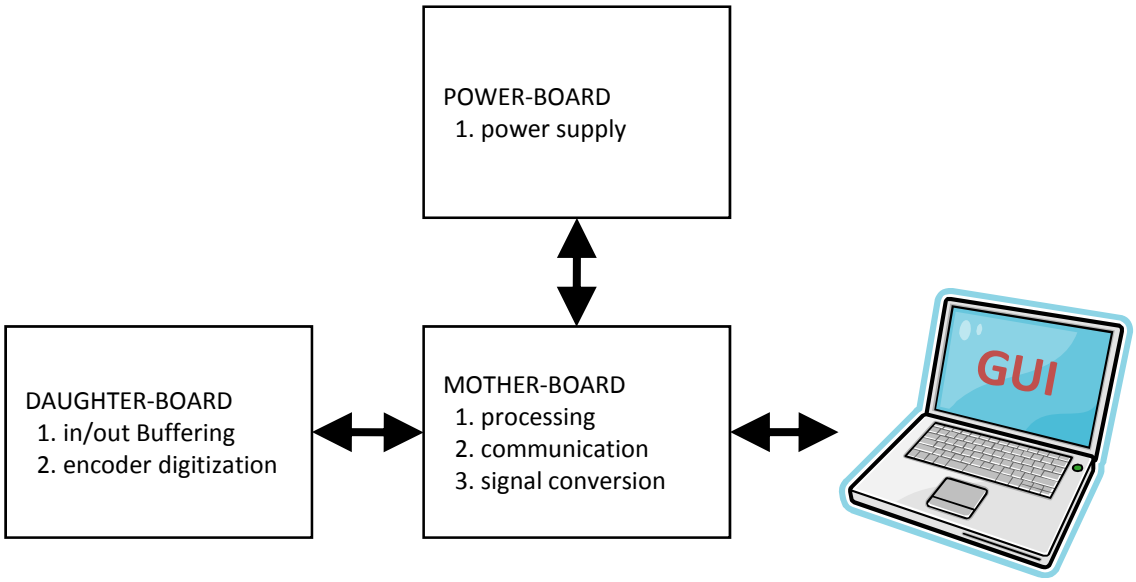

Figure 3-2: The overall layout of custom electronics

3.2.1 Motherboard Architecture Design

The motherboard's architecture design is shown in Figure 3-3. The board is designed around the Virtex-4 FPGA. The virtex-4 FPGA includes programmable and other hard wired dedicated components such as the PowerPC405 (PPC405) processor and the Media Access Controller (MAC) for gigabit ethernet interface. The PPC405 processor can perform the more complex functions such as executing the communication protocol; while, the FPGA fabric can run relatively simple but more frequent parts of the algorithm. The hybrid mix of processor and FPGA fabric allows for very efficient control algorithm implementation. The Media Access Controller is used to communicate with physical layer device for the Ethernet link. The DAC, ADC, and digital inputs/outputs interface with the FPGA fabric. The FPGA fabric is used to design a control engine which obtains, processes, and asserts signals. The FPGA is a configurable device. The configuration memory storing the design is volatile and will be lost if power is disconnected. The platform platform flash and the system ACE are used to permanently store configuration data and to configure the FPGA on power up.

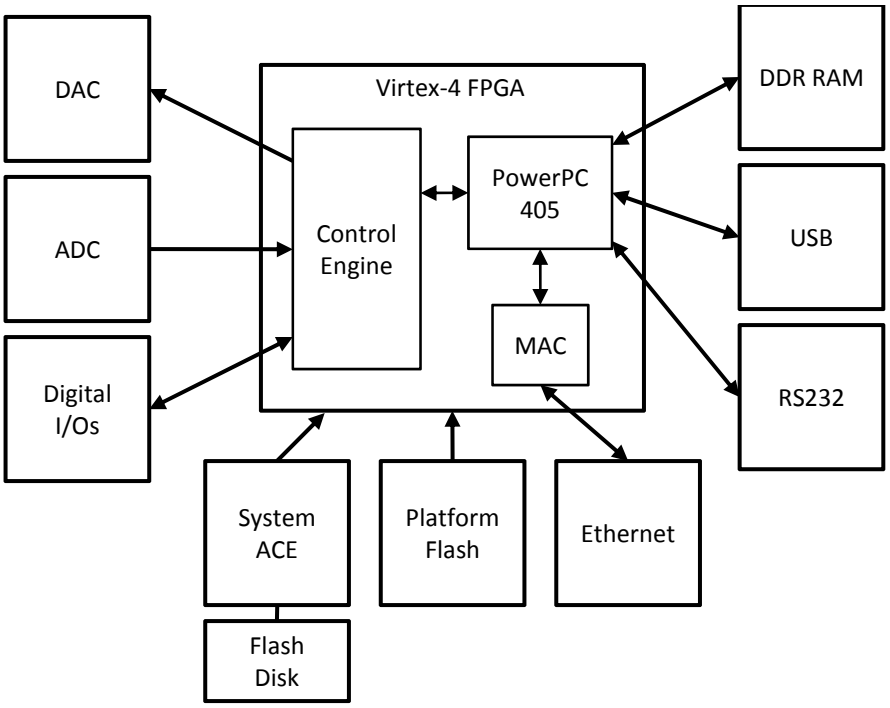

Figure 3-3: Motherboard architecture design

3.2.2 Daughterboard Architecture Design

The daughterboard's architecture design is shown in the Figure 3-4. The daughter board is designed to accommodate for several applications: encoder calibration, digital controller for power amplifiers, rotary position control, and general data acquisition. The daughterboard interfaces the motherboard with the sensors as well as either our in-home developed or commercial power amplifiers. The ADC and DAC switches select which signal to be connected to the ADC and the DAC. Depending on the application each ADC can be monitoring cap probes, auxiliary input, current feedback, or encoder signals; and, each DAC can generate a voltage command, current command, or a reference level to which the encoder signal will be compared to. For encoder calibration, the universal buffers allow averaging and gain adjustment of any specific combination of encoder

signals. The comparator engine consists of high speed comparators and digitizes the buffered encoder signals passed to it.

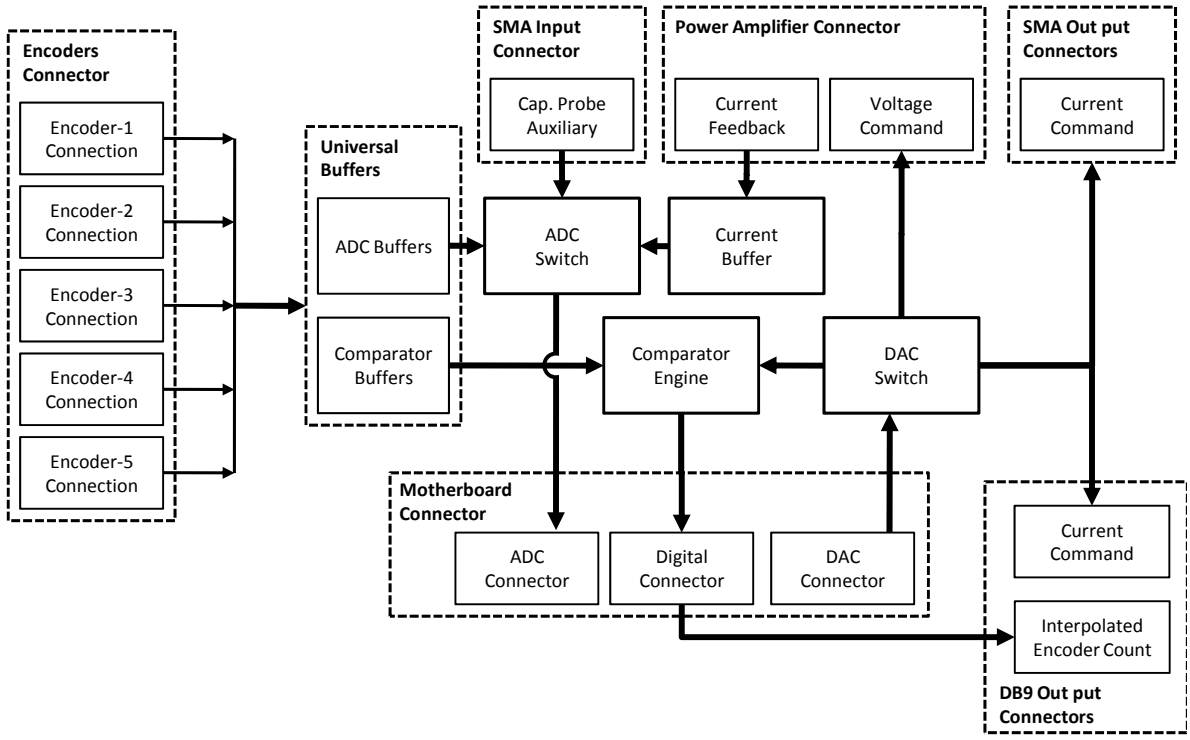

Figure 3-4: Daughterboard architecture design

3.2.3 Powerboard Architecture Design

The power board architecture design is shown in Figure 3-5. The power board supplies the motherboard with its required voltage levels. The analog sections require ±15V, ±5V, and 3.3V levels, while the digital sections require 1.2V, 1.8V, 2.5V, 3.3V, and 5V. Due to the resistive drops of the transmission cable, the voltage levels generated by the powerboard can be different from the levels seen by the motherboard. To eliminate this uncertainty, a separate and load-free feedback line is used for each voltage level, except for ±15V, to monitor the power levels on the motherboard side and to actively compensate for any voltage drops. All the levels for the analog components of the motherboard are generated using linear regulators to ensure low noise. For the voltage

levels used by the digital components of the motherboard, switching regulators are used where large currents are required, and linear regulators are used elsewhere. On the digital power side, to ensure appropriate power up of digital electronics, the switching regulators are synchronized so that their voltage levels follow a common rising ramp. The linear regulators connected to the 5-V supply follow the 5-V level up to their target level.

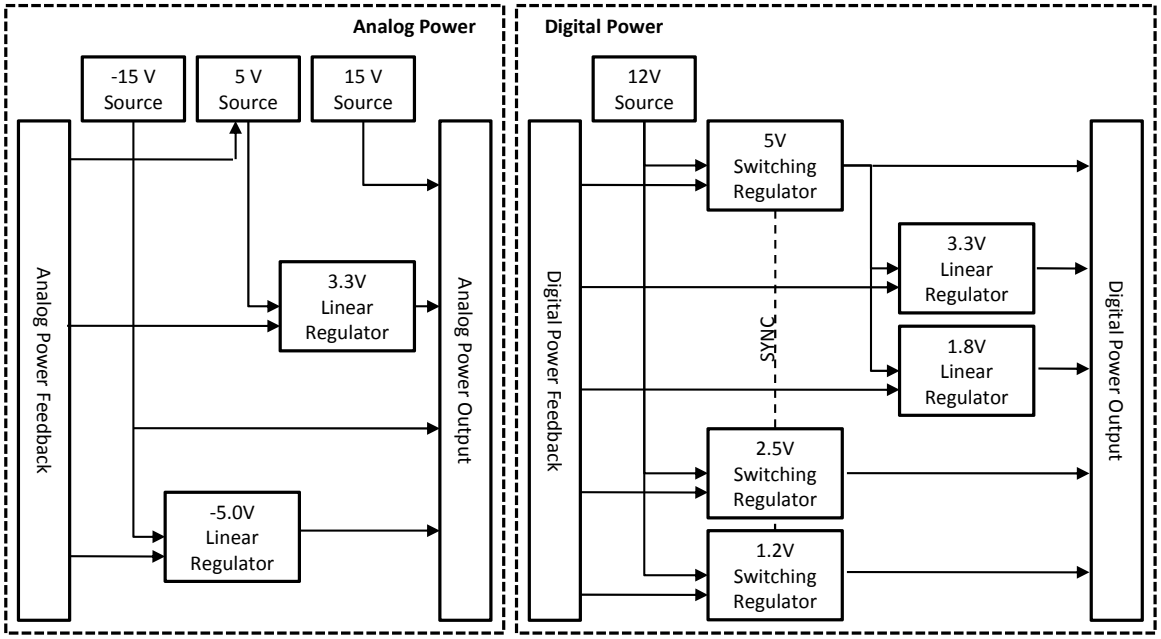

Figure 3-5: Powerboard architecture design

3.3 Motherboard Detailed Design

The manufactured motherboard is shown in Figure 3-6. We named our motherboard as *Avalanche*. It is designed with the Xilinx Virtex-4 FPGA and uses a hybrid mix of the FPGA fabric and the embedded PowerPC405 processor to achieve very high data acquisition and control execution speeds. The processor can carry out complex and less time critical operations while the FPGA can concurrently and very fast carry out relatively simpler and more time critical tasks. By distributing the tasks between the FPGA and the processor, the Avalanche board maximizes speed, minimizes delay, and facilitates implementation of control algorithm.

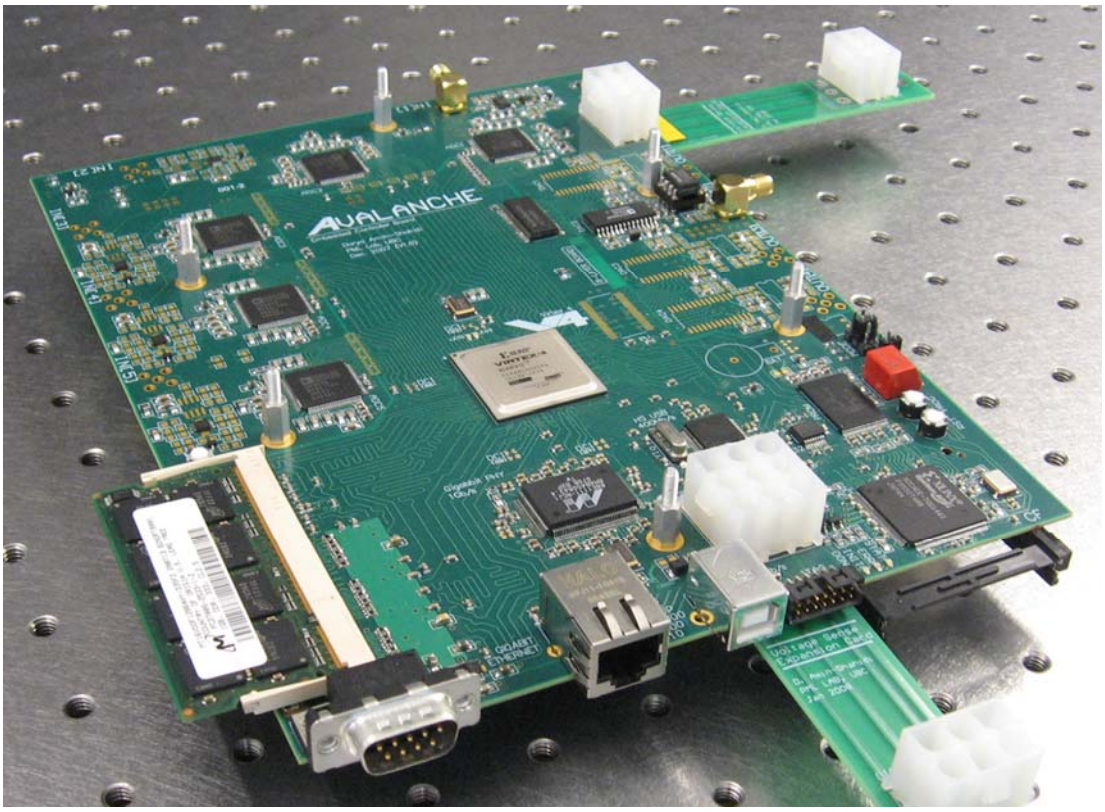

Figure 3-6: The *Avalanche* motherboard

Similar to any other processing machine, the motherboard requires access to a fast accessible memory space. The memory space available inside the Virtex-4 chip is very limited. In our design, a 512-MB DDR RAM memory module with 32-bit data width and a maximum of 333MHz transaction rate is interfaced to the FPGA. Aside from the volatile but very fast RAM memory, the board incorporates nonvolatile memory devices which are used to configure the motherboard upon power-up or user request. The board uses a Xilinx Platform flash device which can configure the hardware and software internal to the FPGA, and a Xilinx ACE System device which can configure hardware and software internal to the FPGA as well as any software loaded to the RAM.

The Avalanche board utilizes state of the art Analog-to-Digital-Converter (ADC) and Digital-to-Analog-Converter (DAC) technology from the Analog Devices to ensure that the board can precisely operate on very fast changing analog signals. The ADCs enables the board to capture an analog signal with 16-bit resolution at up to 100-MHz sampling frequency; while the DACs enable the board to generate an analog signal with 16-bit resolution at maximum 40-MHz sampling frequency.

The Avalanche board incorporates Gigabit Ethernet and High-Speed Universal Serial Bus (HS-USB) to connect to any computer. The Gigabit Ethernet and HS-USB communication media enable data transfer rates up to 1-Gb/s and 480-Mb/s respectively. To add further flexibility an RS232 serial link is added as well. The board also includes 35 digital input or output configurable lines for custom connection to another board.

The following subsections cover the selection and the detailed design of each of the components of the Avalanche board.

3.3.1 Virtex4-fx12 FPGA Component

The Avalanche board is designed to include a fast FPGA which can be configured to efficiently perform a specific set of processes concurrently and very fast. The board also requires having a processor to perform more complex threads. A single FPGA with embedded processor is the most suitable choice for the design because it significantly simplifies the design. At the design stage, the Virtex4-FX12 FPGA from Xilinx was the fastest mature FPGA technology with hardcore processor. The Virtex4 FPGA fabric can support clock rates as high as 500-MHz; however, a clock rate of slightly above 300-MHz would enable the use of all the FPGA resources in different configurations.

Table 3-1: Virtex-4 FX12 FPGA main resources

Embedded Hard Resources	PowerPC Block	1
	Ethernet MAC	2
	Clock Manager	4
Memory Resources	Block RAM	86 kb
I/O resources	Max I/O pins	320
	I/O banks	9
	Global Clock Net I/O	16
	Impedance Control	Yes

The main resources of the Virtex-4 FX12 device are listed in Table 3-1 which is mainly based on information from [20]. The Virtex4 FPGA consists of configurable FPGA fabric and non-configurable blocks. The configurable fabric can be transformed and configured using software to perform a specific task. The flexibility to configure the FPGA fabric requires overhead features at the silicon level and comes at the cost of lower speed performance and larger space requirements. Hard resources are hard wired fixed blocks that are intended for specific functions. The hard blocks can have higher density and can perform faster. Examples of the embedded hard blocks in the Virtex4 device

include the PowerPC405 (PPC405) processor, the Ethernet Media Access Controller (EMAC), and the digital clock managers. The PPC405 processor is used as the Avalanche board's processor, the EMAC is used to establish a Gigabit Ethernet link, and the digital clock manager is used to modify clock signals frequency and phase and to improve the clock signal quality [21].

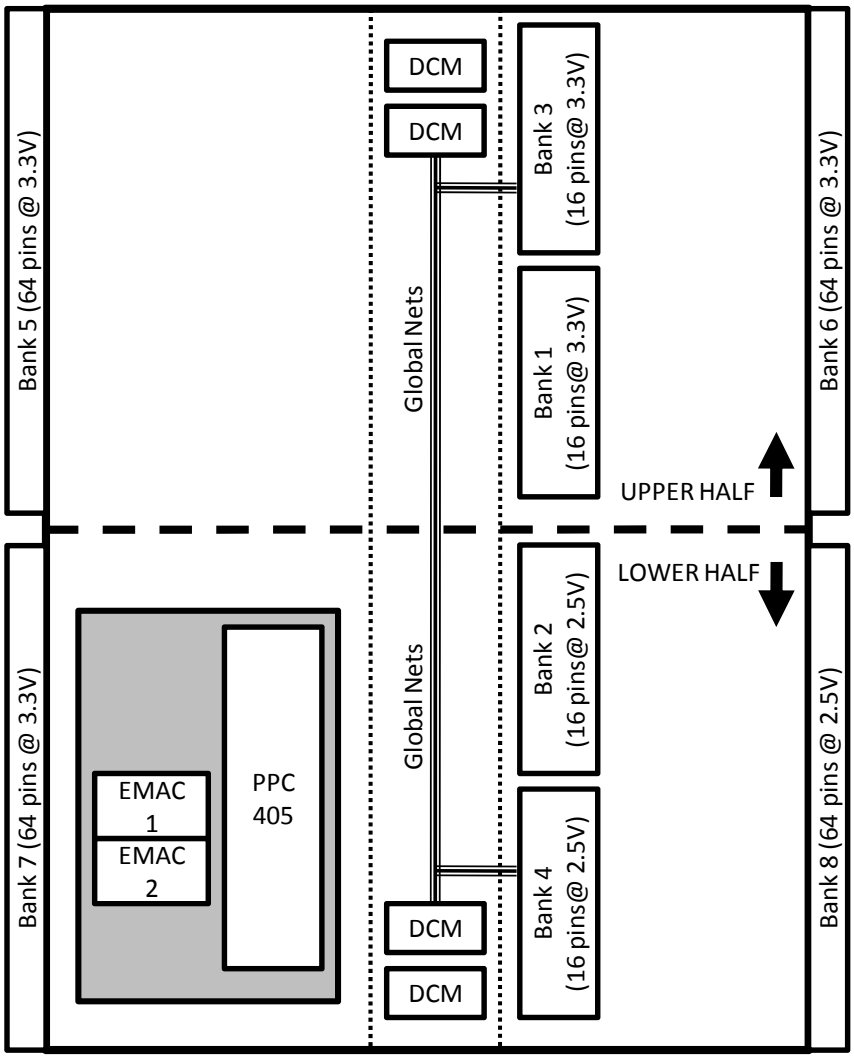

Figure 3-7: Simplified representation of the Virtex4-fx12 chip's architecture

A simplified schematic of the Virtex4 device, in Figure 3-7, shows the layout of the FPGA resources. The FPGA consists of an upper half and a lower half. The FPGA

can be interfaced to the outside through four 64-pin and four 16-pin input/output (I/O) banks. The FPGA can be configured through a separate 16-pin bank, not shown in the Figure 3-7. The device architecture shows the hard-wired PPC405 and the two EMACs as well as four DCMs, two in each half of the device. The DCMs are connected to I/O banks 3 and 4, which are called global clock banks, through the global clock nets.

The eight I/O banks have a total number of 320 pins which may be configured to interface to other digital devices. Each electronic device interfaces with others based on a certain digital standard with specific low and high voltage levels, such as Low Voltage TTL and 2.5-V CMOS. The Virtex4 device supports a wide range of I/O standards that its pins can be configured to. This flexibility enables the designer to use the Virtx4 device with almost any digital device. Although the I/O standards for each pin can be set individually, the driving voltage for all the pins within a single I/O bank must be the same. This introduces a limit on the assignment of the peripheral device I/O pins to the Virtex4 I/O pins. For this design the banks 2, 4, and 8 use a 2.5V driving voltage and the rest of the banks use 3.3V driving voltage.

Included in the 320 I/O pins, the digitally controlled impedance (DCI) pins and the voltage reference pins (V_{ref}) are dual purpose pins which may not be available for use as I/O lines. The Virtex4-fx12 device allows software selectable impedance control of the I/O drivers in all banks except for banks 1 and 2. There are two digitally controlled impedance (DCI) pins in each I/O bank which can be optionally used to specify the termination impedance of I/O drivers within their specific I/O bank. If digital impedance control is not used in a bank, the DCI pins can be used as regular I/Os. Some I/O standards require a high to low transition reference level to operate, such as the SSTL

input standard. To accommodate these I/O standards, the I/O banks have one voltage reference (V_{ref}) pin per 16 I/O pins. If the standards used within a bank do not require a reference voltage, the V_{ref} pins in that bank can be used as regular I/O lines.

In most cases, digital communication or processing needs to be synchronized by a clock signal. Because the timings rely on the clock signal, it must have well defined sharp signal transitions. The Virtex4 device has specific resources for distributing and conditioning the clock signals. In the Virtex4–fx12 device, there are two 16-pin global clock banks, one in each half of the device. The 32 global clock pins can be used as 16 differential clock pairs, or as 16 single-ended clock pins plus 16 general purpose I/Os. The global clock pins connect to the 16 global clock nets which can distribute the clock signals everywhere on the device with matched delay. It is ideal to connect clock signals to the global clock pins because only the global clock nets have direct connection to the main clock resources, such as the digital clock managers. There are four digital clock managers inside the Virtex4-fx12 device, two in each half, which can be used to modify the phase and the frequency of the clock signals. Aside from the global clock nets, the Virtex4 device has local clock capable pins with lower parasitic capacitance. These clock pins are connected to the local clock nets. If global clock pins are not available, the local clock capable pins should be used for clock signals and other signals with critical timing.

The low capacitance global and local clock pins on the Virtex4 device are suitable for connecting to clock signals. Ideally, the clock signals and their related signals should be kept as close as possible to improve timing and to reduce the design's footprint on the FPGA fabric. As shown in Figure 3-7, the Virtex4 device consists of an upper half and a lower half. Although the Virtex4 devices have the resources to compensate for clock

delays, crossing from one half of the device to the other creates a delay on the signal which cannot be compensated. As a result, the designer must keep the critical clock signals and all their related signals in a single half of the device. This introduces another limit on the FPGA I/O pin assignment.

Considering the above guidelines, the Virtex4-fx12 pins were carefully assigned to the pins on the interfacing parts. The pin assignment criteria are to follow the compatibility rules, achieve the best performance, and maximize the number of usable I/Os. The current design appropriately interfaces all the components and utilizes all the remaining 32 pins as 3.3-V general purpose I/Os. The summary of pin allocation is shown in Table 3-2. The different devices and their interfaces' operating voltage level are shown in the two left most columns.

Table 3-2: Virtex4-fx12 pin assignment summary

		FPGA Banks								
		0	1	2	3	4	5	6	7	8
Device	LEVEL (V)	3.3	3.3	2.5	3.3	2.5	3.3	3.3	3.3	2.5
ADC1	3.3									
ADC2	3.3									
ADC3	3.3									
ADC4	3.3									
ADC5	3.3									
DAC 1n2	3.3									
DAC 3n4	3.3									
DDR	2.5									
USB	3.3									
ETHERNET	2.5 (3.3 tolerant)									
General I/Os	3.3									
RS232	3.3									
System ACE	3.3									
Platform Platform flash	3.3									

The two top rows show the FPGA bank numbers and their set operating level. The devices and the banks create a 2-dimenssional matrix. The filled boxes represent a connection between the device on that row and the bank on that column. The operating level of the each device has to match the operating level of the banks that it connects to. This is the case for all the devices except of the Ethernet. The Ethernet device can tolerate 3.3V on most of its inputs but can only generate 2.5V on its outputs. In the design, some of the Ethernet device's input lines are connected to the Virtex4 device's 3.3V outputs. However, all Ethernet outputs connect to the 2.5V Virtex4 inputs only. Although not applied in this design, later it was noticed that the 2.5V Ethernet outputs were compatible with Virtex-4 inputs set to the 3.3V LVTTL standard.

The Virtex4 has many internal resources and can drive up to 320 I/O lines at different configurable levels. As a result it requires power to be supplied to it at several different levels. The Virtex4 device uses $V_{CCINT} = 1.2V$ and $V_{CCAUX} = 2.5V$ for its internal circuits. It also requires power at the selected I/O standards' voltages. In this deign I/O standards with only 3.3V and 2.5V driving voltages are used, so the virtex4 chip requires $V_{CCOi} = 2.5V$ for banks 2, 4, and 8 and $V_{CCOi} = 3.3V$ for banks 1, 3, 5, 6, and 7, where letter *i* should be replaced by the specific bank number. Virtex4 is a high-speed digital device and has fast changing electrical current demands which the power supply cannot directly respond to quick enough. Also, the device will work with high-speed signals which must have a low impedance return path. Decoupling capacitors at the Virtex4 side must be used to create a low impedance power source at all frequencies and a short length and low impedance current return path for the high frequency signals.

As shown in Figure 3-8, any capacitor has non-ideal inductance and resistance in series with their actual capacitive part. The equivalent impedance, Z, of a non-ideal capacitor can be modeled as in equation 3.1. The magnitude plot of a capacitor's effective impedance is plotted in Figure 3-8 according to equation 3.1[22]**Error! Reference source not found.**. The decoupling capacitor is only effective where it has low impedance. As a result, any capacitor is only effective over a certain range of frequencies. The current path inside a smaller capacitor creates a lower loop area, so a smaller capacitor will have lower series inductance. As a result going smaller in the capacitor size, the impedance curve shifts to the right, and the capacitor's effective region moves to higher frequencies. Using a range of capacitor sizes in parallel enables a power supply decoupling design which will be effective a wide operating frequency range.

$$Z = Z_L + Z_R + Z_C = Ls + R + \frac{1}{Cs} \tag{3.1}$$

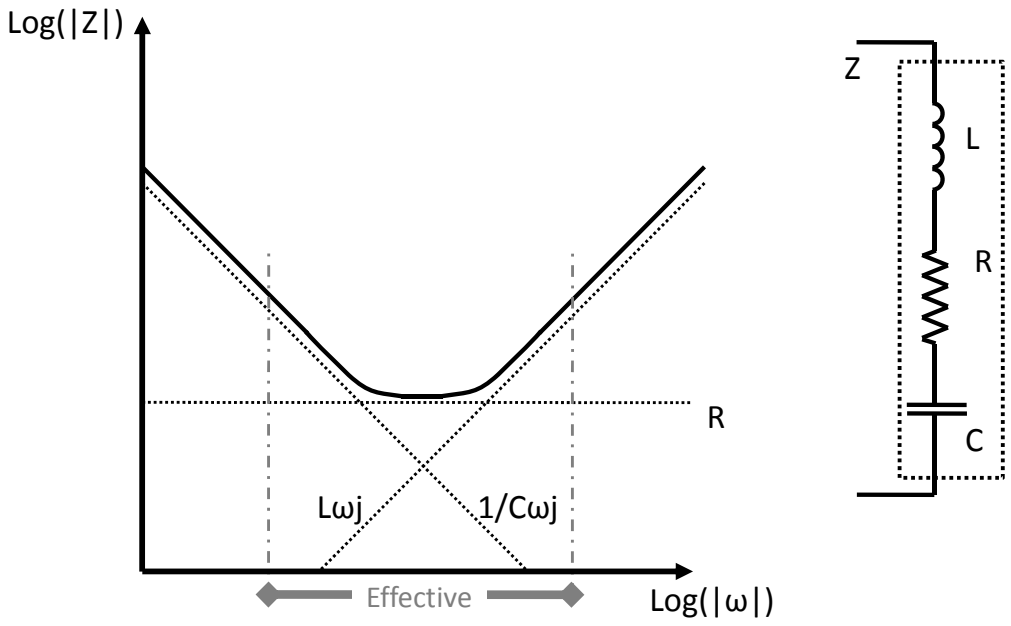

Figure 3-8: Impedance model of a non-ideal capacitor

Xilinx [22] recommends a method for selecting the appropriate number and size of the decoupling capacitors for the Virtex4 device. For each power supply level, Xilinx suggests using as many capacitors as there are FPGA power input pins at that level. The size and the proportional quantity of each capacitor size suggested by Xilinx is summarized in

Table 3-3. Although the voltage reference (VR) pins are not power supply pins, they require capacitors for noise filtering. One pair of 0.1-µF and 0.01-µF capacitor on each VR pin is recommended by Xilinx.

Table 3-3: Supply decoupling capacitor selection rule [22]

Value (µF)	Type	Maximum Size	Percentage
470	Tantalum	N/A	4%
1	Ceramic X7R	0805	14%
0.1	Ceramic X7R	0603	27%
0.01	Ceramic X7R	0402	55%

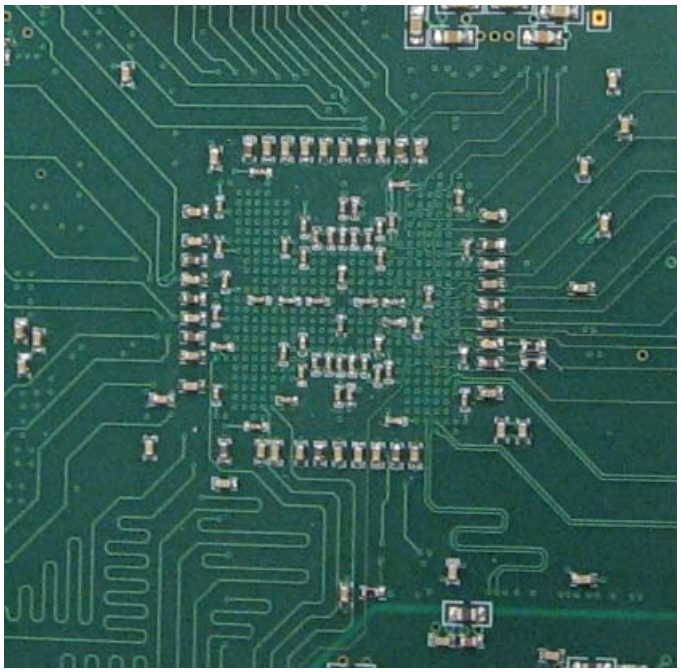

Figure 3-9: Picture of the FPGA decoupling capacitors soldered directly on the back of the FPGA chip. For scale, the spacing between the dots is 1mm.

3.3.2 FPGA Configuration Loading System

A significant portion of the Virtex4 FPGA chip is configurable and can be configured according to the design requirements. For this project, Xilinx's Embedded Development Kit (EDK) and Integrated Software Environment (ISE) software are used together to translate the functional requirements to the internal software and hardware configuration design of the FPGA. The ISE can be used to configure the FPGA fabric to perform a specific task. The ISE environment allows the user to describe the design using VHDL or Verilog language. The EDK software builds on top of the ISE and provides a higher level interface to create a design using certain embedded FPGA hardware modules. The EDK has a large library of already designed hardware modules, called hardware IPs, which can be used to perform a certain function. Examples of hardware IPs are PPC405 processor and its communication buses, DDR memory controller, and Ethernet controller. The EDK interface allows the user to add the hardware IPs to a FPGA hardware design, tailor them for the specific FPGA design, and connect them together. The EDK interface also enables the user to add software drivers for the hardware IPs and to develop software using the C language to run on the processor in the design.

A FPGA design can be developed as a combination of EDK and ISE designs. Any ISE design can be used in the EDK software, and any EDK design can be added as a subsystem inside an ISE project. The ISE and EDK programs translate a higher level design description into a configuration data file which will be used to configure the FPGA. The FPGA configuration pins are used to load the design into the FPGA configuration memory. There are several configuration interfaces available. The mode

pins on the FPGA determine which configuration interface type is used. The different modes are summarized in Table 3-4.

Table 3-4: Virtex4 configuration interfaces [21]

Configuration Mode	M1	M2	M3	Data Width	CCLK Source
Master Serial	0	0	0	1	FPGA
Slave Serial	1	1	1	1	External
Master SelectMAP	0	1	1	8	FPGA
Slave SelectMAP8	1	1	0	8	External
Slave SelectMAP32	0	0	1	32	External
JTAG Only	1	0	1	1	-

In the Master serial mode the data is serially loaded through the D0 FPGA pin and is synchronized by a configuration clock (CCLK) generated by the FPGA. In the slave serial mode the CCLK is generated by an external clock source. The Master SelectMAP interface loads the data through a byte-wide parallel configuration data bus which is synchronized by a FPGA generated CCLK. In the Slave SelectMAP8 mode the CCLK is generated by an external source. SlaveSelecMAP32 uses a 32-bit wide data bus to configure the FPGA and is synchronized by an external clock source. The JTAG interface is an standard interface and is established through separate JTAG data and clock pins on the FPGA. The master configuration modes are simpler and more reliable because they do not require any external clock source. The SelectMAP modes load the configuration data faster; however, use more FPGA I/O pins and are more complex. The JTAG chain is the most versatile interface because it allows other operations aside from loading the FPGA configuration data; for example, the JTAG chain can be used to load, debug, and control any internal software component. The Master Serial and the JTAG interfaces are incorporated in the Avalanche design.

Because the FPGA configuration memory is volatile, the configuration data needs to be loaded every time on power-up. The avalanche design's JTAG interface enables the user to manually load the data using a JTAG programmer connected to a computer. The Avalanche board also includes two non-volatile memory devices which can save the configuration data and load it onto the FPGA upon power up or a user request. The Xilinx platform flash memory device onboard configures the FPGA through the Master Serial interface and the Xilinx system ACE memory device can configure the FPGA through the JTAG chain. The platform flash device was selected because of its simpler implementation and easier use. The System ACE device was selected because of its more versatile functionality and larger memory size. The schematic representation of the Avalanche configuration solution design is shown in Figure 3-10.

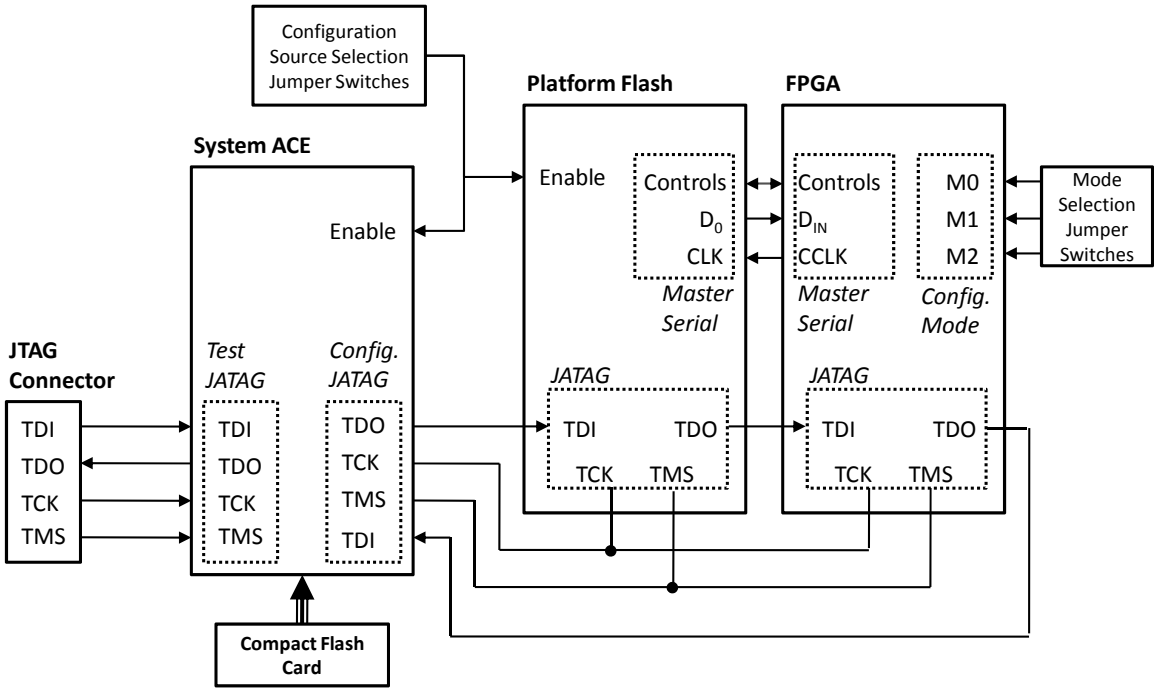

Figure 3-10: Schematic representation of *Avalanche*'s FPGA configuration solution

A JTAG chain is created by the ACE, platform flash, and the FPGA. This JTAG chain enables the user to command or configure any of the three devices manually by connecting a JTAG programmer to the JTAG Connector. The configuration or command data from the JTAG connector will be received by the ACE's Test JTAG interface. The ACE uses the part of the data labeled for the first link of the chain and passes the data to the platform flash. The platform flash chip uses the part of the data labeled for the second link of the chain and passes the data to the FPGA which uses the part of the data labeled for the third link. Using this method, the three devices can be controlled and programmed in a daisy chain. The memory for the Platform flash chip is programmed on the board using the JTAG interface. The ACE uses a Compact flash card (CF card) as its non-volatile memory source. The CF card's memory needs to be programmed using a compact flash drive off the board. Once programmed, both the ACE and the Platform flash can program the FPGA upon startup or a user initiated request. The jumper switches will be used to select which device will be enabled to program the FPGA; at the same time, jumpers are used to select the matching FPGA configuration mode. The Platform flash chip programs the FPGA over the master serial interface and the ACE chip can program the FPGA or the platform flash over the JTAG interface.

Both the platform flash chip and the ACE system are included in the Avalanche board as non-volatile configuration memory solutions. The platform flash is very simple, and hence, reliable configuration solution. Its non-volatile memory can be modified while it is installed on the board through the JTAG chain. However, the platform flash can load the FPGA configuration bit stream only and cannot configure any software external to the FPGA such as software to be loaded to the RAM memory module. The System ACE chip

loads the configuration bits through the JTAG chain. As a result, it can execute JTAG commands to load software which may be located on external memory modules as well. The ACE solution's large memory space is advantageous for large system designs. The ACE CF card in the Avalanche board can hold up to 512MB of configuration data. The main disadvantage of the ACE solution is that its memory cannot be written on the board. The CF card needs to be removed and mounted onto a CF drive.

Figure 3-11 shows the Avalanche board's JTAG chain as identified by iMPACT, the Xilinx ISE software's programming interface. As can be seen, the ACE, the PROM, and the FPGA chip create chain.

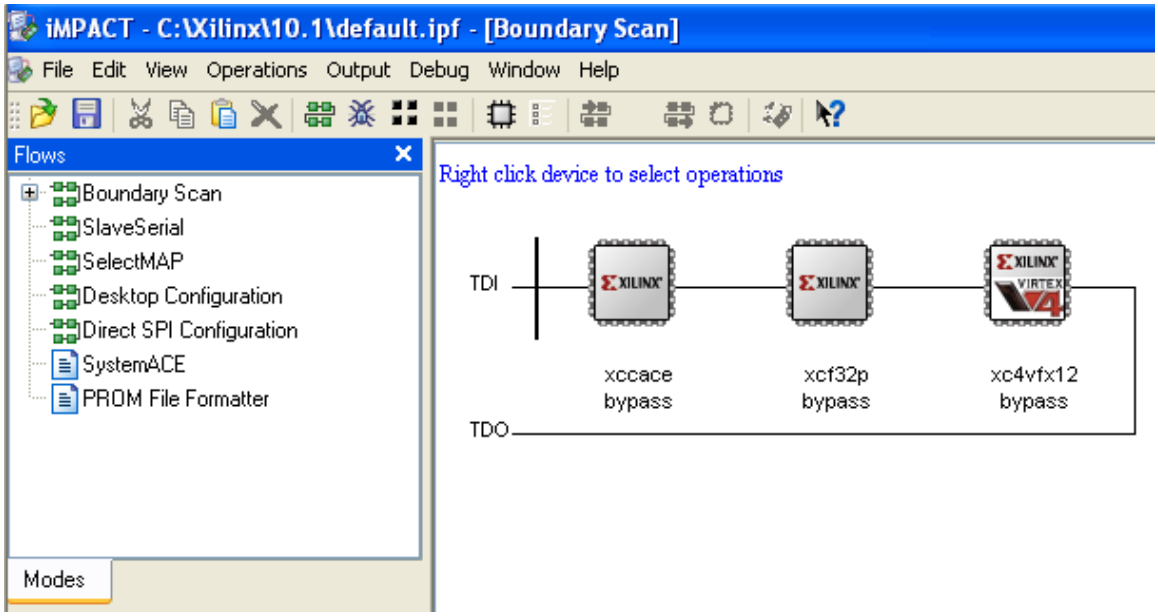

Figure 3-11: Avalanche board's JTAG chain detected by iMPACT, Xilinx ISE software's programming interface

3.3.3 DDR RAM Module

A Small Outline Dual Inline (SODIMM) Double Data Rate (DDR) Random Accessible Memory (RAM) module from Micron Technology is included in the Avalanche board design. The memory module provides 512-MB of fast random accessible memory space which can be used by the Virtex4 device. The Virtex4 can read or write in this memory space in any random order. The large and fast memory space provided by the RAM module is essential to the Avalanche board because it lifts the limits on the software code size as well as the software memory space size. A schematic representation of the RAM Module connection is shown in Figure 3-12.

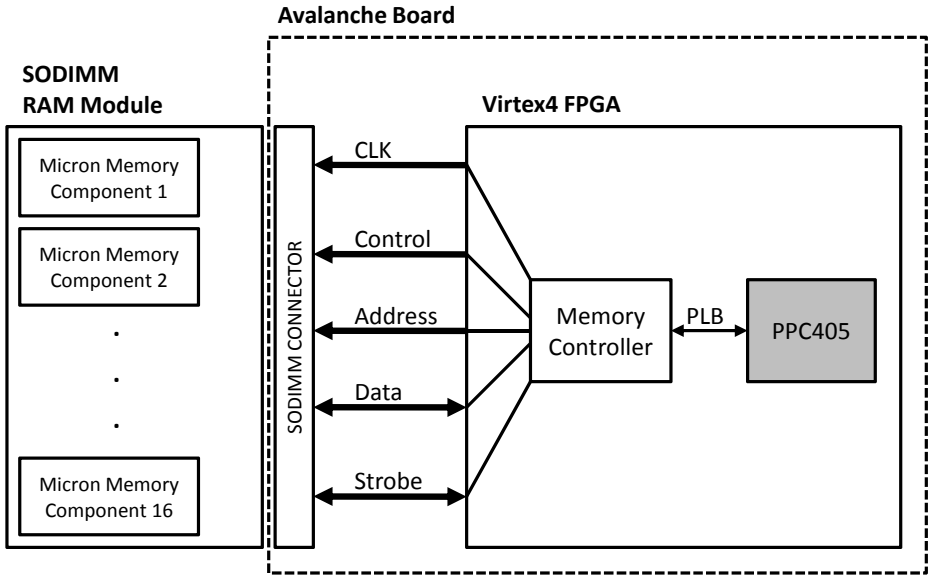

Figure 3-12: Schematic representation of the DDR connection

A memory module is a set of memory components which work together to create a larger memory space. The memory modules are produced in standard packages and are replaceable. Memory modules connect to a board through their specific connector and can be easily exchanged. As a result, instead of incorporating individual components onto

the Avalanche board, the board is designed to accept standard memory modules to add to the flexibility of the RAM memory solution. Double data rate (DDR) memory type is selected to achieve higher data rates. DDR memory devices read/write at both rising and falling edges of the clock; therefore, their throughput is twice the clock frequency.

The memory module to FPGA interface signals can be divided into five broad categories: clock, control, address, data, and strobe. The clock is generated by the FPGA and is used by the DDR internal logic. The clock signal ensures that both the FPGA and the DDR chip operate at the same speed. Control signals enable controlling the memory module's operation. The address lines are used to point to a certain memory location on the memory module, and the data lines are used to receive or send data to that specific location. The strobe lines are used to latch and capture the incoming and outgoing data at the right times. Inside the Virtex4 device, the memory module lines are interfaced to the PPC405 processor through a memory controller peripheral. The memory controller is built using the FPGA's internal programmable fabric. The memory controller connects to the PPC405 processor using the processor's local bus (PLB).

The memory module works at a clock frequency of 100-MHz or faster. Being a DDR device, the signal lines have an effective frequency of 200-MHz or faster which translates to a 5-ns or smaller signal period. At such high speeds, line impedance, line delays, and signal reflections become very significant. Any electrical signal travels at a speed slower than but close to the speed of light. As a result, a signal takes more than 1-ns to travel a 30-cm distance. In the same fashion, variations in the length of the signal lines will result in phase shift between the signals. Significant phase shift can result in improper latching of the signals and will prevent the memory device from functioning

properly. Therefore, in the Avalanche board's design, all the DDR memory signals' lengths are matched to within ±2-mm.

Any electrical signal travels from a source to a receiver. The transmitted signal will reflect back and forth between the receiver and the source. The reflections attenuate every time, and are significant during the first few reflections. If the signal period is much longer than the signal travel time, then reflections can be ignored. With the DDR memory design, the signal period is only five times the signal travel time, so reflections cannot be ignored and to prevent reflections proper source and receiver terminations are required. The termination scheme applied in the Avalanche board's RAM interface is shown in Figure 3-13. A 50-Ω impedance environment is selected where a termination resistance of 50- Ω is created at the end of the line between the signal line and the power plane. The resistor values need to match the signal line's characteristic impedance. The design of PCB traces to exhibit 50-Ω characteristic impedance is discussed in section 3.3.9, *Printed Circuit Board Design*.

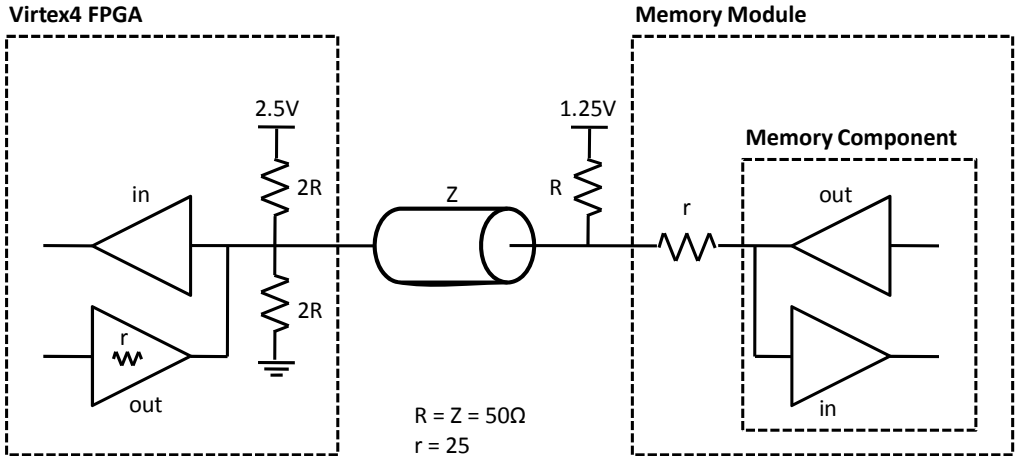

Figure 3-13: FPGA to DDR bi-directional signal termination design

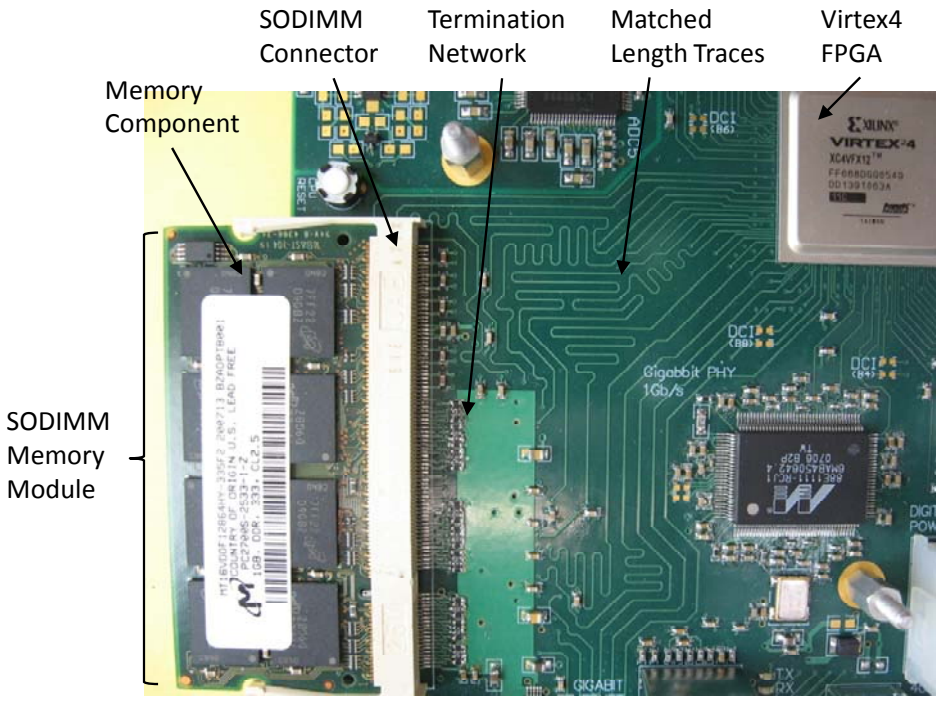

Figure 3-14: Avalanche board's DDR memory module solution

The physical implementation of the Avalanche board's DDR memory design is shown in Figure 3-14. The DDR memory module is connected to the board through the onboard SODIMM connecter. The module connects to the FPGA through the winding signal traces. The shorter signal traces are intentionally made longer to match the length of all the signal traces. Standard DDR SODIMM connectors have two opposing sets of 100 pins spaced with a 0.6-mm pitch. Although with most electronic components the opposing pins are aligned, the SODIMM connectors' opposing pins are offset by half of a pitch, or 0.3-mm. The footprint for the SODIMM connector had mistakenly been designed with aligned opposing pins. To fix this problem, we manually bent all of the connector's pins on one side by 0.3-mm to adapt them to the footprint.

3.3.4 Analog-to-Digital Converter

Aside from processing digital signals, the Avalanche board needs to be able to operate on analog signals. The Avalanche board incorporates five analog-to-digital converters (ADC), which enable it to simultaneously convert up to five analog signals to digital numbers. In order to get a close image of the analog signal, the ADC needs to have a high resolution both in time and voltage domains. The time resolution is dependent on the ADC's sampling frequency, and the voltage resolution depends on the number of effective bits used to represent the analog signal. For control applications, the ADC is required to convert with minimum delay because any delay from the ADC will reduce the achievable controller phase margin. Avalanche board's potential applications can include signals around 1-MHz frequency with signal to noise ratio of about 10,000. The ADC is required to sample at least 10-times faster than the signal. At such sampling rates, ADCs exhibit a pipeline delay which can take around ten clock cycles. As a result, the minimized delay becomes the limiting choice on the ADC's sampling frequency. To achieve tolerable delay, ADC sampling frequency of 100-MHz or higher is required. Also, a minimum 16-bit resolution is targeted to ensure that the ADC resolution is smaller or equal to the noise level.

There exists a trade-off between ADC speed and resolution. Faster ADCs have lower resolution, and higher resolution ADCs have slower sampling rate. AD9446-100 from Analog Devices was found to be the most suitable ADC technology for the Avalanche design. AD9446 achieves a balance between speed and resolution. It can sample a signal at 100-MHz with 16-bit resolution. AD9446 uses a pipeline structure to

speed up the conversion process; therefore, exhibits a 13-cycle delay. At 100-MHz, 13-cycle delay translates to a 130-ns delay.

Similar to most high performance ADCs, AD9446's analog input accepts a pair of differential signals. Differential signaling is used because symmetric design with differential inputs can eliminate even order harmonic distortions. However, most signals of interest are single-ended and need to be translated to the differential format before they can be converted by the ADC. Also, the high-speed ADCs' sample and hold circuit, which holds an analog level to be converted, requires to be connected to a very low impedance signal source. This can be guaranteed only if the signal is buffered onboard and close to the ADC. Furthermore, the ADC inputs are rated for a specific voltage range. Considering all the input requirements, the Avalanche board's ADC requires a buffering stage which accepts a single ended signal and converts it to low impedance differential signal within the ADC's acceptable input range. More specifically, the ADC buffer stage is designed to convert a ±10-V amplitude single ended input signal to a differential pair centered on 3.5-V and each with amplitude of 1.6-V.

Three methods, shown in Figure 3-15, were considered for converting a single ended signal to a differential pair of signals. Solution-A suggested by AD9446's datasheet uses a transformer. Solution-A cannot be used because it only works with alternating signals and will filter any low frequency component of input signals. Also, solution-A is a passive buffer and will not help with reducing the impedance of the signal. Solution-B uses a fully differential operational amplifier. Fully differential amplifiers are similar to regular amplifiers except that they generate two complementary outputs. The fully differential amplifier's a VCOM input can be used to set a common

level for the output differential pair. Solution-B is an active buffer and works for both direct and alternating signals. Solution-C uses two separate operational amplifiers to create a differential pair. Solution-C is not symmetric, and will cause mismatch between the differential pairs both in time and voltage domains. Solution-B is selected for the Avalanche board's design because of the following advantages:

1. Fully symmetric differential design eliminates 2-nd order harmonics
2. Active buffer will provide a low impedance output signal
3. Common mode of the differential pair can easily be set through VOCM

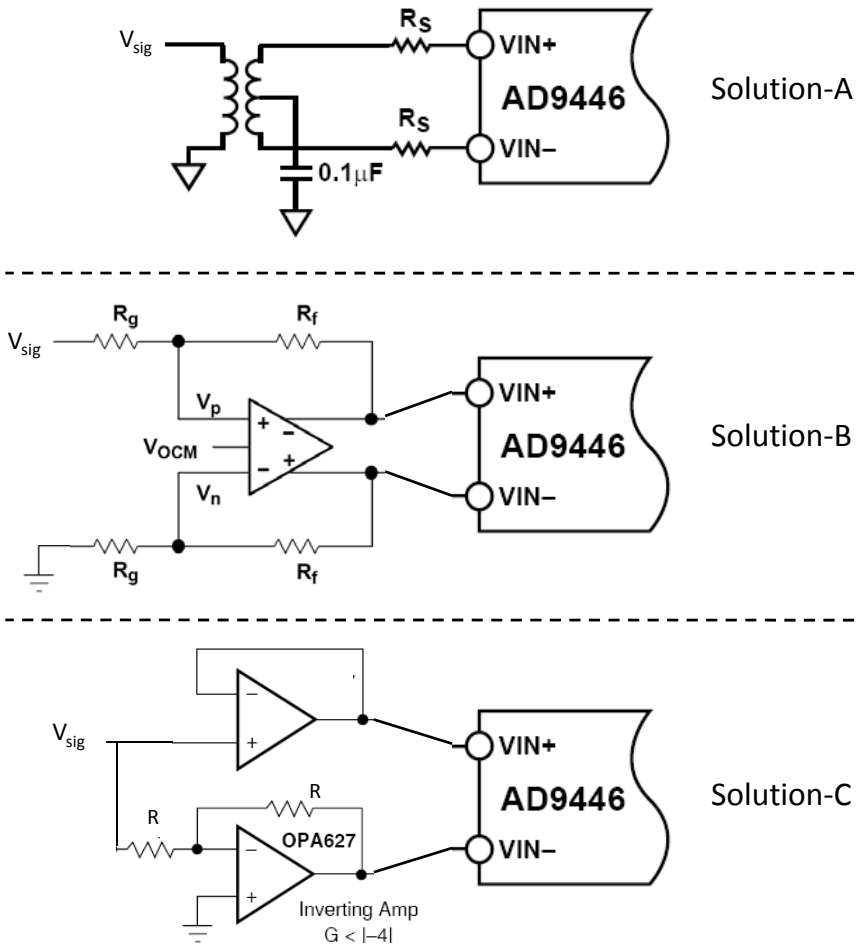

Figure 3-15: Methods for translating a single ended signal to a differential pair

For solution-B, it can be shown that the output signal V_{IN+} and V_{IN-} relate to the input signal V_{SIG} as in equation 3.2. It can be seen that the differential buffer gain is a direct function of the R_F/R_G ratio, and the output common mode voltage is defined by the level at the V_{OCM} terminal. Due to the amplifier's very large gain, its input terminals will be at the same level and can be considered to be virtually connected. The voltage at the amplifier's input terminals is termed as the input common voltage and can be related to the output voltage as in equation 3.3. High values of input common mode voltage will significantly reduce the fully differential amplifier's performance. The maximum value of the input common mode voltage depends on the ADC input level, V_{IN-} and the buffer gain, R_F/R_G. The range of the V_{IN-} signal is imposed by the ADC device and the buffer gain is the only design variable. A single-step down scaling of V_{SIG} to the ADC's acceptable level requires a very small R_F/R_G ratio, and such a small R_F/R_G ratio will result in the input common mode voltage being too high. To solve this problem, an additional down-scaling buffer is added to the ADC buffer stage. A schematic representation of the full ADC input buffer is shown in Figure 3-16.

$$V_{IN+} - V_{IN-} = -\frac{R_F}{R_G}V_{SIG} \;\longrightarrow\; \begin{cases} V_{IN-} = V_{OCM} + \dfrac{R_F}{2R_G}V_{SIG} \\ V_{IN+} = V_{OCM} - \dfrac{R_F}{2R_G}V_{SIG} \end{cases} \tag{3.3}$$

$$V_p = V_n = \left(1 + \frac{R_F}{R_G}\right)^{-1} V_{IN-} \tag{3.4}$$

The ADC buffer stage consists of a down-scaling stage using an operational amplifier and a single-to-differential conversion stage using a fully differential amplifier. Both stages use active devices. It is important to select amplifier devices which are stable

at the required closed loop gain, meet the speed requirements, and have high precision. Amplifiers' datasheets specify a rough estimate of closed loop gains for which the amplifier can be used stably. A quick estimate for a buffer bandwidth is to divide its amplifier's Gain Bandwidth Product by the closed loop gain of the buffer. However, these are only estimates of stability and bandwidth and to find closer values for bandwidth and stability, the buffer's feedback loop needs to be analyzed in the frequency domain. In most cases, the bandwidth information for an amplifier is based on small signal analysis. To ensure a fast enough response to all input signal amplitudes, the amplifier requires to have a slew rate limit above the maximum possible time derivative of output signals. A buffer's precision depends on the precision of the amplifier used. The amplifier needs to have low noise characteristics and low temperature sensitivity.

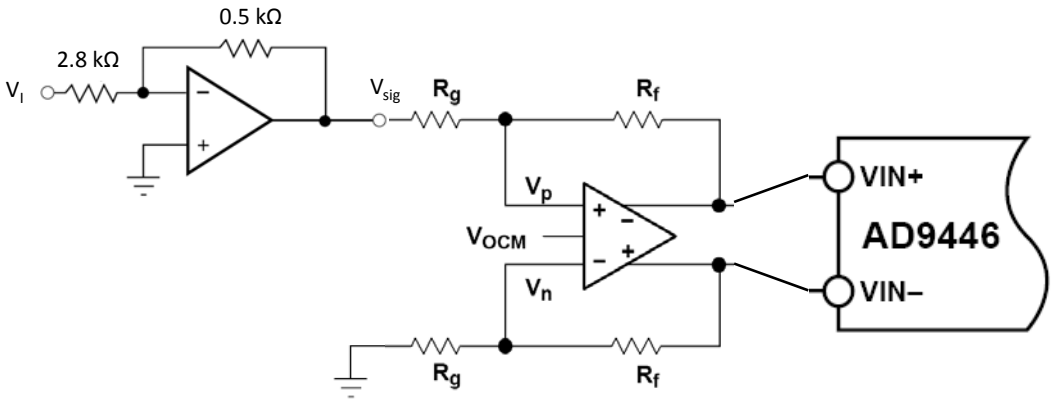

Figure 3-16: Schematic representation view of the ADC input buffer

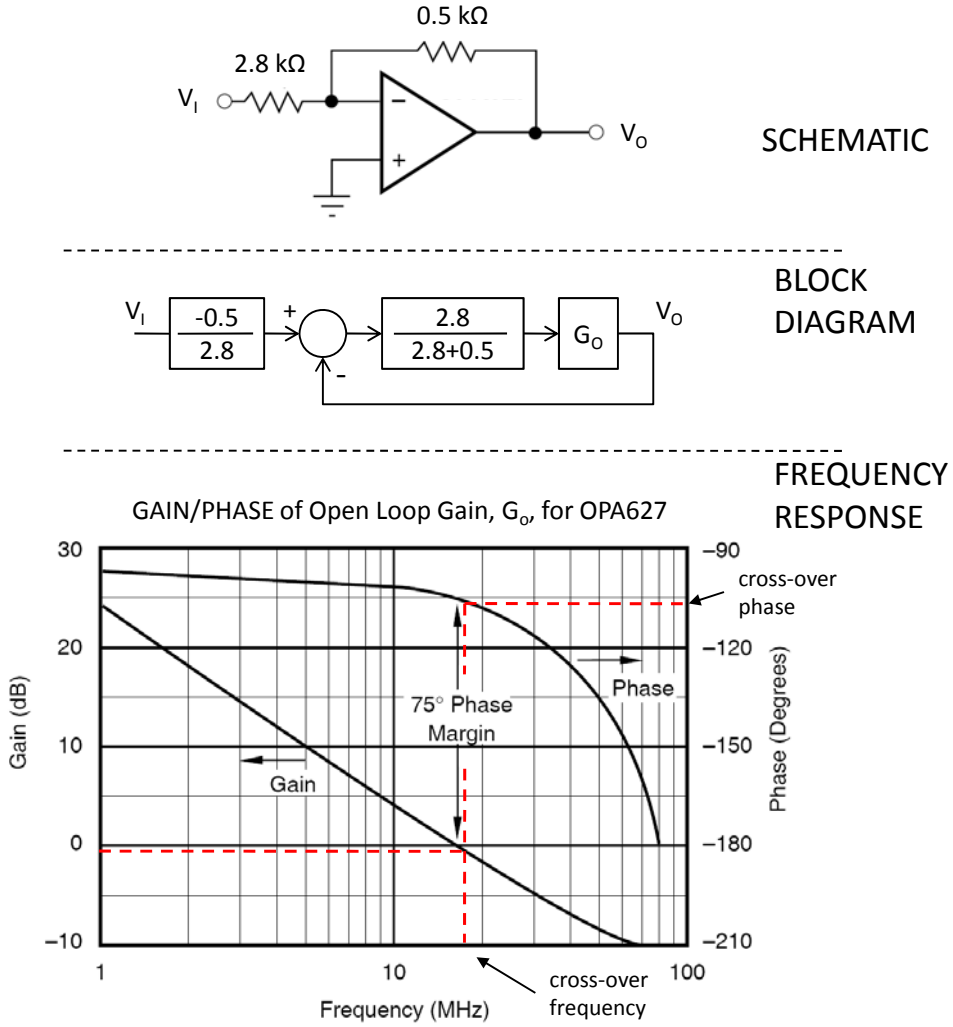

Figure 3-17: Inverting down-scaling buffer design and analysis

OPA627 operational amplifier is selected for the first down-scaling stage and THS4503 is selected for the single-to-differential conversion stage. Each buffering stage will be stable with a bandwidth of above 10-MHz and fast enough slew rate limit. As an example, the frequency response analysis of the OPA627 stage is considered in this paragraph. The buffer configuration and its equivalent block diagram model are shown in Figure 3-17. Defining the negative feedback loop transmission (NLT) as equation 3.4, the closed loop transfer function can be expressed as in equation 3.5. The 3-dB bandwidth frequency for the closed loop system is estimated as NLT's cross over frequency, and the

stability of the closed loop system is analyzed based on the Nyquist criterion. For this specific design, the NLT crosses 0-dB line at around 17-MHz; therefore, we expect a 17-MHz bandwidth. Also at the cross over frequency NLT's phase is about 105-deg, which leaves a 75-deg phase margin to 180-deg phase, and as a result, the buffer is expected to be stable.

$$NLT = \frac{2.8}{2.8 + 0.5} G_O \tag{3.4}$$

$$\frac{V_o}{V_I} = \frac{NLT}{NLT + 1} \tag{3.5}$$

3.3.5 Digital to Analog Converter

In order for the Avalanche board to be used as a controller, it needs to be able to generate analog commands. The Avalanche board uses four digital to analog converters (DAC). Each analog to digital converter can create an analog level based on a digital value from the Virtex4 device. The Avalanche board is designed to enable control frequencies as high as 10-MHz, so the DACs need to have a conversion rate of 10-MHz or faster. Also, higher resolution DACs are preferred because they provide a finer control over the generated analog levels. For the Avalanche board, AD768 device from Analog Devices is selected. AD768 is a 16-bit DAC and has a 40-MHz maximum conversion rate. The Avalanche board's output stage design is shown in Figure 3-18.

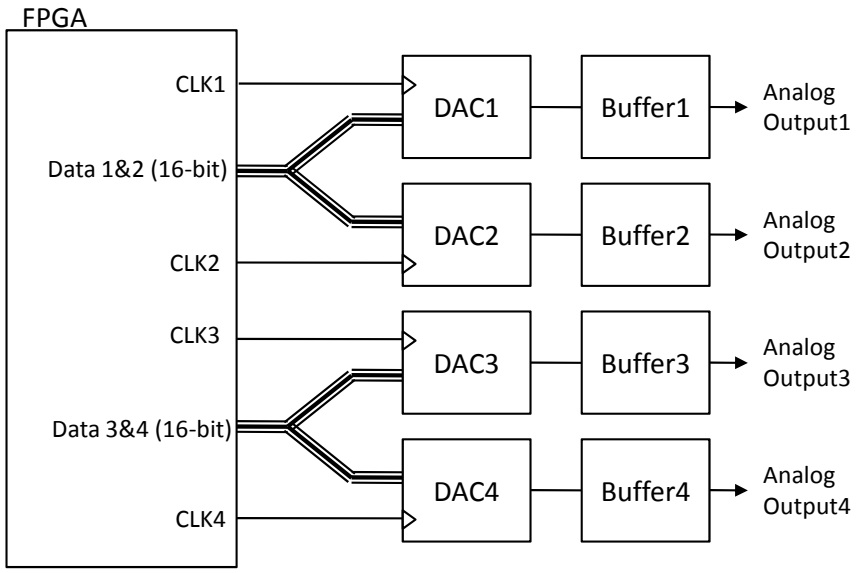

Figure 3-18: Avalanche's analog output stage

Four DACs are interfaced to the FPGA. Each DAC requires a conversion clock and a 16-bit data bus, which specifies the digital level to be converted. However, in the Avalanche board design each data bus is shared between a pair of DACs which have conversion clocks shifted by half a clock period, or 180 degrees. Sharing the data bus will save 32 FPGA digital I/O pins. The AD768 DACs output an analog current proportional to the 16-bit digital data. To practically use the analog output, the current output of each DAC needs to be converted to a voltage signal. A trans-impedance buffer is added on each DAC output which converts the DAC current to voltage signal within ±10V.

The DAC digital interface latches the 16-bit digital data at the rising edge of the conversion clock. The DAC requires that the data be setup on the data bus t_S=10ns before the clock rising edge and be held on the data bus t_H=5ns after the clock rising edge. One example of data and clock pattern for a DAC pair is shown in Figure 3-19. The data bus interlaces DAC1 and DAC2 data, and the CLK1 and CLK2 signals ensure that each DAC

latches to that data at its designated time. To guarantee the hold and setup time requirements for both DACs the CLK frequency has to be 33.3 MHz or slower. This is still above the targeted 10-MHZ conversion rate target. The Avalanche design enables using four DACs at 33-MHz or two DACs at 40-MHz conversion rate.

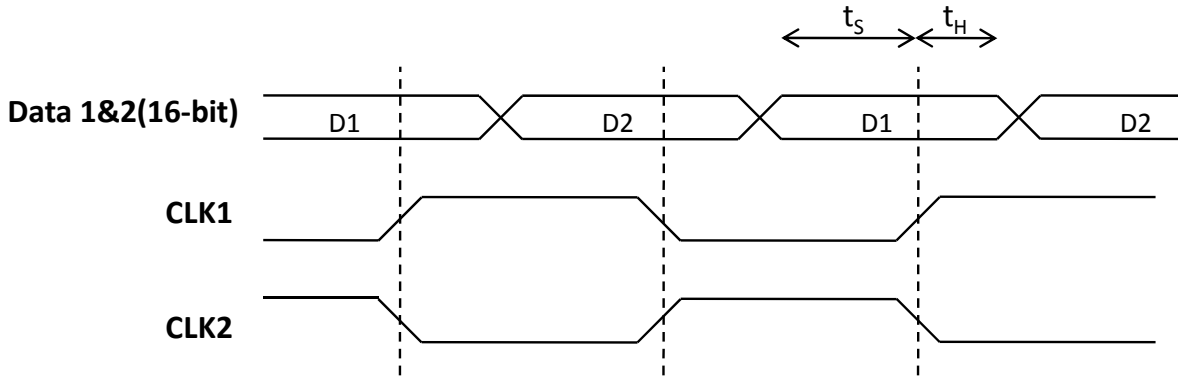

Figure 3-19: Data and clock timing for the DAC1 and DAC2 pair

To ensure basic operation of the DACs, a basic test was performed where a reference signal produced by a commercial signal generator was converted to a digital signal using the Avalanche's ADC at 25 MHz and was converted back to an analog signal using the Avalanche's DAC at 25 MHz. Recovering the same reference signal from the DAC using an oscilloscope confirmed the basic operation of the ADCs and the DACs at 25-MHz. Due to time constraints, no further quantitative tests were performed on the DACs.

3.3.6 Gigabit Ethernet Communication

The Avalanche board can connect to any computer over Ethernet communication media. The Avalanche Ethernet communication media can adapt its operating speed to 10/100/1000 bit per second (bps); therefore, it is termed as Gigabit tri-mode Ethernet. Gigabit Ethernet is selected because of its high data throughput. Also, Ethernet media does not need a shared ground. The connection between the two sides is coupled for alternating signals only. This coupling helps prevent ground loop problems and reduces the unwanted noise from the computer side. The Avalanche board's Ethernet communication media design is shown in Figure 3-20.

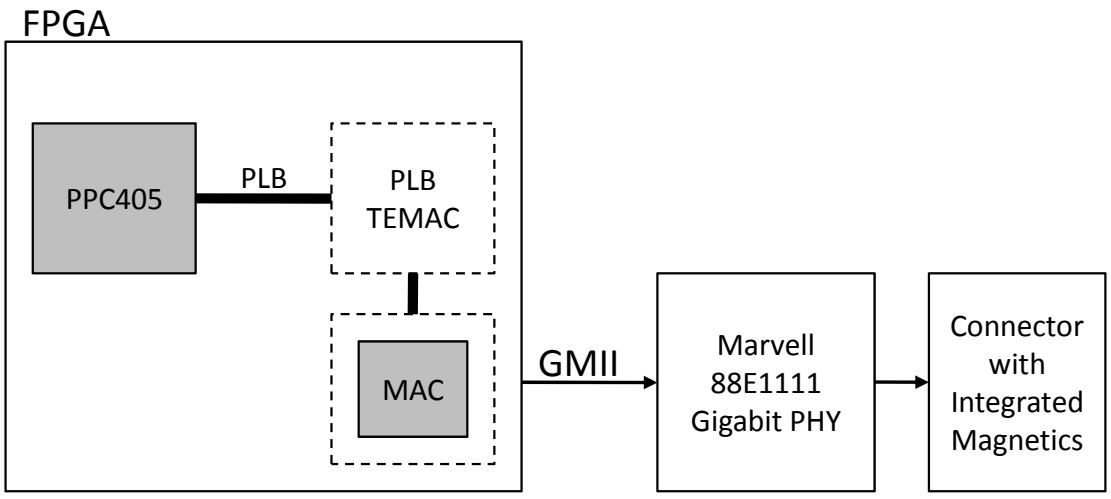

Figure 3-20: Avalanche board's Ethernet communication media design

The Avalanche board's Ethernet media design is shown in Figure 3-20. The Ethernet hardware consists of several layers. The Power PC (PPC) runs the Ethernet software drivers. The hard-wired Media Access Controller (MAC) implements the Giga bit Media Independent Interface (GMII) for connecting to the physical layer. The GMII

interface uses an 8-bit bus at 125MHz to provide an equivalent 1-Gbps data rate. The physical layer, which is implemented by Marvel's 88EE1111 gigabit PHY, performs the very low level functions of the Ethernet protocol and incorporates transreceivers for establishing an Ethernet link. The physical layer connects to the other computer through an RJ45 connector with integrated magnetic. The magnetic creates a high-pass coupling using transformers and includes termination features. The MAC is wrapped around by a Hard Tri-Mode Ethernet Media Access Controller (Hard TEMAC) peripheral, which creates access to the hard-wired MAC. The PLB TEMAC peripheral acts as an adapter which connects the Hard TEMAC to the PPC through the Processor Local Bus (PLB). In this configuration the PPC software can communicate with another PC. The gigabit tri-mode Ethernet was successfully implemented and used as the communication media connecting Avalanche to the host PC running our graphical user interface software. In our tests, the Ethernet link was limited to 100-bps by the host PC's Ethernet card speed. This limitation could be removed by upgrading the host PC's Ethernet card.

3.3.7 High-Speed Universal Serial Bus

As an alternative solution and to add to the flexibility of the communication solution, the Avalanche board is designed to include a high-speed universal serial bus (HS-USB). The HS-USB communication media has a 480Mbps data rate. ISP1581 USB PHY from NXP Semiconductors is selected to implement the physical layer of HS-USB. The NXP PHY device implements the low level but high-speed functions while the PPC implements the high level parts of the USB communication protocol. The ISP1581 PHY uses a 16-bit wide parallel interface for exchanging data with the PPC, which significantly reduces the required communication clock speed on the PPC side. Also, ISP1581 incorporates high-

speed transreceivers which can transfer data at 480-Mbps rate. Due to time limitations, the USB communication media was not tested and is left to be tested as a potential alternative solution over the Ethernet communication media.

3.3.8 RS232 Serial Bus

RS232 Serial communication is very simple to implement and requires very little resources. Furthermore, RS232 is widely used for engineering applications. To further extent the flexibility of the communication solution, RS232 communication standard is added to the Avalanche board. RS232 can be used as very simple but low speed communication solution. To convert the FPGA I/O levels to the RS232 signaling standard levels, Maxim's MAX3227E RS232 driver is used, which uses a capacitor charge pump to generate larger RS232 levels from a lower 3.3-V power supply. Using the MAX3227 RS232 driver, signaling rates of up to 1 Mbps can be achieved. To confirm the operation of the Avalanche's RS232 media, it was used at a baud rate of 115200 to successfully transfer the values of the onboard switches to a Hyper Terminal connection on the host PC.

3.3.9 Printed Circuit Board Design

A printed circuit board (PCB), shown in Figure 3-21, was designed by I for the motherboard design that was discussed in this section. The PCB was 8.9"X6.7" and consisted of 8-layers. The designed PCB was manufactured by Sierra Proto Express[2]. The stack-up of the PCB layers is shown in Figure 3-22 . The following four layers were used for routing signals: Top Layer, Signal Layer2, Signal Layer 3, and Bottom Layer. The

[2] Sierra Proto Express (http://www.protoexpress.com)

other four layers were used as power planes. The four signal layers are separated from each other by the power planes. This way each signal can have two reference power planes close to it to easily ensure controlled impedance; also, the signal layers will be isolated from each other and cross talk among signal layers can be minimized.

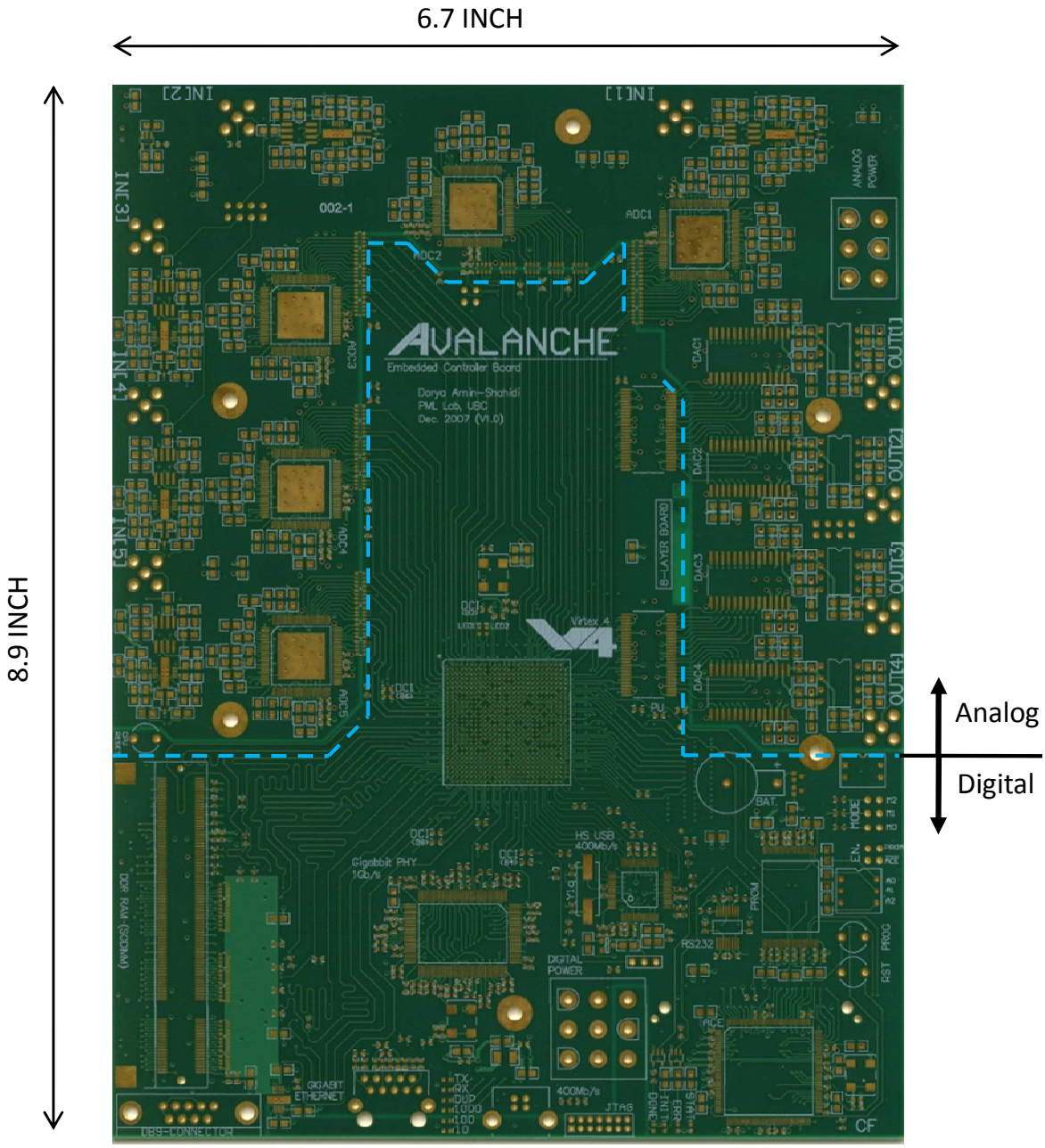

Figure 3-21: Top view of the motherboard's printed circuit board (PCB)

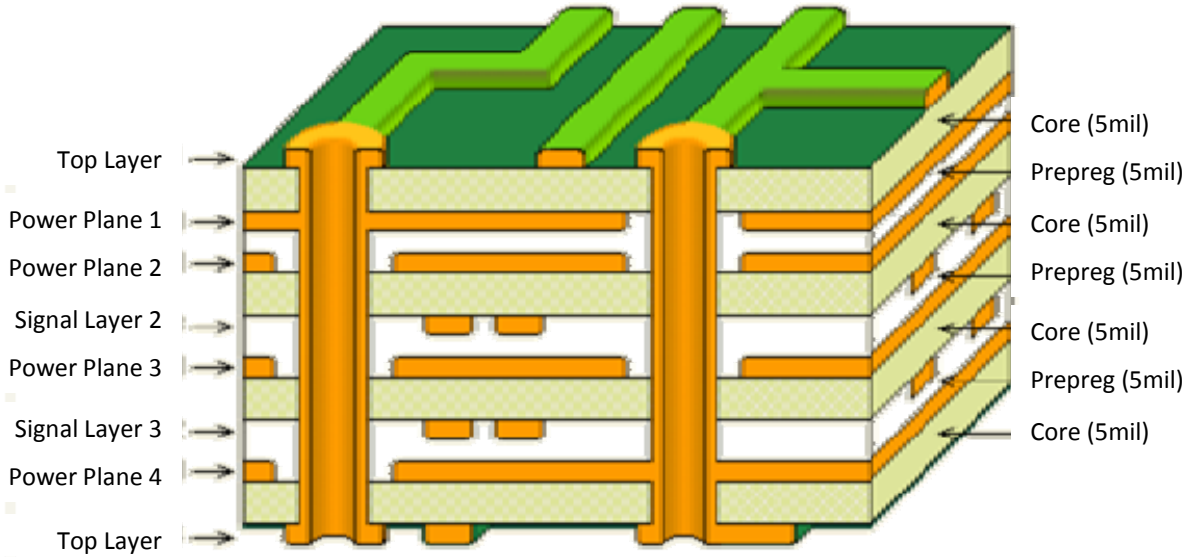

Figure 3-22: motherboard's PCB stack-up design

The PCB design for high-speed signals needs to ensure several requirements. Some of the major requirements are listed here:

1. Each signal needs to have a reference power plane on its adjacent layers. A power plane will qualify as a reference plane if it is by-passed to the ground plane close to the signal source and receivers. In this way, the plane will provide a minimized return path for the signal current and will minimize the inductance of the signal line. Any power plane can be used as a refrence plane as long as by-pass capacitors are added between that plane and the signal's driving power plane.
2. Small spacing between adjacent layers must be used. Large gap between layers will create a large loop area on the signals crossing between layers and will create a significant inductance on the signal lines. Also, the effective capacitance of the power layers, which can be useful for high frequency signals, will be reduced as well.
3. If possible, independent signals should not run in parallel for a long distance which creates mutual capacitance and inductance, and hence, brings cross-talk among them.

4. For cases where signal termination is required, the characteristic impedance of the transmission lines must match with terminating resistors. For the Avalanche board design, the Si9000 impedance solver, donated to UBC by Polar Instrument[3], was used to find trace thickness values which would result in the required characteristic impedance. The analysis was performed based on 0.005" board spacing and for single ended or differential signals with one or two adjacent reference planes. A schematic representation of the different signaling standards is shown in Figure 3-23, and their corresponding trace width and spacing values are listed in Table 3-5.

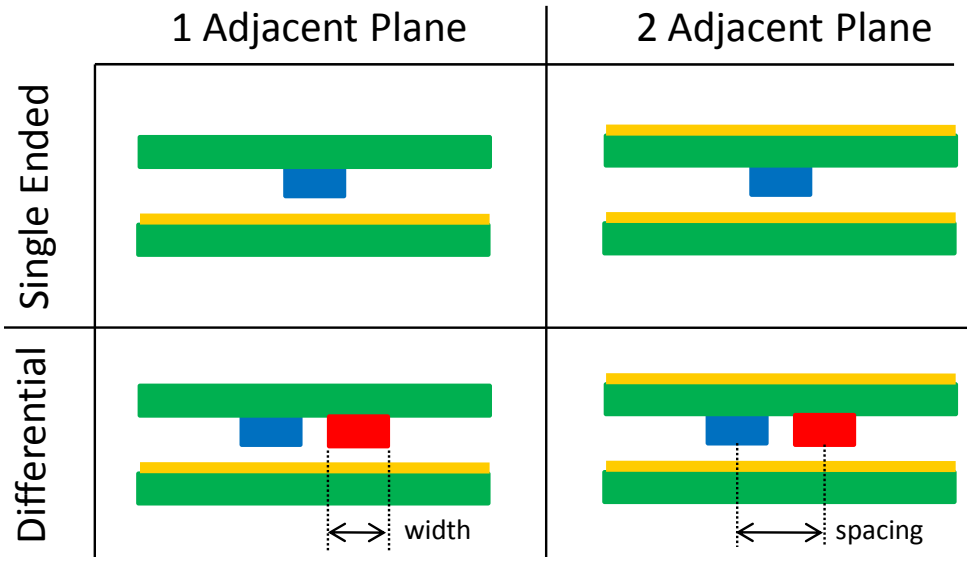

Figure 3-23: Schematic representation of the four signaling standards

Table 3-5: Trace design for 50-Ω impedance matching with 0.005" layer spacing

Signaling Standard	Adjacent Planes	Trace Thickness [in]	Pair Spacing [in]
Single	1	0.009	N/A
Single	2	0.005	N/A
Differential	1	0.009	0.016
Differential	2	0.004	0.011

[3] Polar Instruments Inc (http://www.polarinstruments.com)

The avalanche board operates on both digital and analog signals. As a result it contains both high-speed digital devices and sensitive analog devices. The digital devices operate by fast switching between the zero and one levels. These fast signal edges generate noise, which will affect the power planes as well all other signals. Because of the large threshold between the digital zero and one, the added noise does not affect the digital devices' operation when the signals are routed properly. However, any minor noise will undermine the precision of the analog components and will directly reduce the certainty of any analog operation. Therefore, the analog components need to be isolated from the noise generated by the digital devices; however, the two cannot be completely isolated because there are hybrid components which need to interface with both the digital parts and the analog components, such as the DAC and the ADC.

One typical way to provide isolation without completely disconnecting the analog and the digital sides is to move the digital and analog components to separate areas on the board and separate their power planes. The only common connection between the analog and the digital side, except for the signal lines, will be a single point ground connection between the analog and the digital grounds. The connection is necessary to keep the reference level for both sides the same so they can communicate. A single point connection will enforce a single controlled current return path for interaction between the analog and the digital side. This controlled return path does not allow direct and sharp returning currents; instead, the signals are forced to spread over the plane and find their way through the longer controlled return path. As a result, the noise associated with the signals will be spread out over the plane and its effect will be averaged out or at least minimized. Furthermore, the return path's longer length will result in an inductance

which would further filter out the high frequency noise. As shown in Figure 3-21, in the Avalanche Board's design, the analog components are moved to the top and the digital components are kept at the bottom. The analog and the digital power planes are separated by the blue dashed line. The white dashed line has an opening at the top right which represents the analog and digital grounds' single common connection point.

No digital and analog power planes should run on top of each other. Otherwise, there will be a mutual capacitance between the two planes, which will couple them for high frequency signals and will allow the noise to transfer from the digital plane to the analog plane. This can be easily achieved except where a single component needs both analog and digital power sources. One example of such component is the ADC. The ADC requires clean analog power for its sensitive internal components and needs separate power source for deriving its high-speed digital outputs. Directly connecting the two supplies to the same power plane will cause noise from the ADC's digital part. However, using the board's digital power planes is not an option either because that means that the digital power plane is in the analog region of the board. To address this issue, the digital power input pins of the ADC are connected to the analog power pins through a passive filter. The filter, shown in Figure 3-24, uses an inductor and capacitors to create a second order filter. The 3-dB bandwidth of the filter can be calculated as the following:

$$\frac{i_A}{i_D} = \frac{1}{1 + LCs^2} \tag{3.6}$$

$$\Longrightarrow \omega_{3dB} = \sqrt{\frac{1}{LC}} \tag{3.7}$$

For the selected capacitance and inductance values of C=4.7μF and L=2.2μH the filter will have a 3dB bandwidth of around 50-kHz. As a result, only the frequency

content of i_D below 50-kHz will influence the analog power plane. The frequency components of i_D at frequencies above 50-kHz will be supplied by the capacitors connected to the ADC's digital power pin. This method provides a low impedance voltage source for the ADC and removes the coupling between the ADC digital electronics and the analog power plane for frequencies above 50-kHz. It must be noted that because the ADC is operating at several mega samples per second conversion rate, most of the noise from its electronics will be above 50-kHz.

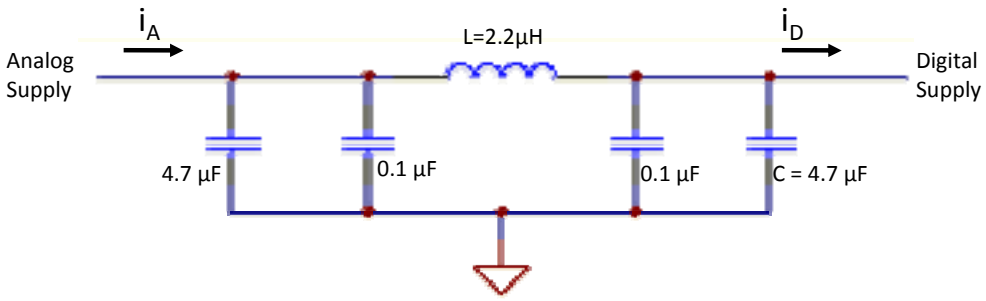

Figure 3-24: Isolating the analog power plane from the ADC digital power pin

3.4 Daughterboard Detailed Design

The manufactured daughterboard is shown in Figure 3-25. The daughterboard is designed to adapt the generic motherboard to the applications within the scope of this research. The daughterboard interfaces sensors and amplifiers used for this research to motherboard's input or output (I/O) resources. The daughterboard's flexible design provides the option to selectively switch the set of signals which connects to the motherboard's limited I/O resources. In this way, a single daughterboard design can be used for several applications within the scope of this research. The applications supported by the daughterboard include the following: encoder calibration, motor calibration, rotation motion control, and digital current control. The basic relation between the daughterboard components is shown in Figure 3-4.

In addition to physical routing of signals, the daughterboard converts the external input and output signals to levels which are compatible with the motherboard. It also includes components to perform operations which are specific to different applications. The following subsections discuss the detailed design of the daughterboard components.

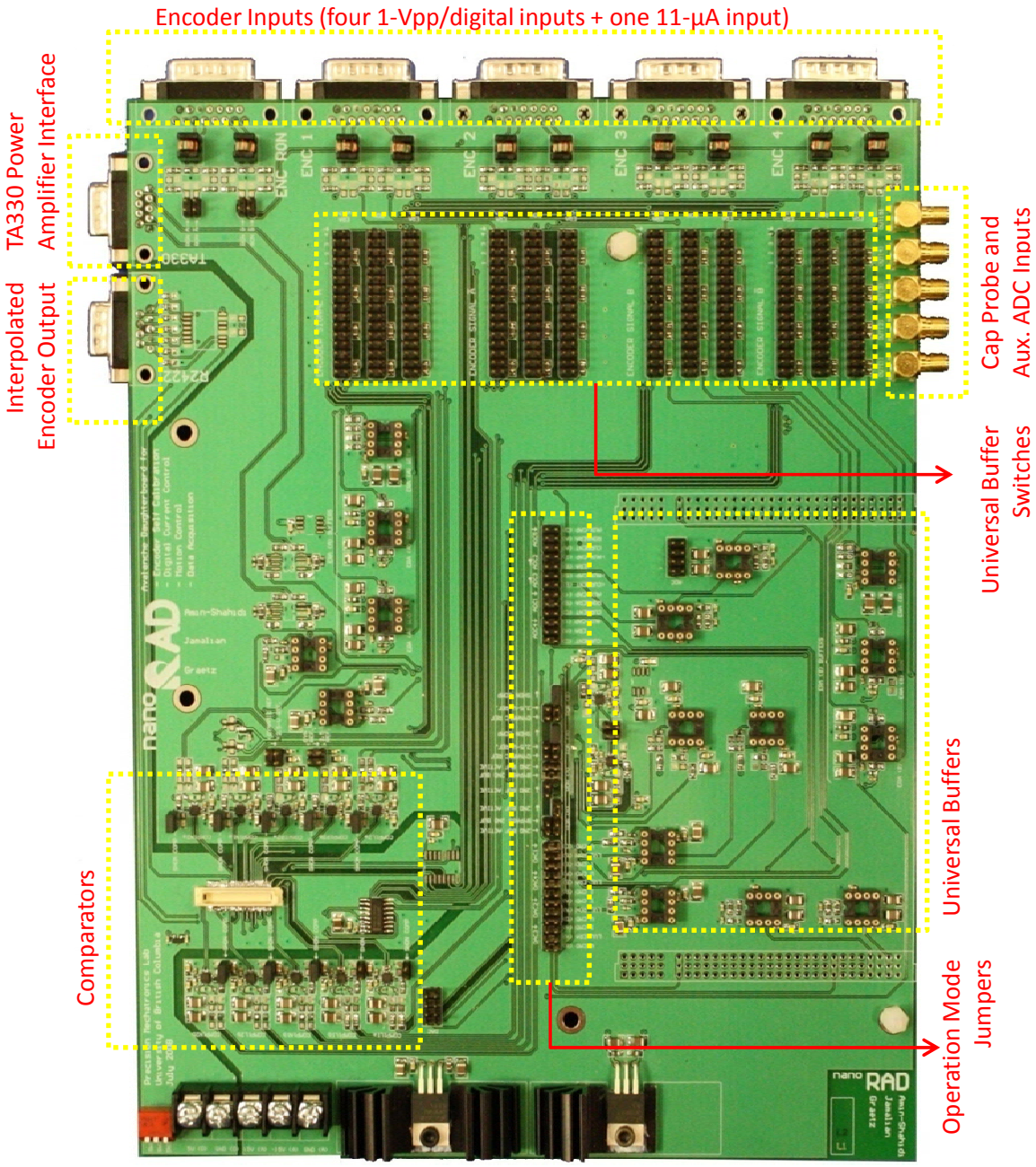

Figure 3-25: NanoRAD daughterboard design

3.4.1 Universal Buffers

The daughterboard will be used for encoder calibration. The encoder output signal can be analog or digital. An analog encoder output signal can be 1-V or 11-µA peak-to-valley differential sinusoidal signal. The daughterboard needs to incorporate analog buffers

which can convert the analog encoder signals to levels compatible with the ADC and the comparators. A digital encoder signal typically follows the RS422 differential signaling standard. The daughterboard requires digital level converters which can convert the RS422 signals to the 3.3-V CMOS/TTL standard compatible with the motherboard's general purpose digital I/Os (GPIO).

As discussed later in section 4.2.3, averaging 2 or 4 equally spaced encoder head signals can remove the cross sensitivity of the rotation measurements to spindle's radial motion. There two possible methods for performing this averaging:

1. Digital Averaging

   The ADCs on the motherboard are used to convert the signals into the digital domain and the digital signals will be averaged by the FPGA.

2. Analog Averaging

   A weighted analog adder is implemented using an operational amplifier which will calculate an average by adding and scaling the input signals.

The digital averaging method is very flexible because the averaging is implemented inside the FPGA and can be easily modified. However, this method will discretize the signal, and any encoder timing data will be valid to only ±5 ns. Also, this method will require many resources, and provided the number of available ADCs on the motherboard, only one 4-signal average can be generated. The analog averaging method will operate in the continuous domain and will contain more precise timing information. However, an

averaging circuit will be fixed; therefore, many adders need to be implemented to generate averages for all different signal combinations.

For easy and precise manipulation of the analog encoder signals, we have designed a universal buffer which accepts four encoder outputs, performs level conversion, and averages the user's selection of four encoder signals. The buffer uses a single operational amplifier and its averaging method can be easily configured by only changing the jumper switches. The ratio of R2 to R1 can be selected to achieve the appropriate level conversion at the same time. The schematic design of the buffer is shown in Figure 3-26. In the configuration shown in Figure 3-26, the buffer is averaging the phase-A signal from encoders 2 and 4. The jumper switches can be used to select other selections of four signals to be averaged. Examples of possible selections can include the average of the signal from all encoders or only the signal from one encoder. The relation between the universal buffer's input and output can be expressed as in equation 3.6.

$$V_{OUT} = \frac{R_2}{R_1}\left[\left(S_{1p} - S_{1n}\right) + \left(S_{2p} - S_{2n}\right) + \left(S_{3p} - S_{3n}\right) + \left(S_{4p} - S_{4n}\right)\right] \qquad (3.6)$$

The universal buffer's single-step conversion and averaging process minimizes the noise by eliminating the need for additional buffers which can add noise to the signal. Furthermore, given the universal buffer's configurable design, there is no longer a need for multiple averaging circuits which can increase the complexity, size, and the cost of the daughterboard design.

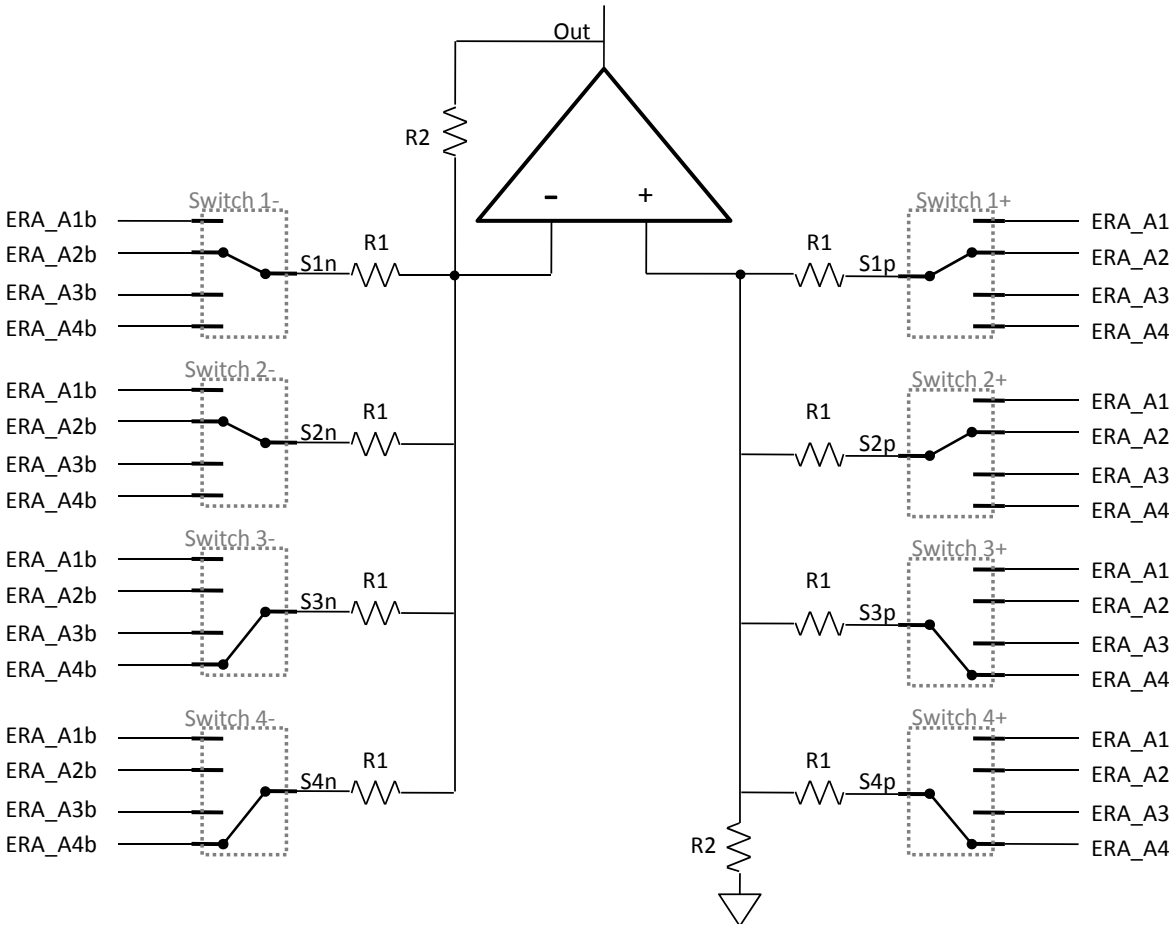

Figure 3-26: Universal buffer schematic design

3.4.2 High-Speed Comparators

To time or process an encoder signal using the FPGA, encoder signals has to be converted to a digital signal. The analog to digital conversion can be performed by the ADC, but that will limit the timing uncertainty to ±5 ns. To perform faster encoder digitization, which can achieve more precise timing information, we have added high speed level comparators to the daughterboard. The comparators create a digital pulse corresponding to the encoder signal crossing a threshold level. Using the daughterboard, the digitization can be performed by three different comparison schemes:

1. Self-Comparison: comparing the negative and the positive sides of the encoder's differential output signal output
2. Ground-Comparison: comparing the average of multiple encoder signals generated by a universal buffer to the ground reference
3. Dynamic-Comparison: comparing an averaged encoder signal to an arbitrary level generated by the DAC

The self-comparison method will directly digitize a single encoder. The second method can digitize the average of multiple encoder signals obtained by the universal buffers. The first two methods are limited to two pulses per encoder signal period. We have purposed a dynamic-comparison method which can generate pulses based on any arbitrary threshold level generated by the DAC. By changing the threshold generated by the DAC one can generated multiple pulses per encoder signal period. As soon as the encoder signal passes the first threshold set by the DAC, an encoder pulse is generated, and the DAC moves the threshold past the encoder signal where an additional pulse will be generated. This process can be repeated several times per encoder cycle.

The performance of the comparator device will set a limit on the achievable digitization precision. Comparator's input noise and dynamic performance are very important in generating a precise digital pulse closely corresponding to the encoder signal crossing a certain threshold. ADCMP605 comparator from Analog Devices is selected for this design. The ADCMP605's high-speed dynamics ensure a good dynamic performance. The comparator's output stage is rated to have a typical rise/fall transition time of 600-ps; such a fast output transition time creates a well defined transition edge and minimizes the cross-sensitivity of the digital pulse's timing to the digital noise. The comparator's input stage and output stage skews are specified, by its datasheet, to be around 70-ps, which is well below our timing precision. There usually exists a trade-off between the comparator's speed and noise specifications; however, ADCMP605's multi-stage design and differential low voltage signaling (LVDS) output standard significantly reduce the noise from the fast output stage. The configuration diagram of the ADCMP605 is shown in Figure 3-27. [23]

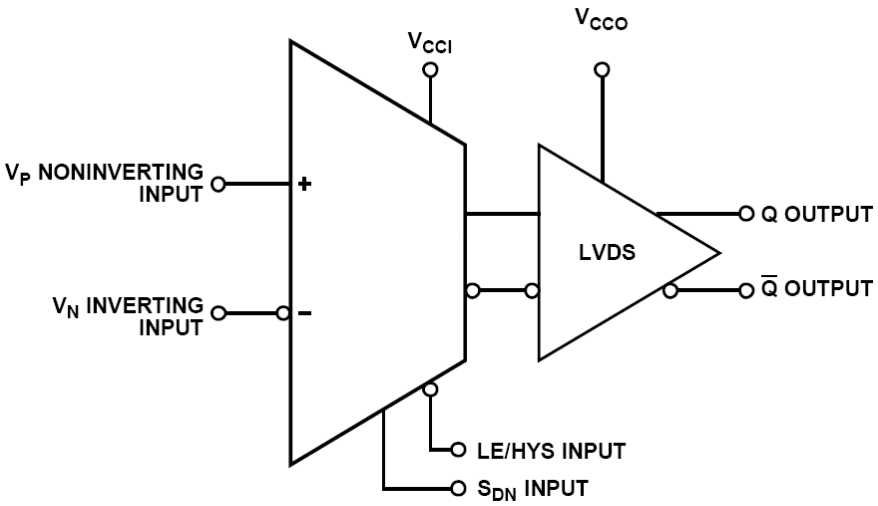

Figure 3-27: ADCMP605 comparator's functional block diagram

As shown in Figure 3-27, the ADCMP605 comparator consists of two sections. The left section, referred to as the analog section, accepts the analog inputs and performs the comparison operation. The right section, referred to as the digital section, derives the digital output lines based on the comparison results from the analog section. ADCMP605 uses differential buffers as its output buffer. This will significantly reduce the noise from the digital section. Because the ADCMP605 uses separated digital and analog power supplies, any residual noise on the digital power supply from the digital switching can be externally isolated from the analog supply. The selected comparator has a continuously configurable hysteresis level which can be set to prevent multiple switching at the zero crossing due to the high frequency noise. The hysteresis level is configured by changing the resistance of the resistor connecting between LE/HYS input pin and the ground. Finally, the ADCMP605's S_{DN} input is used to shut-down the comparators when not in use. In this way, they will not contribute to the noise present on the daughterboard.

3.4.3 Printed Circuit Board Design

The daughterboard's design was implemented on a 4-layer printed circuit board (PCB) shown in Figure 3-29. All layers were used for both power and signal routing. The daughterboard contains both analog and digital components. The analog components must be isolated from the digital components by moving them to separate sections of the board and using different power planes for them. The daughterboard will have its own onboard power supplies which will include: ±15V, 5V, and 3.3V. The daughterboard will use completely different voltage regulators for the digital and analog powers. To properly

communicate with the motherboard the daughterboard's analog and digital grounds connect to the motherboard's analog and digital grounds respectively. Different from the motherboard, the digital and the analog ground planes on the daughterboard are completely separated on the daughterboard; however, they connect together off the daughterboard at the same single common point connection on the motherboard. A schematic representation of this connection is shown in Figure 3-28. With this power plane design, there will be only a single ground connection between all digital and analog ground planes, and the ground connection between the daughterboard and the motherboard is only loaded as a signal return path.

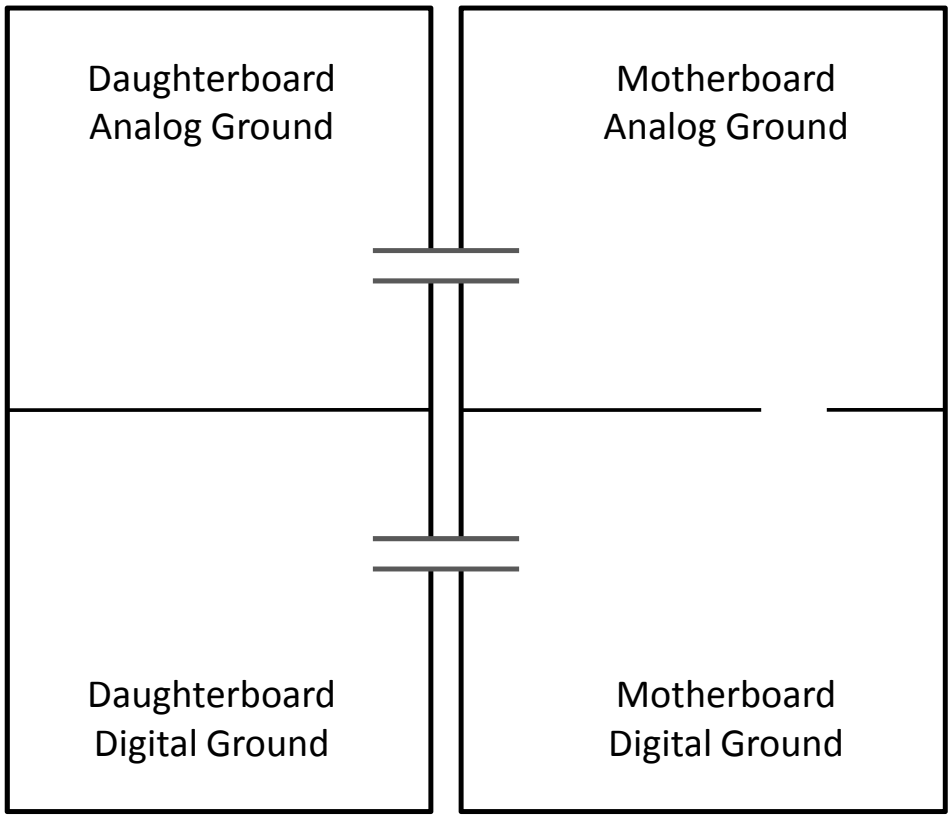

Figure 3-28: Schematic representation of motherboard-daughterboard ground plane partitioning

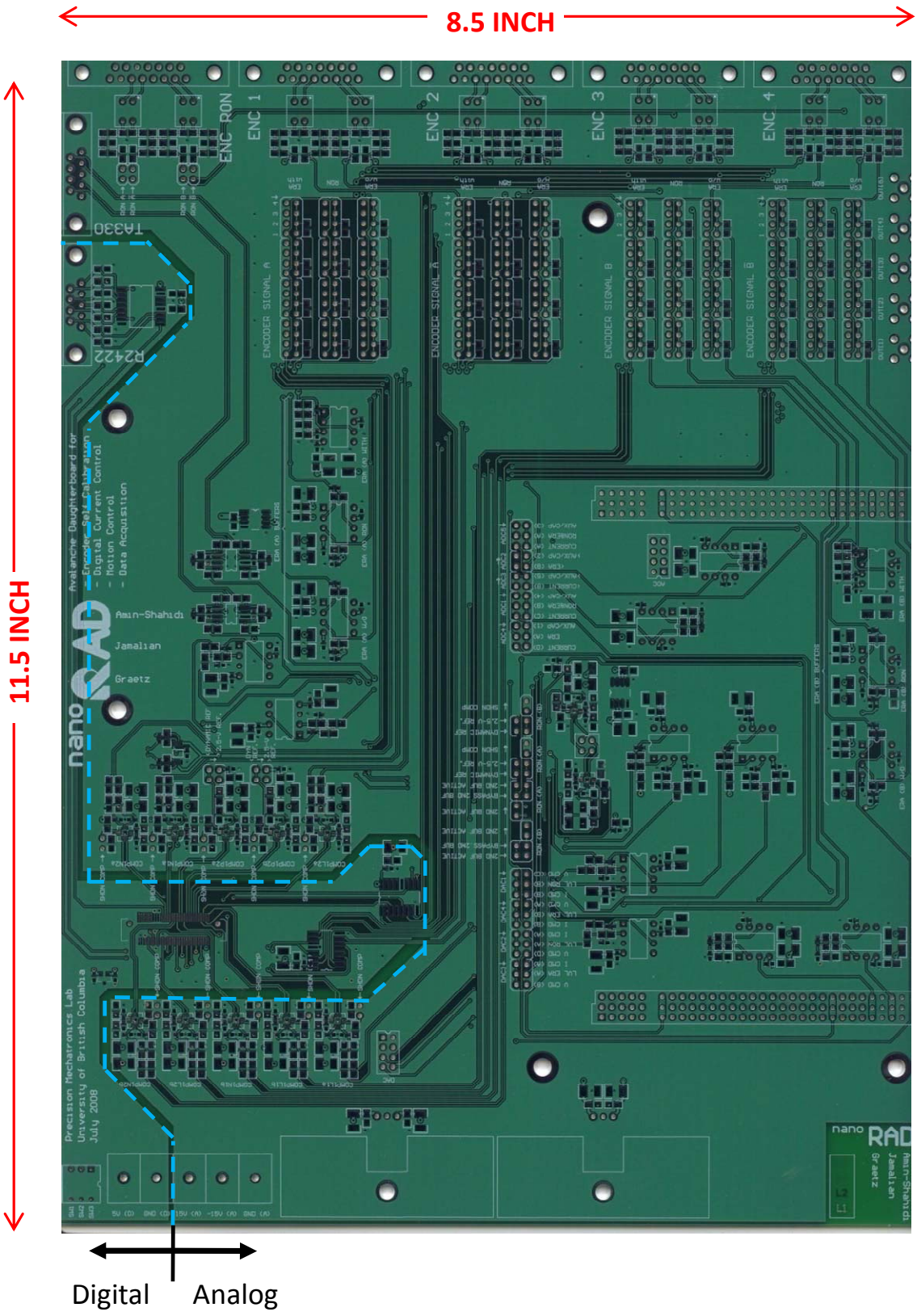

Figure 3-29: Daughterboard's printed circuit board (PCB) showing the ground partition

3.5 Powerboard Detailed Design

The powerboard, shown in Figure 3-30, is designed to supply the many voltage levels required by the motherboard. The summary of the power levels required by the motherboard and their required current limits is provided in Table 3-6. The majority of the levels are regulated by the powerboard. The generated power is passed to the motherboard through a cable connection, and a current-free cable connection, which does not exhibit any additional voltage drops, is used to feedback the actual voltage level on the motherboard. The powerboard is divided into two sections: analog power and digital power. The separation guarantees that the switching noise on the digital power planes does not transfer to the analog power planes.

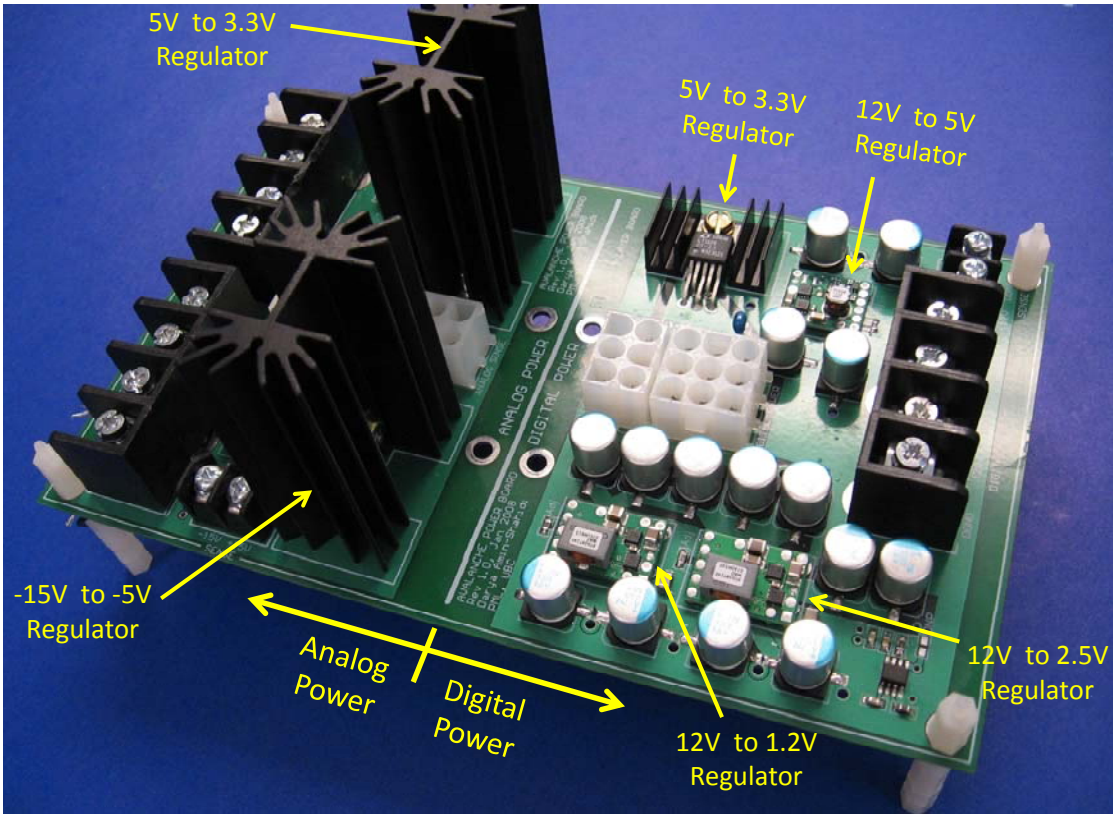

Figure 3-30: Powerboard used as the motherboard's power supply

The power regulation network is shown in Figure 3-31. The triangles represent linear regulators while the circles represent switching regulators. The selected device for each power level is specified in Table 3-6. In general, the power devices can be categorized into two general categories: switching regulators and linear regulators.

Table 3-6: Summary of motherboard power requirement and power supply solution

Supply		Current Capacity (A)		Power Device			
Type	Level (V)	Required	Designed	Type	Part#	Input(V)	Location
Digital	3.3	0.354	0.7	Linear	LT1129	5	Powerboard
	5	3.32	6	Switching	PTH08T230W	12	Powerboard
	2.5	8.97	10	Switching	PTH08T240W	12	Powerboard
	1.8	0.01	0.15	Linear	TPS73118	5	Powerboard
	1.25	1.75	3	Linear	TPS51100	2.5	Motherboard
	1.2	2.51	6	Switching	PTH08T240W	12	Powerboard
	12	6.5	8.5	AC-DC	SWS100	120AC	External
Analog	-5	0.342	0.5	Linear	LT1175	-15	Powerboard
	3.3	2	3	Linear	LT1764A	5	Powerboard
	5	4.485	6	Linear	Lambda HTD1	110 AC	External through the Powerboard
	15	0.5	1.5	Linear			
	-15	0.842	1.5	Linear			

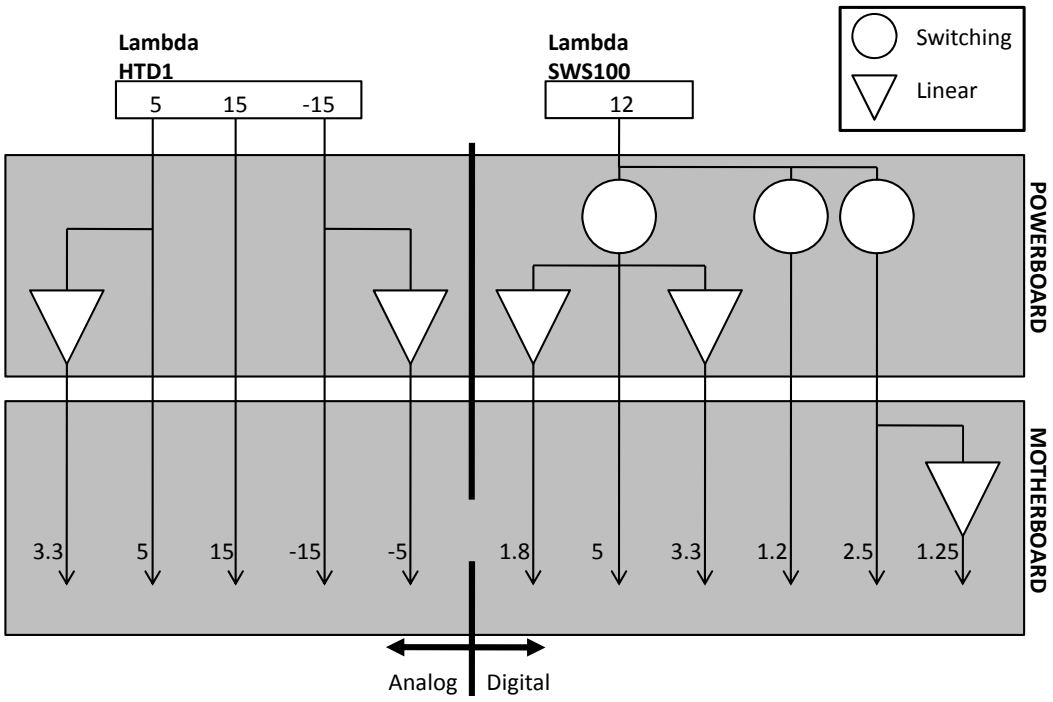

Figure 3-31: Power regulation network implemented on powerboard

The specified current ratings are at the power devices' output. The following subsections discuss the design considerations for using each regulator type in the powerboard and will overview the powerboard's PCB design.

3.5.1 Linear Regulators' Design Consideration

Linear regulators are active devices which generate a regulated output voltage at a specific voltage level. The term linear is used because the active device is operating in its linear region compared to the switching regulators where the transistors are used in their saturated mode. The linear regulators control the output voltage by varying the voltage drop occurring on the regulator. In this way the output voltage will be continuously regulated with high quality. On the down side, the difference between the output and the input voltage will be lost on the linear power device. This will reduce efficiency, but more importantly it will generate heat. By proper thermal design, the designer must ensure that the dissipated heat can be transferred to the ambient without over-heating the power device. Figure 3-32 shows the heat transfer model for a typical linear regulator.

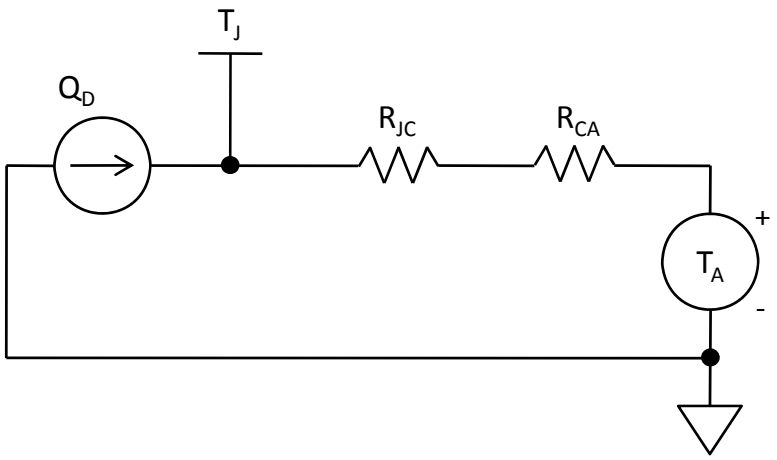

Figure 3-32: Lumped parameter heat transfer model for linear regulator

In Figure 3-32, Q_D represents the heat dissipation power. R_{JC} represents the thermal resistance between the linear regulator's junction and casing, and R_{CA} represents the thermal resistance between the regulator's casing and the ambient. The ground symbol represents the reference to which the ambient temperature, T_A, is defined to. The thermal design criterion is to limit the junction temperature, T_J, to stay below the linear regulator's allowable maximum junction temperature. The thermal resistance between the junction and the casing is fixed by the regulator's design. The ambient temperature depends on the specific application environment, and the dissipated heat is determined by the load. The dissipated heat can be calculated as in equation 3.7, where ΔV represents the voltage difference between the input and the output of the regulator and I_d represents the current sourced by the regulator. The linear regulators require a minimum difference between the input and output voltages; beyond this minimum difference, the closest available power level should be used as the input to minimize heat dissipation. The only design variable left is the thermal resistance between the casing and the ambient, R_{CA}, which can be reduced by using a heat sink. Heat sinks are rated by the effective thermal resistance between the device casing and the environment. The maximum allowable heat sink thermal resistance required for the discussed design parameters can be calculated as in equation 3.8. The designer needs to select a heat sink which is less than the maximum allowable thermal resistance R_{CA}^{max}.

$$Q_d = \Delta V \cdot I_d \tag{3.7}$$

$$R_{CA}^{max} = \frac{Q_d}{T_J^{max} - T_A} - R_{JC} \tag{3.8}$$

It is important to use remote feedback line with the high current linear regulators to compensate for resistive drops on the transmission lines. The regulators selected for the power board incorporate dedicated positive side feedback pin which allows them to compensate any drops on the forward supply line; however, no dedicated negative feedback is included to compensate for the drops on the return path. The linear regulators use their ground pins as the negative feedback line as well as the return path for the currents used by the regulator's small signal control electronics. Because the current drawn by the regulator's control electronics is relatively small, only a few mili-Amperes, we have used the ground as the negative feedback so that the regulator compensates the drops on the return path as well. The feedback configuration is shown in Figure 3-33. The thicker gray lines present load lines with high currents and the thinner black lines present load-free lines. The resistors between the motherboard and the powerboard model the resistance of the cables and the connections.

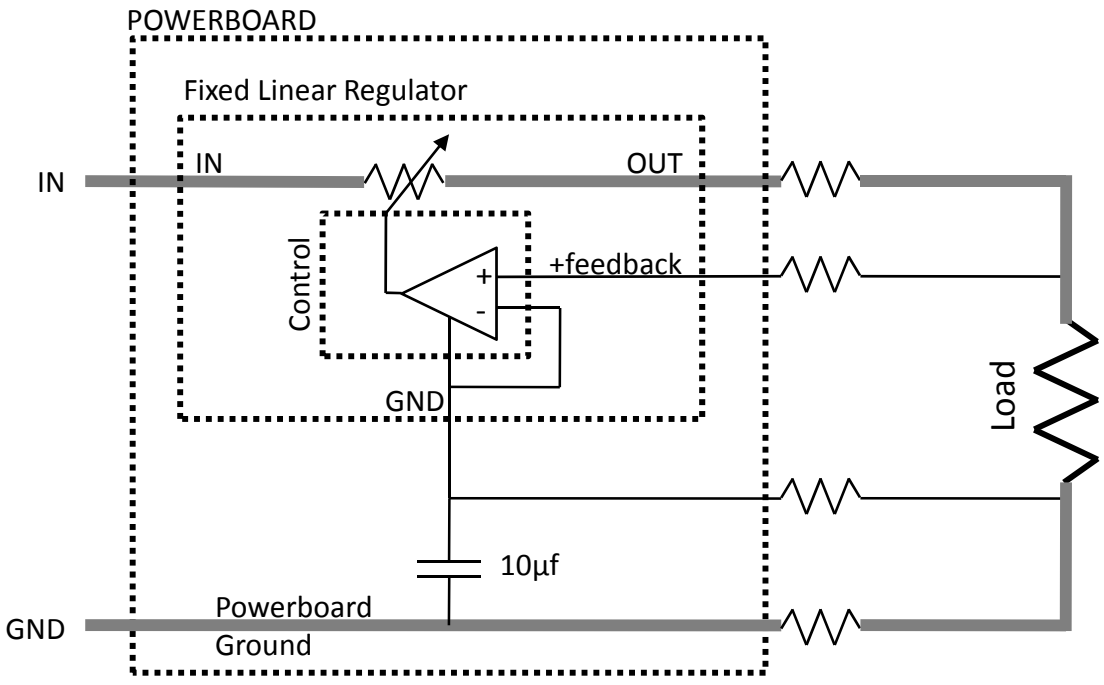

Figure 3-33: Linear regulator's feedback design

To implement negative side feedback with the linear regulators, the regulator's ground pin needs to be connected to the ground on the motherboard through a 10-inch long feedback cable. Because the regulator's ground is supplied through the long cable, it does not serve as a low impedance return path for the control electronics at high frequencies and causes the regulator to become unstable. To resolve this problem, we have added a 10-μF capacitor between the regulator's ground pin and the powerboard's ground. In this way, the regulator's ground pin is connected to the feedback ground for lower frequencies and can use the powerboard ground for higher frequencies. This configuration keeps the feedback functional and ensures the regulator's stability.

3.5.2 Switching Regulators' Design Considerations

Switching regulators generate a regulated output by rapidly switching the input supply. An electronic switch in series with the input and output supplies is used to control the ratio of the ON time to the OFF time of the input supply, called duty ratio. By controlling the duty ratio the supply can control the output level. The target duty ratio is set by small signal control electronics which update the value based on the feedback voltages. There are both step-up and step-down switching regulators, which can increase or decrease the input level respectively. However, only step-down regulators are utilized in the powerboard design. Because the supply is either on or off, the switching regulators will not dissipate much power and can achieve around 90% efficiency. Due to their high efficiency, switching regulators do not generate much heat, and therefore, have the potential to source large currents without requiring any large heat sink solution.

The switching regulator devices used in the power board are listed in Table 3-6. The selected devices have dedicated positive and negative feedback pins which are used in the powerboard design to compensate for the resistive drops on the loaded lines. For optimum dynamic performance the regulators require large output and input capacitors with low equivalent series resistance (ESR). The powerboard uses super low ESR conductive polymer aluminum electrolytic capacitors, part# APXA160ARA331MJC0G, from United Chemi-Con's PXA series, which have a 330-μF capacitance and only 14-mΩ ESR at 100-kHz. To further improve the performance multiple capacitors are added in parallel.

The selected switching regulators have a Track input, which can be used to provide a reference signal which the regulators will follow during the power up. The track input can be used to orchestrate a specific power-up sequence. Random power-up sequences can damage or confuse the complex digital devices which use multiple supply levels. By connecting together the track pins of the switching regulators they power up together and follow a unified ramp signal. The linear regulators used in the digital section do not have a track pin; however, their fast dynamics indirectly ensure that they follow ramp signal by closely following their input supply. The power-up sequence test result for the power board's digital lines is shown in Figure 3-34. As can be seen, all levels rise on a common ramp to their target output level. The 3.3 level, shown in yellow, is generated by a linear regulator from the 5-V line. As soon as the 5-V line reaches a high enough level, the linear regulator becomes enable and the 3.3 line joins the common ramp as well.

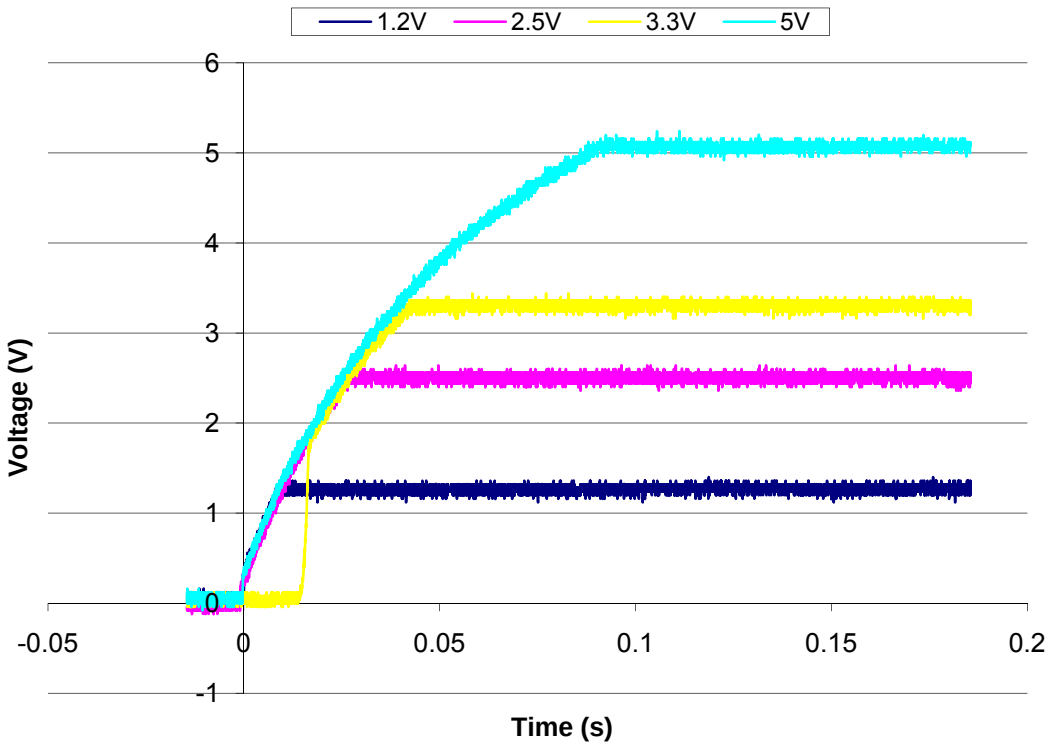

Figure 3-34: Power-up sequence of 1.2-V, 2.5-V, 3.3-V, and 5-V digital supplies

3.5.3 Printed Circuit Board Design:

A four-layer printed circuit board (PCB) was designed for the powerboard. The two middle layers were used for routing power lines only, and the top and bottom layers were used for routing signals as well as power lines. The manufactured PCB is shown in Figure 3-35. The analog power supplies are located on the top and are isolated from the digital power supplies at the bottom. A copper coating connected to the ground is fills any free area on the top and bottom layers of the powerboard. The ground copper layer on the top and bottom surface help with creating signal return paths and can help provide some degree of noise isolation.

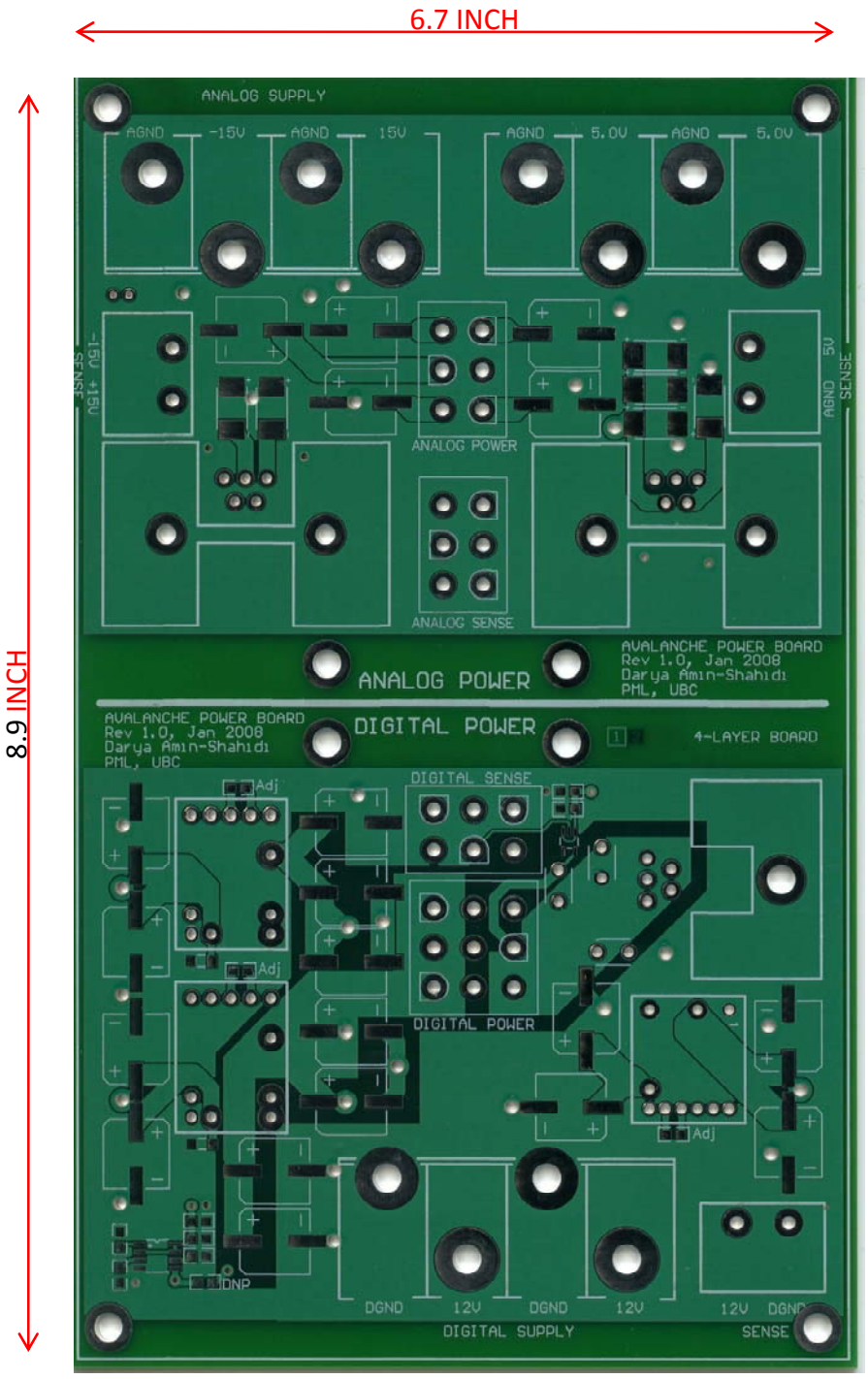

Figure 3-35: Top view of the powerboard's manufactured printed circuit board (PCB)

3.6 Software Design

Software was developed by K. Smeds[4], my colleague, to enable host-target communication with the motherboard and to conifigure the motherboard for encoder calibration. The software architecture, especially at the communication level, was developed such that it could be applied to different applications and boards with minor changes. So far, the software has been successfully used with the Avalanche motherboard for encoder calibration research. The software architecture is shown in Figure 3-36.

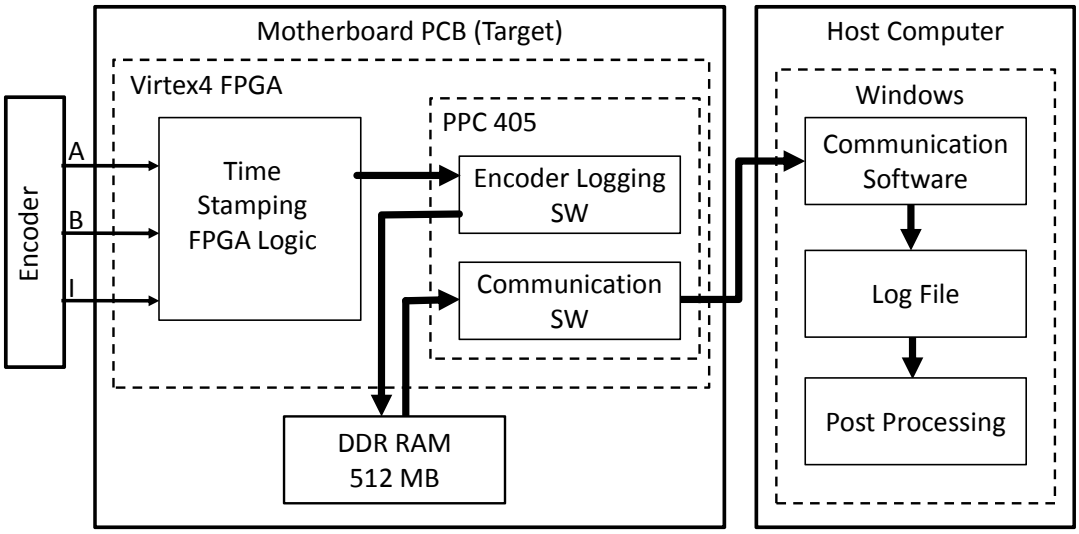

Figure 3-36: Software architecture for encoder calibration

On the motherboard, which serves as the target computer, the digitized encoder signals are processed by the Time Stamping FPGA Logic block which is implemented using the FPGA's logic fabric. The Time Stamping block decodes the quadrature A, B, and I encoder signals into encoder counts and measures their temporal duration (timestamps). The timestamps which are paired with a count number are then compressed

[4] Kris Smeds (ksmeds@interchange.ubc.ca), 2008, graduate student, Precision Mechatronics Laboratory, University of British Columbia.

in size. The Encoder Logging SW, which runs on the motherboard's embedded Power PC 405 processor (PPC 405), receives the compressed timestamp data from Time Stamping block and stores the data into motherboard's DDR RAM memory. The communication SW block, which runs on the PPC 405 processor, is responsible for communicating with the host computer. It requests data from the Encoder Logging SW block and then transfers the data from the DDR SDRAM memory to the host computer via TCPIP.

On the host computer, a graphical-user-interface (GUI), which has been developed using the National Instruments' LabView®, enables the user to request encoder calibration measurements of up to 200-million counts. The requested measurements are received back by the host interface and are logged into a file for post-processing. Currently, the post processing is performed using MATLAB® and involves extracting the compressed calibration measurements and applying the encoder calibration algorithm to them to obtain the encoder's error map.

Figure 3-37 shows a schematic representation of the Time Stamping block. The digital filters prevent detecting noise as encoder pulses. The filters create a fixed time shift but do not change the encoder pulse durations. The filtered encoder signals A and B are decoded to create 4 counts per period of the A & and B signals. The index signal decoder generates a spatially unique reset signal based on the I signal. The encoder count pulses are counted by the Position Calculator block to create a position value for each encoder count. The reset signal is used to reset the position counter once every revolution. In this way each position number is absolute position and consistently corresponds to a certain encoder. The Timestamp Calculator measures the time duration between consecutive counts. The Compression Logic block receives and compresses the

time stamp values and their corresponding position values so they occupy less space on the DDR SDRAM memory and can be sent to the PPC 405 using a single 32bit bus transfer. When using the ERA4282C encoder with 131072 counts per revolution, the compression of data enables encoder-calibration measurement over a longer range of about 1500 revolutions compared to 40 revolutions with no compression.

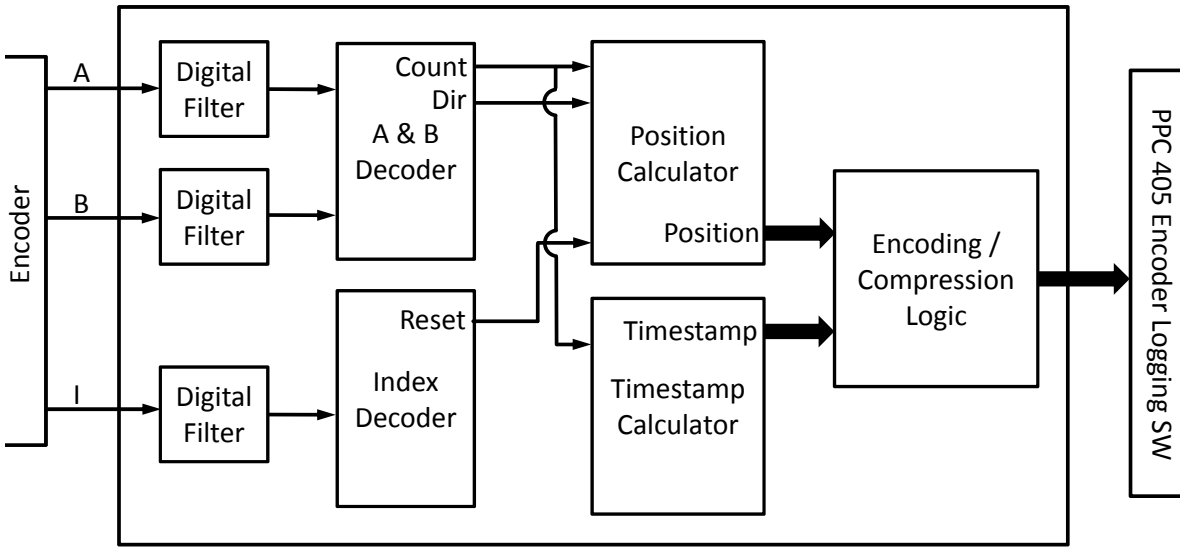

Figure 3-37: Schematic representation of the time stamping block implemented using the FPGA logic fabric

Transferring the data from the FPGA Logic to the host computer is handled by communication software which runs on the motherboard's embedded PPC 405 processor and connects to the host computer. The interface is developed using NI LabView on the host computer side and C on the motherboard (target computer) side. To increase flexibility of the interface, the communication software architecture is designed to have several layers, as shown in Figure 3-38. The transport layer establishes a link between the two sides and is implemented using the Ethernet communication media. The general request layer defines a set of functions which allows the host computer to control and use

the motherboard's resources. As an example, the Get Parameter function enables reading a parameter from the motherboard, and the Set Parameter function enables setting a parameter within the motherboard. The application layer, which is application specific, is used to execute a certain task by calling functions from the request layer and processing the obtained data.

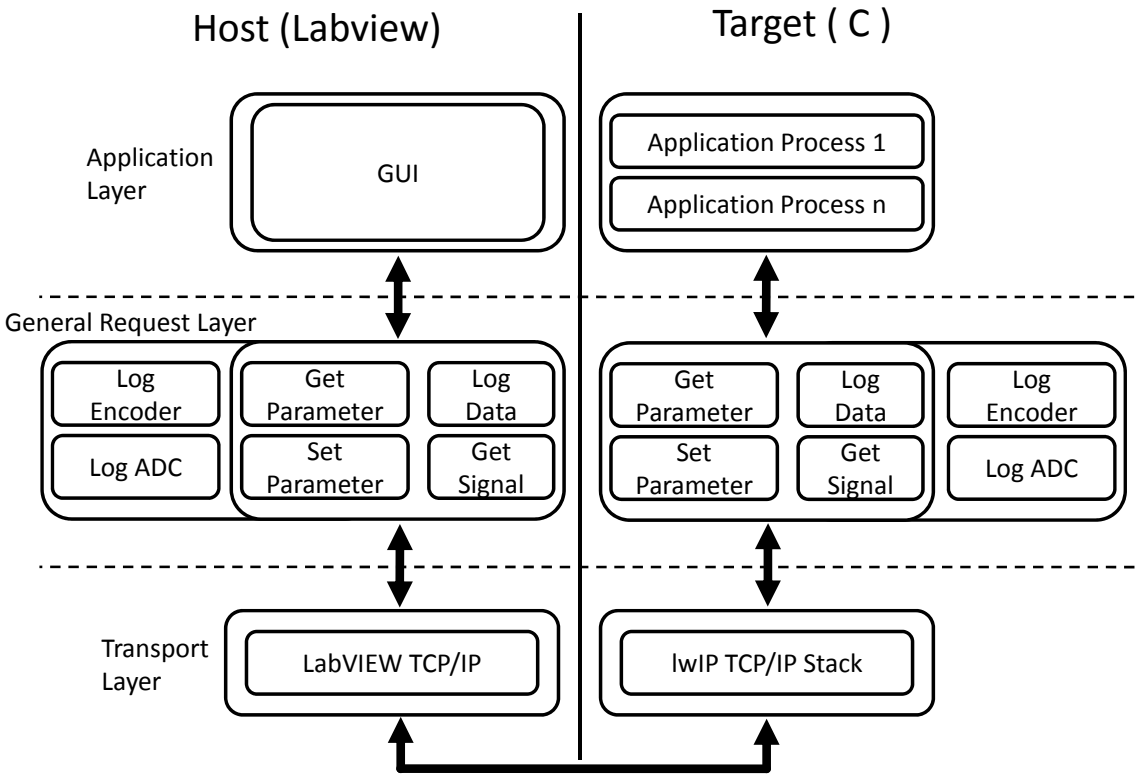

Figure 3-38: Communication Interface Design

The interface software described above was developed by my colleague, K. Smeds, and was successfully used with the Avalanche motherboard to record encoder calibration data for over 500 revolutions.

3.7 Multi-Phase Timer Design

Our encoder calibration method uses time measurements. As explained in section 4.2.1, the accuracy of the time measurements directly influences the accuracy of the encoder calibration results. The precision of a timer is inversely proportional to its timing frequency, and hence faster timers are desired.

Typically the FPGA logic fabric is used to develop timers. Figure 3-39 shows a typical timer design, where a counter counts the rising clock edges. The resulting timing resolution is limited by the FPGA's maximum clocking frequency, which typically ranges between 100 to 500 MHz. As a part of this thesis, I have designed a novel multi-phase timer architecture which can increase the timing frequency by up to 4-times.

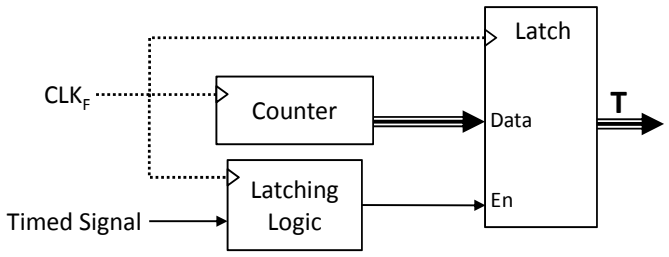

Figure 3-39: Simple digital timer design counting at CLK_f frequency

3.7.1 Multi-Phase Timer Design

The multi-phase timer architecture consists of multiple parallel individual timers with relatively phase-shifted clocks. The overall multi-phase timer's output is calculated as the sum of all the individual timers. While the general idea is applicable to any number of shifted clocks, we present the design using a quad multi-phase timer example, which has four 90-degree phase shifted clocks, shown in figure 3-40. In the multi-phase timer, the individual counter blocks can count at the maximum FPGA clocking frequency. Since the

individual counters' clocks are phase-shifted, the summation of their time-measurements ($T_{MP} = T_0 + T_{90} + T_{180} + T_{270}$) is updated at 4-times the maximum clocking frequency. As a result, this design can be used to increase the timing resolution by four times.

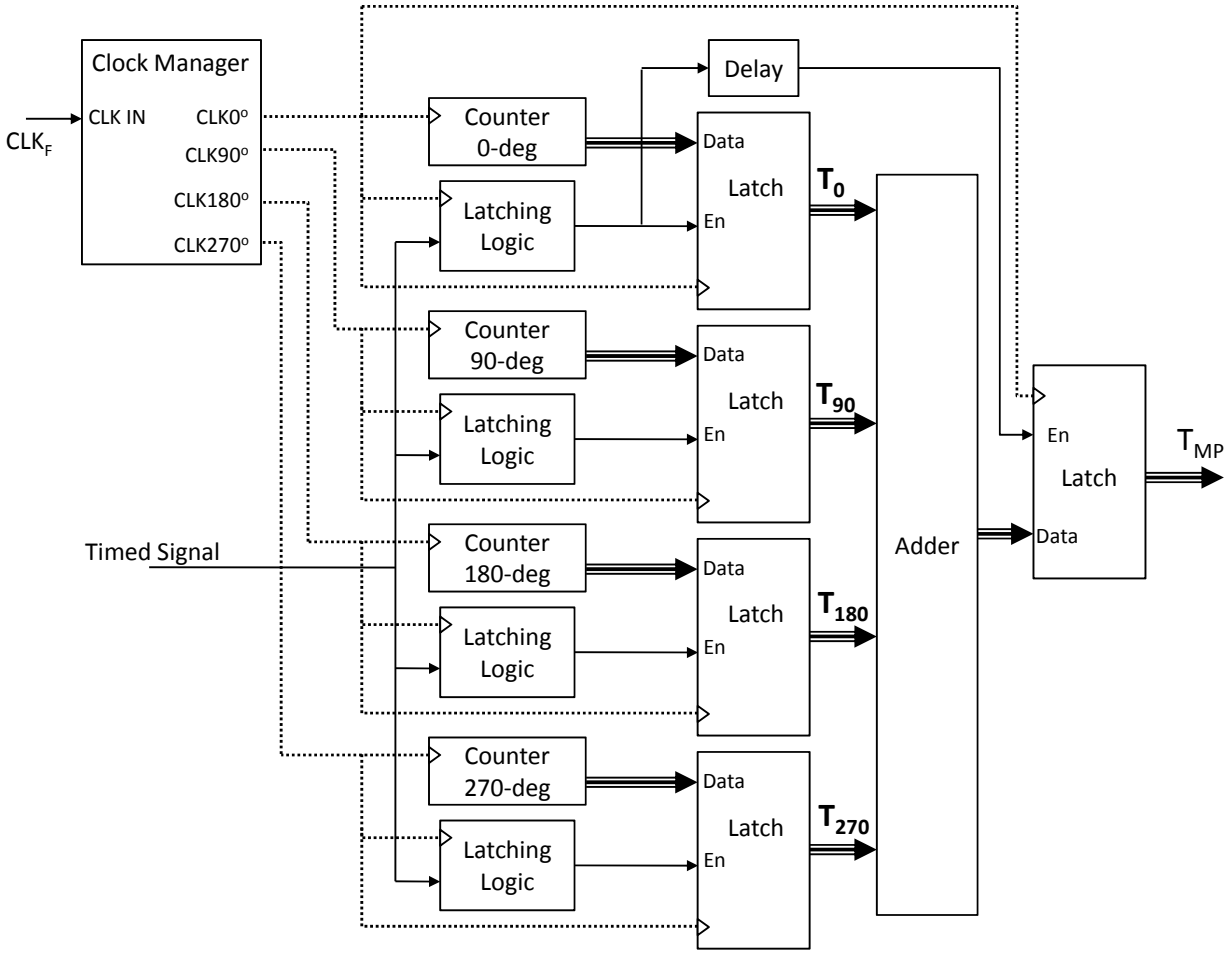

Figure 3-40: Multi-phase digital timer design, with four 90-deg phase shifted clock, counting at four times the CLK_f frequency

A clock manager block generates the four 90-degree phase shifted clocks. The latching logic blocks are used to create latch-enable signals based on the feature of the signal that is being timed. The addition operation is done using combinatorial logic, and its output must also be latched to a clock domain. To allow adequate time for the addition

operation, the latching of T_{MP} is delayed to few clocks after the individual timers' values (T_0, T_{90}, T_{180}, and T_{270}) are latched. In this way, the final measured time (T_{MP}) will have a resolution equivalent to a timer counting at 4-times the fundamental clock (CLK_F).

3.7.2 Performance Test and Calibration

We have developed a performance test for the multi-phase timers which can be used to confirm and enhance the effectiveness of the multi-phase timers. This test is presented here using the quad multi-phase timer example. The multi-phase timer is used to time a randomly varying signal. Considering the time-axis to be divided into segments bounded by the 4 phase-shifted clocks, as shown in figure 3-41, we would expect the random transitions of the timed-signal to have a distribution proportion to the relative width of the segments. Therefore, the distribution of the random transitions can be used to determine how well the quad multi-phase timer divides the time axis into four equally space segments.

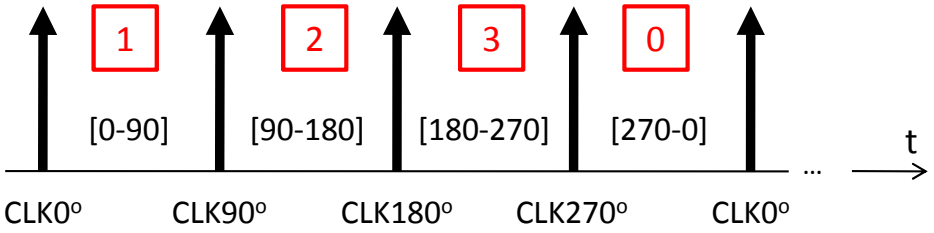

Figure 3-41: Division of the time axis by the 4 phase shifted clocks

The quad multi-phase measured time is determined as

$$T_{MP} = T_0 + T_{90} + T_{180} + T_{270} \tag{3.9}$$

where T_0, T_{90}, T_{180}, T_{270} are the time values from the 0-degree timer to the 270-degree timer respectively. The multi-phase timer is implemented to output the absolute time,

T_{MP}. In other words, its output is not rested to zero every time, and instead, it continually increases. As a result T_{MP} can have either one of the four values:

$$T_{MP} = \begin{cases} 4T_0 - 3 & t \in [0 - 90] \\ 4T_0 - 2 & t \in [90 - 180] \\ 4T_0 - 1 & t \in [180 - 270] \\ 4T_0 & t \in [270 - 0] \end{cases} \tag{3.10}$$

It can be shown that each case can be marked by the modulus of T_{MP} by 4 as

$$T \bmod 4 = \begin{cases} 1 & t \in [0 - 90] \\ 2 & t \in [90 - 180] \\ 3 & t \in [180 - 270] \\ 0 & t \in [270 - 0] \end{cases} \tag{3.10}$$

As a result, the modulus operation can be used determined the segment on the time axis where each time measurement belongs to. The distribution of the time measurement obtained in this way can be used to determine the spacing of the counting events. The proportion of each segment, $P(i)$, can be related to its width as

$$T_i = \frac{n(i)}{n(1) + n(2) + n(3) + n(4)} T_s = P(i) T_s \tag{3.11}$$

where T_i is the width of a the ith gap, n(i) is the number of occurrences of the ith gap's marker, $P(i)$ is the probability of the ith gap, and T_s is the fundamental clock period. A perfectly accurate quad multi-phase timer is expected to have $T_i = 0.25\, T_s$.

3.7.3 Experimental Results

A quad multi-phase timer with a fundamental clock frequency of 300MHz or an effective counting frequency of 1.2GHz has been developed, tested, and used for encoder calibration. To test the quad multi-phase timer, the encoder counts in free revolution were used as the random signal. The multi-phase timer was used to time 65.5-million encoder

counts over 500 free revolutions of the spindle. The distribution of the measured times are shown in figure 3-42.

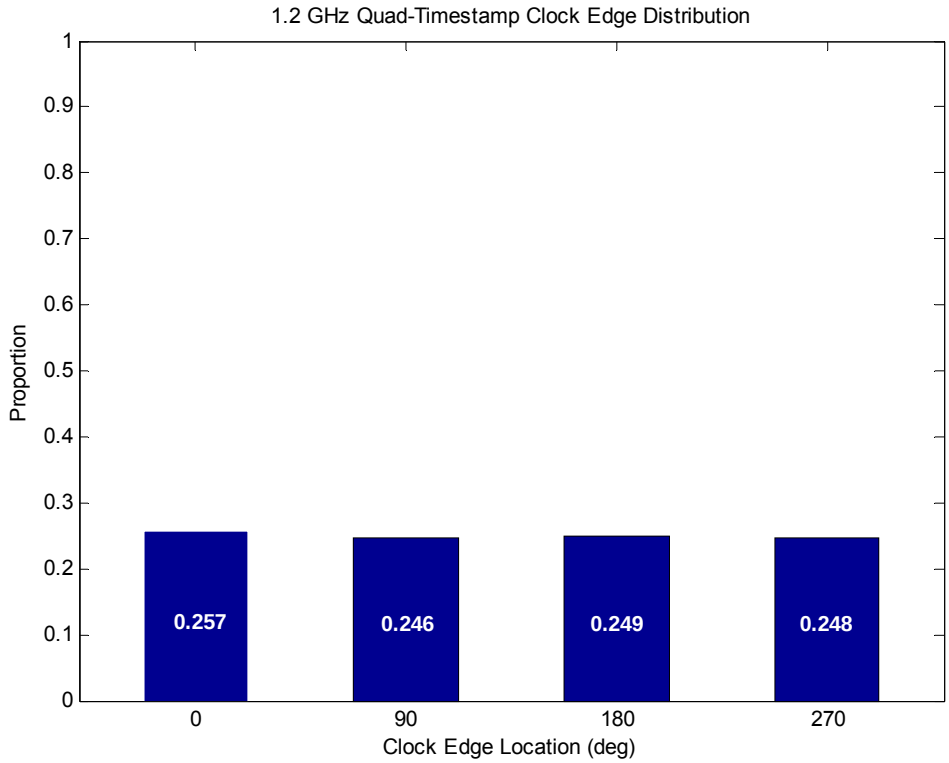

Figure 3-42: Experimental time measurement distribution for a quad multi-phase timer with an equivalent timing frequency of 1.2GHz

Using the distribution in figure 3-42, the average temporal width of the counter segments can be estimated as

$$\begin{cases} T_1 = 0.857\ ns \\ T_2 = 0.820\ ns \\ T_3 = 0.830\ ns \\ T_4 = 0.827\ ns \end{cases} \tag{3.12}$$

The maximum deviation from a 1.2GHz clock is 24ps which is negligible.

3.8 Summary

In this chapter, the need for a set of custom electronics was explained and the overall design of the electronics was discussed. A motherboard, a daughterboard, and a power board were designed and built as a part of this thesis. Details on the design of these boards are provided in this chapter.

The electronics, which were developed to meet the speed and precision requirements of this research, consisted of a generic high-speed motherboard and an application specific daughterboard. The motherboard is an 8-layer high-speed embedded controller designed around a Virtex4 FPGA. It can work at 300-MHz clock frequency and is used for fast digital signal processing. The daughterboard is a 4-layer board used to perform application specific analog processing, signal buffering, and signal routing. The daughterboard is used to adapt the motherboard for a specific application. Together, the motherboard and the daughterboard can be used for encoder calibration as well as digital motion and current control.

CHAPTER 4

Encoder Calibration

The state of the art encoder technology is only accurate to 2000 nano-radians without calibration. To achieve better angular measurement accuracy, the encoders need to be calibrated. In this chapter, we present the encoder calibration work. Section 4.1 overviews the calibration algorithm, developed in [1] by X. Lu, and introduces a recent improvement to the calibration method. In section 4.2 we identify the sources of uncertainty and analyze their effect on our calibration results. Section 4.3 covers the guidelines for selecting an optimal encoder. Simulation results, which are used to verify and evaluate the calibration algorithm, are presented in Section 4.4. The experimental setup and the results obtained are covered in section 4.5.

4.1 Calibration Algorithm

Several methods exist for calibrating rotary encoders, some of which are covered in *Prior Arts* section in Chapter-1. This thesis uses the encoder calibration method developed in [1] by X. Lu. This method uses the system dynamic equations to relate the captured time measurements of the encoder counts to their spatial spacing. In this way, the calibration algorithm can generate an error-map based on only two free revolutions of the platform. In [1], an experimental repeatability of 1 arc-second or 4.85 micro-radians was demonstrated. In thesis we intend to improve and experimentally demonstrate the accuracy limit of this method.

Since most of the work in this chapter is based around the algorithm developed in [1], we explain the details of the original algorithm in the following sub-section 4.2.1. Also, a recent improvement to this method is presented in sub-section 4.2.2.

4.1.1 Original Algorithm – Second Order Method [1]

The calibration algorithm developed by X. Lu in [1] is presented in this section. The calibration algorithm presented here is named as Second-Order method as it uses a second order discrete representation of the system dynamics to find the encoder count spatial widths based on the encoder count temporal durations.

Typical digital encoder signals are shown in Figure 4-1. The channel-A and channel-B signals are phase shifted by 90 degrees. The two channels can be processed using a quadrature decoding scheme to generate the spatial counts displayed as the series of arrows. For encoder which produces N counts per revolution, the ideal spacing

between the counts will be Δ_0=1/N revolution. However the counts spacing is not perfectly accurate. The calibration algorithm's goal is identify these deviations, so they can be compensated.

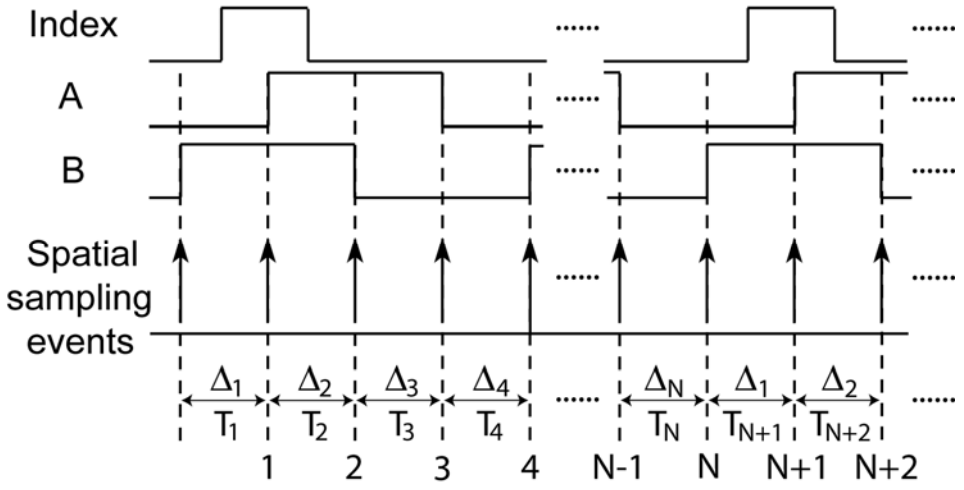

Figure 4-1: Spatial sampling events in rotary encoders [1]

This method relates the time between two successive counts to the count spacing based on the dynamic equations of the system. The dynamic model of the system is used to estimate the angular speed of the rotor. Once angular speed is known, the time measurements can be directly translated into spatial measurements as

$$\boldsymbol{\Delta_k = \omega_k \cdot T_k} \tag{4.1}$$

where Δ_k and $\boldsymbol{T_k}$ are the spatial width and time duration of the k^{th} count respectively. In equation 4.1 it is assumed that the spindle speed stays almost constant within one encoder count.

To achieve most consistent dynamics, the calibration algorithm is developed for an airbearing rotating in free response. The system's dynamics are modeled as an inertial element and a viscous damping element. The time-domain differential equation of the

system is shown in equation 4.2, where c is the normalized damping coefficient (spindle damping divided by the inertia of the rotor).

$$d\omega/dt + c\omega = 0 \tag{4.2}$$

Since the encoder counts are spatial discreet events, the spatial differential equation, equation 4.3, is derived from the time domain differential equation (equation 4.2) as

$$d\omega/d\theta = -c \tag{4.3}$$

As we will discuss later, the calibration algorithm fits a separate model to each individual revolution, and since the change in speed for a precision airbearing is very small during one revolution, the non-linear damping components will be small. To accommodate for speed dependent damping to some extent, the calibration algorithm still defines a first order damping coefficient as

$$c = c_0 + c_1(\omega - \omega_0), \tag{4.4}$$

where c_0 and c_1 are the damping parameters and ω_0 represents the initial speed for the revolution under consideration. The system's spatial-domain differential equation in equation 4.3 can be rewritten as

$$d\omega/d\theta = -c_0 - c_1(\omega - \omega_0). \tag{4.5}$$

The differential equation in 4.5 can be solved to obtain the system's free response equation as

$$\omega(\theta) = \omega_0 + \frac{c_0}{c_1}\left(e^{-c_1(\theta-\theta_0)} - 1\right), \tag{4.6}$$

where θ_0 and $\omega_0$are the initial angular position and speeds. The governing equation of motion, equation 6.6, can be expressed using the Taylor expansion as

$$\omega(\theta) = \omega_0 - c_0(\theta - \theta_0)(1 - c_1(\theta - \theta_0)/2) + \omega_e, \tag{4.7}$$

where ω_e represents the contribution of the third and higher order Taylor expansion terms. The error associated with this approximation is limited as

$$|\omega_e| < c_0 {c_1}^2 (\theta - \theta_0)^3/6 \tag{4.8}$$

To simplify the equation of motion, the contribution of ω_e is neglected for now. Up until now, the analysis has involved continuous functions of position. However, the data from the encoder will be discrete position data. To be used with the encoder data, equation 4.7 needs to be transformed into a discrete function of position as

$$\omega_k = \omega_0 - c_0 \Delta_0 \left(\frac{\theta_k - \theta_0}{\Delta_0}\right) + \frac{c_0 c_1 {\Delta_0}^2}{2} \left(\frac{\theta_k - \theta_0}{\Delta_0}\right)^2 \tag{4.9}$$

$$\omega_k = \omega_0 + a\left(\frac{\theta_k - \theta_0}{\Delta_0}\right) + b\left(\frac{\theta_k - \theta_0}{\Delta_0}\right)^2, \tag{4.10}$$

where $a = -c_0 \Delta_0$ and $b = c_0 c_1 {\Delta_0}^2/2$ are constants related to the system parameters, θ_k is the angular position at the k^{th} encoder count, and Δ_0 is the ideal encoder count spacing. Because the encoder counts are almost equally spaced, the calibration algorithm approximates $\theta_k \approx \theta_0 + k\Delta_0$ and simplifies equation 4.10 as

$$\omega_k = \omega_0 + ak + bk^2. \tag{4.11}$$

Combining equations 4.1 and 4.11, the count widths can expressed as

$$\Delta_k = (\omega_0 + ak + bk^2)T_k\,, \tag{4.12}$$

However, to use the equation 4.12 for encoder calibration, the system parameters *a* and *b* as well as the initial angular speed ω_0 need to be determined. Circular closure guarantees that the sum of the count widths along a full circle will add up to one revolution. The concept of circular closure is used with equation 4.12 to find angular speed as

$$N\Delta_0 = \omega_0 \sum_{S+1}^{S+N} T_i + a \sum_{S+1}^{S+N} T_i i + b \sum_{S+1}^{S+N} T_i i^2, \tag{4.13}$$

$$\Rightarrow \omega_0 = \frac{N\Delta_0}{\sum_{S+1}^{S+N} T_i} - a \frac{\sum_{S+1}^{S+N} T_i i}{\sum_{S+1}^{S+N} T_i} - b \frac{\sum_{S+1}^{S+N} T_i i^2}{\sum_{S+1}^{S+N} T_i}. \tag{4.14}$$

Replacing ω_0 in equation 4.12 with equation 4.14 a relation between the counts' spatial width and temporal duration can be obtained as

$$\frac{N\Delta_0}{\sum_{S+1}^{S+N} T_i} T_k = \Delta_k + a\left(\frac{\sum_{S+1}^{S+N} T_i i}{\sum_{S+1}^{S+N} T_i} - k\right) T_k + b\left(\frac{\sum_{S+1}^{S+N} T_i i^2}{\sum_{S+1}^{S+N} T_i} - k^2\right) T_k\,, \tag{4.15}$$

we rewrite equation 4.15 in a more compact form as

$$m = \Delta + aU + bV\,, \tag{4.16}$$

where

$$\Delta = \begin{bmatrix} \Delta_1 \\ \vdots \\ \Delta_S \\ \Delta_{S+1} \\ \vdots \\ \Delta_N \end{bmatrix};\ m = \frac{N\Delta_0}{\sum_{S+1}^{S+N} T_i} \begin{bmatrix} T_{N+1} \\ \vdots \\ T_{N+S} \\ T_{S+1} \\ \vdots \\ T_N \end{bmatrix};\ U = \frac{\sum_{S+1}^{S+N} T_i i}{\sum_{S+1}^{S+N} T_i} \begin{bmatrix} T_{N+1} \\ \vdots \\ T_{N+S} \\ T_{S+1} \\ \vdots \\ T_N \end{bmatrix} - \begin{bmatrix} (N+1)T_{N+1} \\ \vdots \\ (N+S)T_{N+S} \\ (S+1)T_{S+1} \\ \vdots \\ (N)T_N \end{bmatrix};$$

$$V = \frac{\sum_{S+1}^{S+N} T_i i^2}{\sum_{S+1}^{S+N} T_i} \begin{bmatrix} T_{N+1} \\ \vdots \\ T_{N+S} \\ T_{S+1} \\ \vdots \\ T_N \end{bmatrix} - \begin{bmatrix} (N+1)^2 T_{N+1} \\ \vdots \\ (N+S)^2 T_{N+S} \\ (S+1)^2 T_{S+1} \\ \vdots \\ (N)^2 T_N \end{bmatrix} \tag{4.17}$$

Vectors m, U, and V can be determined from the time measurements and are known. Parameters *a* and *b* are the only unknowns in equation 4.16, and once they are identified, equation 4.16 can be used to find the spatial widths, and hence the error map.

A set of N linear equations is available from the equation 4.16 with only two unknowns. Because of the uncertainties, there are no *a* and *b* that would satisfy all the N equations. As a result, the parameters *a* and *b* need to estimated using a fitting criteria,

such as the least square method. However, unless the variations in the count size are eliminated from the fitting equation, they will introduce error into the estimation of the parameters.

The calibration method uses a novel dynamic-reversal idea to eliminate the encoder count error from the fitting equations. Instead of relying on a single set of equations from one revolution, two sets from two revolutions are used. In Figure 4-2, the spindle speed is decaying with position. The first and the second sets are shown in red and blue respectively. The pulses with the arrows indicate the encoder index signal marking the beginning of each revolution. The equations for the two sets are expressed as

$$\begin{cases} m_1 = \Delta + a_1 U_1 + b_1 V_1 \\ m_2 = \Delta + a_2 U_2 + b_2 V_2 \end{cases}. \tag{4.18}$$

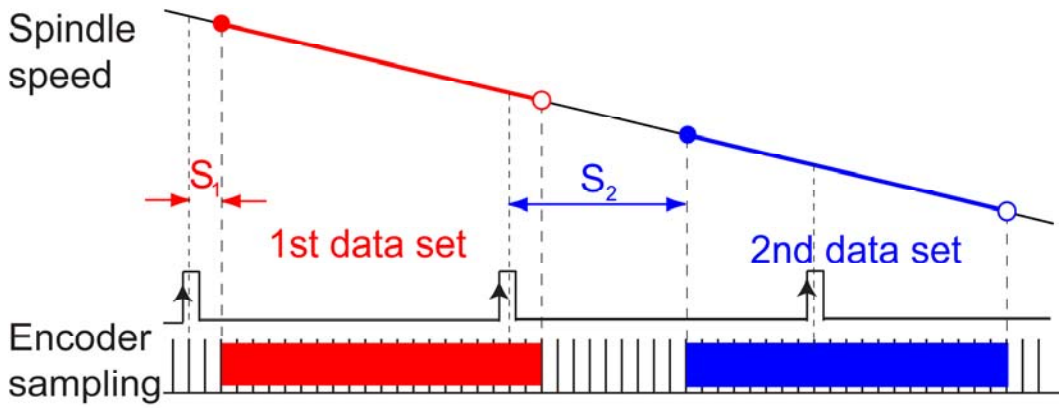

Figure 4-2: Two data sets used for self-calibration [1]

By subtracting the two sets, the artifact Δ can be removed and another equation only in terms of *a* and *b* is obtained as

$$m_1 - m_2 = a_1 U_1 - a_2 U_2 + b_1 V_1 - b_2 V_2, \tag{4.19}$$

This can be represented in matrix form as

$$\underbrace{[U_1 \quad V_1 \quad -U_2 \quad -V_2]}_{M}\begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{bmatrix} = m_1 - m_2 \tag{4.20}$$

As long as matrix M has full column rank, parameters a_1, b_1, a_2, and b_2 can be estimated as the values which would provide the minimum error. To ensure that matrix M has full column rank, vectors U_1, V_1, U_2, and V_2 must be uncorrelated. Since V is a quadratic function of position and U is a linear function of position, any V vector will not be correlated to any U vector. However, the pair of V vectors and the pair of U vectors will be uncorrelated only if shifted starting points are used. If shifted starting points are used, the vectors U_2 and V_2 will have to be circularly shifted so that the encoder counts (Δ) for the two sets end up on the same rows and cancel out. The circular shift results in the position for the elements on the same row to differ from between the two sets, and hence, both the V pair and the U pair of vectors will be uncorrelated. The shifted U vectors from the two sets are shown in Figure 4-3. The solid dots indicate the starting point of the data sets. As can be seen, both data sets have the same pattern with respect to their starting points; however, they do not have the same pattern once they are aligned based on the encoder counts.

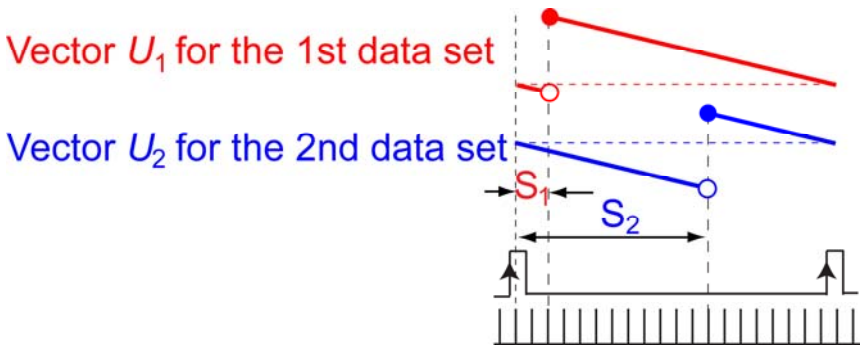

Figure 4-3: Vectors U_1 and U_2 of the two data sets [1]

For a matrix M with full column rank, the system parameters a_1, b_1, a_2, and b_2 are estimated as the values $\overline{a_1}$, $\overline{b_1}$, $\overline{a_2}$, and $\overline{b_2}$ for which the error is minimized. The least-square estimation method is selected, where the sum of the errors squared will be minimized. The estimated parameters obtained by applying the least-square method to the set of the linear equations in equation 4.20 is expressed as

$$\begin{bmatrix} \overline{a_1} \\ \overline{b_1} \\ \overline{a_2} \\ \overline{b_2} \end{bmatrix} = (M^T M)^{-1} M^T [m_1 - m_2] \,. \tag{4.21}$$

In the least square formulation expressed equation 4.18, the damping parameters a and b of can be different for each set. From one set to the other, the initial angular speed, ω_0, varies. For a second order damped system, a is expected to vary with speed but b should stay constant. However, because the actual system may exhibit higher order dynamics, the calibration algorithm is designed such that b is free to vary between the two sets as well.

Once the system parameters are identified, the spatial width of all the encoder counts can be determined as

$$\overline{\Delta_1} = m_1 - \overline{a_1} U_1 - \overline{b_1} V_1 \quad or \quad \overline{\Delta_2} = m_2 - \overline{a_2} U_2 - \overline{b_2} V_2 \,, \tag{4.22}$$

where $\overline{\Delta_1}$ and $\overline{\Delta_2}$ are the two sets of count widths estimated based on the temporal durations of the first and the second sets respectively. Although the two are expected to match closely and only one of them is required as the calibration result, the two sets can be compared as a measure of the calibration's effectiveness. The root-mean-squared (RMS) difference of $\overline{\Delta_1}$ and $\overline{\Delta_2}$ is termed as repeatability and is used as a measure of the calibration's effectiveness. It is evident that if the two sets do not match each other, there will be no way that they both could match the actual error map.

Once the widths of all the counts are determined, they can be processed to find the error associated with encoder reading as

$$\overline{E_k} = \sum_{i=1}^{k} (\overline{\Delta_1} - \Delta_0) - \sum_{k=1}^{N} \sum_{i=1}^{k} (\overline{\Delta_1} - \Delta_0)/N \qquad (4.23)$$

where $\overline{E_k}$ is the encoder's error at the k[th] encoder count. The first term accumulates the count error from the first count up to the k[th] count. To make a universal error-map convention, the second term is used to remove the average of error at all counts.

The calibration algorithm, presented above, makes three approximations in deriving an encoder error map:

1. The equation of motion is simplified using a 2[nd] order Taylor series
2. The spatial samplings are considered to be evenly spaced by Δ_0
3. The angular speed is approximated to be constant within one count

The calibration algorithm revisits these approximations and ensures that their contribution is within our error budget or compensated for by an accuracy-enhancement technique. Based on the experimentally identified damping parameters, it can be shown that the error resulting from the Taylor series approximation is negligible. The calibration process compensates for the error associated with the other two assumptions using an iterative process where the last calibration results are used to obtain an enhanced calibration map.

To obtain the first error map, it is necessary to assume evenly spaced spatial events. However, with a close estimate of the error map available, that approximation can be partially compensated. Equation 4.11can be expressed reformulated as

$$\omega_k = \omega_0 + aP_k + bP_k{}^2, \qquad (4.24)$$

where P_k replaces k and is expressed as

$$P_k = \sum_{i=1}^{k} \frac{\bar{\Delta}_i}{\Delta_0}. \tag{4.25}$$

In equation 4.1, the constant initial speed ω_k is assumed to stay constant during one encoder count. To eliminate error associated with this approximation, a time-varying angular speed is considered within one count. The angular speed is expressed as

$$\omega(\tau) = \omega_k e^{c(T_k - \tau)} \qquad for\ \tau \in [0, T_k), \tag{4.26}$$

where τ is representing time and $c = -\bar{a}/\Delta_0$ is the spindle damping calculated based on the estimated value of $\bar{a}$ obtained from the last calibration results. Equation 4.26 reflects only the constant damping. The first order damping contributions are considered to be relatively small.

Using equation 4.24 instead of ω_k in equation 4.1, the encoder count widths can be expressed as

$$\Delta_k = \int_0^{T_k} \omega(\tau) d\tau = \omega_k \frac{e^{cT_k} - 1}{c}. \tag{4.27}$$

The expression in 4.27 corrects the error introduced by the constant speed assumption. Equation 4.27 can be rewritten as

$$\Delta_k = \omega_k T_k R_k, \tag{4.28}$$

where the R_k can be considered as a correction factor and is defined as

$$R_k = \frac{e^{cT_k} - 1}{T_k c}. \tag{4.29}$$

Combining equations 4.24 and 4.29, an enhanced estimate of the encoder count widths is given as

$$\Delta_k = T_k R_k \left[\omega_0 + aP_k + bP_k{}^2\right], \tag{4.30}$$

To use 4.30, the initial angular speed ω_0, a, and b need to be re-estimated. As before, circular closure is used to estimate ω_0 as

$$\omega_0 = \frac{N\Delta_0}{\sum_{S+1}^{S+N} T_i R_i} - a\frac{\sum_{S+1}^{S+N} T_i R_i P_i}{\sum_{S+1}^{S+N} T_i R_i} - b\frac{\sum_{S+1}^{S+N} T_i R_i {P_i}^2}{\sum_{S+1}^{S+N} T_i R_i}\,. \tag{4.31}$$

Replacing $\omega_0$in equation 4.30 with equation 4.31, an enhanced version of the calibration linear equations can be written as

$$\frac{N\Delta_0}{\sum_{S+1}^{S+N} T_i R_i} T_k R_k = \Delta_k + a\left(\frac{\sum_{S+1}^{S+N} T_i R_i P_i}{\sum_{S+1}^{S+N} T_i R_i} - P_k\right) T_k R_k + b\left(\frac{\sum_{S+1}^{S+N} T_i R_i {P_i}^2}{\sum_{S+1}^{S+N} T_i R_i} - {P_k}^2\right) T_k R_k\,, \tag{4.32}$$

Equation 4.15 can be written in a compact form as

$$m_e = \Delta + aU_e + bV_e\,, \tag{4.33}$$

where

$$\Delta = \begin{bmatrix} \Delta_1 \\ \vdots \\ \Delta_S \\ \Delta_{S+1} \\ \vdots \\ \Delta_N \end{bmatrix};\ m_e = \frac{N\Delta_0}{\sum_{S+1}^{S+N} T_i R_i} \begin{bmatrix} T_{N+1}R_{N+1} \\ \vdots \\ T_{N+S}R_{N+S} \\ T_{S+1}R_{S+1} \\ \vdots \\ T_N R_N \end{bmatrix};$$

$$U_e = \frac{\sum_{S+1}^{S+N} T_i R_i P_i}{\sum_{S+1}^{S+N} T_i R_i} \begin{bmatrix} T_{N+1}R_{N+1} \\ \vdots \\ T_{N+S}R_{N+S} \\ T_{S+1}R_{S+1} \\ \vdots \\ T_N R_N \end{bmatrix} - \begin{bmatrix} P_{N+1}T_{N+1}R_{N+1} \\ \vdots \\ P_{N+S}T_{N+S}R_{N+S} \\ P_{S+1}T_{S+1}R_{S+1} \\ \vdots \\ P_N T_N R_N \end{bmatrix};$$

$$V_e = \frac{\sum_{S+1}^{S+N} T_i R_i {P_i}^2}{\sum_{S+1}^{S+N} T_i R_i} \begin{bmatrix} T_{N+1}R_{N+1} \\ \vdots \\ T_{N+S}R_{N+S} \\ T_{S+1}R_{S+1} \\ \vdots \\ T_N R_N \end{bmatrix} - \begin{bmatrix} {P_{N+1}}^2 T_{N+1}R_{N+1} \\ \vdots \\ {P_{N+S}}^2 T_{N+S}R_{N+S} \\ {P_{S+1}}^2 T_{S+1}R_{S+1} \\ \vdots \\ {P_N}^2 T_N R_N \end{bmatrix}.$$

The estimation of a and b parameters and the calculation of the error map is done the same way as for the first calibration results.

Initially, the calibration algorithm will identify an initial error map based on three main approximations. The calibrated error-map is enhanced iteratively by removing the error caused by the assumptions. The iteration can be carried out until they no longer modify the results. The calibration results after iterative compensation is termed as enhanced calibration results and is expected not to be affected by the initial assumptions.

4.1.2 Modified Calibration Algorithm

Applying the original second order (SO) method to encoder data generated in simulation, it is observed that the system damping estimations are too sensitive to uncertainties caused by limited timing resolution. The damping parameters are found based on a least square fit to equation 4.20. In the equation 4.20, the measurement matrix *M*, which consists of vectors m, V, and U of the two sets of data, is expected to reflect patterns corresponding to the system parameters. Figure 4-4 and Figure 4-5 show the measurement vectors included in *M* for a 100-MHz and a continuous timer respectively. Comparing the two cases together, the uncertainties due to limited time resolution can be seen as a hash imposed on the measurement vectors. Because the size of the uncertainties is comparable to the dynamic range of the patterns in *M*, they influence the least square fit and the damping estimation becomes less accurate.

To eliminate damping estimation error, a modification was made on the basis of the second order encoder calibration method. The modified method, which is named as second order integration method (SOI), uses spatial integration to increase the dynamic range of the measurement patterns with respect to the time resolution. The modification uses the fact that the integral of the time uncertainties will be bounded to the time

resolution while the integral of the patterns will have a much larger dynamic range. The integration method being a linear function, we can integrate both sides of equation 4.20 with respect to position and obtain a new set of linear equations as

$$\underbrace{[int(U_1) \quad int(V_1) \quad -int(U_2) \quad -int(V_2)]}_{int(M)} \begin{bmatrix} a_1 \\ b_1 \\ a_2 \\ b_2 \end{bmatrix} = int(m_1) - int(m_2) \qquad (4.34)$$

where the operator *int* is defined as

$$int(P) = Int\left(\begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \\ \vdots \\ p_N \end{bmatrix}\right) = \begin{bmatrix} p_1 \\ p_1 + p_2 \\ \vdots \\ \sum_{i=1}^{k} p_i \\ \vdots \\ \sum_{i=1}^{N} p_i \end{bmatrix}$$

Applying a least square fit to equation 4.34, the damping parameters can be estimated as

$$\begin{bmatrix} \overline{a_1} \\ \overline{b_1} \\ \overline{a_2} \\ \overline{b_2} \end{bmatrix} = (int(M)^T int(M))^{-1} int(M)^T [int(m_1) - int(m_2)]\,. \qquad (4.35)$$

Figure 4-6 and Figure 4-7 show the measurement vectors of the second order integral method for 100-MHz and continuous clocks. The hash due to time measurement uncertainty is now invisible when compared to the large dynamic range of the measurement patterns. As a result, the damping estimates will be much less sensitive to the time measurement uncertainties.

Later on, in the simulation and experimental work sections, it will be shown how the integral method improves the accuracy limit and consistency of the calibration algorithm. The ratio of the patterns to uncertainties for the integral method is large

enough to eliminate the damping estimates as the limiting factor. Further integration would only cause numeric error and would reduce accuracy.

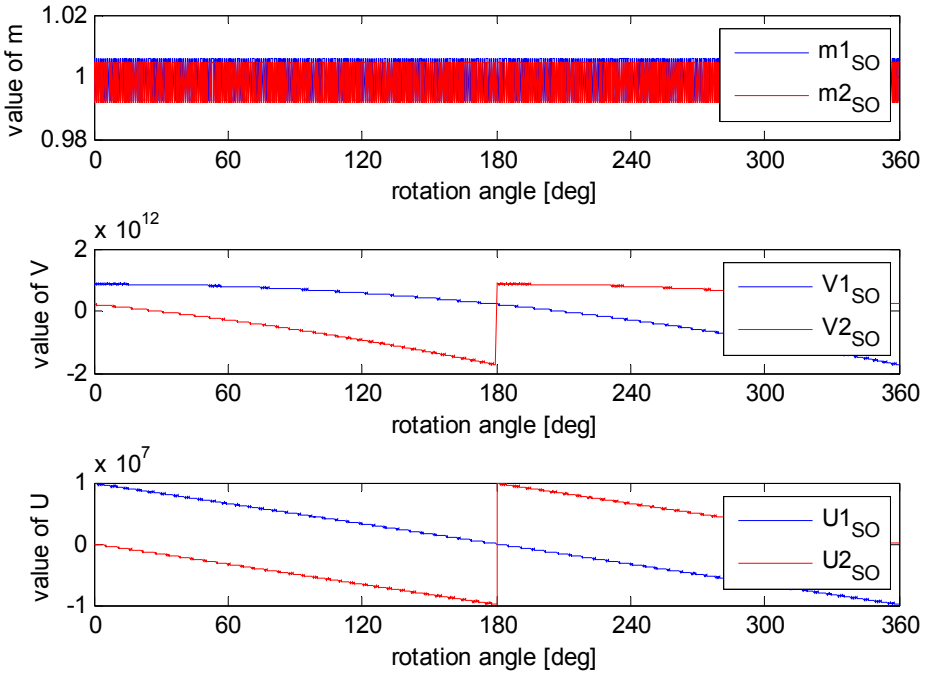

Figure 4-4: Measurement vectors m, V, and U for second order (SO) method at 600-RPM with a 100-MHz clock

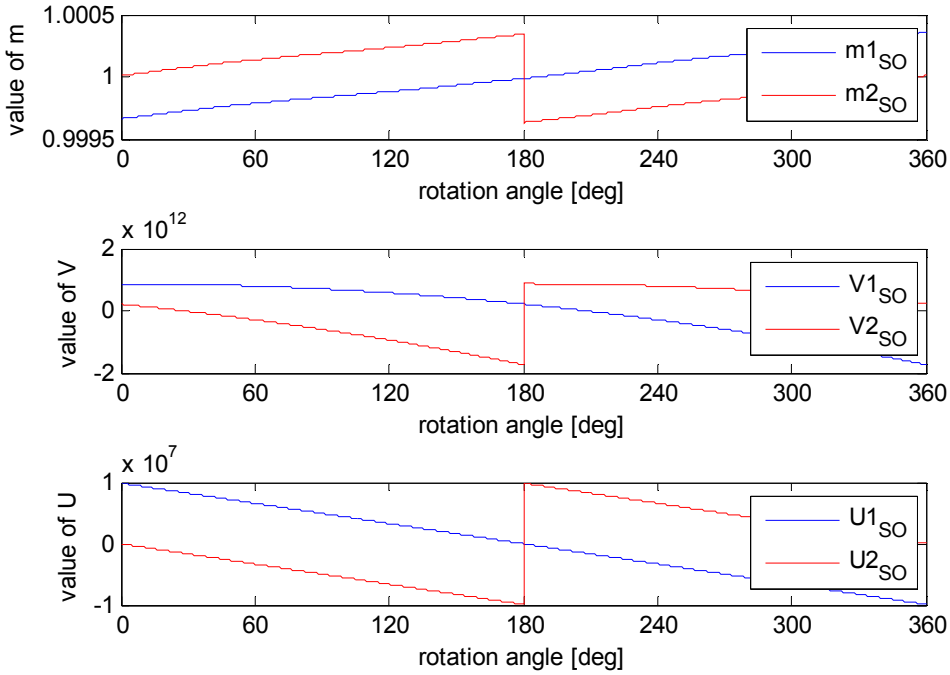

Figure 4-5: Measurement vectors m, V, and U for second order (SO) method at 600-RPM with a continuous clock

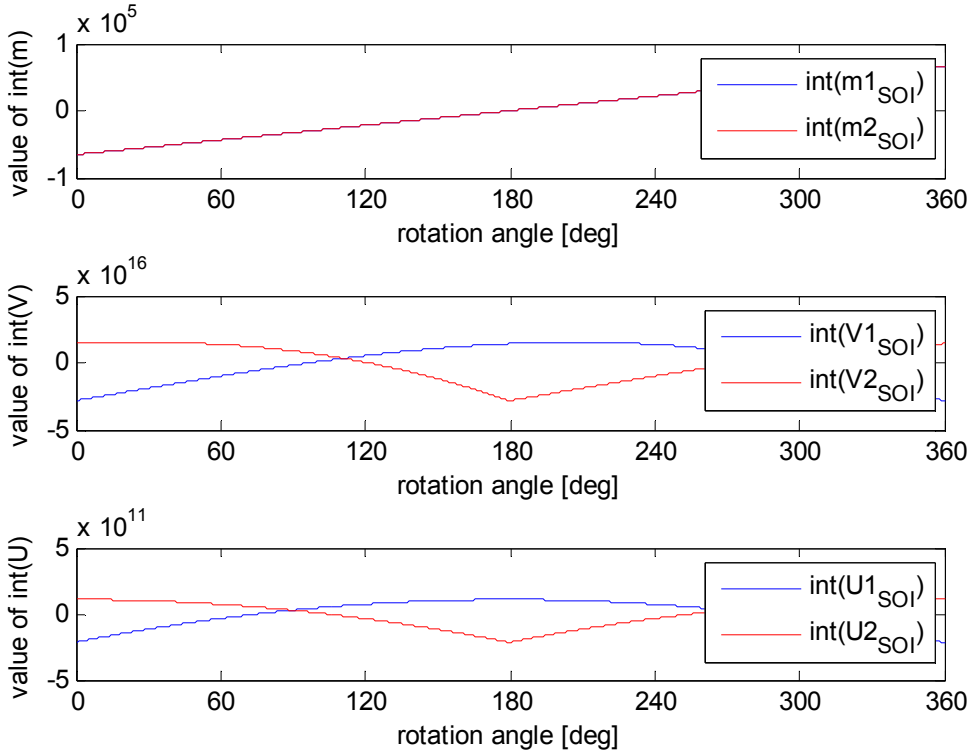

Figure 4-6: Measurement vectors m, V, and U for second order integral (SOI) method at 600-RPM with a 100-MHz clock

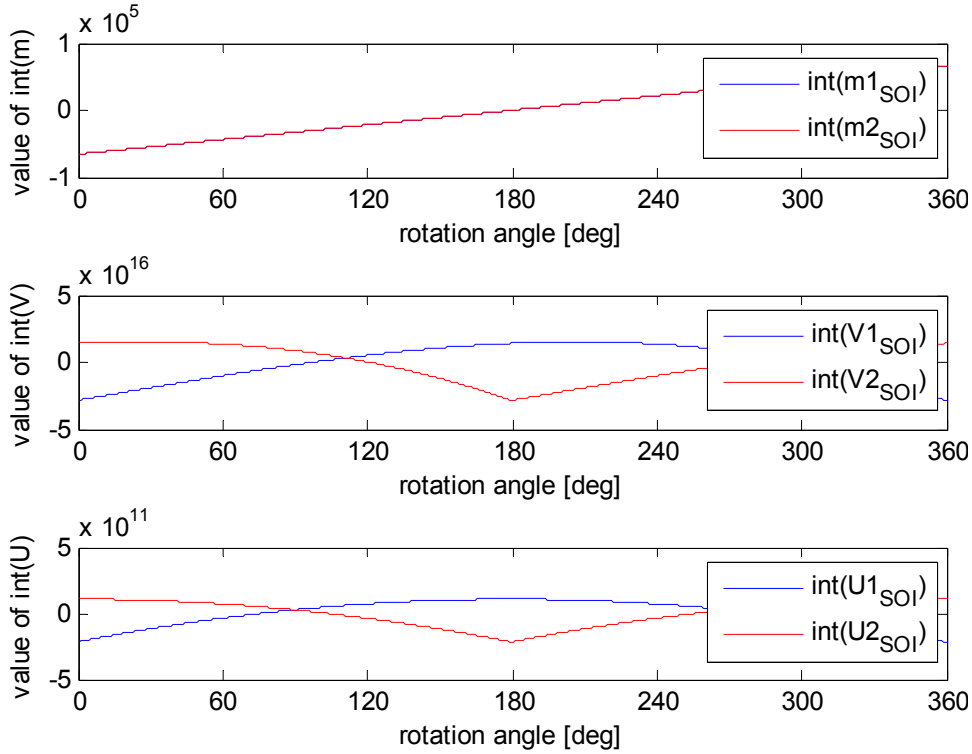

Figure 4-7: Measurement vectors m, V, and U for second order integral (SOI) method at 600-RPM with a continuous clock

4.2 Uncertainty Analysis

In theory, the encoder calibration algorithm is expected to provide a perfectly accurate error map; however, in reality, there are experimental uncertainties which generate non-repeatable error and limit the accuracy of the encoder calibration results. The main sources of uncertainties are time discretization, electrical noise, rotor vibrations, stator vibrations, and spindle dynamics. This section discusses these major sources of experimental uncertainty and their influence on the accuracy of the encoder calibration results.

The non-repeatable encoder signal error causes variations in the encoder error-map. Although this effect can be attenuated by averaging several encoder error-maps obtained under the same working condition, the final averaged error-map will not be able to compensate the random error in the encoder signal during the operation.

4.2.1 Time Discretization

The time measurements are the main input to the calibration algorithm. The calibration algorithm calculates the encoder count widths based on an estimate of angular-speed and the time measurement for the individual encoder counts. As a result, limits on the time resolution will directly result in limited calibration accuracy. The limited timing resolution causes the time measurements for a range of count widths to be the same. Intuitively, if this difference cannot be seen in the time measurements, there is no way for the calibration algorithm to distinguish the difference.

For timing clock frequency of F_S [Hz] the time measurements will be correct within $\pm T_S/2$ where $T_S = 1/F_S$ [sec] is the timing resolution. The associated angular

movement measurement uncertainty of the rotor within this time band can be calculated as

$$\varepsilon_t = \pm(T_S/2) \cdot (n/60), [rev] \tag{4.36}$$

where n is the angular speed of the rotor in rotations per minute. The value of ε_t, as given by equation 4.36 in units of revolutions, is the expected uncertainty of calibration results caused by time measurement discretization.

The expected uncertainty in calibration results caused by the limited time resolution is plotted versus rotation speed for different timer speeds in Figure 4-8. The uncertainty, which is the same as half of the distance traveled in a time step, increases with rotation speed and decrease with timing clock speed. As an example, based on Figure 4-8, the uncertainty caused by the limited time resolution at 200-RPM and with a 100-MHz clock is about 100 nano-radians.

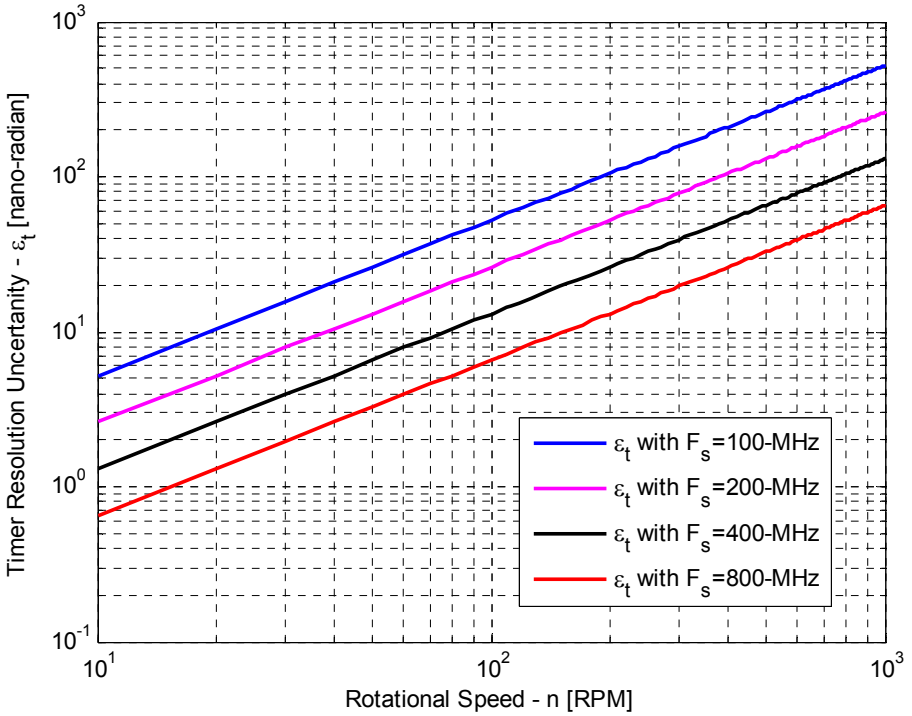

Figure 4-8: Expected encoder calibration uncertainty $\boldsymbol{\varepsilon_t}$ caused by limited time measurement resolution

4.2.2 Encoder Signal Noise

The encoder calibration method identifies the repeatable component of the encoder error. However, the calibration algorithm cannot compensate for any random variations in the encoder error. The noise on the encoder signal influences the location of the encoder counts, and hence, creates random encoder error which cannot be calibrated. The random error caused by signal noise is considered as an uncertainty in the calibration results.

The raw encoder signals are two analog sinusoidal waveforms phase-shifted by 90-degrees, each with $N/4$ periods per revolution, where N is the number of encoder counts generated per revolution. As shown in Figure 4-9, each individual channel is digitized based on its zero crossing point, and then, the two digitized signals are decoded using a quadrature scheme to generate N encoder counts per revolution.

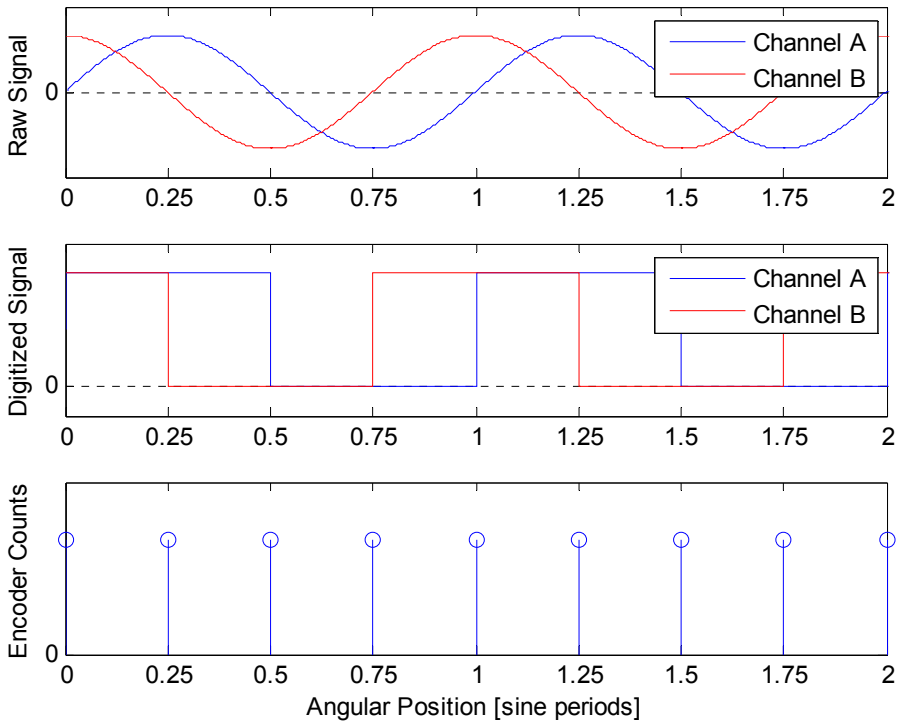

Figure 4-9: Analog encoder signals transformed into digitized signals and encoder counts

To analyze the effect of signal noise on count deviations, a sine function at its zero-crossing can be considered. Because of the limited slope of the raw signal at the zero crossing point, the superimpose signal noise can cause deviations in the zero-crossing point. The raw signal can be expressed as

$$V_A(\theta) = \mathrm{A}\sin\left(\frac{N}{4}\theta\right), \tag{4.37}$$

where A is the amplitude of the encoder signal, N is four times the number of sine period per revolution, and θ represents the rotation angle of the platform. Equation 4.37 can be used to relate variations in angle measurements θ to deviations in signal V_A as

$$\frac{dV_A}{d\theta} = \mathrm{A}\,\frac{N}{4}\cos\left(\frac{N}{4}\theta\right),$$

$$\Rightarrow d\theta = \frac{1}{A}\cdot\frac{4}{N}\sec\left(\frac{N}{4}\theta\right)\cdot dV_A \tag{4.38}$$

The signal to noise ratio (SNR) for any sensor is a measure of the sensor's noise performance and is defined as

$$SNR = \frac{A}{dV_A}, \tag{4.39}$$

where dV_A is the signal variations due to noise. Plugging equation 4.39 in equation 4.38 the uncertainties in angle measurement due to noise can be estimated as

$$d\theta = \frac{1}{SNR}\cdot\frac{4}{N}\sec\left(\frac{N}{4}\theta\right). \tag{4.40}$$

At the zero crossing of the encoder signal the secant term becomes one and the uncertainty in the angle measurement can be expressed as

$$d\theta = \frac{1}{SNR}\cdot\frac{4}{N} \tag{4.41}$$

As can be seen from equation 4.41, the uncertainty due to electrical noise depends on two parameters: the signal-to-noise ratio (SNR) and the number of sine waves generated per revolution for the encoder. Figure 4-10 shows a plot of the uncertainty in calibration results caused by the signal noise. As an example, based on Figure 4-10, for $N=2^{17}$ encoder counts or $N/4=2^{15}$ encoder sine waves per revolution, the uncertainty due to noise will be 10 nano-radians for a SNR of 3000.

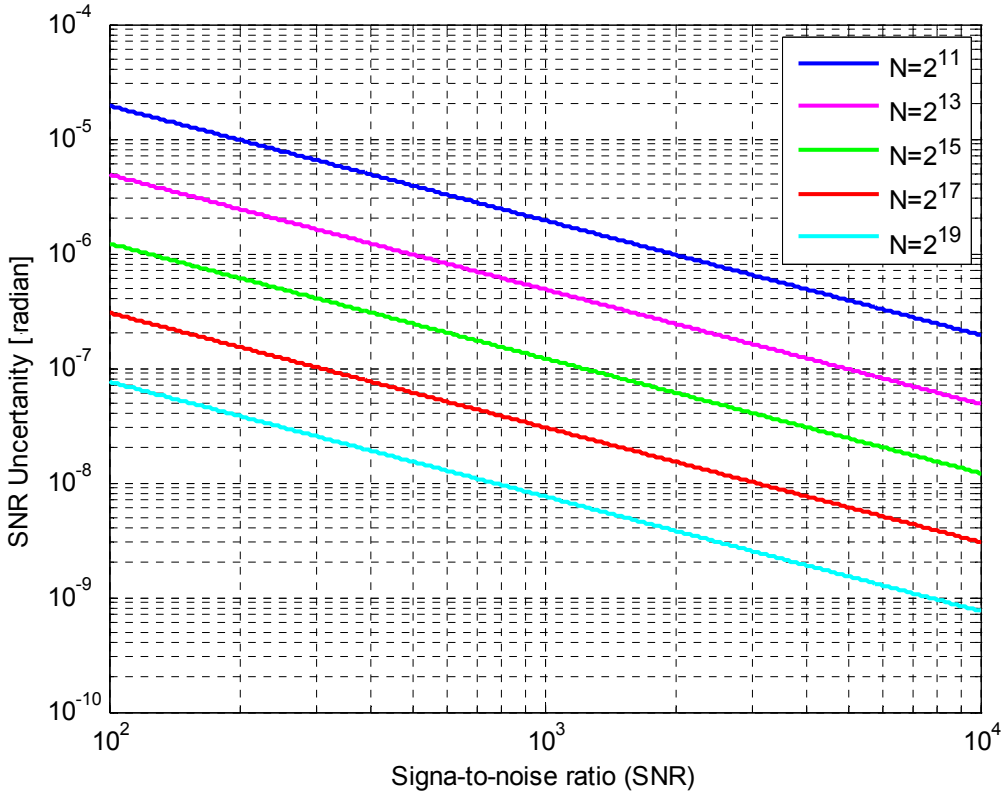

Figure 4-10: Uncertainty due to signal noise versus signal to noise ratio (SNR) for different number of encoder sine waves per revolution (N).

4.2.3 Spindle Radial Error Motion

The airbearing, like any other bearing, will have some error motion as it rotates. Any error motion of the rotor will affect the encoder's angle reading. The portion of the error motion which is repeatable with spindle position will result in repeatable error which can be removed by the calibration algorithm. However, the non-repeatable portion of the error motion will cause random measurement error which cannot be removed. Random rotor vibrations can occur due to random disturbance forces on the rotor, thermal effects, and transmitted stator vibrations. The angle measurement variation due to the asynchronous radial error motion is considered as another major source of calibration uncertainty.

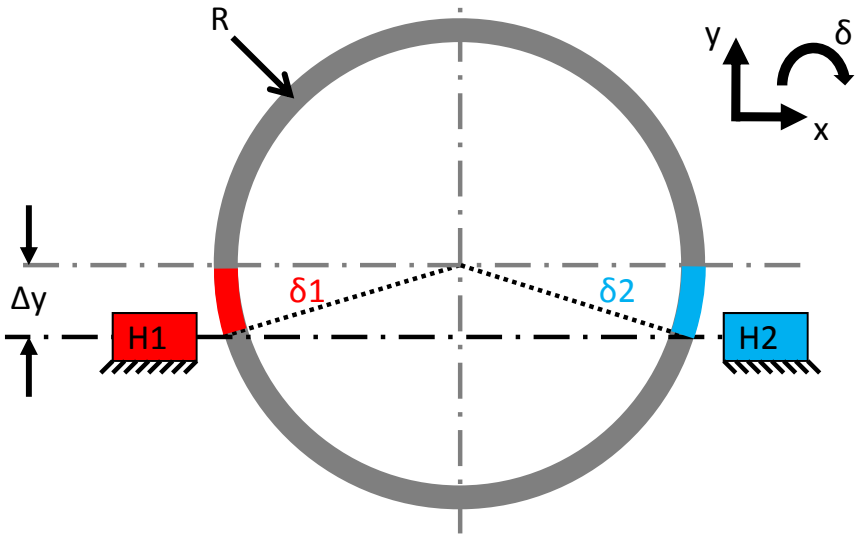

Figure 4-11: Angle measurement error due to lateral error motion

By using a second encoder read-head (H2), which is placed exactly opposite to the first read-head (H1) as shown in figure 4-11, the error due to radial motion can be removed. Figure 4-11 shows the encoder ring, in gray, moved in its plane perpendicular to the read heads by a distance Δy. Moving by Δy in the y-direction, the encoder gratings

on the ring will pass the read heads. This appears to the read-heads as a rotation. The incorrect angle measured be measured as

$$\delta_1 = -\delta_2 = \Delta y/\mathrm{R}\,, \tag{4.42}$$

where δ_1 and δ_2 are the angles measured by read-heads 1 and 2 respectively. It can be seen that the output angle of each read-head is directly influenced by the rotor vibrations. However, because the measuring error is opposite for the two read-heads facing each other, the effect of the lateral vibrations can be simply removed by using the sum of the signals from the two read heads. This is termed as lateral rotor vibration reversal.

The alignment of the opposing read-heads is important. The two read-heads must be separated by a mechanical angle of 180-degree. In reality the two heads will not be on exact opposite sides. Any deviation, α, from 180-deg results in the two heads' response to a lateral movements being different. Figure 4-12 shows an encoder ring movement Δy in front of two read-heads which are misaligned by α. The variations in the summation of the two signals due to this can be derived as

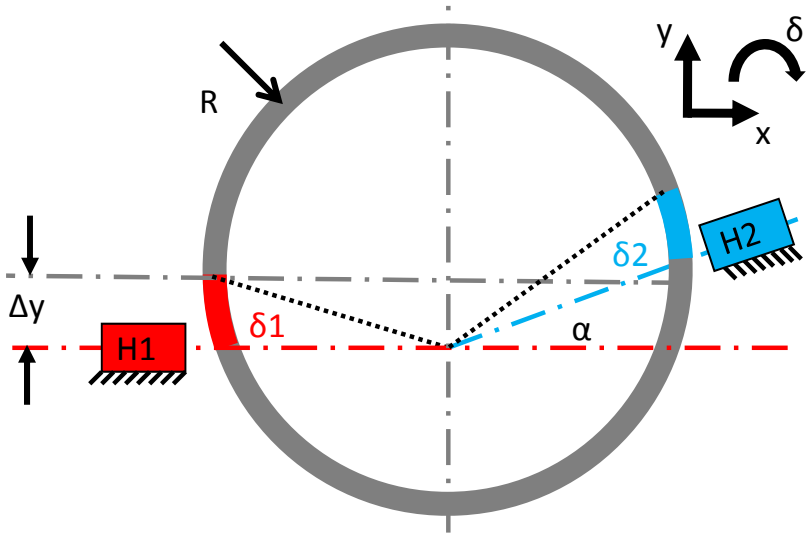

Figure 4-12: Mechanical angle misalignment of multiple read-heads used for lateral rotor vibration reversal

$$\begin{cases} \delta_1 = \Delta y/\mathrm{R} \\ \delta_2 = -\cos(\alpha)\,\Delta y/\mathrm{R} \end{cases}$$

$$\Longrightarrow \delta_1 + \delta_2 = \frac{\Delta y}{\mathrm{R}}[1 - \cos(\alpha)]$$

$$\Longrightarrow \delta_1 + \delta_2 \approx \frac{\Delta y}{\mathrm{R}}\left[\frac{\alpha^2}{2}\right]. \tag{4.43}$$

According to equation 4.43, small misalignment angles will have a second order contribution to the effectiveness of the rotor lateral vibration reversal. Therefore, high mechanical positioning accuracy is not required.

For lateral rotor vibration reversal to work, the raw, sinusoidal, encoder signals from the two heads must be added. In order for the two signals to add up constructively, their electrical phase angle must be close. The two read-head signals can be expressed as

$$\begin{cases} \mathrm{V_{H1}} = \mathrm{A_1}\sin\theta_1 \\ \mathrm{V_{H2}} = \mathrm{A_2}\sin\theta_2 \end{cases}, \tag{4.44}$$

where θ_1 and θ_2 are the electrical angle of the read-heads 1 and 2 respectively. The electrical angles can be represented in terms of the rotor's mechanical angle as

$$\begin{cases} \theta_1 = \dfrac{\mathrm{N}}{4}(\theta_\mathrm{m}) \\ \theta_2 = \dfrac{\mathrm{N}}{4}(\theta_\mathrm{m} + \pi + \alpha) \end{cases}, \tag{4.45}$$

where N is four times the number of sine waves per revolution, θ_m is the mechanical rotation angle, and α is the misalignment in mechanical angle units. Combining equations 4.44 and 4.45, the sum of the two encoder signals can be expressed

$$\mathrm{V_{H1}} + \mathrm{V_{H2}} = 2\mathrm{A}\sin\frac{\mathrm{N}}{4}\left(\theta_\mathrm{m} + \frac{\alpha}{2}\right)\cos\frac{\mathrm{N}}{4}\frac{\alpha}{2} \tag{4.46}$$

The first sine term in equation 4.46 represents the sinusoidal encoder signal and the second term expresses the reduction in amplitude due to an electrical angle misalignment

of $\beta = N\alpha/8$. Because of the large factor of $N/8$ being multiplied by the misalignment angle α, fine adjustment is required to ensure maximum amplitude of sum. The electrical misalignment is important because lower amplitude of sum cause lower SNR and higher uncertainty due to noise.

There are two general methods for performing the addition. The raw encoder signals can be digitized using the analog-to-digital-converters (ADC) and the summation can be performed in the digital domain. However, because the sampling frequency of the ADC is lower than the digital electronics, this method will reduce time resolution and will increase the calibration uncertainties. As an alternative, operational-amplifiers can be used to create an analog adder. This solution can conserve the raw signal's timing information and will not increase calibration uncertainties. In any case the, the two signals must be adjusted to have the same amplitude otherwise the reversal will not be fully effective. Our signal averaging method is described in section 3.4.1.

4.2.4 Stator Vibrations

The reference angle used for calibration with this method is based on the system dynamics and is relative to the inertial frame. However, the read-head measurements are relative to the stator frame. For the stator frame to be the same as the inertial frame, it must be free of vibrations. Any random vibrations on the stator frame will result in random variations in the read-head measurements which cannot be calibrated.

A typical configuration of the mechanical frames is shown in Figure 4-13. The soft vibration mounts support a large piece of granite to minimize the ground vibrations transmitted to the stator-frame. In this configuration, the floor vibrations will not be

transmitted to the stator as long as they do not have any components below the lower frequency limit of the vibration isolation system. The drawback of this configuration is that it is too sensitive to force excitations on the stator or the rotor.

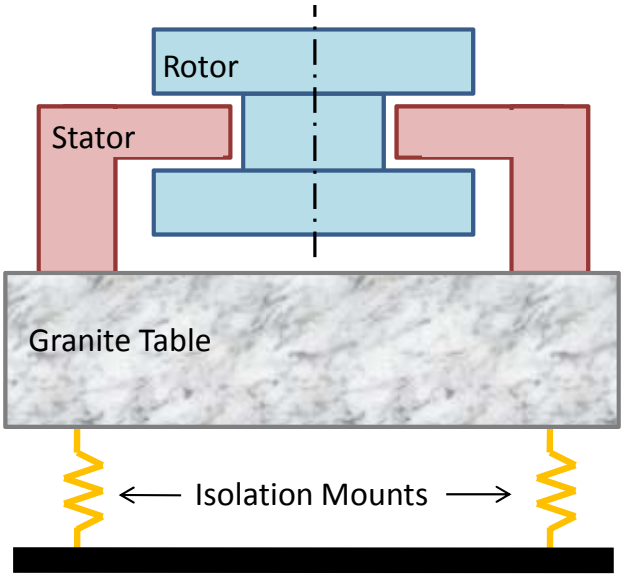

Figure 4-13: Schematic view showing the rotor and the stator frames and the vibration isolation system

4.2.5 Spindle Dynamics

The calibration algorithm used in this thesis assumes an inertia element and a first order damping element. The differences between the actual system's dynamics and this model will introduce error into the calibration results. Although airbearings have excellent linear characteristics due to averaging effect of the thin film, their dynamic model could still exhibit minor non-idealities which are not modeled by the calibration algorithm. External sources, such as the eddy current drag from a motor and the gravitational force on static unbalances, can cause deviations in system dynamic.

The discrepancies between the dynamic model and the system dynamics can be considered as another source of calibration uncertainty. However, the calibration algorithm's dynamic model can be modified to accommodate other systems by considering higher order dynamics.

4.3 Angular Encoder Selection

Many types of rotary encoders are available. The calibration algorithm can be applied to any type of rotary encoder as long as the encoder output can be transformed into a series of spatially distributed pulses. However, the encoder's type and performance directly affects the achievable rotation measurement accuracy. In this section, different types of rotary encoders will be presented; the available high-performance optical encoders will be overviewed; guidelines for encoder selection and sizing will be discussed; finally, the two selected encoders will be presented.

4.3.1 Encoder Types

The rotary encoders can be classified based on their sensing technology into three broad categories: optical and magnetic. Optical encoders are the most widely used type and work by detecting the variations in light using a photodiode. Optical encoders consist of a ring with gratings and a read-head with a photodiode sensor. As the ring rotates relative to the read-head, repeating patterns are created at the read-head which are then sensed by the read-head as a rotation measure. Optical sensing is less sensitive to electromagnetic noise and can achieve high precision; however, it is sensitive to dirt or anything disturbing the optical path. Magnetic encoders work by detecting the changes in the magnetic field using a hall-effect sensor. Magnetic encoders are relatively more sensitive to electromagnetic noise; however, they are less sensitive to dirt or oil, so they can be used in dirtier environments.

Rotary encoders can also be classified based on their bearing configuration. For any rotary encoder, the rotating ring and the stationary read-head need to be constrained

using a rotary bearing. There are encoders with integrated rotary bearing and ones without integrated bearings. Encoders with integrated bearings are less sensitive to mounting misalignment but will introduce friction forces and other uncertainties depending on the bearing characteristics. Encoders without integral bearing do not affect the system dynamics but are very sensitive to mounting misalignments.

4.3.2 Overview of High-Precision Optical Encoders

We focus in optical encoders as they have higher precision. Optical encoders sense the rotation by converting the variations in light using a photo-sensitive diode. In general, two methods are used by the high-end encoders to generate the variations in light: imaging method and interferential method [17].

A typical implementation of the imaging method is shown in Figure 4-14. Inside the encoder read-head, the LED light source is converted to parallel light by the condenser lens. The parallel light passes through the scanning reticle's index gratings and exits the encoder read-head. The scanning index gratings have the same pitch as the gratings of the scale and convert the parallel light into a series of dark and light stripes. The striped light hits the surface of the encoder scale which has reflective gratings with the same pitch as the striped light. Depending on the relative position of the reticle and the scale, a portion of the striped light will be reflected back through the window of the head and will shine on the photovoltaic array. As the scale moves relative to the read-head, the light intensity received at the photovoltaic array is modulated. The photovoltaic array finds a measure of the relative position of the scale and the read-head by measuring the variations in light intensity. The read-head reticle's index gratings have a specific

structure which filters the light to create a sinusoidal modulation of one sine per grating. [17]

The interferential scanning method uses the diffraction and interference of light on a fine graduation to produce modulated light signals. Figure 4-15 shows a typical implementation of the interferential method. A point light source is converted to parallel light using a condenser lens. The parallel light passes a reticle with transparent gratings. The reticle's gratings have the same pitch as the gratings on the scale. The light is diffracted by the reticle into three partial waves with same luminous and with orders of +1, 0, and -1. These waves are reflected back off the scale gratings with most of the intensity in the +1 and -1 orders. The reflected light is diffracted and interferes again at the reticle, and the resulting three waves leave the reticle and are received at the photovoltaic cells. A relative motion of one grating between the scale and reticle causes a phase-shift of one wavelength on the waves. The phase-shift will be positive for the +1 order and negative for the -1 order. As a result, the sum of the two signals will go through a phase shift of two wavelengths. The interferential method creates a sinusoidal variation of light intensity with two sine waves per one grating motion.[17]

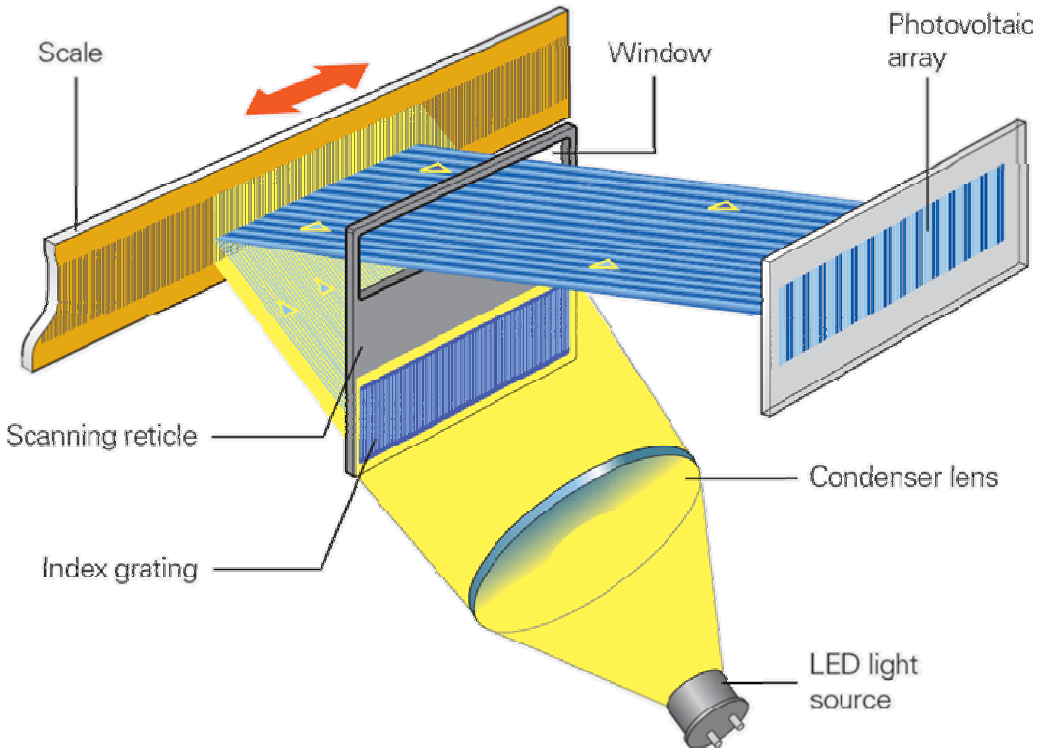

Figure 4-14: Typical configuration of the imaging method for optical encoder sensing [courtesy of Heidenhain] [17]

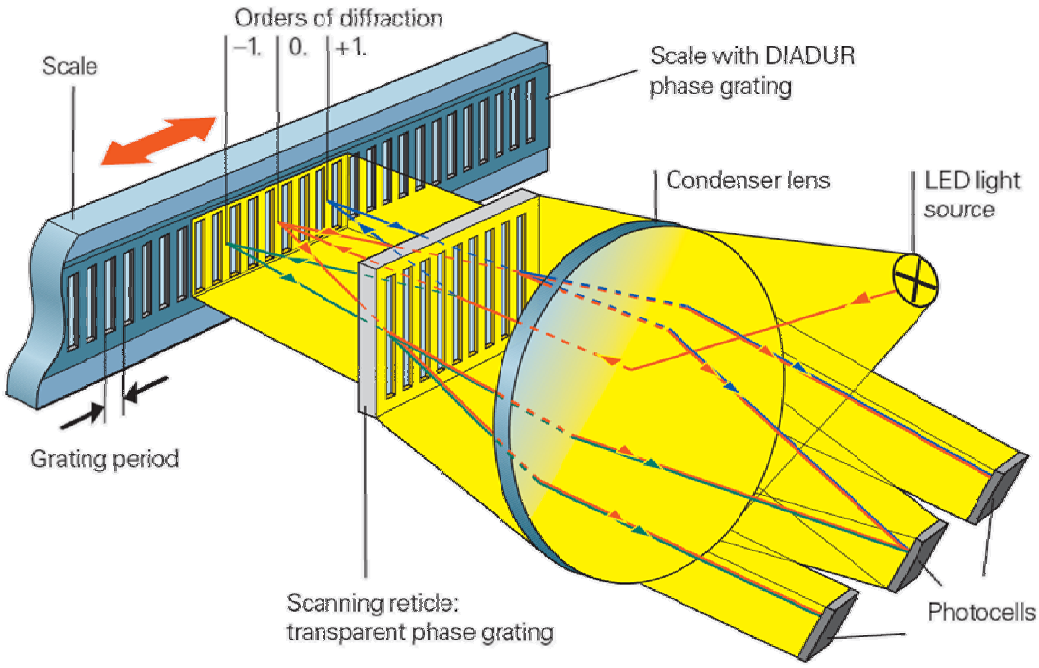

Figure 4-15: Typical configuration of the interferential method for optical encoder sensing [courtesy of Heidenhain] [17]

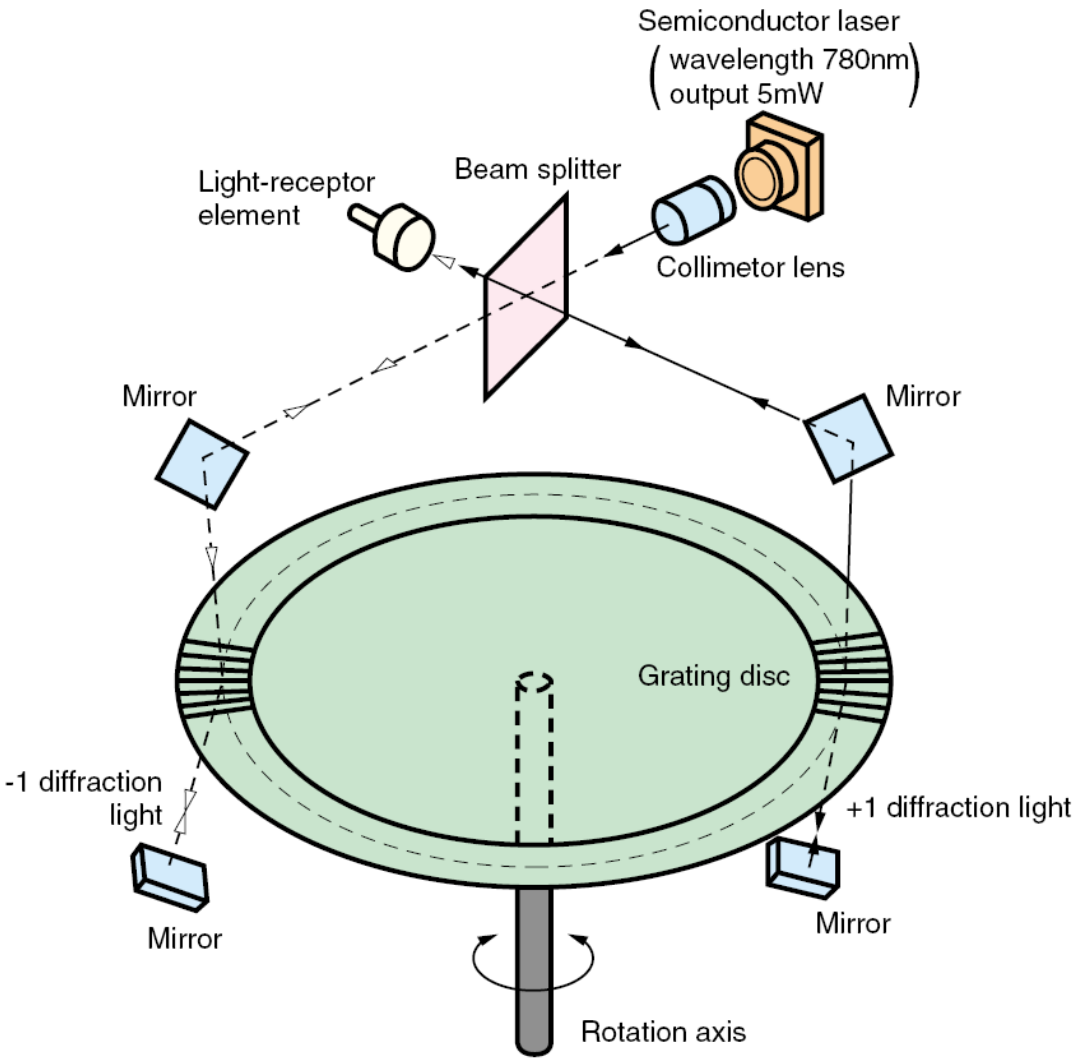

Figure 4-16: Implementation of the interferential method in Canon encoders [courtesy of Canon] [19]

The interferential method utilized by Heidenhain generates two sine periods per grating of movement. However, Canon and Sony's interferential methods can obtain four sine periods per grating. As shown in Figure 4-16, Canon uses a second set of mirrors to reflect back the beam so that each partial wave is applied to the scale gratings twice. In this configuration, the waves shift twice and each will go through a phase shift of two

wavelengths. Because of the opposite order of the two waves, their sum will go through a phase shift of four wave lengths per grating [19].

The imaging method can be used with grating pitch as small as 10 micro-meters, before the mounting tolerances become impractical for the encoder to be used. However, the interferential method has more generous alignment tolerances which enable it to be used with 4-micron or below grating width. Smaller grating pitch enables higher resolution as more encoder counts can be generated per revolution. The interferential method's output signal are naturally sinusoidal whereas the imaging encoders use the structure of the optics to establish a sinusoidal modulation. As a result, the output of the interferential method is a cleaner sinusoid and can be interpolated more accurately compared to the imaging method.[17]

Heidenhain, Canon, Sony, Renishaw, and MicroE are companies which make high-precision optical encoders. The specifications for some of the major encoder products from these companies are summarized in Table 4-1. Absolute accuracy is defined as the accumulative variations in the encoder count widths at the encoder counts. Internal accuracy includes the error associated within the sine wave of each encoder count. System accuracy is the sum of the internal and absolute accuracies and includes both the count location and shape error.

Table 4-1: Overview of high-precision optical encoder products

Manufacturer	Model	# of Gratings	Sines / grating	Sines / rev	Grating					Bandwidth (sines/msec)		BW Speed (RPM)		Disk Diameter (mm)			# of Read Heads	Detection Method
					width		accuracy (")											
					(")	(µm)	absolute	internal	system	3dB	6dB	3dB	6dB	inside	sensor	outside		
Heidenhain	ERP880	90000	2	180000	14.4	4	0.9''	0.1	1	800	1300	267	433	0	126	173	1	interferential
	ERA4282C	12000	1	12000	108.0	20	4	1.1	5.1	350		1750		40	77	77	flex	Imaging
		32768	1	32768	39.6	20	1.9	0.4	2.3			641		150	209	209	flex	Imaging
		52000	1	52000	24.9	20	1.7	0.3	2			404		270	331	331	flex	Imaging
Canon	X-1M	56250	4	225000	23.0	4			1	675		180		51		100	2	interferential
Sony	BH20	37800	4	151200	34.3	1				1400		556		9	12	14	flex	interferential
		59400	4	237600	21.8							354		16	19	21		
		75600	4	302400	17.1							278		21	24	27		
		170100	4	680400	7.6							123		51	54	57		
		226800	4	907200	5.7							93		69	72	75		
Renishaw	SiGNUM RESM	8192	1	8192	158	20	4.0	0.3	4.3	625		4578		30	50	50	flex	
		36000	1	36000	36.0		0.9	0.1	1.0			1042		209	229	229		
		86400	1	86400	15.0		0.4	0.0	0.4			434		510	550	550		
	RESR	8192	1	8192	158	20	4.0	1.6	5.6	225		1652		30	52	52	flex	
		11840	1	11840	109		2.8	1.1	3.9			1146		55	75	75		
		86400	1	86400	15.0		0.4	0.2	0.6			156		510	550	550		
Micro E	M1000 & M1200	16384	1	16384	79.1	20			2.1	360		1318		25	104	108	flex	

4.3.3 Guide Lines for Encoder Sizing

Although larger number of encoder counts per revolution provides more angular resolution, the best encoder size is not simply the largest size. Too many counts results in reaching the encoder bandwidth and reducing the signal amplitude and SNR. The SNR of an encoder as a function of speed changes with the number of encoder counts. Assuming a simple first order model for the encoder's sensing stage, the SNR variations can be expressed as,

$$SNR(n) = \frac{1}{\sqrt{1 + \left(\frac{N/4 \cdot n/60}{f_{3dB}}\right)^2}} SNR_0 \tag{4.47}$$

where n is speed in rotations per minute, SNR_0 is the signal to noise ratio at zero speed, $N/4$ is the number of encoder sine waves per revolution, and f_{3dB} is the sensor's -3dB bandwidth. Plugging equation 4.47 in equation 4.41, the noise accuracy limit can be expressed as

$$d\theta(n) = \frac{\sqrt{1 + \left(\frac{N/4 \cdot n/60}{f_{3dB}}\right)^2}}{SNR_0} \cdot \frac{4}{N} \tag{4.48}$$

Equation 4.48 can be expressed approximated by piece wise function as

$$d\theta(n) = \begin{cases} \frac{1}{SNR_0} \cdot \frac{4}{N} & n < \frac{240 \cdot f_{3dB}}{N} \\ \frac{1}{SNR_0} \cdot \frac{n}{60} \cdot \frac{1}{f_{3dB}} & n > \frac{240 \cdot f_{3dB}}{N} \end{cases} \tag{4.49}$$

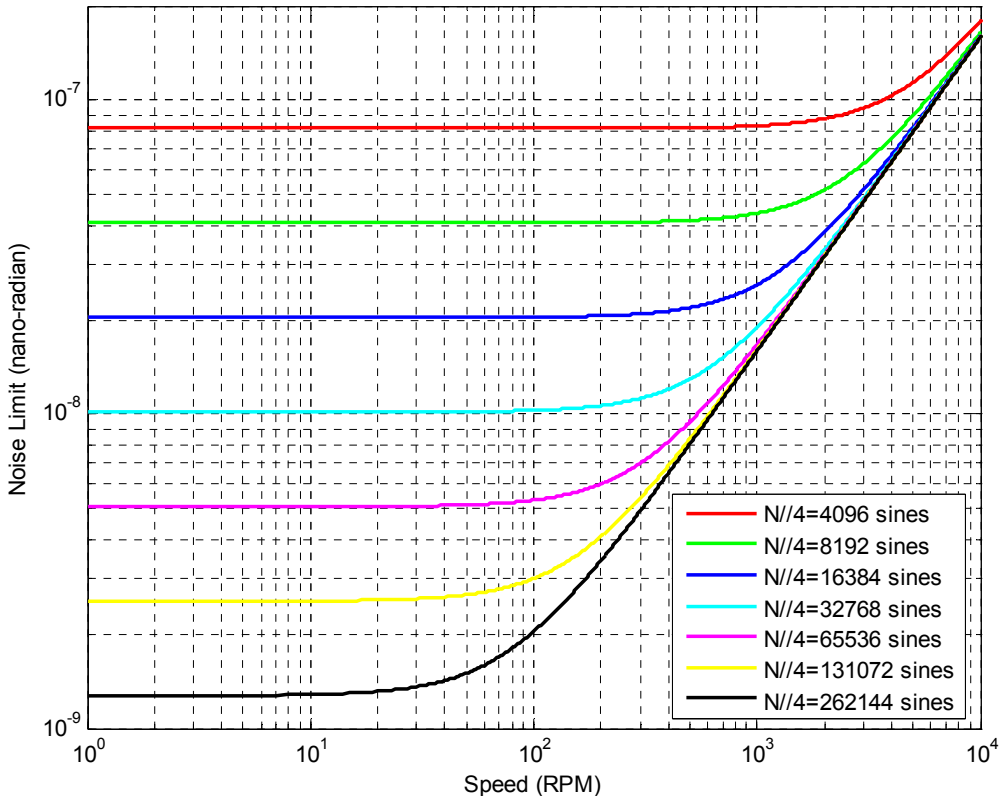

Figure 4-17: Accuracy limit from encoder noise vs. speed with encoder bandwidth of 350-kHz, SNR of 3000, and for different number of encoder sines per revolution (N/4), assuming the same f_{3dB} and SNR_0.

Figure 4-17 shows the accuracy limit set by the encoder signal noise versus the speed of rotation. As expected from equation 4.49, the curves can be approximated by two linear segments. The first segment, which corresponds to before the bandwidth, is dependent on the number of encoder counts; however, the second segment for all encoder counts eventually merge and hence is not as dependent on N. Although increasing the encoder count has improved the encoder accuracy, the benefits are dependent on the operating speed range and the encoder bandwidth. The analysis performed here assumes a first order system, which is true to a certain limit. Perhaps, too far from the bandwidth of the encoder, the amplitude attenuates more rapidly than a first order system and the accuracy will degrade much faster.

Accuracy will not increase indefinitely with the number of the encoder counts. Past the encoder bandwidth, accuracy starts to become independent of the number of encoder counts. In addition, the number of counts needs to be limited to ensure that the encoder is not used too far past its bandwidth where it may start to exhibit faster drops in signal quality. Finally, larger number of counts requires a large encoder scale. Increasing the scale size will reduce its rigidity, and at the same time, increases the centrifugal forces on it. As a result, the encoder becomes more sensitive to the operating conditions and loses consistency. In summary, the minimum encoder size which meets the noise accuracy requirement is perhaps the best solution as it provides maximum rigidity, and maximum signal consistency.

4.3.4 Selected Encoders

Considering the guidelines presented so far, two encoders were selected: ERA4282C and RON905 both from Heidenhain. ERA4282C, shown in Figure 4-18 , is an imaging optical encoder without an integral bearing and generates 32768 sine waves per revolution. RON905, shown in Figure 4-19, is an imaging optical encoder with an integral bearing and generates 36000 sine waves per revolution.

Figure 4-18: ERA4282C encoder from Heidenhain [courtesy of Heidenhain] [17]

Figure 4-19: RON905 encoder from Heidenhain [courtesy of Heidenhain] [5]

The ERA4282C encoder does not have an integral, so its read-head and scale are completely separate parts. The scale is a very rigid steel drum with vertical gratings on its outside surface. Because of the large axial length of the encoder gratings, the encoder read-head is not sensitive to axial misalignment or axial motion. Also, the high rigidity of the encoder drum guarantees minimal deformation in operation. Because the ERA4282C's read-heads are separate parts, different read-head configurations can be tested and all the read head signals will be accessible. The ERA4282C encoder output signals are 2 sine waves each at 1-V peak-to-valley and with 32768 periods per revolution. It has a 3dB bandwidth of above 350-kHz, and with 32768 gratings, it reaches its 3dB bandwidth at 640-RPM. The ERA44282C is selected to be calibrated as the final rotation sensor for the operating speed range of 0 to 1000 RPM.

The RON905 encoder has an integral bearing and comes in one part. The integral bearing can correct misalignments up to a certain extent; therefore, the RON905's

accuracy is not as sensitive to the mounting misalignments. However, to avoid damaging the bearing, tight mounting tolerances must be met and the rotor must be always operated below 100-RPM. The RON905's output signals are 2 sine waves each at 11-μA peak-to-valley and with 36000 periods per revolution. It has a 3dB bandwidth of above 40-kHz, so it reaches its bandwidth at 67-RPM. The RON905 is selected to test the effectiveness of our calibration algorithm on encoders with integral bearings. It is also selected as a way to cross check our calibration results with results from other institutes.

4.4 Simulation Work

Simulation is used to confirm the effectiveness of the calibration algorithm and to evaluate its performance. In simulation, the mechanical plant, the encoder, and the electronics are replaced by dynamic equations. In this way the calibration algorithm can be tested by itself without any uncertainties from the setup. In this section, the equations and the parameters used for simulation will be presented; the simulation results using the second order method will be presented; simulation result using second order integral method will be displayed; and, finally the repeatability and accuracy for both methods will be captured together as a reference for the work in the experimental section.

4.4.1 Simulation Method

A MATLAB m-file script is used to generate simulation data similar to the data that is generated from the experimental setup. The mechanical setup is replaced by the equations of motion for an inertial element and a first order damping element. The differential equation for the system can be expressed as

$$\frac{d\omega}{dt} + \big(c_0 + c_1(\omega - \omega_0)\big)\omega = 0 \tag{4.50}$$

The differential equation above can be solved to find the angular speed, ω, as a function of time. The angular speed can be integrated versus time to find a relation between angular position, θ, and time as

$$t = \frac{1}{\mathrm{c}_1\omega_0 - c_0} \cdot \ln \frac{e^{c_1\theta}(\omega_0 - c_0/c_1) + c_0/c_1}{\omega_0} \tag{4.51}$$

The damping parameters used for simulation are taken from the experimental data and can be represented, using the convention of the encoder calibration section, as

$$\begin{cases} a_s = -3.4902 \times 10^{-11} \\ b_s = -1.8600 \times 10^{-18} \end{cases} \tag{4.52}$$

The parameters c_0 and c_1 can be calculated based on the values a_s and b_s as

$$\begin{cases} c_0 = -\dfrac{a}{\Delta_0} = 3.49 \times 10^{-11} & \left[\dfrac{1}{timer\ count}\right] \\ c_1 = \dfrac{2b}{c_0 {\Delta_0}^2} = -1.1 \times 10^{-7} & [encoder\ count] \end{cases} \tag{4.53}$$

where position is expressed in encoder counts (1/131072 of revolution) and time is expressed in digital timer counts (1/2e8 of a second).

Using equation 4.51, timing information is generated showing the absolute time at each encoder count. The spatial widths of the counts are specified based on an experimental error map shown in Figure 4-20, obtained from our experimental setup. The original timing information obtained from equation 4.51 has a very high resolution and is close to using a continuous clock. To better reflect our setup, the timing information is discretized based on several different clock frequencies. The generated timing information is similar to the experimental setup; however, it only reflects the timing uncertainty. Therefore, it can be used to isolate and test the calibration algorithm and its dependency on the timing uncertainties.

The simulation model is used to generate timing data over the 10 to 1000 RPM spindle speed range, using 10 RPM increments. At each speed, the timing data is stored for continuous as well as 100, 200, 400, and 800-MHz timers.

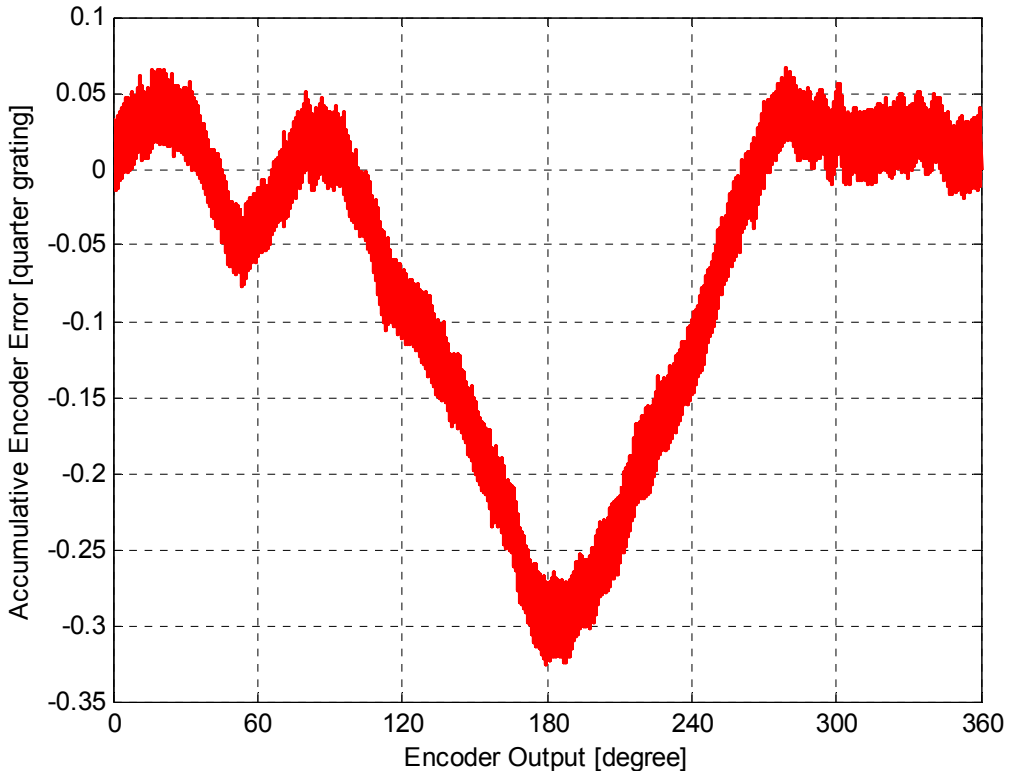

Figure 4-20: Reference encoder error map used for simulation

Our calibration algorithm can be applied to the stored simulation timing data to derive error maps. To quantitatively evaluate the calibration results, two measures are defined: accuracy and repeatability. Accuracy is defined as the root-mean-squared (RMS) variations between the reference and the calibration error maps and can be expressed as

$$\Lambda = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[E_r(i) - E_{c1}(i)]^2}\ , \tag{4.54}$$

where E_r is the reference error map, E_{c1} is the calibrated error map from the first revolution and the parameters a_1 and b_1, and N is the number of encoder counts. Repeatability is defined as the RMS variations between the pair of calibrated error maps and can be expressed as

$$R = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[E_{c2}(i) - E_{c1}(i)]^2}\ , \tag{4.55}$$

where E_{c1} is the error map based on the first revolution and the parameters a_1 and b_1, and E_{c2} is the error map based on the second revolution and the parameters a_2 and b_2.

4.4.2 Second Order Method in Simulation

The second order calibration method with S_1=0-degree and S_2=540-degree is applied to the simulated timing data to generate calibration maps for speeds from 10 to 1000 RPM with continuous as well as 100, 200, 400, and 800-MHz clocks. For each calibration, accuracy and repeatability are calculated according to equations 4.54 and 4.55 and are plotted in Figure 4-21and Figure 4-22 respectively. In each plot the spatial equivalent of the timing uncertainty is calculated using equation 4.36 and is shown using a dotted line for each timer speed.

Because no other uncertainties are modeled in the simulation, the accuracy and the repeatability of the calibration are expected to be around the timing limit. Although the accuracy and repeatability are roughly following such a trend, they are significantly worse than the time limit. As an example, the second order calibration result for 200-RPM speed and 100-MHz timer is shown and is compared to the reference error map in Figure 4-23. The deviations in the calibration result are plotted in green and consists of

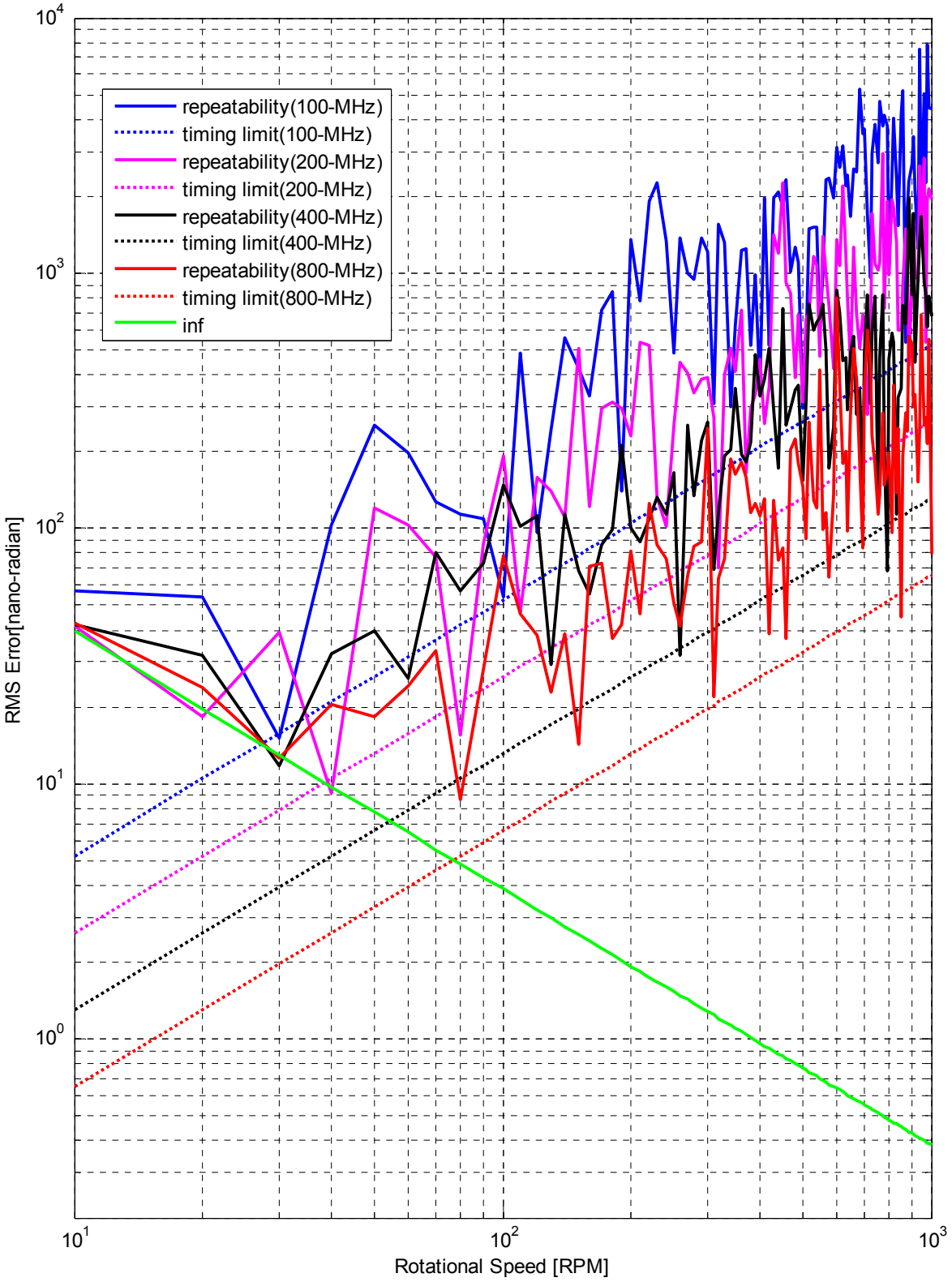

Figure 4-21: Accuracy plot of the second order method results in simulation for 10-1000 RPM and with continuous, 100, 200, 400, and 800 MHz timers

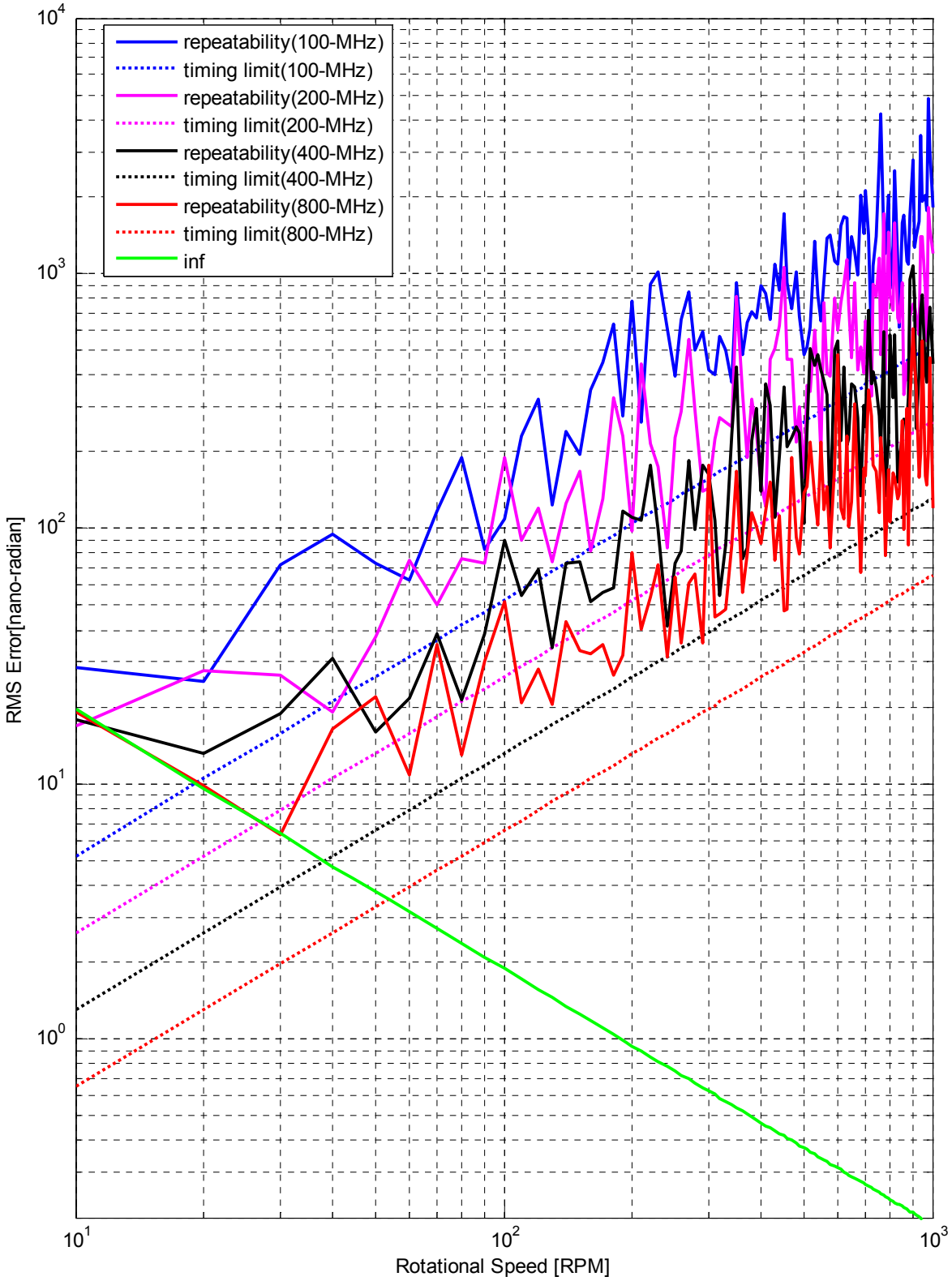

Figure 4-22: Repeatability plot of the second order method results in simulation for 10-1000 RPM and with continuous, 100, 200, 400, and 800 MHz timers

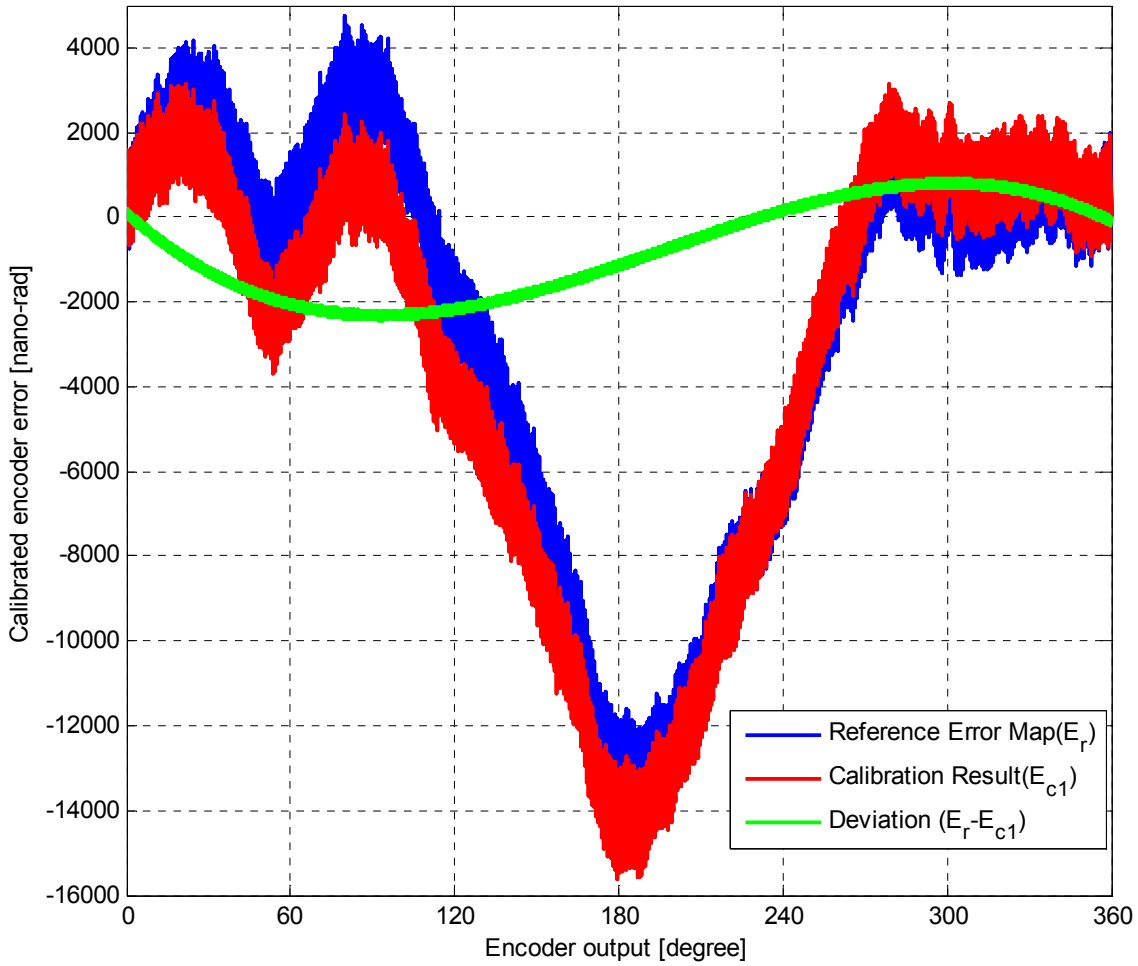

Figure 4-23: Calibration result of second order method at 200-RPM and with a 100-MHz timer

a low frequency component and a high frequency hash. The high-frequency hash is caused by timing uncertainties and is unavoidable, but the low frequency part is due to inaccurate system model estimation and can be avoided by using the integral method.

Looking at Figure 4-21 and Figure 4-22, it is observed that accuracy and repeatability stop following the trend of the time uncertainty limit for speeds below 50-RPM. Instead, they start following the trend indicated by the continuous timer results. The continuous timer does not see any timing uncertainty so its accuracy must be limited by another factor. This other limiting factor which decreases with speed is expected to be due to three approximations used in the second order calibration algorithm. The

enhancement method is required to improve the calibration performance at low angular speed operating region.

4.4.3 Second Order Integral Method in Simulation

The second order integral method is applied to the same collection of simulated timing data. The calibration accuracy and repeatability for the second order integral method is calculated and plotted for a speed range of 10-1000 RPM using continuous, 100, 200, 400, and 800-MHz timers. The accuracy and the repeatability plots are shown in Figure 4-24 and Figure 4-25 respectively. As can be seen, the results closely follow the limit set by the timing uncertainty limit and are much more consistent.

By more accurately estimating the system parameters, the integral estimation method has removed the large low frequency deviations between the reference and the calibrated error maps. The calibration result of the integral method is shown in Figure 4-26 for the same case as the example shown for the second order method in Figure 4-23. As can be seen, the low frequency deviations are reduced significantly to the same amplitude as the high frequency component caused by the limited timing resolution.

The integral method estimates the dynamics more accurately, and hence, significantly reduces the calibration error due to incorrect model estimates. As a result, the accuracy limits created by limited time resolution as well as the approximations in the algorithm can be seen more clearly. The timing limit is dominant at high-speed and increases with speed. The approximation limit is dominant at low-speed and increases with reductions in speed.

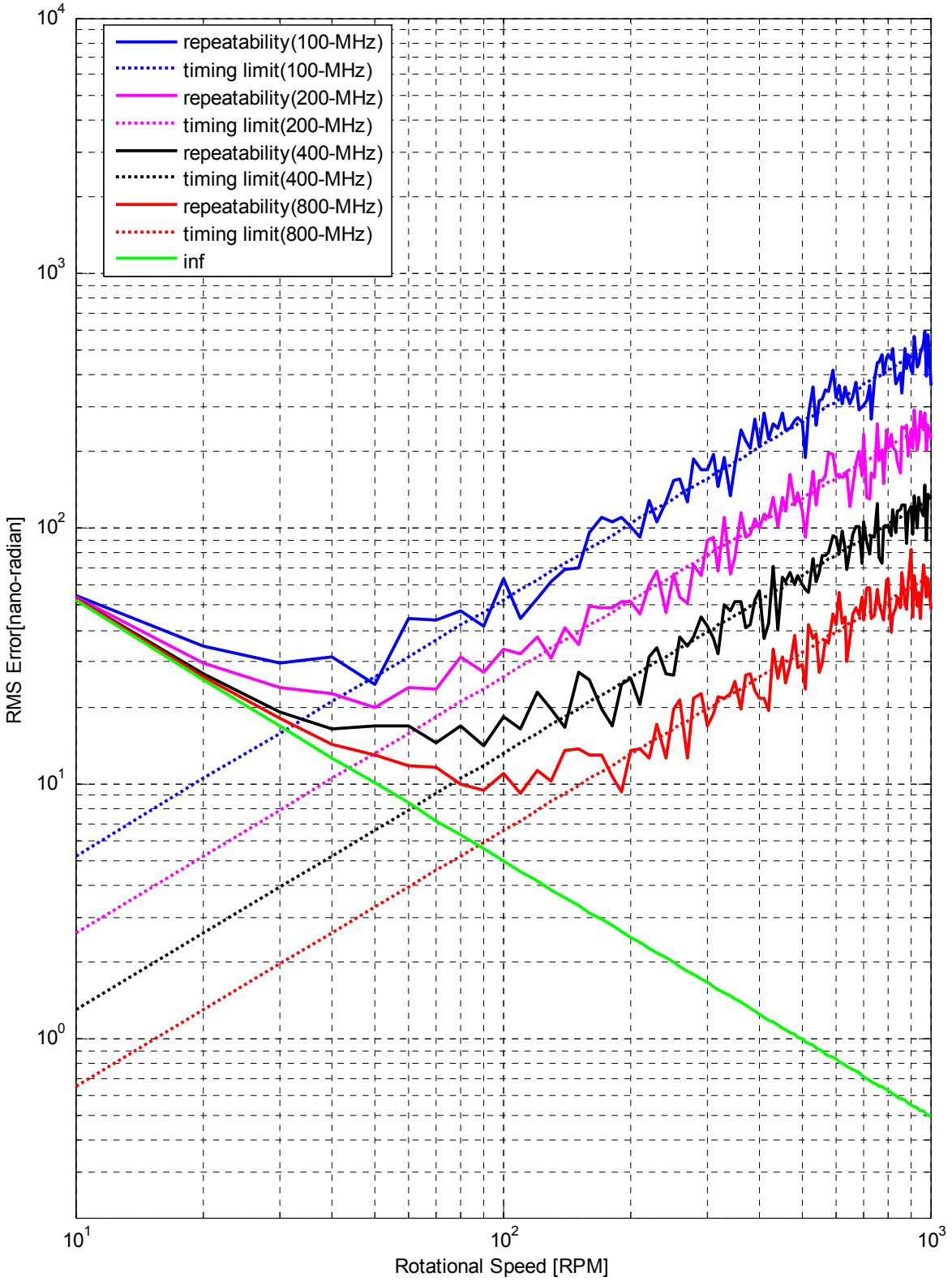

Figure 4-24: Accuracy plot of the second order integral method results in simulation for 10-1000 RPM and with continuous, 100, 200, 400, and 800 MHz timers

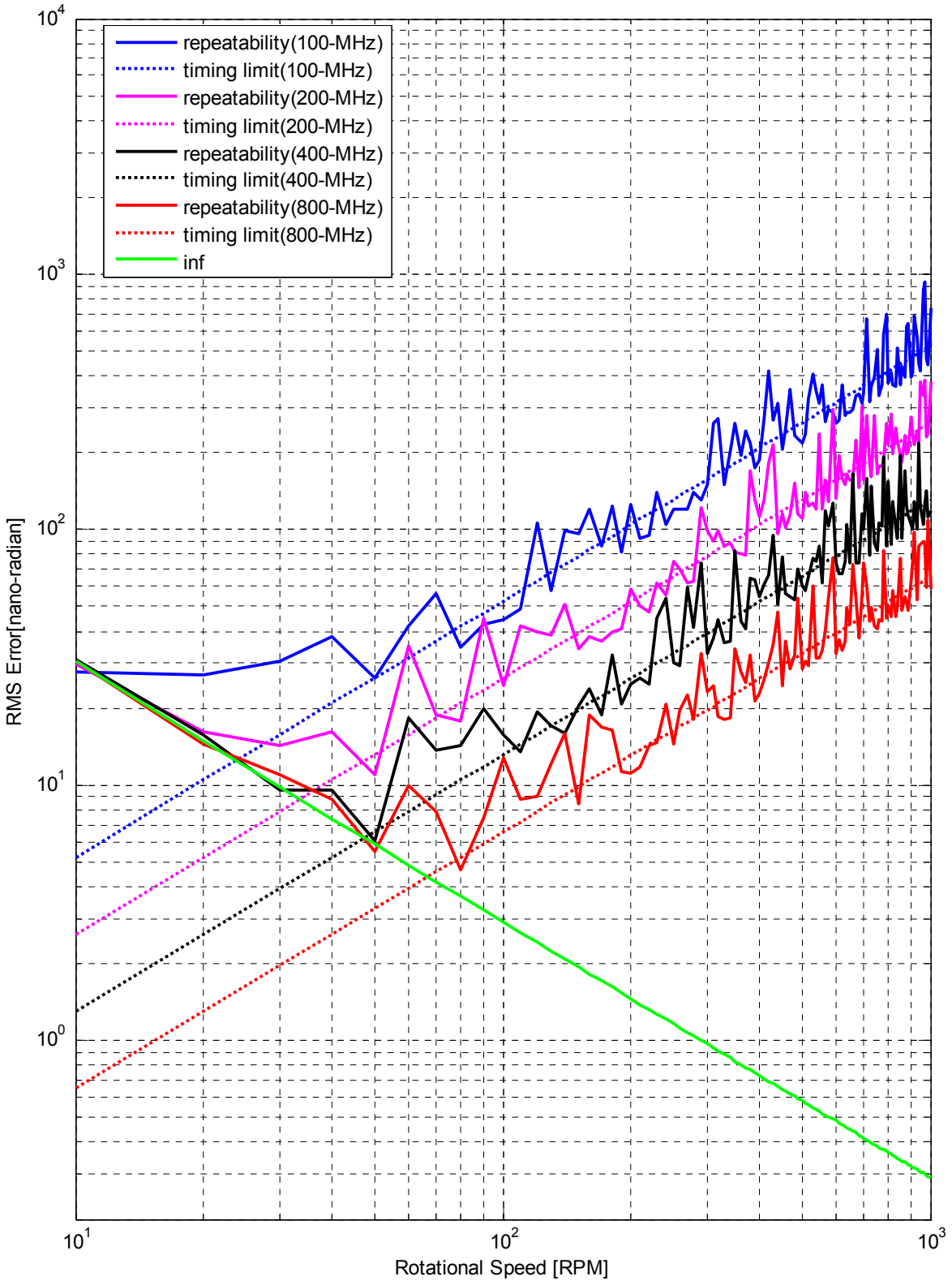

Figure 4-25: Repeatability plot of the second order integral method results in simulation for 10-1000 RPM and with continuous, 100, 200, 400, and 800 MHz timers

Figure 4-26: Calibration result of second order integral method at 200-RPM and with a 100-MHz timer

4.5 Experimental Work

The encoder calibration has been tested experimentally as a part of this thesis. In this section, the experimental setup is introduced; the associated experimental uncertainties are covered; the encoder calibration results are presented; and, the results are discussed in reference to the simulation results.

4.5.1 Experimental Setup

Figure 4-27 shows the experimental setup consisting of a rotary table mounted on a vibration isolation granite table, high-speed electronics, and the user interface. In this sub-section, each component and its role in the setup are overviewed.

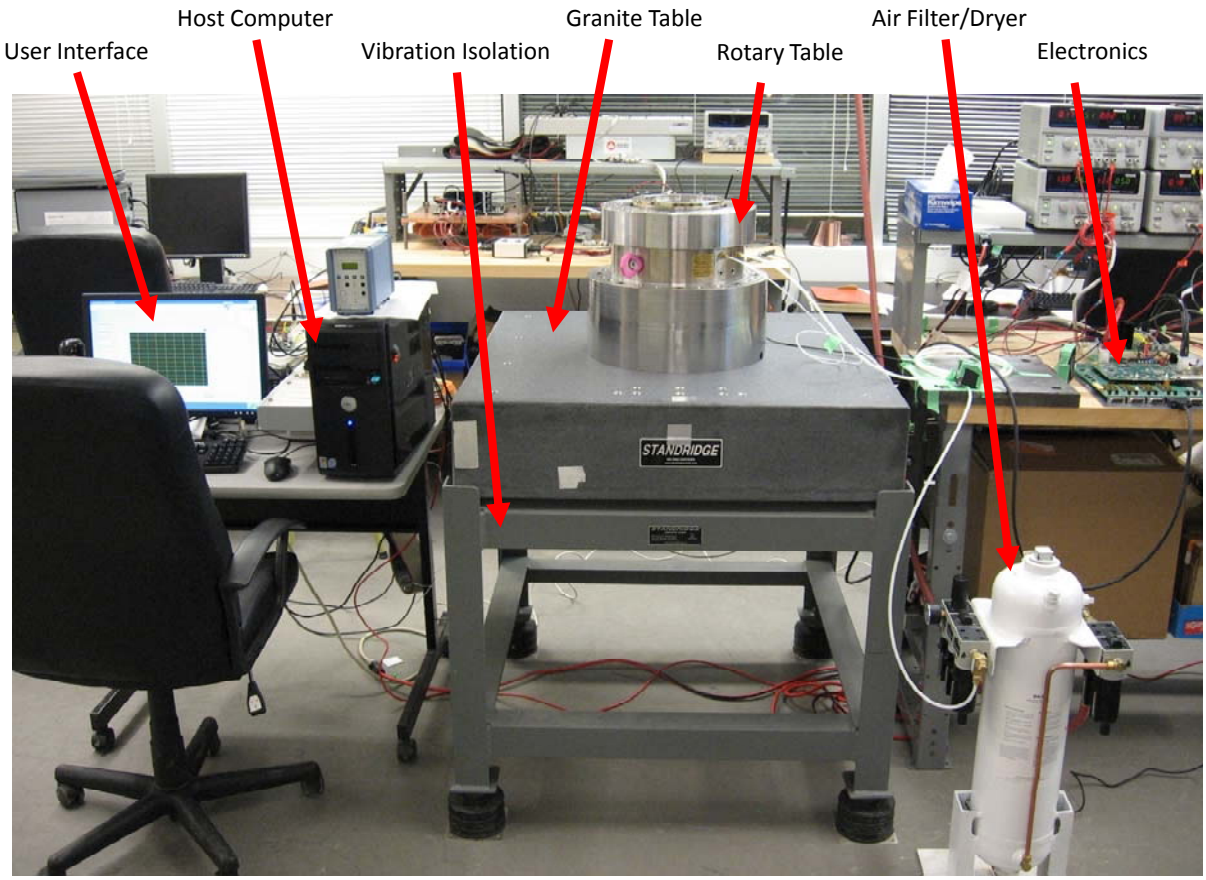

Figure 4-27: The experimental setup used for encoder calibration

The experimental procedure is pictured in Figure 4-28. The rotary table is accelerated to a certain speed and is left to slow down in free-response. The encoder pulses are processed and timed by the custom electronics. The time measurements are then received and stored by the host computer running a graphical user interface developed using Labview. The calibration algorithm, which is implemented using a series of MATLAB m-files, processes the recorded time measurements to generate the error map of the encoder.

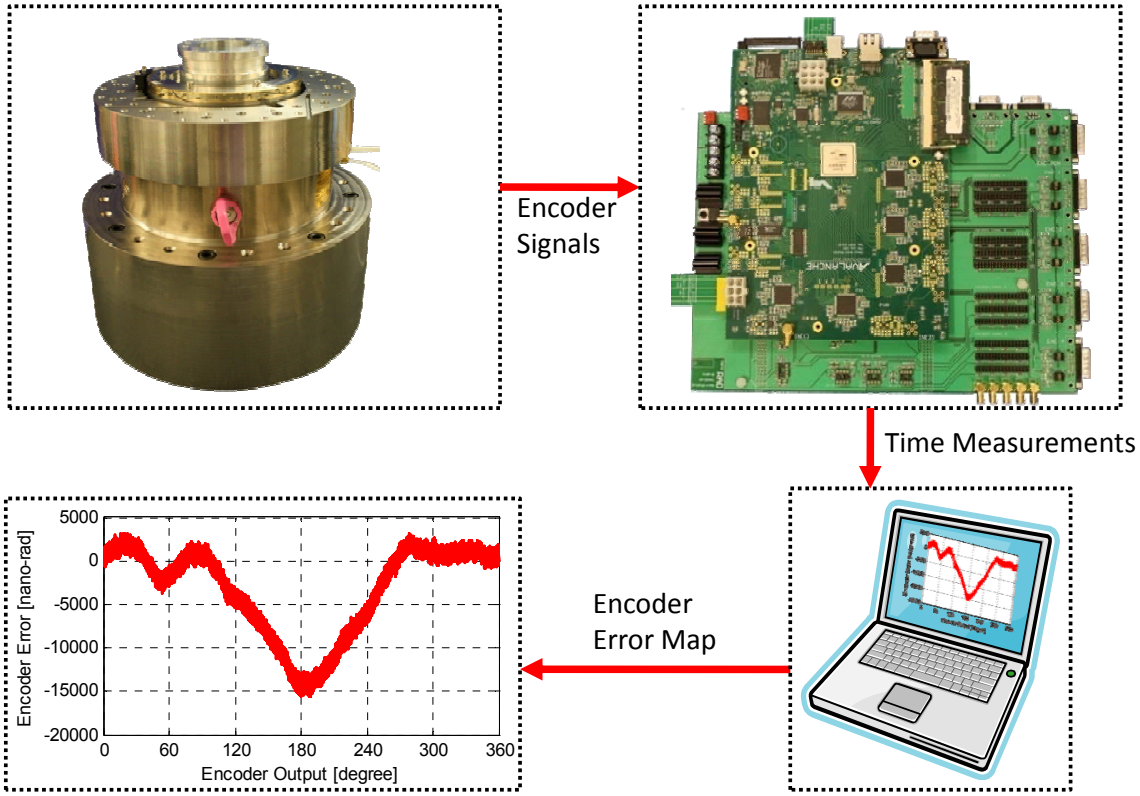

Figure 4-28: Overview of the experimental procedure

The rotary table is shown in Figure 4-29. We designed the table for precision rotary motion and linear dynamics. The rotation axis is created using a 10R Blockhead air-spindle manufactured by Professional Instruments. In this configuration, the rotation is measured using the Heidenhain ERA4282C encoder, which outputs analog sinusoidal

encoder signals. The rotary platform is mounted onto the granite table using a stator support piece. The rotary table's detailed design is covered in Chapter-3, *mechanical design* section.

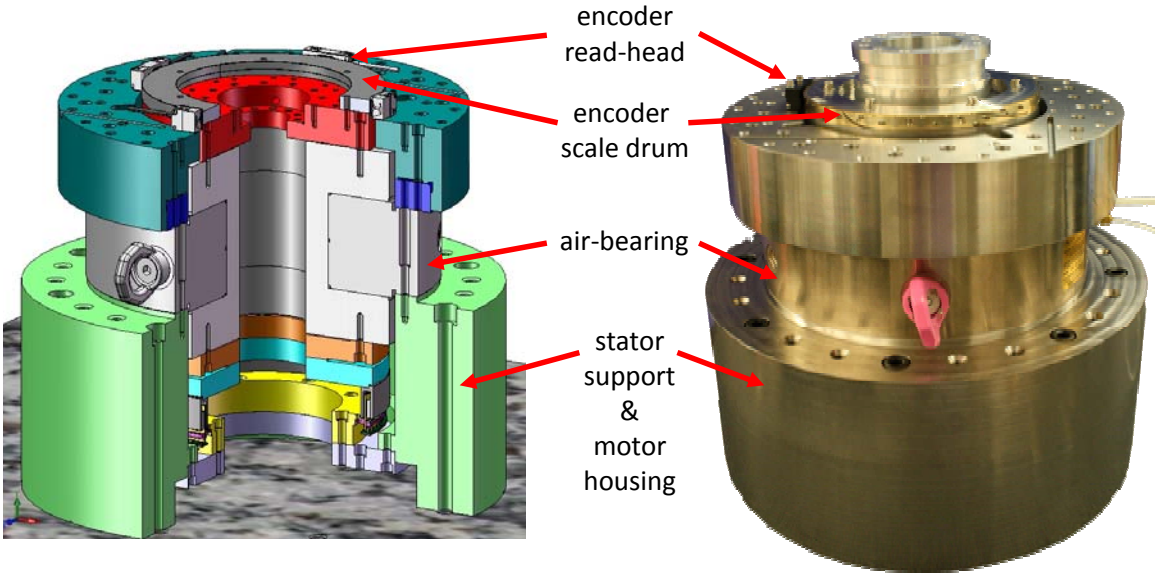

Figure 4-29: Ultra precise rotary table designed and built as a part of the experimental setup

The set of high-speed electronics which were designed and built as a part of this thesis are shown in Figure 4-30. The custom electronics process and time the encoder signals. The daughterboard (NanoRAD) digitizes the encoder signals and routes them to the motherboard's digital inputs. The motherboard (Avalanche) times the digitized encoder pulses and sends the recorded time measurements to the user interface over an Ethernet link. Details on the design and the numerous features of the electronics are provided in Chapter-4, *electrical design*. The time measurements are received and stored at the interface computer by a graphical user interface developed using Labview. The recorded time measurements are then processed using the calibration algorithm described in section 4.1 to obtain an error-map for the encoder.

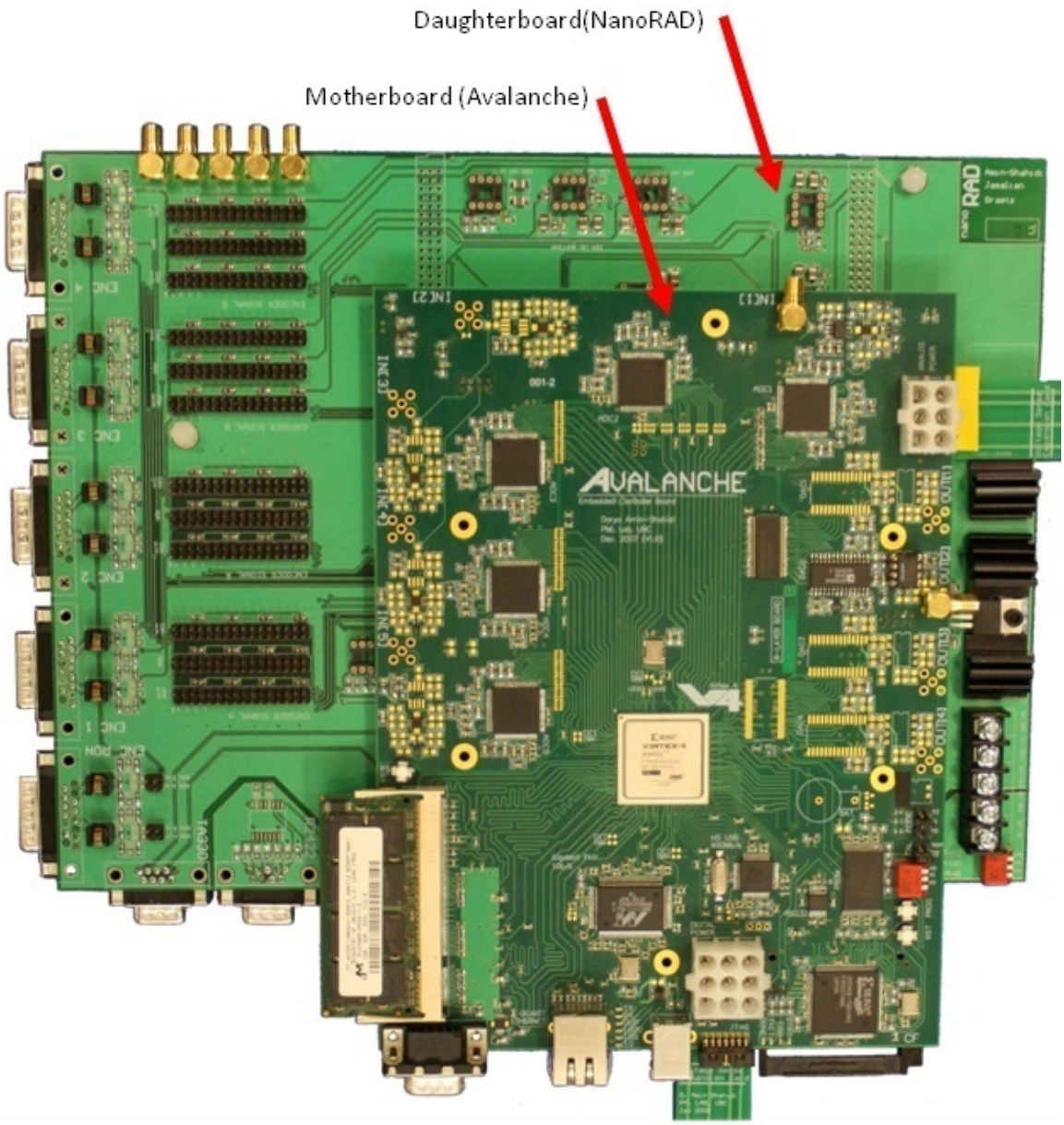

Figure 4-30: Set of electronics designed and built as a part of the experimental setup

There are uncertainties in the experimental data which limit the achievable encoder calibration accuracy. The major sources of uncertainties are indicated and quantified in this sub-section. The major limiting sources of uncertainty include signal noise, vibrations, pressure line fluctuations, and time discretization. The next four sub-sections describe and quantify the uncertainty from each of the uncertainty sources.

4.5.2 Uncertainty from Time Discretization

The electronics included in the experimental setup use a 200-MHz timer to time the encoder signals. As a result, the timing data have an uncertainty of ± 2.5 nano-seconds. As discussed in section 4.2.1, any uncertainty in the timing measurements will introduce uncertainty in the encoder calibration results. The resulting uncertainty can be calculated using equation 4.36. For example, at a speed of 200-RPM, the uncertainty is limited to ± 52 nano-radians.

4.5.3 Uncertainty from Signal Noise

The analog encoder signals contain random electrical noise. The electrical noise limits the repeatability of the spatial locations of the encoder counts, and hence, will reduce the achievable calibration accuracy. In section 4.2.2 it was analyzed how the encoder noise affects the encoder calibration accuracy. Here, we quantify the encoder signal noise existing in the available experimental setup. An encoder-read head, which is kept in a dark box and off the setup, is used to estimate the present electrical noise. Since the encoder read-head is off the setup, any variations on the signal are considered as electrical noise. Figure 4-31 shows time and frequency plots of electrical noise from a single encoder channel. The signal is based on one side of a specific encoder channel's differential pair and is captured after a common mode choke filter which attenuates the noise common to both sides of the pair.

The electrical noise displayed in Figure 4-31 has been captured pas a common mode choke, so does not have any common peak variations of 1 mV and its root-mean-squared (RMS) is calculated as 0.24 mV. However, because we compare the two sides of

the differential pair, the effective noise can be calculated as $0.24\sqrt{2} = 0.34$mV. Given the 500-mV amplitude of the differential encoder signal, the SNR for the encoder can be calculated as 1475. Using equation 4.41, the uncertainty contributions from the electrical noise can be estimated as 24.5 nano-radians RMS. The presented noise data is past a common mode choke and without any other filtering. It is expected that the signal quality can be improved by filtering out higher frequency noise.

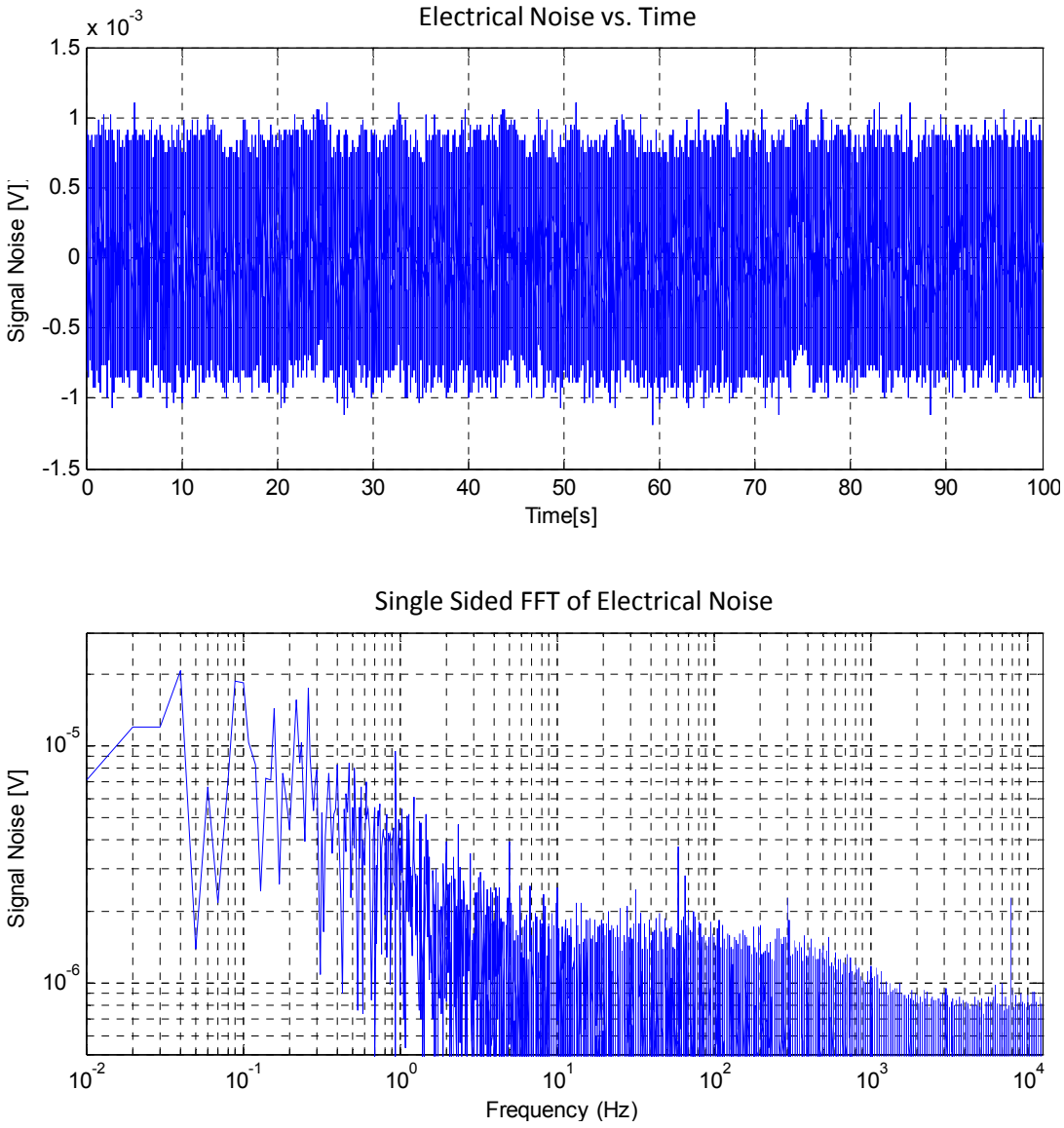

Figure 4-31: Time (top) and frequency (bottom) plots of read-head's positive side channel-A noise captured at 25-kHz.

4.5.4 Uncertainty from Vibrations

The encoder calibration method assumed that the read-heads are mounted on a perfect inertial frame free of vibrations. Therefore, vibrations would reduce calibration accuracy. The effect of vibrations was analyzed in sections 4.2.3 and 4.2.4; here, we quantify the presence and the effect of the vibrations in our experimental results. To capture the vibrations' effect on relative motion of the rotor and the stator, the encoder signal is recorded while the setup is left stationary without any intentional outside excitations. Figure 4-32 shows the encoder reading over 100-seconds with the air supply regulated from 115-psi source pressure down to 80-psi bearing pressure using two separate pressure regulators in series. The rotations due to vibrations are found to have a peak value of 200 nano-radians and an RMS value of 70 nano-radians. The frequency content of the vibration induced rotations shows major peaks at 2.5-Hz and 12-Hz.

These peaks are found to be related to the vibration modes of the granite table. A schematic representation of the table is shown in Figure 4-33. Experimental impact hammer tests, which have been performed by Richard Graetz[5], my colleague, confirm the natural frequency of the lateral (X or Y) at 2.5Hz and the yaw (rotation around Z) modes at 10Hz of the table matching the peaks observed in Figure 4-32. In addition, an analytical model of the table vibrations is developed, which further confirms our hypothesis. To simplify the analytical derivation, the table and the frame are modeled as a single rigid body; all the elements were considered mass-less except for the granite table top; and, the isolation mounts are considered as the only flexible element with a uniform stiffness k in all three X, Y, and Z directions. Due to symmetry, the vibration modes for

[5] Richard Graetz (rgraetz@interchange.ubc.ca), 2008, graduate student, Precision Mechatronics Laboratory, University of British Columbia.

translation and rotation about X and Y are the same; therefore, four distinct rigid body modes are expected for the granite table: axial mode (Z), yaw mode (θ_Z), lateral mode (X/Y), and pitch mode (θ_Y /θ_X).

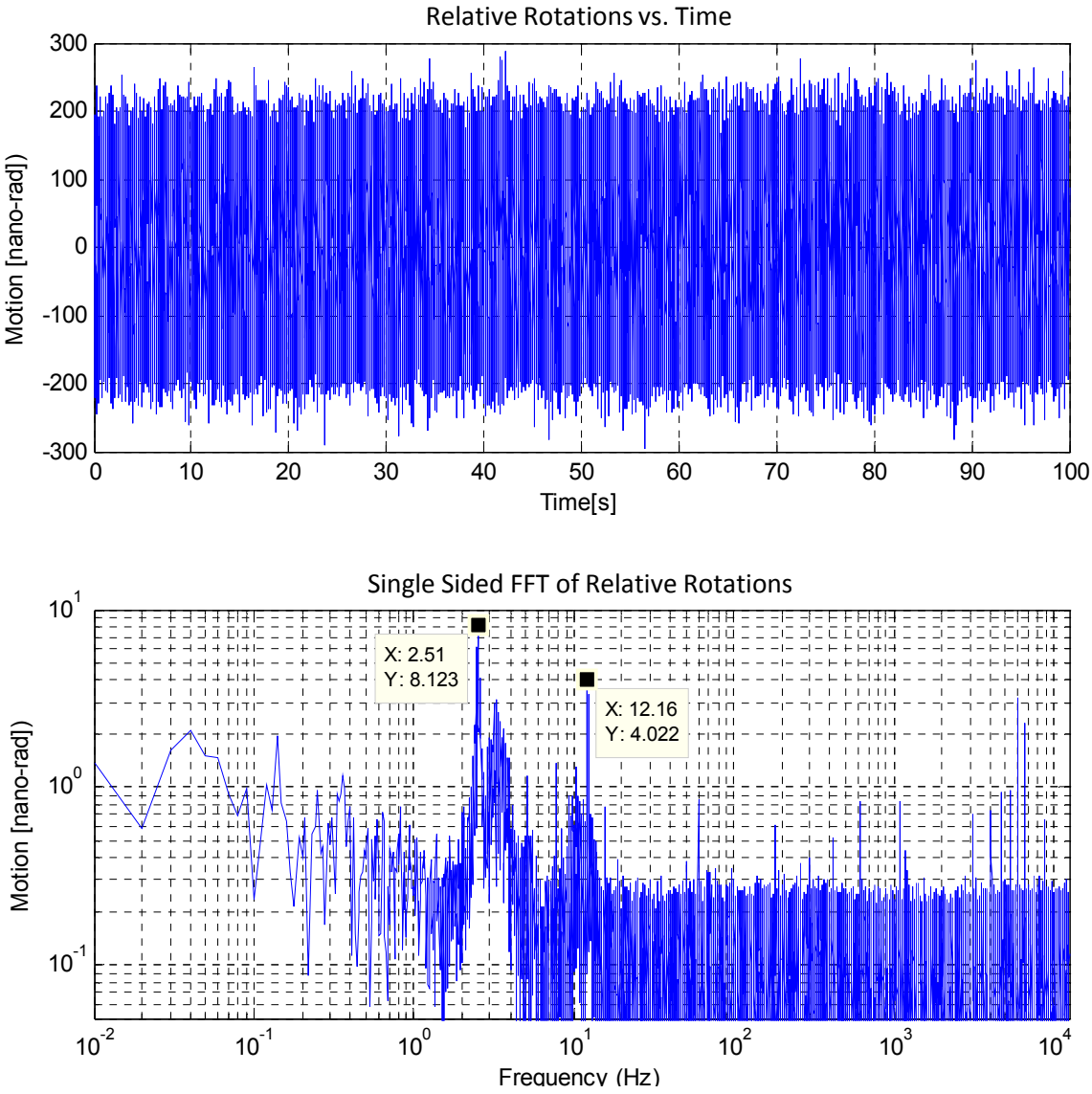

Figure 4-32: Time (top) and frequency (bottom) plots of relative rotations between stator and rotor as seen by the encoder

The governing equations for axial (Z-dir) vibrations can be expressed as

$$m\ddot{z} = -4kz \tag{4.56}$$

where m is the mass of the granite table top and k is the stiffness of one isolation mount. Equation 4.56 can be transformed into the frequency domain and solved for the natural frequency of the axial (Z-dir) vibrations as

$$-m\omega^2 Z = -4kZ$$

$$\Rightarrow \omega = \sqrt{\frac{4k}{m}} = \omega_Z \tag{4.57}$$

where ω_Z represents the natural frequency of the axial vibrations.

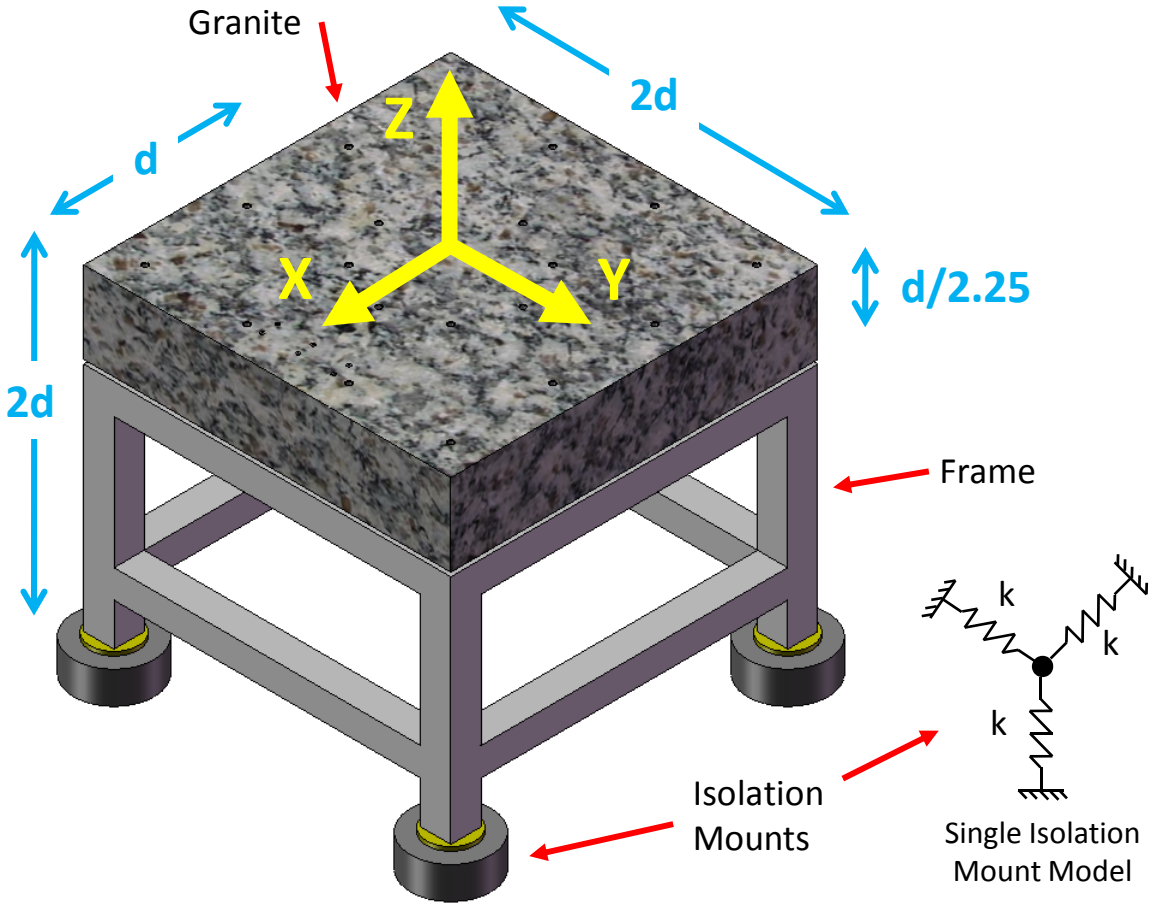

Figure 4-33: Schematic representation of the granite vibration isolation table (d= 18 inch)

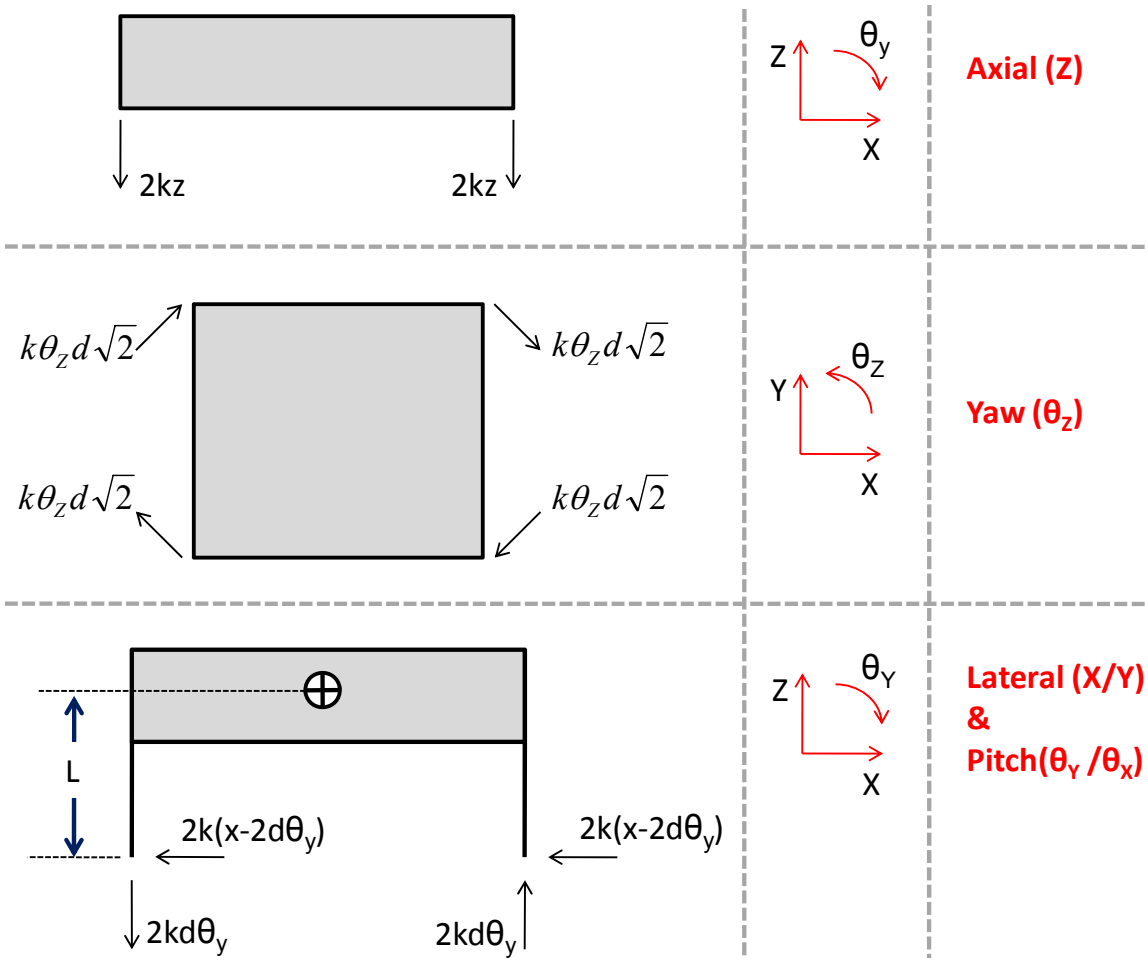

Figure 4-34: Free-body-diagrams used for derivation of the table's four vibration modes

The equations of motion for the yaw (θ_Z) mode can be expressed as

$$I_Z\ddot{\theta}_Z = -8k\ddot{\theta}_Z d^2 \tag{4.58}$$

where d is half of the length of the table top and $I_Z = (8/12)d^2m$ is the inertia of the granite about the z-axis. Equation 4.58 can be represented in the frequency domain and solved for the natural frequency of the table's yaw mode as

$$-I_Z\omega^2\theta_Z = -8kd^2\theta_Z$$

$$\Rightarrow \omega = \sqrt{3}\sqrt{\frac{4k}{m}} = \sqrt{3}\omega_Z \tag{4.59}$$

The natural frequency of the yaw mode is found to be $\sqrt{3}$ times higher than the natural frequency of axial mode.

The equations of motions for the lateral (X/Y) and pitch (θ_Y /θ_X) motions are coupled; hence, their vibration modes will be coupled and must be analyzed as a two degree of freedom system. The equations of motions can be expressed as

$$\begin{cases} mx = -4k(\ddot{x} - 2d\theta_y) \\ I_Y\ddot{\theta}_y = 8kd(x - 2d\theta_y) - 4kd^2\theta_y \end{cases} \tag{4.60}$$

where $I_Y = 0.35md^2$ is the inertia of the granite around the Y-axis. Equation 4.60 can be reformulated in the frequency domain and solved for the natural frequencies of the pitch and lateral modes as

$$\begin{cases} (-m\omega^2 + 4k)X - 8kd\theta_Y = 0 \\ 8kdx + (I_Y\omega^2 - 12kd^2)\theta_Y = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \omega_1 = 0.435\omega_Z & with \quad [x, \theta_y]_1 = [1, 0.89] \\ \omega_2 = 3.89\omega_Z & with \quad [x, \theta_y]_2 = [1, -15.5] \end{cases} \tag{4.61}$$

Although both vibration modes are a mix of lateral and pitch motions, the first mode is mostly lateral motion while the second mode is dominated by pitch motion.

Table 4-2: A summary and comparison of the natural vibration frequencies of the granite table obtained experimentally and analytically

Experimental		**Analytical**	**Mode**			
Frequency (Hz)	Frequency (ω_Z)	Frequency (ω_Z)	X/Y (m)	θ_X/θ_Y (rad)	Z (m)	Θ_Z (rad)
3	0.43	0.435	1	0.89	0	0
7	1.00	1	0	0	1	0
10	1.43	1.7	0	0	0	1
15	2.14	3.89	1	-15.5	0	0

The natural vibration frequencies obtained for the granite table both theoretically and experimentally are summarized and compared in Table 4-2. Since the value of the stiffness *k* is not known, the results are normalized to the natural frequency of the axial (Z) vibrations for comparison. As can be seen, the frequency order of the different vibration modes matches between the experimental and the analytical results. Also, the normalized frequencies match closely except for the highest frequency mode. The discrepancies are expected from our simplistic model, which assumed the stiffness element to be isotropic and ignored the mass of the frame and the rotary table.

4.5.5 Uncertainty from Pressure Fluctuations

The air-bearing needs to be supplied with a pressure source between 75-150 psi. The pressurized air creates a thin film of air between the rotor and the stator of the air-bearing causing the rotor to levitate freely. We experimentally discovered that variations in the air supply's pressure induce rotations. Although the exact cause of this is currently not clear yet, it is expected to be related to the internal structure of the air-bearing. Figure 4-35 shows synchronized time plots of the air-bearing's supply pressure and rotation over one hour with no active pressure-regulators (i.e. the spindle's supply pressure is the same as the building's airline pressure). As can be seen, the line pressure drops by about 13-psi over approximately 700 seconds. The pump pressurizes the line from 102-psi to 115-psi in about 25-seconds. The air-bearing's rotations recorded at the same time reflect the exact same pattern. The pressure fluctuations of 13-psi have resulted in rotation measurements of 3000 nano-radians. This proves that pressure variations can be a significant source of uncertainty.

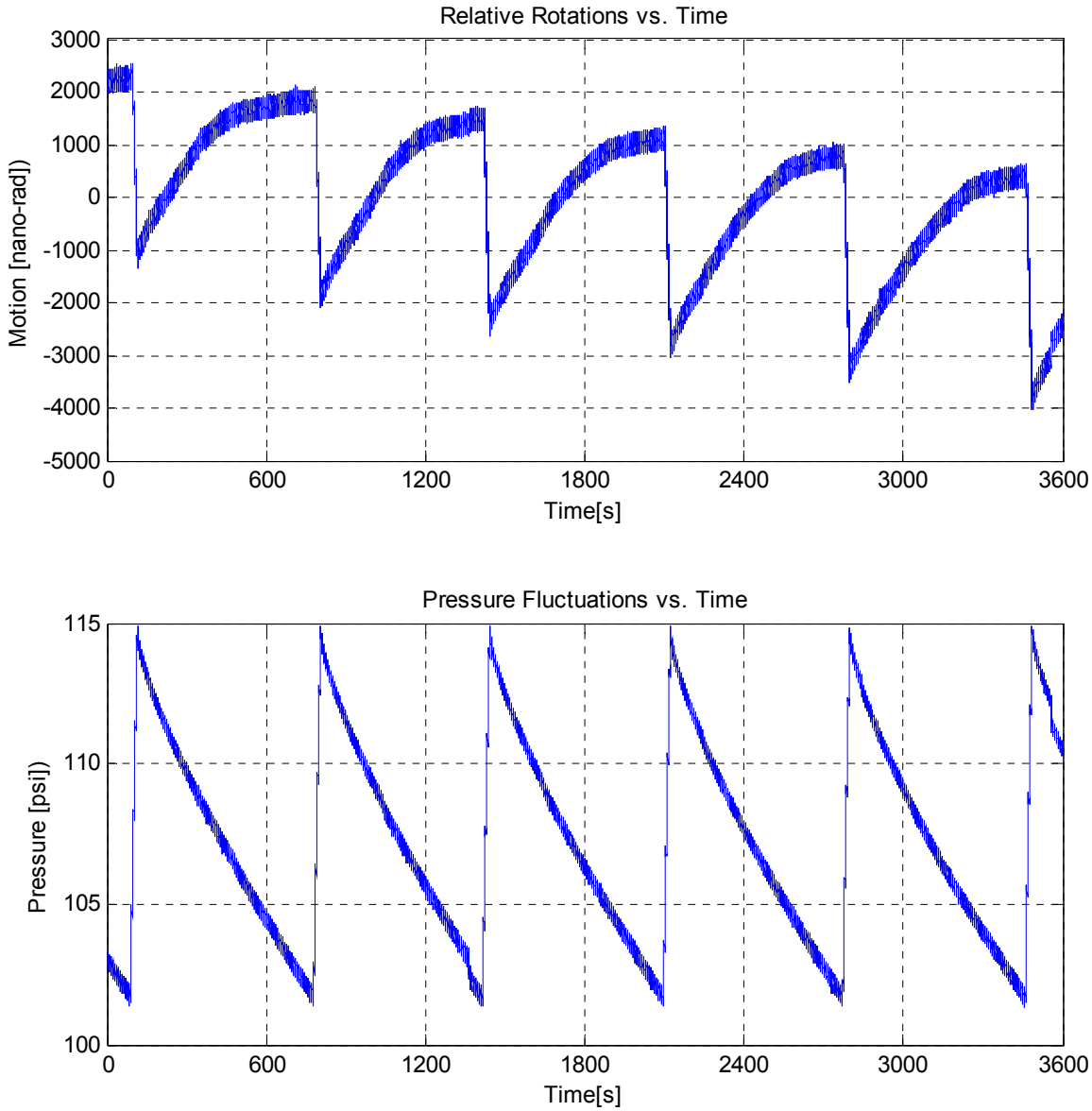

Figure 4-35: Synchronized time plot of relative rotations (top) and supply pressure fluctuations (bottom) for the spindle while left stationary with no active supply regulators over 1-hour sampled at 500-Hz

The frequency content of the pressure and rotation measurements displayed in Figure 4-35 is shown in Figure 4-36. The fundamental peak at 0.00139 Hz and its multiple harmonics represent the repeating pressure cycle happening every 720-seconds. It is also found out that the pressure variations do not exhibit any other particular frequencies.

Therefore, the other peaks must be related to the vibration modes or the air-bearing's internal dynamics.

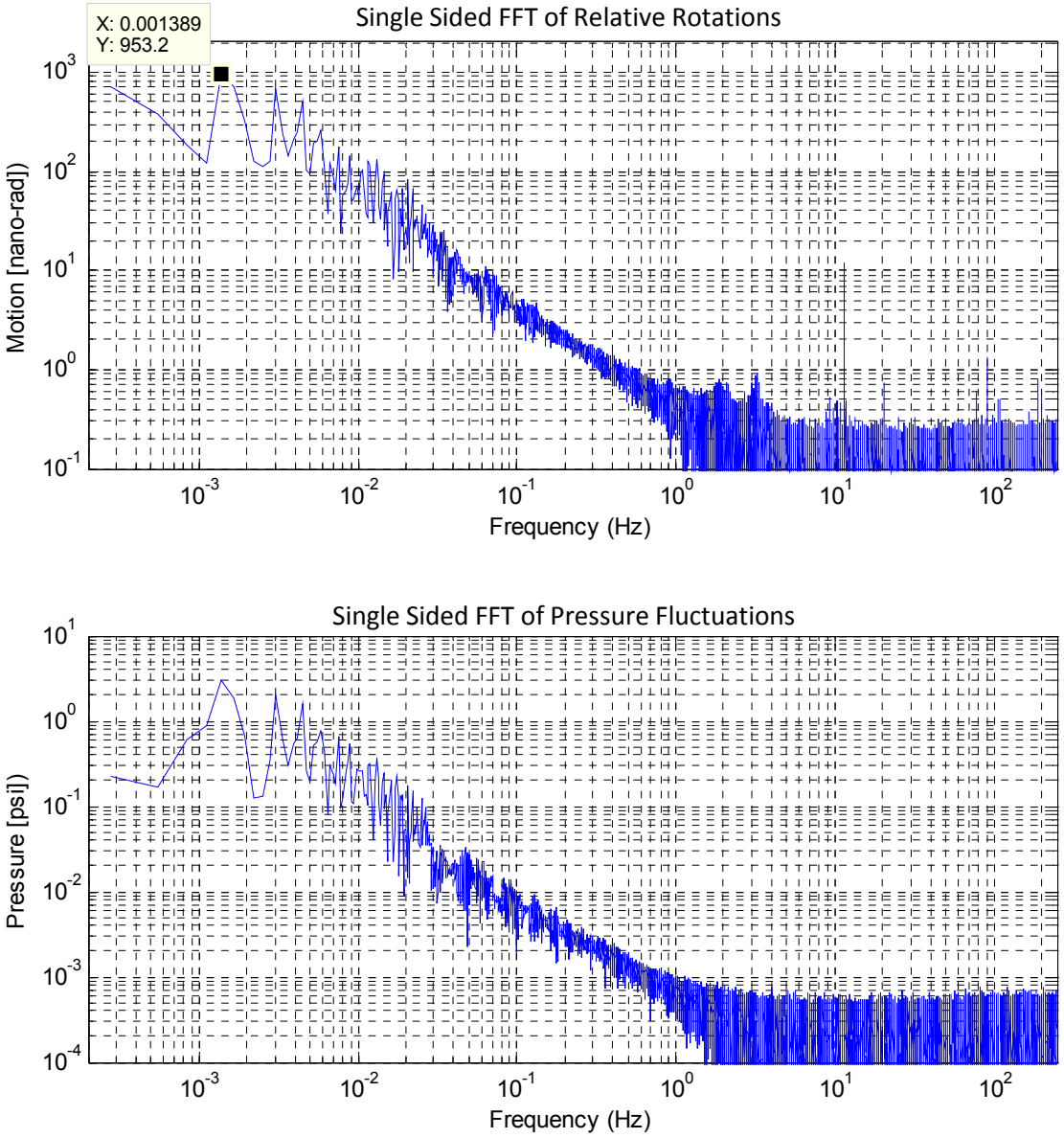

Figure 4-36: Frequency content of the rotation (top) and pressure (bottom) measurements displayed in Figure 4-35

To limit the pressure variations seen by the air-bearing, two regulators are set to reduce the line pressure from 115-psi down to 75-psi in two steps. In this configuration, both regulators are active and filter the variations of the pressure line. The air-bearing

supply's pressure variations and their resulting rotations for this case are shown in Figure 4-37. As can be seen, the pressure fluctuations have reduced by more than 10 times from 13-psi to less than 1-psi. In this case, the induced rotations are too small compared to the vibration induced rotations, and hence, can be ignored.

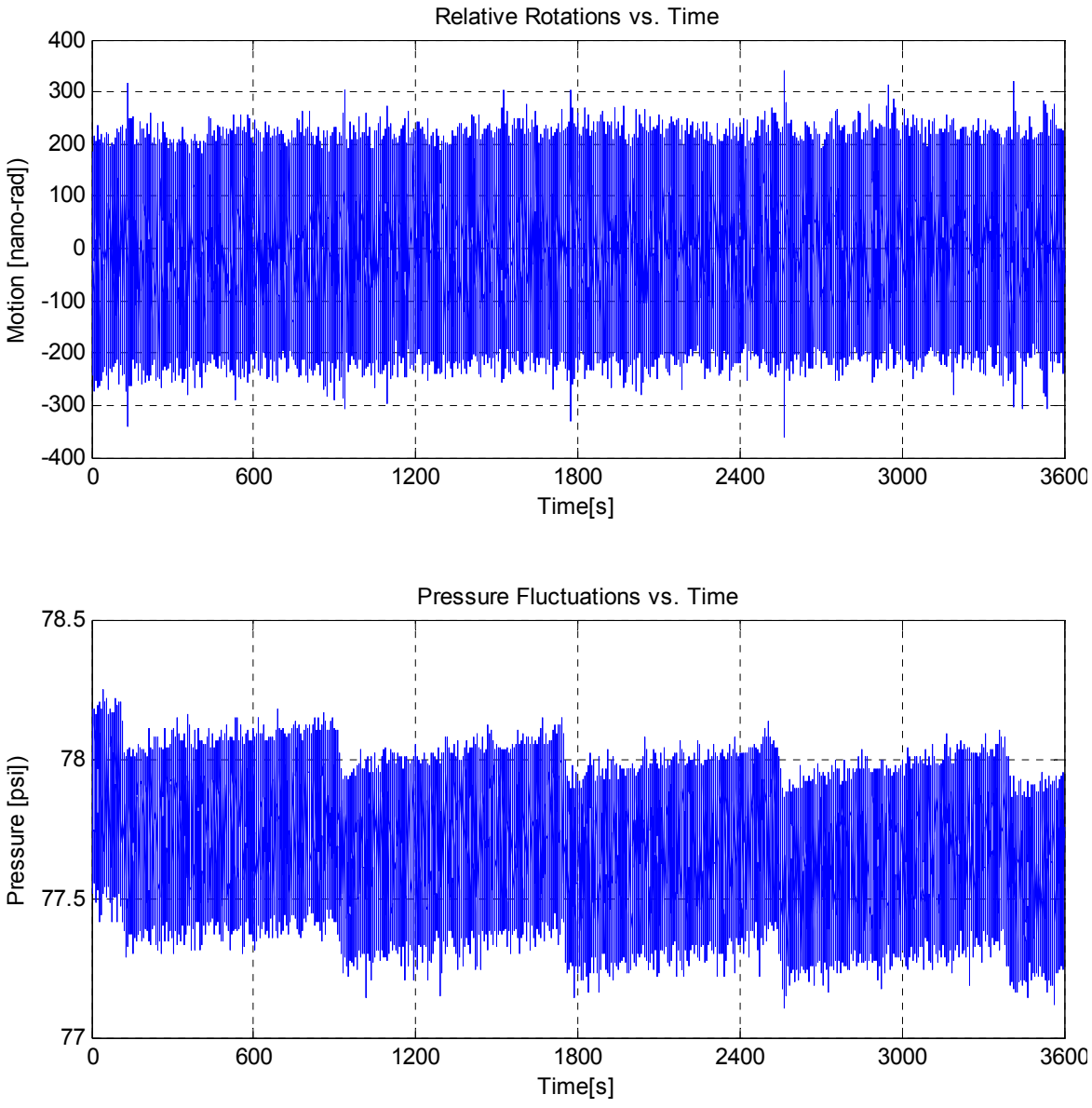

Figure 4-37: Synchronized time plot of relative rotations (top) and pressure fluctuations (bottom) with two active regulators over 1-hour sampled at 500-Hz

4.5.6 Experimental Results

The experimental setup described in section 4.5.1 was used to perform encoder calibration. A representative calibration result based on the integral second order method is shown in Figure 4-38. The error map has been obtained at 220-RPM, using the second-order integral (SOI) method, and with S_1=0 and S_2=1.5N. The calibration error-maps from the first and second revolutions demonstrate a repeatability of 161 nano-radians, where repeatability (R) is defined by equation 4.55.

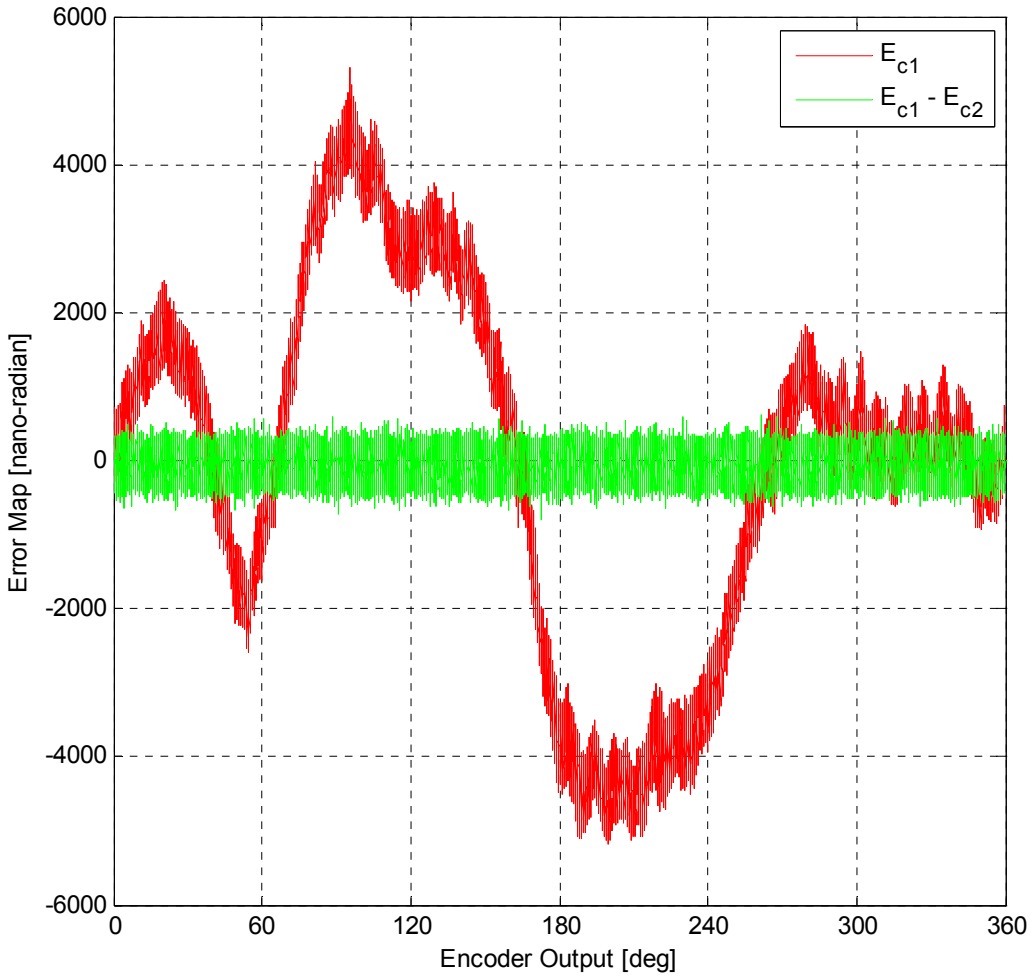

Figure 4-38: Error-Map obtained experimentally (E_{c1}) and the deviations between the two sets (E_{c1} - E_{c2}) generated by a single execution of the algorithm. The error-map was obtained at 220 RPM using the SOI method with S_1=0 and S_2=1.5.

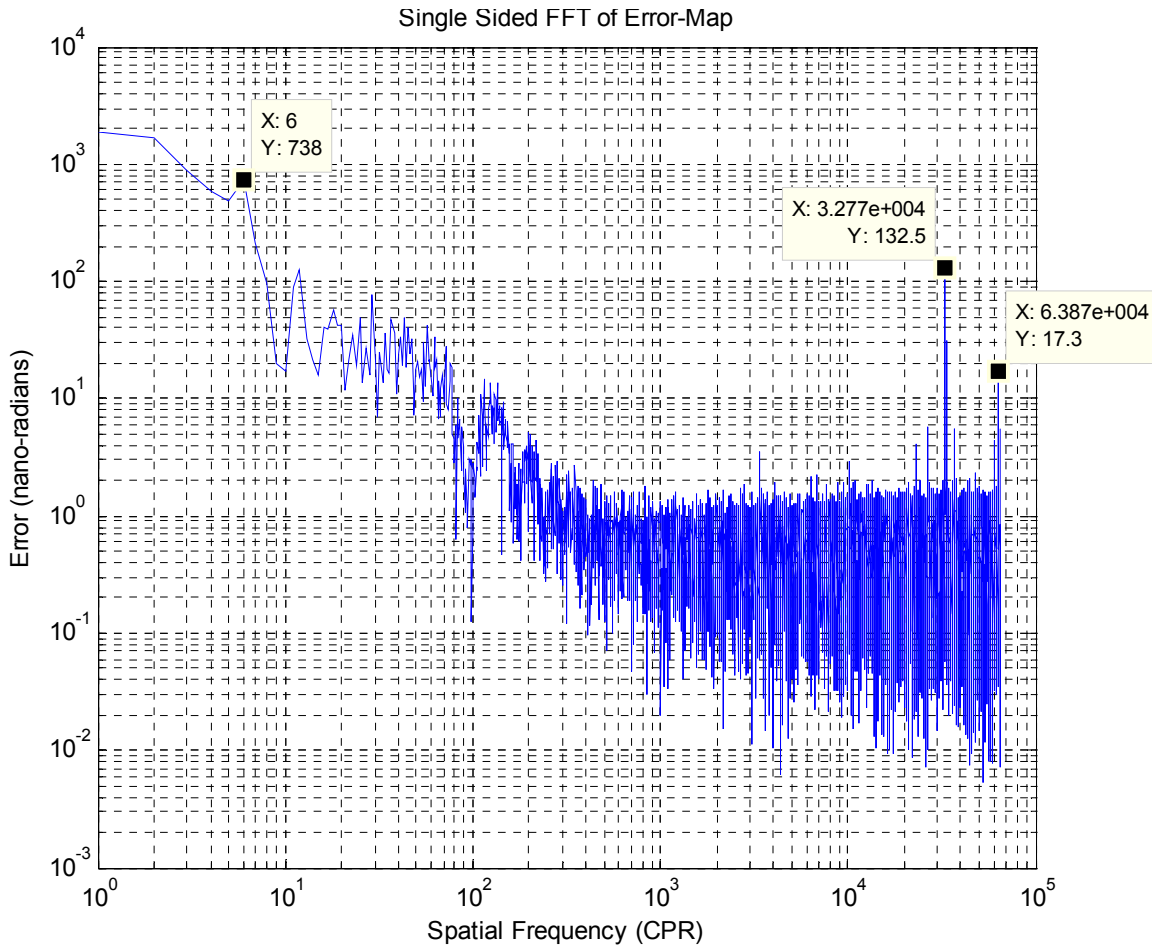

Figure 4-39: Frequency content of the representative error-map shown in figure 4-38.

The frequency content of the representative error-map in figure 4-38 is shown in figure 4-39. As can be seen the majority of the error is as the lower frequency except for the peaks at 32768 CPR (once per scale grating) and 63870 CPR (roughly twice per encoder signal cycle). The lower frequency error could be due to factors such as scale manufacturing error, mounting error, and spindle error-motion. The high frequency peaks could be due to each/both read-heads having unequal on/off widths, the two redheads not being spaced by exactly a quarter of a grating, and the grating size variations. There is peak at 6 CPR which we expect to be related to the encoder being mounted using 6 bolts.

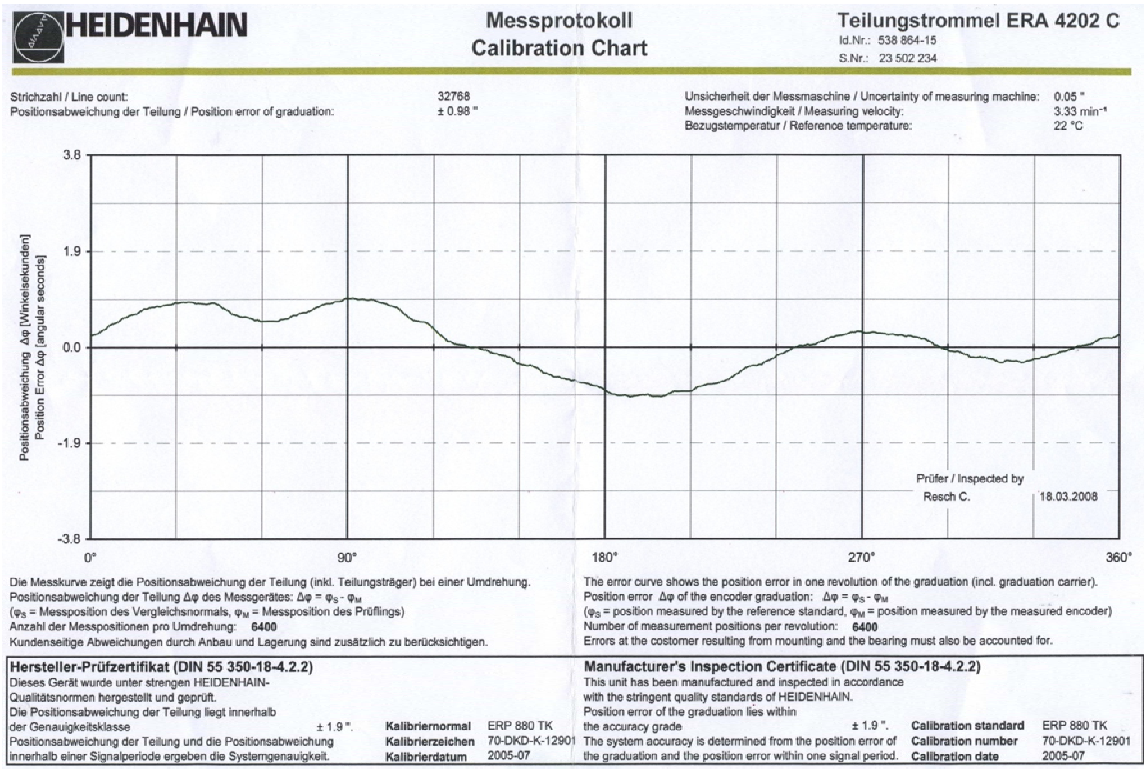

Figure 4-40: Error-map supplied by Heidenhain with our ERA4282C encoder drum obtained at 3.33 RPM (serial number 23-502-234)

Heidenhain's encoders are supplied with a calibration chart indicating the encoder's accuracy. The error-map supplied by Heidenhain for the encoder used in our tests is shown in Figure 4-40. Heindehain's error-map only captures the lower harmonics and is indicated to have a measurement uncertainty of 0.05 arc-second or 240 nano-radians. We have digitized Heidenhain's error-map, using the DigXY software, so that we can compare it with our representative error-map from figure 4-38. To eliminate the effect of eccentricity on the error-map we have removed the first harmonic of the error from both error-maps and have plotted them on the same graph in Figure 4-41. As can be seen, the magnitude and the patterns of the two error-maps match very well; however, the

two are not exactly the same. The reason for the discrepancy could be due to encoder mounting/misalignment variations, operating angular speed differences, and temperature. In fact, it was experimentally shown that the tension on the mounting bolts will influence the encoder error.

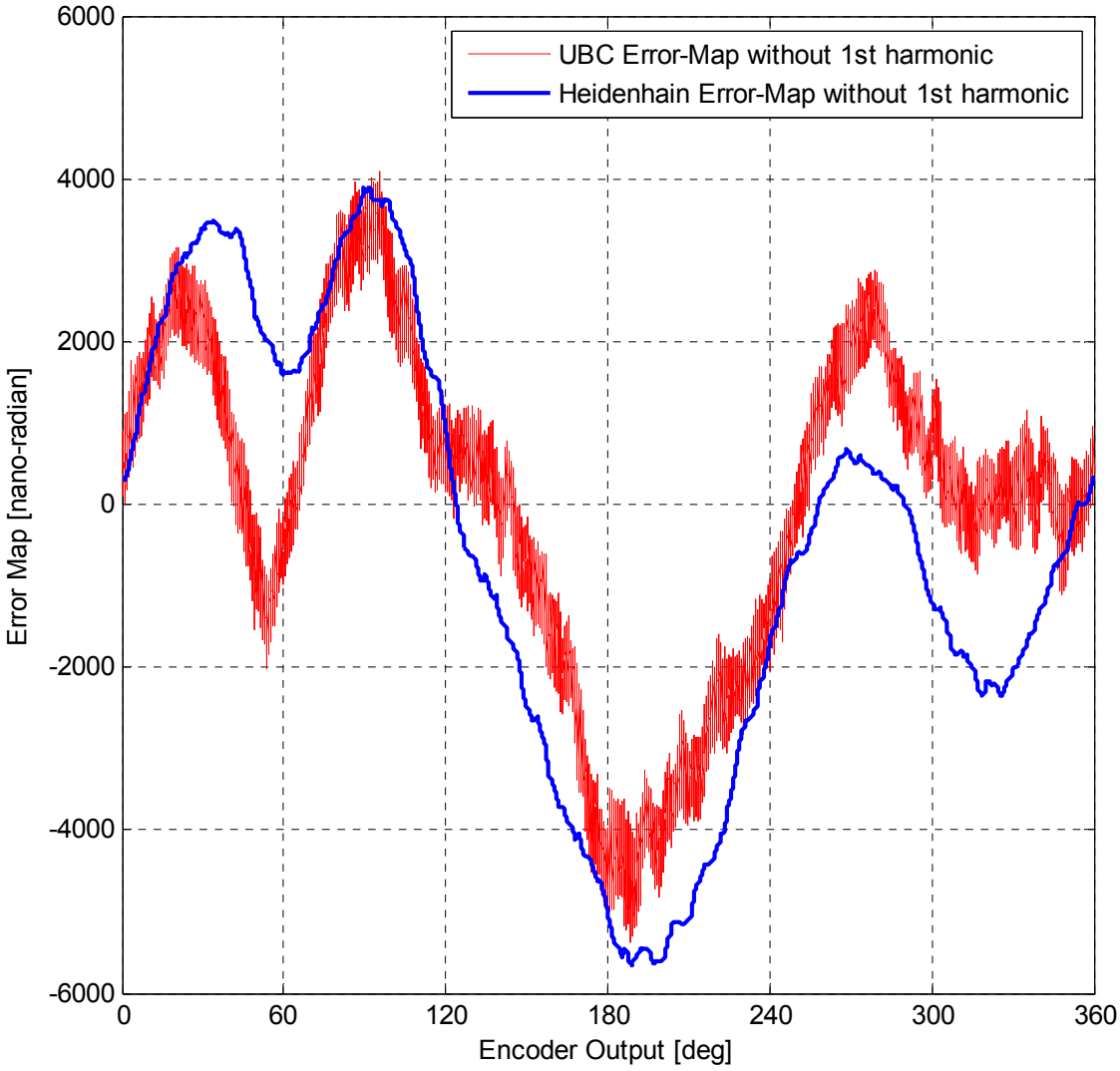

Figure 4-41: Comparison between Heidenhain's and our error map for the ERA4282C drum with the serial number of 23-502-234

So far we have only analyzed a representative error-map obtained at a single speed. To better test the performance and repeatability of the calibration algorithm, time measurements are continuously logged for about 500-revolutions and are then used to

perform encoder calibration over a range of angular speeds. Second order integral encoder calibration method is performed on all the revolutions captured by indexing S_1 from the first revolution to 2 revolutions before the last revolution and keeping $S_2=S_1+1.5(N)$, where N is the number of encoder counts per revolution. Repeatability is calculated for each resulting error-map. The calculated repeatability values are plotted versus the angular-speed at which the error-map was obtained. Figure 4-42 shows a repeatability plot obtained experimentally over the angular-speed range of 50 to 250 RPM under the following conditions:

- Second order integral (SOI) calibration method
- 200-MHz timer speed
- 90-psi airbearing supply pressure regulated down from 115-psi
- The granite table legs all at 80-psi, but not the table was not leveled

Looking at figure 4-42, repeatability decreases with speed, which could be due to the timing uncertainties being reduced. However, below 60-RPM, the repeatability starts to increase with reducing speed. Applying the enhancement method to the logged data did not change the repeatability values at all. As a result, the lower speed limit at 60-RPM is not the enhancement limit. Instaed, it oculd be due to spindle dynamics changing significantly, or outside disturbances becoming more significant at those speed.

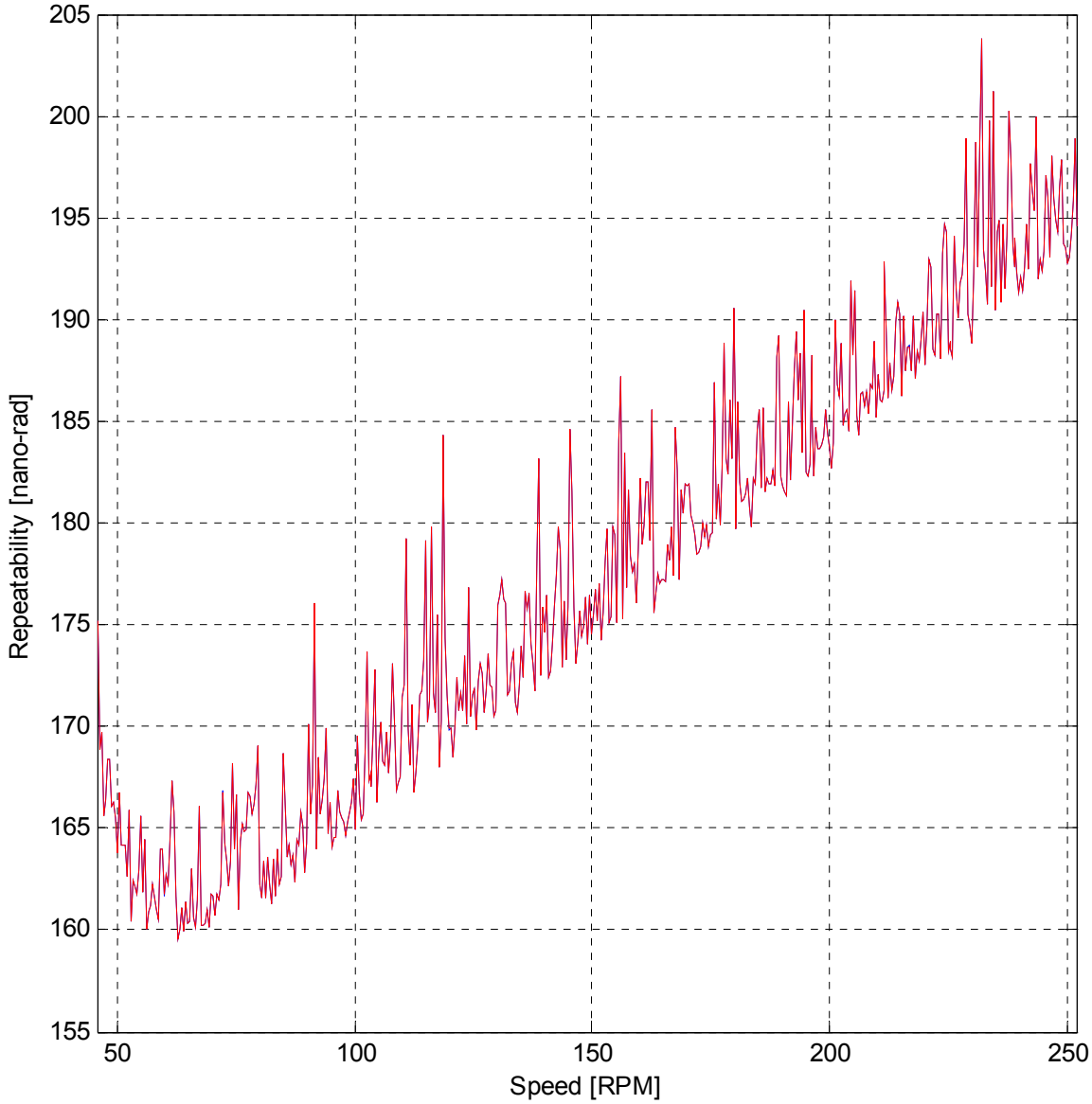

Figure 4-42: Continuous encoder calibration repeatability (R) plot with the second order integral method.

The experimental uncertainties, discussed in sections 4.5.2 through 4.5.5, set a lower limit on calibration repeatability. In section 4.5.4 the uncertainty due to vibrations, noise, and pressure fluctuations is shown to be around 70 nano-radians RMS. In section 4.5.1, it is shown that there is an uncertainty of 52 nano-radians caused by our 200-MHz

timer for a speed of 200-RPM. The combined uncertainty caused on accuracy can be calculated as

$$\sigma_a = \sqrt{{\sigma_v}^2 + {\sigma_t}^2} = \sqrt{70^2 + 50^2} = 86 \tag{4.62}$$

where σ_a, σ_v, and σ_t are the expected RMS variations in nano-radians of calibration accuracy, vibration induced rotations, and time discretization error respectively. Because repeatability is dependent on two calibration results, its variations can be estimated as

$$\sigma_r = \sqrt{{\sigma_a}^2 + {\sigma_a}^2} = \sqrt{2}\sigma_a = 122 \tag{4.63}$$

where σ_r is the expected RMS deviation in repeatability. The expected lower limit on repeatability is about 70 nano-radians lower than the obtained experimental repeatability. One reason for this discrepancy could, perhaps, be that the estimation of the expected repeatability assumes a single time direct influence of the uncertainties on the calibration outcome. Such an assumption only accounts for the each count's temporal duration variations; however, does not account for any biased damping estimates resulting from the uncertainty sources. Also, the uncertainty analysis does not cover other potential sources of uncertainties such as the comparator's imprecise switching, and the bearing's non-ideal dynamics. Further experimental testing is required to remove the limiting uncertainties one by one and improve the achievable repeatability.

In section 4.4.1, equations 4.54 and 4.55 were provided as the definition for error-map accuracy and repeatability. For experimental error-maps, it is not possible to calculate accuracy because the actual encoder error-map is unknown. However, repeatability can be defined between different error-maps and can be used as measure of the calibration quality. The repeatability defined by equation 4.55 is a measure of the difference between a pair of error-maps obtained by a single execution of the calibration

algorithm, and hence, the two error maps are obtained at almost equal speeds. As a result, the repeatability does not reflect any possible variations in the error-map caused by changes in speed. To study the variation of the error-map with speed we define the speed-repeatability performance metric as

$$\mathrm{SR}_{\mathrm{n0}}(n) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[E_{c1}^{n}(i) - E_{c1}^{n0}(i)]^2}\ , \tag{4.55}$$

where $\mathrm{SR}_{\mathrm{n0}}$ is the speed-repeatability of the error-map at n[RPM], using n_0 as the reference speed, and E_{c1}^{n0} and E_{c1}^{n} are the error maps from the first revolution obtained at n_0 [RPM] and n [RPM] respectively.

The same continuously logged data, which was used to generate the repeatability plots, has been processed to create the speed repeatability plot, shown in figure 4-43 based on a reference speed of $n_0 \approx 250$ [RPM]. As we can see from figure 4-43, speed-repeatability is completely different from repeatability. The big variations in speed-repeatability point out unreasonably large variations in the error-map. The flat sections of the speed-repeatability chart above 150RPM show that the calibration is perhaps working fine in that region. However, below 150RPM, the large and increasing slope of the speed-readability curve show that the error-maps are not reliable and the calibration algorithm is not working well. Due to time constraints, the cause for the discrepancies at lower speeds has not been determined; however, few hypotheses have been made and are to be tested as a part of the future work. One possible reason for the variations could be the vibrations of the table induced by the rotating mass. Alternatively, we suspect the problem to be due to pendulation torque caused by the combined effect of rotating unbalance and the table not being leveled.

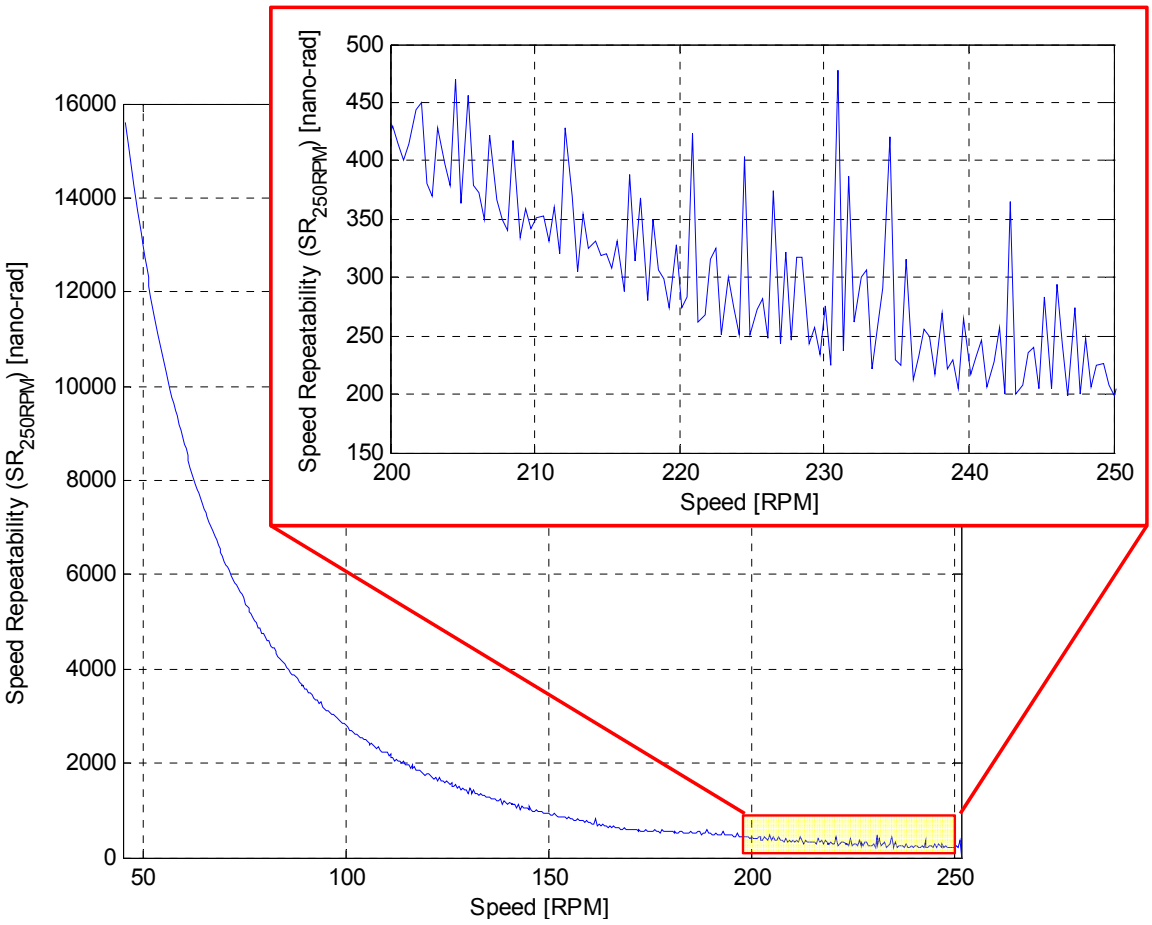

Figure 4-43: Continuous calibration speed-repeatability ($\boldsymbol{SR^{n}_{250RPM}}$) plot with second order integral method, using reference speed of n_0=250RPM, and for $\boldsymbol{n \in [50-250]}$

4.6 Summary

This chapter focused on encoder calibration and provided background on rotary encoder devices, the calibration algorithm, and the calibration tests. In this chapter, the utilized encoder calibration method was described in details and both simulation and experimental encoder calibration results were presented. An integral method was introduced to reduce the model estimation's sensitivity to time uncertainties. The integral method's advantage was proved using simulations. An uncertainty analysis was performed to justify the estimate a lower limit on repeatability and to identify the limiting factors on repeatability. Vibrations were identified as the most significant limiting factor followed by electrical noise and time discretization.

An encoder calibration repeatability of better than 200 nano-radians over 50RPM-250RPM angular speed range was demonstrated experimentally. The repeatability value is close to our 120 nano-radians lower limit estimate, which confirms the consistency of our implementations. On the other hand, experimental speed repeatability value were found to be reasonable for above 200RPM and unreasonably large for lower than 150RPM angular-speeds. Due to time constraints the cause for the discrepancies at lower speeds could not be determined and is considered as future work.

CHAPTER 5

Conclusion and Future Work

5.1 Conclusion

In conclusion, this thesis has developed precision mechanical platform and high-speed electronics which can be used for spindle metrology and control. In this thesis, the developed experimental setup is ued for encoder calibration, where a rotation measurement repeatability of better than 200 nano-radians has been experimentally demonstrated over 50-250RPM. The primary contributions of this thesis include the following:

1. A precision mechanical platform was designed, manufactured, assembled, and tested. The platform can accommodate a range of spindle metrology research including rotation metrology research and error motion identification research.
2. A set of high-speed electronics was designed, manufactured, assembled, and experimentally tested. The high-speed electronics can be used in any application requiring high-speed data processing. In this thesis, the electronics are intended to be used for encoder calibration, digital current control, and digital motion control.

3. Encoder calibration research was conducted where the accuracy limit of the self-calibration method developed in [1] was improved to better than 200 nano-radians and was experimentally demonstrated.

5.2 Future Work

An encoder calibration accuracy limit of 10 nano-radians is projected for this calibration method. Looking into future, we propose the following work to improve the encoder calibration accuracy and reach its 10 nano-radians projected accuracy limit.

5.2.1 Understanding the Discrepancies at Low Speeds

The speed-repeatability results show very large variations of the error-map at lower speeds. Understanding and eliminating the cause of the lower speed variations would enable using the calibration method at lower speeds and can perhaps significantly improve the calibration accuracy at higher speeds.

5.2.2 Reducing Experimental Uncertainties

The existing experimental uncertainties directly limit the performance of the calibration system. Many of the existing uncertainties can be reduced. The vibrations can be attenuated by using a better vibration isolation system, such as the optical-table available in our lab. Time uncertainties can be significantly reduced by utilizing a novel quad-rate timer scheme developed as part of this thesis. Signal noise uncertainties must be tracked more closely and reduced by proper shielding and effective use of filters. A higher performance regulator can be used to remove the remaining pressure variations.

5.2.3 Multiple Read-Head Configuration

The rotary table is designed so it can use multiple read-heads to cancel the effect of air-bearing's lateral vibrations. Two and four read-head configurations can be tested. In addition, using an averaged signal from multiple read-heads reduces the effect of electrical noise and its associated uncertainties.

REFERENCES

[1] X. Lu. (2007, Self-calibration of on-axis rotary encoders. *CIRP annals - manufacturing technology 56(1),*

[2] SIA. (2007, Semiconductor industry association's 2007 annual report. Semiconductor Industry Association, United States. Available: http://www.sia-online.org/downloads/SIA_AR_2007.pdf

[3] SIA. SIA backgrounders - international technology roadmap for semiconductors. *2008(7/18),* Available: http://www.sia-online.org/backgrounders_itrs.cfm

[4] ITRS. (2007, International technology roadmap for semiconductors 2007 edition - lithography. International Technology Roadmap for Semiconductors, United States. Available: http://www.itrs.net/Links/2007ITRS/2007_Chapters/2007_Lithography.pdf

[5] Heindehnain. (2006, Heidenhain's angle encoders with integral bearings brochures. pp. 36. Available: http://wwwpdb.heidenhain.com/ansicht/Heidenhain/media/img/591_109-21.pdf

[6] T. Watanabe, H. Fujimoto, K. Nakayama, T. Masuda and M. Kajitani. Automatic high precision calibration system for angle encoder (II). Presented at SPIE Vo. 5190 REcent Developments in Traceable Dimensional Measurements II.

[7] Probst. (1998, The new PTB angle comparator. *Measurement science technology 9(7),* pp. 1059.

[8] Renishaw. Renishaw's RX10 rotary axis calibration. *2008(09/05),* pp. 2. Available: http://www.renishaw.com/en/6805.aspx

[9] E. W. PALMER. (1988, Goniometer with continuously rotating gratings for use as an angle standard. *Precision engineering 10(3),*

[10] Y. V. Filatov, D. P. Loukianov and R. Probst, "Dynamic angle measurement by means of a ring laser," *Metrologia,* vol. 34, pp. 343, 1997.

[11] Y. V. Filatov, M. Y. Agapov, M. N. Bournachev, D. P. Loukianov and P. A. Pavlov, "Laser goniometer systems for dynamic calibration of optical encoders," in *Optical Measurement Systems for Industrial Inspection,* 2003, pp. 381.

[12] W. T. Elster. (1998, Uncertainty analysis for angle calibrations using circle colsure. *Journal of Research of the National Institute of Standards and Technology 103(2),* pp. 141.

[13] T. Masuda and M. Kajitani. (1989, An automatic calibration system for angular encoders. *Precision engineering 11(2),* pp. 95.

[14] Theodore D. Michaelis, "Apparatus and method for generating calibrated optical encoder pulses," 4593193, 3 Jun 1986,

[15] P. A. Orton, J. F. Poliakoff, E. Hatiris and P. D. Thomas. Automatic self-calibration of an incremental motion encoder. Presented at IEEE Instrumentation and Measurement Technology Conference.

[16] Professional Instruments. Model 10R block-head airbearing spindle. *2008(08/01),* pp. 1. Available: http://www.airbearings.com/10r

[17] Heindehain. (2006, Heidenhain's angle encoders without integral bearings brochures. pp. 56. Available: http://wwwpdb.heidenhain.com/ansicht/Heidenhain/media/img/606_136-22.pdf

[18] ThinGap. (2008, ThinGap motor technology. ThinGap, Available: http://www.thingap.com/pdf/technologypaper.pdf

[19] ThinGap. (2007, ThinGap torque ripple brief. ThinGap, Available: http://www.thingap.com/pdf/torquerippleeffect.pdf

[20] Xilinx. (2008, Virtex-4 FPGA user guide. *(V2.6),* Available: http://www.xilinx.com/support/documentation/user_guides/ug070.pdf

[21] Xilinx. (2008, Virtex-4 configuration guide. Available: http://www.xilinx.com/support/documentation/user_guides/ug071.pdf

[22] Xilinx. (2004, Virtex-4 PCB designer's guide. *Power Distribution System*pp. 29. Available: http://china.xilinx.com/support/documentation/user_guides/ug072.pdf

[23] Analog Devices. (2007, ADCMP605/604 datasheet. Available: http://www.analog.com/static/imported-files/data_sheets/ADCMP604_605.pdf