An extensive study of gradient approximations to the exchange-correlation and kinetic energy functionals

Garnet Kin-Lic Chan[a)] and Nicholas C. Handy
Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, United Kingdom

We formalize the procedure of functional development, in a general theoretical framework. Expansion in a functional basis set, and fitting via an error functional to a data set, casts functional development as a variational problem to obtain the functional basis-set and data-set limits. Overfitting is avoided by defining the optimum number of parameters. We implement our theory for an investigation of first- and second-order generalized gradient approximations (GGA) to the exchange-correlation and kinetic energy functionals, within an *ab initio* model. A variety of functional basis sets, including a general finite-element representation, is constructed to represent both one-dimensional and multidimensional GGA enhancement factors. An extensible data set consisting of 429 atomic and diatomic, neutral and cationic species, at stretched and equilibrium geometries, is constructed from Moller–Plesset level exchange-correlation energies, and Hartree–Fock kinetic energies. The range of chemically relevant density and gradient variables is examined. Exhaustive fitting investigations are carried out, to determine the accuracy of the GGA representation of the *ab initio* models. In the exchange-correlation case we demonstrate that we can reach the functional basis-set and data-set limit, which correspond to a root-mean-square (rms) error of ~ 10 mH (6.3 kcal/mol). Changing the functional basis set, higher-order density variables such as the kinetic energy density, multidimensional enhancement factors, and exact exchange yield *no significant improvement*, and our fits represent an effective solution of the GGA problem for exchange-correlation, at the Møller–Plesset level. In the kinetic energy case, accurate functionals with rms errors of ~ 80 mH (50 kcal/mol) are developed. These exhibit a beautifully simple kinetic energy enhancement factor, and are a step towards orbital-free calculations. © *2000 American Institute of Physics.* [S0021-9606(00)30110-6]

I. INTRODUCTION

Density functional theory is now a mature field. Many applications are routinely found in the literature.[1,2]

Kohn–Sham density functional theory,[3,4] partitions the total energy $E[\rho]$ as

$$E[\rho] = T_s[\rho] + E_{xc}[\rho] + J[\rho] + \int \rho(\mathbf{r}) v(\mathbf{r}) d\mathbf{r}, \quad (1)$$

where $T_s[\rho] = -\frac{1}{2} \min_{\Psi \to \rho} \langle \Psi | \Sigma_i \nabla_i^2 | \Psi \rangle$, $J[\rho] = \frac{1}{2} \int\int \rho(\mathbf{r}_1) \times \rho(\mathbf{r}_2) r_{12}^{-1} d\mathbf{r}_1 d\mathbf{r}_2$, and $v(\mathbf{r})$ is the external potential.

As an approximate theory, the accuracy is limited by the exchange-correlation functional $E_{xc}[\rho]$. Another open question remains an approximation to the Kohn–Sham kinetic energy functional $T_s[\rho]$, which will facilitate orbital-free calculations.

Recent years have seen an increase primarily in the number of exchange-correlation functionals (with notable developments in kinetic energy functionals[5–7]). These have been of the *generalized gradient approximation* (GGA) form,[5,8,9] which are integrals of local functions of density variables ρ, $|\nabla \rho|$, $\nabla^2 \rho$, $\Sigma_i |\nabla \phi_i|^2$ (where ϕ_i is an orbital), or of higher order. Different philosophies in functional development are summarized in the reviews by Perdew and co-workers[10–12] and Becke and others.[13–17] Although the trend has been to-

wards more systematic approaches to functional development (see, for example, Refs. 13 and 17), no general theory of functional development has yet appeared.

In this work, we tackle the problem of functional development afresh. We move away from functionals based on a specific physical model. Instead, in common with recent approaches of Becke and others,[13,17] we rely on extensive fitting. A general theory of functional development, which *systematizes* the procedure of fitting, is developed. In Sec. II, we define the functional basis, data set, and error functional. Within such a language, functional development is reduced to the attainment of the functional basis and data set limits. This is facilitated, in practice, by a rigorous solution of the problem of overfitting.

Our methodology will be applied to the determination of $E_{xc}[\rho]$ and $T_s[\rho]$. However, we need to limit the *form* of the functional. We choose the GGA as our framework to study the question: what is the limiting accuracy of our representation of the exchange-correlation and kinetic energies? In Sec. III we review the density variables, and enhancement factors, which define the GGA functionals. In the case of $E_{xc}[\rho]$, we also review exact exchange.

In Sec. IV, we proceed, within an *ab initio* philosophy, to a practical implementation of our fitting theory. Since exact data sets are generally hard to obtain, we aim instead to choose a *well-defined ab initio* model, and proceed therein to exhaustively examine how exact is a density functional rep-

[a)]Electronic mail: gkc1000@hermes.cam.ac.uk

resentation possible within the GGA framework. We aim to (i) reproduce the exchange-correlation energy of a second-order Møller–Plesset(MP2)–Hartree–Fock(HF) model, and (ii) reproduce the HF kinetic energy. Multiple functional basis sets in ρ, $|\nabla\rho|$, $\nabla^2\rho$, $\Sigma_i|\nabla\phi_i|^2$ are defined. A new feature of this work is the introduction of *finite-element* functional basis sets, which are general and easily extensible to new density variables and multiple dimensions. A simple, extensible data set is constructed from *ab initio* Møller–Plesset and Hartree–Fock calculations, containing 429 atoms and diatomic molecules, charged and neutral species, at equilibrium and stretched geometries.

Our extensive fits are presented in Sec. V. We perform a set of fitting experiments, which tests the ideas developed above, for the exchange-correlation and kinetic energy functionals. By exhaustive optimization, we reach the functional basis-set limit of $E_{xc}[\rho]$ and $T_s[\rho]$ within the GGA form. The exchange-correlation GGA problem is nearly solved within our MP2–HF model. In the case of the kinetic energy, accurate kinetic energy functionals of ~ 80 mH (50 kcal/mol) rms accuracy are developed.

We summarize our work and conclusions in Sec. VI.

II. GENERAL THEORY OF FUNCTIONAL DEVELOPMENT

Here, we introduce the theoretical concepts in the empirical construction of functionals. In what follows, functionals $A[B]$ are often written in the abbreviated form **A**.

A. A formal statement of fitting

The empirical procedure to fit a functional **F** to data set **D** may be stated as follows: given **D**, we wish to find the closest possible functional $F[\rho]$ in a trial functional space **T**. If we define a metric, or error functional $\Delta[\mathbf{F},\mathbf{D}]$, then by minimizing the error Δ with respect to **F**, the optimum functional $\mathbf{F}_0$ is given by

$$\min_{\mathbf{F}\in\mathbf{T}} \Delta[\mathbf{F},\mathbf{D}]\Rightarrow\mathbf{F}\rightarrow\mathbf{F}_0. \tag{2}$$

B. The functional basis and data set

The space of all functionals is very large. Naively, we expect the space of all functionals of ρ to map each point in $\rho\in\mathbf{R}^3$ to $F\in\mathbf{R}$; thus, the dimension of the space of all functionals would be, in some rough sense, $(\mathbf{R}^3)^{|\mathbf{R}|}$. Clearly this is too large a space **T** within which to perform the minimization (2).

Instead, we restrict ourselves to the subspace spanned by general nonlocal functionals of the multi-dimensional integral form

$$\mathbf{F}=\int\cdots\int f[v_1(\mathbf{r}_1),\ldots,v_1(\mathbf{r}_m),$$
$$v_2(\mathbf{r}_1),\ldots,v_2(\mathbf{r}_m),\ldots]d\mathbf{r}_1,\ldots,d\mathbf{r}_m, \tag{3}$$

where v_i is a density variable; that is, $v_i\in\{\mathbf{r},\rho,|\nabla\rho|^2,\nabla^2\rho,\ldots,\}$. The kernel f is a *function* of an arbitrary finite number of density variables.

f can be expanded in a *functional basis* $\{\mu(v_1,v_2,\ldots,)\}$,

$$f=\sum_i c_i\mu_i(v_1,v_2,\ldots,); \tag{4}$$

μ may be truncated to yield a finite functional basis of n functions.

The complete data set **D** will be the exact functional $F^{\text{exact}}[\rho]$. *It is the choice of data for a finite data set which characterizes different philosophies in functional development*. A finite data set is often a set of p data pairs $\{\rho,F\}$, such as the G2 thermochemical data set[18] used by many workers (see, e.g., Refs. 15,17, and 19), but may equally well be constructed from limits and asymptotic conditions, as advocated by Perdew and co-workers.[11,20,21] We shall denote the finite data set as $\mathbf{D}(p)$, with p data pairs or conditions.

C. Fitting and testing sets

Let us consider fitting with a finite n function basis μ to a finite data set $\mathbf{D}(p)$. Our optimum functional $\mathbf{F}_0(n,p,\mu)$, is determined through

$$\min_{\mathbf{F}\in\mathbf{T}} \Delta[\mathbf{F}(n,p,\mu),\mathbf{D}(p)]\Rightarrow\mathbf{F}(n,p,\mu)\rightarrow\mathbf{F}_0(n,p,\mu). \tag{5}$$

The data set $\mathbf{D}(p)$ is known as the *fitting set*, which defines the functional $\mathbf{F}_0(n,p,\mu)$. For brevity, we shall now write the error $\Delta[\mathbf{F}_0(n,p,\mu),\mathbf{D}(p)]$ as $\Delta_0[(n,p,\mu),p]$.

To fully utilize the information in our data set, we can evaluate the error of the functional $\mathbf{F}_0(n,p,\mu)$ over a different *testing set* $\mathbf{D}(p')$. This error is denoted $\Delta_0[(n,p,\mu),p']$.

D. Functional basis set and data-set limits

The data set limit of the error $\boldsymbol{\Delta}_0$ and the optimum functional $\mathbf{F}_0$ are easily defined. If we expand the data set to completion, which we write loosely as the limit $p\rightarrow\infty$, then the corresponding limits of the error $\Delta_0[(n,p,\mu),p]$ and functional $\mathbf{F}_0[n,p,\mu]$ are their data set limits.

We would like the functional basis-set limit to be defined analogously as the limit $n\rightarrow\infty$. However, for finite data sets, if the number of basis functions n is greater than the number of data points p, the error vanishes and the coefficients in the basis expansion (4) are not uniquely defined. This is the well-known problem of *overfitting*. In general, the degeneracy in $\mathbf{F}_0$ may be felt long before $n>p$, resulting in unstable fits, and functionals that perform poorly outside of the fitting set.

We digress to emphasize that *overfitting is not limited to data sets of the form* $\{\rho,F\}$. The exact fitting (or overfitting) of *asymptotic and exact conditions* can also worsen (or not improve) the performance of a functional; such conditions are usually termed too restrictive. For example, it is generally believed that successful GGA exchange functionals do not reduce to the exact gradient correction in the slowly varying electron gas limit.[12] Similarly, the exact reproduction of the asymptotic exchange energy density in finite systems is now not believed to be particularly important for the success of an exchange energy functional.[15] In any case, although the use of exact conditions to construct functionals is

undoubtedly successful and useful,[21] it must still be regarded as a fitting procedure, and consequently a careful consideration of the general validity of the fit is important.

One way forward is to observe that a useful functional basis-set limit is obtained when fitting over a complete data set. Here, we define the functional basis-set limits as

$$\Delta_0[(\infty,\infty),\infty] = \lim_{n\to\infty} \lim_{p\to\infty} \Delta_0[(n,p,\mu),p] \geqslant 0, \tag{6}$$

$$\mathbf{F}_0(\infty,\infty) = \lim_{n\to\infty} \lim_{p\to\infty} \mathbf{F}_0(n,p,\mu), \tag{7}$$

where the dependence on μ is lost in the functional basis-set limit. $\mathbf{F}_0(\infty,\infty)$ may be regarded as the best functional that can be written in the approximate form (3). Since the trial space of the approximate form is smaller than the space of all functionals, we have the inequality (6). Thus, the error $\Delta_0[(\infty,\infty),\infty)]$ is a measure of the severity of the restriction of $\mathbf{F}$ to an approximate form. Informally, we say that functional $\mathbf{F}$ may not be fitted exactly by an approximate form, although any finite set of data may be fitted arbitrarily well.

E. Convergence to the limit and overfitting

Given a finite data set, how can we approximate the data-set and basis-set limits?

To evaluate the data-set limit, consider a suitable sequence $\mathbf{D}(p_1),\mathbf{D}(p_2),\ldots,\mathbf{D}(p)$, with $\mathbf{D}(p_1)\subset\mathbf{D}(p_2)\subset,\ldots,\subset\mathbf{D}(p)$, and the data-set limits of $\mathbf{\Delta}_0$ and $\mathbf{F}_0$ may in principle be extrapolated.

However, more care is needed in taking the double limit that defines the basis-set limit (6). For finite p, taking the limit $n\to\infty$ results in overfitting. Thus, we now introduce the notion of an *optimum number of parameters* n_0, for a finite data set $\mathbf{D}(p)$. The corresponding error $\Delta_0[(n_0,p,\mu),p]$ and functional $\mathbf{F}_0(n_0,p,\mu)$ may loosely be regarded as the ''functional basis-set limits'' for the finite data set $\mathbf{D}(p)$, if the dependence on the type of basis μ is sufficiently weak.

Let us first assume that we do have a complete data set $\mathbf{D}(\infty)$, but we only fit to the finite subset $\mathbf{D}(p)$. Then, let the complete data set be the testing set, and evaluate the error $\Delta_0[(n,p,\mu),\infty]$. The optimum number of parameters n_0 is then seen to be that which minimizes the error,

$$\Delta_0[(n_0,p,\mu),\infty] = \min_{n\to n_0} \Delta_0[(n,p,\mu),\infty]. \tag{8}$$

Equation (8) is the definition of n_0 and does not rely on statistical assumptions common to the usual goodness-of-fit approach.

Now, the above procedure may easily be modified to accommodate the reality of finite data sets. The quantity $\Delta_0[(n_0,p,\mu),\infty]$ is a data-set limit and may be obtained by a suitable extrapolation from a sequence of data sets, as described above. Thus, we may obtain an estimate of n_0, even for the *largest* of our finite data sets $\mathbf{D}(p)$.

So far we have assumed the existence of some error functional $\mathbf{\Delta}$. For finite data sets, the choice of $\mathbf{\Delta}$ may lead to different functionals $\mathbf{F}_0$. We return to this point in Sec. IV C.

Thus, to summarize: for given $\mathbf{\Delta}$, *once the optimum number of parameters has been determined*, the problem of functional determination is a variational minimization (5)

with two basis sets: a functional basis, and a data set. This concludes our discussion of the formal theory of fitting. We now apply our formal machinery, to the determination of $E_{\mathrm{xc}}[\rho]$ and $T_s[\rho]$. First, however, we consider the generalized gradient approximation form.

III. THE GENERALIZED GRADIENT APPROXIMATION FORM

As noted in Sec. II B, if we wish to determine a functional in practice, the minimization procedure must be carried out within some approximate form. The workhorse of density functional theory is the generalized gradient approximation (GGA), a subset of the form (3). We shall use the GGA as the framework for our study of $T_s[\rho]$ and $E_{\mathrm{xc}}[\rho]$.

A. The exchange-correlation functional

Let us now consider the spin-density functional $E_{\mathrm{xc}}[\rho_\alpha,\rho_\beta]$, which admits a separation into like-spin ($\sigma\sigma$) exchange, correlation and opposite-spin ($\alpha\beta$) correlation components[15,22]

$$E_{\mathrm{xc}}[\rho_\alpha,\rho_\beta] = E_{C\alpha\beta}[\rho_\alpha,\rho_\beta]$$
$$+ \sum_{\sigma\in\{\alpha,\beta\}} (E_{X\sigma\sigma}[\rho_\sigma] + E_{C\sigma\sigma}[\rho_\sigma]). \tag{9}$$

Although the above separation (9) admits a particularly simple treatment of spin effects, it must be remembered that as one equation *it does not uniquely define three unknowns* $E_{C\alpha\beta}, E_{X\sigma\sigma}, E_{C\sigma\sigma}$. For example, we may trivially redefine $E_{C\alpha\beta}\to E_{C\alpha\beta}+\lambda E_{C\sigma\sigma}$.

The GGA form, first introduced by Perdew[8] and Becke,[9] attempts an extrapolation from the uniform electron gas (UEG) energy densities, *viz.*

$$\mathbf{E}_{x\sigma\sigma} = \int \epsilon_{X\sigma\sigma}^{\mathrm{UEG}} f_{X\sigma\sigma}(v_{1\sigma},v_{2\sigma},\ldots,)d\mathbf{r}, \tag{10}$$

$$\mathbf{E}_{C\sigma\sigma} = \int \epsilon_{C\sigma\sigma}^{\mathrm{UEG}} f_{C\sigma\sigma}(v_{1\sigma},v_{2\sigma},\ldots,)d\mathbf{r}, \tag{11}$$

$$\mathbf{E}_{C\alpha\beta} = \int \epsilon_{C\alpha\beta}^{\mathrm{UEG}} f_{C\alpha\beta}(v_1,v_2,\ldots,)d\mathbf{r}, \tag{12}$$

where v_i is a density variable. The exchange energy density $\epsilon_{X\sigma\sigma}^{\mathrm{UEG}}$ can be written in closed form $\epsilon_{X\sigma\sigma}^{\mathrm{UEG}} = -\frac{3}{2}(3/4\pi)^{1/3}\rho_\sigma^{1/3}$, while accurate parametrizations exist for the correlation energy densities $\epsilon_{C\sigma\sigma}$ and $\epsilon_{C\alpha\beta}$. In this work we choose the widely implemented Vosko–Wilk–Nusair (Version V) functional (VWN),[23] with the $\sigma\sigma$ and $\alpha\beta$ components extracted following Stoll *et al.*[24]

The choice of density variables v_i specifies the *order* of the GGA. Up to second order, it is conventional to use the density variables $v_i \in \{\rho,x^2,y,\tau\}$, where x,y,τ are defined to be dimensionless gradients

$$x = |\nabla\rho|/\rho^{4/3}, \tag{13}$$

$$y = \nabla^2\rho/\rho^{5/3}, \tag{14}$$

$$\tau = \sum_i^N |\nabla\phi_i|^2/\rho^{5/3}, \tag{15}$$

and the generalization to spin requires adding σ subscripts to all density variables. The inclusion of τ, defined via the

Kohn–Sham orbitals of the system *and here chosen to be dimensionless*, stems from a consideration of the series expansion of the off-diagonal elements of the Kohn–Sham density matrix.[19,25] Higher-order gradients may also be considered (see Neumann and Handy[26]) but lead to numerical difficulties in quadrature.

From the above, we define a class of *first-order* GGA functionals, where the enhancement factors have the form

$$f_{X\sigma\sigma}=f_{X\sigma\sigma}(x_\sigma^2), \tag{16}$$

$$f_{C\sigma\sigma}=f_{C\sigma\sigma}(\rho_\sigma,x_\sigma^2), \tag{17}$$

$$f_{C\alpha\beta}=f_{C\alpha\beta}(\rho,x^2). \tag{18}$$

Second-order GGAs, also sometimes called meta-GGAs,[10] have enhancement factors of the form

$$f_{X\sigma\sigma}=f_{X\sigma\sigma}(x_\sigma^2,y_\sigma,\tau_\sigma), \tag{19}$$

$$f_{C\sigma\sigma}=f_{C\sigma\sigma}(\rho_\sigma,x_\sigma^2,y_\sigma,\tau_\sigma), \tag{20}$$

$$f_{C\alpha\beta}=f_{C\alpha\beta}(\rho,x^2,y,\tau). \tag{21}$$

It is often neglected that even in the first-order GGA, the correlation components are *multidimensional* functions (the exchange component has no ρ dependence from homogeneity under coordinate scaling[27]). Most commonly, however, first-order GGAs are constructed as functions of x^2 only.[15] At the second-order level, all enhancement factors are multidimensional.

In recent years, the GGA form has been extended to include a fraction of exact exchange.[15,28,29] Denoting the GGA component by $E_{\rm xc}^{\rm GGA}[\rho]$ and the exact exchange component by $E_X^{\rm exact}[\rho]$, such hybrid functionals are of the form

$$E_{\rm xc}[\rho_\alpha,\rho_\beta]=E_{\rm xc}^{\rm GGA}[\rho_\alpha,\rho_\beta]+c_X\sum_{\sigma\in\{\alpha,\beta\}}(E_X^{\rm exact}[\rho_\sigma]). \tag{22}$$

B. The kinetic energy functional

We first consider spin,

$$T_s[\rho]=\sum_{\sigma\in\{\alpha,\beta\}}T_s[\rho_\sigma]. \tag{23}$$

Although empirical studies of kinetic energy functionals are scarce, there is a well-known gradient expansion[4] which to second order is

$$T_s[\rho_\sigma]=\int\rho_\sigma^{5/3}\left(C_F+C_Wx_\sigma^2+\cdots\right)d\mathbf{r}, \tag{24}$$

where $C_F=\frac{3}{10}(6\pi^2)^{2/3}$, $C_W=\frac{1}{72}$. The first term is usually referred to as the Thomas–Fermi term, and the second as the von Weizsäcker term. The expansion (24) immediately suggests the GGA form, first investigated by Lee, Lee, and Parr[5]

$$T_s[\rho_\sigma]=\int\epsilon_{T_s}f_{T_s}(v_{1\sigma},v_{2\sigma},\dots,)d\mathbf{r}, \tag{25}$$

where $\epsilon_{T_s}=\rho_\sigma^{5/3}$. From the coordinate scaling of $T_s[\rho]$, the enhancement factor has no explicit ρ dependence; thus, the first-order GGA enhancement factor has the form

$$f_{T_s}=f_{T_s}(x_\sigma^2). \tag{26}$$

A second-order orbital-free GGA approximation will have enhancement factors of the form

$$f_{T_s}=f_{T_s}(x_\sigma^2,y_\sigma). \tag{27}$$

IV. DEVELOPMENT OF A FITTING METHODOLOGY

In the previous sections, we developed a general theory of fitting, and introduced the GGA framework. We now proceed to a practical implementation, with a development of functional basis sets μ, data sets **D**, and an error functional **Δ**.

A. The functional basis

The enhancement factors defined in Secs. III A and III B are one- or multidimensional functions of density variables. We now construct basis sets to expand these functions, including a general purpose finite-element representation. Our approach will be numerical in fashion, i.e., *we do not concern ourselves with a specific physical model to generate our basis sets.*

1. The B97 basis

A widely used basis for the first-order exchange-correlation GGA form was introduced by Becke,[15] where

$$f_{X\sigma\sigma}=\sum_{i=1}^{n_{X\sigma\sigma}}c_i[\gamma_{X\sigma\sigma}x_\sigma^2/(1+\gamma_{X\sigma\sigma}x_\sigma^2)]^{i-1}, \tag{28}$$

$$f_{C\sigma\sigma}=\sum_{i=1}^{n_{C\sigma\sigma}}c_i[\gamma_{C\sigma\sigma}x_\sigma^2/(1+\gamma_{C\sigma\sigma}x_\sigma^2)]^{i-1}, \tag{29}$$

$$f_{C\alpha\beta}=\sum_{i=1}^{n_{C\alpha\beta}}c_i[\gamma_{C\alpha\beta}x_{\rm av}^2/(1+\gamma_{C\alpha\beta}x_{\rm av}^2)]^{i-1}, \tag{30}$$

with $x_{\rm av}^2=x_\alpha^2+x_\beta^2$ (twice $s_{\rm av}^2$ in Becke's work[15]), and $\gamma_{X\sigma\sigma}=0.004$, $\gamma_{C\sigma\sigma}=0.2$, $\gamma_{C\alpha\beta}=0.006$ (preoptimized by Becke). The coefficients in the basis expansion [Eq. (4)] are clearly the coefficients $\{c_i\}$, of which there are $n_{X\sigma\sigma}+n_{C\sigma\sigma}+n_{C\alpha\beta}$ in total. We shall refer to this basis as the B97 basis.

The B97 basis has obvious shortcomings, as it does not span the complete space of first-order GGAs, since the correlation enhancement factors have no explicit ρ dependence, and depend on $x_{\rm av}^2$ as opposed to x^2. However, it is acknowledged as a successful basis, and thus we shall often use it as a *standard basis* against which to compare our other basis sets. Replacing $f_{X\sigma\sigma}$ by f_{T_s} in Eq. (28), we also use the B97 basis to expand the kinetic energy enhancement factor.

2. Finite-element basis F1,F2

Little is known about the general form of the enhancement factor, particularly in the case of the kinetic energy enhancement factor. For this reason, we now introduce finite-element representations.

A simple way to generate a grid in a *local* density variable is to choose grid points that divide the density variable into intervals, which each carry an equal probability of finding an electron within. For example, let us consider a grid in
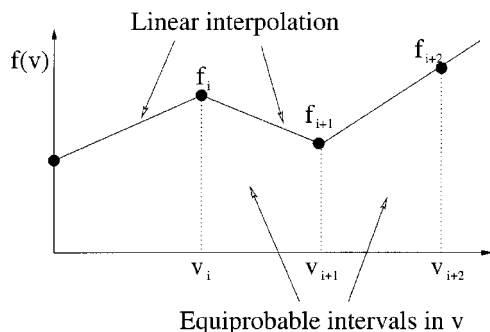
FIG. 1. F1 basis: schematic representation of enhancement factor represented by function values $f(v_i)$ on grid in v_i.

the density variable $v \in \{\rho, x^2, y, \tau\}$, with $i-1$ intervals of the form $[v_i, v_{i+1}]$. Choosing some data set (we shall use our largest data set in this work), at each point in space $\mathbf{r}$, there is a density $\rho(\mathbf{r})$ associated with the corresponding $v(\mathbf{r})$; thus, the probability of an electron with v in the interval $[v_i, v_{i+1}]$, $P_i(v)$ is given by

$$P_i(v) = \int \rho(\mathbf{r}) d\mathbf{r} : v_i \leq v < v_{i+1}. \qquad (31)$$

We then choose the grid points $\{v_i\}$ such that $P_i(v)$ is the same for all i. Such an ''equiprobability'' partitioning is applicable to all density variables ρ, x^2, y, τ and spin counterparts.

Multidimensional enhancement factors are represented on multidimensional grids that are product grids of one-dimensional grids generated in the above fashion; for example, the enhancement factor $f_{C\sigma\sigma}(\rho_\sigma, x_\sigma^2)$ would be represented on a grid formed from the pairs of grid points $\{\rho_{\sigma i}, x_{\sigma j}^2\}$, where $\{\rho_{\sigma i}\}, \{x_{\sigma j}^2\}$ are generated from Eq. (31).

The second feature of a finite-element basis is the choice of interpolating function between the grid points. We investigate two different types of functions:

a. Basis F1/F1[d_1, d_2]- Linear interpolation. Let us denote our generic enhancement factor by f. Then, we represent our enhancement factor simply by the values of f at the grid points, denoted f_i. These form the linear coefficients to optimize in the basis expansion, i.e., $\{c_i\} = \{f_i\}$. In between grid points, we interpolate, e.g., in 1-D, for density variable $v \in \{\rho, x^2, y, \tau\}$ such that $v_i \leq v \leq v_{i+1}$, our basis representation of $f(v)$ is

$$f(v) = f(v_i) + (v - v_i)(f(v_{i+1}) - f(v_i))/(v_{i+1} - v_i). \qquad (32)$$

This representation is illustrated in Fig. 1.

Extensions to higher dimensions are simple. In 2-D, with density variables v, w, with $v_i \leq v < v_{i+1}$, $w_i \leq w < w_{i+1}$,

$$f(v, w) = (1-t)(1-u)f(v_i, w_j) + t(1-u)f(v_{i+1}, w_j) \qquad (33)$$

$$+ tuf(v_{i+1}, w_{j+1}) + (1-t)uf(v, w_{j+1}), \qquad (34)$$

$$t = (v - v_i)/(v_{i+1} - v_i), \qquad (35)$$

$$u = (w - w_i)/(w_{i+1} - w_i). \qquad (36)$$

For *each* enhancement factor represented in this fashion, the *number* of linear coefficients in our basis expansion n, is clearly the number of grid points used to represent the factor. For example, in two dimensions, with d_1, d_2 grid points along each of the two directions, $n = d_1 d_2$.

This conceptually simple basis, denoted by F1 in 1-D, and F1[d_1, d_2] in 2-D, may be considered as the workhorse of the finite-element representation.

b. Basis F2[d] - Polynomial interpolation. Restricting ourselves to a one-dimensional m point grid $\{v_i\}$, we define a transformed variable $z(v) = \gamma_v(v - v_i)/(1 + \gamma_v(v - v_i)): v_i \leq v \leq v_{i+1}$, giving

$$f(v) = \sum_{j=0}^{d} c_{i,j} z(v)^j, \qquad (37)$$

where d is the maximum degree of the interpolating polynomial, and $\gamma_v = 1/\bar{v}$, where $\bar{v}$ is the median of v over the largest data set (listed in Table I). Continuity of the enhancement factor at each of the grid points v_i fixes one coefficient, chosen to be $c_{i,d} = c_{i+1,0} - \sum_{j=0}^{d-1} c_{i,j} z(v_{i+1})^{j-d}$.

The number of linear coefficients n associated with this representation is $d(m-1)+1$. For fixed n, by increasing d, we are balancing an increase in the order of the interpolation against a decrease in the number of points at which we sample. Extensions to higher dimensions are possible, but have not been pursued.

The advantages of the finite-element representation are clear. Enhancement factors in any density variable may be represented, and multidimensional basis sets that span the entire space of GGA functions are easily constructed. A drawback is that the grid points, by their construction (31), formally depend on the choice of data set $\mathbf{D}$, but evidence suggests (see Sec. IV B 2) that such a construction is stable over the range of *chemically relevant* densities. *Finally, we note that our finite-element basis set does not have the usual physical motivations characteristic of other representations.*[14,19]

3. Mixed basis M

The B97 basis provides a global representation of the enhancement factor, while the finite-element basis sets are piecewise in nature. We might thus imagine a mixed basis as being the most flexible representation. For an enhancement factor f, with total linear parameters n, we may choose d_1 terms of the B97 form, and d_2 terms of the F1 form, giving the mixed basis M[d_1, d_2], where $n = d_1 + d_2$.

TABLE I. Median values of density variables over $\mathbf{D}_{18}$.

ρ_α	ρ_β	ρ	x_α^2	x_β^2	x	τ_α	τ_β	τ	y_α	y_β	y
0.515	0.709	1.064	26.977	28.619	17.608	11.847	12.061	7.635	8.932	9.250	5.811

We finish by mentioning that any basis set may be augmented by exact exchange E_X^{exact} through the *ansatz* (22). This introduces the parameter, c_X.

B. The data set

Choosing a data set must remain a somewhat arbitrary procedure. We have been guided by the following *ab initio* aims (i)–(iii):

(i)   it must be systematically extensible;
(ii)   it must cover *chemically relevant* densities;
(iii) it should be well-balanced across the periodic table (i.e., nonbiased, like an *ab initio* method).

1. The MP2–HF model system

To systematically extend a data set [aim (i)], we must be able to generate pairs (ρ, F) at will. This will not be possible with an experimental data set, such as the G2 data set. In principle, *ab initio* methods may be used to generate such data, i.e., to provide a mapping of ρ to the exchange-correlation and kinetic energies.

In this study, we choose the simplest *ab initio* model of correlation, namely MP2 theory.[30] We choose the unrestricted form to handle open-shell systems. The density ρ (and orbitals) and corresponding density variables x, y, τ are generated from unrestricted HF theory, while the exchange-correlation energy corresponding to such an HF density is defined as the sum of the MP2 correlation and HF exchange energies

$$E_{\text{xc}}[\rho] = E_C^{\text{MP2}} + E_X^{\text{HF}}. \tag{38}$$

We shall refer to this as the MP2–HF model of exchange-correlation. Similarly, the kinetic energy corresponding to the HF density is defined as

$$T_s[\rho] = T_s^{\text{HF}}. \tag{39}$$

2. Specification of the data set

The set of *chemically relevant densities* [aim (ii)] is the set of densities of molecular species, with arbitrary numbers of atoms, nuclear charges, electron number, and geometry. Thus, a chemically relevant density may conveniently be indexed by four parameters $\{a, \mathbf{z}, \mathbf{g}, q\}$, these being

a: the number of atoms;
$\mathbf{z}$: a vector denoting the nuclear charges;
$\mathbf{g}$: a vector bundle denoting the geometry of the system;
q: the total charge of the system.

To develop a functional with *ab initio* predictive characteristics [aim (iii)], our data set must be evenly balanced across the periodic table (unlike existing data sets such as the G2 data set). Also, since chemistry can be described primarily on a local level (e.g., in terms of atoms and functional groups), we construct our data set ''from the bottom up'' in terms of system size. This suggests the following algorithm for constructing a class of data sets:
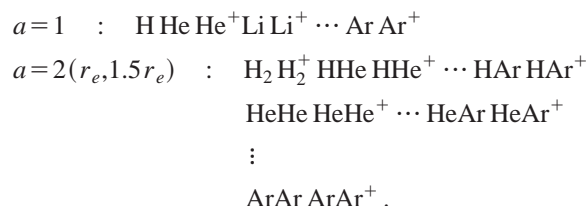
(i)   Choose the maximum a, a_m. If $a_m = 2$, for example, we consider atoms and diatomics.

TABLE II. Number of systems p in data sets $\mathbf{D}_{z_m}$.

z_m	6	9	12	15	18
p	68	144	216	303	429

(ii)   Choose a maximum z, z_m. If $z_m = 18$, for example, we consider all species containing elements up to Ar.
(iii) Choose geometries (e.g., r_e, stretched, etc.) and charges (neutral, cations, etc.) for the systems.

For this work, we chose $a_m = 2$, $z_m = 18$, diatomics at bondlengths $r_e, 1.5r_e$, and neutral and $+1$ cationic species. In principle these parameters generate a total of 719 systems, these being

$$a = 1 \quad : \quad \text{H He He}^+ \text{Li Li}^+ \cdots \text{Ar Ar}^+$$
$$a = 2(r_e, 1.5r_e) \quad : \quad \text{H}_2 \text{ H}_2^+ \text{ HHe HHe}^+ \cdots \text{HAr HAr}^+$$
$$\text{HeHe HeHe}^+ \cdots \text{HeAr HeAr}^+$$
$$\vdots$$
$$\text{ArAr ArAr}^+.$$

Note that by changing the parameters, the data set may be systematically enlarged, in a well-balanced fashion.

We would have liked to increase a_m, as our data set contains only systems of low atomicity. In particular, there are effects which are believed to be more important in larger systems (for example, exact exchange.[31]) However, since the number of systems increases exponentially with a_m, and since we have also included stretched systems, we believe our data set to be a reasonable compromise between universality and practicality. Extensions to triatomics and larger molecules are possible if we limit the type of atoms involved; i.e., choosing some smaller a_m for such systems.

For our analysis in Sec. II E, we will require a partitioning of our data set into a sequence of data sets. Fixing all other parameters, a simple sequence of data sets $\mathbf{D}_{z_m}$ is obtained by varying z_m; thus, our largest data set is written as $\mathbf{D}_{18}$, while the data set $\mathbf{D}_{10}$, for example, would contain all atoms, diatomics, at bondlengths $r_e, 1.5r_e$, neutral and cations, for elements up to Ne. The sizes (the number of data pairs p) of $\mathbf{D}_{z_m}$ for different z_m are listed in Table II.

3. Technical construction

Geometries of the 719 systems in the data set were optimized at the UHF/TZ2P level. The exchange-correlation energy was then computed using Eq. (38) at the UMP2/TZ2P level, and the kinetic energy at the UHF/TZ2P level. The specific Gaussian TZ2P basis sets[40–42] are available as supplementary Table A (see Sec. VII).

Spin multiplicities were assigned using a simple algorithm based on the aufbau principle. Note that data were not collected for systems which

(i)   dissociated, or did not converge in the geometry optimization,
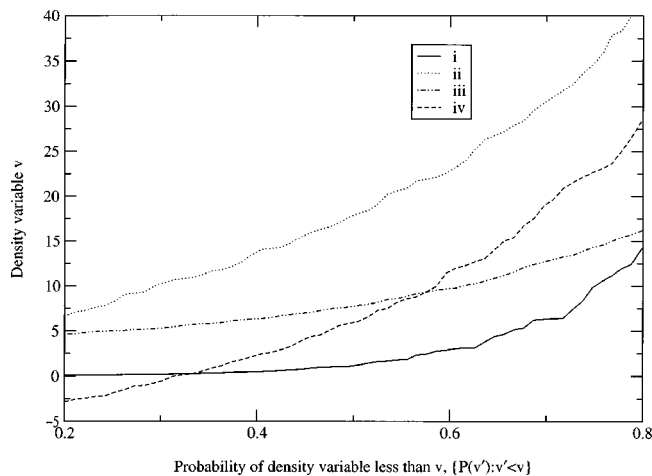(ii)   did not converge in the HF-self-consistent field (SCF) procedure.

FIG. 2. Chemically relevant regions of density variables. (i) $v = \rho$, (ii) $v = x$, (iii) $v = \tau$, (iv) $v = y$.

Of these, (ii) was the greatest difficulty, since the SCF procedure converges poorly for many systems, and does not always converge to a minimum.

After removing systems according to criteria (i) and (ii), our largest data set $\mathbf{D}_{18}$ contains 429 systems. The specific 429 systems are available as supplementary Table B (see Sec. VII).

The density variables ρ, x^2, y, τ (and spin components) were computed at the UHF/TZ2P level on a large Euler–Maclaurin[32] + Lebedev[33] quadrature grid using the CADPAC program package.[34] The grid error is estimated to be better than 1 in 10^5 for all quantities. Our use of Hartree–Fock orbitals, rather than Kohn–Sham orbitals, will lead to a slightly different τ, but for approximating $E_{xc}[\rho]$ (where the contribution of the gradient correction is expected to be small) this is unimportant.

The finite-element basis sets (F1,F2,M) require the partitioning of the density variables into regions of equal probability. These equiprobability grids for each density variable are constructed over the largest data set $\mathbf{D}_{18}$. The finite-element grids yield insight into the ''chemically relevant'' ranges of the density variables ρ, x, y, τ and spin counterparts. Figure 2 plots the variables $v = \{\rho, x^2, y, \tau\}$ against the probability of finding a particle with $v' < v$, $\{P(v'): v' < v\}$. The exponential nature of the density and its gradients can be clearly seen from the plots. The Laplacian y takes on predominantly positive values in the important regions of the data set. Further analysis shows that these plots exhibit similar characteristics if we construct the probability $P(v')$ over some smaller data set $\mathbf{D}_{z_m}: z_m < 18$ (although the range of ρ would be smaller). This demonstrates the dimensionless nature of the density variables (x^2, y, τ), which implies that over *chemically relevant densities, the median values of these density variables are approximately invariant to the choice of system*. For our data set $\mathbf{D}_{18}$, the median values of v are listed in Table I.

C. The error functional

We return to our choice of error functional $\mathbf{\Delta}$. In the following discussion, we will use *energy component* to denote the *kinetic* or *exchange-correlation components*.

A widely used error functional is given by the rms error of the *total* energy components. This *total* error functional, denoted $\mathbf{\Delta}_t$, is defined as

$$\Delta_t[\mathbf{F},\mathbf{D}] = \left[\sum_{i \in \mathbf{D}} (F[\rho_i] - F_i)^2 / p \right]^{1/2}, \tag{40}$$

where $F[\rho_i]$ is the trial (kinetic or exchange-correlation) functional, and F_i is the corresponding ''exact'' input energy component for system i, in the data set $\mathbf{D}$ with p data pairs.

However, it is most convenient to choose some $\mathbf{\Delta}$ which allows relative errors over the sequence of data sets $\mathbf{D}_{z_m}: z_m = 1, \ldots, 18$ to be directly compared. Such a $\mathbf{\Delta}$ should remain relatively invariant in magnitude over the sequence of data sets, characterized by z_m, which is not true of $\mathbf{\Delta}_t$.

Define the *generalized* formation energy F_f, for the different classes of system (diatomics AB and atoms A) in our data set as

$$F_f[AB] = F[AB] - F[A] - F[B]. \tag{41}$$

$$F_f[AB^+] = F[AB^+] - F[A] - F[B], \tag{42}$$

$$F_f[A(Z)] = F[A(Z)] - F[A(Z-1)]. \tag{43}$$

$$F_f[A^+] = F[A^+] - F[A]. \tag{44}$$

Note that the name ''generalized formation energy'' encapsulates several familiar quantities: the first quantity above is the usual atomization energy; the third represents the energy released (''fusion energy'') when a hydrogen atom is fused with a neutral atom of charge $Z-1$, and the last is simply the ionization energy.

Then, we define an approximately invariant error functional $\mathbf{\Delta}_f$[35] as

$$\Delta_f[\mathbf{F},\mathbf{D}] = \left[\sum_{i \in \mathbf{D}} (F_f[\rho_i] - F_{fi})^2 / p \right]^{1/2}. \tag{45}$$

For most of the studies in this work, we shall use $\mathbf{\Delta}_f$ as our error functional.

It remains to mention the least-squares equations. For the functional basis expansion (4) of $F_f[\rho_i]$, minimizing the coefficients $\{c_i\}$ of the error functionals $\mathbf{\Delta}_t$ or $\mathbf{\Delta}_f$ leads to a set of linear least-squares equations, which may be inverted to find the optimum $\{c_i\}$ (which then may be used to construct the enhancement factors, e.g., in the F1 basis, the $\{c_i\} = \{f_i\}$).

V. THE FITTING EXPERIMENTS

Recall from Sec. II, the three important concepts in functional construction are the attainment of the *functional basis set* and *data-set limits*, and the choice of *functional form*.

The above concepts are now examined by the following general investigations:

(i) functional basis-set limit: examination of convergence, determination of the optimum number of parameters, and optimization for each of the basis sets.

(ii) data-set limit: information on the limit is provided by the convergence to our largest data set $\mathbf{D}_{18}$.

(iii) functional form: comparison of representations, e.g., first-order and second-order GGAs and multidimensional forms.

Subsection A is concerned with the exchange-correlation functional, while Subsection B deals with the kinetic energy functional.

Because of the large number of fits performed ($\sim 100\,000$), we give only a representative sample of fitting experiments and results. Our fitting experiments are described in the following format:

opt/eval (description) $\Delta[(\mu, n, z_m), z_m']$.

The notation for the error functional was introduced in Sec. II C. Note that, since our sequence of data sets is conveniently indexed by z_m (see Sec. IV B), we have changed variables to consider the error Δ as a function of z_m, rather than p (the number of data points in D_{z_m}). In addition, opt/eval: opt is a minimization of Δ over the parameters in (description). eval is an evaluation of Δ_0, for the functional **F** determined in the preceding opt.

For example,

opt($\{c\}$,d,d1,n2)$\Delta_f[(\mu, n, z_m), z_m']$,

  for $z_m = z_m' = 18, \mu = \{B97, F1, F2[d], M[d1, d2]\}$

  $\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\} = \{1,1,1\}, \dots, \{16,16,16\}$

indicates: (i) optimize a functional over all expansion coefficients c (and basis parameters d, d1, n2) for some basis $\mu = \{B97, F1, F2[d], M[d1, d2]\}$, by minimizing the error Δ_f in the generalized formation energies over the fitting set $D_{z_m} = D_{18}$. The number of terms used in the expansion of exchange and correlation components ($\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\}$) are the same. (ii) evaluate the error of the optimized functional over the testing set $D_{z_m'} = D_{18}$. (iii) Repeat procedure for different basis sets μ, and different numbers of basis functions (3 to 48).

A. The exchange-correlation functional

We now carry out a set of fitting experiments for the exchange-correlation functional. Sections V A 1–V A 3 deal with one-dimensional fits, where all enhancement factors are functions of x^2 only. Other density variables and multidimensional enhancement factors are treated in Secs. V A 4–V A 5. Exact exchange is treated in Sec. V A 6.

1. General convergence: Functional basis sets

a. Experiment 1: $f_{X\sigma\sigma}(x_\sigma^2), f_{C\sigma\sigma}(x_\sigma^2), f_{C\alpha\beta}(x^2; x_{av}^2)$
Convergence with number of parameters

opt($\{c\}$,d,d1,n2)$\Delta_f[(\mu, n, z_m), z_m']$,

  for $z_m = z_m' = 18, \mu = \{B97, F1, F2[d], M[d1, d2]\}$,

  $\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\} = \{1,1,1\}, \dots, \{16,16,16\}$.

Plotted in Fig. 3 is the error Δ_{f0} against n, for the different basis sets. For $n = 3$, $\mathbf{F}_0$ is a reoptimized local density approximation (LDA) with an error of 20.9 mH (13.10 kcal/mol). The addition of the first gradient terms in ex-
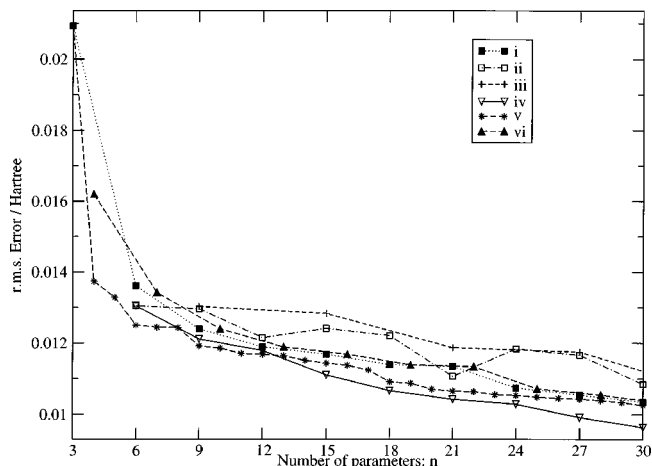

FIG. 3. rms error Δ_{f0} in E_{xc} with different functional basis sets, as a function of the number of parameters n. (i) B97 (experiment 1), (ii) F1 (experiment 1), (iii) F2 [2] (experiment 1), (iv) M [optimized d1,d2] (experiment 1), (v) B97 (optimized $\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\}$) (experiment 2b), (vi) B97 + exact exchange (experiment 6).

change and correlation lead to a rapid decrease in the error. However, there is a plateau at ~ 10 mH (6.27 kcal/mol), for *all four basis sets*—an indication that we are reaching the functional basis-set limit. The polynomial interpolation F2 (line iii) leads to no improvement over the simple F1 basis (line ii). The mixed basis M (line vi) gives the best results, and generally improves on B97 (line i) by about 0.5 mH (0.31 kcal/mol). However, all four basis sets perform similarly, and the error is essentially *independent of basis and a function of the number of parameters n*.

The last point, that the *performance of the functional is often independent of the nature of the functional basis expansion used*, is one which is often overlooked in the literature. It is common to correct a deficient functional by mixing in (with an adjustable parameter) a new term which is ''physically justified.'' However, *one must distinguish between the success of a term due to its physical construction, or due to the additional parameter it entails*. In other words, would another term (together with another adjustable parameter) be just as good? Our results suggest that additional parametrization, rather than physical construction, may lie behind the success of many functional forms.

Previous studies of the B97 functional[15,17] have found plateaus at roughly 5 mH (3.1 kcal/mol). As our plateau is insensitive to the choice of nonlinear parameters $\gamma_{X\sigma\sigma}, \gamma_{C\sigma\sigma}, \gamma_{C\alpha\beta}$ in the B97 basis, and the size of the data set (see Sec. V A 3), we conclude that our relatively high rms error, must reflect an additional difficulty associated with working in the MP2–HF model of exchange-correlation.

If we plot the enhancement factors $f_{X\sigma\sigma}$, $f_{C\sigma\sigma}$, $f_{C\alpha\beta}$ separately for the different functional basis sets, we find that there is great variation in their form. This does not come as a surprise, since the spin separation (9) is not unique. A more useful quantity to examine graphically is the total enhancement factor f_{XC}, defined via

$$E_{xc}[\rho_\alpha, \rho_\beta] = \int E_X^{UEG} f_{XC}(\rho_\alpha, \rho_\beta, x_\alpha^2, x_\beta^2, x^2) d\mathbf{r}. \quad (46)$$
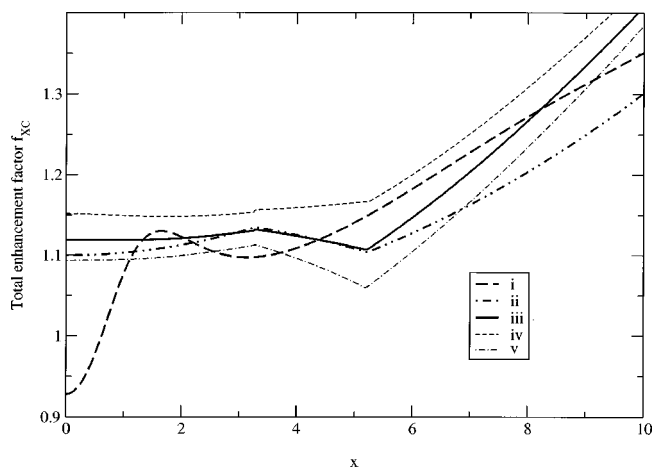
FIG. 4. The closed-shell total enhancement factor $f_{XC}(x)$ at $\rho=1.06$ (experiment 1). (i) B97 (15 parameters), (ii) F1 (12 parameters), (iii) M [1,4] (15 parameters) $\rho=1.06$, (iv) M [1,4] $\rho=0.15$, (v) M [1,4] $\rho=10.0$.

Plotted in Fig. 4 is the total enhancement factor f_{XC} for closed-shell systems ($\rho=1.06$, $x_\alpha=x_\beta=x/2$), determined from experiment 1, with F1 (12 parameters) and B97, M[1,4] (15 parameters). As can be seen, over the range of chemically relevant ρ,x, there is *good convergence to the same form*, which indicates the stability of our fitting methodology. Moreover, the simplicity of the form is encouraging.

We now see whether our enhancement factors obey important exact conditions over the chemically relevant ranges of ρ,x. Plotted are the enhancement factors for M[1,4] at $\rho=0.15$ and $\rho=10.0$. The noncrossing condition[12] $f_{XC}(\rho,x)$ $\leqslant f_{xc}(\rho',x):\rho>\rho'$ is satisfied. All enhancement factors satisfy the Lieb–Oxford bound (now tightened, by Chan and Handy, see Ref. 20) $f_{XC}\leqslant 2.21$. Finally, we note that in the limit of zero gradient x (not in the chemically relevant range), the different basis sets yield LDA coefficients which differ from the exact coefficient, by a modest 7%–10%. Thus, we conclude that to some extent, exact conditions and bounds in the chemically relevant regions of density variables are approximately encapsulated within our data set.

As a representative functional, the coefficients and grids for the 15-parameter M[1,4] functional are given in Table III. Note that the large coefficients at the edge of the grid points are merely an artifact of the last grid point being at such

large x_σ^2, and correspond to very modest values for the enhancement factor over chemically important values of x_σ^2, as can be seen from the previous plots above.

To summarize:

(i)   for all basis sets, we reach the same plateau in the error at roughly 10 mH (6.27 kcal/mol), which is near the functional basis-set limit within the MP2–HF model,

(ii)   the error is a function of the number of parameters, and almost independent of the choice of basis,

(iii) our enhancement factor is a well-converged simple form, and obeys a number of exact conditions for chemically relevant ranges of density variables.

2. Optimization of number of parameters

The determination of n_0 through Eq. (8) requires a complete data set as a testing set. Here, we make the simplifying approximation (!) that for suitably small fitting sets ($\mathbf{D}_{z_m}$, $z_m\leqslant 15$), $\mathbf{D}_{18}$ may be regarded as a complete data set with regard to z_m.

a. Experiment 2a: $f_{X\sigma\sigma}(x_\sigma^2), f_{C\sigma\sigma}(x_\sigma^2), f_{C\alpha\beta}(x^2;x_{av}^2)$
(i) Fit to fitting set

$$\text{opt}(\{c\})\mathbf{\Delta}_f[(\mu,n,z),z_m']\quad\text{for }z_m=z_m'=6,9,\ldots,18,$$

$$\mu=\{B97,F1\},$$

$$\{n_{X\sigma\sigma},n_{C\sigma\sigma},n_{C\alpha\beta}\}=\{1,1,1\},\ldots,\{16,16,16\}.$$

(ii) Optimize with regard to n over testing set $\mathbf{D}_{18}$

$$\text{opt}(n)\mathbf{\Delta}_{f0}[(\mu,n,z_m),z_m'],\quad\text{for }z_m=6,9,12,15,$$

$$z_m'=18,\text{ and other parameters as above.}$$

The error $\mathbf{\Delta}_{f0}$ as evaluated in experiment 2a ii exhibits an early minimum corresponding to n_0 as can be seen in Fig. 5, where we plot $\Delta_{f0}[(B97,n,9),18]$ against n. We tabulate estimated n_0 for the B97 basis, and F1 basis in Table IV. The values for the largest data set $\mathbf{D}_{18}$ are obtained by extrapolation.

The relatively small n_0, even for the largest data set $\mathbf{D}_{18}$ (429 systems), reflects the difficulty of the exchange-

TABLE III. Exchange-correlation functional from experiment 1 using M[1,4] basis: enhancement factors and rms error. The basis sets are described in Sec. IV A.

| B97 polynomial order $i-1$	1			
$f_{X\sigma\sigma}$ coefficient	0.656 25			
$f_{C\sigma\sigma}$ coefficient	−0.182 64			
$f_{C\alpha\beta}$ coefficient	−2.012 29			
Grid points (x_σ^2) (see [39])	0.000 000	16.817 23	42.539 36	$2.396\,744\times10^{25}$
$f_{X\sigma\sigma}(x_\sigma^2)$ at grid points	1.054 720	1.044 054	0.892 291 2	$4.152\,212\times10^{22}$
$f_{C\sigma\sigma}(x_\sigma^2)$ at grid points	0.714 585 8	−0.593 533 1	2.760 651	$-1.806\,301\times10^{23}$
Grid points (x^2)	0.000 00	10.777 98	27.248 52	$9.834\,540\times10^{21}$
$f_{C\alpha\beta}(x^2)$ at grid points	0.633 024 9	1.946 529	0.635 097 1	$3.933\,448\times10^{19}$
Δ_{f0} [(M[1,4],15,18),18]/mH	11.10			
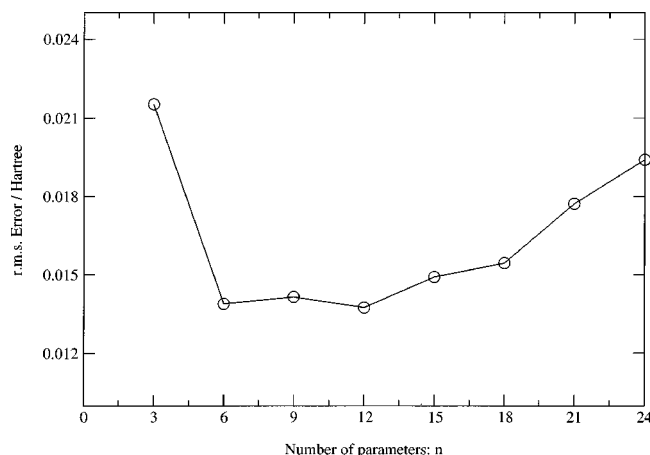
FIG. 5. Determination of optimum number of parameters (E_{xc}): rms error of functional fitted to $\mathbf{D}_9$ evaluated over $\mathbf{D}_{18}$: $\Delta_{f0}[(B97,n,9),18]$ against n (experiment 2).

correlation problem, and is consistent with other results in the field.[15,16] Note that n_0 seems *insensitive* to the choice of basis μ.

For the exchange-correlation functional, there remain additional discrete degrees of freedom: the distribution of parameters between $n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}$. We optimize over these parameters by generalizing experiment 2a. Thus:

b. Experiment 2b: $f_{X\sigma\sigma}(x_\sigma^2), f_{C\sigma\sigma}(x_\sigma^2), f_{C\alpha\beta}(x^2; x_{\mathrm{av}}^2)$

(i) Repeat experiment 2a i with

$$\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\} = \{i,j,k\}, \quad \text{for } i = 1, \ldots, 12,$$

$$j = 1, \ldots, 12, \quad k = 1, \ldots, 12.$$

(ii) Repeat experiment 2 a ii and optimize over $\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\}$, $z_m = 6, 9, \ldots, 18$, $z_m' = 18$.

The n_0 determined from experiment 2b are found to be essentially the same as those from experiment 2a. Listed in Table V are the optimum distributions of parameters $\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\}$ with the (B97, F1) basis sets, for $z_m = z_m' = 18$. Recall from Fig. 3, that there is a large drop in the error between $n = 3$ and $n = 6$. We see that this is probably due to an improvement in the description of exchange.

In Fig. 3 we also plotted the error $\Delta_{f0}[(B97,n,18),18)]$, with an optimized distribution of parameters, as compared with the B97 functional with an even distribution of parameters (see experiment 1). For a total number of parameters $n = 15$, our optimized parameter distribution yields an improvement of roughly 0.25 mH (0.16 kcal/mol).

To summarize:

(i) A minimum according to Eq. (8) may be found to determine n_0. n_0 is small, ~ 15 for our largest data set and insensitive to the basis μ.

TABLE IV. Estimated optimum number of parameters n_0 (exchange-correlation functional) as a function of data set $\mathbf{D}_{z_m}$, as in experiment 2a.

z_m	6	9	12	15	18
B97	9	12	12	15	15
F1	6	12	12	12	12

3. Convergence with data sets

a. Experiment 3a: $f_{X\sigma\sigma}(x_\sigma^2), f_{C\sigma\sigma}(x_\sigma^2), f_{C\alpha\beta}(x^2; x_{\mathrm{av}}^2)$
Convergence with fitting set

$$\mathrm{opt}(\{c\})\Delta_f[(\mu, n_0, z_m), z_m'], \quad \text{for } z_m = z_m' = 6, 9, 12, 15,$$

$$\mu = \{B97, F1\}, n_{X\sigma\sigma} = n_{C\sigma\sigma} = n_{C\alpha\beta}.$$

Plotted in Fig. 6 is the resulting error Δ_{f0} against z_m (line i: using B97). As the fitting data set z_m is increased, the error Δ_{f0} remains stable. We thus conclude that our large data set $\mathbf{D}_{18}$ is *well converged* to the data set limit with respect to z_m. However, to infer the same for the other data set parameters ($a_m, \mathbf{g}, q_m$) will require further investigation, with expensive larger data sets.

b. Experiment 3b: $f_{X\sigma\sigma}(x_\sigma^2), f_{C\sigma\sigma}(x_\sigma^2), f_{C\alpha\beta}(x^2/x_{\mathrm{av}}^2)$
Convergence with testing set: evaluate the error of the functionals determined in experiment 3a, over larger testing sets

$$\text{eval } \Delta_{f0}[(\mu, n_0, z_m), z_m'], \text{for } z_m = 6, 9, \ z_m' = 6, \ldots, 18.$$

Also plotted in Fig. 6 is Δ_{f0} with $z_m = 6$, as a function of the testing set parameter z_m'. With both the B97 (line ii) and F1 basis sets (line iii), the functionals fitted to the small data set D_6, using the optimum number of parameters n_0, are very stable to increasing the test set parameter z_m'. Note that the B97 basis sets and F1 basis sets perform similarly well. Thus, although different in nature, these basis sets indeed offer a similar description over the range of chemical densities, in agreement with our earlier plots of the enhancement factors (Fig. 4).

Finally, in Fig. 6, we also plot the error $\Delta_{f0}[(F1,9,6), z_m']$ (where we have used greater than the optimum number of parameters $n = 9 > n_0$) against the test set z_m'. Here, the error of the functional is very unstable to increasing the size of the testing set, which is a clear sign of overfitting. Since purely by examining a convergence curve such as the one in Fig. 3, it would have been hard to decide whether to stop at $n_0 = 6$ (optimum) or $n = 9$ (as above), our accurate choice of n_0 for the data set $\mathbf{D}_6$ is a success of our formal theory for determining n_{opt}.

To summarize:

TABLE V. Optimized distribution of parameters for exchange and correlation for increasing total number of parameters n (fitted to $\mathbf{D}_{18}$), as in experiment 2b.

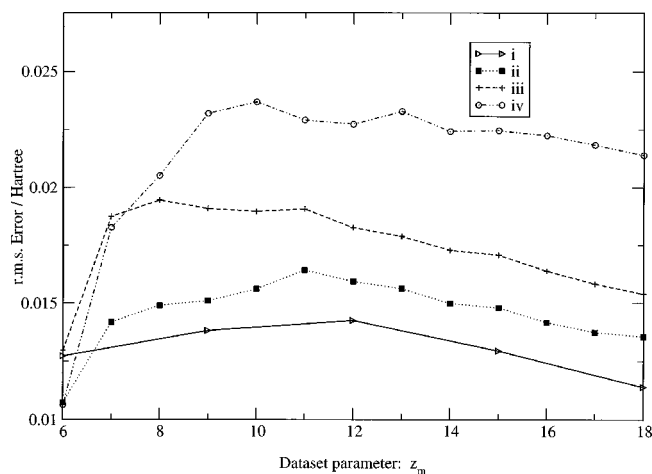|	3	4	5	6	7	8	9	10	11	12	15
B97 $\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\}$	1,1,1	2,1,1	3,1,1	2,1,3	2,2,3	3,2,3	2,4,3	2,5,3	2,6,3	2,7,3	5,7,3
F1 $\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\}$				2,2,2	3,2,2	2,2,4	3,2,4	2,4,4	2,5,4	4,4,4	

FIG. 6. Data set convergence (E_{xc}): rms errors. (i) Increasing fitting set z_m (experiment 3a), (ii) Increasing testing set z_m: B97 (experiment 3b), (iii) Increasing testing set z_m: F1 (experiment 3b), (iv) Increasing testing set z_m: F1 (nonoptimum number of parameters).

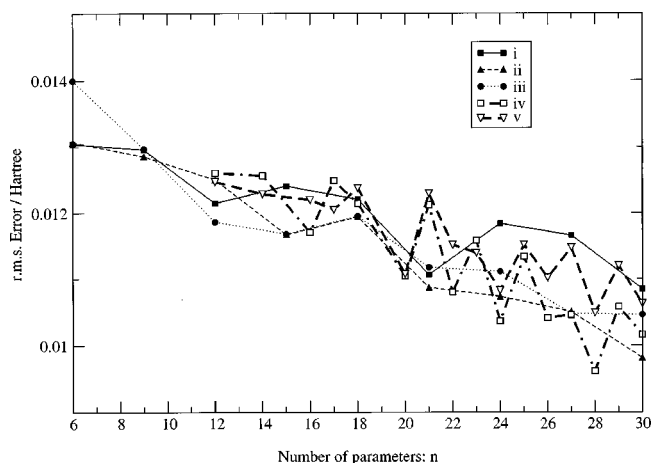FIG. 7. Second-order density variables and two-dimensional enhancement factors (E_{xc}). (i) $f_{X\sigma\sigma}(x_\sigma), f_{C\sigma\sigma}(x_\sigma), f_{C\alpha\beta}(x)$ (experiment 4), (ii) $f_{X\sigma\sigma}(\tau_\sigma), f_{C\sigma\sigma}(\tau_\sigma), f_{C\alpha\beta}(\tau)$ (experiment 4), (iii) $f_{X\sigma\sigma}(y_\sigma), f_{C\sigma\sigma}(\tau_\sigma), f_{C\alpha\beta}(\tau)$ (experiment 4), (iv) $f_{X\sigma\sigma}(x_\sigma, \tau_\sigma), f_{C\sigma\sigma}(x_\sigma, \tau_\sigma), f_{C\alpha\beta}(x, \tau)$ (experiment 5), (v) $f_{X\sigma\sigma}(x_\sigma, \tau_\sigma), f_{C\sigma\sigma}(\rho_\sigma, x_\sigma), f_{C\alpha\beta}(\rho, x)$ (experiment 5).

(i)   Our largest data set is well converged to the data-set limit, with respect to the nuclear charge parameter z_m.

(ii)   Our choice of n_{opt} yields functionals that are stable to increasing the size of the testing set.

4. Density variables

Up till now, we have investigated the commonest form of the first-order GGA where the enhancement factors are functions of x^2 only. Before moving on to multidimensional fits, we investigate one-dimensional enhancement factors in other density variables. Writing the exchange-enhancement factor as $f_{X\sigma\sigma}(v_{1\sigma})$, and the correlation enhancement factors as $f_{C\sigma\sigma}(v_{2\sigma}), f_{C\alpha\beta}(v_2)$, and allowing v_1, v_2 to range over $\{\rho, x^2, y, \tau\}$ we have 16 combinations $(v_1, v_2) = (\rho, \rho), (\rho, x^2), \ldots, (y, \tau), (y, y)$.

a. Experiment 4: $f_{X\sigma\sigma}(v_{1\sigma}), f_{C\sigma\sigma}(v_{2\sigma}), f_{C\alpha\beta}(v_2)$

For each combination (v_1, v_2) repeat experiment 1, for $\mu = $F1.

Plotted in Fig. 7 is the error Δ_{f0} against n, for selected density variable combinations: (x^2, x^2) (line i), (τ, τ) (line ii), (y, τ) (line iii). Although not shown, the density-only (ρ, ρ) enhancement factors perform worse [by about 10 mH (6.27 kcal/mol)] than the gradient-variable enhancement factors. The usual choice of (x, x) is near optimal, particularly near $n_0 \sim 12$–15. The use of second-order gradient variables yields at best a modest improvement: with the choice (y, τ), we gain approximately 0.25 mH (0.16 kcal/mol) improvement over the (x^2, x^2) combination near the optimal number of parameters (12). Other choices of density variables [such as (τ, τ)] give similar curves that lie slightly above the y, τ curve.

We note that recent studies of popular functionals by Cohen and Handy have yielded the similar conclusion that little benefit is to be gained from the use of second-order density variables τ, y.[36]

Such findings contrast strongly with the physical reasons that have been put forth to advocate the use of second-order

density variables.[10,14] For example, the kinetic energy density is believed to be important in describing nondynamical correlation[37] in stretched bonds, and also in the description of so-called congested systems such as N_2. However, we have *not* found noticeable improvements in the description of such systems with our set of optimized functionals.

The enhancement factors in this section, due to their dependence on multiple density variables, cannot be simply compared graphically, and we do not include them here. However, we have found that in general, the dependence on the density variables is encouragingly smooth and simple.

To summarize:

(i) The usual choice of $f_{X\sigma\sigma}(x), f_{C\sigma\sigma}(x), f_{C\alpha\beta}(x)$ is near optimal. Second-order density variables, such as $f_{X\sigma\sigma}(y), f_{C\sigma\sigma}(\tau), f_{C\alpha\beta}(\tau)$ yield a modest improvement at best.

5. Multidimensional fits

As stressed in Secs. III A, the GGA exchange-correlation enhancement factors are multidimensional functions. Thus, for a complete solution to the GGA problem we *must* perform multidimensional fits. We now consider enhancement factors that are two-dimensional functions. Writing the enhancement factors as $f_{X\sigma\sigma}(v_{1\sigma}, v_{2\sigma}), f_{C\sigma\sigma}(v_{3\sigma}, v_{4\sigma})$, $f_{C\alpha\beta}(v_3, v_4)$, and allowing $v_1, \ldots, v_4$ to range over $\{\rho, x^2, y, \tau\}$, we have 18 combinations $(v_1, v_2, v_3, v_4) = (x^2, y, \rho, x), \ldots, (y, \tau, y, \tau)$.

a. Experiment 5: $f_{X\sigma\sigma}(v_{1\sigma}, v_{2\sigma}), f_{C\sigma\sigma}(v_{3\sigma}, v_{4\sigma})$, $f_{C\alpha\beta}(v_3, v_4)$

For each combination (v_1, v_2, v_3, v_4),

$$\text{opt}(\{c\}, d1, d2, d3, d4) \Delta_f[(\mu, n, z_m), z_m']$$

for $z_m = z_m' = 18$, $\mu = F1, n_{X\sigma\sigma} = d_1 d_2$,

$$n_{C\sigma\sigma} = n_{C\alpha\beta} = d_3 d_4,$$

for $d_1, d_2, d_3, d_4 \leqslant 12, d_1 d_2, \quad d_3 d_4 \leqslant 24.$

d_1, d_2, d_3, d_4 denote the number of grid points along the v_1, v_2, v_3, v_4 axes, respectively. The conditions on d_1, d_2, d_3, d_4 set above ensure that (i) no axis has more than 12 grid points, and (ii) each enhancement factor has a maximum of 24 parameters. The optimization over d_1, d_2, d_3, d_4 ensures that for each combination that yields the same n, the one that minimizes the error is selected.

Plotted in Fig. 7 is the error Δ_{f0} against n for some representative combinations (x, τ, ρ, x) (line iv) and (x, τ, x, τ) (line v). Since the smallest two-dimensional grid contains 4 points, the minimum $n = 12$.

As one can see from comparing Fig. 3 with Fig. 7, there is little real improvement in going from one-dimensional to two-dimensional representations of the enhancement factors. At $n = 12$ the error is already in the plateau region; near the previously determined $n_0 \sim 12-15$, no improvement over the one-dimensional enhancement factor results is seen. It might be argued that n_0 is slightly larger for two-dimensional enhancement factors, and that the F1 basis is not optimum, but be that as it may, we do not expect a drastic reduction in Δ_{f0}.

From our two-dimensional studies, it seems that the one-dimensional limit is already close to the limit of the second-order GGA form. Increasing the dimensionality of the enhancement factors (by going to higher-order GGAs) seems *unlikely to yield any improvement*. In principle, for a complete data set, multidimensional enhancement factors must yield some improvement, but this is expected to be small, and appears to be nonexistent for our finite data sets.

From Sec. V A 4, and the current section, we summarize our findings as:

(i)   Little improvement is gained from increasing the dimensionality (and order) of the enhancement factors,
(ii)   The functional basis-set limit of the second-order GGA form can give only a very modest improvement over the first-order GGA form.

We thus conclude that since no benefit is gained from going to higher-order density variables, or higher dimensions, and that convergence with respect to the functional basis has been achieved (Sec. V A 1), we have achieved a near-optimum representation of the GGA for the MP2–HF model.

6. Exact exchange

We assess the extension of the GGA form to include exact exchange.

a. Experiment 6: $f_{X\sigma\sigma}(x_\sigma^2), f_{C\sigma\sigma}(x_\sigma^2), f_{C\alpha\beta}(x_\sigma^2; x_{av}^2), E_X^{HF}$

$\text{opt}(\{c\}, c_X)\Delta_f[(\mu, n, z_m), z_m'],$

for $z_m = z_m' = 18, \mu = B97 + E_X^{HF},$

$\{n_{X\sigma\sigma}, n_{C\sigma\sigma}, n_{C\alpha\beta}\} = \{1,1,1\}, \ldots, \{16,16,16\},$

$n = n_{X\sigma\sigma} + n_{C\sigma\sigma} + n_{C\alpha\beta} + 1.$

Repeat, using Δ_t.

Plotted in Fig. 3 is the error Δ_{f0} against n for the B97+ E_X^{HF} basis, and the B97 basis. Aside from the initial four-

parameter LDA+E_X^{HF} (where $c_X = 0.28$), the improvement from including exact exchange is very small. In agreement with this, c_X in the 13-parameter functional is only -0.0137. Similar conclusions are drawn from a study of the total error Δ_t.

Adamson[38] and van Voorhis and Scuseria,[19] in empirical fits to the G2 data set, have also concluded that exact exchange provides little improvement. However, in recent fits, work by Becke and others[14,15] has demonstrated a significant effect of exact exchange.

This may seem a paradox. However, we note that as n increases and the basis sets saturate, the improvement yielded by exact exchange becomes negligible, as does the magnitude of c_X, and thus the two basis sets B97, B97+E_X^{HF} converge *to roughly the same limit*. This implies exact exchange offers no real improvement to the *functional form* and *functional basis-set limit*, but may yield an improvement in Δ (i.e., quicker convergence to the limit) while the functional basis is not saturated. This then calls into question the importance of the adiabatic connection argument[28] for exact exchange. However, we stress that verification of these ideas is needed for exact exchange-correlation energies (as opposed to our MP2–HF model), perhaps with larger Gaussian basis sets.

To summarize:

(i) Exact exchange does not significantly improve the *functional basis limit*. However, for modest parameter values, convergence to the limit may be accelerated. This may be the case when representing *exact* exchange-correlation energies.

B. The kinetic energy functional

We now turn our attention to the kinetic energy functional. Our treatment in this section will be simpler, as many aspects of our methodology have been discussed in the preceding sections.

1. Functional basis convergence and density variables

a. Experiment 7: $f_{T_s}(v_\sigma)$ for $v = \{x, y\}$

(a) General fit

$\text{opt}(\{c\}, d)\Delta_t[(\mu, n, z_m), z_m']$ for $z_m = z_m' = 18,$

$\mu = F1$, (B97,M[d] for $v = x$), $n = 1, \ldots, 24.$

Repeat for Δ_f.

(b) Determine n_0

Repeat experiment 2a i, 2a ii,   with $n = 1, \ldots, 24.$

For a first investigation of the kinetic energy, the error in the *total kinetic energy* Δ_t is of the most interest. Plotted in Fig. 8 is Δ_{t0} against n, with basis sets F1 (density variables x^2: line i, y: line iv), B97 (line ii), and M (density variable x^2: line iii). There is a rapid convergence of the error to a plateau ~ 100 mH (62.70 kcal/mol). Again, as in Sec. V A 1, the convergence to the same plateau with the B97, F1, and M basis sets indicates that we are near the basis-set limit. There is also little to choose between the x^2 and y representations.
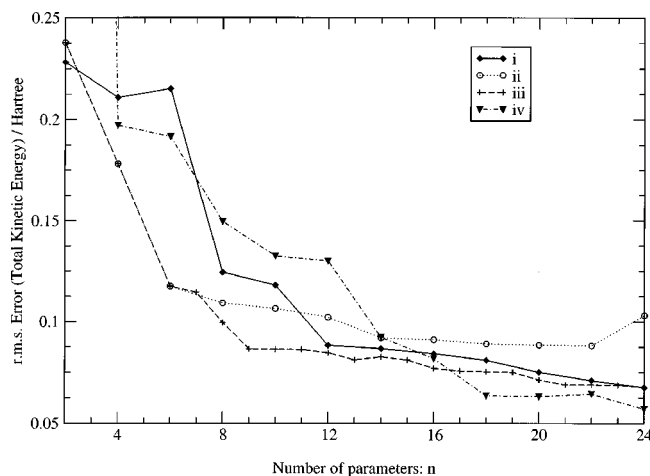
FIG. 8. rms error Δ_{t0} in the total kinetic energy T_s as a function of the number of parameters. (i) F1 (experiment 7a), (ii) B97 (experiment 7a), (iii) M [optimized d1,d2] (experiment 7a), (iv) F1 (y) (experiment 7a).
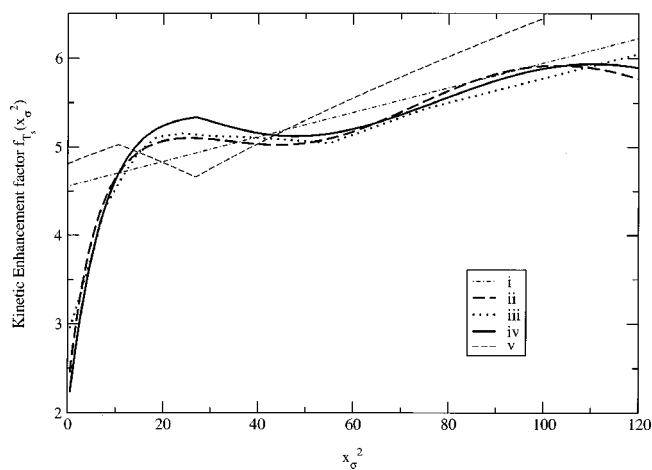

FIG. 9. Kinetic enhancement factor $f_{T_s}(x_\sigma)$. (i) Second-order (von Weizsäcker) gradient expansion, (ii) B97 (12 parameters), (iii) F1 (14 parameters), (iv) M [9,3] (12 parameters), (v) Conjoint exchange $f_{X_{\sigma\sigma}}$ M [1,4].

Although not plotted, as expected the plateau in the error Δ_{f0} occurs at a slightly lower level, ~ 50 mH (31.35 kcal/mol).

The determination of n_0 in experiment 7b is somewhat more difficult than with the exchange-correlation case, as the minimum with respect to n is not easy to locate. The n_0 for the B97 and F1 (density variable x), and M1 representations for D_{18} are estimated to be 12, 14, and 12, respectively. The corresponding errors in Δ_{t0} with n_0 parameters for B97 and F1 (density variable x) basis sets over D_{18} are 102.1 mH (64.01 kcal/mol) and 88.2 mH (55.30 kcal/mol), respectively. With the 12-parameter mixed basis $M[3,9]$, the error is even lower at 84.6 mH (53.04 kcal/mol), which is small given the large magnitude of the total kinetic energy. Such errors are a real improvement over the gradient expansion: for comparison, the error Δ_t of the gradient expansion (24) over D_{18} are 28 141 mH to zeroth order (Thomas–Fermi kinetic energy), and 1788 mH to second order (Thomas–Fermi + von Weizsäcker). Refitting the first two terms in the gradient expansion, we obtain 1568 mH to zero order, and 228.3 mH to second order.

An accuracy of ~ 80 mH (50 kcal/mol) in chemical systems is similar to the accuracy of the local-density approximation for exchange. The individual errors for molecules, available in supplementary Table B, are again good, the maximum error being only ~ 300 mH (188 kcal/mol).

It is natural to consider the second-order GGA $f_{T_s}(x_\sigma, y_\sigma)$. We have carried out fits analogous to experiment 5, but as in the case of the exchange-correlation enhancement factors, have found no improvement.

The relatively low Δ_{t0} is a cause for cautious optimism for orbital-free density functional theory. However, it must be remembered that the real test of a kinetic energy functional is its performance in a *variational calculation*.

The coefficients and grids defining the M[3,9] kinetic energy functional referred to above are listed in Table VI. We remark again that the large coefficients at the edge of the grids are an artifact of those grid points at ''infinity.'' The large values of the B97 coefficients are normalized by the correspondingly small values of $\gamma_{X\sigma}^{i-1}$ in the B97 basis (28). Finally, we need not worry about the oscillating values of the coefficients, as the resulting form of the enhancement factor is well converged and stable, as can be seen in the plots of the next section.

(i) The kinetic energy can be fitted well, with rms errors in the total kinetic energy of only ~ 80 mH (50 kcal/mol).

(ii) The functional must be tested in a variational calculation before further conclusions can be drawn.

2. Kinetic energy enhancement factors

There have been few studies on the form $f_{T_s}(\omega_\sigma)$. In Fig. 9, we plot $f_{T_s}(x_\sigma)$ for F1 (14 parameters: line ii), B97 (12 parameters: line iii), and M[3,9] (12 parameters: line iv). The stability of our fitting approach is exhibited by the beau-

TABLE VI. Kinetic energy functional from experiment 7a using M[9,3] basis: enhancement factor and rms error. The basis sets are described in Sec. IV A.

B97 polynomial order $i-1$	1	2	3	4	5	6	7	8	9
coefficient	72.489 42	−1837.165	15 132.22	−82 593.86	310 818.1	−778 775.3	1 189 505	−981 182.0	331 102.8
Grid points (x_σ^2) (see Ref. 39)	0.000 000	26.977 23	$2.396\,744 \times 10^{25}$						
$f_{T_s}(x_\sigma)$ at grid points	2.002 066	7.002 935	$3.922\,333 \times 10^{24}$						
Δ_{t0} [(M[9,3],12,18),18]/mH	84.56								

tiful convergence of *all basis sets* to a general form: the enhancement factor $f_{T_s}(x_\sigma)$ initially grows with x_σ, has a characteristic kink, and then gradually levels out.

Also plotted in Fig. 9 (line i) is the well-known second order gradient expansion of the kinetic energy,[4] which yields an enhancement factor $C_F + C_W(x_\sigma)^2$. We see that the largest corrections from our empirical enhancement factor occur at small x, but in general, they oscillate around the second-order gradient expansion.

Not plotted is $f_{T_s}(y_\sigma)$. This function is approximately linear in y, but less well characterized than $f_{T_s}(x_\sigma)$. Nonetheless, it is extremely encouraging that the kinetic energy enhancement factor is well approximated by such simple functions.

Lee, Lee, and Parr have conjectured that the exchange-only enhancement factor $f_X(x_\sigma)$ resembles the $f_{T_s}(x_\sigma)$ due to the so-called conjointness of the kinetic and exchange energies.[5] While we cannot strictly extract $f_X(x_\sigma)$ from our work, we may use $f_{X\sigma\sigma}$ as a first-order approximation. We have plotted $C_F f_{X\sigma\sigma}$ for the M[1,4] basis of Sec. V A 1. As can be seen, there is a broad resemblance between the approximate exchange and kinetic energy enhancement factors, which is consistent with the conjointness conjecture.

To summarize:

(i) $f_{T_s}(x_\sigma)$ rises with x and tails off. It is a simple function and behaves like an attenuated second-order gradient expansion. Furthermore, there is some resemblance between the kinetic energy and exchange energy enhancement factors.

VI. CONCLUSIONS

In this work we have demonstrated the power of a systematic and general method to construct functionals. Amongst other things, our experiments have included examinations of functional basis sets, convergence of data sets, optimization of numbers of parameters, and multidimensional fits.

In the exchange-correlation case, we have found the functional basis-set limit within the GGA representation. To the extent to which the MP2–HF model, and our data set of atoms and diatomics, reflects the true exchange-correlation problem, we believe we have obtained a solution of the GGA problem consistent with our original aims (i)–(iii). It may seem disappointing that we could not reduce the rms error below a relatively high plateau, which for the MP2–HF model ~ 10 mH (6.27 kcal/mol). However, let us remember that the GGA approximation proposes a model of exchange-correlation by simple universal functions (the enhancement factors) of a few variables! While the utility of this immense simplification reflects the power of density functional theory, we should not overestimate its abilities.

Of course, we have chosen to study a well-defined *model* of exchange-correlation only, and going to more exact exchange-correlation energies and polyatomic systems will influence our findings. We are currently pursuing these extensions. However, we feel that the broad conclusions will remain similar.

In the case of the kinetic energy functional, we found that kinetic energies could be reproduced relatively accu-

rately, to within ~ 80 mH (50.16 kcal/mol). Moreover, a beautifully simple form was found for the enhancement factor $f_{T_s}(x_\sigma)$. We recognize that the construction of a viable kinetic energy functional for *variational* orbital-free calculations still remains a difficult problem. However, our results indicate that an empirical approach may be the way forward.

Our investigations have pushed the limits of the GGA form, the *local* integrals (10)–(12). We have also shown that little further gain is expected from going to higher-order or higher-dimension enhancement factors. The next level approximation would be to consider *double* integrals of the form (3). In the exchange-correlation case, the incorporation of exact exchange may be regarded as a step in this direction, but as we have shown, it is questionable how successful this is. The functional space is enormous, and the convergence with respect to the dimensionality of the integral to the exact functional may not be rapid.

We finish with some general remarks on our fitting methodology and philosophy. Our aim has been to be systematic and exhaustive, and we have not developed functionals based on any specific physical model. We have also favored a ''brute force'' fitting technique over the incorporation of exact constraints. With respect to physical models, our examinations have in fact shown that physical terms may often owe their improved performance solely to increased parametrization. We have questioned, for example, the importance of ''physically important'' terms such as the kinetic energy density and exact exchange. Of course, we *cannot* underestimate the importance of physical reasoning! Nonetheless, we feel from our investigation that extensive and systematic *empirical* studies form a viable alternative and in many ways complementary way forward, and are often the only way to distinguish the *relative importance*, in practical situations, of different physical arguments and constraints. If a physical constraint is sufficiently important, it will be recovered in an empirical approach.

Finally, we stress that our functional basis sets are general, and we have a well-balanced data set, which is in many ways complementary to the more commonly used data sets. Our method for determining the optimum number of parameters is also superior to less rigorous methods. Thus, we believe that our fitting methodology will advance the systematic development of future functionals.

VII. SUPPLEMENTARY MATERIAL

Additional supplementary data, including supplementary Tables A and B, may be obtained from the authors (G. K-L. Chan, garnet@theor.ch.cam.ac.uk) or on the web at:

http://ket.ch.cam.ac.uk/people/garnet/supplementary.html

ACKNOWLEDGMENTS

[1] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, Phys. Rev. B **46**, 6671 (1992).

[2] S. Kurth, J. P. Perdew, and P. Blaha, Int. J. Quantum Chem. **75**, 889 (1990).

[3] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[4] R. G. Parr and W. Yang, *Density Functional Theory of Atoms and Molecules* (Oxford University Press, Oxford, 1989).

[5] H. Lee, C. T. Lee, and R. G. Parr, Phys. Rev. A **44**, 768 (1991).

[6] D. J. Lacks and R. G. Gordon, J. Chem. Phys. **100**, 4446 (1994).

[7] T. A. Wesolowski, Y. Ellinger, and J. Weber, J. Chem. Phys. **108**, 6078 (1998).

[8] J. P. Perdew, Phys. Rev. B **33**, 8822 (1986); **34**, 7406 (1986).

[9] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[10] J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha, Phys. Rev. Lett. **82**, 5179 (1999).

[11] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996); **78**, 1396 (1997).

[12] J. P. Perdew and K. Burke, Int. J. Quantum Chem. **57**, 309 (1996).

[13] A. D. Becke, J. Comput. Chem. **20**, 63 (1999).

[14] H. L. Schmider and A. D. Becke, J. Chem. Phys. **109**, 8188 (1998).

[15] A. D. Becke, J. Chem. Phys. **107**, 8554 (1997).

[16] H. L. Schmider and A. D. Becke, J. Chem. Phys. **108**, 9624 (1998).

[17] F. A. Hamprecht, A. J. Cohen, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **109**, 6264 (1998).

[18] L. A. Curtiss, K. Raghavachari, G. W. Trucks, and J. A. Pople, J. Chem. Phys. **94**, 7221 (1991).

[19] T. Van Voorhis and G. E. Scuseria, J. Chem. Phys. **109**, 400 (1998).

[20] G. K.-L. Chan and N. C. Handy, Phys. Rev. A **59**, 3075 (1999).

[21] There is considerable disagreement as to whether exact conditions, or exact energies form a more universal data set. We note that while the extent of validity of a bound or inequality may be large, it is not so tight a condition as an exact energy. Thus, we feel that the two approaches should be considered as complementary.

[22] A. D. Becke, J. Chem. Phys. **104**, 1040 (1996).

[23] S. A. Vosko, V. Wilk, and W. Nusair, Can. J. Phys. **58**, 1200 (1980).

[24] H. Stoll, C. M. E. Pavlidou, and H. Preuss, Theor. Chim. Acta **49**, 143 (1978); H. Stoll, E. Golka, and H. Preuss, *ibid*. **55**, 29 (1980).

[25] A. D. Becke and M. R. Roussel, Phys. Rev. A **39**, 3761 (1989).

[26] R. Neumann, R. Nobles, and N. C. Handy, Mol. Phys. **87**, 1 (1996).

[27] M. Levy and J. P. Perdew, Phys. Rev. A **32**, 2010 (1985).

[28] A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).

[29] G. K-L. Chan, D. J. Tozer, and N. C. Handy, J. Chem. Phys. **107**, 1536 (1997).

[30] For a general overview of MP2 theory, see A. Szabo and N. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New York, 1989). In the high density coordinate scaling limit, MP2 theory becomes an exact model of correlation in some systems: see for example, A. Görling and M. Levy, Phys. Rev. B **47**, 13 105 (1993).

[31] G. E. Scuseria (private communication).

[32] C. W. Murray, N. C. Handy, and G. J. Laming, Mol. Phys. **78**, 997 (1993).

[33] V. I. Lebedev, Sibirsk. Mat. Zh. **18**, 132 (1975); V. I. Lebedev and A. L. Skorokhodov, Russ. Acad. Sci. Dokl. Math. **45**, 587 (1992).

[34] CADPAC: Cambridge Analytical Derivatives Package Issue 6.5 Cambridge, UK (1998) A suite of quantum chemistry programs developed by R. D. Amos with contributions from I. L. Alberts, J. S. Andrews, S. M. Colwell, N. C. Handy, D. Jayatilaka, P. J. Knowles, R. Kobayashi, N. Koga, K. E. Laidig, P. E. Maslen, C. W. Murray, J. E. Rice, J. Sanz, E. D. Simandiras, A. J. Stone, M.-D. Su, and D. J. Tozer.

[35] The ''fusion energy'' [see. Eq. (43)] has a $Z^{1/3}$ dependence, but we consider this to be sufficiently weak.

[36] A. J. Cohen and N. C. Handy, Chem. Phys. Lett. **316**, 160 (2000).

[37] G. E. Scuseria (private communication).

[38] R. D. Adamson, P. M. W. Gill, and J. A. Pople, Chem. Phys. Lett. **284**, 6 (1998).

[39] In this case, the equiprobability grid points in x_α and x_β were almost identical; thus, for simplicity we took the x_α grid points to also be the x_β grid points.

[40] T. H. Dunning, Jr., J. Chem. Phys. **55**, 716 (1971).

[41] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **100**, 2975 (1994).

[42] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980); T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. von R. Schleyer, J. Comput. Chem. **4**, 294 (1983); P. M. W. Gill, B. G. Johnson, J. A. Pople, and M. J. Frisch, Chem. Phys. Lett. **197**, 499 (1992); M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984).