Unit 2: Effects arising from the second-order nonlinear susceptibility

Nicholas Rivera

(Dated: February 5, 2026)

Contents

In the previous chapter, we developed the consequences of a nonlinear relationship between polarization and electric field, anticipating phenomena such as harmonic generation, sum- and difference-frequency generation, and self-phase modulation. We also developed formal properties of the nonlinear susceptibility. In this chapter, we develop in detail the consequences of second-order nonlinear polarization governed by $\chi^{(2)}$ for the generation of light at new frequencies, as well as the modification of the refractive index of a material by static voltages.

I. WAVE EQUATION DESCRIPTION OF NONLINEAR OPTICS

Recall the fundamental idea that we described when taking about the anharmonic oscillator: a field with frequency components $\omega_1 \cdots \omega_n$ will lead to the creation of polarization at frequency $\omega_\sigma = \omega_1 + \cdots \omega_n$. Maxwell's equations tell us that polarization at frequency ω_σ generates a field at frequency ω_σ. We will now make this explicit and concrete in this section by discussing Maxwell's equations in dielectric media. Then we will analyze a few canonical cases of second-order nonlinear interaction: sum-frequency generation, second-harmonic generation, difference-frequency generation, and the electro-optic effect. These effects form the modern backbone for *much* contemporary research in nonlinear optics.

In time-domain, Maxwell's equations can be combined to yield:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}, t) = -\mu_0 \partial_t (\mathbf{J}(\mathbf{r}, t) + \epsilon_0 \partial_t \mathbf{E}(\mathbf{r}, t)), \tag{I.1}$$

which, introducing the polarization via $\mathbf{J} = \partial_t \mathbf{P}$, can be written as

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}, t) + \frac{1}{c^2} \partial_t^2 \mathbf{E}(\mathbf{r}, t) = -\mu_0 \partial_t^2 \mathbf{P}(\mathbf{r}, t). \tag{I.2}$$

Using the identity $\nabla \times \nabla \times = \nabla(\nabla \cdot) - \nabla^2$, we have

$$\nabla(\nabla \cdot \mathbf{E}(\mathbf{r}, t)) - \nabla^2 \mathbf{E}(\mathbf{r}, t) + \frac{1}{c^2} \partial_t^2 \mathbf{E}(\mathbf{r}, t) = -\mu_0 \partial_t^2 \mathbf{P}(\mathbf{r}, t). \tag{I.3}$$

Typically in linear electromagnetism you might be accustomed to dropping the $\nabla \cdot \mathbf{E}(\mathbf{r}, t)$ term because it vanishes: in nonlinear optics, it is nonzero due to the *anisotropy* of the linear dielectric properties of the medium[1]. In fact, as we will learn

[1] In vacuum, this is because of Gauss' law, $\nabla \cdot \mathbf{D} = \nabla \cdot \epsilon \mathbf{E} = \rho_{\text{free}}$ which tells us that the divergence

in our discussion of sum-frequency generation, anisotropy is *required* to enforce a certain relation called phase-matching needed for *efficient* sum-frequency generation. *However*, the anisotropy of the linear dielectric properties is usually *small* - differences in the index of refraction in two directions can be less than 5% and so it leads to a small correction to the allowed polarizations of the electric field. Therefore, we will drop the divergence, leaving us with the equation:

$$\nabla^2 \mathbf{E}(\mathbf{r}, t) - \frac{1}{c^2}\partial_t^2 \mathbf{E}(\mathbf{r}, t) = \mu_0 \partial_t^2 \mathbf{P}(\mathbf{r}, t). \tag{I.4}$$

In Fourier domain (in time), this may be written:

$$\nabla^2 \mathbf{E}(\mathbf{r}, \omega) - \frac{1}{c^2}\partial_t^2 \mathbf{E}(\mathbf{r}, \omega) = -\mu_0 \omega^2 \mathbf{P}(\mathbf{r}, \omega). \tag{I.5}$$

Importantly, we see what we were saying: polarization at some frequency leads in general to a field at the same frequency.

It is customary to separate out the linear contribution to the polarization, writing the polarization as

$$\mathbf{P}(\mathbf{r}, \omega) = \epsilon_0(\epsilon(\omega) - 1)\mathbf{E}(\mathbf{r}, \omega) + \mathbf{P}_{\mathrm{NL}}(\mathbf{r}, \omega), \tag{I.6}$$

leaving us with the form of the equations we wish to use:

$$\nabla^2 \mathbf{E}(\mathbf{r}, \omega) + \epsilon(\omega)\frac{\omega^2}{c^2}\mathbf{E}(\mathbf{r}, \omega) = -\mu_0 \omega^2 \mathbf{P}(\mathbf{r}, \omega). \tag{I.7}$$

To proceed, we will need to analyze specific scenarios and geometries.

II. SUM-FREQUENCY GENERATION

Consider a situation in which we have two plane waves of frequencies ω_1 and ω_2 impinge upon a second-order nonlinear crystal at normal incidence. They couple to this medium, and so at the input, there are two forward propagating waves $\mathbf{E}_1(\mathbf{r}, t)$ and $\mathbf{E}_2(\mathbf{r}, t)$. Taking their common direction of propagation as defining the z-direction, we may express the fields as

$$\mathbf{E}_i(\mathbf{r}, t) = \hat{\epsilon}_i \left(A_i e^{ik_i z - i\omega_i t} + A_i^* e^{-ik_i z + i\omega_i t} \right), \tag{II.1}$$

of the displacement field vanishes in the absence of source charges. In an isotropic and uniform medium, $\nabla \cdot \mathbf{D} = \epsilon \nabla \cdot \mathbf{E}$ and so the electric field divergence vanishes. In an anisotropic medium, ϵ is different in at least two principal directions and so it cannot be factored out of the divergence.

where A_i is the complex amplitude and $\hat{\epsilon}_i$ is the polarization. Now suppose that these two waves under sum-frequency generation, producing a third field $\mathbf{E}(\mathbf{r}, t)$ in the z direction with frequency $\omega_3 = \omega_1 + \omega_2$. The sum frequency wave *must* be in the z direction. The reason is that the sum-frequency polarization is proportional to $e^{i(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{r}}$ and so the wavevector of the generated field is $\mathbf{k}_1 + \mathbf{k}_2$. Assuming that fields 1 and 2 propagate in the z direction, that forces the third to also propagate in the z direction. That third-field can be expressed as

$$\mathbf{E}_3(\mathbf{r}, t) = \hat{\epsilon}_3 \left(A_3 e^{ik_3 z - i\omega_3 t} + A_3^* e^{-ik_3 z + i\omega_3 t} \right). \tag{II.2}$$

As this sum-frequency field gets built-up from nonexistence, it must draw energy from the fields at frequency ω_1 and ω_2. We know that the intensity, which is the energy per unit time and area associated with the light field, goes as the square of the complex amplitude, and so the complex amplitudes of the two initial fields (1 and 2) must change.

To capture these changes, we make an ansatz of steady-state operation: the amplitudes are constant in time, but not in space. This describes a situation in which the field generated by sum-frequency generation propagates from the region in which it was generated. Since z is the common direction of propagation, we write all three of our fields as

$$\mathbf{E}_i(\mathbf{r}, t) = \hat{\epsilon}_i \left(A_i(z,t) e^{ik_i z - i\omega_i t} + A_i^*(z,t) e^{-ik_i z + i\omega_i t} \right), \tag{II.3}$$

and pursue an equation of motion for $A_i(z, t)$. In what follows, we will make a very common approximation in wave optics, which is the *slowly-varying envelope approximation* (SVEA): this approximation states that the normalized rate of the change in the amplitude envelope $|\partial_z A_i|/|A| \ll k_i|A|$. This is a reasonable approximation in most cases since the sum-frequency generating interaction is relatively inefficient and the amplitude changes at most by some tens of percent over often millimeters or centimeters of propagation. The SVEA manifests as follows: plug in our ansatz into the Maxwell equations: we then get

$$\left(\partial_z^2 A_i(z) e^{ik_i z} \hat{\epsilon}_i + \epsilon(\omega_i) \frac{\omega_i^2}{c^2} \hat{\epsilon}_i A_i(z,t) e^{ik_i z} \right) = -\mu_0 \omega_i^2 \mathbf{P}_i(z). \tag{II.4}$$

In writing this, we have (a) ignored transverse coordinate dependence of the fields (assuming plane waves in those directions), and (b) expressed our polarization in the form

$$\mathbf{P}_i(r,t) = \left(\mathbf{P}_i(z)e^{-i\omega_i t} + \mathbf{P}_i^*(z)e^{i\omega_i t}\right), \tag{II.5}$$

where I have made no assumption yet about the direction and wavevector of the polarization. Importantly, the second derivative can be expressed as

$$\partial_z^2 A_i(z)e^{ik_i z} = e^{ik_i z}\left(\partial_z^2 A + 2ik_i\partial_z A_i - k_i^2 A\right) \approx e^{ik_i z}\left(2ik_i\partial_z A_i - k_i^2 A\right), \tag{II.6}$$

using the slowly-varying envelope approximation[2].

Putting it all together, we get (after some additional simplification)

$$2ik_i\partial_z A_{z,i} + \left(\hat{\epsilon}_i\cdot(\epsilon(\omega_i)\hat{\epsilon}_i)\frac{\omega_i^2}{c^2} - k_i^2\right)A_i(z) = -\mu_0\omega_i^2\hat{\epsilon}_i\cdot\mathbf{P}(z)e^{-ik_i z}. \tag{II.7}$$

In what follows, we will neglect the anisotropy and pretend that the permittivity is a scalar such that $\epsilon(\omega_i)\omega_i^2/c^2 = k_i^2$ such that we are left with the equation[3].

$$\partial_z A_{z,i} = \frac{i\mu_0\omega_i^2}{2k_i}\hat{\epsilon}_i\cdot\mathbf{P}(z)e^{-ik_i z}. \tag{II.8}$$

We are now left with evaluating the nonlinear polarization. For the wave at frequency ω_3 we may write

$$P_i(\omega_3) = \epsilon_0\chi_{ijk}^{(2)}(\omega_3,\omega_1,\omega_2)\hat{\epsilon}_{3,i}\hat{\epsilon}_{1,j}\hat{\epsilon}_{2,k}A_1A_2e^{i(k_1+k_2)z}, \tag{II.9}$$

while for frequency ω_1 we have

$$P_i(\omega_1) = \epsilon_0\chi_{ijk}^{(2)}(\omega_1,\omega_3,-\omega_2)\hat{\epsilon}_{1,i}\hat{\epsilon}_{3,j}\hat{\epsilon}_{2,k}A_3A_2^*e^{i(k_3-k_2)z}, \tag{II.10}$$

and for frequency ω_2 we have

$$P_i(\omega_2) = \epsilon_0\chi_{ijk}^{(2)}(\omega_2,\omega_3,-\omega_1)\hat{\epsilon}_{2,i}\hat{\epsilon}_{3,j}\hat{\epsilon}_{1,k}A_3A_2^*e^{i(k_3-k_1)z}. \tag{II.11}$$

In what follows, we will consider the most important case of sum-frequency generation which is when the material is *lossless* (approximately) and thus dispersionless. Then, with full permutation symmetry, we find that

---

[2] Why didn't we also drop the first-derivative term?

[3] This is what is done in Boyd (he drops the tensorial nature of the permittivity): strictly speaking not very rigorous. One should have retained the divergence terms and then the non-derivative terms would have cancelled as the correct Maxwell equation for a plane wave polarized along some principle direction follows from the equation $-\mathbf{k}\times\mathbf{k}\times\hat{\epsilon} = \epsilon\hat{\epsilon}\omega_i^2/c^2$

$$\hat{\epsilon}_i \cdot P_i(\omega_3) = 4\epsilon_0 d_{\text{eff}} A_1 A_2 e^{i(k_1+k_2)z}, \hat{\epsilon}_i \cdot P_i(\omega_{1,2}) = 4\epsilon_0 d_{\text{eff}} A_3 A_{2,1}^* e^{i(k_3-k_{2,1})z}. \quad \text{(II.12)}$$

Therefore, we may write

$$\partial_z A_3(z) = \frac{2i\omega_3^2 d_{\text{eff}}}{c^2 k_3} A_1 A_2 e^{i\Delta k z} \quad \text{(II.13)}$$

$$\partial_z A_1(z) = \frac{2i\omega_1^2 d_{\text{eff}}}{c^2 k_1} A_3 A_2^* e^{-i\Delta k z} \quad \text{(II.14)}$$

$$\partial_z A_2(z) = \frac{2i\omega_2^2 d_{\text{eff}}}{c^2 k_2} A_3 A_1^* e^{-i\Delta k z}, \quad \text{(II.15)}$$

where $\Delta k = k_1 + k_2 - k_3$. These are the equations of sum-frequency generation.

A. Sum-frequency generation in the non-depleted approximation

In general these equations are nonlinear and difficult to solve (there is an exact solution). That said, most of the insight comes from the regime when the nonlinearity is weak and the intensity built up in the sum-frequency is much smaller than the intensity in the driving fields. Then, we may approximate the driving fields as constant in space. This is the non-depleted approximation. Then, we may integrate the equation for A_3 directly up to a length $z = L$, finding

$$A_3(L) = \frac{2i\omega_3^2 d_{\text{eff}}}{c^2 k_3} A_1(0) A_2(0) \frac{e^{i\Delta k L} - 1}{i\Delta k} = \frac{2i\omega_3^2 d_{\text{eff}}}{c^2 k_3} A_1(0) A_2(0) e^{i\Delta k L/2} L \text{sinc}\left(\frac{1}{2}\Delta k L\right). \quad \text{(II.16)}$$

The intensity in the sum-frequency wave is then given from Poynting's theorem as $2\epsilon_0 n_3 c|A_3|^2$, in terms of the intensities of the incident waves, as

$$I_3(L) = \frac{2\omega_3^2 d_{\text{eff}}}{n_1 n_2 n_3 \epsilon_0 c^3} I_1(0) I_2(0) L^2 \text{sinc}^2\left(\frac{1}{2}\Delta k L\right). \quad \text{(II.17)}$$

Importantly, the intensity in the sum-frequency field is proportional to the product of the intensities of the incident fields, quadratic in propagation length, and weighted by the sinc factor, which rapidly decays if $\Delta k L \gg 1$. The sinc factor is maximized when $\Delta k = 0$ or equivalently when

$$k_1 + k_2 = k_3. \quad \text{(II.18)}$$

Of course, by construction, it is also the case that

$$\omega_1 + \omega_2 = \omega_3. \tag{II.19}$$

This condition is called the *phase matching* condition and in a photon picture could be thought of as an expression of energy-momentum conservation. It is clear that to have efficient generation of the sum frequency, we must achieve phase-matching. While we have only derived this in the perturbative limit, it holds generally.

B. Phase-matching

The phase-matching condition is not automatically satisfied and in fact takes some doing to satisfy, and it can be satisfied using different techniques. We discuss this here. The wavevector condition can be expressed as

$$n(\omega_1)\omega_1 + n(\omega_2)\omega_2 = n(\omega_3)\omega_3 = n(\omega_3)(\omega_1 + \omega_2). \tag{II.20}$$

Combining this with the frequency-matching condition, we have

$$\frac{n(\omega_1)\omega_1 + n(\omega_2)\omega_2}{\omega_1 + \omega_2} = n(\omega_3). \tag{II.21}$$

This equation is in general *not satisfied*. For example, many materials present *positive dispersion* meaning that the index increases with frequency. In the infrared and visible range, this is typical. Let us assume that the material is isotropic and is thus described only by one index of refraction at each frequency. In a system as described, phase-matching is impossible to satisfy, precluding efficient nonlinearities. Suppose without loss of generality that $\omega_3 > \omega_2 > \omega_1$. Then for positive dispersion, we also have $n(\omega_3) > n(\omega_2) > n(\omega_1)$. The expression on the left-hand side of the above equation is the frequency-weighted average index of refraction between waves 1 and 2. There is no way for this quantity to be larger than $n(\omega_2)$, and therefore cannot be equal to $n(\omega_3)$.

C. Birefringent phase matching: angle and temperature tuning

One way to deal with phase-matching is to use *anomalous dispersion*: near a resonance of a material, the index can decrease with increasing frequency. This is not the most common solution that is used: instead, we typically exploit the natural *anisotropy* of a system, the fact that the index of refraction is different in different directions. Let us analyze the example that Boyd does, of a material in the $3m$ (trigonal) crystal class. For a trigonal system, we have that there are two distinct principal values of the index of refraction, n_e (e for extraordinary) corresponding to field polarization along the crystal or c axis of the system, and n_o in the two directions perpendicular to the c-axis. We call such a system *uniaxial* because there is one principal axis with a distinct index of refraction (the c-axis), while the other two are equivalent. A biaxial system refers to the case where all three principal directions have different indices of refraction.

For a wave propagating along wavevector $\mathbf{k}$ at an angle θ to the c-axis, the effective index seen (as shown on Problem Set 1) is

$$\frac{1}{n^2(\theta)} = \frac{\sin^2\theta}{n_e^2} + \frac{\cos^2\theta}{n_o^2}. \tag{II.22}$$

Importantly, the index of refraction associated with extraordinary polarization and ordinary polarization can be different. Ordinary polarization refers to the field polarization (technically D-field) being orthogonal to the c-axis, while extraordinary polarization refers to the other polarization state which has a nonzero projection with the c-axis. In a *negative* uniaxial system, $n_e < n_o$ while in a *positive* uniaxial system $n_e > n_o$.

How does this help with phase matching? If all three waves in the SFG process have ordinary polarization, we can't phase-match, assuming positive dispersion. If all three waves in the SFG process have extraordinary polarization, we also cannot phase-match in the case where all three waves propagate in the same direction (collinear SFG). It is the same argument as in the isotropic case.

However, we see from the argument we made in the isotropic case that if we could somehow make n_3 lower than n_2, phase-matching could at least be possible. This is the idea behind *type-I phase-matching*, which refers to the case where the two input waves have ordinary polarization and the sum frequency wave has extraordinary polarization (in a negative uniaxial crystal[4]). You might also imagine in the case of

[4] In a positive uniaxial crystal, we reverse the polarizations (inputs are extraordinary, output is ordinary

the negative uniaxial crystal that I could get away with making one of the two input polarizations extraordinary (the lower-frequency input would be easier because it pulls the weighted average down less). That is the case of *type-II phase matching*[5].

Let us now see how this all shakes out mathematically. Let us consider type-I phase matching for second-harmonic generation ($\omega + \omega \to 2\omega$). You can convince yourself that the phase matching condition is simply (in the negative uniaxial case)

$$2\omega n_o(\omega) \to 2\omega n_e(2\omega) \implies n_o(\omega) = n(2\omega, \theta). \tag{II.23}$$

If we take our three waves to have the same direction, which is at an angle θ to the c-axis, then our phase matching condition becomes

$$\frac{1}{n_0^2(\omega)} = \frac{\sin^2\theta}{n_e^2(2\omega)} + \frac{\cos^2\theta}{n_o^2(2\omega)}. \tag{II.24}$$

We may express this as

$$\sin^2\theta = \frac{\frac{1}{n_o^2(\omega)} - \frac{1}{n_o^2(2\omega)}}{\frac{1}{n_e^2(2\omega)} - \frac{1}{n_o^2(2\omega)}}. \tag{II.25}$$

Let's sanity check this. Phase-matching implies that this equation can be solved for some angle, meaning that the right hand side is between 0 and 1. If the right-hand side becomes negative, or larger than 1, we can't phase-match. We know that lots of normal dispersion makes phase-matching harder, and we know that birefringence makes things easier. Note that by assumption that we have normal dispersion, the numerator is positive. Similarly, assuming a negative uniaxial crystal implies that the denominator is positive. If the normal dispersion is strong, then the numerator increases, pushing us away from having a solution. Similarly, if the birefringence is weak, then the denominator gets small, having the same effect of pushing us away from having a solution. So we see that we have a kind of ratio of dispersion to birefringence.

Of course, for a given frequency, the principal indices of refraction as a function of frequency is fixed, and so we can think of the right-hand side as fixed. By rotating the crystal however, we change the angle of the c-axis to our incident light direction, and so we can think of performing *angle tuning* until phase-matching is satisfied.

[5] Again, for positive uniaxial crystals, reverse the polarizations

There is a severe drawback related to angle tuning, which is spatial walk-off (in Problem Set 2, you'll analyze temporal walk off of pulses). In reality, we don't send plane waves at a crystal, we send localized wavepackets. You might imagine then that if the two input field wavepackets and the sum-frequency wavepacket do not overlap in space, there should not be a nonlinear interaction since the nonlinear polarization cares about the three fields at the same location. Let us consider again the case of type-I phase-matching for second-harmonic generation. The fundamental (input) wave has a group velocity along the direction of $\mathbf{k}$. The harmonic however, does not[6]. As a result, the harmonic moves in a different direction and speed from the fundamental and so after some "drift time" they will not overlap, ceasing the nonlinear interaction. This is called *walk-off* for reasons that are sufficiently self-explanatory.

This can be avoided by *temperature tuning*. By changing the temperature of the medium, we can change $n_{o,e}(\omega)$ and $n_{o,e}(2\omega)$ to satisfy the phase-matching equation for angles θ for which walk-off does not occur. Those angles are, as per the previous footnote, $\theta = 0$ and $\theta = \pi/2$. Of course, $\theta = 0$ cannot get you phase-matching by assumption that there is normal dispersion (even with temperature tuning). Importantly, for temperature tuning to work, we need the birefringence to respond strongly to temperature while the dispersion responds less. The opposite case would suppress phase-matching.

D. Quasi-phase matching

There is another powerful trick for achieving phase matching, involving in principle no temperature-tuning, no birefringence, and using collinear fields. This is of course very helpful as it gets around a bunch of the issues mentioned above! It also allows access to elements of the d tensor that are often the largest. For example, in some materials, including lithium niobate, d_{33} is the largest element. However, if the

[6] In an anisotropic medium, the group velocity is $\nabla_{\mathbf{k}}\omega(\mathbf{k})$. For the extraordinary polarization, we can take the result in Problem 1 and write $2\omega_{\mathbf{k}}\nabla_{\mathbf{k}}\omega_{\mathbf{k}} = c^2\left(\hat{k}2k\left(\frac{\sin^2\theta}{n_e^2}+\frac{\cos^2\theta}{n_o^2}\right)+\hat{\theta}k\frac{\sin 2\theta}{n_e^2}-\frac{\sin 2\theta}{n_o^2}\right)$. This clearly has a nonzero component along $\hat{k}$ unless $\theta = 0$ or $\theta = \pi/2$

polarizations of the three waves in SFG or the two polarizations in SHG are not the same, then d_{33} does not contribute to d_{eff}. Of course, if the polarizations are the same, we run into issues related to normal dispersion.

The solution is to periodically flip the sign of d_{eff} (equivalently, flipping the sign of $\chi^{(2)}$. How do we do this? From the standpoint of the cubic oscillator we analyzed, the sign of $\chi^{(2)}$ is controlled by a the constant setting the cubic contribution to the potential felt by the oscillator. It turns out that in centrosymmetry breaking materials like lithium niobate, they are *ferroelectric* having a permanent electric dipole moment not unlike the permanent magnetic moment of a ferromagnet like iron. You can imagine that the presence of frozen permanent dipoles in a material creates a strong potential for electrons, and is solely responsible for breaking centrosymmetry[7]! The direction of these dipoles can be reversed by application of an electric field, thus flipping the sign of the symmetry-breaking potential and this is effectively "flipping a". So our plan is to periodically flip the polarity (in a way to be elaborated) of our ferroelectric: this is called *periodic poling*[8].

Physically, why does this help? Suppose we have a segment of nonlinear material of fixed polarity, and assume we're doing SFG where all three polarizations are the same in a normally dispersive material. As we have established already, we cannot phase match. Therefore, $\Delta k \neq 0$, and we expect to see that, in the case of weak depletion, A_3 is given by (II.16) and will first rise but then come back down. Suppose that when $|A_3|$ reaches its maximum at $z = \pi/\Delta k$, we suddenly flip the sign of d: then A_3 will increase instead of decrease. It will basically "copy" its evolution from 0 to z. When that branch reaches its maximum, flip the sign of d again. This tells us automatically what the period of our poling variation should be. If the period of Λ is

$$\Lambda = \frac{2\pi}{\Delta k}, \tag{II.26}$$

then we're in business. Our graphical argument however suggests that we could let $\frac{1}{2}\Delta k z = (2m+1)\pi/2$ before flipping the poling direction, allowing also for periods

[7] At high temperatures, LN is a cubic paraelectric and is thus centrosymmetric and doesn't have second-order nonlinearity. This also tells us that we can probe symmetry-breaking phases of matter by their second-harmonic / nonlinear response.

[8] Terms in nonlinear optics make a lot of sense about 98% of the time.

$(4m+2)\pi/\Delta k$.

Let us suppose now that we have a periodic d_{eff} in the equations for SFG that we have developed (with period Λ). Then it may be written as

$$d_{\text{eff}}(z) = \sum_{m=-\infty}^{\infty} d_m e^{2\pi i m/\Lambda}. \tag{II.27}$$

In what follows, suppose one particular harmonic q allows for phase-matching. We take only that harmonic[9], thus approximating $d_{\text{eff}}(z) \approx d_q e^{2\pi i q/\Lambda} + d_q^* e^{-2\pi i q/\Lambda}$. The coupled amplitude equations become:

$$\partial_z A_3(z) = \frac{2i\omega_3^2 d_q}{c^2 k_3} A_1 A_2 e^{i\Delta k_q z} \tag{II.28}$$

$$\partial_z A_1(z) = \frac{2i\omega_1^2 d_q^*}{c^2 k_1} A_3 A_2^* e^{-i\Delta k_q z} \tag{II.29}$$

$$\partial_z A_2(z) = \frac{2i\omega_2^2 d_q^*}{c^2 k_2} A_3 A_1^* e^{-i\Delta k_q z}, \tag{II.30}$$

where $\Delta k_m = k_1 + k_2 - k_3 + 2\pi q/\Lambda$. Clearly, in the non-depleted pump case, A_3 has a chance to build up coherently over long distances if one of these Δk_m terms is zero.

E. Manley-Rowe relations

Let us go back to our analysis of the coupled-wave equations for sum-frequency generation. We were able to simply solve these equations in the undepleted approximation. But let us consider what physical statements are true outside of this approximation.

Clearly, we expect energy conservation to hold, which indicates that in a steady-state situation, the total intensity of all three waves is invariant under propagation if no energy builds up in any sub-region of the system. Therefore, we expect that

$$\frac{d}{dz}(I_1 + I_2 + I_3) = 0. \tag{II.31}$$

[9] The other terms give a small oscillatory contribution that could be ignored.

For continuous-wave light, the time-averaged intensity of any one wave is $I_i = 2\epsilon_0 n_i c|A_i|^2$[10] Therefore, we expect to see that

$$\frac{d}{dz}\sum_i n_i|A_i|^2 = 2\sum_i n_i \text{Re } A_i^* \frac{dA_i}{dz} = 0. \tag{II.32}$$

This may easily be checked to be true of the coupled-amplitude equations. In particular, when calculating the z-derivative of the intensities, you would find that

$$\frac{dI_1}{dz} = 2\epsilon_0 c \frac{2\omega_1 d_{\text{eff}}}{c} \text{Re } iA_1^* A_2^* A_3 e^{-i\Delta kz}, \tag{II.33}$$

$$\frac{dI_2}{dz} = 2\epsilon_0 c \frac{2\omega_2 d_{\text{eff}}}{c} \text{Re } iA_1^* A_2^* A_3 e^{-i\Delta kz}, \tag{II.34}$$

and

$$\frac{dI_3}{dz} = -2\epsilon_0 c \frac{2\omega_3 d_{\text{eff}}}{c} \text{Re } iA_1 A_2 A_3^* e^{i\Delta kz}. \tag{II.35}$$

The sum of these terms vanishes if $\omega_1 + \omega_2 - \omega_3 = 0$, which is true by construction. In writing this, we notice another set of exact relations, namely:

$$\frac{dM_{1,2}}{dz} \equiv \frac{1}{\omega_1}\frac{dI_1}{dz} - \frac{1}{\omega_2}\frac{dI_2}{dz} = 0, \tag{II.36}$$

$$\frac{dM_{1,3}}{dz} \equiv \frac{1}{\omega_1}\frac{dI_1}{dz} + \frac{1}{\omega_3}\frac{dI_3}{dz} = 0, \tag{II.37}$$

and

$$\frac{dM_{2,3}}{dz} \equiv \frac{1}{\omega_2}\frac{dI_2}{dz} + \frac{1}{\omega_3}\frac{dI_3}{dz} = 0. \tag{II.38}$$

These relations are called *Manley-Rowe* relations, and indicate that the following quantities are invariant:

$$M_{1,2} = \frac{I_1}{\omega_1} - \frac{I_2}{\omega_2}, M_{1,3} = \frac{I_1}{\omega_1} + \frac{I_3}{\omega_3}, M_{2,3} = \frac{I_2}{\omega_2} + \frac{I_3}{\omega_3}. \tag{II.39}$$

Invariants are physically useful quantities and often correspond to quantities with clear physical interpretation such as energy, momentum, angular momentum, etc. The Manley-Rowe invariants correspond to conservation of photon number. In particular, let us consider the photon picture of sum frequency generation. In sum frequency generation, we have one photon at ω_1 is annihilated, one photon at ω_2 is annihilated, and one photon at ω_3 is created. Let us imagine that as time elapses,

[10] This is assuming that we specify a CW field as $E_i = A_i e^{-i\omega_i t} + A_i^* e^{i\omega_i t}$.

sum-frequency generation processes occur. Then we certainly expect some type of deterministic relation between changes in the number of one type of photons and photons of another. If the number of photons per unit length at frequencies $\omega_{1,2,3}$ is $n_{1,2,3}$ respectively, then we expect that

$$\frac{dn_1}{dt} = \frac{dn_2}{dt} = -\frac{dn_3}{dt}, \tag{II.40}$$

where the number of photons per unit volume at frequencies $\omega_{1,2,3}$ is $n_{1,2,3}$. To relate this to z derivatives, remember that these photons are all moving with their own velocities v_i (the speed of light divided by the index of refraction). Therefore the time derivative is in fact $d/dt = \partial/\partial_t + v_i\partial/\partial_z$. In the steady-state assumption, the partial time derivative vanishes and we are left with $v_1\partial_z n_1 = v_2\partial_z n_2$. But the intensity is simply $\hbar\omega_i v_i n_i$, and so we see simply that this is equivalent to $\partial_z M_{1,2} = 0$, which is the same as above[11]

F. Sum-frequency generation beyond the non-depleted approximation

Let us consider a situation of sum-frequency generation in which the lowest frequency wave at frequency ω_1 is weak and the second input wave at frequency ω_2 is strong. We could imagine the ω_1 field to be an infrared field and the ω_2 field to be a laser field, and they mix to create a wave at frequency $\omega_3 = \omega_1 + \omega_2$. It is possible to solve the equations of sum-frequency generation exactly, in terms of Jacobi elliptic functions. However, when the ω_2 field is strong, we can treat it as undepleted, while still tracking the spatial evolution of the ω_1 and ω_3 fields. For the fields at ω_1 and ω_3, we have

$$\partial_z A_3(z) = \frac{2i\omega_3^2 d_{\text{eff}}}{c^2 k_3} A_1 A_2 e^{i\Delta kz} \tag{II.41}$$

$$\partial_z A_1(z) = \frac{2i\omega_1^2 d_{\text{eff}}}{c^2 k_1} A_3 A_2^* e^{-i\Delta kz} \tag{II.42}$$

When A_2 is treated as constant, we see that we have two linear differential equations with spatially varying coefficients which can be solved. Before doing so, it will help

---

[11] Here I have a partial derivative, that is because I explicitly noted the time dynamics of the photons moving. In the above, there was no notion of time as we took the steady-state limit of the wave equation: thus the total and partial derivatives are interchangeable.

to define $A_3(z) = \tilde{A}_3(z)e^{i\Delta kz/2}$ and $A_1(z) = \tilde{A}_1(z)e^{-i\Delta kz/2}$. It immediately follows that

$$\partial_z \tilde{A}_3(z) + \frac{i\Delta k}{2}\tilde{A}_3(z) = \alpha \tilde{A}_1 \tag{II.43}$$

and

$$\partial_z \tilde{A}_1(z) - \frac{i\Delta k}{2}\tilde{A}_1(z) = \beta \tilde{A}_3 \tag{II.44}$$

with $\alpha = \frac{2i\omega_3^2 d_{\text{eff}}}{c^2 k_3} A_2$ and $\beta = \frac{2i\omega_1^2 d_{\text{eff}}}{c^2 k_1} A_2^*$. This may be written in matrix notation as

$$\partial_z \begin{pmatrix} \tilde{A}_1 \\ \tilde{A}_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}i\Delta k & \beta \\ \alpha & -\frac{1}{2}i\Delta k \end{pmatrix} \begin{pmatrix} \tilde{A}_1 \\ \tilde{A}_3 \end{pmatrix}. \tag{II.45}$$

We know from systems of differential equations such as these that the eigenvalues of the matrix on the right hand side dictate the z-evolution of the fields. The two eigenvalues denoted $\lambda_{1,2}$ can be found to be

$$\lambda_{1,2} = \pm i\sqrt{\frac{4\omega_1^2\omega_3^2 d_{\text{eff}}^2}{k_1 k_3 c^4}|A_2|^2 + \frac{(\Delta k)^2}{4}} \equiv \pm\lambda, \tag{II.46}$$

with λ_1, λ_2 being assigned the positive (negative) roots respectively. We may then write

$$\tilde{A}_{1,3}(z) = J_{1,3}\cos(\lambda z) + K_{1,3}\sin(\lambda z), \tag{II.47}$$

with J, K being constants.

Let us take as initial conditions the typical case in which there is no initial sum-frequency field, and so we have $J_1 = A_1(0)$ and $K_1 = \frac{i\Delta k}{2}A_1(0)$, while $J_3 = 0$ and $K_3 = \frac{2i\omega_3^2 d_{\text{eff}}}{c^2 k_3} A_2 A_1(0)$. Putting it all together, we have

$$A_1(z) = \left[\cos(\lambda z) + \frac{i\Delta k}{2\lambda}\sin(\lambda z)\right] A_1(0)e^{-i\Delta kz/2} \tag{II.48}$$

and

$$A_3(z) = \left[\frac{2i\omega_3^2 d_{\text{eff}}}{c^2 k_3 \lambda} A_2 \sin(\lambda z)\right] A_1(0)e^{i\Delta kz/2}. \tag{II.49}$$

As can be seen from these expressions, the sum-frequency wave builds up from zero, oscillating in space at period $2\pi/\lambda$. The maximum intensity of the sum-frequency field is

$$I_{3,\max} = \frac{n_3}{n_1}\frac{\frac{4\omega_3^2 d_{\text{eff}}^2 |A_2|^2}{c^4 k_3^2}}{\frac{4\omega_1^2\omega_3^2 d_{\text{eff}}^2}{k_1 k_3 c^4}|A_2|^2 + \frac{(\Delta k)^2}{4}} I_1(0). \tag{II.50}$$

It is clear from this expression that the maximum is realized when phase-matching is satisfied ($\Delta k = 0$), and we have

$$I_{3,\max} = \frac{\omega_3}{\omega_1} I_1(0). \tag{II.51}$$

III. SECOND-HARMONIC GENERATION

In second-harmonic generation, we take a wave at frequency ω_1 and convert it using a second-order nonlinear medium into a wave at frequency $\omega_2 = 2\omega_1$. In the photon picture, we take two photons at ω_1 and convert it into a photon at ω_2. The physics of second-harmonic generation is similar to that of sum-frequency generation: indeed we can see second-harmonic generation as a limiting case of sum-frequency generation where the two input frequencies are equal. The coupled-amplitude equations can be written as (Reader: make sure to derive this!)

$$\partial_z A_1(z) = \frac{2i\omega_1^2 d_{\text{eff}}}{c^2 k_1} A_1^* A_2 e^{-i\Delta k z} \tag{III.1}$$

$$\partial_z A_2(z) = \frac{i\omega_2^2 d_{\text{eff}}}{c^2 k_2} A_1^2 e^{i\Delta k z}, \tag{III.2}$$

with $\Delta k = 2k_1 - k_2$[12] It is simple to follow the discussion of sum-frequency generation and find the second-harmonic intensity in the undepleted pump approximation for the fundamental, and the reader is very strongly urged to do this. Here, we will solve this equation without the non-depleted pump approximation.

In what follows, we will consider the simpler case of these equations where the interaction is phase-matched: this gives the most efficient second-harmonic conversion anyway. Therefore we may set $\Delta k = 0$. To solve this equation, we start by identifying useful invariants. One such invariant is the total intensity, which implies that

$$n_1|A_1|^2 + n_2|A_2|^2 = \text{const.}. \tag{III.3}$$

The Manley-Rowe invariant for second-harmonic generation is identical to the total intensity (reader: check that!). There is one other quantity which is invariant when

[12] Note that the "missing" factor of 2 in the equation for the second harmonic field comes from the fact that there is only one instance of A_1^2 upon squaring the electric field, as opposed to the two instances of $A_1^* A_2$.

phase-matching holds. It is

$$\text{Re } A_1^2 A_2^* = |A_1|^2 |A_2| \cos(2\phi_1 - \phi_2) = \text{const.}, \tag{III.4}$$

where we have defined $A_1 = |A_1|e^{i\phi_1}$ and $A_2 = |A_2|e^{i\phi_2}$. This invariant may seem mysterious at first glance (and indeed Boyd just finds this from the coupled amplitude equations directly), but corresponds to constancy of the time-averaged energy density[13].

We will consider the simple but very common case where there is no initial intensity in the second harmonic wave (other cases can easily be done but it involves more steps). In this case, it is clear that the phase of the incident wave does not matter (absolute phases can always be multiplied away!): we set $\phi_1(0)$ to zero without loss of generality. Our nonlinear invariant tells us that $|A_1|^2|A_2|\cos(2\phi_1 - \phi_2) = 0$ which means that $\cos(2\phi_1 - \phi_2) = 0$ unless we are at locations z for which $|A_2| = 0$ in which case the phase ϕ_2 loses any meaning anyway and ϕ_1 can be set arbitrarily. Therefore, we see that

$$2\phi_1 - \phi_2 = (2m+1)\frac{\pi}{2} \tag{III.5}$$

for any integer m. If convert the evolution equation for A_2 into equations for $|A_2|$ and ϕ_2, we find that

$$\partial_z |A_2| + i|A_2|\partial_z \phi_2 = \frac{i\omega_2^2 d_{\text{eff}}}{k_2 c^2} |A_1|^2 e^{i(2\phi_1 - \phi_2)}, \tag{III.6}$$

---

[13] The linear part of the time-averaged energy density is simply the time-average of $\sum_i \epsilon_0 n_i^2 E_i^2$ which is $2\epsilon_0\left(n_1^2|A_1|^2 + n_2^2|A_2|^2\right)$. The nonlinear energy density is proportional to $\chi^{(2)}E^3$ as per our discussion of lossless media in chapter 1. The time-averaged part corresponds to the DC terms in $E^3 = (A_1 e^{ik_1 z - i\omega_1 t} + A_2 e^{ik_2 z - i\omega_2 t} + \text{c.c.})^3$. The frequencies in this expression are $\pm\omega_1 \pm \omega_1 \pm \omega_2 = \pm\omega_1 \pm \omega_1 \pm 2\omega_1$. Clearly there are only two terms for which the sum of these is zero. The time-averaged nonlinear energy density associated with those terms is $A_1^2 A_2^* e^{i\Delta k z} + A_1^{*2} A_2 e^{-i\Delta k z} = 2|A_1|^2|A_2|\cos(2\phi_1 - \phi_2 + \Delta k z)$. The total time-averaged energy density is the sum of these two terms. In the steady-state, constancy of the time-averaged energy density u: $du/dt = 0$ implies that $\partial_z u = 0$. At the same time, we also know that the time-averaged *intensity* is conserved as a function of z: $n_1|A_1|^2 + n_2|A_2|^2 = \text{const.}$. If phase-matching holds, then $n_1 = n_2$ implying $|A_1|^2 + |A_2|^2$ is conserved and so similarly, the linear part of the time-averaged energy density is conserved. Thus, for the time-averaged energy density to be constant, we require $|A_1|^2|A_2|\cos(2\phi_1 - \phi_2)$ to be a constant, as claimed. Note that if phase-matching is not satisfied, then constancy of the intensity does not imply constancy of the linear part of the energy density, and so things get more complicated. It turns out that these invariants are *much* easier to get at quantum mechanically as there it just corresponds to the fact that the expectation value of energy in a time-independent Hamiltonian is conserved!

implying that

$$\partial_z |A_2| = -\frac{\omega_2^2 d_{\rm eff}}{k_2 c^2} |A_1|^2 \sin(2\phi_1 - \phi_2). \tag{III.7}$$

This may be expressed in terms of $|A_2|$ alone via

$$\partial_z |A_2| = -\frac{\omega_2^2 d_{\rm eff}}{k_2 c^2} (C - |A_2|^2) \sin(2\phi_1 - \phi_2), \tag{III.8}$$

where under phase-matching we have $|A_1|^2 + |A_2|^2 = C = |A_1(0)|^2$. Note that the nonlinear invariant vanishing tells us that the sine is either 1 or -1. It is clear that since at first, the intensity in I_2 must increase (intensity is nonnegative), the sine must be -1. We then find that

$$\int_0^{|A_2(z)|} \frac{dx}{|A_1(0)|^2 - x^2} = \frac{2\omega_1 d_{\rm eff}}{n_1 c} z, \tag{III.9}$$

which using standard logarithmic integral identities can be reduced to

$$|A_2(z)| = |A_1(0)| \tanh\left(\frac{\omega_1 d_{\rm eff} |A_1(0)|}{n_1 c} z \right). \tag{III.10}$$

In this case, the second-harmonic field is capable of capturing all of the incident intensity, corresponding to complete conversion of the fundamental wave. This behavior is in some sense non-representative of the full solution space: if the nonlinear invariant is finite, or if phase-matching is not satisfied, one will have an oscillatory power exchange between the fundamental and the second harmonic.

IV. DIFFERENCE-FREQUENCY GENERATION AND PARAMETRIC GENERATION

Maxwell's equations in a lossless medium (even in the nonlinear case) are reversible: playing the solutions backwards generates another valid solution[14]. That suggests that if we take a phenomenon like sum-frequency generation, and play it in reverse, we would see an effect in which a field at frequency ω_3 generates fields at frequencies

[14] Why lossless? Consider an absorbing medium which extracts energy irreversibly from a wave in time. If we play the dynamics backwards, the wave gets amplified, which only occurs if the medium is active. If we watch a movie of the time-reversed solution but know that all of our materials are passive, then the resulting solution is invalid.

ω_1, ω_2 ($\omega_1 < \omega_2 < \omega_3$): this is called difference-frequency generation (or non-degenerate parametric down-conversion, depending on initial conditions). Similarly, if we watch second-harmonic generation in reverse, we would see a field at frequency $2\omega_1$ generate fields at frequency ω_1, an effect called degenerate parametric down-conversion. Let us study the case of non-degenerate parametric down-conversion.

The relevant coupled amplitude equations are *exactly* those describing sum-frequency generation. Let us suppose however that there is a strong incident wave at frequency ω_3 which is treated as un-depleted. The fields A_1 and A_2 are "weak" and their spatial variations are taken to be non-negligible. Then, the resulting linear equations for A_1, A_2 are:

$$\partial_z A_1(z) = \frac{2i\omega_1^2 d_{\text{eff}}}{c^2 k_1} A_3 A_2^* e^{-i\Delta k z} \tag{IV.1}$$

$$\partial_z A_2(z) = \frac{2i\omega_2^2 d_{\text{eff}}}{c^2 k_2} A_3 A_1^* e^{-i\Delta k z}, \tag{IV.2}$$

where $\Delta k = k_1 + k_2 - k_3$. These equations are *almost* directly solvable linear equations. The "issue" is that the fields couple to their conjugates, which are independent variables[15]. This can of course be quickly fixed by conjugating the second equation, yielding:

$$\partial_z A_1(z) = \frac{2i\omega_1^2 d_{\text{eff}}}{c^2 k_1} A_3 A_2^* e^{-i\Delta k z} \tag{IV.3}$$

$$\partial_z A_2^*(z) = -\frac{2i\omega_2^2 d_{\text{eff}}}{c^2 k_2} A_3^* A_1 e^{i\Delta k z}. \tag{IV.4}$$

We can solve this in exactly the same way that we solved sum-frequency generation with one strong undepleted wave. We define $A_1 = \tilde{A}_1 e^{-i\Delta k z/2}$ and $A_2^* = \tilde{A}_2^* e^{i\Delta k z/2}$ and convert into a matrix differential equation with constant coefficients. It immediately follows that

$$\tilde{A}_1(z) = J_1 \cosh(\lambda z) + K_1 \sinh(\lambda z), \tag{IV.5}$$

while

$$\tilde{A}_2^*(z) = J_2 \cosh(\lambda z) + K_2 \sinh(\lambda z), \tag{IV.6}$$

[15] This is very important. We know that for complex variables, we can treat the real and imaginary parts of the variable as independent. Therefore, we can also treat a complex variable and its conjugate as independent *single variables* since they are related to the real and imaginary part by a linear transformation!

with

$$\lambda = \pm\sqrt{\frac{4\omega_1^2\omega_2^2 d_{\text{eff}}^2}{k_1 k_2 c^4}|A_3|^2 - \frac{(\Delta k)^2}{4}}. \tag{IV.7}$$

If we take for example that $A_2(0) = 0$, then matching boundary conditions would give

$$A_1(z) = \left[\cosh(\lambda z) + \frac{i\Delta k}{2\lambda}\sinh(\lambda z)\right] A_1(0) e^{-i\Delta kz/2}, \tag{IV.8}$$

and

$$A_2^*(z) = \left[-\frac{2i\omega_2^2 d_{\text{eff}} A_3^*(0)}{c^2 k_2 \lambda}\sinh(\lambda z)\right] A_1(0) e^{i\Delta kz/2}. \tag{IV.9}$$

In the phase-matched case, this reduces to

$$A_1(z) = A_1(0)\cosh(\lambda z), \tag{IV.10}$$

and

$$A_2(z) = i\sqrt{\frac{n_1\omega_2}{n_2\omega_1}} e^{i\phi_3} A_1^*(0)\sinh(\lambda z), \tag{IV.11}$$

where $\phi_3 = A_3/|A_3|$. For distances $z \gg \lambda$, we may easily see that the intensities of both lower-frequency waves grows exponentially as $e^{2\lambda z}$, representing a type of exponential gain not unlike that of a laser amplifier. Unlike the laser amplifier however, the energy of the amplified wave is not extracted from excited electrons in atoms or semiconductors, but instead by extracting energy from another light wave coherently through nonlinear polarization[16].

V. PARAMETRIC OSCILLATION

The fact that we can amplify waves using parametric interactions forms the basis for a power type of laser-like light source called the optical parametric oscillator (OPO). Such parametric oscillators have become extremely important alternatives to "conventional" lasers based on excited media due to their very widely tunable wavelengths, and have accordingly made an important commercial impact. In this section, we will analyze some of the basic behaviors of the parametric oscillator.

[16] One might wonder how: if this process is related by inverse to sum-frequency generation, why do we have exponential amplification rather than oscillation? We leave this question for the reader to think about!

Let us consider a resonator formed by two mirrors of complex amplitude reflectivities r_1 and r_2, defined such that the transmitted energy fraction is $T_i = 1 - R_i \equiv 1 - |r_i|^2$. Inside this resonator is a second-order nonlinear medium. We assume that the mirrors are perfectly transmissive for the pump field at frequency ω_3. If the mirrors are only highly reflective for one of the lower-frequency waves, we call that wave the signal, and the OPO is a singly-resonant OPO. If the mirrors are highly reflective for both lower-frequency waves, the OPO is doubly-resonant and the name signal is used for the frequency we are more interested in, and idler for the frequency we are less interested in.

If the mirror reflectivities are high for the signal, then the signal can bounce around in this resonator (also called a cavity) many times, effectively experiencing an enhanced length of nonlinear interaction and enabling more efficient extraction of energy from the pump. At the same time, because there is a finite transmission of the mirrors, part of the signal gets lost each round-trip (from each mirror). If the parametric gain is equal to the loss, then the signal can reproduce itself every round-trip, giving a stable oscillation. When this happens, we call the resulting device a *parametric oscillator*. The pump power needed for the gain to *equal* the loss is called the *threshold pump power*. When the pump power exceeds the threshold value, the gain at first exceeds the loss and the signal builds up (as does potentially the idler field if the cavity is doubly-resonant). However, once these fields start to build up appreciably, we can no longer take the pump to be un-depleted. As the pump starts to deplete, the available gain decreases until the gain is equal to the round-trip loss, leading to stable oscillation. This is the mechanism of *gain saturation*.

If you know a little bit about conventional lasers based on the inversion of a gain medium, the story is in fact somewhat similar. Once we pump the gain medium sufficiently to achieve sufficient population inversion, the small-signal gain is greater than the round-trip loss and the field builds up until the population inversion depletes due to strong stimulated emission. At equilibrium, the depleted or saturated gain is equal to the loss. Of course, the major difference with the parametric oscillator is that there is no physical gain medium that we invert with a source of energy! Here, the $\chi^{(2)}$ medium mediates energy transfer between different modes of the electromagnetic field.

In what follows, we will develop a theory of the threshold of the parametric oscillator in the singly- and doubly-resonant regimes. Then, we will work out a theory of saturation. Let us study the behavior of a signal and idler field A_s, A_i that pass through a $\chi^{(2)}$ crystal which also has a pump field A_p. We will consider case of phase-matching for simplicity since this leads to the lowest threshold pump power.

If the pump-field is non-depleted, then a small extension of the previous section would give (mapping $A_1, A_2, A_3 \to A_s, A_i, A_p$):

$$\begin{aligned} A_s(z) &= \left[A_s(0)\cosh(\lambda z) + \frac{2i\omega_s^2 d_{\text{eff}}}{c^2 k_s \lambda} A_p A_i^*(0)\sinh(\lambda z)\right] \\ A_i^*(z) &= \left[A_i^*(0)\cosh(\lambda z) - \frac{2i\omega_i^2 d_{\text{eff}}}{c^2 k_i \lambda} A_p^* A_s(0)\sinh(\lambda z)\right], \end{aligned} \tag{V.1}$$

where $\lambda^2 = \frac{4\omega_1^2\omega_2^2 d_{\text{eff}}^2}{k_1 k_2 c^4}|A_3|^2$.

After passing through the crystal, they propagate to the mirror, and reflect off, each picking up complex amplitudes $-r_{s,1}$ and $-r_{i,1}$ respectively[17]. They then propagate backwards. We assume that there may be absorption losses $\alpha_{s,i}$ per unit length associated with propagation in the cavity. When they propagate backwards through the crystal, there is no gain associated with that (reader: why?). They then bounce off the second mirrors, picking up $-r_{s,2}$ and $-r_{i,2}$, completing their round trip. The electric fields therefore map to:

$$\begin{aligned} E_s(0) &\to r_{s,1} r_{s,2} e^{2ik_s d} e^{-\alpha_s d}\left[A_s(0)\cosh(\lambda L) + \frac{2i\omega_s^2 d_{\text{eff}}}{c^2 k_s \lambda} A_p A_i^*(0)\sinh(\lambda L)\right] \\ E_i^*(0) &\to r_{i,1} r_{i,2} e^{2ik_i d} e^{-\alpha_i d}\left[A_i^*(0)\cosh(\lambda L) - \frac{2i\omega_i^2 d_{\text{eff}}}{c^2 k_i \lambda} A_p^* A_s(0)\sinh(\lambda L)\right], \end{aligned} \tag{V.2}$$

where d is the cavity length and L is the crystal length[18].

---

[17] I am using the convention $\begin{pmatrix} -r & it \\ it & -r \end{pmatrix}$ for the scattering matrix. This is the convention used for example in Waves and Fields in Optoelectronics by Haus. The results of course do not change with a different convention.

[18] You'll notice that I went from envelope to field. This is because in principle, the wavevector can shift due to the mirrors and the crystal, changing the resonance condition for the cavity. We will ignore this effect anyway as it is not important for the gain threshold at this level of treatment, but it is important to know that this is an effect that can occur. We will assume that k_s, k_i correspond to resonances of the cavity so that $2k_i d, 2k_s d \pmod{2\pi} = 0$. This assumes that the ks that we are using take into account the fact that the refractive index is a weighted average of the crystal refractive index and the air/vacuum surrounding it.

Let us define $\ell_s = 1 - r_{s,1}r_{s,2}e^{-\alpha_s d}, \ell_i = 1 - r_{i,1}r_{i,2}e^{-\alpha_i d}$ as the amplitude losses associated with round trip propagation. The threshold condition is that the fields reproduce themselves after this round trip. This gives the conditions

$$\cosh \lambda L = \left[1 + \frac{\ell_i \ell_s}{2 - \ell_i - \ell_s}\right]. \tag{V.3}$$

Typically, the gain per pass is low and we may approximate the left-hand side by $1 + \frac{1}{2}(\lambda L)^2$ giving

$$(\lambda L)^2 \approx \frac{2\ell_i \ell_s}{2 - \ell_i - \ell_s}. \tag{V.4}$$

In the doubly-resonant case, both losses are small compared to one and we can further simplify to

$$(\lambda L)^2 \approx \ell_i \ell_s \tag{V.5}$$

as the resonant condition.

In the singly-resonant case, we can take $A_i(0) = 0$ and it is straightforward to see that

$$1 = (1 - \ell_s) \cosh \lambda L \implies (\lambda L)^2 = 2\ell_s. \tag{V.6}$$

We see then that in the doubly-resonant case, the threshold is lower by a factor of $\sqrt{\ell_i/2}$, and so it is in principle easier to get a doubly-resonant oscillator to oscillate.

However, this comes at a cost. In a real cavity there are many modes with low loss within the gain bandwidth (defined as the range of frequencies such that $\Delta k L < 1$). In the singly-resonant case, the signal mode that oscillates is typically the one with the largest gain. The idler mode automatically satisfies $\omega_i = \omega_p - \omega_s$ and never builds up to an appreciable amplitude so there are no real constraints on the idler (it comes along for free).

In the doubly-resonant case, we want both signal and idler to build up: this means there need to be resonance modes at frequencies ω_s and ω_i such that $\omega_s + \omega_i = \omega_p$. In general, this constraint is hard to satisfy with multiple pairs of ω_s, ω_i and there will typically only be one pair that satisfies the criterion[19]. As a result of this, the

[19] I have assumed the cavity modes are not equally spaced, or that the signal and idler modes have incommensurate mode spacings (free-spectral ranges). In practice, modes are unequally spaced due to the frequency dependence of the refractive index $n(\omega)$ making it such that the frequency spacing between modes satisfies $\omega_m = \frac{ck_m}{n(\omega_m)}$ where $k_m = m\pi/L$. If we take $m \to m+1$ then we get $\Delta\omega_m \frac{d(n\omega)}{d\omega}\Big|_{\omega_m} = \pi c/L \implies \Delta\omega_m = (\pi c/L)/\frac{d(n\omega)}{d\omega}\Big|_{\omega_m}$.

primary determinant of which modes oscillate is less which mode has the highest gain, but which modes can even satisfy the energy-conservation requirement $\omega_s + \omega_i = \omega_p$. Because of the low threshold associated with the doubly-resonant case, the reduced gain by being somewhat phase-mismatched is less problematic. However, what is a problem in the doubly resonant case is that if the cavity properties *change* due to fluctuations in the cavity length of a mechanical or thermal nature, or fluctuations in the pump frequency itself, the cavity frequencies will change. If there is a new pair of signal and idler frequencies ω_s', ω_i' such that $\omega_s' + \omega_i' = \omega_p$, it in general will not be next to ω_s, ω_i due to the unequal cavity mode spacings - leading to mode jumping or hopping, which is undesirable. The doubly resonant geometry amplifies frequency noise [20].

A. Track 2: Above-threshold behavior of the parametric oscillator

This section is focused on a detailed treatment of the parametric oscillator. This material is labeled Track 2 as it is more detailed than we normally are in this class. For even more information on this important topic, see Harris, Stephen E. "Tunable optical parametric oscillators." Proceedings of the IEEE 57.12 (1969): 2096-2113. In what follows, we will deal only with the singly resonant case as such systems are notably more stable than their doubly-resonant counterparts. Assume that the OPO is resonant for the signal: then we expect the idler field to be small. The idea is that the signal builds up in the resonator until there's sufficient gain to build up a significant signal and idler field. Further, we assume, as before, that the OPO is not resonant for the pump. Therefore, the pump field at the entrance facet of the

[20] Suppose that the pump-frequency varies by $\delta\omega_p$. Then of course, by energy conservation $\delta\omega_p = \delta\omega_s + \delta\omega_i$. We expect that the new signal and idler frequency will, at threshold, have the same gain as the original signal and idler frequency. Why? Because, at threshold, the gain must balance the loss and the loss does not change much with frequency. Since the gain is largely determined by the phase-mismatch, we can enforce that $\delta(\Delta k) = \frac{\partial k_i}{\partial \omega_i}\delta\omega_i + \frac{\partial k_s}{\partial \omega_s}\delta\omega_s - \frac{\partial k_p}{\partial \omega_p}\delta\omega_p$. Using $\delta\omega_i = \delta\omega_p - \delta\omega_s$ and $\frac{\partial k}{\partial \omega} = \frac{n_g}{c}$ with n_g the group index derived in the previous footnote, one finds: $\delta\omega_s = \frac{n_{g,p} - n_{g,i}}{n_{g,s} - n_{g,i}}\delta\omega_p$. This result among other things tells us that an operator operated near degeneracy (signal and idler frequencies nearly equal) leads to a huge linewidth. This result is derived in Kovrigin, A., and R. Byer. "Stability factor for optical parametric oscillators." IEEE Journal of Quantum Electronics 5.7 (1969): 384-385 - although it is brief!

crystal can be taken as a constant equal to $A_p(0)$. The coupled amplitude equations are exactly the equations of difference-frequency generation.

Let us analyze the properties of the signal field assuming that the system is oscillating in a steady-state. Then, assuming that the change in the signal field is small as it passes through the crystal, we can take its magnitude at the entrance to the crystal as a known value A_s and ask about the dynamics of the pump and idler fields. The result after a length L of propagation in the crystal is, assuming phase-matching $\Delta k = 0$[21]

$$\begin{aligned} A_p(L) &= A_p(0)\cos(\lambda L) \\ A_i(L) &= \frac{2id_{\text{eff}}\omega_i}{n_i c\lambda} A_s^* A_p(0)\sin(\lambda L). \end{aligned} \tag{V.7}$$

We can use this to then calculate the change in the signal itself using

$$\partial_z A_s = \frac{2i\omega_s d_{\text{eff}}}{n_s c} A_p(z) A_i^*(z) = \frac{2\omega_i\omega_s d_{\text{eff}}^2}{n_i n_s c^2\lambda} |A_p(0)|^2 A_s \sin(2\lambda z) \tag{V.8}$$

Now we will integrate this equation[22], yielding

$$A_s(L) = \exp\left[\frac{2\omega_i\omega_s d_{\text{eff}}^2}{n_i n_s c^2}|A_p(0)|^2\frac{\sin^2\lambda L}{\lambda^2}\right] A_s(0). \tag{V.9}$$

The overall evolution of the signal field after a round trip involves an additional loss such that after a round trip:

$$A_s \to (1-\ell_s)\exp\left[\frac{2\omega_i\omega_s d_{\text{eff}}^2}{n_i n_s c^2}|A_p(0)|^2\frac{\sin^2\lambda L}{\lambda^2}\right] A_s(0). \tag{V.10}$$

We have neglected here the propagation phase assuming $2k_s d \pmod{2\pi} = 0$. The oscillation condition is, assuming weak gain per pass, such that the argument of the exponent can be treated as small:

$$\frac{2\omega_i\omega_s d_{\text{eff}}^2}{n_i n_s c^2}|A_p(0)|^2\frac{\sin^2\lambda L}{\lambda^2} = \ell_s \tag{V.11}$$

This is the condition that amplitude gain equals amplitude loss. Recall that in this expression $\lambda = \frac{2d_{\text{eff}}}{c}\sqrt{\frac{\omega_p\omega_i}{n_p n_i}}|A_s|$.

[21] Recall from the discussion above that for the singly-resonant oscillator, it likes to oscillate at a signal mode where the gain is highest.

[22] This may look fishy since in deriving the pump and idler fields, we said that A_s is constant over the crystal. However, its amplitude changes only weakly over the crystal and so there is little approximation error in the pump and idler fields themselves!

Just above threshold, when the gain is just barely enough to match the loss (in the absence of depletion effects), we expect the steady state signal field to be small (and to go to zero exactly at threshold). In that case, our equilibrium condition is

$$\frac{2\omega_i\omega_s d_{\text{eff}}^2}{n_i n_s c^2}|A_p(0)|^2 L^2 = \ell_s \tag{V.12}$$

which is equivalent to the singly-resonant threshold condition introduced in the previous section. Above threshold, as we increase the pump power, the only way this equality is satisfied is if $\lambda \neq 0$ corresponding to a finite value of the signal field. The signal field is determined as a function of the pump power by the solution of this equation! The origin of this reduction of gain at finite signal field is depletion of the pump. We see that if the λL becomes appreciable, then $A_p(L) = A_p(0)\cos(\lambda L)$ becomes smaller than $A_p(0)$ (hence depletion). We have thus worked out an elementary treatment of the above threshold behavior of the parametric oscillator.

To conclude this analysis, we consider the question of how much energy is extracted from the *idler*. This can be straightforwardly calculated: the intensity of the idler generated is

$$I_i = 2\epsilon_0 n_i c|A_i|^2 = \frac{4\epsilon_0 n_s c\omega_i}{\omega_s}\ell_s|A_s|^2. \tag{V.13}$$

The intensity of the signal exiting the cavity is the sum of the intensities exiting either side of the cavity and so

$$I_s = 2\ell_s \times 2\epsilon_0 n_s c|A_s|^2, \tag{V.14}$$

assuming that all losses from the cavity go into the beam exiting the cavity. This implies that

$$I_i = \frac{\omega_i}{\omega_s} I_s. \tag{V.15}$$

At the same time, we also know from the Manley-Rowe relation that every idler photon generated corresponds to the removal of a photon from the pump, and so we have that

$$\frac{I_p(L)}{\omega_p} - \frac{I_i(L)}{\omega_i} = \frac{I_p(0)}{\omega_p}. \tag{V.16}$$

If the pump power is chosen such that it is fully depleted at $z = L$ then $\frac{\omega_i}{\omega_p} I_p(0)$ is the maximum power that can be in the idler.

VI. ELECTRO-OPTIC EFFECT

The last topic that we'll discuss in our tour of effects mediated by $\chi^{(2)}$ is the electro-optic effect, and in particular, the change in the refractive index of light by static voltages. This is called the *electro-optic* effect, and it allows us to use voltages to change the phase of light accumulated by light of different orthogonal polarizations, which can be used as the basis for *electro-optic modulators.*

The modification of the refractive index by a static voltage can be understood readily in the case of an instantaneous $\chi^{(2)}$ medium, where the polarization is given by

$$P_i^{(2)}(t) = 2\epsilon_0 d_{\mathrm{ijk}} E_j(t) E_k(t). \tag{VI.1}$$

If we consider a field which is a superposition of a DC field and a monochromatic field,

$$E = E_{\mathrm{DC}} + E_0 e^{-i\omega t} + \mathrm{c.c}, \tag{VI.2}$$

then we see that there is a contribution at the same frequency at the monochromatic field to the polarization:

$$4\epsilon_0 d_{\mathrm{ijk}} E_{\mathrm{DC},k} E_{0,j} e^{-i\omega t}, \tag{VI.3}$$

which looks like a change in the susceptibility [23]. Notice however that the change in the susceptibility

$$\delta\chi_{ij}^{(1)} = 4 d_{ijk} E_{\mathrm{DC},k} \tag{VI.4}$$

is tensorial.

In the literature on the electro-optic effect, it is conventional to talk about the change in the inverse permittivity rather than the permittivity itself. The inverse of the relative permittivity $\boldsymbol{\epsilon}$ is denoted $\boldsymbol{\eta}$ and its changes are related to changes in $\boldsymbol{\epsilon}$ by [24]

$$\delta\boldsymbol{\eta} = -\boldsymbol{\eta}\delta\boldsymbol{\epsilon}\boldsymbol{\eta}. \tag{VI.5}$$

---

[23] I have used the fact that for an instantaneous lossless medium, I can interchange indices of d freely.

[24] The result below can be found as follows. Consider $(A+\delta A)^{-1}$ where $A, \delta A$ are matrices. Their inverse corresponds to solving $(A+\delta A)x = b$. Suppose that when $\delta A = 0$, $x_0 = A^{-1}b$. When $\delta A \neq 0$ but is small, we can seek a perturbative solution of the form $(A+\delta A)(x_0+\delta x) \approx Ax_0 + A\delta x + \delta A x_0 = b \implies \delta x = -A^{-1}\delta A A^{-1} b$. Therefore if $x = x_0 + \delta x \approx A^{-1} - A^{-1}\delta A A^{-1} b$, we may say that $\delta A^{-1} = A^{-1}\delta A A^{-1}$.

Using $\delta\chi = \delta\epsilon$ we may write

$$\delta\eta_{ij} = r_{ijk} E_{\mathrm{DC},k}, \tag{VI.6}$$

where

$$r_{ijk} = -4\epsilon_0 \eta_{ia} \eta_{bj} d_{abk}. \tag{VI.7}$$

Similar to the d matrix, there is a contracted notation for r which follows from the symmetry of η and full permutation symmetry of d. In particular, we can interchange the indices i, j meaning that there are only six independent pairs of (i, j). The mapping to contracted notation is $(1,1) \to 1, (2,2) \to 2, (3,3) \to 3, (2,3)/(3,2) \to 4, (3,1)/(1,3) \to 5, (1,2)/(2,1) \to 6$. The r matrix is usually represented as

$$r = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \\ r_{41} & r_{42} & r_{43} \\ r_{51} & r_{52} & r_{53} \\ r_{61} & r_{62} & r_{63} \end{pmatrix}. \tag{VI.8}$$

The nonzero elements are simply related to the nonzero elements of η and d through the relation $r_{ijk} = -4\epsilon_0 \eta_{ia} \eta_{bj} d_{abk}$.

Let us now explore the ramifications of the electro-optic effect. Consider for example a crystal like KDP of the $\bar{4}2m$ crystal class. It has three non-zero r elements: $r_{41} = r_{52} = 8.77$ pm/V and $r_{63} = 10.5$ pm/V. Suppose that the crystal is oriented such that light propagates along the z direction, parallel to one principal axis of KDP. The other two crystal axes are orthogonal to z. The c-axis of the crystal is taken to be in the z direction. Suppose now that we apply a voltage along the z axis ($\mathbf{E}_{\mathrm{DC}} = E_z \hat{z}$ with $E_z = -(V/L)$ where V is the voltage difference and L is the length of crystal.). The change in the inverse permittivity is then

$$\delta\eta_{ij} = \begin{pmatrix} 0 & \Delta & 0 \\ \Delta & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{VI.9}$$

where

$$\Delta = r_{63} E_z. \tag{VI.10}$$

This changes the principal axes in a meaningful way even for infinitesimal E_z. Of course z is still a principal axis of the system. Meanwhile in the xy-plane, the principal axes of η are found by diagonalizing

$$\eta_o I + \begin{pmatrix} 0 & \Delta \\ \Delta & 0 \end{pmatrix}, \tag{VI.11}$$

which has eigenvalues and corresponding eigenvalues

$$\eta_o + r_{63}E_z, \eta_o - r_{63}E_z \leftrightarrow \hat{u}_1 \equiv \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \hat{u_2} \equiv \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{VI.12}$$

We may map these changes in η to changes in the principal values of the index of refraction by using

$$\delta\eta_o = -\frac{2\delta n}{n^3}. \tag{VI.13}$$

Therefore, for the direction 45 degrees counterclockwise to the x-axis, the index is $n_o - \frac{1}{2}n_o^3 r_{63}E_z$ (and a wave propagates faster, making the x-axis the "fast axis"), while for the direction 45 degrees counterclockwise to the y-axis, the index is $n_o + \frac{1}{2}n_o^3 r_{63}E_z$ (making the y-axis the slow-axis)[25]

Therefore, using the electro-optic effect, we can induce birefringence for light polarized in different directions in the xy plane, inducing different phase shifts for light polarized along orthogonal directions. Using this, we can construct a powerful device called an electro-optic modulator that lets us impart a voltage-dependent change in the amplitude or phase of light.

A. Electro-optic modulators

Consider what happens when we send light through a polarizer oriented along the x direction, and then send it into a KDP crystal with a z-oriented electric field. The initial complex field

$$E = E_0\hat{x} = E_0\frac{\hat{u}_1 + \hat{u}_2}{\sqrt{2}} \tag{VI.14}$$

[25] Your instructor strongly believes that Boyd has the wrong sign here. The sign your instructor arrived at is the same as in Haus ("Waves and Fields in Optoelectronics") and Yariv & Yeh ("Optical Electronics in Modern Communications.")

will, after a length L evolve into

$$E(L) = E_0 e^{i\phi_0} \frac{\hat{u}_1 + \hat{u}_2 e^{i\Gamma}}{\sqrt{2}}, \tag{VI.15}$$

where $\phi_0 = n_0 \frac{\omega L}{c} - \frac{1}{2} n_0^3 r_{63} E_z \frac{\omega L}{c}$ and Γ the *retardance* is given by

$$\Gamma = n_0^3 r_{63} E_z \frac{\omega L}{c} = \frac{2\pi}{\lambda} n_0^3 r_{63} E_z L = \frac{2\pi}{\lambda} n_0^3 r_{63} |V|. \tag{VI.16}$$

As a natural result of this birefringence, the components of the polarization along different axes achieve different rotations. Now, if we go back into the xy coordinate system, we get:

$$E_x(L) = E_0 e^{i\phi_0} \frac{1 + e^{i\Gamma}}{2}, \tag{VI.17}$$

and

$$E_y(L) = E_0 e^{i\phi_0} \frac{1 - e^{i\Gamma}}{2}. \tag{VI.18}$$

we see that the polarization is rotated. To get at the rotation angle, suppose we put a polarizer along the x-axis. The resulting intensity, relative to the incident intensity $I_0 \sim |E_0|^2$ would be

$$I = I_0 \cos^2(\Gamma/2). \tag{VI.19}$$

Comparing this to Malus' law, we see that the polarization rotation angle is $\Gamma/2$.

The intensity modulator works by terminating the propagation chain with the polarization in the y direction rather than x, in which the output intensity is

$$I = I_0 \sin^2(\Gamma/2). \tag{VI.20}$$

The principle of the intensity modulator is that changing an applied voltage changes the intensity of light, and if we modulate the voltage in time, we can modulate the intensity in time. For example by having a square voltage wave, we can create a square wave of light intensity that serves as optical bits transmitting "0"s or "1"s. Hence, such intensity modulators are of great importance in optical communications.

In the intensity modulator, it is typically important to operate at a point where a smaller change in the voltage leads to a larger change in the intensity. That requires operating around the voltage such that $\Gamma = \pi/2$, which corresponds to the voltage

$$\frac{4}{\lambda} n_0^3 r_{63} |V| = 1 \implies |V| = \frac{\lambda}{4 n_0^3 r_{63}}. \tag{VI.21}$$

This is related to a figure of merit called the *half-wave voltage* which is the voltage needed to get a π phase shift (a half-wavelength's worth of phase shift). That voltage is simply

$$V_{\lambda/2} = \frac{\lambda}{2n_0^3 r_{63}}. \tag{VI.22}$$

Typical magnitudes of this voltage are on the order of 10 kV and require high-voltage supplies!

Finally, we mention that if we send light polarized along one of the new principal axes, say $\hat{u}_1$, then light just experiences a phase shift proportional to the voltage. By varying that voltage in time, we modulate the *phase* directly, which is important for communication based on the optical phase of light.