Benchmarking density functionals for hydrogen-helium mixtures with quantum Monte Carlo: Energetics, pressures, and forces

Raymond C. Clay III,[1,2,*] Markus Holzmann,[3] David M. Ceperley,[1] and Miguel A. Morales[2]
[1]*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[2]*Lawrence Livermore National Laboratory, Livermore, California 94550, USA*
[3]*LPMMC, UMR 5493 of CNRS, Université Grenoble Alpes, F-38100 Grenoble, France*
(Received 20 August 2015; revised manuscript received 5 November 2015; published 19 January 2016)

An accurate understanding of the phase diagram of dense hydrogen and helium mixtures is a crucial component in the construction of accurate models of Jupiter, Saturn, and Jovian extrasolar planets. Though density-functional-theory-based first-principles methods have the potential to provide the accuracy and computational efficiency required for this task, recent benchmarking in hydrogen has shown that achieving this accuracy requires a judicious choice of functional, and a quantification of the errors introduced. In this work, we present a quantum Monte Carlo (QMC) -based benchmarking study of a wide range of density functionals for use in hydrogen-helium mixtures at thermodynamic conditions relevant for Jovian planets. Not only do we continue our program of benchmarking energetics and pressures, but we deploy QMC-based force estimators and use them to gain insight into how well the local liquid structure is captured by different density functionals. We find that TPSS, BLYP, and vdW-DF are the most accurate functionals by most metrics, and that the enthalpy, energy, and pressure errors are very well behaved as a function of helium concentration. Beyond this, we highlight and analyze the major error trends and relative differences exhibited by the major classes of functionals, and we estimate the magnitudes of these effects when possible.

DOI: 10.1103/PhysRevB.93.035121

I. INTRODUCTION

Since the first exoplanet was discovered in 1988, our understanding of other solar systems has been revolutionized by major advancements in observational techniques. As a result of the Kepler mission and other planetary surveys, there are almost 2000 catalogued extrasolar planets [1]. A large fraction of these are Jovian planets and brown dwarf stars whose compositions are more than 90% hydrogen and helium—similar to Jupiter and Saturn. Understanding the evolution of these extrasolar planetary systems would be greatly aided by accurate models. Current planetary models for Jovian planets use a few observables, such as mass, radius, luminosity, and element compositions, to determine the interior structure and time evolution of the planet. However, in order to construct accurate models, one needs an accurate equation of state for hydrogen-helium mixtures at all temperatures, pressures, and species concentrations relevant in planetary interiors.

In the absence of experimental data at these extreme conditions, perturbative methods [2], chemical models [3], and *ab initio* methods [4–9] have done an excellent job in identifying phases that could be relevant in planetary models. For instance, dense hydrogen is widely believed to be in a metallic liquid phase at temperatures and pressures appropriate to the core regions of Jovian planets, and it should be the source of the planetary dynamo responsible for their large magnetic fields. An important question is whether helium is soluble in dense liquid hydrogen. As a Jovian planet cools, homogeneous H+He mixtures could condense into helium-rich droplets, which would rain down deeper into the planet [10,11]. This process could provide an additional energy dissipation mechanism and alter the mass distribution of helium and heavier elements, and so it should be treated accurately in planetary models.

Unfortunately, there are enough quantitative and qualitative differences between *ab initio* and chemical-model-based equations of state that planetary scientists routinely treat the equation of state as a free parameter in their models—to be varied to match the experimental data [12,13]. Recent models that optimize the phase boundaries for helium immiscibility and hydrogen metallization show an excellent ability to reproduce the observed atmospheric depletion of He and excess luminosity in Saturn, the gravitational moments, and the estimated ages of both Jupiter and Saturn [12,13]. However, this gives rise to uncertainties in other areas of the model, e.g., the core mass and composition, the distribution of heavier elements throughout the planet, etc. These uncertainties would be reduced by an accurate equation of state.

To construct accurate *ab initio* equations of state with well-quantified errors, one needs to understand and accommodate two frequently made approximations. The first is the ideal mixing approximation for the entropy. Used in chemical models and *ab initio* calculations since the 1970s, the validity of ideal mixing has only recently been investigated in the context of *ab initio* simulations using thermodynamic integration (TI) techniques [8,9]. There is currently a quantitative discrepancy between the works of Morales *et al.* and Lorenzen *et al.* [6,7] amounting to approximately 1000 K in the demixing transition at high pressures, and significant qualitative disagreement at lower pressures. With current computational resources, this source of error can be effectively eliminated through the use of TI.

The second and least understood source of error is in the treatment of electronic correlation effects, typically through the use of an approximate exchange correlation functional

*rcclay2@illinois.edu

within density-functional theory (DFT). Various studies have found that the pure hydrogen phase diagram is extremely sensitive to the choice of exchange correlation functional, primarily because of the presence of multiple molecular disassociation and metallization phase transitions [14–21]. In contrast, little is known about the impact that density-functional errors have on the demixing temperature in dense H+He mixtures.

More accurate alternative methods for solving the electronic-structure problem do exist. In particular, projector-based quantum Monte Carlo methods such as diffusion Monte Carlo (DMC) and reptation Monte Carlo (RMC) are known to be highly accurate variational methods. Given a starting guess for a trial wave function $|\Psi_T\rangle$, these methods work by stochastically projecting and sampling the ground-state wave function. In practical calculations, simulating systems with many electrons requires that the nodes of the trial wave function be guessed and fixed as a boundary condition to avoid a sign problem (the "fixed-node approximation"). Experience dictates that the resulting fixed-node wave function $|\Psi_{\mathrm{FN}}\rangle$ dramatically improves the already reasonable trial wave function used at the variational Monte Carlo (VMC) level. This has been shown to be very accurate in low-Z materials, successfully capturing the majority of subtle electronic correlation effects such as dispersion and hydrogen bonding. However, QMC is about two orders of magnitude more expensive than DFT calculations. This greatly restricts its applications toward computing more complex properties relevant to planetary physics, such as electrical conductivity and viscosity.

In this work, we use projector quantum Monte Carlo methods to benchmark a range of density functionals in thermodynamic regimes relevant for helium sedimentation in Jovian planets. Our main objective is to identify and understand qualitative error trends and relative differences between various classes of density functionals when used to estimate thermodynamic quantities in hydrogen-helium mixtures. To achieve this, we benchmark the errors occurring in the energetics, pressures, and forces for each functional, and we note how they change as a function of both density and helium concentration. In Sec. II, we discuss computational details. In Sec. III, we present the benchmarking results for global and local energetics, pressures, enthalpies, and forces. In Sec. IV, we explain the error trends we observe in terms of the underlying exchange functional, after which we conclude.

II. METHOD

In this study, we employ the same general methodology that we used in our work on pure hydrogen [16]; here, we discuss the computational details particular to hydrogen-helium mixtures.

A. Test sets

The relevant thermodynamic variables for describing the H+He phase diagram are the density ρ, the temperature T, and the helium species fraction x_{He}, defined as

$$x_{\mathrm{He}} = \frac{N_{\mathrm{He}}}{N_{\mathrm{H}} + N_{\mathrm{He}}}. \quad (1)$$

We also specify the density by r_s defined as $\Omega/N_e = \frac{4}{3}\pi r_s^3$, where N_e is the total number of electrons and Ω is the volume.

Our goal is to benchmark the accuracy of density functionals for predicting the helium demixing transition in Saturn and Jupiter. Thus, our test sets were chosen to represent the densities $r_s = 1.10$, 1.25, and 1.34, all at a temperature of 7000 K. We considered hydrogen helium mixtures with helium species fractions of 0–20.7 % and 100%. Calculations used cubic cells with 64 electrons with differing numbers of H and He ions to ensure charge neutrality. At each density and helium concentration, 20 statistically independent samples were generated from *ab initio* quantum molecular-dynamics (QMD) simulations in the NVT ensemble. These QMD simulations were performed in VASP [22–25] using the Perdew-Burke-Ernzerhof (PBE) [26] exchange correlation functional and classical nuclei. As discussed in our previous hydrogen benchmarking paper, the choice of functional to generate the test sets is not as important as it is for the configurations to be "physically reasonable" for the temperatures and pressures under consideration. Through its extensive use in this field and its sound construction, PBE meets the criterion of "physically reasonable" for test set purposes, even if other functionals produce more accurate local structures.

B. Density-functional comparison

For all configurations, we calculated the total energy, pressures, and forces using the following functionals in QUANTUM ESPRESSO [27] (note that the flavor of the DFT function is given in italics): LDA [28]; (*GGA*) PBE, revPBE [29], PBEsol [30], BLYP [31], Wu-Cohen [32], (metaGGA) M06L [33], TPSS [34]; (*nonlocal dispersion corrected*) vdW-DF [35], vdW-DF2 [36], vdW-DF-C09, vdW-DF2-C09 [37], vdW-DF-CX [38], vdW-optB86B [39], and vdW-optB88 [40]. The local-density approximation (LDA) and generalized-gradient approximation (GGA) -type functionals require the least computer time, whereas the metaGGA and nonlocal van der Waals functionals are about four times slower than GGA functionals.

For all the above functionals, we used a plane-wave cutoff of 800 Ry and a $7 \times 7 \times 7$ Monkhorst-Pack grid with an offset. We used hard Troullier-Martins pseudopotentials [41] (no core electrons for either H or He) generated with opium [42] using the PBE functional. To ensure no pseudopotential overlap in our test set, we chose real-space cutoffs of $r_c = 0.37a_0$ and $r_c = 0.5a_0$ for the H and He pseudopotentials, respectively.

We also tested the exact-exchange Heyd-Scuseria-Ernzerhof (HSE) functional. However, due to computational and memory limitations, we took the following cost-saving measures. At every density and helium concentration we considered, we performed only 10 HSE calculations, reduced the Monkhorst-Pack grid to $5 \times 5 \times 5$ (the same grid used to evaluate the Fock operator), and ran the calculations with VASP because of its compact projector augmented wave (PAW) formalism. A plane-wave cutoff of 1500 eV and 96 bands were used for all calculations. We found that this gave the desired accuracy for energy and pressure differences within configurations at the same density and helium concentration, however comparisons between different densities and helium concentrations might be slightly underconverged—not

exceeding 1.5 mHa/electron for the energies and 1 GPa for the pressures. Since we had a limited choice of pseudopotentials, we were unable to perform calculations at $r_s = 1.10$ and guarantee the desired level of accuracy for VASP calculations.

C. Quantum Monte Carlo calculations

All quantum Monte Carlo calculations were performed using the QMCPACK [43,44] simulation package with a Slater-Jastrow trial function. Singe-particle orbitals were obtained from QUANTUM ESPRESSO [27] using the PBE functional, a plane-wave cutoff of 200 Ry, and the same Troullier-Martins pseudopotentials described previously. All QMC calculations were "all-electron": we used the bare Coulomb interaction between electrons and electrons, and between electrons and nuclei. The pseudopotentials were only used to generate the trial function orbitals within the DFT calculation.

We used short-ranged one-body and two-body functions of b-spline form for the Jastrow factor. For H and He, the one-body terms were spin-independent. The one-body term for each species was a sum of two functions. The first was a "core" Jastrow, which had a real-space cutoff of $r_c = 1.0a_0$, eight knots, and the suitable electron-ion cusp condition imposed. The second had a cutoff of $L/2$ with eight knots and no cusp-condition imposed. For the two-body functions, we separately included same-spin and opposite-spin *e-e* terms, each with a cutoff of $r_c = L/2$ and correct cusp conditions imposed.

Our wave functions were optimized with the linear method [45]. After obtaining a good initial guess for the Jastrow parameters from a single $r_s = 1.10$ configuration with four He atoms, all variational parameters were simultaneously optimized using an initial variance minimization step, followed by 10 energy minimization steps. Convergence of the minimization procedure was verified.

Energies, pressures, and the structure factor were calculated using reptation Monte Carlo (RMC) [46,47]. Our target statistical error bars for the energies and pressures were 0.008 mHa/electron and 0.3 GPa, respectively. For all but the pure helium configurations, we used a time step of $\tau = 0.0075$ Ha^{-1} and a projection time of $\beta = 4.5$ Ha^{-1}. These choices were found to yield time-step and mixed-estimator errors for the potential energy within the desired error bars. For the pure helium configurations, we fixed the projection time at $\beta = 4.5$ Ha^{-1} and ran with time steps of $\tau = 0.0075$ and $0.003\,75$ Ha^{-1}. All estimated properties were then linearly extrapolated to zero time step.

Forces were computed using the Chiesa, Ceperley, and Zhang estimator [48] adapted to periodic boundary conditions. We used a real-space cutoff of $\mathcal{R} = 1.0a_0$ and a smoothing polynomial of degree $M = 3$. Based on several statistical tests detailed in the supplemental information [49], we found that this choice of parameters yielded systematic errors that were less than the error bar on the hydrogen force components; approximately 2 mHa/bohr. This resolution was sufficient to distinguish between different functionals. We used diffusion Monte Carlo with a time step of $\tau = 0.01$ Ha^{-1} and a population size of 512 walkers, which we found converged the local energy (but not the local potential energy) to within error bars. To correct for the mixed estimator of the density, we used extrapolated force estimates. All systematic errors are expected to be less than the statistical error bar.

To minimize finite-size errors, we first used canonical twist-averaged boundary conditions (TABCs) on a $4 \times 4 \times 4$ Monkhorst-Pack (MP) grid [50]. For the potential energy correction, we used the leading-order Chiesa correction based on pure estimates of $S_{ee}(k)$ [51,52]. Details of the kinetic energy correction are given in the supplemental information. Since most configurations are in the metallic state, there is also a kinetic energy error arising from the fact that we are attempting to reproduce a Fermi surface with only 64 electrons. To correct for this, we used the PBE functional to estimate the energy error between a twist-averaged unit cell and a $7 \times 7 \times 7$ MP grid. This correction scheme was tested against several supercell calculations at $r_s = 1.10$ and 1.34 across all helium concentrations. We expect the absolute energy errors (across all densities and helium concentrations) to be less than 0.5 mHa/electron, and the pressure errors to be approximately 1 GPa. Twist-averaging was also used to minimize of the finite-size error in the forces. There could be a residual error coming from the fact that $\rho(\mathbf{r})$ is not necessarily converged to the thermodynamic limit. By comparing the forces from a twist-averaged Kwee-Zhang-Krakauer density-functional [53] estimate in the supercell with those of a converged $7 \times 7 \times 7$ MP grid, we estimate the standard deviation of the residual finite-size error in the force estimates to be approximately 2 mHa/bohr.

D. Error analysis

1. Scalar quantities

As in our previous paper, we consider a test set S with M configurations, $\{\mathbf{R}_0, \ldots, \mathbf{R}_M\}$. For each configuration $\mathbf{R}_i$, we define the density-functional error $\delta\mathcal{A}(\mathbf{R}_i) = \mathcal{A}^{\mathrm{DF}}(\mathbf{R}_i) - \mathcal{A}^{\mathrm{QMC}}(\mathbf{R}_i)$, where $\mathcal{A}$ is some observable (e.g., total energy, pressure), with "DF" the density functional, and "QMC" the QMC values.

In addition to defining average errors over a test set S, which we denote $\langle \delta\mathcal{A}^{\mathrm{DF}} \rangle_S$, we define a general class of shifted mean absolute errors as

$$\langle |\widetilde{\delta\mathcal{A}}| \rangle_S = \frac{1}{M} \sum_{\mathbf{R}_i \in S} |\delta\mathcal{A}^{\mathrm{DF}}(\mathbf{R}_i) - c^{\mathrm{DF}}|. \tag{2}$$

Here, c^{DF} is a density-functional-dependent offset. The standard "mean absolute error" corresponds to the choice of $c^{\mathrm{DF}} = 0$, which we will label as $\langle |\delta\mathcal{A}| \rangle_S$. However, in this paper we will make other choices of c^{DF} for different observables, specifically for measures of "global" and "local" energetics.

2. Forces

Let $\mathbf{f}_i$ denote the force on ion i, and $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N\}$ the $3N$-dimensional vector of all ionic force components. Because of the large number of force components, we construct a few overall measures of the force errors. One of the simplest is the mean absolute force error $\langle |\delta\mathbf{f}^{\mathrm{DF}}| \rangle_S$: the ensemble average magnitude of the force error vector $\delta\mathbf{f}_i = \mathbf{f}_i^{\mathrm{DF}} - \mathbf{f}_i^{\mathrm{QMC}}$.

The mean force error can be related to how a DFT functional could distort the local structure of the system. Consider the "potential of mean force" $w(r)$, defined in terms of the

pair correlation function by $g(r) = \exp[-\beta w(r)]$. For weak perturbations, changes in $g(r)$ can be related to changes in $w(r)$ and hence to errors in the average forces between ions separated by a distance r. For each density functional, we will define the following error measure. Let i_μ and j_ν denote two particles of species μ and ν, respectively. Denoting $\mathbf{r}_{i_\mu j_\nu} = \mathbf{r}_{i_\mu} - \mathbf{r}_{j_\nu}$, we define

$$\left\langle \delta f^{\mathrm{DF}}_{\mu\text{-}\nu}(r) \right\rangle = \frac{\int d\mathbf{R}\, e^{-\beta E^{\mathrm{PBE}}(\mathbf{R})} \delta\left(r - r_{i_\mu j_\nu}\right) \hat{\mathbf{r}}_{i_\mu j_\nu} \cdot \delta \mathbf{f}^{\mathrm{DF}}_{i_\mu}}{\int d\mathbf{R}\, e^{-\beta E^{\mathrm{PBE}}(\mathbf{R})} \delta\left(r - r_{i_\mu j_\nu}\right)}. \quad (3)$$

Based on this definition, if $\langle \delta f^{\mathrm{DF}}_{\mu\text{-}\nu}(r) \rangle$ is positive (negative), it overbinds (underbinds) species of type μ and ν at a distance r.

Note that we use E^{PBE} in this definition, since our configurations are sampled from QMD using the PBE functional. However, if we could replace E^{PBE} with E^{QMC}, then $\langle \delta f^{\mathrm{DF}}_{\mu\text{-}\nu}(r) \rangle_{\mathrm{QMC}}$ would be related to the density-functional error in the potential of mean force $\delta w^{\mathrm{DF}}_{\mu\nu}(r)$ (relative to the QMC distribution) by

$$\left\langle \delta f^{\mathrm{DF}}_{\mu\text{-}\nu}(r) \right\rangle_{\mathrm{QMC}} = -\frac{\partial}{\partial r} \delta w^{\mathrm{DF}}_{\mu\nu}(r). \quad (4)$$

In any case, given that E^{QMC} and E^{PBE} produce qualitatively similar distributions of ionic configurations, $\langle \delta f^{\mathrm{DF}}_{\mu\text{-}\nu}(r) \rangle$ allows us to estimate when a density functional overbinds or underbinds, and how a density functional will affect the $g(r)$.

III. RESULTS

A. Global energetics

Suppose we are interested in assessing to what extent the average error in the total energy changes as a function of helium concentration. An error in the energy will change both the Helmholtz and Gibbs free energies of mixing, which could then bias the estimated location of the H+He immiscibility transition.

To measure this quantity, we define a measure of "global energetics" as follows: for a given ρ, we build an aggregated test set $S'(\rho)$, which is the union of all test sets at all helium concentrations with a given electronic density ρ. We then choose $c^{\mathrm{DF}}(\rho)$ to be the median of $\{\delta E^{\mathrm{DF}}\}_{S'(\rho)}$. We define the "global energetic error" $\langle |\widetilde{\delta E^{\mathrm{DF}}}| \rangle_{g,S}$ of the test set $S(\rho, x_{\mathrm{He}})$ by using Eq. (2).

In Fig. 1, we show the global energetic error for all functionals at three densities, averaged over all helium concentrations. We see that the best performing functionals over all densities are the meta-GGA functionals TPSS and M06-L, respectively, with global energetic errors that are less than half of those of PBE. After these, the best performing functionals are the semilocal GGA's BLYP and revPBE, followed by optB86b-vdW and vdW-DF. The worst performing functionals are LDA, HSE, PBEsol, and WC, with global energetic errors approximately twice that of PBE. Though PBE has a better than average performance in this regime, one can gain nearly a millihartree in accuracy by switching to a metaGGA or a properly tuned GGA.

It is also useful to consider how accurately the energy difference between systems with different helium concentrations are captured with a density functional. For

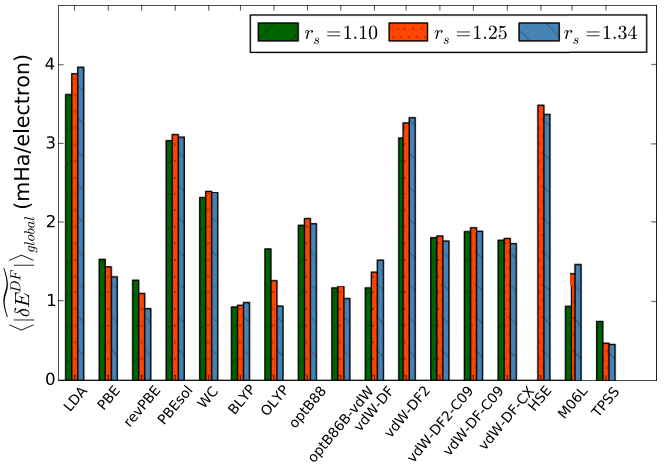

FIG. 1. $\langle |\widetilde{\delta E^{\mathrm{DF}}}| \rangle_{\mathrm{global}}$ averaged over all helium concentrations for all considered functionals. The different bar colors/patterns denote the different densities.

specificity, we look at $\langle \delta[E(x_{\mathrm{He}}, \rho) - E(x_{\mathrm{He}} = 1, \rho)] \rangle$ in Fig. 2 for all considered functionals and helium concentrations. Relative to the pure helium configurations, BLYP, vdW-DF, and vdW-DF2 overestimate the energy difference between the mixed hydrogen/helium configurations, whereas all other functionals underestimate this difference. Additionally, though all curves exhibit noticeable nonlinearity in the $x_{\mathrm{He}} = 0 - 20\,\%$ range, almost all curves have magnitudes that monotonically decrease to zero. The exception is TPSS, which changes sign at least once in the the range $x_{\mathrm{He}} = 0$–$20\,\%$.

In contrast to our previous hydrogen benchmarking study, BLYP performs noticeably better than vdW-DF in global energetics, whereas vdW-DF is now comparable to PBE. This

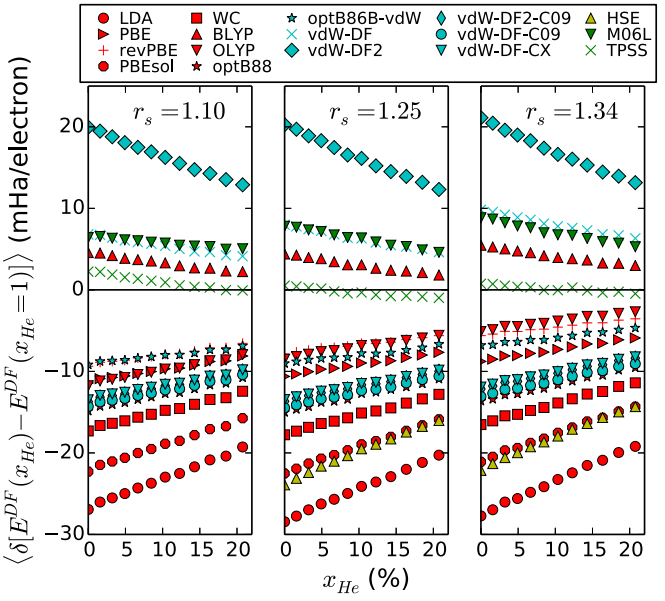

FIG. 2. $\langle \delta[E(x_{\mathrm{He}}) - E(x_{\mathrm{He}} = 1)] \rangle$ vs x_{He} for all considered functionals at $r_s = 1.10$ (left), $r_s = 1.25$ (middle), and $r_s = 1.34$ (right). All energies are measured relative to the average energy of all pure helium configurations at the specified density.

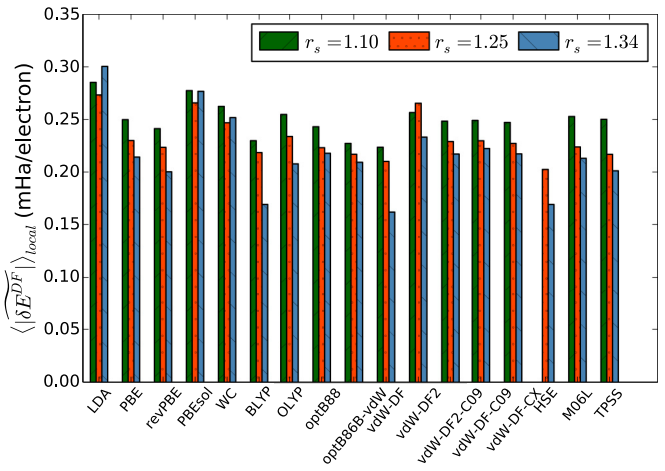

FIG. 3. $\langle|\widetilde{\delta E^{\mathrm{DF}}}|\rangle_{\mathrm{local}}$ averaged over all helium concentrations for all considered functionals. The different bar colors/patterns denote the different densities.

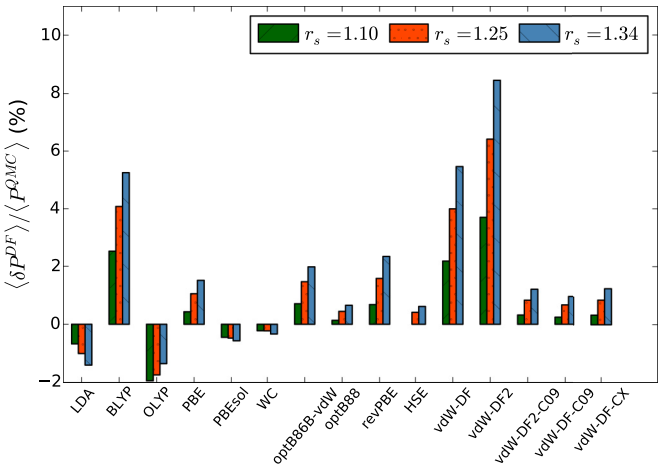

FIG. 4. $\langle \delta P^{\mathrm{DF}} \rangle / \langle P^{\mathrm{QMC}} \rangle$ in units of (%) averaged over all helium concentrations for all considered functionals. The different bar colors/patterns denote the different densities. Not shown: M06-L and TPSS.

does not pose a contradiction to our previous work, which we explain in Sec. IV A.

B. Local energetics

Even if the mean total energy error for a DFT-MD simulation is small, it is still possible for energy differences between similar configurations to have large errors. To measure the spread of the total energy error distribution, we use the following: for each test set $S(\rho, x_{\mathrm{He}})$ corresponding to a given density ρ and helium concentration x_{He}, we set c^{DF} to be the median of $\{\delta E^{\mathrm{DF}}\}_{S(\rho,x_{\mathrm{He}})}$. Using Eq. (2) with this choice gives our "local energy" measure, which we denote as $\langle|\widetilde{\delta E^{\mathrm{DF}}}|\rangle_{\ell,S}$. Note that this is similar to the "global energetic error" measure used previously, except now the test set S and the reference set S' are the same.

In Fig. 3, we show $\langle|\widetilde{\delta E^{\mathrm{DF}}}|\rangle_{\mathrm{local}}$ at three different densities for all functionals considered. We note that the characteristic error scale is now 0.25 mHa/electron, and that the differences between density functionals are significantly less pronounced than in the global energetic case. This cannot be used to discriminate between most functionals, but the outliers are worth mentioning. LDA, PBEsol, vdW-DF2, and M06L are the worst performers, whereas HSE is perhaps the best. These trends are consistent with the global energetic trends, with two notable exceptions. M06L accurately captures global energetics, and HSE exhibits poor global energetic performance. Though it is hard to tell whether BLYP or vdW-DF is better, the fact that they are comparable is consistent with our previous hydrogen-only benchmarking study.

C. Pressures

For all test sets and functionals, we computed $\langle \delta P^{\mathrm{DF}} \rangle_S$ and $\langle|\delta P^{\mathrm{DF}}|\rangle_S$. As previously observed in pure hydrogen, we found that $\langle|\delta P^{\mathrm{DF}}|\rangle_S$ is indistinguishable from the QMC statistical noise. $\langle \delta P^{\mathrm{DF}} \rangle_S$, on the other hand, can be quite sizable. Due to the wide range of pressures that occur across different helium concentrations and densities, we will plot relative mean pressure errors instead of absolute errors for the remainder of this section.

In Fig. 4, we plot $\langle \delta P^{\mathrm{DF}} \rangle_S / \langle P^{\mathrm{QMC}} \rangle$ (averaged over all helium concentrations) at three different densities for all functionals considered. The trend observed is very much the same as in our previous benchmarking studies of pure hydrogen: accurate energetics are compensated by poor pressure estimation. PBESol, Wu-Cohen, and LDA all have pressure errors less than 1% despite their poor energetic performance. The worst functionals for pressures happen to be vdW-DF and BLYP, which were known for their accurate energetics. vdW-DF2 is an exception in that it has poor energy and poor pressure estimation. Note that TPSS and M06-L functionals are not plotted because their pressure errors are so large as to be off the scale: averaged over all three densities, M06-L and TPSS have pressure errors of -31% and -17%, respectively, with errors increasing as the density is decreased.

In Fig. 5, we plot the mean pressure error versus helium concentration for all considered functionals. We observe that the dependence of the pressure errors on x_{He} is very well behaved across all densities and functionals. The magnitude for almost all functionals increases monotonically as x_{He} increases, reaching its maximum error for pure He. The exceptions are OLYP, where the pressure error has a positive slope and changes sign at some nonzero x_{He}, and WC, which reaches a maximum at some nonzero x_{He} and then decreases toward a minimum at $x_{\mathrm{He}} = 1$. Though we did not investigate helium concentrations higher than about 20%, the smoothness of the pressure errors as a function of x_{He} is reassuring, as it opens up the possibility of fitting these errors and correcting for them.

D. Enthalpies

The mean enthalpy error $\langle \delta H^{\mathrm{DF}} \rangle_S$ is given by $\langle \delta H^{\mathrm{DF}} \rangle_S = \langle \delta E^{\mathrm{DF}} \rangle_S + V \langle \delta P^{\mathrm{DF}} \rangle_S$, which we can combine with the results from Secs. III B and III C to study the errors in the predicted DFT enthalpy.

In Fig. 6, we show the average enthalpy error $\langle \delta H^{\mathrm{DF}} \rangle_S$ versus helium concentration for $r_s = 1.10$, 1.25, and 1.34. At fixed density ρ, H^{DF} is measured relative to the enthalpy of

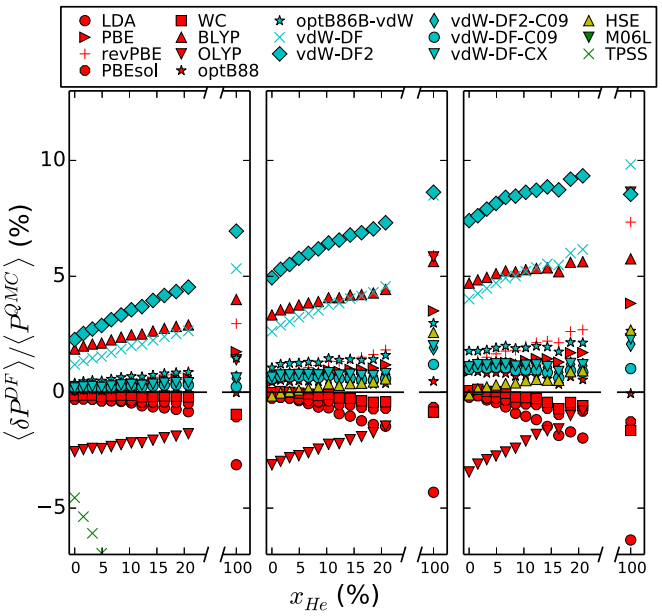

FIG. 5. $\langle \delta P^{\mathrm{DF}} \rangle / \langle P^{\mathrm{QMC}} \rangle$ in units of (%) vs x_{He} for all considered functionals at $r_s = 1.10$ (left), $r_s = 1.25$ (middle), and $r_s = 1.34$ (right). The different colors/shapes are given in the legend and denote the density functional. Note the broken axis, which shows the mean relative pressure error for the pure helium configurations.

pure helium at density ρ. What we see is for the most part qualitatively similar to what we saw for the energy errors in Fig. 2. However, the tested functionals possess varying degrees of error cancellation. Some functionals noticeably benefit from error cancellation, specifically vdW-DF, LDA,

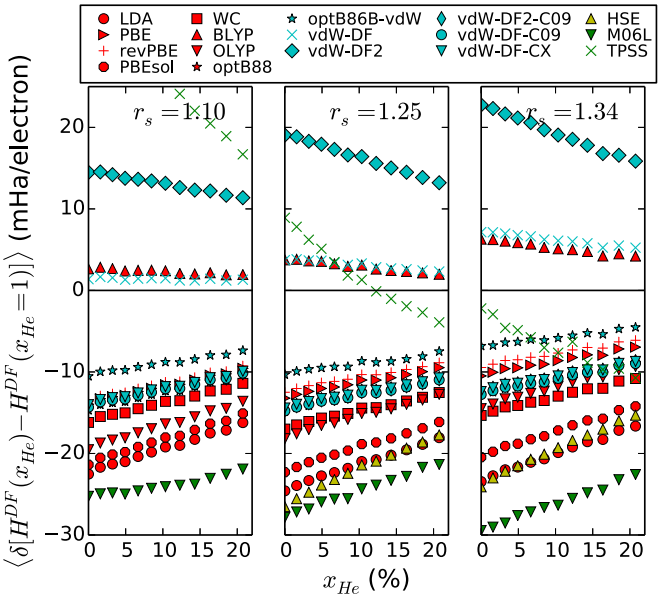

FIG. 6. $\langle \delta[H^{\mathrm{DF}}(x_{\mathrm{He}}) - H^{\mathrm{DF}}(x_{\mathrm{He}} = 1)] \rangle$ vs x_{He} for all considered functionals at $r_s = 1.10$ (left), $r_s = 1.25$ (middle), and $r_s = 1.34$ (right). All energies are measured relative to the average energy of all pure helium configurations at the specified density. Note that the reference enthalpy for each density is taken to be the mean enthalpy of the pure helium configurations at that density.

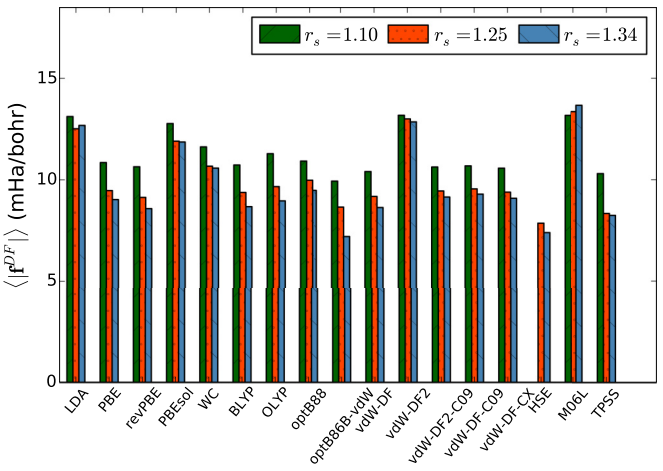

FIG. 7. $\langle |\delta \mathbf{f}_H^{\mathrm{DF}}| \rangle$ aggregated over all helium concentrations for all considered functionals. The different colors denote different densities.

and HSE. This can reduce the absolute enthalpy by as much as 3–4 mHa/electron depending on the functional and density. Others suffer from error *addition*, such as PBE and most dramatically OLYP, which has an enthalpy error almost 10 mHa/electron higher than its corresponding energy error. Lastly, there are some functionals that exhibit neither error cancellation nor addition, namely BLYP and all of the newer van der Waals functionals.

E. Forces

1. Total force errors

The most natural question we can ask ourselves is which functional has the most accurate forces on average. Toward that end, in Fig. 7 we compute $\langle |\delta \mathbf{f}^{\mathrm{DF}}| \rangle$ over all atoms and helium concentrations. Before interpreting Fig. 7, however, we mention an important caveat. Because of the large statistical noise relative to the magnitude of individual force components, any mean absolute error measure is going to be affected by the statistical noise. Thus, the *absolute magnitude* of the mean absolute force error is *not* directly related to the error in the density-functional force. However, because the statistical noise is identical across all functionals, we are able to compare different functionals and establish which ones are better and by how much.

With this in mind, we note in Fig. 7 that the trend in mean absolute force errors is similar to the trends we saw in the local energetic errors. Specifically, HSE is among the best, whereas vdW-DF2, LDA, and M06-L are among the worst. There are some slight differences, however. TPSS and optB86b-vdW seem to perform better for force error measures than the local energetics would suggest. Additionally, though BLYP and vdW-DF still seem to outperform PBE, the difference is not nearly as large as the local energetics would suggest.

2. Local force errors

We now consider how the average density-functional force errors depend on the distance between ions, $\langle \delta f_{\mu\text{-}\nu}^{\mathrm{DF}}(r) \rangle$, as described in Sec. II D 2. In principle, these force errors will not only depend on the density functional, but also on density

and helium concentration. We will address these later two points first, as they will greatly simplify the analysis.

The first question is, how does the average force change as a function of density? We show in Fig. 8 $\langle \delta f^{\mathrm{PBE}}_{\mu\text{-}\nu}(r) \rangle$ versus r/r_s calculated using the PBE functional. We consider H-H (top), H-He (middle), and He-He (bottom) forces. The first two were calculated at a helium fraction of 20.75%, while the He-He forces were computed in pure helium to enhance the statistical sampling. For each plot, we overlay the plots of $\langle \delta f^{\mathrm{DF}}_{\mu\text{-}\nu}(r) \rangle$ at the densities $r_s = 1.10$, 1.25, and 1.34. We observe no statistically significant density dependence in any of the major types of mean force errors.

The second question is, how does the average force change as a function of helium concentration? In Fig. 9, we show the same general plots of $\langle \delta f^{\mathrm{PBE}}_{\mu\text{-}\nu}(r) \rangle$ using the PBE functional as in Fig. 8, but this time overlaying plots of different helium concentrations instead of different densities. All plots were calculated at a density of $r_s = 1.25$. Note that within error bars, $\langle \delta f^{\mathrm{PBE}}_{\mu\text{-}\nu}(r) \rangle$ does not depend on helium concentration.

Given the insensitivity of the average force errors for PBE to both density and helium concentration, we plot in Fig. 10 $\langle \delta f^{\mathrm{DF}}_{\mu\text{-}\nu}(r) \rangle$ for all considered functionals. The helium fraction was chosen to be 1.6% for the H-H plot (top), 20.75% for the H-He plot (middle), and 100% for the He-He plot (bottom). The density was chosen to be $r_s = 1.25$. Recall from Sec. II D 2 that $\langle \delta f^{\mathrm{DF}}_{\mu\text{-}\nu}(r) \rangle > 0$ implies overbinding relative to QMC.

For the H-H forces at the top of Fig. 10, the BLYP, vdW-DF, and vdW-DF2 functionals all exhibit a strong propensity to overbind in the region $1 < r/r_s < 1.5$, with vdW-DF2 overbinding the most. TPSS overbinds the least in the region $1 < r/r_s < 1.2$ but then underbinds slightly up to $r/r_s = 2.2$. All other functionals underbind in the region $1 < r/r_s < 1.5$, with HSE underbinding the least and LDA the most. Though it is hard to tell with the noise, HSE has the lowest absolute error in the region $1 < r/r_s < 1.5$, followed by optB86b-vdW, vdW-DF, and BLYP, and then by the combination revPBE, PBE, vdW-DF-CX, vdW-DF-C09, and vdW-DF2-C09.

For the H-H forces at the top of Fig. 10, there seem to be three distinct regions, defined by when the DF errors cross the r axis. We refer to these as region I ($1 < r/r_s < 1.5$), region II ($1.5 < r/r_s < 2.2$), and region III ($r/r_s > 2.2$), and they correspond roughly to the first, second, and third coordination shells. Notice that with the exception of TPSS and M06-L, if a functional overbinds in region I, it will almost certainly underbind in region II and overbind again in region III. This is not entirely unexpected. The ion-ion force depends only on electron density, and so if two protons overbind because of an increased electronic charge between them, this decreases the electronic charge elsewhere, leading to underbinding in the charge-depleted region.

There are only a few functionals that overbind the H-H interaction in region I: BLYP, vdW-DF, and vdW-DF2, and TPSS (only for $r/r_s \approx 1$). The rest underbind, though to varying degrees. If we try to determine which functionals have the smallest error magnitudes in region I, we find that the trend is very similar to what we saw before in the mean absolute force and local energetic sections. HSE, vdW-DF, BLYP, TPSS, and optB86b-vdW have the smallest errors in regions I, though further discrimination is difficult given the error bars. In region II, on the other hand, HSE and optB86b-vdW seem to have measurably smaller error magnitudes than vdW-DF, BLYP, and TPSS.

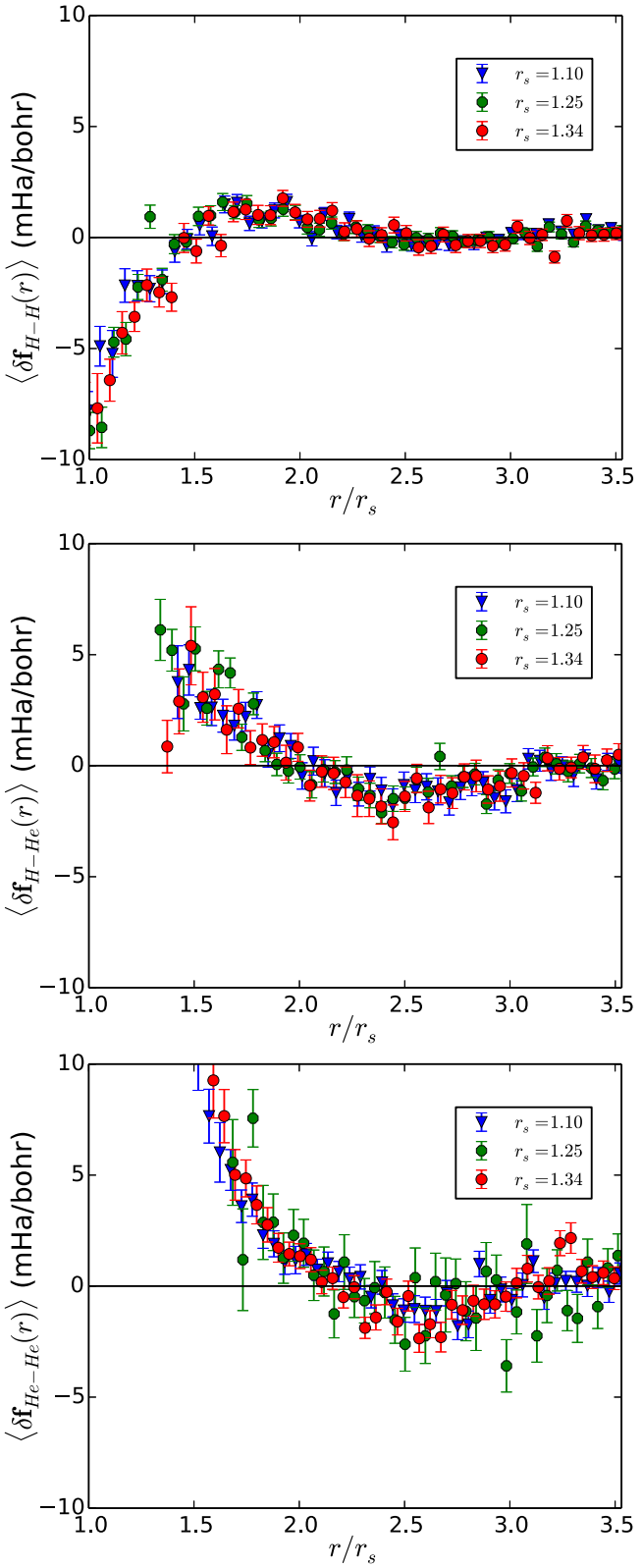

FIG. 8. $\langle \delta f^{\mathrm{PBE}}_{\mu\text{-}\nu}(r) \rangle$ vs r/r_s as density is changed. The different marker colors/styles represent different densities. Top: $\langle \delta f^{\mathrm{PBE}}_{\mathrm{H\text{-}H}}(r) \rangle$ calculated at $x_{\mathrm{He}} = 1.6\%$, middle: $\langle \delta f^{\mathrm{PBE}}_{\mathrm{H\text{-}He}}(r) \rangle$ calculated at $x_{\mathrm{He}} = 20.7\%$, and bottom: $\langle \delta f^{\mathrm{PBE}}_{\mathrm{He\text{-}He}}(r) \rangle$ calculated at $x_{\mathrm{He}} = 100\%$

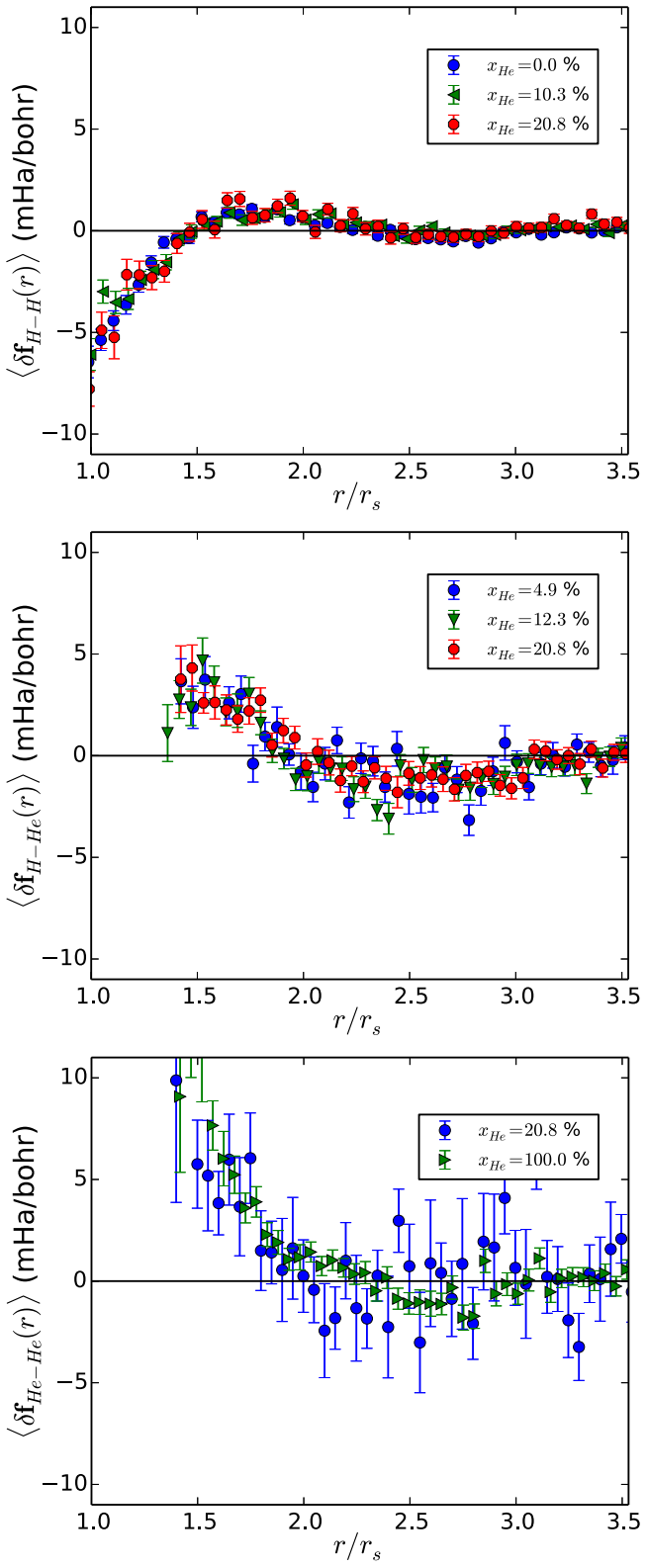

FIG. 9. $\langle \delta f^{\mathrm{PBE}}_{\mu\text{-}v}(r) \rangle$ vs r/r_s as helium concentration is changed. The different marker colors/styles represent different helium concentrations. Top: $\langle \delta f^{\mathrm{PBE}}_{\mathrm{H\text{-}H}}(r) \rangle$, middle: $\langle \delta f^{\mathrm{PBE}}_{\mathrm{H\text{-}He}}(r) \rangle$, and bottom: $\langle \delta f^{\mathrm{PBE}}_{\mathrm{He\text{-}He}}(r) \rangle$. All configurations are at a density of $r_s = 1.25$.

For the H-He forces in the middle of Fig. 10, the differences between different functionals are more striking. HSE and TPSS

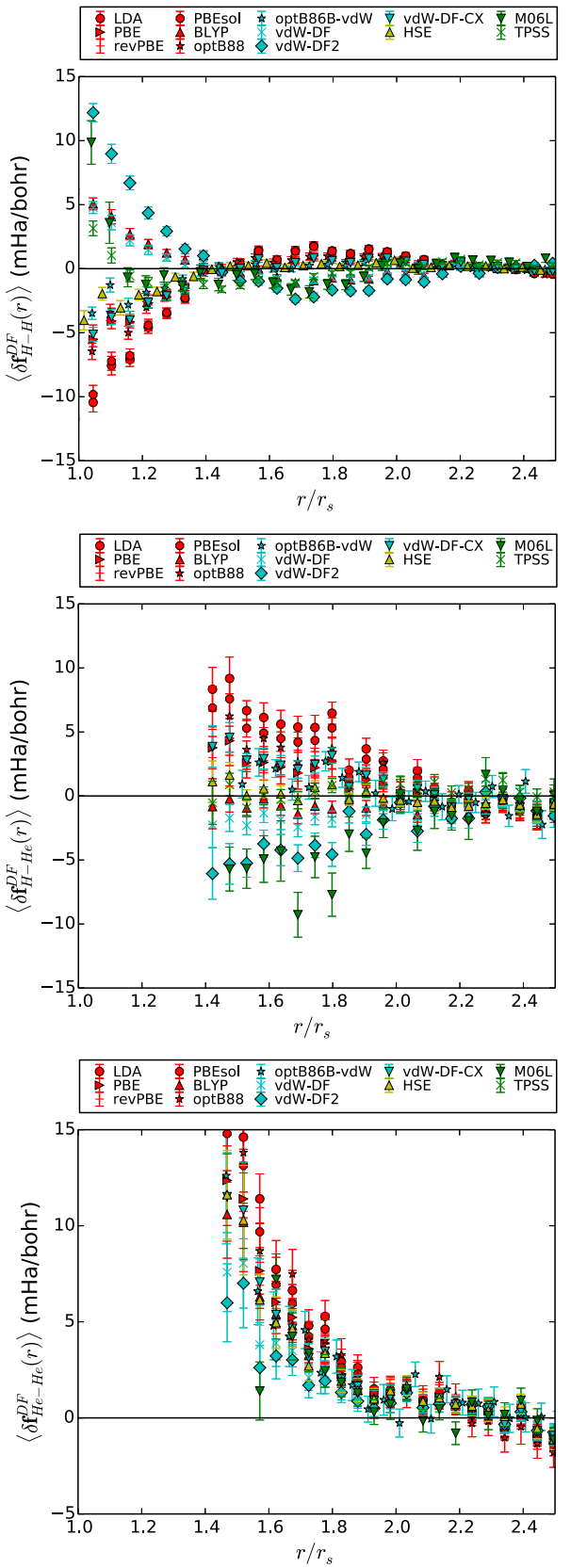

FIG. 10. $\langle \delta f^{\mathrm{DF}}_{\mu\text{-}v}(r) \rangle$ vs r/r_s as the functional is changed. The different marker colors/styles represent different density functionals. Top: $\langle \delta f^{\mathrm{PBE}}_{\mathrm{H\text{-}H}}(r) \rangle$ at $x_{\mathrm{He}} = 1.6\%$, middle: $\langle \delta f^{\mathrm{PBE}}_{\mathrm{H\text{-}He}}(r) \rangle$ at $x_{\mathrm{He}} = 20.7\%$, and bottom: $\langle \delta f^{\mathrm{PBE}}_{\mathrm{He\text{-}He}}(r) \rangle$ at $x_{\mathrm{He}} = 100\%$. All configurations are at a density of $r_s = 1.25$.

have the best average performance in the region $1.5 < r/r_s < 2.0$. However, BLYP also performs exceptionally well, slightly underbinding hydrogen-helium pairs by less than 1 mHa/bohr. The worst performing functionals are LDA, which overbinds the H-He interaction, and vdW-DF2, which underbinds.

Lastly, we consider the He-He forces at the bottom of Fig. 10. The error bars are somewhat large, but there are some obvious trends still visible. All functionals overbind the He-He interaction, although LDA overbinds the most. The functionals that overbind the least are vdW-DF2, HSE, or TPSS, followed by vdW-DF and then BLYP.

IV. DISCUSSION

A. Energetics

In this section, we will try to reconcile the observed differences between the local and global energetic trends. We believe that the trends observed in the local energetic errors are mostly described by the impact that the density functional has on the charge density and forces, as we will discuss in Sec. IV D. For now, we will try to grapple with how little bearing the local energetic errors had on the global energetic errors.

The main point to realize is that the global energetic errors are going to be dominated by errors in the energy differences between configurations with different helium concentrations. Calculating these types of energy differences accurately means having small total energy errors, error cancellation, or both. Error cancellation is possible only if the H-H, H-He, and He-He interactions err in a similar fashion—all overbinding or underbinding by roughly the same magnitude. For example, if the H-H interaction is overbound, the average energy difference between pure hydrogen and pure helium can still be accurately captured if the He-He interaction is also similarly overbound.

We observed from our force discussion that many functionals will overbind one type of interaction while underbinding another. Consider vdW-DF and BLYP. Both have very similar performances for forces and local energetics, with vdW-DF having a slight edge on both. However, BLYP has noticeably smaller global energetic errors than vdW-DF. Looking at the force errors, BLYP and vdW-DF overbind the H-H interaction in almost the exact same way. However, vdW-DF overbinds the He-He interaction noticeably less than BLYP does. This means that the BLYP will predict a higher total energy for pure He configurations than vdW-DF will. However, both vdW-DF and BLYP will overestimate the total energy of sampled pure hydrogen configurations by about the same amount. As a result of error cancellation, $\delta[E(x_{\mathrm{He}} = 0) - E(x_{\mathrm{He}} = 1)]$ will actually be larger for vdW-DF than it will be for BLYP. Extending this argument to intermediate helium concentrations explains why decent local energetics do not imply decent global energetics, since global energetics performance is largely determined by error cancellation.

B. Pressures

As in our previous hydrogen benchmarking work, we observed that favorable energetic errors were anticorrelated with favorable pressure errors—the most dramatic cases being LDA and vdW-DF/BLYP. Though this might seem paradoxical, the recognition of a tradeoff between accurate lattice constants and cohesive energies is well known. Perdew *et al.* [30] argue that this tradeoff is a necessary consequence of the limited form of the GGA functional. Assuming an exchange enhancement factor of the form $F_X(s) = \mu s^2 + \cdots$, accurate binding and atomization energies require a μ that is larger than what would be predicted by perturbation theory on the homogeneous electron gas. We find that almost all the best performing functionals for pressure have a μ that is chosen to ensure that the slowly varying homogeneous electron gas limit is recovered as $s \to 0$. Specifically, the LDA, Wu-Cohen, and PBEsol functionals satisfy this constraint.

The hybrid functional HSE is an exception, as it relies instead on reducing the self-interaction error to achieve better pressure estimates. Not only did we observe reasonably accurate pressure estimations from HSE, but it additionally had some of the smallest errors with local energetics and forces. The global energetics errors were among the worst of the functionals, but this might be due primarily to the lack of error cancellation discussed in the preceding section.

C. Enthalpies

When constructing the equation of state for H+He mixtures, accurate enthalpies are important. One can cut the enthalpy errors by approximately 50–60 % relative to PBE (from 11 to 4 mHa/electron in pure hydrogen) by using either BLYP or vdW-DF. Improving the enthalpy errors beyond this without using postprocessed corrections is difficult. The 4 mHa for vdW-DF and BLYP is due in large part to significant (though noticeably incomplete) error cancellation. Given the inherent tradeoff between energy errors and pressure errors discussed previously, one should be extremely careful about correcting the energy and pressure contributions individually, especially in the absence of some rudimentary error quantification.

D. Forces

From our analysis of local energetic errors and forces, we saw that there is a strong though not perfect correlation between small energetic errors and small force errors. Accurate electron-ion forces depend on the ability of a functional to accurately reproduce the electronic charge density. While this contribution is included in measuring local energetic errors, the treatment of electron-electron correlation also contributes. The superior performance of the HSE functional in minimizing the local energetic errors and force errors most likely stems from its ability to produce a reasonable charge density. The introduction of exact exchange favors charge localization through the reduction of self-interaction errors. After HSE, TPSS seems to produce reasonable charge densities. Among the GGA's and vdW corrected GGA's, the vdW-DF, BLYP, and optB86b-vdW functionals seem to produce reasonable charge densities, most likely because the underlying exchange functionals are skewed to favor bonding and charge localization.

E. Role of exchange

We believe the difference in the performance of the functionals stems from the treatment of exchange, rather than the addition of sophisticated nonlocal correlation effects. This

conclusion follows from two pieces of evidence. First, one would expect that if vdW-type correlation were necessary for an accurate description of dense H+He, no GGA would be able to compete with a properly tuned nonlocal van der Waals functional. However, the two energetically best performing functionals are vdW-DF, which is a nonlocal vdW functional, and BLYP, which is a GGA. Beyond having comparable performance, the total magnitudes and scaling of local and global energy errors with these two functionals are very similar. This strongly suggests that the energetic contributions made by the nonlocal van der Waals effects are small compared to the traditional GGA-type exchange and correlation.

The other piece of evidence comes from comparing the performance of the nonlocal van der Waals functionals. vdW-DF, vdW-DF-C09, and vdW-DF-CX all use the same nonlocal correlation functional, differing only in their choice of exchange functional. The same is true of vdW-DF2 and vdW-DF2-C09. We found that vdW-DF-C09 and vdW-DF2-C09 were virtually indistinguishable energetically, indicating the small role played by the difference in the van der Waals correlation piece. However, there are significant differences between vdW-DF-C09 and vdW-DF, and likewise between vdW-DF2-C09 and vdW-DF2. These pairs demonstrate a propensity to underbind or overbind the H-H interaction, respectively, relative to QMC.

It turns out that the best performing density functionals exhibit some common trends in their underlying exchange functionals. The exchange functionals for GGA's are given by $E_x[\rho] = \int d\mathbf{r}\, \rho(\mathbf{r})\epsilon_x^{\mathrm{hom}}(\mathbf{r})F_x[s(\mathbf{r})]$, where $\epsilon_x^{\mathrm{hom}}$ is the Slater-type exchange for the homogeneous electron gas, F_x is the "enhancement factor," and $s = |\nabla\rho|/[2(3\pi^2)^{1/3}\rho^{4/3}]$ is the "reduced density gradient". Before getting into similarities in F_x responsible for decent or poor energetic or pressure performance, we need to know which values of s are relevant in our system. After analyzing the PBE and BLYP charge densities for a single sample configuration from each density and helium concentration, we conclude that s is bounded by $0 < s < 1.8$ for all configurations of interest ($s \leqslant 0.8$ for pure H configurations). Unsurprisingly, the largest gradients occur in pure helium configurations at low density, whereas the smallest gradients occur in pure hydrogen at high density.

Within the semilocal GGA functionals, we can explain better or worse energetic performance relative to PBE by looking at F_x (see Fig. 11). We saw that the global energetic, local energetic, and force errors followed this progression of decreasing accuracy: BLYP, revPBE, PBE, and PBEsol. Looking at the underlying enhancement factors (BLYP uses B88 exchange), we see the following trend: $F_x^{\mathrm{B88}} > F_x^{\mathrm{revPBE}} > F_x^{\mathrm{PBE}} > F_x^{\mathrm{PBEsol}}$ for all "s" in the relevant range for hydrogen. $F_x^{\mathrm{revPBE}} > F_x^{\mathrm{B88}}$ from about $s = 0.8$ onward (they cross again at much larger s), but this does not affect the description of hydrogen. This implies that the best performing functionals for energies and forces work by lowering the energy contributions coming from the larger reduced gradients, which favors charge localization and bonding.

We also note that the electronic structure around protons is *very* sensitive to the treatment of exchange at these densities. Despite the qualitatively different energetics and H-H force errors between PBE and BLYP, we see that the relative difference in the exchange enhancement factors $F_x^{\mathrm{B88}} - F_x^{\mathrm{PBE}} \approx 0.005$ at $s = 0.4$. In contrast, the electronic structure near the helium nuclei does not seem to be nearly as sensitive to the choice of exchange functional. This conclusion is based on how similar the local energetic errors and He-He forces were in pure helium configurations across different functionals.

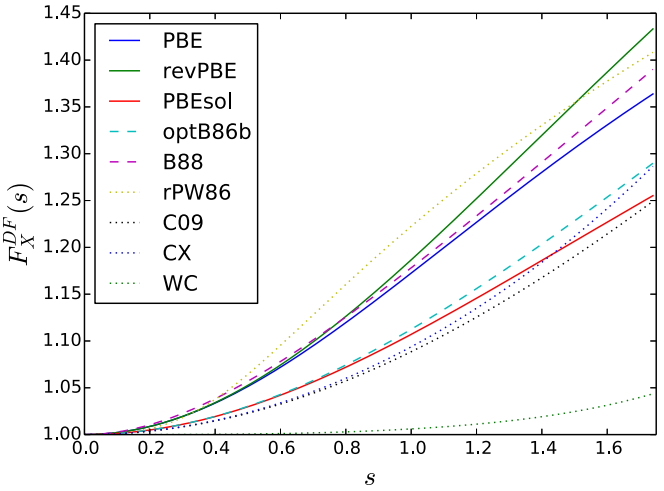

FIG. 11. Plot of enhancement factors $F_x(s)$ for all the best performing functionals.

One can perform the same type of analysis with the vdW-DF-type functionals. vdW-DF, optB86b-vdW, and vdW-DF-CX use the revPBE, optB86b, and CX exchange functionals, respectively. We previously saw that for energetic and force errors, the progression toward decreasing accuracy follows the sequence vdW-DF, optB86b-vdW, and vdW-DF-CX. Looking at the underlying enhancement factors, we find that $F_x^{\mathrm{revPBE}} > F_x^{\mathrm{optB86b}} > F_x^{\mathrm{CX}}$. vdW-DF and optB86b-vdW perform comparably, but vdW-DF overbinds relative to QMC whereas optB86-vdW underbinds. We forgo a direct exchange functional comparison between the vdW functionals and the GGA's, primarily because of the "exchange consistency" complication stemming from the use of a different "outer" and "inner" exchange correlation functional.

Deeper relationships between the functional form of $F_x(s)$ and corresponding errors can be deduced from the previous discussion. However, we leave these considerations to future publications, since our current focus is on hydrogen-helium thermodynamics and not on density-functional development.

V. CONCLUSIONS

In this paper, we have used projector quantum Monte Carlo to benchmark some of the most popular density functionals, ranging from GGA, to nonlocal dispersion corrected, to meta-GGA. We were able to quantify the errors for most quantities that are relevant for constructing an equation of state, specifically the pressures, local, and global energy differences. As a result of our analysis, we can conclude that a significant reduction of enthalpy errors and a much better energetic description of hydrogen helium interactions can be attained by

using the TPSS metaGGA, the BLYP GGA, or the nonlocal vdW-DF functional. For the pressure, on the other hand, there are several functionals that have errors around 1% for all densities and helium concentrations considered: optB88, WC, PBESol, and vdW-DF-C09.

Beyond just identifying the most accurate density functional and quantifying its errors, we have demonstrated the common features of the best performing functionals, specifically in the shape and limiting behavior of the enhancement factors for the exchange functionals. The underlying exchange pieces for both vdW-DF, BLYP, and revPBE tend to emphasize bonding in the energetics, which is well known in the DFT literature. The importance of this work is that it specifies quantitatively just how important this bonding character is for an accurate description of dense hydrogen-helium mixtures. Knowing this, and how much the various exchange correlation functionals overbind or underbind, should facilitate the optimization and deployment of new functionals for mapping out the H+He phase diagram.

ACKNOWLEDGMENTS

M.A.M. was supported by the U.S. Department of Energy at the Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344. M.A.M., R.C., and D.M.C. were supported through the Predictive Theory and Modeling for Materials and Chemical Science program by the Basic Energy Science (BES), DOE. R.C. and D.M.C. were also supported by DOE DE-NA0001789. M.H. and D.M.C. acknowledge support from the Fondation Nanosciences de Grenoble. Computer time was provided by the U.S. DOE INCITE program, Lawrence Livermore National Laboratory through the 7th Institutional Unclassified Computing Grand Challenge program and PRACE Project No. 2011050781.

[1] J. Schneider, C. Dedieu, P. L. Sidaner, R. Savalle, and I. Zolotukhin, Astron. Astrophys. **532**, A79 (2011).
[2] D. Stevenson, Phys. Rev. B **12**, 3999 (1975).
[3] D. Saumon, G. Chabrier, and H. M. van Horn, Astrophys. J. Suppl. Ser. **99**, 713 (1995).
[4] J. E. Klepeis, K. J. Schafer, T. W. Barbee, and M. Ross, Science (New York, NY) **254**, 986 (1991).
[5] O. Pfaffenzeller, D. Hohl, and P. Ballone, Phys. Rev. Lett. **74**, 2599 (1995).
[6] W. Lorenzen, B. Holst, and R. Redmer, Phys. Rev. Lett. **102**, 115701 (2009).
[7] W. Lorenzen, B. Holst, and R. Redmer, Phys. Rev. B **84**, 235109 (2011).
[8] M. A. Morales, E. Schwegler, D. Ceperley, C. Pierleoni, S. Hamel, and K. Caspersen, Proc. Natl. Acad. Sci. (USA) **106**, 1324 (2009).
[9] M. A. Morales, S. Hamel, K. Caspersen, and E. Schwegler, Phys. Rev. B **87**, 174105 (2013).
[10] D. J. Stevenson and E. E. Salpeter, Astrophys. J. Suppl. Ser. **35**, 221 (1977).
[11] D. J. Stevenson and E. E. Salpeter, Astrophys. J. Suppl. Ser. **35**, 239 (1977).
[12] J. J. Fortney and W. B. Hubbard, Icarus **164**, 228 (2003).
[13] N. Nettelmann, Contrib. Plasma Phys. **55**, 116 (2015).
[14] M. A. Morales, J. M. McMahon, C. Pierleoni, and D. M. Ceperley, Phys. Rev. Lett. **110**, 065702 (2013).
[15] M. A. Morales, J. M. McMahon, C. Pierleoni, and D. M. Ceperley, Phys. Rev. B **87**, 184107 (2013).
[16] R. C. Clay III, J. Mcminis, J. M. McMahon, C. Pierleoni, D. M. Ceperley, and M. A. Morales, Phys. Rev. B **89**, 184106 (2014).
[17] J. McMinis, R. C. Clay, D. Lee, and M. A. Morales, Phys. Rev. Lett. **114**, 105305 (2015).
[18] G. Mazzola and S. Sorella, Phys. Rev. Lett. **114**, 105701 (2015).
[19] S. Azadi, B. Monserrat, W. M. C. Foulkes, and R. J. Needs, Phys. Rev. Lett. **112**, 165501 (2014).
[20] N. D. Drummond, B. Monserrat, J. H. Lloyd-Williams, P. L. Rios, C. J. Pickard, and R. J. Needs, Nat. Commun. **6**, 7794 (2015).
[21] S. Azadi and W. M. C. Foulkes, Phys. Rev. B **88**, 014115 (2013).
[22] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).
[23] G. Kresse and J. Hafner, Phys. Rev. B **49**, 14251 (1994).
[24] G. Kresse and J. Furthmüller, Comput. Mater. Sci. **6**, 15 (1996).
[25] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[26] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[27] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo *et al.*, J. Phys. Condens. Matter **21**, 395502 (2009).
[28] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[29] Y. Zhang and W. Yang, Phys. Rev. Lett. **80**, 890 (1998).
[30] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, Phys. Rev. Lett. **100**, 136406 (2008).
[31] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[32] Z. Wu and R. E. Cohen, Phys. Rev. B **73**, 235116 (2006).
[33] Y. Zhao and D. G. Truhlar, J. Chem. Phys. **125**, 194101 (2006).
[34] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, Phys. Rev. Lett. **91**, 146401 (2003).
[35] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, Phys. Rev. Lett. **92**, 246401 (2004).
[36] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, and D. C. Langreth, Phys. Rev. B **82**, 081101 (2010).
[37] V. R. Cooper, Phys. Rev. B **81**, 161104 (2010).
[38] K. Berland and P. Hyldgaard, Phys. Rev. B **89**, 035412 (2014).
[39] J. Klimeš, D. R. Bowler, and A. Michaelides, Phys. Rev. B **83**, 195131 (2011).
[40] J. Klimeš, D. R. Bowler, and A. Michaelides, J. Phys. Condens. Matter **22**, 022201 (2010).
[41] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[42] Opium, http://opium.sourceforge.net.
[43] K. Esler, J. Kim, D. Ceperley, and L. Shulenburger, Comput. Sci. Eng. **14**, 40 (2012).
[44] J. Kim, K. P. Esler, J. McMinis, M. A. Morales, B. K. Clark, L. Shulenburger, and D. M. Ceperley, J. Phys.: Conf. Ser. **402**, 012008 (2012).
[45] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, Phys. Rev. Lett. **98**, 110201 (2007).

[46] S. Baroni and S. Moroni, Phys. Rev. Lett. **82**, 4745 (1999).
[47] C. Pierleoni and D. M. Ceperley, ChemPhysChem **6**, 1872 (2005).
[48] S. Chiesa, D. M. Ceperley, and S. Zhang, Phys. Rev. Lett. **94**, 036404 (2005).
[49] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.93.035121 for details on the finite size correction schemes and force estimator tests.
[50] C. Lin, F. H. Zong, and D. M. Ceperley, Phys. Rev. E **64**, 016702 (2001).
[51] S. Chiesa, D. Ceperley, R. Martin, and M. Holzmann, Phys. Rev. Lett. **97**, 076404 (2006).
[52] N. D. Drummond, R. J. Needs, A. Sorouri, and W. M. C. Foulkes, Phys. Rev. B **78**, 125106 (2008).
[53] H. Kwee, S. Zhang, and H. Krakauer, Phys. Rev. Lett. **100**, 126404 (2008).