2. Ultrashort Laser Pulses

Rick Trebino and Erik Zeek

The Basics

What exactly is an ultrashort laser pulse anyway? Quite simply, it's a very very short burst of electro-magnetic energy.

The pulse, like any light wave, is defined by its electric field as a function of space and time, $\mathcal{E}(x, y, z, t)$. You may be more familiar with a continuous beam, whose electric field is sinusoidal in time. The difference is that an ultrashort pulse comprises only a few cycles of a sine wave (more precisely, less than about a million for visible light). Indeed, our expression for an ultrashort pulse will be the product of a sine wave and a pulse-envelope function. So ultrashort laser pulses are not really much different from other types of laser light, just shorter. A lot shorter.

New issues do arise, however, in dealing with ultrashort pulses, and, in particular, in measuring them. For example, the shorter the pulse, the broader its spectrum, that is, the greater the range of colors (the *bandwidth*) present. And, despite the incredibly short duration of these pulses, the color can change rapidly during one. Indeed, the pulse can begin as one color and end as quite another. Simply passing through a material—even air—can modify the color variation of a pulse in time. We'll need to be able to measure this variation—which is contained in the pulse *phase*—as well as variations in the pulse intensity.

We won't concern ourselves with how such pulses are created, a subject that could fill another entire book (and has! [1–4]). Their measurement will prove adequate subject matter for us.

The Intensity and Phase vs. Time

For the sake of simplicity, we'll treat the electric field as linearly polarized, so we need consider only one component of it. This is called the *scalar approximation*, in which we ignore the pulse electric field's vector character. The electric field of the pulse can potentially be a complicated function of space and time, but, as we're mainly interested in the temporal features of the pulse, we'll ignore the spatial portion of the field and write the temporal dependence of the pulse electric field as:

$$\mathcal{E}(t) = \tfrac{1}{2}\sqrt{I(t)}\,\exp\{i\,[\omega t - \phi(t)]\} + c.c. \tag{2.1}$$

where t is time in the reference frame of the pulse, ω_0 is a carrier angular frequency on the order of 10^{15} sec^{-1}, and $I(t)$ and $\phi(t)$ are the time-dependent intensity and phase of the pulse.

Notice that we've removed the rapidly varying *carrier wave* $\exp(i\omega_0 t)$ from the intensity and phase. This saves us the trouble of plotting all the oscillations of the pulse field.

Sometimes, we refer to $I(t)$ and $\phi(t)$ as the *temporal* intensity and phase of the pulse to distinguish them for their *spectral* counterparts that we'll define next. We assume that, despite their ultrafast nature, $I(t)$ and $\phi(t)$ vary slowly compared to $\exp(i\omega_0 t)$—a good assumption for all but the shortest pulses. As usual, "c.c." means *complex conjugate* and is required to make the pulse field real. But, in this book (as in most other publications), we'll make what's called the *analytic signal* approximation and ignore the complex-conjugate term. This yields a complex pulse field, but it simplifies the mathematics significantly.

We refer to the *complex amplitude* of this wave as:

$$\boxed{E(t) \equiv \sqrt{I(t)}\exp[-i\phi(t)]} \tag{2.2}$$

$E(t)$ is simply $\mathcal{E}(t)$ but without the "Re" and the rapidly varying $\exp(i\omega_0 t)$ factor and multiplied by 2. Equation (2.2) is the quantity we'll be measuring for the rest of this book. Some people refer to $\sqrt{I(t)}$ as the "amplitude," with the word "real" suppressed (see Fig. 2.1).

We can solve for the intensity, given the field:

$$\boxed{I(t) = |E(t)|^2} \tag{2.3}$$

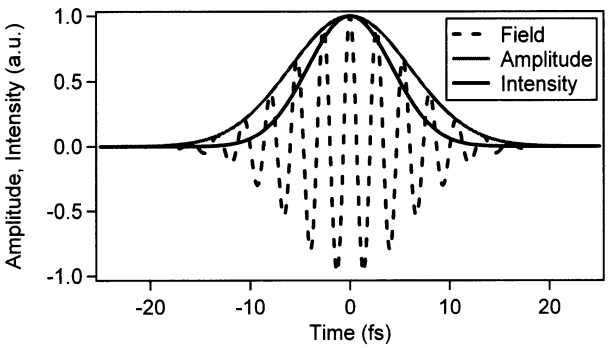

Fig. 2.1: The electric field, intensity, (real) amplitude, and intensity of a Gaussian pulse. The intensity of a Gaussian pulse is $\sqrt{2}$ shorter than its real amplitude. The phase of this pulse is a constant, $\phi(t) = 0$, and is not plotted.

where we don't care about the absolute magnitude of the intensity (the irradiance); instead we only care about the *shape*, so, in Eq. (2.3), we've omitted constants like the permittivity and the speed of light.

We can also solve for the phase:

$$\boxed{\phi(t) = -\arctan\left\{\frac{\mathrm{Im}[E(t)]}{\mathrm{Re}[E(t)]}\right\}} \tag{2.4}$$

An equivalent formula for the phase is:

$$\phi(t) = -\mathrm{Im}\{\ln[E(t)]\} \tag{2.5}$$

The Intensity and Phase vs. Frequency

The pulse field in the frequency domain is the Fourier transform the time-domain field, $\mathcal{E}(t)$:

$$\tilde{\mathcal{E}}(\omega) = \int_{-\infty}^{\infty} \mathcal{E}(t)\exp(-\mathrm{i}\omega t)\,\mathrm{d}t \tag{2.6}$$

where we'll use the tilde ($\tilde{}$) over a function to indicate that it's the Fourier transform. Also, the inverse Fourier transform is:

$$\mathcal{E}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{\mathcal{E}}(\omega)\exp(\mathrm{i}\omega t)\,\mathrm{d}\omega \tag{2.7}$$

Separating $\tilde{\mathcal{E}}(\omega)$ into its intensity and phase yields:

$$\boxed{\tilde{\mathcal{E}}(\omega) = \sqrt{S(\omega)}\exp[-\mathrm{i}\varphi(\omega)]} \tag{2.8}$$

where $S(\omega)$ is the spectrum and $\varphi(\omega)$ is the spectral phase. Note that, while the temporal phase (ϕ) and spectral phase (φ) are both called "phi," we've actually used different Greek characters to distinguish them. The spectrum and spectral phase typically have nonzero regions for both positive and negative frequencies (see Fig. 2.2). Because $\mathcal{E}(t)$ is real, the two regions contain equivalent information, so everyone always ignores the negative-frequency region.

We could've defined the spectrum and spectral phase in terms of the Fourier transform of the complex pulse amplitude, $E(t)$, rather than the entire field, $\mathcal{E}(t)$:

$$\tilde{E}(\omega - \omega_0) = \sqrt{S(\omega - \omega_0)}\exp[-i\varphi(\omega - \omega_0)] \tag{2.9}$$

where $S(\omega - \omega_0)$ would've been the spectrum, and $\varphi(\omega - \omega_0)$ would've been the spectral phase. These are the same functions as above, but the center

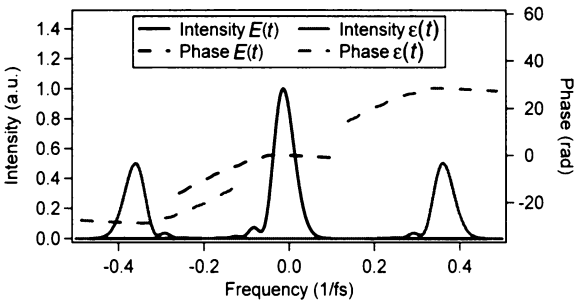

Fig. 2.2: The spectrum and spectral phase corresponding to the real pulse (gray) and the complex amplitude (black). Note that the real pulse spectrum has both positive and negative frequency components, centered on $+\omega_0$ and $-\omega_0$, respectively (in this plot, $\omega_0 \approx 0.38$/fs). The spectrum and spectral phase corresponding to the pulse complex amplitude have only one component, centered on zero frequency.

frequency of the spectrum and spectral phase would've been shifted to zero. Also, the negative-frequency component is *explicitly* removed in Eq. (2.9) because the complex conjugate does not occur in the complex field envelope (see Fig. 2.2). This is done occasionally, and a few plots in this book will use this definition.

Most the time, we won't do this simply because ultrafast optics researchers generally don't. We're sorry if it may be a bit confusing that the time-domain field in general use is the complex field envelope, while the frequency-domain field is the Fourier transform, not of the complex field envelope, but of the full real electric field (in which the negative frequency component is ignored). The reason for this usage is that people like their spectra centered on the actual center wavelength—not zero—but they don't like their temporal waveforms rapidly oscillating, as would be required to be rigorously consistent. Just memorize this, and don't complain; it's a lot easier than remembering all those PIN numbers banks keep sending you.

Notice that the spectrum is given by:

$$\boxed{S(\omega) = |\tilde{\mathcal{E}}(\omega)|^2} \tag{2.10}$$

The spectral phase is given by expressions analogous to those for the temporal phase:

$$\boxed{\varphi(\omega) = -\arctan\left\{\frac{\mathrm{Im}[\tilde{\mathcal{E}}(\omega)]}{\mathrm{Re}[\tilde{\mathcal{E}}(\omega)]}\right\}} \tag{2.11}$$

or, equivalently:

$$\varphi(\omega) = -\mathrm{Im}\{\ln[\tilde{\mathcal{E}}(\omega)]\} \tag{2.12}$$

Finally, the spectrum can also be written in terms of the wavelength. $S_\lambda(\lambda)$ and $S_\omega(\omega)$ can be quite different for broadband functions because, for example, the frequency range extending from zero to some very low frequency extends in wavelength from a finite wavelength out to infinity. So the spectrum plotted vs. wavelength must take on considerably lower values for such large wavelengths to make sense.

We must be able to transform between frequency and wavelength because theoretical work (involving Fourier transforms) uses the frequency, while experiments (involving spectrometers) use the wavelength. The phase vs. wavelength is related to the phase vs. frequency:

$$\varphi_\lambda(\lambda) = \varphi_\omega(2\pi c/\lambda) \tag{2.13}$$

since $\omega = 2\pi c/\lambda$, and where we've added subscripts to indicate the relevant domain (frequency or wavelength). This result simply rescales the phase. But because the frequency scale and wavelength scale aren't linearly related, the phase looks different in the two cases (see Fig. 2.3).

The spectrum is a little trickier. The easiest way to see how these two quantities are related is to note that the spectral energy is equal whether we calculate it vs. frequency or wavelength:

$$\int_{-\infty}^{\infty} S_\lambda(\lambda)\mathrm{d}\lambda = \int_{-\infty}^{\infty} S_\omega(\omega)\mathrm{d}\omega \tag{2.14}$$

Let's now rewrite the left side of this equation by transforming variables, $\omega = 2\pi c/\lambda$, and noting that $\mathrm{d}\omega = -2\pi c/\lambda^2 \mathrm{d}\lambda$. We have:

$$\int_{-\infty}^{\infty} S_\lambda(\lambda)\mathrm{d}\lambda = \int_{\infty}^{-\infty} S_\omega(2\pi c/\lambda)\frac{-2\pi c}{\lambda^2}\mathrm{d}\lambda \tag{2.15}$$

$$= \int_{-\infty}^{\infty} S_\omega(2\pi c/\lambda)\frac{2\pi c}{\lambda^2}d\lambda \tag{2.16}$$

This means that:

$$S_\lambda = S_\omega(2\pi c/\lambda)\frac{2\pi c}{\lambda^2} \tag{2.17}$$

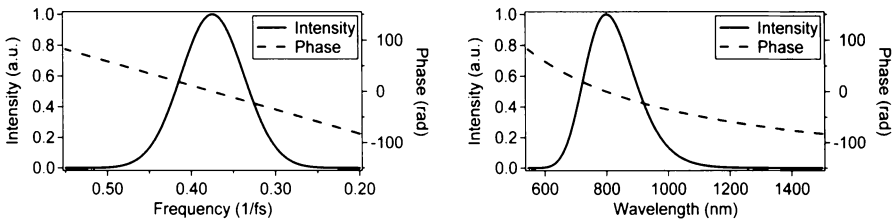

Fig. 2.3: Two identical spectra and spectral phases of a few-fs (i.e., broadband) pulse, plotted vs. frequency (left) and vs. wavelength (right). Note the different shapes of both curves, due to rescaling between frequency and wavelength.

The Phase, Instantaneous Frequency, and Group Velocity

The temporal phase, $\phi(t)$, contains frequency vs. time information, and the pulse *instantaneous angular frequency*, $\omega_{\text{inst}}(t)$, is defined as:

$$\omega_{\text{inst}}(t) \equiv \omega_0 - \mathrm{d}\phi/\mathrm{d}t \tag{2.18}$$

This is easy to see. At some time, t, consider the total phase of the wave. Call this quantity ϕ_0:

$$\phi_0 = \omega_0 t - \phi(t) \tag{2.19}$$

Exactly one period, T, later, the total phase will (by definition) increase to $\phi_0 + 2\pi$:

$$\phi_0 + 2\pi = \omega_0(t + T) - \phi(t + T) \tag{2.20}$$

where $\phi(t + T)$ is the slowly varying phase at the time, $t + T$. Subtracting Eq. (2.19) from Eq. (2.20):

$$2\pi = \omega_0 T - [\phi(t + T) - \phi(t)] \tag{2.21}$$

Dividing by T and recognizing that $2\pi/T$ is a frequency, call it $\omega_{\text{inst}}(t)$:

$$\omega_{\text{inst}}(t) = 2\pi/T = \omega_0 - [\phi(t + T) - \phi(t)]/T \tag{2.22}$$

But T is small, so $[\phi(t + T) - \phi(t)]/T$ is the derivative, $\mathrm{d}\phi/\mathrm{d}t$. So we're done!

Usually, however, we'll think in terms of the *instantaneous frequency*, $\nu_{\text{inst}}(t)$, so we'll need to divide by 2π:

$$\boxed{\nu_{\text{inst}}(t) = \nu_0 - [\mathrm{d}\phi/\mathrm{d}t]/2\pi} \tag{2.23}$$

We can write a Taylor series for the $\phi(t)$ about the time $t = 0$:

$$\phi(t) = \phi_0 + t\phi_1 + t^2\phi_2/2 + \cdots \tag{2.24}$$

where only the first few terms are required to describe well-behaved pulses.

While the temporal phase contains frequency vs. time information, the spectral phase contains time vs. frequency information. So we can define the *group delay vs. frequency*, $t_{\text{group}}(\omega)$, given by:

$$\boxed{t_{\text{group}}(\omega) = \mathrm{d}\varphi/\mathrm{d}\omega} \tag{2.25}$$

A similar derivation to the above one for the instantaneous frequency can show that this definition is reasonable. Also, we'll typically use this result, which is a real time (the rad's cancel out), and never $\mathrm{d}\varphi/\mathrm{d}\nu$, which isn't. Lastly, always remember that $t_{\text{group}}(\omega)$ is *not* the inverse of $\omega_{\text{inst}}(t)$.

It's also common practice to write a Taylor series for $\varphi(\omega)$:

$$\boxed{\varphi(\omega) = \varphi_0 + (\omega - \omega_0)\,\varphi_1 + (\omega - \omega_0)^2\,\varphi_2/2 + \cdots} \tag{2.26}$$

where, as in the time domain, only the first few terms are typically required to describe well-behaved pulses. Of course, we'll want to measure badly behaved pulses, which have higher-order terms in $\phi(t)$ and $\varphi(\omega)$.

Unfortunately, these definitions aren't completely satisfying. In particular, they don't always correspond to our intuitive ideas of what the instantaneous frequency and group delay should be for light. Consider the simple case of light with two frequencies:

$$\mathcal{E}(t) = \exp(\mathrm{i}\omega_1 t) + \exp(\mathrm{i}\omega_2 t) + \text{c.c.} \tag{2.27}$$

Recalling that this is a simple case of "beats," the instantaneous frequency obtained by the definition given above is:

$$\omega_{\text{inst}}(t) = (\omega_1 + \omega_2)/2 \tag{2.28}$$

a frequency that never actually occurs in the beam (only ω_1 and ω_2 do). But, for most ultrashort-pulse applications, there's a broad continuous range of frequencies, and the above definitions prove reasonable.

Phase Distortions in Time and Frequency

Phase Wrapping, Unwrapping, and Blanking

Before we discuss the various phase distortions that occur in ultrashort pulses, we should mention a couple of points that you should always keep in mind when you deal with the phase.

First, because $\exp[\mathrm{i}\phi] = \exp[\mathrm{i}(\phi + 2\pi)] = \exp[\mathrm{i}(\phi + 4\pi)] = \ldots$, the phase could be different by any integer times 2π, and the light pulse will still be exactly the same. What this means is that infinitely many different phases vs. time (or frequency) correspond to precisely the same pulse. So how do we decide which phase to use?

There are two preferred methods. The first is to simply force the phase to always remain between 0 and 2π (or $-\pi$ and $+\pi$). This way, there's only one possible phase that yields a given pulse (once the intensity is determined). This is the method you'll be implementing if you simply ask your computer to compute the phase, given the real and imaginary parts of the pulse using Eqs. (2.4), (2.5), (2.11), or (2.12).

The problem with this approach is that, well, it's ugly. When the phase exceeds 2π, it jumps to zero, and a great big discontinuity opens up in the phase. See Fig. 2.4. And this can happen many times over the pulse's life.

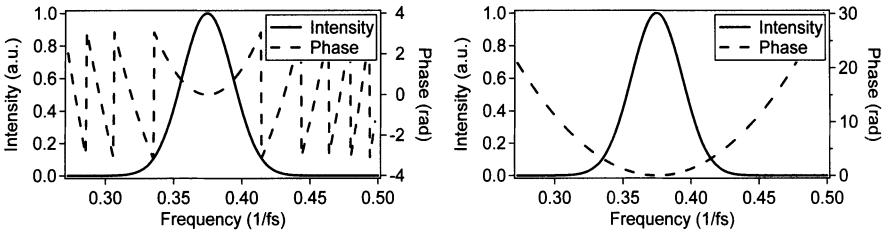

Fig. 2.4: Left: A pulse whose phase has not been phase-unwrapped. Right: The same pulse after phase-unwrapping. Note the different phase scales in each plot.

The solution to this aesthetics problem is to *phase-unwrap*. It involves adding or subtracting the appropriate number of multiples of 2π to the phase at each discontinuity, so that it remains continuous over its entire range. This yields much prettier phases, but the price you pay is the need for a phase-unwrapping routine, which makes these decisions. Fortunately, phase-unwrapping routines work well, and this is the preferred approach in ultrafast optics labs everywhere (including this book).

But be careful, as under-sampling a phase that varies a lot will confuse any phase-unwrapping routine. At a discontinuity, the routine has to decide whether to add 2π to or subtract 2π from the next point. This is easy if the previous two points were 6.276 and 6.280, respectively, and the next point is 0.001: in this case, the routine adds 2π to the 0.001. But if the next point is 2.9 because you didn't sample the points densely enough, it'll just guess. As a result, you could get a really strange-looking phase plot. It'd still be correct, but no one would take you seriously.

Another issue to keep in mind is that, when the intensity goes to zero, the phase is *completely meaningless*. After all, if an arrow has zero length, what possible meaning could there be in its direction? None. Unfortunately, computers are still too dumb to just ignore the phase in this case, and they'll typically simply spew out a blather of random numbers (or worse, error messages) for the phase, even when the intensity is zero.

When this happens, here's something you should *never* do. Do *not* try to fit the resulting random numbers to a polynomial and then call me complaining that your pulse's phase is so complex that even a 500th-order polynomial didn't quite do it (yes, someone did this). Okay, you can do the polynomial fit if you really want to; just don't call me.

The solution to this problem is to *phase-blank*. When the intensity is zero (or so close to zero that it's in the noise), it's customary to simply not plot the phase, instead of plotting random numbers. See Fig. 2.5. The commercial FROG code allows you to decide at what intensity the phase becomes meaningless for your data and hence when to phase-blank. But you can always simply erase these points from your plot.

Finally, there are additional subtleties involving the phase of a pulse. It turns out that a given pulse doesn't necessarily have a unique representation

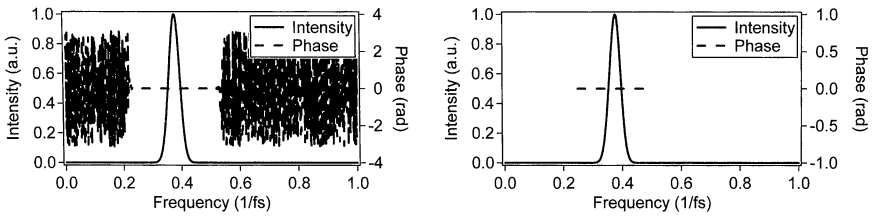

Fig. 2.5: Left: A typical pulse (spectrum and spectral phase) that has not been phase-blanked. The phase takes on random values where the intensity is near zero because the phase is not defined where the intensity is zero. Right: The same pulse after phase-blanking.

in terms of intensity and phase. In other words, different combinations of intensities and phases can yield the same real electric field. Even beyond the above ambiguities, the phase can have additional possible values if we also allow the intensity to vary to compensate. For example, if we artificially modify the intensity slightly by introducing a little bump in it for a very short range of times (think less than one period of the light wave), we can simply adjust the phase at those times to compensate to yield the same real electric field. Don't think too hard about this issue, or you'll have to transfer to a mathematics department.

In fact, to keep us all on the same wavelength, let's all agree to use $S(\omega) = |\mathcal{E}(\omega)|^2$ for the spectrum, $I(t) = |E(t)|^2$ for the intensity, and the corresponding formulas for the phase and spectral phase.

Zeroth-order Phase: The Absolute Phase

First, it's important to realize that the zeroth-order phase is the same in both domains: $\phi_0 = \varphi_0$. This is because the Fourier Transform is linear, and a constant times a function Fourier-Transforms to the same constant times the Fourier Transform of that function. Thus, the zeroth-order phase term, which corresponds to multiplication by a complex constant, is the same in both domains: $E(t)\ \exp(\mathrm{i}\phi_0)$ Fourier-Transforms to $\tilde{E}(\omega)\exp(\mathrm{i}\phi_0)$.

The zeroth-order phase term is often called the *absolute phase*. It's something of a misnomer, as it's really a relative phase: the relative phase of the carrier wave with respect to the envelope. Simply stated, it's the phase of the carrier at the peak of the pulse envelope or some other reference time.

Having said that we desire to measure all orders of the phase, including high ones, we now point out that, in reality, we don't usually care much about the lowest-order term. This is because, when the pulse is many carrier-wave cycles long, variation in the absolute phase shifts the carrier wave from the peak of the envelope to a value only slightly different and hence changes the pulse field very little. Figure 2.6 (top) shows the full real field of a 5-cycle pulse with both a 0 and π values of the absolute phase. Note that it is quite difficult to distinguish the two pulses.

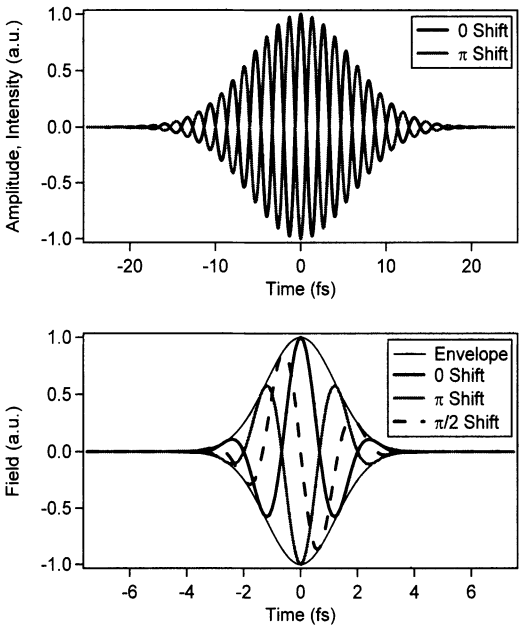

Fig. 2.6: Top: the full real electric field of two 10-fs near-IR pulses, one with zero absolute phase and the other with π absolute phase. Bottom: the full real field of single-cycle near-IR pulses with various absolute phases. Note how different single-cycle pulses look when their absolute phase shifts.

When the pulse is only one cycle long, however, the absolute phase matters. While this effect could be important, we won't consider it in this text.

First-order Phase: A Shift in Time or Frequency

Recall the *Fourier Transform Shift Theorem*, which says that: $E(t - \tau)$ Fourier Transforms to $\tilde{E}(\omega)\ \exp(-i\omega\tau)$. So a linear term in the spectral phase, $\varphi_1 \equiv \tau$, corresponds to a shift in time, i.e., a delay (see Fig. 2.7). Generally, we care only about the pulse's *shape*, not when it arrives. Indeed, if our measurement technique were sensitive a delay of the pulse, we'd have to maintain high stability of its path length, and hence of all beam-steering optics between the source and measurement device. And that would just further complicate our already complicated lives.

Occasionally, the delay is of interest, and interferometric methods can be used in this case (see chapters 22–24). But the first-order term in the spectral phase, φ_1, is generally uninteresting.

Since the Shift Theorem also applies to the inverse Fourier Transform, as well, $\tilde{E}(\omega - \omega_0)$ inverse-Fourier-Transforms to $E(t)\exp(i\omega_0 t)$. So a linear term in the temporal phase, ϕ_1, corresponds to frequency shift (see Fig. 2.7

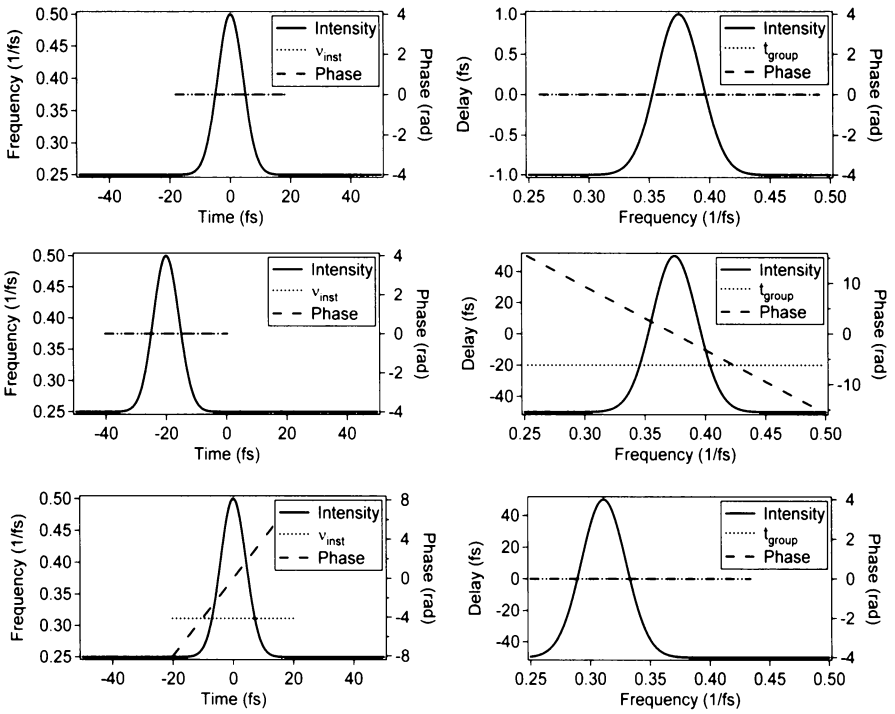

Fig. 2.7: Effect of linear phase. Top row: A Gaussian-intensity, flat-phase pulse. Middle row: the same pulse, but delayed in time, corresponding to a linear spectral phase. Bottom row: the same pulse, but with a linear phase in time, corresponding to a shift of the spectrum. In these plots and all others in this chapter, the frequency scales are measured in *cycles* per fs, not *radians* per fs.

bottom row). A spectral shift is often interesting. It is, however, easily measured with a spectrometer.

Second-order Phase: Linear Chirp

Quadratic variation of $\phi(t)$, that is, a nonzero value of ϕ_2, represents a linear ramp of frequency vs. time and so we say that the pulse is *linearly chirped*. (See Fig. 2.8). Consider a pulse with a Gaussian intensity and quadratic temporal phase:

$$E(t) = [E_0 \exp(-at^2)] \exp(ibt^2) \tag{2.29}$$

where E_0 is a constant, $1/\sqrt{a}$ is roughly the pulse duration, and b is the *chirp parameter*. Here the intensity is:

$$I(t) = |E_0|^2 \exp(-2at^2) \tag{2.30}$$

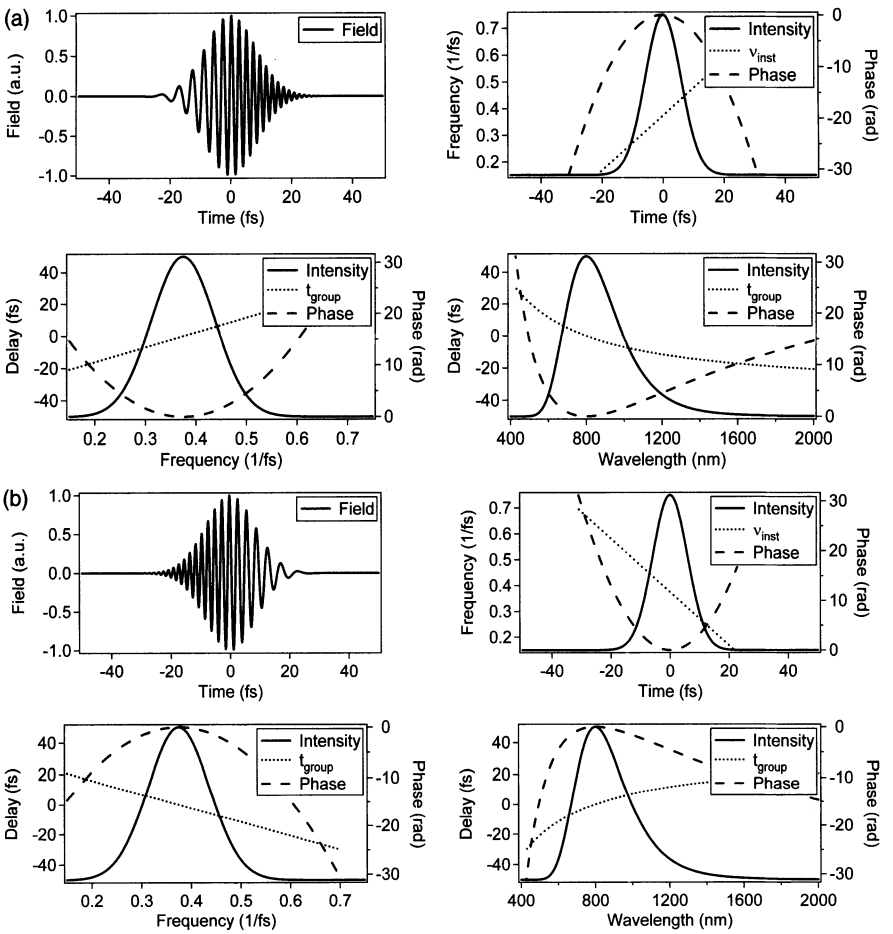

Fig. 2.8: (a) 20-fs Gaussian-intensity pulse w/ quadratic temporal phase, $\phi_2 = -0.032\,\text{rad}\,\text{fs}^2$ or $\varphi_2 = 290\,\text{rad}\,\text{fs}^2$. Here the quadratic phase has stretched what would have been a 3-fs pulse (given the spectrum) to a 13.9-fs one. Top left: the field. Note the increase in frequency with time. Top right: the intensity, phase, and instantaneous frequency vs. time. Bottom row: the spectrum, spectral phase, and group delay vs. frequency and wavelength. Like their time-domain relatives here, the spectrum, spectral phase, and group delay vs. frequency are also Gaussian, quadratic, and linear, respectively, but, plotted vs. wavelength, they are somewhat distorted. (b) Same as Fig. 2.8a, but for a pulse with negative chirp, $\phi_2 = 0.032\,\text{rad}/\text{fs}^2$ or $\varphi_2 = -290\,\text{rad}\,\text{fs}^2$.

and the temporal phase is simply:

$$\phi(t) = -bt^2 \tag{2.31}$$

The Fourier transform of this field is:

$$\tilde{E}(\omega) = \frac{\sqrt{\pi}}{a - \mathrm{i}b} \exp\left[-\frac{\omega^2}{4(a - \mathrm{i}b)}\right] \tag{2.32}$$

Separated into the spectrum and spectral phase, the frequency-domain field can be written:

$$S(\omega) = \frac{\pi}{a^2 + b^2} \exp\left[-\frac{a\omega^2}{2(a^2 + b^2)}\right] \tag{2.33}$$

which is also a Gaussian. And the spectral phase is also quadratic:

$$\varphi(\omega) = \frac{b}{a^2 + b^2}\omega^2 \tag{2.34}$$

As a result, quadratic variation of $\phi(t)$ corresponds to quadratic variation of $\varphi(\omega)$. Note that ϕ_2 and φ_2 have opposite signs. This is a result of the various sign conventions, which are fairly standard.

Propagation through materials usually causes (positive) linear chirp, so if an ultrashort laser pulse doesn't have linear chirp at one point, it will a little further on. In fact, a negatively chirped pulse will shorten as it propagates through material.

Third-order Phase: Quadratic Chirp

Materials have higher-order dispersion, so they also induce higher-order phase distortions, as well. Above second order, distortions in the phase are usually considered in the frequency domain. This is because the spectrum is easily measured, and the intensity vs. time is not, so determination of the spectral phase yields the full pulse field, whereas the temporal phase doesn't. Also, it's quite intuitive to think in terms of how much delay is required for a given frequency to compensate for its distortion in spectral phase.

Third-order spectral phase means a quadratic group delay vs. frequency. This means that the central frequency of the pulse arrives first, say, while frequencies on either side of the central frequency, $\omega_0 \pm \delta\omega$, arrive later. The two slightly different frequencies cause beats in the intensity vs. time, so pulses with cubic spectral phase distortion have oscillations after a main pulse (or before it, if the sign of the third-order coefficient, φ_3, is negative). See Figs. 2.9a and b. Also, you might want to take a peak at Chapter 17, where we'll measure the mother of all cubic-spectral-phase pulses.

Higher-order Phase

Higher-order terms yield additional distortions, which can give rise to extremely complex pulses. Figures 2.10 and 2.11 show pulse shapes with quartic (fourth-order) and quintic (fifth-order) spectral phase.

For example, the nonlinear-optical process, self-phase modulation, yields a temporal phase proportional to the input pulse intensity vs. time. This distortion can be quite complex, especially when considered in the frequency domain (see Figure 2.12).

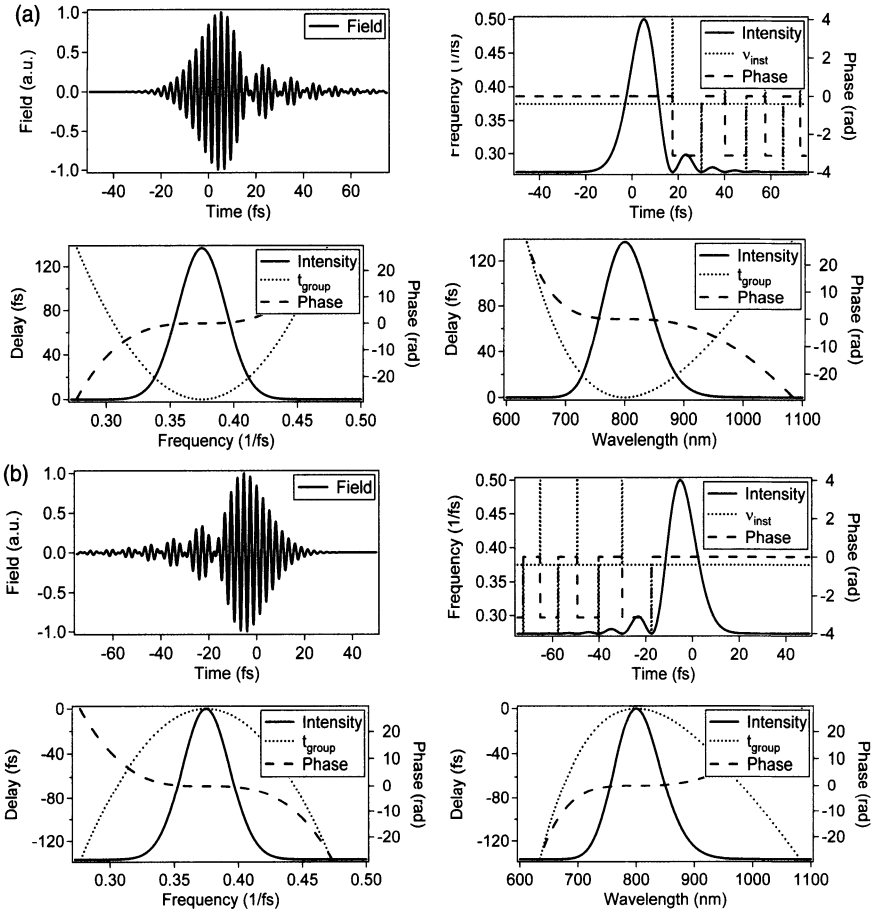

Fig. 2.9: (a) Cubic Spectral phase. Top left: the electric field vs. time for a pulse with a Gaussian spectrum and cubic spectral phase, with $\varphi_3 = 3 \times 10^4$ rad fs^3. Top right: the intensity, phase, and instantaneous frequency vs. time. Note that phase jumps correspond to meaningless discontinuities in the instantaneous frequency. Bottom row: The spectrum, spectral phase, and group delay vs. frequency (left) and wavelength (right). (b) Same as Fig. 2.9a, but with negative cubic spectral phase of the same magnitude as in Fig. 2.9a.

Also, propagation through long distances of fiber can result in higher-order dispersion of the fiber becoming evident in the form of higher-order pulse phase distortions, and nonlinear-optical processes can further distort the pulse phase, as well as the intensity, in both domains.

Finally, to repeat a point we made earlier, it's often tempting to take a phase vs. time or frequency and fit it to a high-order polynomial, as inspired by Eqs. (2.24) or (2.26). While this may be reasonable, it is important to realize that when the intensity is zero, the phase is undefined and hence *meaningless*. And, when the intensity is *near* zero, the phase is *nearly* meaningless, which is probably not too different from *totally* meaningless. Thus, it's important

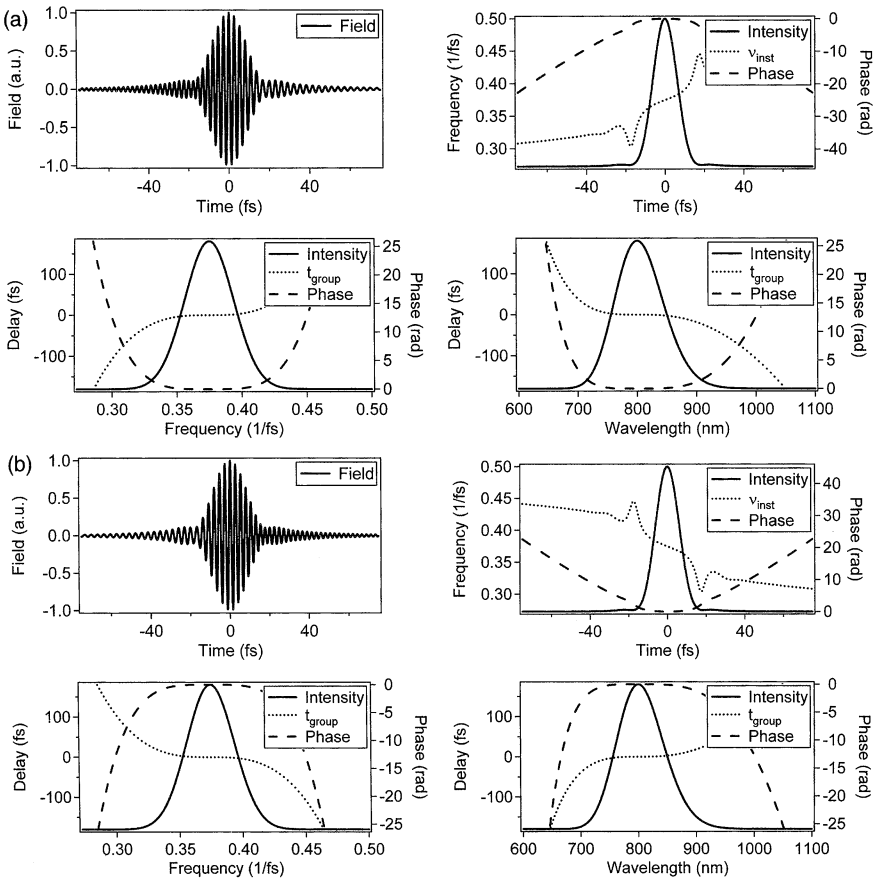

Fig. 2.10: (a) Quartic phase. Top left: the electric field vs. time for a pulse with Gaussian spectrum and positive quartic spectral phase, $\varphi_4 = 4 \times 10^5$ rad fs^4. Top right: The intensity, phase, and instantaneous frequency vs. time. Bottom row: and the spectrum, spectral phase, and group delay vs. frequency (left) and wavelength (right). (b) Same as Fig. 2.10a, but with negative quartic spectral phase of the same magnitude as in Fig. 2.10a.

to crop the phase (to phase-blank) at values of the intensity that are within an error bar of zero, often at about 1% of the peak intensity. Or better, when fitting the phase to a high-order polynomial, use an intensity-weighted fit, which places low emphasis on the phase at times or frequencies where the intensity is weak.

Relative Importance of the Intensity and Phase

Finally, while it's obviously true that both the intensity and phase (in either domain) are required to fully specify a function, in some sense the more

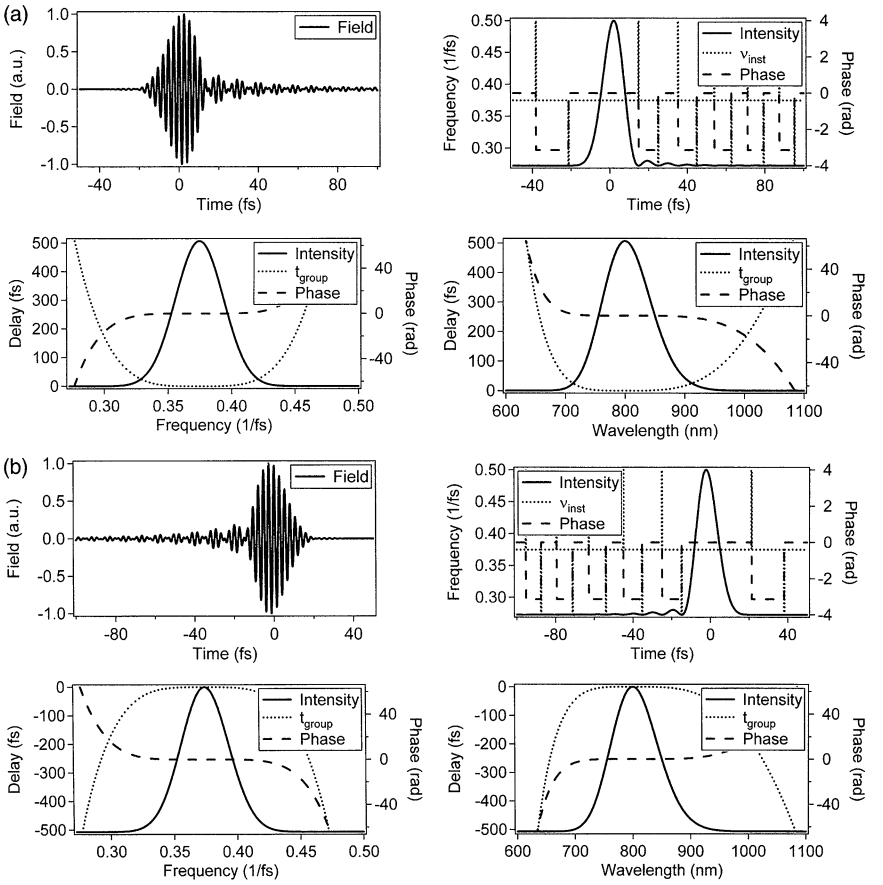

Fig. 2.11: (a) Quintic spectral phase. Top left: the electric field vs. time for a pulse with Gaussian spectrum and $\varphi_5 = 7 \times 10^6$ rad fs^5. Top right: The intensity, phase, and instantaneous frequency vs. time. Bottom: the spectrum, spectral phase, and group delay vs. frequency and wavelength. (b) Same as Fig. 2.11a, but with negative quintic spectral phase of the same magnitude as in Fig. 2.11a.

important of the two quantities is the *phase*. To see this [5], take the *magnitude* of the two-dimensional Fourier Transform of a photograph and combine it with the *phase* from the two-dimensional Fourier Transform of a *different* photograph. This composite image, transformed back to the space domain, tends to look much more like the photograph that supplies the Fourier phase than the photograph that supplies the Fourier magnitude. We've reproduced this example in Fig. 2.13 using different photographs. Note that the composite images look almost nothing like the pictures that supply the Fourier magnitude, and instead both look very much like the picture supplying the Fourier phase!

This fact is also evident in recent work in the generation of near-single-cycle pulses. Spectra of such pulses are often quite structured, but, as long

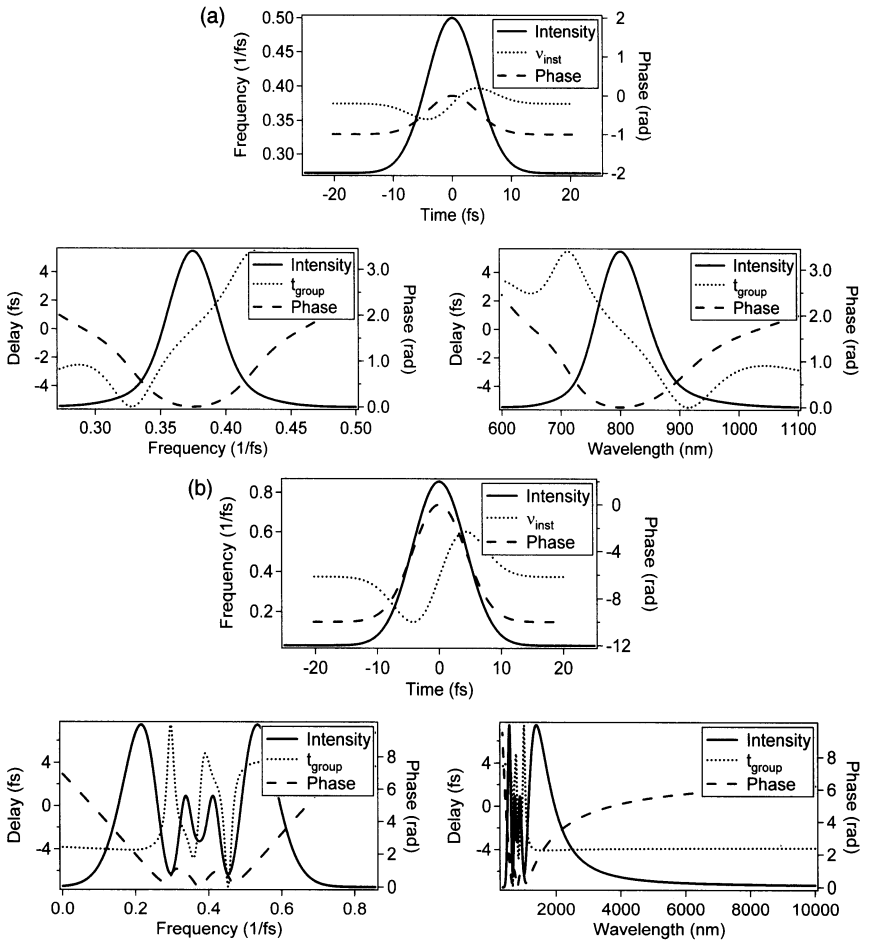

Fig. 2.12: (a) Top: Temporal intensity, phase, and instantaneous frequency of a 10-fs, 800-nm pulse that's experienced self-phase modulation with a peak magnitude of 1 radian. Bottom: spectrum, spectral phase, and group delay vs. frequency and wavelength. All plots use a Gaussian temporal intensity. The pulse is slightly spectrally broadened. (b) Top: Temporal intensity, phase, and instantaneous frequency of a 10-fs, 800-nm pulse that's experienced self-phase modulation with a peak magnitude of 10 radians. Bottom: spectrum, spectral phase, and group delay vs. frequency and wavelength. All plots use a Gaussian temporal intensity. The pulse is massively spectrally broadened.

as a nearly constant spectral phase is achieved, a few-cycle pulse can be produced. The spectral structure causes only small ripples in the wings of the pulse intensity vs. time. See Chapter 14.

Pulse Propagation

We've set up all this terminology to describe potentially very complex ultrashort light pulses. Why have we done this? How do pulses become distorted?

The answer is that light is often created with complex intensity and phase, but, even if it's a simple flat-phase Gaussian pulse to begin with, propagation through materials will distort it.

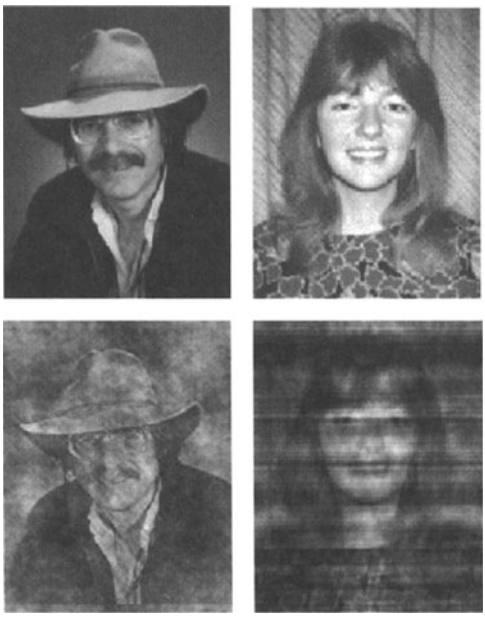

Fig. 2.13: Top: Photographs of the Rick Trebino (left) and his wife, Linda (right). If we 2D-Fourier-transform (FT) each of these pictures, and use the 2D FT magnitude of one photograph in conjunction with the *other photograph's* FT phase, after inverse FT, we make the composite photographs shown on the bottom row. Bottom left: Photograph produced using the FT-magnitude of Linda and FT-phase of Rick. Bottom right: Photograph produced using the FT-magnitude of Rick and FT-phase of Linda. Note that these composite photographs look nothing like the photographs whose FT-magnitude was used, and they look very similar to the photograph whose FT phase was used.

When a pulse propagates through a medium, its various frequencies have different phase and group velocities due to the medium's frequency-dependent refractive index, $n(\omega)$, that is, its *dispersion*. The absorption coefficient, $\alpha(\omega)$, varies also. These effects are easily and accurately modeled. If L is the length of the medium, the frequency-domain output field, $\tilde{E}_{\text{out}}(\omega)$, will be related to the frequency-domain input field, $\tilde{E}_{\text{in}}(\omega)$, by:

$$\tilde{E}_{\text{out}}(\omega) = \tilde{E}_{\text{in}}(\omega) \exp[-\alpha(\omega)L/2] \exp[\text{i}n(\omega)kL] \tag{2.35}$$

$$= \tilde{E}_{\text{in}}(\omega) \exp[-\alpha(\omega)L/2] \exp\left[\text{i}n(\omega)\frac{\omega}{c}L\right] \tag{2.36}$$

Absorption will modify the pulse's spectrum, and dispersion will modify the pulse's spectral phase:

$$S_{\text{out}}(\omega) = S_{\text{in}}(\omega) \exp[-\alpha(\omega)L] \tag{2.37}$$

$$\varphi_{\text{out}}(\omega) = \varphi_{\text{in}}(\omega) + in(\omega)\frac{\omega}{c}L \tag{2.38}$$

Absorption can narrow the spectrum, which could broaden the pulse. On the other hand, occasionally someone attempts to broaden a pulse spectrum by preferentially absorbing its peak frequencies.

We've seen that phase is usually the more interesting quantity. To a reasonably good approximation, propagation through a medium adds first- and second-order terms to the pulse phase. Since, as we have seen, first-order phase vs. ω corresponds to a simple delay, it isn't very interesting. Thus, it's fairly accurate to say that propagation through a material introduces (positive) chirp into a pulse. A flat-phase pulse becomes positively chirped, and a negatively chirped pulse actually shortens. If the pulse is particularly broadband, however, then third, fourth, and possibly fifth-order phase terms must be considered.

Also, if a pulse propagates through some material on its way to your pulse-measurement device, and you really desire to know the pulse's intensity and phase before it propagates through the material, then you can compensate for the distortions introduced by the material using this result. Of course, you can only do this if you're measuring the complete pulse field, $E(t)$ or, equivalently, $\tilde{E}(\omega)$.

The Pulse Length and Spectral Width

Our goal is to measure the pulse complex amplitude $E(t)$ (or $\tilde{E}(\omega)$) completely, that is, to measure both the intensity and phase, expressed in either domain. We must be able to do so even when the pulse has significant intensity structure and highly nonlinear chirp. In addition, we'd like not to have to make assumptions about the pulse.

Unfortunately, this has turned out to be difficult. As a result, researchers have had to make do with considerably less information than they would've liked for many years. A modest request is to be able simply to measure about how long the pulse is. Analogously, we'd like to be able to know how broad the spectrum is. Unfortunately, researchers haven't settled on a single definition of the *pulse length* (also referred to as the *pulse width*) and the *spectral width* (but, for some reason, never referred to as the "spectral length"). Several definitions exist, and each has its advantages and adherents. Here are the most common definitions.

Full-width-half-maximum (τ_{FWHM}): This is the time between the most-separated points that have half of the pulse's peak intensity (see Fig. 2.14). This

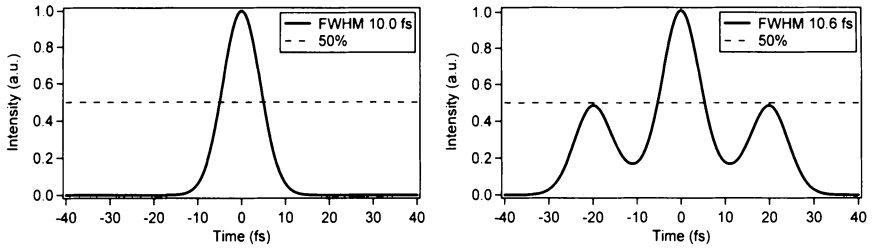

Fig. 2.14: Left: A pulse and its full-width-half-maximum (FWHM). This is a good measure of the pulse width, except when pulse structure exists. Right: A pulse with satellites with 49% of the peak of the pulse, for which this pulse-width definition produces misleading information.

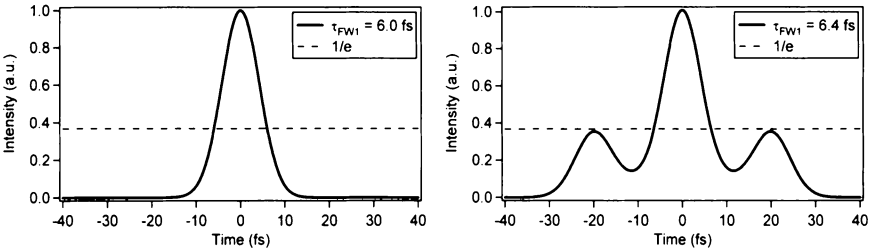

Fig. 2.15: Left: A pulse and its half-width-1/e (HW1/e). Right: This is also a good measure of the pulse width, except when pulse structure exists.

is the most intuitive definition, and it's the rule in experimental measurements, since it's easy to pull τ_{FWHM} off a plot. It's not the most convenient for calculations, however. Also, small variations in the pulse can yield huge changes in τ_{FWHM}. Consider, for example, a pulse with a satellite pulse .49 times as large as the main pulse; if the satellite pulse increases by 1%, the pulse length can increase by a large factor.

For a simple Gaussian-intensity pulse, these issues aren't a problem, and the electric field can be written in terms of τ_{FWHM}:

$$E(t) = E_0 \exp[-2\ln 2(t/\tau_{\mathrm{FWHM}})^2] = E_0 \exp[-1.38(t/\tau_{\mathrm{FWHM}})^2] \quad (2.39)$$

Half-width-1/e ($\tau_{\mathrm{HW1/e}}$): This pulse width (see Fig. 2.15) is the amount of time between the pulse's maximum intensity and the time the intensity drops to 1/e (about 0.36) of the maximum value. Especially useful when the pulse is a Gaussian in time or frequency, this definition allows us to write a simple expression for the pulse, with no messy constants. Theorists like this because it makes it easier to write down expressions in calculations. In terms of this definition, a Gaussian pulse field is written:

$$E(t) = E_0 \exp[-\tfrac{1}{2}(t/\tau_{\mathrm{HW1/e}})^2] \quad (2.40)$$

The factor of 1/2 is required so the intensity will lack such constants:

$$I(t) = |E_0|^2 \exp[-(t/\tau_{\mathrm{HW1/e}})^2] \quad (2.41)$$

Keep in mind that the HW1/e width is considerably less than the FWHM, so be careful to specify which pulse width definition you're using, especially in a conversation between theorists and experimentalists.

Root-mean-squared pulse width (τ_{rms}): This width is the easiest to prove theorems about. It's the second-order moment about the mean arrival time of the pulse:

$$\tau_{\text{rms}}^2 \equiv \left\langle t - \langle t \rangle^2 \right\rangle = \langle t^2 \rangle - \langle t \rangle^2 \tag{2.42}$$

where:

$$\langle t^n \rangle \equiv \int_{-\infty}^{\infty} t^n \, I(t) \, \mathrm{d}t \tag{2.43}$$

and $I(t)$ is assumed normalized so that its time integral is 1 (so it should have dimensions of inverse time). While the FWHM ignores any values of the pulse intensity as long as they're less than one half the pulse maximum intensity, the rms width emphasizes values far from the center of the pulse, and therefore is a good indicator of "wings" in the pulse.

Equivalent pulse width (τ_e): This definition (see Fig. 2.16) considers that the pulse has a width (τ_e) and a height (I_{max}). And the product of these two quantities should be the area under the intensity (the integral of $I(t)$):

$$\tau_e = \frac{1}{I_{\text{max}}} \int_{-\infty}^{\infty} I(t) \, \mathrm{d}t \tag{2.44}$$

This pulse-width definition is most useful when the pulse is complicated, with many sub-pulses and structure.

We define spectral widths, ω_{FWHM}, $\omega_{\text{HW1/e}}$, ω_{rms}, and ω_e, analogously. And spectral widths in cycles per second are $\nu_{\text{FWHM}} = \omega_{\text{FWHM}}/2\pi$, etc.

The Time-Bandwidth Product

Now that we've defined the temporal and spectral widths, we can define the *time-bandwidth product*, or TBP, of a pulse, which is just what it sounds like:

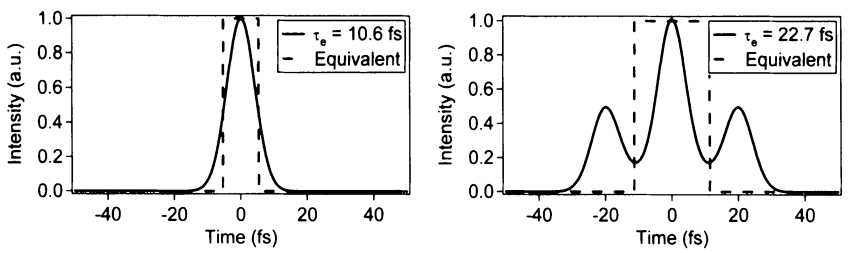

Fig. 2.16: Illustration of the equivalent pulse width for two different pulses. The peak of the dashed rectangular "equivalent" pulse is set equal to the peak of the pulse. The width of the dashed rectangular pulse is then chosen so that its area is equal to that of the solid curve pulse.

the product of the temporal width and the spectral width. If all you can have about a pulse is a single parameter, the parameter you'd like to have is the TBP. Since the units of the pulse width are seconds, and those of the spectral width (ω_{FWHM}) are rad/s, or inverse seconds, the TBP is dimensionless. As a result, it's a good figure of merit for a pulse. The smaller the TBP, the "cleaner" or simpler the pulse. In addition, since the pulse coherence time, τ_c (roughly the length of the shortest structure within a pulse), is the reciprocal of the bandwidth, the TBP is the ratio of the pulse width and the coherence time. So the TBP is the approximate number of sub-pulses in the pulse. For pulses whose main distortion is a low-order phase distortion, however, such as linear chirp, the TBP can be large even when there is no substructure in the pulse. Whatever the source of distortions, laser builders and manufacturers and researchers try very hard to make the simplest pulses with the lowest TBP.

Depending on the definition chosen, the minimum possible TBP ranges from about .1 to 1, and it increases with increasing pulse complexity (see Figs. 2.17 and 2.18).

It would seem reasonable that a pulse with a flat phase would have a smaller TBP than a pulse with a complicated phase. Is this always the case? Or is it possible to have a pulse with, say, a complicated spectrum, for which some complicated spectral phase yields a smaller pulse length and hence a smaller TBP than does a constant phase? It turns out that, for *any* spectrum, the shortest pulse in time, and hence the smallest TBP, *always* occurs for a flat spectral phase. Similarly, for any pulse intensity vs. time, the narrowest spectrum, and hence the smallest TBP, always occurs for a flat temporal phase. These conclusions require that we use the rms temporal and spectral widths and follow easily from the result given by Cohen in his excellent book, *Time-Frequency Analysis* [6,7]:

$$\omega_{\text{rms}}^2 = \int_{-\infty}^{\infty} A'(t)^2 \, \mathrm{d}t + \int_{-\infty}^{\infty} A(t)^2 \, \phi'(t)^2 \, \mathrm{d}t \tag{2.45}$$

where the real amplitude $A(t) = \sqrt{I(t)}$, intensity is assumed normalized to have unity time integral, the prime means the derivative, and the mean frequency is assumed subtracted from $\phi'(t)$.

This result writes the rms bandwidth as something like the Pythagorean sum of a contribution due to variations in the amplitude and a contribution due to variations in the phase (weighted by the intensity). Note that both integrands and integrals are always positive, so variations in the amplitude only increase the bandwidth and, likewise, variations in the phase also only increase the bandwidth.

Since the Fourier Transform is symmetrical, the same holds for the rms pulse width in terms of the spectral variations:

$$\tau_{\text{rms}}^2 = \int_{-\infty}^{\infty} B'(\omega)^2 \, \mathrm{d}\omega + \int_{-\infty}^{\infty} B(\omega)^2 \, \varphi'(\omega)^2 \, \mathrm{d}\omega \tag{2.46}$$

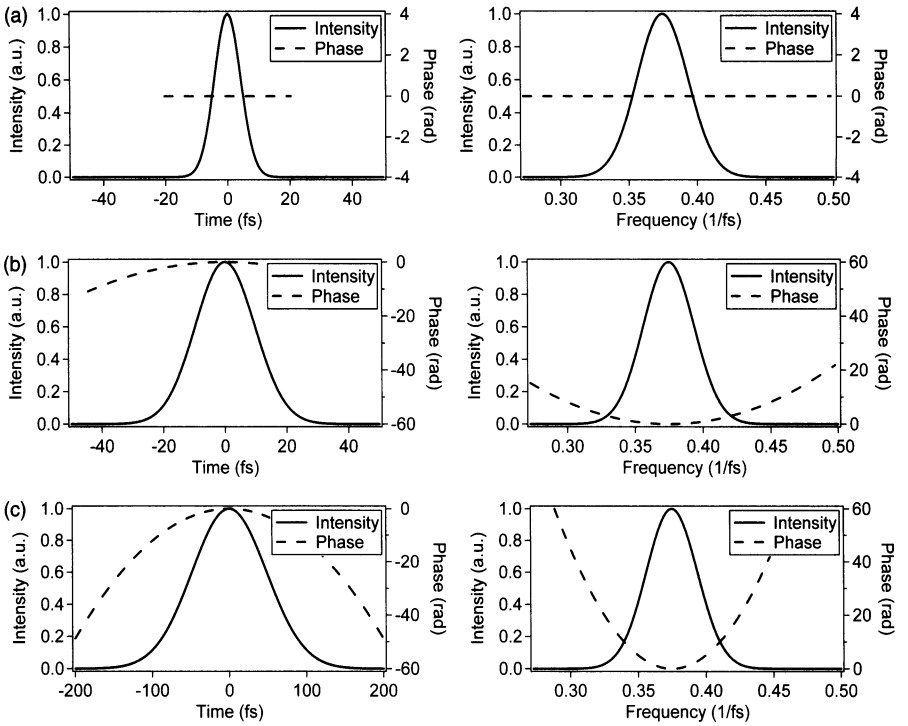

Fig. 2.17: (a) Gaussian-intensity pulse with constant phase and minimal TBP. The intensity and phase vs. time (left); the spectrum and spectral phase vs. frequency (right). For the different definitions of the widths: $TBP_{rms} = \tau_{rms}\,\omega_{rms} = 0.5$, $TBP_e = 3.14$, $TBP_{HW1/e} = 1$, $TBP_{FWHM} = 2.76$. Divide by 2π for $\tau_{rms}\nu_{rms}$, etc. (b) Same as Fig. 2.17a, except a longer pulse (note the change in scale of the phase axis) with chirp and hence a larger TBP. $TBP_{rms} = 1.13$, $TBP_e = 7.01$, $TBP_{HW1/e} = 2.26$, $TBP_{FWHM} = 6.28$. Divide by 2π for $\tau_{rms}\nu_{rms}$, etc. (c) Same as Fig. 2.17a, except an even longer pulse (note the change in scale of the time axis) with more chirp and hence a larger TBP. $TBP_{rms} = 5.65$, $TBP_e = 35.5$, $TBP_{HW1/e} = 11.3$, $TBP_{FWHM} = 31.3$. Divide by 2π for $\tau_{rms}\nu_{rms}$, etc.

where the spectral amplitude is $B(\omega) = \sqrt{S(\omega)}$, $S(\omega)$ is assumed normalized to have unity area, prime means derivative, and the mean pulse time is assumed subtracted from $\varphi'(\omega)$.

Thus, for a given spectrum, $S(\omega)$, variations in the spectral phase can only increase the rms pulse width over that corresponding to a flat spectral phase.

Spatio-Temporal Pulse Characteristics

In writing Eq. (2.1), we've ignored the spatial dependence of the beam. More specifically, we've tacitly assumed that the complex pulse field, which is actually a function of both time and space, separates into the product of spatial and temporal factors, and we have simply ignored the spatial component. This assumption is valid for the fairly smooth pulses emitted by most

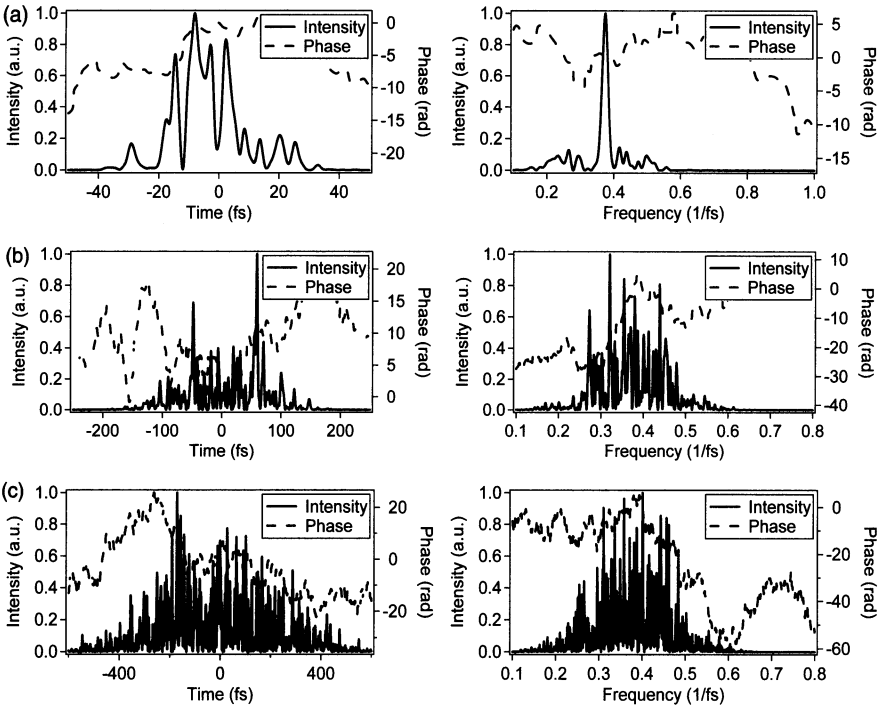

Fig. 2.18: (a) A pulse with random intensity and phase structure. The intensity and phase vs. time (left) the spectrum and spectral phase vs. frequency (right). This pulse has a near-unity TBP. For the various definitions of the pulse and spectral widths, the TBP is: $TBP_{rms} = 6.09$, $TBP_e = 4.02$, $TBP_{HW1/e} = 0.82$, $TBP_{FWHM} = 2.57$. Divide by 2π for $\tau_{rms}\nu_{rms}$, etc. (b) Same as Fig. 2.18a, except a pulse with more structure and hence a larger TBP. $TBP_{rms} = 32.9$, $TBP_e = 10.7$, $TBP_{HW1/e} = 35.2$, $TBP_{FWHM} = 116$. Divide by 2π for $\tau_{rms}\nu_{rms}$, etc. (c) Same as Fig. 2.18a, except a pulse with even more structure and hence an even larger TBP. $TBP_{rms} = 122$, $TBP_e = 44.8$, $TBP_{HW1/e} = 213$, $TBP_{FWHM} = 567$. Divide by 2π for $\tau_{rms}\nu_{rms}$, etc.

ultrafast lasers. It is, however, fairly easy to generate pulses that violate this assumption (for example, pulse compressors and shapers can introduce angular dispersion into the pulse, so the pulse winds up with its redder colors on one side and the bluer colors on the other, a distortion called *spatial chirp*), and nearly all pulse-measurement techniques get confused in this case. We'll talk about how to measure such complicated pulses later when we discuss the spatio-temporal measurement of a pulse (Chapter 22), but in the meantime, we'll ignore this problem. (If you suspect your pulse has this problem before you get to Chapter 22, just aperture it, and measure a small piece of the beam.)

We've also assumed polarized light, but this also is not necessary. We'll get to the measurement of a polarization-varying pulse later (we'll just measure each polarization independently, but we'll have to measure the relative phase of the two polarizations, as well—see Chapter 23).

References

1. Ippen, E.P. and C.V. Shank, *Ultrashort Light Pulses—Picosecond Techniques and Applications*, ed. S.L. Shapiro. 1977, Berlin: Springer-Verlag.
2. Duling, I.N., ed. *Compact Sources of Ultrashort Pulses*. 1995, Cambridge Universty Press: Cambridge.
3. Rulliere, C., ed. *Femtosecond Laser Pulses: Principles and Experiments*. 1998, Springer: Heidelberg.
4. Kaiser, W., ed. *Ultrashort Laser Pulses and Applications*. Topics in Applied Physics. Vol. 60. 1988, Springer-Verlag: Berlin.
5. Stark, H., ed. *Image Recovery: Theory and Application*. 1987, Academic Press: Orlando.
6. Cohen, L., *Time-Frequency Distributions—A Review*. Proceedings of the IEEE, 1989. **77**(7): p. 941–81.
7. Cohen, L., *Time-Frequency Analysis*. 1995, Englewood Cliffs, NJ: Prentice-Hall.

3. Nonlinear Optics

Rick Trebino and John Buck

Linear vs. Nonlinear Optics

The great thing about ultrashort laser pulses is that all their energy is crammed into a very short time, so they have very high power and intensity. A typical ultrashort pulse from a Ti : Sapphire laser oscillator has a paltry nanojoule of energy, but it's crammed into 100 fs, so its peak power is 10,000 Watts. And it can be focused to a micron or so, yielding an intensity of 10^{12} W/cm^2! And it's easy to amplify such pulses by a factor of 10^6!

What this means is that ultrashort laser pulses easily experience *high-intensity effects*—effects that we don't ordinarily see because even sunlight on the brightest day doesn't approach the above intensities. And all high-intensity effects fall under the heading of *nonlinear optics* [1–12]. Some of these effects are undesirable, such as optical damage. Others are very desirable, such as *second-harmonic generation*, which allows us to make light at a new frequency, twice that of the input light. Or like *four-wave mixing*, which allows us to generate light with an electric field proportional to $E_1(t)\ E_2^*(t)\ E_3(t)$, where $E_1(t)$, $E_2(t)$, and $E_3(t)$ are the complex electric-field amplitudes of three different light waves. Whereas linear optics requires that light beams pass through each other without affecting each other, nonlinear optics allows the opposite. This chapter will describe the basics of nonlinear optics for anyone who hasn't experienced this field, so you can understand the basics of FROG, which is an inherently nonlinear-optical phenomenon.

The fundamental equation of optics—whether linear or nonlinear—is the wave equation:

$$\frac{\partial^2 \mathcal{E}}{\partial z^2} - \frac{1}{c_0^2}\frac{\partial^2 \mathcal{E}}{\partial t^2} = \mu_0 \frac{\partial^2 \mathcal{P}}{\partial t^2} \tag{3.1}$$

where μ_0 is the magnetic permeability of free space, c_0 is the speed of light in vacuum, $\mathcal{E}$ is the real electric field, and $\mathcal{P}$ is the real induced polarization. The induced polarization contains the light's effects on the medium and the medium's effect back on the light wave. It drives the wave equation.

The induced polarization contains linear-optical effects (the absorption coefficient and refractive index) and also nonlinear-optical effects. At low intensity (or low field strength), the induced polarization is proportional to the electric field that is already present:

$$\mathcal{P} = \varepsilon_0\, \chi^{(1)}\, \mathcal{E} \tag{3.2}$$

where ε_0 is the electric permittivity of free space, and the linear susceptibility, $\chi^{(1)}$, describes the linear-optical effects. This expression follows from the

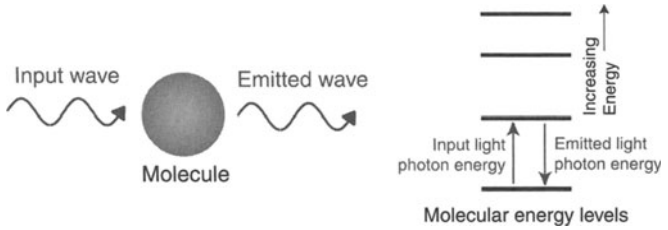

Fig. 3.1: Linear optics. Left: A molecule excited by a light wave oscillates at that frequency and emits only that frequency. Right: This process can be diagrammed by showing the input light wave as exciting ground-state molecules up to an excited level, which re-emits the same frequency.

fact that the light electric field, $\mathcal{E}$, forces electric dipoles in the medium into oscillation at the frequency of the field; the dipole oscillators then emit an additional electric field at the same frequency. The total electric field (incident plus emitted) is what appears as $\mathcal{E}$ in Eqs. (3.1) and (3.2). If we assume a lossless medium, for example, we find that the electric and polarization field expressions, $\mathcal{E}(z,t) \propto E_0 \cos(\omega t - kz)$ and $\mathcal{P} = \varepsilon_0 \chi^{(1)} E_0 \cos(\omega t - kz)$, will solve the wave equation, provided that $\omega = c\,k$, and $c = c_0/(1+\chi^{(1)})^{1/2}$.

In linear optics, (where Eq. (3.2) applies), the wave equation is linear, so if $\mathcal{E}$ is a sum of more than one beam (field), then so is $\mathcal{P}$. As a result, $\mathcal{P}$ drives the wave equation to produce light with *only* those frequencies present in $\mathcal{P}$, and these arise from the original input beams. In other words, light doesn't change color (see Fig. 3.1). Also, with a linear wave equation, the principle of superposition holds, and beams of light can pass through each other and don't affect each other.

Life at low intensity is dull.

Nonlinear-Optical Effects

At high intensity, the induced polarization ceases to be a simple linear function of the electric field. Put simply, like a cheap stereophonic amplifier driven at too much volume, the medium doesn't follow the field perfectly (see Figs. 3.2 through 3.4), and higher-order terms must be included:

$$\mathcal{P} = \varepsilon_0 \left[\chi^{(1)}\mathcal{E} + \chi^{(2)}\,\mathcal{E}^2 + \chi^{(3)}\,\mathcal{E}^3 + \cdots\right] \tag{3.3}$$

where $\chi^{(2)}$ and $\chi^{(3)}$ are called the second- and third-order susceptibilities. $\chi^{(n)}$ is called the nth-order susceptibility.

What do nonlinear-optical effects look like? They're easy to calculate. Recall that the real field, $\mathcal{E}$, is given by:

$$\mathcal{E}(t) = \tfrac{1}{2}E(t)\exp(\mathrm{i}\omega t) + \tfrac{1}{2}E^*(t)\exp(-\mathrm{i}\omega t) \tag{3.4}$$

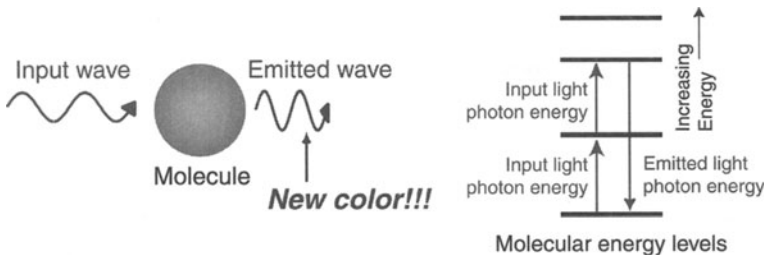

Fig. 3.2: Nonlinear optics. Left: A molecule excited by a light wave oscillates at other frequencies and emits those new frequencies. Right: This process can be diagrammed by showing the input light wave as exciting ground-state molecules up to highly excited levels, which re-emit the new frequencies.

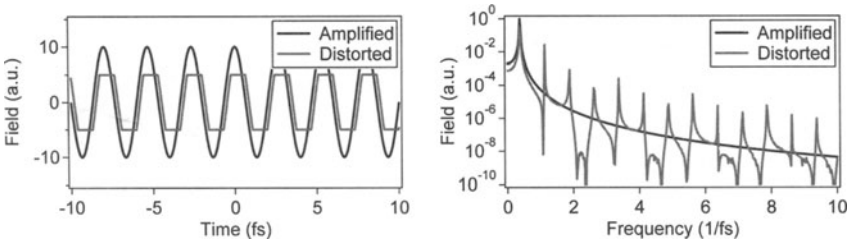

Fig. 3.3: Nonlinear electronic effects in a cheap audio amplifier. The input wave from the audio source is taken here to be a sine wave. In an expensive amplifier, the sine wave is accurately reproduced at higher volume, but, because the cheap amplifier cannot achieve the desired volume, the output wave saturates and begins to look more like a square wave. This produces new frequency components at harmonics of the input wave. Nonlinear-optical effects are analogous: a sine-wave electric wave drives a molecular system, which also does not reproduce the input sine wave accurately, producing new frequencies at harmonics of the input wave. Whereas audiophiles spend a great deal of money to avoid the above nonlinear electronic effects, optical scientists spend a great deal of money to achieve nonlinear-optical effects.

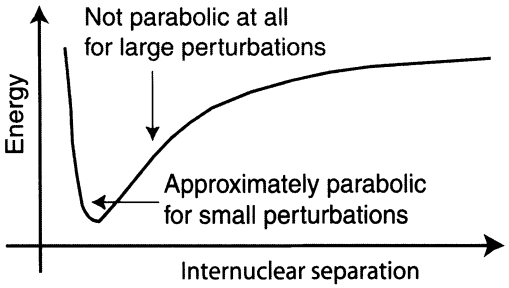

Fig. 3.4: Potential surface of a molecule, showing the energy vs. separation between nuclei. Note that the potential is nearly parabolic near the bottom, but it is far from parabolic for excitations that hit the molecule harder forcing it to vibrate with larger ranges of nuclear separations. This molecule will emit frequencies other than that driving it.

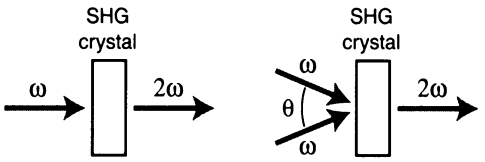

Fig. 3.5: Second-harmonic generation. Left: Collinear beam geometry. Right: Noncollinear beam geometry with an angle, θ, between the two input beams. Such noncollinear beam geometries are possible in nonlinear optics because more than one field is required at the input.

where we have temporarily suppressed the space dependence, and $E(t)$ is the complex field. So squaring this field yields:

$$\mathcal{E}^2(t) = \tfrac{1}{4}E^2(t)\exp(2i\omega t) + \tfrac{1}{2}E(t)E^*(t) + \tfrac{1}{4}E^{*2}(t)\exp(-2i\omega t) \quad (3.5)$$

Notice that this expression includes terms that oscillate at 2ω, the *second harmonic* of the input light frequency. These terms then drive the wave equation to yield light at this new frequency. This process is very important; it's called *second-harmonic generation* (SHG). Optical scientists, especially ultrafast scientists, make great use of SHG to create new frequencies. And it is the single most important effect used to measure ultrashort laser pulses. Figure 3.5 shows a schematic of SHG.

The above expression also contains a zero-frequency term, so light can induce a dc electric field. This effect is called optical rectification; it's generally pretty weak, so we won't say much more about it.

If we consider the presence of two beams and this time don't suppress the spatial dependence, $\mathcal{E}(\vec{r},t) = \frac{1}{2}E_1(\vec{r},t)\exp[i(\omega_1 t - \vec{k}_1\cdot\vec{r})] + \frac{1}{2}E_2(\vec{r},t)\exp[i(\omega_2 t - k_2\cdot\vec{r})] + c.c.$ In this case, we have:

$$\begin{aligned}
\mathcal{E}^2(\vec{r},t) = {} & \tfrac{1}{4}{E_1}^2\exp\left[2i(\omega_1 t - \vec{k}_1\cdot\vec{r})\right] \\
& + \tfrac{1}{2}E_1{E_1}^* + \tfrac{1}{4}{E_1}^{*^2}\exp\left[-2i\left(\omega_1 t - \vec{k}_1\cdot\vec{r}\right)\right] \\
& + \tfrac{1}{4}{E_2}^2\exp\left[2i\left(\omega_2 t - \vec{k}_2\cdot\vec{r}\right)\right] + \tfrac{1}{2}E_2{E_2}^* \\
& + \tfrac{1}{4}E_2^{*^2}\exp\left[-2i\left(\omega_2 t - \vec{k}_2\cdot\vec{r}\right)\right] \\
& + \tfrac{1}{2}E_1E_2\exp\left\{i\left[(\omega_1+\omega_2)t - \left(\vec{k}_1+\vec{k}_2\right)\cdot\vec{r}\right]\right\} \\
& + \tfrac{1}{2}{E_1}^*{E_2}^*\exp\left\{-i\left[(\omega_1+\omega_2)t - \left(\vec{k}_1+\vec{k}_2\right)\cdot\vec{r}\right]\right\} \\
& + \tfrac{1}{2}E_1{E_2}^*\exp\left\{i\left[(\omega_1-\omega_2)t - \left(\vec{k}_1-\vec{k}_2\right)\cdot\vec{r}\right]\right\} \\
& + \tfrac{1}{2}{E_1}^*E_2\exp\left\{-i\left[(\omega_1-\omega_2)t - \left(\vec{k}_1+\vec{k}_2\right)\cdot\vec{r}\right]\right\} \quad (3.6)
\end{aligned}$$

Okay, this looks like a mess. But the first two lines are already familiar; they're the SHG and optical-rectification terms for the individual fields. The next line

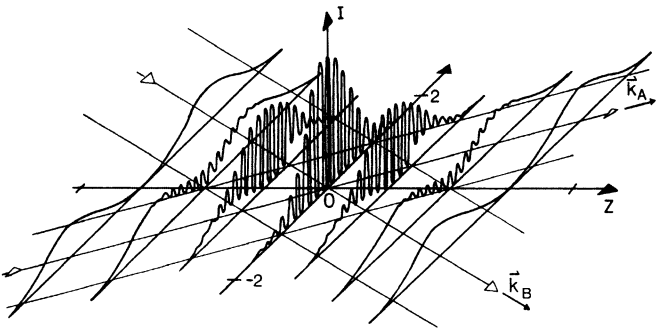

Fig. 3.6: Intensity pattern produced when two beams cross. When the beams cross in a medium, the medium is changed more at the intensity peaks than at the troughs, producing a laser-induced grating [13].

is new: it yields light at the frequency, $\omega_1 + \omega_2$, the sum frequency, and hence is called *sum-frequency generation* (SFG). The last line is also new: it yields light at the frequency, $\omega_1 - \omega_2$, the difference frequency, and hence is called *difference-frequency generation* (DFG). These two processes are also quite important, and they play a key role in techniques to measure pulses, as well.

Notice something else. The new beams are created in new directions, $\vec{k}_1 + \vec{k}_2$ and $\vec{k}_1 - \vec{k}_2$. This can be very convenient if we desire to see these new—potentially weak—beams in the presence of intense input beams that create them.

Third-order effects are collectively referred to as *four-wave-mixing (4WM)* effects because three waves enter the nonlinear medium, and an additional one is created in the process, for a total of four. We won't waste a page and write out the entire third-order induced polarization, but, in third order, as you can probably guess, we see effects including *third-harmonic generation* (THG) and a variety of terms like:

$$\mathcal{P}_i = \tfrac{3}{4}\varepsilon_0 \chi^{(3)} E_1 E_2^* E_3 \exp\left\{i\left[(\omega_1 - \omega_2 + \omega_3)t - \left(\vec{k}_1 - \vec{k}_2 + \vec{k}_3\right)\cdot\vec{r}\right]\right\} \quad (3.7)$$

Notice that, if the factor of the electric field envelope is complex-conjugated, its corresponding frequency and k-vector are both negative, while, if the field is not complex-conjugated, the corresponding frequency and k-vector are both positive. Such third-order effects, in which one k-vector is subtracted, are often called *induced grating* effects because the intensity due to two of the beams, say, E_1 and E_2, has a sinusoidal spatial dependence (see Fig. 3.6). The sinusoidal intensity pattern affects the medium in some way, creating a sinusoidal modulation of its properties, analogous to those of a diffraction grating. The process can then be modeled as diffraction of the third beam off the induced grating.

Third-order effects include a broad range of interesting phenomena (some useful, some irritating), many beyond the scope of this book. But we'll consider a few that are important for pulse measurement. For example, suppose

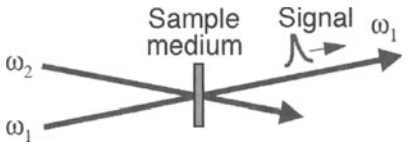

Fig. 3.7: Two-beam coupling. One beam can affect the other in passing through a sample medium. The pulse at the output indicates the signal beam, here collinear with one of the beams and at the same frequency. This idea is the source of a variety of techniques for measuring the properties of the sample medium.

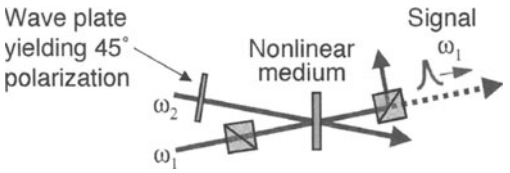

Fig. 3.8: Polarization gating. If the polarizers are oriented at 0° and 90°, respectively, the 45°-polarized beam (at frequency ω_2) induces polarization rotation of the 0°-polarized beam (at frequency ω_1), which can then leak through the second 90° polarizer. The pulse at the output indicates the signal pulse, again collinear with one of the input beams, but here with the orthogonal polarization.

that the second and third beams in the above expression are the same: $E_2 = E_3$ and $\vec{k}_2 = \vec{k}_3$. In this case, the above induced polarization becomes:

$$\mathcal{P}_1 = \tfrac{3}{4}\varepsilon_0 \chi^{(3)} E_1 |E_2|^2 \exp\left\{i\left[\omega_1 t - \vec{k}_1 \cdot \vec{r}\right]\right\} + c.c. \qquad (3.8)$$

This yields a beam that has the same frequency and direction as beam #1, but allows it to be affected by beam #2 through its mag-squared factor. So beams that pass through each other can affect each other! Of course, the strength of all such effects is zero in empty space ($\chi^{(3)}$ of empty space is zero), but the strength can be quite high in a solid, liquid, or gas. It's often called *two-beam coupling* (see Fig. 3.7).

A particularly useful implementation of the above third-order effect is *polarization gating* (see Fig. 3.8), which involves the use of orthogonal polarizations for E_2 and E_3. This typically means that these two co-propagating beams combine together to yield a beam polarized at 45° to that of E_1, which is, say, horizontally polarized. The two vertically polarized beams form a grating, and the horizontally polarized beam diffracts off it, and the diffracted beam maintains horizontal polarization. This creates an induced polarization for the horizontal polarization, i.e., the polarization orthogonal to that of E_1. This new beam is created in the same direction as beam #1, and with the same frequency, too. As a result, crossed polarizers can be used to separate the new beam from the input beam E_1. This beam geometry is convenient and easy to set up, and it's much more sensitive than two-beam coupling.

By the way, another process is simultaneously occurring in polarization gating called *induced birefringence*, in which the electrons in the medium oscillate along with the incident field at +45°, which stretches the formerly spherical electron cloud into an ellipsoid elongated along the +45° direction. This introduces anisotropy into the medium, typically increasing the refractive index for the +45° direction and decreasing it for the −45° direction. The medium then acts like a wave plate, slightly rotating the polarization of the field, E_1, allowing some it to leak through the crossed polarizers.

However you look at it, you get the same answer when the medium responds rapidly.

Another type of induced-grating process is *self diffraction* (see Fig. 3.9). It involves beams #1 and #2 inducing a grating, but beam #1 also diffracting off it. Thus beams #1 and #3 are the same beam. This process has the induced-polarization term:

$$\mathcal{P}_i = \tfrac{3}{8}\,\varepsilon_0\,\chi^{(3)}\,E_1^2\,E_2^*\exp\left\{i\left[(2\omega_1-\omega_2)t-\left(2\vec{k}_1-\vec{k}_2\right)\cdot\vec{r}\right]\right\}+c.c. \quad (3.9)$$

It produces a beam with frequency $2\omega_1-\omega_2$ and k-vector $2\vec{k}_1-\vec{k}_2$. This beam geometry is also convenient because only two input beams are required.

And it is also possible to perform third-harmonic generation using more than one beam (or as many as three). An example beam geometry is shown in Fig. 3.10, using two input beams.

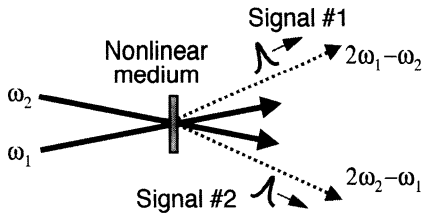

Fig. 3.9: Self diffraction. The two beams yield a sinusoidal intensity pattern, which induces a grating in the medium. Then each beam diffracts off the grating. The pulses at the output indicate the signal pulses, here in the $2\vec{k}_1-\vec{k}_2$ and $2\vec{k}_2-\vec{k}_1$ directions.

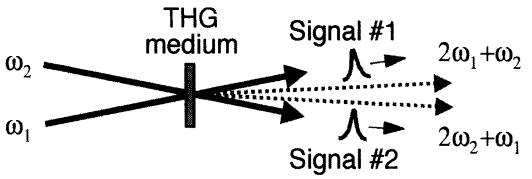

Fig. 3.10: Third-harmonic generation. While each beam individually can produce third harmonic, it can also be produced by two factors of one field and one of the other. These latter two effects are diagrammed here.

Some General Observations about Nonlinear Optics

Nonlinear-optical effects are usually diagrammed as in Fig. 3.11. Upward-pointing arrows indicate fields without complex conjugates and with frequency and k-vector contributions with plus signs. Downward-pointing arrows indicate complex-conjugated fields in the polarization and negative signs in the contributions to the frequency and k-vector of the light created. Unless otherwise specified, ω_0 and k_0 denote the output or *signal* frequency and k-vector.

Notice that, in all of these nonlinear-optical processes, the polarization propagates through the medium just like the light wave does. It has a frequency and k-vector. For a given process of N^{th} order, the signal frequency ω_0 is given by:

$$\boxed{\omega_0 = \pm\omega_1 \pm \omega_2 \pm \cdots \pm \omega_N} \tag{3.10}$$

where the signs obey the above complex-conjugate convention.

The polarization has a k-vector with an analogous expression:

$$\boxed{\vec{k}_0 = \pm\vec{k}_1 \pm \vec{k}_2 \pm \cdots \pm \vec{k}_N} \tag{3.11}$$

where the same signs occur in both Eqs. (3.10) and (3.11).

In all of these nonlinear-optical processes, terms with products of the E-field complex envelopes, such as $E_1^2 E_2^*$, are created. *It is these products that allow us to measure ultrashort laser pulses.* Whether it is simple autocorrelation, FROG, or some new, as yet undiscovered method, it will take advantage of these effects. What we'll be doing, for example, is taking two beams (pulses) and delaying one with respect to the other and considering processes with the

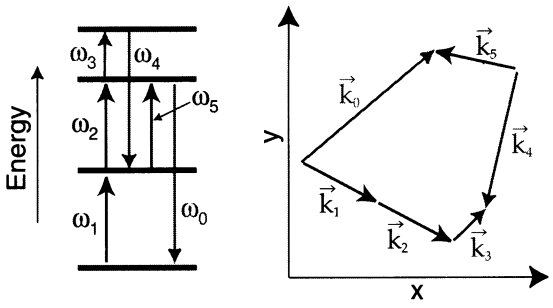

Fig. 3.11: Sample complex nonlinear-optical process, $\mathcal{P} \propto E_1 E_2 E_3 E_4^* E_5$. Here, $\omega_0 = \omega_1 + \omega_2 + \omega_3 - \omega_4 + \omega_5$ and $\vec{k}_0 = \vec{k}_1 + \vec{k}_2 + \vec{k}_3 - \vec{k}_4 + \vec{k}_5$. The k-vectors are shown adding in two-dimensional space, but, in third- and higher-order processes, space's third dimension is potentially also involved. The different frequencies (colors) of the beams are shown as different shades of gray.

product, $E_1(t)\ E_2(t-\tau)$, where τ is the delay. This multiplication of electric fields will allow one pulse to gate out a temporal piece of another.

The Mathematics of Nonlinear Optics

The Slowly Varying Envelope Approximation

Okay, so there are some interesting induced polarizations going on, but how do we calculate what their effects are? Well, we must substitute into the wave equation, Eq. (3.1), and solve the nonlinear differential equation that results. While this is hard to do exactly, a few tricks and approximations make it quite easy in most cases of practical interest.

The first approximation is that we consider only a range of frequencies near one frequency at a time. We'll write the wave equation for one particular signal frequency, ω_0, and only consider a small range of nearby frequencies. Anything happening at distant frequencies will alternately be in phase and then out of phase with the fields and polarizations in this range and so should have little effect. We'll also assume that the nonlinear optical process is fairly weak, so it won't affect the input beams. Thus we'll only consider the one signal field of interest. If you're interested in more complex situations, you're probably not measuring pulses, and you should check out a full text on nonlinear optics (see, for example, the list at the end of this chapter).

The second is the *Slowly Varying Envelope Approximation (SVEA)*, which, despite its name, remains a remarkably good approximation for all but the shortest pulses (we'll see it break down in the chapter on few-femtosecond pulses, but the fix will be remarkably simple). It takes advantage of the fact that, as short as they are, most ultrashort laser pulses are still not as short as an optical cycle (about 2 fs for visible wavelengths). Thus the pulse electric field can be written as the product of the carrier sine wave and a relatively slowly varying envelope function. This is what we've been doing, but we haven't explicitly used this fact; now we will. Since the measure of the change of anything is the derivative, we'll now neglect second derivatives of the slowly varying envelope compared to those of the more rapidly varying carrier sine wave. And the wave equation, which is what we must solve to understand any optics problem, is drowning in derivatives.

Assume that the driving polarization propagates along the z-axis, and write the electric field and polarization in terms of slowly varying envelopes:

$$\mathcal{E}(\vec{r},t) = \tfrac{1}{2}\,E(\vec{r},t)\exp[\mathrm{i}(\omega_0 t - k_0 z)] + c.c. \tag{3.12}$$

$$\mathcal{P}(\vec{r},t) = \tfrac{1}{2}\,P(\vec{r},t)\exp[\mathrm{i}(\omega_0 t - k_0 z)] + c.c. \tag{3.13}$$

where we've chosen to consider the creation of light at the same frequency as that of the induced polarization, ω_0. But we've also assumed that the light field and polarization have the same k-vectors, k_0, which is a big—and often

unjustified—assumption, as discussed above. But bear with us for now, and we'll come clean in a little while.

Recall that the wave equation calls for taking second derivatives of $\mathcal{E}$ and $\mathcal{P}$ with respect to t and/or z. Let's calculate them:

$$\frac{\partial^2 \mathcal{E}}{\partial t^2} = \frac{1}{2}\left[\frac{\partial^2 E}{\partial t^2} + 2i\omega_0 \frac{\partial E}{\partial t} - \omega_0^2 E\right] \exp[i(\omega_0 t - k_0 z)] + c.c. \quad (3.14)$$

$$\frac{\partial^2 \mathcal{E}}{\partial z^2} = \frac{1}{2}\left[\frac{\partial^2 E}{\partial z^2} - 2ik_0 \frac{\partial E}{\partial z} - k_0^2 E\right] \exp[i(\omega_0 t - k_0 z)] + c.c. \quad (3.15)$$

$$\frac{\partial^2 \mathcal{P}}{\partial t^2} = \frac{1}{2}\left[\frac{\partial^2 P}{\partial t^2} + 2i\omega_0 \frac{\partial P}{\partial t} - \omega_0^2 P\right] \exp[i(\omega_0 t - k_0 z)] + c.c. \quad (3.16)$$

As we mentioned above, we'll assume that derivatives are small and that derivatives of derivatives are even smaller:

$$\left|\frac{\partial^2 E}{\partial t^2}\right| \ll \left|2i\omega_0 \frac{\partial E}{\partial t}\right| \ll \left|\omega_0^2 E\right| \quad (3.17)$$

Letting $\omega_0 = 2\pi/T$, we find that this condition will be true as long as:

$$\left|\frac{\partial^2 E}{\partial t^2}\right| \ll \left|2\frac{2\pi}{T}\frac{\partial E}{\partial t}\right| \ll \left|\frac{4\pi^2}{T} E\right| \quad (3.18)$$

where T is the optical period of the light, again about 2 fs for visible light. These conditions hold if the field envelope is not changing on a time scale of a single cycle, which is nearly always true. So we can neglect the smallest term and keep the larger two.

The same is true for the spatial derivatives. We'll also neglect the second spatial derivative of the electric field envelope.

And the same derivatives arise for the polarization. But since the polarization is small to begin with, we'll neglect both the first and second derivatives.

The wave equation becomes:

$$\left[-2ik_0 \frac{\partial E}{dz} - \frac{2i\omega_0}{c^2}\frac{\partial E}{\partial t} - {k_0}^2 E + \frac{\omega_0^2}{c^2} E\right] \exp[i(\omega_0 t - k_0 z)]$$
$$= -\mu_0 \omega_0^2 P \exp[i(\omega_0 t - k_0 z)] \quad (3.19)$$

since we can factor out the complex exponentials.

We can also cancel the exponentials. Recalling that E satisfies the wave equation by itself, $k_0^2 E = (\omega_0^2/c^2)E$, and those two terms can also be canceled.

Then dividing through by $-2ik$ yields:

$$\frac{\partial E}{dz} + \frac{1}{c}\frac{\partial E}{\partial t} = -i\frac{\mu_0\omega_0^2}{2k_0}P \tag{3.20}$$

This expression is actually a bit oversimplified. A more accurate inclusion of dispersion (see Diels' and Rudolph's book) yields the same equation, but with the phase velocity of light, c, replaced with the group velocity, v_g:

$$\frac{\partial E}{dz} + \frac{1}{v_g}\frac{\partial E}{\partial t} = -i\frac{\mu_0\omega_0^2}{2k_0}P \tag{3.21}$$

We can now simplify this equation further by transforming the time co-ordinate to be centered on the pulse. This involves new space and time co-ordinates, z_v and t_v, given by: $z_v = z$ and $t_v = t - z/v_g$. To transform to these new co-ordinates requires replacing the derivatives:

$$\frac{\partial E}{\partial z} = \frac{\partial E}{\partial z_v}\frac{\partial z_v}{\partial z} + \frac{\partial E}{\partial t_v}\frac{\partial t_v}{\partial z} \tag{3.22}$$

$$\frac{\partial E}{\partial t} = \frac{\partial E}{\partial z_v}\frac{\partial z_v}{\partial t} + \frac{\partial E}{\partial t_v}\frac{\partial t_v}{\partial t} \tag{3.23}$$

Computing the simple derivatives and substituting, we find:

$$\frac{\partial E}{\partial z} = \frac{\partial E}{\partial z_v} + \frac{\partial E}{\partial t_v}\left[-\frac{1}{v_g}\right] \tag{3.24}$$

$$\frac{\partial E}{\partial t} = 0 + \frac{\partial E}{\partial t_v} \tag{3.25}$$

The time derivative of the polarization is also easily computed. This yields:

$$\frac{\partial E}{\partial z_v} + \frac{\partial E}{\partial t_v}\left[-\frac{1}{v_g}\right] + \frac{1}{v_g}\left[\frac{\partial E}{\partial t_v}\right] = -i\frac{\mu_0\omega_0^2}{2k_0}P \tag{3.26}$$

Canceling the identical terms leaves:

$$\boxed{\frac{\partial E}{\partial z} = -i\frac{\mu_0\omega_0^2}{2k_0}P} \tag{3.27}$$

where we've dropped the subscripts on t and z for simplicity. This nice simple equation is the SVEA equation for most nonlinear-optical processes in the simplest case. Assumptions that we've made to get here include that: (1) the nonlinear effects are weak; (2) the input beams are not affected by the fact that they're creating new beams (okay, so we're violating Conservation of

Energy here, but only by a little); (3) the group velocity is the same for all frequencies in the beams; (4) the beams are uniform spatially; (5) there is no diffraction; and (6) pulse variations occur only on time scales longer than a few cycles in both space and time. And we've assumed that the electric field and the polarization have the same frequency and k-vector. While the other assumptions mentioned above are probably reasonable in practical situations, this last assumption will be wrong in many cases—in fact it's actually difficult to satisfy, and we go to some trouble in order to do so—and we'll consider it in the next section. But the rest of these assumptions are quire reasonable in most pulse-measurement situations.

Solving the Wave Equation in the Slowly Varying Envelope Approximation

If the polarization envelope is constant, then the wave equation in the SVEA is the world's easiest differential equation to solve, and here's the solution:

$$E(z,t) = -\mathrm{i}\frac{\mu_0\omega_0^2}{2k_0} P\, z \tag{3.28}$$

and we see that the new field grows linearly with distance. Since the intensity is proportional to the mag-squared of the field, the intensity then simply grows quadratically with distance:

$$I(z,t) = \frac{c\mu_0\omega_0^2}{4} |P|^2\, z^2 \tag{3.29}$$

Phase-matching

There is a ubiquitous effect that must always be considered when we perform nonlinear optics and is another reason why nonlinear optics isn't part of our everyday lives. This is *phase-matching*. What it refers to is the tendency, when propagating through a nonlinear-optical medium, of the generated wave to become out of phase with the induced polarization after some distance. If this happens, then the induced polarization will create new light that's out of phase with the light it created earlier, and, instead of making more such light, the two contributions will *cancel out*. The way to avoid this is for the induced polarization and the light it creates to have the same phase velocities. Since they necessarily have the same frequencies, this corresponds to having the same k-vectors, the issue we discussed a couple of sections ago. Then the two waves are always in phase, and the process is orders of magnitude more efficient. In this case, we say that the process is *phase-matched.*

We've been implicitly assuming phase-matching so far by using the variable k_0 for both k-vectors. But because they can be different, let's reserve the variable, k_0, for the k-vector of the light at frequency ω_0 [$k_0 = \omega_0 n(\omega_0)/c_0$,

where c_0 is the speed of light in vacuum], and we'll now refer to the induced polarization's k-vector, as given by Eq. (3.11), as k_P. We must recognize that k_P won't necessarily equal k_0, the k-vector of light with the polarization's frequency ω_0—light that the induced polarization itself creates. Indeed, there's no reason whatsoever for the sum of the k-vectors above, all at different frequencies with their own refractive indices and directions, to equal $\omega_0 \mathrm{n}(\omega_0)/c_0$.

Equation (3.27) now becomes:

$$2ik_0 \frac{\partial E}{dz} \exp[\mathrm{i}(\omega_0 t - k_0 z)] = \mu_0 \omega_0^2 P \exp[\mathrm{i}(\omega_0 t - k_P z)] \tag{3.30}$$

Simplifying:

$$\frac{\partial E}{\partial z} = -\mathrm{i}\frac{\mu_0 \omega_0^2}{2k} P \exp(\mathrm{i}\,\Delta k\, z) \tag{3.31}$$

where:

$$\Delta k \equiv k_0 - k_P \tag{3.32}$$

We can solve this differential equation simply also:

$$E(L,t) = -\mathrm{i}\frac{\mu_0 \omega_0^2}{2k_0} P \left.\frac{\exp(\mathrm{i}\,\Delta k\, z)}{\mathrm{i}\,\Delta k}\right]_0^L \tag{3.33}$$

$$= -\mathrm{i}\frac{\mu_0 \omega_0^2}{2k_0} P \left[\frac{\exp(\mathrm{i}\,\Delta k\, L) - 1}{\mathrm{i}\,\Delta k}\right] \tag{3.34}$$

$$= -\mathrm{i}\frac{\mu_0 \omega_0^2 L}{k_0} P \exp(\mathrm{i}\,\Delta k\, L/2) \left[\frac{\exp(\mathrm{i}\,\Delta k\, L/2) - \exp(-\mathrm{i}\Delta k\, L/2)}{2\mathrm{i}\,\Delta k\, L}\right] \tag{3.35}$$

The expression in the brackets is $\sin(\Delta k L/2)/(\Delta k L/2)$, which is just the function called $\mathrm{sinc}(\Delta k\, L/2)$. Ignoring the phase factor, the light electric field after the nonlinear medium will be:

$$E(L,t) = -\mathrm{i}\frac{\mu_0 \omega_0^2}{k_0} P\, L\, \mathrm{sinc}(\Delta k\, L/2) \tag{3.36}$$

Mag-squaring to obtain the light irradiance or intensity, I, we have:

$$\boxed{I(L,t) = \frac{c\mu_0 \omega_0^2}{4} |P|^2\, L^2\, \mathrm{sinc}^2(\Delta k\, L/2)} \tag{3.37}$$

Since the function, $\mathrm{sinc}^2(x)$, is maximal at $x = 0$, and also highly peaked there (see Fig. 3.12), the nonlinear-optical effect of interest will experience much greater efficiency if $\Delta k = 0$. This confirms what we said earlier, that

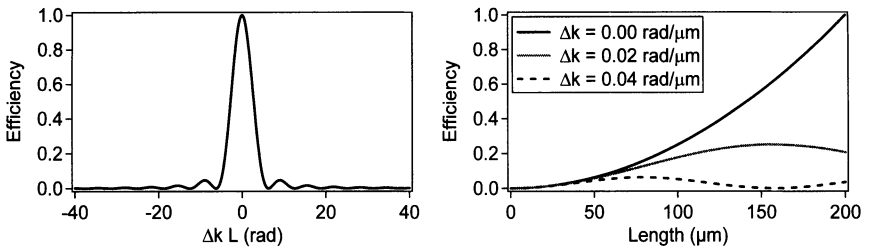

Fig. 3.12: Left: Plot of $\text{sinc}^2(\Delta kL/2)$ vs. ΔkL. Note that the sharp peak at $\Delta k\ L = 0$. Right: Plot of the generated intensity vs. L, the nonlinear-medium thickness for various values of Δk. Note that, when $\Delta k \neq 0$, the efficiency oscillates sinusoidally with distance and remains minimal for all values of L.

the nonlinear-optical efficiency will be maximized when the polarization and the light it creates remain in phase throughout the nonlinear medium, that is, when the process is *phase-matched.*

Phase-matching is crucial for creating more than just a few photons in a nonlinear-optical process. To summarize, the phase-matching conditions for an N-wave-mixing process are (see Fig. 3.11):

$$\omega_0 = \pm\omega_1 \pm \omega_2 \pm \cdots \pm \omega_N \tag{3.38}$$

$$\vec{k}_0 = \pm\vec{k}_1 \pm \vec{k}_2 \pm \cdots \pm \vec{k}_N \tag{3.39}$$

where k_0 is the k-vector of the beam at frequency, ω_0, which may or may not naturally equal the sum of the other k-vectors, and it's our job to make it so.

Note that, if we were to multiply these equations by $\hbar$, they would correspond to energy and momentum conservation for the photons involved in the nonlinear-optical interaction.

Let's consider phase-matching in collinear SHG. Let the input beam (often called the *fundamental beam*) have frequency ω_1 and k-vector, $k_1 = \omega_1 n(\omega_1)/c_0$. The second harmonic occurs at $\omega_0 = 2\omega_1$, which has the k-vector, $k_0 = 2\omega_1\, n(2\omega_1)/c_0$. But the induced polarization's k-vector has magnitude, $k_P = 2k_1 = 2\omega_1 n(\omega 1)/c_0$. The phase-matching condition becomes:

$$k_0 = 2\,k_1 \tag{3.40}$$

which, after canceling common factors $(2\omega_1/c_0)$ simplifies to:

$$\boxed{n(\omega_1) = n(2\omega_1)} \tag{3.41}$$

Thus, in order to phase-match SHG, it's necessary to find a nonlinear medium whose refractive indices at ω and 2ω are the same (to several decimal places). Unfortunately—and this is another reason you don't see things like this everyday—all media have dispersion, the tendency of the refractive index

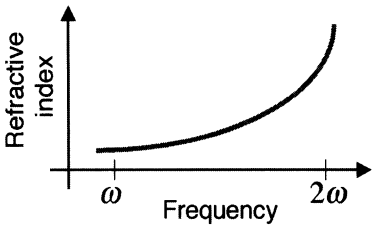

Fig. 3.13: Refractive index vs. wavelength for a typical medium. Because phase-matching SHG requires the refractive indices of the medium to be equal for both ω and 2ω, it is not possible to generate much second harmonic in normal media.

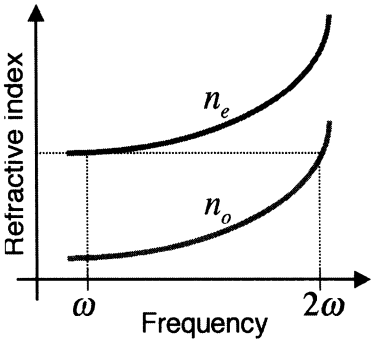

Fig. 3.14: Refractive index vs. wavelength for a typical *birefringent* medium. The two polarizations (say, vertical and horizontal, corresponding to the *ordinary* and *extraordinary* polarizations) see different refractive index curves. As a result, phase-matching of SHG is possible. This is the most common method for achieving phase-matching in SHG. The extra-ordinary refractive index curve depends on the beam propagation angle (and temperature), and thus can be shifted by varying the crystal angle in order to achieve the phase-matching condition.

to vary with wavelength (see Fig. 3.13). This effect quite effectively prevents seeing SHG in nearly all everyday situations.

It turns out to be possible to achieve phase-matching for birefringent crystals, whose refractive-index curves are different for the two orthogonal polarizations (see Fig. 3.14).

In *noncollinear* SHG, we must consider that there's an angle, θ, between the two beams (see Fig. 3.5). The input vectors have longitudinal and transverse components, but, by symmetry, the transverse components cancel out, leaving only the longitudinal component of the phase-matching equation:

$$k_1 \cos(\theta/2) + k_1 \cos(\theta/2) = k_0 \tag{3.42}$$

Simplifying, we have $2k_1 \cos(\theta/2) = k_0$ as our phase-matching condition. Substituting for the k-vectors, the phase-matching becomes:

$$n(\omega_1) \cos(\theta/2) = n(2\omega_1) \tag{3.43}$$

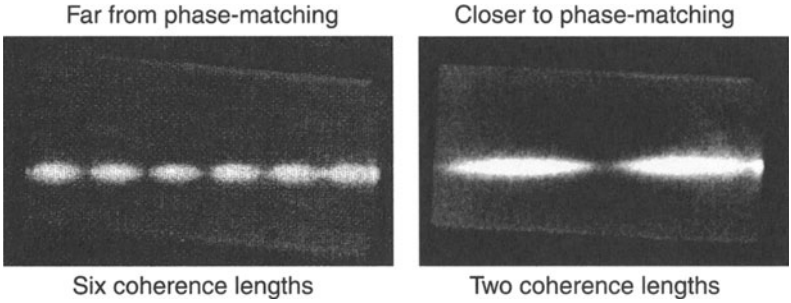

Fig. 3.15: Light inside a SHG crystal for two different amounts of phase-mismatch (i.e., for two different crystal angle orientations). Note that, as the crystal angle approaches the phase-matching condition, the periodicity of the intensity with position decreases, and the intensity increases. At phase-matching, the intensity increases quadratically along the crystal, achieving nearly 100% conversion efficiency in practice [14].

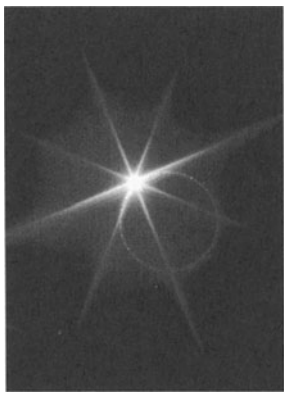

Fig. 3.16: Interesting non-collinear phase-matching effects in second-harmonic generation. (Picture taken by Rick Trebino.)

Figure 3.16 shows a nice display of noncollinear SHG phase-matching processes involving one intense beam and scattered light in essentially all directions. This picture doesn't yield any particular insights for pulse measurement, but it's really pretty, and we thought you might like to see it. By the way, the star isn't really nonlinear-optical; it's just due to the high intensity of the spot at its center (and the "star filter" on the camera lens when the picture was taken). The ring is real, however, and there can be as many as three of them.

Finally, whether a collinear or non-collinear beam geometry, it's also possible to achieve phase-matching using two orthogonal polarizations for the (two) input beams. In other words, the input beam is polarized at a 45° angle to the output SH beam. This is referred to as *Type II phase-matching*, while the above process is called *Type I phase-matching*. Type II phase-matching is

more complex than Type I because the two input beams have different refractive indices, phase velocities, and group velocities, which must be kept in mind when performing measurements using it.

Phase-matching is easier to achieve in third order, largely because we have an extra k-vector to play with. In fact, it can be so easy that it happens automatically. In two-beam coupling and polarization gating, the phase-matching equations become:

$$\omega_0 = \omega_1 - \omega_2 + \omega_2 \tag{3.44}$$

$$\vec{k}_0 = \vec{k}_1 - \vec{k}_2 + \vec{k}_2 \tag{3.45}$$

These equations are *automatically satisfied* when the signal beam has the same frequency and k-vector as beam 1: ω_1 and k_1, respectively.

For other third-order processes, phase-matching is not automatic, but it can be achieved with a little patience. For some processes, however, it can be impossible, as is the case for self-diffraction. In the latter case, sufficient efficiency can be achieved for most purposes, provided that the medium is kept thin to minimize the phase-mismatch.

Phase-Matching Bandwidth

Direct Calculation

While at most one frequency can be exactly phase-matched at any one time, some nonlinear-optical processes are more forgiving about this condition than others. Since it'll turn out to be important in pulse measurement to achieve efficient SHG (or other nonlinear-optical process) for all frequencies in the pulse, *phase-matching bandwidth* is an important issue. Figures 3.17a, b show the SHG efficiency vs. wavelength for two different crystals and for different incidence angles. Notice the huge variations in phase-matching efficiency for different crystal angles and thicknesses.

We can easily calculate the range of frequencies that will be approximately phase-matched in, for example, SHG. Assuming that the SHG process is exactly phase-matched at the wavelength, $\lambda 0$, the phase-mismatch, Δk, will be a function of wavelength:

$$\Delta k(\lambda) = 2k_1 - k_2 \tag{3.46}$$

$$\Delta k(\lambda) = 2\left[2\pi\frac{n(\lambda)}{\lambda}\right] - \left[2\pi\frac{n(\lambda/2)}{\lambda/2}\right] \tag{3.47}$$

$$\Delta k(\lambda) = \frac{4\pi}{\lambda}\left[n(\lambda) - n(\lambda/2)\right] \tag{3.48}$$

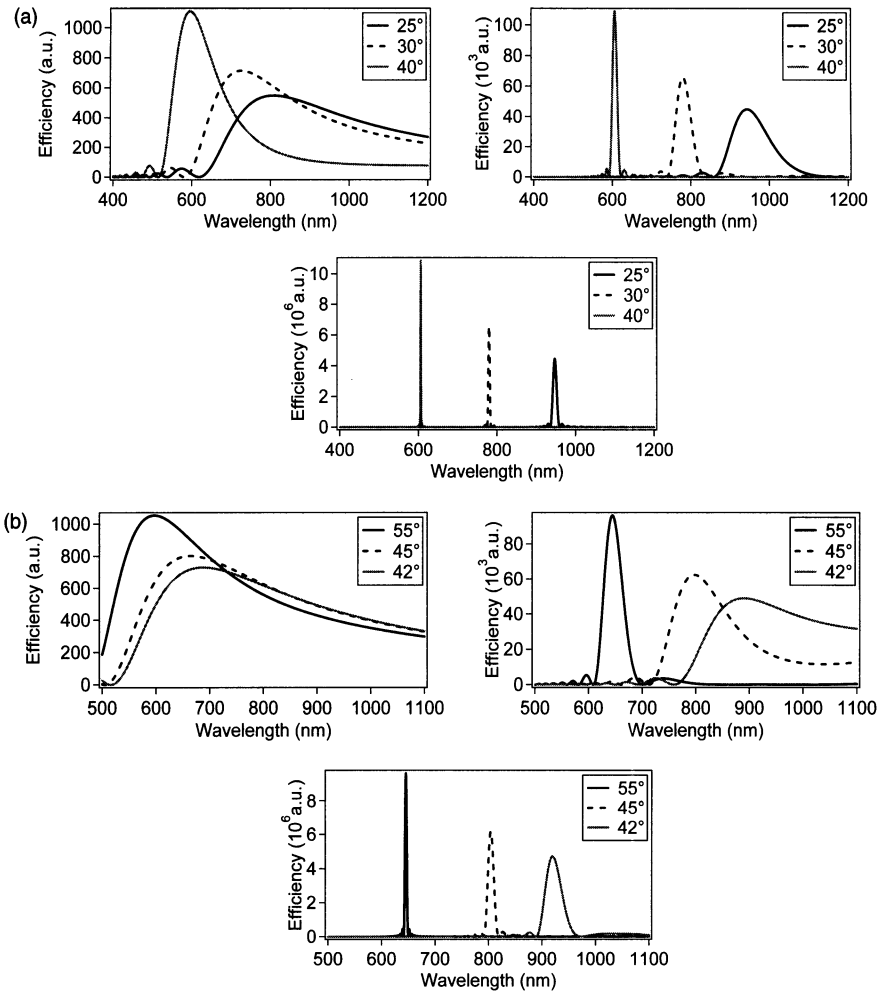

Fig. 3.17: (a) Phase-matching efficiency vs. wavelength for the nonlinear-optical crystal, beta-barium borate (BBO). Top left: a 10 μm thick crystal. Top right: a 100 μm thick crystal. Bottom: a 1000 μm thick crystal. These curves also take into account the ω_0^2 and L^2 factors in Eq. (3.25). While the curves are scaled in arbitrary units, the relative magnitudes can be compared among the three plots. (These curves do not, however, include the nonlinear susceptibility, $\chi^{(2)}$, so comparison of the efficiency curves in Figs. 3.17a and b requires inclusion of this factor.) (b) Same as Fig. 3.17a, except for the nonlinear-optical crystal, potassium di-hydrogen phosphate (KDP). Top left: a 10 μm thick crystal. Top right: a 100 μm thick crystal. Bottom: a 1000 μm thick crystal. The curves for the thin crystal don't fall to zero at long wavelengths because KDP simultaneously phase-matches for two wavelengths, that shown and a longer (IR) wavelength, whose phase-matching ranges begin to overlap when the crystal is thin.

Expanding $1/\lambda$ and the material dispersion to first order in the wavelength,

$$\Delta k(\delta\lambda) = \frac{4\pi}{\lambda_0}\left[1 - \frac{\delta\lambda}{\lambda_0}\right]\left[n(\lambda_0) + \delta\lambda\, n'(\lambda_0) - n(\lambda_0/2) - \frac{\delta\lambda}{2}n'(\lambda_0/2)\right] \tag{3.49}$$

where $\delta\lambda = \lambda - \lambda_0$, $n'(\lambda) = \mathrm{d}n/\mathrm{d}\lambda$ and we have taken into account the fact that, when the input wavelength changes by $\delta\lambda$, the second-harmonic wavelength changes by only $\delta\lambda/2$.

Recalling that the process is phase-matched for the input wavelength, λ_0, we note that $n(\lambda_0/2) - n(\lambda_0) = 0$, and we can simplify this expression:

$$\Delta k(\delta\lambda) = \frac{4\pi}{\lambda_0}\left[\delta\lambda\, n'(\lambda_0) - \frac{\delta\lambda}{2}n'(\lambda_0/2)\right] \tag{3.50}$$

where we have neglected second-order terms.

The sinc^2 curve will decrease by a factor of 2 when $\Delta k\, L/2 = \pm 1.39$. So solving for the wavelength range that yields $|\Delta k| < 2.78/L$, we find that the phase-matching bandwidth $\delta\lambda_{\mathrm{FWHM}}$ will be:

$$\boxed{\delta\lambda_{\mathrm{FWHM}} = \frac{0.44\,\lambda_0/L}{|n'(\lambda_0) - \frac{1}{2}n'(\lambda_0/2)|}} \tag{3.51}$$

Notice that $\delta\lambda_{\mathrm{FWHM}}$ is inversely proportional to the thickness of the nonlinear medium. Thus, in order to increase the phase-matching bandwidth, we must use a medium with dispersion such that $n'(\lambda_0) - \frac{1}{2}n'(\lambda_0/2) \approx 0$, or more commonly decrease the medium's thickness (see Fig. 3.18).

Finally, note the factor of 1/2 multiplying the second-harmonic refractive index derivative in Eq. (3.51). This factor does not appear in results appearing in some journal articles. These articles use a different derivative definition for the second harmonic [that is, $\mathrm{d}n/\mathrm{d}(\lambda/2)$] because the second harmonic necessarily varies by only one half as much as the fundamental wavelength. We, on the other hand, have used the same definition—the standard one, $\mathrm{d}n/\mathrm{d}\lambda$—for both derivatives, which, we think, is less confusing, but it yields the factor

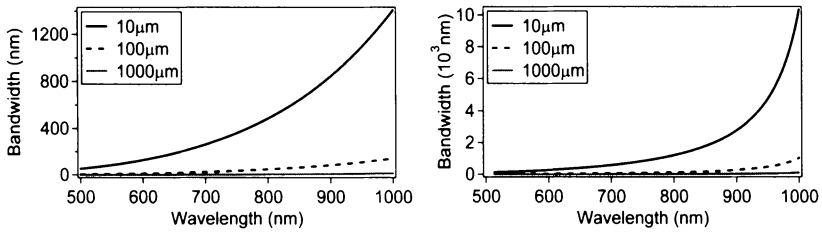

Fig. 3.18: Phase matching bandwidth vs. wavelength for BBO (left) and KDP (right).

of 1/2. It's easy to see that the factor of 1/2 is correct: assuming that the process is phase-matched at λ_0, maintaining a phase-matched process [i.e., $n(\lambda/2) = n(\lambda)$] requires that the variation in refractive index per unit wavelength near $\lambda_0/2$ be twice as great as that near λ_0, since the second harmonic wavelength only changes only half as fast as the fundamental wavelength.

Group-velocity Mismatch

There is an alternative approach for calculating the phase-matching bandwidth, which seems like a completely different effect until you realize that you get the same answer, and that it's just a time-domain approach, while the previous approach was in the frequency domain. Consider that the pulse entering the SHG crystal and the SH it creates may have the same phase velocities (they're phase-matched), but they could have different group velocities. This is called *group-velocity mismatch (GVM)*. If so, then the two pulses could cease to overlap after propagating some distance into the crystal; in this case, the efficiency will be reduced because SH light created at the back of the crystal will not coherently combine with SH light created in the front. This effect is illustrated in Fig. 3.19.

We can calculate the bandwidth of the light created when significant GVM occurs. Assuming that a very short pulse enters the crystal, the length of the SH pulse, δt, will be determined by the difference in light-travel times through the crystal:

$$\delta t = \frac{L}{v_g(\lambda_0/2)} - \frac{L}{v_g(\lambda_0)} = L\,\mathrm{GVM} \tag{3.52}$$

where $\mathrm{GVM} \equiv 1/v_g(\lambda_0/2) - 1/v_g(\lambda_0)$. This expression can be rewritten using expressions for the group velocity:

$$v_g(\lambda) = \frac{c_0/n(\lambda)}{1 - (\lambda/n(\lambda))n'(\lambda)} \tag{3.53}$$

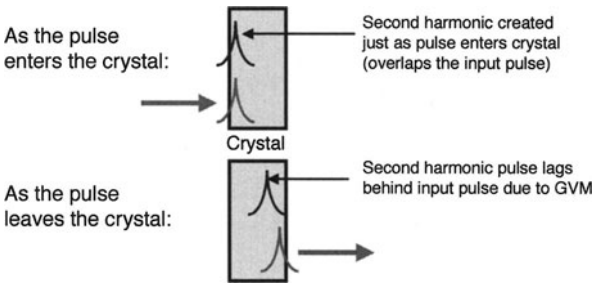

Fig. 3.19: Group-velocity mismatch. The pulse entering the crystal creates SH at the entrance, but this light travels at a different group velocity from that of the fundamental light, and light created at the exit does not coherently add to it.

Substituting for the group velocities in Eq. (3.52), we find:

$$\delta t = \frac{Ln(\lambda_0/2)}{c_0}\left[1 - \frac{\lambda_0/2}{n(\lambda_0/2)}n'(\lambda_0/2)\right] - \frac{Ln(\lambda_0)}{c_0}\left[1 - \frac{\lambda_0}{n(\lambda_0)}n'(\lambda_0)\right] \tag{3.54}$$

Now, recall that we wouldn't doing this calculation for a process that wasn't phase-matched, so we can take advantage of the fact that $n(\lambda_0/2) = n(\lambda_0)$. Things then simplify considerably:

$$\delta t = \frac{L\lambda_0}{c_0}\left[n'(\lambda_0) - \frac{1}{2}n'(\lambda_0/2)\right] \tag{3.55}$$

Take the second-harmonic pulse to have a Gaussian intensity, for which $\delta t\,\delta\nu = 0.44$. Rewriting in terms of the wavelength, $\delta t\,\delta\lambda = \delta t\,\delta\nu\,[\mathrm{d}\nu/\mathrm{d}\lambda]^{-1} = 0.44\,[\mathrm{d}\nu/\mathrm{d}\lambda]^{-1} = 0.44\,\lambda^2/c_0$, where we've neglected the minus sign since we're computing the bandwidth, which is inherently positive. So the bandwidth is:

$$\delta\lambda_{\mathrm{FWHM}} \approx \frac{0.44\,\lambda_0/L}{|n'(\lambda_0) - \frac{1}{2}n'(\lambda_0/2)|} \tag{3.56}$$

Note that the bandwidth calculated from GVM considerations precisely matches that calculated from phase-matching bandwidth considerations.

Phase-matching Bandwidth Conclusions

As we mentioned, in pulse-measurement devices, it's important to achieve efficient (or at least uniform) phase-matching for the entire bandwidth of the pulse. Since ultrashort laser pulses can have extremely large bandwidths (a 10 fs pulse at 800 nm has a bandwidth of over a hundred nm), it'll be necessary to use extremely thin SHG crystals. Crystals as thin as 5 μm have been used to measure few-fs pulses.

But also recall that the intensity of the phase-matched SH produced is proportional to L^2. So a very thin crystal yields very little signal intensity. Thus there is a nasty trade-off between efficiency and bandwidth. Fortunately, we can usually find a compromise—of just enough bandwidth and efficiency simultaneously. But, as with most compromises, we're not happy about it. As a result, we've spent much time thinking of tricks to beat this trade-off. Chapters 11 and 17 will discuss two different approaches.

Nonlinear-Optical Strengths

Just how strong are nonlinear-optical effects? Clearly they're not so strong that sunlight, even on the brightest day, efficiently produces enough of them for us to see. Of course, phase-matching also isn't happening.

Anyway, what sort of laser intensities are necessary to see these effects? We start with Eq. (3.36), which can be rewritten (with $\omega_0 = 2\omega$) in the form:

$$E^{2\omega}(L,t) = -\mathrm{i}\frac{2\mu_0\omega^2 L}{k} P \exp(\mathrm{i}\Delta kL/2)\,\mathrm{sinc}(\Delta kL/2) \tag{3.57}$$

where $P = \frac{1}{2}\varepsilon_0\chi^{(2)}(E^{\omega})^2$. Then, we relate intensity to electric field strength through $I = (n/2\eta_0)|E|^2$, where $\eta_0 = \sqrt{\mu_0/\varepsilon_0}$. With these, we re-write Eq. (3.57) in terms of intensities to find:

$$I^{2\omega} = \frac{\eta_0\omega^2(\chi^{(2)})^2(I^{\omega})^2L^2}{2{c_0}^2n^3}\,\mathrm{sinc}^2(\Delta kL/2) \tag{3.58}$$

Next, suppose we consider the best case, in which the process is phase-matched ($\mathrm{sinc}^2(0) = 1$) and re-write Eq. (3.58) in terms of a *SHG efficiency*:

$$\frac{I^{2\omega}}{I^{\omega}} = \frac{2\eta_0\omega^2d^2I^{\omega}L^2}{{c_0}^2n^3} \tag{3.59}$$

where we define the *d-coefficient* as $d = \frac{1}{2}\chi^{(2)}$. d is what we usually find quoted in handbooks. It will depend not only on the material, but also on the field configuration—how the fields are polarized with respect to the crystal orientation. Again, we refer you to a more detailed treatment of nonlinear optics to fully understand these issues. Our concern now is just to get some feel for the numbers involved and what we can hope to achieve in SHG efficiency in the lab. As a quick calculation, suppose we use beta-barium borate (BBO) as our nonlinear crystal, in which $d \approx 2 \times 10^{-12}$ m/V, and where $n \approx 1.6$ (note that we can get away with approximate values for n when it appears in an amplitude calculation, but we must have *very* accurate values for n when computing phase—or phase mismatch). If we wish to frequency-double an input beam of wavelength, $\lambda = 0.8\,\mu$m, we find from Eq. (3.59):

$$\frac{I^{2\omega}}{I^{\omega}} \approx 5 \times 10^{-8} I^{\omega}L^2 \tag{3.60}$$

where I is in W/m^2 and L is in m.

From the small coefficient in front, some pretty high intensities are needed for modest crystal lengths in order to get anything in the way of a decent efficiency! Suppose we consider an ultrafast laser. Basically, if you have an unamplified Ti : Sapphire laser, which produces nanojoule (nJ) pulses, 100 fs long, you have pulses with intensities on the order of 10^{14} W/m^2 (when focusing to a about a 10 μm spot diameter). But of course when focusing this tightly, the beam doesn't stay focused for long, which limits the crystal length we can use. Additionally, because ultrashort pulses are broadband, the requirement of phase matching the entire bandwidth limits the SHG crystal thickness to

considerably less than 1 mm, and usually less than 100 μm. Choosing a crystal length of 100 μm, and using the other numbers, we would achieve an efficiency of about 5%. This again is best-case for this configuration because 1) the beam does not stay focused to its minimum size throughout the entire length (as the above calculation assumes), and 2) d is reduced somewhat below its maximum value; this is because the fields are not necessarily at the best orientation within the crystal to most effectively excite the anharmonic oscillators. Phase matching decides the field orientation, and the price is paid through a slightly reduced nonlinear coefficient (known as d_{eff}). So we end up trying to optimize all of these parameters until we're satisfied with the SHG power we are getting. Then we stop.

This brings us to $\chi^{(3)}$. To get an idea of its order of magnitude for non-resonant materials, consider glass. Single mode optical fibers, made of glass, guide light with a cross-sectional beam diameter of slightly less than 10 μm. So we can achieve similar intensities that we saw before in our SHG example, but over much longer distances. In silica glass, $\chi^{(3)} \approx 2.4 \times 10^{-22} m^2/V^2$. One can make a comparison to a second order process by calculating the second and third order polarizations that result at a given light intensity. In our 100 fs 1 nJ pulse, focused to 10 μm diameter, the field strength is $E \approx 2.5 \times 10^8$ V/m. Then $\chi^{(3)}E \approx 6 \times 10^{-14}$ m/V. Compare this to $\chi^{(2)} = 2d \approx 4 \times 10^{-12}$ m/V for BBO. From here, the nonlinear polarizations for both processes are found by multiplying these results by the light intensity. As this example demonstrates, third-order processes in non-resonant materials are substantially weaker than second order processes. But this can be made up for sometimes by (1) tuning the frequency of one or more of the interacting waves near a material resonance (but at some cost in higher losses for those waves that are near resonance), or (2) taking advantage of long interactions lengths that may be possible in phase-matched situations (such as in optical fibers). Turning up the intensity will also help. Microjoule pulses can yield more than adequate signal energies from most of the third order nonlinear optical effects mentioned in this chapter. Third order bulk media typically used are fused silica and any glass for the various induced grating effects.

The above illustrations assumed 100 fs pulse intensities on the order of 10^{12} W/cm^2. However, with the less tight focusing that's practical in the lab, intensities more like 10^9 W/cm^2 are typically available. While this seems high, it's only enough to create barely detectable amounts of second harmonic. How about performing third-order nonlinear optics with such pulses? You can just barely do this in some cases, and it's a struggle. It's better to have a stage of amplification, especially from a regenerative amplifier ("regen"). Microjoule pulses can yield more than adequate signal energies from most of the third-order nonlinear-optical effects mentioned in this chapter. Third-order media typically used are fused silica and any glass for the various induced-grating effects. These media are actually not known for their high nonlinearities, but they are optically very clean and hence are the media of choice for pulse measurement applications.

Nonlinear Optics in 25 Words or Less

Okay, that was a lot to digest. So what's the minimum you need to know to understand the basic ideas of ultrashort-pulse measurement? Not much actually. For the next few chapters, we'll assume perfectly phase-matched interactions, and we won't worry about multiplicative constants, so all you need to remember is that the electric field of the nonlinear-optically generated light wave in this case is given by:

$$E_{\text{sig}}(t) \propto P \tag{3.61}$$

which is a simplified version of Eq. (3.28), and we're referring to the generated wave as the *signal field*, $E_{\text{sig}}(t)$. Also, for pulse-measurement applications, we'll typically be splitting a pulse into two using a beam-splitter (usually a 50%-reflecting mirror) and performing nonlinear optics with the pulse, $E(t)$ and another delayed version of itself, $E(t-\tau)$, where τ is the relative delay between the two pulses. For the various processes we've considered so far, the generated field will be:

$$E_{\text{sig}}(t,\tau) \propto \begin{cases} E(t)\,E(t-\tau) & \text{for SHG} \\ E(t)\,|E(t-\tau)|^2 & \text{for PG} \\ E(t)^2\,E^*(t-\tau) & \text{for SD} \\ E(t)^2\,E(t-\tau) & \text{for THG} \end{cases} \tag{3.62}$$

where we've included the delay in the functional dependence of the signal field. Finally, because we'll be mainly interested only in the pulse *shape*, we'll often neglect proportionality constants and just write, for example, $E(t) = E(t)\,E(t-\tau)$ for SHG.

That's all you really need to know. But you may still wish to read more on this fascinating subject, so here's a list of relevant books.

References

1. N. Bloembergen, *Nonlinear Optics*, World Scientific Pub. Co., 1996 (original edition: 1965).
2. R.W. Boyd, *Nonlinear Optics*, Academic Press, 1992.
3. P. Butcher and D. Cotter, *The Elements of Nonlinear Optics* Cambridge University Press, 1991.
4. J.-C. Diels and W. Rudolph, *Ultrashort Laser Pulse Phenomena*, Academic Press, 1996.
5. F.A. Hopf and G.I. Stegeman, *Applied Classical Electrodynamics: Nonlinear Optics*, Krieger Pub. Co., reprinted 1992.
6. D.L. Mills, *Nonlinear Optics: Basic Concepts*, 2nd ed., Springer Verlag, 1998.
7. G.G. Gurzadian et al., *Handbook of Nonlinear Optical Crystals, 3rd ed.*, Springer Verlag, 1999.
8. K.-S. Ho, S.H. Liu and G.S. He, *Physics of Nonlinear Optics*, World Scientific Pub. Co., 2000.
9. E.G. Sauter, *Nonlinear Optics*, Wiley-Interscience, 1996.
10. Y.R. Shen, *The Principles of Nonlinear Optics*, Wiley-Interscience, 1984.
11. A. Yariv, Quantum Electronics, 3rd ed. Wiley, 1989.
12. F. Zernike and J. Midwinter, *Applied Nonlinear Optics*, Wiley-Interscience, 1973 (out of print).
13. H.J. Eichler, et al., *Laser-Induced Dynamic Gratings*, Springer-Verlag, 1986.
14. Christian Rahlff, *Opt. Phot. News*, 8, 4, p. 64 (1997).