Design for enhanced thermo-electric pumping in light emitting diodes

Dodd J. Gray, Jr.,1, 2 Parthiban Santhanam,3 and Rajeev J. Ram1
1Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
2Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA

(Received 27 June 2013; accepted 13 August 2013; published online 17 September 2013)

We present a strategy for optimization of thermo-electric pumping in light emitting diodes (LEDs). We use a finite element model for charge transport in a GaInAsSb/GaSb double hetero-junction LED that is verified experimentally to consider optimal design and operation of low-bias LEDs. The wall-plug efficiency is shown to be enhanced by over 200% at nanowatt power levels and 20× at microwatt power levels. A design for room-temperature operation of a 2.2 μm LED with 100% efficiency is proposed—this represents a 110 °C reduction of the temperature required to observe unity efficiency. © 2013 AIP Publishing LLC. [http://dx.doi.org/10.1063/1.4821266]

A recent experiment1 has demonstrated a semiconductor light-emitting diode (LED) operating with total electrical-to-optical power conversion efficiency (i.e., wall-plug efficiency) η greater than 100%. This effect, known as electro-luminescent (EL) cooling,1–4 electroluminescent refrigeration,5–6 thermophotonic cooling,7 and optothermal cooling8–10 is possible when the work done per injected electron qV (where q is the magnitude of the electron’s charge and V is the applied voltage) is less than the energy hω carried away from the LED by each emitted photon. The wall-plug efficiency of an LED can be expressed as the product of hω/q and the external quantum efficiency ηQL, where ηQ is the number of photons generated per electron injected into the active region and ηQL is the number of photons extracted from the device per photon created in the active region. When the product η = hω/qηQL is greater than one, light power output greater than the electrical power input can be observed. This is allowed thermodynamically by virtue of the facts (1) that thermal energy is extracted from the lattice to produce output light and (2) that the incoherent light from the LED carries with it significant entropy in the form of random phase, polarization, wavelength, and direction. The energy transfer mechanism responsible for EL cooling is thermo-electric Peltier heat exchange between the semiconductor lattice and conducting charge carriers flowing across the diode’s p-i-n junction. Experimental observation of EL cooling resulted from the realization that ηQLηext approaches a finite value at arbitrarily small forward bias voltage. By operating an LED emitting light in the mid-infrared spectrum at a very low applied voltage, we were able to achieve hω/qηQLηext > 1. EL cooling was demonstrated by raising the temperature of the LED’s lattice (using externally supplied heat) to create a high ambient temperature) which exponentially increased the intrinsic carrier concentration in the active region and in turn the output light power at a given applied voltage.1

While this approach did allow for the first experimental observation of EL cooling in an LED, the low light power (~70 pW or ~1.1 × 10−9 W mm−2) and elevated LED temperature (135 °C) has limited the immediate technological relevance of the experiment. The LED with which the effect was observed was not designed with EL cooling in mind nor was it optimized for operation in the high temperature and low bias regimes used in the experiment.10 It is thus of interest to develop a design strategy for optimizing a chosen figure of merit related to EL cooling or thermo-electric pumping in an LED and to give quantitative predictions of achievable improvements over what has been demonstrated. Here, we use the power density (W mm−2) achievable at unity wall-plug efficiency ηQLηext as the figure of merit for thermo-electrically pumped LEDs and consider the doping (species and density), the thickness of the LED’s active region and the operating temperature as design parameters.

As will be seen below, these design parameters affect LED performance interdependently and thus it is difficult to achieve their simultaneous optimization while considering them separately. Thus, we first consider optimizing the active region doping and operating temperature and then active region doping and thickness. We use a finite element model (discussed below) to simulate charge and energy transport in a GaInAsSb/GaSb LED and model L-J-V (i.e., optical power density, current density, and voltage) characteristics as a function of these design parameters. We sweep the parameter values over relevant ranges while keeping the material compositions of the active and barrier regions fixed.

In this work, we use a 1D finite element model developed in Synopsys Sentaurus TCAD for charge transport across double hetero-junction LED’s.11 The model involves calculation of the first three moments of the Boltzmann transport equation, a lattice heat diffusion equation, and an optical transmission calculation. The carrier transport equations are solved self-consistently with the lattice heat diffusion equation. The recombination mechanisms in the carrier transport model are Shockley-Read-Hall (SRH) (both bulk and at the hetero-interfaces), bimolecular (the only radiative mechanism), Auger, and surface recombination at the contacts (modeled with infinite recombination velocity).

The simulation is based on the experimental characterization of a commercial GaInAsSb/GaSb LED presented in Ref. 1 and includes three fitting parameters. A series contact resistance of 4.8 mΩ mm2 (0.78 Ω for a diode with transverse area of .062 mm2) and a non-radiative bulk SRH lifetime τ at 300 K of 95 ns were used to fit experimental I-V curves, and a combined efficiency describing light extraction and collection of 24.5% was used to fit experimental L-I curves. As the low-
bias quantum efficiency is typically small (ηQ ≪ 1), photon recycling is neglected. As is shown below, this seems to be a reasonable approximation in the system being modeled. The model matches experimental L-I-V characteristics of a GaInAsSb/GaSb LED with known active region thickness t (4 μm) and without intentional active region doping over the temperature range 25–135 °C. The predicted I-V curves match experiment to within 50% in I and L-I curves to within a factor of two in L over the temperature range despite needing to track V, I, and L over approximately four (75 μV–750 mV), six (300 nA–300 mA), and seven (70 pW–1 mW) orders of magnitude, respectively. We note that these quoted errors correspond to maximum discrepancy between model and experiment over the experimentally characterized range.

This validated model can be useful for design optimization. This transport model should accurately capture the affects of varied active region thickness and doping insofar as they affect radiative and non-radiative recombination mechanisms and leakage from the active region of the LED as a function of temperature. The accuracy of the model is limited by the availability of only approximate models for dopant-density-dependent band-gap narrowing and the temperature dependence of Auger recombination, neither of which have been experimentally characterized in the GaInAsSb material system. In our model the band gap of GaInSbAsSb is narrowed according to the Jain-Raulston parameters for GaSb and the Auger coefficient is taken to be temperature independent. We perform these simulations using the value for the SRH recombination lifetime τ = 95 ns previously used to fit the model to experiment for liquid-phase epitaxy (LPE)-grown LEDs as well as with a value τ = 1 μs, which we expect to be achievable with more optimized device growth methods.

We demonstrate (1) that an optimal (i.e. with the light output power at unity wall plug efficiency Lq=1 maximized) active region dopant density exists for a given operating temperature 0 °C < T < 150 °C and active region thickness t, (2) that the achievable Lq=1 rises monotonically with operating temperature across the temperature range considered even though ηQ at the η = 1 operating point decreases with temperature over a significant portion of the temperature range considered, (3) that with optimal active region dopant density Lq=1 can be increased by 137× at 20 °C and 4.7× at 150 °C, (4) that for a bulk SRH lifetime 10.5× longer than is used in our experimentally validated model (1 μs as opposed to 95 ns), we expect another improvement in Lq=1 of 3.5× at 20 °C and 2.9× at 150 °C over the existing device studied experimentally, and (5) that for optimized active region thickness the Lq=1 can be further improved by 29% at 25 °C. Thus on aggregate, we expect that active region doping and thickness and cleaner device growth could improve Lq=1 by 621× from 2.7 × 10−13 W mm−2 to 1.7 × 10−10 W mm−2 at 20 °C. Furthermore, we show that optimized devices see more dramatic performance improvements at slightly lower efficiencies as the bias leaves the linear regime (qV ≈ kBT).

An analytic model can be useful in understanding the trends observed in the numerical models. In the low-bias regime of LED operation, we expect that current primarily replenishes carriers lost by carrier recombination in the active region and that quasi-Fermi level separation is approximately uniform across the active region. Thus, if we temporarily neglect carrier leakage from the active region and surface recombination at the hetero-interfaces, we may approximate the current density J = t q Wact, and L = ηQ ηext J act, where t is the active region thickness, q is the electron charge (positive), and Wact (cm−3 s−1) is the total recombination rate per unit volume. We calculate Wact as in Ref. 2.

\[
W_{\text{act}} \approx \frac{1}{\tau(n+p+2n_i)} + B + C(n+p) \left( \frac{e^{qV}}{q} - 1 \right) = M_{\text{act}} n_i^2 \left( \frac{e^{qV}}{q} - 1 \right),
\]

where τ is the SRH lifetime, n and p are the electron and hole concentrations respectively, n_i is the intrinsic carrier concentration, B is the bimolecular (radiative) recombination coefficient, C is the Auger recombination coefficient, k_B is the Boltzmann coefficient, T is the temperature and we introduce M_{\text{act}} (cm^2 s^{-1}) as a parameter describing the total recombination in the active region in the same units as B for comparison. When (1 − η_{\text{ext}}) η_Q ≪ 1, where η_{\text{ext}} is the quantum efficiency of photon extraction from the active region, photon recycling is negligible and we have η_Q ≈ B_{\text{ext}}.

At η = η_{\text{ext}} \eta_Q \frac{h \omega}{q V} \approx 1, we note that

\[
V_{\eta} = \frac{h \omega}{q} \eta_{\text{ext}} \eta_Q = \frac{h \omega}{q} \frac{B}{M_{\text{act}}} \eta_{\text{ext}}.
\]

In the low-bias regime (qV ≪ k_B T ≈ 26 meV), the exponential dependence of current density J on applied voltage V is approximately linear. This can be seen by Taylor expansion of the exponential in the expression for W_{\text{act}} from Eq. (1). In this regime where voltage and current have a linear relationship, we can assign a low-bias resistance R (Ω) to the diode. Assuming a transverse area A (cm^2), we find

\[
R = \frac{V}{J A} \approx \frac{k_B T}{A q^2 n_i^2 t M_{\text{act}}}.
\]

Thus at unity wall-plug efficiency, we have

\[
L_{\eta=1} \approx \frac{V_{\eta=1}}{R A} = \left( \frac{h \omega \eta_{\text{ext}} n_i B}{k_B T M_{\text{act}}} \right)^{\frac{1}{2}} t.
\]

These expressions are useful for thinking about design of an LED optimized for maximal L_{\eta=1}.

As suggested by Heikkila et al.,^2 the low-bias quantum efficiency of a double hetero-junction LED can be improved by doping the active region so that it operates optimally between lower- and higher-carrier density regimes where the non-radiative SRH recombination rate (∼np/(n+p)) and Auger recombination rates (∼n^2p + p^2n) are, respectively, dominant over radiative recombination rates (∼np). Intuitively, upon injection, a given electron is more likely to find a hole for bimolecular recombination in a p-doped active region than in an intrinsic one, while the likelihood it will encounter defect for SRH or an electron-hole pair for Auger recombination remains roughly unchanged.

If we assume that before doping, the bulk SRH recombination process is much faster than the bimolecular and Auger processes (a reasonable assumption in our device at
We see that moderate doping ($N_0 = 3 \times 10^{17}$ cm$^{-3}$) as compared to that between an $n$-type active region decreases the spatial extent over which the quasi-Fermi levels are near mid-gap, where our analysis indicates that the carrier concentration is approximately one order of magnitude more than optimal as a function of temperature for all combinations of $n$- and $p$-type doping (at optimal densities for maximal room temperature $L_{\eta=1}$) and SRH lifetime of $\tau = 95$ ns and $\tau = 1 \mu$s.  We see from Figure 1 that $L_{\eta=1}$ increases monotonically over the temperature range considered and that optimal $p$-doping improves $L_{\eta=1}$ by approximately one order of magnitude more than optimal $n$-doping for a given temperature and SRH lifetime. We associate this latter feature with more dramatic band bending at the interface between a $p$-doped active region and $n$-GaSb as compared to that between an $n$-doped active region and $p$-GaSb. Large band bending at the interface between a $p$-region and an $n$-region decreases the spatial extent over which the quasi-Fermi levels are near mid-gap, where our model assumes the trap states causing SRH recombination lie. We believe that this insight could inform a spatially varied doping scheme further optimized for higher $\eta_Q$ and $L_{\eta=1}$, but we do not attempt that here.

The second parameter considered in this work is the temperature of the LED. We see in Figures 1(a) and 1(b) that $L_{\eta=1}$ rises monotonically with temperature. This is easy to understand in terms of Eq. (4), since $L \sim \eta_Q J_n \sim n_i^2 e^{-\phi_T}$, where $J_n$ varies from $22.7 \mu A/mm^2$ to $13.5 \mu A/mm^2$ at $150^\circ C$, $n_i^2$ varies by almost four orders of magnitude over the temperature range investigated. This temperature dependence is dominant over those of other material parameters in the equations above. For this reason, $L_{\eta=1}$ rises monotonically (and approximately exponentially) with increasing temperature beyond temperatures at which contacting an LED becomes problematic. Figure 2 shows how leakage and the various recombination mechanisms contribute to current at $L_{\eta=1}$ at optimal doping as a function of temperature. We see that as the temperature is increased the internal quantum efficiency at $L_{\eta=1}$, set approximately by the ratio of the bimolecular and SRH recombination rates in the active region at $20^\circ C$, passes a maximum and then begins to drop. This is due to the relative strengths of Auger recombination and carrier leakage increasing with device temperature. We note that, as seen in Eq. (5), the improvement in $\eta_Q$ and $L_{\eta=1}$ achievable by doping the active region decreases as $n_i^{-1}$, since the benefit of introducing extra carriers becomes less significant as the intrinsic carrier density rises.

Note that no significant temperature gradient exists across the LED because we have modeled the $L-J-V$ characteristics at low bias and the LED is thermally coupled to the ambient environment. If instead the LED were thermally insulated from conduction and convection heat exchange with its environment, then heat generated by non-radiative recombination would raise the temperature of the device until a wall-plug efficiency near unity is reached.
The LED’s active region thickness \( t \) is another design parameter that can be optimized. From Eq. (4), we expect that \( L_{\eta=1} \sim J \sim t \), however, we must keep in mind that the extraction efficiency \( \eta_{\text{ext}} \) is also a function of \( t \). The inset in Figure 3 shows the extraction efficiency from the active region calculated for an active region of GaInAsSb for \( 0 < t < 4 \mu m \), where we have used an experimental absorption coefficient\(^{15} \) of 4000 cm\(^{-1} \) for the quaternary. As the active region thickness is reduced, carrier leakage over the potential barriers from the GaInAsSb active region to the GaSb barrier layers (not included in the equations above) will contribute significantly to current at low bias and decrease \( \eta_{\text{Q}} \). We simulated \( L-J-V \) characteristics of LEDs with values of \( t \) between 100 nm and 3.9 \( \mu m \) and active region dopant densities between \( 3 \times 10^{14} \) cm\(^{-3} \) and \( 6 \times 10^{17} \) cm\(^{-3} \) at each thickness. In each case, the modeled LED temperature was 25 °C. Figure 3 shows how \( L_{\eta=1} \) has a maximal value as a function of thickness for a given dopant density. For GaInAsSb/GaSb LEDs modeled at 25 °C, we expect maximal \( L_{\eta=1} \) for an active region dopant density of approximately \( 2 \times 10^{16} \) cm\(^{-3} \) and thickness of approximately 1.5 \( \mu m \) at 25 °C. We expect this optimal thickness to increase as a function of temperature since carrier leakage from the active region should have an exponential dependence on the ratio of the band edge energy differences at the edges of the active region to \( k_B T \). We see that \( L_{\eta=1} \) is enhanced by 29% as compared to the original structure (where \( t = 4 \mu m \)) by optimizing the active region thickness, which while significant is a much weaker improvement than that due to active region doping.

As the applied voltage \( V \) increases beyond \( k_B T \eta_{\text{Q}} \), the LED leaves the linear regime and Eqs. (3) and (4) are no longer valid. We see from Eq. (2) that \( V_{\eta=1} \) rises linearly with the LED’s quantum efficiency and for \( \eta_{\text{Q}} \eta_{\text{ext}} \approx \frac{4k_B}{h} \) (approximately 0.05 in our active material at 300 K) \( V_{\eta=1} \) will be pushed into this “rectifying” regime. In that case, the nonlinearity of the exponential dependence of the recombination rates (and thus both current and output light power) on applied voltage in Eq. (1) will give a much greater differential improvement in \( L_{\eta=1} \) as a function of \( \eta_{\text{Q}} \). While we do not achieve internal quantum efficiencies high enough to reach this regime in this work, we are able to demonstrate this principle by considering slightly lower wall-plug efficiencies.

![FIG. 3. Output optical power density at unity wall-plug efficiency \( L_{\eta=1} \) in W mm\(^{-2} \) for \( p \)-dopant densities of \( N_A = 3 \times 10^{14} \) cm\(^{-3} \) (blue dashed line), \( 2 \times 10^{15} \) cm\(^{-3} \) (red solid line), and \( 6 \times 10^{15} \) cm\(^{-3} \) (black solid line) in the active region at 25 °C. The inset shows calculated extraction efficiency from the GaInAsSb active region as a function of active region thickness.](https://example.com/fig3)

![FIG. 4. Calculated wall plug efficiency \( \eta \) as a function of optical power density \( L \) at 25 °C for an unoptimized device (red dotted line) and a device optimized in active region \( p \)-doping and thickness (dashed blue line). Experimental points corresponding to the unoptimized structure are shown with filled red dots to demonstrate the validity of the model. The black solid curve represents the thermodynamic upper bound for the wall plug efficiency of these devices.](https://example.com/fig4)
system. Our model suggests that improvements feasible through doping and thickness optimization combined with cleaner device growth techniques should allow for improvement by a factor of $621/C^2$ in $L_{\text{g}} = 1$ at $25^\circ C$; for the device area reported in Ref. 1, this represents a change in power at 100% efficiency from 20 fW to over 10 pW. Further improvements of thermo-electric pumping designs could be relevant to the design of ultralow power (sub-μW) pulse-oximetry systems for healthcare applications and low power (>100 μW) gas detection systems, where low-power infrared LED inefficiency creates a bottleneck for minimizing system power consumption.

12We note that some characterization of Auger recombination as a function of pressure has been performed and suggests that the band gap of GaInAsSb narrows faster than the split-off energy gap $\Delta$ between the hole band maxima and that when the band gap narrows to be smaller than $\Delta$ the Auger recombination rate decreases significantly as LHHS Auger processes become suppressed.