The Pennsylvania State University

The Graduate School

Department of Chemical Engineering

OPTIMAL DESIGN AND OPERATION OF COMPLEX MATERIALS PROCESSING WITH APPLICATION TO MICROELECTRONICS

A Thesis in

Chemical Engineering

by

Christopher M. Behrens

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2009

The thesis of Christopher Behrens was reviewed and approved* by the following:

Antonios Armaou
Associate Professor of Chemical Engineering
Thesis Advisor

Themis Matsoukas
Professor of Chemical Engineering

Joan Redwing
Professor of Materials Science and Engineering

Andrew Zydney
Walter L. Robb Chair and Professor of Chemical Engineering
Head of the Department of Chemical Engineering

*Signatures are on file in the Graduate School

ABSTRACT

Spatially distributed multiscale systems have been recently developed to describe transport-reaction processes for phenomena which occur across length scales that differ by several orders of magnitude. The level of detail required to accurately describe the dynamic behavior of the system at smaller length scales cannot be provided by solely using a description from larger length-scales, whereas describing larger length-scale behavior using descriptions from smaller length-scale models would be infeasible due to the associated computational overhead required. Motivated by the above, multiscale models, which combine larger and smaller-length scale models that are intimately connected, have been developed.

Continuous pressure on profit margins has resulted in the desire to improve efficiency to reduce costs, resulting in the need for optimization procedures. Increases in computational efficiency and optimization strategies have allowed for optimization of complex materials processing problems requiring multiscale models to be pursued. Multiscale optimization of such processes, however, is a problem with three levels of hierarchy in computation. In the lower hierarchy, potentially computationally intensive simulations may need to be performed to model the properties of interest at both smaller and larger length-scales. If bi-directional information is required, iterative convergence at all scales may be required; this serves as the intermediate hierarchy. Finally, optimization (gradient-based and black-box) algorithms require several function evaluations, which in this case is at the top of the hierarchy. With the need to perform numerous simulations, solving a multiscale optimization problem could quickly become intractable. One promising idea to circumvent this computational intractability is to obtain the necessary information from linear interpolations whenever possible instead of using costly simulations. This is the basic idea behind *in situ* adaptive tabulation (ISAT). Initially used in solving combustion chemistry problems, ISAT has been subsequently extended to stochastic

systems and combined with black-box optimization for maximizing uniformity and minimizing surface roughness in a gallium nitride thin-film.

Implementation and further refinement of these methods is investigated by modeling the deposition of a thin-film consisting of alternating gallium arsenide and aluminum arsenide layers (GaAs/AlAs.) This example uses macroscale "inputs" from the reactor description to determine properties such as temperature, flow rates, and concentrations, and mesoscale kinetic Monte Carlo (kMC) simulations to measure the previously characterized interfacial properties of the film. The objectives of this problem are to minimize the interfacial step-densities between GaAs and AlAs layers, while also minimizing the temperature and the time spent in-between depositing species (termed annealing time,) as well the macroscopic objective of reducing spatial non-uniformity. This problem only involves adsorption and therefore only requires uni-directional flow of information from the macroscale level to the mesoscale level. Based on this model, we explore the methodology needed to accelerate the computational process by implementing ISAT for each optimization trial. Using this methodology, a multiscale model for a GaAs/AlAs thin-film deposition process was developed for optimization of the interfacial properties, and used within an efficient optimization framework to identify optimal steady-state process conditions for the process. Finally, by implementing this methodology using the Nelder-Mead and Hooke-Jeeves pattern search algorithms, we were able to identify a range of optimal process conditions for minimizing the time with no organometallic flux at the wafer surface and for no organometallic flux at the inlet. With our methodology, we did so while reducing computation time and balancing improvements in step density and thickness uniformity with preferred processing conditions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank several people for making this thesis possible: my thesis advisor, Professor Antonios Armaou of the chemical engineering department at Penn State, Julian Siano for his development of the kinetic Monte Carlo code used for this problem and the background for step density, the other members of the Armaou lab group, the other members of my thesis committee, my parents and other family members for their support throughout my degree, and the numerous classmates and friends that I made while being here at Penn State and elsewhere.

Chapter 1

Introduction

Microelectronics fabrication processes, used in techniques such as rapid thermal chemical vapor deposition (RTCVD), metalorganic vapor phase epitaxy (MOVPE, also known as metalorganic chemical vapor deposition or MOVPE), and plasma etching, amongst others, is an ongoing area of both experimental and computational research [1]. These are complex processes that involve reaction and transport phenomena at multiple length scales, ranging from on the order of nanometers and picometers (10^{-9} - 10^{-12} m) to centimeters and meters. In industrial production, competition will lead to reduced profit margins, which will also be compounded with strict quality control and environmental regulations. To successfully compete within this challenging and sometimes frustrating reality, attention must be paid to methodology used to improve production yield and product quality. As a result, there has been an increasing focus on optimal process design.

Several developments in optimization have corresponded with this increased emphasis. Significant improvements in computational efficiency have made the pursuit of complex modeling strategies feasible, resulting in refinement of multiscale modeling strategies. Also, since many solution techniques, particularly at small length scales, involve calculating expected values of random movements, improvements in stochastic optimization methods have made solving such problems possible. Such approaches are an area of ongoing study across many disciplines and in relation to a myriad of research topics. This work focuses on application of these methodologies to layered heterostructures, which are microelectronics devices that consist of multiple thin layers of at least two different materials alternately stacked in succession.

This work explores multiscale optimization of a layered heterostructure consisting of two materials – gallium arsenide (GaAs) and aluminum arsenide (AlAs.) While the potential applications of the material are interesting, our main goal is not about investigating the specific properties of GaAs and AlAs. Rather, our main objective is using GaAs and AlAs as an example to formulate, illustrate, and extend the methodology used in multiscale optimization of layered heterostructures. The overall problem has many challenges as it contains many seemingly separate components that are brought together by certain common conditions. We must consider the macroscale portion of the problem – designing the observable reactor that will produce the desired species at the desired location, and the associated reaction and transport mechanisms needed to explain its behavior. We also must consider microscale/mesoscale phenomena – how the surface of the thin-film evolves with deposition under certain conditions. We then must consider the conditions themselves, and which ones have the most appreciable effects on desired properties at both the macroscale and mesoscale levels, and determine the most suitable methodology for optimizing these properties. Finally, bringing all of these subcategories together can create a potentially computationally intractable problem, so we consider methodology in improving computational efficiency.

Chapter 2

Literature Overview

Processes used in industrial applications frequently contain phenomena that require mathematical descriptions that span length scales over several orders of magnitude. Such systems, termed multiscale in this work, can range from the smallest (quantum, atomistic, and molecular,) to the intermediate (kinetic Monte Carlo, Lattice-Boltzmann,) to the largest length scale regimes (continuum.) Mathematical modeling in one regime may not be suitable in another due to inadequate level of detail (larger regime descriptions for smaller regime systems) or infeasible computational demands (smaller regime descriptions for larger regime systems, e.g. atomistic simulations to describe CSTR dynamics [2].) In order to solve multiscale problems, the different size systems can be solved using techniques most suitable for their particular length scale, and then connected at an interface. Owing to increasing availability of computational power, significant advancements in the formation of descriptive multiscale models have been achieved in the past 10-15 years. These models have been used to describe such diverse systems as simulating fluid flows [3-4], moving contact line problems [5], crack propagation [6], and chemical vapor deposition of thin films [7-8], amongst others.

Driven by the need to improve profit margins and reduce costs, strategies directed towards process optimization have been developed. Such optimization can present numerous challenges for multiscale systems. Continuum equations for reaction-transport systems typically feature coupled systems of partial differential equations (PDEs.) Traditional practice has been to discretize these PDEs into a finite set of ordinary differential equations (ODEs) using finite difference or finite element approaches. Despite these challenges, optimization involving PDE continuum equation constraints has been achieved recently [9-12]. Additional improvements

involving reduced gradient approaches [13], reduced successive quadratic programming approaches [14], and control vector parameterization [15] have also been implemented to handle the large system of ODEs that develop from discretization methods. The optimization procedure for mesoscale kinetic Monte Carlo (kMC) systems is even more challenging due to the stochastic nature of such problems. Standard procedure has been to treat the mesoscale function evaluator as a black-box, and use pattern search algorithms to find convergence toward a solution. Traditional search algorithms such as Nelder-Mead and Hooke-Jeeves are commonly used for this procedure, although gradient-based algorithms are also possible, and additional methodologies [16-18] have been developed as well. Also complicating optimization of stochastic systems is function noise; recent algorithms for handling such systems include applications to polymerization reactors [19], flowsheet optimization [20], and supply chains [21], although addressing these challenges is still in its early stages. We maintain that such methodologies can be applied towards the optimal operation of microelectronics fabrication processes, such as MOVPE.

MOVPE is one technique commonly used to fabricate thin films for use in various microelectronic applications [22-23]. These films typically have thicknesses on the order of a few microns (roughly $10^2 – 10^3$ molecules.) To obtain maximum performance, mesoscale regime properties, such as the smoothness of thin-film surfaces and interfaces in multilayer films, may require optimization. Previous work in this group [24-26] has concentrated on multiscale optimization for a GaN process, in the realm of surface roughness control, and for catalytic oxidation of CO. In this work, these techniques are extended to layered heterostructure problems, which introduce numerous new challenges, including macroscale reaction control and determining suitable metrics for characterizing interfaces, in addition to the general increase in complexity due to multiple species being present.

Chapter 3

Problem Formulation and Background

In contrast to the previous chapter that primarily focused on general information about developments in multiscale modeling, optimization, and microelectronics fabrication, the emphasis of this chapter is providing an overview of the topics covered and procedures used in this work and the associated supporting literature. These topics will include model formulation, optimization formulation, methods to improve computational efficiency, and information about the specific problem, the fabrication of a layered heterostructure consisting of alternating layers of GaAs and AlAs.

3.1 Problem Formulation

We consider a spatially multiscale process combining macroscale components describing reaction-transport phenomena with mesoscale components that capture molecular properties at interfaces. Equations to describe the macroscale system are in the form of nonhomogeneous, second-order, non-linear PDEs with complex boundary conditions. The solution methodology for the description of the mesoscale system may be even more daunting, as a deterministic solution may not be attainable. The equations describing macroscale and mesoscale systems may be time-varying and have spatial dependence on multiple dimensions. Additionally, these systems are connected at an interface that is dependent on the inputs of both properties.

3.2 Macroscale Model

Reaction-transport systems are typically modeled using coupled PDE systems. For non-isothermal chemical reactions with mobile fluids, the well-known balances for energy, species continuity, and momentum are typically used (this particular formulation taken from the finite element solver package Comsol):

$$\nabla \cdot (-k\nabla T) = -\rho C_p \bar{u} \cdot \nabla T$$

$$\nabla \cdot (-D_i \nabla C_i) = R_i - \bar{u} \cdot \nabla C_i$$

$$\rho \bar{u} \cdot \bar{u} = \nabla \cdot [-PI + \mu(\nabla \bar{u} + (\nabla \bar{u})^T)]$$

where T is the temperature, C_i is the concentration of species i, u is the velocity vector, k is the thermal conductivity, ρ is the density, C_p is the heat capacity, D_i is the diffusivity for species i, R_i is the reaction rate for species i, μ is the dynamic viscosity, P is the pressure, and I is the identity matrix. A total number of m species are described. Standard procedure for solving these systems of equations is discretization of the PDEs through finite difference or finite element methods into a set of ODEs. Approximation through low-order methods, using Karhunen-Loève expansion (KLE, also known as proper orthogonal decomposition) via the method of snapshots, has also been used, particularly if the discretization procedure is a bottleneck in the multiscale simulation process [7, 24, 26-27]. We expect to have conditions specified (Dirichlet) at the inlet, no-slip (Dirichlet) and insulation (Neumann) conditions at the walls, and convective flux (Neumann) at the outlet. At the deposition surface, the conditions at the deposition surface are more complicated. Generally, we can assume no-slip (Dirichlet) conditions for velocity, temperature can either be constant (Dirichlet) or insulated (Neumann), and species concentration is described by the reaction rate (Robin.) Specifically, the reaction rate information described by Robin boundary conditions may be determined by mesoscale simulations or other approximations.

3.3 Mesoscale Model

Solutions to mesoscale systems are typically determined using methods such as molecular dynamics (MD) or kMC. Kinetic Monte Carlo methods provide a stochastic solution to the Master equation:

$$\frac{\partial P(\sigma,t)}{\partial t} = \sum_{\sigma'} W(\sigma',\sigma)P(\sigma',t) - W(\sigma,\sigma')P(\sigma,t)$$

where σ and σ' are system configurations, $P(\sigma,t)$ is the probability that the system exists in state σ at time t, and $W(\sigma,\sigma')$ represents the probability of a transition from state σ to state σ' [24, 28].

Events in kMC simulations have been shown to follow a Poisson process [28], and since processes such as deposition events can be modeled using a Poisson distribution, kMC is a workable methodology for modeling such systems. Mesoscale modeling techniques are especially preferred when information about the positions of molecules is required, yet little detail about the molecules themselves is needed.

3.4 Interface Characterization

Determining the most appropriate methodologies for interface characterization of layered heterostructures is not always as straightforward as it is for purely surface features. Several options for measuring roughness exist; two different metrics of roughness that are frequently employed include root-mean-square (RMS) roughness, derived from the autocorrelation function (ACF), and roughness derived from the height-density correlation function (HDCF.) The equations used for RMS and HDCF roughness are summarized below for a cubic solid-on-solid (SOS) model:

$$R_{RMS} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

$$R_{HDCF} = \frac{\sum_{i=1}^{n-1}\left|x_{i+1} - x_i\right|}{n-1}$$

where x is the height of any given lattice site, and n is the total number of sites [25].

An alternative method for determining interfacial smoothness is step density. This metric simply involves calculating the number of changes that occur between adjacent species, not being concerned with the average interfacial height or the quality of the change. Mathematically, step density S (when the azimuthal angle $\phi = 0$) is defined as the following for a cubic SOS model:

$$S = L^{-1}\sum_{i,j}(1 - \delta(h_{i,j}, h_{i+1,j}))$$

where δ is the Kronecker delta function, $h_{i,j}$ (where i and j represent the lattice position in length and width, respectively) is the height of lattice site (i,j), and L is the number of lattice sites or possible differences [29]. The main advantage of using step density over other metrics for layered heterostructure problems is due to the formation of isolated islands of material (termed "bubbles" here,) which occur when one or more molecules of one species become trapped within a layer of another species. Such layer mixing phenomena have been observed in layered heterostructure systems [30]. Bubble formation can make interface height unclear; however, with step density, only a change between adjacent species at the same height needs to be noted, resolving the bubble formulation complication. While rare in our simulations (about one out of every 2500 sites is a bubble,) they can still cause complications on large lattices.

Step density can be measured as either a spatial or temporal function. Spatial step density is calculated as a function of each layer l in the heterostructure upon completion of a simulation by calculating the number of changes in species between adjacent molecules. For a layer l far

from an interface, the step density S_l of that layer will most likely be zero. If layer l is near an interface, then S_l will take on a positive value between zero and two. (Spatial step density is calculated across both length and width.) The step density of an interface is the sum of the step densities of the individual layers that comprise the interface. For example, the interfacial step density of the lower interface in the one-dimensional example in Figure **3-1** (gray boxes over white boxes) is 0.6, while the middle interface (white boxes over gray boxes) has a step density of 0.4. In contrast, temporal step density is calculated from the number of changes between adjacent molecules of the film surface as a function of time. This metric can prove useful in determining the evolution of surface step density throughout the deposition process. Temporal step density can also be measured experimentally using reflection high-energy electron diffraction (RHEED) data for use in industrial applications [29]. It has the disadvantage, however, of being computationally expensive as its value must be calculated after every event. It is also important to note that temporal and spatial step density are related, since the realization of the next step (for spatial step density) is only dependent on its current state (for temporal step density,) i.e. it is a Markovian process. As a result, any trends that apply to the evolution of temporal step density would also apply to spatial step density.

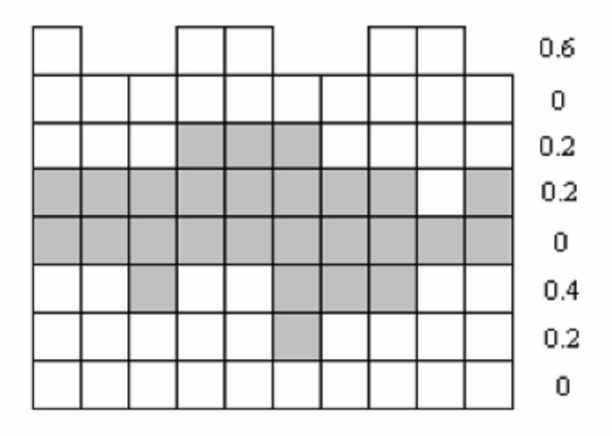

Figure **3-1**: Step densities assuming periodic conditions

3.5 Optimization Search Method

Optimization is an entire subject unto itself, with many different algorithms for possible use. In general, the main objective in optimization is to find extreme points (maxima or minima.) At these points, the following criteria are satisfied: the gradient (the vector of all first-order partial derivatives) at that point is zero, and the Hessian (the matrix of all second-order partial derivatives) is positive for minima and negative for maxima. For simplicity purposes, only minimization will be considered in this thesis. Most of this information is taken from [18].

Algorithms have developed around the idea of a progressive search to determine where such optimal criteria are met. One of the most straightforward is the method of steepest descent. This method uses first-order (gradient) information to determine a proper trajectory. The step size also can be controlled by a parameter λ – this value is often adjusted through various rules, but controlling it can be useful for maintaining the search within reasonable bounds. The update is as follows:

$$x_{i+1} = x_i - \lambda \nabla f(x_i)$$

where x_i is the current point and f represents the function of interest.

Including second-order (Hessian) information is desirable in order to improve the robustness of the search direction and step size. Calculating the Hessian directly, however, may incur considerable computational expense. As a result, methods have been developed that do not calculate the Hessian directly, but rather approximate it from first-order information and an initial guess (often the identity matrix.) The most commonly used method of this type was developed by Broyden, Fletcher, Goldfarb, and Shanno and is referred to as the BFGS algorithm. The BFGS update is the following set of expressions:

$$s = H^{-1}(x_i)\nabla f(x_i)$$

$$x_{i+1} = x_i + s$$

$$y = \nabla f(x_{i+1}) - \nabla f(x_i)$$

$$H(x_{i+1}) = H(x_i) + \frac{yy^T}{y^T s} - \frac{H(x_i)s(H(x_i)s)^T}{s^T H(x_i)s}$$

where *H* is the approximate Hessian and the other variables are the same as in steepest descent.

While gradient-based techniques may be employed, their use is complicated by the unavailability of gradient information in closed form for the mesoscale portion. This difficulty results in the need to calculate gradients numerically, introducing bias into the system and requiring additional computational time as simulations of perturbed systems also must be performed. Therefore, standard practice is to use pattern search algorithms while treating the multiscale system function evaluations as a "black-box." Two commonly used pattern search algorithms are the Nelder-Mead simplex method and the Hooke-Jeeves pattern search. In the Nelder-Mead approach, an initial guess is surrounded by a simplex (essentially a large array of points equidistant from a center that account for as many dimensions as the number of decision variables present.) The function is evaluated at each of the points, and then the point with the worst value is dropped. Based on the evaluations of each of the points, the simplex can be expanded or contracted. Upon subsequent replacements, once the simplex size is sufficiently small, the solution is considered converged. In the Hooke-Jeeves method, an initial point and an initial search pattern, termed a stencil, are established. The algorithm moves in various directions according to the search pattern, finds the best point then aggressively takes a new centered point two times the distance from the initial point. When the centered point becomes the best point, the size of the stencil is reduced and a new search begins; this continues until the stencil is sufficiently small and the solution is considered converged.

Optimization of multiscale systems can become quite complex, as they have macroscale components consisting of possibly time-dependent, nonhomogeneous, and possibly nonlinear second order PDEs, and mesoscale components of time-dependent processes that are most

efficiently modeled using stochastic techniques. Such methodology can quickly become computationally expensive, particularly as the complexity of information flow between directions increases.

3.6 Multiscale Solution Algorithm

As previously noted, spatially multiscale systems are typically initially separated into their different length scale components for ease of solution. These "solutions" are then intimately connected at interfaces that capture properties associated with all sizes in question. In some cases, information flow is unidirectional; this may occur when an input for the mesoscale system may require information from the solution to the macroscale problem, while the solution to the macroscale problem is independent of the results from the mesoscale system, or vice-versa. In this case, the interface calculation is straightforward, as only one evaluation of the multiscale system is required to obtain a solution. (Note that this does not imply that only one macroscale and one mesoscale simulation is required; rather, it specifies that only one evaluation of however many macroscale and mesoscale solutions required to obtain a multiscale solution is needed.) In contrast, bidirectional flow occurs when the solutions to the mesoscale system depend on solutions from the macroscale system, and the solutions to the macroscale system depend on solutions from the mesoscale system. This type of interface is considerably more complicated, since in order to converge to a solution that satisfies the system, solutions must be determined separately at each length scale, and then iteratively converged between the length scales at each time step.

Considerable research has been performed regarding the development of methodologies used in computationally efficient modeling of spatially or temporally multiscale systems. These include the use of timesteppers in equation free computing [31-33], filter application [34], coarse

graining of Monte Carlo simulations [35], using configurations with the greatest contributions [36], and lattice reconstruction techniques [25]; such topics are an ongoing area of study.

3.7 In Situ Adaptive Tabulation

In multiscale optimization, numerous mesoscale simulations, which may be computationally intensive themselves, may be required to obtain a single description of the multiscale system. Combining that with possible iterative convergence requirements and the need for numerous multiscale evaluations for optimization, it becomes apparent that this process can quickly become prohibitively expensive. While some problem specific simplifications can often be made, these usually do not reduce computation time by more than an order of magnitude.

As a result, approaches to reducing computation time have been developed. One such promising approach is known as in situ adaptive tabulation (ISAT.) This technique was first used in combustion chemistry [37] and has been subsequently applied to multiscale systems involving stochastic optimization. ISAT is particularly effective when either the number of required calculations is very large, or when simulations traverse similar regions in state space, as what occurs in optimization. The mathematical justification criteria for ISAT can be found in [37]. More detailed explanations of the ISAT algorithm can be found in [24, 26, 37]. The basics of the method will be presented here for completeness.

Consider a query point, ϕ_q. This query point contains all input properties that determine the dynamic evolution of the important solutions of the process of interest. Our goal is to determine results of interest at ϕ_q, which we would prefer to do by interpolation, but only if it satisfies the necessary requirements for accuracy. Thus, we would like to find a tabulated point, ϕ_0, within the vicinity of ϕ_q such that we can use the tabulated solution information from ϕ_0 to estimate the solution to ϕ_q. To achieve this, we must determine not only the trajectories, R_i, of

any tabulated point, ϕ_0, but also their sensitivities to each of the input variables j in the input matrix. This is mathematically described by the following relation:

$$A_{i,j}(\phi) \equiv \frac{\partial R_i(\phi)}{\partial \phi_j}$$

Any time that a simulation is required, the input properties ϕ_0, the solution information R_i, and the sensitivity matrix $A_{ij}(\phi)$ are tabulated. When another query point ϕ_q enters, it is compared with tabulated points and determined whether it is contained within a region that allows for linear interpolation; this metric of proximity between query and tabulated points is referred to as the ellipsoid of attraction (EOA.) The exact size of the EOA for any given tabulated point ϕ_0 is determined from the sensitivity matrix $A_{ij}(\phi)$ for that point, and a user-specified acceptable error tolerance ε_{tol}. To identify the EOA, define the square matrix A^*, where

$$A^* = \frac{AA^T}{\varepsilon_{tol}^2}$$

Next, perform eigenvalue decomposition of A^* and obtain a set of eigenvectors Q and eigenvalues Λ. If the distance $\delta = \phi_q - \phi_0$ falls within the EOA, given by the following relation:

$$\delta^T Q^T \Lambda Q \delta \leq 1$$

we can simply approximate R_q as:

$$R_q \approx R_0 + A(\phi_q - \phi_0)$$

If the query point does not fall within an EOA for any tabulated point, the process is simulated, the sensitivities are calculated, and a new record is added. The EOA can also be grown if the simulation results indicate that the interpolated solution is within an acceptable error tolerance.

ISAT has the advantage that the table is built as the simulations are performed, and therefore potentially memory intensive table pre-construction techniques are not applied. In order

to avoid extensive computation effort from table lookup, ISAT was initially implemented using a binary tree structure. This practice has traditionally been employed, but it is not necessarily a requirement. Mandating that query points are checked against all tabulated points may potentially reduce lookup error and the number of required table entries, therefore also reducing the computational cost, particularly when relatively few function evaluations are needed. Other improvements also have recently been investigated in table organization [38].

ISAT is an efficient technique primarily when large numbers of simulations are required. Implementation initially incurs additional computational cost be requiring multiple simulations for a single query. The savings occur when a sufficiently large number of queries can be interpolated, such as during the solution of an optimization problem. Another important note is that when using ISAT in stochastic systems, the exact same random number sequence must be used in order to assure that the sensitivity matrix is influenced as little as possible by stochastic noise. This technique, referred to as finite differences with common random numbers (FDCRN), can be outlined in [39]. This ensures that any variation in the solutions between the initial and a perturbed input is due to the perturbation itself and not stochastic noise [40]. While not completely ideal due to the possibility of sequence misalignments, FDCRN is straightforward, easily applicable, and gives increasingly similar answers taking the limit as the perturbation approaches zero. It should be noted that while FDCRN is necessary for determining the trajectories and sensitivities for a particular point, different random number sequences are used for different points; otherwise, the system would contain bias.

3.8 Application to Layered Heterostructure Growth

We maintain that the methodology described in previous sections can be applied towards layered heterostructure growth. While a number of different materials are in theory possible for

consideration, we chose to study the fabrication of a layered heterostructure consisting of alternating layers of gallium arsenide (GaAs) and aluminum arsenide (AlAs.) GaAs and AlAs have the nice property of having nearly identical lattice constants, thus minimizing film strain and defects [1, 41]. Their refractive indices, however, are different (about 3.3 for GaAs and 3.0 for AlAs grown by molecular beam epitaxy [42].) Because of this, such a device may prove useful in Bragg reflectors [43]. Bragg reflectors (usually referring to a distributed Bragg reflectors, or DBRs,) are essentially high-reflectivity mirrors that take advantage of Bragg scattering. Such devices are usually either fabricated using molecular beam epitaxy (MBE) or MOVPE. Both GaAs and AlAs naturally form Zincblende structures. If we only consider the metallic species to be mobile, this allows for use of a cubic solid-on-solid model with no overhangs [44].

For industrial production, the percentage of functional devices can be increased by minimizing the thickness non-uniformity and the interfacial roughness between layers. Thickness uniformity can be calculated macroscopically by solving continuum mass, energy, and momentum balances, but interfacial roughness must be measured at the mesoscale level in order to capture adequate detail.

Chapter 4

Problem Definition and Importance

The main thrust of this research is computationally efficient multiscale optimization with respect to microelectronics fabrication of layered heterostructures. By *computationally efficient*, we would like to develop methodology for problems to be solved accurately and elegantly, but also simply and quickly. Inaccurate or poorly supported methodology is of little use, but methodology that is difficult to understand or extremely time consuming will not provide much value to potential academic and industrial researchers either. This goal of computational efficiency is to be applied to *multiscale optimization*; covering the broad subjects of both "multiscale systems" and "optimization." Optimization will be discussed in the next paragraph, but first, it would not be feasible to develop a detailed procedure to handle all multiscale problems. At one extreme, some systems might have atomistic or quantum length and time scales; at the other, some might involve thousands of kilometers and millions of years. Diverse approaches are expected to be used in determining optimal operation of any multiscale system; our efforts are concentrated towards a specific region (macroscale and mesoscale) in multiscale problems. The specific multiscale application is to *microelectronics fabrication.* Numerous techniques exist within the large field of microelectronics to produce crystalline substrates; a few of them include standard evaporation, electron-beam evaporation, pulsed laser deposition, sputtering, molecular beam epitaxy, and chemical vapor deposition [1]. Each technique has specific macroscopic and microscopic formulations to consider, and while it is possible to create a skeleton where different formulations can be inserted, we chose to use MOVPE. Finally, we are interested in a specific application of microelectronics fabrication – *layered heterostructures*. Similar work has been done in thin-films with one species [24, 26]; however, introducing

multiple species extends the application of the problem to more complex structures, including materials processing with more than two species and types of interfaces.

Computational efficiency, multiscale systems, microelectronics fabrication, and layered heterostructures merely serve as adjectives to the purpose of this problem, which is optimization. Optimization stretches across many disciplines (manufacturing, scheduling, finance,) and covers many types of problems (distribution, traveling salesman, to name a few [45].) Optimization can be linear or nonlinear, continuous or discrete, constrained or unconstrained, single objective or multi-objective, and deterministic or stochastic. Specific solution strategies and methodologies exist for each class of optimization problems. The problem investigated in this work can be considered nonlinear, continuous, constrained, multi-objective, and stochastic, although some adaptations have been made to the constrained and multi-objective categories. Therefore, this work can be summarized as the development of accurate, reliable, well-formulated, easily understood, and relatively fast methods for solving nonlinear, continuous, constrained, multi-objective, and stochastic optimization problems for use in a spatially multiscale system (macroscale and mesoscale lengths) used to grow layered heterostructures from MOVPE.

This problem has a number of important applications. First, as computing power continues to increase, this work helps to further computational methods to model reaction-transport phenomena traditionally found in chemical engineering applications. Even as research moves into nanomaterials and nanostructures, many unanswered questions still exist in microelectronics applications. Research in thin-film production, both in experimental and computational realms, is being performed in order to improve deposition uniformity, interfacial uniformity, impurity incorporation, defects, and film strain, although the last three are not considered in this particular problem. Also, this problem looks at advancements in optimization, especially the challenges associated with including phenomena dependent upon stochastic simulations. As a whole, the motivation behind this problem is driven by production (improving

microelectronics fabrication) and economics (reducing costs and improving profit margins while meeting preset standards and regulations.) The specific motivation is the novel application to layered heterostructure problems.

Chapter 5

Methodology and Results

Previous chapters in the thesis have laid out background information, a review of the literature, and what the purpose behind this problem is. This chapter will establish the particular problem, outline the methodology used to solve the problem, present results, and discuss the meanings of these results. The sections in this chapter are organized approximately in chronological order of when specific tasks were performed, although some liberty is taken when previous background information is needed.

5.1 Mesoscale Modeling of Surface Phenomena

The first consideration in this model was the surface phenomena; much of this work was done previously in [46]. For this problem, we assumed that all gas-phase and surface reactions had already occurred, and were only interested in the properties that these species exhibited upon deposition to the wafer surface. We used a cubic solid-on-solid (SOS) model assuming perfect lattice match and no overhangs in our deposition. The broad model used was very flexible with inputs – a single temperature value was specified, but inputs for species (GaAs, AlAs, or neither – termed annealing), run time, deposition time, and a time period when deposition linear increases (pre-padding) or decreases (post-padding) from adjacent values in the input matrix were allowed to vary throughout the process. The methodology used to simulate deposition was kinetic Monte Carlo (kMC). In this model, two different types of events are possible: deposition, or a nearest-neighbor migration. The likelihood of a migration is based on the number of nearest neighbors a molecule has. Each molecule at lattice position $X_{i,j}$ and height k has twelve neighbors as outlined

in Table **5-1**. Values for the migration rates can be found in [44], and important associated parameters such as prefactors can be found in [47]. Note that [44] only considers metallic migration rates, reducing the problem from its naturally occurring Zincblende structure to a cubic SOS model. These migration rates increase with temperature.

Table **5-1**: Nearest neighbors model: i = x-direction, j = y-direction, k = z-direction

Current position	$X_{i,j,k}$
Adjacent (4)	$X_{i-1,j,k}$, $X_{i+1,j,k}$, $X_{i,j-1,k}$, $X_{i,j+1,k}$
Diagonal (4)	$X_{i-1,j-1,k}$, $X_{i-1,j+1,k}$, $X_{i+1,j-1,k}$, $X_{i+1,j+1,k}$
Adjacent and underneath (4)	$X_{i-1,j,k-1}$, $X_{i+1,j,k-1}$, $X_{i,j-1,k-1}$, $X_{i,j+1,k-1}$

In [46], it was found that step density decreases with increasing temperature and increasing annealing time, and increases (although less strongly) with increasing flux rates. This study involved a number of different cases of inputs to the matrix and lattice sizes. From standardizing the input matrix, we were able to observe a number of helpful trends that could be used to shorten the computation time of performing simulations. Temporal step density, for example, reaches a stationary state value quite rapidly during deposition of one species (usually within one or two monolayers.) Therefore, regardless of the actual thickness of each layer in a Bragg reflector, the average step density will be the same as long as the above requirement is met. As a result, simulating deposition for an extended length of time is not necessary in order to obtain the necessary mesoscale outputs, particularly since average thickness can be calculated macroscopically. Additionally, the step density of a GaAs layer or of an AlAs layer is independent of the number of previous layers that had been deposited. This means that the minimum number of layers required to accurately describe step density is three – a GaAs layer, an AlAs layer, then another GaAs layer. These trends can be illustrated in Figure **5-1**, where GaAs and AlAs processing conditions are kept constant (pre-padding and post-padding are included, however.) Note that because of this, we can consider temporal step density a Markovian process. Since the evolution of spatial step density is dependent on the starting

condition of temporal step density, if this starting condition is the same and all other properties (temperature, time) are kept the same, the spatial step density will be equivalent as well. These simplifications are not sufficient to make the multiscale problem computationally tractable, but they are certainly beneficial, reducing computation time up to an order of magnitude, and are definitely sufficient for the mesoscale problem alone.

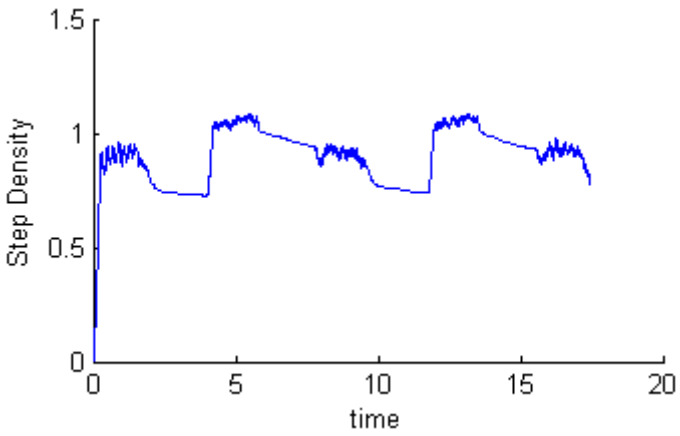

Figure **5-1**: Trends in temporal step density (size of lattice = 100x100)

5.2 Mesoscale Optimization and Results

From this, we were able to formulate a mesoscale optimization problem with the goal of low step densities. We also would like to have as low of a temperature as possible for energy considerations, and as low of an annealing time as possible to maximize production. These are competing objectives according to the findings in [46]. For our optimization problem, we considered the kMC simulations as a "black-box" and avoided the complication of calculating gradients by using the Nelder-Mead simplex method as the optimization algorithm of choice. As a preliminary investigation, since this problem does not have any linear interpolations that could potentially aid convergence, we chose to solve a deterministic cost function with a stochastic

bound. We considered a five-layer model, with two AlAs over GaAs interfaces and two GaAs over AlAs interfaces. Since GaAs over AlAs interfaces have a naturally higher step density than AlAs over GaAs interfaces do, it would not be particularly informative to restrict these interfaces to the same value. Instead, we required that the sum of the step densities of all interfaces be less than four, or a discontinuitous penalty was added. We also added a discontinuitous penalty if improper process conditions (i.e. annealing time less than zero) occurred. Deposition rates and times were kept constant (5 Monolayers/s deposition rate, 1 s deposition time of GaAs, 1.5 s for AlAs), as were pre-padding and post-padding; only temperature and annealing time (which was the same for all layers) were considered design variables. The specific optimization function used was the following:

$$\min 0.01T^2 + 150A^2 + P_1 + P_2$$

s.t.

$$P_1 = \begin{cases} 100{,}000 & \text{if } A < 0 \\ 0 & \text{if } A \geq 0 \end{cases}$$

$$P_2 = \begin{cases} 100{,}000 & \text{if } S_{Al/Ga} + S_{Ga/Al} > 4 \\ 0 & \text{if } S_{Al/Ga} + S_{Ga/Al} \leq 4 \end{cases}$$

where T is the temperature, A is the annealing time, S is step density, and P_1 and P_2 are penalties. Trials were simulated on a 100x100 square lattice, and optimization was performed at several starting conditions, each consisting of five realizations of the problem. Table **5-2** summarizes the optimization results. We observe that the predicted optimal temperatures and annealing times are consistent and reproducible at different starting conditions. Unfortunately, due to the nature of the problem, the starting conditions are required to be in the region where no penalties are levied. Total simulation time was within about 48 hours per optimization trial.

Table **5-2**: Mesoscale optimization results

Starting Temp. (K)	Starting Annealing Time (s)	Mean Optimal Temp. (K)	Mean Optimal Annealing Time (s)	Mean Function Value	Temp. Standard Dev. (K)	Annealing Time Standard Dev. (s)	Function Value Standard Dev.
800	5	767	0.48	5931	9.5	0.33	105
800	8	761	0.67	5866	2.7	0.13	24
800	10	764	0.59	5888	4.5	0.23	38
850	5	761	0.77	5887	4.5	0.24	24
900	5	762	0.69	5889	4.2	0.30	35
Mean		**763**	**0.64**	**5892**	**5.8**	**0.27**	**57**

5.3 Macroscale Reactor Model

Deposition processes have varying conditions across the deposition surface and throughout the reactor. As a result, we must consider macroscale modeling and operating conditions. This process uses a rotating disk reactor, with one inlet at the center of the reactor (pre-mixing is not a concern for the reactants.) A reactor schematic is shown in Figure **5-2**. The reactor dimensions are outlined in Table **5-3**. We use the reactor configuration used here is the one found in [11] (the specific dimensions are not design variables.) Although rotation improves thickness uniformity in the angular direction, these effects to the fluid flow were not considered here. Deposition occurs on a substrate (wafer), which acts as the interface between the macroscale and mesoscale simulations. Throughout the deposition process, GaAs and AlAs are deposited in alternating sequences. Annealing time occurs between depositing each species and allows for the surface of the film to relax. We used a finite element solver, specifically from the commercially available Comsol and Comsol Script software, in our procedure.

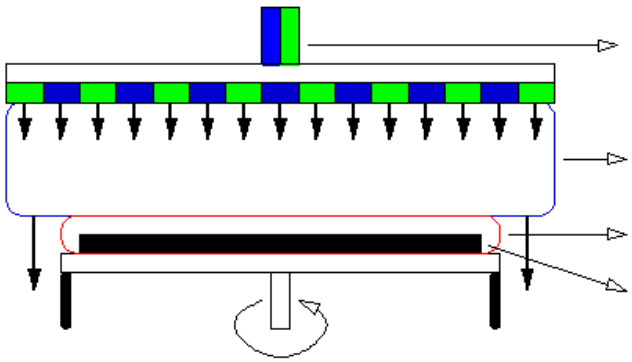

Figure **5-2**: Schematic of proposed reactor.

Table **5-3**: Reactor dimensions

Parameter	**Value**
Radius	5 cm
Height	4.5 cm
Number of inlets	1
Inlet radius	1.25 cm
Inlet temperature	300 K

5.4 Macroscale Reaction and Transport Processes

We use standard precursors for the MOVPE process for depositing GaAs and AlAs thin films – trimethyl gallium (TMG, $Ga(CH_3)_3$) and trimethyl aluminum (TMA, $Al(CH_3)_3$) as the Group III or organometallic source, also abbreviated $M(CH_3)_3$ here when mentioning a nonspecific organometallic species, arsine gas (AsH_3) as the Group V source, and hydrogen as a carrier gas. In our formulation, the inlet concentration of arsine is 7.78 times larger than the inlet concentration of the organometallic sources [48]. The hydrogen concentration is several orders of

magnitude greater than either. Reactions between the organometallic precursors and hydrogen can occur, producing numerous intermediates. We used the reaction scheme found in [49] as the basis of our gas-phase reaction scheme. From investigating relative reaction rates, species concentrations, and applying quasi steady-state assumptions to reactive intermediates, we were able to simplify the gas-phase reaction scheme to the following reactions for TMG and TMA, respectively:

$$Ga(CH_3)_3 + \tfrac{1}{2} H_2 \rightarrow Ga(CH_3)_2 + CH_4$$
$$Ga(CH_3)_2 + \tfrac{1}{2} H_2 \rightarrow GaCH_3 + CH_4$$
$$Ga(CH_3)_3 + H_2 \rightarrow GaH(CH_3)_2 + CH_4$$
$$GaH(CH_3)_2 \rightarrow GaCH_3 + CH_4$$

$$Al(CH_3)_3 + \tfrac{1}{2} H_2 \rightarrow Al(CH_3)_2 + CH_4$$
$$Al(CH_3)_2 + \tfrac{1}{2} H_2 \rightarrow AlCH_3 + CH_4$$
$$Al(CH_3)_3 + H_2 \rightarrow AlH(CH_3)_2 + CH_4$$
$$AlH(CH_3)_2 \rightarrow AlCH_3 + CH_4$$

It should be noted that the concentrations of the dimethyl $M(CH_3)_2$ and monomethyl MCH_3 species are small enough inside the gas phase that surface reactions with these species can essentially be neglected (although they are included in this particular formulation.) Most of the organometallic precursors remain in the trimethyl form throughout, or are converted to the dimethyl hydride $MH(CH_3)_2$.

We are interested in obtaining the flux of organometallic species across the wafer surface where deposition occurs, in order to calculate thickness uniformities and to provide information to the mesoscale simulation for determining interfacial step densities. The flux rates are functions of initial temperatures, concentrations, and velocities; we consider the wafer temperature, inlet velocity when gallium is being deposited, and inlet velocity when aluminum is being deposited to be variables. The inlet temperature and concentrations of all species are kept constant. We

solved the continuity equations found in Section 3.2 using the boundary conditions found in Table **5-4**, where *M* represents a generic organometallic species, *C* is the concentration, and *v* is the inlet velocity. We assumed a constant wafer surface temperature and simplified the traditionally Robin boundary conditions into Neumann conditions by assuming that a organometallic species will immediately react with arsine, effectively reducing the concentration to zero on the surface, and by assuming that hydrogen is in large enough excess that its concentration will not change significantly due to surface chemical reactions. We maintain a constant pressure throughout the reactor, reflected in the concentration of hydrogen, of 0.1 atm. The expressions for thermal conductivity, heat capacity, and viscosity were estimated from [48, 50]. Diffusion coefficients were estimated from Chapman-Enskog theory [51-53]. Density was calculated using the ideal gas law. We assume that during deposition and annealing, pseudo-steady-state operation occurs in the reactor. The modeling of lag time due to transport effects between changing inlet concentrations and that being reflected at the wafer surface is explained in Section 5.6.

Table **5-4**: Boundary conditions

Boundary	**Mass**	**Energy**	**Momentum**
Center	Axial symmetry	Axial symmetry	Axial symmetry
Inlet	$2.5\text{x}10^{-4}$ mol/m^3 $M(CH_3)_3$ $1.944\text{x}10^{-3}$ mol/m^3 AsH_3 4.06 mol/m^3 H_2	300 [K]	v [m/s]
Wafer surface	0 mol/m^3 M_i $\frac{\partial C(AsH_3)}{\partial z} = \frac{\partial(\sum_i M_i)}{\partial z}$ $\frac{\partial C(H_2)}{\partial z} = 0$	800 [K]	0 [m/s]
Outlet	Convective flux	Convective flux	Convective flux
Side and top walls	Insulation	Insulation	0 [m/s]

5.5 Mesoscale Modeling of Surface Reaction

Upon reaching the wafer boundary, organometallic species can adsorb on the surface. Since arsine is in excess, it is assumed that an organometallic molecule will instantly react with an arsine molecule when reaching the surface, forming *MAs* and releasing methane gas. The amount of hydrogen produced in surface reactions is insignificant compared with the total concentration of hydrogen and is therefore neglected. The surface reactions for gallium and aluminum containing species, which can be obtained by performing simple mass balances, are the following:

$$Ga(CH_3)_3 + AsH_3 \rightarrow GaAs + 3CH_4$$
$$Ga(CH_3)_2 + AsH_3 \rightarrow GaAs + 2CH_4 + \tfrac{1}{2} H_2$$
$$GaCH_3 + AsH_3 \rightarrow GaAs + CH_4 + H_2$$
$$GaH(CH_3)_2 + AsH_3 \rightarrow GaAs + 2CH_4 + H_2$$

$$Al(CH_3)_3 + AsH_3 \rightarrow AlAs + 3CH_4$$
$$Al(CH_3)_2 + AsH_3 \rightarrow AlAs + 2CH_4 + \tfrac{1}{2} H_2$$
$$AlCH_3 + AsH_3 \rightarrow AlAs + CH_4 + H_2$$
$$AlH(CH_3)_2 + AsH_3 \rightarrow AlAs + 2CH_4 + H_2$$

The interface between the macroscale and mesoscale involves calculating step densities at various locations across the wafer surface based on their deposition rates, which are converted from *mol/m^2 s* to *Monolayers/s* from GaAs and AlAs lattice constants [1, 41]. The surface model is also quite simplified – all organometallic species that hit the wafer surface are assumed to stick, desorption does not occur, and impurity incorporation is not considered. We continue to use one lattice with the size of 100x100, which is large enough to reduce noise to a workable level, but small enough that mesoscale simulations are not cost prohibitive. To assure the size was sufficiently large, we compared step densities with those predicted from identical operating

conditions for larger lattices (250x250) and found that the values for the 100x100 were within the expected range.

5.6 Dynamic Considerations

Generally, these processes are considered to operate at steady-state. For the most part, this is a reasonable assumption, especially during most of deposition and annealing. Throughout the process, however, step changes (zero to full concentration and vice versa) occur at the inlet. Due to transport effects, these changes are not immediately reflected at the wafer surface. As a result, these time-delay effects (from starting or stopping deposition of *MAs*) should be considered, as they could have significant effects on the evolution of step density at the wafer surface. The most complete methodology is using a finite element solver to determine the solution to a time-dependent system of PDEs. Unfortunately, this approach significantly increases computation time (for comparison, determining the steady-state solution for GaAs deposition with 6208 grid elements takes approximately 4-5 minutes, while determining the time-dependent solution for GaAs deposition for 500 time elements takes about 15-30 minutes.) No simple correlations between time decay rates, input velocities, and wafer position were readily observable. In order to approximate the decay rate, results for inflow velocities of 0.1 *m/s*, 0.2 *m/s*, 0.4 *m/s*, and 0.8 *m/s* were tabulated and linearly interpolated for each wafer point and a constant exponential decay rate was assumed over 0.2 second intervals in the mesoscale (in reality, decay is not simply exponential.) Start-up rates for GaAs and AlAs were found to be inversely proportional to their corresponding decay rates. The main goal with determining decay rates is to more accurately capture wafer surface effects, so while we acknowledge that this model is incomplete, it serves as a springboard to the consideration of the dynamic optimization problem.

This time delay effect brings up an important question – what exactly is annealing time? We define annealing time to be the time period when we assume the steady-state flux of *MAs* to the wafer surface is zero. This is in contrast to "off time," which is when the inlet concentration of *MAs* is zero, which would be the sum of "ramp down" time plus annealing time. Both the minimization of annealing time and "off time" are considered, yielding potentially divergent results (we would expect minimizing "off-time" to favor higher velocities, due to quicker decay rates.) This definition also allows the gallium and aluminum deposition problems to be separated into two independent steady-state systems, as we are only concerned about changes from zero concentration to full concentration of either species. This separation assumption simplifies the computational complexity of the problem into two separate calculations with six species, as opposed to one calculation with ten species. Simultaneous mixing of gallium and aluminum species would also likely lead to poor interface quality, so neglecting the possibility is a reasonable assumption.

5.7 Optimization Problem

The ultimate goal of this research is to determine optimal operation of a thin-film deposition process to produce layered heterostructures. The macroscale solutions can be determined from simulating the reaction-transport PDEs, and the mesoscale solutions from kMC simulations. Then, the optimal operating conditions are to be determined from the metrics for macroscale process yield (thickness uniformity) and mesoscale product quality (interfacial step density.) Also desired are the operations at lower temperatures and annealing times mentioned in Section 5.2. We calculated step densities at 15 points along the wafer surface at 0.25 cm intervals, starting from the center until 0.25 cm from the edge (due to unavoidable edge effects, we did not consider deposition adjacent to the outlet.) All variables, except for annealing time,

were scaled between zero and one. Temperature is kept within a range of 700-950 K (temperatures that exceed 950 K result in rapidly increasing computational time for mesoscale simulations.) Deposition velocities are kept between 0.1-0.8 m/s, to maintain realistic flow rates (below 2500 sccm) at the upper bound and extremely lengthy decay times at the lower bound. Due to the possibility of getting stuck in an unfeasible region, especially using gradient-based methods, linear penalties were used instead of discontinuous. Implementation of these penalties is challenging, particularly for the gradient-based methods. The multi-objective optimization problem is made into a single-objective optimization problem using simple weighting criteria.

Numerous factors were considered in this optimization problem. Two different perturbation sizes are used for ISAT: a smaller size of 0.06, expected to sacrifice noise for enhanced local properties, and a larger size of 0.16, expected to sacrifice local property effects for less noisy general trends. Three different starting conditions are considered, a "low," "medium," and "high" value, outlined in Table **5-5**, used to check for consistency. In terms of formulation, we considered both minimizing "annealing time only" and "off time," which is annealing time plus "ramp down" time. Only annealing time is considered as a decision variable, however: "ramp down" time is estimated from the solution to the macroscale problem. Finally, four different optimization methods – steepest descent, BFGS, Nelder-Mead, and Hooke-Jeeves, were attempted. This gives a total of at least 48 optimization trials needed for this formulation if no other minor changes are added to each system.

Table **5-5**: Starting conditions (scaled)

Type	Temperature	Annealing Time (s)	Gallium Inlet Velocity	Aluminum Inlet Velocity
Low	0.3 (775 K)	0.5 (0.25 steepest descent)	0.3 (0.31 m/s)	0.3 (0.31 m/s)
Medium	0.5 (825 K)	1 (0.5 steepest descent)	0.5 (0.45 m/s)	0.5 (0.45 m/s)
High	0.7 (875 K)	2 (0.75 steepest descent)	0.7 (0.59 m/s)	0.7 (0.59 m/s)

The distinction between "annealing time only" and "off time" requires a difference in cost functions. Less readily apparent are the differences in the optimization function between the gradient-based and pattern search algorithms. This difference is most noticeably reflected in the magnitude of the weights. Even so, the scaling for each optimization function is generally the same, except for the size of the penalty. For the pattern search methods, a large penalty is beneficial, as it will direct the search algorithm back towards the feasible region. With gradient-based algorithms, a large penalty is also desired for this reason, but since the size of the penalty affects the gradient, the next evaluated point could bypass the feasible region and exceed the opposite bound, sometimes severely (this is controllable by step size in steepest descent, but not in BFGS.) The cost functions used for "annealing time only" and "off time," respectively, for pattern-search algorithms are:

$$1000T^2 + 1000A^2 + 10{,}000(U_G^2 + U_A^2) + 1000(S_{A/G}^2 + S_{G/A}^2) + 200{,}000\sum_i P_i$$

$$1000T^2 + 1000F^2 + 10{,}000(U_G^2 + U_A^2) + 1000(S_{A/G}^2 + S_{G/A}^2) + 200{,}000\sum_i P_i$$

where T is the temperature, A is the annealing time, F is the off-time, U_G and U_A are the thickness non-uniformities for gallium and aluminum, respectively, $S_{A/G}$ and $S_{G/A}$ are the step densities of aluminum over gallium, respectively, and P_i are the linear penalties for exceeding i constraints. For gradient-based algorithms, these cost functions for "annealing time only" and "off time," respectively, are:

$$0.2T^2 + 0.1A^2 + (U_G^2 + U_A^2) + 0.1(S_{A/G}^2 + S_{G/A}^2) + W\sum_i P_i$$

$$0.2T^2 + 0.1F^2 + (U_G^2 + U_A^2) + 0.1(S_{A/G}^2 + S_{G/A}^2) + W\sum_i P_i$$

where W are the penalty weights (1 for steepest descent and 0.2 for BFGS) and the other variables are the same as for the pattern search algorithm. Note that these formulations are mostly

equivalent in terms of scaling. Temperature is more strongly weighted in the gradient-based algorithms for computation time consideration purposes, and the relative weights of the penalties are much higher for pattern search methods. The magnitudes of the weights are larger for the pattern search methods to prevent numerical gradient calculation errors. Off-time F is equal to the following:

$$F = (R + A)/4$$

where R is "ramp down" time (off time is divided by four for scaling purposes.) Mathematically, R is average "ramp-down" time of the slower species across the wafer surface. The sum of penalties P_i is equal to the following, where a positive superscript represents a deviation from the maximum (1) and a negative superscript a deviation from the minimum (0):

$$\sum_i P_i = P_T^+ + P_T^- + P_A^- + P_{UG}^+ + P_{UG}^- + P_{UA}^+ + P_{UA}^-$$

Gradients were calculated numerically, with a perturbation size of 0.01 for each decision variable. It was assumed that the vast majority of perturbed calculations would fall within the ellipsoid of attraction of the original point. We acknowledge, however, that this occasionally may not happen, particularly as the ISAT table grows in size, where some perturbed points will fall within a particular EOA while others do not. For steepest descent, step size was constrained such that an evaluation never exceeded 0.05 from a bound (in order to accomplish this for annealing time, the maximum value was fixed at two seconds and scaled by dividing by two.) We used an ε_{tol} value of 0.03, although this was later increased to 0.05 in the Hooke-Jeeves simulations.

Various expectations exist regarding the success of each optimization method and function formulation. We would hypothesize a push towards higher velocities with the "off time" formulation than with the "annealing time only" formulation. This is because lower flux rates are correlated with improved step densities, but this effect would be negated by the increase in "ramp

down" times. In terms of the optimization methods themselves, we predict that the Nelder-Mead method is most likely to have the slowest convergence of the four, although it is less likely to experience difficulties with constraint violations. BFGS would be predicted to most likely converge more quickly than steepest descent, although the latter would be less likely to experience constraint violation difficulties due to the controllability of step sizes. Hooke-Jeeves would be predicted to be more intermediate – a faster search than Nelder-Mead but slower than the gradient-based methods, while somewhat more likely to violate constraints than Nelder-Mead, but less likely than for gradient-based methods.

5.8 Computational Resources

This work was primarily performed using the commercially available software MATLAB. As mentioned in Section 5.3, the finite element solution system was initially modeled using Comsol, then solved with a combination of Comsol Script and its link to MATLAB. Simulations were performed on the High-Performance Computing Cluster here at the Pennsylvania State University with typical computational requirements of approximately two weeks. The Nelder-Mead algorithm was available through fminsearch in MATLAB, the Hooke-Jeeves algorithm through accompanying codes to [18], and the gradient methods were self-programmed.

5.9 Gradient-Based Method Results

Unfortunately, the attempted implementation of gradient-based methods towards this particular multiscale optimization problem was not successful. With BFGS, the uncontrollability of the Hessian resulted in step sizes that frequently exceeded bounds by severe margins. Usually,

this would simply crash the algorithm. The macroscale simulations would fail if the temperature "fell below absolute zero" or was orders of magnitude higher than the expected range. Typically, the mesoscale simulations would fail if velocities were negative (due to inadequate surface deposition) or very low temperatures (but above absolute zero) due to lack of molecular movement, even with a functional macroscale simulation. When temperatures greatly exceeded (but were within the same order of magnitude of) the upper bound, mesoscale simulations simply took too long to converge.

With the limited step size for steepest descent, the above problems were avoided. The failure with steepest descent, however, was the inability to control the size of the penalty. As a result, simulations might appear to be converging towards a solution at first, but the solution may reach a bound and be directed that exceeding the bound even further was the proper direction to proceed. This would cause the simulator to believe the optimal solution was achieved. While increasing the size of the penalty may alleviate this effect, due to gradient noise, the undesired effect of continuously exceeding bounds would likely be introduced. Coupled with explainable results from Nelder-Mead, the steepest descent approach was not further pursued.

5.10 Nelder-Mead Results

Twelve trials as described by the procedure in Section 5.7 were performed using the Nelder-Mead pattern search algorithm. The results are outlined in **Table 5-6**, where Type A trials are "annealing time only," and Type O trials are "off time."

With the exception of Entry 8, a clear outlier, the optimal function values were all within about 500 of each other (note that we cannot compare Type A and Type O function values directly.) A few trends are fairly clear from this formulation. First, annealing time was almost always driven close to zero (or at least significantly lower than the starting condition.) Velocity

appears to be insensitive in the annealing time only formulations. This result is not surprising, as the effects of velocity on deposition rate (and therefore step density) and thickness non-uniformity is rather minor. In the "off-time" formulations, velocity is usually driven towards its maximum, as reducing "off-time" outweighs any other effects. The effects of temperature are a little less clear. This formulation appears to raise temperatures in the annealing time only function; from comparing function values, the optimal temperature seems to be approximately 890 K. At first glance, temperature appears insensitive for "off-time" formulations; however, Trials 5 and 9 are quite close in value, and with further confirmation from Hooke-Jeeves trials, it once again appears that the optimal operating temperature is approximately 890 K. It is interesting to note that the findings of temperature insensitivity here are in direct opposition to the purely mesoscale problem, where a temperature convergence was precise and reproducible from different starting conditions. The most likely explanation is that in the purely mesoscale case, the goal was to minimize temperature while maintaining a maximum allowable step density, as opposed to determining the interplay between the two.

Table **5-6**: Summary of Nelder-Mead optimization results

Trial	**Type**	**Starting Con-dition**	**Pertur-bation Size**	**Temp. (K)**	**Anneal. Time (s)**	**Ga Vel. (m/s)**	**Al Vel. (m/s)**	**Func-tion Value**	**Comments**
1	A	Low	Large	830	0.00	0.33	0.32	4417	
2	A	Low	Small	810	0.02	0.44	0.33	4442	
3	A	Medium	Large	884	0.11	0.80	0.39	3940	
4	A	Medium	Small	---	---	---	---	---	Time failure
5	A	High	Large	917	0.04	0.80	0.60	4223	
6	A	High	Small	954	-0.20	0.82	0.62	50500	Bounds failure
7	O	Low	Large	782	0.03	0.38	0.43	6043	Outlier
8	O	Low	Small	871	0.02	0.74	0.80	4648	Consistent
9	O	Medium	Large	860	0.00	0.63	0.66	4908	
10	O	Medium	Small	822	0.28	0.54	0.79	4942	
11	O	High	Large	---	---	---	---	---	Time failure
12	O	High	Small	893	0.06	0.78	0.80	4636	Consistent

In addition to consistency, we are also interested in the ability for optimization to improve objective function values. Using Trial 3 to represent annealing time only and Trial 12 to represent "off-time," we compared these optimized uniformity and step density profiles with those at their respective starting conditions. The results are shown in a series of figures: Figures **5-3** and **5-5** show uniformity profiles, Figures **5-4** and **5-6** show step density profiles, and Figures **5-7** and **5-8** show a vertical cross-section of the wafer at "optimal" conditions at the center position. Thickness non-uniformities basically remained unchanged – this is expected, since the shapes of surface deposition profiles are insensitive to changes in velocity and temperature, although small improvements were generally observed. Note that relative thickness uniformity is calculated here, not absolute differences; this is done in order to prevent simply minimizing flow rates. Step densities were improved considerably when starting at a moderate temperature, but remained mostly unchanged at high starting temperatures. These values are shown in Table **5-7**.

Table **5-7**: Effects of optimization on thickness non-uniformity and step density, Nelder-Mead search method

	Trial 3 Starting Condition	**Trial 3 Optimal Condition**	**Trial 12 Starting Condition**	**Trial 12 Optimal Condition**
Relative Ga Thickness Non-uniformity	0.4433	0.4122	0.4296	0.4130
Relative Al Thickness Non-uniformity	0.3721	0.3640	0.3827	0.3939
Al over Ga Step Density	0.4688	0.3747	0.3609	0.3355
Ga over Al Step Density	0.7411	0.5485	0.5988	0.5734

In Figures **5-3** and **5-5**, note that the general shapes of the profiles remain similar. The changes in absolute deposition rate indicate that increasing velocity increases deposition rate. Note that for the Al case in the "off-time" formulation, non-uniformity actually increased (higher velocities generally have higher non-uniformities,) but the improvement in decreasing "off-time" overshadows this effect.

In Figures **5-4** and **5-6**, as the temperature increased from starting to optimal conditions, the step densities also improved. The magnitude of the increase is much larger in the annealing time only case, as the optimal operating temperature for the "off-time" formulation is only 18 K higher than the starting operating temperature. Nevertheless, this increase was sufficient to improve the step density in both cases. The profile is indicative of using ISAT to predict step densities at points with similar conditions, which explains the sudden jumps. From the cross-sections in Figures **5-7** and **5-8**, we observe smooth interfaces with little changes in height. Note that green represents a GaAs layer, maroon an AlAs layer, and blue open sites.

Overall computational time was approximately 10-30 days, depending on the starting point of the simulation (ones that started at higher temperatures generally took longer to converge.) Since the limit per cluster submission is 288 hours (12 days), trials were automatically stopped after 280 hours once any set of mesoscale simulations completed, while the maximum allowable time for submission running was 287.5 hours. If necessary, the current evaluation point and the generated ISAT tables were recorded and simulations were restarted at the current evaluation point. Admittedly, a few simulations were lost if mesoscale simulations were still being performed after the 7.5 hour allowance window had passed. A typical number of required function evaluations would be around 3000-9000, with ISAT tables ranging in size from about 40-100 entries. Since seven mesoscale simulations are required from every table entry, it is estimated that this methodology saves about an order of magnitude in computation time.

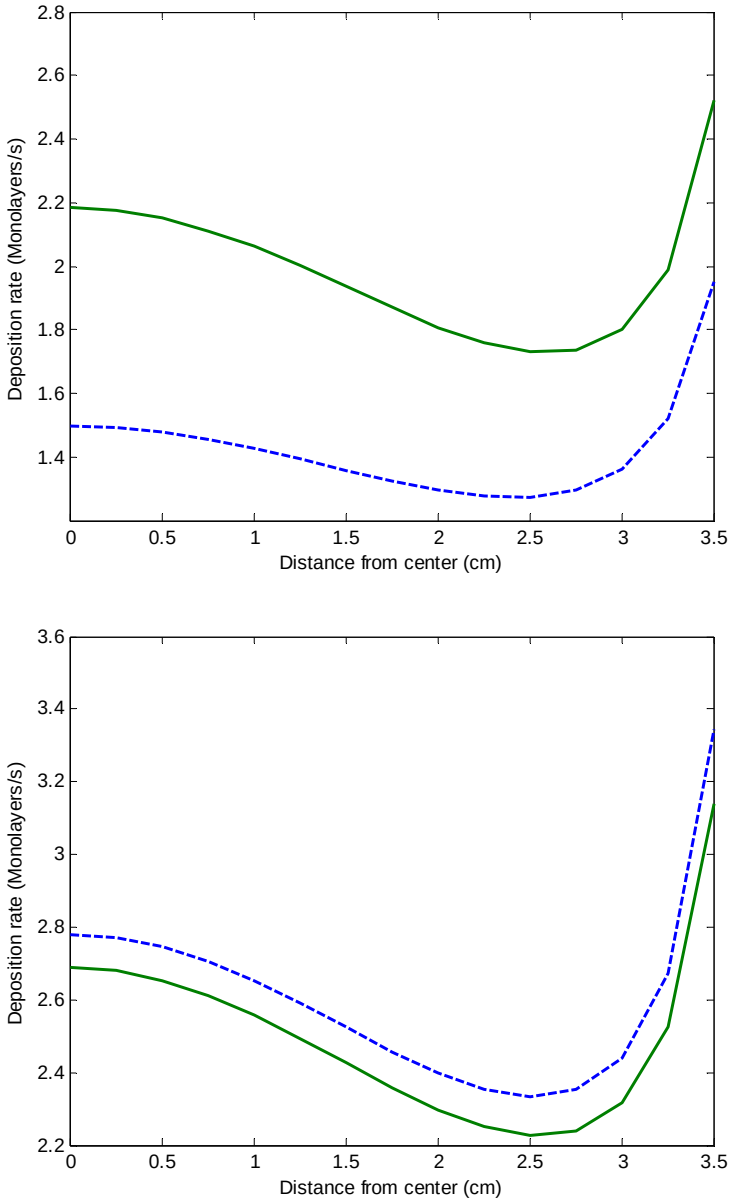

Figure **5-3**: Uniformity profiles at starting conditions (dashed) and optimal conditions (solid) for Ga (top) and Al (bottom) for annealing time only formulation using Nelder-Mead

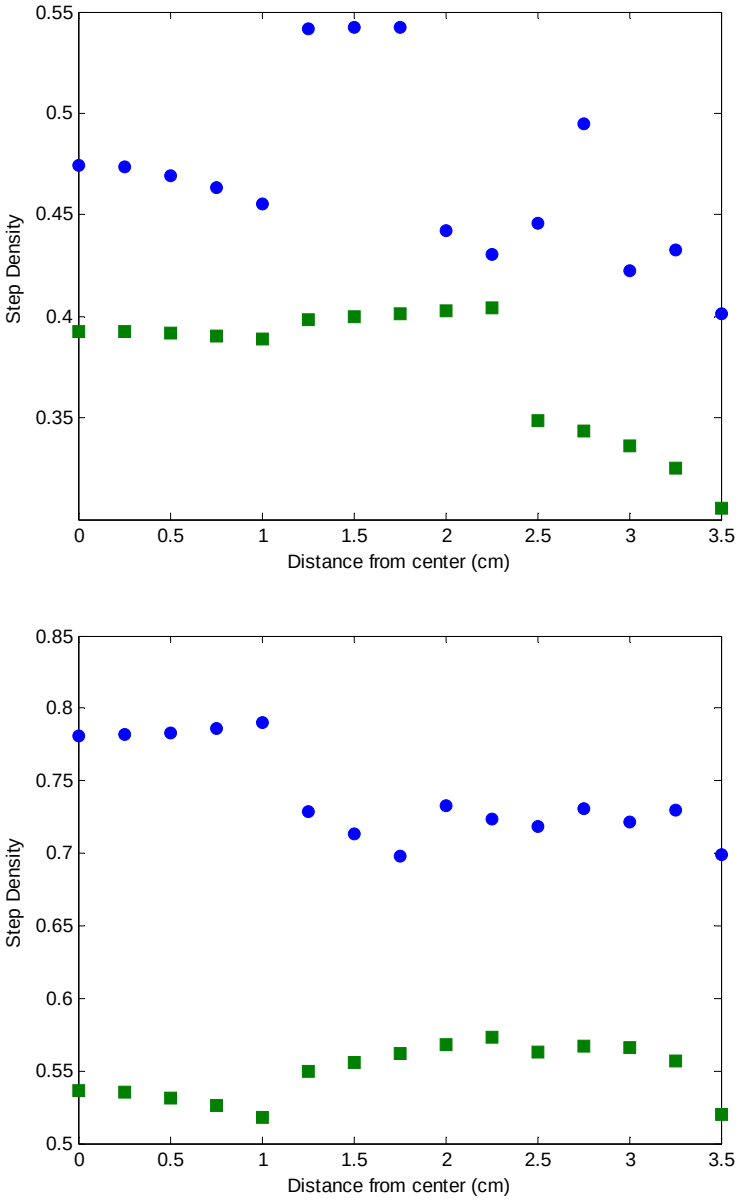

Figure **5-4**: Step density profiles at starting conditions (circles) and optimal conditions (squares) for Al over Ga (top) and Ga over Al (bottom) for annealing time only formulation using Nelder-Mead

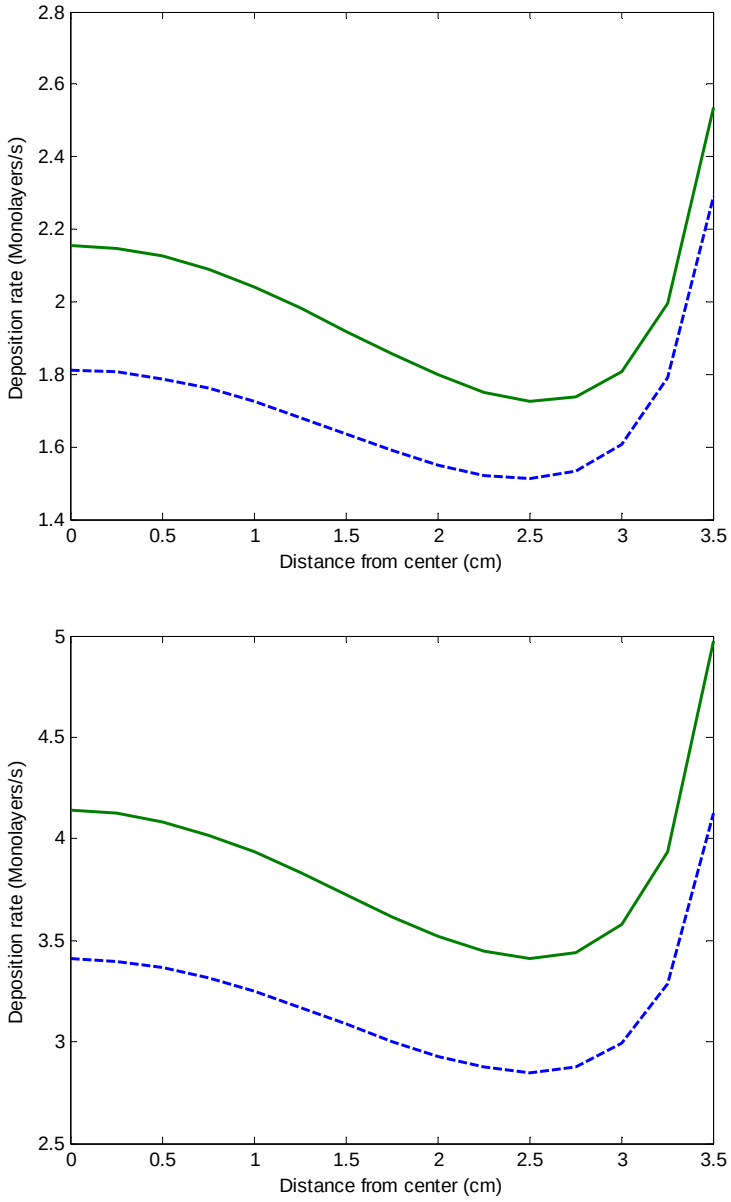

Figure **5-5**: Uniformity profiles at starting conditions (dashed) and optimal conditions (solid) for Ga (top) and Al (bottom) for "off-time" formulation using Nelder-Mead

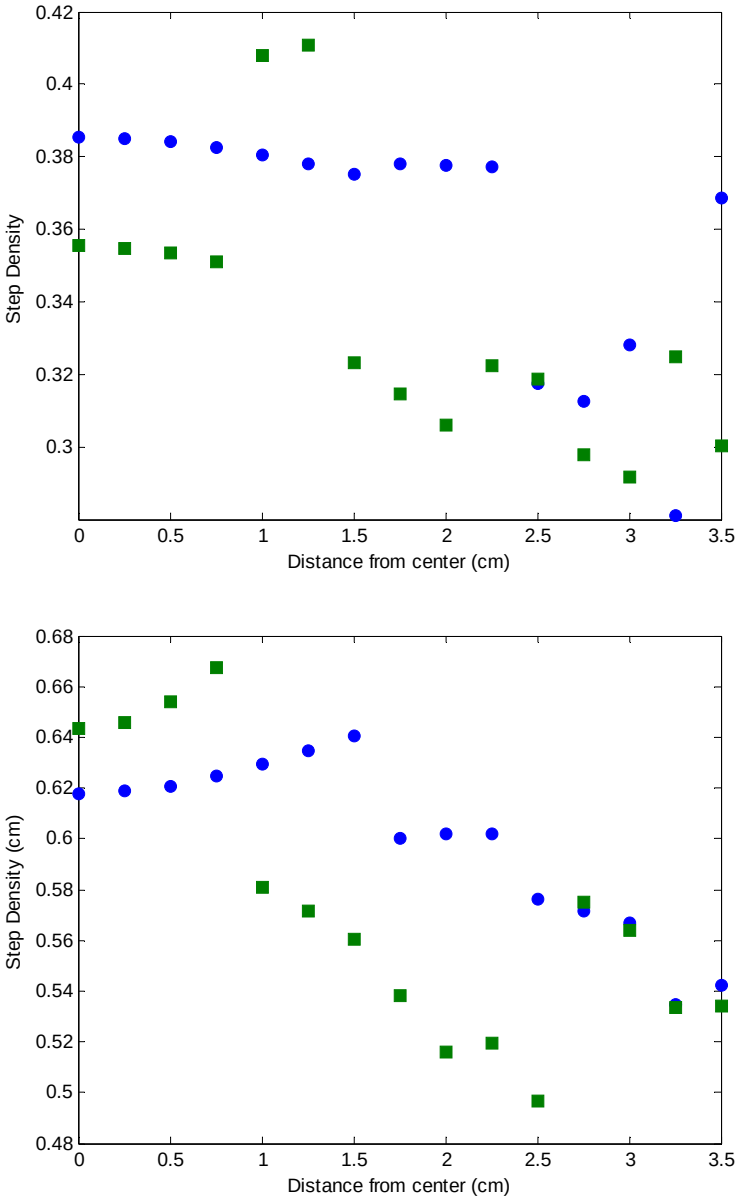

Figure **5-6**: Step density profiles at starting conditions (circles) and optimal conditions (squares) for Al over Ga (top) and Ga over Al (bottom) for "off-time" formulation using Nelder-Mead

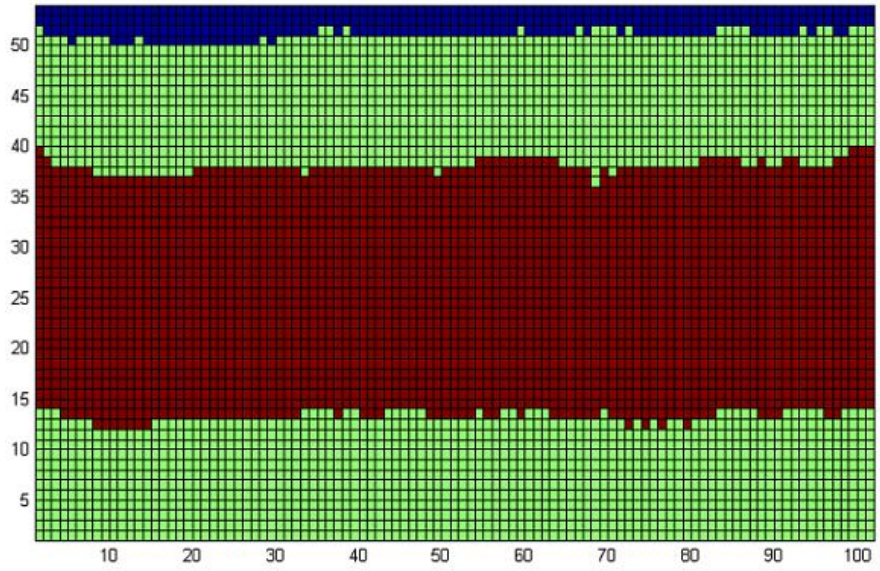

Figure **5-7**: Cross-section of wafer at "optimal" conditions, wafer center, for annealing time only formulation using Nelder-Mead

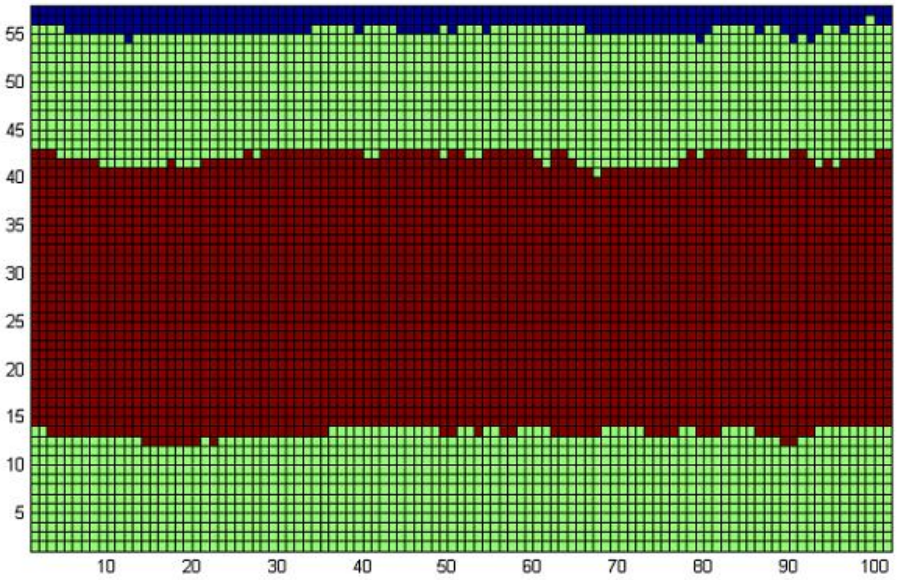

Figure **5-8**: Cross-section of wafer at "optimal" conditions, wafer center, for "off-time" formulation using Nelder-Mead

5.11 Hooke-Jeeves Results

With the somewhat limited success of the Nelder-Mead formulation, we proceeded to perform simulations using the Hooke-Jeeves algorithm. Two different stencil sizes were used: one started with a size of 1/4 and successively reduced to 1/8, 1/16, and 1/32; the other started with a size of 1/8 and successively reduced to 1/16, 1/32, and 1/64. Simulation often required more than the allotted 288 hours of cluster time, often upwards of 3-4 twelve-day sessions, limiting the availability of results, particularly at high temperature starting conditions. Unlike in the Nelder-Mead formulation, however, Hooke-Jeeves trials were not restarted at their current "optimal" point, but rather from the original starting point in order to maintain the integrity of the stencil size (the previously constructed ISAT table was used, however.)

The results using the Hooke-Jeeves search algorithm were more consistent than they were from using the Nelder-Mead search algorithm. Table **5-8** summarizes the results for annealing time only problems, whereas Table **5-9** does so for the "off-time" formulation. A considerable range of temperatures exists for both; as in Nelder-Mead, temperatures generally increased from their starting conditions (although very little high temperature starting condition data is available,) averaging around 870-890 K. In the annealing time only formulation, Ga inlet velocities were typically raised towards the maximum, whereas Al inlet velocities were lowered towards the minimum. In the "off-time" formulation, both velocities were raised towards the maximum as expected. Note that for the "off-time" formulation, the results generally agree with Trials 8 and 12 in the Nelder-Mead formulation. Annealing time is generally reduced towards the minimum, although this does not always occur in the "off-time" formulation. A possible explanation is that at higher inlet velocities, "ramp-down" time is small and the surface might benefit from a short period of relaxation. Also note that roughly 50 percent of the "off-time"

formulations problems are bounds failures. This is likely due to the aggressive nature of the search algorithm occasionally exceeding the constraints and failing to reenter the feasible region.

Table **5-8**: Summary of Hooke-Jeeves optimization results, formulated by annealing time only

Trial	Stencil Size	Starting Condition	Perturbation Size	Temp. (K)	Anneal. Time (s)	Ga Velocity (m/s)	Al Velocity (m/s)	Function Value	Comments
1	1/4	Low	Large	845	0.03	0.75	0.14	3792	
2	1/4	Low	Small	861	0.00	0.77	0.14	3711	
3	1/4	Medium	Large	888	0.03	0.78	0.12	3801	
4	1/4	Medium	Small	---	---	---	---	---	Unavailable
5	1/4	High	Large	---	---	---	---	---	Unavailable
6	1/4	High	Small	---	---	---	---	---	Unavailable
7	1/8	Low	Large	845	0.09	0.74	0.11	3688	
8	1/8	Low	Small	920	0.13	0.11	0.12	3336	
9	1/8	Medium	Large	919	0.02	0.76	0.11	3839	
10	1/8	Medium	Small	---	---	---	---	---	Unavailable
11	1/8	High	Large	844	0.03	0.69	0.12	3659	
12	1/8	High	Small	---	---	---	---	---	Unavailable

Table **5-9**: Summary of Hooke-Jeeves optimization results, formulated by "off-time"

Trial	Stencil Size	Starting Condition	Perturbation Size	Temp. (K)	Anneal. Time (s)	Ga Velocity (m/s)	Al Velocity (m/s)	Function Value	Comments
13	1/4	Low	Large	869	0.13	0.75	0.77	4336	
14	1/4	Low	Small	853	0.13	0.68	0.77	4783	
15	1/4	Medium	Large	927	0.00	0.80	0.82	10966	Bounds failure
16	1/4	Medium	Small	825	0.84	0.76	0.82	10983	Bounds failure
17	1/4	High	Large	859	0.00	0.66	0.77	4686	
18	1/4	High	Small	---	---	---	---	---	Unavailable
19	1/8	Low	Large	880	0.17	0.72	0.802	5351	Bounds failure
20	1/8	Low	Small	900	0.44	0.75	0.74	4892	
21	1/8	Medium	Large	856	0.50	0.80	0.79	4757	
22	1/8	Medium	Small	852	1.13	0.78	0.81	8340	Bounds failure
23	1/8	High	Large	895	1.81	0.80	0.81	7971	Bounds failure
24	1/8	High	Small	---	---	---	---	---	Unavailable

In general, these results show consistency regarding velocity. The expected raising towards maximum values with "off-time" formulation occurs here, and does not warrant further explanation. For the annealing time only formulation, Ga inlet velocity is usually raised towards its maximum, while Al inlet velocity is lowered towards its minimum. An exception occurs in Trial 8, where the Ga inlet velocity is in fact brought towards the lower boundary. Interestingly enough, this trial actually has the best function value out of all of the trials examined. This occurrence raises the possibility of stochastic noise unexpectedly directing the search in the opposite direction from usual, or possibly the existence of multiple optima. Because it was an exception, Trial 8 is not considered for the representative trial for determining uniformity and step density profiles. A few trials, particularly at high temperature starting conditions, likely greatly exceeded the upper bound for temperature, resulting in rapid surface migrations and lengthy simulation times; due to this, some optimal results were unavailable and are labeled as such in Tables **5-8** and **5-9**.

As before, we also wish to determine whether these optimal conditions reduced step density and thickness non-uniformities. We chose two representative trials with the best function values – Trial 7 for the annealing time only formulation (we did not select Trial 11 due to time constraints,) and Trial 13 for the "off-time" formulation. The results are shown in a similar series of figures: Figures **5-9** and **5-11** show uniformity profiles, Figures **5-10** and **5-12** show step density profiles, and Figures **5-13** and **5-14** show a vertical cross-section of the wafer at "optimal" conditions at the center position. Step density and thickness non-uniformity values are shown in Table **5-10**. Similar general trends were observed as in the Nelder-Mead problems: a decrease in step density due to a corresponding increase in temperature, and little change in the non-uniformities, with any instance of a worse uniformity being balanced with improvements in other categories. Both Trials 7 and 13 had the same starting conditions and therefore the same step densities and thickness uniformities.

Computation time for Hooke-Jeeves was comparable to somewhat longer than with Nelder-Mead, often requiring 10-30 days or longer. The main explanation for the sometimes observed increase in simulation time is that the Nelder-Mead algorithm generally samples a larger subspace that includes simulations performed at higher temperatures, sometimes exceeding 950 K. Without ISAT, these simulations would still need to be performed, implying that relative computational savings should not be affected by this. Hooke-Jeeves is a more efficient algorithm, however, and as a result, typically only 1000-2000 mesoscale evaluations needed to be performed for an optimization trial. With table sizes generally ranging from 30-70 entries, the proportional savings was only 2-8 times of not using ISAT. In order to obtain greater computational savings, the tolerance parameter may be increased.

Table **5-10**: Effects of optimization on thickness non-uniformity and step density, Hooke-Jeeves search method

	Trial 7 Starting Condition	**Trial 7 Optimal Condition**	**Trial 13 Starting Condition**	**Trial 13 Optimal Condition**
Relative Ga Thickness Non-uniformity	0.4550	0.4105	Same as Trial 7	0.4119
Relative Al Thickness Non-uniformity	0.3582	0.3524	Same as Trial 7	0.3937
Al over Ga Step Density	0.6617	0.5692	Same as Trial 7	0.4500
Ga over Al Step Density	0.9283	0.6541	Same as Trial 7	0.6239

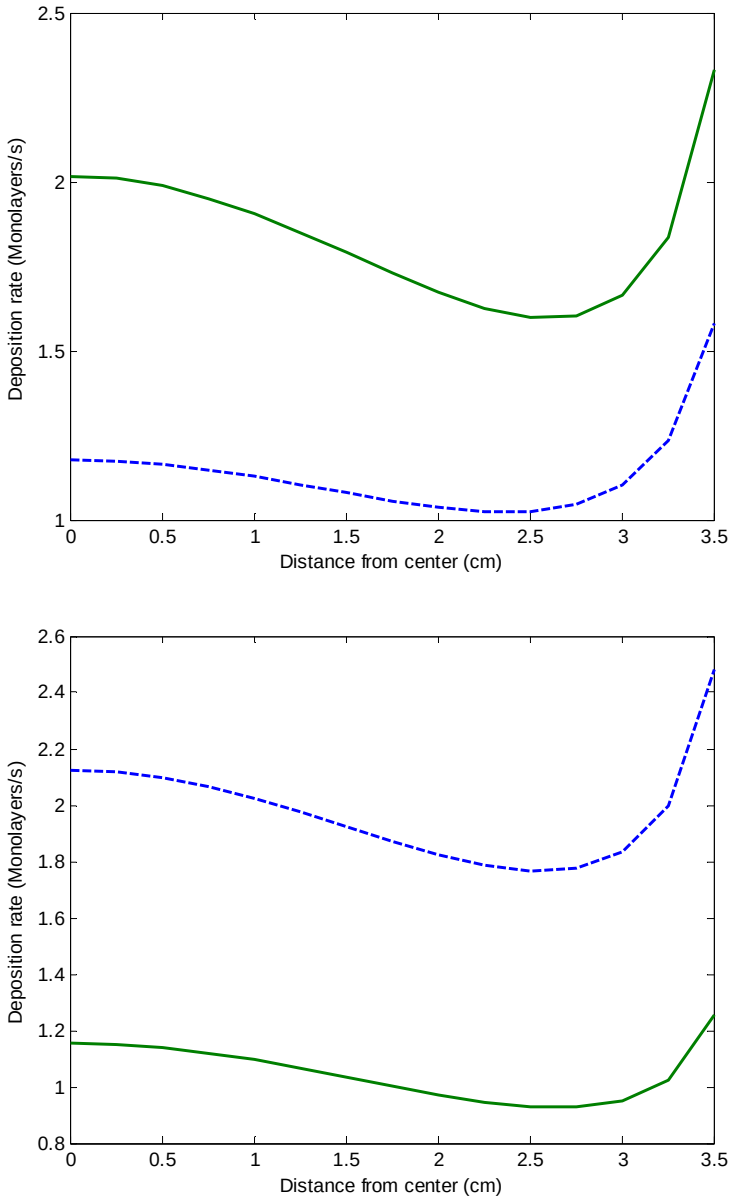

Figure **5-9**: Uniformity profiles at starting conditions (dashed) and optimal conditions (solid) for Ga (top) and Al (bottom) for annealing time only formulation using Hooke-Jeeves

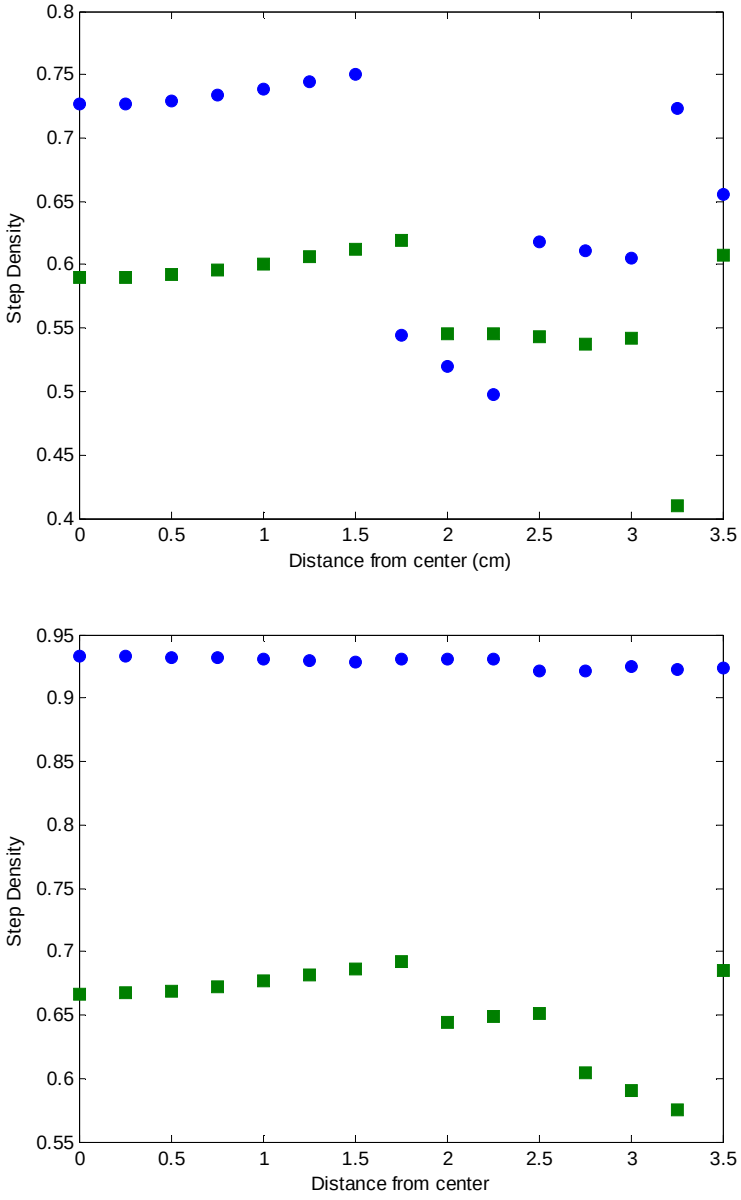

Figure **5-10**: Step density profiles at starting conditions (circles) and optimal conditions (squares) for Al over Ga (top) and Ga over Al (bottom) for annealing time only formulation using Hooke-Jeeves

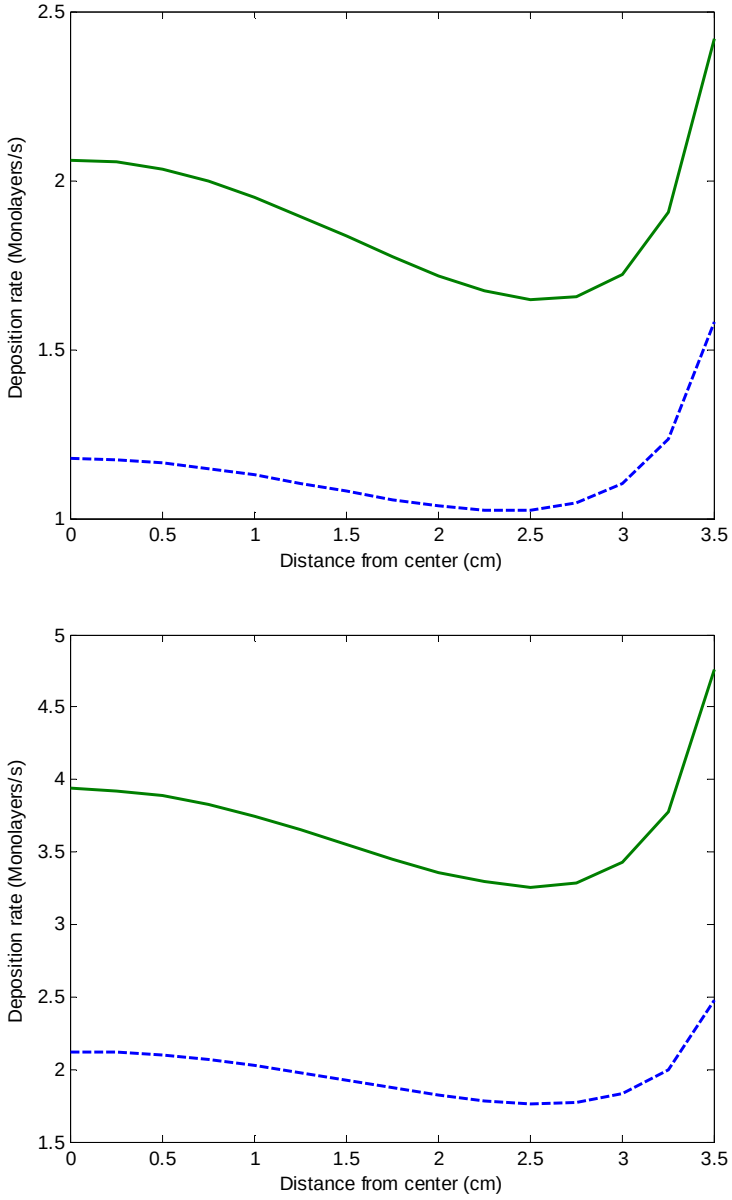

Figure **5-11**: Uniformity profiles at starting conditions (dashed) and optimal conditions (solid) for Ga (top) and Al (bottom) for "off-time" formulation using Hooke-Jeeves

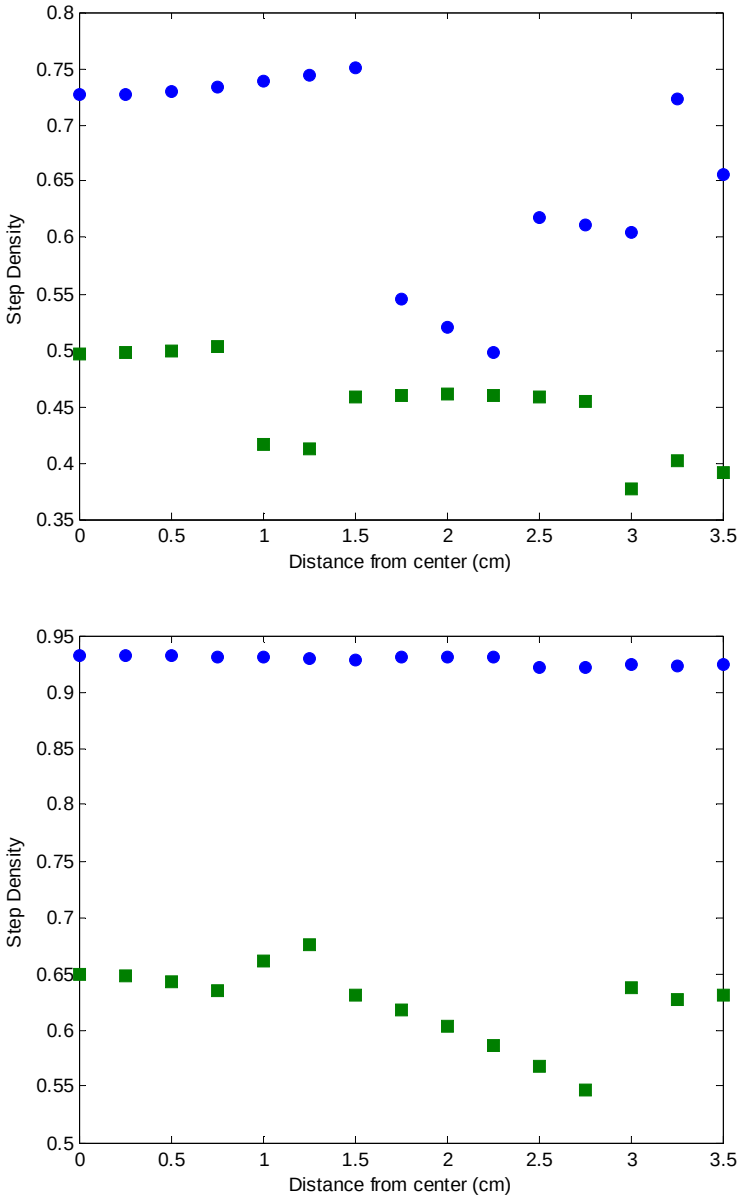

Figure **5-12**: Step density profiles at starting conditions (circles) and optimal conditions (squares) for Al over Ga (top) and Ga over Al (bottom) for "off-time" formulation using Hooke-Jeeves

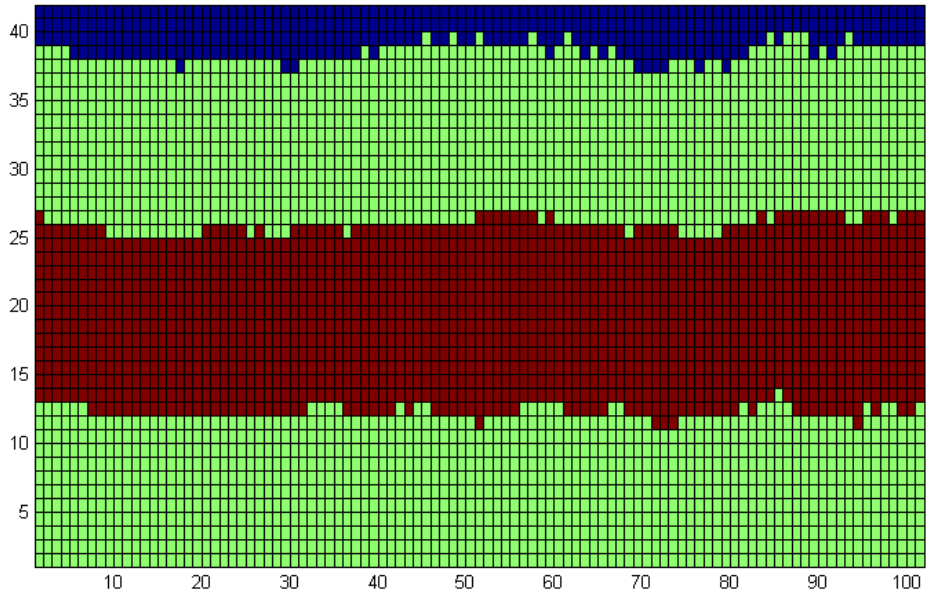

Figure **5-13**: Cross-section of wafer at "optimal" conditions, wafer center, for annealing time only formulation using Hooke-Jeeves

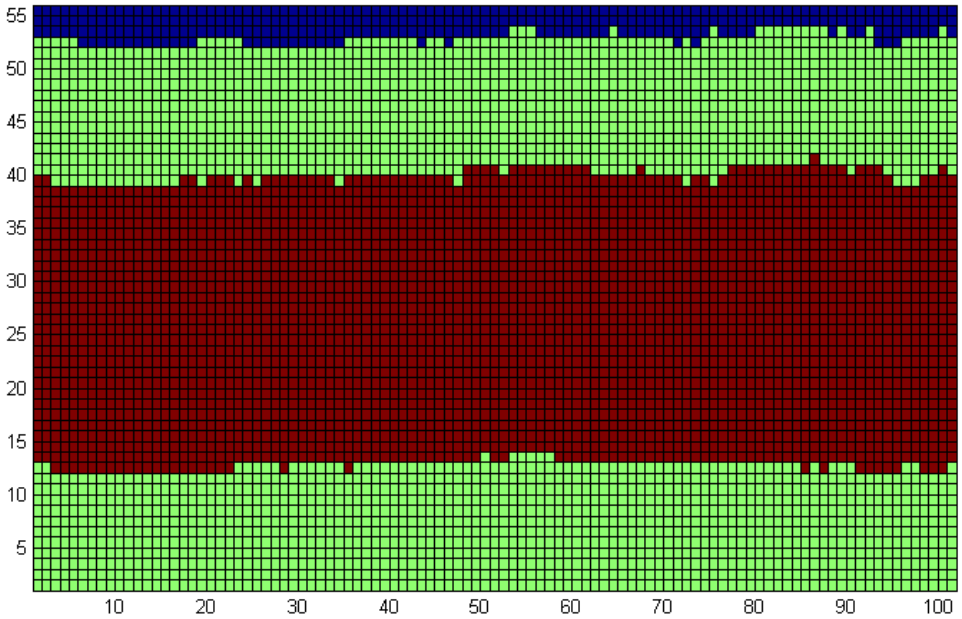

Figure **5-14**: Cross-section of wafer at "optimal" conditions, wafer center, for "off-time" formulation using Hooke-Jeeves

Chapter 6

Conclusions and Future Work

This chapter will focus on what conclusions can be drawn from the proposed methodology and results outlined in Chapter 5. It will also discuss the limitations of this model, the challenges of fixing these limitations, and possible resolutions in future work. Finally, this chapter will discuss many possible directions for this research to expand in, both in applications and in methodology.

6.1 Conclusions

The results in Chapter 5 outline that this methodology is successful in reducing computation time to determine optimal operating conditions. The actual results, however, are limited by three main factors: computational requirements, problem insensitivities, and complications from exceeding constraints. In order to maintain reasonable computational efficiency, the lattice size was not able to exceed 100x100. While sufficient for many of our applications (particularly in the "off-time" formulation,) the stochastic noise was likely too great to handle the insensitivities in temperature and velocity for the annealing-time only formulation. These insensitivities are unfortunately unavoidable, and particularly in stochastic optimization, may be difficult with which to deal. As a result, we are forced to accept that the lack of change from initial function values in the optimization problem in many of the Nelder-Mead simulations and occasionally in the Hooke-Jeeves simulations is partially from the nature of the problem itself and not due to the methodology. Finally, managing constraints proved to be a difficulty in most of the optimization techniques (Nelder-Mead excluded.) This is expected to happen when

applying unconstrained optimization techniques to constrained optimization problems. Further complicating the issue is that if either velocity drops below zero (about -0.14 when scaled,) the algorithm will crash. This can be prevented, however by resetting velocities at the bottom of the scale.

Nevertheless, numerous objectives from this work were accomplished. We developed new methodologies and utilized previous ones to tackle computationally efficient multiscale optimization with respect to microelectronics fabrication of layered heterostructures. We implemented a reaction-transport mechanism and solved its conditions using a finite element solver, coupled with using a kMC code to model the mesoscale surface reactions that represent formation of the layered heterostructure. We further characterized the interfaces to account for specific properties for having multiple interfaces, and then applied this to an optimization framework with computational acceleration from ISAT. Using the Hooke-Jeeves search method, we were generally able to obtain a consistent range of optimal operating conditions for velocity and temperature. Finally, combining all of these items listed above, we were able to reduce thickness non-uniformities and improve interfacial step density, our initial objectives.

6.2 Future Work on Specific Problem

The work here outlines methodology for solving multiscale optimization problems for layered heterostructure problems. It is admittedly somewhat limited in the fact that it only applies to steady-steady optimization problems, where ISAT is used for computational efficiency similar to [24]. Two types of time-dependent optimization problems are possible. The first type does not consider dynamic optimization of process conditions (temperatures and velocities would not change with time) but does consider the dynamic evolution of the macroscale process and connects this to the already time-dependent mesoscale simulations via a dynamic flux rate. The

second type allows for dynamic optimization of the process conditions while solving a process that could be either steady-state or time dependent. The latter is often reflected through changing reactant concentrations or locations during deposition [11, 24]. I am not particularly convinced that dynamic optimization of the process conditions would strengthen our problem, as we only consider one inlet, rendering it meaningless to solve the macroscale problem at steady-state. Even if we extend it to dynamic process conditions, surface deposition rates are driven by diffusion and velocity changes might not have significant effects, and changing the wafer surface temperature would introduce heat transfer complications within the wafer itself. Likely for this problem, the potential improvement in objectives is outweighed by the additional complications and computational cost.

Considering the time-dependent evolution of the macroscale problem, however, could improve our solution. The main improvements would primarily occur during transition periods ("ramp up" and "ramp down" time.) Step density is greatly affected by the composition of species being deposited as well as flux rates, particularly when the flux rates are low. This issue, however, does raise questions about the use of annealing time as a decision variable. The original intent was to use annealing time as a mesoscale control of step density. This objective is currently accomplished by separating annealing time from "ramp down" time, which is determined from macroscale input velocities and is therefore not controllable. If this line is blurred, however, process time becomes a macroscale input, leaving only temperature to control the mesoscale. Also, a time-dependent macroscale solution would also account for changes in velocities in addition to the concentration changes already discussed.

Also of interest is use of ISAT in the mesoscale timestepper. Currently, ISAT is used for entire mesoscale solutions. It can be also used to approximate the dynamic evolution of mesoscale solution by simulating small segments and then using tabulated information to approximate subsequent trajectories whenever possible [26]. This enables considerable

computational savings by not requiring costly simulations to be performed to completion. This advantage, however, is complicated by the following question: if the process evolution during one particular time step is dependent on its initial state, how is the initial state determined? With simulation, the previous state is already present. If some part of the process was not simulated, however, the initial state would have to be reconstructed. Methodology for doing so was outlined in [25] and used in [26]. The technique used was low-order approximations based on accurately capturing the ACF and HDCF of the media. Two lattice locations would then be randomly selected to compare whether switching their locations would decrease the error between the current ACF and HDCF, and the desired value. This technique was briefly explored in this work; generally it was found that obtaining a computationally expedient solution was not achievable. An additional complication for this problem is the multiple interfaces and species. It might not be necessary to capture the detail of all interfaces due to the assumption that the system is Markovian; that is the composition of unexposed layers does not affect the phenomena at the surface. When multiple species (and open sites) are present at the surface, however, the concept of using ACF and HDCF solely based on total height to account for reconstruction properties is no longer sufficient. One possibility might be to split the heights into two values for each species. A possible alternative approach is to reconstruct the media from fractional Brownian motion [54]; two-dimensional construction should be sufficient if it is applied layer-by-layer. This method, however, would necessitate starting with a distribution that already accounts for HDCF, since only ACF is covered in the reconstruction criteria. Other research has focused on fractal dimensions and the Hurst effect [55]. Finally, in order to circumvent the multiple species issue, it might be necessary to reconstruct the lattice based on a single species, then briefly simulate to determine the properties used for the new species.

The main bottleneck associated with implementing these changes is computational efficiency. A steady-state macroscale solution requires about four minutes of simulation time,

whereas a time-dependent solution requires about 15-30 (depending on how many different operating conditions are needed.) Much of this can be circumvented by using low-order methods based on KLE and the method of snapshots [27], the preferred methodology in [7, 24, 26]. Likewise, a single mesoscale simulation (already shortened to approximately its fastest computational state without inducing significant errors) requires about 25-30 minutes of computation time at moderate temperatures (800 K.) While noise is currently at a workable level for our application, in order to accurately predict the trajectories over short time intervals and to calculate the associated gradients, the lattice either needs to be much larger (a minimum of 250x250) or the results need to be averaged across several realizations (5-10) of 100x100 lattices in order to prevent bias. This would increase computation time about tenfold. With optimization trials already requiring several weeks for convergence, this methodology might take months if similar amounts of material need to be simulated (admittedly, this assumption is a worst-case scenario though.) Some reductions in computation time might be achievable through parallelization of the code; other possibilities to reduce computational overhead of the methodology might be explored as well.

It might also be interesting to investigate the problem if certain assumptions were not made. Sticking coefficients are assumed to be unity for all organometallic species here; exploring the effects of temperature and gas-phase composition at the wafer surface on sticking coefficients is an additional possibility. (Note: this may be limited by data availability, particularly for the $MH(CH_3)_2$ species, largely considered a mostly unimportant side product in [49] but found be to produced in large quantities in these macroscale simulations.) Also possibly worth investigating are rotation effects used to decrease thickness non-uniformities. This would introduce another variable (rotation speed,) however, and would increase the complexity of the optimization problem. Desorption effects, while interesting, would force bi-directional flow of information

and mandate iterative convergence between the macroscale and mesoscale solutions, and for that reason, are not likely to be explored.

6.3 Future Research Directions

Two main goals of this research are to propose methodology for handling microelectronics fabrication problems of increasing complexity, and to improve upon methodology currently used to solve such multiscale optimization problems. Several opportunities exist to explore increasingly complex phenomena to model in the microelectronics regime. These include, but are not limited to, films with complex geometries, porous thin films, films with lattice mismatches that form dislocations, and films with a number of different materials. Some research has been performed on control of porosity [56]; the step towards design and optimization is a future direction. The problem involving dislocations is particularly interesting, for a system (gallium nitride) that is notorious for forming them, the density is about $1x10^{8} – 2x10^{10}$ cm^{-2} [57-58]. Since molecular distances are on the order of angstroms, this length scale is too large to realistically capture from kMC simulations alone, possibly requiring the gaptooth method [59] or a similar technique to simulate accurately. Beyond the scope of microelectronics fabrication are those in the nanoscale regime. Multiscale optimization techniques that apply to lithography, etching, and top down versus bottom up approaches, might require considerably different formulations in the mesoscale regime [60].

Also of importance is improvement of the current methodologies. One possibility is reducing stochastic noise by means other than increasing the size of the lattice. A plethora of research exists on variance control and reduction, including some information particularly suited for kMC simulations in multiscale systems [61]. Such techniques would likely require sampling of the current simulations, and then use that information to determine the expectation of the

current trajectory and implement any corrections if the trajectory starts to drift away from its expected value. Other possibilities include further improvements in the lattice reconstruction methodology discussed in Section 6.2, new involvement of ISAT, and overall modeling, amongst others.

6.4 Concluding Remarks

Here, a methodology is outlined for optimization of spatially multiscale problems with macroscale and mesoscale components. It specifically deals with thin-film deposition, particularly to problems with multiple components (i.e. layered heterostructures,) maintaining the finite-element techniques used for macroscale reaction-transport mechanisms and stochastic simulations to model mesoscale surface interaction. Also increased is the computational efficiency by about an order of magnitude through tabulation and linear interpolation.

Appendix

Model Data

Macroscale Reaction Rate Data – Simplified Model

Reaction	*A (s^{-1})*	*E_a (kcal/mol)*
$Ga(CH_3)_3 + ½\ H_2 \rightarrow Ga(CH_3)_2 + CH_4$	$3.6x10^{15}$	68.9
$Ga(CH_3)_2 + ½\ H_2 \rightarrow GaCH_3 + CH_4$	$8.1x10^{14}$	28.2
$Ga(CH_3)_3 + H_2 \rightarrow GaH(CH_3)_2 + CH_4$	$1.1x10^{11}$	35.9
$GaH(CH_3)_2 \rightarrow GaCH_3 + CH_4$	$1.5x10^{13}$	59.5
$Al(CH_3)_3 + ½\ H_2 \rightarrow Al(CH_3)_2 + CH_4$	$3.4x10^{15}$	76.2
$Al(CH_3)_2 + ½\ H_2 \rightarrow AlCH_3 + CH_4$	$1.0x10^{15}$	39.8
$Al(CH_3)_3 + H_2 \rightarrow AlH(CH_3)_2 + CH_4$	$6.0x10^{10}$	29.9
$AlH(CH_3)_2 \rightarrow AlCH_3 + CH_4$	$2.1x10^{13}$	72.1

All reactions are in Arrhenius form: $k = Aexp(-E_a/RT)$. For the full model, see Ref. 49.

Chapman-Enskog Parameters

Species	*ε/k*	*σ*	*Species*	*ε/k*	*σ*
H_2	59.7	2.287	AsH_3	259.8	4.145
$Ga(CH_3)_3$	378	5.52	$Al(CH_3)_3$	458.9	5.04
$GaCH_3$	972	4.92	$AlCH_3$	295.2	3.49

A variety of methods exist for determining diffusion coefficients: we chose to use Equations 11-3.2 and 11-3.6 on p. 582-583 in Ref. 53. We set A=B to calculate one-species diffusion coefficients, since binary diffusion coefficients are calculated through Comsol. Data was unavailable for $M(CH_3)_2$ and $MH(CH_3)_2$ species; as a result, we estimate these diffusion coefficients as the average value of the $M(CH_3)_3$ and MCH_3 species. It may be necessary to solve this problem initially with constant diffusivity values (1 m^2/s) in order to obtain initial conditions that are suitable for the Comsol solver using the diffusivities in the model. Sources: Ref. 51-53.

Other Transport Parameters

Thermal conductivity, k:

$3.777 \times 10^{-4} T + 0.07444$ [W / (m K)]

Heat capacity, C_p:

$13849.04 + 1.69452\ T$ [J / (kg K)]

Dynamic viscosity, μ:

$1 \times 10^{-8}(1.6029\ T + 453.53)$ for H_2 [Pa s]

$1 \times 10^{-8}(5.1657\ T + 82.19)$ for AsH_3 [Pa s]

Sources: Ref. 48 (for thermal conductivity), Ref. 50 (for heat capacity and viscosity.)

Kinetic Monte Carlo Migration Rate Data

$$E^{(M)}(N) = \begin{cases} N \cdot E_1^{(M)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots N = 0,\ldots,4; \\ 4 \cdot E_1^{(M)} + (N-4) \cdot E_2^{(M)} \ldots\ldots\ldots\ldots N = 4,\ldots,12; \end{cases}$$

$$\Gamma^{(M)}(N) = k_0 \cdot \exp\left(\frac{-E^{(M)}(N)}{k_B \cdot T} \right)$$

$k_0 = \Gamma_0 = 8.25e12 \cdot s^{-1}$

$E_1^{Ga} = 0.41eV$

$E_2^{Ga} = 0.24eV$

$E_1^{Al} = 0.45eV$

$E_2^{Al} = 0.26eV$

N is the number of nearest neighbors, and k_B is the Boltzmann constant. Sources: Ref. 44, 46, 47.

Lattice Constant Information

GaAs lattice constant: 5.653 Å

AlAs lattice constant: 5.661 Å

1 Monolayer GaAs = 2.599×10^{-6} mol/m^2 GaAs

1 Monolayer AlAs = 2.591×10^{-6} mol/m^2 AlAs

Sources: Ref. 1 and 41

Table **A-1**: Decay Rate Approximations: 0.2 second intervals: GaAs

Distance from Center (cm)	v = 0.1 m/s	v = 0.2 m/s	v = 0.4 m/s	v = 0.8 m/s
0	0.9054	0.8456	0.6769	0.4624
0.25	0.9059	0.8465	0.6787	0.4634
0.5	0.9072	0.8489	0.6841	0.4668
0.75	0.9095	0.8529	0.6927	0.4724
1	0.9125	0.8580	0.7039	0.4800
1.25	0.9161	0.8641	0.7175	0.4897
1.5	0.9202	0.8710	0.7328	0.5012
1.75	0.9247	0.8783	0.7493	0.5144
2	0.9294	0.8859	0.7666	0.5295
2.25	0.9342	0.8935	0.7843	0.5464
2.5	0.9390	0.9011	0.8021	0.5658
2.75	0.9437	0.9085	0.8196	0.5884
3	0.9483	0.9155	0.8367	0.6153
3.25	0.9527	0.9222	0.8531	0.6478
3.5	0.9573	0.9287	0.8692	0.6876

Table **A-2**: Decay Rate Approximations: 0.2 second intervals: AlAs

Distance from Center (cm)	v = 0.1 m/s	v = 0.2 m/s	v = 0.4 m/s	v = 0.8 m/s
0	0.8768	0.8476	0.7932	0.6107
0.25	0.8771	0.8479	0.7938	0.6122
0.5	0.8779	0.8489	0.7955	0.6167
0.75	0.8791	0.8506	0.7983	0.6241
1	0.8808	0.8528	0.8020	0.6342
1.25	0.8830	0.8556	0.8065	0.6468
1.5	0.8854	0.8588	0.8116	0.6612
1.75	0.8882	0.8623	0.8171	0.6770
2	0.8912	0.8661	0.8230	0.6939
2.25	0.8944	0.8700	0.8290	0.7114
2.5	0.8977	0.8740	0.8351	0.7291
2.75	0.9012	0.8781	0.8410	0.7469
3	0.9048	0.8821	0.8468	0.7643
3.25	0.9089	0.8863	0.8524	0.7812
3.5	0.9138	0.8910	0.8581	0.7979

To use Tables **A-1** and **A-2**, multiply the value in the table by the current deposition rate over the 0.2 s interval. For example, for GaAs at 0.2 m/s at 1.25 cm from the center, if the full deposition rate was 1 Monolayer/s, it would be 0.8641 Monolayers/s from 0-0.2 s once the inlet organometallic flux is turned off, 0.8641^2 = 0.7467 Monolayers/s from 0.2-0.4 s, 0.8641^3 = 0.6452 Monolayers/s from 0.4-0.6 s, and so on. Intermediate value velocities can be interpolated.

References

1. Ohring, M., *Materials Science of Thin Films, Deposition and Structure*. 2nd ed. 2002: Academic Press.
2. Raimondeau, S., Vlachos, D.G., *Recent developments on multiscale, hierarchial modeling of chemical reactors.* Chemical Engineering Journal, 2002. **90**: p. 3-23.
3. O'Connell, S.T., Thompson, P.A., *Molecular dynamics-continuum hybrid computations: a tool for studying complex fluid flows.* Physical Review E, 1995. **52**(6): p. 5792-5795.
4. Li, J., Liao, D., Yip, S., *Nearly exact solution for coupled continuum/MD fluid simulation.* Journal of Computer-Aided Materials Design, 1999. **6**: p. 95-102.
5. Hadjiconstantinou, N.G., *Hybrid atomistic-continuum fomulations and the moving contact line problem.* Journal of Computational Physics, 1999. **154**: p. 245-265.
6. Broughton, J.Q., Abraham, F.F., Bernstein, N., Kaxiras, E., *Concurrent coupling of length scales: methodology and application.* Physical Review B, 1999. **60**(4).
7. Raimondeau, S., Vlachos, D.G., *Low-dimensional approximations of multiscale epitaxial growth models for microstructure control of materials.* Journal of Computational Physics, 2000. **160**: p. 564-576.
8. Jensen, K.F., Rodgers, S.T., Venkataramani, R., *Multiscale modeling of thin film growth.* Current Opinion in Solid State and Materials Science, 1998. **3**: p. 562-569.
9. Christofides, P.D., Armaou, A., *Control and optimization of multiscale process systems.* Computers and Chemical Engineering, 2006. **30**: p. 1670-1686.
10. Itle, G.C., Salinger, A.G., Pawlowski, R.P., Shadid, J.N., Biegler, L.T., *A tailored optimization strategy for PDE-based design: application to a CVD reactor.* Computers and Chemical Engineering, 2004. **28**: p. 291-302.
11. Varshney, A., Armaou, A., *Optimal operation of GaN thin film epitaxy employing control vector parameterization.* Ind. Eng. Chem. Res., 2006. **52**(4): p. 1378-1391.
12. Balsa-Canto, E., Banga, J.R., Alonso, A.A., Vassiliadis, V.S., *Dynamic optimization of distributed parameter systems using second-order directional derivatives.* Ind. Eng. Chem. Res., 2004. **43**(21): p. 6756-6765.
13. Biegler, L.T., Nocedal, J., Schmid, C., *A reduced Hessian method for large-scale constrained optimization.* SIAM Journal of Optimization, 1995. **5**(2): p. 314-347.
14. Vasanthajaran, S., Viswanathan, J., Biegler, L.T., *Reduced successive quadratic programming implementation for a large-scale optimization problems with smaller degrees of freedom.* Computers and Chemical Engineering, 1990. **14**(8): p. 907-915.
15. Vassiliadis, V.S., Sargent, R.W.H., Pantelides, C.C., *Solution of a class of multistage dynamic optimization problems, Part I. problems without path constraints.* Industrial and Engineering Chemistry Research, 1994. **33**: p. 2111-2122.
16. Meyer, C.A., Floudas, C.A., Neumaier, A., *Global optimization with nonfactorable constraints.* Industrial and Engineering Chemistry Research, 2002. **41**(25): p. 6413-6424.
17. Kelley, C.T., Sachs, E.W., *Truncated Newton methods for optimization with inaccurate functions and gradients.* Journal of Optimization Theory and Applications, 2003. **116**(1): p. 83-98.
18. Kelley, C.T., *Iterative Methods for Optimization.* 1999: SIAM.

19. Arora, N., Biegler, L.T., *Parameter estimation for a polymerization reactor model with a composite-step trust-region NLP algorithm.* Industrial and Engineering Chemistry Research, 2004. **43**(14): p. 3616-3631.
20. Caballero, J.A., Grossmann, I.E., *An algorithm for the use of surrogate models in modular flowsheet optimization.* AIChE Journal, 2008. **54**(10): p. 2633-2650.
21. Jung, J.Y., Blau, G., Pekny, J.F., Reklaitis, G.V., Eversdyk, D., *A simulation based optimization approach to supply chain management under demand uncertainty.* Computers and Chemical Engineering, 2004. **28**: p. 2087-2106.
22. Theodoropoulou, A., Adomaitis, R.A., Zafiriou, E., *Model reduction for optimization of rapid thermal chemical vapor deposition systems.* IEEE Transactions on Semiconductor Manufacturing, 1998. **11**: p. 85-98.
23. Theodoropoulou, A., Zafiriou, E., Adomaitis, R.A., *Inverse model-based real time control for temperature uniformity of RTCVD.* IEEE Transactions on Semiconductor Manufacturing, 1999. **12**: p. 87-101.
24. Varshney, A., Armaou, A., *Multiscale optimization using hybrid PDE/kMC process systems with application to thin film growth.* Chemical Engineering Science, 2005. **60**: p. 6780-6794.
25. Varshney, A., Armaou, A., *Idenitification of macroscopic variables for low-order modeling of thin-film growth.* Industrial and Engineering Chemistry Research, 2006. **45**: p. 8290-8298.
26. Varshney, A., Armaou, A., *Reduced order modeling and dynamic optimization of multiscale PDE/kMC process systems.* Computers and Chemical Engineering, 2008. **32**: p. 2136-2143.
27. Sirovich, L., *Turbulence and the dynamics of coherent structures Part I: coherent structures.* Quarterly of Applied Mathematics, 1987. **45**(3): p. 561-571.
28. Fichthorn, K.A., Weinberg, W.H., *Theoretical foundations of dynamical Monte Carlo simulations.* Journal of Chemical Physics, 1991. **95**(2): p. 1090-1096.
29. Shitara, T., Vvedensky, D.D., Wilby, M.R., Zhang, J., Neave, J.H., Joyce, B.A., *Step-density variations and reflection high-energy electron-diffraction intensity oscillations during epitaxial growth on vicinal GaAs(001).* Physical Review B, 1992. **46**(11): p. 6815-6827.
30. Zhou, X.W., Wadley, H.A.G., Johnson, R.A., Larson, D.J., Tabat, N., Cerezo, A., Petford-Long, A.K., Smith, G.D.W., Clifton, P.H., Martens, R.L., Kelly, T.F., *Atomic scale structure of sputtered metal multilayers.* Acta Materiala, 2001. **49**: p. 4005-4015.
31. Armaou, A., Kevrekidis, I.G., *Equation-free optimal switching policies using coarse timesteppers.* International Journal of Robust and Nonlinear Control, 2005. **15**: p. 713-726.
32. Gear, C.W., Hyman, J.M., Kevrekidis, P.G., Runborg, O., Theodoropoulos, K., Kevrekidis, I.G., *Equation-free multiscale computation: enabling microscopic simulators to perform system-level tasks.* Communications in Mathematical Sciences, 2003. **1**: p. 715-762.
33. Siettos, C.I., Armaou, A., Makeev, A.G., Kevrekidis, I.G., *Microscopic/stochastic timesteppers and coarse control: a kinetic Monte Carlo example.* AIChE Journal, 2003. **49**: p. 1922-1926.
34. Lou, Y., Christofides, P.D., *Estimation and control of growth rate and surface roughness in thin film growth using kinetic Monte-Carlo models.* Chemical Engineering Science, 2003. **58**(14): p. 3115-3129.

35. Katsoulakis, M.A., Majda, A.J., Vlachos, D.G., *Coarse-grained stochastic processes and Monte Carlo simulations in lattice systems.* Journal of Computational Physics, 2003. **186**: p. 250-278.
36. Gallivan, M.A., Murray, R.M., *Reduction and identification methods for Markovian control systems, with application to thin film deposition.* International Journal of Robust and Nonlinear Control, 2004. **14**: p. 113-132.
37. Pope, S.B., *Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation.* Combustion Theory and Modelling, 1997. **1**: p. 41-63.
38. Lu, L., Pope, S.B., *An improved algorithm for in situ adaptive tabulation.* Journal of Computational Physics, 2009. **228**(2): p. 361-386.
39. Glasserman, P., Yao, D.D., *Some guidelines and guarantees for common random numbers.* Management Science, 1992. **38**(6): p. 884-908.
40. Zazanis, M.A., Suri. R., *Convergence rates of finite-difference sensitivity estimates for stochastic systems.* Operations Research, 1993. **41**(4): p. 694-703.
41. *Aluminum Arsenide.* National Compound Semiconductor Roadmap 2006; Available from: http://www.onr.navy.mil/sci_tech/31/312/ncsr/materials/alas.asp.
42. Suzuki, Y., Okamoto, H., *Refractive index of GaAs-AlAs superlattice grown by MBE.* Journal of Electronic Materials, 1983. **12**(2): p. 397-411.
43. Chang, S.J., Chang, C.S., Su, Y.K., Chang, P.T., Wu, Y.R., Huang, K.H., Chen, T.P., *Chirped GaAs-AlAs distributed Bragg reflectors for high brightness yellow-green light-emitting diodes.* IEEE Photonics Technology Letters, 1997. **9**(2): p. 182-184.
44. Grosse, F., Zimmermann, R., *Monte Carlo growth simulation for $Al_xGa_{1-x}As$ heteroepitaxy.* Journal of Crystal Growth, 2000. **212**: p. 128-137.
45. Reklaitis, G.V., Ravindran, A., Ragsdell, K.M., *Engineering optimization: methods and applications.* 1983: John Wiley & Sons.
46. Siano, J.D., Armaou, A., *Report on Step Density (untitled)*. 2007.
47. Dekorsy, T., Helm, M., Yakolev, V.A., Seidel, W., Keilmann, F. *Dispersion of the nonlinear susceptibility of GaAs in the THz range*. in *Optical Society of America.*
48. Mountziaris, T.J., Jensen, K.F., *Gas-phase and surface reaction mechanisms in MOCVD of GaAs with trimethyl-gallium and arsine.* Journal of the Electrochemical Society, 1991. **138**(8): p. 2426-2439.
49. Cavallotti, C., Lengyel, I., Nemirovskaya, M., Jensen, K.F., *A computational study of gas-phase and surface reactions in deposition and etching of GaAs and AlAs.* Journal of Crystal Growth, 2004. **268**: p. 76-95.
50. Perry, R.H., Green, D.W., *Perry's Chemical Engineers' Handbook*. 7th ed. 1997: McGraw-Hill.
51. Im, I.T., Sugiyama, M., Shimogaki, Y., Nakano, Y., *A numerical study on heat transfer and film growth rate of InP and GaAs MOCVD process.* Journal of Crystal Growth, 2005. **276**: p. 431-438.
52. Mihopoulos, T.G., Gupta, V., Jensen, K.F., *A reaction-transport model for AlGaN MOVPE growth.* Journal of Crystal Growth, 1998. **195**: p. 733-739.
53. Reid, R.C., Prausnitz, J.M., Poling, B.E., *The Properties of Gases and Liquids*. 4th ed. 1987: McGraw-Hill.
54. Kikkinides, E.S., Burganos, V.N., *Structural and flow properties of binary media generated by fractional Brownian motion models.* Physical Review E, 1999. **59**(6): p. 7185-7194.
55. Gneiting, T., Schlather, M., *Stochastic models that separate fractal dimension and the Hurst effect.* SIAM Review, 2004. **46**(2): p. 269-282.

56. Hu, G., Orkoulas, G., Christofides, P.D., *Stochastic modeling and simultaneous regulation of surface roughness and porosity in thin film deposition.* Industrial and Engineering Chemistry Research, 2009. **48**(14): p. 6690-6700.
57. Harafuji, K., Tsuchiya, T., Kawamura, K., *Molecular dynamics simulation of dislocations in wurtzite-type GaN crystal.* Journal of Applied Physics, 2004. **96**(5): p. 2513-2524.
58. Qian, W., Skowronski, M., De Graef, M., Doverspike, K., Rowland, L.B., Gaskill, D.K., *Microstructural characterization of α-GaN films grown on sapphire by organometallic vapor phase epitaxy.* Applied Physics Letters (American Institute of Physics) 1995. **66**(10): p. 1252-1254.
59. Armaou, A., Kevrekidis, I.G., Theodoropoulos, C., *Equation-free gaptooth-based controller design for distributed complex/multiscale processes.* Computers and Chemical Engineering, 2005. **29**(4): p. 731-740.
60. Madou, M.J., *Fundamentals of microfabrication: the science of miniaturization.* 2nd ed. 2002: CRC Press.
61. Papavasiliou, A., Kevredikis, I.G., *Variance reduction for the equation-free simulation of multiscale stochastic systems.* SIAM: Multiscale Model Simulation, 2007. **6**(1): p. 70-89.