

## Supplementary information

We have carried out *ab-initio* pseudopotential density functional theory (DFT) calculations within the local density approximation (LDA) for a [121]tetramantane molecule on Au(111). We performed these calculations first by using a plane-wave (PW) pseudopotential code to obtain the properties of the isolated molecule. We then modeled the gold surface by three layers of 56 gold atoms each (in a supercell geometry) and calculated the properties of the combined diamondoid/Au(111) system using the SIESTA code, which employs a localized basis set, since the use of plane waves for the combined calculation would be much more expensive. It is very important to verify that the use of localized SIESTA basis set in the calculation of extended molecular orbitals, such as the diamondoid LUMO orbitals, can generate results consistent with the plane-wave method. We have checked this issue very carefully on an isolated diamondoid, and found that the SIESTA results agree with the plane-wave results very well when an appropriate set of ghost atoms are introduced. In this section we present the details of the convergence check.

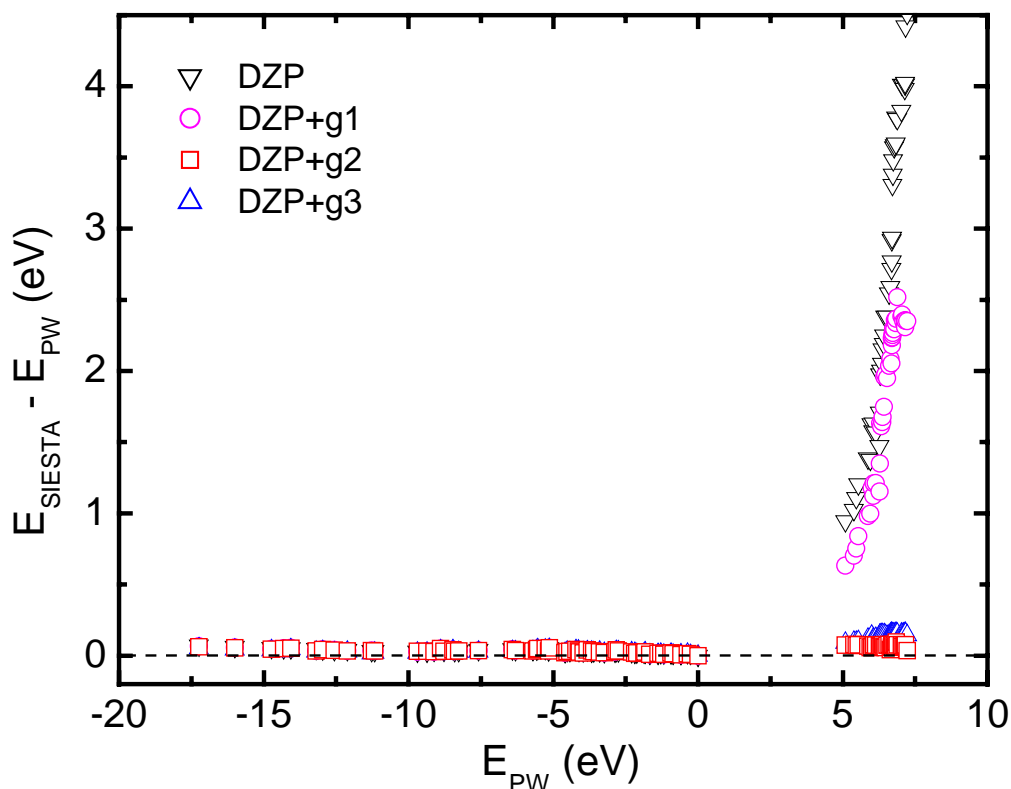


Figure 1: The difference between the energy eigenvalues obtained by SIESTA basis set ( $E_{\text{SIESTA}}$ ) and by plane-wave method ( $E_{\text{PW}}$ ) plotted as a function of the plane-wave eigenvalues ( $E_{\text{PW}}$ ). The four curves represent four different SIESTA bases.

Figure 1 compares the energy eigenvalues of the isolated tetramantane molecule as calculated with plane waves and four different SIESTA bases. The y axis represents the energy difference of the SIESTA eigenvalue ( $E_{\text{SIESTA}}$ ) from the plane wave eigenvalue ( $E_{\text{PW}}$ ), in the unit of eV, for four different SIESTA bases. If the SIESTA bases were complete, then their differences should be zero and all points should be on the dashed line.

The first basis set is the standard SIESTA double- $\zeta$  polarized. The second basis set (DZP+g1) includes a cubic grid of ghost atoms with 1 long ranged *s*-orbital with a spacing of 4 Bohr radii. The third basis set (DZP+g2) is a cubic grid with 2 ghost *s*-orbitals (1 long ranged and one short-ranged), and a spacing equals 4 Bohr radii. The fourth basis set (DZP + g3) is 1 long ranged ghost *s*-orbital in a grid with a spacing of 2 Bohr radii.

As can be seen clearly all basis sets give an excellent description for the occupied orbitals ( $E_{\text{PW}} < 0$ ), which are localized in the molecular region. Plain DZP, however, fails miserably for the unoccupied orbitals ( $E_{\text{PW}} > 0$ ), which are extended outside the molecule and thus cannot be covered by the SIESTA orbitals of the molecule's atoms. This artificially confines the unoccupied wavefunctions and increases their eigenvalues (e.g. the gap is overestimated by about 1 eV).

Increasing the range of the orbitals is not enough to correct for this deviation, since even in this case most of the weight of the orbitals is still in the molecular region. To achieve a better correction, we fill the cell with a grid of ghost *s*-orbitals. This provides more variational freedom and allows the LUMO and higher states to expand. The best scheme is when we add ghost atoms with two *s*-orbitals with different ranges, one short-ranged and one long-ranged. These orbitals are identical to the two that describe the *s*-orbital of the hydrogen atoms (two orbitals are used because it is a double- $\zeta$  set). It can be seen clearly that the (DZP+g2) basis set describes accurately the unoccupied orbital eigenvalues with error less than 0.1 eV compared to PW. Thus we conclude that the (DZP+g2) is the best basis set for the calculation of the extended diamondoid LUMO orbitals. We have used this basis set in the diamondoid on-surface calculations.