OPTICAL PULSE PROPAGATION IN MEDIA EXHIBITING A THIRD ORDER NONLINEAR POLARIZATION

Thesis

by

Jack Clifton Comly, Jr.

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1974

(Submitted May 20, 1974)

This work, as with all my endeavors, is dedicated to

Betsy

ACKNOWLEDGMENTS

The author wishes to express his deep appreciation to his advisor, Professor Amnon Yariv, not only for his technical contributions to this work but also for the encouragement and support he gave throughout the research.

I wish to thank Mr. Desmond Armstrong, Dr. Elsa Garmire, and Dr. Jean-Pierre Laussade for their assistance during the experimental portion of my research. I am also indebted to Dr. Jerry Suydam of the Los Alamos Scientific Laboratory for bringing to my attention the beam moment solutions discussed in Chapter III.

The author is very grateful for the careful typing of this work by Dian Rapchak and Kathy Ellison, and for the financial assistance provided by the National Science Foundation, the Francis J. Cole Special Fellowship, and by the California Institute of Technology.

ABSTRACT

A theoretical investigation is made of two types of distortions, self phase modulation and self-focusing, produced when an intense optical pulse propagates in a medium whose induced polarization contains terms cubic in the field strength. The first analysis indicates that the insertion of such a medium into a laser cavity can result in the generation of a train of very stable, reproducible, bandwidth-limited pulses. By using a "circulating pulse" model, the physical processes involved are clarified, particularly in the case of high gain in the amplifying medium. Examples are given of the Q-switched operation of such a laser, for which the pulse train differs considerably from that of conventional mode-locked systems. A second analysis deals with the steady-state self-focusing of non-axisymmetrical beams; several approaches are used to derive the increase in the threshold power for elliptical-Gaussian beam shapes. As an alternative to a fully numerical solution, a series of increasingly accurate approximate results are obtained in the form of parameterized beam functions. When an Action-Integral minimization technique is employed to optimize these parameters, the method is capable of describing the self-focusing process in some detail.

TABLE OF CONTENTS

INTRODUCTION

The development of laser light sources capable of generating very intense, coherent, narrow bandwidth light beams has allowed the practical observation of many nonlinear optical effects. One class of effects of great interest involves some exchange of optical photon energy with other "distinct" fields, usually by coupling different optical beams; this category would include such effects as Rayleigh[1], Brillouin[2], and Raman[3] scattering, parametric interactions[4], and harmonic generation[5,6]. A typical analysis of such an effect would usually consider the interaction on two or three "scales"; first, a description of a coupling mechanism would be made on a "microscopic" (i.e. molecular) scale, possibly using quantum mechanics, and second, coupling terms (usually involving classical functions) would be derived on a "macroscopic" scale of a few optical wavelengths. An extension of the analysis to include interactions over a "propagation" length scale ($\sim 10^3$ wavelengths or longer) might show significant modifications to the "energy conversion" formulas, but usually would not alter the qualitative explanation of the effect.

A second class of nonlinear optical effects can be considered, in which only a "single" beam interacts with the material; the resultant beam behavior can often be viewed as a nonlinear propagation problem. This category would include many of the "resonant" effects such as saturable absorption[7], self-induced transparency[8], superradiance[9], and the general propagation of intense pulses in amplifying media[10]. Some aspects of this last example will be discussed in this report,

to the extent that it applies to the amplification of ultrashort pulses in laser material. However, the primary subject of the analyses presented here concerns the nonlinear propagation characteristics resulting from "non-resonant" effects such as the Kerr effect[11], and other mechanisms producing a macroscopic third order susceptibility tensor[12].

In the first chapter, the major origins of such effects are described on the "microscopic" and "macroscopic" scales, where their consequences do not appear very dramatic. However, when the nonlinearities act over a sufficiently long "propagation" length, cumulative effects occur which are not apparent in the smaller length scales; it is seen that the most appropriate physical model of the nonlinearity is that it causes an intensity-dependent index of refraction. These propagation effects play a major role in the subsequent chapters, where beam distortions are exemplified by analyzing two distinct physical situations illustrating the phenomena of pulse self-modulation[13], and self-focusing[14], respectively.

The first analysis considers a laser cavity containing this nonlinear medium in addition to the usual amplifying element. The corresponding system employing a saturable absorber has been investigated experimentally[15,16] and theoretically[17,18]; it is well known that this system can lead to "mode-locking"[19], yielding ultrashort optical pulses with very useful characteristics ($\sim 10^{-12}$ seconds duration, $\sim 10^{9}$ Watts peak power in Ruby or Nd:glass lasers.) A previous theory[20] has shown that the Kerr effect can also be used to mode-lock a laser, due to the mode-coupling character of its cubic

polarization; the present propagating-pulse approach extends those results to more realistic conditions, including the high gain case. This analysis predicts the form of the "ultrashort" pulses, and shows the strong stability of the pulsing system; the physical roles played by the amplifier and by the nonlinear material in producing these pulses is clarified. In particular, the intensity-dependent index of refraction is shown qualitatively to produce rapid phase changes within the pulse envelope, and a plane wave analysis is presented to derive quantitative examples.

The second major analysis presented here deals with the steady-state propagation of a beam with a finite cross-section through a medium exhibiting a Kerr type nonlinearity; the beam can produce its own lens and self-focus when its power exceeds a certain threshold value. Experiments[21] have shown that this threshold is increased if the beam cross section is not symmetrical, and a number of semiquantitative arguments are presented to explain this. In general, however, the solution of the propagation equations in this situation cannot be found analytically, and the nonlinearities are too severe to permit permutation expansions. To complement a fully numerical solution, therefore, a Lagrangian formulation of the problem is presented and used to obtain a series of approximate solutions in the form of parameterized beam shapes. This approach is shown to be highly successful in predicting the threshold power, and gives a good approximation to other details of the solution as well.

Chapter I

ORIGINS OF A CUBIC NONLINEARITY

The microscopic description of the interaction of electromagnetic waves with the molecules of a material is clearly a complex many-body problem with many nonlinearities. Even in an idealized case, when one type of nonlinear interaction can be treated separately, a quantitatively accurate analysis is seldom possible for a specific material. On the other hand, the phenomena of interest here involve optical beam propagation over large macroscopic distances, with the material behavior represented primarily through nonlinear "polarization source" terms; therefore, microscopic details are not critical, provided the general form of the macroscopic nonlinearities are preserved.

The lowest order "non-resonant" effects which can contribute to the "self-distortion" of a pulse involve polarizations cubic in the field strengths; this chapter describes some common causes of such nonlinearities in materials which are homogeneous, isotropic, and (usually) lossless. The macroscopic equations for beam propagation are derived, and shown to be consistent in many cases with the concept of an intensity-dependent refractive index.

I.1 Fundamental Equations

The physical situation underlying these investigations involves the propagation of intense, coherent, light beams, produced by lasers which are usually Q-switched and often mode-locked as well. Even though this latter case may involve repetitive pulses, it will be

seen in Chapter II that the pertinent analysis involves the propagation of only a single pulse, which therefore has a continuous, though narrow, bandwidth centered at the optical frequency, ω_0. Under these conditions, the light propagation will be described most conveniently by using classical fields obeying Maxwell's equations, rather than a photon scattering or mode coupling approach[22]. Thus, in e.s.u. units, the effects of interest will be described by the equations[23]:

$$\nabla\cdot\vec{\mathscr{D}} = 4\pi\rho \qquad \nabla\cdot\vec{\mathscr{B}} = 0$$

$$\nabla \times \vec{\mathscr{E}} = -\frac{\partial\vec{\mathscr{B}}}{\partial t} \qquad \nabla \times \vec{\mathscr{H}} = 4\pi\vec{\mathscr{J}} + \frac{\partial\vec{\mathscr{D}}}{\partial t} \tag{I.1.1}$$

$$\vec{\mathscr{D}} = \vec{\mathscr{E}} + 4\pi\vec{\mathscr{P}} \qquad \vec{\mathscr{H}} = c^2\vec{\mathscr{B}} - 4\pi\vec{\mathscr{M}} \quad .$$

We will consider materials in which there are no macroscopic charge densities or currents ($\rho = \vec{\mathscr{J}} = 0$), and in which magnetic effects are negligible ($\vec{\mathscr{M}} = 0$); all linear and nonlinear material interactions will then be represented by the net macroscopic polarization "source" term, $\vec{\mathscr{P}}$.

The analyses in later chapters will deal with laser beams (or pulses) traveling through the material in the +z direction; the field distortions produced by the nonlinearities will also be shown to be very small over a distance of a single wavelength. The standard form for the field will then be in the form of a traveling wave with a slowly varying amplitude and phase; for example, we will write the component of the electric field polarized in the x direction as follows:

$$(\vec{\mathcal{E}})_x = \frac{1}{2}(E_x e^{i(k_o z-\omega_o t)} + E_x^* e^{-i(k_o z-\omega_o t)})\vec{e}_x \ , \qquad \text{(I.1.2)}$$

where $E_x = E_x(x,y,z,t)$ is complex and varies slowly in z and t compared with the exponential, and $k_o = \omega_o n_o(\omega_o)/c$ with $n_o(\omega_o)$ the (linear) refractive index at ω_o. Similar expressions will hold for $(\vec{\mathcal{E}})_y$, $(\vec{\mathcal{E}})_z$ and the components of $\vec{\mathcal{P}}$. Note that a number of assumptions must be implicitly made before the "standard" form in equation I.1.2 can be used: the nonlinearities must be small over a wavelength, third and higher harmonics must be negligible, and no large-angle scattering components (including backward scattered) need to be considered. We will later show that these approximations are consistent with the nonlinearities involved in our investigations.

The relationships giving $\vec{\mathcal{P}}(\vec{r},t)$ complete the set of propagation equations; a description of the microscopic interactions within the material is needed to find this relationship. Some of the most important effects causing a "cubic" polarization response will be analyzed in Section I.2, based on a molecular model described below. Since practical examples of these effects occur in a wide variety of solids, liquids, and, in some cases, gases, our molecular model is too general to allow, for example, quantitative results in any specific case. However, the physical principles and qualitative results should still apply in more detailed analyses, and we will find good order-of-magnitude agreement with experimentally determined "nonlinearity coefficients."

The microscopic interaction is presumed to occur with a set of "active", discrete polarization sources, termed molecules; in real

materials, these might also be atoms, ion-pairs, or "clusters of molecules", and similar analyses could be used if phonon-waves, spin-waves, etc., were involved. For simplicity, we will assume that the entire material is composed of a single "species" of these active molecules, since, to first order, the polarizations of different constituents can simply be summed. Each molecule will be governed, either classically or quantum mechanically, by a Hamiltonian of the form

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}' + \mathcal{H}_{int} \equiv \mathcal{H}^m + \mathcal{H}_{int} \quad , \qquad (I.1.3)$$

where $\mathcal{H}_0$ is the Hamiltonian for an isolated molecule, $\mathcal{H}'$ includes the influences of the surrounding medium, and $\mathcal{H}_{int}$ is the interaction of the molecule with the fields resulting from the propagation beam. The coherent local polarization field, discussed below, will be included in $\mathcal{H}_{int}$, so that $\mathcal{H}'$ thus includes only the short range fields of neighboring molecules. We further assume that the molecule has a fixed "equilibrium" position, due to $\mathcal{H}'$, or at least that molecular migration does not influence the nonlinear effects discussed later.

The interaction of the charges, q_s, comprising the molecule with the electromagnetic fields is given by the Lorentz force, $\vec{\mathcal{F}}_s = q_s(\vec{\mathcal{E}}(\vec{r}_s) + \vec{v}_s \times \vec{\mathcal{B}}(\vec{r}_s))$; when magnetic effects are ignored ($v \ll c$), this implies

$$\mathcal{H}_{int} = -\vec{\mathcal{P}}^m \cdot \vec{\mathcal{E}}^m \qquad (I.1.4)$$

where

$$\vec{\mathcal{P}}^{\,m} = \sum_s q_s \vec{r}_s \quad .$$

The superscript "m" will be used in this chapter to denote microscopic quantities, applying to a single molecule. We make the "dipole approximation", since spatial variations of $\vec{\mathcal{E}}$ over molecular dimensions are negligible; $\vec{\mathcal{E}}^{\,m}$ can then be assigned the constant (spatially) value taken at the molecule's "site." $\vec{\mathcal{P}}^{\,m}$ is the molecular dipole moment; that part which varies at the optical frequency, ω_o, represents the induced moment, and will contribute to the macroscopic polarization when statistically summed over the molecules in a small "unit" volume:

$$\vec{\mathcal{P}} = \sum_m \vec{\mathcal{P}}^{m}(\omega_o) \quad . \tag{I.1.5}$$

The influence of the medium on the response of a single molecule is manifested in two ways, in reference to the Hamiltonian in equation I.1.3. First, in $\mathcal{H}_{int}$, the microscopic electric field, $\vec{\mathcal{E}}^{\,m}$, differs from the macroscopic field, $\vec{\mathcal{E}}$, because of the polarization of the surrounding molecules, taken as a continuum. We will account for this by using the Lorentz-Lorenz local field correction[24],

$$\vec{\mathcal{E}}^{\,m} = \vec{\mathcal{E}} + \frac{4\pi}{3}\vec{\mathcal{P}} \quad . \tag{I.1.6}$$

The second effect involves the dependence of $\mathcal{H}'$ on $\vec{\mathcal{E}}$; thus, even if $\mathcal{H}_{int}$ weren't significant, the "equilibrium" state of the molecule, as influenced in part by the short range forces, would change as neighboring molecules reacted to the field. This indirect interaction

will be ignored, as it would not qualitatively alter the phenomena discussed in the next section; $\mathcal{H}^m$, defined in equation I.1.3, will therefore be assumed to be unperturbed by the optical beam.

We will find that, for isotropic materials, the average induced molecular dipole moment is given by

$$\vec{\mathcal{P}}^m = \bar{\alpha}^m \vec{\mathcal{E}}^m + \vec{\mathcal{P}}^m_{NL} \equiv \vec{\mathcal{P}}^m_{LIN} + \vec{\mathcal{P}}^m_{NL} \quad , \tag{I.1.7}$$

where $\bar{\alpha}^m$ is the (average) molecular polarizability, and $\vec{\mathcal{P}}^m_{NL}$ is the "nonlinear" part of the polarization. Defining N, the number of molecules per unit volume, we can write $\vec{\mathcal{P}} = N\vec{\mathcal{P}}^m$; equations I.1.6 and I.1.7 then show that

$$\vec{\mathcal{E}}^m = (\vec{\mathcal{E}} + \tfrac{4\pi}{3}\vec{\mathcal{P}}_{NL})/(1 - \tfrac{4\pi}{3} N \bar{\alpha}^m) \ .$$

By defining the linear dielectric constant, $\varepsilon = n_o^2$, in the usual manner,

$$\vec{\mathcal{D}}_{LIN} = \vec{\mathcal{E}} + 4\pi \vec{\mathcal{P}}_{LIN} \equiv \varepsilon \vec{\mathcal{E}} \quad , \tag{I.1.8}$$

we obtain, for $|\vec{\mathcal{P}}^m_{NL}| << |\vec{\mathcal{P}}^m_{LIN}|$,

$$\begin{aligned} \varepsilon &= (1 + \tfrac{8\pi}{3} N \bar{\alpha}^m)/(1 - \tfrac{4\pi}{3} N \bar{\alpha}^m) \\ \vec{\mathcal{E}}^m &= (\tfrac{\varepsilon+2}{3}) \vec{\mathcal{E}}^m \ . \end{aligned} \tag{I.1.9}$$

Finally, this local field correction will affect the net macroscopic nonlinear polarization as follows:

$$\vec{\mathcal{D}} = \vec{\mathcal{E}} + 4\pi N\vec{\mathcal{P}}^m = \vec{\mathcal{E}} + 4\pi N(\bar{\alpha}^m\vec{\mathcal{E}}^m + \vec{\mathcal{P}}^m_{NL})$$

$$= \varepsilon\vec{\mathcal{E}} + 4\pi(\frac{\varepsilon+2}{3})N\vec{\mathcal{P}}^m_{NL} \qquad (I.1.10)$$

$$\equiv \varepsilon\vec{\mathcal{E}} + 4\pi\vec{\mathcal{P}}_{NL} ,$$

where $\vec{\mathcal{P}}_{NL} = (\frac{\varepsilon+2}{3})N\vec{\mathcal{P}}^m_{NL}$. When the nonlinear polarization is present, equation I.1.9 is not precisely correct, but the change will be second order when included in $\vec{\mathcal{P}}_{NL}$. Armstrong, et al[6] indicate how the above local field effects can be generalized to more complicated systems of different polarization sources.

I.2 Microscopic Interactions

The microscopic, molecule-field, interaction given in equation I.1.4 is the usual energy of a dipole, which can be written:

$$\mathcal{H}_{int} = -\mathcal{P}^m(t)\,\mathcal{E}^m(\vec{r}^m,t)\cos\theta_{PE}$$

where θ_{PE} is the angle between $\vec{\mathcal{P}}^m$ and $\vec{\mathcal{E}}^m$, and $\vec{r}^m$ is the location of the molecule. Each molecule will thus have a tendency to vary $\mathcal{P}^m$, θ_{PE}, and $\vec{r}^m$ to decrease its energy: a dipole moment, $\mathcal{P}^m$, will be induced which is in phase with $\mathcal{E}^m$, the dipole moment will tend to become aligned with the field $(\theta_{PE} \to 0)$, and the molecule will tend to move into a region of higher field strength. Figure 1 indicates how these three tendencies arise directly from the Lorentz force; the resultant forces are the physical mechanisms producing the nonlinear effects of, respectively, nonlinear (and linear)

electronic polarization, orientational Kerr effect, and electrostriction; each of these is briefly analyzed below.

In order to obtain the proper induced polarization component, we restrict the quantitative analysis in this section to the "steady state" limit, where E_x in equation I.1.2 is constant in time; we will thus use the shortened notation

$$\vec{\mathcal{E}}^m(\vec{r}^m,t) = \frac{1}{2}(\vec{E}^m(\vec{r}^m)e^{ik_o z}e^{-i\omega_o t} + \vec{E}^{m*}(\vec{r}^m)e^{-ik_o z}e^{i\omega_o t})$$
$$\equiv \frac{1}{2}(\vec{E}^m e^{-i\omega_o t} + \vec{E}^{m*} e^{i\omega_o t}) \quad . \tag{I.2.1}$$

The function "$\frac{1}{2}\vec{E}^m$" will therefore represent that part of the field implicitly varying as $e^{-i\omega_o t}$ and will presumably be constant over a volume containing a (statistically) large number of molecules. The purpose of this section is to calculate the average molecular polarization, $\vec{\mathcal{P}}^m$, as given in equation I.1.7, so that, for example, we can find the component, $\vec{\mathcal{P}}^m_{NL}$, where

$$\vec{\mathcal{P}}^m_{NL} = \frac{1}{2}(\vec{P}^m_{NL}e^{-i\omega_o t} + \vec{P}^{m*}_{NL}e^{i\omega_o t}) \quad . \tag{I.2.2}$$

We will find that, in isotropic media, this average microscopic polarization obeys the equation

$$\vec{P}^m = \bar{\alpha}^m\vec{E}^m + \bar{\chi}^m : \vec{E}^m\cdot\vec{E}^m\cdot\vec{E}^{m*} \tag{I.2.3}$$

where $\bar{\alpha}^m$ is a scalar and $\bar{\chi}^m$ is a tensor. For clarity, we will usually write actual tensor products using component notation; in

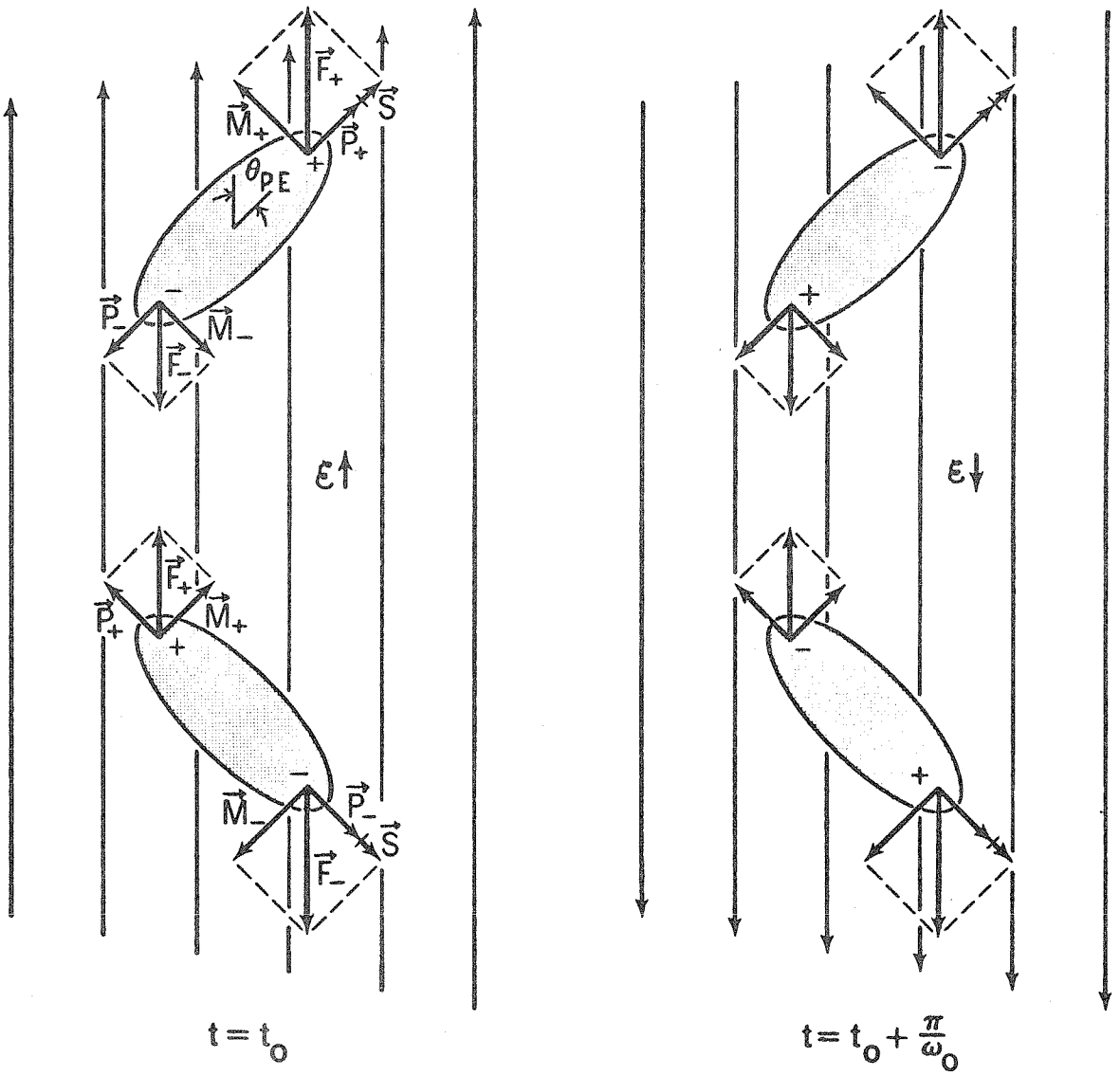


Figure 1. Polarizable, anisotropic molecules in a non-uniform field. At the left are two molecules in a field whose magnitude decreases to the left; an (induced) charge separation is shown. The forces on the charges are $\vec{F}_\pm$, which are resolved into components $\vec{P}_\pm$, $\vec{M}_\pm$, and $\vec{S}$. Here, $\vec{P}_+$ and $\vec{P}_- = -\vec{P}_+$ cause the linear and nonlinear electronic polarization; $\vec{M}_+$ are torques leading to the orientational Kerr effect; and $\vec{S}$ leads to the electrostrictive force, since the net $\vec{S}$ for the two molecules shown is in the direction of the field gradient. On the right are shown the same molecules when the field has reversed sign; the net forces are seen to be the same, leading to time-averaged torques and electrostrictive forces.

these cases, summation over repeated indices will be implied. According to equations I.1.9 and I.1.10, we can then define corresponding macroscopic polarizations analogously to equation I.2.2, and obtain $\vec{\mathcal{D}} = \varepsilon \vec{\mathcal{E}} + 4\pi \vec{\mathcal{P}}_{NL}$, where

$$\varepsilon = 1 + 4\pi \bar{\alpha}$$

$$\vec{P}_{NL} = \bar{\chi} \vdots \vec{E} \cdot \vec{E} \cdot \vec{E}^{*} \qquad \text{(I.2.4a)}$$

with

$$\bar{\alpha} = \left(\frac{\varepsilon+2}{3}\right) N \bar{\alpha}^{m} = N \bar{\alpha}^{m} / \left(1 - \frac{4\pi}{3} N \bar{\alpha}^{m}\right)$$

$$\bar{\chi} = \left(\frac{\varepsilon+2}{3}\right)^{4} N \bar{\chi}^{m} \quad . \qquad \text{(I.2.4b)}$$

For multiple species of active molecules, the formulas in I.2.4b would include sums over each type, with the appropriate N.

Finally we will consider qualitatively the response of the nonlinear polarization when $\vec{E}$ varies slowly with time, to determine a phenomenological "relaxation rate" for each effect.

I.2.i nonlinear electronic polarization

An isolated molecule in a constant electric field will have a polarization induced in it; physically, the shift in the electronic structure would be expected to be a complicated function of the electric field strength, linear only for sufficiently small changes. This "fundamental" source of nonlinearity will be developed here in some generality, using both quantum mechanical and classical approaches. The methods are similar to Armstrong, et al[6], except that we will

deal with the limiting case where every field has the same frequency.

Quantum mechanically, the molecule is described by the Hamiltonian given in equation I.1.3, with the perturbation given in equations I.1.4 and I.2.1; in the more familiar perturbation notation, these become

$$\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_1(t)$$

with

$$\mathcal{H}_1(t) = e^{-i\omega_0 t}\left(-\frac{1}{2}\vec{E}\vec{m}\cdot\vec{\mu}\right) + e^{i\omega_0 t}\left(-\frac{1}{2}\vec{E}\vec{m}^*\cdot\vec{\mu}\right) \qquad \text{(I.2.5)}$$

and

$$\vec{\mu} = \sum_s q_s \vec{r}_s \quad .$$

Here $\vec{\mu}$ is the dipole operator, s labels the molecular constituents, with charges q_s and position operators $\vec{r}_s$. The perturbation, $\mathcal{H}_1(t)$ is of course hermitian, since $\vec{r}_s$ and $\vec{\mu}$ are real operators. Between collisions, it will be assumed that $\mathcal{H}_0$ does not depend on time, so that the molecular state function, $\psi(\hat{r},t)$, can be expanded, in the usual manner, in terms of an orthogonal set of stationary states, u_n:

$$\psi(\hat{r},t) = \sum_n a_n(t)e^{-i\omega_n t}\, u_n(\hat{r}) \qquad \text{(I.2.6a)}$$

where

$$\mathcal{H}_0 u_n = \hbar\, \omega_n u_n \quad . \qquad \text{(I.2.6b)}$$

Here $\hat{r} = \{\vec{r}_1,\vec{r}_2,\ldots,\vec{r}_s,\ldots\}$ is the generalized spatial coordinate for all charges; we will usually consider the molecule to have only one "perturbable" charge (i.e. a valence electron of charge $q = -e$),

in which case $\hat{r}$ is simply the real space vector. We will also label one state as the ground state, u_g, with energy $\hbar\omega_g \equiv 0$, and assume that it is fully occupied at $t = 0$, when the perturbation is turned on. A more detailed analysis would allow for a thermal distribution of ground states, but this will not be done here.

Applying 3rd order, time-dependent, perturbation analysis[25] to the system described by equation I.2.5 yields formulas for $a_n(t)$ and thus $\Psi(\hat{r},t)$; the molecular dipole moment is then given by the formula,

$$\vec{\mathcal{P}}^m(t) = \langle \Psi | \vec{\mu} | \Psi \rangle \ . \qquad \text{(I.2.7)}$$

This polarization source contains all orders of nonlinearity up to cubic and many frequency components. It will be assumed that there is no resonance involved, so that $\omega_n - \omega_{n'} \neq \omega_o$ for any n,n'; this allows the "contributing" components of $\vec{\mathcal{P}}^m(t)$ (those varying at $\pm\omega_o$) to be found, in the form

$$(\vec{\mathcal{P}}^m)_{\omega_o} = \frac{1}{2}(\vec{P}^m e^{-i\omega_o t} + \vec{P}^{m*} e^{i\omega_o t}) \ . \qquad \text{(I.2.8)}$$

The linear polarization found in this way is given by

$$\vec{P}^m_{LIN} = \frac{1}{\hbar} \sum_n \left\{ \frac{(\vec{E}^m \cdot \vec{\mu})_{gn} (\vec{\mu})_{ng}}{(\omega_n + \omega_o)} + \frac{(\vec{\mu})_{gn} (\vec{E}^m \cdot \vec{\mu})_{ng}}{(\omega_n - \omega_o)} \right\} \qquad \text{(I.2.9)}$$

where, for example,

$$(\vec{E}^m \cdot \vec{\mu})_{nn'} \equiv \langle u_n | \vec{E}^m \cdot \vec{\mu} | u_{n'} \rangle = \int u_n^*(\hat{r}) (\vec{E}^m \cdot \vec{\mu}) u_{n'}(\hat{r}) d^3\hat{r} \ .$$

From this we can define the linear molecular polarizability tensor α^m, such that

$$(P^m_{LIN})_i = \alpha^m_{ij} E^m_j \qquad i,j = x,y,z \tag{I.2.10}$$

For the case of a single electron, $\vec{\mu} = -e\vec{r}$, and so equation I.2.9 gives

$$\alpha^m_{ij} = \frac{e^2}{\hbar} \sum_n \left\{ \frac{(r_j)_{gn}(r_i)_{ng}}{(\omega_n + \omega_o)} + \frac{(r_i)_{gn}(r_j)_{ng}}{(\omega_n - \omega_o)} \right\} \tag{I.2.11}$$

where $r_x \equiv x$, etc. We will call a molecule "lossless" if $\omega_o << \omega_n$ for all $n \neq g$ which contribute to polarization formulas. It can be seen that α^m is real (and symmetric) for a "lossless" medium, as expected. In later analyses, we will also find it useful to define a set of "principal molecular axes", (x',y',z'), for which α^m is diagonal; this can always be done[26]. In such a frame (in which we will use primed fields), we can write

$$(P^{m'}_{LIN})_i = \alpha^m_{ij} E^{m'}_j \quad , \tag{I.2.12}$$

with

$$\alpha^m_{ij} = \alpha^m_i \delta_{i,j} \qquad \text{(no sum over i)}$$

The quantities, α^m_i, are called the principal polarizabilities of the molecule.

The next terms in the perturbation expansion which contribute to $\vec{P}^m$ are cubic in the fields, as expected. Most of these involve the three factors $(\vec{E}^m, \vec{E}^m, \vec{E}^{m*})$; however, there also appear to be terms varying as $e^{-i\omega_o t}$ which contain factors $(\vec{E}^m, \vec{E}^m, \vec{E}^m)$ or

($\vec{E}^m, \vec{E}^{m*}, \vec{E}^{m*}$). These can be ignored for two reasons: first, their spatial variation is seen from equation I.2.1 to contain the respective factors, e^{3ik_oz} and e^{-ik_oz} (this latter is a backward scattered wave); their net contribution to $\vec{P}^m$, which varies as e^{ik_oz}, will therefore cancel in a wavelength. Second, the phase of these polarizations, relative to $\vec{E}^m$, depends on the time origin as $e^{\mp 2i\omega_o t}$. Physically, this time origin involves either the leading edge of the pulse or else the most recent relaxation or collision process; therefore, for pulses whose amplitude varies slowly in an optical cycle (the case here), or for times longer than the relaxation rate, these polarizations will average out. This justifies, on a microscopic scale, the neglect of backward scattered waves.

The resultant expression for the cubic polarization is further simplified by assuming that all molecular states, u_n, are non-degenerate; this allows all dipole matrix elements to be taken to be real. Defining a microscopic nonlinear susceptibility tensor, χ^m, such that

$$(\vec{P}^m_{NL})_i = \chi^m_{ijk\ell} E^m_j E^m_k E^{m*}_\ell \quad , \tag{I.2.13}$$

we obtain, for a "single electron" molecule,

$$\chi^m_{ijk\ell} = \frac{e^4}{4\hbar^3} \sum_{n,n',n''} \left\{ \frac{(r_i)_{gn}(r_j)_{nn'}(r_k)_{n'n''}(r_\ell)_{n''g}}{(\omega_{n'} \mp \omega_o)(\omega_{n''} \pm \omega_o)} \cdot \left(\frac{(1-\delta_{n',g})}{\omega_{n'}} - \frac{\delta_{n',g}}{(\omega_{n'} \mp \omega_o)} \right) \right.$$

$$\left. + \frac{(r_j)_{gn}(r_i)_{nn'}(r_\ell)_{n'n''}(r_k)_{n''g}}{(\omega_{n'} \pm \omega_o)(\omega_{n''} \mp \omega_o)} \cdot \left(\frac{(1-\delta_{n',g})}{\omega_{n'}} - \frac{2\delta_{n',g}}{(\omega_{n''} \mp \omega_o)} \right) \right. \tag{I.2.14}$$

$$+\frac{(r_i)_{gn}(r_j)_{nn'}(r_\ell)_{n'n''}(r_k)_{n''g}}{(\omega_n \mp \omega_o)(\omega_{n''} \mp \omega_o)}\cdot\left(\frac{(1-\delta_{n',g})}{\omega_{n'}} - \frac{\delta_{n',g}}{(\omega_n \mp \omega_o)} - \frac{2\delta_{n',g}}{(\omega_{n''} \mp \omega_o)}\right)$$

$$+\frac{(r_j)_{gn}(r_i)_{nn'}(r_k)_{n'n''}(r_\ell)_{n''g}}{(\omega_n \pm \omega_o)(\omega_{n''} \pm \omega_o)}\cdot\left(\frac{(1-\delta_{n',g})}{\omega_{n'}} - \frac{\delta_{n,n''}}{(\omega_n - \omega_{n'} \pm \omega_o)}\right)$$

$$+\frac{(r_i)_{gn}(r_\ell)_{nn'}(r_j)_{n'n''}(r_k)_{n''g}}{(\omega_n \mp \omega_o)(\omega_{n'} \mp 2\omega_o)(\omega_{n''} \mp \omega_o)} + \frac{(r_\ell)_{gn}(r_i)_{nn'}(r_j)_{n'n''}(r_k)_{n''g}}{(\omega_n \mp \omega_o)(\omega_{n''} \mp \omega_o)}\cdot$$

$$\left.\left(\frac{1}{(\omega_{n'} \mp 2\omega_o)} - \frac{\delta_{n,n''}}{(\omega_n - \omega_{n'} \pm \omega_o)}\right)\right\} \cdot \qquad \text{(I.2.14) cont'd.}$$

Here, for example, we have used the notation:

$$\frac{1}{(\omega_n \mp \omega_o)(\omega_{n''} \pm \omega_o)(\omega_n \mp \omega_o)} \equiv \frac{1}{(\omega_n - \omega_o)(\omega_{n''} + \omega_o)(\omega_n - \omega_o)} + \frac{1}{(\omega_n + \omega_o)(\omega_{n''} - \omega_o)(\omega_n + \omega_o)} .$$

Note that "E_j^m", and "E_k^m" are not uniquely defined by equation I.2.13 when $j \neq k$; only the combination, $\chi^m_{ijk\ell} + \chi^m_{ikj\ell}$, has physical significance in this case. Since all matrix elements are real, $\chi^m_{ijk\ell}$ is real; this would hold, of course, even in the presence of degenerate levels for "lossless" materials. In the latter case, it can be seen from equation I.2.14 that

$$\chi^m_{iijj} + \chi^m_{ijij} = 2\chi^m_{ijji} \cdot \qquad \text{(I.2.15)}$$

We will now develop the corresponding classical formulas, based on a "particle in a potential well" model. Such a model would apply to the behavior of an ion in the potential of the surrounding

medium if carried out quantum mechanically, but the classical derivation retains all of the features necessary here, and even yields reasonable results in the case of electronic polarization in a molecule. We will consider the problem in the principal axes reference frame discussed earlier, and assume a slightly anharmonic potential well

$$\mathcal{H} = \frac{\hat{p}\cdot\hat{p}}{2m} + V(\vec{r}) \quad , \tag{I.2.16a}$$

where $\hat{p}$ is the momentum operator, and

$$V(\vec{r}) = V_o + V_{LIN}(\vec{r}) + V_1(\vec{r}) + V_2(\vec{r}) \; ; \tag{I.2.16b}$$

where

$$V_{LIN}(\vec{r}) = \frac{1}{2!} V_{ij} r_i r_j$$

$$V_1(\vec{r}) = \frac{1}{3!} V_{ijk} r_i r_j r_k \tag{I.2.16c}$$

$$V_2(\vec{r}) = \frac{1}{4!} V_{ijk\ell}\, r_i r_j r_k r_\ell$$

with, for example,

$$V_{ijk} \equiv \left(\frac{\partial^3 V(\vec{r})}{\partial r_i \partial r_j \partial r_k}\right)_{\vec{r}=0}$$

and, in the principal axes reference frame,

$$V_{ij} = m\omega_i^2\, \delta_{i,j} \qquad \text{(no sum on i)}. \tag{I.2.16d}$$

The constants, ω_i, are the unperturbed normal frequencies of oscillation. The classical behavior of this system is given by

$$m\frac{d^2\vec{r}}{dt^2} + \nabla V(\vec{r}) = q\vec{\mathcal{E}}^m = \frac{q}{2}\left(\vec{E}^m e^{-i\omega_o t} + \vec{E}^{m*} e^{i\omega_o t}\right) .$$

By assuming that the particle is at rest ($\vec{r} = 0$) in the absence of $\vec{\mathcal{E}}$, the solution for $\vec{r}(t)$ can be found in a perturbation series involving increasing factors of $\vec{E}^m$. As in the quantum mechanical case, the "contributing" terms of the resultant polarization, $\vec{\mathcal{P}}^m(t) = q\vec{r}(t)$ can be found by choosing the terms varying at ω_o. The principal linear polarizabilities are then given by the terms linear in $\vec{E}^m$; these are

$$\alpha_i^m = \frac{q^2}{m(\omega_i^2 - \omega_o^2)} . \qquad \text{(I.2.17)}$$

As before, the cubic field terms also contribute to $\vec{P}^m$; defining the susceptibility tensor, χ^m, as in equation I.2.13, we obtain

$$\chi^m_{ijk\ell} = \frac{q^4}{8m^4(\omega_i^2-\omega_o^2)(\omega_j^2-\omega_o^2)(\omega_k^2-\omega_o^2)(\omega_\ell^2-\omega_o^2)}\left\{-V_{ijk\ell} + \sum_n\left(\frac{V_{i\ell n}V_{njk}}{m(\omega_n^2-4\omega_o^2)} + 2\frac{V_{ijn}V_{nk\ell}}{m\omega_n^2}\right)\right\} . \qquad \text{(I.2.18)}$$

Again, equation I.2.15 is satisfied both in the "lossless" case, and when $V_1 << V_2$.

Before the susceptibilities calculated above can be used to describe the macroscopic polarizations, the isotropy of the material must be taken into account. Such isotropy can occur in two ways; either the individual molecules are highly symmetric with isotropic

polarizabilities, or else the molecules may be oriented at random, as in fluids or glasses. We will now show that the condition of isotropy simplifies the susceptibility tensors considerably.

The case of a symmetrical polarization source will be considered in the classical case; the potential in equation I.2.16 takes the form

$$V(\vec{r}) = V(|r|) = V_o + \frac{1}{2} m\omega_a^2 (x^2+y^2+z^2) + \frac{V_2^o}{4!} (x^2+y^2+z^2)^2 \quad . \tag{I.2.19}$$

The tensor components of equation I.2.18 now reduce to

$$\chi^m_{iiii} = \frac{3}{2} \chi^m_o \quad , \chi^m_{iijj} = \chi^m_{ijij} = \chi^m_{ijji} = \frac{1}{2} \chi^m_o \qquad j \neq i \quad , \tag{I.2.20}$$

where

$$\chi^m_o = \frac{-V_2^o q^4}{12m^4(\omega_a^2-\omega_o^2)^4} \quad .$$

Therefore, for a symmetrical molecule,

$$\vec{P}^m_{NL} = \chi^m_o(\vec{E}^m(\vec{E}^m\cdot\vec{E}^{m*}) + \frac{1}{2} \vec{E}^{m*}(\vec{E}^m\cdot\vec{E}^m)) \quad . \tag{I.2.21}$$

Next the situation of randomly oriented molecules will be considered. The most direct approach would be to keep the field, $\vec{E}^m$, constant while rotating the molecule. Classically, this would mean transforming the apparent $V(\vec{r})$, which would manifest itself in different values of $V_{ijk\ell}$ in the susceptibility tensor elements in equation I.2.18. However, since V_{ij} in equation I.2.16 would no longer be diagonal, the formulas for α^m_{ij} and $\chi^m_{ijk\ell}$ would

have to be generalized. Quantum mechanically, a similar approach would transform the base states, u_n, as the molecule rotated in the surrounding medium; this situation is complicated by the discrete nature of these states, requiring a more complicated definition of "random orientation". These difficulties can be avoided, however, as long as the fields, $\vec{E}^m$ and $\vec{P}^m$, are considered to behave classically; the molecule, whether treated classically or quantum mechanically, can then be treated as a "black box" characterized by linear and cubic susceptibility tensors, such as those of equations I.2.11, I.2.14, I.2.17, and I.2.18.

We will define two reference frames: first, the lab frame, (x,y,z), with fields $\vec{E}^m$, $\vec{P}^m$, and second, the molecular frame, (x',y',z'), in which the fields are $\vec{E}^{m'}$, $\vec{P}^{m'}$. For example, Figure 2 shows how these frames might be related by using the Eulerian angles[27], ϕ,Θ,ψ. We will denote the matrix, $M(\vec{\Omega})$, relating the field components in the two frames for a given orientation, $\vec{\Omega}$, of the molecule; it will be unitary (and real). The field components will therefore transform according to the usual equations

$$E_i^{m'} = M_{ij}(\vec{\Omega})E_j^m$$
$$E_i^m = M_{ij}^{-1}(\vec{\Omega})E_j^{m'} = M_{ji}(\vec{\Omega})E_j^{m'} \quad . \tag{I.2.22}$$

Appendix 1 gives $M(\vec{\Omega})$ for the Euler transformation, and other needed formulas concerning integrations over $\vec{\Omega}$.

We can now find the average polarizability by finding the polarization produced by a molecule characterized by $\vec{\Omega}$.

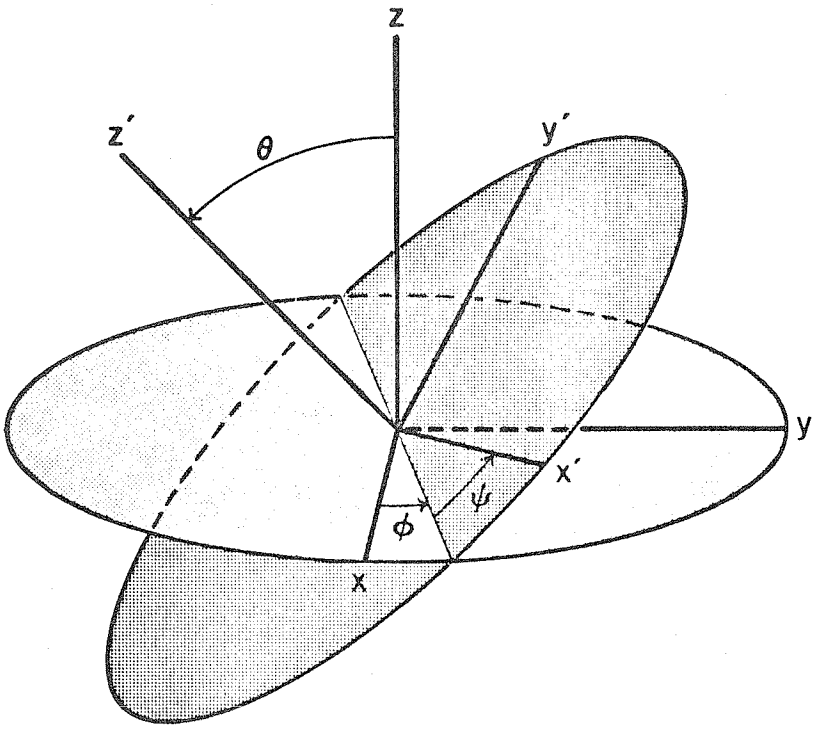


Figure 2. Orientation of molecule as described by Euler angles. The laboratory coordinate frame is labeled (x,y,z), with the (x,y,z=0) plane lightly shaded; the molecular coordinate frame is labeled (x',y',z'), with heavy shading on the (x',y',z'=0) plane. The transformation between the two is described by (ϕ,Θ,ψ) corresponding to: (1) rotation about z by ϕ; (2) rotation about the new x axis by Θ; (3) rotation about z' by ψ.

$$(P^m_{LIN}(\vec{\Omega}))_i = M_{i,i'}(\vec{\Omega})(P^{m'}_{LIN})_{i'} = M_{i,i'}(\vec{\Omega})(\alpha^m_{i',j'}E^{m'}_j) = (M_{i,i'}(\vec{\Omega})\alpha^m_{i',j'}M_{j,j'}(\vec{\Omega}))E^m_j \quad .$$

For randomly oriented molecules, this formula is averaged over $\vec{\Omega}$ giving

$$\langle (P^m_{LIN})_i \rangle = \int (P^m_{LIN}(\vec{\Omega}))_i d^3\vec{\Omega} = \alpha^m_{i',j'}E^m_j \int M_{i,i'}(\vec{\Omega})M_{j,j'}(\vec{\Omega})d^3\vec{\Omega} \quad .$$

The integrals are given in Appendix 1; the resulting formula for $\vec{P}^m$ indicates that the linear response is described by a (scalar) average polarizability, $\bar{\alpha}^m$, as given by

$$\langle (P^m_{LIN})_i \rangle = \bar{\alpha}^m E^m_i \quad ,$$

where (I.2.23)

$$\bar{\alpha}^m = \frac{1}{3}(\alpha^m_{xx} + \alpha^m_{yy} + \alpha^m_{zz}) \quad .$$

The average nonlinear polarization is similarly employed to derive the average cubic susceptibility tensor:

$$(P^m_{NL}(\vec{\Omega}))_i = M_{i,i'}(\vec{\Omega})(P^{m'}_{NL})_{i'} = M_{i,i'}(\vec{\Omega})\chi^m_{i',j',k',\ell'}E^{m'}_j E^{m'}_k E^{m'*}_\ell$$

$$= (\chi^m_{i',j',k',\ell'}M_{i,i'}(\vec{\Omega})M_{j,j'}(\vec{\Omega})M_{k,k'}(\vec{\Omega})M_{\ell,\ell'}(\vec{\Omega}))E^m_j E^m_k E^{m*}_\ell \quad .$$

This gives

$$\langle (P^m_{NL}) \rangle = \bar{\chi}^m_{ijk\ell} E^m_j E^m_k E^{m*}_\ell \quad , \qquad \text{(I.2.24)}$$

where the components of $\bar{\chi}^m$ are related to χ^m components as follows.

Define:

$$\chi_1 \equiv \sum_j \chi^m_{jjjj}$$

$$\chi_2 \equiv \sum_{j,k\neq j} (\chi^m_{jjkk} + \chi^m_{jkjk}) \qquad \text{(I.2.25a)}$$

$$\chi_3 \equiv \sum_{j,k\neq j} \chi^m_{jkkj} \ ,$$

then we find, as the only nonvanishing components of $\overset{-m}{\chi}$,

$$\overset{-m}{\chi}_{iiii} = \frac{1}{5}\chi_1 + \frac{1}{15}\chi_2 + \frac{1}{15}\chi_3$$

$$(\overset{-m}{\chi}_{iijj} + \overset{-m}{\chi}_{ijij}) = \frac{2}{15}\chi_1 + \frac{1}{10}\chi_2 - \frac{1}{15}\chi_3 \qquad j\neq i \qquad \text{(I.2.25b)}$$

$$\overset{-m}{\chi}_{ijji} = \frac{1}{15}\chi_1 - \frac{1}{30}\chi_2 + \frac{2}{15}\chi_3 \qquad j\neq i \ .$$

This shows

$$\langle \vec{P}^m_{NL} \rangle = (\frac{2}{15}\chi_1 + \frac{1}{10}\chi_2 - \frac{1}{15}\chi_3)\vec{E}^m(\vec{E}^m\cdot\vec{E}^{m*}) + (\frac{1}{15}\chi_1 - \frac{1}{30}\chi_2 + \frac{2}{15}\chi_3)\vec{E}^{m*}(\vec{E}^m\cdot\vec{E}^m) \ . \qquad \text{(I.2.25c)}$$

In the "lossless" limit, equation I.2.15 shows that $\chi_2 = 2\chi_3$, and so the average polarization formula becomes

$$\langle P^m_{NL} \rangle_{\text{lossless}} = \overset{-m}{\chi}_E(\vec{E}^m(\vec{E}^m\cdot\vec{E}^{m*}) + \frac{1}{2}\vec{E}^{m*}(\vec{E}^m\cdot\vec{E}^m)) \ , \qquad \text{(I.2.26a)}$$

with

$$\overset{-m}{\chi}_E = \frac{2}{15}(\chi_1 + \chi_3) \ .$$

The subscript "E" indicates that this applies to the electronic

polarization mechanism for causing the nonlinearity. These formulas apply to both the quantum mechanical and the classical cases, by using the corresponding tensors for α^m and χ^m. In a more detailed calculation, a further averaging over the thermal distribution of "ground states" would also have to be performed, but this would not affect the form of the results at this point. Finally, we use equation I.2.4 to obtain the macroscopic polarization, which gives:

$$(\vec{P}_{NL})_E = \bar{\chi}_E(\vec{E}(\vec{E}\cdot\vec{E}^*) + \frac{1}{2}\vec{E}^*(\vec{E}\cdot\vec{E})) \quad ,$$

where

$$\bar{\chi}_E = (\frac{\varepsilon+2}{3})^4 N \; \bar{\chi}_E^m \quad . \qquad \text{(I.2.26b)}$$

We can now form rough estimates of the physical magnitudes involved in these nonlinear coefficients in the following way. First, consider the quantum mechanical results, expressed in equations I.2.11 and I.2.14. If the sums over dipole moments are replaced by one "effective" moment, it is seen that the following approximation holds:

$$\chi_E^m \simeq (\alpha^m)^2 \, (\frac{3}{4\hbar(\omega_a-\omega_o)}) \qquad \text{(Quantum Mechanical)} \quad \text{(I.2.27a)}$$

where $\hbar\omega_a$ is an average "effective" energy gap. A similar approach for the classical formulas yields the approximation

$$\chi_E^m \simeq (\alpha^m)^2 (\frac{-V_2(r_a)}{8m^2 r_a^4 (\omega_a^2-\omega_o^2)^2}) \quad ,$$

where ω_a is a typical resonant frequency, and r_a is some typical

molecular length. Choose r_a to be the distance at which the potential becomes highly nonlinear; an example would be a periodic potential such as $V(r) \simeq V_o(1-\cos(\frac{\pi r}{r_a}))$. Then $m\omega_a^2 r_a^2 \simeq |V_2(r_a)| = V_o$, giving the approximate formula

$$\chi_E^m \simeq (\alpha^m)^2 \left(\frac{\omega_a^4}{8(\omega_a^2-\omega_o^2)^2 V_o}\right) \qquad \text{(Classical)} \quad . \qquad \text{(I.2.27b)}$$

We can now estimate the ionic polarization, relative to the electronic, and see that the former is generally much smaller in contributing to α^m or χ^m . Since ions see a potential $V(\vec{r})$ of about the same magnitude as electrons, equation I.2.16d indicates that the resonant frequency will be smaller by the ratio $(m_{elec}/m_{ion})^{\frac{1}{2}}$, generally leading to resonances in the infrared. Since $\omega_o >> \omega_{ion}$, $\alpha^m_{ion}/\alpha^m_{elec} \simeq (m_{elec}/m_{ion})$, and $\chi^m_{ion}/\chi^m_{elec} \approx (\frac{m_{elec}}{m_{ion}})^4 << 1$. Therefore, in the case of molecular polarizations in the optical range, we can identify α^m_{elec} with $\bar{\alpha}^m$ and consequently $\bar{\alpha}$. Since we can thus use the "lossless" limit, $\omega_o << \omega_a \approx \omega_{elec}$, both formulas in equation I.2.27 show that $\chi^m/(\alpha^m)^2 \approx 1/(\text{typical excited state energy})$. Finally, when the results of the local field effects are included, as indicated in equation I.2.4b, it can be seen that a rough estimate of χ_E is found from

$$\chi_E \approx (\frac{\varepsilon+2}{3})^2(\frac{\varepsilon-1}{4\pi})^2/(4\cdot N\cdot \text{Excited state energy}) \quad . \quad \text{(I.2.28)}$$

In a typical transparent material, the (electronic) excited state resonance, ω_{res}, lies in the ultraviolet. Taking, as an example,

a "typical" optical glass, we find $N \approx 2.7 \times 10^{22} cm^{-3}$, $\varepsilon = n_o^2 = (1.52)^2$, $\lambda_{res} \tilde{=} 2800\text{Å}$; this gives $\chi_E \sim 2.9 \times 10^{-14}$ e.s.u. An experimental value[28] for a laser glass is about 1.4×10^{-14} e.s.u.

When $\vec{E}$ varies in time, the field contains multiple frequencies, and the specific formulas such as I.2.14 do not apply. A general formula which allows for different frequency components, ω_{field}, such as that given in Armstrong et al[6], would have to be summed over all frequency combinations. However, for $\omega_{field} << \omega_{res}$, the response is nearly independent of ω_{field}; this is consistent with the fact that the electronic distribution can respond to perturbations in times on the order of ω_{res}^{-1}, and thus the polarizabilities can be described as following the instantaneous field adiabatically. This view would suggest simply putting the instantaneous fields in equations I.2.13, I.2.21 and I.2.26, with the "monochromatic" values of χ_E. However, since long propagation distances will be considered, even the small frequency dependencies remaining in α and χ might be significant. The largest dispersive effect, of course, is the linear frequency dependence in α. This leads to the usual "group velocity" of pulses, but does not otherwise affect linear propagation; furthermore, it can be shown that this order of dispersion does not qualitatively alter the nonlinear effects under discussion[13]. All other higher order dispersive properties of $\bar{\alpha}$ and $\bar{\chi}_E$ will be ignored in the present investigation, since the cubic polarization itself will "dominate" the pulse distortions. Therefore, the response time of the electronic nonlinearity, τ_E, will be taken to be zero.

I.2.ii orientational Kerr effect

In the preceding section, an effective molecular polarizability, $\bar{\alpha}^m$, was found by averaging the individual polarizabilities, α^m, over all possible molecular orientations, $\vec{\Omega}$, according to equations I.2.22 and I.2.23. In doing this, we implicitly made two related assumptions: first, that the molecular wave functions, u_n, or potentials, $V(\vec{r})$, were defined in a fixed molecular reference frame, unaffected by the field, and second, that all molecular orientations, relative to the lab frame, were equally likely. As shown in Figure 1, however, a molecule with an anisotropic polarizability experiences torques which tend to align the polarization with the applied electric field direction. In a liquid, this results in molecular reorientations which can destroy the isotropy and perturb the polarizability tensor by an amount dependent on the field; this is called the orientational Kerr effect.

We can analyze this effect by modifying our approach toward the linear polarizability in two ways[29]. We will remain in the laboratory frame and consider the molecule to be characterized by $\vec{\Omega}$ in its response; also, we will assume that a thermal equilibrium condition is enforced by the neighboring molecules, so that, in the absence of an applied field, the situation of "random" orientation results. As before, we can avoid a quantum mechanical description[30] of the molecule's response under these conditions by assuming that all fields and polarizations are classical. In order to conserve the total (electromagnetic plus internal) energy, we can state that the

energy of a polarizable molecule in the field is changed by an amount equal to W, with

$$W = W(\vec{\Omega}) = -\frac{1}{2}\vec{\mathcal{P}}^m(\vec{\Omega})\cdot\vec{\mathcal{E}}^m. \qquad (I.2.29)$$

When thermal equilibrium is reached, the probability, $f(\vec{\Omega})$, that the molecule will have an orientation, $\vec{\Omega}$, will be given by

$$f(\vec{\Omega}) = e^{-W(\vec{\Omega})/kT}/\left(\int e^{-W(\vec{\Omega})/kT}d^3\vec{\Omega}\right) . \qquad (I.2.30)$$

We will use the slowly varying part of W, because the molecular inertia prevents responses at $\pm\omega_o$, and because the thermal equilibrium mechanism (collisions) cannot be effective in times as short as ω_o^{-1}. Since the dot product is invarient under $\vec{\Omega}$, we can find, from equations I.2.8, I.2.12, and I.2.22, the function $W(\vec{\Omega})$:

$$\begin{aligned} W(\vec{\Omega}) &= -\frac{1}{8}(\vec{P}^{m'}\cdot\vec{E}^{m'*} + \vec{P}^{m'*}\cdot\vec{E}^{m'}) \\ &= -\frac{1}{4}(\alpha^m_{i'j'}\, E^{m'}_{i'}\, E^{m'*}_{j'}) \\ &= -\frac{1}{4}\,\alpha^m_{i'j'}M_{i'i}(\vec{\Omega})M_{j'j}(\vec{\Omega})E^m_i E^{m*}_j \quad . \end{aligned}$$

Now, from equation I.2.30, we can expand $f(\vec{\Omega})$ out to terms quadratic in $\vec{E}^m$.

$$\begin{aligned} f(\vec{\Omega}) &\simeq (1-W(\vec{\Omega})/kT)\left(\int (1-W(\vec{\Omega})/kT)d^3\vec{\Omega}\right)^{-1} \\ &= (1 - W(\vec{\Omega})/kT)\left(1 + \frac{1}{4kT}\,\alpha^m_{i'j'}E^m_i E^{m*}_j\int M_{i'i}(\vec{\Omega})M_{j'j}(\vec{\Omega})d^3\vec{\Omega}\right)^{-1} \\ &= (1 - W(\vec{\Omega})/kT)\left(1 + \frac{1}{4kT}\,\bar{\alpha}^m(\vec{E}^m\cdot\vec{E}^{m*})\right)^{-1} . \end{aligned}$$

Thus,

$$f(\vec{\Omega}) = 1 - \frac{\bar{\alpha}^{m}(E^{m}\cdot E^{m^*})}{4kT} + \frac{1}{4kT}\,\alpha_{i'''j'''}E^{m}_{i''}E^{m^*}_{j''}\,M_{i'''i''}(\vec{\Omega})M_{j'''j''}(\vec{\Omega}) \quad , \qquad (I.2.31)$$

where $\bar{\alpha}^{m}$ is defined in equation I.2.23. To calculate the effective polarizability of a molecule at $\vec{\Omega}$, defined as $\alpha^{m}_{ij}(\vec{\Omega})$, we write

$$P^{m}_{i}(\vec{\Omega}) = M_{i'i}(\vec{\Omega})P^{m'}_{i'} = M_{i'i}(\vec{\Omega})\alpha^{m}_{i'j'}E^{m'}_{j'} = \alpha^{m}_{i'j'}M_{i'i}(\vec{\Omega})M_{j'j}(\vec{\Omega})E^{m}_{j} \quad ,$$

which implies that the polarizability in the laboratory frame is given by

$$\alpha^{m}_{ij}(\vec{\Omega}) = \alpha^{m}_{i'j'}M_{i'i}(\vec{\Omega})M_{j'j}(\vec{\Omega}) \quad . \qquad (I.2.32)$$

Finally, the thermally averaged polarizability is given by weighting $\alpha^{m}_{ij}(\vec{\Omega})$ by $f(\vec{\Omega})$; using the integral formulas given in Appendix 1, we obtain:

$$\begin{aligned}
\langle \bar{\alpha}^{m}_{ij} \rangle &= \int \alpha^{m}_{ij}(\vec{\Omega}) f(\vec{\Omega}) d^{3}\vec{\Omega} \\
&= \left(1 - \frac{\bar{\alpha}^{m}}{4kT}\,(\vec{E}^{m}\cdot\vec{E}^{m^*})\right)\alpha^{m}_{i'j'}\int M_{i'i}(\vec{\Omega})M_{j'j}(\vec{\Omega})d^{3}\vec{\Omega} \\
&+ \frac{1}{4kT}\,E^{m}_{i''}E^{m^*}_{j''}\,\alpha^{m}_{i'j'}\alpha^{m}_{i'''j'''}\int M_{i'''i''}(\vec{\Omega})M_{j'''j''}(\vec{\Omega})M_{i'i}(\vec{\Omega})M_{j'j}(\vec{\Omega})d^{3}\vec{\Omega} \\
&= \delta_{i,j}\left(\bar{\alpha}^{m} - \frac{\Delta^{2}}{180kT}\,\vec{E}^{m}\cdot\vec{E}^{m^*}\right) + \frac{\Delta^{2}}{120\ kT}\left(E^{m}_{i}E^{m^*}_{j} + E^{m^*}_{i}E^{m}_{j}\right) \\
&\equiv \bar{\alpha}^{m}\delta_{i,j} + (\bar{\alpha}^{m}_{NL})_{ij} \quad , \qquad (I.2.33)
\end{aligned}$$

where

$$\Delta^{2} \equiv (\alpha^{m}_{xx} - \alpha^{m}_{yy})^{2} + (\alpha^{m}_{xx} - \alpha^{m}_{zz})^{2} + (\alpha^{m}_{yy} - \alpha^{m}_{zz})^{2} \quad .$$

When the local field effects, as shown in equation I.2.4, are taken into

account, the orientational Kerr effect can be described by the macroscopic polarization, $(\vec{P}_{NL})_{KERR}$, where

$$(\vec{P}_{NL})_{KERR} = \bar{\chi}_K(\vec{E}(\vec{E}\cdot\vec{E}^*) + 3\vec{E}^*(\vec{E}\cdot\vec{E})) \quad , \qquad (I.2.34)$$

and

$$\bar{\chi}_K = \frac{\Delta^2}{360\ kT}\left(\frac{\varepsilon+2}{3}\right)^4 \quad .$$

This result for the Kerr effect corresponds to equation I.2.26b for the electronic nonlinearity. The analogous equation for the magnitude of $\bar{\chi}_K$ is thus

$$\bar{\chi}_K \approx \left(\frac{\varepsilon+2}{3}\right)^2\left(\frac{\varepsilon-1}{4\pi}\right)^2 \frac{(\Delta^2/(\bar{\alpha}^m)^2)}{90\ kTN} \quad .$$

The factor, $\Delta^2/(\bar{\alpha}^m)^2$, ranges from zero (for an isotropic molecule) to 18 (for an idealized molecule with only one axis of polarizability). As a numerical example, consider carbon disulfide ($T = 293°K$, $N = 1.0 \times 10^{22} cm^{-3}$, $\varepsilon = n_o^2 = (1.6255)^2$), which gives $\bar{\chi}_K \simeq 1.1 \times 10^{-12}(\Delta^2/(\bar{\alpha}^m)^2)$ e.s.u.; this is consistent with a measured value[31] of $\bar{\chi} \approx 2 \times 10^{-13}$ e.s.u. if $\Delta^2/(\bar{\alpha}^m)^2 \gtrsim 0.2$. In general, for a reasonably anisotropic molecule, the Kerr effect nonlinearity is much larger than the corresponding electronic nonlinear polarizability, since $kT << \hbar\omega_{res}$.

In considering the response of the Kerr effect to time variations in $\vec{E}$ (i.e. pulses), the model used above shows that the net nonlinear polarization results from two distinct effects: first, the "average" molecule reorients itself and in the process alters its net polarizability, and second, the normal linear polarization is induced

in these reoriented molecules by the field. We have previously concluded that this last process is effectively instantaneous, but clearly the first response is not, since it involves both a movement of the (relatively massive) molecular nuclei and also a thermalization with the surrounding medium.

The most common model[32] for the rotational response of the molecules is that the surrounding medium produces a viscosity, η, which inhibits rapid rotations; in this view, the molecules will "diffuse", in $\vec{\Omega}$ space, toward their equilibrium distribution. Thus, $f(\vec{\Omega})$ will obey a diffusion equation, driven by the torques resulting from the potential, $W(\vec{\Omega},t) = -\frac{1}{2}\vec{\rho}(\vec{\Omega},t)\cdot\vec{\mathcal{E}}(t)$. Of interest here is the resultant "relaxation" equation obeyed by the distribution, $f(\vec{\Omega})$, and in turn by the polarizability tensor elements, $(\bar{\alpha}^{m}_{NL}(t))_{ij}$:

$$\tau_K \frac{d\bar{\alpha}^{m}_{NL}}{dt} = \bar{\alpha}^{m}_{o}(\vec{E}^{m}) - \bar{\alpha}^{m}_{NL} \qquad (I.2.35)$$

where $\bar{\alpha}^{m}_{o}(\vec{E}^{m})$ is the instantaneous equilibrium value of $\bar{\alpha}^{m}_{NL}$, as given by the part of equation I.2.33 which is quadratic in the field amplitudes. In reality, the effective Kerr relaxation time, τ_K, depends on the polarizations and field magnitudes to some degree[33], but for the cases analyzed later, equations I.2.34 and I.2.35 are sufficiently accurate. In these cases, $\tau_K = \xi/6kT$, with ξ, the Stokes constant, given approximately by $8\pi\eta r_e^3$, and r_e being the effective molecular radius. For carbon disulfide, using $r_e \approx 2\text{Å}$, and $\eta = 0.363$ c.p., we obtain $\tau_K \simeq 3 \times 10^{-12}$sec; since the measured value[34], $\tau_K = 1.9 \times 10^{-12}$sec, is so close to

this, we can assume that the "viscous medium" model is qualitatively accurate for the Kerr effect (on this time scale).

When we later refer to the response time of the nonlinear polarizability, $\bar{\alpha}$, or of the corresponding nonlinear dielectric constant, ε, or refractive index, n, we will assume that an equation of the relaxation type (I.2.35) holds, even though other nonlinear effects, such as electrostriction, may not be accurately described by a simple relaxation rate, τ. Although the steady state value, $\bar{\alpha}_0$, may be relatively large, in equation I.2.35, the nonlinearity will not be important for optical pulse envelopes much shorter than τ since the actual values of $\alpha(t)$ will then be much smaller than $\bar{\alpha}_0$.

I.2.iii electrostrictive nonlinearity

According to Figure 1, a polarizable molecule tends to move into the region of highest electric field amplitude; such a force, in a material of uniform unperturbed density, ρ_0, will create stresses, which are called electrostrictive forces. The resultant strains will lead to density variations,

$$\rho(\vec{r},t) = \rho_0 + \rho_1(\vec{r},t)$$

which, for $\rho_1 << \rho_0$, will be governed by the equations for macroscopic acoustic wave propagation. An accurate description of this effect in solids is complicated by the fact that the electrostrictive stresses and strains are not isotropic, even in an isotropic unperturbed

material[35]. Many of the following equations will therefore hold strictly only for fluids or in the steady state limit.

We have shown that, in an isotropic medium, the linear polarizability, $\bar{\alpha}$, is a scalar such that $\vec{\mathcal{P}} = \bar{\alpha}\vec{\mathcal{E}}$, and that the polarization energy of the medium, per unit volume, is given by $W(\vec{r},t) = -\frac{1}{2}\bar{\alpha}(\vec{r},t)\vec{\mathcal{E}}\cdot\vec{\mathcal{E}}$. By considering the change in this energy caused by the movement of material from one position to another, we find that the electrostrictive force per unit volume depends on the variation of $\bar{\alpha}$ with density, according to the equation

$$\vec{F} = \frac{1}{2}\rho\frac{\partial\bar{\alpha}}{\partial\rho}\nabla(\vec{\mathcal{E}}\cdot\vec{\mathcal{E}}) \approx \frac{1}{4}(\rho_o\frac{\partial\bar{\alpha}}{\partial\rho})\nabla(\vec{E}\cdot\vec{E}^*) \quad .$$

Only the slowly varying forces, as shown, can drive the acoustic waves.

This force drives the density changes, according to the small-signal acoustic wave equation[36], in the following way:

$$\nabla^2\rho_1 - \frac{1}{v_s^2}\frac{\partial^2\rho_1}{\partial t^2} + \tau_v\frac{\partial}{\partial t}(\nabla^2\rho_1) = \frac{1}{v_s^2}\nabla\cdot\vec{F} = \frac{1}{4v_s^2}(\rho_o\frac{\partial\bar{\alpha}}{\partial\rho})\nabla^2(\vec{E}\cdot\vec{E}^*) \quad .$$

Here, $\tau_v = (\frac{4}{3}\eta+\lambda)/\kappa$ and $\eta,\lambda,\kappa, v_s = (\kappa/\rho_o)^{1/2}$ are the coefficient of viscosity, coefficient of expansive friction, compressibility modulus and acoustic velocity, respectively. Defining $\alpha(\vec{r},t) \simeq \bar{\alpha} + (\rho_o\frac{\partial\bar{\alpha}}{\partial\rho})(\frac{\rho_1}{\rho_o}) \equiv \bar{\alpha} + \bar{\alpha}_{NL}(\vec{r},t)$, gives

$$\nabla^2\bar{\alpha}_{NL} - \frac{1}{v_s^2}\frac{\partial^2\bar{\alpha}_{NL}}{\partial t^2} + \tau_v\frac{\partial}{\partial t}(\nabla^2\bar{\alpha}_{NL}) = \frac{1}{4\rho_o v_s^2}(\rho_o\frac{\partial\bar{\alpha}}{\partial\rho})^2\nabla^2(\vec{E}\cdot\vec{E}^*) \quad . \tag{I.2.36}$$

The steady-state response is thus characterized by $\bar{\alpha}_{NL}$ proportional

to $\vec{E}\cdot\vec{E}^*$, leading to a macroscopic nonlinear polarization, $\vec{P}_{NL}$, and nonlinear electrostrictive susceptibility, $\bar{\chi}_{ES}$, given by

$$(\vec{P}_{NL})_{ES} = \bar{\alpha}_{NL}\vec{E} = \bar{\chi}_{ES}\vec{E}(\vec{E}\cdot\vec{E}^*) \qquad (I.2.37)$$

and

$$\bar{\chi}_{ES} = \frac{1}{4\rho_o v_s^2}\left(\rho_o \frac{\partial\bar{\alpha}}{\partial\rho}\right)^2 .$$

The value of the coupling constant, $\left(\rho_o \frac{\partial\bar{\alpha}}{\partial\rho}\right)$, can be found if the assumption is made that the average molecular polarizability, $\bar{\alpha}^m$, does not depend on the density. In this case, from equation I.2.4, we have

$$\rho_o \frac{\partial\bar{\alpha}}{\partial\rho} = N \frac{\partial\bar{\alpha}}{\partial N} = \bar{\alpha}\left(1 + \frac{4\pi}{3}\bar{\alpha}\right) = \left(\frac{\epsilon+2}{3}\right)\left(\frac{\epsilon-1}{4\pi}\right) . \qquad (I.2.38)$$

Using, for glass, $v_s = 2 \times 10^5 \text{cm sec}^{-1}$, $\rho_o = 2.9 \text{ gm cm}^{-3}$, $\epsilon = n_o^2 = (1.52)^2$, we have the estimate $\bar{\chi}_{ES} \approx 5 \times 10^{-14}$ e.s.u., in reasonable agreement with measured values[37].

The electrostrictive response to pulsed fields can be found from equation I.2.36, where the acoustical damping constant, τ_v, can usually be taken to be negligibly small. The electrostrictive nonlinearity, however, differs from the two previously considered, because it is not a "local" response; it is driven by the macroscopic variations in the field intensity. Coupled with the much slower acoustic velocity, this difference greatly complicates any analysis concerning pulses[38]. It is clear, however, that these complications can be ignored in either of the two extreme cases: very long, "steady-state" pulses, or very short ("ultrashort") pulses. In the latter case,

of course, the electrostrictive nonlinearity does not have time to develop, since the effective response time involved, τ_{ES}, is approximately equal to the time it takes for the acoustic wave to transverse the length of the intensity gradient (i.e. transverse beam dimension):

$$\tau_{ES} \approx \frac{\Delta r}{v_s}$$

Typically, $\Delta r \sim 10^{-2}$cm, giving $\tau_{ES} \sim 10^{-7}$ sec.

I.2.iv other sources of cubic nonlinearity

In liquids and solids, the influence of the surrounding medium on a molecule, represented by $\mathcal{H}'$ in equation I.1.3, causes correlations between neighboring polarization sources, so that the "single source" approach in sections I.2.i and I.2.ii is not quantitatively accurate. As an example, there are strong short-range fields which induce polarizations in an "unperturbed" molecule; this form of correlation will then affect the "incremental" polarization induced by an applied field[31]. Also, these correlations may depend on the relative positions of the neighboring molecules, leading to the "distributional" Kerr effect[39].

Our model of the orientational Kerr effect indicates that the torques on an anisotropic molecule can lead to other dynamical molecular behavior, besides an average diffusion toward an equilibrium orientation. First, only a small shift in the distribution function, $f(\vec{\Omega})$, is needed; this can occur even if the molecules are held in very "confining" potentials, such as occur in a solid. This potential

would have an angular dependence, however, and would therefore have to be included in our derivation of the susceptibility. Furthermore, the concept of a microscopic "viscosity" is not accurate for times less than the collision time; on this more rapid time scale, the torques on the molecule would lead to responses governed by inertial forces[40]. Finally, a number of other nonlinearities, or different interpretations of the same ones discussed above, have been investigated[41,42,43,44].

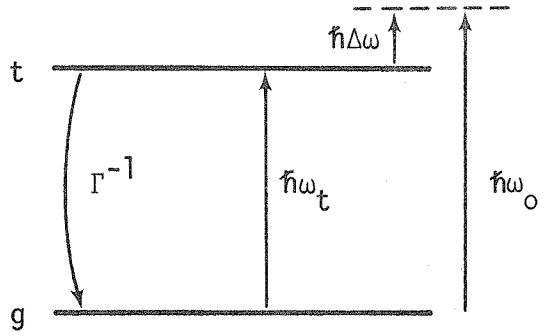
Next, there is a class of nonlinear effects of the "resonant" type which can be termed cubic nonlinearities in some situations; these are caused by resonances in the denominators in equations I.2.9 and I.2.14. These cases cannot be treated by the methods used in section I.2.i for a number of reasons. First, of course, the perturbation approach is not valid if excited levels become significantly occupied, or if exact resonances occur to cause mathematical divergences. Even for perturbations which are small and slightly off-resonance, however, the formulas would have to be modified to include terms which produce frequencies very close to ω_o; such polarizations will be important over short time scales, or if relaxation processes prevent the polarization from "dephasing" with the applied field. For similar reasons, the finite bandwidths of excited levels will allow "exact resonances" to occur in physical situations.

Quantitatively, we will consider the non-degenerate case when only a single excited state has a frequency near ω_o. This resonance can be heuristically accounted for by assigning the resonant level, "t", a complex frequency[12], leading to a bandwidth (and relaxation

rate), Γ, where

$$\omega_t' = \omega_t - i\Gamma \equiv \omega_o - \Delta\omega - i\Gamma \ . \qquad \text{(I.2.39)}$$

Here $\omega_o = \omega_t + \Delta\omega$ describes the relation of the applied frequency to the resonance frequency. Pictorially, the energy level diagram would appear as follows:



Carrying out the perturbation method outlined in section I.2.i, we obtain

$$(\alpha^m_{ij})_{res} = \frac{e^2}{\hbar} \frac{(r_i)_{gt}(r_j)_{tg}}{(\Delta\omega)^2 + (\Gamma)^2} \ (i\Gamma - \Delta\omega) \qquad \text{(I.2.40a)}$$

and

$$(\chi^m_{ijk\ell})_{res} = \frac{e^4}{\hbar^3} \frac{(r_i)_{gt}(r_j)_{tg}(r_\ell)_{gt}(r_k)_{tg}}{((\Delta\omega)^2 + (\Gamma)^2)^2} \ (\Delta\omega - i\Gamma) \ . \qquad \text{(I.2.40b)}$$

The form of the frequency dependence in $(\chi^m)_{res}$ actually depends somewhat on the "recipe" used in substituting for ω_t', indicating that the heuristic "complex frequency" approach may not be adequate for higher order resonances.

These tensors can now be averaged over angles according to equations I.2.23 and I.2.25, multiplied by N_{res}, the number density

of resonant molecules, corrected for local field effects as in equation I.2.4, and averaged over $\Delta\omega$ if inhomogeneous broadening is present. The real part of the resultant tensors then can be used to find the "intensity-dependent refractive index", while the imaginary parts will lead to (saturable) absorption (since our model was based on the ground state being occupied initially). It is also seen that the real part of $(\chi)_{res}$ will be positive when $\Delta\omega > 0$, so that a field "above" resonance will exhibit the same nonlinear effects as caused by the nonlinearities described in previous sections[45].

Because the material will no longer be lossless, and because of the strong frequency dependence of the resonance nonlinearity, the effects described by equation I.2.40 cannot realistically be treated simply by using polarizabilities and susceptibilities if pulses are involved. We will thus exclude "resonant" interactions from our list of cubic nonlinearities; if needed, they can be treated with more general, time-dependent equations[10]. However, if the relevant pulse "area", $e(r_i)_{gt}/\hbar \int_t^{t+1/\Gamma} E(t)dt$, is smaller than π, as will hold here, then equation I.2.40 describes, at least qualitatively, resonant effects.

I.3 Macroscopic Propagation Equations

The linear and nonlinear polarization sources discussed in the last section can all be described by the following steady-state equation for the component varying at $e^{-i\omega_o t}$:

$$\vec{P} = \bar{\alpha}\vec{E} + \bar{\chi}(\vec{E}(\vec{E}\cdot\vec{E}^*) + C_o\vec{E}^*(\vec{E}\cdot\vec{E})), \qquad (I.3.1)$$

where the electric field components are defined in equation I.2.1. From the quantum mechanical formula, I.2.11, for α^m, we can use the local field correction formula (I.2.4) and the isotropy condition (I.2.23) to obtain:

$$\bar{\alpha} = N\bar{\alpha}^m/(1 - \frac{4\pi}{3}N\bar{\alpha}^m) \ ,$$

where

$$\bar{\alpha}^m = \frac{2e^2}{3\hbar} \sum_n \frac{\omega_n}{(\omega_n^2 - \omega_o^2)} \ (\sum_i |(r_i)_{gn}|^2) \ .$$

We have also found that C_o equals 1/2, 3, and 0, for the electronic nonlinearity, orientational Kerr effect, and electrostriction, respectively; the corresponding susceptibilities, $\bar{\chi}_E, \bar{\chi}_K, \bar{\chi}_{ES}$, are given in equations I.2.26, I.2.34, and I.2.37. As shown in Appendix 1, the form of the nonlinear polarization given in equation I.3.1 must hold for any cubic polarization in an isotropic medium, with some C_o. We can thus consider this nonlinearity to be formally equivalent to an intensity-dependent polarizability tensor, $\bar{\alpha}_{NL}$; from equation I.3.1, then, we write

$$(\vec{P}_{NL})_i = (\bar{\alpha}_{NL})_{ij} E_j \qquad \text{(I.3.2)}$$

$$(\bar{\alpha}^o_{NL})_{ij} = \bar{\chi}(1 - C_o)(\vec{E}\cdot\vec{E}^*)\delta_{i,j} + \bar{\chi} C_o (E_i E_j^* + E_i^* E_j) \ .$$

The superscript "o" on $\bar{\alpha}^o_{NL}$ indicates that this is the steady state value, found by assuming $\vec{E}$ does not vary in time. The time dependence of the nonlinearity, to the accuracy required in these

analyses, will be assumed to be given by equation I.2.35; each element of $\bar{\alpha}_{NL}(t)$ thus obeys the relaxation equation:

$$\tau \frac{\partial \bar{\alpha}_{NL}}{\partial t} = \bar{\alpha}^{o}_{NL} - \bar{\alpha}_{NL} \quad . \tag{I.3.3}$$

When a combination of effects is present, the form of $\bar{\alpha}^{o}_{NL}$ must still hold, although C_o will differ from the three values listed earlier[46]. However, equation I.3.3 would not in general apply if the effects had different relaxation rates; we will ignore such "mixed" effects. For a response time, τ, which is short compared to the pulse envelope, we can of course put $\bar{\alpha}_{NL} = \bar{\alpha}^{o}_{NL}$. In our calculations, we will furthermore take $\bar{\alpha}$ and $\bar{\alpha}_{NL}$ to be real; we have seen that this holds when ω_o is far removed from any resonance.

The macroscopic beam propagations are now found by substituting the above polarization into Maxwell's equations. Eliminating $\vec{\mathcal{B}}$ from equation I.1.1 gives

$$\begin{aligned} \nabla \cdot \vec{\mathcal{E}} &= \frac{-4\pi}{n_o^2} \nabla \cdot \vec{\mathcal{P}}_{NL} \\ \nabla^2 \vec{\mathcal{E}} - \frac{n_o^2}{c^2} \frac{\partial^2 \vec{\mathcal{E}}}{\partial t^2} &= \frac{4\pi}{c^2} \frac{\partial^2 \vec{\mathcal{P}}_{NL}}{\partial t^2} - \frac{4\pi}{n_o^2} \nabla (\nabla \cdot \vec{\mathcal{P}}_{NL}) \end{aligned} \tag{I.3.4}$$

where $\frac{c^2 k_o^2}{\omega_o^2} = n_o^2 = \varepsilon = 1 + 4\pi \bar{\alpha}$. Using equations I.2.1 and I.2.2 to express $\vec{\mathcal{E}}$ and $\vec{\mathcal{P}}_{NL}$ in terms of traveling waves, yields equations for the complex amplitude functions, compatible with the nonlinear polarizability definitions:

$$E_z = -\frac{i}{k_o}\nabla\cdot\vec{E} + \frac{4\pi}{n_o^2}(P_{NL})_z - \frac{4\pi i}{n_o^2 k_o}\nabla\cdot\vec{P}_{NL} \qquad \text{(I.3.5)}$$

$$\frac{\partial\vec{E}}{\partial z} + \frac{n_o}{c}\frac{\partial\vec{E}}{\partial t} - \frac{i}{2k_o}\left(\frac{\partial^2\vec{E}}{\partial x^2} + \frac{\partial^2\vec{E}}{\partial y^2}\right) = \frac{2\pi i}{n_o^2}k_o\vec{P}_{NL} - \frac{4\pi}{n_o c}\frac{\partial\vec{P}_{NL}}{\partial t}$$

$$+ \frac{i}{2k_o}\frac{\partial^2\vec{E}}{\partial z^2} - \frac{in_o^2}{2k_o c^2}\frac{\partial^2\vec{E}}{\partial t^2} - \frac{2\pi i}{k_o c^2}\frac{\partial^2\vec{P}_{NL}}{\partial t^2} \qquad \text{(I.3.6)}$$

$$- \frac{2\pi}{n_o^2}\nabla(P_{NL})_z - \frac{2}{n_o^2}(\nabla\cdot\vec{P}_{NL})\hat{e}_z - \frac{2\pi i k_o}{n_o^2}(P_{NL})_z\hat{e}_z + \frac{2\pi i}{n_o^2 k_o}\nabla(\nabla\cdot\vec{P}_{NL}) \quad .$$

The initial conditions of interest here involve beams traveling in the +z direction, according to I.2.1; these beams are assumed to enter the material from free space, and we take $\bar{\alpha}_{NL}(\vec{r},t) = 0$ before the beam arrives. We will ignore the small distortions resulting from the boundary conditions at the material interface[47], and take the field $\vec{E}(0,t)$ as given at some point ($z = 0$) within the material. The beam will have transverse dimensions much larger than a wavelength, and have an envelope, $\vec{E}$, varying slowly in z and t relative to $e^{i(k_o z-\omega_o t)}$; the nonlinearities moreover are small over wavelength distances, so that $|P_{NL}| << |E|$, or $\bar{\alpha}_{NL} << 1$, can be taken. Under these conditions, $(P_{NL})_z << E_z << E_x, E_y$, and so equation I.3.5 gives, to order "∇"$/k_o$, the result:

$$E_z \simeq -\frac{i}{k_o}\left(\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y}\right) \quad . \qquad \text{(I.3.7)}$$

This shows, using equation I.3.6, that the transverse fields are essentially decoupled from E_z, the longitudinal field; $E_z \simeq 0$

can also be used in calculating $\vec{P}_{NL}$. To order "∇"/k_o, we find for E_x the propagation equation

$$\frac{\partial E_x}{\partial z} + \frac{n_o}{c}\frac{\partial E_x}{\partial t} - \frac{i}{2k_o}\left(\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2}\right) = \frac{2\pi i k_o}{n_o^2}(P_{NL})_x - \frac{4\pi}{n_o c}\frac{\partial (P_{NL})_x}{\partial t} \qquad (I.3.8)$$

with a similar equation for E_y.

It should be noted that there is a major qualitative difference between a beam governed by this equation and the true solution to equation I.3.6: $\vec{P}_{NL}$, by itself, "preserves" polarizations, to the extent that, if $E_y \equiv 0$ initially, then $(\vec{P}_{NL})_y = 0$ and $E_y = 0$ will result thereafter. Thus, the "decoupled" equation I.3.8 allows a linearly polarized solution. However, this is not true in the exact equation, I.3.6, since the last term on the right produces a source for E_y whenever $\frac{\partial^2 (P_{NL})_x}{\partial x \partial y}$ is non zero. Our later analyses using finite beams will not show this polarization coupling; qualitatively, however, these effects can be neglected for our purposes, since the "depolarized" fields produced are smaller by $\sim$ "∇"/k_o.

In many cases, the initial beam is sufficiently large, spatially, that it can be approximated by a plane wave, with $E_z = 0$; equation I.3.6 shows that the plane-wave character of the field is preserved exactly in this case. The analysis of the plane wave case leads to the concept of an "intensity-dependent refractive index", which yields a physical model of the macroscopic effects over small areas of a finite beam as well. Consider the case of instantaneous response ($\tau = 0$) in the steady-state problem, which is described by the equations :

$$\frac{\partial E_x}{\partial z} = \frac{2\pi i k_o}{n_o^2}\left[(\bar{\alpha}^o_{NL})_{xx}E_x + (\bar{\alpha}^o_{NL})_{xy}E_y\right]$$
$$\frac{\partial E_y}{\partial z} = \frac{2\pi i k_o}{n_o^2}\left[(\bar{\alpha}^o_{NL})_{yx}E_x + (\bar{\alpha}^o_{NL})_{yy}E_y\right] \qquad (I.3.9)$$

where $\bar{\alpha}^o_{NL} = \bar{\alpha}^o_{NL}(E_x,E_y)$ is given in equation I.3.3.

In the case of a plane polarized wave, $\vec{E} = E_x(z)\hat{e}_x$, setting $E_y = 0$ in equation I.3.9 yields the solution (for real $\bar{\alpha}^o_{NL}$)

$$E_x(z) = E_x(0)\, e^{i\left(\frac{2\pi k_o}{n_o^2}(\bar{\alpha}^o_{NL})_{xx}z\right)} .$$

A comparison with equation I.1.2 shows that the nonlinearity has manifested itself as a change in the wave vector, with $k_o \to k_o' = k_o\left(1 + \frac{2\pi}{n_o^2}(\bar{\alpha}^o_{NL})_{xx}\right)$. We can thus define an intensity-dependent index of refraction, $n(r,t)$, and a nonlinear index coefficient, n_2, as follows:

$$n(r,t) = n_o + \delta n(r,t) = n_o + n_2\langle\vec{\xi}\cdot\vec{\xi}\rangle, \qquad (I.3.10a)$$

where $\langle\vec{\xi}\cdot\vec{\xi}\rangle$ indicates that only the slowly varying part of the field intensity is to be taken. Thus, we have,

$$n(r,t) = n_o + \frac{n_2}{2}\,(\vec{E}\cdot\vec{E}^*) . \qquad (I.3.10b)$$

The refractive index is related to the phase velocity of the traveling wave according to $\frac{c}{n} = \frac{\omega_o}{k}$, which implies $n(r,t) = \frac{c}{\omega_o}k(r,t) = n_o\frac{k}{k_o}$. For the steady state (or instantaneous response) linearly polarized, plane wave case, we thus obtain:

$$n_2 = \frac{4\pi}{n_o} \frac{(\bar{\alpha}_{NL})_{xx}}{|E|^2} = \frac{4\pi}{n_o} \bar{\chi}(1 + C_o) \ . \qquad (I.3.11)$$

We can similarly define an intensity-dependent dielectric constant, $\vec{\mathcal{D}} = \varepsilon' \vec{\mathcal{E}}$, where $\varepsilon' = \varepsilon + \delta\varepsilon = \varepsilon + \varepsilon_2 \langle \vec{\mathcal{E}} \cdot \vec{\mathcal{E}} \rangle$; this gives, for the case discussed here,

$$\varepsilon_2 = 2n_o n_2 = 8\pi\bar{\chi}(1 + C_o) \ . \qquad (I.3.12)$$

When the plane-wave field is not linearly polarized, equation I.3.9 indicates that the orthogonal polarizations are coupled; this can be shown to lead to such phenomena as the rotation of the axes of elliptically polarized beams[12]. Such effects cannot be characterized simply by a single "nonlinear index", and will not be considered in our analyses. An exception is the case of circular polarization, given by $E_y = \pm iE_x$, where a single nonlinear index of the type shown in equation I.3.10 applies, with

$$(n_2)_{circular} = \frac{8\pi}{n_o} \bar{\chi} = (n_2)_{linear} \left(\frac{2}{1 + C_o}\right) \ .$$

Note that this circularly polarized beam carries twice the power density of the linearly polarized beam with equal field amplitudes, so that the ratio of index changes for beams with the same power density is given by $(\delta n)_{circular}/(\delta n)_{linear} = \frac{1}{1 + C_o}$. Otherwise, the propagation equations are identical, even in the presence of temporal or transverse structure (or relaxation effects); the circular polarization will hold everywhere, and only one component needs to be considered in the propagation equation.

In both cases, therefore, $\bar{\alpha}_{NL}$ reduces to an (intensity-dependent) scalar, and the beam propagation will be governed by the following equations, where $E(\vec{r},t)$, and $\delta n(\vec{r},t)$ are scalars:

$$\frac{\partial E}{\partial z} + \frac{(n_o+\delta n)}{c}\frac{\partial E}{\partial t} - \frac{i}{2k_o}\left(\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2}\right) = \frac{ik_o}{n_o}(\delta n)E$$

$$\tau \frac{\partial(\delta n)}{\partial t} = \frac{n_2}{2}EE^* - (\delta n) \ . \qquad \text{(I.3.13)}$$

This same result is obtained simply by setting $n_o \to n_o + \delta n$ in equation I.3.4 and ignoring $\vec{P}_{NL}$, thus justifying the physical model of a local "intensity-dependent refractive index" for all of the phenomena of interest here. Note, however, that, except for linearly or circularly polarized fields, the tensor aspects of $\bar{\alpha}_{NL}$ prevent such a simple view, since the expression for $\vec{P}_{NL}$ in equation I.3.8 does not reduce to the form in equation I.3.13.

We will now consider the consistency of some of our assumptions, and indicate the limitations imposed by them. In the absence of a nonlinearity, the power flow (in the +z direction) per unit area would be given by the quantity: (optical power/area) $= \frac{n_o c}{8\pi}(\vec{E}\cdot\vec{E}^*)$; if $\delta n << n_o$, we can continue to use this expression in the nonlinear case. For $\bar{\alpha}_{NL}$ (and δn) real, equation I.3.13 implies that the total beam power is conserved, and travels with velocity $\approx c/n_o$; therefore, the intensity $|E|^2$ can be estimated, as in the linear case, from the initial power if the beam area is known. A "typical" laser source for the beams of interest can produce a Q-switched pulse of about 1 Megawatt peak power (0.2 Joules in 2×10^{-7} sec.), or a

train of mode-locked pulses of about 1 Gigawatt peak power (10^{-3} Joules in 10^{-12}sec.); the peak (vacuum) electric field amplitudes, $|E|$, if these pulses have a cross section of 0.1 cm^2, are then 2.9×10^2 e.s.u. and 9.2×10^3 e.s.u., respectively. According to equations I.3.10 and I.3.11, the resultant nonlinear refractive index changes, for $n_o = 1.6$, are given by $\bar{\chi}(1+C_o)$ times 3.3×10^5 and 3.3×10^8, respectively. Our previous estimates of $\bar{\chi}$ ranged from about 10^{-14} (electronic nonlinearity) to 4×10^{-13} (Kerr effect), indicating that δn for such a beam would be at most $\sim 10^{-4}$. Since, in fact, the Kerr effect for most materials is usually ineffective for ultrashort pulses, due to $\tau_K > \tau_{pulse}$, we can conclude that $\delta n << n_o$, and $|P_{NL}| << |E|$ are good assumptions except under conditions of extreme amplification and focusing.

Finally, we can now estimate the importance of the production of third harmonic, since this determines if the "carrier waves", $e^{\pm i(k_o z - \omega_o t)}$, accurately describe the rapidly varying fields in the nonlinear material[48]. Of the major nonlinearities discussed in section I.2, only the electronic polarization can respond rapidly enough to produce such a third harmonic, which would then be given by $\vec{\mathcal{P}}_{NL}(3\omega_o) = \frac{1}{2}(\vec{P}_{NL}(3\omega_o)\ e^{3i(k_o z - \omega_o t)} + \text{c.c.})$; in this case, the methods of section I.2.i can be used to show that $\vec{P}_{NL}(3\omega_o) = \bar{\alpha}_{NL}(3\omega_o)\vec{E}$ where $\bar{\alpha}_{NL}(3\omega_o) \approx \frac{1}{2}\bar{\chi}_E(\vec{E}\cdot\vec{E})$. This polarization source, in turn, can be used in Maxwell's equations to yield a "steady-state" third harmonic field size given by:

$$\left|\frac{E(3\omega_o)}{E(\omega_o)}\right| \approx \left|\frac{n_o(3\omega_o)\delta n(3\omega_o)}{n_o^2(3\omega_o)-n_o^2(\omega_o)}\right|$$

with $\delta n(3\omega_o) \approx \delta n(\omega_o) < 10^{-4}$. Here $|n_o(3\omega_o) - n_o(\omega_o)| \gtrsim 10^{-2}$ because of dispersion, since this is equivalent to the problem of non phase-matched third harmonic. This shows that $|E(3\omega_o)| << |E(\omega_o)|$ will hold in the situations analyzed here, justifying the use of equation I.1.2 for the field.

Chapter II

ULTRASHORT PULSE GENERATION IN LASERS THROUGH SELF PHASE MODULATION

The use of saturable absorbers within the cavities of pulsed, solid state lasers has proven to be a very convenient method of "mode-locking" these light sources[16], resulting in the generation of ultrashort pulses with many practical applications[28,49]. In this chapter, we consider the analogous system in which the saturable absorber is replaced by a material exhibiting an intensity-dependent refractive index. For simplicity, we will refer to this material as the "Kerr medium", although any of the nonlinearities considered in Chapter I could be responsible for the effect if the response time is sufficiently short.

Here, we develop a model of such a laser system which is simplified enough to allow analytical solutions; nevertheless, we show that the physical mechanisms involved are well represented in this model, so that the pulsing properties of realistic systems can be predicted. We also examine the "stability" of the pulsing behavior, and find that the system described here has some advantages over the more common mode-locking mechanisms.

II.1 Laser System Model

The laser system under consideration contains three fundamental elements: an amplifier (gain medium), an oscillator feedback system (mirrors), and a "Kerr medium" which produces a cubic polarization

source for the fields in the cavity. The laser model is shown in Figure 3; also shown is a reference plane at which the traveling-wave fields are to be determined. In analyzing this laser model, we will assume that the various effects can be treated separately, so that, for example, the Kerr medium is lossless, and the gain medium does not display any nonresonant cubic polarization. These assumptions are not always valid in real lasers; for example, the refractive index nonlinearity in Nd:glass amplifiers is known to play a role in determining the pulsing characteristics[50]. However, the artificial "separation" of effects should not alter the physical principles involved, and may, in fact, help explain the properties of "complex" media such as Nd:glass.

The ability of the laser system shown in Figure 3 to produce ultrashort pulses can be explained on quite general grounds. The amplifying medium, which is typically Ruby or Nd:glass, has a sufficiently broad amplification bandwidth to produce such pulses, but normally the laser oscillates only over the frequencies having the highest gain. On the other hand, the Kerr medium generates a nonlinear polarization which couples the frequencies together and can thus broaden the oscillating spectrum. If the phases of the different frequency components are properly related, the output of the laser will consist of a train of "mode-locked" pulses[19]. The primary purpose of this chapter is thus to show why such pulses are the "favored" field configuration when an intensity-dependent refractive index is present.

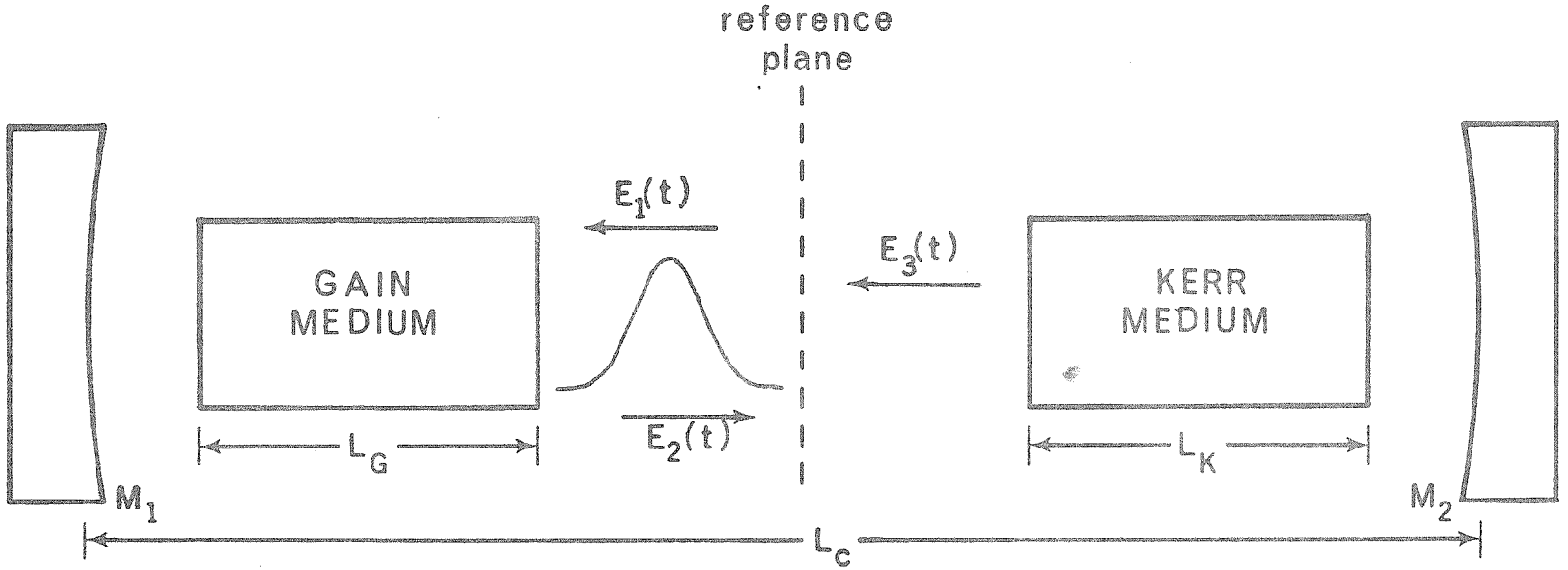


Figure 3. Laser model used in calculations. The optical resonator is formed by mirrors, M_1 and M_2, having reflectivities R_1 and R_2, respectively. The amplification is provided by the gain (laser) medium; also shown is the nonlinear "Kerr medium", which provides the cubic polarization needed to mode-lock the field. The pulses are measured at the reference plane indicated.

The field in the laser cavity will be taken to be linearly polarized, which is the case for most lasers in a "mode-locking" configuration. Furthermore, the analysis will be confined to plane-wave fields; the physical picture discussed later will show that, in general, any transverse structure on the beam would not affect our results, beyond altering some of the constants involved.

II.1.i circulating pulse approach

Because the Kerr medium introduces cubic polarization terms, as indicated in equation I.3.1, we can see that field components at frequencies ω_1, ω_2, and ω_3, will produce a polarization at $\omega_4 = \omega_1 + \omega_2 - \omega_3$ (as well as other frequencies). This observation suggests that an appropriate description of the field within the cavity might be an expansion in terms of resonator "modes", such as those based on the spatially complete set of functions, $e^{\pm im\frac{\pi z}{L_c}}$, as follows:

$$\vec{\xi}(\vec{r},t) = \vec{e}_x \sum_m \left\{ a_m(t) e^{im\frac{\pi}{L_c}(z-ct)} + a_m^*(t) e^{-im\frac{\pi}{L_c}(z-ct)} \right. \tag{II.1.1a}$$

$$\left. + b_m(t) e^{im\frac{\pi}{L_c}(z+ct)} + b_m^*(t) e^{-im\frac{\pi}{L_c}(z+ct)} \right\} .$$

Here, L_c (the cavity length) and z are measured in "optical" lengths, corrected for linear refractive indices, and the "modes" of interest are those near ω_t, the line center of the gain medium; thus $m \approx \frac{\omega_t L_c}{\pi c} \approx 10^6$. By substituting equation II.1.1a into the equations governing the gain medium, Kerr medium, and boundary conditions, and integrating over the cavity length to make use of the

orthogonality of the modes, we can obtain a set of coupled equations involving the derivatives of a_m and b_m. Finally, if we assume that the mode amplitudes vary slowly compared with the intermode frequency,

$$\left|\frac{d\,a_m}{dt}\right| << \frac{\pi c}{L_c}\ |a_m|$$

$$\left|\frac{d\,b_m}{dt}\right| << \frac{\pi c}{L_c}\ |b_m| \quad , \tag{II.1.1b}$$

the equations are "decoupled"; the evolution of the field can be found through a set of equations for $\frac{d\,a_m}{dt}$, $\frac{d\,b_m}{dt}$, in terms of a_n, b_n.

This approach was used in a previous analysis[20], which concluded that a "mode-locked" pulse, with nearly equal mode amplitudes and phases, was indeed a solution to the problem. However, the physical processes occurring in the cavity were not clarified by the analysis, and only very general "energy flow" arguments, similar to those in the introduction to this section, could be given. Furthermore, a numerical solution to the resultant "field evolution" equations was attempted, which indicated that there were two other major drawbacks to the "mode coupling" approach. First, a realistic laser model would require about $\Delta m = \frac{(\Delta\omega)_{pulse}}{(\Delta\omega)_{mode}} = \frac{\tau_{round\ trip}}{\tau_{pulse}} \sim 10^3$ modes in equation II.1.1a, to represent the field; this implies that the cubic polarization would involve nearly 10^9 terms. Computationally, then, the true ultrashort limit is impossible, and only a relatively few modes could be considered in our computer programs. Furthermore, the computer results showed that an even more fundamental

problem existed in the analysis, since they indicated very little tendency toward mode-locking. This contradicted our experiments[51], which had shown that the system shown in Figure 3 was, in fact, capable of producing ultrashort pulses.

This contradiction arises because of the incorrect assumption given in equation II.1.1b. Physically, this equation implies that the field does not change its form during a round trip transit of the cavity, which seems consistent with a mode-locked pulse with constant a_m and b_m. However, we can see from equation II.1.1a that a steady-state pulsing solution is also possible whenever $a_m(t)$ and $b_m(t)$ are periodic, with period $T = \frac{2L_c}{c}$; this condition includes the possibility of strong distortions occurring during the beam propagation, with the pulse reproducing itself after each round trip transit. The approach used in this chapter is designed specifically to allow for this situation; we will find that it also overcomes the other problems associated with the "coupled mode" approach.

The rapid variations in $a_m(t)$ and $b_m(t)$ indicate that the frequencies present in $\vec{\mathcal{E}}(\vec{r},t)$ are not confined to the "normal mode" frequencies, $m(\frac{\pi c}{L_c})$; therefore, we can consider the field at a given reference plane as consisting of repetitive pulses, each of which has a continuous frequency spectrum centered at ω_o. This is consistent with our analysis of the cubic nonlinearity given in Chapter 1. Such a view is called the "circulating pulse" approach, and is particularly suited to the present application, where the pulse physically occupies only a very small part of the cavity. The circulating pulse

techniques used here were developed by Cutler[52] in connection with the microwave regenerative pulse generator, and have been applied to some other laser systems in the past[49,53].

We will assume that a single pulse of light circulates in the system, successively passing through the various elements; in a laser, of course, this is accomplished by the light bouncing back and forth between the cavity reflectors, as shown in Figure 3. The elements inside the cavity are characterized by their effect on the pulse; this temporal "separation" of the pulse shaping process allows approximations to be made in describing each element without affecting the fundamental system behavior.

The circulating light pulse can be described by giving its time dependence as it passes any point, z_{ref}, such as the reference plane in Figure 3. In order to obtain closed-form expressions and analytical solutions, we will assume that this time dependence, as a first approximation, always has a Gaussian envelope, and a phase which may contain terms quadratic in time:

$$\vec{\mathcal{E}}(z_{ref},t) = \frac{1}{2}(E(t)e^{-i\omega_o t} + E^*(t)e^{i\omega_o t})\vec{e}_x$$

where $$E(t) = A\, e^{-(\alpha+i\beta)t^2} \qquad \text{(II.1.2)}$$

and $$\omega_o = \omega_t + \Delta\omega.$$

Here the pulse shape is characterized by four real parameters, A, $\alpha > 0$, $\Delta\omega$, and β, which depend on z_{ref}; ω_t is the laser "transition" frequency, shown in equation I.2.40, and will be discussed

further in the next section. We will ignore any constant phase in E, as implied by A being real; furthermore, the time origin at any z_{ref} is taken to be the peak of the field envelope, so that a time shift will be used to eliminate any terms of the type exp[c't], with c' real. This involves no loss of information, since we are not interested in the loop transit time in these calculations.

The form of the field, as given in equation II.1.2, contains a parameter, β, which will be found to be crucial to the pulse formation process. This parameter determines the "chirp" on the pulse; a pulse with a positive chirp has an instantaneous frequency which increases with time, as illustrated in Figure 4. By defining the generalized amplitude, $\mathcal{A}(t)$, and phase, $\phi(t)$, functions according to

$$\vec{\mathcal{E}}(z,t) = \frac{1}{2}(E_+(z,t) + E_+^*(z,t))\vec{e}_x \qquad \text{(II.1.3a)}$$

$$E_+(z,t) = E(z,t)e^{-i\omega_o t} = \mathcal{A}(z,t)e^{i\phi(z,t)} \quad ,$$

then $\omega(z,t)$, the instantaneous frequency, will be given by

$$\omega(z,t) = -\frac{\partial\phi}{\partial t} \quad , \qquad \text{(II.1.3b)}$$

where we use the negative sign so that the "carrier" frequency is $+\omega_o$, for convenience. For the standard traveling wave field form given in equation I.1.2, we have, for this plane wave case,

$$E_+(z,t) = E_x(z,t)e^{i(k_o z-\omega_o t)} \qquad \text{(II.1.4)}$$

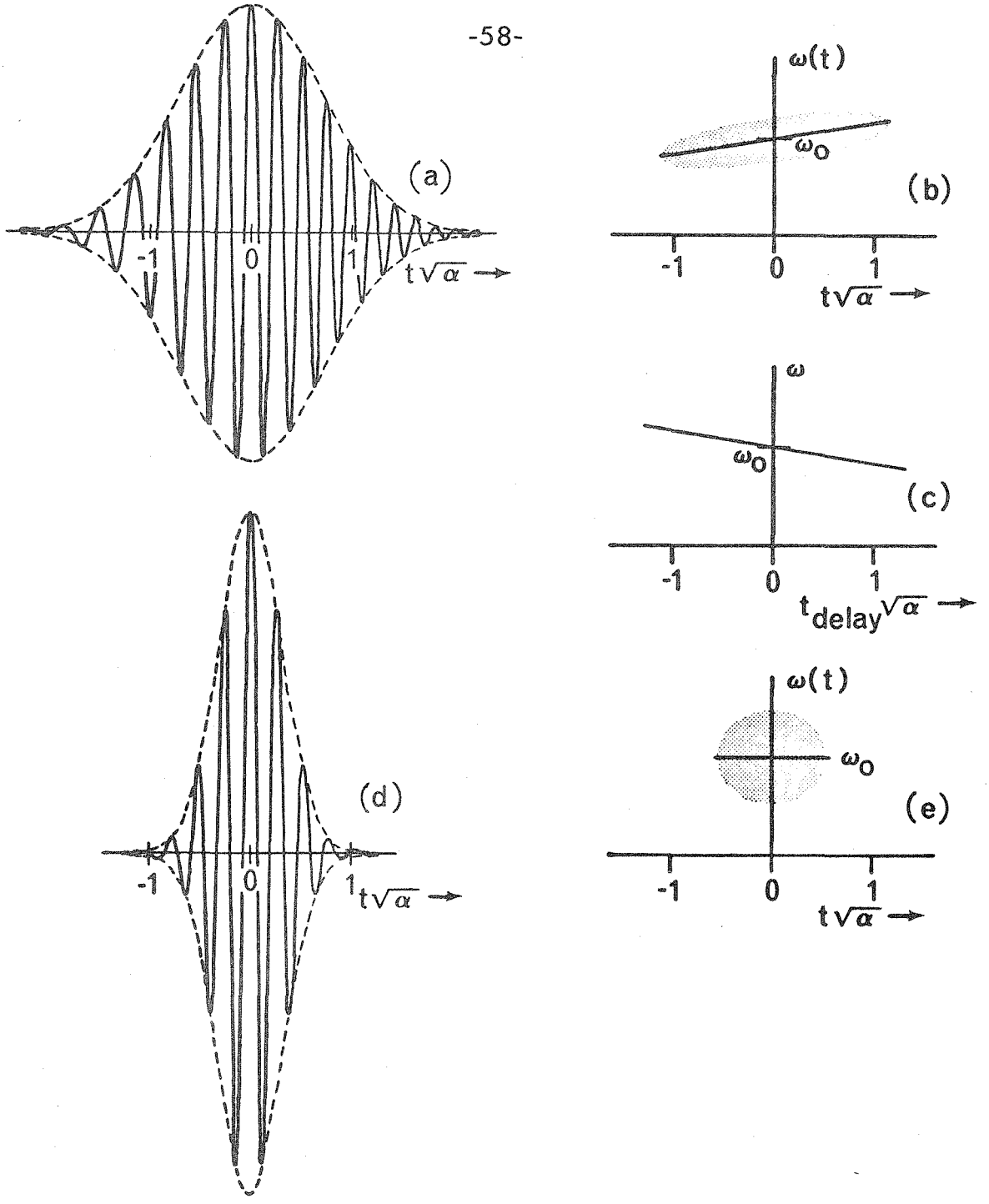


Figure 4. Linearly chirped pulse with Gaussian envelope.
(a),(b) Pulse field and frequency; the shading represents the frequency spread due to envelope modulation. The time scale can be transformed into a length scale, giving an instantaneous "snapshot" of a pulse moving to the left.
(c) Compressor group delay vs. frequency.
(d),(e) Compressed pulse characteristics. The frequency spread is now due entirely to the sharpened envelope, since the chirp has been removed.

with an analogous form without any 'z' dependence for the "reference plane" field of equation II.1.2. We will use the following convention for the (temporal) Fourier transform, $F(\omega)$, of a function, $f(t)$:

$$F(\omega) \equiv (2\pi)^{-1/2}\int f(t')e^{i\omega t'}dt' \qquad \text{(II.1.5a)}$$

which implies the inverse transform as given by:

$$f(t) = (2\pi)^{-1/2}\int F(\omega')e^{-i\omega' t}d\omega' \quad . \qquad \text{(II.1.5b)}$$

With these definitions, we will consider the Fourier transform of $E_+(z,t)$ alone as defining the pulse spectrum at 'z', according to

$$E_+(z,\omega) \equiv (2\pi)^{-1/2}\int E_+(z,t')e^{i\omega t'}dt', \qquad \text{(II.1.6a)}$$

with the corresponding expansion of $E_+(z,t)$ in terms of its frequency components:

$$E_+(z,t) = (2\pi)^{-1/2}\int E_+(z,\omega')e^{-i\omega' t}d\omega'. \qquad \text{(II.1.6b)}$$

The sign of $'\omega'$ in the Fourier transform is thus seen to be consistent with the interpretation that $E_+(z,t)$ contains the positive frequency components of $\mathcal{E}(z,t)$, with frequencies centered at ω_o.

The amplitude and phase definitions in equation II.1.3a are, in fact, not unique, but the form shown in equation II.1.2 is well defined if we choose

$$\begin{aligned} \mathcal{A}(z_{ref},t) &= A\exp[-\alpha t^2] \\ \phi(z_{ref},t) &= \omega_o t + \beta t^2 \end{aligned} \qquad \text{(II.1.7)}$$

giving $\omega(z_{ref},t) = \omega_o + 2\beta t$. Our model of the laser pulse thus includes, at most, a "linear" chirp, with the chirp rate given by "2β". The pulse given in equation II.1.2 has an r.m.s. intensity pulsewidth, τ_p, and an r.m.s. power spectrum bandwidth, $\delta\omega$, given by

$$\tau_p = \frac{1}{2\sqrt{\alpha}}$$

and (II.1.8)

$$\delta\omega = \sqrt{\frac{\alpha^2+\beta^2}{\alpha}} \quad .$$

Thus, even in the absence of any chirp, the envelope modulation causes a frequency uncertainty of $(\delta\omega)_{pulse} = \sqrt{\alpha}$, so that the "frequency sweep", as given by $(\delta\omega)_{chirp} = \tau_p(2\beta) = \frac{\beta}{\sqrt{\alpha}}$, is only a measurable quantity when $\beta \gtrsim \alpha$. Even for $\beta < \alpha$, however, we will refer to the "instantaneous frequency", $\omega(t)$, given by equation II.1.3b, as if it were well-defined.

I.1.ii laser cavity elements

The effect of each cavity element on the circulating pulse can be described in terms of changes in the four parameters, A, α, $\Delta\omega$, β, since we assume that the field always retains the Gaussian form in equation II.1.2. In this section, models will be developed for the gain and Kerr media which meet this requirement, but which nevertheless represent the major physical processes involved. More detailed analyses of plane-wave pulse propagation in such media can be performed[10,13]; however, the refinements indicated in the higher order solutions do not appear to significantly affect the mode-locking model given here.

First, we will discuss the effect of the gain medium on the circulating pulse. From our previous considerations, we know that the primary "mode-coupling" process occurs in the Kerr medium, so that, as a first approximation, we can treat only the linear aspects of the amplification process. The resonant polarizability was discussed in section I.2.iv; from equation I.2.40a, then, we can derive the following result for the "active" polarization in the gain medium, when a monochromatic beam is present:

$$P_{res} = \alpha_{res}(\omega)E \ , \qquad \text{(II.1.9a)}$$

where

$$\alpha_{res}(\omega) = (N_t - N_g)\frac{e^2}{\hbar}\left(\frac{\varepsilon+2}{3}\right) \langle r^2\rangle \left(\frac{(\Delta\omega - i\Gamma)}{(\Delta\omega)^2+\Gamma^2}\right) \qquad \text{(II.1.9b)}$$

$$\langle r^2\rangle = \begin{cases} |(x)_{gt}|^2 & \text{in crystals} \\ \frac{1}{3}\left(|(x)_{gt}|^2 + |(y)_{gt}|^2 + |(z)_{gt}|^2\right) & \text{in liquids or glasses} \end{cases}$$

$$\Delta\omega = \omega - \omega_t$$

$N_t(N_g)$ = number density of molecules originally in upper (lower) transition level .

Here "x" is the polarization direction for both $\vec{\mathcal{P}}$ and $\vec{\mathcal{E}}$; for crystal amplifiers, we thus assume cubic symmetry or else alignment of a principal dielectric axis with the polarization. Also, ε, the dielectric constant, includes only non-resonant molecules and transitions; and local fields effects, due to neighboring resonant

molecules, are neglected. The quantity, $N_t - N_g$, in equation II.1.9b is, of course, the "inversion" density, which is the source of amplification.

If inhomogenous broadening is present, such that a distribution of resonant frequencies, $g'(\omega_t)$, is present, centered on $\bar{\omega}_t$, we can obtain the net polarizability according to the averaging procedure

$$\alpha_{res}(\omega) = (N_t - N_g)\frac{e^2}{\hbar}\left(\frac{\varepsilon+2}{3}\right)\langle r^2\rangle \int \frac{(\omega-\omega_t) - i\Gamma}{(\omega-\omega_t)^2 + \Gamma^2} g'(\omega_t)d\omega_t \Big/ \int g'(\omega_t)d\omega_t \quad .$$

As an example, a normalized Lorentzian distribution for the inhomogeneous broadening would have the form

$$g'(\omega_t) = \frac{\Gamma_1/\pi}{(\omega_t - \bar{\omega}_t)^2 + (\Gamma_1)^2} \qquad \text{(II.1.10)}$$

from which we obtain the same formula for $\alpha_{res}(\omega)$ as in equation II.1.9, with $\omega_t \to \bar{\omega}_t$ and $\Gamma \to \Gamma + \Gamma_1$; we will therefore assume that the polarizability given above holds whether or not there is inhomogeneous broadening.

Equation II.1.9 was derived for a monochromatic source, which thus cannot have a Gaussian envelope, but this presents no difficulty in finding the linear polarization since Fourier transforms can be used. Thus, in equation II.1.9a, we can consider the field to represent a single Fourier component at ω. Since the resonant polarizability was not included in defining ε (or n_o), we can consider it to be a "nonlinear source" and substitute it for P_{NL} in the wave equation; according to the plane-wave version of equation I.3.8, the Fourier component at ω propagates according to

$$\frac{\partial E(z,\omega)}{\partial z} = i\,\frac{2\pi\,\omega}{n_o c}\,\alpha_{res}(\omega)E(z,\omega)\ , \qquad \text{(II.1.11)}$$

from which we obtain

$$E(z,\omega) = E(0,\omega)\exp[i\,\frac{2\pi\,\omega}{n_o c}\,\alpha_{res}(\omega)z]\ , \qquad \text{(II.1.12a)}$$

and the corresponding equation for the net field component:

$$\vec{\mathcal{E}}(z,\omega) = \frac{1}{2}(E(0,\omega)e^{i(k'z-\omega t)} + c.c.)\vec{e}_x$$

where

$$k'(\omega) = k(1 + \frac{2\pi}{n_o^2}\,\alpha_{res}(\omega))\ . \qquad \text{(II.1.12b)}$$

Finally, then, in terms of the Fourier components defined in equation II.1.6, we have a "transfer function" for the gain medium, as given by

$$E_+(z,\omega) = E_+(0,\omega)e^{ik'(\omega)z}\ . \qquad \text{(II.1.13)}$$

Since $\alpha_{res}(\omega)$ is complex, this leads to the well-known exponential amplification (or absorption) law for propagation in a laser medium. However, in order to keep the analytical form shown in equation II.1.2 for the field at any z, we can include only linear and quadratic terms (in ω) in the exponent in equation II.1.13. Defining $g(\ell,\omega)$ as the transfer function for a propagation length, ℓ, in the gain medium, according to

$$E_+(\ell,\omega) = g(\ell,\omega)E_+(0,\omega)\ , \qquad \text{(II.1.14)}$$

we find from the above

$$g(\ell,\omega) = e^{ik'(\omega)\ell}$$

$$= e^{i\frac{n_o\omega_t\ell}{c}} \; e^{i\frac{n_o}{c}(\omega-\omega_t)\ell} \; e^{a_g\ell\left(\frac{\Gamma^2+i\Gamma(\omega-\omega_t)}{(\omega-\omega_t)^2+\Gamma^2}\right)} \quad , \qquad \text{(II.1.15a)}$$

which is approximated for our purposes by

$$g(\ell,\omega) = e^{i\frac{n_o\omega_t}{c}\ell} \; e^{i\frac{n_o}{c}(\omega-\omega_t)\ell} \; e^{a_g\ell\left(1+i\frac{(\omega-\omega_t)}{\Gamma}-\frac{(\omega-\omega_t)^2}{\Gamma^2}\right)} \qquad \text{(II.1.15b)}$$

We have defined the gain constant

$$a_g = \frac{2\pi\omega_t e^2 \langle r^2\rangle}{\hbar c\Gamma}\left(\frac{n_o^2+2}{3n_o}\right)(N_t-N_g) \quad , \qquad \text{(II.1.15c)}$$

with, of course, n_o representing the gain medium refractive index in these equations; we ignore the frequency dependence of a_g (and n_o).

For frequencies outside the gain linewidth, $|\omega-\omega_t| > \Gamma$, the true transfer function in equation II.1.15a indicates that the gain medium is transparent, while our approximate form in equation II.1.15b leads to absorption. We will later see, though, that the presence of losses in the cavity makes the latter form an acceptable approximation.

We will now consider the distortions of the pulse produced by its propagation through the Kerr medium. The field obeys the plane-wave versions of equation I.3.13:

$$\vec{\mathcal{E}}(z,t) = \frac{1}{2}(E(z,t)e^{i(k_oz-\omega_ot)} + \text{c.c.})\vec{e}_x$$

$$\frac{\partial E}{\partial z} + \frac{(n_o+(\delta n))}{c}\frac{\partial E}{\partial t} = \frac{i\omega_o}{c}(\delta n)E \qquad \text{(II.1.16)}$$

$$\tau \frac{\partial(\delta n)}{\partial t} = \frac{n_2}{2} |E|^2 - (\delta n) \quad ,$$

where n_o now represents the Kerr medium refractive index. If we define the intensity, $\mathcal{I} = |E|^2$, then, since (δn) is real, we can see that $\mathcal{I}$ travels with a velocity $v = \frac{c}{n_o+(\delta n)}$, which depends on z and t through $\delta n(z,t)$. Thus, the "intensity" delay time for a given "point", p, on the pulse envelope, after propagating a distance, ℓ, will be given by

$$(\tau_{delay})_p = \frac{1}{c}[n_o \ell + \int_o^{\ell} \delta n(z'_p, t(z'_p))dz'] = t_o + (\delta t)_p \quad .$$

Here $t(z'_p)$ is the implicitly defined time at which the given "point" passes z'_p. Clearly, if $(\delta t)_p << \tau_{pulse}$, for every point, p, then the intensity envelope will not be distorted by the Kerr effect; this holds if

$$\ell < \frac{c\tau_{pulse}}{(\delta n)_{max}} \lesssim \frac{c\tau_{pulse}}{\frac{n_2}{2}\mathcal{I}_{max}} \equiv \ell_o \quad , \qquad \text{(II.1.17)}$$

where ℓ_o is the "envelope distortion distance". Our estimates in section I.3 for a "strong" ultrashort pulse in a "strong" Kerr material give: $(\tau_{pulse} \gtrsim 10^{-12}$ sec, $(\delta n)_{max} \lesssim 10^{-4}) \ \ell_o \gtrsim 300$ cm. This shows that we can take the intensity envelope to be undistorted for our purposes, which shows that $(n_o+(\delta n)) \to n_o$ in equation II.1.16, and

$$E(z,t)E^*(z,t) = E(0,t - \frac{n_o}{c} z)E^*(0,t - \frac{n_o}{c} z) \quad . \qquad \text{(II.1.18a)}$$

This last result can be inserted into the relaxation equation for

$\delta n(z,t)$ to show that this quantity also propagates with velocity $\frac{c}{n_o}$:

$$\delta n(z,t) = \delta n(0,t - \frac{n_o}{c} z) \quad . \qquad \text{(II.1.18b)}$$

Finally, we have an analytic solution for the field; from equation II.1.16, we find

$$E(z,t) = E(0,t - \frac{n_o}{c} z)\exp[i \frac{\omega_o}{c} \delta n(0,t - \frac{n_o}{c} z)z] \qquad \text{(II.1.19a)}$$

with

$$\delta n(0,t') = \frac{n_2}{2\tau} \int_{-\infty}^{t'} E(0,t")E^*(0,t")e^{-(t'-t")/\tau} dt" \quad . \qquad \text{(II.1.19b)}$$

We have assumed that the Kerr medium relaxes completely between round trips of the circulating pulse, so that $\delta n(z,t \to -\infty) = 0$.

A physical interpretation of these results is illustrated in Figure 5. We can see from the figure that the "instantaneous" (freespace) wavelength of the output pulse is changed according to

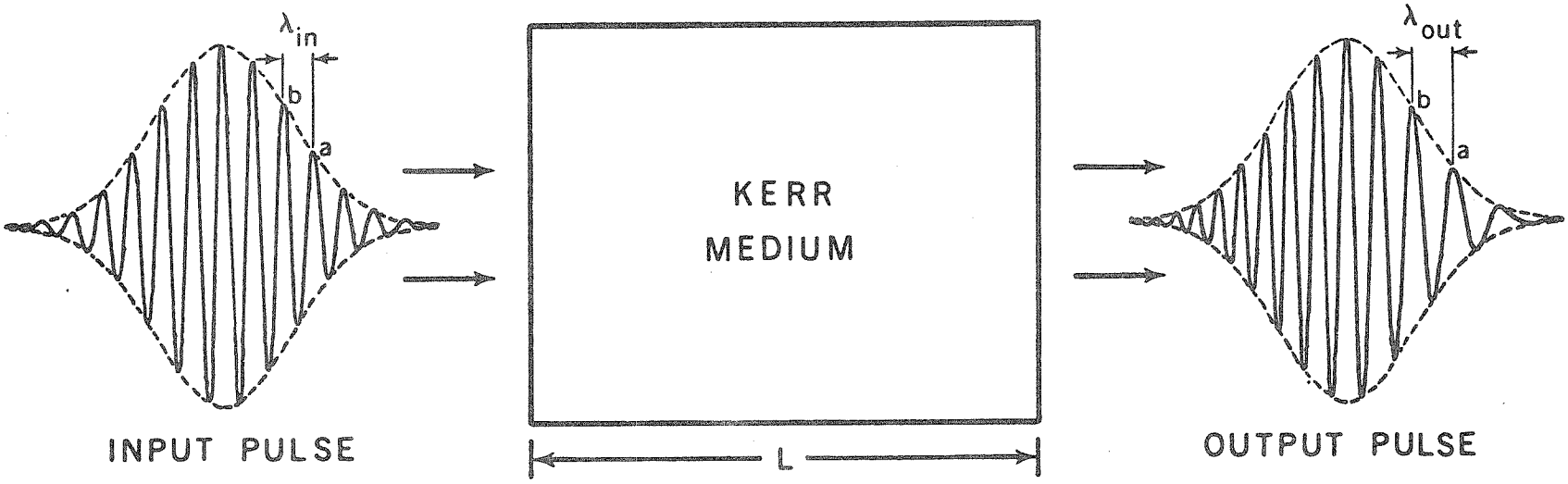
$$\lambda_{out}(t) \simeq \lambda_{in}(t) + \ell \frac{\lambda_{in}(t)}{c} \frac{\partial(\delta n)}{\partial t} = \lambda_{in}(t)[1 + \frac{\ell}{c} \frac{\partial(\delta n)}{\partial t}] \quad ,$$

which implies the following change in the "instantaneous frequency":

$$\omega_{out}(t) = \frac{2\pi c}{\lambda_{out}(t)} \simeq \omega_{in}(t)[1 - \frac{\ell}{c} \frac{\partial(\delta n)}{\partial t}] = \omega_{in}(t) + \delta\omega(t). \qquad \text{(II.1.20a)}$$

Using equation II.1.3, we obtain the following approximate result:

$$\begin{aligned} E_{out}(t) &= E_{in}(t)\exp[-i \int_{-\infty}^{t} \delta\omega(t')dt'] \\ &= E_{in}(t)\exp[i \frac{\omega_o}{c} \delta n(t)\ell] \quad , \end{aligned} \qquad \text{(II.1.20b)}$$



$$\lambda_{out} = \lambda_{in} + L(n_b - n_a)$$

$$\simeq \lambda_{in} + L \frac{d(\delta n)}{dt} \frac{\lambda_{in}}{c}$$

Figure 5. Origin of phase modulation (chirp) on a pulse passing through a medium with a time varying index of refraction, such as that due to an intensity-dependent refractive index ("Kerr medium"). An instantaneous "snapshot" of a pulse is shown before and after passing through a length, L, of the material; point "a" sees an index, n_a, in the Kerr medium, while point "b" sees an index, n_b. The formula for the wavelength change can be found by equating the propagation times for the two points, "a" and "b", between the two "snapshots".

which is equivalent to equation II.1.19a when the time origins are corrected for the pulse delay.

These results show that the Kerr medium has, to first order, produced a phase modulation on the propagating pulse, without affecting the envelope shape. This effect, called self phase modulation, broadens the frequency spectrum, and will be shown in the next section to lead to "mode-locking" in the laser system under discussion. For analytic results, however, we must retain the pulse shape given in equation II.1.2, which means that the index variations, $\delta n(t)$, must be approximated by a quadratic function in time.

For an instantaneous response in the Kerr medium, $\tau\sqrt{\alpha} << 1$, we have

$$\delta n(t) = \frac{n_2}{2} A^2 e^{-2\alpha t^2} \simeq \frac{n_2}{2} A^2(1 - 2\alpha t^2) \quad , \qquad \text{(II.1.21a)}$$

so that the "transfer function" for the Kerr medium is given by

$$E_+(\ell,t) = h(\ell,t)E_+(0,t) \qquad \text{(II.1.21b)}$$

where

$$h(\ell,t) = \exp[\frac{i\omega_o n_2}{2c} \ell A^2(1-2\alpha t^2)] \quad .$$

When $\alpha\tau^2 << 1$, a further expansion leads to the quadratic form:

$$h(\ell,t) \simeq \exp[i \frac{\omega_o n_2}{2c} \ell A^2\{(1-4\alpha\tau^2) + t(4\alpha\tau)-2\alpha t^2(1-12\alpha\tau^2)\}].$$

For our purposes, the instantaneous response case is sufficiently accurate[54], since it displays the physically significant fact that a "chirp" has been added to the pulse. The quadratic expansion is also physically reasonable, since the pulse intensity is very small for $\alpha t^2 \gtrsim 1$; thus, "nonlinear chirp" terms in the tails

will be of little significance.

II.1.iii pulse evolution equations

Equations II.1.15 and II.1.21 constitute simplified expressions for the effects of the gain and Kerr media on a propagating pulse. As shown in Figure 3, the circulating pulse in the laser cavity passes through the various elements in succession; we will assume that the pulse is short enough so that it does not overlap with itself in the Kerr medium while being reflected at M_2. Under these conditions, a "round trip" for a pulse, starting at the reference plane, will consist of six interactions, with the following elements in turn: gain cell, mirror M_1, gain cell, Kerr cell, mirror M_2, and Kerr cell. As shown in Figure 3, we define an initial field, $E_1(t)$, a field $E_2(t)$ before entering the Kerr medium, and a field $E_3(t)$ after the complete round trip.

These fields can be related by the transfer functions, $h(\ell,t)$, $g(\ell,\omega)$, and $R^{1/2}$, for the Kerr medium, gain medium, and mirrors, respectively. The following relations hold for these (we use "operator" algebra, where each operator acts on functions to its right):

$$\begin{aligned} h(\ell_1,t)h(\ell_2,t) &= h(\ell_1+\ell_2,t) \\ R^{1/2}h(R\ell,t) &= h(\ell,t)R^{1/2} \\ g(\ell_1,\omega)g(\ell_2,\omega) &= g(\ell_1+\ell_2,\omega) \\ R^{1/2}g(\ell,\omega) &= g(\ell,\omega)R^{1/2} \quad . \end{aligned} \tag{II.1.22}$$

These relations are not dependent on the "quadratic" forms; the only

restriction is that $\ell < \ell_0$ (i.e. no envelope distortion) must hold for the Kerr medium, in which case the second relation above follows directly from equation II.1.19.

Figure 6 shows how the true cavity configuration can be replaced with an "effective cavity" in which the losses are combined with the gain medium. From Figure 6a, and equation II.1.22, we can show the following:

$$\begin{aligned} E_2(\omega) &= g(L_G,\omega)R_1^{1/2}g(L_G,\omega)E_1(\omega) \\ &= R_1^{1/4}g(L_G,\omega)g(L_G,\omega)R_1^{1/4}E_1(\omega) \end{aligned} \qquad \text{(II.1.23a)}$$

and similarly,

$$\begin{aligned} E_3(t) &= h(L_K,t)R_2^{1/2}h(L_K,t)E_2(t) \\ &= R_2^{1/4}h(R_2^{1/2}L_K,t)h(R_2^{-1/2}L_K,t)R_2^{1/4}E_2(t) \\ &= R_2^{1/4}h(L_K(\frac{R_2^{1/2}+R_2^{-1/2}}{2}),t)h(L_K(\frac{R_2^{1/2}+R_2^{-1/2}}{2}),t)R_2^{1/4}E_2(t) \quad . \end{aligned} \qquad \text{(II.1.23b)}$$

These equations show that the true cavity is physically equivalent to the cavity shown in Figure 6b, in which the mirrors are perfectly reflecting and absorption cells (losses) have been included. Finally, by shifting the reference plane and combining losses, we see that the physical problem can be replaced by that of Figure 6c; it is this "effective cavity" which we will analyze.

The pulse evolution is now described by the net transfer functions, $\bar{g}(\omega)$ and $\bar{h}(t)$, where

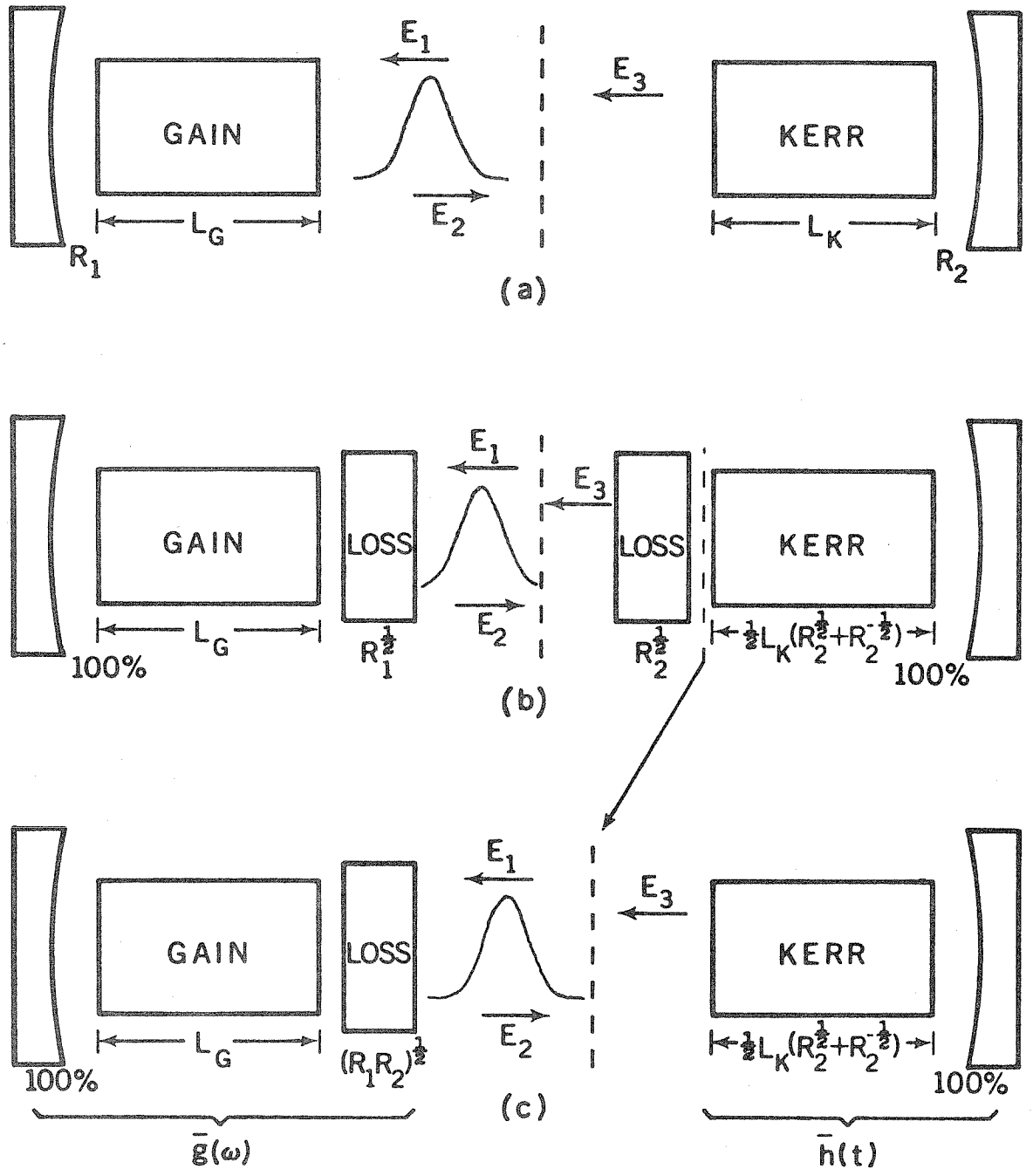


Figure 6. Derivation of effective cavity.
(a) Actual cavity diagram; the reference plane is dashed.
(b) (Mathematically) equivalent cavity, for pulses which are sampled at the reference plane.
(c) Effective cavity used in analysis, in which the mirror losses are combined with the gain. The reference plane has been moved from its position in (b).

$$E_2(\omega) = \bar{g}(\omega)E_1(\omega) \qquad \text{(II.1.24a)}$$

with

$$\bar{g}(\omega) = (R_1R_2)^{1/2}g(2L_G,\omega)$$

$$= (R_1R_2)^{1/2}\exp[2a_gL_G(1 - \frac{(\Delta\omega)^2}{\Gamma^2})] \qquad \text{(II.1.24b)}$$

and

$$E_3(t) = \bar{h}(t)E_2(t) \qquad \text{(II.1.25a)}$$

with

$$\bar{h}(t) = h(L_K(R_2^{1/2}+R_2^{-1/2}),t)$$

$$= \exp[i\,\frac{\omega_o n_2}{2c}\,L_K(R_2^{1/2}+ R_2^{-1/2})A^2(1-2\alpha t^2)] \quad \text{(II.1.25b)}$$

We are ignoring pulse delay times and constant phase factors; the "important" transfer functions can thus be written as follows:

$$\bar{g}(\omega) = G^{1/2}\exp[-\frac{(\Delta\omega)^2}{4\alpha_L}] \qquad \text{(II.1.26a)}$$

$$\bar{h}(t) = \exp[-iD_oA^2\alpha t^2] \qquad \text{(II.1.26b)}$$

where

$$G = R_1R_2e^{4a_gL_G} \qquad \text{(II.1.26c)}$$

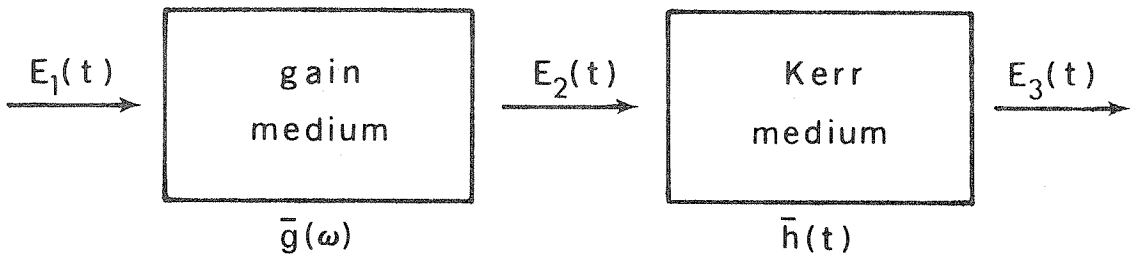
$$\alpha_L = \frac{\Gamma^2}{8\,a_gL_G} \qquad \text{(II.1.26d)}$$

$$D_o = \frac{\omega_t n_2}{c}\,L_K(R_2^{1/2}+ R_2^{-1/2}) \quad . \qquad \text{(II.1.26e)}$$

Here $\Delta\omega = \omega-\omega_t$. Note that G is the line-center round trip power gain in the absence of the Kerr medium, and $\tau_{ML} \equiv (4\alpha_L)^{-1/2}$ is the pulsewidth obtained if there is perfect mode-locking over the entire gain profile, $\bar{g}(\omega)$.

It was noted, in connection with equation II.1.15, that the true gain function has a Lorentzian in the exponent, while we used,

in deriving the Gaussian form in equation II.1.26a, the quadratic approximation. However, the important transfer function is $\bar{g}(\omega)$, which also includes the factor $(R_1R_2)^{1/2}$, as shown in the definition of G. Since $R_1R_2 < 1$. the cavity will have losses for frequencies in the tails of the Lorentzian; Figure 7 shows that the Gaussian approximation used in $\bar{g}(\omega)$ will be physically acceptable in that case.



The pulse, according to the figure, begins as $E_1(t)$ before passing through the gain medium, and has the Gaussian form:

$$E_1(t) = A_1\exp[-\alpha_1 t^2 - i((\omega_t+\Delta\omega_1)t + \beta_1 t^2)] \quad . \quad \text{(II.1.27a)}$$

Its Fourier transform, given by equation II.1.6a, is

$$E_1(\omega) = A_1[2(\alpha_1+i\beta_1)]^{-1/2}\exp[-(\omega-\omega_1)^2/4(\alpha_1+i\beta_1)] \ , \ \text{(II.1.27b)}$$

where, as usual, $\omega_1 = \omega_t+\Delta\omega_1$. Using equations II.1.24a and II.1.26a, the pulse spectrum after the gain medium is given by

$$E_2(\omega) = A_1G^{1/2}[2(\alpha_1+i\beta_1)]^{-1/2}\exp[-(\omega-\omega_1)^2/4(\alpha_1+i\beta_1)-(\omega-\omega_t)^2/4\alpha_L] \ .$$

Therefore, neglecting a constant phase shift and measuring time from

$$\text{(a):}\quad g(\omega) = (R_1R_2)^{\frac{1}{2}} e^{\frac{2a_gL_G}{1+(\omega-\omega_t)^2/\Gamma^2}}$$

$$\text{(b):}\quad g(\omega) = (R_1R_2)^{\frac{1}{2}} e^{2a_gL_G\left(1-\frac{(\omega-\omega_t)^2}{\Gamma^2}\right)}$$

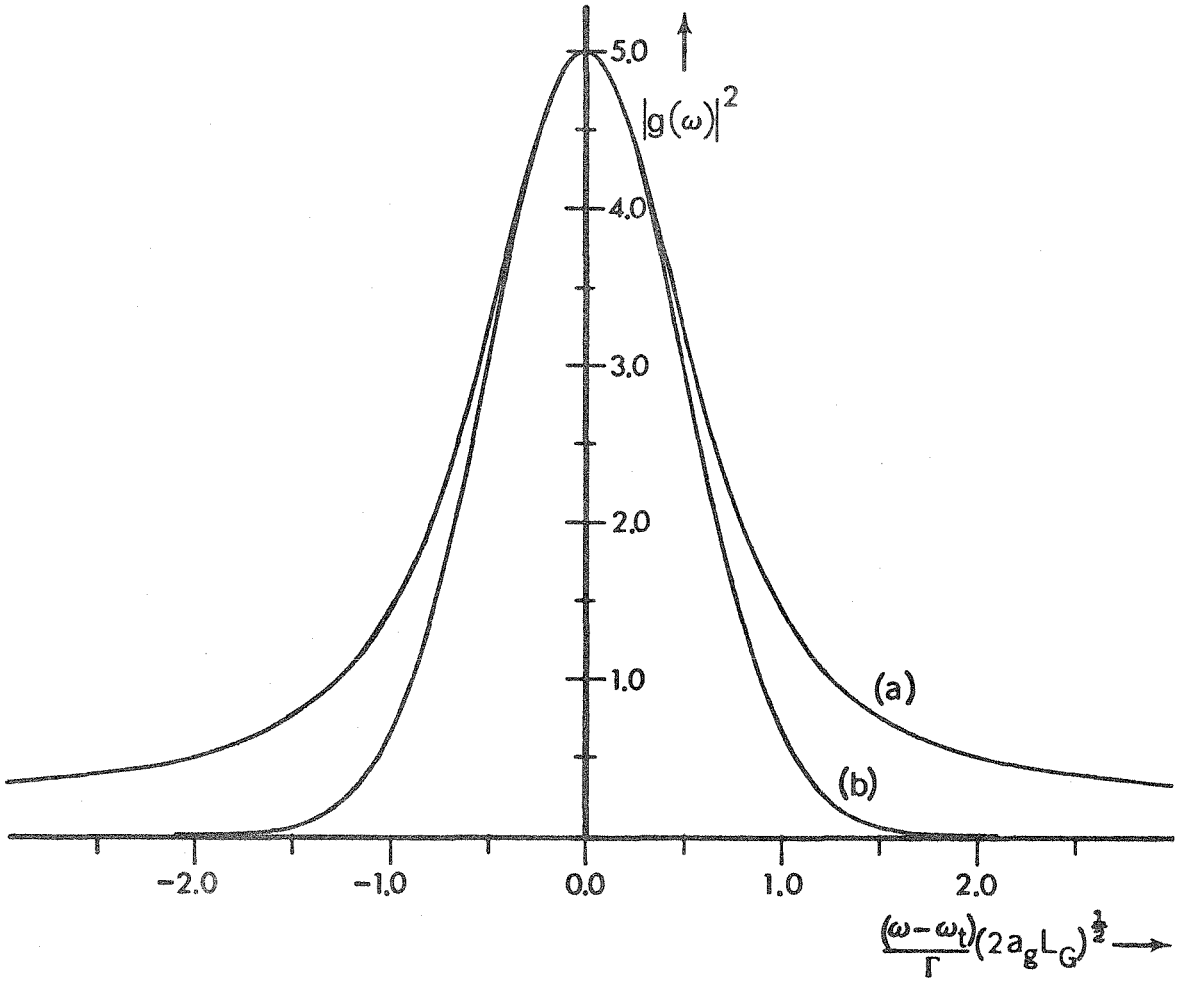


Figure 7. Comparison of power gain lineshapes, showing accuracy of Gaussian approximation for high loss case. (a) Graph of true gain function, based on a Lorentzian frequency response in the amplifying medium. (b) Gaussian approximation used in the calculations. Both curves are based on a "high gain" case, with $G = |g(\omega_t)|^2 = 5.0$, and $R_1R_2 = 0.2$; in each case, frequencies in the "tails" see a net loss, so that they do not play an important role in the laser operation.

the peak of the envelope, we obtain

$$E_2(t) = A_2 \exp[-\alpha_2 t^2 - i((\omega_t+\Delta\omega_2)t + \beta_2 t^2)] \qquad \text{(II.1.28a)}$$

where

$$A_2 = \frac{A_1 G^{1/2}}{\left[\left(1+\frac{\alpha_1}{\alpha_L}\right)^2 + \left(\frac{\beta_1}{\alpha_L}\right)^2\right]^{1/4}} \exp\left[\frac{-(\Delta\omega_1)^2}{4\alpha_L\left(1+\frac{\alpha_1}{\alpha_L}+\frac{\beta_1^2}{\alpha_1\alpha_L}\right)}\right]$$

$$\alpha_2 = \alpha_1 \frac{\left[1+\frac{\alpha_1}{\alpha_L}+\frac{\beta_1^2}{\alpha_1\alpha_L}\right]}{\left[\left(1+\frac{\alpha_1}{\alpha_L}\right)^2 + \left(\frac{\beta_1}{\alpha_L}\right)^2\right]}$$

$$\Delta\omega_2 = \frac{\Delta\omega_1}{\left[1+\frac{\alpha_1}{\alpha_L}+\frac{\beta_1^2}{\alpha_1\alpha_L}\right]} \qquad \text{(II.1.28b)}$$

$$\beta_2 = \frac{\beta_1}{\left[\left(1+\frac{\alpha_1}{\alpha_L}\right)^2 + \left(\frac{\beta_1}{\alpha_L}\right)^2\right]} \quad .$$

After the pulse passes through the Kerr medium, equations II.1.25a and II.1.26b show that the pulse shape is given by:

$$E_3(t) = E_2(t)\exp[-iD_o A_2^2 \alpha_2 t^2] \quad .$$

This affects only the chirp parameter, β. Therefore, the following relations show how the pulse parameters have changed during a complete round trip:

$$E_3(t) = A_3 \exp[-\alpha_3 t^2 - i((\omega_t+\Delta\omega_3)t + \beta_3 t^2)] \qquad \text{(II.1.29a)}$$

where

$$A_3 = \frac{A_1 G^{1/2}}{[(1+\frac{\alpha_1}{\alpha_L})^2 + (\frac{\beta_1}{\beta_L})^2]^{1/4}} \exp\left[\frac{-(\Delta\omega_1)^2}{4\alpha_L(1+\frac{\alpha_1}{L}+\frac{\beta_1^2}{\alpha_1\alpha_L})}\right]$$

$$\alpha_3 = \alpha_1 \frac{[1+\frac{\alpha_1}{\alpha_2}+\frac{\beta_1^2}{\alpha_1\alpha_L}]}{[(1+\frac{\alpha_1}{\alpha_L})^2+(\frac{\beta_1}{\beta_L})^2]}$$

$$\Delta\omega_3 = \frac{\Delta\omega_1}{[1+\frac{\alpha_1}{\alpha_L}+\frac{\beta_1^2}{\alpha_1\alpha_L}]} \qquad \text{(II.1.29b)}$$

$$\beta_3 = \frac{\beta_1}{[(1+\frac{\alpha_1}{\alpha_L})^2+(\frac{\beta_1}{\alpha_L})^2]} + \frac{D_o A_1^2 G\alpha_1[1+\frac{\alpha_1}{\alpha_L}+\frac{\beta_1^2}{\alpha_1\alpha_L}]}{[(1+\frac{\alpha_1}{\alpha_L})^2+(\frac{\beta_1}{\alpha_L})^2]^{3/2}} \exp\left[\frac{-(\Delta\omega_1)^2}{2\alpha_L(1+\frac{\alpha_1}{\alpha_L}+\frac{\beta_1^2}{\alpha_1\alpha_L})}\right].$$

These equations will describe the pulse-to-pulse changes in a train of ultrashort pulses produced by the laser, since we can consider $(A_1,\alpha_1,\Delta\omega_1,\beta_1)$ to describe the n^{th} pulse, and $(A_3,\alpha_3,\Delta\omega_3,\beta_3)$ to correspond to the $(n+1)^{th}$ pulse. Note that we are considering the laser output to be sampled at the reference plane in Figure 6c; however, the pulse shapes at the mirror locations or at the reference plane in Figure 6a can be found, if needed, by applying the proper transfer functions to the fields above.

II.2 Continuous, Mode-Locked Laser Operation

The evolution of a pulse circulating in the laser system is described by equations II.1.27, II.1.28, and II.1.29. These depend on three parameters, G, α_L, and D_0, describing the cavity, as given in equation II.1.26; in a laser operated continuously, these will be constants. We can now solve for the "steady-state" pulse shape which such a laser would produce, by requiring that the pulse reproduce itself after one round trip. This implies that

$$A_3 = A_1, \quad \alpha_3 = \alpha_1, \quad \Delta\omega_3 = \Delta\omega_1, \quad \beta_3 = \beta_1, \qquad \text{(II.2.1)}$$

which, using equation II.1.29, yields solutions determining uniquely the intensity, duration, center frequency, and chirp of the self-reproducing pulse. The resulting "steady-state" parameters, A°, α°, $\Delta\omega^\circ$, and β°, for the pulse at E_1, are given by

$$\begin{aligned} (A^\circ)^2 &= \frac{1}{|D_0|}\,\frac{G^2-1}{G} \\ \alpha^\circ &= \alpha_L\,\frac{G^2-1}{G^2+1} \\ \Delta\omega^\circ &= 0 \\ \beta^\circ &= \alpha_L G\,\frac{G^2-1}{G^2+1}\ . \end{aligned} \qquad \text{(II.2.2)}$$

If $G \leq 1$ there is no steady state solution, and if $D_0 < 0$, β° will change sign. Similarly, the pulse form at E_2, under these steady-state conditions, is described by

$$(A_2^\circ)^2 = \frac{1}{|D_0|}\,\frac{G^2 - 1}{G} = (A^\circ)^2$$

$$\alpha_2^\circ = \alpha_L\,\frac{G^2 - 1}{G^2 + 1} = \alpha^\circ$$

$$\Delta\omega_2^\circ = 0 \qquad = \Delta\omega^\circ \qquad \text{(II.2.3)}$$

$$\beta_2^\circ = \frac{\alpha_L}{G}\,\frac{G^2 - 1}{G^2 + 1} = \frac{\beta^\circ}{G^2} \quad .$$

When the excess line-center gain is small, $G \lesssim 1$, these results are similar to those for active FM mode-locking[53], since the chirp parameter, β°, is equal to the bandwidth parameter, α°, and the pulsewidth, $\tau_p = (4\alpha^\circ)^{-1/2}$ is longer than that of complete mode-locking, $\tau_{ML} = (4\alpha_L)^{-1/2}$. This small gain case also involves very little pulse distortion during the round trip transit, as seen by comparing the fields E_1 and E_2 (equations II.2.2 and II.2.3, respectively); therefore, it most closely corresponds to the results of the "mode-coupling" analysis[20] discussed at the beginning of section II.1.i.

II.2.i stable pulse generation in the high gain limit

We will now examine the laser behavior in the limit of a high excess line-center gain, $G >> 1$; we will find that such a system has some highly desirable characteristics, which will be the primary consideration of the remainder of this chapter. In the limit $G >> 1$, equation II.2.2 shows that the steady-state pulse leaving the Kerr medium is described by:

$$(A^\circ)^2 \simeq \frac{G}{D_o}$$

$$\alpha^\circ \simeq \alpha_L \qquad \text{(II.2.4)}$$

$$\Delta\omega^\circ \simeq 0$$

$$\beta^\circ \simeq \alpha_L G$$

and similarly, equation II.2.3 indicates that this pulse, after passing through the gain medium, differs only in its chirp, which is now $\beta_2^\circ \simeq \alpha_L/G$. These formulas have a relative accuracy of $1/G^2$, which shows that the "high gain limit" holds for $G \gtrsim 4$. Also, in this high gain limit, we have

$$\tau_p \simeq (4\alpha_L)^{-1/2} = \tau_{ML} = \left(\frac{2a_g L_G}{\Gamma^2}\right)^{1/2} . \qquad \text{(II.2.5)}$$

Since $G = R_1 R_2 e^{4a_g L_G}$, we can see that "$2a_g L_G$" is on the order of unity for realistic systems, so that the limiting pulsewidth is approximately equal to the inverse of the laser gain linewidth, Γ. The presence of the Kerr medium inside the laser cavity in combination with the large, profiled, gain is therefore seen to lead to ultrashort pulses with a duration characteristic of "fully mode-locked" pulses.

We can estimate the physical quantities involved, by considering the steady-state value of $(A^\circ)^2$; this shows that

$$\frac{L_K}{G} = \frac{\lambda}{4\pi(R_2^{1/2} + R_2^{-1/2})} \frac{n_o}{\frac{n_2}{2}|E_{peak}|^2} \simeq \frac{\lambda}{8\pi} \frac{n_o}{(\delta n)_{peak}} .$$

Since we estimated earlier that powerful ultrashort pulses could have

$\frac{\delta n}{n_o} \lesssim 10^{-4}$, we obtain, for $\lambda = 1\mu = 10^{-4}$cm, $L_K \gtrsim \frac{G}{8\pi} \gtrsim 0.3$ cm. We will later consider a specific laser system in more detail, but this shows that $\sim$ centimeter sized Kerr cells would usually be sufficient, even if the cubic nonlinearity were caused by the (weak) electronic mechanism.

In general, the pulse evolution is given by equation II.1.29; under constant G, α_L, D_o, we can show that any initial pulse will evolve into the steady-state form, and that this steady-state shape is particularly stable in the high gain limit. This stability can be shown by the following arguments. Assume that the pulse shape is close to the "steady-state" shape, and define the following relative deviation parameters, which will thus be small:

$$(\delta\hat{A}^2)_n = \frac{(A^2 - (A^\circ)^2)}{(A^\circ)^2}$$

$$(\delta\hat{\alpha})_n = \frac{(\alpha-\alpha^\circ)}{\alpha^\circ} \qquad \text{(II.2.6)}$$

$$(\delta\hat{\omega})_n = \frac{\Delta\omega}{(\alpha_L)^{1/2}}$$

$$(\delta\hat{\beta})_n = \frac{(\beta-\beta^\circ)}{\beta^\circ} \quad .$$

Here, the values of A, α, $\Delta\omega$, and β are those characterizing the field E_1, at the end of the n^{th} round trip.

By taking partial derivatives of equation II.1.29, evaluated for A°, α°, $\Delta\omega^\circ$, and β°, the following matrix can be found for the high gain case, where $\varepsilon \equiv 1/G^2 << 1$:

$$\begin{pmatrix} (\hat{\delta A^2})_{n+1} \\ (\hat{\delta\alpha})_{n+1} \\ (\hat{\delta\omega})_{n+1} \\ (\hat{\delta\beta})_{n+1} \end{pmatrix} \approx \begin{pmatrix} 1 & -2\varepsilon & -\frac{1}{2}\varepsilon & -1+4\varepsilon \\ 0 & -\varepsilon & 0 & 4\varepsilon \\ 0 & 0 & \varepsilon & 0 \\ 1-\varepsilon & -3\varepsilon & -\varepsilon & -1+8\varepsilon \end{pmatrix} \begin{pmatrix} (\hat{\delta A^2})_{n} \\ (\hat{\delta\alpha})_{n} \\ (\hat{\delta\omega})_{n} \\ (\hat{\delta\beta})_{n} \end{pmatrix} \quad \text{(II.2.7)}$$

The matrix elements have been kept accurate to order ε.

If the pulse parameters are thus not equal to the steady-state values at the n^{th} transit (i.e. if $(\hat{\delta A^2})_n$, $(\hat{\delta\alpha})_n$, $(\hat{\delta\omega})_n$, and $(\hat{\delta\beta})_n$ are not all zero), then the above matrix shows what the pulse will look like after the $(n+1)^{th}$ transit. The values of ± 1 in the matrix are the crucial point, since they lead to a strong stabilizing tendency; their origin is made clear by the physical picture described below, and is not dependent on the simple cavity model used in our analyses. This stability can be seen from equation II.2.7 by considering the pulse after two transits:

$$\begin{pmatrix} (\hat{\delta A^2})_{n+2} \\ (\hat{\delta\alpha})_{n+2} \\ (\hat{\delta\omega})_{n+2} \\ (\hat{\delta\beta})_{n+2} \end{pmatrix} \approx \begin{pmatrix} 5\varepsilon & \varepsilon & \frac{1}{2}\varepsilon & -8\varepsilon \\ 4\varepsilon & 0 & 0 & -4\varepsilon \\ 0 & 0 & 0 & 0 \\ 8\varepsilon & \varepsilon & \frac{1}{2}\varepsilon & -11\varepsilon \end{pmatrix} \begin{pmatrix} (\hat{\delta A^2})_{n} \\ (\hat{\delta\alpha})_{n} \\ (\hat{\delta\omega})_{n} \\ (\hat{\delta\beta})_{n} \end{pmatrix} \quad \text{(II.2.8)}$$

This shows that all deviations from the steady-state condition are reduced in two round trips by at least a factor $\sim \varepsilon = 1/G^2$.

The general stability in ω and α was demonstrated by Cutler[52] when any highly nonlinear element was present in the system,

but the above additional stability in A and β is a particular characteristic of the present cavity system which employs a Kerr type nonlinearity and a high value of G. As an alternative example, the stability of the low gain case, $G \gtrsim 1$ can be analyzed in a similar way as above, leading to the following "double transit" deviation matrix, where $\varepsilon' = G^2 - 1 << 1$:

$$\begin{pmatrix} (\delta\hat{A}^2)_{n+2} \\ (\delta\hat{\alpha})_{n+2} \\ (\delta\hat{\omega})_{n+2} \\ (\delta\hat{\beta})_{n+2} \end{pmatrix}_{G\gtrsim 1} \approx \begin{pmatrix} 1 & -\varepsilon' & \frac{3}{2}\varepsilon' & 0 \\ 0 & 1-2\varepsilon' & 0 & 2\varepsilon' \\ 0 & 0 & 1-2\varepsilon' & 0 \\ 2\varepsilon' & 0 & -\frac{5}{2}\varepsilon' & 1-2\varepsilon' \end{pmatrix} \begin{pmatrix} (\delta\hat{A}^2)_n \\ (\delta\hat{\alpha})_n \\ (\delta\hat{\omega})_n \\ (\delta\hat{\beta})_n \end{pmatrix} .$$

This shows that the low gain case is only "neutrally" stable, displaying little tendency toward reaching the steady-state condition; this agrees with our computer results for the "coupled-mode" approach, as mentioned in section II.1.i. The general matrix, $\hat{m}$, relating the deviation parameters for the $(n+1)^{th}$ pulse to those of the n^{th} pulse, such as that shown in equation II.2.7 for $G >> 1$, has $DET|\hat{m}| = \frac{1}{G^6} < 1$. This shows that any pulse will eventually reach the steady-state form, for any $G > 1$.

Physically, the above analysis shows that the nonlinear polarization of the Kerr medium feeds energy into frequencies outside the gain linewidth, as shown by the equation for β°; this acts as an effective loss mechanism, reducing the round trip energy gain to unity. Figure 8 shows a qualitative example of the energy balance

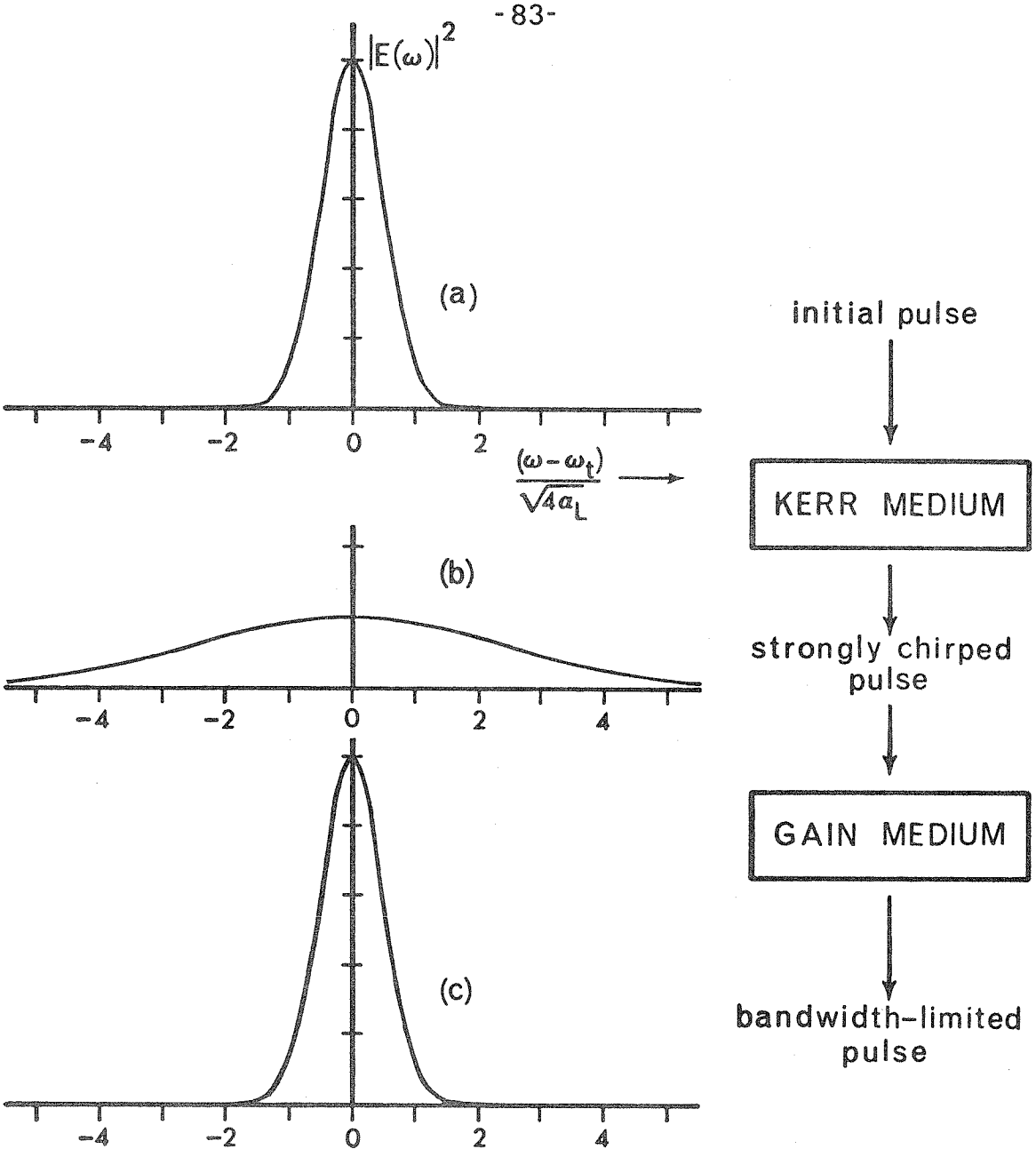


Figure 8. Circulating pulse power spectra for the "high gain" case.
(a) Spectrum of pulse before entering Kerr medium.
(b) Pulse spectrum after Kerr medium. The total energy has not changed, but the spectrum is much broader due to chirp.
(c) Output after amplification of the line center frequencies by a factor of G in the gain medium.
In steady-state pulsing, curves (c) and (a) are the same, since only 1/G of the broadened spectrum in (b) lies within the gain bandwidth. (Curves based on G=5.)

mechanism for the high gain case. This view shows why a pulsed output will occur: only a strong, short pulse traveling through the Kerr medium will generate sufficient new frequencies to create the "loss" which balances the gain for the frequencies near ω_t.

The pulsed output thus occurs independently of any distortion or relaxation-time effects in the Kerr medium, although the strength of the stability for the steady-state pulse depends on the extent of the frequency broadening near ω_t, which should be proportional to the pulse intensity. We will also show that it is desirable to have a "linear" chirp during the peak of the pulse, so that pulse compression can be achieved.

When the pulse leaving the Kerr cell with a broad frequency spectrum passes through the profiled gain medium, the output will be essentially a bandwidth-limited "ultrashort pulse", regardless of the detailed features of the input spectrum. As shown in Figure 7 (page 74) a large loss, $R_1R_2 \ll 1$, coupled with the normal Lorentzian gain curve, yields an acceptably "limited" bandwidth transfer function, justifying equation II.1.25. This requires a very large laser line-center gain, though, so that $G \gg 1$ can still apply. An alternative would be to introduce an independent, bandwidth-limiting device (filter) in the cavity to absorb frequencies outside the gain linewidth. This would reduce the inherent gain required in the laser medium; in addition the value of α_L would then be less dependent on pumping, gain medium temperature, cavity configuration, etc.

The detailed shape of the profiled gain curve would of course

result in an effective linewidth and an effective gain constant in equation II.2.2, differing from the definitions in equation II.1.26, but the physical interpretation is unaffected. It was also assumed that the gain mechanism was linear in the sense that the output pulse amplitude was proportional to the input pulse amplitude; nonlinear (saturation) effects in the laser medium, such as those described by equation I.2.40b, were thus not included. However, the physical picture, as discussed above, of the role played by the gain medium shows that saturation would not significantly alter the processes leading to a stable pulsed output.

Finally, in connection with the physical model of pulse formation, it should be noted that any strongly nonlinear element in a laser cavity, plus a large excess laser gain, could lead to a pulsed output of the type discussed here. The strong stability, however, occurs only when the amount of frequency broadening (as compared, for example, to energy redistribution among frequencies already present) produced by the nonlinearity is proportional to the pulse intensity, as is the case with the cubic polarization effect.

II.2.ii compression of chirped pulses

According to equation II.1.8, the minimum "time-bandwidth" product obtainable in a pulse is given by

$$\tau_p \cdot \delta\omega \geq \frac{1}{2} \qquad \text{(II.2.9a)}$$

where the equality holds for a Gaussian pulse with no chirp. On the

basis of its bandwidth alone, then, a chirped pulse should be capable of producing a "compressed" pulse with a time duration of

$$\tau_{comp} = \frac{1}{2(\delta\omega)} = \frac{1}{2}\sqrt{\frac{\alpha}{\alpha^2+\beta^2}} = (\tau_p)_{chirp}\left(1 + \frac{\beta^2}{\alpha^2}\right)^{-1/2} . \quad \text{(II.2.9b)}$$

It is well known[55] that such a compressed pulse can be produced from a chirped pulse by using a dispersive delay line (i.e. an element whose group delay depends on the frequency); the physical principle is demonstrated in Figure 4 (on page 58) in section II.1.i. In the case of a linear chirp, such as involved in the pulses here, this frequency-dependent group delay requires a quadratic phase shift; we can thus define a "compressor" whose transfer function is given by

$$g_{comp}(\omega) = e^{i\gamma_c(\omega-\omega_c)^2} . \quad \text{(II.2.10)}$$

It can be seen from the Fourier transform in equation II.1.6b that a linear phase shift simply leads to an overall delay, as expected, and is thus not significant. Also a true filter might also have a bandwidth, Γ_c, and a loss factor, B_c, (as well as higher phase terms), so that

$$g_{comp} = B_c e^{-\frac{(\omega-\omega_c)^2}{\Gamma_c^2}} e^{i\gamma_c(\omega-\omega_c)^2}$$

would be more appropriate, but the constant, B_c, is of no fundamental significance, and we will assume that $|\omega_t \pm \delta\omega - \omega_c| << \Gamma_c$; this leads to the form in equation II.2.10. As an example, a

"double diffraction grating system" can be arranged with these characteristics[56].

For an input pulse given by the "standard form":

$$E_{in} = A\, e^{-\alpha t^2}\, e^{-i(\omega_o t + \beta t^2)} \quad , \tag{II.2.11a}$$

the effect of passage through the compressor characterized by $g_{comp}(\omega)$ is to produce an output pulse characterized by

$$E_{out} = A_{out} e^{-\alpha_{out} t^2}\, e^{-i(\omega_o t + \beta_{out} t^2)} \quad , \tag{II.2.11b}$$

where

$$A_{out} = A[1 + 8\beta\gamma_c + 16\,\gamma_c^2(\alpha^2 + \beta^2)]^{-1/4}$$

$$\alpha_{out} = \frac{\alpha}{[1 + 8\beta\gamma_c + 16\,\gamma_c^2(\alpha^2+\beta^2)]}$$

$$\omega_{out} = \omega_o \tag{II.2.11c}$$

$$\beta_{out} = \frac{\beta + 4\gamma_c(\alpha^2 + \beta^2)}{[1 + 8\beta\gamma_c + 16\gamma_c^2(\alpha^2+\beta^2)]} \quad .$$

The optimum compression (i.e. maximum α_{out}) is attained for $\gamma_c = \gamma_{opt}$, where

$$\gamma_{opt} = \frac{-\beta}{4(\alpha^2+\beta^2)} \quad . \tag{II.2.12a}$$

When $\gamma_c = \gamma_{opt}$, we obtain

$$(A_{out})_{opt} = A\left(\frac{\alpha^2+\beta^2}{\alpha^2}\right)^{1/4}$$

$$(\alpha_{out})_{opt} = \frac{\alpha^2+\beta^2}{\alpha} \qquad \text{(II.2.12b)}$$

$$(\beta_{out})_{opt} = 0 \ .$$

We see that $(\tau_p)_{opt} = \frac{1}{2}\left(\frac{\alpha}{\alpha^2+\beta^2}\right)^{1/2} = \frac{1}{2}\ \frac{1}{\delta\omega}$, indicating that the "compressor" has indeed shortened the pulse to its minimum theoretical width.

In the present application, the pulses of interest are the steady-state solutions given in equation II.2.2, which can be passed through a compressor described by

$$\gamma_o = -\frac{G}{4\alpha_L(G^2-1)} \simeq \frac{-1}{4\alpha_L G} \ , \qquad \text{(II.2.13)}$$

where the last value holds in the high gain case. The compressed pulse-width is given by

$$(\tau_o)_{comp} = \frac{1}{2(\alpha_{comp})^{1/2}} = \frac{1}{2[\alpha_L(G^2-1)]^{1/2}} \simeq \frac{1}{2(\alpha_L)^{1/2}}\ \frac{1}{G} = \frac{\tau_{ML}}{G} \ . \qquad \text{(II.2.14)}$$

Therefore, in the high gain case, we can use the chirp on the pulses to compress them by a factor of G, thus generating pulses which are G times shorter than the "fully mode-locked" case. As an example, τ_{ML} for Nd:glass is about 10^{-13}sec, so that a cavity operated with G = 5 will allow the generation of pulses with a width of $\sim 2 \times 10^{-14}$sec. Theoretically, this compression could be performed inside the laser resonator without substantially affecting our analysis,

but the most practical system probably would place the compressor outside the cavity. The high stability against pulse-to-pulse variations in our laser system makes this approach superior to that in which a normally mode-locked laser pulse train is externally passed through a Kerr medium to introduce a chirp, and then passed through a compressor[54].

II.3 Q-Switched Laser Operation

The desirable characteristics of the laser system being analyzed are a result of the highly nonlinear pulse propagation in the Kerr medium; the peak power density required for such operation can be found from the steady-state intensity, given in equation II.2.2. Thus

$$P(\text{erg sec}^{-1}\ \text{cm}^{-2}) = \frac{n_o \lambda c}{128\pi^3\, \bar{\chi}(1+C_o) L_K (R_2^{1/2}+R_2^{-1/2})} \left(\frac{G^2-1}{G}\right) .$$

As an order of magnitude estimate, we use $\lambda = 1\mu$, $G = 5$, $\bar{\chi}(1+C_o) = 2 \times 10^{-14}$, $L_K = 1$ cm, $(R_2^{1/2}+R_2^{-1/2}) = 2$, giving a peak power of $P = 1.4 \times 10^4$ Megawatts/cm^2, with the corresponding average power density in the range of Megawatts/cm^2. This implies that the laser must be Q-switched, so that the steady-state pulsing analysis will not apply exactly.

Our idealized model of Q-switched laser operation will be a very common one[57]. After the gain medium inversion is "pumped" to a high level, the cavity "Q" is "instantly" switched to a high value, so that $G = G_o > 1$ holds initially. We next assume that the pumping

and spontaneous transition in the gain medium are negligible during the time of the Q-switched output. Therefore, as the initial noise field is amplified to form the observed output, the gain medium population inversion is depleted, until $G < 1$ eventually holds, and the output "dies out". In the present case, the net result will be a train of ultrashort pulses modulated by an overall "Q-switched envelope"; we will here find examples of such pulse trains.

In practice, the usefulness of the ultrashort pulses produced by conventional Q-switched lasers is severely limited, both by pulse-to-pulse variations within a given train, and also by the irreproducibility of the train as a whole. However, the present system containing the Kerr cell is capable of shaping the pulses to a "stable" form in only a few round trips, and therefore we expect improvements in both types of "reproducibility". This section will demonstrate some of the advantages, and disadvantages, of our laser system when operated in the Q-switched mode.

II.3.i. initial pulse growth

The original "signals" present in the laser oscillator are the noise fields produced by spontaneous emission in the gain medium; their spectral distribution will thus have a width, Γ. The random temporal fluctuations of such a noise source, at the instant of Q-switching, can be expected to contain at least one "pulse", which will then evolve according to equation II.1.29 to form the "circulating pulse" in the cavity. Because of the strong stability of our system in the high gain case, it is apparent that any such noise pulse will

be shaped to conform to the steady-state form as its intensity becomes equal to the steady-state intensity; to this extent, the initial pulse growth is not crucial to the form of the "output" pulse train.

These arguments, however, suggest a major disadvantage to this laser system, since any noise pulse is amplified and shaped to conform with the steady-state form. This would lead to a very "dirty" pulse train, since a large number of circulating pulses would be present simultaneously; it is far more desirable to obtain an output train corresponding to only a single circulating pulse.

A comparison of this problem with the behavior of the normal laser system containing a saturable absorber indicates a fundamental difference in the roles played by the nonlinearity; the Kerr medium acts as a "pulse shaper", while the saturable absorber acts primarily as a "pulse selector". In the latter case, the strongest (i.e. usually, shortest) spike in the initial cavity noise is preferentially amplified; there is only a little "shaping" done by the saturable absorber during the majority of the output train[17]. This explains why the overall pulse train reproducibility is poor, since the irreproducible characteristics of the initial noise pulse are still present to a large degree in the later output. Furthermore, the experimentally observed[28] pulse-to-pulse variations within a single train show that there remain pulse altering mechanisms (residual Kerr effect, dispersion, gain spectral narrowing, resonant nonlinearities, etc.) which do not stabilize over the time of the Q-switched envelope.

The Kerr medium in our laser will prevent these variations, but an initial "pulse selection" mechanism appears to be lacking.

With this problem in mind, we will examine the evolution equations during the initial pulse growth. First, the amplitude evolution equation in II.1.29 shows that a pulse centered in frequency, $\Delta\omega_1 = 0$, sees the highest gain; since the noise spectrum itself is centered on ω_t, we will assume that $\Delta\omega = 0$ for any noise pulse which is amplified enough to be "important". With this minor simplification, we will now rewrite the "laser evolution equations", relating the $(n+1)^{th}$ pulse parameters to those of the n^{th}:

$$\varepsilon_{n+1} = \varepsilon_n \frac{G}{\left[1+\frac{\alpha_n}{\alpha_L}+\frac{\beta_n^2}{\alpha_n\alpha_L}\right]^{1/2}}$$

$$\alpha_{n+1} = \alpha_n \frac{\left[1+\frac{\alpha_n}{\alpha_L}+\frac{\beta_n^2}{\alpha_n\alpha_L}\right]}{\left[\left(1+\frac{\alpha_n}{\alpha_L}\right)^2+\left(\frac{\beta_n}{\alpha_L}\right)^2\right]} \qquad \text{(II.3.1)}$$

$$\beta_{n+1} = \frac{\beta_n}{\left[\left(1+\frac{\alpha_n}{\alpha_L}\right)^2+\left(\frac{\beta_n}{\alpha_L}\right)^2\right]} + \left(\frac{8D_o(2\pi)^{1/2}}{c}\right)\varepsilon_{n+1}(\alpha_{n+1})^{3/2}$$

where

$$\varepsilon_n \equiv \frac{c}{8(2\pi)^{1/2}} \frac{A_n^2}{(\alpha_n)^{1/2}} . \qquad \text{(II.3.2)}$$

Here, ε_n is the energy density (i.e. ergs/cm^2) in the n^{th} pulse.

The initial noise pulse will be characterized by the values $\varepsilon_o, \alpha_o, \beta_o$. During the early stages of the amplification, the pulse will be much smaller in energy than the steady-state value, and since $\alpha_n < \alpha_L$ will always hold, the second term in the β_{n+1} formula in

II.3.1, which represents the Kerr medium nonlinearity, will be negligible. By also ignoring any depletion of the gain medium inversion during this part of the pulse train, so that G and α_L do not change, we can solve equation II.3.1. We find that the pulse spectrum becomes "gain narrowed" according to

$$(\delta\omega^2)_n \equiv \frac{\alpha_n^2+\beta_n^2}{\alpha_n} = \frac{(\delta\omega^2)_o}{1 + n\,\frac{(\delta\omega^2)_o}{\alpha_L}}, \qquad \text{(II.3.3a)}$$

and that the individual pulse parameters after n transits are given by

$$\varepsilon_n = \varepsilon_o \frac{G^n}{[1 + n(\frac{\alpha_o^2+\beta_o^2}{\alpha_o\alpha_L})]^{1/2}}$$

$$\alpha_n = \alpha_o \frac{[1 + n\,(\frac{\alpha_o^2+\beta_o^2}{\alpha_o\alpha_L})]}{[(1+ n\,\frac{\alpha_o}{\alpha_L})^2 + (n\,\frac{\beta_o}{\alpha_L})^2]} \qquad \text{(II.3.3b)}$$

$$\beta_n = \frac{\beta_o}{[(1 + n\frac{\alpha_o}{\alpha_L})^2 + (n\,\frac{\beta_o}{\alpha_L})^2]} .$$

Note that $\alpha_n \to \frac{\alpha_L}{n}$, independently of α_o, while $\beta_n/\alpha_n \to \sim 1/n$, so that the relative chirp becomes negligible during this stage of the pulse evolution. Because the spectral width of the noise is given by $(\delta\omega^2)_{noise} \sim \alpha_L$, we can expect that

$$\frac{(\delta\omega^2)_o}{\alpha_L} = \frac{\alpha_o^2 + \beta_o^2}{\alpha_o\alpha_L} \approx 1$$

for any noise pulse, although α_0 and β_0 themselves might vary widely. We thus see, from equation II.3.3b, that there is essentially no "pulse selection" capability in the early amplification stages. To obtain a "clean" pulse train, we must add additional components to the system (or else consider more effects in those already present).

A number of methods are possible, by which the Kerr medium nonlinearity can be utilized to yield an effective intensity-dependent loss, so that the laser oscillator preferentially amplifies the strongest noise pulse initially present. One method would employ the ability of the cubic nonlinearity, such as described by equations I.3.2 and I.3.9, to rotate the axes of elliptically polarized beams by an amount proportional to the intensity. This device would act in the same way as the "mode-locking" system of Dahlström[58,59], except that this aspect of the nonlinearity would be designed to act primarily at the lower power levels achieved in the "buildup" portion of the Q-switched pulse train. Another approach, illustrated in Figure 9, would involve placing divergent lenses in the beam before the Kerr cell, and utilizing the intensity-dependent "focusing" properties of the cubic nonlinearity to reduce the diffraction losses for the stronger pulses. Such a system would also help prevent the undesirable effect of complete "self-focusing", as discussed in later chapters.

Probably the simplest method of obtaining a "clean" pulse train would be to insert a normal mode-locking component into the cavity; either an active modulator or passive saturable absorber would

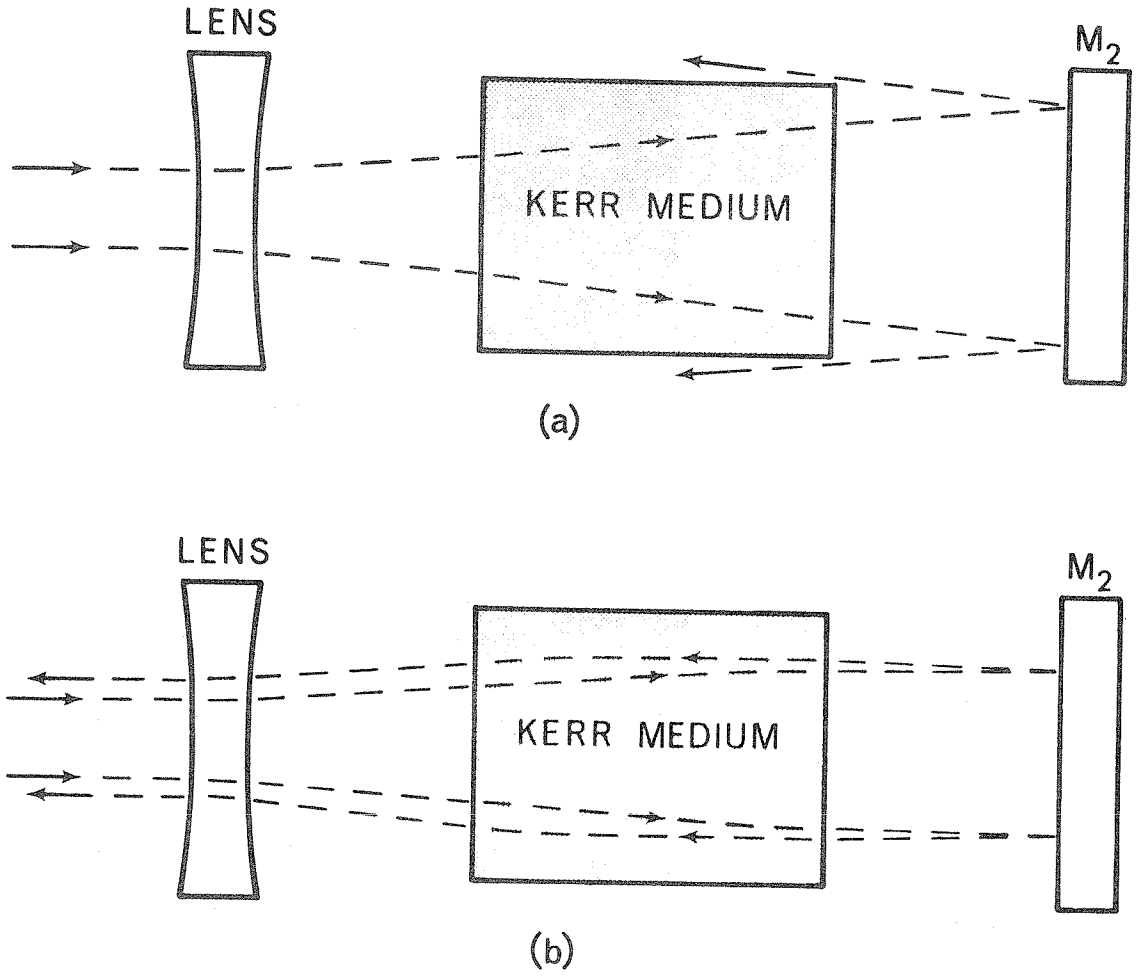


Figure 9. Intensity-dependent loss formed with lens and Kerr cell.
(a) Low intensity pulses are diverged by the lens, leading to high losses.
(b) For high intensity pulses, partial self-focusing in the Kerr cell compensates for the lens divergence.

be appropriate. Finally, there are nonlinear processes in the gain medium itself, which tend to produce mode-locking in many lasers which do not contain any separate nonlinear component[60]. Any of these "effective" intensity-dependent losses would also tend to sharpen the pulse somewhat during the rise of the pulse train; and, as we will show, even a moderate decrease in pulselength over that predicted by equation II.3.3b will greatly improve the overall system operation.

We will now assume that some process occurs to "select" an initial noise pulse; the evolution of this pulse will then be governed by equation II.3.1. We will consider the effects of gain depletion in the next section, but most of the pertinent conclusions concerning the leading edge of the Q-switched pulse train can be seen by keeping G and α_L fixed, and observing the evolution of the pulse towards the steady-state form, as given by

$$\bar{\varepsilon} = \frac{c}{8(2\pi)^{1/2}} \quad \frac{1}{D_o} \quad \frac{(G^4 - 1)^{1/2}}{G(\alpha_L)^{1/2}}$$

$$\bar{\alpha} = \alpha_L \frac{G^2 - 1}{G^2 + 1} \qquad \text{(II.3.4)}$$

$$\bar{\beta} = \alpha_L \, G \, \frac{G^2 - 1}{G^2 + 1} \quad .$$

The initial conditions, for reasons discussed earlier, will have $(\delta\omega^2)_o = \alpha_o + \frac{\beta_o^2}{\alpha_o} \sim \alpha_L$, so that the "spikes" can be expected to have $\alpha_o \gtrsim \alpha_L$, and $\beta_o \lesssim \alpha_L$. In the next section, we will find that $\varepsilon_o/\bar{\varepsilon} \sim (10^{-12} - 10^{-8})$; this means that $n \gg 1$ will hold before the large "output" pulses are reached, and that initial conditions on

α_o and β_o are not crucial.

Figure 10 shows the evolution of a noise pulse toward the steady-state limit for a low gain amplifier, (G = 1.2), based on equation II.3.1. The significant conclusions are that on the order of 10^2 round trip transits are required to reach the "output" portion of the train; the stabilizing ability of the Kerr effect is also seen to be very poor, as expected for low gain, requiring $\sim 10^2$ more transits to damp out the transients. Furthermore, it can be seen that there is a significant "overshoot", in which the pulses become much larger than $\bar{\varepsilon}$, before the Kerr effect can "overcome" the gain.

Figure 11 shows a similar calculation for a typical "high gain" case. Here the pulse size builds up over $\sim$ 20 transits, and the strong stability of the steady-state pulsing condition is apparent. Although only a few transits are required to damp out the transients, it is seen that there again exists a strong "overshoot", during which the pulse energy is many times larger than $\bar{\varepsilon}$. This overshoot feature, which will still occur when gain depletion is included, is undesirable in many laser applications, since the relative importance of these pulses will be increased when nonlinear effects are under investigation. Also, the pulses in the overshoot are "irreproducible", since their characteristics depend on the initial conditions; only the net energy of all the pulses in the overshoot is relatively invarient.

An examination of Figure 11 shows that the overshoot occurs because α_n has become very small during the initial pulse growth, due to the spectral gain narrowing predicted by equation II.3.3. These pulses, which are thus broad in time, must become very intense before the Kerr medium can produce the strong chirp needed to create the "loss"

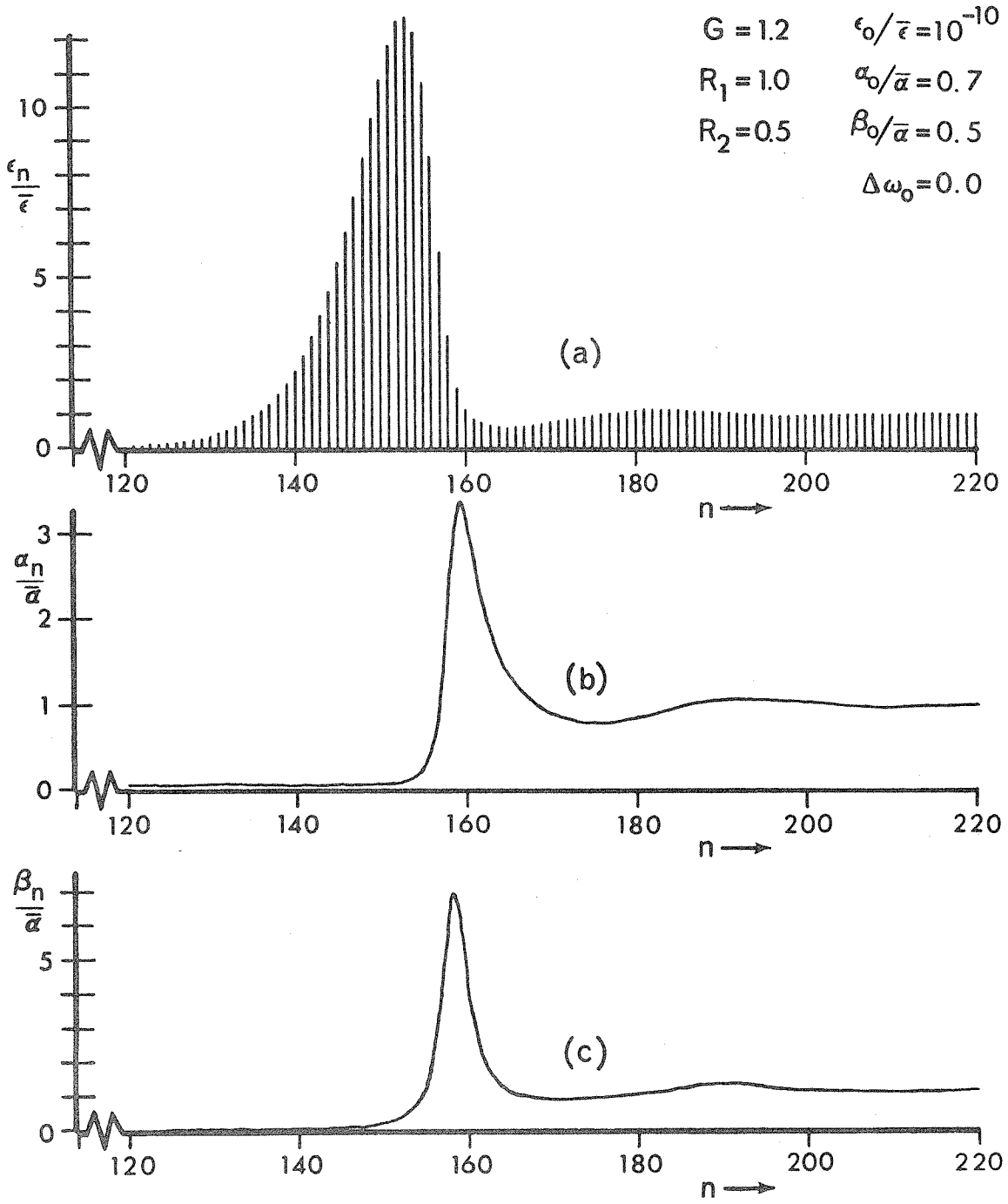


Figure 10. Approach to steady-state pulsing in low gain laser system.
(a) Relative pulse energy, ε_n, vs. pulse number, n.
(b) Variation of pulsewidth parameter, $\alpha_n/\bar{\alpha} = (\bar{\tau}_p/\tau)^2$.
(c) Relative chirp of pulses.

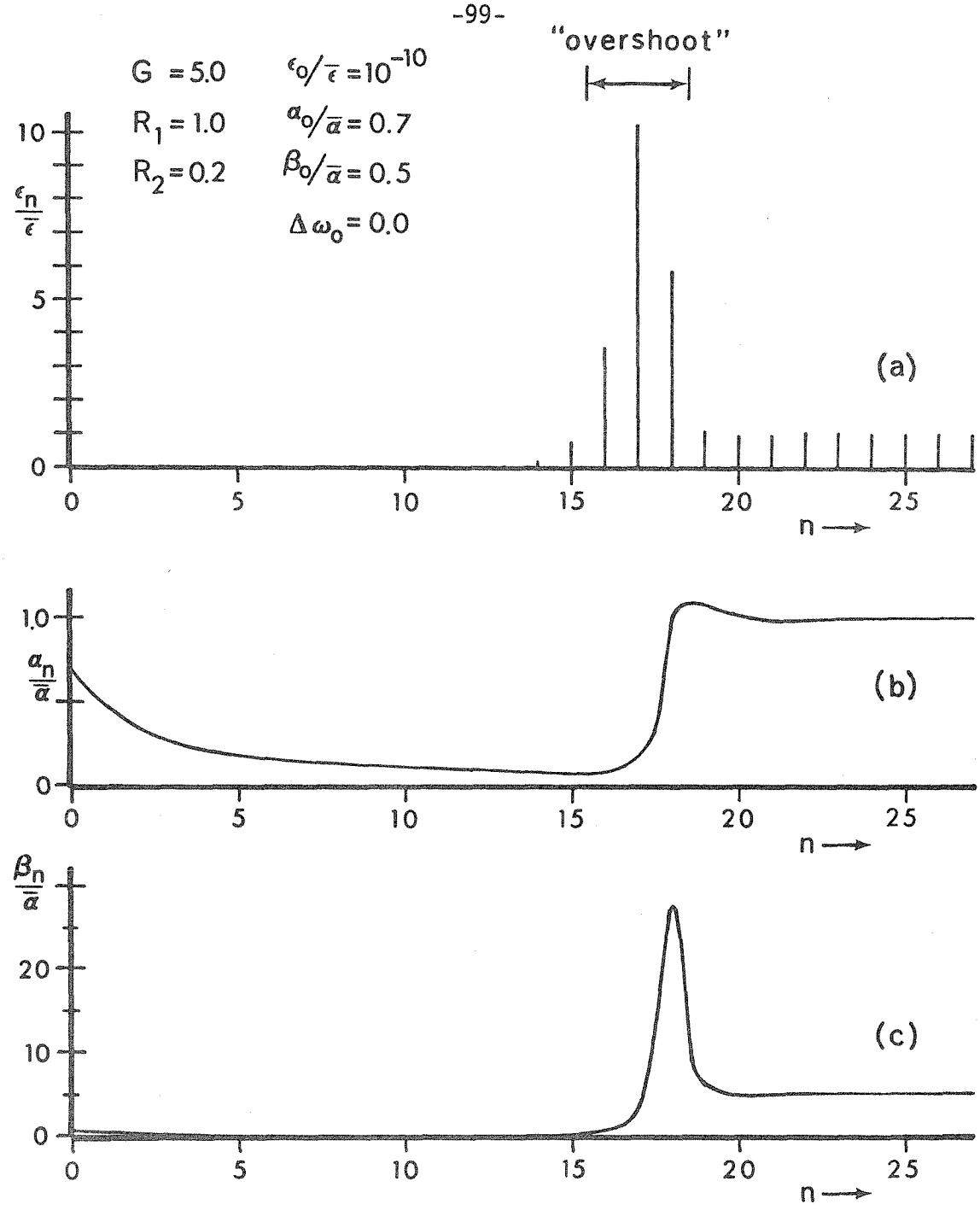


Figure 11. Approach to steady-state pulsing for the high gain case.
(a) Relative pulse energy, ϵ_n, vs. pulse number, n.
(b) Variation of pulsewidth parameter, $\alpha_n/\bar{\alpha} = (\bar{\tau}_p/\tau)^2$.
(c) Relative chirp of pulses.
Note the strong "overshoot" pulses, with extreme chirp.

which, as shown in Figure 8, balances the gain. An examination of the evolution equations, II.3.1, shows that the overshoot would be very small, if α_n were close to $\bar{\alpha}$ at the time when ε_n was approaching $\bar{\varepsilon}$. Physically, this means that it would be very advantageous to include in the cavity a "pulse sharpening" device, such as the nonlinear losses discussed above, which would keep the pulses from broadening too much during the initial amplification stages. For example, the major "pulse sharpening" action of a saturable absorber occurs when the peak pulse intensity is nearly equal to the absorber's saturation intensity, I_s[61]; the ideal saturable absorber would thus have I_s slightly lower than the steady-state pulse peak intensity, $\bar{I} = \frac{1}{D_o}\frac{G^2-1}{G}$.

An accurate description of such a laser system, including both the self phase modulation and the nonlinear loss, is outside the scope of the present analysis; however, we can indicate the benefits of "pulse sharpening" as shown in Figure 12. Here the pulse evolution has been found for the same high gain case as shown in Figure 11, except that the pulse has been "sharpened" when $I_n = 10^{-3}\bar{I}$, so that, at that point, $\alpha_n = \alpha_L$. It should be noted that this "sharpening" itself does not directly interfere with the role of the Kerr nonlinearity, since the self phase modulation is negligible at that point; the residual effect of the sharpening is thus simply to increase "α" (i.e. decrease the pulsewidth) by a relatively small amount at the beginning of the overshoot. A comparison of Figures 11 and 12, however, shows that even this small effect has significantly reduced the overshoot, thus improving the overall approach to the steady-state limit.

II.3.ii Q-switched pulse train

As the amplifying medium adds energy to the circulating

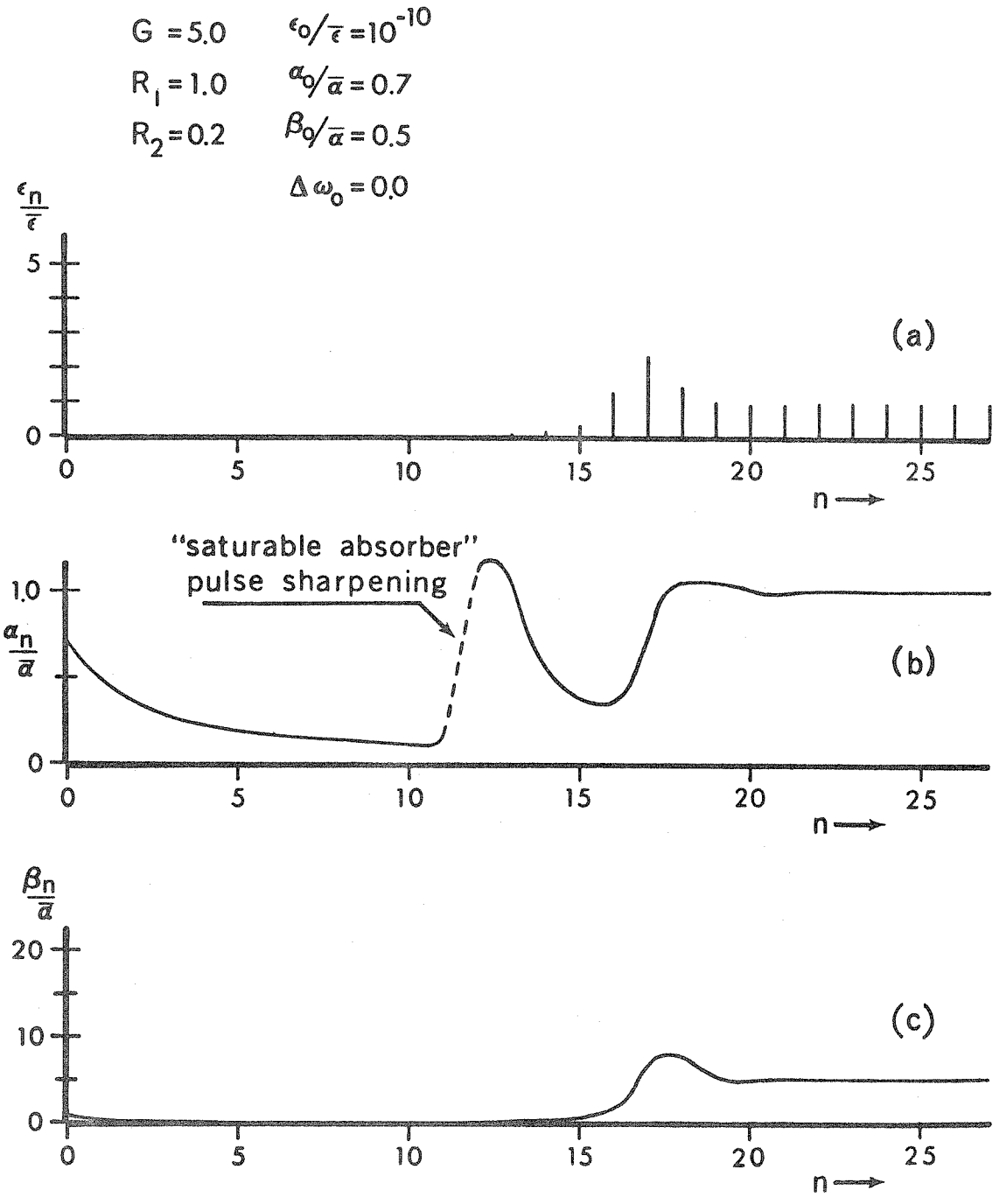


Figure 12. Approach to steady-state pulsing for a high gain laser system including a "pulse sharpener".
(a) Relative pulse energy, ε_n, vs. pulse number, n.
(b) Variation of pulsewidth parameter; $\alpha_n/\bar{\alpha} = (\bar{\tau}_p/\tau)^2$. The idealized pulse sharpening occurs at pulse 12.
(c) Relative chirp of pulses.

pulse, the population inversion is depleted, leading to changes in G (and α_L), and causing the true output to "die out". According to equations II.1.15 and II.1.26, the line-center round trip power gain for the cavity can be written:

$$G = R_1 R_2 e^{2a_c E_G}$$

where

$$E_G = \frac{1}{2}(N_t - N_g)\hbar\omega_t L_G \qquad \text{(II.3.5)}$$

$$a_c = \frac{8\pi e^2 \langle r^2 \rangle}{\hbar^2 \Gamma_c} \left(\frac{n_o^2 + 2}{3n_o}\right) .$$

Here E_G equals the inversion energy available (per cm^2 cross section) of the amplifying medium, since the induced emission of one atom/cm^3 will extract an energy of $(\hbar\omega_t L_G)/cm^2$ and similarly reduce E_G by this same amount. (We ignore the effects of degeneracy in II.3.5, as we did in section I.2.iv.)

We will assume that E_G changes relatively little during any single round trip of the pulse, so that $E_G = E_G(n)$ can be considered a constant of the system during the n^{th} pulse transit. We will then account for gain depletion by changing E_G before the next transit, in accordance with energy conservation. By transforming the fields of the "effective cavity" in Figure 6c to those of the "true cavity" in Figure 6a, we find that the energy added to the circulating pulse during the n^{th} transit is given by

$$\Delta\varepsilon_n = \varepsilon_n \left\{ \frac{R_2^{-1/2} G}{\left[1 + \frac{\alpha_n}{\alpha_L} + \frac{\beta_n^2}{\alpha_n \alpha_L}\right]^{1/2}} \exp\left[- \frac{(\Delta\omega_n)^2}{2\alpha_L \left(1 + \frac{\alpha_n}{\alpha_L} + \frac{\beta_n^2}{\alpha_n \alpha_L}\right)} \right] - R_2^{1/2} \right.$$

$$+\frac{(R_1^{-1/2}-R_1^{1/2})G^{1/2}}{\left[1+\frac{\alpha_n}{2\alpha_L}+\frac{\beta_n^2}{2\alpha_n\alpha_L}\right]^{1/2}}\exp\left[-\frac{(\Delta\omega_n)^2}{4\alpha_L\left(1+\frac{\alpha_n}{2\alpha_L}+\frac{\beta_n^2}{2\alpha_n\alpha_L}\right)}\right]\Bigg\}, \quad \text{(II.3.6a)}$$

where $G = G(n)$ and $\alpha_L = \alpha_L(n)$ are the values of G and α_L during the n^{th} transit. To simplify the formulas, we will usually consider the case where $R_1 = 1$, for which we have

$$\Delta\varepsilon_n = \varepsilon_{n+1}R_2^{-1/2} - \varepsilon_n R_2^{1/2} \quad . \quad (R_1 = 1 \text{ case}) \qquad \text{(II.3.6b)}$$

This energy will be extracted from the gain inversion, giving the "evolution equations" for G and α:

$$G(n+1) = R_1R_2\exp[2a_cE_G(n+1)]$$

$$= G(n)\exp[-2a_c\Delta\varepsilon_n] \quad , \qquad \text{(II.3.7a)}$$

and, from equationII.1.26,

$$\frac{1}{\alpha_L(n+1)} = \frac{1}{\alpha_L(n)} - \frac{4a_c\Delta\varepsilon_n}{\Gamma^2} \quad . \qquad \text{(II.3.7b)}$$

In order to characterize the gain medium energy, we can form an estimate of the number of pulses we expect in the Q-switched train as follows. First, the "useful" energy in the inversion, ΔE_G, can be defined as the amount which is extracted before the gain is reduced to unity:

$$\Delta E_G = \frac{1}{2a_c}\ln(G_o) \quad , \qquad \text{(II.3.8a)}$$

where $G_o = G(0)$ is the "undepleted" gain. We similarly define $\alpha_L^o = \alpha_L(0)$, and $\bar{\varepsilon},\bar{\alpha},\bar{\beta}$ according to equation II.3.4 with $G \to G_o$, $\alpha_L \to \alpha_L^o$. From equation II.3.6a, the energy extraction per transit required in the undepleted steady-state pulsing condition is given by

$$\overline{\Delta\varepsilon} = \bar{\varepsilon}[(R_2^{-1/2} - R_2^{1/2}) + (R_1^{-1/2} - R_1^{1/2})(\frac{2G_o}{G_o^2+1})^{1/2}] \quad . \tag{II.3.8b}$$

The ratio of these, m_t, then gives the number of undepleted steady-state pulses which would be present in the Q-switched train:

$$m_t = \frac{\Delta E_G}{\overline{\Delta\varepsilon}} \quad . \tag{II.3.8c}$$

In the actual Q-switched train, the presence of gain depletion and the departure of the pulses from the steady-state form will cause the actual number of pulses generated to differ from m_t, but equation II.3.8 will still be used to normalize the energy of the gain medium to that of the "standard pulse energy".

With these definitions, we can write

$$G(n+1) = \exp[\ell n(G_o)(1 - \frac{(\Delta\varepsilon_o+\Delta\varepsilon_1+\cdots+\Delta\varepsilon_n)}{m_t\ \overline{\Delta\varepsilon}})] = G(n)\exp\left[-\frac{\Delta\varepsilon_n \ell n(G_o)}{\overline{\Delta\varepsilon}\ m_t}\right]$$

$$\frac{1}{\alpha_L(n+1)} = \frac{1}{\alpha_L^o}\left[1 - \frac{\ell n(G_o)}{\ell n(\frac{G_o}{R_1R_2})}\ (\frac{\Delta\varepsilon_o+\Delta\varepsilon_1+\cdots+\Delta\varepsilon_n}{m_t\ \overline{\Delta\varepsilon}})\right] = \frac{1}{\alpha_L(n)} - \frac{\Delta\varepsilon_n \ell n(G_o)}{\overline{\Delta\varepsilon}\ m_t\ \ell n(\frac{G_o}{R_1\ R_2})} \quad . \tag{II.3.9}$$

These equations, along with the pulse parameters equations, II.3.1, with $G \to G(n)$ and $\alpha_L \to \alpha_L(n)$, form a complete set of

evolution equations for the Q-switched laser system, with $\Delta\epsilon_n$, $\overline{\Delta\epsilon}$, and $\bar{\epsilon}$ defined in equations II.3.6, II.3.8, and II.3.4, respectively. The behavior of the system can be seen to be characterized entirely by the dimensionless parameters: G_o, R_1, R_2, m_t, and by a normalized set of initial conditions, such as $\epsilon_o/\bar{\epsilon}$, $\alpha_o/\bar{\alpha}$, $\Delta\omega_o/(\bar{\alpha})^{1/2}$ and $\beta_o/\bar{\alpha}$.

We will now form an order of magnitude estimate of ϵ_o, the energy in a typical noise pulse, as follows. The average "resonating" intensity of the noise field, I_{ave}, is given by

$$I_{ave} = \left(\frac{\text{Total spontaneous energy/sec}}{\text{beam cross section}}\right)\cdot\left(\frac{\text{solid angle for feedback}}{4\pi}\right)\cdot(\text{cavity Q})$$

$$\simeq \left(\frac{\hbar\omega_t N_t \mathcal{A} L_G}{t_{spont}\, \mathcal{A}}\right)\cdot\left(\frac{\mathcal{A}/(2L_c)^2}{4\pi}\right)\cdot\left(\frac{1}{1-R_1R_2}\right) \qquad \text{(II.3.10)}$$

$$= \frac{E_G\, \mathcal{A}}{8\pi L_c^2 t_{spont}}\left(\frac{N_t}{N_t-N_g}\right),$$

where $\mathcal{A}$ is the beam cross section, and L_c is the cavity length. The "minimum energy noise pulse" would have this peak intensity: $(\epsilon_o)_{min} \simeq I_{ave}(\tau_p) \sim I_{ave}/(4\alpha_L^{\circ})^{1/2}$. Because of phase fluctuations, however, the expected noise "spike" would have a higher peak intensity, ranging up to the "maximum energy noise pulse" which contains the entire cavity power: $(\epsilon_o)_{max} \simeq I_{ave}\left(\frac{2L_c}{c}\right)$. In a 4-level laser, we can take $N_t/(N_t-N_g) \sim 1$ in II.3.10, and in a 3-level laser, we can use the rough approximation, $N_t/(N_t-N_g) = \frac{1}{2}\left(1 + \frac{N_t+N_g}{N_t-N_g}\right) \sim \frac{1}{2} + \frac{N_o}{2(N_t-N_g)}$, where

N_o is the total number density of active gain molecules.

As an example of a realistic laser system, we will consider designing a Ruby laser with $G_o = 5$, $R_1 = 1$, $R_2 = 0.2$, $m_t = 40$. The spontaneous lifetime appearing in equation II.3.10, for an amplifying medium excited state is given by(62)

$$\frac{1}{t_{spont}} = \frac{4e^2 \langle r^2 \rangle \omega_t^2}{\hbar c^3} \quad , \qquad \text{(II.3.11a)}$$

so that we can write

$$a_c = \frac{2\pi c^2}{\hbar \Gamma \omega_t^3 t_{spont}} \left(\frac{n_o^2 + 2}{3n_o}\right) \quad . \qquad \text{(II.3.11b)}$$

For Ruby at 20°C, we will use $t_{spont} = 3 \times 10^{-3}$ sec, $\omega_t = 2.7166 \times 10^{15} sec^{-1}$, $\Gamma = 1.0 \times 10^{12}$ sec^{-1}, $n_o = 1.76$. The following thus describe the system in this case:

$$a_c \simeq 1.3 \times 10^{-7} cm^2/erg$$

$$\Delta E_G \simeq 0.62 \times 10^7 erg/cm^2 \qquad (E_g \simeq 1.2 \text{ Joule}/cm^2)$$

$$\alpha_L^o \simeq 1.6 \times 10^{23} sec^{-1} \qquad (\bar{\alpha} = 1.4 \times 10^{23} sec^{-2};\ \bar{\tau}_p \simeq 1.3 \times 10^{-12} sec)$$

$$\bar{\varepsilon} \simeq 8.7 \times 10^4 erg/cm^2 \quad .$$

The Kerr cell for this example would satisfy the requirement:

$$\frac{n_2 L_K}{4\pi} = \frac{\vec{\chi}(1+C_o)}{n_o} L_K \simeq 5.7 \times 10^{-13} \quad ,$$

which is not a severe restriction. We will take the following parameters

for the gain medium and cavity: $L_c = 10^2$cm, $L_G = 10$ cm, N_o = number density of active molecules = 2×10^{19}cm^{-3} (i.e. dilute Ruby), $\mathcal{A}$ = 0.5 cm^2. Setting $(N_t - N_g) = \frac{2E_g}{\hbar\omega_t L_G}$ and $N_t + N_g = N_o$, in equation II.3.10, we find $I_{ave} \simeq 1.2 \times 10^5$erg sec^{-1}cm^{-2}; a more accurate calculation, including degeneracy effects and the presence of other levels[63] gives $I_{ave} \simeq 1.3 \times 10^5$erg sec^{-1}cm^{-2}, an insignificant change. Using the latter value for I_{ave}, we find

$$(\varepsilon_o/\bar{\varepsilon})_{min} \simeq 1.9 \times 10^{-12}$$

$$(\varepsilon_o/\bar{\varepsilon})_{max} \simeq 1.0 \times 10^{-8} \quad .$$

In Figure 13, we show the Q-switched pulse train for the same "low gain", (G=1.2), case considered in Figure 10. Again, the ineffectiveness of the cubic nonlinearity is apparent, since the curves for the pulse energies are essentially unchanged when the self modulation is neglected. The pulsewidths and chirps, as determined by α_n and β_n, are affected to a larger degree, but their response "lags" behind ε_n, so that the main pulses are not strongly chirped or sharpened.

The high gain (G=5) Q-switched pulse train is depicted in Figures 14 and 15, which show several important features. First, when no Kerr effect is present, the entire train lasts for only a few, very intense, pulses; the Kerr nonlinearity has had the useful effect of lengthening the train to $\sim$ 100 pulses of lower, but more uniform, intensity. However, the undesirable "overshoot", discussed in

connection with Figure 11, still occurs; these overshoot pulses contain about 40% of the total energy in the train, with another 50% occurring in the first 90 pulses of the quasi-steady-state "tail". Furthermore, because of the gain depletion in the overshoot, the "useful" gain for the "train" is given approximately by $G(30) = 2.1$; this decreases, but does not eliminate, the stability of the quasi-steady-state pulsing. Despite the "overshoot", the major benefits of the laser system remain, since the pulses in the "tail" are entirely reproducible, and have uniform enough characteristics to allow the pulse compression shown.

In Figures 16 and 17, we have repeated the high gain calculations with the addition of the "pulse sharpener" discussed in the last section. As expected, the "overshoot" and its gain depletion have been significantly reduced, leading to stronger pulses in the quasi-steady-state "tail". Here, much more of the total train energy is found in the useful "tail" pulses, which also have the higher compression factor shown.

These results indicate that the self phase modulation in the Kerr medium has led, as desired, to long Q-switched trains of ultrashort pulses, with stable and reproducible characteristics. By comparing these results with those of the normal (low gain) laser system, mode-locked with a saturable absorber, we see that the relative advantages of our system are a consequence of the "strong" role played by the nonlinearity; in fact, the entire laser behavior in our case is "dominated" by the stability of the "steady-state pulsing" configuration. We therefore expect that the results found here, although based on a simple mathematical model, will apply semi-quantitatively

to real laser systems.

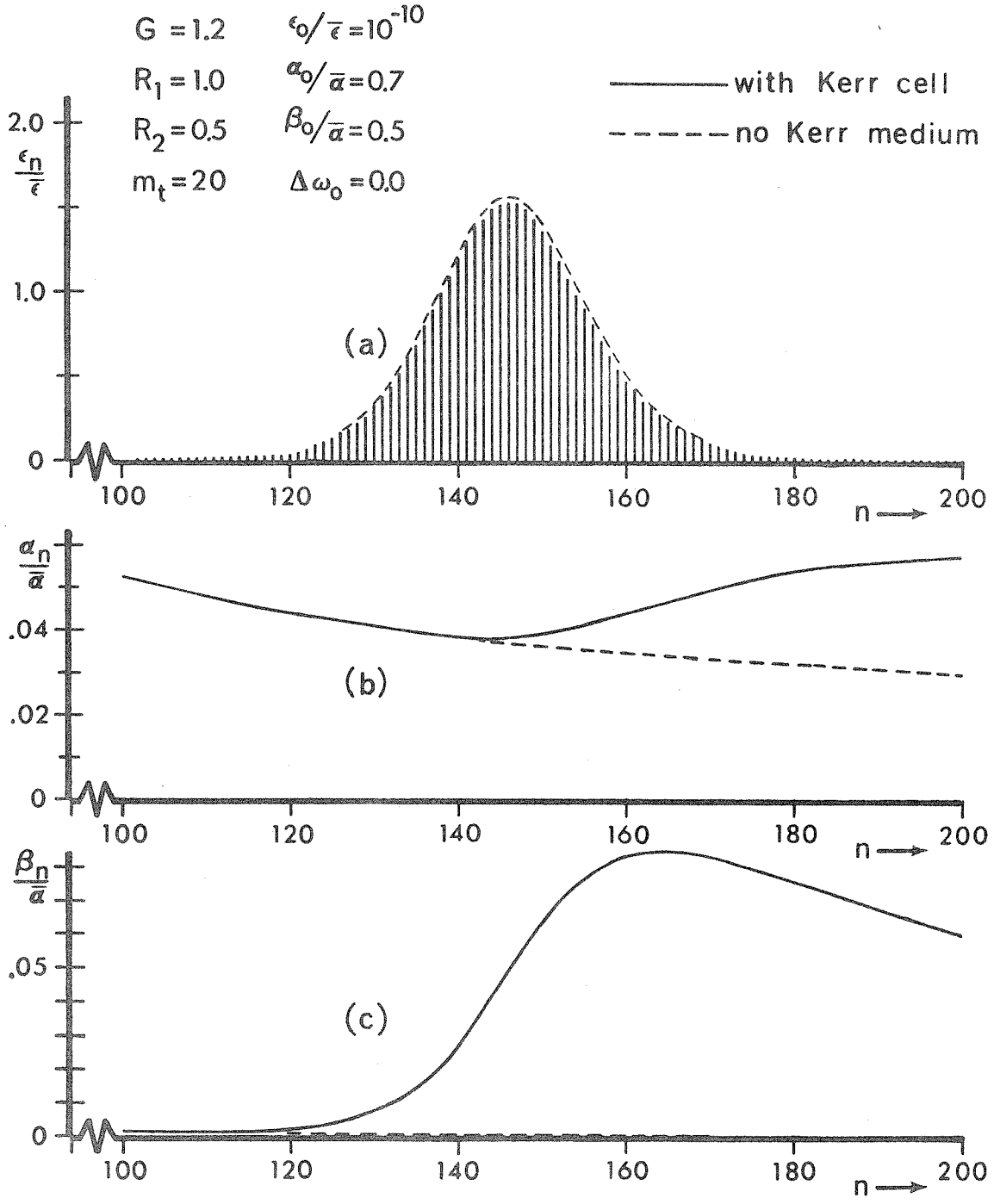


Figure 13. Q-switched pulse train from low gain laser.
(a) Relative pulse energy, ϵ_n, vs. pulse number, n.
(b) Pulsewidth parameter, $\alpha_n/\bar{\alpha} = (\bar{\tau}_p/\tau)^2$.
(c) Relative chirp of pulses.
The dashed curves result when no Kerr medium is present.

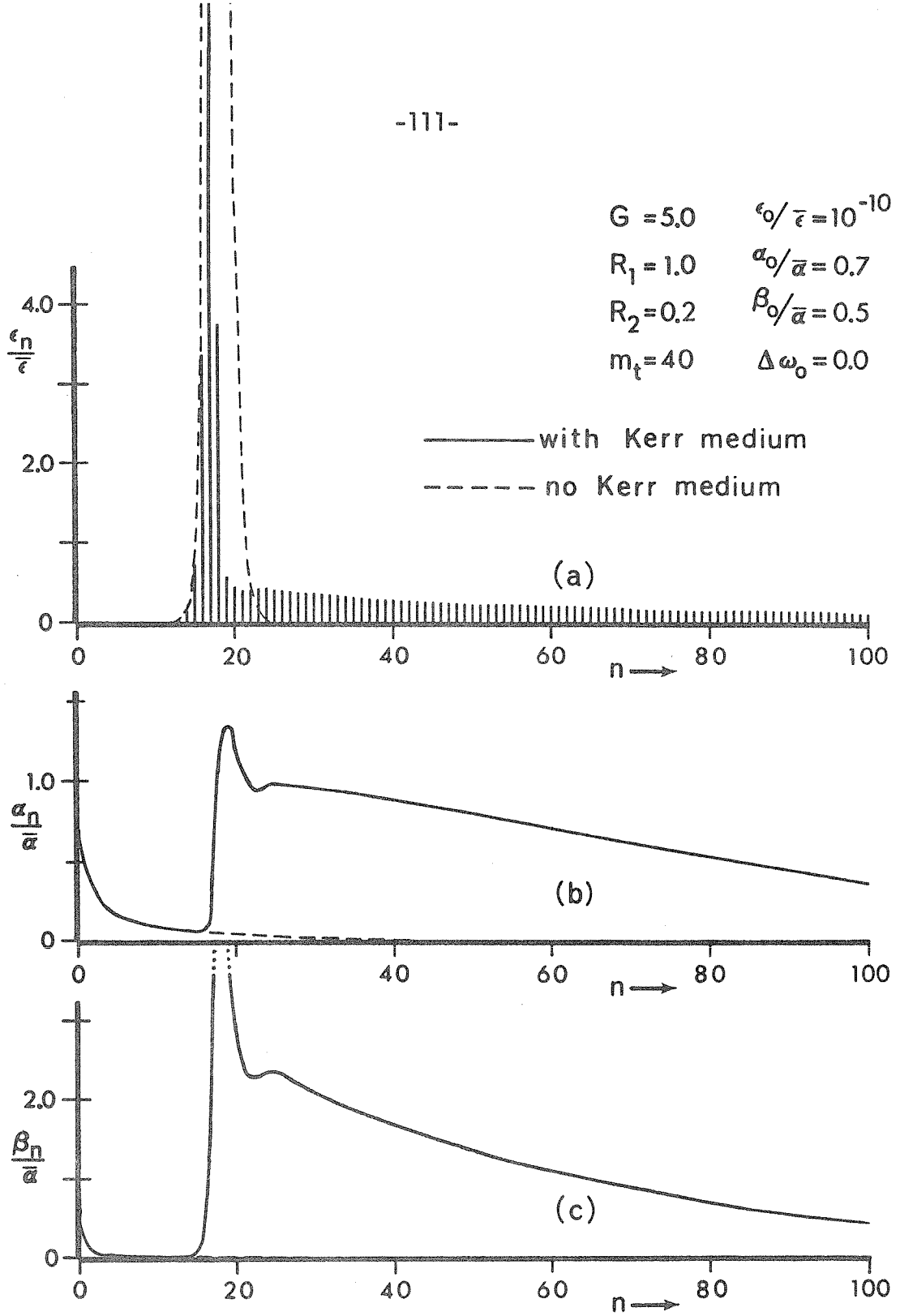


Figure 14. Q-switched pulse train from high gain laser.
(a) Relative pulse energy, ϵ_n, vs. pulse number, n.
(b) Pulsewidth parameter. $\alpha_n/\bar{\alpha} = (\bar{\tau}_p/\tau)^2$.
(c) Relative chirp of pulses.
The dashed curves result when no Kerr medium is present.

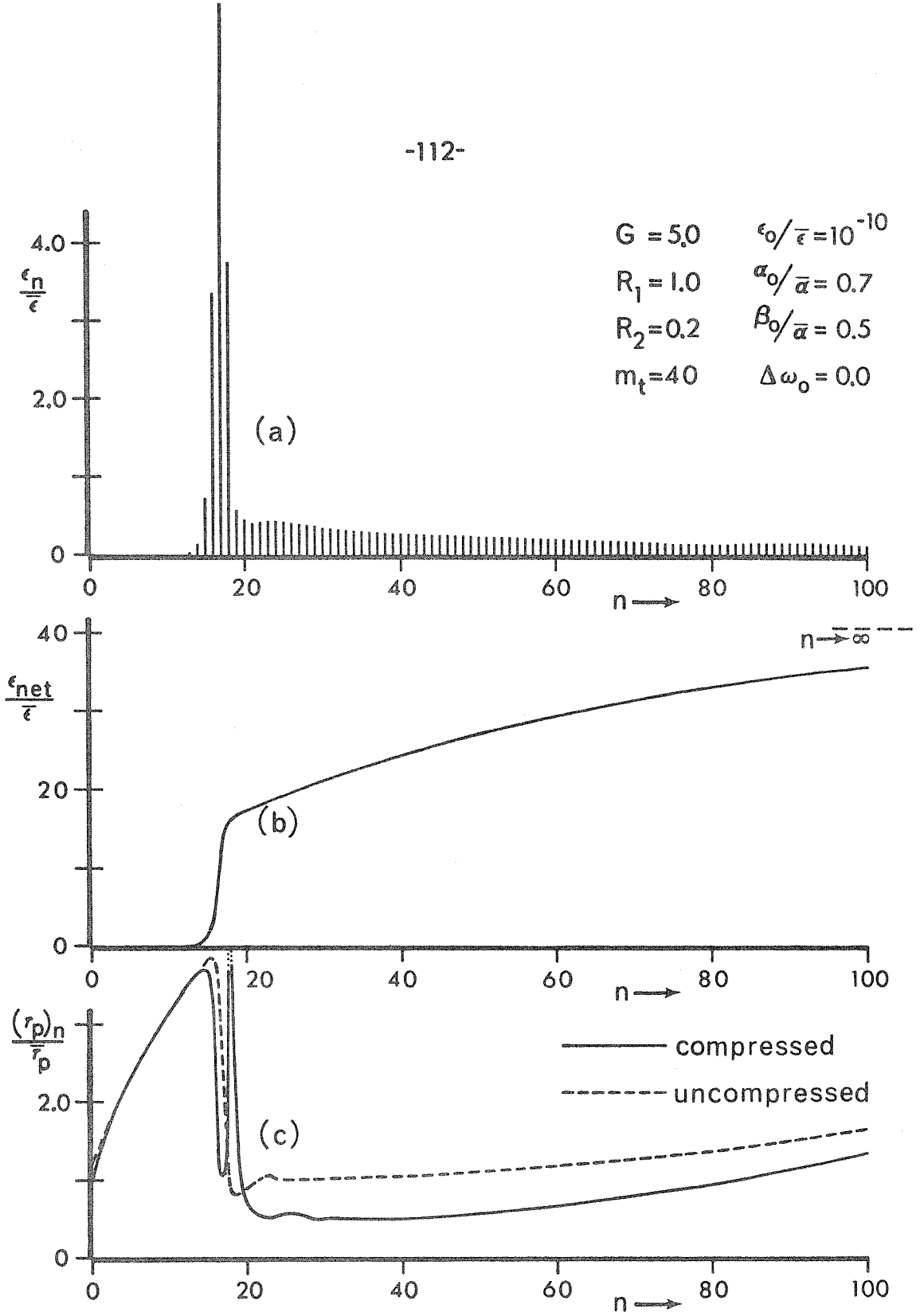


Figure 15. High gain-Q-switched laser, mode-locked with a Kerr-type nonlinearity; further results.
(a) Pulse energies; Figure 14a repeated for reference.
(b) Integrated laser output.
(c) Pulsewidths of uncompressed (dashed) and compressed train, using $g_{comp}(\omega) = \exp\{-i(\omega-\omega_t)^2(0.1275/\bar{\alpha})\}$.

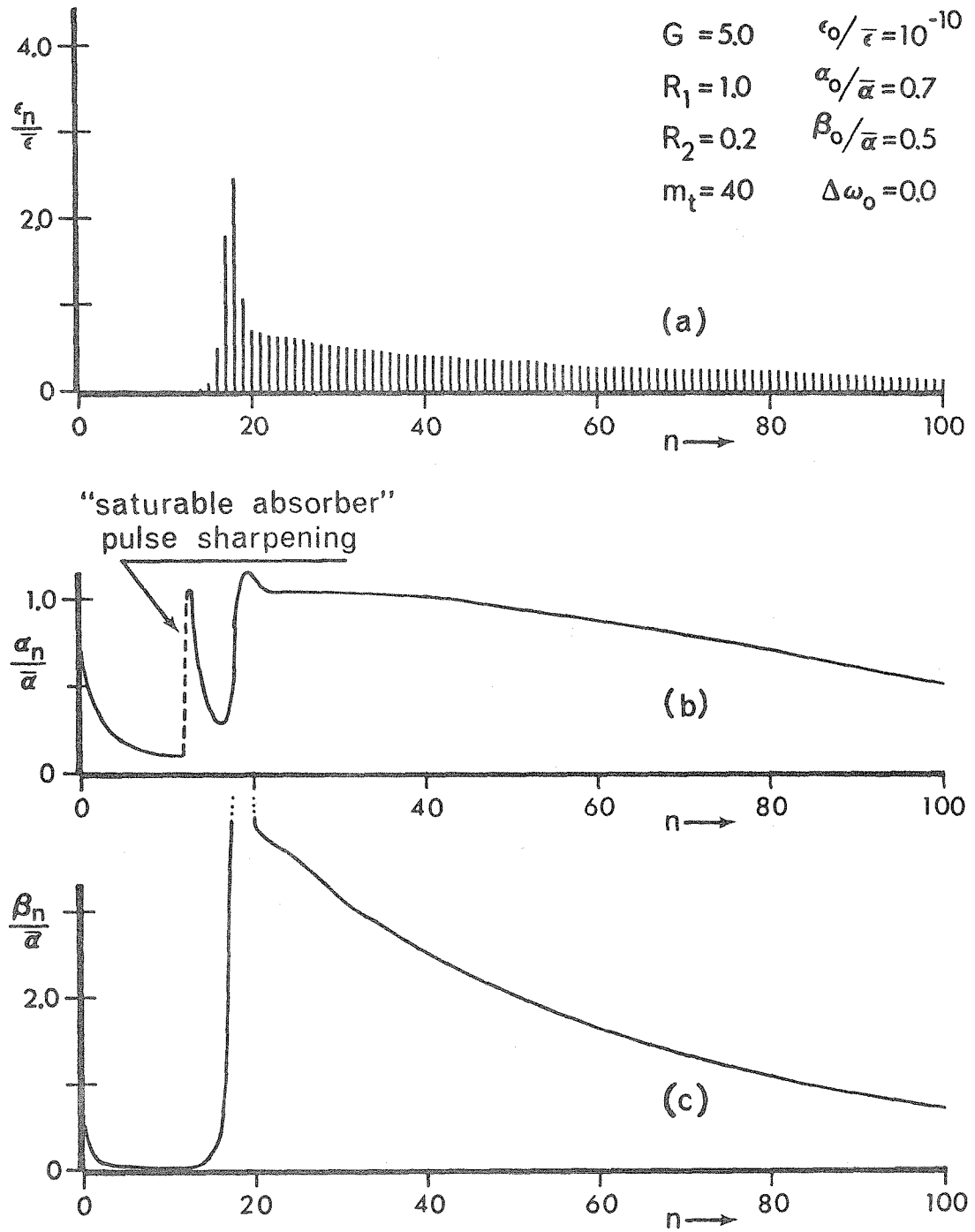


Figure 16. High gain, Q-switched laser, mode-locked with a Kerr-type nonlinearity and including a "pulse sharpening" mechanism.
(a) Relative pulse energies, ϵ_n, vs. pulse number, n.
(b) Pulsewidth parameter. $\alpha_n/\bar{\alpha} = (\tau_p/\tau)^2$. The idealized pulse sharpening occurs at pulse no. 13.
(c) Relative chirp of pulses.

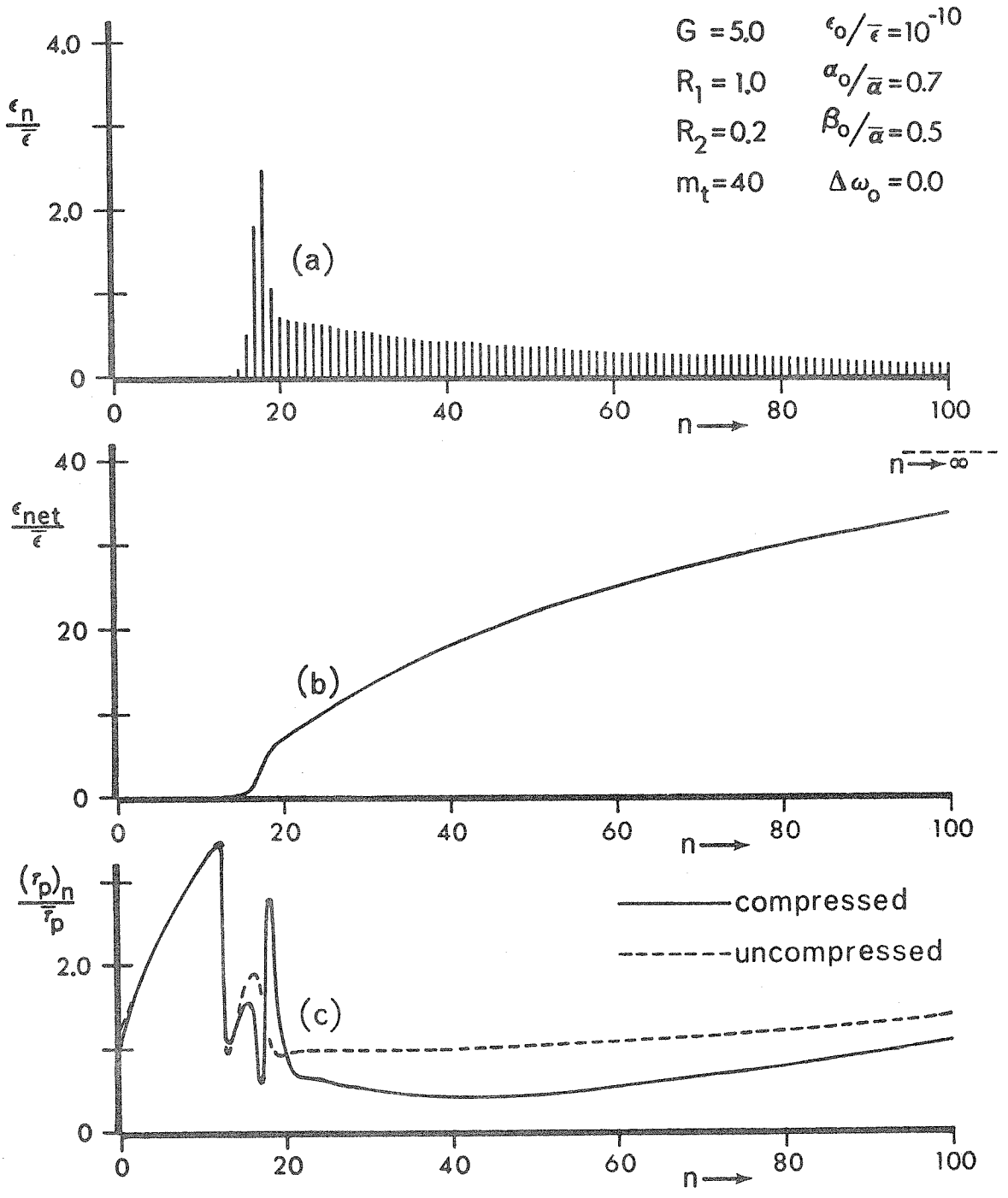


Figure 17. High gain, Q-switched laser system, including a Kerr-type nonlinearity and "pulse sharpener": further results.
(a) Pulse energies; Figure 16a repeated for reference.
(b) Integrated laser output.
(c) Pulsewidths of uncompressed (dashed) and compressed train, using $g_{comp}(\omega) = \exp\{-i(\omega-\omega_t)^2(0.1/\bar{\alpha})\}$.

Chapter III

STEADY-STATE SELF-FOCUSING OF NON-AXISYMMETRIC BEAMS

Our previous considerations of the propagation of intense pulses in a medium exhibiting a cubic polarization have led to a description of the phenomenon of self phase modulation; these analyses were based on the assumption that the beam could be approximated with a plane wave. We now consider the propagation of a beam with a finite cross-section; we will find that the cubic polarization produces distortions in the transverse structure which, for a sufficiently powerful beam, can ultimately lead to the phenomenon of "self-focusing".

Because of the highly nonlinear character of the problem, there are many aspects of self-focusing which are not clearly understood at this time, despite intensive experimental and theoretical efforts. Our analysis here will be limited to showing how the focusing process depends on the shape of the initial beam cross-section; we will consider primarily the steady-state case, with the simplest form for the nonlinearity. Even with these simplifications, however, the three dimensional problem is extremely complex; we will thus be content to consider a number of analytical approaches, each of which provides only a partial solution.

III.1 Description of Self-focusing Problem

In Sec. I.3, we showed that a cubic polarizability manifests itself physically, to lowest order, as an intensity-dependent refractive index. As shown in Fig. 18, this view explains the tendency of a beam to focus

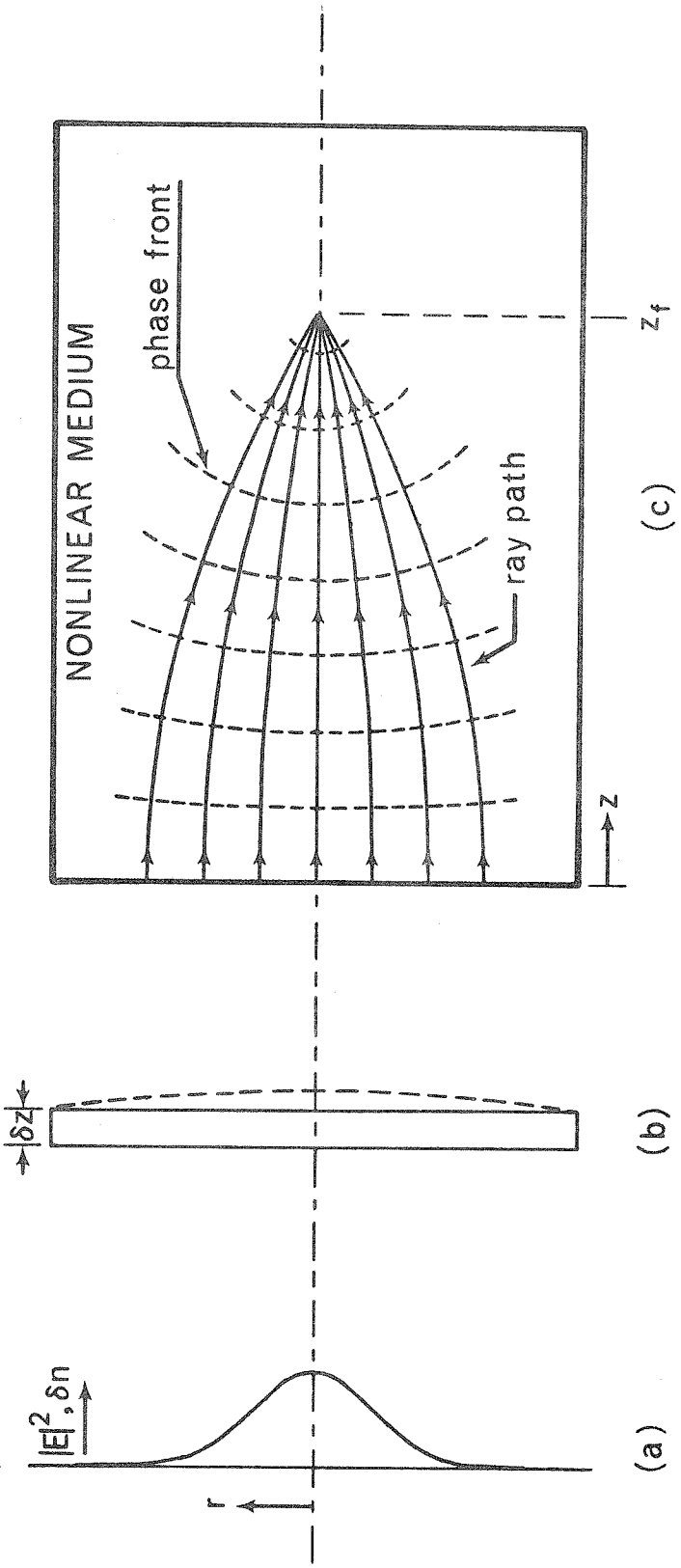


Figure 18. Physical model of self-focusing process.
(a) Plot of incident beam intensity vs. distance from the axis. The increase in the refractive index in the nonlinear material is proportional to $|E|^2$, as shown.
(b) Incremental slab of nonlinear material. The higher index near the beam axis causes the optical length (i.e. phase delay of the field) to be larger at the center, so that the propagating beam sees an effective lens (dotted).
(c) Qualitative example of light rays and phase fronts during self-focusing.

inside such a nonlinear medium: the intense center of the beam increases the refractive index there, so that the material acts as a convergent lens. Since any initial focusing will result in a stronger "lens", the process is unstable, resulting in a "collapse" near some "focus", at z_f. Furthermore, since the beam would tend to diffract in the absence of the nonlinearity, this physical model suggests that the nonlinear "focusing" will dominate only for a sufficiently strong beam. We shall show that this instability can be characterized by a "critical", or "threshold", power, P_{crit}, such that self-focusing will occur for beam powers, P, exceeding P_{crit}. In analyzing this effect, we will obtain P_{crit} and $z_f(P)$, and, in particular, we will find how these parameters depend on the initial shape of the beam cross-section.

To further clarify the physical picture, and form an estimate of P_{crit}, we will consider a beam "at threshold", using the approach that the beam then forms its own "waveguide"[14]. As a first approximation, shown in Fig. 19, we will assume that the beam has a constant intensity, $|E_0|^2$, within the waveguide; the index discontinuity then can cause total internal reflection at the waveguide boundary, which balances the tendency toward diffraction. This reflection will occur if the diffraction angle, θ_{DIF}, shown in the figure, is smaller than the total internal reflection angle, θ_{TIR}, which is given by

$$\theta_{TIR} = \arccos\left(\frac{n_0}{n_0 + \delta n}\right) \approx \left(\frac{n_2|E_0|^2}{n_0}\right)^{\frac{1}{2}} .$$

Using standard Fraunhofer diffraction theory[64], we obtain

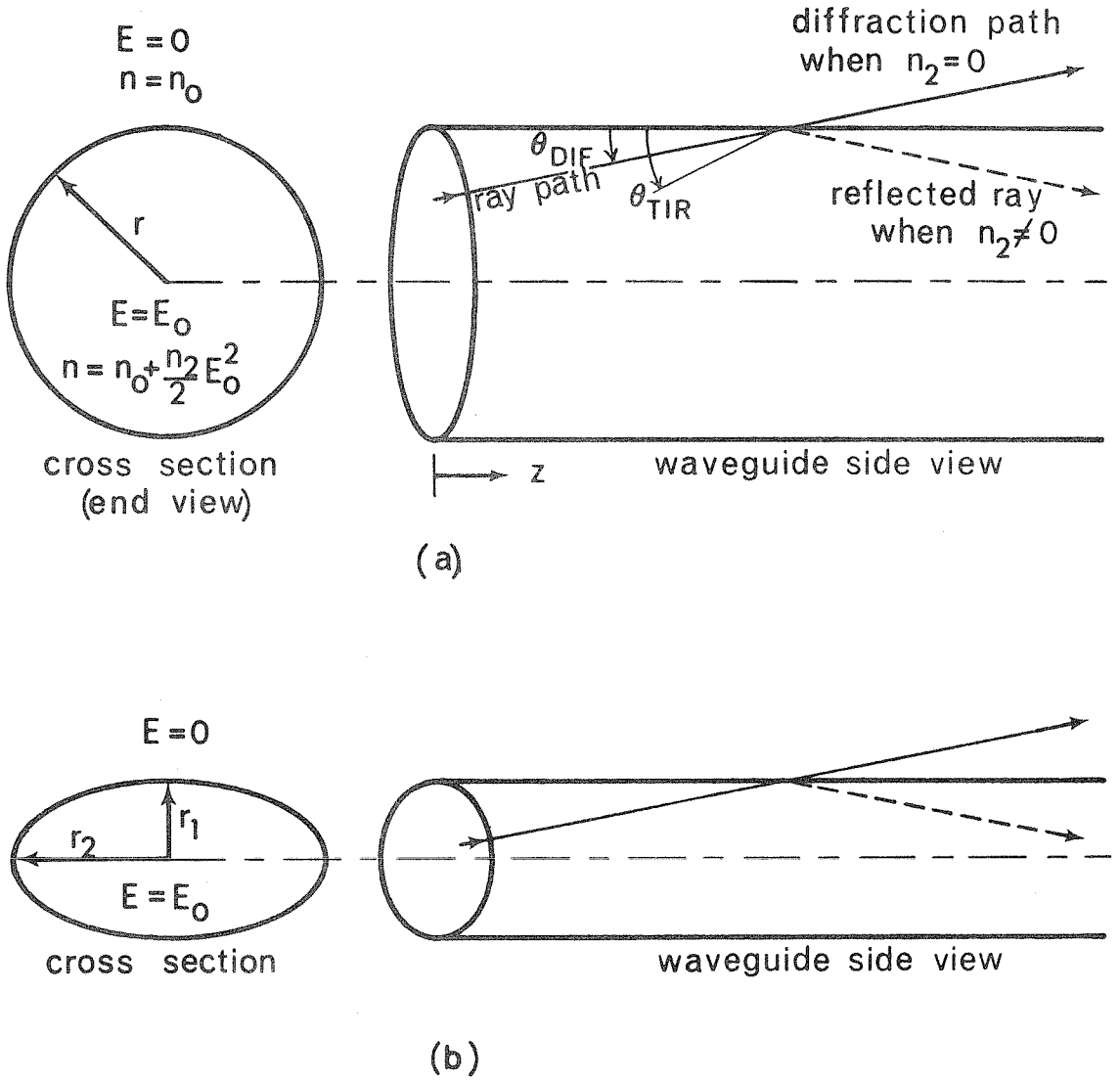


Figure 19. Approximate self-waveguiding model of beams with threshold powers, comparing cross-section shapes. a) Axisymmetric beam. A typical ray tends to diffract at angle θ_{DIF}, as shown, but is reflected back into the waveguide by the index discontinuity, when $\theta_{DIF} < \theta_{TIR}$, where θ_{TIR} is the angle of total internal reflection. b) Same effect as in (a), except that the beam is highly elliptical, with $r_2 > r_1$; only the diffraction across the smaller width is important.

θ_{DIF} as follows. For the axisymmetrical case, which forms the cylindrical waveguide shown in Fig. 19a, the diffraction is equivalent to that of a circular aperture of radius, r, so that

$$\theta_{DIF} \approx m_c \frac{\lambda}{r} \quad ,$$

where $m_c \simeq 0.61$ satisfies $J_1(2\pi m_c) = 0$. Similarly, for the highly elliptic cross-section shown in Fig. 19b, the strongest diffraction occurs across the narrow dimension, so that θ_{DIF} is essentially that of a slit of width, $2r_1$:

$$\theta_{DIF} \approx m_s \frac{\lambda}{r_1} \quad ,$$

where $m_s = 0.5$ satisfies $\sin(2\pi m_s) = 0$. For $\delta n << n_o$, the total power in these beams is given by

$$P = \frac{n_o c}{8\pi} |E_o|^2 a \quad ,$$

where a is the cross-section area. Self-focusing would occur if $\theta_{DIF} < \theta_{TIR}$, which implies the following critical powers for these cases:

$$P_{crit} = \pi^2 m_c^2 P_o = 3.67 P_o \qquad (R = 1) \quad ,$$

$$P_{crit} = \pi^2 m_s^2 \frac{r_2}{r_1} P_o = 2.47 R P_o \qquad (R >> 1) \quad , \qquad \text{(III.1.1)}$$

where

$$P_o \equiv \frac{c^3}{2n_2\omega_o^2} \quad , \qquad \text{(III.1.2a)}$$

and

$$R = \frac{r_2}{r_1} = \frac{\text{major ellipse axis}}{\text{minor ellipse axis}} \quad . \qquad \text{(III.1.2b)}$$

These approximations indicate that the highly elliptic beam shape has a threshold for self-focusing which is larger by the ratio of the major axis to the minor axis, R. The physical model of the "diffraction-focusing balance" shows that this will continue to hold, approximately, for more realistic beam profiles, although the specific numerical coefficients in Eq. III.1.1 will be modified.

As an example, for Ruby laser radiation ($\omega_o = 2.7166 \times 10^{15}$ sec^{-1}) passing through CS_2 ($n_2 \simeq 1.1 \times 10^{-11}$ e.s.u.), $P_0 = 1.7 \times 10^{11}$ e.s.u. = 17 kilowatts; this power level is easily surpassed in Q-switched lasers, even without mode-locking. Experimentally, the threshold for self-focusing of axisymmetric beams is indeed found to be roughly given by P_o [65], and the increase of P_{crit} with R has also been observed[21]. On the other hand, the model used above does not provide information about the beam behavior when $P > P_{crit}$, so that, for example, z_f and the beam shapes near the focal point are still unknown; such information is often crucial, since the field intensities near the focal point are intense enough to produce many interesting nonlinear effects.

The equations for beam propagation are given in Sec. I.3; since the "slowly varying envelope" approximation implies that $\left|\frac{\partial E}{\partial t}\right| << \omega_o |E|$, Eq. I.3.13 gives the following "first order" system:

$$\vec{\mathcal{E}}(x,y,z,t) = \frac{1}{2}\vec{e}_x\left(E(x,y,z,t)\, e^{i(K_o z-\omega_o t)} + c.c.\right)$$

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + 2ik_o\left(\frac{\partial E}{\partial z} + \frac{n_o}{c}\frac{\partial E}{\partial t}\right) = -\frac{2\omega_o{}^2 n_o}{c^2}(\delta n)E$$

$$\tau\frac{\partial(\delta n)}{\partial t} = \frac{n_2}{2}|E|^2 - (\delta n) \qquad \text{(III.1.3)}$$

where $k_o = \omega_o n_o/c$. The boundary conditions of interest are given by

$$E \longrightarrow o \qquad |x|,\ |y| \longrightarrow \infty$$

$$E(x,y,o,t) \text{ given} . \qquad \text{(III.1.4)}$$

The equations with which we will be concerned are given by the "steady-state limit", where $E = E(x,y,z)$. However, the more useful case of a pulsed beam can also be handled if the response of the nonlinearity can be considered "instantaneous", as will now be shown. For $\tau = 0$, we see that $(\delta n) = \frac{n_2}{2}|E|^2$; a change of variables can now be performed to show that

$$E(x,y,z,t) = E'(x,y,z;\ t' = t - \frac{n_o z}{c})$$

where $E'(x,y,z;\ t')$ is the solution of the equations:

$$\frac{\partial^2 E'}{\partial x^2} + \frac{\partial^2 E'}{\partial y^2} + 2ik_o\frac{\partial E'}{\partial z} = -\frac{\omega_o{}^2 n_o n_2}{c^2}|E'|^2 E'$$

$$E'(x,y,0;\ t') = E(x,y,0,t') \quad .$$

Thus, "t" enters only as a parameter indicating the delayed time at which the boundary condition is to be taken, and, mathematically, the propagation equation of interest has reduced to the "steady-state limit". Given the steady-state solution corresponding to every instantaneous value of $E(x,y,0,t)$, the above transformation yields the fields for the case of a time-varying input beam. This approach results in the concept of a "moving focus" for pulsed beams[66, 67, 68].

Our analysis, then, will involve the idealized, three dimensional, propagation of the complex field envelope, $E(x,y,z)$, governed mathematically by the equations:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + 2\, k_o i\, \frac{\partial E}{\partial z} + \frac{\omega_o^2 n_o n_2}{c^2} |E|^2 E = 0$$

$$E(x,y,o) \text{ given}$$

$$E \longrightarrow 0 \qquad \text{for} \qquad |x|\ ,\ |y| \longrightarrow \infty \qquad . \qquad \text{(III.1.5)}$$

The simplest problem based on these equations involves an axisymmetric beam, $E = E(r,z)$ where $r = (x^2 + y^2)^{\frac{1}{2}}$; this two dimensional problem has been investigated in some detail through analytic and numerical methods[69,70]. Values for P_{crit} have been given, and it has been found that beams with $P > P_{crit}$ will self-focus to a mathematical singularity at z_f. However, the "ultimate" goal of accurately describing self-focusing under realistic conditions requires the introduction of at least three types of "higher order" effects. First, the nonlinearity, represented by the last term in the propagation equation, III.1.5, must be altered to include saturation and higher order

polarizations, dispersion and absorption (linear and nonlinear), coupled polarizations, nonlinear scattering losses, and mutual interactions between a number of coherent beams (i.e. Raman or Brillouin components); often, bulk material changes (damage, etc.) can occur near z_f. Second, the steady-state limit will not apply because the powers needed for self-focusing require pulsed beams; for nonzero τ , or when the higher order time derivative terms in the wave equation are included (see Eq. I.3.6), the addition of the variable, t, vastly increases the analytical difficulty. Third, the lack of cylindrical symmetry adds another independent variable, which further complicates the analysis and its interpretation, as we will show.

Because of the fundamental role played by the nonlinearity present in Eq. III.1.5, the above three "complications" will be interrelated; also, the addition of a new effect usually cannot be handled by perturbation methods. As an example, it has been found, through numerical solutions, that altering the form of the nonlinearity, even slightly, leads to quantitatively and qualitatively new phenomena, such as the appearance of multiple focal points[71,72]. For similar reasons, the "moving focal point" theory must be considered to represent only an idealized first-order solution. It appears that many of the uncertainties in the present understanding of symmetrical self-focusing will only be resolved when detailed solutions in (r, z, t) space are available. A few of these analyses have been performed for the simplest cases[73, 74], but the lack of accurate solutions for realistically "complicated" media is an indication of the computational difficulties

involved. In view of these considerations, our analyses of the simplest nonlinearity in (x, y, z) space will represent only the "first step" toward obtaining "practical" results, and will certainly not be physically valid near the focal point; nevertheless, we will be able to predict such "gross" features as P_{crit} and z_f, which are determined primarily by the propagation behavior away from the focal point, and we will be able to suggest some analytical methods which can potentially be extended to realistic problems.

In order to determine the appropriate boundary condition, E(x,y,o), we will consider first the linear propagation problem:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + 2 k_o i \frac{\partial E}{\partial z} = 0$$

$$E \longrightarrow 0 \qquad \text{for} \qquad |x|, |y| \longrightarrow \infty \; . \qquad \text{(III.1.6)}$$

A complete set of solutions can be found in terms of the standard "Gaussian modes":

$$E_{m,n}(x,y,z) = E_o \, e^{i\phi_o} \, e_m(x,z;\, w_x, z_x) \, e_n(y,z;\, w_y, z_y)$$

$$e_m(x,z;\, w_x,z_x) = \left[1 + \left(\frac{2(z-z_x)}{k_o w_x^2}\right)^2\right]^{-\frac{1}{4}} \cdot H_m\left(\sqrt{2}\, x/w_x \left[1 + \left(\frac{2(z-z_x)}{k_o w_x^2}\right)^2\right]^{-\frac{1}{2}}\right)$$

$$\cdot \exp\left[-\frac{x^2}{w_x^2}\left(1 + i\,\frac{2(z-z_x)}{k_o w_x^2}\right)^{-1} - i\left(m + \tfrac{1}{2}\right) \cdot \right.$$

$$\left. \text{arc tan}\left(\frac{2(z-z_x)}{k_o w_x^2}\right)\right] \qquad . \qquad \text{(III.1.7)}$$

Here $H_m(x)$ is the Hermite polynomial; w_x is the "x waist" (i.e. the minimum width in the 'x' transverse direction), z_x is the position where this "x waist" is achieved, and w_y and z_y are similarly defined.

In ideal cases, laser output beams contain primarily the $E_{o,o}$ mode; for computational purposes, such an "incident beam" will be assumed here, so that, by ignoring overall phase factors, we will take

$$E(x,y,0) = E_o e^{-\frac{x^2}{x_o^2}(1 + i\theta_x)} \; e^{-\frac{y^2}{y_o^2}(1 + i\theta_y)} \qquad \text{(III.1.8)}$$

Here, the "x waist" is given by $w_x = x_o(1 + \theta_x^2)^{-\frac{1}{2}}$ and the position of the waist by

$$z_x = \frac{k_o x_o^2}{2} \frac{\theta_x}{1 + \theta_x^2} \quad ,$$

in the absence of the nonlinearity; if $z_x = z_y$, then this point would be the well-known "focal point". Equation III.1.5, with the initial condition in III.1.8, thus constitutes the beam-focusing problem to be solved.

The power in a beam with the form given in Eq. III .1.8 is given by

$$P = \frac{n_o c}{8\pi} \iint |E(x,y,z)|^2 \, dx \, dy = \frac{n_o c}{16} E_o^2 \, x_o y_o \quad . \qquad \text{(III.1.9)}$$

Although this integral has been carried out for the field at z = 0, the propagation Eq. III .1.5 can be used to show that this power will be conserved. We will now write the propagation problem in

"dimensionless" form, using the following normalized variables and functions:

$$x' = \frac{x}{x_o}$$

$$y' = \frac{y}{x_o}$$

$$z' = \frac{z}{2k_o x_o^2}$$

$$E'(x',y',z') = E(x,y,z)/E_o \quad . \qquad \text{(III.1.10)}$$

Our basic problem now takes the (deceptively) simple form:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + i\,\frac{\partial E}{\partial z} + \beta|E|^2 E = 0 \quad , \qquad \text{(III.1.11a)}$$

$$E(x,y,o) = e^{-x^2(1 + i\theta_x)}\, e^{-y^2(1 + i\theta_y)/R^2} \quad , \qquad \text{(III.1.11b)}$$

where β, R, θ_x and θ_y now define the problem parameters, and

$$\beta = \frac{\omega_o^2 n_o n_2 E_o^2 x_o^2}{c^2} = \frac{8}{R}\frac{P}{P_o} \quad , \qquad \text{(III.1.12a)}$$

$$R = y_o/x_o = \text{ellipticity ratio for incident beam} \quad . \qquad \text{(III.1.12b)}$$

P_o is defined in Eq. III.1.2. We have dropped the primes in Eq. III.1.11, since we will usually deal with the dimensionless problem in the remainder of this chapter and the next. Once the dimensionless

solution $E(x,y,z)$ has been found, the physical field envelope will be given by

$$E_{true}(x,y,z) = E_o E\left(\frac{x}{x_o} , \frac{y}{x_o} , \frac{z}{2k_o x_o^2} \right) \quad . \qquad \text{(III.1.13a)}$$

As a measure of the "success" of the various analyses, we will attempt to calculate the critical self-focusing power, P_{crit}, and, for $P > P_{crit}$, the position of the "collapse", $(z_f)_{true}$. In our dimensionless system, these will appear as $\beta_{crit}(R,\theta_x,\theta_y)$ and $z_f(\beta, R,\theta_x,\theta_y)$, with the physical values given by:

$$P_{crit} = \frac{R}{8} \beta_{crit} P_o \quad , \qquad \text{(III.1.13b)}$$

and

$$(Z_f)_{true} = 2 k_o x_o^2 z_f \quad . \qquad \text{(III.1.13c)}$$

It is possible to normalize the variables so that there is no free parameter in the propagation equation ($\beta \longrightarrow 1$); this requires another (scaling) parameter in the initial conditions. We will not do this, usually, because we prefer to have the free-beam limit ($\beta \longrightarrow o$) easily found for comparison, and because the presence of the parameter, 'β', is helpful in recognizing and "ordering" the terms arising from the nonlinearity.

III.2 General Propagation Properties

There are a number of properties common to all solutions of Eq. III.1.11a, which we will find useful in later analyses. It has already been observed, for example, that the propagating beam power is conserved; for our specific "Gaussian" initial conditions, in Eq. III.1.11b, we have, as a "dimensionless" power,

$$P(z) \equiv \iint_{-\infty}^{\infty} E(x,y,z)\, E^{*}(x,y,z)\, dxdy = P(o) = \frac{\pi}{2} R \quad . \tag{III.2.1}$$

It can also be seen that any symmetries present in the initial condition at $z = 0$ will apply for $z > 0$; thus, an axisymmetrical incident beam will retain the cylindrical symmetry. In our case $E(x,y,z)$ will always be an even function of 'x' and 'y', so that, as an example, a solution need only be considered for $x,y \geq 0$.

III.2.i lens transformation

Another general property of the propagation is described as follows: if $E(x,y,z)$ is any solution of Eq. III.1.11a, then the following function will also satisfy that equation

$$\hat{E}(x,y,z) = \hat{A}(z)\, E(\hat{x}(x,y,z),\, \hat{y}(x,y,z),\, \hat{z}(z))\, e^{-i\hat{\phi}(x,y,z)} \tag{III.2.2a}$$

where

$$\hat{A}(z) = \frac{a}{(z_o - z)}$$

$$\hat{x}(x,y,z) = x_2 + \frac{a}{(z_o - z)}[(x - x_1)\cos\theta - (y-y_1)\sin\theta]$$

$$\hat{y}(x,y,z) = y_2 \pm \frac{a}{(z_o - z)}[(x-x_1)\sin\theta + (y-y_1)\cos\theta]$$

$$\hat{z}(z) = z_2 + \frac{a^2}{(z_o-z)}$$

$$\hat{\phi}(x,y,z) = \phi_o + \frac{(x-x_1)^2 + (y-y_1)^2}{4(z_o-z)} \quad . \qquad \text{(III.2.2b)}$$

Here $a,\theta,x_1,y_1,z_o,\phi_o,x_2,y_2,z_2$ are any real constants. These functions are the most general possible form for $\hat{A},\hat{x},\hat{y},\hat{z},\hat{\phi}$, given a transformation of the type in Eq. III.2.2a.

An examination of the transformation above shows that it can be separated into a number of distinct "components", many of which are trivial:

(1) transverse translation of origin: (x_1,y_1,x_2,y_2)

(2) rotation and inversion in transverse plane: (θ,'-' sign in $\hat{y}$)

(3) constant phase shift: (ϕ_o)

(4) transverse scaling, with power conserved: (a)

(5) lens transformation: (z_o,z_2)

The first three are inconsequential for our purposes; we will take $x_1,y_1,x_2,y_2,\theta,\phi_o$ equal to zero. Transformation (4) has already been accounted for, by normalizing $E(o,o,o) = 1$ in our dimensionless

problem derivation; as expected, this shows that the absolute beam size, x_o , does not affect the self-focusing process, except to change the length scale according to Eq. III.1.13c. To conserve the normalization ($\hat{E}(o,o,o) = E(o,o,o)$) and the initial condition location ($\hat{z}(o) = 0$) we use $a = z_o$ and $z_2 = -z_o$ and obtain the solution generated by the "lens transformation"[75]:

$$\hat{E}(x,y,z) = \left(\frac{z_o}{z_o-z}\right) E\left(\frac{xz_o}{z_o-z}, \frac{yz_o}{z_o-z}, \frac{zz_o}{z_o-z}\right) e^{-i\frac{(x^2+y^2)}{4(z_o-z)}} \qquad \text{(III.2.3a)}$$

with the initial condition replaced by

$$\hat{E}(x,y,o) = E(x,y,o)\, e^{-i\frac{(x^2+y^2)}{4z_o}} \quad . \qquad \text{(III.2.3b)}$$

To see why this represents a "lens" transformation, we can consider an incident beam at $z = 0$ which is "focused" at $z = z_f$ (in the geometrical sense); the transformed field, $\hat{E}$, will focus at z_f' , where $\frac{z_f' z_o}{z_o - z_f'} = z_f$. Thus $\frac{1}{z_f'} = \frac{1}{z_f} + \frac{1}{z_o}$, which is the same result obtained by placing a lens of focal length $+z_o$ just in front of the beam at $z = 0_-$; the phase shift which is quadratic in r , as shown in Eq. III.2.3b, is also consistent with a "lens" at $z = 0_-$. (In physical units, the focal length of the lens under consideration would be $(z_o)_{true} = 2\, k_o x_o^2 z_o$). Note that the presence of the nonlinearity does not affect the lens transformation.

Mathematically, the effect of applying Eq. III.2.3 is to reduce

or expand the propagation distance, z. For example, the original problem defined by Eq. III.1.11 involved a solution over the range, z: (o,∞); by simply changing the initial conditions according to Eq. III.2.3b, which for the Gaussian function in Eq. III.1.11b, implies

$$\theta_x \longrightarrow \theta_x' = \theta_x + \frac{1}{4z_o}$$

$$\theta_y \longrightarrow \theta_y' = \theta_y + \frac{R^2}{4z_o} \quad , \qquad \text{(III.2.4)}$$

we will generate a solution for $\hat{E}(x,y,z)$ on the range, z: (o,z_o) which corresponds to E on (o,∞). The remaining solution for $\hat{E}$ in the range (z_o,∞) will not have any physical significance in the original problem.

III.2.ii solutions for beam moments

We will now consider a very powerful "integral" property of the propagation equation, as first reported by Vlasov, et al[76]. These results will be concerned with the evolution, in z, of the beam "moments", such as the "power" (r^o), beam center ($\vec{r}$), and mean square radius (r^2); here,

$$\vec{r} \equiv (x,y) = \vec{e}_x\, x + \vec{e}_y\, y \quad , \qquad \text{(III.2.5a)}$$

is the two dimensional transverse position vector. The following analysis, based on that of Suydam[77], appears indirect, but the overall algebra is simplified (and generalizable to higher dimensions, if desired).

By defining the usual vector differential operator in two dimensions,

$$\vec{\nabla} = \left(\frac{\partial}{\partial x} , \frac{\partial}{\partial y} \right) \equiv \vec{e}_x \frac{\partial}{\partial x} + \vec{e}_y \frac{\partial}{\partial y} \quad , \qquad \text{(III.2.5b)}$$

we can write the field propagation equation, III.1.11a, in the form

$$\frac{\partial E}{\partial z} = i \{ \nabla^2 E + \beta |E|^2 E \} \quad , \qquad \text{(III.2.6)}$$

where $\nabla^2 E \equiv \vec{\nabla} \cdot (\vec{\nabla} E)$ according to the standard notation. From this equation, the following can easily be found to hold:

$$\frac{\partial W}{\partial z} = - \vec{\nabla} \cdot \vec{S} \quad , \qquad \text{(III.2.7a)}$$

where

$$W \equiv |E|^2 \quad , \qquad \text{(III.2.7b)}$$

$$\vec{S} \equiv i \{ E(\vec{\nabla} E^*) - E^* (\vec{\nabla} E) \} \quad . \qquad \text{(III.2.7c)}$$

Next, we can define a tensor form, $\vec{\vec{T}}$, having four components; by considering an analogy with matrices, we will define

$$\vec{\vec{T}} = \vec{A}\,\vec{B} \quad : \quad T_{ij} = A_i B_j$$

$$\vec{C} = \vec{D} \cdot \vec{\vec{T}} \quad : \quad C_i = \sum_j D_j T_{ji}$$

$$\vec{F} = \vec{\vec{T}} \cdot \vec{G} \quad : \quad F_i = \sum_j T_{ij} G_j$$

where $\vec{C}$, $\vec{D}$ are "row" vectors and $\vec{F}$, $\vec{G}$ are "column" vectors. With these conventions, we find that the following holds:

$$\frac{\partial \vec{S}}{\partial z} = \vec{\nabla} \cdot \vec{\vec{T}} \quad , \qquad \text{(III.2.8a)}$$

where

$$\vec{\vec{T}} = -2(\vec{\nabla}E)(\vec{\nabla}E^*) - 2(\vec{\nabla}E^*)(\vec{\nabla}E) + \vec{\vec{I}}\{\nabla^2(|E|^2) + \beta(|E|^2)^2\} \; . \qquad \text{(III.2.8b)}$$

Here, $\vec{\vec{I}}$ is the unit tensor ($I_{ij} = \delta_{ij}$). Finally, defining $\langle T \rangle \equiv \sum_i T_{ii}$, the trace of the tensor $\vec{\vec{T}}$, we obtain

$$\langle T \rangle = 2\{E\nabla^2 E^* + E^*\nabla^2 E + \beta(|E|^2)^2\} \quad , \qquad \text{(III.2.9a)}$$

$$\frac{\partial \langle T \rangle}{\partial z} = \vec{\nabla} \cdot \vec{Q} \quad , \qquad \text{(III.2.9b)}$$

where

$$\vec{Q} = 2i\{E^*(\vec{\nabla}(\nabla^2 E)) - E(\vec{\nabla}(\nabla^2 E^*)) + (\nabla^2 E^*)\vec{\nabla}E - (\nabla^2 E)\vec{\nabla}E^*\} + 4i\beta|E|^2\{E^*(\vec{\nabla}E) - E(\vec{\nabla}E^*)\} \; . \qquad \text{(III.2.9c)}$$

Note that w, $\vec{S}$, $\vec{\vec{T}}$ have analogies with the usual vacuum electromagnetic energy density, power flow, and stress tensor, as seen in Eqs. III.2.7a and III.2.8a. We now assume that we have a beam "confined" in a cross-sectional region, "a", bounded by a curve "Γ"; then Gauss' law, in two dimensions, shows that

$$\iint_a da(\vec{\nabla} \cdot \vec{C}) = \oint_\Gamma d\gamma(\vec{n} \cdot \vec{C})$$

$$\iint_a da(\vec{\nabla} \cdot \vec{\vec{T}}) = \oint_\Gamma d\gamma(\vec{n} \cdot \vec{\vec{T}}) \; .$$

In the present case, $da = dxdy$, and Γ is the curve at "infinity", where all functions approach zero; thus, all line integrals on Γ will vanish.

Defining the "0^{th} moment", P, as in Eq. III.2.1, according to

$$P(z) = \iint da\ w \quad , \qquad \text{(III.2.10a)}$$

we have

$$\frac{dP}{dz} = \frac{d}{dz}\iint da\ w = \iint da\ \frac{\partial w}{\partial z} = -\iint da(\vec{\nabla}\cdot\vec{S}) = -\oint d\gamma(\vec{n}\cdot\vec{S})$$

$$= 0 \quad , \qquad \text{(III.2.10b)}$$

which implies, as expected, that the beam power is conserved:

$$P(z) = P(o) \quad . \qquad \text{(III.2.10c)}$$

Next, consider the "1^{st} beam moment", $\vec{R}(z)$, which gives the "center of gravity" of the beam

$$\vec{R}(z) \equiv \frac{1}{P(z)}\iint da\ \vec{r}w \quad . \qquad \text{(III.2.11a)}$$

We have

$$\frac{d^2\vec{R}}{dz^2} = \frac{d}{dz}\left(-\frac{1}{P(o)}\iint da\ \vec{r}(\vec{\nabla}\cdot\vec{S})\right) = \frac{d}{dz}\left(-\frac{1}{P(o)}\iint da\{\vec{\nabla}\cdot(\vec{S}\vec{r})-\vec{S}\cdot(\vec{\nabla}\vec{r})\}\right)$$

$$= \frac{d}{dz}\left(\frac{1}{P(o)}\iint da\ \vec{S}\right) = \frac{1}{P(o)}\iint da(\vec{\nabla}\cdot\vec{\vec{T}}) = \frac{1}{P(o)}\oint d\gamma(\vec{n}\cdot\vec{\vec{T}}) = 0 \quad .$$

(III.2.11b)

This result corresponds to conservation of momentum, and implies that

$$\vec{R}(z) = \vec{R}(o) + z\left(\frac{1}{P(o)}\iint da\ \vec{S}(x,y,o)\right) \quad . \qquad \text{(III.2.11c)}$$

In the present case, the initial condition is given in Eq. III.1.11b, implying $\vec{R}(z) = 0$, which was already known through symmetry arguments.

Finally, defining the "2nd beam moment", $\langle r^2(z)\rangle$, which, for $\vec{R}(z) = 0$, equals the mean square radius of the beam, according to

$$\langle r^2(z)\rangle \equiv \frac{1}{P(z)} \iint da\ r^2\ w \qquad , \qquad \text{(III.2.12a)}$$

we obtain

$$\begin{aligned}
\frac{d^3\langle r^2\rangle}{dz^3} &= \frac{d^2}{dz^2}\left(-\frac{1}{P(o)}\iint da\ r^2\ \vec{\nabla}\cdot\vec{S}\right) \\
&= \frac{d^2}{dz^2}\left(-\frac{1}{P(o)}\iint da\ \{\vec{\nabla}\cdot(r^2\vec{S}) - \vec{S}\cdot\vec{\nabla}r^2\}\right) \\
&= \frac{d^2}{dz^2}\left(\frac{2}{P(o)}\iint da\ \vec{S}\cdot\vec{r}\right) = \frac{d}{dz}\left(\frac{2}{P(o)}\iint da(\vec{\nabla}\cdot\vec{\vec{T}})\cdot\vec{r}\right) \\
&= \frac{d}{dz}\left(\frac{2}{P(o)}\iint da\ \{\vec{\nabla}\cdot(\vec{\vec{T}}\cdot\vec{r}) - (\vec{\vec{T}}\cdot\vec{\nabla})\cdot\vec{r}\}\right) \\
&= \frac{d}{dz}\left(-\frac{2}{P(o)}\iint da\ \langle T\rangle\right) \\
&= -\frac{2}{P(o)}\iint da\ \vec{\nabla}\cdot\vec{Q} = -\frac{2}{P(o)}\oint d\gamma\ (\vec{n}\cdot\vec{Q}) = 0\ . \qquad \text{(III.2.12b)}
\end{aligned}$$

Except for the specific definitions of w, $\vec{S}$, $\vec{\vec{T}}$, $\vec{Q}$, all of these results can be applied to similar problems in three or more "transverse" dimensions (i.e. Maxwell's equations or Schroedinger's equations, with $z \to t$ and $\vec{r} \to x,y,z$).

We thus find that the mean square radius is precisely quadratic in z for any (confined) solution of the propagation equation:

$$\begin{aligned}
\langle r^2(z)\rangle = &\left[\frac{1}{P(0)}\iint da\ r^2\ w(x,y,o)\right] + z\left[\frac{2}{P(o)}\iint da\ \vec{S}(x,y,o)\cdot\vec{r}\right] \\
&+ z^2\left[-\frac{1}{P(o)}\iint da\ \langle T(x,y,o)\rangle\right] \qquad . \qquad \text{(III.2.12c)}
\end{aligned}$$

We note that this last integral is a "constant of the motion", and can be written:

$$\frac{1}{2}\frac{d^2\langle r^2\rangle}{dz^2} = -\frac{1}{P(o)}\iint da\,\langle T\rangle = \frac{1}{P(o)}\iint da\left\{2(\vec{\nabla}E)\cdot(\vec{\nabla}E^*)-\beta(|E|^2)^2\right\} = \text{cons.} \quad . \qquad \text{(III.2.12d)}$$

For the initial conditions corresponding to a Gaussian-elliptic beam, we obtain

$$P(z) = P(o) = \frac{\pi}{2}R$$

and

$$\langle r^2(z)\rangle = \frac{1}{4}(1+R^2) - 2(\theta_x+\theta_y)z + \left(4\left[1+\theta_x^2+(1+\theta_y^2)/R^2\right]-\beta\right)z^2 \quad . \qquad \text{(III.2.13)}$$

From this last equation, we can find (exactly !) the critical power and self-focusing distance for the problem specified in Eq. III.1.11; since a vanishing mean square beam radius implies that self-focusing has taken place, we obtain:

$$\frac{1}{z_f} = \frac{4(\theta_x+\theta_y)}{1+R^2} + 2\sqrt{\frac{(\beta-\beta_o)}{1+R^2}} \quad , \qquad \text{(III.2.14)}$$

and, for the threshold condition:

$$\beta_{crit} = \begin{cases} \beta_1 \equiv 4\left[1+\theta_x^2+\dfrac{(1+\theta_y^2)}{R^2}\right] & (\theta_x+\theta_y) < 0 \\ \beta_o \equiv 4\left[1+\dfrac{1}{R^2}+\dfrac{R^2}{1+R^2}\left(\theta_x-\dfrac{\theta_y}{R^2}\right)^2\right] & (\theta_x+\theta_y) \geq 0 \end{cases} \quad . \qquad \text{(III.2.15)}$$

Here, $\beta_1 = \beta_o + 4(\theta_x+\theta_y)^2/(1+R^2) \geq \beta_o$. From Eq. III.2.4, it can be seen that placing a lens of focal length z_o in the beam does not affect

the value of β_o ; thus the formula for z_f correctly obeys the "lens formula", $\frac{1}{z_f'} = \frac{1}{z_f} + \frac{1}{z_o}$. However, this also shows that the threshold for a converging beam $(\theta_x + \theta_y \geq 0)$ is not decreased by further pre-focusing the beam with a lens, since β_{crit} is independent of the lens transformation in this case. On the other hand, a divergent lens $(z_o < 0)$ can indeed increase the threshold by forming an initially divergent beam $(\theta_x + \theta_y < 0)$.

Of particular interest as a "test case" is a beam which is initially collimated, which implies $\theta_x = \theta_y = 0$; we then obtain

$$P_{crit} = \frac{R}{8}\beta_{crit} P_o = P_o\left(\frac{R+\frac{1}{R}}{2}\right) \quad , \quad (\theta_x = \theta_y = 0) \qquad \text{(III.2.16a)}$$

$$(z_f)_{True} = 2k_o x_o^2 z_f = \frac{k_o x_o y_o}{2}\left(\frac{P_{crit}}{P-P_{crit}}\right)^{\frac{1}{2}} \quad , \quad (\theta_x = \theta_y = 0) \qquad \text{(III.2.16b)}$$

The relationship between P_{crit} and R for the collimated beam case is illustrated in Fig. 20. For an axisymmetric beam, $R = 1$ and $P_{crit} = P_o$; thus, the "self-focusing" threshold, as usually defined, is exactly $P_o = \frac{c^3}{2n_2\omega_o^2}$ for a Gaussian shaped beam. Similarly, for $\lambda = 1\mu$, $x_o = y_o = 0.1$ cm , the focus will occur at

$$(z_f)_{True} = 314\left(\frac{P_o}{P-P_o}\right)^{\frac{1}{2}} \text{ cm} \quad ;$$

in situations where the threshold power is not greatly exceeded, this usually means that the self-focusing will in practice occur essentially at the focal point, f, of any lens placed in front of the nonlinear medium, since $f << z_f$ will hold.

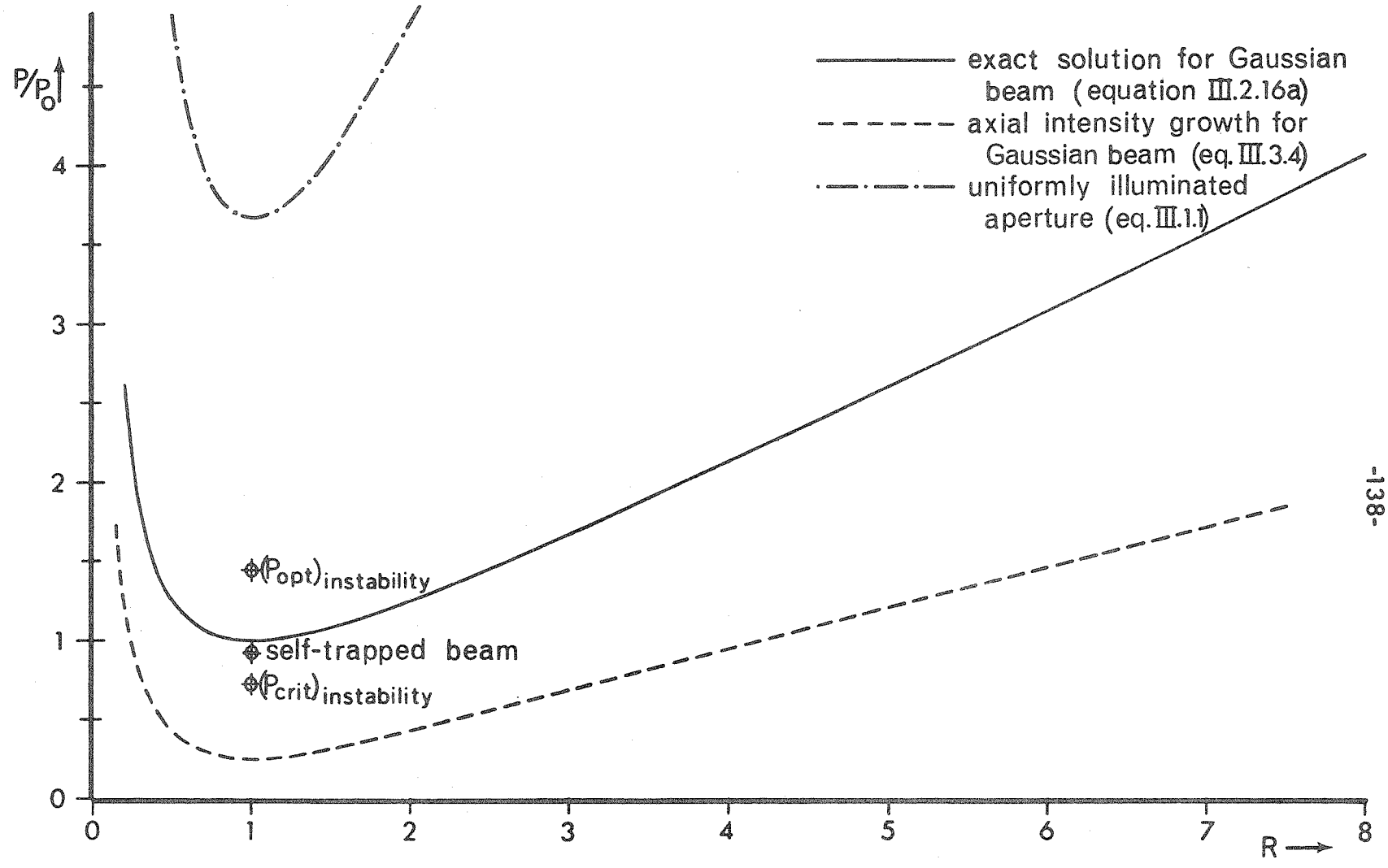


Figure 20. Comparison of various analyses of the self-focusing power vs. ellipticity for collimated beams.

Since the derivation of Eq. III.2.12c was based entirely on the analytic properties of Eq. III.2.6, a comparison with Eq. III.1.5 shows that the following holds for the real field, E, and real variables x, y, z:

$$\langle r^2(z)\rangle = \left[\frac{1}{P(o)}\iint da\; r^2\, w(x,y,o)\right] + z\left[\frac{2}{k_o P(o)}\iint da\;\vec{r}\cdot\vec{S}(x,y,o)\right] + z^2\left[\frac{1}{4k_o^2 P(o)}\iint da\left\{2(\vec{\nabla}E)\cdot(\vec{\nabla}E^*) - \frac{\omega_o^2 n_o n_2}{c^2}(|E|^2)^2\right\}\right]$$

with w, $\vec{S}$, and $P(o)$ defined by Eqs. III.2.7b, III.2.7c and III.2.10a, respectively. For a "collimated" beam at $z = 0$, we can take $E(x,y,o)$ as real, so that $\vec{S}(x,y,o) = 0$; the general threshold condition for such a beam is then given by

$$E = E_o f(x,y,z)$$

$$E_o^2 \geq \frac{2c^2}{\omega_o^2 n_o n_2}\left(\iint da\;\vec{\nabla}f\cdot\vec{\nabla}f^*\right)\Big/\left(\iint da(|f|^2)^2\right),$$

which yields

$$P = \frac{n_o c}{8\pi}\iint |E|^2 da \geq P_{crit} = \frac{P_o}{2}\left(\iint da|E|^2\right)\cdot\left(\iint da(\vec{\nabla}E)\cdot(\vec{\nabla}E^*)\right)\Big/\left(\iint da(|E|^2)^2\right) \quad . \qquad \text{(III.2.17)}$$

This critical power will also hold for an initially "divergent" beam, for which $\left.\frac{d\langle r^2\rangle}{dz}\right|_{z=0} > 0$.

If we now consider again the initial beam shapes illustrated in Fig. 19, we see that P_{crit}, as given by Eq. III.2.17, increases without limit as the boundary becomes sharper, since $\iint da\ |\vec{\nabla}E|^2 \to \infty$; this contradicts the physical model leading to Eq. III.1.1. Physically, the reason is clear; Eq. III.2.17 forms a sufficient condition for self-focusing of the entire beam, since it leads to $\langle r^2(z_f)\rangle = 0$; however, it is not necessary to satisfy this condition if the self-focusing involves only a part of the beam energy. Thus, in reference to our "uniformly illuminated apertures" part of the energy will always diffract away, so that $\langle r^2\rangle > 0$, even though most of the beam will self-focus with the threshold power estimated (roughly) in Eq. III.1.1. These arguments indicate that Eq. III.2.17 gives an accurate threshold condition only when the incident beam is sufficiently "smooth", so that none of the energy "escapes", via diffraction, from the self-focusing process, and when all of the beam focuses at a single location; this appears to be the case (near "threshold") with Gaussian beams, based on the analyses presented later. Thus, in those applications where only P_{crit} or $z_f(P \approx P_{crit})$ are important, the above formulas effectively "solve" the problem, because the onset of the self-focusing instability is adequately described by Eq. III.1.5; the higher-order complexities are usually significant only near the focus.

In addition to the general properties discussed in these last sections, there are also a few "analytic" solutions known for Eq. III.1.11a, either in explicit or implicit form[77,78]. However, we will not consider these solutions for two reasons: first, the implicit

solutions often do not correspond to physically meaningful "beams", or else they apply to special limiting geometries; and secondly, the solution methods are uniquely matched to the precise propagation Eq. III.1.11a, so that they are not applicable when any higher order "complications" are included. Our interests lie, implicitly, in considering methods of analysis which are "extendable" to the more realistic, and "complicated", problems, even though we will not deal with them directly. Thus, for our purposes, the results of Sec. III.2, as well as the particular solution found later in Sec. III.3.iii, will be used to simplify or "test" the more general analyses in Sec. III.4 and in Chap. IV.

III.3 Initial Focusing Behavior

As a first approximation to the behavior of the beam during self-focusing, we will consider the solution near $z = 0$, where the distortion is small enough to allow "expansion" methods. As an example, the discussion leading to Eq. III.1.1 would fall into this category, since the local balance between "self-lensing" and diffraction was used to predict the threshold for self-focusing. Even though such views can yield only limited information about the overall propagation behavior, because of the nonlinearity of the problem, we will find that a number of important characteristics of the "detailed" solutions can be brought out through relatively simple arguments.

III.3.i axial intensity growth

According to Eqs. III.2.7a, III.2.7b, and III.2.8a, we can write

the Taylor expansion of the beam intensity as follows:

$$w = |E|^2 \simeq w(o) + z\left(\left.\frac{\partial w}{\partial z}\right|_{z=0}\right) + z^2\left(\left.\frac{1}{2}\frac{\partial^2 w}{\partial z^2}\right|_{z=0}\right) \quad , \qquad \text{(III.3.1a)}$$

where

$$\frac{\partial w}{\partial z} = -\nabla \cdot \vec{S}$$

$$\frac{\partial^2 w}{\partial z^2} = -\vec{\nabla} \cdot (\vec{\nabla} \cdot \vec{\vec{T}}) \quad . \qquad \text{(III.3.1b)}$$

Here, $\vec{S}$ and $\vec{\vec{T}}$ are defined in Eqs. III.2.7c and III.2.8b, respectively. The present problem involves the Gaussian-elliptic beam function given in Eq. III.1.11b, so that the beam intensity on the axis is given by

$$w(o,o,z) \simeq 1 + z\left[4\left(\theta_x + \frac{\theta_y}{R^2}\right)\right] + z^2\left[4\beta\left(1 + \frac{1}{R^2}\right) - 8\left(1 + \frac{1}{R^4}\right) + 16\left(\theta_x^2 + \frac{\theta_x\theta_y}{R^2} + \frac{\theta_y^2}{R^4}\right)\right] \quad . \qquad \text{(III.3.2)}$$

As a rough guide to the onset of self-focusing, we might expect that the nonlinearity would cause the quadratic term in this expansion to be positive:

$$(\beta_{crit})_{axial} = \left[2\left(1+\frac{1}{R^4}\right) - 4\left(\theta_x^2 + \frac{\theta_x\theta_y}{R^2} + \frac{\theta_y^2}{R^4}\right)\right] \Big/ \left(1 + \frac{1}{R^2}\right) \quad . \qquad \text{(III.3.3)}$$

In the case of an initially collimated beam, this gives:

$$(P_{crit})_{axial} = \frac{P_o}{4}\left(\frac{R^2 + \frac{1}{R^2}}{R + \frac{1}{R}}\right) \qquad (\theta_x = \theta_y = 0) \quad . \qquad \text{(III.3.4)}$$

In Fig. 20, of the previous section, this result is compared with the actual self-focusing threshold. It can be seen that there is a considerable range of beam powers in which

$$(P_{crit})_{axial} < P < P_{crit} \quad ;$$

here, the nonlinearity causes the intensity on the beam axis to increase initially, even though diffraction will eventually prevent self-focusing. This observation shows that the self-focusing process cannot be analyzed accurately by methods which consider only the central portion of the beam, even though the strongest nonlinearity occurs there; the "wings" of the beam play an important part in the transverse power flow.

III.3.ii local instability growth

In Sec. III.1, we indicated that an accurate characterization of pulse propagation in realistic situations would require equations of considerably more complexity than that given in Eq. III.1.5. In addition, of course, the boundary conditions would need to be similarly examined for "realism"; in particular, the idealized Gaussian beam profile given in Eq. III.1.8 is very "smooth", whereas many pulsed lasers produce beams with considerable transverse structure. In any case, the propagation of the pulse before entering the nonlinear medium is likely to introduce some imperfections in the beam profile; we will now investigate how this "noise" on the transverse structure would influence the initial self-focusing process, using a linearized instability analysis[79,80].

In order to linearize the problem, we assume that the local beam form can be approximated to first order as a plane wave; from Eq. III.1.5, we obtain

$$E \simeq \varepsilon_0(z) + \varepsilon_1(x,y,z) \qquad |\varepsilon_1| << |\varepsilon_0| \quad , \qquad \text{(III.3.5a)}$$

where the plane wave solution has the form:

$$\varepsilon_o(z) = E_o \, e^{i\left(\frac{\omega_o n_2}{2c} E_o^2\right)z} \quad . \qquad \text{(III.3.5b)}$$

We take E_o real for convenience. Substituting this into the propagation Eq. III.1.5, and neglecting higher order nonlinear terms, we find that the perturbation field, ε_1, obeys the equation:

$$\varepsilon_1(x,y,z) = \hat{\varepsilon} \, e^{i\left(\frac{\omega_o n_2}{2c} E_o^2\right)z}$$

$$\frac{\partial^2 \hat{\varepsilon}}{\partial x^2} + \frac{\partial^2 \hat{\varepsilon}}{\partial y^2} + \frac{2\omega_o n_o}{c} \, i \, \frac{\partial \hat{\varepsilon}}{\partial z} + \frac{\omega_o^2 n_o n_2}{c^2} E_o^2 \, (\hat{\varepsilon} + \hat{\varepsilon}^*) = 0. \qquad \text{(III.3.6)}$$

This linear equation can be solved by separation of variables, where the transverse variables are included in a function, $u(x,y)$, solving:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + k^2 u = 0$$

Following Suydam[77], we consider a localized, symmetrical, perturbation for simplicity, so that we use

$$\hat{\varepsilon}(x,y,z) = J_o(kr)h(z) \quad , \qquad \text{(III.3.7a)}$$

where $r^2 = x^2 + y^2$ and the complex function $h(z) = h_1(z) + ih_2(z)$ must satisfy:

$$\frac{2\omega_o n_o}{c} \frac{dh_1}{dz} = k^2 h_2 \quad ,$$

$$\frac{2\omega_o n_o}{c} \frac{dh_2}{dz} = \left(- k^2 + \frac{2\omega_o^2 n_o n_2}{c^2} E_o^2 \right) h_1 \quad .$$

Thus, both h_1 and h_2 will obey the equation

$$\frac{d^2h_i}{dz^2} = \frac{c^2}{4\omega_o^2 n_o^2} k^2 \left(\frac{2\omega_o^2 n_o n_2}{c^2} E_o^2 - k^2 \right) h_i \equiv \alpha^2 h_i \ , \qquad \text{(III.3.7b)}$$

For $\alpha^2 < 0$, the solution is stable, since h(z) is oscillatory. However, it can be seen that unstable, exponentially increasing perturbations result if $\alpha^2 > 0$; i.e. for

$$0 \leq k^2 \leq k^2_{LIM} \ , \qquad \text{(III.3.8)}$$

with

$$k^2_{LIM} = \frac{2E_o^2 \omega_o^2}{c^2} n_o n_2 \ .$$

From Eq. III.3.7b, we see that the fastest growing perturbations have a size characterized by k_{opt} , where $k^2_{opt} = \frac{1}{2} k^2_{LIM}$; for a beam with "random" perturbations, "hot spots" of dimensions $\sim \frac{1}{k_{opt}}$ will occur.

These results are clarified by considering the power contained in these unstable perturbations; the growing "hot spot" can be taken to have the area within the central peak of the function $J_o(kr)$, so that

$$P \simeq \frac{n_o c}{8\pi} E_o^2 \cdot \pi \left(\frac{2.405}{k} \right)^2 \simeq 0.723 \frac{n_o c}{k^2} E_o^2 \ .$$

Thus, from Eq. III.3.8, a perturbation will grow if it contains a power satisfying

$$P \geq (P_{crit})_{instability} = 0.723 \frac{n_o c E_o^2}{k^2_{LIM}} = 0.723 \, P_o \ , \qquad \text{(III.3.9a)}$$

with the "most common" hot spots containing

$$(P_{opt})_{instability} = 1.45 \, P_o \ . \qquad \text{(III.3.9b)}$$

Finally, the exponential growth of the perturbation indicates that the "focusing" length will be roughly $\frac{1}{\alpha}$, giving

$$(z_f)_{\text{instability}} \simeq \frac{2k_o}{k^2}\left(\frac{P}{P_{crit}} - 1\right)^{-\frac{1}{2}} \quad , \qquad \text{(III.3.9c)}$$

which compares well with value in Eq. III.2.16b.

This instability is very important when high power ($P >> P_o$) beams are being considered, both physically, when filamentation occurs, and mathematically, since idealized results such as Eq. III.2.12 are "weakened". Thus, if a beam is initially highly asymmetric ($R >> 1$), a "subcritical" power ($P < P_{crit}$) may still produce self-focusing within portions of the cross-section containing powers $\sim P_o$.

III.3.iii shape-preserving solution

As a simple but useful "test" of later analyses involving "full" self-focusing solutions, we will now find one "exact" solution for comparison. First, we consider the "self-trapped" solution, which preserves its shape during propagation; assuming cylindrical symmetry[81], and allowing for, at most, a linear phase shift, Eq. III.1.11a reduces to

$$E(x,y,z) = f(r)e^{i\eta z}$$

$$\frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} - \eta f + \beta|f|^2 f = 0 \quad . \qquad \text{(III.3.10)}$$

Since $\left.\frac{df}{dr}\right|_{r=0} = 0$, we can take $f(o)$ to be real and conclude that $f(r)$ is a real function for all r (by uniqueness); we can also normalize the problem by taking:

$$u(r) = (\beta/\eta)^{\frac{1}{2}} f(r/\eta^{\frac{1}{2}}) \quad , \qquad \text{(III.3.11a)}$$

$$\frac{d^2u}{dr^2} + \frac{1}{r}\frac{du}{dr} - u + u^3 = 0 \quad . \qquad \text{(III.3.11b)}$$

This "propagation" equation must be solved for "eigenfunctions" which are well behaved ($u(r \to \infty) \to K_0(r)$; $\left.\frac{du}{dr}\right|_{r=0} = 0$); the "eigenvalue" has been shifted from β to $u(o)$ through the transformation in Eq. III.3.11a. A numerical calculation yields $u_0(r)$, the (lowest order) solution agreeing with that found by others[14]; it is characterized by $u_0(o) \simeq 2.18$ and an "area" of

$$a_0 \equiv \int_0^\infty u^2 r \, dr \simeq 1.86 \quad .$$

We will refer again to this function, and give its shape, in Chap. IV. In order to correspond with our dimensionless problem statement in Eq. III.1.11, the "power" must equal $P(R = 1) = \frac{\pi}{2}$ (see Eq. III.2.1). Thus

$$P = \int_0^\infty 2\pi r \, f^2(r) \, dr = \frac{2\pi}{\beta} a_0 = \frac{\pi}{2} \quad ,$$

which shows that $\beta = \beta_0 = 4a_0 \simeq 7.43$. This yields a beam with a physical power of

$$(P_{crit})_{min} = \frac{\beta_0}{8} P_0 = 0.928 \, P_0 \quad . \qquad \text{(III.3.12)}$$

The subscript "min" is used, since it can be shown that the "shape-perserving" profile considered here has the lowest possible threshold for self-focusing. The value of η is still arbitrary, since any power-conserving scale change can be made; however, it is convenient

to normalize $E(0,0,0) = 1$, so that $\eta = \eta_0 = \frac{\beta_0}{u^2(0)} \simeq 1.56$; our solution thus takes the form

$$E(x,y,z) = (\eta_0/\beta_0)^{\frac{1}{2}}\, u_0(\eta_0^{\frac{1}{2}}\,(x^2+y^2)^{\frac{1}{2}})\, e^{i\eta_0 z} \quad . \qquad \text{(III.3.13)}$$

Finally, we can use this constant-profiled function to form an exact self-focusing solution, by using the lens transformation given in Sec. III.2.i. Applying Eq. III.2.13a, we find that the following function will self-focus precisely at $z_f = z_0$ for $\beta = \beta_0$:

$$\hat{E}(x,y,z) = (\eta_0/\beta_0)^{\frac{1}{2}} \left(\frac{z_0}{z_0-z}\right) u_0\left(\frac{z_0}{z_0-z}\, \eta_0^{\frac{1}{2}}\, (x^2+y^2)^{\frac{1}{2}}\right) e^{i\eta_0 \frac{zz_0}{z_0-z}}\, e^{-\frac{i(x^2+y^2)}{4(z_0-z)}} \qquad \text{(III.3.14)}$$

where $u_0(r)$ obeys Eq. III.3.11b. This solution is illustrated in Fig. 21. Note that the equiphase surfaces of $\hat{E}(x,y,z)$ differ from those of the physical field due to the "modulation", $e^{ik_0 z}$, in the field definition III.1.3; thus, in terms of the normalized, dimensionless variables, a factor of $e^{i(2k_0^2 x_0^2)z}$ is always presumed to multiply any solutions of Eq. III.1.11.

The function, $\hat{E}(x,y,z)$, actually differs from "typical" self-focusing solutions ($\beta \neq \beta_{crit}$) in two ways. The asymptotic behavior, as $r \to \infty$, goes as $K_0(r)$, which decays as $e^{-r}/\sqrt{r}$, whereas non-shape-preserving solutions, either above or below threshold, would be expected to fall off more rapidly (usually as $e^{-\alpha_1 x^2}\, e^{-\alpha_2 y^2}$ according to Eq. III.1.7). Also, the "relative" phase-front curvature of $\hat{E}$ decreases, since the phase shift from $r = 0$ to $r = \langle r^2(z)\rangle^{\frac{1}{2}}$ goes as $(1 - z/z_0)$; for a beam above threshold ($\beta > \beta_{crit}$), however,

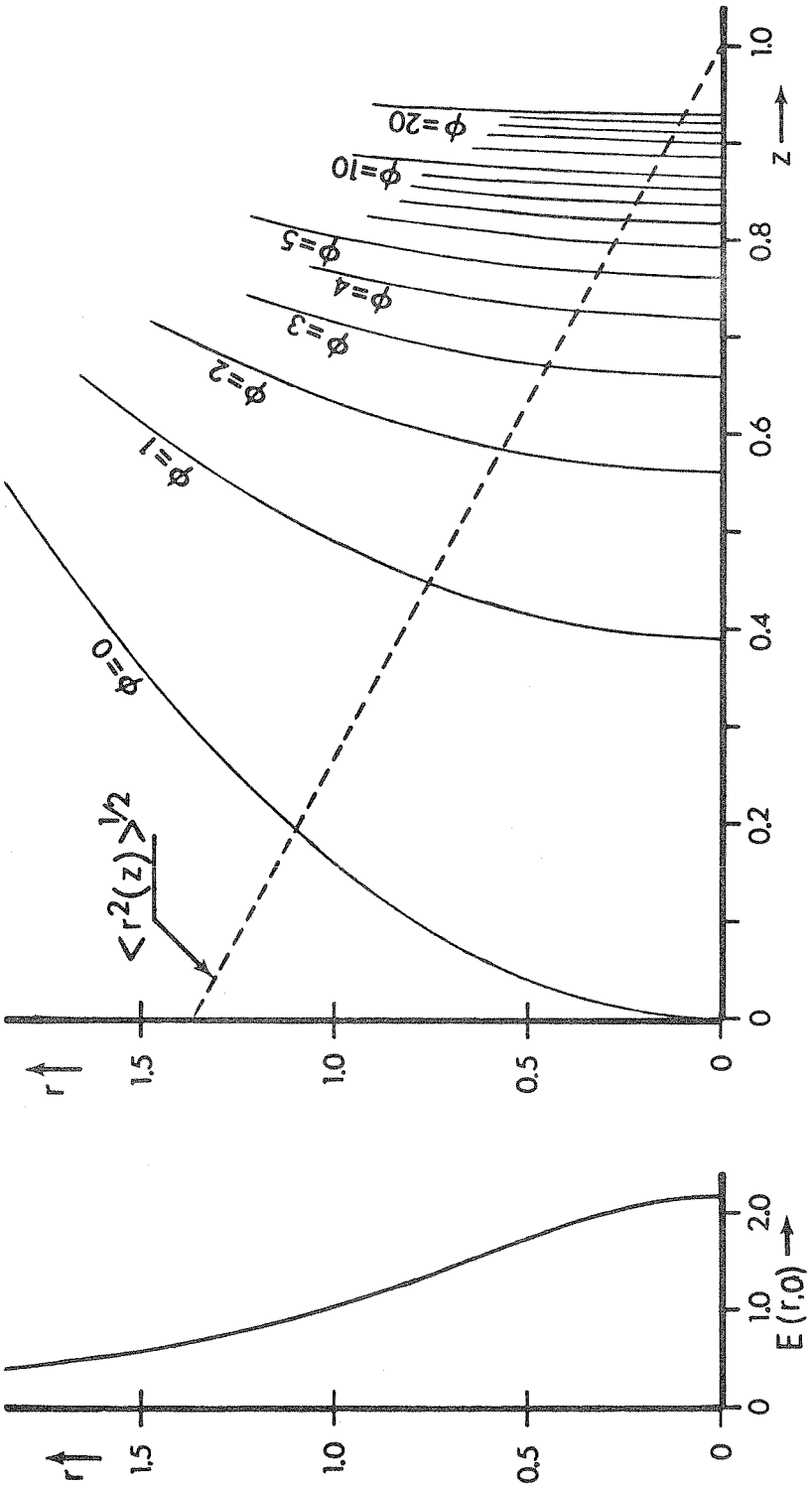


Figure 21. Exact, shape-preserving, self-focusing solution. The curves for $\langle r^2(z)\rangle^{1/2}$ and the constant phase surfaces are based on equation III.3.14 with $z_0 = 1.0$.

$\langle r^2(z)\rangle^{\frac{1}{2}}$ approaches zero as $(1 - z/z_0)^{\frac{1}{2}}$, which leads to strong phase characteristics for any finite region of interest.

III.4 Numerical Approaches

We will now discuss some results obtained for the "full" solution $E(x,y,z)$, of Eq. III.1.11 through numerical analysis; this method appears particularly useful, since the (potential) introduction of further nonlinearities or other steady-state "complications", as described in Sec. III.1, should not involve much additional difficulty. Nevertheless, we have obtained only a few "preliminary" solutions through numerical analysis for the following reasons: first, the analyses of Secs. III.2.ii, III.3.iii, and Chap. IV provide a reasonably complete picture of the properties of self-focusing beams for the "lowest order" equation, so that further numerical results could add only minimal new "information". Second, the results described below, involving partial differential equations in three dimensions, required a rather extensive use of the computing facilities (and caused a correspondingly rapid depletion of our computing budget); it thus seemed prudent to leave "high accuracy runs" to those who require self-focusing solutions for particular applications (in which case the lowest order propagation equation would probably not be sufficiently "detailed" anyway).

Our discussion of the numerical methods will be very brief, concentrating on the general approaches used, rather than on the specific computational schemes. These methods were based directly on

the propagation Eq. III.1.11a, so that, for example, the beam at "z" was described by $E(x_i,y_j,z)$, where the points (x_i,y_j) form a discrete "transverse grid"; the solution at $z' = z + \Delta z$ was then found by some finite-difference approximation to the equation system:

$$\frac{\partial E}{\partial z} = i\,\frac{\partial^2 E}{\partial x^2} + i\,\frac{\partial^2 E}{\partial y^2} + i\,\beta|E|^2\,E \qquad x \geq 0,\ y \geq 0 \tag{III.4.1a}$$

$$\left.\frac{\partial E}{\partial x}\right|_{y=0} = \left.\frac{\partial E}{\partial y}\right|_{x=0} = 0 \tag{III.4.1b}$$

$$E(x,y,o) \quad \text{given} \qquad . \tag{III.4.1c}$$

The use of only one quadrant, $x,y \geq 0$, and the boundary conditions III.4.1b are a result of the symmetry in $E(x,y,o)$ as discussed in the introduction to Sec. III.2. Note that this approach rules out some alternative problem formulations, such as the "ray tracing" method of following the beam[70].

The problem posed in Eq. III.4.1 presents at least three "unusual" difficulties which complicate any "standard approach": the presence of complex functions, the infinite transverse domain, and the potential singularity as $z \to z_f$ when $\beta > \beta_{crit}$. The last problem is, of course, unavoidable for the "lowest order" equation in III.4.1a, so that z_f can never be reached; even for "higher order" propagation equations, however, there is a near-focus for $z \sim z_f$, where the beam becomes highly nonlinear[82,83]. Since any method should be designed to handle such situations, a large number of transverse grid points and a high degree of stability are required.

In order to have the important (high amplitude) portion of the beam well represented in the transverse grid, the variables (x, y) were "scaled" as follows:

$$x = \bar{x}\, x'$$

$$y = \bar{y}\, y'$$

$$\frac{\partial E(x',y',z)}{\partial z} = \frac{i}{\bar{x}^2}\,\frac{\partial^2 E}{\partial (x')^2} + \frac{i}{\bar{y}^2}\,\frac{\partial^2 E}{\partial (y')^2} + i\beta |E|^2\, E \qquad . \qquad \text{(III.4.2)}$$

As the integration proceeded (in z), the r.m.s. moments $\langle (x')^2 \rangle^{\frac{1}{2}}$, $\langle (y')^2 \rangle^{\frac{1}{2}}$ were checked, and rescaling was performed to keep these values approximately constant; this scaling (and variable change) will be presumed to be present in our later discussions of the numerical methods.

Most of the methods were checked by solving at least two "test" cases: the free-beam limit ($\beta = 0$), whose solution can be found from Eq. III.1.7; and the "shape-preserving" solution described in Eq. III.3.14. In addition, power conservation, $\langle 1 \rangle$ = constant , was checked, and the results for $\langle r^2(z) \rangle$ were compared with Eq. III.2.13. In general, the "test" cases showed only moderate accuracy, but this fact is consistent with our use of "low accuracy" computational efforts (single precision arithmetic, low density of transverse grid points, and local error bounds of $\sim 10^{-4}$).

For comparative purposes, we used the initial conditions corresponding to R = 4 with a collimated beam, and applied the lens transformation with $z_0 = 1$:

$$E(x,y,o) = e^{-x^2}\, e^{-y^2/16}\, e^{\frac{-i(x^2+y^2)}{4}} \quad . \qquad \text{(III.4.3)}$$

This reduces the range of interest to z = (0,1), and $\beta = \beta_{crit}$ is indicated by $z_f = 1.0$; in theory, this greatly simplifies the problem of deciding when to stop integrating and whether or not "self-focusing" was occurring. In practice, z_f was never approached very closely, since the presence of self-focusing led to increased transverse phase variations and closely spaced scale changes which caused the computations to quit or run out of time.

III.4.i implicit integration methods

The "stability" of implicit numerical formulas usually makes them very useful in problems of this type; we applied the ADI method[84,85] ("alternating direction implicit"). This method is "simple" and direct, but does not allow much flexibility to cover the particular difficulties mentioned earlier. Thus, to retain the form needed for ADI, the field, E(x,y,z), was integrated directly in terms of its real and imaginary components.

$$E(x,y,z) = u(x,y,z) + iv(x,y,z)$$

The cubic term was linearized according to[71]: $|E|^2 E \rightarrow |E(z)|^2 E(z+\Delta z)$; and the "outer boundary" conditions were replaced by those of a conducting wall at a finite distance (i.e. E had odd symmetry across the outer boundary).

This approach has some severe drawbacks. First, an examination of the asymptotic form of the field (Eq. III.1.7 or III.3.14) shows phase variations of the form

$$E \sim e^{ia_x x^2} \; e^{ia_y y^2}$$

for large $|x|$, $|y|$, thus implying a highly oscillatory character to u and v. The ADI method therefore requires a high "transverse grid density" for accurate field representation; furthermore, the outer boundary is often placed, by necessity, in regions where $|E|^2$ has not "decayed" very far, so that the boundary conditions perturbed the solution. Finally, the fixed location of the outer boundary can cause great difficulty when a divergent beam is being "scaled down" to fit into the transverse grid, since some of the new field values must be extrapolated from points outside the original grid. Fortunately we could avoid such "scaling down" for the beam given in Eq. III.4.3, which is "prefocused" so that only "scaling up" is required; in more general cases, however, this problem would remain.

In Fig. 22, we indicate some results obtained for the ADI method. The theory in Sec. III.2.ii shows that $\beta_{crit} = 4.25$ for this problem, while our results give $(\beta_{crit})_{ADI} \simeq 3.75 \pm 0.25$. This discrepancy appears to be (numerically) real, since it remains when the local error is reduced; it may be caused by the outer boundary conditions, by the limitations of the transverse grid (the result for $\beta = 4.0$ employed the "maximum sized" grid: 102 x 102), or by some "bias" present in the ADI method itself. Note also, with regard to all the methods used, that "scaling up" the beam size to fit the grid involves interpolation; as the beam focuses, this process is used repeatedly, tending to smooth out the beam shape and, thus, reduce β_{crit} according to Eq. III.2.17.

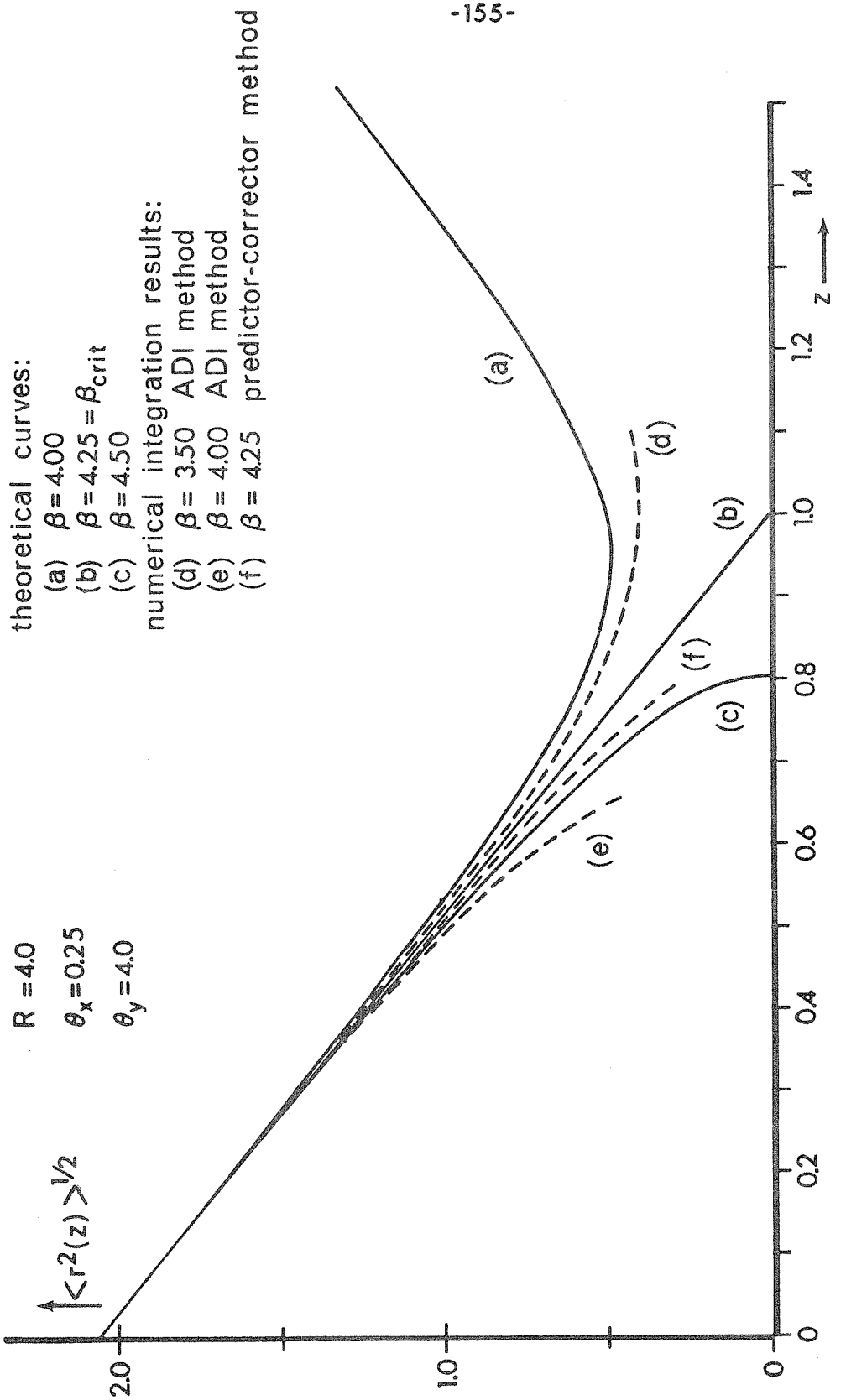


Figure 22. Comparison of numerical self-focusing results with theoretical values.

III.4.ii explicit integration methods

The use of explicit numerical formulas allows a greater flexibility in formulating the propagation problem; we used a number of techniques to alleviate some of the difficulties mentioned earlier with regard to the beam-focusing problem in general. Thus, the field was written in terms of amplitude and phase, each of which varies relatively smoothly in the transverse directions. In addition to the scaling described in Eq. III.4.2, a coordinate transformation was used to map the infinite transverse domain onto a unit square; the field was written as follows:

$$E(x,y,z) = A(s,t,z)\, e^{i\phi(s,t,z)}$$

$$s = \frac{x}{(x^2 + \bar{x}^2)^{1/2}} \qquad 0 \le s < 1$$

$$t = \frac{y}{(y^2 + \bar{y}^2)^{1/2}} \qquad 0 \le t < 1 \qquad \text{(III.4.4a)}$$

The propagation equation now took the form:

$$\frac{\partial A}{\partial z} = \frac{(1 - s^2)^3}{\bar{x}^2}\left[\frac{3s}{1-s^2} A \frac{\partial \phi}{\partial s} - A \frac{\partial^2 \phi}{\partial s^2} - 2 \frac{\partial A}{\partial s} \frac{\partial \phi}{\partial s}\right]$$

$$+ \frac{(1 - t^2)^3}{\bar{y}^2}\left[\frac{3t}{1-t^2} A \frac{\partial \phi}{\partial t} - A \frac{\partial^2 \phi}{\partial t^2} - 2 \frac{\partial A}{\partial t} \frac{\partial \phi}{\partial t}\right] \qquad \text{(III.4.4b)}$$

$$A \frac{\partial \phi}{\partial z} = \frac{(1 - s^2)^3}{\bar{x}^2}\left[\frac{\partial^2 A}{\partial s^2} - \frac{3s}{1-s^2} \frac{\partial A}{\partial s} - A\left(\frac{\partial \phi}{\partial s}\right)^2\right]$$

$$+ \frac{(1 - t^2)^3}{\bar{y}^2}\left[\frac{\partial^2 A}{\partial t^2} - \frac{3t}{1-t^2} \frac{\partial A}{\partial t} - A\left(\frac{\partial \phi}{\partial t}\right)^2\right] + \beta A^3 .$$

The transform in III.4.4a was designed to retain any symmetries present, so that

$$\left.\frac{\partial A,\phi}{\partial s}\right|_{t=0} = \left.\frac{\partial A,\phi}{\partial t}\right|_{s=0} = 0 \qquad \text{(III.4.4c)}$$

The factors $\bar{x}$ and $\bar{y}$ are the scale factors, which are adjusted to keep the beam size in (s,t) relatively constant; note that scale changes require only interpolation, because the unit square always contains the entire beam.

Among the difficulties involved with this approach is the question of modeling the beam "at infinity", which maps into the lines $s=1$, $t=1$. Consideration of the free beam expansion, equation III.1.7, suggests that the dominant behavior would be that of a Gaussian with quadratic phase, at least locally; by using three neighboring "points", the field in the "tails" could be matched to such a function. The approximate propagation solution is then given by the free-beam propagation across Δz, plus an additional phase shift of: $\beta \int_{z}^{z+\Delta z} A^2(z')dz'$. In our numerical solutions, this approach was applied to points lying near the outer boundary (the actual lines $s=1$, $t=1$ were not used); elsewhere, explicit finite difference formulas were applied.

Integration of equation III.4.4 is seen to be relatively complex and, therefore, time consuming on the computer. On the other hand, there are some major benefits to this approach which make it quite competitive with ADI. First, the functions A,ϕ are much less structured than $E = u+iv$, so that our use of a grid of size

(50 x 50) in (s,t) space to represent the field was probably much more accurate than the use of (102 x 102) grids in (x,y) space for ADI. Additionally, the explicit schemes could use higher-order transverse derivative formulas (we used a "9-point" stencil usually, with 5 points in each direction), and required fewer scale changes; overall, these should lead to a considerable improvement in the "transverse" accuracy, except in the "tails".

Similarly, the "longitudinal" integration (in z) can be performed very accurately, and with good stability, through the use of higher-order methods, several of which were tried; the storage requirements for these multiple-z-point formulas were partially offset by the smaller grid sizes. As an example, the result indicated in Figure 22 employed a 4th order predictor-corrector formula (Hamming's method)[86]; admittedly, this involved considerable "overcomputation", but it was chosen as the "opposite" extreme, as compared with the ADI.

From our preliminary results, we cannot conclude which of our approaches is "best" when applied with "high accuracy" to complex beam-focusing problems; however, it appears that the explicit methods are potentially capable of achieving accurate results under more "severe" conditions than the ADI, although the best scheme may not employ all of the "special" techniques described in this section. Neither method, though, with the resources (i.e. budget) available to us, was capable of obtaining the detailed solution needed, for example, to pass "through" a focus in the higher order equations. Finally, our numerical experience indicates that the long-range goal of solving a realistically complex problem in four dimensions (x,y,z,t) appears remote at best.

Chapter IV

PARAMETERIZED BEAM REPRESENTATION OF SELF-FOCUSING

In Chapter III, we discussed some of the conceptual and mathematical difficulties associated with any realistically detailed problem in self-focusing, and presented a number of analyses based on the lowest order, steady-state propagation equation. Of these, only the fully numerical approaches showed a potential for handling the more complicated cases; however, the accuracy attainable in practice was severely restricted by computer limitations and by the (unknown) influence of approximate "outer boundary" conditions.

The analyses presented in this chapter employ a "parameterized beam" approach; here, the transverse beam shape is approximated, at each value of z, by a specific analytical form containing a number of parameters which vary (in z) as the beam propagates[87]. Although these methods are inherently of only moderate accuracy, they avoid most of the computational difficulties of the numerical approaches, while retaining the flexibility to handle (potentially) the more complicated problem formulations. Using this method, we will show that many characteristics of self-focusing are well represented by parameterized beam solutions, when an action-integral minimization technique is used to determine the parameter variations.

IV.1 Near-Axis Parameterized Expansion

The primary differences between various parameterized beam analyses are in the functional forms used to model the beam, and

in the method used to "match" this form to the true beam shape. For reference, we will again state the problem of interest in this chapter: the lowest order, steady-state, propagation equations governing self-focusing, as expressed in dimensionless form

$$\frac{\partial E}{\partial z} = i\left(\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \beta|E|^2 E\right) \qquad z \geq 0 \qquad \text{(IV.1.1a)}$$

$$E(x,y,0) = e^{-x^2} e^{-y^2/R^2} e^{-i\theta_x x^2} e^{-i\theta_y y^2/R^2} \qquad \text{(IV.1.1b)}$$

We will treat this problem abstractly, but the results can be referred to the physical quantities by using equations III.1.12 and III.1.13.

We will use the following functional form, $\hat{E}(x,y,z)$, to approximate the true solution, $E(x,y,z)$, at any given z:

$$\hat{E}(x,y,z) = Ae^{i\phi} e^{-x^2/a^2} e^{-y^2/b^2} e^{iS_x x^2} e^{iS_y y^2} \left\{ \sum_{n=0}^{N} \sum_{m=0}^{N-n} x^{2n} y^{2m} \frac{\tau_{n,m}}{a^{2n} b^{2m}} \right\} , \qquad \text{(IV.1.2a)}$$

where the real parameters, $A(z), \phi(z), a(z), S_x(z), b(z)$, and $S_y(z)$, and the complex parameters, $\tau_{n,m}(z)$, are functions of z only; $\tau_{0,0} \equiv 1$ holds, to normalize the polynomial in the brackets. Note that this corresponds to an elliptical-Gaussian profile (with a quadratic radial phase dependence) multiplied by a (complex) polynomial; the on-axis field is given by $\hat{E}(0,0,z) = A(z)e^{i\phi(z)}$. This expansion has the following useful properties: it automatically satisfies the symmetry and outer boundary conditions; it can be matched to the initial conditions; it is analytic, square integrable,

and computationally "manageable"; it satisfies the free-beam limit as a test case (see equation III.1.7); and, it physically represents the type of localized beam expected to occur. An alternative expansion, including a sum of functions of the form in IV.1.2a, would be appropriate for highly structured initial conditions or for the higher-order equations, where parts of the beam may focus while others do not; this should not be necessary here. We will refer to "N" as the "order" of the solution; for $N \to \infty$, the parameterized solution should approach E exactly, but the algebraic complexity of the problem will prevent using $N \geq 3$.

The initial conditions corresponding to equation IV.1.1b are:

$$\begin{aligned} &A(0) = 1, && \phi(0) = 0 \\ &a(0) = 1, && b(0) = R \\ &S_x(0) = -\theta_x, && S_y(0) = -\theta_y/R^2 \\ &\tau_{n,m}(0) = 0 && (n,m) \neq (0,0) \quad . \end{aligned} \tag{IV.1.2b}$$

The "evolution" of these parameters with z thus describes the beam propagation. In the "near axis" method, the equation for $\hat{E}$, equation IV.1.2a, is substituted directly into the propagation equation, IV.1.1a; the parameter derivatives are then found by equating terms near the "axis", $(x,y) \sim (0,0)$.

The lowest order parameterized beam shape is obtained from equation IV.1.2a when $N = 0$:

$$\hat{E}_0(x,y,z) = A(z)e^{i\phi(z)}e^{-x^2(1/a^2(z)-iS_x(z))}e^{-y^2(1/b^2(z)-iS_y(z))} \tag{IV.1.3}$$

We will carry out the algebra for this case as an example of the near-axis method. Substituting $\hat{E}_o$ into equation IV.1.1a yields:

$$\hat{E}\left\{\frac{1}{A}\frac{dA}{dz} + i\frac{d\phi}{dz} + x^2\left(-\frac{2}{a^3}\frac{da}{dz} + i\frac{dS_x}{dz}\right) + y^2\left(-\frac{2}{b^3}\frac{db}{dz} + i\frac{dS_y}{dz}\right)\right\}$$

$$= \hat{E}_o\left\{i\left(2iS_x - \frac{2}{a^2}\right) + ix^2\left(2iS_x - \frac{2}{a^2}\right)^2 + i\left(2iS_y - \frac{2}{b^2}\right) + iy^2\left(2iS_y - \frac{2}{b^2}\right)^2\right\}$$

$$+ \hat{E}_o\left\{i\beta A^2 e^{-2x^2/a^2} e^{-2y^2/b^2}\right\} \quad .$$

Cancelling the common factor, $\hat{E}_o$, and expanding the nonlinear contribution as a power series about the axis (x=y=0) out to x^2 and y^2, we can equate terms to solve for the parameter derivatives:

$$\begin{aligned}
\frac{1}{A}\frac{dA}{dz} &= -2S_x - 2S_y \\
\frac{d\phi}{dz} &= -\frac{2}{a^2} - \frac{2}{b^2} + \beta A^2 \\
\frac{1}{a}\frac{da}{dz} &= 4S_x \\
a^2\frac{dS_x}{dz} &= \frac{4}{a^2} - 4a^2S_x^2 - 2\beta A^2 \\
\frac{1}{b}\frac{db}{dz} &= 4S_y \\
b^2\frac{dS_y}{dz} &= \frac{4}{b^2} - 4b^2S_y^2 - 2\beta A^2 \quad .
\end{aligned} \tag{IV.1.4}$$

First, it can be seen that these equations are consistent with power conservation, since

$$P = \int\int dxdy|\hat{E}|^2 = \frac{\pi}{2} A^2 ab \qquad \text{(IV.1.5a)}$$

$$\frac{dP}{dz} = \frac{\pi}{2} A^2 ab \left\{ \frac{2}{A}\frac{dA}{dz} + \frac{1}{a}\frac{da}{dz} + \frac{1}{b}\frac{db}{dz} \right\} = 0 \ . \qquad \text{(IV.1.5b)}$$

Using the initial conditions in equation IV.1.2b, we have:

$$A^2 ab = \text{constant} = R. \qquad \text{(IV.1.5c)}$$

Next, we can consider the equation for the mean square radius, which is given by the expression:

$$\langle r^2(z)\rangle \equiv \left(\iint dxdy r^2|\hat{E}|^2 \right) / \iint dxdy|\hat{E}|^2 = \frac{a^2+b^2}{4} \qquad \text{(IV.1.6a)}$$

We now find:

$$\frac{d^3\langle r^2\rangle}{dz^3} = \frac{d^2}{dz^2}\left\{ \frac{a^2}{2}\left(\frac{1}{a}\frac{da}{dz}\right) + \frac{b^2}{2}\left(\frac{1}{b}\frac{db}{dz}\right) \right\} = \frac{d^2}{dz^2}\left\{ 2a^2 S_x + 2b^2 S_y \right\}$$

$$= \frac{d}{dz}\left\{ 4a^2 S_x\left(\frac{1}{a}\frac{da}{dz}\right) + 2\left(a^2\frac{dS_x}{dz}\right) + 4b^2 S_y\left(\frac{1}{b}\frac{db}{dz}\right) + 2\left(b^2\frac{dS_y}{dz}\right) \right\}$$

$$= \frac{d}{dz}\left\{ 8\left(\frac{1}{a^2} + \frac{1}{b^2} + a^2 S_x^2 + b^2 S_y^2 - \beta A^2\right) \right\}$$

$$= 16\left\{ \left(-\frac{1}{a^2} + a^2 S_x^2\right)\left(\frac{1}{a}\frac{da}{dz}\right) + S_x\left(a^2\frac{dS_x}{dz}\right) + \left(-\frac{1}{b^2} + b^2 S_y^2\right)\left(\frac{1}{b}\frac{db}{dz}\right) \right.$$

$$\left. + S_y\left(b^2\frac{dS_y}{dz}\right) - \beta A^2\left(\frac{1}{A}\frac{dA}{dz}\right) \right\}$$

$$= 0 \ , \qquad \text{(IV.1.6b)}$$

which shows that the $N = 0$ near-axis parameterized beam approach yields:

$$\langle r^2(z)\rangle = (1+R^2)/4 - 2(\theta_x+\theta_y)z + (4[1+\theta_x^2 + (1+\theta_y^2)/R^2] - 4\beta)z^2 \ . \tag{IV.1.6c}$$

A comparison with the exact formula in equation III.2.13 shows that the above result has the correct form, but differs by a factor of 4 in the nonlinear contribution. Thus:

$$\left(\frac{1}{z_f}\right)_{\substack{\text{near}\\ \text{axis}}} = \frac{4(\theta_x+\theta_y)}{1+R^2} + 2\sqrt{\frac{(4\beta-\beta_o)}{1+R^2}} \ , \tag{IV.1.7a}$$

$$\left(\beta_{crit}\right)_{\substack{\text{near}\\ \text{axis}}} = (\beta_{crit})/4 = \begin{cases} \beta_1/4 & \theta_x+\theta_y < 0 \\ \beta_o/4 & \theta_x+\theta_y \geq 0 \end{cases} \tag{IV.1.7b}$$

where β_o and β_1 are defined in equation III.2.15. This factor of 4 discrepancy in the threshold is very similar to that found by examining the axial intensity growth in section III.3.i, and can be explained in the same way[83], since the tails of the beam are poorly represented by the power series expansion of the nonlinear term. Nevertheless, this approach (or others equivalent to it) has been applied to the higher-order beam-focusing problems with some qualitative success[87].

When an attempt is made to extend this near-axis parameterized beam approach to any (finite) $N \geq 1$, the method fails. If we consider, for example, the $N = 1$ case, it is seen that the propagation equation

involves (complex) coefficients of terms varying as x^0y^0, x^2, y^2, x^4, x^2y^2, and y^4, which leads to 12 equations; on the other hand, there are only 10 "unknowns", since $\tau_{1,0}$ and $\tau_{0,1}$ introduce 4 new ones beyond those of N = 0. These 12 equations in 10 unknowns are redundant (and solvable) only when $\beta=0$, or when axial symmetry applies; otherwise, there is no consistent way of finding a solution. Furthermore, even for the axially symmetric case, the solutions neither obey the lens transformation nor conserve beam power; we will therefore not pursue this approach further.

IV.2 Least-Action Parameterization

As an alternative approach to "matching" $\hat{E}$ to E, we will now consider in some detail a method which "optimizes" the match over the entire beam cross section. This method is analogous to the use of a variational principle in minimizing a system's energy; here, the propagation is assumed to be governed by the "principle of least action"[88], and it is the action-integral for $\hat{E}$ which is minimized by the parameter evolution equations[89]. To clarify the situation, we will describe the corresponding analysis for the exact solution, E(x,y,z), to equation IV.1.1; the "action-integral" for this system, $\mathcal{A}$, is given by:

$$\mathcal{A} = \int_0^z dz \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \, \mathcal{L}\left(E, E^*, \frac{\partial E}{\partial x}, \frac{\partial E}{\partial y}, \ldots, x, y, z\right), \quad \text{(IV.2.1)}$$

where the functional, $\mathcal{L}$, is the Lagrangian density. The "equations of motion" for the dependent functions, f_n, (in this case there are

two, E and E^*) are found by assuming that the action-integral is minimized; the Euler-Lagrange equations then take the form:

$$\sum_i \frac{\partial}{\partial r_i}\left(\frac{\partial \mathcal{L}}{\partial\left(\frac{\partial f_n}{\partial r_i}\right)}\right) - \frac{\partial \mathcal{L}}{\partial f_n} = 0. \qquad f_n = E, E^* \qquad r_i = x,y,z \tag{IV.2.2}$$

The Lagrangian corresponding to equation IV.1.1a is

$$\mathcal{L} = \frac{\beta}{2}(|E|^2)^2 + \frac{i}{2}\left(E^* \frac{\partial E}{\partial z} - E \frac{\partial E^*}{\partial z}\right) - \frac{\partial E}{\partial x}\frac{\partial E^*}{\partial x} - \frac{\partial E}{\partial y}\frac{\partial E^*}{\partial y} \quad . \tag{IV.2.3}$$

This can be verified by forming the Euler-Lagrange equation for $f_n = E^*$

$$0 = \frac{\partial}{\partial x}\left(-\frac{\partial E}{\partial x}\right) + \frac{\partial}{\partial y}\left(-\frac{\partial E}{\partial y}\right) + \frac{\partial}{\partial z}\left(-\frac{i}{2}E\right) - \left(\beta|E|^2 E + \frac{i}{2}\frac{\partial E}{\partial z}\right)$$

$$= -\left[i\frac{\partial E}{\partial z} + \frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \beta|E|^2 E\right] .$$

By substituting this equation into the Lagrangian, it can be seen that the exact solution will have:

$$\mathcal{L}_{exact} = -\frac{\beta}{2}(|E|^2)^2 - \frac{1}{2}\left(\frac{\partial^2|E|^2}{\partial x^2} + \frac{\partial^2|E|^2}{\partial y^2}\right) \tag{IV.2.4a}$$

$$\mathcal{A}_{exact} = \int_0^z dz \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, \mathcal{L}_{exact} = \int_0^z dz \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\left(-\frac{\beta}{2}(|E|^2)^2\right) \tag{IV.2.4b}$$

Thus, the action, as we have defined it, will be negative, decreasing monotonically with z, the propagation distance. A Taylor expansion of this action, for the initial conditions given in equation IV.1.1b,

yields the following:

$$\mathcal{A}_{exact}(z) = \int_0^z dz \left\{ \left[-\frac{\pi\beta R}{8} \right] + z\left[-\frac{\pi\beta R}{2} (\theta_x+\theta_y/R^2) \right] \right.$$

$$\left. +z^2\left[\beta\pi R\left(-2\left(\theta_x^2+\frac{\theta_x\theta_y}{R^2}+\frac{\theta_y^2}{R^4}\right)+1+\frac{1}{R^4}-\frac{2}{9}\beta\left(1+\frac{1}{R^2}\right)\right)\right]+\ldots \right\}$$

$$= -\frac{\pi\beta R}{8} z \left\{ 1+2(\theta_x+\theta_y/R^2)z+z^2\left[\frac{16}{3}\left(\theta_x^2+\frac{\theta_x\theta_y}{R^2}+\frac{\theta_y^2}{R^4}\right)-\frac{8}{3}\left(1+\frac{1}{R^4}\right)\right.\right.$$

$$\left.\left. +\eta\beta\left(1+\frac{1}{R^2}\right)\right]+\ldots \right\} \qquad \text{(IV.2.5)}$$

where $\eta = \eta_{exact} = \frac{16}{27}$. We will later find that the action-integrals for the parameterized beam functions, $\hat{E}_N$, can be similarly expanded, and differ, to order z^3, only in the value of η. At that time, we will find that $\eta_N < \eta_{exact}$, which implies that $\mathcal{A}_{exact}$ does, in fact, represent the "least" action attainable.

IV.2.i parameter evolution equations

We will apply the action minimization technique to the parameterized beam as follows: the Lagrangian density is evaluated with E replaced by $\hat{E}$ as given in equation IV.1.2a, and the integrations over x and y are carried out; the "optimum" parameter values are then those which minimize the action, where only the integration over z remains. Thus, the three-dimensional function, $E(x,y,z)$, is computationally replaced by a set of one-dimensional functions, $\{\alpha_m\}$ (i.e. the parameters). We will write the parameterized beam in the form:

$$\hat{E}(x,y,z) = \hat{E}(x,y,\alpha_m(z)) \qquad \alpha_m = \{A,\phi,a,S_x,b,S_y,\tau_{k,\ell},\tau^*_{k,\ell}\} \ . \tag{IV.2.6}$$

The action-integral takes the form:

$$\hat{a}(z) = \int_0^z dz\ \hat{L}(\alpha_m, \frac{d\alpha_m}{dz}) \quad , \tag{IV.2.7}$$

where the effective Lagrangian, $\hat{L}$, is given, according to equation IV.2.3, by

$$\hat{L} = \iint dxdy \mathcal{L}(\hat{E},\hat{E}^*, \frac{\partial\hat{E}}{\partial x}, \frac{\partial\hat{E}}{\partial y}, \ldots) \equiv L_1 + L_2 + L_3 \tag{IV.2.8}$$

with

$$L_1(\alpha_m) = \frac{\beta}{2}\iint dxdy(|\hat{E}|^2)^2 \equiv \beta\ell_1(\alpha_m) \tag{IV.2.9a}$$

$$L_2(\alpha_m, \frac{d\alpha_m}{dz}) = \frac{i}{2}\iint dxdy(\hat{E}^*\frac{\partial\hat{E}}{\partial z} - \hat{E}\frac{\partial\hat{E}^*}{\partial z}) \equiv \sum_n \frac{d\alpha_n}{dz} c_n(\alpha_m) \tag{IV.2.9b}$$

$$L_3(\alpha_m) = -\iint dxdy(\frac{\partial\hat{E}}{\partial x}\frac{\partial\hat{E}^*}{\partial x} + \frac{\partial\hat{E}}{\partial y}\frac{\partial\hat{E}^*}{\partial y}) \equiv \ell_3(\alpha_m) \ . \tag{IV.2.9c}$$

Here, for example, the notation "$\ell_1(\alpha_m)$" means that "ℓ_1" is a function of the parameters, $\{\alpha_m\}$, but not of their derivatives. In the above, we have

$$\ell_1(\alpha_m) = \frac{1}{2}\iint dxdy(|\hat{E}|^2)^2$$

$$c_n(\alpha_m) = \frac{i}{2}\iint dxdy(\hat{E}^*\frac{\partial\hat{E}}{\partial\alpha_n} - \hat{E}\frac{\partial\hat{E}^*}{\partial\alpha_n}) \ . \tag{IV.2.9d}$$

The "beam propagation" is determined by the "evolution" equations for α_m, which result from using the Euler-Lagrange equations to minimize $\hat{a}$:

$$\frac{d}{dz}\left(\frac{\partial \hat{L}}{\partial\left(\frac{d\alpha_n}{dz}\right)}\right) - \frac{\partial \hat{L}}{\partial \alpha_n} = 0 \qquad n \to A, \phi, \ldots, \tau^*_{k,\ell} \quad . \tag{IV.2.10}$$

From the form of $\hat{L}$ given above, these Euler-Lagrange equations become:

$$\sum_k \frac{d\alpha_k}{dz} d_{n,k}(\alpha_m) = g_n(\alpha_m) + \beta h_n(\alpha_m) \tag{IV.2.11a}$$

where

$$d_{n,k} = \frac{\partial c_n}{\partial \alpha_k} - \frac{\partial c_k}{\partial \alpha_n} = i \iint dxdy \left(\frac{\partial \hat{E}}{\partial \alpha_n}\frac{\partial \hat{E}^*}{\partial \alpha_k} - \frac{\partial \hat{E}}{\partial \alpha_k}\frac{\partial \hat{E}^*}{\partial \alpha_n}\right)$$

$$g_n = \frac{\partial \ell_3}{\partial \alpha_n} \tag{IV.2.11b}$$

$$h_n = \frac{\partial \ell_1}{\partial \alpha_n} \quad .$$

Formally, then, the problem of finding the "equations of motion" is solved by inverting the matrix $\hat{d} \equiv (d_{n,k})$, giving

$$\frac{d\alpha_n}{dz} = \sum_k (\hat{d}^{-1})_{n,k}(g_k + \beta h_k) \equiv \left(\frac{d\alpha_n}{dz}\right)_{LIN} + \beta\left(\frac{d\alpha_n}{dz}\right)_{N.L.} \tag{IV.2.11c}$$

In later sections, we will carry out these operations, effectively, for $\hat{E}$ parameterized to order N=0,1,2; however, we can now prove some useful results valid for all orders. First, it can be seen that the major objection to the near-axis approach has been

overcome here, since formula IV.2.10 always generates one equation for each parameter. Next, we can consider the Euler-Lagrange equations for A and ϕ, by writing equation IV.1.2a as

$$\hat{E}(x,y,z) = A(z)e^{i\phi(z)}f(x,y,z) \quad .$$

We then have

$$L_1 = A^4\{ \frac{\beta}{2} \iint dxdy(|f|^2)^2\}$$

$$L_2 = A^2 \{\frac{1}{A}\frac{dA}{dz}[0] + \frac{d\phi}{dz}[-\iint dxdy|f|^2] + \frac{i}{2}\iint dxdy(f^*\frac{\partial f}{dz} - f\frac{\partial f^*}{\partial z})\}$$

$$L_3 = A^2 \{-\iint dxdy(\frac{\partial f}{\partial x}\frac{\partial f^*}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial f^*}{\partial y})\} \quad .$$

Setting $\alpha_n = \phi$ in equation IV.2.10 yields

$$\frac{d}{dz}(-A^2\iint dxdy|f|^2) = -\frac{d}{dz}(\iint dxdy|\hat{E}|^2) = 0 \ ,$$

which shows that the parameterized solutions will conserve power. Similarly, the Euler-Lagrange equation generated by $\alpha_n = A$ shows

$$2L_1 + L_2 + L_3 = 0 \Longrightarrow \hat{L} = -L_1 \qquad \text{(IV.2.12a)}$$

which corresponds to the exact result obtained in equation IV.2.4b.

By substituting the lens transformation, equation II.2.3a, into the parameterized beam shape given in equation IV.1.2a, it can be shown, after considerable algebra, that the transformed parameters correctly obey equation IV.2.11a. We will also find that the free-beam

limit ($\beta = 0$) correctly reduces to the form in equation III.1.7; thus, our action-integral minimization procedure leads to solutions which reproduce many important properties of the exact solutions.

We will later find that, for $N \geq 1$, it is essential to obtain an expansion of the action-integral, $\hat{a}(z)$, corresponding at least to the accuracy in equation IV.2.5; we will now derive some formulas useful in such an expansion. Since $\hat{E}(x,y,0) = E(x,y,0)$ a comparison of equations IV.2.4, IV.2.9a, and IV.2.12a shows that the leading term in $\hat{a}$ will equal that in a. Next, we can derive the identity:

$$\frac{d\hat{L}}{dz} = \sum_n \left\{ \frac{\partial \hat{L}}{\partial \alpha_n} \frac{d\alpha_n}{dz} + \frac{\partial \hat{L}}{\partial\left(\frac{d\alpha_n}{dz}\right)} \frac{d}{dz}\left(\frac{d\alpha_n}{dz}\right) \right\} = \sum_n \left\{ \frac{d}{dz}\left(\frac{\partial \hat{L}}{\partial\left(\frac{d\alpha_n}{dz}\right)} \right) \frac{d\alpha_n}{dz} \right.$$

$$\left. + \frac{\partial \hat{L}}{\partial\left(\frac{d\alpha_n}{dz}\right)} \frac{d}{dz}\left(\frac{d\alpha_n}{dz}\right) \right\} = \frac{d}{dz} \sum_n \left\{ \frac{\partial \hat{L}}{\partial\left(\frac{d\alpha_n}{dz}\right)} \frac{d\alpha_n}{dz} \right\} = \frac{d}{dz}(L_2) \,. \qquad \text{(IV.2.12b)}$$

The second step follows from the Euler-Lagrange equations, and the last identity is a result of the linear form of L_2 shown in equation IV.2.9b. Since $\hat{L} = L_1 + L_2 + L_3$, we then have

$$L_1 + L_3 = \iint dxdy \left\{ \frac{\beta}{2}(|\hat{E}|^2)^2 - \left|\frac{d\hat{E}}{dx}\right|^2 - \left|\frac{d\hat{E}}{dx}\right|^2 \right\} = \text{constant}, \qquad \text{(IV.2.12c)}$$

which holds for the exact solution as well (see equation III.2.12d). In the present case, however, this constant cannot be related to $\frac{d^2\langle r^2\rangle}{dz^2}$, so that this conservation law is not of such direct

usefulness. On the other hand, we can now derive other conservation laws by writing $\frac{d\hat{L}}{dz}$ in two ways; from equations IV.2.12a and IV.2.11c, we obtain

$$\frac{d\hat{L}}{dz} = \frac{d}{dz}(-L_1) = -\beta \sum_n \frac{d\alpha_n}{dz} h_n = \beta \left\{ - \sum_n h_n \left(\frac{d\alpha_n}{dz}\right)_{LIN} \right\} + \beta^2 \left\{ -\sum_n h_n \left(\frac{d\alpha_n}{dz}\right)_{N.L.} \right\} .$$

Alternatively, from equation IV.2.12c, we also can write

$$\frac{d\hat{L}}{dz} = \frac{d}{dz}(L_3) = \sum_n \frac{d\alpha_n}{dz} g_n = \left\{ \sum_n g_n \left(\frac{d\alpha_n}{dz}\right)_{LIN} \right\} + \beta \left\{ \sum_n g_n \left(\frac{d\alpha_n}{dz}\right)_{N.L.} \right\} .$$

Since these must hold for all values of the parameters and for any β, we have the results:

$$\sum_n \frac{\partial \ell_3}{\partial \alpha_n} \left(\frac{d\alpha_n}{dz}\right)_{LIN} = \sum_n \frac{\partial \ell}{\partial \alpha_n} \left(\frac{d\alpha_n}{dz}\right)_{N.L.} = \sum_n \left\{ \frac{\partial \ell_1}{\partial \alpha_n} \left(\frac{d\alpha_n}{dz}\right)_{LIN} - \frac{\partial \ell_3}{\partial \alpha_n} \left(\frac{d\alpha_n}{dz}\right)_{N.L.} \right\} = 0$$

$$\frac{d\hat{L}}{dz} = \beta \left\{ -\sum_n \frac{\partial \ell_1}{\partial \alpha_n} \left(\frac{d\alpha_n}{dz}\right)_{LIN} \right\} . \qquad \text{(IV.2.12d)}$$

We will find that $\left(\frac{d\alpha_n}{dz}\right)_{LIN}$ is very easily found, so that equation IV.2.12d greatly simplifies the algebra needed in calculating an expansion for $\hat{L}$.

IV.2.ii lowest order solution (N = 0)

For N = 0 the analysis is algebraically simple; from equation IV.1.2a, the field takes the form:

$$\hat{E}_o(x,y,z) = A e^{i\phi} e^{-x^2\left(\frac{1}{a^2} - iS_x\right)} e^{-y^2\left(\frac{1}{b^2} - iS_y\right)} . \qquad \text{(IV.2.13)}$$

According to equation IV.2.9, the Lagrangian density is given by

$\hat{L} = L_1 + L_2 + L_3$, with

$$L_1 = \frac{\pi}{8}\beta A^4 ab$$

$$L_2 = \frac{dA}{dz}[0] + \frac{d\phi}{dz}\left[-\frac{\pi}{2}A^2ab\right] + \frac{da}{dz}[0] + \frac{dS_x}{dz}\left[-\frac{\pi}{8}A^2a^3b\right] + \frac{db}{dz}[0] + \frac{dS_y}{dz}\left[-\frac{\pi}{8}A^2ab^3\right]$$

$$L_3 = \frac{\pi}{2}A^2ab\left[-\frac{1}{a^2} - \frac{1}{b^2} - a^2S_x^2 - b^2S_y^2\right] . \qquad \text{(IV.2.14)}$$

The Euler-Lagrange equations for A, ϕ, a, S_x, b, and S_y, respectively, take the form (after cancelling common factors of $\pi A^2 ab$, etc.):

$$\frac{d\phi}{dz}(4) + a^2\frac{dS_x}{dz}(1) + b^2\frac{dS_y}{dz}(1) = -\frac{4}{a^2} - \frac{4}{b^2} - 4a^2S_x^2 - 4b^2S_y^2 + \beta A^2(2)$$

$$\frac{1}{A}\frac{dA}{dz}(2) + \frac{1}{a}\frac{da}{dz}(1) + \frac{1}{b}\frac{db}{dz}(1) = 0$$

$$\frac{d\phi}{dz}(4) + a^2\frac{dS_x}{dz}(3) + b^2\frac{dS_y}{dz}(1) = \frac{4}{a^2} - \frac{4}{b^2} - 12a^2S_x^2 - 4b^2S_y^2 + \beta A^2(1)$$

$$\frac{1}{A}\frac{dA}{dz}(2) + \frac{1}{a}\frac{da}{dz}(3) + \frac{1}{b}\frac{db}{dz}(1) = 8S_x$$

$$\frac{d\phi}{dz}(4) + a^2\frac{dS_x}{dz}(1) + b^2\frac{dS_y}{dz}(3) = -\frac{4}{a^2} + \frac{4}{b^2} - 4a^2S_x^2 - 12b^2S_y^2 + \beta A^2(1)$$

$$\frac{1}{A}\frac{dA}{dz}(2) + \frac{1}{a}\frac{da}{dz}(1) + \frac{1}{b}\frac{db}{dz}(3) = 8S_y .$$

Solving these, we obtain the "propagation" equations:

$$\frac{1}{A}\frac{dA}{dz} = -2S_x - 2S_y \qquad A(0) = 1$$

$$\frac{d\phi}{dz} = -\frac{2}{a^2} - \frac{2}{b^2} + \beta A^2\left(\frac{3}{4}\right) \qquad \phi(0) = 0$$

$$\frac{1}{a}\frac{da}{dz} = 4S_x \qquad a(0) = 1 \qquad \text{(IV.2.15)}$$

$$a^2\frac{dS_x}{dz} = \frac{4}{a^2} - 4a^2S_x^2 + \beta A^2\left(-\frac{1}{2}\right) \qquad S_x(0) = -\theta_x$$

$$\frac{1}{b}\frac{dS_x}{dz} = 4S_y \qquad b(0) = R$$

$$b^2\frac{dS_y}{dz} = \frac{4}{b^2} - 4b^2S_y^2 + \beta A^2\left(-\frac{1}{2}\right) \qquad S_y(0) = -\theta_y/R^2$$

A comparison with equation IV.1.4 reveals that the formulas for A, a, S_x, b, and S_y are the same as for the near-axis method, except that each "β" is replaced by "$\beta/4$". Since the mean square radius will follow equation IV.1.6c with a similar change, we find that the N = 0 least-action parameterized beam satisfies the "exact" result given in equation III.2.13; the formulas for $z_f(N = 0)$ and $\beta_{crit}(N = 0)$ will be precisely those of equations III.2.14 and III.2.15. This agreement, however, must be considered "coincidental", since $E(x,y,z)$ is certainly not equal to the Gaussian given by $\hat{E}_\phi$ in equation IV.2.13, and since we will find that solutions for $N > 0$ produce slightly different results for z_f and β_{crit}, despite being more accurate in their overall beam representations.

We will now convert the other "general results" of the last

section into explicit formulas for $N = 0$. First, the second Euler-Lagrange equation shows that power is conserved (see equation IV.1.5b):

$$\frac{\pi}{2} A^2 ab = \text{constant} = \frac{\pi}{2} R \ . \qquad \text{(IV.2.16a)}$$

Next, $2L_1 + L_2 + L_3 = 0$ is seen to hold by direct substitution, showing that $\hat{a}_o = -\frac{\pi}{8}\beta\int_0^z (A^4 ab)dz = -\frac{\pi}{8}\beta R\int_0^z A^2 dz$. From equation IV.2.12c, we have another conservation law:

$$\frac{\beta A^4 ab}{4} - A^2 ab\left(\frac{1}{a^2} + \frac{1}{b^2} + a^2 S_x^2 + b^2 S_y^2\right) = \text{constant} = \frac{\beta R}{4} - R\left(1 + \frac{1}{R^2} + \theta_x^2 + \theta_y^2/R^2\right). \qquad \text{(IV.2.16b)}$$

Equation IV.2.12d reduces to:

$$\frac{d\hat{L}}{dz} = \frac{\pi}{2}\beta A^4 ab(S_x+S_y) = \frac{\pi}{2}\beta R A^2(S_x+S_y) \Longrightarrow \left.\frac{d\hat{L}}{dz}\right|_{z=0} = -\frac{\pi}{2}\beta R(\theta_x+\theta_y/R^2) = \left.\frac{d^2\hat{a}}{dz^2}\right|_{z=0},$$

from which we obtain

$$\left.\frac{d^2\hat{L}}{dz^2}\right|_{z=0} = \pi 2\beta R\left\{-2\left(\theta_x^2 + \frac{\theta_x\theta_y}{R^2} + \frac{\theta_y^2}{R^4}\right) + 1 + \frac{1}{R^4} - \frac{\beta}{8}\left(1 + \frac{1}{R^2}\right)\right\} = \left.\frac{d^3\hat{a}}{dz^3}\right|_{z=0}.$$

Therefore, the action-integral expansion for $N = 0$ takes the form:

$$\hat{a}_o = -\frac{\pi\beta R z}{8}\left\{1 + 2\left(\theta_x + \frac{\theta_y}{R^2}\right)z + z^2\left[\frac{16}{3}\left(\theta_x^2 + \frac{\theta_x\theta_y}{R^2} + \frac{\theta_y^2}{R^4}\right) - \frac{8}{3}\left(1 + \frac{1}{R^4}\right) + \frac{\beta}{3}\left(1 + \frac{1}{R^2}\right)\right] + \ldots\right\}. \qquad \text{(IV.2.17)}$$

This is equivalent to equation IV.2.5 with $\eta_o = \frac{1}{3}$, showing that the $N = 0$ action-integral is larger than that of the exact solution, as expected.

IV.3 Extension of Method to Higher Order (N = 1)

The formal procedure outlined in section IV.2.i suggests that there should be little fundamental difficulty in handling additional parameters present in the higher order beam representation. However, we will find that there are a number of aspects of the problem for $N \geq 1$ which require extensive analysis and result in substantial new features in the solutions. Aside from the massive algebraic efforts required in these higher order solutions, the basic source for concern is the matrix inversion occurring between equations IV.2.11a and IV.2.11c. This matrix, $\hat{d}$, will be shown to be singular for some values of the parameters, so that the general formula IV.2.11c cannot be applied in these cases. The following analysis will illustrate the methods used to obtain the evolution of the $N = 1$ parameterized beam; these methods will apply with only minor changes to higher order solutions as well. Only a few of the specific formulas will be given to illustrate the approaches used, since the ultimate goal is simply to solve for $\alpha_n(z)$, and, thus, $\hat{E}(x,y,\alpha_n(z))$.

Much of the algebra can be performed in terms of the general form for $\hat{E}(x,y,z)$ given in equations IV.1.2a, but some of the solution details require knowledge of the real and imaginary components of $\tau_{m,n}$. We therefore will use an explicit form for $N = 1$ which results from setting $\tau_{1,0}(z) \equiv C(z) + iD(z)$ and $\tau_{0,1}(z) \equiv E(z) + iF(z)$:

$$\hat{E}_1(x,y,z) = Ae^{i\phi}e^{-x^2/a^2}e^{-y^2/b^2}e^{iS_x x^2}e^{iS_y y^2}\left\{1+(C+iD)\frac{x^2}{a^2}+(E+iF)\frac{y^2}{b^2}\right\}, \tag{IV.3.1}$$

where $C(0) = D(0) = E(0) = F(0) = 0$. The form of the equations below should make it clear when "E", for example, refers to the parameter in equation IV.3.1, as compared to the full field in equation IV.1.1; collectively, we will refer to C, D, E, and F as the "minor" parameters, and A, ϕ, a, S_x, b, and S_y as the"major" parameters.

IV.3.i solution for parameter derivatives

Substitution of $\hat{E}_1$, equation IV.3.1, into IV.2.9 yields the Lagrangian density:

$$\hat{L} = L_1 + L_2 + L_3$$

where

$$L_1 = \frac{\pi}{8}\beta A^4 ab \Big\{ 1 + \frac{1}{2}(C+E) + \frac{3}{32}(3C^2+D^2) + \frac{1}{16}(3CE+DF) + \frac{3}{32}(3E^2+F^2)$$

$$+ \frac{15}{128}C(C^2+D^2) + \frac{3}{128}(E(3C^2+D^2)+2CDF) + \frac{3}{128}(C(3E^2+F^2)+2DEF) + \frac{15}{128}E(E^2+F^2)$$

$$+ \frac{105}{4096}(C^2+D^2)^2 + \frac{15}{1024}(C^2+D^2)(CE+DF) + \frac{9}{2048}((C^2+D^2)(E^2+F^2)+2(CE+DF)^2)$$

$$+ \frac{15}{1024}(E^2+F^2)(CE+DF) + \frac{105}{4096}(E^2+F^2)^2 \Big\} \qquad \text{(IV.3.2a)}$$

$$L_2 = \frac{1}{A}\frac{dA}{dz}\{0\} + \frac{d\phi}{dz}(-\frac{\pi}{2}A^2ab)\{1 + \frac{1}{2}(C+E) + \frac{3}{16}(C^2+D^2) + \frac{1}{8}(CE+DF) + \frac{3}{16}(E^2+F^2)\}$$

$$+ \frac{1}{a}\frac{da}{dz}(\frac{\pi}{2}A^2ab)\{\frac{1}{2}D + \frac{1}{8}(DE-CF)\}$$

$$+ a^2\frac{dS_x}{dz}(-\frac{\pi}{8}A^2ab)\{1 + \frac{1}{2}(3C+E) + \frac{5}{16}(C^2+D^2) + \frac{3}{8}(CE+DF) + \frac{3}{16}(E^2+F^2)\}$$

$$+ \frac{1}{b}\frac{db}{dz}(\frac{\pi}{2}A^2ab)\{\frac{1}{2}F + \frac{1}{8}(CF-DE)\}$$

$$+ b^2 \frac{dS_y}{dz} \left(-\frac{\pi}{8} A^2 ab\right)\left\{1 + \frac{1}{2}(C+3E) + \frac{3}{16}(C^2+D^2) + \frac{3}{8}(CE+DF) + \frac{15}{16}(E^2+F^2)\right\}$$

$$+ \frac{dC}{dz} \left(\frac{\pi}{32} A^2 ab\right)\{3D+F\} + \frac{dD}{dz}\left(-\frac{\pi}{8} A^2 ab\right)\left\{1 + \frac{1}{4}(3C+E)\right\}$$

$$+ \frac{dE}{dz} \left(\frac{\pi}{32} A^2 ab\right)\{D+3F\} + \frac{dF}{dz}\left(-\frac{\pi}{8} A^2 ab\right)\left\{1 + \frac{1}{4}(C+3E)\right\} \qquad \text{(IV.3.2b)}$$

$$L_3 = a^2 S_x^2 \left(-\frac{\pi}{2} A^2 ab\right)\left\{1 + \frac{1}{2}(3C+E) + \frac{15}{16}(C^2+D^2) + \frac{3}{8}(CE+DF) + \frac{3}{16}(E^2+F^2)\right\}$$

$$+ b^2 S_y^2 \left(-\frac{\pi}{2} A^2 ab\right)\left\{1 + \frac{1}{2}(C+3E) + \frac{3}{16}(C^2+D^2) + \frac{3}{8}(CE+DF) + \frac{15}{16}(E^2+F^2)\right\}$$

$$+ S_x(-2\pi A^2 ab)\left\{\frac{1}{2}D + \frac{1}{8}(DE-CF)\right\}$$

$$+ S_y(-2\pi A^2 ab)\left\{\frac{1}{2}F + \frac{1}{8}(CF-DE)\right\}$$

$$+ \frac{1}{a^2}\left(-\frac{\pi}{2} A^2 ab\right)\left\{1 + \frac{1}{2}(-C+E) + \frac{7}{16}(C^2+D^2) - \frac{1}{8}(CE+DF) + \frac{3}{16}(E^2+F^2)\right\}$$

$$+ \frac{1}{b^2}\left(-\frac{\pi}{2} A^2 ab\right)\left\{1 + \frac{1}{2}(C-E) + \frac{3}{16}(C^2+D^2) - \frac{1}{8}(CE+DF) + \frac{7}{16}(E^2+F^2)\right\} \qquad \text{(IV.3.2c)}$$

The Euler-Lagrange equations governing beam propagation are then found by applying equation IV.2.10 to the Lagrangian. As expected, the equation generated by "A" gives

$$L_2 = -L_3 - 2L_1 , \qquad \text{(IV.3.3a)}$$

and the "ϕ" equation shows $\frac{dP}{dz} = 0$ where the power, P, is given by

$$P = \frac{\pi}{2} A^2 ab\left\{1 + \frac{1}{2}(C+E) + \frac{3}{16}(C^2+D^2) + \frac{1}{8}(CE+DF) + \frac{3}{16}(E^2+F^2)\right\} \qquad \text{(IV.3.3b)}$$

$$= \text{constant} .$$

As examples of further equations, consider α_n="a" in equation IV.2.10; this shows

$$\frac{1}{A}\frac{dA}{dz}\{\frac{1}{2}D+\frac{1}{8}(DE-CF)\}+\frac{d\phi}{dz}\{\frac{1}{2}+\frac{1}{4}(C+E)+\frac{3}{32}(C^2+D^2)+\frac{1}{16}(CE+DF)+\frac{3}{32}(E^2+F^2)\}$$

$$+\frac{1}{a}\frac{da}{dz}\{0\}+a^2\frac{dS_x}{dz}\{\frac{3}{8}+\frac{3}{16}(3C+E)+\frac{45}{128}(C^2+D^2)+\frac{9}{64}(CE+DF)+\frac{9}{128}(E^2+F^2)\}$$

$$+\frac{1}{b}\frac{db}{dz}\{\frac{1}{4}(D-F)+\frac{1}{8}(DE-CF)\}+b^2\frac{dS_y}{dz}\{\frac{1}{8}+\frac{1}{16}(C+3E)+\frac{3}{128}(C^2+D^2)+\frac{3}{64}(CE+DF)$$

$$+\frac{15}{128}(E^2+F^2)\}$$

$$+\frac{dC}{dz}\{-\frac{3}{32}(D+F)\}+\frac{dD}{dz}\{\frac{3}{8}+\frac{3}{32}(C+E)\}+\frac{dE}{dF}\{\frac{1}{32}(D-3F)\}+\frac{dF}{dz}\{\frac{1}{8}+\frac{1}{32}(-C+3E)\}$$

$$=a^2S_x^2\{-\frac{3}{2}-\frac{3}{4}(3C+E)-\frac{45}{32}(C^2+D^2)-\frac{9}{16}(CE+DF)-\frac{9}{32}(E^2+F^2)\}$$

$$+b^2S_y^2\{-\frac{1}{2}-\frac{1}{4}(C+3E)-\frac{3}{32}(C^2+D^2)-\frac{3}{16}(CE+DF)-\frac{15}{32}(E^2+F^2)\}$$

$$+S_x\{-D-\frac{1}{4}(DE-CF)\}+S_y\{-F-\frac{1}{4}(CF-DE)\}$$

$$+\frac{1}{a^2}\{\frac{1}{2}+\frac{1}{4}(-C+E)+\frac{7}{32}(C^2+D^2)-\frac{1}{16}(CE+DF)+\frac{3}{32}(E^2+F^2)\}$$

$$+\frac{1}{b^2}\{-\frac{1}{2}-\frac{1}{4}(C-E)-\frac{3}{32}(C^2+D^2)+\frac{1}{16}(CE+DF)-\frac{7}{32}(E^2+F^2)\}$$

$$+\frac{\beta A^2}{8}\{1+\frac{1}{2}(C+E)+\frac{3}{32}(3C^2+D^2)+\frac{1}{16}(3CE+DF)+\frac{3}{32}(3E^2+F^2)$$

$$+\frac{15}{128}C(C^2+D^2)+\frac{3}{128}(E(3C^2+D^2)+2CDF)+\frac{3}{128}(C(3E^2+F^2)+2DEF)+\frac{15}{128}E(E^2+F^2)$$

$$+ \frac{105}{4096}(C^2+D^2)^2+ \frac{15}{1024}(C^2+D^2)(CE+DF)+ \frac{9}{2048}\left((C^2+D^2)(E^2+F^2)+2(CE+DF)^2\right)$$

$$+ \frac{15}{1024}(E^2+F^2)(CE+DF) + \frac{105}{4096}\ (E^2+F^2)^2\ \Big\} \ . \qquad \text{(IV.3.4)}$$

The Euler-Lagrange equation generated by the parameter "S_x" is:

$$\frac{1}{A}\frac{dA}{dz}\{1+ \frac{1}{2}(3C+E)+ \frac{15}{16}(C^2+D^2)+ \frac{3}{8}(CE+DF)+ \frac{3}{16}(E^2+F^2)\} + \frac{d\phi}{dz}\{0\}$$

$$+ \frac{1}{a}\frac{da}{dz}\{\frac{3}{2} + \frac{3}{4}(3C+E)+ \frac{45}{32}(C^2+D^2)+ \frac{9}{16}(CE+DF)+ \frac{9}{32}(E^2+F^2)\} + a^2\frac{dS_x}{dz}\{0\}$$

$$+ \frac{1}{b}\frac{db}{dz}\{\frac{1}{2} + \frac{1}{4}(3C+E) + \frac{15}{32}(C^2+D^2)+ \frac{9}{16}(CE+DF)+ \frac{3}{32}(E^2+F^2)\} + b^2\frac{dS_y}{dz}\{0\}$$

$$+ \frac{dC}{dz}\{\frac{3}{4} + \frac{3}{16}(5C+E)\} + \frac{dD}{dz}\{\frac{3}{16}(5D+F)\} + \frac{dE}{dz}\{\frac{1}{4} + \frac{3}{16}(C+E)\} + \frac{dF}{dz}\{\frac{3}{16}(D+F)\}$$

$$= S_x\{4 + 2(3C+E)+ \frac{15}{4}(C^2+D^2)+ \frac{3}{2}(CE+DF)+ \frac{3}{4}(E^2+F^2)\}$$

$$+ \frac{1}{a^2}\{4D + (DE-CF)\} \ . \qquad \text{(IV.3.5)}$$

The Euler-Lagrange equation generated by "C" is:

$$\frac{1}{A}\frac{dA}{dz}\{\frac{1}{4}(3D+F)\} + \frac{d\phi}{dz}\{1+ \frac{1}{4}(3C+E)\} + \frac{1}{a}\frac{da}{dz}\{\frac{3}{8}(D+F)\} + a^2\frac{dS_x}{dz}\{\frac{3}{4} + \frac{3}{16}(5C+E)\}$$

$$+ \frac{1}{b}\frac{db}{dz}\{\frac{1}{8}(3D+F)\} + b^2\frac{dS_y}{dz}\{\frac{1}{4} + \frac{3}{16}(C+E)\} + \frac{dC}{dz}\{0\} + \frac{dD}{dz}\{\frac{3}{4}\}+$$

$$+ \frac{dE}{dz}\{0\} + \frac{dF}{dz}\{\frac{1}{4}\}$$

$$= a^2S_x^2\{-3 - \frac{3}{4}(5C+E)\} + b^2S_y^2\{-1 - \frac{3}{4}(C+E)\} + S_x\{F\} + S_y\{-F\}$$

$$+ \frac{1}{a^2}\{1 - \frac{1}{4}(7C-E)\} + \frac{1}{b^2}\{-1 - \frac{1}{4}(3C-E)\}$$

$$+ \frac{\beta A^2}{4}\left\{1 + \frac{3}{8}(3C+E) + \frac{15}{64}(3C^2+D^2) + \frac{3}{32}(3CE+DF) + \frac{3}{64}(3E^2+F^2)\right.$$

$$+ \frac{105}{512}C(C^2+D^2) + \frac{15}{512}(E(3C^2+D^2) + 2CDF) + \frac{9}{512}(C(3E^2+F^2) + 2DEF) +$$

$$\left. + \frac{15}{512}E(E^2+F^2)\right\} \quad . \qquad \text{(IV.3.6)}$$

Finally, the parameter "D" generates an Euler-Lagrange equation which gives:

$$\frac{1}{A}\frac{dA}{dz}\{1 + \frac{1}{4}(3C+E)\} + \frac{d\phi}{dz}\{-\frac{1}{4}(3D+F)\} + \frac{1}{a}\frac{da}{dz}\{\frac{3}{2} + \frac{3}{8}(C+E)\} +$$

$$a^2\frac{dS_x}{dz}\{-\frac{3}{16}(5D+F)\}$$

$$+ \frac{1}{b}\frac{db}{dz}\{\frac{1}{2} + \frac{1}{8}(3C-E)\} + b^2\frac{dS_y}{dz}\{-\frac{3}{16}(D+F)\} + \frac{dC}{dz}\{\frac{3}{4}\} + \frac{dD}{dz}\{0\} +$$

$$+ \frac{dE}{dz}\{\frac{1}{4}\} + \frac{dF}{dz}\{0\}$$

$$= a^2S_x^2\{\frac{3}{4}(5D+F)\} + b^2S_y^2\{\frac{3}{4}(D+F)\} + S_x\{4+E\} + S_y\{-E\}$$

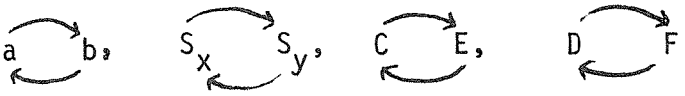
$$+ \frac{1}{a^2}\{\frac{1}{4}(7D-F)\} + \frac{1}{b^2}\{\frac{1}{4}(3D-F)\}$$

$$+ \frac{\beta A^2}{4}\left\{-\frac{1}{8}(3D+F) - \frac{15}{32}CD - \frac{3}{32}(DE+CF) - \frac{3}{32}EF\right.$$

$$\left. - \frac{105}{512}D(C^2+D^2) - \frac{15}{512}(F(C^2+3D^2)+2CDE) - \frac{9}{512}(D(E^2+3F^2)+2CEF) - \frac{15}{512}F(E^2+F^2)\right\}.$$

(IV.3.7)

The equations corresponding to "b", "S_y", "E", and "F" can be found from IV.3.4 through IV.3.7 respectively, by using the symmetry operation:



(IV.3.8)

Equations IV.3.3 through IV.3.8 represent 10 linear equations in the 10 "unknowns", $\frac{1}{A}\frac{dA}{dz}$, $\frac{d\phi}{dz}$, $\frac{1}{a}\frac{da}{dz}$, $a^2\frac{dS_x}{dz}$, $\frac{1}{b}\frac{db}{dz}$, $b^2\frac{dS_y}{dz}$, $\frac{dC}{dz}$, $\frac{dD}{dz}$, $\frac{dE}{dz}$, and $\frac{dF}{dz}$; there appears to be little need for further analysis, since the (10 x 10) matrix of derivative coefficients can, in theory, be inverted by the computer to produce a set of 1st order, ordinary differential equations. In this case, however, such an approach fails to yield reasonable results for two reasons. First, the matrix inversion process is computationally inaccurate, since there will be strong algebraic cancellation in forming the determinant, $DET|\hat{d}|$, for any values of C, D, E, and F. Second, a more fundamental problem occurs at particular values of C, D, E, and F where $DET|\hat{d}| = 0$; the matrix, of course, cannot be inverted under these conditions. We will refer to "points" (i.e. particular sets of parameter values) where $DET|\hat{d}| = 0$ as "degenerate points", since, in the next section, we will show that they actually lead to multiple solutions.

To illustrate both the significance and the unusual nature of this "degeneracy" situation, we can observe that the N = 0 solution defined by equation IV.2.16 satisfies the N = 1 equations above with $C(z) = D(z) = E(z) = F(z) = 0$, $z \geq 0$; the initial conditions are also met by such a solution. Since the above equations are first order and linear (in the derivatives), it would be "natural" to assume that this represents "the unique N = 1 solution". However, the matrix of derivative coefficients when $C = D = E = F = 0$ is singular (for

example, equations IV.3.5 and IV.3.7 have the same coefficients and are, in fact, redundant in that limit); we therefore cannot exclude the possibility of other solutions for $N = 1$ without further examination of the behavior of the parameters when $\mathrm{DET}|\hat{d}| \to 0$.

To simplify the investigation of the "degenerate point solutions", and to provide general formulas for the parameter derivatives which are computationally accurate, we will now algebraically invert the matrix $\hat{d}$. As a preliminary step, the form of equation IV.2.11c and of the Euler-Lagrange equations for $N = 1$ indicates that the "linear" and "nonlinear" contributions to the derivatives can be separated. The linear terms can be found by considering the exact solutions in equation III.1.7, by solving the "near-axis" equations for $N = 1$ (with $\beta = 0$), or by carrying out the analysis described below for the nonlinear terms; in any case, we find that the derivatives can be expressed in the form:

$$
\begin{aligned}
\frac{1}{A}\frac{dA}{dz} &= -2S_x - 2S_y - \frac{2D}{a^2} - \frac{2F}{b^2} && + \beta A^2 W_A \\
\frac{d\phi}{dz} &= \frac{1}{a^2}\{-2+2C\} + \frac{1}{b^2}\{-2+2E\} && + \beta A^2 W_\phi \\
\frac{1}{a}\frac{da}{dz} &= 4S_x && + \beta A^2 W_a \\
a^2\frac{dS_x}{dz} &= \frac{4}{a^2} - 4a^2 S_x^2 && + \beta A^2 W_{S_x} \\
\frac{1}{b}\frac{db}{dz} &= 4S_y && + \beta A^2 W_b
\end{aligned}
\qquad \text{(IV.3.9)}
$$

$$b^2 \frac{dS_y}{dz} = \frac{4}{b^2} - 4b^2 S_y^2 + \beta A^2 W_b$$

$$\frac{dC}{dz} = \frac{1}{a^2}\{8D+4CD\} + \frac{1}{b^2}\{2CF+2DE\} + \beta A^2 W_C$$

$$\frac{dD}{dz} = \frac{1}{a^2}\{-8C-2C^2+2D^2\} + \frac{1}{b^2}\{2DF-2CE\} + \beta A^2 W_D \qquad \text{(IV.3.9) cont'd.}$$

$$\frac{dE}{dz} = \frac{1}{a^2}\{2CF+2DE\} + \frac{1}{b^2}\{8F+4EF\} + \beta A^2 W_E$$

$$\frac{dF}{dz} = \frac{1}{a^2}\{2DF-2CE\} + \frac{1}{b^2}\{-8E-2E^2+2F^2\} + \beta A^2 W_F \quad .$$

Here, W_A, W_ϕ, W_a, etc. are algebraic functions of C, D, E, F only. With these substitutions, the Euler-Lagrange equations reduce to 10 equations in the 10 unknowns, $\frac{d\alpha}{dz} \rightarrow W_\alpha$, with only the nonlinear parts of the right hand side remaining.

Equations IV.3.6, IV.3.7, and their symmetry counterparts, contain the four minor derivatives with constant coefficients; thus, they can be easily combined to "solve" for the minor derivatives in terms of the six "major" derivatives. For example, using equation IV.3.9 and forming the combination: {3/2 x [equation IV.3.7]-1/2 x [symmetrical counterpart to IV.3.7]} yields the solution for W_C,

$$W_C = W_A\{-1-C\} + W_\phi\{D\} + W_a\{-2-\frac{1}{8}(5C+3E)\} + W_{S_x}\{\frac{3}{16}(7D+F)\}$$
$$+ W_b\{-\frac{3}{8}(C-E)\} + W_{S_y}\{\frac{3}{16}(D-F)\} \qquad \text{(IV.3.10a)}$$

$$+ \{-\frac{1}{8}D - \frac{21}{128}CD - \frac{3}{128}(DE+CF) + \frac{3}{128}EF$$

$$- \frac{75}{1024}D(C^2+D^2) - \frac{9}{1024}(F(C^2+3D^2)+2CDE) - \frac{3}{1024}(D(E^2+3F^2)+2CEF)$$

$$+ \frac{15}{1024}F(E^2+F^2)\} \qquad \text{(IV.3.10a) cont'd.}$$

Similarly, from equation IV.3.6 we can obtain:

$$W_D = W_A\{-D\} + W_\phi\{-1-C\} + W_a\{-\frac{1}{8}(5D+3F)\} + W_{S_x}\{-1-\frac{3}{16}(7C+E)\}$$

$$+ W_b\{-\frac{3}{8}(D-F)\} + W_{S_y}\{-\frac{3}{16}(C-E)\}$$

$$+ \{\frac{1}{4} + \frac{3}{8}C + \frac{21}{256}(3C^2+D^2) + \frac{3}{128}(3CE+DF) - \frac{3}{256}(3E^2+F^2)$$

$$+ \frac{75}{1024}C(C^2+D^2) + \frac{9}{1024}(E(3C^2+D^2)+2CDF) + \frac{3}{1024}(C(3E^2+F^2)+2DEF)$$

$$- \frac{15}{1024}E(E^2+F^2)\} . \qquad \text{(IV.3.10b)}$$

Application of equation IV.3.8 then yields the corresponding solutions for W_E and W_F.

Using these expressions to eliminate W_C, W_D, W_E, and W_F from equations IV.3.3b and IV.3.3a leads to similar formulas for W_A and W_ϕ, respectively:

$$W_A = W_a\{\frac{1}{8}(3C+E)\} + W_{S_x}\{-\frac{1}{16}(3D+F)\} + W_b\{\frac{1}{8}(C+3E)\} + W_{S_y}\{-\frac{1}{16}(D+3F)\}$$

$$+ \{-\frac{1}{16}(D+F) - \frac{3}{128}CD - \frac{1}{128}(DE+CF) - \frac{3}{128}EF\} \qquad \text{(IV.3.11a)}$$

and

$$W_\phi = W_a\{\tfrac{1}{8}(3D+F)\} + W_{S_x}\{\tfrac{1}{16}(3C+E)\} + W_b\{\tfrac{1}{8}(D+3F)\} + W_{S_y}\{\tfrac{1}{16}(C+3E)\}$$

$$+ \{\tfrac{3}{4} + \tfrac{3}{16}(C+E) + \tfrac{3}{256}(3C^2+D^2) + \tfrac{1}{128}(3CE+DF) + \tfrac{3}{256}(3E^2+F^2)\}. \qquad \text{(IV.3.11b)}$$

Next, the use of equations IV.3.10 and IV.3.11 in equations IV.3.4 and IV.3.5 yields four equations in the four remaining unknowns, W_a, W_{S_x}, W_b, and W_{S_y}, which we will write symbolically as:

$$\begin{aligned}
W_a\{0\} + W_{S_x}\{\tfrac{1}{2}m_1\} + W_b\{m_2\} + W_{S_y}\{\tfrac{1}{2}m_3\} &= B_1 \\
W_a\{m_1\} + W_{S_x}\{0\} + W_b\{m_3\} + W_{S_y}\{-\tfrac{1}{2}m_2\} &= B_2 \\
W_a\{-m_2\} + W_{S_x}\{\tfrac{1}{2}m_3\} + W_b\{0\} + W_{S_y}\{\tfrac{1}{2}m_4\} &= B_3 \\
W_a\{m_3\} + W_{S_x}\{\tfrac{1}{2}m_2\} + W_b\{m_4\} + W_{S_y}\{0\} &= B_4
\end{aligned} \qquad \text{(IV.3.12)}$$

where direct calculation shows:

$$\begin{aligned}
m_1 &= 6(C^2+D^2) + (E^2+F^2) \\
m_2 &= (DE-CF) \\
m_3 &= (CE+DF) \\
m_4 &= (C^2+D^2) + 6(E^2+F^2) \quad .
\end{aligned} \qquad \text{(IV.3.13)}$$

Here,

$$B_1 = \frac{1}{4}(3C+E) - \frac{9}{32}(3C^2+D^2) + \frac{1}{16}(3CE+DF) - \frac{1}{32}(3E^2+F^2)$$

$$- \frac{117}{256}C(C^2+D^2) - \frac{27}{256}(E(3C^2+D^2)+2CDF) + \frac{11}{256}(C(3E^2+F^2)+2DEF) - \frac{27}{256}E(E^2+F^2)$$

$$- \frac{135}{2048}(C^2+D^2)^2 - \frac{3}{256}(C^2+D^2)(CE+DF) - \frac{9}{1024}((C^2+D^2)(E^2+F^2)+2(CE+DF)^2)$$

$$+ \frac{9}{256}(E^2+F^2)(CE+DF) - \frac{75}{2048}(E^2+F^2)^2 \qquad \text{(IV.3.14a)}$$

and

$$B_2 = \frac{1}{4}(3D+F) - \frac{9}{16}CD + \frac{1}{16}(5DE-3CF) - \frac{1}{16}EF$$

$$- \frac{39}{256}D(C^2+D^2) - \frac{9}{256}(F(C^2+3D^2)+2CDE) + \frac{1}{256}(D(25E^2+11F^2)-14CEF)$$

$$- \frac{9}{256}F(E^2+F^2) - \frac{9}{512}((C^2+D^2)-2(CE+DF)+(E^2+F^2))(CF-DE) \quad . \qquad \text{(IV.3.14b)}$$

B_3 and B_4 are found from symmetry from B_1 and B_2, respectively.

The inverse of the matrix in equation IV.3.12 is easily found. First, the determinant, which differs from the determinant of the original 10 x 10 matrix by only a constant factor, is given by $DET = d_o^2$, where

$$d_o = (m_1 m_4 - m_2^2 - m_3^2)/2 \qquad \text{(IV.3.15)}$$

$$= 3(C^2+D^2)^2 + 18(C^2+D^2)(E^2+F^2) + 3(E^2+F^2)^2 \quad .$$

Note that this determinant contains only eighth "net powers" of the minor variables, (i.e. C^8, C^6D^2, etc.) whereas an examination of the ten original Euler-Lagrange equations suggests that the determinant

could contain powers ranging from zero (i.e. constants) through 12th; this "cancellation" of many "orders of terms" shows why the direct computational inversion of the original matrix would be very inaccurate, particularly if the minor variables were either very small or very large.

Solving equation IV.3.14, we obtain formulas for computing W_a, W_{S_x}, W_b, and W_{S_y}:

$$\begin{aligned} W_a &= \tfrac{1}{2}(m_4B_2 + m_2B_3 - m_3B_4)/d_o \\ W_{S_x} &= (m_4B_1 - m_3B_3 - m_2B_4)/d_o \\ W_b &= \tfrac{1}{2}(-m_2B_1 - m_3B_2 + m_1B_4)/d_o \\ W_{S_y} &= (-m_3B_1 + m_2B_2 + m_1B_3)/d_o \quad . \end{aligned} \tag{IV.3.16}$$

The following "recipe" thus yields a concise computational scheme for calculating the parameter derivatives: first, form W_a, W_{S_x}, W_b, and W_{S_y} from equations IV.3.13, through IV.3.16; next, find W_A and W_ϕ from IV.3.11; then form W_C, W_D, W_E, and W_F from equation IV.3.10 and its symmetrical counterpart; finally, form the derivatives by using equation IV.3.9. Of course, explicit formulas could be obtained by carrying out these steps algebraically; however, the results obtained in this way are so complex, and show so little tendency toward "simplification", that the "recipe" method must be considered equally accurate and far more efficient.

IV.3.ii multiple solutions near degenerate point

The form of the determinant in equation IV.3.15 indicates that the "recipe" fails only at the "degenerate point" given by C = D = E = F = 0; i.e. the propagation of the parameterized beam (for N = 1) follows a well defined formula except when the "instantaneous" (in z) shape is exactly a "$E_{o,o}$ Gaussian". As noted earlier, this "degenerate point" includes the initial conditions at z = 0; from our numerical experience, it also occurs at later points for most initial conditions of interest. Therefore, we will now examine in some detail the solution (s) near such a "degenerate point", as defined by

$$A(z) \to A_o \qquad \phi(z) \to \phi_o$$

$$a(z) \to a_o \qquad b(z) \to b_o$$

$$S_x(z) \to S_x^o \qquad S_y(z) \to S_y^o \qquad \text{as } z \to 0 \,. \qquad \text{(IV.3.17)}$$

$$C(z) \to 0 \qquad E(z) \to 0$$

$$D(z) \to 0 \qquad F(z) \to 0 \quad .$$

We have taken $z_{degenerate} = 0$ for simplicity, since the results can always be generalized by setting $z \to (z - z_{degenerate})$.

A consideration of the "recipe" in the previous section indicates that the parameter derivatives may diverge as $z \to 0$, even though the limiting values given above are all finite; such behavior occurs for functions behaving as z^γ, with $0 < \gamma < 1$.

This is the motivation for the following analysis.

To obtain the parameter behavior near $z = 0$, we assume that C, D, E, and F can all be expanded in powers (perhaps fractional) of z; we then consider the lowest leading power, called γ

$$C(z) = C_0 z^{\gamma} + \cdots, \quad D(z) = D_0 z^{\gamma} + \cdots, \quad E(z) = E_0 z^{\gamma} + \cdots, \quad F(z) = F_0 z^{\gamma} + \cdots$$

$$0 < \gamma \tag{IV.3.18}$$

By definition, at least one of C_0, D_0, E_0, or F_0 is nonzero; we thus exclude the N = 0 solution mentioned earlier, in which $C(z) = D(z) = E(z) = F(z) = 0$, and concentrate on finding any new solutions generated by the N = 1 equations. By substituting these expressions into the "recipe", we can determine the leading terms of each parameter derivative. (The actual leading term may have a zero coefficient, but it cannot have a lower power of z). Thus, from equations IV.3.13 through IV.3.15, we obtain, in turn:

$$m_1(z) = z^{2\gamma}\{6(C_0^2+D_0^2) + (E_0^2+F_0^2)\} + \cdots$$

$$m_2(z) = z^{2\gamma}\{D_0E_0 - C_0F_0\} + \cdots$$

$$m_3(z) = z^{2\gamma}\{C_0E_0+D_0F_0\} + \cdots$$

$$m_4(z) = z^{2\gamma}\{(C_0^2+D_0^2) + 6(E_0^2+F_0^2)\} + \cdots$$

$$B_1(z) = z^{\gamma}\{\tfrac{1}{4}(3C_0+E_0)\} + \cdots$$

$$B_2(z) = z^{\gamma}\{\tfrac{1}{4}(3E_0+F_0)\} + \cdots$$

$$B_3(z) = z^{\gamma}\{\frac{1}{4}(C_o+3E_o)\}+ \cdots$$

$$B_4(z) = z^{\gamma}\{\frac{1}{4}(D_o+3F_o)\}+ \cdots$$

and

$$d_o(z) = z^{4\gamma}\{3(C_o^2+D_o^2)^2 + 18(C_o^2+D_o^2)(E_o^2+F_o^2) + 3(E_o^3+F_o^2)^2\}+ \cdots$$

$$\equiv z^{4\gamma}\bar{d}_o + \cdots \quad . \qquad \text{(IV.3.19)}$$

Since $\bar{d}_o$ must be nonzero, we have, from equation IV.3.16,

$$W_a(z)= z^{-\gamma}\{\frac{3}{8}(D_o(C_o^2+D_o^2)+2F_o(E_o^2+F_o^2)+D_o(7E_o^2+5F_o^2)-2C_oE_oF_o)/\bar{d}_o\}+\ldots\equiv z^{-\gamma}\bar{W}_a+\cdots$$

$$W_{S_x}(z)= z^{-\gamma}\{\frac{3}{4}(C_o(C_o^2+D_o^2)+2E_o(E_o^2+F_o^2)+C_o(5E_o^2+7F_o^2)-2D_oE_oF_o)/\bar{d}_o\}+\ldots \equiv$$

$$\equiv z^{-\gamma}\bar{W}_{S_x} + \cdots$$

$$W_b(z)=z^{-\gamma}\{\frac{3}{8}(F_o(E_o^2+F_o^2)+2D_o(C_o^2+D_o^2)+F_o(7C_o^2+5D_o^2)-2C_oD_oE_o)/\bar{d}_o\}+\ldots\equiv z^{-\gamma}\bar{W}_b+\ldots$$

$$W_{S_y}(z)= z^{-\gamma}\{\frac{3}{4}(E_o(E_o^2+F_o^2)+2C_o(C_o^2+D_o^2)+E_o(5C_o^2+7D_o^2)-2C_oD_oF_o)/\bar{d}_o\}+\ldots \equiv$$

$$\equiv z^{-\gamma}\bar{W}_{S_y}+\cdots \quad . \qquad \text{(IV.3.20)}$$

Next, from equations IV.3.11, we obtain the leading terms:

$$W_A = z^o\{0\}$$

$$W_\phi = z^o\{\frac{53}{64} + \frac{9}{16}[(C_o^2+D_o^2+E_o^2+F_o^2)(C_oE_o+D_oF_o) + (D_oE_o- C_oF_o)^2]\}+\ldots \quad .$$

Finally, equation IV.3.10 yields

$$W_C = z^{-\gamma}\bar{W}_C+ \ldots \qquad \text{where} \quad \bar{W}_C =-2\bar{W}_a$$

$$W_D = z^{-\gamma}\bar{W}_D+ \ldots \qquad \text{where} \quad \bar{W}_D =-\bar{W}_{S_x} \qquad \text{(IV.3.21)}$$

$$W_E = z^{-\gamma}\bar{W}_E + \ldots \qquad \text{where} \quad \bar{W}_E = -2\bar{W}_b$$

$$W_F = z^{-\gamma}\bar{W}_F + \ldots \qquad \text{where} \quad \bar{W}_F = -\bar{W}_{S_y} \qquad \text{(IV.3.21) cont'd.}$$

Here $\bar{W}_a$, $\bar{W}_{S_x}$, etc. are defined in equation IV.3.20.

Equation IV.3.9 can now be employed to yield a set of self-consistent relations which C_o, D_o, E_o, F_o (and γ) must obey: thus, for example, the "C" derivative implies

$$z^{\gamma-1}\{C_o\gamma\} + \ldots = z^{-\gamma}\{\beta A_o^2 \bar{W}_C\} + \ldots ,$$

with similar formulas for D_o, E_o, F_o. It can easily be shown that $\bar{W}_C, \bar{W}_D, \bar{W}_E, \bar{W}_F$ cannot all vanish (unless $C_o = D_o = E_o = F_o = 0$), so that there will be at least one term on the right hand sides with the minimum power, $z^{-\gamma}$; likewise, there will be at least one left hand side with terms starting at $z^{\gamma-1}$. A moment's reflection shows that this can only hold if $\gamma - 1 = -\gamma$ so that $\gamma = \frac{1}{2}$ must apply. We can now write down the algebraic equations which must be satisfied by (real) C_o, D_o, E_o and F_o:

$$\text{(i)} \quad C_o\{\tfrac{2}{3}\bar{d}_o/(\beta A_o^2)\} = 2C_oE_oF_o - D_o(7E_o^2+5F_o^2) - 2F_o(E_o^2+F_o^2) - D_o(C_o^2+D_o^2)$$

$$\text{(ii)} \quad D_o\{\tfrac{2}{3}\bar{d}_o/(\beta A_o^2)\} = 2D_oE_oF_o - C_o(5E_o^2+7F_o^2) - 2E_o(E_o^2+F_o^2) - C_o(C_o^2+D_o^2)$$

$$\text{(iii)} \quad E_o\{\tfrac{2}{3}\bar{d}_o/(\beta A_o^2)\} = 2C_oD_oE_o - F_o(7C_o^2+5D_o^2) - 2D_o(C_o^2+D_o^2) - F_o(E_o^2+F_o^2)$$

$$\text{(iv)} \quad F_o\{\tfrac{2}{3}\bar{d}_o/(\beta A_o^2)\} = 2C_oD_oF_o - E_o(5C_o^2+7D_o^2) - 2C_o(C_o^2+D_o^2) - E_o(E_o^2+F_o^2) \qquad \text{(IV.3.22)}$$

where $\bar{d}_o$ is defined in equation IV.3.19.

The following steps will indicate how these four simultaneous

equations in four unknowns can be solved. First, the combination: $\{E_o\cdot(i)-F_o\cdot(ii)-C_o\cdot(iii)+D_o\cdot(iv)\}$ reduces to the result:

$$0 = 8(C_o^2+D_o^2+E_o^2+F_o^2)(C_oF_o-D_oE_o) \Longrightarrow C_oF_o = D_oE_o \qquad \text{(IV.3.23)}$$

With the help of this relation, it is easy to show that none of the values, C_o, D_o, E_o, or F_o can be zero, unless they all are.

Next, forming the combinations $\{C_o^3(D_o\cdot(i)-C_o\cdot(ii))\}$ and $\{C_o^4(F_o\cdot(iii)-E_o\cdot(iv))\}$ yields, after using equation IV.3.23 to eliminate F_o:

$$0 = (C_o^2+D_o^2)(C_o^2-D_o^2)\{C_o^3+5C_oE_o^2+2E_o^3\}$$

$$0 = E_o(C_o^2+D_o^2)(C_o^2-D_o^2)\{2C_o^3+5C_o^2E_o+E_o^3\}\ .$$

If $C_o^2 \neq D_o^2$, these can hold only if each bracketed expression vanishes; adding them shows that $0 = (C_o+E_o)(3E_o^2+2C_oE_o+3C_o^2) \Longrightarrow$ $E_o = -C_o$, which cannot be reconciled with either bracket vanishing (for nonzero C_o). Thus $C_o^2 = D_o^2$ must hold, and, similarly $E_o^2=F_o^2$. The combination, $\{C_o(E_o\cdot(i)-C_o(iii))\}$, can now be shown to be equivalent to

$$0 = 4D_o(C_o+E_o)^3(C_o-E_o),$$

from which we draw the conclusion:

$$C_o^2 = D_o^2 = E_o^2 = F_o^2\ .$$

It is now a simple matter to test the possible cases in equation IV.3.22 with $\bar{d}_o = 96C_o^4$, and conclude that there are precisely

four solutions:

$$(C_o,D_o,E_o,F_o) = \begin{cases} \text{(I):} & (1,-1,1,-1)\cdot\frac{1}{2}(\beta A_o^2)^{1/2} \\ \text{(II):} & (-1,1,-1,1)\cdot\frac{1}{2}(\beta A_o^2)^{1/2} \\ \text{(III):} & (1,-1,-1,1)\cdot\frac{1}{2\sqrt{2}}(\beta A_o^2)^{1/2} \\ \text{(IV):} & (-1,1,1,-1)\cdot\frac{1}{2\sqrt{2}}(\beta A_o^2)^{1/2} \end{cases} \qquad \text{(IV.3.24)}$$

By substituting values for C_o,D_o,E_o and F_o, into equations IV.3.18 through IV.3.21, with $\gamma = 1/2$, we can find the leading terms of the power series expansions (in $z^{1/2}$) of the parameters near the degenerate point; further terms can be found in the usual way, although the algebra becomes very tedious. Each set of values for (C_o,D_o,E_o,F_o) generates a solution; we will label them $1_I,1_{II},1_{III}$, and 1_{IV} corresponding to the four N = 1 cases in equation IV.3.24, respectively. As an example, it can be shown that the following parameter expansions are produced by solution 1_I.

$$\begin{aligned} A(z) &= A_o[1-2(S_x^o+S_y^o)z + \frac{\sqrt{2}}{16}(\frac{\beta A_o^2}{2}z)^{3/2} + \ldots] \\ \phi(z) &= \phi_o + z(-\frac{2}{a_o^2} - \frac{2}{b_o^2} + \frac{7}{8}\beta A_o^2) + \frac{3\sqrt{2}}{16}(\frac{\beta A_o^2}{2}z)^{3/2} + \ldots \\ a(z) &= a_o[1 - \frac{\sqrt{2}}{4}(\frac{\beta A_o^2}{2}z)^{1/2} + z(4S_x^o + \frac{\beta A_o^2}{32}) + \ldots] \\ S_x(z) &= S_x^o + \frac{\sqrt{2}}{2a_o^2}(\frac{\beta A_o^2}{2}z)^{1/2} + z(\frac{4}{a_o^4} - 4(S_x^o)^2 - \frac{\beta A_o^2}{4a_o^2}) + \ldots \end{aligned} \qquad \text{(IV.3.25)}$$

$$b(z) = b_o[1 - \frac{\sqrt{2}}{4}(\frac{\beta A_o^2}{2} z)^{1/2} + z(4S_y^o + \frac{\beta A_o^2}{32}) + \ldots]$$

$$S_y(z) = S_y^o + \frac{\sqrt{2}}{2b_o^2}(\frac{\beta A_o^2}{2} z)^{1/2} + z(\frac{4}{b_o^4} - 4(S_y^o)^2 - \frac{\beta A_o^2}{4b_o^2}) + \ldots$$

$$C(z) = \frac{\sqrt{2}}{2}(\frac{\beta A_o^2}{2} z)^{1/2} - \frac{\beta A_o^2}{8} z + \ldots$$

$$D(z) = -\frac{\sqrt{2}}{2}(\frac{\beta A_o^2}{2} z)^{1/2} - \frac{\beta A_o^2}{2} z + \ldots \qquad \text{(IV.3.25) cont'd.}$$

$$E(z) = \frac{\sqrt{2}}{2}(\frac{\beta A_o^2}{2} z)^{1/2} - \frac{\beta A_o^2}{8} z + \ldots$$

$$F(z) = -\frac{\sqrt{2}}{2}(\frac{\beta A_o^2}{2} z)^{1/2} - \frac{\beta A_o^2}{2} z + \ldots \quad .$$

Actually, it can be seen that no overall factors of $\sqrt{2}$ are needed, but the above form has the advantage of producing solution 1_{II} by replacing "$\sqrt{2}$" $\rightarrow$ "$-\sqrt{2}$" . When the action integral is expanded according to the methods discussed in section IV.2.i, we obtain, for solution 1_I:

$$\hat{a}(z) = -\frac{\pi}{8}\beta A_o^2 a_o b_o z\{1 - 2(S_x^o + S_y^o)z + z^2[\frac{16}{3}((S_x^o)^2 + S_x^o S_y^o + (S_y^o)^2) - \frac{8}{3}(\frac{1}{a_o^4} + \frac{1}{b_o^4}) + \frac{\beta A_o^2}{2}(\frac{1}{a_o^2} + \frac{1}{b_o^2})] + \frac{\sqrt{2}}{7}(\frac{\beta A_o^2}{2})^{3/2}(\frac{1}{a_o^2} + \frac{1}{b_o^2})z^{5/2} + \ldots\} \qquad \text{(IV.3.26)}$$

This has the general form given in equation IV.2.5 (for the initial conditions in IV.1.2b) with $\eta_{1_I} = \eta_{1_{II}} = 1/2$; from the last term in IV.3.26, we can see that solution 1_I has a smaller action integral than solution 1_{II}, and is thus the "preferred" solution.

Similar expansions can be found for solutions 1_{III} and 1_{IV}; in particular, the action integral expansion again has the

standard form with $\eta_{1_{III}} = \eta_{1_{IV}} = \frac{3}{8}$. Thus, out of all the N = 1 solutions, (including the N = 0 limit where $\eta_0 = 1/3$), solution 1_I has the minimum action.

IV.4 Parameterized Solutions

In the last section, we concluded that the N = 1 "solution", 1_I (and, in fact 1_{II} etc.) represented an improvement over that for N = 0, because it led to an action integral which was closer to the exact value. However, these higher order solutions involved some unusual properties, such as singular matrices, divergent derivatives, multiple solutions, etc., which are rather surprising in a problem which is physically "well-behaved" (away from the focal point). To further confirm that these new solutions are, indeed, significant, the parameterization has been extended to N = 2, leading to the results summarized in Appendix 2. Although the algebra required to derive these N = 2 formulas is far more complex than needed for N = 1, the analysis is seen to follow precisely the same general steps; furthermore, most of the qualitative results are very similar for both orders, so that many of the following comments can be expected to apply to higher order cases as well.

As shown in equation IV.1.2a, the extension from N = 1 to N = 2 requires six new parameters, represented by the real and imaginary parts of $\tau_{2,0}$, $\tau_{1,1}$, and $\tau_{0,2}$; the "recipe" and the "degenerate point" expansions given in Appendix 2 involve solutions in which at least one of these six "N = 2 parameters" is nonzero.

In addition to these solutions, it can be shown that the "N-1 recipe" (and the resultant expansions) provide another solution to the N=2 Euler-Lagrange equations under the condition that $\tau_{2,0}(z) = \tau_{1,1}(z) = \tau_{0,2}(z) = 0$; however, the N=0 solution is no longer consistent with the N=2 Euler-Lagrange equations, even if $\tau_{m,n}(z) = 0$ for all $(m,n) \neq (0,0)$. Thus, even if the "action-integral minimization" arguments in section IV.3 were not sufficient proof that the N=1 solution was an improvement over the N=0 result, we would be forced to admit that this was true, as soon as the N=2 parameterization was attempted.

As described in Appendix 2, the "degenerate point" for N=2 occurs when $\tau_{2,0} = \tau_{1,1} = \tau_{0,2} = 0$; if, at such a point, we also have $\tau_{1,0} = \tau_{0,1} = 0$ (which occurs at the initial conditions), then there are precisely 9 "solutions": $2_I, 2_{II}, \ldots, 2_{IX}$. The leading powers in the expansions are: $z^{1/3}$ for $\tau_{1,0}$ and $\tau_{0,1}$, and $z^{2/3}$ for $\tau_{2,0}, \tau_{1,1}$ and $\tau_{0,2}$; for each solution, the action-integral expansion takes the form in equation IV.2.5, with $\eta_{2_I} = \eta_{2_{II}} = \eta_{2_{III}} = 9/16$, and $\eta_{2_{IV}} = \eta_{2_V} = \ldots = \eta_{2_{IX}} = 41/80$. As shown in Figure 23, we can "order" the solutions according to their action-integrals (ignoring the leading terms which are always the same), and also "classify" the solutions into subsets according to η and certain symmetries.

Before considering quantitative properties of the beam propagation as governed by the parameterized solutions, we can further demonstrate that the higher order solutions display a number of useful properties. First, the field, $\hat{E}(x,y,z)$, given in equation IV.1.2a, is bounded, continuous, well-defined, and other-

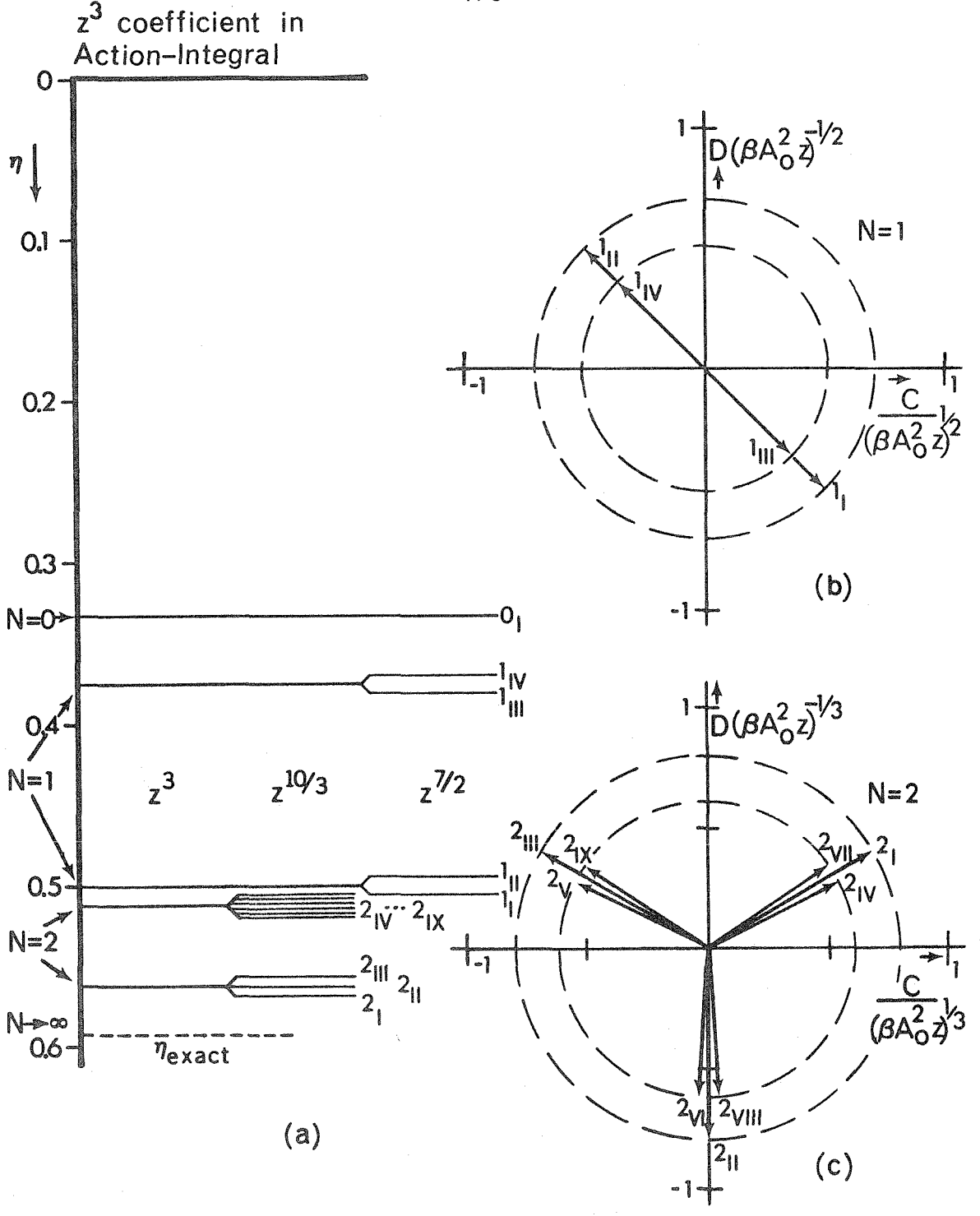


Figure 23. Degenerate "parameterized beam" solutions.
(a) Ordering of solutions through the Action-Integral; the mutual ordering within some subgroups is variable.
(b),(c) Subgroup symmetries as shown by plotting the leading terms of "C" vs. "D".

wise "reasonable", as long as the parameters evolve smoothly; examination of the "recipes" in section IV.3.i and Appendix 2 shows that the only problems occur at the "degenerate points" or when the beam focuses ($a^2, b^2 \to 0$, $A^2 \to \infty$). This latter problem, however, simply reflects the fact that equation IV.1.1 can produce singularities in $E(x,y,z)$; this difficulty, therefore, lies not in the parameterized beam approach, but in the basic problem formulation. The degeneracy point difficulty appears, at first, to be more fundamental, since the parameter derivatives can diverge, and the evolution is not uniquely determined (unless the "least action" criteria is employed). Nevertheless, the following can be shown to hold for a z, including any degenerate points, and for all solutions of a given order: the field intensity, $\hat{I}(x,y,z) = |\hat{E}(x,y,z)|^2$ and its derivative, $\frac{\partial \hat{I}}{\partial z}$, are bounded and continuous, with a bounded derivative. Thus, the mathematical problems associated with the evolution of the individual parameters are not reflected in the physically "measurable" beam, $\hat{E}(x,y,z)$.

Finally, if we consider a beam which is initially symmetrical:

$$a_o = b_o, \qquad S_x^o = S_y^o ,$$

we see that the N=1 solutions 1_I and 1_{II} will continue to be symmetrical, while solutions 1_{III} and 1_{IV} will "break" the symmetry; similarly, for N=2, solutions $2_I, 2_{II}$, and 2_{III} are "symmetrical" while 2_{IV} through 2_{IX} are not. Since, for each order, every symmetrical solution has a lower action-integral than any non-symmetrical solution, we can conclude (by extrapolating

$N \to \infty$) that a symmetrical beam is a stable configuration, on the basis of an "action-minimization" criterion.

IV.4.i self-trapped solutions

A further test of the least-action approach to determining the "best" beam parameters is given by attempting to match the shape-preserving solution discussed in section III.3.iii. Since the lens transformation is satisfied by our solutions, we will consider the "self-trapped" solutions to be defined by parameter values for which

$$\frac{d\hat{E}}{dz} = i\left(\frac{d\phi}{dz}\right)\hat{E} \quad . \tag{IV.4.1a}$$

Thus, a self-trapped solution occurs when all parameter derivatives are zero, except for $\frac{d\phi}{dz}$. Furthermore, to be consistent with the solutions generated by equation III.3.11b, we will take $\hat{E}$ to be symmetrical and real (except for $e^{i\phi}$), with $\frac{d\phi}{dz} = 1$ and $\beta = 1$; so that $\hat{E} = e^{iz}u(r)$, with $r^2 = x^2+y^2$. From the derivation of equation III.3.12, we then have

$$P_{mode}/P_o = \frac{1}{2}\int_0^\infty u^2(r)rdr \quad . \tag{IV.4.1b}$$

Under these conditions, equation IV.2.16 then yields the unique solution for $N = 0$

$$\left.\begin{array}{l} A = 2.0 \\ a^2(=b^2) = 2.0 \\ (S_x = S_y = 0) \end{array}\right\} \quad [u_o(r)]_{N=0} = 2e^{-r^2/2} \tag{IV.4.2}$$

In Figure 24, this shape preserving "mode" is compared with the exact solution, $u_0(r)$, as defined in section III.3.iii.

Next, if we set $\frac{d\alpha}{dz} = \delta_{\alpha,\phi}$ and $\beta = 1$ in equation IV.3.9, with W_α given by the "recipe" in section IV.3.i, we obtain a set of equations for the parameters describing the N = 1 self-trapped modes. One solution to these is given, as expected, by the N = 0 result with C = D = E = F = 0; in addition, there is one other solution, which was found numerically to be described by

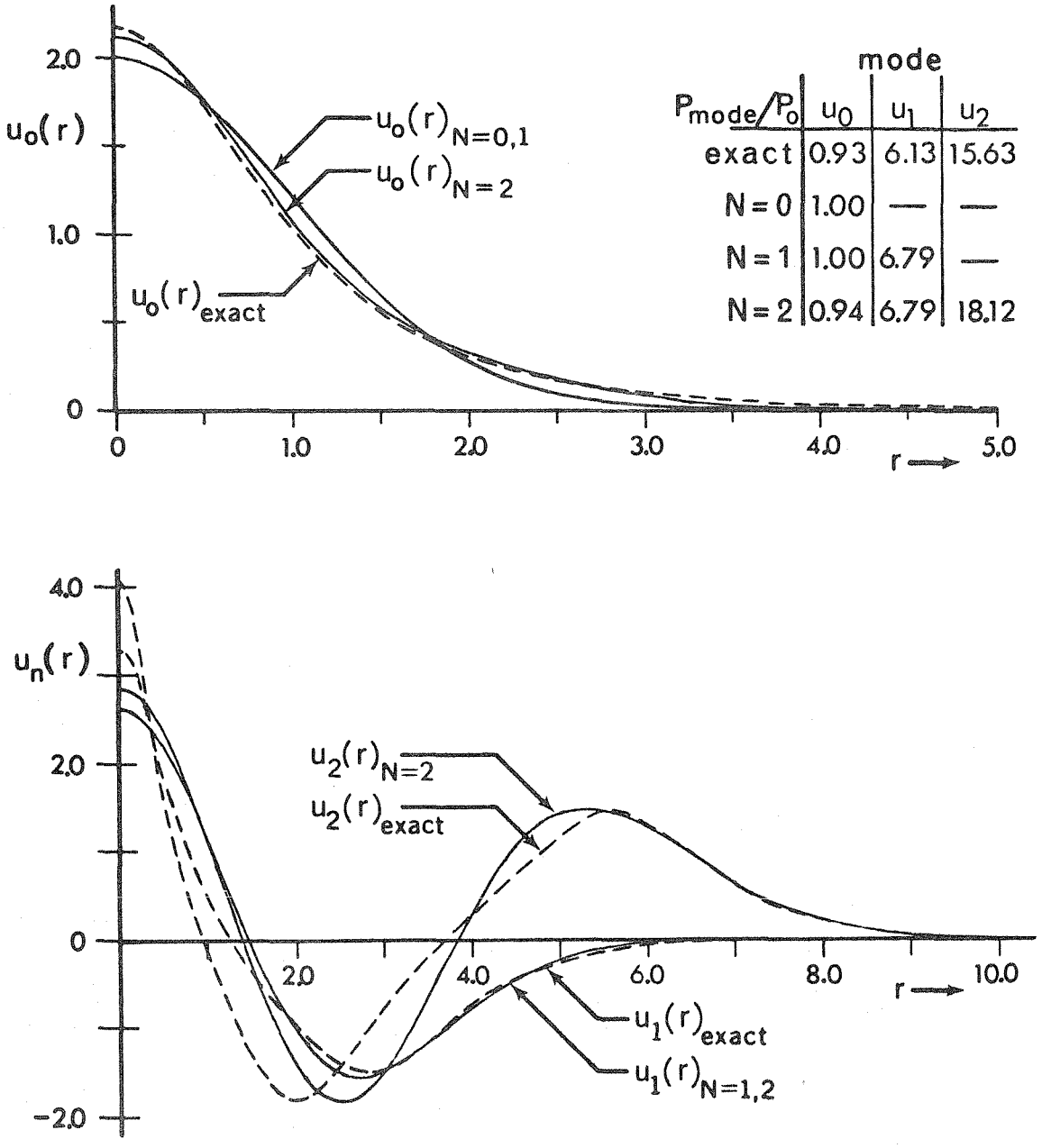
$$\left.\begin{aligned} A_1 &= 2.62536 \\ a_1^2 (=b_1^2) &= 5.20196 \\ C_1 (=E_1) &= -2.42423 \\ (S_x = S_y &= D = F = 0) \end{aligned}\right\} \qquad [u_1(r)]_{N=1} = A_1 e^{-r^2/a_1^2}\{1 + \frac{C_1}{a_1^2} r^2\} \qquad \text{(IV.4.3)}$$

As shown in Figure 24, this solution corresponds to an approximation of $u_1(r)$, the next higher order mode found by Haus[90].

Finally, the N = 2 level of parameterization generates three possible self-trapped solutions, again found numerically. One of these reproduces the same approximation to $u_1(r)$ as found for N = 1, but the other two generate an improved approximation to $u_0(r)$ and a first approximation to $u_2(r)$; these are given by, respectively:

	$[u_0(r)]_{N=2}$	$[u_2(r)]_{N=2}$
$A =$	2.12639	2.85401
$a^2 (=b^2) =$	1.84341	8.87033
$\tau_{1,0} (=\tau_{0,1}) =$	-0.34250	-5.21044
$\tau_{2,0} (= \frac{1}{2}\tau_{1,1} = \tau_{0,2}) =$	0.22593	2.77823

(IV.4.4)



P_{mode}/P_0	mode u_0	u_1	u_2
exact	0.93	6.13	15.63
N = 0	1.00	—	—
N = 1	1.00	6.79	—
N = 2	0.94	6.79	18.12

Figure 24. Self-trapped beams; comparison of exact (numerical) modes with Least-Action parameterized solutions. The data on the exact solutions is taken from Haus (reference 90).

where, for $\tau_{m,n}$ real, we have the N = 2 form given by

$$u(r) = Ae^{-r^2/a^2}\{1 + \frac{\tau_{1,o}}{a^2} r + \frac{\tau_{2,o}}{a^4} r^4\} \ . \qquad \text{(IV.4.5)}$$

As shown in Figure 24, these self-trapped "modes" approximate quite well the exact solutions, further verifying that the N = 2 level of parameterization can approximate the true field in some detail.

IV.4.ii self-focusing examples

The great computational advantage of the parameterized beam method lies in replacing the partial differential equations for the field with a set of first order ordinary differential equations for the parameters, as represented by equation IV.2.16 for N = 0, by equation IV.3.9 and the subsequent "recipe" for N = 1, and by the N = 2 recipe given in Appendix 2. Such equation systems can be integrated with very low errors, primarily limited in this case by the accuracy with which the derivatives can be calculated; however, by using the "recipe" approach to avoid any need for matrix inversions, as well as double precision accuracy in evaluating the resultant formulas, such derivative inaccuracies could be effectively eliminated.

The numerical integration was performed with modified forms of standard routines, employing a fourth-order predictor-corrector formula (Adams-Moulton method); "startup" was achieved with the use of a fourth-order Runge-Kutta method (Runge-Kutta-Gill)[86]. At the initial conditions, or when other "degenerate points" were

encountered (for $N > 0$), the singularities in the derivatives were avoided by "matching" the solution to the known degenerate point expansions. With these methods, the programs were capable of very precise integrations (probably much better than the 10^{-6} to 10^{-7} relative error tests we imposed), so that our results correctly portray the parameterized beam solutions over the entire propagation range. Thus, the discrepancies between $\hat{E}(x,y,z)$ and the true solutions, $E(x,y,z)$, should be fully attributable to the inherent limitations of the parameterized beam approach itself, and should thus decrease rapidly with increasing N. On the other hand, the fully numerical calculations discussed in section III.4 suffer from error sources (local instabilities, imperfect boundary conditions, possible numerical instabilities from the nonlinearity, roundoff accumulation, etc.) which cannot be easily "identified" or alleviated.

The initial conditions used in our parameterized beam solutions corresponded to a collimated beam ($\theta_x = \theta_y = 0$) passed through a "unit lens" ($z_o = 1$); thus, we have, according to equations IV.2.4 and IV.1.2b,

$$\begin{aligned} A(0) &= 1 & \phi(0) &= 0 \\ a(0) &= 1 & b(0) &= R \\ S_x(0) &= -0.25 & S_y(0) &= -0.25 \\ \tau_{m,n}(0) &= 0 & (m,n) &\neq (0,0) \quad . \end{aligned} \tag{IV.4.6a}$$

The resulting solution yields the field parameters as functions of $z = "z_{focused}"$; the "collimated" result of interest actually involves $z = "z_{collimated}"$, where, from equation III.2.3a,

$$z_{collimated} = \frac{z_{focused}}{1 - z_{focused}} \quad . \qquad \text{(IV.4.6b)}$$

The threshold for self-focusing thus occurs at the smallest β for which $(z_f)_{focused} \leq 1$; Figure 25 shows the threshold powers thus obtained, as the ellipticity, R, is varied.

In the remaining figures we show further details of some of the solutions obtained with the parameterized beam approach. These results were obtained by using the least-action degenerate point solutions (i.e. at the initial conditions); however, the "integral properties" of the beam, such as $\langle r^2(z)\rangle$, $\langle x^2(z)\rangle$, $\langle x^4(z)\rangle$, etc., varied only by a few percent, for a given order, when the other degenerate point solutions were employed, even though the individual minor parameters differed in magnitude and sign, in some cases. Thus, the careful use of the correct degenerate point solution seems to be important only when "local" details of the beam are desired (such as field values, $\hat{E}(x,y,z)$), and even then, only if the solution is to be carried for some distance away from the degenerate point.

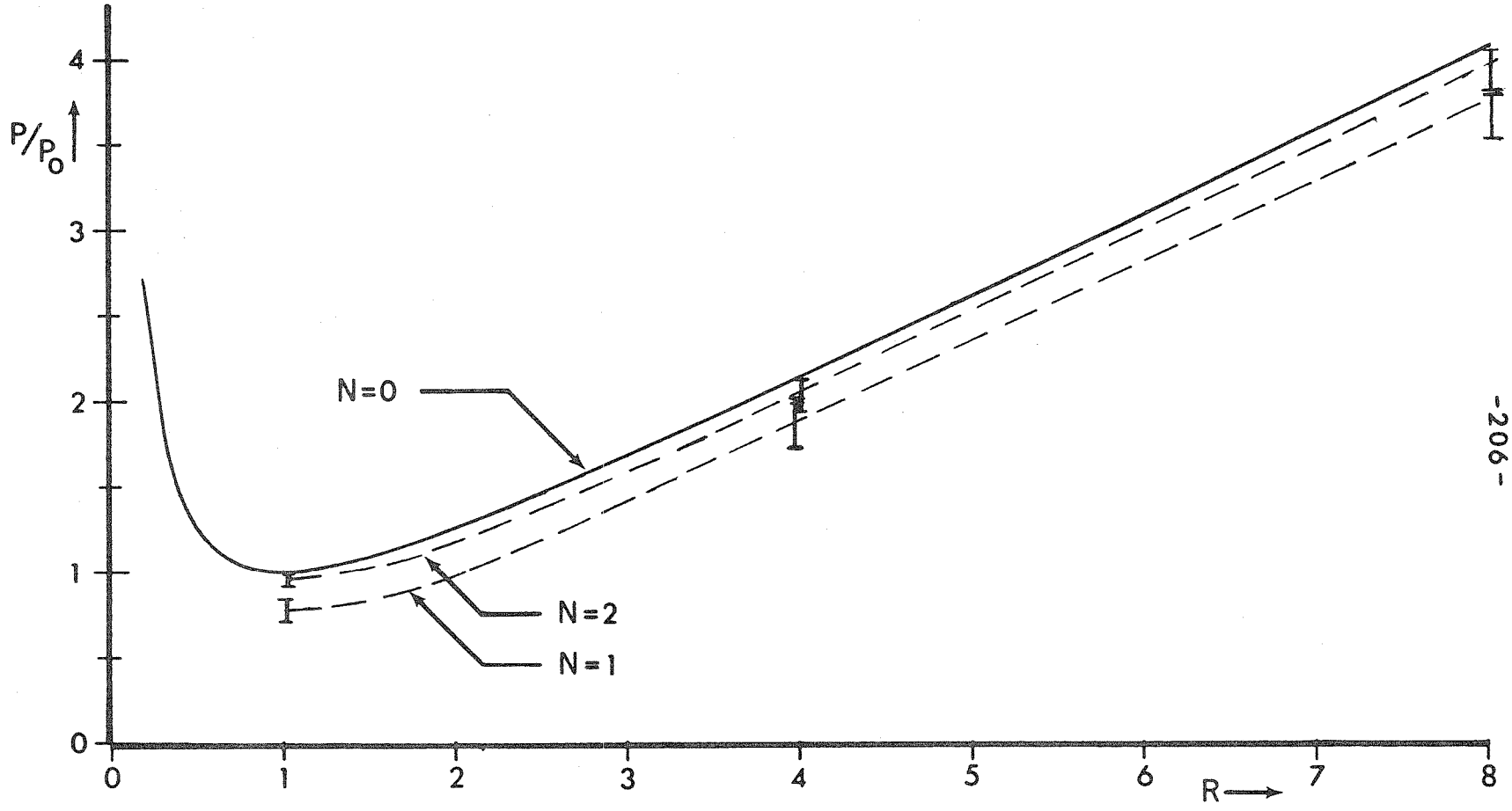


Figure 25. Threshold powers for "collimated" beams, using parameterized beam analysis; the actual initial conditions employed a "unit lens", along with the lens transformation. The N=0 curve is analytic, agreeing with the theoretical result in Figure 20.

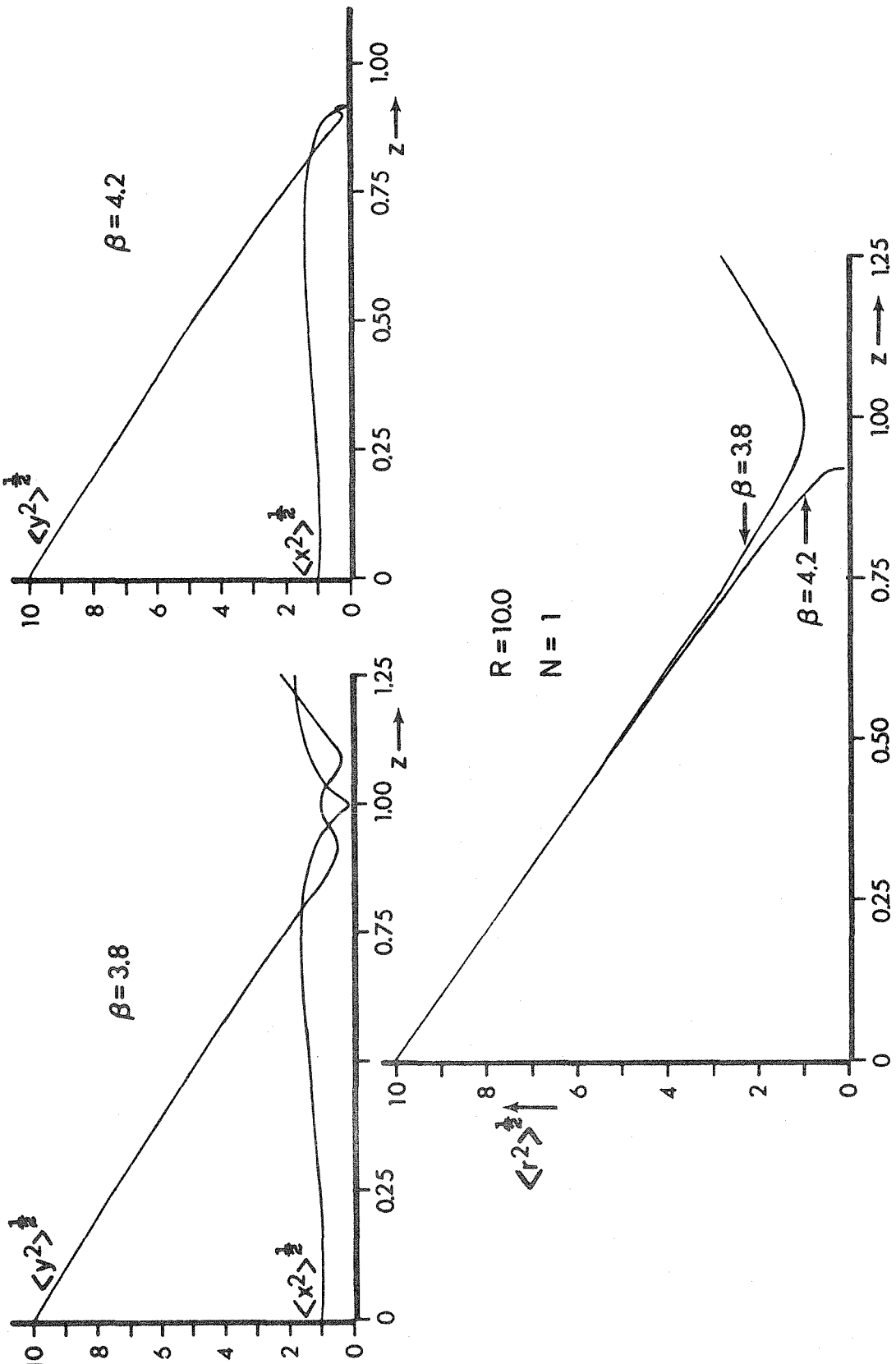


Figure 26. Typical solutions obtained using the N=1 parameterized beam form. Here, R=10.0, and two values of β are compared; focusing occurred, for β=4.2, at z=0.9195.

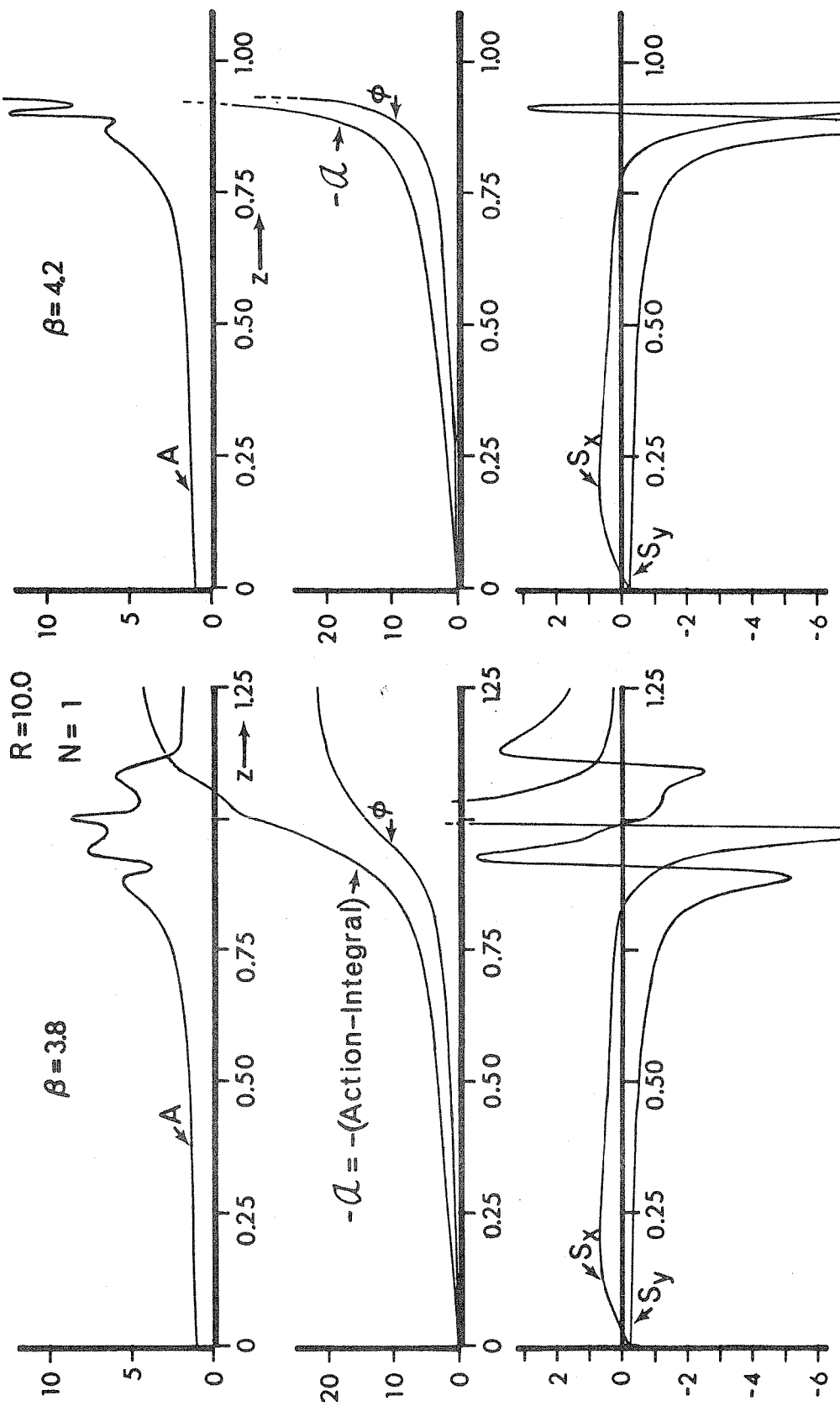


Figure 27. N=1 parameterized beam results for R=10.0 (continued); major parameters. Note that the on-axis intensity A(z) has much more structure than the r.m.s. radii in Figure 26.

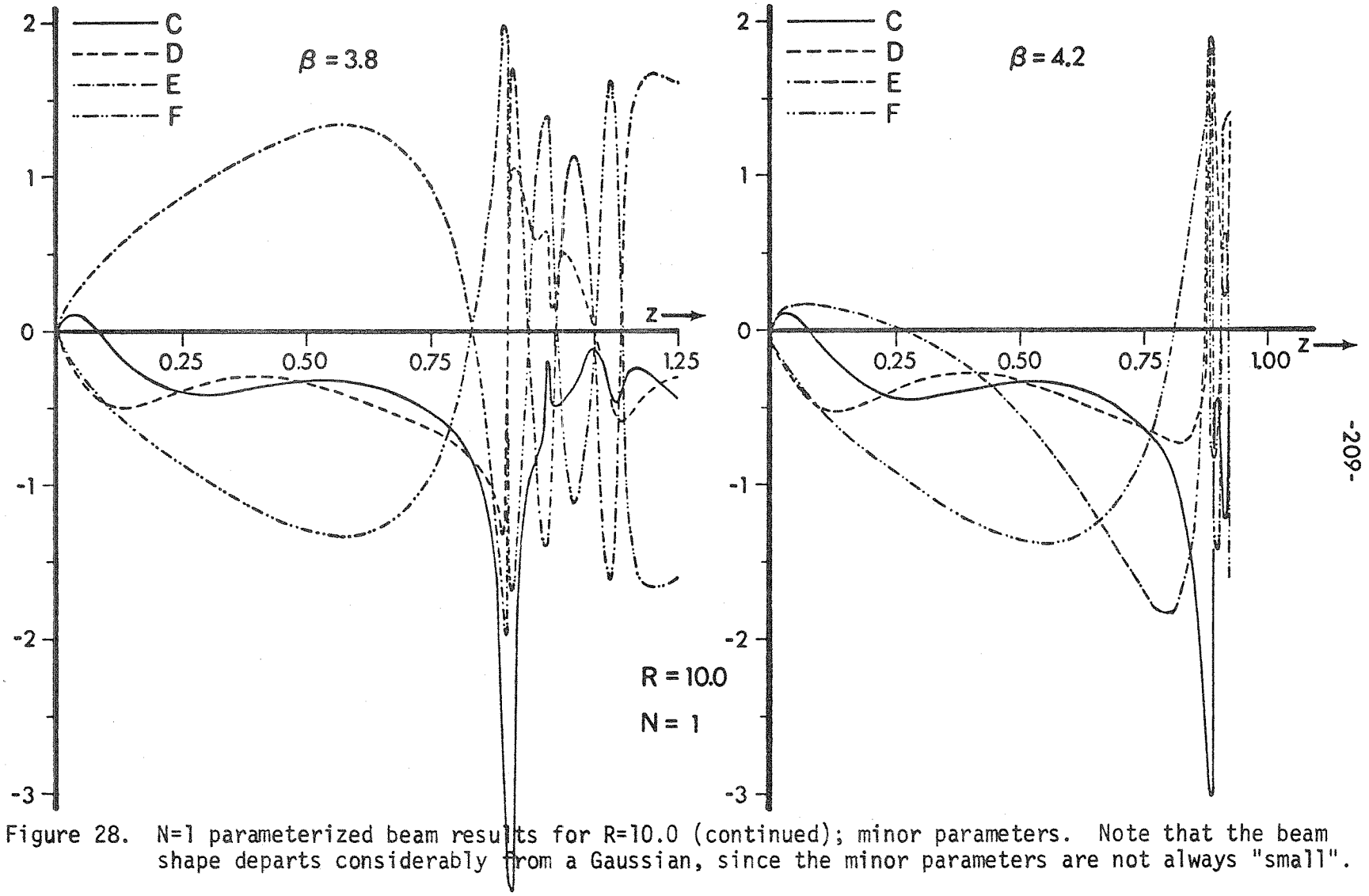


Figure 28. N=1 parameterized beam results for R=10.0 (continued); minor parameters. Note that the beam shape departs considerably from a Gaussian, since the minor parameters are not always "small".

Appendix 1

ANGULAR INTEGRATION FORMULAS AND SUSCEPTIBILITY TENSOR ELEMENTS IN ISOTROPIC MEDIA

In Chapter I, we considered the representation of fields, such as $\vec{E}$, in two reference frames; the first, (x,y,z), is the lab frame, where $\vec{E}$ has components E_i, and the second, (x',y',z') is the molecular frame, in which the field components are E'_i. The transformation matrix, $M(\vec{\Omega})$, connecting the field components is most often given in terms of the Eulerian angles[27], which are pictured in Figure 2, page 23, of Chapter I; in this notation, we have

$$E'_i = M_{ij}(\vec{\Omega})E_j$$
$$E_i = M^{-1}_{ij}(\vec{\Omega})E'_j = M_{ji}(\vec{\Omega})E'_j \quad ; \tag{A1.1}$$

where $\vec{\Omega} = (\phi,\theta,\psi)$ describes the orientation of the mulecule, and

$$M(\vec{\Omega}) = \begin{pmatrix} \cos\psi\cos\phi-\sin\psi\cos\theta\sin\phi & \cos\psi\sin\phi+\sin\psi\cos\theta\cos\phi & \sin\psi\sin\theta \\ -\sin\psi\cos\phi-\cos\psi\cos\theta\sin\phi & -\sin\psi\sin\phi+\cos\psi\cos\theta\cos\phi & \cos\psi\sin\theta \\ \sin\theta\sin\phi & -\sin\theta\cos\phi & \cos\theta \end{pmatrix} \tag{A1.2}$$

Here, we will give tensor products in component form, with a typical subscript, "i", representing x,y, or z; repeated indices in tensor or vector products are assumed to be summed.

We require integrals of functions depending on $\vec{\Omega}$; for example, if $g(\vec{\Omega})$ is some quantity (for example a polarization component)

corresponding to molecules characterized by $\vec{\Omega}$, and $f(\vec{\Omega})$ is the probability of a molecule being oriented at $\vec{\Omega}$, then symbolically we can write

$$\langle g \rangle = \int g(\vec{\Omega}) f(\vec{\Omega}) d^3\vec{\Omega} \ .$$

If we consider the case of a "uniform" distribution $f(\vec{\Omega}) = 1$, we can see from Figure 2 that quantitative evaluation of such integrals, for $\vec{\Omega}$ expressed in Euler angles, involves the substitution:

$$\int d^3\vec{\Omega} = \frac{1}{8\pi^2} \int_0^{2\pi} \Psi \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta \ . \qquad \text{(A1.3)}$$

The necessary integrals involve products of matrix elements, $M_{ij}(\vec{\Omega})$, in the integrand; we will now list these. The results must physically be independent of the specific representation of the rotation transformation, but they can be checked by using equations A1.2 and A1.3. The case of no factors of M is equivalent to the normalization:

$$\int 1 \ d^3\vec{\Omega} = 1 \qquad (0 \ M_{ij} \ \text{factors}) \qquad \text{(A1.4)}$$

Next, we have

$$\int M_{ij}(\vec{\Omega}) d^3\vec{\Omega} = 0 \qquad (1 \ M_{ij} \ \text{factor}) \qquad \text{(A1.5)}$$

For two transformation elements, we have:

$$\int M_{ii'}(\vec{\Omega}) M_{jj'}(\vec{\Omega}) d^3\vec{\Omega} = \frac{1}{3} \delta_{i,j} \delta_{i',j'} \qquad (2 \ M_{ij} \ \text{factors}) \qquad \text{(A1.6)}$$

The case of three elements is covered by:

$$\int M_{ii}(\vec{\Omega})M_{jj}(\vec{\Omega})M_{kk}(\vec{\Omega})d^3\vec{\Omega} = \frac{1}{6}$$

$$\int M_{ii}(\vec{\Omega})M_{jk}(\vec{\Omega})M_{kj}(\vec{\Omega})d^3\vec{\Omega} = -\frac{1}{6} \qquad (3\ M_{ij}\ \text{factors}) \qquad (A1.7)$$

(no sums)

$$\int M_{ij}(\vec{\Omega})M_{jk}(\vec{\Omega})M_{ki}(\vec{\Omega})d^3\vec{\Omega} = \frac{1}{6}$$

Here, (i,j,k) is any permutation of (x,y,z); thus $i \neq j \neq k \neq i$ is needed. All other integrals of three elements of the transformation matrix will vanish.

The enumeration of all nonvanishing integrals involving four matrix elements is more complicated. First, out of the 81 possible distinct combinations of four indices (i,j,k,ℓ), we will form four groups, G_n, each containing certain of these combinations:

$$\begin{aligned} (i,i,i,i) &\longrightarrow G_1 \\ (i,i,j,j) &\longrightarrow G_2 \\ (i,j,i,j) &\longrightarrow G_3 \\ (i,j,j,i) &\longrightarrow G_4 \end{aligned} \qquad (A1.8)$$

Here, $j \neq i$ must hold; G_1 thus contains 3 elements, while each of the other groups contains 6 elements. With these conditions, we find:

$$\int M_{i,i'}(\vec{\Omega})M_{j,j'}(\vec{\Omega})M_{k,k'}(\vec{\Omega})\,M_{\ell,\ell'}(\vec{\Omega})d^3\vec{\Omega} = (I)_{n_1,n_2}$$

when

$$(i,j,k,\ell) \longrightarrow G_{n_1} \qquad (4\ M_{ij}\ \text{factors}) \qquad (A1.9)$$

and

$$(i',j',k',\ell') \longrightarrow G_{n_2}$$

All other four-factor integrals vanish. The values, $(I)_{n_1,n_2}$, are given by the matrix

$$I = \begin{pmatrix} \frac{1}{5} & \frac{1}{15} & \frac{1}{15} & \frac{1}{15} \\ \frac{1}{15} & \frac{2}{15} & -\frac{1}{30} & -\frac{1}{30} \\ \frac{1}{15} & -\frac{1}{30} & \frac{2}{15} & -\frac{1}{30} \\ \frac{1}{15} & -\frac{1}{30} & -\frac{1}{30} & \frac{2}{15} \end{pmatrix} . \qquad \text{(A1.10)}$$

We will now consider the limitations imposed on susceptibility tensors by the isotropy of the medium. Consider expanding the polarization in powers of the electric field components; we will take, as a pertinent example, the cubic polarization at ω_o, which must have the form:

$$P_i = \chi_{ijk\ell} E_j E_k E_\ell^* \quad . \qquad \text{(A1.11)}$$

Now consider reorienting the material using rotations characterized by $\vec{\Omega}$; in the lab system, the medium will appear to have a new susceptibility tensor, $\chi'(\vec{\Omega})$, such that

$$P_i = \chi'_{ijk\ell}(\vec{\Omega}) E_j E_k E_\ell^* \quad . \qquad \text{(A1.12)}$$

By transforming into the material system, however, and denoting with primes the fields seen in this frame, we have

$$P_i = M_{i'i}(\vec{\Omega})P'_{i'}$$

$$= M_{i'i}(\vec{\Omega})(\chi_{i'j'k'\ell'}E_j E_k E^*_{\ell'})$$

$$= M_{i'i}(\vec{\Omega})\ \chi_{i'j'k'\ell'}M_{j'j}(\vec{\Omega})M_{k'k}(\vec{\Omega})M_{\ell'\ell}(\vec{\Omega})E_j E_k E^*_\ell \quad .$$

Comparison with equation A1.12 shows how the susceptibility tensor transforms as the material is rotated:

$$\chi'_{ijk\ell}(\vec{\Omega}) = M_{i'i}(\vec{\Omega})M_{j'j}(\vec{\Omega})M_{k'k}(\vec{\Omega})M_{\ell'\ell}(\vec{\Omega})\chi_{i'j'k'\ell'} \qquad \text{(A1.13)}$$

Next, consider the class of symmetry operations, $\{\vec{\Omega}_s\}$, under which the material is invariant; these operations might include inversions as well, but these are not needed for our purpose. It must then hold, for each $\vec{\Omega}_s$, that

$$\chi'_{ijk\ell}(\Omega_s) = \chi_{ijk\ell} \cdot \qquad \text{(A1.14)}$$

For each crystal symmetry class, a set of $\{\vec{\Omega}_s\}$ is known; equations A1.13 and A1.14 can then be used to reduce the number of possible independent tensor components. The same methods can be used to generalize equations A1.11, A1.13, and A1.14 to any other "order" of nonlinearity; the extension to other harmonics is also obvious, since, for example, the same equations as above must apply to third harmonic, with $E^*_k \to E_k$ in A1.11.

In the particular case of isotropic media, the set $\{\vec{\Omega}_s\}$ contains all $\vec{\Omega}$; this allows the following simple derivation of the general tensor form. Consider averaging $\chi'(\vec{\Omega})$ over all $\vec{\Omega}$; from equation A1.14, we must then have

$$\langle \chi'_{ijk\ell}(\vec{\Omega}) \rangle = \chi_{ijk\ell} = \chi_{i'j'k'\ell'} \int M_{i'i}(\vec{\Omega}) M_{j'j}(\vec{\Omega}) M_{k'k}(\vec{\Omega}) M_{\ell'\ell}(\vec{\Omega}) d^3\vec{\Omega} \ . \tag{A1.15}$$

From equations A1.8 and A1.9, we can immediately see that the only nonzero components of χ are those belonging to the groups, G_n, and that all elements within each group are equal. We can set

$$\begin{aligned} \chi_{iiii} &= \chi_1 \\ \chi_{iijj} &= \chi_2 \\ \chi_{ijij} &= \chi_3 \\ \chi_{ijji} &= \chi_4 \ . \end{aligned} \tag{A1.16}$$

Using equation A1.10, we then obtain, as an example,

$$\chi_1 = \chi_1\left(\frac{3}{5}\right) + \chi_2\left(\frac{6}{15}\right) + \chi_3\left(\frac{6}{15}\right) + \chi_4\left(\frac{6}{15}\right) \ ,$$

with other equations for χ_2, χ_3, χ_4. These equations are all equivalent, giving

$$0 = -\chi_1 + \chi_2 + \chi_3 + \chi_4 \ . \tag{A1.17}$$

By employing equations A1.16 and A1.17 in the definition of χ, equation A1.11, we obtain the general form of the cubic polarization in an isotropic medium:

$$\vec{P} = (\chi_2 + \chi_3)\vec{E}(\vec{E}\cdot\vec{E}^*) + (\chi_4)\vec{E}^*(\vec{E}\cdot\vec{E}) \ . \tag{A1.18}$$

This equation is equivalent to the nonlinear polarization terms in equation I.3.1; it follows directly from equation A1.15 and is thus a necessary condition on the form of $\vec{P}$. It is also sufficient, since

both sides of equation A1.18 transform as simple vectors upon rotation, for any $(\chi_2+\chi_3)$ and (χ_4). We can immediately see that the third harmonic polarization will have the form: $\vec{P}(3\omega_o) = (\chi_2+\chi_3+\chi_4)\vec{E}(\vec{E}\cdot\vec{E})$. Further use of this approach for other orders of nonlinearity gives, for isotropic media,

$$\chi_i = 0 \quad \text{(no permanent polarization)}$$

$$\chi_{ij} = \bar{\alpha}\delta_{i,j} \quad \text{(isotropic linear susceptibility)}$$

$$\chi_{ijk} = 0 \quad \text{(no second harmonic or rectified field)} \quad .$$

Most of these results can be found quite easily through physical arguments, but the methods used here are analytical and generalizable to other orders of nonlinearity.

Appendix 2

EVOLUTION OF SIXTEEN PARAMETER (N=2) BEAM FORM IN SELF-FOCUSING MEDIA

The action-minimization techniques, described in section IV.2 for determining the evolution of a parameterized beam in a medium with a cubic nonlinearity, have been carried out for the case where the parameterization is extended to N=2. The representation of such a beam is given in equation IV.1.2, and contains sixteen parameters, which vary with z; in analogy with equation IV.3.1, we will write this function in the form

$$\hat{E}(x,y,z) = Ae^{i\phi}e^{-x^2/a^2}e^{-y^2/b^2}e^{iS_x x^2}e^{iS_y y^2}\{1+(C+iD)\frac{x^2}{a^2}+(E+iF)\frac{y^2}{b^2}$$

$$+(P+iQ)\frac{x^4}{a^4}+(R+iT)\frac{x^2y^2}{a^2b^2}+(U+iV)\frac{y^4}{b^4}\}\ . \qquad \text{(A2.1)}$$

The six "major" parameters (A,ϕ,a,S_x,b, and S_y) , and the ten "minor" parameters, (C,D,E,F,P,Q,R,T,U, and V) are all functions of z; the form of the equations in this appendix should prevent confusion regarding previous uses of these same symbols in other contexts.

The evolution of the parameters (in z) can be found in terms of first order ordinary differential equations. The procedure is exactly analogous to that used for the N=1 case, described in section IV.3, and the resultant equations have the same general form. However, because of their complexity, we will simply reproduce those results needed for actual computation, expressed in a highly "condensed"

algebraic form; while this condensed form may hinder interpretation, it forms the basis of a very efficient computer routine and also brings out symmetries which are not discernible in the "expanded" formulas.

We will first present a general "recipe" for finding the derivatives, corresponding to the N=1 formulas, IV.3.9 through IV.3.16; we present them in the sequence of use in the "recipe", although their derivation occurs in the opposite order. In general, any formula has a "symmetry counterpart", found by substituting:

$$A \to A, \quad a \rightleftarrows b, \quad C \rightleftarrows E, \quad P \rightleftarrows U, \quad R \to R,$$
$$\phi \to \phi, \quad S_x \rightleftarrows S_y, \quad D \rightleftarrows F, \quad Q \rightleftarrows V, \quad T \to T\ . \tag{A2.2}$$

In forming the derivative functions, it is necessary to compute a number of lengthy polynomials involving only the minor parameters; the following preliminary definitions allow certain subgroups of these functions to be calculated in a recursive manner. First, define the following "lists":

$$E^1_{m,n}: \quad E^1_{0,0} = 1 \qquad E^1_{1,0} = C/8 \qquad E^1_{2,0} = P/64$$
$$E^1_{0,1} = E/8 \qquad E^1_{1,1} = R/64 \tag{A2.3a}$$
$$E^1_{0,2} = U/64$$

$$O^1_{m,n}: \quad O^1_{0,0} = 0 \qquad O^1_{1,0} = D/8 \qquad O^1_{2,0} = Q/64$$
$$O^1_{0,1} = F/8 \qquad O^1_{1,1} = T/64 \tag{A2.3b}$$
$$O^1_{0,2} = V/64\ .$$

Note that, according to equation IV.1.2a, $\tau_{m,n} \equiv (E^1_{m,n}+iO^1_{m,n})\cdot 2^{3(m+n)}$; the superscript "1" implies that the "list" involves only single factors of the minor parameters, including the "parameter" $\tau_{o,o} \equiv 1 + i0$; the subscript pair (m,n) implies that the list element is associated with transverse variations behaving as $(\frac{x^2}{a^2})^m (\frac{y^2}{b^2})^m$. Finally, we have assigned an "even parity" to the elements in "$E_{m,n}$", and an "odd parity" to those elements in "$O_{m,n}$".

We can now recursively define the following lists

$$E^2_{m,n} \equiv \sum_{m'} \sum_{n'} (E^1_{m',n'}E^1_{m-m',n-n'} + O^1_{m',n'}O^1_{m-m',n-n'}) \quad 0\leq m,n,m+n \leq 4$$

$$E^3_{m,n} \equiv \sum_{m'} \sum_{n'} E^1_{m',n'}E^2_{m-m',n-n'} \quad 0\leq m,n,m+n \leq 6$$

$$O^3_{m,n} \equiv \sum_{m'} \sum_{n'} O^1_{m',n'}E^2_{m-m',n-n'} \quad 0\leq m,n,m+n \leq 6 \qquad \text{(A2.4)}$$

Here, of course, only terms with "allowed" subscripts can contribute to each sum; for example, $E^1_{m',n'}$ and $O^1_{m',n'}$ must have m',n' and $m'+n'$ all lying in the range (0,2). Under the symmetry operation defined in equation A2.2, each "list" element, (m,n), is replaced by the element (n,m) in the same "list".

We can now return to the basic problem of forming a "recipe" for N=2. Define

$$B_1 = 8 \sum_m \sum_n \sum_{m'} (E^1_{m',2-m'} E^3_{m,n} + O^1_{m',2-m'} O^3_{m,n})\beta^{m'}_{m,n} \qquad \text{(A2.5a)}$$

$$= -\frac{5}{8} P - \frac{1}{8} R - \frac{1}{8} U + \cdots$$

$$B_2 = 8 \sum_m \sum_n \sum_{m'} (O^1_{m',2-m'} E^3_{m,n} - E^1_{m',2-m'} O^3_{m,n}) \bar{B}^{m'}_{m,n} \qquad \text{(A2.5b)}$$

$$= -\frac{5}{8} Q - \frac{1}{8} T - \frac{1}{8} V + \ldots \quad .$$

These are analogous to the N=1 formulas in equation IV.3.14: B_3 and B_4 are given by the symmetrical counterparts to B_1 and B_2, respectively (with $\bar{B}^{m'}_{m,n} \to \bar{B}^{2-m'}_{n,m}$). The numerical constants, $\bar{B}^{m'}_{m,n}$, are given in Table 1; they also obey the relation

$$\bar{B}^{o}_{m,n} = \bar{B}^{1}_{n,m} \quad . \qquad \text{(A2.5c)}$$

The recipe begins by computing the following four values:

$$\begin{aligned} W_a &= \frac{1}{2} (m_4 B_2 + m_2 B_3 - m_3 B_4)/d_o \\ W_{S_x} &= \quad (m_4 B_1 - m_3 B_3 - m_2 B_4)/d_o \\ W_b &= \frac{1}{2} (-m_2 B_1 - m_3 B_2 + m_1 B_4)/d_o \\ W_{S_y} &= \quad (-m_3 B_1 + m_2 B_2 + m_1 B_3)/d_o \quad , \end{aligned} \qquad \text{(A2.6)}$$

where, for N=2,

$$\begin{aligned} m_1 &= 15(P^2+Q^2) + (R^2+T^2) + (U^2+V^2) \\ m_2 &= \quad (QR-PT) + (TU-RV) \\ m_3 &= \quad (PR+QT) + (RU+TV) \\ m_4 &= \quad (P^2+Q^2) + (R^2+T^2) + 15(U^2+V^2) \end{aligned} \qquad \text{(A2.7a)}$$

$$d_o = (m_1 m_4 - m_2^2 - m_3^2)/2 \quad . \qquad \text{(A2.7b)}$$

m	n	$\bar{B}^2_{m,n}$	$\bar{B}^1_{m,n}$	$\bar{A}_{m,n}$	$\bar{C}_{m,n}$	$\bar{P}_{m,n}$	$\bar{R}_{m,n}$
0	0	-5	-1	-7	3	-4/3	-8/3
1	0	25	3	-3	10	-4	-4
0	1	-5	1	-3	0	0	-4
2	0	45	13	0	42	-13	-18
1	1	25	-3	0	6	-3	2
0	2	-15	9	0	-6	3	-18
3	0	-285	45	30	225	-35	-130
2	1	45	-13	6	27	-11	22
1	2	75	-27	6	9	-7	22
0	3	-75	75	30	-45	25	-130
4	0	-6405	-105	315	1365	210	-1260
3	1	-285	-45	45	135	-30	180
2	2	135	-117	27	45	-30	180
1	3	375	-225	45	15	-30	180
0	4	-525	735	315	-315	210	-1260
5	0	-89775	-7245	2835	7560	8190	-15120
4	1	-6405	105	315	630	210	1680
3	2	-855	-405	135	180	-90	1440
2	3	675	-975	135	90	-150	1440
1	4	2675	-2205	315	0	-210	1680
0	5	-4725	8505	2835	-1890	1890	-15120
6	0	-1091475	-176715	20790	0	183645	-214830
5	1	-89775	7245	1890	0	7875	18270
4	2	-19215	945	630	0	525	13650
3	3	-4275	-3375	450	0	-525	11550
2	4	4725	-9555	630	0	-1155	13650
1	5	23625	-25515	1890	0	-2205	18270
0	6	-51975	114345	20790	0	17325	-214830

Table 1. Numerical constants used in N=2 recipe evaluation.

Next, the values, W_A and W_ϕ, are found from the formulas

$$W_A = W_a\{-\frac{3}{32}(5P+R+U)\}+ W_{S_x}\{\frac{3}{64}(5Q+T+V)\}+ W_b\{-\frac{3}{32}(P+R+5U)\}$$
$$+ W_{S_y}\{\frac{3}{64}(Q+T+5V)\} + \frac{1}{8}\sum_m\sum_n O^3_{m,n}\bar{A}_{m,n} \qquad (A2.8a)$$

$$W_\phi = W_a\{-\frac{3}{32}(5Q+T+V)\} + W_{S_x}\{-\frac{3}{64}(5P+R+U)\}+ W_b\{-\frac{3}{32}(Q+T+5V)\}$$
$$+ W_{S_y}\{-\frac{3}{64}(P+R+5U)\}- \frac{1}{8}\sum_m\sum_n E^3_{m,n}\bar{A}_{m,n}\ , \qquad (A2.8b)$$

where the coefficients, $\bar{A}_{m,n}$, are listed in Table 1. Using these, the remaining "nonlinear" derivative contributions can be found as follows:

$$W_C = W_A\{-2-C\} + W_\phi\{D\} + W_a\{-2+2C+\frac{3}{16}(25P+3R+U)\}$$
$$+ W_{S_x}\{-\frac{3}{32}(25Q+3T+V)\} + W_b\{\frac{3}{16}(3P+R-5U)\} + W_{S_y}\{-\frac{3}{32}(3Q+T-5V)\}$$
$$- \frac{1}{4}\sum_m\sum_n O^3_{m,n}\bar{C}_{m,n} \qquad (A2.9a)$$

$$W_D = W_A\{-D\} + W_\phi\{-2-C\} + W_a\{2D+\frac{3}{16}(25Q+3T+V)\}$$
$$+ W_{S_x}\{-1+\frac{3}{32}(25P+3R+U)\} + W_b\{\frac{3}{16}(3Q+T-5V)\} + W_{S_y}\{\frac{3}{32}(3P+R-5U)\}$$
$$+ \frac{1}{4}\sum_m\sum_n E^3_{m,n}\bar{C}_{m,n} \qquad (A2.9b)$$

$$W_P = W_A\{\tfrac{2}{3}-P\} + W_\phi\{Q\} + W_a\{-2C+4P-\tfrac{1}{16}(115P+7R-U)\}$$

$$+ W_{S_x}\{D+\tfrac{1}{32}(115Q+7T-V)\} + W_b\{\tfrac{1}{16}(-7P+R+5U)\} + W_{S_y}\{-\tfrac{1}{32}(-7Q+T+5V)\}$$

$$- \tfrac{1}{4}\sum_m\sum_n O^3_{m,n}\bar{P}_{m,n} \qquad \text{(A2.9c)}$$

$$W_Q = W_A\{-Q\} + W_\phi\{\tfrac{2}{3}-P\} + W_a\{-2D+4Q-\tfrac{1}{16}(115Q+7T-V)\}$$

$$+ W_{S_x}\{-C-\tfrac{1}{32}(115P+7R-U)\} + W_b\{\tfrac{1}{16}(-7Q+T+5V)\} + W_{S_y}\{\tfrac{1}{32}(-7P+R+5U)\}$$

$$+ \tfrac{1}{4}\sum_m\sum_n E^3_{m,n}\bar{P}_{m,n} \qquad \text{(A2.9d)}$$

$$W_R = W_A\{\tfrac{4}{3}-R\} + W_\phi\{T\} + W_a\{-2E+2R-\tfrac{1}{8}(-5P+23R+23U)\}$$

$$+ W_{S_x}\{F+\tfrac{1}{16}(-5Q+23T+23V)\} + W_b\{-2C+2R-\tfrac{1}{8}(23P+23R-5U)\}$$

$$+ W_{S_y}\{D+\tfrac{1}{16}(23Q+23T-5V)\} - \tfrac{1}{4}\sum_m\sum_n O^3_{m,n}\bar{R}_{m,n} \qquad \text{(A2.9e)}$$

$$W_T = W_A\{-T\} + W_\phi\{\tfrac{4}{3}-R\} + W_a\{-2F+2T-\tfrac{1}{8}(-5Q+23T+23V)\}$$

$$+ W_{S_x}\{-E-\tfrac{1}{16}(-5P+23R+23U)\} + W_b\{-2D+2T-\tfrac{1}{8}(23Q+23T-5V)\}$$

$$+ W_{S_y}\{-C-\tfrac{1}{16}(23P+23R-5U)\} + \tfrac{1}{4}\sum_m\sum_n E^3_{m,n}\bar{R}_{m,n} \quad . \qquad \text{(A2.9f)}$$

The coefficients, $\bar{C}_{m,n}$, $\bar{P}_{m,n}$, and $\bar{R}_{m,n}$ are given in Table 1; W_E, W_F, W_U, and W_V are found by using the symmetrical counterparts to the formulas for W_C, W_D, W_P, and W_Q, respectively.

The last step in the recipe adds the "linear" contributions to the parameter derivatives, as follows:

$$
\begin{aligned}
\frac{1}{A}\frac{dA}{dz} &= -2S_x - 2S_y + \frac{1}{a^2}\{-2D\} + \frac{1}{b^2}\{-2F\} && + \beta A^2 W_A \\
\frac{d\phi}{dz} &= \frac{1}{a^2}\{-2+2C\} + \frac{1}{b^2}\{-2+2E\} && + \beta A^2 W_\phi \\
\frac{1}{a}\frac{da}{dz} &= 4S_x && + \beta A^2 W_a \\
a^2\frac{dS_x}{dz} &= -4a^2 S_x^2 + \frac{1}{a^2}\{4\} && + \beta A^2 W_{S_x} \\
\frac{dC}{dz} &= \frac{1}{a^2}\{8D+4CD-12Q\} + \frac{1}{b^2}\{2CF+2DE-2T\} && + \beta A^2 W_C \\
\frac{dD}{dz} &= \frac{1}{a^2}\{-8C-2C^2+2D^2+12P\} + \frac{1}{b^2}\{2DF-2CE+2R\} && + \beta A^2 W_D \\
\frac{dP}{dz} &= \frac{1}{a^2}\{16Q+2CQ+2DP\} + \frac{1}{b^2}\{2EQ+2FP\} && + \beta A^2 W_P \\
\frac{dQ}{dz} &= \frac{1}{a^2}\{-16P+2DQ-2CP\} + \frac{1}{b^2}\{2FQ-2EP\} && + \beta A^2 W_Q \\
\frac{dR}{dz} &= \frac{1}{a^2}\{8T+2CT+2DR\} + \frac{1}{b^2}\{8T+2ET+2FR\} && + \beta A^2 W_R \\
\frac{dT}{dz} &= \frac{1}{a^2}\{-8R+2DT-2CR\} + \frac{1}{b^2}\{-8R+2FT-2ER\} && + \beta A^2 W_T \quad .
\end{aligned}
\tag{A2.10}
$$

Again, expressions for $\frac{1}{b}\frac{db}{dz}$, $b^2\frac{dS_y}{dz}$, $\frac{dE}{dz}$, $\frac{dF}{dz}$, $\frac{dU}{dz}$ and $\frac{dV}{dz}$ can be found by using symmetry on the proper formulas in equation A2.10.

As with the N=1 case discussed in section IV.3.ii, the above

formulas for the N=2 parameter derivatives will fail at certain "degenerate points". From equations A2.6 and A2.7, these occur when $d_o = 0$, where d_o can be expanded in the form:

$$2d_o = 15(P^2+Q^2+U^2+V^2)^2 + 13(R^2+T^2)(P^2+Q^2+U^2+V^2) + (R^2+T^2)^2$$
$$+196(P^2+Q^2)(U^2+V^2) + (PR-TV)^2 + (PT-RV)^2$$
$$+ (QR-TU)^2 + (QT-RU)^2 + R^2((P-U)^2 + (Q+V)^2) + T^2((P+U)^2+ (Q-V)^2).$$

This expression shows that a degenerate point can only occur when, at some $z = z_{degenerate}$, we have P=Q=R=T=U=V=0; we will now discuss the possible solutions near such points. We can identify two "classes" of the degenerate point solutions; in the first, one or more of C,D,E, or F is nonzero at $z_{degenerate}$, while in the second, all of the minor parameters approach zero as $z \to z_{degenerate}$. We will begin by discussing this latter class of solutions in some detail, since it includes the initial conditions given in equation IV.1.2b.

The following expansions will be assumed to hold near "z" = $z - z_{degenerate} \approx 0$:

$$C(z),D(z),E(z),F(z) \sim z^{\gamma_1}$$
$$P(z),Q(z),R(z),T(z),U(z),V(z) \sim z^{\gamma_2}, \qquad \gamma_1,\gamma_2 > 0$$

where at least one function in each group has a nonzero leading term with the power shown, but no lower powers occur. Applying the "recipe", we obtain, in turn, the leading terms of the expansions of the following functions:

$$B_i \sim z^{\gamma_2}$$

$$m_i \sim z^{2\gamma_2}$$

$$d_o \sim z^{4\gamma_2}$$

$$W_a, W_{S_x}, W_b, W_{S_y} \sim z^{-\gamma_2} \quad ; \; W_A, W_\phi \sim z^o$$

$$W_C, W_D, W_E, W_F \sim z^{-\gamma_2}$$

$$W_P, W_Q, W_R, W_T, W_U, W_V \sim z^{-\gamma_2+\gamma_1} \text{ or } z^o.$$

Finally, from equation A2.10, we have a "balance of powers" for C,D,E,F, and P,Q,R,T,U,V:

$$\frac{dC}{dz}: \; z^{\gamma_1 - 1} \sim z^{-\gamma_2}$$

$$\frac{dP}{dz}: \; z^{\gamma_2 - 1} \sim z^{-\gamma_2+\gamma_1} \text{ or } z^o$$

These equations are consistent only if $\gamma_1 = \frac{1}{3}$, $\gamma_2 = \frac{2}{3}$, so that we can write

$$\begin{aligned}
A(z) &= A_o + \ldots \quad , & \phi(z) &= \phi_o + \ldots \\
a(z) &= a_o + \ldots \quad , & S_x(z) &= S_x^o + \ldots \\
b(z) &= b_o + \ldots \quad , & S_y(z) &= S_y^o + \ldots \\
C(z) &= C_o z^{1/3} + \ldots \quad , & D(z) &= D_o z^{1/3} + \ldots \\
E(z) &= E_o z^{1/3} + \ldots \quad , & F(z) &= F_o z^{1/3} + \ldots \\
P(z) &= P_o z^{2/3} + \ldots \quad , & Q(z) &= Q_o z^{2/3} + \ldots
\end{aligned} \tag{A2.11}$$

$$R(z) = R_o z^{2/3} + \ldots \quad , \qquad T(z) = T_o z^{2/3} + \ldots$$
$$U(z) = U_o z^{2/3} + \ldots \quad , \qquad V(z) = V_o z^{2/3} + \ldots \quad . \qquad \text{(A2.11) cont'd.}$$

From the previous expansions, we have

$$W_a(z) = z^{-2/3}\bar{W}_a + \ldots$$
$$W_{S_x}(z) = z^{-2/3}\bar{W}_{S_x} + \ldots$$
$$W_b(z) = z^{-2/3}\bar{W}_b + \ldots$$
$$W_{S_y}(z) = z^{-2/3}\bar{W}_{S_y} + \ldots \quad ,$$

where $\bar{W}_a, \bar{W}_{S_x}, \bar{W}_b$, and $\bar{W}_{S_y}$ are functions of $C_o, D_o, \ldots, U_o$, and V_o. With these definitions, we can obtain, as "self-consistency" equations, formulas such as

$$\frac{dC}{dz} = z^{-2/3}\{\frac{1}{3} C_o\} + \ldots = \beta A^2 W_c + \ldots = z^{-2/3}\{\beta A_o^2(-2\bar{W}_a)\} + \ldots$$
$$\frac{dD}{dz} = z^{-2/3}\{\frac{1}{3} D_o\} + \ldots = z^{-2/3}\{\beta A_o^2(-\bar{W}_{S_x})\} + \ldots \qquad \text{(A2.12)}$$
$$\frac{dP}{dz} = z^{-1/3}\{\frac{2}{3} P_o\} + \ldots = z^{-1/3}\{\beta A_o^2(-2C_o\bar{W}_a + D_o\bar{W}_{S_x})\} + \ldots \quad .$$

These, and similar equations for the other minor derivatives, show that

$$P_o = \beta A_o^2(-3C_o\bar{W}_a + \frac{3}{2} D_o\bar{W}_{S_x}) = \frac{1}{2}(C_o^2 - D_o^2)$$
$$Q_o = C_o D_o \qquad \text{(A2.13)}$$
$$R_o = C_o E_o - D_o F_o$$

$$T_o = C_oF_o+D_oE_o$$
$$U_o = \frac{1}{2}(E_o^2-F_o^2) \qquad \text{(A2.13) cont'd.}$$
$$V_o = E_oF_o \ .$$

With these substitutions, the leading coefficients in the expressions can be written in terms of C_o,D_o,E_o and F_o:

$$B_1(z) = z^{2/3}\{-\frac{5}{16}(C_o^2-D_o^2)-\frac{1}{8}(C_oE_o-D_oF_o)-\frac{1}{16}(E_o^2-F_o^2)\}+ \ ...$$

$$B_2(z) = z^{2/3}\{-\frac{5}{8}(C_oD_o) - \frac{1}{8}(C_oF_o+D_oE_o) - \frac{1}{8}(E_oF_o)\}+ \ ...$$

$$\vdots$$

$$m_1(z) = z^{4/3}\{\frac{15}{4}(C_o^2+D_o^2)^2 + (C_o^2+D_o^2)(E_o^2+F_o^2) + \frac{1}{4}(E_o^2+F_o^2)^2\}+ \ ...$$

$$m_2(z) = z^{4/3}\{\frac{1}{2}(D_oE_o-C_oF_o)(C_o^2+D_o^2+E_o^2+F_o^2)\}+ \ ...$$

$$\vdots$$

$$d_o(z) = z^{8/3}\bar{d}_o+ \ ... \text{ where } \bar{d}_o = \frac{3}{32}\{5(C_o^2+D_o^2+E_o^2+F_o^2)^4+48(C_o^2+D_o^2)^2(E_o^2+F_o^2)^2\}.$$

Finally, the permissible values of $C_o,D_o,E_o,$ and F_o are those which satisfy the "self-consistency" relations resulting from equation A2.12:

$$C_o\left(\frac{\bar{d}_o}{3\beta A_o^2}\right) = -2\bar{W}_a\bar{d}_o = \frac{1}{32}\{(C_o^2+D_o^2)^2(5C_oD_o-3E_oF_o)+(C_o^2+D_o^2)(E_o^2+F_o^2)(20C_oD_o+8C_oF_o+$$
$$8D_oE_o)+ (E_o^2+F_o^2)^2(75C_oD_o+20C_oF_o+20D_oE_o+15E_oF_o)-20E_oF_o(D_oF_o+C_oE_o)\cdot$$
$$(C_o^2+D_o^2+E_o^2+F_o^2)\} \qquad \text{(A2.14)}$$

$$D_o\left(\frac{\dot{d}_o}{3\beta A_o^2}\right) = \frac{1}{32}\ (C_o^2+D_o^2)^2\left(\frac{5}{2}(C_o^2-D_o^2) - \frac{3}{2}(E_o^2-F_o^2)\right) + (C_o^2+D_o^2)(E_o^2+F_o^2)(10(C_o^2-D_o^2)+$$

$$+ 2(D_oF_o-C_oE_o))+(E_o^2+F_o^2)^2\left(\frac{75}{2}(C_o^2-D_o^2)+10(C_oE_o-D_oF_o)+ \frac{15}{2}(E_o^2-F_o^2)\right)-$$

$$- 20E_oF_o(D_oE_o-C_oF_o)(C_o^2+D_o^2+E_o^2+F_o^2)\}. \qquad \text{(A2.14) cont'd.}$$

The two relations in equation A2.14, with their symmetrical counterparts, form four (nonlinear) algebraic equations in four unknowns. After considerable analysis, they can be shown to yield the following nine sets of values for C_o, D_o, E_o, and F_o:

Solution	$C_o/(\beta A_o^2)^{\frac{1}{3}}$	$D_o/(\beta A_o^2)^{\frac{1}{3}}$	$E_o/(\beta A_o^2)^{\frac{1}{3}}$	$F_o/(\beta A_o^2)^{\frac{1}{3}}$
$^2_{\mathrm{I}}$	$\frac{\sqrt{3}}{2^{4/3}}$	$\frac{1}{2^{4/3}}$	$\frac{\sqrt{3}}{2^{4/3}}$	$\frac{1}{2^{4/3}}$
$^2_{\mathrm{II}}$	0	$-\frac{2}{2^{4/3}}$	0	$-\frac{2}{2^{4/3}}$
$^2_{\mathrm{III}}$	$-\frac{\sqrt{3}}{2^{4/3}}$	$\frac{1}{2^{4/3}}$	$-\frac{\sqrt{3}}{2^{4/3}}$	$\frac{1}{2^{4/3}}$
$^2_{\mathrm{IV}}$	$\frac{4}{(20)^{2/3}}$	$\frac{2}{(20)^{2/3}}$	$-\frac{4}{(20)^{2/3}}$	$\frac{2}{(20)^{2/3}}$
$^2_{\mathrm{V}}$	$-\frac{4}{(20)^{2/3}}$	$\frac{2}{(20)^{2/3}}$	$\frac{4}{(20)^{2/3}}$	$\frac{2}{(20)^{2/3}}$
$^2_{\mathrm{VI}}$	$\frac{-2+\sqrt{3}}{(20)^{2/3}}$	$\frac{-1-2\sqrt{3}}{(20)^{2/3}}$	$\frac{2+\sqrt{3}}{(20)^{2/3}}$	$\frac{-1+2\sqrt{3}}{(20)^{2/3}}$
$^2_{\mathrm{VII}}$	$\frac{2+\sqrt{3}}{(20)^{2/3}}$	$\frac{-1+2\sqrt{3}}{(20)^{2/3}}$	$\frac{-2+\sqrt{3}}{(20)^{2/3}}$	$\frac{-1-2\sqrt{3}}{(20)^{2/3}}$
$^2_{\mathrm{VIII}}$	$\frac{2-\sqrt{3}}{(20)^{2/3}}$	$\frac{-1-2\sqrt{3}}{(20)^{2/3}}$	$\frac{-2-\sqrt{3}}{(20)^{2/3}}$	$\frac{-1+2\sqrt{3}}{(20)^{2/3}}$
$^2_{\mathrm{IX}}$	$\frac{-2-\sqrt{3}}{(20)^{2/3}}$	$\frac{-1+2\sqrt{3}}{(20)^{2/3}}$	$\frac{2-\sqrt{3}}{(20)^{2/3}}$	$\frac{-1-2\sqrt{3}}{(20)^{2/3}}$

(A2.15)

With the use of equations A2.11 and A2.13, the leading terms in each of the solutions can be found; further expansions in powers of $z^{1/3}$ can then be made in the usual way. An expansion of the Action-Integral shows that the solutions above are listed in order of increasing Action, except that the mutual ordering of the last six may depend on the values of a_o, b_o, S_x^o and S_y^o.

The solution 2_I is particularly important because it generates the least-action beam propagation near the initial conditions; a further expansion yields

$$A(z) = A_o[1-(2S_x^o+2S_y^o)z - \frac{3}{64}(\frac{\beta A_o^2}{2} z)^{4/3} + \ldots]$$

$$\phi(z) = \phi_o - (\frac{2}{a_o^2} + \frac{2}{b_o^2} - \frac{15}{16}\beta A_o^2)z - \frac{9\sqrt{3}}{64}(\frac{\beta A_o^2}{2} z)^{4/3} + \ldots$$

$$a(z) = a_o[1 + \frac{\sqrt{3}}{4}(\frac{\beta A_o^2}{2} z)^{1/3} + \frac{9}{32}(\frac{\beta A_o^2}{2} z)^{2/3} + \ldots]$$

$$S_x(z) = S_x^o - \frac{1}{2a_o^2}(\frac{\beta A_o^2}{2} z)^{1/3} - \frac{\sqrt{3}}{4a_o^2}(\frac{\beta A_o^2}{2} z)^{2/3} + \ldots$$

$$b(z) = b_o[1 + \frac{\sqrt{3}}{4}(\frac{\beta A_o^2}{2} z)^{1/3} + \frac{9}{32}(\frac{\beta A_o^2}{2})z)^{2/3} + \ldots] \qquad \text{(A2.16)}$$

$$S_y(z) = S_y^o - \frac{1}{2b_o^2}(\frac{\beta A_o^2}{2} z)^{1/3} - \frac{\sqrt{3}}{4b_o^2}(\frac{\beta A_o^2}{2} z)^{2/3} + \ldots$$

$$C(z) = -\frac{\sqrt{3}}{2}(\frac{\beta A_o^2}{2} z)^{1/3} - \frac{3}{4}(\frac{\beta A_o^2}{2} z)^{2/3} + \ldots = E(z)$$

$$D(z) = \frac{1}{2}(\frac{\beta A_o^2}{2} z)^{1/3} + \frac{\sqrt{3}}{2}(\frac{\beta A_o^2}{2} z)^{2/3} + \ldots = F(z)$$

$$P(z) = \frac{1}{4}(\frac{\beta A_o^2}{2} z)^{2/3} + \frac{\sqrt{3}}{8}(\frac{\beta A_o^2}{2} z) + \ldots = \frac{1}{2} R(z) = U(z)$$

$$Q(z) = -\frac{\sqrt{3}}{4}(\frac{\beta A_o^2}{2} z)^{2/3} + \frac{1}{8}(\frac{\beta A_o^2}{2} z) + \ldots = \frac{1}{2} T(z) = V(z) .$$

(The symmetry conditions: $E(z) = C(z)$, etc. will fail in higher order expansions, unless $a_o = b_o$, $S_x^o = S_y^o$.) Solution 2_{III} can be found from equation A2.16 by setting "$\sqrt{3}$" $\rightarrow$ "$-\sqrt{3}$" .

As mentioned earlier, another class of degenerate point solutions can occur, in which one or more of C,D,E, or F is nonzero at $z_{degenerate}$. By using techniques similar to those above, we find that the minor parameters will behave as

$$\begin{aligned}
C(z) &= C_o + C_1 z^{1/2} + \ldots \quad , & D(z) &= D_o + D_1 z^{1/2} + \ldots \quad , \\
E(z) &= E_o + E_1 z^{1/2} + \ldots \quad , & F(z) &= F_o + F_1 z^{1/2} + \ldots \quad , \\
P(z) &= P_1 z^{1/2} + \ldots \quad , & Q(z) &= Q_1 z^{1/2} + \ldots \quad , \\
R(z) &= R_1 z^{1/2} + \ldots \quad , & T(z) &= T_1 z^{1/2} + \ldots \quad , \\
U(z) &= U_1 z^{1/2} + \ldots \quad , & V(z) &= V_1 z^{1/2} + \ldots \quad .
\end{aligned} \tag{A2.17}$$

Self-consistency equations for P_1, Q_1, R_1, T_1, U_1, and V_1 can be found as usual, except that all coefficients are polynomial functions of C_o, D_o, E_o and F_o (which are presumed known). In practice, this set of nonlinear (algebraic) equations was solved computationally, on a case-to-case basis, since the algebraic complexity of the problem prevented analytic results. Furthermore, this class of degenerate point solutions could usually be "avoided" entirely by altering slightly the initial conditions (i.e. a "physically equivalent" problem could be found for which no degenerate point of this class would occur at any later value of z).

To "order" the degenerate point solutions corresponding to

equation A2.17, when needed, or to verify the analytical results for the degenerate point solutions listed in equation A2.15, the action-integral was calculated along with our numerical solutions. For N=2, equations IV.2.7, IV.2.12a, and IV.2.9a reduce to

$$\hat{a}(z): \quad \frac{d\hat{a}}{dz} = \hat{L}(z) \qquad \hat{a}(0) = 0$$

$$\hat{L}(z) = -\frac{\beta}{2}\iint dx\, dy(|\hat{E}|^2)^2$$

$$= -\frac{\pi\beta A^4 ab}{8}\sum_{m'}\sum_{n'}\sum_{m}\sum_{n}(E^1_{m',n'}E^3_{m-m',n-n'}+O^1_{m',n'}O^3_{m-m',n-n'})\bar{L}_m\bar{L}_n$$

where $\bar{L}_o = 1$ and $\bar{L}_{m+1} = (2m-1)\bar{L}_m$; the functions $E^1_{m,n}, O^1_{m,n}, E^3_{m,n}$, and $O^3_{m,n}$ are defined in equations A2.3 and A2.4.

BIBLIOGRAPHY

1) C. W. Cho, N. D. Foltz, D. H. Rank, T. A. Wiggins, Phys. Rev. Letters 18, 107 (1967).

2) R. Y. Chiao, C. H. Townes, B. F. Stoicheff, Phys. Rev. Letters 12, 592 (1964).

3) R. W. Hellwarth, Phys. Rev. 130, 1850 (1963).

4) A. Yariv and J. E. Pearson, Progress in Quantum Electronics 1, Part 1, 1 (1969).

5) C. K. N. Patel, Phys. Rev. Letters 15, 1027 (1965).

6) J. A. Armstrong, N. Bloembergen, J. Ducuing, P. S. Pershan, Phys. Rev. 127, 1918 (1962).

7) J. D. Macomber, J. Quantum Electronics QE-4, 1 (1968).

8) S. L. McCall, E. L. Hahn, Phys. Rev. 183, 457 (1969).

9) R. H. Dicke, Phys. Rev. 93, 99 (1954).

10) C. L. Tang, B. D. Silverman, Physics of Quantum Electronics, (McGraw-Hill Book Co., 1966), p. 280.

11) N. Bloembergen, P. Lallemand, Phys. Rev. Letters 16, 81 (1966).

12) P. D. Maker, R. W. Terhune, Phys. Rev. 137, 801 (1965).

13) T. K. Gustafson, J. P. Taran, H. A. Haus, J. R. Lifsitz, P. L. Kelley, Phys. Rev. 177, 306 (1969).

14) R. Y. Chiao, E. Garmire, C. H. Townes, Phys. Rev. Letters 13, 479 (1964).

15) H. W. Mocker, R. J. Collins, Appl. Phys. Letters 7, 270 (1965).

16) A. J. DeMaria, D. A. Stetser, H. Heynau, Appl. Phys. Letters 8, 174 (1966).

17) J. A. Fleck, J. of Appl. Phys. 39, 3318 (1968).

18) S. E. Schwarz, J. of Quantum Electronics QE-4, 509 (1968).

19) A. J. DeMaria, D. A. Stetser, W. H. Glenn, Science 156, 1557 (1967).

20) J. P. Laussade, A. Yariv, Appl. Phys. Letters 13, 65 (1968).

21) C. R. Giuliano, J. H. Marburger, A. Yariv, J. of Quantum Electronics QE-8, 588 (1972).

22) R. L. Carman, R. Y. Chiao, P. L. Kelley, Phys. Rev. Letters 17, 1281 (1966).

23) J. D. Jackson, Classical Electrodynamics (Wiley, New York, 1962), p. 618.

24) M. Born and E. Wolf, Principles of Optics (Pergamon Press, Oxford, England, 1970) p. 84-104.

25) R. H. Dicke and J. P. Wittke, Introduction to Quantum Mechanics (Addison-Wesley, Reading, Mass., 1960) p. 237-241.

26) H. Goldstein, Classical Mechanics (Addison-Wesley, Reading, Mass., 1959), p. 154.

27) Ibid., p. 107.

28) M. A. Duguay, J. W. Hansen, S. L. Shapiro, J. of Quantum Electronics QE-6, 725 (1970).

29) D. H. Close, C. R. Giuliano, R. W. Hellwarth, L. D. Hess, F. J. McClung, W. G. Wagner, J. of Quantum Electronics QE-2, 553 (1966).

30) M. Takatsuji, Phys. Rev. 155, 980 (1967).

31) M. Takatsuji, Phys. Rev. 165, 171 (1968).

32) P. Debye, Polar Molecules (Dover Publications, New York, 1958), Chapter V.

33) R. M. Herman, Phys. Rev. 164, 200 (1967).

34) S. L. Shapiro, H. P. Broida, Phys. Rev. 154, 129 (1967).

35) A. A. Chaban, JETP Letters 6, 487 (1967).

36) P. M. Morse and H. Feshbach, Methods of Theoretical Physics (McGraw-Hill, New York, 1953), p. 162.

37) R. H. Stolen, A. Ashkin, J. of Quantum Electronics QE-8, 548 (1972).

38) E. L. Kerr, Phys. Rev. A. 4, 1195 (1971).

39) R. W. Hellwarth, Phys. Rev. 152, 156 (1966).

40) R. Polloni, C. A. Sacchi, O. Svelto, Phys. Rev. Letters 23, 690 (1969).

41) G. A. Askaryan, JETP Letters 6, 157 (1967).

42) Y. P. Raizer, JETP Letters 4, 193 (1966).

43) G. A. Askaryan, V. B. Studenov, JETP Letters 10, 71 (1969).

44) M. S. Brodin, A. M. Kamuz, JETP Letters 13, 119 (1971).

45) A. Javan, P. L. Kelley, J. of Quantum Electronics QE-2, 470 (1966).

46) R. W. Hellwarth, A. Owyoung, N. George, Phys. Rev. A. 4, 2342 (1971).

47) N. Bloembergen, P. S. Pershan, Phys. Rev. 128, 606 (1962).

48) R. D. Small, Ph.D. Thesis, California Institute of Technology, Pasadena, California, 1973, Chapter II.

49) A. J. DeMaria, W. H. Glenn, M. J. Brienza, M. E. Mack, Proceedings of the IEEE 57, 2 (1969).

50) V. V. Korobkin, A. A. Malyutin, A. M. Prokhorov, JETP Letters 12, 216 (1970).

51) J. Comly, E. Garmire, J. P. Laussade, A. Yariv, Appl. Phys.

Letters 13, 176 (1968).

52) C. C. Cutler, Proceedings of the IRE 43, 140 (1955).

53) A. E. Siegman, D. J. Kuizenga, Appl. Phys. Letters 14, 181 (1969).

54) R. Fisher, P. L. Kelley, T. K. Gustafson, Appl. Phys. Letters 14, 140 (1969).

55) J. A. Giordmaine, M. A. Duguay, J. W. Hansen, J. Quantum Electronics QE-4, 252 (1968).

56) E. B. Treacy, J. Quantum Electronics QE-5, 454 (1969).

57) W. G. Wagner, B. A. Lengyel, J. of Appl. Phys. 34, 2040 (1963).

58) L. Dahlstrom, Optics Comm. 5, 157 (1972).

59) L. Dahlstrom, Optics Comm. 7, 89 (1973).

60) H. Statz, G. A. DeMars, C. L. Tang, J. of Appl. Phys. 38, 2212 (1967).

61) J. A. Fleck, Appl. Phys. Letters 13, 365 (1968).

62) A. Yariv, Quantum Electronics (Wiley, New York, 1967), p. 206.

63) Ibid. p. 241.

64) M. Born and E. Wolf, op. cit., p. 393-397.

65) E. Garmire, R. Y. Chiao, C. H. Townes, Phys. Rev. Letters 16, 347 (1966).

66) V. N. Lugovoi, A. M. Prokhorov, JETP Letters 12, 333 (1970).

67) A. A. Abramov, V. N. Lugovoi, A. M. Prokhorov, JETP Letters 9, 419 (1969).

68) Y. R. Shen, M. M. T. Loy, Phys. Rev. A. 3, 2099 (1971).

69) P. L. Kelley, Phys. Rev. Letters 15, 1005 (1965).

70) C. S. Wang, Phys. Rev. 173, 908 (1968).

71) A. L. Dyshko, V. N. Lugovoi, A. M. Prokhorov, JETP Letters 6, 146 (1967).

72) A. L. Dyshko, V. N. Lugovoi, A. M. Prokhorov, ZHETF 61, 2305 (1971).

73) J. A. Fleck, P. L. Kelley, Appl. Phys. Letters 15, 313 (1969).

74) E. L. Kerr, Phys. Rev. A. 4, 1195 (1971).

75) V. I. Talanov, JETP Letters 11, 199 (1970).

76) S. N. Vlasov, V. A. Petrischev, V. I. Talanov, Radiofizika 19, No. 9, 1353 (1971).

77) B. R. Suydam, Los Alamos Scientific Laboratory Report, LA-5003-MS (1973).

78) V. E. Zakharov, A. B. Shabat, Sov. Phys. JETP 34, 62 (1972).

79) V. I. Bespalov, V. I. Talanov, JETP Letters 3, 307 (1966).

80) K. A. Brueckner, S. Jorna, Phys. Rev. 164, 182 (1967).

81) J. Marburger, Optics Comm. 7, 57 (1973).

82) J. H. Marburger, E. Dawes, Phys. Rev. Letters 21, 556 (1968).

83) E. L. Dawes, J. H. Marburger, Phys. Rev. 179, 862 (1969).

84) D. W. Peaceman, H. H. Rachford, J. Soc. Indus. Appl. Math. 3, 28 (1958).

85) W. F. Ames, Numerical Methods for Partial Differential Equations (Barnes & Noble, New York, 1969), p. 246-249.

86) B. Carnahan, H. A. Luther, and T. O. Wilkes, Applied Numerical Methods, (Wiley, New York, 1969), p. 361-404.

87) W. G. Wagner, H. A. Haus, J. H. Marburger, Phys. Rev. 175, 256 (1968).

88) H. Goldstein, op. cit., p. 228-231.

89) V. V. Vorob'yev, IVUZ Radiofiz, No. 12, 1905 (1970).

90) H. A. Haus, Appl. Phys. Letters 8, 128 (1966).