Peer Review File

Manuscript Title: Continuous Mott transition in semiconductor moiré superlattices

Editorial Notes: *none*

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

Two canonical forms of metal-insulator transitions (MITs) are the filling-controlled and bandwidth-controlled transitions. While continuous filling-controlled MITs have been observed in doped semiconductors, tuning the bandwidth in real materials often couples with structural changes that give rise to first-order phase transitions. The manuscript by T. Li et al. reports the observation of a continuous MIT in a TMD-based moiré superlattice as a function of the perpendicular electric field at a fixed carrier concentration. The authors attribute the observations to a bandwidth-controlled MIT where the electric field tunes the moiré subband's width. They also show the absence of long-range magnetic order in the insulating states, hinting at possible spin frustration in the system.

Overall I think the manuscript would be of great interest to researchers in the fields of 2D materials and correlated electron systems. It demonstrates the moiré system as a unique platform that allows for tuning of both filling and bandwidth. However, I think the connection between the observations and the conclusions needs to be strengthened. The transport data is very convincing in showing the MIT, but other possible mechanisms for the MITs shall be considered, and more details on data processing shall be provided. The conclusion of spin frustration also needs to be better supported by the experiments. I would be happy to recommend the paper's publication if the authors could address the following comments.

A. Mechanism of the transition:
The manuscript's main argument is that the applied electric field increases moiré bandwidth, which reduces the U/W ratio and drives a continuous Mott transition at a fixed filling ratio. However, the situation might not be as simple as proposed. I wish the authors could provide additional evidence to support their conclusion. Below I list some scenarios which might also explain the observations.

1. The first scenario is related to the reduced electron confinement along the out-of-plane direction. Because the applied electrical field reduces the band offset and is quite large (~1.3V/nm), it is possible the holes can partly transfer into WSe2 (considering the original band offset is 220 meV and interlayer distance is ~0.7 nm). If this happens, it will lead to a change in the moiré bands' filling ratio because of the additional states coming from WSe2 orbitals.

2. My second concern about the mechanism is that even if we only consider the dispersion in-plane, there seem to be additional effects other than simple bandwidth tuning. Based on Fig. 1d, we see that the first and second moire bands merge at a large field, but each band's bandwidth does not seem to change much (perhaps by ~20%). This suggests that band merging may be a more dominant effect, which changes the filling ratio and induces an MIT. (In fact, Fig. 1c shows that the two bands already merged below E_c, at least for f=2).

B. Analysis and interpretation of transport data:

The values of T_0 and activation gap are extracted to assign the quantum critical regime, which is important to reach the conclusion. I had several comments regarding how one can reliably calculate these parameters.

1. First, it is not clear how the exact values of T_0 are determined from the data. For example, it seems that one can multiply all the T_0 values by a constant and still get the same collapse shown in Fig. 3a. Besides, the choice of T_0 for the metallic and insulating side is also not clear to me. Are they chosen so that Fig. 3a appears symmetric? If this were the case, the T_0 lines in Fig. 3 might not have much significance.

2. I also suggest the authors add more intermediate steps for extracting T_0 as supporting information, such as showing a few representative curves. For example, from raw data in Fig. 2a, it is hard to directly see how the curves at E= 1.306 V/nm and E = 1.309 V/nm can be collapsed onto E = 1.342 V/nm.

3. The authors fit the high-temperature insulating R-T data in terms of thermal activation but attribute the same range of metallic data to electron scattering, although they qualitatively look quite similar in Fig. 2. More evidence is needed to justify this. For example, it would be useful to show that the thermal activation model fails to fit the data above E_c.

4. I am also confused by the attribution of the Pomeranchuk effect in transport. In Fig. 1 and Extended Fig. 4, the metallic R-T is interpreted as Fermi liquid behavior, but similar data in Extended Data Fig. 7 is interpreted as the Pomeranchuk effect. Am I missing something?
In fact, if the increase in R is due to the Pomeranchuck effect, would we expect the higher temperature range of the R-T curves to arise from the insulating states?

5. The extended data Fig. 2 shows the transport properties of f=2 (band insulator and metal). It is quite surprising that the metallic phase does not show any Fermi liquid behavior. This could be concerning, maybe signaling additional temperature-dependent extrinsic factors such as disorder and non-Ohmic contact, etc. Can the authors comment?

6. The transport behaviors reported here are remarkably similar to gate-induced MITs in conventional 2D electron systems. The authors highlight their difference on p. 5, stating that 'In contrast to … conventional 2D electron systems… here … electric field ... modifies … U/W'. However, I am not entirely convinced that there is a dramatic difference: in the conventional 2D systems, one cannot define a 'half-filling. One may think of doping in conventional 2D systems as changing the ratio of Coulomb to kinetic energy, i.e., U/W, which seems to suggest similar physics as here.

C. Magnetic characterizations:
1. The authors find no evidence for spin ordering down to 1.4 K in the insulating state and extract a Curie-Weiss temperature of 30-40 K from optical measurements in the metallic state. Based on the ratio of the two temperatures, the authors suggest possible spin frustration. I am not sure how valid it is to assume that the insulating states would have the same exchange interaction as the metallic states. Because the Wannier function changes as a function of E-field, it may not be surprising if the strength of exchange interaction is different in two states.

2. In addition, probing the magnetic ordering through the optical measurements seems to be a rather indirect probe since the interaction between spin and excitons can be quite complicated. For example, it would seem more natural to probe the magnetic ordering using the excitons in MoTe2 but not WSe2 because holes are in the MoTe2 layer.

3. The MCD data also appears not straightforward to interpret. For example, it seems that one can get different magnetic susceptibility depending on the probe wavelength. Could the authors comment on this? It would also be great if the authors show a few representative I_L and I_R

spectra at different E-field and explain how to eliminate the intrinsic response from WSe2.

Some other minor comments:
Do the authors know the rotation angles between the two TMD layers being 0 or 60 degrees? It would be nice to show the second-harmonic generation data. I imagine the 0-degree and 60-degree would have different moiré potential and electrical dependence thereof.

Figure 2b: it would be useful to specify the meaning of different symbols in the caption (right now only labeled in the figure).

Extended Data Fig. 7: please specify what the vertical dashed line means (f=1?).

Referee #2 (Remarks to the Author):

In this work, the authors study the metal-to-insulator transitions (MIT) in the MoTe2/WSe2 moire superlattice. In particular, they found that the MIT at the half-filling of the first hole moire band is consistent with universal critical theory of a continuous Mott transition from a Fermi liquid to a nonmagnetic Mott insulator in two dimensions. Some other works may have observed similar phenomena in other moire systems in the transport measurements, such as [Nature Physics 15, 237–241(2019)] and [Nature Materials 19, 861–866(2020)], but this work provides compelling evidences for the observed MIT at the half-filling as a Mott transition and provides important characterization of the electronics states across the transition. It demonstrates the advantage of using moire superlattice in simulated critical quantum phenomena and shows promises in further investigation of new exotic state of matter near the transition. Therefore, I think this work is of interest of many people in the field of two-dimensional materials and strongly-correlated physics and I recommend it to be published in Nature. Nevertheless, I have a few concerns of the manuscript as listed below:
1. They claimed that the electric field varies the potential different between the two TMD layers and subsequently the moire potential. I agree with the first half of the statement, but why varying the potential different between the two TMD layers necessarily affect the depth of the moire potential? In the situation that the moire potential is dominated by the moire structure and internal strain as claimed in recent studies (Shabani, et. al. Nature Physics. (arXiv:2008.07696) and Li, et. al. Nature Materials (arXiv:2007.06113)), this is certainly not the case. Their DFT calculations also show that for a out-of-plane electric field up to 2V/nm, the band width barely changes.
2. I am a bit puzzling by the critical behavior in the DFT calculations. Why the band gap between the first moire band and the second moire band suddenly collapses and the band width of the first moire band suddenly increase rapidly when the displacement field is above 2V/nm? The band structures in Figure 1d seems only have a small number of data points (maybe only ~5 kpoints along Gamma-M direction?). Is it enough to sample correctly the band width and the band gap? How to defined the bandwidth of the first moire band when the bandgap between the first and the second moire bands collapses? It might be beneficial to the put all the band structures near the critical point in the supplementary information and add a little more discussion.
3. Moreover, when the displacement field is above 2V/nm, the original ~220 meV band offset between MoTe2 and WSe2 should be significantly reduced. What are the characters of the electronic states in the first and the second moire bands? Are they all contributed from the MoTe2 states under zero displacement field and large displacement field, or one of them is contributed from the WSe2 states?
4. The authors argue that the tunning of U/W by the electric field drive the Mott transition. But the Mott transition seems to happen after the band gap between the first and the second moire bands collapses, it may not be appropriate to neglect the effect of the second moire bands in the discussion.

5. The scales of the x-axis for the upper and the lower panels of figure 1e seems to be different. Is so, they shouldn't share the same labeling.
6. Near at the end of the second paragraph on page 2, it says "…., which changes the size of the localized Wannier function and the bandwidth predominantly." What supports such statement? Have the authors calculated the Wannier function in the system?
7. In the third paragraph on page 2, it says "The valley degeneracy is lifted by valley pseudospin-orbit coupling in small period moiré structures." Could the authors provide more explanation or relevant references?
8. At the end of the page 3, it says "The square resistance exceeds the Mott-Ioffe-Regel limit …". It may be better to provide a few words and link it to the dash line in Fig. 2a to make it clear for the readers.
In Extended Data Figure 4, for the displacement field very close to the critical field (e.g., 0.0019 and 0.0026), the square resistant rises up at very low temperature and deviates from the T square dependence. Is it related to the quantum fluctuations near the MIT? Different from the other figures, the values of the displacement field are listed up to the fourth digit after the decimal point. What is the actual accuracy of the value of the displacement field?

Referee #3 (Remarks to the Author):

In this manuscript, Li et al. demonstrated a continuous Mott transition at f=1 filling in MoTe2/WSe2 moiré superlattices. A quantum criticality is extracted through the scaling behavior of the resistances. Here are several comments below:

1.The evidence of continuous Mott transition is only based on an Arrhenius analysis of the resistivity. In Extended Data Figure 3, the authors extract the activation gaps above 10K. However, in Figure 2, even below the critical electric field, above 10K, there still exist insulating behaviors. In other words, a so-called Arrhenius behavior could be also observed. In this case, can the activation gaps extracted through Arrhenius analysis could be used as evidence to prove the (Mott) insulators?

2.Aside from the temperature dependence of resistivity (Rxx) at f=1, do the author have other evidence, such as Hall density, SdH oscillations to support the reconstruction of Fermi surface, or spectral gaps in compressibility? There could be multiple reasons causing resistance peak, such as charge density waves, van Hove singularities, etc. How could the author exclude the other possible reasons and justify the insulator?

3.The resistance value is the key to the overall analysis of the manuscript. Could the author show the I-V curves of the sample around f=1? If I-V curves are non-linear, the author should clarify the origin.

4.In heavy fermion systems, the non-Fermi liquid behavior is usually due to the presence of a nearby quantum critical point. Here when the electric field (E) approaches critical field (Ec), the temperature dependence of resistivity is still described by the Landau Fermi liquid theory. On one hand, the author gives a trend of diverging effective mass based on the fitting of Fermi liquid. On the other hand, the resulted diverging effective mass is contradicted with the Fermi liquid description for the normal metal. How to reconcile the contradiction? In Figure 2c, when E is getting very close to Ec (E-Ec ~0.003V/nm), the A^1/2 deviates from the power-law fitting. Could the author justify these discussions?

5.In Figure 1c, the width of the resistance peak at f=1 is much wider than that at f=2. Could the author clarify the origin of this difference?

6.In Extended Data Figure 3, at varying electric fields the resistance saturates around 3K. What

causes such saturations at the relatively "high" temperature compared to 0.285K? Is it due to RF noise in the system?

Overall, the study is interesting and there exists novelty in the manuscript, but to reach a positive recommendation, additional data and discussions are needed.

Author Rebuttals to Initial Comments:

Referee #1

Two canonical forms of metal-insulator transitions (MITs) are the filling-controlled and bandwidth-controlled transitions. While continuous filling-controlled MITs have been observed in doped semiconductors, tuning the bandwidth in real materials often couples with structural changes that give rise to first-order phase transitions. The manuscript by T. Li et al. reports the observation of a continuous MIT in a TMD-based moiré superlattice as a function of the perpendicular electric field at a fixed carrier concentration. The authors attribute the observations to a bandwidth-controlled MIT where the electric field tunes the moiré subband's width. They also show the absence of long-range magnetic order in the insulating states, hinting at possible spin frustration in the system.

Overall I think the manuscript would be of great interest to researchers in the fields of 2D materials and correlated electron systems. It demonstrates the moiré system as a unique platform that allows for tuning of both filling and bandwidth. However, I think the connection between the observations and the conclusions needs to be strengthened. The transport data is very convincing in showing the MIT, but other possible mechanisms for the MITs shall be considered, and more details on data processing shall be provided. The conclusion of spin frustration also needs to be better supported by the experiments. I would be happy to recommend the paper's publication if the authors could address the following comments.

A. Mechanism of the transition:

The manuscript's main argument is that the applied electric field increases moiré bandwidth, which reduces the U/W ratio and drives a continuous Mott transition at a fixed filling ratio. However, the situation might not be as simple as proposed. I wish the authors could provide additional evidence to support their conclusion. Below I list some scenarios which might also explain the observations.

1. The first scenario is related to the reduced electron confinement along the out-of-plane direction. Because the applied electrical field reduces the band offset and is quite large (~1.3V/nm), it is possible the holes can partly transfer into WSe2 (considering the original band offset is 220 meV and interlayer distance is ~0.7 nm). If this happens, it will lead to a change in the moiré bands' filling ratio because of the additional states coming from WSe2 orbitals.

Response 1

We would like to thank the reviewer for valuable comments, which have helped to strengthen our manuscript.

We first clarify the definition of the applied electric field that is used in the original manuscript. The electric field in the region above and below the moiré superlattice has different values since the two gates are not symmetric. In the original manuscript, we have quoted the sum of the two fields. Now we see that this potentially could cause confusion. We have redefined the field in the revised manuscript as the average of the two values, which is a more appropriate definition of the electric field in the device. Below in the reply we will refer to the average, not the sum, of the two fields.

We agree that charge transfer between the two layers will change the filling factor in each individual layer. We performed an independent optical study to determine the band offset as a function of applied electric field. We set the doping density in the moiré heterobilayers to be a constant (equal to the moiré density) throughout the measurements. At zero applied electric field, the band offset between the upper valence bands of the two layers is determined to be ~ 300 meV, which is slightly larger than the value from our DFT calculations. The value is reduced to $\gtrsim$ 100 meV at the critical electric field for the Mott transition. These values are obtained using the known applied field and the measured interlayer electric dipole moment ≈ 0.26 e*nm. (This value is smaller than the simple estimate from the interlayer distance of ~ 0.7 nm due to dielectric screening of charges, as reported by Nat. Nanotech. 16, 52–57 (2021)) This result verifies that there is no charge transfer between the layers for the entire range of electric fields investigated in this experiment.

Below we provide details on the optical measurements. The idea is based on the Stark effect for the interlayer exciton transitions (dashed black double-sided arrows in Fig. R1a, c). The energy of the interlayer exciton resonance disperses linearly with applied electric field and the slope gives the interlayer dipole moment. However, the interlayer exciton (spatially indirect exciton) couples very weakly with light because of the spatial separation of electrons and holes; it can be detected in absorption only when the interlayer and intralayer exciton transitions hybridize. We have recently demonstrated the method to study two other moiré heterostructures [Nat. Nanotech. 16, 52–57 (2021)].

As shown in Fig. R1, the $MoTe_2/WSe_2$ heterostructure has a type I band alignment. The lower valence band of $MoTe_2$ and the upper valence band of WSe_2 are close in energy and can be brought near resonance by a vertical electric field E. In 60-degree-aligned samples these bands have the same spins and can hybridize with each other, but in 0-degree-aligned samples these bands have opposite spins and cannot hybridize.

As expected, anti-crossing of the interlayer and intralayer exciton transitions as a function of electric field is observed only in 60-degree samples (Fig. R1c). The strong spectral feature around 1.69 eV is the fundamental intralayer exciton in WSe_2, whose energy nearly does not dependent on applied field.

Hybridization occurs and an anti-crossing feature of the exciton transitions emerges near $E \approx 0.32$ V/nm. The interlayer exciton transition disperses linearly with applied field. We determine the dipole moment d from the slope ≈ 0.26 e*nm. We estimate the offset between the two bands at zero field to be $d*E$ ~ 80 meV. Including the valence band spin splitting energy of $MoTe_2$ (~ 220 meV [Nano Lett. 14, 6231 (2014)]), we determine the upper valence band of $MoTe_2$ is ~ 300 meV above the upper valence band of WSe_2. Similarly, using the measured interlayer dipole moment we estimate the offset of the two top valence bands to decrease to ~ 130 meV at the critical field for the Mott transition $E \approx 0.65$ V/nm (1.3 V/nm in the original manuscript).

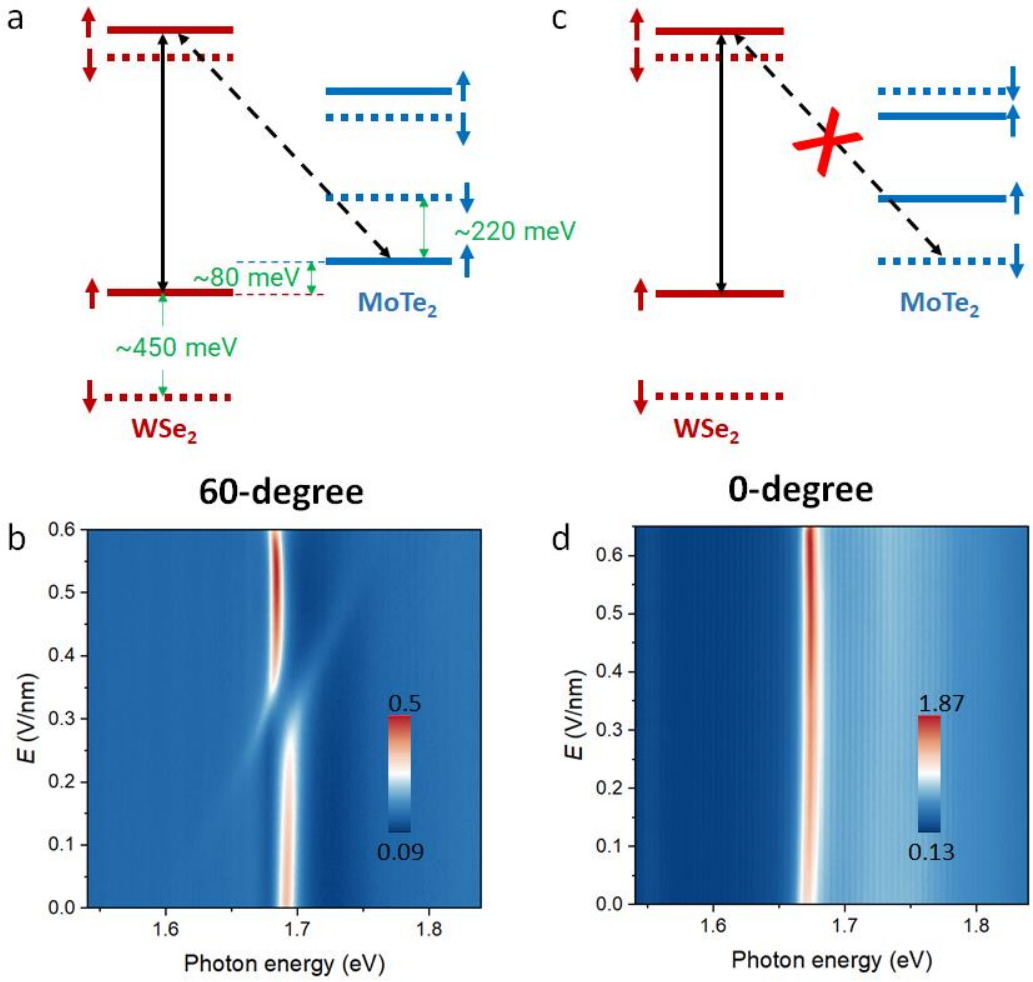

Figure R1. a, c, Spin-resolved band structure of 60- (**a**) and 0- (**c**) degree-aligned $MoTe_2/WSe_2$ heterobilayers under zero electric field. Dashed and solid double-sided arrows denote the relevant interlayer exciton transition and intralayer exciton transition (the fundamental exciton in WSe_2), respectively, for the spectroscopy measurements. Hybridization of the two transitions is possible when the WSe_2 upper valence band becomes nearly degenerate with the $MoTe_2$ lower valence band. This is allowed only in the 60-degree samples since the two valence bands have the same spins. **b, d,** Optical reflection contrast spectrum near the WSe_2 fundamental exciton resonance as a function of applied electric field at fixed doping density of $f = 1$ in 60- (**b**) and 0- (**d**) degree samples. Hybridization of the intralayer exciton resonance (strong feature that does not disperse with field) and the interlayer exciton resonance (weak feature that disperses linearly with field) is observed in 60-degree samples. A dipole moment of ≈ 0.26 e*nm is determined from the electric field dispersion of the interlayer exciton

resonance. The dipole moment allows us to determine the relative band offsets in the heterobilayer as a function of applied electric field.

Changes:

We have included Fig. R1 as Supplementary Fig. S1 in the revised manuscript. We have also commented on the absence of interlayer charge transfer in the revised Methods (page 9).

2. My second concern about the mechanism is that even if we only consider the dispersion in-plane, there seem to be additional effects other than simple bandwidth tuning. Based on Fig. 1d, we see that the first and second moire bands merge at a large field, but each band's bandwidth does not seem to change much (perhaps by ~20%). This suggests that band merging may be a more dominant effect, which changes the filling ratio and induces an MIT. (In fact, Fig. 1c shows that the two bands already merged below E_c, at least for f=2).

Response 2

We thank the reviewer for the comment. We agree that the effect of an electric field on the moiré system is likely more complicated than the simple picture of tuning the bandwidth. However, with all the measurement and calculation results that are available at the moment, we believe that tuning the bandwidth is the dominant effect for the first Hubbard band that is relevant for the reported Mott transition at fixed $f = 1$.

We should also have made it clear in the original manuscript that the DFT calculations were performed under zero doping. Since the moiré system is strongly correlated, the band gap and width of the single-particle moiré minibands from the DFT only serve to provide a qualitative picture of the Hubbard band at $f = 1$. Detailed quantitative agreement is not expected.

We performed a new capacitance measurement to demonstrate the large bandwidth tuning effect for applied electric fields below the critical field. Figure R2a shows the differential top gate capacitance as a function of bottom gate voltage at varying top gate voltages. The gate capacitance measures the electronic compressibility of the moiré superlattice (see arXiv:2102.10823 for details). Each capacitance dip corresponds to an incompressible state. In addition to the insulating states at $f = 1$ and $f = 2$, we observe insulating states at $f = 1/3$ and $f = 2/3$, which are generalized Wigner crystal states [e.g. Nature 579, 359 (2020) and Nature 587, 214 (2020)]. These states were not well resolved in the transport measurement due to the large sample-contact resistance for $f < 1$. The capacitance data shows that with increasing electric field, the $f = 2/3$ state weakens and disappears while the $f = 1$ and $f = 2$ states remain robust (the $f = 1/3$ state is outside the gate voltage range and cannot be accessed).

Since both the $f = 1$ and $f = 2$ states remain robust at the MIT for the $f = 2/3$ state, the filling factor cannot be changed by the applied electric field. The $f = 2/3$ state is stabilized by the nearest-neighbor Coulomb repulsion V, which is nearly independent of the electric field. We estimate $V \sim \frac{U}{2}$ based on the moiré period ~ 5 nm and the size of the Wannier function ~ 2-3 nm [PRL 121, 026402 (2018)]. The observed MIT for the $f = 2/3$ state is consistent with a bandwidth tuned Wigner transition with $W \sim V$. This new result clearly shows a substantial increase in the bandwidth of the first Hubbard band at electric fields well below the critical field at $f = 1$. We expect the bandwidth to further increase with electric field and eventually lead to the Mott transition at $f = 1$ with $W \sim U$.

The reviewer pointed out the relevance of the second moiré band and its impact on the simple description of the single-band Hubbard model. This is relevant for TMD moiré materials in general (even in the absence of any applied electric field). Recent experiments have shown that the second Hubbard band has a substantially larger bandwidth than the first Hubbard band [Nature 587, 214 (2020) and arXiv:2102.10823]. This already suggests mixing of the moiré bands by U. Fortunately, theoretical studies [e.g. PRB 37, 3759(R) (1988) and PRB 102, 201115(R) (2020)] have suggested that the effect of the remote flat bands on the physics near $f = 1$ can be approximated by an effective single-band Hubbard model with renormalized parameters in the Hamiltonian. The situation is similar to that in cuprates [Rev. Mod. Phys. 78, 17 (2006)]. Our experiment is consistent with this view: Extended Data Fig. 1 shows that closing of the $f = 2$ gap at ~ 0.5 V/nm (~ 1.0 V/nm in the original manuscript) does not perturb the $f = 1$ insulating state, which suggests that the band merging at $f = 2$ does not change the filling factor of the Hubbard band. Otherwise, the MIT at $f = 1$ and $f = 2$ would occur at nearly the same electric field.

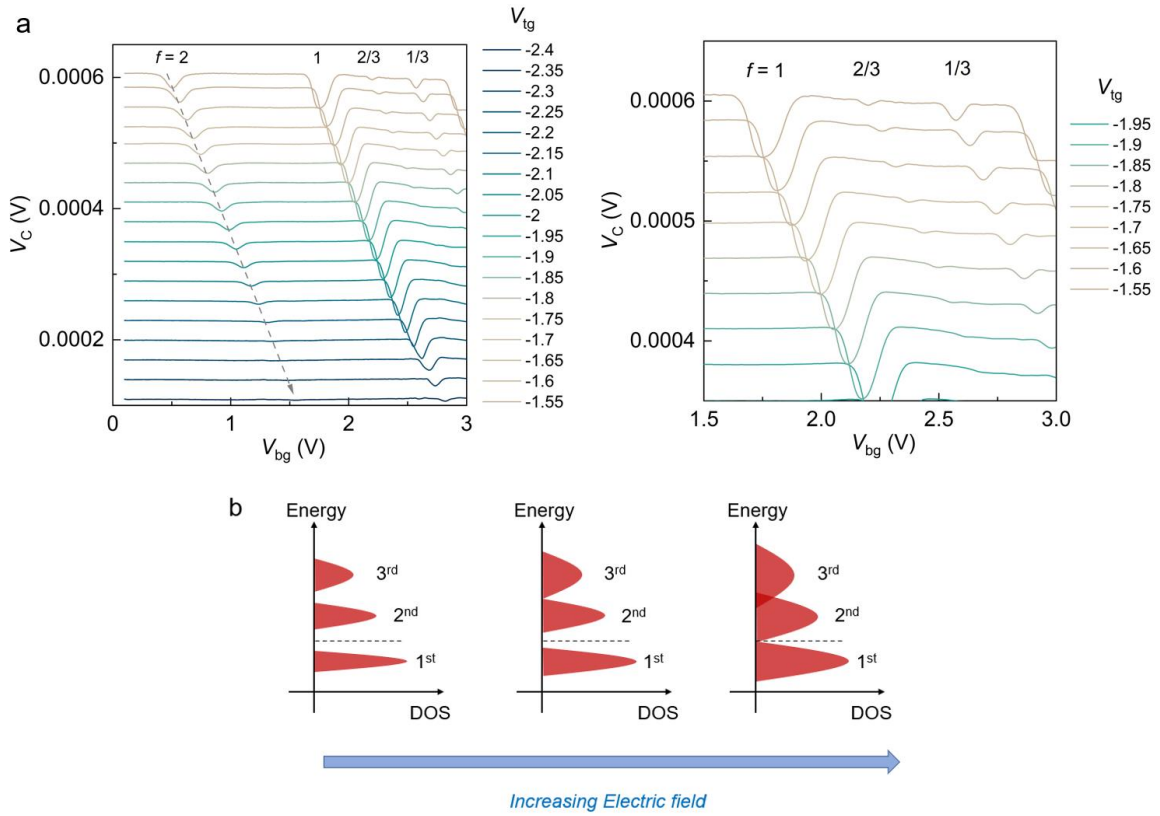

Figure R2. a, Differential top gate capacitance as a function of bottom gate voltage at varying top gate voltages (T = 5 K). The curves are vertically displaced for clarity. The filling factors of the incompressible states are labeled. The vertical electric field increases along the arrowed dashed lines. The right panel shows more details of the left panel, focusing on the f = 2/3 state. **b,** Evolution of the first three Hubbard bands at fixed filling factor of $f = 1$. Because of mixing of the remote moiré bands by the Coulomb interaction, the Hubbard bandwidth increases with the Hubbard band index. With increasing vertical electric field, the $f = 2$ gap closes first, followed by the closing of the $f = 1$ gap near the Mott transition. The closing of the $f = 2$ gap does not change the filling factor of the first Hubbard band.

We propose the schematic picture in Fig. R2b for the first three Hubbard bands at fixed filling factor of $f = 1$. Because of mixing of the remote moiré bands by the Coulomb interaction, the Hubbard bandwidth increases with the Hubbard band index *f*, consistent with recent studies [Nature 587, 214 (2020) and arXiv:2102.10823]. With increasing vertical electric field, the $f = 2$ gap closes first, followed by closing of the $f = 1$ gap near the Mott transition. The closing of the $f = 2$ gap does not change the filling factor of the first Hubbard band.

Changes:

In our revised manuscript, we have included Fig. R2 in the Supplementary Information (Fig. S2). We have also discussed the effect of the remote moiré bands in the revised Methods (page 9).

B. Analysis and interpretation of transport data:

The values of T_0 and activation gap are extracted to assign the quantum critical regime, which is important to reach the conclusion. I had several comments regarding how one can reliably calculate these parameters.

1. First, it is not clear how the exact values of T_0 are determined from the data. For example, it seems that one can multiply all the T_0 values by a constant and still get the same collapse shown in Fig. 3a. Besides, the choice of T_0 for the metallic and insulating side is also not clear to me. Are they chosen so that Fig. 3a appears symmetric? If this were the case, the T_0 lines in Fig. 3 might not have much significance.

Response 3

We thank the reviewer for raising this important question. In our original manuscript, we have scaled the values of $k_B T_0$'s so that they match the measured gap sizes on the insulating side of the phase diagram. We did not choose the values of T_0's independently on the metallic side. The same values of T_0 automatically collapse the resistance curves on the metallic side. The mirror symmetry around the critical point is not handpicked; it reflects the intrinsic property of the quantum criticality.

We totally agree with the reviewer that we can multiply T_0's by a constant and still get the same scaling exponent and the same collapse of resistance curves. The absolute value of T_0 is therefore not very meaningful. We have followed the reviewer's suggestion to remove the T_0 line in the phase diagram of Fig. 3b in the revised manuscript.

Changes:

In our revised manuscript, we have specified that $k_B T_0$ in Fig. 2b is scaled to match the charge gap size on the insulating side and the same values are used to collapse the data on the metallic side (page 4). We have also removed the T_0 lines in Fig. 3b.

2. I also suggest the authors add more intermediate steps for extracting T_0 as supporting information, such as showing a few representative curves. For example, from raw data in Fig. 2a, it is hard to directly see how the curves at E= 1.306 V/nm and E = 1.309 V/nm can be collapsed onto E = 1.342 V/nm.

Response 4

We have added more intermediate steps for the scaling collapse and the extraction of T_0 in Supplementary Fig. S3 as the reviewer suggested. For device 1 the first set of measurements was taken down to 1.6 K. It was reloaded into a He-3 insert to measure down to 300 mK in the second set of measurements. In the revised manuscript, we present results of scaling collapse from both rounds of measurements in Fig. 3. As we discuss below, the disorder density increased slightly in the second round, which is presumably related to reloading the device and multiple thermal cycles.

Below we include the scaling analysis for both rounds of measurements (separated into two columns in Fig. R3a-h). Figure R3a and R3b show the raw data of $R_\square$ versus T in a log-linear plot for varying electric fields. Figure R3c and R3d show the same data in a log-log plot. The critical field and the corresponding resistance $R_C(T)$ are identified from the power-law temperature dependence of resistance (dashed line). In Fig. R3e and R3f we normalize all the resistance curves by $R_C(T)$ and label them by $E - E_C$. We first rescale the curve furthest from the critical point on the insulating side ($E - E_C$ = -16 mV/nm) by setting $T_0 \approx 220$ K, which corresponds to the gap size determined from the thermal activation fit of $R_\square(T)$ at high temperatures. We determine the value of T_0 for the next normalized resistance curves to best collapse it onto the first curve in Fig. R3g and R3h. Figure R3i shows all the T_0's on the insulating side. We then use the same T_0's to scale the resistance curves on the metallic side with the same $|E - E_C|$ without any adjustment in Fig. R3g and R3h. The values of T_0's from both rounds of measurements show good consistency.

For data from the first round measurements (down to 1.6 K), good collapse of resistance curves can be achieved down to $\sim \pm 1$ mV/nm from the critical point. The collapse starts to fail within $\sim \pm 1$ mV/nm from the critical point presumably due to disorder effects (which broaden the transition as shown in Fig. 1f in the main text). It provides an estimate of the disorder density in the system by using the parallel plate capacitor model $\sim 2\varepsilon_{hBN}\varepsilon_0 \times (1\mathrm{mV/nm}) \sim 5 \times 10^{10}$ cm^{-2} ($\varepsilon_{hBN} \approx 3$ is the out-of-plane dielectric constant of hBN and ε_0 is the vacuum permittivity). The disorder density is about two orders of magnitude smaller than the moiré density. For data from the second round of measurements (down to 300 mK), good collapse of curves can only be achieved down to $\sim \pm 2$ mV/nm from the critical point, which corresponds to a higher disorder density $\sim 10^{11}$ cm^{-2}.

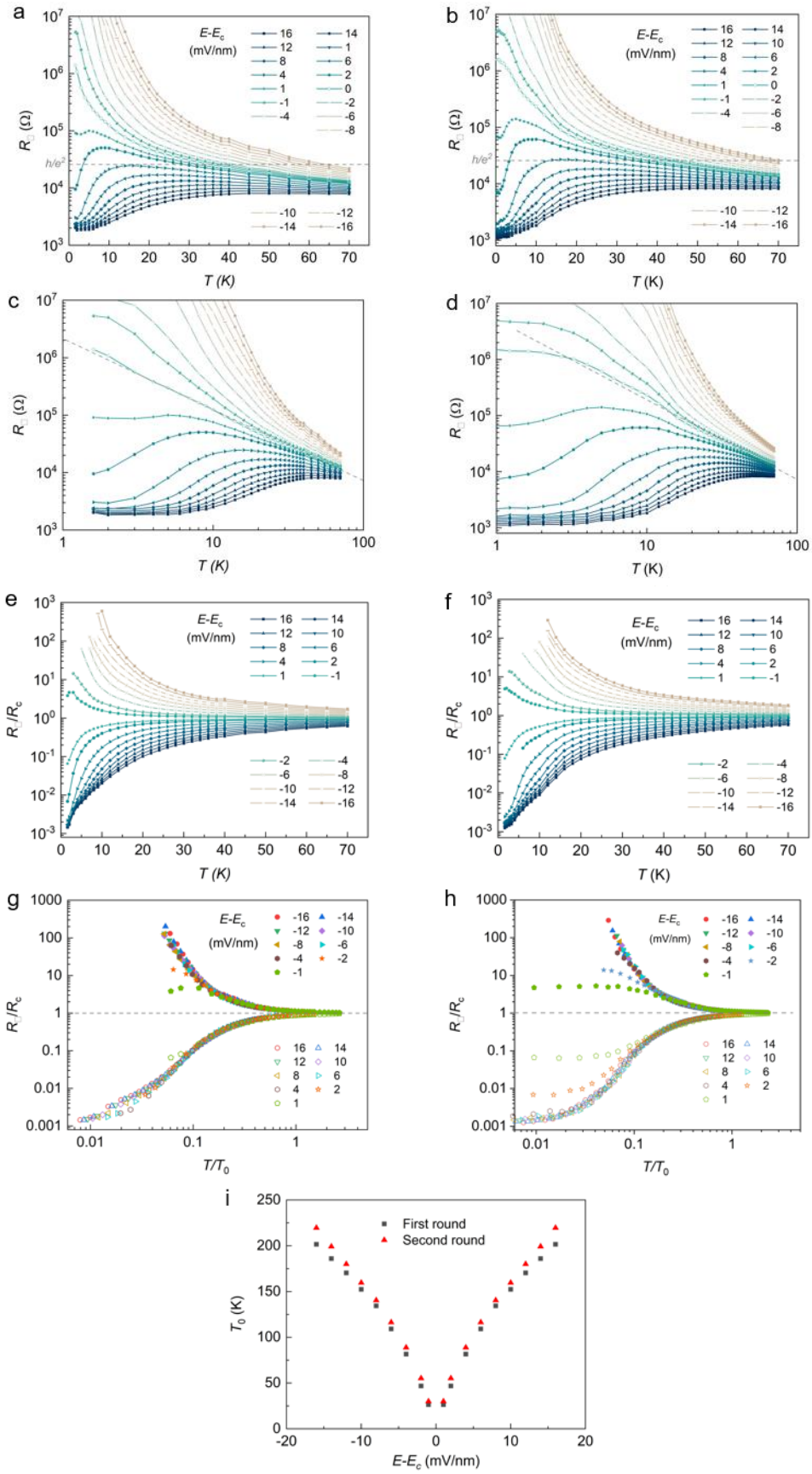

Figure R3. a, b, Temperature dependence of square resistance at f = 1 under varying electric fields from the first (down to 1.6 K, **a**) and second (down to 300 mK, **b**) round measurements. **c** and **d,** same as **a** and **b**, respectively, in log-log plot. The resistance $R_C(T)$ at $E_C \approx 0.652$ V/nm follows a power-law dependence. **e, f,** Temperature dependence of $\frac{R_\square}{R_C}$ under varying electric fields from the first (**e**) and second (**f**) round measurements. **g, h,** $\frac{R_\square}{R_C}$ versus the scaled temperature $\frac{T}{T_0}$ for selected electric fields for clarity. The full range of data is shown in Fig. 3 of the main text. **i,** Electric field dependence of T_0 near the critical point from the two rounds of measurements.

Changes:

In our revised manuscript, we have provided the above discussions and Fig. R3 in the Supplementary Information (Fig. S3).

3. The authors fit the high-temperature insulating R-T data in terms of thermal activation but attribute the same range of metallic data to electron scattering, although they qualitatively look quite similar in Fig. 2. More evidence is needed to justify this. For example, it would be useful to show that the thermal activation model fails to fit the data above E_c.

Response 5

We thank the reviewer for the suggestion. We agree that a more careful comparison between the two sides of the critical point needs to be shown. Figure R4a below is the temperature dependence of the square resistance in a log-log plot. It shows that the resistance at the critical point has a power-law dependence. Figure R4b is the Arrhenius plot of the same data on the insulating side. The resistance changes by more than two orders of magnitude at high temperature. We find a good activation behavior above ~ 10 K (solid lines). Figure R4c is the Arrhenius plot on the metallic side focusing on the high temperature region, where insulating-like behavior is observed. The resistance changes by a few times and there is no obvious thermal activation behavior. The temperature dependence is closer to a power-law dependence. This is better seen in Fig. R4a with a power-law dependence at the critical point, and stronger and weaker temperature dependences below and above the critical field, respectively.

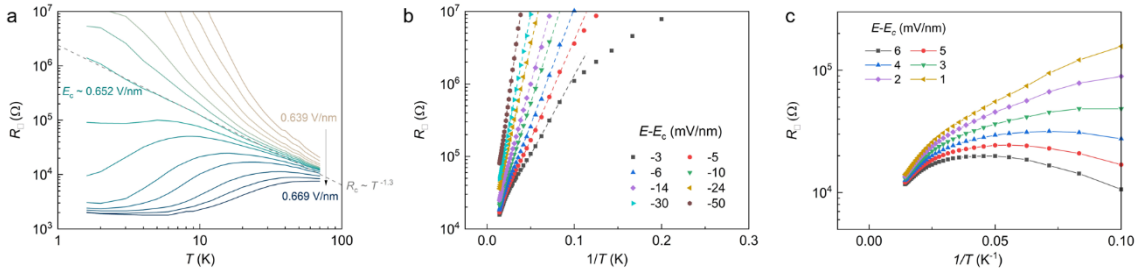

Figure R4. a, Temperature dependence of square resistance at $f = 1$ under varying electric fields in log-log plot. The resistance $R_C(T)$ at $E_C \approx 0.652$ V/nm follows a power-law dependence. **b,** Arrhenius plot of the data in **a** on the insulating side. Thermal activation behavior is seen at high temperatures from which the activation gap can be extracted from the slopes of the linear fits (dashed lines). **c,** Arrhenius plot of the data in **a** on the metallic side. No clear thermal activation behavior is observed in the high-temperature limit.

Changes:

We have specified that the high-temperature behavior of $R_\square$ on the metallic side cannot be described by thermal activation model in the revised manuscript (last paragraph on page 3), and referred to Extended Data Fig. 5.

4. I am also confused by the attribution of the Pomeranchuk effect in transport. In Fig. 1 and Extended Fig. 4, the metallic R-T is interpreted as Fermi liquid behavior, but similar data in Extended Data Fig. 7 is interpreted as the Pomeranchuk effect. Am I missing something? In fact, if the increase in R is due to the Pomeranchuck effect, would we expect the higher temperature range of the R-T curves to arise from the insulating states?

Response 6

We apologize for the confusion. To better understand the unusual temperature dependence of resistance on the metallic side at $f = 1$, we show an example in Fig. R5a for E - E_c = 3.5 mV/nm. The behavior is metallic at low temperature ($\frac{dR_\square}{dT} > 0$) and insulating-like at high temperature ($\frac{dR_\square}{dT} < 0$). The crossover occurs at T^* ~ 16 K, at which $\frac{dR_\square}{dT} = 0$. The Pomeranchuk effect here refers to the change from metallic to insulating-like conduction upon heating. It is analogous to the Pomeranchuk effect in He-3, in which the system changes from a Fermi liquid to a solid upon heating. Such behavior is also analogous to electrical transport in heavy fermion systems and is referred to as the generalized Pomeranchuk effect [e.g. PRB 82, 155102 (2010) and Nat. Mater. 17, 773 (2018)].

More details are provided in the contour plot of square resistance as a function of temperature and bottom gate voltage with a fixed top gate voltage (Fig. R5c). The bottom gate voltage mainly changes the filling factor. The electric field is fixed at 3.5 mV/nm near the $f = 1$ resistance peak. Fig. R5a is the temperature dependent $R_\square$ at $f = 1$ near the same electric field. The resistance peak is absent below ~ 7 K; here the resistance shows Fermi liquid behavior. Above ~ 7 K but below T^*~16 K, the resistance peak emerges; the $R_\square - T$ dependence deviates from Fermi liquid behavior. The emergence of the resistance peak and the deviation from the Fermi liquid behavior are correlated with the emergence of the Curie-Weiss behavior in the magnetic susceptibility (Fig. R5b), indicating the emergence of local moments. Above T^* the system displays insulating-like behavior. But this is not a proper insulator because there is no charge gap and thermal activation transport (see **Response 5**).

A recent theoretical study [arXiv:2102.12904] shows that the insulating-like behavior above T^* on the metallic side reflects a reduction in the doublon density. Such a reduction corresponds to the emergence of localized charges, and local moments or spins associated with localized charges. Charge transport above T^* is incoherent, in contrast to the coherent metallic transport below T^*. The unusual temperature dependence of resistance therefore reflects the release of extensive spin entropy upon heating or the larger (spin) entropy of the incoherent transport state compared to the metallic state, i.e. the Pomeranchuk effect.

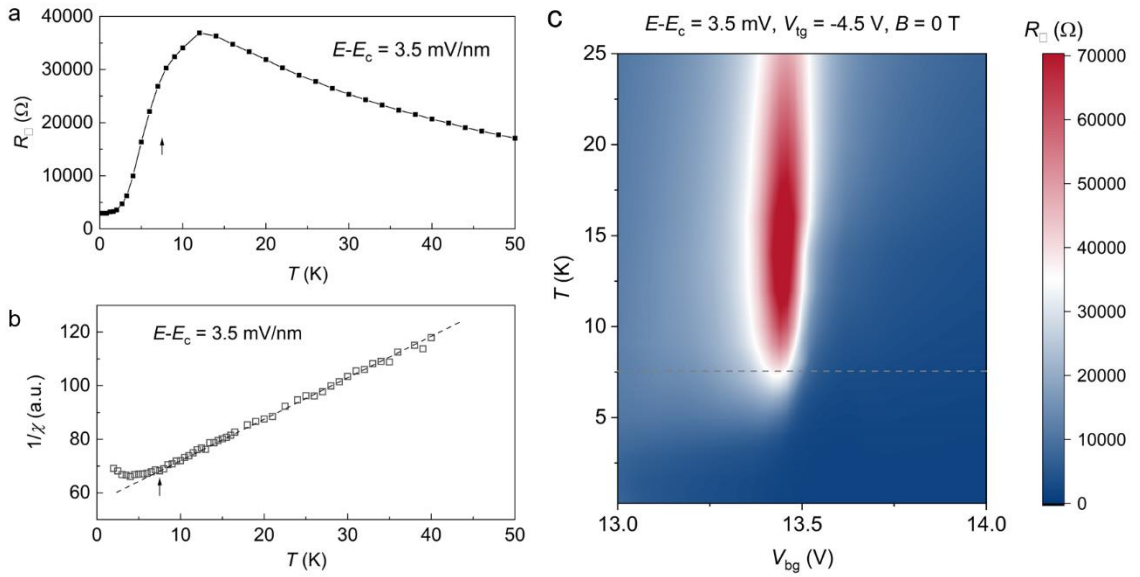

Figure R5. a, Temperature dependence of square resistance at $f = 1$ and near 3.5 mV/nm above the critical field. **b,** Temperature dependence of the inverse magnetic susceptibility under the same condition as **a**. The susceptibility saturates at the low temperatures; it follows the Curie-Weiss dependence (dashed lines) above the crossover from a Fermi liquid to an incoherent metal (denoted by arrow). **c,** Square resistance as a function of temperature and bottom gate voltage at a fixed top gate voltage. The bottom gate voltage mainly changes the filling factor. The electric field is fixed at 3.5 mV/nm near the $f = 1$ resistance peak. The horizontal dashed line marks the onset temperature of the $f = 1$ resistance peak (~ 7 K).

Changes:

In our revised manuscript, we have provided more detailed explanation of the Pomeranchuk effect (first paragraph on page 5). We have also improved Extended Data Fig. 7 and provided more detailed discussions to better connect to the Pomeranchuk effect.

5. The extended data Fig. 2 shows the transport properties of f=2 (band insulator and metal). It is quite surprising that the metallic phase does not show any Fermi liquid behavior. This could be concerning,

maybe signaling additional temperature-dependent extrinsic factors such as disorder and non-Ohmic contact, etc. Can the authors comment?

Response 7

The original Extended Data Fig. 2a includes measurements down to 1.6 K. We performed additional measurements down to 300 mK in the revised figure for $f = 2$. The low-temperature transport now shows Fermi liquid behavior on the metallic side (Fig. R6). In contrast to the MIT at $f = 1$, we do not observe strong effective mass divergence and the Pomeranchuk effect. It remains not understood why transport sets into the Fermi liquid behavior at substantially lower temperatures at $f = 2$ than $f = 1$. Future systematic studies are required to understand the mechanism. We note that with twice the doping density at $f = 2$ compared to $f = 1$, the electrical contacts are better and the disorder effects are less important. We have updated Extended Data Fig. 2a with the new data.

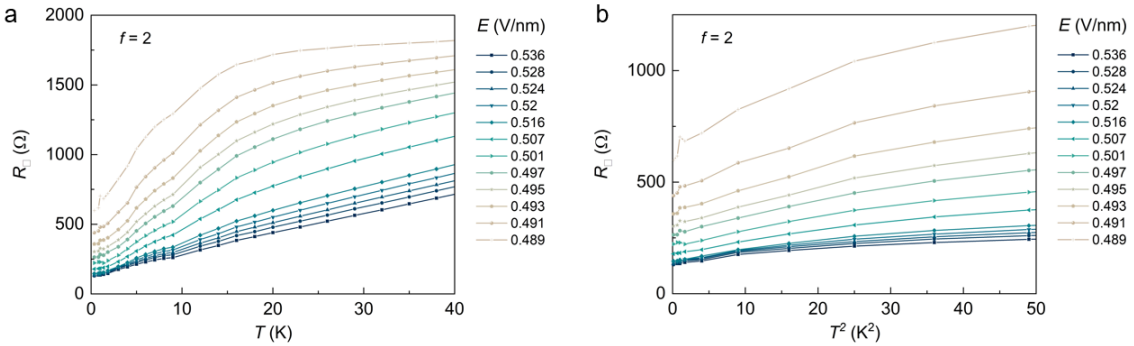

Figure R6. Temperature dependence of the $f = 2$ square resistance at varying electric fields on the metallic side down to 300 mK.

6. The transport behaviors reported here are remarkably similar to gate-induced MITs in conventional 2D electron systems. The authors highlight their difference on p. 5, stating that 'In contrast to … conventional 2D electron systems… here … electric field ... modifies … *U*/*W*'. However, I am not entirely convinced that there is a dramatic difference: in the conventional 2D systems, one cannot define a 'half-filling. One may think of doping in conventional 2D systems as changing the ratio of Coulomb to kinetic energy, i.e., U/W, which seems to suggest similar physics as here.

Response 8

We thank the reviewer for the comment. Indeed, the transport behaviors observed here and in 2D electron gases (2DEGs) are remarkably similar. Our work has been inspired by the early MIT studies in 2DEGs and benefited from the analysis developed there.

Because of the presence of lattices, the moiré system brings in new elements that make the MIT problem richer. For instance, the existence of lattices allows us to define half-band-filling and study the Mott transition. Geometrically frustrated lattices, such as the triangular lattices studied here, favor a pure Mott transition by suppressing long-range magnetic ordering. The moiré system could be a platform to realize the sought-after quantum spin liquid state. From the measurement standpoint, the introduction of lattices enables the observation of the MIT at much higher charge densities and higher temperature (energy) scales, which makes the study of the intrinsic properties of the system more easily accessible.

Changes:

We have modified the last paragraph of the main text (page 5 and 6) to highlight the remarkable similarity between MITs in two different systems and pointed out the new elements and new possibilities with the TMD moiré systems.

C. Magnetic characterizations:

1. The authors find no evidence for spin ordering down to 1.4 K in the insulating state and extract a Curie-Weiss temperature of 30-40 K from optical measurements in the metallic state. Based on the ratio of the two temperatures, the authors suggest possible spin frustration. I am not sure how valid it is to assume that the insulating states would have the same exchange interaction as the metallic states. Because the Wannier function changes as a function of E-field, it may not be surprising if the strength of exchange interaction is different in two states.

Response 9

We apologize for not explaining this part clearly in the original manuscript. We did not assume that the insulating state has a similar exchange interaction as the metallic state. We measured the values on the two sides to be similar using the same optical method (Fig. 4b, also shown in Fig. R7). The high-temperature magnetic susceptibility follows the Curie-Weiss behavior at all electric fields, indicating the emergence of local moments. We extract a Curie-Weiss temperature of ~ 40 K in the vicinity of the critical field on both sides ($E - E_C = \pm$ 3 mV/nm).

The energy scale of the exchange interaction on the insulating side from the Curie-Weiss analysis is consistent with the magnetic saturation field (B^*) that is measured at low temperature (Fig. 4a). The saturation field is determined to be $B^* \sim 4$ T at $E - E_C = -3$ mV/nm (B^* depends weakly on electric field on the insulating side). At temperatures well below the Curie-Weiss temperature, the magnetic saturation field is defined by the antiferromagnetic exchange energy of the local moments. We estimate the exchange energy $J \sim g\mu_B B^* \sim 3$ meV using the valence band g-factor of TMDs [$g \approx 11$, PRL 114, 037401 (2015)], which corresponds to ~ 30-40 K.

We cannot estimate J using the low-temperature B^* on the metallic side due to the emergence of Fermi surface and Pauli paramagnetism. The low-temperature B^* on the metallic side is governed by the renormalized bandwidth W^* of the Fermi liquid, as discussed in the main text.

The observed similar exchange energy scales on the metallic and insulating sides near the critical field are consistent with the dynamical mean-field theory (DMFT) results [Rev. Mod. Phys. 68, 13 (1996)], as well as the more recent numerical simulation results for the Mott transition in a triangular lattice [arXiv:2102.12904]. In the high-temperature limit, the exchange energy scale is determined by the super-exchange mechanism on both sides. Our result indicates that the size of the Wannier function does not change substantially across the MIT. Only the eigenstates of the system built on the Wannier functions are dramatically modified across the transition; the eigenstates change from localized states on the insulating side to extended coherent states built on the Wannier functions on the metallic side.

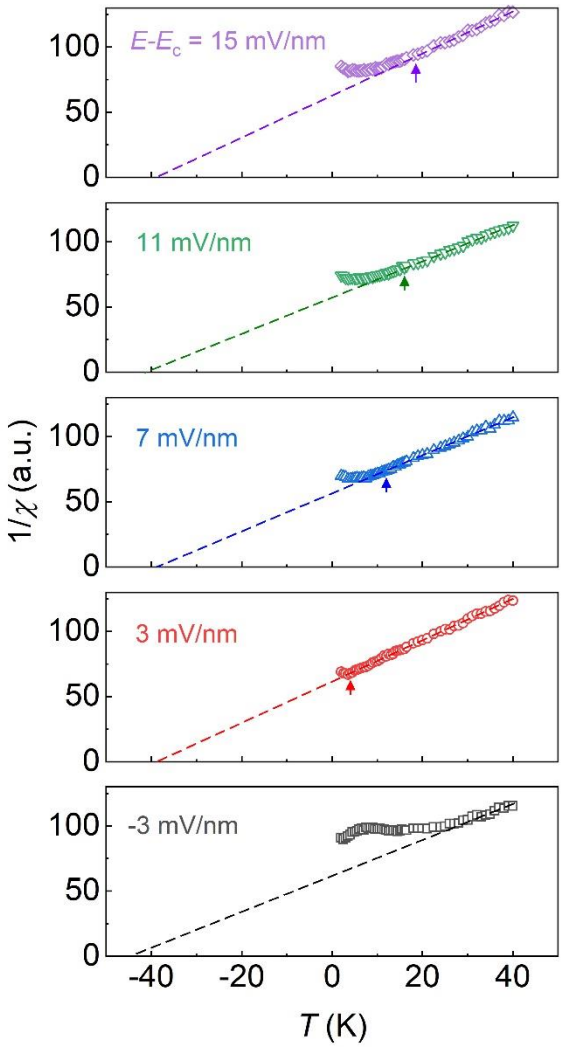

Figure R7. Temperature dependence of the inverse magnetic susceptibility χ^{-1} at different electric fields and Curie-Weiss fits (dashed lines) in the high-temperature limit. The obtained Curie-Weiss temperature is ~ 40 K and is weakly dependent on electric field.

Changes:

We have improved the discussion on the extraction of the exchange energy scale in the revised main text (first paragraph on page 5). We have also included Fig. R7 as Supplementary Fig. S4.

2. In addition, probing the magnetic ordering through the optical measurements seems to be a rather indirect probe since the interaction between spin and excitons can be quite complicated. For example, it would seem more natural to probe the magnetic ordering using the excitons in MoTe2 but not WSe2 because holes are in the MoTe2 layer.

Response 10

We agree with the reviewer that probing the magnetic ordering through the optical measurements is indirect. But unfortunately none of the conventional techniques can be readily applied to the 2D moiré system yet. There is a total of ~10,000 spins of interest; they are also imbedded in the device. Excitons in the TMDs provide the selectivity and sensitivity to probe these spins.

We also agree with the reviewer that excitons in $MoTe_2$ could be more sensitive to magnetic ordering of holes since they reside in the same layer. It is technically inconvenient to probe the fundamental excitons in $MoTe_2$ since they are in the near-infrared spectral region; our lab is currently not equipped with the appropriate light source and detector. However, it has been established that because of the atomic thickness of the TMDs, excitons in monolayer TMDs are generally sensitive to magnetization in adjacent layers, for instance, through exchange interaction. One example is monolayer WSe_2 on CrI_3, in which the excitons in WSe_2 can probe the magnetic states of 2D magnetic insulator CrI_3 [Nat. Nanotech. 15, 187 (2020)]. We have performed control experiments to verify the accuracy of the optical technique whenever possible in this study. More details are provided below in **Response 11**.

3. The MCD data also appears not straightforward to interpret. For example, it seems that one can get different magnetic susceptibility depending on the probe wavelength. Could the authors comment on this? It would also be great if the authors show a few representative I_L and I_R spectra at different E-field and explain how to eliminate the intrinsic response from WSe2.

Response 11

We agree with the reviewer that interpretation of the MCD data is not always straightforward. As we discuss below, one needs to first examine the exciton spectrum under left and right circularly polarized light excitation before choosing an appropriate probe wavelength. The MCD signal is linearly proportional to the sample magnetization; the proportionality constant (i.e. the sensitivity) is dependent

on the probe wavelength; it is usually the largest near an optical resonance. However, the choice of the wavelength does not affect the analysis as long as the spectral feature does not shift appreciably in wavelength as we tune the experimental parameters, such as the temperature, magnetic field or electric field. The results, such as the Curie-Weiss temperature and magnetic saturation field, are independent of the choice of the probe wavelength. Below we describe several control experiments to show more details on the MCD measurements and analysis.

Figure R8 compares the MCD measurement (1.6 K) of a $WSe_2/MoTe_2$ moiré superlattice at filling f = 1 and E - E_C = 3 mV/nm (Fig. R8a, b) and a charge neutral monolayer of WSe_2 (Fig. R8c,d). We show the fundamental exciton spectrum of WSe_2 in two systems under left (black) and right (red) circularly polarized light excitation in Fig. R8a,c, and the MCD spectrum in Fig. R8b,d. The exciton spectrum in the moiré superlattice red shifts and becomes more complex compared to that in the monolayer. (Due to multiple interference, the absolute reflection contrast depends on the exact distance of the sample from the silicon substrate, which acts as a mirror; it is not meaningful to compare the absolute value.) The excitons in the two systems also respond to the applied magnetic field with opposite sign. We track the MCD at the low-energy peak/dip (vertical dashed lines) as a function of magnetic field (Fig. R8e,f). While the magnetic response is linear with field in the monolayer, it increases linearly with field at small fields, saturates above ~ 2 T, and shows a weak linear decrease with field above saturation in the moiré superlattice. The observed linear magnetic response for monolayer TMDs, corresponding to the exciton valley Zeeman effect, is consistent with the literature [e.g. Nat. Phys. 11, 141 (2015)]. The magnetic saturation at a few Tesla is characteristic of local moments. The weak linear reduction in MCD with field above the saturation manifests the contribution of the bare exciton and can be subtracted as a background. The results are fully consistent with Nature 579, 353 (2020), which has also employed magneto-optical measurements to probe the local moments in similar TMD moiré systems.

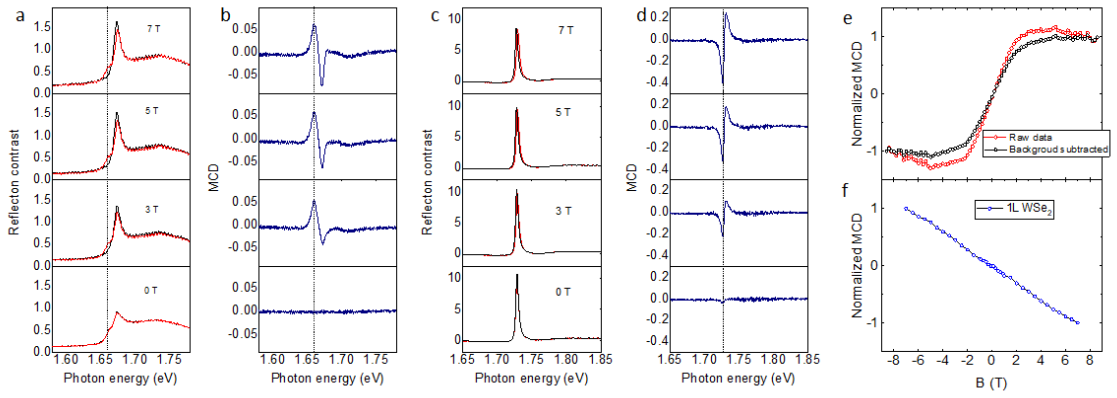

Figure R8. a,c, Reflection contrast spectrum of left- and right-handed polarized light for $MoTe_2/WSe_2$ moiré sample (**a**) and monolayer WSe_2 (**c**) under several representative magnetic fields. The spectrum focuses on the WSe_2 fundamental exciton resonance. The moiré sample is at $f = 1$ and under 3 mV/nm above the critical field. The WSe_2 sample is charge neutral. **b,d,** MCD spectrum obtained from **a** and **c** for $MoTe_2/WSe_2$ moiré sample (**b**) and monolayer WSe_2 (**d**). **e,f,** MCD response as a function of magnetic field for $MoTe_2/WSe_2$ moiré sample (red, **e**) and monolayer WSe_2 (**f**) at the low-energy

peak/dip of the MCD spectra. The black curve in **e** is obtained after subtracting a linear background from the raw data (red).

The low-temperature magnetic saturation field B^* obtained from the MCD measurement is independent of the probing wavelength. In Fig. R9 we compare B^* from the MCD measurement with the saturation field from the transport measurement on the metallic side. They are largely consistent. The result supports that the MCD probes the magnetic response of the moiré material. We note that here B^* is determined at 85% of the saturated MCD value. Because of the smooth magnetic field dependence for the MCD, it is generally harder to determine B^* from the MCD and the error bars are bigger.

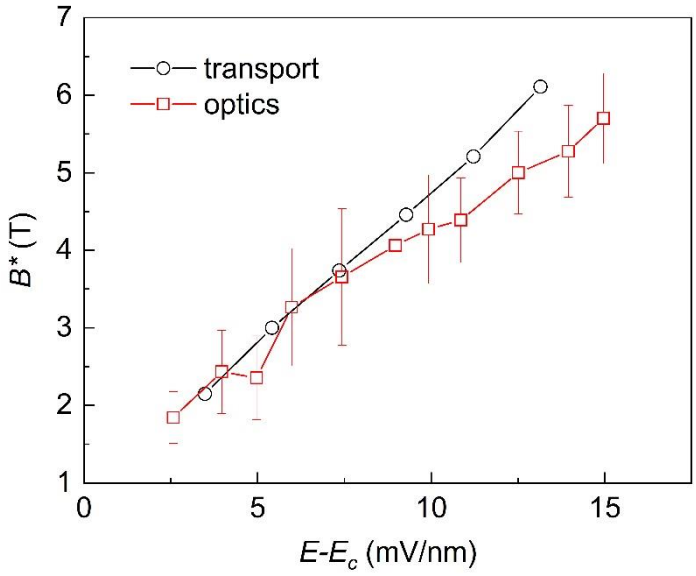

Figure R9. Electric-field dependence of the saturation magnetic fields B^* on the metallic side obtained from transport (black) and MCD (red) measurements. The two are in reasonable agreement.

We further show the MCD measurements at several representative electric fields near the critical field on both sides at a fixed magnetic field of 3 T. Figure R10a shows the left- (black) and right-handed (red) reflection contrast spectrum; Fig. R10b shows the MCD spectrum obtained from Fig. R10a. We choose to probe the system at the low-energy peak, which stays unchanged in wavelength across the MIT. With these additional results, we hope that the reviewer is convinced that the MCD measurement indeed probes the magnetic response of the local moments in the moiré material.

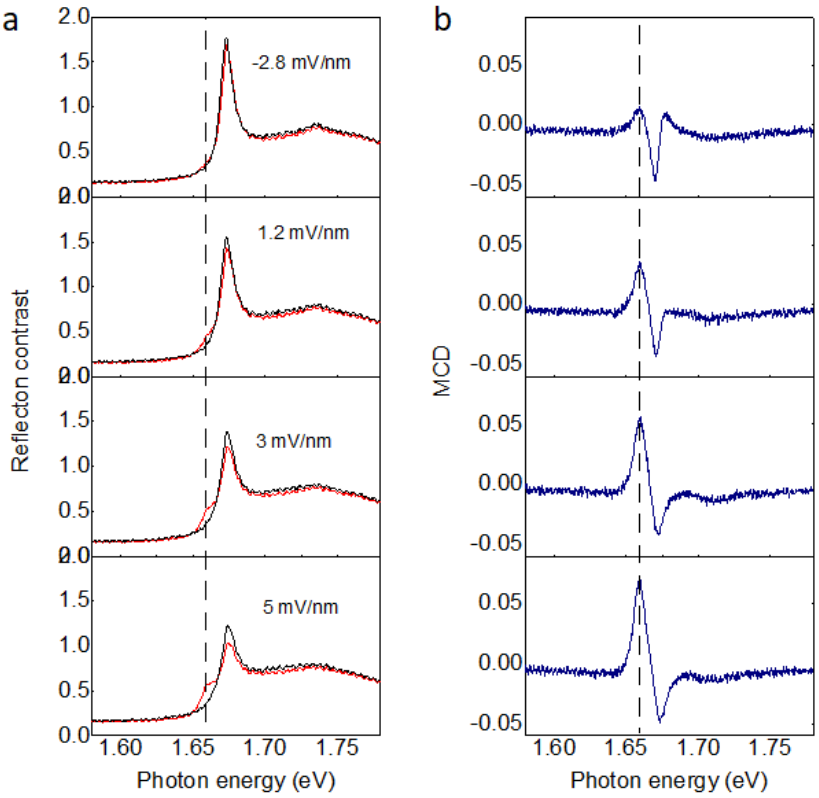

Figure R10. a, Reflection contrast spectrum at 3 T for the $MoTe_2/WSe_2$ moiré sample at $f = 1$ under several representative applied electric fields. Both left-handed (black) and right-handed (red) polarized spectra are shown. **b,** MCD spectrum obtained from **a**. A clear MCD enhancement near the WSe_2 neutral exciton resonance is observed. Although the amplitude changes, the resonance does not shift with electric field near the MIT critical point. It justifies the use of a single wavelength to probe the magnetic response near the critical electric field.

Changes:

We have included Fig. R8-R10 as supplementary figures (Fig. S5-S7) in our revised manuscript. We have also briefly mentioned the control experiments in Methods (page 7).

Some other minor comments:

Do the authors know the rotation angles between the two TMD layers being 0 or 60 degrees? It would be nice to show the second-harmonic generation data. I imagine the 0-degree and 60-degree would have different moiré potential and electrical dependence thereof.

Response 12

In this study we focus on 0-degree-aligned samples. We performed second-harmonic generation (SHG) on each sample. Figure R11 shows the polarization-resolved SHG from two samples. The SHG emissions from the two layers destructively interfere to produce much weaker SHG in the 60-degree-aligned samples. The results are fully consistent with the reports in the literature. We have included Fig. R11 as Supplementary Fig. S8.

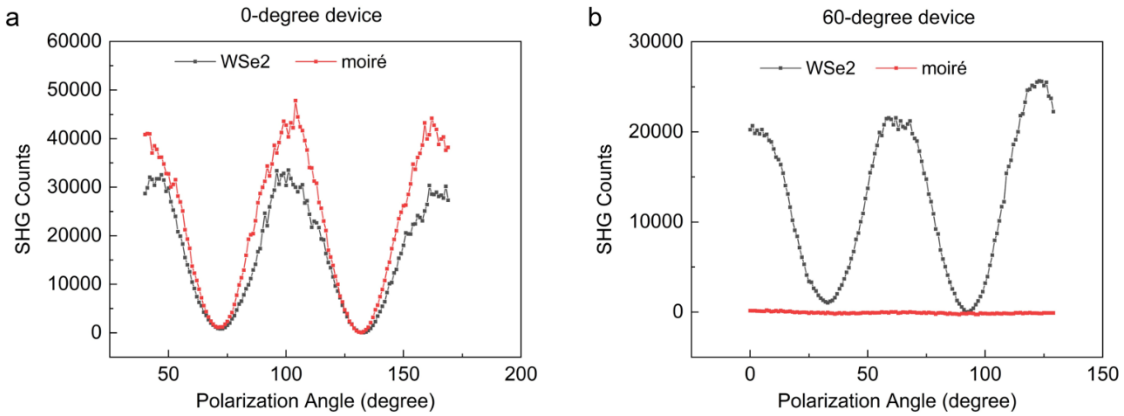

Figure R11. a,b, Second-harmonic intensity as a function of the polarization angle of the excitation light from 0-degree (**a**) and 60-degree (**b**) $MoTe_2/WSe_2$ heterobilayers. Samples are excited by linearly polarized light under normal incidence. The cross-polarized component of the second harmonic emission is detected. The second-harmonic emission from each monolayer is added constructively in the 0-degree sample (**a**) and destructively in the 60-degree sample (**b**).

Figure 2b: it would be useful to specify the meaning of different symbols in the caption (right now only labeled in the figure).

Response 13

We thank the reviewer for the suggestion and have addressed issue in the revised manuscript.

Extended Data Fig. 7: please specify what the vertical dashed line means (f=1?).

Response 14

The dashed line in Ext. Data Fig. 7 denotes constant filling factor $f = 1$. We have added the information in the revised caption.

nature portfolio

Referee #2

In this work, the authors study the metal-to-insulator transitions (MIT) in the MoTe2/WSe2 moire superlattice. In particular, they found that the MIT at the half-filling of the first hole moire band is consistent with universal critical theory of a continuous Mott transition from a Fermi liquid to a nonmagnetic Mott insulator in two dimensions. Some other works may have observed similar phenomena in other moire systems in the transport measurements, such as [Nature Physics 15, 237–241(2019)] and [Nature Materials 19, 861–866(2020)], but this work provides compelling evidences for the observed MIT at the half-filling as a Mott transition and provides important characterization of the electronics states across the transition. It demonstrates the advantage of using moire superlattice in simulated critical quantum phenomena and shows promises in further investigation of new exotic state of matter near the transition. Therefore, I think this work is of interest of many people in the field of two-dimensional materials and strongly-correlated physics and I recommend it to be published in Nature. Nevertheless, I have a few concerns of the manuscript as listed below:

1. They claimed that the electric field varies the potential different between the two TMD layers and subsequently the moire potential. I agree with the first half of the statement, but why varying the potential different between the two TMD layers necessarily affect the depth of the moire potential? In the situation that the moire potential is dominated by the moire structure and internal strain as claimed in recent studies (Shabani, et. al. Nature Physics. (arXiv:2008.07696) and Li, et. al. Nature Materials (arXiv:2007.06113)), this is certainly not the case. Their DFT calculations also show that for a out-of-plane electric field up to 2V/nm, the band width barely changes.

Response 15

We thank the reviewer for the positive assessment of our work and for the valuable suggestions for improvement. The TMD moiré system here has much larger energy scales compared to other moiré systems, which allow us to observe a resistance change of several orders of magnitude and perform quantitative analyses across the Mott transition.

We agree with the reviewer that there are multiple mechanisms that can induce moiré potentials. One of them is the periodic modulation in interlayer hopping. If this is the dominant mechanism, the moiré potential depth is dependent on the relative band offset between the two TMD layers since the interlayer hopping amplitude is sensitive to the relative band offset [PRB 99, 125424 (2019)]. As the reviewer pointed out that the moiré potential could also be induced by lattice reconstruction and periodic strain in the monolayer. If this were the dominant effect, tuning the relative band offset would have minimal effect on the moiré potential depth. This may well be the case in systems with large moiré periods, such as WSe_2/WS_2 (with moiré period ~ 8 nm) and small-angle twisted homobilayer TMDs (with moiré period comparable to or larger than 10 nm). Our system has a rather small moiré period ~ 5 nm. It is conceivable that lattice reconstruction costs more energy in such a small moiré period system and the moiré potential is no longer dominated by periodic strain modulation in the monolayer. Instead, interlayer hopping contributes substantially to the moiré potential depth, which is then tunable by a

vertical electric field. Our DFT calculations, which include lattice relaxations, show that interlayer hopping is important in determining the moiré bandwidth under high electric fields (more details are provided in **Response 16**). Future STM studies on our system are required to better understand the relative importance of the various contributions to the moiré potential.

We apologize for not explaining clearly the DFT result in the original manuscript. The x-axis in Fig. 1e is the displacement field rather than the applied electric field. One needs to know the out-of-plane dielectric constant of the TMD heterobilayer ε_{TMD} to convert them. We estimate $\varepsilon_{TMD} \sim 7-8$ from the interlayer dipole moment $\frac{\varepsilon_{hBN}}{\varepsilon_{TMD}} et \approx 0.26$ e*nm from optical measurements (see **Response 1**). Here $\varepsilon_{hBN} \approx 3$ is the known out-of-plane dielectric constant of hBN, e is the electron charge, and $t \approx 0.7$ nm is the interlayer separation between the two TMD layers. The DFT shows that onset of a large bandwidth change occurs near a displacement field of 2 V/nm, which corresponds to an applied field of ~ 0.3 V/nm. This is within the range of applied electric field ~ 0 – 0.65 V/nm in our experiment.

Because of convergence issues, we were not able to perform reliable DFT calculations beyond displacement field of 3 V/nm (i.e. applied electric field of ~ 0.45 V/nm). But the bandwidth is expected to increase substantially with further increase of the applied electric field. Note that in the original manuscript, we had defined the electric field as the sum of the fields above and below the moiré superlattice. We have redefined the electric field to be the average of the two electric fields. The value is half of the original value. The new definition is more appropriate to represent the applied electric field.

Changes:

We have included an additional discussion on mechanisms of bandwidth tuning in the revised Methods and cited the references that were brought up by the reviewer (page 7 and 8). We have also made it explicit that the x-axis in Fig. 1e is the displacement field, and made a connection between the displacement field and applied electric field in the revised Methods.

2. I am a bit puzzling by the critical behavior in the DFT calculations. Why the band gap between the first moire band and the second moire band suddenly collapses and the band width of the first moire band suddenly increase rapidly when the displacement field is above 2V/nm? The band structures in Figure 1d seems only have a small number of data points (maybe only ~5 kpoints along Gamma-M direction?). Is it enough to sample correctly the band width and the band gap? How to defined the bandwidth of the first moire band when the bandgap between the first and the second moire bands collapses? It might be beneficial to the put all the band structures near the critical point in the supplementary information and add a little more discussion.

Response 16

We thank the reviewer for the question. In our calculation the closing of the moiré gap is correlated with the increase in the bandwidth. It starts to happen when the offset between the upper valence band in $MoTe_2$ and WSe_2 becomes comparable to the interlayer hopping amplitude (which itself is dependent on the band offset). Layer hybridization becomes important which drives gap closing and bandwidth increase near the displacement field $D \sim 2$ V/nm. In principle, this should be a smooth process. We believe the rapid change near $D \sim 2$ V/nm reflects the less accurate DFT calculations under high displacement fields, where the convergence of self-consistent calculations becomes slow and results in the rapid change. The DFT calculations are therefore only able to capture the trend of the dependence that is in qualitative agreement with experiment. Also, since the DFT calculations were performed under zero doping and the moiré system is strongly correlated, we do not expect quantitative agreement between DFT and experiment at $f = 1$. In addition, the onset of layer hybridization is accompanied by a change in the out-of-plane dielectric constant of the heterobilayer. In that case, the displacement field (used in our calculations) and the applied electric field (used in our experiments) do not scale linearly with the onset of layer hybridization.

Although we only have 5 k-points along the $\Gamma - M$ direction and 15 k-points connecting $K - \Gamma - M - K$ in our calculations, the moiré bandwidth can still be determined because we only need energies at the high-symmetry points. The band structure is developed from the K/K' valley of monolayer $MoTe_2$ and WSe_2 so that the moiré band maximum and minimum are expected to be at the Γ and K points of the mini-Brillouin zone, respectively, similar to other TMD moiré systems [PRL 121, 026402 (2018)]. As a result, the first moiré bandwidth can be estimated from the energy difference between the Γ point (maximum) and the K point (minimum). The moiré band gap can also be estimated in a similar manner. When the moiré band gap collapses, there remains only a single touching point for the two moiré bands. The first moiré band remains well defined and its bandwidth can be estimated using the same method.

Changes:

We have included the above discussion in the revised Methods (page 7 and 8). We have also added the calculation results at several other displacement fields near gap closing in the Supplementary Information (Fig. S9).

3. Moreover, when the displacement field is above 2V/nm, the original ~220 meV band offset between MoTe2 and WSe2 should be significantly reduced. What are the characters of the electronic states in the first and the second moire bands? Are they all contributed from the MoTe2 states under zero displacement field and large displacement field, or one of them is contributed from the WSe2 states?

Response 17

We thank the reviewer for the question. We have performed independent optical studies to determine the band offset at filling $f = 1$ as a function of applied electric field (**Response 1**). At zero applied

electric field, the band offset between the upper valence bands of the two layers is determined to be ~ 300 meV, which is slightly larger than the value from our DFT calculations. The value is reduced to ≳ 100 meV at the critical electric field for the Mott transition. This result indicates that both moiré minibands are mainly from the $MoTe_2$ state for the entire range of electric fields investigated in this experiment. (Layer hybridization (**Response 16**) does introduce some WSe_2 characters to the energy eigenstates.)

Changes:

We have included additional discussions on the moiré band structures in the revised Methods (page 7 and 8) and the determination of band offsets by optical spectroscopy measurements in the Supplementary Information (Fig. S1).

4. The authors argue that the tunning of U/W by the electric field drive the Mott transition. But the Mott transition seems to happen after the band gap between the first and the second moire bands collapses, it may not be appropriate to neglect the effect of the second moire bands in the discussion.

Response 18

We agree with the reviewer that the effect of the second moiré band cannot be ignored. In **Response 2**, we have addressed the effect of the second moiré band in details. In short, the effect of the second moiré band is not negligible even in the absence of applied electric field. But the effect on the physics near $f = 1$ can be approximated by an effective single-band Hubbard model with renormalized parameters in the Hamiltonian. Therefore to the first order approximation the MIT near $f = 1$ observed in this study is still close to the Mott transition. We have included discussions on the effects of the second moiré band on the Mott transition in the revised Methods (page 9) and additional data on the tuning of the bandwidth of the first Hubbard band in the Supplementary Information (Fig. S2).

5. The scales of the x-axis for the upper and the lower panels of figure 1e seems to be different. Is so, they shouldn't share the same labeling.

Response 19

The upper and lower panels of Fig. 1e share the same x-axis scale. To improve clarity, we have used independent x-axis for the two panels in the revised manuscript.

6. Near at the end of the second paragraph on page 2, it says "…., which changes the size of the localized Wannier function and the bandwidth predominantly." What supports such statement? Have the authors calculated the Wannier function in the system?

Response 20

We apologize for being not clear about this point in the original manuscript. The Wannier function of the system is calculated in the DFT. But it is easier to see the scaling dependence from a simple model of harmonic moiré trapping potentials with amplitude V_M. Such a simple model has been shown a good approximation for TMD moiré systems [PRB 102, 201115(R) (2020) and arXiv:2011.13558]. The harmonic potential supports an s-orbital with a Gaussian wavefunction of size $\xi = \sqrt{a_M}\left(\frac{\pi^4 m V_M}{\hbar^2}\right)^{-1/4} \propto V_M^{-1/4}$. Here a_M ($\approx$ 5 nm) is the moiré period, m ($\approx 0.5m_0$) is the valence band mass of monolayer $MoTe_2$ (m_0 is the free electron mass) and $\hbar$ denotes the Planck's constant. The on-site Coulomb repulsion is given by $U \approx \frac{e^2}{4\pi\varepsilon\xi} \propto \xi^{-1}$, where e is the electron charge, $\varepsilon \approx 4.5\varepsilon_0$ is the effective permittivity of the background (hBN) and ε_0 denotes the vacuum permittivity. The on-site Coulomb repulsion increases very slowly with V_M. On the other hand, the bandwidth W is proportional to the inter-moiré-site hopping and is defined by the overlap integral of the neighboring Wannier functions. The bandwidth scales exponentially with ξ, a much stronger function on the potential amplitude V_M compared to U. We therefore conclude that changing moiré potential depth predominantly changes the bandwidth.

Changes:

We have included the discussion above in the revised Methods (page 9).

7. In the third paragraph on page 2, it says "The valley degeneracy is lifted by valley pseudospin-orbit coupling in small period moiré structures." Could the authors provide more explanation or relevant references?

Response 21

We thank the reviewer for the question. The literature so far has mainly concerned moiré materials with large moiré periods (i.e., in the limit of infinite moiré period). In such a limit, the K and K' valleys are practically decoupled. The moiré bands are therefore doubly degenerate throughout the entire mini-Brillouin zone.

The moiré period of $MoTe_2/WSe_2$ (~ 5 nm) is relatively small as a result of the large lattice mismatch between these two materials. The small moiré period introduces a weak mixing of the K and K' valleys

and perturbatively splits the valley (or pseudospin) degeneracy in the moiré band structure. Such splitting is similar to splitting of the spin degeneracy in real materials induced by atomic scale antisymmetric spin-orbit coupling. We can therefore think of valley mixing as effective pseudospin-orbit coupling on the moiré length scale. We have included the discussion above in the revised Methods (page 7 and 8).

8. At the end of the page 3, it says "The square resistance exceeds the Mott-Ioffe-Regel limit …". It may be better to provide a few words and link it to the dash line in Fig. 2a to make it clear for the readers.

Response 22

We thank the reviewer for the suggestion and have addressed the issue in the revised manuscript on page 3.

In Extended Data Figure 4, for the displacement field very close to the critical field (e.g., 0.0019 and 0.0026), the square resistant rises up at very low temperature and deviates from the T square dependence. Is it related to the quantum fluctuations near the MIT? Different from the other figures, the values of the displacement field are listed up to the fourth digit after the decimal point. What is the actual accuracy of the value of the displacement field?

Response 23

We believe the data points very close to the critical point do not reflect the intrinsic behavior of the system. The MIT is broadened by disorders (for example, Fig. 1f for 300 mK). This translates to a spread in electric field δE for the critical point, within which disorder effects could be important. We can estimate the broadening in electric field δE from disorder density δn (typically $\sim 5 \times 10^{10} - 10^{11}$ cm^{-2} in our devices) using the parallel plate capacitor model $\delta n \sim 2\varepsilon_{hBN}\varepsilon_0\delta E$. Here $\varepsilon_{hBN} \approx 3$ and ε_0 are the out-of-plane dielectric constant of hBN and the vacuum permittivity, respectively. We obtain $\delta E \sim \pm 1 - 2$ mV/nm, which is close to the range of electric fields within which unexpected behaviors occur.

We also thank the reviewer for the comment on the accuracy of the electric field, which is about ± 0.2 mV/nm. We have modified the key in Extended Data Fig. 4 accordingly.

Changes:

We have included a discussion on the effect of disorders in the revised Methods (page 9 and 10) and in Extended Data Fig. 4.

Referee #3

In this manuscript, Li et al. demonstrated a continuous Mott transition at f=1 filling in MoTe2/WSe2 moiré superlattices. A quantum criticality is extracted through the scaling behavior of the resistances. Here are several comments below:

1.The evidence of continuous Mott transition is only based on an Arrhenius analysis of the resistivity. In Extended Data Figure 3, the authors extract the activation gaps above 10K. However, in Figure 2, even below the critical electric field, above 10K, there still exist insulating behaviors. In other words, a so-called Arrhenius behavior could be also observed. In this case, can the activation gaps extracted through Arrhenius analysis could be used as evidence to prove the (Mott) insulators?

Response 24

We thank the reviewer for his/her comments, which have helped to improve our manuscript.

We agree with the reviewer that a more careful comparison between the two sides of the critical point needs to be shown. This is done in Fig. R4 of this reply. The conclusion is that the insulating-like behavior on the metallic side at high temperatures follows a power-law temperature scaling and cannot be described by the thermal activation model. More detailed discussions can be found in **Response 5**. A recent theoretical study [arXiv:2102.12904] shows that the insulating-like behavior above T^* on the metallic side reflects a reduction in the doublon density upon heating. Such a reduction in double occupancy on moiré sites corresponds to the emergence of localized charges and local moments associated with localized carriers upon heating. Charge transport above T^* thus becomes incoherent (in contrast to the metallic transport below T^*). Such a behavior is analogous to the electrical transport in heavy fermion systems and is regarded as the generalized Pomeranchuk effect in the literature [e.g. PRB 82, 155102 (2010) and Nat. Mater. 17, 773 (2018)]. Please refer to **Response 6** for more details on the Pomeranchuk effect.

In addition to the continuously vanishing charge gap approaching the critical field, several other observations in our experiment also support the presence of a continuous Mott transition. These include 1) a continuously vanishing charge gap from a new capacitance measurement (see **Response 25** below); 2) a diverging quasiparticle effective mass; 3) scaling collapse of resistance curves; and 4) the absence of electric field hysteresis down to $k_BT \sim 30$ μeV, which is much smaller than the characteristic energy scales of the system (e.g. $U \sim W \sim 50$ meV and exchange interaction energy $J \sim 3$ meV). The observed Pomeranchuk effect is also consistent with the most recent numerical simulation of the Mott transition in a triangular lattice [arXiv:2102.12904]. These observations are now summarized in the conclusion paragraph.

Changes:

In our revised manuscript (last paragraph on page 3), we have pointed out the power-law temperature dependence of $R_{\square}$ at high temperatures on the metallic side by referring to Extended Data Fig. 5. We have also provided more detailed explanations of the Pomeranchuk effect (first paragraph on page 5 in Methods and Extended Data Fig. 7).

2.Aside from the temperature dependence of resistivity (Rxx) at f=1, do the author have other evidence, such as Hall density, SdH oscillations to support the reconstruction of Fermi surface, or spectral gaps in compressibility? There could be multiple reasons causing resistance peak, such as charge density waves, van Hove singularities, etc. How could the author exclude the other possible reasons and justify the insulator?

Response 25

Following the reviewer's suggestion, we have performed additional compressibility and Hall measurements to support the presence of a Mott transition.

We performed the Hall measurement on the metallic side. We did not observe any quantum oscillations presumably because the quasiparticle effective mass is very large near the critical point (Fig. 2c) and the highest magnetic field that can be applied is limited to B^* to avoid magnetic field-induced MIT (Fig. 4c). The Hall coefficient as a function of electric field at $f = 1$ is shown in Fig. R12. It is obtained by measuring the Hall resistance at ± 0.2 T, followed by anti-symmetrizing the positive and negative magnetic field data. The data very close to the critical point is unreliable because a very small magnetic field can already induce MIT. The measured Hall coefficient is independent of magnetic field as long as the field is well below B^*.

Figure R12 shows that the Hall coefficient (black line) is nearly a constant on the metallic side. The Hall density (blue line) is largely consistent with the estimated moiré density $\approx 4.8 \times 10^{12}$ cm^{-2} (horizontal dashed line). The small deviation could arise from the anomalous Hall contribution from the Berry curvature of the moiré flatbands that has been ignored in the calculation of the Hall density from R_{xy}. The constant Hall coefficient on the metallic side suggests the existence of a Fermi surface of a constant size all the way to the critical point. The result is consistent with the observed Fermi liquid behavior very close to the critical point.

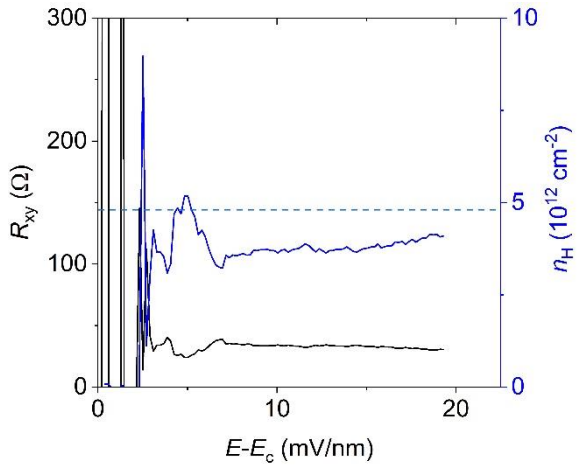

Figure R12. Electric-field dependence of the Hall coefficient (left axis) and Hall density (right axis) near the $f = 1$ MIT. The results are obtained from measurements under ± 0.2 T, followed by anti-symmetrizing the response. The Hall coefficient is largely independent of the electric field. The Hall density is consistent with the expected moiré density $\approx 4.8 \times 10^{12}$ cm^{-2} (horizontal dashed line). The measurement becomes unreliable very close to the critical point because of the magnetic field-induced MIT.

We also performed differential capacitance measurements as a function of filling factor and electric field on the insulating side. Figure R13 shows the differential top gate capacitance as a function of bottom gate voltage at several fixed top gate voltages. In these measurements the gate voltage range is limited because the capacitive signal becomes noisy even with a small leakage current. The top gate capacitance measures the electronic compressibility of the moiré structure [e.g. arXiv:2102.10823]. Each capacitance dip corresponds to an incompressible state. We observe incompressible states at $f = 1$ and $f = 2$ as well as at $f = 1/3$ and $f = 2/3$. The fractional fillings states are generalized Wigner crystal states [e.g. Nature 579, 359 (2020) and Nature 587, 214 (2020)]. With increasing electric field, the $f = 1$ state becomes weaker and vanishes continuously towards the MIT critical point. Overall, the capacitance results are fully consistent with the resistance measurements. They clearly show the presence of a spectral gap at $f = 1$ on the insulating sides that closes continuously approaching the critical point.

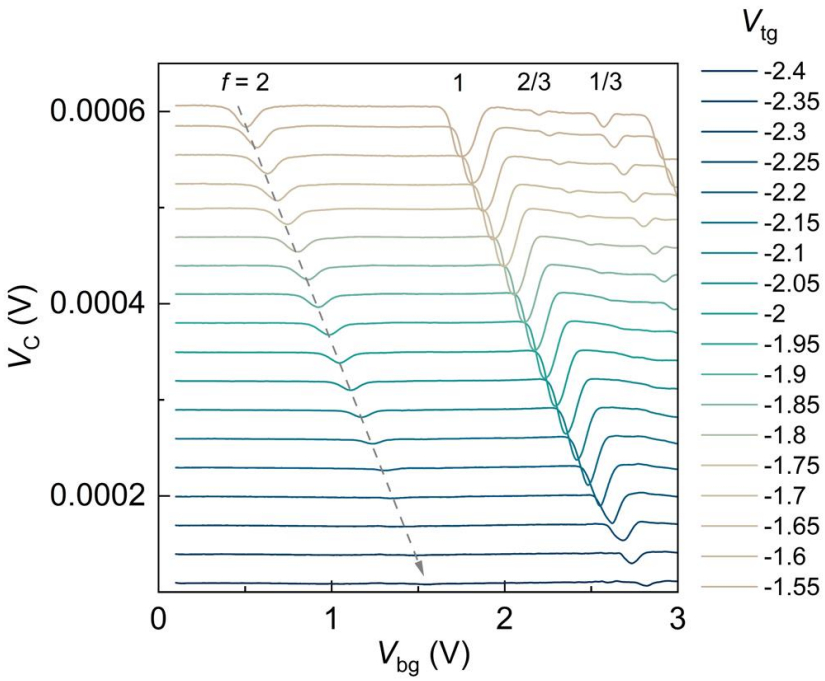

Figure R13. Differential top gate capacitance as a function of bottom gate voltage at varying top gate voltages. The curves are vertically displaced for clarity. The filling factors of the incompressible states are labeled. The arrowed line shows the electric field axis.

The $f = 1$ insulating state is also accompanied by the emergence of local moments, as revealed by the high-temperature Curie-Weiss behavior [Nature 579, 353 (2020) and Fig. 4 in this study]. Meanwhile, there is no evidence of long-range magnetic ordering. These behaviors are consistent with a Mott insulating state in a triangular lattice Hubbard model. In addition, the van Hove singularity is located at $f = 1.5$ for a triangular lattice [PRL 121, 026402 (2018)]. The insulating state at $f = 1$ is therefore unlikely driven by a diverging density of states. There is also no obvious Fermi surface nesting at $f = 1$ to stabilize a charge-density-wave. All these results point to a Mott insulating state at $f = 1$.

Changes:

We have included the Hall data as an inset in Fig. 2c. We have also included the capacitance data in the Supplementary Information (Fig. S2) to further support the existence of a Mott insulating state at $f = 1$.

3.The resistance value is the key to the overall analysis of the manuscript. Could the author show the I-V curves of the sample around f=1? If I-V curves are non-linear, the author should clarify the origin.

Response 26

The I-V curves of our samples around f = 1 are linear for bias voltages $\lesssim 4-5$ mV, except in the region, where the sample is very insulating. Figure R14a shows the differential resistance $\frac{dV}{dI}$ versus bias voltage V at 350 mK under several electric fields on the metallic side. It stays flat for bias voltage up to 4-5 mV for all the applied electric fields. Figure R14b shows the resistance across the MIT measured with three different bias voltages, 1 mV, 2 mV, and 5 mV. The I-V characteristic becomes nonlinear above ~ 2 mV bias when the sample is very insulating; the measured resistance drops at large bias values. This presumably is caused by sample heating at high biases.

In our measurements, we have kept the bias voltage at 1 mV to be safely in the linear regime for all electric fields and temperatures. Furthermore, the activation gap on the insulating side is extracted from the high-temperature data; the sample resistance is much smaller and the I-V curves are linear over a much wider range of bias voltages.

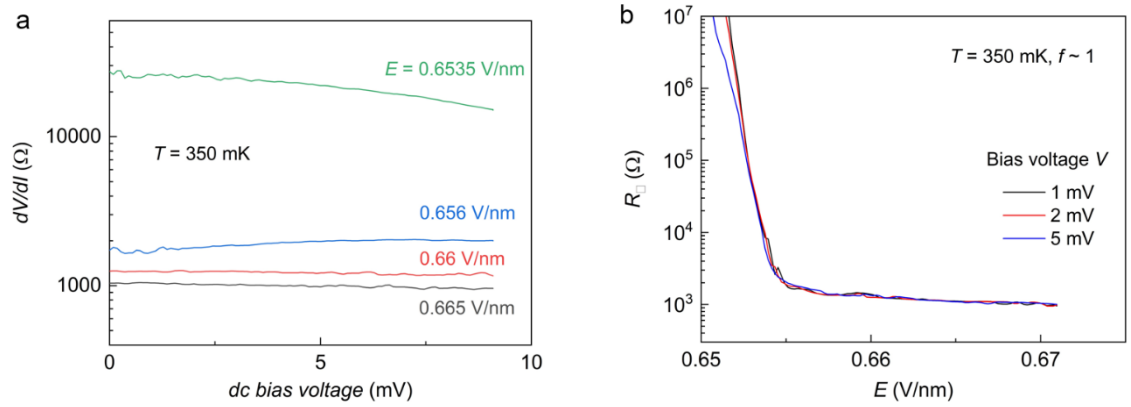

Figure R14. a, Bias voltage (V) dependence of the $f = 1$ differential resistance, $\frac{dV}{dI}$, at 350 mK under varying vertical electric fields on the metallic side. The ac modulation voltage is 0.5 mV. The differential resistance is independent of V for $V \lesssim 4-5$ mV. **b,** Electric-field dependence of square resistance measured at three different bias voltages at 350 mK. The resistance drops slightly on the insulating side at 5 mV bias likely due to heating at high biases.

Changes:

We have included Fig. R14 in the Supplementary Information (Fig. S10) and explicitly stated in the Methods (page 6 and 7) that our measurements have been kept in the linear I-V region.

4.In heavy fermion systems, the non-Fermi liquid behavior is usually due to the presence of a nearby quantum critical point. Here when the electric field (E) approaches critical field (Ec), the temperature dependence of resistivity is still described by the Landau Fermi liquid theory. On one hand, the author gives a trend of diverging effective mass based on the fitting of Fermi liquid. On the other hand, the

resulted diverging effective mass is contradicted with the Fermi liquid description for the normal metal. How to reconcile the contradiction? In Figure 2c, when E is getting very close to Ec (E-Ec ~0.003V/nm), the A^1/2 deviates from the power-law fitting. Could the author justify these discussions?

Response 27

On the metallic side, the measured resistance of our samples is well described by the Fermi-liquid form, $R_{\square} = R_0 + AT^2$, at low temperatures (Extended Data Fig. 4). The prefactor A in the Fermi-liquid description depends on the correlation length, quantified by the effective mass as $A \propto (m^*)^2$ according to the Kadowaki-Woods scaling. In Fig. 2c we show that $A^{1/2}$ from our measurement follows a power-law divergence with $E - E_C$. Similar mass divergence has also been observed in heavy fermion systems [e.g. Fig. 2b and 2c in Nat. Rev. Phys. 3, 9 (2021)]. The diverging effective mass at the critical point does not contradict the Landau Fermi liquid description of the metallic state near the critical point. A diverging effective mass upon approaching the critical point is consistent with the universal critical theory of a continuous MIT in 2D triangular lattices [e.g. PRB 78, 045109 (2008)]. These theoretical studies have predicted a low-temperature Landau Fermi liquid behavior all the way to the critical point on the metallic side. The effective mass, characterizing the charge carriers near the Fermi surface, diverges at the critical point because the charge carriers have the tendency to become localized due to Coulomb repulsions. The Fermi liquid description breaks down only at the critical point.

We do not observe marginal Fermi liquid behavior ($R_{\square} \propto T$) near the quantum critical region in our experiment. The square resistance displays insulating-like behavior ($\frac{dR_{\square}}{dT} < 0$) in our samples at the quantum critical point (Fig. 2a and Extended Data Fig. 5). This is correlated with the Pomeranchuk effect that reflects the large spin entropy of the non-magnetic Mott insulator (see **Response 6** for details). The observed Pomeranchuk effect is consistent with the numerical simulations in arXiv:2102.12904. The behavior here is different from the marginal Fermi liquid behavior in heavy fermion systems near quantum criticality. This is not unexpected because the physics of Mott transition in a triangular lattice is distinct from the Kondo destruction quantum criticality in heavy fermion systems.

There are two possible explanations regarding the deviation of the effective mass from the power-law divergence very close to the critical point. First, disorder effects become relevant very close to the critical point. The MIT tuned by the applied electric field is broadened by disorders. We can estimate the broadening in field δE from the disorder density δn using the parallel plate capacitor model, $\delta n \sim 2\varepsilon_{hBN}\varepsilon_0\delta E$, where $\varepsilon_{hBN} \approx 3$ and ε_0 are the out-of-plane dielectric constant of hBN and the vacuum permittivity, respectively. Using a typical disorder density $\sim 5 \times 10^{10} - 10^{11}$ cm^{-2}, we estimate $\delta E \sim \pm 1 - 2$ mV/nm. This is close to the range of electric fields, in which the effective mass deviates from the power-law dependence. Second, a non-power-law divergence of the effective mass has also been predicted in theoretical work that considers a quantum phase transition from a non-magnetic Mott insulator with spinon Fermi surface to a Landau Fermi liquid [PRB 78, 045109 (2008)]. The quantum fluctuations near the critical point give rise to a logarithmic divergence of the effective mass. The presence of sample disorders in our samples prevents us from experimentally distinguishing these two

scenarios. Future studies with higher quality devices are required to fully understand the region immediately close to the critical point.

Changes:

We have pointed out $\frac{dR_{\square}}{dT} < 0$ in the quantum critical region and its correlation with the Pomeranchuk effect (i.e. no marginal Fermi liquid behavior) in the third paragraph on page 4. We have also referred to the possible disorder origin for the observed deviation from the power-law divergence in Fig. 2c (page 9 and 10 in Methods).

5.In Figure 1c, the width of the resistance peak at f=1 is much wider than that at f=2. Could the author clarify the origin of this difference?

Response 28

The width of the resistance peak of $f = 1$ and $f = 2$ states are comparable. The $f = 1$ state appears much wider in Fig. 1c because the f=1 state has much larger resistance; resistance above 10 MΩ cannot be measured reliably and the log plot caps the value at 10 MΩ (see **Response 29**). The capacitance data in Fig. R13, which does not suffer from orders of magnitude change, clearly shows that the two states have similar widths in filling factor.

Changes:

We have included the capacitance data in the Supplementary Information (Fig. S2).

6.In Extended Data Figure 3, at varying electric fields the resistance saturates around 3K. What causes such saturations at the relatively "high" temperature compared to 0.285K? Is it due to RF noise in the system?

Response 29

We apologize for not showing the data appropriately. The saturation of resistance is caused by the finite impedance (100 MΩ) of the voltmeter used for the resistance measurement. In practice, reliable resistance measurements can only be performed up to about 10 MΩ for the sample resistance (about an order of magnitude lower than the impedance of the voltmeter). We have limited the scale of the y-axis in Extended Data Fig. 3 to 10 MΩ in the revised version and pointed out the maximum reliable resistance in our measurements in the revised Methods (page 6 and 7).

Overall, the study is interesting and there exists novelty in the manuscript, but to reach a positive recommendation, additional data and discussions are needed.

Response 30

We again thank the reviewer for the comments that have helped to improve our manuscript. We hope that with the additional new measurements presented above the reviewer will find our work convincing.

Reviewer Reports on the First Revision:

Referee #1:
Remarks to the Author:
The authors have addressed my comments well. I think the paper is ready for publication in Nature.

Referee #2:
Remarks to the Author:
The authors have made a substantial effort in improving the manuscript and new experimental results are provided. Most issues raised by the referees are well addressed. I have only three more comments left:
1. In Response 1 and the revised supplementary information, the authors provided new experimental results to quantify the band alignment between the MoTe2 and the WSe2 layers. They claimed that the band offset remains larger than 100 meV and the interlayer charge transfer is not expected. Although I think it is nice to have these new experimental results, band offset data along is not sufficient to support that there is no interlayer charge transfer. Whether there will be interlayer charge transfer depends also on the interlayer coupling strength. In Response 16, the authors also stated that "It (the closing of the moire gap) starts to happen when the offset between the upper valence band in MoTe2 and WSe2 becomes comparable to the interlayer hopping amplitude (which itself is dependent on the band offset)." In such situation, the states in the first moire bands should be written as the linear combination of the MoTe2 and the WSe2 states. At a given filling after the moire band gap collapsed at f=2, if the ratio of the WSe2 states in the first moire bands increase (or decrease) upon further increasing top gap voltage, then the charge density distribution on the WSe2 layers will be increased (or decreased). Given the total charge in the whole system is conserved at giving filling, such change of charge density distribution is amount to the charge transfer between the MoTe2 and the WSe2 layers. Without further evidences from DFT calculations or experiments, it may not be appropriate to claim that there is no charge transfer between the layers, in particularly after the moire gap collapsed at f=2 and before the MIT at f=1.
2. Naively, one would expect the interlayer hybridization/hopping is weaker when the band offset between the layers is larger, and vice versa. Assuming the moire potential is originated from the periodic modulation in interlayer hopping as the authors discussed in Response 15, when the band offset is reduced by the displacement field, the interlayer hybridization is stronger, shouldn't the moire potential become stronger? The present study seems to suggest one needs to increase the band offset to increase the depth of the moire potential, or increase the layer hybridization to reduce the depth of the moire potential, which is a bit counterintuitive. Maybe it is better to

provide more explanation.
3. The experimental results give compelling evidences for a continuous Mott transition at half-filling, which is probably due to the reduction of U/W. Although the DFT calculations seem to support the increase of the band width W while increasing the top gate voltage, many of the experimental results can also be explained by the reduction of U or V due to screening (e.g., the MIT at 2/3 filling). As a relatively thin hBN layers (~5nm) is used in the top gate in the experiments, is it possible that the screening by the graphite top gate is also enhanced when the top gate voltage is increased (see Liu, et al., Science 371(6535), 1261-1265) as an example in tuning the screening by changing the Fermi level of a nearby graphene bilayer)?

After the authors address these comments, I'll recommend it to be published in Nature.

Referee #3:
Remarks to the Author:
In Fig. 1c and Extended Data Figure 1, the author attributed the absence of 2/3 and 1/3 state in transport to the large-sample contact resistance for f<1 (Response 2). Could the author estimate the sample contact resistance as a function of filling factor and electrical field? Could the author discuss how much the contact resistance affects the fluctuation of square resistance at different filling?

The authors present a contour plot the square resistance as function of temperature and bottom gate voltage with a fixed top voltage (Fig.R5c, Response 6). In this case, this would change the filling factor and the electric field at the same time. Could the author clarify how much the varying electric field affect the square resistance? Could the author present the data at a fixed electric field?

When the sample is very insulating at f~1, the I-V curves are non-linear (Response 26). The authors argue that the thermal activation gap on the insulating side is extracted from the high-temperature data; the sample resistance is much smaller and I-V curves are linear. I suggest that the relevant data of I-V curves around f~1 on the insulating side at various temperatures should be added into the supplementary information. Especially the onset temperature of non-linear I-V curves showing up should be marked.

Author Rebuttals to First Revision:

Referee #2

The authors have made a substantial effort in improving the manuscript and new experimental results are provided. Most issues raised by the referees are well addressed. I have only three more comments left:

1. In Response 1 and the revised supplementary information, the authors provided new experimental results to quantify the band alignment between the MoTe2 and the WSe2 layers. They claimed that the band offset remains larger than 100 meV and the interlayer charge transfer is not expected. Although I think it is nice to have these new experimental results, band offset data along is not sufficient to support that there is no interlayer charge transfer. Whether there will be interlayer charge transfer depends also on the interlayer coupling strength. In Response 16, the authors also stated that “It (the closing of the moire gap) starts to happen when the offset between the upper valence band in MoTe2 and WSe2

becomes comparable to the interlayer hopping amplitude (which itself is dependent on the band offset)." In such situation, the states in the first moire bands should be written as the linear combination of the MoTe2 and the WSe2 states. At a given filling after the moire band gap collapsed at f=2, if the ratio of the WSe2 states in the first moire bands increase (or decrease) upon further increasing top gap voltage, then the charge density distribution on the WSe2 layers will be increased (or decreased). Given the total charge in the whole system is conserved at giving filling, such change of charge density distribution is amount to the charge transfer between the MoTe2 and the WSe2 layers. Without further evidences from DFT calculations or experiments, it may not be appropriate to claim that there is no charge transfer between the layers, in particularly after the moire gap collapsed at f=2 and before the MIT at f=1.

Response 1

We thank the reviewer for the comment and agree that the statement of no interlayer charge transfer is inappropriate in the presence of finite interlayer hybridization. In that case, as the reviewer pointed out, the energy eigenstates are in general layer-hybridized. Interlayer hybridization, however, does not modify the filling factor of the moiré unit cell or the first moiré band, and thus the physics discussed in this work. We have modified the Methods on Page 12 accordingly.

2. Naively, one would expect the interlayer hybridization/hopping is weaker when the band offset between the layers is larger, and vice versa. Assuming the moire potential is originated from the periodic modulation in interlayer hopping as the authors discussed in Response 15, when the band offset is reduced by the displacement field, the interlayer hybridization is stronger, shouldn't the moire potential become stronger? The present study seems to suggest one needs to increase the band offset to increase the depth of the moire potential, or increase the layer hybridization to reduce the depth of the moire potential, which is a bit counterintuitive. Maybe it is better to provide more explanation.

Response 2

We thank the referee for the comment on the origin of the moiré potential. There are typically multiple mechanisms that contribute to the formation of moiré potential in TMD moiré materials. One of them is the periodic modulation in interlayer hybridization that helps to tune the effective interaction strength in this study. Another one arises from intralayer lattice corrugation as indicated by several recent studies (e.g. Ref. 43 and 44). In angle-aligned $MoTe_2/WSe_2$ heterostructures, the z direction corrugation is about 0.5 Å after lattice relaxation. In the absence of an out-of-plane electric field, the interlayer tunneling is negligible because of the large band offset. The moiré potential is mostly determined by the intralayer potential due to lattice corrugation. Under a large electric field, the contribution to the moiré potential from interlayer hopping increases (which alone would give rise to a larger moiré potential). Both our experimental results and DFT calculations suggest that the two contributions compensate each other (or destructively interfere). As a result, the effective moiré potential actually decreases when the interlayer hopping increases at large electric fields. A similar effect has also been predicted in a recent DFT study of twisted bilayer WSe_2 (Kundu et al. arXiv:2103.07447). We have clarified the different contributions to the moiré potential in the revised manuscript on page 12.

3. The experimental results give compelling evidences for a continuous Mott transition at half-filling, which is probably due to the reduction of U/W. Although the DFT calculations seem to support the increase of the band width W while increasing the top gate voltage, many of the experimental results can also be explained by the reduction of U or V due to screening (e.g., the MIT at 2/3 filling). As a relatively thin hBN layers (~5nm) is used in the top gate in the experiments, is it possible that the screening by the graphite top gate is also enhanced when the top gate voltage is increased (see Liu, et al., Science 371(6535), 1261-1265) as an example in tuning the screening by changing the Fermi level of a nearby graphene bilayer)?

After the authors address these comments, I'll recommend it to be published in Nature.

Response 3

The reviewer raised an interesting question regarding whether the interaction strength can be effectively tuned by screening from the top graphite gate (which is 5 nm away from the sample in our experiment). Screening actually was the first effect we explored to induce the phase transitions in TMD moiré materials but without much success. Our observation is consistent with the result of the paper by Liu et al., that is, the screening effect is rather weak. Liu et al. were only able to perturb the energy scale of the correlated insulating state in twisted bilayer graphene by about 1 K with the screening layer very close to the sample (~ 3 nm). The Mott state in our system has an energy gap of 10's of meV (i.e. 100's K) and cannot be easily perturbed by gate-induced screening. In addition, a pure screening effect from the gate cannot close the single-particle moiré band gap at $f = 2$, as shown in our study. Based on these results, we conclude that the main effect of the electric field is to modify the moiré band and the moiré bandwidth. In the revised manuscript (page 10), we included a sentence on the negligible role of screening.

Referee #3

In Fig. 1c and Extended Data Figure 1, the author attributed the absence of 2/3 and 1/3 state in transport to the large-sample contact resistance for f<1 (Response 2). Could the author estimate the sample contact resistance as a function of filling factor and electrical field? Could the author discuss how much the contact resistance affects the fluctuation of square resistance at different filling?

Response 4

We thank the reviewer for the question on the contact resistance. To illustrate the effect of the contact resistance, in Fig. R1 we compare the 4-point square resistance (R1a) and the 2-point source-drain resistance (R1b) as functions of doping density and applied electric field. We were able to measure 4-point resistance reliably up to ~10 MΩ because the input impedance of the voltage pre-amplifiers used for the measurements is 100 MΩ. We were able to measure the 2-point source-drain resistance up to ~ 1 GΩ, which is limited by the sensitivity of current measurements. We estimate the contact resistance as a function of doping and electric field in Fig. R1c by subtracting the 4-point resistance from the 2-point resistance (taking into account the device geometry). We note that the difference includes contributions both from the contact resistance and sample inhomogeneity, particularly, in the region near the insulating states. We have used the same color scheme in Fig. R1a & c for comparison.

The data shows that the contact resistance decreases substantially at large filling factors and high electric fields. It is particularly large (10MΩ or more) and has significant fluctuations for $f < 1$ for most of the electric fields, reflecting the fact that the contact resistance cannot be precisely determined for $f < 1$. This makes 4-point measurement unreliable for $f < 1$ as can be seen by the large fluctuations in the square resistance in Fig. R1a. On the other hand, as long as the contact resistance is stable (we test the contacts for each device and choose the good ones) and substantially below the input impedance of the voltage pre-amplifiers (conditions satisfied near the Mott transition), the square resistance fluctuations are insignificant. This is evident in Fig. 1c, Extended Data Fig. 1 and Extended Data Fig. 6. In particular, we have negligible square resistance fluctuations near the metal-to-insulator transition at $f = 1$ (Fig. 2a). We have included Fig. R1 as Supplementary Fig. S12 in our revised manuscript.

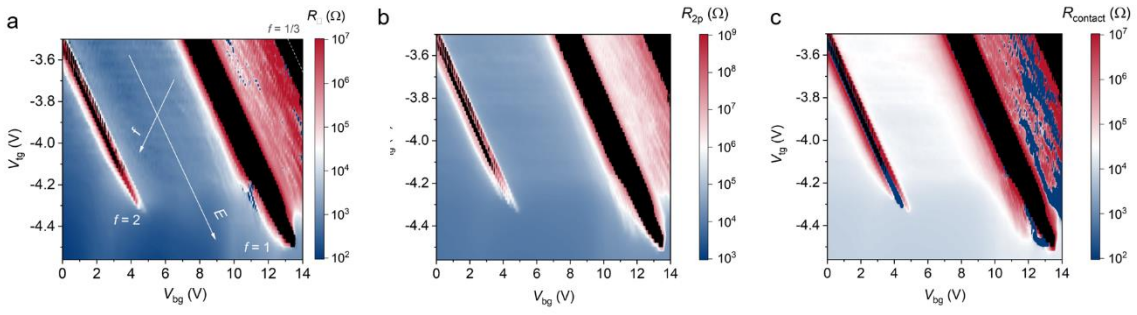

Figure R1. Square resistance, 2-point resistance and contact resistance. 2D map of the top and bottom gate voltage dependences of the square resistance (**a**), 2-point source-drain resistance (**b**) and the estimated contact resistance (**c**) from the difference between **b** and **a** at 300 mK. The estimated resistance in **c** also includes the

sample inhomogeneity effect, particularly, in the region near the insulating states. The contact resistance drops substantially at large filling factors and high electric fields. The contact resistance exceeds 10 MΩ for $f < 1$ for most of the electric fields.

The authors present a contour plot the square resistance as function of temperature and bottom gate voltage with a fixed top voltage (Fig.R5c, Response 6). In this case, this would change the filling factor and the electric field at the same time. Could the author clarify how much the varying electric field affect the square resistance? Could the author present the data at a fixed electric field?

Response 5

Figure R5c (Ext. Fig. 7c in the manuscript) shows the square resistance as a function of temperature and bottom gate voltage at a fixed top gate voltage. The range of gate voltages is chosen to feature the $f = 1$ state for the discussion of the Pomeranchuk effect. The two gates are asymmetric; the thickness of the gate dielectric is ~ 5 and 30 nm in the top and bottom gates, respectively. Both the doping density and electric field are thus much less sensitive to the bottom gate voltage than the top gate voltage. We focus on constant density at $f = 1$ by following the resistance peak. The peak shifts by < 10 mV in bottom gate voltage (as a reference, the half width of the peak is ~ 70 mV) for the entire range of temperature. This corresponds to a shift in the average applied field < 0.2 mV, which is the accuracy of the average applied field for this study. The electric field is therefore practically a constant very close to $f = 1$ throughout the entire temperature range. We have added an explicit statement in the revised caption of Ext. Data Fig. 7.

When the sample is very insulating at f~1, the I-V curves are non-linear (Response 26). The authors argue that the thermal activation gap on the insulating side is extracted from the high-temperature data; the sample resistance is much smaller and I-V curves are linear. I suggest that the relevant data of I-V curves around f~1 on the insulating side at various temperatures should be added into the supplementary information. Especially the onset temperature of non-linear I-V curves showing up should be marked.

Response 6

We agree with the reviewer that it is important to keep the I-V measurement in the linear regime to extract meaningful activation gaps. This issue becomes prominent only when the sample is very insulating at low temperatures; it does not affect our thermal activation analysis on the high-temperature data. We also limit the DC bias voltage to a small value (~ 1 mV) throughout the measurements. To further illustrate this point, we have collected more data from another device, which behaves similarly to the device shown in the main text. The new result is summarized in Fig. R2, which is also included in the revised manuscript as Supplementary Fig. S11.

Figure R2a shows the DC bias voltage dependence of the differential square resistance at varying electric fields from the critical point at 20 K. There is negligible bias dependence of the differential resistance for all electric fields. This confirms the linear I-V dependence at 20 K or above if we limit the bias voltage below 5 mV. Figure R2b shows the DC bias dependence of the differential square resistance at a fixed electric field (17 mV/nm from the critical point) at varying temperatures. The I-V dependence stays linear down to 10 K if we limit the bias to below ~ 1-1.5 mV. Figure R2c shows the temperature dependence of the differential square resistance measured with a DC bias of 1 mV; the vertical line marks the onset temperature for nonlinear I-V (observed at 5 mV). Only the data at higher temperatures is used for the thermal activation fit. Since we extract the activation gap using data above 10 K for all electric fields and keep the DC bias voltage at 1 mV, the nonlinear I-V at higher bias voltages or at lower temperatures does not affect our gap size analysis.

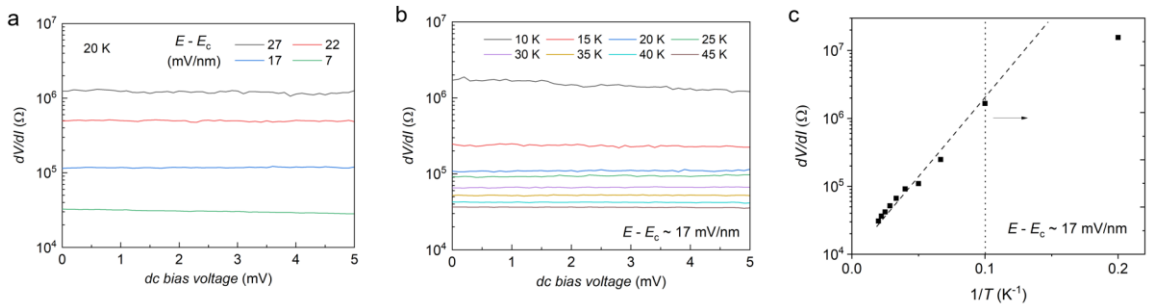

Figure R2. Bias dependence of the sample resistance at $f = 1$ on the insulating side (data from another device). a, Differential resistance, $\frac{dV}{dI}$, as a function of DC bias voltage (V) at 20 K under varying vertical electric fields. The AC modulation voltage is 0.5 mV. The differential resistance is approximately bias independent for bias up to 5 mV. **b,** Same as in **a** for a fixed electric field (17 mV/nm from the critical point) at varying temperatures. The differential resistance is approximately bias independent down to 10 K for bias below 1-1.5. **c,** Arrhenius plot of the square resistance at $V = 1$ mV. Thermal activation behavior is seen above ~ 10 K. The vertical dotted line indicates an onset of nonlinear I-V dependence observed at bias V = 5 mV.

Reviewer Reports on the Second Revision:

Referee #2 (Remarks to the Author):

The authors have nicely addressed all my concerns and I think it is ready for publication in Nature.

Referee #3 (Remarks to the Author):

My questions and comments are well addressed. I recommend the publication of the manuscript in Nature.