Molecular Seismology: An Inverse Problem in Nanobiology

Peter Hinow
Department of Mathematics
Vanderbilt University
Nashville, TN 37240

Erik M. Boczko *
Department of Biomedical Informatics,
Vanderbilt University,
Nashville, TN 37232

Abstract

The density profile of an elastic fiber like DNA will change in space and time as ligands associate with it. This observation affords a new direction in single molecule studies provided that density profiles can be measured in space and time. In fact, this is precisely the objective of seismology, where the mathematics of inverse problems have been employed with success. We argue that inverse problems in elastic media can be directly applied to biophysical problems of fiber-ligand association, and demonstrate that robust algorithms exist to perform density reconstruction in the condensed phase.

Keywords: DNA-protein interaction, DNA binding kinetics, single molecule studies, wave equation, density inversion

1 Introduction

In many situations and applications it is of interest to understand the *dynamics* of interaction between a biopolymer fiber and a ligand. For instance it is of great interest to understand how proteins move along and around DNA fibers and interact with the various sequence elements in space and time. Crystal structures of proteins bound to DNA have shown how certain classes of transcription factors physically recognize specific DNA sequences, but these data contain very few if any clues about the dynamical processes that led to the observed static associations. The suggestion has been made that DNA binding proteins do not

*Correspondence to erik.boczko@vanderbilt.edu, Ph: 615 936–6668, Fax: 615 936–1427

undergo three dimensional diffusion in the nucleus, but rather that they engage in essentially one dimensional diffusion by hoping or crawling on, along or between the fibers, possibly biased toward the DNA by a potential (Gowers *et al.*, 2005; Graneli *et al.*, 2006; Halford & Marko, 2004; von Hippel & Berg, 1989; Wang *et al.*, 2006). How does a DNA binding protein acquire a target site? Do DNA binding proteins home in on their cis regulatory elements (CRE) like a docking space shuttle guided by a potential? Is it possible for a DNA binding protein to collide with a fiber and then corkscrew along the double helix like a monorail, riding either the major or the minor groove? When two or more transcription factors compete for binding sites at a promoter, how do the microscopic dynamics evolve, and what are the dynamical attractors?

The beauty and power of single molecule studies is that they hold the potential to answer questions at the level at which biofolks (scientists of any flavor working on bioproblems) imagine and conceive molecular models. In fact, single molecule studies live in that magical space that lies below the imprecision of ensemble averages and at the same time remains above the uncertainty of quantum effects.

At present there is no equation for genetic signal transduction that is analogous to the Hodgkin–Huxley model that has been so foundational and ubiquitous in understanding neural or ionic signal transduction. The development of such a model is central to ongoing efforts in systems biology to interpret and ultimately predict metabolic, genomic and proteomic data. Such a model cannot be fully developed without a sound physical, mathematical theory of *fiber-ligand interaction*, because the control of genetic signal transduction involves dynamic fiber-ligand interactions along virtually every pathway. The first aim of this manuscript is to describe a novel paradigm with which it may be possible to study the dynamics of fiber ligand association at the single molecule level.

1.1   Inverse Problems

At the core of the approach that we are proposing lies an inverse problem. Many problems in experimental science are inverse problems of one sort or another (Sondhi & Gopinath, 1971), since it is often easy to measure the output of a system as a whole. If the output is uniquely determined by some internal property then it becomes possible to invert the known output to determine the internal property. The following example lies close to the heart of this report and has a long mathematical history (Kirsch, 1996). Consider a taut string of unit length, whose fixed ends correspond with the end points of the unit interval, with a variable mass density $\rho(x)$ along its length. It is well known that, if plucked, the string will vibrate with frequency ω and with a standing wave profile $w(x)$ if and only if the differential equation $w''(x) + \omega^2 \rho(x) w(x) = 0$ is satisfied. The *direct problem* is to compute the frequencies ω and their corresponding profiles, given a density function. The *inverse problem*, in this case, is to determine the density function from a recording of the vibrational frequencies.

The string in this problem is an example of an elastic medium, and the density is an example of one of its material properties. Very often in nature we

are able to record the behavior of a system, or develop devices to do so. The ability to use these recordings to infer material properties is clearly useful. The classic example of this is seismology, where the elastic medium is the ground, and how it vibrates in response to a sound wave, is used to determine its material properties and perhaps the presence of low density pockets of oil. Consequently, there is a large literature and huge interest in inverse problems involving elastic media.

Waves and vibrations have tremendous discriminatory power. This power has been exploited in many disciplines and applications, but surprisingly not in biophysical assays involving polymer association. This omission seems purely historical and not technical. In this manuscript we outline an inverse problem to detect molecular association of single molecules, and describe an algorithm that could conceivably be used for what we shall refer to as *molecular seismology*.

1.2 Application to Biophysics

It is currently of great interest to measure how proteins are binding to an individual DNA fiber in space and time. For this application the DNA can usefully and practically be conceptualized as an elastic fiber along whose length there exists a well defined density profile ρ. As a ligand interacts and becomes associated with a fiber, the density profile will change, in space along the fiber and in time if the ligand moves along the fiber or disassociates. Knowledge of the function $\rho(x,t)$ informs us about the dynamics of the association process between the ligand and the fiber, see Figure 1.

The central idea of molecular seismology is to subject single, solvated fibers to oscillation. How the induced waves propagate and scatter can depend sensitively and robustly on the material properties of the elastic fiber. Conversely, it is possible to determine the material properties of an elastic fiber from a measured trace of its vibration in response to a known impulse, see Figure 2. If one can induce and measure vibrations along a single DNA molecule, incubated with proteins, the dynamics of the density profile will reveal the associated motion of the proteins along the fiber. In principle, elastic deformations that produce compression, shear, or rotational waves could be used reconstruct time and space dependent material properties, any of which could directly report on the dynamics of molecular associations. We focus here on transverse oscillations and the density profile for clarity.

There is strong theoretical and practical evidence to suggest that dsDNA fibers can be accurately modeled as elastic fibers that will support elastic wave propagation. Measurements of such quantities as the persistence length, the length scale beyond which the elastic cost of bending is totally negligible, obtained from experiment are in good agreement with theoretical calculations of $50nm$, with a small sequence dependent variation. The gross elastic parameters of DNA have been measured over the last two decades, see for instance the reviews (Benham & Mielke, 2005; Hagerman, 1988) and many of the most relevant quantities are collected and discussed in the modeling work of (Bishop *et al.*, 2004). The data and modeling support the hypothesis that DNA fibers are

akin to soft materials like polyethylene or rubber. More recently, through single molecule force–extension measurements, it is well established that for tension below roughly 65 pN, dsDNA fibers are perfectly elastic. Many studies (Pant *et al.*, 2004; Sischka *et al.*, 2005; Smith *et al.*, 1992), to name only a few, have used single molecule techniques and the longitudinal overstretching transition to detect and monitor the binding of ligands to DNA. Further, the early landmark experiments of (Hatfield & Quake, 1999; Quake *et al.*, 1997) and (Crut *et al.*, 2003) definitively demonstrated that moderately stretched dsDNA fibers propagate transverse elastic waves. In fact both experiments directly measured the normal modes of individual, solvated, dsDNA fibers. Finally, all atom molecular dynamics simulations have shown that dsDNA fibers support elastic waves (Matsumoto & Olson, 2002) and that all atom molecular dynamics simulations agree with continuum PDE models of elastic rods (Bishop *et al.*, 2004). Thus, many lines of experimental and theoretical evidence strongly support the notion that dsDNA fibers support elastic waves and that these can be accurately modeled with continuum models.

1.3 Experimental Foundations

As we endeavor to investigate the feasibility of the molecular seismology approach and the considerations involved in building a device to measure the dynamics of protein-DNA interactions using this technique, we must investigate and answer at least the following four extended questions.

Which measurements lead to well posed inverse problems?

Well posed problems posses unique solutions that change continuously as a function of the input. From an experimental point of view this is the most crucial question. Which device designs will produce measurements that can robustly be inverted? Setting our initial expectations short of full inversion we might consider experimental designs that lend themselves to simple statistical signal detection schemes that could discriminate between unladen and laden fibers, or that can pattern match pre–recorded control signals.

In seismology, much consideration has been given to reflection problems: the source of vibrations and the detector are placed at the same boundary point (the surface of the earth). While, for finite elastic fibers, transmission problems would appear to be more natural. In a transmission problem vibrations are actuated from one end of the fiber and measured at the other. We have investigated a number of experimentally plausible configurations for molecular seismology.

What theorems are known concerning density reconstruction?

For deterministic, semi–infinite strings as well as finite strings, the reconstruction of material profiles from data have been shown to be well posed under mild smoothness assumptions on the profile of the material property.

Are there algorithms capable of density reconstruction?

Several algorithms have been proposed in the literature. We have extended one of these and use it to examine a set of feasible problems.

What are the effects of stochasticity?

As DNA–protein interactions take place in the condensed phase, thermal fluctuations and viscosity become important and can confound and degrade the ability to measure vibrations. Very little is known theoretically. A contribution of this paper is to examine these effects.

The remainder of this paper is addressed towards answering these questions.

2 Model, Notation and Known Results

The experimental work described in (Crut *et al.*, 2003; Quake *et al.*, 1997) justify the consideration of simple elastic models to describe and investigate the transverse vibrational motion of DNA fibers over length scales that exceed its persistence length. Correspondingly, we consider here perhaps the simplest mathematical description of an elastic fiber. We consider the one dimensional non homogeneous wave equation with damping as a model for the transverse oscillation of the centerline of a dsDNA fiber immersed in an aqueous heat bath,

$$\varrho(x)u_{tt}(x,t) + \nu u_t(x,t) - Eu_{xx}(x,t) = F(x,t). \tag{1}$$

The variable x represents the distance along the centerline of a finite elastic fiber and correspondingly takes values in a finite interval $[0, \ell]$. Here u represents the transverse displacement of the centerline. The mass density, $\varrho > 0$, is assumed to be sufficiently smooth, say twice differentiable (Symes, 1986). The optimal, or weakest, conditions under which inversion theorems can be proven appear to be unresolved. Most of the published algorithms can reconstruct piecewise constant material profiles. The friction coefficient ν and the tension E are assumed to be positive, known constants. Two related quantities are the wave speed $c := \sqrt{E/\varrho}$ and the impedance $\eta := \sqrt{E\varrho}$. The external force F can be either deterministic, stochastic or some combination of both. A large part of this investigation is devoted to demonstrating that density inversion is possible for signals from elastic fibers immersed in a frictive heat bath.

Throughout the remainder of this paper, the function $\varphi(t)$ will denote a boundary source. The function $\psi(t)$ will represent additional information corresponding to a measured response at a boundary point. The boundary data considered here are either of Dirichlet $u(0,t) := \varphi(t)$, or Neumann $u_x(0,t) := \varphi(t)$ type. The function φ is usually chosen to be an approximation of the Dirac δ–function, although it can be chosen quite arbitrarily, since any particular choice determines the response to an impulse $\delta(t)$, and this in turn determines the complete Neumann–to–Dirichlet map (Isakov, V., 1998, section 8.1). Given initial and boundary conditions and a density profile it is possible to solve the wave equation (1) and determine the function ψ. The goal is to study the inversion of this parameter–to–solution map $\varrho \mapsto \psi$.

The dynamics of the wave equation can be considered with a variety of different initial and boundary conditions. Historically, in exploration geology, shear waves are induced into a quiescent elastic medium at a boundary and

the reflected response is measured at the same boundary, the surface of the earth. It is very surprising that this setup, born of necessity, is in fact extremely powerful and robust. Consider a reflection problem in which the fiber is initially undisturbed and motion is excited through the boundary contact force $\varphi(t)$

$$u(x,0) = u_t(x,0) = 0,$$
$$u_x(0,t) = \varphi(t),$$
$$u(\ell,t) = 0,$$
$$\psi(t) = u(0,t).$$

Theoretical results concerning the identifiability of material properties are generally formulated for wave equations that have been transformed into travel time coordinates using the transformation

$$T_\varrho(s) = \int_0^s \varrho^{\frac{1}{2}}(x)\,\mathrm{d}x. \tag{2}$$

Since the density is assumed to be strictly positive, the travel time transformation (2) is injective. This transformation is extremely useful because it "straightens out" the characteristic curves, see for instance (Burridge, 1980; Pivovarchik & van der Mee, 2001; Symes, 1986), and makes it possible to define the parameter–to–solution map as a map between suitable function spaces. As the density ϱ is unknown in the inverse problem, it is not possible in a reflection problem to determine the minimal time $T := T(\ell)$ from which the knowledge of ψ allows the unique determination of ϱ.

An important observation in the reflection problem is that the boundary condition at the far end, ℓ, is immaterial, at least up to twice the travel time, $T(\ell)$, indeed, the string could continue past ℓ. This observation implies that results for the reflection problem while technically proven for semi–infinite strings, apply directly to finite strings in contrast to some existing results for transmission problems.

For the case $\varphi = \delta$, Symes proved that ψ on $[0, 2T]$ determines the impedance $\eta(z)$, provided $\log \eta \in H^1[0, T]$ (the Sobolev space consisting of those square integrable functions that possess a weak derivative that is also square integrable). In addition he characterized the range of the map $\eta \mapsto \psi$ and provided a local Lipschitz estimate of the type

$$||\eta_1 - \eta_2||_{H^1[0,T]} \le C||\psi_1(0,\cdot) - \psi_2(0,\cdot)||_{H^1[0,2T]} \tag{3}$$

where the ψ_i are the responses corresponding to the impedances η_i. An estimate of the type (3) is important with regard to the numerical stability of inversion algorithms as well as any statistical signal detection technique that might be considered to detect deviations from uniform density or impedance.

It has been shown in (Rakesh & Sacks, 1996) and further by (Alekseev & Belonosov, 1999) that for a semi–infinite string whose impedance is constant after a certain point, that the transmission and reflection problems are equivalent for the case of the Neumann–to–Dirichlet map. Further, (Rakesh, 1998) has shown

that impedance inversion from the Dirichlet–to–Neumann map is well posed for a transmission problem. However, the proof of this result depends on the fact that the string is semi–infinite and continues past the point of measurement.

There have been many results concerning density reconstruction for explicitly finite strings, see (Pivovarchik & van der Mee, 2001; Rundell & Sacks, 1992; Rundell, W. & Sacks, 1992) and references therein. Theorems involving finite strings with fixed or freely damped ends are most directly relevant to the physical problem we are interested in solving. In general, existing studies of finite strings have approached the inverse Cauchy problem in the frequency domain. Some recent developments are pointing the way toward developing time domain analysis and algorithms for finite strings (Browning, B. L., 2000).

3 Algorithms

Many algorithms have been proposed to solve inverse problems stemming from one dimensional wave equations similar to equation (1). Roughly speaking they can be divided into three classes. The earliest approaches, whose roots lie in quantum scattering theory, involve solving the classical Sturm–Liouville problem in the frequency domain. Excellent descriptions can be found in (Alekseev & Belonosov, 1999; Rundell, W. & Sacks, 1992) and in the references therein. A second set of algorithms follow from wave theoretic approaches that reformulate the inverse problem as a boundary value problem (Rakesh & Sacks, 1996; Rundell, W. & Sacks, 1992; Symes, 1983; Weston, 1972), and finite difference schemes have been published to solve reflection problems (Bube & Burridge, 1983; Claerbout, 1976; Santosa & Schwetlick, 1982) and references therein. A conceptually third set of methods are based on optimizing a suitably regularized objective cost functional that directly compares the measurement and the model (Symes, 1986; Tadi, 1997).

We have extended and implemented a variational method proposed in (Tadi, 1997). We have chosen this approach because it is flexible with respect to the model, flexible with respect to the boundary data, works directly with signals measured in real time (as opposed to travel time), and is robust with respect to noise. The robustness in the face of noise is not limited simply to additive noise in the measurement, but the method is robust with respect to natural noise arising from a frictive heat bath and modeled directly in the equations governing the elastic media.

Given a measurement ψ, we follow (Tadi, 1997) in defining a regularized, least squares type, cost functional for a pair (u, ϱ) to be

$$J(u, \varrho) = \frac{1}{2} \int_0^T (\psi(t) - u(0,t))^2 \, \mathrm{d}t + \frac{\alpha}{2} \int_0^T \int_0^\ell \varrho_t^2(x,t) \, \mathrm{d}x \, \mathrm{d}t. \qquad (4)$$

The constant $\alpha > 0$ is a regularization parameter (Engl *et al.*, 1996; Kirsch, 1996). The derivation and implementation of an optimization algorithm and its modification for alternative boundary conditions are described in detail in the appendices. The principal idea is to write the density profile as a separable

function of space and time, with a fixed spatial basis and time varying coefficient functions. The time dependent coefficient functions are allowed to evolve from an arbitrary initial state to some final equilibrium that presumably fits the true density. Notice that the regularization term goes to zero when the density profile becomes stationary in time. The Euler–Lagrange equations that result from the cost functional (4) describe the equations of motion for the time dependent coefficient functions and these equations form the basis for an iterative numerical algorithm that is described in the appendix.

The algorithm is of the steepest descent type, and often with non-convex variational problems, multiple local minima confound convergence to a global minimum from an arbitrary initial condition. For the problems that we are considering, a constant density profile is the most natural initial condition. Several algorithms have been proposed to deal with the existence of local minima (Klibanov & Timonov, 2004; Santosa & Schwetlick, 1982; Symes, 1986), prominent among them is the layer stripping technique. However, while it has been possible to describe layer stripping or downward continuation techniques for reflection problems, no such formulation has appeared for transmission problems, apart from the general convexification procedure described by Klibanov (Klibanov & Timonov, 2004). This stems from the important observation that reflections are time ordered according to distance traveled into the fiber, while even the earliest signals from transmission have sensed the entire fiber.

4   Frictive Heat Bath

One of the principle differences between the inverse problem in seismology and biology, apart from the scale, is the addition of a frictive heat bath. Intracellular biology happens in the condensed phase and the DNA fibers we wish to investigate are suspended in aqueous solution. The effects of a heat bath are traditionally and elegantly incorporated into molecular models through the combination of viscous dissipation and random forcing.

The Langevin equation has a long history in physics and mathematics and was developed to model the influence of a frictive heat bath, see for instance (Budini & Caceres, 2004; Chandler, 1987; Kubo, 1966; Zwanzig, 2001). The fluctuation dissipation theorem provides the insight required to relate the fluctuations of the random force to the viscous dissipation with or without memory (Budini & Caceres, 2004; Kubo, 1966). For spatially extended elastic systems like polymers, a widely adopted approach has been to consider the system as a discrete collection of beads, whose motion is described by a coupled system of Langevin equations (Crut *et al.*, 2003; Lamm & Szabo, 1986; Marciano & Brochard-Wyart, 1995). Alternatively, continuum models and algorithms have been developed to explore fluid-structure interactions (Peskin, 2002).

It is possible to analytically solve for the hydrodynamic force exerted on an oscillating but rigid cylinder (Rosenhead, 1963, VII 12). This result has been used recently with the beam equation, motivated by goals similar to ours, to

model the dynamics of cantilevers in solution using the fluctuation dissipation theorem (Clarke *et al.*, 2006; Paul & Cross, 2004; Sader, 1998). An analytical theory akin to that for the Langevin equation does not appear to exists for the wave equation and thus for soft elastic materials oscillating in solution. Of the few analytic results available for the wave equation, Caithamer and Belinskiy have shown that under suitable conditions on the spatial correlation of the noise, $f(x, y)$, the energy of a fiber in frictive heat bath converges exponentially to an equilibrium value that depends on a ratio of the noise to the damping coefficient (Belinskiy, B. P. & Caithamer, 2001*b*, Theorem 2).

In the absence of a broader theory for the wave equation, we model viscous dissipation by adding $\nu u_t(x, t)$ on the left hand side, and a random force $F(x, t) := \dot{G}(x, t)$ on the right hand side of equation (1). More complex models of dissipation involving memory kernels have been considered in the wave equation literature to model damping along a wave guide (Kroller & Probst, 1999) and can be incorporated into the model (1) if and when measurements or evidence support this extension. Typically, G is an (L^2-valued) Brownian motion, whose derivative is informally a stochastic process with expectations:

$$\mathsf{E}[F(x, t)] = 0, \tag{5}$$

$$\mathsf{E}[F(x, t)F(y, s)] = \sigma^2 \delta(t - s) f(x, y), \tag{6}$$

where f is called a spatial correlation function, see (Belinskiy, B. P. & Caithamer, 2001*a*,*b*).

As described in the previous section, and further in the appendix, the cost functional J can be utilized with equations that model a heat bath. We have utilized this approach to study density reconstruction from various boundary and initial conditions corresponding to potential experimental designs for manipulating dsDNA fibers in solution. We have identified two distinct and natural numerical approaches. In an approach denoted as *average over realizations* (AOR), a signal is recorded from the model with a realization of the random force F. The Euler-Lagrange equations (23) directly model this force, but while we have measured the system response $\psi(t)$, we do not have access to the particular realization of the random force. Thus, we perform the numerical reconstruction to convergence, with a random realization. The reconstruction procedure is performed m-times, each reconstruction, ϱ_i, from a realization F_i of the random force. It is natural to consider the pointwise average of these $\bar{\varrho} := \frac{1}{m} \sum_i^m \varrho_i$.

In a second approach denoted as *average over signals* (AOS) we take advantage of the fact that the average of the random force is zero. In this approach we average several recorded signals to produce $\mathsf{E}(\psi) := \bar{\psi}$, and apply the reconstruction algorithm with mean force equal to zero.

5 Results

Computations were performed with MATLAB, and the routines are available upon request. A standard finite–difference leapfrog scheme was used for the numerical

integration of all occurring wave equations (Quarteroni, A. *et al.*, 2000). Spatial discretization was accomplished using piecewise constant functions. The tension parameter was chosen $E = 1$. The random force was implemented using the function `randn`. In this study we do not explore spatially correlated noise and take $f(x, y) = \delta(x - y)$.

We anticipate, at least initially that a single protein will be bombarding a DNA fiber. For this reason, and because it is imagined that signal to noise issues will pose the dominant experimental challenge, it is sufficient for us to confine ourselves to the consideration of a simple density profile. The behavior of the cost functional for a reflection problem, without intrinsic noise and damping, was thoroughly explored on a large variety of density profiles in (Tadi, 1997). Our focus is on alternative boundary conditions and reconstructibility from the condensed phase and in order to meaningfully compare the performance as a function of noise strength it is important to have a gold standard density profile. Therefore, we confine our attention to the reconstructibility of the following smooth density profile:

$$\varrho(x) := 1 + \exp\left(-\frac{(x - 1/2)^2}{(1/5)^2}\right) \tag{7}$$

The natural starting density profile for the iterative algorithm is a constant density. In the AOR procedure it may be desirable to converge solutions from a variety of independent initial conditions, although we do not focus on this possibility in this paper.

A finite width, smooth approximation of a delta pulse was used as excitation in the Neumann–to–Dirichlet transmission and reflection problems. As with most iterative algorithms, one must choose a heuristic stopping criterion. A reasonable criterion for terminating the iteration procedure is to consider the improvement of the cost functional $\mathcal{S}(n) := J(\varrho^n)/J(\varrho^1)$.

The root mean square deviation or $||\varrho - \varrho^*||_2 := \sqrt{\int_0^1 (\varrho - \varrho^*)^2 \, dx}$ norm was used to quantitate the error of the reconstructions from the true density.

Through preliminary reconstructions it was observed that the presence of damping did not prevent convergence but rather only slowed the rate of convergence. Values of $\mathcal{S}$ as a function of damping strength are reported in table 1.

Given the data in table 1 and the theory described in (Belinskiy, B. P. & Caithamer, 2001*a,b*) it is reasonable to fix the damping parameter $\nu = 1$ in subsequent reconstructions and explore a range of noise strength.

5.1   Reconstruction from Reflection

Suppose the source $\varphi(t)$ and the measurement $\psi(t)$ are placed at the same boundary point, 0, of the interval $[0, \ell]$. There are 4 choices of boundary conditions that give rise to a well–posed direct problem. Since the boundary condition at the far end does not influence the response up to the two–way travel time

only two choices are relevant.

$$u_x(0,t) = \varphi(t), \quad u(\ell,t) = 0, \quad \psi(t) = u(0,t), \quad \text{and} \tag{8}$$

$$u(0,t) = \varphi(t), \quad u(\ell,t) = 0, \quad \psi(t) = u_x(0,t). \tag{9}$$

The canonical Neumann–to–Dirichlet boundary setup (8) proved to be the most robust in the sense that it produced the highest quality reconstructions with the fewest iterations, and the algorithm showed excellent convergence toward the true density, under all conditions of noise for both reconstruction strategies and with all density profiles. Reconstructions from signals with damping but without random forcing converge rapidly to the true density, as can be seen in Figure 3. Forty reconstructions from a noisy, damped string, each using a different realization of the random force together with their AOR reconstruction is shown in Figure 5. An AOS reconstruction is shown in Figure 4 in comparison with the true density.

Using a set of 600 independent reconstructions we applied a standard bootstrap algorithm (Efron, B. & Tibshirani, 1986) to estimate confidence intervals about the the true density for the AOR procedure as a function of the number of reconstructions, m, used to compute the average, see Figure 6. The results of this calculation show a $\sqrt{m}$ envelope that is expected from the Chebyshev inequality and the properties of the sample mean, and the derivative of this curve is estimated to be within 17% of stationary when n is 40. Using the results of this analysis we investigated the behavior of the AOS and AOR procedures as a function of increasing noise strength, see Figure 7. The data indicate that the error in the AOR reconstruction increases roughly linearly with the noise strength from approximately 3.5% error and appears to level off below 10% error (percentages are based on L_2 norm of the true density). In contrast the AOS reconstruction remained below 3% error and are seen to be relatively insensative to noise. Because the AOS procedure tacitly assumes that the noise averages out in the signal, this insensitivity is the expected behavior. This behavior is observed with $n = 5$, and the variance becomes small and stable for $n \geq 15$.

In contrast, the Dirichlet–to–Neumann reflection problem (9) was more subtle and we were not able to converge reconstructions through the cost functional optimization using simple Heaviside forcing. This generic forcing was not able to adequately induce discrimination between the constant and true density responses, while in the Neumann–to–Dirichlet problems the induced differences in signal were of the order of the signals themselves, see Figure 8. The failure of the cost functional optimization algorithm in this situation is purely numerical in the sense that theorems exists, as described in the previous sections, that guarantee that the inverse problem has a unique solution. Furthermore, the algorithm of (Santosa & Schwetlick, 1982), that is completely symmetric with respect to φ and ψ and consequently with respect to the two distinct boundary conditions, was able to exactly reconstruct the true density profile from impulse forcing for an undamped, unforced inhomogeneous wave equation in impedance form. However, the algorithm of (Santosa & Schwetlick, 1982) completely failed when even the slightest damping and/or noise were introduced. It is important

to observe that the two algorithms are founded on completely different logic and that the algorithm of (Santosa & Schwetlick, 1982) determines the density directly from a single response through its behavior along a special characteristic, while the cost functional optimization must be able to discriminate between the signals generated from the densities along the path determined by the Euler-Lagrange equations. Given these observations, we were able to converge reconstructions for problem (9) with cost functional optimization using sinusoidal forcing, see Figure 8. The obvious drawback is that in contrast to the generic impulse or Heaviside forcing, sinusoidal forcing requires a choice of frequency. Given that we know the true density in our theoretical experiments, it is straightforward to find such a frequency, but in an actual experiment this will be more difficult unless we can discover a systematic and natural solution to this problem. We are currently exploring this technical issue.

5.2   Reconstruction from Transmission

If the source and receiver are placed at opposite ends of the fiber then four distinct well–posed problems result.

$$u_x(0,t) = \varphi(t), \quad u(\ell,t) = 0, \quad \psi(t) = u_x(\ell,t), \tag{10}$$

$$u_x(0,t) = \varphi(t), \quad u_x(\ell,t) = 0, \quad \psi(t) = u(\ell,t), \tag{11}$$

$$u(0,t) = \varphi(t), \quad u(\ell,t) = 0, \quad \psi(t) = u_x(\ell,t), \quad \text{and} \tag{12}$$

$$u(0,t) = \varphi(t), \quad u_x(\ell,t) = 0, \quad \psi(t) = u(\ell,t). \tag{13}$$

In contrast to reflection problems, because of (17), $\psi(t) = 0$ for $0 \le t \le T_\varrho(\ell)$, the travel time can be determined. It is also important to observe that given any approximation to the true density provides an approximation the the travel time transformation, this fact could be utilized to impose an additional constraint in the cost functional.

The AOS procedure was relatively insensitivity to noise in the transmission problem (11), just as in the reflection problem. In contrast, however, the AOR procedure produced not only more accurate results than the AOS procedure, but the accuracy improved with increasing noise strength, in sharp contrast to the behavior observed in the reflection problem. A partial understanding of this phenomena is provided by observing and contrasting the individual reconstructions used to compute the AOR average. In the reflection problem it was observed that independent of the starting value of the constant density profile, the iterations generally approached the true density from below. While, in the transmission problem, the individual reconstructions, that differ in the particular realization of the random force, were far more uniformly distributed about the true density. In this way the noise appears to have allowed the iterations to approach the global minimizer from different directions on the cost surface.

Figure 6 illustrates the increase in precision of the AOR procedure as the number of reconstructions m, is increased. In general, the accuracy of an AOR reconstruction increases as the number of iterations, n, per realization–reconstruction is increased, as can be seen in Figure 10. The precision and

accuracy of the AOS procedure was seen to be completely analogous, where m is the number of signals averaged over and n the number of iterations in the zero–force reconstruction.

As described for the reflection problem the Dirichlet–to–Neumann transmission problem 12, the Dirichlet–to–Dirichlet problem 13 and the Neumann–to–Neumann problem 10 were less robust and required a careful consideration of the forcing function. Of these the Dirichlet–to–Dirichlet problem is perhaps the most interesting because it is by far the most natural experimental design. This is discussed in greater detail in the conclusions.

5.3   Homogeneous Boundary Conditions

In both the transmission and reflection problems described above, the motion of the fiber was actuated at a boundary point. In this and the following section we consider two problems in which data are collected at the boundaries but the vibrations are induced through initial data.

$$u(x,0) = u_0(x), \quad u_t(x,0) = v_0(x). \tag{14}$$

The problems are homogeneous in the sense that the boundary conditions are the same at both ends of the fiber. Two measurements are taken, the complementary data at both ends. Thus in the Dirichlet case we have

$$
\begin{aligned}
u(0,t) = 0, \quad u(\ell,t) = 0, \\
\psi_1(t) = u_x(0,t), \quad \psi_2(t) = u_x(\ell,t).
\end{aligned}
\tag{15}
$$

The homogeneous Neumann case

$$
\begin{aligned}
u_x(0,t) = 0, \quad u_x(\ell,t) = 0, \\
\psi_1(t) = u(0,t), \quad \psi_2(t) = u(\ell,t).
\end{aligned}
\tag{16}
$$

behaved in a qualitative and quantitatively similar way and we omit these results.

6   Conclusion

We have demonstrated that robust density reconstruction from synthetic signals of an elastic fiber in a viscous heat bath is possible from at least four different boundary conditions. From each of four experimentally relevant conditions, see Figure 12, we have found that not only can a density profile be reconstructed qualitatively, but quantitatively with errors that are less than 10% of the total, and more typically less than 3% of the total with both the AOR and AOS procedures. The AOS procedure was shown to be generally insensitive to strength of the noise. In at least one case we have shown that the presence of noise actually improves the reconstruction process. These data indicate that the process can succeed with measured experimental signals, provided that the model

for the elastic behavior is reasonably accurate. We are currently working on reconstruction from experimental signals.

We have found that Dirichlet–to–Neumann maps in reflection or transmission problems are more subtle and require the selection of nongeneric forcing for the cost functional optimization algorithm to converge to the true density. We are currently exploring this technical issue, however the larger point is that initial experiments should focus on either Neumann–to–Dirichlet or homogeneous problems. We complete the discussion with a physical illustration of the four boundary value problems that we have found to be robustly reconstructible and provide plausible experimental designs for each.

It is possible to design actuators and detectors that induce or measure changes in the position of a microparticle and more recently nanoparticles (Sonnichsen *et al.*, 2005). These changes correspond to Dirichlet conditions. It is not so straightforward to induce or measure Neumann conditions. Designing a method that alters $u_x(0, t)$ but not $u(0, t)$ on the nanoscale is challenging. One possibility is to apply a torque to the center of mass of a magnetic nanoparticle, or a nanoparticle with a non vanishing dipole moment or a birefringent nanoparticle (Friese, M. E. J. *et al.*, 1998*a,b*). Progress in organic chemistry and nanofabrication now provides the tools to chemically and mechanically manipulated DNA fibers to construct devices of the kind required to perform molecular seismology (Holzel *et al.*, 2003). Armed with these tools we describe hypothetical device designs corresponding to each of the four numerical problems that yield robust reconstructions.

Figure 12 contains four drawings corresponding to dsDNA fibers, decorated with nanoparticles and meant to be imagined as suspended in a fluidic chamber. The top setup corresponds to the Neumann–to–Dirichlet transmission problem. In this and other setups we imagine that the tension along the fiber is maintained by the hydrophobic force through a two phase separation. For instance the nanoparticles could be decorated with alkane chains and suspended in hexadecane, while the polar dsDNA fiber lies suspended in an aqueous channel. Two phase separations are among the most robust self organizing systems found in nature. The arrows indicate that the Neumann condition at the left end is actuated through a torque applied to the left nanoparticle while the Dirichlet measurement at the right end is accomplished by monitoring the displacement of the right bead. It has been shown that this could be accomplished by monitoring the plasmon resonance of a gold or silver nanoparticle by darkfield microscopy (Sonnichsen *et al.*, 2005).

The canonical reflection problem is depicted second from the top. The right end of the fiber is covalently attached to a support, the nanoparticle on the left and the fiber are freely floating in solution. Perhaps as a manifestation of Murphy's inviolable law, this setup that displays the most mathematical robustness appears physically to be the most challenging. The bottom two figures depict the homogeneous problems whose motion is not actuated by a boundary force. In the homogeneous Neumann problem, shown at the bottom, the motion of the fiber could be passively agitated by the bath, or the fibers motion could be induced by a short pulse from a laser tweezer. The vibrations

at both ends could be sensed as described above. In the homogeneous Dirichlet setup, both ends of the fiber are fixed and the torque on the beads is used to measure a derivative.

Based on the premise that some material profile along a single molecule of DNA will change as ligands associate with it, we have proposed an inverse problem to image this change in space and time. We have shown that both theoretically and numerically it is possible to invert density profiles from simple elastic models. Moreover we have shown that it is possible to do so from fibers immersed in a frictive heat bath in a natural way. Informed by known mathematical theorems, our numerical results, and the key technical developments in the field of nanofabrication and single molecule detection, we have proposed an experimental strategy with which it may be possible to record the signals required by our theory of molecular seismology.

While we have focused our development around transverse displacements of an elastic fiber, there is nothing intrinsic to inverse problems that would limit them to this case. The molecular seismology approach could be implemented mutatis-mutandis using instead longitudinal compression or rotational waves. The focus in this discussion has been on density profiles, but it is especially important to note that in solution it may well be that the more informative and/or tractable material property is $\nu(x)$. The larger, more important, abstract issue that should not get lost in the details is simply that some material property will change upon association and that a suitable inverse problem at the molecular level can be solved to determine it and track the dynamics of the process. Finally, we note that there is nothing special about DNA fibers or the particular elastic model, in fact the cost functional can accommodate beam equations that model the motion of immersed cantilevers where the theory could be applied to determine the mass distribution of DNA molecules adhered to it in space and time. We are currently exploring many of these alternatives.

7   Acknowledgment

A   Formulation of the variational problem

In this appendix we discuss the cost functional, its modification for alternative boundary conditions and the system of Euler-Lagrange equations that result

from its optimization.

We begin by considering the canonical reflection problem, for which the cost functional (4) and equations

$$u(x,0) = u_t(x,0) = 0, \tag{17}$$
$$u_x(0,t) = \varphi(t), \tag{18}$$
$$u(\ell,t) = 0 \tag{19}$$

define the set of admissible functions, together with the requirement that $\varrho > 0$. The cost functional is constrained to consider a particular model through the following standard construction.

$$G(x,t,u,\varrho) = \varrho(x,t)u_{tt}(x,t) + \nu u_t(x,t) - Eu_{xx}(x,t) - F(x,t) = 0.$$

This same device can be used to consider other models, linear or nonlinear, implicit or explicit, for elastic fibers or cantilevers. The augmented Lagrangian becomes

$$J^*(u,\varrho,\lambda) = \frac{1}{2}\int_0^T (\psi(t) - u(0,t))^2\,\mathrm{d}t \tag{20}$$
$$+ \int_0^T \int_0^\ell \left[\frac{\alpha}{2}\varrho_t^2 + \lambda G(x,t,u,\varrho)\right]\,\mathrm{d}x\,\mathrm{d}t,$$

where $\lambda(x,t)$ is the Lagrange multiplier. If $(\tilde{u},\tilde{\varrho},\tilde{\lambda})$ is an admissible variation then the variation of the augmented cost functional in this direction is

$$\delta J^*(u,\varrho,\lambda;\tilde{u},\tilde{\varrho},\tilde{\lambda}) = \int_0^T -(\psi(t) - u(0,t))\tilde{u}(0,t)\,\mathrm{d}t$$
$$+ \int_0^T \int_0^\ell \left[\alpha\varrho_t\tilde{\varrho}_t + (\varrho u_{tt} + \nu u_t - Eu_{xx} - F)\tilde{\lambda} + \lambda(u_{tt}\tilde{\varrho} + \varrho\tilde{u}_{tt} + \nu\tilde{u}_t - E\tilde{u}_{xx})\right]\,\mathrm{d}x\,\mathrm{d}t.$$

Performing integration by parts on all terms that contain $\tilde{\varrho}_t$, $\tilde{u}_{tt}$, and $\tilde{u}_{xx}$ gives

$$\delta J^*(u,\varrho,\lambda;\tilde{u},\tilde{\varrho},\tilde{\lambda})$$
$$= \int_0^T -(\psi(t) - u(0,t))\tilde{u}(0,t)\,\mathrm{d}t + \alpha \int_0^\ell \left[[\varrho_t\tilde{\varrho}]_0^T - \int_0^T \varrho_{tt}\tilde{\varrho}\,\mathrm{d}t\right]\,\mathrm{d}x$$
$$+ \int_0^T \int_0^\ell [(\varrho u_{tt} + \nu u_t - Eu_{xx} - F)\tilde{\lambda} + \lambda u_{tt}\tilde{\varrho}]\,\mathrm{d}x\,\mathrm{d}t$$
$$+ \int_0^\ell \left[[\lambda\varrho\tilde{u}_t]_0^T - [(\lambda\varrho)_t\tilde{u}]_0^T + \nu\,[\lambda\tilde{u}]_0^T + \int_0^T [(\lambda\varrho)_{tt} - \nu\lambda_t]\tilde{u}\,\mathrm{d}t\right]\,\mathrm{d}x$$
$$- E \int_0^T \left[[\lambda\tilde{u}_x]_0^\ell - [\lambda_x\tilde{u}]_0^\ell + \int_0^\ell \lambda_{xx}\tilde{u}\,\mathrm{d}x\right]\,\mathrm{d}t.$$

The boundary conditions on u require that we add variations which respect these, i.e.

$$\tilde{u}(x,0) = 0, \quad \tilde{u}_t(x,0) = 0,$$
$$\tilde{u}_x(0,t) = 0, \quad \tilde{u}(\ell,t) = 0.$$

This renders some of the boundary terms zero. We impose final conditions

$$\lambda(x,T) = 0, \tag{21}$$

$$\frac{\partial}{\partial t}\left(\lambda(x,t)\varrho(x,t)\right)\Big|_{t=T} = \lambda_t(x,T)\varrho(x,T) + \lambda(x,T)\varrho_t(x,T) = 0. \tag{22}$$

Because of condition (21), condition (22) only implies that

$$\lambda_t(x,T) = 0$$

is necessary. In addition the Lagrange multiplier must satisfy the spatial boundary condition

$$\lambda(\ell,t) = 0.$$

The boundary term containing ϱ must vanish,

$$\varrho_t(x,0)\tilde{\varrho}(x,0) = \varrho_t(x,T)\tilde{\varrho}(x,T) = 0.$$

This can only be achieved by requiring

$$\varrho_t(x,0) = \varrho_t(x,T) = 0,$$

since the alternative, imposing an initial respectively a final condition on ϱ itself is not sensible. We obtain the following necessary condition for a critical point of the augmented cost functional

$$\int_0^T \int_0^\ell \left\{ \left[\left(-\alpha\varrho_{tt} + \lambda u_{tt}\right)\tilde{\varrho} + \left(\varrho u_{tt} + \nu u_t - E u_{xx} - F\right)\tilde{\lambda} + \left((\lambda\varrho)_{tt} - \nu\lambda_t - E\lambda_{xx}\right)\tilde{u} \right] dx \right.$$
$$\left. + \left(-(\psi(t) - u(0,t)) - E\lambda_x(0,t)\right)\tilde{u}(0,t) \right\} dt = 0,$$

for all admissible variations. With the help of the Lagrange lemma (Troutman, 1983) we obtain the following partial differential equations as necessary conditions to a minimizer

$$\varrho u_{tt} + \nu u_t - E u_{xx} = F, \tag{23}$$
$$(\lambda\varrho)_{tt} - \nu\lambda_t - E\lambda_{xx} = 0, \tag{24}$$
$$-\alpha\varrho_{tt} + \lambda u_{tt} = 0. \tag{25}$$

together with the boundary condition

$$-(\psi(t) - u(0,t)) - E\lambda_x(0,t) = 0. \tag{26}$$

These equations are supplemented by the initial, final and boundary conditions.

A.1 Transmission Problem

To adapt the above strategy to the transmission problem (11) it is necessary to modify the cost functional to take into account the boundary conditions

$$J(u, \varrho) = \frac{1}{2} \int_0^T (\psi(t) - u(\ell, t))^2 \, dt + \frac{\alpha}{2} \int_0^T \int_0^\ell \varrho_t^2(x, t) \, dx \, dt.$$

The same partial differential equations (23)–(25) are obtained. The backward adjoint equation has to be supplied with the boundary condition

$$\lambda_x(0, t) = 0, \quad E\lambda_x(\ell, t) - (\psi(t) - u(\ell, t)) = 0.$$

A.2 Homogeneous Boundary Conditions

For the problem with homogeneous Dirichlet boundary conditions (14)–(15) the cost functional J becomes

$$J(u, \varrho) = \frac{1}{2} \int_0^T \left((\psi_1(t) - u_x(0, t))^2 + (\psi_2(t) - u_x(\ell, t)) \right)^2 \, dt$$

$$+ \frac{\alpha}{2} \int_0^T \int_0^\ell \varrho_t^2(x, t) \, dx \, dt.$$

The boundary conditions to the backward adjoint equation are derived

$$E\lambda(0, t) - (\psi_1(t) - u_x(0, t)) = 0, \quad -E\lambda(\ell, t) - (\psi_2(t) - u_x(\ell, t)) = 0.$$

For the inverse problem with homogeneous Neumann boundary conditions, the cost functional J becomes

$$J(u, \varrho) = \frac{1}{2} \int_0^T \left((\psi_1(t) - u(0, t))^2 + (\psi_2(t) - u(\ell, t))^2 \right) \, dt$$

$$+ \frac{\alpha}{2} \int_0^T \int_0^\ell \varrho_t^2(x, t) \, dx \, dt.$$

The boundary conditions to the backward adjoint equation are derived

$$-E\lambda_x(0, t) - (\psi_1(t) - u(0, t)) = 0, \quad E\lambda_x(\ell, t) - (\psi_2(t) - u(\ell, t)) = 0.$$

B The iterative algorithm

For completeness we describe the iterative algorithm used to solve the Euler–Lagrange system that results for the inversion problem from reflection data for the case of the Neumann–to–Dirichlet map. The method of solution for the other cases is completely analogous.

First we argue that there exists a unique optimal solution. Let the true density $\varrho^* = \varrho^*(x)$ be given. Let u^* be the corresponding solution of the wave

equation (1)–(19) with boundary conditions (8). Then clearly the boundary condition (26) implies that the Lagrange multiplier vanishes identically. This in turn implies that $\varrho_{tt} = 0$, and with $\varrho_t(0, x) = 0$ it follows that ϱ is independent of time. In summary, the triple $(u^*, \varrho^*, \lambda^* \equiv 0)$ is the unique solution of the system of partial differential equations (23)–(25). This implies that the pair (u^*, ϱ^*) is a critical point of the regularized cost functional J from equation (4) constrained to the level set $\{G(u, \varrho) = 0\}$, with $J(u^*, \varrho^*) = 0$ as the critical value. Of course, since $\alpha > 0$, this is the global minimal value of J.

Suppose that an approximate density $\varrho^n = \varrho^n(x)$, $n \geq 1$ has already been calculated. During an iteration step this time–independent density is used to determine the solution u^n of the forward wave equation and the solution λ^n of the backward adjoint equation

$$\varrho^n(x)u_{tt}^n(x, t) + \nu u_t^n(x, t) - E u_{xx}^n(x, t) = F,$$
$$\varrho^n(x)\lambda_{tt}^n(x, t) - \nu \lambda_t^n(x, t) - E \lambda_{xx}^n(x, t) = 0,$$

subject to the boundary condition

$$-(\psi(t) - u^n(\ell, t)) + E \lambda_x^n(\ell, t) = 0.$$

Using u^n and λ^n, the equation

$$\varrho_{tt}(x, t) = -\frac{1}{\alpha}\lambda^n(x, t)u_{tt}^n(x, t) \tag{27}$$

is integrated forward in time, with initial conditions $\varrho(0, x) = \varrho^n(x)$, $\varrho_t(0, x) = 0$. To ensure that the value of the cost functional is decreased, we have to switch the "direction" along which $\varrho(x, t)$ evolves, hence the minus sign in equation (27), (Zhdanov, 2002). The density is updated according to

$$\varrho^{n+1}(x) = \varrho^n(x) - \frac{1}{\alpha}\int_0^T \int_0^\tau \lambda^n(s, x)u_{tt}^n(s, x)\, ds\, d\tau.$$

References

Alekseev, A. S. & Belonosov, V. S. (1999). Direct and inverse problems of wave propogation through a one dimensional inhomogeneous medium. *European J. Appl. Math.* **10**, 79–96.

Belinskiy, B. P. & Caithamer, P. (2001*a*). Energy of a string driven by a two–parameter gaussian noise white in time. *J. Appl. Probab.* **38**, 960–974.

Belinskiy, B. P. & Caithamer, P. (2001*b*). Energy of an elastic mechanical system driven by Gaussian noise white in time. *Discrete Contin. Dynam. Systems, Added Volume*, 39–49.

Benham, C. J. & Mielke, S. P. (2005). DNA mechanics. *Annu. Rev. Biomed. Eng.* **7**, 21–53.

Bishop, T. C., Cortez, R. & Zhmudsky, O. O. (2004). Investigation of bend and shear waves in a geometrically exact elastic rod model. *J. Comput. Phys.* **193**, 642–665.

Browning, B. L. (2000). Time and frequency domain scattering for the one-dimensional wave equation. *Inverse Problems,* **16**, 1377–1403.

Bube, K. P. & Burridge, R. (1983). The one dimensional inverse problem of reflection seismology. *SIAM Review,* **25**, 497–559.

Budini, A. A. & Caceres, M. A. (2004). Functional characterization of generalized Langevin equations. *J. Phys. A,* **37**, 5959–5981.

Burridge, R. (1980). The Gelfand–Levitan, the Marchenko, and the Gopinath–Sondhi integral equations of inverse scattering theory, regarded in the context of inverse impulse problems. *Wave Motion,* **2**, 305–323.

Chandler, D. (1987). *Introduction to Modern Statistical Mechanics.* Oxford University Press, New York.

Claerbout, J. F. (1976). *Fundamentals of Geophysical Data Processing : with Applications to Petroleum Prospecting.* McGraw–Hill, New York.

Clarke, R. J., Jensen, O. E., Billingham, J., Pearson, A. P. & Williams, P. M. (2006). Stochastic elastohydrodynamics of a microcantilever oscillating near a wall. *Physcal Review Letters,* **96**, 050801–1–4.

Crut, A., Lasne, D., Allemand, J. F., Dahan, M. & Desbiolles, P. (2003). Transverse fluctuations of single DNA molecules attached at both extremities to a surface. *Phys. Rev. E,* **67**, 051910.

DeGroot, M. H. (1986). *Probability and Statistics.* Addison–Wesley, Reading, MA.

Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science,* **1**, 54–77.

Engl, H. W., Hanke, M. & Neubauer, A. (1996). *Regularization of Inverse Problems.* Kluwer Academic Publishers, Dordrecht, Boston.

Friese, M. E. J., Nieminen, T. A., Heckenberg, N. R. & Rubinsztein-Dunlop, H. (1998*a*). Optical alignment and spinning of laser trapped microscopic particles. *Nature,* **394**, 348–350.

Friese, M. E. J., Nieminen, T. A., Heckenberg, N. R. & Rubinsztein-Dunlop, H. (1998*b*). Optical torque controlled by elliptic polarization. *Optics Letters,* **23**, 1–3.

Gowers, D. M., Wilson, G. G. & Halford, S. E. (2005). Measurement of the contributions of 1D and 3D pathways to the translocation of a protein along DNA. *Proc. Natl. Acad. Sci. USA,* **102**, 15883–15888.

Graneli, A., Yeykal, C. C., Robertson, R. B. & Greene, E. C. (2006). Long distance lateral diffusion of human rad51 on double stranded dna. *Proc. Natl. Acad. Sci.* **103**, 1221–1226.

Hagerman, P. J. (1988). Flexibility of DNA. *Annu. Rev. Biophys. Chem.* **17**, 265–286.

Halford, S. E. & Marko, J. F. (2004). How do site-specific DNA–binding proteins find their targets. *Nucleic Acids Research,* **32**, 3040–3052.

Hatfield, J. W. & Quake, S. R. (1999). Dynamic properties of an extended polymer in solution. *Phys. Rev. Lett.* **82**, 3548–3551.

Holzel, R., Gajovic-Eichelmann, N. & Bier, F. F. (2003). Oriented and vectorial immobilization of linear m13 dsdna between interdigitating electrodes-towards single molecule dna nanostructures. *Biosensors and Bioelectronics,* **18**, 555–564.

Isakov, V. (1998). *Inverse Problems for Partial Differential Equations.* Springer Verlag, New York.

Kirsch, A. (1996). *An introduction to the mathematical theory of inverse problems.* Springer Verlag, New York.

Klibanov, M. V. & Timonov, A. (2004). A unified framework for constructing globally convergent algorithms for multidimensional coefficient inverse problems. *Appl. Anal.* **83**, 933–955.

Kroller, M. & Probst, G. (1999). On the truncation of the infinite sum of exponentials in a memory kernal of the linear wave equation. *Angew. Math. Mech.* **79**, 603–613.

Kubo, R. (1966). The fluctuation dissipation theorem. *Rep. Prog. Phys.* **29**, 255–284.

Lamm, G. & Szabo, A. (1986). Langevin modes of macromolecules. *J. Chem. Phys.* **85**, 7334–7348.

Marciano, Y. & Brochard-Wyart, F. B. (1995). Normal modes of stretched polymers. *Macromolecules,* **28**, 985–990.

Matsumoto, A. & Olson, W. K. (2002). Sequence–dependent motions of DNA: A normal mode analysis at the base pair level. *Biophys. J.* **83**, 22–41.

Pant, K., Karpel, R. L., Rouzina, I. & Williams, M. C. (2004). Mechanical measurement of single–molecule binding rates: Kinetics of DNA helix–destabilization by T4 gene 32 protein. *J. Mol. Biol.* **336**, 851–870.

Paul, M. R. & Cross, M. C. (2004). Stochastic dynamics of nanoscale mechanical oscillators immersed in a viscous fluid. *Physcal Review Letters,* **92**, 23501–1–23501–4.

Peskin, C. S. (2002). The immersed boundary method. *Acta Numerica,* **11**, 479–517.

Pivovarchik, V. N. & van der Mee, C. (2001). The inverse generalized Regge problem. *Inverse Problems,* **17**, 1831–1845.

Quake, S. R., Babcock, H. & Chu, S. (1997). The dynamics of partially extended single molecules of DNA. *Nature,* **388**, 151–154.

Quarteroni, A., Sacco, R. & Saleri, F. (2000). *Numerical Mathematics.* Springer Verlag, New York.

Rakesh (1998). Inversion of spherically symetric potentials from boundary data for the wave equation. *Inverse Problems,* **14**, 999–1007.

Rakesh & Sacks, P. (1996). Impedance inversion from transmission data for the wave equation. *Wave Motion,* **24**, 263–274.

Rosenhead (1963). Laminar boundary layers. In *Fluid Motion Memoirs*, (Rosenhead, L., ed.), Oxford University Press.

Rundell, W. & Sacks, P. (1992). The reconstruction of Sturm–Liouville operators. *Inverse Problems,* **8**, 457–482.

Rundell, W. & Sacks, P. (1992). Reconstruction techniques for classical inverse Sturm–Liouville problems. *Math. Comput.* **58**, 161–183.

Sader, J. E. (1998). Frequency response of cantilever beams immersed in viscous fluids with applications to the atomic force microscope. *Journal of Applied Physics,* **84**, 64–76.

Santosa, F. & Schwetlick, H. (1982). The inversion of acoustical impedence profile by method of characteristics. *Wave Motion,* **4**, 99–110.

Sischka, A., Toensing, K., Eckel, R., Wilking, S. D., Sewald, N., Ros, R. & Anselmetti, D. (2005). Molecular mechanisms and kinetics between DNA and DNA binding–ligands. *Biophys. J.* **88**, 404–411.

Smith, S. B., Finzi, L. & Bustamante, C. (1992). Direct mechanical measurements of the elasticity of single DNA molecules using magnetic beads. *Science,* **258**, 1122–1126.

Sondhi, M. M. & Gopinath, B. (1971). Determination of the vocal tract shape from impulse response at the lips. *J. Acoust. Soc. Amer.* **49**, 1867–1873.

Sonnichsen, C., Reinhard, B. M., Liphardt, J. & Alivisatos, A. P. (2005). A molecular ruler based on plasmon coupling of single gold and silver nanoparticles. *Nature Biotechnology,* **23**, 741–745.

Symes, W. W. (1983). Impedance profile inversion via the first transport equation. *J. Math. Anal. Appl.* **94**, 435–453.

Symes, W. W. (1986). Stability properties for the velocity inversion problem. In *Mathematical and Computational Methods in Seismic Exploration and Reservoir Modeling*, (Fitzgibbon, W. E., ed.), pp. 128–157, SIAM, Philadelphia, PA.

Tadi, M. (1997). Explicit method for inverse wave scattering in solids. *Inverse Problems,* **13**, 509–521.

Troutman, J. (1983). *Variational Calculus with Elementary Convexity.* Springer Verlag, New York.

von Hippel, P. H. & Berg, O. G. (1989). Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678.

Wang, Y. M., Austin, R. H. & Cox, E. C. (2006). Single molecule measurements of repressor protein 1d on dna. *Physical Review Letters,* **97**, 483021–483024.

Weston, V. H. (1972). On the inverse problem for a hyperbolic dispersive partial differential equation. *J. Math. Phys.* **13**, 1952–1956.

Zhdanov, M. S. (2002). *Geophysical Inverse Theory and Regularization Problems.* Elsevier, Amsterdam.

Zwanzig, R. (2001). *Nonequilibrium Statistical Mechanics.* Oxford University Press, Oxford, New York.

C   Figures and Tables

ν	$\mathcal{S}_R(200)$	$\mathcal{S}_R(400)$	$\mathcal{S}_T(200)$	$\mathcal{S}_T(400)$
0	$1.23 \cdot 10^{-5}$	$1.47 \cdot 10^{-6}$	$1.53 \cdot 10^{-4}$	$2.24 \cdot 10^{-5}$
1	$2.2 \cdot 10^{-3}$	$1.18 \cdot 10^{-4}$	$2.96 \cdot 10^{-2}$	$9.11 \cdot 10^{-3}$
2	$3.23 \cdot 10^{-2}$	$1.08 \cdot 10^{-2}$	$5.7 \cdot 10^{-2}$	$3.9 \cdot 10^{-2}$

Table 1: Normalized improvement of the cost functional $\mathcal{S}$ after 200 and 400 iterations for a reflection (R) and a transmission (T) problem.
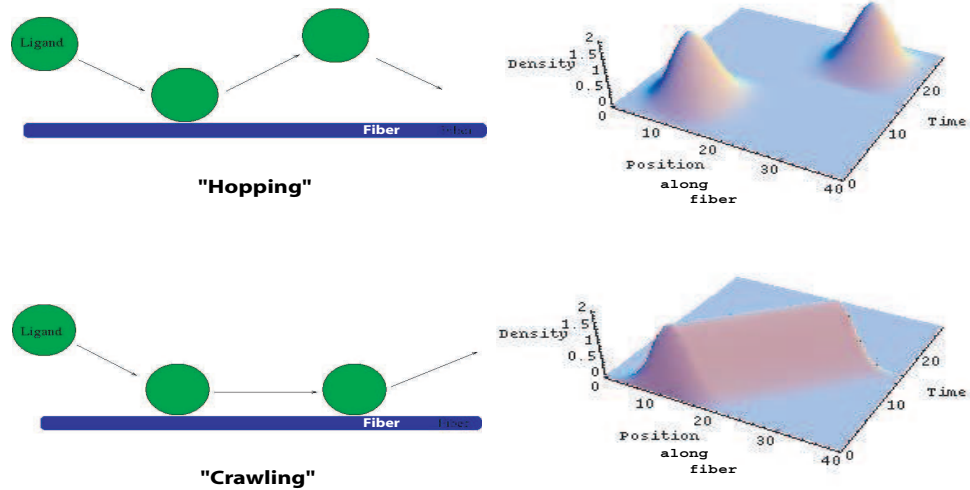
Figure 1: A density profile encodes dynamic temporal and spatial events. Two of the typical DNA association modes discussed in the literature are *hopping* and *crawling*. The top panel depicts a density profile corresponding to hopping, the lower panel describes crawling. During crawling the protein remains in constant contact with the fiber. The shape of the humps encode important kinetic parameters related to the rates of association, residence times, and rates of dissociation. An important point is that the density profiles for these two events are not only distinguishable from each other but their space–time profiles directly encode molecular association dynamics.
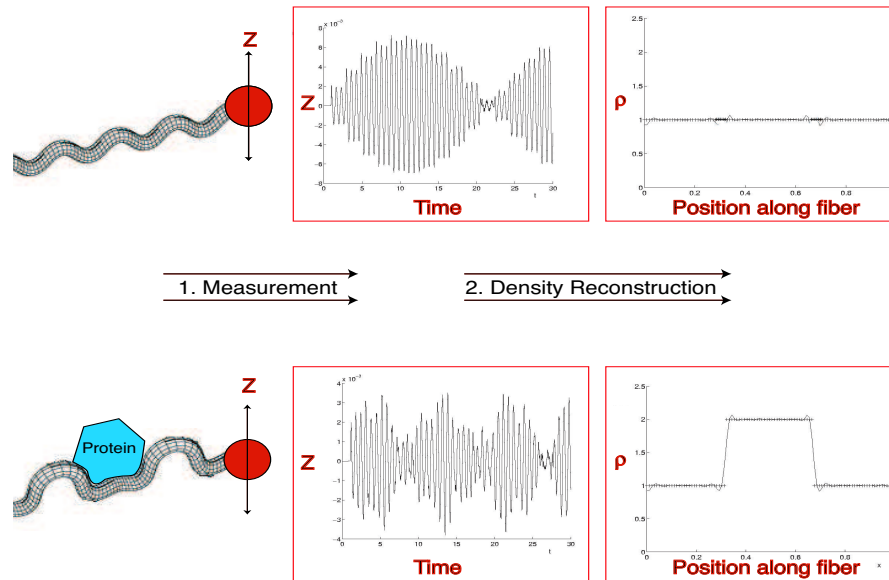
Figure 2: This figure illustrates the density reconstruction procedure that underlies the *molecular seismology* paradigm. The upper left panel shows the oscillations of an unladen single DNA fiber. The vibrations of a bead at the end of the fiber are recorded as a function of time (center panels). The density reconstruction procedure takes recorded oscillations as input and solves an inverse problem to produce a density profile in space along the fiber. For the unladen fiber this profile is nearly flat. The crosses in the figure represent the actual density, while the solid line shows a reconstruction. In contrast, the lower panels show the process for a fiber that has a protein associated at its center. The oscillations are different. The inversion procedure produces a density profile that describes where along the fiber the protein is bound. Through repeated measurements a time and space dependent density profile can be assembled to reveal the molecular dynamics.
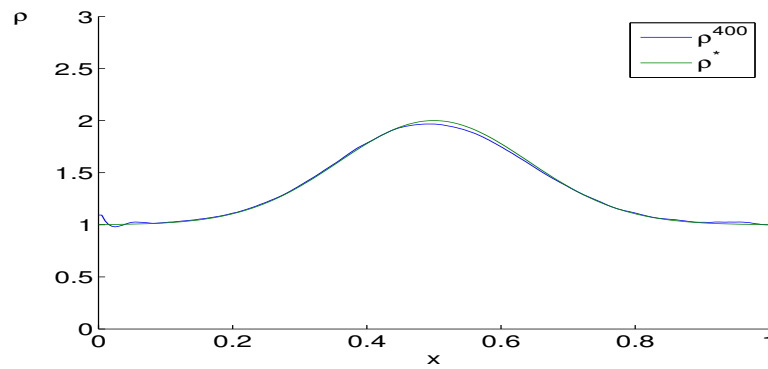
Figure 3: A comparison of the true (as defined in Equation 7) and reconstructed densities profiles after 400 iterations of algorithm described in Appendix B
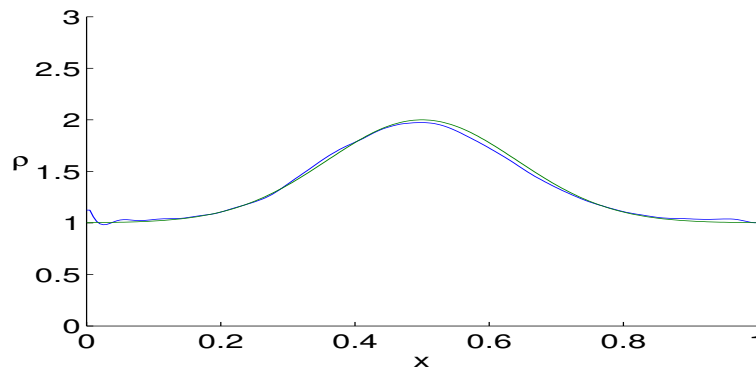
Figure 4: A comparison of the true density profile with that of an AOS reconstruction from an signal averaged over forty independent realizations with parameters $\nu = 1$ and $\sigma^2 = 4$.
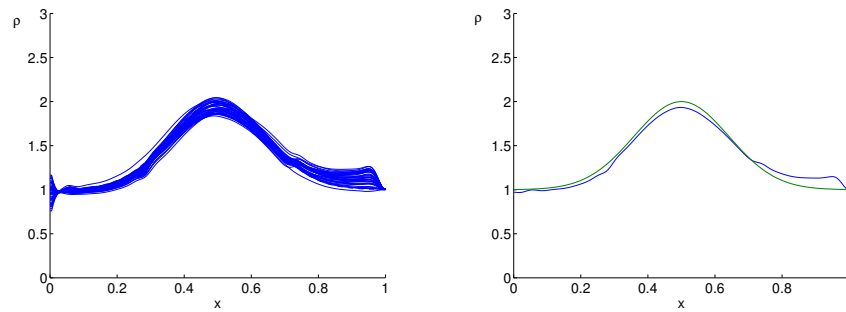
Figure 5: The left panel shows 40 independent reconstructions for the canonical reflection problem 8, each from the same recorded signal, but with different realizations of the random force. The right panel shows their AOR pointwise average density profile in comparison with the true density profile.
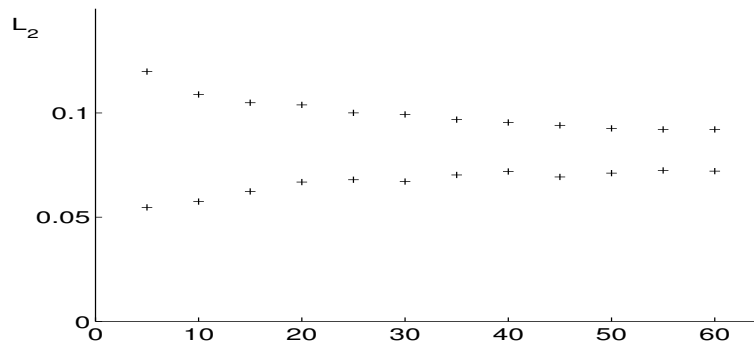
Figure 6: Bootstrapped confidence intervals about the true density for the AOR procedure applied to the canonical reflection problem (8) as a function of m, the number of iid reconstructions used in the average. The data exhibit the behavior expected from the Chebyshev inequality, see (DeGroot, 1986)[Sect 4.8].
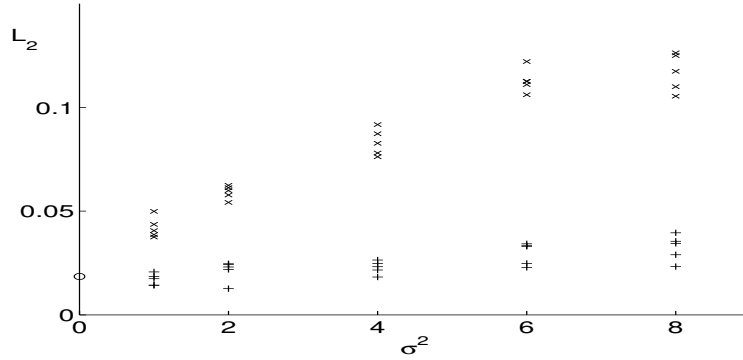
Figure 7: L_2 error in reconstruction as a function of noise strength σ^2 for the reflection problem (8). The data are the results from five independent reconstructions at each of five different noise strengths. Each cross (x) represents an error in an AOR average over a total of 40 independent reconstructions. Each plus (+) represents an analogous result for an AOS reconstruction from 40 independent signals. The open circle corresponds to a reconstruction from a signal generated in the absence of noise.
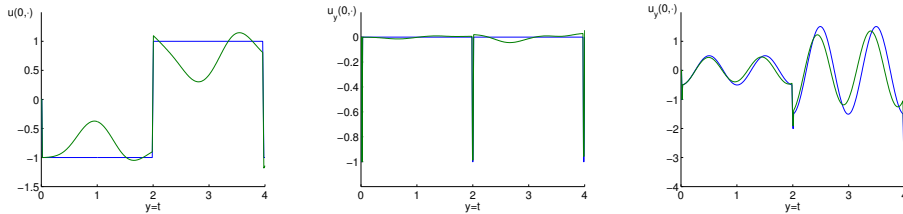
Figure 8: A comparison of the responses of a fiber with uniform impedance profile (shown in blue) with that of the inhomogeneous profile (Equation 7). The leftmost panel shows the difference in the impulse responses for the Neumann–to–Dirichlet problem. The differences in the signal are clearly seen and of the order of the signals themselves. The middle panel shows the differences in the Heaviside responses for the Dirichlet–to–Neumann problem. The differences in the responses are not of the same order as the signals. The direct algorithm of (Santosa & Schwetlick, 1982) is able to reconstruct the true impedance from this signal while the iterative algorithm described in this paper could not. The right panel shows the differences in response to the sinusoidal forcing $u(0,t) = 8\sin(2\pi t)$ for the Dirichlet–to–Neumann problem. These responses are sufficiently different that our iterative algorithm can reconstruct the true impedance or density.
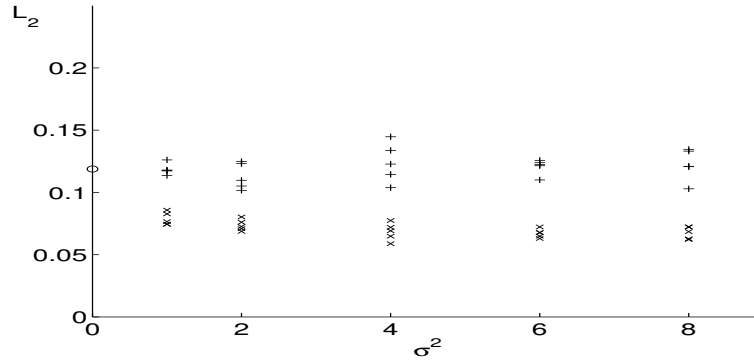
Figure 9: L_2 error in reconstruction as a function of noise strength σ^2 for the transmission problem (11). The data are the results from five independent reconstructions at each of five different noise strengths. Each cross (x) represents an error in an AOR average over a total of 40 independent reconstructions. Each plus (+) represents an analogous result for an AOS reconstruction from 40 independent signals. The open circle corresponds to a reconstruction from a signal generated in the absence of noise.
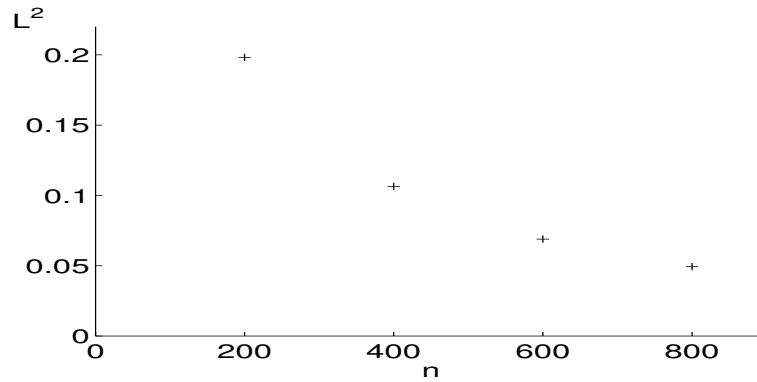

Figure 10: L_2 error of an AOR reconstruction for the transmission problem (11) as a function of the number of iterations per reconstruction. Each data point represents an average over $m = 40$ independent realization–reconstructions. The individual reconstruction were iterated for various lengths, n.
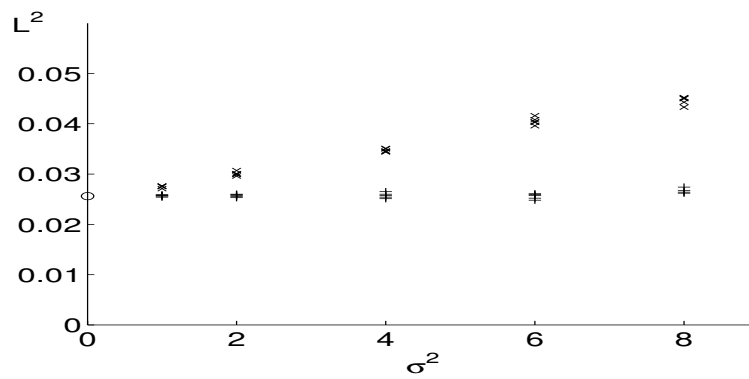
Figure 11: L_2 error in reconstruction as a function of noise strength σ^2 for the reflection problem (15). The data are the results from five independent reconstructions at each of five different noise strengths. Each cross (x) represents an error in an AOR average over a total of 40 independent reconstructions. Each plus (+) represents an analogous result for an AOS reconstruction from 40 independent signals. The open circle corresponds to a reconstruction from a signal generated in the absence of noise.
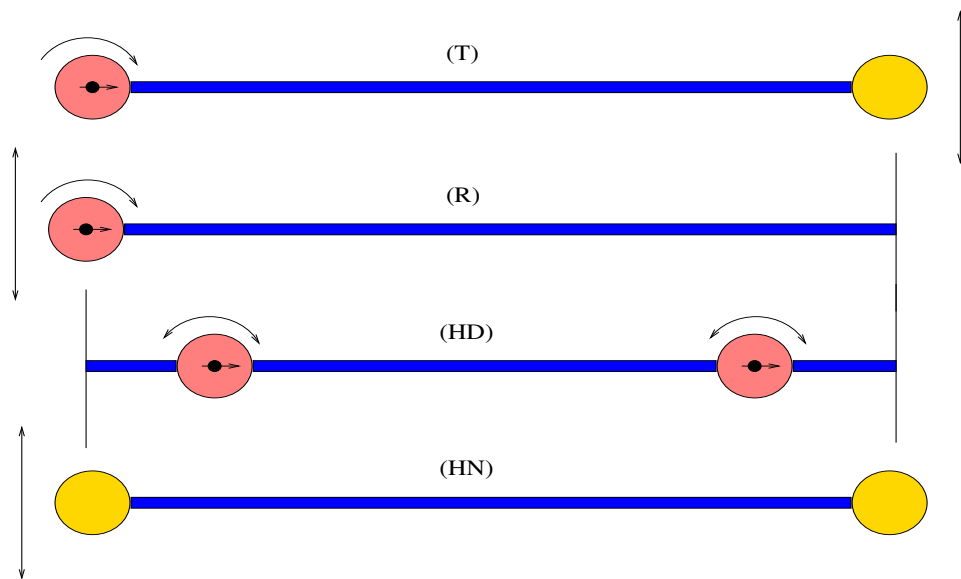
Figure 12: Schematic representation of four experimental designs corresponding to each of the four boundary conditions that produced robust numerical reconstructions. In each figure, the blue fiber represents dsDNA, the spheres represent covalently attached nanoparticles, the vertical arrows imply displacement and the curved arrows imply torque. The design (T) corresponds to the transmission problem (11). (R) diagrams the canonical reflection problem (8). The homogeneous Dirichlet problem (15) is labeled (HD), and the bottom figure corresponds to the homogeneous Neumann problem (16). Actuators and sensors are described in the text.