# for RSC's Electronic Supplementary Information service

## 1 Analysis on Band Folding

The primitive cell area of GALs is $(m_1m_2+m_1n_2+n_1n_2)$ times as big as that of graphene, and the first Brillouin zone area of GALs is $1/(m_1m_2+m_1n_2+n_1n_2)$ times as big as that of graphene, which means the band have been folded. The bandgap of graphene is closed at $\mathbf{K}_1^0$ and $\mathbf{K}_2^0$ points. We deduce the locations of the $\mathbf{K}_1^0$ and $\mathbf{K}_2^0$ in reciprocal space of GALs after band folding. For graphene, the basis vectors are written as $\{\mathbf{a}_1,\mathbf{a}_2\}$ and the reciprocal basis vectors are denoted by $\{\mathbf{b}_1,\mathbf{b}_2\}$. For GALs, the basis vectors and the reciprocal basis vectors are written as $\{\mathbf{A}_1,\mathbf{A}_2\}$ and $\{\mathbf{B}_1,\mathbf{B}_2\}$, respectively. The relationship between $\{\mathbf{A}_1,\mathbf{A}_2\}$ and $\{\mathbf{a}_1,\mathbf{a}_2\}$ is given by Eq.(1):

$$\begin{cases} \mathbf{A_1} = m_1\mathbf{a_1} + n_1\mathbf{a_2} \\ \mathbf{A_2} = -n_2\mathbf{a_1} + (m_2+n_2)\mathbf{a}_2 \end{cases} \tag{S1}$$

If we assume:

$$\begin{cases} B_\mathbf{1} = x_1b_\mathbf{1} + y_1b_\mathbf{2} \\ B_\mathbf{2} = x_2b_\mathbf{1} + y_2b_2 \end{cases} \tag{S2}$$

And there exists:

$$\begin{cases} a_i \cdot b_j = 2\pi\delta_{\mathrm{ij}} \\ A_i \cdot B_j = 2\pi\delta_{\mathrm{ij}} \end{cases} \tag{S3}$$

Then Eqns. (S1)(S2)(S3) are solved for $\{x_1,y_1,x_2,y_2\}$ and substituted in Eq. (S2), which yields:

$$\begin{cases} B_\mathbf{1} = \frac{(m_2+n_2)b_\mathbf{1}+n_2b_\mathbf{2}}{m_1m_2+m_1n_2+n_1n_2} \\ B_\mathbf{2} = \frac{-n_1b_\mathbf{1}+m_1b_\mathbf{2}}{m_1m_2+m_1n_2+n_1n_2} \end{cases} \tag{S4}$$

Through Eq. (S4), the $k$-points $(k_1,k_2)$ corresponding to $k_1\mathbf{b}_1+k_2\mathbf{b}_2$ can converted into $(K_1,K_2)$ corresponding to $K_1\mathbf{B}_1+K_2\mathbf{B}_2$. This conversion is written as:

$$\begin{cases} K_1 = m_1k_1 + n_1k_2 \\ K_2 = -n_2k_1 + (m_2+n_2)k_2 \end{cases} \tag{S5}$$

Coordinates $(K_1,K_2)$ given by Eq. (S5) are lying out of the first Brillouin zone, which can translocate to some equivalent positions in reciprocal space according to Bloch's theorem. Coordinates $(K_1,K_2)$ of the $\mathbf{K}_1^0$ and$\mathbf{K}_2^0$given by Eq. (S5) are written as Eq. (S6) and Eq. (S7), respectively:

$$(\frac{2}{3}m_1+\frac{1}{3}n_1, \frac{1}{3}m_2-\frac{1}{3}n_2) \tag{S6}$$

$$(\frac{1}{3}m_1+\frac{2}{3}n_1, \frac{2}{3}m_2+\frac{1}{3}n_2) \tag{S7}$$

According to reciprocal space translation symmetry, let the locations after band folding be kept in area of $\Lambda_1\mathbf{B}_1+\Lambda_2\mathbf{B}_2$ ($\Lambda_1,\Lambda_2 \in [0,1)$), then a unique coordinate $(\Lambda_1,\Lambda_2)$ will be got from $(K_1,K_2)$ with Eq. (S8) be satisfied:

$$K_1-\Lambda_1, K_2-\Lambda_2 \in Z \tag{S8}$$

Locations of the $\mathbf{K}_1^0$ and $\mathbf{K}_2^0$, mapped to $\Lambda_1\mathbf{B}_1+\Lambda_2\mathbf{B}_2$ ($\Lambda_1,\Lambda_2 \in [0,1)$) area, can be solved from Eqns. (S6)(S7)(S8). $\mathbf{K}_1^0$ and $\mathbf{K}_2^0$ of $\{[m,n],[m,n]\}$ and $\{[m,m],[0,n]\}$ after band folding are expressed in Table S1.

**Table S1** Locations of the $\mathbf{K}_1^0$ and $\mathbf{K}_2^0$ in $\{[m,n],[m,n]\}$ and $\{[m,m],[0,n]\}$ after band folding.

| | Conditions | $K_1^0$ ($\Lambda_1$,$\Lambda_2$) | $K_2^0$ ($\Lambda_1$,$\Lambda_2$) |
|---|---|---|---|
| $\{[m,n],[m,n]\}$ | $m-n$=3$l$ | $\Gamma$ | $\Gamma$ |
| | $m-n$=3$l$+1 | $K_1$ | $K_2$ |
| | $m-n$=3$l$+2 | $K_2$ | $K_1$ |
| $\{[m,m],[0,n]\}$ | $n$=3$l$ | $\Gamma$ | $\Gamma$ |
| | $n$=3$l$+1 | (0,1/3) | (0,2/3) |
| | $n$=3$l$+2 | (0,2/3) | (0,1/3) |

## 2 Models and Methods

Based on density functional theory (DFT) methods and semi-empirical (SE) methods, the electronic structure calculations are performed with the *ab initio* code package, Atomistix ToolKit (ATK). The generalized gradient approximation (GGA) and the Perdew-Burke-Ernzerhof (PBE) exchange-correlation function are chosen on DFT calculations, with a SingleZetaPolarized basis set, to expanding the wavefunctions. The mesh cutoff of carbon atoms is 150 Rydberg and the tolerance of total energy is $10^{-6}$ Hartree. The non-self-consistent extended Hückel tight-binding model in SE is chosen to deal with GALs involving large number of atoms, with Cerda graphite basis set for carbon and default Hoffmann set for hydrogen. Periodic boundary conditions are used for all calculations, and Brillouin-zone integrations are performed using a $11\times11\times1$ Monkhorst-Pack (MP) grid. Each graphene plane is separated by 30 Å of vacuum to avoid the effect of layers. Our work shows reasonable agreement between DFT and SE calculations on considering the rule of bandgap opening/closing.

To clarify the underlying mechanism of the bandgap opening/closing, we consider the four states, namely $\psi_{1\uparrow}$, $\psi_{2\uparrow}$, $\psi_{1\downarrow}$, $\psi_{2\downarrow}$, at Fermi energy of graphene, where 1, 2 represent the different dirac valleys and $\uparrow$, $\downarrow$ represent the different pseudo-spins. These four states are stable and orthogonal in graphene. However, in GALs, they are no longer orthogonal and stable because of antidots. That is, the wave functions will overlap or change themselves, resulting in energy splitting.

In graphene antidot lattices with zigzag-edged hexagonal holes, if ignored the coupling of valleys state and pseudo-spin states, we can draw a concise and lucid explanation to calculation results of bandgaps. That means the system of four states could be simplified from $|1,2,\uparrow,\downarrow\rangle$ to $|1,2\rangle\otimes|\uparrow,\downarrow\rangle$. In Section 2.3, we build models to analyze how the intervalley scattering effect the restructuring of $\psi_1$ and $\psi_2$, and revealed that the bandgap opening mainly results from the intervalley scattering. In addition, energy splitting of $\psi_\uparrow$ and $\psi_\downarrow$ is also discussed, which is attributed to broken symmetry of the pseudo-spin. Other interactions, for example, whether the wave functions of $\psi_\uparrow$ and $\psi_\downarrow$ overlap or not,are neglected in our analysis.

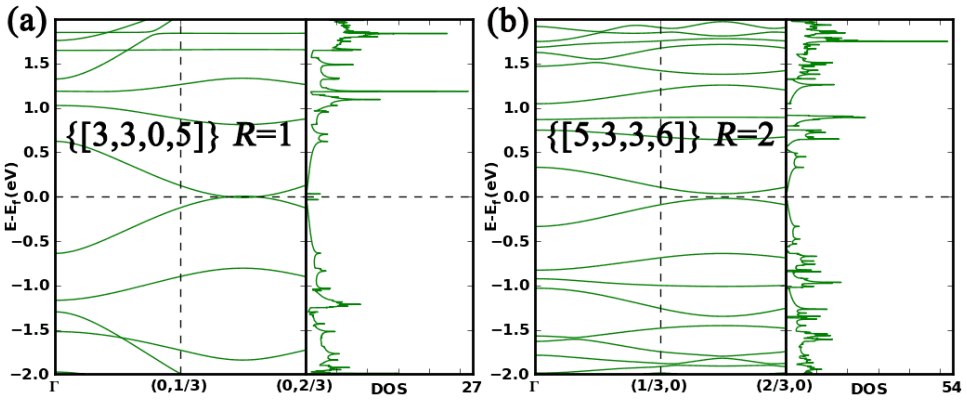


**Fig. S1** Band structures and tetrahedron density of states of (a) ZHS-GALs {[3,3],[0,5] $R$=1 } and (b) ZHS-GALs {[5,3],[3,6] $R$=2 }.

## 3 Simplified *k*-point Paths

Two systems where band gaps close at non-high-symmetry points are given as examples to illustrate how to choose proper simplified *k*-point paths to characterize the electronic structure. According to Table 1, $[r_1,r_2]$ is [0,1] for ZHS-GALs {[3,3],[0,5] *R*=1 }, and [2,0] for ZHS-GALs {[5,3],[3,6] *R*=2 }. So, the $\mathbf{K}_1^0$ and $\mathbf{K}_2^0$ points after band folding are (0,1/3) , (0,2/3) for the former one, and (1/3,0) , (2/3,0) the latter one. DFT results are shown in Fig. S1. The energy interval is set to 0.001 eV when calculate density of states (DOS). The following result illustrated that the simplified *k*-point path $\Gamma\rightarrow \mathbf{K}_1^0 \rightarrow \mathbf{K}_2^0$ is a proper one.

For ZHS-GALs {[3,3],[0,5] *R*=1 }, the bandgap from band calculations would be 0.004 eV with 29 points between (0,1/3) and (0,2/3), and 0.001 eV with 89 points. Under *k*-point density of 80 points per $\text{Å}^{-1}$ ( Monkhorst-Pack grid size being $40\times42\times1$), the gaussian DOS with smearing of 0.005 eV showed a bandgap of 0.015 eV, smearing of 0.001 eV showed 0.053 eV, and the tetrahedron DOS showed 0.062 ev. In this case, the bandgap should be treated as 0 from band calculations. And the DOS calculations require great amount of *k*-points to get an accurate bandgap.

For ZHS-GALs {[5,3],[3,6] *R*=2 }, the bandgap from band calculations would be 0.053 eV with 29 or 89 points between (1/3,0) and (2/3,0). Under *k*-point density of 80 points per $\text{Å}^{-1}$ ( Monkhorst-Pack grid size being $30\times27\times1$), the gaussian DOS with smearing of 0.005 eV showed a bandgap of 0.006 eV, smearing of 0.001 eV showed 0.047 eV, and the tetrahedron DOS showed 0.054 eV. In this case, the bandgap is underestimating through gaussian DOS, and tetrahedron DOS result is consistent with the band calculations with simplified *k*-point paths.