

Efficient Inverse Compensation for Hysteresis Via Homogenized Energy Models

Thomas R. Braun^{*} and Ralph C. Smith[†]

Center for Research in Scientific Computation
Department of Mathematics
North Carolina State University
Raleigh, NC 27695

ABSTRACT

Ferroelectric and magnetic transducers are utilized in large number of applications, including nanopositioning, fluid pumps, high-speed milling, and vibration control/suppression. However, the physical mechanisms which make these materials highly effective transducers inherently introduce nonlinear, hysteretic behavior that must be incorporated in models and control designs. This significantly complicates control designs and limits the effectiveness of linear control algorithms when directly applied to the system. One solution is to employ an exact or approximate inverse model which converts a desired output to the corresponding input. This alleviates the complex input-output relation, allowing a linear control to be applied. Linearization of the actuator dynamics in this manner permits subsequent use of linear control designs to achieve high accuracy, high speed tracking as well as vibration attenuation and positioning objectives.

1. INTRODUCTION TO FERROELECTRIC AND FERROMAGNETIC MATERIALS

Ferroelectric and ferromagnetic materials are employed in a wide variety of applications as both actuators and sensors. They are attractive because the transducer is solid-state and often is very compact. Essentially, the materials present a coupling between a nonmechanical field and a mechanical deformation. Application of an input field yields a shape change (strain) in the material, and conversely an applied stress induces a nonmechanical change (polarization or magnetization) that can be measured. Ferroelectric materials are employed in automobiles to deploy airbags, microphones, accelerometers, fluid pumps, nanopositioning stages, sonar, vibration sensors and attenuators, and ultrasonic sources. Ferromagnetic transducers are typically bulkier than their ferroelectric counterparts, but provide greater force. These are also employed in a wide variety of applications including high-speed milling, ultrasonic sources, sonar, and vibration control. More details on applications of these materials can be found in [5].

Research for many of these applications includes attempts to achieve higher accuracy while simultaneously reducing the size of the actuator or sensor. However, the same physical mechanisms that allow these materials to sense or actuate effectively also introduce nonlinear, hysteretic, and time-dependent behavior. For an accurate controller, this behavior must be accommodated in one of four ways:

1. Operate the material in only a small portion of its theoretical operating range. This reduces much of the capability of the material but preserves nearly linear dynamics.

^{*}Email: tbraun@pobox.com; Telephone: 919-854-2746

[†]Email: rsmith@eos.ncsu.edu; Telephone: 919-515-7552

2. Operate the material over a wider range, but at only a low frequency. If the sufficiently low (compared to the sampling rate of the controller), the controller can update sufficiently fast to follow the nonlinearities and hysteresis, even if it does not possess a model of these effects.
3. Use a model of the material as part of a nonlinear control.
4. Model the material and invert this model to provide a compensator. This allows the construction of a composite system whose dynamics are linear across the entire operating range.

The first two approaches have been taken classically, due to their simplicity. However, it is readily apparent that the price paid for this simplicity is reduced performance from the transducer. The other two approaches seek to preserve this performance, but require a model of the material. Several models for these materials exist, but any model that includes the hysteresis and thermal effects requires a nontrivial amount of computation. As such they have traditionally been limited to relatively low speed regimes, mitigating much of their advantage. Since we desire to build an accurate control even in cases where the desired trajectory is not known *a priori*, we focus on the fourth option, and explore computationally efficient means to compute the inverse model.

2. HOMOGENIZED ENERGY MODEL

Several different models exist for ferroelectric and ferromagnetic materials. These include domain wall models, Preisach models, and homogenized energy models. Information and numerous references for these models may be found in [5]. We employ the homogenized energy approach due to its solid basis in the underlying physics of the material, its ability to characterize effects such as minor loops and creep in a natural manner, and its applicability to both ferroelectric and ferromagnetic materials. We will summarize first the polarization/magnetization model and extend this to strains/displacements in Section 4.

2.1. Theoretical Development

A bulk ferroelectric or ferromagnetic material is comprised of many layers, each of which are generally the same but may have somewhat different parameters. A layer itself is a mesoscopic structure containing several homogenous electric dipoles or magnetic moments that can vary their orientation with the applied field. The homogenized energy model's approach is to model one layer through energy principles, and then extend this to the bulk material through stochastic methods. Specifically, we assume most material parameters are fixed, but that two of them vary in different layers. One of these two parameters is the coercive field where the dipole or moment switches its orientation. For the other, we assume that the field acting on each dipole or moment is not the same as the field applied to the bulk material. Instead, this field is shifted by some effect of interactions between neighboring dipoles or moments. This level of interactions is also assumed to vary in different layers. A stochastic distribution is applied to each of these two parameters, to describe how they vary over the bulk material.

The macroscopic ferroelectric model is then given by integrating over these distributions, namely

$$P(E; x_+) = \int_0^\infty \int_{-\infty}^\infty \nu_c(E_c) \nu_I(E_I) [\overline{P}(E + E_I; E_c; x_+)] dE_I dE_c \quad (1)$$

where P is the polarization, E is the electric field, E_c denotes the coercive point, E_I quantifies dipole interactions, and x_+ is the internal state of the dipoles. Here \overline{P} denotes the model for one layer, which we derive momentarily. For magnetic materials the analogous model

$$M(H; x_+) = \int_0^\infty \int_{-\infty}^\infty \nu_c(H_c) \nu_I(H_I) [\overline{M}(H + H_I; H_c; x_+)] dH_I dH_c \quad (2)$$

is employed, where H is magnetic field and M is magnetization. In either case, the distributions ν_c and ν_I are constrained by the constitutive laws:

1. both ν_c and ν_I are bounded by decaying exponentials,

2. ν_c is strictly positive,
3. ν_I is symmetric about 0, and
4. $\int_0^\infty \nu_c(E_c) dE_c = 1$, $\int_{-\infty}^\infty \nu_I(E_I) dE_I = 1$.

The integrals in (1) or (2) are solved numerically with quadrature, i.e.,

$$P(E; x_+) = \sum_{i=1}^{N_c} \sum_{j=1}^{N_I} \nu_c(E_c[i]) \nu_I(E_I[j]) w_c[i] w_I[j] \bar{P}(E + E_I[j]; E_c[i]; x_+[i, j]) \quad (3)$$

or the magnetic equivalent, where w_c and w_I give the quadrature weights. To simplify the subsequent discussion, we formulate equations solely in terms of the electric field and polarization, with the understanding that the model is unchanged if magnetic field and magnetization/inductance are instead employed.

The kernel \bar{P} describes the polarization in a single homogeneous layer of the material. The Helmholtz energy for this layer may be approximated by the piecewise quadratic function

$$\psi(P) = \begin{cases} \eta(P + P_R)^2/2, & P \leq -P_I \\ \frac{\eta}{2}(P_I - P_R) \left(\frac{P^2}{P_I} - P_R \right), & |P| < P_I \\ \eta(P - P_R)^2/2, & P \geq P_I \end{cases} \quad (4)$$

where P_I denotes the positive inflection point at which the switch occurs, P_R is the local remanence polarization, and η is the reciprocal slope $\frac{\partial E}{\partial P}$. The Gibbs's free energy

$$G = \psi - EP \quad (5)$$

balances this internal Helmholtz energy with the electrostatic energy; i.e., work performed by the applied external field. As detailed [5], where the Legendre transform properties of the Gibbs energy are discussed, G is a function of the independent variable E and the dependent variable P . Note that the local coercive field E_c , local remanence polarization P_R , and inflection point P_I are related by the expression

$$P_I = P_R - \frac{E_c}{\eta}. \quad (6)$$

For regimes in which thermal relaxation is negligible, direct minimization of (5) yields the kernel

$$\bar{P}(E + E_I, x_+) = \frac{E + E_I}{\eta} + 2P_R x_+ - P_R \quad (7)$$

where $x_+ = 1$ for positively oriented dipoles and $x_+ = 0$ for negatively oriented dipoles. In terms of logic flow, x_+ is set to 1 whenever $E + E_I > E_c$ and is set to 0 whenever $E + E_I < -E_c$. Thermal activation is manifested by dipoles having sufficient thermal energy to switch states before a minima of the Gibbs energy is eliminated. To quantify this, the Boltzmann relation

$$\mu(G) = C \exp\left(-\frac{G(E + E_I, P)V}{kT}\right) \quad (8)$$

balances the Gibbs and relative thermal energies, where V is a reference volume, T is the temperature in degrees Kelvin, and k is Boltzmann's constant. The constant C is chosen to ensure integration to unity. It is shown in [5, 6, 8] that the resulting kernel is

$$\bar{P} = x_+ \langle P_+ \rangle + (1 - x_+) \langle P_- \rangle \quad (9)$$

where $x_+ \in [0, 1]$ denotes the fraction of positively oriented dipoles and

$$\langle P_+ \rangle = \frac{\int_{P_I}^{\infty} P \exp\left(\frac{-G(E+E_I, P)V}{kT}\right) dP}{\int_{P_I}^{\infty} \exp\left(\frac{-G(E+E_I, P)V}{kT}\right) dP}, \quad \langle P_- \rangle = \frac{\int_{-\infty}^{-P_I} P \exp\left(\frac{-G(E+E_I, P)V}{kT}\right) dP}{\int_{-\infty}^{-P_I} \exp\left(\frac{-G(E+E_I, P)V}{kT}\right) dP} \quad (10)$$

are the average polarizations associated with positive and negative dipole orientations. The evolution of dipole fractions is governed by the differential equation

$$\dot{x}_+ = -p_{+-}x_+ + p_{-+}(1 - x_+), \quad (11)$$

where the likelihoods p_{+-} and p_{-+} of a dipole switching from positive to negative, or vice-versa, are

$$p_{+-} = \frac{\exp\left(\frac{-G(E+E_I, P_I)V}{kT}\right) dP}{\tau(T) \int_{P_I}^{\infty} \exp\left(\frac{-G(E+E_I, P)V}{kT}\right) dP}, \quad p_{-+} = \frac{\exp\left(\frac{-G(E+E_I, -P_I)V}{kT}\right) dP}{\tau(T) \int_{-\infty}^{-P_I} \exp\left(\frac{-G(E+E_I, P)V}{kT}\right) dP} \quad (12)$$

and τ quantifies the material and temperature-dependent relaxation time.

2.2. Quadrature and the Continuity of P

For model inversion and control applications, we desire the model to be smooth. By definition of the integral, the model (1) is continuous and differentiable. However, the model with quadrature (3) does not give such smoothness by definition. Since (3) is composed of finite sums, P is smooth if and only if \bar{P} is smooth for every quadrature point. Clearly for the negligible relaxation case (7), \bar{P} is discontinuous. Its value changes by $2P_R$ at the coercive point $E + E_I = E_c$. Thus, P contains $N_c N_I$ finite jumps, each of size $2P_R \nu_c(E_c[i]) \nu_I(E_I[j]) w_c[i] w_I[j]$ for $i = 1 \dots N_c, j = 1 \dots N_I$.

With relaxation, the situation is more complicated. Both switching likelihoods and average polarizations are continuously differentiable and thus the polarization is a continuous function. However, a plot of $P(E; x_+)$ (or rather $M(H; x_+)$ since the parameters come from a ferromagnetic material) seems to contradict this result. Figure 1(a) shows a plot of $M(H; x_+)$ for various values of H where the state x_+ was held fixed, i.e., each plotted sample was obtained by inputting the given H to the model with the same value of x_+ (all ones in this case). Inspection shows M is continuously differentiable, as shown in Figure 1(b). However, the resolution needed to fully resolve the smoothness is, in this case, ten orders of magnitude below the field value. Thus for numerical computations (3) cannot be fully resolved and must be treated as discontinuous as well.

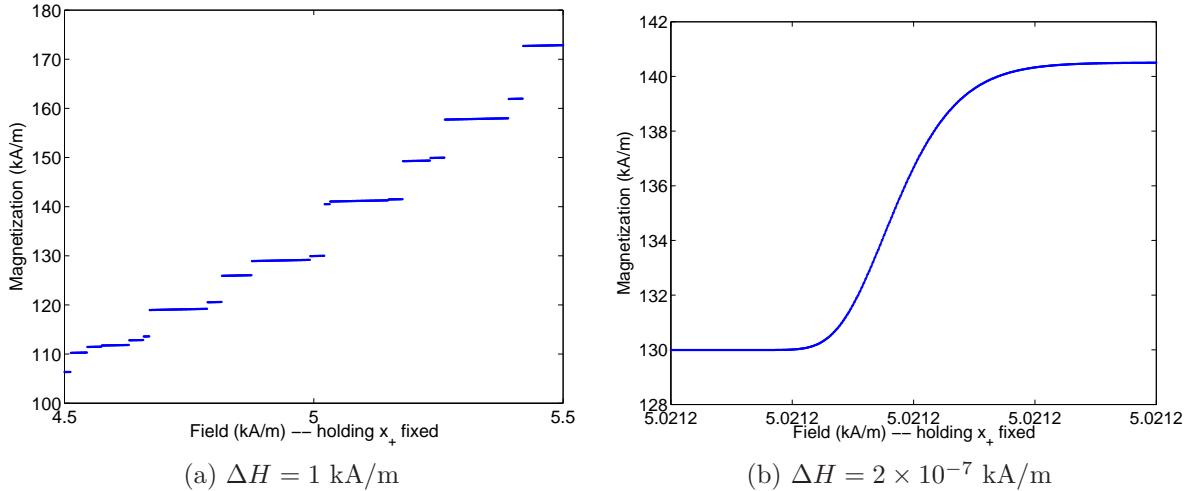


Figure 1. Plot of $M(H; x_+)$ for 6 intervals of composite 4-point Gaussian quadrature on each distribution. For each value of H , x_+ was held fixed; thus the plots represent M as a scalar function of H . The plot in (b) is a much higher resolution view of one of the apparent discontinuities given in (a).

3. MODEL INVERSION

As discussed in Section 1, one method to accurately control a smart material actuator involves the use of an inverse compensator. A prototypical setup for this type of control is depicted in Figure 2. The material is nonlinear, hysteretic, and time-varying, but these details are characterized by the inverse compensator. The composite system to be controlled is then linear and time-invariant. This allows simple control designs including PI/PID and LQR designs to be utilized.

The inverse compensator is essentially just the inverse of the homogenized energy model. Given a value for the polarization or magnetization, we wish to determine the field level necessary to bring the actuator to that value. Extensions of this concept to working with strain or displacement data are addressed in Section 4. Such inverse compensators have previously been explored for the negligible relaxation model in [2, 5]. Here we consider a method that addresses both the relaxation and negligible relaxation models.

A more precise definition of the inverse compensator is the following: given any valid state x_+ and any \hat{P} within the operating range of the material, determine E such that $P = \hat{P}$, where P is the solution of (3). However, due to the nonlinearity and dependence on x_+ , it is not feasible to invert (3) analytically. Thus, we reformulate the problem as a numerical root finding problem, namely determining the value E such that for a given x_+ and \hat{P} ,

$$P(E; x_+) - \hat{P} = 0. \tag{13}$$

To solve this efficiently, it is beneficial to first analyze the structure of (13), which is equivalent to analyzing the structure of (3). Holding x_+ fixed and treating (3) as a function solely of E , we note that it is monotone increasing. However, as discussed in Section 2.2, the function may be discontinuous.

3.1. Discontinuous Root Finding

A number of methods to solve (13) exist in the numerical analysis literature. If the function is continuously differentiable, the secant method or various interpolation schemes give superlinear convergence to E in terms of function evaluations [3, 4]. However, derivative based methods may not converge for discontinuous problems, whereas the convergence of interpolatory methods may be poor [1, 4]. The simplest root finding method, the bisection method, converges comparatively slowly for smooth functions, but will handle the discontinuities appropriately; in fact, for some discontinuous functions it can be shown to be optimal. Methods which handle discontinuities by combining bisection with interpolation schemes have also been developed; see [1] for example. However, we found better performance could be obtained by employing the secant method followed by bisection.

The bisection method, or any other method which utilizes it, requires both that the root be bounded and that the function change signs within the bounds. Since the function being considered is monotone (holding x_+ fixed), the requirement that the function change signs is satisfied whenever the root is bounded. For bounds, theoretical limits on E may be computed by considering the case when all dipoles have switched. If all dipoles are positively oriented, then regardless of whether or not relaxation is considered

$$\hat{P} = \frac{E}{\eta} + P_R \implies E = \eta(\hat{P} - P_R), \tag{14}$$

whereas if all dipoles are negatively oriented

$$\hat{P} = \frac{E}{\eta} - P_R \implies E = \eta(\hat{P} + P_R). \tag{15}$$

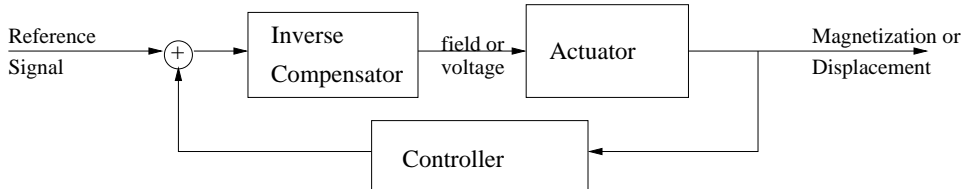


Figure 2. Depiction of an actuator control utilizing an inverse compensator coupled with a linear controller.

Any state will necessarily be between these two extremes and thus we have

$$\eta(\widehat{P} - P_R) \leq E \leq \eta(\widehat{P} + P_R) \quad (16)$$

regardless of the initial state x_+ . These bounds allow one to utilize the bisection method or any other method requiring the root be bounded. However, the bounds are large. The bisection method iterates by dividing the interval between these bounds in half on each iteration, so these large initial bounds result in slow convergence.

We therefore turn to the secant method as a way to calculate more efficient bounds. Like all root-finding methods, the secant method is an iterative approach and seeks by intelligent trials to find a value of E that bring the function value within some tolerance of 0. It utilizes the previous iteration to approximate the derivative of the function, then takes a step equal to where the root would be if the function were linear (see [3] for a more complete description). The reliance on derivatives explains the requirement that the function be smooth. If there is a sufficient amount of relaxation, or if the given value of \widehat{P} does not fall near any of the real or apparent discontinuities, then smoothness is present and the secant method will converge superlinearly to the solution. If this is not the case, the secant method will likely not converge. However, the effort spent in the secant method is not wasted; the secant iterations provide a bound on the root.

Equation (13) is monotone increasing which implies that each iteration of the secant method will step in the correct direction. Either that iteration will not step far enough, in which case the function value is reduced in magnitude, or the secant method will step too far. If it steps too far, the function value may increase or decrease in magnitude, but either way the current approximation for E is now on the opposite side of the root from where it started. In this case, the previous and current iteration provide a bound on the root. We therefore apply the secant method to (13), keeping track of its iterations. As long as the approximation improves (i.e., the function value decreases in magnitude), we continue using it. If the method fails to converge, then we use its iterations to determine a bound on the root. This bound is then given to the bisection method to complete the job. In most cases, the bound computed from the secant method is significantly tighter than (16), which in turn implies bisection will converge more rapidly. For the purposes of our algorithm, failure to converge in the secant method is defined as

$$|P(E_i; x_+) - \widehat{P}| \geq |P(E_{i-2}; x_+) - \widehat{P}|, \quad \text{for all } i \geq 3 \quad (17)$$

where E_i is the value of E determined by the i th iteration of the secant method. Such a definition is a heuristic to detect failure quickly without penalizing a single poor step.

The secant method requires an initial approximation to the derivative to compute its first iteration. One such approximation is $\frac{1}{\eta}$, which is the derivative if no dipoles are switching. The actual derivative may be much larger than this, but will never be smaller. Another approximation involves using the previous two time steps of the model, in a manner analogous to how the secant method itself approximates derivatives. This approximation is simply $\Delta P / \Delta E$, where ΔP is the difference in $P(E; x_+)$ for the previous two time steps, and ΔE is the corresponding differences in E . The latter approximation is deemed more accurate, and is utilized when the previous two time steps are known. When this is not the case, $\frac{1}{\eta}$ is utilized as the initial derivative.

3.2. Numerical Considerations

There are a couple of additional factors that should be considered when implementing the given algorithm. One important issue is that subtraction of nearly equal numbers can lead to large error in any finite precision arithmetic. This can effect the calculation of the approximate derivative. However, these errors can in some sense be caught. The derivative of (13) will never be less than $\frac{1}{\eta}$. Whenever the computed approximate derivative is less than this value, the calculation is known to be inaccurate and $\frac{1}{\eta}$ may be employed instead. An approximate derivative that is too large cannot be caught in such a manner. While not optimal, this situation is deemed acceptable as it does not result in a large magnification of the error (only in a suboptimal step).

Whether caused by numerical error or the approximation of the derivative, other bad steps can at times be caught. Theoretical bounds on E are given in (16). These bounds may be improved as the secant method progresses. If an iteration recommends a value of E outside what is known to be a bound, we know that value to be invalid and can substitute the value at the boundary instead. Since we do not seek to penalize a single poor iteration of the secant method, this can at times help the method recover and either converge or produce a very tight bounding region.

3.3. Validation and Performance

The performance of the inverse model is largely dependent on the material parameters being utilized. As the amount of relaxation increases, $P(E; x_+)$ changes more gradually and the secant method converges better, leading to better overall performance. Likewise, if the variance in ν_c and ν_I is decreased the discontinuities may be grouped in a smaller region and secant may converge for a wider range of values. This also gives better performance. Finally, performance also depends on how quickly the input polarization changes; slower changes typically yield a better initial iterate which also gives better performance. As an example, consider the inverse model for the stepped polarization in Figure 3. The parameters were taken from values for PLZT. Ten intervals with 4-point Gaussian quadrature were utilized for each distribution, and the effect of relaxation is noticeable. As can be seen, the inverse model accounts for the relaxation in the device, decreasing the electric field as appropriate to hold the polarization constant. The error is between the requested and actual values is small; a small tolerance was used so that the error plotted is the least error possible based on the discontinuities given by the quadrature. This level is five orders of magnitude below the remanence polarization.

Accuracy and effort for various tolerance levels can be seen in Table 1. Note that the algorithm makes efficient use of computation; reasonable accuracy levels may be obtained with only 2 function evaluations per iteration on average, or 5 in the worst case scenario. While this computational efficiency is not atypical, it should be

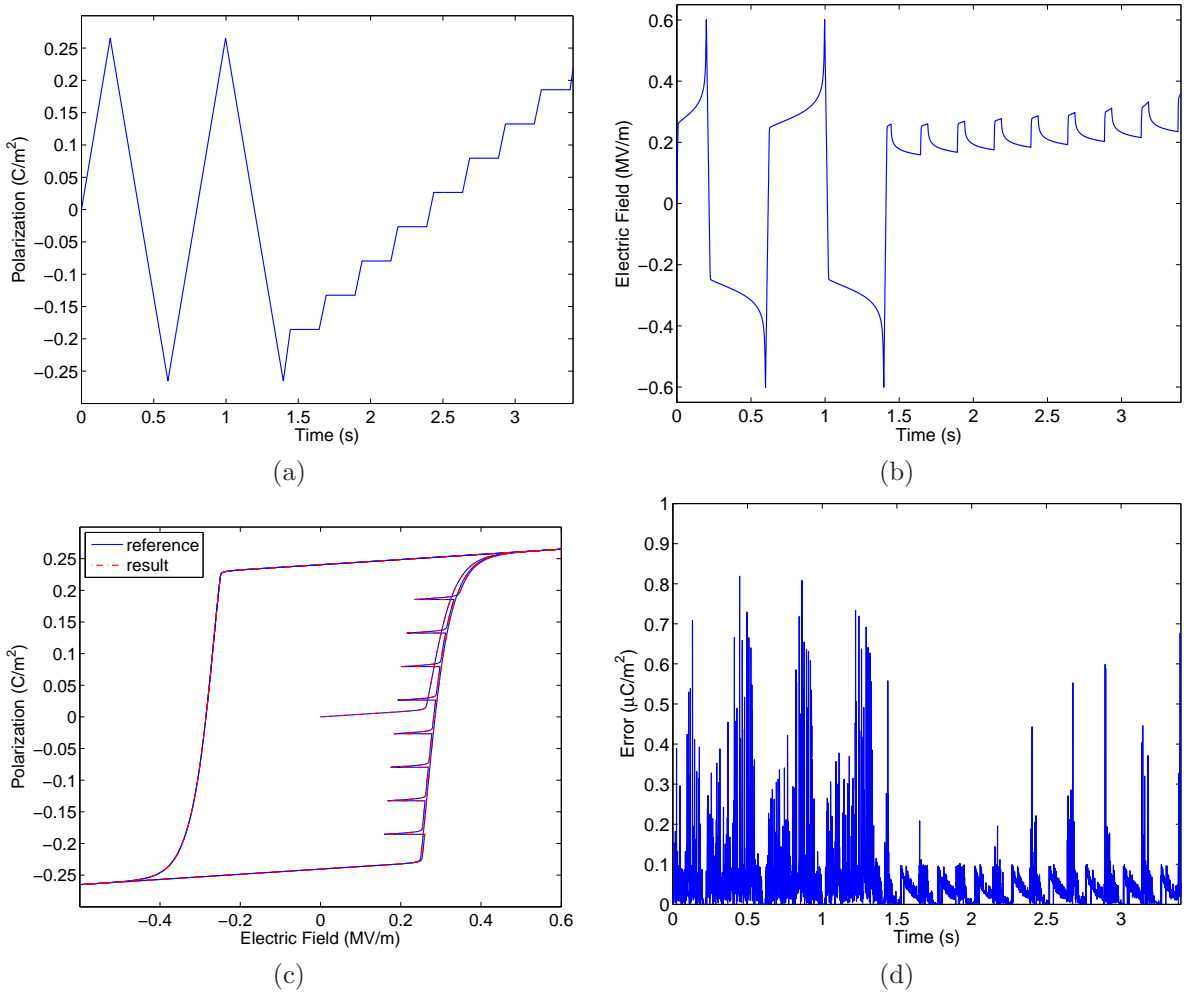


Figure 3. Electric field (b) and error (d) resulting from applying the inverse model to the polarization in (a). A comparison of the requested and resulting polarizations is given in (c). Material parameters were taken from PLZT.

Tolerance	Function Evaluations (Average)	Function Evaluations (Maximum)
$P_R \times 10^{-2} = 0.0024$	1.0855	3
$P_R \times 10^{-3} = 0.00024$	1.2667	4
$P_R \times 10^{-4} = 2.4 \times 10^{-5}$	2.0157	5
$P_R \times 10^{-5} = 2.4 \times 10^{-6}$	3.5423	9
$P_R \times 10^{-6} = 2.4 \times 10^{-7}$	4.8066†	user-defined

Table 1. Effort to compute the inverse model in terms of average and maximum number of function evaluations for various error tolerances. For the final row, some requested polarization values could not be obtained; thus the iterations for these values continued to the user-defined maximum. In this case, the average number of function evaluations is only for those values which converged, which is denoted by †.

noted that parameters for different materials may give worse performance. This can be seen in the Terfenol-D example given in Section 5.

4. DISPLACEMENT MODEL

Whereas polarization or inductance/magnetization can sometimes be measured, it is often necessary to formulate the forward model output or inverse model input terms of displacements instead of polarization. Any such extension requires knowledge of the geometry of the transducer. Rods or beams can be characterized by a one-dimensional transducer model, while more general actuator designs require the model from Section 2 be extended to two or three dimensions. Since a rod is more common, we focus on this geometry. The same principles applied here can be applied to beams or other device models. These extensions may be found in [5].

If we assume applied fields and stresses are uniform along the length of the rod, we may model the rod solely at the free end and produce a simpler model containing solely an ordinary differential equation. Without this assumption, a partial differential equation rod model quantifying the displacement at each point along the rod is needed. The uniform assumption is a reasonable approximate for most applications, and the resulting ordinary differential equation is much more computationally efficient than its partial differential equation counterparts. As such, we focus on this approach.

As detailed in [5], the constitutive relation for the induced stress is given by

$$\sigma = Y\varepsilon + C\dot{\varepsilon} - a_1(P(E; x_+) - P_0) - a_2(P(E; x_+) - P_0)^2 \quad (18)$$

where Y is the Young's modulus of the transducer, C is the Kelvin-Voight damping coefficient, P_0 is the point the material is poled around, a_1 and a_2 are constants of proportionality relating the induced polarization to mechanical force, and $P(E; x_+)$ is the result of (3). We consider a rod of length ℓ and cross-sectional area A as depicted in Figure 4, where the mass m_ℓ , stiffness k_ℓ and damping c_ℓ at the rod tip encompasses the dynamics of the larger system in which the rod is placed. Let u_ℓ denote the displacement at the end of the rod. Then the strain may be given as

$$\varepsilon = \frac{u_\ell}{\ell}. \quad (19)$$

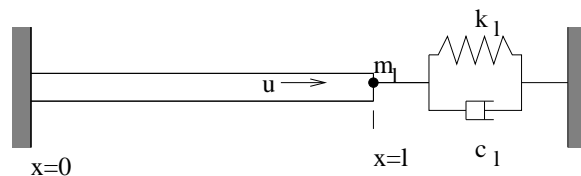


Figure 4. Simplified rod used to construct overall model. The end mass m_ℓ , stiffness k_ℓ and damping c_ℓ represent dynamics imposed by the larger system containing the transducer.

By balancing forces and grouping like terms we obtain

$$m \frac{d^2 u_\ell}{dt^2} + c \frac{du_\ell}{dt} + k u_\ell = \alpha_1 (P - P_0) + \alpha_2 (P - P_0)^2 \quad (20)$$

where ρ is the density of the rod and

$$m = \rho A \ell + m_\ell, \quad c = \frac{CA}{\ell} + c_\ell, \quad k = \frac{YA}{\ell} + k_\ell, \quad \alpha_1 = A a_1, \quad \alpha_2 = A a_2.$$

Further details may be found in [5, 7]. Note that the rod model is decoupled from the homogenized energy model for the device – the homogenized energy model output forms the rod model input, but the two models may be solved independently.

The solution to this differential equation may be obtained with a variety of numerical methods. However, these may be separated into fixed step versus variable step methods. Variable step methods have the potential to be more accurate but also are more complex and may prove problematic in a real-time environment. They also require P be known at arbitrary times, which in turn requires both that E be known and the homogenized energy model be computed at arbitrary times. This is typically not feasible in practice. As such, we focus on fixed time step methods. We assume P is fixed between time steps, and with this assumption solve (20) analytically. This is an approximation, since thermal effects dictate that P will vary with time even if E is constant; however, all other fixed step methods implicitly make similar assumptions and include additional approximation error as well.

The approximation of (20) depends on the relation of $\frac{c^2}{m^2}$ to $\frac{4k}{m}$. To simplify notation, let

$$v = \frac{a_1}{m} (P - P_0) + \frac{a_2}{m} (P - P_0)^2. \quad (21)$$

We note that v is assumed to be a constant in this derivation although the value of the constant changes at each time step. Additionally, let

$$\tilde{k} = \frac{k}{m}, \quad \tilde{c} = \frac{c}{m}.$$

If $\tilde{c}^2 > 4\tilde{k}$, then the solution is given by

$$u(t) = b_1 \exp\left(-\frac{1}{2}(\tilde{c} + \sqrt{\tilde{c}^2 - 4\tilde{k}})t\right) + b_2 \exp\left(-\frac{1}{2}(\tilde{c} - \sqrt{\tilde{c}^2 - 4\tilde{k}})t\right) \quad (22)$$

To simplify notation, let $z = \sqrt{\tilde{c}^2 - 4\tilde{k}}$. Solving for the constants b_1 and b_2 in terms of the previous timestep yields

$$\begin{aligned} u(t + \Delta t) = & \frac{1}{2} \left(\left(1 - \frac{\tilde{c}}{z}\right) \exp\left(-\frac{1}{2}(\tilde{c} + z)t\right) + \left(1 + \frac{\tilde{c}}{z}\right) \exp\left(-\frac{1}{2}(\tilde{c} - z)t\right) \right) \left(u(t) - \frac{v(t + \Delta t)}{\tilde{k}} \right) \\ & + \frac{1}{z} \left(\exp\left(-\frac{1}{2}(\tilde{c} - z)t\right) - \exp\left(-\frac{1}{2}(\tilde{c} + z)t\right) \right) \dot{u}(t) + \frac{v(t + \Delta t)}{\tilde{k}}. \end{aligned} \quad (23)$$

If we assume the step Δt is fixed, then we can define the six constants

$$\begin{aligned} d_1 = & \frac{1}{2} \left(\left(1 - \frac{\tilde{c}}{z}\right) \exp\left(-\frac{1}{2}(\tilde{c} + z)\Delta t\right) + \left(1 + \frac{\tilde{c}}{z}\right) \exp\left(-\frac{1}{2}(\tilde{c} - z)\Delta t\right) \right), \quad d_3 = \frac{1}{k} (1 - d_1), \\ d_2 = & \frac{1}{z} \left(\exp\left(-\frac{1}{2}(\tilde{c} - z)\Delta t\right) - \exp\left(-\frac{1}{2}(\tilde{c} + z)\Delta t\right) \right), \\ d_4 = & \frac{1}{4} \left(\frac{\tilde{c}^2}{z} - z \right) \left(\exp\left(-\frac{1}{2}(\tilde{c} + z)\Delta t\right) - \exp\left(-\frac{1}{2}(\tilde{c} - z)\Delta t\right) \right), \quad d_6 = -\frac{d_4}{\tilde{k}}, \\ d_5 = & \frac{1}{2} \left(\left(1 + \frac{\tilde{c}}{z}\right) \exp\left(-\frac{1}{2}(\tilde{c} + z)\Delta t\right) + \left(1 - \frac{\tilde{c}}{z}\right) \exp\left(-\frac{1}{2}(\tilde{c} - z)\Delta t\right) \right), \end{aligned} \quad (24)$$

In terms of these constants, the rod displacement is approximated by

$$\begin{aligned} u(t + \Delta t) &= d_1 u(t) + d_2 \dot{u}(t) + d_3 v(t + \Delta t), \\ \dot{u}(t + \Delta t) &= d_4 u(t) + d_5 \dot{u}(t) + d_6 v(t + \Delta t). \end{aligned} \quad (25)$$

A similar approach can be taken for the other two cases. In any case, the rod model will be given by (25) and only the values of the constants d_1, \dots, d_6 will vary. If $\tilde{c}^2 = 4\tilde{k}$, then solving the differential equation as before yields

$$\begin{aligned} d_1 &= \left(1 + \frac{\tilde{c}}{2}\Delta t\right) \exp\left(-\frac{\tilde{c}}{2}\Delta t\right), & d_4 &= -k\Delta t \exp\left(-\frac{\tilde{c}}{2}\Delta t\right), \\ d_2 &= \Delta t \exp\left(-\frac{\tilde{c}}{2}\Delta t\right), & d_5 &= \left(1 - \frac{\tilde{c}}{2}\Delta t\right) \exp\left(-\frac{\tilde{c}}{2}\Delta t\right), \\ d_3 &= \frac{1}{k}(1 - d_1), & d_6 &= -\frac{d_4}{k}. \end{aligned} \quad (26)$$

If instead $\tilde{c}^2 < 4\tilde{k}$,

$$\begin{aligned} d_1 &= \exp\left(-\frac{\tilde{c}}{2}t\right) \left(\cos\left(\frac{z}{2}\Delta t\right) + \frac{\tilde{c}}{z}\sin\left(\frac{z}{2}\Delta t\right)\right), & d_4 &= -\frac{1}{2}\left(z + \frac{\tilde{c}^2}{z}\right) \exp\left(-\frac{\tilde{c}}{2}t\right) \sin\left(\frac{z}{2}\Delta t\right), \\ d_2 &= \frac{2}{z}\exp\left(-\frac{\tilde{c}}{2}t\right) \sin\left(\frac{z}{2}\Delta t\right), & d_5 &= \exp\left(-\frac{\tilde{c}}{2}t\right) \left(\cos\left(\frac{z}{2}\Delta t\right) - \frac{\tilde{c}}{z}\sin\left(\frac{z}{2}\Delta t\right)\right), \\ d_3 &= \frac{1}{k}(1 - d_1), & d_6 &= -\frac{d_4}{k} \end{aligned} \quad (27)$$

where $z = \sqrt{4\tilde{k} - \tilde{c}^2}$. Therefore, as long as Δt is fixed, the rod model can be solved for any set of parameters by applying the appropriate constants (24), (26) or (27) to the displacement relation (25).

Unlike the homogenized energy model, the inverse rod model may be compute analytically. Again assuming Δt is fixed, inverting (25) yields

$$v(t + \Delta t) = \frac{1}{d_3}u(t + \Delta t) - \frac{d_1}{d_3}u(t) - \frac{d_2}{d_3}\dot{u}(t). \quad (28)$$

The solution process proceeds in two steps, given a value for $u(0)$, $u(\Delta t)$ and $\dot{u}(0)$, determine $v(\Delta t)$. Use this to determine $\dot{u}(\Delta t)$. This and the desired displacement value $u(2\Delta t)$ allows $v(2\Delta t)$ to be determined, etc. Once v has been determined, the required polarization can be determined from the quadratic formula, namely

$$P(t) = P_0 - \frac{\alpha_1}{2\alpha_2} \pm \sqrt{\left(\frac{\alpha_1}{2\alpha_2}\right)^2 + \frac{mv(t)}{a_2}} \quad (29)$$

as long as $\alpha_2 \neq 0$. If α_2 is 0, the equation is linear and the solution is

$$P(t) = P_0 + \frac{m}{\alpha_1}v(t). \quad (30)$$

Note that for $\alpha_2 \neq 0$ either polarization value is acceptable, as both give equivalent displacements. However, to minimize the effects of modeling error, a single branch should always be utilized. The resulting $P(t)$ may then be fed into the inverse material model from Section 3 to compute the required electric field.

5. RESULTS OF COMBINED INVERSE MODELS

Combining the inverse rod model with the inverse homogenized energy model allows controls to be specified in terms of strain or displacement. This is illustrated in Figure 5 for a Terfenol-D rod. Note that the inverse is

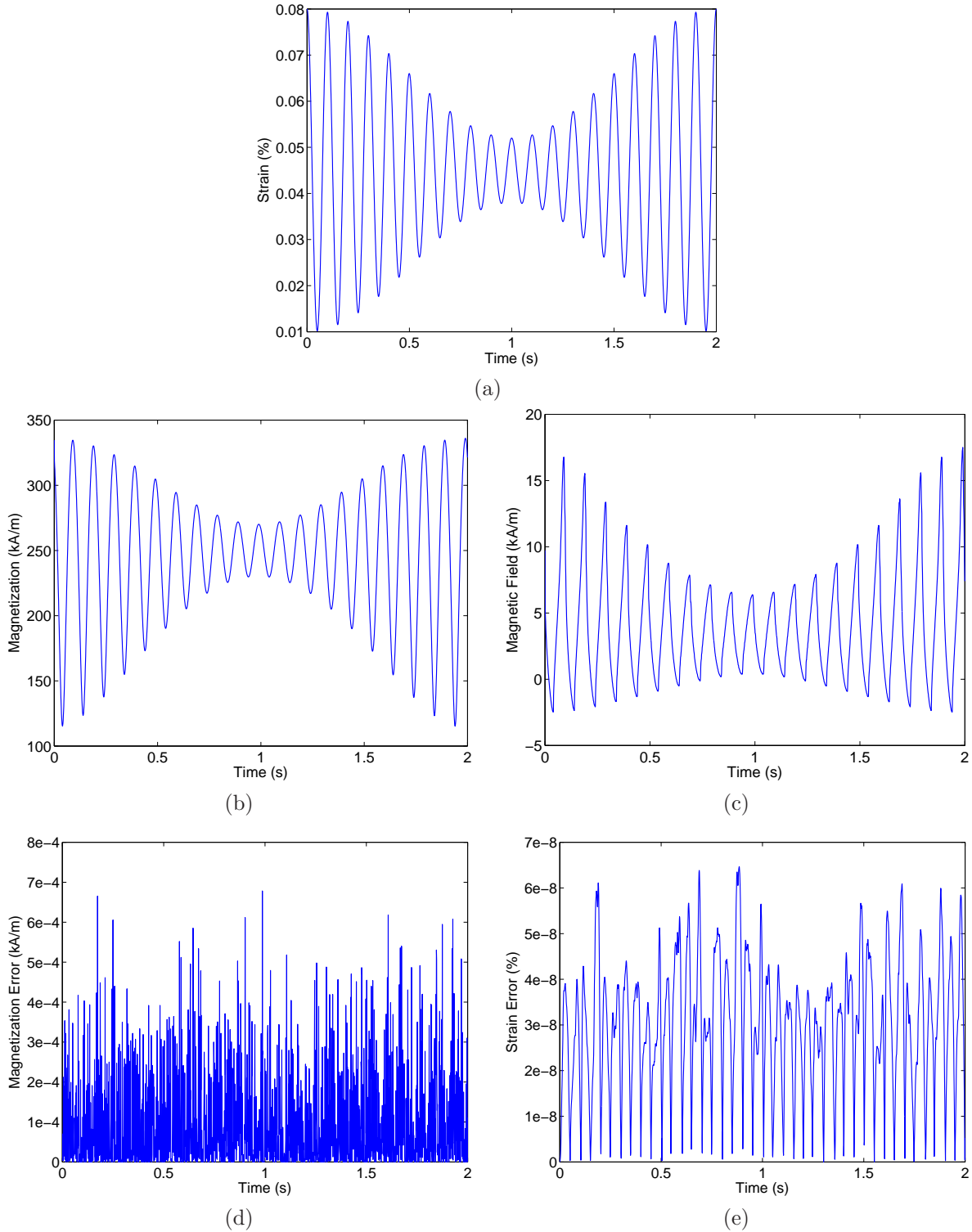


Figure 5. Results of the combined inverse rod and homogenized energy models for a Terfenol-D rod. The strain values in (a) are input to the inverse rod model, yielding the magnetization in (b). This is then input to the inverse homogenized energy model which gives the magnetic field (c) needed to drive the actuator to this displacement. The magnetization error of the homogenized energy model is given in (d), while the error in strain for the combined models is given in (e).

Tolerance	Function Evaluations (Average)	Function Evaluations (Maximum)
$M_R \times 10^{-2} = 3400$	1.8911	9
$M_R \times 10^{-3} = 340$	3.6977	18
$M_R \times 10^{-4} = 34$	12.0440	26
$M_R \times 10^{-5} = 3.4$	15.8381	29
$M_R \times 10^{-6} = 0.34$	18.9040†	user-defined

Table 2. Effort to compute the inverse model in terms of average and maximum number of function evaluations for various error tolerances. For the final row, some requested polarization values could not be obtained; thus the iterations for these values continued to the user-defined maximum. In this case, the average number of function evaluations is only for those values which converged, which is denoted by †.

obtained almost exactly, with a strain error (using a very small tolerance) six orders of magnitude below the desired strain value. Thus, the combined inverse models comprise an accurate open-loop control and can easily be extended to incorporate a feedback controller.

Function evaluations for various tolerance levels are summarized in Table 2. The inverse model is not as efficient for these parameter values as it was for the PLZT parameters in Section 3. This is not an effect of the rod model but instead determined by the material parameters. To allow for an even comparison, the tolerance here was still specified in terms of magnetization. It could also be specified in terms of the strain; although this requires the rod model details be available to the inverse model. Regardless, this inverse compensator approach shows significant potential for real-time model based control of an actuator, by allowing linear control methods to be employed with full accuracy.

ACKNOWLEDGMENTS

This work supported by the United States Department of Education GAANN fellowship and the Air Force Office of Scientific Research under the grant AFOSR-FA9550-04-1-0203.

REFERENCES

- [1] R.P. Brent, *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs, N.J., 1973.
- [2] A.G. Hatch, R.C. Smith, T. De, and M.V. Salapaka, “Construction and experimental implementation of a model-based inverse filter to attenuate hysteresis in ferroelectric transducers,” *IEEE Transactions on Control Systems Technology*, 14(6), pp. 1058-1069, 2006.
- [3] C.T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.
- [4] W.H. Press, S.A. Teukolsky, W.T. Vetterline, B.P. Flanner, *Numerical Recipes in C*, Second Edition, Cambridge University Press, New York, 1992.
- [5] R.C. Smith, *Smart Material Systems: Model Development*, SIAM, Philadelphia, 2005.
- [6] R.C. Smith, M.J. Dapino, T.R. Braun and A.P. Mortensen, “A homogenized energy framework for ferromagnetic hysteresis” *IEEE Transactions on Magnetics*, 42(7), pp. 1474-1769, 2006.
- [7] R.C. Smith, A.G. Hatch, T. De, M.V. Salapaka, R.C.H. del Rosario, and J.K. Raye, “Model development for atomic force microscope stage mechanisms,” *SIAM Journal on Applied Mathematics*, 66(6), pp. 1998-2026, 2006.
- [8] R.C. Smith, S. Seelecke, Z. Ounaies and J. Smith, “A free energy model for hysteresis in ferroelectric materials,” *Journal of Intelligent Material Systems and Structures*, 14(11), pp. 719-739, 2003.