JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

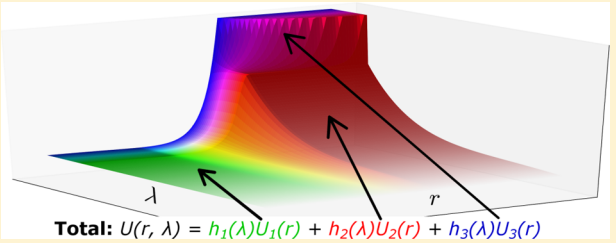
Linear Basis Function Approach to Efficient Alchemical Free Energy Calculations. 1. Removal of Uncharged Atomic Sites

Levi N. Naden,[†] Tri T. Pham,[‡] and Michael R. Shirts*,[†]

[†]Department of Chemical Engineering, University of Virginia, Charlottesville, Virginia 22902, United States

[‡]Biozentrum University of Basel, Klingelbergstrasse, Basel, Switzerland

ABSTRACT: We present a general approach to transform between molecular potential functions during free energy calculations using a variance minimized linear basis functional form. This approach splits the potential energy function into a sum of pairs of basis functions, which depend on coordinates, and 'alchemical' switches, which depend only on the coupling variable. The power of this approach is that, first, the calculation of the coupling parameter dependent terms is removed from inner loop force calculation routines, second, the flexibility in specifying basis functions and alchemical switches allows users to choose transformation pathways that maximize statistical efficiency, and third, it is possible to predict entirely in postprocessing, without any additional energy evaluations, the thermodynamic properties along any alchemical path with moderate overlap from an initial simulation that uses the same basis functions. This allows construction of optimized, minimum variance alchemical switches from a single simulation with fixed basis functions and trial alchemical switching functions. We describe how to construct these linear basis potentials for real molecular systems of different sizes and shapes, considering particularly the problems of eliminating singularities and minimizing variance of particle removal in dense fluids. The statistical error in free energy calculations using the optimized basis functions is lower than standard soft core models, and approach the minimum variance possible over all pair potentials. We recommend an optimized set of basis functions and alchemical switches for standard molecular free energy calculations.



1. INTRODUCTION

Designing new molecules with desired thermodynamic properties requires efficient prediction methods to explore even a small fraction of combinatorial chemistry space.[1] Accurate simulation-based methods allow us to predict thermodynamic properties sensitive to small molecular changes, making them highly valuable for large chemical space screening. Two thermodynamic properties that have been studied most intently by simulation methods are the free energy of solvation and ligand binding free energies, as these quantities can provide insight in to the efficacy of drug candidates.[2] There are also many other properties of physical interest that can be related to free energies, including solubilities[3] and membrane permeabilities.[4]

"Alchemical" transformations are common computational methods for calculating free energy differences. In these methods, the free energy is estimated along a computationally efficient, often nonphysical thermodynamic path connecting the end states of interest. Estimating the free energy along such a path can be performed with several methods each having their own advantages and disadvantages.[5,6] Choosing a statistically efficient pathway, however, is a nontrivial issue.[7−14] In virtually all methods, we must simulate intermediate states along the coupling pathway to obtain a free energy with acceptable statistical error. Some methods require the potential of a configuration to be evaluated at neighboring states along the pathway, such as the Bennett acceptance ratio (BAR)[15] or free energy perturbation. Multistate methods require the potential of a configuration from one state to be evaluated at every other state, such as the Weighted Histogram Analysis Method (WHAM)[16,17] or multistate BAR (MBAR).[18] Thermodynamic integration (TI) requires derivative along the coupling path to be computed at the simulated state.

The challenge for rapid calculation of free energies is that the cost of generating enough samples to obtain sufficiently accurate estimates can become prohibitively large if there are numerous intermediate states along the path or if the systems have long correlation times at some or all of the intermediate states. To help overcome this problem, we can select coupling pathways that have high overlap in configuration space among neighboring states along the path to reduce the amount of simulation required for a given statistical error tolerance.[7,9,13] The number of intermediate states required along a pathway will depend on the specific molecular transformation, the statistical tolerance for the problem, and the computational budget.

Increasing the overlap in configurational phase space along a path is equivalent to minimizing the statistical variance of free energy calculations along the pathway using TI. More precisely, the error in the free energy estimation using TI is proportional to the square root of the variance over the number of samples

Received: October 21, 2013
Published: January 31, 2014

taken.[13,14] The variance is an intrinsic property of the path along which the transformation is performed, independent of the number of samples, and determined by the thermal fluctuations around equilibrium at states along the pathway.[8] The variance of each point along the pathway can also be thought of as the square of the Riemannian metric used to measure thermodynamic length along that pathway.[8,19,20] Minimizing the thermodynamic length between states will maximize their phase space overlap and reduce statistical error in free energy calculations. This relationship is exact for TI but can be shown to be approximately true for other free energy methods.[13]

A common approach which produces a fairly statistically efficient alchemical path for the disappearance or appearance of molecules with Lennard-Jones site potentials is the "soft core" potential, with the most general form[13,21,22]

$$U(r,\lambda) = 4\varepsilon_{ij}\lambda^{a}\left[\left(\frac{1}{\alpha(1-\lambda)^{b} + (r/\sigma_{ij})^{c}}\right)^{12/c} - \left(\frac{1}{\alpha(1-\lambda)^{b} + (r/\sigma_{ij})^{c}}\right)^{6/c}\right] \quad (1)$$

where $U(r,\lambda)$ is the pairwise nonbonded potential depending on radial distance r between two atoms and the alchemical coupling variable λ. a, b, and c are positive, usually integer, constants, α is a positive free parameter that can be optimized, and ε_{ij} and σ_{ij} are the Lennard-Jones (LJ) parameters. Potentials of this type provide a relatively good phase space overlap between neighboring states, which in turn reduces the variance.[13,15,21,22] As λ goes from fully coupled to decoupled, soft core potentials cause the excluded volume region in the pair interaction to gradually decrease in energy, allowing solvent molecules to "leak" into the excluded volume of the solute, until the solute excluded volume disappears completely. The correct choice of constants can reduce the total variance summed over all intermediate states for this class of potentials. The original version of soft core potentials had $a = 1$ or $a = 4$, $b = 2$, $c = 6$, and $\alpha = 0.5$. The choice of $a = 1$, $b = 1$, $c = 6$, and $\alpha = 0.5$ is about 30% more efficient;[13,23] we will refer to this choice as the "1−1−6" potential. A soft core potential that is near minimal variance over all possible pair potentials of r and λ has recently been developed with parameters $a = 1$, $b = 1$, $c = 48$, and $\alpha \approx 0.0025$, which we will refer to as the "1−1−48" potential.[13,14]

Soft core potentials increase the difficulty of implementing free energy methods on new, highly parallelized architectures. Developing new, generalized algorithms on architectures such as GPUs or Many Integrated Core (MIC) platforms has been of large interest to the molecular simulation community. Coding generalized algorithms on these platforms is a nontrivial task and soft core potentials add complexity as configuration and alchemical information must both be passed to the platform calculating the inner force loop due to the nonlinear coupling of the two variables. For free energy methods requiring derivatives, the derivative would also need to be calculated on the platform, further increasing difficulty in quickly writing new code. Eliminating the need to code alchemical transformation-specific information in the accelerated inner loops would reduce the time needed to implement free energy methods on new platforms, which has lagged compared to the implementation of other molecular mechanics functions on these platforms.

We can reduce the difficulty in coding free energy on new platforms and remove the free energy calculations from the inner loops if we write the potential energy as

$$U(r,\lambda) = \sum_{i=1}^{N} h_i(\lambda)u_i(r) + u_{\text{unaffected}}(r) \quad (2)$$

where $h_i(\lambda)$ are alchemical switching functions depending only on λ, and $u_i(r)$ are pairwise potential basis functions depending only on r. In this equation, N is the total number of basis function and alchemical switch combinations, and $u_{\text{unaffected}}$ is the portion of the potential energy of the system not dependent on the coupling parameter. For example, to insert a single solute site into a box of solvent, we could write

$$U(r,\lambda) = h_1(\lambda)\sum_{j=1}^{N_{\text{solvent atoms}}} U_{\text{LJ}}(r) + h_2(\lambda)\sum_{j=1}^{N_{\text{solvent atoms}}} U_{\text{electrostatic}}(r) + U_{\text{solvent-solvent}}(r) \quad (3)$$

with $h_1(\lambda) = (2\lambda)^4$ for $0 < \lambda < {}^1/_2$ and 1 for ${}^1/_2 < \lambda < 1$, and $h_2(\lambda) = 0$ for $0 < \lambda < {}^1/_2$ and $2\lambda-1$ for ${}^1/_2 < \lambda < 1$. This would be equivalent for a workable, though not particularly statistically efficient pathway of first turning on the Lennard-Jones interaction with a λ^4 dependence, and then turning on the electrostatic interaction with linear dependence. We can certainly construct a more efficient path, as we will show, but this provides an example how existing approaches fit into this general formalism.

Using this method, the terms needed for different free energy methods can then be calculated outside of the inner loop. We only need to compute the potential energy of the basis functions $u_i(r)$ once for all intermediate states, with the rest of the required calculations involving manipulations of the coupling functions $h_i(\lambda)$ outside the force calculation of the inner loop. Additionally, no special code for the inner loop is required to compute $\partial\mathcal{H}/\partial\lambda$, the derivative of the Hamiltonian, as it can be calculated directly from the basis function terms with minor postprocessing.

Some instances of such linear potentials have been studied previously, so the functional form itself is not entirely novel. For example, initial free energy applications often multiplied the entire potential energy of the molecule disappearing with a single linear scaling term, of the form $U(\lambda,r) = (1-\lambda)U_{\text{initial}}(r) + \lambda U_{\text{final}}(r)$. However, these simplest linear combination potentials are generally no longer used for problems involving changing numbers of atomic sites because they lead to highly inaccurate, numerically diverging free energies.[7,9,21,22,24] This occurs because r^{-n} terms common in many intermolecular potentials create an infinite potential, or singularity, at $r = 0$ for all λ except $\lambda = 0$. The sudden jump of the potential from infinity to zero when approaching $\lambda = 0$ causes the thermodynamic expectation of $\langle\partial U/\partial\lambda\rangle$ to diverge, giving large variances and uncertainties.[7] This problem is sometimes called the "end point catastrophe."[9,24] As discussed, soft core potentials avoid this problem as the singularity is replaced by a truncated core potential at $r < \sigma$ that gradually disappears as the coupling parameter reaches the decoupled state.

Many simulation packages have moved away from linear interpolation and implement some form of soft core method

for free energy calculations. For example, GROMACS,[25] GROMOS,[10,11] CHARMM,[26] and AMBER[27] all implement a variant of soft core potentials to carry out free energy calculations with an option to do simple linear interpolation. Some of the packages allow interpolation with a power such that $U(\lambda,r) = (1 - \lambda^n)U_{\text{initial}}(r) + \lambda^n U_{\text{final}}(r)$, which removes the end point catastrophe for $n \geq 4$, but can still have simulation stability issues for some systems and time steps.[9] GROMOS allows defining products of an arbitrary function of λ, similar to our $h_i(\lambda)$, times the soft core equation (eq 1). However, these functions still involve the soft core functional form, leaving r and λ coupled. In all packages, the choice of "best" path is left to the user. The basis function approach we present here differs from all soft core methods in that it requires that r and λ be decoupled from each other in separate functions and it differs from traditional linear interpolation since more complex switch functions than λ^n are permitted. Despite this limitation on the basis functions, we will show that a near optimal path can be found to minimize statistical error with this basis function approach.

Recent work has tried to identify new potential formalisms that can avoid the end point catastrophe while still keeping the advantages of a linear basis potential, such as the form described by Buelens and Grubmüller.[12] This approach can be written as single basis function from eq 2 as $U(r, \lambda) = \lambda\sum_{\text{affected pairs}} u(r)$ but avoids the infinite potential at $r = 0$ by setting a maximum finite value for the basis function $u(r)$ and creating a polynomial switch from the LJ function to this cap. However, it is unclear what the statistical efficiency of this approach is, as it was not directly tested, and the fact that significantly more intermediates are required near $\lambda = 0$ than near $\lambda = 1$ indicates it is not a particularly statistically efficient path.[12] Additionally, the approach requires conditionals to be evaluated with every pair potential to check for the needed for caps, which could prove problematic in some highly vectorized implementations.

The goal of this work is to present a formalism for developing low variance pathways for molecular transformations based on linear combinations of more than one coordinate basis set. The present study combines the implementation simplicity of the linear basis potential approach and the statistical efficiency of optimized soft core potentials. We will follow this approach to propose statistically efficient linear basis pathways that work in general cases of removing molecules from dense fluids. We can minimize the variance of families of linear basis potentials with the methods we have previously developed,[13,14] resulting in a family of highly statistically efficient alchemical paths that also reduce the cost of re-evaluating potentials and can simplify implementing free energy code.

2. THEORETICAL BASIS

We can write the most general version of the linear basis potential as

$$U(r, \lambda) = \mathbf{h}(\lambda)\mathbf{u}^T(r) + u_{\text{unaffected}}(r) \tag{4}$$

which is a vectorized version of eq 2. Here, the vector functions $\mathbf{h}(\lambda) = [h_1(\lambda), h_2(\lambda),...,h_N(\lambda)]$, and $\mathbf{u}(r) = [u_1(r),u_2(r),...,u_N(r)]$ replace the individual components in the previous equation. λ can vary from 0 to 1 and each of the N $h_i(\lambda)$ is a monotonic function of λ, which has end points $h_i(0) = 0$, resulting in a fully noninteracting $u_i(r)$ term, and $h_i(1) = 1$, which results in a fully interacting $u_i(r)$ term. The only requirements we place on the $h_i(\lambda)$ is that they are continuous and that $dh_i(\lambda)/d\lambda \geq 0$, that is, they are monotonic, although we allow some of $h_i(\lambda)$ to be fixed while the others change. This formalism means that λ can be thought of as a curve through the N-dimensional alchemical space mapping the single $[0,1]$ domain to the N $h_i(\lambda)$ which span a $N \times [0,1]$ range. A key to our linear basis potential approach is that there are *multiple* $h_i(\lambda)$ and *multiple* basis functions, which together can produce a broad range of functions that can be adjusted to create a highly statistically efficient pathway with minimum variance. We will occasionally leave off the explicit dependence of $h_i(\lambda)$ on λ and write h_i in some equations for simplicity.

Computing the potential energies at different values of λ becomes trivial if the basis function energies are tabulated and can be done either in code (but outside the inner loop) or in postprocessing, as long as the energies of each basis function are stored at each configuration of interest. The derivative $\partial U/\partial\lambda$ can also be computed trivially using the basis function energies, because

$$\frac{\partial U(r, \lambda)}{\partial\lambda} = \sum_i^N \frac{\partial U(r, \lambda)}{\partial h_i}\frac{\partial h_i}{\partial\lambda} = \sum_i^N \frac{\partial h_i}{\partial\lambda}u_i(r) = \mathbf{h}'(\lambda)\mathbf{u}^T(r) \tag{5}$$

where $\mathbf{h}'(\lambda) = [\partial h_1/\partial\lambda,\partial h_2/\partial\lambda,...]$.

There are two main choices that must be made to design a linear basis potential representation of some pathway through alchemical space:

1. Choosing the basis functions, $\mathbf{u}(r)$.
2. Choosing the alchemical switching functions, $\mathbf{h}(\lambda)$.

The basis functions must be chosen to match the potential function at the end states as well as avoid the end point catastrophe. Beyond that, there is significant freedom in the choice of basis functions. Given a set of basis functions satisfying the conditions at the end points, the alchemical switching functions can be chosen to minimize the variance of the path along λ. The statistical uncertainty of a simulation performed along this path is a function of both the choice of basis and the choice of switches. However, given a choice of basis, there exists a single set of switches that minimizes the variance of the path.

There is one complication we encounter while determining the alchemical switches given a set of basis function. As we will see, approaches to minimize the variance over all the switches simultaneously break down when one or more of the $h_i(\lambda)$ functions are zero while the others change. Unfortunately, many of the pathways most straightforward to implement may have this feature; for example, in eq 3, we change the Lennard-Jones and electrostatic energy terms sequentially. We will therefore separately consider the "alchemical schedule," the choice of the ranges of λ where each $h_i(\lambda)$ is changing or constant, and treat this as a third choice we must make to determine the pathway.

To help visualize what we mean by a schedule, Figure 1 shows a few sample schedules that could be taken connecting initial and final states. All schedules will by definition give the same free energy between end states, but some alchemical schedules are larger variance pathways than others, and some may give completely divergent answers. For example, turning on attractive electrostatic interactions before turning on any

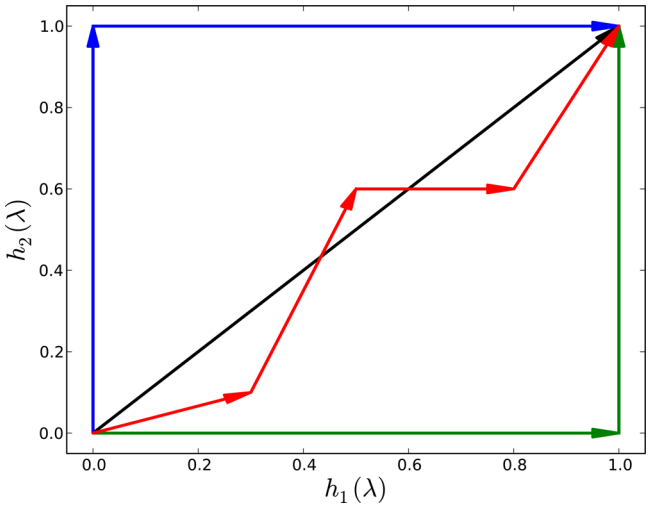


Figure 1. Different alchemical schedules can be used to carry out the same free energy transformation, each with different variances. An example of different choices of the alchemical schedule is shown for when only two $h_i(\lambda)$ are present. Both h_i's could change simultaneously (black) as with attractive and repulsive components of soft core potentials. Each h_i could be changed one at a time (blue or green) such as the common case when decoupling electrostatics before other forces. Alternately, a custom schedule (red) could be chosen to improve sampling at the corresponding states. All pathways yield the same end states and thus the same values for state functions. However, some paths may be more statistically efficient than others.

repulsive Lennard-Jones interactions will lead to negative infinite potential energies and crashing simulations.

Due to the complexity in general design of a linear basis set, we will use a single simple schedule throughout this study and focus on the logic of the other design choices. We will leave the selection of best schedule for a specified choice of basis functions to a future study. We will also focus primarily on determining statistically efficient paths for solute−solvent nonbonded interactions. Solute−solute interactions, both bonded and nonbonded, will behave differently because of the strong couplings in intramolecular degrees of freedom.

2.1. Fixing the Alchemical Schedule. The three main nonbonded interactions used by most pairwise force fields are repulsive, van der Waals dispersion, and electrostatic forces. In this initial study, we assume the electrostatic interactions are fully decoupled using the linear alchemical switch defined by $h_{\text{electrostatic}}(\lambda) \propto \lambda$ before any alchemical modifications of the attractive or the repulsive interactions. This linear electrostatic switch is a common choice that has been found to be quite statistically efficient and often considered best practice.[9,28,23,29] The correct amount of simulation effort to devote to turning off electrostatic interactions will depend on the contribution to the statistical error of the electrostatic versus Lennard-Jones terms, which will depend on the magnitude of the partial charges. We will leave the further optimization of this choice for future studies and instead concentrate on the more challenging removal of Lennard-Jones terms. After the electrostatics are decoupled, $h_{\text{electrostatic}}(\lambda)$ will be 0 for the remainder of the transformation; for example, we would set $h_2(\lambda) = 0$ for all λ shown in our example eq 3.

For this study, we also choose for both the repulsive and the attractive forces to be modified simultaneously with alchemical switches $h_{\text{repulsive}}(\lambda)$ and $h_{\text{attractive}}(\lambda)$, respectively. These switches will change together only after $h_{\text{electrostatic}} = 0$. Because the electrostatics are assumed to already be off, any solvation free energy differences presented in this study will be of the uncharged molecules. This allows us to directly compare to the magnitude of statistical error computed using soft core simulations.[13,14] For simplicity, we will also rescale λ to represent only this nonelectrostatic part of our schedule such that $h_{\text{repulsive}}(\lambda = 0) = h_{\text{attractive}}(\lambda = 0) = 0$ and $h_{\text{repulsive}}(\lambda = 1) = h_{\text{attractive}}(\lambda = 1) = 1$.

This particular choice of schedule and electrostatic alchemical switch is reasonable because the disappearance of Lennard-Jones terms is usually the largest contribution to the statistical uncertainty of free energy calculations.[13] Further optimizing the components of the linear combination potential for molecules with significant electrostatic interaction, as well as examining more general choices for alchemical schedules will be addressed in a future study.

2.2. Selecting Accurate Basis Functions without Singularities. The singularity in the basis functions can be avoided by setting a maximum finite potential referred to here as a "cap." If the cap on the potential is large enough, then the probability of observing atoms within the capped region of the potential energy surface will be effectively zero. This capped potential method was also explored by Buelens and Grubmüller,[12] who examined a linear coupling parameter, but we wish to generalize the potential in our more general linear basis potential formalism.

Using a capped repulsive potential function as one of our basis functions for a minimum variance path is very strongly suggested by inspecting the shape of the optimized soft core potential, as shown in Figure 9 of Pham and Shirts.[13] In the 1−1−48 form of the soft core potential, at intermediate states, for $r < 0.8\sigma_{ij}$, the potential becomes almost completely flat with respect to r, looking very much like a capped potential in r with a capping distance independent of λ. This capped portion of the potential is then scaled to zero as $\lambda \to 0$.

If the capped potential is used as the end state of the simulation, then the capped potential at the fully coupled state must be high enough such that the probability of observing a pair of atoms in the capped region is statistically zero on the time scales of any simulation. We wish to quantify the deviation that would result from replacing the full repulsive potential with a cap. Because the largest contribution to the variance in introducing an atomic interaction site comes from two-particle repulsive interactions,[13] we can approximate the radial distribution function $g(r,\lambda)$ when two particles are very close by the zeroth-order approximation

$$g(r, \lambda) \approx e^{-\beta U(r,\lambda)} \tag{6}$$

where $\beta = 1/k_B T$ and T is the temperature. If this RDF approximation for the capped Lennard-Jones potential and the uncapped potential are sufficiently well matched, the basis functions should be statistically identical to other potentials at the fully coupled end state. We will eventually find that there are problems with using an approximate capped end state. However, we will also show that any deviations caused by the capping procedure can be addressed with minimal additional complexity.

To keep with the basis function formalism, we treat the short-range repulsive forces and the long-range attractive forces separately by applying the Weeks−Chandler−Andersen (WCA) decomposition[30] to our capped potential. The

decomposition provides us two differentiable potential basis functions and is written as

$$U_{wca}(r) = u_R(r) + u_A(r) \tag{7}$$

where the subscripts R and A denote the repulsive and attractive terms respectively. Denoting $u_{LJ}(r)$ as the full Lennard-Jones potential, the individual terms of the WCA potential can be written as

$$\begin{aligned} u_R(r) &= u_{LJ}(r) + \varepsilon_{ij} \quad \text{for} \quad r < 2^{1/6}\sigma_{ij} \\ &= 0 \qquad\qquad \text{for} \quad r \geq 2^{1/6}\sigma_{ij} \end{aligned} \tag{8}$$

$$\begin{aligned} u_A(r) &= -\varepsilon_{ij} \quad \text{for} \quad r < 2^{1/6}\sigma_{ij} \\ &= u_{LJ}(r) \quad \text{for} \quad r \geq 2^{1/6}\sigma_{ij} \end{aligned} \tag{9}$$

The WCA decomposition avoids the unrealistic negative singularity from the attractive van der Waals forces brought on from capping only repulsive interactions. Setting an arbitrary negative cap is an alternative,[12] however, this adds extra parameters, which we wish to avoid for simplicity.

To make the function continuous, the repulsive potential must be capped smoothly. We choose a smooth polynomial function defined over the transition region. This polynomial should be monotonic to avoid creating artificial minima, as well as not imposing an excessive deviation in the force from the normal Lennard-Jones potential. We avoid these problems by setting a quartic polynomial to match the potential energy values and derivatives at either end of the transition region, and the second derivative of the potential at the higher r end point. The second derivative condition on the quartic polynomial was chosen to keep the polynomial monotonic over the transition. Cubic polynomials sometimes fail to be monotonic over the transitions, we tested with different Lennard-Jones parameters. Applying these changes to the repulsive basis function gives

$$\begin{aligned} u_{R,cap} &= u_{cap} && \text{for} && r < f_{cap}\sigma_{ij} \\ &= Ar^4 + Br^3 + Cr^2 + Dr + E && \text{for} && f_{cap}\sigma_{ij} \leq r < f_{switch}\sigma_{ij} \\ &= u_{LJ}(r) + \varepsilon_{ij} && \text{for} && f_{switch}\sigma_{ij} \leq r < 2^{1/6}\sigma_{ij} \\ &= 0 && \text{for} && r \geq 2^{1/6}\sigma_{ij} \end{aligned} \tag{10}$$

where f_{cap} is the fraction of σ_{ij} over which the capping region extends and where the switch starts and f_{switch} is the fraction of σ_{ij} at which the switch ends and the normal Lennard-Jones function resumes. $u_{cap} = u_R(f_{cap}\sigma_{ij})$ is the capped, constant potential, the constants $A-E$ are fit to meet the conditions at either end of the transition region, and the choice of the constants of f_{cap} and f_{switch} will be discussed in the next section. The two basis functions we use in the vector $\mathbf{u}(r)$ are thus eq 10 for the repulsive term and eq 9 for the attractive term.

2.2.1. Deviation of the Capped Potential from a Fully Coupled Lennard-Jones Potential. The thermodynamic difference between the capped WCA potential and a normal Lennard-Jones potential will depend on how large the cap is. The value of the cap will depend on the choice of the transition region between f_{cap} and f_{switch} in eq 10. If the cap is high enough, that is, small f_{cap}, then there will be no statistical difference between the capped potential and the Lennard-Jones potential, as the energy barrier will prevent pairs of atoms from being closer than $f_{cap}\sigma_{ij}$. However, large caps will be less statistically efficient for decoupling, as disappearing a large barrier will have larger variances. Capped potentials with smaller barriers will be more statistically efficient to decouple but will be thermodynamically different than the Lennard-Jones potential.

We can add a step into the alchemical schedule to account for any thermodynamic difference between the capped and the Lennard-Jones potentials. This modifies our alchemical schedule to become a three-step process in turning on the interactions between a molecule and its surroundings. First, we turn on the capped repulsive and van der Waals attractive forces. Second, we linearly transition between the capped repulsive potential and the original repulsive term of the WCA decomposition.[30] Finally, we turn on the electrostatics of the solute. This schedule still fits perfectly within the context of the linear basis function approach and the basis function representation for this additional step is

$$U_{capping}(r, \lambda) = \lambda(u_R(r) - u_{R,cap}(r)) + u_{R,cap}(r) \tag{11}$$

The value of f_{cap} will determine magnitude of the cap and the statistical efficiency of this step. If f_{cap} is too large, the variance for disappearing the infinite potential will be large since the low barrier does not prevent pairs of atoms from 'leaking' into the core.

After some experimentation, we chose a transition region where $f_{cap} = 0.8$ and $f_{switch} = 0.9$ for this study. Examining the 1−1−48 potential with several combinations of ε_{ij} and σ_{ij}, we found that the transition from a scaled Lennard-Jones potential to the capped potential generally takes place in this region. One might expect naively that the ideal height would correspond to a potential energy of several times k_BT, since the probability of single pairwise interactions become negligible with only a few k_BT. However, this neglects the fact that shorter contact distances can easily be overcome by multiple positive interactions in the rest of the system. This means that to be completely thermodynamically indistinguishable from a Lennard-Jones particle, this cap must be higher, between 20 and $50k_BT$ for solute−solvent interactions, as will be detailed below.

The thermodynamic difference between the capped WCA-decomposed potential and a normal Lennard-Jones potential should be small to prevent large variances when switching between the full r^{-12} repulsive term and the capped potential. Figure 2 shows the zeroth-order RDF for the Lennard-Jones potential and the capped potential. The two curves are identical for $r \geq 0.9\sigma_{ij}$ by design and only notably differ in the transition region between $0.8\sigma_{ij}$ and $0.9\sigma_{ij}$. Because this potential energy difference is small, the Boltzmann weights of the two curves will also be small and the curves will have good phase space overlap. Later in the study, we will examine variance of transition regions larger and smaller than the $0.8\sigma_{ij}$−$0.9\sigma_{ij}$ range and will demonstrate problems with using both of these ranges, though we must first introduce additional tools to diagnose this. The remainder of the study will assume this transition region between $0.8\sigma_{ij}$ and $0.9\sigma_{ij}$ unless otherwise noted.

2.3. Designing Low Variance Alchemical Switching Functions. We next must choose the alchemical switches corresponding to the repulsive and attractive basis functions, with the goal to minimize the total variance of the transformation. To minimize the variance along the path, we take the approach presented by Pham and Shirts.[13,14] We wish to minimize the objective function of the total variance, which

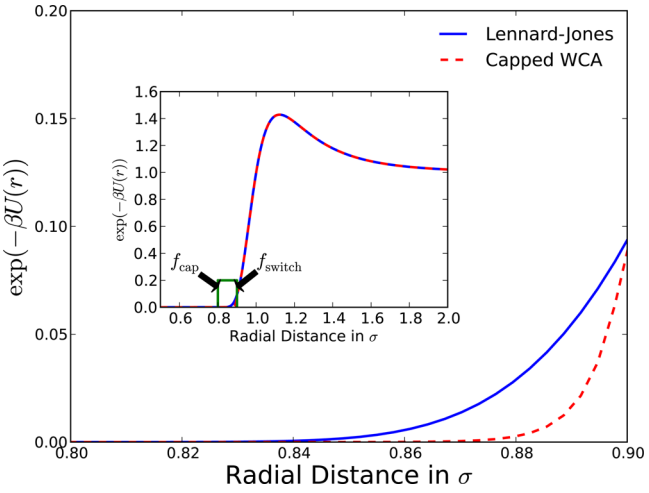


Figure 2. Capped potential should be statistically similar to an uncapped potential at the fully coupled state. The zeroth-order RDF comparison of a Lennard-Jones potential and the capped Weeks–Chandler–Andersen potential for UA Methane interacting with a TIP3P water are shown. Green box in inset highlights the enlarged region. The region of interest is $r = 0.8\sigma_{ij}$ to $r = 0.9\sigma_{ij}$ where the two potentials differ. The capped potential is large enough to be indistinguishable from a normal Lennard-Jones potential below $0.8\sigma_{ij}$ and only slightly different in the switch region. This difference has low enough Boltzmann weight that it should have minimal affect on the simulation.

can first be estimated by evaluating a zeroth-order approximation to the RDF as

$$\mathrm{Var}(\Delta F) \approx \frac{4\pi\rho}{\beta}\int_0^1\int_0^\infty\left(\frac{\partial U(r,\lambda)}{\partial\lambda}\right)^2\exp(-\beta U(r,\lambda))r^2\mathrm{d}r\mathrm{d}\lambda \tag{12}$$

where ρ is the solvent number density, and F is either the Helmholtz or Gibbs free energy, depending on the ensemble sampled, with an additional PV term in the Boltzmann weight if examining the Gibbs free energy. Although this approach assumes thermodynamic integration (TI), previous research has shown that the optimal TI pathway is also (to within statistical error) the minimum variance pathway for BAR and MBAR,[13] so additional theories for those methods need not be developed.

We know that the soft core 1−1−48 potential has near minimum variance among the family of all pairwise paths.[13,14] If the RDF of our linear basis potential pathway closely matches the RDF for 1−1−48 in both λ and r space, then the variances must also closely match as well. The zeroth-order RDF is a suitable alternate to directly matching the shape of $U(r,\lambda)$ because it is fully described by the pairwise potential and finite everywhere. This approximated RDF also allows comparison of capped potentials like our WCA-decomposed potential and soft core 1−1−48.

Equation 12 can be generalized to a set of $h_i(\lambda)$ components by rewriting it as a path integral

$$\mathrm{Var}(\Delta F) \approx 4\pi\rho\int_0^1\int_0^\infty\sum_{i,j}\left(\frac{\partial U}{\partial h_i}\frac{\partial U}{\partial h_j}\exp[-\beta U(r,\lambda)]r^2\right)\times\frac{\partial h_i}{\partial\lambda}\frac{\partial h_j}{\partial\lambda}\mathrm{d}r\mathrm{d}\lambda \tag{13}$$

Up until now, we have used a zeroth-order approximation to the RDF. If the potential can be decomposed into basis functions and alchemical switches, the variance can be derived exactly without any assumptions about the RDF in terms of the potential energy by starting from the equation for TI as

$$\Delta F = \int_0^1\sum_{i=1}\left\langle\frac{\partial\mathcal{H}}{\partial h_i}\right\rangle\frac{\partial h_i}{\partial\lambda}\mathrm{d}\lambda \tag{14}$$

$$\mathrm{Var}(\Delta F) = \int_0^1\sum_{i,j}\mathrm{Cov}\left(\frac{\partial U}{\partial h_i},\frac{\partial U}{\partial h_j}\right)\frac{\partial h_i}{\partial\lambda}\frac{\partial h_j}{\partial\lambda}\mathrm{d}\lambda \tag{15}$$

Defining the $N \times N$ covariance matrix of $\mathbf{u}$ as $\mathrm{Cov}(\mathbf{u},\mathbf{u})$, we can rewrite the variance at a specific λ as

$$\mathrm{Var}(\Delta F)(\lambda) = \mathbf{h}'(\lambda)\cdot\mathrm{Cov}(\mathbf{u},\mathbf{u})(\lambda)\cdot\mathbf{h}'^{T}(\lambda) \tag{16}$$

The total variance over the entire transformation is then the integral:

$$\mathrm{Var}(\Delta F) = \int_0^1\mathbf{h}'(\lambda)\cdot\mathrm{Cov}(\mathbf{u},\mathbf{u})\cdot\mathbf{h}'^{T}(\lambda)\mathrm{d}\lambda \tag{17}$$

These equations are derived in Appendix A. For the derivation of initial alchemical switches, we will focus on the zeroth-order RDF approximation, eq 13, as no simulations are required and optimizations can be carried out in seconds. Once we have an initial estimate, the zeroth-order RDF approximation will not be needed since we will have full potentials and can calculate the variance directly from eq 17.

To obtain the variance at any arbitrary point in state space given simulation data at one or multiple states, we use eq 17, with the covariance matrix estimated using MBAR.[18] The total potential and Boltzmann weights of any point in state space are straightforward to evaluate using eq 4. To compute the covariance matrix with reasonable accuracy, we must have some phase space overlap between the state of interest and the simulated state. For a single type of schedule, we can reparameterize pathways to maximize phase space overlap, as will be shown below, essentially always guaranteeing good phase space overlap as long as our initial simulations also had good phase space overlap.

We know that the soft core 1−1−48 potential has near optimal variance in the space of all pairwise potentials for the disappearance of Lennard-Jones particles[14] and that the two-body interactions that create excluded volume contribute the most to the variance of particle insertion. We therefore examine eq 1 and its zeroth-order RDF in Figure 3 to see if we can identify basis functions and alchemical switches that are near optimal.

Figure 3 shows that the potential energy has two very distinct types of dependence on λ depending on the distance between particles. From the minimum of the potential and larger r, we observe that the potential is linear in λ, but at short-range, we observe a more complicated relation. This decomposition is also clear when directly examining eq 1 and the ratio $(r/\sigma_{ij})^c$. When this ratio is greater than 1, the λ dependence is only significant in the prefactor. When r/σ_{ij} is less than 1, the denominator is controlled by λ resulting in a more complex equation. We also observe that when $r/\sigma_{ij} < 1$, the RDF and the potential of 1−1−48 are mostly insensitive to r.

The two distinct regions of λ dependence suggest two alchemical switch and basis function pairs, one for $r < \sigma_{ij}$ and another for $r > \sigma_{ij}$. The two basis functions naturally come from

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

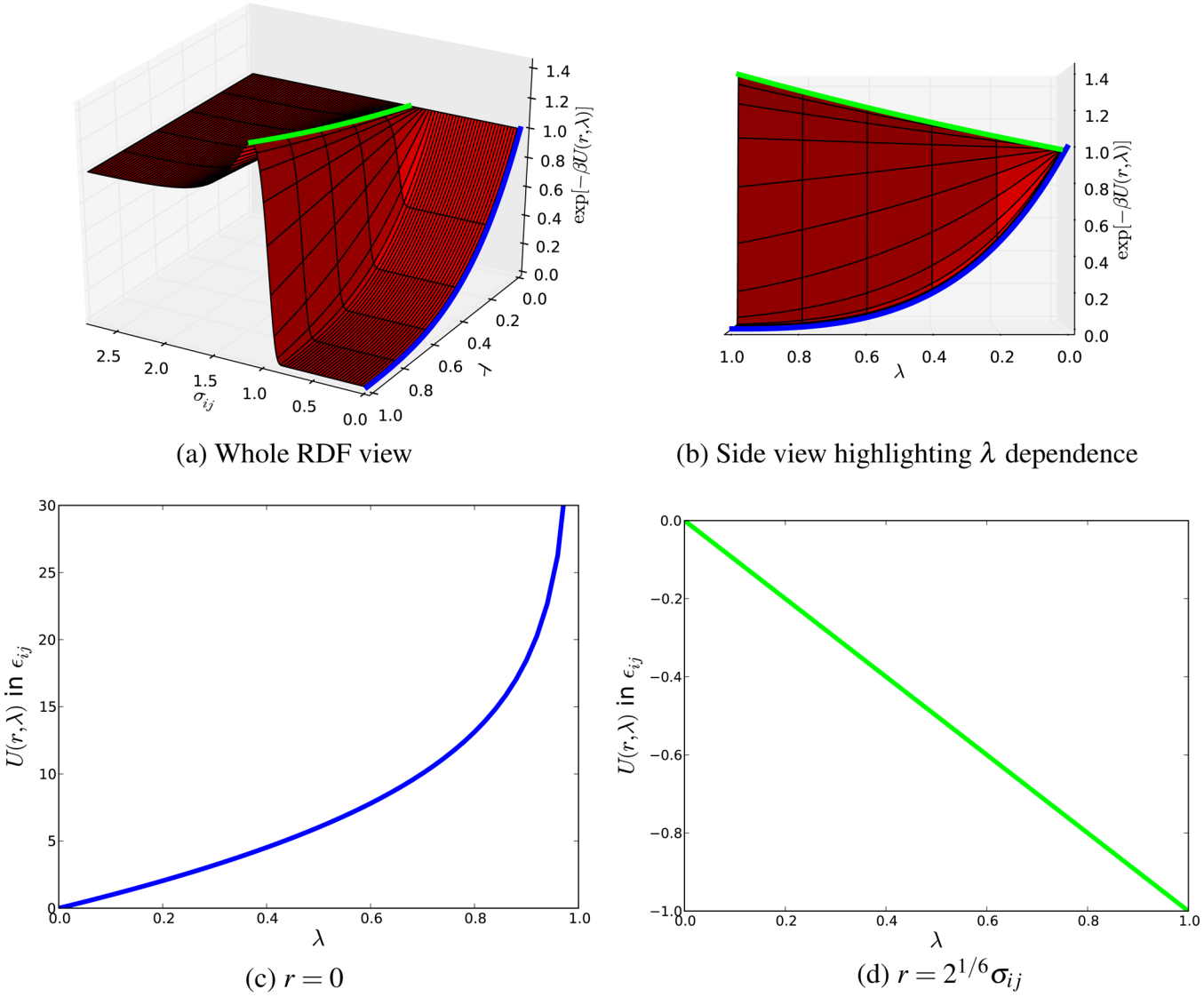


Figure 3. There are two distinct dependencies on λ for soft core 1−1−48 that suggest two basis functions and switches describing low variance paths for removing or inserting Lennard-Jones interactions. Shown are the predicted zeroth-order radial distribution function (RDF, parts a and b) and potential energy at constant r (c and d) along the soft core "1−1−48" pathway for alchemically appearing a united atom (UA) methane site in TIP3P water. The colored lines highlight λ dependence at $r = 2^{1/6}\sigma_{ij}$ for the green line and $r = 0$ for the blue line. In a region where $r/\sigma_{ij} < 1$ (c) the potential increases rapidly approaching infinity since soft core models converge to Lennard-Jones at $\lambda = 1$. When $r/\sigma_{ij} > 1$ (d), there is a linear dependence on λ.

the WCA decomposition, as this decomposition closely matches the behavior in the two regions seen in the 1−1−48 potential. We would prefer simpler alchemical switches for these two basis functions to simplify the overall calculation. Since the long-range dependence seen in Figure 3 is linear in the soft core 1−1−48 potential, a natural choice for the basis derived for the attractive component of the WCA decomposition is simply $h_i(\lambda) = \lambda$.

The repulsive basis must be treated differently, as the behavior of the plateau region in the 1−1−48 potential, shown in Figure 3 as a blue line, is a more complicated function of λ. Additionally, the potential goes to infinity in the limit of $\lambda \to 1$, so we cannot match this potential with a simple alchemical switch as we did the linearly scaled attractive interactions. Considering the physical constraints of the problem, we can derive an approximate formula for the dependence on λ of a low variance pathway as

$$h_i(\lambda) = \frac{K^\lambda - 1}{K - 1} \tag{18}$$

where K is a positive optimization constant greater than 1 (see Appendix B for derivation). Additional trial and error exploration of the inequalities derived in Appendix B yields a second possible alchemical switch:

$$h_i(\lambda) = \lambda\left[p + (1-p)\exp\left(-\left(\frac{1-\lambda}{s}\right)^2\right)\right] \tag{19}$$

where $0 \le p \le 1$, s is positive, and both are parameters that can be adjusted to minimize the variance. This switch was designed to approximate the short-range 1−1−48 dependence on λ by being predominately linear in λ at small values, and transition to a mostly Gaussian shape at larger λ. The first switch, eq 18, will be referred to as "switch A" and the second switch, eq 19, will be labeled "switch B." Many other equations are possible, but

having near optimal switches allows us to start to explore the linear basis pathways. We emphasize that in the end, we will be able to derive the optimal choice of switching function, without the RDF or any approximation to it, and these functions are simply starting points in the process.

3. SIMULATION METHODS

Molecular dynamics simulations of united atom (UA) methane, UA anthracene, and all-atom (AA) 3-methylindole were carried out with YANK[31,32] which was built on GPU accelerated OpenMM v4.1.1[33−36] in explicit TIP3P water. OpenMM allows for rapid and simple deployment of arbitrary nonbonded potentials, making it straightforward to implement arbitrary basis functions. YANK provides the capability to do alchemical solvation free energies as well as Hamiltonian replica exchange to improve sampling[37] with modifications to improve computational efficiency.[38,39]

Molecular input files were constructed using AMBERTOOLS's LEaP[27] with OPLS-AA force field parameters for all atoms except UA anthracene, where parameters were taken from Pitera and van Gunsteren.[40] Starting molecular geometries were acquired from the Supporting Information from Mobley et al.[41] and the molecules were solvated in a periodic cubic box of TIP3P water with boundaries 1.2 nm from the solute in keeping with the source's setup. This lead to 620, 874, and 961 water molecules for UA methane, UA anthracene, and AA 3-methylindole, respectively.

Two additional molecules which were not part of Pham's test set[13] were also tested. AA *n*-decane and a Lennard-Jones (LJ) sphere with a larger radius than Pham evaluated were tested. These two molecules allow validating the basis function approach on more asymmetric as well as larger molecular shapes. The *n*-decane was generated using the same steps as AA 3-methylindole. The large LJ sphere's σ_i was chosen to be 5.23 Å, the radius of buckminsterfullerene (C_{60}); ε_{ii} was chosen the same as UA methane, with the mass for molecular dynamics set to that of C_{60}. Mixing rules were kept the same as the other molecules in AMBERTOOLS such that $\sigma_{ij} = 0.5(\sigma_{ii} + \sigma_{jj})$ and $\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$. Because no previous data are available for variances of solvating molecules with these parameters, we can only compare variances between switches and not to soft core pathways.

All molecules in all pathways except 1−1−48 decoupled their repulsive and attractive forces together with sampling at YANK's default values of $\lambda_{\text{repulsive}} = \lambda_{\text{attractive}} = \{1.0,0.95,0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1,0.0\}$. $(r/\sigma_{ij})^{48}$ of eq 1 exceeds the numerical precision of the OpenMM kernel, which handles arbitrary potentials, so the variances were taken from earlier studies.[13] Because of this, direct comparisons of free energy difference and error in free energy between the 1−1−48 potential and the others will not be possible. However, variance can still be compared, as explained below.

Simulations were carried out under isothermal−isobaric (*NPT*) conditions at 298 K and 1 atm. Every state was simulated for 2 ns with a 2 fs time step, samples collected every 1 ps, and Hamiltonian replica exchange between all states attempted every 1 ps. Although replica exchange is not required for these systems, there is negligible penalty for performing with it in YANK. The potential energy of every state was collected as well as the potential evaluated for every state's configuration re-evaluated to all other state's potentials; this information was analyzed by MBAR[18] for evaluating free energy and expectation values. Errors in variance were obtained using 200 bootstrap samples.[42] 3-methylindole's and *n*-decane's bonded hydrogens were constrained by the SHAKE algorithm[43] and water was constrained by the SETTLE algorithm.[44] Pressure control was handled by a Monte Carlo barostat[45,46] and temperature control through Langevin dynamics. The nonbonded cutoff was 9 Å and interactions outside this cutoff were handled by PME with an error tolerance of 5×10^{-4}. Dispersion corrections to the free energy were not calculated;[47,48] however, they will only shift the free energy differences of a given molecule by a constant amount for all configurations at each value of λ, thus not affecting comparison of variance pathways.

4. RESULTS

We are interested in answering five main questions while examining these linear basis function potentials:

- Is the implementation performed correctly, and do the free energy differences of the linear basis potentials from simulations converge to the same value as using the 1−1−6 pathway?
- What are the variance minimizing parameters for switch A and switch B derived from the zeroth-order RDF?
- Do these new linear basis potentials have comparable or lower variance than the standard 1−1−6 and optimized 1−1−48 soft core potentials?
- Can we identify the set of alchemical switches with the lowest possible variance over all sets of alchemical switches given these basis functions?
- Are these low variance switches specific to the molecules they are developed with, or are they generalizable to other molecular transformations involving removal of Lennard-Jones sites?

4.1. Estimated Variances and Optimal Basis Function Parameters. UA methane is the simplest place to start for estimating optimal function parameters. These parameters can be tested with other molecules to determine if these parameters are sufficiently general. The zeroth-order RDF approximation is the most accurate in the case of UA methane, since there is a single interaction site and the molecule is isotropic. A comparison of the variance and $\langle \partial U/\partial\lambda \rangle$ using the zeroth-order RDF among switch A, switch B, soft core 1−1−6, soft core 1−1−48, and a pure linear decoupling of the UA methane from solvent is shown in Figure 4. The area under the curves gives total variance, the objective function we wish to minimize. The curves shown in Figure 4 were created with the optimized parameters for our proposed switches, and a list of the parameters with the total variances is shown in Table 1. The alchemical switches $h_i(\lambda)$ used in generating Figure 4 are shown in Figure 5 and compared to a scaled 1−1−48 at $r = 0$ to show similarities in shape.

Figure 5 suggests that nearly matching the 1−1−48 potential form at $\lambda < 0.5$ is a key factor in minimizing the variance. The predicted parameters were found to be insensitive to ε_{ij} and σ_{ij} for both switches, meaning the switches are likely to be applicable to a wide range of atom types.

Ideally, the basis functions should have lower overall statistical variance than the soft core 1−1−6 potential to be worth using. Previous work has shown[13] that the ordering of total variances of paths estimated with the zeroth-order RDF approximation is almost always the same as the ordering of total variance in the actual simulations carried out without approximation. According to the prediction, switch A and

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

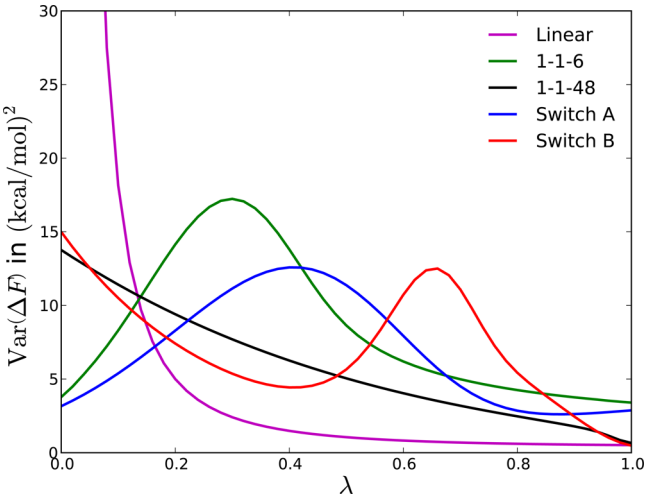


Figure 4. Predicted variances using a zeroth-order RDF approximation for insertion using linear basis function paths are lower than the common soft core 1−1−6 potential, and near the minimal variance soft core 1−1−48 potential. This optimization was carried out for inserting a united atom methane site in TIP3P water for two linear basis potentials, the two soft core potentials, and a purely linear transformation pathway. Optimized parameter(s) for switch A are K = 35 and for switch B are p = 0.22 and s = 0.284. This path only shows Lennard-Jones type interactions and assumes repulsive and attractive forces were appearing together. Total variance is found by integrating under these curves and variance for both switches falls between 1−1−6 and 1−1−48 soft core potentials.

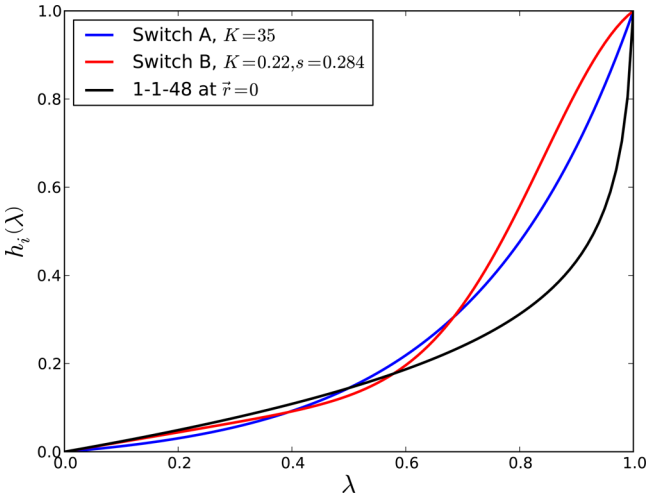


Figure 5. Alchemical switches matching soft core 1−1−48 for $\lambda < 0.5$ is an important key to minimizing total variance. Switch A and switch B are compared to the soft core 1−1−48 pathway at r = 0. The potential for 1−1−48 was normalized to its value at r = 0 and λ = 0.99 to compare the shape of 1−1−48 with the switches. All the curves have a near linear dependence at low λ and a more complex relation at larger λ. Constants for the curves came from minimizing the variance of the zeroth-order RDF approximation for a single UA methane interacting with TIP3P water on a schedule where repulsive and attractive interactions change together.

Table 1. Optimal Parameters and Variances from Zeroth-order RDF Predictions for Basis Functions, Soft Core Potentials, and Linear Alchemical Transformations[a]

alchemical path	parameters	variance
linear		NC
soft core 1−1−6	a = 1, b = 1, c = 6, α = 0.5	8.52
soft core 1−1−48	a = 1, b = 1, c = 48, α = 0.0025	5.84
switch A	K = 35	7.03
switch B	p = 0.22, s = 0.284	7.04

[a]Variances are in units of (kcal/mol)^2. Soft core potential parameters follow convention in eq 1, switch A's parameter is for eq 18, and switch B's parameters are for eq 19. The linear transformation has no parameterization, and its variance was large enough to be considered not converged (NC) by numeric integration. Transformation is for appearing a single united atom methane site in TIP3P water. Simulations of corresponding switch or soft core potentials were run with these parameters.

switch B should have lower variances (and therefore lower statistical error in free energy calculations) than soft core 1−1−6 but not quite as low as 1−1−48.

4.2. Variances and Free Energies from Simulations. *4.2.1. Validating against Previously Studied Molecules.* In Figure 6 we show the $\langle \partial U/\partial\lambda \rangle$ values and variances as a function of λ taken from simulations of the three test molecules for the switches discussed so far ("switch O" curves in Figure 6 will be discussed later). Free energies for the three molecules are reported in Table 2. The soft core variances have been replotted from the figures in Pham[13] as $\partial U/\partial\lambda$ information required for soft core variance is not collected by OpenMM and cannot be evaluated by eq 17. Both Pham's variance and the basis function variance are computed directly from the simulated data. The derivation of both approaches is shown in Appendix A.

The calculated variance is statistically identical between our data collected with OpenMM and Pham's data previously collected with GROMACS.[13,25,49] Differences in the $\langle \partial U/\partial\lambda \rangle$ curves in Figure 6a, c, and e are entirely due to the dispersion correction. The variances in Figure 6b, d, and f are independent of differences in the applied dispersion correction in the two simulations.[47,48]

The linear basis potentials and the linearly scaled variances were only sampled at twelve λ values. However, the properties were then estimated at 101 total λ points to construct the curves, producing smooth curves with low uncertainty. This immediately demonstrates another advantage of the linear basis potential approach: once we collect the values of the basis functions at some values of λ, we can estimate the derivatives and variance at *any* point along the curve without needing to recalculate any additional energy terms. We can make this calculation because we only require the basis function potential energies to calculate $\langle \partial U/\partial\lambda \rangle$ at any state, which can be done using eq 5 as long as the values of the basis functions at each sampled state are recorded. The Boltzmann factor of the configurations changes at each state but depends only on the total potential at that state and can be easily calculated from the computed basis functions outside of any inner loops. MBAR[18] was used to estimate the expectation at each state reweighted to the correct ensemble. The statistical error of these expectation calculations will depend on how close in phase space our sampled states are to the states we estimate the observables at, but we will see that, assuming we are comparing using a single schedule, lack of overlap is never a problem.

The linear basis function implementation gives free energies statistically consistent compared with the 1−1−6 pathway implemented in OpenMM. The ranking of the total variance of the different pathways (linear, 1−1−6, 1−1−48, switch A, and

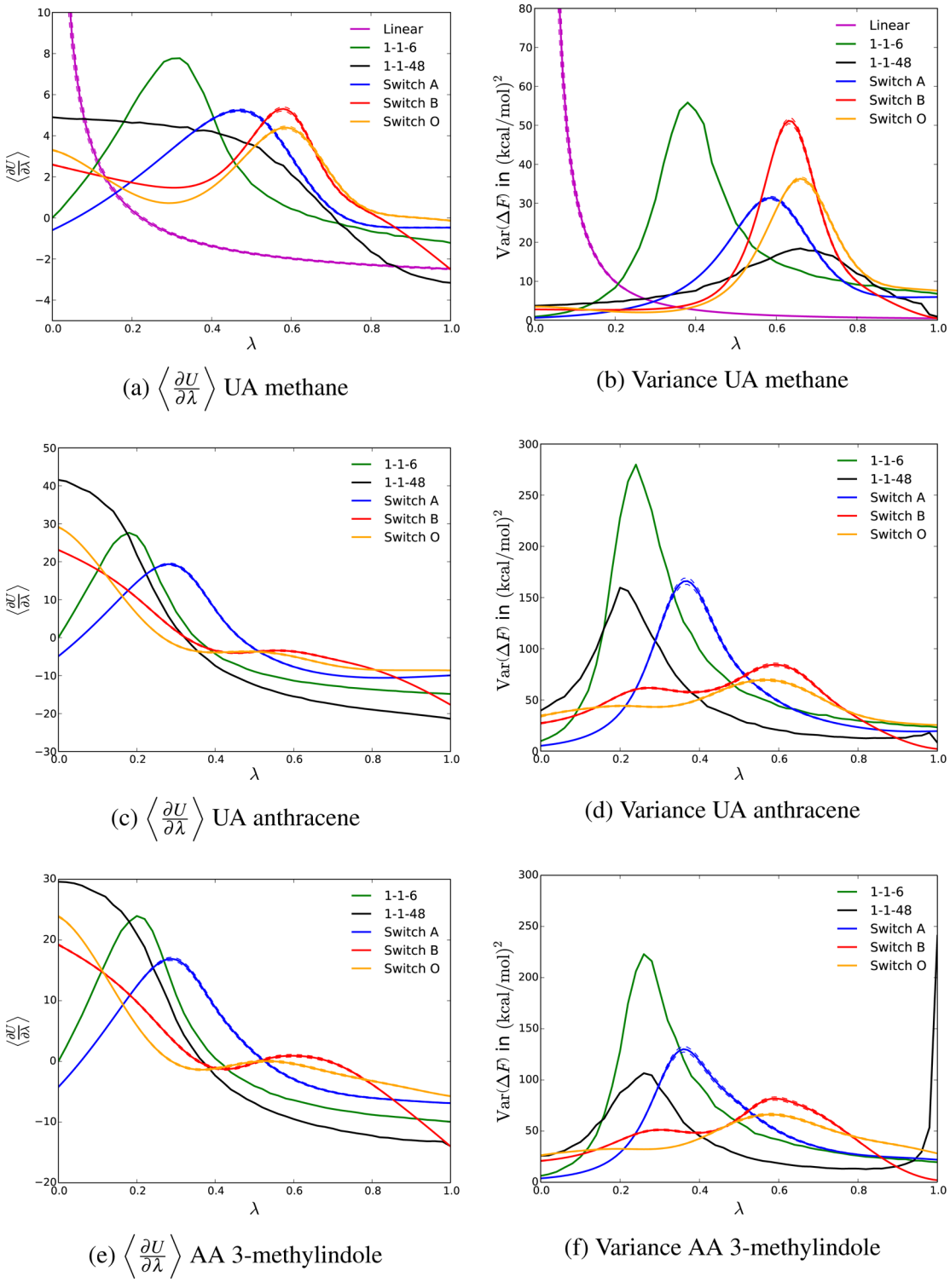


Figure 6. Linear basis function approach has variances lower than common soft core potentials and approach the lower limit among all pairwise potentials. This is demonstrated for multiple molecules by the $\langle \partial U/\partial \lambda \rangle$ and variance of alchemical solvation from simulation in TIP3P water for a series of different coupling pathways for Lennard-Jones decoupling. Errors are shown with dashed lines around the curves and were estimated by 200 bootstrap samples; error is often less than thickness of the curve. Soft core data taken from Pham.[13] Switch A and switch B have comparable or lower statistical error than soft core approaches, with the optimized switch O even lower. Reweighting using the linear basis approach allows the entire smooth curve to be calculated from only 10 to 12 sampled states. Note a strong qualitative similarity to predicted curves in Figure 4.

switch B) using the estimated zeroth-order RDF predictions compared to actual simulations are indeed the same for UA methane, though switches A and B are reversed for UA anthracene and AA 3-methylindole, larger molecules for which the approximation to the radial distribution function approach starts to break down. The zeroth-order RDF thus aids the design process of low variance initial optimal switches, which can later be refined with actual simulation data. Importantly, when performing the actual simulations, the proposed linear basis potential approaches still have lower statistical error and

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

Table 2. Simulated Free Energy of Solvation and Variance in TIP3P Water[a]

alchemical pathway	solvation free energy	variance
UA Methane		
linear	1.761 ± 0.114	NC
soft core 1−1−6	1.727 ± 0.028	16.93 ± 0.09†
soft core 1−1−48		9.03 ± 0.09†
switch A	1.764 ± 0.027	10.80 ± 0.17
switch B	1.742 ± 0.026	11.96 ± 0.19
switch O	1.748 ± 0.026	10.97 ± 0.16
LJ to capped	$-0.002 \pm 4 \times 10^{-4}$	$0.001 \pm 4 \times 10^{-4}$
UA Anthracene		
soft core 1−1−6	−0.277 ± 0.067	78.49 ± 0.53†
soft core 1−1−48		49.29 ± 0.53†
switch A	−0.275 ± 0.060	53.95 ± 1.02
switch B	−0.397 ± 0.056	48.57 ± 0.84
switch O	−0.334 ± 0.054	45.95 ± 0.81
LJ to capped	−0.021 ± 0.001	$0.008 \pm 8 \times 10^{-4}$
AA 3-methylindole		
soft core 1−1−6	1.790 ± 0.065	63.79 ± 0.40†
soft core 1−1−48		40.25 ± 0.40†
switch A	1.751 ± 0.061	48.09 ± 0.99
switch B	1.771 ± 0.055	44.66 ± 0.81
switch O	1.930 ± 0.050	42.92 ± 0.70
LJ to capped	−0.050 ± 0.004	0.26 ± 0.06

[a]Free energies are in kcal/mol and variances are in $(\text{kcal/mol})^2$. Free energies for 1−1−48 path not shown as data are taken from Pham[13] who ran with a Hamiltonian including a λ dependent constant offset resulting in a different overall free energy. However, the variance of the paths can be compared and serves as a target variance. Error estimates for free energy estimated by MBAR[18] and for variance by 200 bootstrap samples. "NC" denotes that the variance did not converge. † indicates data taken from Pham.[13] Statistical error of variance for † involved more samples, but it is the value of variance, not its uncertainty that determines the statistical error of simulations. "LJ to Capped" is the free energy of changing from capped to uncapped potential and *not* included in individual free energy of switches.

Table 3. Simulated Free Energy of Solvation and Variance for *n*-Decane and a Large Lennard-Jones (LJ) Sphere in TIP3P Water to Validate the Approach in the Cases of a Long, Non-hydrogen Bonding, Nonrigid Molecule and a Very Large Molecule[a]

alchemical pathway	solvation free energy	variance
AA *n*-decane		
soft core 1−1−6	4.409 ± 0.072	—
switch A	4.473 ± 0.068	68.56 ± 1.35
switch B	4.368 ± 0.069	63.84 ± 1.26
switch O	4.435 ± 0.060	60.94 ± 1.14
LJ to capped	−0.110 ± 0.005	0.560 ± 0.09
Large LJ Sphere		
soft core 1−1−6	16.127 ± 0.116	—
switch A	15.935 ± 0.103	65.70 ± 1.56
switch B	16.300 ± 0.102	64.53 ± 1.51
switch O	15.956 ± 0.089	59.80 ± 1.41
reoptimized switch O	16.043 ± 0.090†	45.74 ± 1.40†
LJ to capped	−0.006 ± 0.001	$0.007 \pm 3 \times 10^{-4}$

[a]Free energies are in kcal/mol and variances are in $(\text{kcal/mol})^2$. Error in variance was found by 200 bootstrap samples. The large LJ sphere was sampled with three extra intermediate states and every 4 ps instead of every 1 ps due to long correlation times at intermediate λ; † was *not* sampled at the three extra states to show how lower variance can reduce required samples to achieve a target statistical precision. — indicates that no data are available. "LJ to capped" is the free energy of changing from capped to uncapped potential and *not* included in individual free energy of switches.

variance than 1−1−6, though not as low as 1−1−48, as predicted by the zeroth-order RDF. The more precise results using simulation confirm that these alchemical switches can be just as statistically efficient as soft core methods.

The purely linear scaling path for the Lennard-Jones terms ($h(\lambda) = \lambda$) has large, unconverged variances and the largest relative free energy error, as anticipated. The linear transformation was simulated only for UA methane to validate the zeroth-order RDF prediction. Since the linear transformation will virtually always have higher errors and variances for appearing atomic sites in dense fluid,[9,21,22] it was not considered for the other molecules.

4.2.2. Validating against Large and Asymmetric Molecules. The simulated variances and free energies for *n*-decane and the large LJ sphere are shown in Table 3. Since no previous data are available for these molecules, the free energies and variances are shown for the basis function approaches, but only the free energy is shown for the soft core 1−1−6 pathway because the full $\langle \partial U/\partial\lambda \rangle$ curves cannot be generated.

The large LJ sphere was sampled with additional states at $\lambda = \{0.55,0.45,0.35\}$ due to large changes in the variance in that region and with samples taken every 4 ps instead of every 1 ps because correlation times were found to be greater than the original sampling rate of 1 ps. Because we are sampling less frequently to obtain uncorrelated samples, we have larger statistical error in the results, but all the free energies are statistically consistent.

4.2.3. How Significantly Do Capped Potentials Affect the Thermodynamics of the End States? We must examine whether we can use our fully coupled capped end states with transition window between $0.8\sigma_{ij}$ to $0.9\sigma_{ij}$ as the end state of the transformation, or if there are remaining errors. For some molecules, the free energy difference is negligible, less than 0.01 kcal/mol, but for other larger molecules, it is as large as 0.1 kcal/mol. For improved robustness, we thus also recommend including a last step of changing from the capped potential to the full repulsive potential.

Calculating the free energy difference between our initial choice of the capped potential and uncapped potential is statistically efficient. The variance and free energy of turning on the cap are reported in Tables 2 and 3 in the "LJ to capped" entry with a switching region of $0.8\sigma_{ij}$ to $0.9\sigma_{ij}$. Because this transition window is defined in terms of σ_{ij}, the width and height of the cap will depend on the molecular parameters. We thus must compare transition windows by examining the range of cap magnitudes. Interactions with small atoms such as hydrogen will have small caps while larger atoms such as carbon and oxygen will have larger caps based on ε_{ij}. This transition window provided caps between $3.5k_BT$ and $8.8k_BT$ for our systems with an average of $6.2k_BT$. We find that the variance of this final transformation step contributes a negligible amount to the total uncertainty over all molecules, as it is at least 2 orders of magnitude smaller than the insertion. The most efficient pathway guaranteeing correct thermodynamics involves a basis function with a capped potential with switch between $0.8\sigma_{ij}$ and $0.9\sigma_{ij}$ and then changing linearly to the full Lennard-Jones potential.

Entirely eliminating the extra step of switching from a full potential to a capped function would require a repulsive basis

potential with a larger cap, which requires a transition starting at $r < 0.8\sigma_{ij}$. For results that are truly independent of the cap, we would like the free energy difference between the capped potential and uncapped potential at least 1 order of magnitude less than the statistical error of the full solvation free energy. The higher value the cap, the closer the capped WCA potential comes to the unmodified potential, preventing solute atoms from "leaking" into the capped region of the potential. For the molecules studied here, the transition region which provided a statistically identical free energy was between $0.70\sigma_{ij}$ and $0.75\sigma_{ij}$. This provided a cap between 20 and $50k_BT$ with an average of $36.7k_BT$ over all atom types for our systems.

There is a moderate loss in statistical efficiency when using a harder cap over the entire transformation. Simulating with parameters for switch A of $K = 165.5$ and switch B of $p = 0.064$, $s = 0.31$, which have been reoptimized for this harder cap, we found that using this cap increased statistical uncertainty in free energy estimates by 3% for UA methane but up to 41% for the large LJ sphere. Essentially, disappearing an average cap of $\approx 35k_BT$ is too similar to a linear decoupling of an uncapped Lennard-Jones potential.

If harder caps than our initial guess yield larger variance, could we use softer caps than our initial guess as long as the last step coupling to the full potential is retained? Softer caps here are defined by a transition region that starts at $r > 0.8\sigma_{ij}$. However, we find that using a smaller potential energy cap can introduce large errors as the cap is insufficiently repulsive. Figure 7 shows several choices of soft transition regions for 3-methylindole. Very large variances near $\lambda = 0$ start to appear with a very small adjustment to the window spacing, in many cases becoming a significant percentage of the total statistical uncertainty. The average magnitude of the cap at $0.85\sigma_{ij}$ is $2.5k_BT$ and is low enough to have very little phase space overlap with a fully coupled Lennard-Jones potential, which causes the large increase in variance. Softer caps should therefore be avoided as such low caps can cause diverging variances, making any optimization of switch severely error prone. We would need an *additional* capping basis function to bridge this phase space uncertainty, which adds far too much complexity, given that we are already close to the provably minimum variance pathway for most molecules without this additional basis function.

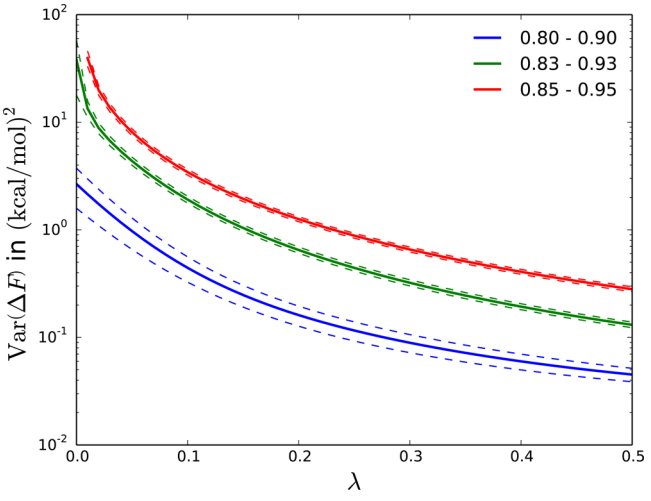


Figure 7. Variances and uncertainty are large if the cap on a potential function is too soft. This is shown here with the variance of linearly transforming from capped to uncapped potential with different transition regions for 3-methylindole. Curves labeled based on fraction of σ_{ij} of the transition region in a capped WCA potential. The variance is shown on a logarithmic scale to better visualize the rapid increase in variance from small changes in the transition region. The variance of the transition region between "0.85−0.95" (red) did not converge for $\lambda < 0.01$ and is truncated. Removing the singularity at $\lambda = 0$ results in more statistical error as the cap becomes softer. Errors estimated by 200 bootstrap samples.

We conclude that the error introduced from disappearing the positive singularity from the fully coupled state must be balanced between bringing the infinite potential to a capped one and decoupling the capped potential from the system. We find that this balance is achieved with a transition region of $0.8\sigma_{ij}$ to $0.9\sigma_{ij}$ with the mean cap of $6.2k_BT$ with respect to atomic pairs in the systems. This choice of capped potential we use for the reminder of the studies.

4.3. Variances and $\langle \partial U/\partial\lambda \rangle$ Predicted over Alternate Paths from a Single Simulation. It is possible to predict the expectations along any pathway, including the total statistical variance, with any arbitrary set of switches, $\mathbf{g}(\lambda)$, given only the data from a single simulation with a known set of monotonic switches, $\mathbf{h}(\lambda)$. A single simulation can therefore be used to find the optimal alchemical switch over the space of all possible $\mathbf{h}(\lambda)$ functions using data from only a single initial simulation. In fact, we did not necessarily need to identify two putative low variance switching functions. Once we performed simulations with a single linear basis potential, we can recalculate all observables of interest with any other alchemical switching function using the same basis functions in postprocessing. If the simulated and predicted simulations involve different alchemical schedules, then our ability to predict the variance of the proposed pathway will depend on the phase space overlap. If the trial simulation involved the same alchemical schedule as the initial simulation, however, then we can always guarantee that we will have good sampling for the trial simulation if the initial simulation also had good sampling. This optimization is possible because all of the thermodynamic information for any set of switches is contained in every other set of switches. We examine two example switches in Figure 8 to demonstrate this logic.

Consider a single basis function simulated with switch $h(\lambda_h)$ at a set of λ values λ_h; also consider an unsampled switch $g(\lambda_g)$ with a different set of λ values λ_g. Because the basis function is independent of the alchemical switch, the potential of a fixed configuration is identical when $h(\lambda_h) = g(\lambda_g)$ as seen in the tie lines of Figure 8. Generalizing the equality to multiple basis functions and alchemical switches, potentials will be the same for fixed basis functions when $\mathbf{h}(\lambda_h) = \mathbf{g}(\lambda_g)$.

Examining eq 16, we must have a map of the $\mathrm{Cov}(\mathbf{u},\mathbf{u})$ of the basis potentials sampled from $\mathbf{h}$ to the covariance computed at the unsampled states $\mathbf{g}$. Expectations and covariances will be dependent explicitly on $\mathbf{h}$ and only implicitly on λ, so the covariance can be denoted as a function of $\mathbf{h}$ as $\mathrm{Cov}(\mathbf{u},\mathbf{u})(\mathbf{h})$. The shared domain of $\mathbf{h}$ and $\mathbf{g}$ means

$$\mathrm{Cov}(\mathbf{u}, \mathbf{u})(\mathbf{h}) = \mathrm{Cov}(\mathbf{u}, \mathbf{u})(\mathbf{g}) \quad \text{if} \quad \mathbf{h} = \mathbf{g} \tag{20}$$

which is the required map. We extended this map to be explicitly defined by λ_h and λ_g since sampling is generally done by explicit statement of λ_h rather than $\mathbf{h}$. The inverse of the components of $\mathbf{h}$ must be evaluated to determine the λ_h corresponding to the state defined by the arbitrary λ_g. This inverse is *not* a matrix inverse, but a component-wise functional inverse, and will be denoted as $\mathbf{h}^{-1}(\lambda) =$

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

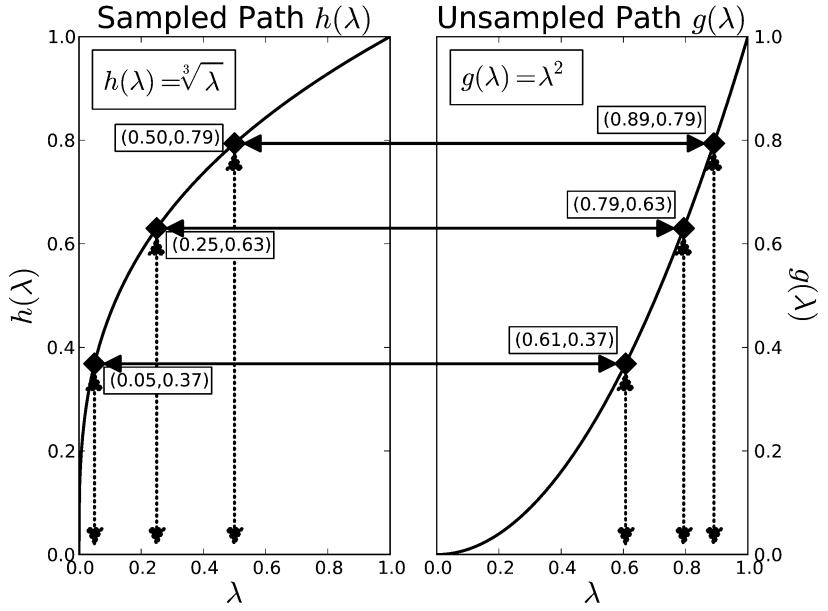


Figure 8. Thermodynamic information about a pathway using an arbitrary alchemical switch is contained within any sampled switch. An example of how to map from a sampled pathway (left) to an unsampled pathway (right) is shown. Monotonic switches both sample the [0,1] range and scale the potential energy over the $\lambda = [0,1]$ domain. As an example, any thermodynamic property, X, of the unsampled path, X_g at the state $\lambda = 0.61$ can be found by evaluating the sampled pathway's X_h at its $\lambda = 0.05$ state; similarly, $X_g(g(\lambda = 0.79)) = X_h(h(\lambda = 0.25))$ and $X_g(g(\lambda = 0.89)) = X_h(h(\lambda = 0.50))$. This mapping is not affected by properties that require derivatives since the derivative with respect to λ do not participate in evaluating expectation values such as the variance. The explicit mapping here is $h = g^6$, but a similar invertible mapping exists for other paths even if not analytically calculable.

$[h_1^{-1}(\lambda), h_2^{-1}(\lambda), \ldots, h_n^{-1}(\lambda)]$; we will assume that $\mathbf{h}^{-1}(\cdot)$ is a complete set of λ instead of writing $\lambda_h = \mathbf{h}^{-1}(\cdot)$ to conserve space in the equations. We can now write the variance of the unsampled path as

$$\begin{aligned}\mathrm{Var}(\Delta F)(\lambda_g) &= \mathbf{g}'(\lambda_g)\cdot\mathrm{Cov}(\mathbf{u},\mathbf{u})(\lambda_g)\cdot\mathbf{g}'^{T}(\lambda_g)\\ &= \mathbf{g}'(\lambda_g)\cdot\mathrm{Cov}(\mathbf{u},\mathbf{u})(\mathbf{h}^{-1}(\mathbf{g}(\lambda_g)))\cdot\mathbf{g}'^{T}(\lambda_g)\end{aligned} \tag{21}$$

where $\mathrm{Cov}(\mathbf{u},\mathbf{u})(\lambda_g)$ is the covariance matrix of the unsampled path and $\mathrm{Cov}(\mathbf{u},\mathbf{u})(\mathbf{h}^{-1}(\cdot))$ is the covariance matrix of the sampled path.

Equation 21 is an important result of this study, as an optimization routine can be written around it that can run entirely in postprocessing and only needs a single simulation's worth of data, not a series of iterative simulations with changing parameters.

A consequence of the linear combination is that the covariance is only explicitly dependent on $\mathbf{h}(\lambda_h)$ and not on $\mathbf{h}'(\lambda_h)$. If the covariance depended explicitly on $\mathbf{h}'(\lambda_h)$ instead, then no map could be made since the domain of $\mathbf{h}'(\lambda_h)$ is not necessarily that of $\mathbf{g}'(\lambda_g)$ and no one-to-one map may exist. $\langle \partial U/\partial\lambda \rangle$ along the alchemical path defined by the set of switches $\mathbf{g}$ can be predicted from eq 5 to get

$$\left\langle \frac{\mathrm{d}U}{\mathrm{d}\lambda} \right\rangle = \mathbf{g}_\lambda{}'\cdot\langle \mathbf{u}^T \rangle(\mathbf{h}^{-1}(\mathbf{g}(\lambda_g))) \tag{22}$$

Predicting the variance of unsampled alchemical switches has some complications. The first inherent problem with this prediction is that the λ_h found by the inverse may not be exactly any of the sampled λ_h. However, the expectation value of the potential energy at this unsampled λ_h can easily be estimated as was done to estimate the intermediate λ points of Figure 6. Assuming that there was reasonable overlap between the sampled λ_h states, then any interpolated values between these will also have good overlap. The second limitation is that the basis functions in $\mathbf{u}(r)$ must be known. One simulation must therefore be run for each unique molecular system before predictions can be made about other alchemical switches; we cannot make exact estimates from simulations of a different molecule. This prediction method clearly takes advantage of the properties of linear basis potentials, meaning variances from a soft core potential such as 1−1−48 at arbitrary points cannot be predicted by this method. Soft core potentials instead require separate evaluations of potential energies or derivatives of the potential at each individual configuration sampled.

The predicted covariance matrix will depend on each alchemical switch described by the alchemical path. The total potential affects the Boltzmann weight, which affects the covariance of predicted path. Each switch must be evaluated at the predicted states, even if it is not being varied. Fixing the alchemical schedule will help determine how unmodified switches are changing with respect to the switch being predicted.

As a demonstration of the power of this approach, we can see that both the variances and $\langle \partial U/\partial\lambda \rangle$ of switch A can be predicted with high accuracy using data from the switch B simulation and vice versa along the entire range of λ, shown in Figure 9. Although switch B does not have an analytical inverse, it is monotonic on the domain so its inverse was found numerically. Clearly, all deviations between the predictions and the validation simulations are very small and are within statistical noise.

We can perform simple optimizations on the parameters of switch A and switch B as an intermediate step in the process of optimizing potentials over all possible alchemical switching functions. We found that the lowest variance of "switch A form" potentials is obtained with $K = 10.8$ when predicting with data

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

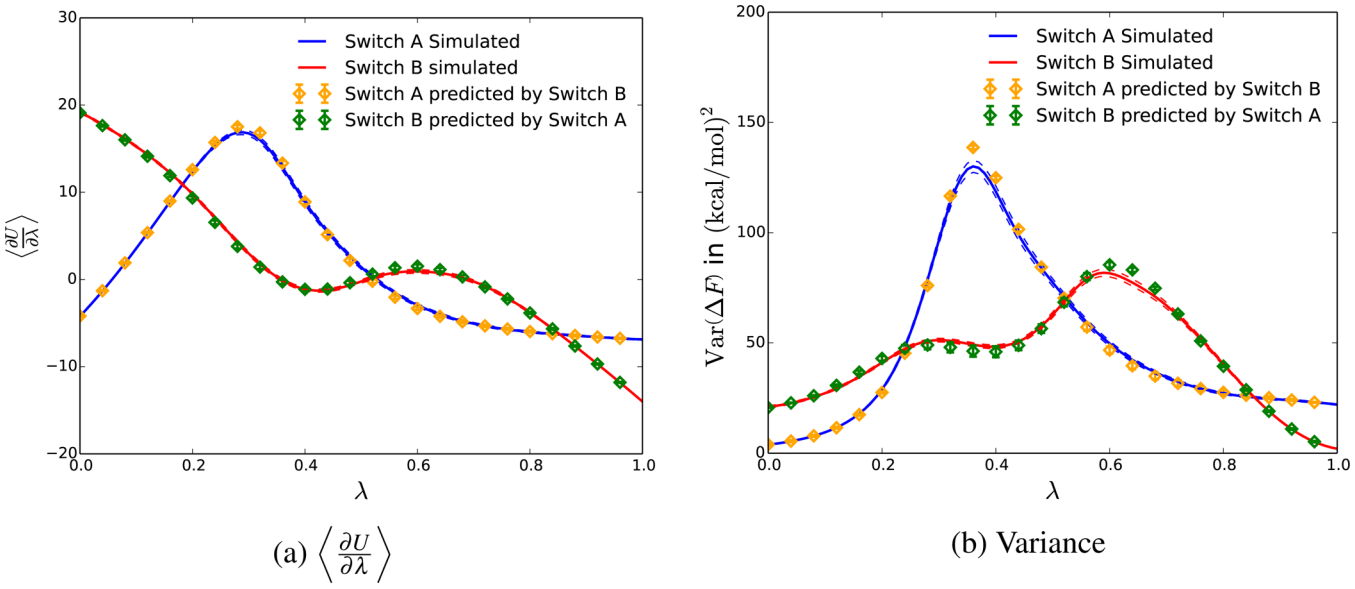


Figure 9. $\langle \partial U/\partial \lambda \rangle$ and variance of an alchemical switch can be accurately predicted from data collected exclusively from another switch. In this case, the properties of switch A and switch B for solvating 3-methylindole are estimated from data taken using the other switch. Error bars for simulation shown as dashed lines around the solid, error for prediction shown as capped vertical bars from the diamond; error is often less than thickness of the curve or diamond. The figures show that variance and $\langle \partial U/\partial \lambda \rangle$ can be predicted with extremely high accuracy from only one set of simulation data, implying that an optimal switch can easily be designed in postprocessing. Only a limited number of prediction values are shown, but they can be generated at any intermediate, just as simulated values can be.

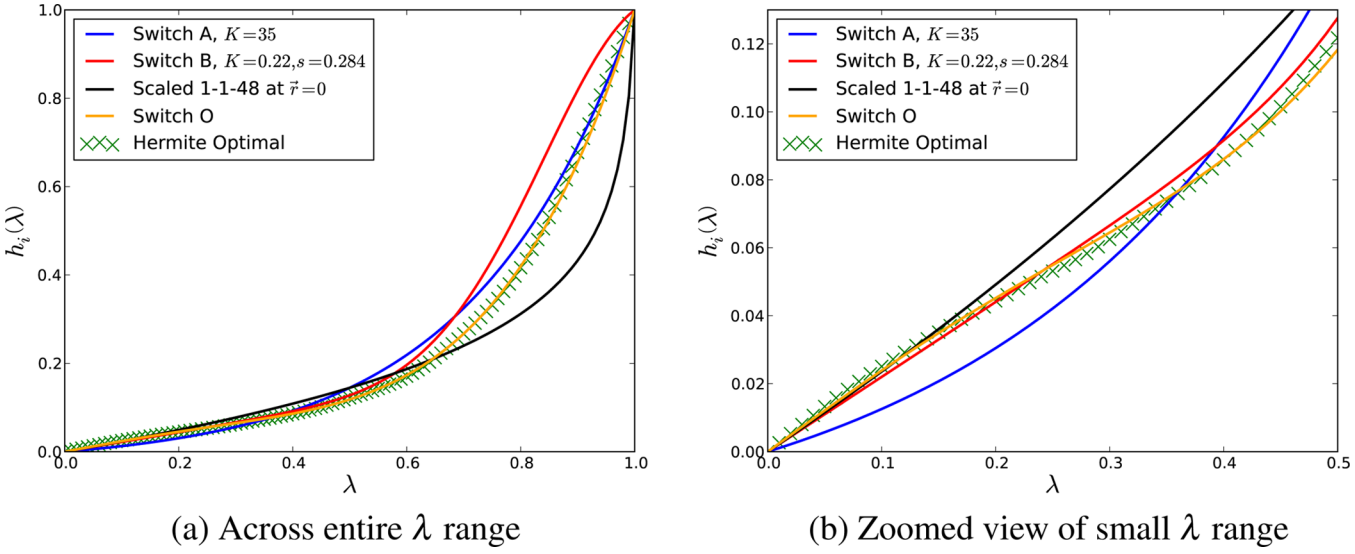


Figure 10. Largest impact to the overall variance is determined by the slope of the alchemical switch when $\lambda < 0.5$. The alchemical switching functions and the optimal Hermite spline compared to soft core 1−1−48 (normalized by the potential at $r = 0$ and $\lambda = 0.99$) are shown. (b) Zoomed in view highlights how the curves differ at small λ. This optimized switch is the best possible given the set of basis functions and fixed alchemical schedule. Different schedules with other basis functions would be required to lower the variance further.

from switch A's simulation, which was performed with the optimized value from the zero-RDF approximation, $K = 35$. When attempting to predict the minimum variance parameters for switch A using a simulation performed with switch B, we obtain minimum variance for switch A with $K \approx 10.9$, validating the optimization of one transformation pathway using data collected from another pathway.

4.3.1. Selecting the Optimal Alchemical Switch. We next examine the case of the lowest variance set of alchemical switch functions **h** for a given set of basis functions. Such functions must satisfy $h_i(0) = 0$, $h_i(1) = 1$, and be monotonic, but may be of arbitrary mathematically complexity, rather than a predefined form as eqs 18 or 19. For the purpose of optimization, we represent the family of possible switches with a monotonic spline.

An optimal $h_i(\lambda_i)$ for the repulsive basis function was created with a constrained optimization using the BY linear approximation (COBYLA)[50] routine to enforce monotonic spline knots using SciPy's[51] optimization module. Monotonic, cubic Hermite splines[52,53] were generated to enforce the monotonic interpolating splines between these knots. After this optimization, we then approximate the splined curve with a single best fit polynomial for implementation simplicity. We select the lowest order polynomial possible, provided that the

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

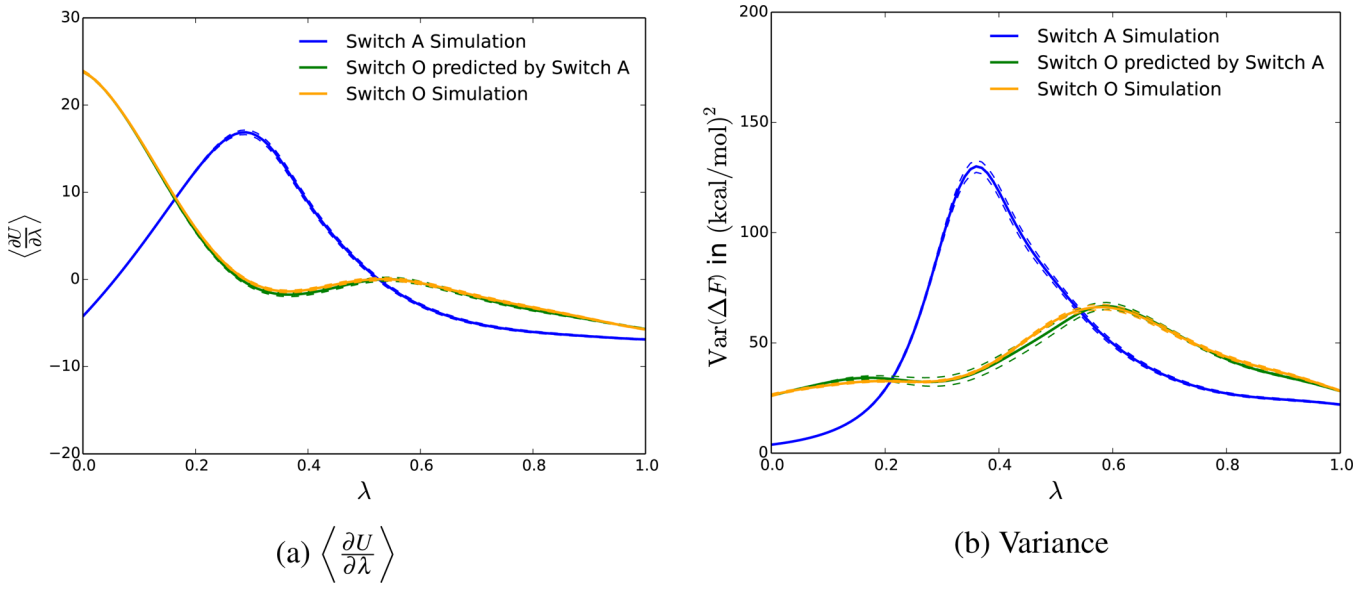


Figure 11. Thermodynamic properties of the optimal switch can be predicted with high accuracy as seen with the minimized variance switch, switch O, for 3-methylindole compared to switch A. Switch O was fit to a constrained quartic polynomial and found to have lower variance than other simulated alchemical switches. Error for each curve was found by 200 bootstrap samples and shown as dashed lines around the solid curve; error is often less than thickness of the curve. The predicted and simulated results are on top of each other, supporting the theory that an optimal switch can be developed from a single simulation.

variance between the spline and the polynomial differed by less than 0.5% and was monotonic on the [0,1] domain. We are able to accurately fit to our splined minimum variance switch with a quartic polynomial of the form

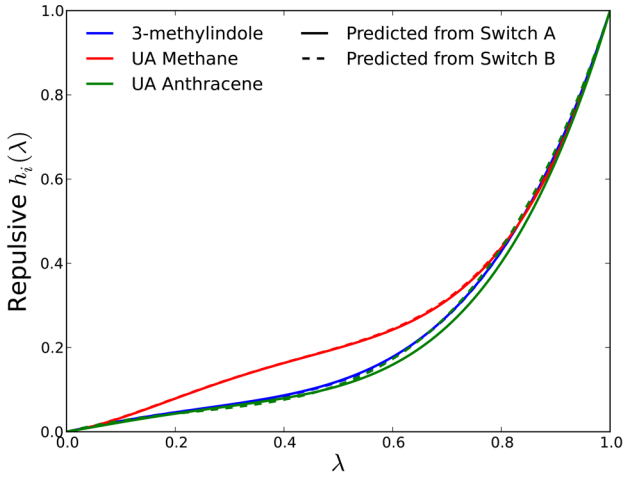
$$h_i(\lambda) = A\lambda^4 + B\lambda^3 + C\lambda^2 + (1 - A - B - C)\lambda \qquad (23)$$

with only three fitting terms because we enforce the conditions $h_i(0) = 0$ and $h_i(1) = 1$. The parameters for this optimized switch function, which we will call "switch O," are $A = 1.62$, $B = -0.889$, and $C = 0.0255$. For comparison, all three switches (A, B, and O) and the optimized Hermite spline are shown in Figure 10. These switch O parameters were found by taking the data from 3-methylindole's simulated switch A with $K = 35$ and optimizing to find the minimum variance switch, which is labeled in Figure 11 as "switch O predicted by switch A."

Results from this predicted switch O (using the switch A simulation) are in excellent agreement compared to observables computed from simulations performed using this switch O with 3-methylindole, labeled "switch O simulation" in Figure 11. The variance of the prediction is 42.53 ± 1.20 (kcal/mol)^2 and the simulated variance of 42.92 ± 0.70 (kcal/mol)^2. The free energy of solvation for this transformation using the optimized switch at 12 intermediate points is 1.930 ± 0.050 kcal/mol, which has the lowest statistical uncertainty of all the reported methods in the table, demonstrating it is indeed an optimal path. Variances and $\langle \partial U/\partial \lambda \rangle$ for switch O are plotted in Figure 6, and all free energies are shown in Table 2.

We can probe the shape of this alchemical pathway to understand more about which regions contribute most to the variance. We find that the region of $0 \leq \lambda \leq 0.5$ shown in Figure 10b, specifically the slope of the function in this region, has the largest impact on total variance. If the slope in this region is too large, then the variance will be significantly increased at that point in the path. If the slope is too low in this region, then a large slope is needed at $\lambda \geq 0.5$ to reach the $h_i(1) = 1$ condition, also resulting in a large overall variance. We find that the exact polynomial constants found from eq 23 are somewhat sensitive to the data set used to perform the optimization, but the shape of the switch itself is much more robust.

We can also test the robustness of the optimized alchemical switch with respect to change of molecule. Figure 12 shows the polynomial fits to the optimized curves found from data generated with switches A and B for all molecules in this study. 3-Methylindole and UA anthracene converge to approximately the same optimized curve with data sets from switches A and B. However, UA methane converges to a slightly different curve. The polynomial fit parameters shown in Figure 12 vary considerably with data set and initial guess points of the optimization, however as can be seen in the figure, the curves remain essentially the same for 3-methylindole and UA anthracene. The optimization also converges to nearly the same curve for the same molecule, regardless of starting point. The RMSD between switch A and switch B for UA methane, UA anthracene, and 3-methylindole in Figure 12 is 0.00097, 0.017, and 0.0055 respectively, and can be qualitatively seen in the overlapping dashed and solid lines for each molecule. Although the optimal alchemical switch for UA methane is different than the switches for the other molecules, the variance-optimized curves derived for the other molecules still give free energies with lower variance than either switch A or switch B for UA methane, as seen in Table 2. Additionally, we note in Figure 6 that all switches behave more robustly than the 1−1−48 switch for 3-methylindole, as there is no large change in the variance near $\lambda = 1$. Keeping the variance consistent across the entire range of molecular shapes is useful for purposes such as Hamiltonian replica exchange that strive to keep the exchange rate roughly constant across the transformation.[19] Indeed, in Table 2, we see that for UA anthracene, the total variance is slightly less than the 1−1−48 curve, which is possible since the 1−1−48 curve is minimal variance only for UA methane.



(a) Comparison of optimized alchemical switch found from different molecules and starting paths

Molecule	Predicting Switch	Coef. A	Coef. B	Coef. C	RMSD
UA Methane	Switch A	3.43	−4.35	1.72	0.0503
	Switch B	3.34	−4.19	1.64	0.0502
UA Anthracene	Switch A	1.86	−1.21	0.128	0.0107
	Switch B	0.796	0.562	−0.679	0.0092
AA 3-Methylindole	Switch A	1.62	−0.889	−0.0255	0.0000
	Switch B	1.28	−0.298	−0.275	0.0055

(b) Switch O coefficients and deviation from simulated Switch O for each optimization

Figure 12. Optimal switch's shape for different molecules is robust, even when the polynomial coefficients are not. A comparison of optimal fitting curves found starting from switch A and switch B for several test molecules. The root-mean-square deviation (RMSD) in dimensionless units is between each switch and the optimized switch O generated from the parameters from 3-methylindole and switch A, over 101 uniformly distributed points on each curve. Switches for the same molecules optimized from differing starting points are almost identical, and switches optimized from different molecules are very similar. Even though UA methane does not have the same optimal alchemical switch, there is little improvement in the variance when using this different curve.

For *n*-decane, switch O from the 3-methylindole data appears to be near optimal, though further optimization reduces the variance by about 7% from switch O. However, for the large LJ sphere (with buckyball radius) we find that switch O has a very large increase in the variance for λ between 0.4 and 0.6 (Figure 13), requiring additional sampling between at λ_i = {0.55,0.45,0.35} to obtain an accurate free energy. We applied the optimization routine to the large LJ sphere using the switch A simulation to obtain a reoptimized switch which significantly reduced the peak in and the total variance (also Figure 13). The reoptimized switch was then simulated *without* the extra sampling at λ_i = {0.55,0.45,0.35} and gave a simulated free energy of 16.043 ± 0.090 kcal/mol and a variance of 45.74 ± 1.40 $(\text{kcal/mol})^2$; this is approximately 25% lower variance compared to other results in Table 3. This reoptimized alchemical switch is flatter in the region λ = 0.4 to λ = 0.6 as seen in Figure 13c, meaning the flat cap in the large excluded volume region changes more slowly to maximize the phase space overlap between neighboring states. This extreme case of a C_{60}-sized sphere is a practical example of both how a single optimized switch (switch O) derived from a single example molecule is better than standard simulation methods and near optimal but also how easily even this pathway can be improved with a small amount of postprocessing.

In summary, although slightly improved paths can be determined for particularly extreme cases (such as very large or very small molecules), a single optimized switch for the repulsive core has nearly minimal variance for all molecules, with a range of sizes from methane to a buckyball, and asymmetries as large as *n*-decane. If required, optimization for individual molecules requires only a single initial simulation (using, for example, the general switch O) and then a simple reoptimization using only this simulation data.

5. DISCUSSION

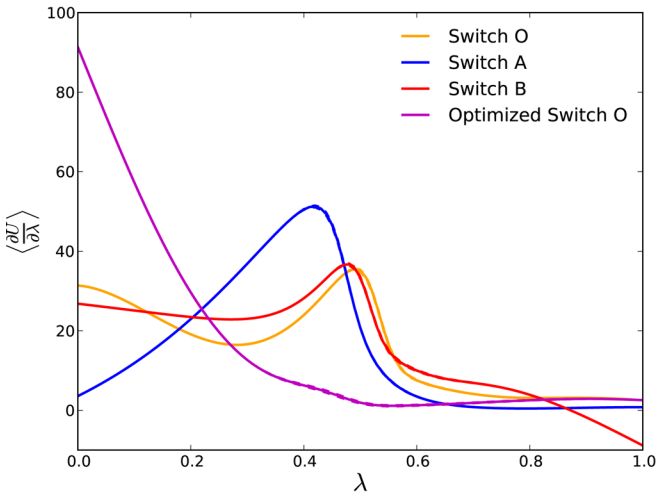
Our optimized linear basis potentials for all tested molecules have total variances, and thus statistical efficiencies between the total variances of the 1−1−6 and 1−1−48 soft core potentials, even for the initial guesses, switch A and switch B, as anticipated.

The most efficient possible alchemical switches are somewhat dependent on the molecular identity, though the switches optimized from single molecule are better for all tested molecules than the initial guessed switches or the 1−1−6 soft core potential. If the calculation is performed for a ligand−protein binding process, instead of solvation, the magnitude of the variance will change, but the relative efficiency of variance pathways will not change significantly. The ranges of λ with large variance will not significantly change along the same path since adding a protein to the solvent will not significantly change the density of the fluid around the ligand. We have illustrated this in Appendix C as an example, though an in-depth comparison of variances in both protein binding and solvation is beyond the scope of this paper.

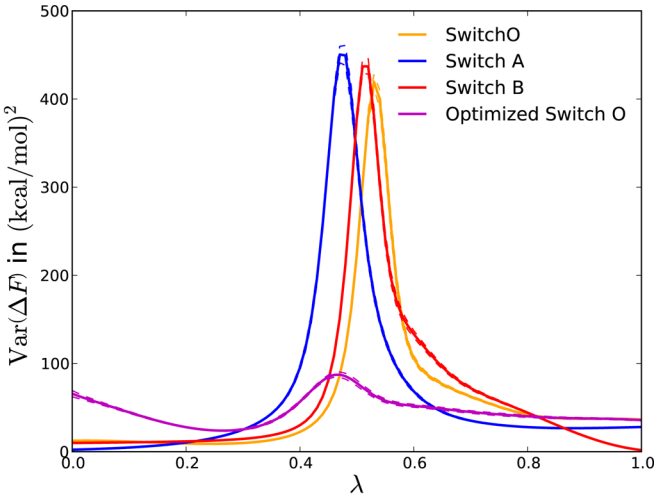
Our variance prediction equation, eq 21, is a powerful tool for optimizing paths of lower variances. Switch O has lower variance with all tested molecules compared to the soft core 1−1−6 pathway, although a reoptimized switch was needed to obtain the lowest possible variance for the large LJ sphere. The extra sampled states and longer correlation times for the large LJ sphere are not unexpected due to the excluded volume being more than 20 times larger than UA methane. Although switch O is slightly different than the minimum variance path in this extreme case, the simplicity of our variance minimization procedure allows a more optimal switch to be easily obtained. In this case, increasing the number of basis functions may allow an even lower variance path for turning off such a molecule and their optimal switches could easily be generated with the methods described in this study.

For general molecular transformations, we recommend a capped potential with a transition region of $0.8\sigma_{ij}$ to $0.9\sigma_{ij}$ and transforming from an uncapped, full Lennard-Jones potential to a capped potential in a single step. This cap with a typical value of $6.2k_BT$ provides a statistically efficient balance of contributions to the variance from these two steps. We find that harder and softer caps do not reduce statistical error over the entire transformation and should be avoided. Application-specific considerations, such as high temperature systems, may need different transition windows to adjust the caps for lower variance in those situations.

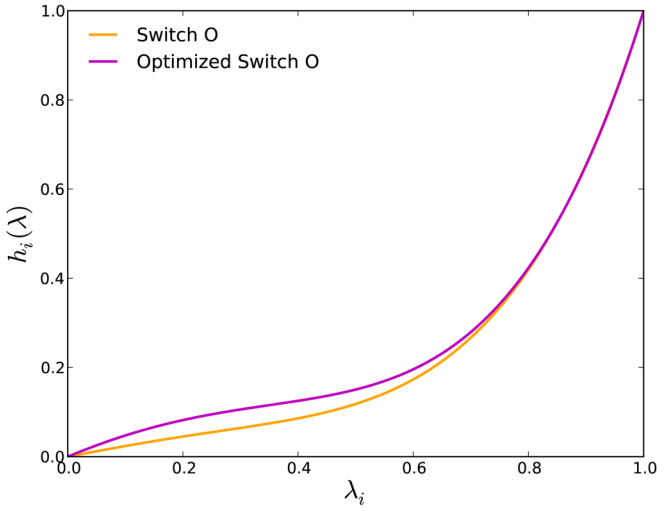
Expanding the variance optimization routine to a high number of basis functions simultaneously is a nontrivial matter. The number of terms in the covariance matrix quickly increases with multiple basis functions. This can cause the optimization routine to fail or be very prone to initial conditions as the covariance becomes less well behaved. These problems will be compounded if multiple switch optimizations are attempted at



(a) $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle$ Large LJ sphere



(b) Variance Large LJ sphere



(c) Comparison of original Switch O and re-optimized switch for large LJ sphere

Figure 13. Extreme peaks in variance can be reduced by a simple postprocessing reoptimization procedure. Here is shown $\langle \partial U/\partial \lambda \rangle$, variance, and the switches for a large Lennard-Jones sphere for switch A, switch B, switch O, and a reoptimized switch O specifically for this molecule. The reoptimized switch was created by predicting the optimal switch from sampled data of switch A. Having a nearly flat $h_i(\lambda)$ in the regions of large variance can drastically reduce the total variance as seen in part c. Errors are shown with dashed lines around the curves and were estimated by 200 bootstrap samples; error is often less than thickness of the curve.

once. For this reason, we do not recommend optimizing more than one or two switches at a time, depending on how many total basis functions there are, and explore more general optimizations in a separate paper.

5.1. Improvements and Implementation of the Basis Function Approach. Our method requires that alchemical switches be monotonic to make it possible to invert them. However, this is not a particularly limiting requirement. When considered as a path, the variance of an alchemical switch is directly proportional to the curvature of the switch and its thermodynamic length,[8,19,20] as is derived in Appendix A. Monotonic switches will have less curvature than an otherwise similar nonmonotonic switch, so intuitively should have lower variance. Additionally, any sequence of sampled λ states, which would require a nonmonotonic switch can be reordered to be monotonic. For example, if samples are taken in the sequence $h_i(\lambda) = \{0.4,0.5,0.3,0.6\}$, one could just as easily sample the sequence $h_i(\lambda) = \{0.3,0.4,0.5,0.6\}$ without loss of information; any repeated h_i value is equivalent to more samples at that state. For these reasons, assuming monotonic switches is much easier and is not a practical limiting factor in the approach.

Although the curves match very well with previous results of Pham and Shirts,[13] the total error in our variance estimations is somewhat larger, as can be seen in Table 2. This is because we sampled with fewer states using shorter simulations and would need about 2 to 6 times more sampling to reach the level of uncertainty obtained in the previous study. However, the uncertainties in our estimates of total variance are still small enough to reach the conclusions in this study.

Specific applications may require changing the basis functions themselves or simulation along other alchemical schedules. Such changes could involve different basis functions

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

instead of the capped WCA decomposition, adding one or more additional basis function terms in eq 4, or using a different schedule with repulsive and attractive terms turned off separately. For example, one could add a second capped repulsive term with a cap at $r > 0.8\sigma_{ij}$ and add a step to transform between different capped basis functions. This may improve phase space overlap as the core is softened but may also require more intermediate states. However, increasing the number of basis functions could lead to overfitting and increased complexity. Any increased complexity in the potential energy function may not be worthwhile to implement if there are not sufficient gains in statistical efficiency. In all of the wide variety of molecular shapes and sizes examined, the procedure presented here gave variances at most 25% larger of the previously found 1−1−48 minimum variance pathway[13,14] and is lower in one case. It is thus unlikely that the increase in complexity would be worth the marginal improvement.

Removing the free energy calculations from the force evaluation loop can make the simulation more easily adaptable to new software or computing paradigms. This linear basis approach makes this removal much easier. Free energy calculations require either the potential energy of configurations evaluated at other thermodynamic states or an analytic computation of the derivative of the Hamiltonian with respect to the coupling parameter. Traditional soft core potentials such as eq 1 require modifying the λ value and evaluating the potential and/or its derivative for each configuration in the inner loop. With the basis function approach, one would only have to evaluate each basis function, $u_i(r)$ in eq 4, once for a configuration and store them temporarily in memory. Evaluating the configuration's potential at other thermodynamic states can then be done by calculating all the $h_i(\lambda)$ of that state, then evaluate the products with the corresponding stored basis function and summing over all terms. If only the total potential is stored, the value of the basis functions can be solved through linear algebra by evaluating configurations at multiple states. $\partial U/\partial\lambda$ can also be computed external to the inner loop using the basis function energies along with knowledge of $\mathbf{h}'(\lambda)$.

6. CONCLUSIONS

This study develops an approach for designing pathways for statistically efficient free energy calculations involving removing or inserting molecules into dense fluids. We seek to retain functions which are easier to parallelize to new architectures than soft core potentials while still keeping the low variance of these approaches. We can achieve both of these objectives using the formalism of linear basis function potentials, consisting of linear combinations of basis functions with alchemical switches to represent the nonbonded potential energy function. The linear basis potential form is shown to greatly simplify the math needed to do potential energy re-evaluation or calculating derivatives of the potential, making it simple to estimate the potential energy at other states. This simplified math allows for potential energy re-evaluation to be carried out entirely in postprocessing. One generally applicable set of basis functions and three sets of alchemical switches, one of which is provably optimal for the choice of basis set, were presented as examples and were all shown to have lower total variance, and therefore lower statistical uncertainty, than the soft core 1−1−6 alchemical pathway for a diverse molecule set.

The simplicity and power of predicting variance and $\langle\partial U/\partial\lambda\rangle$ from the basis potential energies makes it possible to easily find an optimal alchemical switch such as switch O that is simple enough for general usage in other molecular simulation packages. The fact that only one simulation is needed for a given system to find the optimal switch is a large improvement over optimizing pathway parameters requiring multiple, iterative simulations. Indeed, the entire $\langle\partial U/\partial\lambda\rangle$ curve can be generated for arbitrary points along any other alchemical path as long as the original path is sampled sufficiently finely for the initial simulation, with 10−12 states generally enough.

From these studies, we can recommend a four basis function approach for decoupling small molecules from dense fluids such as water. The basis consists of electrostatic, capping, attractive, and repulsive basis functions. The electrostatic basis function is simply the standard electrostatic term and is turned off before the other three basis functions using a simple linear coupling, although the optimal switch for this transformation was not examined in this study. The potential is then transformed in a single step to a capped WCA-decomposed potential composed of the attractive and repulsive basis functions to avoid quickly disappearing a singularity at $r = 0$. These two terms are turned off simultaneously using a linear term for the attractive function, and the three-parameter quartic polynomial of switch O for the repulsive term, or a similarly shaped alchemical switch. Further improved pathways may be possible, but the procedure described in this study gives more efficient pathways compared to standard approaches currently used. Further analysis of the alchemical schedule and alternate ways to include electrostatics will be the focus of future studies.

APPENDIX A: DERIVATION OF LINEAR BASIS FUNCTION VARIANCE

In this section, we derive the variance of the free energy using the linear basis function approach with thermodynamic integration (TI) in the general case of multiple λ_i and multiple basis functions dependent only on coordinates as discussed in sections 2.3 and 4.3. This is an extension of the case with a single λ considered in previous work.[13]

Starting from the TI equation for the total free energy, with all λ_i parametrized by λ,

$$\Delta F = \int_0^1 \frac{\mathrm{d}F}{\mathrm{d}\lambda}\mathrm{d}\lambda = \int_0^1 \sum_{i=1} \frac{\partial F}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial \lambda}\mathrm{d}\lambda$$
$$= \int_0^1 \sum_{i=1} \left\langle \frac{\partial \mathcal{H}}{\partial \lambda_i} \right\rangle \frac{\partial \lambda_i}{\partial \lambda}\mathrm{d}\lambda$$

where $\mathcal{H}$ is the Hamiltonian and F is either the Gibbs or the Helmholtz free energy, depending on the ensemble sampled. At a single point λ along this path, we can write an estimator for $\mathrm{d}F/\mathrm{d}\lambda$ as

$$\frac{\mathrm{d}F}{\mathrm{d}\lambda} = \sum_{i=1} \left\langle \frac{\partial \mathcal{H}}{\partial \lambda_i} \right\rangle \frac{\partial \lambda_i}{\partial \lambda} = \left\langle \sum_{i=1} \frac{\partial \mathcal{H}}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial \lambda} \right\rangle \quad \text{(A.1)}$$

Since we can write the estimator of $\mathrm{d}F/\mathrm{d}\lambda$ in terms of an ensemble average of some quantity, we can write the variance as

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

$$\mathrm{Var}\left(\frac{\mathrm{d}F}{\mathrm{d}\lambda}\right) = \left\langle \left(\sum_{i=1}^{N} \frac{\partial \mathcal{H}}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial \lambda}\right)^2 \right\rangle - \left\langle \sum_{i=1} \frac{\partial \mathcal{H}}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial \lambda} \right\rangle^2$$

$$= \sum_{i,j=1} \mathrm{Cov}\left(\frac{\partial \mathcal{H}}{\partial \lambda_i}, \frac{\partial \mathcal{H}}{\partial \lambda_j}\right)\frac{\partial \lambda_i}{\partial \lambda}\frac{\partial \lambda_j}{\partial \lambda}$$

This can be compared to eq 5 of Crooks,[8] eq 2 of Shenfeld et al.,[19] and eq 42 of Gelman and Meng[20] to see that this total covariance is the square of the Riemannian metric used to measure thermodynamic length as

$$\mathcal{L} = \int_0^1 \sqrt{\mathrm{Var}\left(\frac{\mathrm{d}F}{\mathrm{d}\lambda}\right)}$$

To avoid cross correlation terms from expanding $(\partial \mathcal{H}/\partial \lambda_i)^2$ in terms of pair potentials such as $\sum_{k,l}(\partial u_k/\partial \lambda_i)(\partial u_l/\partial \lambda_i)$, we can rewrite the covariance in terms of pairwise functions. For ease of derivation, we first define the canonical partition function $Q(\lambda) = \int_\Gamma \exp(-\beta \mathcal{H}(\mathbf{x}, \lambda))\mathrm{d}\mathbf{x}$, where coupling variables λ_i are functions of the single parameter λ and we then note that

$$\begin{aligned}
\frac{\partial}{\partial \lambda_i} Q^{-1} &= -Q^{-2}\frac{\partial Q(\lambda)}{\partial \lambda_i} \\
&= -Q^{-2}\int_\Gamma \frac{\partial}{\partial \lambda_i}\exp(-\beta \mathcal{H}(\mathbf{x}, \lambda))dx \\
&= -Q^{-2}\int_\Gamma -\beta\frac{\partial \mathcal{H}(x, \lambda)}{\partial \lambda_i}\exp(-\beta \mathcal{H}(\mathbf{x}, \lambda))\mathrm{d}\mathbf{x} \\
&= \beta Q^{-1}\frac{\int_\Gamma \frac{\mathrm{d}\mathcal{H}(x,\lambda)}{\mathrm{d}\lambda_i}\exp(-\beta \mathcal{H}(\mathbf{x}, \lambda))\mathrm{d}\mathbf{x}}{Q} \\
&= \beta Q^{-1}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_i} \right\rangle
\end{aligned}$$

Although we use the canonical partition function Q, the results are equivalent in NPT or μVT ensembles as long as the paths only change the Hamiltonian, not the external thermodynamic parameters. We then examine the derivative $\frac{\partial}{\partial \lambda_i}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle$ to obtain the covariance

$$\frac{\partial}{\partial \lambda_i}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle = \frac{\partial}{\partial \lambda_i}\left(\int_\Gamma \frac{\partial \mathcal{H}(\mathbf{x}, \lambda)}{\partial \lambda_j}\exp[-\beta \mathcal{H}(\mathbf{x}, \lambda)]Q^{-1}\mathrm{d}\mathbf{x}\right)$$

$$\begin{aligned}
\frac{\partial}{\partial \lambda_i}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle &= \int_\Gamma \frac{\partial^2 \mathcal{H}}{\partial \lambda_i \partial \lambda_j}\exp[-\beta \mathcal{H}(\mathbf{x}, \lambda)]Q^{-1}\mathrm{d}\mathbf{x} \\
&\quad + \int_\Gamma -\beta\frac{\partial \mathcal{H}}{\partial \lambda_i}\frac{\partial \mathcal{H}}{\partial \lambda_j}\exp[-\beta \mathcal{H}(\mathbf{x}, \lambda)]Q^{-1}\mathrm{d}\mathbf{x} \\
&\quad + \int_\Gamma \beta\frac{\partial \mathcal{H}}{\partial \lambda_j}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_i} \right\rangle \exp[-\beta \mathcal{H}(\mathbf{x}, \lambda)]Q^{-1}\mathrm{d}\mathbf{x}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial \lambda_i}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle &= \left\langle \frac{\partial^2 \mathcal{H}}{\partial \lambda_i \partial \lambda_j} \right\rangle - \beta\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_i}\frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle \\
&\quad + \beta\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_i} \right\rangle\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle
\end{aligned}$$

This can then be arranged to form

$$\begin{aligned}
&\beta\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_i}\frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle - \beta\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_i} \right\rangle\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle \\
&\quad = \left\langle \frac{\partial^2 \mathcal{H}}{\partial \lambda_i \partial \lambda_j} \right\rangle - \frac{\partial}{\partial \lambda_i}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle
\end{aligned}$$

$$\mathrm{Cov}\left(\frac{\partial \mathcal{H}}{\partial \lambda_i}, \frac{\partial \mathcal{H}}{\partial \lambda_j}\right) = \beta^{-1}\left(\left\langle \frac{\partial^2 \mathcal{H}}{\partial \lambda_i \partial \lambda_j} \right\rangle - \frac{\partial}{\partial \lambda_i}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle\right)$$

Since $\mathrm{Cov}(x,y) = \mathrm{Cov}(y,x)$ and the partial derivatives of state functions are equal, we must also have

$$\frac{\partial}{\partial \lambda_i}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle = \frac{\partial}{\partial \lambda_j}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_i} \right\rangle$$

Therefore, the total variance for the calculation of ΔF, assuming equal sampling at each point along the path, is

$$\mathrm{Var}(\Delta F) = \int_0^1 \sum_{i,j} \mathrm{Cov}\left(\frac{\partial \mathcal{H}}{\partial \lambda_i}, \frac{\partial \mathcal{H}}{\partial \lambda_j}\right)\frac{\partial \lambda_i}{\partial \lambda}\frac{\partial \lambda_j}{\partial \lambda}\mathrm{d}\lambda \tag{A.2}$$

$$= \beta^{-1}\int_0^1 \sum_{i,j}\left(\left\langle \frac{\partial^2 \mathcal{H}}{\partial \lambda_i \partial \lambda_j} \right\rangle - \frac{\partial}{\partial \lambda_i}\left\langle \frac{\partial \mathcal{H}}{\partial \lambda_j} \right\rangle\right)\frac{\partial \lambda_i}{\partial \lambda}\frac{\partial \lambda_j}{\partial \lambda}\mathrm{d}\lambda \tag{A.3}$$

We will assume from here that the masses of all molecules are independent of λ. This makes the Hamiltonian's dependence on λ entirely through its potential energy, U, so $\langle \partial \mathcal{H}/\partial \lambda_i \rangle = \langle \partial U/\partial \lambda_i \rangle$ as the kinetic energy can be exactly accounted for analytically. The variance can then be rewritten in terms of the complete radial distribution function (RDF) (i.e., not approximated) as

$$\begin{aligned}
\mathrm{Var}(\Delta F) &= \beta^{-1}\int_0^1 \sum_{i,j}\left(\left\langle \frac{\partial^2 U}{\partial \lambda_i \partial \lambda_j} \right\rangle - \frac{\partial}{\partial \lambda_i}\left\langle \frac{\partial U}{\partial \lambda_j} \right\rangle\right)\frac{\partial \lambda_i}{\partial \lambda}\frac{\partial \lambda_j}{\partial \lambda}\mathrm{d}\lambda \\
&= 4\pi\rho\beta^{-1}\int_0^1 \int_0^\infty \sum_{i,j}\left[\frac{\partial^2 U}{\partial \lambda_i \partial \lambda_j}g(r, \lambda)r^2 \right. \\
&\quad \left. - \frac{\partial}{\partial \lambda_i}\left(\frac{\partial U}{\partial \lambda_j}g(r, \lambda)r^2\right)\right]\frac{\partial \lambda_i}{\partial \lambda}\frac{\partial \lambda_j}{\partial \lambda}\mathrm{d}r\mathrm{d}\lambda
\end{aligned} \tag{A.4}$$

where $g(r,\lambda)$ is the RDF and ρ is the solvent number density. Applying the zeroth-order approximation of the RDF of $g(r,\lambda) \approx \exp[-\beta U(r, \lambda)]$ to the second term gives

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

$$\frac{\partial}{\partial\lambda_i}\left\langle\frac{\partial U}{\partial\lambda_j}\right\rangle \approx 4\pi\rho\frac{\partial}{\partial\lambda_i}\int_0^\infty\frac{\partial U}{\partial\lambda_j}\exp[-\beta U(r,\lambda)]r^2\mathrm{d}r$$
$$= 4\pi\rho\int_0^\infty\frac{\partial^2 U}{\partial\lambda_i\partial\lambda_j}\exp[-\beta U(r,\lambda)]r^2 - \beta\frac{\partial U}{\partial\lambda_i}\frac{\partial U}{\partial\lambda_j}\exp[-\beta U(r,\lambda)]r^2\mathrm{d}r$$
$$= \left\langle\frac{\partial^2 U}{\partial\lambda_i\partial\lambda_j}\right\rangle - \int_0^\infty 4\pi\rho\beta\frac{\partial U}{\partial\lambda_i}\times\frac{\partial U}{\partial\lambda_j}\exp[-\beta U(r,\lambda)]r^2\mathrm{d}r \tag{A.5}$$

Substituting eq A.5 into eq A.4 gives the approximated variance equation of

$$\mathrm{Var}(\Delta F) \approx 4\pi\rho\int_0^1\int_0^\infty\sum_{i,j}\left(\frac{\partial U}{\partial\lambda_i}\frac{\partial U}{\partial\lambda_j}\exp[-\beta U(r,\lambda)]r^2\right)\times\frac{\partial\lambda_i}{\partial\lambda}\frac{\partial\lambda_j}{\partial\lambda}\mathrm{d}r\mathrm{d}\lambda \tag{A.6}$$

which reduces to eq 12 for one λ_i.

When the potential is represented with basis functions, we can replace all λ_i directly with h_i. Derivatives in h_i can also be removed from the covariance because they do not participate in the expectation integrals. We define a shorthand such that h_i is equivalent to $h_i(\lambda)$ and

$$\frac{\partial U(r,\lambda)}{\partial h_i} = \frac{\partial h_i}{\partial h_i}u_i(r) + \sum_{j\neq i}\frac{\partial h_j}{\partial h_i}u_j(r) = u_i(r)$$

Equation A.2 then simplifies to

$$\mathrm{Var}(\Delta F) = \int_0^1\sum_{i,j}\mathrm{Cov}\left(\frac{\partial\mathcal{H}}{\partial h_i},\frac{\partial\mathcal{H}}{\partial h_j}\right)\frac{\partial h_i}{\partial\lambda}\frac{\partial h_j}{\partial\lambda}\mathrm{d}\lambda$$
$$= \int_0^1\sum_{i,j}\left(\left\langle\frac{\partial U}{\partial h_i}\frac{\partial U}{\partial h_j}\right\rangle - \left\langle\frac{\partial U}{\partial h_i}\right\rangle\left\langle\frac{\partial U}{\partial h_j}\right\rangle\right)\frac{\partial h_i}{\partial\lambda}\frac{\partial h_j}{\partial\lambda}\mathrm{d}\lambda$$
$$= \int_0^1\sum_{i,j}\left\langle\frac{\partial h_i}{\partial h_i}u_i(r)\frac{\partial h_j}{\partial h_j}u_j(r)\right\rangle - \left\langle\frac{\partial h_i}{\partial h_i}u_i(r)\right\rangle\times\left\langle\frac{\partial h_j}{\partial h_j}u_j(r)\right\rangle\frac{\partial h_i}{\partial\lambda}\frac{\partial h_j}{\partial\lambda}\mathrm{d}\lambda$$
$$= \int_0^1\sum_{i,j}\frac{\partial h_i}{\partial\lambda}\frac{\partial h_j}{\partial\lambda}[\langle u_i(r)u_j(r)\rangle - \langle u_i(r)\rangle\langle u_j(r)\rangle]\mathrm{d}\lambda$$
$$= \int_0^1\sum_{i,j}\frac{\partial h_i}{\partial\lambda}\frac{\partial h_j}{\partial\lambda}Cov(u_i,u_j)\mathrm{d}\lambda \tag{A.7}$$

Defining the covariance matrix for $\mathbf{u}$ as

$$\mathrm{Cov}(\mathbf{u},\mathbf{u}) = \begin{bmatrix}\mathrm{Var}(u_1(r)) & \mathrm{Cov}(u_1(r),u_2(r)) & \cdots & \mathrm{Cov}(u_1(r),u_n(r))\\ \mathrm{Cov}(u_2(r),u_1(r)) & \mathrm{Var}(u_2(r)) & \cdots & Cov(u_2(r),u_n(r))\\ \vdots & \vdots & \ddots & \vdots\\ \mathrm{Cov}(u_n(r),u_1(r)) & \mathrm{Cov}(u_n(r),u_2(r)) & \cdots & \mathrm{Var}(u_n(r))\end{bmatrix}$$

allows the variance for basis functions to be written in condensed, matrix form as

$$\mathrm{Var}(\Delta F) = \int_0^1\mathbf{h}'(\lambda)\cdot\mathrm{Cov}(\mathbf{u},\mathbf{u})\cdot\mathbf{h}'^T(\lambda)\mathrm{d}\lambda \tag{A.8}$$

where $\mathbf{h}'(\lambda) = [\partial h_1/\partial\lambda,\partial h_2/\partial\lambda,...]$. This equation is also eq 17.

$\langle\partial U/\partial\lambda\rangle$ can also be simplified by applying the basis functions to eq A.1 and assuming the Hamiltonian's dependence on λ entirely through its potential energy as was done earlier. The result is

$$\frac{\mathrm{d}F}{\mathrm{d}\lambda} = \left\langle\sum_{i=1}\frac{\partial\mathcal{H}}{\partial h_i}\frac{\partial h_i}{\partial\lambda}\right\rangle = \left\langle\sum_{i=1}\frac{\partial h_i}{\partial h_i}u_i(r)\frac{\partial h_i}{\partial\lambda}\right\rangle = \langle\mathbf{h}'(\lambda)\cdot\mathbf{u}^T\rangle = \mathbf{h}'(\lambda)\cdot\langle\mathbf{u}^T\rangle$$

Extending this equation to predict the expectation of an unsampled switch yields eq 22.

APPENDIX B: DERIVING AN EFFICIENT SHORT-RANGE BASIS FUNCTION

A rule of thumb for the short-range alchemical switch can be derived from eq 12 by assuming that when r is small, $u_i(r)$ for repulsive interactions will be a large, finite, and nearly constant value because of the capped potential. As seen in eq 12, the variance is a function of the product of $(\partial U/\partial\lambda)^2$ and $g(r)$. To minimize this product, we must have, in the two body regime, that

$$\left(\frac{\partial U(r,\lambda)}{\partial\lambda}\right)^2\exp[-\beta U(r,\lambda)]r^2 < V \tag{B.1}$$

where V is approximately constant. If we assume that the potential and thus the variance do not change greatly with r^2 because of the cap, then this will be satisfied with

$$\left(\frac{\partial U(r,\lambda)}{\partial\lambda}\right)^2\exp[-\beta U(r,\lambda)] < V' \tag{B.2}$$

where V' is a new constant. We assume the energy is dominated by the repulsive term near $r = 0$, allowing us to rewrite the equation in terms of the repulsive basis function, u_R, and repulsive alchemical switch, h_R, as

$$\left(\frac{\partial h_R(\lambda)}{\partial\lambda}\right)^2 u_R(r)^2\exp[-\beta h_R(\lambda)u_R(r)] < V' \tag{B.3}$$

We can move the exponential to the right-hand side and, because the variance must always be positive, take the square root considering only the positive inequality. We then apply a

dx.doi.org/10.1021/ct4009188 | J. Chem. Theory Comput. 2014, 10, 1128−1149

Taylor series expansion to the first order of the zeroth-order RDF term and get

$$\frac{\partial h_R(\lambda)}{\partial \lambda} u_R(r) < V'' \left[1 + \frac{\beta h_R(\lambda) u_R(r)}{2} \right] \tag{B.4}$$

where V'' is another modified constant. Since we assumed $u_R(r)$ is large, $1/u_R(r)$ is small, and we divide by $u_R(r)$ to obtain a condition for the $h_R(\lambda)$ of the repulsive term depending only on λ_i:

$$\frac{\partial h_R(\lambda)}{\partial \lambda} < V''' h_R(\lambda) \tag{B.5}$$

with V'' again absorbing other constants. If we make the rule of thumb inequality an equality and apply the boundary conditions on $h_i(\lambda)$, we obtain a solution to eq B.5

$$h_i(\lambda) = \frac{K^\lambda - 1}{K - 1} \tag{B.6}$$

where K is a positive free parameter that can be optimized. This equation is also switch A, eq 18.

APPENDIX C: BRIEF COMPARISON OF VARIANCES IN SOLVENT AND COMPLEX ENVIRONMENT

Low variance paths for alchemical solvation should also be low variance paths for alchemical protein binding. Whether a ligand is surrounded by water or protein, the local density of fluid and therefore the number and size of excluded volume sites is roughly the same. The increased enthalpy of binding configurations would suggest that the variances might be somewhat greater, but the general shape will be determined primarily by the excluded volume and the number density around the solute, which remains roughly independent of the liquid environment.

A full examination of the difference of the variance of pathways in protein and solution does not appear to have been performed. There appears to be a general acceptance that the pathway that is most efficient for one composition will be most efficient for the other. Finding the most statistically efficient pathway to alchemically modify a ligand is frequently done only in one environment[7,9,11−14,24,47] and the other environment is not extensively examined. Many simulation packages with soft core alchemical transformations assume a single pathway for soft core transformations, up to user specifications, regardless of the composition of the system.[10,11,25−27]

We examine a simple comparison of variance between pure solvent and host−guest systems here. This will not be a full study of the minimal variances between complex and solvent, but simply evidence that the low variance pathways developed here for solvent should be low variance paths in complex. We re-examine data from host−guest systems from Monroe and Shirts[54] where the guest molecule is alchemically solvated and alchemically removed from the host structure in separate simulations. Further experimental details and system information are provide in the reference.

Shown in Figure 14 is the variance of the coupling the repulsive and attractive interactions with a soft core 1−1−6 interaction for several guests to the same host. Lines on the figure are drawn connecting data points and, unlike many of the graphs in this paper using the linear basis approach, should not be read as information about unsampled states.

The regions of high variance overlap between solvent and complex environments. In the cases of guest 25 and guest 30,

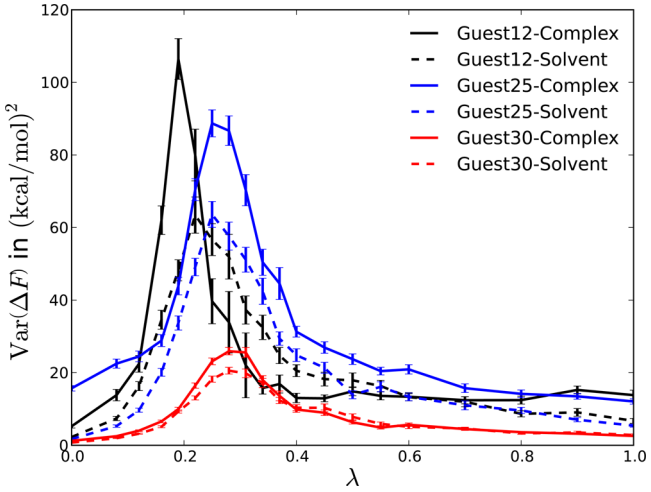


Figure 14. Low variance paths will be similar for solvation and protein binding. The shape of the variance curve will be similar since adding a protein to the solvent will not significantly change the system density or response of fluid to the ligand. Shown here is the variance for solvating and binding three ligands in a host−guest system where the alchemical path and solute are held constant between environments. Lines are drawn to guide the eye and provide a shape of the variance curve but do not represent data between sampled points. Data are only available at points where error is estimated with 200 bootstrap samples. The variance is only shifted slightly in between the environments along λ, so regions of large variance are shared between environments. The magnitude of the variance and number of uncorrelated samples are different in complex than solution but the overlapping peaks means any variance minimization method will minimize the variance in both environments. Data for this figure were generated from Monroe and Shirts.[54]

the maximum in the variance curve is almost at the same λ, and for guest 12, the peak is only shifted by a small $\delta\lambda$. A full demonstration would show that this very close correspondence between the shape of the variance peak holds for all pathways, not just a single pathway. However, the fact that the shapes are the same up to a scaling factor expected because the larger magnitude of ligand binding energies, suggests that the shape of the variance curve for a pure solvent system is similar to that of the ligand bound system.

The variance minimization procedures outlined in this study thus appear to be equally valid for any environment. If minimizing the variance in the solution environment does not provide the desired statistical efficiency, one can run a short, trial simulation with the system of interest and use eq 21 to find a low variance path. Attempting to find a true universally low variance path may not be possible, since changes in the environment can shift the locations of high variance as seen in guest 12 of Figure 14. Even so, the flexibility and power of the linear basis method allows low variance switches to be chosen with ease, simplifying the choice of low variance paths for arbitrary systems.

AUTHOR INFORMATION

Corresponding Author

*E-mail: michael.shirts@virginia.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge funding from the National Science Foundation through grant CHE-1152786 and helpful discussions from John Chodera (Memorial Sloan-Kettering Cancer Center), Peter Eastman (Stanford), and the rest of the OpenMM team. We thank Jacob Monroe (University of Virginia) for making available simulation data on the efficiency of host−guest molecular transformations.

■ REFERENCES

(1) Dobson, C. M. *Nature* **2004**, *432*, 824−828.
(2) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.
(3) Paluch, A. S.; Cryan, D. D.; Maginn, E. J. *J. Chem. Eng. Data* **2011**, *56*, 1587−1595.
(4) Bemporad, D.; Luttmann, C.; Essex, J. W. *Biophys. J.* **2004**, *87*, 1−13.
(5) Paliwal, H.; Shirts, M. R. *J. Chem. Theory Comput.* **2011**, 4115−4134.
(6) de Ruiter, A.; Oostenbrink, C. *J. Chem. Theory Comput.* **2012**, *8*, 3686−3695.
(7) Blondel, A. *J. Comput. Chem.* **2004**, *25*, 985−993.
(8) Crooks, G. *Phys. Rev. Lett.* **2007**, *99*, 10−13.
(9) Steinbrecher, T.; Mobley, D. L.; Case, D. A. *J. Chem. Phys.* **2007**, *127*, 214108.
(10) Hritz, J.; Oostenbrink, C. *J. Chem. Phys.* **2008**, *128*, 144121.
(11) Riniker, S.; Christ, C. D.; Hansen, H. S.; Hünenberger, P. H.; Oostenbrink, C.; Steiner, D.; van Gunsteren, W. F. *J. Phys. Chem. B* **2011**, *115*, 13570−7.
(12) Buelens, F. P.; Grubmüller, H. *J. Comput. Chem.* **2011**, 25−33.
(13) Pham, T. T.; Shirts, M. R. *J. Chem. Phys.* **2011**, *135*, 034114.
(14) Pham, T. T.; Shirts, M. R. *J. Chem. Phys.* **2012**, *136*, 124120.
(15) Bennett, C. H. *J. Comput. Phys.* **1976**, *268*, 245−268.
(16) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(17) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1339−1350.
(18) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.
(19) Shenfeld, D.; Xu, H.; Eastwood, M.; Dror, R.; Shaw, D. *Phys. Rev. E* **2009**, *80*, 1−4.
(20) Gelman, A.; Meng, X.-L. *Stat. Sci.* **1998**, *13*, 163−185.
(21) Beutler, T.; Mark, A.; van Schaik, R. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(22) Zacharias, M.; Straatsma, T. P.; McCammon, J. a. *J. Chem. Phys.* **1994**, *100*, 9025.
(23) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.
(24) Deng, Y.; Roux, B. *J. Phys. Chem. B* **2004**, *17*, 16567−16576.
(25) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(26) Brooks, B. R.; Iii, C. L. B.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(27) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J. et al. *AMBER 12*; University of California: San Francisco, 2012.
(28) Wang, J.; Deng, Y.; Roux, B. *Biophys. J.* **2006**, *91*, 2798−814.
(29) Mobley, D. L.; Dumont, E.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007**, *111*, 2242−54.
(30) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237.
(31) Wang, K.; Chodera, J. D.; Yang, Y.; Shirts, M. R. *J. Comput. Aid. Mol. Des.* **2013**, *27* (12), 989−1007.
(32) YANK. http://simtk.org/home/yank/ (accessed Jul. 8, 2012), available through SimTK.
(33) Friedrichs, M.; Eastman, P.; Vaidyanathan, V.; Houston, M.; LeGrand, L.; Beberg, A.; Ensign, D.; Bruns, C.; Pande, V. S. *J. Comput. Chem.* **2009**, *30*, 864−872.
(34) Eastman, P.; Pande, V. *Comput. Sci. Eng.* **2010**, *12*, 34−39.
(35) Eastman, P.; Pande, V. S. *J. Comput. Chem.* **2009**, *31*, 1268−1272.
(36) OpenMM. http://simtk.org/home/openmm (accessed Jul. 8, 2012), available through SimTK.
(37) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058.
(38) Chodera, J. D.; Shirts, M. R. *J. Chem. Phys.* **2011**, *135*, 194110.
(39) Eastman, P.; Pande, V. *J. Chem. Theory Comput.* **2010**, 434−437.
(40) Pitera, J. W.; van Gunsteren, W. F. *Mol. Simulat.* **2002**, *28*, 45−65.
(41) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350−358.
(42) Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman & Hall/CRC: Boca Raton, FL, 1993; p 436.
(43) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *341*, 327−341.
(44) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952−962.
(45) Chow, K.-H.; Ferguson, D. M. *Comput. Phys. Commun.* **1995**, *91*, 283−289.
(46) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. r. O. *Chem. Phys. Lett.* **2004**, *384*, 288−294.
(47) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740.
(48) Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; Pande, V. S. *J. Phys. Chem. B* **2007**, *111*, 13052−63.
(49) GROMACS. http://www.gromacs.org/ (accessed Jan. 6, 2012).
(50) Powell, M. J. D. In *Advances in Optimization and Numerical Analysis*; Mathematics and Its Applications;Gomez, S., Hennart, J. P., Eds.; Springer: New York, 1994; pp 51−67.
(51) SciPy: Open Source Scientific Tools for Python. http://www.scipy.org/ (accessed Jan. 10, 2012).
(52) Fritsch, F.; Carlson, R. *SIAM J. Numer. Anal.* **1980**, *17*, 238−246.
(53) Fritsch, F.; Butland, J. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 300−304.
(54) Monroe, J. I.; Shirts, M. R. *J. Comput. Aid. Mol. Des.*; DOI: 10.1007/s10822-014-9716-4.