Energy variational analysis of ions in water and channels: Field theory for primitive models of complex ionic fluids

Bob Eisenberg,1,a) YunKyong Hyon,2 and Chun Liu2,3
1Department of Molecular Biophysics and Physiology, Rush University, Chicago, Illinois 60612, USA
2Institute for Mathematics and its Applications, University of Minnesota, Minneapolis, Minnesota 55455, USA
3Department of Mathematics, Pennsylvania State University, University Park, State College, Pennsylvania 16802, USA

(Received 31 May 2010; accepted 16 July 2010; published online 9 September 2010)

Ionic solutions are mixtures of interacting anions and cations. They hardly resemble dilute gases of uncharged noninteracting point particles described in elementary textbooks. Biological and electrochemical solutions have many components that interact strongly as they flow in concentrated environments near electrodes, ion channels, or active sites of enzymes. Interactions in concentrated environments help determine the characteristic properties of electrodes, enzymes, and ion channels. Flows are driven by a combination of electrical and chemical potentials that depend on the charges, concentrations, and sizes of all ions, not just the same type of ion. We use a variational method EnVarA (energy variational analysis) that combines Hamilton’s least action and Rayleigh’s dissipation principles to create a variational field theory that includes flow, friction, and complex structure with physical boundary conditions. EnVarA optimizes both the action integral functional of classical mechanics and the dissipation functional. These functionals can include entropy and dissipation as well as potential energy. The stationary point of the action is determined with respect to the trajectory of particles. The stationary point of the dissipation is determined with respect to rate functions (such as velocity). Both variations are written in one Eulerian (laboratory) framework. In variational analysis, an “extra layer” of mathematics is used to derive partial differential equations. Energies and dissipations of different components are combined in EnVarA and Euler–Lagrange equations are then derived. These partial differential equations are the unique consequence of the contributions of individual components. The form and parameters of the partial differential equations are determined by algebra without additional physical content or assumptions. The partial differential equations of mixtures automatically combine physical properties of individual (unmixed) components. If a new component is added to the energy or dissipation, the Euler–Lagrange equations change form and interaction terms appear without additional adjustable parameters. EnVarA has previously been used to compute properties of liquid crystals, polymer fluids, and electrorheological fluids containing solid balls and charged oil droplets that fission and fuse. Here we apply EnVarA to the primitive model of electrolytes in which ions are spheres in a frictional dielectric. The resulting Euler–Lagrange equations include electrostatics and diffusion and friction. They are a time dependent generalization of the Poisson–Nernst–Planck equations of semiconductors, electrochemistry, and molecular biophysics. They include the finite diameter of ions. The EnVarA treatment is applied to ions next to a charged wall, where layering is observed. Applied to an ion channel, EnVarA calculates a quick transient pile-up of electric charge, transient and steady flow through the channel, stationary “binding” in the channel, and the eventual accumulation of salts in “unstirred layers” near channels. EnVarA treats electrolytes in a unified way as complex rather than simple fluids. Ad hoc descriptions of interactions and flow have been used in many areas of science to deal with the nonideal properties of electrolytes. It seems likely that the variational treatment can simplify, unify, and perhaps derive and improve those descriptions.

I. INTRODUCTION

Variational methods that generalize and optimize energy functionals allow understanding of complex fluids.1–4 Variational methods deal successfully with magnetohydrodynamics systems,5 liquid crystals, polymeric fluids,6 colloids and suspensions,1,7 and electrorheological fluids.8,9 Variational methods describe solid balls in liquids, deformable electrolyte droplets that fission and fuse,1,10 and suspensions of lipoids, including the interfacial properties of these complex mixtures, such as surface tension and the Marangoni effects of “oil on water” and “tears of wine.”1,7,11

Solid charged spheres such as sodium and chloride ions in water seemed to be a simpler fluid than deformable fissioning droplets (in some respects) and so we wondered if...
energy variational methods could help us understand these ionic solutions. We try to create a field theory of ionic solutions that uses only a few fixed parameters to calculate most properties in flow and in traditional thermodynamic equilibrium, both in bulk and in spatially complex domains like pores in channel proteins. We derive the differential equations of the field theory from an energetic variational principle \textit{EnVarA} (energy variational analysis). Ionic solutions are often highly concentrated and so packing effects not present in infinitely dilute solutions are significant and can in fact dominate. Interactions in concentrated environments are responsible for essential characteristics of electrodes, enzymes, and ion channels.

Our variational principle combines the maximum dissipation principle (for long time dynamics) and least action principle (for intrinsic and short time dynamics) into a force balance law that expands the law of conservation of momentum to include dissipation, using the generalized forces in the variational formulation of mechanics (p. 19 of Ref. 20, see Refs. 21 and 22). This procedure is a modern reworking of Rayleigh’s dissipation principle—Eq. (26) of Ref. 24—motivated by Onsager’s treatment of dissipation. Our procedure optimizes both the action functional (integral) of classical mechanics and the dissipation functional. The stationary point of the action is determined with respect to the trajectory of particles. The stationary point of the dissipation is determined with respect to rate functions (such as velocity). Both are written in Eulerian (laboratory) coordinates. These functionals can include entropy and dissipation as well as potential energy, and can be described in many forms on many scales from molecular dynamics calculations of atomic motion, to Monte Carlo (MC) simulations to—more practically—continuum descriptions of ions in water. We use a primitive model of ions in an implicit solvent and adopt self-consistent treatments of electrodiffusion in which the charge on ions help create their own electric field. We introduce the repulsion energy of solid spheres. In this way, the variational calculus extends the primitive model to spatially complex, non-equilibrium time dependent situations, creating a field theory of ionic solutions.

Energy functional integrals and dissipation functional integrals are written from specific models of the assumed physics of a multicomponent system, as did Refs. 1, 2, and 8–10. Components of the potential energy and dissipation functions are chosen so the variational procedure produces the drift diffusion equations of semiconductor physics, called the Vlasov equations in plasma physics, or the similar biophysical Poisson–Nernst–Planck equations—named PNP by Ref. 45—and used since then by many channologists and physical chemists. The energy of the repulsion of solid spheres is included in our functional as (1) Lenard-Jones forces giving (as their Euler–Lagrange equations) a generalization of PNP for solid ions or as (2) that of uncharged spheres in the density functional theory of fluids with similar but not identical results. Boundary conditions tell how energy and matter flow into the system and from phase to phase and are described by a separate variational treatment of the “interfacial” energy and dissipation. The resulting Euler–Lagrange equations are the boundary value problems of our field theory of ionic solutions. They are derived by algebra and solved by mathematics—without additional physical approximations—in spatially complex domains, that perhaps produce flow of nonideal mixtures of ions in solution.

Ionic solutions do not resemble the ideal solutions of elementary textbooks. Indeed, ions such as Na\(^+\) and K\(^+\) have specific properties, and can be selected by biological systems, because they are nonideal and have highly correlated behavior. Screening and finite size effects produce the correlations more than anything else. Solvent effects enter (mostly) through the dielectric coefficient. Ionic solutions do not resemble a perfect gas of noninteracting uncharged particles. Indeed, because of screening, the activity (which is a measure of the free energy) of an ionic solution is not an additive function as concentration is changed (Fig. 3.6 of Ref. 37; Fig. 4.2.1 of Ref. 38) and so does not easily fit some definitions (p. 6 of the book of international standards for physical chemistry) of an extensive quantity.

Some correlations are included explicitly in our models as forces or energies that depend on the location of two particles. Other correlations are implicit and arise automatically as a mathematical consequence of optimizing the functionals even if the models used in the functionals do not contain explicit interactions of components. Kirchhoff’s current law (that implies perfect correlation in the flux of electrical charge) arises this way as a consequence of Maxwell’s equations and does not need to be written separately. Variational analysis produces “optimal” estimates of the correlations that arise from those interactions (and p. 11 of Biot; p. 42 of Gelfand and Fromin) and gives the hope that fewer parameters can be used to describe a system than in models and equations of state of ionic solutions which involve many parameters. These parameters change with conditions and are really functions or even functionals of all the properties of the system. (It is important to understand that in general these coupling parameters need to depend on the type and concentration of all ions, not just the pair of ions that are coupled.)

Nonideal properties are evident in all properties of ionic mixtures and most properties of ionic solutions relevant to biology and electrochemistry. Indeed, nonideal properties are responsible for the defining characteristics of many biological and electrochemical systems of great practical importance. Nonideal properties have been investigated by a generation of chemists and include (much of the lifework of Mayer, Barthe, Friedman, Hansen, Henderson, Pitzer, and Lee, and many others). Nonideal solutions require Onsager reciprocal relations with parameters that depend on the type and concentration of all ions. Models of forces between atoms in molecular dynamics are calibrated for the most part under ideal conditions of infinite dilution and so do not include (for the most part) the complex effects of concentration found in measurements of mixtures (see Ref. 96 and references cited there).

\textit{EnVarA} does not produce a single boundary value prob-
lem or field equation for ionic solutions. Rather, it produces different field equations for different models (of correlations produced by screening or finite size, for example) to be checked by experiment. If a new component is added to the energy or dissipation functionals, the resulting Euler–Lagrange equations change form so that the solutions of the equations reflect all the interactions of the new and old components of the energy without introducing parameters (besides those in the model of the energies and dissipations themselves). Of course, the variational approach can only reveal correlations arising from the physics and components that the functional actually includes. Correlations arising from other components or physics need other models and will lead to other differential equations. For example, ionic interactions that arise from changes in the structure of water would be an example of “other physics,” requiring another model, if they could not be described comfortably by a change in the diffusion coefficient of an ion or a change in the dielectric constant of water. Numerical predictions of EnVarA will be relatively insensitive to the choice of description (of pairwise interactions, for example) because the variational process in general produces the optimal result\(^\text{21,22,76,78}\) for each version of the model. (This is an important practical advantage of the variational approach: compare the success of the variational density functional theory of fluids\(^\text{3,5,6,6,–6,8}\) with the nonvariational mean spherical approximation\(^\text{35–38,90,91,97–100}\) that uses much the same physics.)

All field equations arising from EnVarA optimize both the dissipation and the action integrals. Inadequate functionals can be corrected (to some extent) by adjusting parameters in the functional. Effective parameters are almost always used to describe complex interactions of ions in electrolyte solutions,\(^\text{34,70,90,92,93,106–104}\) e.g., the cross coupling Onsager coefficients\(^\text{93,94}\) or Maxwell–Stefan coefficients.\(^\text{105,106}\) Our work could be extended by inverse methods\(^\text{95,107}\) to provide estimators\(^\text{107}\) of the parameters of EnVarA functionals with least variance or bias, or other desired characteristics.

Our field theory EnVarA represents an ionic solution as a mixture of two fluids,\(^\text{108}\) a solvent water phase and an ionic phase. The ionic phase is a primitive model of ionic solutions,\(^\text{35–38,90,97–99}\) It is a compressible plasma made of charged solid (nearly hard) spheres. The ionic “primitive phase” is itself a composite of two scales, a macroscopic compressible fluid and an atomic scale plasma of solid spheres in a frictional dielectric. Channel proteins are described by primitive (“reduced”) models similar to those used to analyze the selectivity of calcium and sodium channels\(^\text{14,15,17–19,52,109–112}\) and to guide the construction (using the techniques of molecular biology) of a real calcium channel protein in the laboratory.\(^\text{50,115}\) Similar models predicted complex and subtle properties of the RyR channel before experiments were done in >100 solutions and in seven mutations, some drastic, removing nearly all permanent charge from the “active site” of the channel.\(^\text{51,77,114–116}\)

This paper is organized into a (1) biological introduction, (2) a theoretical introduction, (3) a long theoretical section with numbered subsections for clarity in navigation, (4) a computational methods section, (5) results, and (6) discussion. The introductions are more complete than customary as we reach to disparate communities of scientists. We try not to mystify anyone anywhere and regret that we are likely to patronize (and irritate) everyone, somewhere.

**II. BIOLOGICAL SETTING**

One of our motivations is biological. The role of ions has been a central topic in medicine and biology\(^\text{117,118}\) since (at least) Fick (a biologist, actually a physiologist\(^\text{19,120}\), described diffusion. The interacting flows of ions produce the ATP (adenosine-5’-triphosphate) (from photosynthesis and oxidative phosphorylation) that fuel life.\(^\text{115}\) Interacting flows of ions produce the volume regulation that allows animal cells to exist.\(^\text{121}\) Flows of ions (and their interactions) produce signaling in the nervous system, initiation of contraction in muscle, including the coordination of contraction that allows the heart to function as a pump, movement of water into the kidney and out of the stomach and intestine.\(^\text{118}\) Nearly all biological processes depend on ions. Biophysical chemists\(^\text{39,122,123}\) and molecular biologists\(^\text{39,41,124}\) have long dreamed that a physical theory of ions near and in proteins could provide decisive help in understanding biological function.

Biological cells, proteins, and nucleic acids are found in solutions that are plasmas of ions—ultrafiltrates of blood created by the sieving action of macroscopic pores between the cells that form capillaries. Ions in water can be called “the liquid of life” with more color than hyperbole. Ions of the biological plasma surround and pervade the proteins and nucleic acids inside a cell and inside its organelles. Ions are extraordinarily concentrated inside crevices and channels in proteins where much of their function is thought to occur in “active sites.” Concentrations are 20 M or more in the active sites of proteins and selectivity filters of channels, and nearly as large around the double helix of DNA. (Pure water has number density [H\(_2\)O]\(\approx\)55 M.)

Biological plasmas outside cells are mostly sodium (\(~140\) mM) and chloride (\(~102\) mM) and bicarbonate ions (\(~20\) mM) mixed with small but important concentrations of potassium (\(~4\) mM) and calcium and magnesium ions (\(~1\) mM). Different ions carry different “messages” through different channels selective to one type of ion or another. The selectivity of channels for ions is a subject of the greatest biological importance. Hundreds of selective channels are described in the four “Ion Channel Fact Books.”\(^\text{125}\)

Ionic solutions inside cells are rich in potassium (\(~120\) mM) and chloride and organic anions (\(~105\) mM), with smaller concentrations of sodium (\(~15\) mM) and bicarbonate ions (\(~25\) mM) and with trace (\(<10^{-6}\) M) concentrations of calcium and other messenger molecules. Trace concentrations (typically <\(\mu\)M) of ions [e.g., calcium, cyclic AMP (3’-5’-cyclic adenosine monophosphate), inositol-tris-phosphate, even sodium] are signals that act as messengers to control many biochemical systems within cells.\(^\text{126}\) Some important ionic messengers act in concentrations of \(~10^{-11}\) M.\(^\text{127}\) Measurements, simulations, and theories of such a range of concentrations are hard to perform and even harder to calibrate.
The gradient of concentration between the inside and outside of cells is a crucial energy source for the membrane phenomena that control a wide range of biological functions, as are gradients of concentrations across the membranes of many intracellular organelles. Ion concentrations inside cells are controlled by biological (nano) valves called ion channels ("pumps") as ions move in pores within proteins across otherwise impermeable lipid membranes. It is surprising, but true, that many complex properties of channels selective for calcium or sodium ions can be understood using adaptations of the primitive model of ions in bulk solutions fulfilling the early dreams of biophysical chemists, to some extent.

In these reduced models of selectivity, the channel protein enters in a crucial but limited way. The protein contributes no mechanical or chemical energy to the ions inside it (in the simplest version of these models, however, see Ref. 110). The protein determines the size and shape of the pore in which ions are confined. It determines the mechanical and dielectric environment (i.e., the polarization charge) in and around the pore, and it provides side chains (of the amino acids that form the polypeptide backbone of the protein) that mix with the ions and water in a crowded mixture (~20 M) in the pore of the channel. The side chains are often acidic (i.e., have permanent negative charge) or basic (i.e., have permanent positive charge), creating electric fields of great strength (~5 × 10^9 V m^-1). The location of the side chains in these models is an output of the calculations;22 they are not kept at preordained positions in a binding site, for example, because the computed positions change importantly as experimental conditions are changed. The structure of the side chains in these models is their computed positions. The structure is self-organized and the side chains have an induced fit to the mobile ions and vice versa.

The competition between electrostatic forces and crowding (produced by the finite size of the ions and the comparable size of the confining space) can be simulated by classical (originally Metropolis) MC methods developed to study bulk solutions or by quite approximate theories of bulk solutions or by the ad hoc but powerful density functional theory (DFT) of fluids applied to ion channels which is quite different from the better known density functional theory of electrons in orbitals.

These simulations and theories of simple models of ions in crowded confined spaces allow understanding of one of the most important properties of proteins, selectivity, because they compute the nonideal properties of ions that depend on screening, the finite size of the ion, the shape and size of the confining space, and on the concentration of all species of ions. The quite different sodium channels of nerve and calcium channels of the heart are both described well by a single model with the same two fixed parameters in a wide range of solutions of different composition and content. Each channel type is represented only by spheres taking the place of its characteristic amino acid side chains that produce selectivity (Glu Glu Glu Glu for calcium channels; Asp Glu Lys Ala for sodium channels). The simulations are confined to thermodynamic equilibrium, where there are no flows of any kind. The model is subject to all sorts of appropriate objections mostly because of its evident lack of the atomic detail of the protein. The simulations are surprisingly successful, nonetheless. Reduced models seem to describe the kinds of energy used by these channels to create selectivity, probably because they allow the concentrations of every ionic species to change the activity (i.e., free energy per mole) of every other ion. Of course, evolution is likely to use other forms of energy as well in other situations.

One of the Nobel Laureates who founded molecular biology (Aaron Klug) recently said, “There is only one word that matters in biology, and that is specificity. The truth is in the details, not the broad sweeps,” reiterating the common view that selectivity can only be understood in atomic detail. The reduced model of selectivity shows that the broad sweeps of physics (in MC simulations of thermodynamic equilibrium) can compute the biological detail, at least in these calcium and sodium channels at equilibrium.

Variational analysis can extend these equilibrium simulations to nonequilibrium so they can predict current flow, creating a field theory using the same physics. EnVarA automatically calculates the time dependence of currents as the solution of its Euler–Lagrange equations. Of course, channel proteins use specialized structures to produce time dependence, the phenomena called “gating,” and these will not be described by EnVarA unless those structures and their energies are explicitly included in the calculations.

III. THEORETICAL SETTING

In this paper, we only use an energy variational analysis distinct from other variational principles that have been used to analyze the Vlasov equation in general or at thermodynamic equilibrium (the Poisson–Boltzmann equation).

The energy variational treatment of complex fluids starts with the energy dissipation law:

\[
\frac{dE}{dt} + \Delta = 0, ~ (1)
\]

where we use the dissipation function \( \Delta \) of Onsager, which usually includes a linear combination of the squares of various rate functions (such as velocity, rate of strain, etc). In a classical Hamiltonian conservative system, the energy \( E \) is the sum of kinetic and internal energies involving an integral of the energy of individual particles over all space,

\[
E = T + U. ~ (2)
\]

The Lagrangian framework of mechanics writes the energy of a set of \( i \) particles in terms of the motion \( x(X; t) \) of these particles using the action \( A(x(X; t)) \) of these trajectories. The \( i \) particles are labeled by the set of their initial locations \( X = \{X_i\} \) at \( t=0 \).
The first steps of the Legendre transformation\(^{148}\) (pp. 32–39 of Ref. 149) give the action of the trajectories of the particles described by Eq. (2) in terms of the trajectories \(x(\bar{X},t)\),

\[
A = \int (T - U) dt.
\]

The principle of least action optimizes the action \(A\) with respect to all trajectories \(x(\bar{X},t)\) by setting to zero the variation with respect to \(x\), computed over the entire domain \(\Omega\),

\[
\delta A = \delta_x A = \delta_x A(x)
\]

\[
= \int \int_{\Omega_0} \left[ \text{Conservative Force} \right] \cdot \delta x \dot{\bar{X}} dt,
\]

where \(\Omega_0\) is the reference domain of \(\Omega\).

If \(\delta_x A\) is set to zero, this Eq. (4) becomes the weak variation form\(^{20,150}\) of the conservative force balance equation of classical Hamiltonian mechanics, a statement of the conservation of momentum. We use the word “force” in the generalized\(^{11}\) sense of classical Hamiltonian mechanics (p. 19 of Refs. 20 and 21) and extend the classical treatment to include dissipation\(^{20,21,22}\) and then later to microscopic scales [Eq. (8)] so we can deal with the transport energy of ions in ionic channels.

Dissipation will be treated by extending the classical treatment of the Hamiltonian to include dissipation forces\. When classical mechanics adds dissipation into Eq. (2) by the Rayleigh dissipation principle\(^{20,21}\) the physical meaning of the left hand side of Eq. (1) changes. The system is no longer conservative. It is no longer a system constrained to thermodynamic equilibrium. Confusion in the names and physical meaning of variables is likely to result (see pp. 62 and 64 of the classical textbook\(^{20}\).

We treat dissipation by performing a variation with respect to the velocity \(\dot{u}\) in Eulerian (laboratory) coordinates as

\[
\delta_x \left( \frac{1}{2} \Delta \right) = \int_\Omega \left[ \text{Dissipative Force} \right] \cdot \delta \dot{u} \delta \bar{X}.
\]

If \(\delta_x \left( \frac{1}{2} \Delta \right)\) is set to zero, Eq. (5) gives a weak variational form of the dissipative force balance law equivalent to conservation of momentum, using the word force to include the variation of dissipation with respect to velocity. The velocity is sometimes written more explicitly as

\[
\dot{u}(\bar{x}(\bar{X},t);t) = \frac{\partial \bar{x}(\bar{X},t)}{\partial t} = \dot{x}(\bar{X},t).
\]

Ionic solutions satisfy both dissipative force balance and conservative force balance, so we have the following equation, which can be written in either Lagrangian coordinate \(\bar{X}\) or Eulerian coordinate \(x\)\(^{151}\)

\[
\begin{array}{c|c}
\text{Conservative Force} & \text{Dissipative Force} \\
\hline \\
\frac{\partial E}{\partial \bar{x}} & \frac{1}{2} \frac{\partial \Delta}{\partial \dot{u}} \\
\end{array}
\]

One can imagine systems constrained to follow other balance “laws” beyond those in Eq. (7). Such constraints can be included in our variational analysis essentially by adding them into Eq. (7), because the theory of optimal control (Ref. 78 and p. 42 of Ref. 22) uses the variational calculus to apply constraints or penalty functions (p. 120 of Ref. 76). See discussion below, a few paragraphs before Eq. (19).

Equation (7) is nearly identical to Eq. (26) of Rayleigh.\(^{24}\) We go beyond Rayleigh by actually solving the resulting Euler–Lagrange partial differential equations, together with the physical boundary conditions. We use the modern theory of the calculus of variations and corresponding numerical algorithms that reflect the underlying variational structures, all implemented by computational resources not available in the 19th century. Our dissipation function (like Biot’s\(^{21}\) departs from Onsager’s—loosely defined between Eqs. 5.6 and 5.7 on p. 2227 of Ref. 26—because we use variations with respect to two functions\(^{20,150}\) The dissipative—Onsager(ian)—part of the expression uses a variation with respect to the rate function (velocity \(\dot{u}\)\(^{29}\) The conservative—Hamilton(ian)—part of the expression uses a variation with respect to position \(\dot{x}\). The combined results should be expressed in the same coordinate, either Lagrangian coordinates or Eulerian coordinates. A “pushforward” or “pullback” change of variables\(^{151}\) may be needed to write all the physical quantities in the same coordinate, if those quantities were originally defined in different coordinates.

### A. Theoretical model: Transport of ions

The transport of ions through ionic channels is an atomic scale problem, because the diameter of channels is only 2–4\(\times\) the diameters of ions. Valves like ion channels are designed to be much smaller than the systems they control. Biology carries this “to the limit” by making its nanovalves into picochannels with internal diameters about twice as large as the flowing molecules (atoms). A stochastic analysis of the trajectories of atoms is possible\(^{152,153}\) and a multidimensional (nonequilibrium) Fokker–Planck equation can be derived by analysis\(^{39}\) or steepest descent arguments.\(^{154}\) The multidimensional Fokker–Planck equation can be reduced\(^{155}\) to the PNP equations by a closure procedure\(^{39,156,157}\) with no more (than the considerable) arbitrariness of closures of equilibrium systems\(^{32,158}\) (which do not form obviously convergent or uniformly convergent series, for example, and thus have unknown errors). The Fokker–Planck equation includes the continuity equation and so does its derivatives, the drift diffusion or PNP equation. \(\text{Envart}\) of transport starts with an energy law that can be derived from treatments of Brownian motion, as just mentioned, but we prefer an axiomatic approach—guess the law; check the result—in which we treat Eq. (7) as a postulate, valid on the macro-, meso-, and atomic scale for two reasons: (1) present treatments of closure do not allow estimates of their errors and (2) the actual dynamics of charged particles in water and protein channels may not be well described\(^{36}\) by models of the Brownian motion of uncharged particles with independent noise sources, although this has been the custom for some
time. Fluctuations in the number density of ions are likely to produce fluctuating forces not included in the treatment of Brownian motion of uncharged particles.

B. Theoretical model: Energy variational derivation of the Fokker–Planck equation of transport

We define a generalized potential \( \Phi(f(\mathbf{x})) \) for the probability density function \( f(\mathbf{x}) \) of ions and assume that electrodiffusion of particles is only driven by gradients of

\[
E = \int_{\Omega} \Phi(f(\mathbf{x})) d\mathbf{x} = \int_{\Omega} \frac{\partial \Psi}{\partial f} \log f(\mathbf{x}) d\mathbf{x} + \int_{\Omega} \psi(\mathbf{x}) f(\mathbf{x}) d\mathbf{x},
\]

where \( \psi(\mathbf{x}) \) includes both the electrostatic potential \( \phi(\mathbf{x}) \) and also the steric repulsion arising from the finite volume of solid ions [see Eqs. (24)–(27)] and \( \Phi(f(\mathbf{x})) \) includes logarithmic entropy terms; \( \beta = 1/k_B T \) where \( T \) is the absolute temperature and \( k_B \) is the Boltzmann constant. Both electrostatic and steric repulsion forces are global forces depending on boundary conditions, the location of particles everywhere, and spatial variation of parameters like dielectric coefficients. In general, neither force can be written as functions of the position of only two particles.\( ^{14,159,155} \)

When the potential \( \psi(\mathbf{x}) \) is only electrostatic \( \psi(\mathbf{x}) = \phi(\mathbf{x}) \), then \( \Phi(f(\mathbf{x})) \) describes the transport properties of an ideal gas of point charges (including screening) often described by the drift diffusion, Vlasov, or PNP equations, as mentioned previously. Even in this case, without finite size effects, the system is highly nonideal and hardly extensive since screening produces powerful correlations. The excess free energy per mole varies (more or less) as the square root of the concentration and not linearly as assumed in the “independence principle” of classical electrophysiology.\( ^{160} \)

Now, we take the variation of the potential \( \Phi(f(\mathbf{x})) \) with respect to \( \mathbf{x} \), in the Eulerian coordinate. Since the potential \( \Phi(f(\mathbf{x})) \) is a functional of \( f \), not \( \mathbf{x} \), we need to use the chain rule if we want to determine the spatial variation of the potential and thus the force. We write the chain rule here—only for motivation—as if variations were derivatives,

\[
\frac{\delta}{\delta \mathbf{x}} \Phi(f(\mathbf{x})) = \frac{\delta}{\delta f} \Phi(f(\mathbf{x})) \cdot \frac{\delta f}{\delta \mathbf{x}}.
\]

The variation of \( \Phi(f(\mathbf{x})) \) with respect to \( f(\mathbf{x}) \) is

\[
\delta E(f(\mathbf{x})) = \delta \Phi(f(\mathbf{x})) = \int_{\Omega} \frac{\partial \Phi(f(\mathbf{x}))}{\partial f(\mathbf{x})} \delta f(\mathbf{x}) d\mathbf{x}
\]

\[= \int_{\Omega} \mu \delta f(\mathbf{x}) d\mathbf{x}, \quad (9)\]

where the chemical potential \( \mu \) appears because it is the derivative of \( \Phi(f(\mathbf{x})) \) with respect to density \( f(\mathbf{x}) \),

\[
\mu(f(\mathbf{x})) = \frac{\partial}{\partial f(\mathbf{x})} \Phi(f(\mathbf{x})). \quad (10)
\]

We introduce the flux \( \mathbf{J} \) by its definition,

\[
\frac{\partial f(\mathbf{x})}{\partial t} + \nabla \cdot \left( f(\mathbf{x}) \frac{\partial \mathbf{x}}{\partial t} \right) = 0. \quad (11)
\]

Flux \( \mathbf{J} \) is the product of density and the velocity and is the variation of the density \( f(\mathbf{x}) \) with respect to \( \mathbf{x} \), i.e., \( \delta f(\mathbf{x}) = -\nabla \cdot (f(\mathbf{x}) \partial f(\mathbf{x})/\partial \mathbf{x}) \). Substitute this variation into Eq. (9), and integrate by parts, to get

\[
\delta E(f(\mathbf{x})) = \delta J_{\mathbf{J}}(f(\mathbf{x})) = \int_{\Omega} f(\mathbf{x}) \nabla \cdot (f(\mathbf{x}) \nabla \mu(\mathbf{x})). \quad (12)
\]

The long time dynamics are governed by the transport force \( f(\mathbf{x}) \nabla \mu(\mathbf{x}) \) and the transport law. At long times, the velocity is proportional to force, and the divergence of flux is equal to the time rate of change of contents,

\[
\frac{\partial f(\mathbf{x})}{\partial t} = -\xi \nabla \cdot \mathbf{J} = \xi \nabla \cdot (f(\mathbf{x}) \nabla \mu(\mathbf{x})), \quad (13)
\]

where \( \xi \) is a renormalized constant. Equation (13) seems to be a nonlinear equation, but in fact, a little algebra shows that it is the Fokker–Planck equation,\( ^{3,153} \) describing the diffusion and drift of stochastic trajectories of density \( f(\mathbf{x}) \).

\[
\frac{\partial f(\mathbf{x})}{\partial t} = \frac{\xi}{\beta} \nabla^2 f(\mathbf{x}) + \xi \nabla \cdot (f(\mathbf{x}) \nabla \psi(\mathbf{x})), \quad (14)
\]

where \( \nabla^2 \) is the Laplace operator.

Equation (14) is the field equation of ionic transport on the micro (i.e., atomic) scale and \( f(\mathbf{x}) \) is actually a distribution function, i.e., a probability density function that may not have been normalized.\( ^{59,152,155} \) Later we will create a more complete model of the ionic phase by combining the atomic scale description of transport [as the drift diffusion process of Eq. (14)] with a macroscopic description of the flow of a compressible ionic fluid (as a Navier–Stokes process), thereby generalizing the traditional primitive model of ionic solutions into a field theory.

Equation (13) is now written as a variation \( \delta_{\mathbf{x}} \Phi(f(\mathbf{x})) \) with a procedure used often in variational analysis. This procedure shows that Eq. (13) is also a dissipation—a time derivative. First, we multiply Eq. (13) by \( \partial \Phi(f(\mathbf{x}))/\partial f = \mu(\mathbf{x}) \), and integrate,

\[
\int_{\Omega} \frac{\partial \Phi(f(\mathbf{x}))}{\partial f(\mathbf{x})} \frac{\partial f(\mathbf{x})}{\partial t} d\mathbf{x} = \xi \int_{\Omega} \mu(\mathbf{x}) \nabla \cdot (f(\mathbf{x}) \nabla \mu(\mathbf{x})) d\mathbf{x}. \quad (15)
\]

Apply the chain rule to the left hand side and integrate by parts (using the chain rule in the form given on p. 220 of Ref. 28),

\[
\int_{\Omega} \frac{\partial \Phi(f(\mathbf{x}))}{\partial t} d\mathbf{x} = -\xi \int_{\Omega} (\nabla \mu(\mathbf{x})) \cdot (f(\mathbf{x}) \nabla \mu(\mathbf{x})) d\mathbf{x}
\]

\[= -\xi \int_{\Omega} f(\mathbf{x}) |\nabla \mu(\mathbf{x})|^2 d\mathbf{x}. \quad (16)
\]

Equation (16) is the transport dissipation equal to both the
conservative and dissipative terms of Eq. (7).
\[
\frac{d}{dt} \int_{\Omega} \Phi(f) d\vec{x} = -\xi \int_{\Omega} f \nabla \mu \cdot d\vec{x}
\]
= Transport Dissipation. \hspace{1cm} (17)

This equation is a special form of Eq. (1), see Ref. 161. An explicit treatment of this atomic scale model follows to give integro-differential equations for a PNP-like system of hard spheres [see Eqs. (30)–(32)].

C. Theoretical model: Primitive model as a complex fluid

We now treat the entire ionic solution in the spirit of the primitive model but as a complex composite fluid. One component is a macroscopic fluid phase—a purely macroscopic version of the primitive model of ionic solutions. Another component is a plasma, an atomic scale version of the primitive model, in which ions are represented on an atomic scale as charged Lennard-Jones spheres in a frictional dielectric.

\[
E(\text{Primitive Phase}; t) = \int_{\Omega} \left( \frac{1}{2} \rho_{IP} \left\| \vec{u}_{IP} \right\|^2 + \frac{1}{2} \mu \left\| \nabla \phi \right\|^2 + k_{B} T \left( c_n \log c_n + c_p \log c_p \right) + \psi \left( \text{Solid Spheres} \right) \right) d\vec{x}
\]

(18)

where \( \rho_{IP} \) is the mass density, \( \frac{1}{2} \rho_{IP} \left\| \vec{u}_{IP} \right\|^2 \) is the hydrodynamic kinetic energy of the Ionic (primitive) Phase, \( \mu(\rho_{IP}) \) is the hydrodynamic potential energy, \( \xi \) is the dielectric constant (dimensionless), and \( \lambda \) is the coupling constant (coupling scales and physical processes, as we shall see). It is the ratio of hydrodynamic (macroscopic) energy [see Eq. (3)] to microscopic (atomic) energy [see Eq. (8)] and has the role of a Lagrange multiplier. In general, \( \lambda \) must be determined by specific measurements in experiments, as has been done in rheology. \(^{10,164} \) In special cases, \( \lambda \) turns out to be a specific physical meaning, e.g., as a surface tension in the theory of liquid interfaces. \(^{11,164} \) Extensive discussion of the hydrodynamic term of Eq. (18) is found in Refs. 3 and 165. The solid sphere term was first used by Ref. 9, as far as we know.

a. Variational analysis on the macroscopic (fluid dynamics) and atomic (microscopic) scales. Combining the macroscopic and atomic energies as we have here, using just one coordinate \( \vec{x} \), is a drastic simplification not used in more complete and sophisticated multiscale analyses, \(^{3,6} \) see, for example, the “micro-macro” variational analysis \(^{32} \) where the transformation of scales is done explicitly, with two different coordinates, one, say \( \vec{y}_1 \) for the macroscopic scale, and the other, say \( \vec{y}_2 \) for the atomic scale. In the full micro-macro expression that then takes the place of our Eq. (18) an extra nested integral \( \int f \cdots d\vec{y}_2 \) would appear, because an integration must be done over the atomic scale \( \vec{y}_2 \) before the energies of the atomic (microscopic) and macroscopic (hydrodynamic) scales can be combined. In the present paper we identify both the macroscopic and atomic coordinates \( \vec{y}_1 \) and \( \vec{y}_2 \) as the mesoscopic variable \( \vec{x} \). Treatments of solution flow (i.e., solvent plus solute as needed in the study of fluid flow in the kidney, for example \(^{118} \)), or coupled transport of ions in channels (as needed in the study of active transport by protein pumps \(^{118,129} \)), may require a more complete micro/macro
analysis that does this integration explicitly. Our use of the density functional expression for the energy of uncharged hard spheres in Appendix C deals partially with the multiscale problem because it uses nonlocal interactions. Actually performing the integrations over both \( y_1 \) and \( y_2 \) is likely to do even better.

Equation (18) states the physics of our EnVarA analysis of ions in solutions and channels. In the present analysis, the hydrodynamic scale contributes only energy; no entropy terms are involved. The atomic scale, however, contributes both energy and entropy terms. The latter entropy term reflects the thermal fluctuation and the particle Brownian motion. The entropy term in Eq. (18) is on the macroscopic scale \( \bar{x} \) and represents a crude averaging of the energetic consequences of Brownian motion (that occurs on an atomic scale \( y_2 \) not shown here). A full micro-macro treatment would produce much more complete (but complex) results by explicitly averaging of the Brownian trajectories (e.g., as attempted in Refs. 59, 157, and 166). A full micro-macro treatment can embed MC simulations—perhaps even of ions in channels—in the variational integral itself, i.e., as part of the integrand in Eq. (18). The variational approach was used previously in Refs. 167 and 168. A reduced variational approach that inspired ours is used in Ref. 169. The underlying mathematics has been summarized. Simplified approaches using a single coordinate like our Eq. (18) have been published. Full micro-macro analyses are available for polymeric fluids,6,9,16,167 and liquid crystals172 although they do not use the dissipation principle Eq. (7) or EnVarA.

We include an adjustable parameter, the coupling constant \( \lambda \), in our expression (18). \( \lambda \) must be determined from an explicit model describing the (probably atomic scale) origin of the energetics and dissipation.

b. Double counting. The variational approach is helpful here because it prevents us from counting the components of Eq. (18) twice. The mathematics of the variational process used here guarantees that any solution of Eq. (18) will have the minimal values of both dissipation and action, even if the same physical process appears in two components of the integrand. However, there is no magic in the variational method. The variational approach would not prevent other quantities—beyond the least action and dissipation described by Eq. (18)—from being mishandled. Variational approaches only constrain variables that they vary [like \( \Phi \) in Eq. (18)] and use to determine the resulting partial differential equations.

The variational approach is also a natural (albeit approximate) way to combine descriptions of physical phenomena, including those occurring on different scales. The variational approach combines (the variations of the) energy and the variations of the dissipation on both scales. The resulting equations are different from those produced by taking partial differential equation describing each phenomenon and combining them directly.

Combining partial differential equations directly can be problematic. It may not be clear which partial differential equations (or variables) should be connected, and whether they should be added or otherwise combined. If different scales are merged as in Eq. (18), the variables in the different partial differential equations may not be comparable or even have the same units (e.g., the concentrations of species and the distribution function of locations of the atoms of that species59,157,166). Merging differential equations may not be unique and may even violate overarching constraints.1,7,30,32,108,169,171,173 thermodynamic principles, or sum rules, for example. In the variational procedure, the energies and dissipations are usually easily defined. Adding energies is an obvious way to try to combine scales, as is adding dissipations, although more elaborate treatments of dissipation are usually needed, see Eq. (39). Continuum treatments of energy have even been combined with estimations from discrete simulations.32

c. Variational procedure as an optimization: Coupling constant. The variational procedure is a kind of optimal control that produces optimal mixing of the components of the generalized energy function like that shown in Eq. (8) or Eq. (18) and a good starting guess for mixing atomic and macroscopic scales. Equation (38) can be viewed as an optimal control21,22,76,78 as written, without change. The cost function is the macroscopic (hydrodynamic) part of energy. The constraint function is the atomic (microscopic) part of energy, the part multiplied by \( \lambda \). If we want to strictly enforce (control of energy by) the atomic scale, then \( c_p \) becomes a Lagrange multiplier. If we relax the constraint on the atomic scale, then \( \lambda \) can be viewed as the relaxation parameter where the magnitude of \( \lambda \) represents the tolerance allowed for the constraint. Determining the tolerance of the constraint is a major topic for experimental investigation, and cannot be decided by mathematical arguments alone. In our applications to channels, it is clear that focus should be on atomic/mesoscopic scales, and Eq. (18) is written that way. Other formulations of “penalty functions” are possible besides those shown in Eq. (18) and represent different ways of handling different scales of hydrodynamic (macroscopic) and atomic motions. The coupling constant \( \lambda \) could be applied the other way around, in which case the roles of the energies of the two scales are switched. In general, determining the tolerance and penalty functions involves problems of sensitivity, ill-posedness, and overdetermination as do most inverse problems.55,106,175

d. Electrochemical potentials. To apply these ideas to specific problems, we introduce the customary chemical variables, the electrochemical potential \( \mu_i \) (of species \( i \)) often described in channel biology as the “driving force” (for the current of species \( i \)).166 The (electro)chemical potentials of the ions can also be determined as variations, see Eq. (10),

\[
\mu_n(\bar{x}) = \frac{\delta E(f(\bar{x}))}{\delta \phi_n}; \quad \mu_p(\bar{x}) = \frac{\delta E(f(\bar{x}))}{\delta c_p}.
\]

A treatment10,167 without excluded volume gives the classical drift diffusion42–44,53 (partial differential) equations of an ideal gas of point particles176 named the PNP equations by Ref. 45; earlier references to Nerst–Planck and drift diffusion equations are in Refs. 43, 44, 46, 48, and 62. The classical PNP (drift diffusion) equations are the partial differential equations produced from Eq. (18) by the usual variational procedure with respect to \( c_i \), with
\[ \psi(\text{Solid Spheres}) = 0. \]

To illustrate these ideas, we write equations for monovalent salts like NaCl. However, all programming has been done for mixtures of any number of species, with arbitrary valence. They include the permanent charge of the protein (or doping charge of semiconductors, if a variational treatment is made of semiconductors),

\[
\frac{\partial c_n}{\partial t} = \nabla \cdot \left\{ D_n \left( \nabla c_n - \frac{e}{k_B T} c_n \frac{\nabla \phi}{c_n} \right) \right\},
\]

\[
\frac{\partial c_p}{\partial t} = \nabla \cdot \left\{ D_p \left( \nabla c_p + \frac{e}{k_B T} c_p \frac{\nabla \phi}{c_p} \right) \right\},
\]

(Note 20)

\[
e \nabla^2 \phi = -e c_n + e c_p,
\]

\[
\tilde{J}_n = -\xi_n e c_n \nabla \mu_n = -\frac{D_n}{k_B T} c_n \nabla \mu_n;
\]

\[
\tilde{J}_p = -\xi_p e c_p \nabla \mu_p = -\frac{D_p}{k_B T} c_p \nabla \mu_p,
\]

\[
\tilde{I} = e \tilde{J}_p - e \tilde{J}_n + \text{Displacement Current, if present.}
\]

where \( \xi_i = D_i / k_B T \), \( i = n \) or \( p \), and \( t \) is renormalized time. Diffusion coefficients are \( D_i \), \( i = n \) or \( p \) that must not be set equal lest singular simplifications be produced like those that minimize liquid junction potentials in salt bridges when KCl is the sole electrolyte. The chemical forces \( \nabla \mu_i \), \( i = n \) or \( p \) are written in detail in Appendix B.

The flows measured in most experiments differ from the fluxes naturally defined in PNP. Electric current \( \tilde{I} \) is the variable usually controlled or measured in experiments—not flux—and this differs significantly if the measurements include significant displacement (i.e., “capacitive” \( \partial \phi / \partial t \)) current. On the picosecond time scale of Brownian motion, or the (sub) femtosecond time scale of molecular dynamics, the displacement current can be very large indeed. (Roughly speaking, the displacement current and ionic current of a biological solution are equal on a time scale of 10 ps: 1 M NaCl, p. 196 of Ref. 83). The displacement current can be precisely evaluated by the Shockley–Ramo theorem and is a consequence of the continuity of current in the Maxwell equations, if current is suitably defined. In the PNP framework such displacement current may be treated with boundary conditions that describe a specific experimental setup and its stray capacitances, particularly those that shunt the channel and those that couple the baths to ground.

The importance of computing the potential from the charge \( c_n - c_p \) as in Eq. (20) cannot be overstated. Forcing a potential to adopt a value independent of the charge \( c_n - c_p \) requires the injection of energy and charge.

That injection is so likely to substantially perturb and distort the system that it might be called the Dirichlet disaster, if hyperbole is permitted.

A protein cannot be described as a surface of fixed potential—as a Dirichlet boundary condition—or as a rate constant independent of concentrations in the bath if the protein is an isolated system that has no energy source. In particular, a channel protein cannot inject charge into a system. Channels are (chemically) passive devices. They are biological valves, not motors, and do not use the energy of hydrolysis of ATP to move ions. They modulate movements of ions driven by the gradients of electrochemical potential of the ions. The gradients are maintained by other systems, called pumps or active transporters, that do use chemical energy.

\( e \) Solid spheres: Finite size ions. The PNP equations ignore the important effects of the finite size of ions that are thought to determine the nonideal properties of ionic solutions, more than anything else. We could include these in our variational analysis in three different ways: (1) on the macroscopic (hydrodynamic) scale as an equation of state; (2) on the atomic (microscopic) scale, we include a Lennard-Jones term; and (3) also on the atomic (microscopic) scale, we could include a term (for uncharged spheres) from density functional theory (of liquids), in particular the uncharged terms from Refs. 58, 67, and 68 following Refs. In Sec. V, Fig. 1, we compare Lennard-Jones and density functional descriptions. Numerical difficulties prevented us from implementing an equation of state description.

\( f \) Lennard-Jones treatment of solid spheres. The excluded volume of solid spheres can be treated by including the (generalized) energy of an excluded volume term at the atomic scale, with the energy term written as that of Lennard-Jones purely repulsive spheres,

\[
E(\text{Solid Spheres}; t) = \frac{1}{2} \int_{\Omega} \int_{\Omega} \chi_{n,n}(|\vec{x} - \vec{y}|) c_n(\vec{x}) c_n(\vec{y}) d\vec{x} d\vec{y}
\]

\[
+ \frac{1}{2} \int_{\Omega} \int_{\Omega} \chi_{n,p}(|\vec{x} - \vec{y}|) c_n(\vec{x}) c_p(\vec{y}) d\vec{x} d\vec{y}
\]

\[
+ \frac{1}{2} \int_{\Omega} \int_{\Omega} \chi_{p,p}(|\vec{x} - \vec{y}|) c_p(\vec{x}) c_p(\vec{y}) d\vec{x} d\vec{y},
\]

where the repulsion between two balls situated at \( \vec{x}, \vec{y} \) with radius \( a_n, a_p \), respectively, is given by a Lennard-Jones type formula,

\[
\chi_{i,j}(|\vec{x} - \vec{y}|) = e_{i,j} \left( \frac{a_i + a_j}{|\vec{x} - \vec{y}|} \right)^{12} \text{ for } i = n,p,
\]

where \( e_{i,j} \) is a chosen energy coupling constant, not the
dielectric coefficient. Obviously, an attractive term could be added into Eq. (25) if needed. We proceed in the spirit of the discussion of Eqs. (9)–(13). We can derive the drift force by variation of Eq. (24) with respect to \( \mathbf{x} \). This variation determines the components of the flux due to the finite size effect of \( c_n \) [see Eq. (26)] and the finite size effect of \( c_p \) [see Eq. (27)]. Details are in Appendixes A and B. The components of flux are

Component of flux of \( n \)

\[
\begin{align*}
\Delta (\text{Primitive Phase}; t) &= \int \frac{M|\nabla \mathbf{u}_{IP}|^2}{\text{Viscosity}} d\mathbf{x} + \int \lambda \left( \frac{D_n}{k_B T} c_n |\nabla \mu_n|^2 + \frac{D_p}{k_B T} c_p |\nabla \mu_p|^2 \right) d\mathbf{x}.
\end{align*}
\]

\( \text{Component of flux of } p \)

\[
\begin{align*}
 &= c_p(\mathbf{x}) \int_\Omega \frac{12\varepsilon_{s,n}(a_p + a_n)^{12}(\mathbf{x} - \mathbf{y})}{|\mathbf{x} - \mathbf{y}|^{14}} c_n(\mathbf{y}) d\mathbf{y}
 &+ c_p(\mathbf{x}) \int_\Omega \frac{6\varepsilon_{s,n}(a_p + a_n)^{12}(\mathbf{x} - \mathbf{y})}{|\mathbf{x} - \mathbf{y}|^{14}} c_p(\mathbf{y}) d\mathbf{y},
\end{align*}
\]

These components of flux would appear inside the divergence operator in the Fokker–Planck Eq. (13) and as part of the drift term in Eq. (14).

g. Dissipation on the atomic scale. We turn next to the dissipation in the primitive atomic scale phase so we can take its variation with respect to velocity and thus determine the dissipative force in this application of Eq. (7). The dissipation of the primitive atomic scale phase is
described in Appendixes B. Explicit formulas for \( \mu_n \) and \( \mu_p \) are not possible because they are nonlocal, involving the electrical potential and finite volume effects throughout the global system.

Macroscopic conservation of mass is

\[
\frac{\partial \rho_{IP}}{\partial t} + \nabla \cdot (\mathbf{u}_{IP} \rho_{IP}) = 0.
\]

Macroscopic force balance (conservation of linear momentum) is

\[
\rho_{IP} \left( \frac{\partial \mathbf{u}_{IP}}{\partial t} + \mathbf{u}_{IP} \cdot \nabla \mathbf{u}_{IP} \right) + \nabla \mathbf{p}_{IP} = M \nabla^2 \mathbf{u}_{IP} + \left( c_n(\mathbf{x}) - c_p(\mathbf{x}) \right) \nabla \phi
\]

\[
- c_n(\mathbf{x}) \nabla \cdot \left( \chi_{n,n}(\mathbf{x} - \mathbf{y}) c_n(\mathbf{y}) + \frac{1}{2} \chi_{n,p}(\mathbf{x} - \mathbf{y}) c_p(\mathbf{y}) d\mathbf{y} \right)
\]

\[
- c_p(\mathbf{x}) \nabla \cdot \left( \frac{1}{2} \chi_{p,n}(\mathbf{x} - \mathbf{y}) c_n(\mathbf{y}) + \chi_{p,p}(\mathbf{x} - \mathbf{y}) c_p(\mathbf{y}) \right)
\]

\times d\mathbf{y}.
\]
macroscopic ionic phase (fluid). The second and third lines are the body forces due to the finite size effect. Notice that the concentration variables \( c_n(x) \) and \( c_p(x) \) in the body force terms cannot be moved inside the divergence. Their location implies that the divergence theorem (i.e., Green–Gauss formula) cannot put those concentration terms on the boundary. Thus these forces must be evaluated inside the bulk of the system and are given the name body force.

a. PNP for solid spheres. Next we write time dependent PNP equations modified to include finite size effects,

\[
\frac{\partial c_n}{\partial t} + \nabla \cdot (c_n \vec{u}_{IP}) = \nabla \cdot \left[ D_n \left( \nabla c_n - \frac{e}{k_B T} c_n \nabla \phi \right) \right] - \nabla \cdot \left[ c_n(x) \left( \int \frac{12 e_n \rho_n (a_n + a_p)^{12} (\vec{x} - \vec{y}) |\vec{x} - \vec{y}|^{14}}{|\vec{x} - \vec{y}|^{14}} \right) \right] + \int \frac{6 e_n \rho_n (a_n + a_p)^{12} (\vec{x} - \vec{y}) |\vec{x} - \vec{y}|^{14}}{|\vec{x} - \vec{y}|^{14}} c_n(y) dy \right] \right). 
\]

(31)

\[
\frac{\partial c_p}{\partial t} + \nabla \cdot (c_p \vec{u}_{IP}) = \nabla \cdot \left[ D_p \left( \nabla c_p + \frac{e}{k_B T} c_p \nabla \phi \right) \right] - \nabla \cdot \left[ c_p(x) \left( \int \frac{12 e_p \rho_p (a_p + a_p)^{12} (\vec{x} - \vec{y}) |\vec{x} - \vec{y}|^{14}}{|\vec{x} - \vec{y}|^{14}} \right) \right] + \int \frac{6 e_p \rho_p (a_n + a_p)^{12} (\vec{x} - \vec{y}) |\vec{x} - \vec{y}|^{14}}{|\vec{x} - \vec{y}|^{14}} c_p(y) dy \right] \right]. 
\]

(32)

The second terms on the left hand side of Eqs. (31) and (32) represent the transport of the ions by a compressible fluid which is appropriate if the equations are applied only on the atomic scale. If we try to extend these equations to other scales, the proper form of the compressibility becomes an issue that needs to be resolved by experiment. These and the last two terms of Eq. (30) represent the balance of the internal forces (Newton’s third law).

It is important to verify Newton’s third law explicitly in problems of this sort. We deal with the water (solvent on the macroscopic scale), the ionic phase (fluid on the macroscopic scale), and the ionic particles (on an atomic scale). Newton’s third law has to be satisfied for each phase and all interactions among the phases.

3. Theoretical model: Solvent water

We turn now from the ionic phase to the solvent water. The solvent is treated traditionally as an incompressible fluid density \( \rho_s \) and velocity \( \vec{u}_s \) (although treatment as a compressible fluid is possible if needed) using (generalized) energy and dissipation,

\[
E(\text{Incompressible Solvent}) = \int_{\Omega} \frac{1}{2} \rho_s |\vec{u}_s|^2 d\vec{x}. 
\]

(33)

\[
\Delta(\text{Incompressible Solvent}) = \int_{\Omega} M_f |\vec{u}_s|^2 d\vec{x}. 
\]

(34)

Here, \( \nabla \cdot \rho_s = 0 \) for an incompressible fluid. Applying the force balance law “Conservative Force = Dissipative Force” (7) to the equations for the incompressible solvent and Navier–Stokes partial differential equations for an incompressible fluid with dynamic viscosity \( M_f \),

\[
\frac{\partial \rho_s}{\partial t} + \vec{u}_s \cdot \nabla \rho_s = 0, 
\]

(35)

\[
\nabla \cdot \vec{u}_s = 0, 
\]

(36)

\[
\rho_s \frac{\partial \vec{u}_s}{\partial t} + \rho_s \vec{u}_s \cdot \nabla \vec{u}_s + \nabla p_f = M_f \nabla^2 \vec{u}_s. 
\]

(37)

4. Theoretical model: (Entire) Primitive solution

Finally, we can deal with the entire electrolyte, namely, the ionic solution consisting of the solvent and the primitive phase of ions by combining the (generalized) energy and the dissipation of the individual components using the simplest model for the interaction of the components. Later work may need to deal more carefully with the different scales of the components.

The (generalized) energy of the solution is simply the sum of the (generalized) energy of the components, namely, the sum of the energy of the ions in primitive phases (both atomic scale and macroscopic) and of the energy of the incompressible solvent Eq. (34). We do not write it out. We also will not bother to write “(generalized) energy” in every case from now on. It should be clear that “energy” in this paper is not just that defined in classical treatments of the first law of (equilibrium) thermodynamics.

The dissipation of the primitive solution is not just the sum of the dissipation of the components because the solvent can drag the ions and vice versa. Thus, the dissipation of the primitive solution is

\[
\Delta(\text{Primitive Solution}) = \int_{\Omega} \left[ \frac{M_f |\vec{u}_s|^2}{2} + M_{ip} |\vec{u}_{IP}|^2 + \frac{k}{k_B T} c_n |\nabla \mu_n|^2 + \frac{k}{k_B T} c_p |\nabla \mu_p|^2 \right] d\vec{x}, 
\]

(38)

where \( M_f \) is the dynamic viscosity and \( \vec{u}_s \) of the solvent fluid, \( M_{ip} \) is the viscosity, and \( \vec{u}_{IP} \) is the velocity of the Ionic Phase. The last term in Eq. (38) gives rise to an extra drag term \( k (\vec{u}_{IP} - \vec{u}_s) \) that will have to be added into Eq. (37) and also an extra term \(-k (\vec{u}_{IP} - \vec{u}_s)\) that will have to be added into Eq. (30). These two terms again reflect the balance of internal forces enforced by Newton’s third law. The frictional
drag between solvent and ions is described to the lowest order approximation as a Stokes’ drag,

\[ k = \gamma p \text{ or } \gamma(c_n + c_p), \]

where the Stokes drag is \( \gamma = 6\pi M_d u_f \),

(39)

where \( d_f \) is a generic description of the radius of the solvent particle. It is not clear \textit{a priori} how much detail will be needed in describing the drag of the solvent on the ions and the ions on the solvent. This will be determined by solving the problem for specific cases of flow in mixed bulk solutions,\textsuperscript{34,70,90,92,93,100–105} or in ion channels, and seeing whether expressions for drag between water that are \textit{not} specific for individual ions produce correlations similar to those observed experimentally\textsuperscript{188,189} and traditionally attributed\textsuperscript{188,190} to “single filing” (however, see Refs. 114 and 129). Perhaps, specific coefficients will need to be introduced into an \textit{EnVarA} field theory of ionic solutions—as they have been in traditional theories of flux coupling in bulk and ion channels—to describe drag between one type of ion and other types of ions and water, with all the uncertainty that involves \textit{e.g.}, how do the specific coefficients vary with concentration(s) in pure and mixed solutions?\textsuperscript{34,90,92,100–102,104,143}

The total coupled system—involving solvent and macroscopic and atomic scale components of the entire solution—is given below. All the equations need to be solved together, in a simultaneous solution. Note the two physical sources of coupling between the solvent (water) and ion (primitive) phases. The drag term couples these phases dynamically, when there is relative movement. The phases are also coupled at equilibrium, when there is no motion, by Newton’s third law, and this coupling will produce at least some of the complexities normally dealt with “by hand” by the parameters of the equations of state.\textsuperscript{79,81,162} The effect of the ions on the fluid is balanced by the effects of the fluid on the ions.

\textbf{a. Complete ionic (primitive) solution.} \textbf{Solvent water phase} treated as incompressible.

\[ \frac{\partial \rho_f}{\partial t} + \mathbf{u}_f \cdot \nabla \rho_f = 0, \]  

(40)

\[ \nabla \cdot \mathbf{u}_f = 0, \]  

(41)

\[ \rho \frac{\partial \mathbf{u}}{\partial t} + \rho \mathbf{u} \cdot \nabla \mathbf{u} + \nabla p_f = M \nabla^2 \mathbf{u} + k(\mathbf{u}_{IP} - \mathbf{u}_f). \]  

(42)

\textbf{Primitive ionic phases} are macroscopic and atomic scale combined.

\[ \frac{\partial \rho_{IP}}{\partial t} + \nabla \cdot (\mathbf{u}_{IP} \rho_{IP}) = 0, \]  

(43)

\[ \rho_{IP} \left( \frac{\partial \mathbf{u}_{IP}}{\partial t} + \mathbf{u}_{IP} \cdot \nabla \mathbf{u}_{IP} \right) + \nabla \mathbf{p}_{IP} \]

\[ = M \nabla^2 \mathbf{u}_{IP} + k(\mathbf{u}_{IP} - \mathbf{u}_f) + \left( c_n(x) - c_p(x) \right) \nabla \phi \]

\[ - c_n(x) \nabla \cdot \int \sum \chi_{n,n}(\mathbf{x} - \mathbf{y}) c_n(\mathbf{y}) \mathbf{y} + \frac{1}{2} \chi_{n,p}(\mathbf{x} - \mathbf{y}) c_p(\mathbf{y}) d\mathbf{y} \]

\[ - c_p(x) \nabla \cdot \int \sum \chi_{p,n}(\mathbf{x} - \mathbf{y}) c_n(\mathbf{y}) + \chi_{p,p}(\mathbf{x} - \mathbf{y}) c_p(\mathbf{y}) \times d\mathbf{y}. \]  

(44)

\textbf{D. Theoretical model: Generality of the energy variational approach}

An energy variational treatment allows more generality and (possible) complexity than is usual in theories of ionic solution because (1) it includes all the bulk hydrodynamic behavior described by the Navier–Stokes equations; (2) it includes all the bulk hydrodynamic behavior of a compressible phase of ions; (3) it includes the atomic scale behavior of a PNP system including finite size ions that create their own electric field; (4) it allows boundary conditions that can drive flow, for example, when they are spatially nonuniform; and (5) it automatically computes interactions between all components and scales included in the models that describe dissipation and energy. In most models of ionic solutions,\textsuperscript{35,80,81,84,90,92,100–102,162,191} these interactions have to be put in by hand, sometimes with hundreds of parameters.\textsuperscript{79}

The generality of an energetic variational approach—which does not even distinguish between thermodynamic equilibrium and thermodynamic nonequilibrium—is a major potential advantage. For example, energy variational treatments will automatically reveal correlations in the flows of any of these components across all scales, even if pairwise interactions (like force laws of molecular dynamics) are not explicitly included in the energy function. The electrical potential is present on all scales and directly couples atomic and hydrodynamic bulk behavior in the resulting Euler–Lagrange field equations. The energy variational method also deals with double counting better than most \textit{[see discussion of double counting after Eq. (18)]}. It allows (in the future) combination of equations of state\textsuperscript{79,80,81,162,163} and (for example) our models of excluded volume—Lennard-Jones and density functional—each weighted with separate coupling constants, Lagrange multipliers, if we choose to handle them that way. One can choose coupling constants to fit data optimally using methods of inverse problems, if necessary to deal with issues of sensitivity and ill-posedness\textsuperscript{106,175} as we have in fitting PNP-DFT to properties of channels.\textsuperscript{55}
IV. COMPUTATIONAL METHODS

A. Methods of numerical computations

Energetic variational methods produce integro-differential field equations that describe a wide range of systems under many conditions and thus with qualitatively different behaviors. Numerical procedures need to be tuned to the qualitative behavior of the system to be reasonably efficient. EnVarA requires numerical solutions of time dependent equations (because the real world is always time dependent). Steady state phenomena emerge as (hopefully) stable limits of time dependent phenomena, as they often do in the real world, so efficiency and stability are particularly important.

Computation of phenomena of molecular biology poses a stiff computational challenge. Biological phenomena are almost always slow (>10^{-3} s), but are usually controlled by a handful of key atoms in a few molecules, here channel proteins. Ion channels are nanowires. Atomic structures of ion channels control macroscopic functions on macroscopic time scales. Displacements of 10^{11} m in the (time) averaged location of the key atoms of channel proteins have large effects on biological selectivity.15,17,63,112,102 The atoms move more or less at the speed of sound103 and thus time scales from 10^{-18} to 10^1 s are directly involved in the behavior of ion channels.

We use finite element methods that reflect and take advantage of the underlying energetic variational structure of ion channel dynamics building on earlier work,30,171,194 particularly that of Ryham,167,168 working with Liu that described the coupling of ions (PNP equations) and flow (Navier–Stokes equations) using a “mini” finite element to solve the (Navier)–Stokes dynamics. We use generalized mini elements to solve our model,195 namely, the standard elements P_k−P_{k+1}, k=2,3,..., P_k are a set polynomials up to order k. We solve the drift-diffusion equation (Nernst–Planck) Eq. (20) with an efficient finite element method: edge averaged finite elements (EAFE's) that have been proposed196 and studied extensively.167,168 The method exploits the type of monotonicity of the operators in the equations. The Euler method is used to deal with time dependence. Ionic solutions are confined by insulating boundaries in experiments (and in channels). Insulating boundaries are described by no-flux boundary conditions derived by variational procedures. We use the following (pseudo) algorithm based on finite element discretization to solve the coupled Poisson–Nernst–Planck equations that include the effects of the finite volume of ions, for example, Eqs. (31) and (32):

- Step 1. Set initial data \( c_n^{(k)} = c_p^{(k)} = \phi^{(0)} \) and k=1.
- Step 2. Set \( c_n^{(k-1)} = c_p^{(k-1)} = \phi^{(k-1)} \) for k=1,2,...
- Step 3. Solve the following finite dimensional equation for \( c_n^{(k)}, c_p^{(k)} \) with given \( c_n^{(k-1)}, c_p^{(k-1)} = \phi^{(k-1)} \) using EAFE.167,168

\[
\frac{c_n^{(k)} - c_n^{(k-1)}}{dt} = \nabla \cdot D_n \left( \nabla c_n^{(k)} + \frac{z_c F}{k_B T} c_n^{(k)} \frac{\nabla \phi^{(k-1)}}{\nabla c_p^{(k)}} \right) + \nabla \phi_p \left( c_n^{(k-1)}, c_p^{(k-1)} \right),
\]

\[
\frac{c_p^{(k)} - c_p^{(k-1)}}{dt} = \nabla \cdot D_p \left( \nabla c_p^{(k)} + \frac{z_p F}{k_B T} c_p^{(k)} \frac{\nabla \phi^{(k-1)}}{\nabla c_p^{(k)}} \right) + \nabla \phi_p \left( c_n^{(k-1)}, c_p^{(k-1)} \right),
\]

where \( \phi_p \left( c_n^{(k-1)}, c_p^{(k-1)} \right), j=n,p \) are the chemical potential obtained from the finite volume energy and dt is the time step. The equation is written for monovalent ions n and cations p but programming was done for any ions of any charge.

- Step 4. Solve Poisson equation for given \( \phi^{(k-1)} \) with \( c_n^{(k)}, c_p^{(k)} \):

\[
\varepsilon \nabla^2 \phi^{(k-1)} = c_n^{(k)} - c_p^{(k)}.
\]

To prevent oscillatory behavior in this iteration we use a convex iteration scheme that has evolved167,168 from earlier work related to the Gummel iteration180,197 of semiconductor physics,144,198

\[
\phi^{(k)} = c \phi^{(k-1)} + (1-c) \phi^{(k-1)}, \quad 0 < c \leq 1.
\]

- Step 5. Check self-consistency between \( c_n^{(k)}, c_p^{(k)}, \phi^{(k)} \). If consistent, then \( c_n^{(k)} = c_n^{(k-1)}, c_p^{(k)} = c_p^{(k-1)}, \phi^{(k)} = \phi^{(k-1)} \). Otherwise \( \phi^{(k-1)} = \phi^{(k-1)} \) and go to step 3.

- Step 6. Check the error between \( c_n^{(k)}, c_p^{(k)}, \phi^{(k)} \) and \( c_n^{(k-1)}, c_p^{(k-1)}, \phi^{(k-1)} \) with a criterion. If the error is less than a criterion, then print solution \( c_n^{(k)}, c_p^{(k)}, \phi^{(k)} \) and exit. Otherwise, set \( k=k+1 \) and go to step 2.

The numerical scheme for the PNP system has been verified by comparison to theoretical results, and in special cases to known solutions of the Poisson–Boltzmann and renormalized Poisson–Boltzmann equations. We verify by inspection that the numerical scheme has enough resolution to catch the boundary layer behaviors of the electrostatic potential.

V. RESULTS

We implement EnVarA here in a few special cases to show its feasibility. EnVarA yields a time dependent system of Euler–Lagrange equations, even if the properties of interest are stationary, developing after a long time. This is a blessing and a curse. It is a blessing because we learn much more of the system, and can understand or propose experiments in the time domain. It is a curse because the computations of time dependent phenomena produce complex phenomena not emphasized in classical experimental papers that often focus on simpler behavior seen in special steady state conditions.

Experiments are often designed to focus on particular parts of complex phenomena. The conditions that allow that
focus often take many years to discover (consider, for example, sequence of papers needed to discover ionic conductance using the voltage clamp\cite{184}) and the preliminary survey experiments used to design that focus are often not reported in detail. After all, survey experiments give complex results that are usually quite confusing compared to focused experiments designed to illustrate key results.

Variational calculations report all the time dependent properties of the system; they correspond to the survey experiments. So we must survey ranges of parameters before we can focus on important experimental phenomena. The values of effective parameters needed to focus on particular parts of complex phenomena are not known ahead of time, and are hard to determine, particularly in simplified models computed in only one dimension. The numerical solutions of the time dependent Euler–Lagrange equations are slow, particularly since we are usually interested in the eventual steady state. Thus, our survey calculations are incomplete, and certainly have not yet isolated phenomena as clearly as experiments do (using protocols that often have taken decades to work out, we say in our defense).

A. Layering near a charged wall

We first calculate the property of “one dimensional spheres” near a highly charged wall in the presence of divalent and monovalent ions, e.g., hypothetical ions something like Ca$^{2+}$ and Na$^+$ Cl$. More precisely we use center to center interactions and evaluate the forces in one dimension in a long tradition starting with the statistical mechanics of uncharged spheres near walls.\textsuperscript{65} We might be tempted to call these “rods” or one dimensional spheres, but the correct specification is the mathematics. We choose this system because it shows the ability of the variational method to deal with correlations in a highly charged, highly correlated system. Obviously, this idealization should be replaced by computations of spheres in three dimensions as soon as we can do those calculations. Indeed, we cannot quantitatively compare our results with MC simulations of real spheres\textsuperscript{132} until we work in three dimensions.

This reduced test case provides a wealth of complex behavior of great importance—judging from the hundreds of papers devoted to it—in a wide range of applications, from electrochemistry, to biophysics, to material science, where interactions of this type are important in determining the strength of cement\textsuperscript{199} (i.e., calcium-silicate-hydrate). The literature of this field is well reviewed.\textsuperscript{132,199–201} We have particularly used Ref. 200 for an overview of the entire field and Ref. 132 to show the wide variety of behaviors of such systems in MC simulations of the primitive model.

Figure 1 shows the spatial distribution of concentration (really, the number density in molar units) near a highly charged well comparable to those studied in the literature. The wall has charge density of 0.1 C/m$^2$. Charge is shown divided by 1 C/m$^3$. The diameter of the ions is 0.1 nm and they have charges +2e and $-1e$, where $e$ is the charge on a proton. Position is shown divided by nanometer. Dielectric coefficient was 78, temperature was 298 K, and the bulk concentration of ions was 1 M of the divalent cation and 2 M of the monovalent anion. The potential on the wall was set to $-3.1$ kT/e, i.e., $-80$ mV in accord with MC simulations (Ref. 202). Energy coupling coefficients $\lambda$ in EnV arA were 0.5. The dotted circles show the size of the ions in the calculation. Ions are not allowed to overlap with the wall and so the densities are smooth functions until they reach the excluded zone produced by finite diameter of the ions. Calculations were done using (solid lines) the PNP-DFT and PNP-LJ in the form described in the text and Appendixes A and C. The form of the DFT differs in detail (but not spirit) from that in recent literature (Refs. 67, 68, and 203): we use DFT for uncharged interactions (following Refs. 66, 142, and 204) but we use EnV arA to deal with the electrostatics. EnV arA identically satisfies Gauss’ law (which is also one of the sum rules (Ref. 72 of statistical mechanics). This calculation is called PNP-DFT even though we report only equilibrium results: the variational method knows nothing of equilibrium and always computes a nonequilibrium transient response which may in special cases have no flows and a stationary solution. The results are qualitatively similar to the layering reported in MC simulations (Ref. 132). Quantitative differences are expected because the systems are not identical. Most importantly, simulations were of hard spheres (whereas we use Lennard-Jones or DFT in one dimension). There are many other small differences; e.g., MC uses an approximation to the solution of Poisson’s equation (Ref. 205) that produces results independent of system size, but without definite error bounds (Refs. 135 and 206).

**FIG. 1.** The number density (“concentration”) of ions near a charged wall. The wall has charge density of 0.1 C/m$^2$. Charge is shown divided by 1 C/m$^3$. The diameter of the ions is 0.3 nm and they have charges of +2e and $-1e$, where $e$ is the charge on a proton. Position is shown divided by nanometer. Dielectric coefficient was 78, temperature was 298 K, and the bulk concentration of ions was 1 M of the divalent cation and 2 M of the monovalent anion. The potential on the wall was set to $-3.1$ kT/e, i.e., $-80$ mV in accord with MC simulations (Ref. 202). Energy coupling coefficients $\lambda$ in EnV arA were 0.5. The dotted circles show the size of the ions in the calculation. Ions are not allowed to overlap with the wall and so the densities are smooth functions until they reach the excluded zone produced by finite diameter of the ions. Calculations were done using (solid lines) the PNP-DFT and PNP-LJ in the form described in the text and Appendixes A and C. The form of the DFT differs in detail (but not spirit) from that in recent literature (Refs. 67, 68, and 203): we use DFT for uncharged interactions (following Refs. 66, 142, and 204) but we use EnV arA to deal with the electrostatics. EnV arA identically satisfies Gauss’ law (which is also one of the sum rules (Ref. 72 of statistical mechanics). This calculation is called PNP-DFT even though we report only equilibrium results: the variational method knows nothing of equilibrium and always computes a nonequilibrium transient response which may in special cases have no flows and a stationary solution. The results are qualitatively similar to the layering reported in MC simulations (Ref. 132). Quantitative differences are expected because the systems are not identical. Most importantly, simulations were of hard spheres (whereas we use Lennard-Jones or DFT in one dimension). There are many other small differences; e.g., MC uses an approximation to the solution of Poisson’s equation (Ref. 205) that produces results independent of system size, but without definite error bounds (Refs. 135 and 206).
cise treatment including an intermediate scale is necessary in Eqs. (24)–(27) until we can do calculations in three dimensions.

B. Binding in channels

Calculations (Fig. 2) were also done of binding in a simple model of calcium channels that has proven quite successful. In this model, the active site of the calcium channel uses spheres with permanent charge to represent the side chains DEEA (aspartate glutamate glutamate alanine) that are known (from experiments) to mix with the ions and water in a structural motif very different from potassium channels. In our calculations, spheres are uniformly distributed at fixed locations within a cylindrical space of 3 Å long and 7 Å diameter (unlike MC simulations in which the spheres are free to move within that region) to reduce computation time. This and other details in the calculations (see above) mean that the variational method is not expected to give identical results to MC simulations. The ion diameters, geometry, and so on are otherwise as specified previously. Figure 2 shows the relative occupancy of the cylindrical space, that is to say, it shows the ratio of the spatial integral of the density of calcium to the spatial integral of the density of sodium, both within a cylindrical space of 3 Å long and 7 Å diameter. The densities are the stationary solution of the time dependent Euler–Lagrange equations. The concentration of sodium is maintained at 0.1 M on both sides of the channel. The concentration of calcium is varied and is shown on the horizontal axis.

The binding computed is similar to that reported previously for the calcium channel. Figure 3 shows similar calculations for the DEKA (aspartate glutamate lysine alanine) sodium channel. Here occupancy is reported, namely, the spatial integral of the density of either sodium or calcium which is the steady solutions of the Euler–Lagrange equations. The properties are similar to those reported previously with MC simulations of the sodium channel, where the physical and biological implications are extensively discussed.

C. Time dependent phenomena in ion channels

Figure 4 shows the time dependent current calculated after a step function is applied to a DEKA channel specified in Fig. 5 and the caption to Fig. 4. The current and time scales depend on a somewhat arbitrary assignment of effective parameters. The openings at the end of the channel are specified by flared cones as in Fig. 1 of Ref. 109 and subsequent papers so the one dimensional model is not dominated by artifactual electrical or diffusive resistance in the regions outside the channel. The time dependence seen in these records reflects changes of concentration of ions just outside the two ends of the channel. Such phenomena are not thought to be involved in the currents measured from squid sodium channels, but they are present in calcium channels and potassium channels to cite only classical references.

D. Interpretation of current transients

The flow of current through real ion channels and the surrounding proteins and lipids is complex and involves many components. Those components had to be identified and separated before mechanisms could be identified, let alone studied. Indeed, the history of electrophysiology (before recordings were made on currents through single protein
molecules\textsuperscript{212} is in large measure the history of Cole\textsuperscript{213} and Hodgkin’s\textsuperscript{184} successful separation of components of current. Separation was done in preparations of animal and plant cells put in situations designed to focus on and isolate particular components of current. The Anglo-American tradition was to choose preparations (e.g., the squid axon) and experimental methods (the voltage clamp) that isolated components. Workers in that tradition (importantly joined by German colleagues\textsuperscript{212}) depended on biology and experimental design as much as possible to isolate systems and tried to depend on theory and discussion as little as possible. The particular properties of the sodium channels of squid axon allow clear separation of components of currents. These properties ensure that accumulation of ions occurs on a longer time scale than the processes that open the channel (as viewed in macroscopic ensembles of channels). Most channels do not permit such separation, however. For example, in calcium channels,\textsuperscript{210} opening processes and accumulation occur on the same time scale and are more or less inseparable. The channel opening process and accumulation occur on nearly the same scale as well, in squid potassium channels.\textsuperscript{511,214} Indeed, it seems possible that the squid sodium channels are a special case, specialized to reduce change in concentration of ions during natural activity.\textsuperscript{215} Accumulation of sodium ions would reduce inward current and limit the speed of conduction. Speed is what the squid giant axon (and squid itself) are all about from an evolutionary point of view.

Our calculations of the DEKA channel give results such as calcium and potassium channels. We do not know how to reproduce the special properties of squid sodium channels and conductance. In particular, our calculations show the pile-up of salt just outside the channel occurring much (say ten times) faster than in the squid.\textsuperscript{211} (By “salt” we mean neutral combinations of cation and anion, Na\textsuperscript{+} and Cl\textsuperscript{−}, in Figs. 4 and 5.) Our calculations also often show rapid responses to steps in potential that describe the storage (“pile-up”) of charge inside the pore of the channel protein. These rapid pile-ups of charge occur on the same time scale as gating currents\textsuperscript{216} found in real nerve cells\textsuperscript{216} including the very fast component of gating current.\textsuperscript{217} We do not show these rapid responses here because the calculations cannot easily be “corrected” for linear capacitance as are experimental measurements. Our calculated responses are not robust. Detailed comparisons with experiments must await calculations in three dimensions less dependent on assumed values of effective parameters. Thus, we do not know whether our buildup of charge might be a significant component of the gating current observed experimentally in sodium or calcium channels.

VI. DISCUSSION

Ionic solutions have many diverse but specific properties arising from the interactions of their components on all scales and so it seems appropriate to treat them here as complex, not simple fluids. The diversity of properties of ionic solutions means that it is surely pretentious to write a general theory. A general theory must deal with all specific results, with the experiments and interpretations, theories, and simulations of many communities of scientists, painstakingly measured and (often) passionately debated over nearly a century. The authors can certainly not check a wide-ranging theory: we are not even physical chemists. The literature is vast beyond grasp and the many references cited here\textsuperscript{33–38,64,68,70,81,92,99,100,102,123,137,140,142,162,204,218–220} are not
comprehensive, we fear they are not even an unbiased sample, despite our efforts. Our goal is to present enough
detail in enough fields so other workers are motivated to adapt EnVarA to their specific needs and passions.
We propose a general variational analysis here because we suspect that correlations among ions are the key to un-
derstanding ionic solutions in bulk or ions in proteins.\textsuperscript{52} These correlations are the structure of ionic solutions. They
change significantly with conditions. We believe that the structures cannot be assumed, and are difficult to model \textit{a priori} with partial differential equations, because each charge is so correlated with so many other charges on both an
atomic and global (macroscopic) scale. Simulations have dif-
ficulty computing such correlations because atomic motions
must be computed on the $10^{16}$ s time scale but biology
occurs in macroscopic systems on macroscopic time scales
and involves a biological mix of concentrations ranging from
$10^{-11}$ to $10^{1}$ m. These correlations arise naturally (and always
self-consistently) in both physical solutions and in solutions
of EnVarA. The pun on “solutions” is precise. In both the
mathematics and physics, the electrical potential is present
on all scales and directly couples atomic and hydrodynamic
bulk behavior.

The structures of ionic solutions are self-organized\textsuperscript{52} (if we use biologists’ language) and form an induced fit of ion to
ion, ionic atmosphere to ion, and ionic atmosphere and ions
to a protein and its mobile side chains. The solutions of
EnVarA are also self-organized. That is the great advantage
of the variational procedure. Of course, the solutions of En-
VarA only deal with phenomena and constraints that are
present (or are implicit) in our model or energy equations.
For example, the model of energy used here does not include
complex behavior of solvent water. The model of a channel/
transporter here does not include proteins that do work on ions.\textsuperscript{118,120}

Other kinds of mathematical analysis—for example,
analysis of combined partial differential equations in ideal-
ized domains—are likely to (inadvertently) impose unnatural
boundary conditions that constrain the charged particles in
some way or the other and can only be maintained by unnat-
urally artificially imposed flows of charge and energy. These unnatural flows produce unnatural correlations and structures in the mathematical solution that can severely dis-
tort the qualitative properties of the ionic system and so
might be called “Dirichlet disasters” if one likes colorful
language. For example, such a treatment of semiconductors
as systems with constant (electric) fields would have im-
peded or prevented the discovery of transistors,\textsuperscript{221} it seems
safe to say.

Our variational approach to ionic solutions may fail be-
cause of errors in its formulation. We hope not. The theory
may fail because it does not resolve some scales well
enough. We expect so. Proteins have movements (“confor-
mination changes”) over time scales from $10^{2}$ to $10^{-13}$ s judg-
ing from the time scales of protein function, from classical
measurements of dielectric dispersion,\textsuperscript{222} and from molecular
dynamics simulations. Our computations clearly cannot be
expected to capture such a time range using a single time
variable. We expect the theory to fail when it leaves some-
thing out altogether, particularly in biological applications,
where proteins like enzymes\textsuperscript{223} or binding proteins\textsuperscript{224} do
much more than provide a confining volume and electrostatic
environment. In that case, EnVarA may help uncover (and
then compute) the omitted pieces, correcting its own mis-
take.

Even in failure, however, we expect that EnVarA will be
useful in studying ions in solutions and channels, as variational methods are in so many areas of physics, because EnVarA yields a field theory of a chemical system, thus in-
cluding boundary conditions and flow without mathematical
approximation. EnVarA yields specific working hypotheses—partial differential equations and boundary
value problems with only a few adjustable parameters—that
include all the interactions of the components of its energies.
The equations can be tested against experiment in many ap-
lications, and then improved in a mathematically systematic
and physically self-consistent way by adding or modifying
components of the energy. EnVarA will enforce self-
consistency. Experiments will enforce reality.

\section*{APPENDIX A: LENNARD-JONES TREATMENT
OF FINITE SIZE}

The repulsion potential is given by

\begin{equation}
\varphi_{R,i} = \frac{e_{i}a_{i}^{12}}{|x-y|^{12}},
\end{equation}

where $a_{i}$ is the radius of $i$th ion, and $e_{i}$ is the energy constant
(softness) of ion $i$. Then the repulsive energy $E_{i,j}$ for ion $i$
and $j$ is defined by

\begin{equation}
E_{i,j} = \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{e_{i}a_{i}^{12}}{|x-y|^{12}} c_{i}(x)c_{j}(y) dx dy
\end{equation}
or

\begin{equation}
E = \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{e_{n,n}(a_{n} + a_{p})^{12}}{|x-y|^{12}} c_{n}(x)c_{p}(y) dx dy
+ \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{e_{n,p}(a_{n} + a_{p})^{12}}{|x-y|^{12}} c_{n}(x)c_{p}(y) dx dy
+ \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{e_{n,p}(a_{n} + a_{p})^{12}}{|x-y|^{12}} c_{n}(x)c_{p}(y) dx dy.
\end{equation}

Here, we take a variational derivative with respect to the
negative charge density $c_{n}$. Then we have following equa-

\begin{equation}
\delta E_{c_{n}} = \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{e_{n,n}(a_{n} + a_{p})^{12}}{|x-y|^{12}} \delta c_{n}(x)c_{p}(y) dx dy
+ \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{e_{n,n}(a_{n} + a_{p})^{12}}{|x-y|^{12}} c_{p}(x)\delta c_{n}(y) dx dy
+ \frac{1}{2} \int_{\Omega} \int_{\Omega} \frac{e_{n,p}(a_{n} + a_{p})^{12}}{|x-y|^{12}} \delta c_{n}(x)c_{p}(y) dx dy
= \int_{\Omega} \int_{\Omega} \frac{e_{n,p}(a_{n} + a_{p})^{12}}{|x-y|^{12}} \delta c_{n}(x)c_{p}(y) dx dy
\end{equation}
Using the familiar identity
\[ \delta \alpha \nabla \cdot (\mathbf{c}_n \alpha \mathbf{e}) = 0, \]
we take a variation with respect to \( c_n \alpha \) and write it as a variation with respect to \( \mathbf{x} \),
\[ \delta E = \delta \tilde{E} \]
\[ = \int_{\Omega} \int_{\Omega} \left( \frac{e_{n,n}(a_n + p_n)^{12}}{|x - y|^{12}} c_n(x) \delta x c_n(y) d\tilde{x} d\tilde{y} \right) \]
\[ + \int_{\Omega} \int_{\Omega} \left( \frac{e_{n,n}(a_n + p_n)^{12}}{|x - y|^{12}} c_n(x) \delta \mathbf{c}_n \mathbf{c}_n(y) d\tilde{x} d\tilde{y} \right) \]
\[ = -\int_{\Omega} \int_{\Omega} \left( \frac{12 e_{n,n}(a_n + p_n)^{12}}{|x - y|^{14}} c_n(x) c_n(y) \delta x d\tilde{y} \right) \]
\[ + \int_{\Omega} \int_{\Omega} \left( \frac{6 e_{n,n}(a_n + p_n)^{12}}{|x - y|^{14}} c_n(x) c_p(y) \delta \mathbf{c}_n \mathbf{c}_n(y) d\tilde{x} d\tilde{y} \right). \]

Also, for variation on \( \mathbf{x} \) in terms of \( c_p(x) \),
\[ \delta (c_p \mathbf{x}) + \nabla \cdot (c_p \alpha \mathbf{e}) = 0, \]
\[ \delta E = \delta \tilde{E} \]
\[ = -\int_{\Omega} \int_{\Omega} \left( \frac{12 e_{p,n}(a_n + p_n)^{12}}{|x - y|^{14}} c_n(x) c_p(y) \delta x d\tilde{y} \right) \]
\[ -\int_{\Omega} \int_{\Omega} \left( \frac{6 e_{p,n}(a_n + p_n)^{12}}{|x - y|^{14}} c_p(x) c_n(y) \delta \mathbf{c}_n \mathbf{c}_n(y) d\tilde{x} d\tilde{y} \right). \]

Attractive (or other) forces can be included either as another component of the energy in the atomic scale energy Eq. (18) and atomic scale dissipation (28) or by including a traditional description of the excess free energy of ionic solutions into the macroscopic scale equations of the compressible ionic fluid, e.g., using the functionals of the DFT of liquids. An attractive Lennard-Jones force could be easily added since it has the same form as the repulsive term used above Eq. (A1). Other models of interatomic forces could be used, including (for example) estimates of the “potential of mean force” between ions (or molecules) determined from the radial distribution functions in simulations or experiments. Adding an attractive term of any type will yield a different partial differential equation from Eqs. (31) and (32).

**APPENDIX B: THE ELECTROCHEMICAL FORCES \( \nabla \mu_n \) AND \( \nabla \mu_p \) INCLUDING THE EFFECTS OF FINITE SIZE IONS**

In Appendix A we derived just the part of the electrochemical forces \( \nabla \mu_n \) and \( \nabla \mu_p \) related to the finite size of ions. Here, we complete the system. We compute the diffusional and electrostatic terms of the electrochemical forces,
\[ E = \int_{\Omega} \left( k_B T \log c_n + \frac{1}{2} \frac{\phi^2}{\phi^2} \right) d\tilde{x} \]
\[ + \int_{\Omega} \left( k_B T \log c_p + \frac{1}{2} \frac{\phi^2}{\phi^2} \right) d\tilde{x}. \]

Since \( \phi \) is the solution of Poisson equation, we have that
\[ \phi(\mathbf{x}) = \int_{\Omega} G(\mathbf{x},\mathbf{y}) (c_n(\mathbf{y}) - c_p(\mathbf{y})) d\mathbf{y} \]
with Green’s kernel \( G(\mathbf{x},\mathbf{y}) \). Substituting Eq. (B2) with \( \phi \), the generalized energy of Eq. (B1) becomes
\[ E = \int_{\Omega} \left( k_B T \log c_n(\mathbf{x}) + k_B T \log c_p(\mathbf{x}) \right) \]
\[ + (c_p(\mathbf{x}) - c_n(\mathbf{x})) \int_{\Omega} G(\mathbf{x},\mathbf{y}) \frac{c_n(\mathbf{y}) - c_p(\mathbf{y})}{e} d\mathbf{y} \]
\[ (\mathbf{B}3) \]

Hence, the variational derivative employing Eq. (A4) leads us to

\[ \delta E_n = \int_{\Omega} \left( \delta c_n(\mathbf{x}) k_B T \log c_n(\mathbf{x}) + \frac{\delta c_p(\mathbf{x})}{2} \right) \int_{\Omega} G(\mathbf{x},\mathbf{y}) \frac{c_n(\mathbf{y}) - c_p(\mathbf{y})}{e} d\mathbf{y} d\tilde{x} \]
\[ + \int_{\Omega} \left( \frac{\delta c_p(\mathbf{x})}{2} \right) \int_{\Omega} G(\mathbf{x},\mathbf{y}) \frac{c_n(\mathbf{y}) - c_p(\mathbf{y})}{e} d\mathbf{y} d\tilde{x}. \]

(B4)
\[
E(\text{Hard Sphere}) = E_{\text{HS}}^{\text{DFT}} = \int \Phi_{\text{HS}}(\{n_a(\vec{y}), \{\vec{n}_\beta(\vec{y})\}\}) d\vec{y},
\]

(C1)

where

\[
\Phi_{\text{HS}}(\{n_a(\vec{y}), \{\vec{n}_\beta(\vec{y})\}\}) = -n_0 \log(1 - n_3) + \frac{n_1 n_2 - \vec{n}_5 \cdot \vec{n}_5}{1 - n_3} + \frac{n_2^3}{24\pi(1 - n_3^2)^2} \left(1 - \frac{\vec{n}_5 \cdot \vec{n}_5}{n_2^2}\right)^3.
\]

(C2)

and

for \( \alpha = 0, 1, 2, 3 \)
\[
\Phi_{\text{HS}}(\{n_a(\vec{y}), \{\vec{n}_\beta(\vec{y})\}\}) = \sum_i c_i(\vec{y}) \omega_i^{(a)}(\vec{y} - \vec{x})d\vec{y},
\]

(C3)

for \( \beta = 4, 5 \)
\[
\Phi_{\text{HS}}(\{n_a(\vec{y}), \{\vec{n}_\beta(\vec{y})\}\}) = \sum_i c_i(\vec{y}) \omega_i^{(a)}(\vec{y} - \vec{x})d\vec{y},
\]

(C4)

\[
\omega_i^{(2)}(\vec{r}) = \delta(\vec{r} - \vec{x}_i),
\]

(C5)

The chemical potential is given for \( \alpha = 0, \ldots, 3, \beta = 4, 5 \),
\[
\mu_i^{\text{HS}}(\vec{x}) = k_B T \sum \int \frac{\partial \Phi_{\text{HS}}}{\partial n_a}(\vec{y}) \omega_i^{(a)}(\vec{x} - \vec{y})d\vec{y}
\]

(C6)

\[
\mu_i^{\text{HS}}(\vec{x}) = k_B T \sum \int \frac{\partial \Phi_{\text{HS}}}{\partial \vec{n}_\beta}(\vec{y}) \omega_i^{(a)}(\vec{x} - \vec{y})d\vec{y}.
\]

(C7)

\[
\text{force}_{\text{HS}}^{\text{DFT}}(\vec{x}) = -c_i(\vec{x}) \nabla \mu_i^{\text{HS}}(\vec{x}).
\]

ACKNOWLEDGMENTS

We are grateful and fortunate for a conversation with Stuart Rice that motivated our treatment and discussion of optimal control. He and Douglas Henderson reminded us of the utility of equations of state. Thanks to them, Dezső Boda, Jim Fonseca, Dirk Gillespie, Eberhard von Kitzing, and Roland Roth for continually reminding us to compare our results with previous work on the primitive model (particularly near walls). Thanks to Fred Cohen for important advice. The work was supported in part by NIH (Grant No. GM076013). C.L. is supported by NSF (Grant No. DMS-0707594). Much of the work was done at the Institute for Mathematics and Its Applications, University of Minnesota, Minneapolis, MN supported by the NSF.

5C. Liu, in Multi-Scale Phenomena in Complex Fluids: Modeling, Analysis and Numerical Simulation, Series in Contemporary Applied Math,