PRACTICAL AND EFFICIENT SYSTEM IDENTIFICATION OF PRECISION GANTRY SYSTEMS

by

Eric T. Belski

B.S. in Mechanical Engineering, University of Pittsburgh, 2015

Submitted to the Graduate Faculty of

the Swanson School of Engineering in partial fulfillment

of the requirements for the degree of

Master of Science in Mechanical Engineering

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

SWANSON SCHOOL OF ENGINEERING

This thesis was presented

by

Eric T. Belski

It was defended on

April 3, 2018

and approved by

Daniel G. Cole, Ph.D., P.E., Associate Professor

Stephen J. Ludwick, Ph.D., Adjunct Associate Professor

Jeffrey S. Vipperman, Ph.D., Professor, Vice-Chair

Thesis Advisor: Daniel G. Cole, Ph.D., P.E., Associate Professor

PRACTICAL AND EFFICIENT SYSTEM IDENTIFICATION OF PRECISION GANTRY SYSTEMS

Eric T. Belski, M.S.

University of Pittsburgh, 2018

This research develops a procedure to efficiently measure the multi-input multi-output frequency response of precision gantry systems. A precision gantry considered here is defined as two axes moving a bridge in tandem with a third axis mounted to the bridge. An efficient method of measuring the frequency response will decrease the total time required for designing a controller, and higher volumes of gantries can be processed in a set amount of time.

Sequential, random phase multisine signals are chosen as an excitation signal for the system identification. A model-guided methodology based on system parameters determines how many locations at which the frequency response should be measured, as the response varies with axis location. Guidelines are developed and used to determine that the frequency response of the gantry base and workpoint can be measured using accelerometers in order to provide information about system behavior. The frequency response data is processed using multi-input multi-output techniques to provide a meaningful, straightforward presentation. The procedure is then structured into an automated program such that minimal user input is necessary. The result is an automated process that can measure and present the multi-input multi-output frequency response in a deterministic and repeatable manner. The automated procedure is also tested and validated on a sample of industrial gantries.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

The objective of this research is to develop a procedure to efficiently measure the multi-input multi-output (MIMO) frequency response of gantry systems for use in an industrial setting. Precision gantry systems will be considered in this research, which will be classified as two axes that move a bridge in tandem in the x-direction with one axis on the bridge moving in the y-direction, Figure 1. Frequency response measurements are used to design controllers and with the multiple axes on a gantry, a MIMO frequency response should be measured to effectively represent the system under control. The MIMO frequency response of a system is a matrix of transfer functions that each describe how a given input affects some output, in the frequency domain. Various transfer function matrices are used in the analysis and design of MIMO control systems and it is necessary to properly determine these functions to ensure good controller design.

There is a need for a more efficient MIMO frequency response measurement strategy with the prevalence of precision gantry systems in the motion control industry. As applications require faster, more precise motion, gantry systems are often used because of their high force density and structural rigidity. This allows for high accelerations to be commanded without the worry of excessive compliance in the system. With the increased demand for gantries, so too is it is necessary for the industry to optimize the process of identifying system mechanics and designing controllers.

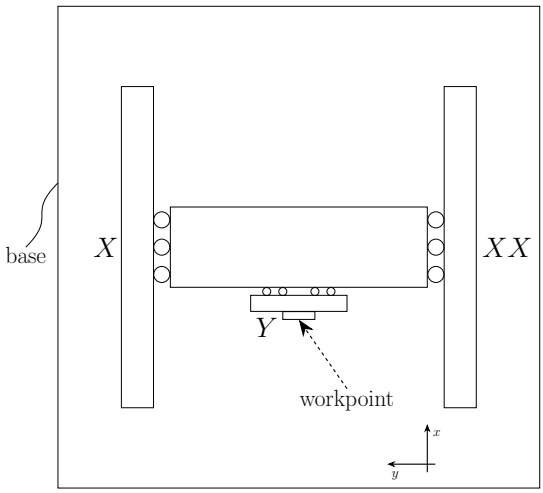


Figure 1: This schematic diagram of a typical gantry system shows two parallel axes (X and XX) free to move in the x-direction, carrying a third axis, Y, that moves in the y-direction. Attached to the Y axis is some payload or tool, commonly referred to as the workpoint.

1.1 STATE OF THE ART

After a gantry is assembled at the factory, a controller is applied to the system in order that a test procedure can be carried out on the system. This test ensures that under normal operating conditions, all system outputs are responding in a desirable and predictable manner. In order to pass the test procedure, the controller is required to be robust such that external disturbances and plant variations do not affect system stability but also maintain a high bandwidth to achieve the throughput requirements of various applications. Achieving a robust controller with the best performance possible necessitates a model of the plant with well defined uncertainty bounds. These uncertainty bounds allow the controller to achieve the highest bandwidth possible while providing the knowledge that robust conditions are not being violated. This is time consuming and difficult with the identification practices currently in use.

In the example system in Figure 1, there are 3 axes, each of which is individually excited at the motor. The output of the axis encoders are measured for each direction. Measurements are typically performed over a 3×3 grid of locations, resulting in 27 individual measurements that need to be conducted. To obtain a useful frequency resolution and range for the frequency responses, typically 200 logarithmically spaced frequencies are excited with a swept sine signal between 20 Hz and 2000 Hz. The measurement at each location takes approximately 80 seconds, verified by actual measurements. This measurement is duplicated 27 times at different locations in travel, yielding a total measurement time of approximately 36 minutes (neglecting move times between locations). Additionally it is useful to input both low- and high-amplitude excitations for each axis to identify nonlinear behavior. This will double the number of responses that need to be collected, consequently doubling the collection time to 1 hour and 12 minutes. Once the data has been collected, a control engineer or technician analyzes it using single-input single-output (SISO) methods and adjusts the controller accordingly. Because the controller of the system has been altered, another grid of responses must be collected to verify stability throughout travel [24]. This process may need to be repeated several times.

When using SISO techniques to analyze a multi-axis control loop, inaccurate stability margins can be reported because the coupling of the system is not properly accounted for [9]. Multiple inputs have an effect on each output, however SISO methods only consider the influence one input at a time. To solve these problems MIMO analysis techniques should be used. In [28], Stoev makes the observation that many industrial controllers are SISO based because:

- they are easily understood and tuned manually
- they can be based on non-parametric frequency response functions, which are inexpensive and accurate
- they require less expensive controller and amplifier hardware

However MIMO controllers can be analyzed with specific tools such that they are easily understood [5], designed with non-parametric frequency response functions [18], and implemented with the same hardware as SISO controllers. An example of how to implement MIMO analysis techniques is necessary to overcome the perception of increased design cost as well as demonstrate the advantages over SISO techniques.

Another source of error when calculating stability margins is measurement error. With the use of a swept sine excitation signal, it is not possible to determine the measurement uncertainty due to the way it is processed to obtain other advantages. If uncertainty bounds are desired, then more periods of each sine wave must be measured, further increasing the process time. Swept sine excitations are used for their flexibility with frequency range and spacing as well as having neither leakage nor harmonic distortion [14]. An improved excitation signal would take less time, contain most of the advantages of swept sine, but also lend itself to determining measurement uncertainty.

The frequency response of a gantry is understood to have dependency on the location of the axes [4]. In order to gauge some estimate of the position dependency, the frequency response function needs to be measured in multiple axis locations. Currently the operator has discretion about how many locations in gantry travel to measure, which can lead to under or over characterized systems. When the operator is determining how many locations to measure, they take into account parameters of the system such as: moving mass, bridge length, gantry axes' length, etc. There is no standard procedure so the number of measurement locations could be inaccurately determined if a system parameter is not properly accounted for. If too few locations are measured, there could be unaccounted variations in the dynamics that will affect the end process because they are not considered. Conversely if the system is over characterized, extra time is spent characterizing more locations in travel than is necessary. Either of these situations could be avoided if intrinsic knowledge of the system is used to systematically determine at how many locations to measure the frequency response. This deterministic process should help to optimize the amount of time needed for a sufficiently characterized system response to be collected.

A frequency response matrix of the plant of a gantry is most often measured from the motor inputs to the encoder (position) outputs. These encoders are normally located close

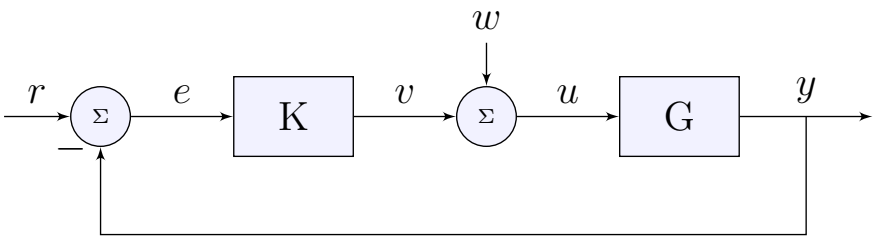


Figure 2: Block diagram describing standard characterization setup where r is the reference signal, e is the error, K is the controller, v is the controller output, w is a generated disturbance, u is the input to the plant, G is the plant, and y is the plant output.

to the motor inputs to create a desirable control loop. While this is convenient from a control design perspective, the encoder is normally located some distance from the system point of interest (workpoint) where the application process is occurring. The larger distance from the system outputs can lead to undesired and unmeasurable behavior at the system workpoint with little to no evidence of this at the system output measurements. A common problem is optimizing performance of the system at the outputs (feedback) of the system while neglecting that the performance is poor at the workpoint. A system will be largely ineffective if the performance at the workpoint is dissimilar to the performance at the outputs. An example of this is high throughput laser contouring applications where high gain is necessary to track command inputs well. If the gain is large at a resonant frequency of the workpoint, but the system outputs are insensitive to the mode shape of the resonance then the mode will be excited without any observable effect. This will translate into good tracking of the profile at the system outputs, however the actual laser path is oscillating about the desired trajectory.

Not only is the workpoint behavior important to the application of a gantry but also the behavior of the base the system is mounted to. If the machine base is susceptible to external disturbances that will affect the servo axes, it can have negative effects on the overall process performance of the gantry. The base may also be prone to vibration from the motion

of the gantry axes which then causes unwanted effects to couple back to the axes. A method of instrumenting both the workpoint and base would aid in the identification of unwanted behavior and also the possibility of generating control effort from these measurements to reject undesired motion.

Inherent in these problems is the need of someone with extensive formal training or experience to solve them and implement the solutions. This is acceptable in research or low volume production, but not in industrial applications with higher production volumes. Therefore some translation into an automated procedure is needed to allow non-experts to utilize the proposed methods in this research.

1.2 OBJECTIVES

Individual objectives are identified and used to form a comprehensive approach to improve the current art of MIMO system identification in industry. The objectives are:

1. **Reduce the overall time required for the measurement process and quantify the uncertainty in the frequency response.**

   Swept sine excitation signals are excellent for obtaining low uncertainty measurements but are the slowest of the excitation signals in consideration for frequency response measurements. White noise excitation is very fast but contains high uncertainty in most multi-axis systems due to spectral leakage [8]. A compromise between the two, which shows good speed improvement over swept sine but less uncertainty than white noise, is multisines. Multisine excitation signals contain the sum of different periodic signals to excite multiple frequencies simultaneously [31]. *The hypothesis is that the multisine signal will bridge the gap between measurement uncertainty and speed improvements.* Typical multisine measurements for these types of systems take about 20 seconds where as swept sine measurements of the same frequency resolution and range take about 80 seconds. Various types of multisine signals are investigated to find the one with the best character-

istics in the frequency ranges of interest as well as the fastest measurement time. This step is critical to the robustness of the controller that will be designed using these responses. Large uncertainty in the frequency response could require overly conservative controller gains to properly satisfy robustness requirements.

2. **Define and incorporate an approach for using a priori knowledge of the gantry system to determine at how many locations the frequency response should be measured.**

   There will be additional time savings and reduced uncertainty by using a priori knowledge of the stage to define how many frequency response measurement locations are needed. The approach here will be to use parameters, such as the ratio of moving mass on the bridge to total bridge mass, to determine how many locations in travel should be measured. If the moving mass of the bridge is small relative to the overall bridge mass, it may only be necessary to collect frequency responses at one location in the travel of the bridge axis. Other parameters that will be taken into account are: the length of the bridge axis, and the length of the gantry axes. In current practice, an expert will take into account a combination of these parameters when determining how many locations to measure in a grid, but there is no defined procedure. The core of this objective is to define a procedure to perform the same task the controls expert does through experience. Through this process there should be time savings in situations where fewer locations are needed than normal and better characterized uncertainty in situations where more excitation locations are required.

3. **Recommend additional sensors for reducing uncertainty and measuring points of interest during the frequency response measurement.**

   The current identification method measures only the encoders of each axis, but this research will investigate how each axis input affects locations of interest, such as the machine base or the workpoint. A sensor, most likely an accelerometer, will be used to determine the response from each motor to the sensor location. This frequency response data will have implications on how the controller is designed and provide better tools for creating a controller that can improve performance for the system

as a whole. With the addition of these sensors, information can be available to help understand behavior at the workpoint, base or any other location of interest. An important aspect of this objective is the ability to measure multiple sensors at once. Measuring multiple sensors together enables a comparison of the phasing of each response to determine a rough approximation of system behavior in different system directions. Investigating which locations have merit to measure and how they are affected by the system inputs are the primary goals of this objective.

4. **Formalize a procedure for measuring the MIMO frequency response of precision gantry systems that does not rely on operator expertise.**

   Once the previous objectives are complete, there will be another iteration of integrating them into an automated program. The system frequency response will need to be presented such that non-experts will be able to interpret the results and make decisions from them. Displaying these items will be useful in making the procedure usable by a broader audience as well as improving the user experience. More relevant data that is readily available to the user will make the interaction more efficient and less time will need to be spent attempting to understand each system. A well-defined procedure will produce a standard set of results that can be compared against other systems of the same design. Operator variability will be removed from the process, and the procedure will eliminate the need for a controls expert to perform the measurements. A large amount of data is produced in collecting a MIMO frequency response and it can easily be overwhelming to even those familiar with such responses. Therefore the presentation will require a selection of key items that provide insight into the system behavior without being overly complicated. This will draw on the theory of robust control and incorporate principles from this field.

The success and effectiveness of this research project will be determined by how much time can be saved in the data collection process and how well the uncertainty of the system frequency response can be estimated to ensure a robust controller. There is ample opportunity to test the robustness of this process due to the number of different industrial gantries available for testing. There is no single system on which this procedure will be de-

veloped. The systems available for use are customer gantries that were proceeding through the standard testing procedure before shipment. There are a few duplicate systems of the same design but there was not an opportunity to conduct multiple tests on the same unique system. While this caused some difficulty during the development of the research objectives, it ultimately benefits the research outcome. The variability demonstrates that the process will work across multiple systems and not be optimized for one specific system, as might be the case if a dedicated system were available for testing.

1.3 IMPACT

Technicians at the factory will experience a faster measurement process as well as a meaningful presentation of the measured system response. No longer will a large portion of time be spent gathering system responses and adjusting servo tuning, only to have to regather system responses to repeat the process. Most users and technicians that work on these systems use tools meant for single-input single-output systems, which leads to inefficient methods of collecting data for multi-axis systems, especially if the method is not automated. The method proposed in this research will not only automate and expedite the system identification process but it will ensure that all relevant responses are collected. Automating the collection process has the added benefit of removing user based variability on what is collected. The presentation of the system response is another important variable for the technicians. With a more intuitive display, better controllers can be designed because the behavior of the system is more apparent. Ultimately a fully-automated tuning procedure will be developed from the data gathered which will not require human design input apart from the intended goal (contouring, pick and place, etc.).

As a trickle down effect from the factory, customers will receive a system that has robust stability margins. Therefore the system should be less sensitive to changes that occur during shipping or environmental effects at the customer facility. The intent currently is for the customer to be able to initialize the system with little extra work other than plugging in all relevant cables. If a new disturbance is present, then the performance margins of the system

might not be the same as it was at the factory. Better identification processes enable the design of a controller that contains lower uncertainty and therefore better defined stability margins. Not only will it benefit the customer from the factory but also it may assist in on-site controller design if the customer makes a change to the system at their facility. The automated process can be run by a technician or the customer and the system can be tuned on-site or the data sent back to the factory where a controller can then be designed if necessary.

Design engineers will also find use from this procedure in identifying structural resonances and cross-coupling from axes. Currently, full system identification through modal analysis is a long process that requires extensive data interpretation to provide results to the design engineer. With a faster frequency response procedure that presents the data in a meaningful way, future mechanical development on systems can be aided by information on dynamic performance of existing systems without the time requirement of a full modal analysis. Especially beneficial to this process is the extra data collected by sensors at the workpoint or base. This information will help the engineers refine designs to achieve the intended dynamic behavior at points of interest.

2.0 SYSTEM IDENTIFICATION BACKGROUND

The goal of system identification is to develop a mathematical model that approximates the deterministic behavior of a system [22]. Critical to developing this model is minimizing the influence of stochastic effects (noise) on the measurement of the deterministic portion. Noise minimization can be achieved by carefully selecting the locations where the system will be measured as well as what excitation signals are chosen as inputs. Included in all system identification schemes is the design of a signal that is input to the system under consideration. This input signal is compared to an output signal that is measured somewhere else on the system. The correlation in the frequency domain between these two signals is the basic theory behind the system identification considered here.

2.1 FREQUENCY DOMAIN IDENTIFICATION

A common method of describing a system's input/output relationship is the frequency response function or $G(\omega)$, where ω is frequency. The frequency response function is defined as the Fourier transform of the system output divided by the Fourier transform of the system input [19],

$$G(\omega) = \frac{Y(\omega)}{U(\omega)}. \tag{2.1}$$

The Fourier transform is defined as

$$U(\omega) = \mathscr{F}\{u(t)\} = \frac{1}{2\pi}\int_{-\infty}^{\infty} u(t)e^{-i\omega t}dt \tag{2.2}$$

where $U(\omega)$ is a complex quantity. The Fourier transform of the correlation function R_{uu} will then be defined as $S_{uu}(\omega)$ or the two-sided auto-spectral density of the signal $u(t)$.

This function ceases to be complex valued and instead only contains magnitude information. Similarly the Fourier transform of the cross-correlation function R_{uy} can be defined as $S_{uy}(\omega)$ or the two-sided cross-spectral density of the signals $u(t)$ and $y(t)$. Both $S_{uu}(\omega)$ and $S_{uy}(\omega)$ are two-sided or range in frequency from $-\infty$ to ∞. A common or more useful way to define the auto-spectral density and the cross-spectral density is

$$S_{uu}(\omega) = \mathscr{F}^*\{u(t)\}\mathscr{F}\{u(t)\} \tag{2.3}$$

$$S_{uy}(\omega) = \mathscr{F}^*\{u(t)\}\mathscr{F}\{y(t)\} \tag{2.4}$$

It is more convenient to deal with functions that range in frequency from 0 to ∞, therefore one-sided function are defined as

$$S_{uu}^{single}(\omega) = 2S_{uu}(\omega) \quad 0 \leq \omega < \infty \tag{2.5}$$

$$S_{uy}^{single}(\omega) = 2S_{uy}(\omega) \quad 0 \leq \omega < \infty \tag{2.6}$$

Thus the one-sided auto and cross spectral-density functions can be used to define the frequency response of a system as

$$G(\omega) = \frac{S_{uy}^{single}(\omega)}{S_{uu}^{single}(\omega)} \tag{2.7}$$

which is more useful than Equation 2.1 because the effects of measurement noise are reduced [19]. Noise reduction is a rather important topic when performing system identification as noise will add to the uncertainty of the measured frequency response function.

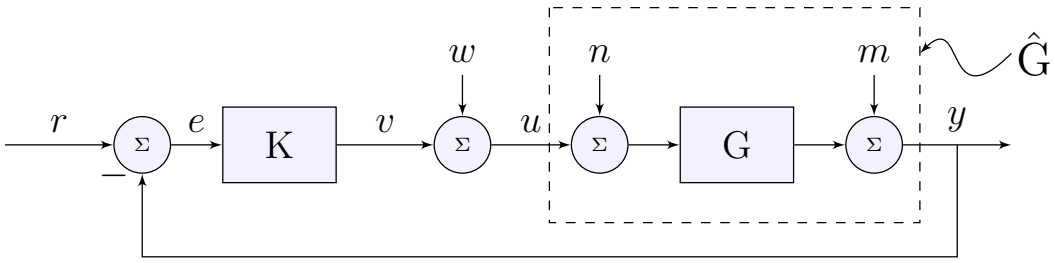


Figure 3: Block diagram with input measurement noise n and output measurement noise m. The frequency response function of G cannot be measured directly and instead $\hat{\text{G}}$ is measured, which includes effects from the input and output noise.

2.2 UNCERTAINTY CONSIDERATIONS

When measuring a frequency response, there is always an uncertainty associated with it. This is due to imperfect sensors, external disturbances and non-linearity in plant mechanics (we assume linearity but this is never the case). Consider the simplified block diagram in Figure 3 which represents a simple linear motion stage. There are sensors that measure the motor current u and the position output y, each exhibiting some type of quantization error and noise. The noise associated with the input and the output are represented as n and m respectively. The noise from the sensors measuring the input and output translates into an uncertain value for the quantity $\hat{G}$, the approximate frequency response function, due to the inability to separate the error and noise from the "true" frequency response. This uncertainty can be visually represented as an error bound around the bode plot of $\hat{G}$, as in Figure 4. This indicates that the true plant G lies somewhere in this bound, but where is exactly is unknown. When using a measurement that contains uncertainty, all possible combinations must be considered in order to ensure a stable controller that satisfies given design conditions. In SISO systems the controller can be designed with increased stability margins in an attempt to account for the unknown uncertainty. This strategy can also be applied to MIMO systems but is much less effective due to coupling between multiple inputs and outputs. For example in a 3 axis system, 9 uncertainty values must be compensated for

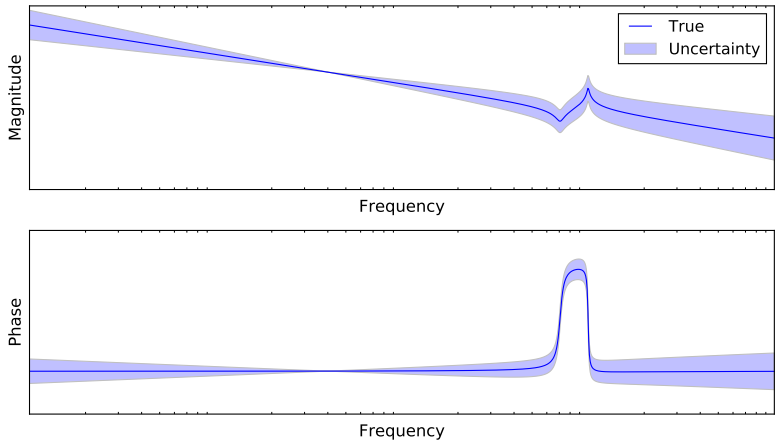


Figure 4: Example frequency response with uncertainty bounds. The true frequency response function may exist anywhere within these bounds.

with increased stability margins. Without knowing the amount of uncertainty or degree of coupling the axes have, the system can easily have an overly conservative controller design. Therefore it becomes necessary to estimate the uncertainty of the frequency response to create confidence in controller design choices.

To estimate the uncertainty bounds of the frequency response of a system, multiple measurements of the system's input and output must be obtained. In this way a variance of the measurement at each frequency can be obtained and translated into uncertainty bounds about a frequency response plot. One such method of analyzing the uncertainty of a frequency response is to input a periodic signal to a system and measure both the input and output for multiple periods. Subdivide the measured signal into distinct periods and then calculate the frequency response of each period. At this point there will be N values of $\hat{G}$, where N is the number of signal periods analyzed. As found in [22] the sample mean and sample variance can be determined for the collection of $\hat{G}_N$ values

$$\hat{G}_s(\omega) = \frac{1}{N}\sum_{k=1}^{N}\hat{G}_k(\omega) \tag{2.8}$$

$$\hat{\sigma}_s^2(\omega) = \frac{1}{N(N-1)}\sum_{k=1}^{N}|\hat{G}_k(\omega) - \hat{G}_s(\omega)|^2 \tag{2.9}$$

where $\hat{G}_s(w)$ is the sample mean and $\hat{\sigma}_s^2(\omega)$ is the sample noise variance of the sample mean. Now a sample mean over multiple periods of an excitation signal has been established with uncertainty in the form of the sample noise variance. While this is helpful to measure and average out noise in the measurements, there are still nonlinearities that can pervade through a single realization of a periodic signal.

To properly account for nonlinearities, it is sometimes necessary to create another realization of the periodic signal that is independent from the first periodic signal. For signals with random phase, this implies unique random phase generated for subsequent realizations. This signal is then applied independently from the first periodic signal in order that a different set of nonlinear effects can be captured in the measurement. Using a separate realization helps to reduce the variance of the frequency response due to nonlinearities by a factor of the number of realizations used [23]. More accurate estimates of the noise distortion can be obtained as well by using an increased number of realizations. The sample mean and sample variance over multiple realizations are calculated in the same way as for multiple periods.

However if N periods and M realizations are used in conjunction, the sample mean and sample variance over the periods should be calculated first for each realization as in Equations 2.8 and 2.9. Then the sample mean and variances for M realizations are combined to yield the total sample mean and variance

$$\hat{G}_T(\omega) = \frac{1}{M}\sum_{j=1}^{M}\hat{G}_{s,j}(\omega) \tag{2.10}$$

$$\hat{\sigma}_T^2(\omega) = \frac{1}{M(M-1)}\sum_{j=1}^{M}|\hat{G}_{s,j}(\omega) - \hat{G}_T(\omega)|^2 \tag{2.11}$$

$$\hat{\sigma}_{T_n}^2(\omega) = \frac{1}{M^2}\sum_{j=1}^{M}\hat{\sigma}_{s,j}^2(\omega) \tag{2.12}$$

where the subscript T denotes total sample mean and variance, $\hat{\sigma}^2_{T_n}(\omega)$ is the sample noise variance and $\hat{\sigma}^2_T(\omega)$ is the sample total variance. The overall uncertainty associated with the measured frequency response is the primary concern therefore only the sample total variance, Equation 2.11, will be considered. This type of uncertainty analysis regarding system identification measurements is common and can also be found in [22], [23], [31], and [33]. Representation of the uncertainty on the plot can take the form of Figure 4 with a shaded uncertain region about the sample mean frequency response as see in [20] or the form of the frequency response estimate $\hat{G}_T(\omega)$ plotted on the same plot as $\hat{\sigma}^2_T(\omega)$ to show the difference in magnitude of the two quantities as in [33].

3.0 EXCITATION SIGNAL DESIGN

Key in designing an excitation signal is to consider the end goal of the system identification. In this research nonparametric system identification is the goal and therefore it is suggested to use excitation signals that will maximize the accuracy obtained in a fixed measurement time for a specified maximum peak value of the excitation signal [22]. In system identification, commonly used excitation signals include swept sine, random noise and multisine signals. Multisine excitation signals contain distinct advantages over the other signals when examining speed and accuracy in tandem and are therefore adopted for the process being developed in this research.

3.1 CANDIDATE SIGNALS

The signals considered have different characteristics that need to be described before the strengths and weaknesses can be examined.

3.1.1 Swept Sine

An excitation signal that is commonly used in system identification is swept sine. A sinusoid containing only one frequency is input for a few signal periods before stepping to a slightly higher frequency sinusoid. This process is repeated until enough frequencies are measured within the frequency range of interest. Multiple periods of each input frequency are necessary

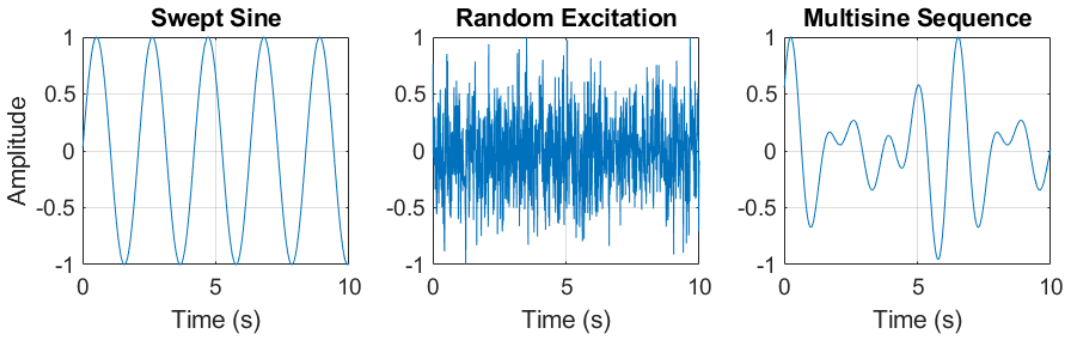


Figure 5: Examples of excitation signals that are considered for use in system identification. The swept sine is a single frequency component of the many frequencies composing an excitation. The random excitation contains all the frequencies of interest but needs to be averaged many times. The multisine sequence is one of multiple sequences needed to measure the entire frequency range.

to obtain a good representation of the signal as well as to let transients decay before signals are measured. The swept sine excitation can be characterized by the sine wave

$$u(t) = A\sin(\omega t) \tag{3.1}$$

where ω varies in some frequency range and is logarithmically spaced and A is the signal amplitude. The amplitude and frequency spacing of the swept sine excitation can be adjusted to better suit the system being identified. An example of one component of the swept sine excitation is shown in Figure 5.

3.1.2 Random Excitation

Another type of excitation signal that is commonly used in system identification is random noise or so called white noise. Wide band random noise is a signal that contains approximately uniform spectral density from low frequencies up to half the sampling frequency. When this signal is input to the system, the entire frequency spectrum contained in the

random noise is excited. Thus the full power of the signal is not contained within the frequency range of interest and other frequencies are present in the signal. In an attempt to reconcile this, windowing functions can be applied to the measured data to limit the spectral leakage, however there will always be some leakage present. The adjustable parameters in a random noise excitation are the frequency range, the RMS amplitude of the signal (peak values as well), and the amplitude distribution. In this work the RMS amplitude of the random noise is controlled and a Gaussian amplitude distribution is used. An example of a random excitation is seen in Figure 5.

3.1.3 Multisine Periodic Excitation

Multisine signals are a classification of signals that are the sum of multiple sine waves of varying frequencies and phases. The formal definition of a multisine signal is

$$u(t) = \sum_{k=1}^{N} A_m \cos(\omega_k t + \phi_k) \tag{3.2}$$

where A_m is the amplitude of each signal being summed, ω_k is the frequency content of the k^{th} sine component and ϕ_k is the phase of the k^{th} sine component. There are multiple classifications of multisine signals that specify different requirements for the amplitude, frequency and phase of each component. A prominent class of this signal is the random phase multisine. This implies that the quantity ϕ_k from Equation 3.2 is randomly distributed across the set of k summed signals. This type of signal is evaluated and used in [6], [7], [8], [31], and [33].

Another subset of multisine signals that function well, while also maintaining compatibility with random phase, is the sequential multisine [14]. A sequential multisine is defined by multiple multisine signals that are applied to the system in sequence. Each multisine sequence contains a subset of the frequency range of interest and once all sequences have been applied to the system, a full range of frequencies is measured. For example if 10-100 Hz needes to be measured, the first sequence could contain a number of frequencies between 10-19 Hz, the next sequence 20-39 Hz, then 40-79 Hz and lastly 80-100 Hz. In this manner the

entire frequency range is covered but harmonics do not interfere with the frequencies measured in each sequence. Each sequence is applied one after another with time for transients to decay between sequences. One sequence of a multisine excitation is found in Figure 5.

According to N. Duncan (personal communication, August 17, 2017) the design of a random phase sequential multisine should begin by dividing the frequency range of interest into non-overlapping sections where the largest frequency in each section is less than twice the minimum frequency in the same section. This prevents harmonic interference in analyzing the frequency spectrum of the signals. With the frequency range subdivided into sections, these become the sequence of multisine signals. The frequencies in each section are chosen such that all frequency components are exactly periodic within the sampling frequency to prevent leakage. Then the sinusoids in each sequence are summed, with each sinusoid having randomly generated phase. The amplitude of each sine wave component is also normalized such that the root mean square (RMS) amplitude of each multisine sequence is equivalent to a specified amplitude. These equations are as follows, using the notation from Equation 3.2

$$A_m = \frac{A_{set}}{\sqrt{N_f}} \tag{3.3}$$

$$\phi_k \in [0, 2\pi) \tag{3.4}$$

where A_{set} is the desired RMS amplitude, N_f is the number of frequencies in the multisine sequence, and ϕ_k is a random phase between 0 and 2π. Using Equations 3.3 and 3.4, and applying them to Equation 3.2, each multisine sequence can be calculated to contain equivalent power per frequency, maintain a set RMS current, contain randomized phase, and have non-overlapping, non-harmonic frequency content.

3.2 EXCITATION ACCURACY AND UNCERTAINTY

The swept sine excitation is an accurate method of measuring the frequency response of a system. Each segment of the excitation only contains a single frequency to analyze. This makes determining the relationship between the excitation input and system output straightforward with a discrete Fourier transform (DFT). The frequency bin of the excited signal is

the only bin of importance, while all other frequency bins contain effects from nonlinearities, noise or harmonics. By comparing the power of the excited frequencies in the output to the power contained in the rest of the frequency bins, a measure of the linearity and signal to noise ratio can be obtained. If the power of the excited frequency dominates the spectrum, then a good measurement is assumed to have been made. The uncertainty can be calculated by repeating multiple measurements and calculating the variance. With the systems studied here, the standard deviation (square root of the variance) is shown to be low, as in Figure 6.

Also shown in Figure 6 are the magnitudes and standard deviations of random excitation and multisine excitations. The random excitation contains more measurement uncertainty than either the swept sine or multisine up until about 300 Hz. The random excitation exhibits larger uncertainty due to spectral leakage and higher crest factors. The spectral leakage allows higher frequencies to alias into the frequency range of interest and contribute energy to the measured frequencies. In addition harmonics can be excited within the frequency range and create further distortion of the measured response. The higher crest factors distort the amplitude spectrum of the signal, leading to uneven power distribution across the excited frequencies [22]. Because the excitations are random, the power of each frequency varies between measurements, leading to an average response that might be close to the actual linear response but has a large standard deviation.

The multisine measurement in Figure 6 contains comparable uncertainty to the swept sine. This is partially due to the elimination of spectral leakage by making all of the excited frequencies exactly periodic with the measurement. Each multisine sequence also contains multiple periods of the frequencies that compose it. So although one period of the multisine is excited, multiple periods of each frequency are being excited. This enables a more accurate measurement to be obtained as more signal periods are included.

Additionally the random phase of each multisine component creates a multisine sequence that has a Gaussian distribution of signal amplitude [23]. In a Gaussian distribution most of the probability mass is centered about zero. A sine wave's amplitude distribution is heavily distributed towards the max and min amplitude, both positive and negative, and random noise has a roughly Gaussian amplitude distribution over the range of signal amplitudes. Similarly with a random phase multisine, the amplitude distribution is centered about the

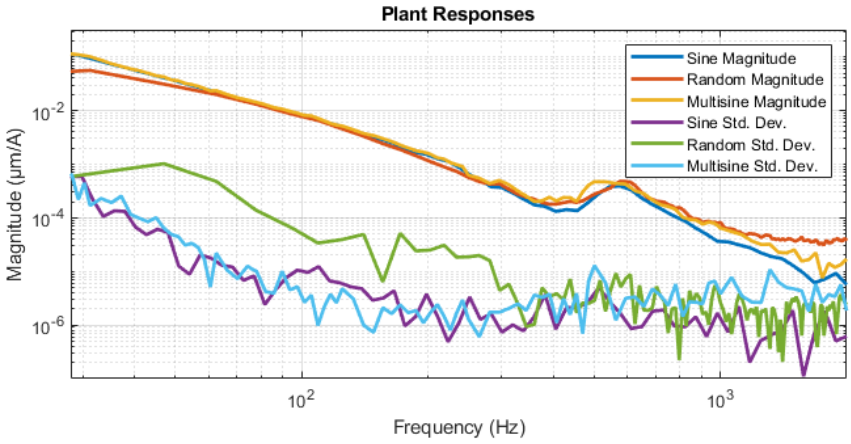


Figure 6: The average frequency response magnitude using various excitation signals and the standard deviation of the measurements. Three measurements of each were averaged to determine an estimate of the sample standard deviation and mean. Swept sine and multisine have a smaller standard deviation than random noise until approximately 300 Hz where noise and external disturbances seem to dominate the response for all excitation signals.

lower amplitudes and therefore better approximates the response at lower, more linear amplitudes. This becomes necessary because non-linearities in a response can occur and often vary with input amplitude. By distributing the signal density in a Gaussian distribution about zero, the linear portion of the response is better approximated rather than being skewed by high amplitude non-linearities [23].

3.3 EXCITATION DURATION

In order to excite a full range of frequencies with swept sine, the total measurement time can be approximated by the amount of time required to excite each frequency for a minimum of one full period and the amount of time required to wait between each frequency excitation

to allow transients to decay. The total measurement time can be approximated by

$$T_{ss} = \sum_{k=1}^{F} \frac{N+M}{f_k} \tag{3.5}$$

where $\frac{1}{f_k}$ is the signal period of the excitation frequency, N is the number of excited periods, and M is the number of periods to wait for transients to decay, as in [22]. This measurement time is clearly increased if the signal to noise ratio is not large enough and multiple periods of measurement are required. T_{ss} typically averages around 80-90 seconds for a single measurement.

The issue of measurement time is exacerbated when measuring the frequency response of a MIMO system. There are not only multiple axes to excite and measure but also position dependencies to consider for the system frequency response. Certain position configurations of the axes alter the frequency response of the system, thus multiple measurement locations are often needed. In the gantry systems under consideration, there are often 9 locations in travel where the frequency response is measured, the center and ends of travel of each axis. At each location, each of the three axes needs to be measured individually, totaling to 27 swept sine measurements. Adding to this time is the movement time between locations, however this should be constant regardless of the excitation signal used. In experiments performed during this research, the time for a complete swept sine measurement totaled 36 minutes, which contained 200 frequencies from 20 Hz to 2000 Hz. The behavior of the system at low signal amplitudes should also be measured, which doubles the measurement time. The low amplitude excitation is useful to understand amplifier and friction nonlinearities and their effects on the gantry. Two complete location grids are measured, one with high amplitude and one with low amplitude for a total measurement time of 72 minutes, which is quite a long time for an industrial system. Due to this long excitation time, the minimum number of signal periods possible is collected at each frequency which does not lend itself to calculating uncertainty because multiple periods are required to obtain a good result for a single DFT. Therefore the number of signal periods would need to be doubled to obtain an uncertainty with respect to two measurements. This has previously not been considered due to the extra measurement time required. Without including some type of uncertainty of the frequency

response, some interpretation is left to the user on the validity of the frequency response measured.

A random noise excitation signal contains the entire frequency spectrum of interest so measurement times of random noise can be fast. In order to obtain data with acceptable levels of accuracy, about 8 to 12 averages over the signal are necessary. This however only takes about 10 seconds per individual frequency response measurement. For a grid with the same size as swept sine and containing a high and low amplitude, the total measurement time is about 9 minutes.

A multisine excitation is a compromise between the quick response of the random noise excitation and the sluggish swept sine. Using the predetermined signal characteristics of the multisine signal (sequential, random phase) the measurement time can be approximated as

$$T_{ms} = \sum_{i=1}^{P} T_P(N + M) \tag{3.6}$$

where T_P is the period of each multisine sequence, P is the number of multisine sequences, N is the number of executed periods, and M is the number of periods to wait for transients to decay. Measuring a full grid of responses as with swept sine and random excitation takes 9 minutes or 18 minutes if two full grids of measurements are taken with high and low amplitudes. The second grid of responses is no longer necessary for multisines because the nonlinearities of the system can be examined with just one excitation amplitude. The harmonics and nonlinear effects can be studied by examining the other frequency bins in the DFT [23]. The ratio of power in the excited frequency bins to the power in all other bins will determine how linear the system is, with a high ratio signifying greater linearity. Without this second grid of excitations, the measurement time is again reduced to 9 minutes. To further reduce the measurement time, simultaneous axis excitations can be performed at each location in travel. Instead of needing to excite each axis individually, now all axes can be simultaneously excited so that only one measurement per location is necessary. This again reduces the measurement time by 66%. However in order to measure all axes simultaneously, no frequencies can overlap between axes. This results in a so called zippered multisine where the frequencies of a sequential multisine are divided between each axis [23]. Figure 7 is an example of how a zippered multisine is composed. Following this composition, the

frequency resolution of the response of each axis is reduced. In order to maintain a suitable frequency resolution, the excitation time must increase to measure more frequencies. In order to measure more frequencies, the period length of each sequence must increase. The increased sequence period length allows more frequencies to have exact periodicity within a sequence period, thus satisfying previous multisine design criteria. Increasing the number of frequencies to a suitable level effectively doubles the excitation time. This is done by doubling the frequency resolution, however the zippered resolution is two-thirds of the original frequency resolution. From an original measurement time of 9 minutes, simultaneous excitation reduces this to 3 minutes upon which an increased frequency resolution increases this again to 6 minutes. This proves that indeed multisine is faster than swept sine and comparable to the measurement duration of a random noise excitation.

3.4 AVERAGING METHODS

For certain systems that exhibit large nonlinearities it is sometimes deemed necessary to collect multiple repeated periods of an input signal and the response or to conduct independent experiments. The need to average over multiple periods or independent experiments arises when systems are exhibiting non-linearities as mentioned in Section 2.2. Figure 8 shows the different variances associated with each method. The system measured here is a good representation of the gantry style this procedure will mainly be used on, mechanical bearing linear stages. The slight differences in uncertainty are not significant enough to justify doubling the excitation time by averaging over independent realizations. Averaging over periods is necessary to obtain an estimate of the uncertainty value and for this reason, averaging over at least two multisine periods is recommended. However the ability to increase the number of periods or realizations averaged will be built into the measurement procedure.

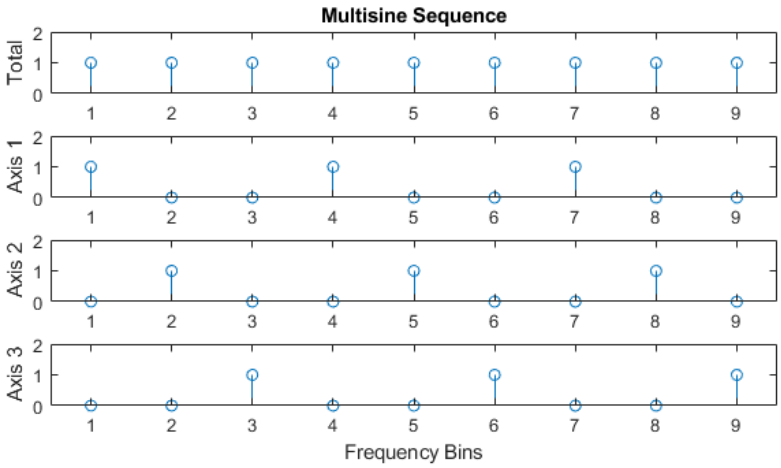


Figure 7: An illustration of how a zippered multisine excitation is composed. If a multisine sequence originally contains 9 frequency components, the frequencies are alternated between each axis until the excitation signal for each axis contain 3 unique frequencies. In this way all excitations can be conducted and measured simultaneously without interfering with one another.

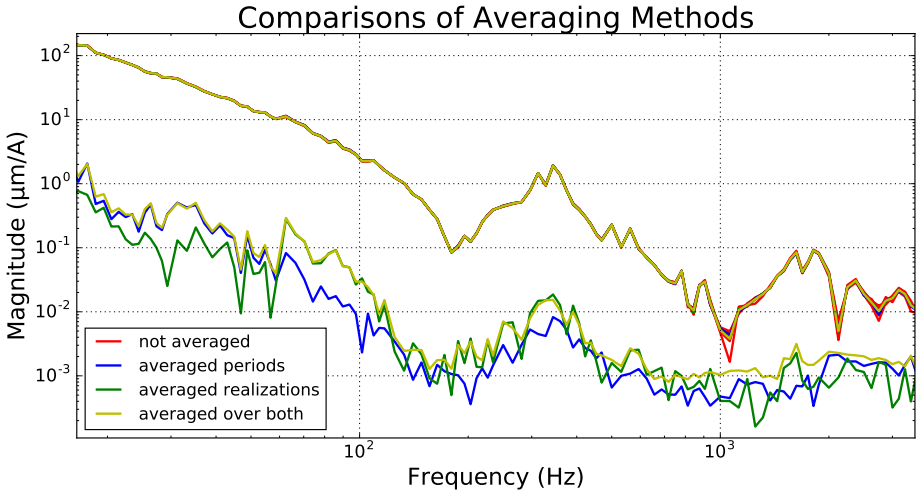


Figure 8: Comparison of averaging multisine excitation, plant frequency response over 3 periods, over 2 realizations, or over both. Standard deviation is shown for each method below the magnitude.

Table 1: Comparison of the necessary measurement duration with different excitation signals. All are compared to the current practice of swept sine. Multisine out performs the other two signals when considering time reduction and measurement uncertainty.

Excitation Signal	**Measurement Duration**	**Time Reduction**	**Uncertainty**
Swept Sine	86 minutes	-	Low
Random Excitation	9 minutes	90%	High
Multisine	6 minutes	93%	Low

3.5 SIGNAL COMPARISON

Here it is shown that the multisine excitation signal has appealing characteristics when compared to swept sine and random noise excitations. The multisine excitation produces a measurement with less uncertainty than random noise and comparable uncertainty to swept sine, as in Figure 6. The measurement with multisines is also faster than swept sine and comparable to random noise, even slightly faster depending on frequency spacing. Table 1 compares the measurement duration and uncertainty of the three excitation signals. It is apparent that the multisine signal outperforms the other two signals with regards to measurement duration and uncertainty and therefore is used in the identification process being developed.

4.0 DATA PROCESSING AND PRESENTATION

For a MIMO system, system identification and analysis generates a large amount of data. Wrapped within this data is information about system stability and performance but various forms of processing are required to transform the data into something meaningful. As in a SISO system, identifying the plant, loop gain, and sensitivity transfer functions provides an array of information with which to understand the MIMO gantry system. This chapter discusses the considerations for accurate plant identification, decoupling techniques, and closed loop analysis methods. These topics are all useful when creating a presentation of measurements that are readily understood and useful to an operator.

4.1 PLANT IDENTIFICATION

Plant identification is fundamental when controlling and analyzing the performance of a system. Ideally when identifying a system's plant, the identification is done in open loop, a condition where there is no feedback from the system outputs. Sometimes it is not possible to identify a system in open loop because the system will incrementally drift away from the reference position towards an unsafe operating position [3]. This could be an end of travel or potential crash condition. Therefore the control loop must stay closed and use feedback to maintain safe operating conditions.

In SISO systems, measuring a frequency response of the system in closed loop is simple to perform. Apart from external disturbances, there is a single, well defined loop that the

system operates within, Figure 3. If access to each of these signals is available, identification is as straightforward as computing the cross spectral density of the input with the output and dividing by the auto spectral density of the input as in Equation 2.7.

In MIMO systems measuring the frequency response of the system in closed loop is much more complicated. It becomes difficult to isolate an input signal's influence on an output due to multiple signals affecting each output. Take for example a 3-input 3-output system where the plant frequency response function is being measured. If the notation from Figure 3 is used and noise is neglected, the equation

$$y = Gu \tag{4.1}$$

represents the relationship between the input signals $[u]$ and the output signals $[y]$ for axes X, XX and Y respectively. Expanding this and solving for y yields the linear system of equations

$$\begin{aligned} y_1 &= G_{11}u_1 + G_{12}u_2 + G_{13}u_3 \\ y_2 &= G_{21}u_1 + G_{22}u_2 + G_{23}u_3 \\ y_3 &= G_{31}u_1 + G_{32}u_2 + G_{33}u_3 \end{aligned} \tag{4.2}$$

where G_{ij} denotes the transfer function of input j to output i. The dependency of the outputs y on all inputs u makes it difficult to isolate one output with respect to just one input. The normal method of measuring the outputs with respect to one input is operating open loop and only input one excitation while keeping the other inputs zero. In this manner the outputs y_1, y_2, or y_3 can be measured with influence from only one input, u_1, u_2, or u_3. In open loop it is straightforward to calculate each element of the plant transfer function matrix G.

However as mentioned before, it is necessary to operate the system in closed loop so that safe operating conditions are maintained. Now it is not so simple to separate individual input effects from the outputs. In an attempt to populate the plant transfer matrix G, an excitation can first be applied to input 1 and the outputs are measured. The excitation is applied to the input of the plant through signal w, as in Figure 3, which allows direct access to alter the input to the plant. Although inputs 2 and 3 do not have an applied disturbance,

the controller applies current to hold the axes in place and reject disturbances which is the cause of the interference. This is repeated in the same manner with input 2 and 3 until a total of 9 measurement pairs are generated. By calculating the frequency response of each measurement pair, the corresponding transfer function elements can be calculated for the plant transfer matrix G. The resulting 3x3 matrix is the estimated plant transfer matrix $\hat{G}$, where

$$\hat{G} = G \pm \Delta_g \tag{4.3}$$

G is the "true" plant transfer matrix and Δ_g is the uncertainty matrix associated with the plant. Δ_g can have both positive and negative values for which $G + \Delta_g$ is the upper bound of the frequency response function and $G - \Delta_g$ is the lower bound of the frequency response function. For example the element in the first row and column

$$\hat{G}_{11} = G_{11} \pm \Delta_g^{11}$$

includes uncertainty, and lumped into this uncertainty term are noise, nonlinearities and coupling from other axes. Element $\hat{G}_{11}$ is measured by correlating input u_1 with output y_1, however as shown in Equation 4.1, output y_1 contains influence from the inputs u_2 and u_3 due to plant coupling dynamics $\hat{G}_{12}$ and $\hat{G}_{13}$. Inputs u_2 and u_3 are non-zero because the system is operating in closed loop. Therefore when the disturbance is input to the system through input 1, not only does the controller reject the disturbance through input 1 but also through inputs 2 and 3.

For some systems the cross-coupling components $\hat{G}_{12}$ and $\hat{G}_{13}$ are small enough that no matter the values of u_2 and u_3, the primary influence on y_1 is $\hat{G}_{11}u_1$ and the system is effectively three independent SISO systems. For the gantry systems considered here, the cross coupling components are not negligible, especially between the two gantry axes, X (axis 1) and XX (axis 2). Although the coupling terms $\hat{G}_{12}$ and $\hat{G}_{13}$ may not be as large as the component $\hat{G}_{11}$, if the controller for each axis has high gain, this can cause all inputs to be large in an effort to reject the disturbance. This would yield an output y_1 that is affected by all of the terms in Equation 4.1. Large inputs multiplied by non-negligible cross-coupling

components yields an effect on the output that cannot be ignored. Thus when correlating the input u_1 with y_1, a large influence from other axes is associated with frequencies at which the controller has high gain. This follows for the other outputs y_2 and y_3.

Figure 9 plots the diagonal terms of a gantry plant frequency response matrix or $\hat{G}_{11}$, $\hat{G}_{22}$, and $\hat{G}_{33}$ respectively. At low frequencies for the plot of the gantry axes $\hat{G}_{11}$ and $\hat{G}_{22}$ (green and blue) the magnitude is flat before assuming the expected slope of -40 dB/decade. This spring like response at low frequencies for $\hat{G}_{11}$ and $\hat{G}_{22}$ is due to high controller gain at low frequencies on axis 1 and 2 (X and XX respectively) and the coupled nature of the gantry axes (large $\hat{G}_{12}$ and $\hat{G}_{21}$ terms). Therefore the influence from the opposing gantry axis on each of the gantry axes plant responses is large until the frequency at which the cross coupling terms become small with respect to the diagonal term $\hat{G}_{ii}u_i$, where i = 1,2,3. In Figure 9 this can be considered to be about 40 Hz. $\hat{G}_{13}$ represents the effect of the bridge input on the output of one gantry axis and is assumed to be small relative to $\hat{G}_{12}$ and $\hat{G}_{11}$. This is assumed due to the bridge being orthogonal to the gantry axes as well as from analyzing the plot of $\hat{G}_{33}$ in Figure 9. There is little to no interference in the bridge plant response from the gantry axes at frequencies where the controller is known to have large gain. This implies that for the response of u_3 to y_3 the values of $\hat{G}_{31}$ and $\hat{G}_{32}$ are small. $\hat{G}_{13}$ and $\hat{G}_{23}$ can be assumed to be small as well due to reciprocity. Thus the majority of the influence on each gantry axis (X and XX) plant measurement in closed loop originates from the opposing gantry axis.

A useful strategy to minimize the contributions from axis coupling is to minimize the values of u_2 and u_3. This can be done by designing a controller that is able to maintain safe operating conditions but has low gain, relatively. The controller is able to maintain stability and hold the system to a set point, however the frequency range of the system where the controller gain is large is minimal. This technique was adopted and yielded an improvement in minimizing the controller influence on the measured plant frequency response however some influence is still present, just at lower frequencies.

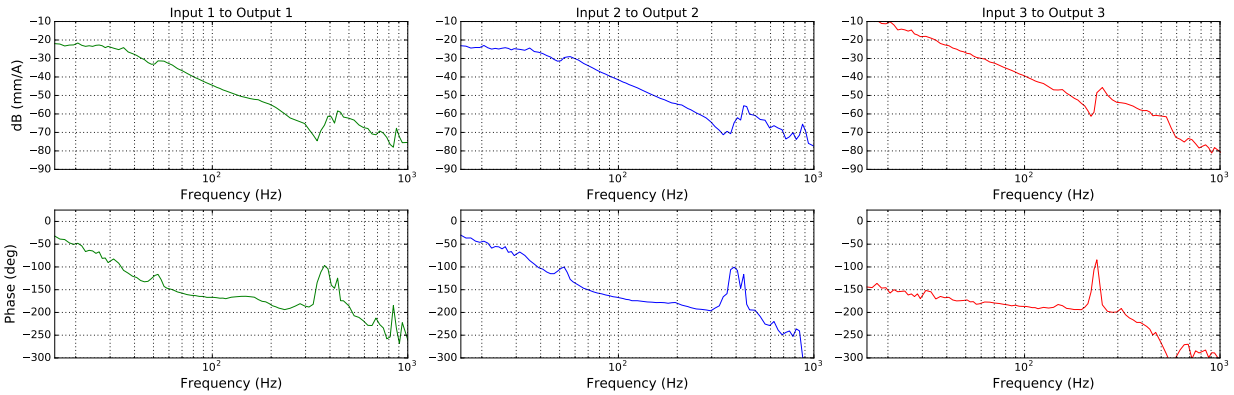


Figure 9: Diagonal responses showing the influence of axis coupling on the identified plant. Large coupling terms influence the output of axes 1 and 2 (X and XX respectively) from low frequencies until about 40 Hz. The phase roll off seen here is due to the sampling delay of the system.

4.2 AXIS DECOUPLING

Coupling between axes is present in most gantry systems and complicates system identification and performance analysis. The primary coupling occurs between the two gantry axes because they move along the same axis of motion and are rigidly linked. The rigid coupling of these two axes is intentional as it reduces yaw errors at the workpoint. With these benefits though, the control of the two gantry axes X and XX is also affected. Each gantry axis experiences influence not only from its own input but from the other gantry axis input. This influence causes uncertainty in stability and performance measurements of each axis when measured in closed loop. The uncertainty in the measurements necessitates a conservative controller to meet required metrics, otherwise an iterative identification approach is required to search for an optimal controller to meet these requirements [24].

Another approach that improves controller synthesis is the design of a coordinate transform to diagonalize the MIMO plant matrix [26]. A diagonalized plant matrix ideally contains negligible off diagonal or coupling terms. The plant matrix then signifies that each output

is only affected by a single input, which is referred to as a decoupled system. Decoupling via kinematic transformation [13], [15], [11], singular value decoupling of symmetric subsystems [26], and tensor decoupling methods [28] are the methods of decoupling considered in this research. To examine the effects of the various decoupling methods, first a simplified gantry model will be studied. This is emulated by connecting two motors with a flexible coupling and driving the motors in the same manner as linear gantry axes (X and XX).The dynamics are greatly simplified and only exhibit a torsional resonance to recreate the effect of a yaw resonance on a linear gantry system. Figure 10 shows a model of the setup used to demonstrate the decoupling methods.

In the kinematic decoupling of a gantry system, a simple kinematic gantry model is needed to create the decoupling matrices. A 2×2 subset of the plant matrix containing elements G^{11}, G^{12}, G^{21}, and G^{22} is considered because as discussed before, the coupling with the bridge axis is negligible for the gantry systems being considered, at least in the frequency range of interest. This assumption is also made at this moment to demonstrate the effect of decoupling on highly coupled axes. In Figure 11 the simple model of the two gantry axes X and XX driving the bridge is shown. It is desirable to find matrices that transform the standard coordinate system into a different physical coordinate system that attempts to diagonalize the plant matrix. As in [15] an R, Θ, Y transformation will be used to attempt to achieve minimal coupling of the axes. The new coordinates R and Θ are also shown in Figure 11, while the Y axis maintains the same coordinates. The original plant matrix

$$\hat{G} = \begin{bmatrix} \hat{G}_{X/X} & \hat{G}_{X/XX} & \hat{G}_{X/Y} \\ \hat{G}_{XX/X} & \hat{G}_{XX/XX} & \hat{G}_{XX/Y} \\ \hat{G}_{Y/X} & \hat{G}_{Y/XX} & \hat{G}_{Y/Y} \end{bmatrix} \tag{4.4}$$

can be transformed by generalizing that the bridge axis is always midway between the gantry axes as well as assuming that the bridge axis motor applies a force at the center of mass of the bridge axis. From these assumptions the transformation matrices are generated to be

$$T_y = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad T_u = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.5 & -0.5 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.5}$$

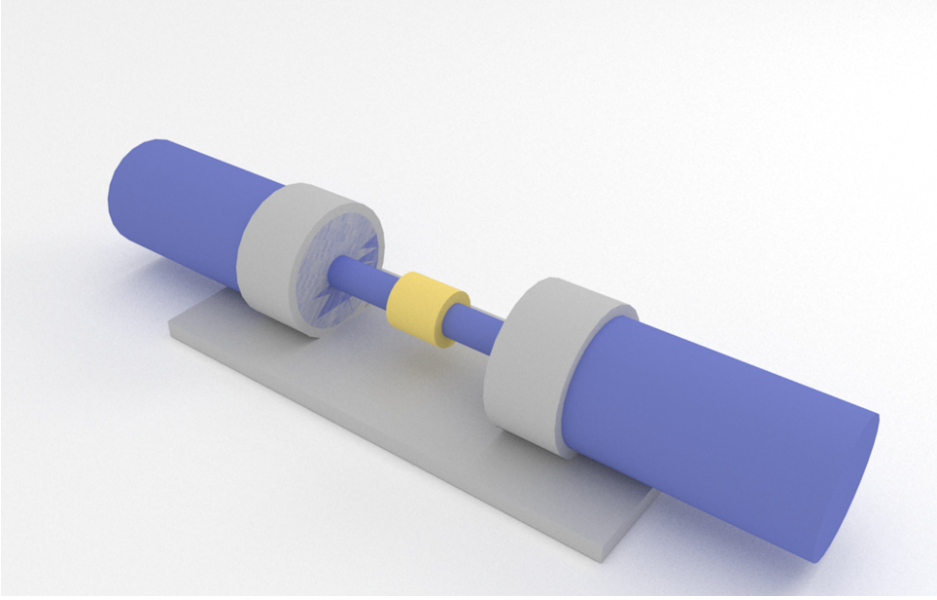


Figure 10: Gantry axes X and XX simulated with two rotary motors connected by a flexible coupling. This generates ideal dynamics with a torsional mode simulating the yaw mode of linear gantries.

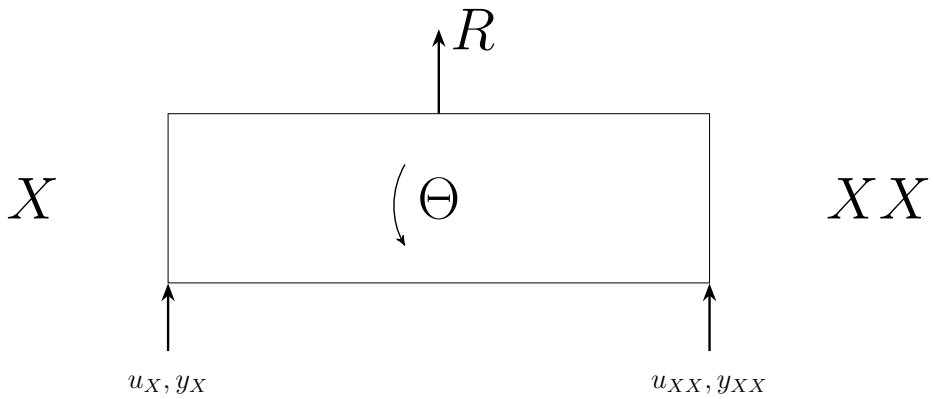


Figure 11: Schematic model depicting the gantry axes X and XX and the kinematically transformed coordinates R and Θ

where T_y is the output transformation matrix and T_u is the input transformation matrix. These two matrices are formed from the relationship of the existing coordinate system to the new coordinate system, as presented in [15]. The equations below demonstrate how the matrices transform the existing coordinates into a new coordinate system,

$$\begin{bmatrix} y_R \\ y_\Theta \\ y_Y \end{bmatrix} = T_y \begin{bmatrix} y_X \\ y_{XX} \\ y_Y \end{bmatrix} \qquad \begin{bmatrix} u_R \\ u_\Theta \\ u_Y \end{bmatrix} = T_u \begin{bmatrix} u_X \\ u_{XX} \\ u_Y \end{bmatrix}. \tag{4.6}$$

Multiplying these two matrices with the original plant matrix yields the transformed plant matrix

$$\hat{G}_{decoupled} = \begin{bmatrix} \hat{G}_{R/R} & \hat{G}_{R/\Theta} & \hat{G}_{R/Y} \\ \hat{G}_{\Theta/R} & \hat{G}_{\Theta/\Theta} & \hat{G}_{\Theta/Y} \\ \hat{G}_{Y/R} & \hat{G}_{Y/\Theta} & \hat{G}_{Y/Y} \end{bmatrix} = T_y \hat{G} T_u. \tag{4.7}$$

Applying this transformation yields a coordinate system where R is the motion in the direction of the X and XX axes, Θ is now the rotation about the center of the bridge axis (in linear units) [15] and Y is still the bridge axis. More complicated decoupling schemes can account for the location of the bridge axis but in the operational frequency range of this system, the coupling between the bridge and other axes can be assumed to be acceptable for now.

Another transformation that can be used to decouple the highly coupled gantry axes is based on singular value decomposition. Singular value decomposition is a process by which a matrix can be diagonalized [18]. These diagonalized matrix elements are known as singular values and are very useful in MIMO system analysis. The transformation is done much the same way as the previous method however the plant decoupling matrices are generated using singular value decomposition at a specific frequency. This frequency is normally chosen to be within the expected bandwidth of the system to maximize the decoupling within the frequency range of interest [26]. The singular value decomposition is defined as

$$\hat{G}_o = U_o \Sigma_o V_o^T \tag{4.8}$$

where $\hat{G}_o$ is the value of the plant matrix $\hat{G}$ at the chosen frequency within the bandwidth, and U_o and V_o are the left and right singular vector matrices of the singular value matrix Σ_o. For this particular example $\hat{G}_o$ is the plant transfer matrix at 48 Hz, which is within the operational bandwidth of this system and should generate a useful decoupling. So Σ_o ends up being the diagonalized plant matrix at 48 Hz. The matrices U_o and V_o are multiplied with the plant matrix $\hat{G}$ to yield the decoupled matrix

$$\hat{G}_{decoupled} = U_o^{-1}\hat{G}V_o^{-T} \tag{4.9}$$

$\hat{G}_{decoupled}$ is now the decoupled frequency response which was transformed with the left and right singular value matrices generated at 48 Hz which are

$$U_o^{-1} = \begin{bmatrix} -0.6133 + 0.3511i & -0.6144 + 0.351i \\ 0.388 + 0.5918i & -0.3869 - 0.5914i \end{bmatrix} \quad V_o^{-T} = \begin{bmatrix} 0.7091 & 0.7051 - 0.0025i \\ 0.7051 & -0.7091 + 0.0026i \end{bmatrix}.$$

The tensor decoupling method proposed by Stoev, Oomen and Schoukens is useful in that neither a kinematic model of the system nor a chosen frequency are needed. Tensor decoupling is performed with canonical polyadic decomposition which can be considered a generalization of singular value decomposition, further detail is found in [28]. The 2×2 plant matrix, along with a randomized first attempt at a factor matrix is input to the function `cpd3_sd` from the TensorLab toolbox [32]. This results in a decoupling of the plant as in

$$\hat{G}_{decoupled} = T_y\hat{G}_{coupled}T_u \tag{4.10}$$

where $\hat{G}_{decoupled}$, T_y and T_u are produced from the function `cpd3_sd` in the TensorLab toolbox. The two decoupling matrices are defined as

$$T_y = \begin{bmatrix} -0.7067 & -0.7203 \\ -0.7075 & 0.6937 \end{bmatrix} \qquad T_u = \begin{bmatrix} -0.6963 & -0.7057 \\ -0.7177 & 0.7085 \end{bmatrix}. \tag{4.11}$$

It is interesting to compare the matrices produced from the tensor decoupling method to those produced from a singular value decomposition. There is similarity especially between V_o^{-T} and T_u but further study is needed to understand the exact mechanism for similar

transformation matrices. This method has proved to be comparable to the decoupling based on a kinematic model for gantry systems. One major difference from kinematic decoupling though is that the tensor method does not necessarily decouple the system into axes that have a physical representation.

The tensor decoupling method, kinematic decoupling method, and singular value decomposition method all yield similar results as shown in Figure 12. The plot is a comparison of each decoupling method to the original 2x2 subset of the gantry plant. The response from input 1 and output 1 and from input 2 to output 2 are all very similar, so much so that the lines for the singular value response and kinematic response are covered by the response from the tensor method. This suggests that the kinematic decoupling model for R and Θ is the closest physical model configuration to achieving a diagonalized plant matrix. All decoupling methods yield a minimum of 25 dB more decoupling on the off diagonal terms until the torsional resonance of the system. At this point the axes become coupled and the decoupling methods lose effectiveness. This is not an issue as the bandwidth of the system lies below this resonance. Therefore all methods are proven effective on an ideal system.

With the methods proven in theory, it is now important to test them on actual gantry systems. Figure 13 is the result of testing the methods on an actual gantry. The kinematic transformation matrices are the same as testing on the ideal system, however the singular value and tensor matrices are recalculated for this new system. All methods successfully decouple the response from input 1 to output 2 and achieve roughly 30 dB more decoupling. The main differences in the system occur from input 2 to output 1 where the tensor decoupling method is not as successful as the kinematic or singular value method. This comparison of systems is performed on multiple other gantry systems and similar results were obtained. The singular value method is useful if there is not model knowledge of the system and similar results to kinematic decoupling are desired. However a frequency within the bandwidth of the system must be chosen at which to calculate the decoupling matrices, which requires some knowledge of the system. If no knowledge of the system is available, the tensor method can generate decoupling matrices that decouple the system but are not guaranteed to function as well as singular value or kinematic model methods. The gantries that will be identified by this procedure are all of a similar type and can be described by the

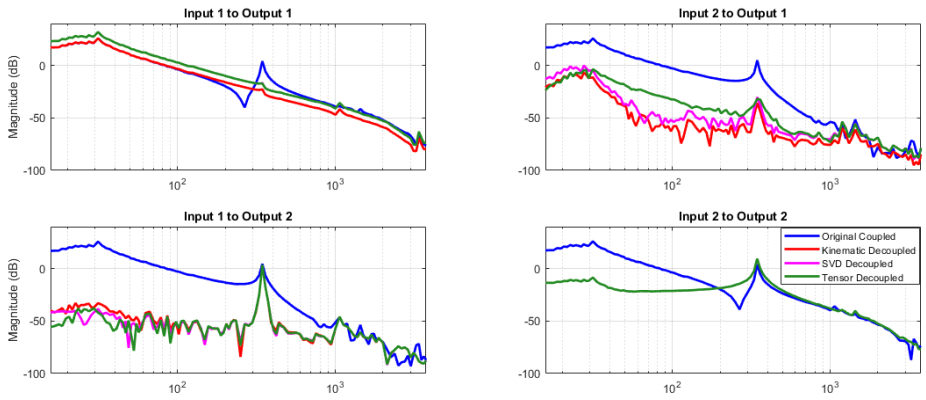


Figure 12: 2×2 plant decoupling method comparison as tested on the ideal setup shown in Figure 10. All methods yield similar decoupling results.

kinematic model considered here. Therefore the kinematic transformation is preferred as it is grounded in a physical model and creates a new coordinate system with which the system can still be readily understood.

If the gantry axes X and XX are sufficiently coupled within the frequency range of interest then a kinematic transformation can be used to simplify the controller design for the system. In all gantry systems tested in this research, there is significant coupling inherent in the mechanical design of the system. So it becomes assumed that X and XX are sufficiently coupled but what is unknown is the effectiveness of the decoupling scheme. Therefore a metric is needed to signify the effective decoupling by kinematic transformation. This can take the form of a plot of the ratio of magnitudes of the diagonal terms to the non-diagonal terms before and after the decoupling transformation. Figure 14 is a comparison on the ideal setup which shows that after the decoupling the response of the diagonal terms is 30 dB greater than the response of the non-diagonal terms in the operational frequency range, where as the diagonal and non-diagonal terms were equivalent in magnitude before decoupling. This again should be demonstrated on an actual gantry system, which is seen in Figure 15. Here the response is more complicated but it can be seen that the non-diagonal

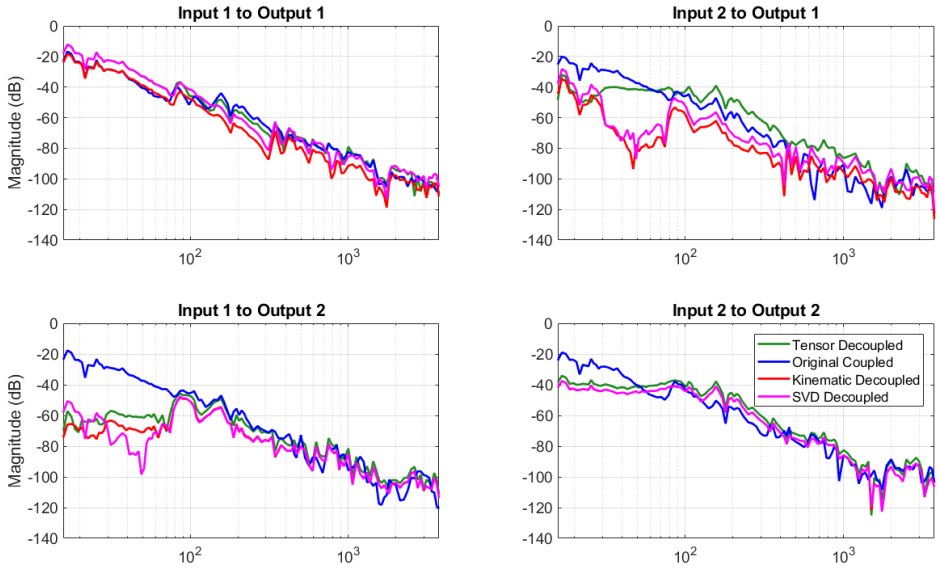


Figure 13: 2×2 plant decoupling method comparison as tested on an industrial gantry. The kinematic and singular value based methods out perform the tensor based method.

terms of the gantry system are 20 dB lower or about 10 times smaller than the diagonal terms. This is an improvement over the 1 to 1 ratio of the original system. A suggested metric for determining if a decoupled transformation is effective is to verify that within the operational bandwidth of the system, the ratio of non-diagonal terms to diagonal terms are 20 dB lower for the decoupled system than the original system. If the decoupled transformation is useful, then all other procedures may be carried out using the new decoupled coordinates. The only consideration that must be made is to apply the transformations to the inputs and outputs of the gantry once a controller has been designed. This is outside the scope of this research but is left as a relevant note for future work.

Identifying the plant transfer matrix is fundamental in the system identification of a MIMO system. The dynamics of the system with regards to the inputs and outputs are contained within this matrix. The plant transfer matrix is also useful in forming other transfer functions that describe the system. The controller matrix is also assumed to be

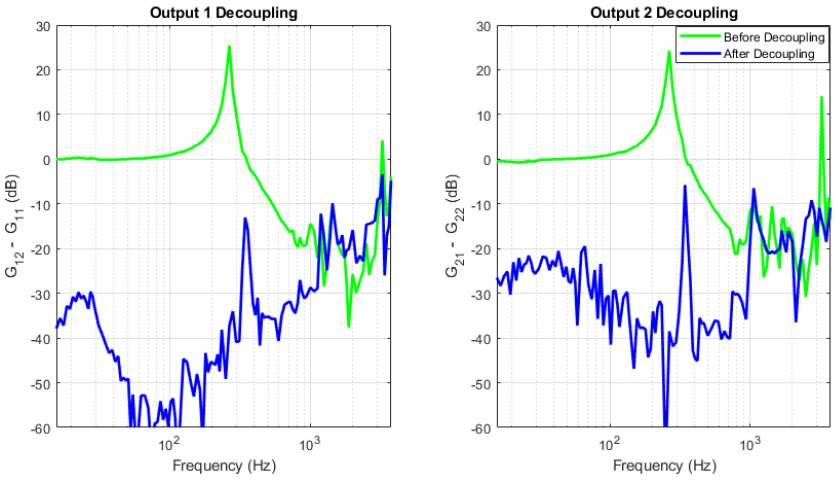


Figure 14: Amount of decoupling achieved with a kinematic transformation on the ideal setup in Figure 10. The ratio of the diagonal terms and non-diagonal terms shows the reduced influence of coupling on the diagonal terms of the plant matrix. The ratio being small shows that a more decoupled system has been achieved.

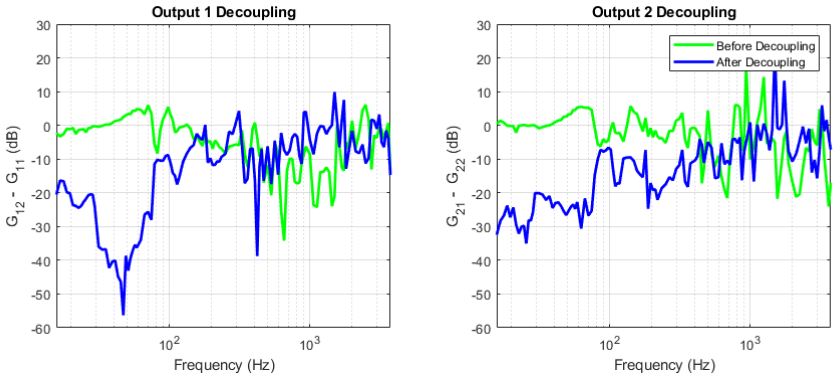


Figure 15: Amount of decoupling achieved with a kinematic transformation on the same system as Figure 13. The ratio of the diagonal terms and non-diagonal terms shows the reduced influence of coupling on the diagonal terms of the plant matrix. The responses are not as decoupled as the ideal setup however significant decoupling is achieved through the use of a kinematic transformation.

known and between these two matrices many functions can be calculated that describe the robustness and performance of the system. However if either of the two matrices have associated uncertainty, this gets carried through the calculations into the robustness and performance measurements. As discussed previously in this section, the plant measurement is influenced at frequencies where the controller has large gain. Apart from operating open loop, which was eliminated as a possibility, one solution is to reduce the controller bandwidth as low as safety constraints will allow in order to minimize the frequency range where the controller has large gain and thus exerts influence on the plant frequency response matrix through coupling terms. However there are still other sources of uncertainty that need to be characterized before analysis can begin.

4.3 UNCERTAINTY ANALYSIS

Measurement uncertainty was the primary consideration for sources of uncertainty in this research. It was expected that a large portion of the uncertainty in the frequency response measurements would be due to sensor noise, external disturbances and amplifier nonlinearities. However the measurement uncertainty yields negligible contributions to the frequency response, except for high frequencies. Analyzing the response collected in Figure 8, the standard deviation of the response is 0.5% of the magnitude up until about 500 Hz. This is a difference of 46 dB which proves that measurement uncertainty is clearly a minimal influence on the variability of frequency response measurements for these systems. At higher frequencies the standard deviation begins to approach the magnitude of the response but this is to be expected [9]. As frequencies increase, the amplitude of induced motion decreases resulting in a lower signal to noise ratio. These frequencies are much higher than the bandwidth of the systems though and as a result can be filtered to minimize their influence without affecting system performance significantly.

The primary source of uncertainty for gantry systems is the variability in the frequency response throughout travel. The frequency of resonances shift slightly as the axes move throughout travel and the inertia experienced by the two gantry axes X and XX changes as

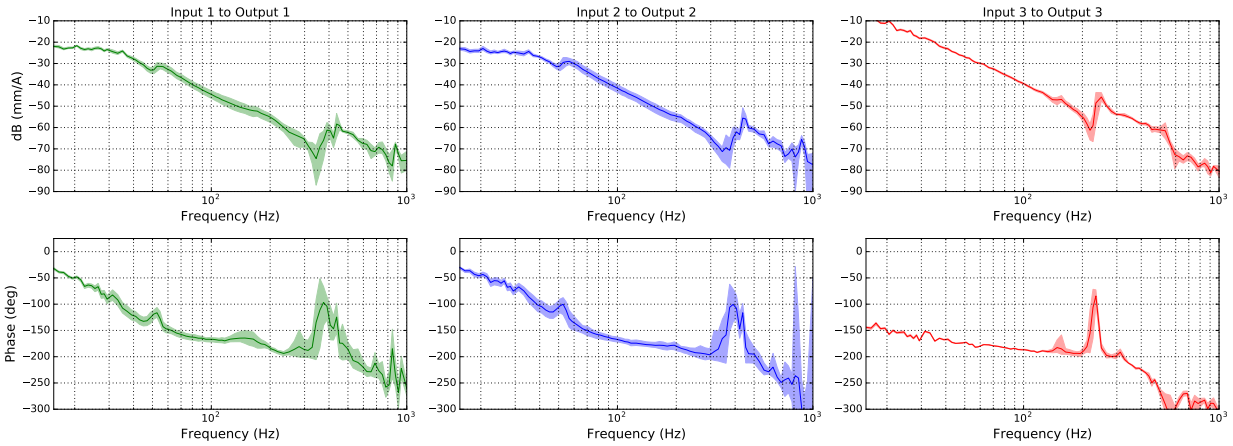


Figure 16: The uncertainty of the plant frequency response due to variations in gantry travel. Resonant frequencies shift and inertias change which create a band of possible plant responses for the system.

the location of the bridge axis Y changes. If all of the various responses are overlain, a band of responses, as in Figure 16 is created. This band of responses represents an estimation of all possible frequency responses for the plant of the gantry. This is useful in that a controller can be designed with a multitude of plant responses in order to study the performance of the entire system. If SISO methods were to be used here, this would entail designing a controller and applying it to the measured plant responses. Whether the loop gain, sensitivity, or other responses were analyzed, a total of 9 plots with uncertainty bounds would need to be studied. This is a large amount of data to consume which will make controller design tedious and as will be seen in the next section is not always accurate. Therefore other methods need to be adopted to simplify the analysis of the performance and robustness of a precision gantry.

4.4 PERFORMANCE AND ROBUSTNESS ANALYSIS

The end goal of system identification is to be able to make a statement with a certain level of confidence about system performance. In MIMO system identification a few metrics can be used to ensure the desired performance is being achieved. The sensitivity function and the complementary sensitivity, defined in Equations 4.12 and 4.13 [12], become useful in evaluating performance.

$$\text{Sensitivity} = S = (I + GK)^{-1} \tag{4.12}$$

$$\text{Complementary Sensitivity} = T = GK(I + GK)^{-1} \tag{4.13}$$

First begin with Figure 17 where the signals d and n are disturbances and sensor noise respectively. In order to make sure there is sufficient rejection of the disturbance d, the sensitivity $(I + GK)^{-1}$ must be small. If the sensitivity is small this implies that the return difference or $I + GK$ is large. With I being the identity matrix, GK can be considered to be large which leads to high loop gain and disturbance rejection. A supplemental method of considering the need for low sensitivity in the frequency range of interest is to consider the system output[26].

$$y = Tr + Sd - Tn \tag{4.14}$$

The output y depends on the complementary sensitivity with respect to the reference input and sensor noise, as well as the sensitivity with respect to disturbances. It is desirable for the output to be dominated by the Tr term which requires the Sd and Tn term to be small. Therefore for good disturbance rejection in the system loop, S must be small within the operational frequency range. This same output equation is useful in motivating the requirements for noise attenuation as well as reference tracking. In order for the contribution from n to be negligible at the higher frequencies noise usually contains, the complementary sensitivity at these frequencies should be nominally zero. This normally translates to frequencies higher than the system bandwidth and reduces the effect of high frequency noise on the system output. In the frequency range of operation, from DC to the system bandwidth, T should be approximately 1 in order track the reference input r.

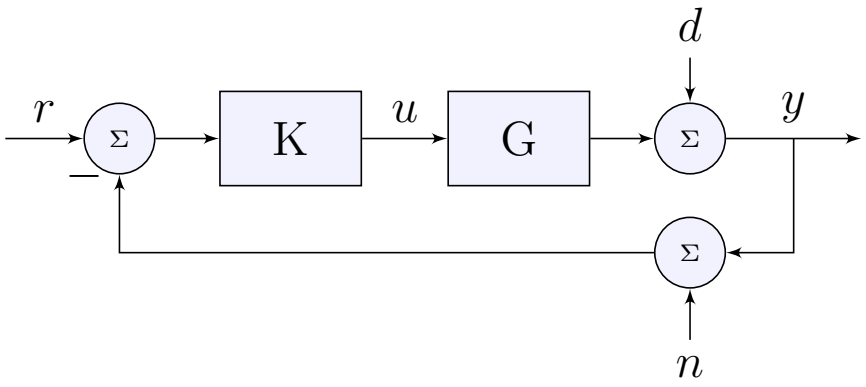
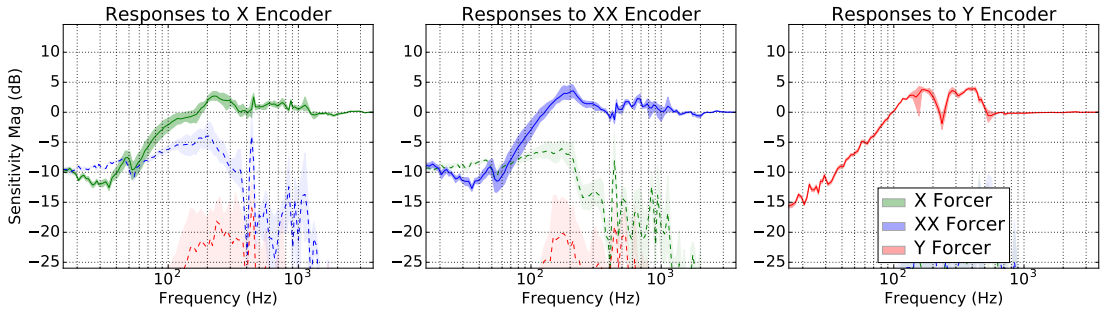


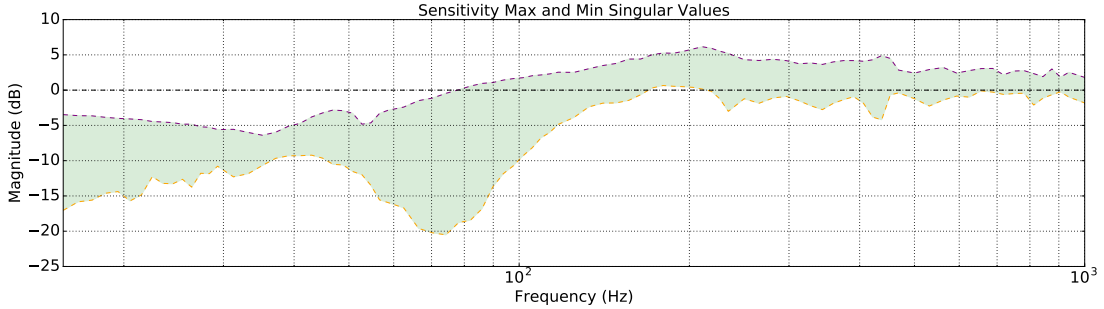
Figure 17: Simplified block diagram to illustrate sensitivity usefulness.

The values of of the sensitivity functions not only range over frequencies but also are dependent on multiple inputs. For the gantry system, there are 9 unique sensitivity functions that relate each input to each output. To examine both the sensitivity and complementary sensitivity completely would require the analysis of 18 different transfer functions. Not only is this time consuming, the analysis can lead to false stability margins. As mentioned in [9] what is considered stable from a SISO analysis can be unstable if considered in MIMO context, similar to the actual operating conditions of the system. One tool which enables a MIMO analysis is the concept of singular values.

Singular values are the diagonals of the matrix Σ_o in Equation 4.8 where U_o and V_o^T are the left and right singular vectors of the matrix $\hat{G}_o$. The singular values of $\hat{G}_o$ provide a very detailed description of how $\hat{G}_o$ acts on a vector [1]. This property makes singular values very useful in analyzing the effect of a matrix on input vectors and can be generalized for any matrix. The benefit for MIMO systems is that the singular value decomposition can identify system behavior that is not apparent using SISO methodology [26]. An example of this is the sensitivity function for a gantry. The sensitivity when viewed from a SISO perspective appears as in Figure 18a. It's clear that from the SISO perspective the sensitivity appears acceptable for the system, although it peaks near 200 Hz. However when the singular value decomposition of the sensitivity matrix is taken, Figure 18b is the result. Here the maximum



(a)



(b)

Figure 18: (a) Sensitivity when viewed from a SISO perspective (b) Sensitivity when viewed from a MIMO perspective, maximum and minimum singular values

Table 2: MIMO functions of system performance

Performance Metrics	Criteria	Frequency Range
Disturbance Rejection	$\sigma_{max}(S) \approx 0$	$0 \leq \omega \leq \omega_{bw}$
Noise Attenuation	$\sigma_{max}(T) \approx 0$	$\omega > \omega_{bw}$
Trajectory Tracking	$\sigma_{max}(T) \approx 1$ $\sigma_{min}(T) \approx 1$	$0 \leq \omega \leq \omega_r$
Low Control Energy	$\sigma_{max}((I + KG)^{-1}K) \approx 0$	$\omega(d) \quad \& \quad \omega(n)$

sensitivity magnitude is even larger and represents a configuration where the system performance could be affected by disturbances near 200 Hz. Individually the inputs and outputs of a system might satisfy the sensitivity requirements, but the inputs and outputs never occur individually during gantry operation. In addition to occurring simultaneously, the system responses do not naturally occur orthogonal to one another. This means that each system output can be larger in magnitude when multiple inputs are excited simultaneously, than any single output response to an individual input. The singular values of the response are the magnitude of these resultant responses. The maximum singular value can be considered as the worst case of the system inputs aligning to form the system response of the greatest magnitude. Conversely the minimum singular value is the condition of the system with the least amount of combination between the various input vectors to the system. Together the minimum and maximum singular values describe the maximum and minimum response possible from a given system. A caveat of the singular value is that it might not be physically possible for the maximum singular value to be realized and can create conservative estimates of system robustness and performance. Which is advantageous when viewed from the perspective that the system will perform within the expected bounds. However the controller design can only be made using the measurements and information available which are conservative by nature. In some cases a conservative controller design is the penalty of using the convenience of the singular value decomposition as an analysis tool.

Consulting Table 2, [5] and [26], the metrics listed are useful in analyzing the robustness and performance of a MIMO system. The first 3 rows are useful as described previously however the singular value is now used in the analysis. For disturbance rejection, the maximum singular value of the sensitivity must be small in the operational bandwidth of the system, or up to ω_{bw}. At frequencies larger than this, the maximum singular value of the complementary sensitivity should be small to minimize the influence of noise on the system output. And lastly for good tracking, the maximum and minimum singular of the complementary sensitivity should be approximately 1 so that the reference is followed one to one. The fourth row $\sigma_{max}((I + KG)^{-1}K) \approx 0$ is desirable to minimize at the frequencies where disturbances and sensor noise dominate. This relationship is used in minimizing the input energy required for a system, which is not considered in the design of gantry systems. The amplifiers that drive the motors on a gantry are suitably sized for the current demands of the motors as well as the process application. Therefore if input energy becomes an issue, a redesign of the amplifiers, motors, or application is normally required.

In addition to satisfying the requirements list in Table 2 it can be useful to draw analogies between MIMO and SISO system analysis due to the intuitive nature of SISO controller design. In [10] there are guaranteed gain and phase margin criteria that can be analyzed using the return difference of the MIMO system or $(I + GK)$. The minimum value of the minimum singular value of the return difference, α, is used to calculate these margins

$$GM = \frac{1}{1 \pm \alpha} \tag{4.15}$$

$$PM = \pm 2 \sin^{-1}(\frac{\alpha}{2}) \tag{4.16}$$

As with other analysis tools related to the SVD, this is a conservative estimate of the gain and phase margins but these values are guaranteed. In supplement to the sensitivity value analysis, the phase and gain margins are useful in providing a familiar relationship to stability. Figure 19 is a plot of the maximum and minimum singular values of the return difference of a gantry system, where the minimum value is -6.21 dB. The guaranteed gain margin of this system is in the range of -3.43 dB to 5.8 dB from Equation 4.15, which

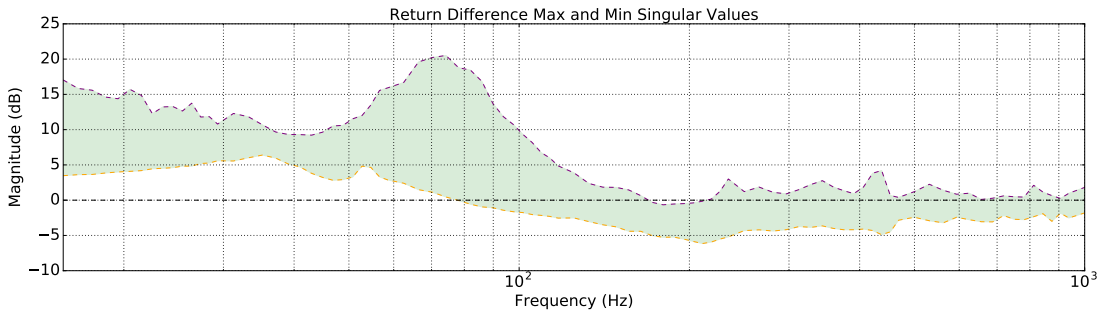


Figure 19: Return difference of a gantry system, maximum and minimum singular values.

shows that the system is stable but is not very robust near 200 Hz. The phase margin comparatively is guaranteed to be 28.3 degrees, which is an acceptable phase margin. While these are conservative estimates, it is useful to determine if the stability margins meet the rule of thumb of 6 dB of gain margin and 30 degrees of phase margin.

5.0 MODEL GUIDED MEASUREMENTS

The variability of a gantry system's plant frequency response is partially due to uncertainty in the measurement but is mainly due to position dependency [4]. It becomes necessary to measure the frequency responses in multiple locations to ensure robustness and adequate performance by defining the uncertainty bounds. When designing a controller for a gantry, it is easy to design the controls in one location of xy travel only to move to a different location and have the system go unstable. Thus system frequency responses from multiple locations in xy travel should be collected and tested with a proposed controller to determine system robustness. The level of variability would also guide a designer towards position-dependent gain scheduling if necessary. Currently it is not established how to determine the number of locations at which to measure the system frequency response. It is left up to the operator to determine this, and normally defaults to 9 measurement locations, varying between 3 different positions in x travel and 3 different positions in y travel. Through the number of gantries analyzed, this is enough data to create a robust controller, however there are systems where the frequency response does not vary significantly with travel. Conversely there are also systems where the response varies greatly with travel and requires the numerous measurement locations.

Different models will be considered to examine if a defined number of measurement locations can be determined from a discrete number of parameters. The cost (in time and complexity) of the models studied outweighs the insights that can be gained. Instead a simplified method is proposed that uses only the length of travel of the axes and the change in inertia experienced by the gantry axes as the bridge axis moves through travel.

5.1 MODEL CONSIDERATIONS

A lumped parameter model gives a designer insight into possible frequency response variations. As in [30] the model can be a 3 degree of freedom model where the dynamics are modeled as a rigid bridge with a variably located moving mass representing the bridge carriage. Using Lagrangian based modeling, the equations of motion can be derived to represent a precision gantry system. Another model, as in [11], can be used to develop similar equations of motion by including bridge stiffness and damping in a different model formulation.

$$[M]\ddot{q} + [H]\dot{q} + [C]\dot{q} + [K]q = f \tag{5.1}$$

Where M, H, C, and K are the inertia, Coriolis and centripetal, viscous damping and stiffness matrices respectively, f is the force vector and q is the coordinate vector. Both models produce roughly equivalent dynamic equations by which the system response to an input vector could be estimated. By including an uncertainty value with the various parameters included in the dynamic equation, the maximum and minimum responses from the model could be estimated. If a large enough variability in responses was expected, then more measurement locations could be added.

While this model-based method has some appealing characteristics, there are a few issues with how the model could be developed for each gantry system. Firstly it is time-consuming (i.e. expensive) to develop a unique dynamic model to describe every gantry that is produced in industry. For example, Aerotech, Inc. develops many custom gantry solutions, each of which would require an independent model to accurately model dynamics. Time is required to develop each of these models. Assuming each of these models could be created economically, the next challenge would be to determine uncertainty values for the parameters used in the dynamic equations. Testing of various components and multiple gantry systems of the same design would be necessary to determine the uncertainty values. Parametric uncertainty in gantry models is due to both component tolerances and variations in assembly procedure. Small batch sizes of components and gantries limit the confidence on uncertainty values.

Although there are challenges to develop a model, one can be developed with enough time [30]. Once the model is developed and used, the information that describes where to

add extra measurement locations reduces to inertial, stiffness, and damping changes over travel. This can be considered in terms of a frequency response plot of the plant. The overall magnitude of the plot is determined by the effective inertia of the axis, the frequency of the system resonant modes depends on the stiffness and inertia, and the magnitude of the resonant mode peaks depend on the damping at each resonance. In order to further examine the model parameters, the bridge axis should be considered first and then the analysis broadened to include the gantry axes.

With regard to the bridge axis, the mass of the payload on the bridge carriage does not change with bridge or gantry travel. The stiffness or damping of the bearings and carriage structure do not change over travel either for single axis configurations, apart from small variations in the bearing rail surface and alignment. Therefore the bridge frequency response is largely the same over the full travel. The only variations throughout bridge travel are those that couple through from the gantry axes to the bridge axis. Figure 20 shows the variation of the MIMO plant frequency response over multiple locations in gantry and bridge travel. The uncertainty of the bridge frequency response is largely due to the coupling with the gantry axes as well as the decreased signal to noise ratio at high frequencies. The coupling is apparent by examining the magnitude of the solid red line (bridge response) and comparing it to the dashed blue and green lines (gantry coupling to the bridge) on the right most plot in Figure 20. The locations where all responses nearly meet, around 80 Hz, is the frequency at which the bridge frequency response contains the most uncertainty. Here the dynamics from the gantry axes are coupling to the bridge axis in a location dependent manner.

The gantry axes however have uncertainty in the frequency response that is dependent on the location of the bridge carriage. As the bridge carriage moves closer to one of the gantry axes, X for example, the inertia experienced by the X axis increases while the inertia experienced by the XX axis decreases. The effect of this shift in inertia is two fold in that not only does the overall magnitude of the frequency response change but so too do the system natural frequencies.

Another factor that can alter the frequency response of the gantry axes is alignment throughout gantry travel. The precision gantries in this study all have very high stiffness bearings used to connect the bridge to the gantry axes. The resulting connection between the

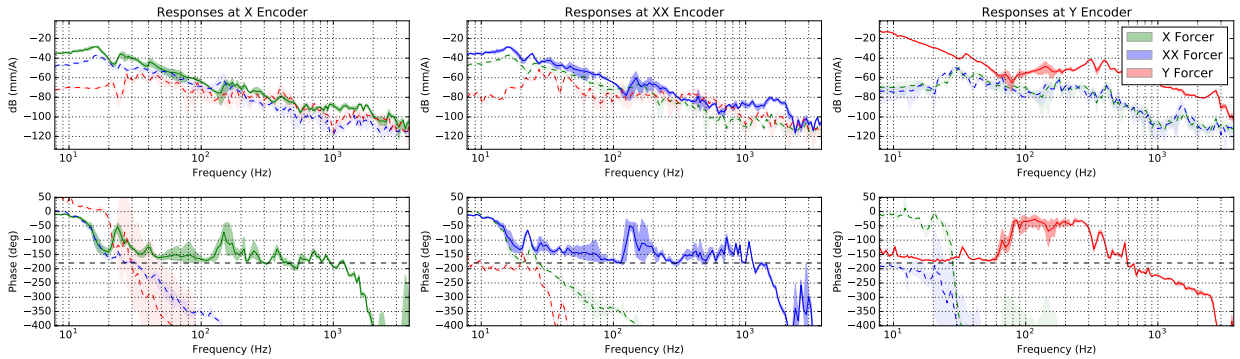


Figure 20: Gantry plant frequency response showing variation over travel

gantry axes is very stiff and therefore the alignment of the two gantry axes needs to be of good quality. Any small misalignment can result in a large lateral force on the gantry axes. For example, most bearings used on precision gantries have lateral stiffnesses of approximately 300 $\frac{\text{N}}{\mu\text{m}}$ which would result in a lateral force of 150 N if the axes are misaligned even 500 nm. Therefore when dealing with gantry systems, small changes in alignment over travel can cause significant changes in the frequency response of the gantry axes. Figure 20 shows the variation in the gantry axes plant gain and natural frequencies over bridge travel and gantry travel. The frequency responses in this figure were collected in an equally spaced 3x3 grid of xy travel. The plots are sorted by the input of the measurement where X and XX are gantry axes and Y is the bridge axis. The solid lines correspond to the diagonal terms of the plant matrix and the dashed lines are the off-diagonal terms; the off-diagonal terms are the coupling in the system. The shaded area about each response is the uncertainty due to travel and the line in the middle of the shade is the average response. Resulting from this model analysis are two parameters that are important for determining how many locations in gantry travel to measure the frequency response. The difference in inertia experienced by the gantry axes from one bridge carriage end of travel to the other and the length of travel of the gantry axes. Together these two parameters should contain the information needed to determine the number of measurement locations. The length of gantry travel is known by the operator and the gantry axes inertial difference is something that can be calculated.

5.2 BRIDGE AXIS MEASUREMENT LOCATIONS

A method for determining the number of locations to measure along the bridge axis is proposed and uses the effective inertia driven by each of gantry axis. Determining the experienced inertia of each of the two gantry axes can be done with one of two methods. CAD software makes it simple to estimate the center of mass of the gantry and calculate the division of mass to each of the individual gantry motors. This determines the machined component inertia but there is also influence due to the system cables and cable carrier. The more cables that are routed to the upper axes of the gantry and the longer the travel of the gantry axes, the higher the mass of the cables and carrier that are often not known or not generally included in sufficient detail in a typical assembly model. On some gantries the routed cables are distributed evenly between each of the gantry axes, creating approximately the same added mass on either side of the gantry. However on most gantries, the cable carrier is only attached to one side of the gantry which creates a differential in the inertia each individual gantry axis experiences. The difference in inertia between gantry axes is not static from system to system due to the variability in the size and number of cables that need to be transported in the cable carrier. It might seem that the cables and carrier are a small percentage of the gantry moving mass, but in most systems the carrier accounts for 5% of the moving mass. Figure 21 is a gantry with a cable carrier located only on one side and is a typical representation of a precision gantry.

Due to all of the variables involved in estimating the moving mass of the system, measuring the inertia of each gantry axis is more economical and accurate. In order to measure the inertia of each gantry axis, a stable controller is created for the gantry so that closed loop motion can occur. First the bridge carriage is moved closest to the gantry axis X where incremental harmonic motion is applied to each gantry axis simultaneously as in Figures 22 and 23. During the harmonic motion, the current and velocity signal are measured on both the X and XX axes so that the inertia can be fit using a least squares algorithm. These calculations are done separately for each gantry axis. The process is then repeated with the bridge carriage at the other end of bridge travel. The inertial terms from either end of bridge travel are then compared for each axis. A ratio of the inertia terms is calculated for each

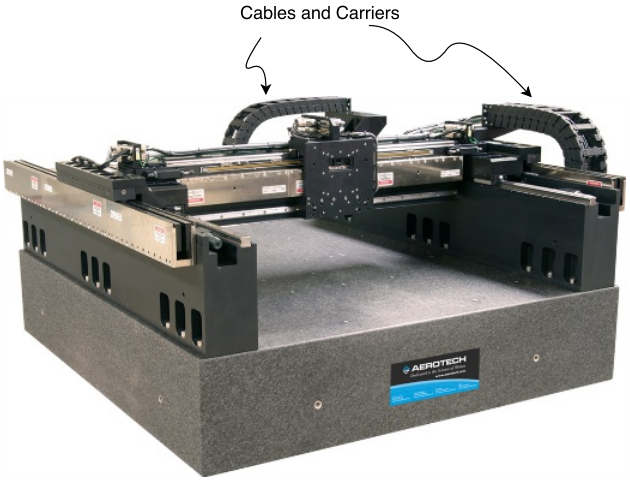


Figure 21: Example of a common precision gantry with a cable carrier on one side of the gantry. (AGS15000 gantry Aerotech, Inc.)

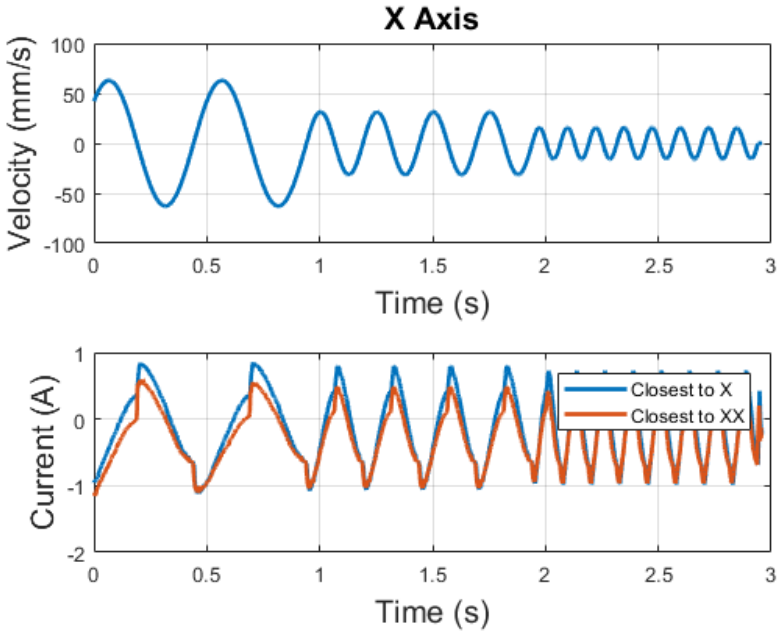


Figure 22: X axis data collected to calculate inertial difference

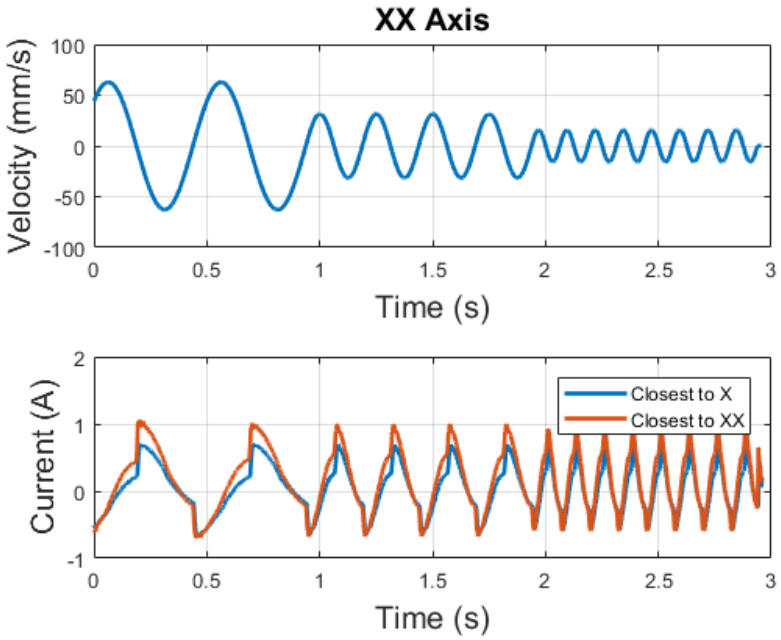


Figure 23: XX axis data collected to calculate inertial difference

axis in the form of $\frac{J_1}{J_2}$. Even before the ratio of inertias is calculated, it is apparent that each gantry axis experiences a change in inertia as the bridge carriage moves from one end of travel to the other. Figures 22 and 23 not only show a significant difference in current due to bridge carriage location but also have different peak to peak currents for each gantry axis. This indicates that the inertia of the X axis does indeed differ from the XX axis. The cable carrier for this system was attached to the X axis which explains why the X axis measured inertia is larger.

The step from measuring an inertial difference to calculating how many frequency response measurement points are needed is not readily apparent. Therefore an empirical method for determination is taken. This process was tested on 12 different gantry systems and the resulting inertial differences were measured. Initially it was presumed that the larger the travel the bridge axis, the larger the inertial difference would be on the gantry axes. This turned out not to be the case because as the bridge travel increases, the mass of the bridge carriage does not increase but the mass of the bridge itself does. Therefore the bridge car-

riage becomes a smaller percentage of the gantry axes' moving mass and therefore does not affect the frequency response as significantly. Additionally the payload on the bridge carriage has a large effect on inertial differences and customer applications require different payloads the vary in mass. Figure 24 shows the inertial ratio plotted against the bridge length and there is a slight upward trend. However larger bridge travels do not trend with the rest of the data. This is due to the large mass of the bridge required to achieve the stiffness to span such a large distance.

Using this test set, a threshold inertial ratio was established from which the number of frequency response measurement locations can be determined. Nominally 2 locations at either end of the bridge should always be measured to account for any small changes in inertial differences as well as any small mechanical variations over bridge travel. From the measured data, it was found that if the inertial ratio exceeded 1.2 on either gantry axis then the natural frequency variation over travel is large enough to warrant a third equally spaced measurement location along the bridge axis. More than three measurement locations along the bridge axis was found to provide little value except in the situation of an extremely large travel bridge, approximately 2 m, which is not a common gantry style in the scope of this research.

5.3 GANTRY AXIS MEASUREMENT LOCATIONS

The number of locations to measure the frequency response along the two gantry axes also needs to be determined. Throughout gantry axis travel there are no inertial changes, but there are still frequency response variations that occur throughout travel due to the stiffness of the gantry bearings and bridge connection. These variations are large enough that in order to guarantee a robust controller, more than one location in gantry travel must be measured. As with the bridge, 2 locations in gantry travel will be measured, one at either end of gantry travel. In the 12 unique gantries measured with this process, it was found that for gantries with travel under 600 mm, 2 locations will suffice. For gantries greater than 600 mm in travel, an extra measurement location is needed. The use of gantry travel for determining

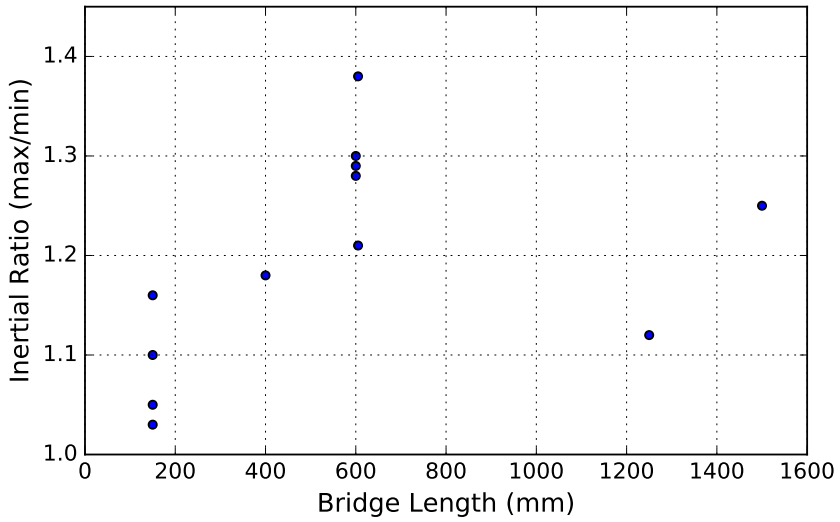


Figure 24: Plot of the various inertial ratios against the bridge length of gantries. There is a slight trend upwards as length increase, however the dependency on payload mass cannot be ignored. Very large bridge travels do not follow this trend.

the number of measurement locations needed is a valid method due to bearing consistency and alignment criteria. The linear bearings used on the gantry axes have specific preloads to ensure consistent stiffness of the bearing trucks. The alignment of the gantry axes follows a specific procedure to ensure the axes X and XX are aligned within a certain tolerance. This establishes a threshold for which the variation in bearings and alignment will not exceed and ensures this threshold is accurate at all gantry travel locations.

The methodology for determining the number of locations to measure along the bridge and gantry axes has diverged from a complex model guided method in favor of a parameter based strategy. The bridge axis defaults to two measurement locations unless the inertial ratio experienced by the gantry axes is greater than 1.2, in which case a third measurement location along the bridge axis is added. Similarly the number of measurements along the gantry axis defaults to two locations unless the travel exceeds 600 mm, upon which a third measurement location will be added. This methodology has proven successful on the small sample of gantries evaluated in this research.

6.0 MEASURING ADDITIONAL POINTS OF INTEREST

In gantry systems the axis encoders, which are located near the system inputs or axis motors, provide the position measurements. This is convenient for control and a good design choice because few dynamic effects can occur between the input and output. In other words, the input is always directly affecting the output and there is little compliance between the two. This configuration leaves out important information about the gantry workpoint or other points of interest. If the payload of the the gantry is rigid, this is probably not a concern because the assembly can be treated like a rigid body in the operational frequency range. There are however gantry payloads that are not rigid or are mounted in cantilever orientations that create resonant modes at frequencies lower than the first natural frequency of the unloaded gantry. In these situations, the behavior at the payload is not always apparent from taking a frequency response measurement or collecting time data on the axis in question. Figure 25 is a frequency response of the loop gain of one gantry axis. There is a co-located resonance around 25-28 Hz that seems to have high damping and is not destabilizing. Most technicians would observe this resonance and not consider it to be a problem. Looking at Figure 26 this is not the case because the workpoint or tool is oscillating even after the input reports no motion. This is a situation where there is a payload resonance that is not observable by the system outputs but can affect the process of the system. If this resonance becomes known, the input can be designed such that the frequency content of the system mode is avoided.

Along with workpoint vibration, base vibration can have a negative impact on system performance. There are scenarios where the base rocks in reaction to high accelerations of axes on the gantry. This rocking creates disturbances that the control loop must reject

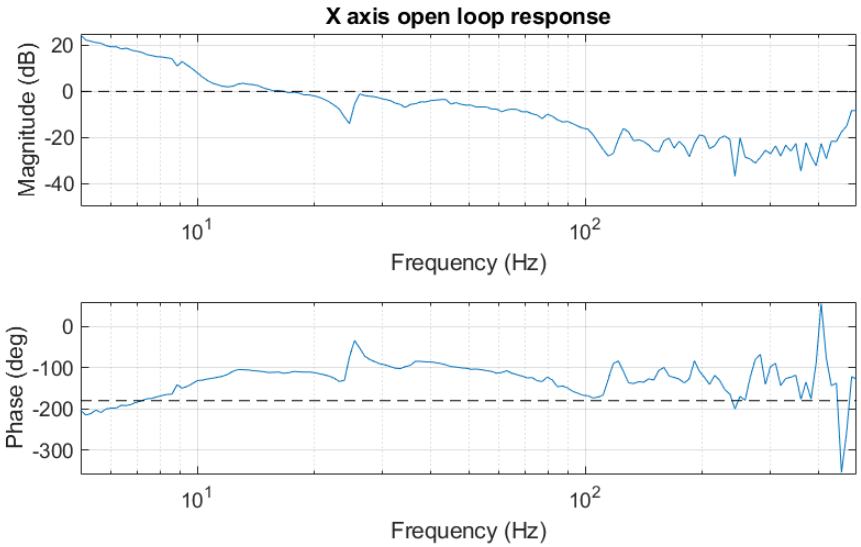


Figure 25: Loop gain of one gantry axis showing low frequency resonance around 25 Hz, a workpoint oscillation. A resonant peak on the frequency response indicates only that some element on the structure is vibrating. It does not indicate whether that vibration will be detrimental to the process in some way.

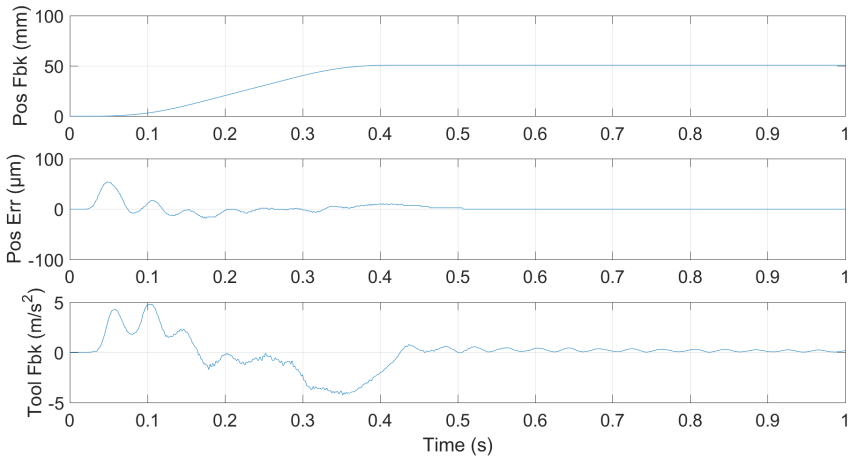


Figure 26: Time trace showing low frequency resonance around 25 Hz observable only at workpoint sensor, not at axis encoder

while following the commanded profile. This motion is normally unobservable in a frequency response because there is not enough energy input into the system to have a noticeable effect at the axis encoders.

In order to detect this undesired motion, there needs to be a sensor or method of inferring the workpoint or base behavior from known inputs [21]. Modeling issues can arise from inferring the behavior of the base or workpoint from just the 3 axis encoders. Due to the accessibility of these locations in most situations, especially at the factory, it makes more sense to directly instrument them to measure motion. A direct measurement has the advantage of not containing any modeling errors but is only as accurate as the sensors used. The frequency response of the point of interest will be calculated from these measurements and used to identify resonances that will impact system performance.

6.1 TYPE OF SENSOR

To measure the motion of points of interest (e.g. the workpoint or base) a sensor must be placed there. It is desirable for the sensor to minimally influence the behavior of the location (e.g. minimal mass loading). The time required to setup the sensor and the sensor accuracy should also be considered. These metrics facilitate the selection of a preferred sensor.

When high precision feedback from a point of interest is required, laser interferometers are often chosen for their accuracy and utility [17]. There are environmental considerations and measurement lengths that need to be considered in the setup of the instrument but interferometer measurements are useful for verification and feedback. The downside of using an interferometer is the stringent environmental constraints that are needed to minimize measurement noise. The wavelength of the interferometer light varies with temperature, pressure, humidity, and local air composition [17] which necessitates minimizing the fluctuations of these variables as well as the path length of the laser to minimize noise. Suitably sized mirrors also need to be located at the point of interest and can cause significant mass to be added. This additional mass, if located at the workpoint, will change the dynamics of the system. The cost of an interferometer is rather high as well which makes it unappealing

for companies purchasing the system, unless the interferometer is absolutely necessary. As a temporary setup for verification it is time consuming to align the interferometer mirrors to minimize the measurement error. This adds extra time to the overall measurement process because of the need to reposition devices between measurement locations, as well as needing to capture 3 axes of motion at each location. Normal interferometer setups usually capture 1 to 2 degrees of freedom at a time. Through all of the steps required to obtain measurements with an interferometer, the only real metric to consider is if the point of interest is still rigidly coupled to the measurement locations. In other words, it is the dynamic errors that are of importance.

Accelerometers are convenient for conducting dynamic measurements and have been used in multiple applications for performance improvements and position verification [16], [2], and [25]. Even the highest end accelerometers are significantly less expensive than a laser interferometer but are well suited for measuring vibration. The response of an accelerometer is an absolute measurement and can be very accurate with a low noise level. To counteract measurement noise, a low pass filter is normally applied to reduce high frequency noise, although if the signal is integrated to a position measurement the integration also acts as a low pass filter. An issue with integrating measured acceleration to position is the low frequency drift that appears in the signal, which can be minimized by a high pass filter. This results in a band pass filter over the frequency range of interest to the accelerometer measurement to ensure a position measurement with minimal drift and noise. Accelerometers normally come in a small package size (about 25 mm x 25 mm x 25 mm) and are a fraction of the mass of most components on a gantry (50 grams). This means that mass loading effects can largely be ignored and by using modal wax the accelerometer can be applied to most surfaces. This creates versatility in attaching an accelerometer directly where a measurement is desired.

6.2 SENSOR LOCATION

Two important locations on a gantry are the workpoint, where the tool or sensor is located, and the surface the workpoint interacts with, the base. The workpoint is important for the obvious reason that the application originates from this point and the base or medium the workpoint interacts with is important relative to the workpoint. That is to say the base should not be moving with respect to the workpoint, but it might be moving with respect to some external coordinate system, for example the factory floor.

Modal wax works well enough to attach accelerometers for temporary applications and provides good transfer of energy up to 7500 Hz [29], which is well above the gantry operational frequency range. Once the appropriate number of accelerometers are mounted to the gantry, care needs to be taken to make sure the cables are properly managed. This entails strain relieving the cords so that there is a short, low tension distance between the accelerometer and the first place the cable is secured. This is important to prevent any excessive cable movement from interfering with the normal motion of the stage. The data from the accelerometers should be collected with the same sample rate and under the same clock as the input and other outputs to ensure the measurements are aligned in time. With the equipment used in this research, there are analog input ports that can read an accelerometer signal in time with the other measured signals.

The workpoint should ideally have an accelerometer oriented in all three axes to measure all planes of movement during the frequency response measurements, as on the diagram in Figure 27. If the application the gantry is intended for has an insensitive direction, it is not necessary to measure that direction although it might be useful for further characterizing the system frequency response. Many times this insensitive direction is the z direction or the direction perpendicular to the plane of the gantry base. The frequency response function from the input to the accelerometers is then calculated to compare against the system response in the same axis of motion. Figure 28 is an example of a frequency response measured at the workpoint. There is a resonance present at the workpoint that is not visible at the encoder. This is a worst case scenario because most times there is some indication of the resonance in the encoder response, however even if it appears in the encoder response there is no indication

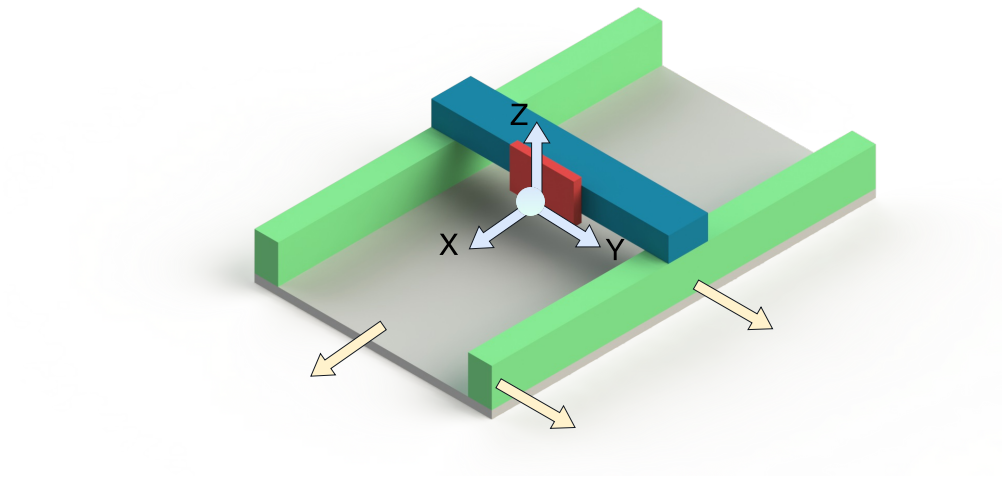


Figure 27: Diagram showing accelerometer mounting locations for gantry

if the resonance affects the workpoint. Figure 28 shows that this resonance clearly would have an effect. An example using actual data shows amplified motion in Figure 29 at 19 Hz in the x and y direction and can be used to inform the operator of amplified motion at the workpoint. This can be compared to the response at the encoders, Figure 20, where the motion is also observable. It now becomes known that resonance observed at the encoders will affect the workpoint performance. The stage in Figure 29 has a very tall, massive payload that creates this amplified, low frequency motion. It is flexible, as seen from the response, in the x and y direction and relatively stiff in the z direction. In other works, the accelerometer signal can be used to improve the performance of the stage as well as reject undesirable frequency content [2]. Here it is useful as a diagnostic tool and might be considered for other roles in future work.

A gantry base can almost always be considered as a rigid body for the purposes of determining base motion. This is a safe assumption for the gantries considered in this work. The bases these gantries are mounted to are normally 150 mm thick and at least as wide and long as gantry travel. The base then is a massive body that does have flexible modes but are of no consequence to the frequency ranges considered here. This implies that if the base is

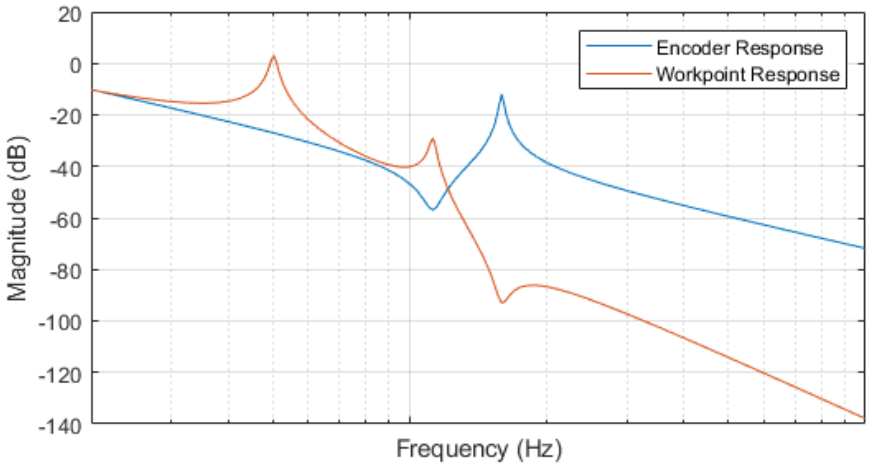


Figure 28: Frequency response magnitude showing the measurement of an accelerometer integrated to position. It is compared to the magnitude response as measured from the encoder. A resonance is present at the workpoint that is not visible at the encoder.

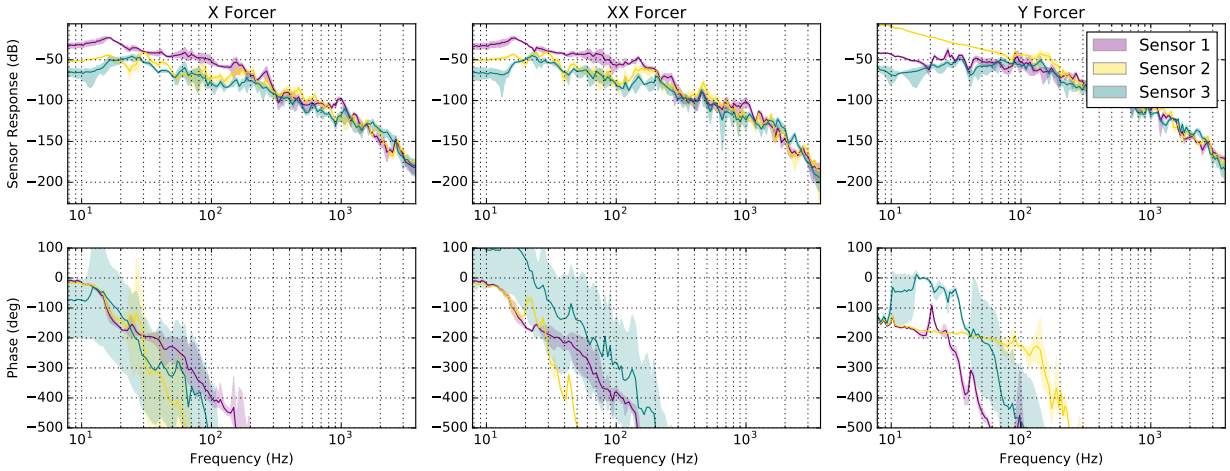


Figure 29: Frequency response showing acceleration integrated to position from the gantry workpoint. Units are in $\frac{\text{mm}}{\text{A}}$ before the dB conversion. Sensors 1, 2, and 3 measure the x, y, and z directions, respectively. The phase roll off is due to a 250 Hz low pass filter that was enabled on the accelerometer inputs as well as a 1000 Hz low pass enabled on the control loop.

moving, the entire base as a single mass. Taking this into consideration, three accelerometers are needed for a base measurement as well. There are the two directions of translation, the x and y direction, as well as the pitch, yaw, and roll of the gantry. Where pitching is rotation about the y axis of motion, yaw is rotation about the z axis, and roll is rotation about the x axis. Because the base is treated as a rigid body, if pitching or rolling is occurring translation will occur in either the x or y direction, respectively. The yaw mode can be measured by spacing two accelerometers apart along the same axis, one along the centerline of mass and the other offset some distance. This will enable translation in this axis as well as base yaw to be detected. The third accelerometer should be mounted approximately at the center of mass in a perpendicular direction to measure primarily translation. Therefore all degrees of freedom except for motion in the z direction can be measured and due to the massiveness of the base this is acceptable for most gantries. Figure 27 is an example of where to mount accelerometers on a gantry base. The two sensors on opposing sides of the same axis should be compared. If both signals are in phase, this means the base is translating without rotating. The more interesting dynamics occur if the sensors are 180 degrees out of phase which implies the base is rotating about the center of mass. This can occur at resonant frequencies if enough energy is input into the system. Figure 30 shows the base of a gantry not only translating at 50 Hz, the natural frequency of the Z stage mounted to this gantry, but also rotating at the primary yaw mode of the gantry, at 400 Hz. Only two sensors are considered here to illustrate sensing yaw motion and translation and the placement is depicted in Figure 31. The two sensors were both mounted along the same axis of motion. Sensor Y1 was mounted off to one side and sensor Y2 was mounted along the centerline of the same axis. The translation at 50 Hz is evident as both sensors show equal magnitude and aligned phase. The rotation at 400 Hz can be seen because the magnitude of the off center accelerometer peaks while the centered accelerometer shows no change in acceleration. The phase of the two accelerometers at this frequency also is indicative of the two sensors moving out of phase of one another. In this scenario the yaw mode of the gantry is well removed from the operational bandwidth of the gantry, however the 50 Hz translation could possibly pose an issue. Further analysis of this mode is necessary but the operator is now aware of its effect on certain parts of the system.

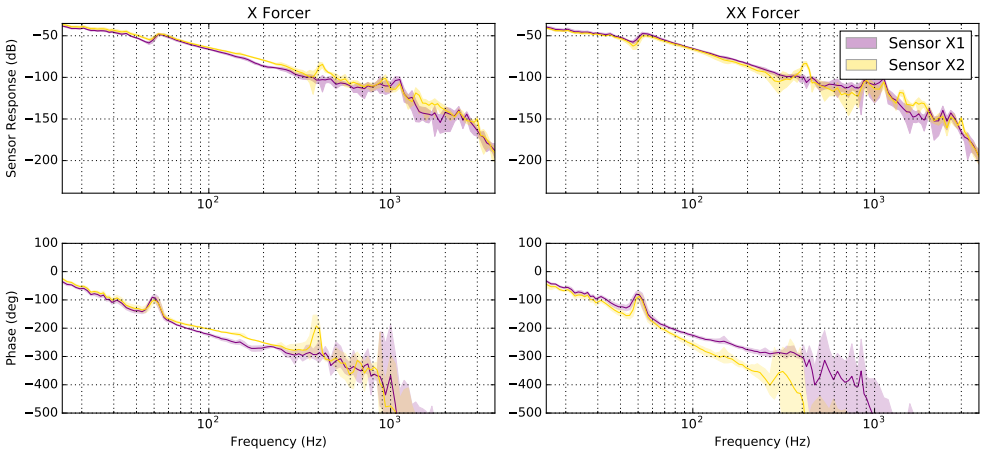


Figure 30: Plot showing accelerometer frequency response from the gantry base, using two sensors aligned along the same axis. A translational base mode at 50 Hz and system yaw mode at 400 Hz are present. Phase roll off is from low pass filters on the accelerometer signals and control loop.

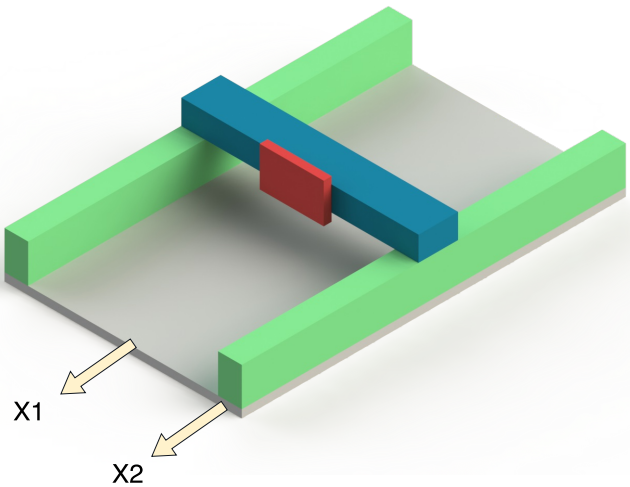


Figure 31: Diagram depicting sensor locations from Figure 30

Adding sensors to a gantry system enables the measurement of additional frequency response functions at locations that are not currently instrumented. Accelerometers are temporarily attached during the frequency response measurements to provide information about the behavior of points of interest. The base and workpoint of gantries are both locations of interest that benefit from being measured. The base measurement determines if the base is translating or rotating and potentially affecting the performance of the system. The workpoint measurement is important to verify that there is similarity between the encoder measurements and the workpoint. If there is amplified motion at the workpoint that is not apparent from the encoder response, care must be taken to not excite that frequency. It requires little effort to attach accelerometers to the base and workpoint of the gantry and therefore should be done every time a series of measurements is done. In this way there will be beneficial data to diagnose issues and support conclusions about system performance.

7.0 PROCEDURE DEVELOPMENT

The primary objective of this research has been to create a procedure that can be used by non-experts to measure the MIMO frequency response of precision gantry systems. The procedure should only require minimal input from the operator and handle most of the decision making based on predefined logic. This makes the operators job easier as well as removing variability from operator to operator. The process will always be carried out in the same way and the same data always collected. This is rather important because the current procedure is more or less dependent on the operator and what data they wish to collect.

A prototype program written in the programming language AeroBasic was used to test different excitation signals and the various parameters associated with those. AeroBasic is a programming language developed and maintained by Aerotech, Inc for interacting with and controlling precision motion systems. It allows for direct command of motion stages as well as collecting data from the motion system sensors. Specifically this was used for testing multisine variations and comparing test times as well as the signal uncertainty. From this testing and the research done in Section 2, it was decided that sequential random phase multisine signals provided the most benefit for the excitation time required. As the excitation on each axis is started, the disturbance signal is slowly scaled up to the desired amplitude and allowed to oscillate for one signal period to let transients decay. Then the multisine disturbance is input for the desired number of periods, where the measurement takes place, before it is again ramped down to prevent large steps in current that can act like an impulse to the system. Two periods minimum of each multisine sequence are repeated in the measurement phase to enable averaging between the two. In testing it was found that averaging over more than two periods provided little to no improvement in the measured frequency response. This is most likely due to the multiple complete periods of each frequency

component contained in each period of the multisine. So when an average is taken over two multisine periods, each frequency is averaged over many periods of itself. The ability to change this parameter easily is desired in the event a system can benefit from more averaging over signal periods.

Similarly, using multiple independent realizations of a multisine signal did not produce a significant improvement in the measured frequency response of gantries. Independent realizations are useful for calculating the uncertainty from nonlinearities [23], whereas many of the gantries only exhibit nonlinearities at low frequencies from friction [34] and high frequencies from amplifier characteristics [27]. Conveniently the frequency range of interest is primarily in the linear behavior range and therefore does not require techniques to detect uncertainty from nonlinearities. As with the signal periods, this is a parameter that should be easily adjustable in the event a system exhibits nonlinear behavior.

With the type of excitation signal determined and a nominal number of signal periods to excite, the next piece to establish is the measurement of multiple inputs and outputs. Again AeroBasic was used to create a program that moves the gantry to multiple locations in travel and collects time data from an excitation at each input to the gantry system. The signals from all outputs are collected along with the generated disturbance and system inputs. This totals three independent measurements at each axis location with data collection of all relevant control loop signals.

After the excitation signal prototype program was complete, the measurement locations in gantry travel need to be determined. Two locations along the gantry axis and two locations along the bridge axis are measured at minimum, for a total of four measurements. This was found to be optimal in time savings as well as still capturing uncertainty due to travel. The length of travel of the gantry axis is the determining metric for adding locations along the gantry axis. If the gantry axis travel is greater than 600 mm, an extra measurement location is added. For adding measurement locations to the bridge axis, the maximum and minimum inertia experienced by each gantry axis due to the bridge carriage location is measured. If the ratio of maximum to minimum is greater than 1.2, an extra measurement location is added along the bridge travel. It is worth noting that each extra measurement location adds

3 extra frequency responses to be measured, one for each axis. The inertia determination and gantry travel length are used to add more measurement locations before the excitation program would begin execution.

Next the number of accelerometers and their location on the gantry are determined. As mentioned in Chapter 6, the workpoint and the base are normally the most useful locations to gather extra information. Having the extra frequency response data from these outputs is useful currently in diagnosing structural resonances that can affect the end process. Sometimes excessive base motion or workpoint error is the cause of robustness or performance issues but having sensors located at either point is a convenient way to determine these influences.

The final consideration is the presentation of the different responses that have been gathered and how to effectively display the most important information. After the data collection is complete, the data is processed to calculate the transfer matrices of interest to the user. In the interest of improving robustness and performance, the sensitivity criteria from Table 2 in Chapter 4 is useful for an operator to determine robustness and performance from sensitivity plots. Due to the high number of responses collected and the MIMO nature of the system, the singular values of the sensitivity functions are displayed to show the maximum and minimum possible responses from the system. While this may result in some conservative design choices, it will guarantee stability and a certain level of performance, which may or may not be adequate for the application at hand. The amount of decoupling possible from a kinematic decoupling matrix will also be displayed to inform the operator if a decoupled control scheme would be advantageous to implement.

The steps of the research process can now be arranged into a process flow for the development of an automated procedure. Figure 32 shows a top level organization of the procedure and the order of each step. This flow chart is the structure upon which the automated procedure will be built.

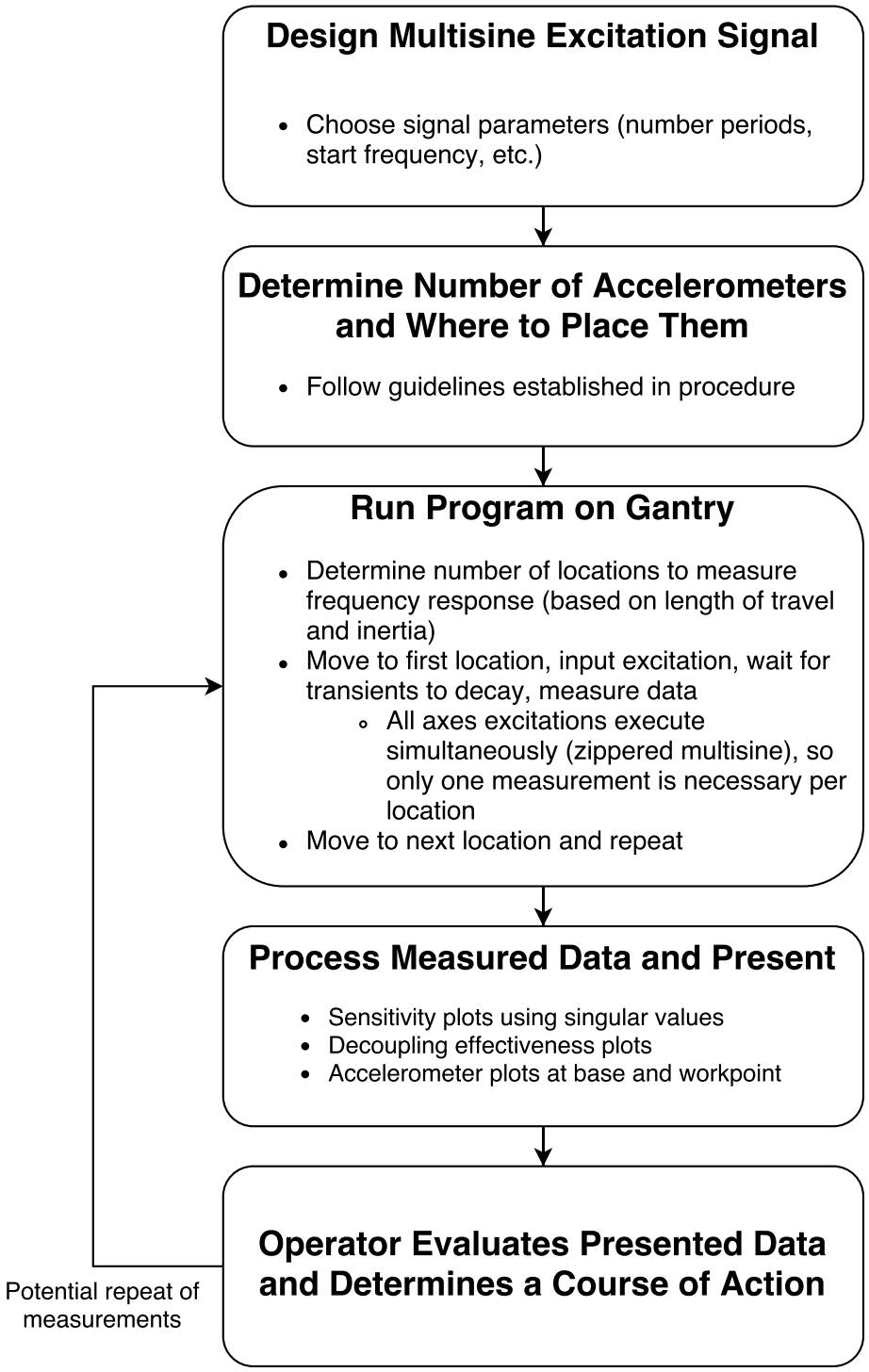


Figure 32: Flowchart of system identification process

7.1 PROGRAM APPLICATION

The procedure is now defined and in order to automate it, a program that interfaces with industrial software is necessary. This is achieved with the Python programming language. A small Python application with a graphical user interface is developed to create an interface for the operator to interact with. Figure 33 is a screenshot of the user interface for both generating the machine code as well as processing the collected data. All data fields to the left of the vertical dashed line are variables for the data collection of the frequency response. The start and stop frequency can be controlled, along with the frequency spacing of the frequency response which is controlled by the field "Spacing Weight". This field also controls how many periods of each frequency are contained in one period of a multisine sequence [14]. The maximum current percentage is based on the gantry amplifier maximum current and controls the RMS current of the multisine disturbance. And as mentioned before there are fields to control the number of independent experiments or realizations and the number of multisine periods. The names of the gantry and bridge axes are required as well as the length of travel. The data fields at the bottom of the window are used for entering information about the accelerometers attached to the gantry. It is important for the fields to be fully populated for the accelerometers so the signal can be converted from voltage to acceleration as well as properly associated to the axis each is measuring. Once all of the fields are defined, clicking the "Generate PGM" button will generate an AeroBasic file that can be run on the gantry system. The file runs, determines how many locations in travel to measure and begins performing the frequency response measurements on the gantry. When the program is finished, all data files are saved and processed by the same Python program. The Python program can automatically process the files once the data collection is complete or processing can be performed at a later time.

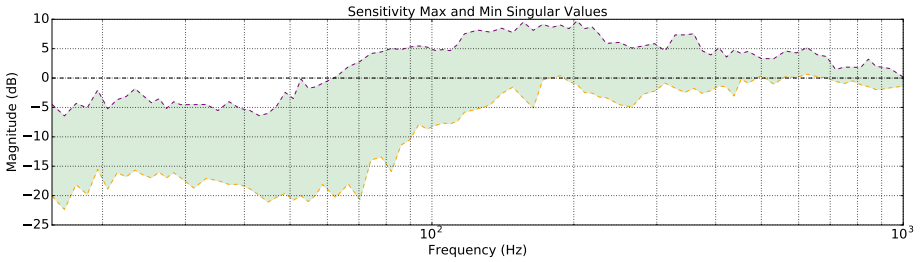
When the processing begins, the raw time data is gathered by the program and the appropriate frequency response matrices are calculated. The program currently shows the maximum and minimum singular values of the sensitivity and complementary sensitivity matrices, Figure 34a and 34b respectively. From the criteria presented in Table 2, it is guaranteed that the system will have good disturbance rejection up to 65 Hz and good

Figure 33: Screenshot of gantry program user interface

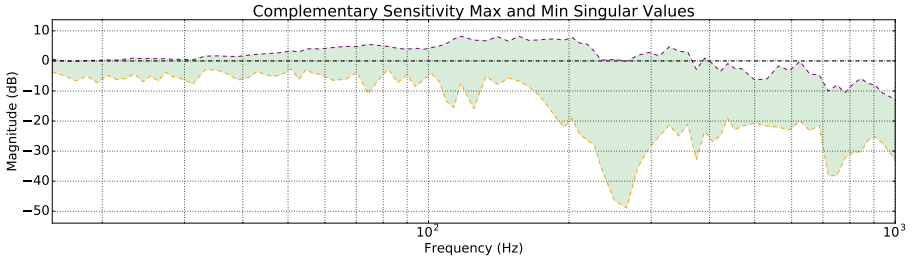
trajectory tracking up to about about 40 Hz. This does not mean the system will not function above these frequencies nor does it imply poor performance above these frequencies. What this guarantees is good performance and robustness where the responses satisfy the sensitivity and complementary sensitivity requirements.

In addition, the amount of gantry axis decoupling that can be achieved with kinematic decoupling is shown in a plot, Figure 35. By showing this plot, the operator can decide if it would useful to transform the axes in order to obtain a simplified controls configuration. The difference in magnitude between the diagonal and off diagonal terms represents the ratio of influence each input exerts on an output. The goal is for the diagonal terms to be much larger than the off diagonal terms which would yield negative dB ratios for the plot. This would indicate that the primary influence on each axis output is from only one input (e.g. a diagonalized plant transfer matrix). The threshold for good decoupling is set at a -20 dB difference or a 10 times decrease in the coupling of the two gantry axes, X and XX. Figure 35 is an example of a system where kinematic decoupling does not achieve any improvement in the frequency range of interest, 0-80 Hz, but actually increases the degree of coupling.

As a final plot to assist the operator, the frequency response of the accelerometers, either on the base or the workpoint, is shown. This plot is useful in identifying the behavior of a previously unmeasured point and relating it to the system behavior. Figure 36 shows the base and workpoint frequency responses of the same gantry considered thus far. Only two accelerometers were available for this testing when ideally a third accelerometer would be used to measure yaw of the base and the z direction of the workpoint. The base response, Figure 36a, shows in the Y direction (the direction of the gantry axes) there is a resonant mode at 20 Hz that does not appear in the plant or open loop response from the gantry axis Y and YY encoders. This implies the base is resonating in the frequency range of interest and might be affecting the process. It was discovered when testing this system with a customer supplied program, that indeed something was hindering the performance of the system. After running this program, the base was isolated as the limiting factor. With confirmation from this program, the design of the base was altered and the performance was improved to meet the customer specifications.



(a)



(b)

Figure 34: (a) Maximum and minimum singular value bounds of gantry sensitivity function for the full 3×3 system(b) Maximum and minimum singular value bounds of gantry complementary sensitivity function for the full 3×3 system

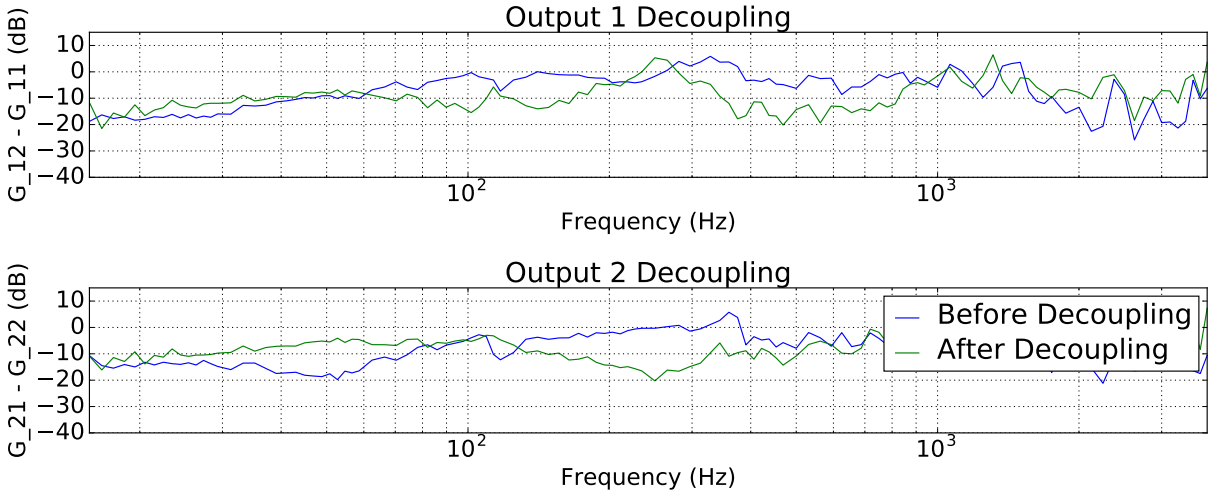
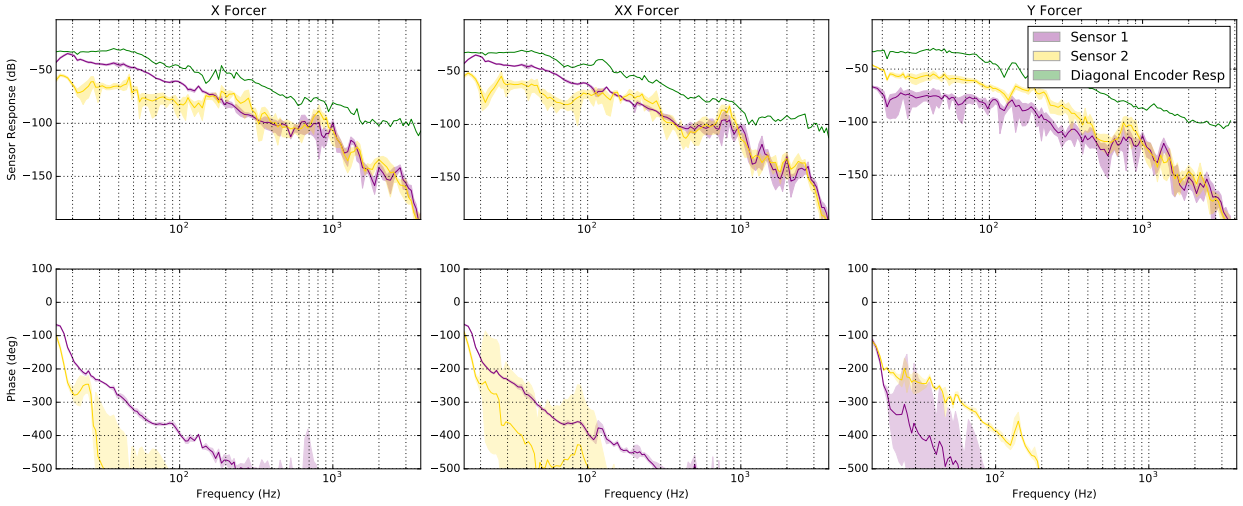


Figure 35: Amount of decoupling achieved with a kinematic transformation on the same gantry shown in Figure 34. It is not decoupled in the frequency range of interest therefore this decoupling scheme is not beneficial and should not be adopted for this particular system.

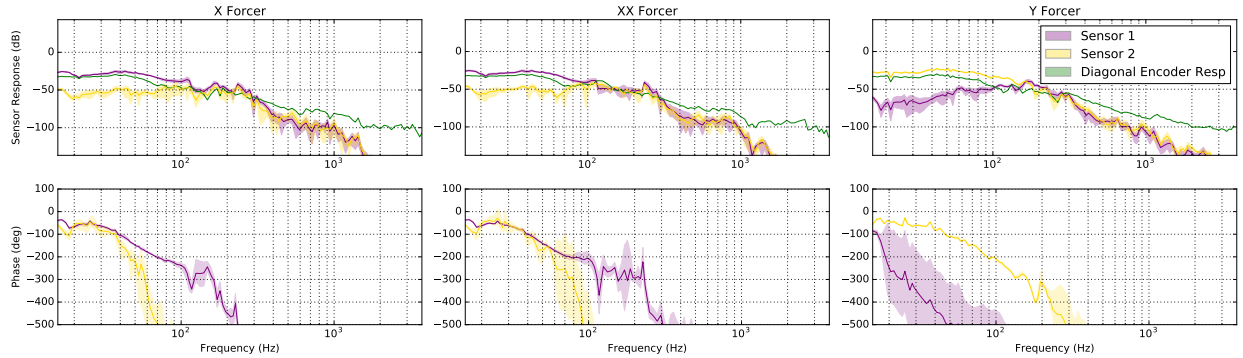
The workpoint response in Figure 36b of the same system, shows what is to be expected from the system. The yaw mode of the gantry, around 120 Hz, appears on the Y accelerometer response. Lower in frequency there is a flat region that indicates a constant transfer of force from the motor to the workpoint. At 40 Hz and lower however, the magnitude of the response at the workpoint decreases due to aforementioned coupling of gantry axes and friction. The X axis acceleration follows a similar trend with low frequency attenuation due to friction with good force transfer up to the first system resonance. This would be an example of a workpoint that is behaving as predicted by the encoder responses.

7.2 SYSTEMS TESTED

An important aspect of this research is the ability to test and develop the automated procedure on multiple different gantry systems. The systems that were available for testing



(a)



(b)

Figure 36: (a) Gantry base accelerometer response from Gantry in Figure 34, Sensor 1 measures X direction, Sensor 2 measures Y direction. Comparison with diagonal responses at encoder. (b) Gantry workpoint accelerometer response from Gantry in Figure 34. Sensor 1 measures X direction, Sensor 2 measures Y direction. Comparison with diagonal responses at encoder. Phase roll off is from filtering on accelerometers and control loop.

were actual customer systems that were tested before being shipped to the customer. This is useful in that a representative sample of systems is tested but also has the difficulty of not being able to retest a system after it has shipped. Enough gantry systems were produced during this research that availability was never a problem and any changes made to the process were tested out on subsequent systems. Throughout this research the following gantry systems were tested, Table 3. The gantry styles are the closest approximation to a standard Aerotech gantry system because many of these systems are customized to specific applications. The inertia ratio is the value calculated in Section 5.1 which is the ratio of maximum inertia to minimum inertia experienced by an individual axis of the gantry.

Table 3: Gantry testing sample

Gantry Style	Gantry Travel	Bridge Travel	Inertia Ratio
AGS15000	605 mm	605 mm	1.21
AGS10000	1250 mm	1250 mm	1.12
AGS15000	605 mm	455 mm	1.38
AGS15000	500 mm	500 mm	-
AGS15000	500 mm	600 mm	1.288
Custom	1800 mm	1500 mm	1.248
AGS1500	620 mm	150 mm	1.1
AGS1500	620 mm	150 mm	1.03
AGS1500	500 mm	400 mm	1.18
AGS1500	620 mm	150 mm	1.16
AGS15000	500 mm	600 mm	1.28
AGS15000	500 mm	600 mm	1.3
AGS1500	620 mm	150 mm	1.05

8.0 SUMMARY

The identification of precision gantry systems for industry requires the use of MIMO identification techniques, specifically frequency response measurement and analysis tools. The current identification methods take a long time, neglect the cross coupling of axes and do not characterize uncertainty. These are problems when needing to identify many unique systems to ensure robustness and performance. In an attempt to solve these issues, objectives were established to create a better method for performing MIMO system identification on gantry systems. In this research the following objectives were accomplished:

1. **Reduce the overall time required for the measurement process and quantify the uncertainty in the new measurement of the frequency response.**

   The time needed to fully identify the MIMO frequency response of a gantry has decreased by 90 % in typical test cases. The original method required 72 minutes to complete a full measurement, while the new method takes 11 minutes to accomplish the same measurements. This is not accounting for the decreased number of locations that need to be measured when using a priori knowledge of a gantry system. If this was taken into account only 6 locations would need to be measured and the time would be decreased to only 7 minutes and 20 seconds. In addition to improving the speed, the uncertainty of the measurements are now also included. The uncertainty includes the variation based on location as well as the measurement uncertainty. This creates a band of possible gantry responses and enables the uncertainty to be considered when designing a controller for a gantry.

2. **Define and incorporate an approach for using a priori knowledge of the gantry system to determine at how many locations the frequency response should be measured.**

   In order to optimize the number of locations measured, a priori knowledge of the gantry, inertia ratios and length of travel, is used to determine the number of locations to measure the frequency response. Along the direction of the gantry axis, the number of measurement locations is determined by the length of travel of the gantry axis. If the length of travel is greater than 600 mm, an extra, equally spaced measurement location is added. Along the direction of the bridge axis, the number of measurement locations is determined by the ratio of maximum and minimum inertia experienced by each gantry axis based on the bridge carriage location. If the maximum inertia measured is greater than 1.2 times the minimum inertia, an extra equally spaced measurement location is added. In both the directions of the gantry and bridge axes, the minimum number of measurements is limited to 2x2 grid or 4 locations total. This will account for any variations over travel as well as provide a minimum number of measurements with which to average over. The maximum number of locations is a 3x3 grid or 9 total measurement locations.

3. **Define the number and type of sensors necessary for gathering additional information about the system during the frequency response.**

   By placing accelerometers at the workpoint and on the base, a measurement is now available to verify dynamic behavior at previously unmeasured locations. Three axes of acceleration should be used at the workpoint, unless the end process is insensitive to one axis of motion. In this way the behavior of the workpoint can be correlated to the measurements from the system outputs. The base can also be characterized by placing three accelerometers on the XY plane. Two accelerometers should be oriented in one axis, spaced apart to be able to measure translation as well as rotation of the base and the third accelerometer should be placed along the perpendicular axis on the center of rotation to only measure translation in that axis of motion. From these three measurements the dynamic behavior of the base can be characterized.

4. **Create a procedure that can be used by non-experts for measuring the MIMO frequency response of precision gantry systems.**

   An automated procedure has been developed to measure the MIMO frequency response of gantry, process the data and present the results. A program that allows the operator to adjust frequency response measurement variables has been created that suggests default values but also allows user input. Upon running this program, a complete system identification is performed on the gantry in the form of a series of frequency responses. Once the data collection is finished, the data is processed to calculate various transfer matrices. The singular values of the sensitivity and complementary sensitivity are determined and plotted to make observations about the robustness and performance of a gantry. Along with the sensitivities a plot of the ability to decouple the two gantry axes with a kinematic transformation is shown to help in determining if decoupling will simplify the control design. Lastly the frequency response from any attached sensors is plotted to show if the instrumented point is exhibiting the predicted dynamic behavior. This program combines all the steps of nonparametric MIMO system identification into an automated procedure that enables use by non-experts.

8.1 FUTURE WORK

The work presented here is the first step in a process for automating MIMO system identification and analysis. From this point there are several ways to build upon this research:

1. Further explore simultaneous axis excitations.

2. Design controller based on system identification data collected.

3. Generalize the system identification method to function on any MIMO system.

8.1.1 Multi-axis Excitation

One of the large time commitments of this system identification process is the need to perform 3 individual excitations at each location in system travel. Simultaneous excitation was briefly explored in this work but there is more that can be done to optimize the process. In order to improve this, the effects on uncertainty and spectral characteristics should be further studied. Different techniques to design orthogonal signals should also be included.

8.1.2 Controller Design

The next logical step from this research is studying a method to design a controller using the non-parametric models obtained from this research. There are multiple control methods that use non-parametric data but it would be useful to study and choose a preferred method as well as working towards automating the controller design. This would not only allow non-experts to implement procedures to design MIMO controllers but also create a systematic method for controller design. In this manner, controller design would follow a set process and produce similar controller structures for similar systems, unlike the current methods.

8.1.3 Method Generalization

This research could also be expanded by generalizing the procedure created here to study a larger subset of MIMO systems. More inputs and outputs could be considered as well as different methods of decoupling complex systems. It would be useful to consider other MIMO transfer matrices to consider control energy or input sensitivities. Although these quantities are not as functional for the gantries considered in this research, that is not the case for all MIMO systems. In some circumstances the control energy could be the primary concern if energy is a limited quantity. Conversely the process could also be altered to consider a different subset of MIMO systems in more detail. In either case, the application of the process could be altered to consider different types of MIMO systems.

APPENDIX

NOMENCLATURE

$\frac{1}{f_k}$	signal period
$\hat{\sigma}_s^2$	sample noise variance
$\hat{\sigma}_T^2$	total sample variance
$\hat{\sigma}_{T_n}^2$	total sample noise variance
$\hat{G}$	approximate frequency response function
$\hat{G}_s$	average frequency response function
$\hat{G}_T$	total average frequency response function
ω	frequency
ϕ	phase
Σ	singular value matrix
A	amplitude
d	output disturbance
e	error
F	number of frequencies
G	plant
$G(\omega)$	frequency response function
K	controller
M	number of realizations
N	number of periods
N_f	number of frequencies in multisine signal

r reference

S sensitivity

$S_{uu}(\omega)$ input auto-spectral density

$S_{uu}^{single}(\omega)$ single sided input auto-spectral density

$S_{uy}(\omega)$ cross-spectral density

$S_{uy}^{single}(\omega)$ single sided cross-spectral density

T complementary sensitivity

T_u input transformation matrix

T_w waiting time for transient decay

T_y output transformation matrix

T_{ss} total measurement time

U left singular vector matrix

u system input

$U(\omega)$ input Fourier transform

V right singular vector matrix

v control output

w disturbance

y system output

$Y(\omega)$ output Fourier transform

BIBLIOGRAPHY

[1] Principles of optimal control, 16.31 Feedback Control Systems, March 2008. Available online. MIT OpenCourseWare <https://ocw.mit.edu/courses/aeronautics-and-astronautics/16-30-feedback-control-systems-fall-2010/lecture-notes/>.

[2] Riccardo Antonello, Kazuaki Ito, and Roberto Oboe. Acceleration measurement drift rejection in motion control systems by augmented-state kinematic kalman filter. *IEEE Transactions on Industrial Electronics*, 63(3):1953–1961, 2016.

[3] Johan Boot, MJG van de Molengraft, and PWJM Nuij. Frequency response measurement in closed loop: brushing up our knowledge. *TUe Eindhoven, April*, 2003. Available online. <http://repository.tue.nl/615430>.

[4] Dennis Bruijnen and Stan van der Meulen. Faster computation of closed loop transfers with frequency response data for multivariable loopshaping. *IFAC-PapersOnLine*, 49(13):87–92, 2016.

[5] Mohammed Dahleh, Munther A. Dahleh, and George Verghese. Lectures on dynamic systems and control, 6.241j/16.338j, March 2011. Available online. MIT OpenCourseWare.

[6] Tadeusz P Dobrowiecki and Johan Schoukens. Practical choices in the frf measurement in presence of nonlinear distortions. *IEEE Transactions on Instrumentation and Measurement*, 50(1):2–7, 2001.

[7] Tadeusz P Dobrowiecki and Johan Schoukens. Linear approximation of weakly nonlinear MIMO systems. *IEEE Transactions on Instrumentation and Measurement*, 56(3):887–894, 2007.

[8] Tadeusz P Dobrowiecki, Johan Schoukens, and Patrick Guillaume. Optimized excitation signals for MIMO frequency response function measurements. *IEEE Transactions on Instrumentation and Measurement*, 55(6):2072–2079, 2006.

[9] John Doyle and Gunter Stein. Multivariable feedback design: Concepts for a classical/modern synthesis. *IEEE Transactions on Automatic Control*, 26(1):4–16, 1981.

[10] Bernard Friedland. *Control system design: An introduction to state-space methods*. Courier Corporation, 2012.

[11] Iván García-Herreros, Xavier Kestelyn, Julien Gomand, Ralph Coleman, and Pierre-Jean Barre. Model-based decoupling control method for dual-drive gantry stages: A case study with experimental validations. *Control Engineering Practice*, 21(3):298–307, 2013.

[12] Graebe Goodwin and Salgado. Analysis of MIMO control loops, 2000. Available online, <http://csd.newcastle.edu.au/book_slds_download/>.

[13] Dan J Gordon and Kaan Erkorkmaz. Precision control of a t-type gantry using sensor/actuator averaging and active vibration damping. *Precision Engineering*, 36(2):299–314, 2012.

[14] Jesse B Hoagg, Seth L Lacy, Vit Babuska, and Dennis S Bernstein. Sequential multisine excitation signals for system identification of large space structures. In *American Control Conference, 2006*, pages 6–21. IEEE, 2006.

[15] Stephen J. Ludwick. Dynamic modeling of an H-type precision gantry including base motion effects. *ASPE 2015 Annual Meeting Proceedings*, 30:293–298, 2016.

[16] Alexander Keck, Jörg-Uwe Pott, and Oliver Sawodny. Accelerometer-based online reconstruction of vibrations from delayed measurements. In *Control Applications (CCA), 2015 IEEE Conference on*, pages 424–429. IEEE, 2015.

[17] William Land. Machine positioning uncertainty with laser interferometer feedback, June 2016. Available online. <https://www.aerotech.com/resources/white-papersapplication-notes>.

[18] W S Levine. *The Control Handbook*. IEEE Press, 1996.

[19] Michael Peter Norton and Denis G Karczub. *Fundamentals of noise and vibration analysis for engineers*. Cambridge University Press, 2003.

[20] Tom Oomen, Erik Grassens, and Ferdinand Hendriks. Inferential motion control: Identification and robust control framework for positioning an unmeasurable point of interest. *IEEE Transactions on Control Systems Technology*, 23(4):1602–1610, 2015.

[21] Tom Oomen, Robbert van Herpen, Sander Quist, Marc van de Wal, Okko Bosgra, and Maarten Steinbuch. Connecting system identification and robust control for next-generation motion control of a wafer stage. *IEEE Transactions on Control Systems Technology*, 22(1):102–118, 2014.

[22] Rik Pintelon and Johan Schoukens. *System identification: a frequency domain approach*. John Wiley & Sons, 2012.

[23] Johan Schoukens, Mark Vaes, and Rik Pintelon. Linear system identification in a nonlinear setting: Nonparametric analysis of the nonlinear distortions and their impact on the best linear approximation. *IEEE Control Systems*, 36(3):38–69, 2016.

[24] Ruud JP Schrama. Accurate identification for control: The necessity of an iterative scheme. *IEEE Transactions on Automatic Control*, 37(7):991–994, 1992.

[25] Hyunchul Shim, Michael Kochem, and Masayoshi Tomizuka. Use of accelerometer for precision motion control of linear motor driven positioning system. In *Industrial Electronics Society, 1998. IECON'98. Proceedings of the 24th Annual Conference of the IEEE*, volume 4, pages 2409–2414. IEEE, 1998.

[26] Sigurd Skogestad and Ian Postlethwaite. *Multivariable feedback control: analysis and design*, volume 2. Wiley New York, 2007.

[27] Jean-Jacques E Slotine, Weiping Li, et al. *Applied nonlinear control*. Prentice Hall Englewood Cliffs, NJ, 1st edition, 1991.

[28] Julian Stoev, Tom Oomen, and Johan Schoukens. Tensor methods for MIMO decoupling using frequency response functions. *IFAC-PapersOnLine*, 49(21):447–453, 2016.

[29] Agilent Technologies. The fundamentals of modal testing 243-3. *Agilent Technologies, Application Note*, 1986.

[30] CS Teo, KK Tan, SY Lim, S Huang, and EB Tay. Dynamic modeling and adaptive control of a h-type gantry stage. *Mechatronics*, 17(7):361–367, 2007.

[31] Rick van der Maas, Annemiek van der Maas, Johan Dries, and Bram de Jager. Efficient nonparametric identification for high-precision motion systems: A practical comparison based on a medical X-ray system. *Control Engineering Practice*, 56:75–85, 2016.

[32] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer. Tensorlab 3.0, Mar. 2016. Available online. <https://www.tensorlab.net>.

[33] Erik Wernholt and Svante Gunnarsson. Estimation of nonlinear effects in frequency domain identification of industrial robots. *IEEE Transactions on Instrumentation and Measurement*, 57(4):856–863, 2008.

[34] Jun Young Yoon and David L Trumper. Friction modeling, identification, and compensation based on friction hysteresis and Dahl resonance. *Mechatronics*, 24(6):734–741, 2014.