Supporting Information for

Unconventional Current Scaling and Edge Effects for Charge Transport through Molecular Clusters

Veronika Obersteiner,[†,‡] *Georg Huhs,*[§,+,‡] *Nick Papior,*[#,**] *and Egbert Zojer*[†,] *

[†] Institute of Solid State Physics, NAWI Graz, Graz University of Technology, Petersgasse 16, 8010 Graz, Austria

[§] Barcelona Supercomputing Center (BSC), C/ Jordi Girona 29, 08034-Barcelona, Spain

[+] Humboldt-Universität zu Berlin, Zum Großen Windkanal 6, 12489 Berlin, Germany

[#] Department of Micro- and Nanotechnology (DTU Nanotech) and Center for Nanostructured Graphene, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

** Institut Català de Nanociència i Nanotecnologia (ICN2), UAB Campus, Bellaterra, Spain

Corresponding Author

*egbert.zojer@tugraz.at, Tel.: +43 316 8738475

1. System Setup

The geometries for the differently sized clusters of molecules bridging the electrodes were obtained by consecutively adding molecules and simultaneously increasing the size of the unit cell. Figure S1 shows the different unit cells for the monolayer and for 1, 2, 3, 4, 9, and 16 molecules in the cluster. For all clusters, the molecules are kept in the equilibrium geometry found for the monolayer. The full monolayer is simulated by a p(2x2) unit cell, see red box in Figure S1a. The single molecule, see blue box, is mimicked by 1/16 coverage of the full monolayer. Whenever a molecule is added to the cluster, the size of the unit cell is increased in order to ensure a constant distance between molecular clusters of consecutive cells. For the explanations in this text we express geometrical relations relative to the p(2x2) unit cell, with the 3rd dimension (out of plane) as transport direction.

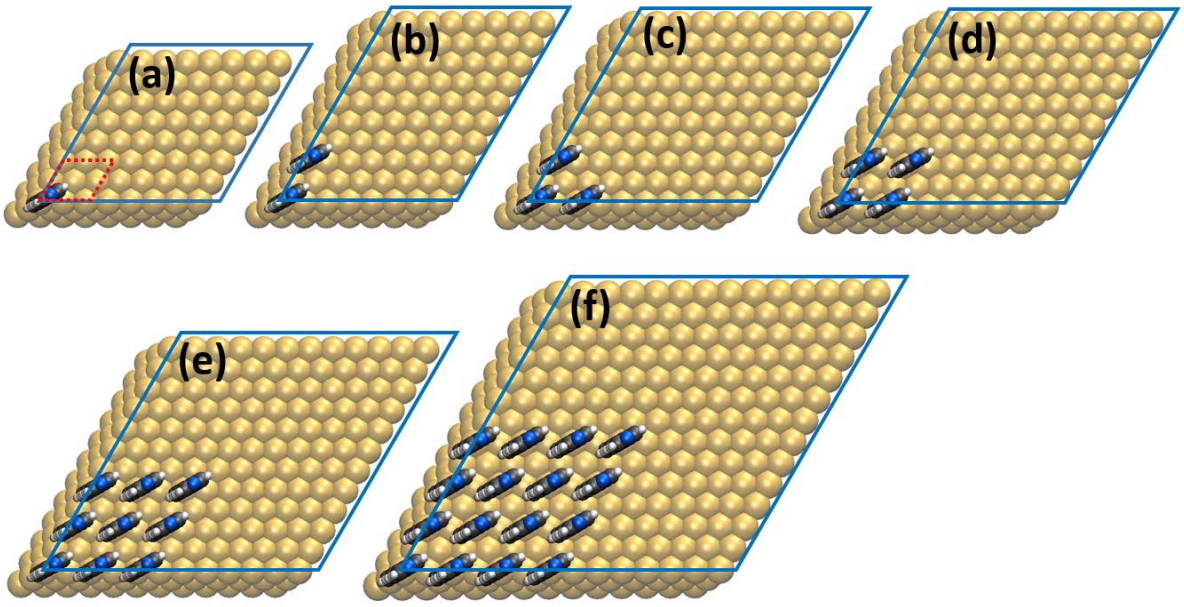


Figure S1. *Top view of the molecular junctions comprising clusters of 1(a), 2(b), 3(c), 4(d), 9(e), and 16(f) molecules. The unit cells for the periodic calculations are indicated as blue boxes. The upper electrode is removed.*

The non-equilibrium Greens function (NEGF) method requires a specific setup of the system's geometry, which in turn requires a clear nomenclature. In a physical description the *full junction* consists of the molecule attached to Au surfaces. In the NEGF calculation setup these surfaces are modelled as – in transport direction - semi-infinite leads. For the calculations the *full junction* is separated into the *central region*, consisting of the molecule

and a thin surface-layer of the leads, and the *electrodes*, which in the NEGF method describe the semi-infinite bulk parts, see Figure S2. For any geometry optimization of the molecule it is sufficient to consider the central region only.

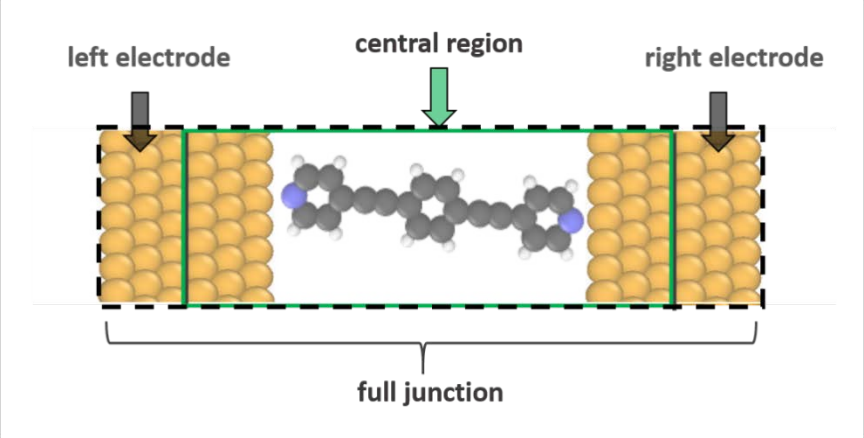


Figure S2. *Setup of the geometry for the NEGF calculations and definition of the wording used in the text. The whole geometry is called full junction, which consists of the electrodes, representing the bulk leads in the NEGF scheme, and the central region. The latter includes besides the molecule under investigation also some Au layers as transition between the molecule and the bulk part of the leads.*

2. Transport Calculations

In this work we investigate the transport characteristics of molecular junctions within the Landauer-Büttiker formalism:

$$I(V) = \frac{2e}{h} \int T(E)[f(E - \mu_{\text{left}}) - f(E - \mu_{\text{right}})]dE, \quad (1)$$

where $f(x) = \frac{1}{1+e^{\left[\frac{x}{k_B \tau}\right]}}$ is the Fermi-Dirac occupation function at an electronic temperature τ of 300K and $\mu_{\text{left/right}} = E_F \pm \frac{eV}{2}$ with e the elementary charge, E_F the Fermi level and V the applied voltage. No external voltage is considered for determining the transmission function *T(E)*, such that this quantity is evaluated as

$$T(E) = \sum_k w_k \, Tr[\Gamma_{\text{left}} G_{\text{central}} \Gamma_{\text{right}} (G_{\text{central}})^{\dagger}]. \quad (2)$$

G_{central} is the retarded Green's function of the central region. Further $\Gamma_{\text{left/right}} = i(\boldsymbol{\Sigma}_{\text{left/right}} - (\boldsymbol{\Sigma}_{\text{left/right}})^{\dagger})$ with $\boldsymbol{\Sigma}_{\text{left/right}}$ being the self-energies of the leads.

For the calculations we used the TranSIESTA 4.1-beta release and TBTrans software packages. At the time of the investigations the latest versions of these programs were still under development by one of the coauthors (NP). The final calculations presented in the main text were all done with the same development version of the packages. Due to the system size (up to 2900 atoms) we only present transmission and current calculations using the Kohn-Sham Hamiltonians as calculated by SIESTA as input to the transport calculations in TBtrans (i.e. we omit the self-consistent TranSIESTA scheme). We have, however, performed equilibrium Green function calculations with TranSIESTA to assert that the physics are unchanged due to sufficient screening towards the bulk gold electrodes, see below.

2.1 Computation workflow

The sequence of calculations for getting the electronic transport characteristics of one specific system is given by first calculating the bulk-state of the electrodes followed by computing the electronic structure of the full junction. The results serve as input for the calculation of the transmission function, which finally determines the quantities presented in this work. For a better understanding, these three steps are presented in reverse order, providing a top-down view.

Transmission calculation

The transmission characteristics are calculated by the post-processing tool TBTrans, which is part of the SIESTA software package and is in general applicable to any local orbital based set of Hamiltonian and overlap matrices. This also implies, that results from various stages of the computational workflow can be used, a feature we take advantage of as described below.

Electronic structure calculation with TranSIESTA

The presented results in the paper are all calculated from the Kohn-Sham Hamiltonians. We have, however, repeated the full NEGF calculations with TranSIESTA to assert that the physics are unchanged. In this section we discuss the intricacies of the TranSIESTA calculations performed to verify the transport calculations.

The NEGF scheme implemented in TranSIESTA needs an electronic configuration as starting point for the self-consistent cycle. This initialization is given by a ground-state DFT calculation of the full junction. For the *zero-bias* calculations of the examined systems we found that the final transport characteristics are rather sensitive to the quality of the mentioned

starting point, while the update of the electronic structure when applying the NEGF scheme for calculating the transport is negligible. This can be seen in Figure S3, where the effect of the NEGF correction on the transmission function is plotted for the transmission peaks associated with the LUMO of the pyridine-linked single molecule junction. This was found to be the most critical system, since for others the differences are even smaller. As can be seen, the deviation between only calculating the KS groundstate and applying the TranSIESTA correction amounts to 0.04 eV. Thus we decided to use the computationally much simpler approach. Note that this is only valid for equilibrium calculations where no voltage between the electrodes is applied when determining the transmission function.

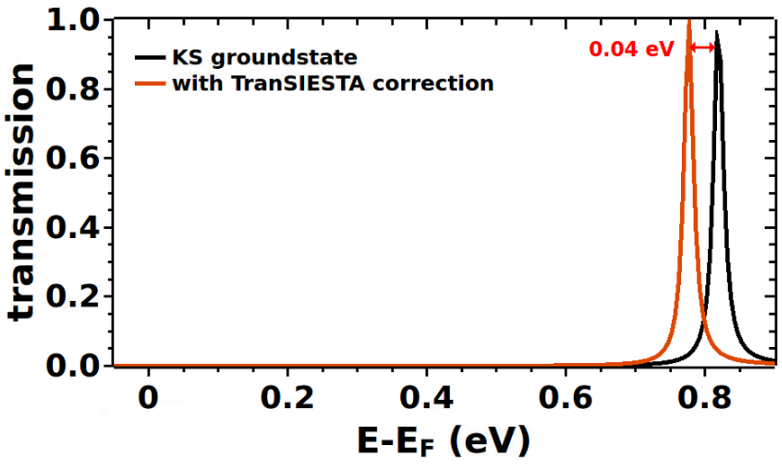


Figure S3. *Comparison of transmission functions for the pyridine-linked single molecule junction based on different electronic structure calculations. The first is a pure Kohn-Sham ground-state calculation performed with SIESTA ("KS ground state"; black line), the second a self-consistent NEGF calculation using the electronic structure from the first calculation as starting point ("with TranSIESTA correction"; orange line). The single molecule system shown here represents the sparsest case. For the other extreme, the densely packed monolayer, the two calculations were virtually identical (peak shifts <0.005eV). The differences in the relevant peaks of junctions with other docking groups are smaller; moreover those peaks are much less pronounced.*

Electrode calculation

The NEGF scheme requires calculating the ground state electronic structure of the electrodes (as shown in Figure S2) in a bulk configuration. To save computational effort, TranSIESTA and TBTrans implement the Bloch's theorem for electrodes with periodicity transverse to the transport direction. For example, as shown in Figure S1, the electrode surface area (cross section perpendicular to the transport direction) of the single molecule system consists of 8x8 atoms. In transport direction the three layers are arranged in ABC stacking. Thus, the electrode can be expressed as an 8x8x1 repetition of a cell containing 1x1x3 atoms. The k-point grids used for the calculations of the small and large cell are related: for each cell direction $N\_kpt_{small} = N\_kpt_{large} \times N\_repetitions$ has to be fulfilled.

Finally this means, that a SIESTA electronic structure calculation of the 1x1x3 atoms cell is executed, writing the final Hamiltonian and overlap matrices (`TSHS` file format). In a subsequent TranSIESTA or TBTrans calculation this is integrated as electrode by the following configuration in the input file:

```
TS.Elecs.Bulk                .true.

%block TS.Elecs
  Left
  Right
%endblock TS.Elecs

%block TS.Elec.Left
  TSHS            ../electrode/AuBulk_1x1x3.TSHS
  chem-pot        Left
  semi-inf-dir    -a3
  elec-pos begin  1
  bloch           8 8 1
%endblock TS.Elec.Left
```

In all our systems these electrodes have the same structure, but since N_kpt_{large} is the same for all clusters (see chapter 2.2), the electrode calculation has to be adapted to each cluster size, while amongst systems that feature different docking groups the results are transferrable.

2.2 Numerical setup

All performance-relevant parameters shown in the following text have been examined and optimized. For most of them the sweet spot for the tradeoff between accuracy and computational cost can be found in a straight-forward manner. The criterion for this type of convergence was that the differences of the locations of peaks in the DOS or transport curves

have to be below 0.1 eV. Figure S3 demonstrates that this is fulfilled for the junction with the largest deviations.

The situation is different regarding the pseudopotentials and basis set, for which prior work for aligning DOS data with highly converged (with a plane wave cutoff of 274 eV) VASP calculations existed.[1] To verify also the PAW potentials, we compared to FHI-AIMS[2] all-electron calculations, which quantitatively agreed to the VASP results. We also tested a more efficient basis for gold, which has been optimized for surface properties,[3] but did not observe a considerable reduction of computational costs or an improved correlation with the VASP results.

Basis and functional

SIESTA's build in standard configuration of a double-zeta with polarization (DZP) basis was used, with changing the energy shift to 0.001 Ry (this is necessary in order to correctly reproduce the level alignment obtained from highly converged VASP calculations, see the Supporting Information of Ref.[1]) Also for the grid parameters the defaults were used. The exchange-correlation functional was described by the Perdew-Burke-Ernzerhof (PBE) variant of the generalized gradient approximation (GGA). This is configured in the `fdf` input files as follows:

```
PAO.BasisType                 split
PAO.BasisSize                 DZP
PAO.EnergyShift               0.001 Ry
XC.functional                 GGA
XC.authors                    PBE
MeshCutoff                    200 Ry
```

Pseudopotentials

The headers of the pseudopotential files used for the various atomic species are as follows:

```
Au pb rel pcec
 ATM 3.2.2 26-JUN-07 Troullier-Martins
 6s 1.00r r= 2.08/6p 0.00r r= 2.47/5d10.00r r= 1.31/5f 0.00r r= 1.88/
   4  3 1212  0.313766098312E-04  0.125000000000E-01   11.0000000000

C  pb rel pcec
 ATM 3.2.2 20-JUL-07 Troullier-Martins
 2s 2.00r r= 1.17/2p 2.00r r= 1.31/3d 0.00r r= 1.25/4f 0.00r r= 1.25/
   4  3 1006  0.413125362778E-03  0.125000000000E-01   4.00000000000

H  pb rel nc
 ATM 3.2.2 12-MAY-07 Troullier-Martins
```

```
 1s 1.00r r= 2.04/2p 0.00r r= 1.55/3d 0.00r r= 1.25/4f 0.00r r= 1.19/
   4  3  863  0.247875217667E-02  0.125000000000E-01   1.00000000000

N  pb rel pcec
 ATM 3.2.2 20-JUL-07 Troullier-Martins
 2s 2.00r r= 1.42/2p 3.00r r= 1.61/3d 0.00r r= 1.52/4f 0.00r r= 1.55/
   4  3 1018  0.354107453809E-03  0.125000000000E-01   5.00000000000

S  pb rel pcec
 ATM 3.2.2 20-JUL-07 Troullier-Martins
 3s 2.00r r= 1.34/3p 4.00r r= 1.37/3d 0.00r r= 1.29/4f 0.00r r= 1.27/
   4  3 1084  0.154922011042E-03  0.125000000000E-01   6.00000000000
```

k-point grids

For the electrode bulk calculation, 20 k-points in transport direction were used, perpendicular to it we employed grids from 16x16 up to 28x28 k-points, depending on the size of the leads as explained above.

The TranSIESTA electronic structure calculations used converged grids generated according to the Monkhorst-Pack scheme with the following number of grid points:

Monolayer:	8x8x1
Clusters:	2x2x1

Converged transmission functions from TBTrans are requiring higher grid densities:

Monolayer:	32x32x1
1 molecule cluster:	6x6x1
2 molecule cluster:	6x6x1
3 molecule cluster:	6x6x1
4 molecule cluster:	6x6x1
9 molecule cluster:	4x4x1
16 molecule cluster:	4x4x1

In all configurations no offset was used; thus, the Γ-point is always included.

SCF mixing and convergence parameters

The following configuration was found to give reasonable convergence for all systems:

```
DM.MixingWeight              0.02
DM.NumberPulay               4
DM.MixSCF1                   .false.
MixHamiltonian               .true.
DM.Tolerance                 0.0001
```

The convergence behavior during the SCF loop depends also on the system size, thus, it could have been optimized further for every cluster size. However, as it turns out that these optimization procedures would be more costly than the possible gains in SCF iterations, we refrained from pursuing that further.

Typical durations of the SCF loop are given below in chapter 2.3.

Further options for electronic structure calculations with SIESTA

The following options are necessary for SIESTA to write the TSHS files required for TBTrans:

```
TS.SIESTA.Only               .true.
TS.SaveHS                    .true.
TBT.Elecs.Neglect.Principal   .true.
```

The last option is used because otherwise SIESTA stops, reporting problems in the electrode setup. This arises due to the self-energy calculations, which require only nearest-cell interactions, while the far-reaching basis functions caused by the very small PAO energy shift increases the range of the basis.

Further parameters for the transmission calculation with TBTrans

TBTtrans has to be given the offset from the real axis for the integration, and the energy points for computing the transmission, specified by an energy range and resolution. The latter is chosen such that the finest peaks in the transport characteristics are resolved well.

```
TBT.Elecs.Eta    0.0000010000 Ry
%block TBT.Contours
  neq
%endblock TBT.Contours
%block TBT.Contour.neq
  part line
   from   -3.00000 eV to    3.00000 eV
    delta    0.00500 eV
     method mid-rule
%endblock TBT.Contour.neq

TBT.Elecs.Neglect.Principal   .true.
```

The last option is needed because otherwise TBTrans stops, reporting problems in the electrode setup. This arise due to the self-energy calculations, which require only nearest-cell interactions, while the far-reaching basis functions caused by the very small PAO energy shift increases the range of the basis.

One can define which atoms are considered for the calculation of the transmission, which we did for reducing computational cost and memory needs. An example is:

```
%block TBT.Atoms.Device
```

```
    atom from 49 to 62
%endblock
```

A recommendation for the atoms to include (smallest set for correct calculations) can be extracted from the output of a TBTrans run with specifying the option

```
TBT.analyze    .true.
```

2.3 Computational setup and cost

As some of the computations presented in the current article are computationally extremely costly and demanding, in the following we provide an overview of the computational details of our studies to aid readers in performing similar simulations. The electronic transport calculations were executed at the MareNostrum3 supercomputer, hosted by the Barcelona Supercomputing Center. It is built upon 48896 Intel Sandy Bridge processors, providing a peak performance of 1.1 Petaflops. Each of its 3056 nodes features 16 computing cores and at least 32GB of RAM. The nodes with larger main memory were not used for the calculations presented here.

Electrode calculations

Due to using the repetition scheme described in Chapter 2.2, the electrode system, consisting of only three atoms, causes only minor effort. Using 8 cores with pure MPI parallelization, each calculation lasts only a few minutes.

Electronic structure calculations with SIESTA

For simulating the molecular cluster junctions correctly, leads that are quite wide and include six layers of Au each are needed. The resulting systems contain up to a few thousand atoms, yielding considerable computational effort and memory needs, which require high performance computing (HPC) resources. We used the standard diagonalization scheme based on ScaLAPACK routines and MPI parallelization. Thus the cost for a single SCF iteration scales cubically with the system size. Those routines become more inefficient when increasing the number of cores, so for each system size a tradeoff between reducing the time to solution and computational efficiency was found by executing a simple scaling test and applied to all similar systems. The given system sizes and determined numbers of cores are given in Table S1.

Table S1. *Sizes of all systems presented and the number of MPI processes used for the SIESTA electronic structure calculations.*

	Number of atoms			Number of MPI processes
	Pyridine	**Thiolate**	**Isocyanide**	
Monolayer	82	84	86	32
1 molecule	802	804	806	256
2 molecules	1028	1032	1036	256
3 molecules	1302	1308	1312	256
4 molecules	1336	1344	1352	256
9 molecules	2034	2052	2070	512
16 molecules	2896	2928	2960	1024

The resulting costs differ slightly amongst the systems due to variations in the numbers of atoms and SCF iterations needed, complemented by non-deterministic fluctuations in the platform. Prototypical data from one set of calculations are given in Table S2.

Table S2. *Computational effort for the SIESTA electronic structure calculations of various cluster sizes for the same molecule on the example of the pyridine-docked junction. The effort for a single SCF iteration is mainly determined by the ScaLAPACK eigensolver routine, depending on the number of atomic orbitals, which in turn is given by the number of atoms and their basis cardinality (defined by the DZP scheme). Finally the time per iteration is determined by the number of processors and efficiency of the routine. The differences to the other molecules investigated in this paper concerning time per iteration as well as the number of iterations are minor.*

	Electronic structure			
	Number of MPI processes	**Number of SCF iterations**	**Time per iteration** [hh:mm:ss]	**Total Time** [hh:mm]
Monolayer	32	44	00:01:04	00:48
1 molecule	256	57	00:08:31	08:10
2 molecules	256	41	00:15:39	10:50
3 molecules	256	37	00:27:08	16:57
4 molecules	256	38	00:28:34	18:19
9 molecules	512	36	00:49:18	30:09
16 molecules	1024	38	01:30:38	60:12

Transmission calculations with TBTrans

TBTrans implements a hybrid parallelization scheme, using MPI for energy points, and OpenMP for distributing orbitals. Multithreading allows using more cores as well as memory per MPI process. The latter is needed, since larger systems would not fit in the memory of a single core; thus we scaled the resources by increasing the number of threads per MPI process whenever needed, see Table S3.

Table S3. *Configuration and effort for the transmission calculations with TBTrans. The program calculates the transmission for each energy point, as defined by the energy range and distance between points in the fdf file, and k-point. The times are prototypical for a resolution of 0.005 eV and an energy range of -3 to 3 eV, which is a wider range than shown in Figure 2 in the main text. The number of k-points are listed above on page 8. The times are similar for all molecules.*

	Number of transmission calculations *Energy-points* x *k-points*	**Hardware configuration** *Processes* x *Treads*	**Total number of cores**	**time** [hh:mm]
Monolayer	614400	64x1	64	00:14
1 molecule	21.600	128x2	256	01:02
2 molecules	21.600	128x2	256	01:51
3 molecules	21.600	128x4	512	01:51
4 molecules	21.600	128x4	512	01:57
9 molecules	9600	128x8	1024	03:18
16 molecules	9600	128x16	2048	04:49

Total effort and calculation limits

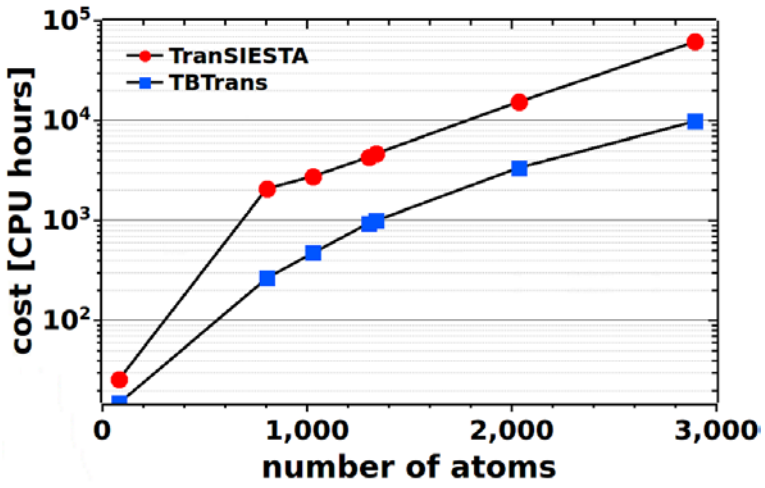


Figure S4. *Computational effort for an increasing number of atoms in the unit cell for the*

pyridine-linked junction, i.e. starting from the monolayer (82 atoms per unit cell) up to 16 molecules in the cluster (2896 atoms); for detailed numbers, see Tables S2 and S3. The effort is given separately for the SIESTA electronic structure and TBTrans transmission calculations.

The resulting computational costs for different system sizes are plotted in Figure S4. Two things can be observed: First, the total effort is mainly determined by the electronic structure calculation, which is also more limited than the transmission calculation regarding the number of processors that can be used. Second, the total cost for the monolayer amounts to around 90 CPU-hours, but grows for the 16 molecule cluster to around 70.000 CPU-hours. In total, the calculation of all different clusters together with the monolayer for one particular docking group accumulates to around 110.000 CPU-hours.

Tests with a 25 molecule cluster showed that the calculation time for the SIESTA electronic structure calculation would double to about 5 days.

3. Additional Data

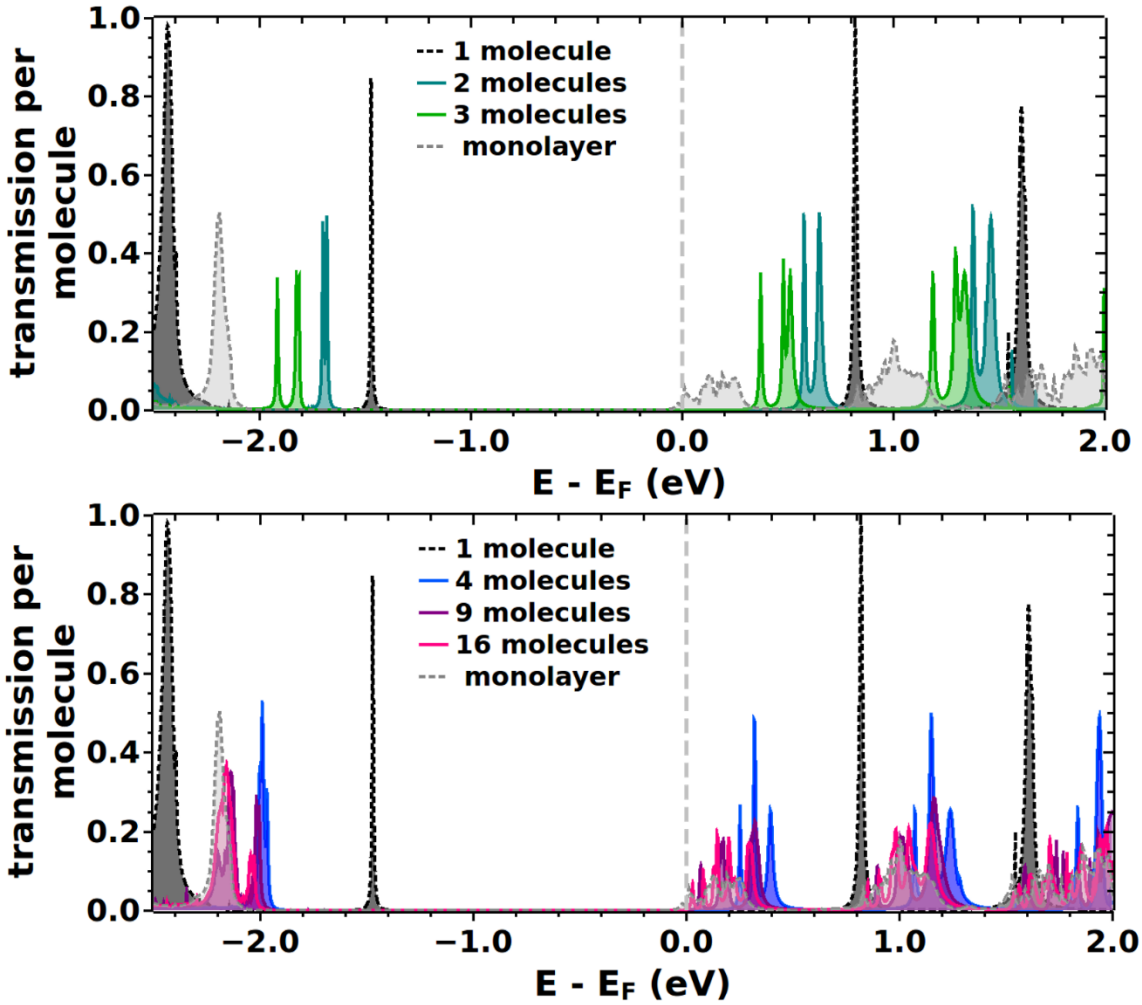


Figure S5. *Calculated (zero-bias) transmission function for the pyridine linked junction for different cluster sizes, i.e. 2, 3 molecules (upper plot) and 4, 9, 16 molecules (lower plot) compared to the single molecule and monolayer junctions. The Fermi level, E_F, is used as the reference energy.*

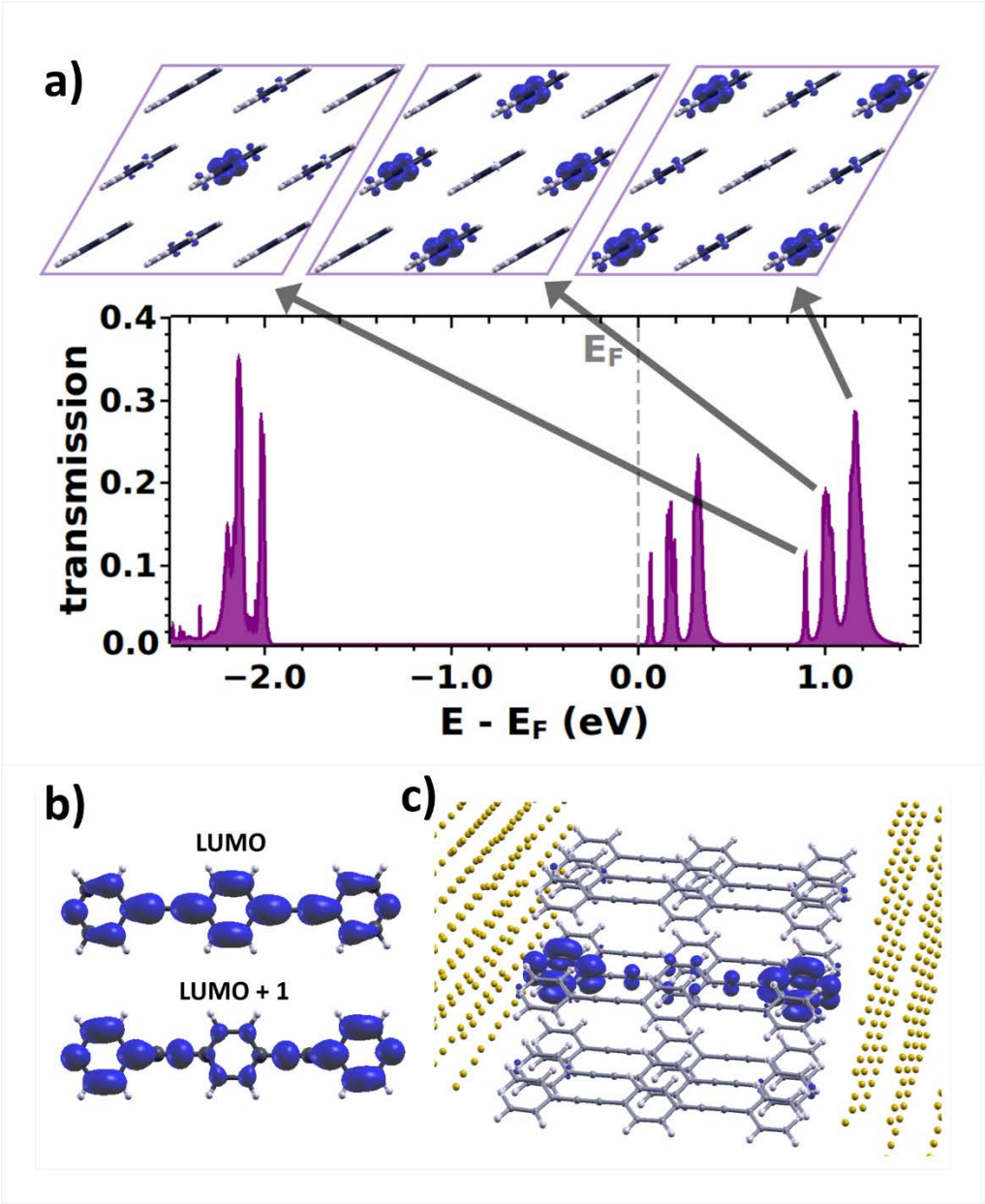


Figure S6. ***(a)*** *Transmission function for the pyridine-linked cluster containing nine molecules. The insets show a top view of the LDOS (local density of states), as obtained from VASP with an isovalue of 0.03 per Å^3, associated to the indicated unoccupied transmission channels (calculated for the following energy windows: 0.8-0.95 eV, 0.95-1.1 eV and 1.1-1.3 eV).* ***(b)*** *LDOS of the LUMO (lowest unoccupied molecular orbital) and the LUMO+1 for a single porphyrin molecule in gas phase.* ***(c)*** *LDOS associated with the transmission features derived from the molecular LUMO+1 for the full junction.*

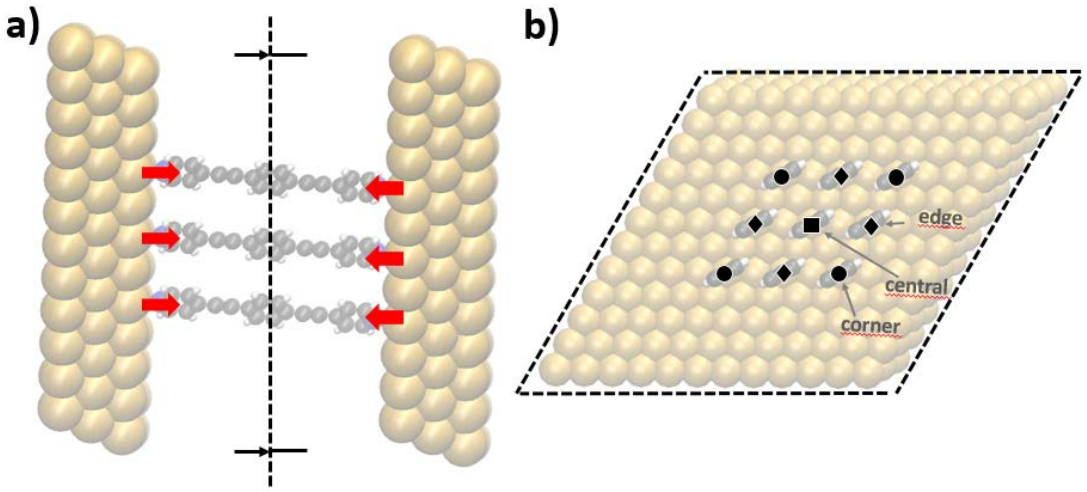


Figure S7. *Visualization of the electrostatic model. The electrostatic situation is described by two opposite square 2D arrays of dipoles generated by the docking groups and charge redistributions due to the bonding to the leads. The N point dipoles mimicking this situation are shown as red arrows in* ***(a)***. *We calculate the shift in the electrostatic energy an electron would experience in the middle of the two arrays, see plane indicated in (a). The positions of the central, corner and edge molecules (dipoles) are indicated in* ***(b).***

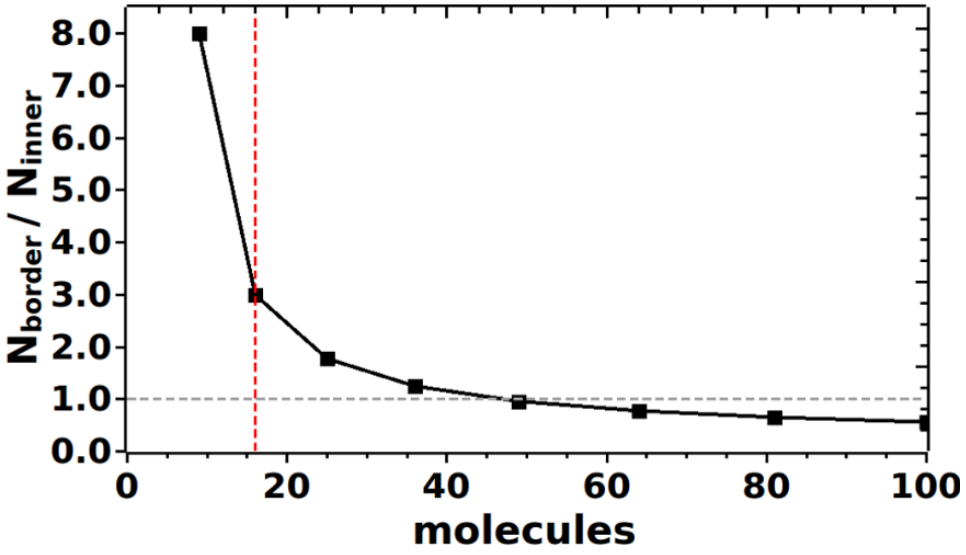


Figure S8. *Ratio of the number of border molecules to molecules inside the cluster for increasing number of molecules (dipoles) in the cluster.*

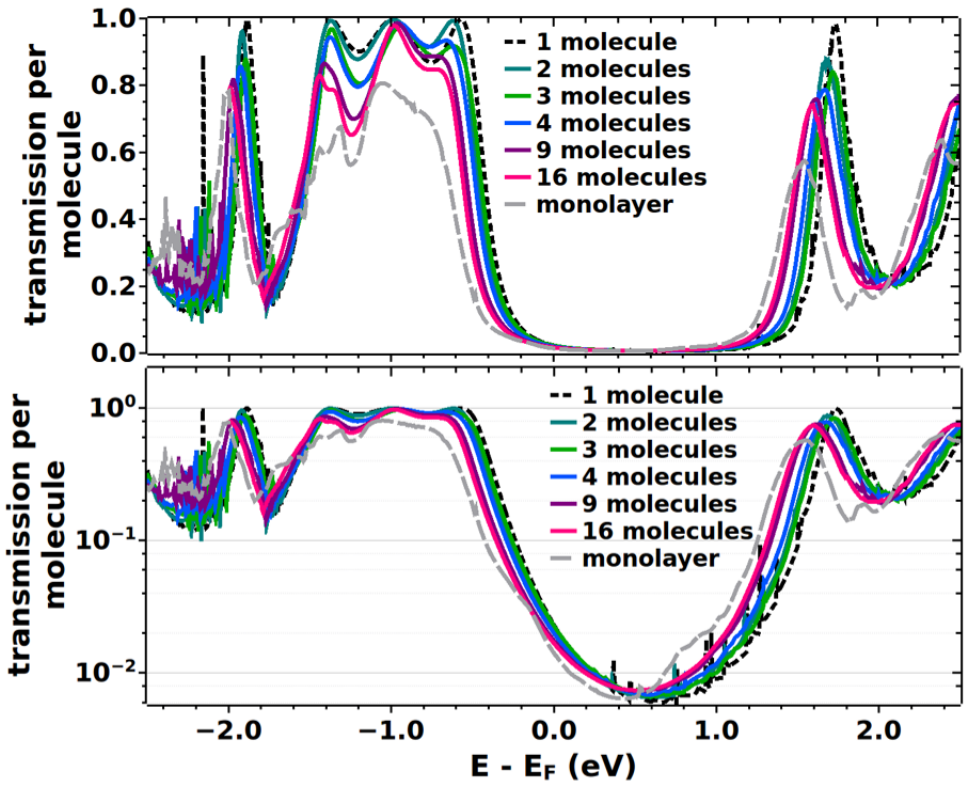


Figure S9. *Calculated (zero-bias) transmission function for the thiolate-linked junction for different cluster sizes (1, 2, 3, 4, 9, and 16 molecules) and the monolayer junction; on a linear (upper panel) and logarithmic (lower panel) scale. The Fermi level, E_F, is used as the reference energy.*

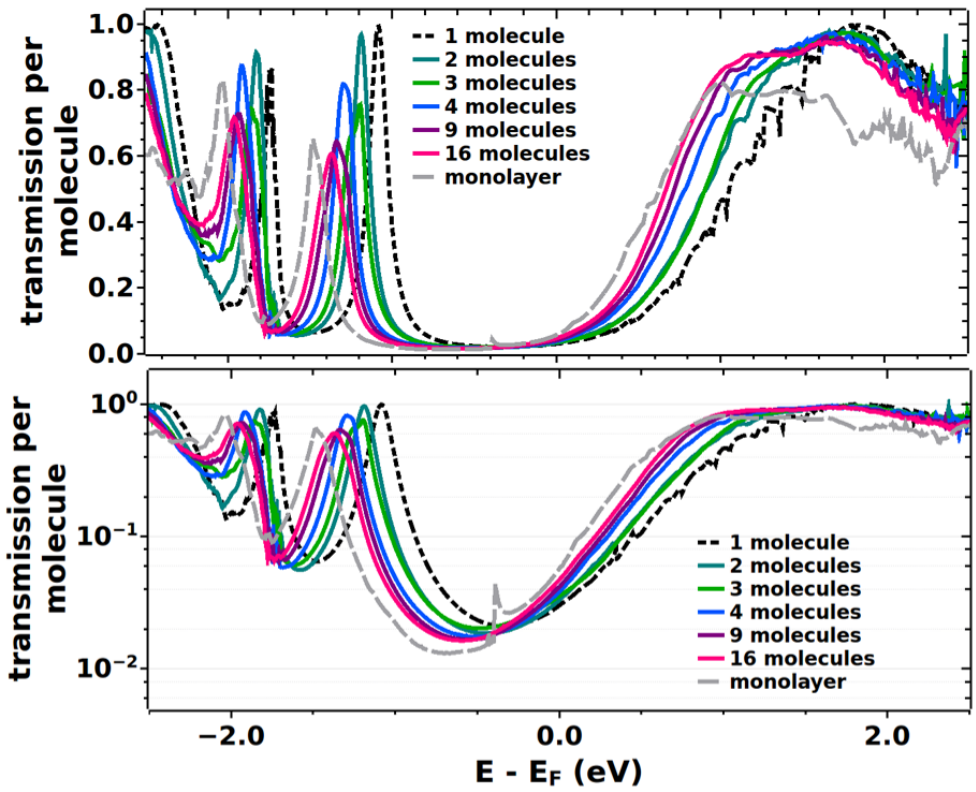


Figure S10. *Calculated (zero-bias) transmission function for the isocyanide-linked junction for different cluster sizes (1, 2, 3, 4, 9, and 16 molecules) and the monolayer junction; on a linear (upper panel) and logarithmic (lower panel) scale. The Fermi level, E_F, is used as the reference energy.*

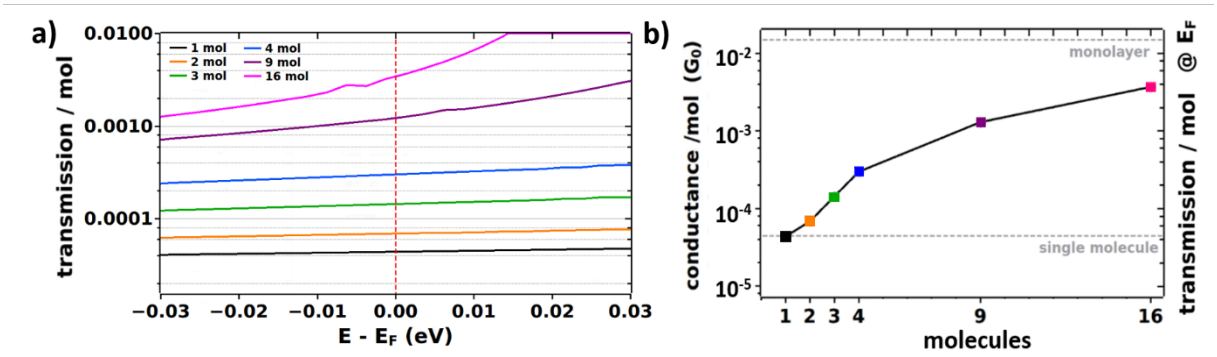


Figure S11*. (a) Calculated (zero-bias) transmission function for the pyridine-linked junction for different cluster sizes (1, 2, 3, 4, 9, and 16 molecules) on a logarithmic scale zoomed into the region around the Fermi level, E_F, which is used as the reference energy. (b) Calculated (zero-bias) conductance for the pyridine-linked junction for different cluster sizes compared to the monolayer junction. The conductance was obtained from the calculated transmission function shown in (a) as $G(EF) = T(EF).G_0$ where $G_0 = 2e^2/h$ refers to the quantum of conductance.*

4. Identification of the origin of the edge effects

In order to clarify to what extent the observed edge effects are related to the quantum-mechanical interactions between the molecules and to what extent they arise from differences in the local potential, we pursued a dual approach. First, we designed and parameterized a tight-binding model, which in a straightforward manner allows distinguishing between the quantum-mechanical coupling (expressed by the hopping element, *t*) and variations in the local energy landscape due to electrostatic edge effects (via changes in the site energy). This was combined with calculating the electronic states of the molecular projected self-consistent Hamiltonian (MPSH) by removing all couplings to the leads. This retains the effective electrostatics used in the further analysis. Secondly, we calculated the electronic states of molecular clusters in the absence of the electrodes but manipulating the magnitude and orientations of the molecular dipoles at the ends of the molecules.

4.1 Tight-binding model

To qualitatively analyze the energetics of the 9-molecule cluster discussed in detail in the main manuscript (c.f., Figure 3), we designed a simple tight-binding model as sketched in Figure S12, where we only use a single orbital to describe each molecule. Only nearest neighbor coupling via the hopping element, *t*, and variations in the onsite-energies relative to the central molecule expressed by Δ_{ab} and Δ_{ac} were considered. a, b, and c, here denote the three types of molecules in the cluster (a = center, b = edge, and c = corner). The next-nearest neighbor couplings, *t'*, are not considered in this simplistic model aiming at providing mostly qualitative insights. This is reasonable considering the symmetry and spatial shapes of the π-orbitals on the individual molecules (compare charge densities in Figure 3a of the main manuscript).

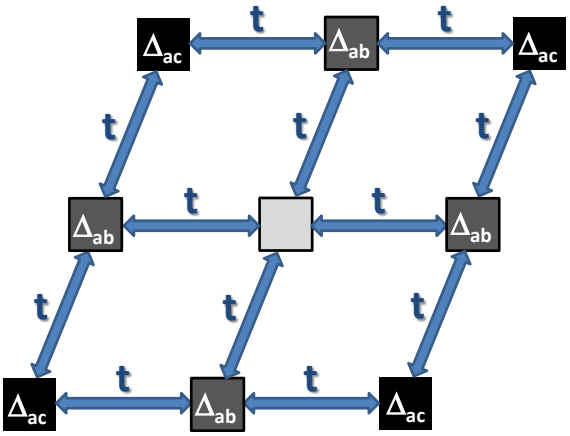


Figure S12: *Setup of the tight-binding model used to analyze edge effects in a 9-molecule cluster. Each molecule is only described by a single orbital.*

Determination of the tight-binding parameters:

To obtain the parameters for the tight-binding model, as a first step we extracted the MPSH Hamiltonian describing the nine molecules of the cluster, H_M, from the DFT calculations described in the main manuscript. H_M provides the eigenstates of the 9-molecule system in the electrostatic environment of the electrodes. The corresponding eigenvalue equation can be written as:

$$H_M \left| M_i \right\rangle = \varepsilon_i^M S_M \left| M_i \right\rangle \tag{3}$$

Notably, in absence of coupling between the leads and the cluster, the molecule-localized eigenstates of the full system Hamiltonian will coincide with eigenstates of H_M. As a second step, to evaluate the impact of electrostatic effect for each of the molecules we partition H_M into Hamiltonians (H^{α}_M) for each of the 9 molecules in the junction. This yields the following eigenvalue equation

$$H_M^{\alpha} \left| M_i^{\alpha} \right\rangle = \varepsilon_i^{M,\alpha} S_M^{\alpha} \left| M_i^{\alpha} \right\rangle, \tag{4}$$

where α refers to one of the 9 molecules. The eigenvalues obtained in this way allow identifying the local effect of the global electrostatic environment. From these we determine the *effective* electrostatic potential by taking the LUMO level for each of these 9 molecules (surrounded by an effective electrostatic potential). The shifts between the LUMO energies then yield the variations in the site-energies relative to the central molecule, Δ_{ab} and Δ_{ac},

where for molecules at the edges and corners average values are taken. In this way, one obtains $\Delta_{ab} = \varepsilon^{b}_{LUMO} - \varepsilon^{a}_{LUMO} = 66meV$ and $\Delta_{ac} = \varepsilon^{c}_{LUMO} - \varepsilon^{a}_{LUMO} = 193meV$. Finally, the hopping term t is chosen such that the spread between the 9 eigenvalues in the tight-binding model and the 9 LUMO levels of H_M is minimized. This yiels t = 34 meV, which is approximately 1/2, respectively, 1/6 of the differences in onsite energies.

Results of the molecular-cluster Hamiltonian and the tight-binding eigenstates:

In the following we will compare the eigenvalues and eigenstate localizations on the molecules by comparison of the tight-binding model and the molecular-cluster Hamiltonian (MPSH). The results obtained from diagonalizing the molecular cluster Hamiltonian, H_M, are shown in Figure S13, while those for the tight-binding model are contained in Figure S14.

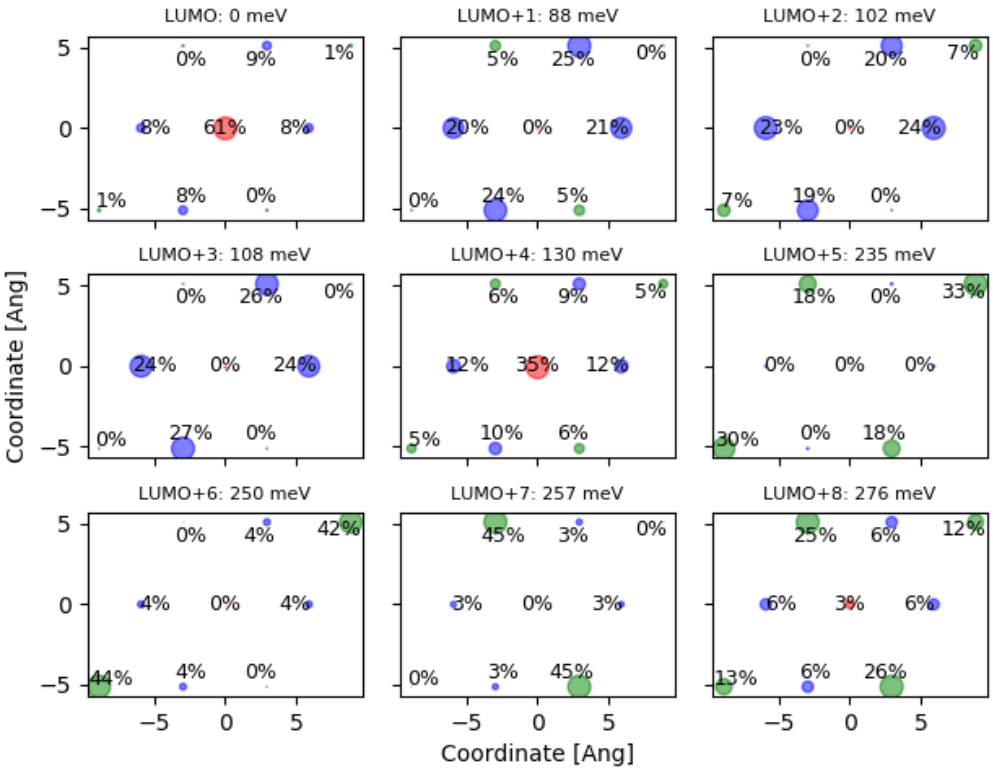


***Figure S13**: Energy eigenvalues and weights of the eigenstates on the individual molecules for the nine lowest unoccupied states in the 9-molecule cluster as obtained from diagonalizing H_M. The diameter of the circles scales with the weight and red, blue and green circles denote central, edge, and corner molecules respectively. All energies are given relative to the LUMO.*

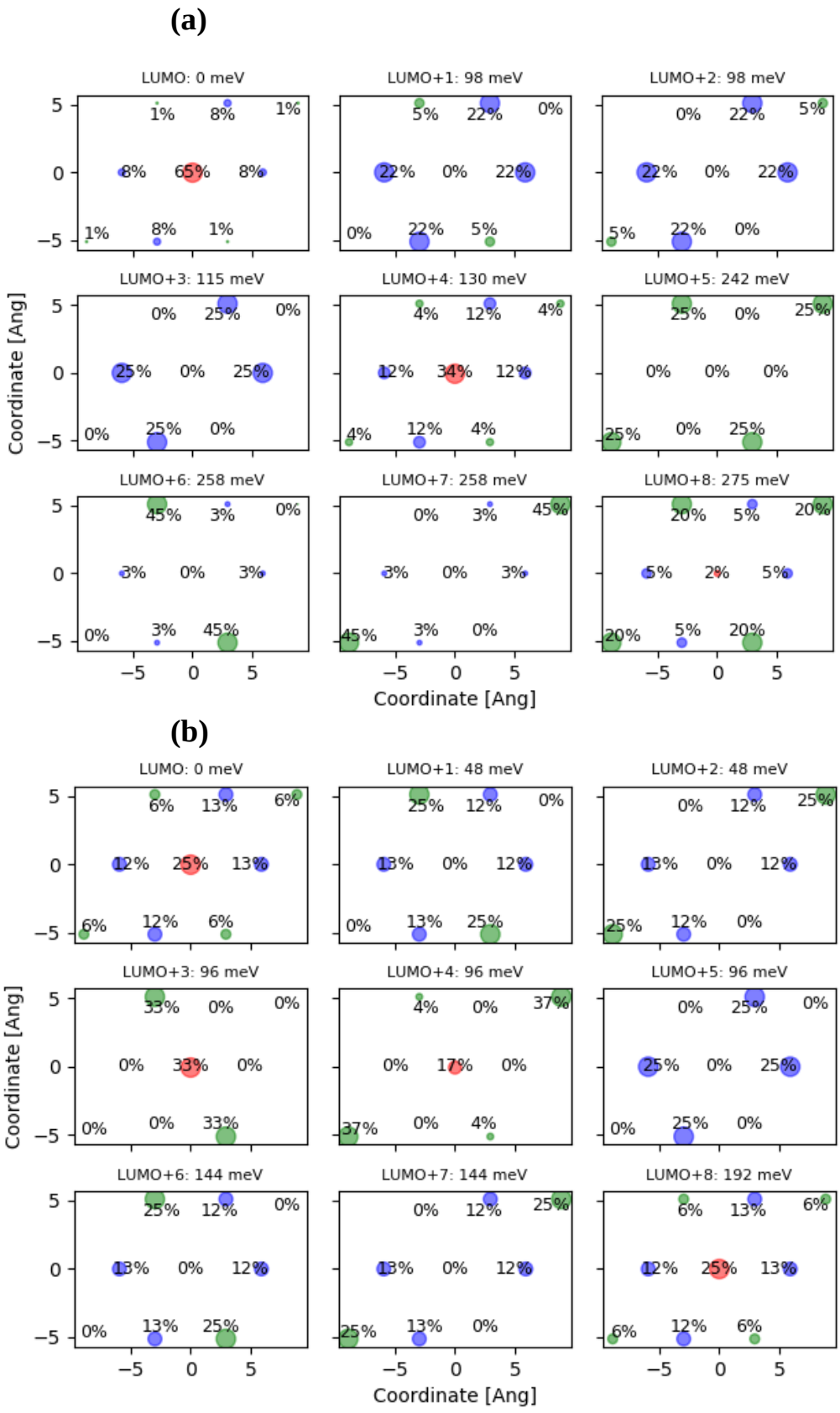


Figure S14: Energy eigenvalues and weights of the eigenstates on the individual orbitals for the nine states as obtained from diagonalizing the tight-binding Hamiltonian. In (a) the optimum parameters [Δ_{ab} = 66 meV and Δ_{ac} = 193 meV, and t = 34 meV] have been used,

while in (b) all on-site energies have been chosen to be equal [Δ_{ab} = 0 meV and Δ_{ac} = 0 meV, and t = 34 meV]. The diameter of the circles scales with the weight and red, blue and green circles denote central, edge, and corner molecules respectively. All energies are given relative to the LUMO.

The DFT simulations (Figure S13) and the full tight-binding model (Figure S14 a) yield equivalent pictures with the lowest energy state localized largely on the central molecule and the next four states mainly found on the edge molecules (with a significant additional contribution on the central molecule for the LUMO+4). The LUMO+5 to LUMO+8 are localized largely on the corner molecules, fully consistent with the local densities of states for the full calculations discussed in Figure 3a. Overall, the eigenstates can be described via {1, 4, 4} degenerate modes corresponding to the electrostatic environments. If, however, the onsite energies are all the same, a different degeneracy pattern emerges {1, 2, 3, 2, 1}, which does not fit the DFT calculations.

The impact of the inter-molecular coupling can be understood from the consequences of setting t to 0 meV in the tight-binding model (such that the eigenstates correspond to the onsite-energies). In that case the {1, 4, 4} degeneracy prevails. The overall splitting of the states is somewhat reduced (in the tight-binding model from 276 meV to 193 meV) and also the minor variation of energies within the groups is gone. Nevertheless, **these considerations clearly show that the edge effects prevail also in the absence quantum-mechanical coupling between the individual molecules** (albeit at a somewhat reduced degree).

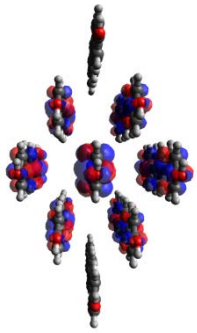
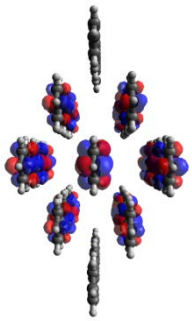
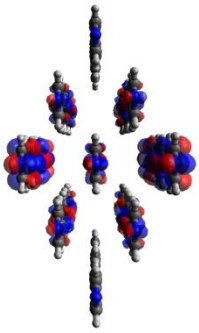
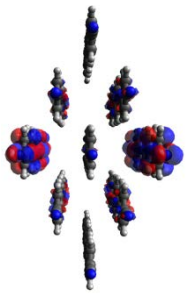
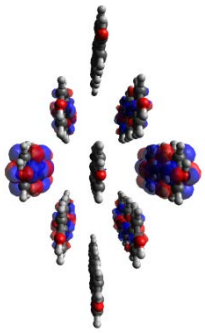
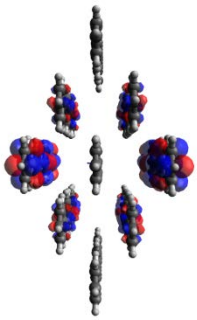
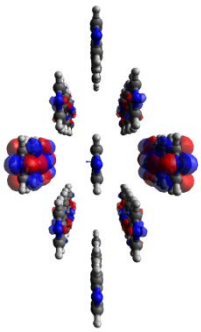
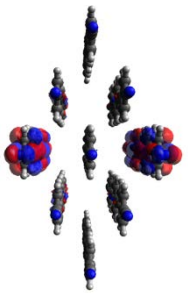
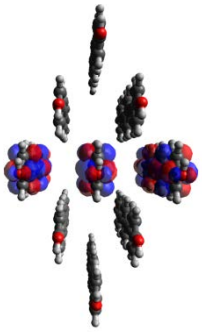
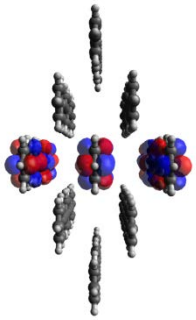
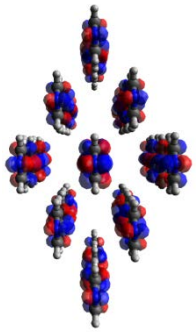
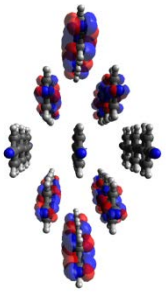
In contrast, “switching off” the onsite-energy differences (i.e., the electrostatic effects; [Δab = 0 meV and Δac = 0 meV), a qualitatively different picture evolves (see Figure S14b): Then the eigenstate degeneracy is {1, 2, 3, 2, 1} (vide supra, as expected from the symmetry of the way the tight-binding model has been set up). The highest and lowest energy groups are delocalized over all molecules, the two groups containing two (degenerate) states are delocalized over corner and edge states, and for the three (degenerate states) at intermediate energy one finds one localized on each sub-group (a, b, and c). The splitting between each sub-group amounts to 48 meV yielding an overall spread of the states of 192 meV, which is somewhat smaller than the 8×t one finds for an infinitely extended 2D tight-binding model.

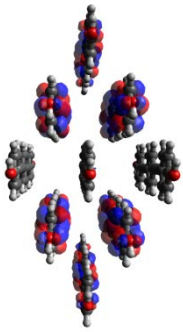
These considerations show that a certain localization of states in specific regions of the cluster is obtained also in the absence of electrostatic effects (e.g., the groups of two-fold degenerate states have no weight on the central molecule). Moreover, also the **energetic splitting of states derived from the molecular LUMOs in the neglecting differences in the onsite energies is comparable to that obtained when they are included. The nature of the states and especially their localization patterns are, however fundamentally affected by differences in the onsite-energies.**

4.2 Calculations on molecular clusters

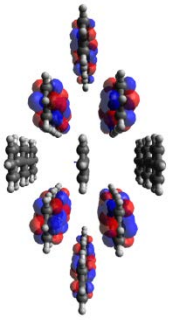
To approach the situation from a different direction, we also studied the electronic eigenstates (orbitals) of 9-molecule clusters bearing tail groups of different polarity. These were constructed using the geometry of the 9-molecule cluster considered in the calculations discussed in Figure 3 of the main paper as a starting point. That cluster in the following table is denoted as “pyridine”. It is characterized by small inward-pointing dipoles at the ends of the molecules (note that these are not the dipoles present in the clusters sandwiched between electrodes as, as they do not account for the charge rearrangements arising from the metal-molecule bonding). Very weak (outward-pointing dipoles) are realized by replacing the pyridine rings by phenyls (where we optimized only the coordinates of the “new” C-H groups; in the same spirit –OH terminal groups have been added in the “phenyl-OH” and “phenyl-CN” clusters. These calculations were performed using Gaussian 09, Revision D.01[4] employing the PBE functional (consistent with the main text) and a 6-31g(d,p) basis set.

Table S4: *cluster eigenstates (energies and isodensity-plots) of 9-molecule clusters bearing different terminal substituents (for details see main text). Note that in Figure 3 local densities of states are plotted, which essentially correspond to charge densities (i.e., squares of orbitals); consequently, localization effects might appear somewhat less pronounced here. The symmetry breaking in some of the close to degenerate orbitals is very likely a consequence of minor imperfections in the geometries (only three of the four standard convergence criteria of Gaussian fulfilled), which can hardly be avoided when optimizing clusters of the size considered here.*

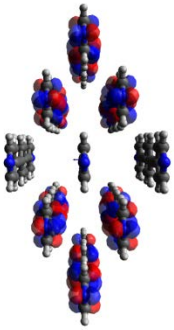
phenyl-OH	phenyl	pyridine	phenyl-CN
LUMO+8: -1,96 eV	LUMO+8: -2,26 eV	LUMO+8: -3,14 eV	LUMO+8: -3,62 eV
			
LUMO+7: -2,00 eV	LUMO+7: -2,29 eV	LUMO+7: -3,15 eV	LUMO+7: -3,63 eV
			
LUMO+6: -2,03 eV	LUMO+6: -2,33 eV	LUMO+6: -3,23 eV	LUMO+6: -3,70 eV
			
LUMO+5: -2,07 eV	LUMO+5: -2,35 eV	LUMO+5: -3,23 eV	LUMO+5: -3,70 eV



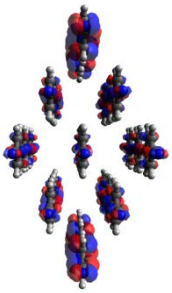
LUMO+4: -2,09 eV



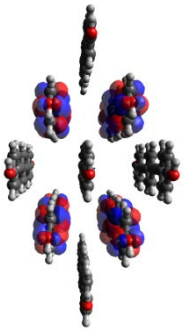
LUMO+4: -2,38 eV



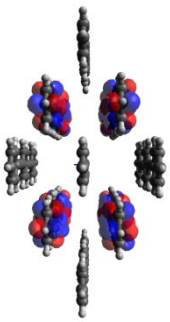
LUMO+4: -3,28 eV



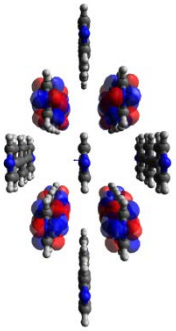
LUMO+4: -3,77 eV



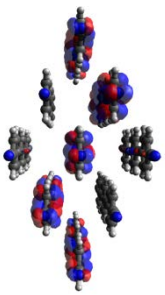
LUMO+3: -2,12 eV



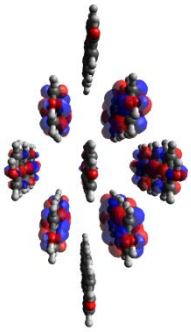
LUMO+3: -2,41 eV



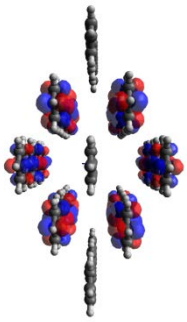
LUMO+3: -3,30 eV



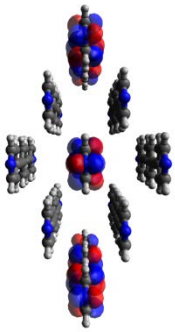
LUMO+3: -3,78 eV



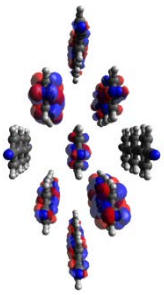
LUMO+2: -2,12 eV



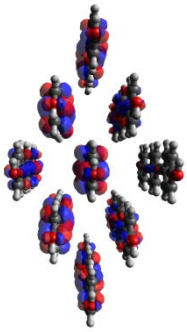
LUMO+2: -2,41 eV



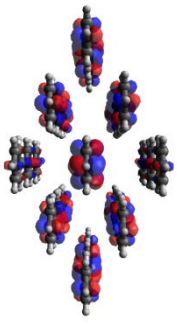
LUMO+2: -3,30 eV



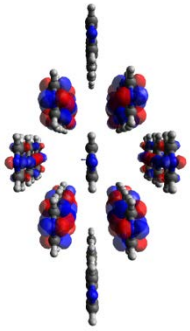
LUMO+2: -3,79 eV



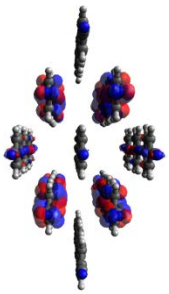
LUMO+1: -2,21 eV



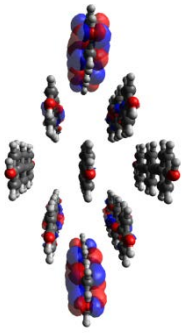
LUMO+1: -2,50 eV

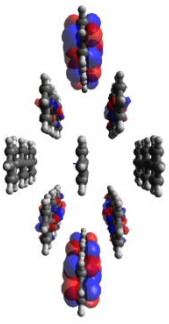


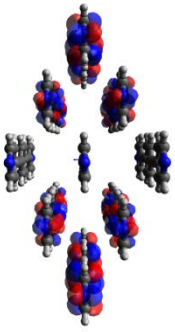
LUMO+1: -3,33 eV

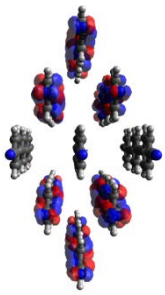


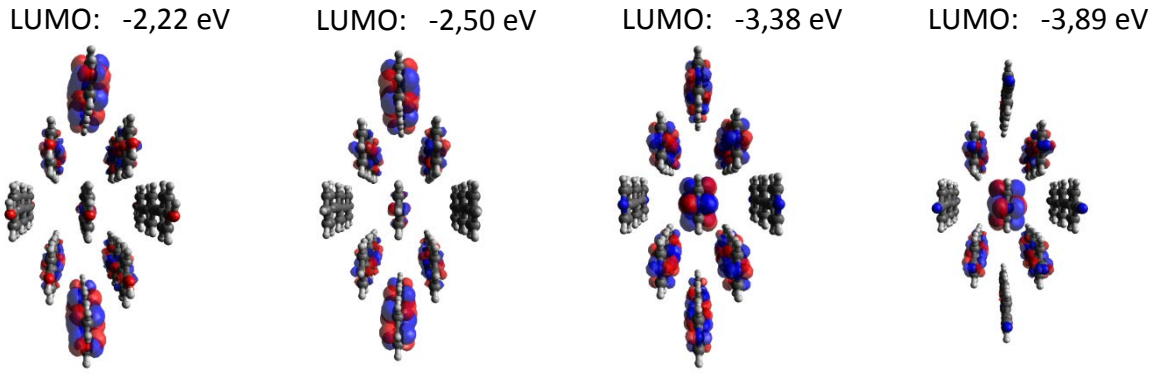
LUMO+1: -3,80 eV











The purpose of this comparison is to see the impact of changes in the electrostatic energy (tuned by the tail-group substituents), while at the same time leaving the quantum-mechanical coupling between the molecules essentially unchanged. The results in Table 3 clearly highlight the impact of electrostatic effects on the localization patterns of the orbitals. For example, while for the "pyridine" cluster and even more pronounced for the more polar "phenyl-CN" cluster the LUMOs are largely localized on the central molecule (consistent with the situation depicted in Figure 3a when including the interfaces), for the other two clusters the LUMO is localized on the top and bottom corner molecules. The situation is reversed when considering the highest orbital derived from the molecular LUMO (the LUMO+8 of the cluster). It is largely localized on the central molecule for the "phenyl-OH" cluster, while it is found at the (left and right) corner molecules for the "pyridine" and "phenyl-CN" clusters. Besides those two examples, the order of orbitals is not exactly inverted between the extreme cases (inward-, respectively, outward-pointing dipoles), as one would expect if all that counted were electrostatic effects. This is, however, not unexpected bearing in mind that in the above calculations the quantum-mechanical coupling between the molecules cannot be switched off like in the tight-binding case discussed earlier. Nevertheless the general trends support the picture discussed already there.

Notably, the orbital structure of the "phenyl-CN cluster" closely resembles that found for the pyridine cluster bonded to the electrodes (Figure 3a), namely that the LUMO is localized on the central molecule, the next four orbitals have dominant weights at the edge molecules (albeit with some corner contributions), while the four highest-lying orbitals are found on the corner molecules. I.e., there is again a 1-4-4 (or maybe rather a 1-4-(2+2)) pattern of eigenstates. This implies that the magnitude of the dipoles of the phenyl-CN terminal groups is similar to that of the pyridines bonded to the Au substrate (i.e., including the bond dipoles).

As far as the energetic spread between the LUMO and the LUMO+8 is concerned it is interesting to note that it is of comparable magnitude for all clusters in spite of the different terminal dipole moments and orientations (0.26 eV for the “phenyl-OH” cluster, 0.25 eV for the “phenyl” cluster, 0.24 eV for the “pyridine” cluster, and 0.27 eV for the “phenyl-CN” cluster). This is in-line with the above-discussed tight-binding model, where the impact of varying the site-energies on the overall spread of energies was also comparably small.

These results further confirm the conclusions from section 4.1 that **an energetic spreading of the states arises already from the quantum-mechanical coupling, while the actual shape and order of the orbitals is determined by electrostatics.**

References

(1) Obersteiner, V.; Egger, D. A.; Heimel, G.; Zojer, E. *J. Phys. Chem. C* **2014**, *118*, 22395–22401.

(2) Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M. *Comput. Phys. Commun.* **2009**, *180*, 2175–2196.

(3) García-Gil, S.; García, A.; Lorente, N.; Ordejón, P. *Phys. Rev. B* **2009**, *79*, 075441.

(4) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.; Izmaylov, A.; Bloino, J.; Zheng, G.; Sonnenberg, J.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J.; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J.; Klene, M.; Knox, J.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Martin, R.; Morokuma, K.; Zakrzewski, V.; Voth, G.; Salvador, P.; Dannenberg, J.; Dapprich, S.; Daniels, A.; Farkas; Foresman, J.; Ortiz, J.; Cioslowski, J.; Fox, D. *Gaussian 09 Revis. D01 Gaussian Inc Wallingford CT* **2013**.