In-plane electric polarization of bilayer graphene nanoribbons by interlayer bias voltage

Ryo Okugawa,1 Junya Tanaka,1 Takashi Koretsune,2 Susumu Saito,1,3 and Shuichi Murakami1,3

1Department of Physics, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8551, Japan
2RIKEN Center for Emergent Matter Science, Hirosawa 2-1, Wako, Saitama 351-0198, Japan
3TIES, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo 152-8551, Japan

We theoretically show that an interlayer bias voltage in the AB-stacked bilayer graphene nanoribbons with armchair edges induces an electric polarization along the ribbon. Both tight-binding and ab initio calculations consistently indicate that when the bias voltage is weak, the polarization shows opposite signs depending on the ribbon width modulo three. This nontrivial dependence is explained using a two-band effective model. A strong limit of the bias voltage in the tight-binding model shows either one-third or zero polarization, which agrees with topological argument.

PACS numbers: 72.80.Vp,73.63.-b,77.22.Ej,73.23.-b

Monolayer graphene nanoribbons (GNRs) show various energy bands depending on the edge orientation and the width of the nanoribbons.1,2 When the GNRs have armchair edges, the energy bands become gapped or gapless, depending on the width. Like monolayer GNRs, the energy bands in the AB-stacked bilayer GNRs also depend sensitively on the edges. Namely, the zigzag bilayer GNRs show localized edge states, whereas the armchair bilayer GNRs vary from insulator to metal by changing the width $N$ in a tight-binding (TB) model. In addition, external fields play scientifically and technologically important roles in atomic-layer materials, such as bilayer graphene. The external electric field opens up the fundamental gap in the AB-stacked bilayer graphene, which otherwise possesses massive and gapless parabolic bands in the low energy region.

In this Letter, we theoretically show that an external interlayer bias voltage in the AB-stacked bilayer GNR with armchair edges induces a polarization along the ribbon direction. We use two methods: calculation on the TB model and ab initio calculations. Both two methods consistently show that when the bias voltage is weak, the polarization shows a nontrivial dependence on the ribbon width, having opposite signs depending on the width modulo three. A strong limit of the bias voltage in the tight-binding model shows either one-third or zero polarization, which agrees with topological argument.

First we numerically calculate the polarization for a spinless TB model:

$$H = \sum_{i,j} t_{ij} c_i^\dagger c_j + \frac{U}{2} \sum_i \xi_i c_i^\dagger c_i. \tag{1}$$

The first term describes the hoppings with the amplitude $t_{ij}$, for which we only consider the nearest-neighbor intralayer hopping $t$ and the interlayer hopping $t_\perp$ within a “dimer”. Here, we set $t_\perp = 0.13t \ (t > 0)$ according to Ref.3. The second term represents the interlayer bias $U$ and $\xi_i$ takes +1 (-1) for the upper (lower) layers.
From the TB model, we calculate the electronic contribution of the polarization $P$ in terms of the Berry connection within the modern theory of polarization.

It is calculated as a change of polarization $P(U) - P(0)$ by changing the interlayer bias voltage $U$. This calculation works only for insulators, and therefore we restrict ourselves to the insulating GNRs whose width is $N = 3l$ or $3l + 1$ ($l$: integer). Because $P(0) = 0$ by inversion symmetry, we obtain $P(U)$ numerically.

Using this method, we find that in the bilayer GNRs with the armchair edges, the polarization arises for nonzero interlayer bias voltage $U$. Figures 2(a)(b) are our numerical results for various widths $N$. In reality, feasible values of $U$ may be limited to about $|U| < 0.15t$; nevertheless we show the results for much larger $U$ in the figure, to show consistency for the large $U$ limit. Interestingly, the behavior of the polarization is classified into two classes, $N = 3l$ and $N = 3l + 1$. For $N = 3l$, the polarization goes to zero for $U \to \pm \infty$, while for $N = 3l + 1$ the polarization goes to $\pm e/3$, where $-e$ ($e > 0$) represents the electron charge. Furthermore, the slope around $U \sim 0$ has opposite signs between the two classes. The dependence of the asymptotic behavior at $U \to \pm \infty$ is the lattice constant. Here the polarization is $P/N$ of minibands from a finite-size effect.

The strong limit of the interlayer bias voltage $U$ totally agrees with our numerical calculations. It is calculated as a change of polarization $P(U) - P(0)$ by changing the interlayer bias voltage $U$. This calculation works only for insulators, and therefore we restrict ourselves to the insulating GNRs whose width is $N = 3l$ or $3l + 1$ ($l$: integer). Because $P(0) = 0$ by inversion symmetry, we obtain $P(U)$ numerically.

Using this method, we find that in the bilayer GNRs with the armchair edges, the polarization arises for nonzero interlayer bias voltage $U$. Figures 2(a)(b) are our numerical results for various widths $N$. In reality, feasible values of $U$ may be limited to about $|U| < 0.15t$; nevertheless we show the results for much larger $U$ in the figure, to show consistency for the large $U$ limit. Interestingly, the behavior of the polarization is classified into two classes, $N = 3l$ and $N = 3l + 1$. For $N = 3l$, the polarization goes to zero for $U \to \pm \infty$, while for $N = 3l + 1$ the polarization goes to $\pm e/3$, where $-e$ ($e > 0$) represents the electron charge. Furthermore, the slope around $U \sim 0$ has opposite signs between the two classes. The dependence of the asymptotic behavior at $U \to \pm \infty$ is the lattice constant. Here the polarization is $P/N$ of minibands from a finite-size effect.

The slope, we construct a simple two-band (2B) effective Hamiltonian for the weakly biased GNRs, by retaining only the highest occupied band and the lowest unoccupied band. To this end, we begin with the analytic forms of the eigenstates of the TB model at $k = 0$ and $U = 0$. The eigenvalue equation at $k = 0$, $U = 0$ is written

$$
\begin{bmatrix}
0 & t(2\cos\theta + 1) \\
\pm t_\perp & B^\pm
\end{bmatrix} = \varepsilon^\pm \begin{bmatrix} A^\pm \\ B^\pm \end{bmatrix}
$$

where $A^\pm$ corresponds to the sum and the difference between the amplitudes at the A and B’ sublattices, respectively, and $B^\pm$ is defined similarly for the B and A’ sublattices (see the Supplemental Material [1] for details). $\theta$ is a phase difference of an electronic wave between the neighboring rows, forming a standing wave in the ribbon. Its eigenvalues are

$$
\varepsilon^\pm a = \pm \frac{t_\perp}{2} + q\sqrt{\left(\frac{t_\perp}{2}\right)^2 + t^2(2\cos\theta + 1)^2},
$$

where $q = \pm 1$. From the boundary condition, we get $\theta = \theta_N^r = \frac{r\pi}{N+1}$, $r = 1, 2, \ldots, N$. Here, response of the polarization $P$ to an external perturbation is given by the Berry curvature $\mathbf{R}$ [4, 12]. Therefore, the eigenstates close to $k = 0$, where the band structure has a direct gap when $U \sim 0$, contributes considerably to the polarization. Hence, from the analytic forms of the eigenstates of the TB model at $k = 0$ and $U = 0$ [1], we retain only the lowest unoccupied state $|+\rangle$ and the highest occupied state $|-\rangle$. Their energy eigenvalues are given by $\pm g_0$, where $g_0 = -t_\perp/2 + d$, $d = \sqrt{(t_\perp/2)^2 + t^2(2\cos\theta_{2l+1}^N + 1)^2}$, and $\theta_{2l+1}^N = \frac{\pi(2l+1)}{(N+1)}$ for both $N = 3l$ and $N = 3l + 1$.

By using these two eigenstates, we construct a 2B model which describes the energy bands around the Fermi energy when $k \sim 0$ and $U \ll t_\perp, t_\parallel$. The 2B Hamiltonian to the first order in $k$ and $U$ is derived as

$$
H_{\text{eff}} = h_1 U \sigma_x + h_2 k \sigma_y + g_0 \sigma_z,
$$

where $h_1 = t_\perp/(4d)$, $h_2 = a t^2(\cos\theta_{2l+1}^N + 1)(2\cos\theta_{2l+1}^N + 1)/(3d)$ and $\sigma_{x,y,z}$ are the Pauli matrices for the space spanned by the eigenstates $|\pm\rangle$ [11]. We note that width...
dependence appears through $\theta_{n, l}^{N}$. From the 2B Hamiltonian, we calculate the polarization $P(U)$ for small $U$, as shown in Fig. 5. They well agree with the results of the DFT model in the $U \sim 0$ region, including the sign of the slope, for $N > 20$. For $N < 20$, the gap in the TB model is non-monotonous as a function of the interlayer bias $U$; for small $U$ the gap decreases as a function of $U$. It is not reproduced in the 2B model where the gap always increases with $U$. This leads to differences between the two models.

The width dependence is understood from the analytic formula for the slope of $P(U)$:

$$\frac{\partial P}{\partial U} = \frac{ie}{2\pi} \int_{-\pi}^{\pi} dk \frac{\langle u_{k-} | \frac{\partial H}{\partial U} | u_{k+} \rangle}{(E_{k-} - E_{k+})^2} + c, \xi(5)$$

Here $E_{kj}$ and $|u_{kj}\rangle$ are the eigenvalue and the eigenstate of the 2B Hamiltonian for the $j$ band. Therefore, the slope at $U = 0$ is

$$\frac{\partial P}{\partial U} \bigg|_{U=0} = -\frac{e}{2a} \hbar^2 \frac{1}{g_0 \sqrt{g_0} + (h_2^2 \pi/a)^2},$$

and its sign is given by $-\text{sgn}(h_2)$. From $2\pi/3 - \theta_{n, l}^{N+1} = \mp \pi/(3N + 3)$ for $N = 3l$ and $3l + 1$, the sign of the slope is negative for $N = 3l$ and positive for $N = 3l + 1$, in agreement with the results for various widths of GNRs. Asymptotic form for a wider ribbon is evaluated as

$$\frac{\partial P}{\partial U} \bigg|_{U=0} \sim \pm \frac{3t_1 e}{8\pi^2 t} (N + 1)^2.$$ (7)

where the signs $\pm$ is $-$ for $N = 3l$ and $+$ for $N = 3l + 1$.

These behaviors are confirmed by ab initio calculations based on density functional theory (DFT). We perform the electronic structure calculation of hydrogen terminated AB-stacked bilayer GNRs within the local-density approximation (LDA) [10, 11] based on DFT using QUANTUM ESPRESSO package [13]. We use ultrsoft pseudopotentials [14] and plane-wave basis sets to describe the charge densities and wave functions with cutoff energies of 30 Ry and 300 Ry, respectively. The supercell approach is used and the distances of neighboring bilayer GNRs along the $x$-axis and the $z$-axis are at least 10 and 30 Å, respectively. The geometries are fully optimized.

To experimentally measure this proposed effect, one
needs a bilayer nanoribbon with well-defined edges and width. For single-layer graphene nanoribbons, well-defined edge orientations have been demonstrated [20–23], and it might be realized also for bilayer graphene. For the bilayer graphene, interlayer electric field up to 0.3V/Å has been achieved [24], and therefore the proposed effect with polarization up to $\sim -0.12e$ per spin for $N = 15$ is expected to be realizable experimentally. We also have calculated the effect of periodic modulations of the width to check the edge disorder effect via supercell approach and confirmed that the polarization survives the weak modulations considered [11]. Nevertheless, since the effect is sensitive to the ribbon width, the proposed effect will disappear in the presence of strong disorder. We note here that the in-plane polarization by an interlayer bias can be expected for a wide variety of atomic-layer compounds, as long as the symmetry criterion for its emergence is satisfied. As an example, a bilayer armchair ribbon of transition metal dichalcogenides in the 2H stacking satisfies this criteria. Moreover, our calculation show induced polarization in AA'-stacked bilayer boron nitride nanoribbons [11]. Such a wide choice of candidate materials provides us with many chances for experimental verifications of our theory.

To conclude, we theoretically show that the AB-stacked graphene nanoribbon with armchair edges has a polarization along the ribbon direction, when interlayer bias voltage is applied. This is shown both by the simple tight-binding model and the ab initio calculations. In particular, the linear response to the interlayer voltage shows different signs for the cases $N = 3l$ and $N = 3l + 1$, and it is fully understood by means of a simple two-band model.

This work is partially supported by Grant-in-Aid from MEXT, Japan (No. 26287062, No. 25107005 and No. 25104711), JSPS Research Fellowships for Young Scientists, and MEXT Elements Strategy Initiative to Form Core Research Center (TIES).

[1] M. Fujita, K. Wakabayashi, K. Nakada, K. Kusakabe, J. Phys. Soc. Jpn. 65, 1920 (1996).
[2] K. Wakabayashi, M. Fujita, H. Ajiki, M. Sigrist, Phys. Rev. B 59, 8271 (1999).
[3] E. V. Castro, N. M. R. Peres, J. M. B. Lopes dos Santos, A. H. Castro Neto, F. Guinea, Phys. Rev. Lett. 100, 026802 (2008).
[4] B. Sahu, H. Min, A. H. MacDonald, S. K. Banerjee, Phys. Rev. B 78, 045404 (2008).
[5] E. McCann, V. I. Fal’ko, Phys. Rev. Lett. 96, 086805 (2006).
[6] E. McCann, Phys. Rev. B 74, 161403 (2006).
[7] H. Min, B. Sahu, S. K. Banerjee, A. H. MacDonald, Phys. Rev. B 75, 155115 (2007).
[8] R. D. King-Smith, D. Vanderbilt, Phys. Rev. B 47, 1651 (1993).
[9] R. Resta, Ferroelectrics 136, 51 (1992).
[10] R. Resta, Rev. Mod. Phys. 66, 899 (1994).
[11] See the Supplemental Material.
[12] M. V. Berry, Proc. Roy. Soc. London Ser A 392, 45 (1984).
[13] D. M. Ceperley, B. J. Alder, Phys. Rev. Lett. 45, 566 (1980).
[14] J. P. Perdew, A. Zunger, Phys. Rev. B 23, 5048 (1981).
[15] P. Giannozzi et al., J. Phys.: Condens. Matter 21, 395502 (2009).
[16] D. Vanderbilt, Phys. Rev. B 41, 7892 (1990).
[17] N. Marzari, D. Vanderbilt, Phys. Rev. B 56, 12847 (1997).
[18] I. Souza, N. Marzari, D. Vanderbilt, Phys. Rev. B 65, 035109 (2001).
[19] A. A. Mostofi et al., Comput. Phys. Commun. 178, 685 (2008).
[20] G. Z. Magda et al., Nature 514, 608 (2014).
[21] M. Y. Han, B. ¨Ozyilmaz, Y. Zhang, P. Kim, Phys. Rev. Lett. 98, 206805 (2007).
[22] J. Cai et al., Nature 466, 470 (2010).
[23] D. Kosynkin et al., Nature 458, 872 (2009).
[24] Y. Zhang et al., Nature 459, 820 (2009).
FORMULA OF THE POLARIZATION IN TERMS OF THE BLOCH WAVEFUNCTIONS

In our calculation of the polarization $P$ in terms of the Bloch wavefunctions, we used the formula (2–4)

$$ P = -\frac{ie}{2\pi} \int_{-\pi}^{\pi} \frac{dk}{2\pi} \sum_{n} \langle u_{kn} | \frac{\partial}{\partial k} | u_{kn} \rangle, $$

where $|u_{kn}\rangle$ is the Bloch wavefunction satisfying the cell-periodic gauge condition

$$ u_{k,n}(r) = e^{iGy} u_{k+N,n}(r), $$

and the summation is taken over the occupied states below the Fermi energy. Here $k$ is the Bloch wavenumber and $G \equiv 2\pi/a$ is a reciprocal lattice vector. In numerical calculation, the differentiation in terms of $k$ in Eq. (S1) should be replaced by a difference in $k$. Such a formula with this replacement is discussed in detail in Ref. 3, and we followed this formalism for the calculation of the polarization.

POLARIZATION CALCULATED FROM THE TIGHT-BINDING MODEL FOR VARIOUS WIDTHS

We show numerical results of the polarization induced by the interlayer bias for graphene nanoribbons (GNRs) with various widths $N$, calculated from the tight-binding model. Some results are shown in Fig. 2 in the main text, and we show more examples for various $N$ in Fig. S1. For $N = 3l$ the results are shown in Fig. S1a, c ($N = 9, 15, 21$), and in Fig. S1a, b ($N = 27, 39$). For $N = 3l + 1$ the results are shown in Fig. S1b, d ($N = 10, 16, 22$), and in Fig. 2c, d ($N = 28, 40$). In the wide range of $U$ (Fig. S1), the polarization $P$ oscillates as a function of $U$, which is attributed to crossings of minigaps. There are more oscillations for larger $N$, which reflects the fact that there are a larger number of minibands for wider ribbons. On the other hand, the oscillation amplitude becomes gradually smaller for wider ribbons (see Fig. 2) because the contribution from each miniband becomes relatively smaller. On the other hand, in the regime $U \sim 0$, we showed from the two-band model that the polarization is linear in $U$ with its slope scales with $(N+1)^2$. This is roughly reproduced for the results shown in Fig. S1c and d.

![Graph showing polarization vs interlayer bias](image)

FIG. S1. Polarization induced by the interlayer bias $U$ for GNRs with various widths $N$, calculated from the tight-binding model. It is shown as a function of the interlayer bias $U$. a and c are for the class $N = 3l$ and b and d are for the class $N = 3l + 1$.

It may look strange that for a large $N$ limit, i.e. the 2D graphene limit, the polarization has a different asymptotics for $N = 3l$ and $N = 3l + 1$. It is in fact reasonable; the polarization per area is proportional to $P$ divided by the width, and therefore goes to zero for $N \to \infty$. 
EIGENVALUES AND EIGENVECTORS OF THE TIGHT-BINDING MODEL

Here we derive the two-band low-energy effective Hamiltonian from the eigenstates of the tight-binding model of the bilayer GNR with armchair edge, without the interlayer bias voltage:

\[ H_t = \sum_{<i,j>} t_{ij} c_i^\dagger c_j. \] (S3)

Firstly, to obtain the two-band Hamiltonian, we diagonalize the tight-binding model at \( k = 0 \) according to Ref. 1. We set the eigenvector of the Hamiltonian

\[ |\Phi(k = 0)\rangle = \sum_{m=1}^{N} a_m |mA\rangle + \sum_{m=1}^{N} b_m |mB\rangle + \sum_{m=1}^{N} a'_m |mA'\rangle + \sum_{m=1}^{N} b'_m |mB'\rangle, \] (S4)

where \( A \) and \( B \) represent sublattices in the lower layer, and \( A' \) and \( B' \) in the upper layer. Here, \( a_m, b_m, a'_m, \) and \( b'_m \) are expansion coefficients. From the tight-binding model at \( k = 0 \), we obtain

\[
\varepsilon a_m = t(b_{m-1} + b_m + b_{m+1}), \quad \varepsilon b_m = t(a_{m-1} + a_m + a_{m+1}) + t\perp a'_m, \\
\varepsilon a'_m = t(b'_{m-1} + b'_m + b'_{m+1}) + t\perp b_m, \quad \varepsilon b'_m = t(a'_{m-1} + a'_m + a'_{m+1}),
\]

where \( \varepsilon \) represents the energy eigenvalue. To solve the above equations (S5)-(S8), we introduce new coefficients

\[ a_m^{\pm} = \frac{1}{\sqrt{2}} (a_m \pm b'_m), \quad \beta_m^{\pm} = \frac{1}{\sqrt{2}} (b_m \pm a'_m). \] (S9)

Then, we obtain

\[ \varepsilon a_m^{\pm} = t (\beta_m^{\pm} + \beta_{m+1}^{\pm}), \quad \varepsilon \beta_m^{\pm} = t (a_m^{\pm} + a_{m+1}^{\pm}) \pm t\perp \beta_m^{\pm}. \] (S10, S11)

Since \( \alpha_0^{\pm} = \beta_0^{\pm} \), the solutions have the form \( \alpha_m^{\pm} \propto A^{\pm} \sin(m\theta) \) and \( \beta_m^{\pm} \propto B^{\pm} \sin(m\theta) \), where \( \theta \) is a constant \((0 < \theta < \pi)\). We rewrite the equations (S10) and (S11) in the matrix form

\[ \begin{pmatrix} 0 & t(2\cos\theta + 1) \\ t(2\cos\theta + 1) & \pm t\perp \end{pmatrix} \begin{pmatrix} A^{\pm} \\ B^{\pm} \end{pmatrix} = \varepsilon^{\pm} \begin{pmatrix} A^{\pm} \\ B^{\pm} \end{pmatrix}, \] (S12)

and its eigenvalues are obtained analytically.

\[ \varepsilon^{\pm,q} = \pm \frac{t_\perp}{2} + q \sqrt{\left(\frac{t_\perp}{2}\right)^2 + t^2(\cos\theta + 1)^2}, \] (S13)

where \( q = \pm 1 \). From the boundary condition, the coefficients must vanish when \( m = 0 \) and \( N + 1 \) and we get

\[ \theta_r = \frac{r}{N + 1}\pi, \quad r = 1, 2, \ldots, N. \] (S14)

Thus, we get \( 4N \) eigenvalues and eigenvectors,

\[ \varepsilon_{r,q}^{\pm} = \pm \frac{t_\perp}{2} + q d_r, \] (S15)

\[ \begin{pmatrix} A_{r,q}^{\pm} \\ B_{r,q}^{\pm} \end{pmatrix} = \frac{1}{\sqrt{2d_r(d_r \pm q t_\perp)}} \begin{pmatrix} \mp \frac{t_\perp}{2} + q d_r \\ t(2\cos\theta_r + 1) \end{pmatrix}. \] (S16)

We put \( d_r = \sqrt{\left(\frac{t_\perp}{2}\right)^2 + t^2(\cos\theta_r + 1)^2} \) for notational simplicity. These energy eigenvalues can become zero only when \( \cos\theta_r = -1/2 \). When \( N = 3l + 2 \) it can be satisfied for \( r = 2l + 2 \), and the energy bands are gapless at \( k = 0 \) because \( \varepsilon_{2l+2}^{\pm} = \varepsilon_{2l+2}^{-\pm} = 0 \). On the other hand, when \( N = 3l \) or \( 3l + 1 \), \( \cos\theta_r = -1/2 \) cannot be satisfied and the energy bands are gapped. Therefore, the polarization \( P_y(U) \) can be defined in the nanoribbons with widths \( N = 3l \) or \( 3l + 1 \). In these cases, the energy eigenvalues closest to zero are \( \varepsilon_{2l+1}^{-\pm} \) and \( \varepsilon_{2l+1}^{\pm} \), and the corresponding eigenstates
\(|+\rangle \equiv |r = 2l + 1, - , +\rangle \) and \(|-\rangle \equiv |r = 2l + 1, + , -\rangle \) considerably contribute to the polarization. For brevity, we write \(\pm g_0 = \varepsilon_{2l+1}^{\pm}\) and
\[
g_0 = -\frac{t_1}{2} + d_{2l+1}, \quad (S17)
\]
\(|\pm \rangle = \frac{1}{\sqrt{2}} \sum_{m=1}^{N} \sin(m\theta_{2l+1}) \left( A_{2l+1}^{\pm} |mA\rangle \mp |mB\rangle \right) + B_{2l+1}^{\pm} |mB\rangle \mp |mA\rangle \right) \). \quad (S18)
Hereafter, we omit subscripts \(2l + 1\) except for that in \(\theta_{2l+1}\); for example, we write \(d = d_{2l+1}\).

**CONSTRUCTION OF THE TWO-BAND HAMILTONIAN**

To elucidate the polarization in the weak interlayer bias voltage, we construct a two-band low-energy effective Hamiltonian for the space spanned by \(|\pm\rangle\). Therefore, we add the interlayer bias voltage \(U\) to the tight-binding model as a perturbation;
\[
H_U = \frac{U}{2} \sum_i \xi_i c_i^\dagger c_i, \quad (S19)
\]
We retain terms up to the linear order in \(k\) and the nonzero matrix elements to this order are given by
\[
\langle mA | H_t | nB \rangle = \langle mA' | H_t | nB' \rangle = t \delta_{m,n} \delta_{m,n+1} + \frac{a}{3} \delta_{m,n} - \frac{1}{2} \frac{1}{3} \delta_{m,n+1}, \quad (S20)
\]
\[
\langle mA | H_t | nB \rangle = t \delta_{m,n} \delta_{m,n+1}, \quad (S21)
\]
\[
\langle mA | H_U | nA \rangle = \langle mA | H_U | nB \rangle = -U \delta_{m,n}, \quad (S22)
\]
\[
\langle mA' | H_U | nA' \rangle = \langle mA' | H_U | nB' \rangle = U \delta_{m,n}. \quad (S23)
\]
Therefore, the nonzero matrix elements of \(H_t\) and \(H_U\) are
\[
\langle \pm | H_t | \pm \rangle = \pm g_0, \quad (S24)
\]
\[
\langle + | H_t | - \rangle = i k t^2 \frac{a}{3d} (1 - \cos \theta_{2l+1})(2 \cos \theta_{2l+1} + 1), \quad (S25)
\]
\[
\langle + | H_U | - \rangle = \frac{U t_1}{4d}. \quad (S26)
\]
Hence, we obtain the two-band Hamiltonian \(H_{\text{eff}} = H_t + H_U\),
\[
H_{\text{eff}}(k, U) = h_1 U \sigma_x + h_2 k \sigma_y + g_0 \sigma_z, \quad (S27)
\]
where
\[
h_1 = t_1 \frac{U}{4d}, \quad h_2 = -t^2 \frac{a}{3d} (1 - \cos \theta_{2l+1})(2 \cos \theta_{2l+1} + 1). \quad (S28)
\]
The eigenvalues and eigenstates of this two-band Hamiltonian are
\[
E_{k,\pm} = \pm \sqrt{(h_1 U)^2 + (h_2 k)^2 + g_0^2} = \pm g, \quad (S29)
\]
\[
|u_{k,\pm}\rangle = \frac{1}{\sqrt{2g(g \pm g)}} \left( \begin{array}{c} g_0 \pm g \\ h_1 U \pm i h_2 k \end{array} \right). \quad (S30)
\]
POLARIZATION FROM THE TWO-BAND HAMILTONIAN

We focus on the region of the weak interlayer bias voltage to clarify the difference of the slope of the polarization at $U \sim 0$ for two classes $N = 3l$ and $N = 3l + 1$. The polarization $P$ is given by

$$P = \int_{0}^{U} dU' j(U'),$$

(S31)

where

$$j(U) = \frac{ie}{2\pi} \int_{-\pi}^{\pi} dk \sum_{\text{occ.}} \sum_{\text{unocc.}} \langle u_{kn} | \frac{\partial H}{\partial k} | u_{km} \rangle \langle u_{km} | \frac{\partial H}{\partial U} | u_{kn} \rangle + \text{c.c.}$$

(S32)

Here, $j(U) = \partial P/\partial U$, and $n$ and $m$ are band indices for occupied bands and unoccupied bands, respectively. $E_{kj}$ and $|u_{kj}\rangle$ are the $j$th eigenvalue and the eigenstate of the Hamiltonian. In the present case $n = -$ and $m = +$, and therefore $P$ and $j(U)$ are calculated by using the two-band Hamiltonian derived in section .

$$P(U) = -\frac{e}{2\pi} \arctan \left( \frac{h_1 h_2 U \pi / a}{g_0 \sqrt{g_0^2 + (h_1 U)^2 + (h_2 \pi / a)^2}} \right),$$

(S33)

$$j(U) = \frac{e}{2a} h_1 h_2 \frac{g_0}{((h_1 U)^2 + g_0^2) \sqrt{(h_1 U)^2 + g_0^2 + (h_2 \pi / a)^2}}$$

(S34)

In particular, the slope of the polarization at $U = 0$, i.e. $j(0)$, is given by

$$j(0) = -\frac{e}{2a} h_1 h_2 \frac{1}{g_0 \sqrt{g_0^2 + (h_2 \pi / a)^2}}.$$

(S35)

Because $t_\perp > 0$, $h_1$ and $g_0$ are positive, and we have

$$\text{sgn}(j(0)) = -\text{sgn}(h_2) = \text{sgn} \left( \frac{2}{3} \pi - \theta_{2l+1} \right).$$

(S36)

Here,

$$\frac{2}{3} \pi - \theta_{2l+1} = \begin{cases} -\frac{\pi}{3(N+1)} & N = 3l \\ \frac{\pi}{3(N+1)} & N = 3l + 1 \end{cases}$$

(S37)

Therefore, the sign of $j(0) = \partial P/\partial U|_{U=0}$ is given by

$$\left. \frac{\partial P}{\partial U} \right|_{U=0} \begin{cases} < 0 & N = 3l \\ > 0 & N = 3l + 1 \end{cases}$$

(S38)

which well agrees with the numerical results of the tight-binding model. Asymptotic behavior of $j(0)$ (Eq. S35) is evaluated for large $N$, where $g_0 \ll h_2 \pi / a$;

$$j(0) = -\text{sgn}(h_2) \frac{h_1 e}{2\pi g_0} \sim \pm \frac{t_\perp e}{8\pi d(d - \frac{2\pi}{3})},$$

(S39)

where $\pm$ is $-$ for $N = 3l$ and $+$ for $N = 3l + 1$. By using equation $\text{(S37)}$ it is approximated as

$$j(0) \sim \pm \frac{3t_\perp e}{8\pi^3 d^2} (N + 1)^2.$$

(S40)

EFFECT OF PERIODIC MODULATION OF THE RIBBON WIDTH

We numerically calculate the polarization in bilayer GNRs with weak periodic modulations of the ribbon widths by using the tight-binding model when the interlayer bias voltage is weak. We consider two cases of periodic modulations by changing the width of each layer in various ways.
Firstly, we discuss an effect of the difference of the widths of the upper and lower layers. Figure S2 shows the polarization in the armchair bilayer GNRs when the upper and lower layers have the different widths. Then, we calculate the polarization by changing the width for the upper layer $N_U$ while fixing that for the lower layer $N_L$, except that $N_L = 2$ or $N_U = 2 \pmod{3}$ since the energy bands are gapless at $U = 0$. When $N_L = N_U$, the system is the perfectly stacked bilayer armchair GNRs in Fig. 1. The results in Fig. S2b and c correspond to the polarization in Fig. S2 have the same sign regardless of the width of the upper layer $N_U$. Furthermore, even if the bilayer is composed of two layers with the width $0$ and $1 \pmod{3}$, we find that the sign of the slope is equal to that of the perfectly stacked bilayer GNRs with the narrower width although the magnitude becomes small. Therefore, when the widths of the upper and lower layers are different, the polarization behaves like the perfectly stacked armchair GNRs with the width $\min(N_U, N_L)$.

Secondly, we calculate the polarization of bilayer GNRs when the widths of the upper and lower layers alternates between two values $N_1$ and $N_2$ ($N_1 \geq N_2$), as shown in Fig. S3a having no dangling bonds. In this system, the primitive translation vector doubles. In particular, when $N_1 = N_2$, the system corresponds to the armchair bilayer GNRs. In this case, when we calculate the polarization, we change only $N_1$ and fix $N_2$. The results of the polarization for various widths $(N_1, N_2)$ are shown in Fig. S3. Figure S3b and c show the polarization for $N_2 = 12 \equiv 0$ and $N_2 = 13 \equiv 1 \pmod{3}$, respectively. As a result, as $N_1$ becomes larger, we find that the magnitudes of the polarization for $N_2 \equiv 0 \pmod{3}$ are enhanced, while those for $N_2 \equiv 0 \pmod{3}$ are suppressed. Nevertheless, the sign of the slope of the polarization is unchanged from that of the perfect armchair GNRs with the width $N_1 = N_2$. Therefore, we can see that the polarization in the weak interlayer bias voltage is dominated by the narrow part of GNRs, which is similar to the previous case.

From the above results, we find that small variations of the width do not affect the sign of the slope of the polarization in the weak interlayer bias voltage. In other words, even though the edges of the bilayer graphene nanoribbons are not completely perfect, the nontrivial dependence of the polarization on the width appear like the perfectly stacked bilayer GNRs with armchair edges. Thus, the in-plane polarization in response to the interlayer voltage survives even if the edges have weak periodic modulation of the ribbon width, as long as the energy bands are gapped. Nevertheless, since the effect is sensitive to ribbon width, the proposed effect will disappear in the presence of strong disorder.

FIG. S2. Polarization in the bilayer composed of two layers with different widths. a shows structure of the GNRs with the different upper and lower layer. The unit cell is described by the red dashed-line box. b and c show the polarization for $N_U = 12$ and $N_U = 13$, respectively. b is for $N_U = 3l$, and c is for $N_U = 3l + 1$. 

Secondly, we calculate the polarization of bilayer GNRs when the widths of the upper and lower layers alternates between two values $N_1$ and $N_2$ ($N_1 \geq N_2$), as shown in Fig. S3a having no dangling bonds. In this system, the primitive translation vector doubles. In particular, when $N_1 = N_2$, the system corresponds to the armchair bilayer GNRs. In this case, when we calculate the polarization, we change only $N_1$ and fix $N_2$. The results of the polarization for various widths $(N_1, N_2)$ are shown in Fig. S3. Figure S3b and c show the polarization for $N_2 = 12 \equiv 0$ and $N_2 = 13 \equiv 1 \pmod{3}$, respectively. As a result, as $N_1$ becomes larger, we find that the magnitudes of the polarization for $N_2 \equiv 0 \pmod{3}$ are enhanced, while those for $N_2 \equiv 0 \pmod{3}$ are suppressed. Nevertheless, the sign of the slope of the polarization is unchanged from that of the perfect armchair GNRs with the width $N_1 = N_2$. Therefore, we can see that the polarization in the weak interlayer bias voltage is dominated by the narrow part of GNRs, which is similar to the previous case.

From the above results, we find that small variations of the width do not affect the sign of the slope of the polarization in the weak interlayer bias voltage. In other words, even though the edges of the bilayer graphene nanoribbons are not completely perfect, the nontrivial dependence of the polarization on the width appear like the perfectly stacked bilayer GNRs with armchair edges. Thus, the in-plane polarization in response to the interlayer voltage survives even if the edges have weak periodic modulation of the ribbon width, as long as the energy bands are gapped. Nevertheless, since the effect is sensitive to ribbon width, the proposed effect will disappear in the presence of strong disorder.
FIG. S3. **Polarization in the bilayer consisting of the stacked monolayer with two widths.** a shows the structure of the bilayer nanoribbons with two widths. The red dashed-line box describes the unit cell. b and c are examples of the polarization for \( N_U = 12 \) and \( N_U = 13 \), respectively. b is for \( N_U = 3l \), and c is for \( N_U = 3l + 1 \).

**POLARIZATION IN THE BILAYER BN NANORIBBON**

We explained the emergence of the polarization in bilayer graphene nanoribbons by symmetry argument. Therefore, other nanoribbons of atomic-layer compounds can have a finite polarization along the ribbon direction in response to the interlayer voltage, when the symmetry criterion for such a response is satisfied. To confirm this, we compute the polarization of hydrogen terminated bilayer BN nanoribbons from first-principles calculations. Here, we consider so-called AA’-stacked bilayer BN nanoribbons with the armchair edges and the geometry is fully optimized (see Fig. S3 (a)). As in the case of the bilayer armchair GNRs, \( xz \)-mirror symmetries are broken in this structure. When the interlayer bias voltage is zero, inversion symmetry is preserved and the polarization is zero. The interlayer voltage breaks the inversion symmetry, leading to nonzero polarization along the ribbon, as we see in the following. Figure S3 (b) shows calculated in-plane polarization as a function of the interlayer voltage. As we can see, a finite polarization appears as expected. Note that the size of the polarization is rather small compared to those in the GNRs. One reason is that the band gaps of these nanoribbons are relatively large (\( \sim 4.3 \) eV both for \( N = 7 \) and \( N = 9 \) at \( U = 0 \)).

[1] Sahu, B., Min, H., MacDonald, A. H. & Banerjee, S. K. Energy gaps, magnetism, and electric-field effects in bilayer graphene nanoribbons. Phys. Rev. B 78, 045404 (2008).

[2] Resta, R. Theory of the electric polarization in crystals.

[3] King-Smith, R. D. & Vanderbilt, D. Theory of polarization of crystalline solids. Phys. Rev. B 47, 1651–1654 (1993).

[4] Resta, R. Macroscopic polarization in crystalline dielectrics: the geometric phase approach. Rev. Mod. Phys. 66, 899–915 (1994).
FIG. S4. **Polarization in the BN nanoribbons.** a Top view of the hydrogen terminated bilayer BN nanoribbon with the armchair edges for $N = 7$. B (N) atoms on the lower layer are just below the N (B) atoms on the upper layer. b Electric field dependence of the polarization obtained by DFT calculations for bilayer BN nanoribbons with the armchair edges.