Available online at www.sciencedirect.com
SCIENCE DIRECT®
Control Engineering Practice 12 (2004) 1241–1252
CONTROL ENGINEERING PRACTICE
www.elsevier.com/locate/conengprac

Data-based hybrid modelling of the component placement process in pick-and-place machines ☆

A.Lj. Juloski[a,*], W.P.M.H. Heemels[a], G. Ferrari-Trecate[b]

[a] Department of Electrical Engineering, Eindhoven University of Technology, EH 4.15, P.O. Box 513, 5600MB Eindhoven, The Netherlands
[b] INRIA, Domaine de Voluceau, Rocquencourt-B.P.105, 78153 Le Chesnay, Cedex, France

Received 1 October 2003; accepted 3 April 2004

Abstract

In this paper, an experimental study in the identification of the electronic component placement process in pick-and-place machines is presented. Unilateral contact and saturation phenomena characterize the hybrid dynamics of the system. Furthermore, the mode switch cannot be measured and identification algorithms for hybrid systems, that are capable of reconstructing both the modes and the switching law, must be used. Piece-wise autoregressive exogenous (PWARX) models, which consist of a number of ARX modes together with the partition of the regressor space into regions where each model is valid are identified. Reconstructed models are able to capture the relevant dynamics of the experimental setup. Practical insights on hybrid system identification and comments on possible improvements of the identification algorithm complement the quantitative results.

Keywords: Pick-and-place machine; Hybrid systems; Piecewise affine systems; Identification

1. Introduction

In this paper, an experimental study in the identification of the electronic component placement process in pick-and-place machines is presented. Pick-and-place machines are used to automatically place electronic components on printed circuit boards (PCBs), and form a key part of an automated PCB assembly line. A pick-and-place machine works as follows: the PCB is placed in the working area of the mounting head; the mounting head, carrying an electronic component (using, for instance, a vacuum pipette), is navigated to the position where the component should be placed on the PCB; the component is placed, released, and the process is repeated with the next component. A fast component mounter, consisting of 12 mounting heads working in parallel, is shown in Fig. 1. The throughput of such a configuration can be up to 96.000 placed components per hour (Assembleon, 2002).

Consider the subtask of the component placement on the PCB. Assuming that the mounting head, carrying the component, is in the right position above the PCB, the component is pushed down until it comes in contact with the PCB and then released. The PCB is not rigid, but, depending on the material, has certain elasticity properties. The whole operation should be as fast as possible (to achieve maximal throughput), while satisfying technological and safety constraints (e.g. the exerted forces must not damage the component).

As detailed in Section 2, during the placement process, switching between several different modes of operation occurs. This motivates the search for the model in the form of a hybrid system (Juloski, Heemels, & Ferrari-Trecate, 2003). Research on hybrid systems identification has been mainly concerned with the reconstruction of piece-wise autoregressive exogenous (PWARX) models, which can be further recast into piecewise affine (Sontag, 1981), mixed logic dynamics

☆This work has been financially supported by STW/PROGRESS Grant EES.5173, and EU Grant SICONOS (IST-2001-37172). Part of the material appearing in this paper has been presented in the paper by Juloski et al. (2003) at the IFAC Conference on Analysis and Design of Hybrid Systems, Saint Malo, France, 2003.

*Corresponding author. Tel.: +31-40-247-3142; fax: +31-40-243-4582.

E-mail addresses: a.juloski@tue.nl (A.Lj. Juloski), w.p.m.h.heemels@tue.nl (W.P.M.H. Heemels), giancarlo.ferrari-trecate@inria.fr (G. Ferrari-Trecate).

0967-0661/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.conengprac.2004.04.004

Fig. 1. Fast component mounter (courtesy of Assembleon).

(Bemporad & Morari, 1999) or linear complementarity (Heemels, Schumacher, & Weiland, 2000) systems. A PWARX model consists of a finite number of ARX modes, together with a polyhedral partition of the regressor space into regions where each mode is active. Several techniques for identification of PWARX systems have been proposed (Ferrari-Trecate, Muselli, Liberati, & Morari, 2003; Bemporad, Roll, & Ljung, 2000a; Bemporad, Garulli, Paoletti, & Vicino, 2003; Vidal, Soatto, Ma, & Sastry, 2003). In this work, the algorithm developed by Ferrari-Trecate et al. (2003) is adopted. The algorithm is briefly summarized in Section 3.

The identification of the component placement process aims at providing models for analysis and control design. In particular, model predictive control (MPC) and verification algorithms for PWA and MLD systems (Bemporad & Morari, 1999; Bemporad, Torrisi, & Morari, 2000b) can be straightforwardly applied to hybrid systems in the PWARX form. Generally speaking, MPC and verification algorithms for PWA systems can be recast into mixed-integer programming problems where the process model enters the constraints. Therefore, the application of these techniques is subordinated to the availability of dynamic models of the system considered. From one side, it is of interest to obtain models with the smallest number of states and modes, as the computational complexity of control and verification methods may increase dramatically with model complexity (Bemporad & Morari, 1999; Bemporad et al., 2000b). From the other side, it is important to note that such methods require the *simulation* of the process over a horizon of $N > 1$ steps. Therefore, models optimized for one-step-ahead predictions may not be satisfactory since an accurate simulation performance is required.

The fact that models tailored to one-step-ahead prediction may perform poorly in simulation is well known in the context of ARX systems (Ljung, 1999), and it can become more noticeable for PWARX systems. Indeed, prediction errors due to imperfect estimates in mode parameters and regions, may lead to a wrong choice of the next mode, so triggering large-error propagation phenomena. Since all the existing identification algorithms for hybrid systems seek for one-step-ahead predictive models, it is of paramount importance to check the model quality in simulation.

In Sections 4 and 5, the identification results are presented. Models of increasing complexity are considered and their simulation properties discussed. It is shown that the quality of the simpler models, capturing a subset of the system modes, is satisfactory. Moreover, despite the use of black-box algorithms that do not exploit any physical knowledge on the system, the reconstructed modes admit a physical interpretation. The more complex models, that take into account an increasing number of system modes, are less satisfactory. Their responses, together with some physical insight on the system, allow to single out the model elements that are not correctly identified. In particular, this information may be of help in designing new identification experiments or devising new identification strategies for improving the model quality.

2. Experimental setup

In order to study the placement process, an experimental setup was made as depicted in Fig. 2. A scheme

Fig. 2. Photo of the experimental setup.

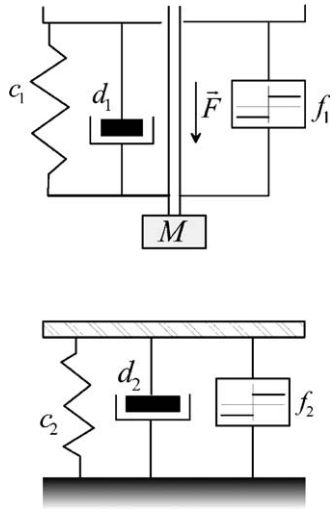

Fig. 3. Schematic representation of the experimental setup.

of the setup is presented in Fig. 3. The setup consists of the mounting head, from an actual pick-and-place machine, which is fixed above the impacting surface (the small disc in Fig. 2). The impacting surface is in contact with the ground via the spring (the spring c_2 in Fig. 3, within the outer tube in Fig. 2). The mechanical construction under the impacting surface is such that only the movement on the vertical axis is enabled (inner tube, which can slide inside the outer tube in Fig. 2). This construction exhibits linear and dry friction phenomena, represented in Fig. 3 by the damper d_2 and the block f_2, respectively. The chosen design of the impacting surface simulates the elasticity properties of the PCB as well as hard mechanical constraints due to saturations. It also introduces some side effects, such as dry friction.

The mounting head contains: a vacuum pipette which can move on the vertical axis (the mass M in Fig. 3) and which is connected via the spring to the casing (the spring c_1 in Fig. 3); an electrical motor which enables the movement (represented by force F in Fig. 3); and a position sensor, which measures the position of the pipette, relative to the upper retracted position. The position axis is pointed downward, i.e. the value of the position increases when the pipette moves downward. The motion of the pipette is also subject to friction phenomena (the damper d_1 and the dry friction block f_1 in Fig. 3).

The dynamics of the experimental setup exhibits, in a first approximation, four different modes of operation:

upper saturation: the pipette is in the upper retracted position (i.e. cannot move upward, due to the physical constraints);

free mode: the pipette is not in contact with the impacting surface, but is not in the upper saturation;

impact mode: the pipette is in contact with the impacting surface, but is not in lower saturation;

lower saturation: the pipette is in the lower extended position, (i.e. cannot move downward due to the physical constraints).

The control input is the voltage applied to the motor, which is converted up to a negligible time constant to the force F. The input signal for the identification experiment should be chosen in a way that all modes are sufficiently excited. However, conditions for the design of persistently exciting inputs are not available for hybrid models. To obtain the data for identification, the input signal $u(t)$ is chosen as

$$u(t) = a_k \quad \text{when } t \in [kT, (k+1)T), \tag{1}$$

where $T > 0$ is fixed, and the amplitude a_k is a random variable, with uniform distribution in the interval $[a, b]$. By properly choosing the boundaries of the interval $[a, b]$ only certain modes of the system are excited. For instance, only free and impact modes can be excited, without reaching upper and lower saturations.

Some features of the data sets obtained for the input signal (1) are shown in Fig. 4. In Fig. 4a the effect of dry friction damping can be noticed. In Fig. 4b small changes in the input signal produce no change in position, because of the dry friction in stick phase. In Fig. 4c, system is excited so that the lower saturation is reached. The lower saturation effectively acts as a velocity reset map, active when certain position is reached (≈ 25 in Figs. 4c and d). Fig. 4d shows the system behavior when both upper and lower saturations are reached. The bouncing effect can be observed when reaching upper saturation, due to the elastic impact with the mechanical constraints.

Waveforms in the experimental setup were sampled at 4 kHz. Control hardware in the pick-and-place machines enables sampling and control at much lower frequencies. Hence, models of the process at lower sampling frequencies are of interest. Data used for identification are obtained by re-sampling the original signals at 50 Hz. All plots in the paper have samples as units on the time axis. In all plots, the original input signals are multiplied by -100, in order to show them together with the system responses. However, original input signals are used in the identification experiments.

Having a physical representation of the setup, like the one depicted in Fig. 3, one would use a white-box modeling technique for identifying the value of the physical parameters (c_1, c_2, etc.). However, there are major difficulties that hamper this goal. In order to illustrate them, consider a simplified representation of the system. If saturations are not reached, dry friction is absent and the springs and dampers in Fig. 3 are linear, the equations relating the force F to the

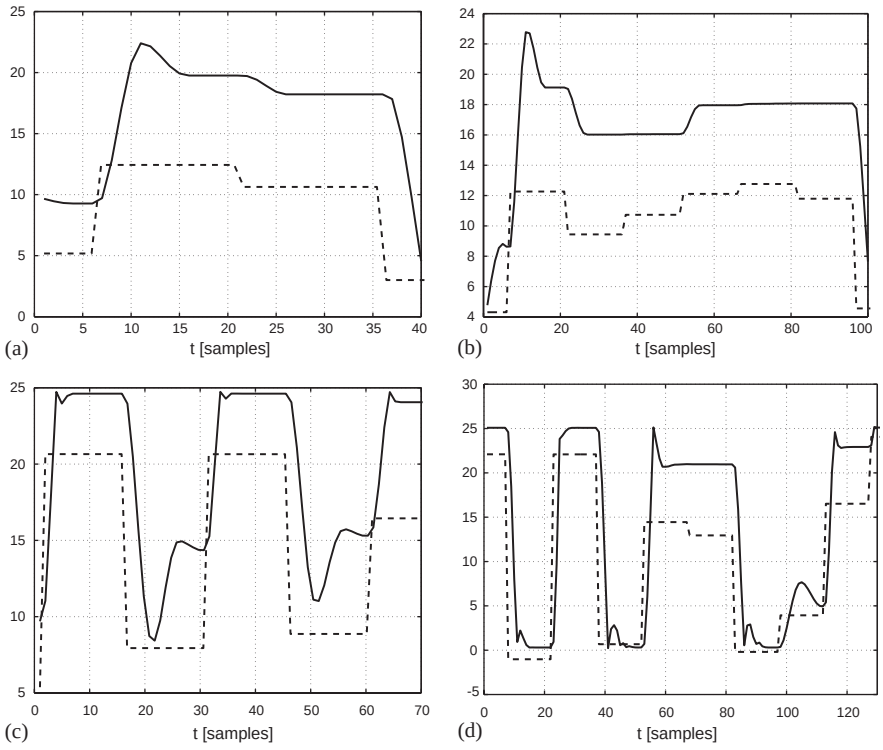

Fig. 4. Some features of the experimental data set: (a), (b) effects of the dry friction, (c) lower saturation, (d) upper saturation (solid: system response, dashed: scaled input).

position y of the head are

$$\ddot{y} = -\frac{c_1}{M}y - \frac{d_1}{M}\dot{y} + \frac{c_1\bar{y}_1}{M} + \frac{F}{M} \quad \text{(free mode)}, \tag{2a}$$

$$\ddot{y} = -\frac{c_1 + c_2}{M + m}y - \frac{d_1 + d_2}{M + m}\dot{y} + \frac{c_1\bar{y}_1 + c_2\bar{y}_2}{M + m} + \frac{F}{M + m} \quad \text{(impact mode)}, \tag{2b}$$

where $y = 0$ corresponds to the head in the upper retracted position, positive forces are directed downward, the springs c_1 and c_2 are at rest for $y = \bar{y}_1$, $\bar{y}_2$, respectively, and m is the mass of the impacting surface. Since data are available in discrete-time, one would like to sample system (2) and relate the parameters of the discrete-time model to those of model (2) and to the sampling time. Unfortunately, a general methodology for discretizing continuous-time hybrid systems is not available (Çamlibel, Heemels, & Schumacher, 2002), the key difficulty being to account, in discrete-time, for the effects produced by mode switches happening within the sampling interval. For this reason, one resorts to black-box identification by considering an ARX representation for each mode

$$y(k) = \theta_{1,1}y(k-1) + \cdots + \theta_{1,n_a}y(k-n_a) + \theta_{1,n_a+1}F(k-1) + \cdots + \theta_{1,n_a+n_b}F(k-n_b) + \theta_{1,n_a+n_b+1} \quad \text{(free mode)}, \tag{3a}$$

$$y(k) = \theta_{2,1}y(k-1) + \cdots + \theta_{2,n_a}y(k-n_a) + \theta_{2,n_a+1}F(k-1) + \cdots + \theta_{2,n_a+n_b}F(k-n_b) + \theta_{2,n_a+n_b+1} \quad \text{(impact mode)}. \tag{3b}$$

Even if the relations between the parameters in (3) and in (2) are unclear, note that both (3) and (2) are affine models in y and F. Moreover, if $n_a = 2$ and $n_b = 1$, Eqs. (3) can be obtained by sampling separately each mode of the original system.

Consider now the switching mechanism. Let $\tilde{y}(t)$ be the position of the impacting surface. In (2), the modes "free" and "impact" are active if $\tilde{y}(t) > y(t)$ and $\tilde{y}(t) = y(t)$, respectively. However, $\tilde{y}(t)$ is not measured, and it is not even a constant signal, because of the dynamics of the impacting surface. These facts have two major consequences on the switching rules for (3). First, it is impossible to associate a priori a mode of operation to each data point $(y(k), F(k))$. This lack of information constitutes the key difficulty of hybrid system identification (Ferrari-Trecate et al., 2003) and obliges one to adopt an identification algorithm that is capable to reconstruct all the modes at the same time. Second, the switching mechanism has to be reconstructed from y and F only. For these reasons, a black-box approach for the switch reconstruction is adopted, by assuming that modes (3a) and (3b) are active if the regressors lie in $\mathscr{X}_1$ and $\mathscr{X}_2$, respectively, where the sets $\mathscr{X}_1$ and $\mathscr{X}_2$ are

disjoint polyhedra that have to be estimated. Model (3), endowed with this switching law, define a PWARX system, whose identification is discussed in the next section. Obviously, differences in the true and reconstructed switching mechanism may completely destroy the analogy between (2) and (3) since each one of the reconstructed modes can possibly represent a mixture of the original modes. Then, the possibility of associating the estimated modes to the true ones, has to be assessed a posteriori.

Remark 1. An alternative way for reconstructing the system dynamics is to consider general nonlinear ARX models of the type $y(k) = f(y(k-1), \ldots, y(k-n_a), F(k-1), \ldots, F(k-n_b))$ that can be estimated by resorting to nonlinear identification procedures. For control, nonlinear models can be exploited in MPC schemes (Mayne, Rawlings, Rao, & Scokaert, 2000) where the optimal input is computed by solving a *nonlinear* programming problem. However, it is important to note that due to nonlinearities, convergence to the optimal solution is not guaranteed. This is in sharp contrast with MPC for PWA systems, that hinges on mixed-integer linear or quadratic optimization for which the optimal solution is always achieved (Bemporad & Morari, 1999).

3. Identification algorithm

In this section, the skeleton of the identification algorithm proposed by Ferrari-Trecate et al. (2003) is summarized.

A PWA map $f : \mathbb{X} \mapsto \mathbb{R}$ is defined by the equations

$$f(x) = f_q(x) \quad \text{if } x \in \bar{\mathscr{X}}_q, \tag{4}$$

$$f_q(x) = [x^\top \ \ 1]\bar{\theta}_q, \tag{5}$$

where $\mathbb{X} \subset \mathbb{R}^n$ is a bounded polyhedron, $\{\bar{\mathscr{X}}_q\}_{q=1}^s$ is a polyhedral partition of $\mathbb{X}$ in s regions and $\bar{\theta}_q \in \mathbb{R}^{n+1}$ are parameter vectors (PVs). Therefore, a PWA map is composed of s affine modes defined by the pairs $(\bar{\theta}_q, \bar{\mathscr{X}}_q)$. The data set $\mathscr{N}$ collects the samples $(x(k), y(k))$, $k = 1, \ldots, N$, generated according to the equation

$$y(k) = f(x(k)) + \eta(k), \tag{6}$$

where $\eta(\cdot)$ represents the measurement noise. It is assumed that the number s of modes is known. The aim of PWA regression is to estimate the PVs and the regions by using the information provided by $\mathscr{N}$.

When considering hybrid systems, an input/output description of a PWA system with inputs $u(k) \in \mathbb{R}^m$ and outputs $y(k) \in \mathbb{R}$ is provided by PWARX models that are defined by Eq. (6) where k is the time index and the vector of regressors $x(k)$ is given by

$$x(k) = [y(k-1)y(k-2)\ldots y(k-n_a)u^\top(k-1)u^\top \times (k-2)\ldots u^\top(k-n_b)]^\top.$$

It is apparent that, if the orders n_a and n_b are known, the identification of a PWARX model amounts to a PWA regression problem.

The essential hybrid feature of PWARX models is the ability to capture regressor-dependent mode switches. This switch mechanism is quite general. For instance, ARX models with input/output PWA static nonlinearities are PWARX models. Moreover, PWARX models provide an input–output description for a fairly large class of discrete-time PWA systems without logic states (Vidal et al., 2003).

The regression algorithm is structured in three steps.

(1) *Local regression*: For $j = 1, \ldots, N$, a local data set (LD) $\mathscr{C}_j$ is formed. It collects $(x(j), y(j))$ and the samples $(x, y) \in \mathscr{N}$ corresponding to the $c-1$ nearest neighbors x to $x(j)$. The cardinality c of an LD is a parameter of the algorithm satisfying $c > n+1$. LDs collecting only data points associated to a single mode are referred to as *pure* LDs. The remaining LDs are termed *mixed*. Linear regression is performed on each LD $\mathscr{C}_j$ to obtain the local parameter vectors (LPVs) θ_j and their empirical variance V_j. The LD centers $m_j = (1/c)\sum_{(x,y)\in\mathscr{C}_j} x$ are also computed together with the associated scatter matrix $Q_j = \sum_{(x,y)\in\mathscr{C}_j}(x-m_j)(x-m_j)^\top$. The information about the jth local model is collected in the feature vector (FV) $\xi_j = [\theta_j^\top, m_j^\top]^\top$. As for the LDs, FVs are either pure or mixed. FV ξ_j are interpreted as the realization of a Gaussian random variable with variance

$$R_j = \begin{bmatrix} V_j & 0 \\ 0 & Q_j \end{bmatrix}. \tag{7}$$

Intuitively, both V_j^{-1} and Q_j^{-1} are related to the confidence one should have on the fact that $\mathscr{C}_j$ is pure. In fact, in the noiseless case, $\|V_j^{-1}\|$ becomes infinite for pure LDs while remaining finite for mixed LDs. Analogously, if $\|Q_j^{-1}\|$ is "small" this indicates that the regressors in $\mathscr{C}_j$ are scattered and therefore the LD is likely to be mixed.

The key point is that pure FVs associated to the same mode are expected to be similar and then to form s distinct and dense clouds in the FV-space. Thus, as detailed below in steps 2 and 3, the problem of finding the data points associated to the same mode can be recast into the problem of finding s dense clouds of pure FVs. However, one should be warned about the presence of mixed FVs that do not carry any useful information on the true modes and form a pattern of isolated points in the FV-space.

(2) *Clustering*: The FVs are partitioned in s groups $\{\mathscr{D}_q^*\}_{q=1}^s$ through clustering. For this purpose, a

K-means algorithm (Duda & Hart, 1973) is used in order to minimize the cost functional

$$J(\{\mathscr{D}_q\}_{q=1}^s, \{\mu_q\}_{q=1}^s) = \sum_{i=1}^{s} \sum_{\xi_j \in \mathscr{D}_i} \|\xi_j - \mu_q\|^2_{R_j^{-1}}, \quad (8)$$

where $\{\mathscr{D}_q\}_{q=1}^s$ are the clusters and $\{\mu_q\}_{q=1}^s$ are the cluster centers. In principle, one would be able to collect all pure FVs characterizing the same mode in a single set $\mathscr{D}_q$. However, since mixed FVs are not a priori known, they will be assigned to some cluster as well. The best results are expected when they do not spoil the accuracy in clustering pure FVs. To this aim, we highlight that the confidence measures R_j^{-1} in (8) allow to assign less influence to mixed FVs than to pure FVs. Thus, the final clusters will mainly depend on pure FVs.

(3) *Estimation of the modes*: By using the bijective maps $(x(j), y(j)) \leftrightarrow \mathscr{C}_j \leftrightarrow \xi_j$, sets $\{\mathscr{F}_i\}_{q=1}^s$ of data points are built according to the rule: $(x(j), y(j)) \in \mathscr{F}_q \Leftrightarrow \xi_j \in \mathscr{D}_q^*$. This means that each set $\mathscr{F}_q$ collects data associated to similar local features. The data points in each final set $\mathscr{F}_q$ are then used for estimating the mode PVs through weighted least squares. The regions are reconstructed on the basis of the final sets by resorting to multicategory pattern recognition algorithms (Bredensteiner & Bennett, 1999) that find the hyperplanes separating $\{x : (x, y) \in \mathscr{F}_q\}$ and $\{x : (x, y) \in \mathscr{F}_{q'}\}$ for all indexes $q \neq q'$. This allows to find the matrix H_q and the vector h_q, for $q = 1, \ldots, s$, representing the polyhedra $\mathscr{X}_q$ in terms of the linear inequalities $H_q x \leqslant h_q$.

For the practical use of the algorithm, some potential pitfalls are highlighted. First, the method is expected to perform poorly if the ratio between the number of mixed and pure LDs is high. Note that the number of mixed LDs increases with c. Thus, it is desirable to keep c as small as possible. On the other hand, for high noise levels large values of c may be needed in order to filter the noise corrupting pure LPVs. Another reason for choosing large values of c is to "average out" small nonlinearities affecting the modes. This point is illustrated in Section 4.

Second, bad identification results can also be obtained because of the inherent sub-optimality of K-means (Duda & Hart, 1973) that is a computationally cheap but approximate method for minimizing the cost J in (8). In Ferrari-Trecate et al. (2003), some strategies to alleviate this problem are given.

Finally, data points suspected to be attributed to the wrong mode can be detected a posteriori through residuals analysis or via the procedure proposed by Ferrari-Trecate and Schinkel (2003). A discussion about how to exploit this information for improving the identification results is provided in Ferrari-Trecate and Schinkel (2003).

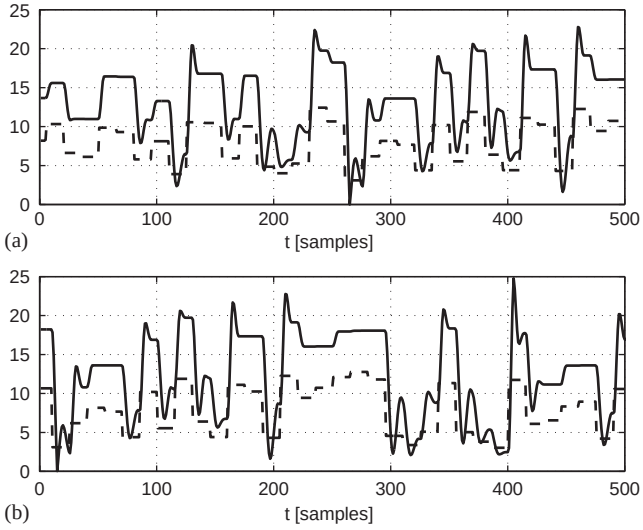

Fig. 5. Excitation of the impact and free modes. Data set used for (a) identification, (b) validation (solid: system response, dashed: scaled input).

4. Identification with free and impact modes

In order to study the applicability of the described identification procedure to our experimental setup, several data sets have been collected. They consist of 750 samples (15 s), divided in two overlapping sets of 500 points. The first set is used for identification, and the second for validation purposes. Note that 250 samples are used both for validation and identification. This allows to show, on the same picture, the fitting and the generalization properties of the reconstructed models.

In the first experiment, the parameters a and b characterizing the input signal are chosen so that only the free and impact modes are excited. The identification and validation data sets are depicted in Fig. 5. The effect of the dry friction in stick phase is clearly visible in Fig. 5b on the time interval $(200, 300)$.

PWARX models with two modes have been identified. For the first model the parameters $n_a = 2$, $n_b = 1$ and $c = 80$ have been used. One-step-ahead predictions on validation data are depicted in Fig. 6 together with the mode active at each time instant. The coefficients θ_q and the matrices H_q and h_q defining the regions are given in Fig. 9. The difference between the predicted and the actual system response is small and visible only on certain time intervals (e.g. $(200, 300)$). To what concerns the reconstructed modes, they capture the different physical modes of operation. In fact, from Fig. 6, one can observe that modes 1 and 2 are active when the system is likely to be in the impact and free modes, respectively.

Results of *simulation* with the identified model are shown in Fig. 7a. The response is still similar to the one of the real system, but now differences are visible.

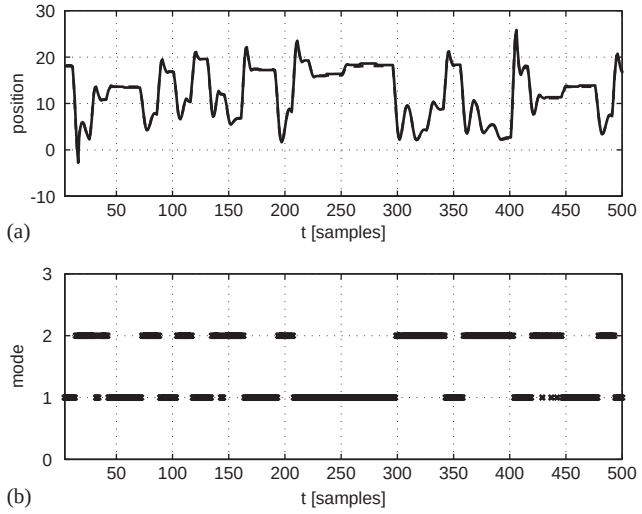

Fig. 6. Validation of the model with $n_a = 2$, $n_b = 1$, 2 modes, $c = 80$. (a) One-step-ahead prediction for validation data (solid: model prediction, dashed: system response). (b) Active mode at each time instant.

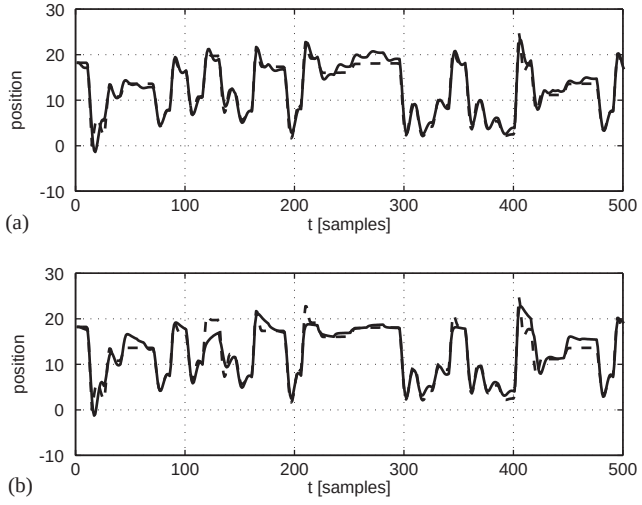

Fig. 7. Simulation of the bimodal PWARX systems: (a) model with $n_a = 2$, $n_b = 1$, $c = 80$, (b) model with $n_a = 2$, $n_b = 2$, $c = 80$ (solid: model output, dashed: system output).

A second model, characterized by parameters $n_a = 2$, $n_b = 2$, $c = 80$ is identified, and its simulated response is shown in Fig. 7b. Despite the increased order, the physical behavior is still not perfectly represented.

From the results shown in Fig. 7, it is clear that the differences are mainly due to the effects of the dry friction, which are not captured by the models. For instance, on the time interval (200, 300) a linear response is predicted to small step excitations, while in the physical system no movement occurs. The first model accurately simulates responses to large step excitations, while responses to small step excitations are not correct. The second model attempts to make a compromise between small and large step excitations—small step responses are better represented by the second model, but large step responses are worse.

The best bimodal PWARX model requires the use of LDs with large cardinality. For minimal theoretical values of c ($c = 4$, resp. $c = 5$) the obtained models are not usable, because the simulated output is completely dissimilar to the measured one. By increasing c, models of different quality are obtained, and the best results correspond to $c \geqslant 40$. It is interesting to note that even for large values of c (i.e. $c = 90$, for a data set of 500 points) good models can be still obtained.

A possible explanation is the following. Because of the presence of dry friction (see Figs. 4a and b) responses in both modes are nonlinear. Therefore, LDs with small c produce scattered LPVs, and the clustering step is not successful in separating FVs associated to the contact and the free modes. On the other hand, LDs collecting a large number of data points produce LPVs corresponding to "averaged" linear models. Such estimates produce distinct clusters in the feature vector space, but filters out the effect of the dry friction. The effect of "averaging" is noticeable in Fig. 7b, on time interval (200, 300), where a compromise is reached between responses to large and small step signals.

The previous discussion motivates the attempt to identify a PWARX model with more modes by using the same data set. The simulated response of the model with four modes, $n_a = 2$, $n_b = 2$, $c = 70$ is depicted in Fig. 8a, and the active modes during the simulation are shown in Fig. 8b. From Fig. 8b, one can observe that modes 2 and 4 correspond to the free and impact modes, respectively, while modes 1 and 3 represent the behavior on the boundaries between free and impact modes. The overall

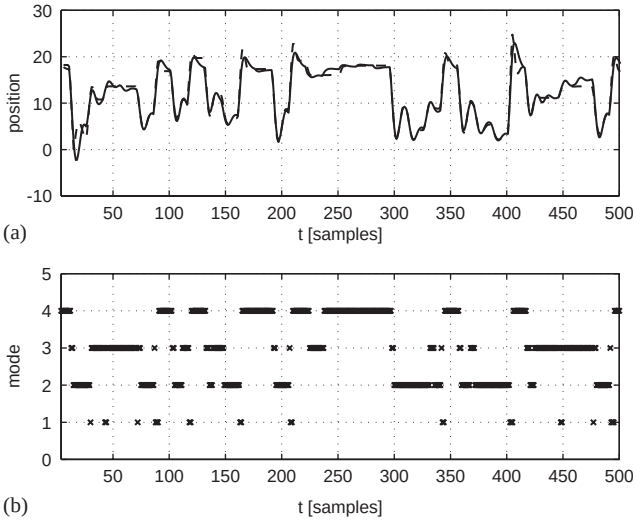

Fig. 8. (a) Simulation of the four-mode PWARX system $n_a = 2$, $n_b = 2$, $c = 70$ (solid: model output, dashed: system output). (b) Active mode at each time instant.

model performance in simulation is better in comparison to the bimodal case, at the price of increased model complexity. Identification with higher model orders and with more modes shows no significant improvements on the response quality.

Remark 2. For assessing the quality of results, Neural Networks (NN) are used for identifying the NARX model $y(k) = f(y(k-1), y(k-2), u(k-1))$. The best results were obtained by considering a 2-layer network with four neurons in the hidden layer having hyperbolic tangents as activation functions. The network was trained by using an *output error* algorithm exploiting weight decay and optimal brain surgeon strategy, for avoiding over-parametrization. The experiment was performed by using the Matlab toolbox written by Nørgaard (1997). The resulting NN has 18 parameters and the simulation results are analogous to those of Fig. 7a, i.e. the NARX model suffers from the same problems as the PWARX model with $n_a = 2$, $n_b = 1$ and $c = 80$. As discussed in Ferrari-Trecate and Muselli (2002), the high number of NN parameters is needed for approximating model discontinuities with a continuous NN.

5. Identification with saturations

In order to get a model which is valid in wider range of operating conditions, a new experiment is performed. The input was chosen so that impacts between the head and the spring occur and the lower saturation of the spring is reached. The data set was again divided into two parts (depicted in Fig. 10), where the first part is used for identification, and the other for validation.

In order to highlight the difficulties in reconstructing the lower saturation, consider again the scheme of Fig. 3. Lower saturation occurs at a fixed but unknown position y_s and resets the velocity to zero. In principle, one would enhance the models obtained in Section 4 with the additional dynamics $y(k) = y_s$, active when either

$$x(k-1) \in \mathscr{X}_c \quad \text{and} \quad y_c(k) \geqslant y_s \tag{9}$$

or

$$y(k-1) = y_s \quad \text{and} \quad F(k-1) \geqslant -\bar{F}(k-1), \tag{10}$$

where $\mathscr{X}_c$ is the region associated to the contact mode, $y_c(k)$ is the predicted position in contact mode and $\bar{F}(k-1)$ is the reaction force of the impacting surface when $y = y_s$. Condition (9) characterizes the switching between the contact and lower saturation, whereas (10) captures the requirements for staying in saturation at the instants $k-1$ and k. Note that it is assumed that when the saturation is reached, the impacting surface

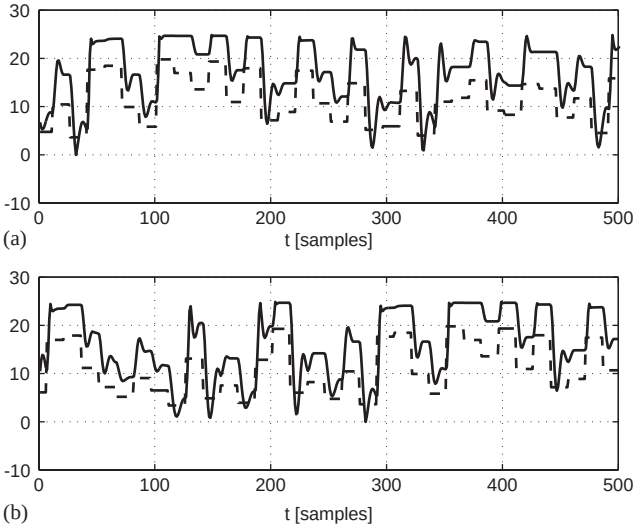

Fig. 10. Excitation of the free, impact and lower saturation modes. Data set used for (a) identification, (b) validation (solid: system response, dashed: scaled input).

Parameters	Model
$n_a = 2$, $n_b = 1$, 2 modes, $c = 80$	$\theta_1 = \begin{bmatrix} 1.4006 & -0.6299 & -34.4116 & 0.3031 \end{bmatrix}^\top$, $\theta_2 = \begin{bmatrix} 1.5872 & -0.7682 & -44.8641 & -0.7184 \end{bmatrix}^\top$, $H_1 = \begin{bmatrix} -1.0778 & 0.26430 & 5.7984 \\ 1.0 & 0 & 0 \\ 0 & -1.0 & 0 \\ 0 & 1.0 & 0 \\ 0 & 0 & -1.0 \\ 0 & 0 & 1.0 \end{bmatrix}$, $H_2 = \begin{bmatrix} 1.0778 & -0.26430 & -5.7984 \\ -1.0 & 0 & 0 \\ 0 & -1.0 & 0 \\ 0 & 1.0 & 0 \\ 0 & 0 & -1.0 \\ 0 & 0 & 1.0 \end{bmatrix}$, $h_1 = \begin{bmatrix} -10.282 & 30.0 & 30.0 & 30.0 & 0.2 & 0.2 \end{bmatrix}^\top$, $h_2 = \begin{bmatrix} 10.282 & 30.0 & 30.0 & 30.0 & 0.2 & 0.2 \end{bmatrix}^\top$

Fig. 9. Parameters of identified model.

remains in contact with the head, so neglecting bouncing phenomena of the impacting surface on the head.

Despite the fact that the above dynamics defines a PWARX model, it is not evident how to include it in the identification algorithm for several reasons. First, the region defined by (9) depends on the polyhedron $\mathscr{X}_c$ that has to be estimated by itself. The identification procedure of Section 3 does not allow to specify constraints between different regions. Second, the reaction force $\bar{F}(k)$ is not measured and it has to be deduced from the available signals. Third, for having $y(k) = y_s$ all entries of the mode parameter, except the displacement must be constrained to zero. Conceptually speaking, the introduction of a partially specified dynamics falls within the domain of *gray-box* identification, a topic that, to the authors' knowledge, has not yet been considered for hybrid systems.

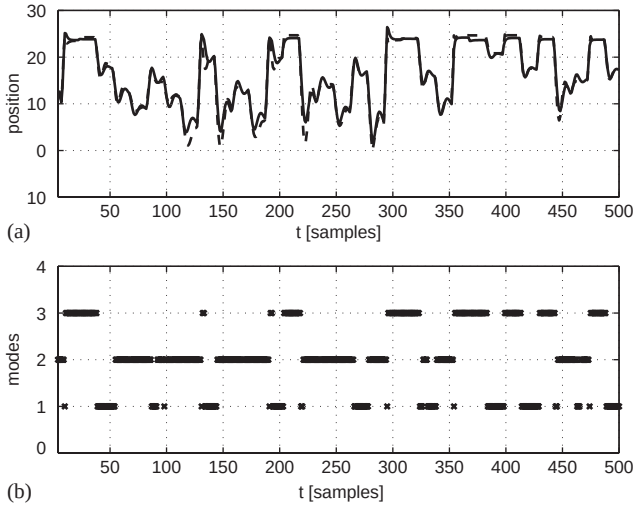

Fig. 11. (a) Simulation of the tri-modal PWARX system $n_a = 2$, $n_b = 2$, $c = 40$ (solid: model response, dashed: system response). (b) Active mode at each time instant.

Therefore, once more black-box identification approach is used and a general PWARX model with three modes and parameters $n_a = 2$, $n_b = 2$, $c = 40$ is identified. The response of the model in simulation is depicted in Fig. 11a, and the active modes are shown in Fig. 11b. The model parameters are reported in Fig 12. From Fig. 11, it is possible to observe that the third mode is triggered by the lower saturation. An analysis of the parameter vector θ_3 and the region $\mathscr{X}_3 = \{x : H_3 x \leqslant h_3\}$ reveals that the "lower saturation" mode can be interpreted as a mixture of the true saturation and contact (in proximity of saturation) modes. From the dynamics of the third mode, it is possible to compute the estimated saturation value y_s. Consider the case where

Parameters	Model
$n_a = 2$, $n_b = 2$, 3 modes, $c = 40$	$\theta_1 = \begin{bmatrix} 1.3112 & -0.60069 & -32.820 & -2.1941 & 1.2052 \end{bmatrix}^\top$, $\theta_2 = \begin{bmatrix} 1.4693 & -0.73980 & -36.712 & -8.1766 & 0.12148 \end{bmatrix}^\top$, $\theta_3 = \begin{bmatrix} 0.68555 & -0.097315 & -25.474 & 17.555 & 8.3912 \end{bmatrix}^\top$, $H_1 = \begin{bmatrix} -0.38975 & -0.13906 & -5.4530 & 16.011 \\ 4.3814 & -0.72624 & -40.376 & -52.533 \\ 0 & -1.0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 \\ 0 & 0 & -1.0 & 0 \\ 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & -1.0 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$, $H_2 = \begin{bmatrix} 0.38975 & 0.13906 & 5.4530 & -16.011 \\ 4.7711 & -0.58718 & -34.923 & -68.544 \\ 0 & -1.0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 \\ 0 & 0 & -1.0 & 0 \\ 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & -1.0 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$, $H_3 = \begin{bmatrix} -4.3814 & 0.72624 & 40.376 & 52.533 \\ -4.7711 & 0.58718 & 34.923 & 68.544 \\ 0 & -1.0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 \\ 0 & 0 & -1.0 & 0 \\ 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & -1.0 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$, $h_1 = \begin{bmatrix} -9.2035 & 98.496 & 30.0 & 30.0 & 0.2 & 0.2 & 0.2 & 0.2 \end{bmatrix}^\top$, $h_2 = \begin{bmatrix} 9.2035 & 107.70 & 30.0 & 30.0 & 0.2 & 0.2 & 0.2 & 0.2 \end{bmatrix}^\top$, $h_3 = \begin{bmatrix} -98.496 & -107.70 & 30.0 & 30.0 & 0.2 & 0.2 & 0.2 & 0.2 \end{bmatrix}^\top$

Fig. 12. Parameters of identified model.

the system is in saturation at time $k-2$ and $k-1$ and the constant force $F \sim u$ balances the reaction of the impacting surface. Then, the system will be in saturation also at time k. According to the ARX dynamics of mode 3, y_s and u must verify the linear constraints

$$y_s(1-\theta_{3,1}-\theta_{3,2}) = (\theta_{3,3}+\theta_{3,4})u + \theta_{3,5} \quad \text{and}$$
$$H_3[y_s y_s \ u \ u]^\top \leqslant h_3. \tag{11}$$

The minimal and maximal predicted values of saturation can be found by solving the linear programs $\min y_s$ and $\max y_s$, respectively, in the unknowns y_s and u subject to the constraints (11). We found $\min y_s = 23.2$, $\max y_s = 24.2$. Note that the latter value matches quite well the saturation level visible in Fig. 11a.

Remark 3. If the PWARX model has to be used for designing an MPC control scheme, a detailed model of the process of entering and leaving lower saturation is not necessary. In fact, by knowing only y_s, one can add to the MPC problem the constraints $y(k) \leqslant y_s$ and the resulting control law will take care of not driving the system in saturation. Therefore, for MPC, it is of primary importance to have a reliable description of the system in free and contact mode (as done in Section 4) and complement it with the estimate of y_s derived before, or better by conducting ad hoc experiments.

Modes 1 and 2 correspond to non-saturated modes in the experimental setup. The majority of the points associated to non-saturated behaviors is attributed to the second mode (this includes the points from both the impact and the free modes), while the points associated to the first mode mainly characterize the impact phenomena. The model output is more similar to the output of the physical system in parts of the data set that correspond to the saturated and impact behaviors.

In the last experiment, parameters of the input were chosen such that both upper and lower saturations are reached. The data sets used for identification and validation are represented in Fig. 13.

A PWARX model with four modes and parameters $n_a = 2$, $n_b = 3$, $c = 55$ is reconstructed. The corresponding response in simulation is shown in Fig. 14a, and the active modes are depicted in Fig. 14b. From Figs. 14a and b, one can conclude that modes 3 and 4 correspond to upper and lower saturations respectively. Modes 1 and 2 correspond to non-saturated behaviors. Moreover, almost all data points in non-saturated modes are attributed to mode 1, while mode 2 corresponds to large negative steps in the input signal, i.e. to the large transients of the output in the direction of the upper saturation.

A careful analysis of Fig. 14 reveals that:

- Responses in the portions of the data set associated to mode 1 are not satisfactory; one can conclude that the parameter vector of the mode 1 is poorly estimated.
- Responses in the portions of the data set associated with the fourth mode are satisfactory.
- By observing that the model response is good in the time instants when a switch from modes 1 to 3 and from 1 to 4 occurs, one can conclude that the boundaries between the modes 1/3 and 1/4 are correctly identified.
- The large undershoots (negative position values, which are not physically possible due to the mechanical constraints) in time instants after mode switch from 2 to 3 (e.g. time instants around 140 and 320)

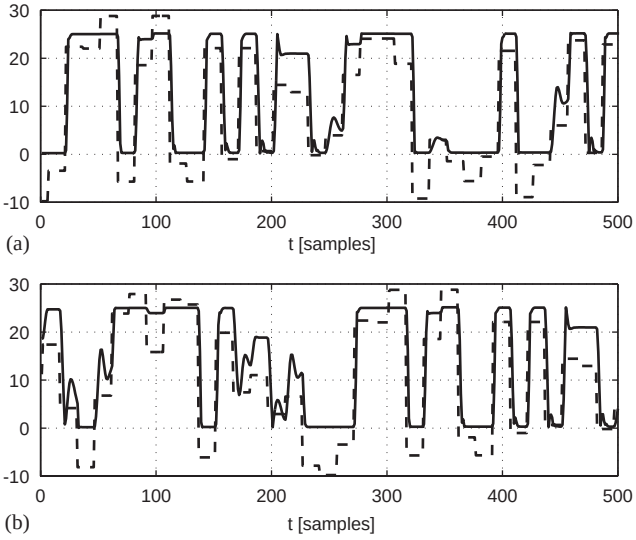

Fig. 13. Excitation of free, impact, lower saturation and upper saturation modes. Data used for (a) identification, (b) validation (solid: system output, dashed: scaled input).

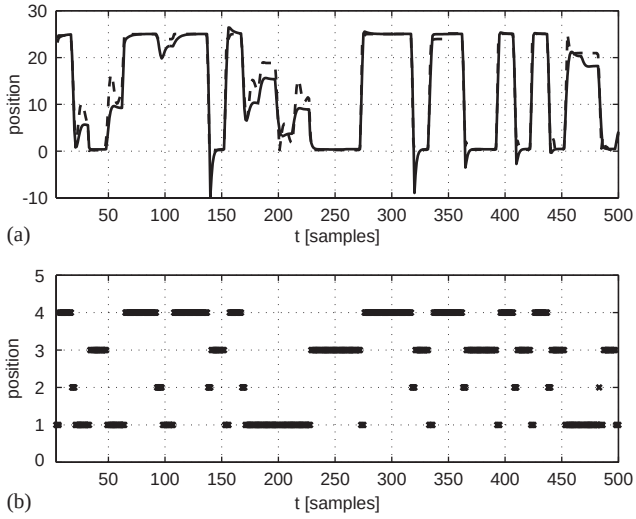

Fig. 14. (a) Simulation of the four-modal PWARX model $n_a = 2$, $n_b = 3$, $c = 55$ (solid: model response, dashed: system response). (b) Active mode at each time instant.

and the incorrectly predicted response around the time 100 (mode change between 2 and 4), allows to conclude that neither the parameter vector in mode 2, nor the boundaries between modes 2/3 and 2/4 are correctly identified.

- The simulated response is correct in the periods after mode switches between 1 and 3 (e.g. time interval around 230). Therefore, the PV of the third mode is acceptable. Large undershoots in mode 3, after the mode changes from 2 to 3 (e.g. time instants around 140 and 320) are due to the incorrect estimation of the mode 2.

Based on physical insights one would expect that the data points are classified in the four modes, corresponding to upper saturation, lower saturation, free mode and impact mode. The previous remarks highlight that the modes corresponding to the saturations are well reconstructed, whereas free and impact modes are not distinguished. From Fig. 13a, one can see that the data set used for identification contains a small number of points corresponding to free and impact modes, compared to the number of points representing saturations. This suggests that in order to improve the results, new identification experiment should be conducted to achieve a better excitation of the non-saturated modes.

6. Conclusion

In this paper, an experimental study in the identification of the electronic component placement process in pick-and-place machines is presented. An experimental setup was made, using the component placement head from the actual pick-and-place machine and the impacting surface, designed so as to mimic the elasticity properties of the real PCB. PWARX models of the process have been obtained by using the methodology introduced in Ferrari-Trecate et al. (2003).

Despite the fact that the models are optimized for one-step-ahead prediction their performance in simulation is satisfactory, when the free and contact modes are excited. Thus, models are adequate also for model-predictive control and verification. When attempting to identify also saturations, the quality of the resulting models deteriorates and limitations of the identification algorithms yielding to unsatisfactory behaviors are highlighted.

By combining physical insights on the process operation with the model responses it is possible to single out model features (i.e. mode parameters and/or regions) which are incorrectly reconstructed. This information may be used in order to design targeted identification experiments that allow to refine the results. Moreover, an accurate analysis of the physical behaviors in saturation conditions provides a relevant amount of a priori information, neglected by the black-box procedure adopted. Better results could be achieved by devising a *gray-box* identification procedure capable to exploit such knowledge.

The possibility of exciting an increasing number of modes with different experiments suggests also to study *incremental* identification algorithms. The basic idea would be to reconstruct first the modes visible in the simpler experiments, and then enhance the model with additional behaviors appearing in richer data sets, by keeping fixed the dynamics already identified. This idea will be developed in future research.

Acknowledgements

Authors would like to thank Ben Smeets from Assembleon, Eindhoven for providing experimental setup and technical help.

References

Assembleon, (2002). Web site: www.assembleon.com.

Bemporad, A., Garulli, A., Paoletti, S., & Vicino, A. (2003). A greedy approach to identification of piecewise affine models. In O. Maler & A. Pnueli (Eds.), *Hybrid systems: Computation and control*, *Lecture notes in computer science*, Vol. 2623, Prague, Czech Republic (pp. 97–112). Berlin: Springer.

Bemporad, A., & Morari, M. (1999). Control of systems integrating logic, dynamics and constraints. *Automatica*, *35*(3), 407–427.

Bemporad, A., Roll, J., & Ljung, L. (2000a). Identification of hybrid systems via mixed-integer programming. Technical Report AUT00-28, ETH Zurich, http://control.ethz.ch.

Bemporad, A., Torrisi, F., & Morari, M. (2000b). Optimization-based verification and stability characterization of piecewise affine and hybrid systems. In N. Lynch & B. Krogh, (Eds.) *Proceedings of the third international workshop on hybrid systems*, Pittsburgh, PA, USA. *Lecture notes in computer science*, Vol. 1790 (pp. 45–58). Berlin: Springer.

Bredensteiner, E. J., & Bennett, K. P. (1999). Multicategory classification by support vector machines. *Computational Optimizations and Applications*, *12*, 53–79.

Çamlibel, M., Heemels, W., & Schumacher, J. (2002). Consistency of a time-stepping method for a class of piecewise-linear networks. *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, *49*(3), 349–357.

Duda, R., & Hart, P. (1973). *Pattern classification and scene analysis*. New York: Wiley.

Ferrari-Trecate, G., & Muselli, M. (2002). A new learning method for piecewise linear regression. In *Artificial neural networks—ICANN 2002: International conference*, Madrid, Spain. *Lecture notes in computer science*, Vol. 2415 (pp. 444–449). New York: Springer.

Ferrari-Trecate, G., Muselli, M., Liberati, D., & Morari, M. (2003). A clustering technique for the identification of piecewise affine systems. *Automatica*, *39*(2), 205–217.

Ferrari-Trecate, G., & Schinkel, M. (2003). Conditions of optimal classification for piecewise affine regression. In O. Maler & A. Pnueli (Eds.), *Proceedings of the sixth international workshop on hybrid systems: Computation and control*, *Lecture notes in computer science*, Vol. 2623 (pp. 188–202). New York: Springer.

Heemels, W., Schumacher, J., & Weiland, S. (2000). Linear complementarity systems. *SIAM Journal of Applied Mathematics*, *60*(4), 1234–1269.

Juloski, A., Heemels, W., & Ferrari-Trecate, G. (2003). Identification of an experimental hybrid system. In *Proceedings of IFAC conference on analysis and design of hybrid systems*, Saint Malo, France.

Ljung, L. (1999). *System identification—theory for the user*. Englewood Cliffs, NJ: Prentice-Hall.

Mayne, D., Rawlings, J., Rao, C., & Scokaert, P. (2000). Constrained model predictive control: Stability and optimality. *Automatica*, *36*, 789–814.

Nørgaard, M. (1997). Neural networks based system identification toolbox. Technical Report 97-E-851, Technical University of Denmark, http://www.iau.dtu.dk/research/control/nnsysid.html.

Sontag, E. D. (1981). Nonlinear regulation: The piecewise linear approach. *IEEE Transactions on Automatic Control*, *26*(2), 346–358.

Vidal, R., Soatto, S., Ma, Y., & Sastry, S. (2003). An algebraic geometric approach to the identification of a class of linear hybrid systems. In *Proceedings of 42nd IEEE conference on decision and control*, Maui, HI, USA (pp. 167–172).