Electron and hole transport in disordered monolayer MoS$_2$: Atomic vacancy induced short-range and Coulomb disorder scattering

Kaasbjerg, Kristen; Low, Tony; Jauho, Antti-Pekka

Published in:
Physical Review B (Condensed Matter and Materials Physics)

Link to article, DOI:
10.1103/PhysRevB.100.115409

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Electron and hole transport in disordered monolayer MoS$_2$: Atomic vacancy induced short-range and Coulomb disorder scattering

Kristen Kaasbjerg 1, 4, Tony Low 2 and Antti-Pekka Jauho 1

1 Center for Nanostructured Graphene (CNG), Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
2 Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA

(Received 18 July 2019; revised manuscript received 26 August 2019; published 6 September 2019)

Atomic disorder is a common limiting factor for the low-temperature mobility in monolayer transition-metal dichalcogenides (TMDs; MX$_2$). Here, we study the effect of often occurring atomic vacancies on carrier scattering and transport in p- and n-type monolayer MoS$_2$. Due to charge trapping in vacancy-induced in-gap states, both neutral and charged vacancies resembling, respectively, short-range and combined short-range and long-range Coulomb scatterers must be considered. Using the T-matrix formalism, we demonstrate a strong renormalization of the Born description of short-range scattering, manifested in a pronounced reduction and a characteristic energy dependence of the scattering rate. As a consequence, carrier scattering in TMDs with charged vacancies is dominated by the long-range Coulomb disorder scattering, giving rise to a strong screening-induced temperature and density dependence of the low-temperature carrier mobility. For TMDs with neutral vacancies, the absence of intrinsic Coulomb disorder results in significantly higher mobilities as well as an unusual density dependence of the mobility which decreases with the carrier density. Our work illuminates the transport-limiting effects of atomic-vacancy scattering relevant for high-mobility TMD devices.

DOI: 10.1103/PhysRevB.100.115409

I. INTRODUCTION

Two-dimensional (2D) monolayers of transition-metal dichalcogenides (TMDs; MX$_2$) hold great promise for future electronics and optoelectronics [1–4]. In addition, their spin-valley coupling [5,6] makes them potential candidates for spin- and valleytronics applications, which among other things has sparked interest in TMD-based quantum-dot qubits [7–11]. For such purposes, high-mobility samples with long spin and valley lifetimes are essential.

Like in conventional 2D semiconductor heterostructure systems, disorder sets the ultimate limit for the achievable low-temperature mobility in monolayer TMDs [12], most often limited by short-range and Coulomb disorder scattering [13–21]. Only recently have mobilities exceeding $\sim$1000 cm$^2$/V$\cdot$s$^{-1}$ been achieved [11,22–25].

On the theoretical side, studies of the transport properties have focused on semiclassical transport including electron-phonon [26–29] and charged impurity scattering [30,31], microscopic descriptions of atomic point defects within the Kubo formalism [32–34], and quantum-transport studies [34–41]. However, the impact of atomic disorder on the transport properties is still poorly understood.

In experimental STM studies on monolayer TMDs, atomic monovacancies have been found to be among the dominating sources of intrinsic lattice disorder [43–48]. Their stability and electronic structure have been studied in great detail theoretically [49–57], demonstrating that they often introduce in-gap states, as illustrated in Fig. 1 for a S vacancy in MoS$_2$, which are localized at the vacancy site. In-gap states play a crucial role in the transport properties of 2D TMDs as they can trap charges. At low gate-induced doping levels, electrons become trapped in empty in-gap states and variable-range hopping between the defect sites dominates the transport [16,58–61]. With increasing doping, the material becomes n-type with free carriers in the conduction band. At this point, a transition from the insulating state with thermally activated hopping transport to a metallic conduction regime with a conductivity (resistivity) that decreases (increases) with increasing temperature takes place [13,14,16,17,58,62–64]. Usually, metallic behavior is attributed to electron-phonon scattering, but this is suppressed at low temperature, and disorder scattering by, e.g., atomic vacancies becomes the mobility-limiting factor [16–20].

FIG. 1. Atomic vacancies in 2D TMDs. (a) Top and side views of the atomic structure of a sulfur vacancy in 2D MoS$_2$. The $C_{3v}$ symmetry of the vacancy site results in a protection against inter-valley scattering [42]. (b) Sulfur-vacancy-induced in-gap states. In the undoped material ($V_g = 0$), the state close to the valence band maximum is occupied with two electrons from the missing sulfur anion. With an applied gate voltage, $V_g > 0$, one electron is trapped in the upper in-gap states before carriers are introduced into the conduction band ($V_g \gg 0$).
In this work, we provide an in-depth study of the effect of atomic vacancies on carrier scattering and transport in p- and n-type 2D MoS$_2$. Conventionally, atomic vacancies in 2D TMDs are treated as neutral point defects (see, e.g., Ref. [33]) which act as short-range scatterers. However, in the presence of vacancy-induced in-gap states, vacancies may acquire a dual character due to charging of the vacancy site, and should be treated as combined short-range and long-range Coulomb (i.e., charged impurity) scatterers.

The situation for an S vacancy in 2D MoS$_2$ [49–57] is illustrated in Fig. 1, where the vacancy gives rise to three in-gap states. The lowest state is doubly occupied in the undoped, charge-neutral material ($V_g = 0$), and hence behaves as a donor state which traps holes in p-doped materials. On the other hand, the upper states are empty and behave as deep single-electron acceptors which trap electrons in n-doped ($V_g > 0$) samples [16,50,51,59] (doubly charged vacancy sites are expected to be prohibited by a large on-site Coulomb repulsion energy). Thus, S vacancies are expected to acquire a net charge in p- or n-doped MoS$_2$ (positive and negative, respectively) due to charging of the in-gap states. The same holds for Mo vacancies in MoS$_2$ (see, e.g., Ref. [55]). However, while vacancies in 2D TMDs, in general, seem to introduce empty in-gap states, occupied states above the valence band edge are not always present [55,65]. In the latter case, the vacancy will remain neutral in the p-doped material.

Based on the $T$-matrix formalism [66,67], we here demonstrate that the scattering amplitude for the short-range component of the scattering potential is strongly renormalized with respect to its value in the Born approximation. For charged vacancies, this renders the short-range potential irrelevant in comparison to the Coulomb contribution to the scattering potential, and effectively reduces the vacancies to charged impurities.

The dominance of Coulomb disorder scattering gives rise to a strong temperature and density dependence of the mobility which stems from the temperature-dependent screening of the Coulomb potential. In the degenerate low-temperature ($T \ll T_F$) regime, Coulomb disorder resembles short-range disorder due to the efficient carrier screening in 2D TMDs, and yields a $\mu \sim n^0, T^0$ behavior. At the crossover to the nondegenerate ($T \gtrsim T_F$) regime, the screening efficiency is strongly reduced due to the temperature dependence of the 2D screening function, and the experimentally observed metallic behavior with a mobility that decreases with increasing temperature and increases with the carrier density prevails.

While this behavior is inherent to any 2D semiconductor systems at the quantum-classical crossover ($T = T_F$) between the degenerate and nondegenerate regimes [68], the temperature and density interval where the crossover takes place is highly dependent on material parameters (spin and valley degeneracy, effective mass). For typical carrier densities ($10^{11}$–$10^{13}$ cm$^{-2}$) in 2D TMDs, the Fermi temperature falls in the range $1 \, \text{K} \lesssim T_F \lesssim 100 \, \text{K}$, thus placing the quantum-classical crossover in an easily accessible temperature range where phonon scattering is weak [26,27].

In 2D TMDs where the vacancies remain neutral in the p-doped material, the mobility is demonstrated to be significantly higher and shows much weaker temperature dependence as well as a qualitatively different density dependence with a mobility that decreases with the carrier density. This behavior is inherent to atomic-vacancy-limited transport in 2D TMDs with charge-neutral vacancies.

Our findings are of high relevance for the understanding of the microscopic factors governing the mobility, magnetotransport, and quantum transport in 2D TMDs [11,19–25,69–71].

II. LOW-ENERGY HAMILTONIAN AND BOLTZMANN TRANSPORT THEORY

In 2D TMDs, the $K, K'$ valence and conduction band valleys illustrated in Fig. 2 can be described by the effective low-energy Hamiltonian [5]

$$\hat{H}_\nu(k) = \frac{\Delta}{2} \hat{\sigma}_z + \frac{\Delta_{\text{SO}}}{2} \hat{\sigma}_z \otimes \hat{\delta}_z,$$

where $a$ is the lattice parameter, $t$ a hopping parameter, $\tau = \pm 1$ is the $K, K'$ valley index, $\Delta$ is the band gap, $\Delta_{\text{SO}}$ ($=148$ meV in 2D MoS$_2$) is the spin-orbit (SO) induced spin splitting at the top of the valence band [72], and $\hat{\sigma}_z, \hat{\tau}_x, \hat{\tau}_y$ are the identity ($\sigma = 0$) and Pauli matrices ($i = x, y, z$) in the symmetry-adapted orbital basis, valley, and spin, respectively. The orbital basis $|\phi_{\nu \tau}\rangle$ is spanned by the $M$ d-orbitals $|\phi_{\nu \tau}\rangle = 1/\sqrt{2}(|d_{\nu \tau}^+\rangle + i\tau|d_{\nu \tau}^-\rangle)$ and $|\phi_{\nu \tau}\rangle = |d_{\nu \tau}\rangle$.

In the continuum description of Eq. (1), the wave function can be expressed as (suppressing the spin index $s = \pm 1$ for brevity) $\psi_{\nu \sigma \tau}(r) = \sum_A \frac{1}{A} e^{i \phi_{\nu \sigma \tau}} \chi_{\sigma \tau}$, where $A$ is the sample area and the valence ($\sigma = v$) and conduction ($\sigma = c$) band eigenspinors $\chi_{\sigma \tau}$ are dominated by, respectively, $|\phi_{\nu \sigma}\rangle$ and $|\phi_{\nu \sigma}\rangle$ near the band edges [36]. In this energy range, the bands are parabolic $E_{\sigma \tau} = E_{\sigma \tau}^0 \pm \hbar^2 k^2/2m^*_c$, where $E_{\sigma \tau} = \pm \Delta/2 + \delta_{\nu \tau} \tau \Delta_{\text{SO}}/2$ is the band-edge position (the spin dependence is shown in Fig. 2). The small SO splitting of 2–3 meV in the conduction band of MoS$_2$ [73,74] is here neglected and spin degeneracy is assumed. For the effective masses, we use our calculated DFT values, $m^*_c = 0.59$ and $m^*_c = 0.48$ for 2D MoS$_2$ [75], which are in good agreement with previously reported values [26,73,79].

![FIG. 2. Band structure and disorder scattering in the $K, K'$ valleys of 2D TMDs. While intravalley scattering is allowed (not shown), intervalley scattering is strongly suppressed by (i) the large spin-orbit splitting in the valence band (for some TMDs also in the conduction band), and (ii) a symmetry-induced selection rule for defects with $C_3$ symmetry and centered at the $M$ and $X$ sites in both the valence and conduction bands [42].](115409-2)
A. Disorder-limited transport

In Boltzmann transport theory, the carrier mobility \( \mu_{xx} = \sigma_{xx}/ne \), where \( \sigma_{xx} \) is the conductivity, limited by elastic disorder scattering can be expressed as \( \mu_{xx} = e(\tau_{B})/m^* \), where \( m^* \) is the effective mass and \( \tau_{B} \) is the total energy-dependent momentum relaxation time given by the sum over contributions from different types of disorder, \( \tau^{-1} = \sum_i \tau_i^{-1} \). The brackets \( \langle \cdot \rangle \) denote the energy-weighted average defined as

\[
\langle A \rangle = \frac{1}{n} \int d\varepsilon \rho_{\varepsilon}(\varepsilon) A(\varepsilon) \left(-\frac{\partial f}{\partial \varepsilon} \right).
\]

Here, \( n \) is the two-dimensional carrier density, \( \rho_{\varepsilon} = g_{\varepsilon}g_{\varepsilon}^* / 2\pi\hbar^2 \) is the constant density of states in 2D, \( g_{\varepsilon} \) and \( g_{\varepsilon}^* \) are the spin (\( g_{\varepsilon} = 1, 2 \) for \( v = u, c \)), and valley degeneracy, respectively, \( f(\varepsilon) = [1 + \exp\{ (\varepsilon - \mu)/k_B T \} ]^{-1} \) is the equilibrium Fermi-Dirac distribution function, and \( \mu \) is the chemical potential. For a degenerate electron gas, scattering is limited by elastic disorder, \( \tau \sim \tau_{\text{el}} \), where \( \tau_{\text{el}} \) is the momentum relaxation time given by the sum over contributions to spin-conserving scattering off nonmagnetic defects, \( \hat{V}_i \), and valley Green’s function, \( \hat{G}_{\sigma}(\varepsilon) \) is the intervalley coupling due to defects in 2D TMDs with C3 symmetry is suppressed by a symmetry-induced selection rule (for a given spin) is given by the integral equation

\[
T_{kk}^{\sigma\tau}(\varepsilon) = V_{kk}^{\sigma\tau} + \int \frac{d\varepsilon}{(2\pi)^2} V_{kk}^{\sigma\tau} G_{kk;kk}^{\sigma\tau}(\varepsilon) T_{kk}^{\sigma\tau}(\varepsilon),
\]

where \( G_{kk}^{\sigma\tau}(\varepsilon) \) is the bare Green’s function and the integral is over the first BZ.

With the band structure described by the low-energy model in Eq. (1), we confine the \( k \)-integral in (6) to the \( K, K' \) valleys, i.e., \( \int_{BZ} \frac{d\varepsilon}{(2\pi)^2} \rightarrow \sum_k \int_{BZ} \frac{d\varepsilon}{(2\pi)^2} \), and parametrize the defect matrix element by a band- and valley-dependent disorder strength \( V_{kk}^{\sigma\tau}(\varepsilon) \). For a general defect characterized by the matrix element \( V_{kk}^{\sigma\tau} \), which is small when \( \mu_{xx} \) is dilute disorder, i.e., \( n_i \ll 1/A_{\text{cell}} \), where \( A_{\text{cell}} \) is the unit-cell area.

In the following subsections, we analyze the scattering properties of atomic vacancies (Mo and S) in 2D TMDs.

B. Short-range potential: \( V_{ij} \)

Atomic point defects often act as strong short-range scatterers, and must be treated with the T-matrix formalism which describes multiple scattering off the same defect to infinite order in the scattering potential [66,67]. Here, we combine atomistic calculations with an analytic approach to get a realistic description of the T-matrix scattering rate.

1. Low-energy T-matrix model

For a general defect characterized by the matrix element \( V_{kk}^{\sigma\tau} \) of its scattering potential between the Bloch states with wave vectors \( k, k' \) in 1st Brillouin zone (BZ), the T matrix (for a given spin) is given by the integral equation

\[
T_{kk}^{\sigma\tau}(\varepsilon) = V_{kk}^{\sigma\tau} + \int \frac{d\varepsilon}{(2\pi)^2} V_{kk}^{\sigma\tau} G_{kk;kk}^{\sigma\tau}(\varepsilon) T_{kk}^{\sigma\tau}(\varepsilon),
\]

where \( G_{kk}^{\sigma\tau}(\varepsilon) = \delta(\varepsilon - \varepsilon_{k,k'})^{-1} \) is the bare Green’s function and the integral is over the first BZ.

With the band structure described by the low-energy model in Eq. (1), we confine the \( k \)-integral in (6) to the \( K, K' \) valleys, i.e., \( \int_{BZ} \frac{d\varepsilon}{(2\pi)^2} \rightarrow \sum_k \int_{BZ} \frac{d\varepsilon}{(2\pi)^2} \), and parametrize the defect matrix element by a band- and valley-dependent disorder strength \( V_{kk}^{\sigma\tau}(\varepsilon) \) like in Eq. (4). With this, the T-matrix equation in (6) reduces to a \( k, k' \)-independent algebraic \( 2 \times 2 \) matrix equation in the valley indices,

\[
T_{\sigma}^{ij}(\varepsilon) = \left( \delta_{\sigma} - \hat{G}_{\sigma}(\varepsilon) V_{0,\sigma}^{\sigma} \right)^{-1} V_{0,\sigma}^{ij},
\]

where \( \hat{G}_{\sigma}(\varepsilon) = \text{diag} [ \hat{G}_K(\varepsilon), \hat{G}_{K\varepsilon}(\varepsilon) ] \) is diagonal with elements given by the \( K \)-integrated valley Green’s function, \( \hat{G}_{\sigma}(\varepsilon) = \int \frac{d\varepsilon}{(2\pi)^2} \hat{G}_{\sigma}(\varepsilon) \), and the diagonal (off-diagonal) elements of \( V_{0,\sigma}^{ij} \) correspond to intravalley (intervalley) couplings.

In a recent work [42], we have shown that the \( K \leftrightarrow K' \) intervalley coupling due to defects in 2D TMDs with C3 symmetry is suppressed by a symmetry-induced selection rule, except for \( M \)-centered defects where intervalley coupling is possible in the conduction band as illustrated in Fig. 2.
the intra- and inter-valley matrix elements for \( M \) vacancies are comparable [42], a single disorder strength can therefore be considered for both \( M \) and \( X \) vacancies. The defect potential can hence be written

\[
V^{i}_{0,\sigma} = V^{i}_{0,\sigma} \times \begin{cases} \hat{\varepsilon}_0 + \hat{\varepsilon}_\nu, & \sigma = c, \ i = M, \\ \hat{\varepsilon}_0, & \text{otherwise}, \end{cases}
\]

(8)

which captures the defect and band dependent suppression of the inter-valley coupling described above.

With the form of the defect potential in (8), the \( T \) matrix in (7) simplifies considerably. For the two cases in (8), we get

\[
\hat{T}^{i}_\sigma(\varepsilon) = \begin{cases} T^{i}_{cK}(\varepsilon)(\hat{\varepsilon}_0 + \hat{\varepsilon}_\nu), \\ T^{i}_{\sigma K}((\varepsilon-\bar{\varepsilon}_0) - \hat{\varepsilon}_\nu)/2 + T^{i}_{\sigma K}((\varepsilon-\bar{\varepsilon}_0) + \hat{\varepsilon}_\nu)/2, \end{cases}
\]

(9)

where

\[
T^{i}_{\sigma \tau}(\varepsilon) = \frac{V^{i}_{0,\sigma}}{1 - g_{\sigma i}V^{i}_{0,\sigma}A_{\sigma \tau}(\varepsilon)}
\]

(10)

is the band and valley dependent \( T \) matrix with \( g_{\sigma i} \) denoting a valley multiplication factor given by the number of inequivalent valleys coupled by the defect type \( i \), i.e., \( g_{cM} = 2 \) and \( g_{\sigma i} = 1 \) otherwise. For the parabolic bands in the \( K, K' \) valleys, the \( \mathbf{k} \)-integrated Green’s function \( \hat{G}_{\sigma \tau} \) for \( \sigma = c, \nu \) becomes

\[
\hat{G}_{\sigma \tau}(\varepsilon) = \pm \bar{\rho}_{\sigma} \ln \left| \frac{\varepsilon - E_{\sigma \tau}}{\varepsilon - E_{\sigma \tau} \mp \Lambda} \right| - i\pi \bar{\rho}_{\sigma} \theta(\varepsilon \mp E_{\sigma \tau}),
\]

(11)

where \( \bar{\rho}_{\sigma} = m^*_\sigma/2\pi\hbar^2 \) is the density of states excluding spin and valley degeneracy, \( E_{\sigma \nu} \) is the valley (and spin) dependent band extrema, and \( \Lambda \) is an ultraviolet cutoff to be determined below. Note that due to the neglected spin-orbit splitting in the conduction band, i.e., \( E_{\sigma K} = E_{\sigma K'} \), the elements of the \( T \) matrix in the upper equation in Eq. (9) become independent of the valley index.

Via the optical theorem [66,67], \( T^{i}_{\sigma \tau}(\varepsilon) \) in Eq. (10) can be identified as the scattering amplitude in the \( T \)-matrix approximation. Together with Eq. (11), the formal condition for disorder to be weak (strong) thus becomes \( V^{i}_{0,\sigma} \bar{\rho}_{\sigma} \ll 1 \) (\( \gg 1 \)). In monolayer MoS\(_2\), \( \bar{\rho}_{\nu} \approx m^*_{\nu} \times 0.02 \ eV^{-1} \AA^{-2} \), implying that \( V^{i}_{0,\sigma} \gg / \ll 100 \ eV \AA^2 \) corresponds to strong/weak disorder. For weak disorder, the Born approximation for the scattering amplitude, \( T^{i}_{\sigma \tau}(\varepsilon) \approx V^{i}_{0,\sigma} \), is recovered, whereas the \( T \) matrix becomes independent of the disorder potential, \( T^{i}_{\sigma \tau}(\varepsilon) \approx -1/\hat{G}_{\sigma \tau}(\varepsilon) \), for strong disorder in the unitary limit, \( V^{i}_{0,\sigma} \rightarrow \infty \). In the intermediate regime, the \( T \) matrix should also be considered as it gives rise to a nonnegligible renormalization of the Born scattering amplitude.

\[
\tau_{i,\sigma \tau}(\varepsilon) = \frac{2\pi}{\hbar} \frac{n_{\sigma i} \bar{\rho}_{\sigma} \left| V^{i}_{0,\sigma} \right|^2}{\left( 1 + g_{\sigma i} \bar{\rho}_{\sigma} V^{i}_{0,\sigma} \ln (|\varepsilon - E_{\sigma \tau}|/\Lambda) \right)^2 + \left( \pi g_{\sigma i} \bar{\rho}_{\sigma} V^{i}_{0,\sigma} \right)^2}
\]

(12)

which for the constant defect potentials considered here is identical to the quantum scattering rate related to the imaginary part of the \( T \) matrix [see Eq. (13) below], and given by Eq. (3) with the replacement \( (1 - \cos \theta_{kk}) \rightarrow 1 \). Thus \( \tau_{\nu,\sigma \tau}^{-1}(\varepsilon) = -\frac{\pi}{\hbar} n V^{i}_{0,\sigma} T^{i}_{\sigma \tau}(\varepsilon)^2 \), where the factor \( g_{\sigma i} \bar{\rho}_{\sigma} \) stems from the \( k' \) integral in Eq. (3).

Figure 3 shows the energy dependence of the inverse relaxation time in Eq. (12) (full lines) in terms of \( -\text{Im} T^{i}_{\sigma \tau}(\varepsilon) \) for parameters corresponding to Mo and S vacancies in MoS\(_2\) (see Sec. II B 2 below). The characteristic behavior of the scattering rate, which increases away from the band edge where it drops sharply, can be traced back to \( \hat{G}_{\sigma \tau} \) in (11) which diverges for \( |\varepsilon - E_{\sigma \nu}| \rightarrow 0 \) due to the logarithm in its real part. This behavior is inherent to intermediate and strong short-range disorder, such as vacancies and point defects, in 2D materials with parabolic bands, and in stark contrast to the constant relaxation time in the Born approximation, \( \tau_{\nu,\sigma \tau}^{-1} \approx n V^{i}_{0,\sigma} \), which is recovered for weak disorder.

Also, the above is in contrast to the situation for vacancies in graphene where the scattering rate is predicted to have a nonmonotonic energy dependence due to resonance scattering off quasibound defect states in the vicinity of the Dirac point [80–84].

2. Effective disorder parameters for vacancies

To determine the \( T \)-matrix parameters for Mo and S vacancies which enter Eq. (10), we have used the atomistic method outlined in Refs. [42,84] to calculate the \( T \) matrix for atomic
vacancies in 2D MoS$_2$ based on Eq. (6) with DFT inputs for the band structure (including spin-orbit interaction) and defect matrix elements sampled in the entire BZ [75].

In order to quantify our atomistic calculations, we consider the quasiparticle scattering rate $\tau^{-1}_{\sigma,k} = \gamma^{-1}_{\sigma,k}/\hbar$, where

$$\gamma^{-1}_{\sigma,k} = -2\text{Im} \Sigma^{-1}_{\sigma,k}(\varepsilon_{\sigma,k}) = -2n_i \text{Im} \mathcal{T}^{	ext{on-shell}}_{\sigma,k}(\varepsilon_{\sigma,k})$$  \hspace{1cm} (13)

is the lifetime broadening given by the imaginary part of the on-shell ($\varepsilon = \varepsilon_{\sigma,k}$) $T$-matrix self-energy $\Sigma^{-1}_{\sigma,k} = n_i \mathcal{T}^{	ext{on-shell}}_{\sigma,k}$ of the disorder-averaged Green’s function [66,67].

Figure 3 shows the imaginary part of the calculated on-shell $T$ matrix for Mo (top) and S (bottom) vacancies [42,84] as a function energy in the $K$ valley of the valence (left) and conduction (right) bands. The red dots and blue squares correspond to the $T$ matrix for the spin-orbit split spin-up and spin-down bands, respectively. In the plots for the valence band, the offset in energy between the red dots and blue squares is due to the large spin-orbit splitting between the spin-up and spin-down bands, respectively. In the plots for the conduction band, the small spin-orbit splitting of $\sim 2–3$ meV is hardly discernible, and the imaginary part of the $T$ matrix is almost identical for the spin-up and spin-down bands. The $T$-matrix results in Fig. 3 can be converted to a lifetime broadening via Eq. (13), and correspond to a broadening of the order of $\gamma^{-1}_{\sigma,k} \sim 0.01–0.1$ meV for a disorder density of $n_i = 10^{11}$ cm$^{-2}$. Furthermore, the atomic $T$-matrix results confirm the characteristic energy dependence of the lifetime broadening, or scattering rate, predicted by Eq. (12).

The $T$-matrix parameters for vacancies are obtained by fitting the analytic expression (10) to the atomistic $T$ matrix in Fig. 3 [we fit the imaginary parts to achieve an optimal description of the scattering rate; cf. Eq. (13)]. In the fitting procedure, we first fix the disorder strength $V_{0,\sigma}^i$ to the value of the atomistic matrix elements $V_{0,\sigma}^i$ $T$ matrix at the $K$, $K'$ points (see Refs. [42,84]), and treat the ultraviolet cutoff $\Lambda$ as a fitting parameter. This is justified as $\Lambda$ does not have an immediate physical interpretation in the low-energy $T$-matrix model. Instead, it should be regarded as an effective parameter which compensates for approximating in Eq. (6) (i) the band structure in the full BZ with parabolic bands in the $K$, $K'$ valleys, and (ii) the defect matrix element by a constant disorder strength.

The resulting fits are shown with solid lines (the dashed lines show the unitary limit) in Fig. 3, and are seen to match the atomistic calculations almost perfectly, in particular at energies near the band edges relevant for transport. The corresponding fitting parameters are summarized in Table I.

As witnessed by Fig. 3, the $T$-matrix is highly dependent on the band. This is due to the $\pm 1$ factor in front of the real part of $G$ in Eq. (11), which even for an electron-hole symmetric band structure gives rise to different scattering rates for electrons and holes [cf. Eq. (12)]. For $V_{0}^i > 0$ (the reverse holds for $V_{0}^i < 0$), this leads to a hole scattering rate which is larger than the electron scattering rate as seen in Fig. 3. For the same reason, the electron and hole rates are, respectively, larger and smaller than the rate for unitary scattering.

Finally, we note that the $T$-matrix results in Fig. 3 correspond to a significant renormalization of the Born scattering amplitude given by the bare disorder strengths in Table I.

<table>
<thead>
<tr>
<th>Valence band</th>
<th>Conduction band</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mo vacancy</td>
<td>S vacancy</td>
</tr>
<tr>
<td>$V_{0,\sigma}^i$ = 145 eV Å$^2$</td>
<td>$V_{0,\sigma}^i$ = 155 eV Å$^2$</td>
</tr>
<tr>
<td>$\Lambda$ = 8 eV</td>
<td>$\Lambda$ = 20 eV</td>
</tr>
<tr>
<td>$g_{\sigma,i}$ = 1</td>
<td>$g_{\sigma,i}$ = 2</td>
</tr>
<tr>
<td>$V_{0,\sigma}^i$ = 24 eV Å$^2$</td>
<td>$V_{0,\sigma}^i$ = 15 eV Å$^2$</td>
</tr>
<tr>
<td>$\Lambda$ = 400 eV</td>
<td>$\Lambda$ = 100 eV</td>
</tr>
<tr>
<td>$g_{\sigma,i}$ = 1</td>
<td>$g_{\sigma,i}$ = 1</td>
</tr>
</tbody>
</table>

From the numbers in Fig. 3, the $T$-matrix scattering amplitudes are found to be up to an order of magnitude smaller than the Born scattering amplitudes. The renormalization is most pronounced for Mo vacancies which are strong (almost unitary) scatterers, and still noticeable for the weaker S vacancies. As the scattering rate depends on the square of the scattering amplitude, the Born approximation therefore severely overestimates the scattering rate.

To summarize, the strong $T$-matrix renormalization of the Born scattering amplitude results in (i) the characteristic energy dependence of the scattering rates in Fig. 3, as well as (ii) a significant reduction of the scattering rate relative to the Born result. These findings point to a concomitant breakdown of the Born approximation for atomic-vacancy scattering in 2D TMDs.

C. Long-range Coulomb potential: $V_C^i$

The relaxation time in Eq. (5) due to trapped charges in the in-gap states is governed by a long-range Coulomb potential. Due to the localized nature of the in-gap states, this can be approximated by the screened Coulomb potential from a point charge located at the vacancy site, and the matrix element becomes

$$V_C^i(q) = \frac{e^2}{2\epsilon_0 q \epsilon(q)}$$  \hspace{1cm} (14)

where $\epsilon(q) \equiv \epsilon(q,T,\mu)$ is the static dielectric function of the 2D material which includes both intra- and interband screening as discussed in further detail in Sec. II C 1 below.

For scattering off the long-range Coulomb potential, we use the Born approximation which is justified due to the screening of the Coulomb potential [30,31,68,85,86]. For a degenerate 2DEG, the dielectric function is given by $\epsilon(q) = \epsilon(1 + qT_{\text{TF}}/q) < 2k_F$, where $qT_{\text{TF}} = \frac{4\pi\epsilon_0\epsilon^*}{m^*}$ is the Thomas-Fermi (TF) wave vector and $\varepsilon$ is a background dielectric constant. In 2D TMDs, $qT_{\text{TF}} \sim \pi/\alpha \gg k_F$, where $\alpha$ is the lattice constant, implying that carrier screening is very efficient. This changes qualitatively the $1/q^4$ dependence of the unscreened Coulomb potential, with the screened Coulomb matrix element in (14) becoming

$$V_C^i(q) \approx \frac{e^2}{2\epsilon_0 q T_{\text{TF}}} = \frac{2\pi \hbar^2}{g_s g_i m^*} \equiv V_C^i, \quad q < 2k_F,$$  \hspace{1cm} (15)
which is independent of \( q \), and thus resembles short-range disorder with an effective disorder strength \( V_0^C \). For \( q > 2k_F \), the screening efficiency is reduced and the unscreened \( 1/q \) potential is recovered. In 2D MoS\(_2\), the effective disorder strengths in the valence and conduction bands are \( V_0^C \approx 41 \text{ eV Å}^2 \) and \( V_0^C \approx 25 \text{ eV Å}^2 \), respectively.

In passing, we note that scattering off an unscreened 2D Coulomb potential was studied with a partial-wave analysis (which is equivalent to a \( T \)-matrix treatment) in Ref. [87]. However, as has been shown in numerous works [30, 31, 68, 85, 86], it is important to take into account screening for a correct description of the density and temperature dependence of the low-temperature conductivities in 2D systems.

\[ \text{I. Dielectric function} \]

The long-wavelength dielectric function of a doped 2D semiconductor has two contributions: the first is due to the gate-induced extrinsic carriers and is normally accounted for in transport studies, and the second from the intrinsic screening associated with virtual electron-hole (e-h) pair excitation between the valence and conduction bands. Contrary to 3D bulk systems where the latter is well described by the dielectric constant of the material, intrinsic screening in 2D semiconductors is strongly \( q \) dependent [88, 89].

The total dielectric function thus becomes

\[
\epsilon(q, T, \mu) = 1 - V(q)\chi_{\text{ext}}(q, T, \mu) - V(q)\chi_{\text{int}}(q),
\]

where \( V(q) = e^2/2\epsilon_0q \) is the 2D Fourier transform of the Coulomb interaction and \( \chi_{\text{ext}}(\chi_{\text{int}}) \) is the polarizability due to intraband (interband) e-h pair processes involving extrinsic (intrinsic) carriers. Dielectric screening from the environment is included with the replacements \( \epsilon \rightarrow \kappa \epsilon \) and \( V \rightarrow V/\kappa \), where \( \kappa = 4.1 \) is an effective background dielectric constant for encapsulation in \( h \)-BN [31].

The screening from the extrinsic carriers in the valence/conduction band is described with finite-temperature RPA theory where the static polarizability at finite temperatures is given by [90, 91]

\[
\chi_{\text{ext}}(q, T, \mu) = \int_0^\infty d\mu' \frac{\chi_{\text{ext}}(q, 0, \mu')}{4k_B T} \frac{\cos^2(\sqrt{\mu' - \mu}/2m^*_F)}{m^*_F}.\]

Here, \( \chi_{\text{ext}}(q, 0, \mu) \) is the zero-temperature polarizability and the integral is evaluated following Ref. [92]. For a degenerate 2DEG, Eq. (17) yields the above-mentioned TF result for the extrinsic screening. In the nondegenerate regime, extrinsic carrier screening is described by Debye-Hückel theory and is strongly reduced. Thus, the strength of the Coulomb potential increases dramatically, and the transport is expected to become dominated by Coulomb disorder scattering at low carrier density and high temperature.

For the intrinsic screening due to interband e-h pair excitations between the valence and conduction bands, we use the dielectric function [88, 89]

\[
\epsilon_{\text{int}}(q) = 1 - \frac{e^2}{2\epsilon_0q} \chi_{\text{int}}(q) = 1 + \alpha q,
\]

which is valid in the long-wavelength limit and where \( \alpha \) is given by the interband polarizability \( \chi_{\text{int}}(q) \propto q^2 \), and \( \alpha \approx 40 \text{ Å} \) for MoS\(_2\) [93, 94]. Note that the intrinsic screening vanishes, \( \epsilon_{\text{int}} \rightarrow 1 \), for \( q \rightarrow 0 \) in 2D semiconductors. As it originates from e-h pair excitations between bands separated by a large band gap, the intrinsic screening is not expected to depend on temperature.

\[ \text{2. Extrinsic vs intrinsic screening} \]

The dielectric function in Eq. (16) is similar to the one in gated graphene [95]. However, due the sizable band gaps in 2D TMDs, intrinsic screening is much weaker than in graphene, and the long-wavelength dielectric function is dominated by intraband processes involving extrinsic carriers.

Already from the \( q \) dependence of the polarizabilities it is clear that extrinsic screening dominates the total dielectric function in Eq. (16) in the long-wavelength limit. For a degenerate carrier distribution, the wave vector \( q^* \) at which the two screening mechanisms contribute equally to the total dielectric function can be estimated by comparing the products of the Coulomb interaction and the polarizabilities,

\[
q_{TF}/q^* = \alpha q^* \rightarrow q^* = \sqrt{q_{TF}/\alpha}.
\]

For monolayer MoS\(_2\), we estimate \( q^* \approx 0.3 \pi/a \) based on the conduction band TF wave vector. Thus, extrinsic screening dominates at temperatures where the scattering wave vector is limited by \( q < 2k_F \ll q^* \), whereas at (i) high temperatures \( T > T_F \) where extrinsic carrier screening is reduced and at (ii) high carrier densities where \( k_F \) becomes appreciable compared to \( q^* \), intrinsic screening becomes important.

\[ \text{III. RESULTS} \]

In the following, we present our results for the density, \( \mu \sim n^2 \), and temperature, \( \mu \sim T^{-\gamma} \), dependencies of the mobility limited by atomic vacancies. As the inverse relaxation time scales directly with the disorder density, we restrict here the discussion to \( n_{\text{dis}} = 10^{11} \text{ cm}^{-2} \) which corresponds to a dilute concentration \( (\epsilon_{\text{dis}} \approx 0.01\%) \) of defects.

From our considerations in the previous section, we anticipate a density scaling of the mobility with (i) \( \alpha < 0 \) if the short-range potential dominates as the scattering rate increases with the electron and hole energies [cf. Fig. 3 and Eq. (12)], and (ii) \( \alpha \gtrsim 0 \) if the Coulomb potential dominates [85, 86].

\[ \text{A. Charged vacancies: Coulomb-disorder-dominated transport} \]

In Fig. 4 we show the density and temperature dependence of the mobility limited by charged vacancies. As we explain below, the results for charged Mo and S vacancies are almost identical; hence only the results for S vacancies are shown. The scaling exponents \( \alpha = d \log \mu/d \log n \) and \( \gamma = -d \log \mu/d \log T \) are shown in the bottom panels. To emphasize the effect of the intrinsic interband screening, we show the results both with (full lines) and without (dashed lines; \( \epsilon_{\text{int}} = 1 \) ) intrinsic screening, and the dots indicate the densities and temperatures where the Fermi temperature is equal to the actual temperature, \( T = T_F \), and hence mark the quantum-classical crossover between the degenerate \( (T < T_F) \) and nondegenerate \( (T > T_F) \) regimes.

Overall, the mobilities limited by charged S vacancies show a strong density and temperature dependence with
Electron mobility

Hole mobility

FIG. 4. Hole and electron mobilities limited by charged S vacancies. The plots show the hole (left) and electron (right) mobilities vs carrier density and temperature for a disorder density of \( n_s = 10^{11} \) cm\(^{-2}\). The corresponding density, \( \alpha = d \log \mu / d \log n \), and temperature, \( \gamma = -d \log \mu / d \log T \), exponents of the mobilities are shown in the lower row. The weak dashed lines show the mobility in the absence of intrinsic screening. The dots (•) indicate the temperatures and densities where \( T = T_F \).

\( \alpha > 0 \) and \( \gamma > 0 \), except at high temperatures and low densities where \( \gamma < 0 \). These observations indicate that the vacancy-independent Coulomb potential scattering dominates the inverse relaxation time for charged vacancies, thus resulting in almost identical mobilities for charged Mo and S vacancies. Based on the observations regarding the energy dependence of the relaxation time in Eq. (12), a \( \alpha < 0 \) density scaling would have been expected if the short-range potential was dominating.

This is also consistent with the observed strong temperature dependence of the mobility which stems from the temperature-dependent screening of the Coulomb potential. Moving from the quantum-classical crossover into the nondegenerate low-density/high-temperature regimes, the strength of the Coulomb potential increases dramatically due to a strong reduction in the efficiency of extrinsic carrier screening. In turn, this leads to a pronounced reduction of the mobility of up to two orders of magnitude.

The effect of intrinsic screening on the Coulomb potential is evident at high temperatures \( T > T_F \) where it becomes comparable to extrinsic screening. Here, it gives rise to an unusual upturn in the mobility with increasing temperature (compare full and dashed lines). Also, at high densities the mobilities show a pronounced upturn with increasing density. This is again due to the intrinsic screening which becomes weaker with increasing density \( (k_F) \) due to its linear \( q \) dependence in (18). Overall, the differences between the full and dashed lines show that intrinsic screening gives rise to a significant boost of the mobility in the nondegenerate regime and for high densities in the degenerate regime.

Finally, we note that the hole mobilities are in general lower than the electron mobilities in the degenerate regime. This can be attributed to the lack of spin degeneracy in the valence band which results in a weaker screening of Coulomb scattering for holes (cf. Sec. II C).

The density and temperature range where the pronounced screening-induced change in the mobilities in Fig. 4 takes place is summarized in the transport diagram shown for the S-vacancy-limited conductivity in \( n \)-type MoS\(_2\) in Fig. 5. The diagram shows the conductivity \( \sigma = e^2 n(t)/m^* \) (here \( \log \sigma \)) in the \( (n, T) \) plane and the black dashed line indicates the quantum-classical crossover at \( T = T_F \). As is evident, the crossover takes place at experimentally easily accessible densities and temperatures.

The quantum-classical crossover is clearly visible in the conductivity data. In the degenerate high-density regime \( (T \ll T_F) \), the conductivity scales as \( \sigma \sim n \) and is almost independent of temperature due to the effective short-range nature of the Coulomb potential in Eq. (15) caused by carrier screening. At \( T \gtrsim T_F \), the conductivity acquires a stronger density and temperature dependence due to the reduced screening of the Coulomb potential.

\[ \text{FIG. 5. Transport diagram for } n \text{-type 2D MoS}_2 \text{ showing the conductivity } \sigma \text{ limited by charged S vacancies (} n_s = 10^{11} \text{ cm}^{-2}\text{) as a function of carrier density } n \text{ and temperature } T. \]
B. Neutral vacancies: Short-range disorder limited transport

In this section, we consider the situation where the vacancies remain neutral in doped samples. This is expected to occur in 2D TMDs where vacancies do not introduce occupied in-gap states (e.g., X = S, Se vacancies in WX$_2$ [55,65]), implying that the vacancy will remain overall neutral in p-doped samples and only the short-range contribution to the relaxation time in Eq. (5) remains.

To demonstrate the impact on the hole mobility in TMDs where the above-mentioned situation is realized, we show in Fig. 6 the hole mobility in MoS$_2$ limited by neutral S vacancies at different temperatures. The dashed line shows the zero-temperature mobility $\mu_{0,i} = e\tau_{0,i}(E_F)/m^*_i$ with $\tau_{0,i}$ given by its $T$-matrix form in Eq. (12).

![FIG. 6. Hole mobility in 2D MoS$_2$ limited by neutral S vacancies, $n_S = 10^{11}$ cm$^{-2}$, at different temperatures. The dashed line shows the zero-temperature mobility $\mu_{0,i} = e\tau_{0,i}(E_F)/m^*_i$ with $\tau_{0,i}$ given by its $T$-matrix form in Eq. (12).](image)

The qualitative difference between the screening-induced and $T$-matrix-induced density and temperature dependencies of the mobility emerges clearly from the transport diagrams for charged and neutral vacancies in Figs. 5 and 7, respectively. While in the latter case the conductivity $\sigma \sim n^\alpha$ scales roughly as $\alpha \approx 1$ on both sides of the quantum-classical crossover, the screening-induced density dependence of the scattering rate in the former case results in a stronger $\alpha \approx 2$ behavior in the nondegenerate regime at $T \gtrsim T_F$.

These qualitative differences in the transport characteristics allow us to clearly distinguish Coulomb-disorder-limited transport from short-range-limited transport due to, e.g., charged and neutral vacancies, respectively.

IV. DISCUSSION

Reported experimental low-temperature mobilities in 2D TMDs have thus far been rather low [13,14,16,17,19,63], and have only recently exceeded $\sim$1000 cm$^{-2}$ V$^{-1}$ s$^{-1}$ [11,22–25]. This is still slightly lower than our calculated mobilities limited by charged vacancies and much lower than the calculated mobilities limited by neutral vacancies. This indicates that the concentration of vacancies in experimental samples could be higher than $n_{\text{dis}} = 10^{11}$ cm$^{-2}$ used here. Several reports of high concentrations of S vacancies [16,59] suggest that this is indeed a likely reason for the low experimental mobilities. Only recently have defect densities of $n_{\text{dis}} \sim 10^{11}$ cm$^{-2}$ been demonstrated [47]. However, experimental transport properties of such high-quality TMDs have so far not been reported.

The often observed strong metallic temperature and density dependence of experimental low-temperature mobilities [13,14,16,17,19,62,63] resemble best our results for charged vacancies. This indicates that a combination of short-range and Coulomb disorder scattering is limiting the low-temperature mobility in 2D TMDs. As we here suggest, vacancies may be the source of both short-range and Coulomb disorder scattering. Hence, there is no need to introduce additional Coulomb disorder [30,31] (e.g., charged impurities in the substrate) in order to account for the metallic transport behavior.

V. CONCLUSIONS

In conclusion, we have studied the effect of atomic vacancies on carrier scattering and transport in p- and n-type monolayer MoS$_2$. Due to the presence of both filled and empty
vacancy-induced in-gap states, the vacancies can be expected to become charged in p- and n-doped MoS$_2$, and thereby give rise to both short-range and Coulomb disorder scattering. The situation is similar for vacancies in many other 2D TMDs, but cases lacking filled in-gap states have been reported, implying that the vacancy remains neutral in the p-doped material.

Studying the short-range scattering properties of vacancies with the T-matrix formalism, we show that multiple scattering gives rise to a strong renormalization of the Born scattering amplitude, which results in a pronounced reduction as well as a characteristic energy dependence of the scattering rate. As a result, the Coulomb contribution to the scattering potential for charged vacancies by far dominates carrier scattering. This results in a strong screening-induced temperature and density dependence of the mobility in 2D TMDs hosting charged vacancies. For TMDs with neutral vacancies the mobility is significantly higher and shows an unusual behavior with a decreasing density dependence. Thus, TMDs in which vacancies remain neutral in the doped material are better candidates for high-mobility devices.

ACKNOWLEDGMENTS

K.K. acknowledges support from the Carlsberg Foundation and the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Grant Agreement No. 713683 (COFUNDfellowsDTU). T.L. acknowledges partial support from NSF ECCS-1542202. The Center for Nanostructured Graphene (CNG) is sponsored by the Danish National Research Foundation, Project No. DNRF103.


[75] The DFT calculations of the band structure, effective masses, and defect matrix elements have been performed with the electronic structure code GPAW [76–78] within the projector augmented wave method, using the LDA xc functional and a LCAO double-zeta polarized (DZP) basis set. The defect potentials were obtained using a 11 × 11 supercell with 10 Å of vacuum between the MoS 2 sheet and the cell boundaries in the vertical direction imposed with Dirichlet boundary conditions in order to avoid spurious interlayer interactions. Lastly, the defect matrix elements and T matrix were calculated on a 135 × 135 BZ grid using a numerical broadening of η = 10 meV.


