

# DESIGN OF AN ITERATIVE LEARNING CONTROL FOR A SERVO SYSTEM USING MULTI-DICTIONARY MATCHING PURSUIT

Iuliana Rotariu

ARC Centre of Excellence for Autonomous Systems, University of Sydney  
NSW2006, Sydney, Australia

Erik Vullings

MELCOE, Macquarie University  
NSW2109, Sydney, Australia

**Keywords:** Iterative Learning Control, System Dynamics, Mechatronics, Time-frequency analysis, Wigner distribution, Atomic decomposition, Matching Pursuit.

**Abstract:** Many motion systems repeatedly follow the same trajectory. However, in many cases, the motion system does not learn from tracking errors obtained in a previous cycle. Iterative Learning Control (ILC) resolves this issue by compensating for previous tracking errors, but it suffers from not being able to distinguish between tracking errors caused by machine dynamics versus errors caused by noise, and by trying to 'learn' the noise, additional errors are introduced.

In this paper we address this issue by using the servo error signal by identifying the time-varying nonlinear effects, which can be learned and therefore improve the position accuracy, versus the stochastic effects, which cannot be learned. The identification of these effects is performed by means of time-frequency analysis of the servo error and therefore our goal is to obtain a high-resolution time-frequency energy distribution of the analyzed signal. Here we compare the servo error energy distribution by three means: (1) Wigner distribution; (2) adaptive signal decomposition over one dictionary of modulated versions of wavelets (*simple atomic dictionary*); (3) and by means of combining several simple atomic dictionaries into a *complex atomic dictionary*. We show that the latter approach leads to the highest-resolution energy distribution and tracking performance.

## 1 INTRODUCTION

The wafer scanner mechatronic motion system is an opto-mechanical machine for producing Integrated Circuits (ICs) on a silicon wafer using a photolithographic process. One of the main components of a wafer scanner is the six degrees of freedom (DOF's) wafer stage (Rotariu et al., 2003a). This is an electro-mechanical servo system that positions the wafer (200-300mm diameter) with respect to the imaging optics. The wafer stage largely determines the *throughput* (80-100 wafers/h, 80-200 ICs/wafer) and the *accuracy* of the products, and they are both subject to severe performance requirements. Normal scan speeds and accelerations are 0.5 m/s and 10 m/s<sup>2</sup>, respectively. In order to maximize the throughput and minimize the servo error of such a complex dynamical system, advanced intelligent identification and control schemes are preferred to standard linear or robust non-linear techniques (Casalino and Bartolini, 1984).

One of such advanced intelligent control schemes, Iterative Learning Control (ILC), is an effective technique to reduce systematic control errors that occur in

systems that repetitively perform the same motion or operation (Moore, 1993).

Although the time-domain ILC results can be extended to time-varying and nonlinear systems (Goh, 1994), the *time-domain* analysis does not give useful frequency domain insights for the learning design. In addition, the time-domain analysis results do not address the issue of good transients and long-term stability, and while different schemes for tuning of the learning gain (Chang et al., 1992) have been proposed, in (Wirkander and Longman, 1999) it has been pointed out that the learning gain is not a critical factor to bandwidth. On the other hand, many *frequency domain* analysis ILC algorithms have been proposed based on frequency response methods and iteration varying filter schemes (Tang et al., 2000), (Norrlöf, 2002), but these do not give insightful time-domain information for the learning design or they depend heavily on the system model.

A logical advance of the above-mentioned time-based and frequency-based ILC methods is to use an ILC based on *time-frequency* analysis of control signals. This has been first proposed in (Chen and

Moore, 2001), where an adaptive scheme of learning feedforward control based on a B-spline network is presented. In (Zhang et al., 2005) the use of wavelet packet transform for time-frequency analysis and design of a cutoff frequency tuning for the ILC scheme is proposed. In (Zheng and Alleyne, 2001), (Rotariu et al., 2003a) continuous Wigner transform is used to analyze the signals and in (Tharayil and Alleyne, 2004) and (Rotariu et al., 2006) an adaptive robustness filter based on quadratic time-frequency analysis (Wigner distribution) of the control signals is proposed. In case of the Wigner distribution, extensive studies have been made and methods (Cappellini and Constantinides, 1984), (Rotariu et al., 2006) devised to remove the cross-terms in some way, but these methods do not improve the frequency resolution. Therefore, we have to focus on alternative methods that do increase the frequency resolution and reduce the cross-terms as well.

In this paper we propose an adaptive ILC based on high-resolution time-frequency analysis of the control signals which is performed by means of signal decomposition over a simple versus a complex time-frequency atomic dictionary. We shall show that our analysis in these two cases leads to a better understanding of the systems dynamics and more insightful learning information than (piecewise) Wigner-based adaptive ILC, while achieving a very good tracking performance.

In Section II, the results of the signal analysis by means of Wigner distribution will be used to find a suitable profile for the bandwidth of the time-varying robustness filter. In the end of this section it will become clear that by increasing the resolution and accuracy of the time-frequency distribution, the tracking performance of the ILC improves. In order to achieve this, in Section 2.2 we shall introduce another two different time-frequency analysis methods of the servo error signal than Wigner distribution; as consequence, in Section 2.2 we shall show that the design of the adaptive ILC changes and that this leads to very good tracking performance of the proposed learning algorithm.

## 2 TIME-FREQUENCY ADAPTIVE ILC

This section discusses the time-frequency adaptive ILC, i.e. time-varying adaptive ILC whose design is based on quadratic time-frequency representation of the control signals. Intuitively, this means that the ILC 'gain' is governed by the dynamics still present in the error signal after the previous iteration: high dynamic behavior leads to a large gain, low dynamic behavior (noise) is ignored.

### 2.1 Time-varying Adaptive ILC

In this section the concept of time-varying adaptive ILC is introduced (see Figure 1). We restrict the study to the case where the plant is a causal, LTI dynamical system  $P$ .  $C$  is a feedback controller which insures the stability of the closed loop system.

We suppose that the desired response  $r$  is defined on the interval  $(t_0, t_f)$ , where  $t_f \leq \infty$  and the initial conditions are the same at the beginning of each iteration.

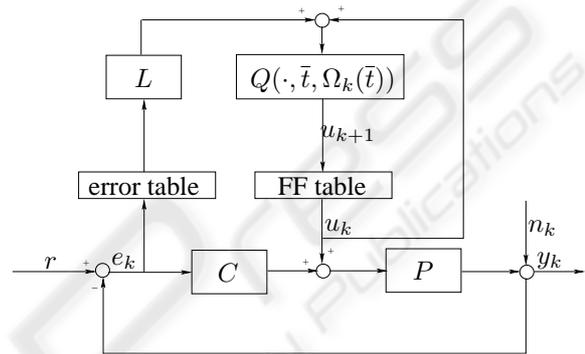


Figure 1: A block schematic for adaptive ILC.

The goal of the ILC design is to find the feedforward signal  $u^*$  such that  $r = Pu^*$ . We seek a sequence of inputs  $u_k$  with the property that  $\lim_{k \rightarrow \infty} u_k = u^*$ , where the index  $k$  is the iteration.

The adaptive ILC design consists of the design of the  $L$  and  $Q$  filters. The learning filter  $L$  is the same as for standard ILC, i.e. it has to approximate a stable inverse of the modeled process sensitivity function  $P_s(s) = \frac{P}{1+PC}$  (Rotariu et al., 2004).

The process sensitivity function as steady-state transfer function is measured at the center wafer position and it does not account for system's position dependent dynamics within the scanning trajectory (Rotariu et al., 2004) (Rotariu et al., 2003a) (Rotariu et al., 2003b). We replace the fixed  $Q$  robustness filter (steady-state filter, not changing from one iteration to the other) of standard ILC with a time-varying  $Q$ -filter  $Q_k(s, \bar{t}, \Omega_k(\bar{t}))$ , namely a zero-phase low-pass Butterworth filter of order  $n$  and cut-off frequency  $\Omega_k(\bar{t})$ , where  $\bar{t} \in [t_{0(k)}, t_{0(k)} + T]$ ,  $t_{0(k)}$  is the initial time of the  $k^{th}$  iteration, and  $T$  is the time required to perform the trajectory. The cut-off frequency  $\Omega_k = \Omega_k(\bar{t})$  may vary throughout the length of each iteration. In what follows, we denote by  $\Gamma_k(t, \bar{t}, \Omega_k(\bar{t}))$  the inverse Fourier transform of the Butterworth filter  $Q_k(s, \bar{t}, \Omega_k(\bar{t}))$  as a function in the variable  $s$ :

$$Q_k(s, \bar{t}, \Omega_k(\bar{t})) \xrightarrow{\mathcal{F}^{-1}} \Gamma_k(t, \bar{t}, \Omega_k(\bar{t})). \quad (1)$$

A converging (Rotariu et al., 2006) adaptive ILC update law is given by

$$e_k = e^r - P_s u_k - S n_k, \quad (2)$$

$$u_{k+1}(\bar{t}) = \int_{-\infty}^{\infty} \Gamma_k(\tau, \bar{t}, \sigma_k(\bar{t})) (u_k + L e_k)(\bar{t} - \tau) d\tau, \quad (3)$$

where  $e_k$  is the error signal,  $u_k$  the feedforward signal and  $n_k$  an output disturbance (see Figure 1). The formula (3) is known as the nonstationary convolutional integral (Margrave, 1998) which is an extension of the convolutional method to nonstationary processes. We refer to (Zheng and Alleyne, 2003), (Tharayil and Alleyne, 2004) and (Rotariu et al., 2006) for a rigorous convergence analysis of our approach.

Next we present the design of the time-varying Q-filter that is based on the time-frequency analysis of the error signal.

## 2.2 Time-frequency Analysis of the Servo Error Signal

We will present three alternative approaches to determine the time-frequency analysis of the servo error: using Wigner, and matching pursuit with a single and multiple dictionaries. Intuitively, the better the time-frequency representation, the better will we be able to reduce the servo error signal, and we therefore strive for the best TF representation.

### 2.2.1 Wigner Distribution

The Wigner distribution (Mecklenbräuker et al., 1997) is defined by

$$W_h(t, f) = \frac{1}{2\pi} \int_{-\infty}^{\infty} h^* \left( t - \frac{\tau}{2} \right) h \left( t + \frac{\tau}{2} \right) e^{-2\pi j f \tau} d\tau, \quad (4)$$

where time  $t \in \mathbf{R}$  in [s], the frequency  $f \in \mathbf{R}$  in [Hz], and  $h^*$  is the complex conjugate of the analyzed time-signal  $h$ . The distribution for real signals is real-valued and can – due to its quadratic form – be physically interpreted as the distribution of the signal's energy over both time and frequency. Although the Wigner distribution is especially appropriate for the analysis of non-stationary multi-component signals (Cohen, 1989), its main deficiency is the cross-term interference: each pair of signal components or signal component and noise creates one additional cross-term in the spectrum, thus the resulting time-frequency representation may be confusing (see Figure 2).

Next we show how to eliminate the cross-terms while maintaining a high-resolution energy distribution of the servo error.

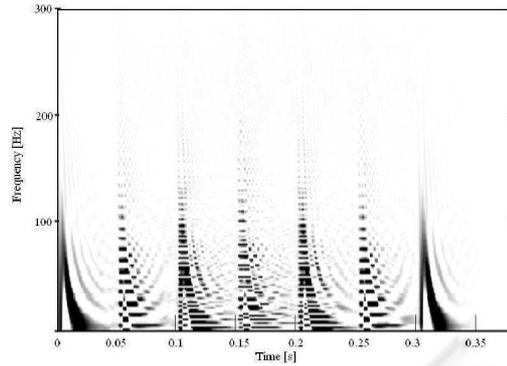


Figure 2: The absolute value of the Wigner distribution of the servo error signal. Darker means higher relative energy. Seven intervals can be identified with significant energy content, of which three spurious ones (2, 4, and 6) due to cross-terms. Note that also the non-spurious third and fifth peaks are distorted by cross-terms.

### 2.2.2 Atomic Decomposition of the Servo Error and Matching Pursuit in $L_2(\mathbf{R})$

Decomposition of signals over window Fourier transforms and wavelet transforms are the best known examples of signal decomposition over a family of functions that are well localized in time and frequency. In this section, we shall discuss time-frequency atomic decomposition, also known as adaptive decomposition, of a signal and we shall describe a general iterative decomposition algorithm known as matching pursuit (MP) (Mallat and Zhang, 1993). We shall also explain why this atomic decomposition is well fitted for servo error signal decomposition. Based on this decomposition we shall show the servo error energy distribution over a simple and complex atomic dictionary.

MP is based on a family  $\mathcal{D}$  of time-frequency atoms that can be generated by scaling, translating and modulating a single window function  $g \in L^2(\mathbf{R})$ . We suppose that the function  $g$  is real, continuously differentiable, non-zero,  $g(0) \neq 0$  and  $g(t) \sim \mathcal{O}(\frac{1}{1+t^2})$ . In addition, we impose that  $\|g\| = 1$  and that  $\int_{\mathbf{R}} g(t) dt \neq 0$ . For any scale  $s > 0$ , frequency modulation  $\phi$  [Hz] and translation  $u$ , one defines

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) e^{2\pi j \phi t}, \quad (5)$$

with  $\gamma = (s, u, \phi) \in \mathbf{R}^+ \times \mathbf{R}^2$ . The factor  $\frac{1}{\sqrt{s}}$  normalizes to 1 the  $L_2$  norm of  $g_\gamma$ . We also observe that atoms defined in (5) look like  $\phi$ -modulated versions of a doubly-indexed family of wavelets  $\psi_{s,u} = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right)$ , with the difference that for atoms one imposes the condition  $\int_{\mathbf{R}} g(t) dt \neq 0$  while the admissibility condition for wavelets requires  $\int_{\mathbf{R}} \psi(t) dt =$

0. If  $g$  is even as in most situations, then  $g_\gamma(t)$  is well concentrated in time and frequency (Mallat and Zhang, 1993).

The family  $\mathcal{D} = \{g_\gamma\}_\gamma$  is extremely redundant (Torresani, 1991). To represent any function  $h$  efficiently, we must select an appropriate countable subset of atoms  $\{g_{\gamma_n}\}_n$  so that

$$h = \sum_{n=-\infty}^{\infty} a_n g_{\gamma_n}. \quad (6)$$

Consequently, for signals that include scaling and highly oscillatory structures one cannot define a priori appropriate constraints in scale and modulation parameters of the atoms  $g_{\gamma_n}$  used in expansion (6). The elements of the dictionary  $\mathcal{D} = \{g_{\gamma_n}\}_n$  need to be selected adaptively, depending on local properties of the function  $h$ .

In our case, the servo error contains non-stationary random structures (machine vibrations and sensor noise), deterministic chaotic vibrations, and subharmonic oscillations that are known to demonstrate narrow high frequency support (Yen and Lin, 2000). For this reason, the decomposition of error signals over triple-indexed time-frequency atoms (5) enables the extraction of signal features that combine non-stationary, deterministic chaotic and transient chaotic characteristics

Next we'll give an outline of the MP algorithm in the Hilbert space  $L^2(\mathbf{R})$ . We first approximate  $h \in L_2(\mathbf{R})$  with linear projections on elements of  $\mathcal{D}$ . Then it follows that

$$h = \langle h, g_{\gamma_0} \rangle g_{\gamma_0} + Rh, \quad (7)$$

where  $Rh \in L^2(\mathbf{R})$  is the residual after approximating  $h$  in the direction  $g_{\gamma_0}$ . Since  $g_{\gamma_0}$  is orthogonal on  $Rh$ , it follows that

$$\|h\|^2 = |\langle h, g_{\gamma_0} \rangle|^2 + \|Rh\|^2.$$

To minimize  $\|Rh\|$ , we chose  $g_{\gamma_0} \in \mathcal{D}$  such that  $|\langle h, g_{\gamma_0} \rangle|$  is maximum. After the first step decomposition (7), we continue iteratively by sub-decomposing the residual  $Rh$  by projecting it on a vector of  $\mathcal{D}$  that matches  $Rh$  the best, as we have done for  $h$ . Therefore, we inductively obtain the  $m^{\text{th}}$  order decomposition of  $h$  over the dictionary  $\mathcal{D}$ ,

$$h = \sum_{n=0}^{m-1} \langle R^n h, g_{\gamma_n} \rangle g_{\gamma_n} + R^m h, \quad (8)$$

where we denote by  $R^m h$  the residual obtained at the  $m^{\text{th}}$  order decomposition of  $h$ . Using (8), one can easily obtain the following important result:

*Theorem.*(Mallat and Zhang, 1993) If  $\mathcal{D}$  is complete ( $\text{span}(\mathcal{D}) = L_2(\mathbf{R})$ ) then

$$h = \sum_{n=0}^{\infty} \langle R^n h, g_{\gamma_n} \rangle g_{\gamma_n} \quad (9)$$

and

$$\|h\| = \sum_{n=0}^{\infty} |\langle R^n h, g_{\gamma_n} \rangle|^2. \quad (10)$$

*Remark.* Finite linear expansions of time-frequency atoms (5) are dense in  $L^2(\mathbf{R})$  and therefore this dictionary is complete.

The smallest complete dictionaries are bases. By decomposing a signal onto an orthonormal bases of compactly supported wavelets having a certain number of vanishing moments (Daubechies, 1991), correlation between scales is avoided. One-dimensional well localized in time (compactly supported) wavelets are of the greatest interest for applications because of the simplest numerical realization of expansion and synthesis algorithms. The number of the vanishing moments is especially important when one wants to quickly compress large data sets. By using compactly supported wavelets that have a relatively high number of vanishing moments, the  $L_2$  norm of the residual will decrease faster than when using other wavelets that have less vanishing moments. On the other hand, for feature extraction tasks, choosing a too high number of vanishing moments is not desirable as we are interested in the non-redundant high frequency components of the signal (Chandroth, 1999). In other words, we are interested to decompose the signal into time-frequency atoms that describe the non-smooth (nonlinear and nonstationary) behavior well while preserving regular components of the servo error (Struzik and Siebes, 1998).

Next we show in Figure 3 the servo error decomposition with respect to a simple atomic dictionary  $\mathcal{D}$  which is built with Symmlets. By applying the MP algorithm, we found that the decomposition of the analyzed servo error signal with respect to Symmlets with 9 vanishing moments provide the best results: for a given number of iterations, the  $L_2$  norm of the residuals given by formula (8) becomes smaller than when using other simple atomic dictionaries, like those that are built with Symmlets  $\{4, 5, 6, 7, 8, 10\}$ , Daubechies, Coiflets, and Haar wavelets.

At the end of this section we consider a complex atomic dictionary  $\mathcal{D}$  built with the asymmetric Daubechies' wavelets and by their more symmetric and larger supported closely related cousins, i.e. Symmlets and Coiflets (Daubechies, 1991). We apply the MP algorithm for the servo error decomposition with respect to this dictionary and we obtain a smaller  $L_2$  norm of the residuals (8) than when the servo error is decomposed with Symmlets, see Figure 3. Through our numerical experiments we use a quadrature mirror filter bank MP algorithm, see (Mallat and Zhang, 1993), (Buckheit et al., 1995), (Rioul and Vetterli, 1991).

Based on the decomposition (9) of any  $h \in L^2(\mathbf{R})$  over a simple or complex dictionary, and the defini-

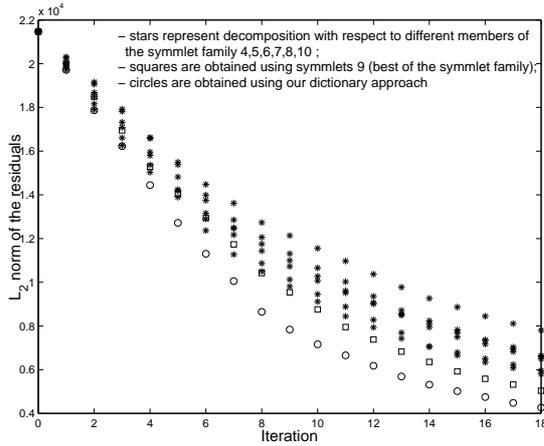


Figure 3: The  $L_2$  norm of the residuals (8) using a simple atomic dictionary built with Symmlets. After a relatively small number of iterations, Symmlets 9 show the best signal decomposition. The best results, however, are obtained using our multi-dictionary approach.

tion of the Wigner distribution in (4), we obtain

$$W_h(t, f) = \sum_{n=0}^{\infty} |\langle R^n h, g_{\gamma_n} \rangle|^2 W_{g_{\gamma_n}}(t, f) + \sum_{n=0}^{\infty} \sum_{m=0, m \neq n}^{\infty} |\langle R^n h, g_{\gamma_n} \rangle| |\langle R^m h, g_{\gamma_m} \rangle| W_{n,m},$$

where

$$W_{n,m} = \frac{1}{2\pi} \int_{\mathbf{R}} g_{\gamma_n}(t + \frac{\tau}{2}) \overline{g_{\gamma_m}(t - \frac{\tau}{2})} e^{-2\pi j f \tau} d\tau$$

denotes the cross Wigner distribution of the atoms  $g_{\gamma_n}$  and  $g_{\gamma_m}$ .

The double sum corresponds to the cross-terms of the Wigner distribution that we try to remove in order to obtain a clear time-frequency distribution of the signal  $h$ . We only keep the first sum and define the energy distribution of the signal  $h$  over the time-frequency plane as

$$E_h(t, f) = \sum_{n=0}^{\infty} |\langle R^n h, g_{\gamma_n} \rangle|^2 W_{g_{\gamma_n}}(t, f). \quad (11)$$

By taking the absolute value of the energy distribution defined in formula (11) with  $f = e_0$ , when the dictionary  $\mathcal{D} = \{g_{\gamma_n}\}$  is built with a complex atomic dictionary generated by Symmlets, Coiflets and Daubechies, we obtain the time-frequency energy distribution plotted in Figure 4.

### 2.3 Design of a Bandwidth Profile

Next we shall design a bandwidth profile for the time-varying robustness filter introduced in (1).

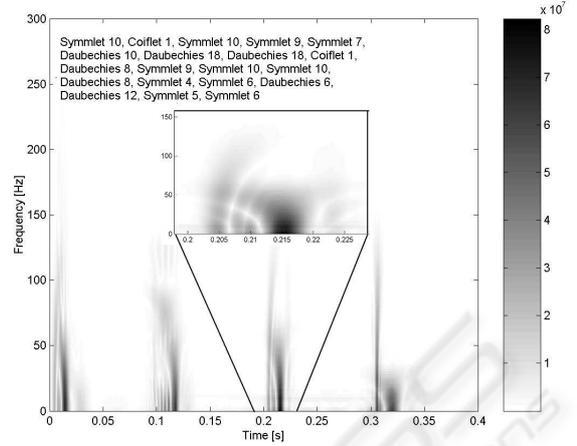


Figure 4: Servo error energy distribution using a multi-dictionary wavelet packet generated by combining the Daubechies, Symmlets and Coiflets atoms (selected MP atoms are listed). The small insert zooms in to show the 'banana' shapes, i.e. they consist of narrow-band subharmonic oscillations and chaotic vibrations.

Consider the time vector  $\bar{t} = (t_i)_{i \in \mathbb{N}}$ ; the initial feedforward signal  $u_0$  is chosen identically zero and the error signal  $e_0$  is the measured servo error when a rigid body acceleration feed-forward is applied. The elements of the vector  $\Omega_0(\bar{t})$  are chosen equally small values, i.e. the initial time-varying robustness filter is a steady-state filter whose bandwidth is high enough (about 500 [Hz]) such that it does not filter the deterministic content of the servo error. Because of this choice and because the learning filter  $L \sim P_s^{-1}$  behaves as a 1000 [Hz] low-pass filter, by (2) it follows that the energy distribution of error signal  $e_0$  and feedforward signal  $u_1$  are similar in shape and for the design of the bandwidth  $\Omega_1(\bar{t})$  it is not important whether we analyze the servo error  $e_0$  or its high bandwidth low-pass filtered version  $u_1(\bar{t})$ .

The adaptive update law  $\Omega_k(\bar{t}) \rightarrow \Omega_{k+1}(\bar{t})$  for the design of the time-varying bandwidth  $\Omega_k(\bar{t})$  of the robustness filter  $Q_k(s, \bar{t}, \Omega_k(\bar{t}))$  for any iteration  $k$  contains the frequency envelope  $F_{max,k}(\bar{t})$  as gain. This encompasses the frequencies of all signal components at each time-instant whose energy exceeds  $C_e$ , the value of the noise during standstill:

$$F_{max,k}(\bar{t}) = \max(\omega_k(\bar{t})), \text{ for } H_{u_k}(\bar{t}, \omega_k(\bar{t})) \geq C_e, \quad (12)$$

where  $H_{u_k}$  is the time-frequency energy distribution of the feedforward signal  $u_k$ ,  $\omega_k(\bar{t})$  is the cross-section of  $H_{u_k}$  and  $C_e$  height plane (see Figure 5).

The envelope  $F_{max}(\bar{t})$  is used as a gain in an adaptive update law. This law changes the bandwidth profile  $\Omega(\bar{t})$  after each iteration, when the effects of the

previous change on the measured error are evaluated. After a new bandwidth profile has been implemented, its benefit is evaluated by the function  $\Delta N_k(\bar{t})$ , which compares the local  $\ell_2$  norm of the current error to that of the error at the previous iteration, such that

$$\Delta N_k(\bar{t}) = N_k(\bar{t}) - N_{k-1}(\bar{t}) \quad (13)$$

where

$$N_k(t_i) = \sum_{j=i-T_w/2}^{i+T_w/2} e_k^2(t_j), \quad (14)$$

$T_w > 0$  gives the width of the window where the signals are locally compared.

After introducing the terms  $F_{max,k}(\bar{t})$  and  $\Delta N_k(\bar{t})$ , we are now ready to give the bandwidth update rule

$$\begin{aligned} \Omega_{k+1}(\bar{t}) &= \Omega_k(\bar{t}) + \Delta \Omega_k(\bar{t}), \\ \Delta \Omega_k(\bar{t}) &= F_{max,k}(\bar{t}) \cdot \Delta N_k(\bar{t}) \cdot K_k(\bar{t}) \end{aligned} \quad (15)$$

where the term  $K_k(\bar{t}) = -\text{sign}(\Delta \Omega_{k-1}(\bar{t}))$  is introduced to add the following logic to the mechanism: if the bandwidth was previously increased ( $\Delta \Omega_{k-1}(t_i) > 0$ ), while the error decreased ( $\Delta N_k(t_i) < 0$ ), this change was beneficial and the bandwidth may be further increased. On the other hand, if an increase in the bandwidth resulted in a larger error, this was obviously not the case and the bandwidth should be lowered again. The combination  $\Delta N_k(\bar{t}) \cdot K_k(\bar{t})$  results in this kind of update behavior.

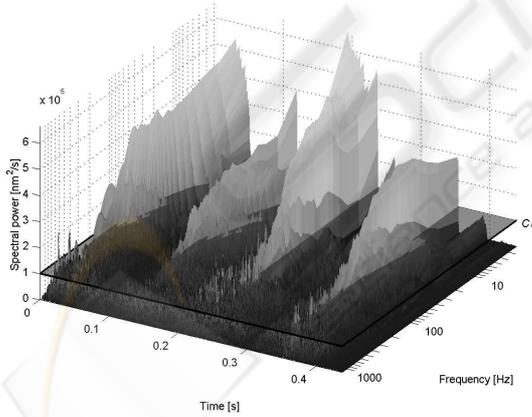


Figure 5: A 3D-plot of the Wigner distribution for  $k = 1$  (therefore of  $u_1$ ). The horizontal plane at energy value  $C_e$  discriminates deterministic signal components from noise. As the learning filter  $L \sim P_s^{-1}$  behaves as a 1000 [Hz] low-pass filter, by (2) and (3) it follows that the energy distribution of error signal  $e_0$  and feedforward signal  $u_1$  are similar in shape.

Applying the above algorithm in this section with  $H_{u_k}(\bar{t}, \omega_k(\bar{t})) \stackrel{def}{=} E_{u_k}(\bar{t}, \omega_k(\bar{t}))$  to the multi-

dictionary approach, we obtain the bandwidth profile plot in Figure 6.

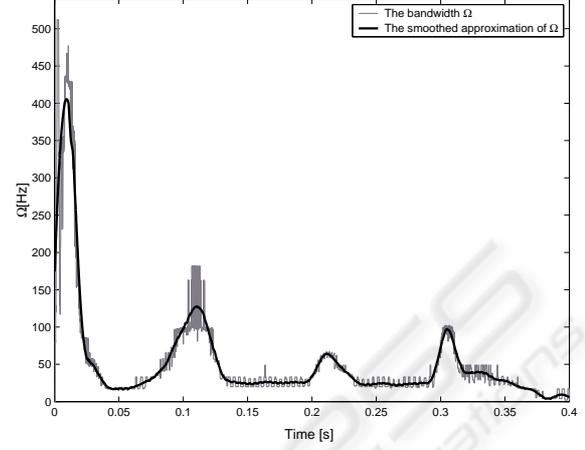


Figure 6: The bandwidth profile  $\Omega(\bar{t})$  at iteration  $k = 1$ ; design based on the servo error energy distribution over the complex time-frequency atomic dictionary found in the end of Section 2.2.

Finally, in Figure 7 we show the servo error signal when Adaptive ILC with the bandwidth of the Q-filter as shown in Figure 6 is implemented on the wafer stage test rig.

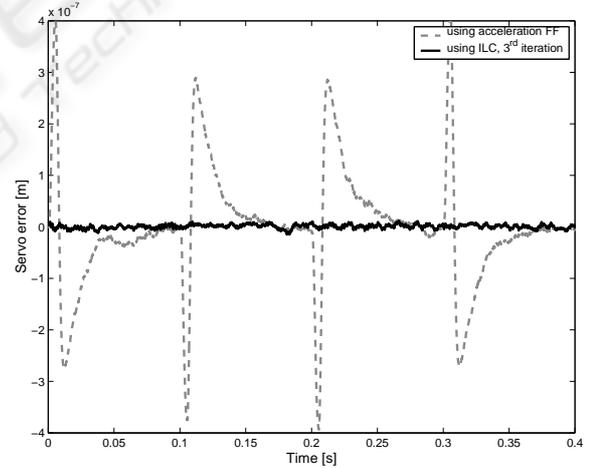


Figure 7: Error signal when rigid body acceleration feedforward is applied and Adaptive ILC with the bandwidth of the Q-filter as shown in Figure 6.

### 3 CONCLUSIONS

When comparing the resulting time-frequency energy distributions in Figures 2 and 4, we note the following:

The energy distribution based on the complex atomic dictionary in Figure 4 is better localized in time and frequency than the other. Also its resolution is much higher, and its high-energy deterministic components can be much better separated from low-energy stochastic components.

As we have announced, we are not only interested in separating deterministic and stochastic components in control signals, but also in separating their constituents, i.e. periodic, subharmonic, chaotic and transient for deterministic nonlinear time-varying components versus non-stationary and stationary for stochastic components. The non-stationary and deterministic chaotic vibrations present in the servo error signal have narrow high-frequency support and are difficult to be detected using Wigner distribution. By applying the adaptive decomposition algorithm described in Section 2.2.2, we are also able to identify such narrow-band frequency characteristics, as can be seen from the ‘banana’ shape of the dark patterns in Figure 4, which are much more expressive than the black patterns in Figure 2.

In Figure 2 the local cross-terms around  $t = \{0, 0.1, 0.2, 0.3\}$  are not eliminated, while in Figure 4 they are not present anymore.

The energy plot in Figure 4 does not contain cross-terms and any numerically undesired effects, as in Figure 2.

Additionally, our multi-dictionary approach leads to a clearer identification of time-varying non-linear and stochastic servo error components and therefore, to an improved adaptive ILC design.

We further observe that an overall smaller bandwidth than in Figure 2 is obtained (not higher than about 400 [Hz] in the beginning of the acceleration profile and 100 [Hz] around other jerk moments). Therefore, the ILC needs to learn only around these jerk moments up to smaller frequencies than those found when Wigner distribution was applied for the computation of the bandwidth profile.

This means that the stochastic effects will not be amplified unnecessarily (high bandwidth means good tracking performance and noise amplification) while all existent deterministic effects will be learned. Also, unlike in Figure 2, we do not obtain any increased bandwidth because of cross terms: the bandwidth of the Q-filter needs to be increased just around the jerk moments while, in between, a small bandwidth can be maintained.

Also, as seen in Figure 4, the bandwidth profile has a very good smooth approximation. Therefore, fast switching between the cut-off frequency of Q-filters that corresponds to different time instances is avoided and the stability of the switched ILC system is not an issue anymore.

## REFERENCES

- Buckheit, J., Chen, S., Donoho, D., Johnstone, I., and Scargle, J. (1995). Wavelab. In *Tech. Rep.*, Stanford University.
- Cappellini, V. and Constantinides, A. (1984). *Interference terms in the Wigner distribution*. Digital Signal Proc.
- Casalino, G. and Bartolini, G. (1984). A learning procedure for the control of movements of robotic manipulators. In *Proc. 4th IASTED Symp. Robotics Automation*, pages 108–111.
- Chandroth, G. O. (1999). Diagnostic classifier ensembles: Enforcing diversity for reliability in the combination. In *University of Sheffield, UK*.
- Chang, C.-K., Longman, R. W., and Phan, M. Q. (1992). Techniques for improving transients in learning control systems. *Adv. Astronautic. Sci.*, 76:20352052.
- Chen, Y. and Moore, K. (2001). Frequency domain adaptive learning feedforward control. In *Proc. of the 2001 IEEE Int. Symp. Comp. Intell. Robotics and Automation*, pages 396–401.
- Cohen, L. (1989). Time-frequency distributions – a review. *Proc. IEEE*, 77(7).
- Daubechies, I. (1991). *Ten lectures on wavelets*. Series in Applied mathematics, SIAM.
- Goh, C. J. (1994). A frequency domain analysis of learning control. *J. Dynamic Syst., Measurement, Contr.*, 116:781–786.
- Mallat, S. and Zhang, Z. (1993). Matching pursuit with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 41(12):3397–3415.
- Margrave, G. F. (1998). Theory of nonstationary linear filtering in the fourier domain with applications to time-variant filtering. *Geophysics*, 63(1):244 – 259.
- Mecklenbräuker, W., Hlawatsch, F., and Janssen, A. (1997). The wigner distribution, theory and applications in signal processing. *Elsevier*.
- Moore, K. L. (1993). *Iterative learning control for deterministic systems*. Adv. in Ind. Control. Springer-Verlag.
- Norrlöf, M. (2002). Iteration varying filters in iterative learning control. In *Proc. 4th Asian Control Conf.*, page 21242129.
- Rioul, O. and Vetterli, M. (1991). Wavelets and signal processing. *IEEE Signal Processing Magazine*, 8(4):14–38.
- Rotariu, I., Dijkstra, B., and Steinbuch, M. (2004). Comparison of standard and lifted ilc on a motion system. In *3rd IFAC Symp. Mechatronic Systems, Sydney, Australia*, pages 211–6.
- Rotariu, I., Ellenbroek, R., and Steinbuch, M. (2003a). Time-frequency analysis of a motion system with learning control. In *Proc. ACC Conf.*, pages 3650 – 3654. American Control Conf.

- Rotariu, I., Ellenbroek, R., and Steinbuch, M. (2006). Adaptive iterative learning control with application to a wafer stage. *submitted to IEEE Transactions on Control Systems Technology*.
- Rotariu, I., Ellenbroek, R., van Baars, G., and Steinbuch, M. (2003b). Scan-length independent iterative learning control applied to a wafer stage motion system. In *Proc. ECC Conf.*
- Struzik, Z. and Siebes, A. (1998). Wavelet transform in similarity paradigm i and ii. In *Centrum voor Wiskunde en Informatica, Report, INS-R9802*.
- Tang, X., Cai, L., and Huang, W. (2000). A learning controller for robot manipulators using fourier series. *IEEE Trans. Robot. Automat.*, 16(1):3645.
- Tharayil, M. and Alleyne, A. (2004). Design and convergence of a time-varying iterative learning control law. In *Proc. IMECE Int. Mech. Eng. Congr. and R&D Expo.*
- Torresani, B. (1991). Wavelets associated with representations of the affine weyl-heisenberg group. *J. Math Physics*, 32(5):1273–9.
- Wirkander, S.-L. and Longman, R. (1999). Limit cycles for improved performance in self-tuning learning control. *Adv. Astronautic. Sci.*, 102:763–781.
- Yen, G. and Lin, K. (2000). Wavelet packet feature extraction for vibration monitoring. *IEEE Trans. Ind. Electronics*, 47(3):650–67.
- Zhang, B., Wang, D., and Ye, Y. (2005). Wavelet transform-based frequency tuning ilc. *IEEE Trans. Systems, Man & Cybernetics, Part B: Cybernetics*, 35(1):107–114.
- Zheng, D. and Alleyne, A. (2001). Adaptive iterative learning control for systems with non-smooth nonlinearities. In *Proc. ASME Conf.*, Pittsburgh, Pennsylvania. ASME Int. DETC.
- Zheng, D. and Alleyne, A. (2003). Stability of a novel iterative learning control scheme with adaptive filtering. In *Proc. ACC Conf.*

