Feedforward control of a closed-loop piezoelectric translation stage for atomic force microscope

Yang Li and John Bechhoefer[a)]
Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

(Received 4 October 2006; accepted 7 November 2006; published online 11 January 2007)

Simple feedforward ideas are shown to lead to a nearly tenfold increase in the effective bandwidth of a closed-loop piezoelectric positioning stage used in scanning probe microscopy. If the desired control signal is known in advance, the feedforward filter can be acausal: the information about the future can be used to make the output of the stage have almost no phase lag with respect to the input. This keeps in register the images assembled from right and left scans. We discuss the design constraints imposed by the need for the feedforward filter to work robustly under a variety of circumstances. Because the feedforward needs only to modify the input signal, it can be added to any piezoelectric stage, whether closed or open loop. © *2007 American Institute of Physics.*
[DOI: 10.1063/1.2403839]

I. INTRODUCTION

The problem of positioning objects rapidly and accurately is one that experimentalists encounter frequently and is required by many nanoscience technologies. For example, scanning probe microscopes (SPM) build up an image by rastering the tip (or sample) through a two-dimensional scan, while tracking sample topography in the third dimension, as well.[1,2] In nanolithography, an active tip must be controlled to follow some designed path.[3] In single-molecule biophysics, probes often must be positioned to nanometer accuracy.[4]

For displacements at the nanoscale, the typical technological solution uses piezoelectric actuators. Because piezoelectric materials have a nonlinear, hysteretic response, they are often used in closed-loop control systems, where a sensor records the actual displacement and feeds back an error signal to an analog or digital controller that attempts to track a given setpoint. The resulting closed-loop system has a nearly linear response at low frequencies, meaning that slow control signals can be accurately tracked. In addition, the kind of stage to be discussed here guides the motion using a flexure system that greatly reduces unwanted motions. One limitation of such stages, however, is that at higher frequencies, the feedback is deliberately reduced and the beneficial effects of feedback linearization are lost. The feedback control systems is limited in bandwidth for two reasons:[5] First, the inevitable mechanical resonances of the displacement stage will add phase shifts, which tend to destabilize feedback loops. While it is in principle possible to compensate for such resonances, the more straightforward solution is simply to limit the bandwidth of the controller to frequencies well below that of the first significant resonance. Second, as the feedback bandwidth increases, the noise in the displacement sensor, which feeds through to the physical displacement, will be more and more significant. These two limitations mean that practical closed-loop translation stages tend to have closed-loop bandwidths that are 1/10 to 1/20 the frequency of the first resonance. Because the usual tracking waveforms have harmonics—SPMs commonly use triangle waves so that the scans are at constant velocity—most currently available closed-loop stages can track movements at frequencies that are at most 1/20 to 1/40 of the first resonant frequency. The lowest resonant frequency depends on the purpose of the stage (load, displacement range, etc.) but typically ranges from 100 Hz to 10 kHz. Such low frequencies are one obstacle in developing higher-speed SPMs.

The purpose of this article is to show that a simple implementation of a technique known as feedforward control can significantly increase the usable bandwidth of piezoelectric displacement stages. While feedforward has been applied by the groups of Devasia and of Stemmer to piezoelectric tube scanners not under closed-loop control,[6–11] the design presented here is particularly simple and robust and does not rely on advanced concepts such as $\mathcal{H}_\infty$-metric design.[9] Because the stage considered here is already under closed-loop control, we can obtain good performance using simple methods. Another design that combines feedforward with feedback presented in this journal[12] focused on improving accuracy at low bandwidth, while we focus here on speeding up the stage response.

The basic idea of all the feedforward strategies is to use the known closed-loop response of the stage to design a prefilter for the desired input stage. Loosely, to the extent that the distortion of a desired control signal is known in advance, one can design a waveform with compensating distortions so that the stage's response "undoes" the distortions, leaving an approximation to the desired waveform. Because all that is done is a prefiltering of the desired input, one may add such a feedforward control to any existing displacement stage, whether open or closed loop. Other advantages of the technique will be given below.

Feedforward is a standard technique in control theory[5,13] where one uses the known dynamical response of the system to design an input that will lead to a desired output. Classic

a)Electronic mail: johnb@sfu.ca

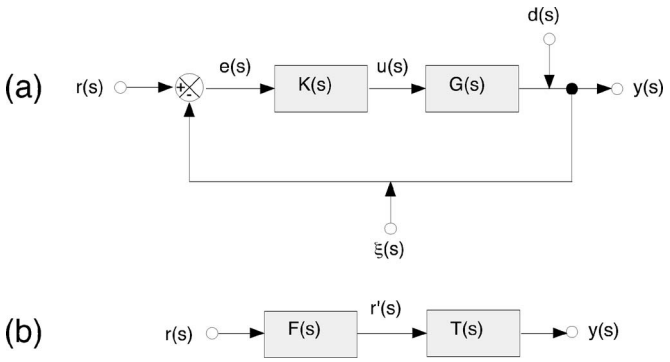

FIG. 1. (a) Block diagram of a typical feedback control system showing the control input r, direct-acting disturbances d, and sensor noise ξ. The physical system is represented by G and the control algorithm by K. (b) Feedforward prefilter F modifies the input signal from r to r', which is then substituted for the control signal r in the closed-loop system in (a). The closed-loop system of (a) is represented in (b) by the block T.

applications focus on avoiding the excitation of softly damped vibration modes. They include the related problems of using a crane to displace a heavy load without making it swing[14] and that of moving an open container of fluid without making it slosh.[15] Feedback, by contrast, uses trajectory information about the measured deviations of the output from the desired output. While the potential advantages of feedforward are well-understood by engineers,[16] they have been less appreciated in the physics community. In addition to the work by the groups of Devasia and Stemmer cited above, a commercial manufacturer has recently begun offering feedforward as an option for its own piezoelectric stages.[17] Although the commercial feedforward control has some features in common with the one proposed here, the details are somewhat different.[18] More generally, we shall see that feedforward—like all engineering design—involves a number of tradeoffs (e.g., accuracy versus speed, optimization for a particular scan waveform, etc.). The advantage of designing one's own system is that one can weight the necessary tradeoffs in a way that best reflects one's own goals.

II. FEEDFORWARD CONTROL

As closed-loop positioning stages have a nearly linear response, subject to range and slew-rate limitations, a linear analysis will suffice, although extensions to nonlinear systems are also possible. Figure 1(a) illustrates the signal flow in a single-input, single-output (SISO) control system. In our case, the input will be the desired position of the stage, $r(t)$, and the output $y(t)$ will be the stage's actual position. Using Laplace transforms and working in the frequency domain, we denote the physical system's Laplace transform as $G(s)$, and the controller's as $K(s)$. The signal $d(s)$ represents mechanical and electrical disturbances (assumed, for simplicity to affect the output directly), while $\xi(s)$ represents sensor noise. In addition, the actuator signal $u(s)$ is the output from the controller that is sent to the input of the physical system. For the translation stages discussed here, the physical system $G(s)$ includes a high-voltage amplifier, piezoelectric ceramic actuators, a mechanical flexure guide that constrains motion to one dimension, and any load placed on the stage.

Solving the signal flow in Fig. 1 gives

$$y(s) = \frac{K(s)G(s)}{1+K(s)G(s)}[r(s)-\xi(s)] + \frac{1}{1+K(s)G(s)}d(s)$$
$$\equiv T(s)[r(s)-\xi(s)] + [1-T(s)]d(s). \quad (1)$$

In Eq. (1), $T \equiv KG/(1+KG)$ is the closed-loop response function, also known as the "complementary sensitivity function."[5] In general, one chooses the feedback gain K to be large ($\gg 1/G$), implying that $T \approx 1$. As Eq. (1) then shows, $T \to 1$ suppresses the effect of disturbances d, while the output y will tend to track the input r. Unfortunately, two problems limit this solution. First, if the denominator $1+K(s)G(s)$ ever vanishes, there will be an infinite response to a finite input, i.e., an instability. This will occur when $|KG|$ has unit gain and a 180° phase shift. Since physical systems will develop large phase shifts at high frequencies, the gain of K must be reduced at those frequencies so that $|KG|<1$. This implies a feedback bandwidth, ω_b, defined to be the lowest frequency where $|T|=1/\sqrt{2}$. In addition, even if stability is not an issue, Eq. (1) shows when $T=1$, $y=r-\xi$. In other words, the noise ξ feeds through to the actual output. Because this noise is injected by the feedback loop itself, it represents a deterioration of performance. The usual solution to this dilemma is again to limit the feedback bandwidth, with the idea that disturbances are typically low frequency while sensor noise extends to very high frequencies. More sophisticated approaches (Kalman and Wiener filtering) calculate the optimal feedback bandwidth in the presence of sensor (and actuator) noise with known statistical properties.[19]

Both of the limitations discussed above imply that the closed-loop response function $T(s)$ will resemble that of a low-pass filter with bandwidth ω_b (with perhaps some complicated dynamics in its roll-off). As mentioned in the Introduction, in SPM applications the desired control signal is often a triangle waveform (representing constant-velocity scans to the right and left). The corner of the triangle implies high-frequency components, resulting in significant distortions that occur at control frequencies well below ω_b.

The feedforward approach to increasing the effective bandwidth of the closed-loop system modifies the loop-structure of Fig. 1(a) by adding a prefilter. Figure 1(b) shows this arrangement: the prefilter $F(s)$ changes $r(s)$ to $r'(s)$, which is then sent to the closed-loop system T, with the goal that $T(s)r'(s)$ is close to the desired signal $r(s)$. One advantage of this configuration, known as a "two degree of freedom" controller (see Ref. 13, Chap. 10), is that one can choose the controller $K(s)$ to limit disturbances and noise feedthrough while at the same time choosing $F(s)$ to track $r(s)$. Ordinary "one degree of freedom" controllers that lack F must set the frequency-dependent gains in K in a way that compromises between sensor noise feedthrough and tracking. Note that, while our discussion assumes closed-loop control of the physical system G, it clearly applies, too, if there is no closed-loop control of the physical system, in which case $T=G$.

There are many strategies for choosing the feedforward controller $F(s)$. The group of Seering and collaborators (used

in the commercial design of Ref. 16) advocates a particular strategy where $F(s)$ is a finite-impulse-response (FIR) filter that can be expressed in discrete-time form as

$$r'_n = a_0 r_n + a_1 r_{n-1} + a_2 r_{n-2} + \cdots . \quad (2)$$

In Eq. (9), r'_n is the output at time n (in units of a sampling time T_s), r_n is the input, and the set of a's are filter coefficients whose value must be determined. One also commonly imposes constraints on the coefficients so that the range of the output signal $r'(t)$ never exceeds that of the input $r(t)$ (considered over the whole waveform). In our case, we shall use an "infinite impulse response" (IIR) filter and somewhat different constraints, which are more suited for typical SPM applications. (See Sec. III C, below.)

We begin by noting that the obvious naive strategy is to choose F to be the inverse of T: $F(s)=T^{-1}(s)$. The combined response function is then unity and it would seem that y will track r perfectly. There are two difficulties (see Ref. 13, Chap. 15). First, the strategy is not robust: If the physical system G were known perfectly and were unchanging, its inverse would be well-defined. However, this is rarely the case. In the application discussed below, for example, the various models of the stage dynamics are accurate at low frequencies but become inaccurate at higher frequencies. The second difficulty is that the feedforward signal implied by $F(s)$ may be physically unachievable, as physical inputs are limited in both amplitude and frequency (for example, by the sampling time of the discrete output). These limitations make it impossible to reproduce the dynamics of T beyond some cutoff frequency.

III. APPLICATION TO A PIEZOELECTRIC STAGE

In this section, we show how the general ideas about feedforward discussed above apply to the specific case of a piezoelectric stage.

A. Characterization of stage dynamics

We used a commercially available closed-loop, two-axis translation stage[20] that is part of a home-built atomic force microscope (AFM). The stage is controlled by an analog proportional-integral feedback loop. Figure 2 shows the closed-loop transfer function $T(s)$ for the X-stage,[21] measured directly using a lock-in amplifier.[22] The light and heavy solid curves show, respectively, the response of the unloaded stage and that of a 113 g. load (our sample support). Notice that the first resonance is lowered (from 525 to 442 Hz), but the low-pass filter, the second resonant peak, and the zero are all essentially unchanged. We can identify the features of the transfer function physically, as follows: The two-pole low-pass filter at 42 Hz is imposed by the analog feedback electronics of the stage's controller. The 525/442-Hz peak is the mechanical resonance of the moving stage itself. The alternation of peaks and zeros (660, 950 Hz, ...) are typical of mechanical systems where energy is input and measurements are made locally.[23] The resonances represent different mechanical modes of the system, while the antiresonance frequencies depend on the placement of actuator and sensor. (Essentially, zero response is measured

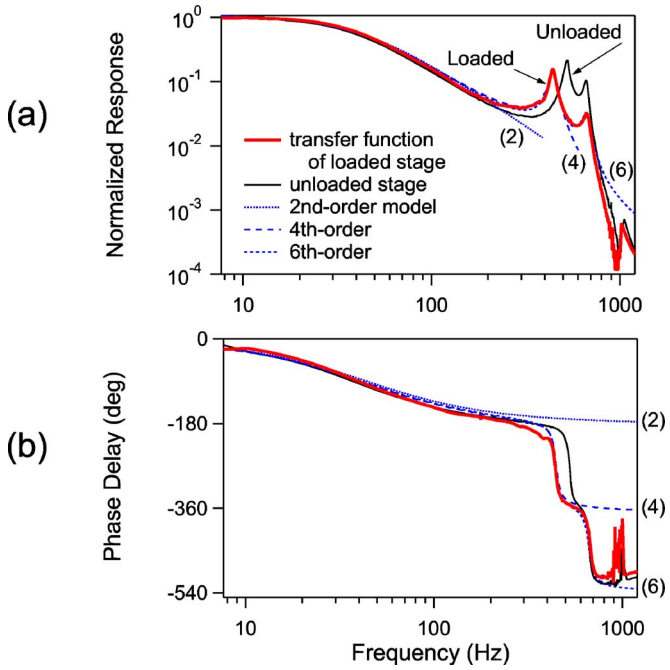

FIG. 2. Measured frequency response (solid lines) of loaded and unloaded translation stage, overlaid with fits (dashed lines) to the loaded response by three different models (two, four, and six poles). (a) Magnitude response of the Bode plot. (b) Phase response.

when the sensor happens to be at a node of the excited system.) In any case, because of the low-pass filter imposed by the stage's controller, the amplitudes of these dynamics will be small ($<10^{-3}$ of the driving amplitude).

We also show in Fig. 2 a series of fits to the loaded transfer function that represents models that capture the behavior to higher frequencies. The simplest is a fit to a two-pole, low-pass filter of the form

$$T_2(s) = \frac{1}{\left(1 + \frac{s}{\omega_0}\right)^2}, \quad (3)$$

with $\omega_0/(2\pi)=42$ Hz, corresponding to a bandwidth of $42\sqrt{\sqrt{2}-1} \approx 27$ Hz.[24] We refer to this as the "two-pole" model. In order to explore the benefits of including higher-frequency dynamics, we also fit a "four-pole" model by multiplying Eq. (3) by a second-order denominator, corresponding to the main mechanical resonance at $\omega_1/(2\pi)=442$ Hz. Explicitly,

$$T_4(s) = T_2(s)\left(\frac{1}{1 + 2\zeta_1\left(\frac{s}{\omega_1}\right) + \left(\frac{s}{\omega_1}\right)^2}\right). \quad (4)$$

Similarly, the six-pole model $T_6(s)$ multiplies an additional second-order denominator of frequency ω_2 and damping factor ζ_2. Associated with the two-, four-, and six-pole models are asymptotic phase shifts of 180°, 360°, and 540°, respectively.

B. Continuous-time feedforward filter

We explored a series of different feedforward filters $F_2(s)$, $F_4(s)$, and $F_6(s)$, based on the different models of the

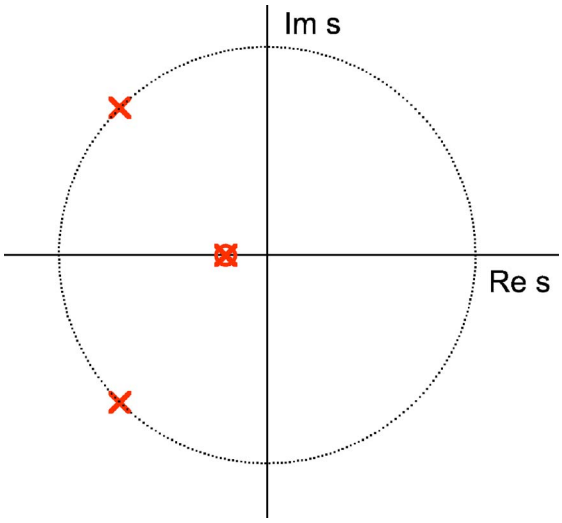

FIG. 3. Pole-zero plot of the system $F_2(s)T_2(s)$, showing how the prefilter "cancels out" poles of the system with zeros and then adds new poles at higher frequencies. The crosses indicate poles; the circles, zeros.

system dynamics discussed in the previous section. They have the form $F_j(s)=T_j^{-1}(s)L_j(s)$ for $j=\{2,4,6\}$. Here, L_2, L_4, and $L_6(s)$ are low-pass filters of orders 2, 4, 6 that set the bandwidth of the dynamics of the combined prefilter and physical system. We set it to about 250 Hz. This is greater than 50% of the first resonant frequency and is 9.3 times greater than the bandwidth of the normal stage (27 Hz). In principle, $L(s)$ could have the same form as $T_2(s)$, but a two-pole Butterworth filter (which has the flattest amplitude response) performed slightly better.[25] Thus,

$$F_2(s)=\frac{\left(1+\frac{s}{\omega_0}\right)^2}{1+\sqrt{2}\frac{s}{\omega_{lp}}+\left(\frac{s}{\omega_{lp}}\right)^2}. \tag{5}$$

The combination $F_2(s)T_2(s)$ in effect moves the two poles of the closed-loop system from ω_0 on the real axis to $\pm 45°$ on the circle of radius ω_{lp}, as shown in Fig. 3. Feedforward can thus "cancel out" physical poles by placing zeros on top of them in the complex s-plane. At the same time, one adds other poles farther to the left (more stable).

The feedforward filters corresponding to the fourth- and sixth-order models are constructed in a similar way. For each pair of poles corresponding to a resonance, one puts a corresponding zero in the prefilter. Adding a zero over a resonance pole ensures that no component of the control signal $r(t)$ will be able to excite the mechanical resonance. One then compensates by increasing the order of the Butterworth cutoff filter by 2. Thus, the fourth-order model uses a fourth-order Butterworth filter, etc.

C. Discrete-time approximation to the feedforward filter

The next step in the frequency design is to approximate the continuous filter $F(z)$ by a discrete equivalent . This is done by substituting[5]

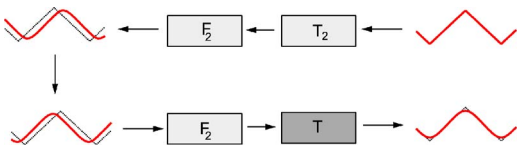

FIG. 4. Signal flow in the design of an acausal feedforward filter, using the second-order model. The light-shaded boxes represent dynamics that are coded on computer. The dark-shaded box represents the physical system, including its analog closed-loop control.

$$z=e^{T_s s}\approx\frac{1+sT_s/2}{1-sT_s/2} \tag{6}$$

or, equivalently,

$$s=\frac{2}{T_s}\frac{1-z^{-1}}{1+z^{-1}}, \tag{7}$$

which is known as the bilinear (Tustin) transform.[26] Because our sampling rate $T_s^{-1}=10$ kHz is much faster than the highest frequencies in the prefilter (250 Hz), the discretization algorithm is not crucial.[27] For the stage we have been discussing, we find

$$F_2(z)=\frac{a_0+a_1z+a_2z^2}{b_0+b_1z+z^2}, \tag{8}$$

with $a_0=26.5$, $a_1=-54.6$, $a_2=28.1$, $b_0=0.801$, and $b1=-1.78$. To calculate these coefficients, it is helpful to use a computer-algebra program or equivalent to carry out the substitution of Eq. (7) into Eq. (5). We used the open-source program SCILAB, which has many signal-processing and control algorithms.[28] Finally, by noting that z^{-1} has the interpretation "delay by T_s," we can divide by z^2 in Eq. (8) and convert our discrete filter into an "infinite impulse response" (IIR) digital filter of the form

$$r_n'=a_2r_n+a_1r_{n-1}+a_0r_{n-2}-b_1r_{n-1}'-b_0r_{n-2}', \tag{9}$$

where r_n' is the modified input at time nT_s, r_n is the desired output signal, and the a and b coefficients are taken from Eq. (8). Equation (9) is used to calculate the signal fed to the stage's input.

D. Acausal filtering

Feedforward, at least implicitly, makes use of prior knowledge about the system under control. In our case, the prefilter contains a partial inverse of the system's dynamics, which was explicitly measured beforehand. If we also know the desired future behavior of the control signal $r(t)$, we can use this knowledge to design a filter whose output has no phase lag with respect to the input—an "acausal filter."[29] A simple technique for designing acausal filters that is well-known to engineers[25] is illustrated in Fig. 4. One starts at the upper right of the figure, with the time-reversed version of the desired output signal. One passes this signal through a model of the physical closed-loop system (with transfer function T_2) and then through the prefilter F_2. [Because F_2 inverts the modeled dynamics of the physical system G, the product F_2T_2 is just the low-pass filter $L_2(s)$ with cutoff ω_{lp} in Eq. (5).] The output is illustrated at the upper left of the figure and is a low-pass-filtered, phase-

lagged version of the original signal (dotted line). One then time-reverses this output, as illustrated by the signal in the lower left. This time reversal converts phase lags to leads. When the time-reversed signal is passed again through the prefilter and then through the physical closed-loop system (T, in the dark-shaded box), the phase leads lag just enough to produce a net-zero phase shift. Because the signals pass through F_2 twice, the low-pass filtering is effectively four-pole, rather than two. (Compensating for this double filtering by using a single-pole filter in F leads to other complications because F is then an improper transfer function, with a frequency response that goes to infinity at high frequencies. It was simpler to use the same two-pole filter in both the causal and acausal cases.)

In principle, perfect acausal feedforward requires knowledge of the desired control signal into the indefinite future. In practice, one needs to know it only a short time ahead. In Fig. 4, we see that signals propagated backward in time will decay at a time scale $\tau_{\max}$ set by the longest decay time of the combined system $F_2(s)T_2(s)$. Thus, one need calculate only several times $\tau_{\max}$ into the future. Implicitly, we have done this in the work reported here, as we calculate waveforms over three periods of the driving signal. The beginning of the first period and the end of the last show transients. We extract the middle period and use it as the basis for a periodically repeated waveform sent to the stage. In the engineering literature, rather more complicated designs lead to explicit formula to compute the response in real time, with a finite amount of "preview" or "lookahead" required.[30,31] In our application, we do not vary the desired waveform during a scan, and the simpler method described here suffices.

IV. RESULTS

A. Stage performance

Figure 5 shows the stage response for three different driving frequencies (10, 40, and 100 Hz). The feedforward filter was calculated using the two-pole model. The goal was for the stage's motion to reproduce a triangular waveform (dotted line) with 1 μm amplitude. Figure 5(a) shows the stage's normal response to the driving signal. Because its bandwidth is less than 30 Hz, there is a noticeable phase lag already at 10 Hz. By 100 Hz, the response is completely unusable. Figure 5(b) shows the response to the modified input signal given by the causal feedforward algorithm (dashed line), which largely reproduces the desired signal, even at 100 Hz. The main differences are a small phase lag and a rounding of the corners consistent with the 250 Hz low-pass filter in the feedforward algorithm. Figure 5(c) shows the response of the acausal filter, which is similar to the causal case but without the phase lag.

In Fig. 5, at 100 Hz, one can observe a small amount of oscillation about the desired trajectory that comes from excitation of the mechanical resonance of the stage. In the 1 μm range shown in the figure, the oscillations have an amplitude of $\approx$3%, or 30 nm. Their amplitude is reduced to about 20 nm for the acausal filter. In that case, since the signal passes through the 250 Hz low-pass filter twice, the response is that of a four-pole filter rather than a two-pole,

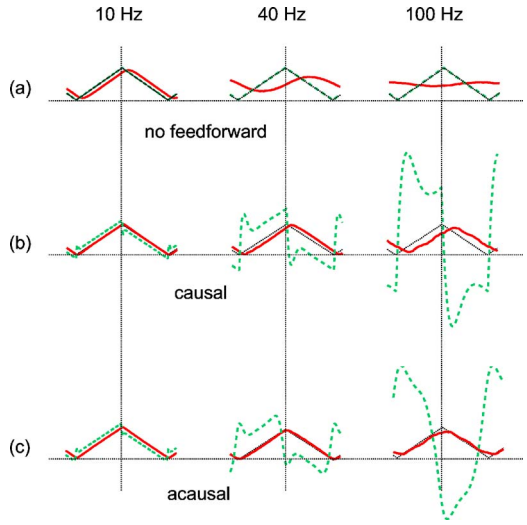

FIG. 5. Collage showing time series of measured stage responses using the two-pole dynamical model T_2 and its associated prefilter F_2. Dotted-line triangular waveform represents the desired stage response (1 μm amplitude). Dashed lines represent the signal fed to the stage. Solid lines represent the measured sensor signal. (a) Normal stage operation; (b) causal feedforward algorithm; (c) acausal feedforward algorithm. The phase shifts in the stage response observed in (a) and (b) are removed in (c). (Use the vertical dotted line as a reference.)

which reduces excitation of the resonance at the cost of more rounding of the waveform "corners." The resonance peaks are not compensated for in the second-order model. In Sec. IV B, below, we explore the performance of the fourth- and sixth-order models, which remove the first and second resonant peaks, respectively. Since our actual applications are more typically at around 10 Hz, where phase delays and rounding of the triangular waveform are an issue but excitation of the resonances is negligible, the second-order model will turn out to be sufficient. Another reason for favoring a lower-order model is that the mechanical resonance frequency changes significantly with load. Thus, a correction calculated for one load would be less effective at another load. (Devasia[32] has done a formal calculation of the effect of uncertainty on feedforward schemes. The conclusion matches the reasoning advanced here: feedforward is helpful only if the uncertainty in the dynamics is small. Since the system is linear, the size may be assessed at each frequency, with the overall conclusion being that in this case feedforward is useful at low frequencies but much less so at high frequencies.)

Another general issue is the choice of bandwidth for the feedforward stage. (In the example discussed here, the feedforward bandwidth $\omega_{lp}=9.3\omega_0$.) As ω_1 is increased, so will the magnitude of the modified input increase. The dashed curves showing the modified input signal r' in Fig. 5 illustrate this point clearly. As one goes from 10 to 40 to 100 Hz, the ratio of the range of the modified signal to the original increases from 1.2 to 5.7. A similar increase would occur (for fixed input frequency) as ω_{lp} is increased. The basic point is that requiring high-frequency motion requires large-amplitude inputs. For our applications, these limitations were not too severe because we are mostly interested in scans that are small (1–10 μm) compared to the overall range of the

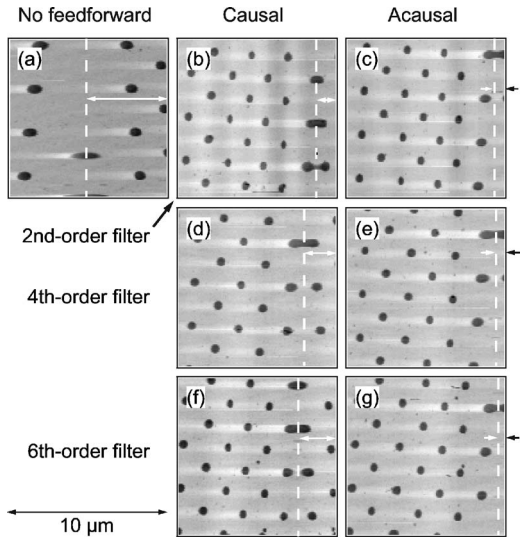

FIG. 6. AFM images of a calibration grating. All images are taken at 40 Hz ($v=800$ μm/s), scanning to the left. The white vertical dashed lines indicate the "turnaround" point in the image, with the distance to the right side of the image proportional to the phase lag. (a) Image taken by the scanner in its normal mode, without feedforward. (b), (d), (f): Images taken with causal feedforward filters of second, fourth, and sixth orders, respectively. (c), (e), (g): Same, with acausal filters.

stage (100 μm). The large overall range is important in being able to position the origin of the scanned images. Thus, even a nearly tenfold increase in amplitude reduces the available offset by only 6% for 1 μm scans. By contrast, large scans are more severely constrained: with a 9.3-fold increase, the largest possible scan is only about 16 μm. For scan frequencies limited to 10 Hz, the ratio between applied and nominal scan ranges is 1.2, which reduces the maximum scan range to about 80 μm. These limitations are peculiar to working with a stage under closed-loop control. The large required amplitudes in the control signal must "fight the low-pass filter" imposed by the closed-loop regulation. In the work on open-loop piezotube scanners,[6,9] there is no such low-pass filter and the control signals do not have to become large. On the other hand, one must then compensate for the nonlinear, hysteretic open-loop properties of the piezotube scanners.[7]

B. AFM images

We have tested the performance of the feedforward algorithm in our AFM. Figure 6 shows a series of images of a standard calibration sample. The left-hand image 6(a) shows an image taken at 40 Hz, not using the feedforward techniques discussed here. Because the stage has a two-pole, 42 Hz low-pass response, there is significant phase lag. The usual practice of AFM controllers is to hide the effects of phase lags by "overscanning"—i.e., one scans an image larger than desired, so that one can show only the central portion. In Fig. 6(a), the phase shift is roughly 90°, as indicated by the vertical white dashed line, which shows the "turnaround" point on the image. With no phase shift, this would coincide with the right edge of the image (which is scanned to the left). Here, the shift is about half the size of the image (or one-quarter the back-and-forth length). Figures 6(b), 6(d), and 6(f) show the results of applying causal feedforward filters of the form discussed in Sec. III C, with two, four, and six poles, respectively. The models for each are shown in the Bode plots of Fig. 2. In each case, we regularized the behavior of the prefilter by using a Butterworth filter of two, four, and six poles, with bandwidth set at 250 Hz. The phase lag is significantly reduced from 6(a). It increases with the order of the filter, simply because the phase shift in an nth-order filter is asymptotically $(\pi/2)n$.

Figures 6(c), 6(e), and 6(g) show images taken by acausal versions of the filters in Figs. 6(b), 6(d), and 6(f). Notice that the phase shift is reduced significantly but not eliminated. Notice, too, that it improves with the order of the filter. If the model of the system's dynamics were perfect, one would expect no phase shift. Since each model is accurate only up to some frequency limit, there is a residual phase shift that reflects the contributions from unmodeled dynamics. As the model improves, this residual phase shift decreases.

To summarize these results, we list in Table I the phase shifts measured from the images in Figs. 6(b)–6(g), for causal and acausal feedforward filters of second, fourth, and sixth orders, respectively, as well as the shifts predicted for Butterworth filters of the corresponding order. Numbers in the first column are phase shifts for the causal filters [Figs. 6(b), 6(d), and 6(f)]. With no modeling error, the phase delays of the causal filters should match those of the corresponding Butterworth filter, and we would expect no phase delay for the acausal filter. In Table I, we see extra phase shifts in both cases, which are due to the unmodeled dynamics. As the order of the dynamic model increases, these residual phase shifts decrease (the model is accurate to higher frequencies). Since the same unmodeled dynamics is present in both the causal and acausal cases, the difference between their phase lags (column 3) should match that of an ideal Butterworth filter (column 4). These differences agree reasonably well, within the accuracy of the phase-shift estimates.

Our feedforward algorithm also shows a dramatic improvement in the amplitude response, as demonstrated by comparing the (horizontal) spacing of dots on images scanned with and without feedforward. In Table II, we list

TABLE I. Phase shift (degrees) produced by different types of filters.

Filter order	Causal	Acausal	Difference	Butterworth
Second	−21	−10	−11	−13
Fourth	−36	−8	−28	−24
Sixth	−42	−7	−35	−36

TABLE II. Average spacing (μm) of dots measured on images with different types of causal filters, compared to actual spacing and that expected without feedforward.

Filter order	Causal	Expected
Second	2.73	2.84
Fourth	2.97	3.03
Sixth	2.72	2.70
Actual spacing		2.90
No feedforward		4.83

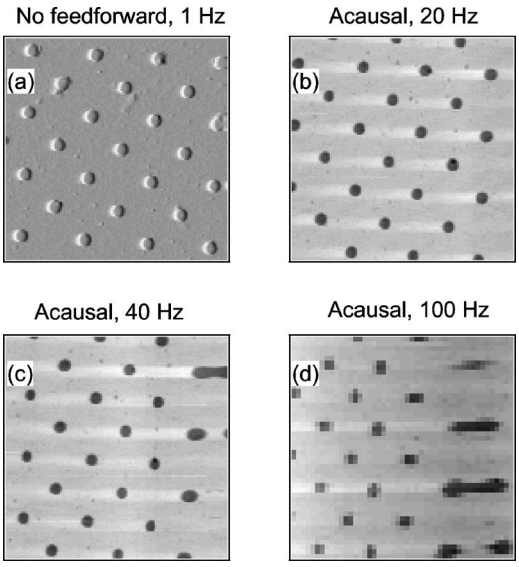

FIG. 7. AFM images of calibration grating. (a) "Standard" image, 1 Hz, without feedforward. (b)–(d) Images taken with a fourth-order feedforward filter, at indicated scan rates. All images are scanned over a $10\times10\ \mu m^2$ area.

the average horizontal spacings measured from Figs. 6(b), 6(d), and 6(f), together with those expected using the causal prefilters. (The acausal case is similar.) The expected values are obtained from the measured amplitude response [Fig. 2(a)] of the physical system, combined with the analytical forms of the causal prefilters, $F_2(s)$, $F_4(s)$, and $F_6(s)$. The actual spacing of 2.90 μm is estimated from Fig. 7(a), an image scanned at a low frequency under feedback. The uncertainty of ±0.08 μm for all measurements comes from the pixel resolution. The spacings observed using causal feedforward agree, within uncertainties, with the expected values. For the second-order filter, vibration excited by turnaround motion of the stage has distorted the part of image near the turnaround; therefore, the spacings of dots in this part are not counted toward the average value shown in the table. By contrast, if no feedforward is applied, the limited bandwidth implies a large distortion, to 4.83 μm at 40 Hz. [We cannot measure the spacing from the no-feedforward image in Fig. 6(a). The turnaround point of the image is shifted into its middle, and the horizontally related dots are actually images of the same structure.]

We see, then, the clear improvement in the amplitude response produced by all of the feedforward filters. The improvement, however, does not monotonically increase with the order of the filter. This results from an imperfect cancellation of resonances by the zeros in the various filters. The residuals slightly alter the gain at a given scan frequency. Usually, in AFM operation, one can calibrate measurements relative to features of known size in the images. In the rarer cases when absolute accuracy is important, one can do the kind of calculation done here to correct the scale at any given scan frequency. Since the response is predictable, it can be corrected easily.

Finally, Fig. 7 shows the performance of the acausal filter at different scan frequencies. Figure 7(a) shows a typical "normal" AFM image, scanned at 1 Hz, without feedforward. Figures 7(b)–7(d) show the results of scans at 20, 40, and 100 Hz using the fourth-order acausal filter discussed in the previous paragraph. Although the contributions of the unmodeled dynamics begin to be significant at 40 Hz, the image (away from the turnaround point) is accurate even at 100 Hz. At that frequency, the sampling rate of our AFM for the z dynamics (0.1 ms) limits the image resolution. This limitation has nothing to do with the translation stage or the feedforward dynamics.

V. DISCUSSION

We have shown that a simple feedforward algorithm can increase the usable bandwidth of a closed-loop piezoelectric translation stage by nearly an order of magnitude. Of course, if one starts with a stage with higher resonant frequencies, a higher scanning bandwidth could be obtained. There are stiffer versions of the design we use that claim a fivefold increase in resonant frequency.[20] We would expect that scan rates of up to 1 kHz would be possible on such stages. Other, more rigid stage designs with higher resonant frequencies[33] would allow even higher scanning frequencies.

Our choice for an upper-bandwidth limit was a compromise between the desire to improve the bandwidth and the desire to formulate a robust solution that would work for different mechanical loads and different laboratory temperatures. The exact tradeoffs between performance and robustness, though, should be set by the individual user—much as a motor is tuned differently for a race than for city driving. Croft and Devasia[6] show that one can optimize feedforward algorithms to obtain significant improvement in performance over simpler implementations. This optimization can be done by evaluating an integral over the square of the error between actual and desired control signal plus a term involving the control effort, with appropriate frequency-dependent relative weights. Since the actual solution depends critically on the choice of weight between small control effort and solution accuracy, the optimal method leads to waveforms similar to the ones used here, which were calculated in a more informal (simpler) way. In more recent work, Zou *et al.*[11] use an interesting variation where they numerically invert the measured transfer function rather than use an analytic fit. They then use the numerical inversion as the starting point for an iteration scheme where the candidate input waveform is sent through the system (a piezoelectric stick-slip rotary motor in their case), the output is measured, and the input is corrected in proportion to the difference between the actual and desired outputs. The algorithm converges in about ten iterations and gives roughly a fivefold improvement over the numerical inverse. (A numerical inverse will perform less well than an analytic fit, as measurement noise is included in the inverse dynamics.) The tradeoff is that the optimized input must be calculated (with the iterative process) for each situation (frequency, amplitude, load, etc.). More recently, Leang and Devasia have extended this technique by using adaptive methods to "learn" the characteristics of the physical system and adjust the system parameters accordingly.[34] Such a technique can handle slowly varying conditions. With all of these

more advanced techniques, one must decide whether the gain in performance is worth the additional effort in any given situation.

From a broader point of view, feedforward techniques can be helpful in many other situations. For example, feedback algorithms that are optimized to regulate a fixed value (of temperature, pressure, etc.) will not perform well when changing the setpoint. A feedforward prefilter can address this issue. Another category of applications is in the cancellation of disturbances. If independent measurements of the disturbance are available, they may be used to cancel their effect on the controlled system. In short, feedforward is a technique that deserves wider use, and combining feedforward with standard feedback control gives a particularly simple and robust approach.

ACKNOWLEDGMENTS

This research was supported by NSERC (Canada). The authors thank Connie Roth and Russ Greenall for comments and for help with testing the displacement stage on the atomic force microscope.

[1] D. Sarid, *Scanning Force Microscopy: With Applications to Electric, Magnetic, and Atomic Forces*, revised ed. (Oxford University Press, New York, 1994).
[2] M. A. Poggi, E. D. Gadsby, L. A. Bottomley, W. P. King, E. Oroudjev, and H. Hansma, Anal. Chem. **76**, 3429 (2004).
[3] B. D. Gates, Q. B. Xu, J. C. Love, D. B. Wolfe, and G. M. Whitesides, Annu. Rev. Mater. Res. **34**, 339 (2004).
[4] C. Bustamante, Y. R. Chemla, N. R. Forde, and D. Izhaky, Annu. Rev. Biochem. **73**, 705 (2004).
[5] J. Bechhoefer, Rev. Mod. Phys. **77**, 783 (2005).
[6] D. Croft and S. Devasia, Rev. Sci. Instrum. **70**, 4600 (1999).
[7] D. Croft, G. Shed, and S. Devasia, J. Dyn. Syst., Meas., Control **123**, 35 (2001).
[8] H. Perez, Q. Zou, and S. Devasia, J. Dyn. Syst., Meas., Control **126**, 187 (2004).
[9] G. Schitter and A. Stemmer, IEEE Trans. Control Syst. Technol. **12**, 449 (2004).
[10] G. Schitter, F. Allgöwer, and A. Stemmer, Nanotechnology **15**, 108 (2004).
[11] Q. Zou, C. Vander Giessen, J. Garbini, and S. Devasia, Rev. Sci. Instrum. **76**, 023701 (2005).
[12] C. Ru and L. Sun, Rev. Sci. Instrum. **76**, 095111 (2005).
[13] G. C. Goodwin, S. F. Graebe, and M. E. Salgado, *Control System Design* (Prentice-Hall, Upper Saddle River, NJ, 2001).
[14] W. E. Singhose, L. J. Porter, and W. P. Seering, "Input Shaped Control of a Planar Gantry Crane with Hoisting," American Control Conference, Albuquerque, NM, 1997.
[15] J. T. Feddema, C. R. Dohrmann, G. G. Parker, R. D. Robinett, V. J. Romero, and D. J. Schmitt, IEEE Control Syst. Mag. **17**, 29 (1997).
[16] The group at MIT led by W. P. Seering has been a particular champion of feedforward (or "input shaping") techniques. For a detailed and lucid introduction, see W. E. Singhose, Ph.D. thesis, MIT (1997).
[17] "Dynamic Digital Linearization" and "Input Shaping" for Polytec PI (Karlsruhe, Germany) translation stages. See www.polytecpi.com. The algorithms are licensed from Convolve, Inc.; see www.convolve.com.
[18] We implement an acausal feedforward filter to eliminate phase shifts. In addition, we use an infinite-impulse-response (IIR) filter, rather than a finite-impulse-response (FIR) filter.
[19] P. S. Maybeck, *Stochastic Models, Estimation, and Control* (Academic, New York, 1979), Vol. 1.
[20] Mad City Labs, Madison, WI. Stage Nano-H100, with a range of 100 μm. Another model, the Nano-PDQ series, is specified to have a resonant frequency five times that of our stage (the H100). See madcitylabs.com.
[21] For simplicity, we discuss only the *X*-stage response. The *Y*-stage has a similar response, with a lower mechanical resonance frequency, since the *Y*-stage supports the *X*-stage.
[22] Stanford Research Systems, SR850. The excitation was an 8 mV rms sine wave (equivalent to about 220 nm, peak-to-peak). The time constant was 1 s, with 8 s/data point in the sweep. See thinksrs.com.
[23] J. He and Z.-F. Fu, *Modal Analysis* (Butterworth-Heinemann, Oxford, 2001).
[24] The bandwidth is here defined, in the standard way, as being the frequency where the response amplitude first falls below $1/\sqrt{2}$. The stage's manufacturer claims a bandwidth of 200 Hz. Apparently, this discrepancy stems from their use of a nonstandard definition of bandwidth.
[25] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, 2nd ed. (California Technical Publishing, San Diego, 1999). Available on the web at DSPguide.com.
[26] G. F. Franklin, J. D. Powell, and M. Workman, *Digital Control of Dynamic Systems*, 3rd ed. (Addison-Wesley, Menlo Park, CA, 1998).
[27] One subtlety of going from the continuous to the discrete domain is that the discrete filter constructed using Eq. (7) has a different bandwidth. But, because the sampling rate is over 20 times the filter bandwidth, the bandwidth of the discrete filter differs from that of its continuous counterpart by only $\sim$0.3 Hz. Thus, although one can compensate for the shift by using "prewarping," we did not. See, for example, Ref. 26, Chap. 6.1.
[28] SCILAB is available from scilab.org.
[29] The term "noncausal" is also used. Both terms refer to the feature that a desired future behavior of the system is achieved by modifying the inputs in the present.
[30] Q. Zou and S. Devasia, IEEE Trans. Control Syst. Technol. **12**, 375 (2004).
[31] D. N. Hoover, R. Longchamp, and J. Rosenthal, Automatica **40**, 155 (2004).
[32] S. Devasia, IEEE Trans. Autom. Control **47**, 1865 (2002).
[33] G. E. Fantner *et al.*, Ultramicroscopy **106**, 881 (2006).
[34] K. K. Leang and S. Devasia, Mechatronics **16**, 141 (2006).