SIAM J. APPL. MATH.
Vol. 67, No. 4, pp. 1183–1201

ON THE SHOCKLEY–READ–HALL MODEL: GENERATION-RECOMBINATION IN SEMICONDUCTORS*

THIERRY GOUDON†, VERA MILJANOVIĆ‡, AND CHRISTIAN SCHMEISER§

Abstract. The Shockley–Read–Hall model for generation-recombination of electron-hole pairs in semiconductors based on a quasi-stationary approximation for electrons in a trapped state is generalized to distributed trapped states in the forbidden band and to kinetic transport models for electrons and holes. The quasi-stationary limit is rigorously justified both for the drift-diffusion and for the kinetic model.

Key words. semiconductor, generation, recombination, drift-diffusion, kinetic model

AMS subject classification. 78A35

DOI. 10.1137/060650751

1. Introduction.

The Shockley–Read–Hall (SRH) model was introduced in 1952 [15], [9] to describe the statistics of recombination and generation of holes and electrons in semiconductors occurring through the mechanism of trapping.

The transfer of electrons from the valence band to the conduction band is referred to as the generation of electron-hole pairs (or pair-generation process), since not only is a free electron created in the conduction band, but also a hole in the valence band which can contribute to the charge current. The inverse process is termed recombination of electron-hole pairs. The bandgap between the upper edge of the valence band and the lower edge of the conduction band is very large in semiconductors, which means that a big amount of energy is needed for a direct band-to-band generation event. The presence of trap levels within the forbidden band caused by crystal impurities facilitates this process, since the jump can be split into two parts, each of them "cheaper" in terms of energy. The basic mechanisms are illustrated in Figure 1: (a) hole emission (an electron jumps from the valence band to the trapped level), (b) hole capture (an electron moves from an occupied trap to the valence band, and a hole disappears), (c) electron emission (an electron jumps from the trapped level to the conduction band), (d) electron capture (an electron moves from the conduction band to an unoccupied trap).

Models for this process involve equations for the densities of electrons in the conduction band, holes in the valence band, and trapped electrons. Basic for the SRH model are the drift-diffusion assumption for the transport of electrons and holes, the assumption of one trap level in the forbidden band, and the assumption that

*Received by the editors January 24, 2006; accepted for publication (in revised form) March 7, 2007; published electronically May 29, 2007. This work was supported by the European IHP network "HYKE-Hyperbolic and Kinetic Equations: Asymptotics, Numerics, Analysis" under contract HPRN-CT-2002-00282, and by the Austrian Science Fund under project W008 (Wissenschaftskolleg "Differential Equations").

http://www.siam.org/journals/siap/67-4/65075.html

†Team SIMPAF–INRIA Futurs et Laboratoire Paul Painlevé UMR 8524, CNRS–Université des Sciences et Technologies Lille 1, Cité Scientifique, F-59655 Villeneuve d'Ascq cedex, France (thierry.goudon@math.univ-lille1.fr).

‡Wolfgang Pauli Institut, Universität Wien, Nordbergstraße 15C, 1090 Wien, Austria (vera.miljanovic@univie.ac.at).

§Fakultät für Mathematik, Universität Wien, Nordbergstraße 15C, 1090 Wien, Austria, and RICAM Linz, Österreichische Akademie der Wissenschaften, Altenbergstr. 56, 4040 Linz, Austria (christian.schmeiser@univie.ac.at).

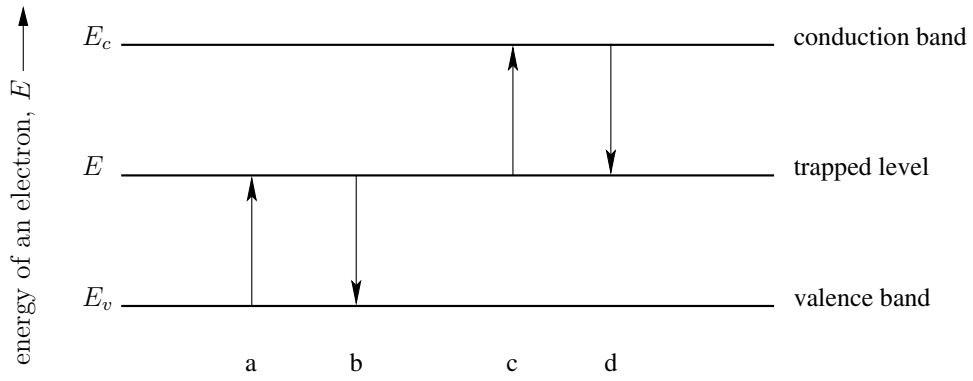

FIG. 1. *The four basic processes of electron-hole recombination.*

the dynamics of the trapped electrons is quasi-stationary, which can be motivated by the smallness of the density of trapped states compared to typical carrier densities. This last assumption leads to the elimination of the density of trapped electrons from the system and to a nonlinear effective recombination-generation rate, reminiscent of Michaelis–Menten kinetics in chemistry. This model is an important ingredient of simulation models for semiconductor devices (see, e.g., [10], [14]).

In this work, two generalizations of the classical SRH model are considered: Instead of a single trapped state, a distribution of trapped states across the forbidden band is allowed, and, in a second step, a semiclassical kinetic model including the fermion nature of the charge carriers is introduced. Although direct band-to-band recombination-generation (see, e.g., [13]) and impact ionization (e.g., [2], [3]) have been modelled on the kinetic level before, this is (to the best of the authors' knowledge) the first attempt to derive a "kinetic SRH model." (We also mention the modelling discussions and numerical simulations in [7], [8].)

For both the drift-diffusion and the kinetic models with self-consistent electric fields existence results and rigorous results concerning the quasi-stationary limit are proven. For the drift-diffusion problem, the essential estimate is derived similarly to [6], where the quasi-neutral limit has been carried out. For the kinetic model Degond's approach [4] for the existence of solutions of the Vlasov–Poisson problem is extended. Actually, the existence theory already provides the uniform estimates necessary for passing to the quasi-stationary limit.

In the following section, the drift-diffusion based model is formulated and nondimensionalized, and the SRH model is formally derived. Section 3 contains the rigorous justification of the passage to the quasi-stationary limit. Section 4 corresponds to section 2, dealing with the kinetic model, and in section 5 existence of global solutions for the kinetic model is proven, and the quasi-stationary limit is justified.

2. The drift-diffusion Shockley–Read–Hall model. We consider a semiconductor crystal with a forbidden band represented by the energy interval (E_v, E_c) with the valence band edge E_v and the conduction band edge E_c. The constant (in space) number density N_{tr} of trapped states is obtained by summing up contributions across the forbidden band:

$$N_{tr} = \int_{E_v}^{E_c} M_{tr}(E)\, dE.$$

Here $M_{tr}(E)$ is the energy dependent density of available trapped states. The position density of occupied traps is given by

$$n_{tr}(f_{tr})(x,t) = \int_{E_v}^{E_c} M_{tr}(E) f_{tr}(x,E,t)\, dE,$$

where $f_{tr}(x,E,t)$ is the fraction of occupied trapped states at position $x \in \Omega$, energy $E \in (E_v, E_c)$, and time $t \geq 0$. Note that $0 \leq f_{tr} \leq 1$ should hold from a physical point of view.

The evolution of f_{tr} is coupled to those of the density of electrons in the conduction band, denoted by $n(x,t) \geq 0$, and the density of holes in the valence band, denoted by $p(x,t) \geq 0$. Electrons and holes are oppositely charged. The coupling is expressed through the following quantities:

$$S_n = \frac{1}{\tau_n N_{tr}}\big[n_0 f_{tr} - n(1-f_{tr})\big], \qquad S_p = \frac{1}{\tau_p N_{tr}}\big[p_0(1-f_{tr}) - p f_{tr}\big], \tag{1}$$

$$R_n = \int_{E_v}^{E_c} S_n M_{tr}\, dE, \qquad R_p = \int_{E_v}^{E_c} S_p M_{tr}\, dE. \tag{2}$$

Indeed, the governing equations are given by

$$\partial_t f_{tr} = S_p - S_n = \frac{p_0}{\tau_p N_{tr}} + \frac{n}{\tau_n N_{tr}} - f_{tr}\left(\frac{p_0+p}{\tau_p N_{tr}} + \frac{n_0+n}{\tau_n N_{tr}}\right), \tag{3}$$

$$\partial_t n = \nabla \cdot J_n + R_n, \quad J_n = \mu_n(U_T \nabla n - n \nabla V), \tag{4}$$

$$\partial_t p = -\nabla \cdot J_p + R_p, \quad J_p = -\mu_p(U_T \nabla p + p \nabla V), \tag{5}$$

$$\varepsilon_s \Delta V = q(n + n_{tr}(f_{tr}) - p - C). \tag{6}$$

For the current densities J_n, J_p we use the simplest possible model, the drift diffusion ansatz, with constant mobilities μ_n, μ_p, and with thermal voltage U_T. Moreover, since the trapped states have fixed positions, no flux appears in (3).

By R_n and R_p we denote the recombination-generation rates for n and p, respectively. The rate constants are $\tau_n(E)$, $\tau_p(E)$, $n_0(E)$, $p_0(E)$, where $n_0(E)p_0(E) = {n_i}^2$ with the energy independent intrinsic density n_i.

Integration of (3) yields

$$\partial_t n_{tr} = R_p - R_n. \tag{7}$$

By adding (4), (5), (7), we obtain the continuity equation

$$\partial_t(p - n - n_{tr}) + \nabla \cdot (J_n + J_p) = 0, \tag{8}$$

with the total charge density $p - n - n_{tr}$ and the total current density $J_n + J_p$.

In the Poisson equation (6), $V(x,t)$ is the electrostatic potential, ε_s the permittivity of the semiconductor material, q the elementary charge, and $C = C(x)$ the given doping profile.

Note that if τ_n, τ_p, n_0, p_0 are independent of E, or if there exists only one trap level E_{tr} with $M_{tr}(E) = N_{tr}\delta(E - E_{tr})$, then $R_n = \frac{1}{\tau_n}[n_0 \frac{n_{tr}}{N_{tr}} - n(1 - \frac{n_{tr}}{N_{tr}})]$, $R_p = \frac{1}{\tau_p}[p_0(1 - \frac{n_{tr}}{N_{tr}}) - p\frac{n_{tr}}{N_{tr}}]$, and (4), (5) together with (7) are a closed system governing the evolution of n, p, and n_{tr}.

We now introduce a scaling of n, p, and f_{tr} in order to render (4)–(6) dimensionless:

Scaling of parameters:

i. $M_{tr} \to \frac{N_{tr}}{E_c - E_v} M_{tr}$.

ii. $\tau_{n,p} \to \bar{\tau}\tau_{n,p}$, where $\bar{\tau}$ is a typical value for τ_n and τ_p.

iii. $\mu_{n,p} \to \bar{\mu}\mu_{n,p}$, where $\bar{\mu}$ is a typical value for $\mu_{n,p}$.

iv. $(n_0, p_0, n_i, C) \to \bar{C}(n_0, p_0, n_i, C)$, where $\bar{C}$ is a typical value of C.

Scaling of unknowns:

v. $(n, p) \to \bar{C}(n, p)$.

vi. $n_{tr} \to N_{tr} n_{tr}$.

vii. $V \to U_T V$.

viii. $f_{tr} \to f_{tr}$.

Scaling of independent variables:

ix. $E \to E_v + (E_c - E_v)E$.

x. $x \to \sqrt{\bar{\mu} U_T \bar{\tau}}\, x$, where the reference length is a typical diffusion length before recombination.

xi. $t \to \bar{\tau} t$, where the reference time is a typical carrier life time.

Dimensionless parameters:

xii. $\lambda = \sqrt{\frac{\varepsilon_s}{q\bar{C}\bar{\mu}\bar{\tau}}} = \frac{1}{\bar{x}}\sqrt{\frac{\varepsilon_s U_T}{q\bar{C}}}$ is the scaled Debye length.

xiii. $\varepsilon = \frac{N_{tr}}{\bar{C}}$ is the ratio of the density of traps to the typical doping density, and will be assumed to be small: $\varepsilon \ll 1$.

The scaled system reads as follows:

(9)
$$\varepsilon \partial_t f_{tr} = S_p(p, f_{tr}) - S_n(n, f_{tr}), \quad S_p = \frac{1}{\tau_p}\big[p_0(1 - f_{tr}) - p f_{tr}\big],$$
$$S_n = \frac{1}{\tau_n}\big[n_0 f_{tr} - n(1 - f_{tr})\big],$$

(10)
$$\partial_t n = \nabla \cdot J_n + R_n(n, f_{tr}), \quad J_n = \mu_n(\nabla n - n\nabla V), \quad R_n = \int_0^1 S_n M_{tr}\, dE,$$

(11)
$$\partial_t p = -\nabla \cdot J_p + R_p(p, f_{tr}), \quad J_p = -\mu_p(\nabla p + p\nabla V), \quad R_p = \int_0^1 S_p M_{tr}\, dE,$$

(12)
$$\lambda^2 \Delta V = n + \varepsilon n_{tr} - p - C, \quad n_{tr}(f_{tr}) = \int_0^1 f_{tr} M_{tr}\, dE,$$

with $n_0(E)p_0(E) = n_i^2$ and $\int_0^1 M_{tr}\, dE = 1$.

By letting $\varepsilon \to 0$ in (9) formally, we obtain $f_{tr} = \frac{\tau_n p_0 + \tau_p n}{\tau_n(p + p_0) + \tau_p(n + n_0)}$, and the reduced system has the following form:

(13) $$\partial_t n = \nabla \cdot J_n + R(n, p),$$

(14) $$\partial_t p = -\nabla \cdot J_p + R(n, p),$$

(15) $$R(n, p) = ({n_i}^2 - np) \int_0^1 \frac{M_{tr}(E)}{\tau_n(E)(p + p_0(E)) + \tau_p(E)(n + n_0(E))\, dE},$$

(16) $$\lambda^2 \Delta V = n - p - C.$$

Note that if τ_n, τ_p, n_0, p_0 are independent of E or if there exists only one trap level, then we would have the standard SRH model, with $R = \frac{{n_i}^2 - np}{\tau_n(p+p_0)+\tau_p(n+n_0)}$. Existence and uniqueness of solutions of the limiting system (13)–(16) under assumptions (21)–(25) stated below is a standard result in semiconductor modelling. A proof can be found in, e.g., [10].

3. Rigorous derivation of the drift-diffusion Shockley–Read–Hall model. We consider the system (9)–(12) with the position x varying in a bounded domain $\Omega \in \mathbb{R}^3$ (all our results are easily extended to the one- and two-dimensional situations), the energy $E \in (0,1)$, and time $t > 0$, subject to initial conditions

$$n(x,0) = n_I(x), \quad p(x,0) = p_I(x), \quad f_{tr}(x,E,0) = f_{tr,I}(x,E) \tag{17}$$

and mixed Dirichlet–Neumann boundary conditions

$$n(x,t) = n_D(x,t), \quad p(x,t) = p_D(x,t), \quad V(x,t) = V_D(x,t), \quad x \in \partial\Omega_D \subset \partial\Omega, \tag{18}$$

and

$$\frac{\partial n}{\partial \nu}(x,t) = \frac{\partial p}{\partial \nu}(x,t) = \frac{\partial V}{\partial \nu}(x,t) = 0, \quad x \in \partial\Omega_N := \partial\Omega \setminus \partial\Omega_D, \tag{19}$$

where ν is the unit outward normal vector along $\partial\Omega_N$. We permit the special case that either $\partial\Omega_D$ or $\partial\Omega_N$ is empty. More precisely, we assume that either $\partial\Omega_D$ has positive $(d-1)$-dimensional measure, or it is empty. In the second situation ($\partial\Omega_D$ empty) we have to assume total charge neutrality; i.e.,

$$\int_\Omega (n + \varepsilon n_{tr} - p - C)\, dx = 0 \qquad \text{if } \partial\Omega = \partial\Omega_N. \tag{20}$$

The potential is then determined only up to a (physically irrelevant) additive constant.

The following assumptions on the data will be used: For the boundary data, given any $0 < T < \infty$,

$$n_D, p_D \in W^{1,\infty}(0,T; W^{1,\infty}_{loc}(\Omega)), \quad V_D \in L^\infty(0,T; W^{1,6}(\Omega)); \tag{21}$$

for the initial data

$$n_I, p_I \in H^1(\Omega) \cap L^\infty(\Omega), \quad 0 \le f_{tr,I} \le 1, \tag{22}$$

$$\int_\Omega (n_I + \varepsilon n_{tr}(f_{tr,I}) - p_I - C)\, dx = 0 \qquad \text{if } \partial\Omega = \partial\Omega_N; \tag{23}$$

for the doping profile

$$C \in L^\infty(\Omega); \tag{24}$$

and for the recombination-generation rate constants

$$n_0, p_0, \tau_n, \tau_p \in L^\infty((0,1)), \quad \tau_n, \tau_p \ge \tau_{min} > 0. \tag{25}$$

With these assumptions, a local existence and uniqueness result for the problem (9)–(12), (17)–(19) for fixed positive ε can be proven by a straightforward extension of the approach in [5] (see also [10]). In the following, local existence will be assumed, and we shall concentrate on obtaining bounds which guarantee global existence and

which are uniform in ε as $\varepsilon \to 0$. For the sake of simplicity, we consider that the data in (21), (22), and (24) do not depend on ε; of course, our strategy works when dealing with sequences of data bounded in the mentioned spaces.

The following result is a generalization of [6, Lemma 3.1], where the case of homogeneous Neumann boundary conditions and vanishing recombination was treated. Our proof uses a similar approach.

LEMMA 3.1. *Let the assumptions* (21)–(25) *be satisfied. Then, the solution of* (9)–(12), (17)–(19) *exists for all times and satisfies* $n, p \in L^\infty(0,T;L^\infty(\Omega)) \cap L^2(0,T;H^1(\Omega))$ *uniformly in* ε *as* $\varepsilon \to 0$ *as well as* $0 \le f_{tr} \le 1$.

Proof. Global existence will be a consequence of the following estimates. Introducing the new variables $\widetilde{n} = n - n_D$, $\widetilde{p} = p - p_D$, $\widetilde{C} = C - \varepsilon n_{tr} - n_D + p_D$, (10)–(12) take the following form:

$$\partial_t \widetilde{n} = \nabla \cdot J_n + R_n - \partial_t n_D, \quad J_n = \mu_n\big[\nabla\widetilde{n} + \nabla n_D - (\widetilde{n} + n_D)\nabla V\big], \tag{26}$$

$$\partial_t \widetilde{p} = -\nabla J_p + R_p - \partial_t p_D, \quad J_p = -\mu_p\big[\nabla\widetilde{p} + \nabla p_D + (\widetilde{p} + p_D)\nabla V\big], \tag{27}$$

$$\lambda^2 \Delta V = \widetilde{n} - \widetilde{p} - \widetilde{C}. \tag{28}$$

As a consequence of $0 \le f_{tr} \le 1$, $\widetilde{C} \in L^\infty((0,\infty)\times\Omega)$ holds. For $q \ge 2$ and even, we multiply (26) by $\widetilde{n}^{q-1}/\mu_n$ and (27) by $\widetilde{p}^{q-1}/\mu_p$, and add:

$$\begin{aligned}
\frac{d}{dt}\int_\Omega \left[\frac{\widetilde{n}^q}{q\mu_n} + \frac{\widetilde{p}^q}{q\mu_p}\right] dx \;=\; & -(q-1)\int_\Omega \widetilde{n}^{q-2}\nabla\widetilde{n}\nabla n\,dx - (q-1)\int_\Omega \widetilde{p}^{q-2}\nabla\widetilde{p}\nabla p\,dx \\
& + (q-1)\int_\Omega \left[\widetilde{n}^{q-2} n\nabla\widetilde{n} - \widetilde{p}^{q-2} p\nabla\widetilde{p}\right]\nabla V\,dx \\
& + \int_\Omega \frac{\widetilde{n}^{q-1}}{\mu_n}(R_n - \partial_t n_D) + \int_\Omega \frac{\widetilde{p}^{q-1}}{\mu_p}(R_p - \partial_t p_D) \\
=:\; & I_1 + I_2 + I_3 + I_4 + I_5.
\end{aligned} \tag{29}$$

Using the assumptions on n_D, p_D and $|R_n| \le C(n+1)$, $|R_p| \le C(p+1)$, we estimate

$$I_4 \le C\int_\Omega |\widetilde{n}|^{q-1}(n+1)\,dx \le C\left(\int_\Omega \widetilde{n}^q\,dx + 1\right), \quad I_5 \le C\left(\int_\Omega \widetilde{p}^q\,dx + 1\right).$$

The term I_3 can be rewritten as follows:

$$\begin{aligned}
I_3 = & \int_\Omega \left[\widetilde{n}^{q-1}\nabla\widetilde{n} - \widetilde{p}^{q-1}\nabla\widetilde{p}\right]\nabla V\,dx \\
& + \int_\Omega \left[\widetilde{n}^{q-2}\nabla\widetilde{n}\right](n_D\nabla V)\,dx - \int_\Omega \left[\widetilde{p}^{q-2}\nabla\widetilde{p}\right](p_D\nabla V)\,dx \\
= & -\frac{1}{\lambda^2 q}\int_\Omega \left[\widetilde{n}^q - \widetilde{p}^q\right](\widetilde{n} - \widetilde{p} - \widetilde{C})\,dx \\
& - \frac{1}{\lambda^2(q-1)}\int_\Omega \widetilde{n}^{q-1}\left(\nabla n_D\nabla V + n_D(\widetilde{n} - \widetilde{p} - \widetilde{C})\right)dx \\
& + \frac{1}{\lambda^2(q-1)}\int_\Omega \widetilde{p}^{q-1}\left(\nabla p_D\nabla V + p_D(\widetilde{n} - \widetilde{p} - \widetilde{C})\right)dx.
\end{aligned}$$

The second equality uses integration by parts and (28). The first term on the right-hand side is the only term of degree $q+1$. It reflects the quadratic nonlinearity of

the problem. Fortunately, it can be written as the sum of a term of degree q and a nonnegative term. By estimation of the terms of degree q using the assumptions on n_D and p_D as well as $\|\nabla V\|_{L^q(\Omega)} \le C(\|\widetilde{n}\|_{L^q(\Omega)} + \|\widetilde{p}\|_{L^q(\Omega)} + \|\widetilde{C}\|_{L^q(\Omega)})$, we obtain

$$\begin{aligned} I_3 &\le -\frac{1}{\lambda^2 q}\int_\Omega [\widetilde{n}^q - \widetilde{p}^q]\,(\widetilde{n}-\widetilde{p})\,dx + C\left(\int_\Omega (\widetilde{n}^q + \widetilde{p}^q)\,dx + 1\right) \\ &\le C\left(\int_\Omega (\widetilde{n}^q + \widetilde{p}^q)\,dx + 1\right). \end{aligned}$$

The integral I_1 can be written as

$$I_1 = -\int_\Omega \widetilde{n}^{q-2}|\nabla n|^2\,dx + \int_\Omega \widetilde{n}^{q-2}\nabla n_D \nabla n\,dx. \tag{30}$$

By rewriting the integrand in the second integral as

$$\widetilde{n}^{q-2}\nabla n_D \nabla n = \widetilde{n}^{\frac{q-2}{2}}\nabla n \widetilde{n}^{\frac{q-2}{2}}\nabla n_D$$

and applying the Cauchy–Schwarz inequality, we have the following estimate for (30):

$$\begin{aligned} I_1 &\le -\int_\Omega \widetilde{n}^{q-2}|\nabla n|^2\,dx + \sqrt{\int_\Omega \widetilde{n}^{q-2}|\nabla n|^2\,dx \int_\Omega \widetilde{n}^{q-2}|\nabla n_D|^2\,dx} \\ &\le -\frac{1}{2}\int_\Omega \widetilde{n}^{q-2}|\nabla n|^2\,dx + C\|\widetilde{n}\|_{L^q}^{q-2} \le -\frac{1}{2}\int_\Omega \widetilde{n}^{q-2}|\nabla n|^2\,dx + C\left(\int_\Omega \widetilde{n}^q\,dx + 1\right). \end{aligned}$$

For I_2, the same reasoning (with n and n_D replaced by p and p_D, respectively) yields an analogous estimate. Collecting our results, we obtain

$$\begin{aligned} \frac{d}{dt}\int_\Omega \left[\frac{\widetilde{n}^q}{q\mu_n} + \frac{\widetilde{p}^q}{q\mu_p}\right]dx &\le -\frac{1}{2}\int_\Omega \widetilde{n}^{q-2}|\nabla n|^2\,dx - \frac{1}{2}\int_\Omega \widetilde{p}^{q-2}|\nabla p|^2\,dx \\ &\quad + C\left(\int_\Omega (\widetilde{n}^q + \widetilde{p}^q)\,dx + 1\right). \end{aligned} \tag{31}$$

Since $q \ge 2$ is even, the first two terms on the right-hand side are nonpositive, and the Gronwall lemma gives

$$\int_\Omega (\widetilde{n}^q + \widetilde{p}^q)\,dx \le e^{qCt}\left(\int_\Omega (\widetilde{n}(t=0)^q + \widetilde{p}(t=0)^q)\,dx + 1\right).$$

A uniform-in-q-and-ε estimate for $\|n\|_{L^q}$, $\|p\|_{L^q}$ follows, and the uniform-in-ε bound in $L^\infty(0,T;L^\infty(\Omega))$ is obtained in the limit $q \to \infty$. The estimate in $L^2(0,T;H^1(\Omega))$ is then derived by returning to (31) with $q = 2$. ∎

Now we are ready to prove the main result of this section.

THEOREM 3.2. *Let the assumptions of Theorem* 3.1 *be satisfied. Then, as $\varepsilon \to 0$, for every $T > 0$, the solution (f_{tr}, n, p, V) of* (9)–(12), (17)–(19) *converges with convergence of f_{tr} in $L^\infty((0,T)\times\Omega\times(0,1))$ weak*, n and p in $L^2((0,T)\times\Omega)$, and V in $L^2(0,T;H^1(\Omega))$. The limits of n, p, and V satisfy* (13)–(19).

Proof. The L^∞-bounds for f_{tr}, n, and p, which are uniform with respect to ε, and the Poisson equation (12) imply ∇V is bounded in $L^2((0,T)\times\Omega)$. From the definition of J_n, J_p (see (4), (5)), it then follows that $J_n, J_p \in L^2((0,T)\times\Omega)$. Then (10) and (11) together with $R_n, R_p \in L^\infty((0,T)\times\Omega)$ imply $\partial_t n, \partial_t p \in L^2(0,T;H^{-1}(\Omega))$.

The previous result and the Aubin lemma (see, e.g., Simon [16, Corollary 4, p. 85]) give compactness of n and p in $L^2((0,T)\times\Omega)$.

We already know from the Poisson equation that $\nabla V \in L^\infty(0,T;H^1(\Omega))$. By taking the time derivative of (12), one obtains

$$\partial_t \Delta V = \nabla\cdot(J_n + J_p),$$

with the consequence that $\partial_t \nabla V$ is bounded (uniformly with respect to ε) in $L^2((0,T)\times\Omega)$. Therefore, the Aubin lemma can again be applied as above to prove compactness of ∇V in $L^2((0,T)\times\Omega)$.

These results and the weak compactness of f_{tr} are sufficient for passing to the limit in the nonlinear terms $n\nabla V$, $p\nabla V$, nf_{tr}, and pf_{tr}. Let us also remark that $\partial_t n$ and $\partial_t p$ are bounded in $L^2(0,T;H^{-1}(\Omega))$, so that n,p are compact in $C^0([0,T];L^2(\Omega)$ weak). With this remark the initial data for the limit equation make sense. By the uniqueness result for the limiting problem (mentioned at the end of section 2), the convergence is not restricted to subsequences. ∎

4. A kinetic Shockley–Read–Hall model. In this section we replace the drift-diffusion model for electrons and holes by a semiclassical kinetic transport model. It is governed by the system

$$\partial_t f_n + v_n(k)\cdot\nabla_x f_n + \frac{q}{\hbar}\nabla_x V\cdot\nabla_k f_n = Q_n(f_n) + Q_{n,r}(f_n,f_{tr}), \tag{32}$$

$$\partial_t f_p + v_p(k)\cdot\nabla_x f_p - \frac{q}{\hbar}\nabla_x V\cdot\nabla_k f_p = Q_p(f_p) + Q_{p,r}(f_p,f_{tr}), \tag{33}$$

$$\partial_t f_{tr} = S_p(f_p,f_{tr}) - S_n(f_n,f_{tr}), \tag{34}$$

$$\varepsilon_s\Delta_x V = q(n + n_{tr} - p - C), \tag{35}$$

where $f_i(x,k,t)$ represents the particle distribution function (with $i=n$ for electrons and $i=p$ for holes) at time $t\ge 0$, at the position $x\in\mathbb{R}^3$, and at the wave vector (or generalized momentum) $k\in\mathbb{R}^3$. All functions of k have the periodicity of the reciprocal lattice of the semiconductor crystal. Equivalently, we shall consider only $k\in B$, where B is the Brillouin zone, i.e., the set of all k which are closer to the origin than to any other lattice point, with periodic boundary conditions on ∂B.

The coefficient functions $v_n(k)$ and $v_p(k)$ denote the electron and hole velocities, respectively, which are related to the electron and hole band diagrams by

$$v_n(k) = \nabla_k\varepsilon_n(k)/\hbar, \quad v_p(k) = -\nabla_k\varepsilon_p(k)/\hbar,$$

where $\hbar$ is the reduced Planck constant. The elementary charge is still denoted by q.

The collision operators Q_n and Q_p describe the interactions between the particles and the crystal lattice. They involve several physical phenomena and can be written in the general form

$$Q_n(f_n) = \int_B \widetilde{\Phi}_n(k,k')[M_n f_n'(1-f_n) - M_n' f_n(1-f_n')]\,dk', \tag{36}$$

$$Q_p(f_p) = \int_B \widetilde{\Phi}_p(k,k')[M_p f_p'(1-f_p) - M_p' f_p(1-f_p')]\,dk', \tag{37}$$

with the primes denoting evaluation at k', with the nonnegative, symmetric scattering cross sections $\widetilde{\Phi}_n(k,k')$ and $\widetilde{\Phi}_p(k,k')$, and with the Maxwellians

$$M_n(k) = c_n\exp(-\varepsilon_n(k)/k_BT), \quad M_p(k) = c_p\exp(-\varepsilon_p(k)/k_BT),$$

where k_BT is the thermal energy of the semiconductor crystal lattice and the constants c_n, c_p are chosen such that

$$\int_B M_n \, dk = \int_B M_p \, dk = 1.$$

The remaining collision operators $Q_{n,r}(f_n, f_{tr})$ and $Q_{p,r}(f_p, f_{tr})$ model the generation and recombination processes and are given by

$$Q_{n,r}(f_n, f_{tr}) = \int_{E_v}^{E_c} \hat{S}_n(f_n, f_{tr}) M_{tr} \, dE, \tag{38}$$

with

$$\hat{S}_n(f_n, f_{tr}) = \frac{\Phi_n(k, E)}{N_{tr}} \left[n_0 M_n f_{tr}(1 - f_n) - f_n(1 - f_{tr}) \right],$$

and

$$Q_{p,r}(f_p, f_{tr}) = \int_{E_v}^{E_c} \hat{S}_p(f_p, f_{tr}) M_{tr} \, dE, \tag{39}$$

with

$$\hat{S}_p(f_p, f_{tr}) = \frac{\Phi_p(k, E)}{N_{tr}} \left[p_0 M_p(1 - f_p)(1 - f_{tr}) - f_p f_{tr} \right],$$

and where $\Phi_{n,p}$ are nonnegative and $M_{tr}(x, E)$ is the same density of available trapped states as for the drift-diffusion model, except that we allow for a position dependence now. This will be commented on below. The parameter N_{tr} is now determined as $N_{tr} = \sup_{x \in \mathbb{R}^3} \int_0^1 M_{tr}(x, E) \, dE$.

The right-hand side in the equation for the occupancy $f_{tr}(x, E, t)$ of the trapped states is defined by

$$S_n(f_n, f_{tr}) = \int_B \hat{S}_n \, dk = \lambda_n[n_0 M_n(1 - f_n)] f_{tr} - \lambda_n[f_n](1 - f_{tr}), \tag{40}$$

with $\lambda_n[g] = \int_B \Phi_n g \, dk$, and

$$S_p(f_p, f_{tr}) = \int_B \hat{S}_p \, dk = \lambda_p[p_0 M_p(1 - f_p)](1 - f_{tr}) - \lambda_p[f_p] f_{tr}, \tag{41}$$

with $\lambda_p[g] = \int_B \Phi_p g \, dk$.

The factors $(1 - f_n)$ and $(1 - f_p)$ take into account the Pauli exclusion principle, which therefore manifests itself in the requirement that the values of the distribution function have to respect the bounds $0 \le f_n, f_p \le 1$.

The position densities on the right-hand side of the Poisson equation (35) are given by

$$n(x, t) = \int_B f_n \, dk, \quad p(x, t) = \int_B f_p \, dk, \quad n_{tr}(x, t) = \int_{E_v}^{E_c} f_{tr} M_{tr} \, dE.$$

The following scaling, which is strongly related to the one used for the drift-diffusion model, will render (32)–(35) dimensionless:

Scaling of parameters:

i. $M_{tr} \to \frac{N_{tr}}{E_v - E_c} M_{tr}$.
ii. $(\varepsilon_n, \varepsilon_p) \to k_B T(\varepsilon_n, \varepsilon_p)$, with the thermal energy $k_B T$.
iii. $(\Phi_n, \Phi_p) \to \tau_{rg}^{-1}(\Phi_n, \Phi_p)$, where τ_{rg} is a typical carrier life time.
iv. $(\widetilde{\Phi}_n, \widetilde{\Phi}_p) \to \tau_{coll}^{-1}(\widetilde{\Phi}_n, \widetilde{\Phi}_p)$.
v. $(n_0, p_0, C) \to \overline{C}(n_0, p_0, C)$, where $\overline{C}$ is a typical value of $|C|$.
vi. $(M_n, M_p) \to \overline{C}^{-1}(M_n, M_p)$.

Scaling of independent variables:

vii. $x \to k_B T \sqrt{\tau_{rg}\tau_{coll}}\overline{C}^{-1/3}\hbar^{-1}\, x$.
viii. $t \to \tau_{rg} t$.
ix. $k \to \overline{C}^{1/3} k$.
x. $E \to E_v + (E_c - E_v)E$.

Scaling of unknowns:

xi. $(f_n, f_p, f_{tr}) \to (f_n, f_p, f_{tr})$.
xii. $V \to U_T V$, with the thermal voltage $U_T = k_B T/q$.

Dimensionless parameters:

xiii. $\alpha^2 = \frac{\tau_{coll}}{\tau_{rg}}$.
xiv. $\lambda = \frac{\hbar}{q\sqrt{\tau_{rg}\tau_{coll}}\overline{C}^{1/6}} \sqrt{\frac{\varepsilon_s}{k_B T}}$.
xv. $\varepsilon = \frac{N_{tr}}{\overline{C}}$, where again we shall study the situation $\varepsilon \ll 1$.

Finally, the scaled system reads as follows:

$$\alpha^2 \partial_t f_n + \alpha v_n(k) \cdot \nabla_x f_n + \alpha \nabla_x V \cdot \nabla_k f_n = Q_n(f_n) + \alpha^2 Q_{n,r}(f_n, f_{tr}), \tag{42}$$

$$\alpha^2 \partial_t f_p + \alpha v_p(k) \cdot \nabla_x f_p - \alpha \nabla_x V \cdot \nabla_k f_p = Q_p(f_p) + \alpha^2 Q_{p,r}(f_p, f_{tr}), \tag{43}$$

$$\varepsilon \partial_t f_{tr} = S_p(f_p, f_{tr}) - S_n(f_n, f_{tr}), \tag{44}$$

$$\lambda^2 \Delta_x V = n + \varepsilon n_{tr} - p - C = -\rho, \tag{45}$$

with $v_n = \nabla_k \varepsilon_n$, $v_p = -\nabla_k \varepsilon_p$, with Q_n and Q_p still having the form (36) and (37), respectively, with the scaled Maxwellians

$$M_n(k) = c_n \exp(-\varepsilon_n(k)), \qquad M_p(k) = c_p \exp(-\varepsilon_p(k)), \tag{46}$$

and with the recombination-generation terms

$$Q_{n,r}(f_n, f_{tr}) = \int_0^1 \hat{S}_n M_{tr}\, dE, \qquad Q_{p,r}(f_p, f_{tr}) = \int_0^1 \hat{S}_p M_{tr}\, dE, \tag{47}$$

with

$$\hat{S}_n = \Phi_n[n_0 M_n f_{tr}(1-f_n) - f_n(1-f_{tr})], \quad \hat{S}_p = \Phi_p[p_0 M_p(1-f_{tr})(1-f_p) - f_p f_{tr}]. \tag{48}$$

The right-hand side of (44) still has the form (40), (41). The position densities are given by

$$n = \int_B f_n\, dk, \qquad p = \int_B f_p\, dk, \qquad n_{tr} = \int_0^1 f_{tr} M_{tr}\, dE. \tag{49}$$

The system (42)–(44) conserves the total charge $\rho = p + C - n - \varepsilon n_{tr}$. With the definition

$$J_n = -\frac{1}{\alpha} \int_B v_n f_n\, dk, \quad J_p = \frac{1}{\alpha} \int_B v_p f_p\, dk$$

of the current densities, the following continuity equation holds formally:

$$\partial_t \rho + \nabla_x \cdot (J_n + J_p) = 0.$$

Formally setting $\varepsilon = 0$ in (44), we obtain

$$\overline{f}_{tr}(f_n, f_p) = \frac{p_0 \lambda_p[M_p(1-f_p)] + \lambda_n[f_n]}{p_0 \lambda_p[M_p(1-f_p)] + \lambda_p[f_p] + \lambda_n[f_n] + n_0 \lambda_n[M_n(1-f_n)]}.$$

Substituting $\overline{f}_{tr}$ into (47) leads to the kinetic SRH recombination-generation operators

$$\begin{aligned} \overline{Q}_{n,r}(f_n, f_p) &= \overline{g}_n[f_n, f_p](1-f_n) - \overline{r}_n[f_n, f_p] f_n, \\ \overline{Q}_{p,r}(f_n, f_p) &= \overline{g}_p[f_n, f_p](1-f_p) - \overline{r}_p[f_n, f_p] f_p, \end{aligned} \tag{50}$$

with

$$\begin{aligned} \overline{g}_n &= \int_0^1 \frac{\Phi_n M_n n_0 \Big(p_0 \lambda_p[M_p(1-f_p)] + \lambda_n[f_n] \Big) M_{tr}}{p_0 \lambda_p[M_p(1-f_p)] + \lambda_p[f_p] + \lambda_n[f_n] + n_0 \lambda_n[M_n(1-f_n)]}\, dE, \\ \overline{r}_n &= \int_0^1 \frac{\Phi_n \Big(\lambda_p[f_p] + n_0 \lambda_n[M_n(1-f_n)] \Big) M_{tr}}{p_0 \lambda_p[M_p(1-f_p)] + \lambda_p[f_p] + \lambda_n[f_n] + n_0 \lambda_n[M_n(1-f_n)]}\, dE, \\ \overline{g}_p &= \int_0^1 \frac{\Phi_p M_p p_0 \Big(n_0 \lambda_n[M_n(1-f_n)] + \lambda_p[f_p] \Big) M_{tr}}{p_0 \lambda_p[M_p(1-f_p)] + \lambda_p[f_p] + \lambda_n[f_n] + n_0 \lambda_n[M_n(1-f_n)]}\, dE, \\ \overline{r}_p &= \int_0^1 \frac{\Phi_p \Big(\lambda_n[f_n] + p_0 \lambda_p[M_p(1-f_p)] \Big) M_{tr}}{p_0 \lambda_p[M_p(1-f_p)] + \lambda_p[f_p] + \lambda_n[f_n] + n_0 \lambda_n[M_n(1-f_n)]}\, dE. \end{aligned}$$

Of course, the limiting model still conserves charge, which is expressed by the identity

$$\int_B \overline{Q}_{n,r}\, dk = \int_B \overline{Q}_{p,r}\, dk.$$

Pairs of electrons and holes are generated or recombine, however, generally not with the same wave vector. This absence of momentum conservation is reasonable since the process involves an interaction with the trapped states fixed within the crystal lattice.

5. Rigorous derivation of the kinetic Shockley–Read–Hall model. The limit $\varepsilon \to 0$ will be carried out rigorously in an initial value problem for the kinetic model: From now on we work with $x \in \mathbb{R}^3$ (and we avoid any discussion on boundary conditions and possible boundary layers). Concerning the behavior for $|x| \to \infty$, we shall require the densities to be in L^1 and use the Newtonian potential solution of the Poisson equation; i.e., (45) will be replaced by

$$\mathcal{E}(x,t) = -\nabla_x V = \lambda^{-2} \int_{\mathbb{R}^3} \frac{x-y}{|x-y|^3} \rho(y,t)\, dy. \tag{51}$$

We define Problem (K) as the system (42)–(44), (51) with (36), (37), (47)–(49), (40), and (41), subject to the initial conditions

$$f_n(x,k,0) = f_{n,I}(x,k), \quad f_p(x,k,0) = f_{p,I}(x,k), \quad f_{tr}(x,E,0) = f_{tr,I}(x,E).$$

We start by stating our assumptions on the data. For the velocities we assume

$$v_n, v_p \in W^{1,\infty}_{per}(B), \tag{52}$$

where here and in the following, the subscript *per* denotes Sobolev spaces of functions of k satisfying periodic boundary conditions on ∂B. Further we assume that the cross sections satisfy

$$\widetilde{\Phi}_n, \widetilde{\Phi}_p \geq 0, \quad \widetilde{\Phi}_n, \widetilde{\Phi}_p \in W^{1,\infty}_{per}(B \times B), \tag{53}$$

and

$$\Phi_n, \Phi_p \geq 0, \quad \Phi_n, \Phi_p \in W^{1,\infty}_{per}(B \times (0,1)). \tag{54}$$

A finite total number of trapped states is assumed:

$$M_{tr} \geq 0, \quad M_{tr} \in W^{1,\infty}(\mathbb{R}^3 \times (0,1)) \cap W^{1,1}(\mathbb{R}^3 \times (0,1)).$$

The L^1-assumption with respect to x is needed for controlling the total number of generated particles. For the initial data we assume

$$\begin{aligned} &0 \leq f_{n,I}, f_{p,I} \leq 1, \quad f_{n,I}, f_{p,I} \in W^{1,\infty}_{per}(\mathbb{R}^3 \times B) \cap W^{1,1}_{per}(\mathbb{R}^3 \times B), \\ &0 \leq f_{tr,I} \leq 1, \quad f_{tr,I} \in W^{1,\infty}_{per}(\mathbb{R}^3 \times (0,1)). \end{aligned} \tag{55}$$

We also assume

$$n_0, p_0 \in L^\infty((0,1)), \qquad C \in W^{1,\infty}(\mathbb{R}^3) \cap W^{1,1}(\mathbb{R}^3). \tag{56}$$

Finally, we need an upper bound for the life time of trapped electrons:

$$\int_B (\Phi_n \min\{1, n_0 M_n\} + \Phi_p \min\{1, p_0 M_p\})\, dk \geq \gamma > 0. \tag{57}$$

The reason for the various differentiability assumptions above is that we shall construct smooth solutions by an approach along the lines of [13], which goes back to [4].

An essential tool relies on the following potential theory estimates:

$$\|\mathcal{E}\|_{L^\infty(\mathbb{R}^3)} \leq C \|\rho\|^{1/2}_{L^1(\mathbb{R}^3)} \|\rho\|^{1/2}_{L^\infty(\mathbb{R}^3)}, \tag{58}$$

$$\|\nabla_x \mathcal{E}\|_{L^\infty(\mathbb{R}^3)} \leq C\big(1 + \|\rho\|_{L^1(\mathbb{R}^3)} + \|\rho\|_{L^\infty(\mathbb{R}^3)} \big[1 + \log(1 + \|\nabla_x \rho\|_{L^\infty(\mathbb{R}^3)})\big]\big). \tag{59}$$

This kind of estimate was already crucial in [17]; for the sake of completeness, we recall the proof in the appendix. We start by rewriting the collision and recombination-generation operators as

$$Q_i(f_i) = a_i[f_i](1 - f_i) - b_i[f_i] f_i, \quad i = n, p,$$

and

$$Q_{i,r}(f_i, f_{tr}) = g_i[f_{tr}](1 - f_i) - r_i[f_{tr}] f_i, \quad i = n, p,$$

with

$$\begin{aligned} a_i[f_i] &= \int_B \widetilde{\Phi}_i M_i f_i'\, dk', \quad b_i[f_i] = \int_B \widetilde{\Phi}_i M_i' (1 - f_i')\, dk', \quad i = n, p, \\ g_n[f_{tr}] &= \int_0^1 \Phi_n n_0 M_n f_{tr} M_{tr}\, dE, \quad g_p[f_{tr}] = \int_0^1 \Phi_p p_0 M_p (1 - f_{tr}) M_{tr}\, dE, \\ r_n[f_{tr}] &= \int_0^1 \Phi_n (1 - f_{tr}) M_{tr}\, dE, \quad r_p[f_{tr}] = \int_0^1 \Phi_p f_{tr} M_{tr}\, dE. \end{aligned}$$

In order to construct an approximating sequence $(f_n^j, f_p^j, f_{tr}^j, \mathcal{E}^j)$ we begin with

$$f_i^0(x,k,t) = f_{i,I}(x,k), \quad i = n,p, \quad f_{tr}^0(x,E,t) = f_{tr,I}(x,E). \tag{60}$$

The field always satisfies

$$\mathcal{E}^j(x,t) = \int_{\mathbb{R}^3} \frac{x-y}{|x-y|^3} \rho^j(y,t)\, dy. \tag{61}$$

Let $(f_n^j, f_p^j, f_{tr}^j, \mathcal{E}^j)$ be given. Then the ${f_i}^{j+1}$ are defined as the solutions of the following problem:

(62)
$$\begin{aligned}
&\alpha^2 \partial_t f_n^{j+1} + \alpha v_n(k) \cdot \nabla_x f_n^{j+1} - \alpha \mathcal{E}^j \cdot \nabla_k f_n^{j+1} \\
&\qquad = (a_n[f_n^j] + \alpha^2 g_n[f_{tr}^j])(1 - f_n^{j+1}) - (b_n[f_n^j] + \alpha^2 r_n[f_{tr}^j]) f_n^{j+1}, \\
&\alpha^2 \partial_t f_p^{j+1} + \alpha v_p(k) \cdot \nabla_x f_p^{j+1} + \alpha \mathcal{E}^j \cdot \nabla_k f_p^{j+1} \\
&\qquad = (a_p[f_p^j] + \alpha^2 g_p[f_{tr}^j])(1 - f_p^{j+1}) - (b_p[f_p^j] + \alpha^2 r_p[f_{tr}^j]) f_p^{j+1}, \\
&\varepsilon \partial_t f_{tr}^{j+1} = (p_0 \lambda_p[M_p(1 - f_p^j)] + \lambda_n[f_n^j])(1 - f_{tr}^{j+1}) - (n_0 \lambda_n[M_n(1 - f_n^j)] + \lambda_p[f_p^j]) f_{tr}^{j+1},
\end{aligned}$$

subject to the initial conditions

(63)
$$f_n^{j+1}(x,k,0) = f_{n,I}(x,k), \quad f_p^{j+1}(x,k,0) = f_{p,I}(x,k), \quad f_{tr}^{j+1}(x,E,0) = f_{tr,I}(x,E).$$

For the iterative sequence we state the following lemma, which is very similar to Proposition 3.1 from [13].

LEMMA 5.1. *Let the assumptions* (52)–(56) *be satisfied. Then the sequence* $(f_n^j, f_p^j, f_{tr}^j, \mathcal{E}^j)$ *defined by* (60)–(63) *satisfies the following for any time* $T > 0$:

(a) $0 \le {f_i}^j \le 1$, $i = n,p,tr$.

(b) f_n^j *and* f_p^j *are uniformly bounded with respect to* $j \to \infty$ *and* $\varepsilon \to 0$ *in* $L^\infty(0,T; L^1(\mathbb{R}^3 \times B))$.

(c) $\mathcal{E}^j$ *is uniformly bounded with respect to* j *and* ε *in* $L^\infty((0,T) \times \mathbb{R}^3)$.

Proof. The first two equations in (62) are standard linear transport equations, and the third equation is a linear ODE. Existence and uniqueness for the initial value problems is therefore a standard result.

Note that the a_i, b_i, g_i, r_i, and λ_i in (62) are nonnegative if we assume that (a) holds for j. Then (a) for $j+1$ is an immediate consequence of the maximum principle.

To estimate the L^1-norms of the distributions, we integrate the first equation in (62) and obtain

$$\|f_n^{j+1}\|_{L^1(\mathbb{R}^3 \times B)} \le \|f_{n,I}\|_{L^1(\mathbb{R}^3 \times B)} + \int_0^t \left\| a_n[f_n^j] \frac{1}{\alpha^2} + g_n[f_{tr}^j] \right\|_{L^1(\mathbb{R}^3 \times B)} (s)\, ds. \tag{64}$$

The boundedness of $\widetilde{\Phi}_n$, Φ_n, and f_{tr}^j and the integrability of M_{tr} imply

$$\|a_n[f_n^j]\|_{L^1(\mathbb{R}^3 \times B)} \le C \|f_n^j\|_{L^1(\mathbb{R}^3 \times B)}, \quad \|g_n[f_{tr}^j]\|_{L^1(\mathbb{R}^3 \times B)} \le C. \tag{65}$$

This is now used in (64). Then an estimate is derived for f_n^j by replacing $j+1$ by j and using the Gronwall inequality. Finally, it is easily seen that this estimate is passed from j to $j+1$ by (64). An analogous argument for f_p^j completes the proof of (b).

A uniform-in-ε ($L^1 \cap L^\infty$)-bound for the total charge density $\rho^j = n^j + \varepsilon n_{tr}^j - p^j - C$ follows from (b) and from the integrability of M_{tr}. Statement (c) of the lemma is now a consequence of (58). ∎

For passing to the limit in the nonlinear terms some compactness is needed. Therefore we prove uniform smoothness of the approximating sequence.

LEMMA 5.2. *Let the assumptions* (52)–(57) *be satisfied. Then for any time* $T > 0$ *the following hold:*

(a) f_n^j *and* f_p^j *are uniformly bounded with respect to* j *and* ε *in* $L^\infty(0,T; W^{1,1}_{per}(\mathbb{R}^3 \times B) \cap W^{1,\infty}_{per}(\mathbb{R}^3 \times B))$.

(b) f_{tr}^j *is uniformly bounded with respect to* j *and* ε *in* $L^\infty(0,T; W^{1,\infty}(\mathbb{R}^3 \times (0,1)))$.

(c) $\mathcal{E}^j$ *is uniformly bounded with respect to* j *and* ε *in* $L^\infty(0,T; W^{1,\infty}(\mathbb{R}^3))$.

Proof. We start by introducing $\nu^j = \nabla_{x,k} f_n^j = (\nu_x^j, \nu_k^j)$, $\pi^j = \nabla_{x,k} f_p^j = (\pi_x^j, \pi_k^j)$, $\phi^j = \nabla_x f_{tr}^j$, and by differentiating the last equation in (62) with respect to x:

$$\begin{aligned}\varepsilon \partial_t \phi^{j+1} = {} & (-p_0 \lambda_p[M_p \pi_x^j] + \lambda_n[\nu_x^j])(1 - f_{tr}^{j+1}) - (-n_0 \lambda_n[M_n \nu_x^j] + \lambda_p[\pi_x^j]) f_{tr}^{j+1} \\ & - (p_0 \lambda_p[M_p(1 - f_p^j)] + \lambda_n[f_n^j] + n_0 \lambda_n[M_n(1 - f_n^j)] + \lambda_p[f_p^j])\phi^{j+1}.\end{aligned}$$

The coefficient of ϕ^{j+1} on the right-hand side is bounded from below by the term appearing in assumption (57) and, thus, bounded away from zero. The maximum principle implies

$$\sup_{(0,t)} \|\phi^{j+1}\|_\infty \le C \left(\sup_{(0,t)} \|\nu_x^j\|_\infty + \sup_{(0,t)} \|\pi_x^j\|_\infty + 1 \right),$$

where here and in the following we use the symbol $\|\cdot\|_\infty$ for the L^∞-norm on $\mathbb{R}^3$, on $\mathbb{R}^3 \times B$, and on $\mathbb{R}^3 \times (0,1)$. The gradient of the first equation in (62) with respect to x and k can be written as

$$\alpha^2 \partial_t \nu^{j+1} + \alpha v_n \cdot \nabla_x \nu^{j+1} - \alpha \mathcal{E}^j \cdot \nabla_k \nu^{j+1} + (a_n + b_n + \alpha^2 g_n + \alpha^2 r_n)\nu^{j+1} = S_n^j,$$

where it is easily seen that, using our assumptions,

$$\|S_n^j\|_\infty \le C \left(1 + \|\nu^j\|_\infty + \|\phi^j\|_\infty + \|\nu^{j+1}\|_\infty (1 + \|\nabla_x \mathcal{E}^j\|_\infty)\right)$$

holds. The analogous treatment of the second equation in (62), the potential theory inequality (59), and the definition

$$\gamma^j(t) = \sup_{(0,t)} (\|\nu^j\|_\infty + \|\pi^j\|_\infty + \|\phi^j\|_\infty)$$

lead to

$$\gamma^{j+1} \le C \left(1 + \int_0^t (\gamma^j + \gamma^{j+1}(1 + \log(1 + \gamma^j)))\, ds \right),$$

implying boundedness of γ^j on arbitrary bounded time intervals (as in [4]). This proves (c) and the L^∞-part of (a). The equation for $\partial_E f_{tr}^{j+1}$ can be treated as above, completing the proof of (b).

By $\int_{\mathbb{R}^3} n_{tr}\, dx \le \int_{\mathbb{R}^3} M_{tr}\, dx$, it is trivial that the total number of trapped electrons is bounded. Therefore, the L^1-estimates in (a) follow the line of [13] since no coupling with the equation for the trapped electrons is necessary. ∎

With the previous results, the first two equations in (62) also give uniform bounds for the time derivatives of f_n^j and f_p^j. Thus, subsequences converge strongly locally in x and t. In the same way, the right-hand side of the time derivative of the Poisson equation is bounded in L^1 and in L^∞, and (58) implies boundedness of the time derivative of the field. So the field also converges strongly. This and the (obvious) weak convergence of f_{tr}^j are sufficient for passing to the limit in the quadratic nonlinearities. Note also that we have enough bounds on the time derivative to define the trace at time $t = 0$. Existence of a global solution of Problem (K) follows. By the same argument, the limit $\varepsilon \to 0$ can be justified, since all estimates are also uniform in ε.

THEOREM 5.3. *Let the assumptions* (52)–(57) *be satisfied. Then Problem* (K) *has a global solution* $(f_n, f_p, f_{tr}, \mathcal{E})$ *with* $f_n, f_p \in L^\infty(0,T; W^{1,\infty}_{per}(\mathbb{R}^3 \times B))$, $f_{tr} \in L^\infty(0,T; W^{1,\infty}(\mathbb{R}^3 \times (0,1)))$, $\mathcal{E} \in L^\infty(0,T; W^{1,\infty}(\mathbb{R}^3))$. *For* $\varepsilon \to 0$, *a subsequence of solutions converges to the formal limit problem. The convergence of* f_n *and* f_p *is in* $L^\infty((0,\infty) \times \mathbb{R}^3 \times B)$, *that of* $\mathcal{E}$ *in* $L^\infty((0,\infty) \times \mathbb{R}^3)$, *and that of* f_{tr} *in* $L^\infty((0,\infty) \times \mathbb{R}^3 \times (0,1))$ *weak*.*

6. Relation between macroscopic and kinetic models. In this section the relation between the two models in sections 2 and 4 is clarified on a formal level. The drift-diffusion model of section 2 can be derived from the kinetic model of section 4 by two simplification steps: a macroscopic and a low density limit.

Starting with the macroscopic limit, i.e., the limit when the Knudsen number α tends to zero in (42), (43), the solutions are expanded in terms of powers of α:

$$f_n = f_n^0 + \alpha f_n^1 + \mathcal{O}(\alpha^2), \qquad f_p = f_p^0 + \alpha f_p^1 + \mathcal{O}(\alpha^2), \tag{66}$$

$$f_{tr} = f_{tr}^0 + \mathcal{O}(\alpha), \qquad V = V^0 + \mathcal{O}(\alpha). \tag{67}$$

The limit of (42), (43) as $\alpha \to 0$ leads to $Q_n(f_n^0) = Q_p(f_p^0) = 0$. With the (frequently used) simplifying assumption that the cross sections $\widetilde{\Phi}_n$ and $\widetilde{\Phi}_p$ are strictly positive, the limiting distributions are of Fermi–Dirac type (see [13]):

$$f_n^0(x,k,t) = \frac{1}{1 + e^{-\mu_n(x,t)}/M_n(k)}, \qquad f_p^0(x,k,t) = \frac{1}{1 + e^{\mu_p(x,t)}/M_p(k)},$$

where the scaled Maxwellians M_n, M_p are given by (46) and the chemical potentials μ_n and μ_p are yet to be specified. Note the one-to-one relations between the chemical potentials and the macroscopic electron and hole densities:

$$n(\mu_n) = \int_B \frac{dk}{1 + e^{-\mu_n}/M_n(k)}, \qquad p(\mu_p) = \int_B \frac{dk}{1 + e^{\mu_p}/M_p(k)}.$$

Now (42) is divided by α, and then again the limit $\alpha \to 0$ is carried out (formally):

$$v_n \cdot \nabla_x f_n^0 + \nabla_x V^0 \cdot \nabla_k f_n^0 = LQ_n(f_n^0) f_n^1, \tag{68}$$

where LQ_n is the linearization of Q_n:

$$LQ_n(f_n^0) f_n^1 = \int_B \widetilde{\Phi}_n \big[(M_n(1 - f_n^0) + M_n' f_n^0) f_n^{1'} - (M_n f_n^{0'} + M_n'(1 - f_n^{0'})) f_n^1 \big] \, dk'.$$

For the following we shall need two facts about the linearized collision operator $LQ_n(f_n^0)$ (see, e.g., [1]): It has a one-dimensional kernel spanned by $f_n^0(1 - f_n^0)$, and

its range consists of functions whose integral with respect to k vanishes. Therefore, for solvability of (68), seen as an equation for f_n^1, the integral with respect to k of the left-hand side has to vanish. This is obvious for the second term $\nabla_x V^0 \cdot \nabla_k f_n^0$ by the periodicity with respect to k. Since the first term can be written as

$$v_n \cdot \nabla_x f_n^0 = \nabla_k \varepsilon_n \cdot \nabla_x \frac{M_n}{M_n + e^{-\mu_n}} = -\nabla_k \cdot \nabla_x \log\left(M_n + e^{-\mu_n}\right),$$

it also satisfies the solvability condition. Now (68) is written as

$$\frac{M_n e^{-\mu_n}}{(M_n + e^{-\mu_n})^2} \nabla_k \varepsilon_n \cdot (\nabla_x V^0 - \nabla_x \mu_n) = LQ_n(f_n^0) f_n^1. \tag{69}$$

Note that the factor in parentheses is independent of k. Thus, choosing a solution $h_n(k, \mu_n)$ of

$$LQ_n(f_n^0) h_n = -\frac{M_n e^{-\mu_n}}{(M_n + e^{-\mu_n})^2} \nabla_k \varepsilon_n, \tag{70}$$

the solution of (69) can be written as

$$f_n^1 = h_n(k, \mu_n) \cdot (\nabla_x V^0 - \nabla_x \mu_n) + \mu_n^1 f_n^0 (1 - f_n^0).$$

Analogously,

$$f_p^1 = h_p(k, \mu_p) \cdot (\nabla_x V^0 + \nabla_x \mu_p) + \mu_p^1 f_p^0 (1 - f_p^0) \tag{71}$$

is obtained (with $\mu_n^1(x,t)$ and $\mu_p^1(x,t)$ not specified, and not needed in the following). Finally, (42), (43) are divided by α^2 and integrated with respect to k, and the limit $\alpha \to 0$ is carried out:

$$\partial_t n + \nabla_x \cdot \int_B v_n f_n^1 \, dk = \int_B Q_{n,r}(f_n^0, f_{tr}^0) \, dk = \int_0^1 S_n(f_n^0, f_{tr}^0) \, dE, \tag{72}$$

$$\partial_t p + \nabla_x \cdot \int_B v_p f_p^1 \, dk = \int_B Q_{p,r}(f_p^0, f_{tr}^0) \, dk = \int_0^1 S_p(f_p^0, f_{tr}^0) \, dE. \tag{73}$$

With the formulas for f_n^1 and f_p^1, we obtain the drift-diffusion fluxes

$$\int_B v_n f_n^1 \, dk = D_n(\mu_n)(\nabla_x V^0 - \nabla_x \mu_n), \qquad \int_B v_p f_p^1 \, dk = D_p(\mu_p)(\nabla_x V^0 + \nabla_x \mu_p),$$

with the diffusion matrices

$$D_n = \int_B v_n \otimes h_n \, dk, \qquad D_p = \int_B v_p \otimes h_p \, dk.$$

For the recombination-generation terms, we obtain

$$S_n(f_n^0, f_{tr}^0) = \lambda_n \left[\frac{e^{-\mu_n}}{1 + e^{-\mu_n}/M_n}\right] (n_0 f_{tr}^0 - e^{\mu_n}(1 - f_{tr}^0)),$$

$$S_p(f_p^0, f_{tr}^0) = \lambda_p \left[\frac{e^{\mu_p}}{1 + e^{\mu_p}/M_p}\right] (p_0(1 - f_{tr}^0) - e^{-\mu_p} f_{tr}^0).$$

Finally, we consider the small densities situation, when μ_n is large and negative and μ_p large and positive. This gives $n(\mu_n) \approx e^{\mu_n}$ and $p(\mu_p) \approx e^{-\mu_p}$. The above recombination-generation terms can then be approximated by the terms in (9) with $1/\tau_n = \lambda_n[M_n]$ and $1/\tau_p = \lambda_p[M_p]$.

Equation (70) for h_n can be approximated by

$$\int_B \widetilde{\Phi}_n \big[M_n h_n' - M_n' h_n\big]\, dk' = -nM_n \nabla_k \varepsilon_n,$$

implying $h_n = n\tilde{h}_n(k)$ and, thus, $D_n = n\tilde{D}_n$. With this and the analogous approximation for holes, the macroscopic model becomes the drift-diffusion model from section 2.

It is worth pointing out that the drift-diffusion SRH model has been obtained from the kinetic model by a two-step approximation procedure: At first, the hydrodynamic limit leads to a more nonlinear system, and we perform additionally the small densities asymptotics. This remark appeals to further mathematical questions:

• It could be interesting to investigate the intermediate macroscopic model that comes directly from the Fermi–Dirac statistics.

• It could be tempting to reverse the limits. Roughly speaking, it means that we do not take into account the Pauli exclusion principle in the kinetic equations, and the collision operator is replaced by a linear Boltzmann operator which relaxes to a Maxwellian (instead of a Fermi–Dirac distribution). Mathematically, this leads to additional difficulties since we lose the natural L^∞-estimate given for free with the exclusion terms. Rigorous derivation of the diffusion regime for the corresponding Boltzmann–Poisson system in a bounded domain, with only one species of charged particles, has been obtained only very recently by using a tricky renormalization argument; see [11] (and [12] for an earlier work on renormalized solutions).

Appendix. Proof of (58) and (59). We recall that the fundamental solution of $-\Delta$ in $\mathbb{R}^N$, $N \geq 3$, reads $E(x) = C_N|x|^{2-N}$. For a given function $\rho : \mathbb{R}^N \to \mathbb{R}^+$, we set

$$\Phi = E * \rho, \qquad \nabla_x \Phi(x) = C_N(2-N) \int_{\mathbb{R}^N} \frac{x-y}{|x-y|}\, \frac{\rho(y)}{|x-y|^{N-1}}\, dy.$$

For any $0 < R < \infty$, we have

$$\int_{\mathbb{R}^N} \frac{\rho(y)}{|x-y|^{N-1}}\, dy = \int_{|x-y|\leq R} \dots\, dy + \int_{|x-y|\geq R} \dots\, dy \leq \|\rho\|_\infty \frac{\Omega_N R^2}{2} + \frac{1}{R^{N-1}} \|\rho\|_1,$$

where Ω_N stands for the surface of the N-dimensional sphere. Optimizing with respect to R yields

$$\|\nabla \Phi\|_\infty \leq K_N \|\rho\|_\infty^{(N-1)/(N+1)} \|\rho\|_1^{2/(N+1)},$$

where K_N is the constant depending only on the dimension.

Since $|x|^{N-1}$ is locally integrable, we compute the second derivatives of the potential as follows. For any $\varphi \in C_c^\infty(\mathbb{R}^N)$, we have

$$\big\langle \partial_{ij}^2 \Phi; \varphi \big\rangle = C_N(2-N) \int_{\mathbb{R}^N} \frac{x_j}{|x|^N}\, \partial_j \varphi(x)\, dx = C_N(2-N) \lim_{\eta \to 0} \int_{|x|\geq \eta} \frac{x_j}{|x|^N}\, \partial_j \varphi(x)\, dx$$

$$= C_N(N-2) \lim_{\eta \to 0} \left(\int_{|x|\geq \eta} \left(\frac{\delta_{ij}}{|x|^N} - N \frac{x_i x_j}{|x|^{N+2}} \right) \varphi(x)\, dx + \int_{|x|=\eta} \frac{x_j}{|x|^N} \frac{x_i}{|x|} \varphi(x)\, d\sigma(x) \right).$$

The second integral in the right-hand side can be recast as

$$\int_{\mathbb{S}^{N-1}} \varphi(\eta\omega)\omega_i\omega_j \, d\omega,$$

and therefore it converges to

$$\frac{\Omega_N}{N}\delta_{ij}\,\varphi(0).$$

Let us introduce the matrix

$$\mathbb{K}_{ij}(x) = C_N \frac{N-2}{|x|^N}\left(\delta_{ij} - N\frac{x_i x_j}{|x|^2}\right).$$

Then, in the sense of distribution the Hessian matrix of E satisfies

$$D^2E(x) = C_N(2-N)\frac{\Omega_N}{N}\mathbb{I}\,\delta(x=0) + \lim_{\eta\to 0}\mathbb{K}(x)\chi_{|x|\geq\eta}.$$

In particular we remark that

$$\int_{\mathbb{S}^{N-1}} \mathbb{K}(r\omega)\, d\omega = 0, \qquad \operatorname{Tr}\mathbb{K} = 0. \tag{74}$$

Accordingly, the Hessian matrix of the potential Φ is given by

$$D^2\Phi(x) = C_N(2-N)\frac{\Omega_N}{N}\mathbb{I}\,\rho(x) + \lim_{\eta\to 0}\int_{|x-y|\geq\eta}\mathbb{K}(x-y)\rho(y)\,dy.$$

Let us discuss the last term. Consider $0<\eta<R_1<R_2<\infty$. Using the notation C to stand for any constant depending only on the dimension, we get

$$\begin{aligned}\left|\int_{|x-y|\geq\eta}\mathbb{K}(x-y)\rho(y)\,dy\right| &\leq C\left(\int_{|x-y|\geq R_2}\frac{\rho(y)}{|x-y|^N}\,dy + \int_{R_1\leq|x-y|\leq R_2}\frac{\rho(y)}{|x-y|^N}\,dy \right.\\ &\qquad \left. + \left|\int_{\eta\leq|x-y|\leq R_1}\mathbb{K}(x-y)\rho(y)\,dy\right|\right)\\ &\leq C\frac{1}{R_2^N}\|\rho\|_1 + C\|\rho\|_\infty \ln\left(\frac{R_2}{R_1}\right) + C\left|\int_{\eta\leq|x-y|\leq R_1}\mathbb{K}(x-y)\big(\rho(y)-\rho(x)\big)\,dy\right|\end{aligned}$$

where we used (74). We deduce the following estimate:

$$\left|\int_{|x-y|\geq\eta}\mathbb{K}(x-y)\rho(y)\,dy\right| \leq C\left(\frac{1}{R_2^N}\|\rho\|_1 + \|\rho\|_\infty\ln\left(\frac{R_2}{R_1}\right) + \|\nabla_x\rho\|_\infty R_1\right).$$

Then, we choose $R_2 = 1 > R_1 = (1+\|\nabla_x\rho\|_\infty)^{-1}$ and conclude that

$$|D^2\Phi(x)| \leq C\Big(1 + \|\rho\|_1 + \|\rho\|_\infty\big(1+\ln(1+\|\nabla_x\rho\|_\infty)\big)\Big)$$

holds. A similar analysis can be done in dimension two; see [17].

Acknowledgment. Part of this work was carried out while the second and third authors enjoyed the hospitality of the Université des Sciences et Technologies Lille 1.

REFERENCES

[1] N. Ben Abdallah and P. Degond, *On a hierarchy of macroscopic models for semiconductors*, J. Math. Phys., 37 (1996), pp. 3306–3333.

[2] I. Choquet, P. Degond, and C. Schmeiser, *Energy-transport models for charge carriers involving impact ionization in semiconductors*, Transport Theory Statist. Phys., 32 (2003), pp. 99–132.

[3] I. Choquet, P. Degond, and C. Schmeiser, *Hydrodynamic model for charge carriers involving strong ionization in semiconductors*, Commun. Math. Sci., 1 (2003), pp. 74–86.

[4] P. Degond, *Global existence of smooth solutions for the Vlasov-Fokker-Planck equation in* 1 *and* 2 *space dimensions*, Ann. Sci. École Norm. Sup. (4), 19 (1986), pp. 519–542.

[5] H. Gajewski, *On existence, uniqueness, and asymptotic behaviour of solutions of the basic equations for carrier transport in semiconductors*, Z. Angew. Math. Mech., 65 (1985), pp. 101–108.

[6] I. Gasser, C. D. Levermore, P. A. Markowich, and C. Schmeiser, *The initial time layer problem and the quasineutral limit in the semiconductor drift-diffusion model*, European J. Appl. Math., 12 (2001), pp. 497–512.

[7] P. González, A. Godoy, F. Gámiz, and J. A. Carrillo, *Accurate deterministic numerical simulation of p-n junctions*, J. Comput. Electron., 3 (2004), pp. 235–238.

[8] P. González, J. A. Carrillo, and F. Gámiz, *Deterministic numerical simulation of* $1d$ *kinetic descriptions of bipolar electron devices*, in Scientific Computing in Electrical Engineering, Math. Ind. 9, Springer, Berlin, Heidelberg, 2006, pp. 339–344.

[9] R. N. Hall, *Electron-hole recombination in Germanium*, Phys. Rev., 87 (1952), p. 387.

[10] P. A. Markowich, C. A. Ringhofer, and C. Schmeiser, *Semiconductor Equations*, Springer, Vienna, New York, 1990.

[11] N. Masmoudi and M. L. Tayeb, *Diffusion limit of a semiconductor Boltzmann–Poisson system*, SIAM J. Math. Anal., 38 (2007), pp. 1788–1807.

[12] S. Mischler, *On the initial boundary value problem for the Vlasov-Poisson-Boltzmann system*, Comm. Math. Phys., 210 (2000), pp. 447–466.

[13] F. Poupaud, *On a system of nonlinear Boltzmann equations of semiconductor physics*, SIAM J. Appl. Math., 50 (1990), pp. 1593–1606.

[14] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer, Vienna, New York, 1984.

[15] W. Shockley and W. T. Read, *Statistics of the recombinations of holes and electrons*, Phys. Rev., 87 (1952), pp. 835–842.

[16] J. Simon, *Compact sets in the space* $L^p(0,T;B)$, Ann. Mat. Pura Appl. (4), 146 (1987), pp. 65–96.

[17] S. Ukai and T. Okabe, *On the classical solution in the large time of the two dimensional Vlasov equation*, Osaka J. Math., 15 (1978), pp. 245–261.