Dynamic Doppler Frequency Shift Errors: Measurement, Characterization, and Compensation

Chen Wang, Graduate Student Member, IEEE, and Jonathan D. Ellis

Abstract—Positioning calibration under dynamic conditions is becoming increasingly of interest for high precision fields, such as additive manufacturing and semiconductor lithography. Heterodyne interferometry is often used to calibrate a stage’s position because interferometry has a high dynamic range and direct traceability to the meter. When using heterodyne interferometry, filtering is routinely performed to process and determine the measured phase change, which is proportional to the displacement from one target location to another. The filtering in the signal processing introduces a phase delay dependent on the detection frequency, which leads to displacement errors when target velocity is non-constant as is the case in dynamic calibrations. This paper presents a phase delay compensation method by measuring instantaneous detection frequency and solving for the corresponding phase delay in a field-programmable gate array (FPGA) in real time. The FPGA hardware-in-the-loop simulation shows that this method can significantly decrease the displacement error from ±100’s nm to ±3 nm in dynamic cases and it will still keep subnanometer resolution for quasi-static calibrations.

Index Terms—Displacement measurement, dynamic response, field-programmable gate arrays (FPGAs), interferometry, phase measurement.

I. INTRODUCTION

Displacement measuring interferometry is a widely used technique for displacement metrology, position feedback sensing, and position sensor calibration for, e.g., capacitive sensors, linear variable differential transducers, and linescales. Displacement interferometry is often used in applications that require its high bandwidth, high dynamic range, and direct traceability to the meter [1]. During stage calibration, measurements are typically performed in a quasi-static configuration where either the velocity of the moving stage is low or the stage is stepped to discrete positions and held statically for the calibration to occur. The calibration data for the stage’s positioning can then be determined independently from dynamic effects, including non-rigid body motions in the stage and distortions from forces generated by the stage’s motion into the frame, which can propagate into the metrology loop.

In practice, many systems require a moving stage traveling along a pre-programmed path to perform a repetitive function. Calibration methods employing quasi-static motion profiles often do not encompass the ultimate intended use of the stage. The existing machine tool standards [2]–[4] define explicit procedures for quasi-static calibrations. Recently, there is an initiative to establish a standard for performing dynamic stage calibrations to complement the existing standards for quasi-static calibrations [5] and [7]. In this paper, we investigate how the signal processing for heterodyne interferometry systems is affected by the non-constant velocity profiles found when dynamically calibrating positioning systems.

Interferometer architectures can vary depending on the desired target geometry and intended application [7]. Traditionally, displacement interferometers have either a retroreflecting target that is insensitive to target angular rotation or a plane mirror target with a topology that utilizes the angular sensitivity or is tilt insensitive for small angles. There are two types of displacement interferometers: 1) homodyne, meaning a single optical frequency is used and 2) heterodyne, meaning two optical frequencies are used. In all cases, the metrology principle is the same: target position changes are recorded as a Doppler frequency shift in the optical frequency, which manifests as a measured phase change. The target displacement is determined by the measured phase change. As the velocity of the measurement target changes, the instantaneous frequency of the detected signal changes. The relationship between the Doppler frequency shift and the velocity of target is

\[ f_D = \frac{1}{2\pi} \frac{d\phi}{dt} = \frac{Nn}{\lambda} \frac{dx}{dt} \] (1)

where \( \phi \) is the phase difference between the reference and measurement signals, \( N \) is the interferometer fold factor (two for the interferometer used in this paper), \( n \) is the nominal index along the optical path difference, \( \lambda \) is the nominal wavelength of the laser light, and \( x \) is the displacement of the target.

For heterodyne interferometry, the two optical frequencies from the laser source provide a quasi-static heterodyne
frequency, $f_s$, typically in the 1–4-MHz range for Zeeman lasers and 20 MHz for single acousto-optic modulator (AOM) lasers [8], [9]. Alternatively, two differentially driven AOMs can be used to generate heterodyne frequencies of arbitrary values ranging from kilohertz to tens of megahertz [10].

The combination of a nominal heterodyne frequency, $f_s$, with a moving target inducing a Doppler frequency shift, $f_D$, means the frequency band of the measurement signal is $f_s \pm f_D$, as shown in Fig. 1. As the target velocity changes, the instantaneous frequency of the measurement signal changes, leading to a different local phase response as a function of detection frequency due to filtering in the measurement [11]. This imparts a distortion in the measured phase change, leading to errors when there are rapid position changes and non-constant velocity motions of the target. It should be noted that this error is not limited to interferometry as linescales would exhibit a similar behavior as a function of frequency for their phase measurements as well.

There has been limited reporting of this effect in the literature for phase measurements. Demarest [1] described dynamic measurements, but these are in the context of constant velocities, where the Doppler shift is nominally constant and this effect is not present. Holmes [11] briefly mentions this error in his thesis work, but it was only in the context of estimating the error contribution. Given that his work was limited to velocities around 1 mm/s, this error had a negligible impact and warranted no further investigation for compensation techniques. Djokic and So [12] mention the phase shift introduced by filters will influence the result in phase-angle measurement, but they state that the phase shift does not cause appreciable change the phase difference between two input signals, implying the frequencies of input signals are almost same. Vandenbussche, et al. [13] analyzed the inaccuracies introduced by the architecture of the phase measurement system, including limited attenuation and equivalent noise bandwidth of the filters, but do not investigate the impact of non-constant phase response of filters on the accuracy of the phase measurement system. In the area of global positioning system (GPS) signal monitoring, the dynamic effects of large accelerations and jerk on frequency measurements are widely discussed [14]. However, for GPS signals, the general parameter of interest is the instantaneous frequency rather than a continuous high-speed phase measurement as is the case in this paper.

In this paper, we investigate the source and magnitude of this error, model the effects of this error, and propose methods for measuring and compensating it for heterodyne interferometry applications. We present compensation methods using a Fourier transform approach, as well as a discrete derivative method. We discuss the limitation of each method and present the simulation results to validate the proposed methods.

II. HETERODYNE INTERFEROMETRY SIGNAL PROCESSING

Prior to being processed, two optical signals are detected and converted to voltage-level representations of interference amplitude, typically called the measurement signal and reference signal. The measurement signal has an instantaneous frequency of $f_s \pm f_D$, and its phase is measured and continuously tracked against the reference signal at $f_s$.

Typically, the analog signal is processed prior to converting it to a digital signal to maximize the signal and quality in the analog-to-digital converter (ADC). Fig. 2 shows the analog processing used in this paper: a transimpedance amplifier (TIA) is used to convert the current generated in the photodiode to a voltage signal, followed by a high-pass filter (HPF), high-gain inverting amplifier (IA), and low-pass filter (LPF) before being converted to a digital signal in a high-speed 14-bit ADC. For heterodyne system, this is a typical processing configuration prior to performing the digital phase processing algorithm.

The top path of the analog system in Fig. 2 is for the reference signal, while the bottom path is for measurement signals; it could be one channel from single-element photodiode, or four channels from our quadrant photodiode, based on our custom interferometer topology [15]–[17]. Fig. 3 shows the modeled phase response for this analog system. In this paper, we used two separate heterodyne frequencies depending on the anticipated target velocity with low-speed applications using a 70-kHz heterodyne frequency (shown) and high-speed applications using a 5-MHz heterodyne frequency (not shown). The whole signal processing system in Fig. 2 is designed for a 70-kHz heterodyne frequency, with a maximum Doppler frequency shift of ±20 kHz [18].

There are three general phase measurement techniques: 1) phase-locked loop [19], [20]; 2) single-bin discrete Fourier transform (SBDFT) [21], [22]; and 3) time interval analysis [11], [19]. Fig. 2 shows an example of an SBDFT technique (the digital processing part) that we used as part of this paper with our heterodyne interferometer. The SBDFT generates in-phase and quadrature signals using a voltage-controlled oscillator (VCO) nominally at a frequency of $f_s$. Each VCO-generated in-phase and quadrature signal is multiplied by the measurement signal, and using the trigonometric relationship $2 \cos A \cos B = \cos(A + B) + \cos(A - B)$, a pair of the sum and difference signals with a 90° phase difference is generated. The summed frequency (high frequency) component is removed using an LPF (Fig. 4).\(^1\)

\(^1\)The digital processing part is not an linear time-invariant system, which does not have phase response. The phase response of SBDFT algorithm here means that of digital filters.
Fig. 2. Schematic of the heterodyne interferometry signal processing system. It consists of two parts, an analog processing part and a digital processing part. The analog processing system employs several op-amp circuits, 1-kHz first-order HPFs, 200-kHz Sallen–Key LPFs, TIAs, and IAs, to process the analog signal and improve the quality of the signal. The digital processing system used an SBDFT technique and was implemented in an FPGA in this design, in which the LPFs are fourth-order, 20-kHz cutoff frequency, IIR filters.

Fig. 3. Modeled phase response of the analog processing system. It delays the phase of signal as a function of frequency, which is the superimposition of each circuit’s phase response. In the band of $70 \pm 20$ kHz, the phase difference caused by this is over 30° [6].

An arctangent operation is then used to determine the phase difference between the measurement signal and the VCO signals. Other phase demodulation methods also could be performed to extract the phase from the in-phase and quadrature component pair, such as differentiate and cross multiply [23], [24] and Kalman filters [23]. A similar process is also computed for the reference signal, ultimately measuring the phase difference between the reference signal and the VCO signals. Both phase signals are modulo-$2\pi$ and must be unwrapped first [25], prior to computing the difference. Common phase differences from the VCO signals cancel, resulting in the desired phase difference between the measurement and reference signals to be computed (detailed derivation in the next section). Once the phase is known, the displacement can be calculated from the phase difference from the start of a measurement, scaled with the interferometer fold factor, $N$, refractive index, $n$, and wavelength, $\lambda$.

III. DYNAMIC DOPPLER FREQUENCY SHIFTS

When the desired signal passes through a filter with varied frequency, a frequency-dependent phase delay will be introduced due to the non-constant phase response of the filters. This problem will occur when the target, which is measured by the interferometer, moves with a varied velocity. However, this phenomenon widely exists in phase measurement. Because filters are used to remove noise or demodulate, signals always have a non-constant phase delays even over narrow passbands, and the filters are irreplaceable in most phase measurement algorithms. This delay can be a critical phase measurement error in any application where the velocity is varied continuously, such as in semiconductor equipment and inspection, dynamic position calibration, and pick-and-place machines. Thus, compensating the phase delay due to non-constant phase response of filters is crucial to accurate phase measurements, especially when there are high velocities and accelerations in the system.

The phase measurement in heterodyne displacement interferometry systems consists of two main sources of dynamic frequency shift error: 1) filtering in the high-frequency detection regime and 2) filtering after the phase has been converted to a quasi-DC level signal. The high-frequency regime consists of any processing that would cause a frequency-dependent phase delay prior to the signal being multiplied by the signals from the VCO, which is the analog processing system in this paper (Fig. 2). Similarly, the low-frequency regime is for phase shifts induced after the multiplication with the VCO, which is the digital processing system in this paper. In the high-frequency regime, HPFs and LPFs are typically used to remove the DC component and high-frequency noise of the signals, contributing to the phase delay. If the input measurement and reference signals are

$$u_m(t) = U_m \sin(2\pi f_st + \phi_m) + U_{DC} + U_{Noise}$$

(2)
and

\[ u_r(t) = U_r \sin(2\pi f_s t + \phi_r) + U_{DC} + U_{Noise} \]  

(3)

where \( U_m \) and \( U_r \) are the amplitudes of the measurement and reference signals, \( f_s \) is the nominal heterodyne frequency of the laser source, \( \phi_m \) and \( \phi_r \) are the phases of the two signals, and \( U_{DC} \) and \( U_{Noise} \) are the DC offset and noise, which may differ in the two signals. After processing in the high-frequency regime

\[ u_m(t) = \sin \left( 2\pi f_s t + \phi_m + \phi_{h,m}(f_m) \right) \]  

(4)

and

\[ u_r(t) = \sin \left( 2\pi f_s t + \phi_r + \phi_{h,r}(f_r) \right) \]  

(5)

where the amplitudes \( U_m \) and \( U_r \) have been scaled by the amplifier circuits, \( \phi_{h,m}(\cdot) \) and \( \phi_{h,r}(\cdot) \) are the high-frequency regime phase responses (Fig. 3) for the measurement and reference signals, respectively, and \( f_m \) and \( f_r \) are the instantaneous frequency of the two signals. The underlined terms represent the Doppler-shift-induced errors that are typically ignored in most traditional treatments. Because the instantaneous frequency is a function of the heterodyne frequency and Doppler frequency shift, \( \phi_{h,m}(f_m) = \phi_{h,m}(f_s, df_m/dt) \) and \( \phi_{h,r}(f_r) = \phi_{h,r}(f_s, df_r/dt) \). Based on this, the output signals of the high-frequency regime have varied phases, when the Doppler frequency changes and there is an additional phase component that is dependent on the instantaneous frequency. In practice, the reference signal is held at a nominally constant heterodyne frequency for traditional displacement interferometry, but we include it here for the sake of completeness and because there are some interferometer architectures where reference frequency is not constant [26], [27].

After converting to digital representations in the ADCs, the signals are typically processed in an FPGA-based digital signal processing module. The specific algorithms used to measure the phase can be varied, but if there is any filtering present, then it will exhibit this error. For this paper, we used the SBDFT process shown in Fig. 2, but this compensation methodology can be adapted for other processing algorithms.

The FPGA processing is used to extract the phase difference between the reference and measurement signals in the heterodyne interferometer. The VCO or numerically controlled oscillator (NCO) blocks in the algorithm generate a constant frequency in-phase, \( u_{n,i} \), and quadrature, \( u_{n,q} \), sinusoidal signals

\[ u_{n,i}(t) = \sin(2\pi \hat{f}_s t) \]  

(6)

and

\[ u_{n,q}(t) = \cos(2\pi \hat{f}_s t) \]  

(7)

where \( \hat{f}_s \) indicates a frequency selected that is close to the nominal heterodyne frequency. The measurement and reference signals, (4) and (5), are separately multiplied by each of two NCO output signals. Based on the trigonometric product-to-sum identity, the products equal to a sum of one high-frequency \((4\pi f_s)\) component and one low-frequency or quasi-DC \((\phi)\) component. For the measurement signal, the products are

\[ u_{m,i}(t) = u_m(t) \cdot u_{n,i}(t) \]  

(8)

\[ u_{m,q}(t) = u_m(t) \cdot u_{n,q}(t) \]  

(9)

which, if the LPF is assumed to filter the high-frequency term at \((4\pi f_s)\), simplify to

\[ u_{m,i}(t) = \frac{1}{2} \sin \left( \phi_m + \phi_{h,m}(f_m) + \phi_{l,m}(f_{d,m}) \right) \]  

(10)

and

\[ u_{m,q}(t) = \frac{1}{2} \cos \left( \phi_m + \phi_{h,m}(f_m) + \phi_{l,m}(f_{d,m}) \right) \]  

(11)

where \( u_{m,i} \) and \( u_{m,q} \) are the in-phase and quadrature products of the measurement signal, respectively, \( \phi_{l,m}(\cdot) \) is the phase delay in the low-frequency regime for the measurement signal, which is the function of frequency according to Fig. 4, and \( f_{d,m} \) is the Doppler frequency shift, \( f_m - f_s \). This formulation ignores the minor frequency difference between the actual heterodyne frequency, \( f_s \), and the VCO assumed heterodyne frequency, \( \hat{f}_s \).

After that, an inverse tangent operation of \( u_{m,i}/u_{m,q} \) will extract the absolute phase of the measurement signal, \( \phi'_m \), resulting in

\[ \phi'_m = \phi_m + \phi_{h,m}(f_m) + \phi_{l,m}(f_{d,m}) \]  

(12)

A similar process can be used to determine the absolute phase of the reference signal

\[ \phi'_r = \phi_r + \phi_{h,r}(f_r) + \phi_{l,r}(f_{d,r}) \]  

(13)

Differencing the two absolute phases extracts the relative phase difference \( \Delta \phi' \) between the measurement and reference signals

\[ \Delta \phi' = \phi'_m - \phi'_r = \phi_m - \phi_r + \left( \phi_{h,m}(f_m) + \phi_{l,m}(f_{d,m}) - \phi_{h,r}(f_r) - \phi_{l,r}(f_{d,r}) \right) \]  

(14)

The ideal result should be \( \phi_m - \phi_r \); however, the extra phase terms, \( \phi_{h,m}(f_m), \phi_{l,m}(f_{d,m}), \phi_{h,r}(f_r), \) and \( \phi_{l,r}(f_{d,r}) \), contribute some error and are a function of the Doppler frequency shift. For quasi-static measurements, these added phase terms can be largely ignored. For dynamic measurements, these added phase errors can contribute significant error and must be considered.

Simulations were performed to illustrate how these extra phase terms or phase delays impact the final result when the target is moving with constant and non-constant velocity profiles. As shown in Fig. 5, filters introduce constant phase delay when target moving in constant velocity, which does not cause an error in the measured phase value because the relative displacement between any two time points is the same in theory and in practice. However, when the target is moving with a non-constant velocity, the phase delay and the relative displacement change along with the Doppler frequency.
delay. The high- and low-frequency regime phase responses
From (14), the four varied terms contribute to the phase
is to subtract the phase delay from the measurement result.
Although the filter parameters are nominally known,
analog system superimposing on that of the lock-in amplifier.
A potential problem of switching to a lock-in amplifier is that
A. Phase Response Measurement
The preliminary step in the process is to determine the phase
responses of the system, specifically from analog and digital
filtering. Although the filter parameters are nominally known,
sweeping from 1 to 100 kHz\textsuperscript{2} and the phase was recorded as a function of frequency. For this paper, we used the output signal of the function generator as the input reference signal and assumed the AOM has a negligible phase response as it has a 6-nS response rate [29].

The method to measure the phase response of digital filters is similar to that of analog system. The digital signal process model was designed in MATLAB/Simulink and the Altera DSP Builder toolbox, which was simulated on a PC. A chirp signal block generates the input signal, and the digital filters introduce phase delay to output signal. By comparing the phase difference of the input and output signals, the phase response of digital filters could be identified.

2) Phase Response Curve Parameterization: The sampling characteristics limit the number of points that the phase is known in the frequency domain. To solve the phase delay for any of the potential frequencies of interest, the expressions of the curves must be known. Thus, curve fitting can be used to parameterize the expression of the phase response curve.

The phase response has an exact expression, derived from the basic equations based on its topology. For example, the phase response of a second-order LPF that is widely used in this design is

\[
\phi(\omega) = -\tan^{-1}\left(\frac{1}{\alpha}\left(2 \frac{\omega}{\omega_0} + \sqrt{4 - \alpha^2}\right)\right) \\
-\tan^{-1}\left(\frac{1}{\alpha}\left(2 \frac{\omega}{\omega_0} - \sqrt{4 - \alpha^2}\right)\right)
\]

where \(\omega\) represents the angular frequency, \(\omega_0\) denotes the angular cutoff frequency of the filter, and \(\alpha\) is the damping ratio of the filter, which is the inverse of the quality factor \(Q\) of the filter.

The curve for an HPF has a similar form of expression. However, it is more complicated to fit the phase response curve of a complete system, like the analog system in Fig. 2, whose phase response is Fig. 3.

Since only a specific frequency band is of interest (50–90 kHz in Fig. 3) and the actual response will vary from the theoretical value, the phase response curve can simply be treated as normal curve and polynomial curve fitting can be used to parameterize the phase response using

\[
\phi(f) = \sum_{i=0}^{n} p_i f^{n-i+1}
\]

where \(\phi(f)\) is the phase response, \(f\) is the measured frequency, \(p_n\) are the coefficients for polynomial fitting, and \(n\) is the order.

Fig. 8 shows a ninth polynomial fitting of the modeled phase response of the analog processing system in band 50–90 kHz. The error between the original curve and the fitted curve is in nanoradian, which is negligible.

For the digital filters, there is an additional way to parameterize the phase response, which is using the MATLAB System Identification Toolbox. It can identify the system and create the transfer function from the measured input–output data directly. By knowing the transfer function, the exact expression of phase response is known. For the phase response of the digital filters in this design, the System Identification Toolbox can create the transfer function with a 99.41% fit to the data, and the RMS error is approximately 2.23 mrad. Based on the variability, the RMS error is approximately 0.11 nm over the \(\pm15\)-kHz Doppler frequency shift band, assuming a two-pass interferometer using a wavelength of 633 nm.

Hence, the methods to parameterize phase response curve through polynomial fitting for analog system, or through System Identification Toolbox for digital filters, are effective and precise to achieve the goal.

B. Frequency Measurement

To compensate the phase in real time, the frequency measurement should be performed in real time. Frequencies \(f_m\) and \(f_r\) are the frequencies of the input measurement and reference signals, and \(f_{d,m}\) and \(f_{d,r}\) are the Doppler frequency shifts in the measurement and reference signals. Since \(f_m\) (\(f_r\)) is the sum of split frequency \(f_s\) and Doppler frequency shift \(f_{d,m}\) (\(f_{d,r}\)) and the split frequency is nominally known, only one set of frequencies is necessary to measure \(f_m\) (\(f_r\)) or \(f_{d,m}\) (\(f_{d,r}\)).

There are two potential methods to measure the frequencies. One is the short-time Fourier transform (STFT). This would be suitable to measure \(f_m\) and \(f_r\) of the measurement and reference signals at the beginning of the digital processing in the FPGA. Another method is to directly compute the discrete derivative of the raw phase. This would measure \(f_{d,m}\) and \(f_{d,r}\) after digital processing getting uncompensated raw phase.

1) Short-Time Fourier Transform Method: The Fourier transform can display the spectrum of a signal, and it can be used to locate the primary signal frequency. However, the Fourier transform deals with the data set in the entire sample period. It shows the whole scope of the spectrum, but ignores the time dependence on a signal where the frequency may change, as is the case in this paper. In this application, the frequency is varied temporally, and the frequency at each time point or over a short time period is preferred.

The STFT employs a moving window function by multiplying it with the signal; the window is nonzero only in a short period. Thus, it can measure the spectrum in every local section, whose period is the width of the window, localizing

\(\text{Fig. 8. Curve fitting for modeled analog phase response. In the range 50–90 kHz, the error between the original phase response and the calculated phase by ninth polynomial fitting coefficients is in the range of } \pm100 \text{ mrad. The original and fitted data are offset for clarity.}\)
the signal with varied frequency throughout a larger time period. The STFT is

\[ F(f, \tau) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j2\pi ft}dt \] (17)

where the \( w(t) \) is the window function, \( \tau \) is the hop size, which determines the speed of the window moving along the time axis, and \( F(f, \tau) \) is the spectrum of the signal in every short period; the frequency which is corresponding to the maximum value of \( F(f, \tau) \) is the primary frequency of the signal in this local section. There are a variety of window functions available, commonly Hanning, Hamming, cosine, and Gaussian that can be applied to establish the bin size over which the STFT is computed [30].

However, the STFT has a fixed time-frequency resolution. This means there is a tradeoff between high time resolution and high frequency resolution. The STFT corresponding to a shorter window length produces a sharper time resolution, but a poorer frequency resolution than those corresponding to a longer window length [30]. Fig. 9 shows simulations of measuring frequencies by the STFT using windows with three different lengths. From the figure, it can be seen how the STFT can cause problems in real-time measurement. Using a 256-point window for the STFT has a lower frequency resolution, resulting in a staircase type of response. On the contrary, a 1024-point window has a better frequency resolution. Since it needs a longer interval, 1024 points, to compute the fast Fourier transform (FFT) once, it is time sensitive and introduces significant delay.

2) **Phase Derivative Method:** The frequency is the first derivative of the phase, and computing the derivative of the final measured phase can obtain the Doppler frequency shift by

\[ f_D = \frac{1}{2\pi} \frac{d\phi}{dt} \] (18)

Fig. 10 shows a simulation of measuring the frequency by this method. The theoretical signal is same as that in Fig. 9. The measured result compares nicely, the algorithm is straightforward to understand, and the computation is readily achievable in real time, although noise can be amplified. However, this method and STFT method have an inherent problem. The frequencies are computed based on signals [(4), (5), or (14)], which have suffered from the phase delay. From Fig. 5, it can be clearly seen that the filters influence not only the value of phase (displacement) but also influence its slope (frequency), so the frequency based on the inaccurately measured phase will also be inaccurate to some degrees.

3) **STFT Versus Derivative:** Since it is impractical to measure the frequencies using the unprocessed analog signal at the entry of the phasemeter, which is free from phase delay from any filters, the STFT or derivative methods are still the practical means to implement corrections. From Figs. 9 and 10, the frequency error from the STFT (several kilohertz) is much more than that from the derivative method (hundreds of hertz). This is specifically assuming the measured oscillating frequency \( f \) is at 20 Hz (Fig. 9). Different errors will occur, and potentially be greater, when the measured oscillating frequency is higher. From those simulations, the frequency errors from the STFT method and the measured Doppler frequency shift \( (f_{d,m} \text{ and } f_{d,r}) \) have the same kilohertz order of magnitude. With \( f \) increasing, the errors become comparable with the measured frequency. The reason is that the STFT in this simulation always needs 256 to 1024 point samples to operate the FFT; when the period of the measured signal approaches to time interval of those samples, the STFT will lose its ability to measure the local frequency. However, the derivative method is less related with this problem, and the error is always hundreds of hertz.

In addition to the resolution, the properties of real-time implementation and resource usage must be considered. The STFT method requires large amounts of multipliers and adders. For an \( N \)-point FFT, it needs \( 2 \times N \times \log_2 N \) times real multiplications and \( 2 \times N \times \log_2 N \) times real additions. It also spends lots of time on the calculation, which impacts the real-time measurement. On the contrary, the derivative is relatively light and simple and only needs subtractors and shifters. Hence, the derivative method is chosen to monitor the instantaneous Doppler frequency shifts in this design.

4) **Performance:** Simulations were performed to determine how the displacement and the velocity of a target’s position were affected by the frequency measurement. Assuming the target displacement is

\[ x = A \sin(2\pi ft) \] (19)

then the velocity and acceleration are derived by

\[ v = 2\pi fA \cos(2\pi ft) \] (20)

\[ a = -4\pi^2 f^2 A \sin(2\pi ft) \] (21)
even if the frequency \( f \) about 3\% of the maximum measured frequency. The measurement error is this frequency measurement mechanism has its own inherent delay to calculate the frequency. The measurement error depends on the amplitude of acceleration. This can be the effect that is mentioned in the previous section: the displacement amplitude \( x \) is constant. From the two top subfigures in this figure, both the amplitude of acceleration \( a \) is constant, the frequency \( f \), and the maximum Doppler frequency is about 15 kHz. In the second subfigure, the motion profile has a linearly increasing frequency from 0 to 200 Hz and a linearly increasing displacement amplitude from 0 to 3.8 \( \mu \)m, and the maximum Doppler frequency is about 15 kHz. In the third subfigure, the motion profile has a linearly increasing frequency from 0 to 200 Hz and a constant velocity amplitude about 4.75 mm/s, and the Doppler frequency amplitude is about 15 kHz constantly. In the fourth subfigure, the motion profile has a linearly increasing frequency from 20 to 200 Hz and a constant acceleration amplitude of 0.6 m/s\(^2\). Based on these results, the measured frequency error differs depending on the oscillation amplitude, velocity, and acceleration of the target.

According to (1), the Doppler frequency is given by

\[
 f_D = \frac{N n}{\lambda} 2\pi A f \cos(2\pi f t). \tag{22}
\]

Fig. 11 compares the frequency measurement errors from the derivative method when the amplitude of \( x, f, v, \) or \( a \) is constant. From the two top subfigures in this figure, both the displacement amplitude \( x \) and the frequency \( f \) of the target affect the error using the derivative method. From the two bottom subfigures, the error would change along with frequency \( f \), although the amplitude of velocity \( v \) (Doppler frequency shift) is constant; the amplitude of the error would be constant when the amplitude of acceleration \( a \) is constant, even if the frequency \( f \) is changed. That means the frequency measurement error depends on the amplitude of acceleration. This can be the effect that is mentioned in the previous section: this frequency measurement mechanism has its own inherent limitation when using the phase, which suffers from the phase delay to calculate the frequency. The measurement error is about 3\% of the maximum measured frequency.

C. Phase Delay Computation

The last main step to compensate the phase delay is solving for the phase delays \( \phi_m(f_m), \phi_d(f_{d,m}), \phi_m(f_r), \) and \( \phi_d(f_{d,r}) \). The phase responses \( \phi_m() \) and \( \phi_d() \) and frequencies \( f_m(f_r) \) or \( f_{d,m}(f_{d,r}) \) are already known, thus there are two ways to solve for the phase delays. One method is a calculation in real time, while another is looking up a table that stores the previously calculated phase delays.

1) Direct Calculation: This way resembles the inverse process of curve fitting, which calculates the phase delay using the instantaneous frequency and parameters of the curve.

In practice, the parameters are stored in the FPGA. For each measured frequency, it operates additions and multiplications (polynomial) or arctangents and divisions (principle expression) to calculate the phase in FPGA.

The resolution of this method is restricted by the precision of curve fitting itself, because the precision of the parameters determines the calculation. Furthermore, the operations in an FPGA are fixed point, which means it will also lose some resolution. The following section will study the precision this method could achieve. Direct calculation in the FPGA occupies a large amount of on-chip resources, especially multipliers. Furthermore, the calculation costs significant time, limiting real-time performance.

2) Lookup Table: To avoid using significant on-chip resources and computation time, an alternative method is to build a lookup table (LUT) that stores the phase delay in the FPGA. For each measured frequency, it can look up the table and find the corresponding phase delay. The phase delays are previously determined, and the number and the word length of phase delays are limited by the volume of the table.

The precision of this method mainly relies on the size of LUT. Since LUT can store a limited number of phase points, it will introduce a quantification error when the measured frequency is located between two frequency samples. In that case, the frequency will be approximated to the nearest sample. The larger the table is, the less narrow the interval is, leading to a more precise phase approximation. The following section will study how the quantification error impacts the performance of this method. This method requires memory elements but not multipliers. It is relatively less resource intensive with less calculation delay than the direct calculation.

3) Direct Calculation Versus LUT: Simulations were performed to compare the precision of the two methods. The decimal length in both methods is 8 bits for fixed-point operations and storage in the LUT. The table stores 4096 phase delay samples for the Doppler frequency shift from \(-2.4\) to \(2.4\) kHz. From Fig. 12, the error caused by direct solving is less than 4 mrad, which is about half of that of LUT. Both of the errors are less than 0.5 nm in displacement. However, as the decimal length and the volume of the LUT are increased, the error from LUT method will decrease proportionally.

Since the LUT quantification error is minimal, controllable, and fewer resources are needed, the LUT was implemented for this paper.
Fig. 12. Comparison of error introduced by LUT method and direct calculation. The errors are in the range of 8 mrad (LUT) and 4 mrad (direct calculation).

Fig. 13. Measurement errors before phase compensation and after phase compensation with or without filters, when the velocity of the target is zero and the Doppler frequency shift is zero as well. The data are offset for clarity.

V. OVERALL VERIFICATION

An overall verification was done to determine the performance of this phase compensation solution and error contribution of the three main parts. A technique called FPGA hardware-in-the-loop (HIL) was used to verify the functionality and the performance of the overall design.

Traditional software simulations for such FPGA fixed-point digital signal processing algorithms have a long simulation runtime, and application-specific availability and accuracy. Conversely, HIL simulation allows data to be processed in real time by the FPGA hardware rather than by the software. The stimulus data are generated by MATLAB and fed to the FPGA, and the FPGA computational results are read back by the MATLAB software for further analysis and display. This approach accelerates simulation time, and also ensures that the algorithm will behave as expected in the real world, so will the availability and accuracy [31].

Several simulations were performed to verify the static and dynamic characteristics of the phase compensation module and whole phasemeter model, which simulate different target motions, including static, constant velocity, and non-constant velocity motions.

A. Static Simulation

In theory, when the velocity of target is zero or constant, the Doppler frequency shift should be zero or constant. According to the curve of filter’s phase response, the error due to phase delay is zero or constant as well, which does not impact the relative displacement measurement. However, in practice, the precision of the fixed-point operation, the high-frequency residuals after filters, and other sources influence the precision of the phase (displacement) measurement. Fig. 13 shows the static error (red), which is about $\pm 15$ pm.

It is expected to compensate zero phase when the velocity of target is zero or constant. However, from Fig. 13, the error after compensation (blue) is much larger than that before compensation, which is about $\pm 300$ pm. The reason is that the phase error couples into the frequency measurement, which was discussed in the previous sections. The frequency measurement introduces high-frequency components due to the derivative, which is exhibited in the final measurement result and increases the noise.

To improve the quasi-static measurements, an LPF was applied after the frequency measurement module to remove the high-frequency components. Fig. 13 shows the effects of the added filter in the static test where the error after compensation with the filter (green) is almost same as the red one. Like other filters in this design, this filter certainly introduces extra phase delay. Because the phase for compensation is much smaller than phase measurement result, a tiny distortion of the phase for compensation has only a minor influence on the whole measurement result. This will be discussed in the following section.

B. Dynamic Simulation

The dynamic tests consist of constant velocity and non-constant velocity tests.

In the constant velocity test, the motions of the target with two different velocities, $4.75 \times 10^{-4}$ and $4.75 \times 10^{-3}$ m/s (the limits of the phasemeter for this instance), were simulated. The displacement errors before and after the phase compensation are shown in Fig. 14.

Theoretically, the error caused by the phase delay should be constant when the target is moving with a constant velocity. For low-velocity motions, the error should be near zero because the Doppler frequency shift is low. From Fig. 14(a), the measurement errors (red) do exist even for small Doppler
frequencies but are still less than ±0.5 nm. However, the measurements with phase compensation (blue) get worse, even though the Doppler velocity is low, reaching just over ±1 nm. In Fig. 14(b), the increased Doppler frequency still leads to a dynamic error (red) that gets magnified from the derivative measurement (blue). The reason is that when the velocity increases, the accuracy of the frequency measurement goes down, and the profile of the phase response at higher frequencies is not as accurate as in lower frequencies.

By adding a filter on the derivative measurement, the errors (green) can decrease by half. However, for high Doppler frequency shifts, the error cannot get back to the level before compensation [Fig. 14(b)].

In the non-constant velocity test, we simulated the motion of the target oscillating in a linearly increasing frequency from 0 to 200 Hz, and the amplitude of velocity remains at 4.75 × 10^{-3} m/s (the amplitude of Doppler frequency shift maintains at 15 kHz). The profile of the motion is shown in the third subfigure in Fig. 11. The phase errors before and after the phase compensation are shown in Fig. 15.

In this situation, the phase delay does influence the phase (displacement) measurement. From Fig. 15, the displacement error before compensation (red) is about ±120 nm. The phase compensation module decreases this error to ±3 nm (blue), a reduction of about 97%. However, the added filter harms the dynamic response of the phase compensation. After 100 Hz (at 25 ms), the error with filter (green) increases more rapidly than that without filter (blue). The reason is that the phase delay of the added filter actually starts to play a role.

From these simulations, it can be found this solution has subnanometer static performance with a filter after frequency measurement module. In dynamic simulations, the phase compensation method effectively reduces the phase error from filtering effects. The filter on the derivative measurement slightly impacts the accuracy of the compensation when the motion of target has a high frequency (higher than 100 Hz). It will be a tradeoff between the static and dynamic performance.

VI. Conclusion

The filter-introduced phase delay in interferometry displacement measurements was not extensively investigated in the literature. To the best of the authors’ knowledge, there was no clear and direct method published to eliminate or compensate the displacement error due to the phase delay. However, this effect becomes significant but often overlooked when there are rapid position changes and non-constant velocity motions of target, whose value could be up to hundreds of nanometers depending on the non-constant phase response of filters and the Doppler frequency shift.

In this paper, we have investigated the source and magnitude of this error, modeled the effects of this error, proposed methods for measuring and compensating it for heterodyne interferometry applications, and implemented the phase compensation module in an FPGA. The synthesizable phase compensation module consists of a Doppler frequency shift measurement part using an instantaneous derivative method and a phase delay computation part using LUT. The FPGA HIL simulation results show this method can decrease the error from ±100’s nm to ±3 nm in dynamic cases and still keep subnanometer resolution in quasi-static cases. This increases the dynamic accuracy of the phasemeter, thus increases the accuracy of the displacement interferometry measurement. However, this method is not limited to only heterodyne interferometry but rather it can be applied to other frequency shifting measurement techniques like homodyne interferometry and linescale signal processing.

At present, this paper demonstrates the FPGA HIL simulations using a designed heterodyne frequency at 70 kHz, with a maximum Doppler frequency shift of 20 kHz (maximum velocity of 6.33 mm/s). Future work will verify the design through the experiments with interferometry in practice and demonstrate the phase compensation for higher speed application using a 5-MHz heterodyne frequency with a maximum Doppler frequency shift of 2 MHz (633 mm/s).

ACKNOWLEDGMENT

The authors would like to thank R. Fesperman, J. A. Tarbutton, T. L. Schmitz, S. R. Gillmer, X. Yu, and C. Lu for their fruitful discussions regarding this paper.

REFERENCES


2004 IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT, VOL. 64, NO. 7, JULY 2015


Chen Wang (GSM’13) received the B.E. degree in measurement and control technique and instruments from the Harbin Institute of Technology, Harbin, China, in 2011, and the M.S. degree in electrical engineering from the University of Rochester, Rochester, NY, USA, in 2013, where he is currently pursuing the Ph.D. degree in electrical engineering. He has been a Research Assistant with the Precision Instrumentation Group, University of Rochester, since 2011. His current research interests include high speed, real time, precise electrical/electronic measurement, and control system design, mainly including signal processing, dynamic control algorithm analysis, development and implementation on field-programmable gate array, and embedded system.

Jonathan D. Ellis received the B.S. and M.S. degrees in mechanical engineering from the University of North Carolina at Charlotte, Charlotte, NC, USA, and the Ph.D. degree in mechanical engineering from the Delft University of Technology, Delft, The Netherlands. He is currently an Assistant Professor with the University of Rochester, Rochester, NY, USA, with a joint appointment between the Department of Mechanical Engineering and the Institute of Optics. His current research interests include dimensional metrology, precision measurement, and instrumentation for biological application.