Unit 3: Effects arising from the third-order nonlinear susceptibility

Nicholas Rivera

(Dated: March 13, 2026)

Contents

In this chapter, we will study a handful of important effects arising $\chi^{(3)}$. In the previous chapter, we focused on various frequency-conversion effects such as harmonic generation, sum- and difference-frequency generation, and electro-optic modulation. While many of these effects can also be realized by $\chi^{(3)}$, we will focus on the most unique effects. At the core of these effects is the idea of an *intensity-dependent index of refraction* [1].

I. NONLINEAR INDEX OF REFRACTION

Consider a third-order nonlinear medium. The relationship between electric field and polarization, in frequency-domain, can be written as

$$P_i(\mathbf{r},\omega) = \int \frac{d\omega_1}{2\pi}\frac{d\omega_2}{2\pi}\frac{d\omega_3}{2\pi}\delta(\omega-\omega_\sigma)\chi^{(3)}(\omega_\sigma;\omega_1,\omega_2,\omega_3)E_j(\omega_1)E_k(\omega_2)E_l(\omega_3). \quad \text{(I.1)}$$

Needless to say, this expression is quite complicated due to the 81 different combination of indices (in three dimensions) for a given set of frequency labels. Symmetries simplify this behavior, but only to an extent. We will, therefore, consider two types of simplifications for our analysis, but do keep in mind they do not always hold. We will start with the extraordinarily simple (but one of the most practically useful) cases of a lossless material, in which we may write the time-domain relation

$$P_i(\mathbf{r},\omega) = \epsilon_0\chi^{(3)}_{ijkl}E_j(\mathbf{r},t)E_k(\mathbf{r},t)E_l(\mathbf{r},t). \quad \text{(I.2)}$$

Let us now consider what happens when we have a monochromatic field of the form

$$E(\mathbf{r},t) = E(\mathbf{r})e^{-i\omega t} + \text{c.c.} \quad \text{(I.3)}$$

There are polarization terms at $\pm 3\omega$ and $\pm\omega$. The $\pm 3\omega$ terms lead to third-harmonic generation, which typically is hard to phase-match for reasons discussed in the previous unit. The $\pm\omega$ polarization terms however lead to an effective correction to the refractive index that is always relevant. The corresponding polarization is

$$P_i(\mathbf{r},\omega) = \epsilon_0\chi^{(3)}_{ijkl}\left(E_j(\mathbf{r})E_k(\mathbf{r})E_l^*(\mathbf{r}) + E_j(\mathbf{r})E_k^*(\mathbf{r})E_l(\mathbf{r}) + E_j^*(\mathbf{r})E_k(\mathbf{r})E_l(\mathbf{r})\right). \quad \text{(I.4)}$$

[1] These effects are also realizable via second-order nonlinear media, where under some circumstances $\chi^{(2)}$ can act as an effective $\chi^{(3)}$ but there are provisos here and the effects we will describe are much more commonly explored in the third-order case.

In a lossless material, we may use Kleinman symmetry to write

$$P_i(\mathbf{r},\omega) = \epsilon_0 \left(\chi^{(3)}_{ijkl} E_j(\mathbf{r}) E_k(\mathbf{r}) E^*_l(\mathbf{r}) + \chi^{(3)}_{ijkl} E_j(\mathbf{r}) E^*_k(\mathbf{r}) E_l(\mathbf{r}) + \chi^{(3)}_{ijkl} E^*_k(\mathbf{r}) E_j(\mathbf{r}) E_l(\mathbf{r}) \right). \tag{I.5}$$

This can be written as an effective anisotropic index of refraction, via

$$P_i(\mathbf{r},\omega) = \epsilon_0 \left(\chi^{(3)}_{ijkl} E_j(\mathbf{r}) E_k(\mathbf{r}) E^*_l(\mathbf{r}) + 2\chi^{(3)}_{ijkl} E_j(\mathbf{r}) E^*_k(\mathbf{r}) E_l(\mathbf{r}) \right) \equiv \epsilon_0 \chi_{\text{eff},ij}(\mathbf{r}) E_j(\mathbf{r}), \tag{I.6}$$

with

$$\chi_{\text{eff},ij}(\mathbf{r}) = \chi^{(3)}_{ijkl} E_k(\mathbf{r}) E^*_l(\mathbf{r}) + 2\chi^{(3)}_{ijkl} E^*_k(\mathbf{r}) E_l(\mathbf{r}) = 3\chi^{(3)}_{ijkl} E^*_k(\mathbf{r}) E_l(\mathbf{r}). \tag{I.7}$$

The exact anisotropy depends on the nonzero indices of the third-order susceptibility, but we can see already at this stage schematically that the polarization at the same frequency acts almost like an intensity-dependent index of refraction. To simplify further, we will make use of symmetry.

A. Effect of symmetries

Symmetries strongly reduce the number of independent components. A detailed group theoretic analysis of which crystal classes have how many independent elements is beyond the scope of this work, but we will consider one example, which is the isotropic case. For an isotropic system, the number of independent elements is one (rather than 81!). This can be understood using the quartic oscillator as an example. We derived the quartic oscillator nonlinear susceptibility assuming an isotropic potential. The result was:

$$\chi^{(3)}_{ijkl}(\omega_\sigma;\omega_1,\omega_2,\omega_3) = -\frac{nq^4}{3m^3\epsilon_0} \frac{\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}}{D(\omega_\sigma)D(\omega_1)D(\omega_2)D(\omega_3)}. \tag{I.8}$$

Clearly all elements are proportional to $-\frac{nq^4}{3m^3\epsilon_0}\frac{1}{D(\omega_\sigma)D(\omega_1)D(\omega_2)D(\omega_3)}$. In fact, it turns out that in an isotropic system, the susceptibility *must* have the tensor structure of the quartic oscillator we analyzed. We may therefore write a generic isotropic lossless third-order susceptibility as

$$\chi^{(3)}_{ijkl} = \frac{1}{3}\chi^{(3)}(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}), \tag{I.9}$$

where $\chi^{(3)}$ corresponds to an element such as $\chi^{(3)}_{1111}$. Plugging this relationship into the nonlinear polarization at frequency ω, we get

$$P_i(\mathbf{r},\omega) = \epsilon_0\chi^{(3)}(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})E_k^*(\mathbf{r})E_l(\mathbf{r})E_j(\mathbf{r}), \tag{I.10}$$

which may be written as

$$\mathbf{P}(\mathbf{r},\omega) = \epsilon_0\chi^{(3)}\left(2(\mathbf{E}(\mathbf{r})\cdot\mathbf{E}^*(\mathbf{r}))\mathbf{E}(\mathbf{r}) + (\mathbf{E}(\mathbf{r})\cdot\mathbf{E}(\mathbf{r}))\mathbf{E}^*(\mathbf{r})\right). \tag{I.11}$$

A term like this acts as an effective polarization rotation, which you will explore on the problem set.

Let us however now consider the very simple case in which the field is polarized in a single direction, e.g., $\hat{x}$, so that we get

$$P(\mathbf{r},\omega) = 3\epsilon_0\chi^{(3)}|E(\mathbf{r})|^2E(\mathbf{r}), \tag{I.12}$$

where the scalarized form of $\mathbf{P}, \mathbf{E}$ should be understood e.g., as $\hat{x}\cdot\mathbf{P}$. Let us now explore the effective index of refraction for an x-polarized wave as follows. The scalar Helmholtz equation in the presence of a source is given by

$$\left(\nabla^2 + \left(\frac{n_0\omega}{c}\right)^2\right)E(\mathbf{r}) = -\mu_0\omega^2P_{\mathrm{NL}} = -3\frac{\omega^2}{c^2}\chi^{(3)}|E(\mathbf{r})|^2E(\mathbf{r}), \tag{I.13}$$

where $n_0 = n(\omega)$ is the linear index of refraction at frequency ω. Clearly, by moving the right-hand side over to the left-hand side, we have

$$\left(\nabla^2 + \left(\frac{n(I)\omega}{c}\right)^2\right)E(\mathbf{r}), \tag{I.14}$$

where

$$n(I)^2 = n_0^2 + 3\chi^{(3)}|E(\mathbf{r})|^2 \implies n(I) = n_0 + \frac{3\chi^{(3)}|E(\mathbf{r})|^2}{2n_0} = n_0 + \frac{3\chi^{(3)}}{4n_0^2\epsilon_0 c}I. \tag{I.15}$$

This is typically expressed as

$$n(I) = n_0 + n_2 I, \tag{I.16}$$

where

$$n_2 = \frac{3\chi^{(3)}}{4n_0^2\epsilon_0 c}. \tag{I.17}$$

The quantity n_2 is called the nonlinear index of refraction and has units of inverse intensity. In MKS units, there are m^2/W.

II. EFFECT OF THE INTENSITY-DEPENDENT REFRACTIVE INDEX ON SHORT PULSES

The third-order nonlinear contribution is much weaker than the second-order nonlinear contribution (in a centrosymmetry breaking material) in the following sense: we know that the polarization is a perturbation expansion and that the contribution of the third-order term is like $P^{(3)} \sim P^{(1)}(E/E_0)^2 \sim P^{(2)}(E/E_0)$ and so for $E \ll E_0 \sim 0.1$V/pm, the third-order contribution to the polarization is very small in magnitude. Said in terms of the nonlinear index of refraction, consider light with an intensity of 1 GW/cm^2. In a material like amorphous silica, where $n_2 \sim 10^{-20}$m^2/W, the corresponding change in the refractive index is only 10^{-7} which is very small. It is possible to have small changes like this lead to large effects, particularly when the light propagates over long distances. To achieve these intensities in the first place, it helps to have both well-focused light (confined to a small lateral dimension) and to compress the energy of the light into a short pulse, which enhances the instantaneous electric field. Since we are considering instantaneous materials, their polarization responds to the instantaneous field and intensity. This brings us to an analysis of how short pulses propagate in third-order nonlinear media.

Let us consider a simple case of short pulses propagating in $\chi^{(3)}$ media in which the light can be described effectively as one-dimensional, propagating in a single direction z with fixed polarization. In that case, the relevant self-polarization terms are along the same direction as the electric field and we may consider a nonlinear scalar wave treatment. Although this treatment sounds artificial, it well-describes a situation in which light is confined to propagate along an optical waveguide. One of the most technologically relevant examples of an optical waveguide is an optical fiber. An example of an optical fiber includes a high-index "core" region of some radius surrounded by a lower index cladding region. The difference in index can be small, on the order of 0.02. We know from Maxwell's equations that a region of high-index acts as a region of negative potential [2]. This core-cladding difference is sufficient to trap

[2] As a rudimentary example, consider the Helmholtz equation for a scalar wave propagating in a variable refractive index: $\nabla^2 E(\mathbf{r},\omega) = -\frac{\omega^2}{c^2} n^2(\mathbf{r}) E(\mathbf{r},\omega)$. This is formally equivalent to a

a small number of bound states. Since they are bound transversally, their transverse dynamics are effectly frozen and we can focus purely on field variations in the z direction. The result of this treatment is effectively to recover the one-dimensional pulse propagation equation that we derived on the homework. To see this, let us start by writing the full spatiotemporal electric field, corresponding to a particular transverse mode $u(\mathbf{r})$ polarized in the $\hat{x}$ direction. This is possible to achieve by using polarization-maintaining fibers that exploit anisotropy to break the degeneracy between $\hat{x}$ and $\hat{y}$ polarized modes. The field may be written as [3]

$$\mathbf{E}(\mathbf{r}, t) = \hat{x} u_k(\mathbf{r}) A(z, t) e^{ikz - i\omega t} + \text{c.c.} \tag{II.1}$$

The corresponding field, in the frequency domain, could be expressed as:

$$\mathbf{E}(\mathbf{r}, \omega) = \hat{x} u_k(\mathbf{r}) A(z, \omega) e^{ikz}. \tag{II.2}$$

Let us consider the Maxwell equations in the absence of nonlinear polarization for the time being. The source-free frequency domain Maxwell equation is

$$\left(\nabla^2 + n^2(\mathbf{r}) \frac{\omega^2}{c^2}\right) \mathbf{E}(\mathbf{r}, \omega) = 0 \tag{II.3}$$

The left-hand side simplifies as follows. The Maxwell eigenmode $u_k(\mathbf{r})$ satisfies

$$\nabla_\perp^2 u_k(\mathbf{r}) + \left(n^2(\mathbf{r}) \frac{\omega^2}{c^2} - k^2(\omega)\right) u_k(\mathbf{r}) = 0, \tag{II.4}$$

where $\nabla_\perp^2$ denotes the Laplacian with respect to the transverse coordinates. In writing $k(\omega)$, I've used the fact that the ω acts as an eigenvalue which is dictated by k so that $\omega = \omega(k)$. Alternatively, if we fix ω then there is a specific $k = k(\omega)$ for which $u_k(\mathbf{r})$ has eigenfrequency ω. Using this fact, we are left with the left-hand side of the Maxwell equation being

$$u_k(\mathbf{r}) \left(\frac{d^2}{dz^2} + k^2(\omega)\right) A(z, \omega) e^{ikz} = 0 \tag{II.5}$$

Schrodinger equation with $\frac{\hbar^2}{2m} \to \frac{c^2}{\omega^2}$ and $V = -n^2$. Hence, a high index region can act as a potential well that supports confined states of light. The presence of a propagation wavevector in some direction replaces the mass term $\frac{c^2}{\omega^2}$ with $\frac{1}{\omega^2/c^2 - k^2}$.

[3] As we will see below, the eigenfunction depends on k.

By the same set of steps that we followed in Problem Set 1 (slowly varying envelope approximation, writing $k(\omega)$ in a Taylor series), we have that

$$u_k(\mathbf{r})e^{ikz}\left(2ik\frac{dA}{dz}+2k\sum_m\frac{\beta_m}{m!}(\omega-\omega_0)^m\right)A(z,\omega)=0. \tag{II.6}$$

In time-domain, this may be expressed as

$$u_k(\mathbf{r})e^{ikz-i\omega t}\left(2ik\partial_z A(z,t)+2k\sum_m\frac{i^m\beta_m}{m!}\partial_t^m\right)A(z,t)=0. \tag{II.7}$$

Let us now consider the effect of the nonlinear polarization. Its effect is to add to the right hand side $\mu_0\partial_t^2 P_x(\mathbf{r},t)$. This polarization may be written as

$$P_x(\mathbf{r},t)=3\epsilon_0\chi^{(3)}\left(u_k(\mathbf{r})A(z,t)e^{ikz-i\omega t}+\text{c.c}\right)^3. \tag{II.8}$$

The only relevant parts are those which oscillate near frequency ω by construction $A(z,t)$ is considered to be an envelope field tightly centered spectrally around ω. That term is

$$P_x(\mathbf{r},t)=3\epsilon_0\chi^{(3)}|u_k(\mathbf{r})|^2u_k(\mathbf{r})e^{ikz-i\omega t}|A(z,t)|^2A(z,t). \tag{II.9}$$

The second time-derivative of this term, since the envelope varies in time slowly compared to the oscillation at ω, is approximately $-\omega^2$ and so the resulting equation is then

$$u_k(\mathbf{r})\left(2ik\partial_z A(z,t)+2k\sum_m\frac{i^m\beta_m}{m!}\partial_t^m\right)A(z,t)=-3\frac{\omega^2}{c^2}\chi^{(3)}|u_k(\mathbf{r})|^2u_k(\mathbf{r})|A(z,t)|^2A(z,t). \tag{II.10}$$

To complete the derivation, we note that since $u_k(\mathbf{r})$ is an orthonormal eigenmode of the Maxwell equations [4], we can project it out, writing

$$\left(\partial_z A(z,t)-i\sum_m\frac{i^m\beta_m}{m!}\partial_t^m\right)A(z,t)=\frac{3i\omega^2\chi_{\text{eff}}^{(3)}}{2c^2k}|A(z,t)|^2A(z,t), \tag{II.11}$$

where

$$\chi_{\text{eff}}^{(3)}=\chi^{(3)}\int d\mathbf{r}|u_k(\mathbf{r})|^4. \tag{II.12}$$

[4] The orthonormality condition is $\int d\mathbf{r}\epsilon(\mathbf{r})|u_k(\mathbf{r})|^2=1$

This equation is typically re-arranged in the form:

$$\left(\partial_z + \beta_1 \partial_t\right) A(z,t) = \sum_{m=2} \frac{i^{m+1}\beta_m}{m!} \partial_t^m A(z,t) + \frac{3i\omega\chi_{\text{eff}}^{(3)}}{2cn_0} |A(z,t)|^2 A(z,t). \tag{II.13}$$

We may further simplify the coefficients on the right hand side by identifying $\frac{3i\omega\chi_{\text{eff}}^{(3)}}{2cn_0} \equiv \gamma$, yielding:

$$\left(\partial_z + \beta_1 \partial_t\right) A(z,t) = \sum_{m=2} \frac{i^{m+1}\beta_m}{m!} \partial_t^m A(z,t) + i\gamma |A(z,t)|^2 A(z,t). \tag{II.14}$$

You'll notice that we have separated the first time-derivative term from the remaining. Indeed, the two terms on the left hand side represent a "one-way" wave equation which have as a solution any function $f(t - \beta_1 z)$: i.e., a function moving at the group velocity β_1. It is conventional to go into a moving frame which removes this motion: it leads to simpler forms of the solutions. To make this transformation, take $z' = z, t' = t - \beta_1 z = t - \beta_1 z'$. The chain rule tells us then that

$$\partial_z = \partial_{z'} \frac{\partial z'}{\partial z} + \partial_t' \frac{\partial t'}{\partial z} = \partial_{z'} - \beta_1 \partial_{t'} \tag{II.15}$$

and

$$\partial_t = \partial_{z'} \frac{\partial z'}{\partial t} + \partial_t' \frac{\partial t'}{\partial t} = \partial_t'. \tag{II.16}$$

Plugging these substitutions in gives us

$$\partial_{z'} A(z',t') = \sum_{m=2} \frac{i^{m+1}\beta_m}{m!} \partial_{t'}^m A(z',t') + i\gamma |A(z',t')|^2 A(z',t'). \tag{II.17}$$

This brings us to the final form of the so-called nonlinear short pulse propagation equation that we will use in this class. Before analyzing it, let us discuss the physical significance of this transformation. The new variable z' being the same as z is the physical distance along the waveguide. The new variable t' is like a *retarded time* from electrodynamics. It is related to the "absolute" time by subtracting the time it takes the pulse to arrive to a particular location. That has the following physical interpretation. Let's park at some particular location in the fiber z' and watch a pulse go by us. Suppose we are at a z' such that it takes 1 ns for the peak of the pulse to reach that location (do not worry about dispersion and nonlinearity right now). Suppose the pulse starts propagating in the waveguide from $z' = 0$ at $t = 0$. Then at $t = 1$ ns (absolute time), at our $z' \neq 0$, we see the pulse passing by. The corresponding t' for the center of the pulse is zero. The leading edge of the pulse corresponds to negative t' and trailing edge to positive t'.

In the presence of other effects like nonlinearity and dispersion, we will see the pulse profile in time at some fixed z get distorted. For example, if the pulse peak arrives at positive t' then it slowed down upon its propagation. More generally, the time profile we see, parked at some z is more or less the same as what happens when we take a snapshot at fixed time and look at the spatial dependence of the pulse (for short enough pulses not undergoing very rapid dynamics in z. That is because in the absence of other effects, the solution is $f(z,t) = f(0, t - \beta_1 z)$ which translates rigidly at the group velocity [5].

As one last important point about the moving frame. Suppose there is no higher-order dispersion $\beta_{m\geq 2} = 0$ and no nonlinearity $\gamma = 0$. Then $\partial_z' A(z', t') = 0$ and so $A(z', t') = A(0, t') = A(0, t - \beta_1 z)$. The pulse, in this moving frame, does not evolve at all: it has the same shape as it had at $z = 0$. That is the benefit of the moving frame: it moves out these trivial translational dynamics. Hopefully this discussion provides you with a lot of intuition for how z', t' work. Now we will now study the consequences of this equation systematically.

In what follows, since $z' = z$ we will use z instead of z' for notational simplicity.

A. Linear dynamics

Let us start by considering the linear limit, which is obtained by setting $\gamma = 0$. Our equation is therefore

$$\partial_z A(z, t') = \sum_{m=2} \frac{i^{m+1}\beta_m}{m!} \partial_{t'}^m A(z, t'). \tag{II.18}$$

[5] More explicitly, if we pursue the line or argument we above, let's fix z and watch the pulse go by us in time. That is described by fixing z and letting t vary in $f(z,t) = f(0, t - \beta_1 z)$. If we take a snapshot in time and look at the spatial dependence of the pulse, that's described by the very same function with t fixed. As you can see, these two are related by a reflection and scaling. The reflection is why the trailing edge of the pulse is positive t'. The scale is just a change of units from space to time.

This is trivially solved in Fourier domain. Fourier transforming in time, we get

$$\partial_z A(z,\Omega) = \sum_{m=2} \frac{i^{m+1}\beta_m}{m!}(-i\Omega)^m A(z,\Omega) = i\beta(\Omega)A(z,\Omega), \tag{II.19}$$

where we have defined

$$\beta(\Omega) = \sum_{m=2}^{\infty} \frac{\beta_m}{m!}\Omega^m. \tag{II.20}$$

We have defined the frequency variable as Ω to remind us that this frequency describes time variations of the *envelope*. Time variations of the total field are related by a shift by ω. In particular

$$E(\omega+\Omega) \sim A(\Omega). \tag{II.21}$$

In Fourier domain, we then see that

$$A(z,\Omega) = A(0,\Omega)e^{i\beta(\Omega)z}. \tag{II.22}$$

Each spectral (Fourier) component picks up a (relative) frequency dependent phase. Let's explore the consequences of this in the time-domain by considering a simple situation in which the only non-negligible dispersion is second-order, which is called group velocity dispersion (GVD). In this case, the time-domain field is

$$A(z,t') = \int \frac{d\Omega}{2\pi} e^{-i\Omega t' + \frac{1}{2}i\beta_2\Omega^2 z} A(0,\Omega). \tag{II.23}$$

From this form, it is clear why β_2 represents a group velocity dispersion: each spectral component moves as a plane wave with an Ω dependent phase velocity which is, in the lab frame, $(\beta_1 + \frac{1}{2}\beta_2\Omega)^{-1}$. Alternatively, we could say that the inverse group velocity is $dk/d\omega = \beta_1 + \beta_2\Omega$. Let us consider now what happens when the initial pulse is a Gaussian, with $A(0,t) = A_0 e^{-t^2/(2\tau^2)}$ with τ the pulse duration. We solved this problem on the first problem set and we know that the solution for the time-domain field at finite z' is

$$A(z,t') = A_0\sqrt{\frac{\tau^2}{\tau^2 - i\beta_2 z}} \exp\left[-\frac{t'^2}{2(\tau^2 - i\beta_2 z)}\right]. \tag{II.24}$$

The instantaneous pulse intensity [6], which is given by $I(\mathbf{r},z,t') \sim |E(z,t')|^2 = |u_k(\mathbf{r})|^2|A(z,t')|^2$. Typically we don't care too much about the variations in the

[6] Note that this is still time-averaged over an optical cycle which is assumed to be short relative to the envelope variations in time

transverse plane as we typically collect all of the light within the transverse plane, and so we average over transverse coordinates. Therefore, the instantaneous optical *power* $P(z,t') \sim |A(z,t')|^2$. Therefore, when we talk about optical power, we will use it interchangeably with $|A(z,t')|^2$ and in some cases even normalize A such that its square has dimensions of power[7].

The optical power is then in the appropriate units

$$P(z,t') = P_0 \frac{1}{\sqrt{1+\frac{\beta_2^2 z^2}{\tau^4}}} \exp\left[-\frac{t^2}{\tau^2+\frac{\beta_2^2 z^2}{\tau^2}}\right]. \tag{II.27}$$

To simplify the interpretation, define a *dispersion length* $L_D = \tau^2/\beta_2$, which lets us write the above as

$$P(z,t') = P_0 \frac{1}{\sqrt{1+(z/L_D)^2}} \exp\left[-\frac{t^2}{\tau^2(1+(z/L_D)^2)}\right]. \tag{II.28}$$

This equation describes a *spreading* of the pulse and a concomitant dilution of the energy over a longer time period (causing a reduction in power). The pulse duration increases by a factor of $\sqrt{2}$ over $z = L_D$. Over a very long distance of propagation, the pulse duration eventually becomes $\tau z/L_D$ and the peak power decreases as $1/z$.

Beyond the reduction of the optical power, another subtle effect that occurs is *chirping*. If we look at a particular portion or time-segment of the pulse, we will see that depending on the sign of the dispersion, the trailing edge is redder than

[7] Note that if we elect for such a normalization, we must change the expression for γ. As we derived the equation $|A|^2$ has dimensions of $[|E|^2]L^2$ which has dimensions of $[P/\epsilon_0 c]$. To change the units of $|A|^2$ to be power, we must redefine γ to be

$$\gamma = \frac{3i\omega\chi^{(3)}_{\text{eff}}}{2\epsilon_0 c^2 n_0} = 2n_0 n_2 \frac{\omega}{c} \int d\mathbf{r}\ |u(\mathbf{r})|^2 \tag{II.25}$$

. The integral has dimensions of inverse area, and is typically defined as an effective inverse area A^{-1}_{eff}. To match the literature, we also adjust the scale of $|A|$ by a factor of $\sqrt{2}$ so that $A \to \sqrt{2}A$. The factor of 2 can be seem as coming from enforcing that $|A|^2$ itself is the optical power. Note that any adjustment of scale (even dimensionless adjustments) requires a change in the coefficients due to the nonlinearity. This lets us write

$$\gamma = \frac{2\pi n_0 n_2}{\lambda A} \tag{II.26}$$

which has dimensions of $[P^{-1}]L^{-1}$, which is a standard and frequently used expression relating the nonlinear index to the coefficient in the nonlinear pulse propagation equation.

Since the exponent oscillates very rapidly for t' comparable to the dispersed pulse duration, i.e, for $t' \sim \sqrt{z/L_D}\tau$, we can perform a stationary phase approximation. The stationary phase approximation tells us that the t' values that determine the integral are the ones that make the derivative of the phase vanish: $i\Omega - iL_D t'/(\tau^2 z) = 0 \implies \Omega = t'/(\beta_2 z)$. Not worrying about overall factors, this gives

$$A(\Omega, t_0) \sim \sqrt{P\left(z, \beta_2 z\Omega\right)} \text{ if } \beta_2\Omega z \in [t_0 - \Delta t/2, t_0 + \Delta t/2], \text{ else } 0. \tag{II.34}$$

The if is because if $\Omega = t'/(\beta_2 z)$ is satisfied for a t' outside of our integration window then we don't pick up the stationary phase point in the integral and we expect to get a much lower value which may be approximated as zero. That says that time and frequency get locked together: the frequency at that time slice is very concentrated around $t_0/(\beta_2 z)$. This is precisely the result we were looking for. The spectrum around time slice at t_0 only has a meaningful value if $\beta_2\Omega z = t_0$ (up to the width of the window function).

Of course, this could have been anticipated by our group velocity analysis! If we have a spectrum of the initial pulse which consists of temporally localized photons, and photons of different colors move at different group velocities $v_g \approx \frac{1}{\beta_1} - \frac{\beta_2\Omega}{\beta_1^2}$ [8], then after a long distance, photons of different colors will be separated from each other. Photons at (relative) frequency Ω will at a distance z be lag the photons at $\Omega = 0$ by a distance $\delta z = \frac{\beta_2\Omega}{\beta_1^2}(\Delta t_{\text{lab}}) = \frac{\beta_2\Omega}{\beta_1^2}\beta_1 z$. Converting the spatial separation into a temporal separation, we get a temporal separation of $\beta_2\Omega z$ as expected: a higher frequency component will be slower and will end up at the trailing edge of the pulse (positive t'). This is precisely the same type of thing that occurs via a prism or a grating except that a prism separates colors in angle, and this "time prism" separates the colors temporally (or longitudinally in space).

[8] This from inverting $\beta_1 + \beta_2\Omega$.

B. Nonlinear dynamics without dispersion: self-phase modulation

Let us now look at what the nonlinear terms do. To make things simple, we will consider what happens in the absence of dispersion $\beta_{m\geq 2} = 0$. In this case, we have

$$\partial_z A(z, t') = i\gamma |A(z, t')|^2 A(z, t'). \tag{II.35}$$

This is in fact exactly solvable. That is because the instantaneous intensity is conserved by this equation. Consider

$$\partial_z |A(z, t')|^2 = 2\text{Re } A^*(z, t')\partial_z A(z, t') = 2\text{Re } i\gamma |A(z, t')|^4 = 0. \tag{II.36}$$

Therefore $|A(z, t')|^2 = |A(0, t')|^2$ and

$$\partial_z A(z, t') = i\gamma |A(0, t')|^2 A(z, t'). \tag{II.37}$$

This can therefore be directly integrated, yielding

$$A(z, t') = \exp\left[i\gamma z |A(0, t')|^2\right] A(0, t'). \tag{II.38}$$

This equation has the interpretation that each time slice of the pulse (which is equivalent to points on the pulse) picks up a phase proportional to the intensity of the pulse at that time slice. This phase is called the *nonlinear* phase and is given instantaneously by

$$\phi_{\text{NL}}(z, t') = \gamma z |A(0, t')|^2. \tag{II.39}$$

This is, appropriately, called *self-phase modulation.* As we see, the instantaneous power does not change at all, but the spectrum does change.

Self-phase modulation also leads to a chirp. To see this, let us consider our windowed Fourier transform again. Let us assume that the nonlinear phase is sufficiently large such that we can use stationary phase techniques again to compute the windowed Fourier transform with a square windowing function. The windowed Fourier transform is equal to

$$A(\Omega, t_0) = \int\limits_{t_0-(\Delta t)/2}^{t_0+(\Delta t)/2} dt\ e^{i\Omega t' + i\phi_{\text{NL}}(z,t')} \sqrt{P(0, t')} e^{i\phi(0,t')}, \tag{II.40}$$

where we have expressed the initial pulse ($z = 0$) in polar form in anticipation of using stationary phase and we have defined the phase of the initial pulse as $\phi(0, t')$.

Let us consider a purely real and thus unchirped Gaussian pulse as an example. Then the stationary phase condition is

$$\Omega = -\partial_{t'}\phi_{\mathrm{NL}}(z,t') = \gamma z|A_0|^2\frac{2t}{\tau^2}e^{-t^2/\tau^2}. \tag{II.41}$$

We see in this case that the frequency at some particular time slice is redder for $t < 0$ (leading edge) and bluer for $t > 0$ (trailing edge).

C. Solitons

We now consider what happens when dispersion and nonlinearity are simultaneously present in our system. This is a rich regime with a lot of complicated physics, some of which remains even today to be an active form of research. There are typically very few analytical solutions in this regime. However, under specific conditions with only second-order dispersion and nonlinearity, the dynamics are analytically soluble. Although these solutions have been known for decades, small changes of the equations evade analytical solutions.

To give some intuition for the fundamental idea, we notice that self-phase modulation pushes high frequencies to $t > 0$, while group velocity dispersion with $\beta_2 > 0$ does the same thing, and the two effects reinforce each other. On the other hand, if the dispersion is *anomalous* and $\beta_2 < 0$ then the two effects could even cancel out. It turns out this cancellation is perfect and a pulse can propagate in a fully shape-preserving manner. We can get at this idea as follows. There is a way to find this shape-preserving solution, called an *soliton* exactly but it is very involved and for this course we will content ourselves with stating the solution.

The following solution is an example of a soliton.

$$A(0,t') = A_0\mathrm{sech}(t/\tau). \tag{II.42}$$

Remarkably, its z evolution is

$$A(0,t') = A_0\mathrm{sech}(t/\tau)e^{\frac{1}{2}i\gamma|A_0|^2 z}. \tag{II.43}$$

It propagates and just picks up a nonlinear phase.

To see this, let's compute the right-hand side. Using $\text{sech}(x)'' = \text{sech}(x) - 2\text{sech}^3(x)$, we have that the right hand side of the pulse propagation equation:

$$\partial_z A(z,t') = -i\frac{\beta_2}{2}\partial_{t'}^2 A(z,t') + i\gamma|A(z,t')|^2 A(z,t'), \tag{II.44}$$

which in this case is called the *nonlinear Schrodinger equation* is,

$$\left(-\frac{i\beta_2}{2\tau^2}A_0\left(\text{sech}(t'/\tau) - 2\text{sech}^3(t'/\tau)\right) + i\gamma|A_0|^2 A_0\text{sech}^3(t'/\tau)\right)e^{\frac{1}{2}i\gamma|A_0|^2 z}. \tag{II.45}$$

While the right hand side is simply

$$\frac{1}{2}i\gamma|A_0|^2 A_0\text{sech}(t/\tau)e^{\frac{1}{2}i\gamma|A_0|^2 z}. \tag{II.46}$$

To get both sides to be equal, one requires

$$\frac{\beta_2}{\tau^2} + \gamma|A_0|^2 = 0 \implies |A_0|^2 = -\frac{\beta_2}{\gamma\tau^2}, \tag{II.47}$$

which requires $\beta_2 < 0$ as is consistent with the argument we made about SPM and dispersion imparting opposite chirp. With this constraint, we see that the left-hand side and right-hand side are equal to each other. We can understand the condition above as a constraint on the amplitude given known dispersion, pulse duration, and nonlinearity. Or we can understand it as a constraint on the parameters for a given amplitude.

Beyond the soliton solution described here, there is a continuous family of soliton solutions which are related by translations, phases, and initial velocities. I will not show it here, but you can show (with a bit more algebra) that the following is also a solution to the nonlinear Schrodinger equation:

$$A(z,t') = A_0\text{sech}\left(\frac{t - t_0 + |\beta_2|\Omega_0 z}{\tau}\right)\exp\left[i\left(\phi_0 - \Omega_0(t - t_0) - \frac{1}{2}|\beta_2|\Omega_0^2 z + \frac{1}{2}\gamma|A_0|^2 z\right)\right]. \tag{II.48}$$

Compared to the soliton solution we evaluated previously, this one is translated by a time offset t_0 and has an overall phase ϕ_0. Additionally there is a frequency offset Ω_0 which gives the pulse an overall group velocity related to $|\beta_2|\Omega$ in the moving frame.

There also exist higher-order solitons which exhibit periodic breathing behavior that may be understood as collections of more than one soliton which nonlinearly interact.

III. NONLINEAR REFRACTION OF BEAMS

The treatment of nonlinear effects thus far has frozen out the transverse dynamics of the electromagnetic field: we have assumed that the light occupies effectively a single-mode waveguide. There are however very important effects that become apparent when we consider the transverse dynamics, including self-focusing, spatial solitons, and beam collapse. In what follows, we develop these effects. We will pursue a simplified treatment in which we consider propagation of scalar monochromatic beams. The topic of vectorial spatiotemporal pulses with transverse and longitudinal dynamics is an area of still active research.

The starting point is the Helmholtz equation for the positive frequency part of a monochromatic field at frequency ω. We will consider the light to be propagating in a medium with uniform linear index n_0. Including nonlinear polarization, the result is

$$\left(\nabla^2 + k_0^2\right) E(\mathbf{r}) = -3\frac{\omega^2}{c^2}\chi^{(3)}|E(\mathbf{r})|^2 E(\mathbf{r}), \tag{III.1}$$

where we have defined $k_0 = n_0\omega/c$.

To proceed, let us consider a beam which propagates primarily along z with wavevector k_0 with slow residual variations in z. While equivalent to our standard slowly varying envelope approximation, in the spatial case, this carries an additional implication. In particular, our slowly varying envelope approximation is equivalent to saying that there only negligible spatial Fourier components with large angles to the z axis. This is called the *paraxial approximation* because it says that the wavevector components of the beam are mostly parallel to the propagation axis . This is because, if we consider the Fourier transform of the field in space

$$E(\mathbf{k}) = \int d\mathbf{r}\; e^{i\mathbf{k}\cdot\mathbf{r}} E(\mathbf{r}), \tag{III.2}$$

the only $\mathbf{k}$ components we expect to see are ones satisfying

$$\mathbf{k}_\perp^2 + k_z^2 = k_0^2. \tag{III.3}$$

Therefore, if we insist that the z-dependence has Fourier components $k_z \approx k_0$, then we have that $|k_\perp| \ll k_0$, and so the angular spread of the beam must be small. This is because the wavevector components are associated with transverse propagation angles (in a small angle approximation)

$$\theta(k_\perp) \approx |k_\perp|/k_0 \ll 1, \tag{III.4}$$

establishing the equivalence of the spatially varying envelope approximation to the paraxial approximation.

Therefore we write

$$E(\mathbf{r}, \omega) = A(\boldsymbol{\rho}, z)e^{ik_0 z}, \tag{III.5}$$

where $A(\boldsymbol{\rho}, z)$ is a slowly varying envelope in z. Dealing with the z derivatives as usual, we have that

$$\left(\nabla_\perp^2 + 2ik_0\partial_z\right) A(\boldsymbol{\rho}, z) = -3\frac{\omega^2}{c^2}\chi^{(3)}|A(\boldsymbol{\rho}, z)|^2 A(\boldsymbol{\rho}, z). \tag{III.6}$$

Rearranging terms, we find:

$$\partial_z A = \frac{i}{2k_0}\nabla_\perp^2 A + \frac{3i\omega\chi^{(3)}}{2n_0 c}|A|^2 A. \tag{III.7}$$

We then change units such that $|A|^2$ has dimensions of intensity rather than electric field squared (but we still refer to the variable as A). The resulting equation becomes

$$\partial_z A = \frac{i}{2k_0}\nabla_\perp^2 A + \frac{3i\omega\chi^{(3)}}{4\epsilon_0 n_0^2 c^2}|A|^2 A, \tag{III.8}$$

which we write as

$$\partial_z A = \frac{i}{2k_0}\nabla_\perp^2 A + i\kappa|A|^2 A, \tag{III.9}$$

where

$$\kappa = \frac{3\omega\chi^{(3)}}{4\epsilon_0 n_0^2 c^2} = \frac{n_2 k_0}{n_0}. \tag{III.10}$$

This equation will be the basis for our investigations of spatial nonlinear phenomena. Like the case where we studied nonlinear pulse propagation, we will start with the linear behavior. Then, we will consider nonlinearity without diffraction. And finally we will consider what happens when the two effects come together.

A. Diffraction of paraxial beams

In the absence of nonlinearity, we have

$$\partial_z A = \frac{i}{2k_0}\nabla_\perp^2 A. \tag{III.11}$$

To understand the dynamics of this equation, let us consider what happens when we have a known field $A(\boldsymbol{\rho}, 0)$ and we would like to know what the field is for $z > 0$. From Fourier considerations, we could write this as

$$A(\boldsymbol{\rho}, z) = \int \frac{d^2k}{(2\pi)^2} e^{i\mathbf{k}\cdot\boldsymbol{\rho} - i\frac{k^2}{2k_0}z} A(\mathbf{k}, 0), \tag{III.12}$$

where I have taken $\mathbf{k}_\perp \to \mathbf{k}$ for notational ease. A particularly important case is that of the *Gaussian beam.* This is described by a situation where

$$A(\boldsymbol{\rho}, 0) = A_0 e^{-\frac{\rho^2}{w_0^2}}, \tag{III.13}$$

where w_0 the *beam waist* is a measure of the width of the beam in the transverse directions. The resulting field at nonzero z is given by the resulting Gaussian integral which is just a two-dimensionally separable form of the same integral we encountered when looking at the effect of second-order dispersion on pulses[9]. The result can be written as

$$A(\boldsymbol{\rho}, z) = A_0 \frac{1}{1 + iz/b} e^{-\frac{\rho^2}{w_0^2(1+iz/b)}}, \tag{III.14}$$

where $b = k_0 w_0^2/2$ is called the confocal parameter. The expression above for the so-called *Gaussian beam* is often expressed in a non-complex form as

$$A(\boldsymbol{\rho}, z) = A_0 \frac{1}{\sqrt{1 + z^2/b^2}} e^{-i\psi(z)} e^{-\frac{\rho^2}{w^2(z)}} e^{\frac{ik_0\rho^2}{2R(z)}}, \tag{III.15}$$

where

$$w(z) = w_0 \sqrt{1 + \left(\frac{z}{b}\right)^2}, \tag{III.16}$$

$$R^{-1}(z) = \frac{z}{z^2 + b^2} \tag{III.17}$$

and

$$\psi(z) = \tan^{-1}\left(\frac{z}{b}\right). \tag{III.18}$$

This is the standard form of the Gaussian beam[10] Let us now discuss the physics of this solution. I should mention that this is one of the most important free-space

[9] In general, the math of beams is quite analogous to that of pulses in the linear and nonlinear regimes even when one case has analytical solutions and the other does not

[10] If you compare this form to that in Haus' *Waves and Fields in Optoelectronics* you will find that ours maps to his with $\psi \to -\psi$ and $R \to -R$. This is a straightforward consequence of him using e^{-ikz} to represent forward propagation while we use e^{ikz}. Additionally, you will find an additional factor of $\sqrt{\frac{2}{\pi w_0^2}}$ in his result because he defines a normalized Gaussian beam such that its square integral over x, y is unity.

propagating solutions to Maxwell equations as it very well approximates the fields emitted by laser cavities and thus many beams we work with are well-described by the Gaussian beam. Study it well!

The first thing to point out is that the beam width in the transverse direction, described by the beam waist $w(z)$, increases for $|z| > 0$, this is diffraction and is essentially equivalent to what occurs when we have light undergo single slit diffraction. The scale over which the beam waist changes is the confocal parameter $kw_0^2/2$. At large distance z the waist expands linearly, implying that the light expands in a cone-like manner. The half-angle θ of that cone can be approximated as

$$\theta = \lim_{z\to\infty} \frac{w(z)}{z} = \frac{w_0}{b} = \frac{2}{k_0 w_0} = \frac{\lambda}{n_0 \pi w_0}. \tag{III.19}$$

This same form for the angle could be inferred from simple Fourier uncertainty principle considerations. Consider a wave in one dimension which is localized to a size $\Delta x \sim w_0$. Then there must also be a spread in wavevectors $\Delta k_x \sim \frac{1}{\Delta x} = \frac{1}{w_0}$, corresponding to an angular spread[11]. That angular spread is simply $\Delta\theta = \frac{\Delta k_x}{k_0} \sim \frac{\lambda}{n_0 w_0}$. Hence, we identify this Gaussian beam spreading as diffraction. Importantly, we see that the angle of expansion depends on the minimum beam waist. The more localized we are to start, the more the beam diverges as it propagates. This can be straightforwardly recalled using the Fourier uncertainty principle. We see that when $w_0 \gg \lambda$, the angular spread is low, and the paraxial approximation is satisfied. For beam waists close to the wavelength, the angular spread is large and this treatment becomes invalid. We need to consider vector diffraction theory in that case. Diffraction is expected by energy conservation to dilute the amplitude of the beam. This is exactly what the factor $(1 + z^2/b^2)^{-1/2}$ is doing.

The next factors we would like to understand are the phase factors. Let us start with the factor $e^{ik_0\rho^2/2R(z)}$. This describes a wavefront with a z-dependent radius of curvature $R(z)$. At large distances the surfaces of constant phase are given by parabolas with $z = \frac{k_0\rho^2}{2}$. The remaining factor $e^{-i\psi(z)}$ called the *Gouy phase* describes an additional effective wavevector for light at the beam waist (at large

[11] The Fourier uncertainty principle is $\Delta x \Delta k_x \geq \frac{1}{2}$ which is straightforwardly inferred from taking the quantum mechanical version of the statement $\Delta x \Delta p_x \geq \frac{\hbar}{2}$ and taking $p_x = \hbar k_x$. I am dropping all prefactors since this argument is meant to just explain the scaling.

z this factor is equivalent to an overall phase). That additional wavevector is of order $-1/b$ and essentially reflects that sharp transverse variations associated with localization must take away from the longitudinal wavevector.

Similar to our analysis of pulse propagation, we can define an instantaneous *spatial frequency* (also known as, wavevector) for the beam. Recall in the temporal case that we defined our instantaneous frequency in the context of finding the spectrogram of self-phase modulation. This instantaneous frequency arises from stationary phase considerations. The instantaneous wavevector is given by

$$\mathbf{k}(\mathbf{r}) = \nabla\left(k_0 z + \frac{ik_0\rho^2}{2R(z)} - \psi(z)\right) = \hat{z}\left(k_0 - \frac{b}{z^2+b^2} + \frac{k_0\rho^2}{2}\frac{z^2-b^2}{(z^2+b^2)^2}\right) + \frac{k_0}{R(z)}\boldsymbol{\rho}. \tag{III.20}$$

We will not worry about the two corrections to k_0 which are of the same order and are Gouy-like. Thus we will approximate the local wavevector as

$$\mathbf{k}(\mathbf{r}) \approx k_0\hat{z} + \frac{k_0}{R(z)}\boldsymbol{\rho}. \tag{III.21}$$

The local angle of propagation (if we think of our expanding beam in terms of a diverging bundle of rays) is, in a small-angle approximation:

$$\theta(\mathbf{r}) \approx \frac{\rho}{R(z)}, \tag{III.22}$$

which, at the beam waist, and for large z approximates to $w(z)/z = w_0/b$ as consistent with our previous analysis. The instantaneous wavevector picture, besides giving a rigorous and intuitive ray-optics description of our beam, also will be indispensable for understanding the effect of nonlinearity, which we now turn to.

B. Self-phase modulation in space

Now that we understand the linear behavior somewhat, let us now consider the nonlinear terms alone. We are left with

$$\partial_z A = i\kappa|A|^2 A, \tag{III.23}$$

which can be seen to be solved by

$$A(\boldsymbol{\rho}, z) = A(\boldsymbol{\rho}, 0)\exp\left[i\kappa|A(\boldsymbol{\rho},0)|^2 z\right] A(\boldsymbol{\rho}, 0). \tag{III.24}$$

As we shall now show, this corresponds to changing the instantaneous wavevector of the beam, leading to a remarkable effect called self-focusing.

Let us consider what happens in the case of a Gaussian beam. Let us assume that the Gaussian beam is *collimated*: in other words, $k_0 w_0 \gg 1$ and so the confocal parameter is very large compared to the length of propagation. In this limit, we can approximate z as zero for the Gaussian beam. In this case, this collimated beam experiences an additional phase (a phase modulation) of

$$\kappa |A_0|^2 e^{-2\rho^2/w_0^2}. \tag{III.25}$$

The instantaneous wavevector is therefore given by (for $z \ll b$)

$$\mathbf{k}(\mathbf{r}) = k_0 \hat{z} - 4\kappa z |A_0|^2 \frac{\boldsymbol{\rho}}{w^2} e^{-2\rho^2/w_0^2}. \tag{III.26}$$

You can already see it: the additional wavevector points inwards towards the center of the beam at all points along the beam. That means that the light rays want to bend inwards towards the center, making the beam smaller. This is the essence of self-focusing. If it is possible for example for all rays to converge onto a point, then the beam will undergo a catastrophic collapse. In particular: the beam having some intensity causes it to bend inwards, and increase its intensity. That however causes more bending, until the beam focuses to a point. At this stage, the beam intensity can become so high that few materials can withstand it and undergo damage.

Before we provide a condition for self-focusing, it is worthwhile to describe another perspective on how we see inward focusing that does not rely on the instantaneous wavevector. The intensity-dependent phase modulation is strongest at the center corresponding to a phase advance for rays propagating on-axis relative to rays propagating (along z) which are displaced from the beam axis. If we draw surfaces of constant phase, they will be curved, and the curvature will be such that they appear to emanate from a focus at positive z. This is suggestive that the nonlinearity is acting as a focusing lens.

In what follows, we will pursue a significantly oversimplified analysis of self-focusing which should not be taken to be quantitative. However, it will yield the correct scaling laws for the so-called critical power for self-focusing, and the length over which self-focusing occurs. We see that the inward bending of rays depends on the phase gradient. Let us approximate this phase gradient as $\kappa Iz/w$ where I is some characteristic intensity and w is some characteristic length. For a Gaussian beam, I could be the intensity at the beam center and w the minimum beam waist. In this case, the downward bending angle of an off-axis ray is given by

$$\theta_{\mathrm{NL}} = \frac{\kappa Iz}{wk_0} = \frac{n_2 Iz}{n_0 w}. \tag{III.27}$$

We shall (very coarsely) assume that this angle is maintained for a distance z until the ray crosses the propagation axis, therefore traversing a lateral distance w. By construction, this ray always meets the central on-axis ray since from our considerations above for the Gaussian beam, the central ray experiences no deflection. Then $\theta \approx w/z$ in a small-angle approximation and we have

$$\theta_{\mathrm{NL}}^2 = \frac{n_2 I}{n_0} \implies \theta_{\mathrm{NL}} = \sqrt{\frac{n_2 I}{n_0}}. \tag{III.28}$$

The distance required to focus is given by

$$L_f = w/\theta_{\mathrm{NL}} = w\sqrt{\frac{n_0}{n_2 I}}. \tag{III.29}$$

Now one thing our analysis has neglected so far is the fact that a Gaussian beam will naturally spread outwards with some asymptotic angle $\theta_{\mathrm{D}} = \frac{\lambda}{\pi n_0 w_0}$. One can expect that if $\theta_{\mathrm{NL}} < \theta_{\mathrm{D}}$ then the beam will still spread out, just at a reduced angle. Meanwhile, if $\theta_{\mathrm{NL}} > \theta_{\mathrm{D}}$, then one gets catastrophic self-focusing. The boundary case is when self-focusing balances diffraction, and in this case it is possible to get a shape-preserving beam called an *spatial soliton*. This effect is called self-trapping and this occurs for a critical intensity I_c and corresponding critical power $P_c = \frac{\pi}{2}w^2 I$ equal to[12]:

$$\sqrt{\frac{n_2 I_c}{n_0}} = \frac{\lambda}{\pi n_0 w} \implies P_c = \frac{\lambda^2}{2\pi n_0 n_2}. \tag{III.30}$$

Expressed in terms of the critical power, we can write the self-focusing length as[13]

$$L_f = \frac{\pi n_0 w^2}{\lambda}\frac{1}{\sqrt{P/P_c}}. \tag{III.31}$$

---

[12] The prefactor for this critical power is different than both what is in Boyd and what is in Yariv's *Quantum Electronics*. This is to say that we should only take the scaling seriously.

[13] This also slightly differs from Boyd in overall prefactor, as is to be expected.

For $P = P_c$ and a 100 micron beam waist at a wavelength of 1 micron with a characteristic index of 1.5, the self-focusing length is roughly 5 cm. The corresponding critical power in a material such as silica is on the order of 10 MW.

C. Filamentation of an optical beam

We conclude this chapter by discussing what happens when the beam power is significantly larger than the critical power for self-focusing. The short answer is that the beam will break up into beamlets or filaments each which carry approximately the critical power for self-trapping. We will pursue a simplified analysis of this problem as follows. Let us crudely approximate our beam as a strong plane wave. The rough idea is that the beam will break up into other beams propagating in other directions. So let us describe the growth of these beamlets as follows. Let us take our field as

$$A(\boldsymbol{\rho}, z) = \left(A_0(z) + \delta A(\boldsymbol{\rho}, z)\right), \tag{III.32}$$

where $|\delta A| \ll |A_0|$. In other words we have a strong forward-propagating plane wave representing the main beam[14] which seeds a transverse varying field $\delta A(\boldsymbol{\rho}, z)$.

$$\partial_z A_0 = i\kappa |A_0|^2 A_0 \implies A_0(z) = A_0(0) e^{i\kappa |A_0(0)|^2 z} \equiv A_0 e^{iQz}. \tag{III.33}$$

In what follows, we will refer to $A_0(0)$ as A_0 and $Q = \kappa |A_0|^2$. At order δA, we have two coupled equations

$$\begin{aligned}
\partial_z \delta A &= \frac{i}{2k_0} \nabla_\perp^2 \delta A + i\kappa \left(2|A_0(z)|^2 \delta A + A_0^2(z) \delta A^*\right) \\
\partial_z \delta A^* &= -\frac{i}{2k_0} \nabla_\perp^2 \delta A^* - i\kappa \left(2|A_0(z)|^2 \delta A^* + A_0^{*2}(z) \delta A\right).
\end{aligned} \tag{III.34}$$

which may be simplified as

$$\begin{aligned}
\partial_z \delta A &= \frac{i}{2k_0} \nabla_\perp^2 \delta A + i\kappa \left(2|A_0|^2 \delta A + A_0^2 e^{2iQz} \delta A^*\right) \\
\partial_z \delta A^* &= -\frac{i}{2k_0} \nabla_\perp^2 \delta A^* - i\kappa \left(2|A_0|^2 \delta A^* + A_0^{*2} e^{-2iQz} \delta A\right).
\end{aligned} \tag{III.35}$$

[14] Remember that the overall electric field has an $e^{ik_0 z}$.

By defining envelope variables $\delta A = \delta\tilde{A}e^{iQz+i\arg(A_0)}$, we can get rid of the nonlinear oscillations and write

$$\begin{aligned}
\partial_z \delta\tilde{A} &= \frac{i}{2k_0}\nabla_\perp^2 \delta\tilde{A} + i\kappa\left(|A_0|^2\delta\tilde{A} + |A_0|^2\delta\tilde{A}^*\right) \\
\partial_z \delta\tilde{A}^* &= -\frac{i}{2k_0}\nabla_\perp^2 \delta\tilde{A}^* - i\kappa\left(|A_0|^2\delta\tilde{A}^* + |A_0|^2\delta\tilde{A}\right).
\end{aligned} \tag{III.36}$$

Let us now go into the Fourier domain in the transverse wavevector space, defining $A(\boldsymbol{\rho}, z) = \int \frac{d^2q}{(2\pi)^2} e^{i\mathbf{q}\cdot\boldsymbol{\rho}} A(\mathbf{q}, z)$. We see then that we may write

$$\partial_z \begin{pmatrix} \tilde{A}(\mathbf{q}, z) \\ \tilde{A}(-\mathbf{q}, z)^* \end{pmatrix} = i \begin{pmatrix} K & Q \\ -Q & -K \end{pmatrix} \begin{pmatrix} \tilde{A}(\mathbf{q}, z) \\ \tilde{A}(-\mathbf{q}, z)^* \end{pmatrix}. \tag{III.37}$$

In writing this, we have used the fact that $(A^*)(\mathbf{q}) = (A(-\mathbf{q}))^*$. We have also defined $K = Q - \frac{q^2}{2k_0}$.

As with typical linear systems of equations, a good way to analyze it is by looking at the eigenvalues and eigenvectors of the matrix. The eigenvalues $\lambda_{1,2}$ are found to be[15]

$$\lambda_{1,2} = \pm\sqrt{Q^2 - K^2} = \pm\sqrt{2\frac{q^2}{2k_0}Q - \left(\frac{q^2}{2k_0}\right)^2} = \sqrt{\frac{q^2}{2k_0}}\sqrt{2Q - \frac{q^2}{2k_0}}. \tag{III.38}$$

We see that in the absence of nonlinearity, the eigenvalues are purely imaginary, corresponding to oscillatory solutions. For finite nonlinearity, we see that for small transverse wavevectors, there is a positive eigenvalue corresponding to an exponentially growing solution. This corresponds to the exponential growth of perturbations from initial noise. The maximum growth rate occurs for a transverse wavevector $q^2 \equiv 2k_0 x$ which maximizes

$$x(2Q - x), \tag{III.39}$$

for which the maximum is simply $x = Q$, corresponding to a maximum gain at wavevector $q_{\max}$ equal to

$$q_{\max} = \sqrt{2k_0\kappa|A_0|^2}, \tag{III.40}$$

[15] The quick way to do this is to note that the matrix is $iK\sigma_z - Q\sigma_y$ and to use the standard result for eigenvalues of sums of Pauli matrices. Note however that because of the imaginary coefficient of σ_z, the matrix is not Hermitian.

corresponding to the harmonic mean of the central wavevector and the "nonlinear wavevector".

The picture we have developed is now this: a plane wave propagating in a nonlinear medium is unstable and seeds the exponential growth of waves propagating off-axis. We can go back to the beam picture by considering each of these plane waves as beamlets carrying a localized region of light around it[16]. These beamlets in reality will diffract while growing. This puts a natural cap on the growth because at some point the power in each beamlet will approach the critical power for self-trapping. Thus the initial beam develops modulations which grow into beams or filaments which carry power approximately equal to P_c. This is called filamentation, and the growth of transverse variations (modulations) from noise is called *modulation instability*, because if there is any initial modulation present in the field (nonzero δA in our analysis), then the perturbations will grow from noise.

We mentioned that the perturbations grow from noise. This can correspond to imperfections such as aberrations in the optical beam, or propagation through even a weakly disordered medium. Even in an ostensibly clean beam with field A_0 only, with no initial perturbation, it turns out these perturbations can continue to grow. This happens due to *quantum noise*: Heisenberg's uncertainty principle for the electric and magnetic field states that the fields cannot be exactly zero, only on average both zero. So we can think there is a small fluctuating perturbation called quantum noise which seeds this filamentation. That brings us to the second part of the course, on the *quantum* description of nonlinear optical phenomena.

[16] The linearized analysis above is much more complicated with an optical beam although it can still be done