Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information of *Electron-vibrational renormalization in fullerenes through ab initio and machine learning methods*

Pablo García-Risueño*

Independent scholar, Barcelona, Spain

Eva Armengol and Àngel García-Cerdaña

Artificial Intelligence Research Institute,

(IIIA, CSIC) Carrer de Can Planes, s/n,

Campus UAB, 08193, Bellaterra, Catalonia, Spain

Juan María García-Lastra

Department of Energy Conversion and Storage,

Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

David Carrasco-Busturia†

DTU Chemistry, Technical University of Denmark,

DK-2800 Kgs. Lyngby, Denmark and

Division of Theoretical Chemistry and Biology,

School of Engineering Sciences in Chemistry, Biotechnology and Health,

KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden

(Dated: June 13, 2024)

Abstract

In this document we present information to supplement the main article.

PACS numbers: 63.22.Kn, 71.38.-k, 81.05.uj, 65.80.-g

CONTENTS

I. CONVERGENCE

In Tab. I we display the ab initio-calculated renormalizations of HOMO, LUMO and gap of several randomly chosen trial molecules for different values of the plane-wave cutoff and different values of the displacement parameter[1] (h). The former parameter informs about the size of the basis set used in the DFT calculations and is measured in Rydberg. The latter parameter gives the size of the finite-difference displacements of nuclear coordinates for the frozen-phonon calculations and has units of $1/\sqrt{\omega_\nu}$. The displayed results indicate that $h = 2$ provides converged results. They also indicate that the differences between renormalizations calculated using 30 Ry and 80 Ry are much lower than typical errors in electronic eigenvalues which arise because of choosing GGA xc functionals. Therefore we can consider that calculations are converged using a cutoff of 30 Ry. These conclusions agree with those presented in Ref. [1], where other small carbon-based

molecules (diamondoids) were analysed.

Fullerene	Pl-wave cutoff [Ry]	h	HOMO renorm. [meV]	LUMO renorm. [meV]	Gap renorm. [meV]
C30-C2v	30	2	26.32	43.01	16.69
C30-C2v	30	4	26.45	42.60	16.15
C30-C2v	30	6	26.46	42.53	16.07
C30-C2v	30	8	26.36	42.47	16.11
C30-C2v	80	2	29.85	44.71	14.86
C44-D3d	30	2	21.22	-15.92	-37.14
C44-D3d	60	2	26.77	-15.94	-42.71
C44-D3d	80	2	24.66	-16.17	-40.83
C80-Ih	30	2	-17.00	-29.30	-12.30
C80-Ih	40	1	-15.68	-31.99	-16.31
C80-Ih	40	2	-15.75	-32.29	-16.54
C80-Ih	40	4	-15.81	-31.27	-15.46
C84-D6h	30	2	73.12	-37.76	-110.88
C84-D6h	30	4	72.39	-38.95	-111.34

TABLE I: Renormalizations for test molecules for different values of the displacement parameter h and of the plane-wave cutoff.

II. INPUT VARIABLES

In this section we list the variables used in our regression analyses. These input variables are grouped into eleven sets of features: *Electronic structure features (selected, few, all)*, *Geometric features*, *Phonon features*, *Bond length features* and *Bond order features (Mayer, GJ, NM1, NM2, NM3)*.

Electronic structure features

The employed *electronic structure features* are listed in the table below.

Gap_{PBE}	$\varepsilon_{HOMO} - \varepsilon_{HOMO-1}$	$(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})^{-1}$	$\varepsilon_{LUMO+1} - \varepsilon_{LUMO}$	$(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$
Gap_{B3LYP}	$\varepsilon_{HOMO} - \varepsilon_{HOMO-2}$	$(\varepsilon_{HOMO} - \varepsilon_{HOMO-2})^{-1}$	$\varepsilon_{LUMO+2} - \varepsilon_{LUMO}$	$(\varepsilon_{LUMO+2} - \varepsilon_{LUMO})^{-1}$
AvgOcc	$\varepsilon_{HOMO} - \varepsilon_{HOMO-3}$	$(\varepsilon_{HOMO} - \varepsilon_{HOMO-3})^{-1}$	$\varepsilon_{LUMO+3} - \varepsilon_{LUMO}$	$(\varepsilon_{LUMO+3} - \varepsilon_{LUMO})^{-1}$
AvgEmpty	$\varepsilon_{HOMO} - \varepsilon_{HOMO-4}$	$(\varepsilon_{HOMO} - \varepsilon_{HOMO-4})^{-1}$	$\varepsilon_{LUMO+4} - \varepsilon_{LUMO}$	$(\varepsilon_{LUMO+4} - \varepsilon_{LUMO})^{-1}$
Deg_{HOMO}	$\varepsilon_{HOMO} - \varepsilon_{HOMO-5}$	$(\varepsilon_{HOMO} - \varepsilon_{HOMO-5})^{-1}$	$\varepsilon_{LUMO+5} - \varepsilon_{LUMO}$	$(\varepsilon_{LUMO+5} - \varepsilon_{LUMO})^{-1}$
Deg_{LUMO}	$\varepsilon_{HOMO} - \varepsilon_{HOMO-6}$	$(\varepsilon_{HOMO} - \varepsilon_{HOMO-6})^{-1}$	$\varepsilon_{LUMO+6} - \varepsilon_{LUMO}$	$(\varepsilon_{LUMO+6} - \varepsilon_{LUMO})^{-1}$

$(\varepsilon_{LUMO+1} - \varepsilon_{HOMO})^{-1}$	$\langle\, \varepsilon \,\rangle$
$(\varepsilon_{LUMO+2} - \varepsilon_{HOMO})^{-1}$	σ^2_ε
$(\, \varepsilon_{LUMO} - \varepsilon_{HOMO-1})^{-1}$	$\text{Skewness}_\varepsilon$
$(\varepsilon_{LUMO} - \varepsilon_{HOMO-2})^{-1}$	$\text{Kurtosis}_\varepsilon$.

Deg_{HOMO} and Deg_{LUMO} stand for the degeneracy of the HOMO and the LUMO, respectively. All the electronic structure variables were calculated with the PBE functional except Gap_{B3LYP}, which was calculated using the B(3)LYP functional. The used pseudopotentials are HGH[2]. The gaps are defined as the difference between the lowest unoccupied and the highest occupied eigenvalues $\varepsilon_{LUMO} - \varepsilon_{HOMO}$. Despite the well-known fact that gaps calculated using the GGA-PBE functional are underestimated, we use the gap from PBE calculations assuming that its errors are systematic, and hence the gap from PBE can still be used as an appropriate regressor. AvgOcc is the average of the inverse of the difference between the eigenvalues and the HOMO, this is:

$$\text{AvgOcc} \;:=\; \frac{1}{(N_e - \text{Deg}_{HOMO})} \left(\sum_{i=1}^{N_e - \text{Deg}_{HOMO}} (\, \varepsilon_i - \varepsilon_{HOMO}\,)^{-1} \right), \tag{1}$$

where N_e stands for the number of calculated eigenvalues; all the occupations of electronic eigenvalues are 2. AvgEmpty is calculated in an analogous manner, using the lowest unoccupied states:

$$\text{AvgEmpty} \;:=\; \frac{1}{20} \left(\sum_{j=N_e+\text{Deg}_{LUMO}+1}^{N_e+20} (\, \varepsilon_j - \varepsilon_{LUMO}\,)^{-1} \right). \tag{2}$$

Both AvgOcc and AvgEmpty, as well as the quantities with superscript -1 in the table above, are measured in eV^{-1}. The rest of the energies are measured in eV. The average, standard deviation,

skewness and Kurtosis coefficient of the eigenvalues were calculated taking into account just occupied states and the GGA-PBE exchange-correlation functional. Note that a simple comparison of eq. (1) of the main paper with the definitions of AvgOcc, AvgEmpty above indicates that AvgOcc, AvgEmptycan be considered to be proportional to $\Sigma^{Fan}(T = 0)$ in doing an approximation consisting in considering all Fan electron-phonon matrix elements (g) equal (with a severe truncation of states for the LUMO). The fact that this coarse-grain approximation has predictive power indicates that such hypothetical constant g would have a similar size for all fullerenes.

We stress that all relevant features listed in sec. III-B of the main article are easy to get from ground state calculations and a simple post-processing of their results. To this end one just needs to run the executions that follow in `Quantum Espresso`: a) A relaxation (geometry optimization); b) A ground state calculation using GGA-PBE with the relaxed positions; c) A ground state calculation using B3LYP with the relaxed positions. Example input files for (a), (b) and (c) can be found in `https://zenodo.org/records/10059442`. The names of the corresponding input files are `geomopt.in`, `scf.in` and `scf-gs-b3lyp.in`, respectively. The post-processing of the generated information generated can be done by executing the code `module_writing_input_data.py` which can be found at:
`https://github.com/pablogr/ML_fullerenes/blob/main/CODE_FOR_FORECASTING/`.
In addition, the reader can carry out the whole frozen-phonon calculations of arbitrary molecules following the methods presented in Ref. [1] and using the code and tutorials available at:
`https://github.com/pablogr/PHYS_Correction_of_anticrossings_frozen_phonon`.

	Δ^{B3LYP}	Δ^{PBE}	AvgOcc	AvgEmpty
HOMO renorm.	0.42	0.46	-0.92	-0.12
LUMO renorm.	-0.59	-0.60	0.03	-0.69
Gap renorm.	-0.70	-0.74	0.75	-0.33

TABLE II: Correlations between the output variables and some of the inputs.

In tab. II we display the Pearson's correlation of the renormalizations displayed in Tab. 2 of the main article (removing the C_{28} and $C_{58} - C_3$ outliers, as well as the 4-carbons-ring fullerene and the fullerene derivatives) with some of the input variables (the gaps from B3LYP and PBE

calculations, Δ^{B3LYP}, Δ^{PBE}, as well as AvgOcc and AvgEmpty). The correlations between AvgOcc and AvgEmpty (which are indeed quantities defined *ad hoc*) with the output variables reaches high sizes. Despite the fact that B3LYP calculations of the gap are considered to be more accurate than calculations using the GGA-PBE functional, the sizes of the correlations of the output variables are higher if Δ^{PBE} is used. Note that the units are omitted because they are irrelevant for the correlations.

The 38 regressors listed in the table above in this section form the feature set called *electronic structure (all)*. The feature set called *electronic structure (few)* is a subset of it, which is formed by the ten regressors listed in the first and last columns of the table. The feature set called *electronic structure (selected)* is different for HOMO, LUMO and gap, and is different for the ordinary least squares linear regression (LR) and for the machine learning analysis (RF, NN and kNN stand for random forests, neural networks and k-nearest neighbors, respectively). For the HOMO, it is:

LR, RF: Gap_{PBE}, AvgOcc, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})$, ..., $(\varepsilon_{HOMO} - \varepsilon_{HOMO-5})$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})^{-1}$, ..., $(\varepsilon_{HOMO} - \varepsilon_{HOMO-5})^{-1}$, $(\varepsilon_{LUMO} - \varepsilon_{HOMO-1})^{-1}$, $(\varepsilon_{LUMO} - \varepsilon_{HOMO-2})^{-1}$, $(\varepsilon_{LUMO+1} - \varepsilon_{HOMO})^{-1}$, $(\varepsilon_{LUMO+2} - \varepsilon_{HOMO})^{-1}$.

NN: Gap_{PBE}, AvgOcc, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})^{-1}$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-2})^{-1}$.

k-NN: AvgOcc, AvgEmpty.

For the LUMO, it is:

LR: Gap_{PBE}, AvgEmpty, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})$, $(\varepsilon_{LUMO+2} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$.

RF: Gap_{PBE}, Gap_{B3LYP}, AvgEmpty, $(\varepsilon_{LUMO+4} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+3} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+2} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$.

NN: Gap_{PBE}, AvgEmpty, AvgOcc, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})$, $(\varepsilon_{LUMO+2} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$.

k-NN: AvgOcc, AvgEmpty.

For the gap it is:

LR: Gap_{PBE}, AvgOcc, AvgEmpty, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})^{-1}$, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$.

RF: Gap_{PBE}, Gap_{B3LYP}, AvgEmpty, AvgOcc , $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})^{-1}$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-2})^{-1}$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-3})^{-1}$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-4})^{-1}$, $(\varepsilon_{LUMO+4} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+3} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+2} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$.

NN: Gap_{PBE}, Gap_{B3LYP}, AvgEmpty, AvgOcc, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})^{-1}$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-2})^{-1}$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-3})^{-1}$, $(\varepsilon_{LUMO+3} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+2} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$.

k-NN: AvgOcc, AvgEmpty.

Geometric features

These are the number of atoms of the fullerene ($N_{\text{⚛}}{}^3$), its number of hexagons ($N_{\hexagon}$), the area of the surface of the molecule, its volume and the quotient between both. The area and volume can be found in the output of the code which generates the fullerene topologies[4].

$N_{\text{⚛}}$
$N_{\hexagon}$
Area of the molecule
Volume of the molecule
Area / Volume

Phonon features

These are properties of the phonon frequencies, obtained from solving the dynamical equation[1,5]. The calculated frequencies of phonons in fullerenes typically lie between 200 and 1800 cm^{-1}.

$\min(\omega_\nu)$
$\max(\omega_\nu)$
$\langle\, \omega \,\rangle$
σ^2_ω
Skewness_ω
Kurtosis_ω

Bond length features

We denote bond lengths with B. We consider their average, variance, skewness and Kurtosis coefficients. In addition, we consider the 4 maximum and minimum values.

$\langle B \rangle$	highest B	lowest B
σ_B^2	2^{nd} highest B	2^{nd} lowest B
Skewness$_B$	3^{rd} highest B	3^{rd} lowest B
Kurtosis$_B$	4^{th} highest B	4^{th} lowest B

Bond orders calculations

All fullerenes were geometry optimized at the B3LYP level of theory, and triple-zeta with two polarization functions (TZ2P) basis set as implemented in ADF 2019[6]. The structures were optimized until the forces converged to less than $1{\cdot}10^{-4}$ Hartree Å^{-1} and the energy change between two consecutive steps was less than $1{\cdot}10^{-4}$ Hartree. The break condition for the electronic self-consistent loop was set to $1{\cdot}10^{-6}$ Hartree. A bond order analysis was performed on each optimized structure following the Nalewajski-Mrozek[7–11] approach. In this approach, there are three different definitions of the valence and bond order indices based on the way the density matrix is partitioned (referred to in the paper as the *bond order features NM1*, *NM2* and *NM3*). In addition, the Gopinathan-Jug[12] and Mayer[11] approaches (i.e. *GJ* and *Mayer*) have been utilized for comparison.

Bond orders - Mayer features

This feature set includes the bond length variables (listed in the previous section) together with the Mayer bond orders. For the latter we also consider the four statistics and the four maximum and minimum values.

lowest B	lowest O_M
2nd lowest B	2nd lowest O_M
3rd lowest B	3rd lowest O_M
4th lowest B	4th lowest O_M
highest B	highest O_M
2nd highest B	2nd highest O_M
3rd highest B	3rd highest O_M
4th highest B	4th highest O_M
$\langle B \rangle$	$\langle O_M \rangle$
σ_B^2	$\sigma_{O_M}^2$
Skewness$_B$	Skewness$_{O_M}$
Kurtosis$_B$	Kurtosis$_{O_M}$

Bond orders - GJ features

In this feature set we consider the bond length variables mentioned above as well as the GJ bond orders (statistics, 4 lowest values and 4 highest values). In addition, we consider the number of bonds ($N^{\rightharpoonup}$) and ratios of bonds ($R^{\rightharpoonup}$, i.e. the previous number divided by the total number of bonds of the molecule) whose GJ bond orders (O_{GJ}) are below 1.1 (i.e they are not hybridized) and whose bond orders are below 1.2, 1.3, 1.4 and 1.5, as well as above 1.5. We also consider the number and ratio of all the hybridized bonds ($N_{hyb}^{\rightharpoonup}$, $R_{hyb}^{\rightharpoonup}$). All the regressors of this feature set are listed below.

lowest B	lowest O_{GJ}	$N_{\overrightarrow{O_{GJ}}<1.1}$	$R_{\overrightarrow{O_{GJ}}<1.1}$
2nd lowest B	2nd lowest O_{GJ}	$N_{\overrightarrow{O_{GJ}}<1.2}$	$R_{\overrightarrow{O_{GJ}}<1.2}$
3rd lowest B	3rd lowest O_{GJ}	$N_{\overrightarrow{O_{GJ}}<1.3}$	$R_{\overrightarrow{O_{GJ}}<1.3}$
4th lowest B	4th lowest O_{GJ}	$N_{\overrightarrow{O_{GJ}}<1.4}$	$R_{\overrightarrow{O_{GJ}}<1.4}$
highest B	highest O_{GJ}	$N_{\overrightarrow{O_{GJ}}<1.5}$	$R_{\overrightarrow{O_{GJ}}<1.5}$
2nd highest B	2nd highest O_{GJ}	$N_{\overrightarrow{O_{GJ}}>1.5}$	$R_{\overrightarrow{O_{GJ}}>1.5}$
3rd highest B	3rd highest O_{GJ}	$N_{\overrightarrow{hyb}}$	$R_{\overrightarrow{hyb}}$
4th highest B	4th highest O_{GJ}		
$\langle B \rangle$	$\langle O_{GJ} \rangle$		
σ_B^2	$\sigma_{O_{GJ}}^2$		
Skewness$_B$	Skewness$_{O_{GJ}}$		
Kurtosis$_B$	Kurtosis$_{O_{GJ}}$		

Bond orders - NM1, NM2, NM3 features

These three feature sets consist in the same regressors listed in the section above (GJ), but using the data from the NN1, NM2 or NM3 bond orders (O_{NMX}).

lowest B	lowest O_{NMX}	$N_{\overrightarrow{O_{NMX}}<1.1}$	$R_{\overrightarrow{O_{NMX}}<1.1}$
2nd lowest B	2nd lowest O_{NMX}	$N_{\overrightarrow{O_{NMX}}<1.2}$	$R_{\overrightarrow{O_{NMX}}<1.2}$
3rd lowest B	3rd lowest O_{NMX}	$N_{\overrightarrow{O_{NMX}}<1.3}$	$R_{\overrightarrow{O_{NMX}}<1.3}$
4th lowest B	4th lowest O_{NMX}	$N_{\overrightarrow{O_{NMX}}<1.4}$	$R_{\overrightarrow{O_{NMX}}<1.4}$
highest B	highest O_{NMX}	$N_{\overrightarrow{O_{NMX}}<1.5}$	$R_{\overrightarrow{O_{NMX}}<1.5}$
2nd highest B	2nd highest O_{NMX}	$N_{\overrightarrow{O_{NMX}}>1.5}$	$R_{\overrightarrow{O_{NMX}}>1.5}$
3rd highest B	3rd highest O_{NMX}	$N_{\overrightarrow{hyb}}$	$R_{\overrightarrow{hyb}}$
4th highest B	4th highest O_{NMX}		
$\langle B \rangle$	$\langle O_{NMX} \rangle$		
σ_B^2	$\sigma_{O_{NMX}}^2$		
Skewness$_B$	Skewness$_{O_{NMX}}$		
Kurtosis$_B$	Kurtosis$_{O_{NMX}}$		

III. HOMO-LUMO GAPS

The difference between the gaps calculated using non-hybrids and hybrid xc functionals that we display in Tab. III is in line with results from the literature. This can be viewed in Ref. 13) for $C_{60} - I_h$ and in[14] for $C_{32} - D_3$ and $C_{36} - D_{6h}$. Note that DFT is known to underestimate gaps[13,15].

IV. FEATURE IMPORTANCE

In this section (Figs. 1 - 3) we present the results of Feature Importance analyses of our whole data using Random Forests. In order to collect the displayed data we trained the RF algorithm 100 times (using a different random seed each time). For each trained RF, we performed 10 different sets of permutations for the permutation-based analysis. For every displayed figure the set of regressors is the same that we presented in sec. III.B.1 of the main article.

Fullerene - Symmetry	HOMO - LUMO gap PBE (B3LYP) [meV]	Fullerene - Symmetry	HOMO - LUMO gap PBE (B3LYP) [meV]	Fullerene - Symmetry	HOMO - LUMO gap PBE (B3LYP) [meV]
C28-D2	487.0 (1495.2)	C56-C2v	523.7 (1264.8)	C84-D2	511.3 (1241.8)
C30-C2v-a	1206.9 (2200.9)	C56-Cs	848.1 (1727.6)	C84-D2d	1068.8 (1936.4)
C30-C2v-b	300.9 (1270.4)	C56-D2	650.5 (1521.9)	C84-D3d	408.5 (1232.8)
C32-C2	967.4 (1959.4)	C56-Td	882.9 (1707.2)	C84-D6h	1330.7 (2198.9)
C32-D3	1488.5 (2535.8)	C58-C1	236.0 (996.4)	C84-Td	1570.7 (2493.3)
C32-D3h	1375.6 (2541.5)	*C58-C3*	164.5 (909.1)	C86-C1	381.2 (1091.4)
C34-C2	520.3 (1420.6)	C58-Cs	200.0 (911.2)	C86-C2	620.0 (1408.6)
C36-C2	421.6 (1341.7)	C60-C1	752.2 (1611.5)	C86-C2v	357.1 (1017.1)
C36-D2	619.8 (1544.5)	C60-C2	523.3 (1389.4)	C86-D3	360.9 (964.3)
C36-D2d	408.0 (1290.1)	C60-C2v	871.6 (1795.6)	C88-C1	268.7 (927.6)
C36-D6h	492.7 (1159.1)	C60-C3v	1034.7 (1951.6)	C88-C2	90.8 (632.4)
C38-C2-a	755.4 (1675.0)	C60-Cs	952.0 (1852.0)	C88-T	111.2 (725.9)
C38-C2-b	828.2 (1662.7)	C60-D2h	554.0 (1446.7)	C90-C1-a	429.1 (1156.2)
C38-C2v	292.4 (1012.2)	C60-D5	143.4 (887.8)	**C90-C1-b**	**279.3 (987.7)**
C40-C1	535.9 (1441.0)	**C60-Ih**	**1592.5 (2634.6)**	C90-C2	715.4 (1535.6)
C40-D2	842.7 (1859.2)	C60-S4	381.8 (1134.8)	C90-C2v	822.0 (1645.1)
C40-D5d	961.1 (1983.4)	C62-C1	460.7 (1206.2)	C90-Cs	826.9 (1635.8)
C40-Td	417.3 (1352)	C62-C2	268.8 (931.7)	C92-C1	243.8 (859.7)
C42-C1	331.8 (1197.5)	*C62-C2v*	821.1 (1704.2)	C92-C2	442.4 (1099.5)
C42-C2	210.7 (1012.1)	C64-C2	1317.1 (1729.1)	C92-Cs	659.8 (1428.4)
C42-Cs	249.4 (1074.1)	C64-C3v	237.3 (1043.2)	C92-D2	630.6 (1420.0)
C42-D3	859.9 (1854.3)	C64-D2	1203.9 (2098.3)	C92-D2h	324.5 (1029.5)
C44-C1	498.0 (1317.4)	C66-C2v	307.3 (1020.2)	C92-S4	363.4 (995.2)
C44-C2	565.2 (1379.4)	C66-Cs	914.0 (1811.1)	C92-T	542.9 (1111.2)
C44-C2v	1099.3 (1955.0)	C68-C1	500.9 (1207.0)	C92-Td	528.5 (1158.5)
C44-D2	808.4 (1741.4)	C68-C2	916.3 (1744.3)	C94-C2	502.4 (1063.9)
C44-D3	234.6 (1011.9)	C68-D2	493.3 (1271.9)	C94-C2v	348.7 (1022.7)
C44-D3d	815.3 (1704.4)	C68-S4	709.0 (1411.6)	C96-C2	585.7 (1239.3)
C44-T	1150.4 (1990.3)	C68-S6	885.7 (1581.2)	C96-C2v	342.8 (972.1)
C46-C1	505.7 (1338.1)	C68-T	435.0 (1158.5)	C96-Cs	585.3 (1168.5)
C46-C2	509.8 (1244.3)	C70-C1	298.7 (968)	C96-D2h	1139.7 (1927.7)
C46-C2v	606.0 (1546.4)	C70-C2	784.3 (1538)	C96-D3h	1361.3 (2259.9)
C46-Cs	690.9 (1522.8)	**C70-D5h**	**1673.6 (2622.6)**	C96-D6h	601.5 (1409.3)
C48-C1	598.0 (1451.9)	C72-C2v	677.7 (1459.2)	C98-C1-a	294.7 (923.4)
C48-C2-a	740.9 (1475.8)	C72-D6d	1376.5 (2326.8)	C98-C1-b	643.8 (1199.2)
C48-C2-b	1007.3 (1923.9)	C74-C1	263.3 (887.8)	C98-C2	698.9 (1286.1)
C48-Cs	598.5 (1490.0)	C74-C2	581 (1364.1)	C98-C2v	411.9 (989.3)
C48-D2	719.7 (1634.1)	C74-Cs	389.5 (1133.1)	C98-C3	806.5 (1385.7)
C48-D2h	596.5 (1244.3)	C76-C1	451.9 (1100.9)	C98-D3	382.1 (867.8)
C48-D6d	713.6 (1732.9)	C76-C2	307.9 (944.4)	C100-C2	441.9 (995.7)
C50-C1-a	382.9 (1215.7)	C76-C3v	140.9 (813.1)	C100-C2v	950.2 (1747.7)
C50-C1-b	679.7 (1551.2)	C76-D2	1038.9 (1855.2)	C100-Cs	251.8 (834.0)
C50-C2	722.4 (1543.5)	C76-S4	519.9 (1288.9)	C100-D2	380.6 (1077.0)
C50-Cs	555.2 (1494.1)	C78-C2v	750.8 (1545.1)	C100-S4	347.3 (1069.8)
C50-D3	1345.7 (2230.3)	C78-D3	712.9 (1505.2)	C100-T	121.6 (721.9)
C50-D5h	454.8 (1193.9)	C78-D3h	589.0 (1406.9)	C100-Td	245.8 (957.2)
C52-C1	433.3 (1242.0)	C80-C2v	260.9 (848.8)	C104-C1	338.7 (967.0)
C52-C2	327.3 (1062.3)	C80-D2	438.1 (1217.6)	C104-S4	378.5 (1099.5)
C52-Cs	443.3 (1302.2)	C80-Ih	139.7 (781.6)	C104-T	750.1 (1357.7)
C52-D2	354.4 (1115.8)	C80-S4	417.2 (1148.7)	C180-Ih	1436.6 (2244.0)
C52-D2d	379.3 (1080.2)	C82-C2	444.7 (1179.9)		
C54-C1	379.7 (1144.0)	C82-C3v	103.4 (770.0)		
C54-C2	279.5 (990.8)	C82-Cs	680.4 (1451.5)	**PC60BM**	**1479.9 (2475.5)**
C56-C1	261.6 (1051.2)	C84-C2v	414.2 (1172.3)	**PC70BM**	**1568.3 (2444.0)**
C56-C2	690.5 (1480.2)	C84-Cs	655.3 (1426.5)	**IC60BA**	**1489.6 (2489.4)**

TABLE III: HOMO-LUMO gap obtained with PBE and B3LYP the fullerenes and derivatives investigated.

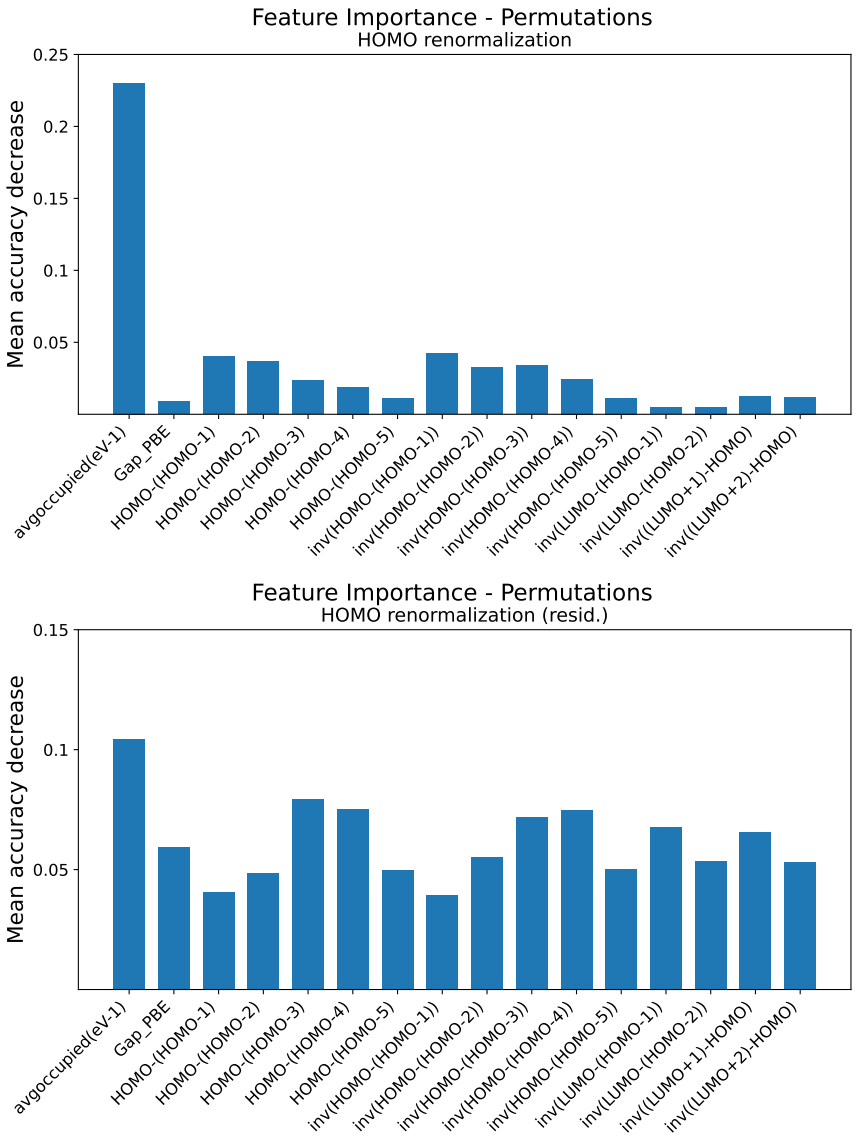

FIG. 1: Importance of the different regressors (features) in the ML-Random Forests analysis of the HOMO renormalization. Top: ML applied to the renormalizations; Bottom: ML applied to the residuals of the renormalizations (after having applied linear regression).

Note that, for the sake of clarity, in Figs. 1 to 3 we have not followed the usual convention of ordering features by decreasing importance, but gathered related variables (e.g. all (HOMO-(HOMO-n)) together) and presented them as given by their indices (e.g. first (HOMO-(HOMO-1)), then (HOMO-(HOMO-2)), etc.).

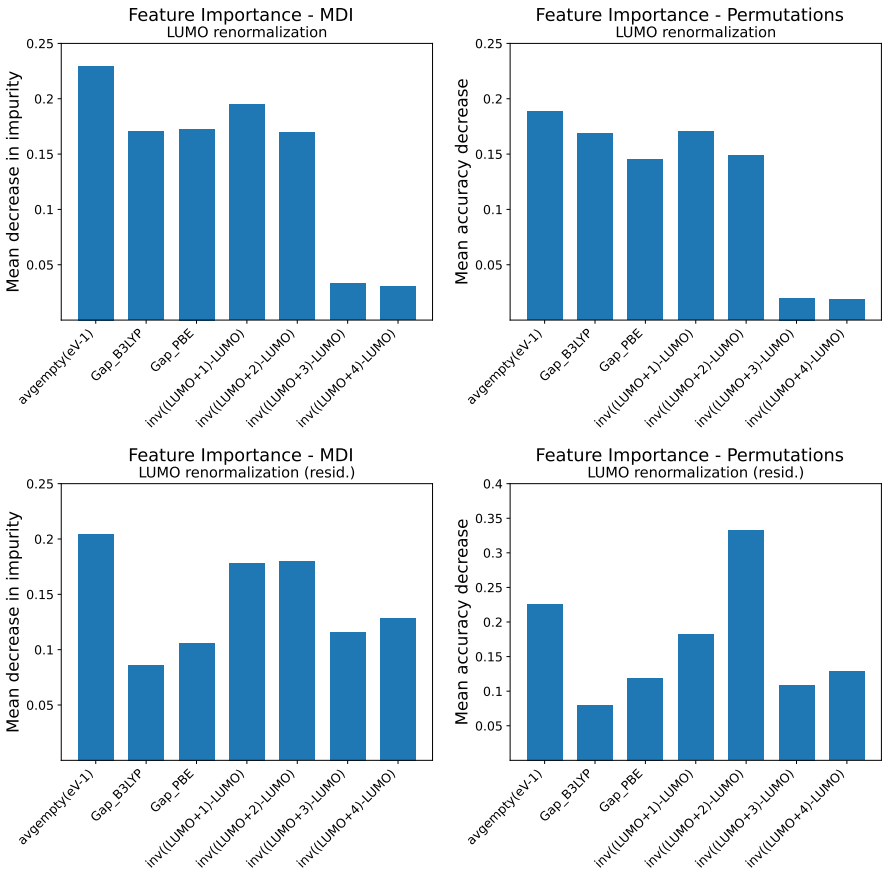

FIG. 2: Importance of the different regressors (features) in the ML-Random Forests analysis of the LUMO renormalization. Top: ML applied to the renormalizations; Bottom: ML applied to the residuals of the renormalizations (after having applied linear regression). Left: Mean Decrease in Impurity; Right: From permutations.

In Fig. 1 we notice that the most significant regressor is AvgOcc. Its importance is specially remarkable if ML is applied on a standalone basis (not on top of linear regression). In that case (Fig. 1-top), energy differences of HOMO with other electronic states gradually lose importance for farther-lying states, as one could expect (more distant electrons are expected to have a lower influence on the HOMO). When ML is applied on top of linear regression (Fig. 1-bottom), the importance of the different regressors becomes more similar in size. Despite the clear relationship between renormalizations and gaps, as displayed in Fig. 4 (see next section), Fig. 1 indicates that the importance of this feature is relatively low.

The average of the eigenvalues of the empty states is the most important feature for the LUMO renormalization (Fig. 2-top). However, its importance is lower than that of AvgOcc for the HOMO renormalization.

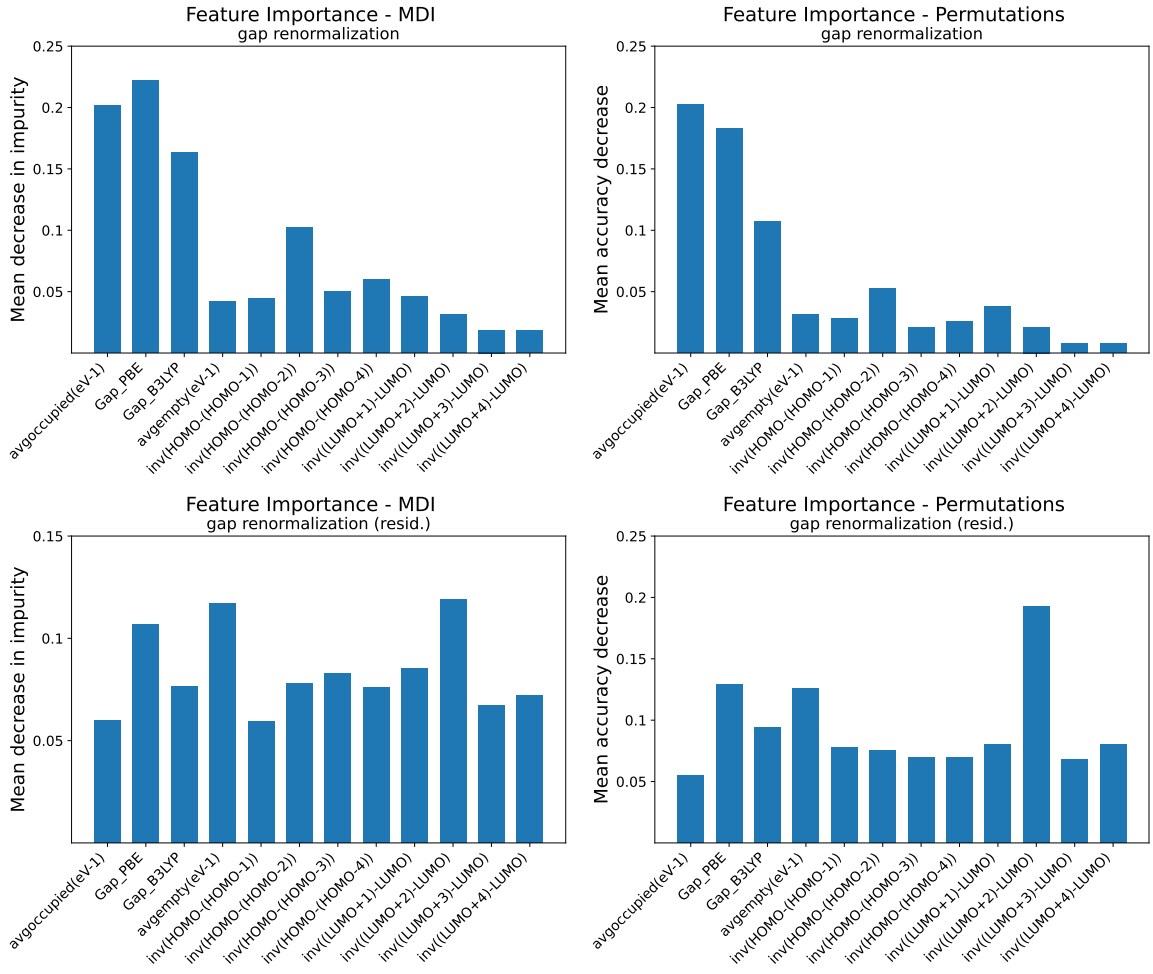

FIG. 3: Importance of the different regressors (features) in the ML-Random Forests analysis of the HOMO-LUMO gap renormalization. Top: ML applied to the renormalizations; Bottom: ML applied to the residuals of the renormalizations (after having applied linear regression). Left: Mean Decrease in Impurity; Right: From permutations.

Figs. 2-top and 3-top indicate that the gap is an important feature for LUMO and gap renormalizations, in contrast with the HOMO renormalization.

Figs. 1, 3 indicate that the average of the inverse eigenvalues for occupied states (avgoccupied(eV-1)) has a strong predictive power. Moreover, we notice that its importances are higher when the ML algorithm is forecasting the whole renormalization instead of forecasting the residual of the renormalization after a linear regression. This indicates that the relationship between this feature and the renormalizations is mostly linear.

V. LINEAR REGRESSION

In the main article, we displayed the renormalizations of the HOMO-LUMO gaps as a function of the gap calculated with B3LYP. As a complement we present fittings of the renormalizations as a function of the gap from GGA-PBE (figs. 4, 5).

Fig. 4 indicates that the gap renormalizations lie between 2% and 18% of the PBE gap if the fullerenes with T-symmetry are excluded. Figs. 4 and 5 also indicates that the renormalizations of phonon derivatives are larger than the renormalizations of most of the bare fullerenes.

For illustrative purposes we perform the fitting of the clouds of points of the renormalizations vs gaps with univariate linear regressions. To this, we take the fullerenes displayed in tab. 2 of the main article (excluding the two outliers which were also excluded in the analysis based on decision trees, i.e. C_{28} and $C_{58}-C_3$, as well as the fullerene which contains a ring made of 4 carbons and the fullerene derivatives). We fit the renormalizations of HOMO, LUMO and gap ($r_\Delta^{fit,HOMO}$, $r_\Delta^{fit,LUMO}$ and $r_\Delta^{fit,gap}$) as a function of the gaps calculated with PBE and B3LYP ($\Delta^{PBE}, \Delta^{B3LYP}$). The results (in meV) are:

$$
\begin{aligned}
r_\Delta^{fit,HOMO} &= +0.027255 \cdot \Delta^{B3LYP} - 14.4\,; \quad R^2 = 0.18 \\
r_\Delta^{fit,LUMO} &= -0.029642 \cdot \Delta^{B3LYP} + 15.1\,; \quad R^2 = 0.35 \\
r_\Delta^{fit,gap} &= -0.056897 \cdot \Delta^{B3LYP} + 29.5\,; \quad R^2 = 0.49
\end{aligned}
$$

$$
\begin{aligned}
r_\Delta^{fit,HOMO} &= +0.036848 \cdot \Delta^{PBE} + 0.9\,; \quad R^2 = 0.21 \\
r_\Delta^{fit,LUMO} &= -0.037204 \cdot \Delta^{PBE} + 3.3\,; \quad R^2 = 0.36 \\
r_\Delta^{fit,gap} &= -0.074053 \cdot \Delta^{PBE} - 4.3\,; \quad R^2 = 0.54
\end{aligned}
$$

The coefficients of determination of LUMO and gap renormalizations improve if the fullerenes with T and I symmetries are also excluded:

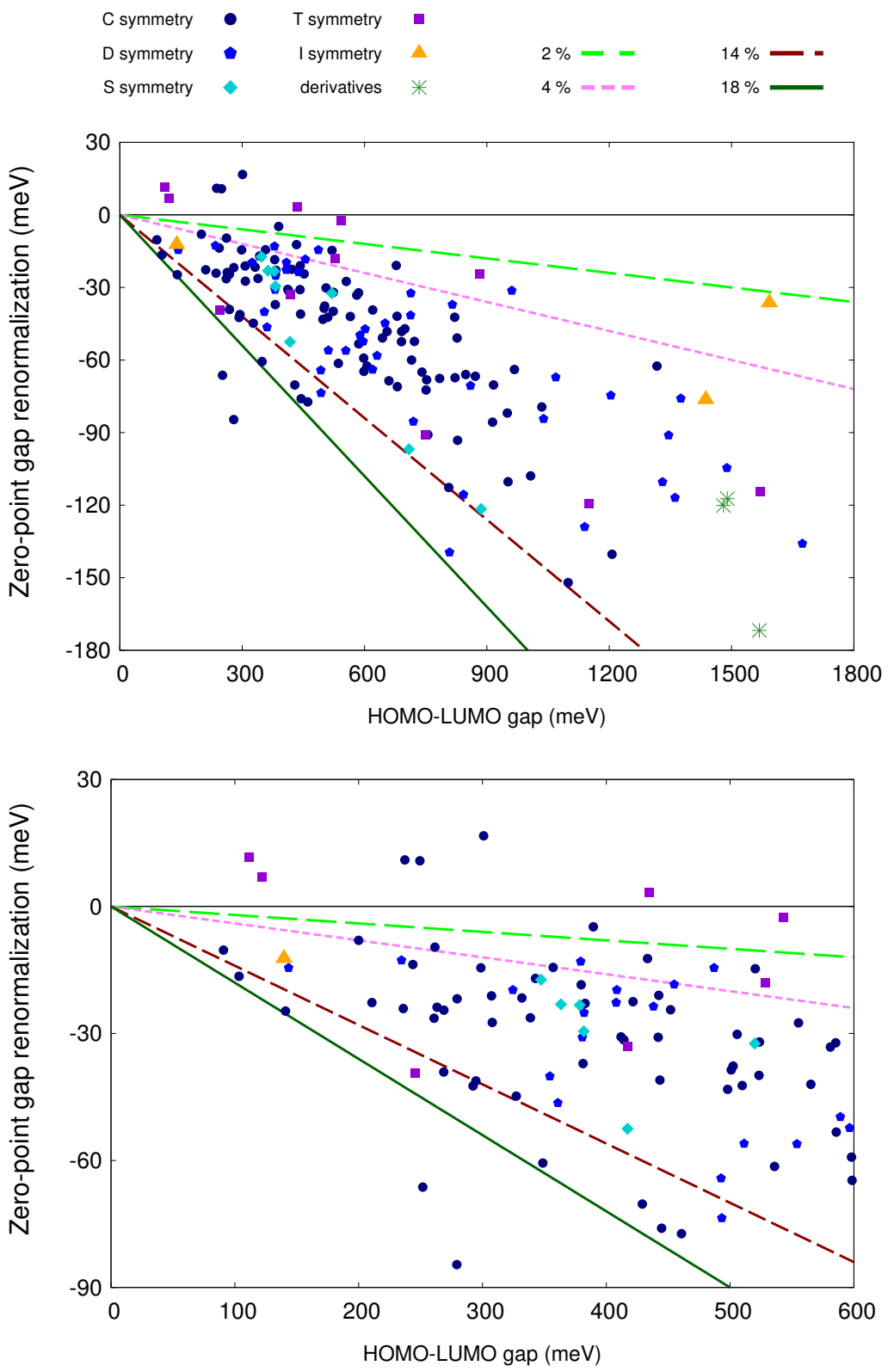

FIG. 4: Frozen-phonon renormalization of the HOMO-LUMO gap of fullerenes as a function of the gap from a ground state calculation (all calculations using GGA-PBE). The bottom figure is a zoom of the top figure.

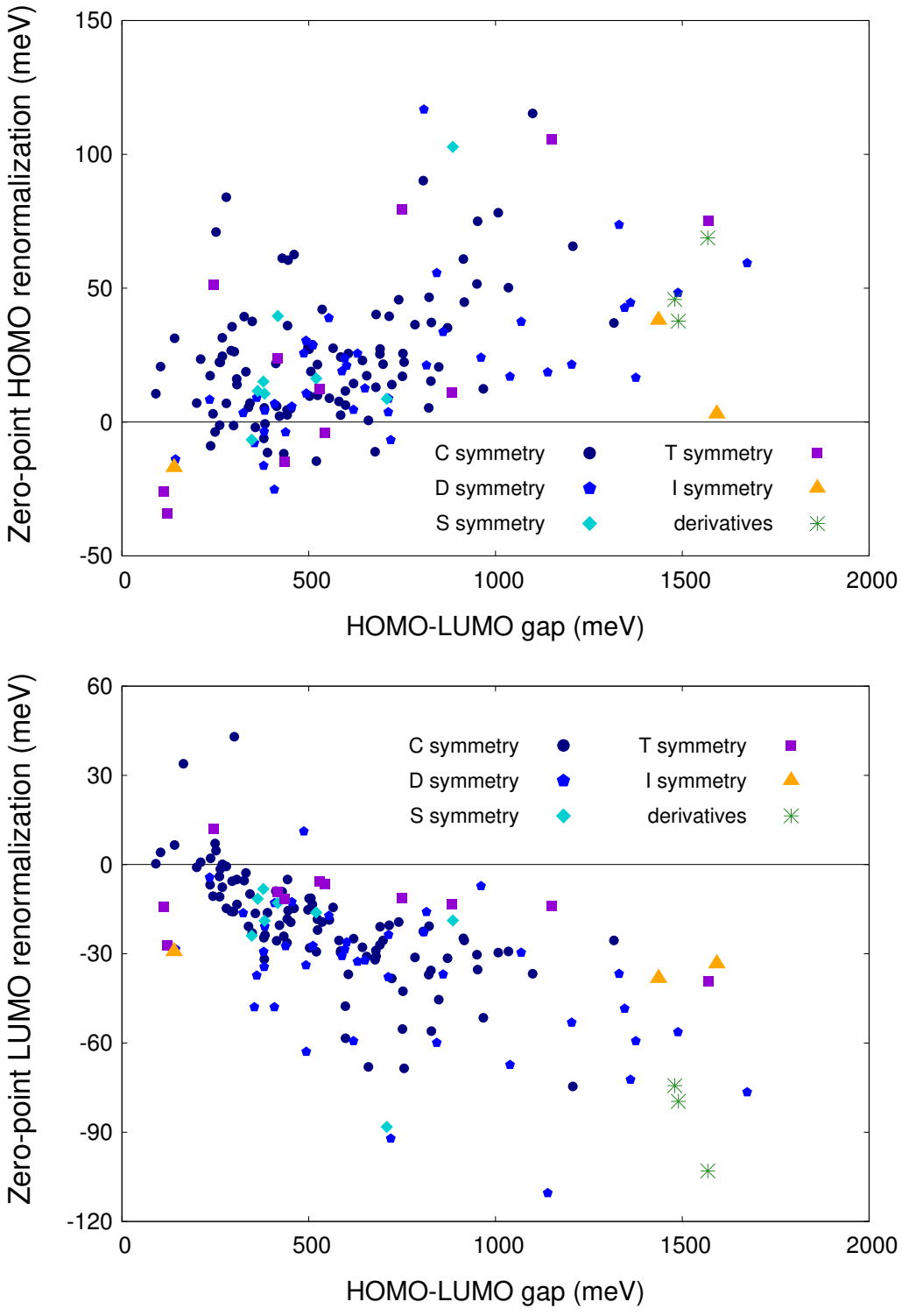

FIG. 5: Frozen-phonon renormalization of the HOMO (top) and LUMO (bottom) of fullerenes as a function of the gap (all calculations using GGA-PBE).

$$r_{\Delta}^{fit,HOMO} = +0.025556 \cdot \Delta^{B3LYP} - 11.7\,; \quad R^2 = 0.16$$
$$r_{\Delta}^{fit,LUMO} = -0.034571 \cdot \Delta^{B3LYP} + 20.8\,; \quad R^2 = 0.40$$
$$r_{\Delta}^{fit,gap} = -0.060127 \cdot \Delta^{B3LYP} + 32.5\,; \quad R^2 = 0.52$$

$$r_{\Delta}^{fit,HOMO} = +0.035582 \cdot \Delta^{PBE} + 2.2\,; \quad R^2 = 0.20$$
$$r_{\Delta}^{fit,LUMO} = -0.044509 \cdot \Delta^{PBE} - 0.2\,; \quad R^2 = 0.42$$
$$r_{\Delta}^{fit,gap} = -0.080091 \cdot \Delta^{PBE} - 2.4\,; \quad R^2 = 0.58$$

In fig. 6 we show two examples of the gaps and gap renormalizations for two isomers (the fullerenes with 60 and 84 carbon atoms; we have omitted the buckminsterfullerene because it has I symmetry). An approximately linear trend can be noticed.

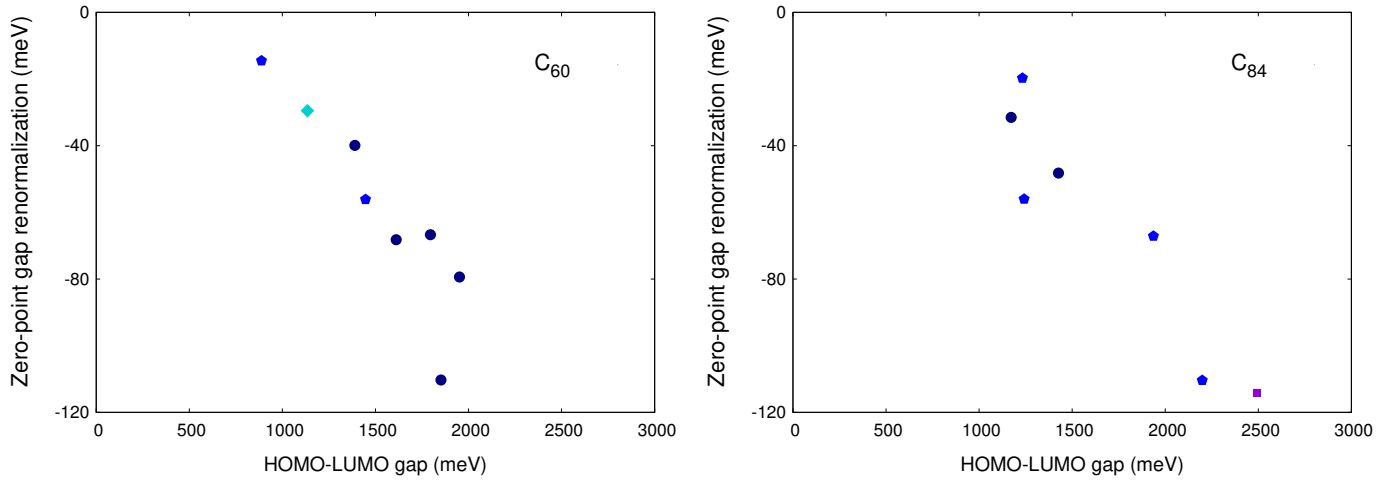

FIG. 6: Frozen-phonon renormalization (from PBE) of the HOMO-LUMO gap of C_{60} (top) and C_{84} (bottom) fullerenes as a function of the HOMO-LUMO gap (from B3LYP).

VI. REGRESSION USING EQUAL REGRESSORS FOR LINEAR AND NONLINEAR REGRESSIONS

The table with the average errors of the ML methods that we presented in the main article used different sets of regressors for the linear regression (LR) and the machine learning methods (KNN, NN, RF), with the only exception of HOMO-RF. Therefore one may wonder whether the accuracy gained through machine learning methods was not indeed due to the more sophisticated

calculation mode, but to the mere usage of difference (e.g. more) regressors. In order to provide some insight to answer this question we present the table IV. For the HOMO renormalization, the list of regressors is the one presented in the main article for LR and RF. For the LUMO renormalization, the list of regressors consists in Gap_{PBE}, AvgEmpty, AvgOccupied, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+2} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})$. For the gap renormalization, the list of regressors consists in Gap_{PBE}, AvgEmpty, AvgOccupied, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})^{-1}$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-2})^{-1}$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-3})^{-1}$, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+2} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+3} - \varepsilon_{LUMO})^{-1}$.

Tab. IV indicates that the RF algorithm itself does improve the forecast with respect to LR. For NN and KNN –for the chosen sets of regressors– this does not hold. Nevertheless, note that the use of many regressors makes the results of the KNN solver less accurate (the optimal set of regressors to apply KNN on top of LR being the AvgEmpty, AvgOccupied pair). The NN seems especially sensitive to the noise introduced by not-carefully-selected regressors.

Renorm. of	LR	KNN @LR	NN @LR	RF @LR
HOMO	5.80	5.80	5.81	5.56
LUMO	6.77	6.92	6.67	6.12
Gap	8.80	8.81	8.81	8.16

TABLE IV: Results of the regression tests of the renormalizations using linear regression (LR) and k-nearest neighbours, neural networks and random forests on top of linear regression (KNN@LR, NN@LR and RF@LR, respectively). For each quantity (HOMO, LUMO, gap) the set of regressors was equal for LR and KNN@LR, NN@LR, RF@LR. The numbers indicate the average absolute error in the test datasets measured in meV.

VII. REGRESSION USING NEURAL NETWORKS AND K-NEAREST NEIGHBOURS

In Figs. 7, 8 below we present the renormalizations of HOMO, LUMO and gap calculated ab initio (x axis) and forecasted using linear regression and applying ML methods (neural networks and k-nearest neighbours) on top of it (y axis). These graphs are analogous to those presented in the main paper. This is, the plots of the left column correspond to the *electronic structure selected* feature set, while the plots of the right column correspond to 11 different feature sets. The calculation methods for the results displayed in Figs. 7 , 8 are those described in the main article.

Though Figs. 7, 8 reveal larger errors for the presented methods than the errors of the random forests method presented in the main article, the forecasts are still reasonable. Nevertheless, this is mainly due to the forecast of the linear regression. The feature sets which do not include *elec-*

tronic structure information are also inaccurate using neural networks and k-nearest neighbours. We can therefore conclude that these features are not useful for forecasting renormalizations due to electron-phonon interaction in fullerenes.

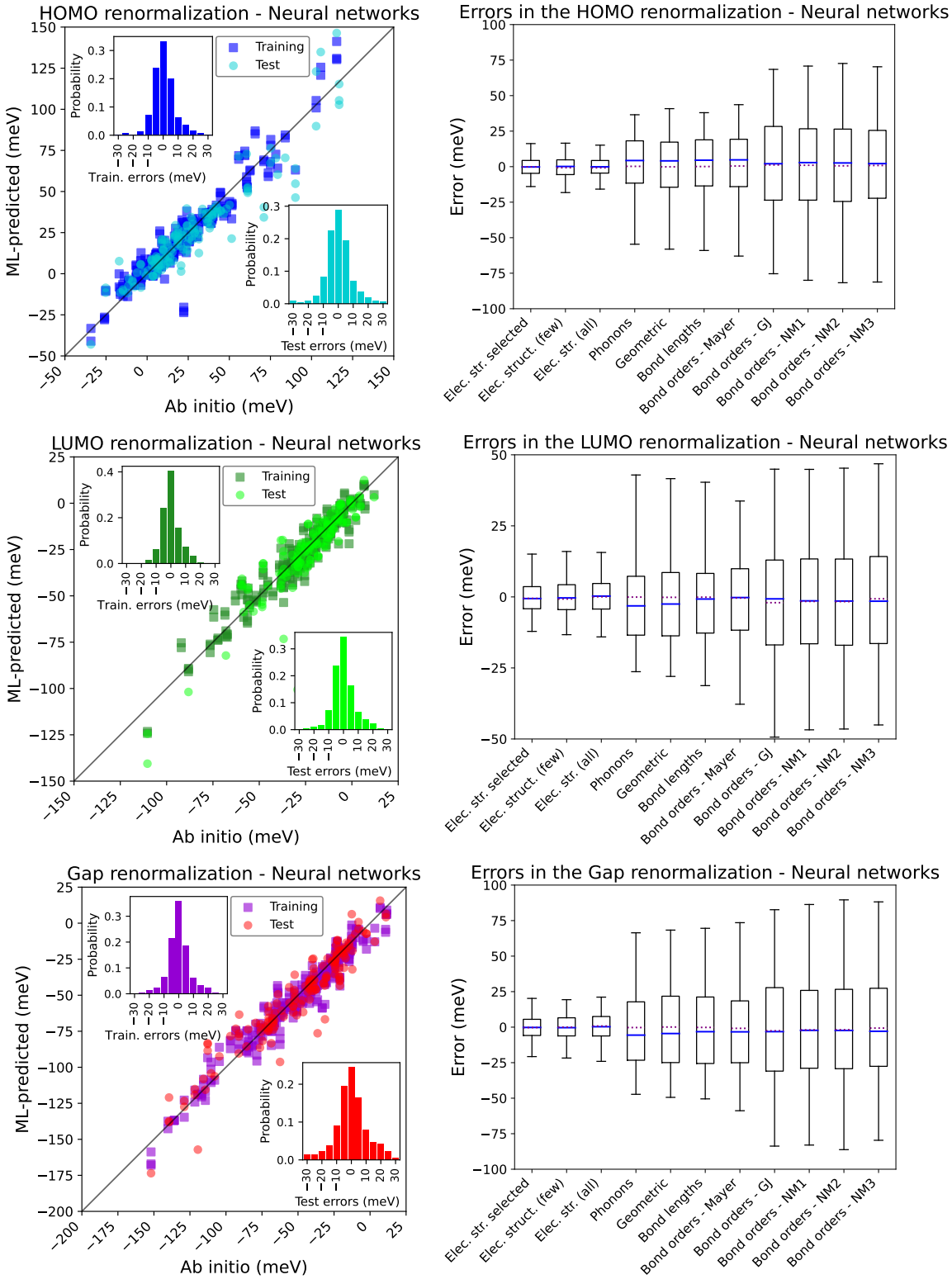

FIG. 7: Predictions using neural networks on top of linear regression. Left (selected feature set): Scatter plots: Predicted vs ab initio; Histograms: training and test errors. Right: Test errors of predictions using different feature sets. Top: HOMO; Center: LUMO; Bottom: gap.

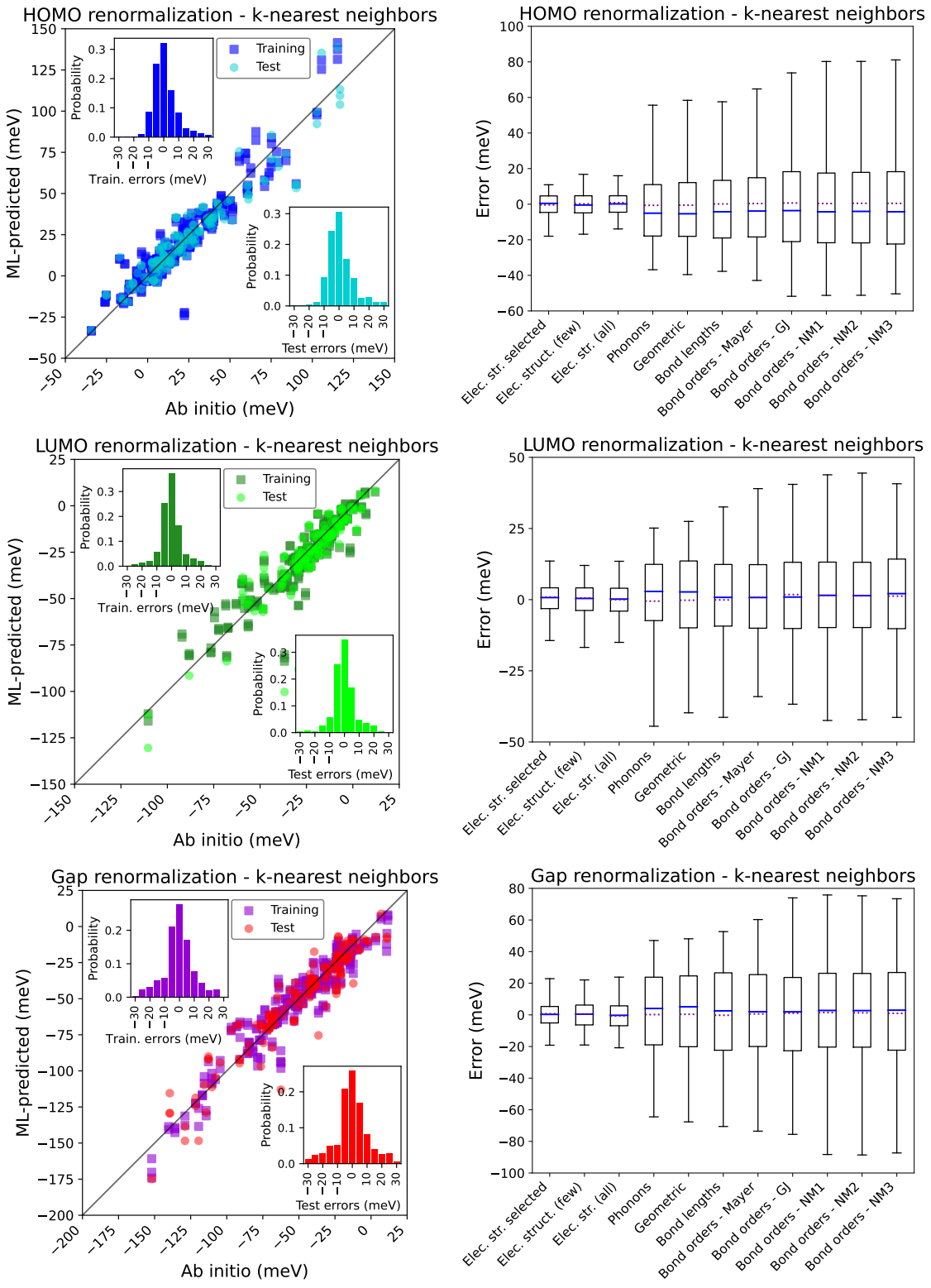

FIG. 8: Predictions using k-nearest neighbours on top of linear regression. Left (selected feature set): Scatter plots: Predicted vs ab initio; Histograms: training and test errors. Right: Test errors of predictions using different feature sets. Top: HOMO; Center: LUMO; Bottom: gap.

VIII. REGRESSION USING MIXED SETS OF REGRESSORS

In the main article we presented two sections with results of i) sheer regression, ii) ML-based classification performed on top of linear regression. The calculations of i) and ii) were based on different sets of regressors. For the sake of completeness, in this section we display the calculations of regression (i, carried out as indicated in sec. II.B.1 of the main article) performed using the regressors used for the LR of ii (sec. II.B.2 of the main article). This is, the regressors for LR are:

- For the HOMO renormalization: $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-2})$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})^{-1}$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-2})^{-1}$ and AvgOcc.
- For the LUMO renormalization: An ensemble of two regression models whose regressors are, respectively:
  - $\{Gap_{PBE}$, AvgEmpty, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-1})$, $(\varepsilon_{HOMO} - \varepsilon_{HOMO-2})\}$
  - $\{Gap_{PBE}$, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}$, $(\varepsilon_{LUMO+2} - \varepsilon_{LUMO})^{-1}\}$
- For the Gap renormalization:
  - $\{Gap_{PBE}$, AvgOcc, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}\}$
  - $\{Gap_{B3LYP}$, AvgOcc, $(\varepsilon_{LUMO+1} - \varepsilon_{LUMO})^{-1}\}$

In the cases where there is an ensemble the value outputted by the linear regression (LR) is simply the average between the quantity calculated with each of the two sets of regressors.

The regressors for the ML-based calculations are those listed in sec. II of this document.

The results are displayed in Tab. V. They differ very slightly from the results presented in Tab. II of the main article for the renormalizations of LUMO and gap; the difference is higher for the renormalization of the HOMO.

IX. DECISION TREES ON TOP OF ML-BASED REGRESSION

In the main paper (sec. III.B.2) we presented the results of the analysis of the effect of decision trees to discard unreliable forecasts made by linear regression (LR). For the sake of completeness, in this section we perform the same analysis using regression from machine learning methods instead. Our results are presented in Tab. VI. Its columns represent the same quantities as their counterparties of the table which appears in the main article.

Renorm. of	LR	KNN	KNN @LR	NN	NN @LR	RF	RF @LR
HOMO	6.60	11.58	6.22	8.04	6.49	7.62	6.53
LUMO	6.51	11.69	6.35	7.17	6.57	6.95	5.52
Gap	9.19	17.40	7.81	14.02	8.66	10.48	7.69

TABLE V: Results of the regression tests of the renormalizations using linear regression (LR), k-nearest neighbours (KNN), neural networks (NN) and random forests (RF), as well as using the ML methods on top of linear regression (KNN@LR, NN@LR and RF@LR). The regressors for LR are those used in the main article in the section on classification; the regressors for ML are those used in the main article for sheer regression. The numbers indicate the average absolute error in the test datasets measured in meV.

Quantity	ML meth.	$\mathfrak{E}_{LR}$ (%)	$\mathfrak{E}_{ML@LR}$ (%)	$\mathfrak{E}_{tree_1}$ (%)	Marked$_1$ (%)	$\mathfrak{E}_{marked_1}$ (%)	$\mathfrak{E}_{tree_2}$ (%)	Marked$_2$ (%)	$\mathfrak{E}_{marked_2}$ (%)
HOMO	RF	14.37	13.50	12.45	6.87	38.72	12.44	5.43	22.38
LUMO	RF	17.95	12.34	10.88	3.387	36.67	10.97	5.08	52.28
Gap	RF	32.48	25.41	24.56	2.34	28.33	22.64	2.77	16.67
HOMO	NN	14.37	15.64	14.78	2.99	36.37	14.17	16.54	5.89
LUMO	NN	17.95	18.37	15.33	5.34	27.40	12.40	17.48	63.92
Gap	NN	32.48	29.93	24.18	43.09	41.19	23.33	11.14	59.73
HOMO	KNN	14.37	13.70	12.21	2.96	15.87	11.99	5.93	28.09
LUMO	KNN	17.95	17.73	14.12	8.24	51.11	13.49	8.23	38.33
Gap	KNN	32.48	28.42	17.43	24.47	55.42	18.93	18.86	43.52

TABLE VI: Summary of results for all the renormalizations using a regression model (or an ensemble in the case of the LUMO and gap renormalizations) and then a regression tree to assess the acceptability of the result.

X. ANALYSED FULLERENES

In figs. 9 and following we present plots with the aspect of the simulated systems. They were made with `Avogadro`[16]. The corresponding geometry files can be downloaded in [17]. The geometries were obtained from the literature[18–25] or were generated using the code of [4].

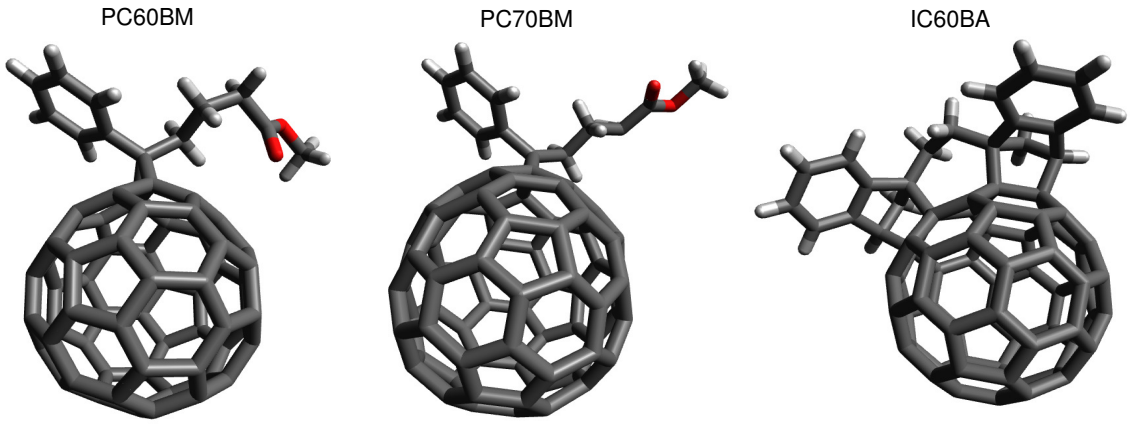

FIG. 9: Fullerene derivatives analysed in this work.

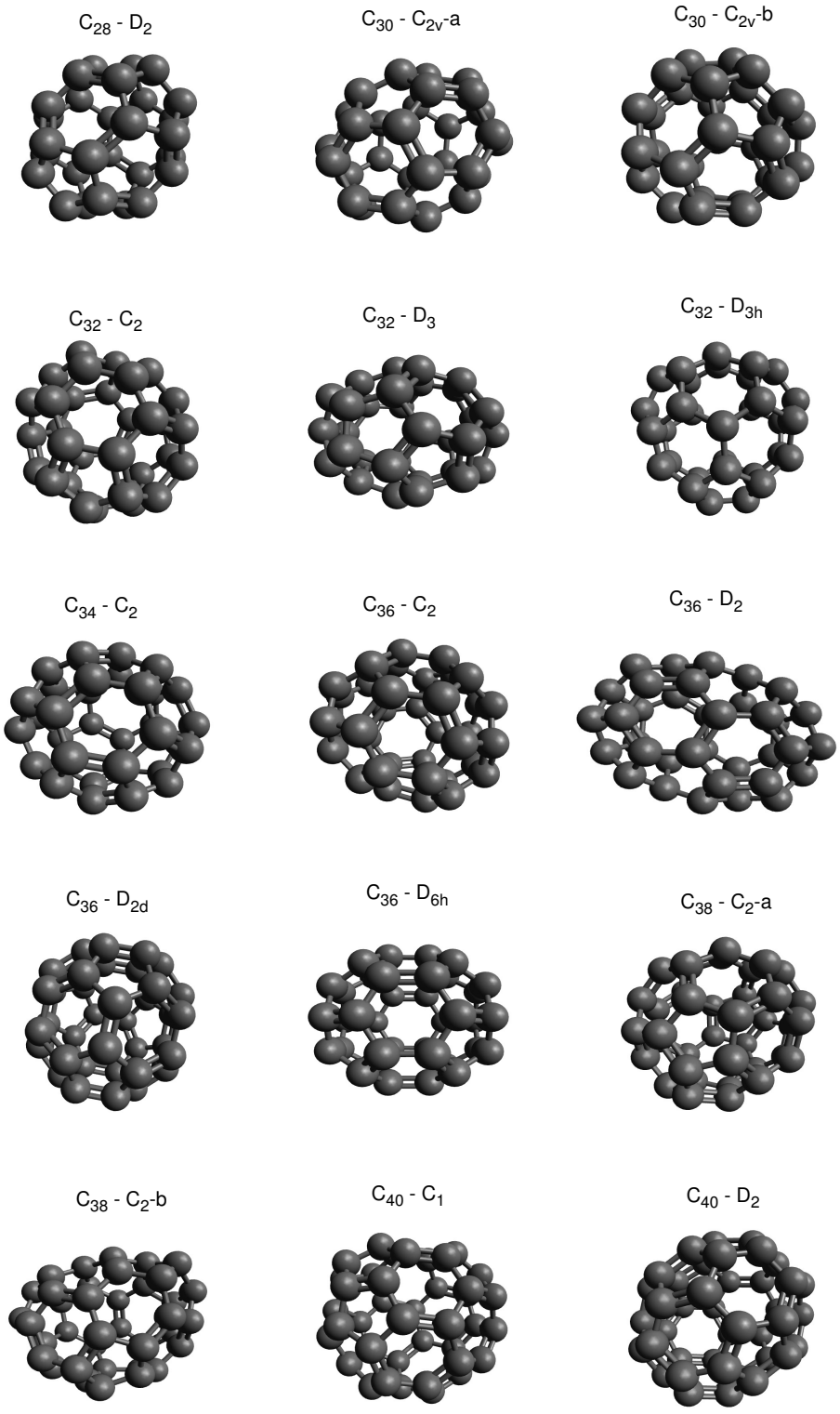

FIG. 10: Fullerenes analysed in this work.

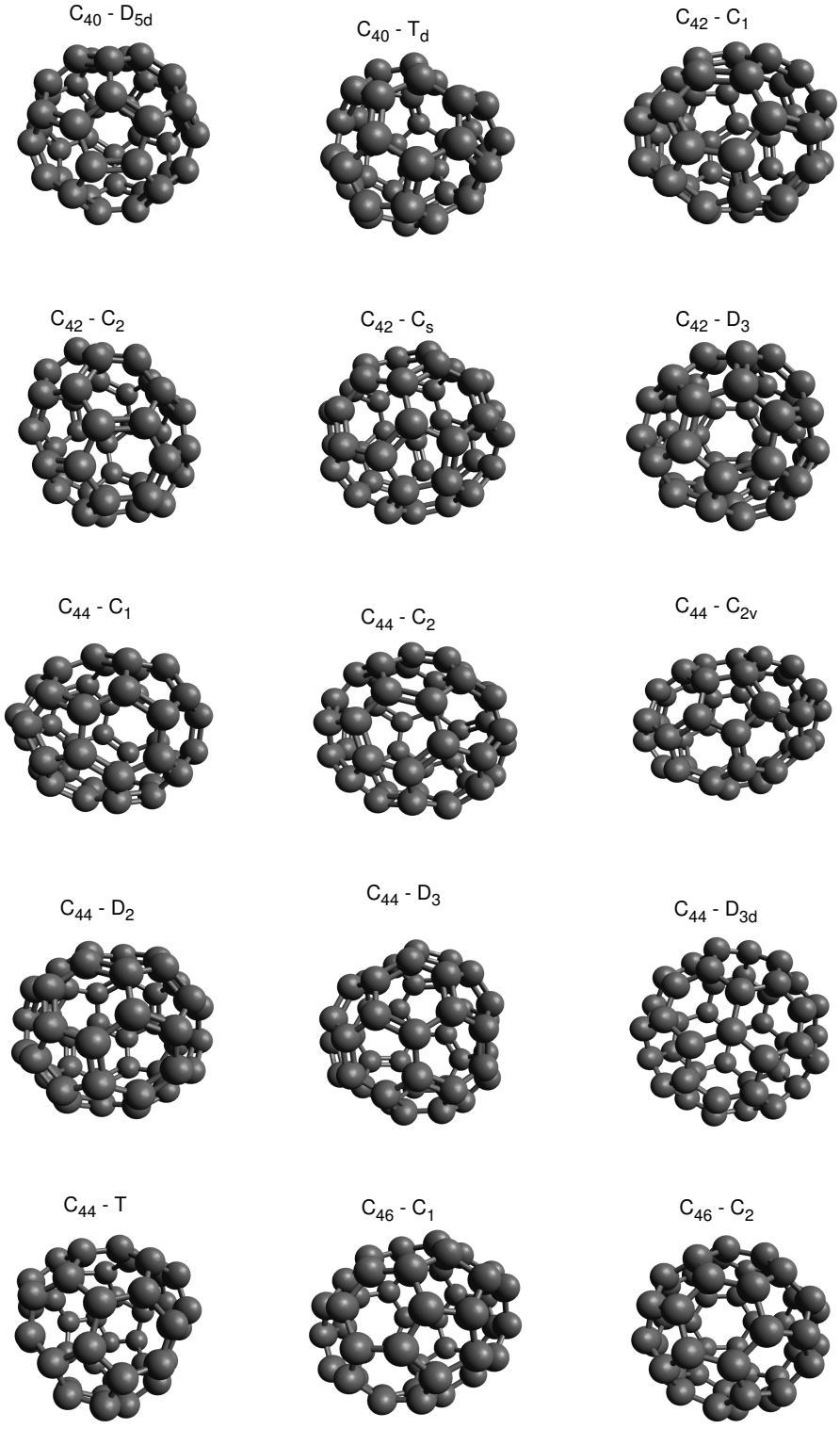

FIG. 11: Fullerenes analysed in this work.

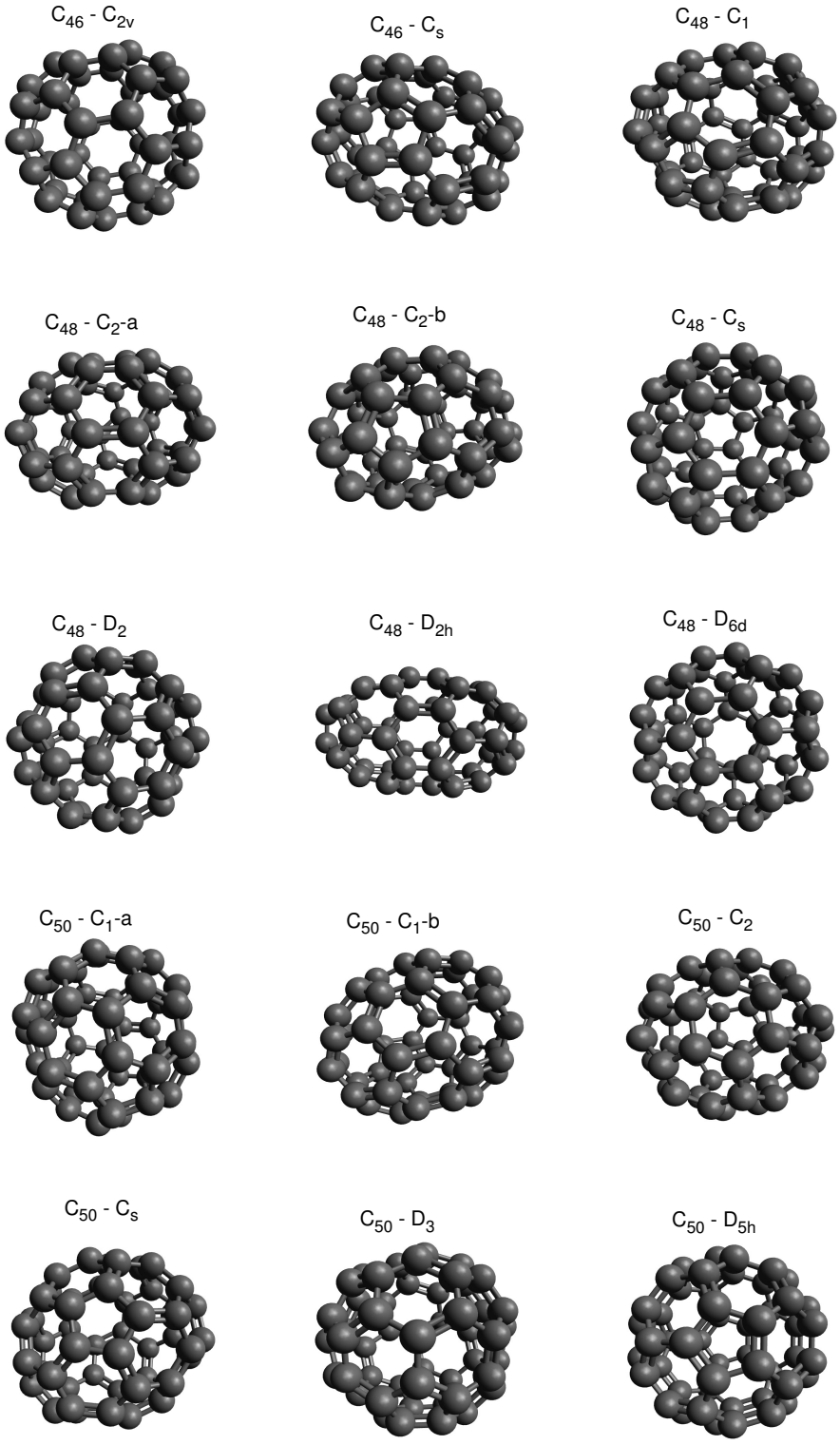

FIG. 12: Fullerenes analysed in this work.

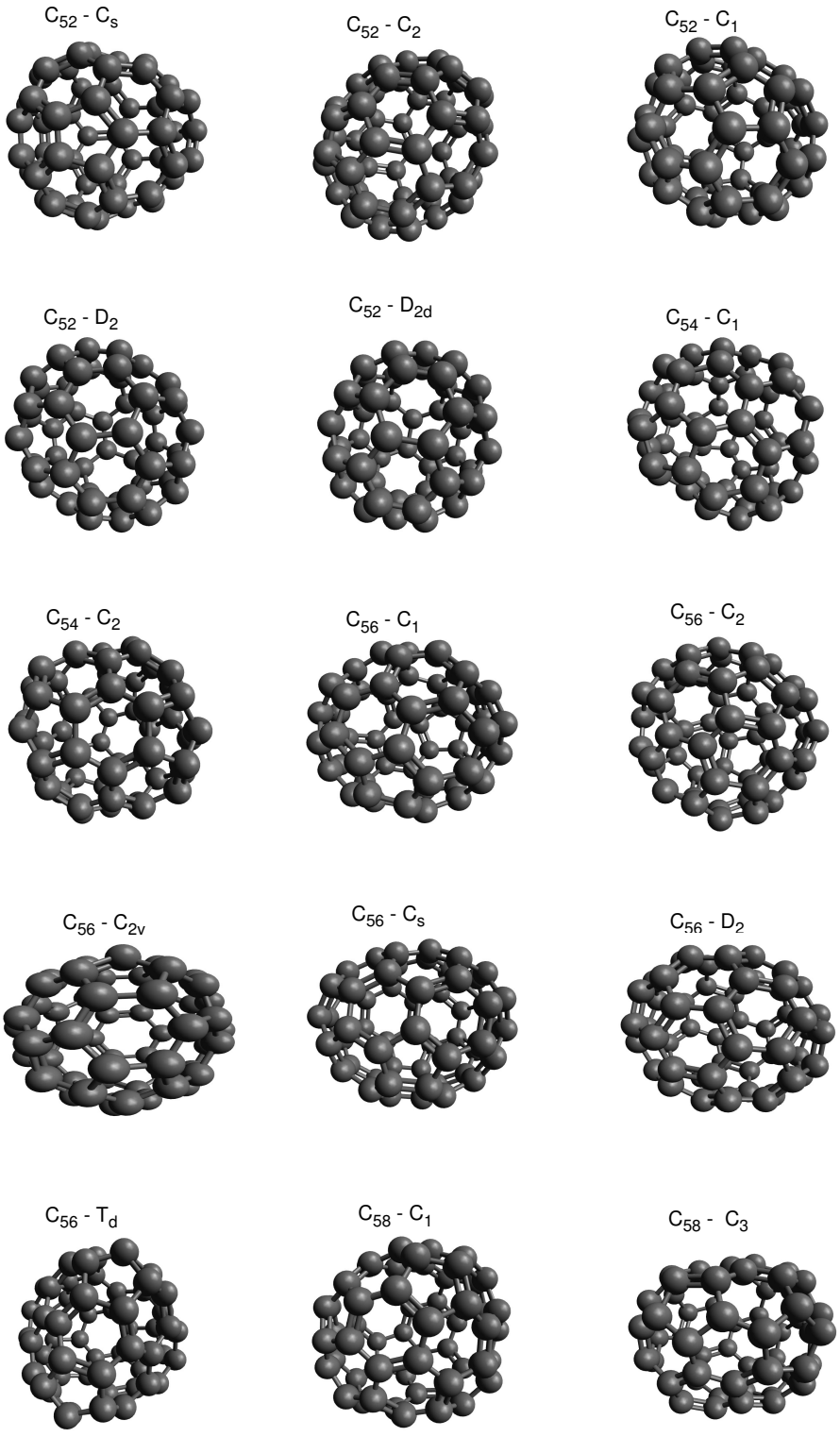

FIG. 13: Fullerenes analysed in this work.

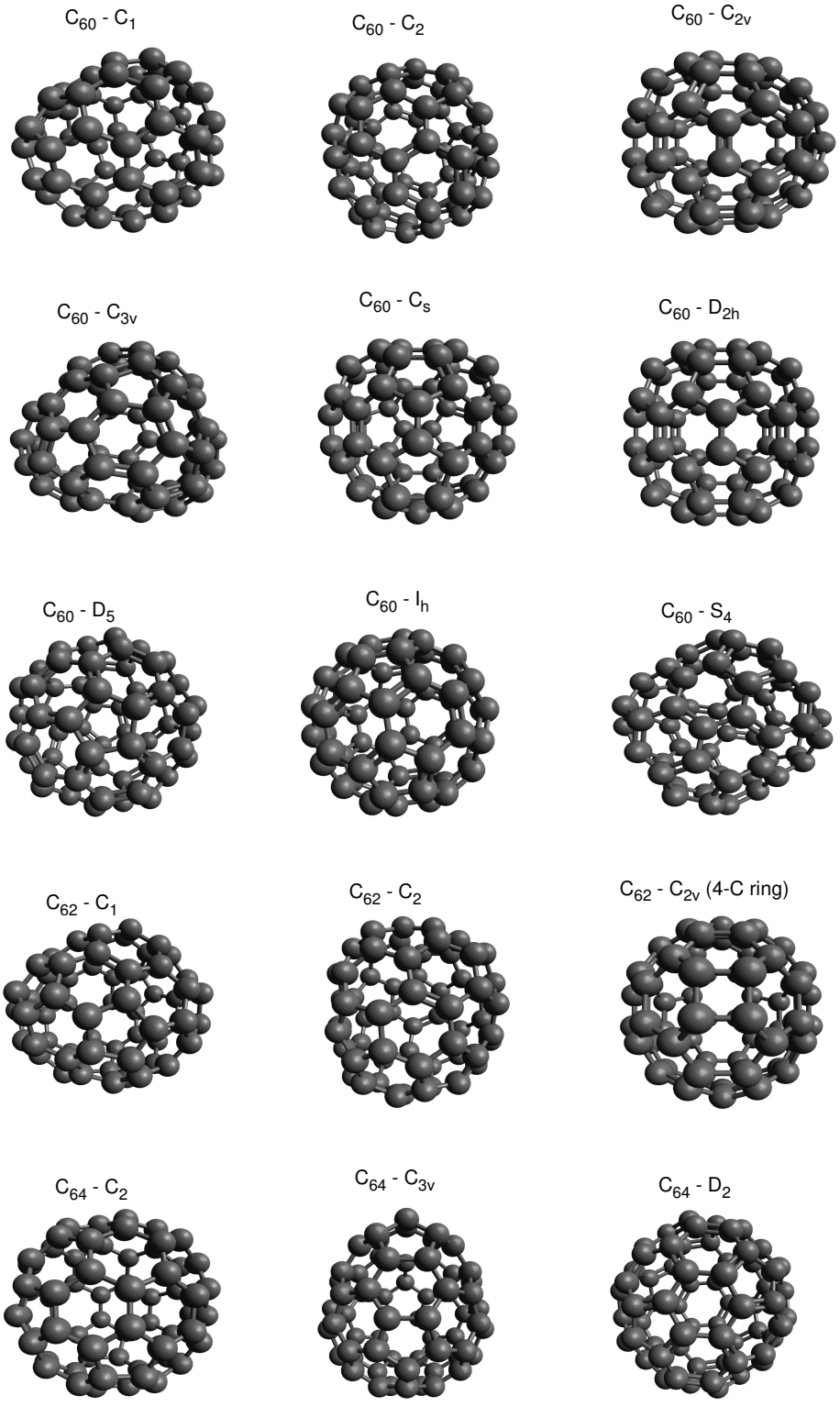

FIG. 14: Fullerenes analysed in this work.

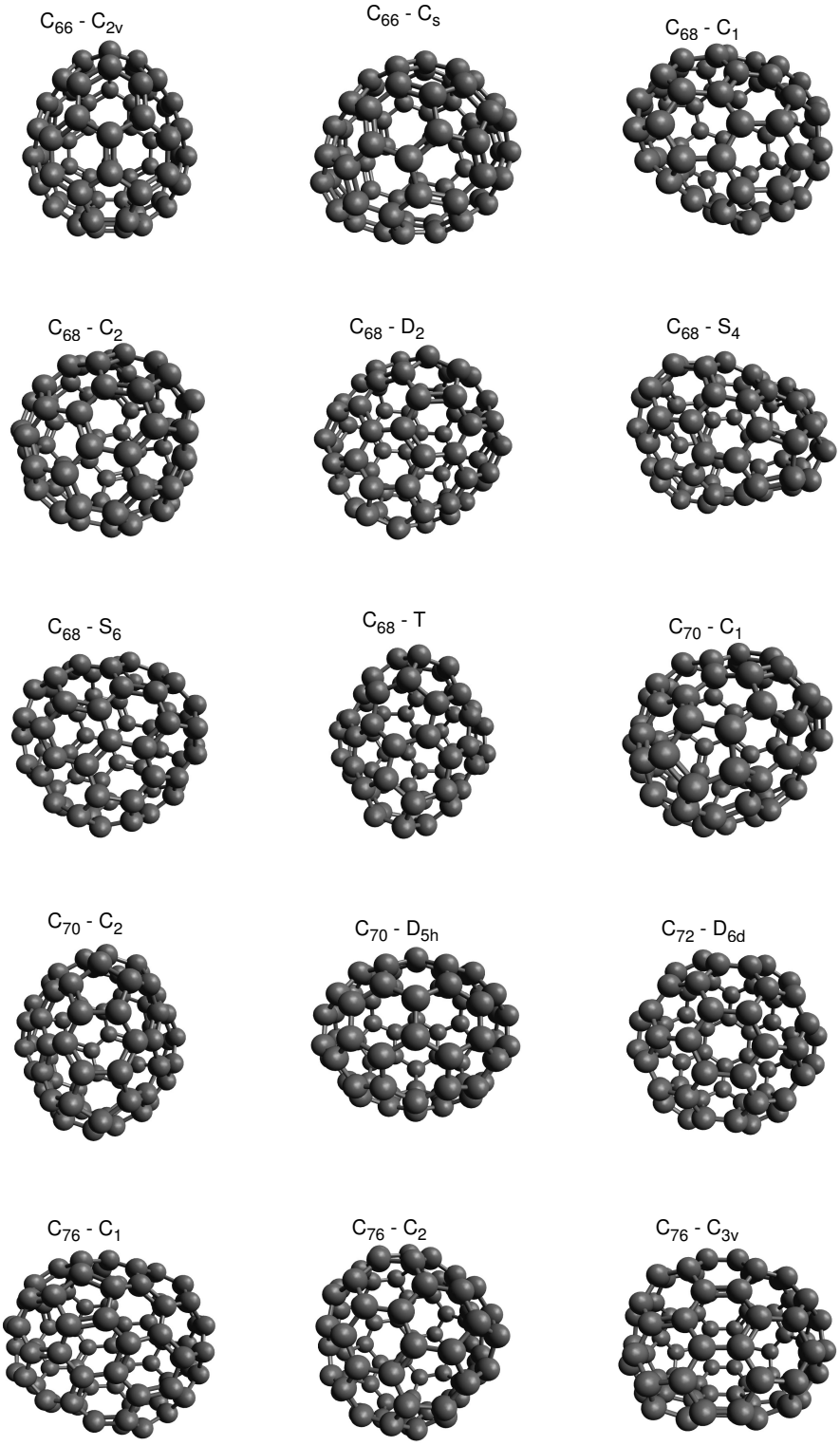

FIG. 15: Fullerenes analysed in this work.

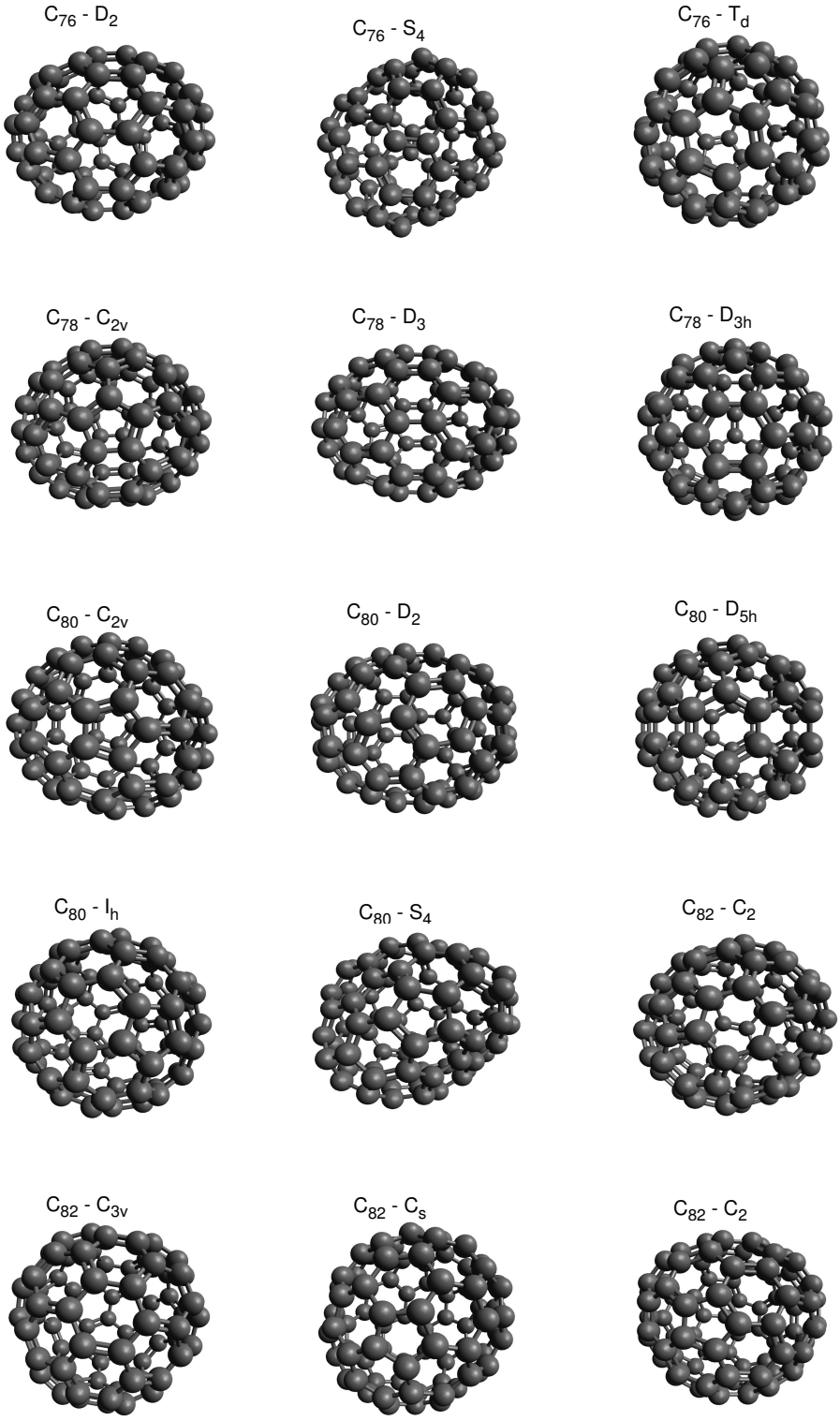

FIG. 16: Fullerenes analysed in this work.

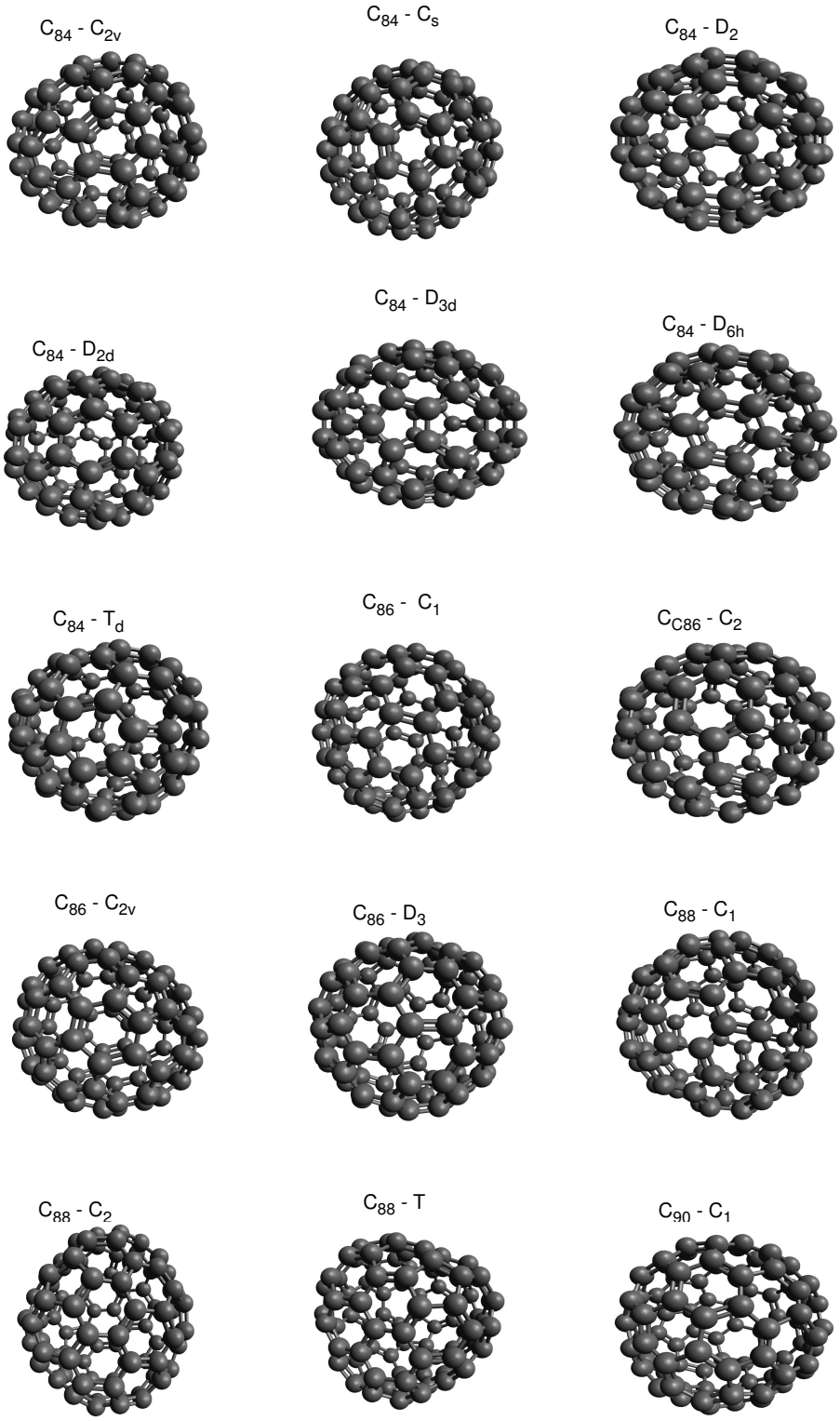

FIG. 17: Fullerenes analysed in this work.

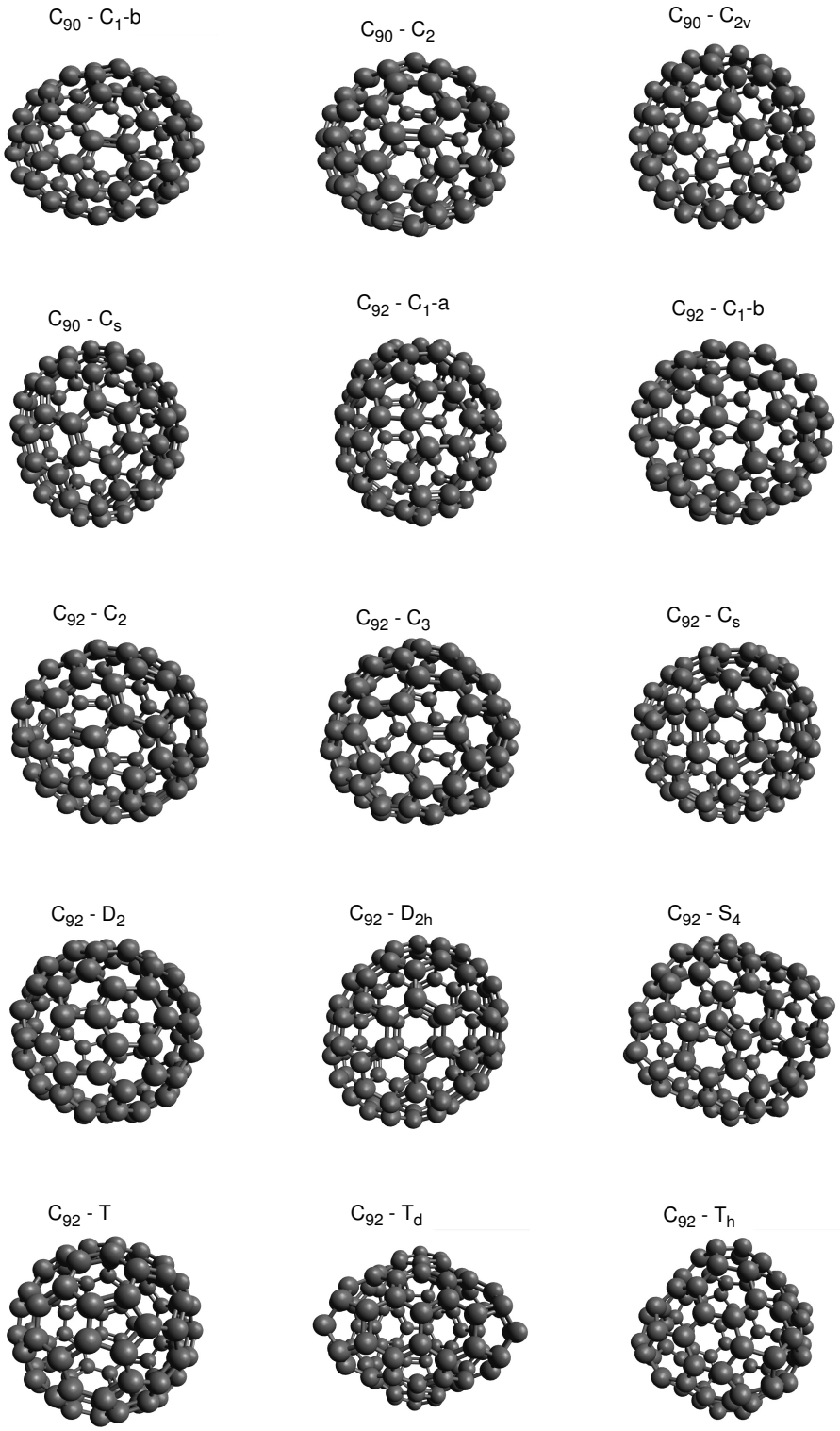

FIG. 18: Fullerenes analysed in this work.

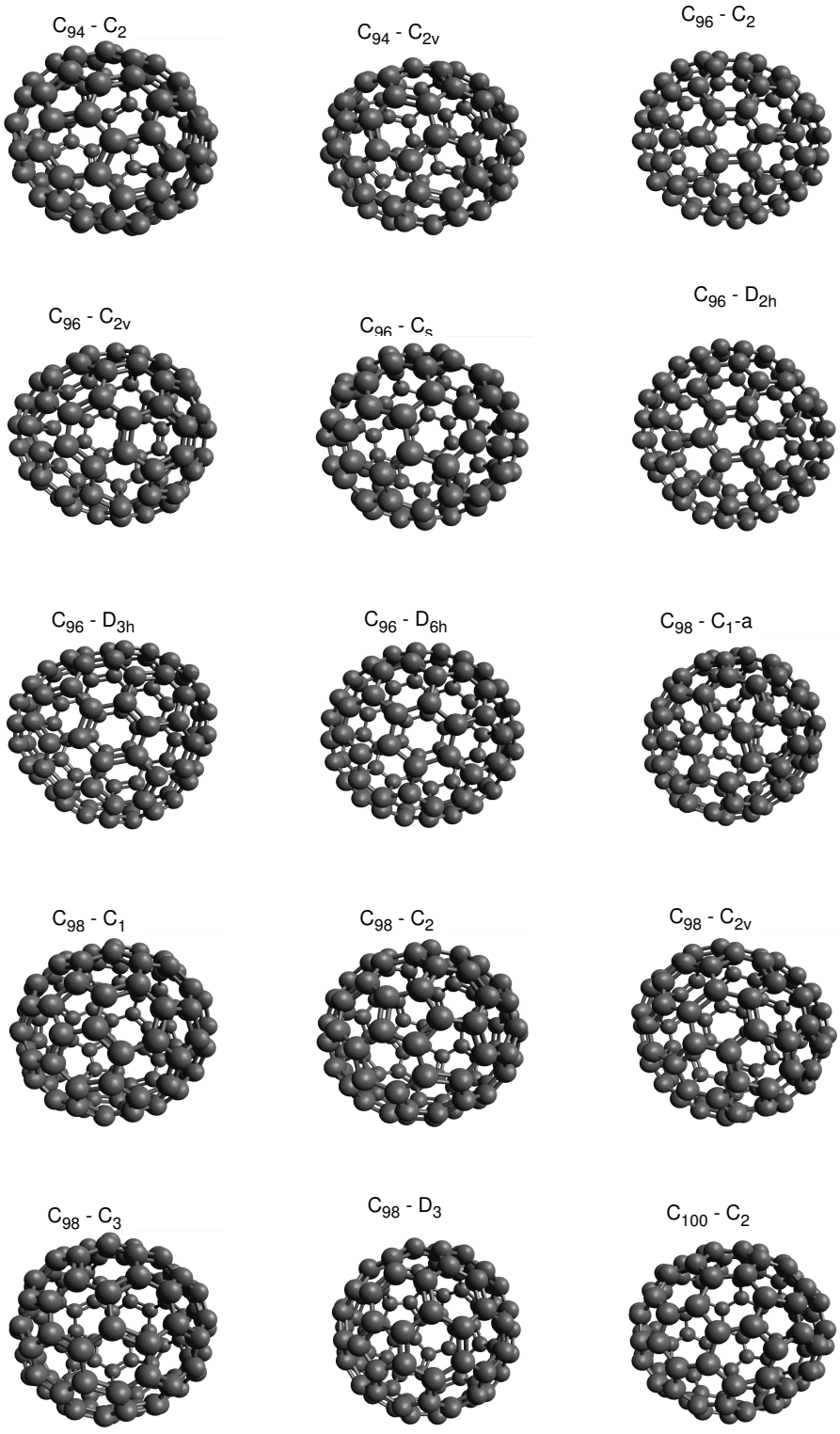

FIG. 19: Fullerenes analysed in this work.

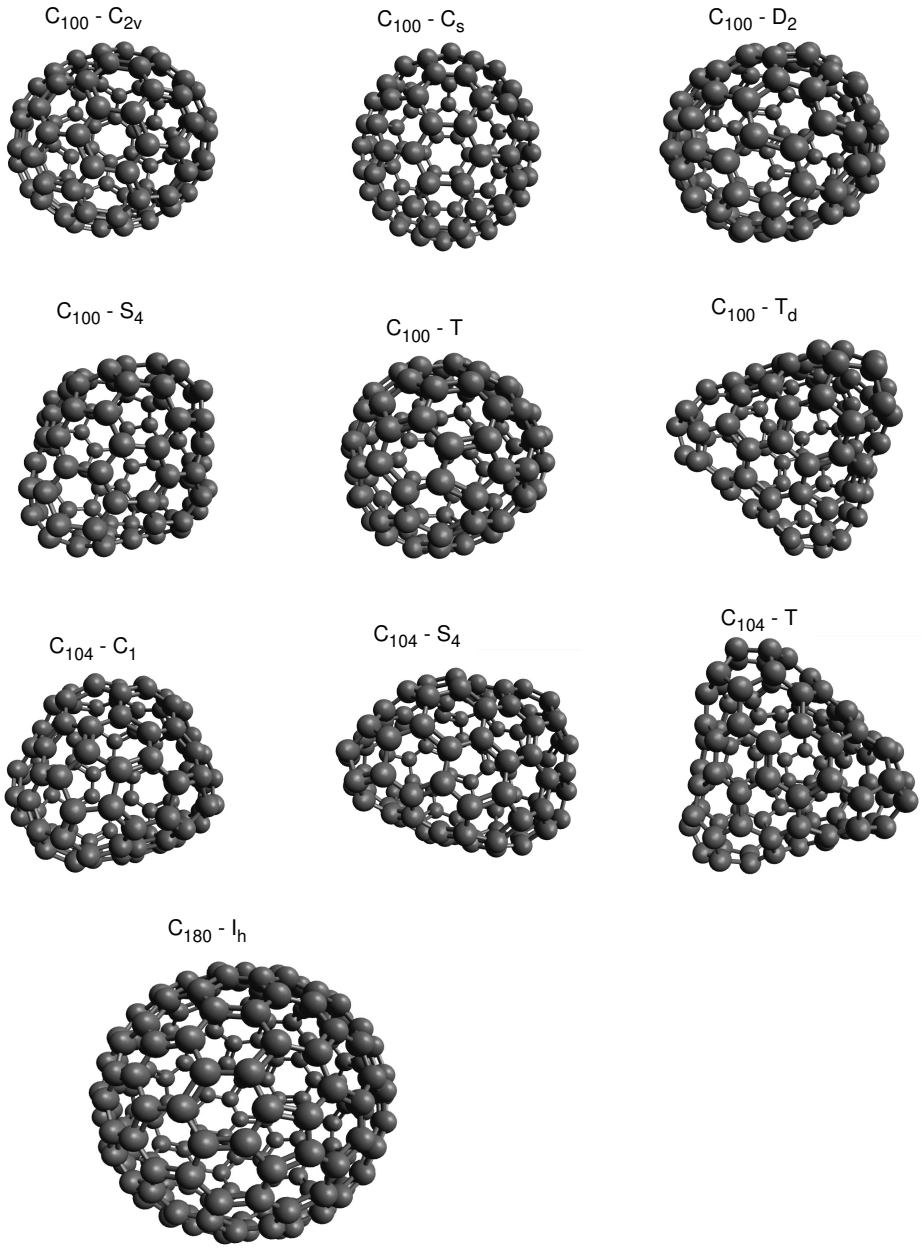

FIG. 20: Fullerenes analysed in this work.

* risueno@unizar.es
† davidcdb@kth.se

[1] García-Risueño P, Han P, Kumar S, Bester G. Frozen-phonon method for state anticrossing situations and its application to zero-point motion effects in diamondoids. Phys Rev B. 2023 Sep;108:125403. Available from: https://link.aps.org/doi/10.1103/PhysRevB.108.125403.

[2] Hartwigsen C, Goedecker S, Hutter J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. Phys Rev B. 1998;58:3641-62.

[3] García-Risueño P, Syropoulos A, Vergés N. New ideograms for physics and chemistry. N Found Phys. 2016;46:1713.

[4] Schwerdtfeger P, Wirz L, Avery J. Fullerene – A Software Package for Constructing and Analyzing Structures of Regular Fullerenes. J Comput Chem. 2013;34:1508-26.

[5] Poncé S, Antonius G, Gillet Y, Boulanger P, Laflamme Janssen J, Marini A, et al. Temperature dependence of electronic eigenenergies in the adiabatic harmonic approximation. Phys Rev B. 2014 Dec;90:214304. Available from: http://link.aps.org/doi/10.1103/PhysRevB.90.214304.

[6] te Velde G, Bickelhaupt FM, Baerends EJ, Fonseca Guerra C, van Gisbergen SJA, Snijders JG, et al. Chemistry with ADF. Journal of Computational Chemistry. 2001;22(9):931-67. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.1056.

[7] Michalak A, DeKock RL, Ziegler T. Bond Multiplicity in Transition-Metal Complexes: Applications of Two-Electron Valence Indices. The Journal of Physical Chemistry A. 2008;112(31):7256-63. PMID: 18627137. Available from: https://doi.org/10.1021/jp800139g.

[8] Nalewajski RF, Mrozek J, Michalak A. Two-electron valence indices from the Kohn-Sham orbitals. International Journal of Quantum Chemistry. 1997;61(3):589-601.

[9] Nalewajski RF, Mrozek J. Modified valence indices from the two-particle density matrix. International Journal of Quantum Chemistry. 1994;51(4):187-200. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560510403.

[10] Mrozek J, Nalewajski RF, Michalak A. Exploring bonding patterns of molecular systems using quantum mechanical bond multiplicities. Polish Journal of Chemistry. 1998;Vol. 72, nr 7S:1779-91.

[11] Nalewajski RF, Mrozek J, Mazur G. Quantum chemical valence indices from the one-determinantal difference approach. Canadian Journal of Chemistry. 1996;74(6):1121-30. Available from: https://doi.org/10.1139/v96-126.

[12] Gopinathan MS, Jug K. Valency. I. A quantum chemical definition and properties. Theoretica chimica acta. 1983;63:497-509. Available from: https://doi.org/10.1007/BF02394809.

[13] Faber C, Janssen JL, Côté M, Runge E, Blase X. Electron-phonon coupling in the C_{60} fullerene within the many-body *GW* approach. Phys Rev B. 2011;84:155104.

[14] Chen Z, Jiao H, Bühl M, Hirsch A, Thiel W. Theoretical investigation into structures and magnetic properties of smaller fullerenes and their heteroanalogues. Theoretical Chemistry Accounts. 2001;106:352-63.

[15] Paulus B. Electronic and structural properties of the cage-like molecules C20 to C36. Phys Chem Chem Phys. 2003;5:3364-7.

[16] Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. Journal of Cheminformatics. 2012 Aug;4(1):17.

[17] ;. https://github.com/pablogr/ML_fullerenes/tree/main/FULLERENE_XYZ_FILES.

[18] Zuo T, Olmstead MM, Beavers CM, Balch AL, Wang G, Yee GT, et al. Preparation and Structural Characterization of the Ih and the D5h Isomers of the Endohedral Fullerenes Tm3N@C80: Icosahedral C80 Cage Encapsulation of a Trimetallic Nitride Magnetic Cluster with Three Uncoupled Tm3+ Ions. Inorganic Chemistry. 2008;47(12):5234-44.

[19] Zhao S, Zhao P, Cai W, Bao L, Chen M, Xie Y, et al. Stabilization of Giant Fullerenes $C_2(41)$-C_{90}, $D_3(85)$-C_{92}, $C_1(132)$-C_{94}, $C_2(157)$-C_{96}, and $C_1(175)$-C_{98} by Encapsulation of a Large La2C2 Cluster: The Importance of Cluster-Cage Matching. Journal of the American Chemical Society. 2017;139(13):4724-8.

[20] Tao R, Umeyama T, Higashino T, Koganezawa T, Imahori H. A single cis-2 regioisomer of ethylene-tethered indene dimer-fullerene adduct as an electron-acceptor in polymer solar cells. Chem Commun. 2015;51:8233-6. Available from: http://dx.doi.org/10.1039/C5CC01712B.

[21] Rispens MT, Meetsma A, Rittberger R, Brabec CJ, Sariciftci NS, Hummelen JC. Influence of the solvent on the crystal structure of PCBM and the efficiency of MDMO-PPV: PCBM 'plastic'solar cells. Chemical Communications. 2003;(17):2116-8.

[22] Umeyama T, Miyata T, Jakowetz AC, Shibata S, Kurotobi K, Higashino T, et al. Regioisomer effects of [70]fullerene mono-adduct acceptors in bulk heterojunction polymer solar cells. Chem Sci. 2017;8:181-8. Available from: http://dx.doi.org/10.1039/C6SC02950G.

[23] Qian W, Bartberger MD, Pastor SJ, Houk KN, Wilkins CL, Rubin Y. C62, a Non-Classical Fullerene Incorporating a Four-Membered Ring. Journal of the American Chemical Society. 2000;122(34):8333-4.

[24] Cui YH, Tian WQ, Feng JK, Chen DL. Structures, stabilities, electronic, and optical properties of C_{64} fullerene isomers, anions (C_{64}^{2-} and C_{64}^{4-}), metallofullerene $Sc_2@C_{64}$, and $Sc_2C_2@C_{64}$. Journal of Computational Chemistry. 2008;29(16):2623-30.

[25] Vukicevic D, Cataldo F, Ori O, Graovac A. Topological efficiency of C66 fullerene. Chemical Physics Letters. 2011;501(4):442 445.