Title

Linear Combination of Atomic Dipoles to Calculate the Bond and Molecular Dipole Moments of Molecules and Molecular Liquids

Journal

The Journal of Physical Chemistry Letters, 12(51)

ISSN

1948-7185

Authors

Chen, Kaixuan

Li, Wan-Lu

Head-Gordon, Teresa

Publication Date

2021-12-30

DOI

10.1021/acs.jpclett.1c03476

Linear Combination of Atomic Dipoles to Calculate the Bond and Molecular Dipole Moments of Molecules and Molecular Liquids

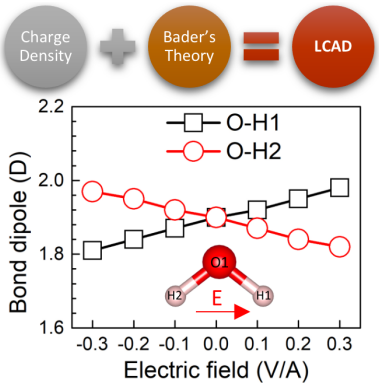
Kaixuan Chen[1−3], Wan-Lu Li[1−3], Teresa Head-Gordon[1−5*]

[1]*Chemical Sciences Division, Lawrence Berkeley National Laboratory,* [2]*Kenneth S. Pitzer Center for Theoretical Chemistry,* [3]*Department of Chemistry,* [4]*Department of Chemical and Biomolecular Engineering,* [5]*Department of Bioengineering, University of California, Berkeley, Berkeley, California 94720, USA*

E-mail: thg@berkeley.edu

Abstract

We report a a linear combination of atomic dipole (LCAD) method for calculating the bond dipole moments of molecules. We show that the LCAD method reproduces the known molecular dipole moments of small to large molecules with small error with respect to experiment and benchmark ab initio calculations, and molecular dipole distributions of bulk water that agree with maximally localized Wannier functions. The bond dipole moments derived from LCAD are also chemically interpretable in terms of trend in bond ionicity in going from neutral to charged molecules. Moreover, the LCAD method accurately captures the influence of electric fields, supported by the correct trend in the change of the dipole moment under a uniform external electric field. The better grounding of bond dipole calculations indicates that it should also serve as a useful approach to bond dipole-field models used in catalysis or to reconstruct the small dipole of a H-terminated graphene flake.



In the field of natural and synthetic (bio)catalysis, electric fields are sensitive to the interplay of local positioning at the active site and long-ranged interactions from the surrounding chemical environment.[1–3] In order to interpret the influence of electric fields on the catalysis process,[2,4,5] a bond-dipole-field model is often used[6]

$$\Delta G_{elec} = -(\overrightarrow{\mu}_2 \cdot \overrightarrow{E}_2 - \overrightarrow{\mu}_1 \cdot \overrightarrow{E}_1) \tag{1}$$

where $\overrightarrow{\mu}$ is the bond dipole of the breaking (or making) bond of interest and $\overrightarrow{E_i}$ is the electric field at two sites and/or two states such as the reactant and transition state.[6] In recent work our lab has used this model to propose site mutations that stabilize the transition state for a designed Kemp Eliminase KE15,[7] and how the mechanochemical coupling of the electrostatic environment with side chain rotamer transitions effect the turnover step for the natural enzyme ketosteroid isomerase.[8] We have also used the model to determine that an encapsulated water molecule within a supramolecular assembly $Ga_4L_6^{12-}$ in water generates electric fields that contributes most to the reduction in the activation free energy, overcoming the poor bulk water organization outside the nanocage.[3] In spite of these successful outcomes, Eq. (1) relies inherently on a definition of the bond dipole moments of a given state to provide accurate changes in free energy.[6] Although the electric field is readily available and fairly accurate when calculated from ab initio or many-body derivatives of the potential on a standard grid, a sensible and reliable definition of the bond dipole is necessary for making quantitative connections to (bio)catalysis outcomes.[9,10]

Molecular dipole moments are a direct measure of polarity of the charge density within a molecule; analogously, the bond dipole moment is an intrinsic characteristic of the polarity of the charge density within the bonding region. While the molecular dipole moment can be determined either from experimental dielectric spectroscopy[11] or from electronic structure theory[12] the bond dipole moment is a heuristic quantity, and is easily interpretable for small, neutral, and symmetric molecules, and early attempts have decomposed the molecular dipole moment into bond contributions for systems such as hydrocarbon molecules.[13] In fact there has been a long standing interest to predict the molecular dipole moment of large molecules or even their distributions in solutions

through group contribution methods.[14,15] For example, Burnelle et al.[16] successfully decomposed the total dipole moment into lone-pair moments and bond moments calculated from molecular wave functions of single H_2O and NH_3 molecules. However, Borst et al. tested the same assumption that the total dipole moment of a polyatomic molecule is the vector sum of bond dipole moments, and concluded that it works approximately in excited states but not for ground states.[17] Bader[18–21] proposed to decompose the molecular dipole as shown in Eq. 2,

$$\mu_{molecule} = \sum_{\Omega} [q(\Omega) X_{\Omega} + \mu(\Omega)] \tag{2}$$

where the first term is a charge transfer term, X_Ω is the position vector of the nucleus of atom Ω, and the bond dipole can be estimated by evaluating the charge transfer of each bond. The second term is an atomic dipole that denotes the polarization of the charge distribution that is centered around each individual atom that then completes the prediction for the molecular dipole moment. In addition, the atomic dipole has been used to impose a correction to the Hirshfeld charge in the ADCH charge partitioning scheme so that it can reproduce the correct molecular dipole.[22] The ADCH charge partitioning scheme has been adopted to calculate the bond dipole of simple AB_2- and AB_3-type polar molecules by manually counting the charge transfer for each bond[23,24] but has not been shown to be broadly extensible to large molecules or molecular liquids.

While these earlier attempts to decompose molecular dipole moments based on group contributions are instructive, it is not straightforward or obvious to define consistently for asymmetric molecules, charged molecules, or condensed phase systems. In this Letter we seek a straightforward and efficient way to calculate bond dipoles directly from the charge density and at the same time fulfill the assumption that the molecular dipole is the vector sum of bond dipoles. We propose a linear combination of atomic dipole moment (LCAD) to approximate bond dipoles. One advantage of the LCAD method is that the vector sum of all bond dipole moments matches the molecular dipole moment from experiment or ab initio calculations with careful choice of a reference position vector $\overrightarrow{r}_{ref}$ and a weighting scheme that estimates the charge transfer of the atomic

dipoles. We demonstrate that the method performs well for not only small and symmetric neutral molecules, but also asymmetric large molecules, charged species, bulk water dipole moment distributions, and bond dipole distortions under applied electric fields. The resulting LCAD method used to define bond dipoles are physically interpretable and intuitive, offer an alternative to maximally localized Wannier functions (MLWFs) in liquids, and have the potential to better quantify catalytic mechanisms through bond dipole-field analysis as we show for a catalytic gold complex.

Molecular dipole moments can be calculated from classical point charges, q_i

$$\overrightarrow{\mu}_{molecule} = \sum_i q_i(\overrightarrow{r_i} - \overrightarrow{r}_{ref}) \tag{3}$$

where $\overrightarrow{r}_i$ is the position vector with respect to the reference position vector $\overrightarrow{r}_{ref}$ for atom *i*, or using an integrated form based on the charge density $\rho(\overrightarrow{r})$

$$\overrightarrow{\mu}_{molecule} = \int \rho(\overrightarrow{r})(\overrightarrow{r} - \overrightarrow{r}_{ref})d\overrightarrow{r} \tag{4}$$

where $\rho(\overrightarrow{r})$ combines the contribution from both the atomic core and electrons. Correspondingly, calculating the bond dipole moment can be estimated by limiting the integral range to a region of the bond of interest

$$\overrightarrow{\mu}_{bond} = \int_{bond\,area} \rho(\overrightarrow{r})(\overrightarrow{r} - \overrightarrow{r}_{ref})d\overrightarrow{r} \tag{5}$$

For a neutral molecule, the choice of $\overrightarrow{r}_{ref}$ doesn't make any difference in the calculation of $\overrightarrow{\mu}_{molecule}$, since the influence of the reference vector will be cancelled in the vector sum. Of course this will not be true for either charged molecules nor for any fragmented molecular density such as $\overrightarrow{\mu}_{bond}$. Another disadvantage of using Eq. 5 to define $\overrightarrow{\mu}_{bond}$ is the possible multiple counting of charge density regions when summing over multiple $\overrightarrow{\mu}_{bond}$ to evaluate $\overrightarrow{\mu}_{molecule}$. For example, when we add up two O-H bond dipole moments of a water molecule, the charge density grid at and near the oxygen is counted twice, resulting in a large mean absolute deviation (MAD) in $\overrightarrow{\mu}_{molecule}$. This is not unique to a water molecule as shown in Table S1 when we approximate

$\overrightarrow{\mu}_{molecule} = \sum \overrightarrow{\mu}_b$ and compare it to the experimental molecular dipole moments of a large range of small molecules.

To address all of these issues for bond dipoles, we propose a methodology that avoids an overcount of the same charge region by using a linear combination of atomic dipole (LCAD) moments. We define atomic dipole moments $\overrightarrow{\mu}_{atom}$ for each individual atom *i* using a localization scheme (Eq. 6),

$$\overrightarrow{\mu}_{atom} = \int_{atomic\,volume} \rho(\overrightarrow{r})(\overrightarrow{r} - \overrightarrow{r}_{ref})d\overrightarrow{r} \tag{6}$$

based on Bader's theory,[18,25,26] and the subsequently formed atomic dipoles, $\overrightarrow{\mu}_A$ and $\overrightarrow{\mu}_B$, are weighted to formulate the bond dipole $\overrightarrow{\mu}_{A-B}$ as given in (Eq. 7). Note that the atomic dipole defined in this article has a dependence on reference point, different from some previous definitions which generally adopts the nuclear coordinates.[19,22]

$$\overrightarrow{\mu}_{A-B} = x_A \overrightarrow{\mu}_A + x_B \overrightarrow{\mu}_B \tag{7}$$

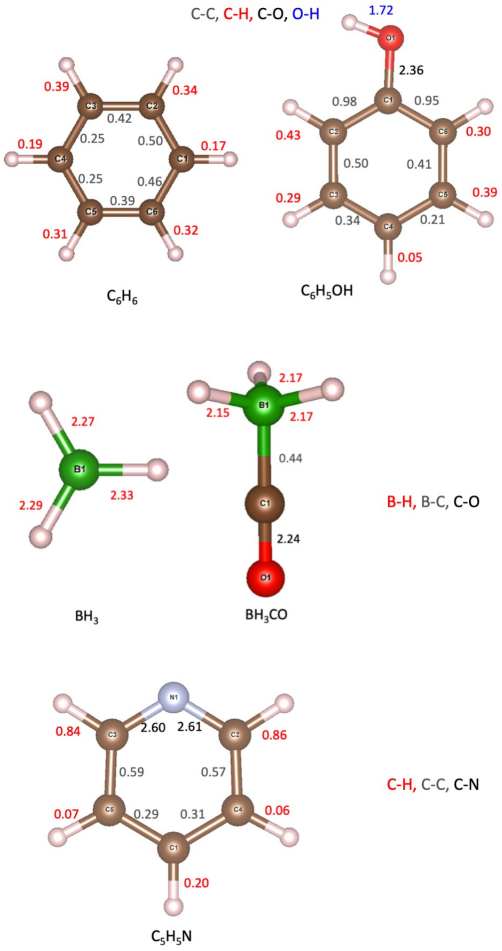
The ratio x_A denotes the portion of atomic dipole of atom A that is assigned to the bond A-B, when atom A forms chemical bonds with X atom neighbors, as expressed in Eq. 8. This is based on the physical idea that the distribution of the portion of $\overrightarrow{\mu}_A$ to the connecting bonds is *inversely* proportional to the square of bond length *l*, inspired by the relationship between charge transfer and distance.

$$x_A = \frac{1/l_{A-B}^2}{\sum_X 1/l_{A-X}^2} \tag{8}$$

One advantage for the LCAD method is that the vector sum of all bond dipole moments matches the molecular dipole moment, as long as we choose the same reference position vector $\overrightarrow{r}_{ref}$ when calculating the atomic dipoles. We test the LCAD approach in the following section on reproducing the molecular dipole moments from experiment or *ab initio* calculations under different scenarios.

The bond dipole derived from the LCAD method relies on the choice of the ratio $x_{A/B}$ as described in Eq. 8 as well as the vector origin $\overrightarrow{r}_{ref}$. In this work, the center of mass of the molecule (CMM) is adopted as $\overrightarrow{r}_{ref}$.[27] Table 1 and the Figure insert show that the experimental

Table 1: The calculated bond dipole moments using the LCAD method with the strategy of distributing atomic dipole moments inversely proportional to the square of the bond length. The charge density is derived from the PBE functional with the DZVP basis set. The corresponding result of large molecules are shown in the Figure. The unit is Debye.



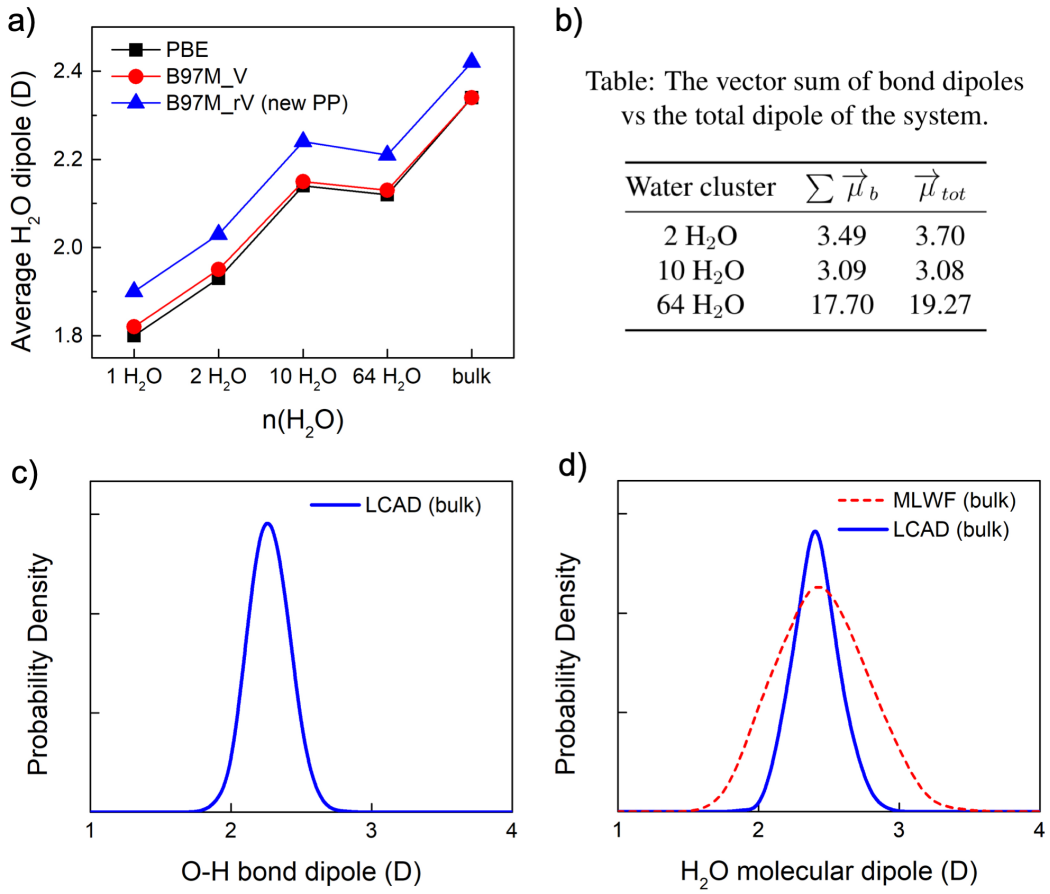
Molecule	Bond	$\overrightarrow{\mu}_b$	$\sum \overrightarrow{\mu}_b$	$\overrightarrow{\mu}_m$(ref)[b]
	C-H[a]	0.27		
CH_3OH	C-O	1.14	1.59	1.70
	O-H	1.71		
CH_4	C-H	0.13	0.00	0.00
H_2CO	C-O	3.06	2.25	2.33
	O-H	0.41		
H_2O	O-H	1.83	1.84	1.85
H_2O_2	O-H	1.78	1.64	1.77
HCN	C-H	0.19	2.99	2.98
	C-N	2.85		
HNC	N-H	0.92	2.87	3.05
	C-N	1.97		
HNO	O-N	0.72	1.67	1.67
	N-H	1.06		
NF_3	N-F	1.72	0.11	0.23
NH_3	N-H	1.25	1.70	1.47
OF_2	O-F	0.98	0.25	0.30
C_6H_6			0.00	0.00
C_6H_5OH			1.27	1.22
BH_3			0.00	0.00
BH_3CO			1.03	1.70
C_5H_5N			2.31	2.19
MAD			0.10	

[a]The average of the calculated dipole moments for all same-type bonds within a molecule is used. For example, the three C-H bonds within a CH_3OH molecule. [b]The experimental data is obtained from the Computational Chemistry Comparison and Benchmark DataBase.[28]

molecular dipole moment is well reproduced using the LCAD approach, and the vector sum of bond dipoles ($\sum \overrightarrow{\mu}_b$) is close to the experimental molecular dipole for each molecule with an overall mean average deviation (MAD) of 0.10 D. By contrast the use of point charges or charge density as given in Eq. 5 give much larger error as seen in Table S1, due to the aforementioned overcounting problem. The one outlier in comparison to the experimental value is BH_3CO, which we find is due to the choice of *ab initio* density using the PBE functional with the DZVP basis

set, since it is known that the molecular properties such as dipoles are sensitive to the quality of the underlying electron density when comparing to experiment.[12] We note that the magnitude of the bond dipole moment of the same type of bond remains clustered in predictable ranges within different molecules: $\overrightarrow{\mu}_{C-H}$ (0.1-0.7 D), $\overrightarrow{\mu}_{N-H}$ (0.3-1.3 D), $\overrightarrow{\mu}_{O-H}$ (1.6-1.9 D), $\overrightarrow{\mu}_{C-N}$ (1.9-2.8 D), and $\overrightarrow{\mu}_{C-O}$ (1.0-3.1 D). In addition, the bond dipole increases qualitatively by a factor proportional to the bond ionicity, for instance C-H $<$ N-H $<$ O-H, and C-N $<$ C-O, and is consistent with chemical intuition that larger charge transfer occurs for bonds with more ionic character, and thus leading to to a larger dipole moment.

Next we turn to a more complex system, namely water clusters consisting of 1 to 64 water molecules as shown in Figure 1. The calculations were performed using either PBE or the B97M-rV density functional with the TZVP basis set, the latter DFT functional is one which we found to be an excellent description of bulk water and correspondingly the electron density of water in the condensed phase.[29,30] To use the LCAD method, we divide the whole system into subsystems, each consisting of one individual water molecule and the bond dipoles are calculated as per the small molecule case. Figure 1A shows that the average water molecular dipole moment increases gradually as the size of cluster increases, which is consistent with previous theoretical findings,[31] and the small but perhaps not surprising effect in the choice of pseudopotential (PP), in which we recently reported an optimized set of "small core" PPs for main group and transition metal atoms for the B97M-rV functional.[32] In addition, the vector sum of all bond dipoles within the water cluster is a good approximation to the total calculated dipole moment of the cluster from ab initio as shown in Figure 1B, which supports the rationale of the LCAD method. For bulk water, we demonstrate the statistical dipole distribution in Figures 1C and D, examined from 10 snapshots of AIMD simulation at an NVT ensemble (300 K), where the average dipole for each snapshot is shown in Figure S1. We find that both the O-H bond dipole moment and water molecular dipole are Gaussian distributed, which agrees nicely with the result using the MLWF method[31,33] at the same level of DFT theory, although the MLWF result exhibits a wider distribution range.[30]



b)

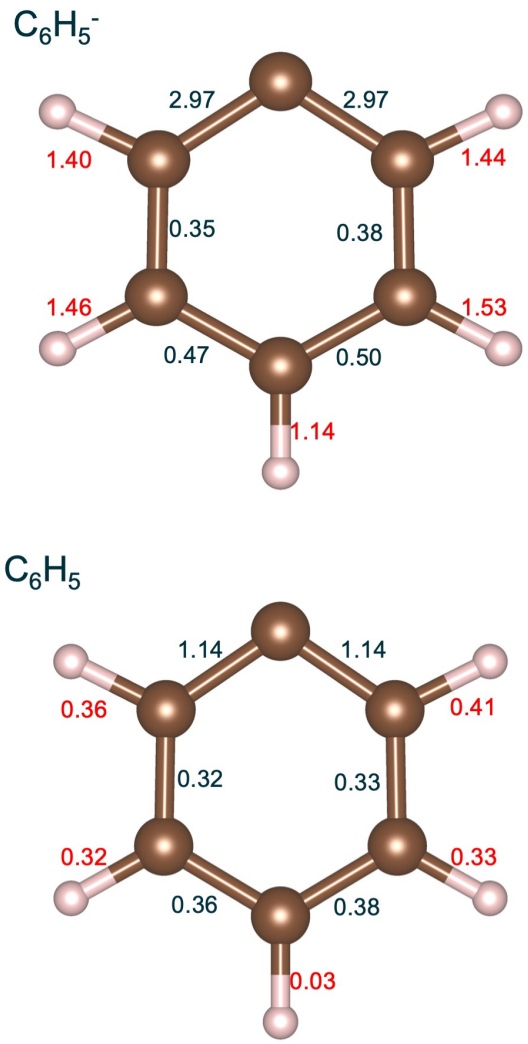
Table: The vector sum of bond dipoles vs the total dipole of the system.

Water cluster	$\sum \vec{\mu}_b$	$\vec{\mu}_{tot}$
2 H_2O	3.49	3.70
10 H_2O	3.09	3.08
64 H_2O	17.70	19.27

Figure 1: The calculated molecular dipole moment of water clusters from bond dipoles using LCAD. (A) Comparison of LCAD calculations using different density functionals: PBE,[34] meta-GGA B97M-rV[35] with the PBE pseudopotential and newly optimized pseudopotential.[32] (B) The vector sum of bond dipoles vs the calculated total dipole moment of the system. The statistical distribution of (C) the O-H bond dipole and (D) H_2O molecular dipole of bulk water is shown, comparing LCAD and MLWF,[33] examined from 10 AIMD snapshots after the system reaches thermal and energetic balance at 1 ps. The dipole unit is Debye.

For charged systems, the choice of reference center can give rise to uncertainity in the predicted bond dipoles and their sum to reproduce the molecular dipole moment, and these systems provide a further test as to the validity of CMM as $\vec{r}_{ref}$. Table 2 compares the data produced using the LCAD and MLWF method using B97M-rV functional with TZV2P basis set in CP2k, as bench-

Table 2: The calculated dipole moments for typical ions, molecules and radicals (B97M-rV functional with TZV2P basis). The dipole unit is Debye.



Ions	$\overrightarrow{\mu}_b{}^a$	$\sum \overrightarrow{\mu}_b$	$\overrightarrow{\mu}_m$(MLWF)	$\overrightarrow{\mu}_m$(ref)^b
H_2O^+	2.59	2.10	2.38	2.37
H_2O	1.89	1.81	1.83	1.90
H_2O^-	1.07	1.53	1.76	1.41
OH^+	2.03	2.03	2.29	2.28
OH	1.62	1.62	1.63	1.68
OH^-	1.19	1.19	1.81	1.09
NH_3^+	2.53	0.54	0.01	0.00
NH_3	1.30	1.48	1.82	1.57
NH_3^-	0.96	0.60	0.62	0.67
NH^+	1.76	1.76	1.97	1.99
NH	1.50	1.50	1.50	1.54
NH^-	1.29	1.29	0.90	0.99
CH_4^+	1.46	0.54	0.04	0.00
CH_4	0.28	0.00	0.00	0.00
CH_4^-	2.23	0.54	0.00	0.00
CH_3^+	1.15	0.54	0.01	0.00
CH_3	0.15	0.00	0.01	0.00
CH_3^-	1.41	1.50	1.28	1.31
CH_2^+	1.63	0.57	0.69	0.68
CH_2	0.78	1.55	1.61	1.60
CH_2^-	1.50	1.81	1.50	1.50
CH^+	2.15	2.15	1.81	1.73
CH	1.46	1.46	1.47	1.44
CH^-	0.93	0.93	1.20	1.17
$C_6H_5^-$		5.12	4.8	4.64
C_6H_5		0.86	0.84	0.85
MAD		0.22	0.15	

[a]The average of the calculated dipole moments for all same-type bonds within an ion (or a radical/molecule) is used. [b]Reference data is from ab initio calculations using Q-Chem.[36]

marked against the same DFT functional and basis set using the Q-Chem code as its definition of reference center is close to the CMM we have defined for LCAD.[36] As seen from Table 2, there is a relatively small MAD of around 0.22 D for the LCAD and 0.15 D for MLWF in estimating the molecular dipole moment, showing the equivalence of the two methods.

When observing the trends in bond dipoles, there is a decrease in bond dipole from cation to neutral to anion, with the exception of the alkane series CH_2, CH_3 and CH_4. Unlike the CH

fragment alone, as more and more C-H bonds form on C, it leads to smaller charge transfer between C and H, such that the bond dipole decreases as n increases in CH_n. Therefore, for CH_n (n = 2, 3, 4), either the removal or the addition of an electron will increase the charge transfer between C and H, thereby resulting in an increase in C-H bond dipole in either a cationic CH_n^+ or an anionic CH_n^- form. In the insert figure of Table 2, we can see an increase of C-H bond dipoles as well in a relatively more complex systems from C_6H_5 to $C_6H_5^-$.

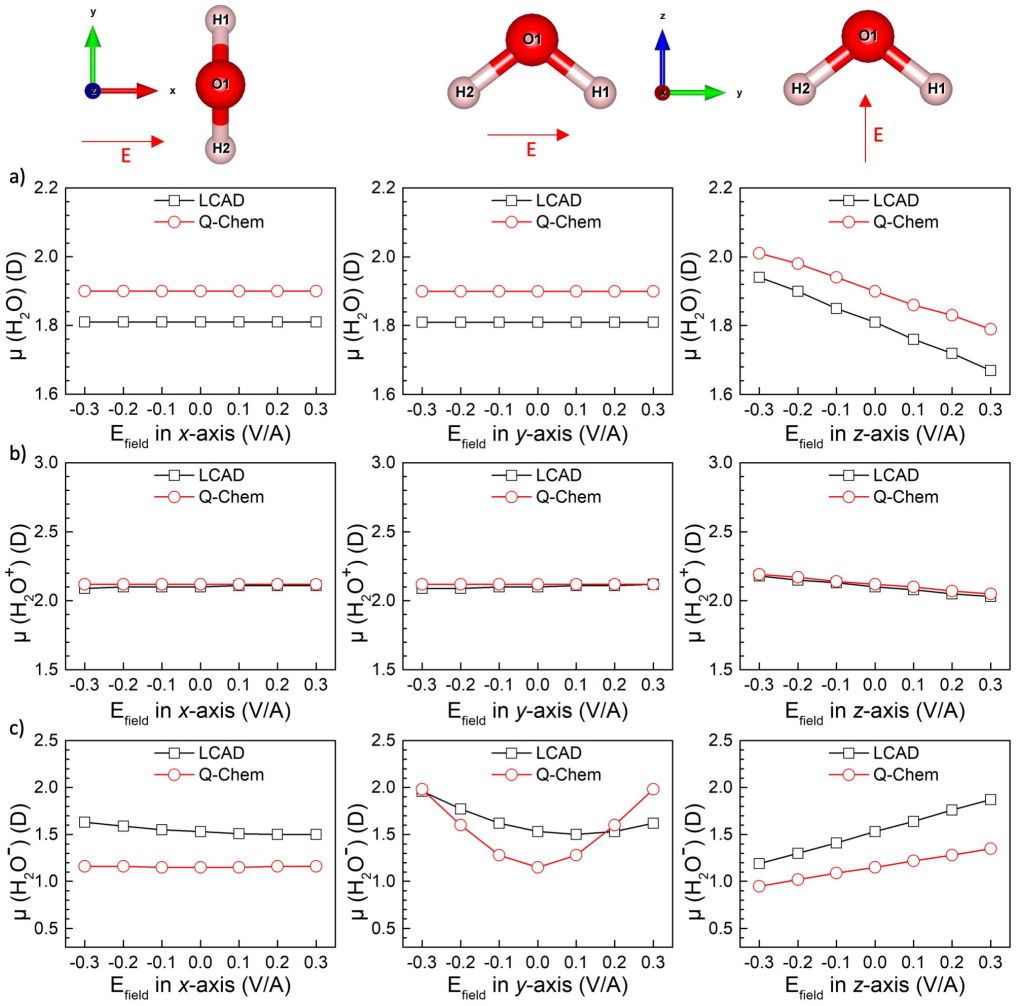


Figure 2: The molecular dipole moment in a) H_2O molecule, b) H_2O^+ cation and c) H_2O^- anion calculated using LCAD method and Q-Chem under different uniform external electric fields.

Given the interest in the bond dipole-field model, we also examine whether the LCAD can accurately capture the response of charge density under the perturbation of an electric field. In the case of H_2O and H_2O^+ (Figure 2A and 2B), the total molecular dipole moment points roughly

in the *negative* z direction, such that the electric field in either x or y direction hardly influences the magnitude of the molecular dipole, while the z-axis carries all the electric field response and changes the dipole value with trends in excellent agreement with the Q-Chem benchmark. For H_2O^-, however, the extra electron disperses such that total dipole moment points in the opposite y-z direction relative to the neutral and cationic water species, and leading to a more complicated molecular dipole response as seen in Figure 2C), a trend that agrees well with the Q-Chem reference. Similar plots for CH_4, CH_4^- and CH_4^+ are shown in Figure S2 in which the molecular dipole response to the electric fields using LCAD method is in good accord with the Q-Chem results.

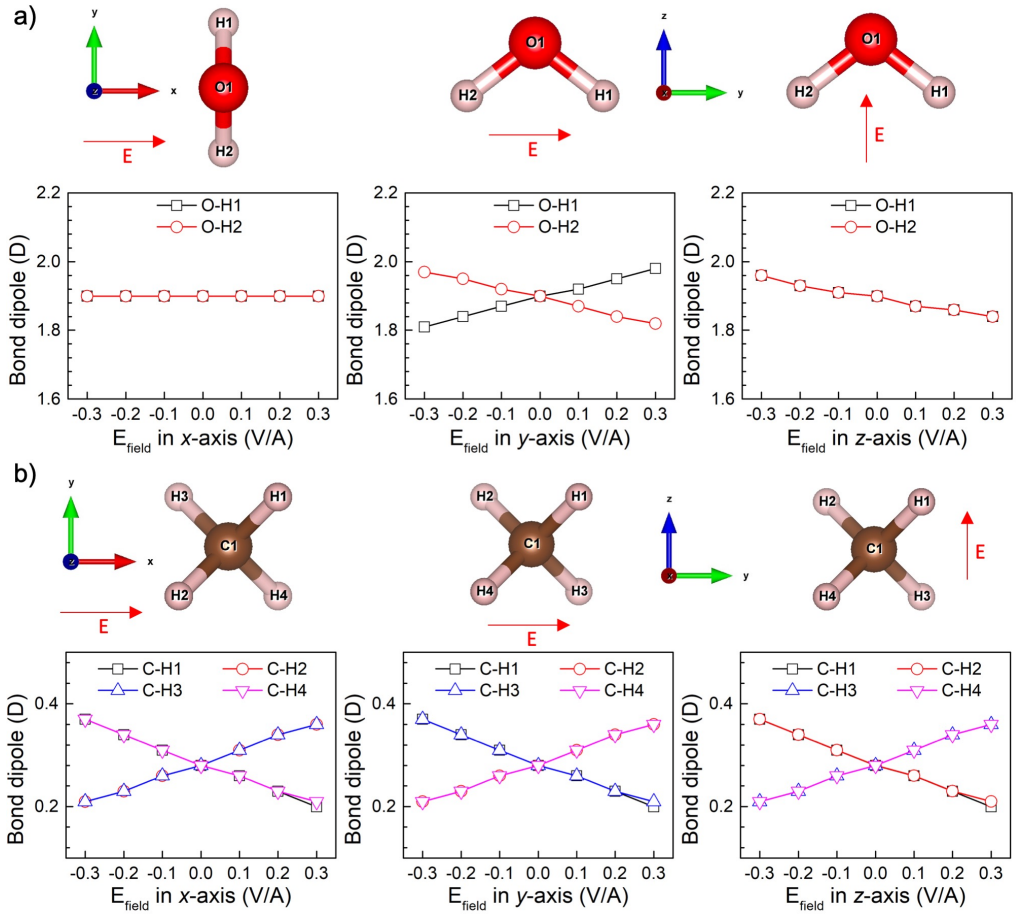


Figure 3: The calculated bond dipoles of a) two O-H bonds within a H_2O molecule and b) four C-H bonds within a CH_4 molecule, respectively, in the presence of uniform external electric fields.

Figure 3 shows the analogous influence of electric fields on the bond dipole moments for the O-H bonds in a H_2O molecule and C-H bonds for the CH_4 molecule. For water the two O-H bonds

possess the same magnitude of the bond dipole moment but which differ in direction from the O atom to H atom. When an electric field is applied which is orthogonal to both O-H bonds, no significant change is observed, whereas under an applied electric field in which the direction senses asymmetry, the two bonds have either completely opposite change but with same magnitude, or both bonds respond identically, demonstrating a deceasing dipole moment as electric field increases (Figure 3A). Another example is given in Figure 3B to demonstrate how the structurally equivalent C-H bond dipoles of the CH_4 molecule respond asymmetrically to directional electric fields that can weaken or strengthen bonds depending on external field direction.

Next we consider an application of LCAD to both neutral and charged gold complexes that when encapsulated in the supramolecular capsule $Ga_4L_6^{12-}$,[37] have been shown to catalyze the carbon-carbon reductive elimination reaction.[6] This provides an example for which the bond dipoles would be utilized in the bond dipole-field model using Eq. 1 to estimate reactant state destabilization or transition state stabilization that could help explain the catalytic rate.[3,38] Here we only consider the reactant ground state and its charge based on whether a halogen ligand is or is not present (denoted as RS or RS^+, respectively), and how well the molecular dipole moments agree with MLWF and *ab initio* benchmarks. Table 3 shows that the molecular dipole moments calculated for RS and RS+ using the LCAD method is in close agreement in magnitude and direction to that determined from the MLWF method, both of which utilize the B97M-rV functional,[35] recently optimized pseudopotentials,[32] and the TZV2P basis set, all performed in the CP2k code.[39]

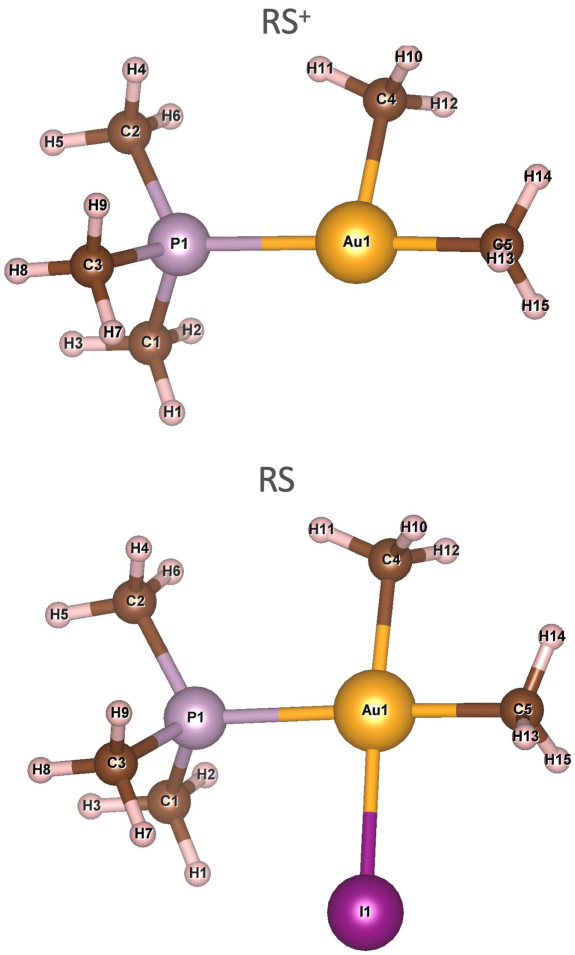
Table 3: The calculated molecular dipole for charged (RS^+) and neutral (RS) state of gold complex. Both dipole vector (x, y, z) and dipole norm are shown in the units of Debye. In all cases we use the B97M-rV functional, but the underlying electron density depended on basis set with TZV2P basis set in CP2K.

Method	RS^+				RS			
	x	y	z	Norm	x	y	z	Norm
LCAD	-2.90	-2.16	0.31	3.63	-7.23	-1.22	0.00	7.34
MLWF	-3.20	-2.50	0.00	4.50	-7.41	-1.13	0.20	7.49
def2-TZVPP (all-electrons)	-1.79	-2.50	0.00	3.07	0.60	-5.36	0.00	5.40
def2-TZVPPD (all-electrons)	-1.51	-2.53	0.00	2.95	-1.00	-4.60	0.00	4.71
CRENBL PP + def2-TZVPP	-0.20	-4.25	0.00	4.26	-6.19	-1.27	0.00	6.31

We also predict the molecular dipole moments of the neutral and charged gold complexes using Q-Chem with all-electron vs. pseudopotential (PP) for the underlying electron density, or adopting different basis sets (def2-TZVPP and def2-TZVPPD), given known evidence of the sensitivity of molecular dipole moments to the electron density representation.[12] For RS^+, the LCAD and MLWF result is close to the all-electrons calculation with the same basis level (def2-TZVPP), while for RS, the LCAD and MLWF results seem to agree better with the Q-Chem result using the CRENBL pseudopotential.

Table 4: The calculated bond dipoles for charged (RS^+) and neutral (RS) state of gold complex using the LCAD method. The dipole moment unit is Debye.

Bond	Atom 1	Atom 2	$\overrightarrow{\mu}_b$ (RS^+)	$\overrightarrow{\mu}_b$ (RS)
Au-P	Au	P	0.13	0.20
Au-C	Au	C4	0.83	0.71
	Au	C5	0.79	0.71
P-C	P	C1	0.96	1.00
	P	C2	0.95	0.96
	P	C3	0.90	0.94
C-H	C1	H1	0.70	0.93
	C1	H2	0.68	0.90
	C1	H3	0.60	0.98
	C2	H4	0.66	0.53
	C2	H5	0.59	0.38
	C2	H6	0.67	0.53
	C3	H7	0.51	0.63
	C3	H8	0.48	1.07
	C3	H9	0.48	0.63
	C4	H10	0.88	0.52
	C4	H11	0.27	0.92
	C4	H12	0.88	0.51
	C5	H13	0.15	0.53
	C5	H14	0.17	0.68
	C5	H15	0.14	0.47
Au-I	Au	I		5.06



Having been validated against the molecular dipole moment of the complex, we next consider the bond dipoles and how they change in the presence of charge in (Table 4). The introduction

of iodine for the neutral RS leads to a highly polar Au-I bond of 5.06 D. This neutral RS is the precursor state in the catalytic mechanism, but the catalytically active form is the RS^+ charged state with iodine removed. It is notable that the Au-C bond dipoles elongate by 20% and the C-H bonds of the methyl groups also noticeably elongate on average in the catalytic state RS^+. With good electric field alignment, these changes could destabilize the reactant state.

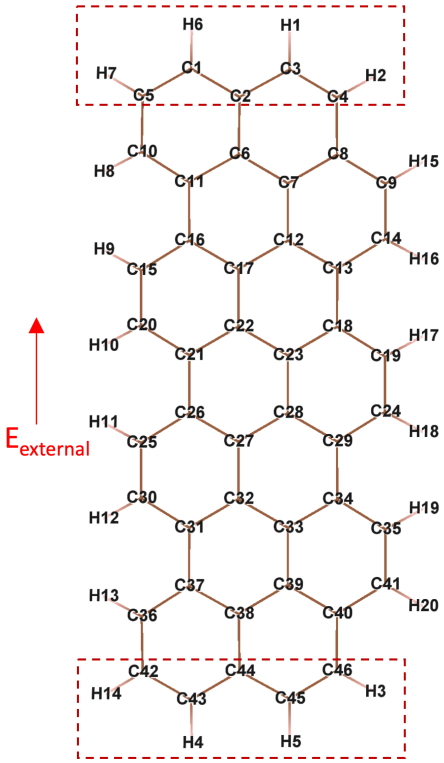


Figure 4: The molecular structure of a graphene flake $C_{46}H_{20}$. The dotted boxes denote outermost edges where the C-H bond dipoles are significantly affected by an external electric field (E_y = 2.06 GV/m).

Finally, we consider a graphene flake $C_{46}H_{20}$ [40] under external electric field, as shown in Figure 4. The molecular dipole moments calculated using B97M-rV functional changes from 0.03 D to 13.36 D when an electric field (E_y = 2.06 GV/m) is applied along the y-axis, which resembles the result using B3LYP hybrid functional (0.00 D and 12.16 D, respectively) in previous reported literature.[40] As regards bond dipoles, the electric field has a small influence on the dipoles of C-C bonds, but a rather significant effect on C-H bonds, especially those at the edge of the graphene flake, as shown in Figure 4 and Table S2. The LCAD method well captures the increase in dipoles

of the outermost C-H bonds along the electric field direction, an example to demonstrate its validity in organic systems.

In conclusion, the LCAD method can accurately calculate molecular dipole moments from localized atomic densities that can be formulated in terms of physically motivated bond dipole moments, starting with the charge density matrix that is produced in a typical first-principle calculation and electron partition information from Bader's theory. For the evaluation of the bond dipole moments, it is suggested to choose the center of mass of the molecule as the reference point, so that the vector sum of bond dipole moment reproduce the molecular dipole moment for not only neutral molecules, but charged molecules or molecules in the presence of directional electric fields. For larger condensed phase systems, one should divide each molecule as a subsystem and use the center of mass for each molecule as the reference point independently.

We note that the orientation of bond dipoles calculated by Eq. 7 may deviate from the chemical bond direction, which arises from the inclusion of the polarization term in Eq. 2. In other words, the chemical environment of a certain bond has an influence on both its dipole norm and direction. While the length criteria in Eq. 8 can correctly capture the ratio as regards single and double bonds, it may show some weakness when dealing with complicated bond multiplicity, such as bonds with low charge transfer. Further improvement may be taken into consideration in the future, such as the electronegativity difference of the element (the bond ionicity). Nonetheless, using the LCAD method, we showed that we can predict the dipole distribution in water clusters with accuracy comparable to maximally localized Wannier functions, and offers a better foundational approach to using bond dipole-field models in catalysis applications, as illustrated here of how bond and molecular dipole moments change in the presence of external electric fields, and providing valuable insights into investigation of chemical bond rearrangements for reactive systems as shown for the gold complex that plays the role of the active site within a supramolecular cage.[6]

Density Functional Theory. The DFT calculations were performed with the Perdew–Burke–Ernzerhof (PBE)[41] and meta-GGA B97M-rV[35] functional using CP2K software.[39,42] The geometric param-

eters of simulated molecules are obtained from the Computational Chemistry Comparison and Benchmark DataBase,[28] after which the coordinates for each molecule were re-organized so that the center of mass of the new coordinates being the zero point. And then we placed them into a cubic box of size L = 12 Å. The charge density distribution is obtained from self-consistency calculation with the energy converged to at most 10^{-6} hartree. For the sake of comparison, we have evaluated the molecular dipole moment results using Q-Chem[36] at the same theory level with the built-in B97M-rV functional and def2-TZVPP basis.

Especially for Figure 1, 1 H_2O and 2 H_2O (10 H_2O and 64 H_2O) molecules are placed into a cubic box of size L = 12.42 Å (25 Å) to avoid periodic interaction. For bulk water, we adopted the model of 64 water molecules in a cubic box of size L = 12.42 Å, corresponding to the experimental density of bulk water.[30] Structural relaxations were performed using B97M-rV functional[35] for these systems with all atomic forces smaller than 10^{-3} hartree/bohr. In addition, ab initio molecular dynamics (AIMD) simulation was performed for bulk water in an NVT ensemble at 300 K, with a step size of 0.5 fs. The system reached temperature and energetic balance at 1 ps, and 10 snapshots were collected each 0.05 ps ever after. The statistic dipole distribution of these 10 snapshots was used in Figure 1C and D. Note that in order to get the MLWF result for bulk water, a larger cubic box of size L = 25 Å has been used to store the cooridnates of 64 water molecules from AIMD snapshots.

Maximally localized Wannier functions (MLWF). MLWF's are analogs of localized orbitals generated from quantum chemical calculations of small molecules, but relevant to the condensed phase where electron density divisions between molecules are more ambiguous without a localization scheme.[43] A unique transformation U_{kl} of the Kohn-Sham orbital Φ_l which can be obtained by maximally diagonalizing a set of matrices with a certain criterion produced the orbitals called MLWF as shown in Eq. 9.

$$w_k(\overrightarrow{r}) = \sum_l U_{kl}\Phi_l(\overrightarrow{r}) \tag{9}$$

And the expectation value of the position operator $\overrightarrow{r}$ of such MLWF is defined as:

$$\overrightarrow{r}_k = -\frac{L}{2\pi} Im \ln z_k \tag{10}$$

which is also called as a Wannier function center (WFC). Here in, the complex number $z_{\alpha,k}$ is expressed as:

$$z_{\alpha,k} = \langle w_k | e^{-i\overrightarrow{G}_\alpha \cdot \overrightarrow{r}} | w_k \rangle \tag{11}$$

where $\overrightarrow{G}_\alpha$ is a reciprocal lattice vector depending on the supercell shapes.[44,45] Based on this concept, the molecular dipole moment can be determined from Eq. 12 over all WFC points *k* in the molecular region *I*.[46]

$$\mu_I^{el} = -2e \sum_{k \in I} \overrightarrow{r}_k \tag{12}$$

Availability of the LCAD code *cubedipole.x*. This code adopts the charge density data from DFT calculation using the CP2K software,[39,42] combined with the density partitioning method from Bader's theory.[25,26] The current version of the code can perform calculations of both the molecular dipole moment (Eq. 4) and bond dipoles of a molecules using the LCAD method proposed in the main text (Eq. 6-7). It is applicable to neutral and charged systems and with the presence of a uniform electric field.

Supporting Information Available

Supporting information: water molecular dipole for all AIMD snapshots, molecular dipole change of CH_4, CH_4^+ and $CH_4^-$with respect to external electric field, the bond dipoles in graphene flake $C_{46}H_{20}$.

Acknowledgement

The methodological work received support by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) program, with catalytic applications supported by the CPIMS program by the Director, Office of Science, Office of Basic Energy Sciences, Chemical Sciences Division of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This work used computational resources provided by the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

References

(1) Bhowmick, A.; Sharma, S. C.; Head-Gordon, T. The Importance of the Scaffold for de Novo Enzymes: A Case Study with Kemp Eliminase. *J. Am. Chem. Soc.* **2017**, *139*, 5793–5800.

(2) Welborn, V. V.; Ruiz Pestana, L.; Head-Gordon, T. Computational optimization of electric fields for better catalysis design. *Nat. Catal.* **2018**, *1*, 649–655.

(3) Welborn, V. V.; Li, W. L.; Head-Gordon, T. Interplay of water and a supramolecular capsule for catalysis of reductive elimination reaction from gold. *Nature Communications* **2020**, *11*, 1–6.

(4) Warshel, A. Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J. Biol. Chem.* **1998**, *273*, 27035–27038.

(5) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **2006**, *106*, 3210–3235.

(6) Li, W. L.; Head-Gordon, T. Catalytic Principles from Natural Enzymes and Translational Design Strategies for Synthetic Catalysts. *ACS Cent. Sci.* **2021**, *7*, 72–80.

(7) Vaissier, V.; Sharma, S. C.; Schaettle, K.; Zhang, T.; Head-Gordon, T. Computational Optimization of Electric Fields for Improving Catalysis of a Designed Kemp Eliminase. *ACS Catalysis* **2018**, *8*, 219–227.

(8) Welborn, V. V.; Head-Gordon, T. Fluctuations of Electric Fields in the Active Site of the Enzyme Ketosteroid Isomerase. *Journal of the American Chemical Society* **2019**, *141*, 12487–12492, PMID: 31368302.

(9) Wang, L.; Fried, S. D.; Boxer, S. G.; Markland, T. E. Quantum delocalization of protons in the hydrogen-bond network of an enzyme active site. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 18454–18459.

(10) Fried, S. D.; Boxer, S. G. Measuring electric fields and noncovalent interactions using the vibrational stark effect. *Acc. Chem. Res.* **2015**, *48*, 998–1006.

(11) Kortschot, R. J.; Bakelaar, I. A.; Erné, B. H.; Kuipers, B. W. M. A differential dielectric spectroscopy setup to measure the electric dipole moment and net charge of colloidal quantum dots. *Rev. Sci. Instrum.* **2014**, *85*, 033903.

(12) Hait, D.; Head-Gordon, M. How Accurate Is Density Functional Theory at Predicting Dipole Moments? An Assessment Using a New Database of 200 Benchmark Values. *J. Chem. Theory Comput.* **2018**, *14*, 1969–1981.

(13) Syrkin, Y. K.; Dyatkina, M. E. *Structure of Molecules and the Chemical Bond (English Translation)*; Dover Publications: New York, 1964.

(14) Müller, K.; Mokrushina, L.; Arlt, W. Second-order group contribution method for the determination of the dipole moment. *J. Chem. Eng. Data* **2012**, *57*, 1231–1236.

(15) Cho, M.; Sylvetsky, N.; Eshafi, S.; Santra, G.; Efremenko, I.; Martin, J. M. The Atomic Partial Charges Arboretum: Trying to See the Forest for the Trees. *ChemPhysChem* **2020**, *21*, 688–696.

(16) Burnelle, L.; Coulson, C. A. Bond dipole moments in water and ammonia. *J. Chem. Soc. Faraday Trans.* **1957**, *53*, 403–405.

(17) Borst, D. R.; Korter, T. M.; Pratt, D. W. On the additivity of bond dipole moments. Stark effect studies of the rotationally resolved electronic spectra of aniline, benzonitrile, and aminobenzonitrile. *Chem. Phys. Lett.* **2001**, *350*, 485–490.

(18) Bader, R. F. W.; Beddall, P. M.; Cade, P. E. Partitioning and characterization of molecular charge distributions. *J. Am. Chem. Soc.* **1971**, *93*, 3095–3107.

(19) Bader, R. F. W.; Larouche, A.; Gatti, C.; Carroll, M. T.; MacDougall, P. J.; Wiberg, K. B. Properties of atoms in molecules: Dipole moments and transferability of properties. *J. Chem. Phys.* **1987**, *87*, 1142–1152.

(20) Bader, R. F.; Matta, C. F. Properties of atoms in crystals: Dielectric polarization. *Int. J. Quantum Chem.* **2001**, *85*, 592–607.

(21) Bader, R. F. Dielectric polarization: A problem in the physics of an open system. *Mol. Phys.* **2002**, *100*, 3333–3344.

(22) Lu, T.; Chen, F. Atomic dipole moment corrected Hirshfeld population method. *J. Theor. Comput. Chem.* **2012**, *11*, 163–183.

(23) Wei, M. J.; Jia, D. Q.; Chen, F. W. Geometric structures, excitation energies and dipole moments of the ground and excited states of TiO_2. *Acta Phys. -Chim. Sin.* **2013**, *29*, 1441–1452.

(24) Cao, J. S.; Wei, M. J.; Chen, F. W. Relationship between the bond dipole moment and bond angle of polar molecules. *Acta Phys. -Chim. Sin.* **2016**, *32*, 1639–1648.

(25) Henkelman, G.; Arnaldsson, A.; Jónsson, H. A fast and robust algorithm for Bader decomposition of charge density. *Comput. Mater. Sci.* **2006**, *36*, 354–360.

(26) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, 1990.

(27) Cramer, C. J. *Essentials of computational chemistry*; Wiley, 2004.

(28) Johnson III, R. D., Ed. *NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101 Release 21*; 2020.

(29) Ruiz Pestana, L.; Mardirossian, N.; Head-Gordon, M.; Head-Gordon, T. Ab initio molecular dynamics simulations of liquid water using high quality meta-GGA functionals. *Chem. Sci.* **2017**, *8*, 3554–3565.

(30) Ruiz Pestana, L.; Marsalek, O.; Markland, T. E.; Head-Gordon, T. The Quest for Accurate Liquid Water Properties from First Principles. *J. Phys. Chem. Lett.* **2018**, *9*, 5009–5016.

(31) Zhu, T.; Van Voorhis, T. Understanding the Dipole Moment of Liquid Water from a Self-Attractive Hartree Decomposition. *J. Phys. Chem. Lett.* **2021**, *12*, 6–12.

(32) Li, W.-l.; Chen, K.; Rossomme, E.; Head-Gordon, M.; Head-Gordon, T. Optimized Pseudopotentials and Basis Sets for Semiempirical Density Functional Theory for Electrocatalysis Applications. *J. Phys. Chem. Lett.* **2021**, *12*, 10304–10309.

(33) Silvestrelli, P. L.; Parrinello, M. Water molecule dipole in the gas and in the liquid phase. *Phys. Rev. Lett.* **1999**, *82*, 3308–3311.

(34) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(35) Mardirossian, N.; Ruiz Pestana, L.; Womack, J. C.; Skylaris, C.-K.; Head-Gordon, T.; Head-Gordon, M. Use of the rVV10 Nonlocal Correlation Functional in the B97M-V Density Functional: Defining B97M-rV and Related Functionals. *J. Phys. Chem. Lett.* **2017**, *8*, 35–40.

(36) Epifanovsky, E. et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*, 084801.

(37) Morimoto, M.; Bierschenk, S. M.; Xia, K. T.; Bergman, R. G.; Raymond, K. N.; Toste, F. D. Advances in supramolecular host-mediated reactivity. *Nat. Catal.* **2020**, *3*, 969–984.

(38) Vaissier Welborn, V.; Head-Gordon, T. Electrostatics Generated by a Supramolecular Capsule Stabilizes the Transition State for Carbon-Carbon Reductive Elimination from Gold(III) Complex. *J. Phys. Chem. Lett.* **2018**, *9*, 3814–3818.

(39) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. CP2K: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 15–25.

(40) Huang, L.; Massa, L.; Matta, C. F. A graphene flake under external electric fields reconstructed from field-perturbed kernels. *Carbon* **2014**, *76*, 310–320.

(41) Paier, J.; Hirschl, R.; Marsman, M.; Kresse, G. The Perdew–Burke–Ernzerhof exchange-correlation functional applied to the G2-1 test set using a plane-wave basis set. *J. Chem. Phys.* **2005**, *122*, 234102.

(42) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103–128.

(43) Wannier, G. H. The Structure of Electronic Excitation Levels in Insulating Crystals. *Phys. Rev.* **1937**, *52*, 191–197.

(44) Silvestrelli, P. L. Maximally localized Wannier functions for simulations with supercells of general symmetry. *Phys. Rev. B* **1999**, *59*, 9703–9706.

(45) Berghold, G.; Mundy, C. J.; Romero, A. H.; Hutter, J.; Parrinello, M. General and efficient algorithms for obtaining maximally localized Wannier functions. *Phys. Rev. B* **2000**, *61*, 10040–10048.

(46) Kirchner, B.; Hutter, J. Solvent effects on electronic properties from Wannier functions in a dimethyl sulfoxide/water mixture. *J. Chem. Phys.* **2004**, *121*, 5133–5142.