INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2015; **00**:1–16
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/acs

Overview and New Results in Disturbance Observer based Adaptive Vibration Rejection With Application to Advanced Manufacturing

Xu Chen[1]*and Masayoshi Tomizuka[2]

[1]*Department of Mechanical Engineering, University of Connecticut, Storrs, CT, U.S.A.*
[2]*Department of Mechanical Engineering, University of California, Berkeley, CA, U.S.A.*

SUMMARY

Vibrations with unknown and/or time-varying frequencies significantly affect the achievable performance of control systems, particularly in precision engineering and manufacturing applications. This paper provides an overview of disturbance-observer based adaptive vibration rejection schemes; studies several new results in algorithm design; and discusses new applications in semiconductor manufacturing. We show the construction of inverse-model-based controller parameterization, and discuss its benefits in decoupled design, algorithm tuning, and parameter adaptation. Also studied are the formulation of recursive least squares and output-error based adaptation algorithms, as well as their corresponding scopes of applications. Experiments on a wafer scanner testbed in semiconductor manufacturing prove the effectiveness of the algorithm in high-precision motion control.

Received . . .

KEY WORDS: adaptive vibration rejection, narrow-band disturbances, advanced manufacturing, inverse-based control, semiconductor manufacturing, disturbance observer

1. INTRODUCTION

Identification and rejection of vibrations are of fundamental importance in science and engineering. Many problems—such as attenuation of repeatable and nonrepeatable runout disturbances in hard disk drives (HDDs) [1], rejecting fan noise from cooling systems [2, 3], and suspension control [4]—fall into the category of vibration rejection. Although many conventional problems have been relatively well solved, new and emerging challenges continuously occur. For instance, to meet the ever-increasing demand of storage volume in the era of big data, the disk drive industry is moving towards the new bit patterned media (BPM) recording. Each bit on BPM is stored in a fundamentally different principle, to reach a 20-30 times reduction in the unit storage size [5]. In the new architecture, vibration-related issues have become more important than ever before, with various new challenges such as significant increase of harmonics, much wider span of disturbance frequencies, etc [6].

Fundamentally, the importance of high-performance vibration rejection roots in the fact that practical systems, as long as they involve rotational or periodic motions, are inevitably subjected to periodic commands and/or disturbances. The problem of adaptive narrow-band disturbance[†] rejection has thus attracted great research attention, and recently been

*Correspondence to: Xu Chen, 191 Auditorium Road, Unit 3139, University of Connecticut, Storrs, CT, 06269-3139, U.S.A. E-mail: <xchen@engr.uconn.edu>
[†]Disturbances whose energies are concentrated within narrow frequency bands.

extensively studied in the multi-band situation [7]. Many control design algorithms have been studied in the related literature. Adaptive noise cancellation [8] uses sensors and stochastic-gradient-based adaptation to cancel the disturbance effect. Adaptive feedforward cancellation [9] composes an estimate of the sinusoidal disturbance using trigonometric functions. In feedback control, it has been understood that for asymptotic disturbance rejection, controllers should be customized to incorporate the model of the disturbance. This internal-model-principle approach has been essential for the achieved results in [4, 10, 11, 12, 13, 14, 2]. In the scope of direct adaptive controls, [12, 11] used state-space designs; [4, 10, 13, 14, 2] applied Youla Parameterization with adaptive finite-impulse-response (FIR) filters. Alternatively, indirect adaptive control can be formulated after online identification of the disturbance frequencies. Among the related literature on frequency identification, MUSIC and ESPRIT [15, 16] are two of the most commonly used subspace spectrum estimation methods. Both algorithms are batch processes that require an entire sequence (of sufficient length) of data to find the spectral peaks of the signals. Compared to Fast Fourier Transform (FFT), MUSIC and ESPRIT require less data points to generate accurate identification in noisy environments. For adaptive control, more common than the batch process are the adaptive notch filters and phase lock loops, where improved frequency identification is provided at each time step; and in addition, it is possible to track the change of the disturbance characteristics online. Algorithms of this kind include [17, 18, 19, 20, 21, 22, 23, 24], among which [18, 19, 20, 21, 23] are state-space constructions suitable for different computation complexities; [22] is a phase locked loop method that is commonly combined with adaptive feedforward cancellation; [17] is a popular adaptive notch filter based on recursive prediction error methods.

Recently, focusing on the concept of selective model inversion, the authors designed a class of adaptive narrow-band disturbance observers for spectral peak suppression. Successful applications have been achieved on HDDs in information storage systems [25, 26], electrical power steering [27], and the recent benchmark on active suspension [28, 29]. Comparison to algorithms of peer researchers has also been made in the benchmark summary report [7].

The purposes of this paper are twofold. First, we provide an overview and new discussions on different design options in the adaptive narrow-band disturbance observer. New results about the selection (and the corresponding designs) between FIR and infinite impulse response (IIR) filters are discussed. We study the use of an extended internal model principle for IIR-based disturbance rejection, and show the reduction of error amplification from the waterbed effect, which has often been theoretically overlooked previously. Also discussed are the parameter adaptation in non-ideal practical implementations, especially for situations where the number of disturbance components are unknown. Another contribution of the paper is the new implementation results in advanced manufacturing, where vibration and disturbance rejection have been key for next-generation technology development, due to the continuously updating precision requirement that is driven by Moore's law.

The remainder of the paper is organized as follows. Section 2 provides an overview of the deterministic control structure and discusses the design options. Section 3 discusses the parameter adaptation under different application environments. Experimental results on a wafer scanner in semiconductor manufacturing are provided in Section 4. Section 5 concludes the paper.

2. DETERMINISTIC SELECTIVE MODEL INVERSION

2.1. Disturbance cancellation in disturbance observers

The narrow-band disturbance observer is an inverse-based disturbance cancellation scheme as shown in Fig. 1. Independent from the outer-loop feedback controller $C(z^{-1})$, the thick red line builds an internal feedback of the disturbance $d(k)$.

DOI: 10.1002/acs

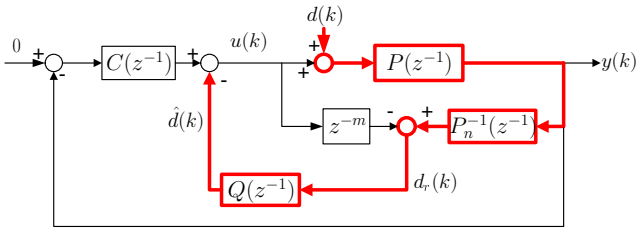


Figure 1. Disturbance estimation and feedback in disturbance observers: $P(z^{-1})$–the controlled plant, $C(z^{-1})$–feedback controller, $P_n^{-1}(z^{-1}) = z^{-m}\hat{P}^{-1}(z^{-1})$–nominal inverse of the plant, z^{-m}–relative degree of the plant, $Q(z^{-1})$–Q filter.

More specifically, focus first on the signal flow from $d(k)$ to $\hat{d}(k)$. If $P_n^{-1}(z^{-1})$ correctly inverts the plant dynamics $P(z^{-1})$, then ideally letting $Q(z^{-1}) = 1$ will create a internal negative feedback of $d\,(k)$, hence the cancellation of the disturbance at the input of the plant.

To analyze the flows of the control command $u(k)$, notice first, that practical systems inevitably contain uncertainties that are excluded in the transfer function model $P(z^{-1})$. In addition, $P^{-1}(z^{-1})$ is mostly acausal and not directly implementable. Hence $P_n^{-1}(z^{-1}) = P^{-1}(z^{-1})$ is practically unachievable. In narrow-band disturbance observer, $P_n^{-1}(z^{-1})$ is set as $z^{-m}\hat{P}^{-1}(z^{-1})$, where m is the relative degree of $P(z^{-1})$ for realizability of $P_n^{-1}(z^{-1})$, and $\hat{P}(z^{-1})$ is a nominal *stable* inverse that captures the main dynamics of $P(z^{-1})$. This way, $P_n^{-1}(z^{-1})P(z^{-1})$ approximately equals z^{-m}. Observing now the two internal paths from $u(k)$ to $Q(z^{-1})$—through $P_n^{-1}(z^{-1})P(z^{-1})$ and z^{-m}, respectively—one can see that $u(k)$ actually has null influence on $\hat{d}(k)$.

From the discussions in the last paragraph, one may observe that the disturbance observer is build on the idea of rejecting $d(k)$ without influencing the output of $C(z^{-1})$. $Q(z^{-1})$ can thus be designed independently from the feedback controller $C(z^{-1})$. This is the *decoupled design* principle in disturbance-observer-based feedback design.

Remark. Ultimately, the control goal is to eliminate the effect of disturbance from the plant output. Certainly, actual disturbances can impact the system internally and externally, from both the input and the output of the plant. Yet the control algorithm can only influence the command at the input side of the plant. In Fig. 1, we have lumped the disturbance at the plant input as a mathematical extraction. The algorithm also works if the disturbance enters via the more general Box-Jenkins model $y(k) = P(z^{-1})u(k) + M_d(z^{-1})d(k)$ (see [28, 29]), where $M_d(z^{-1})$ is a rational transfer function modeling the disturbance dynamics. In this case, the disturbance observer maintains its effect by estimating and canceling an *equivalent input disturbance*.

Although the inverse architecture provides rich information of $d(k)$ in the raw estimate $d_r(k)$, the signal-to-noise ratio in $d_r(k)$ is low due to uncertainty-induced model mismatch (common at high frequencies) and sensor noises. Conventionally, $Q(z^{-1})$ was chosen as a low-pass filter to avoid high-frequency noise amplifications. This is sufficient for low-frequency servo enhancement, yet infeasible for high-performance vibration rejection. To see this, notice that $P_n^{-1}P \approx z^{-m}$ and hence $d_r(k) \approx z^{-m}d(k)$ (z^{-1} as a delay operator here). Therefore $d_r(k)$ is actually a *delayed* estimate of $d(k)$. For vibration rejection, especially at high frequencies, the delay will introduce large mismatch between $d_r(k)$ and $d(k)$ (see one example in [26]). To cope with the difficulty, design of $Q(z^{-1})$ in narrow-band disturbance observer is based on the cancellation criterion:

$$d(k) - \hat{d}(k) = d(k) - Q(z^{-1})P_n^{-1}(z^{-1})P(z^{-1})d(k) \approx \left[1 - z^{-m}Q(z^{-1})\right] d(k). \tag{1}$$

Based on (1), for vibration signals satisfying

$$A(z^{-1})d(k) = 0 \text{ (asymptotically)}, \tag{2}$$

DOI: 10.1002/acs

asymptotic perfect disturbance rejection is achieved if the feedback loop is constructed to contain the internal model:

$$1 - z^{-m}Q(z^{-1}) = K_c(z^{-1})A(z^{-1}) \tag{3}$$

where $K_c(z^{-1})$ is a realizability filter for $Q(z^{-1})$ to be causal.

2.2. Extended internal model principle

The most intuitive internal-model based approach is perhaps to design from the time domain. For instance, consider the case where $m = 1$ and $A(z^{-1}) = 1 - 2\cos\omega_o z^{-1} + z^{-2}$, i.e. $d(k)$ is a single-frequency vibration in the form of $C_o \sin(\omega_o k + \zeta_o)$. Letting $K_c(z^{-1}) = 1$ and $1 - z^{-1}Q(z^{-1}) = 1 - 2\cos\omega_o z^{-1} + z^{-2}$ yields the first-order FIR solution of (3):

$$Q\left(z^{-1}\right) = 2\cos\omega_o - z^{-1}. \tag{4}$$

Although satisfying (3), the FIR solution will be seen to be not practical in general. The narrow-band disturbance observer assigns instead

$$1 - z^{-1}Q(z^{-1}) = \frac{1 - 2\cos\omega_o z^{-1} + z^{-2}}{1 - 2\alpha\cos\omega_o z^{-1} + \alpha^2 z^{-2}} \tag{5}$$

and for more general values of m,

$$1 - z^{-m}Q(z^{-1}) = K(z^{-1})\frac{A(z^{-1})}{A(\alpha z^{-1})} \tag{6}$$

where $A(\alpha z^{-1})$ $[\alpha \in (0, 1)]$ is a damped polynomial of z^{-1}, obtained by replacing every z^{-1} in $A(z^{-1})$ with αz^{-1}. Based on (2) and (3), the disturbance is still asymptotically rejected under the extended internal model $A(z^{-1})/A(\alpha z^{-1})$. The new solution to the example at the beginning of this subsection is

$$Q(z^{-1}) = \frac{-2\left(\alpha - 1\right)\cos\omega_o + \left(\alpha^2 - 1\right)z^{-1}}{1 - 2\alpha\cos\omega_o z^{-1} + \alpha^2 z^{-2}}. \tag{7}$$

The choice of an IIR design (7) over the FIR solution (4) is based on the frequency-domain properties of the filters. As shown in the example in Fig. 2, the damped IIR design provides a bandpass filter characteristics. As the magnitude of $A(z^{-1})$ is normalized by $A(\alpha z^{-1})$, the frequency response of $A(z^{-1})/A(\alpha z^{-1})$ is approximately unity at low and high frequencies. Hence from (5), the magnitude of $Q(z^{-1})$ must be close to zero at the corresponding frequency regions. On the contrary, the FIR design has no default constraints on the filter magnitude, and gives a scaled high-pass filter response—significantly amplifying the add-on noises in $d_r(k)$. The involved magnitude constraint in the IIR design provides the fundamental control of the waterbed effect in narrow-band disturbance observers.

Indeed, the FIR filter in (4) satisfies $Q\left(e^{-j\omega}\right)\big|_{\omega=0} = 2\cos\omega_o - 1$ (DC gain), $Q\left(e^{-j\omega}\right)\big|_{\omega=\pi} = 2\cos\omega_o + 1$ (gain at Nyquist frequency). Regardless of the value of ω_o, $|Q\left(1\right) - Q\left(-1\right)|$ equals 2, namely, it is not possible to keep the gains at DC and Nyquist frequency small at the same time. For disturbances below one half of Nyquist frequency (typical in practice), $\cos\omega_o$ is positive and $Q\left(-1\right)$ is thus larger than one, hence a high-pass $Q(z^{-1})$.

The bandpass Q-filter structure forms the idea of selective model inversion: only the plant dynamics at the disturbance frequencies is actually inverted. At frequencies outside of the passband of $Q(z^{-1})$, the effect of the inverse structure is neutralized by the small gains of $Q(z^{-1})$. Thus, the previously discussed plant uncertainties is acceptable in practice, as long as the magnitude of $Q(z^{-1})$ is sufficiently small at the corresponding frequencies. This selective model inversion principle is central for high performance of the algorithm in practice.

DOI: 10.1002/acs

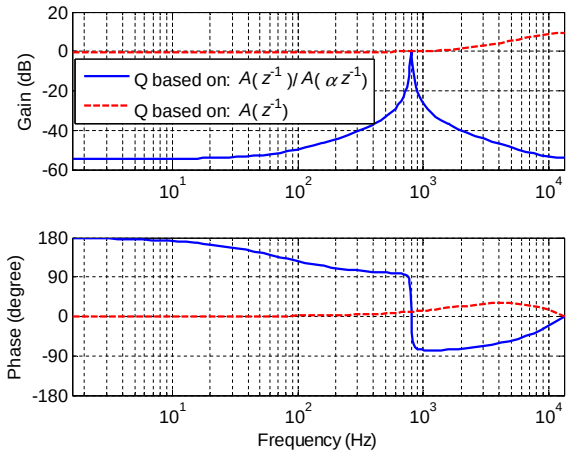


Figure 2. Example frequency responses of FIR and IIR Q-filter designs (design parameters: $m = 1$, sampling time $T_s = 1/26400$ sec, vibration frequency $\omega_o = 2\pi T_s \times 800$ rad, $\alpha = 0.998$).

For general narrow-band vibrations in the form of

$$d(k) = \sum_{i=1}^{n} C_i \sin(\omega_i k + \psi_i) + n_o(k), \tag{8}$$

[‡]the internal model is extended to

$$A\left(z^{-1}\right) = \prod_{i=1}^{n} \left(1 - 2\cos(\omega_i) z^{-1} + z^{-2}\right) \tag{9}$$

$$= 1 + a_1 z^{-1} + a_2 z^{-2} \cdots + a_n z^{-n} + \cdots + a_2 z^{-2n+2} + a_1 z^{-2n+1} + z^{-2n}. \tag{10}$$

Equations (9) and (10) are respectively the (nonlinear) cascaded and the (linear) direct forms of the polynomial $A\left(z^{-1}\right)$.

Depending on the design of $K(z^{-1})$ in (6), $Q(z^{-1})$ takes two forms of structures.

Direct Q design approach: Letting $K\left(z^{-1}\right)$ be an FIR filter $K\left(z^{-1}\right) = k_o + k_1 z^{-1} + \cdots + k_{n_K} z^{-n_K}$ and selecting $Q\left(z^{-1}\right) = B_Q\left(z^{-1}\right)/A\left(\alpha z^{-1}\right)$ in (6) yield

$$K\left(z^{-1}\right) A\left(z^{-1}\right) + z^{-m} B_Q\left(z^{-1}\right) = A\left(\alpha z^{-1}\right). \tag{11}$$

Matching coefficients of z^{-i} on each side of (11), we can obtain $K\left(z^{-1}\right)$ and $B_Q\left(z^{-1}\right)$.

More generally, (11) is in the form of a polynomial Diophantine equation. Solutions exist as long as the greatest common factor of $A(z^{-1})$ and z^{-m} divides $A(\alpha z^{-1})$, which is always true.

Example 1. Consider the case of $m = 2$, $n = 1$, and $A\left(z^{-1}\right) = 1 - 2\cos\omega_o z^{-1} + z^{-2}$. Let $a = -2\cos\omega_o$. We have $A\left(\alpha z^{-1}\right) = 1 + \alpha a z^{-1} + \alpha^2 z^{-2}$. The Diophantine equation is then $K\left(z^{-1}\right)\left(1 + a z^{-1} + z^{-2}\right) + z^{-2} B_Q\left(z^{-1}\right) = 1 + \alpha a z^{-1} + \alpha^2 z^{-2}$. Letting $K\left(z^{-1}\right) = k_0 + k_1 z^{-1}$ gives $z^{-2} B_Q\left(z^{-1}\right) = 1 - k_0 + (\alpha a - a k_0 - k_1) z^{-1} + \left(\alpha^2 - k_0 - a k_1\right) z^{-2} - k_1 z^{-3}$. To have a realizable $B_Q\left(z^{-1}\right)$, we need $1 - k_0 = 0$ and $\alpha a - a k_0 - k_1 = 0$. Solving the two equations for k_0 and k_1 gives (after simplifications)

$$Q\left(z^{-1}\right) = \frac{\left(\alpha^2 - 1 - a^2(\alpha - 1)\right) - (\alpha - 1) a z^{-1}}{1 + \alpha a z^{-1} + \alpha^2 z^{-2}}. \tag{12}$$

[‡]Here $n_o(k)$ is the additive noise; $\omega_i = 2\pi\Omega_i T_s$ is the frequency in rad (Ω_i is the frequency in Hz, T_s is the sampling time); $C_i\,(\neq 0)$ and ψ_i are respectively the unknown magnitude and phase of each sinusoidal component.

DOI: 10.1002/acs

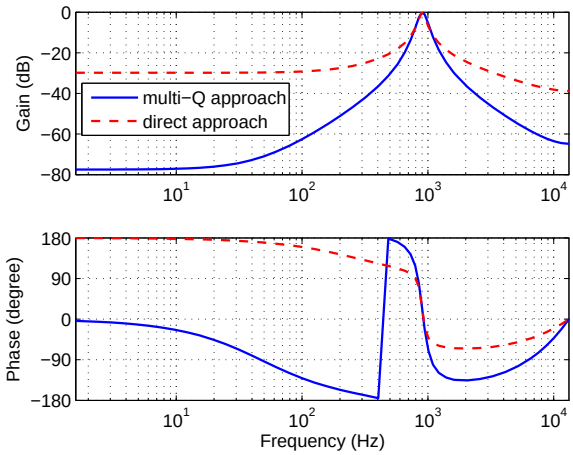


Figure 3. Internal-model based Q-design example: frequency response of $Q\left(z^{-1}\right)$.

Multi-Q approach: For the case of IIR design on $K(z^{-1})$, in [28, 29] we have obtained the solution:

$$Q(z^{-1}) = \left[\frac{\sum_{i=1}^{2n}(\alpha^i - 1)a_i z^{-i+1}}{A(\alpha z^{-1})}\right]^m \; ; \; a_i = a_{2n-i}, \; a_{2n} = 1 \tag{13}$$

$$1 - z^{-m}Q(z^{-1}) = \frac{A(z^{-1})}{A(\alpha z^{-1})}\sum_{i=1}^{m}\left(\begin{array}{c} m \\ i \end{array}\right)\left[-\frac{A(z^{-1})}{A(\alpha z^{-1})}\right]^{i-1} \; ; \; \left(\begin{array}{c} m \\ i \end{array}\right) = \frac{m!}{i!(m-i)!}. \tag{14}$$

where $\{a_i\}_{i=1}^n$ is as defined in (10).

Equation (13) is a cascaded version of a baseline filter in the square bracket, hence the name of the solution approach. Recall that m is the delay in the disturbance estimate. Intuitively, instead of solving the Diophantine equation, the multi-Q approach cascades m standardized sub filters in (13). Each sub filter is a special bandpass filter, and compensates one step of delay,[§] contributing to a total of m-step phase advance for perfect disturbance cancellation.

2.3. Comparison of different Q designs

To see the similarity and differences between the two design approaches, we provide now a case study assuming that: the sampling time $T_s = 1/26400$ sec; the plant delay $m = 2$; the number of vibration components $n = 1$; and the disturbance frequency $\Omega = 900$ Hz.

From Example 1, the Q filter in the direct approach is given by (12). Letting $\alpha = 0.9882$, we obtain a $Q\left(z^{-1}\right)$ whose frequency response is shown in the dashed line in Fig. 3. Correspondingly in Fig. 4, the dashed line provides the magnitude response of $1 - z^{-m}Q\left(z^{-1}\right)$. Recalling that $1 - z^{-m}Q(z^{-1})$ characterizes the dynamics of disturbance rejection in (1), we observe the strong attenuation (zero gain) at 900 Hz, with a stop-band width that is about 100 Hz. For the multi-Q approach, we use (13) to get $Q\left(z^{-1}\right) = \left[Q_{m=1}\left(z^{-1}\right)\right]^m$ where $Q_{m=1}\left(z^{-1}\right)$ is from (7). Letting $\alpha = 0.9765$ yields the solid lines in Figs. 3 and 4.

Both approaches give a desired notch shape of $1 - z^{-m}Q\left(z^{-1}\right)$ in Fig. 4, so that disturbance attenuation is specifically focused at 900 Hz, and the waterbed effect is well controlled. In the phase response in Fig. 3, both filters have non-zero phases at 900 Hz—the center frequency of the pass band. This is for compensating the delay effect z^{-m}. The multi-Q approach provides smaller magnitude response outside the pass band of $Q\left(z^{-1}\right)$, which implies stronger filtering

[§]Indeed, if $m = 1$ and $n = 1$, the multi-Q solution simplifies to (7).

DOI: 10.1002/acs

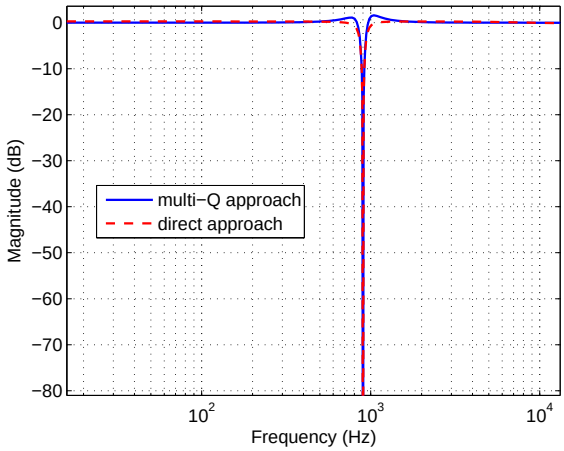


Figure 4. Internal-model based Q-design example: magnitude response of $1 - z^{-m}Q\left(z^{-1}\right)$.

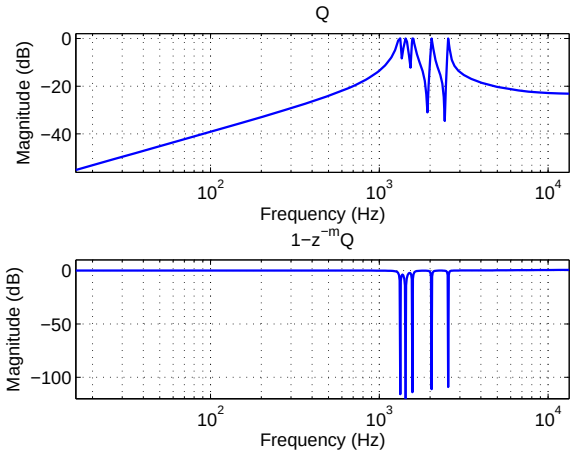


Figure 5. Multiple narrow-band Q-design example.

of noises and model mismatches by $Q\left(z^{-1}\right)$. Hence, under the same bandwidth of disturbance attenuation, the multi-Q approach is less sensitive to plant uncertainties than the direct solution.

Remark (Choosing α in practice). In the example, we designed the two Q filters to approximately have the same pass band in Fig. 3. The width of the pass band is controlled by the parameter $\alpha\,[\in (0,1)]$. A smaller α gives a wider pass band. For narrow-band disturbance rejection, α is usually selected to be close to one. Note that to achieve the same pass band in the example above, a smaller α is used in the multi-Q approach. This is because the multi-Q solution consists of a set of cascaded bandpass filters and α directly controls the pass band of the *individual* sub filters.

The design naturally extends to the case for multiple narrow-band disturbance rejection. Fig. 5 shows the use of the direct-Q approach to attenuate five narrow-band disturbances based on (11). Such complex bandpass $Q\left(z^{-1}\right)$ is challenging to obtain using conventional approaches, especially when the frequencies of different bands are very close to each other.

Finally, examining Fig. 6—an enlarged version of Fig. 4—we see that $1 - z^{-m}Q\left(z^{-1}\right)$ has slightly different behaviors outside the narrow band at 900 Hz: in the multi-Q approach, the magnitude is almost strictly 1 (0 dB) for the majority of the frequency region, and slightly

DOI: 10.1002/acs

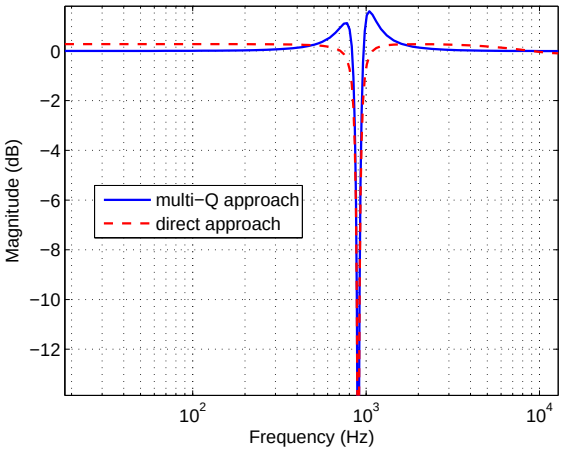


Figure 6. Q-design example: magnitude response of $1 - z^{-m}Q\,(z)$ (enlarged view).

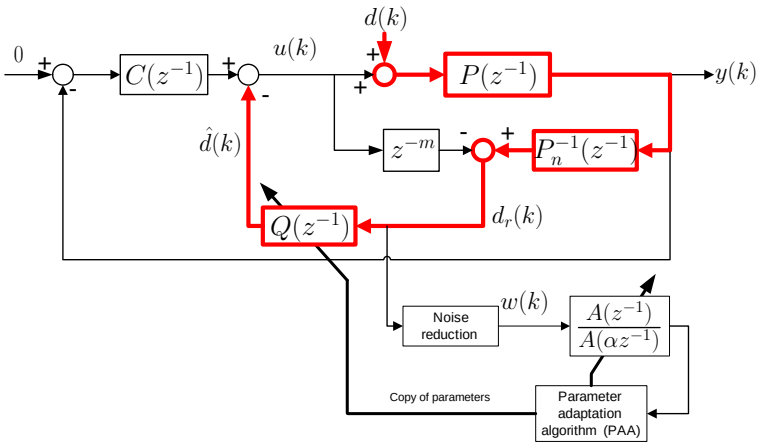


Figure 7. Structure of adaptive narrow-band disturbance observer.

more amplified near 800 and 1000 Hz [about 1.185 (1.5dB) maximum]; in the direct approach, $1 - z^{-m}Q\left(z^{-1}\right)$ has flatter magnitude response but slightly less robustness against noise and uncertainties (recall Fig. 3). For narrow-band loop shaping, the impact of the difference is quite small. When the desired attenuation region is wider, designers will need to make engineering judgments between the two Q designs for best control of the waterbed effect.

3. PARAMETER ADAPTATION

Knowledge of the disturbance frequencies is needed in the disturbance model $A(z^{-1})$. This section studies the adaptive formulation when such information is not *a priori* available.

The overall adaptive narrow-band disturbance observer scheme is shown in Fig. 7. Recall that the input to the Q filter is a delayed and noisy estimate of the actual disturbance. We can use this $d_r(k)$ signal for parameter adaptation on $Q\left(z^{-1}\right)$. To see the reason of the noise-reduction block, one can derive the output dynamics from the block diagram:

$$y(k) = \frac{P(z^{-1})}{1 + P(z^{-1})C(z^{-1})}\left[d(k) - \hat{d}(k)\right]$$

DOI: 10.1002/acs

which yields, after substituting in (1),

$$y(k) \approx \frac{\left[1 - z^{-m}Q(z^{-1})\right] P(z^{-1})}{1 + P(z^{-1})C(z^{-1})} d(k) \triangleq \left[1 - z^{-m}Q(z^{-1})\right] y_o(k).$$

Here $y_o(k) = \{P(z^{-1})/[1 + P(z^{-1})C(z^{-1})]\}d(k)$ is the baseline output without the disturbance observer. Using $P\left(z^{-1}\right) / \left[1 + P\left(z^{-1}\right) C\left(z^{-1}\right)\right]$ (or a nominal version of it) as the noise-reduction block makes $w\left(k\right) \approx y_o\left(k\right)$ in Fig. 7. The parameter adaptation algorithm is thus directly based on $y_o\left(k\right)$, which reflects the actual effect of the disturbance on the output. By tuning the coefficients of $Q(z^{-1})$, the adaptation aims to directly minimize the system output under vibration.

Remark. Of course, if prior knowledge about a coarse region of the disturbances is available, additional filtering—using e.g., standard bandpass filters from the filter design toolbox in MATLAB—should be applied in the noise-reduction block. This will further improve the signal-to-noise ratio during adaptation.

In the case of $m = 1$, $K\left(z^{-1}\right)$ equals unity for both designs of the Q filter [recall (5)]. Hence we have $1 - z^{-1}Q(z^{-1}) = A(z^{-1})/A(\alpha z^{-1})$. Adaptation can be directly constructed based on

$$v\left(k\right) = \frac{A\left(z^{-1}\right)}{A\left(\alpha z^{-1}\right)} w\left(k\right). \tag{15}$$

Note that when $w(k)$ is a pure sinusoidal signal, the output $v(k)$ is zero for tuned value of the filter. Thus, the goal of parameter adaptation is to drive the adaptive prediction of $v(k)$ toward zero.¶

Equation (15) turns out to be also applicable for the cases when $m > 1$, as $1 - z^{-m}Q(z^{-1})$ in (6) always contains the cascaded factor $A(z^{-1})/A(\alpha z^{-1})$.

Under the direct-filter form (10), (15) has the realization

$$v\left(k\right) = \psi\left(k-1\right)^T \theta + w\left(k\right) + w\left(k-2n\right) - \alpha^{2n} v\left(k-2n\right) \tag{16}$$

where the parameter vector θ and the regressor vector $\psi\left(k-1\right)$ are

$$\theta = \left[a_1, a_2, \ldots, a_n\right]^T \tag{17}$$

$$\psi\left(k-1\right) = \left[\psi_1\left(k-1\right), \psi_2\left(k-1\right), \ldots, \psi_n\left(k-1\right)\right]^T \tag{18}$$

$$\psi_i\left(k-1\right) = w\left(k-i\right) + w\left(k-2n+i\right) - \alpha^i v\left(k-i\right) - \alpha^{2n-i} v\left(k-2n+i\right);\ i = 1, ..., n-1 \tag{19}$$

$$\psi_n\left(k-1\right) = w\left(k-n\right) - \alpha^n v\left(k-n\right). \tag{20}$$

The linear filter form is suitable for various parameter adaptation algorithms for different application environments. Recall that $v(k)$ will be null if $w(k)$ is composed of pure sinusoidal signals and $A\left(z^{-1}\right)/A\left(\alpha z^{-1}\right)$ has the correct frequency parameters. Based on (16), the simplest recursive least squares (RLS) PAA can then be constructed to find

$$\hat{\theta}\left(k\right) = \arg\min_{\theta^*} \sum_{i=1}^{k} \left\{ \left(\prod_{j=i}^{k-1} \lambda\left(j\right) \right) \left[\psi\left(i-1\right)^T \theta^* + w\left(i\right) + w\left(i-2n\right) \right]^2 \right\} \tag{21}$$

where $\lambda(k)$ is a forgetting factor—close or equal to 1—that puts different weightings on the quadratic terms at different time steps. In designing the cost function, we used the result that $v\left(k\right)$ equals zero under the pure sinusoidal disturbance assumption. This observation also simplifies the regressor vectors in (19) and (20), by annihilating all the terms related to $v(k)$.

¶Absolute zero convergence is achievable only under ideal sinusoidal assumptions.

DOI: 10.1002/acs

In other words, ψ_i simplifies to

$$\psi_i(k-1) = w(k-i) + w(k-2n+i);\ i = 1, \ldots, n-1$$
$$\psi_n(k-1) = w(k-n).$$

The cost function (21) is quadratic, and actually convex, in θ^*. An analytic solution can be found for the global minimum. The iterative solution is

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{F(k-1)\psi(k-1)e^o(k)}{1+\psi(k-1)^T F(k-1)\psi(k-1)} \tag{22}$$

$$e^o(k) = -\psi(k-1)^T \hat{\theta}(k-1) - (w(k) + w(k-2n)) \tag{23}$$

$$F(k) = \frac{1}{\lambda(k)}\left[F(k-1) - \frac{F(k-1)\psi(k-1)\psi(k-1)^T F(k-1)}{\lambda(k)+\psi(k-1)^T F(k-1)\psi(k-1)}\right]. \tag{24}$$

The RLS algorithm has the advantages of simplicity and guaranteed stability, yet also relies heavily on the noise-free assumption on $w(k)$. In practice the RLS PAA provides accurate parameter convergence only when $w(k)$ has a high signal-to-noise ratio. A more accurate algorithm is to construct adaptation based on the output error

$$e(k) = \frac{\hat{A}\left(\hat{\theta}, z^{-1}\right)}{\hat{A}\left(\hat{\theta}, \alpha z^{-1}\right)} w(k). \tag{25}$$

Here, $e(k)$ is a residual error signal that we aim to minimize. In the associated difference equation,

$$e(k) = \phi(k-1)^T \hat{\theta}(k) + w(k) + w(k-2n) - \alpha^{2n} e(k-2n) \tag{26}$$

$$\phi_i(k-1) = w(k-i) + w(k-2n+i) \tag{27}$$
$$- \alpha^i e(k-i) - \alpha^{2n-i} e(k-2n+i);\ i = 1, ..., n-1$$

$$\phi_n(k-1) = w(k-n) - \alpha^n e(k-n), \tag{28}$$

all terms about $e(k-i)$ are no longer treated as 0. Constructing PAAs based on such output errors provides more accurate parameter convergence, but requires more signal processing to guarantee stability, as the adaptation changes poles of the IIR transfer function in (25). Based on extensive simulation and experiments in [28, 25], the output-error PAA (aka parallel PAA) with a fixed compensator is one algorithm that provides accurate convergence and computational efficiency. The main equations are:

- *a priori* estimation error

$$e^o(k) = -\left(-\phi(k-1)\right)^T \hat{\theta}(k-1) + w(k) + w(k-2n) - \alpha^{2n} e(k-2n) \tag{29}$$

  where $-\phi(k-1)$ indicates the negative of the gradient direction.
- *a posteriori* adaptation error

$$\epsilon(k) = C\left(z^{-1}\right) e(k) = e(k) + \alpha^{2n} e(k-2n) + \varphi(k-1)^T \theta_c \tag{30}$$

$$\theta_c = [c_1, c_2, \ldots, c_n]^T \tag{31}$$

$$\varphi_i(k-1) = \alpha^i e(k-i) + \alpha^{2n-i} e(k-2n+i);\ i = 1, ..., n-1 \tag{32}$$

$$\varphi_n(k-1) = \alpha^n e(k-n) \tag{33}$$

  where $C(z^{-1})$ and θ_c are, respectively, estimates of $A(\alpha z^{-1})$ and θ.
- *a priori* adaptation error

$$\epsilon^o(k) = e^o(k) + \alpha^{2n} e(k-2n) + \varphi(k-1)^T \theta_c \tag{34}$$

DOI: 10.1002/acs

- PAA

$$\hat{\theta}(k) = \hat{\theta}(k-1) + \frac{F(k-1)\left(-\phi(k-1)\right)\epsilon^o(k)}{1+\phi(k-1)^T F(k-1)\phi(k-1)} \tag{35}$$

$$F(k) = \frac{1}{\lambda(k)}\left[F(k-1) - \frac{F(k-1)\phi(k-1)\phi^T(k-1)F(k-1)}{\lambda(k)+\phi^T(k-1)F(k-1)\phi(k-1)}\right] \tag{36}$$

We focus next on several important aspects about practical implementations of the PAA algorithm. Above all, the two types of PAAs can be combined, to take advantage of the simplicity and global convergence of RLS, as well as the accuracy of the output-error method under noisy environments. This is central for successful applications when a good idea of the disturbance characteristics is unavailable in practice. The basic principle is to run the RLS PAA as an initialization of the output-error PAA. After convergence of RLS, the identified $\hat{a}_i$ can be assigned to the coefficients c_i in the fixed compensator. Some engineering tuning has to be performed to confirm convergence of RLS. A prediction-error-based auto-switching algorithm is provided in [25].

Additional implementation aspects during adaptation are discussed next.

Online Adaptation and Offline Monitoring: The frequencies and the identified parameters are mapped by

$$\prod_{i=1}^{n}\left(1-2\cos(\omega_i)z^{-1}+z^{-2}\right) = 1+a_1z^{-1}+\cdots+a_nz^{-n}+\cdots+a_1z^{-2n+1}+z^{-2n}. \tag{37}$$

Although a_i is the actual parameter updated online, knowing the frequencies of the vibrations is often useful for understanding the problem and for algorithm tuning. For the simplest case where $n=1$, we have $a_1 = -2\cos\omega_1$, from which we can compute $\omega_1 = 2\pi\Omega_1 T_s$. The parameter a_1 is online updated and Ω_1 can be computed offline for algorithm tuning.

For $n>1$, as $1-2\cos(\omega_i)z^{-1}+z^{-2} = \left(1-e^{j\omega_i}z^{-1}\right)\left(1-e^{-j\omega_i}z^{-1}\right)$ in (37), the values of ω_i can be computed offline via calculating the angle of the complex roots of $1+a_1z^{-1}+\cdots+a_nz^{-n}+\cdots+a_1z^{-2n+1}+z^{-2n}=0$.

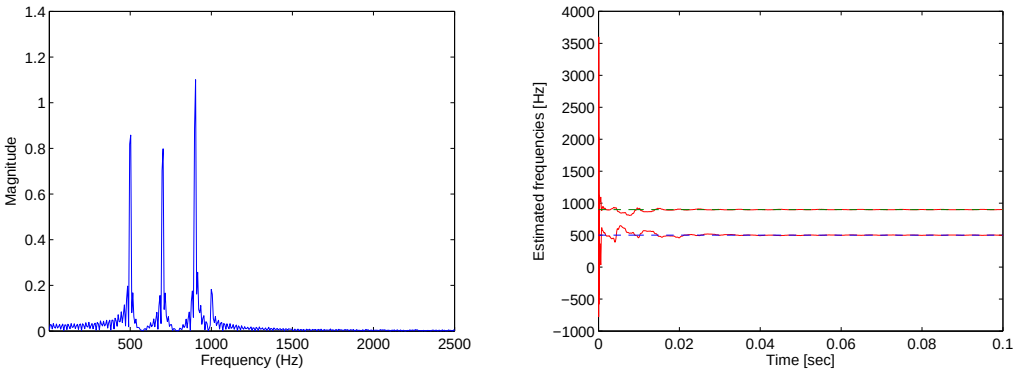
Parameter Initialization: If the output-error PAA is run without the RLS initialization, proper assignments should be placed on $\hat{\theta}(0)$ and c_i's. One can first estimate the disturbance frequencies $\{\hat{\omega}_i(0)\}_{i=1}^n$ based on physics or existing data, and then expand the product on the left side of (37), to obtain $\{\hat{a}_i(0)\}_{i=1}^n$ for implementation in the PAA.

Underdetermined and Overdetermined Adaptation: We have been assuming that the number of the adaptation parameters in $\{\theta_i\}_{i=1}^n$ is the same as the number of frequency components in the actual disturbance signal. If there are actually more than n, say $r\,(>n)$, narrow-band signals in $w(k)$, the parameters in the output-error PAA will converge to a local optimal point,[‖] as demonstrated in the example in Fig. 8. Here there are four independent vibration components between 500 Hz and 1000 Hz. The adaptation was constructed using only $n=2$ in the PAA. From Fig. 8b, the PAA correctly found the two main components at 500 Hz and 900 Hz.

If, on the contrary, $r<n$ (which is rare in practice), we will still be able to identify r frequency components in $w(k)$. Notice however, that in this case the adaptation model is

[‖]The location of the local optima depends on the initialization of the parameter estimate.

DOI: 10.1002/acs



(a) Spectrum of actual disturbance (b) Identified frequencies ($n = 2$)

Figure 8. Narrow-band disturbance identification example: adaptation input has more frequency components than the adaptation model.

over-determined and the optimum parameter estimate will not be unique. We will have

$$1 + a_1\alpha z^{-1} + \cdots + a_n\alpha^n z^{-n} + \cdots + a_1\alpha^{2n-1}z^{-2n+1} + \alpha^{2n}z^{-2n} = \prod_{i=1}^{r}\left(1 - 2\alpha\cos(\omega_i)z^{-1} + \alpha^2 z^{-2}\right) \times \prod_{i=r+1}^{n}\left(1 - 2\alpha\cos(\omega_i)z^{-1} + \alpha^2 z^{-2}\right). \quad (38)$$

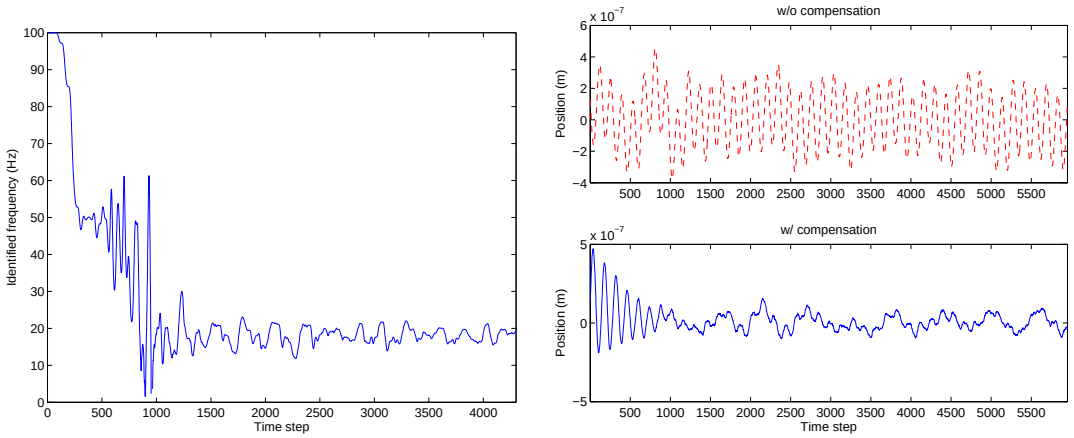
The PAA will conduct the adaptation to form a factor of $\prod_{i=1}^{r}\left(1 - 2\alpha\cos(\omega_i)z^{-1} + \alpha^2 z^{-2}\right)$ after convergence, and ω_i's in the last $n - r$ terms on the right side of (38) will converge to some region around the minimum.

4. EXPERIMENTAL RESULTS ON A WAFER SCANNER

Wafer scanners are key components for motion control in the photolithography process in semiconductor manufacturing. Billions of transistors need to be printed on the integrated circuit in silicon wafers. Wafer scanners are hence very sensitive to vibration disturbances. This section provides implementation and evaluation of the narrow-band disturbance observer for vibration rejection on a laboratory wafer scanner testbed. The testbed is controlled via LabVIEW programming language; equipped with air bearings; and uses DC linear permanent magnet motors for precision motion. Additional hardware description and system modeling have been described in [30]. All results in this section are from experiments on the actual system.

Fig. 9 shows the rejection of external vibrations. Such disturbances come from the operation environments such as cooling systems, adjacent machines, etc. On the testbed, an 18Hz vibration presents as shown in Fig. 9b. There are also other low-frequency disturbances after a spectrum analysis in Fig. 10. RLS alone thus has limited achievable performance here. The adaptive direct-Q filter based disturbance observer was implemented for improved precision. From Fig. 9, the vibration component was successfully identified online. After a transient period of about 1000 time steps (1 time step = 0.4 ms), the converged PAA provides a 72.8% reduction (from 1.764×10^{-7} m to 4.800×10^{-8} m) of the 3σ (σ is the standard deviation) value of the position errors. The strong disturbance rejection can also be observed from the frequency spectra in Fig. 10 (there is a small residual peak at 18 Hz, from the transient response).

DOI: 10.1002/acs



(a) Calculated frequency adaptation (b) Online time trace of position error

Figure 9. Rejection of environmental disturbance.

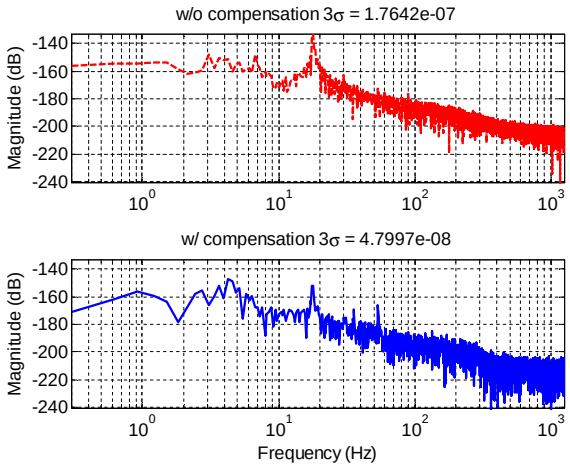


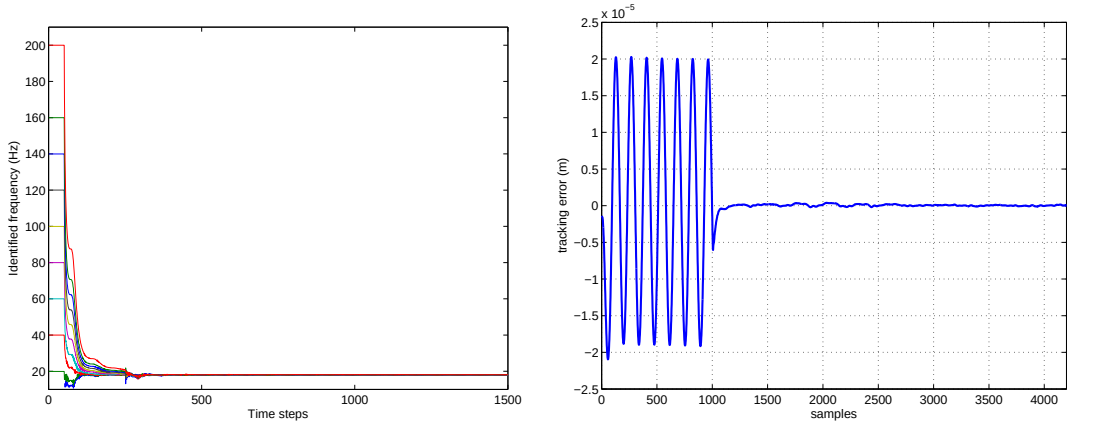
Figure 10. Frequency spectra of the position errors in rejection of environmental disturbance.

Table I. Three sigma values of the position errors before and after turning on the proposed narrow-band disturbance observer

	3σ: before	3σ: after			
freq. (Hz)	**overall/steady-state** (m)	**overall** (m)	**reduction**	**steady-state** (m)	**reduction**
40	2.0E-4	1.8E-5	**90.8%**	4.0E-7	**99.8%**
53	2.5E-4	2.7E-5	**89.4%**	5.2E-7	**99.8%**
80	8.6E-5	1.1E-5	**87.1%**	3.2E-7	**99.6%**
90	9.3E-5	1.2E-5	**86.8%**	3.6E-7	**99.6%**

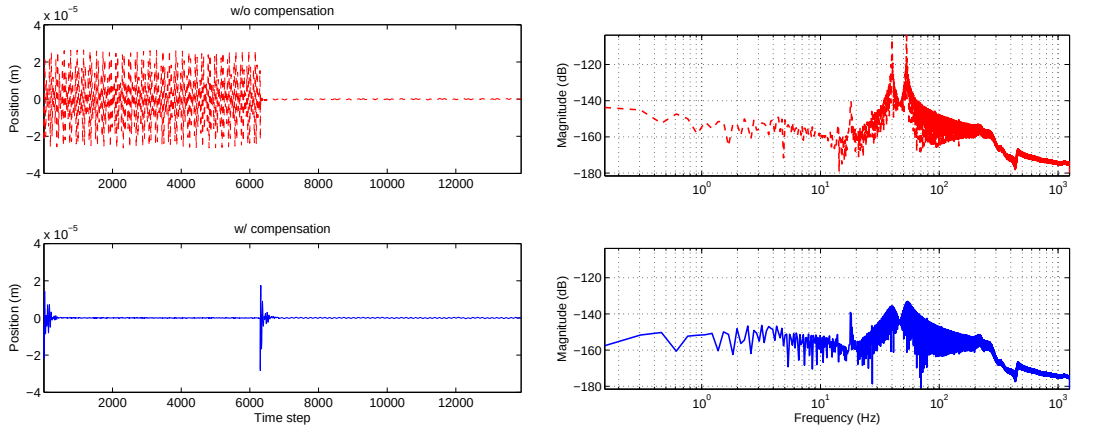
To evaluate the algorithm performance at different frequencies, a series of external disturbances were added to the system. Table I summarizes the reductions of 3σ values. The overall 3σ values measure the time-domain error reduction over the entire experiment length. The steady-state results were computed after the initial transient period. All numbers indicate strong vibration rejection at different frequencies. Based on the steady-state results of more than 99.6% error reduction, one can say that practically all disturbances were fully removed.

DOI: 10.1002/acs



(a) Effects of different initial conditions (b) Online Position error (compensation starts at the 1000-th sample)

Figure 11. Performance test under different PAA initializations.



(a) Time traces (b) Frequency spectra

Figure 12. Example rejection of two vibration components.

Fig. 11 studies performance of the adaptive algorithm under different parameter initializations. From Fig. 11a, the algorithm maintained accurate convergence even when the PAA was initialized far away from the actual optimal point. From Fig. 11b, the disturbance compensation successfully rejected the vibrations in full. The signal-to-noise ratio in this case is sufficiently large to allow a more aggressive adaptation gain. We were able to obtain a faster convergence (less than 200 time steps) than the results in Fig. 9. In practice, it is recommended to always increase the signal-to-noise ratio as much as possible, by prefiltering during adaptation.

Fig. 12 extends the evaluation to the rejection of two vibration components. In Fig. 12a, external vibrations were injected in the first 6000 time steps. Similar to the single-frequency case, the vibrations were significantly attenuated. In the frequency domain, the two spectral peaks at 40 Hz and 53 Hz were removed without virtually distinguishable error amplifications at other frequencies. There were some residual peaks due to the transient at 0th and 6000th time steps. After removal of the transient, Table II again reveals that the disturbances were fully rejected at steady state. Also shown in Table II are the performance evaluations at different frequencies. In all cases, the algorithm achieved more than 87% overall 3σ reduction and 99.4% steady-state improvement.

DOI: 10.1002/acs

Table II. Performance evaluation at different frequencies: two vibration component.

	3σ: **before**	3σ: **after**			
freq. (Hz)	**overall/steady-state** (m)	**overall** (m)	**reduction**	**steady-state** (m)	**reduction**
18, 40	2.1E-4	1.4E-5	**93.2%**	3.0E-7	**99.9%**
40, 53	4.0E-5	3.7E-6	**90.7%**	2.6E-7	**99.4%**
53, 80	5.3E-5	5.6E-6	**89.4%**	2.3E-7	**99.6%**
80, 90	6.3E-5	8.1E-6	**87.0%**	2.6E-7	**99.6%**

5. CONCLUSIONS AND DISCUSSIONS

This paper has discussed the design of adaptive narrow-band disturbance observers and the implementation for vibration rejection in semiconductor manufacturing. Overall, the algorithm is an (inverse) model-based controller parameterization, with a special internal model absorbed in the Q-filter design. Model-based design has been well recognized to be essential for high-precision high-performance control systems. The inverse-based design further decouples the baseline control system with the adaptive vibration rejection scheme. This provides conveniences in controller tuning in practice, especially for vibration rejection applications using bandpass-type Q filters.

The practical importance and historical investigations have formed adaptive vibration rejection as a rich research field. Many challenges and new applications remain to be explored. For instance, current results have been mostly considering adaptations of no more than three frequency components. Depending on the algorithm configurations, order of the adaptation vector ranges from three to six for the case of rejecting three vibration components [7]. Various engineering considerations (e.g. stability, algorithm tuning, monitoring, and the effects of numerical quantization) are needed for online adaptation of high-order IIR filters. Also, control of the transient response in closed-loop adaptive control schemes remains an unsolved issue.

References

1. A. Al Mamun, G. Guo, and C. Bi, *Hard disk drive: mechatronics and control.* CRC Press, Taylor & Francis Group, London, 2007.
2. C. Kinney, R. de Callafon, E. Dunens, R. Bargerhuff, and C. Bash, "Optimal periodic disturbance reduction for active noise cancelation," *Journal of Sound and Vibration*, vol. 305, no. 1-2, pp. 22–33, Aug. 2007.
3. R. de Callafon, J. Zeng, and C. E. Kinney, "Active noise control in a forced-air cooling system," *Control Engineering Practice*, vol. 18, no. 9, pp. 1045–1052, Sep. 2010.
4. I. D. Landau, A. Constantinescu, and D. Rey, "Adaptive narrow band disturbance rejection applied to an active suspension–an internal model principle approach," *Automatica*, vol. 41, no. 4, pp. 563–574, Apr. 2005.
5. R. A. Griffiths, A. Williams, C. Oakland, J. Roberts, A. Vijayaraghavan, and T. Thomson, "Directed self-assembly of block copolymers for use in bit patterned media fabrication," *Journal of Physics D: Applied Physics*, vol. 46, no. 50, p. 503001, 2013.
6. B. Shahsavari, E. Keikha, F. Zhang, and R. Horowitz, "Repeatable runout following in bit patterned media recording," in *Proceedings of the ASME 2014 Conference on Information Storage and Processing Systems, Santa Clara, California, USA*, June 2014, p. V001T03A001.
7. I. D. Landau, A. C. Silva, T.-B. Airimitoaie, G. Buche, and M. Noe, "Benchmark on adaptive regulation—rejection of unknown/time-varying multiple narrow band disturbances," *European Journal of Control*, vol. 19, no. 4, pp. 237 – 252, 2013.
8. B. Widrow, J. Glover Jr, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, E. Dong Jr, and R. Goodlin, "Adaptive noise cancelling: principles and applications," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.
9. M. Bodson, "Adaptive algorithms for the rejection of sinusoidal disturbances with unknown frequency," *Automatica*, vol. 33, no. 12, pp. 2213–2221, Dec. 1997.
10. R. de Callafon and C. E. Kinney, "Robust estimation and adaptive controller tuning for variance minimization in servo systems," *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, vol. 4, no. 1, pp. 130–142, 2010.

DOI: 10.1002/acs

11. Q. Zhang and L. Brown, "Noise analysis of an algorithm for uncertain frequency identification," *IEEE Transactions on Automatic Control*, vol. 51, no. 1, pp. 103–110, 2006.
12. W. Kim, H. Kim, C. Chung, and M. Tomizuka, "Adaptive output regulation for the rejection of a periodic disturbance with an unknown frequency," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 5, pp. 1296–1304, 2011.
13. F. Ben-Amara, P. T. Kabamba, and a. G. Ulsoy, "Adaptive sinusoidal disturbance rejection in linear discrete-time systems–part I: Theory," *ASME Journal of Dynamics Systems, Measurement, and Control*, vol. 121, no. 4, pp. 648–654, 1999.
14. ——, "Adaptive sinusoidal disturbance rejection in linear discrete-time systems–part II: Experiments," *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 121, no. 4, pp. 655–659, 1999.
15. R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984 –995, Jul. 1989.
16. R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276 – 280, Mar. 1986.
17. A. Nehorai, "A minimal parameter adaptive notch filter with constrained poles and zeros," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 4, pp. 983–996, Aug. 1985.
18. G. Obregon-Pulido, B. Castillo-Toledo, and A. Loukianov, "A globally convergent estimator for n-frequencies," *IEEE Transactions on Automatic Control*, vol. 47, no. 5, pp. 857 –863, May. 2002.
19. L. Hsu, R. Ortega, and G. Damm, "A globally convergent frequency estimator," *IEEE Transactions on Automatic Control*, vol. 44, no. 4, pp. 698 –713, Apr. 1999.
20. M. Mojiri and A. Bakhshai, "An adaptive notch filter for frequency estimation of a periodic signal," *IEEE Transactions on Automatic Control*, vol. 49, no. 2, pp. 314 – 318, Feb. 2004.
21. X. Xia, "Global frequency estimation using adaptive identifiers," *IEEE Transactions on Automatic Control*, vol. 47, no. 7, pp. 1188 –1193, Jul. 2002.
22. B. Wu and M. Bodson, "A magnitude/phase-locked loop approach to parameter estimation of periodic signals," *IEEE Transactions on Automatic Control*, vol. 48, no. 4, pp. 612 – 618, Apr. 2003.
23. R. Marino and P. Tomei, "Global estimation of n unknown frequencies," *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1324 – 1328, Aug. 2002.
24. R. Marino, G. L. Santosuosso, and P. Tomei, "Robust adaptive compensation of biased sinusoidal disturbances with unknown frequency," *Automatica*, vol. 39, no. 10, pp. 1755 – 1761, 2003.
25. X. Chen and M. Tomizuka, "A minimum parameter adaptive approach for rejecting multiple narrow-band disturbances with application to hard disk drives," *IEEE Transactions on Control Systems Technology*, vol. 20, pp. 408 – 415, March 2012.
26. X. Chen, W. Xi, Y. Kim, and K. Tu, "Methods for closed-loop compensation of ultra-high frequency disturbances in hard disk drives and hard disk drives utilizing same," January 2014, uS Patent 8,630,059.
27. A. Oshima, X. Chen, and M. Tomizuka, "Control design for cancellation of unnatural reaction torque and vibrations in variable-gear-ratio steering system," in *Proceedings of 2013 ASME Dynamic Systems and Control Conference, Stanford, CA*, October 2013, p. V001T11A003.
28. X. Chen and M. Tomizuka, "Selective model inversion and adaptive disturbance observer for time-varying vibration rejection on an active-suspension benchmark," *European Journal of Control*, vol. 19, no. 4, pp. 300 – 312, 2013.
29. ——, "Selective model inversion for rejection of time-varying vibrations on an active suspension," in *Proceedings of European Control Conference, Zurich, Switzerland*, July 2013, pp. 2897 – 2903.
30. ——, "New repetitive control with improved steady-state performance and accelerated transient," *IEEE Transactions on Control System Technology*, vol. 22, no. 2, pp. 664–675, March 2014.

DOI: 10.1002/acs