Preview-Based Optimal Inversion for Output Tracking: Application to Scanning Tunneling Microscopy

Qingze Zou, *Member, IEEE,* and Santosh Devasia, *Member, IEEE*

***Abstract*—Optimal inversion of system dynamics can be used to design inputs that achieve precision output tracking. However, a challenge in implementing the optimal-inversion approach is that the resulting inverse input tends to be noncausal. The noncausality of the optimal inverse implies that the desired output trajectory must be pre-specified and cannot be changed online. Therefore, the optimal inverse can only be used in trajectory-planning applications (where the desired output is known in advance for all future time). The main contribution of this article is the development of a technique to compute the noncausal optimal inverse when the desired output trajectory is known in advance for only a finite time interval. This future time interval, during which the desired output trajectory is specified, is referred to as the *preview time*. Additionally, this article develops a time-domain implementation of the optimal inverse and quantifies the required preview time in terms of the specified accuracy in output tracking, the system dynamics, and the cost function used to develop the optimal inverse. The proposed approach is applied to precision (subnanoscale) positioning of a scanning tunneling microscope (STM), which is a key enabling tool in emerging nanotechnologies. Experimental results are presented which show that finite preview of the desired output trajectory is sufficient to operate the STM at high speeds.**

***Index Terms*—Nanotechnology, output tracking, scanning, scanning tunneling microscope (STM), system inversion.**

I. Introduction

MODEL-BASED inversion of system dynamics [1]–[3], can be used to find inputs that achieve high-precision output tracking; this input is referred to as the inverse input. The inversion technique has been applied to a number of output-tracking applications; for example, in the precision control of flexible manipulators [4], [5], aircraft control [6], and high-precision positioning of piezo probes for nanoscale imaging using scanning-probe microscopy [7]. However, the model-based inversion approach suffers from two problems: 1) the inverse input will be erroneous if the modeling uncertainty is large and 2) the inverse input will be unacceptable if it violates input energy or bandwidth limitations. These two problems have been addressed by the development of the optimal-inversion technique in [8]. In particular, the optimal–inversion technique can be used to account for modeling errors by only inverting the system model in frequency regions where the modeling uncertainty is sufficiently small [9]. Additionally, actuator constraints such as input energy and bandwidth limitations can be accounted for by trading off the precision needed in output tracking. Thus, the optimal-inversion technique extends the standard-inversion theory to design inverse inputs in the presence of modeling uncertainties and actuator limitations. However, a challenge in implementing the optimal-inversion approach is that the resulting inverse input tends to be noncausal [9]. (Even the exact inverse input is noncausal when the system is nonminimum phase [2].) The noncausality of the optimal inverse implies that the desired output trajectory must be pre-specified for all time and cannot be changed online. Therefore, the optimal inverse can only be used in trajectory-planning applications (where the desired output is known in advance for all future time).

The main contribution of this article is the development of a technique to compute the noncausal optimal inverse when the desired output trajectory is known, in advance, for only a finite-time interval. This future time interval, during which the desired output trajectory is specified, is referred to as the *preview time*. We note that such a preview-based implementation enables the online specification of the desired output trajectory (see, e.g., [10]). Additionally, this article develops a time-domain implementation of the preview-based optimal inverse as opposed to the frequency-domain computation in [8]. The time-domain representation enables the quantification of the required preview time in terms of the specified accuracy in output tracking, the system dynamics and the cost function used to develop the optimal inverse.

The finite preview based optimal-inversion technique is illustrated by using it for precision positioning of a probe during high-speed surface imaging with a scanning tunneling microscope (STM). It is noted that the STM is a key enabling tool in emerging technologies such as nanofabrication [11]. However, in spite of its atomic-level resolution, the operating-speed (throughput) limitation prevents the STM-based nanofabrication from competing (at least at present) with more established techniques like electron beam (EB) and X-ray lithography [11]. A limitation to increasing STM's throughput is the positioning error caused by movement-induced vibrations in the piezo-based positioning system (i.e., piezo scanner). As the STM's scanning frequency (also called scan rate) is increased relative to the smallest resonant-vibrational frequency of the piezo scanner, the vibrational modes of the piezo scanner are excited. These movement-induced vibrations increase with

Manuscript received May 15, 2002; revised December 24, 2002. Manuscript received in final form June 17, 2003. Recommended by Associate Editor A. Kelkar. This work was supported by NASA Ames Research Center Grant NAG 2-1450 and NSF Grants CMS 0196214 and CMS 0301787.

The authors are with the Mechanical Engineering Department, University of Washington, Seattle, WA 98195-2600 USA (e-mail: qzouatuw@u.washington.edu; devasia@u.washington.edu).

Digital Object Identifier 10.1109/TCST.2004.824797

scan rate and result in errors in the STM probe's positioning, thereby limiting the maximum operating speed of the STM [7]. In practice, the achievable scan rate is substantially smaller (around 100 times smaller) than the smallest resonant-vibrational frequency of the STM scanner. This inability to operate the STM at high speed currently hinders the investigation and manipulation of ultrafast processes at the nanoscale. Therefore, there is a need to develop high-speed precision-positioning techniques for STM.

In general, the tracking performance of piezo-based positioning systems can be improved by using feedback control, for example, to reduce positioning errors due to creep and hysteresis (see, e.g., [12]–[18]). However, a problem with using feedback-based approach is the low-gain margin, of piezo-based positioners, that limits the achievable improvements because high-gain feedback tends to destabilize piezo-based STM scanners [19]. (*The low gain margin is due to low structural damping in piezo-actuators that results in high-quality factor* Q, i.e., *a sharp-resonant peak accompanied by a rapid-phase drop in the frequency response*.) In practice, a compromise is sought between performance and instability; feedback gains are adjusted to improve performance without instability. Thus, the tendency to become unstable at high gains (due to low-gain margins) has limited the success of typical feedback-based techniques to achieve high-speed positioning in STM applications. Furthermore, conventional sensors have relatively low resolution during high-speed operation (because the sensor noise tends to increase with operating speed); therefore, currently available sensors cannot be used to implement feedback controllers for subnanometer scale positioning when operating STMs at high speeds (under normal room temperatures).

Feedforward control approaches have been successful in increasing the operating speed of STM [20]. However, when the inverse input is noncausal, the computation of the inverse input requires prespecification of the entire output trajectory. This prevents the use of the inversion-based technique in online-STM applications such as nanofabrication, where the desired output trajectories may not be completely pre-specified and may have to be changed online. This problem is addressed by the development of the preview-based approach, which enables the online implementation of the optimal inverse. The preview-based optimal inversion technique is applied to a STM system and experimental results are presented to show that the STM can be operated at high speeds when only a finite preview of the desired output trajectory is available.

The paper is organized in the following format. The optimal-inversion approach and its finite-preview-based implementation is presented in Section II. In Section III, the preview-based approach is applied to a STM system and results (simulation and experimental) are presented and discussed in Section IV. Our conclusions are in Section V.

II. Preview-Based Optimal-Inversion

The optimal-inversion problem is presented as the minimization of a quadratic-cost function and the optimal inverse is obtained as a filter $G_{\text{opt}}(j\omega)$ (as developed in [8]). Properties of the optimal-inverse filter $G_{\text{opt}}(j\omega)$ are analyzed and a state-space representation of the optimal inverse is developed for a preview-based implementation of the optimal inverse.

A. Formulation and Solution of the Optimal Inversion Problem

Consider the following linear-time-invariant (LTI) system:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \tag{1}$$

where $x(t) \in \Re^n$ is the system state, and the number of inputs is the same as the number of outputs, $u(t) \in \Re^p$ and $y(t) \in \Re^p$, i.e., the system is square. The Laplace-domain representation of the system is given by

$$y(s) = C(sI - A)^{-1}Bu(s) \triangleq G(s)u(s). \tag{2}$$

We make the following assumptions.

Assumption 1: System (1) is invertible, i.e., there exists transfer matrix $G^{-1}(s)$ such that $G^{-1}(s)G(s) = G(s)G^{-1}(s) = I$, where I is the identity matrix. Moreover, system (1) and its inverse $(G^{-1}(s))$ are hyperbolic, i.e., $G(s)$ does not have poles or zeros on the imaginary axis $(s = j\omega)$ of the complex plane.

Remark 1: The invertibility of system (1) can be guaranteed if system (1) has a well-defined relative degree [3]. The requirement that the system (1) has no zeros on the imaginary axis is needed for computation [3] and robustness [9] of the exact inverse. The requirement of no poles on the imaginary axis is satisfied if the system is stable. (Alternatively, the optimal inversion problem could be posed after the system is stabilized, if the system is stabilizable.)

Optimal Inversion Problem—For a given sufficiently smooth, desired output trajectory, $y_d(\cdot) \in \mathcal{L}_{\mathbf{2}}$, the optimal-inversion problem is to minimize the following cost function [8]:

$$\begin{aligned} J = \int_{-\infty}^{\infty} \{ & u(j\omega)^* R(j\omega) u(j\omega) + [y(j\omega) - y_d(j\omega)]^* Q(j\omega) \\ & \times [y(j\omega) - y_d(j\omega)] \} \, d\omega \end{aligned} \tag{3}$$

where the superscript * denotes complex conjugate transpose and the dependence on ω indicates Fourier transforms of the signals. In the above cost function, $R(j\omega)$ is a frequency-dependent weight on the input energy and $Q(j\omega)$ is a frequency-dependent weight on the output-tracking error and are chosen such that Assumption 2 is satisfied.

Assumption 2: At each frequency ω, the weights $R(j\omega)$ and $Q(j\omega)$ are chosen to be Hermitian, symmetric, semi-definite matrices of $j\omega$ such that $G_{rq}(j\omega) \triangleq [R(j\omega) + G^*(j\omega) Q(j\omega)G(j\omega)]$ is positive-definite.

Remark 2: The requirement that the matrix $G_{rq}(j\omega)$ is positive-definite at each ω is less stringent than requiring $R(j\omega)$ to be positive-definite at each ω; as in standard optimal control theory [21]. The cost function (3) can be used to account for model uncertainty and actuator constraints by choosing appropriate weights $R(j\omega)$ and $Q(j\omega)$ [9], [20]. The approach can also be extended to nonsquare systems (actuator-redundant or actuator-deficient systems) as shown in [22].

Remark 3: A similar frequency-dependent quadratic performance index has been used in the past (e.g., [21]). However, the method in [21] aims at finding causal-control laws for output regulation ($Y_d = 0$) and causal-control laws may not minimize the above cost function. In contrast, we allow noncausal-control inputs to solve the optimal-inversion problem; these noncausal solutions are then implemented using a preview-based approach (as discussed in Section II-C).

Lemma 1: Solution to the Optimal Inversion Problem [8]—The optimal inverse can be found using a filter ($G_{\rm opt}(j\omega)$) as

$$\begin{aligned} u_{\rm opt}(j\omega) &= [R(j\omega) + G^*(j\omega)Q(j\omega)G(j\omega)]^{-1} \\ &\quad \times G^*(j\omega)Q(j\omega)y_d(j\omega) \\ &\triangleq G_{\rm opt}(j\omega)y_d(j\omega) \end{aligned} \tag{4}$$

where $G(j\omega)$ is the transfer function of the system (2) evaluated at $s = j\omega$, i.e., $G(j\omega) = G(s)|_{s=j\omega}$. Under the optimal control law (4), the output of the system tracks the following optimal desired trajectory:

$$y_{\rm opt}(j\omega) = G(j\omega)G_{\rm opt}(j\omega)y_d(j\omega). \tag{5}$$

Proof: See [8]. ∎

The optimal inverse was implemented offline in [8] by using the filter given in (4). The offline computations involved: 1) Fourier transformation of the pre-specified desired output-trajectory to obtain $y_d(j\omega)$; 2) multiplication by the optimal-inverse filter to find $u_{\rm opt}(j\omega)$; and 3) inverse Fourier transformation to find the optimal inverse in the time domain. This frequency-domain implementation requires the desired output trajectory to be pre-specified for all time $t \in (-\infty, \infty)$. In the following, a time-domain implementation is developed that enables a finite preview-based online implementation of the optimal inverse.

B. Properties of the Optimal Inverse

Properties of the optimal-inverse filter affecting its computation are investigated next. In particular, conditions are developed under which the optimal-inversion filter is hyperbolic and it will be shown that the optimal inverse tends to be noncausal.

Assumption 3: In the following, we assume that control-energy weight $R(s)$ and tracking-error weight $Q(s)$ are chosen as matrices whose entries are proper, real-coefficient, minimal, rational functions of the Laplace variable s, such that they satisfy Assumption 2 on the imaginary axis ($s = j\omega$). Furthermore, we assume that the weighting matrices $R(s)$ and $Q(s)$ are invertible (e.g., [23]).

Remark 4: Invertibility of the weighting matrices $R(s)$ and $Q(s)$ in Assumption 3 is not necessary for solving the optimal inversion problem. Rather, it enables us to analyze the properties of the optimal-inverse filter $G_{\rm opt}(s)$ defined in Lemma 1. For previous use of rational functions to design frequency-dependent weighting matrices, see, e.g., [21].

Lemma 2: Optimal-Inverse Filter in Transfer Function Form—Let the control-energy and tracking-error weights satisfy Assumption 3. Then the optimal-inverse filter can be represented by the following transfer function $G_{\rm opt}(s)$:

$$\begin{aligned} G_{\rm opt}(s) &= \left[R(s) + G^T(-s)Q(s)G(s)\right]^{-1} G^T(-s)Q(s) \\ &\triangleq G_{rq}^{-1}(s)G^T(-s)Q(s) \end{aligned} \tag{6}$$

where $G^T(s)$ denotes the transpose of the matrix $G(s)$.

Proof: This follows from (4) because $G^T(-s)$ evaluated at $s = j\omega$ is equal to $G^*(j\omega)$. ∎

The following analysis shows that the optimal-inverse filter $G_{\rm opt}(s)$ tends to be unstable; in such cases the bounded optimal inverse is necessarily noncausal. We begin by showing that $G_{\rm opt}(s)$ cannot have poles on the imaginary axis (of the complex plane), i.e., it is hyperbolic.

Lemma 3: Hyperbolicity of Optimal-Inverse Filter—Under Assumptions 1 and 3, the optimal-inverse filter $G_{\rm opt}(s)$ is hyperbolic.

Proof: From (6), $G_{\rm opt}(s)$ will have no poles on the imaginary axis if $G_{rq}^{-1}(s)$ [defined in (6)], $G^T(-s)$ and $Q(s)$ have no poles on the imaginary axis (i.e., they are hyperbolic). First, note that the poles of $G_{rq}^{-1}(s)$ (with finite values) are the zeros of $G_{rq}(s)$. Since $G_{rq}(j\omega)$ is positive-definite (by Assumption 2), $G_{rq}(s)$ cannot have zeros on the imaginary axis. Therefore, $G_{rq}^{-1}(s)$ can not have poles on the imaginary axis. Second, to show that $G^T(-s)$ is hyperbolic, note that the pole-location pattern of $G^T(-s)$ is a mirror image of the pole-location pattern of $G(s)$ with respect to the imaginary axis. This symmetric pole location pattern implies that poles of $G^T(-s)$ cannot be on the imaginary axis (by Assumption 1). Third, hyperbolicity of Q follows from Assumption 2 since Q is finite-valued at each point on the imaginary axis and, therefore, Q cannot have poles on the imaginary axis because it is minimal by Assumption 3. Thus, the hyperbolicity of the optimal-inverse filter follows from the hyperbolicity of $G_{rq}^{-1}(s)$, $G^T(-s)$ and $Q(s)$. ∎

Next, we consider how the poles of $G_{\rm opt}(s)$ move when the input energy weight is changed with respect to the tracking-error weight.

Lemma 4: Asymptotic Properties of the Optimal-Inverse Filter—Let the control-energy weight be of the form $R(s) = \rho R_o(s)$ where ρ is a positive constant. Furthermore, let Assumptions 1 and 3 be satisfied. Then

1) as the input energy weight is increased from zero to infinity (i.e., ρ: $0 \to \infty$), the poles of the optimal inverse filter $G_{\rm opt}(s)$ that have finite values move from the zeros of $G(s)$, $G^T(-s)$, $Q(s)$, and the poles of $R(s)$, to the poles of $G(s)$, $G^T(-s)$, $Q(s)$, and the zeros of $R(s)$; furthermore, the root locus of $G_{\rm opt}(s)$ does not intersect the imaginary axis for $0 \le \rho < \infty$;
2) when the input energy weight is small (i.e. $\rho \to 0$), the optimal-inverse filter tends to the inverse of the system, i.e., $G_{\rm opt}(s) \to G^{-1}(s)$; when the input energy weight becomes large (i.e., $\rho \to \infty$), the optimal-inverse input tends to zero.

Proof: We start the proof by defining the following coprime polynomial fraction of matrices $G^T(-s)Q(s)$, $R_o(s)$ and $G(s)$ (e.g., [23]):

$$\begin{aligned} G^T(-s)Q(s) &= D_{GQ}^{-1}(s)N_{GQ}(s) \\ R_o(s) &= N_R(s)D_R^{-1}(s) \\ G(s) &= N_G(s)D_G^{-1}(s). \end{aligned} \tag{7}$$

Statement 1: Using the coprime fraction defined in (7), (6) can be rewritten as (to simplify the notation, we omit the dependence on s)

$$\begin{aligned} G_{\text{opt}}(s) &= \left[\rho N_R D_R^{-1} + D_{GQ}^{-1}N_{GQ}G\right]^{-1} D_{GQ}^{-1}N_{GQ} \\ &= D_R\left[\rho D_{GQ}N_R + N_{GQ}GD_R\right]^{-1} N_{GQ}. \end{aligned} \tag{8}$$

Note that $G^T(-s)Q\ (s)$ is invertible (i.e., $N_{GQ}^{-1}(s)$ exists) because both $G(s)$ and $Q(s)$ are invertible (Assumption 1 and Remark 4). Therefore, by taking the limits on both sides of (8), we have

$$\begin{aligned} \lim_{\rho\to 0} G_{\text{opt}}(s) &= \lim_{\rho\to 0} D_R[\rho D_{GQ}N_R + N_{GQ}GD_R]^{-1}N_{GQ} \\ &= D_R(N_{GQ}GD_R)^{-1}N_{GQ} \\ &= G^{-1}(s). \end{aligned} \tag{9}$$

Equations (7) and (9) show that as $\rho \to 0$, the poles of $G_{\text{opt}}(s)$ move toward the zeros of $Q(s)$, $G(s)$, $G^T(-s)$ and the poles of $R(s)$, and the optimal-inverse filter approaches to the inverse of the system $G_{\text{opt}}(s) = G^{-1}(s)$. Note that pole-zero cancellations occur when $\rho = 0$.

Similarly, to find the locations of the poles of $G_{\text{opt}}(s)$ when $\rho \to \infty$, we rewrite (6), by using the coprime fraction defined in (7) as

$$\begin{aligned} G_{\text{opt}}(s) &= \left[\rho R_o + D_{GQ}^{-1}N_{GQ}N_G D_G^{-1}\right]^{-1} D_{GQ}^{-1}N_{GQ} \\ &= \frac{D_G}{\rho}\left[D_{GQ}R_oD_G + \frac{1}{\rho}N_{GQ}N_G(s)\right]^{-1} N_{GQ}. \end{aligned} \tag{10}$$

Note the term $D_{GQ}R_oD_G$ in (10) is invertible since $R(s) = \rho R_o(s)$ is invertible (Remark 4). Therefore, by taking the limits on both sides of (10) with $1/\rho \to 0$ (i.e., $\rho \to \infty$), we have

$$\begin{aligned} \lim_{\frac{1}{\rho}\to 0} G_{\text{opt}}(s) &= \lim_{\frac{1}{\rho}\to 0} \frac{1}{\rho} D_G[D_{GQ}R_oD_G]^{-1}N_{GQ} \\ &= \lim_{\frac{1}{\rho}\to 0} \frac{1}{\rho} D_G D_G^{-1} D_R N_R^{-1} D_{GQ}^{-1} N_{GQ} \\ &= 0. \end{aligned} \tag{11}$$

As $1/\rho \to 0$ (i.e. $\rho \to \infty$), the poles of $G_{\text{opt}}(s)$ move toward the poles of $G(s)$, $G^T(-s)$, $Q(s)$, and the zeros of $R(s)$. The hyperbolicity of the root locus of $G_{\text{opt}}(s)$ follows from Lemma 3. This completes the proof of Statement 1 of the Lemma.

Statement 2 of the Lemma follows from (9) and (11). Equation (11) shows that as $\rho \to \infty$, $G_{\text{opt}}(s)$ is zero for all s but a finite number of complex number s (at the poles of $G_{\text{opt}}(s)$), hence, the optimal-inverse input obtained from the inverse-Laplace transform of $u_{\text{opt}}(s) = G_{\text{opt}}(s)y_d(s)$ is zero. ■

Remark 5: Inputs from standard linear quadratic (LQ) optimal control theory with cheap control (i.e., low weight on the input energy) yields perfect tracking for minimum phase systems [24]. However, standard LQ controllers have performance limitations when the system is nonminimum phase [25]. In contrast, the noncausal optimal-inversion approach finds bounded inputs that can yield perfect output tracking in the cheap control case—even for the nonminimum phase case. The following corollary describes the root locus of the optimal-inverse filter $G_{\text{opt}}(s)$ for a single-input single-output (SISO) system, which follows directly from Lemma 4.

Corollary 4.1: SISO Case—Let assumptions in Lemma 4 be satisfied, let the system (1) be SISO and let the weights on the input energy and tracking error be constants in the frequency domain, i.e., $R(s) = \rho$, $Q(S) = 1$. Then, as the input energy weight tends to infinity (ρ: $0 \to \infty$), the root locus of the optimal-inverse filter $G_{\text{opt}}(s)$ starts at the zeros of $G(s)$ and $G(-s)$ and end at the poles of the $G(s)$ and $G(-s)$. Therefore, the root locus is symmetric with respect to both the real and imaginary axes and does not intersect the imaginary axis.

Remark 6: The root locus of the optimal inverse (described in Corollary 4.1) includes the root locus of standard optimal control that uses causal inputs. For example, when optimal constant feedback $u = -k_o x$ is used in a SISO system $G(s) = n(s)/d(s)$, where k_o is obtained by minimizing

$$J = \int_0^\infty \rho u(t)^2 + [y(t) - y_d(t)]^2\, dt$$

the poles of closed-loop system are the left-half-plane roots of the Chang–Letov Equation (summarized in [23, Ch. 3])

$$n(-s)n(s) + \rho d(-s)d(s) = 0.$$

Therefore, the root locus of the closed-loop system by using $u = -k_o x$ starts at the zeros of $G(s)$ and $G(-s)$ that are on the left half of the complex plane and end at the stable poles of the $G(s)$ and $G(-s)$. Thus, the root locus of the closed-loop system, obtained by using the causal control law $u = -k_o x$, is the stable part of the root locus of the optimal-inverse filter $G_{\text{opt}}(s)$.

Remark 7: Optimal inverse filter is unstable—Under the assumptions of Lemma 4, the poles of the optimal filter $G_{\text{opt}}(s)$ approach to the zeros of $G(s)$ and $G^T(-s)$ as the input energy weight becomes small, i.e., as high-precision output tracking is required. Moreover, since the optimal inverse is hyperbolic, root locus branches associated with the nonminimum-phase zeros of $G(s)$ and $G^T(-s)$ will remain on the right hand side of the imaginary axis (in the complex plane) as the weight on the input energy is increased (from Lemma 4). Therefore, for general choice of weights $R(s)$ and $Q(s)$, the optimal inverse filter tends to be unstable if the system (1) has some internal dynamics (i.e., has at least one zero).

C. Preview-Based Optimal Inverse

The noncausality of the optimal-inverse filter motivates the development of the preview-based implementation approach that is discussed in this subsection. We begin by rewriting the

optimal-inverse filter as a proper transfer-function matrix to develop a state-space representation of the optimal inverse.

Lemma 5: Optimal Inversion using a Proper Transfer Function Matrix—Let the desired output trajectory $y_d(t)$ be sufficiently smooth. If the optimal-inverse filter $G_{\rm opt}(s)$ is not proper, then the optimal inverse can be rewritten as

$$u_{\rm opt}(s) = \hat{G}_{\rm opt}(s)\hat{y}_d(s) \tag{12}$$

where $\hat{G}_{\rm opt}(s)$ is proper, and $\hat{y}_d(s)$ denotes the Laplace transform of the linear combination of the output and its time derivatives. Furthermore, $\hat{G}_{\rm opt}(s)$ is hyperbolic if $G_{\rm opt}(s)$ is hyperbolic.

Proof: If $G_{\rm opt}(s)$ is proper, then we define $\hat{G}_{\rm opt}(s) = G_{\rm opt}(s)$, $\hat{y}_d(s) = y_d(s)$. Otherwise, the optimal inversion filter is rewritten as

$$G_{\rm opt}(s) = \frac{N(s)}{den(s)} \tag{13}$$

where $den(s)$ is the least common denominator polynomial of $G_{\rm opt}(s)$ and each entry in the $p \times p$ numerator matrix $N(s) = N_{i,j}(s)(i,j \le p)$ is a polynomial of the Laplace variable s. Let the order of polynomials in the numerator matrix N be $O_{\rm num}$. Furthermore, let the order of the denominator polynomial $\mathrm{den}(s)$ be $O_{\rm den}$. Then, each entry $N_{i,j}(s)$ in the numerator matrix is rewritten as

$$N_{i,j}(s) = a_{i,j}(s)b_{i,j}(s) \qquad \text{for all } i,j \le p \tag{14}$$

where $a_{i,j}(s)$ and $b_{i,j}(s)$ are polynomials of s with real coefficients and are chosen such that the order of $a_{i,j}(s)$ is less than or equivalent to $O_{\rm den}(s)$. Then, the modified proper optimal-inverse filter $\hat{G}_{\rm opt}(s)$ can be obtained as

$$\begin{aligned} u_{\rm opt}(s) &= G_{\rm opt}(s)y_d(s) = \frac{N(s)}{\mathrm{den}(s)}y_d(s) \\ &= \frac{1}{\mathrm{den}(s)}\begin{bmatrix} A_1(s) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & A_2(s) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & A_p(s) \end{bmatrix} \\ &\quad \times [B_1(s)\, B_2(s)\, \cdots\, B_p(s)]^T\, y_d(s) \\ &\triangleq \hat{G}_{\rm opt}(s)\,[B_1(s)\, B_2(s)\, \cdots\, B_p(s)]^T\, y_d(s) \\ &\triangleq \hat{G}_{\rm opt}(s)\hat{y}_d(s) \end{aligned} \tag{15}$$

where

$$\begin{aligned} A_i(s) &\triangleq [a_{i,1}(s),\, a_{i,2}(s),\, \cdots,\, a_{i,p}(s)], \\ B_i(s) &\triangleq \begin{bmatrix} b_{i,1}(s) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & b_{i,2}(s) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & b_{i,p}(s) \end{bmatrix} \end{aligned} \tag{16}$$

for $1 \le i \le p$. Note that the poles of $G_{\rm opt}(s)$ and $\hat{G}_{\rm opt}(s)$ will be the roots of $den(s)$ with possible repetitions (see, e.g., [23]). Therefore, the hyperbolicity of the optimal-inverse filter $G_{\rm opt}(s)$ implies the hyperbolicity of $\hat{G}_{\rm opt}(s)$. ■

Remark 8: The order of the output's time derivative used to define $\hat{y}_d(s)$ is the difference between the order of the numerator polynomials $(O_{\rm num})$ and the order of the denominator polynomial $(O_{\rm den})$ of the optimal-inverse filter, i.e., $\max(0, (O_{\rm num} - O_{\rm den}))$. It is noted that a similar redefinition of the desired output (in terms of its time derivatives) was used previously to develop a state-space realization of the exact inverse in [1].

Remark 9: The input energy weight $R(s)$ can be chosen to make $G_{\rm opt}(s)$ a proper rational function matrix, for example, by choosing the order of the numerator polynomials in $R(s)$ to be sufficiently large.

State-space representation of the Optimal Inverse—The modified optimal-inverse filter $\hat{G}_{\rm opt}(s)$ is hyperbolic (from Lemmas 3 and 5) and can be decoupled into stable and unstable parts by partial fraction expansion as

$$\hat{G}_{\rm opt}(s) = G^s_{\rm opt}(s) + G^u_{\rm opt}(s) \tag{17}$$

where $G^s_{\rm opt}(s)$ has its poles on the left half plane, and $G^u_{\rm opt}(s)$ has its poles on the right-half plane. Furthermore, let the minimal realization (e.g., [23]) of the stable and unstable portions of the optimal inversion filter be given by

$$\begin{aligned} \left[G^s_{\rm opt}(s)\right] \qquad \dot{x}_s(t) &= A_s x_s(t) + B_s\hat{y}_d(t) \\ u^s_{\rm opt}(t) &= C_s x_s(t) + D_s\hat{y}_d(t) \\ \left[G^u_{\rm opt}(s)\right] \qquad \dot{x}_u(t) &= A_u x_u(t) + B_u\hat{y}_d(t) \\ u^u_{\rm opt}(t) &= C_u x_u(t) + D_u\hat{y}_d(t). \end{aligned} \tag{18}$$

The bounded solution to the optimal inversion problem is then obtained by flowing the stable portion forward in time and flowing the unstable portion backward in time [3] as stated in the next lemma.

Lemma 6: Bounded Optimal Inverse—Let desired output and its time derivatives $\hat{y}_d(t)$ be bounded uniformly in time $t \in (-\infty, \infty)$. The bounded solution to the optimal-inverse input $u_{\rm opt}(t)$ for all time $t \in (-\infty, \infty)$ is found as

$$\begin{aligned} u^s_{\rm opt}(t) &= C_s \int_{-\infty}^{t} e^{A_s(t-\tau)}B_s\hat{y}_d(\tau)d\tau + D_s\hat{y}_d(t) \\ u^u_{\rm opt}(t) &= -C_u \int_{t}^{\infty} e^{-A_u(\tau-t)}B_u\hat{y}_d(\tau)d\tau + D_u\hat{y}_d(t) \\ u_{\rm opt}(t) &= u^u_{\rm opt}(t) + u^s_{\rm opt}(t). \end{aligned} \tag{19}$$

Proof: This follows from arguments used to find bounded solutions to unstable internal dynamics in [3]. ■

Remark 10: The requirements that $\hat{y}(\cdot)$ be bounded in the above Lemma can be relaxed to the requirement that the following terms are bounded: $B_s\hat{y}_d(\cdot)$, $B_u\hat{y}_d(\cdot)$, $D_s\hat{y}_d(\cdot)$, and $D_u\hat{y}_d(\cdot)$.

Finite-preview-based Implementation of the Optimal-Inverse Filter—The computation of the optimal-inverse input, at any time t, requires all future values of the desired output trajectory. In particular, the computation of $u^u_{\rm opt}(t)$ in (19) requires knowledge of all future values of the desired output $\hat{y}_d(t)$ for time interval $[t, \infty)$. However, $u^u_{\rm opt}(t)$ can be approximated by

truncating the integral in (19), by only using information of the desired output trajectory during a finite time interval $[t, t+T_p]$, as

$$\tilde{u}^u_{\rm opt}(t) = -C_u \int_t^{t+T_p} e^{-A_u(\tau-t)} B_u \hat{y}_d(\tau) d\tau + D_u \hat{y}_d(t) \quad (20)$$

and the finite-preview based optimal-inverse input is given by

$$\tilde{u}_{\rm opt}(t) = u^s_{\rm opt}(t) + \tilde{u}^u_{\rm opt}(t). \quad (21)$$

Error due to Finite-preview-based Implementation—The output-tracking error due to the finite-preview based implementation [i.e., using input found from (21) rather than infinite preview (19)] is quantified, next. We first define the following condition 1.

Condition 1: Positive constants K_1, K_2, α, β satisfy condition 1 if

$$\|e^{At}\|_2 \le K_1 e^{-\beta t}, \quad \|e^{-A_u t}\|_2 \le K_2 e^{-\alpha t}, \quad \forall t \ge 0. \quad (22)$$

Remark 11: The above Condition 1 is equivalent to system (1) being stable (Assumption 1) and its corresponding optimal inverse $\hat{G}_{\rm opt}(s)$ being hyperbolic (Lemma 3). In this case, K_1 and K_2 satisfying (22) can always be found if β and α are chosen such that $\beta < \inf_i |Re(\lambda_i(As))|$ and $\alpha < \inf_i Re(\lambda_i(A_u))$, where $Re(\lambda_i(A))$ denotes the real part of the $i^{\rm th}$ eigenvalue of matrix A.

Next, let the output obtained by applying the finite-preview based optimal-inverse input $\tilde{u}_{\rm opt}(t)$ be denoted by $\tilde{y}_{\rm opt}(t)$, and let the resulting error in output tracking be denoted by

$$e_y(t) = y_{\rm opt}(t) - \tilde{y}_{\rm opt}(t). \quad (23)$$

The tracking error $e_y(t)$ can be made arbitrarily small by choosing a sufficiently large preview time T_p.

Lemma 7: Let K_1, K_2, α, and β satisfy condition 1. Then for any given number $\epsilon > 0$, there exists a finite time T_p^*

$$T_p^* \triangleq \frac{1}{\alpha} \ln\left(\frac{\bar{K}}{\alpha\beta\epsilon}\right) \quad (24)$$

such that the output-tracking error is smaller than ϵ

$$\|e_y(t)\|_2 \le \epsilon = \frac{\bar{K}}{\alpha\beta} e^{-\alpha T_p^*}, \qquad \text{for all } t \in (-\infty, \infty) \quad (25)$$

for the preview time $T_p \ge T_p^*$, where

$$\bar{K} \triangleq K_1 K_2 K_y \|C_u\|_2 \|B_u\|_2 \|C\|_2 \|B\|_2$$

is a finite positive constant, $\|F\|_2$ denotes the induced matrix 2-norm of matrix F, $K_y = \sup_{t\in\Re} \|\hat{y}_d(t)\|_2$ is the bound on the desired trajectory, B, C are as defined in (1) and B_u, C_u are as defined in (18).

Proof: The proof of this lemma follows from arguments similar to those in [10] and is therefore omitted. ■

The above Lemma shows that: 1) the tracking error exponentially decays with the increase of preview time and 2) the minimum preview time T_p^* needed for precision output tracking within a specified output-tracking error (ϵ) decreases as α (see Remark 11) increases. It is noted that the maximum value of α is less than the minimum distance

$$L_{\rm min} \triangleq \inf_i {\rm Re}\,(\lambda_i(A_u))$$

of the unstable poles of the optimal-inverse filter from the imaginary axis, as illustrated in Fig. 1.

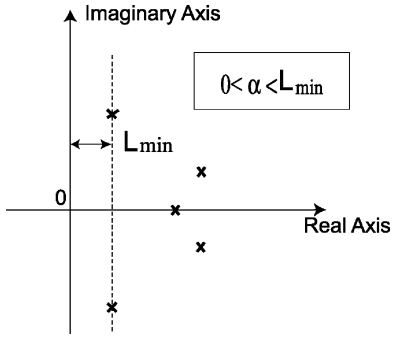


Fig. 1. Dependence of the preview-time T_p^* on the unstable poles of the optimal inversion filter $G_{\rm opt}(s)$.

Remark 12: Due to the exponential decay of the tracking error with the distance of the unstable poles [of the optimal inversion filter $G_{\rm opt}(s)$] from the imaginary axis [see (25)], a rule of thumb is to choose the preview time T_p to be greater than four times the time constant of the dominant unstable pole (i.e., four times the inverse of the dominant unstable pole's distance from the imaginary axis).

III. Example: Application to Scanning Tunneling Microscope (STM)

The preview based optimal-inversion approach is applied to a STM (Metris-1000, Burleigh Instruments, Inc.). The goal is to demonstrate that the use of finite preview-based optimal-inversion approach enables high-speed STM imaging.[1] We begin by describing the STM system and its model.

A. Burleigh Metris-1000 STM Scanner

Principle of STM operation—The principle of STM operation is shown in Fig. 2. A piezo scanner moves the STM probe across the sample surface as the tunneling current between the STM probe and the sample (for an applied voltage) is measured. The tunneling current depends on the distance between the STM probe and the sample surface. Therefore, the measured tunneling current is directly related to the sample's surface topology and is used to form images of the sample's surface topology. A critical component of a STM system is the piezo scanner that positions the STM probe over the sample surface [see Fig. 2(a)]. The piezo scanner moves the STM probe parallel to the sample surface (x–y axes) as well as perpendicular to the sample surface (z–axis). Errors in positioning the STM probe can lead to distortions in the acquired images, since the position error leads to the incorrect mapping of the tunneling current signal to the x–y position of the STM probe. As STM's operation speed increases, movement-induced vibrations lead to significant positioning errors and image distortion. It is noted that

[1]The MATLAB code can be obtained by e-mail to the authors.

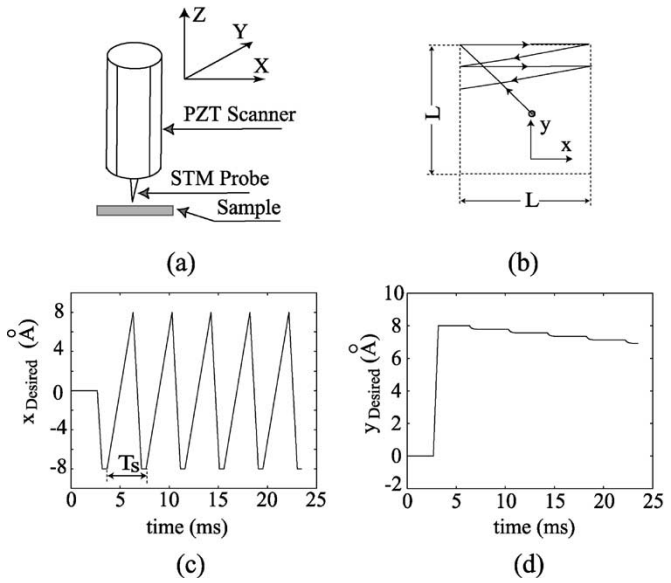


Fig. 2. (a) STM scanner scheme and (b) scan trajectory, both drawings are not to scale. The corresponding desired trajectory for 250 Hz scanning in (c) x-direction and (d) y-direction.

post-correction techniques can be applied to remove the effects of positioning error in imaging applications [19]. However, such post-processing approaches cannot be used when online correction is critical, for example, in nanofabrication applications because post-compensation cannot prevent distortions in the fabricated components. Moreover, the positioning error, if large, can also lead to sample and/or tip damage (even in imaging applications). Therefore, online precision positioning of the STM probe over the sample is important.

B. Scanning of the STM Probe Over the Sample Surface

The scanning motion of the STM probe is illustrated in Fig. 2(b). To start the scanning process, the STM probe is moved from the center to the top-left point of the image area. Then, the tip is moved in a raster pattern to image a square area (L Å $\times$ L Å, where Å represents Angstrom). During the forward scan (from left to right), the tunneling current signal is measured at a constant sampling rate, and the y-position is fixed. During the return scan (from right to left), the y-position is incremented and x-position is returned to the left. The tunneling current signal is not collected during the return scan. The back and forth scanning pattern is repeated until the entire desired area is imaged. During the imaging operation the z-position of the STM probe is not changed (this is referred to as *constant height mode*) and the measured tunneling current is plotted against the desired x–y position to generate an image of the surface topology. (*Images are plotted against the desired x–y position rather than the actual x–y position because direct measurement of the STM probe's tip-position is not available during imaging, see* Section IV.) The desired x–y trajectory is shown in Fig. 2(c) and (d). The scan rate of the imaging process is defined as $s_r = 1/T_s(Hz)$, where T_s is the time to finish one complete scan as indicated in Fig. 2(c).

C. STM Model

The dynamic model of the STM system was obtained experimentally with a dynamic signal analyzer (DSA). To find the dynamic model of the STM probe positioning in x-direction (referred to as the x-dynamics for simplicity), we applied an input voltage $u_x(t)$ generated by the DSA to drive the x-dynamics in the micrometer (μm) displacement range and fed the position signal (measured with an inductive eddy-current sensor that has a static resolution of five Å) back to the DSA. Experimental frequency response was obtained from the DSA and used to identify the system model. The transfer function $G_x(s)$ of the *x-Dynamics*, from the input voltage u_x (in volt) to the piezo-position x (in Å), was found as

$$G_x(s) = \frac{x(s)}{u_x(s)} = K_{xp}\frac{\prod_{q=1}^{4}(s-z_q)}{\prod_{r=1}^{6}(s-p_r)} \tag{26}$$

where $K_{xp} = 29.28$, $z_q = \{9.274 \pm 41.659, -2.484 \pm 30.434i\}$, $p_r = \{-0.188 \pm 31.326i, -0.857 \pm 24.570i, -7.263, -3.198\}$, and the Laplace transform variable s is in rad/ms. Similarly, the transfer function $G_y(s)$ of *y-Dynamics*, from the input voltage u_y (in volts) to the piezo-position y (in Å), was found as

$$G_y(s) = \frac{y(s)}{u_y(s)} = K_{yp}\frac{\prod_{q=1}^{4}(s-z_q)}{\prod_{r=1}^{6}(s-p_r)} \tag{27}$$

where $K_{yp} = 15.26$, $z_q = \{-0.7135 \pm 26.6719i, 44.8250 \pm 63.1009i\}$, $p_r = \{-0.3722 \pm 25.6469i, -0.2239 \pm 31.3441i, -8.566, -3.866\}$.

D. Root Locus of the Optimal-Inverse Filter

In this section, the optimal-inversion technique is applied to the STM and properties of the resulting optimal-inverse filter are investigated.

Choice of Weights in the Cost Function—The optimal-inverse filter was chosen such that perfect tracking is achieved at low frequencies (in the operating bandwidth of 444 Hz for the x-dynamics) and to reduce the input required at high frequencies (at which the model also tends to become less accurate). This implies that the tracking-error weight $q(j\omega)$ should be larger than the input energy weight $r(j\omega)$ in the low-frequency range and vice versa ($q(j\omega) < r(j\omega)$) in the high-frequency range. For this example, $q(s)$ and $r(s)$ were chosen as

$$\begin{aligned} r(s) &= \rho\frac{\left(1-\frac{s}{4}\right)\left(1+\frac{s}{4}\right)}{\left(1-\frac{s}{80}\right)\left(1+\frac{s}{80}\right)} \triangleq \rho r_o(s) \\ q(s) &= \frac{1}{r_o(s)} \end{aligned} \tag{28}$$

where ρ was used to adjust the relative magnitude of the input energy and tracking-error weights. In this example, $\rho = 1/400$ is chosen, which yields a large tracking-error weight $|q(j\omega)| > 10|r(j\omega)|$ for low frequencies $\omega \leq 1$ KHz and a large input energy weight $|r(j\omega)| > 10|q(j\omega)|$ for high frequencies $\omega \geq 10$ KHz, as shown in Fig. 3.

Root Locus of the Optimal-inverse Filter—For the above choice of the cost function ($r(j\omega)$ and $q(j\omega)$), the optimal-inverse filter can be written as [see (4)]

$$G_{\text{opt},x}(s) = \frac{G_x(-s)q(s)}{\rho r_o(s) + G_x(-s)q(s)G_x(s)}. \tag{29}$$

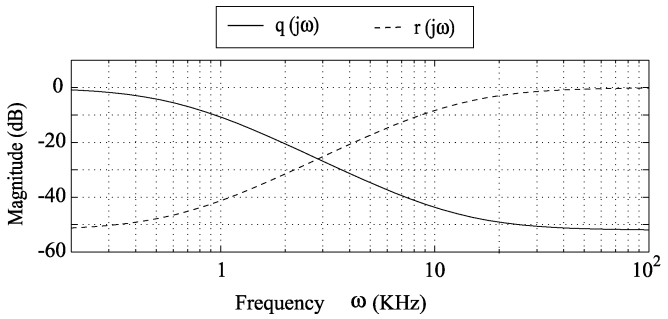


Fig. 3. Magnitude bode plots of the input energy weight $r(j\omega)$ (dash line) and the tracking-error weight $q(j\omega)$ (solid line).

As stated in Lemma 4, when ρ is changed from zero to ∞, the poles of the optimal-inverse filter $G_{\text{opt},x}(s)$ move continuously from (i) the zeros of $G_x(-s)q(s)G_x(s)$ and the poles of $r(s)$ to (ii) the poles of $G_x(-s)q(s)G_x(s)$ and the zeros of $r(s)$. The root locus of $G_{\text{opt},x}(s)$ with respect to ρ is shown in Fig. 4, where the arrows indicate the orientation of root locus as ρ: $0 \rightarrow \infty$. It is noted that the root locus is symmetric with respect to both the real and imaginary axis. The unstable poles of $G_{\text{opt},x}(s)$ move towards the imaginary axis as ρ is increased, which implies an increase in minimum preview time (see Remark 12). It is noted that the piezo dynamics is nonminimum phase [has zeros on the right-hand side, see (26) and (27)]. Therefore, from Lemma 4 and Remark 7, the optimal-inverse filter is unstable for all values of ρ, as can be seen from Fig. 4.

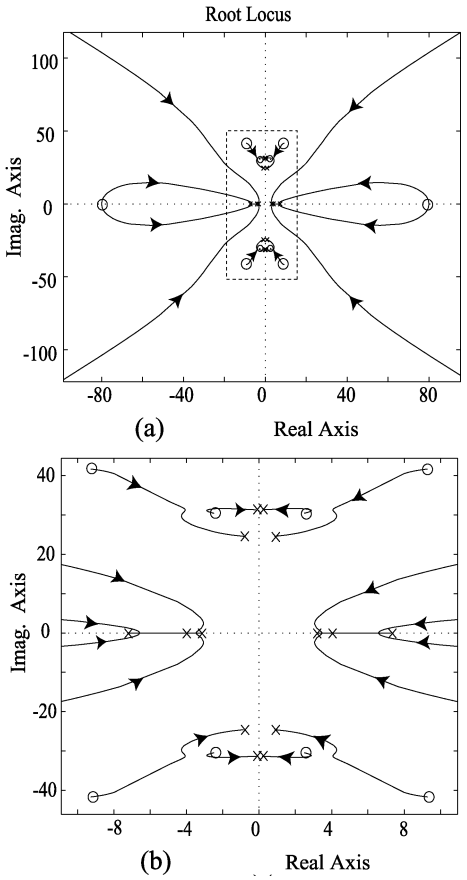


Fig. 4. *Symmetric root locus* of the optimal-inverse filter $G_{\text{opt},x}(s)$ as $\rho = 0 \rightarrow \infty$, where (b) is the zoomed-in view of the region inside the dashed line in (a) and the unit is in rad/ms.

IV. Results and Discussions

Simulations in MATLAB and experimental implementations were carried out to show that a finite-preview time is sufficient for on-line implementation of the optimal inversion.

A. Choice of Experiment

The main advantage of the STM is that it enables the investigation and manipulation of surface phenomena with subnanoscale spatial resolution. Therefore, the ability to image subnanoscale features at high speeds when using the proposed optimal-inversion technique was tested using the experimental STM system. In particular, the STM was used to image carbon atoms in highly oriented pyrolytic graphite (HOPG), in which the spacing of the imaged atoms is less than 1 nm (2.46 Å). It is noted that the crystal symmetry and its low thermal expansion coefficient ($< 10^{-6}K^{-1}$) of HOPG makes it a standard sample used to characterize STM systems (e.g., [26], [27]). Since the features on the calibration sample (HOPG) do not change with the choice of the scan trajectory, distortions in the acquired image, if any, are caused by the inability to precisely position the STM probe along the scan trajectory. Thus, distortions in the acquired image provides a measure of the STM scanner's tracking performance when imaging subnanoscale features.

The use of image distortions to quantify positioning error is necessary when imaging and manipulating at the subnanometer scale using the STM. It is noted that conventional sensors such as capacitive, inductive, and laser sensors can be used to calibrate and model SPMs when the scan size (i.e., lateral positioning range) is in the hundreds of nanometers [12]. However, such sensors cannot be used to calibrate or provide feedback to correct the SPM when imaging with subnanometer resolution. For example, the static resolution of the inductive sensor (eddy-current sensor, that we used for modeling) is 5.0 Å, which is not sufficient to provide position measurements when imaging HOPG, in which the spacing of the imaged atoms is less than one nanometer (2.46 Å). Furthermore, even with the use of sensors that have better resolution, we note that typically at normal room temperature, the measurement noise increases with scan frequency and results in poor resolution during high-speed STM operation. Thus, external sensor-based measurements of STM probe's lateral position (X, Y) are not available when imaging HOPG samples (with subnanoscale features) at high-scan frequencies. Therefore, we verify the ability to track the output trajectory by imaging standard surfaces (HOPG) and then quantifying the distortion in the acquired image. The simulation and experimental results are presented next.

B. Simulation Results

Optimal Redesign of Desired Output Trajectories—For the desired trajectory shown in Fig. 2(c), the optimal-inverse input for x-dynamics was found from (19); the weights $r(j\omega)$ and $q(j\omega)$ were as given in (28) (with $\rho = 1/400$). Similarly, the optimal-inverse input for the y-dynamics was computed [with the same weights in the cost function as in (28)] for the desired trajectory, as shown in Fig. 2(d). Simulations were performed

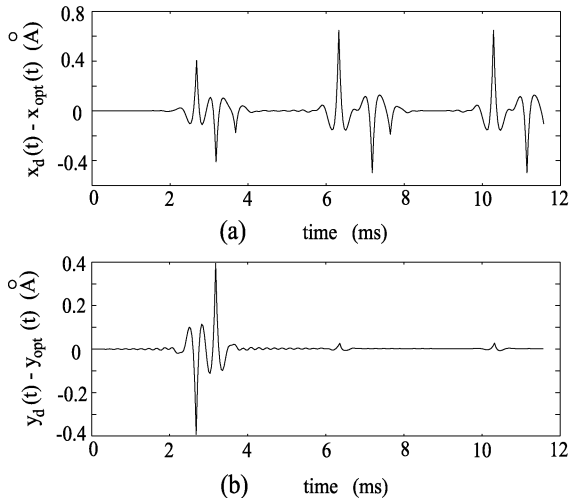


Fig. 5. Output trajectory difference between the desired trajectory and the optimal trajectory for scan rate $s_r = 250$ Hz in (a) x-direction and (b) y-direction.

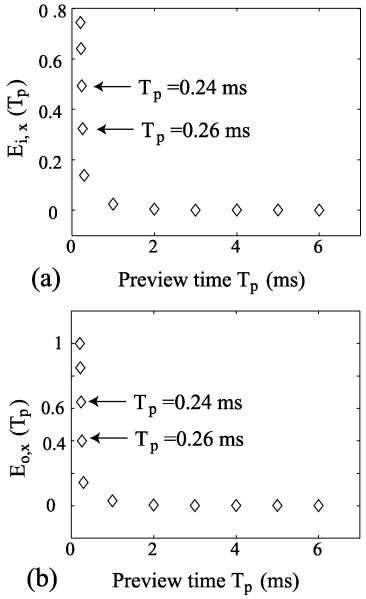


Fig. 6. Simulation results. (a) The relative input error $E_{i,x}(T_p)$ and (b) the relative output error $E_{o,x}(T_p)$ for different preview times T_p.

in MATLAB to find the optimal inputs $u_{\mathrm{opt},x}(t)$ and $u_{\mathrm{opt},y}(t)$ along with the optimal desired trajectories, $x_{\mathrm{opt}}(t)$ and $y_{\mathrm{opt}}(t)$ that were computed by applying the optimal inputs $u_{\mathrm{opt},x}(t)$ and $u_{\mathrm{opt},y}(t)$ to the dynamics models $G_x(s)$ and $G_y(s)$, respectively. We note that the differences between the optimal desired trajectories and the original desired trajectories are small (as shown in Fig. 5)

Preview-based Implementation—The optimal-inverse input was computed for different values of preview time. Simulations were performed to find the error in output tracking with the use of these preview-based optimal-inverse inputs. The error in computing the optimal-inverse input ($\hat{E}_{i,x}(T_p)$) and the error in output tracking due to the computational error ($\hat{E}_{o,x}(T_p)$) are shown in Fig. 6, where

$$\hat{E}_{i,x}(T_p) = \frac{\|u_{opt,x}(\cdot) - \tilde{u}_{opt,x}(\cdot)\|_\infty}{\|u_{opt,x}(\cdot)\|_\infty}$$
$$\hat{E}_{o,x}(T_p) = \frac{\|x_{\mathrm{opt}}(\cdot) - \tilde{x}_{\mathrm{opt}}(\cdot)\|_\infty}{\|x_{\mathrm{opt}}(\cdot)\|_\infty} \tag{30}$$

and for the signal $x(t)$

$$\|x(\cdot)\|_\infty = \sup_t \|x(t)\|.$$

The errors, in computing the optimal-inverse input and the resulting error in the output tracking, decay exponentially as the preview time increases, as shown in Fig. 6 [also see (25)]. Fig. 6 also indicates that preview information of the desired output is only needed for about 2–4 ms. Knowledge of the future desired output during this 2–4 ms time interval leads to substantial reduction of the error in the computation of the optimal inverse using the preview-based computation scheme. We note that Remark 12 gives a thumb rule to choose a large enough preview time based on the distance of the unstable poles of the optimal inverse filter from the imaginary axis. For this example, the unstable poles closest to the imaginary axis are at $p_{u,m} = -0.202 \pm 31.327\ \mathrm{rad/ms}$, therefore, the required preview time from Remark 12 is around $4/0.202 = 20$ ms. However, these poles are not the *dominant* unstable poles of the optimal inverse system; the output vector for the unstable subsystem of the optimal inverse filter [C_u in (19)] is almost perpendicular to the eigenvector associated with these two unstable poles $p_{u,m}$, i.e., $C_u \cdot V_p \approx 0$, where $A_u V_p = p_{u,m} V_p$. The next unstable poles closest to the imaginary axis are at $1.0488 \pm 24.645\ \mathrm{rad/ms}$, the required preview time from Remark 12 is $4/1.0488 = 3.81$ ms, which is close to the 4 ms preview time chosen for the experiments.

To illustrate that insufficient preview time leads to large output-tracking error, the optimal-inverse inputs were computed using three different preview times, $T_p = 4$ ms, $T_p = 0.26$ ms, and $T_p = 0.24$ ms. For each of the above preview times, the preview-based procedure given in Subsection II-C was used to find the optimal-inverse input [see (19)–(21)], which are compared in Fig. 7(a). The responses of the system for these inputs were found using simulations and are compared in Fig. 7(b). The error in computing the optimal-inverse input with a preview time $T_p = 4$ ms is very small (maximum error is 5×10^{-3} V). Therefore, the optimal-inverse input (dash–dotted line) and the preview-based optimal-inverse input for preview time $T_p = 4$ ms (solid line) overlap each other in Fig. 7(a); the maximum difference in the corresponding outputs is also small (less than 0.012 Å). Thus, the simulation results show that a 4 ms of preview time is sufficient to compute the optimal-inverse input without significant error in output tracking.

C. Experimental Results

The optimal-inverse inputs computed using different preview times were applied to the STM; the images were acquired one after the other using the same experiment setup. The image acquired without dynamics compensation in which the input was generated by scaling the desired trajectory with the dc gain of the STM model [20], is shown in Fig. 8(a). This approach that does not compensate for the STM dynamics is referred to as the dc-gain approach. Next, the image acquired with dynamics compensation using the optimal-inverse input is shown in Fig. 8(b).

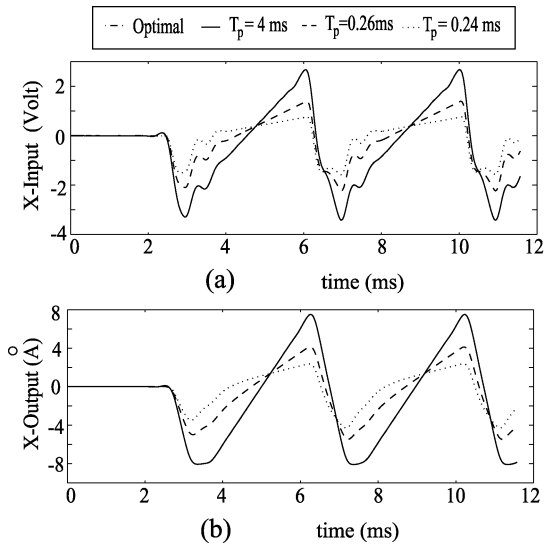


Fig. 7. Simulation results for $s_r = 250$ Hz. (a) Comparison of preview-based optimal-inverse inputs and (b) output trackings for different preview times.

Additionally, images acquired using finite-preview optimal-inverse inputs with preview times $T_p = 4$ ms, 0.26 ms and 0.24 ms are shown in Fig. 9(a), (b), and (c), respectively.

Image-based Quantification of Output-Tracking Error—In the following, we quantify the output-tracking error in terms of atomic-spacing deviations in the acquired images. We note that HOPG is a layered material with hexagonal lattice structures in each layer. However, not all atoms in the hexagonal lattice are imaged with the STM. Instead, only every other carbon atom in the hexagonal structure appears in the images, which results in an equidistant atom spacing of $d_{a,n} = 2.46$ Å between every adjacent atom in the acquired image (e.g., see [28, Ch. 26] for detailed discussion). Therefore, the tracking error in positioning of the STM probe over the sample surface can be quantified from the experiment images by measuring the distortions in the atom spacing. To quantify distortion in the STM-acquired images, the centers of each atoms were found using numerical image processing [29] [see Fig. 8(c)], and the distance μ_d between the adjacent atoms were computed from the centers. We obtained the mean value μ_d, the standard deviation σ_d and the maximum deviation $\delta_{\max}$ of the measured atom spacing $d_{a,i}$ of adjacent atoms from the nominal atom spacing of $d_a = 2.46$ Å as

$$\begin{aligned} \mu_d &\triangleq \frac{1}{N}\sum_{i=1}^{N} d_{a,i} \\ \sigma_d &\triangleq \sqrt{\sum_{i=1}^{N} \frac{(d_{a,i}-\mu_d)^2}{(N-1)}} \\ \delta_{\max} &\triangleq \max_i |d_{a,i} - d_a|. \end{aligned} \tag{31}$$

The statistical properties of atom spacing in the acquired images were computed for the experimental images acquired with different preview times and are compared in Table I.

TABLE I
COMPARISON OF ATOM-SPACING MEASURED FROM THE EXPERIMENTAL IMAGES AT SCAN RATE OF 250 HZ

Method	μ_d (Å)	σ_d (Å)	δ_{max} (Å)	δ_{max}/d_a
DC-Gain	2.66	0.52	2.66 [a]	108%
Optimal-Inverse	2.44	0.17	0.51	21%
$T_p = 4ms$	2.44	0.19	0.59	24%
$T_p = .26ms$	4.42	1.07	5.12	208%
$T_p = .24ms$	6.84	2.59	8.88	361%

D. Discussion

The experimental results show that a finite-preview can be used to compute optimal-inverse inputs that achieve precision STM probe positioning for imaging at high speeds. The use of the optimal inverse enables precision positioning of the STM probe at 250 Hz. For example, note that the distortion of atom spacing is small with the use of the optimal inverse, the mean value of the atom spacing by using optimal inverse at 250 Hz is 2.44 Å, which is close to the expected value of 2.46 Å (see Table I). It is noted that precision positioning of the STM probe at such a high scan rate cannot be achieved without accounting for the dynamics of this STM system, as shown in Fig. 9(a) and Table I (see [20] for a detailed study of high-speed STM imaging). The experimental results (Fig. 9(b) and Table I) also show that such precision output tracking can be achieved by using a finite preview of the desired output when computing the optimal-inverse input. Note that the image quality using finite preview of $T_p = 4$ ms [Fig. 9(a)] is very similar to the image quality using optimal inverse [Fig. 8(b)]. Furthermore, the atom spacing obtained by using preview time $T_p = 4$ ms is similar to the atom spacing obtained by using optimal inverse, as shown in Table I (compare the third line with the fourth line).

Insufficient preview can lead to substantial output-tracking error when using the optimal-inverse input. Note that significant input error and output-tracking error were found using simulation results with preview time of 0.26 ms and 0.24 ms (Fig. 7). These errors cause large image distortion in the experimental results, as shown in Fig. 9(b) and (c). Such large image distortions are also evident in the computed atom spacing in Table I: the mean error and the maximum error in atom spacing are significantly large when compared to the case in which the preview time T_p is 4 ms. It is also noted that the input error and output tracking error for $T_p = 0.24$ ms is substantially larger than those for $T_p = 0.26$ ms: $E_{i,x}(T_p)$ and $E_{o,x}(T_p)$ [defined in (30)] were increased from 32% and 40% to 49% and 64%, respectively, when the preview time is decreased from $T_p = 0.26$ ms to $T_p = 0.24$ ms (also see Figs. 6 and 7). Therefore, the corresponding image distortion is also significantly more severe for $T_p = 0.24$ ms than for $T_p = 0.26$ ms. This large difference due to a seemingly small increase of preview time can be explained by Fig. 6, which shows that the input error and output error (due to finite preview time) decay exponentially as the preview time increases [also see (25)]. Moreover, the input error and output error curves have a sharp drop for $T_p \leq 1$ ms. Therefore, within this region ($T_p \leq 1$ ms), a small increase of preview time (i.e., T_p from 0.24 to 0.26 ms) results in large decrease of input-computation error and output-tracking error. However, the exponen-

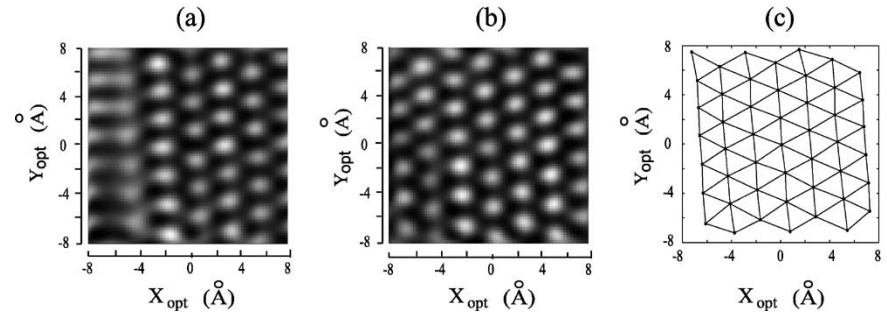


Fig. 8. Experimental STM images of graphite. (a) Image acquired without dynamics compensation and (b) with dynamics compensation using optimal inversion. The graph connecting atom centers from the image in (b) is shown in (c).

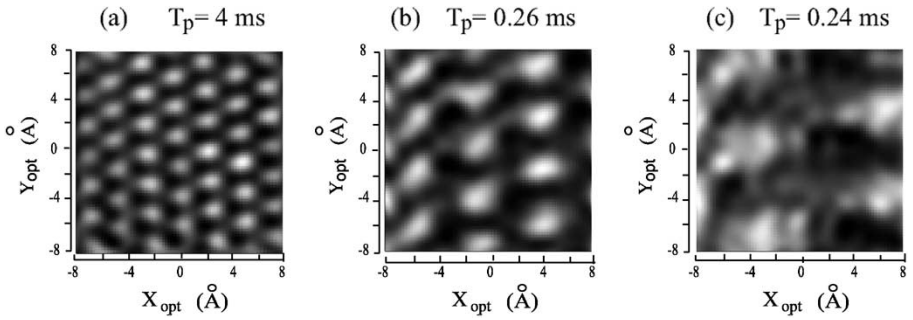


Fig. 9. Experimental results: the STM image of graphite acquired by using preview-based optimal-inverse inputs for preview time (a) $T_p = 4\ \mathrm{ms}$; (b) $T_p = 0.26\ \mathrm{ms}$; and (c) $T_p = 0.24\ \mathrm{ms}$.

tial decrease of the tracking error with the preview time implies that the tracking error can be made small by increasing the preview time; precision output tracking can be achieved by using optimal-inverse inputs computed with a finite preview time of 4ms for this experimental STM.

V. Conclusion

This paper showed that the noncausal optimal inverse can be computed using a preview-based approach to achieve precision output tracking. Properties of the optimal inversion were analyzed and the amount of preview time needed was quantified in terms of the required accuracy in output tracking, the poles and zeros of the system, and cost function used to develop the optimal inverse. The technique was illustrated by applying it to achieve high-speed nanoscale imaging with an STM. Experimental results were presented, which show that on-line precision output tracking can be achieved by using finite preview-based optimal inverse.

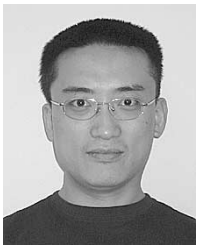
Acknowledgment

The authors would like to thank G. Clayton for his help with implementing imaging algorithms.

References

[1] L. Silverman, "Inversion of multivariable linear systems," *IEEE Trans. Automat. Contr.*, vol. AC–14, pp. 270–276, Mar. 1969.
[2] E. Bayo, "A finite-element approach to control the end-point motion of a single-link flexible robot," *J. Robot Syst.*, vol. 4, no. 1, pp. 63–75, 1987.
[3] S. Devasia, D. Chen, and B. Paden, "Nonlinear inversion-based output tracking," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 930–942, July 1996.
[4] D. Kwon and W. J. Book, "A time-domain inverse dynamic tracking control of a single-link flexible manipulator," *J. Dynamic Syst. Measurement Control*, vol. 116, no. 2, pp. 193–200, 1994.
[5] B. Paden, D. Chen, R. Ledesma, and E. Bayo, "Exponentially stable tracking control for multi-joint flexible manipulators," *ASME J. Dynamic Syst. Measurement Control*, vol. 115, no. 1, pp. 53–59, 1993.
[6] C. Tomlin, J. Lygeros, and S. Sastry, "Output tracking for a nonminimum phase dynamic CTOL aircraft model," in *Proc. IEEE Conf. Decision and Control*, New Orleans, LA, 1995, pp. 1867–1872.
[7] D. Croft, G. Shedd, and S. Devasia, "Creep, hysteresis, and vibration compensation for piezoactuators: atomic force microscopy application," *ASME J. Dynamic Syst. Measurement Control*, vol. 123, no. 35, pp. 35–43, Mar. 2001.
[8] J. S. Dewey, K. Leang, and S. Devasia, "Experimental and theoretical results in output-trajectory redesign for flexible structures," *ASME J. Dynamics Syst. Measurement Control*, vol. 120, pp. 456–461, Dec. 1998.
[9] S. Devasia, "Should model-based inverse input be used as feedforward under plant uncertainty?," *IEEE Trans. Automat. Contr.*, vol. 47, no. 11, pp. 1865–1871, Nov. 2002.
[10] Q. Zou and S. Devasia, "Preview-based stable-inversion for output tracking of linear systems," *ASME J. Dynamic Syst. Measurement Control*, vol. 121, no. 1, pp. 625–630, Dec. 1999.
[11] G. M. Whitesides and H. C. Love, "The art of building small," *Sci. Amer.*, vol. 285, no. 3, pp. 39–47, Sept. 2001.
[12] A. Daniele, S. Salapaka, M. V. Salapaka, and M. Dahleh, "Piezoelectric scanners for atomic force microscopes: design of lateral sensors, identification and control," in *Proc. American Control Conf.*, San Diego, CA, June 1999, pp. 253–257.
[13] J. A. Main and E. Garcia, "Piezoelectric stack actuators and control system design: strategies and pitfalls," *J. Guid. Control Dyn.*, vol. 20, no. 3, pp. 479–485, May–June 1997.
[14] D. Halim and S. O. R. Moheimani, "Experiments on spatial H_2 control for vibration suppression of a piezo-electric laminate beam ," in *Proc. IEEE Conf. Decision and Control*, Orlando, FL, Dec. 2001, pp. 3860–3865.
[15] N. Hidehiko and M. Tomizuka, "Setting control and performance of dual-actuator systems for hard disk drives," in *Proc. American Control Conf.*, Arlington, VA, June 2001, pp. 2779–2785.
[16] V. Gosavi, V. Shekhar, and A. G. Kelkar, "Passivity-based robust control of piezo-actuated flexible beam," in *Proc. American Control Conf.*, Arlington, VA, June 2001, pp. 2492–2497.

[17] P. Ge and M. Jouaneh, "Tracking control of a piezoceramic actuator," *IEEE Trans. Contr. Syst. Technol.*, vol. 4, pp. 209–216, May 1996.
[18] J. M. Cruz-Hernandez and V. Hayward, "Phase control approach to hysteresis reduction," *IEEE Trans. Contr. Syst. Technol.*, vol. 9, no. 1, pp. 17–26, Jan. 2001.
[19] R. C. Barrett and C. F. Quate, "Optical scan-correction system applied to automatic force microscopy," *Rev. Sci. Instrum.*, vol. 62, pp. 1393–1399, 1991.
[20] D. Croft and S. Devasia, "Vibration compensation for high speed scanning tunneling microscopy," *Rev. Sci. Instrum.*, vol. 70, no. 12, pp. 4600–4605, Dec. 1999.
[21] N. K. Gupta, "Frequency shaped cost functionals: extension of linear-quadratic-gaussian design methods," *J. Guid. Control*, vol. 3, no. 6, pp. 529–535, Nov. 1980.
[22] R. Brinkerhoff and S. Devasia, "Output tracking for actuator deficient/redundant systems: multiple piezoactuator example," *J. Guid. Control Dyn.*, vol. 23, no. 2, pp. 370–373, 2000.
[23] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.
[24] H. Kwakernaak and R. Sivan, "The maximally achievable accuracy of linear optimal regulators and linear optimal filters," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 79–85, Feb. 1972.
[25] L. Qiu and E. J. Davison, "Performance limitations of nonminimum phase systems in the servomechanism problem," *Automatica*, vol. 29, pp. 337–349, Mar. 1993.
[26] C. Z. Cai, X. Y. Chen, Q. Q. Shu, and X. L. Zheng, "Computer correction for distorted STM images," *Rev. Sci. Instrum.*, vol. 63, no. 12, pp. 5649–5652, Dec. 1993.
[27] K. G. Vandervoort, R. K. Zasadzinski, G. G. Galicia, and G. W. Crabtree, "Full temperature calibration from 4 to 300 k of the voltage response of piezoelectric tube scanner pzt-5a for use in scanning tunneling microscopes," *Rev. Sci. Instrum.*, vol. 64, no. 4, pp. 896–899, Apr. 1993.
[28] H. J. Guntherodt and R. Wiesendanger, Eds., *Scanning Tunneling Microscopy I*, 2nd ed. Heidelberg, Germany: Springer-Verlag, 1994, Surface Sci.
[29] I. Pitas, *Digital Image Processing Algorithms Applications*. New York: Wiley, 2000.
[30] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill, 1966.



Qingze Zou (S'02–M'03) received the B.S. degree in automatic control from the University of Electronic Science and Technology of China, Chengdu, China, in 1994, the M.S. degree in mechanical engineering from Tsinghua University, Beijing, China, in 1997, and the Ph.D. degree in mechanical engineering from the University of Washington, Seattle, WA, in 2003.

He is currently a Researh Associate at the University of Washington. His research interests include inversion-based control theory, precision positioning of piezo–actuator for nanotechnology, biomedical applications, and combinational chemistry applications.

Dr. Zou is a member of ASME .

Santosh Devasia (M'94) received the B.Tech. (Hons.) degree from the Indian Institute of Technology, Kharagpur, India, in 1988 and the M.S. and Ph.D. degrees in mechanical engineering from the University of California, Santa Barbara, CA in 1990 and 1993, respectively.

He has taught in the Mechanical Engineering Department, University of Utah, Salt Lake City. He is currently an Associate Professor in the Mechanical Engineering Department, University of Washington, Seattle. He is an Associate Editor for *ASME Journal of Dynamic Systems, Measurement and Control*. His current research interests include inversion-based control theory, high-precision positioning systems for nanotechnology and biomedical applications, and the control of complex distributed systems such as air traffic management (these projects are funded through NSF, NIH and NASA grants).

Dr. Devasia is a member of AIAA and ASME and is currently an Associate Editor of IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY.