1-14-2020

Determination of the Free Energies of Mixing of Organic Solutions through a Combined Molecular Dynamics and Bayesian Statistics Approach

Shi Li
University of Kentucky

Balaji Sesha Sarath Pokuri
Iowa State University, balajip@iastate.edu

Sean Ryno
University of Kentucky

Asare Nkansah
University of Kentucky

Camron De'vine
University of Kentucky

See next page for additional authors

Determination of the Free Energies of Mixing of Organic Solutions through a Combined Molecular Dynamics and Bayesian Statistics Approach

Abstract

As new generations of thin-film semiconductors are moving towards solution-based processing, the development of printing formulations will require information pertaining to the free energies of mixing of complex mixtures. From the standpoint of in silico materials design, this move necessitates the development of methods that can accurately and quickly evaluate these formulations in order to maximize processing speed and reproducibility. Here, we make use of molecular dynamics (MD) simulations in combination with the two-phase thermodynamic (2PT) model to explore the free energy of mixing surfaces for a series of halogenated solvents and high boiling point solvent additives used in the development of thin-film organic semiconductors. While the combined methods generally show good agreement with available experimental data, the computational cost to traverse the free-energy landscape is considerable. Hence, we demonstrate how a Bayesian optimization scheme, coupled with the MD and 2PT approaches, can drastically reduce the number of simulations required, in turn shrinking dramatically both the computational cost and time.

Disciplines

Chemical Engineering | Other Chemistry

Authors

Shi Li, Balaji Sesha Sarath Pokuri, Sean Ryno, Asare Nkansah, Camron De'vine, Baskar Ganapathysubramanian, and Chad Risko

Determination of the Free Energies of Mixing of Organic Solutions through a Combined Molecular Dynamics and Bayesian Statistics Approach

Shi Li,[§] Balaji Sesha Sarath Pokuri,[†] Sean M. Ryno,[§] Asare Nkansah,[§] Camron De'vine,[§] Baskar Ganapathysubramanian[†,*] & Chad Risko[§,*]

[§] *Department of Chemistry & Center for Applied Energy Research (CAER) University of Kentucky Lexington, KY 40506*

[†] *Department of Mechanical Engineering Iowa State University Ames, Iowa 50011*

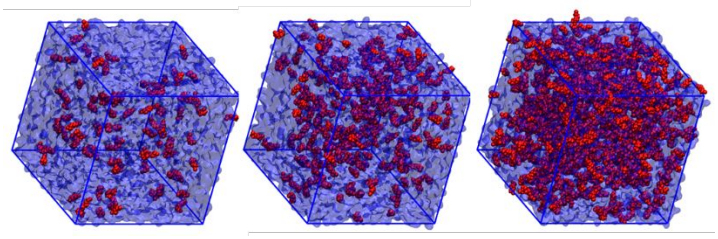
Corresponding Authors
Chad Risko, chad.risko@uky.edu
Baskar Ganapathysubramanian, baskarg@iastate.edu

ORCID
Shi Li – 0000-0003-0505-6751
Balaji Sesha Sarath Pokuri – 0000-0002-5816-0184
Sean M. Ryno – 0000-0002-8359-2659
Baskar Ganapathysubramanian – 0000-0002-8931-4852
Chad Risko – 0000-0001-9838-5233

Abstract

As new generations of thin-film semiconductors are moving towards solution-based processing, the development of printing formulations will require information pertaining to the free energies of mixing of complex mixtures. From the standpoint of *in silico* materials design, this move necessitates the development of methods that can accurately and quickly evaluate these formulations in order to maximize processing speed and reproducibility. Here, we make use of molecular dynamics (MD) simulations, in combination with the two-phase thermodynamic (2PT) model, to explore the free energy of mixing surfaces for a series of halogenated solvents and high boiling point solvent additives used in the development of thin-film organic semiconductors. While the combined methods generally show good agreement with available experimental data, the computational cost to traverse the free-energy landscape is considerable. Hence, we demonstrate how a Bayesian optimization scheme, coupled with the MD and 2PT approaches, can drastically reduce the number of simulations required, in turn shrinking both the computational cost and time.

TOC Graphic



Introduction

In addition to the geometric structures and chemical compositions of the individual molecules or polymers, the solid-state morphologies of thin-film organic semiconductors (OSC) are determined by the processing conditions. While the current application of organic light-emitting diodes (OLED) in display technologies relies on vapor deposition, there is a move to optimize solution processing (*i.e.* printing) capabilities to enable low-cost production at large scale and realize conformal additive manufacturing. Depending on the OSC and intended application, the formulation of the printing solution (*i.e.* ink) can vary in complexity. In bulk heterojunction (BHJ) organic photovoltaics (OPV), for instance, controlling the degree of phase separation of the binary or ternary material blend is critical to material performance.[1, 2] While much of the early efforts to print BHJ OPV relied solely on halogenated solvents, Bazan and co-workers demonstrated that the addition of a small fraction of a high boiling point solvent to the solution could enable a dramatic change to the nanoscale phase separation in the OSC thin film.[3] Across the literature, OSC deposited with these low-volatility solution additives have facilitated generally improved OPV performance when compared to OSC active layers printed from halogenated solvents on their own, though this is not always the case.[4-9] Notably, the choice of primary solvent and additive to implement in such ink formulations remains highly empirical, and it is often through trial-and-error approaches that appropriate ratios of the solution components are determined. If such Edisonian tactics are to be overcome and *a priori* processing design guidelines established, there is a pressing need to understand at a fundamental level how the primary solvent and additive work in concert.

Developing this requisite insight requires knowledge of the free energy of mixing (ΔG_m) for all components that comprise the solution. From the standpoint of molecular dynamics (MD)

simulation methods, the determination of ΔG_m is complicated by the fact that the determination of the entropy of mixing (ΔS_m) is not straightforward. Methods such as umbrella sampling,[10] Widom particle insertion,[11] and thermodynamic integration,[12] can effectively probe the entropy of an ensemble, though each has potential drawbacks depending on the system under investigation. Another means to determine entropy is to evaluate the system vibrations. Here, the thermodynamic characteristics can be evaluated under the premise that the vibrational density of states (DoS) is comprised of harmonic oscillators. While such an assumption is generally valid for solids, the anharmonic nature of low-frequency and diffuse vibrational modes in liquids and gases limits the hypothesis.[13] To overcome this constraint for liquids, Lin, Blanco, and Goddard proposed the two-phase thermodynamic (2PT) model,[14] where the thermodynamic characteristics of a liquid are determined based on the premise that the total DoS (referred to here as $DoS(v)$) is the sum of two components: (i) a gas component that contains the anharmonic diffusive vibrations ($DoS_{diffuse}(v)$), and (ii) a solid component that contains the harmonic vibrations ($DoS_{solid}(v)$), as shown in Equation 1,[15]

$$DoS(v) = f \times DoS_{diffuse}(v) + (1 - f) \times DoS_{solid}(v) \qquad \text{Equation 1}$$

A key feature of 2PT theory is the fluidicity parameter (*f*), which is a function of the system properties (*e.g.* self-diffusion, density, and temperature) that is solved self-consistently from the MD simulation.[16] The system thermodynamics are recovered by applying statistical weighting functions to each respective component. Since the inception of the 2PT method, it has been shown to provide accurate determinations of the absolute entropies and free energies for a variety of chemical systems, including those with considerable chemical and physical complexity.[13, 14, 17-32]

Returning to the concept of printing, when considering solution drying, *e.g.* as the solvent in an ink formulation dries to leave behind a thin film, the relative concentrations of the primary solvent and all solution components are continuously changing. Hence, it is important to determine ΔG_m over a wide range of concentrations. Charting the entire free-energy surface, therefore, could require performing many MD simulations over a dense sampling of mass/volume fraction configurations. From the perspective of computation, such sampling is resource intensive, and even infeasible for many-component systems. An attractive option instead of uniform (and dense) sampling of the configuration space is to perform adaptive sampling that minimizes the total number of simulations required to represent the free-energy surface. Here, we specifically use ideas from Bayesian sampling techniques that identify regions of the configuration space that have the most information to construct the free energy surface. Based on minimal assumptions of smoothness (*i.e.* continuity assumptions) of the free energy landscape curve, Bayesian statistics provides a formal, rigorous approach to sequentially identify informative configurations and assimilate the resulting MD simulations at these configurations to produce a representation of the free-energy surface. Additionally, Bayesian statistics provides a confidence estimate on the constructed surface.[33-35] It has been shown that the constructed surface (called the posterior estimate) can model arbitrarily large degree polynomials and complex surfaces with little error.[36-39] This combination of estimate along with the confidence in the estimate enables natural incorporation of user-specified bounds on sampling. The availability of easy-to-use software[40] opens the possibility of integrating Bayesian statistics with molecular simulations for the efficient, automated and accurate construction of the free energy landscape.

Given this backdrop, we are broadly interested in exploring *in silico* OSC solution processing. Here, as a first step in this process, we report on the application of the 2PT approach in combination

with Bayesian statistics to determine ΔG_m for a variety of organic solvent−high-boiling-point additive solutions. Specifically, the solvents chosen are chlorobenzene (CB), 1,2-dichlorobenzene (DCB), and 1,2,3-trichlorobenzene (TCB), while the additives are chloronaphthalene (CN) and 1,8-diiodooctane (DIO), see Figure 1; each of these solvents and additives have found wide use in solution-based OSC processing. By varying the additive concentrations in the binary solutions, we aim to chart the entire ΔG_m surface the solution may follow during the drying process. The methods developed here provide a framework for further investigations of such multicomponent systems across a range of applications.

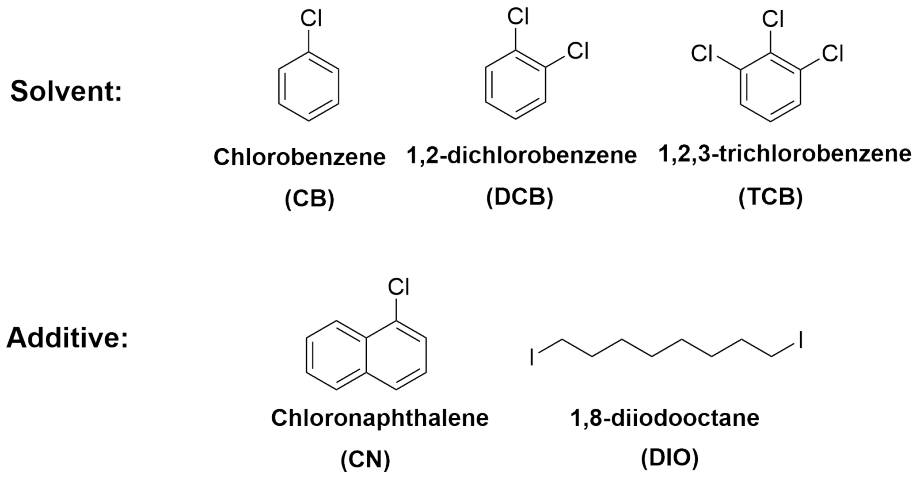


Figure 1. Primary solvents and high boiling-point solvent additives studied in this work.

Computational Methods

Atomic charges. Density functional theory (DFT) calculations were performed at the ωB97X-D/6-311g(d,p) level of theory in the Gaussian 09, Revision A.02 program.[41] The CM5 model[42] was used to determine the atomic charges. For DIO, a split-basis set was used in order to calculate the

charge on the iodine atom. An example of the Gaussian input for DIO molecule can be found in the Supporting Information (SI).

System generation and equilibration. All MD simulations were performed using the GROMACS 2018.1 software package.[43, 44] The binary solvent−additive blends were built as a function of additive weight percent; Tables S1 and S2 show the makeup of the binary combinations considered in this study. The intra- and inter-molecular interaction parameters were built from the OPLS-AA (optimized potentials for liquid simulations-all atom) force field.[45, 46]

The equilibration of each system was completed through a workflow that included an initial energy minimization process using the steepest descent algorithm followed by an *NVT* (constant-number, constant-volume, and constant-temperature) and two *NPT* (constant-number, constant-pressure, and constant-temperature) ensembles that made use of the leap-frog integrator with a time step of 2 fs. Three-dimensional periodic boundary conditions (PBC) were applied using the velocity rescaling thermostat with a coupling time of 0.04 ps. Then, an *NPT* ensemble using the Berendsen[47] barostats was applied to stabilize the volume, followed by the Parrinello-Rahman[48] barostat that allows additional dynamic shape change (both with a coupling time of 1.0 ps). A compressibility of 4.5×10^{-5} bar^{-1} was used for both *NPT* ensembles. A spherical cut-off of 1.4 nm for the summation of van der Walls (vdW) interaction and the particle-mesh Ewald (PME)[49] solver for long-range Coulomb interactions (with a cut-off of 1.4 nm) were used throughout all simulations. Each *NVT* and the first *NPT* ensemble was carried out for 5 ns, while the second *NPT* ensemble was carried out for 10 ns, and the enthalpy of each equilibrated system was extracted for analysis. We note that the densities of each single-component solvent and binary solutions were determined, and a plot of mixture density vs. additive concentration is shown in Figure S4. As one might expect,

the densities of the mixtures follow a trend line that connects the densities of the pure host solvent and additive.

For the 2PT evaluations, 20 ps MD runs were run in the *NVT* ensemble. Here, we made use of the DoSPT code developed by Caro and co-wokers, which can be found at http://dospt.org/. [21, 50] We note that minor modifications were made to the DoSPT code as distributed to accommodate the large system sizes used here. These modifications include vectorization of the position and volume data arrays and parallelization via OpenMP of the Voronoi cell calculations for each atom. The modified code is available at https://github.com/sryno/DoSPT.

Solubility parameters. A key physical characteristic used to empirically design solutions is the solubility parameter. In particular, the Hildebrand[51] and Hansen[52] solubility parameters have found wide-spread use. For a pure substance, the Hildebrand solubility parameter is defined as the square root of the cohesive energy density,[53]

$$\delta = \sqrt{CED} = \left[\frac{\Delta H_{vap} - RT}{V_m}\right]^{1/2} \quad \text{Equation 2}$$

where CED is the cohesive energy density, ΔH_{vap} is the heat of vaporization, R is the gas constant, T is the temperature, and V_m is the molar volume. Hansen proposed an extension of the Hildebrand parameter to estimate the relative miscibility of polar and hydrogen bonding systems,

$$\delta^2 = \delta_d^2 + \delta_p^2 + \delta_h^2 \quad \text{Equation 3}$$

where δ_d, δ_p and δ_h represent the dispersion, dipolar and hydrogen-bond intermolecular forces among molecules, respectively. However, the determination of hydrogen-bonding component requires a separation of the Coulombic interaction into separate dipole-dipole and hydrogen-bond

interactions, while in MD simulations this is determined through an arbitrary cutoff length.[54] Therefore, in this study, we did not extract individual dipolar and hydrogen bond contributions, rather we report a global Coulombic interaction ($\delta_p + \delta_h$). To calculate the solubility parameters, the heat of vaporization was determined from potential energy difference of the gas and liquid phase, while the Coulombic interaction term was extracted directly from the simulation.

Solvation free energy. The solvation free energy was determined using the Bennett acceptance ratio (BAR)[55] perturbation free energy method as implemented in the GROMACS software suite. Here, the state coupling parameter λ was modified from zero to one through equidistant values of 0.05, and the free energy difference (ΔG_{AB}) was calculated with *gmx bar*. To validate the BAR method, the solvation energy of several common solvents in water were evaluated and compared with literature values, each showing reasonable agreement with experiment (Table S3).[56]

Diffusion coefficients. The mean square displacement (MSD) for the additive present in the host solvent can be calculated from a set of initial positions during the equilibrium process. The MSD is defined as

$$MSD = \left\langle (r - r_0)^2 \right\rangle = \frac{1}{N}\sum_{i=1}^{N} (r_i(t) - r_i(0))^2 \qquad \text{Equation 4}$$

where $r_i(t)$is the location of the center-of-mass (COM) of additive i at time t and N is the total number of additive in the system. Therefore, the diffusion constant of the additive can be computed via the Einstein relation[57] by fitting a straight line

$$MSD(t) = D \times t + c \qquad \text{Equation 5}$$

through the $MSD(t)$, where D is the diffusion constant and t is the time from the reference positions.

Bayesian Sampling: We represent the free energy as $y = f(\bar{x})$, where $\bar{x}$ is the configuration of the system. $\bar{x}$ includes fractions of each component, temperature of the system and any other factors that determine the free energy surface. Using Bayesian sampling, we approximate this free energy representation as $y \approx \widehat{y(\bar{x})} = \sum_{i=1}^{N} w_i g(\bar{x}, \bar{x}_i)$, where g is (kernel) basis, w_i are weights (to be computed), and $i = 1,..,N$ are the free energies at N configurations that have been computed. The function y is approximated using a set of basis functions. This approximated function, which we refer to as the surrogate, is an exact interpolant of the original free energy at the known (N) configurations. The weights w_i are determined by solving a system of linear equations for the existing N configurations, through the calculation of the covariance matrix as follows:

$$\hat{y} = \boldsymbol{K}\, w, \textit{ where } K_{i,j} = g(\bar{x}_i, \bar{x}_j) \quad \forall \; i,j \in [1,N]) \qquad \text{Equation 6}$$

The mean and variance of the surrogate can be calculated, respectively, as

$$\boldsymbol{\mu}(\bar{x}) = \boldsymbol{k}^T \boldsymbol{K}^{-1} y_{1:N} \qquad \text{Equation 7}$$

$$\boldsymbol{\sigma}^2 = g(\bar{x}, \bar{x}) - \boldsymbol{k}^T \boldsymbol{K}^{-1} \boldsymbol{k} \qquad \text{Equation 8}$$

where $\boldsymbol{k} = [g(\bar{x}, \bar{x}_1),\ g(\bar{x}, \bar{x}_2)..., g(\bar{x}, \bar{x}_N)]$, the correlation vector with respect to every existing point. We assume C^{∞} continuity of the free energy function, and choose a kernel function of the form $(\bar{x}_i, \bar{x}_j) = \exp\left(-\frac{||\bar{x}_i - \bar{x}_j||^2}{l^2}\right)$; where l represents the average correlation length of free energy function (y).

After using the N points to construct the current best representation of the free energy surface, the Bayesian approach suggests the next set of configurations at which the free energy must be evaluated. The choice of the next evaluation configurations can be informed by the estimate (μ) and the confidence in the estimate (σ) using an acquisition function. Popular choices of acquisition functions include confidence bounds (CB) and expectation of improvement (EI). Configurations that maximally increase these acquisition functions are chosen next for simulation.

Results and Discussion

Pure solvent equilibration. Prior to the examination of ΔG_m, it is important to establish the appropriate system sizes for the simulations. Since we are particularly interested in following the mixing processes at small concentrations of the additives, as the additives are often included at sub 1% concentrations in the initial solution, it is imperative that the simulation boxes be large enough to allow for such scenarios. We begin with evaluations of the pure liquids that are subjected to the 2PT analysis; relevant data is summarized in Figure S1 and S2 of the SI. With increasing system size, the variations in ΔS plateau, as summarized in Figure S1. Notably, as the system sizes are extended, the computational cost for the DoSPT increases exponentially (Figure S3). To balance the computational cost of the exponential system size dependence of the DoSPT method with enough additive molecules to ensure ΔS convergence, each pure and binary system studied here contains 10,000 total molecules.

Table 1 summarizes the results from six individual simulations for each pure solvent. In general, the liquid densities and the standard molar entropies of CB and DCB show good agreement with experiment. DIO does present a larger standard deviation of 6.5% with respect to the experimental

density; this deviation could be due to the fact that the non-bonded parameter for iodine was derived for aryl halides,[58] which present slightly different chemistries than one might expect for alkyl halides. We note that two different non-bonded parameters for iodine were evaluated and the input files are included in the SI, with the one that presented the best agreement with the experimental DIO density being used in all subsequent calculations. Although the 2PT method has been widely applied to liquids and even mixtures, most of the simulation targets have been ions, non-organic liquids,[11-26] or water/organic mixtures.[13, 16-31] The validation of using 2PT on pure organic solvents here provides confidence in the application of the method for binary organic mixtures.

We also determined the solubility parameters of each organic solvent and additive. The calculated Hildebrand and Hansen solubility parameters are in good agreement with previous reference values.[53] The solubility parameters of each pure solvent fall within 9.4 to 10 $cal^{1/2}$ $cm^{-3/2}$, suggesting from an enthalpic standpoint that each additive should readily mix with the host solvent.

Table 1. Hildebrand and Hansen solubility parameter, density and standard molar entropy of solvent systems calculated from simulation.

	Density (g/cm^3)		Standard molar entropy (J/mol K)		Hildebrand ($cal^{1/2}\,cm^{-3/2}$)		Hansen ($cal^{1/2}\,cm^{-3/2}$)			
	ρ		S_0		δ		δ_d		$\delta_p + \delta_h$	$\delta_p + \delta_h$
	Ref.[a]	MD	Ref.[a]	MD	Ref. [b]	MD	Ref. [b]	MD	Ref. [b]	MD
CB	1.11	1.14	197.48	196.97 ± 0.64	9.50	9.42 ± 1.64	10.30	8.35 ± 1.34	3.08	4.36 ± 0.04
DCB	1.30	1.33	211.29	224.15 ± 0.39	10.02	10.04 ± 1.14	8.94	9.02 ± 0.65	5.13	4.40 ± 0.07
TCB	1.46	1.51	N/A	254.95 ± 0.43	N/A	10.21 ± 0.26	N/A	9.22 ± 0.04	N/A	4.38 ± 0.06
CNP	1.19	1.21	N/A	228.73 ± 0.35	9.80	9.61 ± 0.78	9.73	8.79 ± 0.52	3.62	3.88 ± 0.06
DIO	1.84	1.96	N/A	312.41 ± 5.26	N/A	10.12 ± 1.01	8.94	9.39 ± 0.29	3.57	3.79 ± 0.20

[a]Density and entropy data were retrieved from the NIST WebBook at www.nist.gov. [b]Experimental solubility parameters were retrieved from Reference [59] and the Hansen Solubility Parameters handbook.[60]

Solutions – Free Energy of Solvation and Diffusion. Since experimental solvation energies of our target systems are not available, we selected several common organic–aqueous solutions and determined the solvation energy for these reference solutions as a means to validate the approach. The results, presented in Table S3, demonstrate that the calculated solvation energies agree reasonably well with reference experimental values.[56] In turn, we determined the solvation free energy of each additive in the three host solvents. All solvation free energies are negative, as shown in Figure 2, suggesting that the solvation of the additive into the host solvent is an exothermic process. For both CN and DIO, the solvation free energy is similar across all three solvents, indicating that the energy differences associated with dissolving the additive into the different solvents are negligible. Solutions with DIO present solvation energies about 10 kJ/mol larger than the solutions containing CN, suggesting that DIO is more readily solvated.

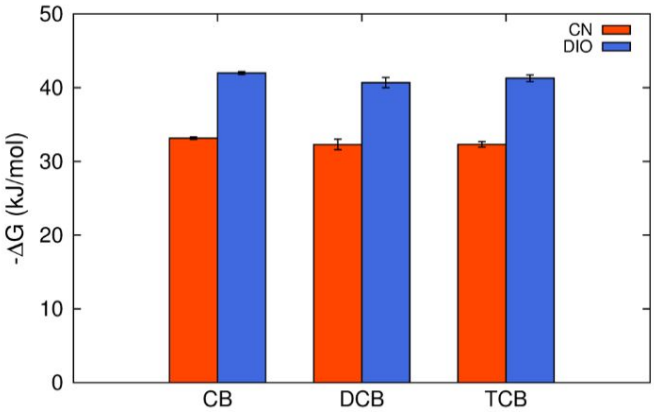


Figure 2. Free energy of solvation of CN or DIO in each of the three host solvents.

Turning to diffusion, there is large variation in the additive diffusion coefficients at low additive concentrations, Figure 3. In these scenarios, there are only a few additive molecules present in the system, resulting large standard deviations. The diffusion constant stabilizes in all mixtures at additive concentrations larger than 10%. Generally, the additive diffusion constants in each host solvent follow the order CB > DCB > TCB, which is expected in part due to the increasing size of the host solvent. Note that when CB is the host solvent and at low additive concentrations, both CN and DIO have large diffusion coefficients, around 0.8×10^{-5} cm^2/s, which then decrease significantly with increasing concentration; this result suggests that there are more solvent–additive interactions at lower additive concentration in CB solution that decrease the additive diffusion rates. In DCB and TCB, the additive diffusion constants are less affected by additive concentration.

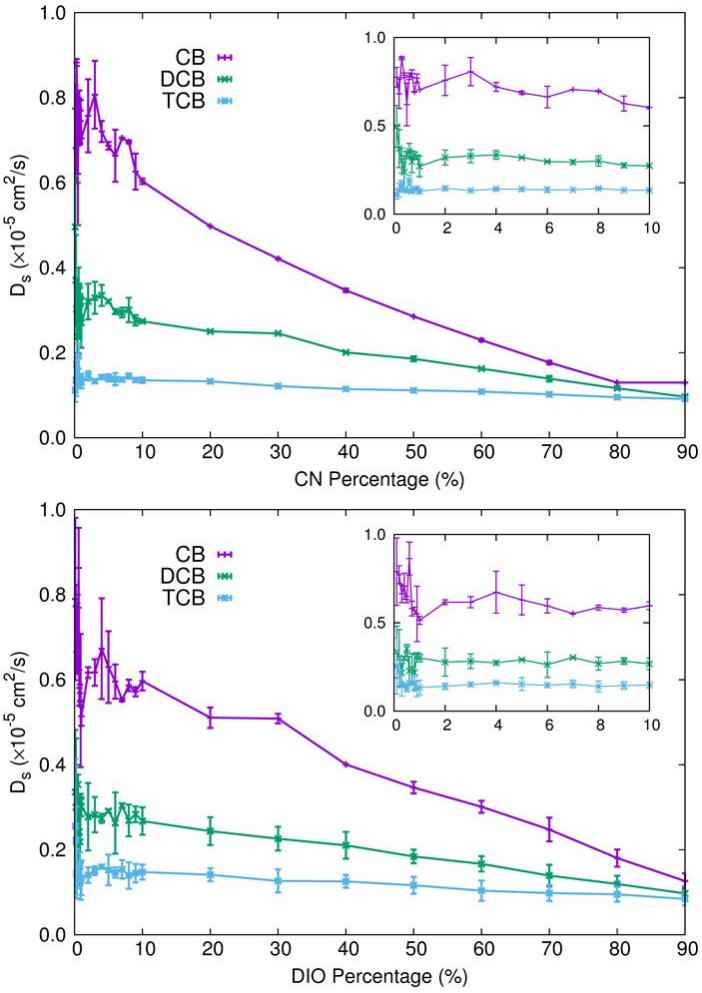


Figure 3. Diffusion constants of the two additives, CN (top) and DIO (bottom), in the three host solvents as a function of the additive concentration.

Gibbs free energy of mixing

All binary solvent–additive mixtures present ΔG_m less than zero across all additive fractions, Figure 4, indicating the additives are generally soluble with three host solvent. The minimum of ΔG_m occurs between 0.2 – 0.4 additive mole fraction. Interestingly, when DIO is the additive, the enthalpy terms are all positive, suggesting that the mixing of DIO with the host solvent is enthalpically unfavorable. For mixtures containing CN as the additive, the enthalpy is nearly

unchanged and zero across all mole fractions. In every case, ΔG_m follows the trends set by $-T\Delta S$, suggesting that the mixing between the additive and host solvents are entropy driven. Notably, when CB and DCB are the host solvent, both DIO and CN show a slightly higher ΔG_m than when TCB is the host.

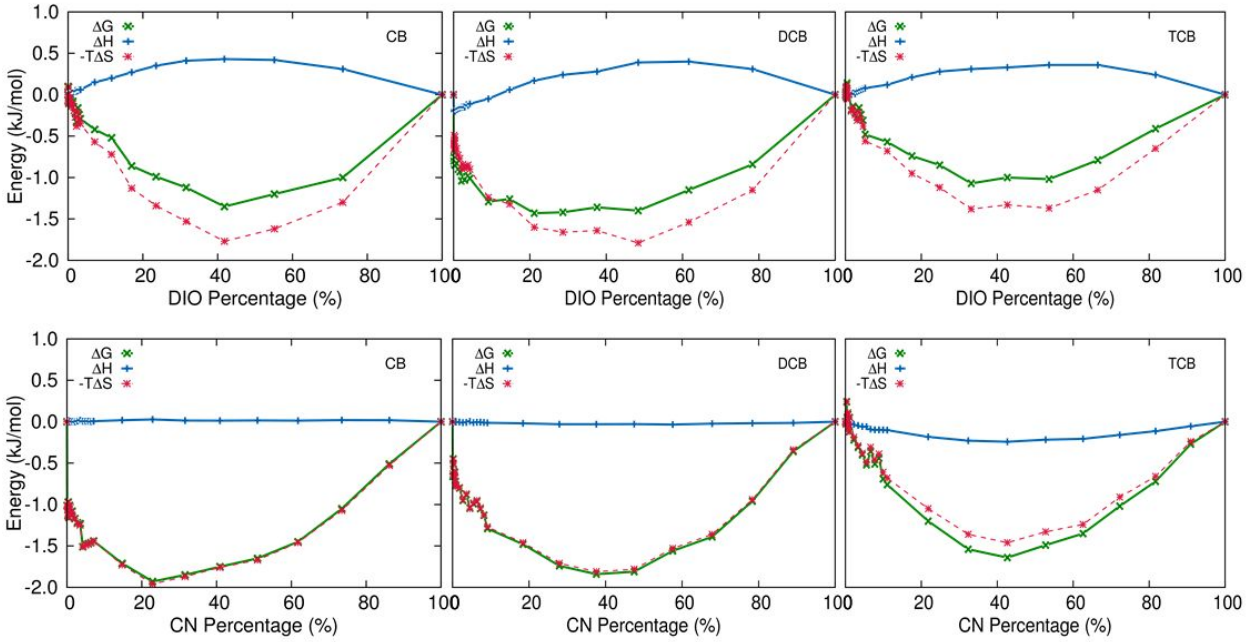


Figure 4. Enthalpy, entropy and Gibbs free energy of mixing as function of the additive mole fraction of the binary solutions.

For the purpose of solution printing OSC, there has been a move towards developing non-halogenated "green" solutions. To further show the generality of the method, we also examined the mixing of 1,2,4-trimethylbenzene (TMB) as the host solvent and 1-phenylnaphthalene (PN) as the additive (Figure 5).[61] Here, the minimum of ΔG_m occurs between 0.4 – 0.6 additive mole fraction. The enthalpy term is again slightly positive, suggesting that the mixing of PN with the TMB is enthalpically unfavorable. The overall entropy and free energy of mixing are similar to

the halogenated solutions using CN as the additive, which is expected due to the chemical similarity of the CB–CN and TMB–PN mixtures.

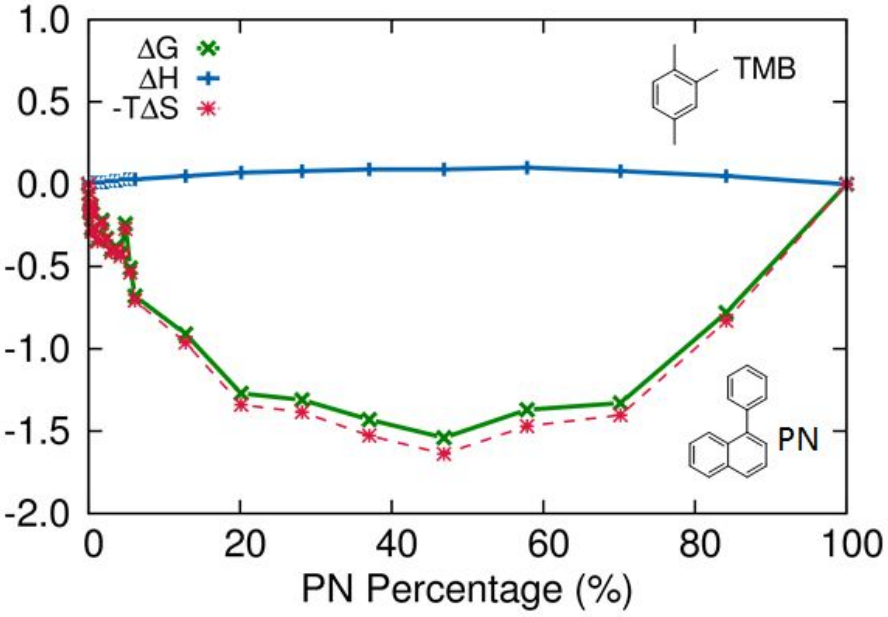


Figure 5. Enthalpy, entropy and Gibbs free energy of mixing as function of the additive mole fraction of the TMB–PN solutions.

Comparisons with Bayesian Optimized Free-Energy Profiles. Spanning the entire free energy surface for mixing the solvents and high boiling point additives required rather considerable computational cost. Here we are interested in exploring how Bayesian models can be implemented to lower the number of computations required to examine the full potential energy surface. We applied the Bayesian optimization method to the six halogenated binary mixtures, then used the first ten Bayesian predictions to recreate three independent systems on each binary mixture. For each system, the change of enthalpy was calculated and compared with our original simulation results. Because of the decrease in the number of sampling systems, the computational core hours

were significantly decreased. As an example, the original CB-DIO simulation includes 27 independent mixtures and required 27,338 core hours. After Bayesian optimization, the core hours were reduced by 68%, with only 8,895 core hours used.

Figure 6 reveals that the Bayesian method provides an excellent estimate of the enthalpy surface, where only 10 fractions are needed to reproduce the potential energy surfaces originally obtained from simulations with 29 data points. The standard approach resulted in a selection of points that are equally spaced (logarithmically) in the range of [0 1], [1 10], and [10 100]. This is agnostic to the behavior of the energy surface. In contrast, the Bayesian method (adaptively) chooses points that are maximally informative about the local behavior of the energy surface. Thus, points are selectively placed in regions where there are steep(er) gradients, with very few points in regions of the energy surface exhibiting constant or (near) linear behavior. We also observe that there were more fluctuations between independent runs in the CN systems, however, due to the scale of the enthalpy change in CN mixtures was relatively small, such fluctuation was acceptable and will not affect the final energy of mixing.

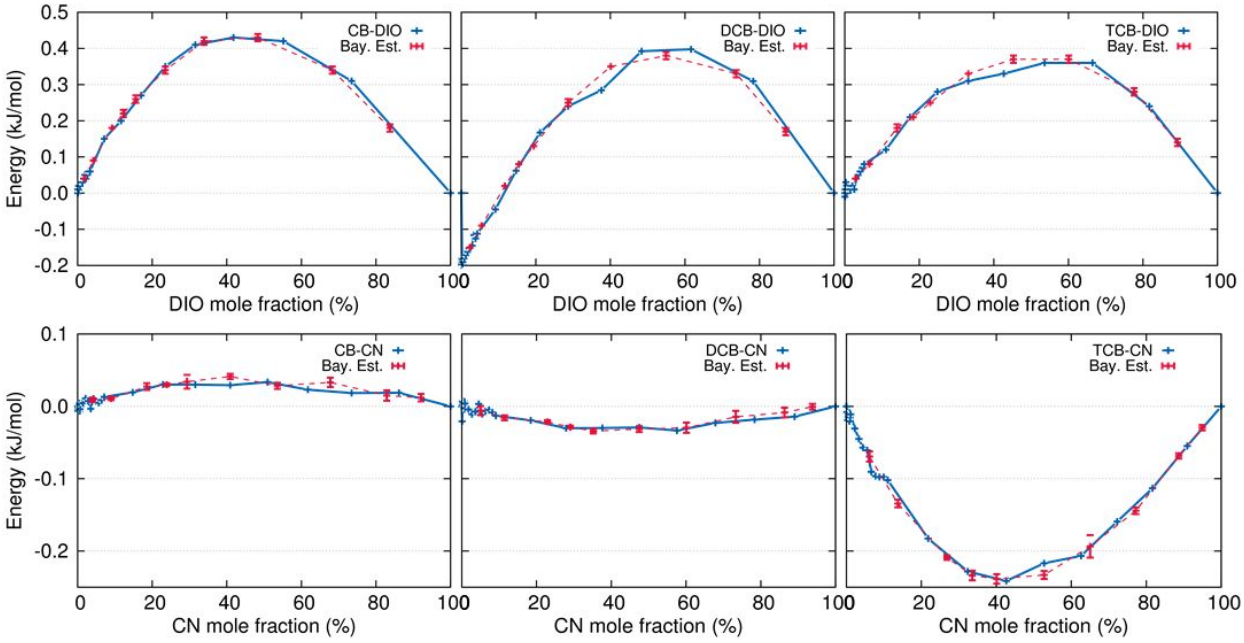


Figure 6. Change of enthalpy as function of the additive mole fraction from Bayesian optimization of the binary solutions. Each dashed line was an average of three independent simulations. The original trend was used as solid lines for reference.

The reduction in computational effort, for even this one-dimensional function representation, is noteworthy considering that each molecular simulation in principle take several days to run on a computing cluster. Importantly, the reduction in the computational effort improves significantly with increasing dimensionality, *i.e.* with increasing number of components in the system (ternary, quaternary, and quintenary systems). We illustrate this in the Figure S6 by reconstructing an analytical free energy surface for a ternary system. Here, using Bayesian sampling to approximate the function requires only ≈ 40 configurational evaluations, whereas dense sampling will require $\approx 0.5 * 30^2 = 450$ configurational evaluations, which is a nearly an order of magnitude reduction in computational effort. We anticipate significantly larger computational reductions with more complex systems.

Conclusion

Accurate and expedient determinations of the free energy of mixing can hold important consequence across a number of fields, including the development of emerging technologies. Here, with a focus on beginning to develop *in silico* protocols to understand and design printing inks for OSC, we make use of MD simulations coupled with 2PT to determine the free energies of mixing for commonly used halogenated solvents and additives across a range of concentrations. We showed that the mixing in these systems is predominantly entropically driven, which highlights the importance of needing robust methods to determine entropic terms. However, spanning the entire profile of concentrations that one might expect to observe during the ink drying process required considerable computational resources and time. Implementation of a Bayseian optimization scheme was shown to greatly reduce the number of simulations required, leading to a tremendous reduction in computational cost and time to determine the free energies as a function of concentration. This work represents a first step along the path to implementing *in silico* design of OSC that includes the processing conditions.

Acknowledgements

The work at the University of Kentucky (UK) was supported by the National Science Foundation (NSF, Award No. CMMI 1563412). A.N. and C.D. were supported in part by the NSF Broadening Participation in Engineering program (Award No. EEC 1444779). Supercomputing resources at UK on the Lipscomb High Performance Computing Cluster were provided by the UK Information Technology Department and Center for Computational Sciences (CCS). The work at Iowa State University was supported by the National Science Foundation (Award No. CMMI 1563359).

Supporting Information Available

The Supporting Information is available free of charge on the ACS Publications website at DOI:

MD binary system combinations; comparisons among available experimental thermodynamic data with those derived from the simulations; estimate of the speed-up made available by the Bayesian model; modified DoSPT code used in this work.

Notes

The authors declare no competing financial interest.

References

1. Halls, J. J. M.; Walsh, C. A.; Greenham, N. C.; Marseglia, E. A.; Friend, R. H.; Moratti, S. C.; Holmes, A. B., Efficient photodiodes from interpenetrating polymer networks. *Nature* **1995**, 376, 498.

2. Yu, G.; J.Gao; C.Hummelen, J.; Wudl, F.; Heeger, A. J., Ploymer Photovoltaic Cells: Enhanced Efficiencies via a Network of Internal Donor-Acceptor Heterojunctions. *Science* **1995**, 270, 1789.

3. Peet, J.; Kim, J. Y.; Coates, N. E.; Ma, W. L.; Moses, D.; Heeger, A. J.; Bazan, G. C., Efficiency enhancement in low-bandgap polymer solar cells by processing with alkane dithiols. *Nat. Mater.* **2007**, 6, 497-500.

4. Yao, Y.; Hou, J.; Xu, Z.; Li, G.; Yang, Y., Effects of Solvent Mixtures on the Nanoscale Phase Separation in Polymer Solar Cells. *Adv. Funct. Mater.* **2008**, 18, 1783-1789.

5. An, T. K.; Kang, I.; Yun, H. J.; Cha, H.; Hwang, J.; Park, S.; Kim, J.; Kim, Y. J.; Chung, D. S.; Kwon, S. K.; Kim, Y. H.; Park, C. E., Solvent additive to achieve highly ordered nanostructural semicrystalline DPP copolymers: toward a high charge carrier mobility. *Adv. Mater.* **2013**, 25, 7003-9.

6. Liao, H.-C.; Ho, C.-C.; Chang, C.-Y.; Jao, M.-H.; Darling, S. B.; Su, W.-F., Additives for morphology control in high-efficiency organic solar cells. *Mater. Today* **2013**, 16, 326-336.

7. Perez, L. A.; Rogers, J. T.; Brady, M. A.; Sun, Y.; Welch, G. C.; Schmidt, K.; Toney, M. F.; Jinnai, H.; Heeger, A. J.; Chabinyc, M. L.; Bazan, G. C.; Kramer, E. J., The Role of Solvent Additive Processing in High Performance Small Molecule Solar Cells. *Chem. Mater.* **2014**, 26, 6531-6541.

8. Kwon, S.; Kang, H.; Lee, J.-H.; Lee, J.; Hong, S.; Kim, H.; Lee, K., Effect of Processing Additives on Organic Photovoltaics: Recent Progress and Future Prospects. *Adv. Energy Mater.* **2017**, 7, 1601496.

9. Zhao, W.; Zhang, S.; Zhang, Y.; Li, S.; Liu, X.; He, C.; Zheng, Z.; Hou, J., Environmentally Friendly Solvent-Processed Organic Solar Cells that are Highly Efficient and Adaptable for the Blade-Coating Method. *Adv. Mater.* **2018**, 30.

10. Zhang, Y.; Voth, G. A., Combined Metadynamics and Umbrella Sampling Method for the Calculation of Ion Permeation Free Energy Profiles. *J. Chem. Theory Comput.* **2011**, 7, 2277-2283.

11. Widom, B., Some Topics in the Theory of Fluids. *J. Chem. Phys.* **1963**, 39, 2808-2812.

12. Kofke, D. A., Gibbs-Duhem integration: a new method for direct evaluation of phase coexistence by molecular simulation. *Mol. Phys.* **1993**, 78, 1331-1336.

13. Lai, P. K.; Hsieh, C. M.; Lin, S. T., Rapid determination of entropy and free energy of mixtures from molecular dynamics simulations with the two-phase thermodynamic model. *PCCP* **2012**, 14, 15206-13.

14. Lin, S.-T.; Blanco, M.; Goddard, W. A., The two-phase model for calculating thermodynamic properties of liquids from molecular dynamics: Validation for the phase diagram of Lennard-Jones fluids. *J. Chem. Phys.* **2003**, 119, 11792-11805.

15. Michael P. Allen, D. J. T., *Computer simulation of liquids*. 2 ed.; Oxford: Oxford University Press, 2017.

16. Pascal, T. A.; Goddard, W. A., Hydrophobic Segregation, Phase Transitions and the Anomalous Thermodynamics of Water/Methanol Mixtures. *J. Phys. Chem. B* **2012**, 116, 13905-13912.

17. Zhang, H.; Duquesne, M.; Godin, A.; Niedermaier, S.; Palomo del Barrio, E.; Nedea, S. V.; Rindt, C. C. M., Experimental and in silico characterization of xylitol as seasonal heat storage material. *Fluid Phase Equilib.* **2017**, 436, 55-68.

18. Minakov, D. V.; Levashov, P. R.; Fokin, V. B., Vibrational spectrum and entropy in simulation of melting. *Computational Materials Science* **2017**, 127, 42-47.

19. Vadhana, V.; Ayappa, K. G., Structure and Dynamics of Octamethylcyclotetrasiloxane Confined between Mica Surfaces. *J. Phys. Chem. B* **2016**, 120, 2951-67.

20. Lim, H. K.; Lee, H.; Kim, H., A Seamless Grid-Based Interface for Mean-Field QM/MM Coupled with Efficient Solvation Free Energy Calculations. *J. Chem. Theory Comput.* **2016**, 12, 5088-5099.

21. Caro, M. A.; Laurila, T.; Lopez-Acevedo, O., Accurate schemes for calculation of thermodynamic properties of liquid mixtures from molecular dynamics simulations. *J. Chem. Phys.* **2016**, 145, 244504.

22. Wang, J.; Chakraborty, B.; Eapen, J., Absolute thermodynamic properties of molten salts using the two-phase thermodynamic (2PT) superpositioning method. *PCCP* **2014**, 16, 3062-9.

23. Pascal, T. A.; Goddard, W. A., 3rd, Interfacial thermodynamics of water and six other liquid solvents. *J. Phys. Chem. B* **2014**, 118, 5943-56.

24. Lai, P.-K.; Lin, S.-T., Rapid determination of entropy for flexible molecules in condensed phase from the two-phase thermodynamic model. *RSC Advances* **2014**, 4.

25. Chen, M.; Pendrill, R.; Widmalm, G.; Brady, J. W.; Wohlert, J., Molecular Dynamics Simulations of the Ionic Liquid 1-n-Butyl-3-Methylimidazolium Chloride and Its Binary Mixtures with Ethanol. *J. Chem. Theory Comput.* **2014**, 10, 4465-79.

26. Li, C.; Medvedev, G. A.; Lee, E.-W.; Kim, J.; Caruthers, J. M.; Strachan, A., Molecular dynamics simulations and experimental studies of the thermomechanical response of an epoxy thermoset polymer. *Polymer* **2012**, 53, 4222-4230.

27. Jeon, J.; Kim, H.; Goddard, W. A., 3rd; Pascal, T. A.; Lee, G. I.; Kang, J. K., The Role of Confined Water in Ionic Liquid Electrolytes for Dye-Sensitized Solar Cells. *J. Phys. Chem. Lett.* **2012**, 3, 556-9.

28. Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D., Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, 8, 61-74.

29. Pascal, T. A.; Lin, S. T.; Goddard, W. A., 3rd, Thermodynamics of liquids: standard molar entropies and heat capacities of common solvents from 2PT molecular dynamics. *PCCP* **2011**, 13, 169-81.

30. Pascal, T. A.; He, Y.; Jiang, S.; Goddard, W. A., Thermodynamics of Water Stabilization of Carboxybetaine Hydrogels from Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2011**, 2, 1757-1760.

31. Huang, S. N.; Pascal, T. A.; Goddard, W. A., 3rd; Maiti, P. K.; Lin, S. T., Absolute Entropy and Energy of Carbon Dioxide Using the Two-Phase Thermodynamic Model. *J. Chem. Theory Comput.* **2011**, 7, 1893-901.

32. Berens, P. H.; Mackay, D. H. J.; White, G. M.; Wilson, K. R., Thermodynamics and quantum corrections from molecular dynamics for liquid water. *J. Chem. Phys.* **1983**, 79, 2375-2389.

33. Brochu, E.; Cora, V. M.; de Freitas, N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. arXiv e-prints2010; https://ui.adsabs.harvard.edu/abs/2010arXiv1012.2599B (accessed December 01, 2010).

34. Frazier, P. I. A Tutorial on Bayesian Optimization. arXiv e-prints2018; https://ui.adsabs.harvard.edu/abs/2018arXiv180702811F (accessed July 01, 2018).

35. Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R. P.; Freitas, N. d., Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE* **2016**, 104, 148-175.

36. Duvenaud, D.; Lloyd, J. R.; Grosse, R.; Tenenbaum, J. B.; Ghahramani, Z. Structure Discovery in Nonparametric Regression through Compositional Kernel Search. arXiv e-prints2013; https://ui.adsabs.harvard.edu/abs/2013arXiv1302.4922D (accessed February 01, 2013).

37. Oh, C.; Gavves, E.; Welling, M. BOCK : Bayesian Optimization with Cylindrical Kernels. arXiv e-prints2018; https://ui.adsabs.harvard.edu/abs/2018arXiv180601619O (accessed June 01, 2018).

38. Duvenaud, D. Automatic model construction with Gaussian processes. University of Cambridge, 2014.

39. Carl Edward Rasmussen, C. K. I. W., In; The MIT Press: 2006, pp 1-272.

40. Sesha Sarath Pokuri, B.; Lofquist, A.; Risko, C. M.; Ganapathysubramanian, B. PARyOpt: A software for Parallel Asynchronous Remote Bayesian Optimization. arXiv e-prints2018; https://ui.adsabs.harvard.edu/abs/2018arXiv180904668S (accessed September 01, 2018).

41. M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox *Gaussian 09, Revision A.02*, Gaussian, Inc.: Wallingford CT, 2016.

42. Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G., Charge Model 5: An Extension of Hirshfeld Population Analysis for the Accurate Description of Molecular Interactions in Gaseous and Condensed Phases. *J. Chem. Theory Comput.* **2012**, 8, 527-541.

43. Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R., GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computational Physics Communications* **1995**, 91, 43-56.

44. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, 4, 435-447.

45. Bernardes, C. E.; Joseph, A., Evaluation of the OPLS-AA Force Field for the Study of Structural and Energetic Aspects of Molecular Organic Crystals. *J. Phys. Chem. A* **2015**, 119, 3023-3034.

46. William L. Jorgensen, D. S. M., and Julian Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, 118, 11225-11236.

47. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, 81, 3684-3690.

48. Parrinello, M.; Rahman, A., Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, 52, 7182-7190.

49. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, 103, 8577-8593.

50. Caro, M. A.; Lopez-Acevedo, O.; Laurila, T., Redox Potentials from Ab Initio Molecular Dynamics and Explicit Entropy Calculations: Application to Transition Metals in Aqueous Solution. *J. Chem. Theory Comput.* **2017**, 13, 3432-3441.

51. Hildebrand, J. H.; Scott, R. L., The Entropy of Solution of Nonelectrolytes. *J. Chem. Phys.* **1952**, 20, 1520-1521.

52. Hansen, C. M., The Three Disensional Solubility Parameter and Solvent Diffusion Coefficient. *Journal of Paint Technology* **1967**, 39.

53. Belmares, M.; Blanco, M.; Goddard, W. A., 3rd; Ross, R. B.; Caldwell, G.; Chou, S. H.; Pham, J.; Olofson, P. M.; Thomas, C., Hildebrand and Hansen solubility parameters from molecular dynamics with applications to electronic nose polymer sensors. *J. Comput. Chem.* **2004**, 25, 1814-26.

54. Tummala, N. R.; Mehraeen, S.; Fu, Y.-T.; Risko, C.; Brédas, J.-L., Materials-Scale Implications of Solvent and Temperature on [6,6]-Phenyl-C61-butyric Acid Methyl Ester (PCBM): A Theoretical Perspective. *Adv. Funct. Mater.* **2013**, 23, 5800-5813.

55. Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976**, 22, 245-268.

56. Ben-Naim, A.; Marcus, Y., Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **1984**, 81, 2016-2027.

57. Einstein, A., Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* **1905**, 322, 549-560.

58. Jorgensen, W. L.; Schyman, P., Treatment of Halogen Bonding in the OPLS-AA Force Field; Application to Potent Anti-HIV Agents. *J. Chem. Theory Comput.* **2012**, 8, 3895-3801.

59. Machui, F.; Abbott, S.; Waller, D.; Koppe, M.; Brabec, C. J., Determination of Solubility Parameters for Organic Semiconductor Formulations. *Macromol. Chem. Phys.* **2011**, 212, 2159-2165.

60. Hansen, C. M., *Hansen Solubility Parameters (A User's Handbook)*. 2 ed.; CRC Press: 2007.

61. Zhao, J.; Li, Y.; Yang, G.; Jiang, K.; Lin, H.; Ade, H.; Ma, W.; Yan, H., Efficient organic solar cells processed from hydrocarbon solvents. *Nature Energy* **2016**, 1, 15027.