AN INTEGRATED OPTICS

PULSE SHAPING DEVICE

by

Scott Roger Shepard

B.S. EE Kansas State University
B.S. Physics Kansas State University
(1979)

SUBMITTED TO THE DEPARTMENT OF
ELECTRICAL ENGINEERING AND COMPUTER
SCIENCE IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1981

© Scott R. Shepard 1981

The author hereby grants to M.I.T. the permission to reproduce and to distribute copies of this thesis document in whole or in part.

Signature of Author _____
Department of Electrical Engineering
and Computer Science May 28, 1981

Certified by _____
Herman A. Haus
Thesis Supervisor

Accepted by _____
A.C. Smith
Chairman, Departmental Graduate Committee

Archives
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
OCT 20 1982
LIBRARIES

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

AN INTEGRATED OPTICS

PULSE SHAPING DEVICE

by

SCOTT ROGER SHEPARD

Submitted to the Department of Electrical Engineering and Computer Science on May 28, 1981 in partial fulfillment of the requirements for the Degree of Master of Science in Electrical Engineering and Computer Science

ABSTRACT

A theoretical study of a proposed travelling wave pulse shaping device is performed. The device consists of two evanescently coupled single mode dielectric waveguides. One waveguide is amplifying, the other is lossy, and both may exhibit saturation. Four nonlinear partial differential equations are derived from Maxwell's equations and Schodinger's equation, in the semiclassical manner, to define the system's dynamic model. The steady state limit is then defined and described by a second order nonlinear autonomous differential equation. Critical point analysis of the steady state equations reveal three real valued critical points for which stability criteria are given. Steady state solutions for ten different characteristic behaviors are obtained by two different algorithms. One involves the use of symbolic manipulation software to provide Taylor series expansions of the solutions in the phase plane and the other is a numerical method. The former algorithm may be used to provide approximate solutions of any second order autonomous system for any phase plane region with any specified accuracy. The results correlate well with the critical point predictions and physical interpretations are offered for each case.

A simplifying transformation is performed on the dynamic equations. These transformed equations are then solved by the use of nonlinear finite difference equations. Conditions necessary for the convergence and stability of the finite difference scheme are derived. These conditions are empirically shown to be sufficient as well as

necessary. The dynamic results demonstrate the two major pulse shaping applications: 1) pulse height standardization (i.e. optical limiting), and 2) pulse width shortening. On the latter application it is shown that pulses with widths smaller than the medium's relaxation time may be generated by the saturation effects of either waveguide by itself (i.e. when the other waveguide is linear).

Device parameters are optimized for the pulse height standardizer (PHS) in the steady state limit. When the input amplitude varies uniformly over a 2 to 1 range (an input signal to noise ratio of 14.31 dB) the output of a realizable "optimal" PHS device has a signal to noise ratio of 23.3dB. The effects of pulsewidth and risetime on the PHS performance are then demonstrated.

Other applications include the use of the device as an amplifier, oscillator, logic element, comparator, and general pulse shaper.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Herman A. Haus for the original conception of the proposed device and for his patient and willing assistance in the development and supervision of this thesis. I am also in debt to the Mathlab Group of the Laboratory for Computer Science at M.I.T. for the use of the MACSYMA software which was developed under contract number E (11-1)-3070 of the U.S. Energy Research and Development Administration and grant NSG 1323 of NASA. I would also like to thank the administration of Bell Telephone Laboratories (especially at the Merrimack Valley Laboratory) for their support. I also thank Jan Hammond for her constant support.

CHAPTER I - PULSE SHAPING AT OPTICAL FREQUENCIES

In recent years a wide variety of optical devices have been considered for the possible development of several high speed discrete and analog signal processing components--logic elements, memory elements, differential amplifiers, and the like. To insure the proper operation of any of these components within an actual system, however, it is often essential that some type of signal filtering, or pulse shaping, be performed. One important example of this is the necessity of pulse height standardization (i.e., optical limiting) to make feasible the use of various optical devices as logic elements within an integrable system. There are, of course, numerous other examples.

This chapter briefly reviews some conventional approaches to pulse shaping at optical frequencies, introduces the more recent use of integrated optical devices as signal processors, and concludes with the presentation of a novel travelling wave device which will be modeled, analyzed and developed in the remainder of this thesis.

1.1 Some Conventional Optical Pulse Shaping Devices

An example of a conventional optical pulse shaping device is the saturable optical resonator (SOR). This is simply an optical resonator which may be characterized by several parameters, one or more of which may saturate (i.e., approach

a limiting value) due to the presence of the resonating optical field. The saturation phenomena cause the resonator transmissivity to become a nonlinear function of the incident optical intensity. Under certain conditions this nonlinear function may become multiple-valued so that the resonator exhibits multistability. While multistability is essential for memory element applications, it is the highly nonlinear transmission characteristic that makes the saturable resonator useful in optical pulse shaping.

The earlier saturable optical resonators were implemented "in bulk", i.e., the resonators were physically large and unsuited to any scheme of integration. One of the first types of SOR analyzed was that which Szöke et. al. discussed in 1969.[1] The SOR consisted of a Fabry-Perot type cavity resonator that contained a saturably absorptive medium. Saturably absorptive media have an absorption that decreases with increasing intensity[2] (decreasing absorption with increasing intensity is often referred to as bleaching). The incident intensity was provided by a Q-switched CO_2 laser. Since the gain medium was external to the SOR, the SOR could be thought of as a passive "cavity dumping" device.

Szökes' model of the SOR consisted primarily of the combination of the steady state field equations of the Fabry-Perot with the steady state solution of the dipole rate

Portions of the text on the
following page(s) are not legible
in the original.

equation for the absorbtion coefficient. From this model he obtained steady state solutions for the resonator intensity versus incident intensity via both computer calculations and an approximate formula valid for large mirror reflectivity (R>1). A reproduction of these results obtained from reference [1] is presented in figure 1.

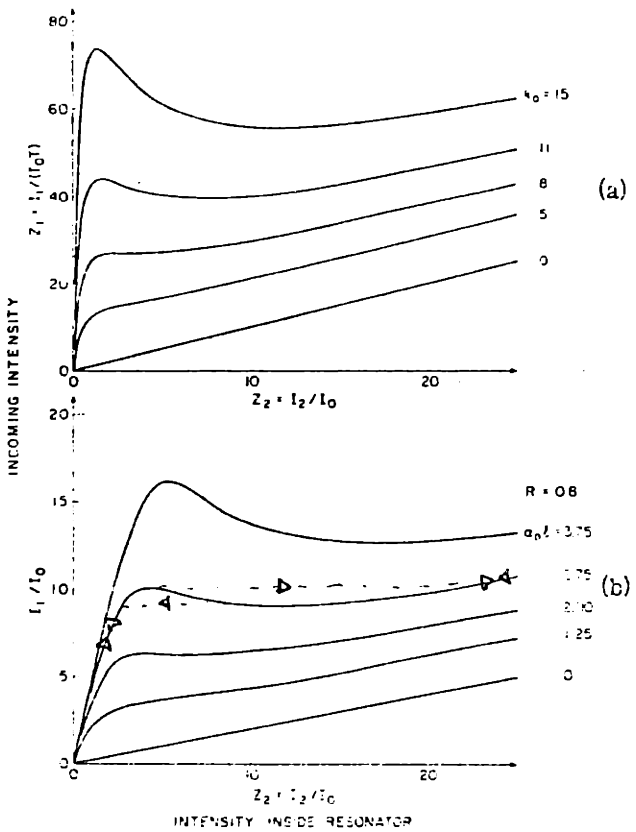


FIGURE 1- SOR Transfer Curves

a) approximate formula
b) numerical solution

The figure shows solutions of incident intensity versus resonator intensity for various values of the unperturbed absorption coefficient, α_0 . The Fabry-Perot is assumed to be on resonance. Notice that when the initial absorption coefficient is zero (lossless medium) the transfer

characteristic is a linear one. When the medium within the SOR is lossy ($\alpha_0 \neq 0$) some input intensity will be required to bleach the absorber before energy can penetrate into the resonator. The resulting nonlinearity of the transfer characteristic becomes more pronounced with increasing α_0 until α_0 exceeds a particular value for which the characteristic curve obtains both a maximum and a minimum. When this occurs the resonator intensity becomes a multiple-valued "function" of the incident intensity. The SOR is then said to be operating in its bistable mode where hysterisis effects may be observed as I have illustrated by superimposing dotted lines and arrows on one of the curves of figure 1.

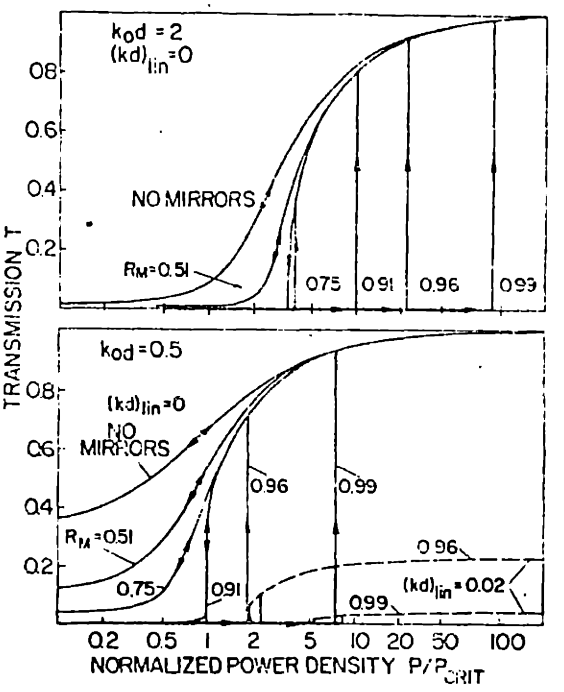
Experimentally Szoke et. al. could not demonstrate bistable operation of the SOR since the values of α_0 that were achievable for the gaseous absorptive medium used were less than the required value. Pulses with a full width at half maximum pulse width of 135 nanoseconds were shortened to 100 nanoseconds by use of the SOR*. The SOR also performed additional pulse shaping by substantially

*Various other applications of the SOR suggested by Szoke include Q-switching, laser output stabilization, pulse train generation, logical operations (by use of a probe beam), and memory functions. Descriptions of SOR mode-locking and pulse shaping may be found in references [3] and [4].

reducing the energy in the tail of the pulse. The two main limitations of the analysis, as recognized by Szöke were the neglect of the effects of linear loss (the residual absorption due to the inadequacy of the two level absorber model) and the neglect of the effects of coupling between forward and backward waves within the saturable resonator. This last effects is expected to result in Bragg reflections from a periodic "grating" produced by standing waves. It should also be emphasized that the steady state model used is only valid when the incident intensity is slowly varying over time intervals comparable to the absorbers' relaxation time.

Several other authors have analyzed saturable optical resonators of similar construction. Notably, in 1971 Spiller considered the effects of both residual (linear) loss and cavity detuning on the SOR.[5] The same steady state model was used. Spiller obtains the same results as Szöke for the case of no linear loss and no cavity detuning. These results were presented as plots of the resonator transmissivity versus incident power density for different mirror reflectivities, R_M, and are reproduced from reference [5] in figure 2.

FIGURE 2 - SOR Transmission for Different Mirror Reflectivity



Notice that when $R_M = 0$ the result is simply the single pass steady state transfer characteristic of a saturable absorber. When the mirror reflectivity is increased the result is a steepening of the transfer characteristic as one would expect from a cascade of single pass sections. When R_M exceeds a certain value, hysteresis loops are formed and bistable operation of the SOR is achieved.

Cavity detuning and residual losses both degrade SOR performance in a similar fashion. The presence of either will reduce both the maximum transmissivity of the SOR and the area encompassed by the hysteresis loops. This

area decreases continuously until it vanishes thus eliminating the probability of bistable operation.

Spiller presents some transient response data although the steady state equations used can not be expected to remain valid. These results are therefore of dubious value but it is claimed that the SOR will change states due to an instantaneous input switching in 5 picoseconds after an initial "punch-through" delay of 10 nanosecond for the numerical example considered.

Experimentally Spiller could not operate in the bistable region due to the linear loss of the saturable absorber used. Pulses of 2 nanosecond pulse widths (FWHM) were shortened to 1 nanosecond pulse widths by use of the SOR.

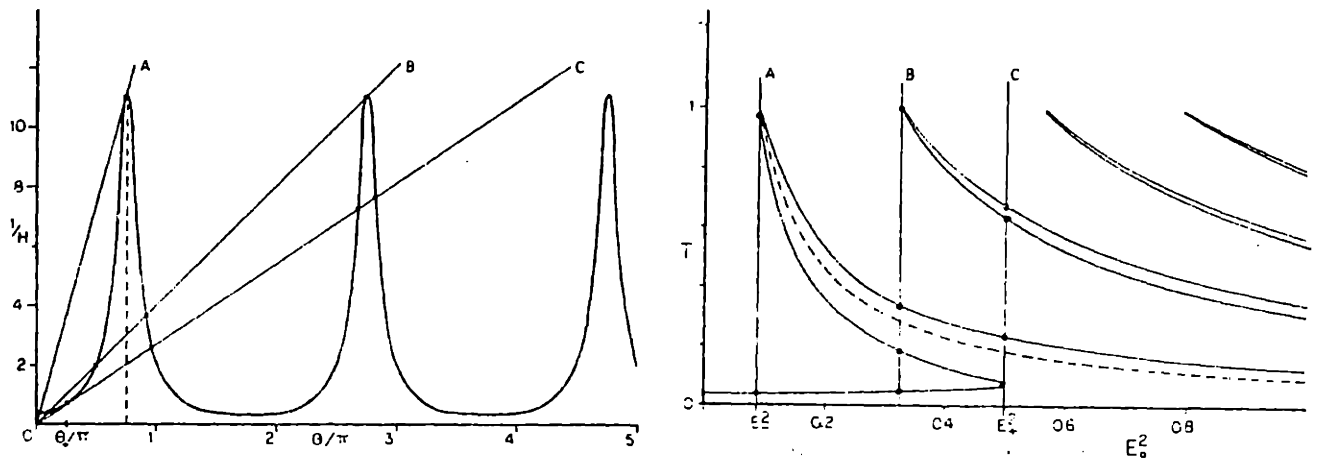
A more theoretical treatment of some similar processes is offered by McCall in reference [6]. Bistable operation of a SOR was demonstrated experimentally by McCall et. al. in 1975 [7], [8].

Other examples of conventional optical pulse shaping devices are often comprised of the same nonlinear Fabry-Perot structure as the SOR but the nonlinearity is achieved through the exploitation of a different physical phenomenon. Stark,[9] Raman, and Brillouin[10] effects have been examined as potential sources of the required nonlinearity. A large number of nonlinear Fabry-Perots utilize an intensity dependent (nonlinear) phase retardation. These devices often consist

of a Kerr medium within a cavity resonator. The nonlinearity is introduced by the intensity dependent refractive index of the medium.

The first analysis of this type of device was performed by Felber and Marburger in 1976.[11] The results are presented in figure 3.

FIGURE 3 - Kerr-filled Nonlinear Fabry-Perot Interferometer



a) Transmission vs. Phase

b) Transmission vs. Incident Intensity

The familiar Fabry-Perot resonator transmissivity, T, versus phase delay, θ, characteristic is presented in part (a) of figure 3 along with three lines that show the linear dependence of phase on intensity, $\theta = v \langle E^2 \rangle$, for three different values of incident intensity, E_0^2 (note that T is

proportional to $\langle E^2 \rangle / E_0^2$). The system operating point is located at the intersection of the F-P curve* and the line that corresponds to the present value of incident intensity. Keeping in mind that the slopes of the straight lines are inversely proportional to the incident intensity, the device operation may be illustrated as follows. As the incident intensity increases from some initially small value the resonator operating point will move slightly upward (increasing T) along the first part of the F-P curve (about half way to the first F-P peak). Thus the operating point would move from the first intersection with line A to the first intersection with line B to the first intersection with line C. At this first intersection with line C, a further increase in incident intensity would reduce the slope of the phase line so that the operating point would have to jump discontinuously to the other side of the first F.P. transmission peak where the transmissivity is somewhat larger. Further increase in E_0^2 would decrease the value of T whereas a decrease in input intensity at this point would increase T. As E_0^2 is decreased at this point the operating point would move upward along the right hand side of the first F-P peak through another intersection with line B until it intersects line A at the top of the F-P peak. Further decrease in input intensity would cause the operating point to jump back to the low transmissivity

*I.e. the Fabry-Perot transfer characteristic.

region where this discussion originated. These results may be presented as a plot of T versus E_0^2 which is done in part (b) of figure 3. Feasibility of multistable and bistable operation is apparent due to the hysteresis effects as described above. By affecting the index of refraction of the Kerr medium with a control or probe beam the device may be operated as an amplifier. Due to the nonlinear transfer characteristic the device may also be operated as an optical limiter.

In 1977 Bischofberger and Shen[12] obtained the first set of experimental results for the Kerr filled resonator. They also developed models for description of the devices transient behavior that are in excellent agreement with the empirical results. Operation of the device as a power limiter, a differential amplifier, and a bistable element was demonstrated for input pulses of widths on the order of tens of nanoseconds. The transient response of such a device has been considered in greater detail most recently (March 1981) by Goldstone and Garmire[13]. In this article the device is operated in the bistable mode and conditions are derived for overshoot switching (i.e., switching from one bistable state to another at intensities smaller than the switching intensity indicated by the steady state transfer curve). Overshoot switching is dependent on the rise time of the input pulse and the ratio of cavity roundtrip time,

τ_c, to the response time of the nonlinear medium, τ_n. Overshoot switching is optimal for $\tau_c/\tau_n = 1$ and requires an input rise time less than a few τ_c. Also considered is the effect of dead time (i.e., the duration of zero input intensity) between pulses in a pulse train input. It is shown that pulses are processed independently (as far as overshoot switching is concerned) when dead times are $\geq 3\tau_c$. For dead times less than $3\tau_c$ there are four different pulse train transmission modes for four different ranges of dead time.

1.2 Integrated Optical Signal Processors

If any of the aforementioned optical signal processing devices are to be of practical use they must be implemented in an integrable form. Several integrated optical devices have been considered and fabricated, a few of which will be discussed here.

In 1976 Okuda et.al. proposed the use of distributed Bragg-reflectors on each side of an optical waveguide made of either a saturable absorber[14] or a Kerr medium[15] as integrated implementations of the two major types of nonlinear resonators. The Bragg-reflectors, which are simply rectangular (in this case square) periodic corrugations in one surface of the waveguide, produce reflections so that the physics of the devices is expected to be very similar

to that of the bulk versions. The spatial distribution of propagation modes (forward and backward) is assumed to be time independent so that conventional coupled mode theory may be used. These coupled mode equations used in conjunction with the steady state saturable loss equation or the intensity dependent refractive index equation (depending of course on which device is considered) define the systems steady state response. The results are exactly analogous to those obtained for the bulk resonators with the exception that the mirror reflectivity parameter, , has been replaced by the coupling coefficient which describes the amount of coupling between the forward and backward propagating modes as produced by the corregations. Experimental results are not presented nor is an attempt at dynamic analysis made. The dynamic results of Goldstone and Garmire for the bulk version of the Kerr medium resonator are probably applicable to Okuda's integrated version; however, there may be some differences in the dynamic response of the two different types of "mirrors".

Another interesting integrated optical signal processor is that proposed in 1978 by Gray and Casperson.[16] Although this device is normally operated as an optical transistor, and is therefore a pulse shaping device only in the most general sense, it is discussed here since the structure is somewhat similar to the one that is to be

proposed and developed in this thesis. Gray and Casperson's device employs the cross-saturation effects of two coupled propagation modes within a gain medium to produce an optical inverter. The operation of the inverter may be simply described as follows. The two modes will experience two different values of net gain in the waveguide direction (since photons of the higher order mode travel through more medium for the same distance in waveguide directions they experience more net gain). The higher gain mode will be more sensitive to gain fluctuations so that it may be modulated by the saturation effects produced by the lower order mode. Amplifiers, inverters, comparators, and logic devices may be fabricated in this fashion. Although the basic process does not demonstrate bistability; memory elements, controlled oscillators, etc. may be fabricated by optically feeding back the various output signals to the appropriate inputs in a manner exactly analogous to techniques common in electronics. The coupled mode equations used in the analysis, however, were again time independent so that the systems dynamic response was not considered. No experimental results were reported.

An analysis of the dynamic characteristics of this type of inverter has recently (January 1981) been performed by Otsuka.[17] This analysis is based on a lumped model which utilizes the mean value theorem and the assumption of

small device lengths. It is shown that the inverter will operate on rectangular pulse trains with pulse widths comparable to the medium relaxation time for a required dead time of about three times the pulse width. The amplifier is used in conjunction with an integrated resonator to produce a saturable oscillator for the implementation of an optical limiter; however, the dynamic response of the limiter is not considered (i.e., steady state equations for the oscillator are used).

The device to be proposed, analyzed, and developed in this thesis is an integrated optics active traveling wave structure. Physically the device consists of two single-mode slab dielectric waveguides embedded in a substrate and separated by a distance small enough to permit some evanescent mode coupling between the guides. One waveguide is amplifying (for which an inversion pump is supplied) and the other is lossy. Both the gain and loss media may exhibit saturation. The device is similar to a saturable resonator except that pulse energy is coupled from one waveguide to the other and back again as it travels through an initially unperturbed medium rather than travelling through the same medium as it reflects between resonator mirrors (i.e., the device is a travelling wave device rather than a resonator). The nature of the mode coupling is considerably different from that of Gray and Casperson's device since in the

proposed structure each mode saturates only its own waveguide medium. Some of the anticipated advantages of this type of pulse shaping device are:

1. the degrading effects of absorption (both linear and saturable loss) experienced in the saturable optical resonator should be compensated by the presence of gain;

2. any initial "punch-through" delay for penetration of the first mirror of a nonlinear Fabry-Perot should be eliminated by the traveling wave structure;

3. the traveling wave nature of the device should allow considerable reduction of the requisite dead time between pulses for independent processing;

4. the pulse response should be significantly faster than the characteristic response times of either medium due to the combined saturation effects which are easily controlled.

These advantages and several others will be demonstrated later. The device shall normally be operated as a single input single output device since the primary application of interest is the shaping of optical pulses. Other types of signal processing may be achieved by use of this pulse shaper in conjunction with available components or by

operating the device in a triode mode. A time dependent coupled mode theory will be developed so that the systems dynamic response may be analyzed.

CHAPTER II - SEMICLASSICAL DERIVATION OF THE DYNAMIC EQUATIONS

The equations governing the most general behavior of the proposed travelling wave pulse shaping device are derived in this chapter. These equations describe the transient or dynamic response of the system. The derivation of the equations is exact except for the semiclassical assumption which neglects the quantum nature of the electromagnetic radiation. This approximation yields highly accurate results when applied to systems operating at frequencies within the visible spectrum and is thus valid for the actual devices at intensities capable of saturating a nonlinear medium. In order to derive these equations a time dependent coupled mode theory is developed as a simple extension of the conventional theory. These coupled mode equations are then used in conjunction with the appropriate rate equations (essentially Schödinger's equation cast in the density matrix formalism) to form the system's mathematical model. A dimensionless form of the dynamic equations is then defined. From the dynamic equations, two sets of "dominant saturation" equations are derived for two limiting cases of practical importance.

2.1 Theory of Coupled Modes in Space and Time

Yariv[18] has developed a formalism to describe coupling between the various modes or propagation of

electromagnetic radiation within a guided-wave structure when the mode amplitudes are time invariant. To describe the transient behavior of guided-wave devices a theory of coupling between mode amplitudes that are time variant as well as space dependent is required. The extension from Yariv's formulation is a straightforward one.

If the polarization within the medium may be expressed as

(Eq. 2.1-1) $$\bar{P}(\bar{r},t) = (\varepsilon(\bar{r}) - \varepsilon_0)\bar{E}(\bar{r},t) \equiv \bar{P}_0(\bar{r},t)$$

then there will be no coupling of modes as is the case, for example, of any linear isotropic medium.

The polarization, however, will not always be expressable in the above form. If, for example, the medium is anisotropic or nonlinear, or if external sources are present then the polarization must be described as

(Eq. 2.1-2) $$\bar{P}(\bar{r},t) = \bar{P}_0(\bar{r},t) + \bar{P}_{pert}(\bar{r},t)$$

where $\bar{P}_{pert}(\bar{r},t)$ is the so called perturbing polarization due to the external source, the bianisotropy of the medium, corrugations in the waveguide surface, or whatever else is causing the deviation between $\bar{P}$ and $\bar{P}_0$. Systems in which polarizations of this more general form occur may be analyzed in terms of coupled modes.

From Maxwell's equations the following wave equation may be derived for each component of the electric field.

(Eq. 2.1-3) $$\nabla^2 E_i(\bar{r},t) = \mu\varepsilon(\bar{r})\frac{\partial^2 E_i(\bar{r},t)}{\partial t^2} + \mu\frac{\partial^2}{\partial t^2}\left(\bar{P}_{pert}(\bar{r},t)\right)_i$$

Where i = x, y, or z indicates the specific vector component. Since the possible coupling of guided modes to radiation modes is to be ignored (radiation losses for the structures considered are small and may, if necessary, be included phenomenologically) the electric field may be expanded in terms of the confined propagation modes. For example, considering only plane waves that travel in the +z direction,

(Eq. 2.1-4)

$$E_y(\bar{r},t) = (1/2)\mathrm{Re}\left\{\sum_m A_m(z)E_y^{(m)}(x)e^{i(\omega t - \beta_m z)}\right\}$$

where the mode amplitudes A_m are allowed to vary with z. This expression may be used to describe the fields within a system that has achieved temporal equilibrium or steady state. For such a system the distribution in mode amplitudes at any point along the waveguide does not vary in time (it may of course vary in distance). To consider a dynamic situation, i.e., to determine the transient behavior of the mode amplitudes due to the turning on or off, or temporal variation, of the perturbation, one must allow for the time dependence of the mode amplitudes, thus

(Eq. 2.1-5)

$$E_y(\bar{r},t) = (1/2)\mathrm{Re}\left\{\sum_m A_m(z,t)E_y^{(m)}(x)e^{i(\omega t - \beta_m z)}\right\}$$

Using this expression and the "slowly" varying envelope approximations $\left|\frac{\partial^2 A_m(z,t)}{\partial z^2}\right| << \left|\beta_m \frac{\partial A_m(z,t)}{\partial z}\right|$

and $\left|\frac{\partial^2 A_m(z,t)}{\partial t^2}\right| << \left|\omega \frac{\partial A_m(z,t)}{\partial t}\right|$

which simply require that the mode amplitudes vary only slightly over a distance of an optical wavelength and within a time interval of an optical period, the wave equation becomes

(Eq. 2.1-6)

$$\mu \frac{\partial^2}{\partial t^2}\{\bar{P}_{pert}(\bar{r},t)\}_y = \mathrm{Re}\left\{\sum_m \left(-i\beta_m \frac{\partial A_m}{\partial z} E_y^{(m)}(x) e^{i(\omega t-\beta_m z)} - i\omega\mu\epsilon \frac{\partial A_m}{\partial t} E_y^{(m)}(x) e^{i(\omega t-\beta_m z)}\right)\right\}$$

Multiplying this equation by some normalized field of another mode, say $E_y^{(\ell)}(x)$, and using the orthogonality property of the modes (i.e. $\int_{-\infty}^{\infty} E_y^{(m)}(x) E_y^{(\ell)}(x)dx = \frac{2\mu\omega}{\beta_m}\delta_{\ell,m}$)

integration of both sides of the equation yields

(Eq. 2.1-7)

$$\frac{\partial A_m}{\partial z} + \frac{\omega\mu\epsilon}{\beta_m}\frac{\partial A_m}{\partial t} = \frac{i}{2\omega}\frac{\partial^2}{\partial t^2}\left\{\int_{-\infty}^{\infty}[\bar{P}_{pert}(\bar{r},t)]_y E_y^{(m)}(x)dx\right\}$$

The term on the right hand side of the above equation is a source term, i.e., it represents energy that is either imparted to or taken from the energy of the mth mode due to the presence of the perturbation. The source term may be attacked analytically, but a phenomenological derivation of the relevant terms is simpler and provides greater insight

to the underlying physical processes.

As an example, consider the case of a single isotropic waveguide. Suppose a perturbation is induced in the form of an external pump so that a population inversion takes place causing light wave amplification. To examine the steady state behavior of the system the mode amplitudes are assumed not to vary in time. The resulting equation would be

$$\frac{\partial A_m}{\partial z} = \frac{i}{2\omega} \frac{\partial^2}{\partial t^2} \left\{ \int_{\infty}^{\infty} [\bar{P}_{pert}(\bar{r},t)]_y E_y^{(m)}(x)dx \right\}$$

Neglecting gain saturation, the source term would take the form GAm, where G is the amplitude gain (of the mth mode) per unit length. Thus $\frac{dAm}{dZ}$ = G Am (G>O) would describe the exponential gain in a nonsaturating gain medium.

Similarly if the medium had nonsaturating losses that were not incorporated into the dielectric constant ε, then the following equation would hold $\frac{dAm}{dZ}$ = - L Am (L > O) where L is the loss per unit length.

To account for gain and loss saturation effects the above equations would still be used except that G and L would become functions of Z. The solution could be obtained by combining the above mode equations with the steady state rate equations that describe the medium saturation effects.

As another example consider two waveguides that are adjacent to each other closely enough that the evanescent

fields from each waveguide extend slightly into the other waveguide. Although this evanescent field does not propagate any power by itself,* it does have energy that may be viewed as an external source term that perturbs, or couples energy to, the propagating modes of the waveguide that it has penetrated into. The amount of energy that is added or subtracted is proportional to the perturbing evanescent field which is proportional to its total source, the mode amplitude of the other waveguide. Thus if we consider two closely adjacent single-mode waveguides in the time invariant case, we would obtain the equations $\frac{dA1}{dz} = KA2$, $\frac{dA2}{dz} = -KA1$, where A1 and A2 are the amplitudes of the single modes of propagation (one for each of the two waveguides) and K is the coupling constant. The signs of the source terms are arbitrary as long as one is positive and the other negative, as may be proved by power conservation arguments for modes traveling in the same direction, or from evaluating the time derivatives of the polarization integral.

The dynamic behavior of two closely adjacent single mode waveguides, one of which is amplifying and the other of which is lossy (not described by ε) must be governed by

(Eq. 2.1-8)

$$\frac{\partial A1}{\partial z} + \frac{1}{v}\frac{\partial A1}{\partial t} = G\ A1 + K\ A2$$

$$\frac{\partial A2}{\partial z} + \frac{1}{v}\frac{\partial A2}{\partial t} = -L\ A2 - K\ A1$$

*I.e. in the direction perpendicular to the waveguide interface. The evanescent field does propagate a surface wave at the interface but this is not the coupled energy term being described.

where the unperturbed speed of propagation $v = \frac{\beta}{\omega\mu\varepsilon}$ is assumed the same for the two waveguides. This result may of course be derived more rigorously. Notice that the time dependence of the mode amplitudes does not affect the form of the source terms with the exception that they may now be time dependent also. Therefore in general we would have

$L = L(z,t)$, $G = G(z,t)$.

2.2 Gain and Loss Saturation Rate Equations

To analyze saturation of both the loss and the gain media the semiclassical approach is used. Thus the media are treated quantum mechanically and the fields are treated classically. The interconnection between media and fields is provided by the electric dipole interaction.

Consider a classical electromagnetic plane wave incident upon a quantum mechanical two-level electronic system (e.g., an atom or molecule). According to the density matrix formulation of quantum mechanics a rate equation may be written for the difference in probabilities of upper and lower energy eigenstate occupation as follows

(Eq. 2.2-1)
$$\frac{\partial(\rho_{11}-\rho_{22})}{\partial t} + \frac{(\rho_{11}-\rho_{22}) - (\rho_{11}-\rho_{22})^{e}}{T1} = \frac{[H',\rho]_{11}-[H',\rho]_{22}}{i\hbar}$$

where

ρ_{11} is the (1,1) matrix element of the density operator which is the probability of being in the upper energy eigenstate,

ρ_{22} is the probability of being in the lower energy eigenstate,

H' is the perturbing Hamiltonian (for our case this is the electric dipole interaction Hamiltonian),

$[H',\rho]_{ii}$ is the (i,i)th element of the commutator of the perturbing Hamiltonian operator H', and the density operator ρ,

$(\rho_{11}-\rho_{22})^e$ is the equilibrium (i.e. unperturbed) population inversion,

and $T1$ is the characteristic relaxation time that incorporates the effects of spontaneous emission, lattice interactions, inelastic collisions, etc.

Now we consider a homogenous ensemble of these electronic systems. We define macroscopic energy level population densities by simply multiplying the appropriate

density matrix element by the average number of atoms/molecules per unit volume, Nv. Thus $N_2 = N_v \rho_{11}$ is the average number of electronic systems in the lower energy eigenstate, similarily $N_1 = N_v \rho_{22}$. Now using the electric dipole interaction energy for the perturbing Hamiltonian it can be shown that Eq. 2.2-1 becomes

(Eq. 2.2-2)

$$\frac{\partial(N_1-N_2)}{\partial t} + \frac{(N_1-N_2)-(N_1-N_2)^e}{T1} = \frac{-2}{\hbar\Omega}(\dot{\bar{P}}\cdot\bar{E})$$

where $\dot{\bar{P}}$ is the time derivative of the polarization

$\bar{E}$ is the local electrical field

and $\hbar\Omega$ is the energy difference between the two levels.

The instantaneous loss seen at the photon-electron interaction site is proportional to the probability of absorption, which is proportional to $\rho_{22} - \rho_{11}$. Therefore the macroscopic loss is proportional to $N_2 - N_1$. Similarly the macroscopic gain is proportional to $N_1 - N_2$. Thus the rate equation describing the population level saturation (and other) effects for the macroscopic loss, L, is

(Eq. 2.2-3)

$$\frac{\partial L(z,t)}{\partial t} + \frac{L(z,t) - L^e}{T1} = \left(\frac{2}{\hbar\Omega}\dot{\bar{P}}\cdot\bar{E}\right)\times(\text{some positive constant})$$

Similarly the rate equation for the macroscopic gain, G, is

(Eq. 2.2-4)

$$\frac{\partial G(z,t)}{\partial t} + \frac{G(z,t)-G^{e}}{T1} = -\left(\frac{2}{h\Omega}\,\dot{\bar{P}}\cdot\bar{E}\right)\text{x(some positive constant)}$$

The source terms on the right hand side of these rate equations may each be expressed in terms of the mode amplitude of the appropriate waveguide. Noting that the polarization is proportional to the product of the susceptibility and the electric field, and that the imaginary part of the susceptibility is proportional to $N_2 - N_1$, it may be shown that Eq. 2.2-3 and Eq. 2.2-4 can be written as

(Eq. 2.2-5)

$$\frac{\partial L(z,t)}{\partial t} + \frac{L(z,t)-L^{e}}{T1} = -\frac{(A2)^{2}L(z,t)}{T1_{(L)}PS_{(L)}}$$

and

(Eq. 2.2-6)

$$\frac{\partial G(z,t)}{\partial t} + \frac{G(z,t)-G^{e}}{T1} = -\frac{(A1)^{2}G(z,t)}{T1_{(G)}PS_{(G)}}$$

where the constant PS is often referred to as the saturation intensity level. Note the subscripts reflect the fact that the two different waveguide media may be characterized by different parameters, PS, and T1.

2.3 A Dimensionless Form of the Dynamic Equations

Four partial differential equations have been derived which model the dynamic behavior of the mode amplitudes of two single-mode waveguides (one of which exhibits saturable gain, the other exhibits saturable loss) that are evanescently coupled. These equations are repeated here for completeness.

(Eq. 2.3-1)

$$\frac{\partial A1}{\partial z} + \frac{1}{v}\frac{A1}{\partial t} = G\ A1 + K\ A2$$

$$\frac{A2}{\partial z} + \frac{1}{v}\frac{A2}{\partial t} = -L\ A2 - K\ A1$$

$$\frac{\partial L}{\partial t} + \frac{L-L^{e}}{T1_{(L)}} = -\frac{(A1)^{2}L}{T1_{(L)}PS_{(L)}}$$

$$\frac{\partial G}{\partial t} + \frac{G-G^{e}}{T1_{(G)}} = -\frac{(A2)^{2}G}{T1_{(G)}PS_{(G)}}$$

A dimensionless form of the above system of equations is desired. For the present assume $T1_{(L)} \simeq T1_{(G)} \equiv T1$ and $PS_{(L)} \simeq PS_{(G)} \equiv PS$. Later this restriction may be removed. Now define the following set of dimensionless parameters:

(Eq. 2.3-2)

$$a1 \equiv A1/\sqrt{PS} \qquad a2 \equiv A2/\sqrt{PS}$$

$$\ell \equiv L/L^{e} \qquad g \equiv G/G^{e} \qquad k \equiv v\ T1\ K$$

$$y \equiv t/T1 \qquad x \equiv z/(v\ T1)$$

$$\ell_{0} \equiv v\ T1\ L^{e} \qquad g_{0} \equiv v\ T1\ G^{e}$$

So that a dimensionless form of the dynamic equations is

(Eq. 2.3-3)

$$al_x + al_y = g_0\, g\, al + k\, a2$$
$$a2_x + a2_y = -\ell_0\, \ell\, a2 - k\, al$$
$$\ell_y = 1 - \ell(1 + (al)^2)$$
$$g_y = 1 - g(1 + (a2)^2)$$

where the subscripts, of course, indicate partial derivatives.

2.4 Dominant Saturation Equations

Systems are often encountered in which the saturation of either the loss or the gain (but not both) is the predominant effect. In other words either $PS_{(G)}$ or $PS_{(L)}$ is so large (in comparison to the other) that it is essentially infinite. When this happens, the corresponding parameter, L or G, will change very little with time and may be assumed to remain at its unperturbed value. If the loss saturation effects are dominating then the following set of equations would be appropriate.

(Eq. 2.4-1)

$$al_x + al_y = g_0\, al + k\, a2$$
$$a2_x + a2_y = -\ell_0\, \ell\, a2 - k\, al$$
$$\ell_y = 1 - \ell(1 + (a2)^2)$$

Similarily, if the effects of gain saturation are the dominating phenomenon then we would arrive at the following set of equations.

(Eq. 2.4-2)

$$a1_x + a1_y = g_0\, g\, a1 + k\, a2$$

$$a2_x + a2_y = -\ell_0\, a2 - k\, a1$$

$$g_y = 1 - g(1 + (a1)^2)$$

CHAPTER III - STEADY STATE RESPONSE AND STABILITY CRITERIA

This chapter deals entirely with the steady state behavior of the proposed travelling wave pulse shaping device. Steady state is first defined, then the appropriate equations are derived from the dynamic equations. The applicability of the steady state limit to several real situations is discussed. It is shown that the steady state equations define a set of critical points for which a linearized analysis is performed. Following a brief introduction to the theory of critical points, critical point stability criteria are given. Approximate and numerical phase plane solutions demonstrating all possible types of steady state behavior are generated. The approximate algorithim which utilizes the symbolic manipulation software of MACSYMA,[19] may be applied to any second order nonlinear autonomous differential equation. Physical interpretations of the steady state results are offered and potential device applications are enumerated.

3.1 Derivation of the Steady State Equations

The systems steady state response may be defined as the asymptotic limit as time goes to infinity of the device response due to constant input amplitudes. Since steady state is a limiting concept it may never be achieved exactly in practice. It will however yield an accurate approximation of the response due to input amplitudes that are essentially

constant over a time interval of several characteristic time constants, T1. For constant input amplitudes, $a1(0,y)$ = constant 1 and $a2(0,y)$ = constant 2, the loss and gain saturation equations show that the loss and gain parameters at the input, $l(0,y)$ and $g(0,y)$, will exponentially approach steady state values that are time invariant. Due to the continuity of the mode amplitudes in space, temporal invariance will propagate through the waveguides as steady state ensues. In the steady state limit, therefore, all variables become independent of time for all x. Setting the partial derivatives with respect to y equal to zero in the loss and gain saturation equations results in the steady state loss and gain equations

(Eq. 3.1-1)

$$\ell_{ss}(x) = \frac{1}{(1 + [a2_{ss}(x)]^2)} \qquad g_{ss}(x) = \frac{1}{(1 + [a1_{ss}(x)]^2)}$$

Similarily, the steady state limit of the mode amplitude equations is

(Eq. 3.1-2)

$$\frac{d\, a1_{ss}(x)}{dx} = g_0\, g_{ss}(x)\, a1_{ss}(x) + k\, a2_{ss}(x)$$

$$\frac{d\, a2_{ss}(x)}{dx} = -\ell_0\, \ell_{ss}(x)\, a2_{ss}(x) - k\, a1_{ss}(x)$$

Combining the above two sets of equations results in the folowing two nonlinear autonomous ordinary differential equations that describe the steady state behavior of the system.

(Eq. 3.1-3)
$$\frac{d\,al_{ss}}{dx} = \frac{g_0}{(1 + al_{ss}^2)}\, al_{ss} + k\, a2_{ss}$$
$$\frac{d\,a2_{ss}}{dx} = \frac{-\ell_0}{(1 + a2_{ss}^2)}\, a2_{ss} - k\, al_{ss}$$

Analogously, the steady state response of the dominant loss saturation device can be shown to be governed by

(Eq. 3.1-4)
$$\frac{d\,al_{ss}}{dx} = g_0\, al_{ss} + k\, a2_{ss}$$
$$\frac{d\,a2_{ss}}{dx} = \frac{-\ell_0}{(1 + a2_{ss}^2)}\, a2_{ss} - k\, al_{ss}$$

The steady state limit of the dominant gain saturation equations results in

(Eq. 3.1-5)
$$\frac{d\,al_{ss}}{dx} = \frac{g_0}{(1 + al_{ss}^2)}\, al_{ss} + k\, a2_{ss}$$
$$\frac{d\,a2_{ss}}{dx} = -\ell_0\, a2_{ss} - k\, al_{ss}$$

For the remainder of this chapter the "s.s." subscript will not be used since all further discussions will refer to the steady state.*

3.2 Critical Point Analysis of the Steady State Equations

3.2.1 Theory of Critical Points

Consider a nonlinear system of coupled first order differential equations of the following form [20]

(Eq. 3.2.1)
$$\dot{\phi}_1(z) = f_1(\phi_1, \phi_2, \cdots \phi_n)$$
$$\dot{\phi}_2(z) = f_2(\phi_1, \phi_2, \cdots \phi_n)$$
$$\vdots$$
$$\dot{\phi}_n(z) = f_n(\phi_1, \phi_2, \cdots \phi_n)$$

*Notice that the steady state equations depend only on two device parameters (divide by k). The two parameters used may be written in several different forms. This matter is discused in chapter 5.

This type of system is autonomous since the independent variable does not appear explicitly. The solutions of such a system describe trajectories (with Z as the parameter) in an n-dimensional "phase" space. If n = 2, as in the case of the steady state equations, the space is referred to as the phase plane. Any given trajectory will exhibit one of five possible asymptotic behaviors as Z goes to infinity. The five possibilities are

1. the trajectory may approach infinity,
2. the trajectory may describe a closed cycle,
3. the trajectory may approach a closed cycle (by spiraling inward or outward),
4. the trajectory may remain motionless at a critical point for all Z, or
5. the trajectory may approach a critical point.

The critical points are defined to be the points in phase space where the dependent variables do not change. The location of the critical points may be obtained from the solution of the following system of equations

(Eq. 3.2.1-2)
$$\begin{aligned} 0 &= f_1(\phi_1,\phi_2, \cdots \phi_n) \\ 0 &= f_2(\phi_1,\phi_2, \cdots \phi_n) \\ &\vdots \\ 0 &= f_n(\phi_1,\phi_2, \cdots \phi_n) \end{aligned}$$

The global behavior of a system is determined to a great extent by the local behavior near the systems critical points. Trajectory behavior near a critical point must be

one of the four following types:

1. a center, trajectories are closed paths or cycles around the critical point;
2. a saddle point, trajectories approach the critical point and then go away from it;
3. a spiral point, trajectories spiral towards (stable) or away from (unstable) the critical point; or
4. a node, trajectories are asymptotic to straight lines through the critical point moving towards (stable) or away from (unstable) the critical point.

3.2.2 Analysis of Dominant Loss Saturation

This subsection presents critical point analysis of the steady state equations when loss saturation effects dominate. For this system there are three critical points, specifically: $\left(\frac{k}{g_0} a_{(hi)}, -a_{(hi)}\right)$ $(0,0)$ $\left(-\frac{k}{g_0} a_{(hi)}, a_{(hi)}\right)$

where $a_{(hi)} \equiv \left(\frac{g_0 \ell_0}{k^2} - 1\right)^{1/2}$. Existence of the two non-origin (hereafter referred to as "off-center") critical points within the real valued phase plane requires that $k^2 < g_0 \ell_0$. If this is not the case then there will only be one critical point, the origin.

Classification of the system behavior near the critical points is achieved by examining the approximate

solutions obtained from linearizing the equations in the vicinity of the critical points. Near the origin the behavior is approximated by

(Eq. 3.2.2-1) $$a2(x) \simeq A \exp(s_1 x) + B \exp(s_2 x)$$

$$a1(x) \simeq -\frac{(s_1 + \ell_0)}{k} A \exp(s_1 x) - \frac{(s_2 + \ell_0)}{k} B \exp(s_2 x)$$

where

$$s_{1,2} = -\frac{(\ell_0 - g_0)}{2} \pm \left[\frac{(\ell_0 - g_0)^2}{4} - k^2 + g_0 \ell_0\right]^{1/2}$$

and A and B are arbitrary constants.

Solutions of the form $a1(z) = A_1 \exp(s_a z) + B_1 \exp(s_b z)$, $a2(z) = A_2 \exp(s_a z) + B_2 \exp(s_b z)$, will produce phase plane trajectories (Z is the "parameter") that trace:

1. a <u>node</u>, if S_a and S_b are real with the same sign (if they are both negative the node is stable, if they are both positive the node is unstable);
2. a <u>saddle point</u>, if S_a and S_b are real with opposite sign; or
3. a <u>spiral point</u>, if S_a and S_b are complex conjugates (if the real parts of S_a and S_b are negative the spiral point is stable, if the real parts of S_a and S_b are negative the spiral point is unstable).

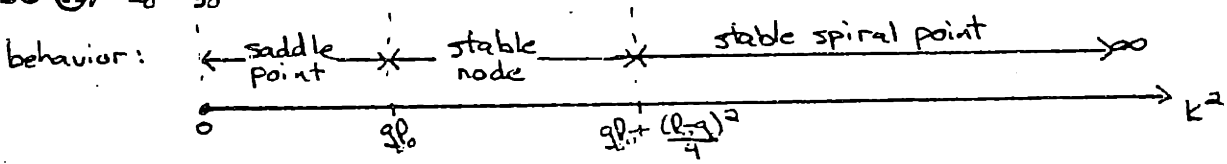
If S_a and S_b were imaginary then the trajectories would trace closed cycles around the critical point. For this to happen, however, the physical parameters of the system, l_o, g_o, and k, would have to take on specific values with exact precision. Due to fluctuations and uncertainties this is

physically impossible and mathematically must occur with zero probability. Although this precludes the existence of limit cycles around any particular critical point it does not prohibit the existence of global limit cycles (i.e., limit cycles encompassing more than one critical point).

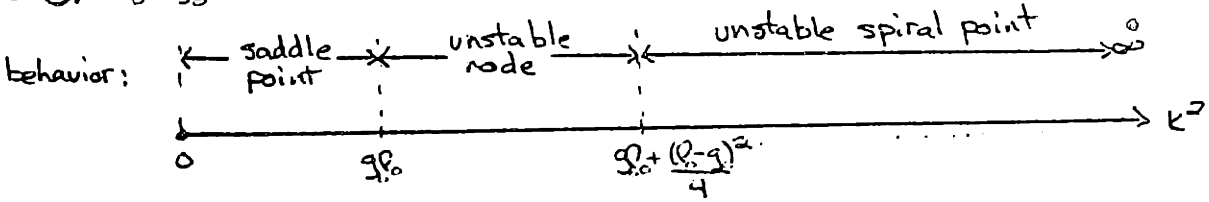
Applying the statements above to Eq. 3.2.2-1 yields stability criteria for the critical point at the origin in terms of the device parameters, l_o, g_o, and k. These criteria are presented in the following graphical format *

FIGURE 4 - Saturable Loss Central Critical Point Stability

Case Ⓐ, $l_o > g_o$



Case Ⓑ, $l_o < g_o$



The system behavior near either of the two off-center critical points is approximated by the linearized solution

(Eq. 3.2.2-2) $a2(x) \simeq A \exp(s_3 x) + B \exp(s_4 x) \pm a_{(hi)}$

$$a1(x) \simeq \left(\frac{k}{g_0}\left(1 - \frac{2k^2}{g_0 l_0}\right) - \frac{s_3}{k}\right) A \exp(s_3 x) + \left(\frac{k}{g_0}\left(1 - \frac{2k^2}{g_0 l_0}\right) - \frac{s_4}{k}\right) B \exp(s_4 x) - (\pm)\frac{k}{g_0} a_{(hi)}$$

*Notice that the stability of this critical point is dependent only upon two parameters, $k^2/l_0 g_0$ and g_0/l_0 (the latter dependence being simply whether greater than or less than one). This is true for all real critical points of all three structures.

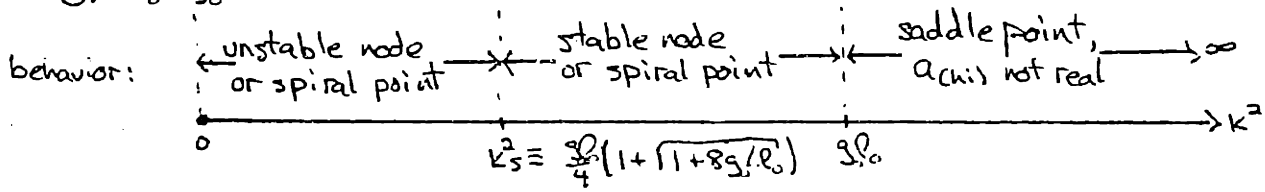
$$s_{3,4} = \frac{g_0}{2} + \frac{k^2}{2g_0^2\ell_0}(g_0\ell_0 - 2k^2)$$

where $$\pm\left[\left(\frac{g_0}{2} + \frac{k^2}{2g_0^2\ell_0}(g_0\ell_0 - 2k^2)\right)^2 - 2k^2\left(1 - \frac{k^2}{g_0\ell_0}\right)\right]^{1/2}$$

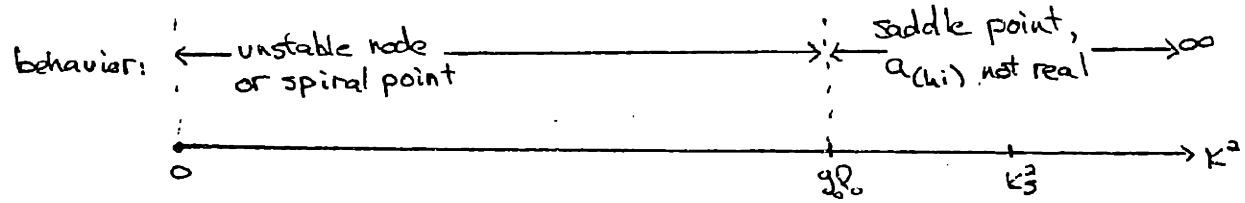
and A and B are again arbitrary constants. From these solutions the following stability criteria are obtained for the off-center critical points.

FIGURE 5 - Saturable Loss Off-center Critical Point Stability

Case Ⓐ, $\ell_o > g_o$



Case Ⓑ, $\ell_o < g_o$



Notice that the distinction between nodes and spiral points has not been specified. Actual closed form solutions for the values of k^2 at which the system makes transitions between nodal and spiral point behavior have been obtained. These formulae are so extremely complex, however, that they are of no practical value and are in fact too lengthy to be presented here. Useful formulae may be obtained for various limiting cases. At this point it is sufficient, however, to simply note that the existence (or non-existence) of spiral points within these regions is easily controlled. The numerical values of k^2 for which the behavior changes from node to spiral point (providing these values exist)

are easily obtained for any particular set of l_o and g_o.

Behavior of the off-center critical points when $k^2 > g_o\, l_o$, has not been described since these points are only of interest when they exist in the real valued mode amplitude phase plane.

3.2.3 Analysis of Dominant Gain Saturation

This subsection presents the critical point analysis of the steady state equations when gain saturation effects dominate. The same procedures used in the previous subsection to obtain stability criteria for the dominant loss saturation device may be repeated for the dominant gain saturation equations. The results, however, may be obtained in a different manner which reveals the symetries and similarities of the two devices. This approach also allows for a reduction in the number of possible types of behavior for which solutions must be computed.

Application of the transformation; $a1 \rightarrow a2$, $a2 \rightarrow a1$, $l_o \rightarrow g_o$, $g_o \rightarrow l_o$, to the dominant loss saturation equations results in

(Eq. 3.2.3-1)

$$\frac{d\,a1}{dx} = \frac{-g_0}{[1+(a1)^2]}\, a1 - k\, a2$$

$$\frac{d\,a2}{dx} = \ell_0\, a2 + k\, a1$$

These equations are the "space-reversed" version of the dominant gain saturation equations.

Therefore the phase plane trajectories of the dominant gain saturation equations for a specific parameterization: ℓ_o = constant 1, g_o = constant 2, k = constant 3, will be identical to the trajectories of the dominant loss saturation equations for the parameterization: ℓ_o = constant 2, g_o = constant 1, k = constant 3, except that the dependence on the parameter x has been reversed (i.e., the trajectories of one progress in a direction opposite of the other with increasing x). An analysis of the dominant loss saturation equations (in steady state) simultaneously analyzes the dominant gain saturation equations. Thus the critical point analysis of the dominant gain saturation equations has already been performed. The stability criteria may be obtained from those of the previous subsection by exchanging ℓ_o and g_o and performing space reversal. This is equivalent to the following changes:

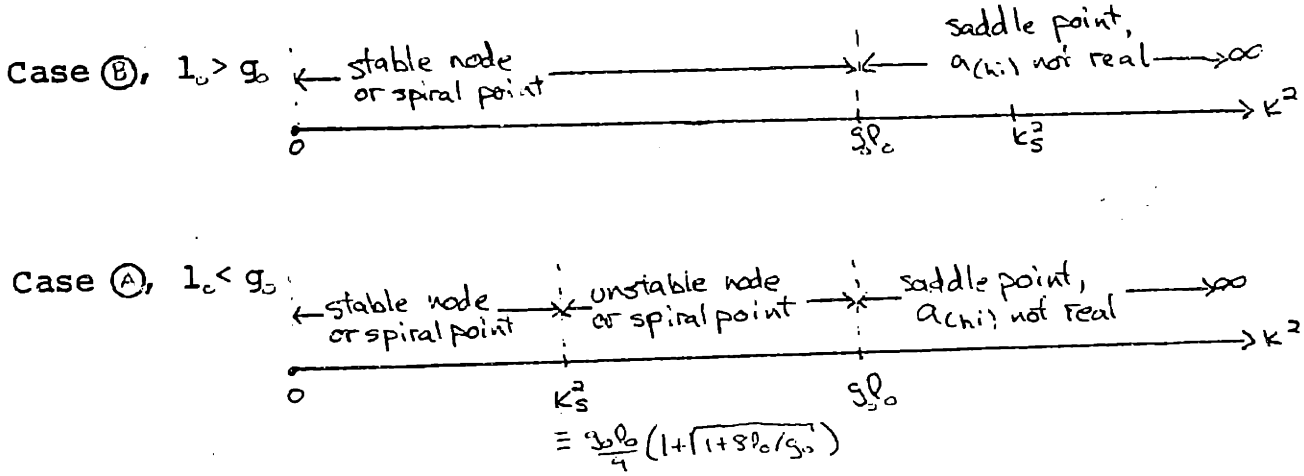
stability → instability instability → stability

$$k_s^2(\text{saturable loss}) \equiv \frac{g_0\ell_0}{4}(1 + \sqrt{1 + 8g_0/\ell_0})$$

$$\rightarrow k_s^2(\text{saturable gain}) \equiv g_0\ell_0(1 + \sqrt{1 + 8g_0/\ell_0})/4$$

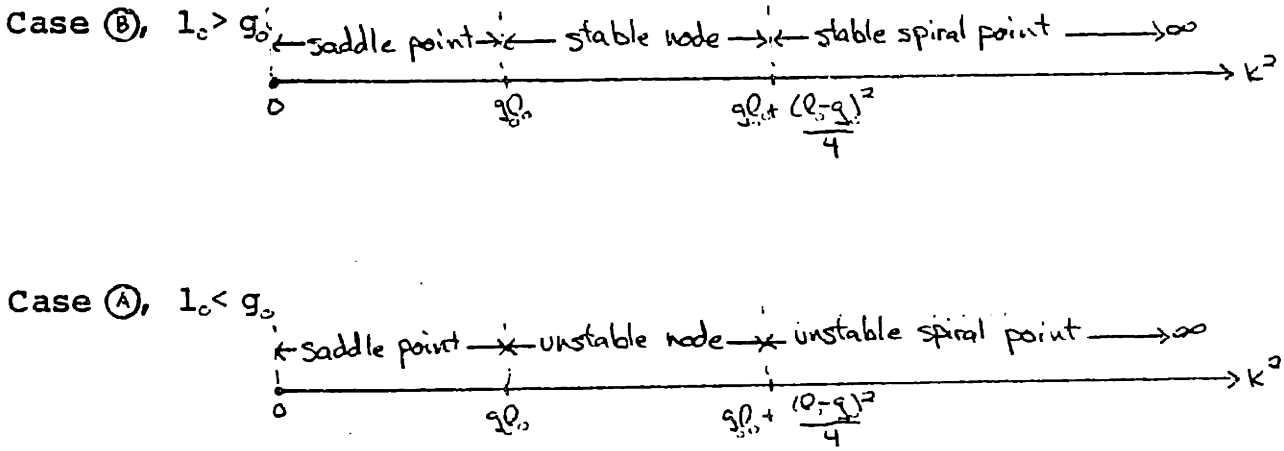
The resultant stability criteria for the off-center critical points of the dominant gain saturation device are as follows.

FIGURE 6 - Saturable Gain Off-center Critical Point Stability



The stability criteria for the critical point at the origin are

FIGURE 7 - Saturable Gain Central Critical Point Stability



Note that this is exactly the same result as was obtained for the central critical point in the previous subsection. This follows from the fact that near the origin, where the mode amplitudes are small, both devices are not experiencing saturation effects and will therefore exhibit the same unperturbed behavior.

3.2.4 Analysis of Dual Saturation

This subsection deals with the critical point analysis of the steady state equations when both loss and gain saturation effects are considered. There are five critical points for this system, two of which are never real valued and will therefore not be analyzed. The three possible real critical points are :

$$(-a1_{(hi)}, a2_{(hi)}); \qquad (0,0); \qquad (a1_{(hi)}, -a2_{(hi)}).$$

where $$a1_{(hi)} \equiv \left[\left(\frac{(\ell_0 - g_0)\, g_0}{2k^2} - 1 \right) + \left[\left(\frac{(\ell_0 - g_0) g_0}{2k^2} - 1 \right)^2 + \frac{g_0 \ell_0}{k^2} - 1 \right]^{1/2} \right]^{1/2}$$

and $$a2_{(hi)} \equiv \left[\left(\frac{(g_0 - \ell_0)\, \ell_0}{2k^2} - 1 \right) + \left[\left(\frac{(g_0 - \ell_0) \ell_0}{2k^2} - 1 \right)^2 + \frac{g_0 \ell_0}{k^2} - 1 \right]^{1/2} \right]^{1/2}$$

The stability criteria for the critical point at the origin may be shown to be the same as those for the two previous devices described. Again this is to be expected since all three devices have the same unperturbed response near the origin.

The off-center critical points again exist in the real phase plane only when $k^2 < g_o \ell_o$. Linearizing the equations near either of the off-center critical points results in

(Eq. 3.2.4-1) $$\ddot{\varepsilon}2 + b\, \dot{\varepsilon}2 + c\, \varepsilon 2 = 0$$

$$\varepsilon 2 = \varepsilon 2(x) \equiv a2(x) \pm a2_{(hi)}$$

where $$b \equiv \frac{\ell_0(1-a2^2_{(hi)})}{(1+a2^2_{(hi)})^2} - \frac{g_0(1-a1^2_{(hi)})}{(1+a1^2_{(hi)})^2}$$, and

$$c \equiv k^2 - \left[\frac{\ell_0(1-a2^2_{(hi)})}{(1+a2^2_{(hi)})^2}\right]\left[\frac{g_0(1-a1^2_{(hi)})}{(1+a1^2_{(hi)})^2}\right]$$

The algebraic complexity of b and c on 1_o, g_o and k^2 renders closed form solutions of the stability transition points, k_s^2 , etc., useless. Limiting cases may be examined and it may be shown that all types of critical point behavior are possible. Stability criteria may be obtained for specific values and 1_o and g_o by plotting the k^2 dependence of b and c.

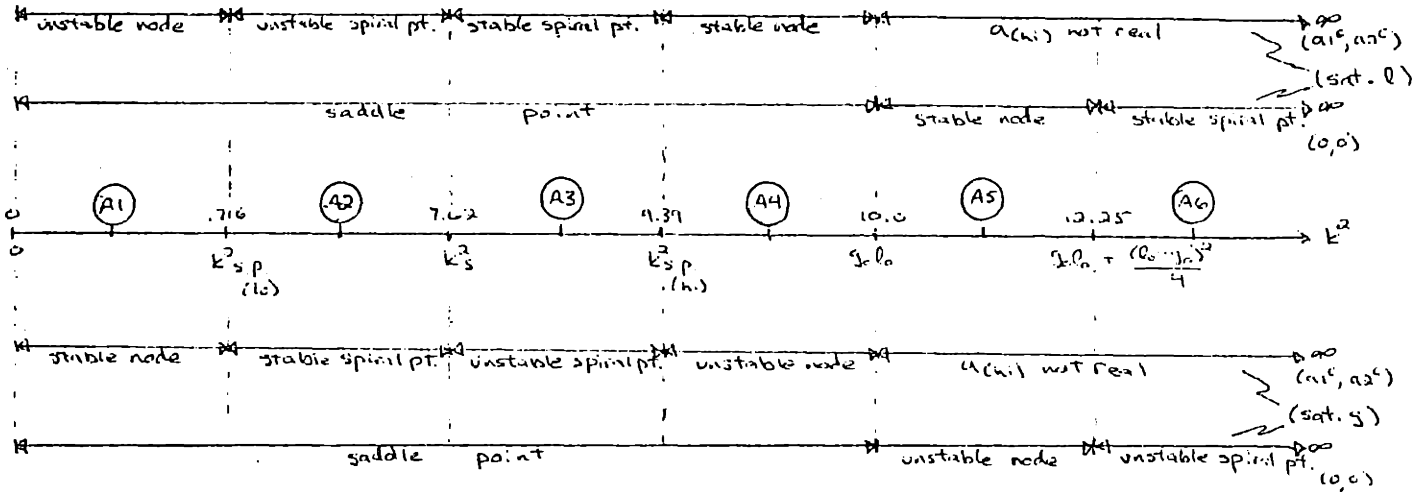
3.3 Parameterizations and Cases of Interest

Numerical values of 1_o, g_o, and k which demonstrate all possible types of steady state behavior are now chosen. Steady state solutions will be generated for these examplimary parameterizations.

The dominant loss and dominant gain saturation devices are examined simultaneously. For case A, i.e. when the unperturbed value of the saturating parameter is larger than the value of the linear parameter, the numerical values chosen for the saturable loss device are, 1_o=5, g_o=2, which corresponds to 1_o=2, g_o=5, for the saturable gain dvice. The resultant stability criteria for case A are:

Portions of the text on the following page(s) are not legible in the original.

FIGURE 8 - Critical Point Stability for Case A ; Saturable Loss ($l_o = 5$, $g_o = 2$) and Saturable Gain ($g_o = 5$, $l_o = 2$)



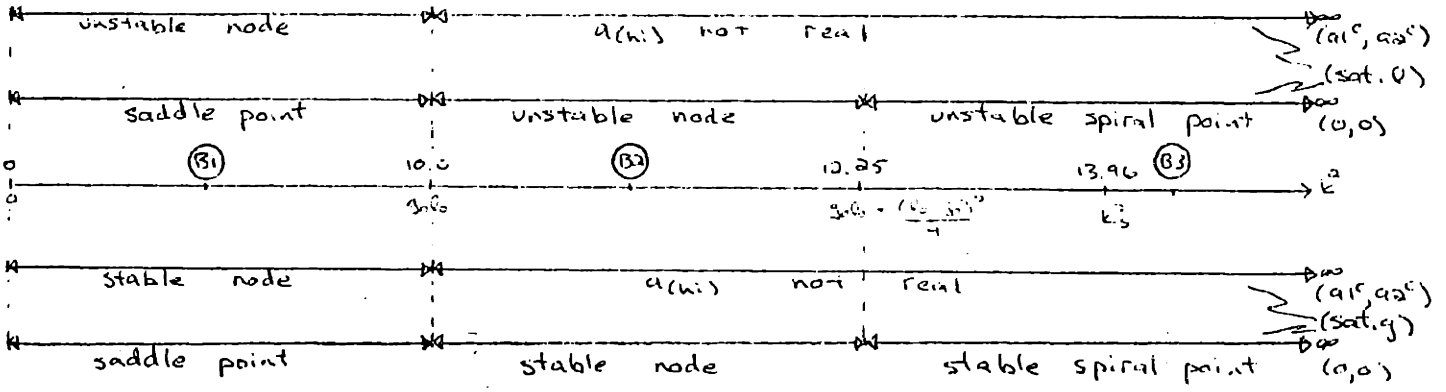
Six values of k^2 corresponding to the six different stability regions are chosen and labeled (A1) through (A6). These numerical values are:

$k^2(A1) = .3583$ $\quad k^2(A2) = 4.1683$ $\quad k^2(A3) = 8.505$

$k^2(A4) = 9.695$ $\quad k^2(A5) = 11.125$ $\quad k^2(A6) = 15.0$

To examine Case B, the parameterization $l_o = 2$, $g_o = 5$, is chosen for the saturable loss device corresponding to the choice, $l_o = 5$, $g_o = 2$, for the saturable gain device. The resulting stability criteria are:

FIGURE 9 - Critical Point Stability for Case B; Saturable Loss ($l_o = 2$, $g_o = 5$) and Saturable Gain ($g_o = 2$, $l_o = 5$)



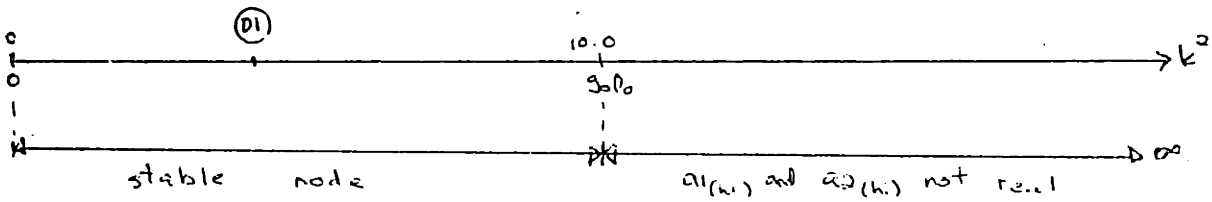
The values of k^2 corresponding to the three different stability regions are chosen and labeled (B1) through (B3). These numerical values are:

k^2 (B1)=5.0 k^2 (B2)=11.125 k^2 (B3)=15.0

Notice that the transition point k^2_s is irrelevent for Case B since $a_{(hi)}$ is not real for $k^2 > g_o l_o$.

For the dual saturation device we arbitrarily select the parameterization, l_o=5, g_o=2, and plot the k^2 dependence of b, c, and $\frac{b^2}{4}$ - c. From these plots the following off-center critical point stability predictions are obtained.

FIGURE 10 - Off-center Critical Point Stability for Dual Saturation (ℓ_0=5,g_0=2)



There is only one case of interest for this parameterization, labeled D1, where k^2(D1)=5. Other types of behavior could be allowed for different parameterizations. For instance, if l_o and g_o are more equal, then stable and unstable nodes and spiral points are allowed at the off-center critical points. When l_o is substantially greater than g_o then only stable off-center spiral points are achieved. When g_o is

substantially greater than l_o then only unstable off-center spiral points are allowed. The case, D1, however, will be sufficient to demonstrate the differences in dual and dominant steady state saturation.

3.4 Functional Approximate and Numerical Solutions in the Phase Plane

3.4.1 Initial Value Problems and Spontaneous Singularities

Whether or not a given trajectory will enter a neighborhood close enough to a critical point that its behavior is accurately described by the aforementioned linearized solution is, of course, dependent on the initial conditions that uniquely define that trajectory. The singularities of a nonlinear system may not be derived from inspection of the coefficients of the differential equation (as in the case of a linear system). Furthermore, these singularities are dependent on the initial conditions (thus the term "spontaneous" singularities because for a given parameterization some trajectories may be stable, (i.e.,

decay into a critical point) while others are singular (i.e., run off to infinity).

Obviously some type of global analysis is required. We could simply pick various initial conditions and numerically solve the steady state equations but it would be difficult to know how to insure that we have exhausted all types of behavior. A more useful approach is to use functional approximations of the solutions at various grid points over large regions of the phase plane.

3.4.2 Tangential Field Analysis and Taylor Series Approximations of Multiple Valued Functions

Formulae describing the slope of a trajectory at any point in the phase plane may be obtained by writing the steady state equations in terms of $\frac{da2}{da1}$. In the dominant loss saturation case, for example, the trajectory slope formula would be

$$\frac{da2}{da1} = \frac{-\left(\frac{\ell_0 a2}{(1+a2^2)} + k\ a1\right)}{g_0 a1 + k a2}$$

For global indications of stability small lines could be plotted at various points in the phase plane, each having a slope equal to the tangent of the trajectory at that point. The eye could then interpolate the general behavior of a trajectory for any initial condition from this plot of the "tangential field". A more elaborate and more useful extension of this approach is to expand the trajectory around each selected point in a Taylor series approximation. The

terms are easily calculated* by taking successive derivatives of the trajectory slope formula, but one must be careful in using Taylor series expansions of multiple valued "functions".

The current form of the algorithm used will now be explained. Grid points around which expansions are calculated are evenly spaced and determined from the given maximum and minimum values of a1 and a2 and the given number of steps between them. For each point the program decides whether to expand a1 as a function of a2 or a2 as a function of a1. (If the absolute value of the slope is less than one then a2 is expanded as a function of a1.) A reasonable number of terms is calculated by expanding until the last term adds a negligable change in the plot when the independent variable runs to the next grid point. (The number of terms is limited to a maximum value which is usually set equal to seven.) Then the limit of the double ended (plus or minus) increment of the independent variable is selected (with the number of terms set to its "reasonable" value) so that the error in the expansion is just on the verge of being noticeable as seen from the plot (which is usually set to - 5mm). To calculate this error exactly one would have to evaluate the substitution of

*The terms are calculated and simplified (if possible) by the symbolic manipulation capabilities of the MACSYMA software.

the series expansion for the dependent variable in the formula for the next derivative and find the maximum of this derivative as the independent variable varies over some arbitrarily chosen maximum range (in the present implementation the next grid point is this maximum range). When this is done the resulting expressions in terms of the independent variable become extremely complicated and the computer rapidly runs out of memory space (and calculation times become ridiculously long). Since, however, these higher derivatives are small we may assume that the next derivative does not change very much over our selected maximum range and approximate its maximum by selecting the maximum of this derivative (with, of course, the expansion substituted for the dependent variable) evaluated at the grid point and the two end points of the independent variables' maximum range. The range of the independent variable is then restricted to that which produces the allowable error unless that range exceeds the chosen ("to the next grid point") maximum range. There is, however, one more possible failure of this type of expansion. The actual solution may represent a double or multiple valued function over the determined independent variable's range. In other words, the true solution may "double back" on itself while the expansion cannot do this and the error formula does not take this into account. For this to occur in an expansion of a2

as a function of a1 the slope must sometime become infinite, i.e., the expansion must intersect the infinite slope locus (e.g. for dominant loss saturation $g_0 a1 + k\, a2 = 0$ For this to occur in an expansion of a1 as a function of a2, the slope must sometime become zero, i.e., the expansion must intersect the zero slope locus (e.g. for dominant loss saturation $\frac{\ell_0 a2}{(1+a2^2)} + k\, a1 = 0$. Since these "double-valued errors" rarely occur, the algorithm is to test each expansion for "double-valued error candidacy" (i.e., if the slope at the end points has changed sign from the original or if the absolute value of the slope at the end points has changed from being less than one to greater than one) and only then is the expansion tested for intersection with the infinite or zero slope locus. If an intersection exists (within the length of the expansion) then the proper end (or possibly both ends) is shortened to ensure a valid functional representation. The program works well, using little time and memory while conveying a lot of information. The program can be trivially modified to approximate phase plane solutions of any second order nonlinear autonomous differential equation. Obvious extensions to this method include both the use of adaptive grid point selection to produce continuous approximations and the use of more than one pair of orthogonal axes for more efficient functional expansions.

3.4.3 Numerical Analysis

Once the global behavior of a system has been examined via the aforementioned "trajectory field" plots, continuous trajectories may be obtained from numerical integration of the steady state equations for the initial conditions of interest. The numerical solutions should provide the final verification of the validity of the critical point stability predictions and the trajectory field approximations.

The algorithm employed utilizes a straightforward "first order backward difference" formula to integrate, with respect to x, each of the two equations of a steady state system. When the distance in the phase plane between the last plotted point and the running sums becomes larger than a specified value then the values of the sums are stored as a point to be plotted and the summation resumes. Thus, information about the x dependence of the mode amplitudes is lost. If, however, the x dependence is desired it is a simple matter to retain the "number of steps taken" information.

3.5 Steady State Results and Interpretations

3.5.1 General Remarks

This section presents results for the steady state limit under the ten parameterizations of interest as originally enumerated in section 3.3. For each case of interest three plots will be presented and physical interpretations will be offered. Two of these three plots will be the discontinuous trajectory field plots generated from Taylor series approximations. These two will cover two different regions of the phase plane at two different levels of globality. One will present information for all four quadrants for amplitudes comparable to and somewhat larger than the off-center critical point amplitudes. The other will focus on the behavior near the critical points. The third plot for each case of interest will be the set of continuous phase plane trajectories that were numerically generated from the set of initial conditions as selected from and indicated on the first two plots of that case of interest. All three plots are presented on a single page along with a legend containing pertinent information.

At this point the author digresses somewhat to explain the large amplitude ($A >> \sqrt{PS}$ or $a >> 1$) response. Discontinuous trajectory field plots covering large amplitude regions of the phase plane have been

generated but will not be presented here since simple physical arguments render these results blatantly obvious. For example, when considering the saturable loss device, large amplitudes will saturate the lossy waveguide while the gain waveguide remains linear. Therefore, if the waveguides were uncoupled (i.e. $k^2 = 0$) then a2 would become constant while a1 experiences exponential gain. For increased coupling the net effect would still be that the system energy would increase without bound. Thus for the saturable loss device large amplitudes "blow up", i.e. the trajectories approach infinity. Obviously infinite energy is unobtainable in any actual physical system. What really happens is that the system energy increases until our model is no longer valid, i.e. other mechanisms that were neglected begin to dominate the behavior. Some of these various mechanisms are: the gain saturation effects due to non-infinite $PS_{(G)}$, non-linear effects such as self-induced transparency, thermal effects, and others. Similarly for the saturable gain device, large amplitudes will saturate the gain medium while the lossy waveguide remains linear. Therefore, if the waveguides were uncoupled then a1 would become constant while a2 experiences exponential loss. For increased coupling the net effect would still be a reduction in the system energy. Thus for saturable gain, large

amplitudes attenuate, i.e. the trajectories move towards the smaller amplitude regions of the phase plane. The model violation mechanisms for large amplitudes are the same except that for saturable gain it is the non-infinite $PS_{(L)}$ that was neglected. Practical devices, however, shall always be designed to operate with system energies much less than those experienced for the large amplitude region, therefore, this region will not be considered further.

Results for the ten parameterizations of interest shall be presented sequentially. Correlation between results of the two computational methods and the predicted critical point behavior shall be examined and physical interpretations will be discussed for each case. Critical points are indicated by encircled stars and the directions of the trajectories are indicated by arrowheads. For case A the trajectory directions are indicated for the saturable loss device while the directions for case B are indicated for the saturable gain device. The initial conditions used for the numerical solutions are indicated on the trajectory field plots by stars with an arrow extending in the direction of anticipated initial progression. Also indicated on the first trajectory field plot is a rectangle indicating the phase plane region that will be covered in the second trajectory field plot.

A few comments on the accuracy and extrapolation techniques of the trajectory field plots will now be made. Due to the nature of the Taylor series approximation, as used in this algorithm, the errors at the endpoints of an expansion (i.e. the deviations between the actual solution and the plotted expansion) are kept within a certain tolerance. For the plots presented here that tolerance was specified to be about 5 mm*. Although the endpoints are kept within the specified tolerance the derivatives at the endpoints may be quite different than those of the actual solution so that it is more accurate to extrapolate from adjacent expansions rather than to extrapolate by continuing with the same behavior (i.e. using derivative information) from the endpoint of an expansion. Occasionally an expansion with a double valued error will escape detection from the double valued error candidacy tests resulting in an erroneous expansion. This happens so rarely, however, that it is not worthwhile (for our purposes) to explore more elaborate candidacy tests or to waste computer time testing for double valued errors directly.

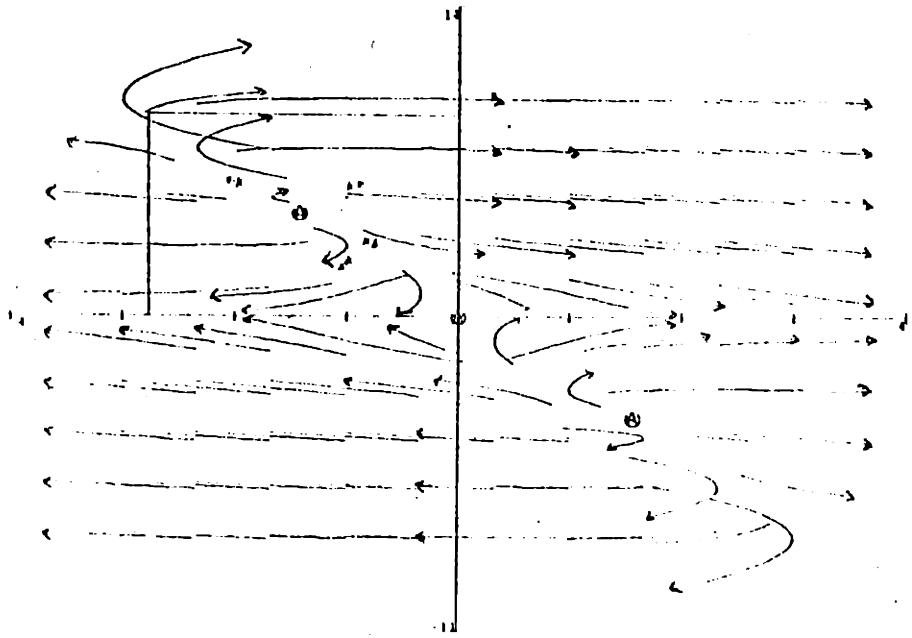
3.5.2 Results and Interpretations

The first case of interest is labeled A1 and represents the saturable loss device for $l_0=5$, $q_0=2$,

* Before a photo-reduction of about 40%.

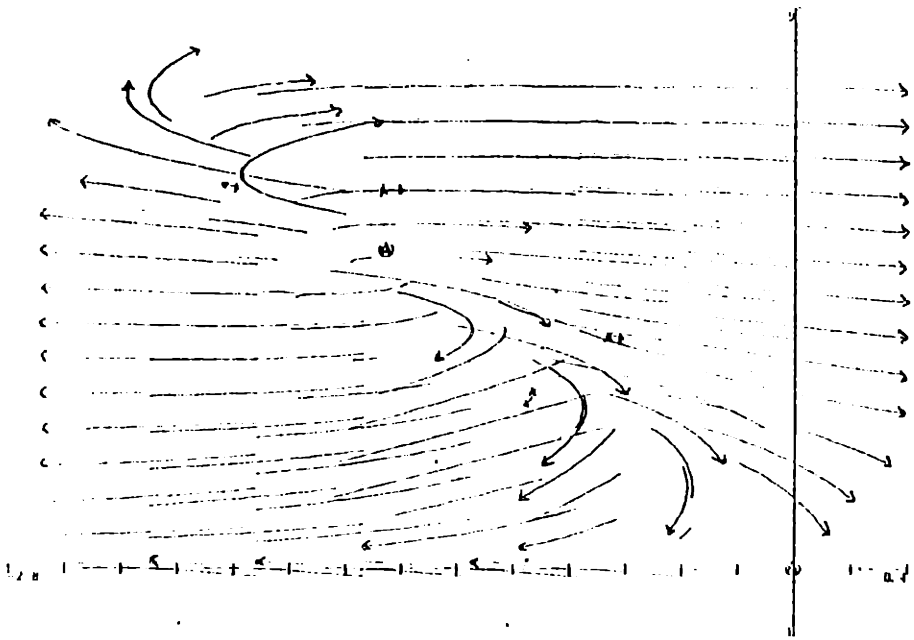
k^2=.3583, which is predicted to have an unstable node at the off-center critical point and a saddle point at the origin. This case also represents (with directions reversed for increasing x) the saturable gain device for l_0=2, g_0=5, k^2=.3583, which is predicted to have a stable node at the off-center critical point and a saddle point at the origin. The computed results are presented in PL. 3.5.2-1, which correlate well within the predicted behavior, i.e. the origin is a saddle point for both devices and the off-center critical point is stable for saturable gain and unstable for saturable loss. Note that although the rapid variation of the trajectories close to the off-center critical point yield some inaccuracy it is possible to discern that we do in fact have a node rather than a spiral point. This may be seen by noting that a trajectory starting below the critical point will not end up above it and similarly for those starting above, to the right, or to the left.* These results are verified by the numerical results which show four continuous trajectories. A physical interpretation

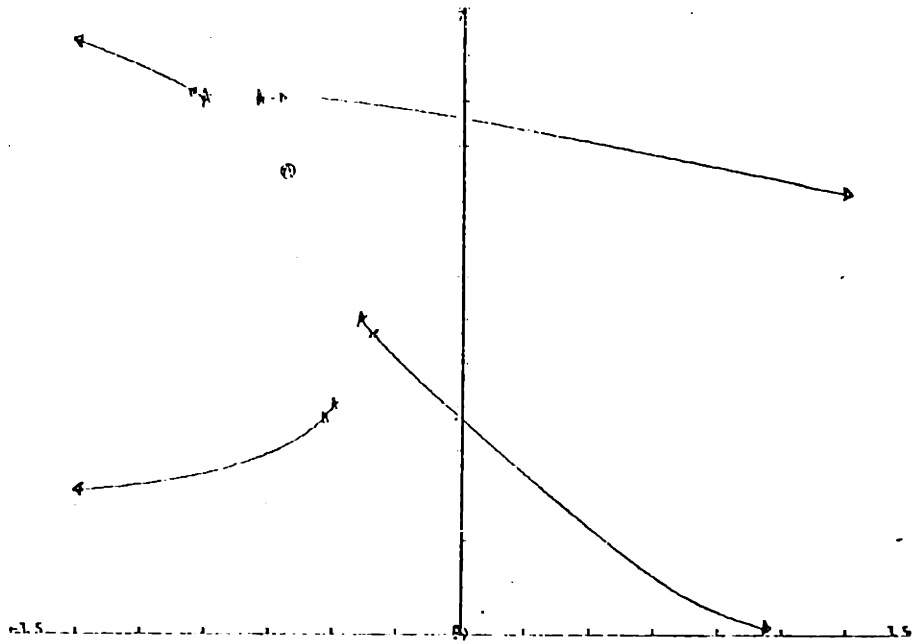
*The distinction between nodes and spiral points is a slight one, however, and depends on both the location of k^2 in that stability region and the initial conditions. The distinction is considered here, however, since in Chapter IV it will be shown that this distinction leads to two different device applications.



PL.3.5.2-1 Case A1
x-coordinate: a1 (saturable loss)
a2 (saturable gain)
y-coordinate: a2 (saturable loss)
a1 (saturable gain)

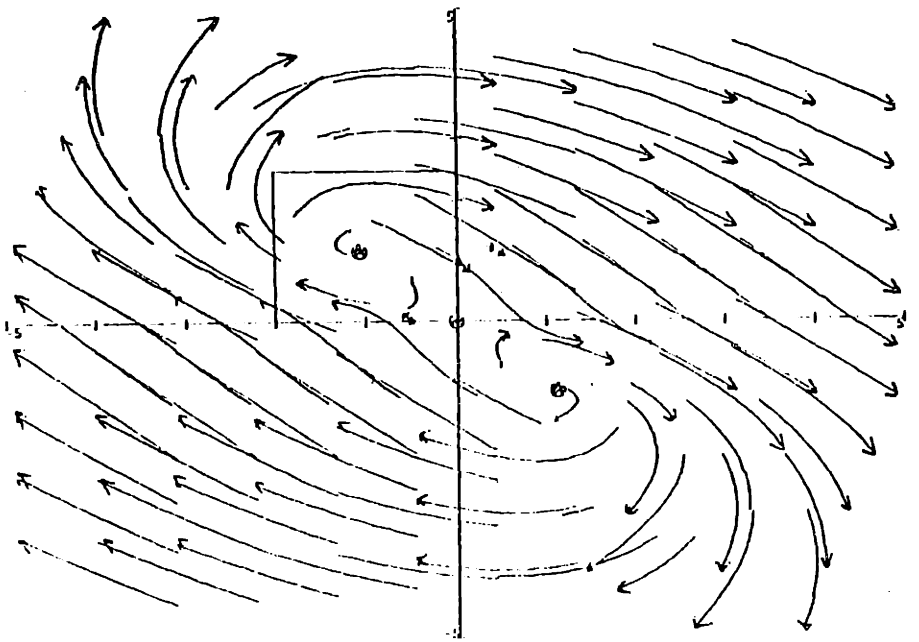
arrows for saturable loss





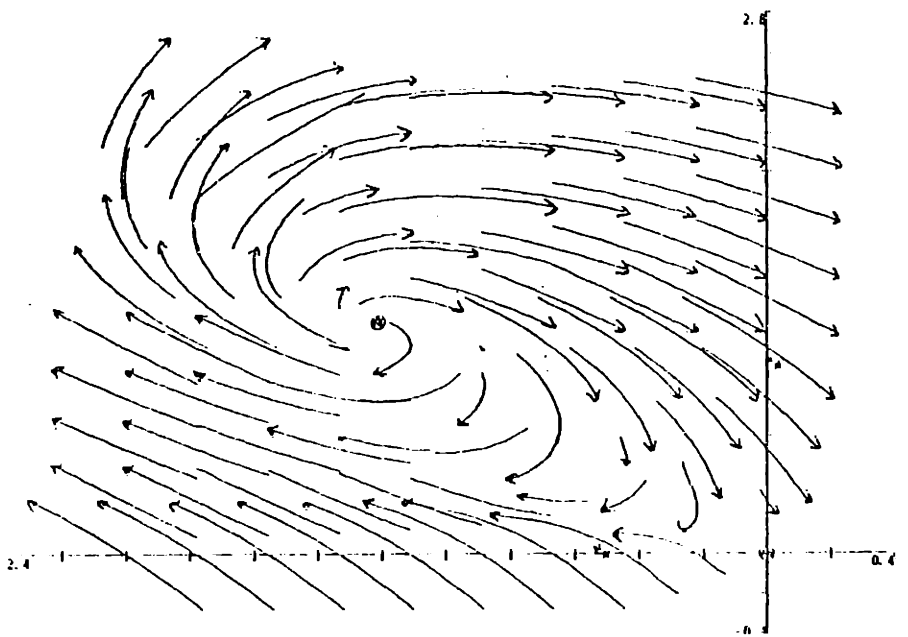
of these results is as follows. For the saturable loss device when $l_0 > g_0$, weak coupling (small k^2) means that the electromaganetic energy will be amplified in the gain waveguide and the small amount of energy that is coupled into the other waveguide will start to saturate the loss so that the net effect is that the amplitude a1 will increase without bound while amplitude a2 will approach a constant value. For the saturable gain device when $l_0 < g_0$, weak coupling is no problem for stability since the gain will saturate until loss and gain effects balance each other and an equilibrium amplitude is achieved at the off-center critical point.

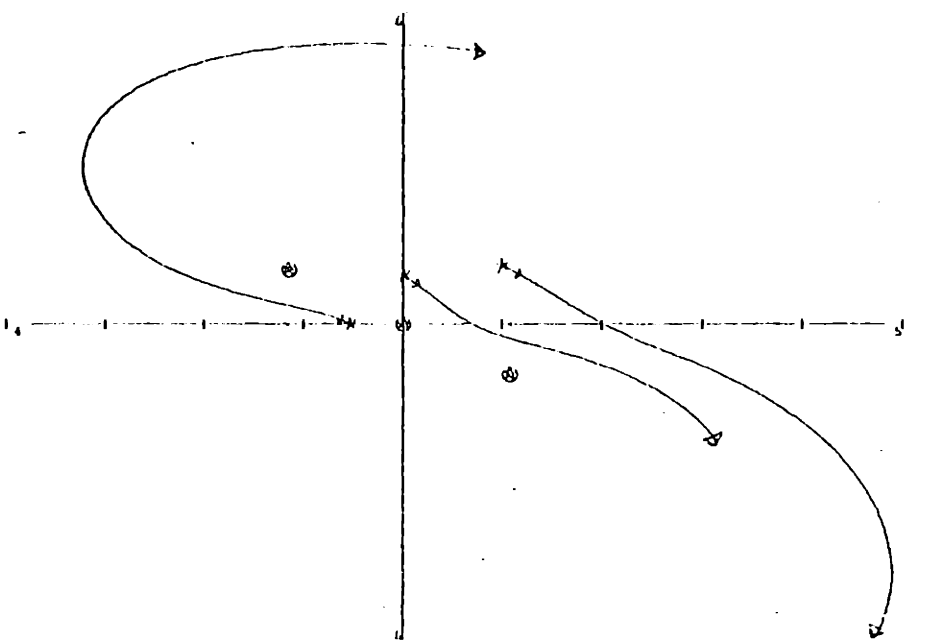
For the next case, A2, we have increased the coupling constant slightly (k^2=4.1683) so that the predicted behavior of the off-center critical point is now a spiral point which should be unstable for saturable loss and stable for saturable gain. PL.3.5.2-2 demonstrates this. Note that for a spiral point we have the possibility of circling around the critical point before running off to infinity (saturable loss) or converging into the critical point (saturable gain). A physical interpretation is as follows. We have increased the coupling constant by an amount such that the mode amplitudes are more interdependent (thus the "curvature"



PL.3.5.2-2 Case A2
x-coordinate: a1 (saturable loss)
a2 (saturable gain)
y-coordinate: a2 (saturable loss)
a1 (saturable gain)

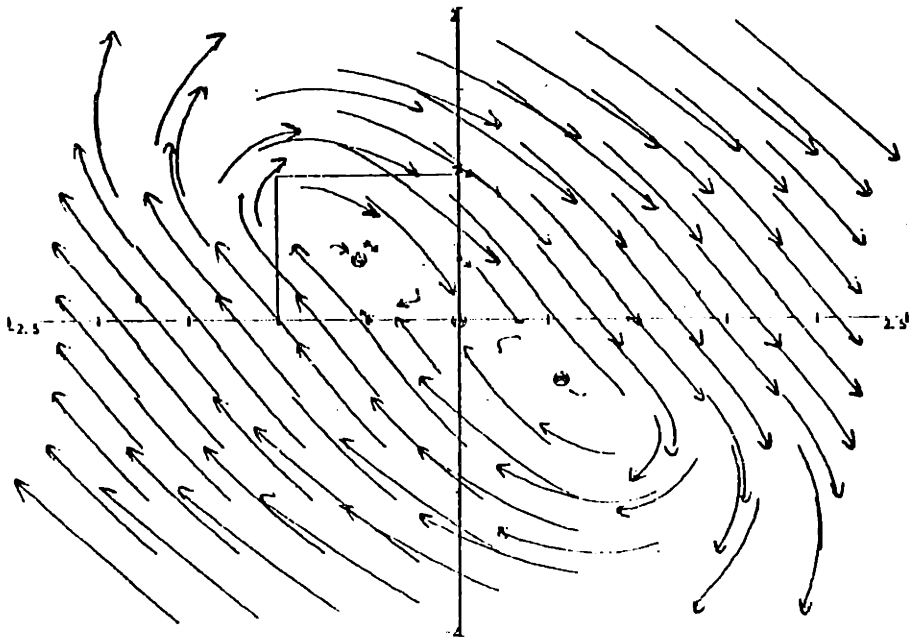
arrows for saturable loss





of the phase plane trajectories has increased) but we have not increased it enough to change the basic stability (i.e. saturable loss response still blows up and saturable gain response still converges to the off-center critical point). This excellent agreement between theory and results is further verified by the three continuous trajectories generated by numerical analysis.

For the next case, A3, the coupling constant has increased (k^2=8.505) so that we predict the saturable loss device to have a stable spiral point at the off-center critical point and the saturable gain device to have an unstable spiral point. Note the critical point at the origin is not yet predicted to have changed from its saddle point behavior. PL.3.5.2-3 shows that for initial conditions much larger than the off-center critical point values the saturable loss trajectories may still run off to infinity but for smaller initial conditions the trajectories will spiral in towards the critical point. This convergence (or divergence for saturable gain) is detailed by the second Taylor field plot and numerically verified for four different initial conditions. Again there is excellent correlation between analytical predictions and the results of the two computational methods. Physically this continuation of increased "trajectory folding over" with increased coupling may be interpreted

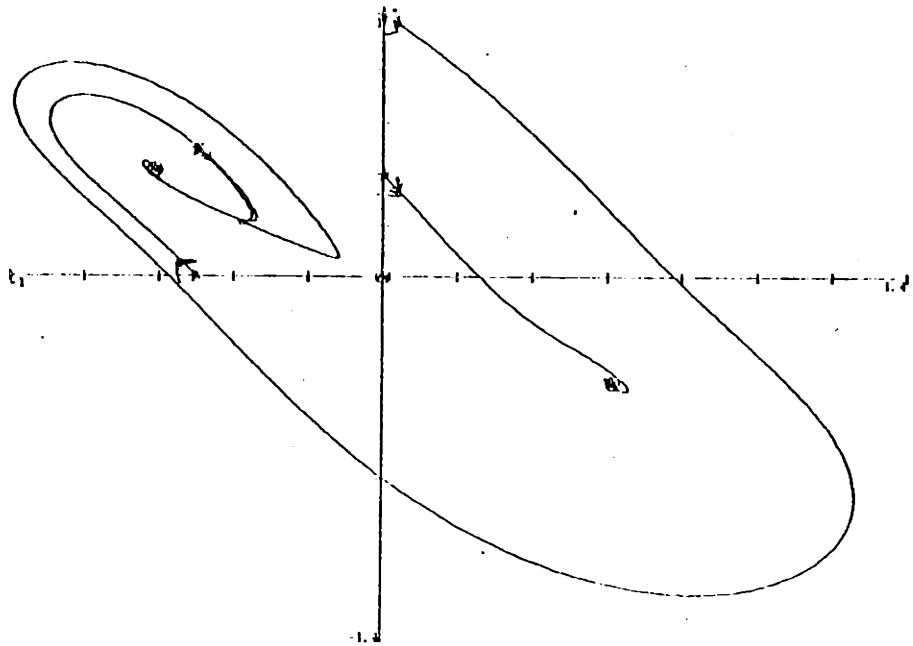


PL.3.5.2-3 Case A3

x-coordinate: a1 (saturable loss)
a2 (saturable gain)

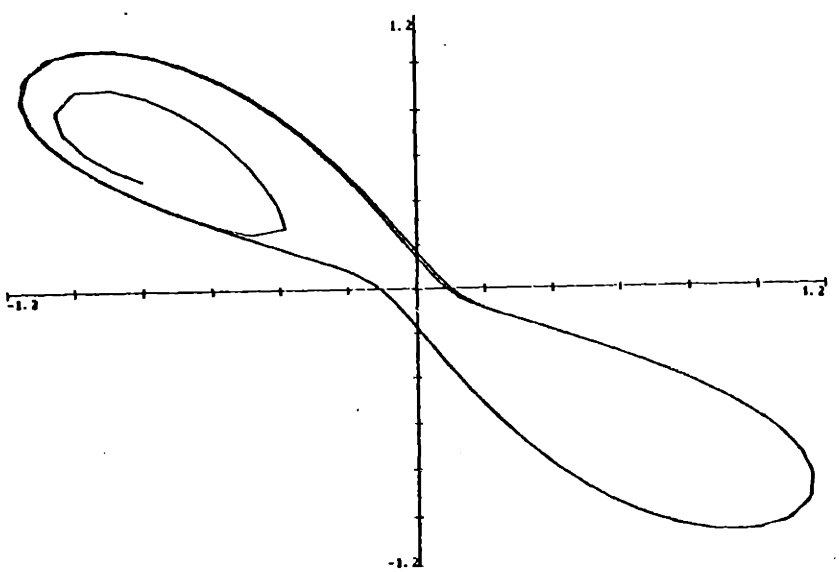
y-coordinate: a2 (saturable loss)
a1 (saturable gain)

arrows for saturable loss



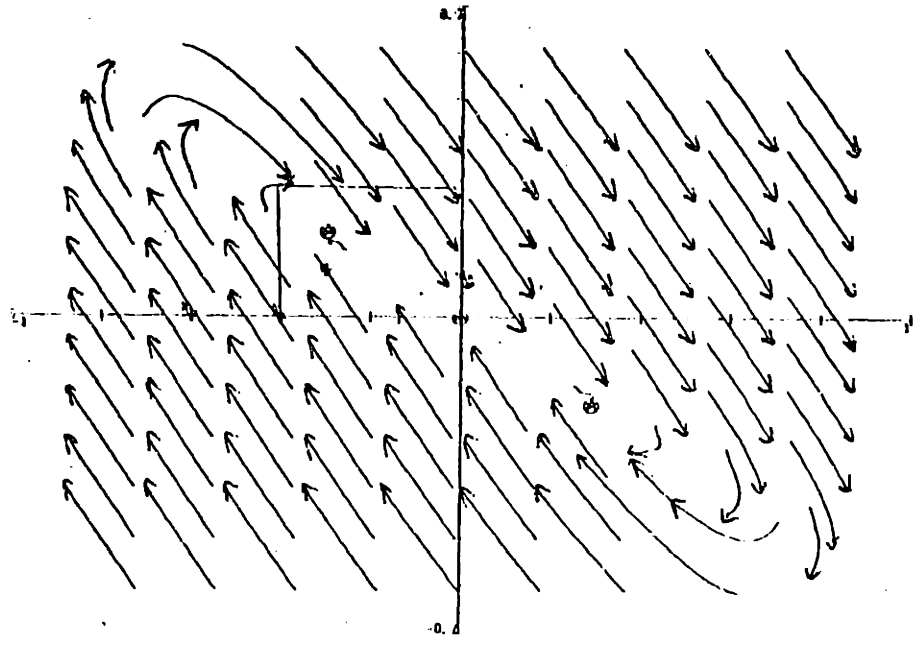
as follows. For the saturable loss device the coupling has increased enough that the energy in the gain waveguide (if it is small enough initially) will not just increase without bound since a large amount is coupled into the loss waveguide. Although the loss is greater than the gain, the loss begins to saturate until loss and gain effects balance each other and equilibrium amplitudes are achieved at the off-center critical points. For the saturable gain device this increased coupling means that enough energy from the lossy waveguide is coupled into the gain waveguide that the amplitudes will begin to increase and gain saturation effects are not significant enough to cause convergence near the critical point. Gain saturation will cause attenuation for large amplitudes however, and this opposition to the small amplitude blow-up results in global limit cycles. An example of these global limit cycles is presented in Figure 11.

FIGURE 11 - A Global Limit Cycle



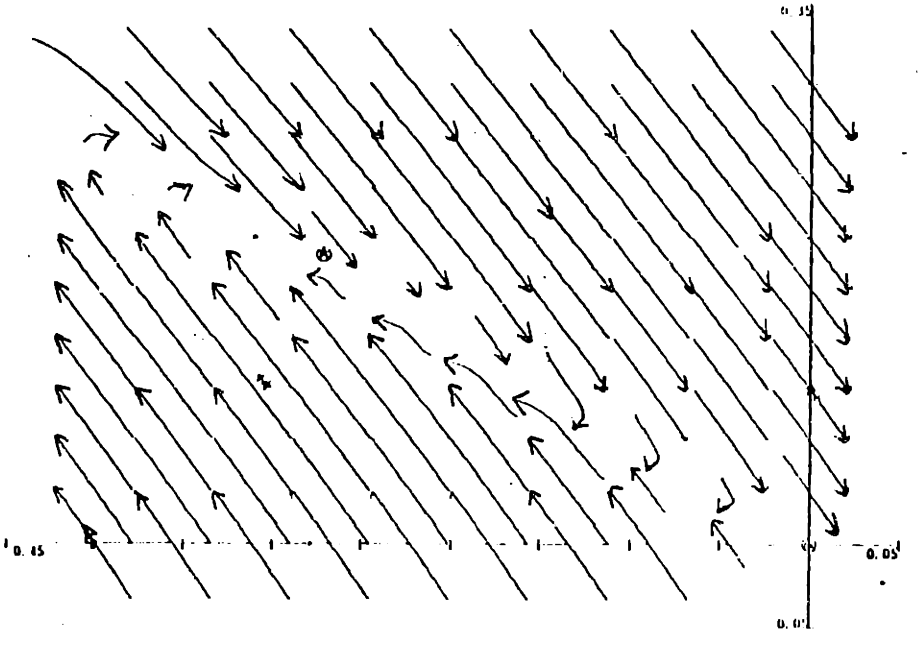
Increasing the coupling constant further (k^2=9.695) brings us to our next case, A4, which is predicted to have an off-center stable node for the saturable loss device and an unstable node for saturable gain. PL.3.5.2-4 shows that the trajectories have been "folded-over" even more, so that there are trajectories with extremely large curvature in the regions where this increased interaction is significant (i.e., near the critical points). Although it is in this region that both the series truncation and double-valued errors are larger, it is still clear from the plots that critical point convergence (or divergence, for saturable gain) is asymptotic to straight lines, i.e. this is a node rather than a spiral point. The set of numerically generated continuous trajectories for four initial conditions correlates well with the predicted behavior and the trajectory-field plots. A physical interpretation is that we have not changed the basic stability (from A3) yet, but we have increased the interaction enough so that the curvature of the trajectories is more abrupt, thus the final critical point convergence or divergence must be asymptotic to straight lines.

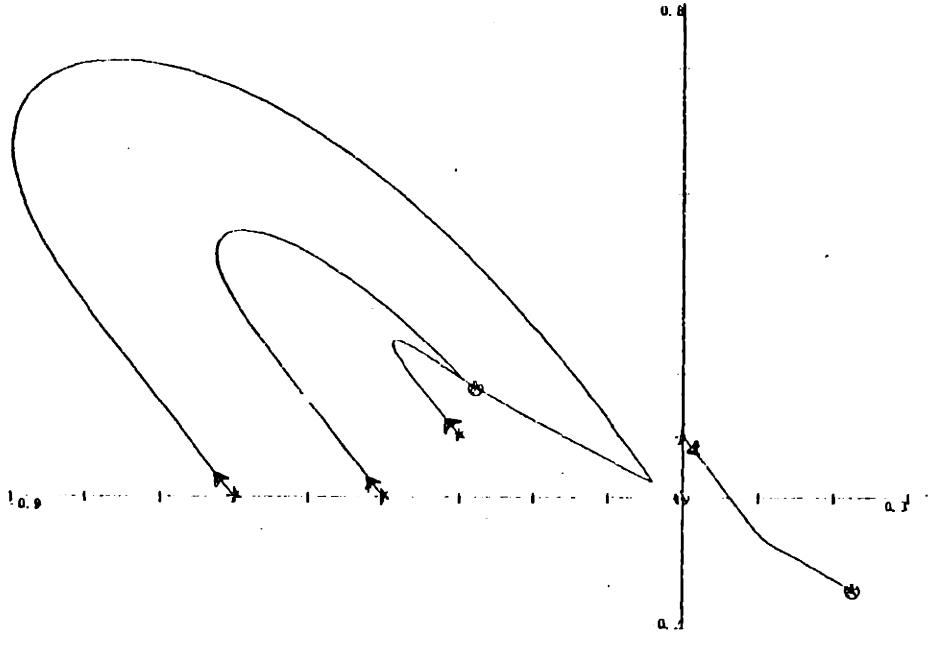
For the next case, A5, the coupling constant has increased enough (k^2=11.125) to change the predicted central critical point behavior to a stable node, while $a_{(hi)}$ has become imaginary so that we now have only one real



PL.3.5.2-4 Case A4
x-coordinate: a1 (saturable loss)
a2 (saturable gain)
y-coordinate: a2 (saturable loss)
a1 (saturable gain)

arrows for saturable loss

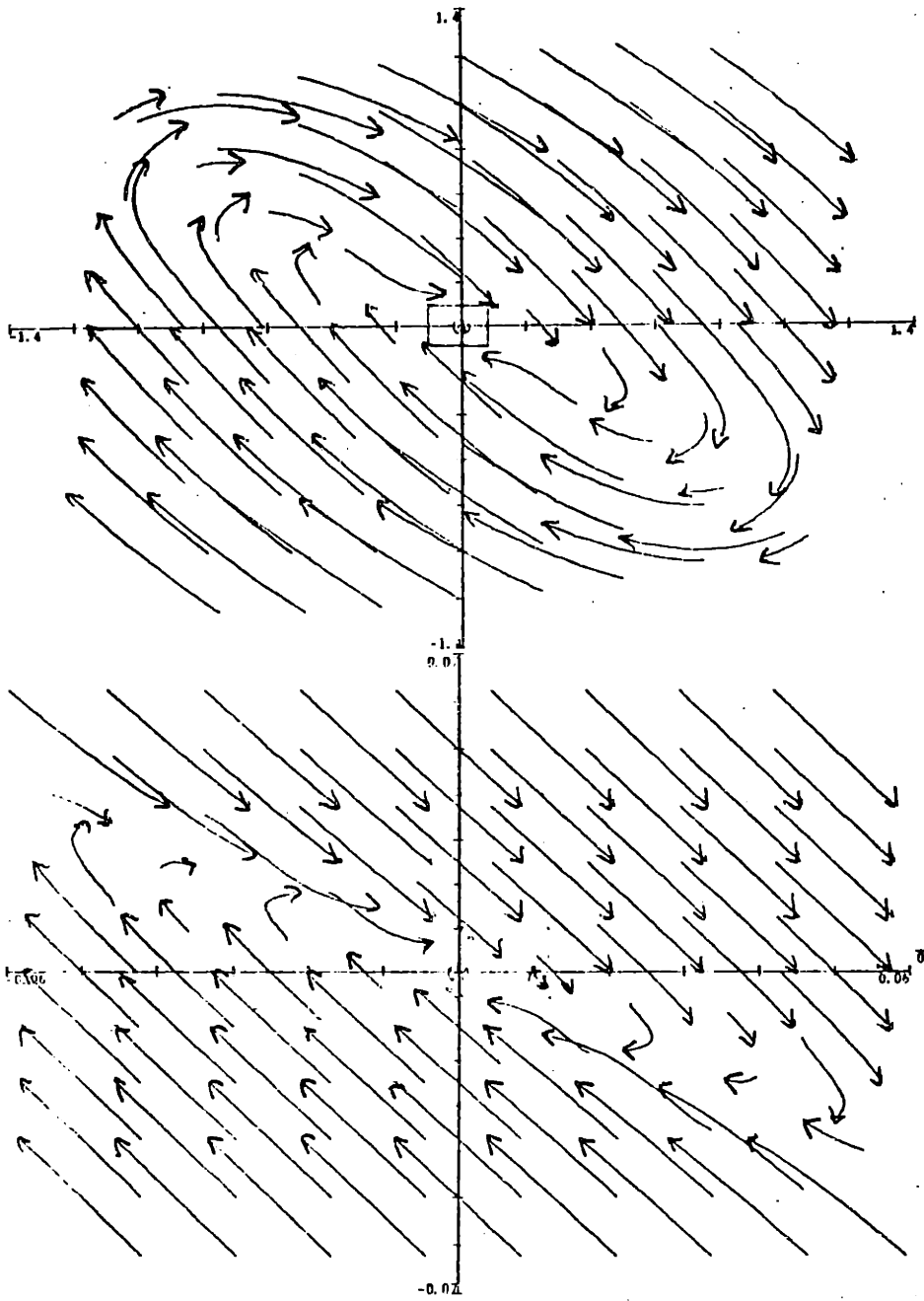




critical point, the origin. This critical point is predicted to be a stable node for saturable loss and an unstable node for saturable gain. PL.3.5.2-5 presents the trajectory fields for relatively large amplitudes and for the behavior near the critical point. Also shown are three numerically generated continuous trajectories. Once again the results and the predictions are in agreement. The physical interpretation for this large coupling is particularly simple. Around the origin the amplitudes are small enough that the system is relatively unperturbed. Since the coupling is large the stability is determined by the relative magnitudes of l_0 and g_0. Thus for saturable loss (A5: $l_0 > g_0$) the loss dominates and the amplitudes attenuate making the origin stable. For saturable gain (A5: $g_0 > l_0$) the gain will dominate causing amplitudes around the origin to increase thereby making the critical point unstable.

For the next case, A6, the coupling constant has increased (k^2=13.125) slightly so that the only critical point at the origin is now predicted to be a spiral point (stable for saturable loss and unstable for saturable gain). This slight modification in behavior is shown in PL.3.5.2-6.

Results for the next three cases of interest were generated from the dominant gain saturation equations,

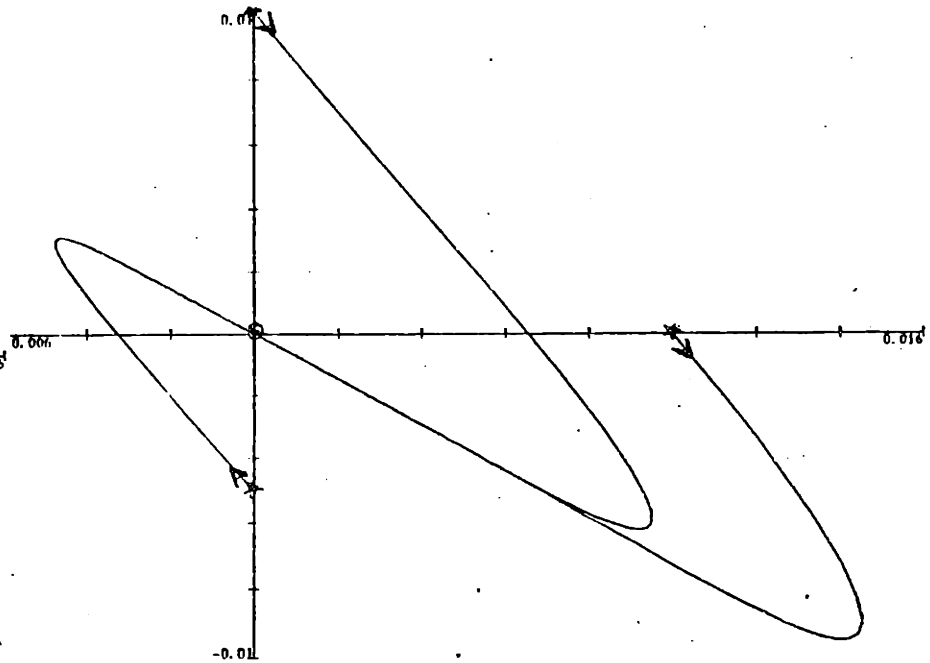


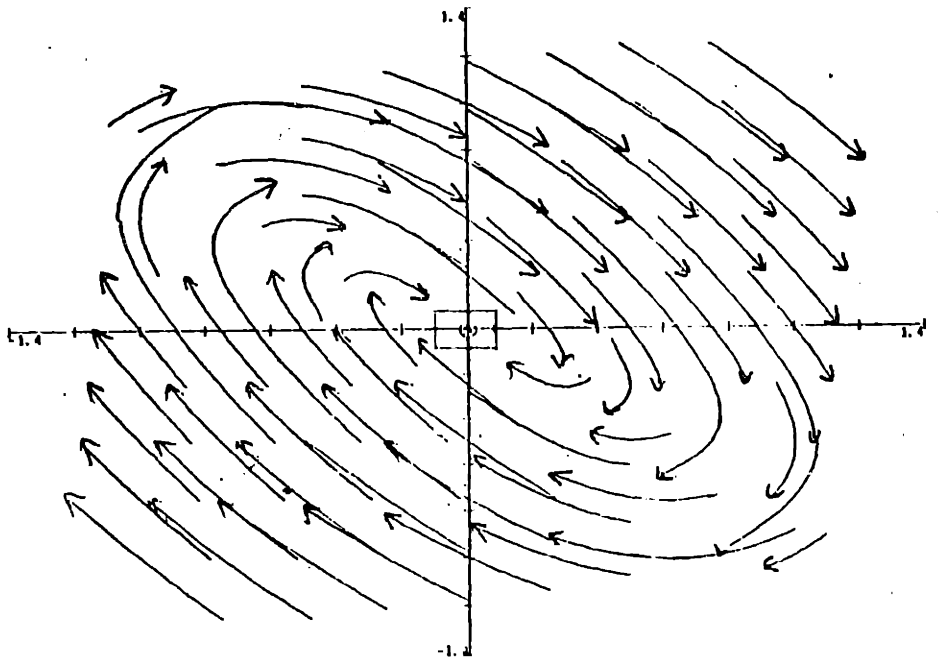
PL.3.5.2-5 Case A5

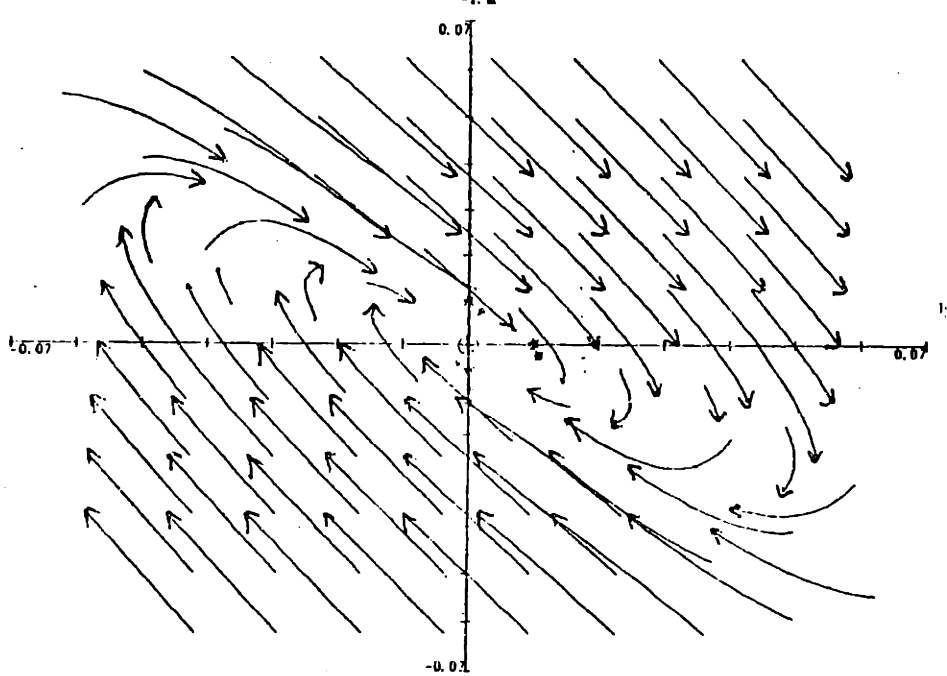
x-coordinate: a1 (saturable loss)
a2 (saturable gain)

y-coordinate: a2 (saturable loss)
a1 (saturable gain)

arrows for saturable loss

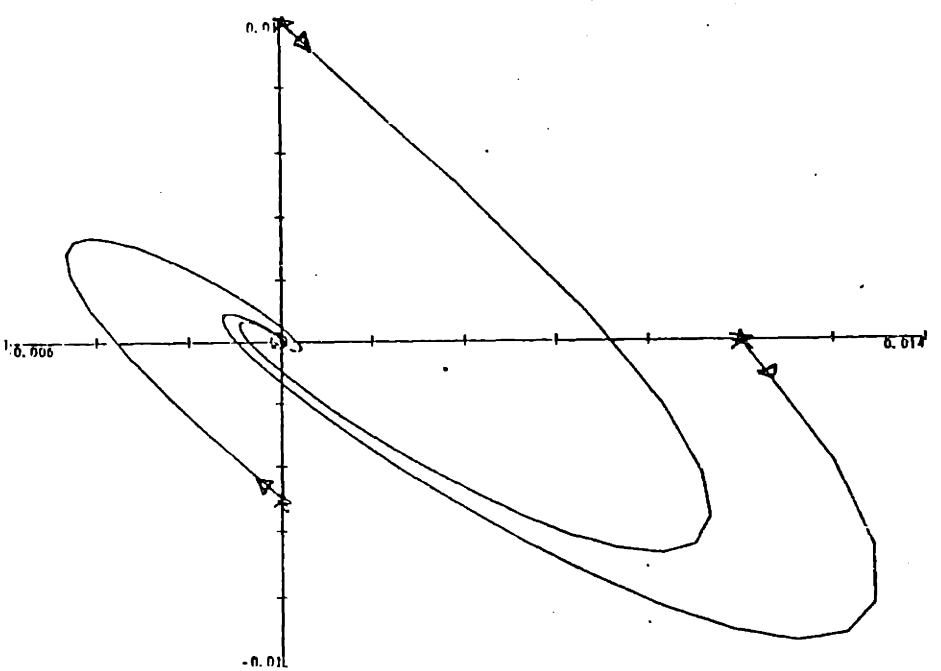






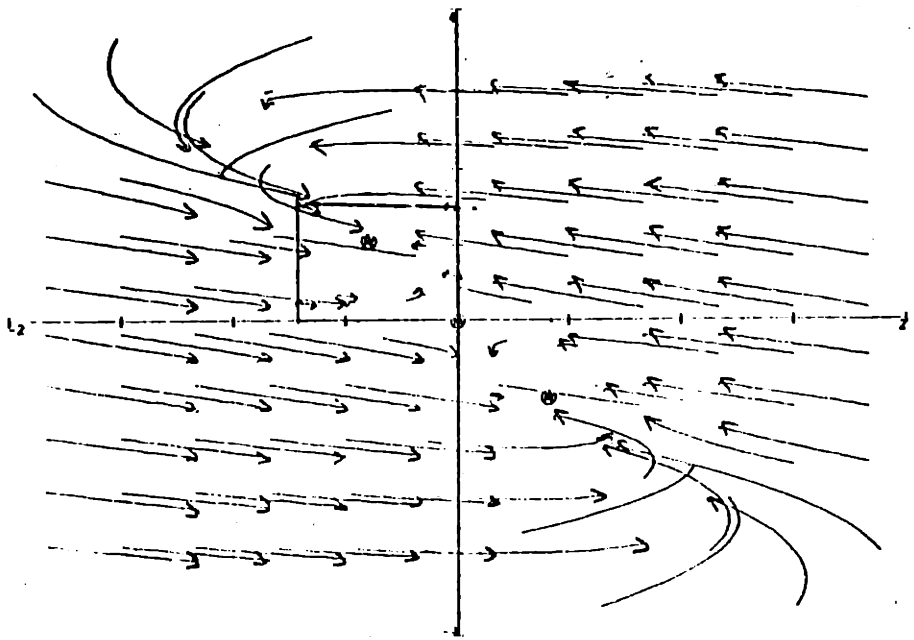
PL.3.5.2-6 Case A6
x-coordinate: a1 (saturable loss)
a2 (saturable gain)
y-coordinate: a2 (saturable loss)
a1 (saturable gain)

arrows for saturable loss



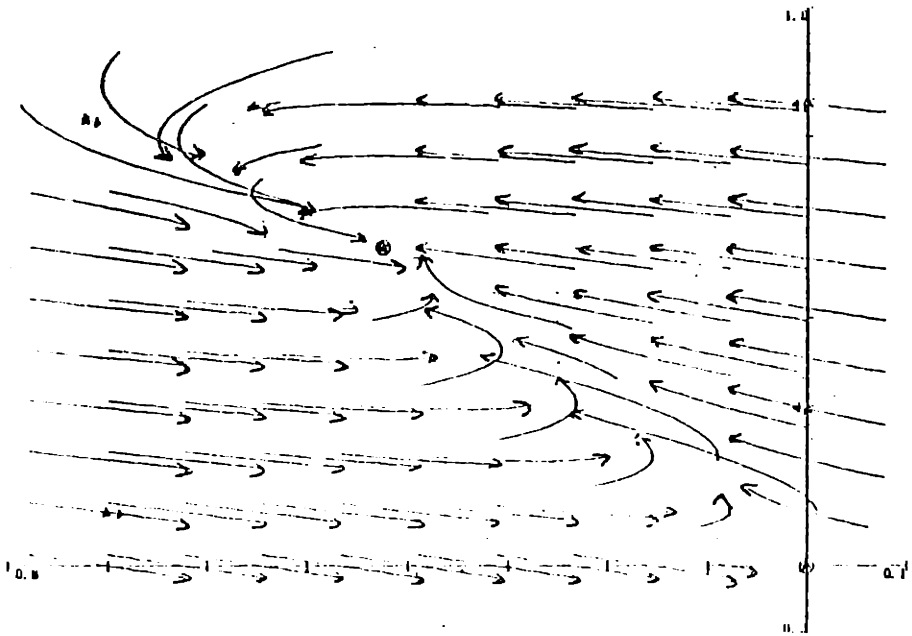
for l_0=5, g_0=2 (notice that the same results could have been obtained from the dominant loss saturation equations for l_0=2, g_0=5). For the case B1, k^2=5, the predicted behavior from critical point analysis is that the off-center critical point is a stable node (or an unstable node for the corresponding saturable loss device: l_0=2, g_0=5) and the central critical point is a saddle point. PL.3.5.2-7 demonstrates the basic stability. The second trajectory field shows a node-like behavior which is verified numerically for four different initial conditions. All results are in excellent agreement with the critical point predictions. Physically the interpretation is as follows. Weak coupling for saturable gain means that the waveguides do not affect each other very much so amplitude a1 saturates the gain thereby approaching some constant value while the loss causes attenuation of a2 reducing the system energy until equalibrium is achieved at the off-center critical points. For saturable loss, however, weak coupling means that the saturable loss can not prevent the linear gain from increasing the system energy without bound.

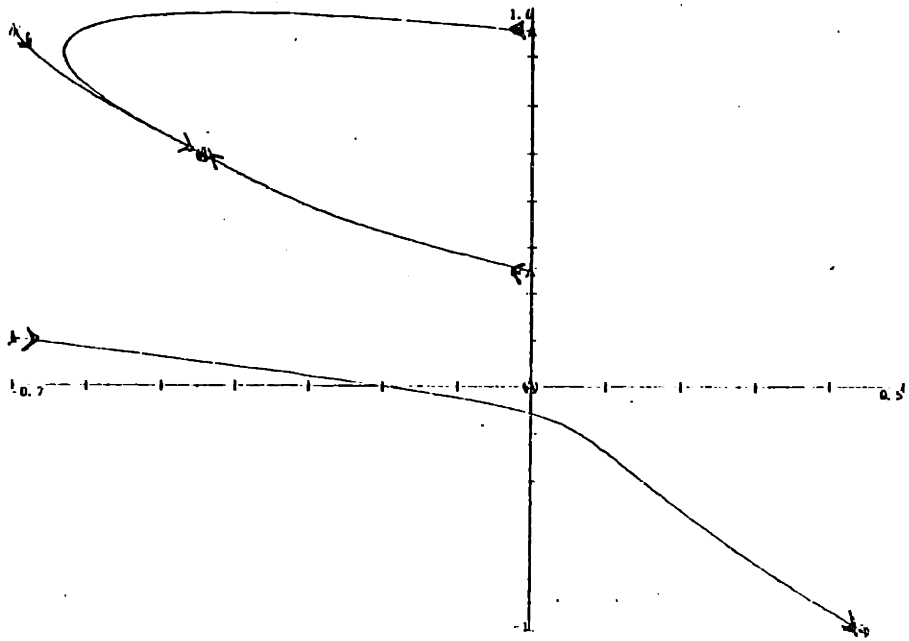
For B2, the coupling constant has increased enough (k^2=11.125) that $a_{(hi)}$ has become imaginary and the only critical point is the origin which is predicted to be a stable node for saturable gain and an unstable node for



PL.3.5.2-7 Case B1
x-coordinate: a1 (saturable gain)
a2 (saturable loss)
y-coordinate: a2 (saturable gain)
a1 (saturable loss)

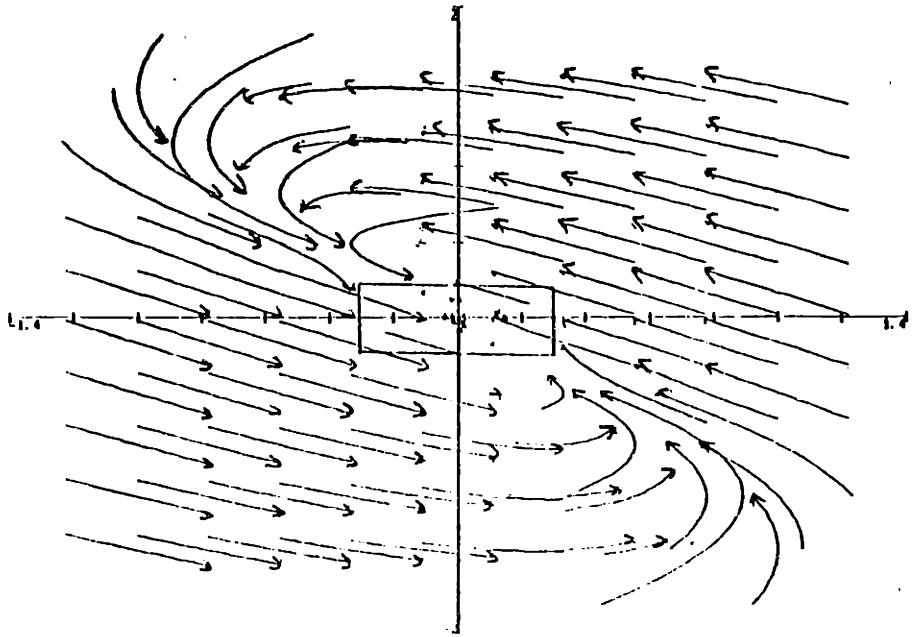
arrows for saturable gain

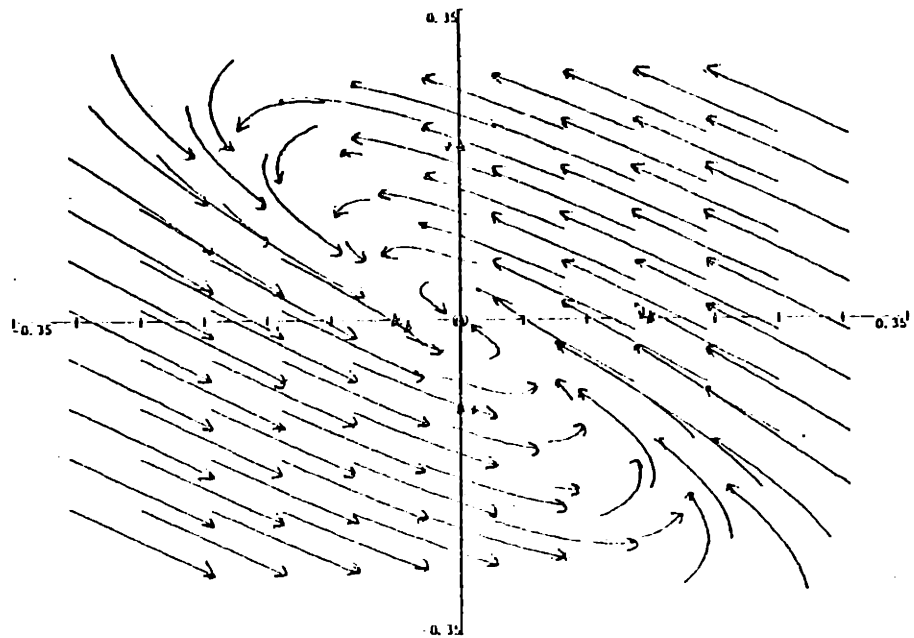




saturable loss. These predictions are verified in PL.3.5.2-8. Physically the interpretation is again extremely simple since near the origin the device is non-saturated or unperturbed. Therefore, in the saturable gain case(B2: $1_0 > g_0$) the loss is dominating causing stability at the origin and in the saturable loss case (B2: $1_0 < g_0$) the gain is dominating causing instability at the origin.

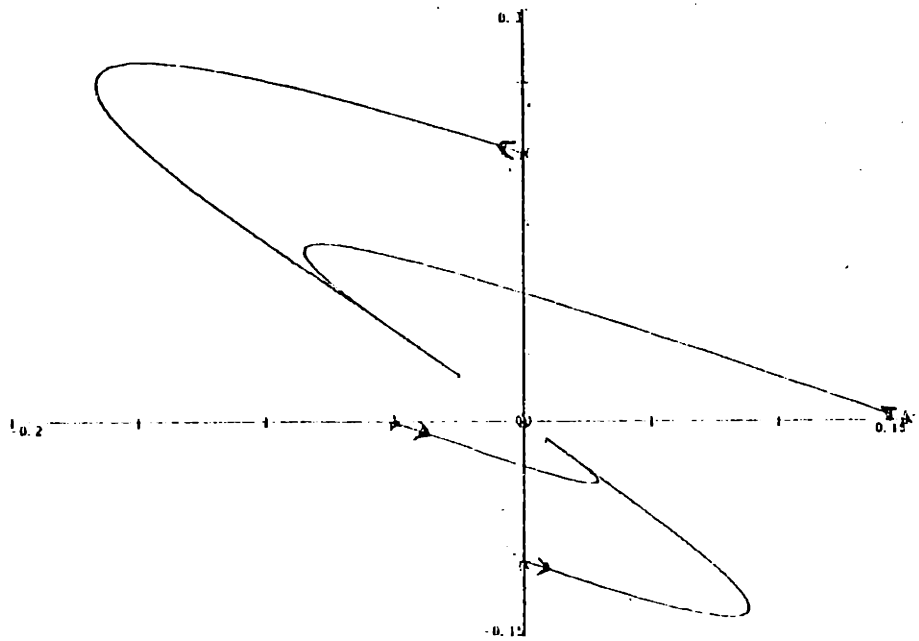
For B3, the coupling constant has increased ($k^2=15$) so that the origin is now predicted to be a stable spiral point for saturable gain and an unstable spiral point for saturable loss. This is verified and illustrated in PL.3.5.2-9. Results for the dual saturation case, D1 ($1_o=5$, $g_o=2$, $k^2=5$), are presented in PL.3.5.2-10. Again there is excellent agreement between theoretical predictions and the results of the two computational methods. Notice that dual saturation causes the trajectories to become more symetric (about a1=a2) and that large amplitudes saturate (i.e. produce trajectories perpendicular to the axes) for both axes, rather than for one axis as in the dominant saturation cases.

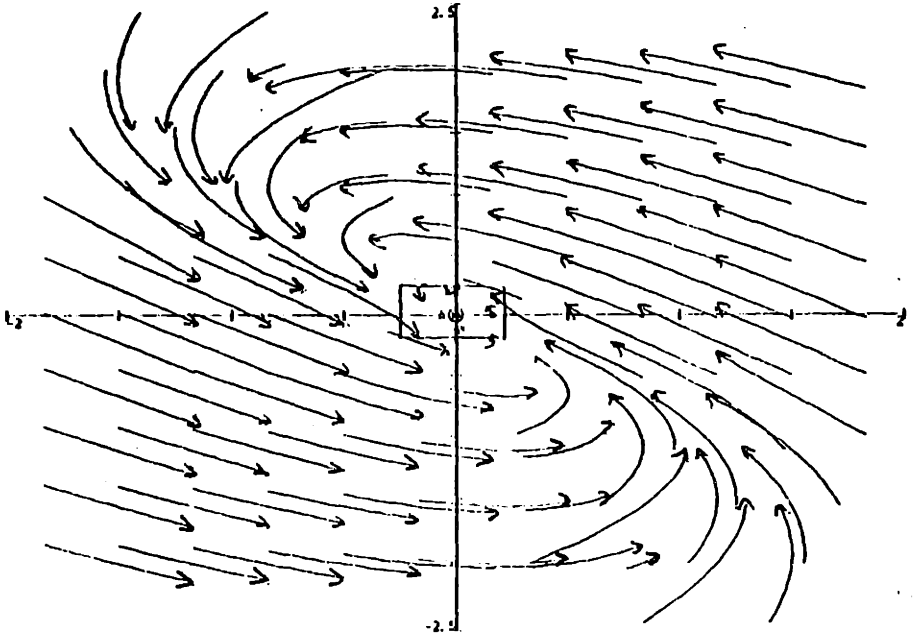


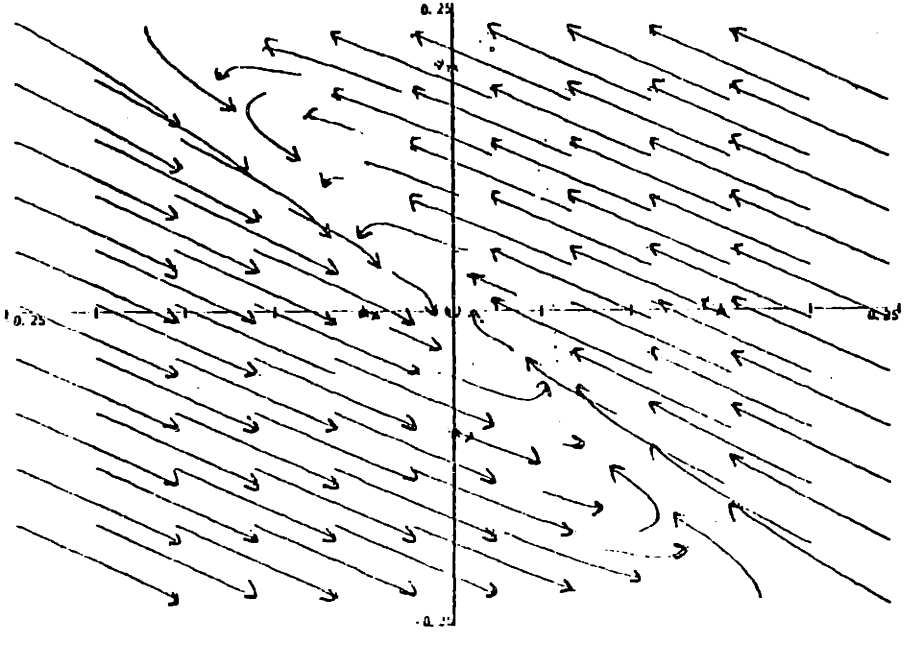


PL.3.5.2-8 Case B2
x-coordinate: a1 (saturable gain)
a2 (saturable loss)
y-coordinate: a2 (saturable gain)
a1 (saturable loss)

arrows for saturable gain

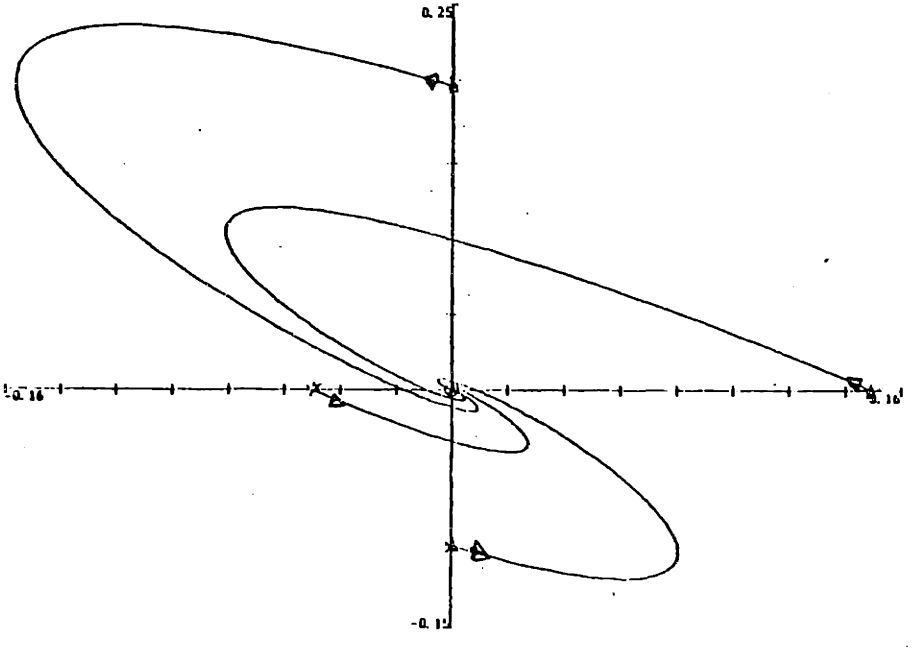


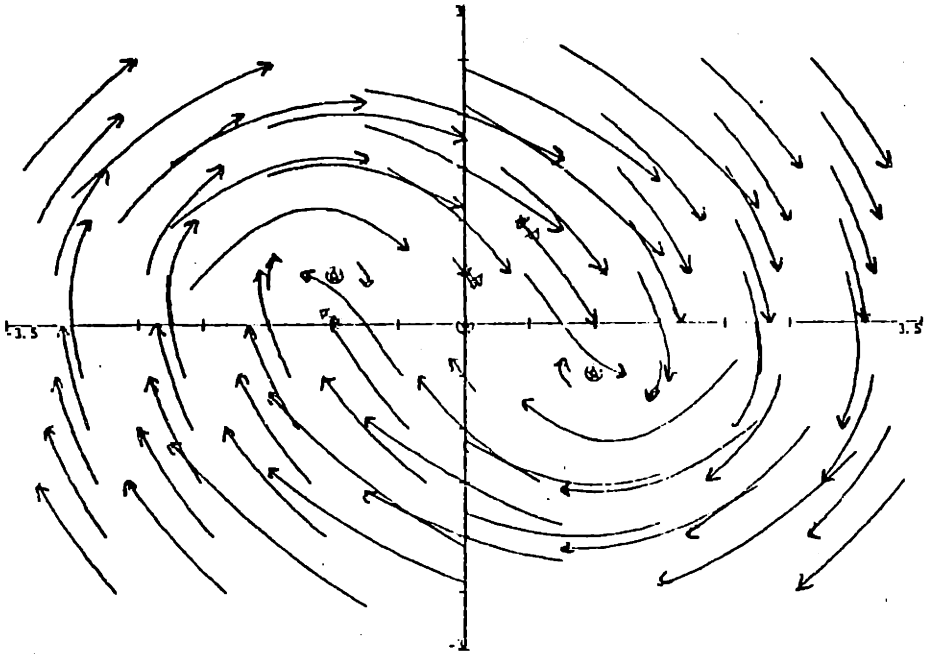




PL.3.5.2-9 Case B3
x-coordinate: a1 (saturable gain)
a2 (saturable loss)
y-coordinate: a2 (saturable gain)
a1 (saturable loss)

arrows for saturable gain

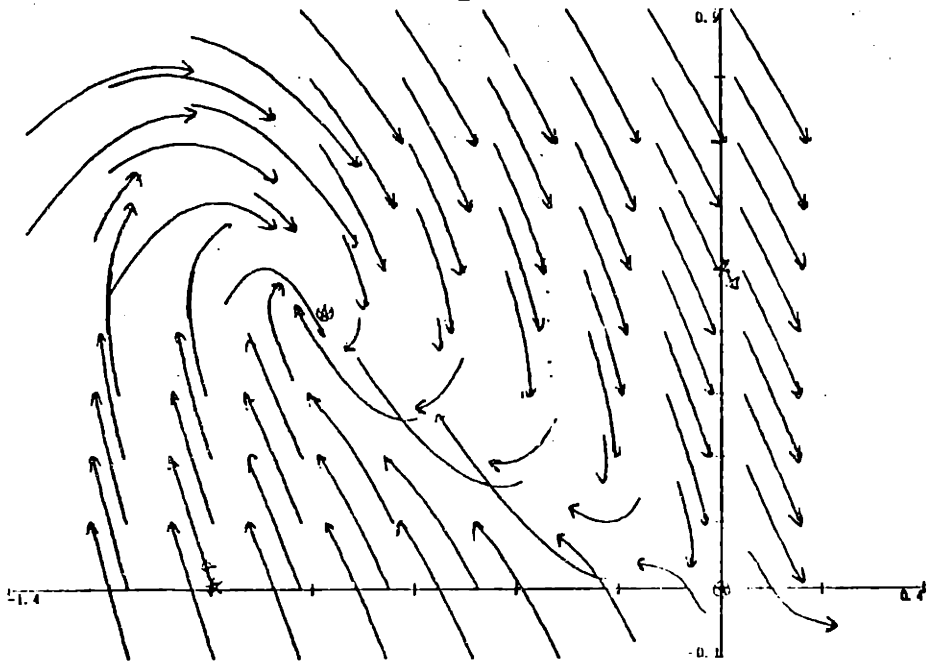




PL.3.5.2-10 Case D1

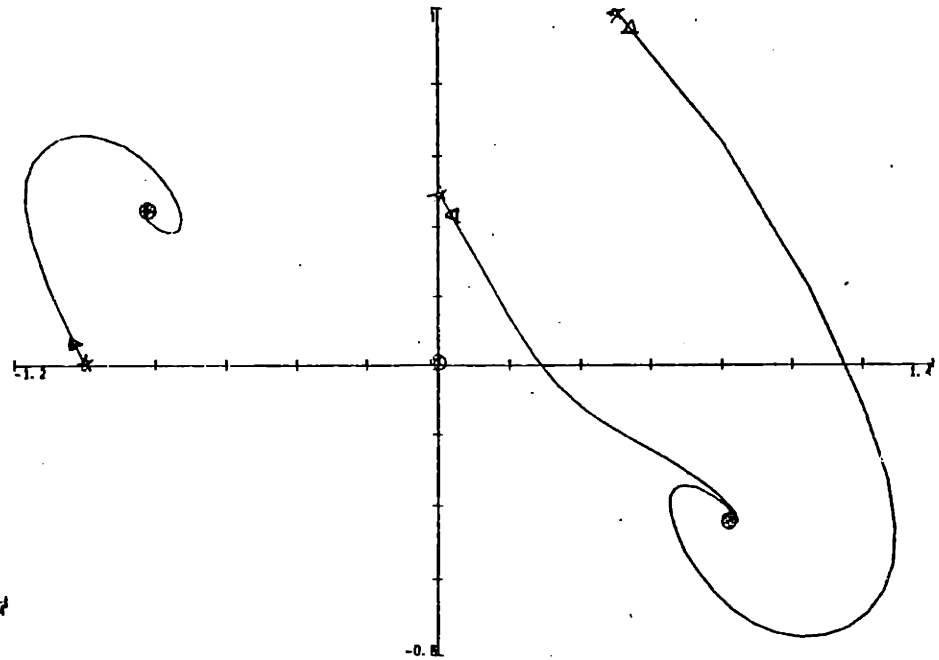
x-coordinate: a1

y-coordinate: a2



PL.3.5.2-10 Case D1

x-coordinate: a1, y-coordinate: a2



3.6 Device Applications Based Upon Steady State Results

Although steady state is a response limit that may never be achieved exactly, we may anticipate certain device applications based upon the steady state results for instances where steady state is approximately valid. One of these instances is when we have an input excitation that is slowly varying over a period of time that is large compared to the media response time, T1 (just how much larger for how close an approximation will be determined by the dynamic response).

The most obvious application (and the original impetus for consideration of this type of structure) is the use of these devices as an optical limiter or pulse height standardizer. An optical limiter is not only essential for optical logic devices but would also be extremely useful for optical communications and various types of physical measurements where optical pulse height and wave shape control are necessary.

To fabricate a pulse height standardizer we simply select a parameterization for one of the three types of waveguide structure (saturable loss, saturable gain, or dual saturation) such that there is stability at the off-center critical point. We then expect that if we can fabricate a device with this parameterization, an input pulse of arbitrary (within reason) amplitude and

substantial width should eventually (in distance that is) produce an output pulse with an amplitude that is always equal to the off-center critical point value. In theory, this output pulse height could be made as small as desired or as large as the value obtained for $k^2=k_s^2$ (e.g. for the dominant loss saturation case

$$a_{(hi)max} = a_{(hi)}\Big|_{k^2=k_s^2} = \left(\frac{4}{(1+\sqrt{1+8\,g_0\ell_0}\,)} - 1 \right)^{1/2}$$

In actuality, however, there will be a limit as to how small this standardized amplitude may be made due to parameter uncertainties and fluctuations. For example, if we were designing a pulse height standardizer for a small standardized height using a loss saturation device, we would want k^2 to be close to $g_0 l_0$, but we must keep $k^2 < g_0 l_0$. If incorporated all our uncertainty information into an effective deviation in k^2, say Δk^2, for some assumed l_0 and g_0, then the smallest we could make $a_{(hi)}$ can be shown to be

$$a_{(hi)min} = \left(\frac{g_0\ell_0}{g_0\ell_0 - \Delta k^2} - 1 \right)^{1/2} \simeq \sqrt{\Delta k^2/g_0\ell_0}$$

In actuality the limit as to how large we can make $a_{(hi)}$ is somewhat less than

$$a_{(hi)}\big|_{k^2=k_s^2}$$

since as shown later, when we consider the dynamic response, that for pulse height standardizers it is preferred to operate in the stable node region. Thus

$$a_{(hi)max} = a_{(hi)}\big|_{k^2=k^2_{spiral}}$$

but this is not very restrictive since it is anticipated that we would usually prefer to operate at a fraction of the medium saturation intensity anyway. If larger power levels are desired a different medium with a higher saturation intensity, could be used. A more restrictive practical constraint is that we can not presently fabricate devices with simultaneously large values of coupling constant, k, and device length, d. We may take this constraint as being $kd < 5$. Presently, however, we shall not be concerned with this since the matter will be considered extensively in Chapter V when we present actual device designs. Note that the fact that we can not simultaneously make both the origin and off-center critical points stable does not necessarily prohibit the

operation of the device as a pulse height standardizer since the steady state is not valid for the leading edge of a pulse or for the response due to noise, etc.

Another possible application based upon steady state results is the use of device as a mode selector. If the coupling of modes between multimode waveguides is predominantly just between modes of the same order then our model would describe the behavior of each coupled pair individually. The coupling constant will be different for the different mode pairs. Thus we could conceivably design a device that would force different ordered mode pairs to different (order dependent) fixed amplitudes (including of course zero amplitude). Mode selectors could be constructed in this fashion.

Another possible application involves the difference in phase between the two off-center critical points (since one has amplitudes that are the negative of the other they are 180°out of phase with each other). There are difficulties in using any phase state device, however, since sampling the output at a distance that deviates from the desired distance by half a wavelength will be seen as a change of state. Thus, we would have to be able to fabricate distances accurate to within half a wavelength. Operating in the millimeter wavelength range, however, may make phase state devices feasible. If so, we could use the

device as an amplitude discriminator. This is so since for stable off-center critical points some input amplitudes may converge to one critical point (say the one in the second quadrant) while other input amplitudes will converge to the other (say the one in the fourth quadrant). Detection of the state of phase could be simply achieved through interference techniques. It may be possible that phase state memory elements could also be fabricated. Note, however, that all the applications mentioned so far could only be expected to operate on time scales larger than the medium response time, Tl. To examine faster response behavior we must analyze the dynamic equations.

CHAPTER IV - DYNAMIC RESPONSE

Device response to arbitrary input excitation is now considered. The input pulses need not be slowly varying (with respect to T1) and transient responses are included.

In this chapter a transformation is performed on the dynamic equations which results in their simplification. Some algorithms for solution of the transformed equations are discussed. It is pointed out that a numerical approach is the most efficient. This numerical scheme is presented in detail after a brief discussion of the various finite difference options. Results from excitation by two different waveshapes for various heights and widths are presented for two different device parameterizations. Physical interpretations of the results are offered. Device applications based on the dynamic results are then presented.

4.1 Pulse Frame Transformation

The four nonlinear coupled partial differential equations that govern the general (dynamic) device response, EQ. 1.3.3-3, may be classified as a parabolic system that has one characteristic direction along the lines $y=x+b$, where y is the dimensionless time variable (normalized to T1), x is the dimensionless space variable (normalized to vT1), and b is the y intercept. The dominant loss or gain

saturation equations of course fall into the same classification with the same characteristic. Instead of trying to solve the equations in their present form, however, there is a transformation which may be performed that results in a significant reduction in mathematical complexity and computational difficulty. Examination of the dynamic equations shows that there are actually two basic directions of information propagation. One of these directions is the y direction along which the loss and gain parameters saturate. In other words the loss and/or gain parameter at any point in space-time has a value that depends only on the values of that parameter and its saturating mode amplitude at that space point for all previous time points. The other direction is the characteristic direction (i.e., the lines y=x+b) along which the mode amplitudes propagate. In other words the mode amplitudes at any point in space-time depend only on the values of the mode amplitudes and the loss and gain parameters at previous points along the line y=x+b, where b is the y intercept of a line of unit slope that intersects that space-time point. Therefore the loss and gain parameters may be evaluated by integrating the saturation equations in the y direction for each specific value of x over the interval in y for which values of a1 and a2 are known (from previous calculation or specified boundary conditions). A similar integration of

the coupled mode equations may be performed along the y=x+b direction if we simply parameterize x and y for any specified value of b. Thus we let, $x=x(s)=s$; $y=y(s)=s+b$

then $\frac{\partial}{\partial s} = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}$ and

$l(x,y)=l(s)$, $g(x,y)=g(s)$, $a1(x,y)=a1(s)$, $a2(x,y)=a2(s)$

so that the coupled mode equations become (for any specific value of b):

$$\frac{da1(s)}{ds} = g_0 g(s) a1(s) + k\, a2(s)$$

(Eq. 4.1-1)

$$\frac{da2(s)}{ds} = -\ell_0 \ell(s) a2(s) - k\, a1(s)$$

The total derivatives were used since b was assumed to be some particular value. In general, therefore, the coupled mode equations may be written as: *

(Eq. 4.1-2)

$$\frac{\partial a1(s,y)}{\partial s} = g_0 g(s,y) a1(s,y) + k\, a2(s,y)$$

$$\frac{\partial a2(s,y)}{\partial s} = -\ell_0 \ell(s,y) a2(s,y) - k\, a1(s,y)$$

This transformation shall be referred to as the pulse frame transformation since it has the following physical interpretation. Select a little bit of input pulse energy and travel along the waveguide in the frame of that infinitesimal "energy chunk". Mathematically this is equivalent to selecting some value of b and traveling along the line y=x+b with increasing parameter s. Note that this transformation results in substantial simplification since Eq. 4.1-2 involves derivatives with respect to a single independent variable only.

*Notice that these equations depend only on two independent device parameters (say ℓ_0/k and g_0/k) as was the case for the steady state equations. This is considered in 5.1.

There are two types of computational methods that have been developed for the solution of the dynamic equations under the pulse frame transformation. One involves a combined numeric and analytic algorithm. This algorithm uses linear finite difference formulae for the integration of the saturation equations thus generating approximations of $l(s)$ and $g(s)$. These $l(s)$ and $g(s)$ are approximate since the linear finite difference formulae merely predict l and g along the line $y=x+b$ from the behavior along the previous line $y=x+(b-\delta b)$, thereby neglecting the instantaneous saturation effects (which are the source of the nonlinearity). The algorithm then generates an approximate analytic solution of Eq. 4.1-1 using interpolation polynomials for the functions $l(s)$ and $g(s)$ (these polynomials are produced from the numerical values of l and g obtained by the finite difference formulae). This algorithm required a lot of time and elaborate mathematics for its development. It can be used to generate correct solutions but the computational times necessary make it impractical. The main difficulty lies in the fact that the difference equations neglect the instantaneous (nonlinear) saturation effects so that extremely small time increments (δb) must be used or iterations must be employed. There may be certain cases (long device lengths or rapid mode amplitude variations) where this algorithm is superior and

future research may produce modifications that eliminate the time step size difficulties. Details of this algorithm and of various methods of its implementation are available from the author but will not be discussed further here.

The other computational method employed circumvents the aforementioned difficulties by using a totally numeric approach that allows for the instantaneous field-media interaction by retaining the nonlinearities involved. This algorithm is discussed in the next section and is the method used to generate all of the dynamic results presented.

4.2 Nonlinear Finite Difference Equations

In choosing a quantization scheme for the numerical solution of a differential equation one must consider the problems of convergence, stability, and computation time. Ascertaining the convergence and stability properties of a finite difference approximation is usually difficult and often impossible for nonlinear differential equations. It is shown later in this section, however, that a simple probabilistic error analysis exists which provides criteria necessary for convergence and stability although it does not rigorously prove that these criteria are sufficient for the existence of these properties. Despite this fact, the arguments for the existence of convergence and stability are intuitively palatable and empirically correct. Before

delving into this, however, let us first consider some of the options available in selecting a finite difference scheme.

For the dynamic equations (as is the case for most systems) it is preferred to leave the equations as a system of coupled first order equations rather than combining some or all of the equations. Thus finite difference formulae for first ordered derivatives only need be considered. Consider a function, $\phi(r)$, for which a quantized solution is desired at the points $r=i(\delta r)$, $\forall i \in [0, i_{max}]$, where δr is the "step size' and i takes on integer values. The first order derivative of with respect to r evaluated at is given by the exact relationship[22]

(Eq. 4.2-1) $$\left.\frac{d\phi}{dr}\right|_i = \frac{1}{(\delta r)} [\log(1+\Delta)]\phi_i \ .$$

Where $\phi_i \equiv \phi(i\delta r)$, and Δ is the forward difference operator defined by $\Delta\phi_i = \phi_{i+1} - \phi_i$. Expanding the logarithm one obtains

(Eq. 4.2-2) $$\left.\frac{d\phi}{dr}\right|_i = \frac{1}{(\delta r)} [\Delta - (1/2)\Delta^2 + (1/3)\Delta^3 - (1/4)\Delta^4 + \ldots]\phi_i$$

Where $\Delta^2\phi_i = \Delta(\Delta\phi_i) = \Delta(\phi_{i+1} - \phi_i) = \phi_{i+2} - 2\phi_{i+1} + \phi_i$

etc. Since it is only possible to perform a finite number of operations in a finite amount of time, the above formula must be truncated somewhere thereby yielding truncation errors. The order of a finite difference approximation refers to the highest ordered difference operator retained.

Thus

$$\text{(Eq. 4.2-2)}\quad \left.\frac{d\phi}{dr}\right|_i \simeq \frac{1}{(\delta r)}\left[\Delta - (1/2)\Delta^2 + (1/3)\Delta^3 - \ldots + (1/n)\Delta^n\right]\phi_i$$

is an n th order forward difference formula which will have a truncation error that is proportional to $(\delta r)^n$.

Other formulae may be obtained by the use of shifting operators as follows,

$$\text{(Eq. 4.2-3)}\quad \left.\frac{d\phi}{dr}\right|_i = \frac{1}{(\delta r)}\left[\log(1+\Delta)\right]\phi_i = \frac{1}{(\delta r)}\left[\log(1+\Delta)\right](1+\Delta)\phi_{i-1}$$

$$= \frac{1}{(\delta r)}\left[\Delta + (1/2)\Delta^2 - (1/6)\Delta^3 + (1/12)\Delta^4 - \ldots\right]\phi_{i-1}$$

or by the use of backward difference operators as in

$$\text{(Eq. 4.2-4)}\quad \left.\frac{d\phi}{dr}\right|_i = \frac{-1}{(\delta r)}\left[\log(1-\nabla)\right]\phi_i$$

$$= \frac{1}{(\delta r)}\left[\nabla + (1/2)\nabla^2 + (1/3)\nabla^3 + (1/4)\nabla^4 + \ldots\right]\phi_i$$

where $\nabla\phi_i \equiv \phi_i - \phi_{i-1}$. Obviously there are several other formulae that may be obtained through the use of various shifting and averaging operations.

Although truncation errors are proportional to the step size raised to the order of the finite difference formula, blind selection of higher order formulae will not necessarily lead to improved behavior since some of these formulae either will not converge to the original (continuous) equations or will become numerically unstable. Convergence and stability of the use of first order formulae for the dynamic equations has been proved (this is usually the case, due to the limit definition of a derivative) however, the requisite smallness of the step sizes results in

computation times that are intolerable.

Note that the application of forward difference formulae to the medium saturation equations produces equations that involve the medium parameter (l and/or g) at a point that is ahead (in time) of the point at which the nonlinear term ($a1^2$ and/or $a2^2$) occurs. Therefore, formulae of this type linearize the equations as alluded to in the discussion of the semianalytic algorithm in the previous section. This is unfavorable, however, since saturation is an instantaneous process. Consequently, forward difference formulae for the media saturation equations converge only for very small time steps regardless of the order of approximation used.

The use of backward difference formulae eliminates this problem by allowing for the instantaneous field-medium interaction. This results in a nonlinear system of algebraic equations but the nonlinearity is simple enough that the MACSYMA software may be used to obtain numerical solutions. Higher order formulae may be used for the time derivatives so that larger step sizes may be employed resulting in a substantial reduction of computation time.

After considering a large number of applicable formulae the optimal choice appears to be the use of a third order backward difference formula for the time derivatives and a first order backward difference formula for the derivatives with respect to s. The third order formula obtained from

from Eq. 4.2-3 may be written as

$$\left.\frac{d\phi}{dr}\right|_i \simeq \frac{1}{6(\delta r)} [11\phi_i - 18\phi_{i-1} + 9\phi_{i-2} - 2\phi_{i-3}]$$

and the first order formula is simply

$$\left.\frac{d\phi}{dr}\right|_i \simeq \frac{1}{(\delta r)} [\phi_i - \phi_{i-1}].$$

Using these formulae results in the following quantization of the dynamic equations for the dual saturation case:

(Eq. 4.2-5)
$$al_{i,j} - al_{i,j-1} - (\delta s) g_0 g_{i,j} al_{i,j} - (\delta s) k\, a2_{i,j} = 0$$

$$a2_{i,j} - a2_{i,j-1} + (\delta s) \ell_0 \ell_{i,j} a2_{i,j} + (\delta s) k\, al_{i,j} = 0$$

$$11g_{i,j} - 18g_{i,j-1} + 9g_{i,j-2} - 2g_{i,j-3} + 6(\delta y) g_{i,j} (1+(al_{i,j})^2) - 6(\delta y) = 0$$

$$11\ell_{i,j} - 18\ell_{i,j-1} + 9\ell_{i,j-2} - 2\ell_{i,j-3} + 6(\delta y) \ell_{i,j} (1+(a2_{i,j})^2) - 6(\delta y) = 0.$$

Where δs is the step size in the s direction (associated with the index i) and δy is the step size in the y direction (associated with the index j). In other words, $s=i\ (\delta s)$, $y=s+j\ (\delta y)$.

Similarily the quantization of the dynamic equations for the dominant loss saturation case is:

(Eq. 4.2-6)
$$al_{i,j} - al_{i,j-1} - (\delta s) g_0 al_{i,j} - (\delta s) k\, a2_{i,j} = 0$$

$$a2_{i,j} - a2_{i,j-1} + (\delta s) \ell_0 \ell_{i,j} a2_{i,j} + (\delta s) k\, al_{i,j} = 0$$

$$11\ell_{i,j} - 18\ell_{i,j-1} + 9\ell_{i,j-2} - 2\ell_{i,j-3} + 6(\delta y) \ell_{i,j} (1+(a2_{i,j})^2) - 6(\delta y) = 0$$

and the quantilization for the dominant gain saturation case is:

(Eq. 4.2-7)
$$al_{i,j} - al_{i,j-1} - (\delta s)g_0 g_{i,j} al_{i,j} - (\delta s)k\, a2_{i,j} = 0$$

$$a2_{i,j} - a2_{i,j-1} + (\delta s)\ell_0 a2_{i,j} + (\delta s)k\, al_{i,j} = 0$$

$$11g_{i,j} - 18g_{i,j-1} + 9g_{i,j-2} - 2g_{i,j-3} + 6(\delta y)g_{i,j}(1+(al_{i,j})^2) - 6(\delta y) = 0$$

As an example of the imposition of boundary conditions suppose the saturable loss device is operated as a single input pulse shaping device. An input pulse, say pulse(y), is fed into the gain waveguide while the lossy waveguide input is kept at zero. The system is initially unperturbed so the boundary conditions are:

$$al(x,0) = 0, \quad a2(x,0) = 0, \quad \ell(x,0) = 1$$

$$al(0,y) = \text{pulse}(y), \quad a2(0,y) = 0, \quad \ell(0,y) = 1.$$

These conditions are consistent with the differential equations. In imposing these conditions on the finite difference equations it is valid to use the third order media equation for the first two calculations under the assumption that all previous conditions were unperturbed also (i.e., $\ell_{i,0} \equiv 1$, $\ell_{i,-1} \equiv 1$, $\ell_{i,-2} \equiv 1$. Mathematically and physically this assumption is allowed since it is equivalent to the assumption that the input pulse begins with a zero amplitude of duration greater than or equal to $2(\delta y)$.

Conditions necessary for the convergence and stability of the finite difference approximations shall now be derived. Rather than presenting the error analysis for all

the equations, the theory shall be demonstrated for a single equation and results will be cited for the others. Consider the loss saturation equation. Let $L_{i,j}$ represent the exact solution for ℓ, evaluated at $s = i(\delta s)$, $y = s + j(\delta y)$, and $\varepsilon^{(\ell)}_{i,j}$ represent the total quantization error for ℓ at that point. Thus, $\ell_{i,j} = L_{i,j} - \varepsilon^{(\ell)}_{i,j}$. Using this notation in the loss saturation equation one obtains in the limit as $\delta s \to 0$ and $\delta y \to 0$,

(Eq. 4.2-8) $$11\varepsilon^{(\ell)}_{i,j} = 18\varepsilon^{(\ell)}_{i,j-1} - 9\varepsilon^{(\ell)}_{i,j-2} + 2\varepsilon^{(\ell)}_{i,j-3} + 6(\delta y)\left(\frac{\partial L}{\partial y}\bigg|_{i,j}\right)$$

The last term vanishes in the limit providing that L is differentiable (as it must be by physical arguments). This equation describes the propagation of errors for ℓ in the limit as $\delta s \to 0$ and $\delta y \to 0$. Ordinarily one would attempt to prove boundedness of the magnitude of the jth error by use of the Schwartz inequality, however, this method is useless for the present case. Instead let us consider the effects of a single quantization error. Say the first quantization error occurs at $j = 0$ and there are no independent errors at $j = 1$, $j = 2$, or $j = 3$. In other words; $\varepsilon^{(\ell)}_{i,3}$, $\varepsilon^{(\ell)}_{i,2}$, and $\varepsilon^{(\ell)}_{i,1}$ represent the quantization errors due solely to the presence of $\varepsilon^{(\ell)}_{i,0}$. From Eq. 4.2.-8 it may be shown that all of these errors

must be of the same sign so that Eq. 4.2-8 may be written as

$$11|\varepsilon_j^{(\ell)}| = 18|\varepsilon_{j-1}^{(\ell)}| - 9|\varepsilon_{j-2}^{(\ell)}| + 2|\varepsilon_{j-3}^{(\ell)}|$$

where the arbitrary index, i, has been dropped. There will be some maximum error, say $|E|$, over the points j-1, j-2 and j-3, so that $11|\varepsilon_j^{(\ell)}| \leq (18 - 9 + 2)|E|$ or $|\varepsilon_j^{(\ell)}| \leq |E|$.

If the last inequality did not hold then the errors at later points due to a single error would be unbounded and convergence and/or stability would not exist.

Formulae for which similar inequalities do not hold have been demonstrated to generate nonsensical results (as they must since they would be proven non-convergent). Applying similar arguments to the other equations of Eq. 4.2-5, Eq. 4.2-6, and Eq. 4.2-7, results in the stipulation $(\delta s) < (1/g_0)$. The equations are proven to be non-convergent when this condition is violated and empirical results bear this out.

Note that this condition is necessary but not sufficient. Convergence and stability may not be proved by this argument since it describes only the effects of a single error. There are several non-rigorous arguments, however, that make this type of convergence "proof" palatable. One might argue that superposition holds for the effects of multiple errors since the equations

are quasilinear (i.e., the highest ordered derivatives appear linearly). One might also argue that all of the quantization errors will be of the same sign anyway since the proof is for the limit as $\delta s \to 0$ and $\delta y \to 0$. Probabilistically, one might argue that on the average (over j) the errors will be of the same sign since the parts due to previous errors are of the same sign and the independent parts are of either sign with equal likelihood (there is an initial swelling effect in error propagation that strengthens this argument even more). Perhaps the best argument for the validity of the single error analysis as a convergence proof is the fact that the empirical results make physical sense when the "proof" holds.

4.3 Dynamic Results and Interpretations

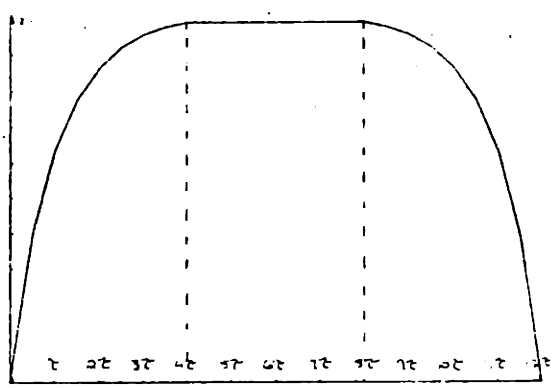
Since the main objective of this thesis is to demonstrate the operation of the proposed travelling wave structure as a practical and useful optical pulse shaping device, dynamic response results are presented only for the dominant loss saturation case rather than attempting to enumerate all possible cases and behaviors. The two parameterizations considered are those referred to as A3 (l_o=5, g_o=2, k^2=8.505) and A4 (l_o=5, g_o=2, k^2=9.695) for which the steady state response has already been examined. These cases are off-center stable spiral point and off-center stable node examples of the saturable loss device.

The input pulses for which the dynamic responses are computed are of two basic types. One is an exponential edged rectangular pulse. The other is the hyperbolic secant pulse. For both of these wave forms three different pulse widths are used. One of these pulse widths is long compared to the medium response time, T1, and shall be referred to as an l-type pulse. Another pulse width is short (or narrow) compared to T1 and is called an n-type pulse. A pulse of medium width is also considered which has a pulse width comparable to T1 and is referred to as an m-type pulse.

Portions of the text on the following page(s) are not legible in the original.

The exponantial edged rectangular pulse, intended to be representative of a typical binary logic pulse, is presented in figure 11.

FIGURE 11 - Logic Pulse



The pulse may be defined as follows.

$$\text{pulse}(y) \equiv \begin{cases} A\,(1 - \exp(-y/\tau)) & (0 \leq y \leq 4\tau) \\ A\,(1 - \exp(-4)) & (4\tau \leq y \leq 8\tau) \\ A\,(1 - \exp((y-12\tau)/\tau)) & (8\tau \leq y \leq 12\tau) \end{cases}$$

where A defines the pulse amplitude. Thus the leading edge is exponentially increasing for four exponential time constants, (i.e., 4τ), the pulse is then flat for four more time constants, and the trailing edge falls exponentially during the last four time constants. For l, m, and n type exponential edged logic pulses we somewhat arbitrarily choose, $\tau = 4.0$, $\tau = 0.5$ and $\tau = (1/16)$ respectively (recall that y is normalized, i.e., $t = \tau \rightarrow y = 1$). The value of A used is (0.2).

The hyperbolic secant pulse is presented in figure 12.

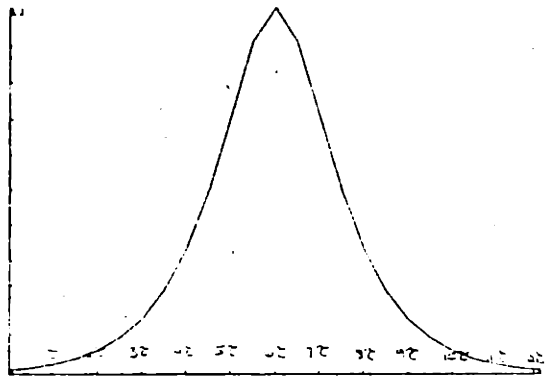


FIGURE 12 - Hyperbolic Secant Pulse

The pulse may be defined as follows:

$$\text{pulse}(y) \equiv A\ \text{sech}(.8844(y-6\tau)/\tau)$$

Where again A defines the amplitude. Thus the hyperbolic secant pulse power is down by 3dB when y deviates from 6τ (the location of the maximum power) by a factor of 2τ. In other words, 2τ is the 3dB power pulsewidth (FWHM). Note that 2τ is also a good definition of rise time for the hyperbolic secant pulse since it is the time required for the pulse energy to rise form about one ninth of its maximum to its maximum value. In addition to the well known soliton properties of the hyperbolic secant pulse in self-induced transparency phenomena, the pulse is useful since it is analytic and displays both types of convexity (which is important for the problem at hand). For l, m, and n type hyperbolic secant pulses we somewhat arbitrarily choose, $\tau = 2$, $\tau = (1/4)$ and $\tau = 1/32$, respectively. Again, A = .2 is used.

The dynamic results are presented in two different sets of plots. The first set demonstrates pulse

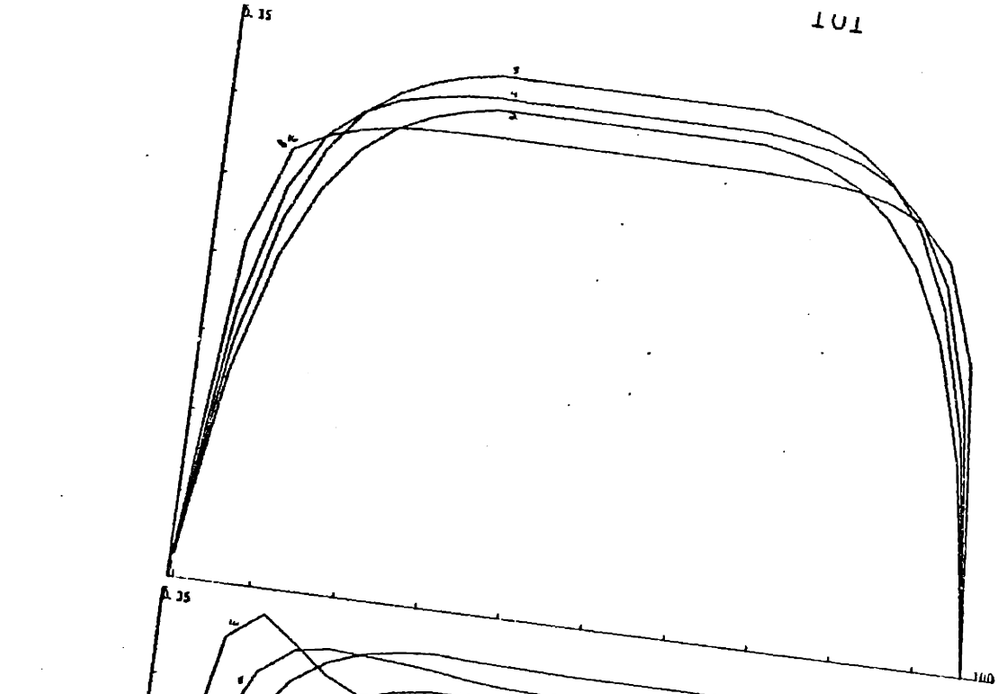
evolution as it presents the gain waveguide amplitude, al, versus time, y, at four different values of distance, x. These four different distances are labeled 2, 4, 8, and 16 (from the δs step index, i), corresponding to the waveguide distances x = .5657, 1.132, 2.263, and 4.525 (δs=.4 was used) respectively. The second set of plots presents al versus y at x = 4.525 (i=16) along with both the unperturbed response (see Eq. 3.2.2-1) and the steady state response evaluated at the same distance for the same input pulse. These plots facilitate the physical interpretations of the results. For both types of plot sets the l, m, and n pulse width responses are presented on one page for each waveshape and parameterization.

A summary of the parameters used for all cases plotted is presented in figure 13.

FIGURE 13 - Dynamic Plot Parameter Summary

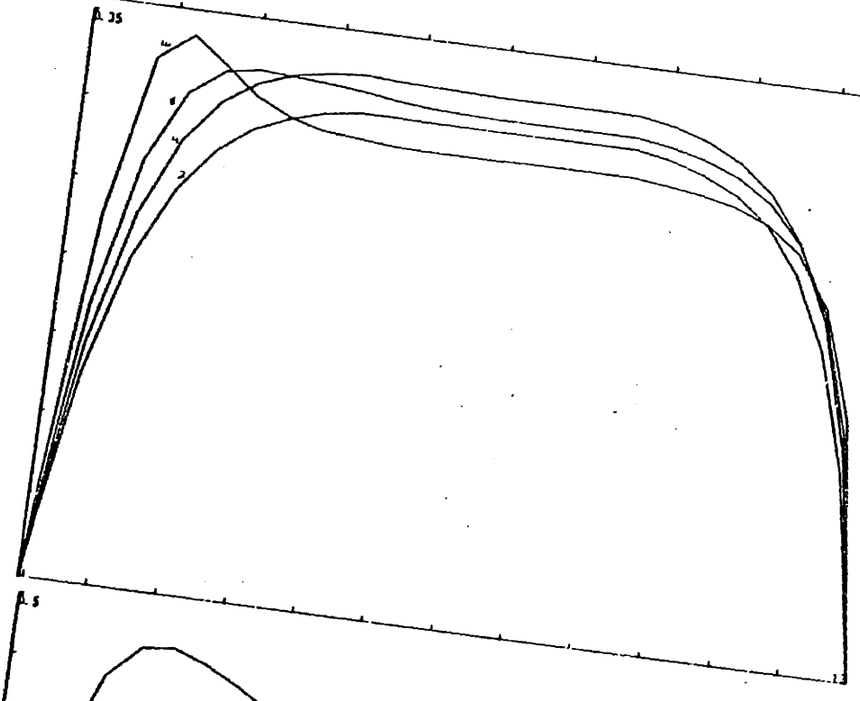
Parameterization	Pulse	τ	Plot 1	Plot 2
A4	logic	4	PL.4.3-1	PL.4.3-5
		1/2		
		1/16		
A3	logic	4	PL.4.3-2	PL.4.3-6
		1/2		
		1/16		
A4	hyp.sec.	2	PL.4.3-3	PL.4.3-7
		1/4		
		1/32		
A3	hyp.sec.	2	PL.4.3-4	PL.4.3-8
		1/4		
		1/32		

Pl. 4.3-1 presents the evolution of the exponential edged logic pulse for the off-center stable node parameterization, A4. We see that for the 1-type pulse the amplitude becomes close to the critical point value for even small distances and then gradually achieves the critical point value exactly. Once this value is achieved, a further increase in distance results in a steepening of the leading and trailing edges along with a widening of the standardized pulse height region. The m-type pulse also converges to the critical point but the leading edge has a slight initial overshoot. The n-type pulse has a greater initial overshoot and does not actually converge to the critical point. These results will be interpreted later.

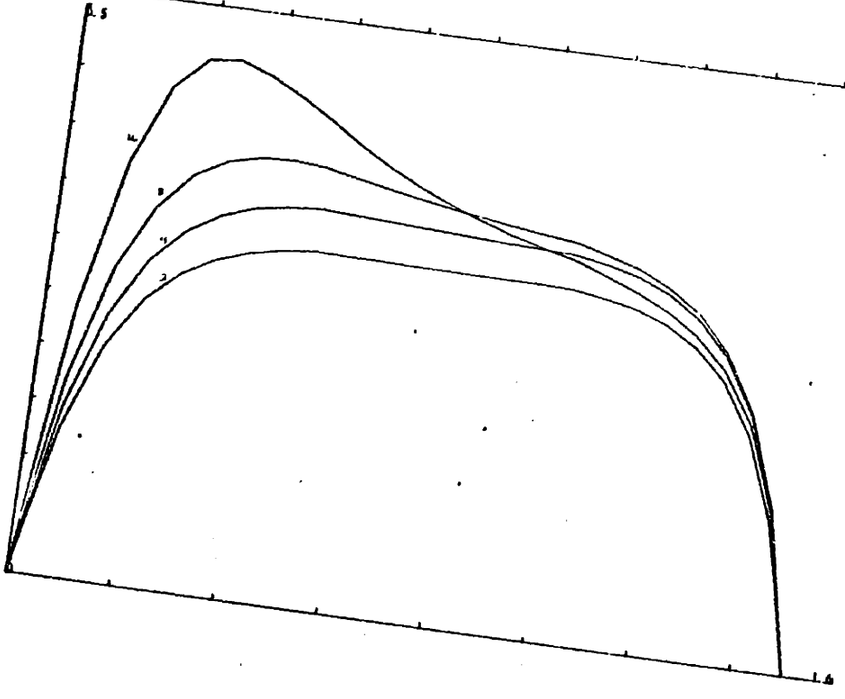


PL.4.3-1

l-type



m-type



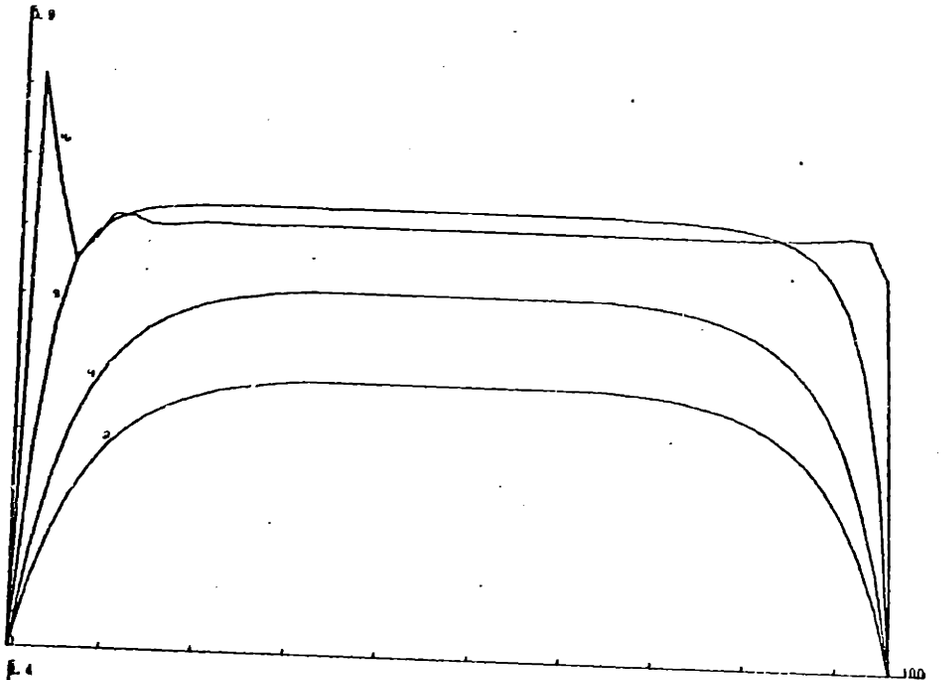
n-type

The evolution of the logic pulse for the stable spiral point parameterization, A3, is presented in PL. 4.3-2. The l-type pulse takes longer in x to converge to the critical point (than for A4) and once the edges start getting steeper the leading edge begins to ring (oscillate). The initial ring for the m-type pulse is larger but the amplitude still converges to the critical point value. The initial ring for the n-type pulse is larger still and the pulse disappears before the amplitude can recover to the critical point value.

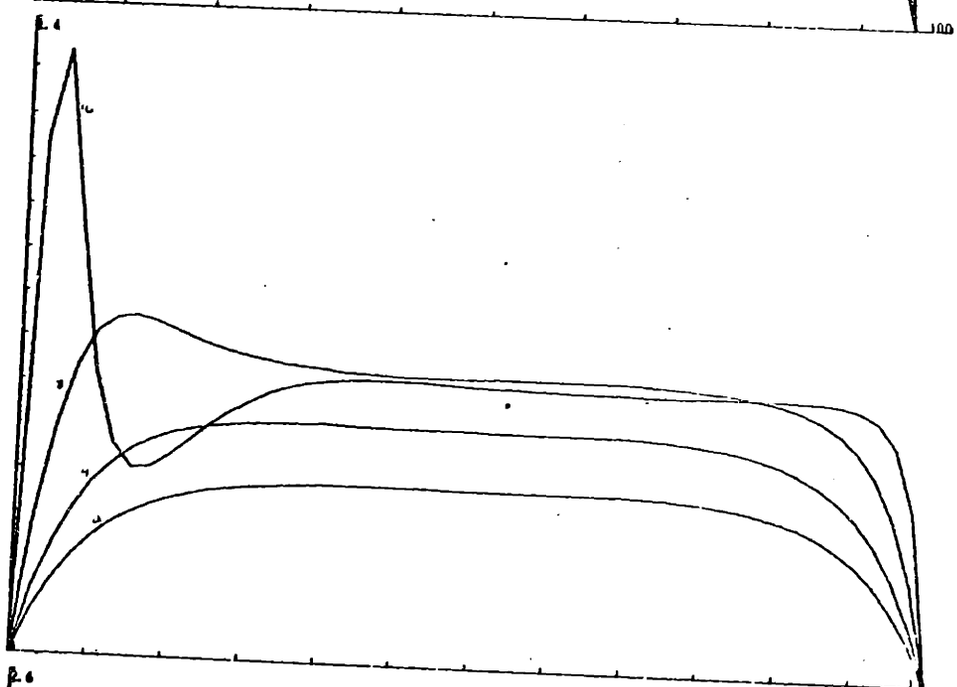
The hyperbolic secant pulse evolution for A4 is presented in PL.4.3-3. The l-type pulse exhibits some pulse height standardization and pulse edge sharpening. The m-type pulse has a leading edge that is unaffected but the trailing edge begins to "feel" the critical point influence. The n-type pulse is relatively unperturbed for these distances.

In PL.4.3-4 the evolution of the hyperbolic pulse for A3 is presented. The l-type pulse does exhibit substantial pulse height standardization and pulse edge sharpening at the expense of some oscillation. For the m-type pulse there is a substantial overshoot before a brief convergence to the critical point value. The n-type

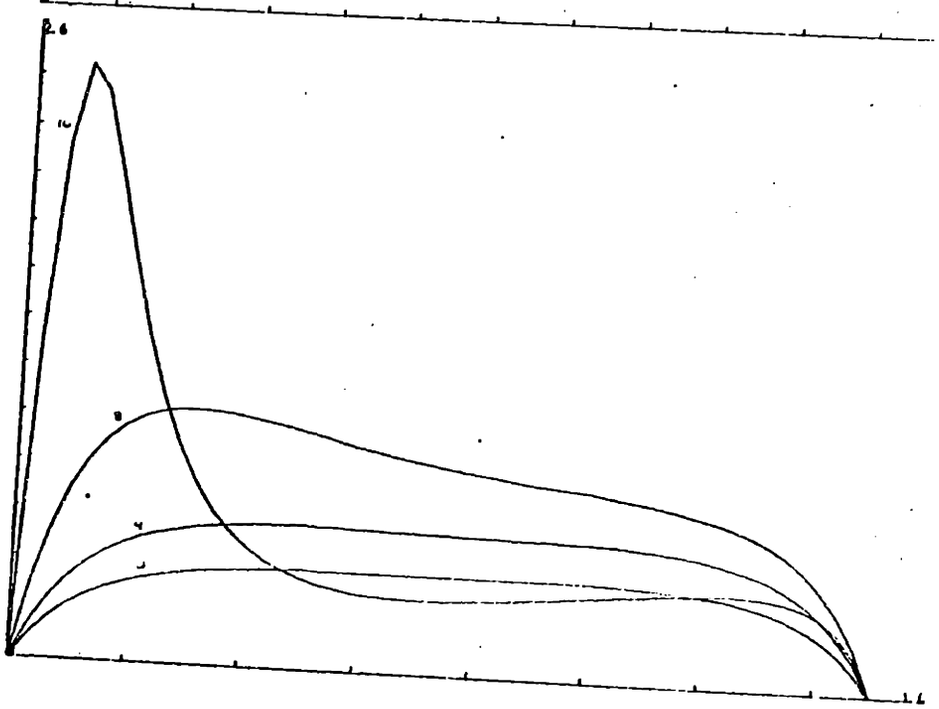
PL.4.3-2



l-type

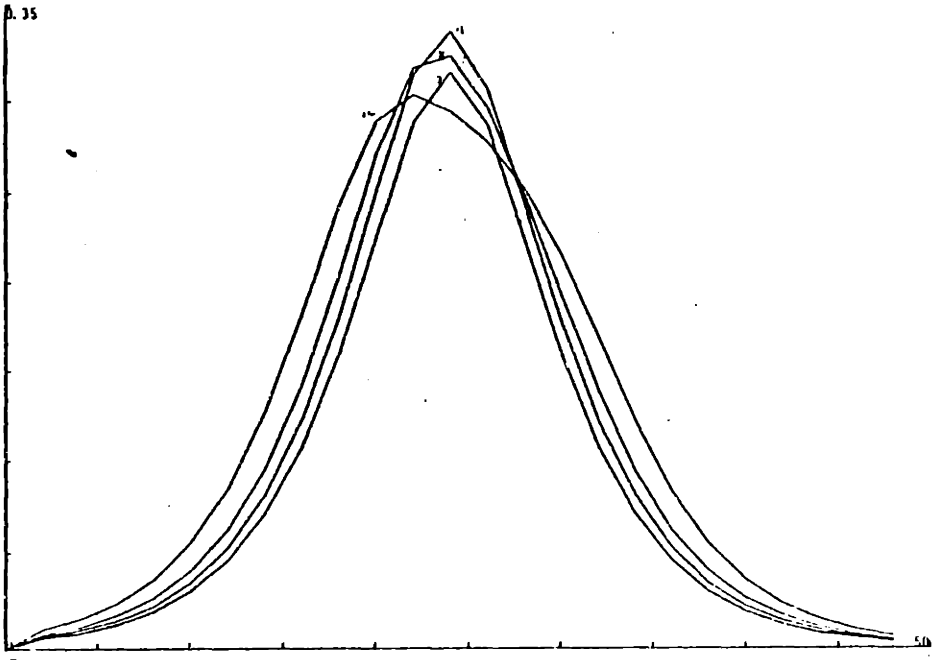


m-type

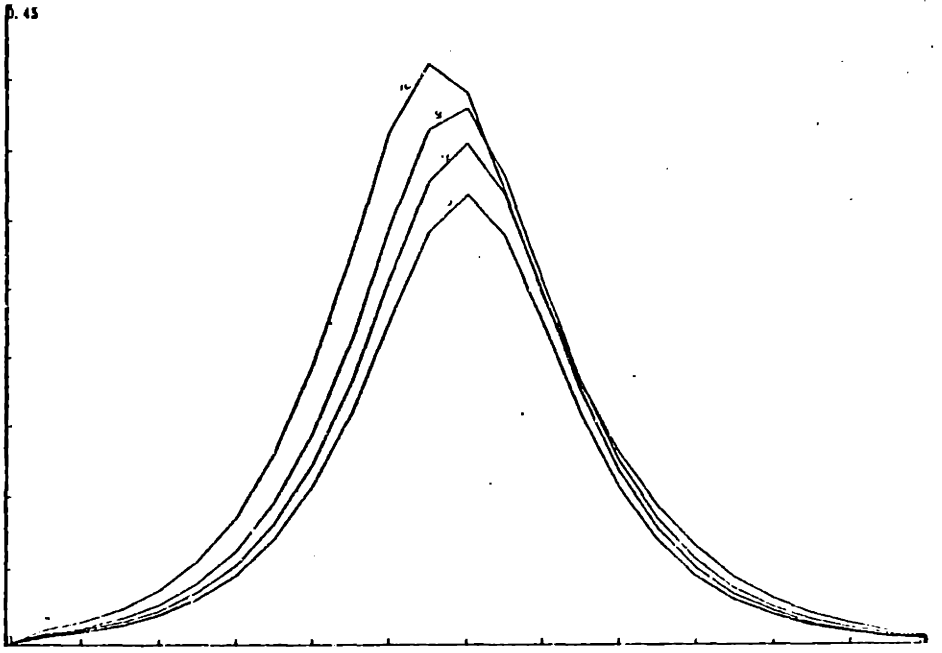


n-type

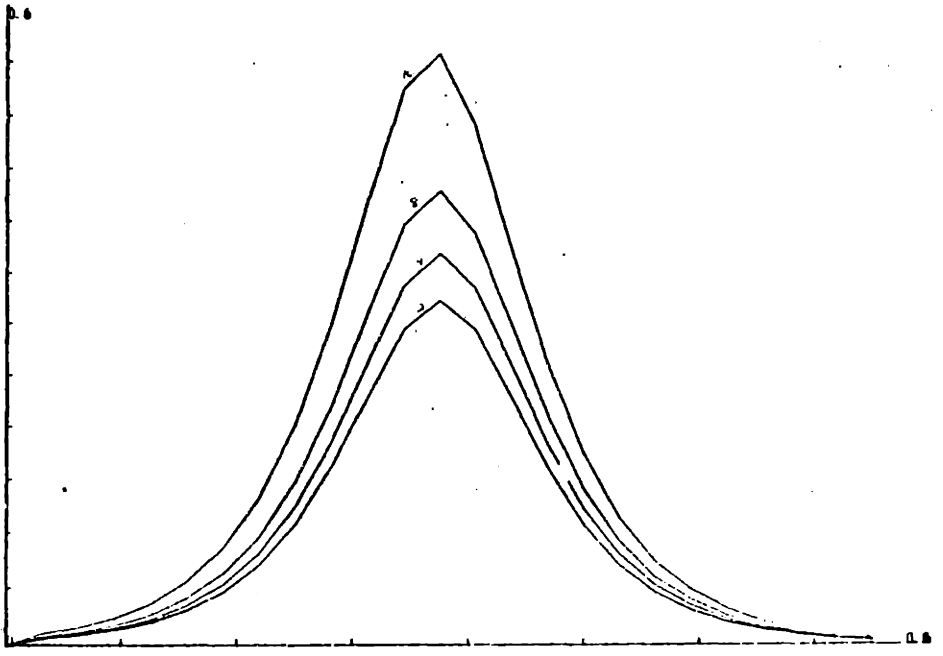
PL.4.3-3



l-type

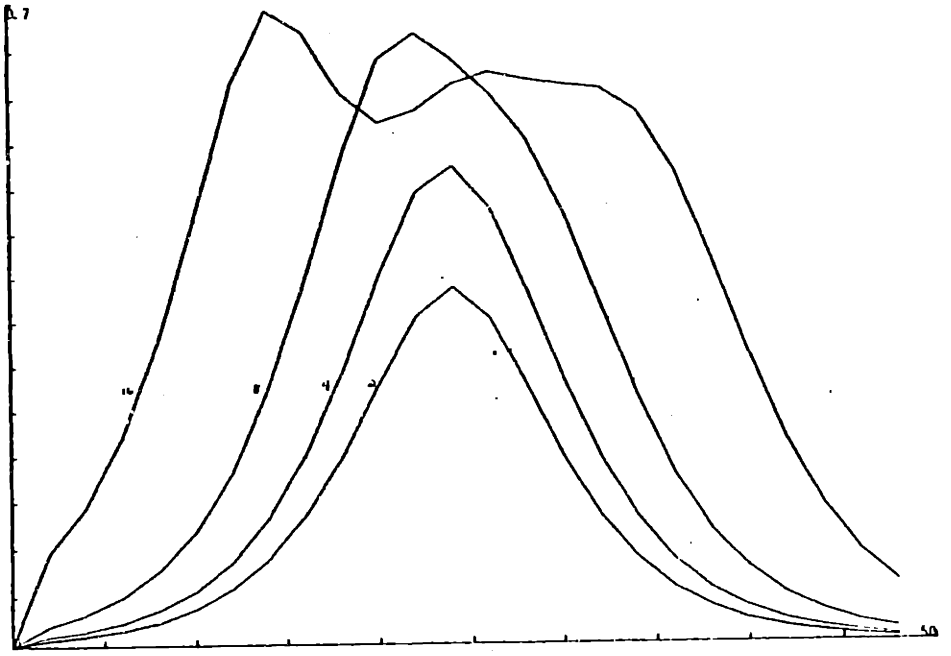


m-type

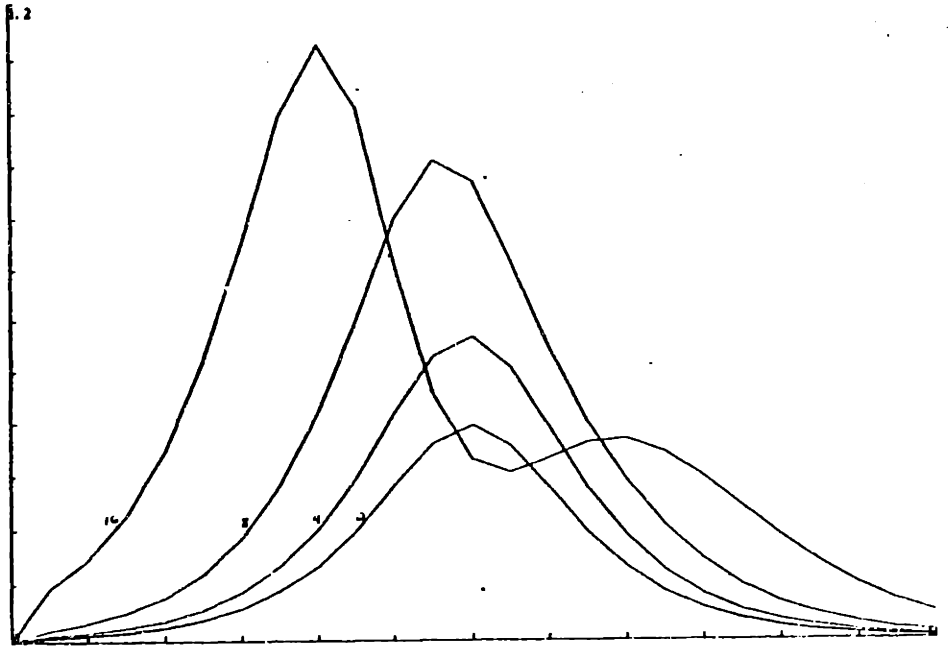


n-type

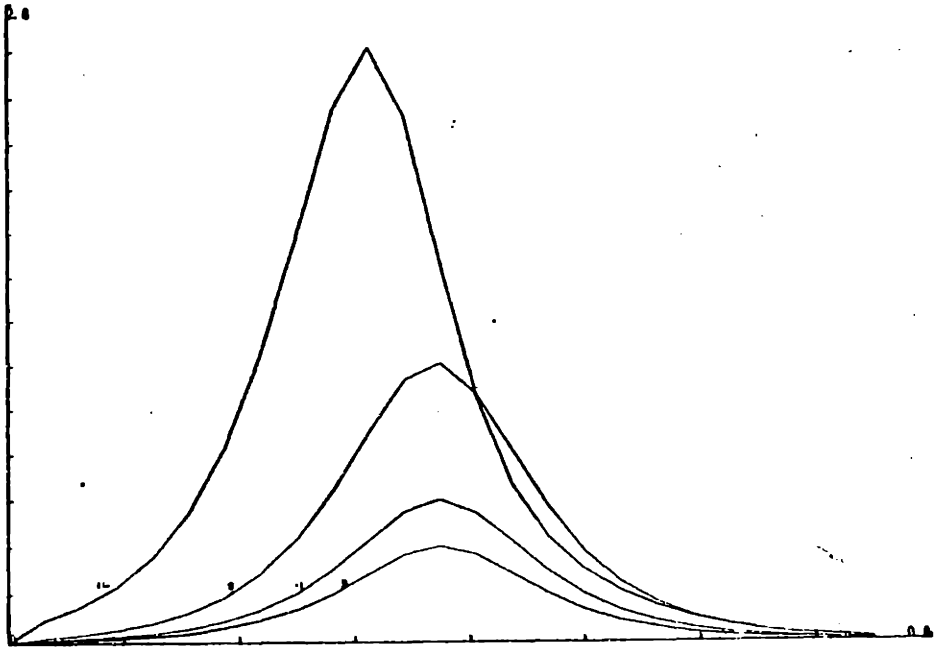
PL.4.3-4



l-type



m-type



n-type

pulse is again relatively unperturbed.

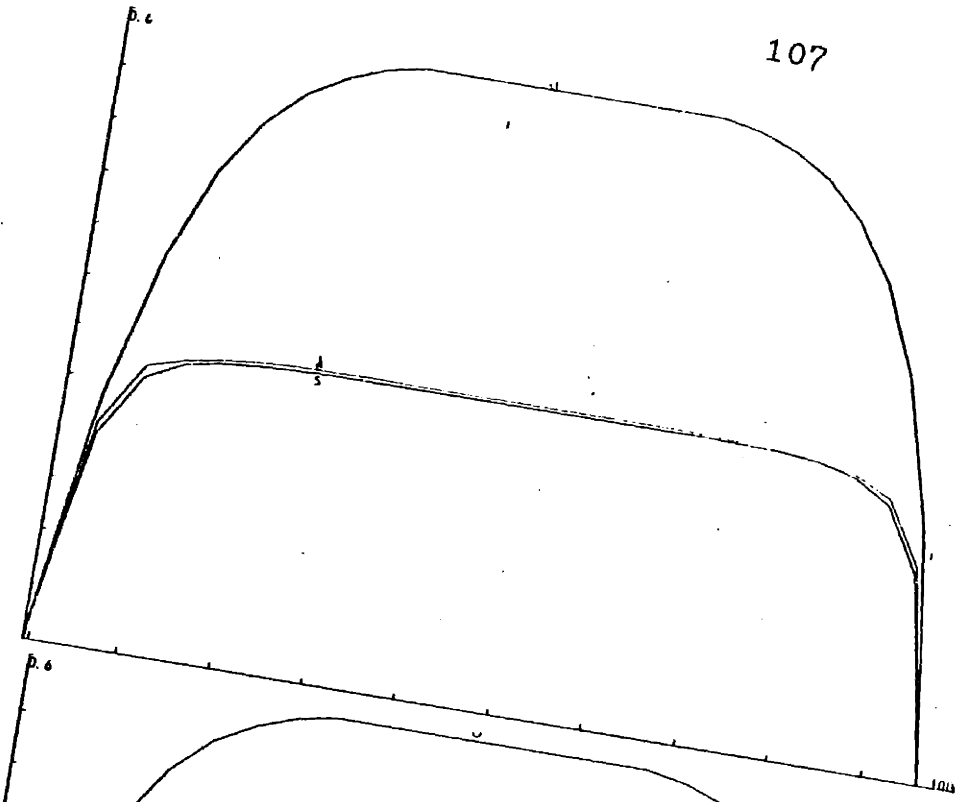
The second plot set demonstrates that the dynamic response for large pulse widths is essentially the steady state response and that for narrow pulse widths it is essentially the unperturbed response. For pulse widths in between these two limits the leading edge will tend to be unperturbed and the trailing edge will tend towards steady state. The differences between parameterizations A3 and A4 will become apparent.

In PL.4.3-5 the logic pulse for A4 at i = 16 is presented along with the unperturbed response, u , and the steady state response, s. For the l-type pulse steady state is approximately valid. For the m-type pulse the leading edge tends to be briefly unperturbed before approaching steady state. A larger amount of the leading edge of the n-type pulse is unperturbed and the pulse input decays before steady state is achieved.

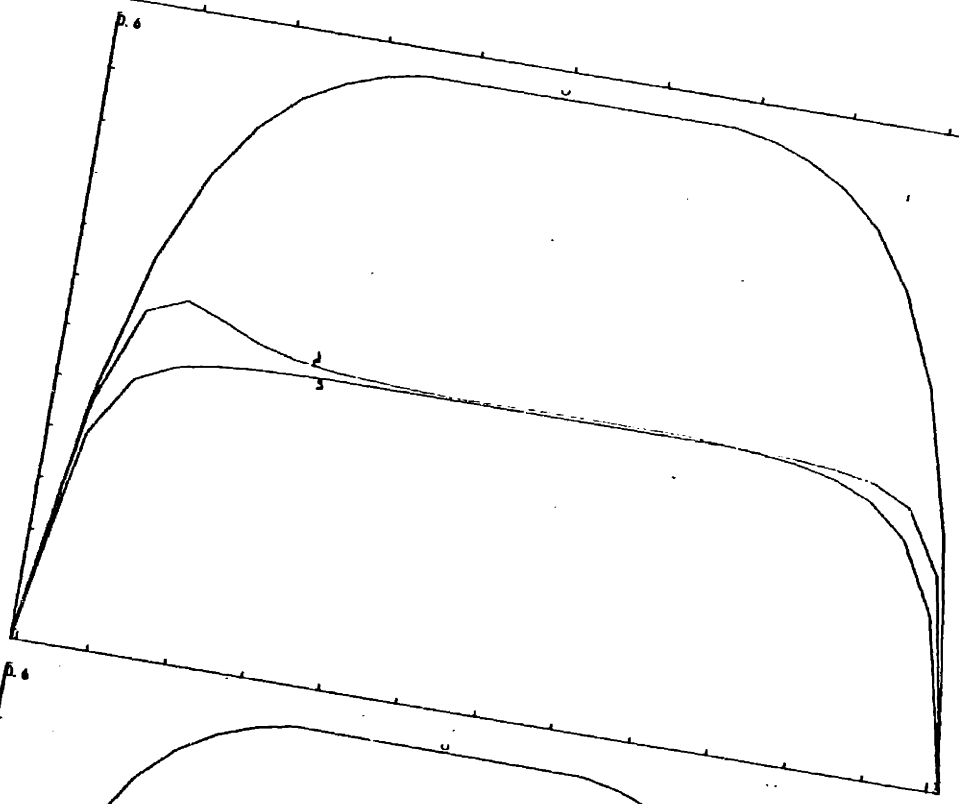
PL.4.3-6 presents similar results for the logic pulse response under parameterization A3. The difference is that the unperturbed response for A3 is larger in relation to the steady state response than it is for A4. The result being a wider standardized pulse height at the expense of substantial ringing.

PL.4-3-7 and PL.4.3-8 demonstrate these same conclusions for the hyperbolic secant pulses of A4 and A3

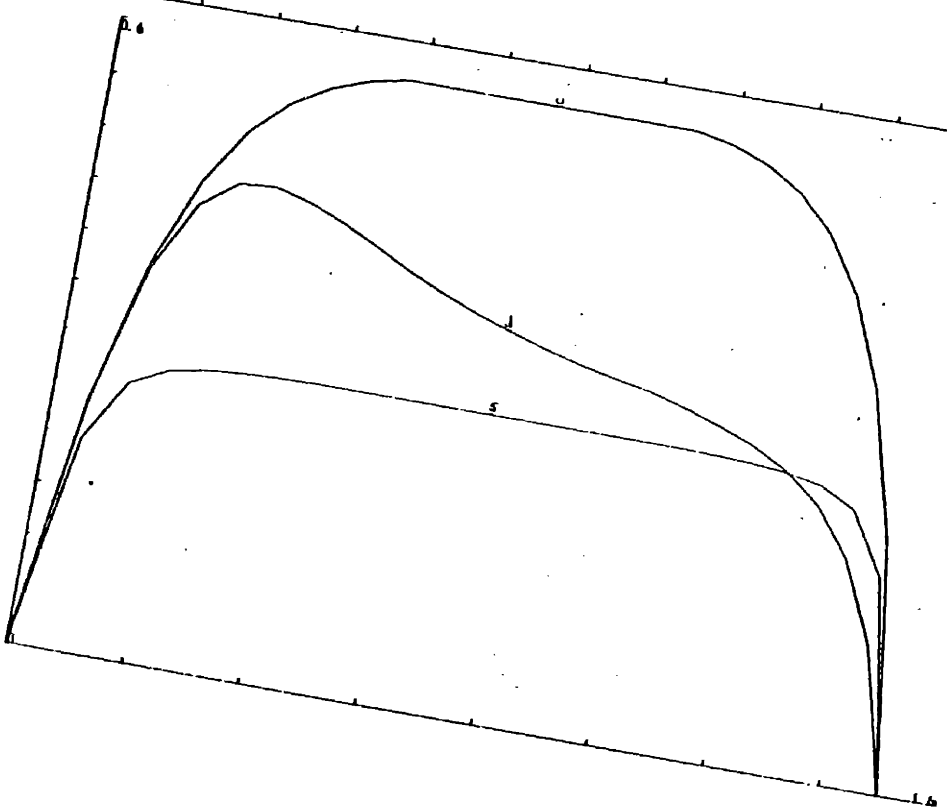
PL.4.3-5



l-type

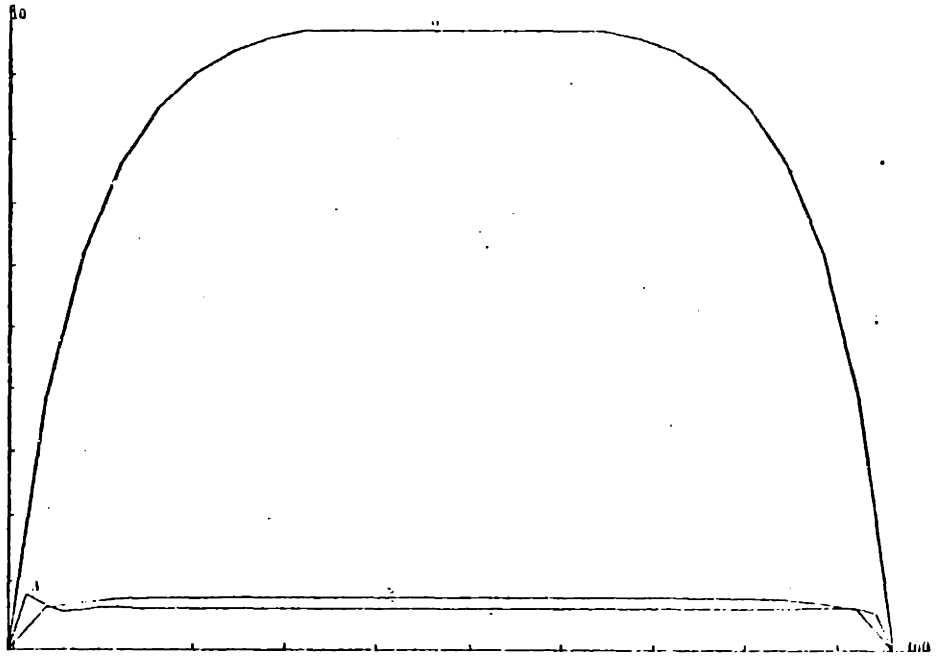


m-type

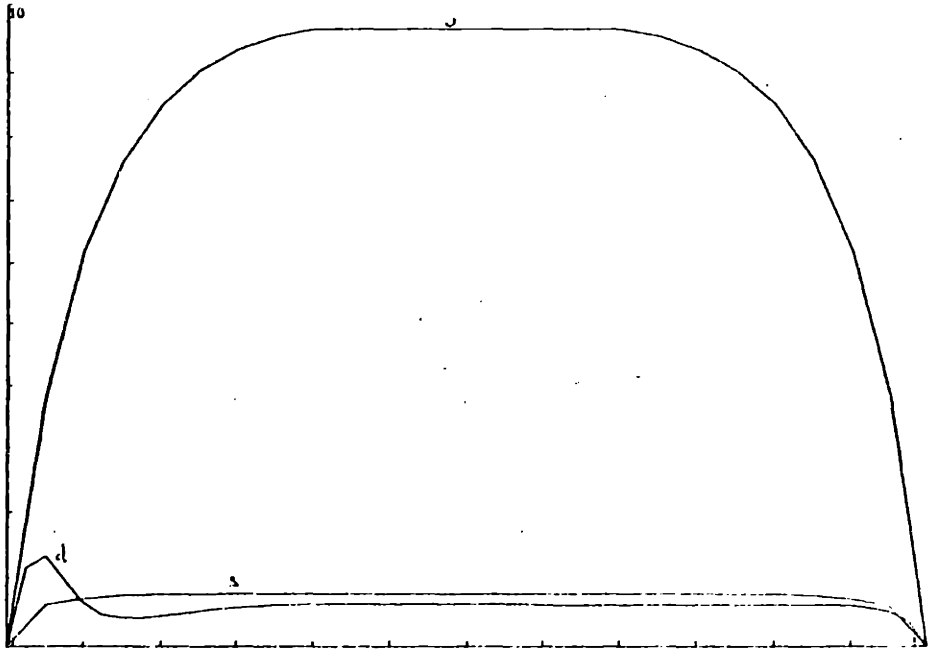


n-type

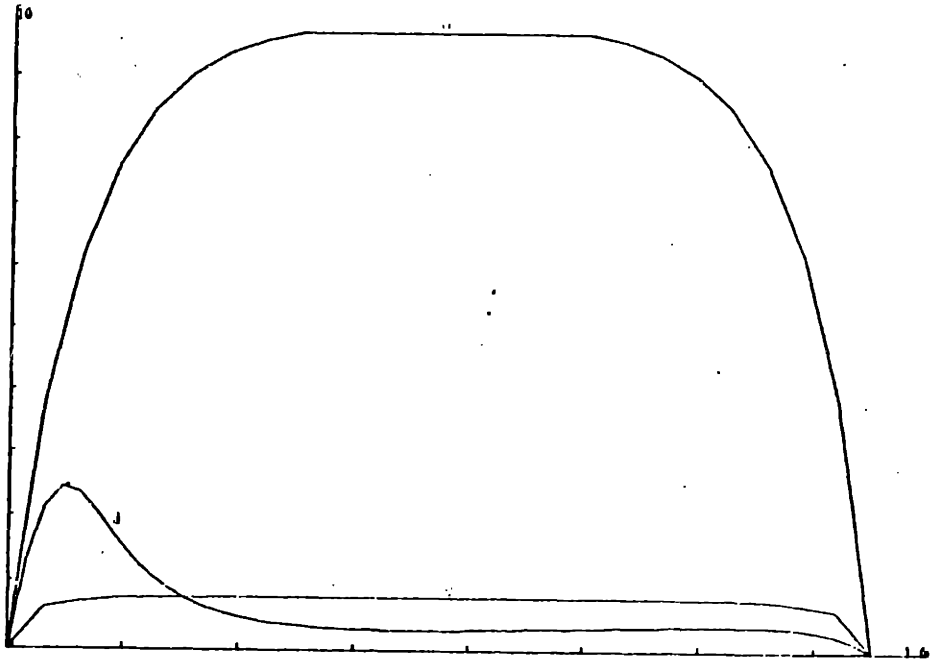
PL.4.3-6



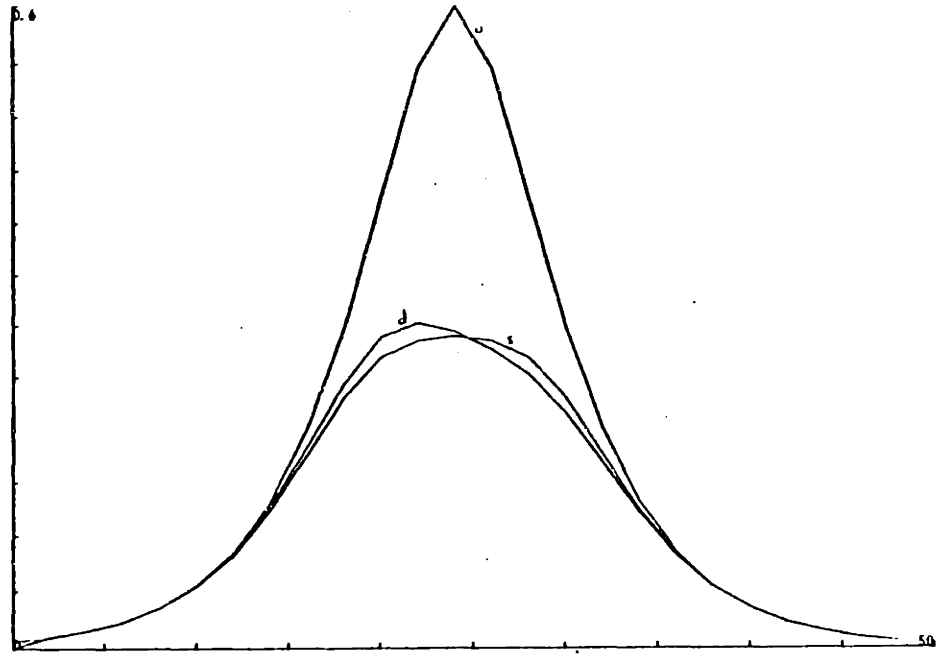
1-type



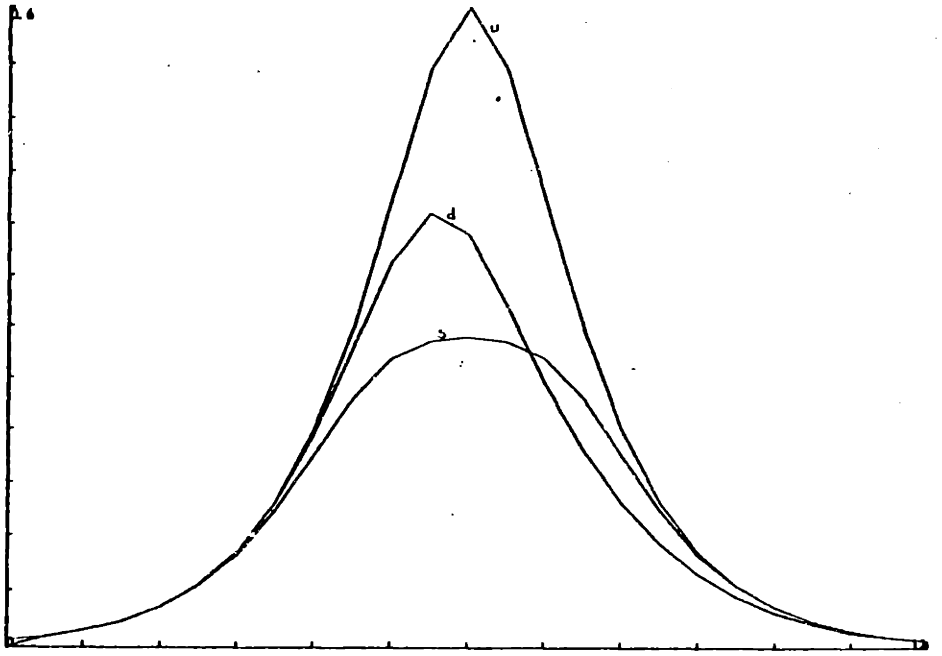
m-type



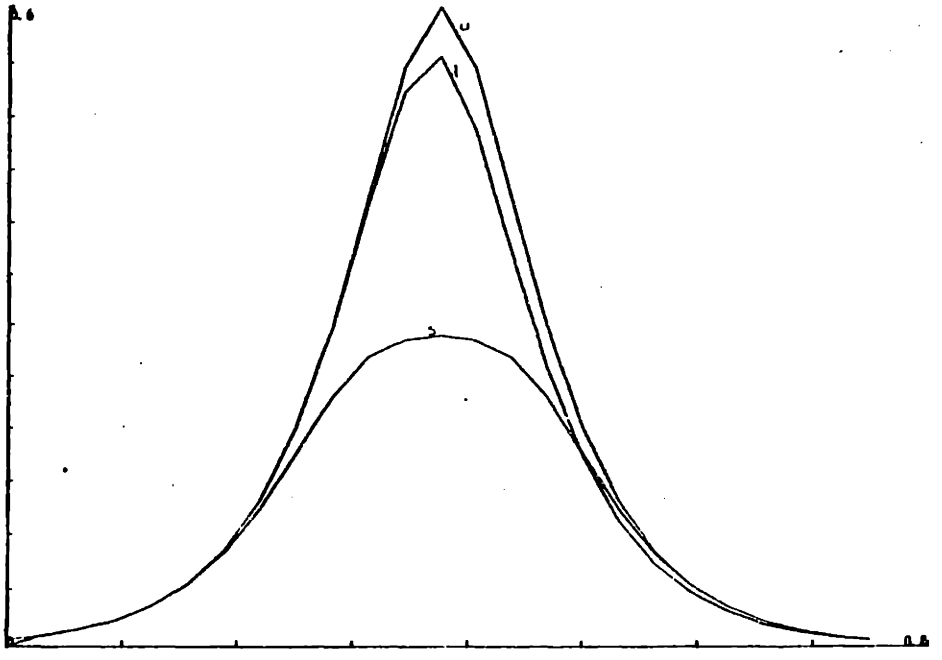
n-type



l-type

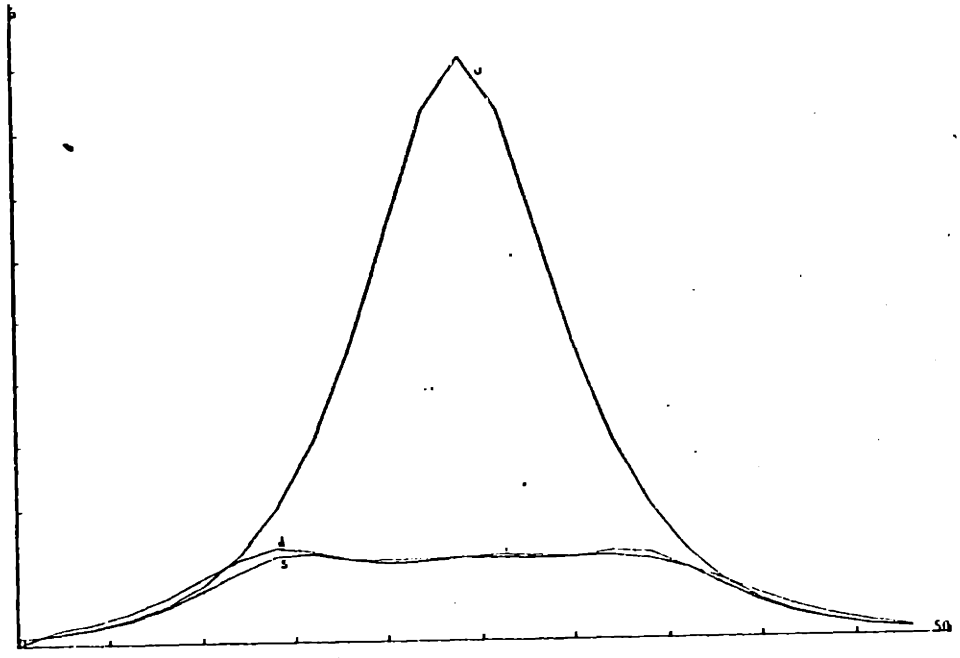


m-type

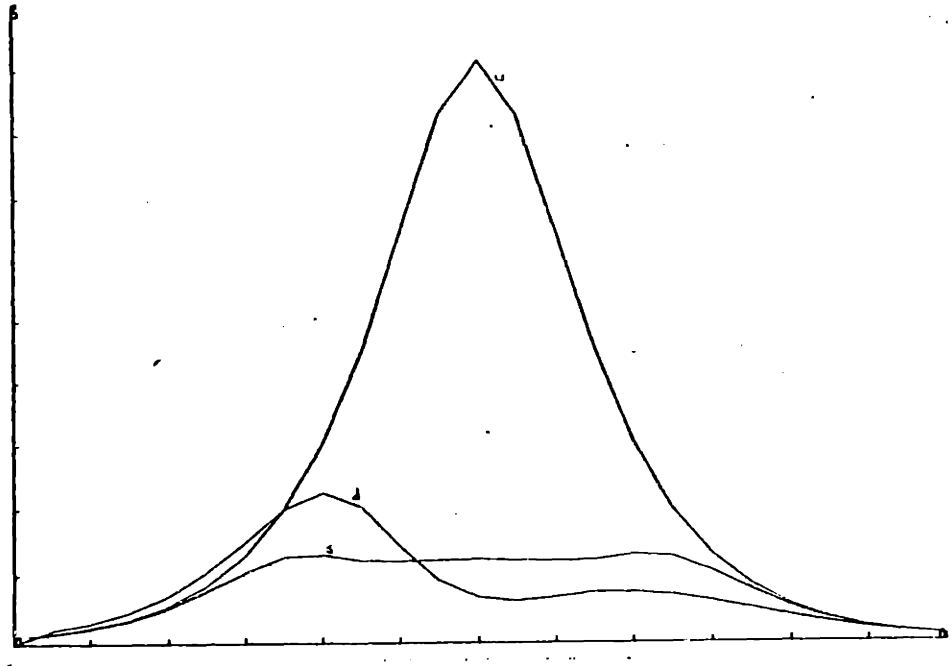


n-type

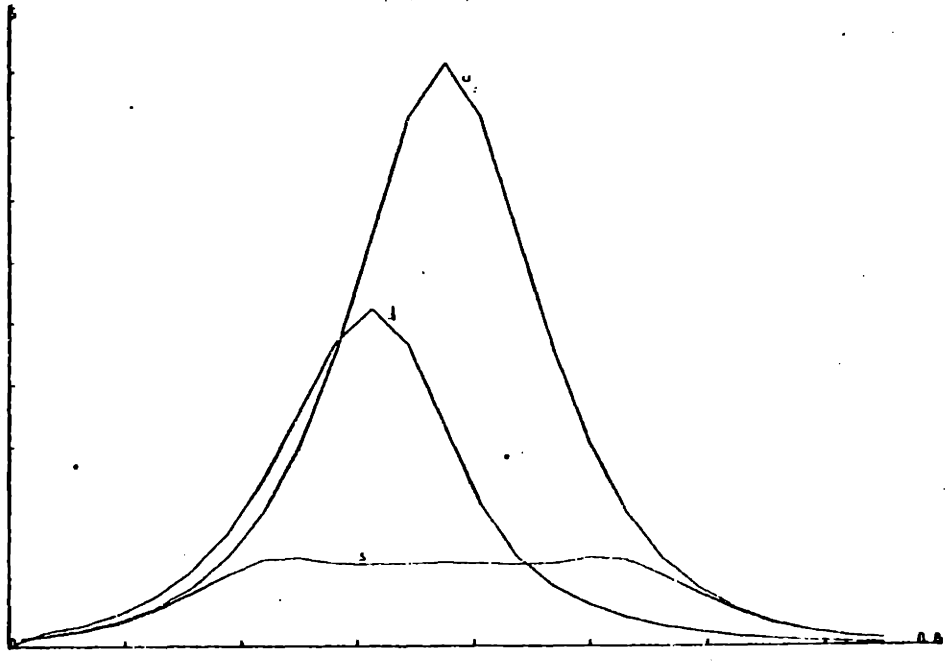
PL.4.3-8



l-type



m-type



n-type

respectively. Note that for A3 the leading edge of the dynamic response may be larger than the unperturbed response. This may also be shown by expanding the steady state response in a perturbation series in terms of The first term of the series is the unperturbed response and the summation of the infinite series is the steady state response. The dynamic response (in x) of any part of an input pulse may be approximated by the different orders of this perturbation series. For the "slightly perturbed" response it may be shown that the amplitudes may exceed those of the unperturbed response. This is physically reasonable since an initial perturbation of the loss should simply increase the effective gain.

4.4 Device Applications

Now that the systems dynamic response has been demonstrated, potential device applications may be more seriously considered. The applications of primary concern may be categorized as optical pulse shaping. Within this group two useful devices have already been demonstrated.

One of these is the optical pulse height standardizer (PHS) or optical limiter. System applications for such a device are seemingly endless. The most obvious way to construct a PHS is to exploit the convergence of mode amplitudes to a stable off-center critical point in the steady state limit. This approach shall be used in Chapter V which deals with the semi-optimal design of a realizable PHS. Non-critical point convergent schemes might also provide a means of obtaining a PHS. These methods would utilize the fact that in the steady state limit several different initial conditions may have trajectories that end (after a certain distance) at a "point" (or on a line tangential to one of the axes) in the phase plane without actually converging to a critical point. The usefulness of this method in the dynamic case, however, has not been demonstrated and will therefore not be considered further in this thesis.

The other useful device that has been demonstrated may be termed a pulse width shortener (PWS). These devices

could be used to shorten the width of an input pulse to time intervals substantially smaller than the medium response times. Since this phenomenum depends on the ringing of a pulses unperturbed leading edge, a practical PWS device would probably use a PHS for an input stage in order to steepen the edges of the input pulse before being processed by the PWS itself. Device lengths for the PWS could be kept small by using this technique. It is interesting to note that the generation of optical pulses with widths smaller than the medium response times in, for example, a laser diode has usually been attributed to the combined effects of loss and gain saturation. For the saturable loss coupled travelling wave structure, however, it has been shown that saturation effects of the loss medium alone may produce this phenomenum. Gain saturation for such a device ($PS_{(G)} > PS_{(L)}$) would tend to limit the height of the shortened pulse.

Other pulse shaping applications may be referred to as general filtering. These could be realized by a cascade of several PHSs and PWSs, each designed to perform a specific part of the required pulse shaping. One method of fabricating these cascades in a continuous fashion would be to induce a prescribed spatial variation in k and/or l_o and g_o. The spatial variance of k could be produced either by varying the refractive index of the separating

medium (by some method of impurity implantation) or by varying the distance between waveguides.

Other potential applications may be classified as general signal processing devices. These include single and double input amplifiers (i.e., transistors and differential amplifiers). Logic devices could conceivably be made also, either in a direct fashion or by appropriate interconnection of optical amplifiers. Gated mode amplitude oscillators that exploit the global limit cycles of the structure could also be made. The structure might also provide an easy way of obtaining a highly stable light wave amplifier or a highly stable laser. The only application that will be considered and developed in detail here will be the PHS device which is dealt with in the next chapter.

parameter pairs. Some of these pairs are $(\ell_0/k, g_0/k)$; $(g_0/\ell_0, \ell_0/k)$; $(\ell_0/g_0, k/g_0)$; and $(k^2/(g_0\ell_0), g_0/\ell_0)$ etc. In keeping with the fashion in which steady s ate stability criteria were presented, the following definitions are made: $\gamma \equiv k^2/(g_0\ell_0)$, $\beta \equiv g_0/\ell_0$. In terms of these parameters the steady state equations for the saturable loss device become:

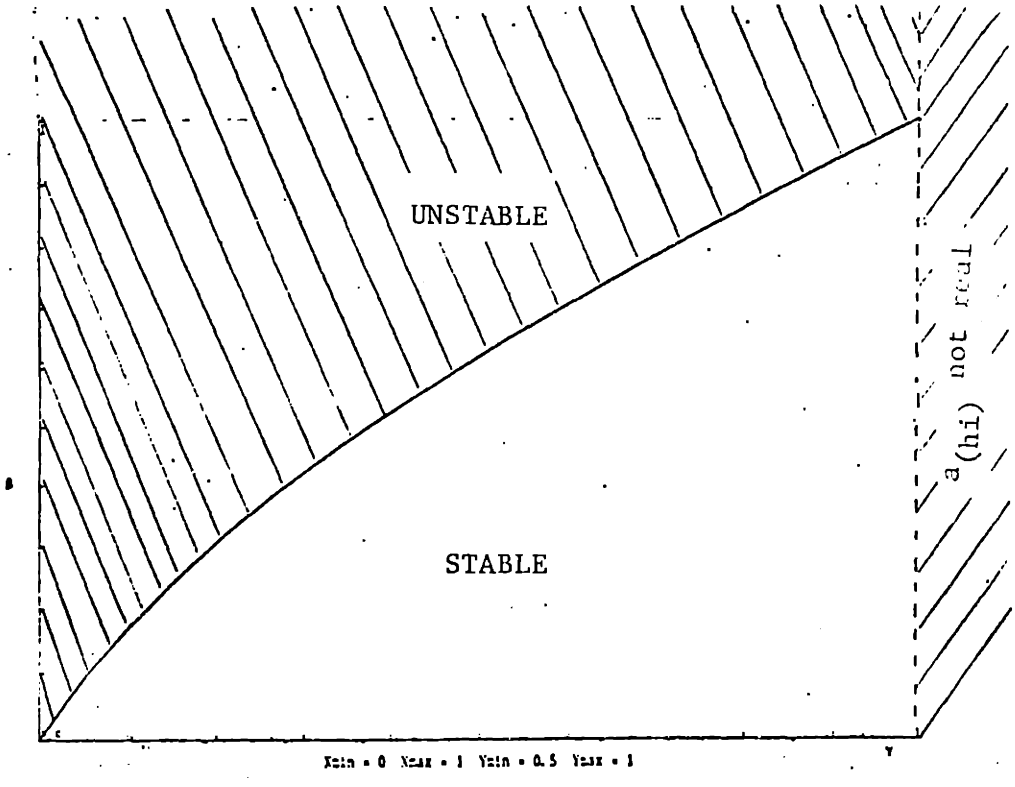
$$\frac{d\ a1}{d\eta} = \sqrt{\beta/\gamma}\ a1 + a2$$

$$\frac{d\ a2}{d\eta} = \frac{-a2}{\sqrt{\gamma\beta}(1 + a2^2)} - k\ a1$$

where $\eta \equiv kx$. The criteria for the stability of the off-center critical points are:

$$\beta < 1 \ ; \qquad \gamma < \frac{1}{4}\left(1 + \sqrt{1 + 8\beta}\right).$$

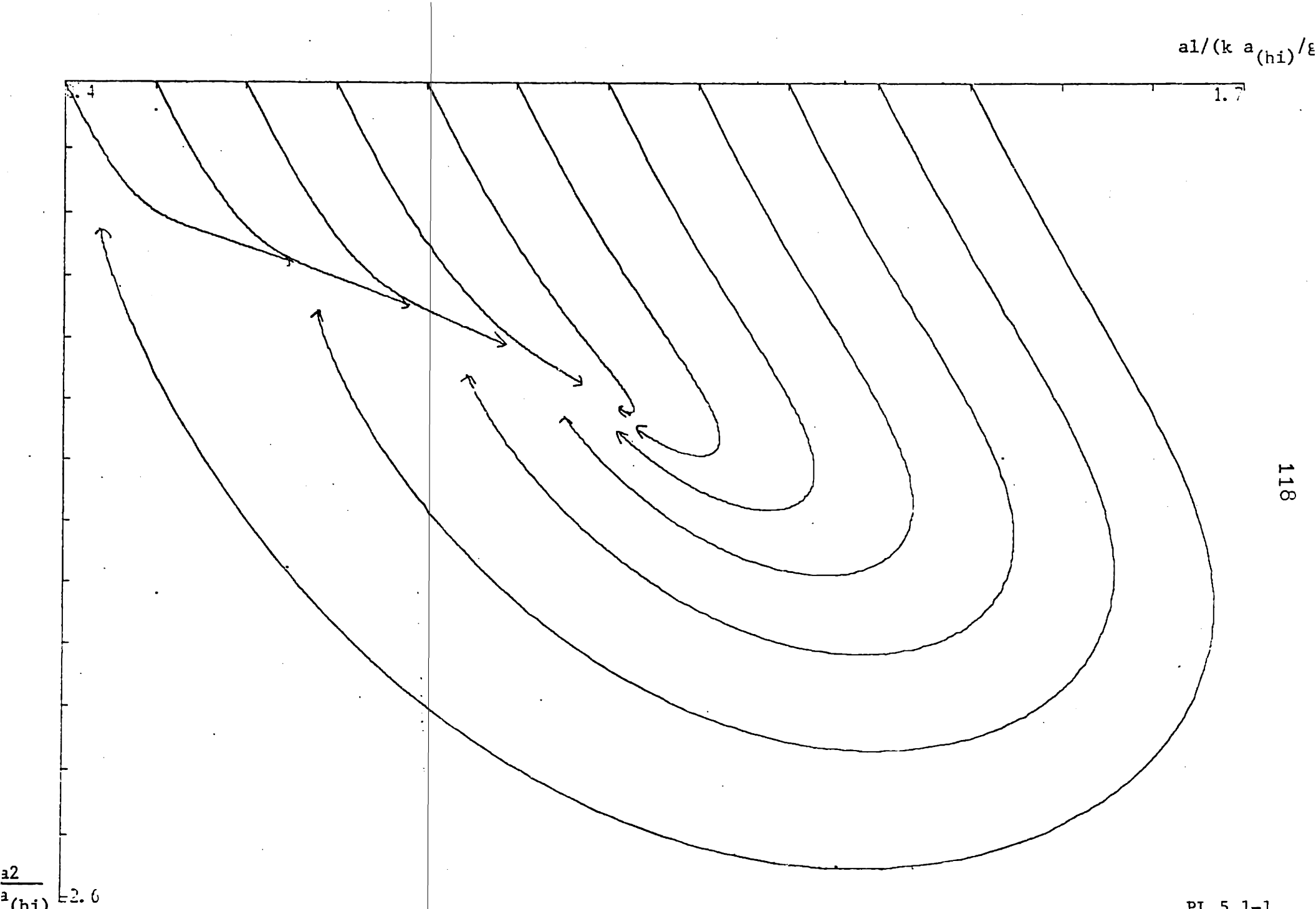
For the off-center critical point to be real requires $\gamma < 1$. These statements are displayed in the γ-β plane of figure 13.



The stable region was searched for the parameterization which yields the largest output (either a1 or a2) signal to noise ratio when the input amplitude varies uniformly over its optimal 2 to 1 range. Thus the input signal to noise ratio is 14.31dB.

The output signal to noise ratio (at kx= kd=5) is computed for various 2:1 input ranges for each parameterization trial point in the stable region of the γ-β plane. The optimal parameterization chosen is γ =.82 , β=.15, for which the a1 output signal to noise ratio is 23.3dB when the input varies uniformly from $.6\frac{k}{g_0} a_{(hi)}$ to $1.2\frac{k}{g_0} a_{(hi)}$. Output signal to noise ratios of 24.5dB are achievable for other γ,β points but this one was chosen since the output signal to noise ratio near γ =.82, β=.15 is essentially constant over a larger region of the γ-β plane. Output signal to noise ratios of 30dB can be achieved for "non-critical point convergent" schemes where the outputs are much larger than the critical point values. The dynamic response of these non-convergent schemes, however, is inferior.

PL.5.1-1 presents the phase plane trajectories (up to kx = kd = 5) for the parameterization: γ =.82 , β =.15 for the inputs: a1 = .4, .5, .6, .7, .8, .9, 1.0, 1.1, 1.2, 1.3, 1.4, in normalized coordinates



Xmin = 0.4 Xmax = 1.7 Ymin = -2.6 Ymax = 0

PL.5.1-1

(a1 normalized to $\frac{k}{g}\, a_{(hi)}$, a2 normalized to $a_{(hi)}$). Notice that the output value of a1 is the same for inputs that deviate from 1 (or a little less than 1) by the same amount. Thus for an input which has small fluctuations around three amplitudes: A, 2A, and 0 (like the output of a two input interferometer operating on binary data), the a1 output of the PHS would be tightly centered around some value for the A and 2A inputs and tightly centered around 0 for the 0 input. Operation of the PHS in this fashion would make feasible the use of interferometers as exclusive--or logic gates from which entire digital computers could be constructed. The speed of such a logic gate would be limited only by the propagation delay corresponding to the interferometer and PHS device lengths.

5.2 Effects of Pulsewidth on Performance of Optimal PHS

This section presents dynamic response information for the optimal PHS of the previous section. Since this PHS was optimized for the steady state limit it is expected that the operation will deteriorate for small pulse widths. This does not, however, mean that a faster PHS could not be achieved with the traveling wave device under a different parameterization.

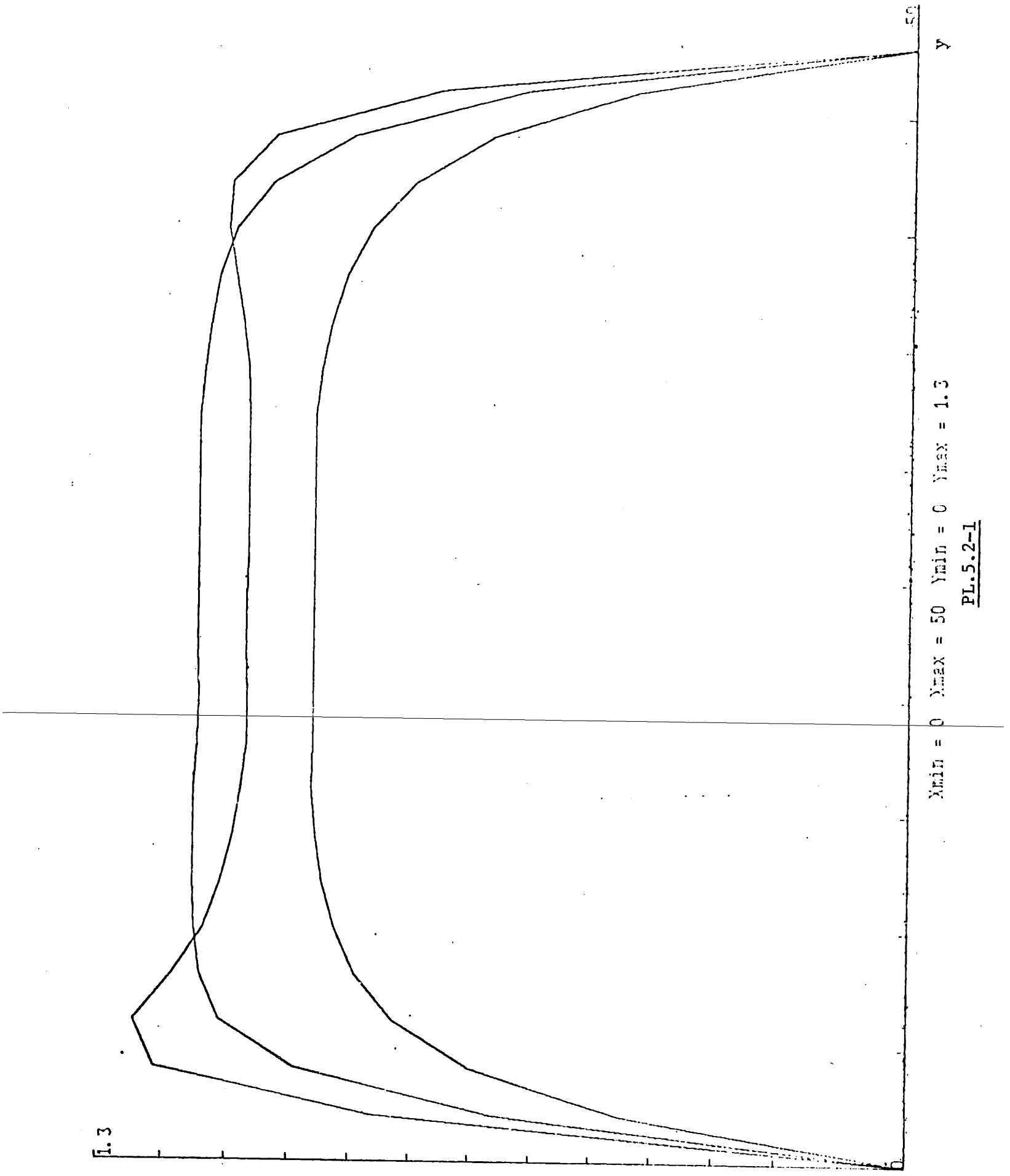
The waveform used to test the dynamic response is the logic pulse as defined in 4.3. The exponential rise and fall times, τ , used are: 4; 1; and 1/2. For each of these rise times (or pulsewidths, since pulsewidth = 4 τ) the input amplitudes: al = .6, al = .9, al = 1.2 (normalized coordinates) are tested.

PL.5.2-1 presents the dynamic response at the device output for the three input amplitudes when $\tau = 4$. Notice that there is some overshoot for the largest input amplitude.

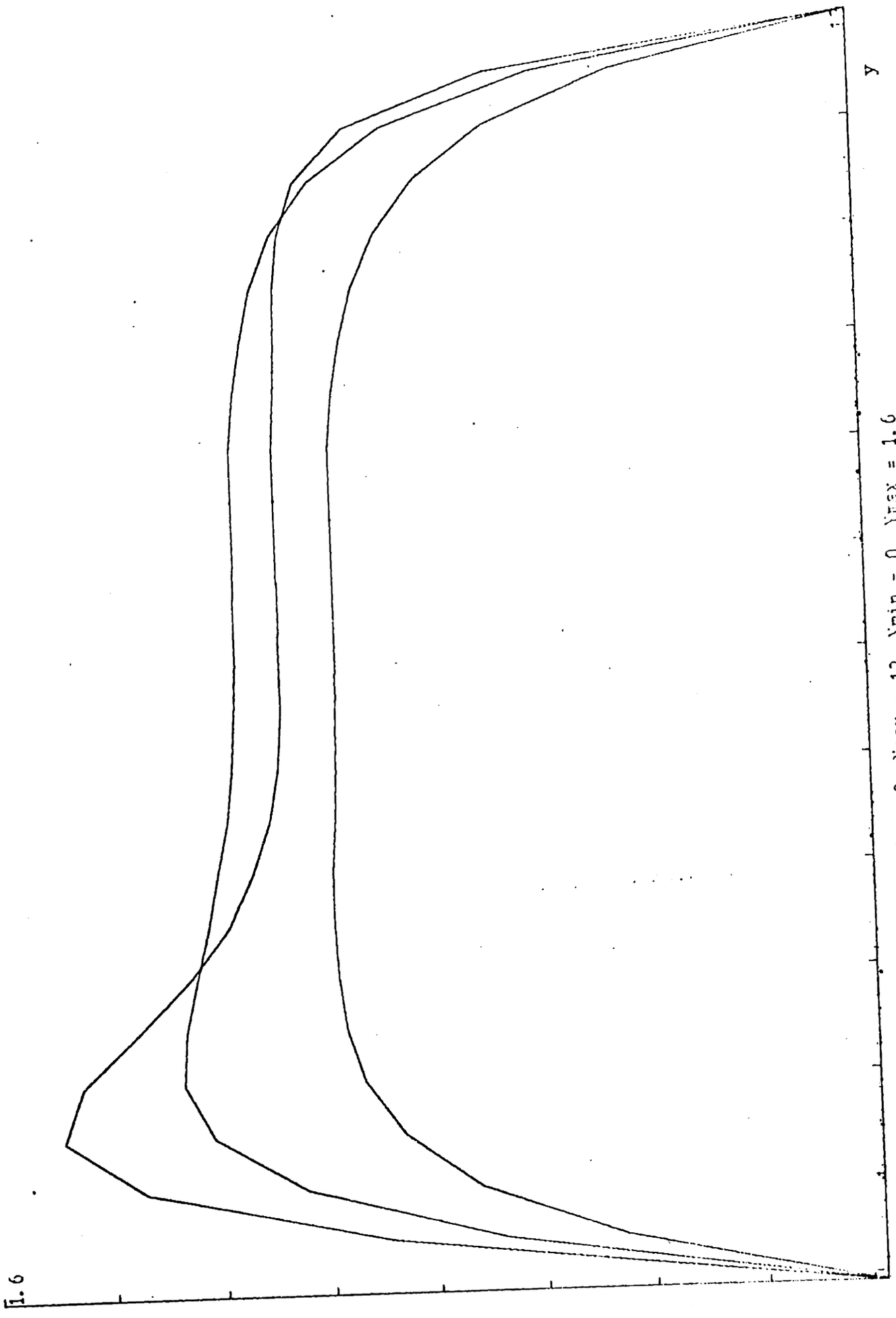
PL.5.2-2 presents the dynamic response from the three input amplitudes when $\tau = 1$. Pulse height standardization is still achieved although waveform fidelity has deteriorated somewhat (predominantly for the largest input amplitude response).

PL.5.2-3 presents the dynamic response for the three input amplitudes when $\tau = 1/2$. The PHS is still performing its basic function but the pulse shapes are deteriorated further with a substantial overshoot for the response from the input amplitude of 1.2.

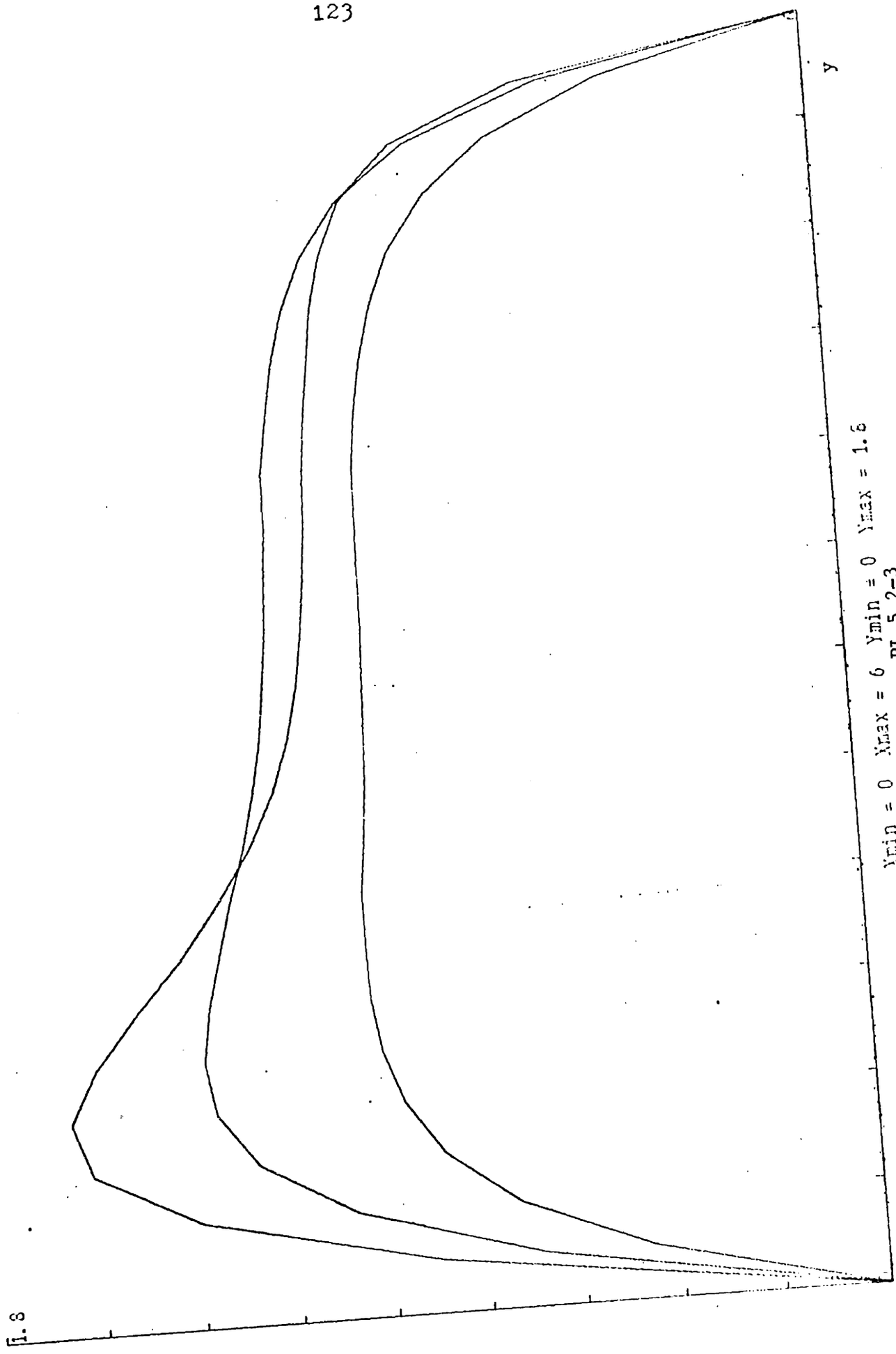
If a fast PHS is desired one could try optimizing and for the dynamic response. Another method might be to predistort the pulse with a device of some γ and β such that the pulse edges are slowed before inputting



PL.5.2-1



PL.5.2-2



PL.5.2-3

the waveform to the PHS. Since, however, in a system (for either computer or communication applications) the data is sampled at the center of the pulse, initial overshoots are not as critical as one might expect.

CHAPTER VI - Suggestions for Future Research

There are a number of interesting problems remaining that relate to the proposed travelling wave pulse shaping device and similar devices. A multimode model could be easily formulated to describe the coupling of two multimode waveguides or several single mode waveguides. This model could be used to explore the possibility of fabricating mode selectors/converters as suggested in 3.6.

A more detailed analysis of the dynamic response would be a fruitful endeavor. Comparisons of the saturable loss response with the responses of the saturable gain and dual saturation cases would be interesting. It would also be useful to examine the effects of different $T1_{(L)}$, $T1_{(g)}$, $PS_{(L)}$, and $PS_{(g)}$ for the dual saturation case. This could possibly provide some insight into the mechanisms of picosecond pulse generation in laser diodes as well as lead to more pulse shaping applications as mentioned in 4.4. This detailed study of the dynamic response would probably involve characterization of various input waveforms and might view the dynamic response as a perturbation of the steady state case. The various orders of this perturbation series have closed form representations (available from the author) which have been shown to be useful in describing the unperturbed, the slightly perturbed, and the totally perturbed (i.e. steady state) cases. With this analytic

handle on might be able to optimize the design of PHSs and PWSs, etc., for the dynamic case. The existence of solitons might be examined by a asympotically expanding an integral formulation of the equations although the $kd \leq 5$ constraint would probably render this study to be an academic one.

The $kd \leq 5$ constraint should be more carefully evaluated. Improvements (larger upper bound) of this restriction would certainly be useful.

General pulse shaping by a device with a spatially variant coupling constant would be a challenging and interesting problem. One would attempt to be able to solve for the required $k(z)$ to produce a given output waveform from a given input waveform.

"Non-critical point convergent" applications such as amplifiers and oscillators should also be investigated.

Phase state logic and decision circuits (amplitude comparators) may prove to be fabricable from the pulse shaping structure.

PHSs could also be optimized for specific input distributions for different system applications as mentioned in 5.1.

Bibliography

[1] A. Szoke et.al., "Bistable Optical Element and its Applications," Appl. Phys. Lett., vol. 15, pp. 376-379, Dec. 1969.

[2] R.H. Pantell and H.E. Puthoff, Fundamentals of Quantum Electronics. New York: Wiley, 1969.

[3] M.S. Sargent, M.O. Skully, and W.E. Lamb, Laser Physics. Reading, MA: Addison-Wesley, 1974.

[4] Nonlinear Optics; proceedings. Scottish Universities Summer School, 16th, 1975, Heriot-Watt University: Academic, 1977.

[5] E. Spiller, "Saturable Optical Resonator," J. Appl. Phys., vol. 43, pp. 1673-1681, April 1972.

[6] S.L. McCall, "Instabilities in Continuous-wave Light Propagation in Absorbing Media," Phys. Rev. A, vol. 9, pp. 1515-1523, April 1974.

[7] S.L. McCall, "Optical Transistor and Bistability", J. Opt. Soc. Am. Vol. 65, Oct. 1975.

[8] S.L. McCall, "Self-Induced Transparency", Phys. Rev. Vol. 183, July 1969.

[9] D.M. Pepper, "Observation of Mirrorless Optical Bistability and Optical Limiting Using Stark Turable Gases," I.E.E.E. J. Quantum Electron., vol. QE-15, pp. 1362-1368, Dec. 1979.

[10] V.N. Lugovoi, "On the Theory of a Nonlinear Optical Resonator," Opt. Acta, vol. 24, pp. 743-756, 1977.

[11] F.S. Felber and J.H. Marburger, "Theory of Nonresonant Multistable Optical Devices," Appl. Phys. Lett., vol. 28, pp. 731-753, June 1976.

[12] T. Bischofberger and Y.R. Shen, "Transient Behavior of a Nonlinear Fabry-Perot," Appl. Phys. Lett., vol. 32, pp. 156-158, Feb. 1978..

Bibliography (cont.)

[13] J.A.Goldstone and E.M. Garmire, "On the Dynamic Response of Nonlinear Fabry-Perot Interferometers," I.E.E.E. J. Quantum Electron., vol. QE-17, March 1978.

[14] M. Okuda et.al., "Saturable Optical Resonators with Distributed Bragg-Reflectors," Opt. Commun., vol. 19, Oct. 1976

[15] M. Okuda and K. Onaka, "Bistability of Optical Resonators with Distributed Bragg-Reflectors by Using the Kerr Effect," Jap. J. Appl. Phys., vol. 16, pp. 769-773, Nov. 1976.

[16] R.W. Gray and L.W. Casperson, "Optooptic Modulation Based on Gain Saturation," I.E.E.E. J. Quantum Electron., vol. QE-14, pp. 893-900, Nov. 1978.

[17] K. Otsuka, "Proposals and Analyses on Laser Amplified Based Integrated Optical Circuits, "I.E.E.E. J. Quantum Electron., vol. QE-17, pp. 23-28, Jan. 1981.

[18] A. Yariv, "Coupled Mode Theory for Guided-Wave Optics," I.E.E.E. J. Quantum Electron, vol. QE-9, pp. 919-933, Sept. 1973.

[19] R. Bogen, MACSYMA Reference Manual. Cambridge, Massachusetts Institute of Technology, 1977.

[20] C.M. Bender and S.A. Orzag, Advanced Mathematical Methods for Scientists and Engineers. New York: McGraw-Hill, 1978.

[21] W.J. Cunningham, Nonlinear Analysis. New York: McGraw-Hill, 1958.

[22] W.F. Ames, Numerical Methods for Partial Differential Equations, New York: Academic Press, 1977.