3

ULTRAFAST-PULSE MEASUREMENT METHODS

Up until now we have only discussed how to generate ultrashort pulses, not how to measure them. Measurement of pulses on the femtosecond time scale is an important issue, since the speed required is considerably faster than existing photodetectors and oscilloscopes. In this chapter we cover several measurement techniques capable of resolving femtosecond pulses. We begin by discussing correlation measurements, which are widespread experimentally but which offer only partial information about the optical pulse. We then cover several approaches that can reveal complete intensity and phase information, including the frequency-resolved optical gating approach based on time–frequency analysis concepts as well as techniques based on frequency filtering and on interferometry. Finally, we discuss methods for characterizing the intensity noise and timing jitter of mode-locked lasers based on radio-frequency spectral analysis.
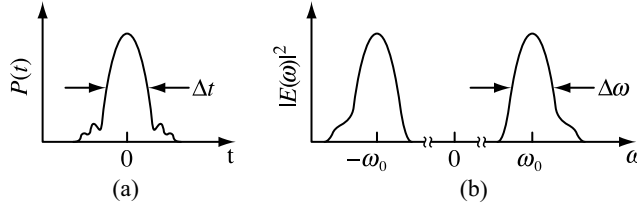
3.1 TERMINOLOGY AND DEFINITIONS

We start with some definitions. As usual we write the field as

$$e(t) = \text{Re}\{a(t)e^{j\omega_0 t}\} \tag{3.1}$$

with $a(t)$ normalized so that the power averaged over an optical cycle is given by

$$P(t) = |a(t)|^2 \tag{3.2}$$

Figure 3.1   Time profile of (a) power $P(t)$ and (b) power spectrum. Δt and $\Delta\omega$ are the full width at half maximum pulse width and bandwidth.

The intensity $I(t)$ is, of course, simply proportional to $P(t)$. The Fourier transform and the power spectrum are given by [see eq. (1.86a)]

$$E(\omega) = \frac{1}{2}\left[A(\omega - \omega_0) + A^*[-(\omega + \omega_0)]\right] \tag{3.3}$$

and

$$|E(\omega)|^2 = \frac{1}{4}\left[|A(\omega - \omega_0)|^2 + |A[-(\omega + \omega_0)]|^2\right] \tag{3.4}$$

where $A(\omega)$ is the Fourier transform of $a(t)$. In writing the power spectrum, we have assumed that the positive and negative frequency components of the spectrum do not overlap. The power and the power spectrum are sketched in Fig. 3.1.

The most widely used measure of the pulse duration in ultrafast optics is the intensity full width at half maximum (FWHM), which we denote Δt. The bandwidth is usually measured as the FWHM of the positive frequency part of the power spectrum, denoted $\Delta\nu$ (in hertz), where $\Delta\nu = \Delta\omega/2\pi$ and $\Delta\omega$ is the FWHM in angular frequency units. Δt and $\Delta\omega$ are sensitive only to the central portion of a pulse or its spectrum. Root-mean-square (rms) pulse widths or bandwidths have also been introduced for use in ultrafast optics [94]. The rms pulse width is defined

$$\Delta t_{\text{rms}} = 2\sqrt{\langle t^2 \rangle - \langle t \rangle^2} \tag{3.5}$$

where

$$\langle f(t) \rangle = \frac{\int_{-\infty}^{\infty} dt \, f(t) \, |a(t)|^2}{\int_{-\infty}^{\infty} dt \, |a(t)|^2} \tag{3.6}$$

Similarly, the rms bandwidth is defined

$$\Delta \nu_{\text{rms}} = \frac{1}{\pi}\sqrt{\langle \omega^2 \rangle - \langle \omega \rangle^2} \tag{3.7}$$

where

$$\langle G(\omega) \rangle = \frac{\int_{0}^{\infty} d\omega \, G(\omega) \, |A(\omega - \omega_0)|^2}{\int_{0}^{\infty} d\omega \, |A(\omega - \omega_0)|^2} \tag{3.8}$$

Table 3.1   Time–Bandwidth Products for Three Transform-Limited Pulse Shapes[a]

$a(t)$	$\|A(\omega)\|^2$	Δt	Δt_{rms}	$\Delta \nu\, \Delta t$	$\Delta \nu_{\text{rms}}\, \Delta t_{\text{rms}}$
$e^{-(t/t_p)^2}$	$e^{-(\omega^2 t_p^2/2)}$	$1.177 t_p$	t_p	0.441 $(2\ln 2/\pi)$	0.318 $(1/\pi)$
$\text{sech}\left(\dfrac{t}{t_p}\right)$	$\text{sech}^2\left(\dfrac{\pi\omega t_p}{2}\right)$	$1.763 t_p$	$1.814 t_p$ $\left(\pi t_p/\sqrt{3}\right)$	0.315	$\dfrac{1}{3}$
$\dfrac{1}{1+(t/t_p)^2}$	$e^{-2\|\omega t_p\|}$	$1.287 t_p$	$2 t_p$	0.142	0.450 $\left(\sqrt{2}/\pi\right)$

[a] Δt and $\Delta \nu$ are full widths at half maximum of $|a(t)|^2$ and $|A(\omega)|^2$, respectively. Δt_{rms} and $\Delta \nu_{\text{rms}}$ are the corresponding rms pulse width and spectral bandwidth. $|A(\omega)|^2$ is normalized to one.

In eq. (3.8) only the positive frequency part of the power spectrum is included in the integrals; this is necessary to give a result for $\langle \omega \rangle$ that is in the vicinity of ω_0. The rms pulse width (bandwidth) has sensitivity to energy in the wings of a pulse (spectrum) and therefore provides complementary information to the FWHM widths.

In measuring ultrashort pulses, one is interested not only in the intensity profile but also in any phase variations in either the frequency or time domains. Conversely, one is often interested in knowing when a pulse is free of such phase variations. We use $\phi(t)$ and $\psi(\omega)$, respectively, to define the temporal and spectral phases, as follows:

$$a(t) = |a(t)|\, e^{j\phi(t)} \tag{3.9}$$

and

$$A(\omega) = |A(\omega)|\, e^{j\psi(\omega)} \tag{3.10}$$

When $\psi(\omega)$ is a constant, we speak of the pulse as *bandwidth limited*.[1] When $\phi(t)$ is a constant, we say the pulse is free of *phase or frequency modulation*.[2] When $\phi(t) = 0$, $a(t)$ is real, and the Fourier transform gives $A(\omega) = A^*(-\omega)$. If, in addition, $a(t)$ is even, one finds $A(\omega) = A^*(\omega)$, which means that the pulse is bandwidth limited. Conversely, when $A(\omega)$ is real, one finds that $a(t) = a^*(-t)$. If, in addition, $A(\omega)$ is even, then $a(t) = a^*(t)$ and the pulse is free of phase modulation. We note that for a pulse to be simultaneously bandwidth limited and free of phase modulation [i.e., $\psi(\omega) = 0$ and $\phi(t) = 0$], it must also be an even function of both time and frequency. Therefore, it is possible for a bandwidth-limited pulse to show a phase modulation or a pulse free of any phase modulation to show a spectral phase variation. We speak of a pulse as *chirped* when it has a nontrivial variation in both $\phi(t)$ and $\psi(\omega)$.

One simple metric for characterizing the degree of chirp of a pulse is the time–bandwidth product $\Delta \nu\, \Delta t$, which is written in terms of the FWHMs. Values of $\Delta \nu\, \Delta t$ are given in Table 3.1 for three different simple chirp-free pulse shapes [94,95]. For a given pulse shape, an increased value of $\Delta \nu\, \Delta t$ is an indicator of chirp. The rms time–bandwidth product, written $\Delta \nu_{\text{rms}}\, \Delta t_{\text{rms}}$, is also useful [94]. Unlike the time–bandwidth product written in

---

[1] A pulse with $\psi(\omega)$ linear in ω is also considered bandwidth limited, since a linear spectral phase shift corresponds to a simple delay in the time domain.

[2] A pulse with $\phi(t)$ linear in t may also be considered as free of phase modulation, since a linear temporal phase is simply a shift of center frequency.

terms of FWHMs, it has the important property that it can be given a lower bound without specifying the pulse shape. Specifically, the *uncertainty relation*

$$\Delta \nu_{\text{rms}} \, \Delta t_{\text{rms}} \geq \frac{1}{\pi} \tag{3.11}$$

is always satisfied, independent of pulse shape. The minimum value occurs when the pulse shape is Gaussian [11].

3.2 ELECTRIC FIELD AUTOCORRELATION MEASUREMENTS AND THE POWER SPECTRUM

Since the ultrashort pulse itself is often the shortest event available, a common measurement strategy is to try to let the pulse sample itself. A simple setup for doing this is sketched in Fig. 3.2, where a Michelson interferometer is used to split the input pulse into two identical pulses which are then recombined with a time delay τ. The delay can be varied simply by moving one of the stages forward or backward, often using a computer-controlled translation stage with submicrometer step sizes. The simplest measurement consists of recording the average power (using a slow, time-averaging detector) emerging from the interferometer as a function of delay τ. As we shall see, this linear (in power) measurement technique is actually not capable of determining ultrashort pulse intensity or phase profiles. However, before proceeding to related nonlinear techniques which can yield the needed information, it is first important to understand what the linear methods tell us.

We now analyze the measurement outlined above [9,96]. The output field from the interferometer can be written

$$e_{\text{out}}(t) = \frac{1}{2} \text{Re}\left\{ a(t - t_1) e^{j\omega_0(t - t_1)} + a(t - t_2) e^{j\omega_0(t - t_2)} \right\} \tag{3.12}$$
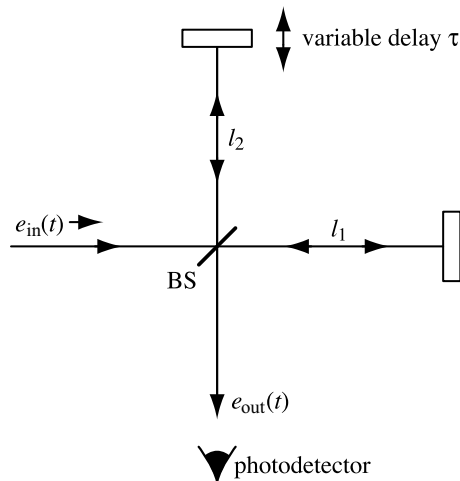
Figure 3.2   Michelson interferometer setup for measuring the electric field autocorrelation function.

The delays t_1 and t_2 are given by $t_1 = 2l_1/c$ and $t_2 = 2l_2/c$, respectively. The factor of 1/2 takes into account one reflection from and one pass through a 50–50 splitter. (The field amplitude is multiplied by the factor $1/\sqrt{2}$ after a single reflection from or transmission through the beamsplitter, resulting in 50% power reflected and 50% transmitted.) The output power is written

$$
\begin{aligned}
P_{\text{out}}(t) = \frac{1}{4} \big\{ &|a(t-t_1)|^2 + |a(t-t_2)|^2 \\
&+ \big[a(t-t_1)a^*(t-t_2)e^{j\omega_0(t_2-t_1)} + \text{c.c.} \big] \big\}
\end{aligned}
\tag{3.13}
$$

where "c.c." stands for *complex conjugate*. If we assume that the detector is slow compared to the pulse, so that it measures the total time-integrated intensity (i.e., the pulse energy), the measured quantity is given by

$$
\int dt \, P_{\text{out}}(t) = \frac{1}{2} \int dt \, |a(t)|^2 + \left[\frac{e^{j\omega_0\tau}}{4} \int dt \, a(t)a^*(t-\tau) + \text{c.c.} \right]
\tag{3.14}
$$

Here we have introduced the delay $\tau = t_2 - t_1$ and assumed a *time-stationary process* (i.e., the time-integrated intensity is independent of the pulse delay, and the integral of the cross terms depends only on the relative delay between the two pulses).

At this point we are ready to define the first-order correlation function $\Gamma_a(\tau)$ of the complex electric field amplitude $a(t)$, given by

$$
\Gamma_a(\tau) = \frac{1}{\Delta T} \int_{-\Delta T/2}^{\Delta T/2} dt \, a(t)a^*(t-\tau) = \big\langle a(t)a^*(t-\tau) \big\rangle
\tag{3.15}
$$

where ΔT is assumed to approach ∞, and the angular brackets $\langle \cdots \rangle$ denote the time average. In the ultrafast optics community, the quantity given in eq. (3.15) is often called simply the *field* or *amplitude autocorrelation*. The term *autocorrelation* means that the two pulses being correlated are identical. In the case of an *ergodic process*, the time average is the same as the ensemble average (a concept that is particularly useful in the case of a random or noise input electric field). Unless otherwise noted, we assume ergodicity, and therefore we do not need to distinguish between time and ensemble averages.

The output of the photodetector in Fig. 3.2 is proportional to the time-average power, which we now rewrite using eqs. (3.14) and (3.15) as follows:

$$
\langle P_{\text{out}} \rangle = \frac{1}{2}\Gamma_a(0) \left\{ 1 + \frac{1}{2} \left(\frac{\Gamma_a(\tau)e^{j\omega_0\tau} + \Gamma_a^*(\tau)e^{-j\omega_0\tau}}{\Gamma_a(0)} \right) \right\}
\tag{3.16}
$$

Since $a(t)$ is in general complex, $\Gamma_a(\tau)$ is also complex, and therefore we write $\Gamma_a(\tau) = |\Gamma_a(\tau)| \, e^{j\Phi(\tau)}$. Equation (3.16) now becomes

$$
\langle P_{\text{out}} \rangle = \frac{1}{2}\Gamma_a(0) \left[1 + G_1(\tau) \right]
\tag{3.17}
$$

where we have introduced the first-order correlation function of the electric field, $G_1(\tau)$, which is related to $\Gamma_a(\tau)$ by

$$G_1(\tau) = \frac{|\Gamma_a(\tau)|}{\Gamma_a(0)} \cos(\omega_0\tau + \Phi(\tau)) \tag{3.18}$$

To interpret these equations, we first note that $\Gamma_a(\tau) = \Gamma_a^*(-\tau)$, as one can easily verify by substitution into eq. (3.15). Consequently, $G_1(\tau)$ is even. We also note that $|\Gamma_a(\tau)| \, / \, |\Gamma_a(0)| \leq 1$. The maximum value occurs at $\tau = 0$. In the case of pulses, this corresponds to perfect temporal overlap between the recombined pulses emerging from the interferometer. For nonzero delays the overlap decreases and the value of the correlation function is reduced. It is typical to define a correlation time τ_c as the delay for which the correlation function is reduced to half of the peak value. A different way of defining τ_c uses a weighted average of τ defined by

$$\tau_c = \frac{\int_0^\infty \tau \, |\Gamma_a(\tau)| \, d\tau}{\int_0^\infty |\Gamma_a(\tau)| \, d\tau} \tag{3.19}$$

With either definition the overlap is very small for large delays ($|\tau| \gg \tau_c$), and the correlation approaches zero.

As one simple example, consider a Gaussian pulse with a complex amplitude given by $a(t) = \exp(-t^2/t_p^2)$. By direct integration, one can easily show that $\Gamma_a(\tau)/\Gamma_a(0) = \exp(-\tau^2/2t_p^2)$. An example of the time-average output power for a Gaussian pulse is sketched in Fig. 3.3. The power oscillates rapidly as a function of τ, with a period corresponding to one optical wavelength path-length change. Near $\tau = 0$ the power swings between zero and $\Gamma_a(0)$, which is the full input power. Thus, for phase differences of 0 or π, the power emerges completely in the output direction indicated in Fig. 3.2 or is reflected completely back toward the source, respectively. The average power (averaged over a few cycles of delay τ) is $\frac{1}{2}\Gamma_a(0)$ in either direction. The oscillations are damped as the
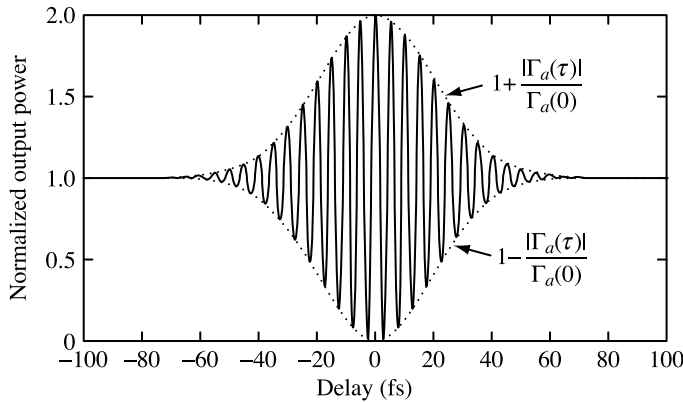
Figure 3.3 Example of time-average interferometer output power as a function of delay τ. The power is normalized by the factor $\frac{1}{2}\Gamma_a(0)$. This figure corresponds to a Gaussian input pulse with $t_p = 20$ fs and an optical period equal to 5 fs.

delay is increased, and the envelope of the oscillations, normalized by the average power, is $1 \pm |\Gamma_a(\tau)|$.

We see from this example that in the case of a coherent short-pulse input, the width of the electric field autocorrelation function is related in a simple way to the actual pulse width. The following question then arises: Can we determine the pulse width by measuring $G_1(\tau)$? In general, the answer is NO! To see this, we calculate the Fourier transform of $\Gamma_a(\tau)$ as follows:

$$\mathcal{F}\{\Gamma_a(\tau)\} \sim \int d\tau \, e^{-j\omega\tau} \int dt \, a(t)a^*(t-\tau) \tag{3.20}$$

By substituting $\tau = t - \eta$, we obtain

$$\mathcal{F}\{\Gamma_a(\tau)\} \sim \int dt \, a(t)e^{-j\omega t} \int d\eta \, e^{j\omega\eta}a^*(\eta) = |A(\omega)|^2 \tag{3.21}$$

Similarly, we can calculate the Fourier transform of $G_1(\tau)$, with the result

$$\mathcal{F}\{G_1(\tau)\} \sim |E_{\text{in}}(\omega)|^2 \tag{3.22}$$

These results say that the electric field correlation function is the inverse Fourier transform of the power spectrum, which can be measured, for example, by using a spectrometer. This has several important implications:

- The information obtained from measuring $\Gamma_a(\tau)$ and measuring the optical power spectrum is identical.
- The correlation time τ_c is roughly the inverse of the optical bandwidth.
- The correlation measurement gives no information about the spectral phase.
- Therefore, on the basis of these measurements, one cannot distinguish a bandwidth-limited pulse from a longer chirped pulse with the same bandwidth. In fact, a coherent ultrashort pulse and continuous-wave incoherent light (i.e., noise) give the same result, provided again that the optical spectra are the same.

On the other hand, if we somehow know a priori that the spectrum has a constant phase, we can determine the input pulse uniquely. We would start with the power spectrum, measured either directly using a monochromator or indirectly by measuring $\Gamma_a(\tau)$ and then doing the Fourier transform. Assuming for simplicity that the spectral phase is zero, we would take the square root of the power spectrum to get $A(\omega)$. Performing an inverse Fourier transform would then yield the input pulse $a(t)$.

3.3   ELECTRIC FIELD CROSS-CORRELATION MEASUREMENTS AND SPECTRAL INTERFEROMETRY

The situation changes when the correlation measurement involves two different input pulses, especially when one of the pulse shapes is known. This situation typically arises when the shape of the original optical signal is manipulated or distorted as the result of some linear or

nonlinear process occurring in the experiment. We refer to the known pulse as the *reference* and to the unknown pulse as the *signal*. If one wishes to characterize such pulse shape changes, one can perform a cross-correlation measurement by splitting off a portion of the original light from the laser to use as a reference. In some cases the shape of the reference pulse is measured independently using one of the methods described later in the chapter.

3.3.1   Electric Field Cross-Correlation

We denote the reference and signal pulses as $a_r(t)$ and $a_s(t)$, respectively, and we assume that the reference pulse is delayed by time τ relative to the signal. The total electric field after the pulses are combined by a beamsplitter is given by

$$e_{\text{out}}(t) = \frac{1}{\sqrt{2}} \, \text{Re}\Big\{ \big[a_s(t) + a_r(t-\tau)e^{-j\omega_0\tau} \big] \, e^{j\omega_0 t} \Big\} \tag{3.23}$$

The time-integrated output power becomes

$$\int dt \, P_{\text{out}}(t) = \frac{1}{2} \int dt \, \Big\{ |a_r(t)|^2 + |a_s(t)|^2 + \big[a_s(t)a_r^*(t-\tau)e^{j\omega_0\tau} + \text{c.c.} \big] \Big\} \tag{3.24}$$

or

$$\langle P_{\text{out}}(t) \rangle = \frac{1}{2} \Big\{ U_r + U_s + \big[e^{j\omega_0\tau} \langle a_s(t)a_r^*(t-\tau) \rangle + \text{c.c.} \big] \Big\} \tag{3.25}$$

where U_r and U_s are the energies of the reference and signal pulses, respectively. The factor of 1/4 in eq. (3.13) is changed to 1/2 since we assume that the reference and signal are already separate at the input to the cross-correlation setup; therefore, the beamsplitter is needed only once (to combine the beams). In the case where the reference pulse is known to be much shorter than the signal pulse, it can be approximated as a delta function. The time-integrated output power then reduces to

$$\langle P_{\text{out}}(t) \rangle = \frac{1}{2} U_r + \frac{1}{2} U_s + \text{Re}\big\{ a_s(\tau)e^{j\omega_0\tau} \big\} \tag{3.26}$$

Thus, recording the oscillatory part of the output signal as a function of τ gives a direct measurement of the signal field. The envelope of the oscillations gives $|a_s(t)|$; measuring the phase of the fringes yields the phase of $a_s(t)$. When $a_r(t)$ is not truly a delta function but is still much narrower than $a_s(t)$, the cross-correlation yields an approximate but somewhat broadened measurement of the signal field.

It is also useful to analyze the electric field cross-correlation measurement in the frequency domain. The Fourier transform of eq. (3.24) with respect to τ is given by

$$\mathcal{F}\left\{ \int dt \, P_{\text{out}}(t) \right\}$$

$$= \cdots + \frac{1}{2} \big[A_s(\omega-\omega_0)A_r^*(\omega-\omega_0) + A_s^*(-\omega-\omega_0)A_r(-\omega-\omega_0) \big] \tag{3.27}$$

where the noninterferometric terms have been omitted. Provided that the reference spectrum is known and is broad compared to that of the signal, the signal spectrum and therefore the signal pulse can be obtained by deconvolution [i.e., by dividing the first term on the right-hand side of eq. (3.27) by $A_r^*(\omega - \omega_0)$]. The requirement of a broad reference spectrum is equivalent to the requirement that the reference pulse be short compared to the signal pulse.

From an experimental perspective, measuring the phase of the fringes requires a great deal of care. In particular, an online calibration of the sweep in the length of the interferometer is usually required in order to account for the possibly irregular motion of the delay stage. This can be achieved by counting the fringes made by a narrowband reference laser, such as a He–Ne laser, which is simultaneously passed through the interferometer [97,98].

3.3.2 Spectral Interferometry

In another measurement technique called *spectral interferometry*, τ is kept fixed, and a spectrometer is used to measure the output power spectrum [99–101]. In addition to its application in ultrafast optics, this method has also been applied for characterization of short optical fibers [102,103]. In terms of the envelope functions, the power spectrum can be written

$$|A_{\text{out}}(\tilde{\omega})|^2 = \frac{1}{2}\left\{|A_s(\tilde{\omega})|^2 + |A_r(\tilde{\omega})|^2\right\}$$

$$+ \frac{1}{2}\left\{\left[A_s(\tilde{\omega})A_r^*(\tilde{\omega})e^{j(\omega_0+\tilde{\omega})\tau} + \text{c.c.}\right]\right\} \quad (3.28)$$

Here $|E_{\text{out}}(\omega)|^2$ is related to $|A_{\text{out}}(\tilde{\omega})|^2$ through eq. (3.4) and we use the symbol $\tilde{\omega} = \omega - \omega_0$ to emphasize that $A_{\text{out}}(\tilde{\omega})$ is a baseband function peaked around zero frequency. If we now write $A_s(\tilde{\omega}) = |A_s(\tilde{\omega})|\, e^{j\psi_s(\tilde{\omega})}$ and $A_r(\tilde{\omega}) = |A_r(\tilde{\omega})|\, e^{j\psi_r(\tilde{\omega})}$, eq. (3.28) becomes

$$|A_{\text{out}}(\tilde{\omega})|^2 = \frac{1}{2}\left\{|A_s(\tilde{\omega})|^2 + |A_r(\tilde{\omega})|^2\right\}$$

$$+ |A_s(\tilde{\omega})|\, |A_r(\tilde{\omega})| \cos\left((\omega_0 + \tilde{\omega})\tau + \psi_s(\tilde{\omega}) - \psi_r(\tilde{\omega})\right) \quad (3.29)$$

The overall power spectrum exhibits oscillations with period $2\pi/\tau$, which are modulated by the spectral phase variations of the pulses. If one knows the spectral amplitude and phase of the reference pulse, one can derive the spectral phase of the signal pulse from the oscillatory part of the power spectrum. The spectral amplitude of the signal pulse $|A_s(\tilde{\omega})|$ can also be recovered from the spectral interferometry signal; however, it is often simpler to determine $|A_s(\tilde{\omega})|$ from the signal power spectrum $|A_s(\tilde{\omega})|^2$, which can be measured directly by blocking the reference beam.

Thus, spectral interferometry can be used to determine the complete amplitude and phase of the signal pulse, provided that a well-characterized reference pulse is available. For accurate results it is important that the reference spectrum be as broad or broader than that of the pulse being measured (otherwise, the oscillatory cross-term in the power spectrum will vanish at some frequencies of interest). This requirement is consistent with the requirement that the reference pulse must be shorter than the signal pulse, which was already mentioned in connection with time-domain electric field cross-correlation measurements.

One important form of spectral interferometry arises when the signal pulse is related to the reference pulse through a linear filtering operation. As shown in Fig. 3.4, both pulses
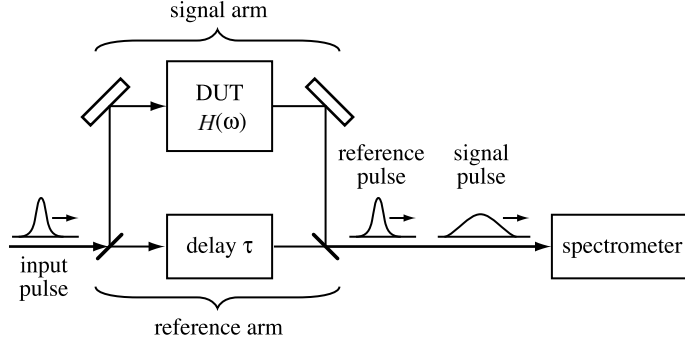
Figure 3.4   Spectral interferometry setup for characterizing the frequency response of linear optical elements. DUT, device under test.

are created from a common input pulse $a_{in}(t)$ and recombined after passing through an imbalanced interferometer in which the signal arm contains an optical component or device under test. Let us call the frequency response of the device under test $H(\omega)$, defined by $E_{out}(\omega) = H(\omega)E_{in}(\omega)$. Equivalently, the device under test has an impulse response function $h(t) = \mathcal{F}^{-1}\{H(\omega)\}$. Then the reference and signal pulses are given, respectively, by

$$A_r(\tilde{\omega}) = \frac{1}{\sqrt{2}}A_{in}(\tilde{\omega}) \qquad (3.30a)$$

and

$$A_s(\tilde{\omega}) = \frac{1}{\sqrt{2}}H(\omega_0 + \tilde{\omega})A_{in}(\tilde{\omega}) \qquad (3.30b)$$

Equation (3.28) for the output power spectrum now becomes

$$|A_{out}(\tilde{\omega})|^2 = \frac{1}{4}\left\{\left[1 + |H(\omega)|^2\right] + H(\omega)e^{j\omega\tau} + H^*(\omega)e^{-j\omega\tau}\right\}|A_{in}(\tilde{\omega})|^2 \qquad (3.31)$$

If we separate the phase and amplitude response of the component under test, $H(\omega) = |H(\omega)|\,e^{j\psi(\omega)}$, the total power spectrum becomes

$$|A_{out}(\tilde{\omega})|^2 = \frac{1}{4}\left\{1 + |H(\omega)|^2 + 2\,|H(\omega)|\cos(\omega\tau + \psi(\omega))\right\}|A_{in}(\tilde{\omega})|^2 \qquad (3.32)$$

Thus, the oscillatory part of the spectral interferometry signal allows one to determine the spectral phase and amplitude response of the component under test. This makes spectral interferometry a valuable tool for measuring the dispersion of fibers, mirrors, prism sequences, and many other optical elements.

We briefly outline one algorithm for recovering the transfer function $H(\omega)$ based on Fourier transform processing of the spectral interferometry data. The result of taking the inverse Fourier transform of eq. (3.31) is sketched schematically in Fig. 3.5 for the case $\tau > 0$. The feature centered at $t = 0$, marked a, arises from the term proportional to $\left[1 + |H(\omega)|^2\right]|A_{in}(\tilde{\omega})|^2$ in eq. (3.31) and is simply the incoherent superposition of the field
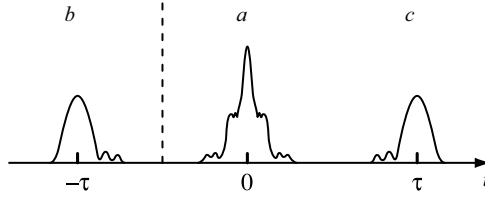
Figure 3.5   Fourier transform of spectral interferometry data for the case $\tau > 0$. Features marked a, b, and c correspond to the first, second, and final terms, respectively, in eq. (3.31). Feature b contains the desired information.

autocorrelation envelopes of the reference and signal pulses. This term does not contain any phase information. The term marked b is located at $t = -\tau$ and has the form

$$\Gamma_a(t) * \left[h(t + \tau)e^{-j\omega_0 t} \right]$$

which represents a baseband version of the impulse response function convolved with the autocorrelation of $a(t)$. $h(t)$ and $H(\omega)$ can be extracted from this term provided that the autocorrelation is sufficiently short (this is equivalent to the requirement of a sufficiently broad spectrum). The third term, marked c and located at $t = \tau$, involves a time-reversed version of $h(t)$ but is otherwise similar to b. The algorithm proceeds as follows:

1. The spectral interferometry data is first inverse-Fourier-transformed.
2. Terms a and c are set to zero. Provided that $\tau > 0$, as in our example, this means that everything to the right of the dashed vertical line in the figure is replaced by zero. Note that one needs to know the sign of τ in order to distinguish correctly between terms b and c.
3. Now the result is Fourier transformed back into the frequency domain and the delay is suppressed. There are two equivalent procedures for this:
   (a) Term b in the time-domain data can simply be shifted to $t = 0$ before inverse Fourier transforming back to the frequency domain.
   (b) Alternatively, term b can be Fourier-transformed back into the frequency domain without any temporal shifting, yielding $H(\omega) |A_{in}(\tilde{\omega})|^2 e^{j\omega\tau}$. The linear spectral phase term (which can be obtained by curve fitting) is then subtracted off.
4. Finally, the result is divided by $|A_{in}(\tilde{\omega})|^2$. This leaves $H(\omega)$, as desired.

It should be clear from Fig. 3.5 that this algorithm places a constraint on τ; namely, $|\tau|$ should be sufficiently large to avoid overlap between terms a and b. In the frequency domain, this constraint ensures that the period of the spectral fringes is finer than the fastest spectral variation in $H(\omega)$ and $|E_{in}(\omega)|^2$ (i.e., the spectral fringes are able to adequately sample the desired spectral phase information). As an extreme case, consider the result of setting $\tau = 0$. In this case the last term in eq. (3.32) becomes proportional to $\cos(\psi(\omega))$. Since the cosine is an even function, information is clearly lost. On the other hand, $H(\omega)$ can be recovered even for $\tau = 0$ provided that both $\sin(\psi(\omega))$ and $\cos(\psi(\omega))$ are available. The sine function can be obtained by repeating the spectral interferometry measurement after introducing an additional $\pi/2$ relative phase shift between the signal and reference

beams. From an experimental perspective, however, it is usually preferred to acquire a single spectral interferometry trace at sufficiently large τ.

A few additional points are worth mentioning:

- Since electric field cross-correlation and spectral interferometry are linear techniques, they can be applied to low power signals (unlike some of the nonlinear optical methods discussed later in the chapter).
- The phase of the input signal does not appear in eq. (3.32). Therefore, in order to characterize the complex spectral response of a linear optical system, as shown in Fig. 3.4, the input field need not be a short pulse. Incoherent light can also be used to characterize the complex spectral response of the component under test as long as the input spectrum covers the spectral range of interest. On the other hand, if one wishes to extract the signal field waveform itself, this can be done only if the reference pulse is known. These remarks also apply to time-domain electric field cross-correlation measurements.
- In spectral interferometry there is no need for an accurate calibration of the lengths of the interferometer arms; the measurement itself yields the relative delay. This makes spectral interferometry somewhat simpler to implement than electric field cross-correlation, where an independent online length calibration is required.

3.3.3   Application: Optical Coherence Tomography

In addition to waveform measurement, interferometry techniques have important applications in imaging and depth profiling. Here we briefly introduce one such application, *optical coherence tomography* (OCT), which has found extensive application for imaging within biological samples [104,105].

A basic setup for OCT is sketched in Fig. 3.6a. The apparatus is based on electric field cross-correlation and is usually implemented using fiber optic components. A broadband light source is directed to a 50–50 coupler which splits the light into a reference arm equipped with a retroreflecting mirror and a sample arm. Back-scattered light from the sample and retroreflected reference light are recombined in the coupler and measured using a slow detector as the reference mirror is scanned longitudinally. The measurement principle

(a)   (b)
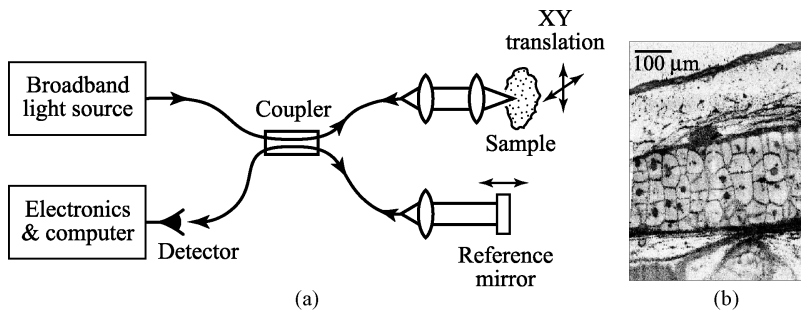
Figure 3.6   (a) Setup for optical coherence tomography. Detection of interference between back-scattered light from the sample and retroreflected light from a scanned reference mirror provides depth-resolved data on scattering microstructure. An image is formed by scanning the beam across the sample. (b) OCT image of live African frog tadpole tissue, showing subcellular resolution. Image from [106].

is that interference between reference and sample arms is observed only when the delay difference is less than the coherence time. Consequently, detecting the interference acts as a coherence gate. The depth into the sample corresponding to zero delay depends on the delay of the reference arm. As a result, the interferogram recorded as the reference mirror is scanned provides a profile of scattering strength vs. depth. Images are built up by using a focused beam and recording a sequence of depth scans at different transverse positions on the sample. The transverse position on the sample may be varied either by translating the sample, as depicted in the figure, or by deflecting the beam.

OCT has been used extensively to image microstructure in both transparent and non-transparent tissue. The eye is the prime example of transparent tissue, whereas most other biological tissues are highly scattering. In either case the ability to select time slices of the back-scattered field through coherence gating is the key factor. The dynamic range of such measurements can reach 110 dB, which means that the system is sensitive to reflected or scattered powers as low as 10^{-11} of the input power [106]. In highly scattering media where attenuation is very strong, this allows imaging to millimeter depths, which is not possible using standard optical microscopy techniques. An example of an OCT image demonstrating subcellular spatial resolution is shown in Fig. 3.6b.

Usually, very good spatial resolution is desired. An advantage of OCT is that transverse resolution and depth resolution are essentially decoupled: Transverse resolution is determined by the focused spot size, while depth resolution is determined by the coherence time of the light source. In particular, depth resolution Δz is given by

$$\Delta z = \frac{v_g \tau_c}{2} \tag{3.33}$$

where τ_c here is the FWHM of the electric field autocorrelation function, v_g is the group velocity in the sample, and a factor of 2 arises due to the double pass through the sample in a back-scattering geometry. Because coherence time is inversely proportional to optical bandwidth, sources with broad bandwidth are preferred. Note that as in other electric field correlation measurements, coherent short pulses are not required. Temporally incoherent sources are also acceptable as long as they have broad bandwidth and good spatial properties. Furthermore, dispersion compensation is not required. As long as the individual arms of the interferometer are adjusted to have identical dispersion, there is no effect on the interferometry measurement. It is this point that makes OCT compatible with fiber optic implementations.

These points are illustrated in Fig. 3.7, which shows optical spectra and interference fringes obtained in an autocorrelation geometry (sample replaced by a second retroreflecting mirror) [106]. Data are shown for two different light sources: a mode-locked Ti:sapphire (Ti:S) laser with a 260-nm bandwidth and a superluminescent diode (SLD) with a 32-nm bandwidth. The SLD is a semiconductor diode device similar to a semiconductor injection laser but with lasing suppressed. Hence, the SLD provides light that is spatially coherent but temporally incoherent. The interference fringes give the electric field autocorrelation functions, which are equal to the inverse Fourier transforms of the optical power spectra. The depth resolutions are 11.5 and 1.5 μm for the SLD and Ti:S laser, respectively, in inverse proportion to the optical bandwidths.

A few additional points are worthy of comment. First, unlike the simulated data of Fig. 3.3, the interferometry data of Fig. 3.7 have zero mean value. This occurs because the results shown in Fig. 3.7 were acquired using a differential measurement scheme, not
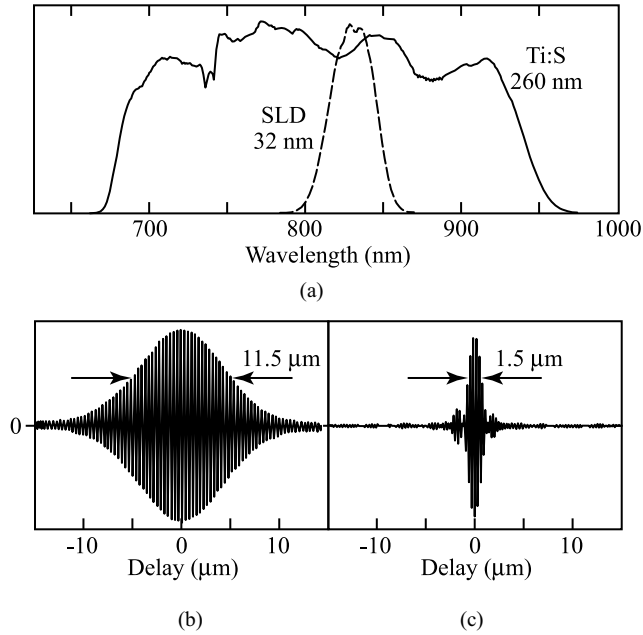
Figure 3.7   Comparison of optical power spectra and resulting OCT spatial resolution for superluminescent diode (SLD) and mode-locked Ti:S laser light sources: (a) power spectra; (b,c) electric field autocorrelation interference fringes from (b) SLD and (c) Ti:S laser. Adapted from [106].

depicted in Fig. 3.2 or 3.6. Referring to Fig. 3.2, there are actually two interferometer outputs, one in the direction of the photodetector, as portrayed, the other in a direction back toward the source. By detecting both outputs with a pair of photodetectors and differencing the result, one obtains a zero-mean signal in which only the fringe terms remain. This differential scheme has the practical advantage of suppressing intensity noise for increased sensitivity. A second point is that although autocorrelations are supposed to be symmetric, close inspection of the Ti:S correlation signal in Fig. 3.7 reveals a slight asymmetry. Such asymmetry may arise from small differences in the dispersion of the two interferometer arms. Thus, although compensation of dispersion common to both interferometer arms is not required, for very broadband signals great care must be exercised to equalize the dispersion experienced by the interfering signals.

A number of different modalities exist for processing of OCT signals. Arguably, the most basic difference between various schemes is the choice of coherent vs. incoherent demodulation. These two main approaches are described briefly below.

- In *incoherent demodulation* approaches, one is concerned only with the envelope of interference fringes. The positions of peaks in the interferogram envelope reveal the depths of strongly back-scattering structures, much as delays of return signals in a pulsed radar system yield a target range. Phase information associated with the interference fringes is not used.

- In *coherent demodulation* approaches, one acquires and analyzes the full interferometric signal, including phase and frequency information associated with the fringes themselves. In this mode, OCT measurements become fully analogous to electric field

cross-correlation measurements described in Section 3.3.1. With full fringe data one may extract new information through Fourier transform analysis. For example, in Doppler OCT one assigns shifts in interferogram spectra to shifts in the mean frequency of back-scattered light [107,108]. Such frequency shifts provide information on scatterer velocities, projected in the direction of light propagation, through the Doppler effect. Furthermore, by analyzing Doppler OCT data in terms of joint time–frequency distributions (see Section 3.5), it becomes possible to acquire depth-resolved velocity images. This approach has enabled mapping of flow in subsurface blood vessels in living tissues.

3.4 INTENSITY CORRELATION MEASUREMENTS

3.4.1 Correlation Measurements Using Second-Harmonic Generation

We have already seen that in the case of a single unknown pulse, a common measurement strategy is to arrange for the pulse to sample itself. Mathematically, this means that one measures a correlation function of the pulse. We have also seen that measurement of the electric field autocorrelation function using a linear Michelson interferometer does not provide any new information about the pulse. To perform a more useful measurement, one can insert a nonlinear element into the interferometer. The most common approach involves second-harmonic generation (SHG), in which a nonlinear crystal is used to generate light at twice the input optical frequency (i.e., at the second harmonic $2\omega_0$ of the input light at the fundamental frequency ω_0). The measurement procedure is to record the time-averaged second-harmonic power as a function of the relative delay τ between the two identical versions of the input pulse [109–112]. Due to the nonlinearity, the total energy in the second-harmonic pulse is greater when the two pulses incident on the nonlinear crystal overlap in time. Therefore, the peak in second-harmonic power plotted as a function of τ contains information about the pulse width. For an excellent discussion of SHG and other nonlinear correlation techniques, see [113].

We now analyze this SHG measurement in the collinear geometry shown in Fig. 3.8. We assume (1) that the intensity is low enough to avoid saturation associated with pump
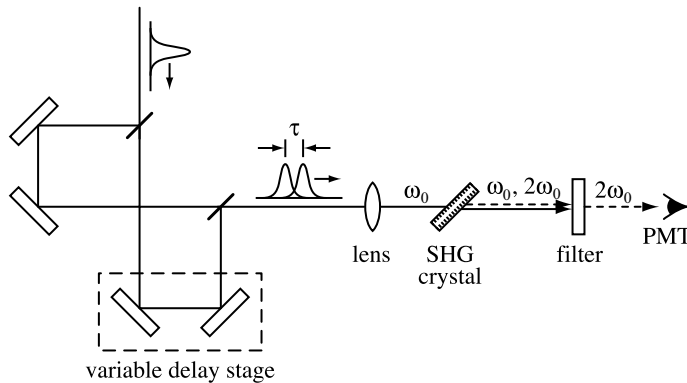
Figure 3.8 Collinear second-harmonic generation geometry for measurement of the intensity autocorrelation function. PMT, photomultiplier tube.

depletion and (2) that the bandwidth of the SHG process is large enough to accommodate the entire spectrum. The first assumption is easily met in practice. The second criterion requires the use of thin nonlinear crystals, especially for very short pulses, so that second-harmonic generation is *phase matched* over a sufficiently broad spectrum. The concept of phase matching and its implications for the bandwidth of the SHG process are discussed in Chapter 5.

The field emerging out of the interferometer and incident on the SHG crystal is written

$$e_{\text{out}}(t) = \frac{1}{2}\text{Re}\{a(t)e^{j\omega_0 t} + a(t-\tau)e^{j\omega_0(t-\tau)}\} \tag{3.34}$$

The field at the second-harmonic frequency, which is proportional to the square of input field, is given by

$$
\begin{aligned}
e_{\text{SHG}}(t) &\sim e_{\text{out}}^2(t) \\
&\sim \text{Re}\{[a^2(t) + a^2(t-\tau)e^{-2j\omega_0\tau} + 2a(t)a(t-\tau)e^{-j\omega_0\tau}]e^{2j\omega_0 t}\}
\end{aligned} \tag{3.35}
$$

The actual measurement records the time-integrated (or equivalently, the time-averaged) second-harmonic power (not the instantaneous power, since this would require an ultrafast detector, which we have assumed is not available). After some algebra one obtains the following expression:

$$
\begin{aligned}
\langle P_{\text{SHG}}(t)\rangle \sim{} & 2\langle|a(t)|^4\rangle + 4\langle|a(t)|^2|a(t-\tau)|^2\rangle + \{\{2\langle(|a(t)|^2 + |a(t-\tau)|^2) \\
& \times a(t)a^*(t-\tau)\rangle e^{j\omega_0\tau} + \langle[a(t)a^*(t-\tau)]^2\rangle e^{2j\omega_0\tau}\} + \text{c.c.}\}
\end{aligned} \tag{3.36}
$$

In arriving at this expression, we have assumed a time-stationary process, as previously. The first term in eq. (3.36) is independent of τ, and the second term varies relatively slowly with τ. The third and fourth terms represent interferometric contributions at frequencies ω_0 and $2\omega_0$, respectively, which oscillate rapidly as τ is varied over one optical period.
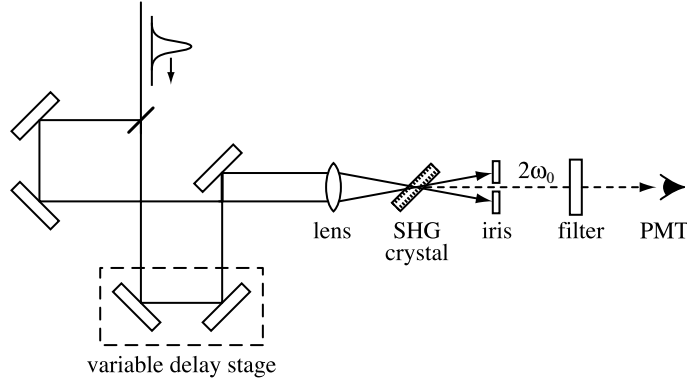
Fringe-Averaged Autocorrelation   Often, one suppresses the interferometric terms by averaging the SHG signal over a few optical periods. The measured time-average second-harmonic power then becomes

$$\langle P_{\text{SHG}}(t)\rangle \sim \left\langle|a(t)|^4\right\rangle\left(1 + \frac{2\langle|a(t)|^2\,|a(t-\tau)|^2\rangle}{\langle|a(t)|^4\rangle}\right) \tag{3.37}$$

At this point we define the normalized intensity autocorrelation function $G_2(\tau)$:

$$G_2(\tau) = \frac{\langle I(t)I(t-\tau)\rangle}{\langle|I(t)|^2\rangle} \tag{3.38}$$

Figure 3.9 Noncollinear second-harmonic generation geometry for background-free intensity autocorrelation function measurement. PMT, photomultiplier tube.

where $I(t) \sim |a(t)|^2$ as usual. Equation (3.37) for the time-average second harmonic power in a collinear measurement geometry then becomes

$$\langle P_{\text{SHG}}(t) \rangle \sim 1 + 2G_2(\tau) \tag{3.39}$$

The measured second-harmonic power is the sum of a constant background term (which represents the SHG from each pulse individually) and twice the intensity autocorrelation function (which arises from the interaction of the two time-delayed pulses).

   Another common experimental arrangement uses a noncollinear geometry (Fig. 3.9), in which the second harmonic due to the interaction between the time-delayed pulses is emitted in a different direction than that due to the individual pulses. In this case, the constant background terms can be suppressed, both by using an iris and by adjusting the SHG phase-matching conditions. The output signal in the background-free noncollinear geometry becomes

$$\langle P_{\text{SHG}}(t) \rangle \sim G_2(\tau) \tag{3.40}$$

Note that the noncollinear geometry automatically averages out the fringe terms in eq. (3.36).

   It is useful at this point to comment on the properties of $G_2(\tau)$. Several important points are listed below.

- The normalized intensity correlation function assumes its maximum value of unity at $\tau = 0$.
- In the case of finite-duration pulses, the pulses exhibit no overlap for delays much greater than the pulse width, and the correlation function goes to zero. The width of the correlation peak gives information about the pulse width.
- The correlation function is an even function of τ, independent of the symmetry of the actual pulse. Therefore, one cannot uniquely recover the pulse intensity profile from $G_2(\tau)$.

- If one somehow knew a priori that the intensity profile were symmetric, one could indeed recover $I(t)$ from $G_2(\tau)$. We note that the Fourier transform of $G_2(\tau)$ is $\left|\tilde{I}(\omega)\right|^2$, where $\tilde{I}(\omega)$ is the Fourier transform of $I(t)$. If $I(t)$ is symmetric, $\tilde{I}(\omega)$ must be real; therefore, $\tilde{I}(\omega)$ can be obtained by taking the square root of $\left|\tilde{I}(\omega)\right|^2$.
- Conversely, in the usual case where $I(t)$ is not known to be symmetric, $\tilde{I}(\omega)$ cannot be assumed real. Therefore, $G_2(\tau)$ only gives $\left|\tilde{I}(\omega)\right|$ and not its phase, which would be needed to reconstruct $I(t)$.
- By definition, the intensity autocorrelation function is also insensitive to the phase of the electric field itself. Therefore, the fringe-averaged intensity autocorrelation $G_2(\tau)$ cannot be used to characterize chirps or phase modulations on the pulse.

It is also worth discussing the behavior of $G_2(\tau)$ in the case of continuous-wave noise [113]. The intensity correlation function still takes on its maximum value of unity at $\tau = 0$. However, for sufficiently large values of τ, the two relatively delayed versions of the same noise signal are uncorrelated. Since the intensity is always nonnegative, the correlation never drops to zero, but the value is reduced below unity. This results in a coherence peak in the intensity autocorrelation function. It is important to note that this coherence peak has nothing to do with the actual pulse width; the width of the peak is governed by the correlation time of the intensity fluctuations, which can be as short as the inverse of the optical bandwidth. For the specific common case of a Gaussian random field (i.e., the real and imaginary parts of the electric field are independent, zero-mean random variables described by Gaussian probability distributions with equal variances), the probability distribution $p_I(I)$ for the intensity is given by

$$p_I(I) = \frac{1}{\langle I \rangle} e^{-I/\langle I \rangle} \tag{3.41}$$

where $\langle I \rangle$ is the average intensity. At delays large enough so that the two noise signals are independent, we get

$$G_2(\tau) = \frac{\langle I(t) \rangle^2}{\langle I^2(t) \rangle} = \frac{1}{2} \tag{3.42}$$

where we have used $\langle f(I) \rangle = \int dI \, f(I) p_I(I)$.

These findings are summarized in Fig. 3.10. In the collinear measurement geometry, the correlation function for a single coherent pulse decays smoothly to a constant background level, with a $3:1$ contrast ratio between the peak at $\tau = 0$ and the background. This theoretical contrast ratio can be used as a check for proper experimental alignment. In the case of continuous-wave noise, the only feature is the coherence spike centered near $\tau = 0$, with a contrast ratio of $3:2$. The intermediate case of a finite-duration noise burst has a coherence spike centered on top of a finite-duration pedestal, with a contrast ratio of $3:2:1$. It is important not to use the duration of the coherence spike to estimate the pulse duration, and it also important to note that the intensity autocorrelation of a pulse with a smooth intensity profile but containing excess bandwidth in the form of a chirp will not exhibit a coherence spike, which arises only from substructure in the intensity. One can estimate the pulse duration from the full width at half maximum (FWHM) of the correlation trace (excluding coherence spike), denoted $\Delta\tau$ in the figure.
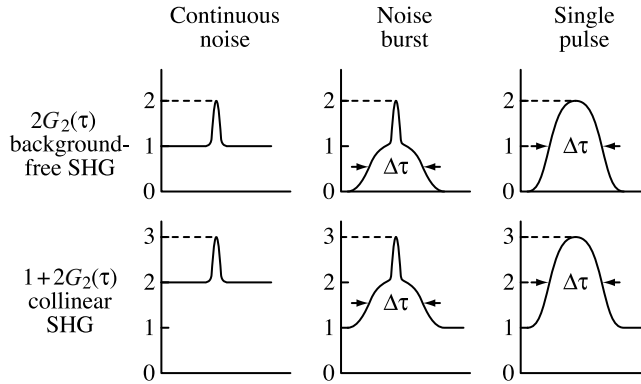
Figure 3.10 Theoretical traces for collinear and noncollinear SHG autocorrelation measurements. From Fig. 3.3 of [113], with permission from Springer Science and Business Media.

The situation is similar in the noncollinear measurement geometry, except that there is no constant background for finite-duration pulses. The zero background makes it easier to identify small amounts of energy in the wings of the pulse. The contrast ratio is 2 : 1 for continuous-wave noise and 2 : 1 : 0 for a finite-duration noise burst.

To estimate the intensity FWHM pulse width from the autocorrelation, one must assume a specific pulse shape. The intensity FWHM (Δt) is then determined by dividing the intensity autocorrelation FWHM ($\Delta \tau$) by a pulse-shape-specific deconvolution factor. See Table 3.2 for autocorrelation functions and deconvolution factors for a few different pulse shapes [113,114]. For smoothly varying pulse shapes (e.g., Gaussian or sech^2 pulses), the variation in the deconvolution factor is of order 10%, and therefore the pulse duration can be obtained approximately, even though the pulse shape is not, in fact, known. For less-well-behaved pulses (e.g., rectangular or single-sided exponential pulses), the deconvolution factor can vary by a larger amount. To do a consistency check, one can fit the autocorrelation data to the functional form of $G_2(\tau)$ corresponding to the assumed pulse shape. It is also useful to compare the measured time–bandwidth product $\Delta \nu \, \Delta t$ with that corresponding to the assumed pulse shape. A discrepancy between the experimental and theoretical time–bandwidth product indicates either a deviation from the assumed pulse shape or the presence of chirp. Note, however, that these consistency checks do not

Table 3.2   Autocorrelation Functions and Deconvolution Factors for Four Pulse Shapes[a]

$I(t)$	$G_2(\tau)$	$\dfrac{\Delta \tau}{\Delta t}$
$\text{sq}\,(t)$	$1 - \lvert\tau\rvert$ for $\lvert\tau\rvert \leq 1$, 0 otherwise	1
e^{-2t^2}	$e^{-\tau^2}$	$\sqrt{2}$
$\text{sech}^2\,(t)$	$\dfrac{3\,(\tau\cosh(\tau) - \sinh(\tau))}{\sinh^3(\tau)}$	1.543
e^{-t} for $t \geq 0$, 0 otherwise	$e^{-\lvert\tau\rvert}$	2

[a] Δt and $\Delta \tau$ are full widths at half maximum of the intensity $I(t)$ and the intensity autocorrelation function $G_2(\tau)$, respectively. $\text{sq}(u)$ is defined as a unit square pulse such that $\text{sq}(u) = 1$ for $\lvert u\rvert \leq \frac{1}{2}$ and 0 otherwise; and t and τ are in normalized units.

assure uniqueness (i.e., it is still possible that the actual pulse shape deviates from the shape assumed).

Additional information is available using the root-mean-square (rms) pulse width Δt_{rms}, as discussed in [94]. The definition of the rms pulse width was given in Section 3.1. A very useful property of the rms pulse width is that it can be found directly from the intensity autocorrelation function without any assumption about the pulse shape using the following formula:

$$(\Delta t_{\text{rms}})^2 = \frac{2 \int d\tau \, \tau^2 G_2(\tau)}{\int d\tau \, G_2(\tau)} \tag{3.43}$$

It was also demonstrated that for a given power spectrum $|E(\omega)|^2$, Δt_{rms} assumes its minimum value $\Delta t_{\text{rms}}^{(0)}$ when $E(\omega)$ has a constant spectral phase [i.e., when $E(\omega)$ is bandwidth-limited]. A similar relation is often assumed for the FWHM pulse width Δt, but there is no proof of this assumption. Based on this minimum rms pulse width theorem, Sorokin et al. [94] also introduced an rms chirp parameter

$$C = \sqrt{\langle \tau^2(\omega) \rangle - \langle \tau(\omega) \rangle^2} \tag{3.44}$$

which they showed could also be expressed in terms of the rms pulse widths:

$$C = \frac{1}{2} \sqrt{(\Delta t_{\text{rms}})^2 - \left(\Delta t_{\text{rms}}^{(0)} \right)^2} \tag{3.45}$$

In these expressions $\tau(\omega)$ is the frequency-dependent delay as defined in eq. (2.82), and eq. (3.8) is used. An increase in Δt_{rms} over the bandwidth-limited value $\Delta t_{\text{rms}}^{(0)}$ gives a direct indication of spectral chirp, as quantified through the C parameter. An attractive feature of this definition is that both Δt_{rms} and $\Delta t_{\text{rms}}^{(0)}$ can be obtained directly from commonly available experimental diagnostics: Δt_{rms} from intensity autocorrelation through eq. (3.43) and $\Delta t_{\text{rms}}^{(0)}$ by taking the square root of the power spectrum, assuming constant phase, and Fourier transforming to get $I(t)$.
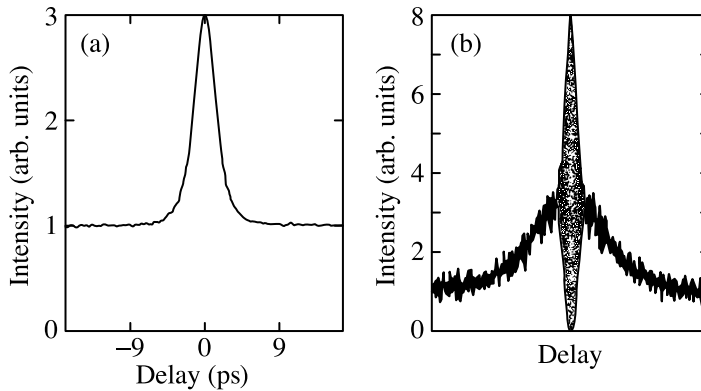
Interferometric Autocorrelation   We now return to the full expression describing the SHG pulse-width measurement including interferometric terms, eq. (3.36). Detailed information on interpreting interferometric autocorrelation data is given in [95,114]. The interferometric terms include some phase and coherence information that is lacking in $G_2(\tau)$ itself. These terms may be recorded by using a collinear measurement geometry and exercising care to avoid averaging over the rapid interferometric oscillations. As an example, consider the data plotted in Figs. 3.11 and 3.12, which correspond to pulses from a self-mode-locked Ti:S. Figure 3.11 corresponds to pulses obtained when a prism pair is used for intracavity dispersion compensation. The intensity autocorrelation function is smooth and yields a 60-fs intensity FWHM pulse duration assuming a sech^2 intensity profile. The width of the power spectrum is such that the time–bandwidth product is $\Delta \nu \, \Delta t \approx 0.33$, in close agreement with the theoretical value of 0.315 for secant hyperbolic pulses. Thus, these pulses appear to be close to *bandwidth limited*, meaning that (1) the pulse duration is as short as it can be given the mode-locked bandwidth, and (2) there is little or no chirp. The most notable
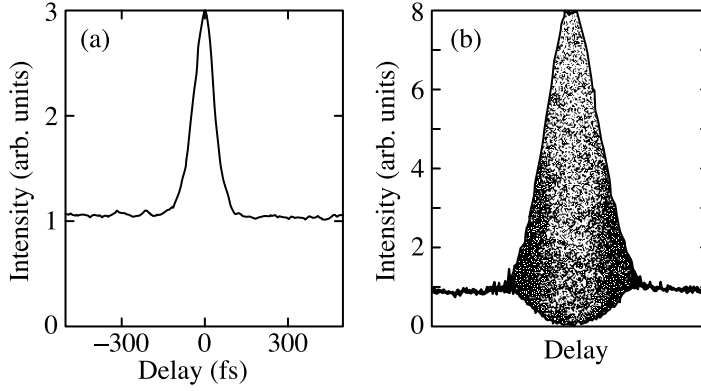
Figure 3.11 (a) Intensity autocorrelation and (b) interferometric autocorrelation for nearly bandwidth-limited pulses from a mode-locked Ti:S laser. A 60-fs pulse width and a time–bandwidth product $\Delta\nu\,\Delta t = 0.33$ are calculated assuming a $sech^2$ intensity profile. For (b), the delay axis is zoomed by roughly a factor of 3 compared to (a). From [52].

feature of the interferometric autocorrelation is that the fringes last throughout the entire pulse duration. This is another sign that the pulse is bandwidth limited.

The situation is rather different in Fig. 3.12, which corresponds to data obtained when the prism pair is removed from the laser. In this case the dispersion in the laser cavity is dominated by the Ti:S crystal and is large and positive. Comparing with Figs. 2.19 and 2.20, one expects relatively long mode-locked pulses with a strong chirp. This expectation is confirmed by the measurements. The intensity autocorrelation remains smooth (no coherence spike), but with the intensity FWHM pulse duration broadened to 2 ps (again assuming $sech^2$ pulses). At the same time the spectral width is narrowed from 14 nm to 3.5 nm. The calculated time–bandwidth product increases to 2.7, indicating that the pulses are far from the bandwidth limit. The appearance of the interferometric autocorrelation is changed dramatically: The fringes now occur only near the center of the trace and die off

Figure 3.12 (a) Intensity autocorrelation and (b) interferometric autocorrelation for highly chirped pulses from a mode-locked Ti:S laser. A 2-ps pulse width and a time–bandwidth product $\Delta\nu\,\Delta t = 2.7$ are calculated assuming a $sech^2$ intensity profile. For (b), the delay axis is zoomed by roughly a factor of 3 compared to (a). From [52].

much faster than the nonoscillating portion of the experimental trace. The envelope of the fringes is related to the inverse of the spectral bandwidth (but in a more complicated way than in electric field autocorrelation). The data indicate that the coherence time of the pulses is much shorter than the duration of the intensity profile, consistent with the large $\Delta\nu\,\Delta t$ product and a strong chirp. This behavior should not be confused with coherence spikes which can be observed in measurements of $G_2(\tau)$, which arise due to noise substructure on the intensity profile and are insensitive to chirp.

It is also worth discussing the contrast ratio observed in interferometric autocorrelation measurements. The upper envelope of the fringes results when the two time-delayed fields are exactly in phase and add to produce a maximum total field; the lower envelope results when the two fields are exactly out of phase and cancel to produce the minimum total field. For the case where the coherence time is essentially equal to the pulse width (i.e., the temporal phase is essentially constant), the upper and lower envelopes are proportional to

$$\text{upper envelope} \sim \left\langle\, [|a(t)| + |a(t-\tau)|]^4 \,\right\rangle \tag{3.46a}$$

$$\text{lower envelope} \sim \left\langle\, [|a(t)| - |a(t-\tau)|]^4 \,\right\rangle \tag{3.46b}$$

where the two pulses are also assumed to have equal amplitudes. When there is significant phase modulation (i.e., the coherence time is shorter than the pulse width), this equation applies only for $|\tau| \ll \tau_c$; for $|\tau| \gg \tau_c$ the fringes disappear and only the nonoscillating terms given by eq. (3.39) remain. In all cases the background level observed for large delays when the pulses no longer overlap is proportional to

$$\text{background} \sim 2\langle|a(t)|^4\rangle \tag{3.47}$$

From these formulas we can easily see that the contrast ratio between the peak of the upper fringe envelope and the background level should be 8 : 1, as observed in the data shown above. The lower envelope is at the zero level near $\tau = 0$, also as observed. If one averages over the fringes, one gets back to $1 + 2G_2(\tau)$ as indicated in eq. (3.39), with a contrast ratio of 3 : 1 between the peak and the background. (This last statement is not obvious from Figs. 3.11 and 3.12 but is true nevertheless!)

Note that the presence or absence of strong chirp in Figs. 3.11 and 3.12 could be deduced from the time–bandwidth product alone, without using the interferometric autocorrelation. The interferometric autocorrelation is useful because it can serve as a clear visual indicator of moderate to large chirp and because it provides additional data for quantitative pulse-shape analysis. It has been shown that if one measures the term varying as $\exp\{2j\omega_0\tau\}$ in the interferometric autocorrelation, that is,

$$\Xi(\tau) = \left\langle\, \left[a(t)a^*(t-\tau)\right]^2 \,\right\rangle e^{2j\omega_0\tau}$$

then together with the fringe-averaged autocorrelation $G_2(\tau)$ and the power spectrum of the input pulse $|A(\omega - \omega_0)|^2$, this is enough information to determine the complete electric field profile uniquely[3] [115]. Reconstruction of the electric field from these data has been

---

[3] The electric field is determined uniquely except for a time-reversal ambiguity; $a(t)$ and $a^*(-t)$ are both allowed solutions. Furthermore, the proof assumes that the field has finite support (i.e., it is strictly zero outside a finite time window).
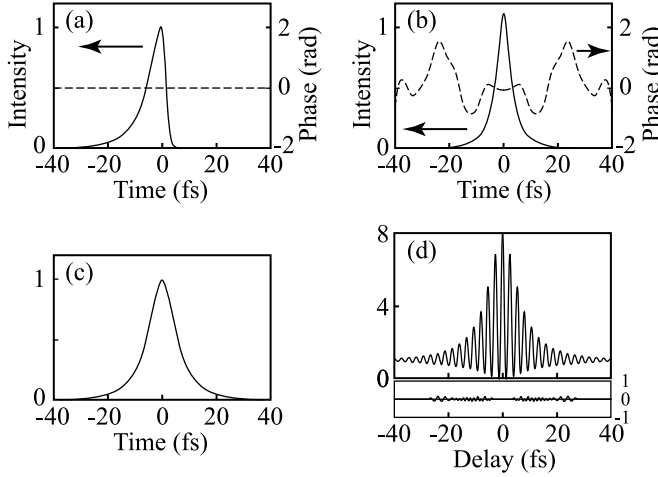
Figure 3.13 Calculation results for a pair of pulses constructed to have identical fringe-averaged autocorrelation traces and power spectra: (a) asymmetric pulse; (b) symmetric pulse; (c) fringe-averaged autocorrelation trace, identical for both pulses; (d) interferometric autocorrelation of asymmetric pulse (top) and difference in asymmetric and symmetric pulse interferometric autocorrelations (bottom). Adapted from [116]. Copyright ©2001, IEEE.

demonstrated by using iterative phase retrieval techniques related to those described later in Section 3.6.5. Practically, this technique has seen limited use, possibly because of the need for high measurement accuracy to ensure valid reconstruction of $a(t)$. It is also interesting that in the case of a pulse free of frequency modulation, the term multiplying $\exp\{2j\omega_0\tau\}$ in $\Xi(\tau)$ is identical to $G_2(\tau)$. Therefore, by carefully comparing these two terms, one can directly check whether $a(t)$ is real. Finally, we note that $\Xi(\tau)$ is simply the electric field autocorrelation of the second-harmonic pulse $a^2(t)$ produced in response to a single pulse illuminating the second-harmonic crystal. Therefore, measurement of the power spectrum of a single second-harmonic pulse gives equivalent information.

These points are illustrated in Fig. 3.13, which shows calculation results for a pair of pulses constructed to have identical fringe-averaged autocorrelation traces and power spectra. One of the pulses (Fig. 3.13a) has an asymmetric intensity profile with a 28-fs FHWM pulse width and a flat temporal phase, while the other (Fig. 3.13b) has a symmetric intensity profile with a 37-fs FWHM pulse width and phase modulation. $G_2(\tau)$ is shown in Fig. 3.13c and is identical for both pulses, as stated. This demonstrates the ambiguity inherent in fringe-averaged intensity autocorrelation measurements. Figure 3.13d shows the interferometric autocorrelation of the asymmetric pulse. The interferometric autocorrelation of the symmetric pulse is almost identical and is not included in the main frame of Fig. 3.13d. Instead, the difference between the two interferometric autocorrelations is plotted in the lower part of Fig. 3.13d. The rms difference between the two traces is less than 1%. Thus, although the interferometric autocorrelations of the two pulses are in fact distinct, as predicted [115], the differences are subtle. The conclusion is that interferometric autocorrelation data may be only weakly sensitive to rather significant changes in pulse shape and may therefore be quite challenging to apply for unambiguous pulse-shape retrieval in a practical experimental context.

Intensity Cross-Correlation   In addition to intensity autocorrelation measurements, one can also perform intensity cross-correlation measurements in which the two pulses being correlated have different intensity profiles. If we denote the intensity profiles of reference and signal pulses as $I_r(t)$ and $I_s(t)$, respectively, one measures

$$\langle P_{\text{SHG}}(\tau) \rangle \sim \langle I_s(t) I_r(t - \tau) \rangle \sim \int dt\, I_s(t) I_r(t - \tau) \tag{3.48}$$

where we have assumed a background-free noncollinear measurement geometry. When the reference pulse is much shorter than the signal pulse, the measurement returns $I_s(t)$ directly, including any asymmetry. To estimate the amount of broadening due to the finite duration of the reference pulse, we consider the case of Gaussian pulses [i.e., $a_r(t) \sim e^{-t^2/t_r^2}$ and $a_s(t) \sim e^{-t^2/t_s^2}$]. The intensity cross-correlation measurement then results in

$$\langle P_{\text{SHG}}(\tau) \rangle \sim \exp\left(-\frac{2\tau^2}{t_s^2 + t_r^2} \right) \tag{3.49}$$

The measured pulse width t_s' is proportional to

$$t_s' = \sqrt{t_s^2 + t_r^2} \tag{3.50}$$

If the reference pulse duration is 50% that of the signal pulse, the cross-correlation trace is broadened by only 11%. The broadening rapidly becomes very small as the reference pulse duration is further decreased relative to the signal pulse. Intensity cross-correlation measurements are particularly useful when the signal pulse is known to be broader than the reference, for example, in the case when the signal is generated from the reference pulse using a pulse-shaping apparatus (see Chapter 8). If, in addition, the intensity profile of the reference pulse is already known, $I_s(t)$ can be extracted from the cross-correlation data by deconvolution.

3.4.2   Experimental Procedures

Repetitive Pulse Measurements   In terms of experimental procedure, with CW mode-locked systems [113] one commonly uses a chopper to modulate one of the interferometer arms and a lock-in amplifier to detect the component of the photomultiplier (PMT) current modulated at that frequency. A stepper motor–driven delay stage is stepped under computer control (typically, with a 0.1- or 1-μm step size) in order to vary the delay, and the lock-in output is recorded as a function of delay. The delay range can be scanned many times to allow signal averaging for improved sensitivity. In this procedure the measurement is the average over a large number (millions or billions) of laser pulses. This is valid provided that all the pulses in the ensemble are essentially identical, which is usually the case for modern mode-locked sources. However, when the measurement averages over pulses with widely varying pulse durations, it has been shown that the experimental correlation trace assumes a double-sided exponential shape, independent of the actual pulse shape [117].

Single-Shot Measurements   In high-power amplified laser systems with low pulse repetition rates, one often prefers a different measurement setup, one allowing single-shot
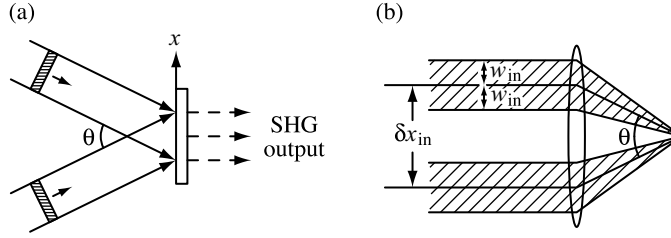
Figure 3.14   (a) Transverse variation of the delay in noncollinear intensity autocorrelation measurements; (b) focusing geometry for noncollinear intensity autocorrelation.

measurement capability. This is usually achieved by crossing the two beams (which typically are not focused) at a relatively large angle in the second-harmonic crystal [118]. For a sufficiently large beam size and crossing angle, this leads to an appreciable variation in delay across the transverse beam profile, where the delay is given by

$$\tau = \frac{2x \sin(\theta/2)}{c} \tag{3.51}$$

As illustrated in Fig. 3.14, x is the lateral spatial coordinate on the surface of the crystal and θ is the full angle between the incident beams (in air). The single-shot measurement is accomplished by recording the spatial profile of the generated second harmonic (e.g., using a CCD camera).

Transverse Delay Effects in Noncollinear Repetitive Pulse Measurements   This same transverse variation of the delay, which is desirable for single-shot measurements, can give rise to an undesirable temporal broadening in noncollinear autocorrelation measurements for high-repetition-rate laser systems, in which the entire second-harmonic beam is incident on a single photodetector. From eq. (3.51), the range of delay is clearly on the of order $\theta d/c$, where d is the beam diameter and θ has been assumed small. To make this more precise, we follow a procedure similar to [119] and write down an expression for the measured SHG power in the noncollinear experiment in the case of a Gaussian beam profile [i.e., $E(r, t) \sim a(t)e^{-r^2/w^2}$]. The measured second-harmonic power in the direction of the detector, integrated over both time and the spatial profile, takes the form

$$\langle P_{\text{SHG}} \rangle \sim \iint dx\, dt \, \left| a \left(t + \frac{x}{c} \sin(\theta/2) \right) \right|^2 \left| a \left(t - \tau - \frac{x}{c} \sin(\theta/2) \right) \right|^2$$
$$\times \, \exp \left(\frac{-4x^2 \cos^2 (\theta/2)}{w^2} \right) \tag{3.52}$$

These integrals can be evaluated analytically when the pulse shape is also Gaussian [i.e., $a(t) \sim e^{-t^2/t_p^2}$]. After some algebra one obtains the result:

$$\langle P_{\text{SHG}} \rangle \sim \exp \left(-\frac{\tau^2}{t_p^2 + w^2 \tan^2(\theta/2)/c^2} \right) \tag{3.53}$$

Thus, the measurement yields a pulse width proportional to t'_p, where

$$\left(t'_p\right)^2 = t_p^2 + (\delta T)^2 \qquad \text{with} \qquad \delta T = \frac{w \tan(\theta/2)}{c} \qquad (3.54)$$

To evaluate this expression, we consider Fig. 3.14b, which depicts the typical beam geometry used for focusing into the SHG crystal in noncollinear autocorrelations. The input pulses are assumed to have Gaussian beams with radii w_{in} and center-to-center separation δx_{in}. Assuming that the input beams are parallel prior to the lens (focal length f), the angle between the focused beams is given by $\tan(\theta/2) = \delta x_{\text{in}}/2f$. The beam radius at the focus of the lens is given by $w = f\lambda/\pi w_{\text{in}}$, which can be verified[4] by using the Gaussian beam formulas and $ABCD$ matrices summarized in Section 1.3.3. Inserting into eq. (3.54) gives

$$\delta T = \frac{\lambda}{2\pi c}\frac{\delta x_{\text{in}}}{w_{\text{in}}} \qquad (3.55)$$

Since $\delta x_{\text{in}} \geq 2w_{\text{in}}$ is required in order that the input beams are actually separated, $\delta T \geq \lambda/\pi c$. The same result is obtained if one requires instead that the beam-crossing angle at the crystal exceeds the far-field beam divergence angle (which can also be verified using the Gaussian beam formulas). The minimum value for the time-broadening parameter is the optical period divided by π, which gives about 0.9 fs for $\lambda = 0.8$ μm. Thus the time broadening inherent in noncollinear SHG is quite small, even for pulses in the 10-fs regime, provided that the crossing angle is kept near the minimum.

3.4.3   Correlation Measurements Using Two-Photon absorption

In addition to SHG, other nonlinear effects, such as two-photon fluorescence (TPF) and two-photon absorption (TPA), can also be used to measure the intensity autocorrelation. In TPA the measurement depends on an intensity-dependent absorption, so that there is increased absorption when the pulses are overlapped in time. The absorption coefficient α for a TPA process is given by

$$\alpha = \beta I(t) \qquad (3.56)$$

where β is the TPA coefficient in cm/W. It is assumed here that the TPA process occurs essentially instantaneously (or at least very fast compared to the pulse being measured), which is valid for electronic two-photon transitions with no intermediate single-photon resonance. The linear absorption is also assumed to be negligible. The absorbed energy ΔU can then be written

$$\Delta U \sim \int dt\, \alpha I(t) = \int dt\, \beta I^2(t) \qquad (3.57)$$

If one uses the field of a delayed pulse pair from eq. (3.34) to compute the intensity in this equation, one finds that the absorbed energy as a function of delay, $\Delta U(\tau)$, has the same

---

[4] This formula applies exactly when the input beams have their waists a distance f in front of the lens.

form as eq. (3.36). Therefore, measurement of $\Delta U(\tau)$ yields the same results as the SHG measurement.

In the TPF technique [120,121], which was introduced in the early days of ultrafast optics, the pulses are passed through a medium with fluorescence proportional to the square of the intensity (but ideally, with no fluorescence possible via single-photon excitation). The two pulses are usually arranged in a counterpropagating geometry, so that the relative delay varies along the sample thickness. Recording the fluorescence signal as a function of position yields an intensity autocorrelation measurement (with background). In this technique the time-integrated fluorescence provides a measure of the energy absorbed. TPF is no longer widely used in current ultrafast optics practice, although it can be useful for autocorrelation measurements at very short wavelengths where efficient second-harmonic crystals are not available.

TPA correlation measurements are usually performed in semiconductor materials [122–127], typically using photodetector structures. The material composition and optical wavelength are chosen such that the light is below bandgap, so that linear absorption is not allowed, but above half-bandgap, so that TPA is allowed. TPA leads to the generation of charge carriers which are collected with the aid of the electric field present in the photodetector structure. The resulting time-average photocurrent is a direct measure of the energy absorbed. TPA correlators have two attractive features compared to SHG: (1) for some wavelengths it is possible to use commercially available photodetectors, which can be very inexpensive compared to an SHG crystal and the associated photomultiplier; and (2) the TPA process alleviates some of the difficulties associated with measurement of very short pulses. With respect to the latter point, two-photon detectors intrinsically have a quite broad optical bandwidth, which can only be obtained in SHG using very thin crystals; furthermore, the electrically active region of conventional surface-normal TPA photodetector structures usually has a depth on the order of micrometers, so that pulse broadening due to dispersion within the nonlinear material is also excluded. Another interesting experimental implementation of a TPA correlator uses a two-photon optical waveguide photodetector as the nonlinear medium [123,127]. The waveguide structure keeps the light confined to a diameter of a few micrometers over a device length on the order of 1 mm, which gives high sensitivity, enabling measurement of very low intensity pulses (e.g., from semiconductor diode lasers with pulse durations on a time scale of tens of picoseconds.

3.4.4   Higher-Order Correlation Techniques

Finally, we mention briefly higher-order correlation measurements which depend on nonlinearities stronger than the SHG or TPA used for second-order correlations. Because these methods depend on higher powers of the electric field, they offer greater pulse shape sensitivity but at the cost of higher peak power requirements. The possibility of higher-order correlation measurements has been known for some time [113,128] and has seen some use for the characterization of pulses from high-power amplifier systems. It has also been demonstrated that third-order correlations can be performed with high-repetition-rate femtosecond oscillator systems using either third-harmonic generation (THG) [129] or three-photon absorption (3PA) [130]. Both sets of experiments were performed with 130-fs, 1.5 to 1.6-μm pulses from an optical parametric oscillator system with average powers from a few tens to a few hundreds of milliwatts at an 80-MHz repetition rate.

In the THG experiments, the pulses were focused by a high-numerical-aperture micro-scope objective onto the surface of a glass plate. Theory shows that in the case of tight focusing, efficient THG occurs only when the input beam is focused at an interface [131]. This means that the THG interaction is confined to within a few micrometers of the surface, which eliminates the requirement for phase matching, resulting in a broad optical bandwidth for the harmonic generation process. The time-integrated third-harmonic power is given by

$$\langle P_{\text{THG}}(t) \rangle \sim \int dt \left| \left[a(t) + \eta a(t - \tau) e^{-j\omega_0 \tau} \right]^3 \right|^2 \tag{3.58}$$

where we have introduced η to allow for the possibility that the time-delayed pulse pair have different amplitudes. Meshulach et al. [129] give the full functional form of the THG correlation signal and demonstrate that when $\eta = 1$, the fringe-averaged third-order au-tocorrelation gives a contrast ratio of $10:1$, while the interferometric (fringe-resolved) autocorrelation gives a $32:1$ contrast ratio.

The 3PA experiments used a photodiode with a bandgap near 680 nm, so that both single- and two-photon absorption were disallowed [130]. The expression for the three-photon absorption photocurrent is given by

$$\langle P_{\text{3PA}}(t) \rangle \sim \int dt \left\{ \left| a(t) + \eta a(t - \tau) e^{-j\omega_0 \tau} \right|^2 \right\}^3 \tag{3.59}$$

Since eqs. (3.58) and (3.59) are easily shown to be equal, we conclude that THG and 3PA measure the same third-order correlation function, just as SHG and TPA yield the same second-order correlation function. Experimentally, the 3PA measurements yielded the same fringe-averaged and fringe-resolved contrast ratios as for THG. Furthermore, it was also demonstrated in the 3PA experiments that by setting $\eta \ll 1$, the correlation measurements can reveal asymmetries in the intensity profile (when $\eta = 1$, the correlation trace remains symmetric).

3.5   CHIRPED PULSES AND MEASUREMENTS IN THE TIME–FREQUENCY DOMAIN

We now begin to discuss methods for determining completely the phase and intensity pro-files of ultrashort pulses. These methods are particularly important in the case of chirped or phase-modulated pulses, since intensity autocorrelation methods do not provide much chirp information. Before covering these pulse measurement methods, let us first review the char-acter of chirped pulses and introduce powerful time–frequency techniques for visualizing such pulses.

Consider a Gaussian pulse written in the time and frequency domains as follows:

$$e(t) = \text{Re} \left\{ a(t) e^{j\omega_0 t} \right\} \tag{3.60a}$$

with

$$a(t) = e^{-\Gamma t^2} \tag{3.60b}$$

and

$$A(\omega) = \left(\frac{\pi}{\Gamma}\right)^{1/2} e^{-\omega^2/4\Gamma} \tag{3.60c}$$

This represents a linearly chirped pulse if we choose a complex pulse-width parameter Γ. Let us write

$$\frac{1}{\Gamma} = \alpha + j\beta \tag{3.61a}$$

with

$$\Gamma = \frac{\alpha - j\beta}{\alpha^2 + \beta^2} = \Gamma_r + j\Gamma_i \tag{3.61b}$$

Here α and β are both real numbers; α affects the spectral amplitude, and β is a chirp parameter. The time-domain pulse envelope can now be rewritten

$$a(t) = e^{-\Gamma_r t^2} e^{-j\Gamma_i t^2} \tag{3.62}$$

The pulse duration is determined by Γ_r, while Γ_i is related to the chirp. For a fixed spectral width (fixed α), the minimum pulse width occurs for $\beta = 0$. The pulse width for arbitrary β is given by

$$\frac{\Delta t(\alpha, \beta)}{\Delta t(\alpha, \beta = 0)} = \sqrt{1 + \frac{\beta^2}{\alpha^2}} \tag{3.63}$$

Let us examine this chirped pulse in more detail. We can first calculate the frequency-dependent delay $\tau(\omega)$, given by

$$\tau(\omega) = \frac{-\partial\psi(\omega)}{\partial\omega} = \frac{\beta(\omega - \omega_0)}{2} \tag{3.64}$$

The spectral phase $\psi(\omega)$ and the frequency-dependent delay $\tau(\omega)$ are sketched in Fig. 3.15a for $\beta > 0$. $\psi(\omega)$ is a parabola and $\tau(\omega)$ is a straight line with positive slope. Since positive frequencies experience greater delays, this represents a positive linear chirp (or an up-chirp).

Similarly, we can calculate the instantaneous frequency $\omega_{\text{inst}}(t)$, given by

$$\omega_{\text{inst}}(t) = \omega_0 + \frac{\partial\phi(t)}{\partial t} = \omega_0 - 2\Gamma_i t$$

$$= \omega_0 + \frac{2\beta t}{\alpha^2 + \beta^2} \tag{3.65}$$

The temporal phase $\phi(t)$ and the instantaneous frequency are sketched in Fig. 3.15b, again assuming that $\beta > 0$. This plot is similar to Fig. 3.15a: $\phi(t)$ is a parabola and $\omega_{\text{inst}}(t)$ is again a straight line with positive slope. Larger times correspond to higher frequencies, so this still represents a positive chirp.
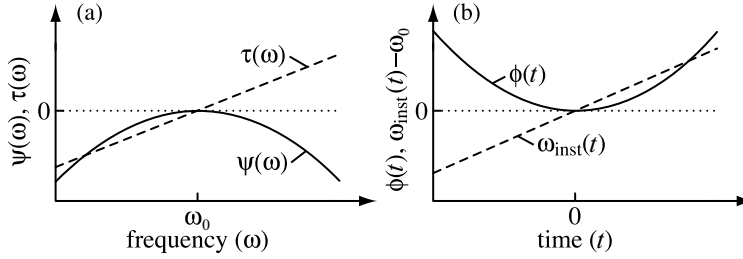
Figure 3.15 (a) Phase $\psi(\omega)$ (solid line) and frequency-dependent delay $\tau(\omega)$ (dashed line) vs. frequency for $\beta > 0$; (b) Phase $\phi(t)$ (solid line) and instantaneous frequency $\omega_{\text{inst}}(t)$ (dashed line) vs. time for $\beta > 0$. All units are arbitrary.

Such plots of frequency-dependent delay and instantaneous frequency give a good visual indication of the chirp, and they are useful not only for linear chirps as in this example but for nonlinear chirps as well. Note, however, that these plots give no direct indication of how close to or far from bandwidth limited the pulse is. It turns out that when $\beta \gg \alpha$, the pulses are highly chirped and far from the bandwidth limit. In this limit the pulse duration is proportional to the bandwidth:

$$\Delta t \approx \frac{1}{2} \beta \, \Delta\omega \tag{3.66}$$

Equation (3.66) clearly breaks down when the chirp is small ($\beta \leq \alpha$). The pulse width does not approach zero as $\beta \to 0$; rather, it approaches a limiting value proportional to $\alpha^{1/2}$. In the case of small chirp, the $\Delta\nu \, \Delta t$ product is close to the bandwidth limit (0.44 for a Gaussian). Plots of $\tau(\omega)$ and $\omega_{\text{inst}}(t)$ as in Fig. 3.15 are most useful for the case of relatively large chirp.

Let us consider one further example: a pulse with a cubic spectral phase corresponding to a nonlinear chirp. Here we assume that the spectrum is given by

$$A(\omega) = e^{-\omega^2/4\Gamma} e^{j\gamma\omega^3} \tag{3.67}$$

where γ represents the size of the cubic phase variation and Γ is now real. The frequency-dependent delay is

$$\tau(\omega) = -3\gamma(\omega - \omega_0)^2 \tag{3.68}$$

The spectral phase and the frequency-dependent delay are sketched in Fig. 3.16 for $\gamma < 0$ and Γ real. $\tau(\omega)$ is a parabola, and all the frequencies have a delay greater than or equal to zero. The highest frequencies have the highest delay and lead to an asymmetric tail (for $t > 0$) in the time-domain intensity profile. The central frequencies contribute to the main peak, which is somewhat broadened and slightly delayed. Figure 3.16 also shows the calculated intensity profile of a pulse with a significant amount of cubic spectral phase. The shift of the central peak and the long asymmetric tail are clearly evident, as predicted. In addition, rapid oscillation is observed. The reason for this oscillation can also be identified with the help of Fig. 3.16a. We see from the figure that for every time $t > 0$,
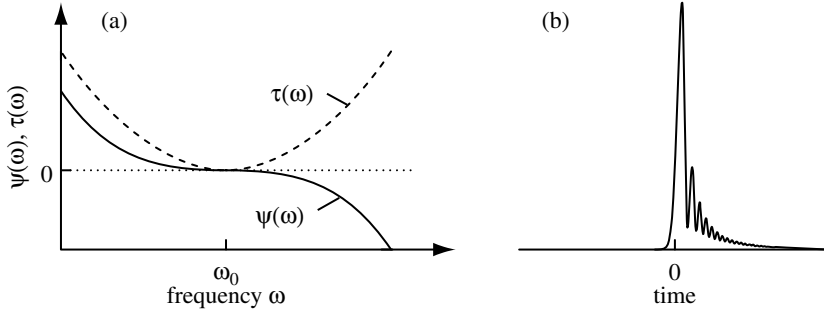
Figure 3.16 (a) Phase $\psi(\omega)$ (solid line) and frequency-dependent delay $\tau(\omega)$ (dashed line) vs. frequency for a cubic spectral phase with $\gamma < 0$; (b) calculated intensity profile. All units are arbitrary.

two different optical frequencies contribute to the signal. When these two frequencies are in phase, they add constructively to give a large field at the corresponding time instant, but when they are out of phase, they interfere destructively to give a weak field or zero. The phase difference is a cubic function of frequency, leading to the oscillations observed.

Let us now discuss how one might measure $\omega_{\text{inst}}(t)$ or $\tau(\omega)$. In the case of the instantaneous frequency, one could multiply the actual field by a suitable gating function $g(t - \tau)$ centered at $t = \tau$, resulting in a signal field given by

$$e_{\text{sig}}(t, \tau) = e(t)g(t - \tau) \tag{3.69}$$

One would then determine the instantaneous frequency by measuring the power spectrum of the time-gated pulse. This procedure, depicted in Fig. 3.17, yields a two-dimensional function $S_e(\omega, \tau)$, given by

$$S_e(\omega, \tau) = \left| \int dt \, e_{\text{sig}}(t, \tau)e^{-j\omega t} \right|^2 = \left| \int dt \, e(t)g(t - \tau)e^{-j\omega t} \right|^2 \tag{3.70}$$

$S_e(\omega, \tau)$ is known as the *spectrogram* of the electric field $e(t)$. For a constant τ, $S_e(\omega, \tau)$ yields a one-dimensional function of ω which provides information about the instantaneous frequency distribution at that τ. For a constant ω, $S_e(\omega, \tau)$ yields a one-dimensional function of τ which provides information about the times at which frequency ω appears.
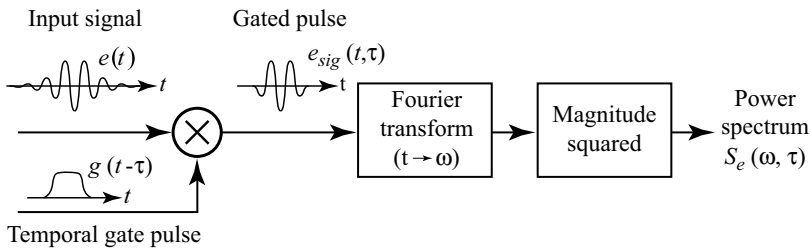
Figure 3.17 Spectrogram time–frequency distribution function.

Spectrograms are not unique to the field of ultrafast optics. They are used extensively to characterize acoustic signals (e.g., speech [132]). Musical scores, which specify a series of musical notes as a function of time, may also be regarded as a type of spectrogram. Spectrograms are an example of a joint time–frequency distribution, an analysis tool used to characterize signals whose spectral content is varying in time. Time–frequency distributions are themselves the subject of a rich and still evolving field, and many different types of time–frequency distributions have been investigated. For a review, see [133]. Here we discuss only the spectrogram and the closely related sonogram, which are the time–frequency distributions that have played the most important role in the measurement of ultrafast optical signals. We note, however, that the Wigner function has also seen some use for the description of ultrafast pulses in the time–frequency domain [134–136].

In Section 3.6 we discuss ultrafast optical methods for generating a gate function and measuring the spectrogram as well as methods for retrieving the actual pulses shape from the spectrogram data. For now we simply note the importance of choosing a proper gate function! If the $g(t)$ is too short, the gated function $e_{\text{sig}}(t, \tau)$ will have a spectrum much wider than the original spectrum. In this case any variations in the instantaneous frequency will be much smaller than the spectrum of $e_{\text{sig}}(t, \tau)$ and will therefore be difficult to detect. Similarly, if $g(t)$ is much longer than the input pulse, $e_{\text{sig}}(t, \tau)$ will be identical to $e(t)$, and the spectrogram will reveal only the original power spectrum. Therefore, one should generally select a gate function whose duration is on the order of the inverse bandwidth of the pulse to be characterized.

Instead of measuring $\omega_{\text{inst}}(t)$ as in eqs. (3.69) and (3.70), one can also imagine measuring $\tau(\omega)$. In this case one would first use a narrowband filter to select a specific frequency band. The frequency-dependent delay is determined by measuring the temporal intensity profile of the filtered waveform and then repeating as the passband of the narrowband filter is slid across the spectrum of the pulse. This procedure is sketched in Fig. 3.18. Mathematically, it turns out that such a measurement is equivalent to a spectrogram. This can be seen by substituting $e(t) = (1/2\pi) \int d\omega\, E(\omega)e^{j\omega t}$ and $g(t) = (1/2\pi) \int d\omega\, G(\omega)e^{j\omega t}$ into eq. (3.70), where $E(\omega)$ and $G(\omega)$ are the Fourier transforms of $e(t)$ and $g(t)$, respectively. The result is

$$\hat{S}_e(\omega, t) = \left| \frac{1}{2\pi} \int d\omega'\, E(\omega')H(\omega' - \omega)e^{j\omega' t} \right|^2 \tag{3.71}$$

which describes the temporal intensity profile of the original waveform after filtering by a sliding filter of the form $H(\omega) = G(-\omega)$. $\hat{S}_e(\omega, t)$ is sometimes called a *sonogram*.

The sonogram and the spectrogram of a given signal are identical provided that the sonogram filter function is equal to the frequency-reversed Fourier transform of the spectrogram
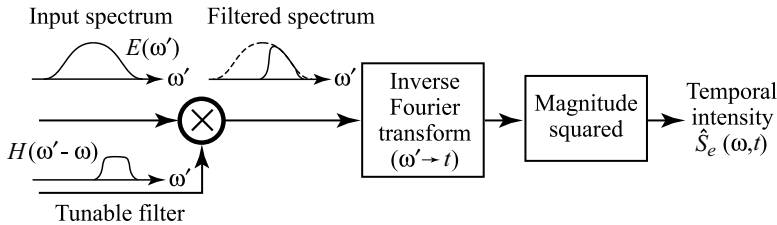
Figure 3.18   Sonogram time–frequency distribution function.

gate function. Furthermore, due to the Fourier transform relationship between the spectrogram gate function and the sonogram filter function, the rms widths of these functions in the time domain (Δt_{rms}) and in the frequency domain ($\Delta \omega_{rms}$) must satisfy the uncertainty relationship, eq. (3.11). This helps to quantify our discussion about selecting the duration of the gate function. High time resolution in the spectrogram can be obtained by using a narrow gate function, but unavoidably at the cost of spectral resolution.

We illustrate these concepts by discussing the form of the spectrogram for the case where the input pulse $s(t)$ and the gate pulse $g(t)$ are a chirped Gaussian centered at frequency ω_0 and a baseband chirped Gaussian, respectively, written as

$$s(t) = e^{-a_s t^2/2} e^{jb_s t^2/2} e^{j\omega_0 t} \tag{3.72}$$

$$g(t) = e^{-a_g t^2/2} e^{jb_g t^2/2} \tag{3.73}$$

$s(t)$ has the same form as the chirped Gaussian considered earlier in this section if we identify $\Gamma_r = a_s/2$ and $\Gamma_i = -b_s/2$ from eq. (3.61b). The resulting spectrogram of $s(t)$ can be written in two equivalent forms, as follows [133]:

$$S_s(\omega, \tau) \sim P_\tau(\tau) e^{-[\omega - \langle\omega\rangle_\tau]^2/2\sigma_\omega^2} \tag{3.74}$$

and

$$S_s(\omega, \tau) \sim P_\omega(\omega) e^{-[\tau - \langle\tau\rangle_\omega]^2/2\sigma_\tau^2} \tag{3.75}$$

Here $P_\tau(\tau)$ and $P_\omega(\omega)$ are known as the *marginal delay distribution* (delay marginal) and *marginal frequency distribution* (frequency marginal), obtained by integrating the spectrogram over either frequency or delay, respectively, and are given by

$$P_\tau(\tau) = \int d\omega \, S_s(\omega, \tau) \sim e^{-a_g a_s \tau^2 / (a_g + a_s)} \tag{3.76}$$

$$P_\omega(\omega) = \int d\tau \, S_s(\omega, \tau) \sim e^{-a_g a_s (\omega - \omega_0)^2 / \left[a_s \left(a_g^2 + b_g^2 \right) + a_g \left(a_s^2 + b_s^2 \right) \right]} \tag{3.77}$$

with

$$\langle\omega\rangle_\tau = \frac{1}{P_\tau(\tau)} \int d\omega \, \omega S_s(\omega, \tau) = \omega_0 + \frac{(a_g b_s - b_g a_s)\tau}{a_g + a_s} \tag{3.78}$$

$$\langle\tau\rangle_\omega = \frac{1}{P_\omega(\omega)} \int d\tau \, \tau S_s(\omega, \tau) = \frac{(a_g b_s - b_g a_s)(\omega - \omega_0)}{a_s \left(a_g^2 + b_g^2 \right) + a_g \left(a_s^2 + b_s^2 \right)} \tag{3.79}$$

$$\sigma_\omega^2 = \frac{(a_g + a_s)^2 + (b_g + b_s)^2}{2(a_g + a_s)} \tag{3.80}$$

$$\sigma_\tau^2 = \frac{(a_g + a_s)^2 + (b_g + b_s)^2}{2 \left[a_s \left(a_g^2 + b_g^2 \right) + a_g \left(a_s^2 + b_s^2 \right) \right]} \tag{3.81}$$

$\langle\omega\rangle_\tau$ and $\langle\tau\rangle_\omega$ are known as the *first conditional moments*, which provide an *estimate* of the instantaneous frequency and frequency-dependent delay, respectively.

Several points are worth discussing:

- Writing the time–frequency distribution in two different forms emphasizes the equivalence of the spectrogram and the sonogram. Equation (3.74) may be considered the spectrogram form, while eq. (3.75) may be considered the sonogram form.

- The spectrogram takes the form of an ellipse, with an inclination that depends on the chirp of both the input pulse and the gate pulse. For a real gate pulse ($b_g = 0$), the inclination depends only on the input pulse chirp, and unchirped input pulses have major and minor axes aligned with the τ and ω axes. Thus, an intuitive interpretation of the spectrogram is more easily accomplished in the case of a real gate pulse, although both chirped or unchirped gate pulses are permissible.

- The area of the ellipse remains finite, even when the duration of the input pulse or the gate pulse becomes extremely short. This is a manifestation of the uncertainty relationship, eq. (3.11).

- From eq. (3.74), for a fixed delay the distribution is centered at $\langle\omega\rangle_\tau$. For an accurate estimation of the instantaneous frequency, one should use a very short gate pulse. For an unchirped gate pulse with $a_g \gg a_s$, one obtains $\langle\omega\rangle_\tau = \omega_0 + b_s\tau$. This is identical to eq. (3.65) for ω_{inst} after we substitute in $b_s = -2\Gamma_i$. Furthermore, $P_\tau(\tau)$ reduces exactly to the intensity profile $|s(t)|^2$ in this case. However, at the same time the frequency spread parameter σ_ω^2 becomes large.

- Conversely, from eq. (3.75), for a fixed frequency the distribution is centered at $\langle\tau\rangle_\omega$. In order that $\langle\tau\rangle_\omega$ give an accurate estimate of the frequency-dependent delay, one should use a very long gate pulse with a narrow spectrum. For an unchirped gate pulse with $a_g \ll a_s$, one obtains $\langle\tau\rangle_\omega = b_s(\omega - \omega_0)/(a_s^2 + b_s^2)$, which reduces to eq. (3.64) for $\tau(\omega)$ after we substitute in for a_s and b_s in terms of Γ_r and Γ_i. At the same time the temporal spread parameter σ_τ^2 becomes large, while $P_\omega(\omega)$ reduces exactly to the power spectrum.

- The need for a short gate pulse to portray temporal information (e.g., ω_{inst}) accurately and a long gate pulse to portray spectral information [e.g., $\tau(\omega)$] accurately is another manifestation of the uncertainty relation, which dictates an inverse relation between the temporal resolution and the spectral resolution of the information displayed by a spectrogram.

Pulse measurement techniques based on spectrograms and sonograms are discussed in the following two sections, respectively. For simplicity of understanding, in ultrafast optics we generally use the term *spectrogram* to refer to measurements involving multiplication by a relatively short time gate and reserve the term *sonogram* to refer to measurements involving the use of a series of relatively narrow filters. However, from a mathematical point of view, this distinction is not necessary.

3.6   FREQUENCY-RESOLVED OPTICAL GATING

In this section we discuss methods for measuring spectrograms in ultrafast optics and for retrieving complete amplitude and phase information about the electric field from the

spectrogram. The techniques we discuss are called *frequency-resolved optical gating*, also known by the acronym FROG [137–140]. The main experimental requirement in measuring ultrafast optical spectrograms is in implementing an ultrafast gate function. In FROG, as in other ultrashort-pulse measurement techniques, the pulse is used to gate itself via a nonlinear optical interaction. A number of nonlinear interactions, including polarization gating, self-diffraction, transient gratings, and second- and third-harmonic generation, have been demonstrated for FROG measurements. In the following we first describe three of the more important FROG geometries and discuss general features of the FROG spectrograms. We then discuss the procedure for recovering a complete electric field profile from the FROG spectrograms.

3.6.1   Polarization-Gating FROG

One common FROG geometry, called *polarization gating FROG* (PG-FROG) [138,139], is shown in Fig. 3.19. Ultrafast gating is achieved through the optical Kerr effect, in which the phase modulation and nonlinear birefringence induced by the gating pulse are used to rotate the probe pulse polarization. The probe pulse is linearly polarized and is passed through a pair of crossed polarizers. In the absence of the gating pulse, the probe pulse is rejected by the second polarizer and does not pass through to the rest of the measurement apparatus. The gate pulse, which has its polarization oriented at an angle of 45° with respect to the probe, intersects the probe pulse in a thin nonlinear medium with an ultrafast third-order optical nonlinearity (the optical Kerr effect). We discussed previously nonlinear refractive index changes induced by an intense ultrashort pulse. Here we note that the optical Kerr effect also results in a nonlinear birefringence, since the nonlinear index change seen by a probe pulse with polarization parallel to that of a linearly polarized gate pulse is different than that seen by a probe pulse with polarization perpendicular to that of the gate pulse. For an electronic Kerr effect, the nonlinear birefringence is proportional to the instantaneous intensity of the gate pulse. This time-dependent birefringence leads to a time-dependent polarization rotation of the probe pulse, which is therefore transmitted in part through the second polarizer. The power spectrum of the probe pulse transmitted is measured using a spectrometer. By recording the spectrogram as a function of delay τ between gate and probe pulses, one obtains the spectrogram.
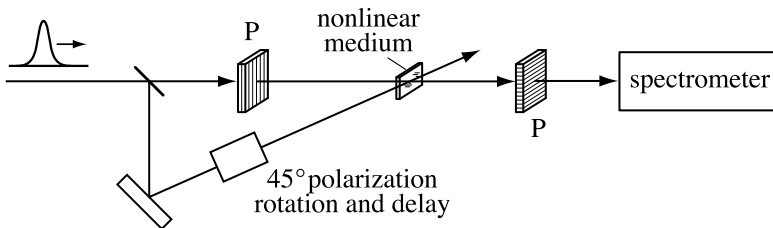
Figure 3.19   Experimental geometry for FROG using polarization gating. Usually, this measurement is performed in a single-shot geometry, where the delay varies across the nonlinear medium. The output of the spectrometer is captured using a CCD camera, which displays a two-dimensional delay–frequency plot. P, polarizer.

We can analyze the nonlinear polarization rotation by writing

$$n_\| = n_0 + \Delta n_\|(t) = n_0 + n_{2,\|} I_{\text{gate}}(t) \tag{3.82a}$$

and

$$n_\perp = n_0 + \Delta n_\perp(t) = n_0 + n_{2,\perp} I_{\text{gate}}(t) \tag{3.82b}$$

Here $n_\|$ and $n_\perp$ are the index of refraction seen by light polarized parallel or perpendicular to the linearly polarized gate pulse, respectively, and $I_{\text{gate}}(t)$ is the gate pulse intensity. For the electronic Kerr effect in an isotropic medium, $n_{2,\perp} = \frac{1}{3} n_{2,\|}$ [37]. We assume that the input probe and gate pulses can be written

$$\mathbf{e_{probe}}(t) = E_{\text{pr}} e(t) \hat{\mathbf{x}}$$

$$\mathbf{e_{gate}}(t) = \frac{1}{\sqrt{2}} E_g e(t - \tau)(\hat{\mathbf{x}} + \hat{\mathbf{y}}) \tag{3.83}$$

E_{pr} and E_g are the amplitudes of the probe and gate pulses, respectively, and $e(t)$ is the electric field temporal profile (assumed the same for both pulses). The probe pulse after the nonlinear medium is written

$$\mathbf{e_{probe}}(t) = \frac{1}{\sqrt{2}} E_{\text{pr}} e(t) \left\{ \frac{\hat{\mathbf{x}} + \hat{\mathbf{y}}}{\sqrt{2}} e^{-j(\omega/c)\Delta n_\|(t)L} + \frac{\hat{\mathbf{x}} - \hat{\mathbf{y}}}{\sqrt{2}} e^{-j(\omega/c)\Delta n_\perp(t)L} \right\} \tag{3.84}$$

where the linear phase shift corresponding to passage through the nonlinear medium of length L has been neglected. After some algebra, eq. (3.84) can be rewritten

$$\mathbf{e_{probe}}(t) = E_{\text{pr}} e(t) e^{-j(\omega/2c)\left(\Delta n_\| + \Delta n_\perp\right)L}$$

$$\times \left\{ \hat{\mathbf{x}} \cos\left[\frac{\omega}{2c}\left(\Delta n_\| - \Delta n_\perp\right)L\right] - j\hat{\mathbf{y}} \sin\left[\frac{\omega}{2c}(\Delta n_\| - \Delta n_\perp)L\right] \right\} \tag{3.85}$$

Finally, after passage through a polarizer passing $\hat{\mathbf{y}}$-polarized light, the gated probe field is

$$\mathbf{e_{probe}}(t) = -jE_{\text{pr}} e(t) \hat{\mathbf{y}} e^{-j(\omega/2c)\left(\Delta n_\| + \Delta n_\perp\right)L} \sin\left[\frac{\omega}{2c}(\Delta n_\| - \Delta n_\perp)L\right] \tag{3.86}$$

The probe transmitted is gated according to the time-dependent birefringence $\Delta n_\|(t) - \Delta n_\perp(t)$ and also experiences a phase modulation induced by the gate pulse. The appearance of the phase modulation together with the amplitude modulation desired is related to the appearance of both self-amplitude-modulation and self-phase-modulation terms in our treatment of solid-state laser mode-locking in Section 2.4. If we assume that the induced birefringence in eq. (3.86) remains small, we can expand the sine to first order and neglect the SPM term, which will also be small. Substituting in for the nonlinear birefringence, we obtain

$$\mathbf{e_{probe}}(t) \sim \hat{\mathbf{y}} E_{\text{pr}} e(t) \left|e(t - \tau)\right|^2 \tag{3.87}$$

This gives us the gated field $e_{\text{sig}}(t, \tau)$ used in calculating the spectrogram. Thus, we make the identification

$$e_{\text{sig}}(t, \tau) \sim e(t) |e(t - \tau)|^2 \qquad (3.88)$$

The apparatus measures the power spectrum, resulting in a spectrogram that we call $I_{\text{FROG}}(\omega, \tau)$, written as

$$I_{\text{FROG}}(\omega, \tau) = \left| \int dt\, e_{\text{sig}}(t, \tau) e^{-j\omega t} \right|^2 = \left| \int dt\, e(t) |e(t - \tau)|^2 e^{-j\omega t} \right|^2 \qquad (3.89)$$

It is important to remember that we assumed small birefringence in our derivation. This results in the specific functional form of eq. (3.89) and also justifies neglect of the gate pulse–induced phase modulation term, which would otherwise modify the probe pulse phase profile. Of course, this means that the efficiency of the Kerr gating process must be low. We have also assumed that the probe pulse intensity is low enough that there is negligible self-phase modulation and that the nonlinear medium is thin enough that we can ignore dispersion.

From the experimental point of view, we note that PG-FROG is used mainly with amplified femtosecond laser systems with sufficient intensity to perform single-shot measurements. For single-shot measurements the probe and gate pulses, which are at most weakly focused, cross at a relatively large angle in the nonlinear material. This leads to a variation in the relative delay across the beam cross section, as given by eq. (3.51). The spectrometer disperses the frequency spectrum in a direction perpendicular to the beam-crossing direction, resulting in a two-dimensional image (one axis for frequency, one axis for delay) which is recorded using a CCD camera. In principle, PG-FROG can also be used with CW mode-locked lasers, but due to the lower peak powers provided by such sources, PG-FROG signal levels are usually too low for practical measurements.

PG-FROG traces calculated for unchirped and linearly chirped Gaussian pulses are shown in Fig. 3.20. Following our earlier discussion of conventional spectrograms with independent gate functions, the FROG traces are ellipses rather than lines, reflecting the minimum time–bandwidth product property of the Fourier transform. The tilt (or lack of tilt) of the plots is an obvious visual indication of the magnitude and sign of the chirp. Another example of a FROG trace, this time calculated for a pulse with cubic spectral phase, is also shown in Fig. 3.20. The shape of the trace reveals a roughly parabolic delay vs. frequency dependence, as expected. Thus, even though FROG is a time-domain measurement technique, it is able to provide information such as $\tau(\omega)$ which is more intuitively connected with sonogram measurements. This illustrates once again the duality between spectrograms and sonograms.

In each of these examples, the FROG traces give a qualitative visual indication of the type of chirp present. Note, however, that the instantaneous frequency or frequency-dependent delay cannot be taken directly from the FROG trace (without first retrieving the electric field profile itself). As a simple illustration, consider the PG-FROG distribution for the case of a linearly chirped Gaussian, according to eq. (3.72). Since the gate pulse is $|e(t)|^2$, the form of the FROG trace is then given by eqs. (3.74) to (3.81) with $a_g = 2a_s$ and $b_g = 0$. From eq. (3.78) we obtain $\langle\omega\rangle_\tau = \omega_0 + \frac{2}{3}b_s\tau$. Comparing our earlier discussion on spectrograms, we find that $\langle\omega\rangle_\tau - \omega_0 = \frac{2}{3}(\omega_{\text{inst}} - \omega_0)$. The slope of the FROG trace is only two-thirds of the
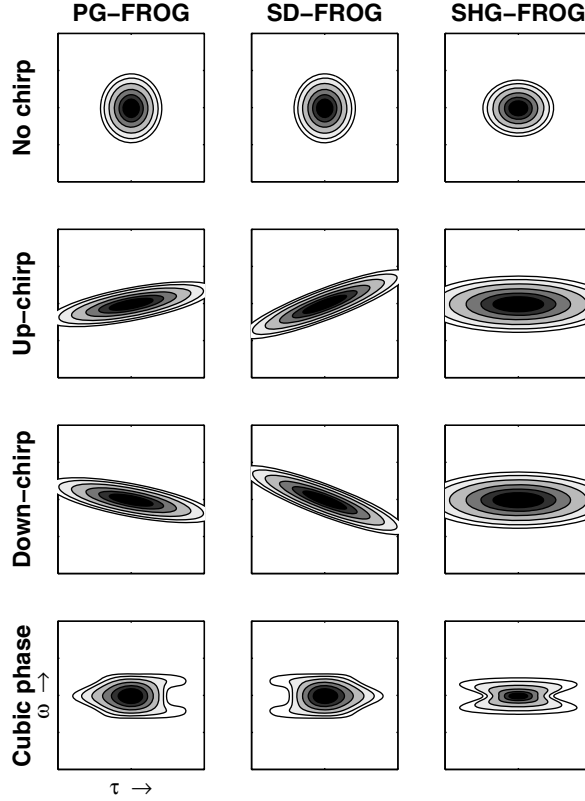
Figure 3.20 Calculated spectrogram traces for PG-FROG (left column), SD-FROG (middle), and SHG-FROG (right). The FROG traces are shown as shaded contour plots, with darker regions indicating higher intensity. All traces correspond to identical Gaussian power spectra, but with different spectral phase profiles. Row one, bandwidth-limited; row two, linear up-chirp; row three, linear down-chirp; row four, cubic spectral phase ($\gamma < 0$ case).

slope in a corresponding $\omega_{\mathrm{inst}}(t)$ plot. The reason for this can readily be understood: The quantity $e(t)\,|e(t-\tau)|^2$ is centered at $t = 2\tau/3$, and therefore the PG-FROG spectrogram with delay τ samples the frequency content around $2\tau/3$, which accounts for the difference between $\langle\omega\rangle_\tau$ and ω_{inst} [139]. Nevertheless, since the gate function is strictly real, PG-FROG still provides relatively intuitive traces.

3.6.2 Self-Diffraction FROG

Another common FROG geometry for amplified pulse measurement is based on self-diffraction in a thin nonlinear refractive index medium [139,141]. The geometry is sketched in Fig. 3.21. Two identical and usually co-polarized pulses with a relative time delay τ intersect at an angle in the nonlinear material. The interference between the two pulses induces a transient refractive index grating varying in the transverse direction of the form

$$\Delta n_{\mathrm{grating}} = n_{2,\parallel}\left[a(t)a^*(t-\tau)e^{-j(\mathbf{k_1}-\mathbf{k_2})\cdot\mathbf{r}} + \mathrm{c.c.}\right] \tag{3.90}$$
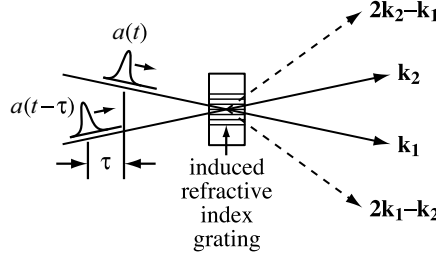
Figure 3.21   Geometry used in self-diffraction experiments.

where $\mathbf{k_1}$ and $\mathbf{k_2}$ are as shown in the figure and $a(t)$ is the complex amplitude function. Each of the pulses can diffract off this grating, giving light in the new directions $2\mathbf{k_1} - \mathbf{k_2}$ and $2\mathbf{k_2} - \mathbf{k_1}$. This provides the self-diffraction FROG (SD-FROG) signal fields

$$e_{\text{sig}(1)}(t, \tau) \sim a^2(t)a^*(t - \tau)e^{j[\omega_0 t - (2\mathbf{k_1} - \mathbf{k_2}) \cdot \mathbf{r}]} \tag{3.91a}$$

and

$$e_{\text{sig}(2)}(t, \tau) \sim a^2(t - \tau)a^*(t)e^{j[\omega_0 t - (2\mathbf{k_2} - \mathbf{k_1}) \cdot \mathbf{r}]} \tag{3.91b}$$

where low diffraction efficiency is assumed. The FROG trace is obtained by measuring the power spectrum of the signal field as a function of τ. The FROG trace for the $2\mathbf{k_1} - \mathbf{k_2}$ direction is written

$$I_{\text{FROG}}(\omega, \tau) = \left| \int dt \, e_{\text{sig}(1)}(t, \tau)e^{-j\omega t} \right|^2 = \left| \int dt \, a^2(t)a^*(t - \tau)e^{-j\omega t} \right|^2 \tag{3.92}$$

The FROG trace using direction $2\mathbf{k_2} - \mathbf{k_1}$ is easily shown to have the same form with the substitution $\tau \to -\tau$ and therefore carries no additional information.

From a practical point of view, one advantage of SD-FROG is that no polarizer is needed, which may be helpful for measurements at very short wavelengths where high-quality polarizers are difficult to realize. On the other hand, self-diffraction is not phase matched, which means that the angle between the two beams and the thickness of the nonlinear medium must be kept small. This limits the sensitivity of SD-FROG and in principle can introduce an undesirable wavelength dependence to the generation process. However, these difficulties can be avoided with appropriate experimental care.

Calculated SD-FROG traces for unchirped and linearly chirped pulses (Fig. 3.20) show behavior similar to those for PG-FROG. The trace for a bandwidth-limited pulse is an untilted ellipse. The chirped pulses are tilted ellipses, where the sign of the tilt indicates the sign of the chirp. The magnitude of the tilt is different than for PG-FROG. The form of the SD-FROG trace for a complex Gaussian pulse, eq. (3.72), is given by eqs. (3.74) to (3.81), provided that we make the following substitutions: $a_s \to 2a_s$; $b_s \to 2b_s$; $a_g \to a_s$; $b_g \to -b_s$. As a result, we find that $\langle \omega \rangle_\tau - \omega_0 = \frac{4}{3}b_s\tau = \frac{4}{3}(\omega_{\text{inst}} - \omega_0)$. This can be understood by noting that the FROG signal field peaks at $\tau/3$, and therefore the contribution to $\langle \omega \rangle_\tau - \omega_0$ from the three factors involved in $e_{\text{sig}(1)}(t, \tau)$ is $2\omega_{\text{inst}}(\tau/3) - \omega_{\text{inst}}(-2\tau/3) = 4b_s\tau/3$. Thus, the slope of SD-FROG traces for linearly chirped pulses is twice that of PG-FROG.

Note that in SD-FROG the gate function is not purely real; it also carries phase information about the pulse. As a result, the phase information in the signal and gate functions can interact, leading to spectrogram traces that are in many cases more difficult to interpret than for PG-FROG. One example is the case of a spectral cubic phase. As shown in Fig 3.20, the trace is still qualitatively consistent with a parabolic $\tau(\omega)$; however, the direction of the trace with respect to positive τ is reversed compared to the PG-FROG case and compared to the actual frequency-dependent delay.

3.6.3   Second-Harmonic-Generation FROG

PG-FROG and SD-FROG are difficult to apply with CW mode-locked sources, since the combination of low peak powers together with a third-order optical nonlinearity leads to polarization gating or self-diffraction efficiencies which are too low for convenient operation. By using second-harmonic generation (SHG), which is a second-order nonlinearity, as the gating mechanism, one can obtain sufficient FROG signal levels with CW mode-locked systems. The geometry for SHG-FROG is similar to the noncollinear SHG intensity autocorrelation setup shown in Fig. 3.9, except that a spectrometer is placed in front of the detector. The cost of this approach is that SHG-FROG traces are less amenable than PG-FROG traces to simple qualitative interpretation. The SHG-FROG approach has been described, for example, in [138,142,143].

In SHG-FROG the gated field and the FROG trace are given, respectively, by

$$e_{\text{sig}}(t, \tau) \sim e(t)\,e(t - \tau) \tag{3.93}$$

and

$$I_{\text{FROG}}(\omega, \tau) = \left| \int dt\, e(t)e(t - \tau)e^{-j\omega t} \right|^2 \tag{3.94}$$

By making the substitution $t' = t - \tau$, one can easily show that

$$I_{\text{FROG}}(\omega, \tau) = I_{\text{FROG}}(\omega, -\tau) \tag{3.95}$$

That is, the SHG-FROG trace is a symmetric function of τ. This is not surprising, since the pulse and the gating function are identical, and therefore one cannot tell whether the pulse is ahead of the gating function or the gating function is ahead of the pulse. In contrast, the gating function in PG-FROG, $|e(t - \tau)|^2$, is distinguishable from the pulse itself, and therefore FROG traces that are not symmetric in delay are possible. A further consequence of eqs. (3.93) to (3.95) is that SHG-FROG has a time-reversal ambiguity not present in PG-FROG. In particular, the fields $e(t)$ and $e(-t)$, corresponding to complex amplitude functions $a(t)$ and $a^*(-t)$, respectively, have identical SHG-FROG traces. To determine the pulse shape uniquely, one must perform additional measurements. One possibility that can be used to resolve the ambiguity is to first measure the FROG trace of the initial pulse and then measure the FROG trace of the pulse after transmission through a known dispersive medium.

One other nontrivial ambiguity has been identified for SHG-FROG. In the case of two isolated pulses, relative phases between the pulses of ϕ and $\phi + \pi$ lead to identical $I_{\text{FROG}}(\omega, \tau)$. Therefore, these two cases cannot be distinguished on the basis of the FROG trace.

The third column of Fig. 3.20 shows calculated SHG-FROG traces for transform-limited and linearly chirped Gaussian pulses. The trace for the transform-limited pulse is a horizontal ellipse and is qualitatively similar to FROG traces shown earlier. The SHG-FROG traces for the linearly chirped pulses, however, show qualitative differences compared to the earlier results. As the chirp is increased, the trace grows wider in the horizontal (delay) direction, but the major and minor axes of the ellipse remain in the horizontal and vertical directions. This is a consequence of the symmetry of the FROG trace with respect to τ. Identical traces are obtained for equal-magnitude positive and negative linear chirps; this further illustrates one of the disadvantages of SHG-FROG: the inability to determine the sign of a chirp. Additionally, one clearly sees that the area contained under the SHG-FROG trace increases substantially for the chirped pulse. This is a consequence of the increased time–bandwidth product of the chirped pulse.

The formula for the SHG-FROG trace for a linearly chirped Gaussian pulse is given by eqs. (3.74) to (3.81), with the assignments $a_g \rightarrow a_s$ and $b_g \rightarrow b_s$, resulting in the following particularly simple form:

$$I_{\text{FROG}}(\omega, \tau) \sim e^{-a_s \tau^2/2} \, e^{-\omega^2/2\sigma_\omega^2} \qquad (3.96)$$

where $\sigma_\omega^2 = a_s + b_s^2/a_s$. These expressions conform to our discussion above: The trace retains its untilted orientation independent of the chirp parameter (b_s), while the width of the trace in the frequency direction increases with increasing frequency modulation.

An example of a more complicated SHG-FROG trace, corresponding to a pulse with cubic spectral phase, is also shown in the figure. Compared to corresponding PG- and SD-FROG plots, the symmetrization of the SHG-FROG trace as a function of τ is clearly evident.

It is interesting to consider the marginal frequency distribution $P_\omega(\omega)$ of the SHG-FROG trace. It turns out that $P_\omega(\omega)$ is given by

$$P_\omega(\omega) = \int d\tau \, I_{\text{FROG}}(\omega, \tau) = \frac{1}{2\pi} \int d\omega' \, \left| E(\omega') \right|^2 \left| E(\omega - \omega') \right|^2 \qquad (3.97)$$

This can be derived [132] by replacing $I_{\text{FROG}}(\omega, \tau)$ in the above eq. (3.97) with the sonogram form of the SHG-FROG trace, eq. (3.71), where $g(t)$ is set equal to $e(t)$. $P_\omega(\omega)$ is given by the autoconvolution of the optical power spectrum at the fundamental wavelength. This offers a useful consistency check for measurement data: The SHG-FROG trace can be compared to the independently measured input power spectrum using eq. (3.97). Since the FROG data and input power spectrum are acquired in different wavelength ranges and often using different equipment, it is likely that any data sets corrupted through systematic measurement errors will fail to satisfy the frequency marginal. Therefore, this procedure can be used to test the reliability of the FROG data.

3.6.4   Frequency-Resolved Optical Gating Using Temporal Phase Modulation

In principle, FROG can also use a phase-only gate function. In one early approach the spectra of a pulse before and after *self-phase modulation* (SPM) in a nonlinear refractive index material were measured and analyzed. Reconstruction of electric field profiles was attempted using only a single spectrum after SPM, but with limited success due to the

one-dimensional nature of this phase-retrieval problem [144]. Later, temporal modulation on the input signal was imposed using *cross-phase modulation* (XPM) from a pump pulse with a variable time delay. Analysis of the resulting two-dimensional data set, consisting of XPM spectra corresponding to different delays, yielded more consistent success in electric field reconstruction [145]. FROG has also been demonstrated using SPM and XPM of a delayed pulse pair in optical fibers; like SHG-FROG, this technique offers the possibility of using relatively low pulse energies [146]. At present, such pulse measurements based on temporal phase modulation are not used extensively in ultrafast optics; therefore, they are not considered further.

3.6.5   Signal Recovery from FROG Traces

Let us now discuss how to retrieve electric field pulse shapes from FROG traces [138,140,143]. Note that if $e_{sig}(t, \tau)$ is known, $e(t)$ can be obtained by direct integration:

$$\int d\tau \, e_{sig}(t, \tau) = \int d\tau \, e(t)g(t - \tau) \sim e(t) \tag{3.98}$$

where the proportionality constant is not important. We therefore ask the question: is it possible to determine $e_{sig}(t, \tau)$?

To answer this, we express the FROG trace in terms of $\tilde{E}_{sig}(t, \Omega)$, the Fourier transform of $e_{sig}(t, \tau)$ with respect to τ:
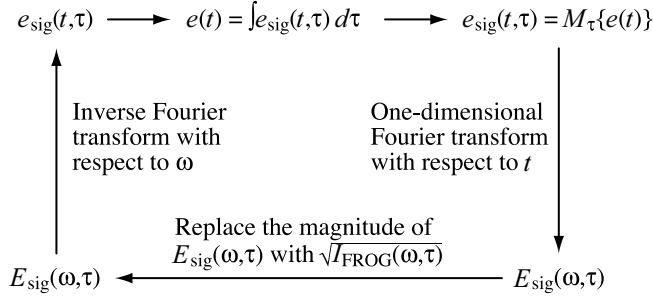
$$I_{FROG}(\omega, \tau) \sim \left| \iint dt \, d\Omega \, \tilde{E}_{sig}(t, \Omega)e^{-j\omega t}e^{j\Omega\tau} \right|^2 \tag{3.99}$$

where

$$\tilde{E}_{sig}(t, \Omega) = \int d\tau \, e_{sig}(t, \tau)e^{-j\Omega\tau} \tag{3.100}$$

Equation (3.99) represents a phase-retrieval problem: We wish to recover $\tilde{E}_{sig}(t, \Omega)$ [and hence $e_{sig}(t, \tau)$] from its two-dimensional power spectrum. We have already seen that in one dimension this problem does not have a unique solution; that is why we cannot determine $e(t)$ from its power spectrum $|E(\omega)|^2$. It turns out, however, that in the two-dimensional phase-retrieval problem [i.e., eq. (3.99)], one can almost always recover a unique solution.[5] This fact has been the focus of considerable activity in the image science community since the early 1980s [147–150] and is the basis for pulse retrieval in FROG. We can rationalize why unique phase retrieval is possible by noting that what we actually wish to recover is the one-dimensional complex function $e(t)$. If we represent $e(t)$ by a complex vector of length N, the FROG data can be represented by an $N \times N$ two-dimensional array with N^2

---

[5] Existing proofs of such uniqueness depend on the assumption that the signal to be retrieved, here e_{sig}, has compact support (i.e., it is nonzero only within a bounded region of space). Although in a strict sense this is generally not true in ultrafast optics, in a practical sense $e(t)$ and $e_{sig}(t, \tau)$ generally decay rapidly and disappear into the noise. Also, the mathematical form constraint provides extra information not available in the usual phase-retrieval problem, which helps enforce uniqueness. Experience strongly suggests that except for certain known ambiguities, FROG returns unique solutions.

$$e_{\rm sig}(t,\tau) \longrightarrow e(t) = \int e_{\rm sig}(t,\tau)\,d\tau \longrightarrow e_{\rm sig}(t,\tau) = M_\tau\{e(t)\}$$

Inverse Fourier
transform with
respect to ω

One-dimensional
Fourier transform
with respect to t

Replace the magnitude of
$E_{\rm sig}(\omega,\tau)$ with $\sqrt{I_{\rm FROG}(\omega,\tau)}$

$$E_{\rm sig}(\omega,\tau) \longleftarrow \qquad\qquad E_{\rm sig}(\omega,\tau)$$

Figure 3.22 Iterative Fourier transform algorithm for retrieving the ultrashort-pulse amplitude and phase from FROG data. $M_\tau\{e(t)\}$ stands for the mathematical form constraint [e.g., for PG-FROG, $M_\tau\{e(t)\} = e(t)\,|e(t-\tau)|^2$]. Adapted from [138].

data values. Since $e(t)$ has only $2N$ degrees of freedom (real and imaginary parts for each element of the $e(t)$ vector), there is considerable redundancy in the FROG data, and this makes phase retrieval possible.

Trebino and co-workers have published several algorithms for recovering the electric field profile from FROG data [138]. Their original iterative Fourier transform algorithm is indicated in Fig. 3.22. One starts at the top center with a guess for $e(t)$. Moving in a clockwise direction, the next step is to use this guess for $e(t)$ to generate $e_{\rm sig}(t,\tau)$ using

$$e_{\rm sig}(t,\tau) = M_\tau\{e(t)\} \qquad (3.101)$$

where $M_\tau\{\cdots\}$ represents the mathematical form of the nonlinear process being utilized [e.g., eq. (3.88), (3.91a), or (3.93) for PG-, SD-, or SHG-FROG, respectively] [139]. This is called a *mathematical form constraint*. By performing a series of one-dimensional Fourier transforms with respect to t, one obtains $E_{\rm sig}(\omega,\tau)$. At this point one applies the experimental data (the intensity constraint) by replacing the magnitude of $E_{\rm sig}(\omega,\tau)$ with the magnitude (square root) of the FROG data while leaving its phase unchanged; that is,

$$E_{\rm sig}(\omega,\tau) \rightarrow \frac{E_{\rm sig}(\omega,\tau)}{\left|E_{\rm sig}(\omega,\tau)\right|}\sqrt{I_{\rm FROG}(\omega,\tau)} \qquad (3.102)$$

The algorithm continues by inverse Fourier transforming with respect to ω to get a new estimate of $e_{\rm sig}(t,\tau)$ and then integrating with respect to τ to obtain an updated estimate of $e(t)$. This completes one iteration. The algorithm continues by using the updated version of $e(t)$ as the input for the mathematical form constraint on the second iteration; and so on. Additional iterations are performed until the solution converges. This algorithm is related to iterative Fourier transform algorithms used previously in image science [148,151], with the extra feature of the mathematical form constraint.

Let $e^{(k)}(t)$ and $E_{\rm sig}^{(k)}(\omega,\tau)$ denote the computed arrays representing $e(t)$ and $E_{\rm sig}(\omega,\tau)$ during iteration number k. One obvious way to check for convergence is to compare these results to the corresponding results from the previous iteration. Another way is to compute the error between the computed and experimental values. Since $e(t)$ is not known directly from the experiments, one cannot compute the error in $e^{(k)}(t)$. However, one can use $E_{\rm sig}^{(k)}(\omega,\tau)$ for comparison with actual FROG data. In particular, it is useful to compute the rms FROG

error, defined by

$$\varepsilon_{\text{FROG}}^{(k)} = \left\{ \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left[I_{\text{FROG}}^{(k)}(\omega_i, \tau_i) - I_{\text{FROG}}(\omega_i, \tau_i) \right]^2 \right\}^{1/2} \tag{3.103}$$

where

$$I_{\text{FROG}}^{(k)}(\omega, \tau) = \left| E_{\text{sig}}^{(k)}(\omega, \tau) \right|^2 \tag{3.104}$$

The algorithm should be run until the FROG error cannot be made any smaller. Note that $\varepsilon_{\text{FROG}}$ will not generally reach zero, due to noise on the experimental trace.[6] Note also that because of the redundancy built into the FROG trace, in general there is no $e(t)$ that corresponds precisely to the noisy experimental FROG data. As long as the noise in the input data is not excessive, the final computer value for $E_{\text{sig}}^{(k)}(\omega, \tau)$ may be a better estimate of the actual FROG trace than the experimental data [152].

It is important to distinguish between convergence and uniqueness [148]. Convergence means that the algorithm reduces the FROG error to a very small value, ideally limited only by experimental or computer noise, whichever is larger. Upon convergence the electric field waveform retrieved is consistent with the FROG trace. Uniqueness means that the waveform retrieved is the only waveform consistent with the FROG trace. Due to the uniqueness of the two-dimensional phase-retrieval problem, convergence of the FROG algorithm implies that a unique electric field solution has been obtained (except for some known ambiguities, as noted below). However, it is important to remember that for other problems, such as one-dimensional phase-retrieval, convergence need not imply uniqueness. Therefore, as new pulse measurement techniques are developed, results from such techniques should be taken with caution until their uniqueness is established.

In addition to the iterative Fourier transform algorithm, other more powerful FROG algorithms which tend to give more robust convergence have also been developed. One of the most widely used of these is called *generalized projections* [140,143]. The idea is pictured in Fig.3.23, in which the upper and lower elliptical regions represent the sets of all possible $e_{\text{sig}}(t, \tau)$ functions that satisfy the mathematical form constraint and the intensity constraint, respectively. According to the uniqueness of the two-dimensional phase retrieval problem, these regions intersect at a single point, which is the solution for $e_{\text{sig}}(t, \tau)$. The algorithm starts with an initial guess and then projects onto (moves to the closest point on) one of the constraint sets. At this point the algorithm iteratively projects from one constraint set onto the other until the solution is reached. Operationally, the algorithm has a flowchart that looks very much like Fig. 3.22, with one exception: At the top of the flowchart, the new estimate for $e(t)$ in each iteration is obtained by projecting onto the mathematical form constraint. This is achieved at the $(k + 1)$th iteration by finding the new field $e^{(k+1)}(t)$ which minimizes the difference (Z) between $M_\tau\{e^{(k+1)}(t)\}$ and the most recent (kth) estimate of

---

[6] The FROG error must obviously be computed prior to the magnitude update of eq. (3.102).
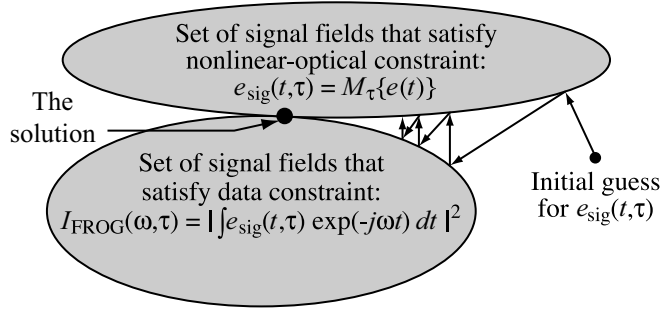
Figure 3.23 Generalized projections algorithm for retrieving the ultrashort-pulse amplitude and phase from FROG data. Adapted from [140], with permission. Copyright ©1997, American Institute of Physics.

$e_{\mathrm{sig}}(t, \tau)$; that is,

$$Z = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| e_{\mathrm{sig}}^{(k)}(t_i, \tau_j) - M_{\tau_j}\{e^{(k+1)}(t_i)\} \right|^2 \tag{3.105}$$

In practice, minimization of Z has been performed approximately, usually by using a simple gradient-based approach [140]. More recently, generalized projections has been implemented by using an outer product matrix method [153]. For more details the reader is directed to the references.

There are a number of practical issues that must keep track of in implementing FROG pulse recovery algorithms [152]. Some of these are listed below.

- The original guess for $e(t)$ is a matter of choice and can range from an intelligent guess based perhaps on the power spectrum and any additional information available to random noise. Using noise as the starting point is often regarded as a favorable choice.
- It is useful to normalize the experimental FROG data to a peak magnitude of 1, and to do the same for the computer value of $|E_{\mathrm{sig}}^{(k)}(\omega, \tau)|^2$ during each iteration. This facilitates computation of the rms FROG error $\varepsilon_{\mathrm{FROG}}$. As a slight improvement, which makes a difference in the case of real data with noise present, one could instead adjust the normalization of $|E_{\mathrm{sig}}^{(k)}(\omega, \tau)|^2$ to minimize the FROG error.
- Although as stated, the shape of the electric field $e(t)$ is essentially uniquely determined from the FROG trace, it is important to note that there are some simple ambiguities: The measurement is not sensitive to the time delay of the pulse or to a constant multiplicative phase factor. In the computer algorithm this means that the time delay or the phase of the pulse may wander as subsequent iterations are performed. In addition, it has already been pointed out that SHG-FROG has some additional ambiguities: a time-reversal ambiguity and a twofold ambiguity in the relative phases of a pair of well-separated pulses.
- It is important to avoid truncation of the FROG trace in either the ω or τ dimensions. This means that the spectral range should be sufficiently broad and the delay scan

sufficiently long that the FROG trace disappears into the noise. The matrix containing the FROG data should contain the valid data completely surrounded by zeros (or noise).

- As noted in Section 1.5.2, one should recall in performing Fourier transforms that FFT algorithms typically have the convention that $t = 0\,(\omega = 0)$ corresponds to the leftmost (first) point in the data arrays. If the actual field or spectrum is centered in the array, the FFT interprets this as a time (or frequency) shift of half a period, resulting in a $(-1)^n$ phase factor in the Fourier transform. One must be aware of such FFT idiosyncracies in implementing pulse retrieval algorithms.

It is worth repeating that the redundancy inherent in the FROG trace substantially overdetermines the electric field $e(t)$. As a result, when the phase retrieval algorithm converges, it is virtually guaranteed that the correct $e(t)$ has been found (except for known ambiguities). Conversely, for the more robust FROG algorithms, lack of convergence can be taken as an indicator of measurement error (excessive noise, incorrect calibration, truncation, etc.) [152]. The marginal distributions can also serve as a check in the consistency of the FROG data. The utility of comparing the SHG-FROG frequency marginal with a curve independently derived from the input power spectrum has already been discussed. Similar marginals exist for other forms of FROG [139]. An important point is that many one-dimensional pulse measurement schemes, such as intensity autocorrelation, lack the redundancy or internal consistency checks of FROG. Therefore, the experimentalist must exercise greater care with such one-dimensional techniques to ensure that any important sources of error are eliminated.

Pulse-shape recovery using FROG has been explored most extensively with relatively simple pulse shapes, such as those emitted from femtosecond modelocked lasers. Although FROG has also demonstrated success for some more highly structured pulses, such as those generated using femtosecond pulse shapers or resulting from nonlinear propagation in fibers, there is also the following alternative [154]. FROG can be used, for example, to completely characterize the relatively simple pulses from the mode-locked laser which provides the input to the pulse shaping apparatus. These characterized pulses can then be used as the reference for measurement of the shaped pulses via spectral interferometry or electric field cross-correlation.

3.7 PULSE MEASUREMENTS BASED ON FREQUENCY FILTERING

We have already seen that the sonogram of a short pulse could be obtained by first filtering the pulse to select a certain frequency band and then measuring the temporal intensity profile of the filtered pulse as the filter is set to different center frequencies. We have also seen that sonograms and spectrograms are equivalent mathematically. Therefore, it should be possible to determine ultrashort-pulse electric field profiles from sonogram measurements, just as FROG retrieves complete pulse information from spectrogram measurements. One important difference is that tunable filters can be constructed independent of the pulse to be measured, while in FROG the gate function is always derived from the pulse. This leads to the new possibility of determining the spectral phase without the need for iterative calculations.

A block diagram depiction of the measurement approach is shown in Fig. 3.24. The input pulse first passes through a filter characterized by its frequency response $H(\omega - \Omega)$,
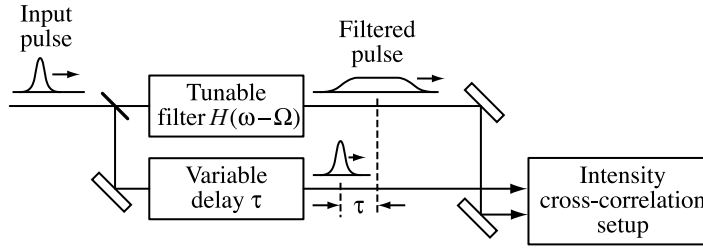
Figure 3.24   Sonogram approach to ultrashort-pulse measurement.

where the filter center frequency Ω is tunable. Often, the filter is implemented using a femtosecond pulse-shaping apparatus [155] constructed from a pair of diffraction gratings and lenses. The first grating and lens serve to separate different optical frequency components, as in a spectrometer, and the second grating and lens reassemble all the optical frequencies into a single collimated beam. This operation is performed in such a way that the output pulse shape is identical to that of the input, without distortion or dispersion. Let us use $\hat{\mathbf{x}}$ to denote the direction along which the optical frequencies are spread. A mask with a transparency function spatially patterned along $\hat{\mathbf{x}}$ is placed at the plane where the frequencies are separated. To a good approximation this yields an optical filter function which is simply a scaled version of the spatial mask pattern. The output pulse from the filter is broadened and reshaped in a manner related to the Fourier transform of the spatial masking pattern. Measurement of the output intensity profile, usually performed via intensity cross-correlation, comprises the final step in the measurement scheme. Either an unfiltered version of the input pulse, or any other sufficiently short pulse that may be available, may be used as the reference for the cross-correlation. Note that femtosecond pulse shaping is discussed in detail in Chapter 8. For the pulse measurement application it provides a means of conveniently tuning the shape and center frequency of the filter function.

3.7.1   Single-Slit Approaches

Single-Slit Sonogram Approach   Let us first consider the case where the spatial mask is a simple slit, so that $H(\omega)$ is approximately a rectangle function with spectral width $\delta\omega$ and constant spectral phase. $\delta\omega$ is determined by the width of the slit, and the center frequency Ω can be adjusted by translating the slit along $\hat{\mathbf{x}}$. Let us also assume that the following three approximations are valid:

1. The power spectrum of the input pulse remains essentially constant over a frequency range of width $\delta\omega$.
2. The spectral phase $\psi(\omega)$ of the input pulse varies sufficiently slowly over a frequency $\delta\omega$ that it may be expanded to first order.
3. The duration of the filtered pulse, which is proportional to $\delta\omega^{-1}$, is sufficiently broad that the intensity cross-correlation yields a good estimate of the filtered intensity profile.

These conditions can all be satisfied by choosing a sufficiently narrow filter width. We find then that measurement of the output temporal profile provides direct information about $\psi(\omega)$ [156,157]. This is demonstrated by using these approximations to simplify the expression for the filtered intensity profile, eq. (3.71), which gives

$$\hat{S}_e(\Omega, t) \approx |E(\Omega)|^2 \left| h \left[t + \psi'(\Omega) \right] \right|^2 \qquad (3.106)$$

where $h(t)$ is the inverse Fourier transform of $H(\omega)$ and $\psi'(\Omega) = \partial\psi(\omega)/\partial\omega$ evaluated at $\omega = \Omega$. The output intensity profile is shifted in time by $-\psi'(\Omega)$, which is exactly equal to the frequency-dependent delay $\tau(\omega)$! Thus, $\tau(\omega)$ is obtained directly by recording the shift as a function of Ω, which can either be read directly from the data, or determined by calculating the first temporal moment:

$$\tau(\omega = \Omega) = \frac{\int dt \, t \, \hat{S}_e(\Omega, t)}{\int dt \, \hat{S}_e(\Omega, t)} \qquad (3.107)$$

where $\hat{S}_e(\Omega, t)$ refers to the actual data. Integrating $\tau(\omega)$ gives $\psi(\omega)$.

This result is independent of the shape of the filter function, provided that $\delta\omega$ is narrow enough to satisfy our assumptions. The case of a Gaussian filter function with a linearly chirped Gaussian input pulse was worked out exactly in Section 3.5, where it was shown that the first conditional moment of the delay, $\langle\tau\rangle_\omega$, approaches $\tau(\omega)$ in the limit of a narrow filter function. However, as was pointed out at the same time, it is undesirable to make the filter much narrower than necessary; the reduced filter throughput and increased output pulse duration both lower the signal-to-noise ratio of the measurement.

This method has the advantage that the spectral phase information desired is obtained directly without the need for iterative phase retrieval. A disadvantage is that the accuracy of the phase information obtained depends on the degree to which the assumptions enumerated above are satisfied and is difficult to assess quantitatively. On the other hand, it is also possible to use the full information inherent in the sonogram, eq. (3.106), together with phase-retrieval algorithms to recover $e(t)$ [158]. This is essentially a sonogram version of FROG. One new feature which this provides is the possibility of substituting a two-photon absorption photodetector in place of the second-harmonic crystal in the intensity correlator, which is not possible in SHG-FROG.

Spectrally and Temporally Resolved Upconversion Technique   An alternative measurement approach is depicted in Fig. 3.25. Compared to the previous technique, this spectrally and temporally resolved upconversion technique (STRUT) [159,160] uses a fixed filter which we denote $F(\omega - \Omega)$, where Ω is fixed. Since tunability is not required, an interference filter can be utilized instead of a pulse shaper, if desired. The broadened filtered pulse is controlled with a variable delay and then interacts with a replica of the input pulse in a second harmonic crystal to produce an upconverted signal, which is spectrally resolved using a spectrometer. This procedure produces a two-dimensional data set as a function of delay τ and upconverted frequency ω_{up}. The measurement can be performed in either a repetitively scanned or a single-shot geometry. The arrangement is similar to SHG-FROG, except that the gate is now given by a filtered version of the input pulse. As we shall see, due to this difference, the measurement yields direct information on $\psi(\omega)$ without the need for iterative phase retrieval.
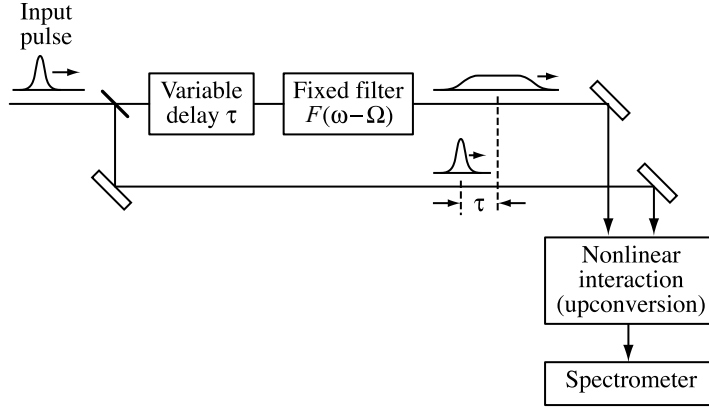
Figure 3.25 Spectrally and temporally resolved upconversion technique (STRUT) for ultrashort-pulse measurement.

We now analyze STRUT assuming that the filter bandwidth is sufficiently narrow that approximations 1 and 2 from earlier in this section still hold. Condition 3 is not needed, since the measurement does not require knowledge of the actual intensity profile of the filtered pulse. The quantity measured by the apparatus is given by

$$I_{\text{STRUT}}(\omega_{\text{up}}, \tau) = \left| \int dt \, e(t) g(t - \tau) e^{-j\omega_{\text{up}}t} \right|^2 \tag{3.108}$$

where $g(t)$ is the filtered version of the pulse. Using the equality between eqs. (3.70) and (3.71), we can rewrite this as

$$I_{\text{STRUT}}(\omega_{\text{up}}, \tau) = \left| \frac{1}{2\pi} \int d\omega' \, E(\omega') E(\omega_{\text{up}} - \omega') F(\omega_{\text{up}} - \omega' - \Omega) e^{j\omega'\tau} \right|^2 \tag{3.109}$$

where we have used $G(\omega) = E(\omega)F(\omega - \Omega)$. This can be simplified if we introduce a new variable $\delta\omega'$ such that $\omega' = \omega_{\text{up}} - \Omega + \delta\omega'$ and expand the spectral phase $\psi(\omega)$ to first order in $\delta\omega'$ (the power spectra are written only to zero order in $\delta\omega'$). After some algebra, this gives the result

$$I_{\text{STRUT}}(\omega_{\text{up}}, \tau) \approx \left| E(\Omega) E(\omega_{\text{up}} - \Omega) \right|^2 \left| f \left[\psi'(\Omega) - \psi'(\omega_{\text{up}} - \Omega) - \tau \right] \right|^2 \tag{3.110}$$

where $f(t)$ is the inverse Fourier transform of $F(\omega)$. The curve shifts along the delay coordinate by an amount given by $\psi'(\omega_{\text{up}} - \Omega)$. Measuring the temporal shift with the spectrometer set for upconverted frequency ω_{up} gives $\tau(\omega)$ for frequency $\omega_{\text{up}} - \Omega$ within the spectrum of the original pulse. The $\psi'(\Omega)$ term is a constant that does not influence the measurement. The fact that $f(\cdots)$ appears in a time-reversed form in eq. (3.110) arises simply because of the sign convention chosen for τ in our analysis and is not consequential.

As described earlier, STRUT has the advantage that the spectral phase information desired is obtained directly, at least approximately, with the concomitant disadvantage that the accuracy of the phase information obtained is not directly known. Of course, it should be

possible to run phase retrieval algorithms on the full STRUT data set, but this would offset the main benefit of this technique.

3.7.2   Double-Slit Approach

Let us now consider another technique based on the scheme of Fig. 3.24, but with the single-slit filter of Section 3.7.1 replaced by a double slit. This technique was first used for measurement of picosecond pulses in optical pulse compression experiments [161] and then developed more fully for femtosecond pulse characterization, where it was termed *direct optical spectral phase measurement* (DOSPM) [162,163]. The frequency filter is now written $H(\omega - \omega_1) + H(\omega - \omega_2)$. We assume that the following approximations are valid:

1. Both the power spectrum and the spectral phase of the input pulse remain essentially constant within the bandwidth $\delta\omega$ of a single slit.
2. The reference pulse is sufficiently well behaved that it can measure the intensity profile of the filtered pulse with good accuracy.

Under these conditions, the measured output intensity contains a beat note of the form

$$I_{\text{DOSPM}} \sim |h(t)|^2 \cos^2 \left(\frac{\Delta\omega t + \psi(\omega_2) - \psi(\omega_1)}{2} \right) \qquad (3.111)$$

where $\Delta\omega = \omega_2 - \omega_1$. The phase of the $\cos^2$ function gives directly the phase difference between the two optical frequencies selected. For example, if the relative phase is zero, an intensity peak occurs at $t = 0$, while for a relative phase of π, there is an intensity minimum at $t = 0$.

There are several strategies for exploiting this temporal interference phenomenon to determine the full spectral phase profile $\psi(\omega)$:

1. The measurement is repeated several times, first for frequencies ω_1 and $\omega_1 + \Delta\omega$, then for frequencies $\omega_1 + \Delta\omega$ and $\omega_1 + 2\Delta\omega$, and so on. Taking $\psi(\omega_1)$ as an arbitrary constant, the first measurement yields $\psi(\omega_1 + \Delta\omega) - \psi(\omega_1)$, the second measurement gives $\psi(\omega_1 + 2\Delta\omega) - \psi(\omega_1 + \Delta\omega)$, and so on. The overall spectral phase function is obtained by adding (or integrating over) the series of phase differences thus obtained. This procedure is a proper sonogram method, since the two-slit frequency filter is slid across the spectrum.
2. Instead of keeping the slit separation fixed, one can instead keep the position of one slit fixed while the second slit is slid across the spectrum. Thus, ω_1 is held constant while $\Delta\omega$ and hence ω_2 are varied. Each measurement yields $\psi(\omega_2)$ directly for the corresponding value of ω_2. This method is no longer a proper sonogram, since the variation of the two-slit filter function from one trace to the next is no longer a simple frequency translation.
3. Instead of a two-slit filter, a $2N$-slit filter can be utilized for simultaneous measurement of $2N - 1$ independent spectral phase values [163]. For example, $2N - 1$ of the slits can be arranged with a constant frequency separation $\Delta\omega$ [i.e., at frequencies $\omega_0 - (N-1)\Delta\omega, \omega_0 - (N-2)\Delta\omega, \ldots, \omega_0 + (N-2)\Delta\omega,$

$\omega_0 + (N-1)\Delta\omega$]. The last slit is deliberately offset (e.g., at $\omega_0 + \Delta\omega/3$). The filtered pulse has a complicated intensity profile in which $N(2N-1)$ individual interference terms are superimposed. However, with the slit positions given above, the terms corresponding to the interference between the single offset slit with each of the other $2N-1$ regularly spaced slits all occur at unique beat frequencies. Hence, each of these terms can be separated by Fourier analysis of the measured intensity profile, which yields the individual spectral phases.

It is interesting to note that although this DOSPM technique is clearly related to sonogram concepts, unlike the single-slit approaches it also makes obvious use of interference phenomena. This brings us to a discussion of interference-based pulse characterization techniques.

3.8   SELF-REFERENCING INTERFEROMETRY

In Section 3.3 we discussed temporal and spectral interferometry techniques in which the complete electric field could be determined via interference with a *completely characterized* reference pulse. Here we discuss two self-referencing interferometric methods in which the need for a fully characterized reference pulse is removed. Like the spectral filtering techniques discussed in Section 3.7, these interferometry-based methods provide a direct measurement of the phase profile without the need for iterative inversion and therefore permit rapid reconstruction of the electric field.

3.8.1   Time-Domain Interferometry of Chirped Pulses

We first describe a time-domain interferometry method applicable to highly chirped pulses when a much shorter (but not fully characterized) reference pulse is available [164]. This technique was developed for characterization of chirped pulses that had been spectrally broadened via self-phase modulation in an optical fiber and subsequently compressed in the context of fiber and grating pulse compression experiments (see Chapter 6). In this context the compressed pulses serve as the short reference pulses needed for measurement of the longer chirped pulses at the fiber output.

The experimental setup is shown schematically in Fig. 3.26. The chirped input pulse is split and sent into the arms of an interferometer, one of which contains a narrowband filter $H(\omega - \omega_0)$ of bandwidth $\delta\omega$ centered at frequency ω_0. The light from each of the two arms is recombined at the output of the interferometer. $\delta\omega$ should be sufficiently small that the spectral phase is essentially constant within the filter bandwidth and the filtered pulse is at least as long as the input pulse; the relative delay of the interferometer should also be set so that the filtered pulse and the input pulse have good temporal overlap. Physically, the output intensity profile exhibits a temporal interference pattern reflecting the time-varying phase difference between the essentially monochromatic (constant phase) filtered pulse and the chirped input pulse. This temporal interference pattern is recorded by means of an intensity cross-correlation using a short reference pulse. In our subsequent analysis we assume that the reference pulse is sufficiently short that the cross-correlation provides a good estimate of the actual intensity profile.
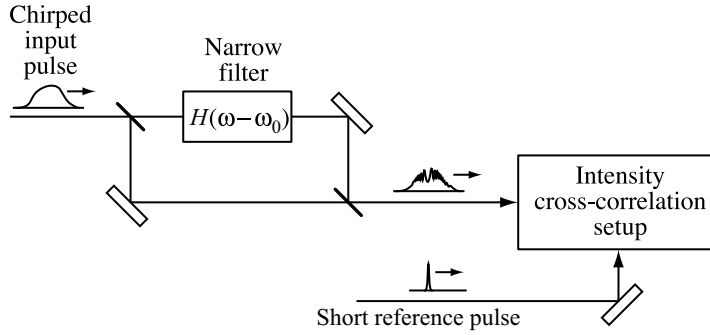
Figure 3.26   Time-domain interferometry of chirped pulses.

Mathematically, the output intensity profile can be written

$$I_{out}(t) \sim \left| |a(t)|e^{j[\omega t + \phi_{in}(t)]} + \alpha|h(t)|e^{j[\omega_0 t + \phi_h(t) + \phi_r]} \right|^2 \tag{3.112}$$

where $\phi_h(t)$ is the temporal phase associated with the impulse response function of the filter, ϕ_r is a constant reference phase which can be controlled by fine tuning the interferometer, and α (assumed real) accounts for the different amplitudes of the two terms. We now assume that $\phi_h(t)$ is a constant, which for simplicity we set to zero. This is valid provided that $H(\omega) = H^*(-\omega)$ in addition to our earlier assumption that $\delta\omega$ is sufficiently narrow. The resulting intensity profile is given by

$$I_{out}(t) \sim |a(t)|^2 + \alpha^2 |h(t)|^2 + 2 |a(t)h(t)| \cos [\phi_{in}(t) - \phi_r] \tag{3.113}$$

The desired information is extracted by subtracting $|a(t)|^2$ and $\alpha^2 |h(t)|^2$; these terms can be measured by performing cross-correlations from one arm of the interferometer at a time, with the other arm blocked. Let us call the result I'_{out}. The results for ϕ_r values of 0 and $\pi/2$ are given, respectively, as follows:

$$I'_{out}(t; \phi_r = 0) \sim 2\alpha |a(t)h(t)| \cos [\phi_{in}(t)] \tag{3.114a}$$

$$I'_{out}(t; \phi_r = \frac{\pi}{2}) \sim 2\alpha |a(t)h(t)| \sin [\phi_{in}(t)] \tag{3.114b}$$

The temporal phase $\phi_{in}(t)$ is obtained by dividing eq. (3.114b) by eq. (3.114a) and taking the arctangent of the result.

Experimental demonstrations of this technique used an etalon to implement $H(\omega)$. An interference filter or a pulse shaper could clearly also be used. It is interesting that in earlier experiments [165], passage of a single beam through a resonant atomic medium was used instead of the interferometer and filter shown in Fig. 3.26. The temporal reshaping caused by a single sharp absorption resonance can be described as arising from the temporal interference between the original input pulse and the nearly monochromatic free induction decay signal emitted by the atomic medium subsequent to its excitation. Thus, these two techniques are very similar. However, the earlier experiments based on atomic media suffered the disadvantages that ϕ_r could not conveniently be varied and that the temporal

interference began only partway through the pulse when the input frequency swept through the absorption line.

3.8.2 Self-Referencing Spectral Interferometry

We now discuss a self-referencing spectral interferometry technique based on the interference between two spectrally and temporally shifted but otherwise identical versions of the pulse to be measured. Although this approach has much in common with the spectral interferometry method discussed in Section 3.3, the introduction of a spectral shift (also called a *spectral shear*) allows the pulse shape to be determined without the need for a separate characterized reference pulse. This technique is called *spectral phase interferometry for direct electric field reconstruction* (SPIDER) [166,167].

We write the two pulses that interfere in the SPIDER apparatus as

$$e_1(t) = \text{Re}\{a(t)e^{j\omega_0 t}\} \tag{3.115a}$$

and

$$e_2(t) = \text{Re}\{a(t-\tau)e^{j(\omega_0+\Omega)(t-\tau)}\} \tag{3.115b}$$

We discuss later how the spectral shear by Ω is imposed. In terms of the envelope functions, the output power spectrum can be written

$$|A_{\text{out}}(\tilde{\omega})|^2 \sim |A(\tilde{\omega})|^2 + |A(\tilde{\omega}-\Omega)|^2$$
$$+ \left[A(\tilde{\omega})A^*(\tilde{\omega}-\Omega)e^{j(\omega_0+\tilde{\omega})\tau} + \text{c.c.}\right] \tag{3.116}$$

We are using the symbol $\tilde{\omega} = \omega - \omega_0$ since $A(\tilde{\omega})$ is a baseband function. If we now write $A(\tilde{\omega}) = |A(\tilde{\omega})|\, e^{j\psi(\tilde{\omega})}$, eq. (3.116) becomes

$$|A_{\text{out}}(\tilde{\omega})|^2 \sim |A(\tilde{\omega})|^2 + |A(\tilde{\omega}-\Omega)|^2$$
$$+ |A(\tilde{\omega})|\, |A(\tilde{\omega}-\Omega)| \cos\left[(\omega_0+\tilde{\omega})\tau + \psi(\tilde{\omega}) - \psi(\tilde{\omega}-\Omega)\right] \tag{3.117}$$

For a flat spectral phase the overall power spectrum exhibits oscillations with period $2\pi/\tau$. The period is modulated by the spectral phase difference $\psi_{\text{diff}}(\tilde{\omega}, \Omega) = \psi(\tilde{\omega}) - \psi(\tilde{\omega}-\Omega)$. Provided that the power spectrum is measured with sufficient resolution, this allows determination of the spectral phase.

The algorithm for reconstructing the spectral phase is similar to spectral interferometry. The SPIDER data is first inverse Fourier transformed into the time domain, which leads to terms located around $t = 0$ and $t = \pm\tau$ (refer back to Fig. 3.5). The term at $t = -\tau$ is extracted by setting the terms at $t = 0$ and $t = \tau$ to zero. The result is then Fourier transformed back into the frequency domain. The phase in the frequency domain is equal to $(\omega_0+\tilde{\omega})\tau + \psi_{\text{diff}}(\tilde{\omega}, \Omega)$. Subtracting the term linear in τ gives $\psi_{\text{diff}}(\tilde{\omega}, \Omega)$. After assigning a single arbitrary phase to one frequency, say ω_0, the measured phase differences then specify the spectral phase at frequencies $\omega_0 \pm m\Omega$, where m is a positive integer.

In order that this procedure provide an accurate reconstruction of $\psi(\omega)$ and hence $e(t)$, two conditions must be satisfied:

1. The delay τ must be sufficiently large that the terms at $t = 0$ and $t = \pm\tau$ can be clearly separated in the Fourier-transformed SPIDER data. This implies that the $2\pi/\tau$ period of the fringes in the power spectrum is fast compared to the variation of $\psi_{\text{diff}}(\tilde{\omega}, \Omega)$ and of the power spectrum $|A(\tilde{\omega})|^2$. This also implies that the duration of the pulse to be measured, t_p, satisfies $t_p \ll \tau$.

2. The spectral shear Ω must be sufficiently small that it adequately samples the power spectrum and $\psi_{\text{diff}}(\tilde{\omega}, \Omega)$. The latter condition can be restated as $|\psi_{\text{diff}}(\tilde{\omega}, \Omega)| \ll 1$. This implies that $\Omega \ll 2\pi/t_p$. In the unusual case of a pulse known to have finite support (i.e., the intensity is strictly zero outside a region $-T/2 < t < T/2$), this condition would be relaxed to $\Omega \le 2\pi/T$, which is the familiar sampling theorem of electrical engineering [168]. On the other hand, one should not make Ω so small that $\psi_{\text{diff}}(\tilde{\omega}, \Omega)$ becomes too small to measure with good signal-to-noise ratio.

A schematic SPIDER apparatus is shown in Fig. 3.27. The pulse to be measured is split into two beams. The first is transformed into a pulse doublet with delay τ, which can be performed by passing through either an etalon or a Michelson interferometer. The relative phase between the two pulses in the pulse pair must remain constant during the course of the measurement. The second beam passes through a pulse stretcher, which broadens the pulse and gives it a chirp. To lowest order, pulse stretchers operate by imposing a quadratic spectral phase onto the pulse; for a description of a common pulse stretcher based on a pair of diffraction gratings, see Chapter 4. The pulse pair and stretched pulse then interact in a second harmonic crystal. Each pulse in the pulse pair interacts with a different frequency from within the chirped stretched pulse. This results in a pair of upconverted pulses, still separated by τ but with the required spectral shift. Practically the experiment is usually performed using a phase-matching geometry where the second harmonic is generated only by the interaction between the pulse pair and the stretched pulse, not by either the stretched pulse (or the pulse pair) interacting with itself alone. Finally, the upconverted power spectrum is measured using a spectrometer.

In the case of a Gaussian pulse, the action of the pulse stretcher is described by the formalism of Section 3.5. If we write the spectrum after stretching as $A(\omega) \sim \exp\left\{-(\alpha + j\beta)\omega^2/4\right\}$, the instantaneous frequency is given by eq. (3.65), which reduces to $\tilde{\omega} \approx 2t/\beta$ in the limit $|\beta| \gg \alpha$. This is the limit where the stretched pulse is much longer than the inverse of the bandwidth of the input pulse. The spectral shear is related to the
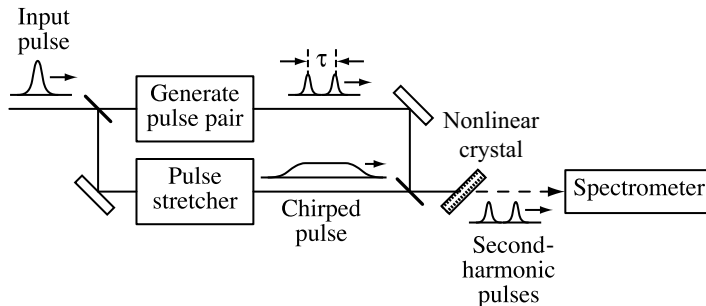
Figure 3.27   Self-referencing spectral shearing interferometry apparatus (SPIDER). Although not depicted explicitly, a background-free geometry is assumed where second harmonic is generated only through the interaction of the pulse pair with the stretched pulse.

delay through $\Omega \approx 2\tau/\beta$. The chirp parameter β should be chosen to satisfy the following constraints:

1. The stretched pulse duration should exceed the pulse pair delay τ.
2. The instantaneous frequency of the stretched pulse should be essentially constant within the duration of a single input pulse. This ensures that the upconversion process faithfully produces a pair of spectrally shifted but otherwise undistorted versions of the complex input spectrum. If the input pulse is close to bandwidth limited, the condition $|\beta| \gg \alpha$ is sufficient to enforce this constant instantaneous frequency constraint. Adherence to this constraint can also be checked experimentally by observing the upconverted spectrum resulting when only one of the pulses from the delayed pulse pair is present (the other pulse is blocked).

Finally, some calibrations are necessary to use the SPIDER technique. The most important is to determine τ accurately, since a linear phase term proportional to τ must be subtracted from the SPIDER data. An error in τ leads to an erroneous quadratic phase term in $\psi(\tilde{\omega})$, even for pulses without any chirp. One calibration method is to use spectral interferometry to determine τ. Spectral interferometry can be performed, with the stretched pulse blocked, either on the delayed pulse pair itself in the fundamental frequency band, or on the pair of second-harmonic pulses produced individually by each of the delayed pulses. In the latter case the second-harmonic crystal orientation is adjusted to permit upconversion of each pulse individually. The accuracy with which τ must be known may be especially demanding for measurement of pulses in the few-cycle regime. A variation of spectral shearing interferometry designed for robust measurement of few-cycle pulses is reported in [169].

3.9   CHARACTERIZATION OF NOISE AND JITTER

The characterization methods described so far in this chapter focus on measurement of isolated pulses. In principle, a perfectly mode-locked train of such pulses would be periodic, with all pulses having the same shape, amplitude, frequency, and timing. In practice, however, these quantities can fluctuate. In this section we discuss characterization of amplitude and timing fluctuations based on radio-frequency (RF) spectral analysis of the mode-locked pulse train after detection via a fast (subnanosecond) photodetector [170]. Using this technique, picosecond timing jitters can be measured using subnanosecond photodetectors!

Consistent with von der Linde's original treatment, we analyze this technique in the context of active mode-locking. Subtle differences that would arise in the passive mode-locking case are mentioned at the end of this section. We write the power of the mode-locked pulse train as

$$P(t) = P_0(t) + \delta P(t) \tag{3.118}$$

Here, $P_0(t)$, written as

$$P_0(t) = \sum_m f_0(t - mT) \tag{3.119}$$

represents the ideal, fluctuation-free pulse train, $f_0(t) = |a(t)|^2$ is the power of a single pulse, T is the period of the pulse train, and $\delta P(T)$ represents the fluctuations. To take into account amplitude and timing variations, we write

$$P(t) = [1 + \mathcal{A}(t)] \sum_m f_0[t - mT - \delta T(t)] \tag{3.120}$$

Here $\mathcal{A}(t)$ and $\delta T(t)$ represent fluctuations in the pulse energy and timing jitter, respectively. Note that changes in the pulse shape itself are not accounted for in this analysis; therefore, pulse amplitude and pulse energy are considered to be equivalent variables. We rewrite this expression in terms of a first order Taylor series expansion:

$$P(t) = [1 + \mathcal{A}(t)] \, P_0(t) - T \mathcal{J}(t) \dot{P}_0(t) \tag{3.121}$$

where $\mathcal{J}(t) = \delta T(t)/T$ is the normalized timing jitter and $\dot{P}_0(t)$ is the first derivative of $P_0(t)$. The first-order series expansion is strictly valid only when δT is much less than the pulse width. However, in the actual experiment, the optical pulse train is detected by a photodetector. The photodetector current response $\mathcal{I}_{\mathrm{PD}}$ can be written

$$\mathcal{I}_{\mathrm{PD}}(t) = P(t) * g_{\mathrm{PD}}(t) \tag{3.122}$$

$g_{\mathrm{PD}}(t)$ is the photocurrent impulse response function. Although $g_{\mathrm{PD}}(t)$ should be chosen to be fast compared to the mode-locked period T, it is usually quite slow compared to the duration of a pulse in ultrafast optics. Once the photocurrent response time is accounted for, the first-order expansion is valid as long as δT is much less than the photodetector impulse response, a much less restrictive requirement.

To relate the fluctuations to the RF spectrum, we first form the correlation function $\langle P(t) P(t - \tau) \rangle$. In contrast to our discussion earlier in this chapter, we are now interested in the behavior of the correlation function for $\tau \geq T$, which is the time scale in which the amplitude and timing fluctuations become important. Assuming that $\mathcal{A}(t)$ and $\delta T(t)$ are independent, zero-mean random variables, we obtain

$$\langle P(t) P(t - \tau) \rangle = \langle P_0(t) P_0(t - \tau) \rangle \left[1 + \langle \mathcal{A}(t) \mathcal{A}(t - \tau) \rangle \right]$$
$$+ T^2 \left\langle \dot{P}(t) \dot{P}(t - \tau) \right\rangle \langle \mathcal{J}(t) \mathcal{J}(t - \tau) \rangle \tag{3.123}$$

The power autocorrelation function depends on the autocorrelation functions of the amplitude and timing fluctuations. Using our earlier results on the Fourier transform relationship between the autocorrelation and the power spectrum, eq. (3.21), we obtain

$$\left| \tilde{\mathcal{I}}_{\mathrm{PD}}(\omega) \right|^2 = \left| \tilde{G}_{\mathrm{PD}}(\omega) \tilde{P}_0(\omega) \right|^2 * \left[2\pi \delta(\omega) + \left| \tilde{\mathcal{A}}(\omega) \right|^2 \right]$$
$$+ \omega^2 T^2 \left| \tilde{G}_{\mathrm{PD}}(\omega) \tilde{P}_0(\omega) \right|^2 * \left| \tilde{\mathcal{J}}(\omega) \right|^2 \tag{3.124}$$

where $\tilde{\mathcal{I}}_{\mathrm{PD}}$, $\tilde{G}_{\mathrm{PD}}$, $\tilde{P}_0$, $\tilde{\mathcal{A}}$, and $\tilde{\mathcal{J}}$ are the Fourier transforms of the corresponding variables without tildes. Substituting for $\tilde{P}_0(\omega)$ using the Fourier transform of eq. (3.119),

we obtain

$$\left|\tilde{\mathcal{I}}_{PD}(\omega)\right|^2 = \left(\frac{2\pi}{T}\right)^2 \sum_k \left|\tilde{G}_{PD}(\omega_k)\tilde{F}_0(\omega_k)\right|^2$$

$$\times \left[2\pi\delta(\omega - \omega_k) + \left|\tilde{\mathcal{A}}(\omega - \omega_k)\right|^2 + \omega_k^2 T^2 \left|\tilde{\mathcal{J}}(\omega - \omega_k)\right|^2\right] \quad (3.125)$$

where $\tilde{F}_0$ is the Fourier transform of f_0 and $\omega_k = 2\pi k/T$.

The photocurrent power spectrum, which can be measured using an RF spectrum analyzer, is pictured in Fig. 3.28. The spectrum consists of a series of lines at DC and at the harmonics ω_k of the mode-locking frequency. The sharp central lines occurring right at the ω_k correspond to the ideal mode-locked train, while the broader low-intensity pedestals are related to the fluctuations. The width of the sharp central lines is limited by the resolution bandwidth of the RF spectrum analyzer. Although not indicated in the figure, the peaks begin to roll off at high frequencies, due to the limited photodetection bandwidth. To analyze the fluctuations, we note the k^2 factor, which multiplies $|\tilde{\mathcal{J}}(\omega)|^2$ but not $|\tilde{\mathcal{A}}(\omega)|^2$. Therefore, we can determine the amplitude noise spectrum $|\tilde{\mathcal{A}}(\omega)|^2$ from the low intensity pedestal centered at DC (or at frequencies ω_k with k sufficiently low that $|\tilde{\mathcal{A}}(\omega)|^2$ dominates). The timing jitter term grows quadratically with k and therefore is the dominant contribution to the noise pedestal at sufficiently large k. This allows determination of $|\tilde{\mathcal{J}}(\omega)|^2$. This behavior is evident in data for the mode-locked argon ion laser shown in Fig. 3.29.

Once $|\tilde{\mathcal{A}}(\omega)|^2$ and $|\tilde{\mathcal{J}}(\omega)|^2$ have been determined, they can be inverse-Fourier-transformed to yield the correlation functions of the amplitude noise and of the timing jitter. The correlation functions characterize the size of the amplitude and timing variations as well as the time scale over which the amplitude and timing fluctuate. The size of the variations is characterized by the variances, which are equal to the correlation functions evaluated at $\tau = 0$. Thus, we have

$$\left\langle \mathcal{A}^2(t) \right\rangle = \frac{1}{2\pi} \int d\omega \left|\tilde{\mathcal{A}}(\omega)\right|^2 \quad (3.126)$$
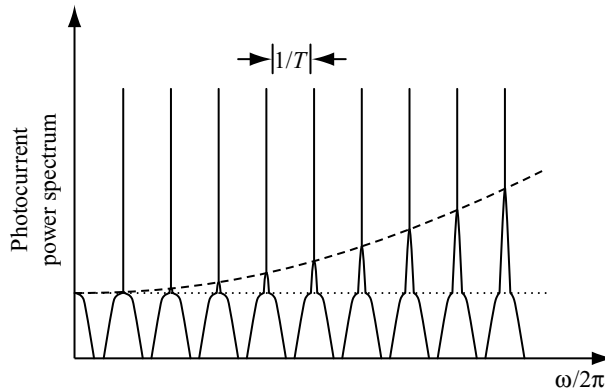
Figure 3.28 Radio-frequency spectrum of a fluctuating mode-locked train. The photocurrent response is assumed to be flat over the frequency range pictured. From Fig. 1 of [170], with permission from Springer Science and Business Media.
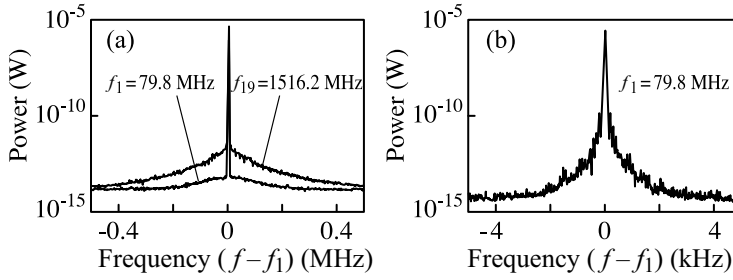
Figure 3.29 (a) RF photocurrent power spectra at the fundamental mode-locking frequency and nineteenth harmonic of a mode-locked argon laser (1-kHz resolution bandwidth); (b) photocurrent power spectrum at fundamental mode-locking frequency with a 30-Hz resolution bandwidth. From Figs. 5 and 6 of [170], with permission from Springer Science and Business Media.

and

$$\left\langle \left(\frac{\delta T(t)}{T} \right)^2 \right\rangle = \frac{1}{2\pi} \int d\omega \left| \tilde{\mathcal{J}}(\omega) \right|^2 \tag{3.127}$$

Practically, in evaluating the area under the noise pedestal, one usually integrates only over a certain range of frequency offsets, $\delta\omega_{\min} \leq |\omega - \omega_k| \leq \delta\omega_{\max}$. There are two reasons for this. First, one may be interested only in the fluctuations occurring within a certain frequency range. Second, there is the practical problem that at sufficiently low frequency offset, the noise pedestal merges with the spike corresponding to the ideal mode-locked train. The frequency at which the spike and pedestal merge depends on the resolution bandwidth $\Delta\omega_{RB}$ of the RF spectrum analyzer. This effect can have a very significant effect on the measurement, as illustrated by Fig. 3.29, which shows RF spectra taken from the same mode-locked laser, with resolution bandwidths of 1 kHz and 30 Hz, respectively [170]. The higher-resolution trace clearly reveals the structure of additional noise energy not visible in the lower-resolution data. Consequently, it is important to specify the frequency range over which the broad pedestal is integrated. The amplitude variance can be computed using

$$\left\langle \mathcal{A}^2(t) \right\rangle = \frac{\left[\int_{-\delta\omega_{\max}}^{-\delta\omega_{\min}} + \int_{\delta\omega_{\min}}^{\delta\omega_{\max}} \right] d\tilde{\omega} \left| \tilde{\mathcal{I}}_{PD}(\omega_k + \tilde{\omega}) \right|^2}{\int_{-\Delta\omega_{RB}}^{\Delta\omega_{RB}} d\tilde{\omega} \left| \tilde{\mathcal{I}}_{PD}(\omega_k + \tilde{\omega}) \right|^2} \tag{3.128}$$

The integrals involve only a single harmonic (single k value), with k chosen sufficiently small that the amplitude fluctuations dominate. The integral over the central peak [denominator of eq. (3.128)] should be carried out over a range sufficient to determine its full area (i.e., over a range at least equal to the resolution bandwidth of the spectrum analyzer). Clearly, $\delta\omega_{\min} \geq \Delta\omega_{RB}$ should be satisfied. The resulting rms amplitude noise given by the square root of eq. (3.128) should be specified in terms of both the percentage value and the corresponding frequency range ($\delta\omega_{\min}, \delta\omega_{\max}$). Computation of the timing jitter follows a similar procedure, but using a higher k value such that the timing jitter dominates the area

under the pedestal. The timing variance is then given by

$$\left\langle (\delta T(t))^2 \right\rangle = \frac{T^2}{4\pi^2 k^2} \frac{\left[\int_{-\delta\omega_{\max}}^{-\delta\omega_{\min}} + \int_{\delta\omega_{\min}}^{\delta\omega_{\max}} \right] d\tilde{\omega} \left[\left| \tilde{\mathcal{I}}_{PD}(\omega_k + \tilde{\omega}) \right|^2 - \left| \tilde{\mathcal{A}}(\omega_k + \tilde{\omega}) \right|^2 \right]}{\int_{-\Delta\omega_{RB}}^{\Delta\omega_{RB}} d\tilde{\omega} \left| \tilde{\mathcal{I}}_{PD}(\omega_k + \tilde{\omega}) \right|^2}$$

$$(3.129)$$

For sufficiently large k, the $|\tilde{\mathcal{A}}(\omega)|^2$ term subtracted in the upper integral is only a small correction, which can be ignored without much loss of accuracy. The rms timing jitter must also be specified together with the frequency range. The photodetector frequency response does not appear in eqs. (3.128) and (3.129) and therefore has no direct effect on the measurement, provided, of course, that the response is sufficiently broadband to allow measurement of the desired harmonics with the desired signal-to-noise ratio. Note that truncation of the RF spectrum for frequency offsets below $\omega_{\min}$ means that information about the full correlation functions $\langle \mathcal{A}(t)\mathcal{A}(t-\tau) \rangle$ and $\langle \mathcal{J}(t)\mathcal{J}(t-\tau) \rangle$ is lost for $|\tau| \overset{>}{\sim} 2\pi/\omega_{min}$. Similarly, truncation for offsets above $\delta\omega_{\max}$ means that information about the correlation functions is lost for $|\tau| \overset{<}{\sim} 2\pi/\omega_{\max}$.

In some experiments one is interested in the relative timing jitter between two pulses emitted from the laser at a specified delay offset. For example, it is the relative timing jitter that would determine broadening of an experimental cross-correlation trace involving interaction of the pulses emitted from a mode-locked laser with delayed earlier pulses emitted from the same laser. The quantity of interest here would be the variance of $[\delta T(t) - \delta T(t-\tau)]$, which is easily shown to be

$$\left\langle [\delta T(t) - \delta T(t-\tau)]^2 \right\rangle = 2 \left\langle [\delta T(t)]^2 \right\rangle \{1 - G_{\mathcal{J}}(\tau)\} \qquad (3.130a)$$

where

$$G_{\mathcal{J}}(\tau) = \frac{\langle \mathcal{J}(t)\mathcal{J}(t-\tau) \rangle}{\left\langle [\mathcal{J}(t)]^2 \right\rangle} = \frac{\int d\omega \left| \tilde{\mathcal{J}}(\omega) \right|^2 e^{j\omega\tau}}{\int d\omega \left| \tilde{\mathcal{J}}(\omega) \right|^2} \qquad (3.130b)$$

is the autocorrelation of $\mathcal{J}$, which can be computed using the inverse Fourier transform of $|\tilde{\mathcal{J}}(\omega)|^2$ extracted from the RF spectrum. For delays less than the correlation time of $\mathcal{J}$, this relative timing jitter can be much less than the absolute timing jitter given by eq. (3.127).

Studies of fundamental amplitude noise and timing jitter properties of actively and passively mode-locked lasers may be found in [171–173]. One key difference between actively and passively mode-locked lasers is that in the actively mode-locked laser, the externally driven modulator acts as a restoring force which damps timing fluctuations of the mode-locked pulses with respect to the external timing source. As a result, in the actively mode-locked case, $\left\langle [\delta T(t) - \delta T(t-\tau)]^2 \right\rangle$ approaches a constant for large τ; the size of the timing fluctuations is bounded. In contrast, for the passively mode-locked case, there is no timing reference or restoring force. The timing of the mode-locked pulses is governed by a random walk process. Therefore, $\left\langle [\delta T(t) - \delta T(t-\tau)]^2 \right\rangle$ grows asymptotically as τ. The size of the timing variations with increased observation time is unbounded; therefore, the first-order Taylor series expansion used to describe the timing jitter in eq. (3.121) is not valid for very large observation times. As a consequence, the form of the photocurrent spectrum is modified for passive mode-locking. In particular, the delta function, which is the signature

of a perfectly mode-locked pulse train in eq. (3.125), is replaced by a central feature of finite width and amplitude. For passively mode-locked lasers, then, it is fundamentally not possible to distinguish the noise and jitter pedestal from the central feature in the RF spectrum at very low frequency offsets. However, except at very low frequencies, eq. (3.125) remains a good description.

Finally, we note that although independent timing and energy fluctuations have been assumed, in principle these fluctuations can have correlations. For example, energy fluctuations lead to gain changes, which imply changes in the refractive index and hence the cavity round-trip time. One consequence of correlations between $\mathcal{A}$ and δT is the possibility of photocurrent power spectra which are not symmetric with respect to the various ω_k. Pulse-width fluctuations, which we have not considered here, are also possible. For further discussion of these effects, see [172,174,175].

PROBLEMS

3.1. Either numerically or analytically, construct a pulse whose time–bandwidth product in terms of FWHM duration and bandwidth is very small. Can you realize a $\Delta \nu \, \Delta t$ value of less than 0.1? Less than 0.01? Also compute the product of the rms pulse duration and bandwidth for your pulse. Comment on your results.

3.2. (a) Spectral interferometry is used to characterize a lossless optical device which imparts a quadratic spectral phase $\Delta \psi(\omega) = A(\omega - \omega_0)^2$ onto the signal field, where $A = 1.5 \times 10^{-26}$ s^{-2}. The input field is a bandwidth-limited Gaussian pulse with center frequency $\omega_0/2\pi = 2 \times 10^{14}$ Hz and FWHM power spectral bandwidth $\Delta \nu = 5 \times 10^{12}$ Hz. Plot the spectral interferometry signal assuming that the reference field is delayed relative to the signal field by (i) 2 ps and (ii) -2 ps. Comment on the main features in your data.

(b) Repeat assuming a chirped Gaussian input, with spectral phase $-A(\omega - \omega_0)^2/2$.

3.3. Figure 3.7 shows power spectra and electric field autocorrelation data for superluminescent diode and mode-locked Ti:S laser light sources. Compare the correlation widths to those expected from the spectra. To relate delay in micrometers, as shown in the figure, to delay in femtoseconds, you may assume that $v_g = c$ (i.e., the group velocity is equal to the speed of light in free space).

3.4. Derive eq. (3.43) relating the rms pulse width to the autocorrelation. [*Hint*: Use the substitution $\eta = t - \tau$ in evaluating the numerator on the right-hand side of eq. (3.43).]

3.5. Consider a frequency-modulated Gaussian pulse with center frequency ω_0 and complex envelope function $a(t) = \exp(-t^2/t_p^2) \exp(jt^2/t_c^2)$. Produce plots of the power spectrum, electric field autocorrelation envelope, (fringe-averaged) intensity autocorrelation, and interferometric intensity autocorrelation, for $t_p = 200$ fs, $\omega_0/2\pi = 5 \times 10^{14}$ Hz, and $t_c = 40$, 200, and 800 fs. Comment on the trends in your plots. For the interferometric autocorrelation, it is sufficient to plot the envelopes of the fringes.

3.6. (a) Verify the expression for the time broadening of the autocorrelation trace for a Gaussian pulse in a noncollinear geometry, eq. (3.53).

(b) By evaluating eq. (3.55) in light of the requirement that the input beam separation (at the lens) exceed the input beam diameter, $\delta x_{in} \geq 2w_{in}$, the text finds a minimum temporal broadening of $\lambda/\pi c$. Referring to the Gaussian beam formulas of Chapter 1, show that the same result is obtained if one instead imposes the requirement that the angular separation of the beams exceed the beam divergence (both evaluated at the nonlinear crystal).

3.7. Using eq. (3.58) for the correlation function measured by third-harmonic generation, give an expression for the contrast ratio of the interferometric correlation trace as a function of η for $0 \leq \eta \leq 1$ [i.e., give the ratios of the maximum (minimum) value at the top (bottom) of a fringe near $\tau = 0$ to the value of the power for large delays where the pulses do not overlap].

3.8. Demonstrate that spectrogram and sonogram time–frequency distributions, eqs. (3.70) and (3.71), are identical, provided that the sonogram filter function and spectrogram gate function satisfy $H(\omega) = G(-\omega)$.

3.9. Consider a pulse whose spectral amplitude function is given by $A(\tilde{\omega}) \sim \exp(-\tilde{\omega}^2 t_p^2/4) \exp(jA\tilde{\omega}^2) \exp(jB\tilde{\omega}^4)$, where $t_p = 100$ fs and (a) $A = \pi \times 10^4$ fs^2, $B = 0$; (b) $A = 0$, $B = (-\pi/16) \times 10^8$ fs^4; (c) $A = \pi \times 10^4$ fs^2, $B = (-\pi/16) \times 10^8$ fs^4. For each case use a real Gaussian gate function to generate a spectrogram, choosing the duration of the gate function to bring out interesting information in your spectrogram plot. Comment on the relationship between your spectrograms and the instantaneous frequency and frequency-dependent delay functions.

3.10. In PG-FROG there is a trade-off between efficiency and nonlinear phase modulation (the latter is usually kept low to avoid distortion of the pulse to be measured). Define efficiency as the fraction of the input power that is passed by the analyzer after the nonlinear index material. How large an efficiency is possible if we restrict the peak nonlinear phase shift to less than 0.1π? To less than 0.01π? To keep the problem simple, you may calculate the instantaneous efficiency at the peak of the gate pulse.

3.11. Derive the formula for the SHG-FROG frequency marginal, eq. (3.97).

3.12. In self-referencing spectral interferometry (SPIDER), an unknown input pulse is split into a pair of identical pulses with relative delay τ and spectral shear Ω. As described in the text, the algorithm for extracting the spectral phase from the SPIDER data requires exact knowledge of τ.

(a) Although the actual delay is τ_0, the experimentalist erroneously believes the delay to be $\tau_0 + \delta\tau$, where $\delta\tau$ is a small error. The erroneous delay value $\tau_0 + \delta\tau$ is used to extract the spectral phase from the SPIDER data. Assuming that all other conditions outlined in the text are satisfied, derive a formula for the error in the estimated spectral phase.

(b) The input pulse is a bandwidth-limited Gaussian with duration Δt (intensity FWHM). Using your result from above, obtain a formula for the estimated pulse duration.

(c) Plot the estimated pulse duration as a function of delay error for $\Delta t = 50$ fs, $\tau = 500$ fs, and $\Omega/2\pi = 400$ GHz. What is the maximum permissible delay error if the pulse duration is to be determined to better than 5% accuracy?

3.13. Again considering self-referencing spectral interferometry, assume that the experimentalist has perfect knowledge of the delay ($\delta\tau = 0$ in the nomenclature of Problem 3.12), but believes the spectral shear to be $\Omega + \delta\Omega$, where Ω and $\delta\Omega$ are, respectively, the actual spectral shear and a small error. Discuss the error introduced into the estimated spectral phase as a result of processing the SPIDER data using the erroneous value of the spectral shear.