The Journal of
Chemical Physics

Ultrafast estimation of electronic couplings for electron transfer between pi-conjugated organic molecules. II

Accepted Manuscript: This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination, and proofreading process, which may lead to differences between this version and the Version of Record.

Cite as: J. Chem. Phys. (in press) (2021); https://doi.org/10.1063/5.0076555
Submitted: 26 October 2021 • Accepted: 06 December 2021 • Accepted Manuscript Online: 07 December 2021

Orestis George Ziogos and **Jochen Blumberger**

openaccess

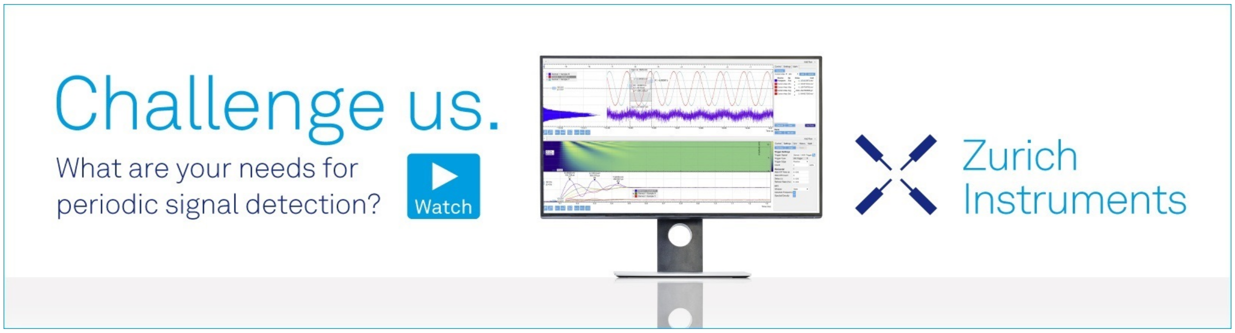
View Online
Export Citation
CrossMark

ARTICLES YOU MAY BE INTERESTED IN

HAB79: A New Molecular Dataset for Benchmarking DFT and DFTB Electronic Couplings Against High-Level Ab-initio Calculations
The Journal of Chemical Physics (2021); https://doi.org/10.1063/5.0076010

DFTB+, a software package for efficient approximate density functional theory based atomistic simulations
The Journal of Chemical Physics **152**, 124101 (2020); https://doi.org/10.1063/1.5143190

CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations
The Journal of Chemical Physics **152**, 194103 (2020); https://doi.org/10.1063/5.0007045


Challenge us.
What are your needs for periodic signal detection?
Watch
Zurich Instruments

AIP Publishing

J. Chem. Phys. (in press) (2021); https://doi.org/10.1063/5.0076555

(c) 2021 Author(s).

Ultrafast estimation of electronic couplings for electron transfer between pi-conjugated organic molecules. II

Orestis George Ziogos[1,a)] and Jochen Blumberger[1,b)]

[1] Department of Physics and Astronomy and Thomas Young Centre, University College London, Gower Street, London WC1E 6BT, United Kingdom

a) Electronic mail: o.ziogos@ucl.ac.uk

b) Electronic mail: j.blumberger@ucl.ac.uk

Abstract

The development of highly efficient methods for the calculation of electronic coupling matrix elements between electron donor and acceptor is an important goal in theoretical organic semiconductor research. In previous work we introduced the analytic overlap method (AOM) for this purpose, which is an ultrafast electronic coupling estimator parametrized to and orders of magnitude faster than density functional theory (DFT) calculations at a reasonably small loss in accuracy. In this work we reparametrize and extend AOM to molecules containing nitrogen, oxygen, fluorine, and sulfur heteroatoms using 921 dimer configurations from the recently introduced HAB79 dataset. We find again a very good linear correlation between the frontier orbital overlap, calculated ultrafast in an optimized minimum Slater basis, and DFT reference electronic couplings. The new parametrization scheme is shown to be transferable to sulfur-containing polyaromatic hydrocarbons in experimentally resolved dimeric configurations. Our extension of AOM enables high-throughput screening of very large databases of chemically diverse organic crystal structures and the application of computationally intense non-adiabatic molecular dynamics methods to charge transport in state-of-the-art organic semiconductors, e.g., non-fullerene acceptors.

I. Introduction

The calculation of electronic coupling matrix elements, or transfer integrals, is at the very heart of charge transfer simulations in biology and materials science [1-4]. Computational techniques for electronic couplings have significantly matured over the last 20 years and, depending on the requirements of the problem at hand, a large number of methods can now be employed ranging from high accuracy but time consuming ab-initio calculations [5-7] to density functional theory (DFT) calculations, e.g. time-dependent DFT (TDDFT) [8], constrained DFT (CDFT) [9-14], projector operator-based diabatization (POD) [15-17], fragment-orbital DFT (FODFT) [18, 19], and frozen density embedding [20, 21], to fast semi-empirical density functional tight binding (DFTB) [6, 7, 22], and the ultrafast analytic overlap method (AOM) [23].

For the modelling of charge carrier transport in soft condensed media (e.g., organic and biological semiconductors) it is important to properly account for the strong thermal fluctuations of the molecules and several methods have been developed to do that, e.g. Kinetic Monte Carlo [24-27], transient localization theory [28, 29] or non-adiabatic molecular dynamics (MD) simulations [30-31]. A common feature of these methods is the very large number of transfer integrals that need to be evaluated to converge charge mobilities. As a result, one needs to select an appropriate electronic coupling calculation scheme as to strike balance between computational efficiency and chemical accuracy.

The analytic overlap method for electronic coupling calculation was designed with these requirements in mind [23]. The idea behind this method is that the computationally expensive DFT calculation of charge transfer integrals can be effectively substituted by very rapid estimations of frontier molecular orbital (FMO) overlap integrals, multiplied by an appropriate scaling coefficient. The FMOs are reconstructed very efficiently in an optimized minimum Slater type orbital (STO) basis set allowing the ultrafast analytic calculation of FMO overlaps and electronic couplings for a large number of dimer configurations [32].

The parametrization of AOM relies on the compilation of a molecular training set and a series of single molecule DFT calculation for the determination of the FMOs and dimer electronic coupling calculations using an appropriate level of theory as to establish a linear approximation between overlap and charge transfer integrals. The original version of the AOM relied on planewave DFT calculations for both single molecule and dimeric configurations using a limited molecular training set comprised of small pi-conjugated molecules and C60 fullerenes, primarily focusing on carbon-rich compounds with minimal heteroatom contribution to the FMO electronic density [23].

From a materials point of view, the AOM has been successfully utilized for the modeling of charge transfer mechanisms for homocyclic pristine and functionalized polyaromatic hydrocarbons (PAHs) for organic electronics [33-36]. Furthermore, from a theoretical and method development perspective, the AOM is at the heart of the fragment orbital-based surface hopping non-adiabatic MD simulation methodology for charge transport modeling [34], while simultaneously constituting a robust electronic couplings estimator for pi-conjugated systems [7].

Two issues limit the applicability of our previous implementation of the AOM [23]. First, reference electronic couplings were obtained with planewave DFT calculations in the gas phase, which generated a significant computational overhead for large pi-conjugated organic molecules compared to atomic orbital-based DFT calculations. Second, the chemical diversity of the training molecular dataset with regard to types of PAH heteroatoms and dimer structural variability was very limited. The focus of this work is to effectively address both of these issues. With regard to the first issue, we take advantage of the recently implemented [16] POD method [15] for reference electronic coupling calculations from DFT with atom-centered Gaussian-type orbitals (GTO). AOM is recast here to support the DFT FMO orbitals in a GTO basis as input for FMO reconstruction in a minimum Slater basis. As regards to the diversity of the chemical space covered by the AOM, we utilize the newly introduced HAB79 dataset [37] containing nitrogen, oxygen, fluorine, and sulfur heteroatoms. This database contains 79 single molecules and 921 dimer configurations, providing a wealth of data for the reparametrization of key parameters of AOM (that is, Slater decay coefficients), hence enabling ultrafast electronic coupling calculation for a plethora of organic semiconductors found in state-of-the-art organic electronics and photovoltaics applications (e.g. non-fullerene acceptors).

The revised AOM framework provides reparametrized optimal Slater decay coefficients for carbon and nitrogen chemical species and new parameters for oxygen, fluorine, and sulfur (oxygen and sulfur parameters in the original AOM implementation relied on unoptimized Clementi-Raimondi (C-R) reference values [38] and fluorine was not included). Using the new parametrization, a universal scaling constant satisfying the linear relationship between electronic coupling and overlap integrals is established and successfully tested on a molecular dataset taken from experimentally resolved molecular crystal structures.

The layout of the paper is organized follows: all computational details are outlined in Sec. II, followed by the results in Sec. III. The conclusions of this study are presented in Sec. IV.

II. Computational details

The AOM asserts that the electronic coupling H_{ab} between two non-orthogonal diabatic charge transfer electronic states, Ψ_{a} and Ψ_{b}, is, to a good approximation, proportional to the overlap integral S_{ab} between the two states, thus satisfying the linear relationship

$$|H_{\mathrm{ab}}| = C|S_{\mathrm{ab}}| \qquad (1)$$

where

$$S_{\mathrm{ab}} = \langle \Psi_{\mathrm{a}} | \Psi_{\mathrm{b}} \rangle \qquad (2)$$

We note that a similar relation is used in Hückel theory to approximate the resonance integral. The difference is that in the latter the resonance integral refers to the electronic interaction of atomic orbitals centered at neighboring atoms on the same molecule, whereas in AOM the coupling matrix element is between two diabatic wavefunctions centered on neighboring molecules.

In most methods for coupling matrix element calculation, including POD, the two wavefunctions Ψ_{a} and Ψ_{b} are approximated by localized FMOs obtained by diabatization of a dimer electronic Hamiltonian (highest occupied FMOs on donor and acceptor for hole transfer and lowest unoccupied FMOs for excess electron transfer). These diabatized FMOs are very similar to the FMOs of the isolated molecules, the highest occupied electronic orbitals (HOMO) or lowest unoccupied molecular orbitals (LUMO), but not identical with them due to donor-acceptor interactions in the dimer Hamiltonian (if included) and the presence of orthogonalization tails. Hence, if the (non-orthogonal) FMOs of the isolated donor and acceptor are used for the calculation of S_{ab}, as is done in this work, "a" and "b" on the left- and right-hand side of Eq. 1 refer to slightly different orbitals. In practice, this is not a problem because any such differences are absorbed in the C constant.

In the following we present explicit expressions of the FMOs in terms of an optimized minimal Slater basis set that allows for an ultrafast calculation of overlap, and of electronic coupling via Eq. 1. The FMOs of donor and acceptor are initially resolved by means of DFT calculations on the isolated molecules using an atom-centered GTO basis set. For instance, the HOMO of the donor molecule, $|\phi'^{N}_{\mathrm{D}}\rangle$ (notation as in Ref. [23]), is expressed as

$$|\phi'^{N}_{\mathrm{D}}\rangle = \sum_{i=1}^{N_{\mathrm{GTO}}} d_i |\psi_i\rangle, \qquad (3)$$

where

$$|\psi_i\rangle = \psi_{l_x l_y l_z}(\alpha;\vec{r}) = \left(\frac{2\alpha}{\pi}\right)^{3/4} \sqrt{\frac{(8\alpha)^{l_x+l_y+l_z} l_x! l_y! l_z!}{(2l_x)!(2l_y)!(2l_z)!}}\, x^{l_x} y^{l_y} z^{l_z} \exp(-\alpha r^2) \quad (4)$$

is a primitive GTO in its cartesian form and N_{GTO} the number of GTO basis functions. Similar expansions are made for the HOMO of the acceptor or for the LUMO of donor and acceptor in case of excess electron transfer. The FMO is subsequently projected into a STO basis set only with s- and p-type angular momentum contributions:

$$|\phi'^{N}_{\mathrm{D}}\rangle \approx |\tilde{\phi}'^{N}_{\mathrm{D}}\rangle = \sum_{i=1}^{N_{\mathrm{STO}}} c_i |\chi_i\rangle \qquad (5)$$

where N_{STO} the number of STO basis functions and

$$|\chi_k\rangle = \chi_{nlm}(\mu;\vec{r}) = R_n(\mu;r) Y_{lm}(\theta,\varphi) \qquad (6)$$

with $R_n(\mu;r)$ the radial part of the STO orbital:

$$R_n(\mu;r) = \frac{(2\mu)^{n+1/2}}{\sqrt{(2n)!}} r^{n-1} \exp(-\mu r) \qquad (7)$$

and $Y_{lm}(\theta,\varphi)$ an orthonormal real spherical harmonic:

$$Y_{lm}(\theta,\varphi) = (-1)^m \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}} P_l^{|m|}(\cos\theta) \times \begin{cases} \sqrt{2}\cos(m\varphi) & m > 0 \\ 1 & m = 0 \\ \sqrt{2}\sin(|m|\varphi) & m < 0 \end{cases} \qquad (8)$$

The GTO-to-STO FMO projection is effectively reduced to the solution of the following linear system of equations

$$\sum_{i=1}^{N_{\mathrm{GTO}}} d_i \langle \chi_j | \psi_i \rangle = \sum_{i=1}^{N_{\mathrm{STO}}} c_i \langle \chi_j | \chi_i \rangle, \forall j \in [1, N_{\mathrm{STO}}] \quad (9)$$

For the sake of computational simplicity and efficiency, STO-GTO overlap integrals $\langle \chi_j | \psi_i \rangle$ are calculated using the Gaussian expansion method by O-ohata et al. [39] using expansions of 10, 6, and 8 GTOs for 1s and 2s, 2p, 3s and 3p STO orbitals, respectively.

Defining the projection completeness as

$$p(\{\mu_i\}) = \langle \tilde{\phi}'^{N}_{\mathrm{D}} | \tilde{\phi}'^{N}_{\mathrm{D}} \rangle, \quad (10)$$

the set of STO exponential coefficients $\{\mu_i\}$ per chemical species and angular momentum channel is chosen so as to maximize $p(\{\mu_i\}) \leq 1$. In order to achieve an augmented level of chemical diversity, the projection completeness optimization procedure with respect to $\{\mu_i\}$ is carried out considering an ensemble of N_{mol} different molecules. To this end, we define a GTO-to-STO projection goodness metric function

$$p_{\mathrm{M}}(\{\mu_i\}) = \sqrt{[\mathrm{mean}(p) - 1]^2 + [\mathrm{var}(p)]^2} \qquad (11)$$

where $\mathrm{mean}(p) = \frac{1}{N_{\mathrm{mol}}}\sum_{j=1}^{N_{\mathrm{mol}}} p_j$ and $\mathrm{var}(p) = \frac{1}{N_{\mathrm{mol}}}\sum_{j=1}^{N_{\mathrm{mol}}}\left[p_j - \mathrm{mean}(p)\right]^2$. Due to the intricate nature of the multidimensional objective function $p_{\mathrm{M}}(\{\mu_i\})$, the minimization procedure is carried out by means of a derivative-free Nelder-Mead simplex optimization algorithm [40]. Using the optimized μ_i Slater decay coefficients, we can arrive at a set of $\{c_i\}$ expansion coefficients for every isolated molecule under study, thus concluding the GTO-to-STO projection procedure.

The expansion of the FMOs, Eq. 5, can be simplified further. The FMOs of pi-conjugated organic molecules are usually formed of linear combinations of p orbitals pointing in the direction perpendicular to the plane of pi-conjugation. Hence, we project the Slater p orbitals of each atom i participating in pi-conjugation on p-orbitals orthogonal to the plane, $p_{\pi,i}$, and on two p-orbitals parallel to the plane, $p_{\sigma 1,i}$ and $p_{\sigma 2,i}$. As shown before, the expansion coefficients of the p_σ orbitals and s orbitals can be neglected to a very good approximation [23]. Hence the expansion Eq. 5 simplifies to,

$$\left|\tilde{\phi}'^{N}_{\mathrm{D}}\right\rangle \approx \left|\bar{\phi}'^{N}_{\mathrm{D}}\right\rangle = \sum_{i=1}^{N_{\mathrm{STO}}} c_{p\pi,i}\left|p_{\pi,i}\right\rangle \quad (12)$$

A similar projection is carried out for the HOMO of the acceptor. After renormalization, these orbitals are then used to calculate the overlap in an optimal, minimalistic representation of the FMOs,

$$|S_{\mathrm{ab}}| \approx |\bar{S}_{\mathrm{ab}}| = \left|\left\langle \bar{\phi}'^{N}_{\mathrm{D}} \middle| \bar{\phi}'^{N}_{\mathrm{A}} \right\rangle\right|. \quad (13)$$

The AOM estimate for electronic coupling is then,

$$|H_{\mathrm{ab}}| = \bar{C}|\bar{S}_{\mathrm{ab}}| \qquad (14)$$

where $\bar{C}$ is a constant of proportion obtained by a linear fit of $|\bar{S}_{\mathrm{ab}}|$ to $|H_{\mathrm{ab}}|$ values from POD DFT reference calculations. Given a set of dimeric configurations, in order to obtain optimal linear correlation between reference electronic coupling values $|H_{\mathrm{ab}}|$ and AOM calculated overlap integrals $|\bar{S}_{\mathrm{ab}}|$, a further tuning of the 2p and 3p STO exponential coefficients $\{\bar{\mu}_i\}$ per chemical species is required. This optimization is carried out using the Nelder-Mead simplex algorithm, aiming to minimize the sum of squared residuals during a least-squares linear fit between $|H_{\mathrm{ab}}|$ and $|\bar{S}_{\mathrm{ab}}|$ in a logarithmic scale, effectively calculating the intercept $\bar{C}' = \log \bar{C}$ and minimizing the residual in equation

$$\log|H_{\mathrm{ab}}| = \bar{C}' + \log|\bar{S}_{\mathrm{ab}}|. \qquad (15)$$

As regards technical details of DFT single molecule calculations, FMOs of isolated molecules are calculated at the PBE/DZVP-GTH minimal level of theory [41, 42]. Electronic couplings from scaled POD (sPOD) calculations at the PBE/DZVP-GTH level of theory were taken from the HAB79 dataset [37]. The POD couplings were scaled by a factor of 1.282, as dictated by the correlation analysis using reference ab-initio values at the minimal active space CASSCF/NEVPT2 level of theory, corresponding to a mean relative signed error of -3.3% and a mean relative unsigned error of 17.9% [37]. All DFT calculations are carried out using the CP2K package [43].

III. Results and discussion

The training of the revised AOM scheme in this work is carried out using the newly introduced HAB79 dataset: a selection of planar heterocyclic PAHs with non-degenerate FMOs, containing nitrogen, oxygen, fluorine, and sulfur heteroatoms with significant contribution to the electronic density, and capable of either n- or p-type semiconducting behavior [37].

The first step in the training procedure is dedicated to isolated molecules and aims at the determination of the optimal STO decay exponents $\{\mu_i\}$ for atomic species and orbital type and molecule-specific expansion coefficients $\{c_i\}$, both optimizing the statistical distribution of the GTO-to-STO projection completeness value via the metric described in Eq. 11, with the STO decay exponents explicitly treated during the optimization procedure and the expansion coefficients derived via the solution of Eq. 9 for every $\{\mu_i\}$ set. We have carried out two types of computational procedures for $\{\mu_i\}$. In the first procedure we group the HAB79 molecules in subsets that contain the same atom types actively participating to the FMO density and examine the landscape of the goodness metric of Eq. 11 by means of 2D $\{\mu_i\}$ grid evaluations. Doing so, we validate the global minima resolved by the Nelder-Mead simplex optimization on selected lower dimensional 2D grids. In the second procedure we carry out a simultaneous multi-dimensional optimization of $\{\mu_i\}$ for all molecules of HAB79.

In the first procedure, the HAB79 dataset is divided into 5 subsets, as depicted in Fig. 1, 2, and 3. Molecules 1-14 correspond to heterocyclic PAHs with nitrogen heteroatoms having significant FMO density contribution. In the same fashion, molecules 15-23, 24-30, and 31-53 bear oxygen, fluorine, and sulfur heteroatoms with sizable contribution to the electronic density. Finally, molecules 54-79 have multiple heteroatom contributions to the FMOs and are excluded from the 2D landscape analysis of the projection goodness metric.

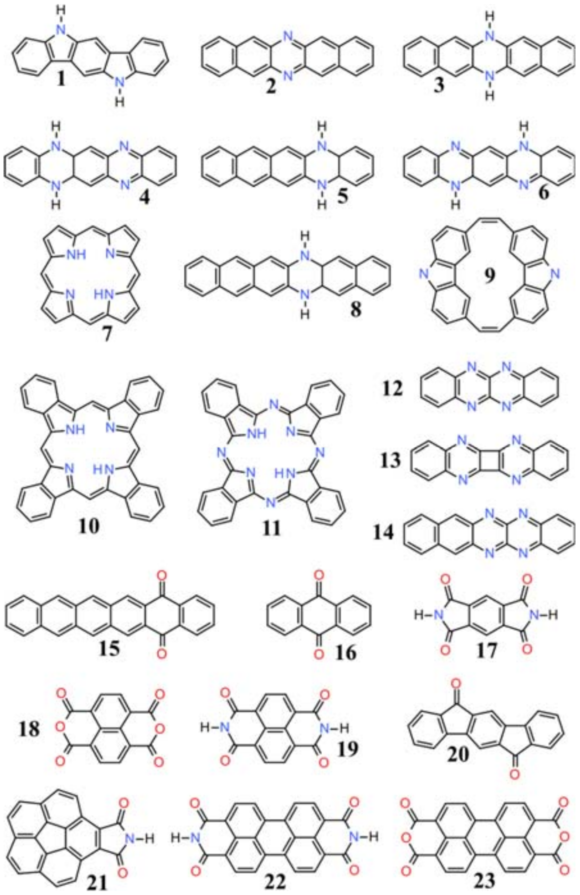

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Fig. 1. Nitrogen- and oxygen-based PAHs with significant heteroatom FMO density contribution. Molecules 1-11 and 15 are treated as hole conductors whereas 12-14 and 16-23 as electron conductors.

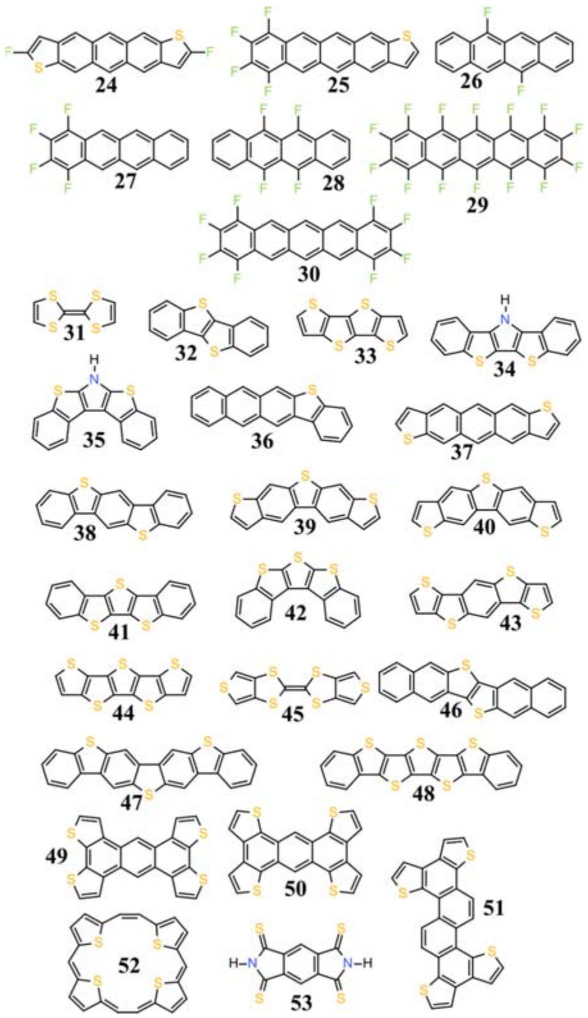

24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

Fig. 2. Fluorine and sulfur-based PAHs with significant heteroatom FMO density contribution. Molecules 24, 25 and 31-52 are treated as hole conductors whereas 26-30 and 53 as electron conductors.

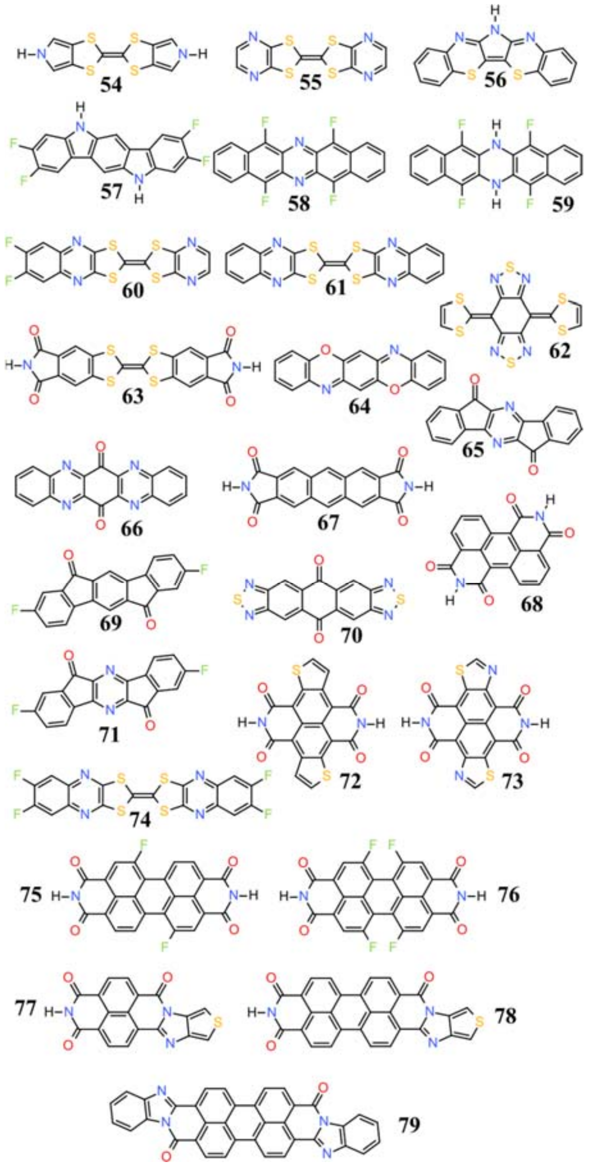


Fig. 3. PAH molecules with multi-heteroatom FMO density contributions. Molecules 54-63 are treated as hole conductors whereas 64-79 as electron conductors.

Considering the cases of X-based PAHs, where X corresponds to a N, O, F, or S heteroatom, the projection goodness metric is evaluated on a grid of $(\mu_{\mathrm{C}}, \mu_{\mathrm{X}})$ STO exponents. This procedure returns unimodal surfaces with a distinctive global minimum, indicating the existence of a unique set of decay exponents that can minimize the projection goodness metric.

The contour diagrams of the projection goodness metric, along with projection completeness histograms utilizing the C-R and optimized STO exponential coefficients for the aforementioned HAB79 subsets are depicted in Fig. 4 and Fig. 5. As a reminder, the C-R coefficients were obtained by Clementi and Raimondi [38] by means of SCF single atom calculations for the expression of atomic orbitals using a STO basis set – hence may not be optimal for our purpose. In our previous work [23], the C-R coefficients were used for oxygen and sulfur as the optimization of these coefficients was not possible due to a lack of molecules in the previously used training set with significant contributions to the FMO density on these atom types. The optimization procedure clearly shifts all projection completeness distributions towards unity, and reduces the standard deviation relative to C-R.

In the second procedure, a simultaneous minimization using the complete HAB79 dataset (molecules 1-79) is carried out by means of a Nelder-Mead simplex, resulting in the optimized values reported in Table I.

The performance of the Nelder-Mead derivative-free optimization procedure in minimizing the projection goodness metric objective function is illustrated in Fig. 6. Using the C-R STO exponential decay coefficients as a reference, the new set of optimized $\{\mu_i\}$ parameters correspond to a more complete GTO-to-STO projection, exhibiting tighter distributions closer to 1.0. In what follows below we use the set of $\{\mu_i\}$ values from the Nelder-Mead simplex optimization procedure.

Table I. Optimized STO decay exponential coefficients from simultaneous optimization using the Nelder-Mead simplex method (second procedure, see main text; 2p for N, O, and F heteroatoms and 3p for S). Literature values are also listed for comparison. All values are expressed in Bohr radii.

	C	N	O	F	S
This work; projection $\{\mu_i\}$	1.4427	1.6467	1.8588	2.1364	1.6517
This work; AOM $\{\bar{\mu}_i\}$	1.3856	1.6171	1.5051	1.6652	1.6411
Gajdos et al[*]; projection $\{\mu_i\}$	1.3120	1.7000	2.2266	N/A	1.8273
Gajdos et al[*]; AOM $\{\bar{\mu}_i\}$	1.0000	1.5000	2.2266	N/A	1.8273
C-R[†]	1.5679	1.9170	2.2266	2.5500	1.8273

[*] : Ref. [23]

[†] : Ref. [38]

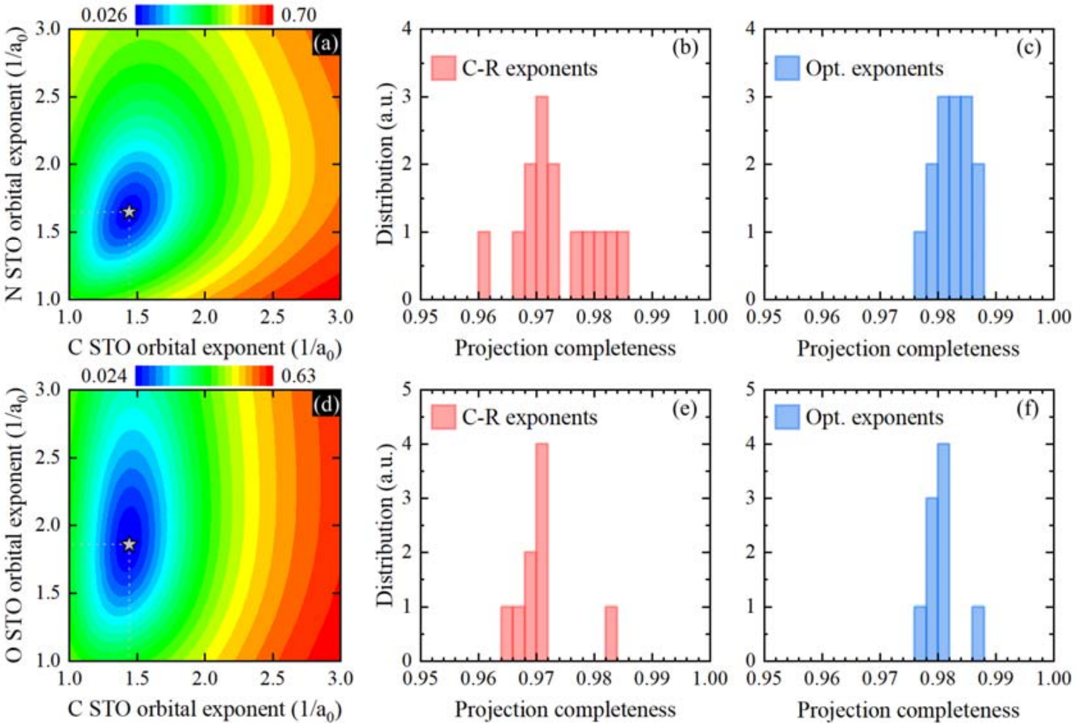


Fig. 4. (a, d) Projection goodness metric contour diagrams on a two-dimensional STO orbital exponents grid, along with projection completeness histograms based on (b, e) unoptimized C-R and (c, f) simplex-optimized STO orbital exponents (first procedure, see main text). The highlighted points are the metric minima resolved by the Nelder-Mead optimization procedure (second procedure, see main text). Diagrams (a-c) correspond to nitrogen-based PAHs 1-14 while (d-f) to oxygen-based PAHs 15-23, as depicted in Fig. 1.

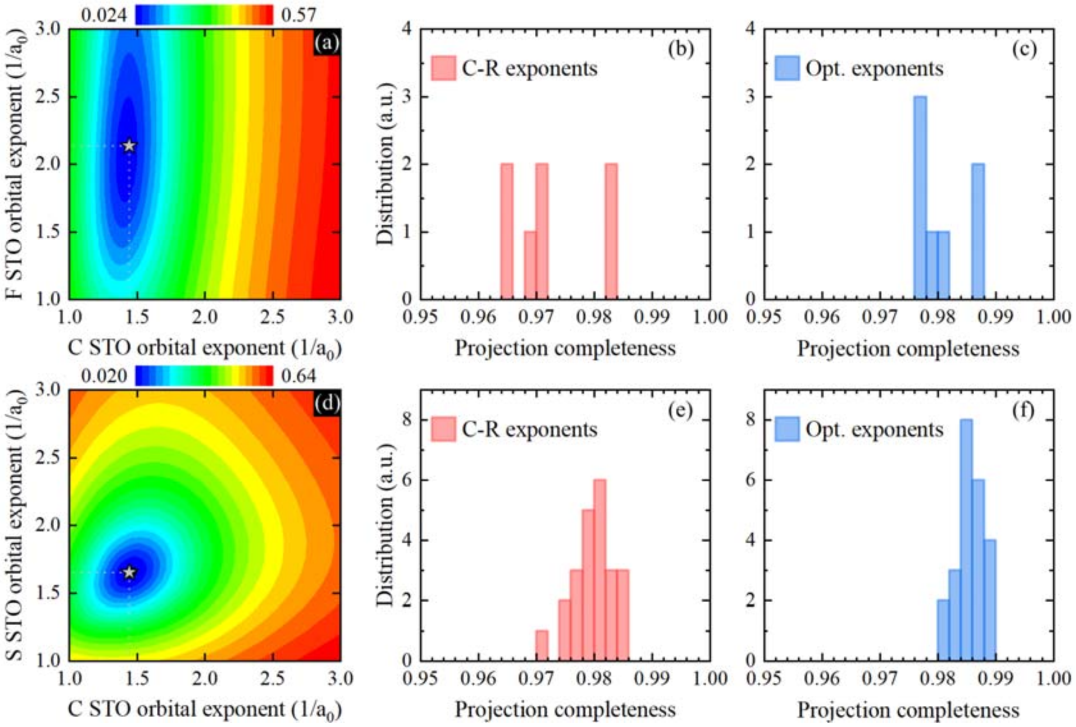


Fig. 5. (a, d) Projection goodness metric contour diagrams on a two-dimensional STO orbital exponents grid, along with projection completeness histograms based on (b, e) unoptimized C-R and (c, f) simplex-optimized STO orbital exponents (first procedure, see main text). The highlighted points are the metric minima resolved by the Nelder-Mead optimization procedure (second procedure, see main text). Diagrams (a-c) correspond to fluorine-based PAHs 24-30 while (d-f) to sulfur-based PAHs 31-53, as depicted in Fig. 2.

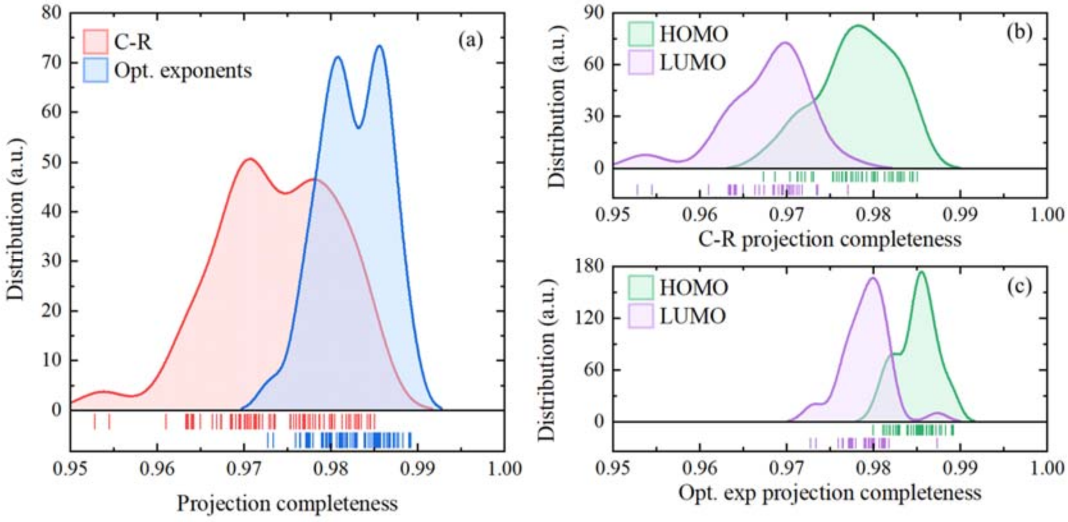


Fig. 6. (a) Projection completeness distribution diagrams of the full HAB79 dataset (molecule 1-79) when utilizing the original C-R and the Nelder-Mead optimized (second procedure; see

main text) STO orbital exponents. Panels (b) and (c) depict the HOMO/LUMO decomposition of the distributions.

After the calculation of the STO decay coefficients $\{\mu_i\}$ reported in Table I that optimize the projection procedure, we arrive at a set of STO-based FMO expansion coefficients $\{c_i\}$ for every molecule in the dataset. Previous work has shown that the set of $\{\mu_i\}$ coefficients derived from the projection part of this particular methodology cannot fully account for a sufficient linear relationship between overlap and electronic coupling values [23]. This is also the case for the current work, as demonstrated by the correlation plots depicted in Fig. 7.

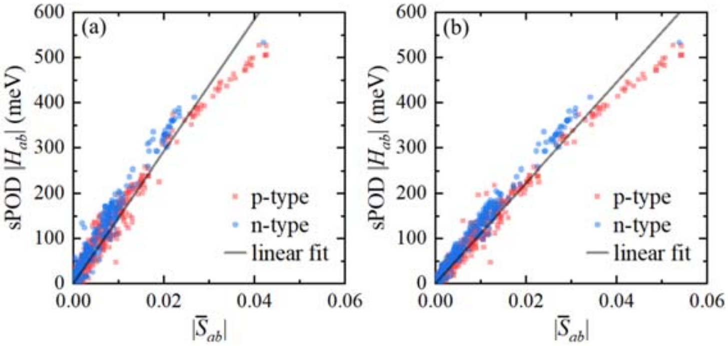


Fig. 7. Correlation diagrams between sPOD electronic couplings and AOM overlap values using (a) the C-R STO orbital exponents and (b) the exponents from the single molecule Nelder-Mean simplex projection optimization procedure for the calculation of the overlap.

As in Ref. [23], we introduce a second set of STO decay coefficients $\{\bar{\mu}_i\}$ to optimize the linear correlation between electronic coupling and overlap, while retaining the optimized expansion coefficients $\{c_i\}$ obtained from the GTO-to-STO projection with optimized decay coefficients $\{\mu_i\}$. Due to the complexity of the optimization problem at hand, we utilize a Nelder-Mead simplex algorithm aiming to minimize the sum of squared residuals during a least-squares linear fit of $\log|H_{\mathrm{ab}}|$ versus $\log|\bar{S}_{\mathrm{ab}}|$ according to Eq. 15. The set $\{\bar{\mu}_i\}$ reported in Table I corresponds to the global minimum attained by the simplex optimization procedure. Contour plots of the residual objective function calculated on two-dimensional grids of STO decay coefficients for HAB79 subsets are illustrated in Fig. 8, along with the simplex-derived global minimum position. The optimal linear correlation using the set $\{\bar{\mu}_i\}$ between overlap and charge coupling is shown in Fig. 9 with a universal scaling constant $\bar{C}$ equal to 9463 meV. We find that that the linear correlation is excellent for large coupling values (> 100 meV) and still rather good for values down to about 10 meV, while we notice a significant spread for smaller coupling values.

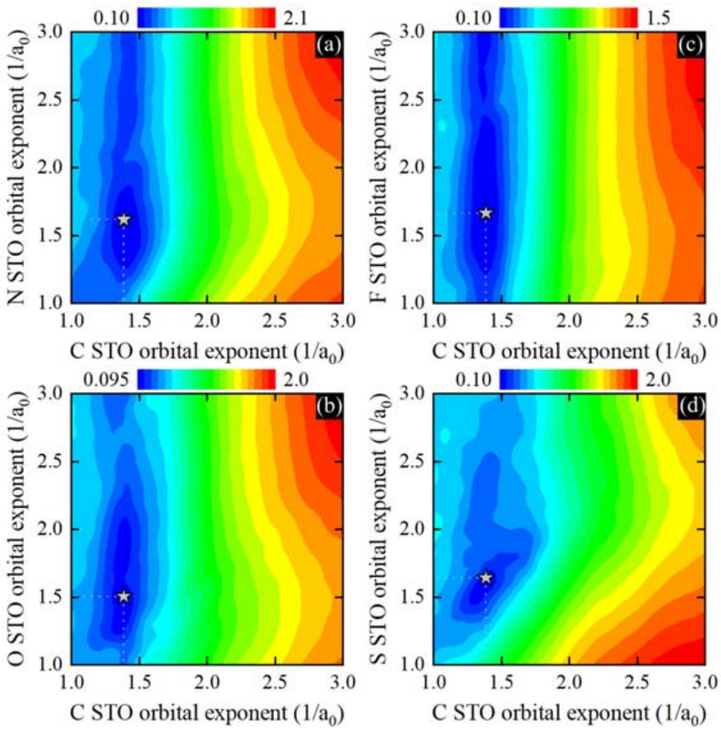


Fig. 8. Overlap-electronic coupling linear correlation metric contour diagrams on a two-dimensional AOM STO orbital exponents grid for (a) nitrogen, (b) oxygen, (c) fluorine, and (d) sulfur containing PAHs. The highlighted points are the metric minima for the set $\{\bar{\mu}_i\}$ obtained by the Nelder-Mead simplex optimization procedure.

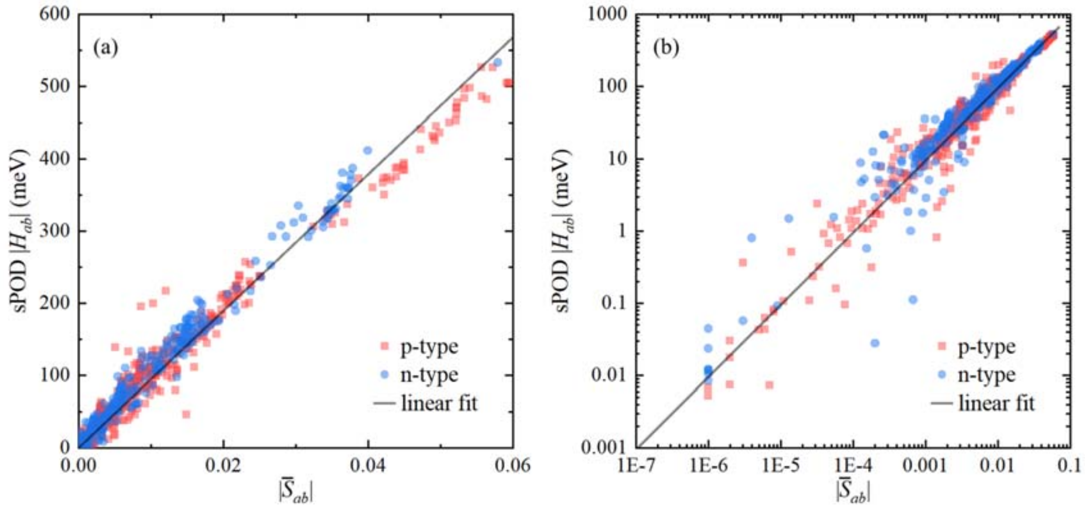


Fig. 9. Correlation between scaled POD electronic couplings and AOM overlap values using the set $\{\bar{\mu}_i\}$ for the calculation of the overlap and the set $\{\mu_i\}$ for the GTO-to-STO FMO projection in (a) linear and (b) logarithmic scale.

The reparametrization of the AOM based on the HAB79 dataset is tested against a set of 30 sulfur-based PAHs. The testing molecular set contains 7 molecules already present in the training HAB79 dataset, namely molecules 31, 32, 38, 40, 41, 42, and 48. The molecules on the test dataset not belonging to the training set are depicted in Fig. 10.

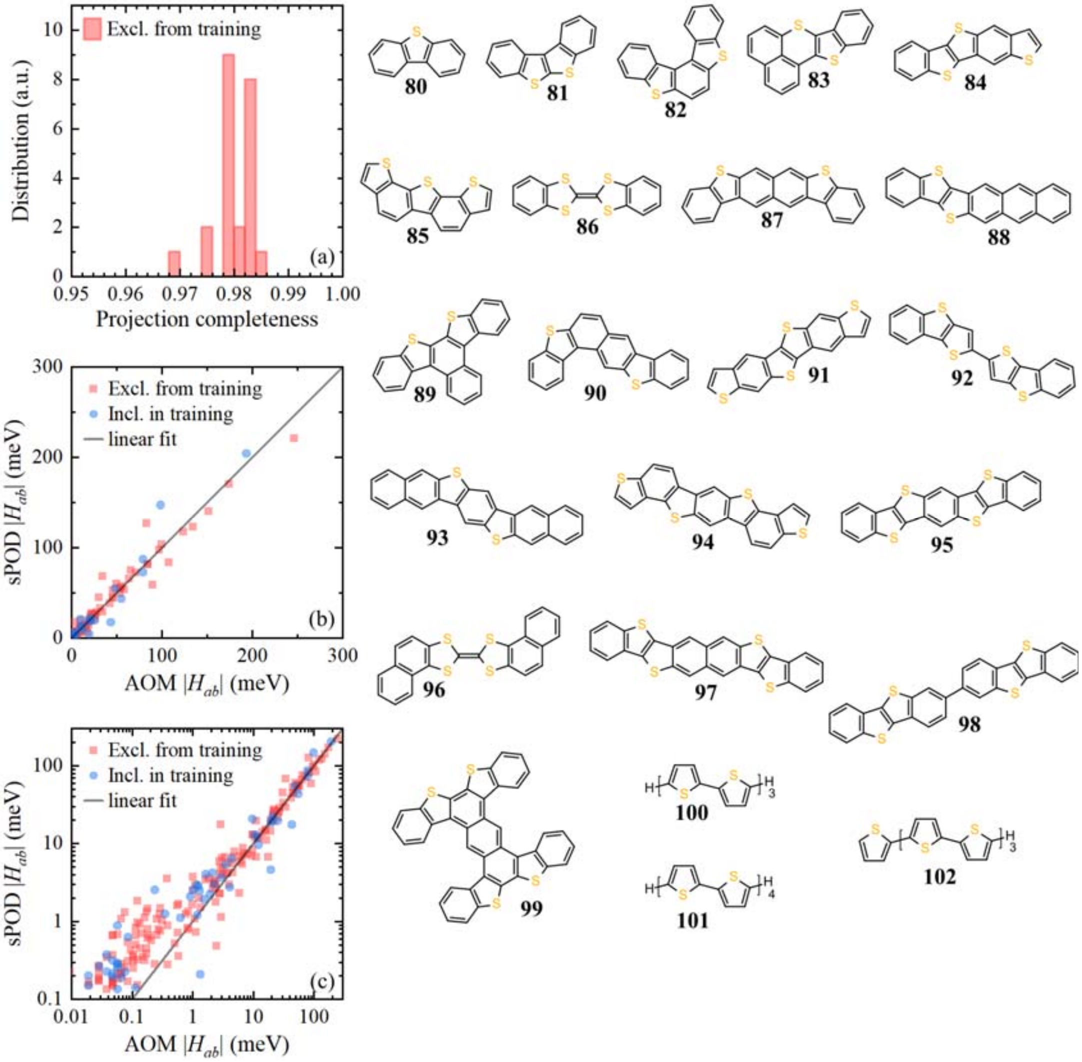


Fig. 10. Performance of the reparametrized AOM on a test set of 30 sulfur-containing PAHs. (a) Projection completeness histogram for molecules 80-102, not belonging to the AOM

training set. Correlation of AOM estimated and POD calculated charge transfer integrals in (b) linear and (c) logarithmic scale.

The Cambridge Crystallographic Data Centre (CCDC) deposition numbers for all compounds in the testing dataset are listed in the supplementary material [44]. It should be noted that, although consisting of a smaller number of configurations compared to the training set, the testing set constitutes a quite challenging benchmark, since it contains dimeric structures taken directly from experimentally resolved structures that are not present in the training set.

As a first test at the single molecule level of description, the GTO-to-STO projection completeness is evaluated for all 23 test molecules not belonging to the training dataset. The statistical distribution of these values using the optimized parameters $\{\mu_i\}$ from Table I is plotted in Fig. 10. All except 1 fall in the range obtained for the training set 0.97-0.99, showing excellent transferability of $\{\mu_i\}$ to sulfur-containing PAH outside the training set.

Moving on to molecular dimers, a set of test configurations of dimers is extracted from the experimentally-resolved molecular crystal structures reported in the CCDC records for the testing dataset. In order to keep dimers with significant charge transfer characteristics, scaled POD coupling calculations at the PBE/DZVP-GTH level of theory are carried out for all dimers having a center-of-mass separation smaller than 1.5 nm, discarding dimers with charge coupling values smaller than 0.1 meV. As a result, we obtain 233 dimer configurations.

The correlation between AOM and scaled POD coupling values is shown in Fig. 10, with the values of the dimers of the 23 test molecules not present in the training set depicted using square symbols and the 7 molecules belonging to the training set (but in different dimer configurations than in the HAB79 dataset) depicted using circles. The linear fitting procedure results in a slope of 0.99, thus perfectly correlating the AOM electronic coupling estimate with the reference sPOD DFT calculations. As for the training set, the agreement with reference sPOD couplings is excellent for couplings larger than 10 meV. For lower values, AOM underestimates sPOD couplings by up to an order of magnitude. The results on the training data (Fig. 9) suggest that the underestimation of small coupling values is not a systematic feature of AOM but specific to the molecules in the testing set.

Finally, a statistical error analysis is carried out on AOM electronic coupling values, calculating the mean unsigned error (MUE), mean relative signed error (MRSE), mean relative unsigned error (MRUE), and maximum absolute error (MAX). The results of this analysis are summarized in Table II using scaled POD and ab-initio values for the HAB79 dataset.

Table II. Statistical error analysis on AOM electronic couplings for the training and testing datasets.

Dataset	Reference	MUE (meV)	MRSE (%)	MRUE (%)	MAX (meV)
HAB79	sPOD	12.4	11.2	44.0	114.4
HAB79	ab-initio[*]	17.8	-8.4	36.6	182.5
Testing	sPOD	4.4	-10.5	32.4	49.1

[*] : Ref. [37]

The MRUE of AOM relative to sPOD is reasonable considering that this is an average error over electronic couplings that span 5-6 orders of magnitude. A further breakdown of the AOM error for each order of magnitude of reference sPOD coupling is summarized in Table III. As apparent from Fig. 7, the MRUE for couplings larger than 100 meV is very small, 8.5%, and still reasonably small for couplings between 10-100 meV, 24.7%. For couplings smaller than 10 meV, AOM appears to be less reliable. However, we would like to point out that also for DFT electronic structure calculations it becomes increasingly challenging to accurately calculate small electronic coupling values and converge them with respect to basis set and grid size, so some of the error could derive from inaccuracies of the reference calculations. Fortunately, the electronic couplings that determine transport in organic (opto-)electronic materials are typically on the order of 10-200 meV, which is the regime that is very well captured by the present AOM parametrization.

Table III. Interval statistical error analysis on AOM electronic couplings for the training dataset with respect to sPOD electronic coupling reference values.

sPOD H_{ab} interval (meV)	MUE (meV)	MRSE (%)	MRUE (%)	MAX (meV)
(0, 1]	0.7	299.8	357.5	12.9
(1, 10]	3.9	24.0	69.3	27.0
(10, 100]	12.2	-13.2	24.7	94.4
(100, 1000]	18.3	-4.8	8.5	114.4

The fact that the MRUE of AOM relative to ab-initio couplings is smaller than relative to sPOD is most likely a coincidence. The small statistical deviations reported for the sulfur-based PAH testing dataset constitute a promising feature as regards the predictive capabilities of AOM calculations.

IV. Conclusions

In this work we have reparametrized the analytic overlap method (AOM) for ultrafast electronic coupling calculation, we extended the parametrization to heteroatoms commonly found in semiconducting organic molecules, namely nitrogen, oxygen, fluorine and sulfur, and we made AOM compatible with DFT electronic coupling calculations that employ Gaussian-type basis sets. The original AOM implementation relied on unoptimized values for oxygen and sulfur, taken directly from the work of Clementi and Raimondi [38], and a parametrized value for nitrogen based only on the pyrrole molecule; hence the need for (re)parametrization in this work. The good linear correlation between frontier orbital overlap and DFT electronic coupling that we reported in our first paper [23] is retained for the much larger and more diverse database investigated in this work. The couplings of the vast majority of the 921 dimers investigated here can be well described by a single linear fit, i.e. by a single constant $\bar{C}$. The transferability of the revised parametrization was tested on a set of application relevant sulfur-containing PAHs that was not included in the parametrization. The error was not larger (in fact, smaller) than for the training set, which makes us confident that transferability is a general feature of AOM.

The major advantage of AOM is the sheer speed with which electronic couplings can be estimated as it requires only the analytic evaluation of overlap integrals using a small number of Slater type orbitals. One AOM coupling calculation on a dimer of a 152-atom molecule (e.g. a porphyrin dimer) typically takes 14 milliseconds on a single compute core, whereas the explicit POD electronic coupling calculations takes 527 seconds on 144 cores (Intel Xeon CPU E5-2630 v3 @ 2.40GHz), a speed-up of a factor of 10^4-10^5 and a saving in computing resources of a factor of 100.

The error of AOM, while reasonably small given the large chemical space and range of coupling values covered, can be reduced by parametrizing the linear relation, i.e. the $\bar{C}$ constant, specifically for a given molecule in question. This usually requires only a small number of dimer configurations, typically not more than a few tens [34, 35]. Another approach to reduce the error, which could be investigated in future work, is to use state-of-the-art machine learning methods (ML) to estimate and correct the error between AOM and DFT coupling estimates (ΔML). Compared to ML methods, AOM has the advantage that it already knows the complicated nodal shape of the FMOs, which determines electronic coupling, so this important feature no longer needs to learned. Hence, we expect that ΔML will need a much smaller number of reference data than ML models alone.

The AOM reparametrization procedure employed here is based on a well-defined derivative-free optimization scheme, the Nelder-Mead simplex method, and an appropriate choice of the objective functions. As a result, the method can be easily extended to cover additional application-relevant heteroatoms, such as other peripheral halogen atoms or PAH core boron, phosphorus, and heavy chalcogen additions – provided the existence and curation of appropriate training molecular datasets.

Finally, the new code for GTO-to-STO projection and for calculation of AOM electronic couplings is publicly available at https://github.com/oziogos/pyAOMlite, along with all reference single molecule and dimer information needed to validate the results presented in this work.

Supplementary material

Refer to the supplementary material for a full list of CCDC deposition numbers for all molecules belonging to the AOM sulfur-based PAH testing set.

Acknowledgments

This work was supported by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (Grant Agreement No. 682539/ SOFTCHARGE). Via our membership of the UK’s HEC Materials Chemistry Consortium, which is funded by EPSRC (Grant Nos. EP/L000202 and EP/R029431), this work used the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk), as well as the UK Materials and Molecular Modeling (MMM) Hub, which is partially funded by EPSRC (Grant No. EP/P020194), for computational resources. The authors also acknowledge the use of UCL Grace and Kathleen High Performance Computing Facilities.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] H. Oberhofer, K. Reuter, and J. Blumberger, Chem. Rev. **117**, 10319 (2017).

[2] G. Gryn'ova, A. Nicolaï, A. Prlj, P. Ollitrault, D. Andrienko, and C. Corminboeuf, J. Mater. Chem. C **5**, 350 (2017).

[3] G. Gryn'ova, K.-H. Lin, and C. Corminboeuf, J. Am. Chem. Soc. **140**, 16370 (2018).

[4] T. Nematiaram, S. Ciuchi, X. Xie, S. Fratini, and A. Troisi, J. Phys. Chem. C **123**, 6989 (2019).

[5] R. J. Cave and M. D. Newton, J. Chem. Phys. **106**, 9213 (1997).

[6] A. Kubas, F. Hoffmann, A. Heck, H. Oberhofer, M. Elstner, and J. Blumberger, J. Chem. Phys. **140**, 104105 (2014).

[7] A. Kubas, F. Gajdos, A. Heck, H. Oberhofer, M. Elstner, and J. Blumberger, Phys. Chem. Chem. Phys. **17**, 14342 (2015).

[8] D. Manna, J. Blumberger, J. M. L. Martin, and L. Kronik, Mol. Phys. **116**, 2497 (2018).

[9] Q. Wu and T. V. Voorhis, J. Chem. Phys. **125**, 164105 (2006).

[10] A. de la Lande and D. R. Salahub, J. Mol. Struct. THEOCHEM **943**, 115 (2010).

[11] H. Oberhofer and J. Blumberger, J. Chem. Phys. **133**, 244105 (2010).

[12] K. P. McKenna and J. Blumberger, Phys. Rev. B **86**, 245110 (2012).

[13] J. Blumberger and K. P. McKenna, Phys. Chem. Chem. Phys. **15**, 2184 (2013).

[14] N. Gillet, L. Berstis, X. Wu, F. Gajdos, A. Heck, A. de la Lande, J. Blumberger, and M. Elstner, J. Chem. Theory Comput. **12**, 4793 (2016).

[15] I. Kondov, M. Čížek, C. Benesch, H. Wang, and M. Thoss, J. Phys. Chem. C **111**, 11970 (2007).

[16] Z. Futera and J. Blumberger, J. Phys. Chem. C **121**, 19677 (2017).

[17] S. Ghan, C. Kunkel, K. Reuter, and H. Oberhofer, J. Chem. Theory Comput. **16**, 7431 (2020).

[18] K. Senthilkumar, F. C. Grozema, F. M. Bickelhaupt, and L. D. A. Siebbeles, J. Chem. Phys. **119**, 9809 (2003).

[19] H. Oberhofer and J. Blumberger, Angew. Chem. - Int. Ed. **49**, 3631 (2010).

[20] M. Pavanello, T. V. Voorhis, L. Visscher, and J. Neugebauer, J. Chem. Phys. **138**, 054101 (2013).

[21] P. Ramos, M. Papadakis, and M. Pavanello, J. Phys. Chem. B **119**, 7541 (2015).

[22] T. Kubař and M. Elstner, J. Phys. Chem. B **114**, 11221 (2010).

[23] F. Gajdos, S. Valner, F. Hoffmann, J. Spencer, M. Breuer, A. Kubas, M. Dupuis, and J. Blumberger, J. Chem. Theory Comput. **10**, 4653 (2014).

[24] J. Cornil, D. Beljonne, J.-P. Calbert, and J.-L. Brédas, Adv. Mater. **13**, 1053 (2001).

[25] J.-L. Brédas, D. Beljonne, V. Coropceanu, and J. Cornil, Chem. Rev. **104**, 4971 (2004).

[26] V. Coropceanu, J. Cornil, D. A. da Silva Filho, Y. Olivier, R. Silbey, and J.-L. Brédas, Chem. Rev. **107**, 926 (2007).

[27] A. Troisi, D. L. Cheung, and D. Andrienko, Phys. Rev. Lett. **102**, 116602 (2009).

[28] S. Ciuchi, S. Fratini, and D. Mayou, Phys. Rev. B **83**, 081202 (2011).

[29] S. Fratini, D. Mayou, and S. Ciuchi, Adv. Funct. Mater. **26**, 2292 (2016).

[30] T. Kubař and M. Elstner, J R Soc Interface **10**, 20130415 (2013).

[31] T. Kubař and M. Elstner, Phys. Chem. Chem. Phys. **15**, 5794 (2013).

[32] R. S. Mulliken, C. A. Rieke, D. Orloff, and H. Orloff, J. Chem. Phys. **17**, 1248 (1949).

[33] H. Yang, F. Gajdos, and J. Blumberger, J. Phys. Chem. C **121**, 7689 (2017).

[34] S. Giannini, A. Carof, M. Ellis, H. Yang, O. G. Ziogos, S. Ghosh, and J. Blumberger, Nat Commun **10**, (2019).

[35] S. Giannini, O. G. Ziogos, A. Carof, M. Ellis, and J. Blumberger, Adv. Theory Simul. **3**, 2000093 (2020).

[36] O. G. Ziogos, S. Giannini, M. Ellis, and J. Blumberger, J. Mater. Chem. C **8**, 1054 (2020).

[37] O. G. Ziogos, A. Kubas, Z. Futera, W. Xie, M. Elstner, and J. Blumberger, J. Chem. Phys. accepted for publication; http://dx.doi.org/10.1063/5.0076010

[38] E. Clementi and D. L. Raimondi, J. Chem. Phys. **38**, 2686 (1963).

[39] K. O-ohata, H. Taketa, and S. Huzinaga, J. Phys. Soc. Jpn. **21**, 2306 (1966).

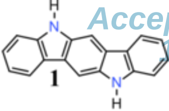
[40] J. A. Nelder and R. Mead, Comput. J. 7, **308** (1965).

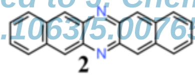
[41] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

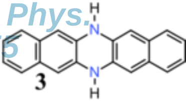
[42] J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996).

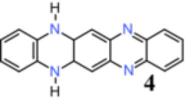
[43] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, WIREs Comput Mol Sci **4**, 15 (2013).

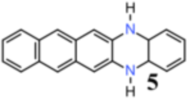
[44] C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, Acta Cryst. B **72**, 171 (2016).

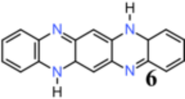

1

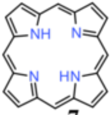

2

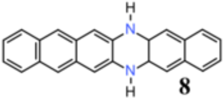

3

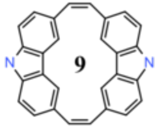

4

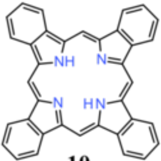

5

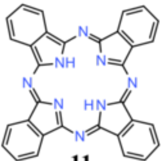

6

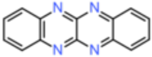

7

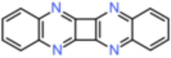

8

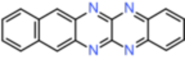

9

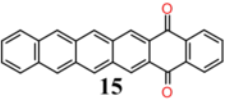

10

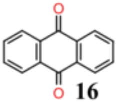

11

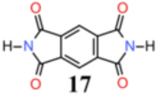
12


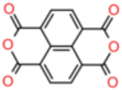
13


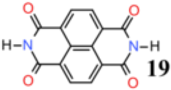
14


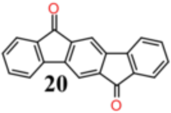

15

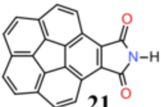

16

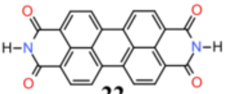

17

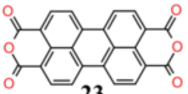
18



19

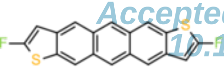

20

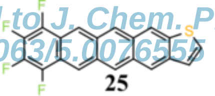

21

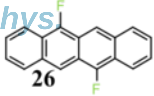

22

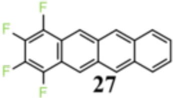

23

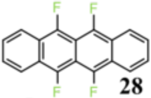
Accepted to J. Chem. Phys.
10.1063/5.0076555

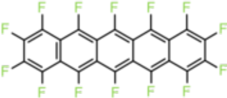

24

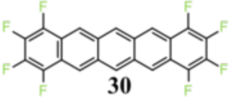

25

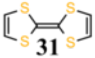

26

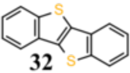

27

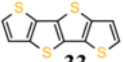

28

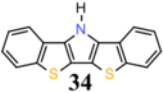

29

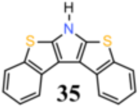

30

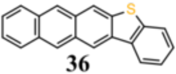

31

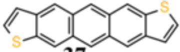

32

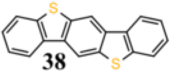

33

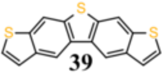

34

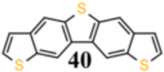

35

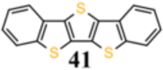

36

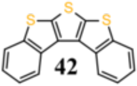

37

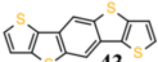

38

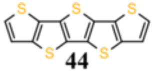

39

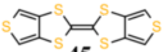

40

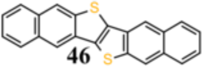

41

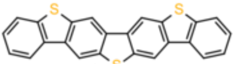

42

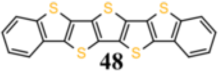

43

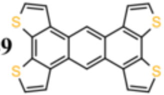

44

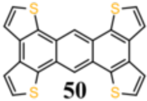

45

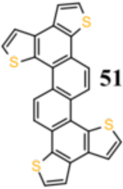

46

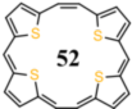

47

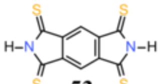

48


49

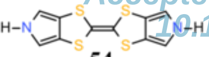

50

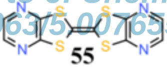

51

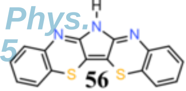

52

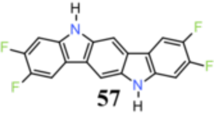

53

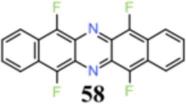
Accepted to J. Chem. Phys.
10.1063/5.0076555

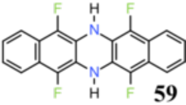

54

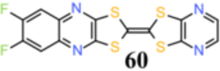

55

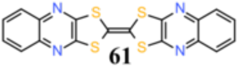

56

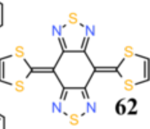

57

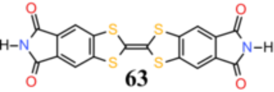

58

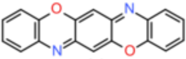

59

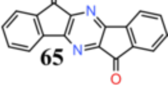

60

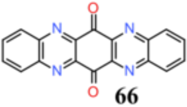

61

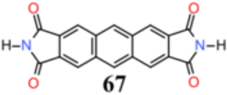

62

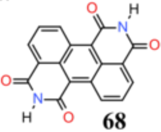

63

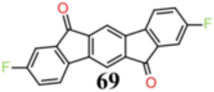

64

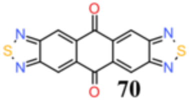

65

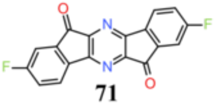

66

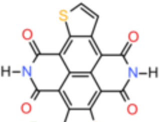

67

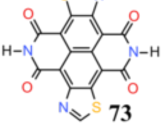

68

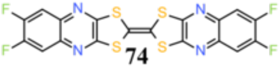

69

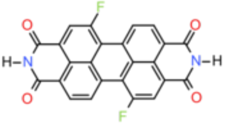

70

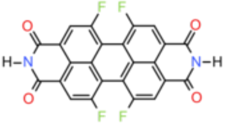

71

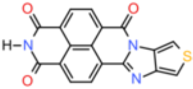

72

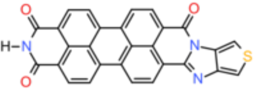

73

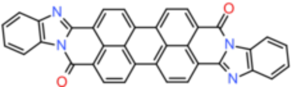

74

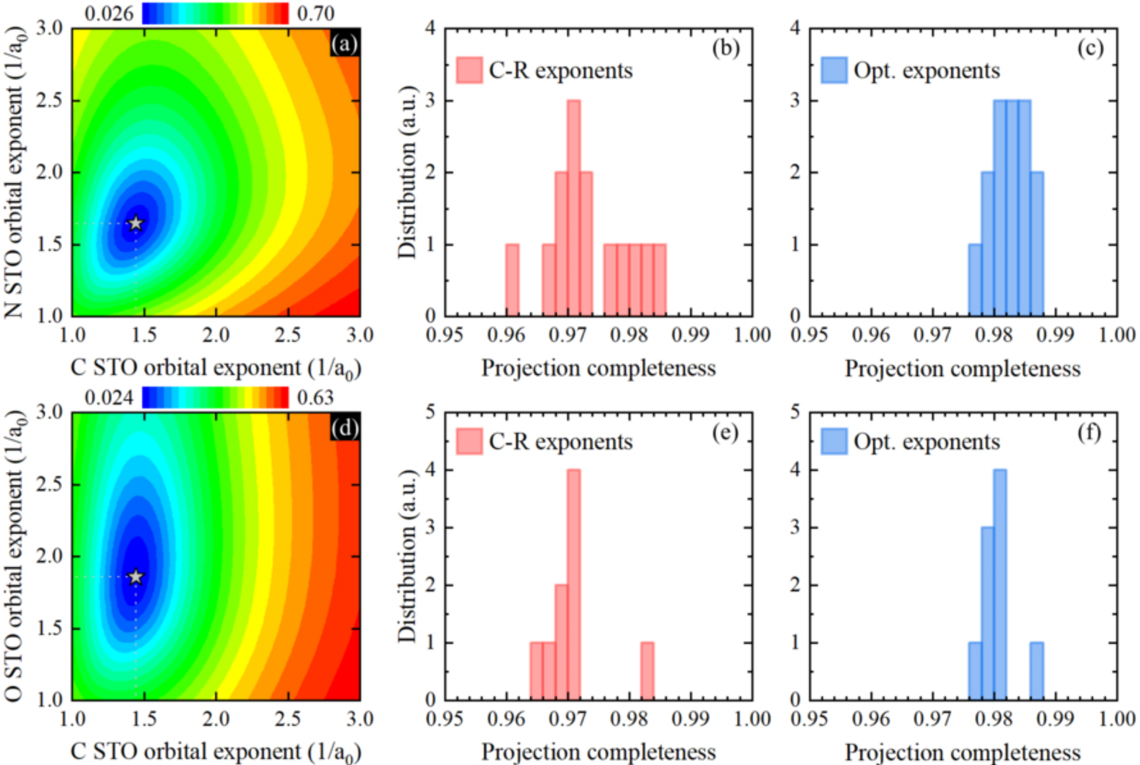

75

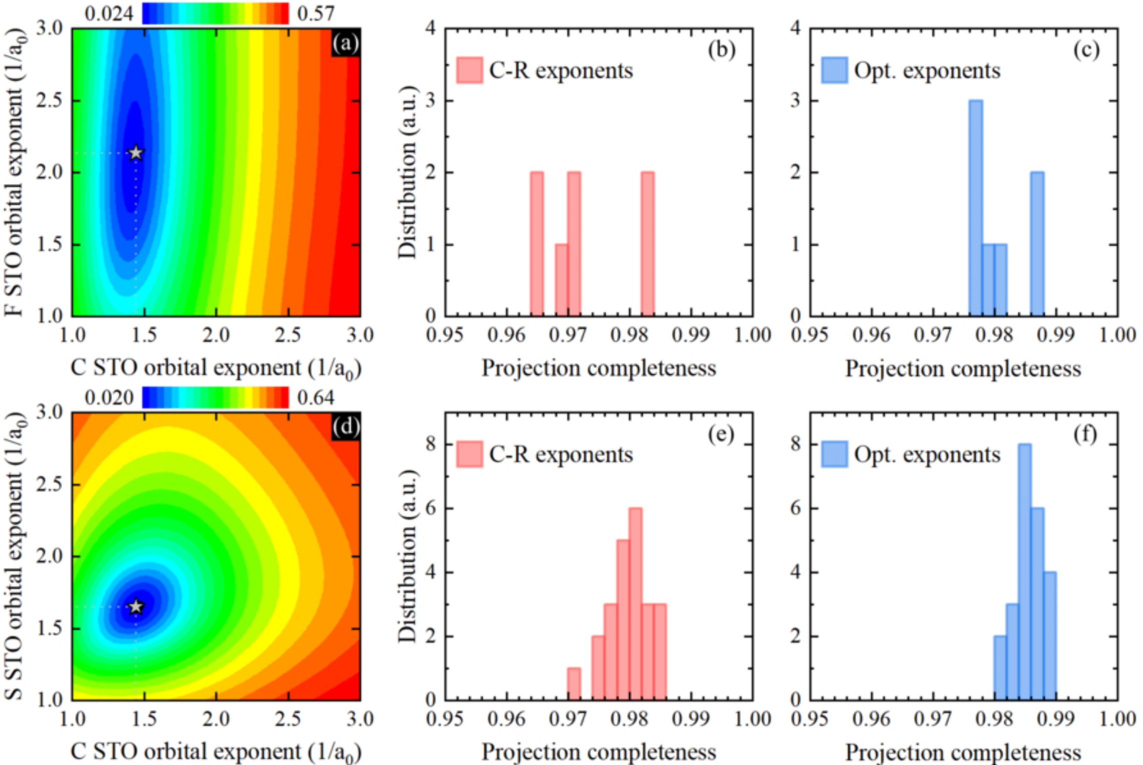

76

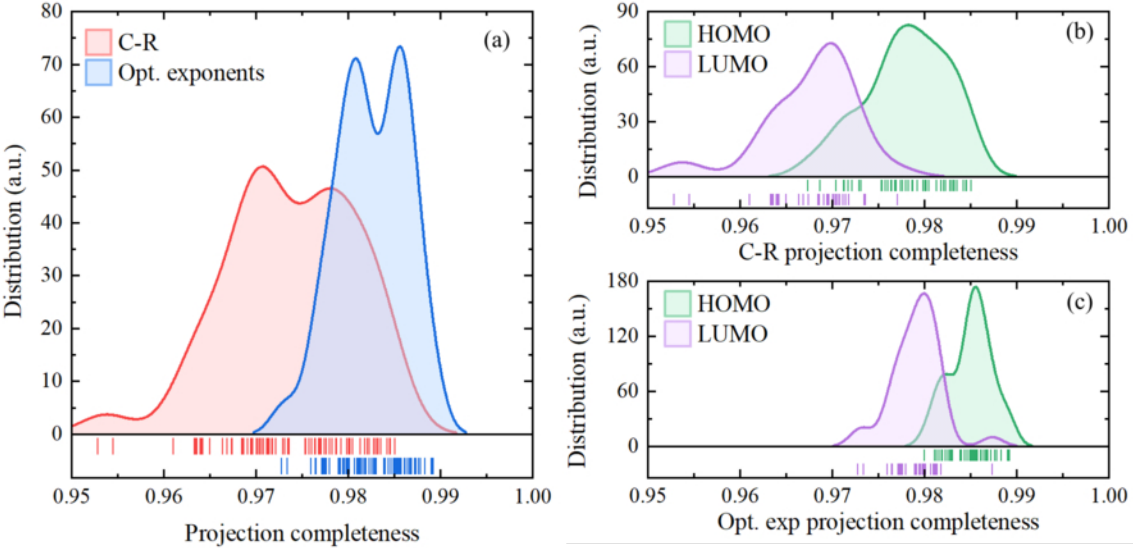

77

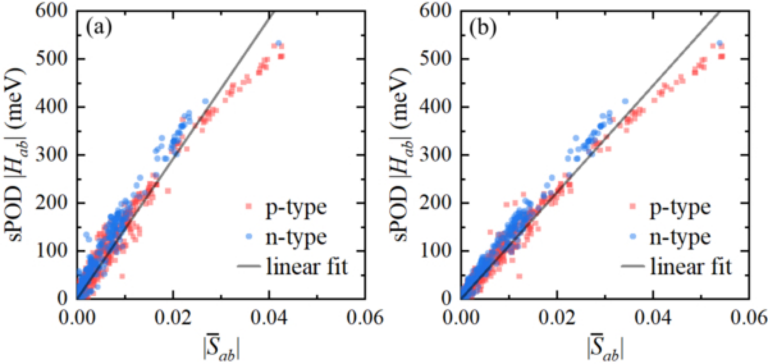

78

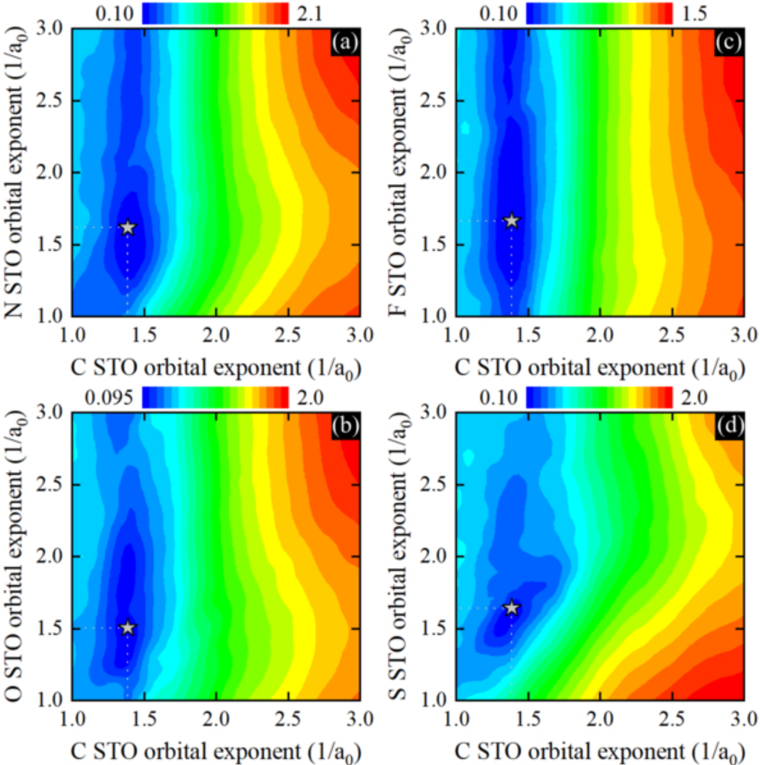

79

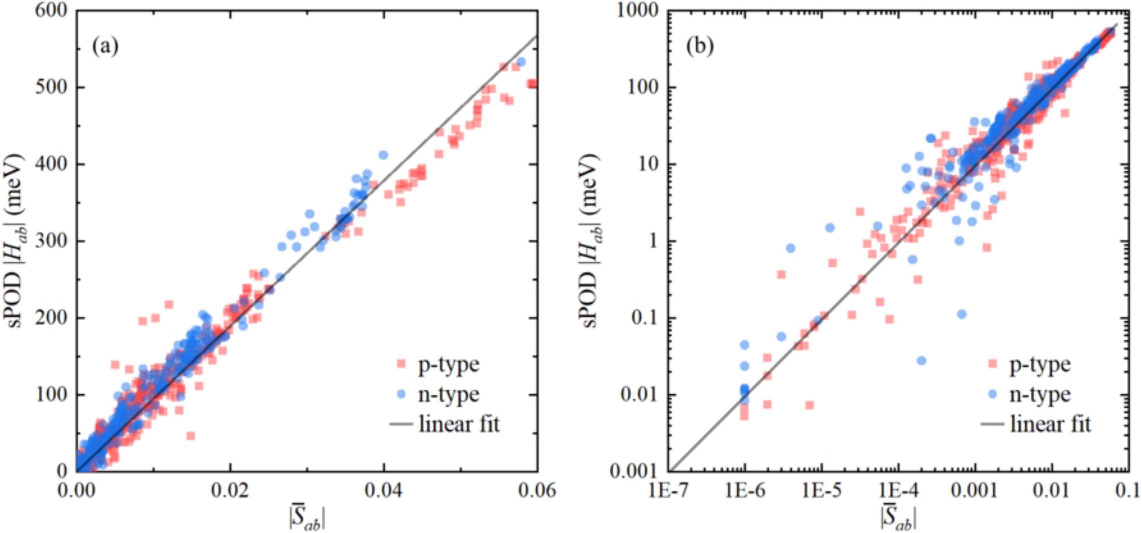

0.026
0.70
(a)
N STO orbital exponent (1/a$_0$)
C STO orbital exponent (1/a$_0$)
3.0
2.5
2.0
1.5
1.0
1.0
1.5
2.0
2.5
3.0
(b)
C-R exponents
Distribution (a.u.)
Projection completeness
0
1
2
3
4
0.95
0.96
0.97
0.98
0.99
1.00
(c)
Opt. exponents
Projection completeness
0.024
0.63
(d)
O STO orbital exponent (1/a$_0$)
C STO orbital exponent (1/a$_0$)
(e)
C-R exponents
Distribution (a.u.)
Projection completeness
5
(f)
Opt. exponents
Projection completeness

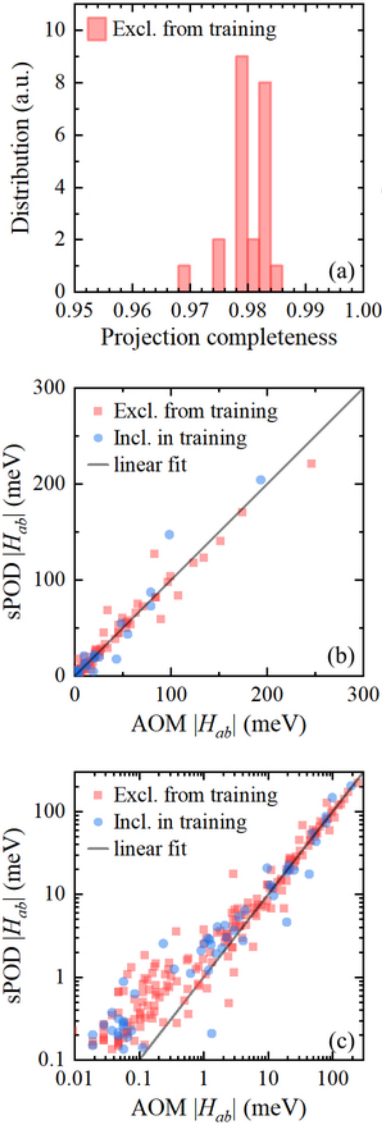

0.024
0.57
(a)
F STO orbital exponent ($1/a_0$)
C STO orbital exponent ($1/a_0$)
1.0
1.5
2.0
2.5
3.0
(b)
C-R exponents
Distribution (a.u.)
Projection completeness
0.95
0.96
0.97
0.98
0.99
1.00
(c)
Opt. exponents
Projection completeness
0.020
0.64
(d)
S STO orbital exponent ($1/a_0$)
C STO orbital exponent ($1/a_0$)
(e)
C-R exponents
Distribution (a.u.)
Projection completeness
(f)
Opt. exponents
Projection completeness
0
1
2
3
4
6
8

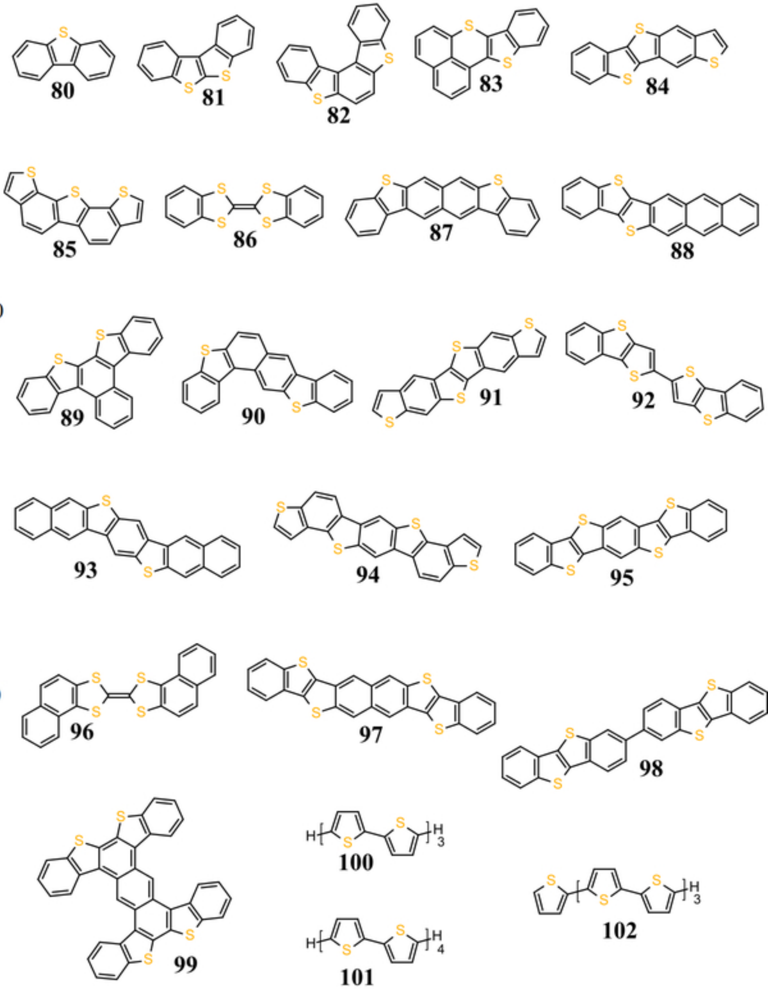

(a)
C-R
Opt. exponents
Distribution (a.u.)
80
70
60
50
40
30
20
10
0
0.95
0.96
0.97
0.98
0.99
1.00
Projection completeness
(b)
HOMO
LUMO
Distribution (a.u.)
90
60
30
0
C-R projection completeness
(c)
HOMO
LUMO
Distribution (a.u.)
180
120
60
0
Opt. exp projection completeness


(a)
(b)
600
500
400
300
200
100
0
sPOD $|H_{ab}|$ (meV)
0.00
0.02
0.04
0.06
$|\bar{S}_{ab}|$
p-type
n-type
linear fit


0.10
2.1
(a)
N STO orbital exponent (1/a$_0$)
C STO orbital exponent (1/a$_0$)
1.0
1.5
2.0
2.5
3.0
0.10
1.5
(c)
F STO orbital exponent (1/a$_0$)
0.095
2.0
(b)
O STO orbital exponent (1/a$_0$)
0.10
2.0
(d)
S STO orbital exponent (1/a$_0$)


(a)
sPOD $|H_{ab}|$ (meV)
600
500
400
300
200
100
0
0.00
0.02
0.04
0.06
$|\bar{S}_{ab}|$
p-type
n-type
linear fit
(b)
sPOD $|H_{ab}|$ (meV)
1000
100
10
1
0.1
0.01
0.001
1E-7
1E-6
1E-5
1E-4
0.001
0.01
0.1
$|\bar{S}_{ab}|$
p-type
n-type
linear fit


Excl. from training
Distribution (a.u.)
0
2
4
6
8
10
0.95
0.96
0.97
0.98
0.99
1.00
Projection completeness
(a)
Excl. from training
Incl. in training
linear fit
sPOD |H_{ab}| (meV)
0
100
200
300
AOM |H_{ab}| (meV)
(b)
Excl. from training
Incl. in training
linear fit
sPOD |H_{ab}| (meV)
0.1
1
10
100
0.01
AOM |H_{ab}| (meV)
(c)


80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
S
H
3
4