Reprinted from Journal of the Optical Society of America B, Vol. 1, page 857, December 1984

Optical bistability and differential gain resulting from absorption increasing with excitation

D. A. B. Miller

AT&T Bell Laboratories, Holmdel, New Jersey 07733

Received January 16, 1984; accepted July 16, 1984

An extended theoretical discussion of a novel general method of obtaining optical bistability is given. The method uses absorption that increases as the material becomes more excited and requires no mirrors or other external feedback. Criteria for bistability and differential gain are derived for a simple case, and specific illustrative analytic results are given. Limiting relations are derived for the switching powers and the width of the bistable region. It is shown that this general class of bistability encompasses at least six previous independent and unrelated investigations of mirrorless bistability in specific physical systems. The common general requirements for this bistability are discussed together with its attributes.

1. INTRODUCTION

Optical bistability (OB) and the associated switching and differential gain are phenomena that can be obtained in a variety of ways. Several of these ways have been categorized[1]; it is usual for bistability to arise from the combination of a microscopic nonlinearity (such as intensity-dependent refractive index) and a macroscopic feedback (such as mirrors and cavities). However, as has recently been pointed out,[2] there have been several independent discussions[3–6] of OB in specific physical systems without mirrors or other external or macroscopic feedback that all appear to belong to a further, separate class of OB.[2] Indeed, there have been several more independent discussions that I now believe also belong to this class.[7–9] Extensions of some of these discussions are now available[10,11] as is a discussion of a related general class.[12] Recently, with the invention of the self-electro-optic effect device (SEED), a hybrid version of the same bistability has been demonstrated.[13] The principal unifying characteristic of this class is that the positive-feedback mechanism that leads to OB is internal to the material and results from optical absorption that increases as the material becomes more excited; the resulting increase in absorption makes the material yet more excited, giving yet more absorption, and so on. Consequently, this phenomenon has been called OB due to increasing absorption.[2] The inverting logic operation and the absence of mirrors or of resonant cavities make this class of OB particularly interesting from a practical standpoint.

The historical development of this class appears to be as follows: Kaplan[3] analyzed theoretically the behavior of a relativistic electron gas under optical illumination and deduced optically bistable behavior, although the extension of a new general principle was not made and the feedback mechanism resulting from increasing absorption was not explicitly recognized. Independently, Hajto and Janossy,[4] in investigations of optical effects in amorphous $GeSe_2$, observed several hysteretic and switching effects that were a result of a variety of mechanisms in this material. However, they identified a thermal mechanism for OB that was in qualitative agreement with some of their results, and they provided a simple model identifying the increasing-absorption mechanism. Again independently, Rossmann *et al.*[7] observed a hysteresis phenomenon in CdS platelets that they tentatively ascribed to an increasing absorptive-feedback mechanism. Further research reinforced this interpretation,[10] independently deriving a model similar to that in Ref. 4. Almost simultaneously with the research by Rossmann *et al.*, Bohnert *et al.*[6] independently observed a qualitatively similar hysteresis effect in CdS that they also independently ascribed to OB resulting from increasing absorption; this was followed by further analysis[11] (the connection with Ref. 7 was then noted). In yet another independent discussion, Hopf *et al.*,[5] in considering theoretically the effect of local field corrections on the behavior of a dense collection of two-level systems, found a possible mirrorless bistability. Then Miller *et al.*[2] pointed out that many of these previous discussions[3–6] seemed to be concerned with the same basic mechanism and discussed some of the general features of the class. In their paper,[2] they also presented a thermal-bistability experiment designed to test the mechanism, with other bistability mechanisms specifically excludable, and showed clear quantitative agreement between experiment and theory. This was followed by a hybrid demonstration of the same class of bistability in the SEED[13] with good agreement between theory and experiment. In other recent work, Dagenais[8] also independently discovered thermal bistability resulting from increasing absorption in CdS; Tooley and Seaton[9] observed bistability in InSb that is likely also to be thermal bistability resulting from increasing absorption, and Goldstone and Garmire[12] analyzed mirrorless refractive bistability considering also a general formalism in which their refractive bistability and the present absorptive bistability were special cases. This bistability mechanism has therefore apparently been described independently at least six times. Ironically, the reason for this may be the universality of the mechanism; for the most part, the descriptions have been in response to considerations of particular and apparently exceptional physical systems, so that the universality was unrecognized or at least unpublished.

The purpose of this paper is threefold. First, since nearly all the discussions of this OB have been independent, with most researchers unaware of the connection with other dis-

cussions, it is important to show the connection and the universality of the mechanism; in many cases, the connections are not obvious, and the previous discussion[2] was necessarily too brief to explain any of these connections in sufficient detail. Second, given the universality of the mechanism, it is important to establish what the necessary physical attributes of the material are (insofar as it is possible to prescribe these in the general case) and what the general characteristics of the bistable and differential-gain behavior are. Third, with the possibility of practical devices, it is important to understand the origins and the limits of the amplifying and the switching behavior. In this paper, we attempt to address all three of these issues and draw some conclusions in each case.

In Section 2 of this paper, the general theory of this bistability is described, and general relations for the existence of bistability and differential gain are derived. In Section 3, I give detailed exemplary results for a simple, analytically solvable example, and, in Section 4, I derive various simple limiting relations for switching powers and the width of the bistable region. In Section 5, the various other discussions of this class of bistability[2–13] are compared to show the connections and to identify the common features. Finally, in Section 6, I draw general conclusions from this research.

2. GENERAL FORMALISM

Suppose that we have a material whose absorption A increases as it becomes excited. We presume for the present analysis that the degree to which the material is excited can be adequately described by one parameter N; this is, of course, a simplification of any real physical situation, although it is a good approximation in many of the situations discussed. In the various systems discussed, N takes many different forms, such as temperature rise,[2–4,8,9] excited population density,[5–7,10,11] and voltage change,[13] and I will discuss this point more fully in Section 5. For the moment, we use the most general form and simply express A as a function of N:

$$A = A(N). \tag{1}$$

Now we assume that the degree of excitation depends in the steady state proportionately on the rate of absorption of energy. If a power P_{in} is incident upon the medium, we have

$$N = \eta A P_{in}. \tag{2}$$

Here, η is a proportionality constant. Note that we are assuming that all the absorption A contributes to exciting the medium and that the degree of excitation is directly proportional to the absorbed power. Neither of these is a necessary restriction: The current theory could readily be extended to account for background absorption that does not contribute to N and for other functional relations between N and absorbed power by appropriately altering Eq. (2). Such sophistications will not, however, be treated in this paper. Also, in this paper I treat only the steady-state solutions and do not discuss transient effects or stability; the stability of states in the bistable region has been checked experimentally,[2] and critical slowing down has also been seen.[2]

The general solution for A as a function of P_{in} is simple: If A is known as a function of N [Eq. (1)], solve Eqs. (1) and (2) simultaneously to eliminate N. This is a simple generalization of the model of Hajto and Janossy.[4] This model is readily visualized graphically. It is more useful and easier to compare

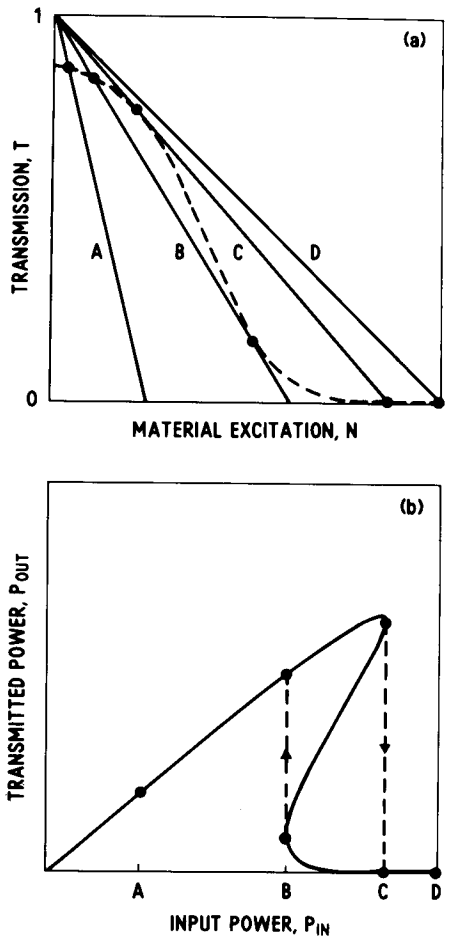

Fig. 1. Graphical solution of OB resulting from a hypothetical absorption increasing with increasing excitation. (a) The dashed line is $T = 1 - A(N)$ [Eq. (1)]. Lines A–D correspond to increasing power in Eq. (2). Lines A and D intersect only once with the curve, indicating only one solution for these powers. Lines B and C, each showing two intersections, represent the critical powers for the switch to low transmission and the switch to high transmission. Any lines between B and C would show three intersections as required for OB. (b) P_{out}, the transmitted power, is plotted against P_{in}, the incident power, using the solution method from (a). $\kappa, \eta = 1$ for simplicity.

with other types of bistability if we use transmission $T \equiv \kappa(1 - A)$ rather than absorption; κ is a constant allowing for all nonabsorptive losses (e.g., reflection and scattering). Because it does not affect the form of the solutions, we set $\kappa = 1$ for convenience in the rest of this paper. Output power P_{out} is $P_{out} = TP_{in}$. Figure 1 shows the graphical-solution method for a hypothetical $A(N)$. The transmission is deduced from the intersection points of the straight lines [from Eq. (2)] with the curve [from Eq. (1)]. Triple intersections of a line with the curve imply a bistable region at the power associated with this line.

The reason for the existence of bistability and/or differential gain can be explained through the following regenerative mechanism. Increasing power P_{in} results in increasing N [through Eq. (2)]. Increasing N gives increasing A [through Eq. (1)], which results in a further increase N [through Eq. (2)]. This process clearly feeds on itself (i.e., it is regenerative) and can result in enhancement of absorption changes (leading to differential gain) or in switching (giving bistability). The

general conditions for differential gain and bistability are derived below.

Note that A must depend on N and not merely on P_{in}; otherwise the regenerative loop is broken. This means that no conventional optical nonlinearity that can be rigorously expanded in a power series of electric field (in which the coefficients are the susceptibilities $\chi^{(n)}$) can give rise to these effects, unless at least one $\chi^{(n)}$ depends on the real state of excitation of the medium, N.[14] In the situations analyzed in this paper, we assume that the imaginary part of $\chi^{(1)}$ depends on N. Thus, for example, the nonlinearity associated with two-photon absorption could not cause this kind of bistability unless the two-photon absorption coefficient also depended on N. Such a situation could be handled by generalizing the present theory to allow for $A = A(N, P_{in})$. Nonlinearities in which the optical response depends on the real state of excitation rather than directly on the electric field are sometimes called dynamic,[15] and it is crucial that the nonlinearity be dynamic for this bistability. Another way of describing such optical nonlinearities would be to call them indirect, as the optical response does not depend directly on the electric field.

A. Condition for Amplifying Differential Gain

The differential gain is

$$\frac{dP_{out}}{dP_{in}} = \frac{d}{dP_{in}}[P_{in}(1 - A)]. \tag{3}$$

Now we have the identity

$$\frac{dA}{dP_{in}} = \frac{dA}{dN} \Big/ \frac{dP_{in}}{dN}.$$

Therefore, using Eq. (2) to obtain a sustitution for dP_{in}/dN, we obtain

$$\frac{dA}{dP_{in}} = \frac{\eta A^2}{\dfrac{A}{dA/dN} - N}. \tag{4}$$

By using Eq. (4) we therefore obtain from Eq. (3)

$$\frac{dP_{out}}{dP_{in}} = (1 - A) - \frac{NA}{\dfrac{A}{dA/dN} - N}. \tag{5}$$

Note that this is independent of η. The condition for amplifying inverted differential gain is $dP_{out}/dP_{in} < -1$. After rearrangement this condition becomes

$$N\frac{dA}{dN} > \frac{A(2 - A)}{2}. \tag{6}$$

This differential inequality can be solved to give the equivalent condition for amplifying inverted differential gain:

$$A_2 > \frac{2}{1 + \dfrac{N_1}{N_2}\left(\dfrac{2}{A_1} - 1\right)} \tag{7}$$

for some $A_1 = A(N_1)$, $A_2 = A(N_2)$, with $N_2 > N_1$ and $A_2 > A_1$. This condition means that if there is a range of N from N_1 to N_2 for which this condition is satisfied, then there will be differential gain overall in this region.

Although it is difficult to give a simple interpretation to condition (6) or (7), the requirement on $A(N)$ is sublinear in N, i.e., A proportional to N will give infinite differential gain, so the requirement for finite differential gain is A less than proportional to N.

One other simple analytic $A(N)$ is worth mentioning here if only because it is ultimately uninteresting, namely,

$$A(N) = 1 - \exp[-(\gamma N + B)], \tag{8}$$

where B and γ are positive constants. This represents an absorption coefficient (as distinct from the total fractional absorption A) that increases linearly with N. Since an absorption $A \propto N$ shows infinite differential gain at the critical power, it might be thought that this related nonlinearity ($A \propto N$ asymptotically for $B = 0$) might also show at least differential gain. In fact, it does not, as can be verified by substitution in relation (6). Even for $B = 0$, the maximum (negative) differential gain is -1.[16] Thus, in a system in which the absorption coefficient depends on N, a necessary condition even for amplifying differential gain is that the absorption coefficient be more than proportional to N.

B. Condition for Bistability

It is clear from the graphical visualization that the borderline condition for multiple intersections of straight line and a curve is that in which the straight line is tangent to the curve. If the curve is any steeper, multiple intersections must result. We can write this condition as

$$\frac{dA}{dN} > \frac{A}{N}. \tag{9}$$

This differential inequality is readily solved to give the equivalent condition

$$A_2 > \frac{N_2}{N_1}A_1, \tag{10}$$

where $A_2 > A_1$ and $N_2 > N_1$. That is, if there is a region from N_1 to N_2 for which this condition is satisfied, then there will be bistability in this region. This condition has a simple interpretation: For bistability, A must be more than linearly proportional to N over some region of N.

The borderline condition can also be derived from the expression for differential gain [Eq. (5)]. If $A \propto N$, the differential gain becomes infinite as expected for the onset of switching action.

3. ANALYTIC EXAMPLE

To illustrate the theory in a relatively clear way, we must choose a form for $A(N)$ that is not unphysical and yet is simple enough to be analytically solvable. We choose

$$A(N) = \frac{\beta}{1 + (\rho N - \sigma)^2}, \tag{11}$$

where β, ρ, and σ are constants with $0 < \beta < 1$ so that $0 < A < 1$. This corresponds to a Lorentzian absorption profile that shifts with increasing excitation N. ρ is a proportionality constant, and σ represents the original detuning of the operating point from line center. In the limit of small β and $N = 0$, this is the actual absorption that would be seen for a classical Lorentzian line. This is not meant to model any particular situation exactly, although it can, in practice, model the experiment in Ref. 2 and probably also that of Ref. 8 fairly well.

Substituting for N using Eq. (2), we obtain

$$A = \frac{\beta}{1 + (\rho\eta A P_{in} - \sigma)^2}. \quad (12)$$

This can be solved exactly for A in terms of P_{in} and hence for P_{out} in terms of P_{in}; the solution involves solving a cubic equation whose roots determine the one or three intersection points of straight line and curve of the graphical method. This can always be done but is tedious. Instead, we merely calculate analytically the switching points by using relation (9). At the switching points ($dA/dN = A/N$),

$$\rho N = (2\sigma/3)(1 \pm R), \quad (13)$$

where

$$R = \left[1 - \frac{3}{4}\frac{(1+\sigma^2)}{\sigma^2}\right]^{1/2}.$$

The corresponding absorption at the switching points is [using relation (10)]

$$A = \frac{\beta}{1 + (\sigma^2/9)(1 \pm 2R)^2}, \quad (14)$$

and the switching powers are [using Eqs. (2), (11), and (13)]

$$P_s = \frac{2}{3}\frac{\sigma}{\beta\rho\eta}(1 \pm R)\left[1 + \frac{\sigma^2}{9}(1 \mp 2R)^2\right]. \quad (15)$$

The switching powers are plotted in Fig. 2 as a function of σ. There are no solutions for $\sigma < \sqrt{3}$; the square-root term becomes imaginary for $|\sigma| < \sqrt{3}$, and negative solutions are excluded for positive N [Eq. (13)]. For $\sigma > \sqrt{3}$ (i.e., by moving $\sqrt{3}$ or more linewidths below the line), bistability appears, and the width of the bistable region (the difference between the two switching powers) becomes progressively larger.

Figure 3 shows calculated input–output characteristics for this shifting Lorentzian system. The trend from no differential gain for $\sigma \ll \sqrt{3}$ through differential gain and eventually into broad bistability as σ increases is clearly seen. These

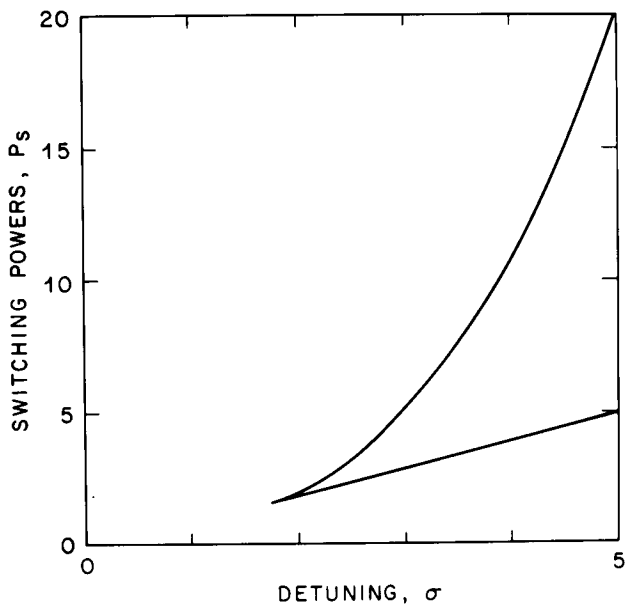

Fig. 2. Switching powers for a shifting Lorentzian line as discussed in the text. Dimensionless units are used ($\rho = \beta = \eta = 1$).

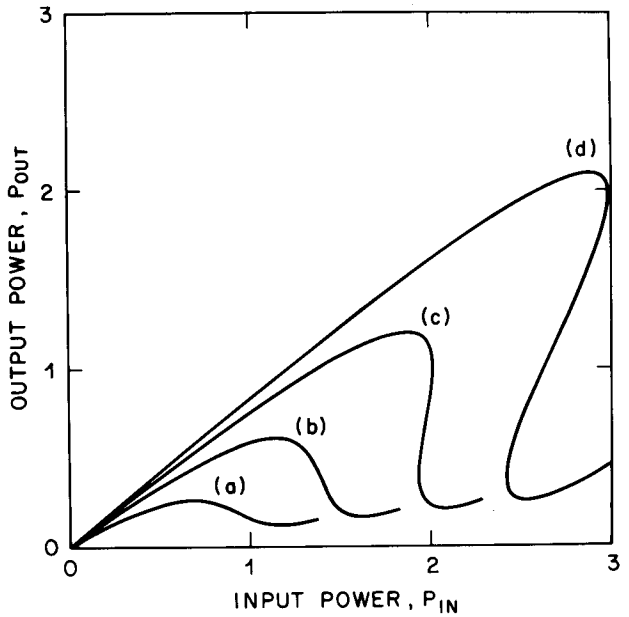

Fig. 3. Calculated input–output characteristics for the shifting Lorentzian example showing (a) no differential gain or bistability ($\sigma = 1.1$), (b) differential gain near $P_{in} = 1.4$ ($\sigma = 1.5$), (c) bistability ($\sigma = 1.9$), and (d) bistability with a broad bistable region ($\sigma = 2.3$). Dimensionless units are used ($\rho = \eta = 1$). β is chosen as 0.9.

curves are calculated by evaluating A for a range of N and for each N deducing P_{in} from Eq. (2). P_{out} is then deduced from A and P_{in}. This simple technique can be used for any function $A(N)$.

I have performed simulations for a variety of functions $A(N)$ (e.g., shifting Gaussian functions, linear functions, quadratic functions, step functions). All those that display bistability or strong differential gain show qualitatively similar sawtoothlike characteristics, such as those shown in Figs. 3(b)–3(d); this characteristic seems to result from any absorption that monotonically increases with N and that is sufficiently nonlinear to display differential gain and/or bistability.

4. LIMITING RELATIONS

In the above, I have given general analytic solutions and a specific analytically solvable example. The aim of this section is to derive further general relations that give physical insight into such characteristics as the switching power and the width of the bistable region. Ideally, such relations should give enough information to predict the main features of the operation of a given system without requiring detailed knowledge of, for example, $A(N)$; in what follows, relations in two limiting cases are derived for switching powers that go some way toward this ideal.

First, I derive an exact limit on switching or operating power. Suppose that we have a system that we know displays bistability. We restrict ourselves to an $A(N)$ that increases monotonically with d^2A/dN^2 positive (i.e., the absorption bends up) up to the first switching point (i.e., the switch to low transmission). The condition for the critical power for switching is [relation (9)] $dA/dN = A/N$. Since d^2A/dN^2 is positive, $dA/dN \geq dA/dN|_{N=0}$ up to the first switching point. Therefore we obtain

$$\frac{A(N_1)}{N_1} \geq \left.\frac{dA}{dN}\right|_{N=0}, \tag{16}$$

where N_1 is the excitation when the first switching point is reached. From Eq. (2), therefore, the switching power P_1 at this first switching point can be written as

$$P_1 \leq \frac{1}{\eta \, dA/dN|_{N=0}}. \tag{17}$$

This condition can be used by itself. However, by using the identity

$$\frac{dA}{dN} = \frac{dA}{dP_{in}} \Big/ \frac{dN}{dP_{in}}$$

[substituting Eq. (2) for dN/dP_{in}] we obtain

$$P_1 \leq \frac{A(0)}{dA/dP_{in}|_{P_{in}=0}} \tag{18a}$$

or

$$P_1 \leq \frac{1}{d \log A/dP_{in}|_{P_{in}=0}}, \tag{18b}$$

where we have used the fact that the conditions $N = 0$ and $P_{in} = 0$ are equivalent for $A \neq 0$ [from Eq. (2)]. Relations (18) therefore give a simple upper estimate of switching power (if the system is going to switch) in terms of measurable parameters at low power; the only requirement is that the absorption bend upward with increasing N up to this switching point.

In some situations, relations (18) greatly overestimate the switching power. This is especially true in situations in which A increases rather abruptly with N. An extreme case is the step function (where $dA/dP_{in}|_{P_{in}=0} = 0$)[17]:

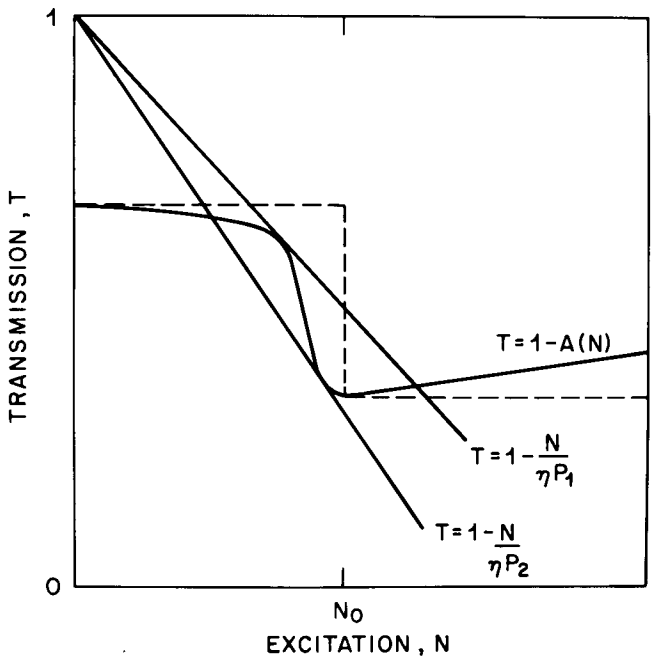

Fig. 4. Realistic abrupt function with associated critical power (straight lines). For comparison, the step function (dashed line) with the same peak absorption and N_0 is shown. The powers P_1 and P_2 are lower for the realistic function than those for the step function because the corresponding lines have steeper slopes than those for the step function (which must pass through the corners).

$$A = A_1, N < N_0,$$
$$A = A_2, N > N_0 \ (A_2 > A_1). \tag{19}$$

This function is shown as the dashed line in Fig. 4.

The solutions for the switching powers here are trivial:

$$P_1 = N_0/\eta A_1, P_2 = N_0/\eta A_2. \tag{20}$$

These powers correspond to straight lines $T = 1 - (NA_1/N_0)$ and $T = -NA_2/N_0$, which would pass through the vertices of the dashed line in Fig. 4. To interpret the usefulness of the step-function solutions, consider the more realistic function shown as the solid curve in Fig. 4. This function is still relatively abrupt and has an absorption peak at excitation N_0; the mathematical restrictions that we will put on this function for the following argument are the following: (1) The function is monotonically increasing in absorption (decreasing in transmission) up to N_0 and (2) the function has an absorption maximum at N_0 (with absorption A_p). It is clear from Fig. 4 that P_1 and P_2 are both lower than the corresponding powers for the abrupt nonlinearity (this is deduced directly from the relative slopes of the straight lines). This is a general result that can readily be rigorously proved given the above restrictions on the function. Therefore we have two new limits. For a function $A(N)$ that shows bistability and increases monotonically up to a peak A_p at $N = N_0$, the switching powers obey the relations

$$P_1 \leq N_0/\eta A(0), \tag{21}$$

$$P_2 \leq \frac{N_0}{\eta A_p}. \tag{22}$$

Result (21) is clearly an upper bound that may be well above the actual P_1. However, in practice, relation (22) gives a close estimate of switching power from low to high transmission (P_2); it is asymptotically correct for large σ in the Lorentzian case analyzed above and agrees within 10% over the whole range of $\sigma > \sqrt{3}$. These results [relations (21) and (22)] will also be approximately correct even if the function $A(N)$ does not have a peak but merely has a sharp knee at N_0 with absorption A_p.

The step function $A(N)$ also gives insight into the width of the bistable region. Clearly, for the step function, the width W expressed as a fraction of P_1 is

$$W \equiv \frac{P_1 - P_2}{P_1} = 1 - \frac{A_1}{A_2}. \tag{23}$$

We can extend the argument one step further to show that this W is an upper limit for the realistic function in Fig. 4. It is clear from Fig. 4 that the excitation that corresponds to the actual P_2 for the realistic function is not N_0 but a slightly lower value, which we now call N_0'. The actual absorption at which the switching takes place is $A' < A_p$. Therefore $P_2 \equiv N_0'/\eta A' > N_0'/\eta A_p$. Because the function $A(N)$ is single valued (i.e., it cannot double back on itself), the excitation corresponding to the actual P_1 for the realistic function must be less than N_0' so that $P_1 \leq N_0'/\eta A(0)$. Therefore the width W of the bistable region for this realistic function becomes after some manipulation

$$W < 1 - \frac{A(0)}{A_p}. \tag{24}$$

This general form of relation is borne out by simulations, with relative bistable width increasing with decreasing $A(0)/A_p$.

The accuracy is improved if $A(0)$ is replaced by the actual absorption A_1 just before the switch to low transmission (if this number is known). For the Lorentzian profile, a relation $W = 1 - \epsilon A_1/A_p$ is accurate for $\epsilon = 1$ for $\sigma = \sqrt{3}$ to $\epsilon = 3$ as $\sigma \rightarrow \infty$.

5. COMPARISON WITH OTHER DISCUSSIONS

In this section, I attempt to point out the relation between the general principle in this paper and the previous specific and largely independent discussions of what I believe to be the same general class of bistability.

Kaplan[3] proposed theoretically that bistability could arise in a relativistic electron gas under appropriate illumination. The mechanism appears to be as follows: The photon energy of excitation is chosen just below the cyclotron-resonance frequency of the unexcited gas, where there is still significant optical absorption resulting from broadening of the cyclotron resonance. As the incident power is increased, the power absorbed in the electron gas results in an increase in the kinetic energy (or, equivalently, in the temperature) of the gas. This increased electron energy results in a higher average mass for the electrons (or a higher average effective mass in the analogous quasi-relativistic nonparabolic semiconductor band case) and hence a lower cyclotron-resonance frequency. This movement of the cyclotron resonance toward the exciting photon energy gives increased absorption. Thus the increasing-absorption regenerative loop can be established, and bistability can result. The degree of excitation N could be chosen as (1) the total energy of the electron gas, (2) the average energy of an electron, or (3) the temperature of the electron gas, and a simple comparison can be seen with Eq. (2), where N is proportional to absorbed power. It is also possible to define N as, for example, the average electron velocity (magnitude) or the average electron effective mass; N would then no longer be linearly related to absorbed power (given curves instead of straight lines in the graphical solutions), but the functional form of $A(N)$ would change accordingly, giving identical final solutions.

In the case of Hopf *et al.*,[5] the specific system considered is a dense gas of two-level systems. Inclusion of the local field gives a Stark shift of the optical transition to lower photon energy. However, as the gas is optically excited, this local field is progressively reduced as more atoms are raised to the excited state. Thus, if the photon energy is chosen on the high side of the absorption line at lower excitation, then with increasing excitation the Stark shift is progressively reduced, and the absorption line shifts up toward the photon energy, giving increased absorption and hence the possible regenerative mechanism. The most obvious choice for the degree of excitation N is the population inversion[5] or the number of excited systems, which is likely to be linearly proportional to the absorbed power.

The cases of thermal bistability resulting from the shrinkage of optical band gap with increasing temperature[2,4,8,9] are particularly simple to relate to this general mechanism. In all cases, the photon energy is chosen below the band-gap energy so that the increased temperature resulting from absorbed power shrinks the band gap, giving increased absorption and hence establishing the regenerative mechanism; the obvious choice for N is temperature rise, which is again likely to be approximately proportional to absorbed power.

In the SEED,[13] a quantum-well electroabsorptive modulator is connected in series with a large resistor to a reverse-bias supply. The photon energy is chosen so that decreasing voltage on the modulator gives increasing absorption. Furthermore, over a large range of voltages, the modulator also works as a photodetector of constant internal quantum efficiency, so that a photocurrent that is directly proportional to absorbed power is generated. Thus increased incident optical power gives more photocurrent, which gives more voltage drop across the series resistor, less voltage across the quantum well, and hence greater absorption. Thus the increasing absorption regenerative loop is established. The obvious choice of N is photocurrent, although voltage change across the quantum well would do equally well. Because the quantum efficiency falls off at low voltage, the strict proportional relationship between photocurrent and optical absorption breaks down, and the simple relation of Eq. (2) is then inaccurate. However, Eqs. (1) and (2) become exact descriptions of the SEED if absorption A is replaced by responsivity S (current/incident optical power), although transmission T is no longer simply related to S.

The case of CdS considered by Bohnert *et al.*[6] and by Schmidt *et al.*[11] is slightly more complex in that it is believed that, at low intensities for the photon energies used below the band-gap energy, there will initially be negligible linear absorption. However, the intensities used are sufficiently high that two-photon absorption will be appreciable, and some photoexcited carriers will be created. Once this happens, the band-gap energy will start to renormalize to lower energies because of the change in self-energy of carriers in the presence of a plasma. Consequently, at sufficient incident intensity, the linear absorption becomes significant because of this renormalization, giving yet more excited carriers, more renormalization, and more absorption, thus establishing the basic increasing-absorption regenerative mechanism. N is basically the carrier density with an approximately linear relationship between N and absorbed power expected at high powers if a constant carrier lifetime is appropriate.[11] However, in the steady state, in general, the relations of Eqs. (1) and (2) would have to be further sophisticated to allow for N as a function of both linear absorption A and the two-photon absorption.

In the case of the research of Rossmann *et al.*[7] and of Henneberger and Rossmann,[10] the suggested microscopic mechanism for increase of absorption in CdS is different from that given in Refs. 6 and 9, with broadening of the exciton line being proposed. However, the basic mechanism is phenomenologically similar; increased carrier density (presumably resulting in this case from only single-photon absorption) results in an increasing absorption in an appropriate spectral region below the band-edge energy and hence the required regenerative mechanism. The theory of bistability[10] uses a similar graphical method to that used here and in Refs. 2 and 4, and the steady-state models are all essentially equivalent.[18] Henneberger and Rossmann[10] also consider some of the consequences of including or neglecting diffusion of the excitation. The simple model in this paper implicitly assumes total diffusion of the excitation so that the entire material can be characterized by one parameter N (e.g., temperature resulting from full thermal diffusion or carrier concentration resulting from carrier diffusion). If diffusion is totally neglected, the switching action disappears,[10] although hysteresis

remains; however, including even limited diffusion restores the switching transitions. This situation is reminiscent of the existence of whole-beam switching in Fabry–Perot bistability with Gaussian input beams; simple plane-wave theory neglecting diffraction and diffusion predicts hysteresis but no switching, but, in practice, switching is seen.

Goldstone and Garmire[12] do not consider explicitly the case of bistability resulting from increasing absorption. However, they do consider an interesting general formalism in which the optical field E is a single-valued function of the state of the medium as expressed by the polarization P rather than the more normal expression, P as a function of E. In this formalism, it is possible to obtain bistability without apparent feedback.[19] It is easily seen that the mechanism of bistability resulting from increasing absorption is a special case of this formalism. The internal optical field (e.g., as represented by the output power) is a single-valued function of the state of the medium (e.g., as represented by the absorption A or by the degree of excitation N). However, the converse is not true; the state of the medium is many valued (bistable) as a function of the internal optical field. This point is equivalent to the point made in Section 2 of this paper: that OB resulting from increasing absorption cannot result from a polarization expandable as a power series in the optical field (with constant coefficients).

6. CONCLUSIONS

The conclusions that can be drawn from the above exposition can be separated into three parts, corresponding to the three stated purposes of this paper.

First, the discussion in Section 5 demonstrates that there is a common underlying mechanism of OB resulting from increasing absorption in a large number of discussions of mirrorless bistability.[2–13] These discussions represent at least six independent descriptions of the same principle. Many of the theoretical models used[2–5,8,9,13] appear to be exactly reducible to the simple case described in general terms in Section 2 of this paper. Others require transient effects,[6,7,10,11] limited diffusion,[10] or more-elaborate absorption mechanisms,[11] although the basic principle is similar.

Second, various universal aspects are apparent. Most fundamental is the restriction that the nonlinearity that gives rise to this bistability cannot be expanded as a power series in the optical field (with constant coefficients) so that (constant) nonlinear susceptibilities are not an adequate description. In all cases of absorptive bistability discussed in this paper, the change in optical properties is a function of the real state of excitation of the material rather than of the optical field. In the simplest steady-state case in which all absorption of optical power contributes equally to the degree of excitation, the requirement on the material for bistability to exist is simply that the absorption be more than linearly proportional to the degree of excitation. A general condition for differential gain can also be derived. As to the general behavior of the bistability, the switching with increasing power is always to a state of higher absorption with an increased internal energy of the medium. Except when diffusion is totally neglected,[10] the systems show hysteresis and switching. In all the numerical simulations that I have performed and in all discussions in which the dependence of the bistability on other parameters of the system is considered in detail, there is a gradual evolution of the input–output characteristic through differential gain into bistability with increasing width, as an appropriate parameter (e.g., detuning) is altered; this is similar to other types of bistability and to first-order phase transitions in general. Qualitatively, it seems to be easier for one to obtain this bistability with a shift of an absorption feature[2–6,8,9,11,13] rather than with broadening (only Refs. 7 and 10 suggest broadening). This is in agreement with my numerical simulations and is easy to understand from the graphical solution in the simple case analyzed in this paper, in which absorption must increase more than linearly with the degree of excitation for bistability; if the absorption shifts linearly with the degree of excitation, there will always be some initial detuning for which the bistability condition is satisfied regardless of the shape of the absorption feature, whereas for broadening, the existence of bistability is critically dependent on the precise functional form of the broadening.

Third, for the simple case analyzed in this paper, it is possible to derive limiting relations for switching powers with only minor restrictions on the functional form of $A(N)$. In general, larger absorptions lead to lower switching powers. One of the more important general conclusions of the discussion in Section 4 is that the width of the bistable region tends to be controlled by the contrast ratio between the high-transmission and the low-transmission states, with higher contrast permitting a broader bistable region.

In conclusion, the great variety of physical systems that show this bistability suggests that there may be many other diverse systems capable of showing it, and the simplicity of the mechanism together with the absence of mirrors or other external feedback makes it attractive for practical applications. It also does not necessarily require coherent light.[2,13]

ACKNOWLEDGMENTS

I am grateful to C. M. Bowden, M. Dagenais, E. Garmire, C. Klingshirn, and S. W. Koch for many stimulating discussions.

REFERENCES

1. D. A. B. Miller, "Bistable optical devices: physics and operating characteristics," Laser Focus 18 (4), 79–84 (1982); "Optical bistability," in *Proceedings of the Fifth General Conference of the European Physical Society*, I. A. Dorobantu, ed. (European Physical Society, Geneva, 1982), pp. 239–249.
2. D. A. B. Miller, A. C. Gossard, and W. Wiegmann, "Optical bistability due to increasing absorption," Opt. Lett. **9**, 162–164 (1984).
3. A. E. Kaplan, "Hysteresis in cyclotron resonance based on weak relativistic-mass effects of the electron," Phys. Rev. Lett. **48**, 138–141 (1982).
4. J. Hajto and I. Janossy, "Optical bistability observed in amorphous semiconductor films," Philos. Mag. **B47**, 347–366 (1983).
5. F. A. Hopf, C. M. Bowden, and W. Louisell, "Mirrorless optical bistability with the use of the local field correction," Phys. Rev. A (to be published).
6. K. Bohnert, H. Kalt, and C. Klingshirn, "Intrinsic absorptive optical bistability in CdS," Appl. Phys. Lett. **43**, 1088–1090 (1983).
7. H. Rossmann, F. Henneberger, and J. Voigt, "Memory effect in the excitonic transmission of CdS," Phys. Status Solidi B **115**, K63–K67 (1983).
8. M. Dagenais, *Proceedings of Royal Society Discussion Meeting on Optical Bistability, Dynamical Nonlinearity and Photonic*

Logic, London, 1984, Philos. Trans. R. Soc. London Ser. A (to be published); M. Dagenais, W. F. Sharfin, and H. G. Winful, "Low-power optical bistability in cadmium sulfide platelets," J. Opt. Soc. Am. B **1**, 476–477 (1984).

9. F. A. P. Tooley and C. T. Seaton, Department of Physics, Heriot-Watt University, Edinburgh, U.K. (personal communication).
10. F. Henneberger and H. Rossmann, "Resonatorless optical bistability based on increasing nonlinear absorption," Phys. Status Solidi B **121**, 685–693 (1984).
11. H. E. Schmidt, H. Haug and S. W. Koch, "Theoretical explanation of the absorptive optical bistability in semiconductors due to band-gap shrinkage," Appl. Phys. Lett. **44**, 787–789 (1984).
12. J. A. Goldstone and E. Garmire, "Macroscopic manifestations of microscopic optical bistability," J. Opt. Soc. Am. B **1**, 466–467 (1984); E. Garmire, C. D. Poole, and J. A. Goldstone, *Proceedings of Royal Society Discussion Meeting on Optical Bistability, Dynamical Nonlinearity and Photonic Logic, London, 1984*, Philos. Trans. Roy. Soc. London Ser. A (to be published).
13. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, "A novel hybrid optically bistable switch: the quantum well self electro-optic effect device," Appl. Phys. Lett. **45**, 13–15 (1984).
14. Sometimes it is convenient to treat susceptibilities that depend on N as nonlinear susceptibilities that depend on the internal optical electric field E, since N depends on E. For example, a $\chi^{(1)}$ (linear susceptibility) whose real part (refractive index) depends linearly on N can be rigorously expressed as a power series in E in which the nonlinear term is $\chi^{(3)}$ when N is proportional to absorbed power (provided that the change in refractive index does not affect the absorption). The dynamics of N are then reflected in the frequency dependence of $\chi^{(3)}$. However, it is important for the present discussion to recognize the distinction between a nonlinearity that can be expressed only in terms of a power series in E and one that can be so expressed in some situations but can also always be expressed as a function of N. In the present situation, the expression in terms of N is correct, whereas because of the feedback the expression in terms of E alone is demonstrably incorrect. The response of this system is not even a single-valued function of E in the bistable region and hence cannot even be expressed as a power series in E. (By contrast, in a Fabry–Perot bistable system with a Kerr nonlinearity the response of the nonlinear medium remains a constant, single-valued function of the internal electric field at all times.)
15. For a discussion of dynamic nonlinear optics, see, for example, A. Miller, D. A. B. Miller, and S. D. Smith, "Dynamic nonlinear optical processes in semiconductors," Adv. Phys. **30**, 697–800 (1981); D. A. B. Miller, "Dynamic nonlinear optics in semiconductors," Laser Focus **19** (7), 61–68 (1983).
16. Care must be taken with the power-series expansions to prove this, requiring at least a second-order expansion.
17. A similar step function has also been considered by F. Henneberger and H. Rossmann.[10]
18. The condition derived for the existence of bistability in the simplest case (i.e., N proportional to AP_{in}) in this paper [expressions (9) and (10)] and in Ref. 2 is actually equivalent to that given in Ref. 10, Eq. (4) under the same conditions except that, in our case, we use total absorption A, and in Ref. 10 the absorption coefficient α is used. This appears to contradict the conclusion in this paper [see the discussion in Section 2.A concerning Eq. (8)] that the condition for absorption coefficient α is different from that for total absorption A. However, the difference is that the present treatment assumes implicitly that there is strong diffusion of excitation (e.g., A is a function of one well-defined temperature or carrier concentration of the whole material), whereas Eq. (4) of Ref. 10 assumes totally local response (i.e., no diffusion).
19. There is no physical contradiction between the contention of Goldstone and Garmire[12] that their class of bistability requires no feedback and my identification of a feedback mechanism for OB resulting from increasing absorption. What is true in both cases is that no external feedback is required.