nature portfolio

https://doi.org/10.1038/s41586-022-05685-y

Supplementary information

The quantum twisting microscope

In the format provided by the authors and unedited

Supplementary materials for:

The Quantum Twisting Microscope

A. Inbar, J. Birkbeck, J. Xiao, T. Taniguchi, K. Watanabe, B. Yan, Y. Oreg, Ady Stern, E. Berg, and S. Ilani

Table of Content

S1. Fabrication of vdW-device-on-tip 3

S2. Fabrication of the flat devices 4

S3. Configuring a commercial AFM for QTM experiments 5

S4. Calculation of momentum resolved tunneling in a QTM junction. 6

S4.1. Solving the electrostatics of the QTM junction self-consistently 6

S4.2 Momentum-resolved tunneling current calculation 7

S4.3 Band structure of TBG based on Bistrizer-MacDonald model 9

S4.4 Nordheim-Fowler Contribution and WSe_2 barrier breakdown 9

S5. Understanding MLG-MLG momentum resolved tunneling experiments 10

S5.1 Momentum-resolved onset condition 11

S5.2 Nesting condition 12

S6. Additional experiments: back-gate dependence of MLG-MLG k-resolved tunneling 14

S7. Electron lifetime in MLG-MLG k-resolved tunneling experiment 18

S8. Understanding MLG-TBG momentum resolved tunneling experiments 19

S8.1 First Toy Model – tracing energy bands with a Dirac cone 19

S8.2 Second toy model - explaining the k-resolved 'onset' features 20

S8.3 Temperature dependence of 'onset' and 'Dirac matching' features 22

S8.4 Comparing the second toy model with the experiment 22

S8.5 The full calculation – origin of the features in the MLG-TBG k-resolved tunneling 23

S9. MLG-TBG tunneling experiments 25

S10. Line traces from the MLG-MLG and MLG-TBG tunneling experiments 26

S11. Tunneling experiments of TBG with different twist angles 27

S12. Experiments with different tips and different measurement schemes 27

S12.1 Tip Shapes 28

S12.2 Comparing measurements of the same TBG sample with two different tips 28

S12.3 Comparing measurements with different contact modalities 28

S13. Quantum coherence of the QTM junction and a lower bound on the coherence length 29

S14. Additional details 30

S15. Supplementary Video 1 – Imaging TBG bands under pressure 32
S16. Supplementary Video 2 – Tracing the TBG flat bands with the Dirac point of MLG 32
S17. Supplementary Video 3 – Tracing the MLG probe bands using the TBG Dirac point. .. 32
S18. Supplementary Video 4 – Energy-momentum conserving states in MLG-MLG junction vs. twist angle.. 33
S19. Supplementary Video 5 – Energy-momentum conserving states in MLG-TBG junction vs. bias.. 33

S1. Fabrication of vdW-device-on-tip

To fabricate the vdW-device-on-tips, we start with commercial tipless AFM cantilevers with a nominal spring constant of $K = 40\ [N/m]$, length $L = 232\ [\mu m]$, width $W = 37[\mu m]$ and thickness $T = 7[\mu m]$. We coat the bottom side of the cantilevers with 4/100nm Cr/Au using an e-beam evaporator. This coating serves as the electrical connection for the active vdW layer and the buried graphite, such that they are at the same electrical potential. Next, we deposit a ~ 1.2-1.6 um tall platinum (Pt) pyramid with a ~ 2 um square base using focused ion beam (FIB) deposition (Fig. S1a-c). The nominal size of the pyramids' apex is ~ 50nm.

To assemble the vdW-device-on-tip, we use the polymer membrane transfer technique[1]. Briefly, we spin-coat polypropylene carbonate (PPC) onto a Si/(90nm) SiO_2 wafer, followed by exfoliating graphene/hBN crystals. We then cut a window into a piece of adhesive tape and use it to pick up the PPC film with the target crystal. The crystal is then aligned and transferred on top of the platinum pyramid, followed by melting (~ 120 °C) and washing-off (acetone) of the PPC film to leave behind the vdW crystal. We assemble the vdW heterostructure by successive transferring and cleaning of each

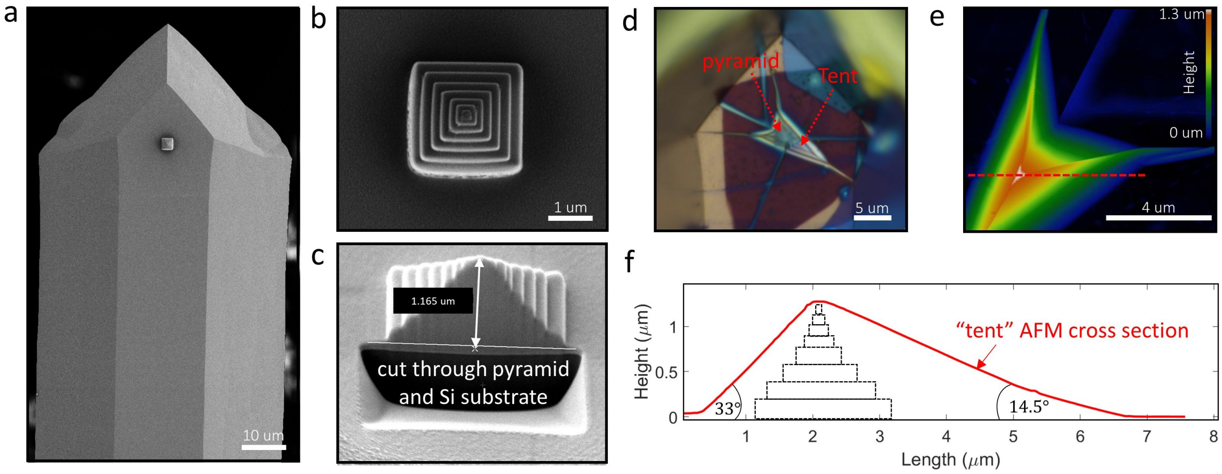

Figure S1: **QTM tip characterization**. **a.** SEM image of FIB deposited Pt pyramid on a tipless AFM cantilever **b.** Top view of Pt pyramid **c.** Cross-section of a Pt pyramid cut using local FIB cutting **d.** Optical image of a MLG/hBN/graphite vdW heterostructure on top of a pyramid **e.** AFM topography of vdW heterostructure in panel d. **f.** A linecut from the AFM topography of the vdW tent (red, taken along the dashed red line in panel e) overlayed with a schematic drawing of the Pt pyramid to scale. The top part of the pyramid is smaller than the plateau formed at the top part of the vdW tent.

layer. For example, for the experiments done in this paper where the active vdW layer on the tip is monolayer graphene, we start by transferring a graphite flake, ~ 20-40 nm thick, on top of the Pt pyramid, followed by an insulating hBN flake, ~ 20-40 nm. These two layers act as an atomically flat mechanical support for the active vdW layer (monolayer graphene), which is transferred at the last step. The thicker underlayers also define the shape of the plateau obtained at the tip apex. By varying the thickness of these layers, we can obtain plateaus of variable sizes in the range of 50nm to 1μm.

S2. Fabrication of the flat devices

In this paper, we performed two types of experiments: In-situ twistronics experiments, in which two MLG layers are brought into direct contact (without an insulating barrier in between) and the conductance of the combined system is measured as a function of the twist angle (Fig. 1h, main text), and momentum resolved tunneling, in which we put a WSe_2 barrier between the MLG probe and the system (MLG or TBG). Here, rotation is used to scan the momentum axis. For the former, the devices consist of MLG placed on top of an hBN substrate and a graphite gate (Fig. S2b). For the latter, we use MLG or TBG on top of hBN and graphite, but now cap it from the top with 2-3 layer WSe_2 (Fig. S2c, d). All our devices are assembled using the standard dry transfer technique. Briefly, a polypropylene carbonate (PPC) coated polydimethylsiloxane (PDMS) is used to pick up the hBN, MLG (or TBG created using laser cut and stack method), and WSe_2 layers. We then flip the stacks by depositing them onto a PMMA-coated PDMS, followed by placing them on the bottom graphite gate. The final heterostructure is transferred on a ~100um wide and ~80um tall Si/SiO_2 pillar, with pre-pattered gold contacts reaching close to its edge (Fig. S2a). The reason for placing the stack on a pillar rather than on a flat Si/SiO_2 wafer is that this provides mechanical and geometrical flexibility for the tip rotation with respect to the sample, preventing them from touching each other at any point other than at the tip's apex. Next, we create the electrical contacts by transferring thin (~10nm) flakes of graphite to bridge the pre-patterned contacts and the MLG or TBG. Fig. S2b,c,d show the three variants of heterostructures used in the main text (MLG without tunnel barrier, MLG with tunnel barrier, TBG with tunnel barrier). After assembly, the heterostructures are cleaned by scanning them with an AFM in contact mode to remove polymer residues.

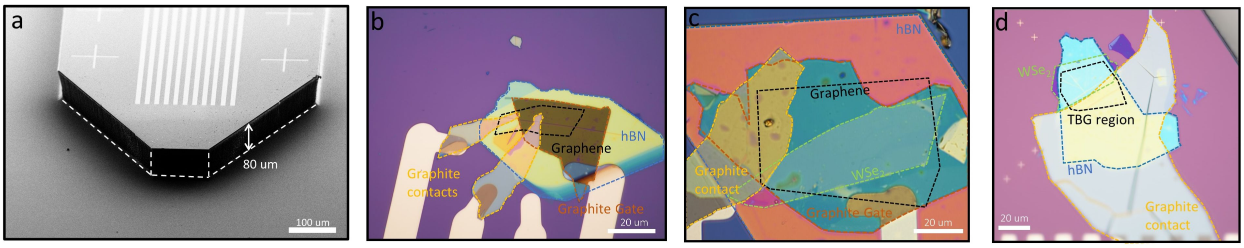

Figure S2: **QTM flat side substrate and devices**. **a.** SEM image of a Si/SiO_2 pillar with pre-patterned gold contacts. **b.** Optical image of a MLG/hBN/graphite sample used for in-situ twistronics experiments (Fig. 1h). **c.** Optical image of a WSe_2/MLG/hBN/graphite sample used for momentum resolved MLG-MLG tunneling experiments (Fig. 2). **d.** Optical image of a WSe_2/TBG/hBN sample used for momentum resolved MLG-TBG tunneling experiments (Fig. 3 and 4).

S3. Configuring a commercial AFM for QTM experiments

Our QTM setup comprises of two parts (1) A custom-built rotating stage housing the bottom devices (2) commercial atomic force microscope (Bruker Icon) for performing the AFM operations (Force feedback and lateral positioning).

The custom-built rotating stage comprises (from bottom to top): a goniometer, piezoelectric rotator (attocube), X and Y piezoelectric nanopositioners (attocube) and a sample holder with multiple electrical contacts (Fig. S3c). A conventional AFM has a large inclination angle, $\sim 12°$, which is problematic for the QTM experiments. QTM tips have much shorter pyramids ($\sim 1 - 1.5\mu m$) than conventional AFM tip pyramids ($\sim 20\mu m$). In addition, for the QTM tip to work, the cantilever should have a few tens of μm of flat cantilever surface in all directions around the pyramid, such that the vdW heterostructure can adhere well to the cantilever all around the pyramid. With conventional $12°$ inclination, such a tip will not reach the sample surface because the front of the cantilever will touch it instead. The QTM experiments are therefore designed such that the inclination of the tip with respect to the sample is small, $\sim 1°$. In a commercial AFM we cannot change the inclination angle of the tip since the laser illumination and quadrature detector are designed to work at this angle. Instead, we tilted the sample by $\sim 11°$ such that the relative inclination between sample and tip is reduced to about $1°$. This is achieved using the goniometer at the bottom of the custom-built rotating stage (Fig. S3c). Importantly, the rotator and X,Y translators are placed on top of this goniometer, assuring that the rotation axis is perpendicular to the sample plane, and the sample surface rotates parallel to itself to an excellent approximation.

The piezoelectric rotator has an optical encoder allowing for closed-loop angular control with ~1 milli-degree precision. The two linear nanopositioners, mounted on top of the piezoelectric rotator, position the device's point-of-interest in the center of rotation. This custom stage is attached using a rigid frame to an existing AFM stage (because of height limitations, we had to connect it on a side facet of the AFM stage). This adapter allows us to use the integrated optics and all standard AFM operations. The AFM enables us to bring the two vdW surfaces (in the tip and the sample) into contact. Additionally, we use the active force feedback to maintain a constant force during rotation and measurement with a sub-nano-Newton precision.

With this QTM setup, we employ two modalities for the rotation measurements. (1) **Continuous contact**: Here, we bring the tip and sample together and rotate whilst maintaining the two surfaces in contact throughout the entire experiment. (2) **"Woodpecker"**: Here, for every point in the measurement as a function of rotation angle, we bring the tip into contact with the sample, perform the measurement,

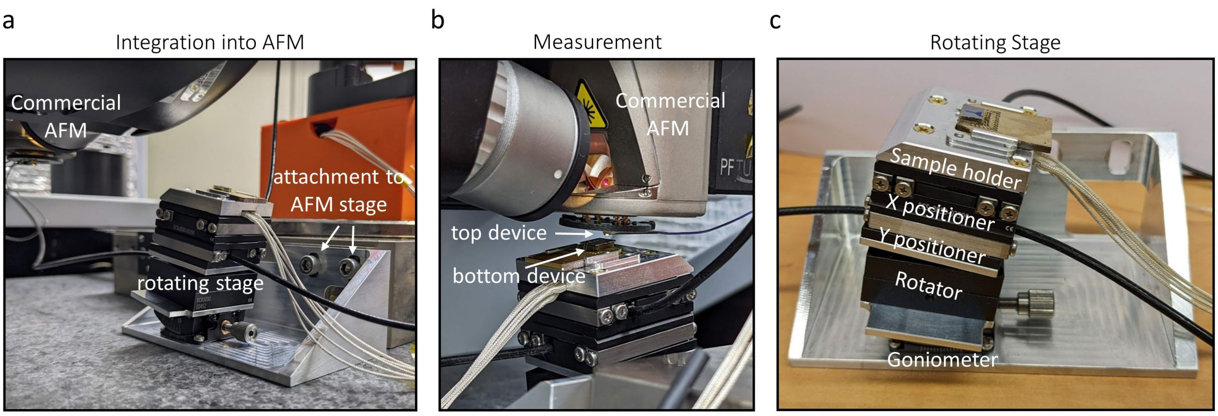

Figure S3: **Room-temperature QTM measurement setup a.** Custom-built rotator stage integrated into the side of a commercial AFM stage. **b**. Side view of the rotator stage facing the AFM scanner and optics. **c**. The components of the custom-built rotating stage.

disengage, then rotate by a small angle, and then repeat this sequence to get the full measurement. We note that we obtain almost identical results from both modalities, and both are highly reproducible.

S4. Calculation of momentum resolved tunneling in a QTM junction.

In this section we present the model used for calculating the theory panels in Figs. 2,3 in the main text. In addition to the parameters that are scanned in the experiment that are shown in the main text – the twist angle, θ, and bias, V_b, we also include in the model the effect of back gate voltage, V_{BG}, such that it also describes the experiments at finite gate voltages shown in section S6 below. The tunneling current is calculated separately for each combination of the parameters θ, V_b, V_{BG}. We first determine the chemical potential of top and bottom layers and the electrostatic potential shift between them (μ_T, μ_B and ϕ) by solving the electrostatic equations (S4.1, S4.2) self-consistently with the equations-of-state of the individual layers: $\mu_T(n_T)$ and $\mu_B(n_B)$. This is described in section S4.1 below. We then sum up the momentum-conserving tunneling rates for all states within the energy window opened by the bias, between μ_T and μ_B. This is described in subsection S4.2 below. The equation of state for the TBG is obtained from the Bistrizer-Macdonald continuum model, detailed in section S4.3. Finally, we add a Norhdeim-Fowler[2] term due to the breakdown of the WSe_2 barrier, described in section S4.4. All expressions include the effects of finite temperature and finite lifetime.

S4.1. Solving the electrostatics of the QTM junction self-consistently

We describe the electrostatics of the QTM tunnel junction by a three-plate-capacitor model. The first two plates are the top and bottom active vdW layers (in the tip and sample), which are separated by a tunnel barrier (2 or 3 layers of WSe_2). The third layer is the graphite back-gate, separated by an hBN spacer from the bottom vdW system. We derive the two following equations[3]:

$$eV_b = \mu_B - \mu_T - \frac{e^2 n_T}{C_g}, \tag{S4.1}$$

$$eV_{BG} = \mu_B + \frac{e^2(n_B + n_T)}{C_{BG}}, \tag{S4.2}$$

where V_b is the applied bias between the top and bottom layer, V_{BG} is the back gate voltage with respect to the bottom layer, and μ_B, n_B, μ_T and n_T are the chemical potential and carrier densities of the bottom and top layer respectively. The geometric capacitance per unit area between top and bottom layers, C_g, and between the back-gate and bottom layer, C_{BG}, are given by:

$$C_g = \frac{\varepsilon_{WSe_2}\varepsilon_0}{d_{WSe_2}} \tag{S4.3}$$

$$C_{BG} = \frac{\varepsilon_{BN}\varepsilon_0}{d_{BN}} \tag{S4.4}$$

where d_{WSe_2} is the barrier thickness (taken to be $0.58nm$ per monolayer of WSe2), d_{BN} is the hBN thickness ($70nm$ in the MLG-MLG experiment) in Fig. 2, and $\varepsilon_{WSe_2} \approx 4.63, \varepsilon_{BN} \approx 3.2$ are taken to be the corresponding dielectric constants. For MLG, the dependence of carrier density on chemical potential is given by $n_{Gr} = sign(\mu_{Gr})\frac{\mu_{Gr}^2}{\pi\hbar^2 v_F^2}$, where v_F is the Fermi velocity. For TBG, $n_{TBG}(\mu_{TBG})$ is calculated numerically from the Bistritzer-MacDonald continuum model (section S4 below). We solve together the above equations for μ_B, μ_T and ϕ, self-consistently. Due to the low carrier density of graphene, the quantum capacitance of the QTM junction is generally of a similar order of magnitude as its geometric capacitance. Consequently, the applied bias changes the chemical potentials of the layers as well as the electrostatic potential between them, $\phi = V_b - (\mu_B - \mu_T)$. In Fig. S4 we plot the dependence of ϕ, μ_B and μ_T vs. V_b calculated for the parameters of the MLG-MLG experiment in Fig. 2 in the main text. We can see that in this experiment, a change of V_b indeed leads to a same order of magnitude of change in all these quantities. It can also be seen that $\mu_B = -\mu_T$ due to the assumption of overall charge neutrality, which describes well the experiment in Fig. 2 with $V_{BG} = 0$. As explained in section S5, experimentally, we can directly determine the dependence $\phi(V_b)$ from the nesting line in Fig. 2c. Specifically, we see the nesting feature in Figs. 2c, f fits well the experiment if we take $C_g = 2.35\mu F/cm^2$, consistent with thickness and dielectric constant values mentioned above.

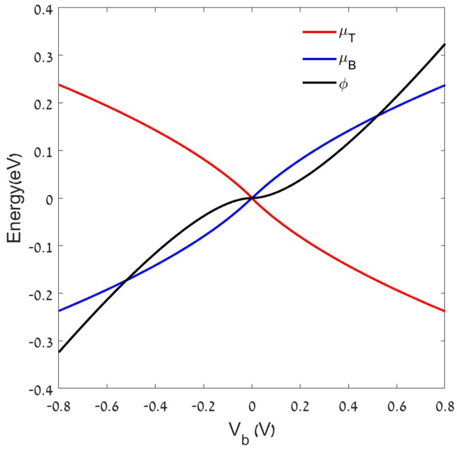

Figure S4. The bias dependence of top/bottom chemical potential $\mu_{T/B}$ and electrostatic potential ϕ for MLG-MLG tunneling junction.

S4.2 Momentum-resolved tunneling current calculation

After solving the electrostatics and obtaining the chemical potential and band alignment between top and bottom layers, we proceed to calculate the tunneling current. The theoretical model is based on the Bistritzer-MacDonald expression for momentum resolved tunneling between two layers[4]. The model assumes that the tunneling is well described by Fermi's golden rule, namely, that a carrier tunnels coherently only once across the junction and not multiple times back and forth. This assumption is valid for all the experiments in the paper that involve a tunneling barrier between the two vdW layers and is implied from the fact that the resistance of the tunnel junction is much larger that the resistance along the vdW layers themselves. With these approximations, the tunneling current is given by:

$$I = \frac{eg_s}{2\pi}\int dE\left(f_B(E) - f_T(E+\phi)\right) \times \sum_{\boldsymbol{k_T},\boldsymbol{k_B}} \left|T^{\alpha\beta}_{\boldsymbol{k_T}\boldsymbol{k'_B}}\right|^2 A_B(\boldsymbol{k'_B}, E) A_T(\boldsymbol{k_T}, E+\phi) \tag{S4.5}$$

Here $g_s = 2$ accounts for the spin degeneracy. The functions $f_B(E)$ and $f_T(E+\phi)$ are the Fermi-Dirac distribution functions of the bottom and top layers, while ϕ is the electrostatic potential that shifts the relative position of energy bands. $T^{\alpha\beta}_{\boldsymbol{k_T}\boldsymbol{k'_B}}$ is the tunneling matrix element between a state with momentum $\boldsymbol{k_T}$ on the top layer to one with momentum $\boldsymbol{k'_B}$ in the rotated Brillouin zone of bottom layer. The functions A_B and A_T are the spectral functions for the bottom and top layers: $A_T(E, \boldsymbol{k_T}) = \frac{2\gamma_T}{(E-\epsilon_T(\boldsymbol{k}))^2+\gamma_T^2}$ where $\epsilon_T(\boldsymbol{k})$ is the energy bands of the top layer. γ_T is the corresponding inverse electron lifetime (we elaborate on this lifetime in section S7). Similar formula applies to the bottom layer.

Eq. (S4.5) requires three conditions to be met for a tunneling event to be allowed. The Pauli principle require the tunneling electron to originate from a full state and arrive to an empty state. This is imposed by the difference between the two Fermi functions. Energy conservation requires the energy difference between the state of origin and the state of destination to differ by ϕ. This is encoded in the energy arguments of the Fermi functions and the spectral functions. And momentum conservation requires the states of origin and destination to have the same momentum. These three requirements are typically met by a curve in momentum space, whose characteristics are determined by μ_b, μ_t, ϕ and, most importantly, by the energy spectra of the two layers. The contribution of the states along that curve is determined by the density of states and by the value of the tunneling matrix element at that momentum.

Following Bistritzer-MacDonald[4], the tunneling matrix element is given by:

$$T^{\alpha\beta}_{\boldsymbol{k_T k'_B}} = \frac{1}{\Omega_0}\sum_{s,s'}\sum_{\boldsymbol{G_T},\boldsymbol{G_B}} t_{\boldsymbol{k'_B}+\boldsymbol{G_B}} \left\langle u_{\boldsymbol{k_T},s}\middle|T_{\boldsymbol{G_T},\boldsymbol{G_B}}\middle|u_{\boldsymbol{k'_B},s'}\right\rangle \delta_{\boldsymbol{k_T}+\boldsymbol{G_T},\boldsymbol{k'_B}+\boldsymbol{G_B}} \tag{S4.6}$$

where $\left|u_{\boldsymbol{k_T},s}\right\rangle$ and $\left|u_{\boldsymbol{k'_B},s'}\right\rangle$ are the wavefunctions on the top and bottom layers, s and s' are the band indices, primed wave vectors are rotated with the twist angle, $\boldsymbol{G_T}, \boldsymbol{G_B}$ are reciprocal lattice vectors of the top and bottom layers, α, β are sublattice indices, and $t_{\boldsymbol{k'_B}+\boldsymbol{G_B}}$ is the Fourier transform of the finite-range interlayer hopping amplitude (discussed below). In a matrix form in the sublattice basis the tunneling matrix element is given by:

$$\mathrm{T}_{\mathbf{G_1},\mathbf{G_2}} = \begin{pmatrix} \mathrm{e}^{\mathrm{i}\mathbf{G_T}\tau_t^A - \mathrm{i}\mathbf{G_B}\tau_b^A} & \mathrm{e}^{\mathrm{i}\mathbf{G_T}\tau_t^A - \mathrm{i}\mathbf{G_B}\tau_b^B} \\ \mathrm{e}^{\mathrm{i}\mathbf{G_T}\tau_t^B - \mathrm{i}\mathbf{G_B}\tau_b^A} & \mathrm{e}^{\mathrm{i}\mathbf{G_T}\tau_t^B - \mathrm{i}\mathbf{G_B}\tau_b^B} \end{pmatrix} e^{i\boldsymbol{G_T}\cdot\boldsymbol{d_T} - i\boldsymbol{G_B}\cdot\boldsymbol{d_B}} \tag{S4.7}$$

Here $\boldsymbol{\tau}_t^A$ and $\boldsymbol{\tau}_t^B$ are the sublattice positions of atoms A and B within the unit cell of the top layer. $\boldsymbol{d_T}$ is the lateral displacement vectors for these two layers. The BM model assumes that the tunneling amplitude $t_{\boldsymbol{k'_B}+\boldsymbol{G_B}}$ decays exponentially with the absolute value of momentum measured with respect to the origin in momentum space. Thus, we consider tunneling events within the first Brillouin zone and approximate $t_{\boldsymbol{k'_B}+\boldsymbol{G_B}} \approx t_0$, which is independent of momentum. Consequently, in the summation of $\boldsymbol{G_T}$ and $\boldsymbol{G_B}$ reciprocal lattice vectors to obtain $T_{\boldsymbol{G_T},\boldsymbol{G_B}}$ only three processes are relevant:

$$\mathrm{T}_{\mathbf{1}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} e^{i\boldsymbol{G_T}\cdot\boldsymbol{d_T} - i\boldsymbol{G_B}\cdot\boldsymbol{d_B}} \tag{S4.8}$$

$$\mathrm{T}_{\mathbf{2}} = \begin{pmatrix} 1 & \mathrm{e}^{+i\frac{2\pi}{3}} \\ \mathrm{e}^{-i\frac{2\pi}{3}} & 1 \end{pmatrix} e^{i\boldsymbol{G_T}\cdot\boldsymbol{d_T} - i\boldsymbol{G_B}\cdot\boldsymbol{d_B}} \tag{S4.9}$$

$$\mathrm{T}_{\mathbf{3}} = \begin{pmatrix} 1 & \mathrm{e}^{-i\frac{2\pi}{3}} \\ \mathrm{e}^{+i\frac{2\pi}{3}} & 1 \end{pmatrix} e^{i\boldsymbol{G_1}\cdot\boldsymbol{d_T} - i\boldsymbol{G_2}\cdot\boldsymbol{d_B}} \tag{S4.10}$$

In addition, the tunneling amplitude t_0 is affected by the energy of tunneling electrons and the tunneling barrier shape $V(z)$. This is described by the WKB approximation[2,3]:

$$V(z) = \Delta_c - \frac{|\phi|}{d} z \tag{S4.11}$$

$$t_0 = \exp\left\{-\frac{1}{\hbar}\int_0^d \left(2m^*(V(z)-E)\right)^{\frac{1}{2}} dz\right\}$$
$$= \exp\left\{-\frac{1}{\hbar}\frac{2d}{3|\phi|}(2m^*)^{\frac{1}{2}}\left((\Delta_c - E)^{\frac{3}{2}} - (\Delta_c - E - |\phi|)^{\frac{3}{2}}\right)\right\} \quad \text{(S4.12)}$$

Δ_c is the distance from the Dirac point to the conductance band edge of WSe_2 relative to the graphene. E is the energy of the tunneling electron. d is the barrier width and m^* is the effective mass in the z direction of WSe_2, where we assume a parabolic dispersion. We consider here only the conduction and not the valence band edge since it is significantly closer in energy to the Dirac point[5]. We also ignore the in-plane momentum dependence of Δ_c.

S4.3 Band structure of TBG based on Bistrizer-MacDonald model

We use the continuum model to calculate the band structure of TBG[6]. For each graphene layer, the lattice vectors are $\boldsymbol{a_1} = \sqrt{3}a_0\left(\frac{1}{2},\frac{\sqrt{3}}{2}\right), \boldsymbol{a_2} = \sqrt{3}a_0\left(-\frac{1}{2},\frac{\sqrt{3}}{2}\right)$. $a_0 = 0.142nm$ is the carbon-carbon bond length and the sublattice vectors A, B are $\boldsymbol{\delta_A} = a_0(0,0), \boldsymbol{\delta_B} = a_0(0,1)$. The Hamiltonian is given by

$$H = H_b + H_t + H_{inter} \quad \text{(S4.13)}$$

$$H_b = \sum_{q,\xi,s} a^\dagger_{b,s,\xi}(\boldsymbol{q})\xi\hbar v_f \boldsymbol{R}_+\boldsymbol{q}\cdot\boldsymbol{\sigma} a_{b,s,\xi}(\boldsymbol{q}) \quad \text{(S4.14)}$$

$$H_t = \sum_{q,\xi,s} a^\dagger_{t,s,\xi}(\boldsymbol{q})\xi\hbar v_f \boldsymbol{R}_-\boldsymbol{q}\cdot\boldsymbol{\sigma} a_{t,s,\xi}(\boldsymbol{q}) \quad \text{(S4.15)}$$

$$H_{inter} = \sum_{\boldsymbol{q},\boldsymbol{q}',\xi,s} a^\dagger_{t,s,\xi}(\boldsymbol{q})\left(T_{q_b,\xi}(\boldsymbol{q},\boldsymbol{q}') + T_{q_{tr},\xi}(\boldsymbol{q},\boldsymbol{q}') + T_{q_{tl},\xi}(\boldsymbol{q},\boldsymbol{q}')\right) a_{b,s,\xi}(\boldsymbol{q}') \quad \text{(S4.16)}$$

$a^\dagger_{b,s,\xi}$ is the electron creation operator for electrons on the bottom layer with sublattice index s and valley index ξ. $\boldsymbol{R}_\pm$ is the rotation matrix of top and bottom layer by twist angle $\pm\frac{\theta}{2}$: $\cos\left(\frac{\theta}{2}\right) + i\sigma_y \sin\left(\frac{\theta}{2}\right)$. The interlayer coupling matrix is given by:

$$T_{\boldsymbol{q_b},\xi}(\boldsymbol{q},\boldsymbol{q}') = \begin{pmatrix} w_{AA} & w_{AB} \\ w_{AB} & w_{AA} \end{pmatrix} \delta_{\boldsymbol{q}-\boldsymbol{q}',\boldsymbol{q_b}} \quad \text{(S4.17)}$$

$$T_{\boldsymbol{q_{tl}},\xi}(\boldsymbol{q},\boldsymbol{q}') = \begin{pmatrix} w_{AA} & w_{AB}e^{-i\xi\frac{2\pi}{3}} \\ w_{AB}e^{+i\xi\frac{2\pi}{3}} & w_{AA} \end{pmatrix} \delta_{\boldsymbol{q}-\boldsymbol{q}',\boldsymbol{q_{tl}}} \quad \text{(S4.18)}$$

$$T_{\boldsymbol{q_{tr}},\xi}(\boldsymbol{q},\boldsymbol{q}') = \begin{pmatrix} w_{AA} & w_{AB}e^{+i\xi\frac{2\pi}{3}} \\ w_{AB}e^{-i\xi\frac{2\pi}{3}} & w_{AA} \end{pmatrix} \delta_{\boldsymbol{q}-\boldsymbol{q}',\boldsymbol{q_{tr}}} \quad \text{(S4.19)}$$

where $\boldsymbol{q_b} = \frac{8\pi sin\left(\frac{\theta}{2}\right)}{3\sqrt{3}d}(-1,0), \boldsymbol{q_{tl}} = \frac{8\pi sin\left(\frac{\theta}{2}\right)}{3\sqrt{3}d}\left(-\frac{\sqrt{3}}{2},\frac{1}{2}\right), \boldsymbol{q_{tr}} = \frac{8\pi sin\left(\frac{\theta}{2}\right)}{3\sqrt{3}d}\left(\frac{\sqrt{3}}{2},\frac{1}{2}\right)$, $\boldsymbol{d}$ is the interlayer distance. w_{AA} is smaller than w_{AB} due to the lattice reconstruction in the AA region, and we adopt the value: $w_{AA} = 88meV, w_{AB} = 110meV$[7].

S4.4 Nordheim-Fowler Contribution and WSe_2 barrier breakdown

In the MLG-MLG junction, the tunneling spectrum shows a breakdown of the WSe_2 tunneling barrier. This is modeled within the WKB approximation[2,3]. At large bias, the bending of the WSe_2 bands is large enough such that electrons can directly tunnel from the graphene into the conduction band states of WSe_2. This is the Nordheim-Fowler field emission regime, described by:

$$t_{WSe_2} = \exp\left\{-\frac{1}{\hbar}\int_0^{d_0}\left(2m^*(V(z)-E)\right)^{\frac{1}{2}}dz\right\} = \exp\left\{-\frac{1}{\hbar}\frac{2d}{3|\phi|}(2m^*)^{\frac{1}{2}}(\Delta_c - E)^{\frac{3}{2}}\right\} \quad \text{(S4.20)}$$

$$d_0 = (\Delta_c - E) * \frac{d}{|\phi|} \quad \text{(S4.21)}$$

Assuming that Δ_c is momentum independent, then the tunneling current is also momentum independent, and is given by:

$$I_{WSe_2} = \frac{e g_s}{2\pi}\int_{\Delta_c - |\phi|}^{\infty} dE\ t_{WSe_2}^2 f_{MLG}(E) DOS_{MLG} DOS_{WSe_2} \quad \text{(S4.22)}$$

We found the values $m^* = 1.07m_e$ and $\Delta_c = 0.55eV$ fit the experiments, consistent with the values reported in the literature[5].

S5. Understanding MLG-MLG momentum resolved tunneling experiments

In the MLG-MLG momentum-resolved tunneling experiment we identified two main features in the measured d^2I/dV^2 vs. V_b, and θ (Fig. 2c): lines of momentum resolved 'onset' condition, appearing along a straight-X feature, and 'nesting' lines, appearing along a curved-X feature. In this section, we derive the equations that describe these conditions for the case of a charge-neutral system. The experiments in the main text, performed with $V_{BG} = 0V$, are to a very good approximation a charge-neutral system. By comparing these experiments to the formulas derived here, we extract the density dependence of the Fermi velocity and electronic compressibility in this system. In the next section we will show additional measurements done with non-zero V_{BG}, and derive the corresponding equations for the general case of a non-charge-neutral system.

In the charge-neutral case, the carrier density of the bottom layer is equal in magnitude and opposite in sign to that in the top layer:

$$n_T = -n_B \quad \text{(S5.1)}$$

Since MLG is electron-hole symmetric to a good approximation in the energy range probed in our experiments, Eq. S5.1 translates to:

$$\mu_T = -\mu_B \equiv \mu \quad \text{(S5.2)}$$

The electrostatic equation that connects the applied bias, chemical potentials in the two layers, and their relative energy band shift (=electrostatic potential, ϕ) reads (illustrated in top panel, Fig. S5):

$$V_b = \phi + 2\mu \quad \text{(S5.3)}$$

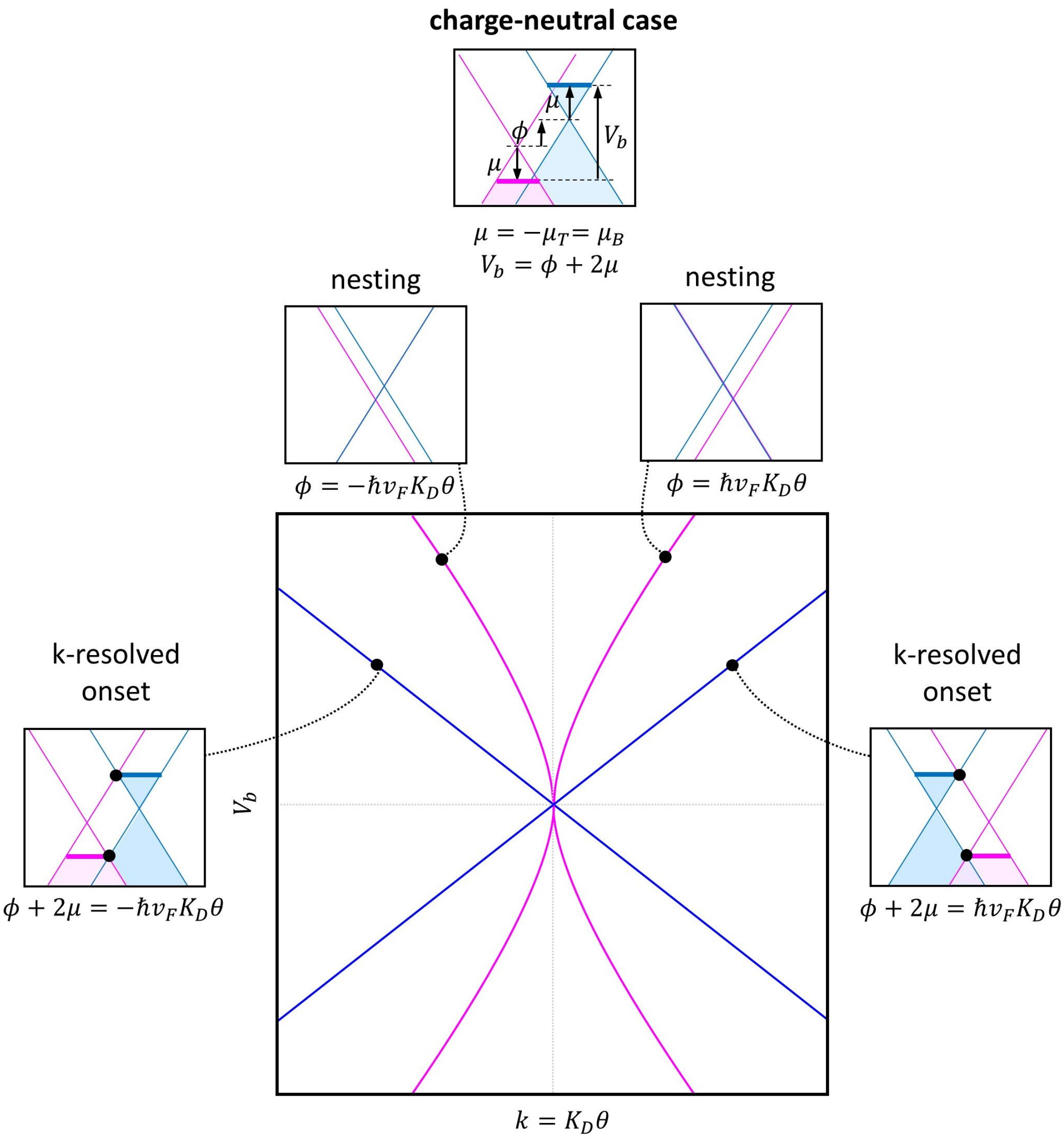

Figure S5. Schematic of the charge-neutral MLG-MLG tunneling experiment (Fig. 2 in main text). The main panel shows the analytically derived lines in the $\theta - V_b$ plane, which correspond to the momentum-resolved 'onset' and 'nesting' conditions. Right and left insets show the theoretically calculated energy band alignments at select points along these lines. Top inset shows how the bias divides between the chemical potential of the two layers, $\mu = -\mu_T = \mu_B$, and the relative electrostatic energy shift of the bands, ϕ.

S5.1 Momentum-resolved onset condition

At a given bias voltage, momentum conserving tunneling onsets when the Fermi surface of one layer touches the empty bands of the other layer. For the charge-neutral case, this condition happens simultaneously for the Fermi surfaces of the top and bottom layers (right and left insets, Fig. S5), when:

$$\phi + 2\mu = \pm\hbar v_F K_D \theta \tag{S5.4}$$

Combining equations S5.4 and S5.3, we get a simple relation for the onset condition,

$$V_b = \pm\hbar v_F K_D \theta \tag{S5.5}$$

Effectively, this relation is the Dirac equation of graphene, but with energy E and momentum k replaced by V_b and $K_D\theta$, respectively.

This simple relation appears only in the charge neutral case (we will discuss the general case below), which is a good approximation for the experiments performed at $V_{BG} = 0V$, allowing us to directly compare them with equation S5.5. In Fig. S6a we plot the measured d^2I/dV^2 in the MLG-MLG (reproduced from Fig. 2c) and trace the peaks (dips) that occur at positive (negative) bias at the onset conditions. The points obtained from this tracing, shown by black dots in Fig. S6b, fit well a linear Dirac dispersion (blue) with a Fermi velocity of $v_F = (1.05 \pm 0.02) * 10^6 m/s$, consistent with previous measurements in graphene[8]. In principle, the Fermi velocity in graphene is known to be renormalized by interactions and becomes logarithmically faster as the energy of the electrons approaches the Dirac point[8,9]. This renormalization depends on the fine structure constant in graphene $\alpha = e^2/\epsilon\hbar v_F$. Since α depends on the dielectric coefficient of the surrounding environment, when graphene is encapsulated in a dielectric material (in our case WSe_2 and hBN at its opposite sides), the value of the fine structure constant is close to one, and the velocity renormalization is rather small,

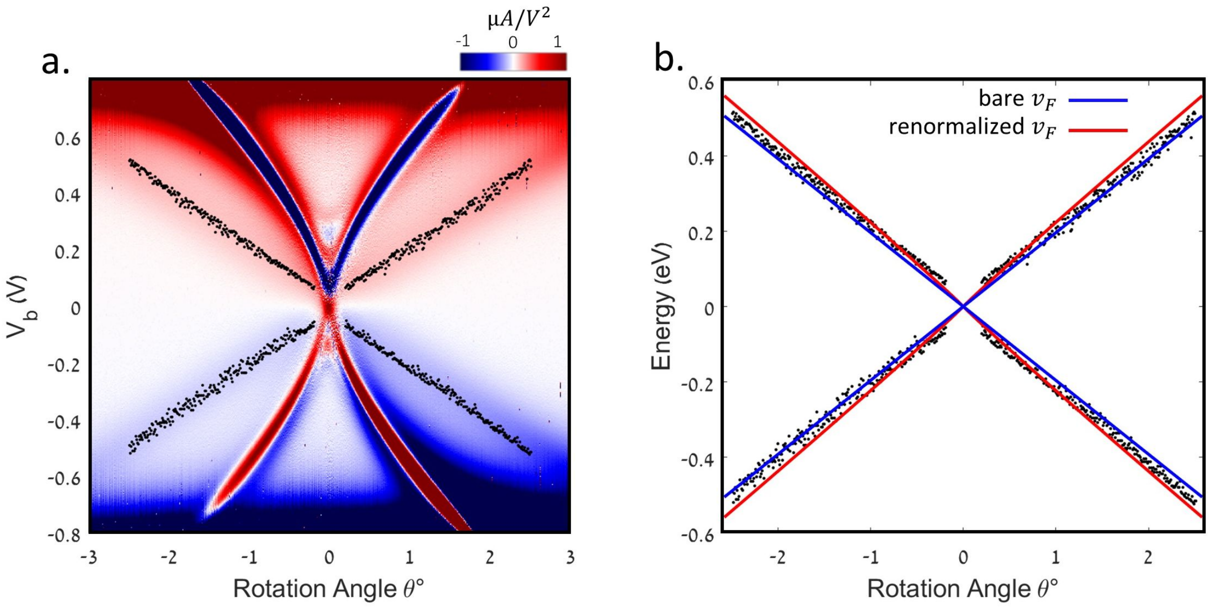

Figure S6. a. Measured $\frac{d^2I}{dV^2}$ from Fig. 2c, where we traced the local maximum (minimum) at positive (negative) bias, along the 'onset' features (black dots). **b**. Plot of the black dots from panel a together with bare Dirac dispersion (Eq. S5.5) with $v_F = 1.05 \cdot 10^6 m/s$ (blue) and Dirac dispersion that is renormalized by interactions[8,9].

about 10% change in velocity per one decade in energy[9]. Our experiments have two MLG layers, which screen better than the single MLG in ref. [9]. On length scales larger than the separation of the two MLG layers ($\sim 1nm$ in our experiment), they act as a single layer with double the density of states, effectively dividing the relevant fine structure constant, and the velocity renormalization by another factor of 2. The red curve in Fig. S6b shows the renormalized velocity calculated for our case. Visibly, the effect of renormalization is small, and within the errors in the determination of the onset peaks, both renormalized and un-renormalized expressions for the Fermi velocity fit well.

S5.2 Nesting condition

Nesting between the Dirac bands of the top and bottom layers happens when these bands overlap along a line of the Dirac dispersion. In this situation, a large number of energy-momentum conserving states become available to tunnel between the layers. Unlike the onset condition, the nesting condition is independent of the band filling, and therefore it directly probes the energy dispersion. At a given twist angle θ, the Dirac dispersion of one of the MLG is shifted in momentum by $k = K_D\theta$ relative to the other. Shifting the Dirac bands in energy by

$$\phi = \pm \hbar v_F K_D \theta \qquad \text{(S5.6)}$$

will bring the two MLG Dirac dispersions into nesting. This condition is similar to the onset condition in equation S5.5, except that here ϕ replaces V_b.

The nesting line in the measurement allows us to read out directly the energy shift between the bands for every V_b. Figure S7a plots the measurement after rescaling the rotation angle axis using equation S5.6. In these new axes the nesting line gives directly $\phi(V_b)$.

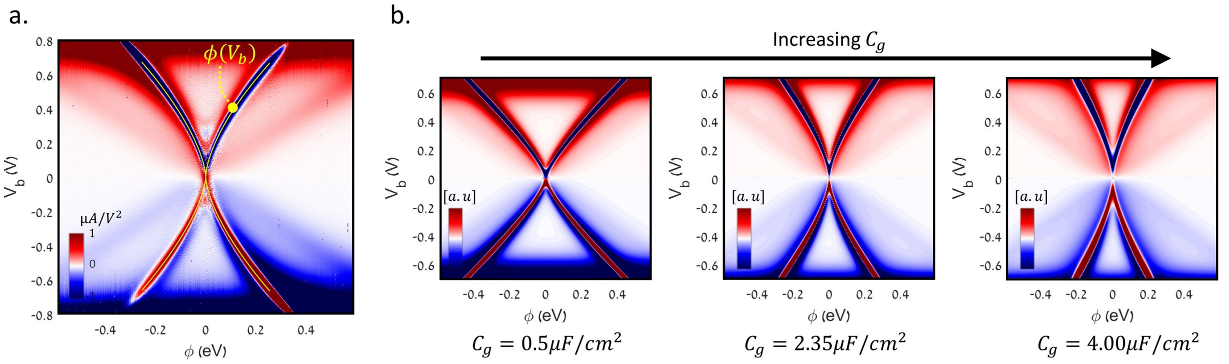

Figure S7. Rescaling rotation angle θ to the electrostatic potential ϕ. a. $\frac{d^2I}{dV^2}$ measurement from Fig. 2c with θ axis converted to ϕ using the relation $\phi = \hbar v_F K_D \theta$. After rescaling, the nesting line (yellow extracted from peak fitting) gives directly $\phi(V_b)$. **b**. Calculated $\frac{d^2I}{dV^2}$ spectrum for geometric capacitances $C_g = 0.50, 2.35\ and\ 4.00\ \mu F/cm^2$

Once the dependence $\phi(V_b)$ is determined from the measurement, it is straightforward to obtain from it the geometric capacitance, C_g, and quantum capacitance, C_q, of the system. By differentiating Eq. S5.3 with respect to ϕ, we obtain,

$$\frac{dV_b}{d\phi} = 1 + 2\frac{d\mu}{dn}\frac{dn}{d\phi} = 1 + 2\frac{C_g}{C_q} \qquad \text{(S5.7)}$$

which provides a direct relation between the local slope of the nesting line and the ratio of geometric to quantum capacitances. We see that for system with different geometric capacitances the nesting line will have a different shape. This is demonstrated in Fig. S7b, which shows the theoretically calculated graphs with increasing value of C_g between the different panels. For the case of a smaller geometric capacitance (left panel, Fig. S7b), the slope of the nesting line approaches 1, as required by Eq. S5.7. In this limit, which corresponds to the case of a thick tunneling barrier, most of the applied bias goes into shifting the bands, and only a small fraction of it goes to increasing the chemical potentials. In the opposite limit (right panel), where the geometric capacitance is large (corresponding to a thin barrier), the bias hardly shifts the energy bands but mostly changes their occupation. The center panel in Fig. S7b has a nesting line that matches closely the dependence observed in the experiment (Fig. S7a). This allows us to determine the geometric capacitance in this experiment to be $C_g = 2.35\mu F/cm^2$. We can see that for a typical width of tunneling barriers (2-4 monolayers) and with the typical density of states of MLG, the experiments are generally in an intermediate regime in which the geometric and quantum capacitances are of comparable magnitude. Eq. S5.7 also shows that once C_g is determined, the local slope of the nesting line provides the quantum capacitance of the system. Eq. S5.7 was derived for the symmetric (charge-neutral) case of two MLG layers. In the general case when two systems are different, or they are the same but have different carrier densities, then $2/C_q$ in Eq. S5.7 should be replaced by the in-series sum of the quantum capacitances of the two systems, $1/C_q^T + 1/C_q^B$.

S6. Additional experiments: back-gate dependence of MLG-MLG k-resolved tunneling

In Fig. 2 we showed momentum resolved tunneling experiments between two MLG layers. Here we present additional data, showing how these measurements evolve with back gate voltage. Figures S8-S10 show measurements of I, dI/dV and d^2I/dV^2, respectively, vs. θ and V_b, now as a function of back gate voltage. The top row in each figure shows measurements done at three values of the back-gate voltage, $V_{BG} = -10V, 0V, +10V$. The bottom row shows the corresponding theoretical calculations, based on the BM expression for momentum resolved tunneling between two parallel layers[4]. Whereas the measurement at $V_{BG} = 0V$ is symmetric between positive and negative bias, the measurements at $V_{BG} = -10V, +10V$ are asymmetric in bias. Focusing on the measurements of d^2I/dV^2 (Fig. S10), in which the features are the sharpest, we can see that while for $V_{BG} = 0V$ the crossing point of the curved-X is at zero bias, for finite V_{BG} it shifts to finite bias (positive bias for $V_{BG} = +10V$ and negative bias for $V_{BG} = -10V$). In addition, the symmetric X shape at $V_{BG} = 0V$ evolves into a "bell-shape" that is asymmetric with respect to bias. However, if we flip the bias axis and the sign of the current in the $V_{BG} = -10V$ measurement it becomes almost identical to the measurement at $V_{BG} = +10V$. Namely, the data obeys the following symmetry: $I(\theta, V_b, V_{BG}) = -I(\theta, -V_b, -V_{BG})$ and similarly $d^2I/dV^2(\theta, V_b, V_{BG}) = -d^2I/dV^2(\theta, -V_b, -V_{BG})$.

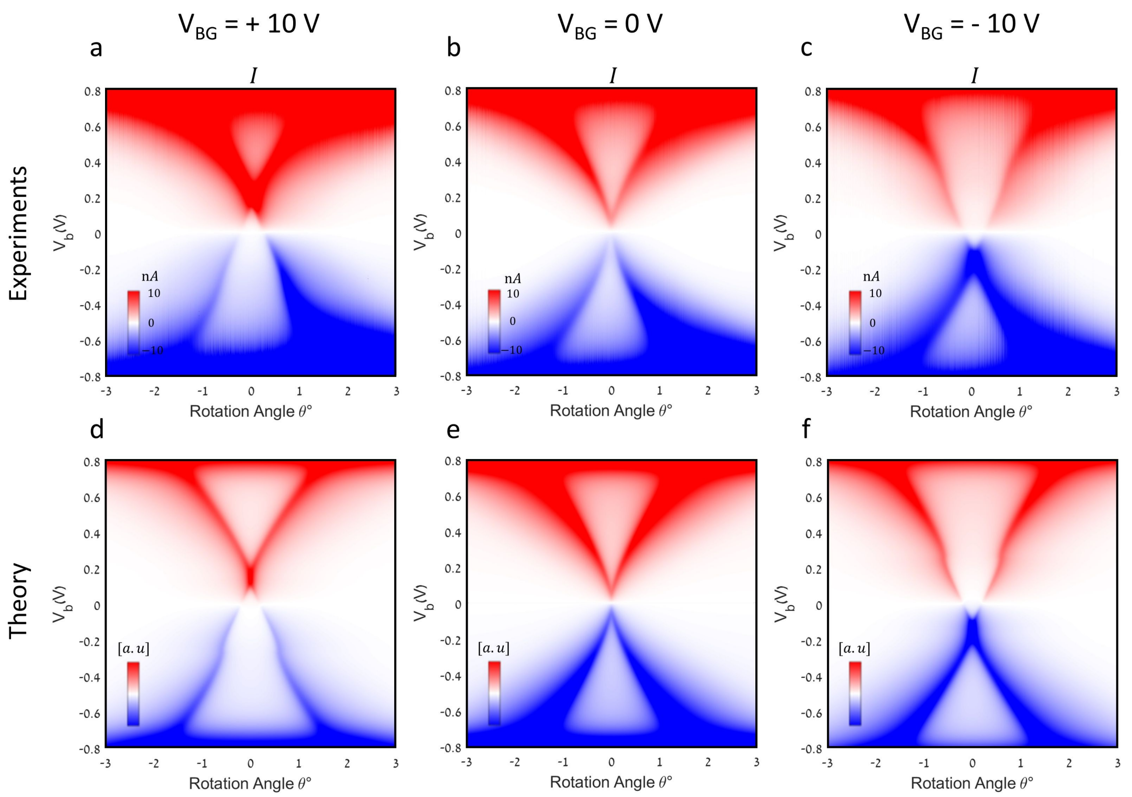

Figure S8: **Back-gate voltage dependence of momentum-resolved tunneling current in a MLG - MLG junction a-c.** Tunneling current, I, measured in the junction in Fig. 2 (MLG/ trilayer WSe_2/MLG) as a function of θ and V_b, for three values of the back-gate voltage $V_{BG} = +10V, 0V, -10V$ respectively. **d-e.** Corresponding theoretical calculations

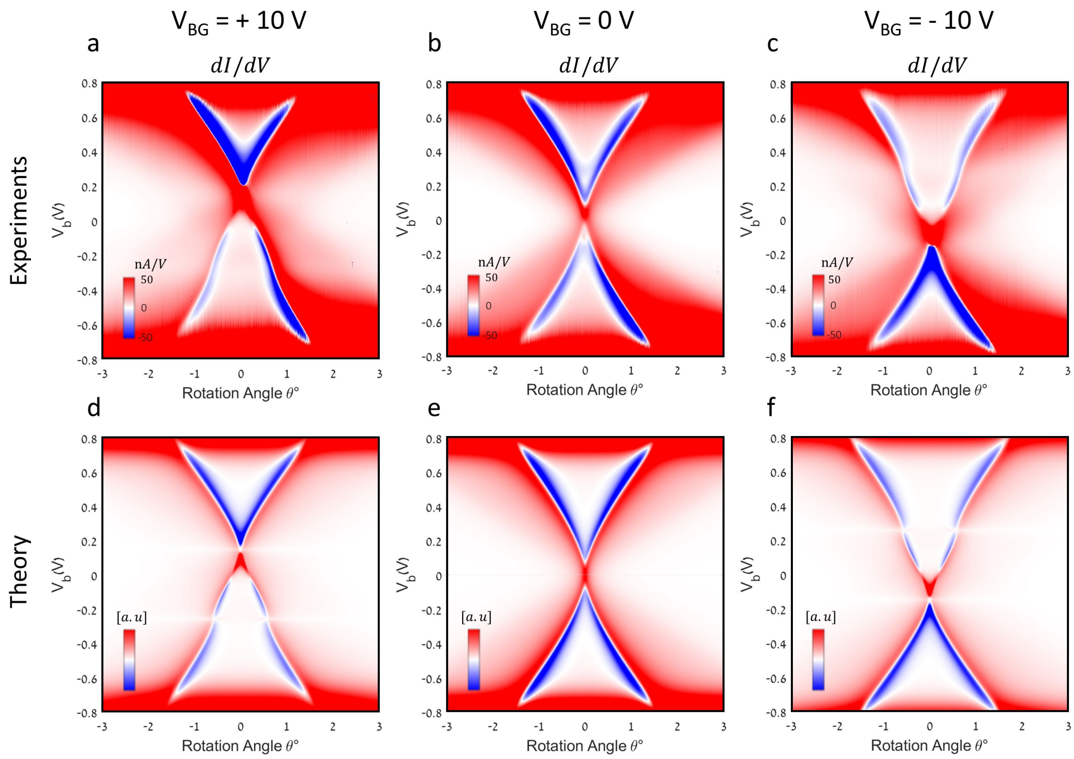

Figure S9: **Back-Gate voltage dependence of dI/dV.** The same measurement as in Fig. S8 but now showing the (simultaneously measured) conductance, dI/dV. **a-c.** Experiment. **d-e.** Theory.

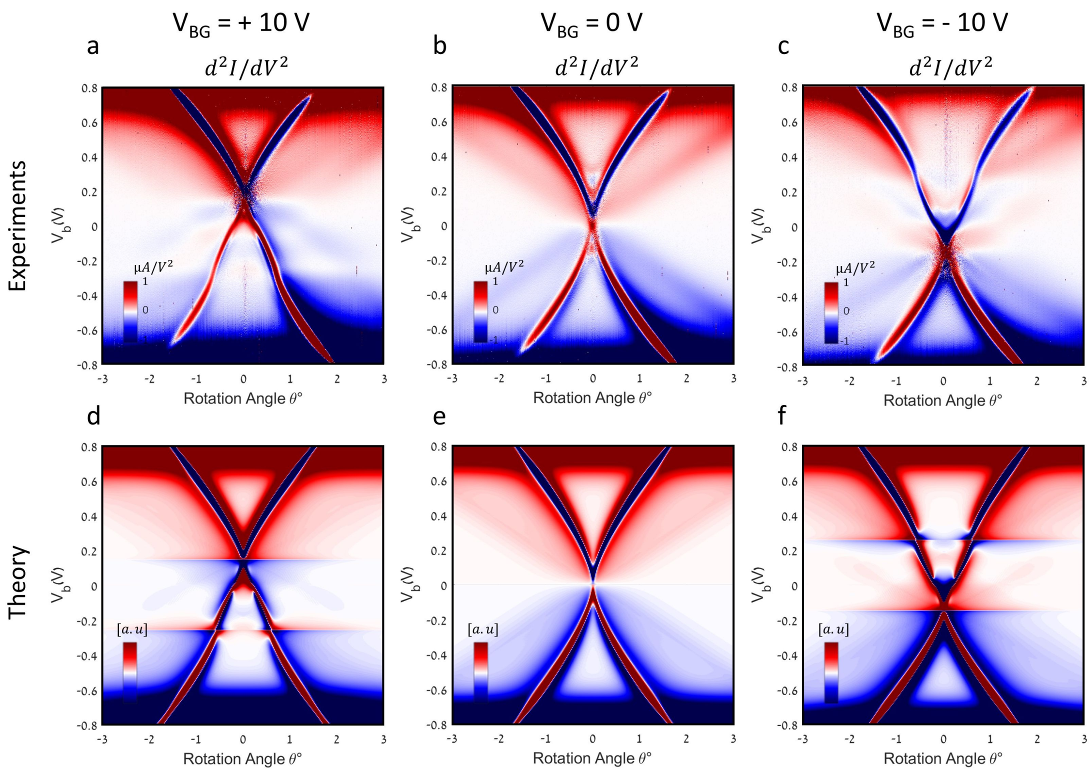

Figure S10: **Back-Gate voltage dependence of d^2I/dV^2.** The same measurement as in Fig. S8,9 but now showing the second derivative, d^2I/dV^2, obtained by numerically deriving the data in Fig. S9. **a-c.** Experiment. **d-e.** Theory.

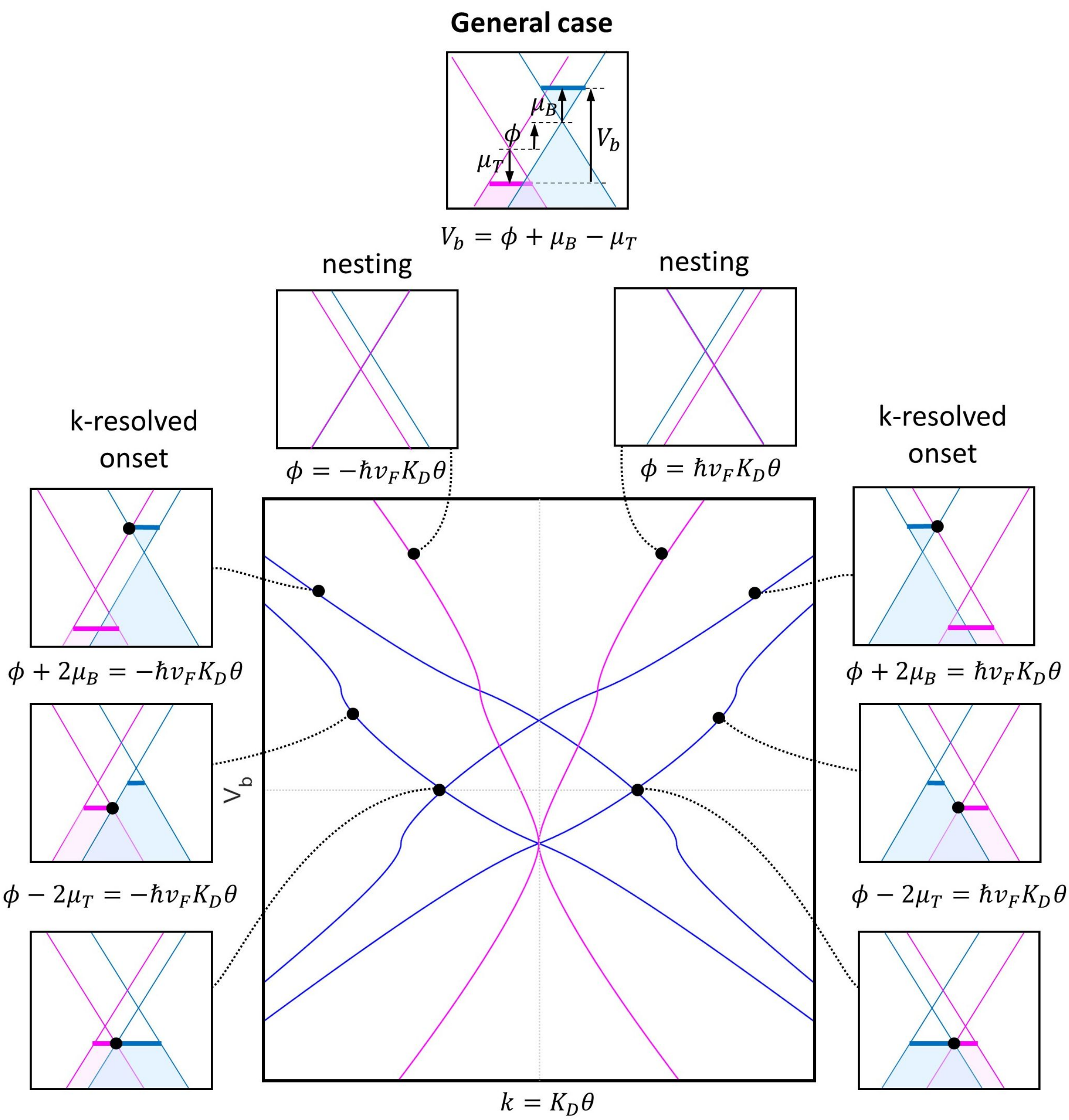

Figure S11: Schematic of the MLG-MLG tunneling junction for experiments with a finite back gate voltage. The main panel shows the theoretically derived lines in the $\theta - V_b$ plane, which correspond to the momentum-resolved 'onset' and 'nesting' conditions. Compared to the charge-neutral case, the 'nesting' line becomes "bell-shaped" and asymmetric in bias, and the 'onset' lines splits to four lines. Right and left insets show the theoretically calculated energy band alignments at select points along these lines. Top inset shows how the bias divides between the chemical potential of the two layers, μ_T and μ_B, and the relative electrostatic energy shift of the bands, ϕ.

To understand the experiments with finite back-gate voltages, we allow for the chemical potentials of the two layers to be different and generalize the analysis to include a finite V_{BG}, following equations S4.1-S4.4. The resulting 'onset' and 'nesting' lines in this more general case are shown in Fig S11. We see that the theory recovers the asymmetric bell-shaped 'nesting' line, observed experimentally. This shape generalizes the functional dependence of ϕ vs. V_b (Eq. S5.7) to now include two unequal inverse quantum capacitances in the layers, which add in series. In addition, we see that now each 'onset' line splits in two, since the Fermi surfaces in both layers have a different size and they touch the empty bands in the other layer at different twist angles.

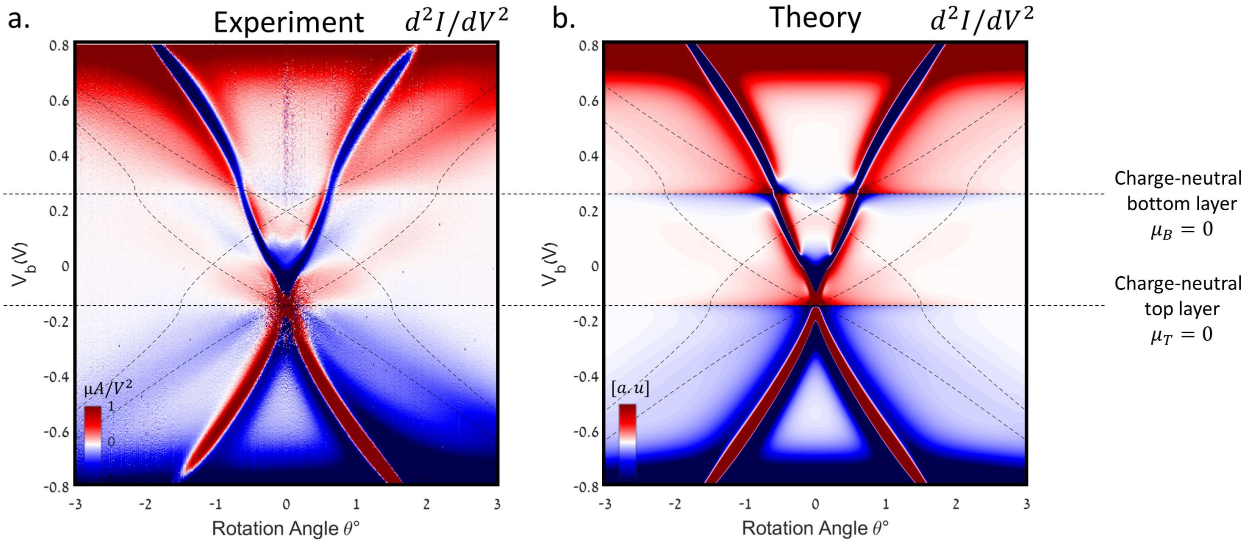

Figure S12. a. Measured and **b.** Calculated d^2I/dV^2 vs. θ and V_b, at $V_{BG} = -10V$ (similar to Figs. S10c,f). The overlaid black dotted lines correspond to the theoretically calculated k-resolved 'onset' conditions, as shown in Fig. S11. Horizontal dashed lines indicate the two biases at which each one of the MLG crosses through charge neutrality ($\mu_B = 0$ or $\mu_T = 0$).

Fig S12 overlays the calculated 'onset' lines on the measured d^2I/dV^2 at $V_{BG} = -10V$, as well as on the corresponding theoretical calculation. Although room temperature smearing has almost eliminated all the detailed structure related to the 'onset' conditions, we can still resolve in the experiment faint peaks that follow the predicted 'onset' lines. Another feature that is clearly visible both in theory and experiment are the horizontal lines (dashed black) along which the d^2I/dV^2 flips sign. Each such line corresponds to one of the MLG layers going across charge neutrality, namely either $\mu_T = 0$ or $\mu_B = 0$. Overall, we can see that the theory captures most of the details in the experiment rather accurately.

S7. Electron lifetime in MLG-MLG k-resolved tunneling experiment

The dominant mechanism setting the width of the features in our measurement is thermal smearing. All measurements are performed at room temperature, leading to $k_BT{\sim}25meV$ smearing entering via the Fermi-Dirac functions in Eq. S4.5. To quantitatively explain our data, we find that we should also include smaller smearing due to finite electron lifetime[10–13]. The electron lifetime is encoded in the spectral functions of the bottom and top layers, $A_B(\epsilon,\boldsymbol{k}) = \frac{2\gamma_B}{(\epsilon-\epsilon_B(\boldsymbol{k}))^2+\gamma_B^2}$ and $A_T(\epsilon,\boldsymbol{k}) = \frac{2\gamma_T}{(\epsilon-\epsilon_T(\boldsymbol{k}))^2+\gamma_T^2}$, where $\epsilon_B(\boldsymbol{k})$ and $\epsilon_T(\boldsymbol{k})$ are the energy bands of these layers, and γ_B, γ_T are their corresponding inverse electron lifetimes. Finite lifetime leads to a Lorentzian broadening of the spectral functions, which could be directly observed in the width of the features in the measurements.

Fig. S13 plots the measured I vs. θ at small bias, $V_b = 40mV$, (black line, same as the inset in Fig. 2a in the main text). Compared with theoretical curves broadened by $k_BT = 25meV$ and with various inverse lifetime values, γ_0, we see that the value that best fits the measurement is $\gamma_0 \approx 4meV$. At higher biases we (of the order of the Fermi energy) we generally observe shorter lifetimes. The dependence of the lifetime on energy is expected to be complicated in this energy range. We find a reasonably good fit to the phenomenological expression used previously[13], where $\gamma = \gamma_0 + \gamma_1|\epsilon - \mu|$, with the γ_0 above and $\gamma_1 \approx 0.035$, but this subject requires more elaborate study, which is beyond the scope of the present paper.

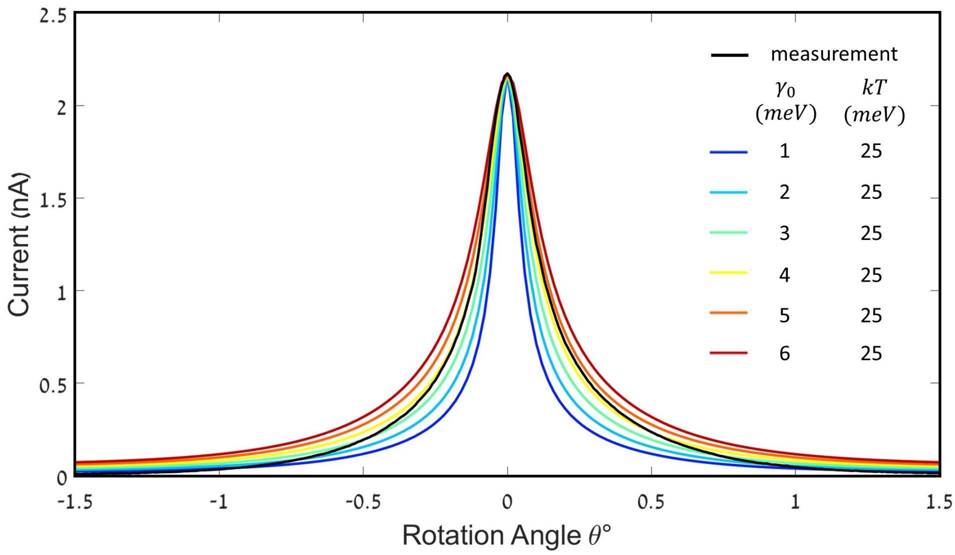

Figure S13: **Determining the electron lifetime at low bias.** Black curve: Measured tunneling current, I vs. rotation angle θ, at small bias $V_b = 40mV$, for the junction in Fig 2 (same as he right inset in Fig. 2a). Colored curves are the theoretically calculated k-resolved tunneling current (see text) with room temperature broadening $k_BT = 25meV$ and various values of inverse electron lifetime, γ_0 (see key).

S8. Understanding MLG-TBG momentum resolved tunneling experiments

In this section, we explain the various features observed in the momentum resolved tunneling experiments between MLG and TBG, shown in Fig. 3. We first use a simple toy model to explain the main mechanism leading to the quantitative energy band mapping in our experiment – tracing the bands of the sample with a Dirac point of the probe. As we shall show, this mechanism directly probes the dispersion of the bands and is insensitive to their filling. We then expand the discussion to include additional momentum-resolved tunneling 'onset' features that appear when the Fermi level of one side (tip/sample) crosses the energy bands of the other side. We end by showing how these features appear in the full momentum resolved calculation, which includes the multitude of energy bands obtained within the BM model of TBG as well as finite temperature effects.

S8.1 First Toy Model – tracing energy bands with a Dirac cone.

Here we show the ability of using a Dirac cone to trace the energy bands of a sample, i.e., a method to read off the samples band structure. We conceive a toy model of a probe layer with a Dirac dispersion scanning a system with arbitrary energy bands. For simplicity, we set the tunneling matrix element to be a momentum-independent constant. The results of the toy model depend on the relative Fermi velocity of the probe and system: $v_{F_{probe}}$ and $v_{F_{syst}}$. Close to the intersection point with the probe's dispersion, the system's band can be linearized with Fermi velocity $v_{F_{syst}}$. The energy offset between the Dirac point of the probe and the system's band is denoted by $\Delta\varepsilon$. This offset can be tuned by changing the bias voltage between the probe and the system. We consider three situations: 1. $v_{F_{probe}} > v_{F_{syst}}$; 2. $v_{F_{probe}} = v_{F_{syst}}$; 3. $v_{F_{probe}} < v_{F_{syst}}$. The tunneling occurs in an energy window set by the bias voltage, where the states in the system are empty and the states in the probe are full (or vice versa). We denote this energy window by $\varepsilon_1 \le \varepsilon \le \varepsilon_2$. For simplicity, we set $T = 0$ and neglect the electron lifetimes (i.e., we treat the electrons as free).

In the case where $v_{F_{probe}} > v_{F_{syst}}$, the intersection between the energy surfaces of the probe and the system is an ellipse. We assume that the entire ellipse is contained within the energy window $[\varepsilon_1, \varepsilon_2]$ where the tunneling takes place. Then, using Eq. (S4.5), we obtain:

$$I \propto |T_{TB}|^2 \frac{|\Delta\varepsilon|}{4\pi v_{F_{probe}}^2} \frac{\left(v_{F_{probe}}/v_{F_{syst}}\right)^3}{\left[\left(v_{F_{probe}}/v_{F_{syst}}\right)^2 - 1\right]^{3/2}}, \qquad (v_{F_{probe}} > v_{F_{syst}}) \qquad \text{(S10.1)}$$

where T_{TB} is the tunneling matrix element. Hence, the current is minimal when the Dirac point coincides the system's dispersion ($|\Delta\varepsilon| \to 0$). At this point, d^2I/dV^2 is strongly peaked. Thus, in this situation, the peak in d^2I/dV^2 can be used to directly read off the system's band structure (see Fig. S14a, b and Supp. Video 2). In the limit $v_{F_{probe}} \to v_{F_{sys}}$, the size of the elliptical intersection between the two dispersion diverges, and it extends outside the window $[\varepsilon_1, \varepsilon_2]$. Hence, Eq. (S10.1) is not applicable in this limit.

If $v_{f_{probe}} = v_{f_{syst}}$, the intersection between the energy surfaces is a parabola. The tunneling current is

$$I \propto |T_{TB}|^2 \frac{\sqrt{2}|\Theta(\Delta\varepsilon)\varepsilon_2 + \Theta(-\Delta\varepsilon)\varepsilon_1|^{3/2}}{3(2\pi)^2 v_{F_{probe}}^2 |\Delta\varepsilon|^{1/2}}, \quad (v_{F_{probe}} = v_{F_{syst}}) \qquad \text{(S10.2)}$$

Interestingly, the current diverges as $|\Delta\varepsilon| \to 0$. This divergence corresponds to the nesting condition in the MLG-MLG tunneling junction. The divergence is regularized by introducing a curvature to the bands, or a finite lifetime of the electrons in the system and the probe.

Finally, for $v_{F_{probe}} < v_{F_{syst}}$, the intersection is a hyperbola. In this case, the singular part of the current at small $|\Delta\varepsilon|$ is

$$I \propto -|T_{TB}|^2 \frac{|\Delta\varepsilon| \ln |\frac{\Theta(\Delta\varepsilon)\varepsilon_2+\Theta(-\Delta\varepsilon)\varepsilon_1}{\Delta\varepsilon}|}{4\pi v_{f_{probe}}^2} \frac{\left(v_{F_{probe}}/v_{F_{syst}}\right)^3}{\left[1-\left(v_{F_{probe}}/v_{F_{syst}}\right)^2\right]^{3/2}}. \; (v_{F_{probe}} < v_{F_{syst}}) \tag{S10.3}$$

I.e., in this case, the current has a maximum at $\Delta\varepsilon = 0$. At this point, d^2I/dV^2 shows a minimum. This is the case for the Dirac point of TBG - with slower Fermi velocity - tracing the MLG dispersion.

S8.2 Second toy model - explaining the k-resolved 'onset' features

In the previous section we explained the main features in our experiment that visualize the energy bands as resulting from tracing with a Dirac point. In addition to these, we see features in the measurement (Fig. 3) that correspond to 'onset' conditions of the momentum resolved tunneling. These occur whenever the Fermi surface of one side (tip/sample) crosses the energy bands of the other side. In this section, we use a more elaborate model that captures these 'onset' features and identifies their origin. We will further show that these 'onset' features are strongly smeared by temperature, whereas the 'Dirac matching' features discussed in the previous subsection are largely insensitive to temperature.

To capture the essential physics, we use a second toy model that includes Dirac bands that represent the MLG and two parabolic bands, $\epsilon_{k\pm} = \pm(Ak^2 + W/2)$, which mimic the conduction and valence flat TBG bands along the $K_{bot} - M - K_{top}$ trajectory in momentum space. In addition, the model also keeps track of the chemical potentials in both systems. The Dirac and Parabolic bands are shown explicitly in Fig. S14a. For simplicity, we focus only on a relative rotation angle of $\theta = 0°$ between the MLG and TBG and calculate the current and its derivatives as a function of bias. Figure S14b, shows I vs. V_b calculated within this model, assuming a low temperature ($k_BT = 1meV$). The corresponding d^2I/dV^2 is shown in Fig. S14c. For the electrostatics, we assume a simple linear relation between the chemical potential of the layers and the applied bias, $\mu_T = -0.4V_b$ and $\mu_B = 0.2V_b$, which approximates rather well the full electrostatics of the real MLG-TBG junction.

At small V_b, the current is zero since there are no energy-momentum conserving states in the tunneling energy window (represented by a white window in Fig.14a and insets to Fig. 14b) between the chemical potentials of the two layers (red and blue dashed lines). Then, at a finite V_b ($\sim 0.18V$ for the parameters of Fig. 14b), the current exhibits a sharp step increase, that occurs when the Fermi level of the Dirac band (red dashed) crosses the occupied parabolic energy bands (left inset, Fig. 14b, crossing indicated by the black arrows). This is the 'onset' point where momentum conserving current first becomes available. At this bias, at zero temperature, the current discontinuously jumps to a finite value because a full circle of energy-momentum conserving states shared by the Dirac and parabolic bands becomes abruptly available. In the second derivative, d^2I/dV^2, the step of I shows up as a plus-minus dipole (Fig. 14c). With a further increase of V_b, the Dirac cone shifts down with respect to the parabolic bands and the circle of energy-momentum matched states (now deep below the Fermi energy) shrinks. Consequently, the current decreases linearly with V_b. It reaches a minimum when the Dirac point exactly crosses the parabolic band and the circle shrinks to a single point (center inset, Fig. S14b). In the second derivative, this minimum translates to a peak. This is the 'Dirac matching' condition, discussed in the previous section. Upon a further increase of V_b, the energy-momentum conserving circle of states grows back again, leading to a linear increase of I, until a second onset condition occurs, where the Fermi level in the parabolic bands (dashed blue) crosses the empty Dirac bands (right inset, Fig. S14b). This second onset leads to another step in I and to a second diploe in d^2I/dV^2.

The calculation in Fig S14 was performed with a low temperature, such that it would be easy to identify the various features. Before relating the predicted features with our experiment, we discuss the smearing of different features at elevated temperatures.

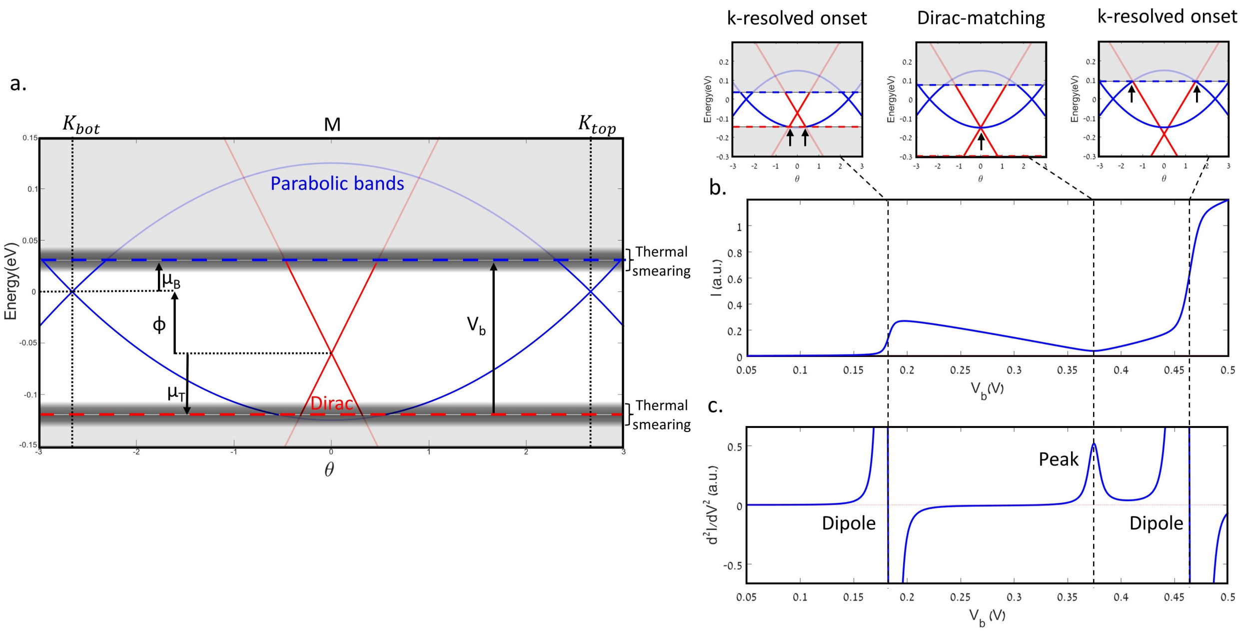

Figure S14. A second toy model, explaining the momentum resolved features. a. Schematic band alignment of a toy model comprising of Dirac (red) and parabolic (blue) bands. The Fermi levels in the Dirac and parabolic bands are marked by dashed red and blue lines, correspondingly. The bias, V_b, gives the distance between these two Fermi energies. Also indicated are the chemical potentials of the top and bottom layers, μ_T and μ_B, and the energy band shift (= the electrostatic potential), ϕ. Darker regions around the Fermi energies represent the thermal smearing due to room temperature. **b**. and **c**. are the calculated I and $\frac{d^2I}{dV^2}$ as a function of V_b at $\theta = 0°$. Three distinct features are marked by dashed lines. Top insets show the calculated relative band alignment between the Dirac and parabolic bands, that correspond to these features. Black arrows represent the energy-momentum matching states at every condition.

S8.3 Temperature dependence of 'onset' and 'Dirac matching' features

Figure S15 shows how the 'onset' and 'Dirac matching' features evolve between low temperature ($k_BT = 1meV$, blue) and room temperature ($k_BT = 25meV$, dashed red). Clearly, the steps in I that correspond to the 'onset' conditions gets smeared with temperature (panel a). The corresponding dipoles in d^2I/dV^2 (panel b) are also smeared. Thus, their amplitude is significantly reduced. In contrast, the d^2I/dV^2 peak that correspond to the 'Dirac matching' condition is hardly affected by the temperature. This highlights the different mechanisms at play – the 'onset' conditions occur when a Fermi energy crosses an available band. Temperature broadens the Fermi-Dirac distribution around the Fermi energy, which leads to smearing of these features. In contrast, the 'Dirac matching' feature occurs deep inside the bias window. If the temperature is smaller than the distance between the Dirac point and the edges of the bias window, then these features would be unaffected by the thermal smearing. Instead, their width represents the intrinsic electronic lifetime.

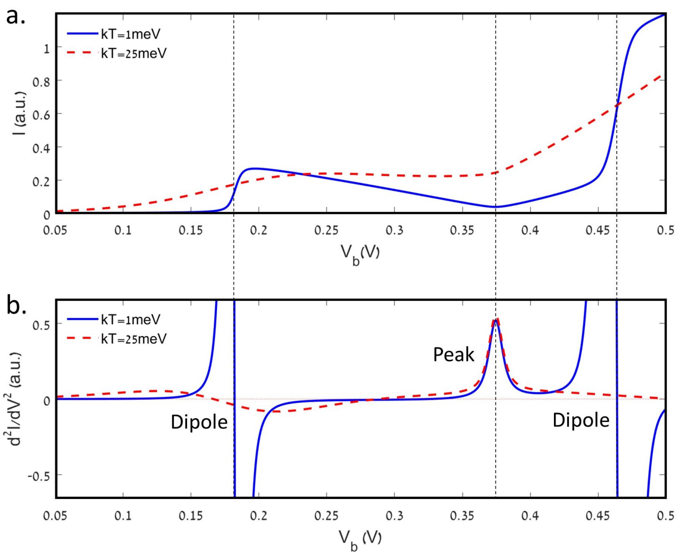

Figure S15. Temperature dependence of tunneling vs. bias in the second toy model. A. and **b**. plot the current, I, and its second derivative, $\frac{d^2I}{dV^2}$, at $\theta = 0°$ at two temperatures (see key). The dipole feature in panel b results from the 'onset' of current and is smeared with the temperature due to the smeared Fermi-Dirac distribution. In contrast, the peak in $\frac{d^2I}{dV^2}$ that originates from energy-momentum matching of the Dirac point and parabolic bands, is not affected by temperature as it occurs deep inside the bias window.

S8.4 Comparing the second toy model with the experiment

We can now compare the second toy model with the measurements. Fig. S16a shows a zoom-in of the central part of one of the flat bands in the measurement in Fig. 3d, where it exhibits a parabolic shape. We can see that a single band generates three curved features in d^2I/dV^2, marked A, B, and C in the figure, corresponding to a maximum/minimum/maximum (red/blue/red) of d^2I/dV^2. The toy model, calculated with $k_BT = 25meV$ (Fig. S16b) qualitatively reproduces all three features. Figs. S16c, d

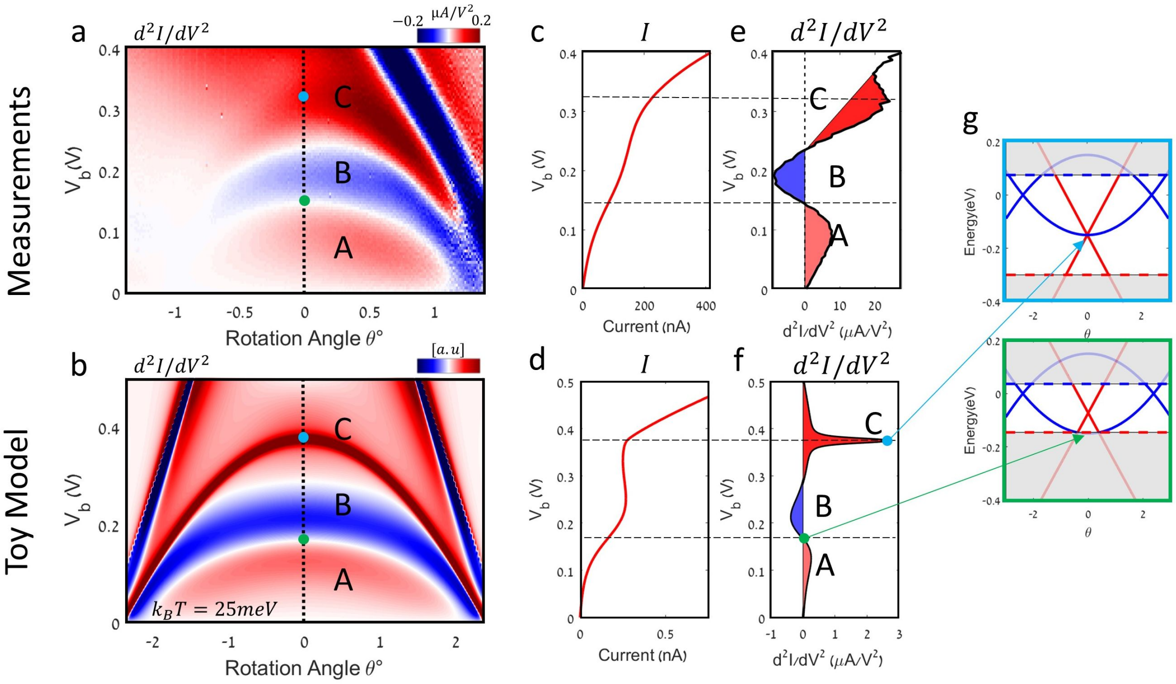

Figure S16. a. Measured d^2I/dV^2 vs θ and V_b from Fig. 3d, zoomed-in around features originating from the TBG flat bands. Three features are marked by A, B and C (red, blue, red). **b**. Corresponding calculation based on the toy model in section S8.2 with $k_BT = 25meV$. **c**. I vs. V_b along a line cut at $\theta = 0°$ in the measurement (dashed black line in panel a). **d.** Similar linecut through the model calculation in panel b. **e**. and **f**. corresponding second derivative, d^2I/dV^2, vs. V_b traces. **g**. The band alignments that correspond to the V_b points marked in panel f. Red/blue bands correspond to the Dirac and parabolic bands. Red/blue dashed lines are the Fermi energies in these two systems. The gray area marks the energies outside the bias window.

show I vs. V_b along a line cut at $\theta = 0°$ (black dashed line in panels a, b) in the measurement and model, respectively. Figs. S16e,f plot the corresponding d^2I/dV^2 traces. By comparing this cut to Fig. S15 we can see that the pair of A/B (maximum/minimum) features correspond to the d^2I/dV^2 dipole that results from the first 'onset', and that feature C (maximum) is the peak that corresponds to the 'Dirac matching'. To see this more directly, we plot in Fig. S16g the band alignment in the model for the transition point between the A and B features, and for the peak of the C feature. Indeed, we see that the assignment fits well these two conditions, even in this higher temperature calculation. In the experiment (Fig. S16e), we see the same features, at similar biases, although the relative amplitudes are somewhat different. We also note that in the experiment these features are sitting on top of a background current that increases with V_b. This background corresponds to thermally activated current above the WSe_2 barrier, which is present in the experiment and not in the simple toy model.

S8.5 The full calculation – origin of the features in the MLG-TBG k-resolved tunneling

Having identified the main features in our experiments using toy models, we now end by showing how these features appear in the full calculation (Fig. 3e), which includes the multiple TBG bands (flat and remote) of the BM continuum model[6].

In Fig. S17a we reproduce the theory plot of Fig. 3e, showing d^2I/dV^2 as a function of θ and V_b, and highlight special points at which we show the calculated band alignment. Fig. 17b plots the same points

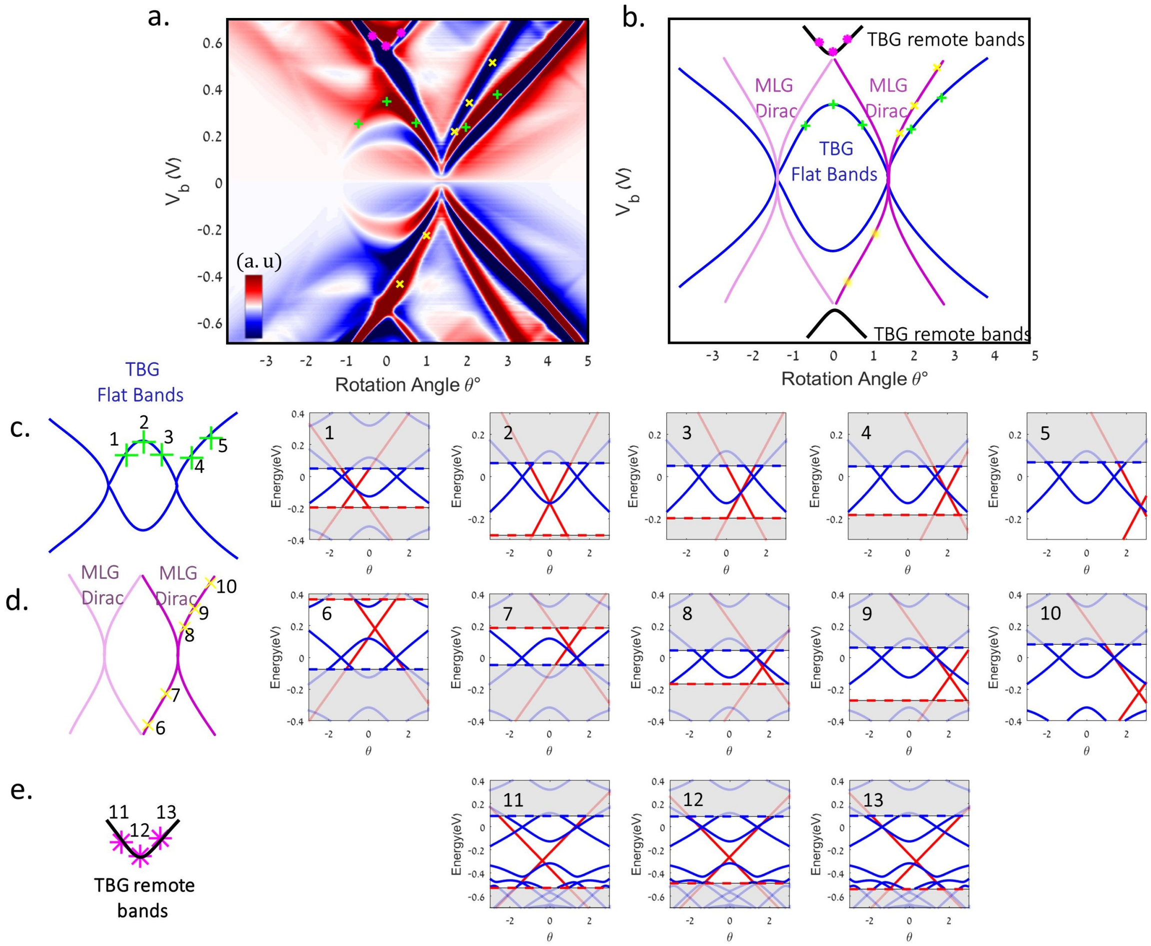

Figure S17. **a**. Calculated d^2I/dV^2 vs. θ and V_b similar to Fig. 3e, but with overlayed points for which we present the band alignment. **b**. tracing of the major features in panel a, which include the TBG flat bands (blue), MLG Dirac bands (purple) and TBG remote bands (black) **c.** band alignment for five points along the TBG flat bands features. **d.** Same, for five points along the MLG Dirac features **e**. Same, for three points along the remote band feature. In the individual panes the MLG and TBG bands are plotted by red and blue, their Fermi energies are plotted by correspondingly colored dashed line, and the greyed-out area shows the region that is outside the bias window.

over a traced-out version of the measurements. In the rest of the panels, we present the calculated band alignments along each of the traced features. In panel c we show five points along the flat band feature (blue). The theory-derived band alignments at these points are shown in the right panels. We clearly see that along this feature the Dirac point of the MLG traces the TBG flat band (seen in more details in Supp. Video 2). In panel d, we follow the band alignment along the traced-out purple feature. Notably, here the right Dirac cone of the TBG flat bands is tracing the bands of the MLG (Supp. Video 3). In panel e, we show points along the 'remote bands' features. The band alignment diagrams show that this feature appears when the MLG Dirac point is tracing these bands, similar to the mechanism through which the flat bands are traced.

We end by noting that the MLG Dirac feature that appears prominently in our measurement (indicated by purple lines in Fig. S17b) allows us to convert the applied V_b to the actual energy shift between the bands. The procedure is similar to that in the MLG-MLG experiment: given that we know the v_F in our MLG probe, we can relate the angle along which the feature appears to the energy shift between the bands by $\phi = \hbar v_F K_D \theta$ (we use here one of the two Dirac features observed in the measurement, and measure θ from its crossing point). The $\phi(V_b)$ extracted in this way is shown in Fig. S18 (red dots) and compared to the theoretically calculated curve based on the inverse compressibility of TBG and MLG, which add up in series in the electrostatic solution. We see that these agree well. This method of using the MLG probe dispersion, which inherently appears in our measurements, allows us to convert the V_b

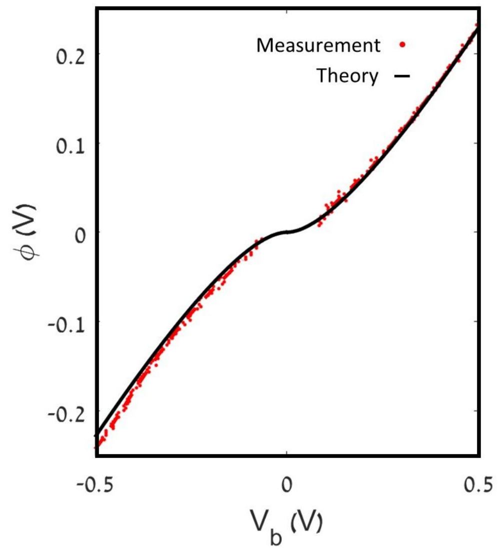

Figure S18. The energy band shift, ϕ, as a function of V_b, (red dots) extracted from the feature that follows the MLG bands in the MLG-TBG measurement in Fig. 3d (see text above). The black line is the theoretically calculated $\phi(V_b)$ using the inverse compressibilities of MLG and TBG and a geometric capacitance $C_g = 1.90\mu F/cm^2$.

axis into an energy shift and plot directly the $E(k)$ relation of the imaged bands. We have used this in Figs 3 and 4 to directly compare the bands with the theoretical predictions.

S9. MLG-TBG tunneling experiments

In Fig. 3d we presented the measured second derivative d^2I/dV^2 vs. θ and V_b in the MLG-TBG tunneling experiment, since this derivative showed most prominently the various momentum-resolved features. For completeness, we present below also I and dI/dV measured in this experiment.

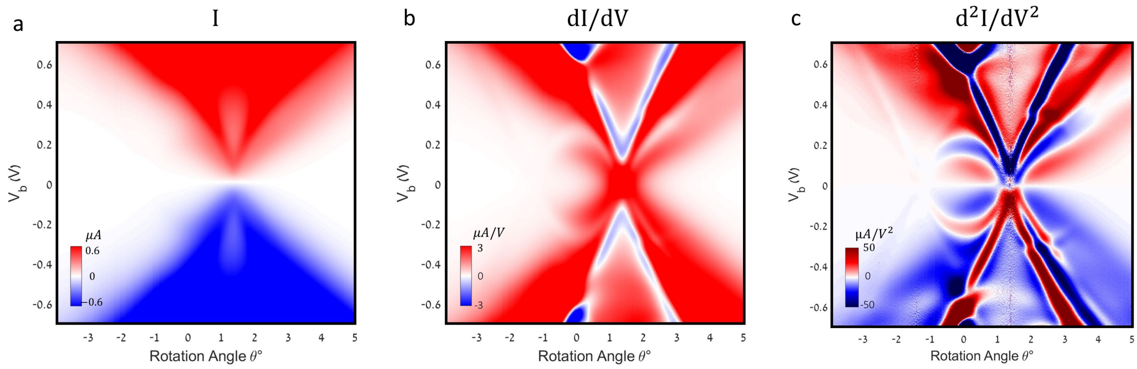

Figure S19. The tunneling current and its derivatives in the MLG-TBG momentum-resolved experiment of Fig 3: a. I **b.** dI/dV and **c.** d^2I/dV^2 vs. θ and V_b. Panel 3 is reproduction of Fig. 3d in the main text. The first two are measured directly (DC and lock-in), the third is obtained by a numerical derivative of the second.

S10. Line traces from the MLG-MLG and MLG-TBG tunneling experiments

Below we present line traces through the MLG-MLG and MLG-TBG tunneling experiments in Figs. 2 and 3 in the main text. Fig. S20b-d present the measured I, dI/dV and d^2I/dV^2 vs. θ, taken along the horizontal dashed line in the MLG-MLG experiment shown in Fig20a. Fig S20e-g present the measured I, dI/dV and d^2I/dV^2 vs. V_b along the vertical dashed line in Fig20a. Fig S21 presents similar linecuts through the MLG-TBG experiment.

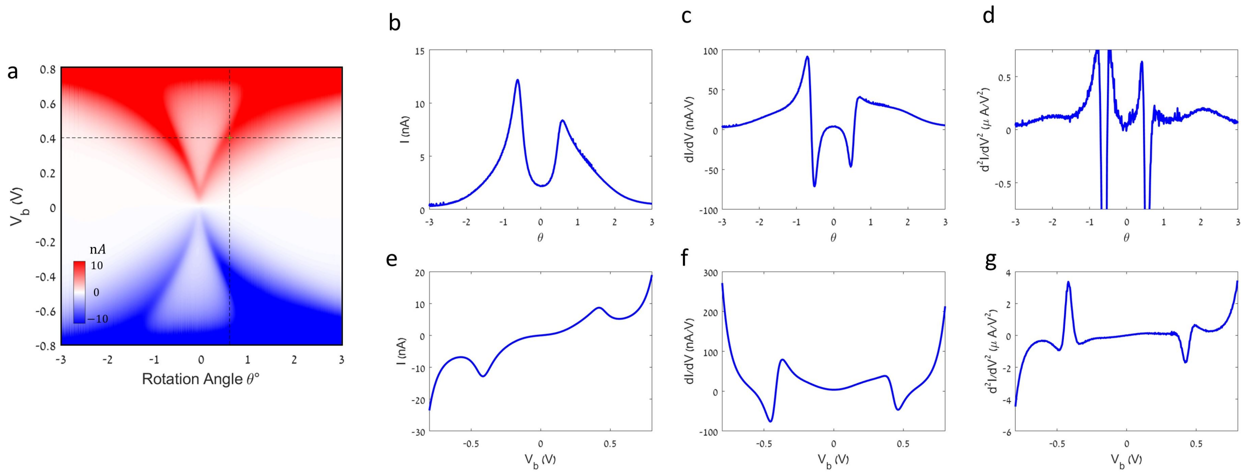

Figure S20. Linecuts through the MLG-MLG tunneling experiment: a. The tunneling current, I, vs. rotation angle, θ, and bias, V_b, in the MLG-MLG experiment (reproduced from Fig. 2a). **b-d** Traces of I, dI/dV, and d^2I/dV^2 vs. θ, taken from a linecut along the horizontal dashed line in panel a. **e-g** Traces of I, dI/dV, and d^2I/dV^2 vs. V_b, taken from a linecut along the vertical dashed line in panel a.

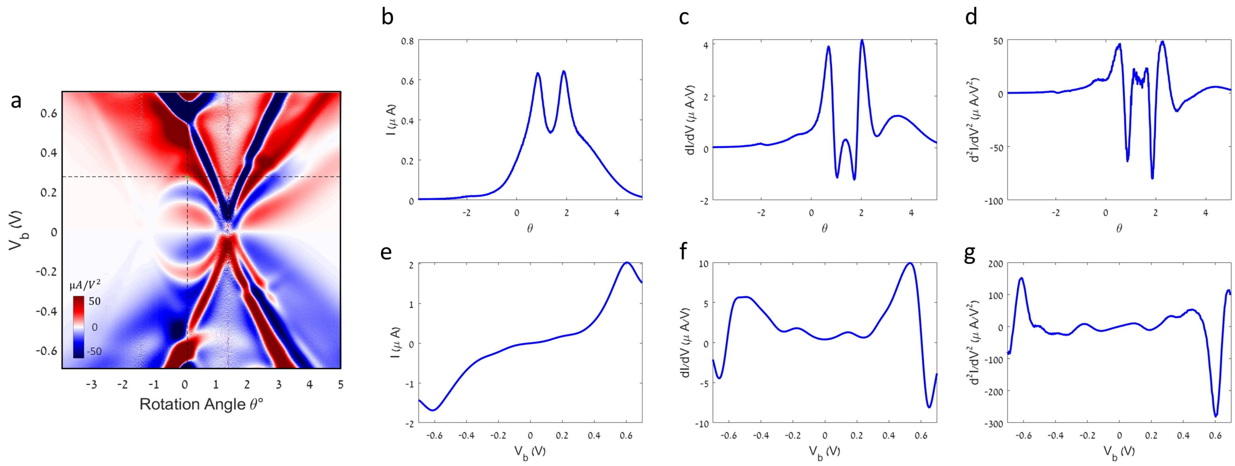

Figure S21. Linecuts through the MLG-TBG tunneling experiment: a. The tunneling d^2I/dV^2, vs. rotation angle, θ, and bias, V_b, in the MLG-TBG experiment (reproduced from Fig. 3d). **b-d** Traces of I, dI/dV, and d^2I/dV^2 vs. θ, taken from a linecut along the horizontal dashed line in panel a. **e-g** Traces of I, dI/dV, and d^2I/dV^2 vs. V_b, taken from a linecut along the vertical dashed line in panel a.

S11. Tunneling experiments of TBG with different twist angles

In this section, we present a tunneling experiment of an additional TBG sample, having a different twist angle than the sample presented in the main text. Fig. S22a shows the tunneling d^2I/dV^2 from MLG to a 3.4° TBG, measured vs. θ and V_b. The two layers are separated by a bilayer WSe_2 tunneling barrier. For comparison, we plot in panel b the measurement with the 2.7° TBG, reproduced from the main text (Fig. 3d). Visibly, both experiments exhibit similar features: in both cases we observe the "flat" bands of the TBG, superimposed with the MLG Dirac spectrum. The separation of the two Dirac points of the "flat" bands in the θ axis is different in the two measurements: it is 3.4° in panel a, and 2.7° in panel b, directly reflecting the twist angle of the corresponding TBG. In both cases the amplitude of the observed features gradually decreases when going from θ in which the MLG probe is aligned with the top layer of the TBG to the angle where it is aligned with the bottom layer. As explained in the main text, this reflects the momentum-dependent layer polarization of the wavefunctions. Visibly, the measured 3.4° TBG "flat" bands extend to larger bias than the 2.7° TBG bands, reflecting the difference of the bandwidth in the two twist angles. Correspondingly, also the remote band features, that are apparent in the 2.7° TBG measurement are pushed to higher bias in the 3.4° TBG measurement, outside our measurement window. However, one can still see in the measurement a strong positive d^2I/dV^2 feature at positive V_b, which is a precursor of these remote bands' features. Overall, the features observed in this experiment reproduce very nicely the features seen in the sample presented in the main text.

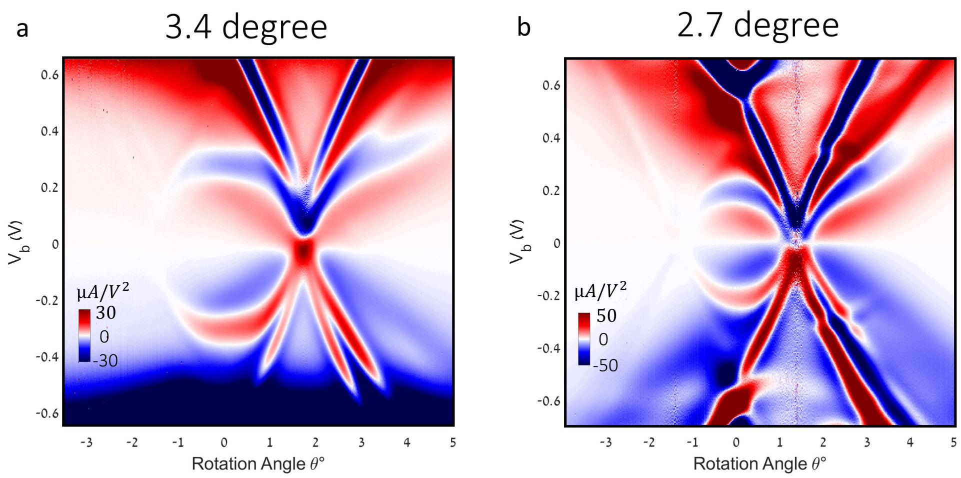

Figure S22: **Comparing momentum-resolved experiments on TBG samples with different twist angles: a.** Measured d^2I/dV^2 vs. θ and V_b between MLG and a 3.4° TBG. **b.** The experiment with 2.7° TBG, reproduced from Fig. 3d. In both cases the barrier is bilayer of WSe_2.

S12. Experiments with different tips and different measurement schemes

An important question is to what extent the experiments that we showed depend on specific details of the specific tip that measured them. In the current work we did not attempt to precisely control the shape and size of the plateau formed at the tip axes. We did produce more than a dozen tips that worked well (success rate larger than 50%) which had different plateau areas and differently oriented edges. In this section we provide more data that includes AFM pictures of various tips, show measurements of the same sample with two different tips, and compare the results of different measurement modalities.

S12.1 Tip Shapes

Figure S23 shows five tips whose geometry we carefully examined by imaging them using a standard AFM. The top row shows the AFM topographies of the vdW tents, composed in all cases of MLG/hBN/graphite. The bottom row zooms-in on the top part of the pyramids and shows the AFM's "error" signal, which highlights the different angles of different regions. Note that these **AFM images were taken after the tips participated in the QTM experiment**. In some of these images one can clearly see "junk" that was accumulated during the experiment in the sloped regions of the tent adjacent to the plateau (e.g. top right in Tip C).

The tips have common features, but they also have variability in some of the details. Specifically, we see in all cases folds that climb up the tent to the top (dashed white lines). Quite generally, these folds determine the corners of the formed plateau. In some cases, the AFM images of the tips show a clear flat plateau (e.g. Tip B). In other cases, there is a clear plateau, but it is composed of two regions with slightly different tilt angle (e.g. Tip A), and in some cases, a plateau is not even clearly discernable (e.g. Tip E). **Strikingly, all the tips shown here have led to clean data, of comparable quality to the one shown in the main text**. This highlights the fact that, most likely, upon contact with the sample, even with minimal force, these small imperfections are ironed out and a clear plateau is formed.

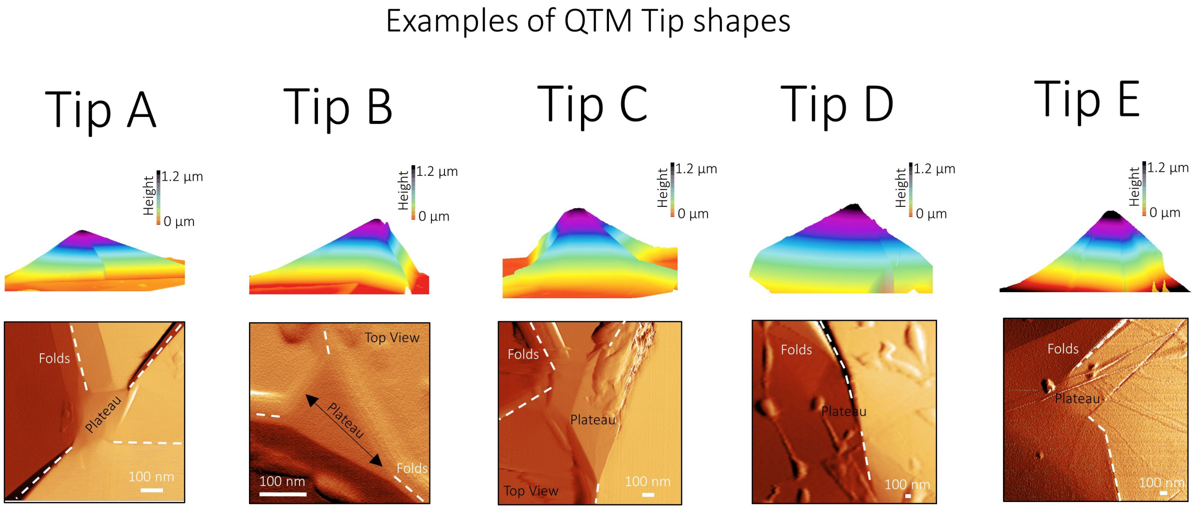

Figure S23: **Various QTM tips.** Top row shows the AFM topography of the vdW tent, consisting in all cases of MLG/hBN/graphite. The bottom row zooms-in on the top part of the pyramids and plots the AFM's "error" signal, which highlights the different angles of different regions. The folds climbing up the tent are marked by dashed white lines, and the top plateau is labeled. Note that these images were taken after the QTM experiment was performed.

S12.2 Comparing measurements of the same TBG sample with two different tips

Panels a and b in Fig. S24 compares the measurements of the same 3.4° TBG sample with two different tips (Tips D and E from Fig S23). Although the tip shape, size and plateau edges are different between the two tips, the measurements are almost identical. This includes the position and relative magnitude of the TBG bands, the MLG bands, and the dependence at high bias. In panel b we see mild doubling of the MLG Dirac feature with an angular shift between the two copies. This most likely reflects the existence of a second region in the tip's plateau with a slightly different twist angle.

S12.3 Comparing measurements with different contact modalities

Figure 24 also compares a measurement performed with two different contact modalities: panels a and b show measurements performed using a "continuous contact" method, namely the tip remains in contact with the sample throughout the experiment, specifically, when we rotate from one angle to another. Panel c shows measurement done with the same tip and sample as in panel b but in a "woodpecker" modality, namely, the tip and sample are brought into contact, measurement is performed as a function of V_b, then only after the tip is retracted (detached) from the sample, we perform the angular rotation. The tip is then re-approached into contact with the same force (determined by the AFM feedback) and this sequence is looped. Remarkably, the measurements using the two methods are practically identical. A tiny difference is that in the "woodpecker" scheme we see very few isolated (random) angles where the signal decreases, most likely because the detachment in this scheme allows in rare cases for contaminants to penetrate into the junction area. From this perspective, the "in contact" method is practically perfect – we do not see even a single line where this happens, emphasizing that the junction remains clean throughout the entire experiment. Also note that in the "woodpecker" scheme the tip approaches and withdraws hundreds of times and still maintains a perfectly continuous tunneling spectrum. This shows that the shape of the tip is not modified by detaching and reconnection.

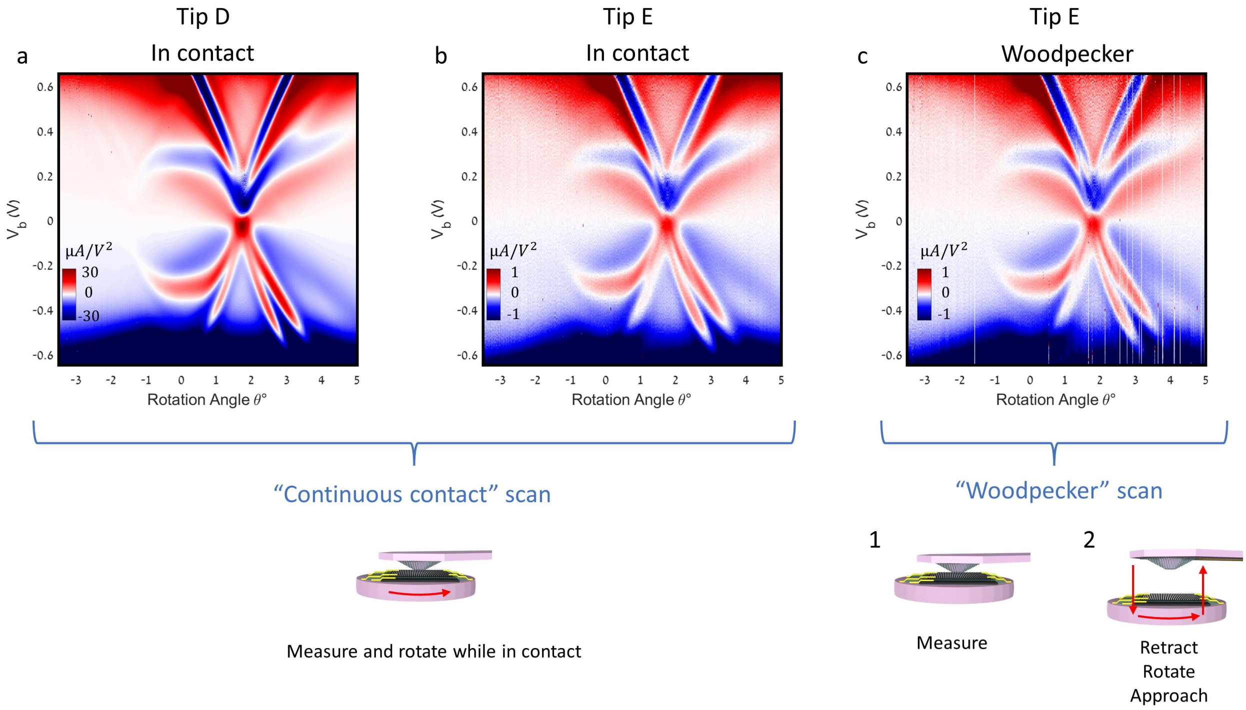

Figure S24. Measurements of the same TBG sample with two different tips and two different contact methods: a. d^2I/dV^2 vs. θ and V_b for a 3.4° TBG sample probed with a tip D from Fig. S23. **b.** The same measurement, done on the same sample, but now with tip E. Both these measurements are in a "continuous contact" modality, were the tip remains in contact with the sample throughout the entire experiment, and specifically during the rotation from one angle to another (see bottom illustration) **c.** Same measurement and tip as in panel b, but now done in the "woodpecker" modality (see bottom illustration). Here we measure when the tip and sample are in contact, then retract the tip, rotate the sample, and reapproach into contact. This sequence is repeated for every step of the rotation angle. Remarkably, the measured data is almost identical in all three panels (see text).

S13. Quantum coherence of the QTM junction and a lower bound on the coherence length

Momentum resolved tunneling relies on interference between multiple tunneling paths. A prerequisite for this tunneling is that the wavefunctions on both sides of the junction are quantum coherent over a certain length. Momentum-resolved tunneling can therefore be suppressed by decoherence, which is generally caused by electron-electron scattering, electron-phonon scattering, or any other inelastic scattering channel.

In addition to decoherence, momentum-resolved tunneling can be suppressed also by elastic scattering. For example, lattice defects or other impurities can provide momentum kicks and break the momentum conservation in the tunneling.

The observation in this paper of momentum-conserving tunneling at room temperature therefore directly demonstrates minimal disorder and phonon scattering as well as quantum coherence in the junction, even at this temperature.

Quantitatively, the measurements provide a lower bound on the elastic scattering length, $l_{elastic}$, and coherence length, l_{ϕ}. At low bias, the level to which momentum is conserved in the tunneling can be estimated by the angular width ($\sim 0.2°$) of the tunneling peak around $\theta = 0°$, presented in Fig. S13. By fitting this peak to a theory that includes the broadening by room temperature ($K_B T = 25meV$) we obtain a lifetime broadening of $\gamma_0 = 4meV$. Using the Fermi velocity of graphene ($v_F = 1 \cdot 10^6 m/s$) we can convert this lifetime to a scattering length of $l = 150nm$. This length is given by the minimal length out of $l_{elastic}$, l_{ϕ} and the tip diameter, L. Since the diameter of our tip is of the order of $150nm$ it is likely that this is the critical length that limits the momentum resolution, even at room temperature. This measurement therefore provides a lower bound on the elastic scattering length and the coherence length at room temperature, $l_{elastic}$ & $l_{\phi} > 150nm$.

S14. Additional details

Translation invariance of the experiments – all the experiments described in the main text show translational invariance to an excellent degree. We can determine that in two independent ways:

a) Repeating the experiment at different locations in a large sample resulted in almost identical results (in this study, the samples were relatively simple and therefore were large in size > $10\mu m \times 10\mu m$).
b) In the experiments, we use X and Y nano-positioners to bring the point of interest in the sample to the center of rotation. However, commercial piezo-rotators are generally not perfect – upon rotation they displace in the lateral directions. In our setup this displacement was of the order of $\sim 1\mu m$. Although it is possible to reduce this displacement through various technical solutions, we found that it was not necessary to do so in the experiments presented in the main text. The fact that this displacement is not visible in the experiment attests to the high degree of translation invariance of the samples and the measured physics.

The above observation is especially interesting when referred to the measurements done on TBG. This system notoriously suffers from twist angle disorder, namely, a variation of the twist angle with the lateral position. The momentum-resolved data, measured for 2.7° TBG (Fig. 3, 4), allows us to determine the local twist angle by measuring the separation between the two copies of the MLG Dirac cones. From such measurement done at points separated a few microns apart we see that the twist angle remains constant across the sample to within a $\sim 0.1°$ accuracy.

Why we observe a conductance peak only for the 1st commensurate angle $\mathbf{21.8°}$ (and in its symmetric partner, $\boldsymbol{\theta = 38.2°}$) – Ref. [4] predicted a series of conductance peaks at various commensurate angles, whereas our experiment sees only a single peak at $\theta = 21.8°$ (and its symmetric partner at $\theta = 38.2°$). We do not observe the additional peaks because their tunneling amplitudes are orders-of-magnitude smaller. The peaks at $\theta = 27.8°, 32.2°$ are predicted to be two orders of magnitude weaker than that at $\theta = 21.8°$ and those at $\theta = 13.2°, 46.8°$, are four orders of magnitude smaller. At room temperature, these peaks are well buried within the gradually changing background due to inelastic scattering by thermal phonon, seen in our measurements for θ varying from 0° to 30° .

The exponential decay of the peak conductance along the series of commensurate angles is deeply rooted in their momentum-resolved nature: As explained in ref [4] .whenever the lattices are commensurate in real space, their Brillouin zones and wavefunctions are commensurate in momentum space. For instance, at $\theta = 21.8°$ the two latices are commensurate in real space with a $\sqrt{7} \times \sqrt{7}$ unit cell, and concomitantly their Dirac cones are overlapping in the third Brillouin zone (Fig. 1i). If the conductance peaks were a real-space phenomena – namely they would reflect a better real-space overlap of atoms of the two layers at commensurate angles, one would expect a weak algebraic decay of the peak amplitude, proportional to the fraction of atoms that overlap between the two layers ($f = \frac{1}{2}, \frac{1}{14}, \frac{1}{26}, \frac{1}{38}$ for $\theta = 0°, 21.8°, 27.8°, 13.2°$). However, as shown in Ref. [4], the peak amplitude is related to the matrix element in k-space that describes the coupling of the wave functions between two layers, i.e., the k-dependent interlayer tunneling amplitude squared, $|t_k|^2$. The tunneling amplitude, t_k, decays exponentially with k. Physically, this is because the components of the wavefunction with larger momentum in the plane decay much faster in the barrier in the perpendicular direction to the plane. The $21.8°$ peak, corresponds to a matching of Dirac cones at the corner of the 3rd BZ. The corresponding $|t_k|^2$ is predicted[4] to be two orders of magnitude smaller than the $|t_k|^2$ that corresponds to zero twist angle, where the matching happens at the corner of the 1st BZ. For the next commensurate angle, at $27.8°$, $|t_k|^2$ is expected to be smaller by an additional two orders of magnitude, and this trend continues for the next commensurate angles. Thus, the peaks at these angles are expected to be tiny, and to be swamped by the gradually decreasing background due to inelastic tunneling by thermal phonons, observed in our measurements.

Considerations for future QTM experiments with other layered materials

Materials that can be compatible with QTM: In this work we examined devices that consist of graphene, hBN and WSe_2 layers. The method, however, should be applicable to a wider range of 2D materials. In principle, any layered material that could be stacked using conventional transfer techniques should possibly be compatible with the QTM. The list includes semiconductor TMDs such as $WS_2, WSe_2, MoS_2, MoSe_2$ superconductors such as $NbSe_2$, 2D magnets, and possibly even layered high temperature superconductors such as BSCCO. Naturally, experiments with different material systems pose different challenges and require different considerations, few of which are discussed below.

Geometry: The geometry of the QTM tips and samples are generally compatible with the broad range of cleavable/exfoliatable layered materials. The flat side (Fig. 1a) is practically identical to standard devices that are routinely prepared with these materials. The tip side (Fig. 1b) requires the vdW stack to form a sloped tent over a pyramid. TMD materials are often more fragile than carbon-based materials and hBN, however, since the bending angle between the plateau and the sloped part of the tent is very mild (often as small as 6deg) the TMD layers can conform to that without any damage to the layers.

Air sensitivity: Some of the interesting materials to study in QTM geometry are air sensitive (e.g. $NbSe_2$, 2D magnets, or BSSCO). In these cases, tips and samples should be prepared in a glove box and transferred under vacuum to the measurement setup. Our tip fabrication process works well in a glove box. An even simpler option, in the case of tunneling experiments that naturally require the presence of a tunneling barrier, is to prepare the air-sensitive TMD sample and tip and cap them with a thin layer of hBN in the glove box. This way they can be transferred more easily, in ambient room conditions, into a measurement setup. Even more sensitive materials (e.g. BSCCO) require to be exfoliated also at low temperatures. A possible, though not trivial way to experiment with such materials in a QTM junction is to use cold cleaving in a cryogenic-QTM, like is customarily done in many STM experiments.

Measurement principle – In this paper we demonstrated momentum-resolved measurements that used a MLG probe to image the band structure of another MLG layer or a TBG. In these measurements the Dirac points of the probing MLG traced an arc in momentum space of the sample, along which the energy bands were imaged. Probably the easiest way to do similar measurements in TMD materials would be to have the same material on the system and the probe. This will ensure that the relevant parts of the Fermi surface will be probed. To measure a twisted TMD structures on the sample side, one can still use a monolayer of the same TMD as the probe. This should work in the same way that TBG was probed using MLG (the cutting in k space naturally passed through the important points of the energy bands). The interpretation of the measurements becomes very simple if the probe has a small Fermi surface, ideally just a single point in k-space. This can be easily achieved with MLG graphene probe, by bringing it to its Dirac point. This could be similarly achieved in TMDs: since these materials usually have gapped parabolic bands at their K or Q points in the BZ, for low enough carrier densities, the Fermi surfaces can be very small, close to being just a point. TMDs are more susceptible to disorder than graphene, and this disorder might limit the lowest density one can achieve while still observing good momentum resolved data, and thus cleaner TMDs will allow to get closer to this ideal limit. Another important point is that experiments probing low energy physics would be typically done at small bias. In this case, adding carriers to the sample using a back gate will also add some carriers to the probe, owing to the finite compressibility of the former. In such experiments, it would be beneficial to have an additional top gate that will allow bringing the probe back to the neutrality point for each gating of the sample. Although this was not explored in the current paper, our QTM tips have a built-in graphite layer that can be used as such top gate.

S15. Supplementary Video 1 – Imaging TBG bands under pressure

The video shows the measured dI^2/dV^2 vs. θ and V_b for the junction shown in Fig. 3 (MLG - 2.7° TBG), under a full sequence of applied pressures from P = 0.01 to 0.68 GPa (9 separate measurements in total, as shown in Fig. 4d). The movie is looped to run back and forth from low to high pressure.

If there are gradients of pressure across the QTM tunnel junction, they will affect the tunneling spectrum. Most simply, such gradients will cause a broadening of the momentum-resolved features due to the spatial inhomogeneity. However, as can be seen in the attached video the momentum-resolved features remain narrow even at the highest pressure in our measurements, implying that the pressure gradients are not substantial.

S16. Supplementary Video 2 – Tracing the TBG flat bands with the Dirac point of MLG

The video shows the relative band alignment between the MLG and TBG bands when scanning along a path in the $\theta - V_b$ plane that traces a flat band feature, as outlined in blue in Fig. 3f. The band alignment is determined from the theoretical calculation in Fig. 3e. We see that this feature appears whenever the Dirac point of the MLG matches in energy and momentum a point on the TBG flat bands.

S17. Supplementary Video 3 – Tracing the MLG probe bands using the TBG Dirac point.

The video shows the relative band alignment between the MLG and TBG, when scanning along a path in the $\theta - V_b$ plane that traces the MLG feature, as outlined in purple in Fig. 3f. We see that this feature appears whenever one of the two Dirac points of the TBG bands matches in energy and momentum a point on the MLG bands.

S18. Supplementary Video 4 – Energy-momentum conserving states in MLG-MLG junction vs. twist angle

The video shows the band alignment as a function of rotation angle between two MLG layers, with a constant applied bias. The Fermi surfaces of the two MLG layers are marked by red and blue circles. The energy-momentum conserving states within the bias window are marked in yellow. The video visualizes the origin of 'onset' and 'nesting' conditions in our measurements.

S19. Supplementary Video 5 – Energy-momentum conserving states in MLG-TBG junction vs. bias

The video shows the band alignment as a function of bias between MLG and TBG, for $\theta = 0$. The Fermi surfaces in the MLG and TBG layers are marked as red and blue, respectively. The energy-momentum conserving states within the bias window are marked in yellow. The video visualizes the origin of 'onset' and 'Dirac-matching' conditions in our measurements.

References

1. Frisenda, R. *et al.* Recent progress in the assembly of nanodevices and van der Waals heterostructures by deterministic placement of 2D materials. *Chem Soc Rev* **47**, 53–68 (2018).

2. Fowler, R. H. & Nordheim, L. Electron emission in intense electric fields. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **119**, 173–181 (1928).

3. Britnell, L. *et al.* Field-Effect Tunneling Transistor Based on Vertical Graphene Heterostructures. *Science* **335**, 947–950 (2012).

4. Bistritzer, R. & MacDonald, A. H. Transport between twisted graphene layers. *Phys Rev B* **81**, 245412 (2010).

5. Kim, K. *et al.* Band Alignment in WSe_2 –Graphene Heterostructures. *ACS Nano* **9**, 4527–4532 (2015).

6. Bistritzer, R. & MacDonald, A. H. Moiré bands in twisted double-layer graphene. *Proceedings of the National Academy of Sciences* **108**, 12233–12237 (2011).

7. Koshino, M. *et al.* Maximally Localized Wannier Orbitals and the Extended Hubbard Model for Twisted Bilayer Graphene. *Phys Rev X* **8**, (2018).

8. Elias, D. C. *et al.* Dirac cones reshaped by interaction effects in suspended graphene. *Nat Phys* **7**, 701–704 (2011).

9. Yu, G. L. *et al.* Interaction phenomena in graphene seen through quantum capacitance. *Proceedings of the National Academy of Sciences* **110**, 3282–3286 (2013).

10. Prasad, N. *et al.* Quantum Lifetime Spectroscopy and Magnetotunneling in Double Bilayer Graphene Heterostructures. *Phys Rev Lett* **127**, 117701 (2021).

11. Lin, K. A. et al. Emergence of Interlayer Coherence in Twist-Controlled Graphene Double Layers. *Phys Rev Lett* **129**, 187701 (2022).

12. Mishchenko, A. *et al.* Twist-controlled resonant tunnelling in graphene/boron nitride/graphene heterostructures. *Nat Nanotechnol* **9**, 808–813 (2014).

13. Wallbank, J. R. *et al.* Tuning the valley and chiral quantum state of Dirac electrons in van der Waals heterostructures. *Science* **353**, 575–579 (2016).