Mathematical modeling of thermal runaway in semiconductor laser operation

W. R. Smith

Department of Mathematics and Computing Science, Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 18 August 1999; accepted for publication 14 March 2000)

A mathematical model describing the coupling of electrical, optical and thermal effects in semiconductor lasers is introduced. Through a systematic asymptotic expansion, the governing system of differential equations is reduced to a single second-order boundary value problem. This highly nonlinear equation describes the time-independent maximum temperature in the boundary layer adjacent to the mirror facet. The solution of the problem is a multi-valued function of current. The graph of the maximum steady-state temperature as a function of current gives a fold-shaped response curve, which indicates that no bounded steady state exists beyond a critical value of current. For certain device parameters and initial conditions, thermal runaway occurs. A mechanism for the sudden mode of semiconductor laser failure is described in terms of thermal runaway.

I. INTRODUCTION

The basic property of thermal runaway is that the temperature blows up at some point in the domain at a finite time for the appropriate choices of parameters and initial conditions. The most celebrated example of thermal runaway occurs in combustion theory (see Ref. 1 and references therein). The physical reason for this is the positive feedback associated with the exponential source term, which occurs in the large activation-energy limit of the Arrhenius law of chemical kinetics. The use of microwaves to sinter or join ceramics is another application of thermal runaway. The physical process providing the positive feedback here is the highly nonlinear temperature dependence of the photon absorption.

A schematic cross section of a typical semiconductor laser is shown in Fig. 1. Semiconductor lasers have been developed for many applications, for example, optical fiber transmission systems. However, these devices suffer from new types of degradation and failure mechanisms. These degradation modes have been split into three categories, based on the rate of change of the device characteristics, namely rapid, gradual and sudden (see Ref. 4 and references therein). The gradual mode is attributed to an increase in the density of point defects at the mirror facet. This gradual degradation, which takes place on a time scale of ten years, is responsible for a loss of efficiency. Moreover, all three modes are associated with thermal effects, but this link is not well understood.

There has recently been a considerable amount of work directed at the understanding of this heating process. Two-dimensional (transverse and lateral) steady-state results have been obtained for the system of equations given by electrical models (electron continuity, hole continuity and Poisson’s equation), optical models (wave equation and photon rate equation) and thermal models (heat equation); see, for example, Refs. 5 and 6. These space-dependent models are currently exclusively solved by numerical approaches, despite the wide variation of length scales. The numerical simulations require extensive computer resources, especially when parameter studies are required.

A fully lumped model, which comprises four ordinary differential equations, has recently been reported.7 Asymptotic solutions of this lumped model explain the link between thermal effects and loss of efficiency. However, the steady-state temperature is a monotonic function of the current and spatial variation of thermal effects cannot be described. Subsequently we introduced a one-dimensional model, which comprises four partial differential equations and one parametric ordinary differential equation, to describe the longitudinal variation of thermal effects.8 Asymptotic solutions of this model have been obtained on the short and long thermal time scales as with the lumped model. The increase in the density of point defects at the mirror (corresponding to degradation) produces a hot spot over the short length scale (in comparison to the cavity length) associated with the thermal conduction. The maximum steady-state temperature (in the hot spot) is a monotonic function of current.

The purpose of this article is to gain a better understanding of the failure mechanisms for semiconductor lasers. We extend the work done on hot spots by considering the effect of a scaling for the higher current, the higher photon density, the higher density of defects and the higher temperature. These scalings result in a highly nonlinear boundary value problem to describe the steady-state temperature in the hot
spot. This equation admits a maximum temperature which is a multi-valued function of current, a result which explains the runaway phenomenon seen experimentally when the current is increased.4

This article is not concerned with the reduction of light intensity which may be explained by the temperature dependence of the threshold current.

The contents of this article will now be outlined. A one-dimensional model is formulated in Sec. II. These equations are nondimensionalized in Sec. III. In Sec. IV, an asymptotic analysis of the model on thermal time scales is undertaken. Section V examines the highly nonlinear boundary value problem which describes the maximum temperature at steady state. Section VI describes a numerical solution of the full system of equations to illustrate the predictions of the analysis. Finally, Sec. VII briefly draws some conclusions.

II. PROBLEM FORMULATION

This section outlines the inclusion of thermal effects into the traveling-wave rate equations (see Ref. 9). We have

$$\frac{\partial N}{\partial t} = \frac{J}{ed} - a(N - N_i \exp(-E_{\text{int}}/k_BT)) (I^+ + I^-)$$

$$- N(A_m \exp(-E_1/k_BT_1) + B N \exp(-E_2/k_BT_1) + CN^2 \exp(-E_3/k_BT_1)),$$

$$\frac{\partial I^z}{\partial t} + \frac{c}{\mu_e} \frac{\partial I^z}{\partial z} = \left[ a(N - N_i \exp(-E_{\text{int}}/k_BT_1)) - \frac{c \alpha_{\text{int}}}{\mu_e} \right] I^z$$

$$+ B N^2 \exp(-E_2/k_BT_1),$$

where $N(z,t)$ is the concentration of electrons in the conduction band, $I^+(z,t)$ is the light intensity of the wave traveling in the positive $z$ direction and $I^-(z,t)$ is the light intensity of the wave traveling in the negative $z$ direction. The term $t$ is time, $z$ is the axial length, the constant $J$ is the current density per unit axial length, $e$ is the charge on an electron, $d$ is the thickness of the active layer, $c$ is the velocity of light in a vacuum and $\mu_e$ is the group refractive index of the active layer.

In Eqs. (1) and (2) the optical gain has been modeled by an expression of the form $G(N,T_1) = a(N - N_i \exp(-E_{\text{int}}/k_BT_1))$ where $a$ is the linear gain rate, $k_B$ is Boltzmann’s constant, $T_1$ is the temperature per unit axial length of the active layer, $E_{\text{int}}$ is the activation energy and $N_i \exp(-E_{\text{int}}/k_BT_1)$ is the electron density at transparency. The linear dependence of the optical gain on the carrier density and the Arrhenius temperature dependence of absorption are adopted on the basis of the experimental evidence (see Refs. 10 and 11, respectively). We neglect any temperature dependence of $a$ (cf. Ref. 5).

The carrier lifetime, $\tau_c$, is given by (see Ref. 7) $1/\tau_c = A_m(z) \exp(-E_1/k_BT_1) + B N \exp(-E_2/k_BT_1) + CN^2 \exp(-E_3/k_BT_1)$ where $A_m(z)$ is a function (discussed below) and $B$ and $C$ are constant. The first term on the right-hand side represents surface and defect recombination with activation energy $E_1$, the second radiative recombination with activation energy $E_2$ and the third Auger recombination with activation energy $E_3$. The radiative recombination can be split into spontaneous emission which enters the lasing mode and spontaneous emission which is absorbed in the surrounding as heat. The constant $B$ represents the fraction of spontaneous emission which enters the lasing mode. The electrons lost to the valence band by nonradiative and Auger recombination are assumed to have converted their energy into heat in the active layer.

The function $A_m(z)$ represents the defects in the semiconductor caused by degradation. The rate of degradation is governed by mechanisms such as oxidation and fatigue. In general these processes are temperature dependent and take place over much larger time scales (ten years) than those considered here; a typical model being an Arrhenius law [see (3.37) of Ref. 4]. The association of degradation with thermal effects indicates that it is appropriate to consider it on the thermal length scale. This is in contrast to surface defects...
which are usually assumed to be on the molecular length scale. The function $A_{\text{int}}(z)$ will be assumed to be of the form

$$A_{\text{int}}(z) = A_{e} \left\{ \begin{array}{ll} 1 + S(z) & z \leq \gamma, \\ 1 & \gamma < z < L - \gamma, \\ 1 + S(L - z) & L - \gamma \leq z, \end{array} \right.$$  

where $L$ is the cavity length, $S(z)$ represents the increased density of defects in the neighborhood of the surface, $\gamma$ is the penetration depth of these defects and $A_{e}$ is the density of defects in the cavity of the laser. The defects will be taken to be of the form $S(z) = \alpha(\gamma - z)/\gamma$ where $\alpha$ is a measure of the maximum density of defects. In the analysis which follows, we also take the slightly simpler form $S(z) = \alpha$ in order to obtain asymptotic solutions.

The rate of scattering loss at heterostructure interfaces is given by $c c_{\text{int}}/\mu_{g}$ where $c_{\text{int}}$ is the absorption constant. The photons in the lasing mode which are scattered at heterostructure interfaces are all assumed to be turned into heat in the surround. In the absence of experimental data, we neglect any temperature dependence of $\beta$ and $\mu_{g}$.

Thermal effects can then be coupled into the single-mode rate equations via

$$\rho_{1}c_{1} \frac{\partial T_{1}}{\partial t} - \frac{\partial}{\partial z} \left( k_{1} \frac{\partial T_{1}}{\partial z} \right)$$

$$= (A_{e} \exp(-E_{1}/k_{B}T_{1}) + CN^{2} \exp(-E_{1}/k_{B}T_{1}))$$

$$\times (1 - 2\beta) \frac{\kappa_{1}}{\Omega_{1}} (T_{2} - T_{1}),$$

$$\rho_{2}c_{2} \frac{\partial T_{2}}{\partial t} - \frac{\partial}{\partial z} \left( k_{2} \frac{\partial T_{2}}{\partial z} \right)$$

$$= c \frac{\kappa_{1}E_{g}\Omega_{1}}{\mu_{g}\Omega_{2}} (I^{+} + I^{-}) + (1 - 2\beta) \frac{BN^{2}E_{g}\Omega_{1}}{\Omega_{2}}$$

$$\times \exp(-E_{2}/k_{B}T_{1}) + \frac{\kappa_{1}}{\Omega_{2}} (T_{1} - T_{2}) + \frac{\kappa_{2}}{\Omega_{2}} (T_{a} - T_{2}),$$

where $\rho_{1}$ is the density, $c_{1}$ the specific heat capacity and $k_{1}$ the thermal conductivity of the active layer, $E_{g}$ is the bandgap energy, $\kappa_{1}$ is the heat transfer coefficient between the active and the surrounding layers, $T_{2}$ is the temperature per unit axial length, $\rho_{2}$ the density, $c_{2}$ the specific heat capacity and $k_{2}$ the thermal conductivity of the surrounding layer, $\kappa_{2}$ is the heat transfer coefficient between the surrounding layer and the material outside the laser, $T_{a}$ is the (constant) ambient temperature, $\Omega_{1}$ is the cross-sectional area of the active layer and $\Omega_{2}$ is the cross-sectional area of the surround. The first term on the right-hand side of Eq. (3) is the energy generated by the nonradiative recombination of electrons and holes in the active region. The first term on the right-hand side of Eq. (4) represents the heat generated from the absorption of the scattering loss at heterostructure interfaces. The second term on the right-hand side of Eq. (4) represents the heat generated from the absorption of the spontaneous emission which deviates outside the active region. The remaining terms on the right-hand side of Eqs. (3) and (4) model the transport of heat.

The boundary conditions for the intensities are given in terms of the reflectivities $R^{(1)}$ at $z = 0$ and $R^{(2)}$ at $z = L$ by the expressions

$$I^{+}(0, t) = R^{(1)}I^{-}(0, t), \quad I^{-}(L, t) = R^{(2)}I^{+}(L, t).$$

Radiation and convection through the sides of the laser are assumed to be negligible, so we have

$$\frac{\partial T_{1}}{\partial z}(0, t) = \frac{\partial T_{2}}{\partial z}(0, t) = \frac{\partial T_{2}}{\partial z}(L, t) = 0.$$  

We neglect any temperature dependence of $R^{(1)}$ and $R^{(2)}$. Further details of the modeling are given in Ref. 8.

The threshold current $[J_{\text{th}}(T_{1})]$ and electron concentration at threshold $[N_{\text{th}}(T_{1})]$ are given by the solutions of the equations

$$\frac{c}{2\mu_{g}} \ln \left( \frac{1}{R^{(1)}R^{(2)}} \right) = \int_{z = 0}^{L} a(N_{\text{th}}(T_{1}) - N_{t})$$

$$\times \exp(-E_{\text{int}}/k_{B}T_{1}) - \frac{c \alpha_{\text{int}}}{\mu_{g}} dz,$$

$$\frac{J_{\text{th}}(T_{1})}{ed} = A_{e}J_{\text{th}}(T_{1}) \exp(-E_{1}/k_{B}T_{1}) + 2B_{n}N_{\text{th}}(T_{1})$$

$$\times \exp(-E_{2}/k_{B}T_{1}) + CN_{\text{th}}(T_{1})^{2}$$

$$\times \exp(-E_{3}/k_{B}T_{1}).$$

The ratio of two threshold currents at different temperatures is given by

$$\frac{J_{\text{th}}(T_{1})}{J_{\text{th}}(T_{a})} = \frac{A_{e}N_{\text{th}}(T_{1}) \exp(-E_{1}/k_{B}T_{1}) + 2B_{n}N_{\text{th}}(T_{1})^{2} \exp(-E_{2}/k_{B}T_{1}) + CN_{\text{th}}(T_{1})^{3} \exp(-E_{3}/k_{B}T_{1})}{A_{e}N_{\text{th}}(T_{a}) \exp(-E_{1}/k_{B}T_{a}) + 2B_{n}N_{\text{th}}(T_{a})^{2} \exp(-E_{2}/k_{B}T_{a}) + CN_{\text{th}}(T_{a})^{3} \exp(-E_{3}/k_{B}T_{a})}.$$
III. NONDIMENSIONALIZATION

We define \( N_e \) and \( I_e \) to be representative values for the average electron concentration and the average photon density at fixed temperature \( T_1 = T_o \) and current greater than the threshold current (described in Ref. 8). The values of \( N_e \) and \( I_e \) are calculated using the formulas:

\[
N_e = N_e(\exp(-E_i/k_BT_o) + C\exp(-E_2/k_BT_o))^{1/2} \times \exp(-E_i/k_BT_o)
\]

\[
I_e = \frac{J}{e} - A_m \exp(-E_i/k_BT_o)N_e - B \exp(-E_2/k_BT_o)N_e^2 - C \exp(-E_2/k_BT_o)N_e^3
\]

\[
I_e = \frac{2(J/e - A_m \exp(-E_i/k_BT_o)N_e - B \exp(-E_2/k_BT_o)N_e^2 - C \exp(-E_2/k_BT_o)N_e^3)}{2a(N_e - N_e(\exp(-E_i/k_BT_o)))}
\]

\( \lambda = \frac{\partial \hat{N}}{\partial t} = A - \frac{B}{\Gamma} \hat{g}(\hat{N}, \hat{T}_1)(\hat{i}^+ + \hat{i}^-) - (A_1 \hat{A} \hat{N})(-A_1, \hat{T}_1)
\]

\[
+ a_2 \hat{N}^2 \hat{f}(-A_2, \hat{T}_1) + a_3 \hat{N}^3 \hat{f}(-A_3, \hat{T}_1)
\]

\( \frac{\partial \hat{I}}{\partial t} = \frac{D}{\nu} \frac{\partial \hat{I}}{\partial z} = \nu \hat{g}(\hat{N}, \hat{T}_1) - \epsilon \hat{I}
\]

\[
+ \gamma \hat{N}^2 \hat{f}(-A_2, \hat{T}_1)
\]

\( \frac{\partial \hat{T}_1}{\partial t} = \frac{\partial^2 \hat{T}_1}{\partial z^2} = \gamma \hat{K} \hat{N} \hat{f}(-A_1, \hat{T}_1)
\]

\[
+ \gamma \hat{N}^2 \hat{f}(-A_3, \hat{T}_1) \hat{T}_2 - \hat{T}_1
\]

\( \frac{\partial \hat{T}_2}{\partial t} = \frac{\partial^2 \hat{T}_2}{\partial z^2} = \gamma \hat{S} \hat{N} \frac{\partial \hat{T}_1}{\partial z} \hat{T}_2 - e \hat{T}_2
\]

where \( \hat{f}(\Lambda, \hat{T}_1) = \exp[N(1 + \delta \hat{T}_1)] \) and \( \hat{g}(\hat{N}, \hat{T}_1) = \hat{N} - N^* \hat{f}(\Gamma \hat{T}_1, \hat{T}_1) \) with the boundary conditions:

\( \hat{I}(0, \hat{i}) = R(1)^{\hat{i}}(0, \hat{i}), \quad \hat{T}_1(0, \hat{i}) = R(2)^{\hat{i}}(0, \hat{i}) \)

\( \frac{\partial \hat{T}_1}{\partial z}(0, \hat{i}) = \frac{\partial \hat{T}_2}{\partial z}(0, \hat{i}) = \frac{\partial \hat{T}_2}{\partial z}(1, \hat{i}) = 0 \).

The dimensionless constants \( A, a, A_1, a_2, A_2, A_3, B / \Gamma, N^*, \Gamma, \delta, \nu, D / \nu, \hat{\nu}, \gamma / \nu, \Gamma, C, \hat{\nu}, \gamma, \gamma / L \), \( \kappa / L \), \( A_{\nu} / \nu, \gamma_{\nu} / \nu, \gamma_{\nu} / L \) are defined in Table I; the time scale chosen being \( \tau_e = 1/A_{\nu} \) (one of the two time scales on the right-hand side of Ref. 8). The conditions \( \nu \ll 1, \Gamma \ll 1, \nu / \nu \ll 1, \delta / \Gamma \ll 1 \) typically hold in practice.

The small parameters are \( \nu \), representing the ratio of the optical time scale to the electrical time scale; \( \lambda \), the ratio of the electrical to the short thermal time scale; \( \epsilon \), the ratio of the short thermal time scale to the long thermal time scale; \( \sigma \), the ratio of the thermal boundary layer length-scale to the cavity length; \( \delta \), the typical temperature rise in the active region relative to the ambient temperature; and \( \Gamma \), the dimensionless activation energy for absorption.

IV. ASYMPTOTIC ANALYSIS

A. Introduction

There are several disparate time scales and two length scales in the problem (8)–(12). We study these equations on

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Typical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A/\Gamma )</td>
<td>( J \tau_e / N_e \epsilon d )</td>
<td>8</td>
</tr>
<tr>
<td>( a )</td>
<td>( 3 \times 10^2 )</td>
<td></td>
</tr>
<tr>
<td>( a_1 )</td>
<td>( A_{\tau} \gamma )</td>
<td>1</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>( B_{\tau} \gamma )</td>
<td>1</td>
</tr>
<tr>
<td>( a_3 )</td>
<td>( C \gamma \tau )</td>
<td>0.3</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>( E_1 / k_BT_o \gamma )</td>
<td>1</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>( E_2 / k_BT_o \gamma )</td>
<td>1</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>( E_3 / k_BT_o \gamma )</td>
<td>1</td>
</tr>
<tr>
<td>( B/\Gamma )</td>
<td>( a_\nu / \nu \gamma )</td>
<td>10</td>
</tr>
<tr>
<td>( N^* )</td>
<td>( N_e \exp(-E_m/k_BT_o) \gamma / \nu )</td>
<td>0.7</td>
</tr>
<tr>
<td>( \nu )</td>
<td>( 1 / a_{\nu} \gamma )</td>
<td>( 1 \times 10^{-4} )</td>
</tr>
<tr>
<td>( D/\nu )</td>
<td>( c_{\nu} / \mu_L \gamma )</td>
<td>( 2 \times 10^3 )</td>
</tr>
<tr>
<td>( \hat{\nu} / \gamma )</td>
<td>( c_{\nu} / \gamma \nu )</td>
<td>( 1 \times 10^8 )</td>
</tr>
<tr>
<td>( \gamma / L )</td>
<td>( \beta \gamma \nu \tau \gamma )</td>
<td>( 8 \times 10^{-2} )</td>
</tr>
<tr>
<td>( \kappa / L )</td>
<td>( A_{\gamma}N_e \gamma_{\gamma} / \nu \gamma )</td>
<td>( 4 \times 10^{-3} )</td>
</tr>
<tr>
<td>( A_{\nu} / \nu )</td>
<td>( A_{\gamma}N_e \gamma_{\gamma} / \nu \gamma )</td>
<td>( 2 \times 10^{-4} )</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>( N_e \gamma_{\gamma} )</td>
<td>( 3 \times 10^{-4} )</td>
</tr>
<tr>
<td>( \tau_e / \nu )</td>
<td>( (1 - 2\beta) \beta \gamma \nu \tau \gamma )</td>
<td>( 2 \times 10^{-4} )</td>
</tr>
<tr>
<td>( \gamma_{\nu} / \nu )</td>
<td>( \kappa_{\gamma} \gamma_{\gamma} / \nu \gamma )</td>
<td>( 1 \times 10^{-4} )</td>
</tr>
<tr>
<td>( R(1) )</td>
<td>( 0.3 )</td>
<td></td>
</tr>
<tr>
<td>( R(2) )</td>
<td>( 0.3 )</td>
<td></td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>( k_\tau / \nu )</td>
<td>( 2 \times 10^{-5} )</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>( \kappa_{\gamma} \gamma_{\gamma} / \nu \gamma )</td>
<td>( 6 \times 10^{-4} )</td>
</tr>
</tbody>
</table>
the short $[\hat{t} = O(1)]$ and long $[\hat{t} = O(1/\epsilon)]$ thermal time scales. Both of the length scales are required in the analysis which follows.

The leading order term for temperature in Ref. 8 satisfied an outer problem and an inner problem close to the mirror facet. The extra term introduced in the boundary layer corresponding to axial thermal conduction. The temperature was much higher in the boundary layer than in the body of the semiconductor due to the heating mechanism associated with defects in the neighborhood of the surface, the boundary layer being described as a hot spot. As the current is increased, the temperature in the hot-spot also increases. The temperature dependence of the absorption, which was previously treated as a lower order term, now enters the leading order problem. This new balance requires a new layer in which the temperature and the number of defects are scaled in terms of the small parameter $\Gamma$. In summary, the leading order problem on the thermal time scales comprises an outer problem and two layers close to the mirror facet.

**B. Short thermal time scale**

1. **Outer expansion**

The required expansions take the form $\hat{N} - N_0, \hat{I} \sim F_0, I^- \sim B_0, \hat{T}_1 - T_0(1)$ and $\hat{T}_2 = O(\epsilon)$. We obtain

$$F_0 = C \exp \left[ \int_{x=0}^{1} \left[ N_0 - N^* - \mathcal{E} \right] dx / D \right],$$

$$B_0 = C \sqrt{\frac{R(2)}{R(1)}} \exp \left[ \int_{x=z}^{1} \left[ N_0 - N^* - \mathcal{E} \right] dx / D \right]$$

and $N_0$ given by

$$B(N_0 - N^*)(F_0 + B_0) = A,$$

where $(C \neq 0$ because $A \neq 0$).

$$\frac{1}{2} \ln \left( \frac{1}{R(1)R(2)} \right) = \int_{x=0}^{1} \left[ N_0 - N^* - \mathcal{E} \right] dx / D. \quad (13)$$

Equation (10) implies the general solution $T_0^{(1)}$ = $[\mathcal{K} \hat{N}_0 e^{-A_1} + \mathcal{L} N_0^3 e^{-A_3}] + g(\hat{z}) e^{-\hat{z}}$. Therefore, the steady state in this outer problem is stable for all choices of initial condition consistent with these scalings. The first term in the expansion for $\hat{T}_1$ does not satisfy the boundary conditions, in general, and there are boundary layers at $\hat{z} = 0$ and $\hat{z} = 1$. We consider the layers at $\hat{z} = 0$ in the following two subsections, the layers at $\hat{z} = 1$ being similar.

2. **Layer I at $\hat{z} = 0$**

We perform the stretching transformation $\hat{z} = \sigma Z_1$ in the boundary layer and let $\hat{N} - n_0, I^+ - f_0, I^- - b_0, \hat{T}_1 - \theta_0/\Gamma, \hat{T}_2 = O(\epsilon)$ and $\hat{A} - A_0 / \Gamma$. The scaling on $\hat{A}$ corresponds to the increased density of defects in the neighborhood of the surface. Equation (9) implies $f_0 = F_0(0), b_0 = B_0(0)$ and Eq. (8) then gives

$$n_0 = \frac{A + B N^* \exp(\theta_0)(f_0 + b_0)}{a_1 A_0 \exp(-A_1) + B (f_0 + b_0)}.$$

We thus obtain

$$\frac{\partial \theta_0}{\partial \hat{t}} + \theta_0 - \frac{\partial^2 \theta_0}{\partial \hat{z}^2} = \mathcal{K} A_0 \exp(-A_1) \times \frac{A + B N^* \exp(\theta_0)(f_0 + b_0)}{a_1 A_0 \exp(-A_1) + B (f_0 + b_0)}$$

(15)

with one boundary condition at $Z_1 = 0$ given by $\partial \theta_0 / \partial \hat{z} = 0$ and the other to be determined by matching. We note that the exponential nonlinearity on the right-hand side of Eq. (15) was not present in the asymptotic analysis of Ref. 8.

3. **Layer II at $\hat{z} = 0$**

We perform the stretching transformation $\hat{z} = \sigma Z_2$ in the boundary layer and let $\hat{N} - \tilde{n}_0, I^+ - \tilde{f}_0, I^- - \tilde{b}_0, \hat{T}_1 - \tilde{\theta}_0/\Gamma$ and $\hat{T}_2 = O(\epsilon)$. Equation (8) implies $\tilde{n}_0 = [A + B N^*(f_0 + b_0)] / B (f_0 + b_0)$ and Eq. (10) then gives

$$\frac{\partial \tilde{\psi}_0}{\partial \hat{t}} + \psi_0 - \frac{\partial^2 \psi_0}{\partial \hat{z}^2} = \mathcal{K} \tilde{n}_0 \exp(-A_1) + \mathcal{L} \tilde{n}_0^3 \exp(-A_3)$$

with boundary conditions determined from matching with layer I and the outer expansion.

C. **Long thermal time scale**

This corresponds to the time scale of conduction in the surround and is the longest time scale in the problem. In Eqs. (8)–(11) we scale $\hat{t} = \tau / \epsilon$ to give

$$\gamma e \frac{\partial \tilde{N}}{\partial \tau} = \frac{A + B}{\mathcal{G} \tilde{N}} (\tilde{N}, \tilde{T}_1)(I^+ + I^-) - (a_1 \hat{A} \hat{N} \hat{f}(\hat{A}_1, \hat{T}_1)) + a_2 \hat{N}^2 \hat{f}(-\hat{A}_2, \hat{T}_1) + a_3 \hat{N}^3 \hat{f}(-\hat{A}_3, \hat{T}_1), \quad (16)$$

$$\gamma e \frac{\partial \tilde{I}^z}{\partial \tau} + \frac{D}{\nu} \frac{\partial \tilde{I}^z}{\partial \hat{z}} = \frac{1}{\nu} [g(\tilde{N}, \tilde{T}_1) - \mathcal{E}] \tilde{I}^z$$

$$+ \mathcal{G} \hat{N}^2 \hat{f}(-\hat{A}_2, \hat{T}_1), \quad (17)$$

$$e \frac{\partial \tilde{T}_1}{\partial \tau} - \sigma^2 \frac{\partial^2 \tilde{T}_1}{\partial \hat{z}^2} = \mathcal{K} \hat{A} \hat{N} \hat{f}(-\hat{A}_1, \hat{T}_1) + \hat{N}^3 \hat{f}(-\hat{A}_3, \hat{T}_1)$$

$$+ \hat{T}_2 - \hat{T}_1, \quad (18)$$

$$e \frac{\partial \tilde{T}_2}{\partial \tau} - \frac{D^2}{\nu^2} \frac{\partial^2 \tilde{T}_2}{\partial \hat{z}^2} = \mathcal{N} (I^+ + I^-) + \mathcal{P} \hat{N}^2 \hat{f}(-\hat{A}_2, \hat{T}_1)$$

$$+ \mathcal{Q} (\hat{T}_1 - \hat{T}_2) - \hat{T}_2. \quad (19)$$
1. Outer Expansion

The required expansions take the form \( \hat{N} \sim N_0, \quad I^+ \sim F_0, \quad I^- \sim B_0, \quad \hat{T}_1 \sim \Phi_0^{(1)} \) and \( \hat{T}_2 \sim \Phi_0^{(2)} \). Equation (18) implies \( \mathcal{K} \hat{N}_0 e^{-A_1} + \mathcal{L} n^0 e^{-A_3} + \phi_0^{(2)} - \phi_0^{(1)} = 0 \), and Eq. (19) then gives

\[
\frac{\partial \phi_0^{(2)}}{\partial \tau} + \phi_0^{(2)} - \mathcal{S} \frac{\partial^2 \phi_0^{(2)}}{\partial z^2} = -p(\hat{z})
\]

with

\[
\frac{\partial \phi_0^{(2)}}{\partial z}(0, \tau) = \frac{\partial \phi_0^{(2)}}{\partial z}(1, \tau) = 0,
\]

where

\[
G(\xi, \hat{z}) = \frac{1}{\sqrt{\xi}} \begin{cases} \cosh(\xi/\sqrt{S})\sinh(\hat{z}/\sqrt{S}) - \cosh(\hat{z}/\sqrt{S})\coth(1/\sqrt{S}) & \xi < \hat{z}, \\
\cosh(\hat{z}/\sqrt{S})\sinh(\xi/\sqrt{S}) - \cosh(\xi/\sqrt{S})\coth(1/\sqrt{S}) & \hat{z} < \xi 
\end{cases}
\]

and \( A_n \) are determined by matching with the short thermal time scale. The steady state in this outer problem is stable for all choices of initial condition consistent with these scalings.

We now consider the boundary layers at \( \hat{z} = 0 \) in a similar manner to the short thermal time scale.

2. Layer A at \( \hat{z} = 0 \)

We perform the stretching transformation \( \hat{z} = \alpha Z_1 \) in the boundary layer and let \( \hat{N} \sim \eta_0, \quad I^+ \sim f_0, \quad I^- \sim b_0, \quad \hat{T}_1 \sim \Theta_0/\Gamma, \quad \hat{T}_2 = O(1) \) and \( \hat{A} \sim A_0/\Gamma \). Equation (16) implies

\[
\eta_0 = \frac{A + B n^* \exp(\Theta_0)(f_0 + b_0)}{a_1 A_0 \exp(-A_1) + B(f_0 + b_0)}.
\]

We thus obtain

\[
\Theta_0 - \frac{\partial^2 \Theta_0}{\partial Z_1^2} = \mathcal{K} A_0 \exp(-A_1)
\]

\[
\times \left\{ \frac{A + B n^* \exp(\Theta_0)(f_0 + b_0)}{a_1 A_0 \exp(-A_1) + B(f_0 + b_0)} \right\}
\]

with one boundary condition at \( Z_1 = 0 \) given by \( \partial \Theta_0 / \partial Z_1 = 0 \) and the other to be determined by matching. We again note the exponential nonlinearity on the right-hand side of Eq. (20).

3. Layer B at \( \hat{z} = 0 \)

We perform the stretching transformation \( \hat{z} = \sigma \ln(1/\Gamma) + Z_2 \) in the boundary layer and let \( \hat{N} \sim \eta_0, \quad I^+ \sim f_0, \quad I^- \sim -b_0, \quad \hat{T}_1 \sim \Psi_0^{(1)} \) and \( \hat{T}_2 \sim \Psi_0^{(2)} \). Equation (18) implies \( \Psi_0^{(2)} = \psi_0^{(2)}(0, \tau) \) and Eq. (18) then gives

\[
\psi_0^{(1)} - \frac{\partial^2 \psi_0^{(1)}}{\partial Z_2^2} = \mathcal{K} \hat{N}_0 \exp(-A_1) + \mathcal{L} n^0 \exp(-A_3) + \psi_0^{(2)}
\]

with boundary conditions determined from matching with layer A and the outer expansion.

V. THERMAL RUNAWAY

A. Introduction

In this section we examine the maximum value of temperature at steady state. Moreover, we are interested in treating \( \mathcal{A} \) as a bifurcation parameter which corresponds to varying the current. In Ref. 8, the maximum temperature at steady state is a monotonic function of current and this does not explain the phenomenon of thermal runaway. The additional nonlinearities, which are introduced in layer I and layer A, admit a maximum temperature which is a multi-valued function of current.

We now derive the equation which describes the dependence of the maximum temperature on \( \mathcal{A} \), utilizing the equations for the short or long thermal time scales in Sec. IV. We use Eq. (13) to eliminate the photon densities from Eqs. (15) or (20) and note that the function \( A_0(Z_1) \) is only nonzero over a finite length, namely \([0, W])\). For \( Z_1 \geq W \), the steady-state temperature, \( \theta(Z_1) \), is given by \( \theta = P e^{-Z_1} + Q e^{Z_1} \) where \( P \) and \( Q \) are constants. Matching with layer II implies that \( Q = 0 \). Therefore we can prescribe the boundary condition at \( Z_1 = W \) by \( d \theta / d Z_1 = -\theta \). We summarize the nonlinear boundary value problem which describes the maximum temperature as follows:

\[
\frac{d^2 \theta}{d Z_1^2} = \theta - \bar{\mathcal{K}} A_0 A \left\{ \frac{N + N^* \exp(\theta)}{\bar{a} N A_0 + \mathcal{A}} \right\}
\]

where \( \bar{a} = a_1 \exp(-A_1), \quad \bar{\mathcal{K}} = \mathcal{K} \exp(-A_1) \) and \( \bar{N} = N_{0}(0) \).
The upper branch of the response curve becomes of Arrhenius temperature dependence in the expression for form intensity within the boundary layer and the assumption critical current is mathematically a consequence of the uni-
duction. The nonexistence of bounded solutions beyond the by nonradiative recombination than is lost by thermal con-
bounded steady state exists; that is, more power is deposited
in Fig. 2. Typical steady-state response diagram for the maximum temperature \( \theta_{\text{max}} \) plotted against the bifurcation parameter \( A \). The parameter values are given in Table I with \( A_0 = 4(W-Z_1) \) and \( W = 1 \).

The quantity \( N_0(0) \) is taken to be fixed because the physical solutions of Eqs. (13) and (14) correspond to \( A/C \) constant and \( N_0 \) independent of the choice of \( A \). In this section, we consider four examples of Eqs. (21) and (22) for different functions \( A_0(Z_1) \) and constants \( W \). In two cases, we will obtain numerical results with the AUTO bifurcation package and, in the other two cases, we will make asymptotic simpli-
ifications.

**B. \( W = O(1), A_0 = \chi(W-Z_1) \)**

First, we consider a particular version of Eqs. (21) and (22) with \( A_0 = 4(1-Z_1), W = 1 \) and the data given in Table I (the reason for the choice \( \chi = 4 \) will become apparent). The maximum value of \( \theta \) at steady state as a function of the bifurcation parameter \( A \) is shown in Fig. 2. It is clear from the figure that there is a current \( A_c \), beyond which no bounded steady state exists; that is, more power is deposited by nonradiative recombination than is lost by thermal con-
duction. The nonexistence of bounded solutions beyond the critical current is mathematically a consequence of the uni-
form intensity within the boundary layer and the assumption of Arrhenius temperature dependence in the expression for absorption. The upper branch of the response curve becomes invalid as \( \theta \rightarrow \infty \) because at these high temperatures the absor-
ption becomes very large and the intensity may no longer be taken as uniform across the boundary layer. Numerical simulations of Eqs. (8)–(12), with the appropriate choice of initial conditions, indicate that the lower branch is stable and the upper branch unstable.

We now examine the variation of the current, \( A_c \), beyond which no bounded steady state exists in response to changes in \( W \) and \( \chi \). The results of simulations with \( W = 1 \) and several values for \( \chi \) are shown in Fig. 3. For \( \chi \ll 1 \), the figure indicates that no fold bifurcation takes place, that is a bounded steady-state solution always exists. However, for \( \chi \gg 1 \), the current \( A_c \) decreases significantly only leaving a narrow range of currents available for stable operation. We now consider the results of simulations with \( \chi = 4 \) and se-
veral values for \( W \) (shown in Fig. 4). For \( W \ll 1 \), the figure indicates that a bounded steady state always exists. In the case \( W \gg 1 \), only a narrow range of currents is available for stable operation. This behavior is consistent with experimen-
tal observations of the sudden mode of failure.4

We also note that the maximum temperature is a multi-
valued function of the density of defects in the neighbor-
hood of the surface. The response diagram for fixed \( A = 1 \) with the bifurcation parameter \( \chi \) is shown in Fig. 5. The sudden mode of failure may be explained in terms of the slow increase in \( \chi \) which takes place during aging. The sudden failure occurs as \( \chi \) passes through \( \chi_c \), the critical defect density beyond which no bounded steady state exists.

The current is proportional to the optical power at leading order. The response diagram with bifurcation parameter taken to be optical power is a scaled version of Fig. 2.

**C. \( W = O(1), A_0 \) constant**

We consider an autonomous version of Eqs. (21) and (22) with \( A_0 = 4 \) and \( W = 1 \) for comparison. The response

![FIG. 2. Typical steady-state response diagram for the maximum temperature \( \theta_{\text{max}} \) plotted against the bifurcation parameter \( A \). The parameter values are given in Table I with \( A_0 = 4(W-Z_1) \) and \( W = 1 \).](image1)

![FIG. 3. The decrease in the current beyond which no bounded steady state exists in response to an increase in the density of defects in the neighborhood of the surface, using the data in Table I and \( W = 1 \). The parameter region above the curve corresponds to thermal runaway for all initial conditions.](image2)

![FIG. 4. The decrease in the current beyond which no bounded steady state exists in response to an increase in the length of the region containing a high density of defects, using the data in Table I and \( \chi = 4 \). The parameter region above the curve corresponds to thermal runaway for all initial conditions.](image3)
The fold point, \( \mathcal{A} \), is given by the unique positive real root of the equation \( N^\ast \exp(Y+1) = 1 \) where \( Y = kA_0 \mathcal{A} / (aNA_0 + \mathcal{A}) \). This outer solution does not satisfy the boundary condition at \( Z = 1 \) and there exists a boundary layer. The transverse and lateral thermal conduction balances with the thermal source term in this outer expansion; the axial thermal conduction will appear in the boundary layer.

Inner Expansion. We rescale the independent variable \( Z = 1 - \xi / \mathcal{W} \) and introduce an expansion of the form \( \bar{\theta} = \Phi_0(\xi) \). We obtain
\[
\frac{d^2 \Phi_0}{d \xi^2} = \Phi_0 - \bar{k}A_0 A \left( \frac{\bar{N} + N^\ast \exp(\Phi_0)}{aNA_0 + \mathcal{A}} \right)
\]
with
\[
\frac{d \Phi_0}{d \xi} = \Phi_0, \quad \frac{d \Phi_0}{d \xi} \to 0 \quad \text{as} \quad \xi \to \infty.
\]
The response diagram for Eqs. (21) and (22) with \( W = 10 \) and \( A_0 = 4 \) is again qualitatively similar to Fig. 2. The increase in \( W \) produces a fold at a lower value of current than in Sec. V C (\( \mathcal{A}_c = 0.1 \)).

VI. NUMERICAL SOLUTION

We discretize the spatial variable in Eqs. (8)–(11) leading to a system of ordinary differential equations. The convection terms in Eq. (9) are represented by a first-order upwind discretization. The diffusion terms in Eqs. (10) and (11) are approximated by the standard conservative central difference. The thermal diffusion coefficient in Eq. (10) is sufficiently small to only become significant in a thin region of space. We select a mesh which (i) accurately models the boundary layers at \( \tilde{z} = 0 \) and \( \tilde{z} = 1 \) and (ii) varies continuously throughout the domain. We chose a Bakhvalov mesh (see Ref. 12 and references therein) in the regions \( 0 \leq \tilde{z} < \gamma \mathcal{L} \) and \( 1 - \gamma \mathcal{L} < \tilde{z} \leq 1 \) to satisfy (i). We must be careful, in addressing (ii), not to significantly increase the number of mesh points. Define a uniform mesh \( x_1, x_2, \ldots, x_{N+1} \) on \( [-\pi/2, \pi/2] \) as the starting point. Then define a mapping \( \tilde{z} = f(x_j) \), where \( f(x) = 1/2 + (1/2 - \gamma \mathcal{L}) \sin(x) \), to produce a progressive refinement towards both boundary layers. The parameter values in Table I indicate that a number of very different time scales is present and the stiff ordinary differential equation solver D02EAF was accordingly selected from the NAG library. Numerical results have been obtained for a variety of meshes but it was not possible to obtain grid independence in the simulation of thermal runaway owing to the ill-posed nature of the problem. However, the results did remain qualitatively the same on the various mesh refinements and we present the results to illustrate the analysis. The numerical simulations, which attained bounded steady states, proved to be grid independent.

We present the stable steady state attained for the parameter values given in Table I. The steady-state temperature rise in the active region is shown in Fig. 6, higher temperatures being attained there than the surround. We note that the temperature rise in the surround does not exhibit the sharp maxima seen in the temperature rise of the active region. The maximum steady-state temperature rise in the active region predicted by the lower branch of the response diagram in Fig. 2 is 13 K, similarly a value of 13 K is predicted by the...
solution; the two approaches show good agreement. As a result of the temperature rise, the light intensity (shown in Fig. 7) decreases slightly. The electron concentration increases in response to these temperature rises with maxima being observed in the neighborhood of both mirrors.

We now seek numerical results for the parameter values given in Table I except that we take $A = 1$. This choice of parameter values corresponds to the regime where no bounded steady-state solutions exist (cf. Fig. 2). The simulation is started from the initial condition $\hat{N}(\hat{z},0) = 0$, $\hat{I}^-(\hat{z},0) = 0$, $\hat{I}^+(\hat{z},0) = 0$, and $\hat{T}^1(\hat{z},0) = 0$. Numerical results are given for two spatial meshes: one in which there are 100 points within each Bakhvalov mesh and 280 points in total (mesh A) and one in which there are 80 points within each Bakhvalov mesh and 220 points in total (mesh B). The results are qualitatively the same for the two meshes. The temperature rise of the active region on the mirror facet as a function of time is shown in Fig. 8. The temperature of the active region shows a gradual increase and this is accompanied by a gradual increase in electron concentration and a slow decrease in light intensity (shown in Fig. 9). The rate of change of temperature in the active region increases rapidly as the temperature in the active region attains a value where nonlinear effects dominate. The high temperatures result in a sudden reduction of photon density and increase in electron concentration at the mirror due to the strong temperature dependence of the absorption. The temperature rise in the active region and the surround as a function of longitudinal length is shown in Fig. 10. The temperature rise in the boundary layer is significantly higher than inside the cavity of the laser.

VII. SUMMARY AND CONCLUSIONS

A mathematical model has been introduced to describe the thermal runaway observed during semiconductor laser operation. The equations are scaled to reflect the effect of high current, photon density, density of defects and temperature. An asymptotic analysis results in a stable steady state in...
the body of the laser and in a highly nonlinear boundary value problem to describe the steady-state temperature in the hot spot. The response diagram for this boundary value problem takes the form of a fold in terms of the bifurcation parameter which represents current. Numerical simulations of the full time-dependent system determines that the lower branch is stable and the upper branch is unstable. 

We may also choose the bifurcation parameter to be the density of defects in the neighborhood of the surface and obtain a fold in the response diagram. We note that fold bifurcations have structural stability. The fold bifurcations obtained for the leading-order equations will therefore be characteristic of the complete system of equations. The subsequent terms in the asymptotic expansions merely represent imperfections. Moreover, this general behavior should also be observed in other (higher dimensional) models for semiconductor lasers with strong temperature dependence of the photon absorption.

The length scale over which the increased density of defects stretches is an important parameter in determining the critical current beyond which no bounded steady states exist. If this distance is significantly less than the thermal length scale \( (\sqrt{k_1\Omega_1/k_1}) \) then there is a stable and an unstable steady state. Thermal runaway only takes place when extreme initial conditions are prescribed. In contrast if this length scale is significantly greater than the thermal length-scale then there is only a narrow range of currents over which a stable steady state can be obtained. Similar comments apply to the magnitude of the increased density of defects in the neighborhood of the facet relative to the reference value \( A_k k_B T^2 / E_{int} \).

We conclude by offering an explanation of the sudden mode of failure in terms of thermal runaway. The sudden mode corresponds to the device setup initially lying inside the region where bounded steady states are allowed but the slow increase in defects (aging) moves the device parameters out of this region. The instant the device crosses this boundary a sudden increase in temperature takes place at the mirror facet and a sudden decrease occurs in laser output. The temperature at the facet can exceed the melting point of the laser crystal leading to the so-called catastrophic optical damage.

ACKNOWLEDGMENTS

Financial support for this work was provided by the EPSRC and the TMR network entitled "Differential Equations in Industry and Commerce." This project has benefited from the generous advice of J. R. King, B. Tuck and R. M. M. Mattheij.