Enhancement of Numerical Computations of the Wigner-Poisson Equations for Application to the Simulation of THz-Frequency RTD Oscillators

Matthew S. Lasater^a, C. T. Kelley^a, Andrew G. Salinger^b, Dwight Woolard^c, and Peiji Zhao^d

ABSTRACT

Resonant tunneling diodes (RTDs) are ultra-small semiconductor devices that have potential as very high frequency oscillators. To describe the electron transport within these devices, the Wigner-Poisson Equations are used. These equations incorporate quantum mechanics to describe how the electron distribution changes in time due to kinetic energy, potential energy, and scattering effects. To study the RTD, we apply numerical continuation methods to calculate the steady-state electron distribution as the voltage difference across the RTD varies. To implement the continuation methods, the RTD simulator is interfaced to LOCA (Library of Continuation Algorithm), a software library that is a part of Sandia National Laboratories' parallel solver package, Trilinos. With more sophisticated numerical solvers, we are able to calculate solutions on finer grids that were previously too computationally intensive. This is very important to allow for detailed studies of correlation effects which may dramatically influence oscillatory behavior in RTD-based devices. The more accurate results derived from this work reveal new physical effects that were absent in prior studies. Hence, these physics-based and more refined numerical simulations will provide new insights and greatly facilitate the future optimization of RTD-based oscillator sources and thus has important relevance to THz-frequency-regime based spectroscopic sensing technology.

Keywords: Resosant tunneling diodes, THz current oscillation, parameter study

1. INTRODUCTION

Resonant tunneling diodes (RTDs) have been widely investigated for 30 years due to their unique physical properties and their potential application as very high-speed components. The primary motivation for the studies presented here is to provide insight into the time-dependent behavior of RTDs, since they have relevance to sensors and systems that operate at terahertz (THz) frequencies. In particular, new and more robust THz sources are needed for the development of spectroscopic sensors and to enable high-speed systems that can be used for data-processing and communications. For example, there are long-standing and ongoing research efforts^{1,2} that have demonstrated the potential of THz sensing technology for the detection and identification of biological and chemical agents, along with the serious need for making advances to the state-of-the-art in THz electronics. Therefore, a modeling and simulation capability that can be used to accurately and efficiently study the oscillatory behavior of RTDs can be extremely valuable for enabling new and novel types of THz sensors and systems.

Further author information: (Send correspondence to Matthew Lasater)

Matthew Lasater: E-mail: mslasate@unity.ncsu.edu

Sandia is a multiprogram laboratory operated by Sandia corporation, a Lockheed-Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

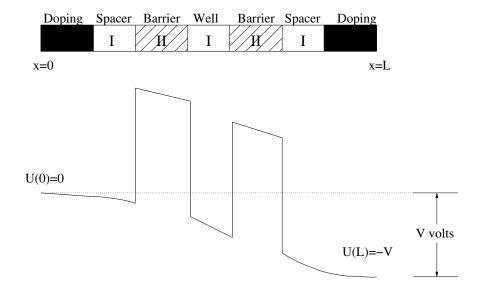


Figure 1. Diagram of RTD and the Electric Potential U(x) Inside

The basic building blocks of all resonant tunneling structures and the RTD device is a double-barrier quantum-well structure. Specifically, the basic RTD is constructed by placing a thin layer of a narrow bandgap semiconductor between two thin layers of a different semiconductor that has a wider electronic bandgap characteristic. Since the outer semiconductor layers have a larger band-gap, this naturally leads to a device with a potential well in the center of the conduction band as depicted in the bottom of Figure 1. Here, the material types for both the well and spacer regions (i.e., that define the conduction paths between each barrier and the external contacts) are labeled as "I" and the barriers are labeled as "II" see top of Figure 1. At some distance from the barriers, the "I-type" semiconductor is artificially doped (represented by the darker lines) to define an electronic reservoir which allows for electronic injection into the quantum mechanical system when an electric potential is applied to the device. Here, doping is where atoms that contain more (or less) electrons than the base semiconductor itself are embedded into the semiconductor to create (or take away) extra electrons in the structure. For these studies, n-type doping is considered where extra electrons are available for the transport processes.

Classically, if a particle is incident to one of the potential barriers and it does not have enough total energy (i.e., speed and/or potential), it will be reflected backwards Since this is a quantum mechanical system and the electrons exhibit wave-like behavior, electrons with energies less than the barriers still has some probability of passing through the barrier. This effect is known as "quantum tunneling" and is the basis of this device. If a voltage difference is applied across the device, electrons will start to move along the device, tunnel through the barriers, and reach the other side, thus creating a current.

In addition, numerical simulations^{3,4} have previously shown that current oscillation can be expected for certain voltage differences, and that these current oscillations have frequencies in the terahertz (THz) regime. However, there are know correlation effects that can have strong influence on both the amplitude and frequency of the resulting oscillations. Therefore, this research seeks to establish enhanced computational capabilities that can be used by engineers and scientists to fully understand what physical mechanism creates these intrinsic oscillations and to determine what physical parameters (i.e. doping profile, barrier height and width, well width, etc.) are conducive to sustaining and controlling these oscillations in hopes of producing a viable high frequency power source. In particular, this work seeks to establish a faster and more accurate RTD simulator to aid the engineers in these goals.

2. MODEL DESCRIPTION

The model used to describe the electron transport in these devices is the Wigner-Poisson Equations.⁵ These equations consist of a nonlinear PDE that describes the time-evolution of the distribution of the electrons in the

device coupled with Poisson's equation which incorporates the potential effects of the electrons into the model. The first of these equations can be given by

$$\frac{\partial f}{\partial t} = W(f) = K(f) + P(f) + S(f) \tag{1}$$

Here, f = f(x, k, t), is the distribution of the electrons. It is a function of the position of the electron, x, the momentum of the electron, k, and time, t. The position variable x ranges from 0 to L, the length of the device, and the momentum variable k ranges from $-\infty$ to ∞ . The time-derivative of f is comprised of three terms. The first term, K(f), represents the kinetic energy effects on the distribution and is given by

$$K(f) = \frac{hk}{2\pi m^*} \frac{\partial f}{\partial x} \tag{2}$$

Here, h is Planck's constant and m^* is the effective mass of the electron. The second term, P(f), is the nonlinear term in the equation and is for the potential energy effects on the distribution

$$P(f) = -\frac{4}{h} \int_{-\infty}^{\infty} f(x, k') T(x, k - k') dk'$$
(3)

The function T(x,z) is defined by

$$T(x,z) = \int_0^{\frac{L_c}{2}} [U(x+y) - U(x-y)] \sin(2*x*z) dy \tag{4}$$

In this equation, U(x) is the electric potential as a function of position, and L_c is the correlation length. This term is nonlinear in f since U(x) depends on f through Poisson's equation. The final part of the time-derivative describes the scattering processes that occur between the electrons

$$S(f) = \frac{1}{\tau} \left[\frac{\int_{-\infty}^{\infty} f(x, k') dk'}{\int_{-\infty}^{\infty} f_0(x, k') dk'} f_0(x, k) - f(x, k) \right]$$
 (5)

Here, τ is the relaxation time, and $f_0(x, k)$ is the equilibrium Wigner distribution. This is the steady state solution to Equation (1) when there is no voltage difference across the device. The boundary conditions for f impose what the incoming electron distributions are. That is, at x = 0 and for k > 0 (electrons with positive momentum that are moving right) we have

$$f_0(0,k) = \frac{4\pi m^* k_B T}{h^2} ln(1 + exp\left[\frac{1}{k_B T} \left(\frac{h^2 k^2}{8\pi^2 m^*} - \mu_0\right)\right])$$
 (6)

and at x = L and for k < 0 (electrons with negative momentum that are moving left) we have

$$f_0(L,k) = \frac{4\pi m^* k_B T}{h^2} ln(1 + exp\left[\frac{1}{k_B T} \left(\frac{h^2 k^2}{8\pi^2 m^*} - \mu_L\right)\right])$$
 (7)

 k_B is Boltzmann's constant, T is the temperature, μ_0 is the Fermi energy at x=0, and μ_L is the Fermi energy at x=L.

The electric potential U(x) is made up of two parts. The first part is from the electrostatic potential created by the electrons in the device. We will denote this part by u(x). The second part is from the potential barriers in the device created from the heterojunction of the two different semiconductor materials. We will denote this part by $\Delta_c(x)$. To get u(x), we must solve Poisson's equation

$$\frac{d^2u}{dx^2} = \frac{q^2}{\epsilon} \left[N_d(x) - \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x, k') dk' \right]$$
 (8)

q is the charge of the electron, ϵ is the dielectric constant, and $N_d(x)$ is the doping profile. The boundary conditions for Poisson's equation are where the voltage difference across the device is incorporated. We have that

$$u(0) = 0, u(L) = -V (9)$$

where $V \ge 0$ is the applied voltage. Once we have solved for u(x), we have $U(x) = u(x) + \Delta_c(x)$.

3. DISCRETIZATION

To numerically solve for the distribution, we discretize both the domain and equations using a finite difference method. For the x-domain, we use N_x grid points where $x_i = (i-1)*\Delta x$, $i=1,2,\ldots,N_x$ and $\Delta x = \frac{L}{N_x-1}$. These grid points are evenly spaced across [0,L]. For the k-domain, we first truncate from $-\infty$ to ∞ to $-K_M$ to K_M , where K_M is a maximum momentum we consider. We use N_k grid points where $k_j = (2j-N_k-1)*\frac{\Delta k}{2}$, $j=1,2,\ldots,N_k$ and $\Delta k = \frac{2K_M}{N_k}$. These grid points are evenly spaced across $(-K_M,K_M)$. So numerically we want to compute an approximation to the distribution at each grid point. That is for each $i=1,2,\ldots,N_x$ and $j=1,2,\ldots,N_k$, calculate a f_{ij} such that $f_{ij}\approx f(x_i,k_j)$.

To approximate the spatial derivative term in Equation (2), we use a second-order upwind differencing scheme. For the integral terms in Equation (3), Equation (4), and Equation (5), we use quadrature formulae for their approximations. Finally, for solving Poisson's equation, we use a standard three-point central differencing scheme. This discretization converts the continuous nonlinear PDE problem to a finite dimensional nonlinear ODE problem for the solution of f at the grid points. The dimension of the ODE approximations is $m = N_x * N_k$.

4. CONTINUATION METHODS

4.1. Analyzing the RTD's Behavior for Various Applied Voltage Differences

Physicists and electrical engineers want to exploit the quantum tunneling effects present in an RTD to create very fast current oscillations. From previous numerical simulations,³ researchers expect the oscillation frequency to be in the THz regime. Given a voltage difference V across the RTD, the Wigner-Poisson model can predict current output from the device in this operating condition. This is determined by computing the steady-state solution to the Wigner equation, which corresponds to finding a distribution of electrons f such that $\frac{\partial f}{\partial t} = W(f) = 0$. Since the operating conditions of the RTD need to be understood for a range of voltage differences V, the collection of steady-state distributions for each of these voltages will be denoted by f(V). The collection of f(V) can be thought of as a curve of electron distributions that are parameterized by the scalar parameter V, which is the voltage difference across the device. Ideally, we want to efficiently compute f(V) for various applicable voltage differences to get an understanding for how the current output changes as the voltage varies. The collection f(V) satisfies the equation

$$\frac{\partial f}{\partial t} = W(f(V), V) = 0 \tag{10}$$

In Equation (10), we explicitly write out the Wigner equation's dependence on the voltage difference V through Poisson's equation. For a fixed voltage difference, call it V = V*, a nonlinear equation must be solved to locate the steady-state distribution f(V*). We solve the nonlinear equation with Newton's method.

4.2. Newton's Method

Newton's method solves the nonlinear equation G(z) = 0, where z and G(z) are vectors in R^m . We will denote by G'(z) the Jacobian matrix of G(z), which is a m by m matrix that represents the derivative of G with respect to the vector variable z. If there is a solution z^* , if G is a well-behaved function, and if an initial vector z^0 that is close to z^* is given, then Newton's method produces iterates z^n that converge to z^* . The iterates are given by

$$z^{n+1} = z^n + s^n = z^n - G'(z^n)^{-1}G(z^n), n = 0, 1, 2, \dots$$
(11)

The Newton step, s^n , is the update or correction to the previous iterate z^n to create the new iterate z^{n+1} and is the solution to the linear equation

$$G'(z^n)s^n = -G(z^n) \tag{12}$$

To solve this linear system in practice, we use an iterative method called GMRES.⁶ This method is a matrix-free method. The advantages of using a matrix-free method are that the coefficient matrix is not needed, only the action of the coefficient matrix on the vector. This saves us from computing the Jacobian matrix at each nonlinear solve, factoring and inverting the Jacobian, and storing this matrix and its inverse. We accelerate the convergence of GMRES my multiplying both sides of Equation (12) my another matrix M (called a preconditioner), so that the new coefficient matrix $MG'(z^n)$ is an "easier" matrix for GMRES to handle.

The normal criteria for terminating the Newton iteration is when a maximum number of iterations have been taken or when the size of the vector $G(z^n)$ is below a specified tolerance.

4.3. Natural Continuation

Continuation methods attempt to trace out the curve of steady-state solutions f(V) for a range of parameters V, where f(V) is the solution to the nonlinear equation given in Equation (10). After discretizing the Wigner-Poisson equations to numerically solve the model, the computed f(V) are the steady-state solutions to the nonlinear m-dimensional differential equation for the approximation of the distribution at nodal points in (x,k) space. Here, $m = N_x * N_k$ is the total number of unknowns or the number of nodal points. Each continuation method discussed in this paper uses Newton's method to solve the nonlinear equation required at each parameter value V. The continuation methods differ in how the parameter is varied as f(V) is computed. The parameter that is varied, in this case V, is also called the continuation parameter since the steady-state solution collection is being continued in the parameter V. Numerically, continuation methods produce a sequence of parameter values V_n and a corresponding sequence of steady-state solutions f_n . Given an initial parameter value V_0 , an initial steady-state distribution f_0 which satisfies Equation (10) for when $V = V_0$, the natural continuation method increases or decreases the parameter V by some known increment ΔV so that $V_{n+1} = V_n + \Delta V$ and f_{n+1} is the solution to Newton's Method applied to Equation (10) where the parameter V is set to V_{n+1} . The initial iterate for Newton's method is the previously computed distribution f_n .

While the natural continuation method is relatively easy to implement, it is not designed to handle the curve f(V) turns back around on itself. The natural continuation methods either monotonically increases or monotonically decreases the parameter V. When the curve f(V) turns around on itself, the parameter value at which the turn occurs, here denoted by V_{turn} , is called a turning point. For some parameter values nearby the turning point, that is $V \approx V_{turn}$, there are two portions of f(V) since the curve has just turned around, and this degrades the performance of the nonlinear solver. The linear solves in Newton's method become harder to perform since as V approaches the critical value V_{turn} , the Jacobian matrix becomes more ill-conditioned. At $V = V_{turn}$, the Jacobian matrix is not even invertible. Since hysteric effects in the current / voltage relationship have been previously established with the RTD, the multiple solutions at a fixed parameter value imply turning points can be expected. To handle the just described numerical problems that arise in natural continuation, we change continuation parameters.

4.4. Pseudo-Arc Length Continuation

Pseudo-arc length continuation attempts to follow the curve f(V) in terms of its arc length from an initial point on the curve, specified by an initial parameter V_0 and initial distribution f_0 . We will let s denote the arc length of the curve f(V), and it will be the new continuation parameter. Since the voltage V and steady-state distribution curve f(V) are parameterized by s, given by V(s) and f(s), V is now a variable that must be solved in the nonlinear equation

$$W(f(s), V(s)) = 0 (13)$$

Let the m+1 dimensional vector z represent the combined m dimensional distribution f and the 1 dimensional scalar variable V. Applying Newton's method to Equation (13), gives the iteration

$$z^{n+1} = z^n + \Delta z^n \tag{14}$$

where the Newton step Δz^n solves the linear equation

$$W'(z^n)\Delta z^n = -W(z^n) \tag{15}$$

Since $W(z^n)$ is a m dimensional vector and Δz^n is a m+1 dimensional vector, $W'(z^n)$ is a m by m+1 matrix. Since $W'(z^n)$ is not a square matrix, the linear system does not have a unique solution. The system is in fact underdetermined, and an additional constraint is needed on the linear system. In pseudo-arc length continuation⁸ the additional constraint is the arc length equation

$$n(z) = t^T z = 0 (16)$$

The vector t is the unit vector in the director of the vector $\frac{\partial z}{\partial s}$. Geometrically, this constraint enforces the Newton step to be perpendicular to the current tangent vector of the curve f(s) at the previously computed

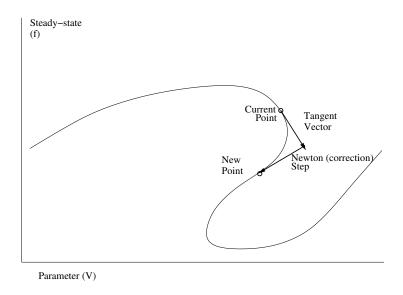


Figure 2. Example of One Step of Pseudo-Arc Length Continuation

point f_n . Figure 2 shows a contrived example to illustrate a step of pseudo-arc length continuation in the case of m = 1.

The augmented nonlinear system has m+1 constraints when n(z)=0 is added to the m dimensional vector W(z). The new square linear system of size m+1 to solve for the Newton step is given in matrix - vector form as

$$n(z) = t^T z = 0 (17)$$

here are two ways of handling the solution of this augmented linear system. The two methods will be discussed in the next section along with numerical results.

5. NUMERICAL ALGORITHM COMPARISON

5.1. LOCA: Library of Continuation Algorithms

The continuation algorithms discussed in the previous section are implemented with our computational model using LOCA, Library Of Continuation Algorithms. This C++ library is part of the parallel solver packages developed at Sandia National Laboratories called Trilinos.⁸ This package makes use of several other packages within Trilinos. These include a nonlinear solver NOX, a linear solver AztecOO, and an eigensolver Anasazi. LOCA allows users to choose between the two linear system algorithms when using pseudo-arc length continuation.

5.2. First Method: Bordering Method

To use the bordering method to solve the linear system in Equation (17), two linear solves involving W'(f) as a coefficient matrix are needed. The resulting two vector solutions are used to compute the step.⁹ As mentioned in the natural continuation section, the Jacobian matrix will become ill-conditioned near the turning point. So these two linear solves will be computationally hard to perform if the voltage parameter V is near V_{turn} . The extra constraint defined in Equation (16) should enable the algorithm to jump over the turning point.

5.3. Second Method: Householder Method

Define a projection from P from R^m to R^{m+1} that preserves the length of the vector, but includes no component in the span of the tangent vector t. This can be done quickly by using Householder reflections. So for any vector in R^m , it projection under P will automatically be perpendicular to the tangent vector t. So now the composition of W'(z) with the projection P will be an invertible map. It will only require one linear solve with

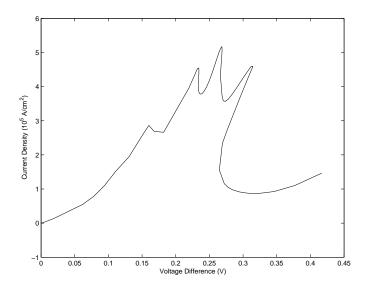


Figure 3. Current - Voltage Plot for RTD

W'(f) as the coefficient matrix.¹¹ Since this method requires half as many linear solves as the bordering method, the only way the Householder method could be slower if the construction and application of the projection P is inefficient.

5.4. Numerical Comparision

The numerical experiments comparing the two linear solve methods were performed on 20 processors of a Linux cluster at Sandia National Laboratories. This cluster has 236 nodes; each node is a duel 3.06 GHz Xenon processor with 2 GB of RAM. The example shows a 1 million unknown problem ($N_x = 512$, $N_k = 2048$) going through 7 continuation steps in arc length. The table below compares the two linear solver method through different criteria. These include number of linear iterations, number of nonlinear iterations, and total time of execution.

 Table 1. Comparison Of Linear Solve Methods

Algorithm	Total Linear Iterations	Total Nonlinear Iterations	Total Time (min)
Bordering	8826	21	98.4
Householder	3026	15	33.6

From this comparison, it is apparent that the Householder method for solving the linear systems arising in pseudo-arc length continuation is much more efficient for this problem. There is a factor of 3 savings in computational time when the total runtimes are compared for the 7 nonlinear solves. The current density/voltage plot for the 1 million unknown problem is shown in Figure 3. The boxed in area is the 7 continuation steps analyzed in the numerical comparison. This area is blown up and presented in Figure 4. The current is measured in 10^5 amperes per square centimeter, and the voltage is measured in volts.

The region in Figure 4 includes a turning point, so the linear solves involving the Jacobian matrix W'(f) should have been hard for part of the time, but the average number of iterations was around 200 iterations to solve one linear system. The current / voltage plot differs from the one previous shown.³ A coarser computational grid was used, and there were only 6 thousand unknowns. The current / voltage plot found in³ is shown in Figure 5.

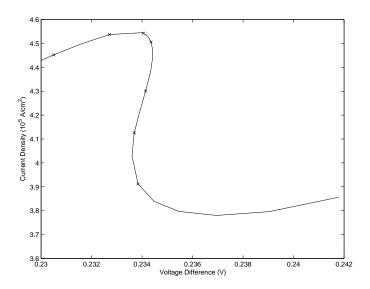


Figure 4. Seven Continuation Steps of the Numerical Study

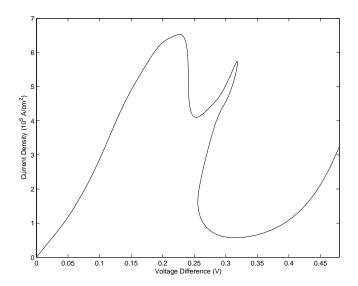


Figure 5. Coarser Grid Current - Voltage Relationship

The simulations using the finer grid reveal new physical effects that differ significantly from those that had been obtained from earlier investigations that employed coarser grids. For example, in Figure 3 one can see two local maximum in the current characteristic for voltages less than 0.25 volts when a high-level of unknowns are used. This can be contrasted to the results in Figure 5 that shows only one local maximum in current for voltages less than 0.25 volts on the coarser grid. Therefore, the more accurate numerical results of the refined grid studies suggest the presence of additional resonant peaks that are most probably a result of subband coupling. However, it is important to note that multiple resonant peaks have not been routinely observed in experimental measurement. In addition, our group has independent modeling results that show a very strong dependence of these characteristics on the amount of correlation (i.e., the value of L_c) that is included into the calculation of T(x, z). Therefore, these results show that it is very important to prescribe the proper amount of correlation into the Wigner-based simulation of electron transport in RTDs if one seeks an accurate calculation of the current density. Furthermore, it can be expected that correlation will also influence the dynamical behavior of RTD oscillators, and this will be the focus of future applications of this research.

5.5. Conclusion

This research effort has combined our parallelized Wigner code with robust numerical continuation algorithms. This allows for the efficient simulation of RTD characteristics using significantly refined grids. The more sophisticated numerical solvers that allow for generating highly accurate solutions in reasonable amounts of time, has revealed new physical effects in the current/voltage characteristics. Currently, the authors are analyzing the stability of the steady-state electron distributions generated on the finer grids. This analysis will inform us of the transient operating behavior of the RTD, in particular, telling us if oscillations will develop. The results will be presented by the authors in a later work.

The numerical tool developed during this research has enabled efficient simulation of RTDs, dramatically speeding up coarse grid computations and allowing fine grid computations that were previously infeasible to perform. The speed-up gained with this implementation makes multiple computational experiments possible in a reasonable amount of time, allowing for a detailed study of RTD characteristics. This simulation tool will be applied in the future to the study of RTD-based oscillators and will be useful for the study and design of THz frequency sources.

6. ACKNOWLEDGEMENTS

MSL, CTK, and PZ were supported by a US Army DURINT grant and by ARO grant W911NF-04-1-0276.

REFERENCES

- 1. D. Woolard, W. Loerop, and M. Shur, eds., *Electronic Devices and Advanced Systems Technology*, vol. 1 of *Terahertz Sensing Technology*, World Scientific, Hong Kong, 2003.
- 2. D. Woolard, W. Loerop, and M. Shur, eds., Emerging Scientific Applications and Novel Device Concepts, vol. 2 of Terahertz Sensing Technology, World Scientific, Hong Kong, 2003.
- 3. P. Zhao, H. L. Cui, and D. L. Woolard, "Dynamical instabilities and I-V characteristics in resonant tunneling through double barrier quantum well systems," *Phys. Rev. B* **63**, p. 75302, 2001.
- 4. P. Zhao, H. L. Cui, D. L. Woolard, K.L.Jensen, and F.A.Buot, "Simulation of resonant tunneling structures: Origin of i-v hysteresis and plateau-like structure," *Journal of Applied Physics* 87, pp. 1337–1349, 2000.
- 5. E. Wigner, "On the quantum correction for thermodynamic equilbrium," Phys. Rev. 40, pp. 749–759, 1932.
- 6. C. T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, vol. 16 of *Frontiers in Applied Mathematics*, SIAM, Philadelphia, PA, first ed., 1995.
- 7. Y. A. Kuznetsov, *Elements of Applied Bifurcation Theory*, vol. 112 of *Applied Mathematics Sciences*, Springer-Verlag, New York, second ed., 1998.
- 8. A. G. Salinger, N. M. Bou-Rabee, R. P. Pawlowski, E. D. Wilkes, E. A. Burroughs, R. B. Lehoucq, and L. A. Romero, "Loca 1.0 library of continuation algorithms: Theory and implementation manual," Technical Report SAND2002-0396, Sandia National Laboratory, March 2002.
- 9. H. B. Keller, Lectures on Numerical Methods in Bifurcation Problems, Springer-Verlag, New York, 1987.

- 10. J. W. Demmel, Applied Numerical Linear Algebra, SIAM, Philadelphia, PA, 1997.
- 11. H. F. Walker, "Timely communication: An adaption of krylov subspace methods to path following problems," SIAM J. Sci. Comput. 21, pp. 1191–1198, 1999.