Degeneracy and bandgap narrowing in the semiconductor electron-hole product

Lachlan E. Black

Citation: Journal of Applied Physics 121, 105701 (2017); doi: 10.1063/1.4977200
View online: http://dx.doi.org/10.1063/1.4977200
View Table of Contents: http://aip.scitation.org/toc/jap/121/10
Published by the American Institute of Physics

Articles you may be interested in

Resonant optical properties of AlGaAs/GaAs multiple-quantum-well based Bragg structure at the second quantum state
Journal of Applied Physics 121, 103101 (2017); 10.1063/1.4978252

Photoluminescence enhancement of ZnO via coupling with surface plasmons on Al thin films
Journal of Applied Physics 121, 103104 (2017); 10.1063/1.4977954

Carrier capture in InGaN/GaN quantum wells: Role of electron-electron scattering
Journal of Applied Physics 121, 123107 (2017); 10.1063/1.4979010

Wealth inequality: The physics basis
Journal of Applied Physics 121, 124903 (2017); 10.1063/1.4977962

Mechanism of yellow luminescence in GaN at room temperature
Journal of Applied Physics 121, 065104 (2017); 10.1063/1.4975116

Theoretical study of time-resolved luminescence in semiconductors. IV. Lateral inhomogeneities
Journal of Applied Physics 121, 085103 (2017); 10.1063/1.4976102
Degeneracy and bandgap narrowing in the semiconductor electron-hole product

Lachlan E. Black

Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 30 October 2016; accepted 10 February 2017; published online 10 March 2017)

It is shown that previously proposed expressions for the semiconductor electron-hole product, which purport to separate the influence of carrier degeneracy and bandgap narrowing, fail to properly delineate these effects. A reformulation of the expression in which the two effects are successfully separated is proposed, valid when the majority carrier concentration is independent of bandgap narrowing, as occurs in the common case of low injection and quasi-neutrality. It is shown that under other conditions, the two effects may not be treated in isolation but that only their combined effect or the marginal effect of one or the other on the total electron-hole product may be assessed. Convenient expressions are provided for the latter case. The revised expression provides both conceptual and computational advantages for semiconductor device modelling. Published by AIP Publishing. [http://dx.doi.org/10.1063/1.4977200]

I. INTRODUCTION

In non-degenerate intrinsic or non-heavily doped semiconductors in equilibrium, the free electron and hole concentrations, \( n \) and \( p \), are simply related by \( np = n_i^2 \), where \( n_i \) is the intrinsic carrier concentration. Under non-equilibrium conditions, or when carrier degeneracy and/or bandgap narrowing are significant, as occurs in heavily doped material, this relationship is no longer valid. However, we can continue to conveniently express the resulting \( np \) product in terms of \( n_i^2 \) by introducing additional multiplicative factors that account for these effects.

It was proposed by Altermatt et al.,\(^1\) based on an earlier treatment of the equilibrium case by Schenk,\(^2\) that when the minority carrier concentration is non-degenerate, the \( np \) product may be written in such a way that the effects of degeneracy, bandgap narrowing, and non-equilibrium conditions are separated into independent factors \( \gamma_{\text{deg}} \), \( \gamma_{\text{BGN}} \), and \( \gamma_{\text{neq}} \), such that

\[
np \approx n_i^2 \times \gamma_{\text{deg}} \times \gamma_{\text{BGN}} \times \gamma_{\text{neq}}. \tag{1}
\]

Such a division is clearly advantageous from the point of view of promoting conceptual understanding. The expression of Ref. 1 is employed for example, in the software program EDNA, which is widely used to simulate heavily doped regions in silicon solar cells.\(^3\)

In this letter, it is shown that the division of terms made by Refs. 1 and 2 fails to properly delineate the effects of degeneracy and bandgap narrowing. A reformulation of the expression for the \( np \) product that avoids this problem in the common case of quasi-neutrality and low injection is proposed, and the conditions under which such a division is or is not possible are described.

II. DERIVATION OF THE \( np \) PRODUCT

We begin by describing the derivation of the expression for the non-equilibrium \( np \) product including degeneracy and bandgap narrowing. In order to ease comparison with Ref. 1, we take the case of an \( n \)-type semiconductor with net donor concentration \( N_D \). Analogous expressions apply in the case of \( p \)-type material. It is assumed that the minority hole concentration is non-degenerate, such that it can be described by Boltzmann statistics. This will be the case in heavily doped material except under very high injection. It is further assumed that a rigid-band approximation is applicable (i.e., \( N_c \) and \( N_v \) are independent of bandgap narrowing) and that the band structure is parabolic. The latter merely determines the form of the Fermi integrals and is non-essential. The majority electron concentration is then given by

\[
n = N_e F_{1/2} (\tilde{E}_{Fn} - \tilde{E}_{e0} + \Delta \tilde{E}_e), \tag{2}
\]

and the minority hole concentration by

\[
p = N_v \exp (\tilde{E}_{fp} - \tilde{E}_{e0} + \Delta \tilde{E}_e), \tag{3}
\]

where \( N_c \) and \( N_v \) are the effective densities of states of the conduction and valence bands, \( F_{1/2} \) is the Fermi integral of order 1/2, \( \tilde{E}_{Fn} \) and \( \tilde{E}_{fp} \) are the electron and hole quasi-Fermi energies, \( \tilde{E}_{e0} \) and \( \tilde{E}_{e0} \) are the intrinsic energies of the conduction and valence band edges, and \( \Delta \tilde{E}_e \) and \( \Delta \tilde{E}_v \) are the energy shifts of the conduction and valence band edges towards midgap due to bandgap narrowing, where the bar notation indicates that all these energies are normalised to \( kT \), where \( k \) is Boltzmann’s constant and \( T \) is the absolute temperature. All energies are defined relative to the intrinsic Fermi energy. Combining Eqs. (2) and (3), we have

\[
np = N_D N_e F_{1/2} (\tilde{E}_{Fn} - \tilde{E}_{e0} + \Delta \tilde{E}_e) \times \exp (\tilde{E}_{fp} - \tilde{E}_{e0} + \Delta \tilde{E}_e), \tag{4}
\]

which can be rewritten as

\[
pn = n_i^2 F_{1/2} (\tilde{E}_{Fn} - \tilde{E}_{e0} + \Delta \tilde{E}_e) \times \exp (\tilde{E}_{fp} - \tilde{E}_{e0} + \Delta \tilde{E}_e) \times \exp (\tilde{E}_{e0} - \tilde{E}_{e0}), \tag{5}
\]
using the definition of \( n_i \)
\[
n_i^2 = N_e N_c \exp \left( \bar{E}_{c0} - \bar{E}_{c0} \right).
\]

### III. PREVIOUS DIVISION OF TERMS

In Eq. (5), the terms relating to degeneracy, bandgap narrowing, and non-equilibrium effects are mixed together, so that it is difficult to discern their relative influence on the resulting \( pn \) product. It was proposed in Ref. 1 that Eq. (5) may be rewritten in the form of Eq. (1), with \( \gamma_{\text{deg}} \) and \( \gamma_{\text{BGN}} \) defined as listed in Table I, and \( \gamma_{\text{neq}} = \exp \left( \bar{E}_{Fn} - \bar{E}_{Fp} \right) \).

\[ pn = n_i^2 \frac{F_{1/2}(\bar{E}_{Fn} - \bar{E}_{c0})}{\exp(\bar{E}_{Fn} - \bar{E}_{c0})} \times \frac{F_{1/2}(\bar{E}_{Fn} - \bar{E}_{c0} + \Delta E_c)}{\exp(\Delta E_c)} \times \exp(\bar{E}_{Fn} - \bar{E}_{Fp}). \]

The advantage to be expected from expressing the \( pn \) product in this way is primarily conceptual. The effects of degeneracy, bandgap narrowing, and non-equilibrium conditions are intended to be divided into separate, independent terms, thereby making their relative influence on the \( pn \) product more readily apparent.

Unfortunately, the separation of terms embodied in Eq. (7) fails to achieve this aim, for reasons that shall now be discussed.

First, it is apparent that the inclusion of the Fermi integral terms in \( \gamma_{\text{BGN}} \) of Eq. (7) necessarily makes this term dependent on the degree of majority carrier degeneracy. To see this clearly, consider that to be independent of degeneracy, \( \gamma_{\text{BGN}} \) should be independent of \( n/N_e = F_{1/2}(\bar{E}_{Fn} - \bar{E}_{c0} + \Delta E_c) \), which is equal to the upper Fermi integral in \( \gamma_{\text{BGN}} \) of Eq. (7). This will occur only in the trivial case that either \( \Delta E_c = 0 \), or \( n \) is non-degenerate, such that the Fermi integrals may be replaced by exponential functions (we note that in either case, \( \gamma_{\text{BGN}} \) reduces to the alternative form, which we shall propose in Eq. (8)).

What is perhaps less readily apparent is that \( \gamma_{\text{deg}} \) of Eq. (7) is also dependent on the bandgap narrowing, despite the absence of explicit bandgap-narrowing terms, through the implicit dependence of \( \bar{E}_{Fn} \) on \( \Delta E_c \). Specifically, \( \gamma_{\text{deg}} \) is independent of \( \Delta E_c \) only when \( n \) is non-degenerate (in which case \( \gamma_{\text{deg}} = 1 \)), or when \( \bar{E}_{Fn} = \bar{E}_{F0} \), where \( \bar{E}_{F0} \) is the normalised electron quasi-Fermi energy in the absence of bandgap-narrowing (in other words, when \( \bar{E}_{Fn} \), defined with respect to the intrinsic Fermi energy, is independent of bandgap narrowing). In Ref. 1, Eq. (7) is explicitly applied to the case of the quasi-neutral diffused region in a silicon solar cell, for which its definition of both \( \gamma_{\text{deg}} \) and \( \gamma_{\text{BGN}} \) is certainly invalid.

### IV. PROPOSED ALTERNATIVE DIVISION

It is possible to make a different division of terms in Eq. (5) in order to arrive at an expression in the form of Eq. (1), which is valid in cases of practical interest. However, in order to maintain generality, it is necessary to introduce an additional term \( \gamma_{\text{deg} \times \text{BGN}} \), which accounts for interactions between degeneracy and bandgap narrowing which therefore depends on both effects. We propose the following expression, in which \( \gamma_{\text{deg}}, \gamma_{\text{BGN}}, \) and \( \gamma_{\text{deg} \times \text{BGN}} \) are defined as listed in Table I, and \( \gamma_{\text{neq}} \) is the same as in Eq. (7).

\[
\begin{align*}
pn &= n_i^2 \frac{F_{1/2}(\bar{E}_{F0} - \bar{E}_{c0})}{\exp(\bar{E}_{F0} - \bar{E}_{c0})} \times \exp(\Delta E_c + \Delta E_c) \\
& \quad \times \frac{F_{1/2}(\bar{E}_{F0} - \bar{E}_{c0} + \Delta E_c)}{\exp(\bar{E}_{F0} + \Delta E_c)} \\
& \quad \times \exp(\bar{E}_{F0} - \bar{E}_{Fp}) \\
&= n_i^2 \times \gamma_{\text{deg}} \times \gamma_{\text{BGN}} \times \gamma_{\text{deg} \times \text{BGN}} \times \gamma_{\text{neq}}.
\end{align*}
\]

Eqs. (7) and (8) are mathematically identical, but the separation of terms made between \( \gamma_{\text{deg}} \) and \( \gamma_{\text{BGN}} \) (and \( \gamma_{\text{deg} \times \text{BGN}} \) in the latter case) is different. It is obvious that \( \gamma_{\text{BGN}} \) of Eq. (8) depends only on the bandgap narrowing, unlike \( \gamma_{\text{BGN}} \) of Eq. (7), and that similarly \( \gamma_{\text{deg}} \) of Eq. (8) is independent of bandgap narrowing. However, this is achieved by splitting off the terms that are dependent on both degeneracy and bandgap narrowing into \( \gamma_{\text{deg} \times \text{BGN}} \). Only when \( \gamma_{\text{deg} \times \text{BGN}} = 1 \), therefore, is it possible to neglect the interaction between degeneracy and bandgap narrowing and consider the influence of the two effects on the \( pn \) product independently. As expected, this will occur in the trivial case that either bandgap narrowing is negligible (\( \Delta E_c = \Delta E_e = 0 \), \( \bar{E}_{Fn} = \bar{E}_{F0} \)), or \( n \) is non-degenerate, these cases corresponding to \( \gamma_{\text{BGN}} = 1 \) or \( \gamma_{\text{deg}} = 1 \), respectively.

The other, non-trivial case in which \( \gamma_{\text{deg} \times \text{BGN}} = 1 \) occurs when
\[
\bar{E}_{F0} = \bar{E}_{c0} + \Delta E_c = \bar{E}_{F0} - \bar{E}_{c0}.
\]

From Eq. (2), this implies that the majority carrier concentration is independent of bandgap narrowing. This is practically significant, since it corresponds to the common case of quasi-neutral low injection, which often prevails in heavily doped device regions. Consequently, under these conditions, Eq. (8) may be simplified to

**TABLE I.** Definitions of \( \gamma_{\text{deg}}, \gamma_{\text{BGN}}, \) and \( \gamma_{\text{deg} \times \text{BGN}} \) proposed by Altermatt et al.¹ in Eq. (7) and in Eq. (8) of the present work. Note that \( \gamma_{\text{deg} \times \text{BGN}} = 1 \) under conditions of quasi-neutrality and low injection.

<table>
<thead>
<tr>
<th>( \gamma_{\text{deg}} )</th>
<th>( \gamma_{\text{BGN}} )</th>
<th>( \gamma_{\text{deg} \times \text{BGN}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>This work</td>
<td>( F_{1/2}(\bar{E}<em>{F0} - \bar{E}</em>{c0}) ) ( F_{1/2}(\bar{E}<em>{F0} - \bar{E}</em>{c0} + \Delta E_c) ) ( F_{1/2}(\bar{E}<em>{F0} - \bar{E}</em>{c0} + \Delta E_c) )</td>
<td>( \exp(\Delta E_c + \Delta E_e) ) ( \exp(\bar{E}_{F0} + \Delta E_e) )</td>
</tr>
</tbody>
</table>
\[ pn = n_i^2 \frac{F_1(\Delta E_g)}{\exp(\Delta E_g)} \times \exp(\Delta E_g) \times \exp(\Delta E_g) \times \exp(\Delta E_g), \]

(10)

where \( \Delta E_g = \Delta E_c + \Delta E_v \) is the total reduction in bandgap energy due to bandgap narrowing, and the effects of degeneracy and bandgap narrowing on the \( pn \) product are independent. Note that, as expected, the expression for \( \gamma_{\text{BGN}} \) in Eq. (10) is the same as that for non-degenerate semiconductors. We note that an expression for the \( pn \) product featuring a similar division of terms to that in Eq. (10) was given by Marshak et al.\textsuperscript{4} although the physical significance of this division was not explicitly identified.

Significantly, Eq. (10) makes clear that under conditions of quasi-neutrality and low injection, \( \gamma_{\text{deg}} \), \( \gamma_{\text{BGN}} \), and the total \( pn \) product depend only on the total bandgap narrowing \( \Delta E_g \) and not on the distribution of this narrowing between the conduction and valence bands, even when degeneracy is significant. This is obscured by the arrangement of terms in Eq. (7). While this result has been noted before,\textsuperscript{1,5} it is worth reiterating as it is quite important, the scarcity of data on the degree of asymmetry in bandgap narrowing mirroring (not coincidentally) the fact that it is rarely relevant in practice.

To illustrate the advantage of Eqs. (8) or (10) over Eq. (7), in Fig. 1, \( \gamma_{\text{deg}} \) and \( \gamma_{\text{BGN}} \) calculated from both Eqs. (7) and (10) are plotted as a function of \( N_D \) for \( n \)-type silicon, under conditions of quasi-neutrality and low injection. The total bandgap narrowing \( E_g \) and the effective conduction band density of states \( N_c \) are treated as adjustable parameters in order to highlight the different dependence of the two definitions of \( \gamma_{\text{deg}} \) and \( \gamma_{\text{BGN}} \) on bandgap narrowing and degeneracy. Fig. 1 shows that changes in \( \Delta E_g \) affect both \( \gamma_{\text{deg}} \) and \( \gamma_{\text{BGN}} \) of Eq. (7), while \( \gamma_{\text{deg}} \) of Eq. (10) is independent of bandgap narrowing, as intended. Similarly, changes in \( N_c \) affect not only \( \gamma_{\text{deg}} \) but also \( \gamma_{\text{BGN}} \) of Eq. (7), while \( \gamma_{\text{BGN}} \) of Eq. (10) is independent of such changes. We note in passing that the values given for \( \gamma_{\text{deg}} \) and \( \gamma_{\text{BGN}} \) in Fig. 1 of Ref. 1 in fact appear to match those given by Eq. (10) rather than the expression given in that work.

It is worth pointing out that an expression for the \( pn \) product in which the influence of different physical mechanisms is contained in independent terms, as realised in Eq. (10), also has a computational advantage, in that it is not necessary to re-evaluate the entire expression when updating only one of the relevant parameters in simulation. Finally, we note that an additional practical advantage of Eq. (10) over Eq. (7) is that solution of the former only requires the evaluation of a single Fermi integral, compared to two in the case of the latter.

Eq. (10) allows the influence of degeneracy and bandgap narrowing on the \( pn \) product to be treated independently when the majority carrier concentration is independent of bandgap narrowing, as is the case in a quasi-neutral, lowly injected semiconductor. Under other conditions, it is not possible to treat the two effects in isolation when both are significant. In such cases, the best we can do is to evaluate the
marginal effects of either degeneracy or bandgap narrowing relative to the case in which only the other is present. This will be given by the product of either $\gamma_{\text{deg}}$ or $\gamma_{\text{BGN}}$ with $\gamma_{\text{deg} \times \text{BGN}}$, depending on what is taken to be the reference case. So the marginal influence of degeneracy when bandgap narrowing is already present is given by

$$\gamma_{\text{deg}} \gamma_{\text{deg} \times \text{BGN}} = \frac{F_{1/2}(\epsilon_F - \epsilon_c + \Delta \epsilon_c)}{\exp(\epsilon_F - \epsilon_c + \Delta \epsilon_c)},$$

(11)

while in the reverse case, the following expression applies:

$$\gamma_{\text{BGN}} \gamma_{\text{deg} \times \text{BGN}} = \frac{F_{1/2}(\epsilon_F - \epsilon_c + \Delta \epsilon_c)}{F_{1/2}(\epsilon_{F0} - \epsilon_c)} \times \exp(\epsilon_{F0} - \epsilon_F + \Delta \epsilon_c).$$

(12)

Note that in the case that $\epsilon_F = \epsilon_{F0}$, Eq. (12) reduces to $\gamma_{\text{BGN}}$ of the expression of Ref. 1 given in Eq. (7) and $\gamma_{\text{deg}}$ of Eq. (8) reduces to $\gamma_{\text{deg}}$ of Eq. (7) (recall that this was the condition under which $\gamma_{\text{deg}}$ of Eq. (7) was independent of bandgap narrowing).

Finally, it should be remembered that ultimately it is the total $pn$ product that matters for semiconductor device performance and that this is independent of the particular form chosen for its expression. The separation of the $pn$ product into $\gamma_{\text{deg}}$, $\gamma_{\text{BGN}}$, and $\gamma_{\text{neq}}$ factors does however provide a valuable conceptual (and sometimes computational) aid.

V. CONCLUSION

It was shown that previously proposed expressions\textsuperscript{1,2} for the $pn$ product in heavily doped semiconductors, which were intended to separate the influence of carrier degeneracy and bandgap narrowing, fail to properly delineate these effects. An alternative formulation was proposed that fulfils this requirement when the majority carrier concentration is independent of bandgap narrowing, as occurs in the common case of quasi-neutrality and low injection. It was shown that under other circumstances it is not possible to assess the influence of degeneracy and bandgap narrowing on the $pn$ product independently, but only their combined effect or the marginal influence of one or the other, and simplified expressions to evaluate the latter were provided. The revised expression provides a correct delineation of degeneracy and bandgap narrowing, thereby aiding conceptual understanding and computational implementation.

ACKNOWLEDGMENTS

The author gratefully acknowledges Keith R. McIntosh for stimulating discussions and for providing valuable feedback on draft versions of this manuscript, and Pietro Altermatt for his encouragement and insights.


