Electro-Luminescent Cooling: Light Emitting Diodes Above Unity Efficiency

Parthiban Santhanam\textsuperscript{a}, Duanni Huang\textsuperscript{a}, Dodd J. Gray Jr.\textsuperscript{a,b}, and Rajeev J. Ram\textsuperscript{a}

\textsuperscript{a}Research Lab of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139
\textsuperscript{b}Department of Electrical Engineering, Stanford University, Stanford, CA

ABSTRACT

Experimental demonstration of net electro-luminescent cooling in a diode, or equivalently electroluminescence with wall-plug efficiency greater than unity, had eluded direct observation for more than five decades. We review experiments demonstrating light emission from a light-emitting diode in which the electron population is pumped by a combination of electrical work and heat.

1 INTRODUCTION

It has long been known that in theory a light-emitting diode (LED) may emit optical power in excess of the electrical power required to drive it, with the remainder drawn from lattice heat. This phenomenon has been referred to in the literature by various authors as electro-luminescent cooling \cite{1, 2}, electroluminescence refrigeration \cite{3, 4}, opto-thermionic cooling \cite{5, 6}, the operation of a “Thermischer Konverter” \cite{7}, and thermo-photonic cooling \cite{8}.

Electro-luminescent cooling is possible when a light-emitting diode is placed under a small forward bias \( qV < \hbar \omega \), where \( q \) is the magnitude of the electronic charge, \( V \) is the bias voltage, and \( \hbar \omega \) is the average energy of an emitted photon. Light emission at these low voltages \( (qV < \hbar \omega) \) was first observed in 1964 \cite{9} and is readily observable in commercial devices today. Nevertheless, the presence of parasitic non-radiative recombination had not permitted experimental observation of net cooling until our recent work \cite{10}. To observe the phenomenon, we investigated much lower bias operation \( (qV \ll k_B T) \) of infrared LEDs at elevated temperature. Here we discuss the theoretical basis of our experimental approach and summarize the experiments reported to date.

2 THE LED AS A THERMODYNAMIC MACHINE

In Statistical Mechanics, the word “heat” is used to refer to any form of energy which possesses entropy \cite{11}. This usage applies equally to forms of energy referred to colloquially as “heat,” such as the kinetic energy in the relative motion of the molecules in a gas or the constant vibrations of atoms in a crystal lattice, as well as those for which the entropy is frequently less relevant, such as the kinetic energy in the relative motion of electrons and holes in a semiconductor or the thermal vibrations of the electromagnetic field in free space. Critically, the Laws of Thermodynamics which govern the flow of heat are formulated independently of the Laws which govern the deterministic trajectories of mechanical systems, be they classical or quantum. As a result concepts such as the Carnot limits for the efficiency of various energy conversion processes apply equally well to the gases and solid cylinder walls of an internal combustion engine as to the electrons, holes, and photons in a modern LED.

An LED is an electronic device which takes entropy-free electrical work as input and emits incoherent light which carries entropy. Instead of irreversibly generating the entropy that it ejects into the photon reservoir, an LED may absorb it from another reservoir at finite temperature, such as the phonon bath. As the diagram in Figure 1 suggests, the device may absorb heat from the phonon bath and deposit it into the photon field in much the same way as a Thermo-Electric Cooler (TEC) absorbs heat from the cold side of the module and deposits it on the hot side \cite{12, 13}. In the reversible limit the flows of energy and entropy are highly analogous for an LED and a TEC. Moreover, in both the LED and TEC, the Peltier effect is responsible for the absorption of lattice heat by electrons and holes \cite{14, 15, 16, 17}. Electrical work is being used to pump entropy from one reservoir to another instead of simply creating it through irreversible processes. The devices are thermodynamic heat pumps.
Figure 1: Diagrams depicting energy and entropy flows in two types of thermodynamic heat pumps: TECs (top row) and LEDs (bottom row). The left column shows the theoretical energy and entropy flows in Carnot-efficient devices. The right column shows the same in devices with common sources of irreversibility.
For each bit of entropy $\delta S$ absorbed on net from the phonon reservoir at finite temperature, an amount of heat $T_{\text{lattice}} \Delta S$ comes with it. Since input and output power must balance in steady-state, the rate at which this heat and the input electrical work enter the system (both measured in Watts) must exactly equal the rate at which heat is ejected into the photon reservoir (also measured in Watts). That is to say, when lattice heat is being absorbed on net an LED’s wall-plug efficiency $\eta$ (or equivalently its heating coefficient of performance), defined as the ratio of output optical power to input electrical power, must exceed unity.

The Second Law of Thermodynamics (i.e. non-deletion of entropy) places a clear limit on the maximum efficiency of an LED in this framework. To understand this limit, we must first understand the thermodynamics of the photon and phonon fields at finite temperature.

### 3 RESERVOIR TEMPERATURES

When an LED is operating above unity efficiency, heat is continuously extracted from the lattice by the electronic system. Due to the large heat capacity of the lattice relative to the electron-hole system, the phonon field of the semiconductor diode acts very nearly as a perfect reservoir. That is to say, the other statistical subsystems which it interacts with may deposit or withdraw any amount of energy $\Delta U$ from it so long as it is accompanied by a proportional amount of entropy $\Delta S$. The constant of proportionality is given by the lattice temperature, yielding the following equation for a perfect phonon reservoir:

$$\Delta U = T_{\text{lattice}} \delta S \quad (1)$$

In reality, due to the large but finite heat capacity of the phonon bath and its large but finite thermal conductance to ambient, the lattice remains slightly cooled compared to its surroundings, so that heat is continuously conducted into the device from the environment in steady-state. Rather than self-heating, the LED is experiencing self-cooling.

In contrast, the heat capacity of the photon field in the relevant range of wavelengths is not much larger than that of the electron-hole system. Furthermore, when the LED is emitting light, the optical field of the outgoing radiation modes is not in equilibrium at ambient temperature. Nevertheless, we can analyze it thermodynamically as follows.

Incoherent electromagnetic radiation which originates in an LED is equally capable of carrying entropy with it as electromagnetic radiation from a hot blackbody. All incoherent light is therefore, in the statistical-mechanical sense mentioned above, a type of heat. The ratio of the rate at which radiation carries away energy to the rate at which it carries away entropy gives its flux temperature \[18\]:

$$T_F = \frac{\dot{U}}{\dot{S}} = \frac{dU/dt}{dS/dt} \quad (2)$$

Although this notion of temperature may be used to calculate the thermodynamic limits of power-conversion efficiency, the rate of entropy flux in light is difficult to measure directly. Fortunately a more intuitive definition of the temperature of light is described in Figure 2. Consider two bodies that are each perfectly thermally isolated from their environments (i.e. by adiabatic walls) and similarly isolated from each other. Suppose body 1 has energy $U_1$ and entropy $S_1$ and likewise the second body has energy $U_2$ and entropy $S_2$. If the insulating boundary between bodies 1 and 2 is replaced with one which permits the flow of energy, the total energy $U_1 + U_2$ will flow to rearrange itself in the way which maximizes the total entropy. The flow will stop only when the addition of a differential amount of energy $\delta U$ to either body results in the same fractional increase in the number of available micro-states for that body (i.e. the same increase in its entropy). Equivalently, we may say that the flow of energy stops when the bodies have equal temperature \[11\]:

$$\frac{\partial S_1}{\partial U_1} \equiv \frac{1}{T_1} = \frac{1}{T_2} \equiv \frac{\partial S_2}{\partial U_2} \quad (3)$$

Now consider a similar scenario in which body 1 is an LED and body 2 is a perfect blackbody radiator. To begin, both bodies are adiabatically isolated from their environments and each other. In the case of the blackbody, the walls are a surrounding surface made of mirrors, such that the radiator has zero emissivity. In...
Figure 2: The figures in the left column depict the motivating logic behind the definition of temperature in the micro-canonical ensemble [11]. The figures at right depict a similar logic for incoherent light. The brightness temperature of an incoherent source (here, an LED) may be defined as follows. For each optical frequency, light ray direction, and position on the emitting surface, consider the temperature at which a perfect blackbody would emit with the same spectral intensity (i.e. power per unit area per unit frequency per unit solid angle). At this temperature, the optical fields would be in thermal equilibrium. If the light is entirely incoherent, the photon field is maximally disordered and this temperature indicates the ratio of the rate of energy flux to the rate of entropy flux carried by the radiation in that band. The weighted average of these temperatures over the intensity spectrum of the emitter gives the brightness-temperature of the source.
the case of the LED, the adiabatic walls must form a cavity with perfect reflectivity, such that each photon emitted reflects off a mirror and returns to the active region to generate a quantum of reverse-current. Assume no non-radiative recombination occurs. The LED is ‘on’, but is in steady-state and consumes no power. Assume that the bodies have no means of exchanging energy other than through photons and that to begin the boundary between them is also a perfectly-reflective mirror.

If the walls surrounding the bodies are modified to include transmitting pinholes along the axis between them and the mirror is modified to permit transmission over a narrow range of wavelengths around \( \lambda_0 \), energy will flow on net from the body with higher spectral power density in the direction of the other pinhole (i.e. \( I(\lambda) \) in W m\(^{-2}\) nm\(^{-1}\) str\(^{-1}\)) to the body with lower \( I(\lambda) \) at \( \lambda_0 \). If we assume the LED is perfectly incoherent, the flow of photons in either direction is equally capable of carrying entropy, and therefore equally justified in being termed ‘heat.’ Since heat may only flow from high temperature to low, the equilibrium condition for the two bodies may only be satisfied when \( I_1(\lambda_0) = I_2(\lambda_0) \). Since the relationship between intensity and temperature for a perfect blackbody is given by the Planck radiation law, we may define the brightness temperature \( T_B \) of an incoherent source as the temperature of blackbody whose spectral intensity equals that of the emitter at the wavelength and emission direction of interest [18, 19]:

\[
I_{\text{emitter}}(\lambda_0) = I_{\text{blackbody}}(\lambda_0; T_B) = \frac{4\pi^2c^2}{\lambda_0^5} \exp\left(\frac{\hbar(2\pi\lambda_0 I_0)}{k_B T_B}\right) - 1
\]

(4)

Note that unlike the color temperature of radiation commonly used in the lighting and display spaces, a longer-wavelength emitter is not necessarily cooler than a short-wavelength emitter. The linewidth, angular extent, wavelength, and intensity of the source all matter and may result in thermodynamically-cold emission from a blue LED or thermodynamically-hot emission from a red one. That is to say, the flux temperature \( T_F \) and brightness temperature \( T_B \) of a source may be cool, even when the radiation is blue. A note to the reader: a more detailed discussion of the distinction between the flux \( (T_F) \) and brightness \( (T_B) \) photon temperatures can be found in [18] and [19].

Since the temperature of an incoherent photon flux is essentially a measure of its spectral intensity \( I(\lambda) \), the Second Law places a different efficiency constraint on emitters of different spectral intensity. As a function of lattice temperature and emitter intensity, the Carnot limit may be expressed compactly as follows:

\[
\eta \leq \eta_{\text{Carnot}} = \frac{T_{\text{photon}}(I)}{T_{\text{photon}}(I) - T_{\text{lattice}}}
\]

(5)

For bright sources \( (I(\lambda) \gg I_{\text{blackbody}}(\lambda; T_{\text{lattice}})) \), the LED must pump heat against the large temperature difference between the lattice and the outgoing photon field. This results in a maximum efficiency, even for a perfect Carnot-efficient LED, which exceeds unity but only slightly. For dim sources \( (I(\lambda) \approx I_{\text{blackbody}}(\lambda; T_{\text{lattice}})) \), the LED must only pump heat against a small temperature difference. As a result, efficiencies far in excess of unity are possible.

Examination of Equation 5 at fixed spectral intensity \( I \) reveals another counter-intuitive aspect of the heat-pump regime. As \( T_{\text{lattice}} \) is increased, the temperature difference against which the LED must pump becomes smaller, and the maximum allowable efficiency increases.

Thus, the basic thermal physics of an LED in the heat pump regime is the reverse of the conventional thermal physics:

- Above-unity efficiency results in self-cooling that decreases the device’s operating temperature.
- For a desired spectral intensity, a higher lattice temperature means that the device can be more efficient.

These differences may result in practical consequences for both the device-level design of LEDs and the thermal design of their packaging.

4 ELECTRONS AS THE WORKING FLUID

To quantify the heat pump picture for a particular LED, we must use the dynamics of the carriers to calculate the entropy flows to and from each reservoir. Based on the Peltier effect, a picture of net heat flow may be drawn from a band diagram as shown in Figure 3.
For simplicity, consider the processes of carrier injection and recombination separately. When electrons flow from a metal into a lightly $n$-doped semiconductor, the average energy of the carriers involved in conduction increases from around the Fermi level to an energy above it. This increase in energy is supplied by lattice heat in steady state through the Peltier effect. Generalizing this principle and applying it to the conventional double hetero-junction LED in Figure 3, we may conclude that as electrons flow from the negative contact towards a typical recombination site in the active region, lattice heat is absorbed as the electron energy increases. Likewise, as holes enter from the positive contact and diffuse toward the recombination site, they too must draw energy from the lattice. Since the lattice is a thermodynamic reservoir, this energy also has entropy associated with it. Thus in forward bias, during injection the carriers absorb entropy and energy from the lattice with a flux proportional to the slope of the relevant band edge.

Likewise, recombination results in the flow of energy and entropy out of the electron-hole system. Critically, although recombination and generation events take place continually even when there is no current, it is the net recombination which determines these flows.

As with the majority carriers in the doped regions, even when the device is off the electrons and holes in the active region are perpetually experiencing generation and recombination as the result of their interaction with other reservoirs. These processes can be thought of in terms of the following chemical reaction equation:

$$e^- + h^+ \leftrightarrow U_{\text{bandgap}}$$

where $e^-$ is an electron, $h^+$ is a hole, and $U_{\text{bandgap}}$ denotes some excitation with energy (and other conserved quantities) equal to that of the electron-hole pair. As with a typical chemical reaction, the reactants and products are in equilibrium at some concentrations. When carriers are injected into the active region by a forward bias voltage, the concentration of electrons $n$ and holes $p$ exceeds these values (i.e. when $np$ exceeds the squared intrinsic carrier concentration $n_i^2$). Net recombination occurs and the reaction in Equation 6 is driven from left to right.

Each time that an electron-hole pair is annihilated, both energy and entropy are removed from the electron and hole gases. That is to say, the number of microscopic configurations in which the conduction and valence bands can be occupied is reduced. However, this entropy cannot disappear entirely as doing so would violate the Second Law. Instead, the entropy which is removed from the electronic sub-system (i.e. the degrees of freedom from excitations of the conduction and valence band states) is transported to another sub-system at the same location in the device. Which sub-system that is depends on where the electron-hole pair’s energy went. As seen in Figure 3, for non-radiative recombination, the destination is the lattice. For radiative recombination, the destination is the photon field.

Let us now abstract away the internal dynamics of the electronic system and consider just the flows of entropy and energy between the three sub-systems in Figure 3. For each quantum of charge that flows through the device, one net recombination event occurs. We would like to know how much entropy enters and leaves each system. Knowledge of the energy flows between the sub-systems combined with Equation 1 and Equation 2 determines the entropy flows in and out of the lattice and photon fields respectively. However, because the electronic sub-system is not in equilibrium at any fixed temperature, we must examine it more closely.

We begin with a simple model for the electronic degrees of freedom at a single point in space. Consider the statistical two-level system shown in Figure 4. Define $f_e$ to be the probability of occupancy for the higher energy state, $f_v$ to be the occupancy of the lower state, and take the states to be separated by energy $\Delta E$. In terms of these quantities then, we may write expressions for the total energy and entropy of the system:

$$U = f_e \cdot \Delta E + (\text{const}) \quad \text{and}$$

$$S = -k_B \left[ (f_e \ln f_e + (1 - f_e) \ln(1 - f_e) + (f_v \leftrightarrow f_e) \right].$$

$$f_e$$ denotes some excitation with energy (and other conserved quantities) equal to that of the electron-hole pair.
Figure 3: Simplified band diagram depicting energy and entropy flux in a conventional LED structure.

Figure 4: A statistical two-level system.
If we define a degree of freedom corresponding to excitation from the lower state to the upper state, we may find the amount of entropy change in the system per unit energy change for distortions of this type. This ratio can be expressed conveniently as the inverse temperature $T^{-1}$ of the electronic system:

$$ T^{-1} = \frac{\partial S}{\partial U} = \frac{dS}{df_c} - \frac{dS}{df_v} $$

$$ \frac{dS}{df_c} = -k_B \ln f_c + 1 - \ln(1 - f_c) - 1 $$

$$ \frac{dS}{df_v} = -k_B \ln \left( \frac{f_c}{1 - f_c} \right) $$

$$ T^{-1} = \frac{\partial S}{\partial U} = \frac{-k_B \ln \left( \frac{f_c}{1 - f_c} \right) + k_B \ln \left( \frac{f_c}{1 - f_c} \right)}{\Delta E} $$

If we constrain the probability for occupancy of either state $f_c + f_v$ to be 1 so that the Fermi level $E_F$ falls halfway between the states in energy, the equation above can be rearranged to recover the expression for Fermi-Dirac occupancy in equilibrium at temperature $T$:

$$ \exp \left( \frac{\Delta E}{k_B T} \right) = \frac{f_c}{1 - f_c} \cdot \frac{1 - f_c}{f_c} = \left( \frac{1 - f_c}{f_c} \right)^2 $$

$$ \exp \left( \frac{\Delta E/2}{k_B T} \right) = \left( \frac{1}{f_c} - 1 \right)^{-1} $$

$$ f_c = \left( \exp \left( \frac{E_{\text{upper-state}} - E_F}{k_B T} \right) + 1 \right)^{-1} $$

The preceding result is unsurprising, but clarifies an important point. The inverse temperature of a Fermionic system, meaning the amount of entropy that is added to it when a unit of energy is added, can be calculated purely from the occupation of the states. That is to say, two situations which are described differently must still have the same temperature if their occupancies are the same.

Consider now the similar but more physical example of an ensemble of homogeneous quantum dots, each with one low-energy electron state and one high-energy state as before, but each also possessing a lattice with temperature $T_{lattice}$. Let the total charge between the states be $f_c + f_v = 1$ to ensure charge neutrality. If $T_{lattice}$ is kept at 300K and no electrical excitation is applied, the statistical two-level system will have a Fermi level at exactly halfway between the two states and the occupancies $f_c$ and $f_v$ can be determined by the Fermi-Dirac distribution. This situation is described by the diagram in Figure 5a.

Let us now excite this system. Since a recombination event removes an electron from a higher energy state and places it in a lower energy state (and vice versa for a generation event), let us explore the degree of freedom corresponding to $f_c \rightarrow f_c + \delta f$ and $f_v \rightarrow f_v - \delta f$. Note that this is the same degree of freedom that we used in Equation 9 and corresponds to excitations that retain charge neutrality. Figure 5b and Figure 5c show two physically different types of excitations that result in the same values of $f_c$ and $f_v$. In Figure 5b, the electrical system has been taken out of equilibrium with the lattice by an applied voltage $qV = \Delta E/2$. In Figure 5c, the absolute temperature of the lattice has been doubled. In both situations, the number of $k_B T$’s of between each state and its quasi-Fermi level has been halved. Consequently the Fermi-Dirac occupation of the states in both situations is equal (i.e. $f_c$ and $f_c$ are the same in both). Since the total entropy $S$ and energy $U$ of the electron-hole system are determined entirely by $f_c$ and $f_v$ (see Equation 7), these quantities are also equal.

As a result, the effective temperature $T^* = (\partial S/\partial U)^{-1}$ with which the electron-hole system interacts through
inter-band processes is also the same for the electrical (Figure 5b) and thermal (Figure 5c) excitation conditions. From these examples, we may follow [20] to a general expression for $T^*$ in a semiconductor whose quasi-Fermi levels are separated by an energy $\Delta E_F$ in a region with bandgap energy $E_{\text{gap}}$:

$$T^* \equiv \frac{T_{\text{lattice}}}{1 - \frac{\Delta E_F}{E_{\text{gap}}}}$$  \hspace{1cm} (16)$$

From here, we may simplify the internal dynamics of the electronic system into a simple thermodynamic model. For inter-band processes in which the electronic system loses energy to another reservoir (e.g., recombination), the corresponding loss of entropy is determined by $T^*$ from Equation 16. By contrast, for the intra-band electron-phonon scattering processes that comprise thermally-assisted injection, the amount of entropy exchanged during an energy exchange is given by $T_{\text{lattice}}$.

If we modify Figure 3 by consolidating all flows of entropy together, and we also include the corresponding flows of energy from the various sources, the picture becomes the canonical diagram for a thermodynamic heat pump as shown in Figure 6. As electrons and holes are injected into the active region, they absorb heat from the phonons at $T_{\text{lattice}}$. When the electrons and holes undergo radiative recombination in the active region, they deposit that heat into the photon field at temperature $T^* > T_{\text{lattice}}$. Thus the electrons and holes act as a working fluid in a heat pump operating between these two temperatures. Additionally, in the optically thick limit (i.e., when light rays travel many absorption lengths as they pass through the active region), the active region will radiate like a blackbody with unity emissivity, so that $T_{\text{photon}} = T^*$. Finally, in this case of only radiative recombination, since no irreversible processes are taking place, the heat pump is Carnot efficient.

5 PREVIOUS WORK TOWARD UNITY EFFICIENCY

For several decades it has been theoretically understood that the presence of entropy in incoherent electromagnetic radiation theoretically permits semiconductor light-emitting diodes (LEDs) to emit more optical power than they consume in electrical power [21, 18, 22, 23]. Moreover, starting very early on the phenomenon has drawn the attention of the applied community. In 1959 a US Patent was granted for a refrigeration device based on the principle [24]. In the last decade, the applied literature on the subject has expanded to include more realistic modeling and more recent advances in device fabrication technologies [1, 2, 3, 4, 5, 6, 7] and at least one attempt to demonstrate practical cooling is currently underway [8]. Nevertheless, prior to this work, the basic phenomenon of electrically-driven light emission above unity efficiency had never been experimentally verified.
Figure 6: The flows of entropy and energy between various sub-systems in an LED may be organized in the canonical picture of a thermodynamic heat pump. At left is an idealized picture. The irreversible contributions shown in the picture at right can be quantified for any real LED using the arguments from §4.

The experimental literature on electro-luminescent cooling stretches back more than five decades, beginning before even the early theoretical work of Tauc [21] in 1957 and Weinstein [18] in 1960. A summary of this work appears in Table 1.

<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>$V_{\text{min}}$</th>
<th>$qV_{\text{min}}/k_B T$</th>
<th>$e^{-\hbar \omega / k_B T}$</th>
<th>Max Reported $\eta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1953</td>
<td>Lehovec, et al. [25]</td>
<td>1.8 V</td>
<td>70</td>
<td>$\leq 2.5 \times 10^{-34}$</td>
<td>Not Published</td>
</tr>
<tr>
<td>1966</td>
<td>Nathan, et al. [26]</td>
<td>1.1 V</td>
<td>6380</td>
<td>$10^{-3630}$</td>
<td>6 %</td>
</tr>
<tr>
<td>2005</td>
<td>Wang, et al. [27]</td>
<td>0.36 V</td>
<td>14.2</td>
<td>$3.8 \times 10^{-11}$</td>
<td>Not Published</td>
</tr>
<tr>
<td>2011</td>
<td>Oksanen, et al. [8, 28]</td>
<td>0.5 V*</td>
<td>19.3*</td>
<td>$4 \times 10^{-13}$</td>
<td>Not Published</td>
</tr>
<tr>
<td>2012</td>
<td>Our Work [10, 29]</td>
<td>5.8 $\mu$V</td>
<td>0.00016</td>
<td>$1.04 \times 10^{-6}$</td>
<td>2134 $\pm$ 177 %</td>
</tr>
</tbody>
</table>

Table 1: Summary of previous experiments towards electro-luminescent cooling (i.e. electro-luminescence with $\eta > 1$). The asterisk (*) indicates that these figures were taken from simulation data. The quantity $qV_{\text{min}}/k_B T$ highlights the primary difference between the approach taken in this work and previous experiments. The quantity $e^{-\hbar \omega / k_B T}$ provides a scale for the optical power available in the low-bias regime.

As early as 1953, Lehovec et al. speculated on the role of thermo-electric heat exchange in SiC LEDs [25]. The authors were motivated by their observation of light emission with photon energy $\hbar \omega$ on the order of the electrical input energy per electron, $qV$. In 1964, Dousmanis et al. demonstrated that a GaAs diode could produce electro-luminescence with an average photon energy 3% greater than $qV$ [9]. Still, net cooling was not achieved due to non-radiative recombination [20] and the authors concluded that the fraction of current resulting in escaping photons, typically called the external quantum efficiency $\eta_{\text{EQE}}$, must be large to observe net cooling:

“Diodes with high quantum yield are required for direct experimental observation of the cooling effect.”

Dousmanis, et. al.
Physical Review, 1964
A similar observation was made two years later in a cryogenic GaAs LED (\(\hbar \omega = 1.44 \text{eV}\)) by Nathan, et al. [26]. Then after several decades of minimal experimental activity, multiple modeling and design efforts were undertaken and concluded that \(\eta_{\text{EQE}}\) could be raised toward unity by maximizing the fraction of recombination that is radiative [1, 2, 5] and employing photon recycling to improve photon extraction [1, 2, 4]. Along with these efforts, at least one experiment was performed by Wang, et al. in 2005 [27], but no optical power or wall-plug efficiency data was found to be published. At least one effort to observe electro-luminescent cooling with \(\eta_{\text{EQE}}\) above 50% continues to be active [8], although early results suggest problems with shunts in the emitting diode [28].

All of these experiments followed the logic of the quote above from Dousmanis, et al. by attempting to raise \(\eta_{\text{EQE}}\) toward unity. In contrast, \(\eta > 1\) was observed in this work with \(\eta_{\text{EQE}} \approx 3 \times 10^{-4}\). Since each electron that passes through the device consumes \(qV\) of work and leads to the emission of a photon of energy \(\hbar \omega\) with probability \(\eta_{\text{EQE}}\), the wall-plug efficiency \(\eta\) of a diode may be expressed as follows [21, 1]:

\[
\eta = \frac{\hbar \omega}{qV} \cdot \eta_{\text{EQE}},
\]

in order to achieve above-unity \(\eta\) with small \(\eta_{\text{EQE}}\) requires \(V \ll \hbar \omega / q\). Multiple authors have dismissed such operating regimes in the past because of the low output power available in this regime, but the present work has found it’s consideration worthwhile for 3 main reasons:

- Regardless of the power requirements for a practical cooling system, lower power may be sufficient for specific applications and/or experimental confirmation.

- The greatest deviations from conventional \(\eta < 1\) operation (i.e. highest coefficients of performance) always occur at low power. This is a general property of endo-reversible heat pumps.

- The decrease in power from lowering \(V\) can be compensated by increasing the ratio \(k_B T / h \omega\).

The third observation above was made in 1985, when Paul Berdahl presented an analysis of semiconductor diodes as radiant heat engines [20]. In that work, he showed that the available cooling power decreased exponentially with the ratio of the diode material’s bandgap \(E_{\text{gap}}\) to the thermal energy \(k_B T\), in accordance with the blackbody emission power integrated over the absorptive/emissive band.

6 LEDs IN THE LOW-BIAS REGIME

As described previously, it has long been known that at low output power an LED may in principle operate with wall-plug efficiency \(\eta\) far in excess of unity [18, 21]. That is, its optical output power (\(L\), measured in Watts) may be a large multiple of its input electrical power (\(IV\), also measured in Watts) in steady-state. In fact, the Second Law of Thermodynamics permits an arbitrarily large value of \(\eta\) at low power. This is the situation in the low-bias regime we will discuss here.

We pause briefly to address a question of terminology. Typically the ratio of the rate that heat (in this case, photons) is emitted to the rate at which work is consumed is called the heating coefficient of performance \(COP_H\), but in this work we refer to this quantity as the wall-plug efficiency \(\eta\). We note that in other electrically-driven sources of incoherent light for which \(\eta < 1\), the output energy also has entropy associated with it, so that \(L/(IV)\) would be most appropriately termed \(COP_H\) in this case as well. Nevertheless, convention dictates that \(L/(IV)\) is referred to as the wall-plug efficiency \(\eta\). For this reason, we follow several previous authors [21, 18, 1, 2] in referring to this quantity as the wall-plug efficiency (or simply efficiency) \(\eta\), which we allow to exceed unity.

Decomposing the expression for the external quantum efficiency, our basic expression for wall-plug efficiency becomes:

\[
\eta = \frac{\hbar \omega}{qV} \cdot \eta_{\text{extract}} \cdot \eta_{\text{inject}} \cdot \frac{\langle R_{\text{radiative}} \rangle_{\text{active}}}{\langle R_{\text{SRH}} \rangle_{\text{active}} + \langle R_{\text{radiative}} \rangle_{\text{active}} + \langle R_{\text{Auger}} \rangle_{\text{active}}}.
\]
Here \( \eta_{\text{extract}} \) is the efficiency with which generated photons are extracted from the device, \( \eta_{\text{inject}} \) is the efficiency with which injected electrons from the cathode and injected holes from the anode fall into the narrow-gap active region and remain confined there until recombination, and the final term expresses the fraction of that active-region recombination which is radiative.

Although different device structures and material systems lead to various types of recombination whose rates (both relative and absolute) can vary widely, here we will consider three processes: trap-assisted Shockley-Reed-Hall, bimolecular radiative, and Auger recombination. The rates of SRH, bimolecular, and Auger recombination are typically expressed in terms of the electron and hole concentrations, \( n \) and \( p \) respectively, while all other dependences are captured by some phenomenological rate constant (here \( A, B, \) and \( C \)). It is worth noting that these constants are intended to be independent of the magnitude of the local electrical excitation; the \( n \) and \( p \) dependences capture that physics. The most common form of these expressions appears below:

\[
R_{\text{SRH}} = \frac{(np - n_i^2)}{(n + n_i)\tau_n^{-1} + (p + p_i)\tau_p^{-1}} \tag{19}
\]

\[
\approx A(n - n_0) \text{ or } A(p - p_0) \tag{20}
\]

\[
R_{\text{rad}} = B(np - n_i^2) \tag{21}
\]

\[
R_{\text{Auger}} = C(n(np - n_i^2) + (np - n_i^2)p) \tag{22}
\]

Instead of the carrier concentrations \( n \) and \( p \), these rates can be rewritten in terms of the Fermi level separation, taken to be equal to the applied voltage \( qV \). In the dilute Boltzmann limit, the product \( np \) rises exponentially as with \( qV \) so that

\[
np = n_i^2 \left( e^{qV/k_B T} \right) . \tag{23}
\]

When the majority carriers are electrons, the fractional increase in this product is due to the minority carrier density. That is to say, the quasi-Fermi level of the majority species is relatively fixed while the quasi-Fermi level of the minority species is moved closer to the band edge states, increasing that carrier density. Thus, to a good approximation

\[
p = p_0 \text{ and } n = n_0 \left( e^{qV/k_B T} \right) \text{ where } p \gg n \text{ and } \tag{24}
\]

\[
n = n_0 \text{ and } p = p_0 \left( e^{qV/k_B T} \right) \text{ where } n \gg p. \tag{25}
\]

Substituting these expressions into Equation 19, where \( p \gg n \) we have

\[
R_{\text{SRH}} = \frac{n_i^2 \left( e^{qV/k_B T} - 1 \right)}{(p_0 + p_i)\tau_n^{-1}} = \frac{n_i^2 \left( e^{qV/k_B T} - 1 \right)}{2n_0^2 \tau_n^{-1}} \tag{26}
\]

\[
= A n_0 \left( e^{qV/k_B T} - 1 \right) \tag{27}
\]

\[
R_{\text{rad}} = B n_i^2 \left( e^{qV/k_B T} - 1 \right) \tag{28}
\]

\[
R_{\text{Auger}} = C p_0 n_i^2 \left( e^{qV/k_B T} - 1 \right) , \tag{29}
\]

where we have assumed the states contributing to SRH recombination are near the zero-bias equilibrium Fermi level and the trap lifetimes \( \tau_n^{-1} \) and \( \tau_p^{-1} \) were on the same order. A similar expression for \( R_{\text{SRH}} \) can be derived when \( n \gg p \). At some point along the junction, \( n \) is on the order of \( p \). Here we can write simple expressions for \( n \) and \( p \) in terms of the carrier asymmetry \( \alpha \equiv n_0/(n_0 + p_0) \).

\[
p = p_0 \left( e^{\alpha qV/k_B T} \right) \text{ and } n = n_0 \left( e^{(1-\alpha)qV/k_B T} \right) \text{ where } p \sim n \tag{30}
\]
Substituting as before, for regions with \( p \sim n \), we have:

\[
R_{SRH} = \frac{n_i^2}{n_0} \left( e^{qV/k_B T} - 1 \right) \quad (31)
\]

\[
R_{rad} = B n_i^2 \left( e^{qV/k_B T} - 1 \right) \quad (32)
\]

\[
R_{Auger} = C \cdot \left[ n_0 \left( e^{(1-\alpha)qV/k_B T} \right) + p_0 \left( e^{\alpha qV/k_B T} \right) \right] \cdot n_i^2 \left( e^{qV/k_B T} - 1 \right) . \quad (33)
\]

Now let us imagine a device in which the active region extends from \( x = 0 \) to \( x = L \) and consider three separate regions: \( p \gg n \) over \((0,X_p)\), \( p \sim n \) over \((X_p,X_n)\), and \( n \gg p \) over \((X_n,L)\). Now if we assume the extraction and injection efficiencies to be independent of applied bias, we can capture the voltage dependence of the quantum efficiency:

\[
\eta_{EQE} \propto \frac{\langle R_{rad} \rangle_{active}}{\langle R_{SRH} \rangle_{active} + \langle R_{rad} \rangle_{active} + \langle R_{Auger} \rangle_{active}} \quad (34)
\]

From the above equations for \( R_{SRH} \), \( R_{rad} \), and \( R_{Auger} \), it is clear that all three recombination processes have nonzero contributions at linear order in \( qV/k_BT \).

This may at first seem counter-intuitive, because we typically think of defect-based SRH recombination as a one-particle process, radiative bimolecular recombination as a two-particle process, and non-radiative Auger recombination as a three-particle process. While this is true, not all of the particles in these processes need to be excess particles. Some can be thermally-generated equilibrium carriers that exist when the device is off but at finite temperature. In fact, if we were to consider the situation at 0 Kelvin with no thermally-generated equilibrium carriers, we should not expect \( \eta > 1 \) operation to be possible, since the low-temperature reservoir would be at absolute zero and have no entropy.

The fact that radiative bimolecular recombination has a finite contribution at linear order in the dimensionless electrical excitation \( qV/k_BT \) implies that the external quantum efficiency of a very general class of LEDs remains a nonzero constant as \( V \to 0 \):

\[
\lim_{V \to 0} \eta_{EQE} \neq 0 . \quad (35)
\]

From this it follows that arbitrarily high wall plug efficiency is achievable at low voltage:

\[
\lim_{V \to 0} \eta = \lim_{V \to 0} \frac{\hbar \omega}{q V} \eta_{EQE} = \infty . \quad (36)
\]

This type of behavior, where unbounded coefficient of performance for heat pumping is available at arbitrarily low power is a general feature of thermodynamic heat pumps.

### 7 SUMMARY OF EXPERIMENTS

Here we present a summary of experimental evidence of the behavior of real LEDs in the low-bias regime. In Figure 7, we see that the basic claim about constant quantum efficiency at low bias is observed in a mid-infrared LED (\( \lambda \approx 2.1 \mu m \)) at 25°C and at various temperatures.

In Figure 8 and Figure 9, we see that the final conclusion from § 6 is confirmed. Since \( \eta_{EQE} \) is independent of voltage in this regime, \( \eta \propto 1/L \). At elevated temperatures, \( \exp \left[ -\hbar \omega/k_BT \right] \) is larger and sufficient power is available in this regime to observe \( \eta > 1 \) operation. Although the cooling power is very limited, these measurements constitute observation of electro-luminescent cooling.

### References

Figure 7: The quantum efficiency of a conventional LED approaches a constant as the applied voltage falls below \( k_B T/q \) (\( \approx 26\text{mV} \) at room temperature). This figure presents results from the experiment described in [10].

Figure 8: The wall-plug efficiency of a conventional LED scales inversely with the output power at low voltages. At elevated temperatures, the LED tested here produced a sufficient flux of 2.42\(\mu\text{m} \) photons at 135\(^\circ\text{C} \) to confirm that this scaling continues beyond the conventional limit of 100\% wall-plug efficiency. Figure from [10].
Figure 9: At further elevated temperatures, the same device from Figure 8 was observed to generate light with efficiency $\eta \gg 1$ (at left). In this low-bias cooling regime, the cooling power versus current (at right) is parabolic in analogy to a thermo-electric cooler. Figure from [29].


