LOW-ORDER DYNAMICS IN A LATTICE MODEL OF THIN FILM DEPOSITION, USING NONLINEAR PRINCIPAL COMPONENT ANALYSIS

Martha A. Gallivan *

* *School of Chemical and Biomolecular Engineering*
Georgia Institute of Technology, Atlanta, GA 30332
email: martha.gallivan@chbe.gatech.edu

Abstract: Linear and nonlinear principal component analysis is used to characterize the state space of a kinetic Monte Carlo simulation of thin film deposition. The film's surface is first characterized using spatial correlation functions. This high-dimensional representation is reduced using a combination of linear and nonlinear projection. When nonlinear projection is used, the dynamics of the training and test data can be captured within 2% and 7%, respectively, using three dimensions. In constrast, a three-dimensional linear reduction does not adequately describe the relationship between the training and test data.

Keywords: Nonlinear principal component analysis, low-order dynamics, thin film deposition, kinetic Monte Carlo

1. INTRODUCTION

To manufacture faster computers, device size is approaching the atomic scale, and to create new materials, structure at the nanometer scale is being designed and exploited. In both cases, the manufacturing dynamics are defined by the interactions among many atoms. To optimize and control such processes, dynamic models are needed, which may take the form of molecular dynamics or Monte Carlo simulations (Nieminen, 2002; Starrost and Carter, 2002). The focus of this paper is on the fabrication of ultra-thin films, using lattice Monte Carlo simulations to describe the processing dynamics. Time-varying processes histories are sometimes beneficial in depositing very thin films (Markov *et al.*, 1991; Rosefeld *et al.*, 1995), but they are difficult to optimize due to the extremely high dimension of the dynamic models.

Model reduction provides a framework to reduce the internal dimension of these models, while preserving the map between inputs and outputs of interest. Analytical approaches to reduction have been developed for linear systems (Moore, 1981), but when the system is nonlinear, a data-driven approach is more often used. Snapshots of data are obtained from experiment or simulation, and are then reduced, yielding either a linear subspace via principal component analysis (PCA) (Holmes *et al.*, 1996), or a nonlinear manifold using nonlinear principal component analysis (NLPCA). The former method is computationally efficient, but when the relationship among the states is nonlinear, may not yield a small state dimension. Nonlinear PCA provides an alternative to linear PCA, and may be accomplished through the method of principal curves, neural net training, or a combination of both methods (Bolton *et al.*, 2003; Dong and McAvoy, 1996; Kramer, 1991). Recent application of NLPCA to high-dimensional process dynamics includes the reduction of a polymerization reaction (McLain and Henson, 2000).

Recent research has focused on the reduction of process models for thin film deposition. Straightforward application of existing methods is possible for continuum models of reacting flow (Banks *et al.*, 2002), while the surface evolution may require a non-continuum model. Principal component analysis has been applied to the surface height profile in lattice Monte Carlo simulations (Gallivan *et al.*, 2000; Raimondeau and Vlachos, 2000), but does not provide a predictive model in part due to the stochastic nature of the simulations. An alternative approach to reduction directly addresses this stochastic feature, by describing the evolution of a probability distribution (Gallivan and Murray, 2004), but requires that the set of "typical" surfaces be known. In this work, a systematic reduction method is applied to a Monte Carlo simulation of film growth, using a combination of PCA and NLPCA to identify a reduced set of state variables. This reduced state space could then be used for system identification. Section 2 describes the original high-dimensional model. The reduction method is described in Section 3, and the results of the Monte Carlo reduction are presented in Section 4. Sections 5 and 6 provide a discussion of the method and results, as well as directions for future research.

2. MODEL AND SIMULATIONS

In this work, the physical model for the processing dynamics is a lattice model. Atoms may only take spatial positions on this rigid lattice and, in this work, a cubic lattice is used. Atoms are allowed to take positions in the lattice from the gas phase, or to hop from one lattice site to an adjacent one. The rates of these transitions and there temperature dependence reflect the evolution of germanium films, as described in a previous work (Gallivan and Atwater, 2004).

The lattice model describes the evolution of a probability distribution for the occupancy of the entire lattice. However, due to its extremely high dimension, individual stochastic realizations are instead performed using the kinetic Monte Carlo (KMC) method. While Monte Carlo simulations typically describe equilibrium properties, kinetic Monte Carlo simulations capture the correct evolution in time (Bortz *et al.*, 1975). Periodic boundary conditions are used to eliminate edge effects, since the actual growth surface is much larger than the simulation domain. A KMC simulation of a 300×300 site surface is shown in Figure 1. Only a portion of the full simulation domain is shown so that the individual atoms can be resolved. Dark-colored atoms are on the edge of an atomic-height step, while light atoms have four side neighbors and comprise atomically-smooth terraces.

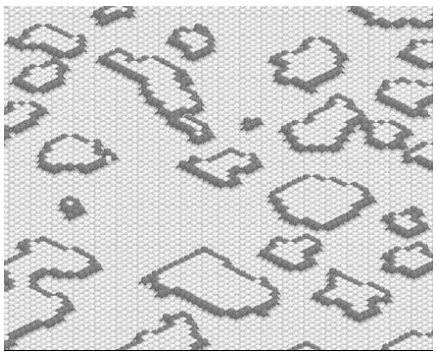

Fig. 1. Picture of a KMC simulation.

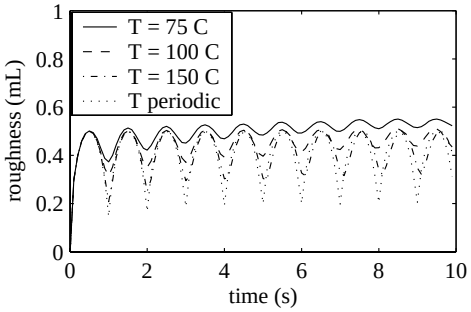

Fig. 2. Step density oscillations during epitaxial growth.

The model problem considered in this work is the description of the surface height profile, or surface morphology. As atoms adsorb onto the surface from the surrounding gas, clusters first nucleate on the surface, grow, and then coalesesce. This oscillatory behavior is illustrated by Figure 2, in which the time to deposit one monolayer (mL) of material is 1 s. When the temperature is high, the hopping rate of atoms increases, leading to fewer clusters, and a smoother surface. The oscillatory behavior decays faster at lower temperatures, because clusters nucleate on the second layer prior to full coalescence of the first layer. Periodic growth temperature can be used to achieve a smoother surface than any constant temperature in this range, as shown in Figure 2. The temperature is lowered to 75 C during the first 20% of each layer, and then raised to 150 C during the remaining 80%. As a result, many smaller clusters nucleate during the beginning of the layer, after which temperature is raised to fill in the gaps between clusters (Markov *et al.*, 1991; Rosefeld *et al.*, 1995). The goal of this paper is to generate a reduced set of coordinates that describes the dynamics of both constant and time-varying growth temperature.

While each KMC simulation generates a surface height profile, the location of individual clusters on the surface changes from realization to realization because the simulations are stochastic. Thus, alternate measures are needed to characterize each realization, and to compare simulations generated

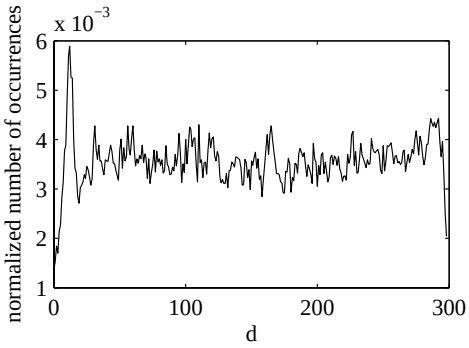

Fig. 3. Step-step correlation function (up-down) for 0.5 mL at 150 C.

under the same or different process conditions. One simple solution is to perform a Fourier transform of each surface to identify key length scales associated with cluster size and spacing. Alternatively, spatial correlation functions be used. In fact, the height-height correlation function can be computed directly from the Fourier transform (Tong and Williams, 1994). In this work the spatial correlation function for surface steps is used instead, since the films are extremely thin and are best characterized by the location and spacing of atomic-height steps. Defining the surface height to be $h(x, y)$ for site (x, y), the presence of an "up" step in the x and y directions is computed as

$$s_{ux}(x,y) = 1 \quad \text{if } h(x+1,y) > h(x,y) \tag{1}$$
$$= 0 \quad \text{otherwise} \tag{2}$$
$$s_{uy}(x,y) = 1 \quad \text{if } h(x,y+1) > h(x,y) \tag{3}$$
$$= 0 \quad \text{otherwise.} \tag{4}$$

"Down" steps are computed similarly, by reversing the inequalities. The step-step correlation function then is defined to be

$$ssc_{kl}(r) = \frac{1}{n_x\ n_y} \times \sum_{i=1}^{n_x}\sum_{j=1}^{n_y}\{s_{kx}(x_i,y_j)s_{lx}(x_i+r,y_j) + s_{kx}(x_i,y_j)s_{lx}(x_i,y_j+r)\}, \tag{5}$$

where k and l may each take values of u or d to denote the up and down directions, respectively. Due to the discrete nature of the simulations, x, y, and r are positive integers. An example of a step-step correlation function is shown in Figure 3. The peak in the correlation function represents the typical cluster size, since a cluster is composed of an up-step followed by a down-step. To characterize a surface during a KMC simulation, all four step-step correlation functions are computed, and thus represent the full state of the system. In the case of the 300×300 ($n_x = n_y = 300$) surfaces used here, the state dimension is thus 1200. The underlying physical assumption is that the evolution of two surfaces with the same initial step-step correlation function is the same, as measured by the step-step correlation function, and is justified based on physical intuition only.

3. METHOD FOR STATE SPACE REDUCTION

The method used here for reduction is analogous to other data-driven methods for state-space reduction (Holmes *et al.*, 1996; McLain and Henson, 2000). Snapshots are collected from a high-dimensional simulation, and are then projected onto a space of reduced dimension. The four steps are outlined and briefly described:

- Step 1: Collect snapshots.
  The snapshots of the surface height profile are collected at regular intervals. These snapshots *must* represent the entire range of surfaces to be considered (which is true for any black-box system identification method).
- Step 2: Express in physically meaningful coordinates.
  Earlier attempts to reduce the state dimension of KMC simulations (Gallivan *et al.*, 2000; Raimondeau and Vlachos, 2000) focused on the surface height profile. This set of coordinates changes from realization to realization because of the spatial randomness of the simulations, and does not yield sufficiently predictive modes. Here, a spatial correlation function is used instead, which is sensitive to the distribution of features, but not to their spatial locations. This provides a set of physically meaningful (although high-dimensional) coordinates to classify the state of the system.
- Step 3: Perform linear projection.
  A linear principal component analysis is first performed on the matrix of snapshots to identify a reduced dimensional linear subspace. One could theoretically skip this step and perform the nonlinear projection on the full snapshot matrix, but in practice it is advantageous to perform a linear projection first. The singular value decomposition used for PCA can easily handle large data matrices, while the nonlinear optimization used in Step 4 is more computationally demanding.

  The linear modes determined by PCA are ordered in terms of the energy associated with each mode. However, it is not clear in this example that the highest energy mode is also the most important in predicting the evolution of the state. Consequently, the PCA modes are rescaled by their singular values, so that the projection *coefficients* for each mode have a typical value of order 1.

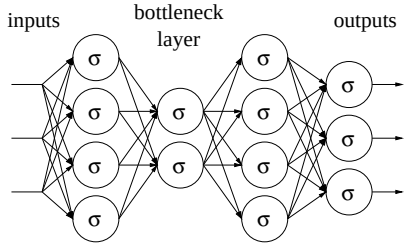

Fig. 4. Neural net with 3 inputs and outputs, 4 nodes in the mapping and demapping layers, and 2 nodes in the bottleneck layer. This net is used to reduce data from 3 dimensions to 2 dimensions.

These rescaled coefficients are then used as inputs for the nonlinear projection.

- Step 4: Perform nonlinear projection.

  The NLPCA method of Kramer (Kramer, 1991) is used to perform the nonlinear projection. This method relies on a neural net to describe the principal components, as illustrated in Figure 4. The net is trained to approximate the identity map for the data matrix, despite a hidden bottleneck layer with a reduced number of nodes. Once the net is successfully trained, the outputs of the bottleneck layer are the reduced set of coordinates.

4. RESULTS

The reduction approach presented in Section 3 is now applied to the model of Section 2. In the previous study of germanium film growth (Gallivan and Atwater, 2004), it was noted that the surfaces accessed during constant temperature growth also represented the typical surfaces seen under time-varying temperature. The benefit of the time-varying input was simply to access the smoother states after 10 s of growth, which could only be accessed under constant temperature at earlier times. Thus, constant temperature simulations were run at temperatures of T = 75, 87, 100, 125, and 150 C for 10 s of growth. They were sampled at intervals of 0.1 s, yielding 500 snapshots.

After performing the KMC simulations and collecting the snapshots, the step-step correlation function of equation (5) was computed for each snapshot. All computations associated with the reduction were performed using Matlab. The data was collected into a 1200×500 training matrix and its singular value decomposition was computed to obtain the linear principal component analysis. The corresponding singular values are shown in Figure 2. The singular values quickly decay by two orders of magnitude, and then level off. Thus, the first six linear modes are used to represent the surface structure, while the remainder are neglected and assumed to be noise. This issue will

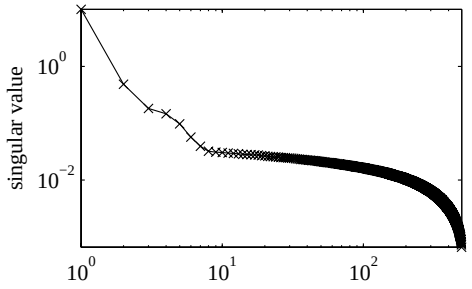

Fig. 5. Singular values of the training matrix.

be revisited in Section 5. The error associated with this truncation is quantified as

$$e = \frac{\sum_{i=1}^{n_s}\sum_{j=1}^{n_i}(y(i,j)-\hat{y}(i,j))^2}{\sum_{i=1}^{n_s}\sum_{j=1}^{n_i}y(i,j)^2}, \qquad (6)$$

where n_s is the number of snapshots, n_i is the number of elements in the data vector, $y(i,j)$ is the (i,j) element of a data matrix, and $\hat{y}(i,j)$ is a reconstruction of $y(i,j)$. This is the same metric used to assess the *energy* captured in PCA.

The first six linear modes retained here capture 0.9993 of the energy of the constant temperature data matrix, and also 0.9992 of the energy for the periodic input pictured in Figure 2. In the remainder of the paper, the former matrix will be referred to as the training matrix, and the latter as the test matrix. It should also be noted that the first three modes capture 0.9990 of the energy of the training matrix and the test matrix. However, it is not clear that the energy metric is the most appropriate one, since small energy features in the step-step correlation function could be critical in determinig its future evolution. Thus, before performing NLPCA, each of the six PCA modes is multiplied by its corresponding singular value, so that the projection coefficients for all six modes will be (typically) of order 1.

The nonlinear principal component analysis is performed using Matlab's Neural Net toolbox. The training matrix is used to compute a weight and a bias for a sigmoid function associated with each node, using a scaled conjugate gradient method and a maximum of 5000 iterations. Random initial guesses were used for the weights and biases. In all cases, 10 mapping and 10 demapping nodes were used for the first and third layers of the neural net, while the number of bottleneck nodes was varied. The number of inputs and outputs is six, corresponding to the number of linear principal components retained. The goal of the training is to obtain the identify map. If the identity can be well-approximated using a reduced number of bottleneck nodes, then the outputs of the bottleneck nodes may be taken as a reduced set of coordinates for the training matrix. Furthermore, if the error is also small for the test matrix, then

Table 1. Reconstruction error associated with linear and nonlinear PCA.

Data:	training		test	
Projection:	linear	nonlinear	linear	nonlinear
Dimension				
2	0.48	0.17	0.76	0.44
3	0.29	0.02	0.64	0.07
4	0.05	0.01	0.10	0.03
6	0	0.00	0	0.02

the net can be viewed as predictive and may be used to interpolate the reduced coordinates for intermediate data points not used in the training.

The error, as defined by equation (6), is shown in Table 1. The error for the two-dimensional NLPCA is high, but is greatly reduced when a third bottleneck node is added. Adding additional nodes improves the reconstruction, and should theoretically bring the error to zero when the number of bottleneck nodes equals the number of inputs (6 in this case). The error for the linear reconstruction is associated with a second PCA performed on the rescaled coefficients of the original PCA. Since each mode should contribute equally, it should not be expected that a reduced number would provide a good approximation.

Figure 6(a) shows the 3-dimensional NLPCA coordinates for the training and test data. Consistent with the small error, these coordinates reflect the physical relationship among the data points. The various temperatures are ordered from top to bottom, and the oscillatory behavior of Figure 2 is reflected in a circular arrangement of the data. As the oscillations decay, the surfaces spiral out from the center. Note that some of these cycles obscurred by the two-dimensional rendering of the data. This nonlinear projection also captures the test data, in which the system moves back and forth between the training data at the minumum and maximum temperatures.

The linear projection shown in Figure 6(b) corresponds to the errors reported in Table 1. The constant temperature training data is not clearly organized as with NLPCA, and the error is substantial (64%) for the test data. For comparison, a third plot is shown as Figure 6(c), representing the linear reconstruction of the data on the *original* three principal components, prior to the rescaling. The reconstruction does not clearly show a physically meaningful relationship among the snapshots. Perhaps more significantly, the test data does not interpolate among the training data, but instead is extrapolating out into a region not used for identification.

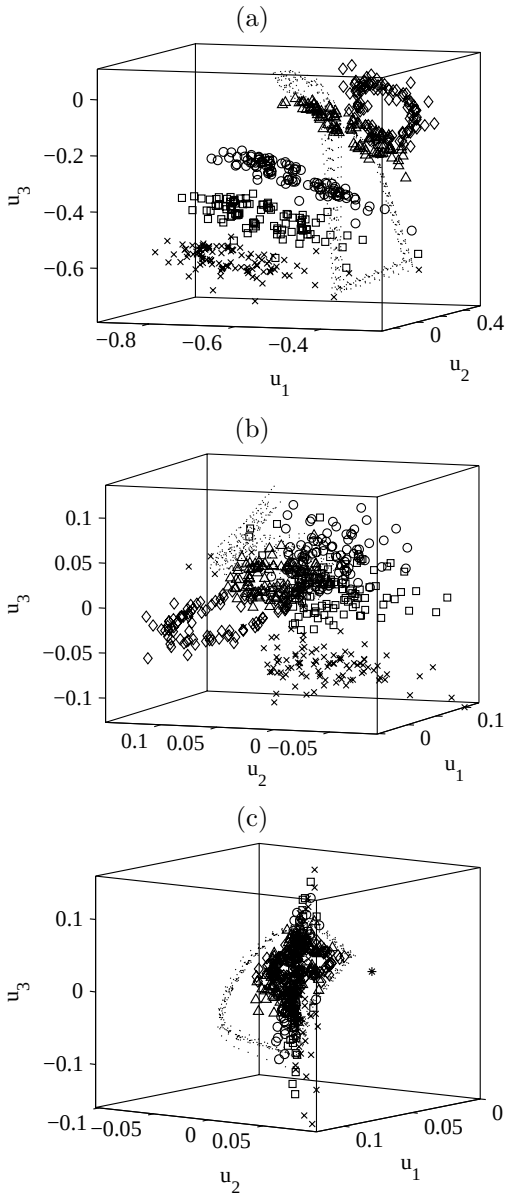

Fig. 6. 3-D projections (a) nonlinear, (b) linear with scaled coefficients, (c) linear with unscaled coefficients. The constant temperature snapshots used in the training matrix are marked by symbols—75 C (×), 87 C (square), 100 C (○), 125 C (△), and 150 C (◇). The evolution under the test data (periodic temperature) is represented by the dashed line.

5. DISCUSSION AND FUTURE WORK

Nonlinear projection captures the behavior observed in the KMC simulations, using a smaller set of coordinates than is needed for linear PCA. This reduction of the state space may lead to improved understanding of the dominant dynamics of thin film morphology evolution, and also could be used for system identification. For example, the reduced state space could be discretized, with each point representing an important group of surfaces with similar properties. The transitions between groups could then be computed using KMC simulation data (Gallivan and Murray, 2004). Other identification methods could also be applied to this low-order system.

This reduction method is largely automated, although some user input is required to select the number of modes in the linear projection. Open issues include the role of noise, and the importance of small energy modes which may be obscurred by noise. To reduce the noise, ensembles of KMC simulations could be collected. However, the noise can never be completely eliminated, so one will generally need to truncate modes from the PCA that have nonzero singular values. This issue remains a subject for future study.

6. CONCLUSIONS

A set of reduced-order coordinates was computed for a kinetic Monte Carlo simulation of thin film deposition. The film surface was characterized with a spatial correlation function describing the relationship between atomic-height surface steps. Principal component analysis was used to compute linear subspaces of reduced dimension for the correlation function, while nonlinear principal component analysis was accomplished using neural nets. A dimension of three provided a good approximation of the original dynamics when the nonlinear principal component analysis was used (2% error in training data, 7% in test data), while the dynamics were not adequately captured by a 3-dimensional *linear* projection. This preliminary study indicates that low-order behavior exists in the stochastic, high-dimensional Monte Carlo simulations, but it may not always be best captured using linear principal component analysis.

ACKNOWLEDGMENTS

The author thanks Georgia Tech for financial support, and I.G. Kevrekidis for fruitful discussions.

REFERENCES

Banks, H.. T., S. C. Beeler, G. M. Kepler and H. T. Tran (2002). Reduced order modeling and control of thin film growth in an HPCVD reactor. *SIAM Journal on Applied Mathematics* **62**, 1251–1280.

Bolton, R. J., D. J. Hand and A. R. Webb (2003). Projection techniques for nonlinear principal component analysis. *Statistics and Computing* **13**, 267–276.

Bortz, A. B., M. H. Kalos and J. L. Lebowitz (1975). A new algorithm for Monte Carlo simulations of Ising spin systems. *Journal of Computational Physics* **17**, 10–18.

Dong, D. and T. J. McAvoy (1996). Nonlinear principal component analysis—based on principal curves and neural networks. *Computers and Chemical Engineering* **20**, 65–78.

Gallivan, M. A. and H. A. Atwater (2004). Design of a film surface roughness-minimizing molecular beam epitaxy process by reduced-order modeling of epitaxial growth. *Journal of Applied Physics* **95**, 483–489.

Gallivan, M. A. and R. M. Murray (2004). Reduction and identification methods for markovian control systems, with application to thin film deposition. *International Journal of Robust and Nonlinear Control* **14**(2), 113–132.

Gallivan, M. A., R. M. Murray and D. G. Goodwin (2000). The dynamics of thin film growth: a modeling study. In: *CVD XV: Proceedings of the Fifteenth Symposium on Chemical Vapor Deposition* (M. D. Allendorf and M. L. Hitchman, Eds.). Vol. 616. Electrochemical Society. pp. 168–175.

Holmes, P., J. L. Lumley and G. Berkooz (1996). *Turbulence, Coherent Structures, Dynamical Systems, and Symmetry*. Cambridge University Press. Cambridge, UK.

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociate neural networks. *AIChE Journal* **37**, 233–243.

Markov, V. A., O. P. Pchelyakov, L. V. Sokolov, S. I. Stenin and S. Stoyanov (1991). Molecular beam epitaxy with synchronization of nucleation. *Surface Science* **250**, 229–234.

McLain, R. B. and M. A. Henson (2000). Principal component analysis for nonlinear model reference adaptive control. *Computers and Chemical Engineering* **2000**, 99–110.

Moore, B. C. (1981). Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Transactions on Automatic Control* **26**, 1732.

Nieminen, R. M. (2002). From atomistic simulation toward multiscale modeling of materials. *Journal of Physics: Condensed Matter* **14**, 2859–2876.

Raimondeau, S. and D. G. Vlachos (2000). Low-dimensional approximations of multiscale epitaxial growth models for microstructure control of materials. *Journal of Computational Physics* **160**, 564–576.

Rosefeld, G., N. N. Lipkin, W. Wulfhekel, J. Kliewer, K. Morgenstern, B. Poelsema and G. Comsa (1995). New concepts for controlled epitaxy. *Applied Physics A* **61**, 455–466.

Starrost, F. and E. A. Carter (2002). Modeling the full monty: Baring the nature of surfaces across time and space. *Surface Science* **500**, 323–346.

Tong, W. M. and R. S. Williams (1994). Kinetics of surface growth—phenomenology, scaling, and mechanisms of smoothening and roughening. *Annual Review of Physical Chemistry* **45**, 401–438.