The role of dispersion in ultrafast optics

Ian Walmsley
The Institute of Optics, University of Rochester, Rochester, New York 14627

Leon Waxer
Laboratory for Laser Energetics, University of Rochester, 250 East River Road, Rochester, New York 14623

Christophe Dorrer
The Institute of Optics, University of Rochester, Rochester, New York 14627

(Received 22 February 2000; accepted for publication 12 October 2000)

In this article, we review the phenomenon of dispersion, paying particular attention to its impact in the optics of ultrashort pulses, as well as its measurement and management. At present, lasers generating coherent bandwidths of several hundred nanometers have been demonstrated and correspondingly short pulses of 10 fs or so are quite usual. The limits to the breadth of optical spectra and brevity of pulse durations that may be achieved are often set by the dispersive properties of the linear optical elements of which the source is constructed. Progress in ultrafast optics to date has therefore relied extensively on the development of ways to characterize and manipulate dispersion. The means by which this can be accomplished are significantly different for laser oscillators and laser amplifiers, as well as for nonlinear interactions that are used to extend the range of frequencies at which short optical pulses are available, but in all cases it is this phenomenon that determines the output of current optical sources. © 2001 American Institute of Physics.

I. INTRODUCTION

A. General considerations

The work of Edward Muybridge and Harold Edgerton is compelling, as art as well as science, because it provides a view into a world that is outside our everyday experience, and in it reveals a drama that is unanticipated. It is perhaps not too great a leap to argue that the continuation of the legacy of these past masters is to be found currently in the science of ultrafast phenomena. This is the business of looking into the realm of events that take place on femtosecond time scales ($10^{-15}$ s = 1 fs).

The technology required for this activity uses ideas not developed at all even at the time of Edgerton, but nonetheless the basic principle is still the same as that developed by him—scatter a short burst of light from the dynamic entity and detect the reflection. The ability of modern scientists to observe events such as the ballistic transport of electrons in solids, the first events in photosynthesis, the primary dynamics in the biology of the human visual response, and the gyrations of a molecule undergoing a chemical reaction have all been predicated on the ability to manipulate short bursts of light. And this ability has largely been one of being able to manage dispersion, since it is this ubiquitous phenomenon that often limits the duration of the burst that can be generated or delivered effectively to the target.

Another, equally important, branch of ultrafast optics is the generation of very large optical fields using lasers of modest scale. In this branch, too, dispersion plays a central role in both the amplification and frequency conversion of ultrashort pulses. But there are the added technological problems of damage of the optical elements, and the consequent need for large-aperture optics.

In this article, we will discuss the role of dispersion in ultrafast optics in a broad way: how it arises, what its effect on optical pulse propagation is, how it is measured, and how it is manipulated. The subject divides quite naturally into the techniques necessary for the management of dispersion in laser sources, in amplifiers, and in nonlinear optical frequency generation, each of which has special conditions associated with them, such as the requirement for extremely low loss or high-peak power handling capability.

Dispersion is the phenomenon that the phase velocity of a wave depends on its frequency. For optical fields, the basic vacuum dispersion relation is quite simple: $\omega = ck$, where $\omega$ is the angular frequency of the radiation, $k$ the wave number (the modulus of the wave vector), and $c$ a constant velocity. In this case, the phase velocity $\omega/k$ and the group velocity $\partial \omega/\partial k$ of a wave packet centered at frequency $\omega$ are both constant, equal to $c$. In all optical media, the presence of absorptive resonances modifies this dispersion relationship so that the phase and group velocities are different, even in regions of little absorption far away from resonances. The most important consequence of this is that both velocities are
functions of frequency, having the effect that a short pulse will change shape as it propagates. One can think of a short pulse as being made up of a set of much longer wave packets of narrow spectrum, all added together coherently. In a dispersive system, these wave packets all travel at different velocities, and consequently the initial short pulse may broaden. Of course, in reality, things are more complicated than this simple picture: Both diffraction of finite beams and their behavior at interfaces between media of different dispersion, to say nothing of nonlinear effects, are important.

Nonetheless, it is instructive to look at a particular example. The shortest pulses that can be generated by optical sources to date are in the region of 4–6 fs in duration.3,4 For these pulses with spectra centered near 800 nm, this means a 1 fs duration pulse contains many cycles. However they are violet and x-ray regions of the spectrum. In this regime, even this simple situation that keeping a short pulse localized in time at a target, especially when it must be focused onto a target, is not an easy task.

Using Parseval’s theorem, this can be rewritten as

$$\Delta t^2 = \langle (t - \langle t \rangle)^2 \rangle = \int (t - \langle t \rangle)^2 I(t) dt$$

$$= \int |(t - \langle t \rangle)| E(t)^2 dt.$$  (1)

Despite its detrimental effects when using short laser pulses for experiments, dispersion has a very important role to play in actually generating such pulses. The primary source for ultrashort optical pulses is the mode-locked laser oscillator. In this device, the many longitudinal modes of the laser resonator must be kept phase locked and equally spaced in frequency. Dispersion causes the modes to be spaced unequally in frequency, and leads to output pulse durations that are longer than the minimum that can be supported by the amplifier bandwidth. An equivalent picture is that a pulse propagating around the laser cavity must have the same shape after one round trip. Clearly dispersion will cause the pulse to broaden. The nonlinear mode-locking mechanism to some extent balances this broadening tendency, though minimizing linear dispersion is still necessary. Although it has proven possible to generate very short pulses using mode-locked lasers with a variety of gain media, by far the most common source of such pulses are lasers which use titanium sapphire (Ti:sapphire). b This is for purely technological reasons, which are compelling enough to warrant a technology based around a wavelength of 800 nm. However, it is often desirable to generate wavelengths away from this region, and this is the domain of ultrafast nonlinear optics. In this arena, too, dispersion plays a central role. The bandwidth of the nonlinear process in phase matched configurations, such as harmonic generation and other parametric interactions, is limited by the dispersive properties of the nonlinear medium or the geometry of the interaction, or some combination of these.

### B. Linear systems in ultrafast optics

The purview of optics as specified by the wavelength ranges it supposedly incorporates currently extends from the near ultraviolet, around 200 nm, to the mid-infrared, around 2 μm, give or take. Some would include the x-ray and Terahertz regimes. Even optical materials that are transparent over large fractions of the optical region of the spectrum generally have absorption resonances in the near ultraviolet and far infrared, and therefore the change of phase velocity with frequency cannot be ignored for pulses whose spectra are broad enough. In this region of the spectrum, the realm of the ultrafast is similarly malleable, but is generally taken to mean pulses in the picosecond (10^{-12} s), femtosecond (10^{-15} s), or perhaps in the near future attosecond (10^{-18} s) range of durations, with spectra correspondingly of 1 nm to more than 200 nm width. Over such bandwidths, neither material nor other sources of dispersion can be ignored.

It is common to consider a rather simple picture of pulse propagation in first-order analysis of ultrafast optical systems. That is, the electric field \(E_{\text{out}}(t)\) at the output of a linear optical system is related to the input field \(E_{\text{in}}(t)\) via

$$E_{\text{out}}(t) = \int_{-\infty}^{\infty} dt' S(t,t') E_{\text{in}}(t'),$$

where \(S(t,t')\) is the impulse response function of the system. Dispersive elements in ultrafast optical systems are usually taken as phase-only linear filters with time-stationary re-
sponse functions. In that case, taking the Fourier transform of Eq. (3), the input and output field spectra are related by

\[ \tilde{E}_{\text{out}}(\omega) = S(\omega) \tilde{E}_{\text{in}}(\omega). \] (4)

Note that, unlike the case of continuous quasimonochromatic fields, there is no difficulty in defining the spectrum of an optical pulse, which is, by definition, taken to have finite support in the time domain. Nonetheless, there are difficulties in defining the ensemble and time averages in a consistent manner when this approach is taken. The issue is not yet completely resolved formally, but for most practical purposes, the approach taken here appears to give satisfactory results.

The system transfer function \( S(\omega) \) is defined by

\[ S(\omega) \delta(\omega - \omega') = \int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} dt' S(t, t') e^{i\omega t} e^{-i\omega' t'}. \] (5)

The argument of the transfer function \( S(\omega) \) is the spectral phase transfer function \( \phi(\omega) \). This is the phase accumulated by the spectral component of the pulse at frequency \( \omega \) upon propagation between the input and output reference planes that define the optical system. This function plays a central role in ultrafast optics.

It is important to understand the limitations of this model. It derives from the Helmholtz equation under the condition that the system response function does not couple the spatial and frequency variables. This means that the transverse structure of the pulse may be completely suppressed. Thus the possibility that each frequency could be displaced or imaged in a different way by the optical system has been ignored. This might seem a rather drastic reductionist stance, especially when one learns that a common way to control the dispersion of a system is to make use of the spatiotemporal coupling inherent in refraction and diffraction, but it is also understood that the vast majority of useful ultrafast optical systems have the feature that even though the pulse may be spatially dispersed inside the system, at the input and output all frequencies are spatially overlapped. Systems that do not satisfy this criterion are said to generate "spatial chirp," one of the banes of current ultrafast technology. Such a simple model should be used only to implement a first-order design or analysis of a system for the purposes of evaluating dispersion. Despite this limitation, the model works remarkably well in specifying the dispersive character of a system.

Usually a system is designed to first order using a standard set of equations that suppose ideal conditions. Next the well-known connection between paraxial optics and Gaussian beams is used to evaluate, and perhaps optimize, the parameters of an optical system. Beyond this, more sophisticated ray-tracing methods are used to evaluate aberrations. The conclusion of this design is usually to provide a specification for the layout and elements of a system operating within the paraxial regime, whose dispersive properties are given by the coefficients of a polynomial approximation to the spectral phase function \( \phi(\omega) \):

\[ \phi(\omega) = \sum_{n} \frac{1}{n!} \phi^{(n)}(\omega_0)(\omega - \omega_0)^n. \] (6)

where \( \phi^{(n)}(\omega_0) \) is the \( n \)th derivative of the phase function with respect to frequency evaluated at a reference angular frequency \( \omega_0 \). This quantity is specified in units of \([\text{rad}^{-1}\cdot\text{fs}^2]\), most of the time simplified to \([\text{fs}^2]\). The terms have the following interpretation:

The absolute phase \( \phi^{(0)}(\omega_0) \) measures the phase accumulated at the reference frequency \( \omega_0 \). In terms of optical pulse propagation, it specifies, for simple pulse shapes, where the optical reference is located with respect to the peak of the envelope of the pulse. It is usually ignored in the design of a system because it does not play a role in the interaction of pulses with matter at durations longer than one or two optical cycles. Its measurement is currently a field of intense research.\(^7\text{--9}\)

The group delay \( \phi^{(1)}(\omega_0) \) is the time taken for a wavepacket centered at the reference frequency to propagate between the reference planes. It is likewise not considered important. All of the present techniques for the characterization of ultrashort pulses that are self-referencing (such as FROG\(^10\) and SPIDER\(^11\)) cannot measure the absolute phase or the group delay at the reference frequency, since this requires precise knowledge of a reference. The required reference information is usually another pulse that has not traversed the optical system under study, or the precise positions of the input and output reference planes.

The lowest order term that is considered useful is \( \phi^{(2)}(\omega_0) \), known as the group delay dispersion (GDD) or the quadratic phase. The terms "positive dispersion" and "negative dispersion" are defined regarding the sign of this quantity. It is specified in units of \([\text{fs}^2]\) and measures the rate at which a pulse centered at the reference frequency will increase in duration upon propagation through the system. A transform-limited Gaussian pulse with duration \( \tau_0 \) (specified by the FWHM of the intensity) centered at frequency \( \omega_0 \) is changed to a Gaussian pulse with duration \( \tau_0 \) such as

\[ \tau_0 = \sqrt{1 + \frac{16(\ln 2)^2(\phi^{(2)}(\omega_0))^2}{\tau_0^2}} \] (7)

by a system with group delay dispersion only. The pulse is also said to be "chirped." That is, its instantaneous frequency, defined as \( \partial \arg[E(t)]/\partial t \), is a function of time.

Higher-order coefficients are known as the cubic, quartic, quintic or, in general, the prosaic but accurate \( n \)th-order spectral phase.

In fact the polynomial approximation to the spectral phase function is rather too simple, in a strict sense, and proposals have been made for the tests of its inadequacy for optical materials.\(^12\) Despite this, designs for dispersive elements based on the natural dispersion of optical materials are often based on this approximation, with the group delay dispersion and higher order spectral phase often derived from a Sellmeier equation for the refractive index of the material. One-dimensional models have proven effective in both design and in assessing the effects of system parameter changes on the pulse propagation.
Finally, it is worth noting that the specification of the dispersion given here is unique to the ultrafast optical community. In fiber optics, for example, where the quadratic phase is the dominant factor, other terms are used. In particular the $\beta$ parameter of a fiber is related to the group delay dispersion via

$$\beta = \frac{\phi^{(2)}(\omega_0)}{L} \frac{\text{ps}^2}{\text{km}},$$

where $L$ is the length of the fiber, and $1 \text{ ps} = 1$ picosecond $= 10^{-12}$ s.

Likewise the fiber $D$ coefficient measures the group delay dispersion per unit bandwidth and is related to the modulus of $\beta$ by

$$D = \frac{\omega_0^2}{2 \pi c} |\beta| \frac{\text{ps}}{\text{nm} \cdot \text{km}}.$$  

II. PHYSICAL ORIGINS OF DISPERSION

There are three sources of dispersion in optics, arising from refraction, diffraction, and interference. The first of these is the most pervasive, and the best known: the variation of the refractive index of matter with frequency. The second arises from the coupling of the wave vector direction and frequency that occurs at dielectric interfaces at non-normal incidence or upon diffraction. The third has its origin in the interference of waves in periodic structures. Although there is a little flexibility in choosing the first of these, often the dispersion arising from materials is set by other constraints in the system: the damage threshold, or whether the material is an appropriate gain medium, for example. The second and third mechanisms are usually used to counteract certain parts of the dispersion arising from materials, since they can be adjusted. The price for this adjustability is often increased complexity in fabrication or alignment.

A. Material dispersion

When an electromagnetic wave is incident on an atom, the atom is polarized. The coherent superposition of many such dipoles in a crystalline or amorphous material leads to the linear response from which is defined the bulk refractive index. The Lorentz atom provides a simple model for this effect. Consider an electron bound to a heavy nucleus. The electron is taken to behave as if bound to its equilibrium position with respect to the nucleus in a harmonic potential of frequency $\omega_0$. When driven with an optical field at frequency $\omega$, a polarization is induced. In the Lorentz model this polarization is always proportional to the applied field strength, with the proportionality constant being the linear susceptibility. The susceptibility of the medium consisting of a dilute mixture of such noninteracting atoms is then related to the strength of the field (assuming that the density of polarizable atoms in the sample is $N$)

$$\chi^{(1)}(\omega) = \frac{N e^2}{m} \frac{1}{(\omega_0^2 - \omega^2) + i \omega \Gamma},$$

where $\omega$ is the frequency of the applied optical field, $m$ is the reduced mass of the electron, $e$ its charge, and $\Gamma$ a phenomenological damping rate for the electronic vibrational motion. The real and imaginary parts of the susceptibility are plotted in Fig. 1(a). The refractive index $n$ is related to the susceptibility by the Lorentz–Lorenz formula, which approximates for weakly absorptive media as

$$n^2(\omega) - 1 = 4 \pi \chi^{(1)}(\omega).$$  

The square root of this function near an absorption resonance is plotted in Fig. 1(b). For most optical materials, the resonance frequency $\omega_0$, at which the material is absorptive is in the ultraviolet region of the spectrum, corresponding to wavelengths in the range 150–300 nm. Thus the interesting portion of the curve in many applications in ultrafast optics is the region of positive group delay dispersion for frequencies lower than resonance. Typical values of GDD are about 500 fs$^2$/cm.

The Lorentz model is easily extended to the case of several resonance frequencies, which is more common. Moreover, the Lorentz–Lorenz formula has the same form as Eq. (10) even in strongly absorbing (but nonetheless linear) media provided the resonance frequencies are modified in the manner prescribed by Sellmeier. It is usual to use Sellmeier’s formula

$$n^2(\omega) - 1 = \sum_j \frac{A_j}{\omega_0^2 - \omega_j^2}$$

to determine the refractive index of material, and values of the parameters $A_j$ and $\omega_j$ are tabulated for many optical glasses and crystals.

We point out that another way of describing the index variation of a material with the wavelength is by the use of the Schott coefficients. These coefficients, which are the coefficients of a power series, are calculated to give the best accuracy when fitting an experimental determination of the index with the calculated index on a broad range of wavelengths. They thus do not take explicitly into account properties of the material’s resonance frequencies.

An important limitation of the Lorentz model atom is that an optical material can act only as an absorber or scatterer, never as an amplifier. For this, one must use a quantized model of the atom. The simplest such model represents the atom as a two-level system for purposes of its interaction with laser light. In the linear regime, in which the atomic populations of the two states are fixed, the polarizability of a sample of atoms of this type is

$$\chi^{(1)}(\omega) = \frac{2 d_{12}^2}{\hbar} \frac{(N_2 - N_1)}{(\omega_0^2 - \omega^2) + i \Gamma},$$

where $d_{12}$ is the quantum mechanical dipole matrix element, $N_2$ is the density of atoms in the upper state, and $N_1$ the density in the lower state. The difference in these numbers is the inversion density, and is negative for absorbers, positive for amplifiers. Importantly, the sign of the GDD changes compared to that of an absorber when the material is used as an amplifier. The gain $G_0$ of a linear amplifier is related to the inversion density via

$$\ln(G_0) = \sigma (N_2 - N_1)L,$$  

where $\sigma$ is the absorption cross-section.
where \( \sigma \) is the stimulated emission cross section and \( L \) the length of the amplifier (a similar relation holds for the linear absorption).

When a pulse traverses a sample of material, say a block of glass used at normal incidence, the phase transfer function is

\[
\phi(\omega) = \frac{\omega}{c} n(\omega)L
\]

(15)

with the refractive index usually evaluated from the susceptibility using the approximate Lorentz–Lorenz or Sellmeier formula.

One way to effect minor adjustments of dispersion is to use materials that exhibit birefringence. These materials have different refractive indices for waves polarized in the plane containing the symmetry axis of the material and the wave vector of the light (the extraordinary wave) and waves polarized orthogonally to this plane. While this phenomenon is critical for ultrafast nonlinear optics, it is not commonly used for the linear systems.

**B. Interference**

A second source of dispersion is the interference of waves in periodic or waveguiding structures. Significantly, the dispersion depends on the topology of the periodic structure and can therefore sometimes be specified over a wide range, including the region of negative dispersion. This flexibility of design makes such interferometrically induced dispersion a very useful tool in ultrafast optics.

The underlying principle by which dispersion occurs via interference is this: A periodic structure will transmit waves of certain frequencies, and reflect others. This occurs usually for wavelengths comparable to the periodicity of the structure, which undergo strong Bragg-type scattering. Thus the periodicity effectively induces a resonance in the transfer function of the system, which has dispersion associated with it.

Perhaps the simplest of such devices will serve to illustrate the point. This is the Gires–Tournier interferometer, (GTI) as shown in Fig. 2a.\(^{14}\) It is a variant of the better known Fabry–Perot interferometer, consisting of two parallel, planar mirrors. One of these has a reflectivity of 100\%, and the other is partially reflective with reflectivity \( R \). The separation of the reflective surfaces is \( L \). Clearly the modulus of the transfer function is unity for this device, yet the phase transfer function is not constant.

The output field can easily be written in terms of the input field and the parameters of the interferometer, leading to a transfer function of the form \( H(\omega) = e^{i\varphi(\omega)} \), where

\[
\varphi(\omega) = -\tan^{-1}\left[ \frac{(1-R)\sin(\psi - \omega T)}{-2\sqrt{R + (1+R)\cos(\psi - \omega T)}} \right]
\]

(16)

is the phase transfer function. In this formula, \( \psi \) is the phase change of the field at the highly reflective surface, and \( T = 2L/c \) the round-trip time for the light in the GTI. A sketch of the function is shown in Fig. 2b, indicating that for optical frequencies near the resonance frequencies of the interferometer numbered by an integer

\[
\omega_m = m\frac{c}{2L} = m\frac{1}{T}
\]

(17)

there can be a significant phase accumulation. This can be interpreted as an extra path length that fields at these frequencies experience—they are effectively trapped in the structure for several round trips, whereas waves at frequencies between the resonances escape the device after a single round trip.

The GTI has few design parameters, so it is not possible to fabricate interferometers with complicated phase transfer functions using only two reflective surfaces. It is possible to add several more interfaces, however, and to fabricate a mirror that will have almost arbitrary dispersion. The design of such multilayer interferometers is rather involved, nonetheless, the same principle is at work as in the GTI—each frequency is constrained by interference to spend a different time in the structure. Commonly these consist of several layers of thin inorganic films on a glass or crystal substrate and are designed so that waves at some frequencies are reflected from the surface layers, whereas those at other frequencies must penetrate the structure more deeply before they are reflected. This geometrical picture works quite well for estimating the GDD experienced by each frequency.

Dispersion in a waveguide can be introduced using the example of an optical fiber as shown Fig. 3a, consisting of a thin cylinder of glass a few microns in diameter, the core, surrounded by a thin tube of glass of lower refractive index,
the cladding. The dispersive properties of such a structure have three components: modal dispersion and chromatic dispersion arising from material dispersion and waveguide dispersion.

The fiber supports several modes, whose propagation constants (analogous to wave numbers) are different. This is the origin of modal dispersion. A geometrical picture similar to that suggested for interferometers can be used to understand this source of dispersion. In terms of ray trajectories, one can imagine that the lowest order mode corresponds to the propagation of light along a trajectory through the center of the fiber. A higher order mode follows a trajectory that bounces off the ‘‘walls’’ of the fiber core. Increasingly higher order modes have larger and larger numbers of bounces, meaning that their optical path length is actually much longer than the physical fiber length.

Chromatic dispersion in fibers has two sources; the material of which the fiber is fabricated, and the details of its construction. The effects of material dispersion are the same as for bulk materials. In the visible and near infrared, the material dispersion is positive. This can be counteracted to some extent by negative dispersion due to waveguiding. The confinement of the fundamental mode of the fiber depends on the difference in refractive index between core and cladding, and the boundary conditions which lead to this also cause the dispersion to be negative near the waveguide cutoff frequency. This is exactly opposite to the case in bulk materials, and waveguide dispersion can lead to dispersion flattened or dispersion compensated fibers at telecommunications wavelengths [see Fig. 3(b)]. A structure having a zero-group-velocity dispersion wavelength close to 800 nm has recently been reported, using the cancellation of the material dispersion due to silica by waveguide dispersion.15

The details of the waveguide dispersion are critically dependent upon the way in which the fiber is fabricated. The wavelength at which the GDD is zero can therefore be shifted by proper tailoring of the refractive index profile across the fiber. Such fibers are called dispersion-shifted fibers, since the minimum of the GDD is shifted away from 1330 nm, which is the minimum for silica. The cross section of a fiber whose dispersion has been shifted to locate a minimum for wavelengths near 1550 nm (a standard telecommunications wavelength) is shown in Fig. 4(a). The core of the fiber is doped with germanium by diffusion. This leads to a radial gradient of refractive index that modifies the waveguide dispersion according to the size of the gradient and the overall Ge concentration. The GDD of this fiber is shown in Fig. 4(b), and is negative over the entire range of telecommunications wavelengths.

C. Geometrical dispersion

The bending of light at the interface of two dielectrics or by diffraction at a periodic interface leads to a very powerful method for engineering GDD. The basic idea is illustrated in Fig. 5, which shows a pair of prisms arranged in opposition at minimum deviation, with parallel faces. A pulse of light incident from the left is angularly dispersed by refraction through the prism. The angle of deviation between the input and output rays is set by the refractive index of the prism material. Since the material has inherent dispersion, the angle of deviation depends on the input wavelength. The action of the second prism is to exactly cancel the angular dispersion of the first, so that the rays for different wavelengths emerge mutually parallel to the input ray. They are, however, laterally displaced according to wavelength, which is known as spatial chirp.

Consider the time taken for wave packets centered at several different frequencies within the original pulse spectrum to reach a plane perpendicular to the outgoing rays. It is clear that the shorter wavelengths traverse a longer geometrical path than the longer wavelengths. The longer wavelengths, however, experience a longer path in the prisms than the shorter wavelengths, and a straightforward calculation of the total phase accumulation shows that the group delay may be longer for longer wavelength wave packets, provided the separation of the prisms is large enough. The symmetry of the system can be used to eliminate the spatial chirp, which allows the use of a spectral phase function. If a mirror is placed at the location of the reference plane, then each ray retraces its path, and the outgoing beam has every ray directed antiparallel to the input, and collocated on top of it. This arrangement can provide negative or positive dispersion, which can be adjusted by translating the prisms into or out of the beam to insert more or less positively dispersive material.16 It is also possible to adjust the prism separation (while keeping the faces parallel) but this is significantly more difficult, and seldom used.

Diffraction gratings can also be used to assemble an optical system with negative dispersion, in the manner shown in Fig. 6. Again a pair of elements is used, with the grating vectors parallel. A pulse incident from the left on the first grating is diffracted, so that the wave packets making up the pulse are angularly dispersed, with the longer wavelengths deviated more than the shorter ones. The energy scattered into one of the first-order beams can be 90% or more of the incident energy if the correct geometry is used. The second grating recollimates the dispersed beam, and a mirror placed at the right-hand reference plane will retroreflect the different wave packets so that they are again spatially superimposed at the output. By inspection one can see that in this system the

![Image](https://example.com/image.jpg)
longer wavelength wave packets traverse a longer path than the shorter wavelength ones. In this case, though, there is no material dispersion to compensate this negative dispersion, and this arrangement of gratings always has negative GDD, in an amount proportional to the grating spacing.

The origin of dispersion in both optical systems is the angular dispersion that arises from refraction or diffraction. The amount of dispersion is easily calculated for systems of this class, by calculating the phase accumulated between the input and output reference planes, as shown in Fig. 7. A beam with input wave vector \( \mathbf{k}_{\text{in}} \) in the direction \( \mathbf{l} \) is scattered by element 1 into a direction \( \mathbf{k}_{\text{out}} \). The beam passes between the first and second elements and is scattered back into its original direction. The input and output reference planes are taken perpendicular to the input wave vector at its intersection with the first and second elements. The phase accumulated by the scattered beam relative to the (unscattered) reference is simply

\[
\phi(\omega) = \mathbf{k}_{\text{out}}(\omega) \cdot \mathbf{l}.
\]  

For propagation in free space between the two elements, we may take \( |\mathbf{k}_{\text{out}}| = \omega/c \), so the phase transfer function becomes

\[
\phi(\omega) = \frac{\omega}{c} \mathbf{l} |\cos(\gamma - \alpha(\omega))| = \frac{\omega}{c} \cos(\gamma - \alpha(\omega)),
\]

where \( \gamma \) is the angle between the incident wave vector and the normal to element 1, and \( \alpha \) the angle of the outgoing wave vector. The latter is a function of frequency, whose functional form depends on the details of the scattering mechanism. \( D \) is the spacing between the scattering elements along a direction parallel to their normal.

To be concrete, here we treat the case when both elements are diffraction gratings. In that case the input beam is diffracted into a direction corresponding to a wave vector \( \mathbf{k}_{\text{out}} = \mathbf{k}_{\text{in}} + m\mathbf{k} \), where \( \mathbf{k} \) is the grating wave vector, and \( m \) is the order of diffraction. \( \mathbf{k} \) lies in the plane of the grating and has magnitude \( |\mathbf{k}| = 2\pi/d \), where \( d \) is the groove spacing of the grating. Straightforward algebra then gives the grating equation relating the input wave frequency and the diffracted angle \( \alpha \):

\[
m \frac{2\pi c}{\omega} = m\lambda = d[\sin(\alpha(\omega)) - \sin \gamma].
\]

Together with Eq. (19), this allows analytic expressions for the group delay, GDD, and higher-order spectral phase to be calculated. The expressions for the quadratic (GDD), cubic, and quartic spectral phases are (for single pass)

\[
d^{(2)}(\omega) = -\frac{4\pi^2 cD}{\omega^2 d^2 \cos^3(\alpha(\omega))} = -\frac{\lambda^3 D}{2\pi c d^2 \cos^3(\alpha)},
\]

\[
d^{(3)}(\omega) = \frac{3(4\pi^2 c D)}{\omega^3 d^2 \cos^3(\alpha(\omega))} \left[ 1 + \frac{2\pi c \sin(\alpha(\omega))}{\omega d \cos^2(\alpha(\omega))} \right]
\]

\[
= \frac{3\lambda^4 D}{4\pi^2 c^3 d^2 \cos^3(\alpha)} \left[ 1 + \frac{\lambda \sin(\alpha)}{d \cos^2(\alpha)} \right],
\]

\[
d^{(4)}(\omega) = -\frac{3(4\pi^2 c D)}{\omega^4 d^2 \cos^3(\alpha(\omega))} \left[ 4 + \frac{2\pi c \sin(\alpha(\omega))}{\omega d \cos^2(\alpha(\omega))} \right] + \frac{(2\pi c)^2}{\omega^2 d^2} \left[ 1 + \tan^2(\alpha(\omega))(6 + 5 \tan^2(\alpha(\omega))) \right]
\]

\[
= -\frac{3\lambda^5 D}{8\pi^2 c^3 d^2 \cos^3(\alpha)} \left[ 4 + \frac{8\lambda \sin(\alpha)}{d \cos^2(\alpha)} \right].
\]

FIG. 4. Chromatic dispersion characteristics of a dispersion-compensating fiber (DCF). (a) Cross-section of the fiber refractive index profile, for a fiber with cutoff wavelength (minimum supported wavelength) of 1.55 \( \mu \)m. A radial gradient in the germanium doping of the core changes the refractive index by about 1% across the core. (b) The dispersion function \( \beta \) of the fiber shown in (a).

FIG. 5. Arrangement of prisms providing adjustable group-delay dispersion. The dispersion of the refractive index of the prism material leads to a geometrical dispersion of broadband input light. In the optical region of the spectrum, longer wavelength radiation (dashed line) travels a shorter path in air than shorter wavelength radiation (solid line). In the material, however, it travels a longer path, so that by adjusting the separation of the prisms in such a way that these effects compensate, the total dispersion can be set to either positive or negative values.
A. Introduction

The design of dispersive systems calls for ways in which dispersion may be measured. Since dispersion is generally considered a property of optical systems with time-shift-invariant response functions (the output pulses are the same no matter at what time they enter the system) certain important relationships hold. For example, the real part of the transfer function $\tilde{S}(\omega)$ is related to the imaginary part by a Hilbert transform:

$$\text{Re}[\tilde{S}(\omega)] = \frac{1}{\pi} P \int_{-\infty}^{\infty} d\omega' \frac{\text{Im}[\tilde{S}(\omega')]}{\omega - \omega'}. \quad (24)$$

Since it is not simple to conceptualize a method for measuring one or other of these components, determining its partner by means of this transform appears moot. Nonetheless, there are certain classes of systems for which an analogous relationship between the modulus and argument of the transfer function holds.

It is straightforward to measure the absorption coefficient spectrum of an optical system. This is defined as $\alpha(\omega) = -\text{Re}[\ln(\tilde{S}(\omega))]$. For materials this provides a way in which the spectral phase transfer function can be determined from measured data—a sort of generalized Kramers–Kronig relationship.

However, for certain other systems, no such relationship exists. If the complex transfer function $\tilde{S}_L(p)$ (the Laplace transform of the response function) has zeroes at some value of $p$, then the transform cannot be performed, unless these singularities in the generalized absorption coefficient $\alpha(p)$ are subtracted from the function. The elimination of these so-called Blaschke factors allows the reconstruction of the phase transfer function from a measurement of the amplitude transfer function.\(^18\) However, it is not usually possible to identify the location of the singular points from a measurement of absorption: Their positions need to be known, at least approximately, beforehand. Generally the applicability of the generalized Kramers–Kronig transform is predicated on a more or less detailed knowledge of the system transfer function from analysis rather than from measurement.\(^19\) But

III. MEASUREMENT OF DISPERSION

A. Introduction

The design of dispersive systems calls for ways in which dispersion may be measured. Since dispersion is generally considered a property of optical systems with time-shift-invariant response functions (the output pulses are the same no matter at what time they enter the system) certain important relationships hold. For example, the real part of the transfer function $\tilde{S}(\omega)$ is related to the imaginary part by a Hilbert transform:

$$\text{Re}[\tilde{S}(\omega)] = \frac{1}{\pi} P \int_{-\infty}^{\infty} d\omega' \frac{\text{Im}[\tilde{S}(\omega')]}{\omega - \omega'} \quad (24)$$

Since it is not simple to conceptualize a method for measuring one or other of these components, determining its partner by means of this transform appears moot. Nonetheless, there are certain classes of systems for which an analogous relationship between the modulus and argument of the transfer function holds.

It is straightforward to measure the absorption coefficient spectrum of an optical system. This is defined as $\alpha(\omega) = -\text{Re}[\ln(\tilde{S}(\omega))]$. For materials this provides a way in which the spectral phase transfer function can be determined from measured data—a sort of generalized Kramers–Kronig relationship.

However, for certain other systems, no such relationship exists. If the complex transfer function $\tilde{S}_L(p)$ (the Laplace transform of the response function) has zeroes at some value of $p$, then the transform cannot be performed, unless these singularities in the generalized absorption coefficient $\alpha(p)$ are subtracted from the function. The elimination of these so-called Blaschke factors allows the reconstruction of the phase transfer function from a measurement of the amplitude transfer function.\(^18\) However, it is not usually possible to identify the location of the singular points from a measurement of absorption: Their positions need to be known, at least approximately, beforehand. Generally the applicability of the generalized Kramers–Kronig transform is predicated on a more or less detailed knowledge of the system transfer function from analysis rather than from measurement.\(^19\) But
such arguments are specious if there exists a way to measure
the phase transfer function directly.\textsuperscript{20}

Another reason for wanting direct measurement is that
the structure of the phase function may depend on reso-
nances that are located at frequencies quite far from the
spectral region of interest. Performing the inverse trans-
mform therefore requires quite accurate measurements of the
absorption over an extended range of wavelengths, some of
which might be difficult to access. Direct measurements of
the phase transfer, on the other hand, require only the same
light sources with which the optical systems are to be used.

**B. Measurement of dispersion of an optical element**

The standard way in which dispersion is measured is via
interferometry. Experimentally, the requirement is only to
have a broadband source of arbitrary coherence, but using a
coherent source, such as a laser, usually simplifies measure-
ments for highly scattering or weakly transmissive systems.
In the context of ultrafast optics, the earliest attempt to char-
acterize relevant optical systems was the compensated Mich-
elson interferometer of Knox \textit{et al.}\textsuperscript{21} shown in Fig. 9(a).
In this device, the system whose dispersion is to be measured is
placed in one arm of the interferometer. A light source, fil-
tered by a passband spectral filter whose transmission range
is centered at \(\omega_0\), is used to illuminate the interferometer,
and an integrating square-law detector monitors the power at
the output port as a function of the delay between the two
arms of the interferometer. The delay can be adjusted by
moving one of the mirrors (M\textsubscript{2} in Fig. 9). The detected signal
\(I(\tau)\) is

\[
I(\tau) = \int dt \left| E_i(t + \tau) + E_i(t) \right|^2 = 2\varepsilon_{\text{in}} + 2 \Re \int d\omega \left| \tilde{E}_i(\omega) \right|^2 \tilde{F}(\omega) \tilde{S}(\omega) e^{i\omega\tau},
\]

where \(\varepsilon_{\text{in}}\) is the energy of one of the input pulses and
\(\left| \tilde{E}_i(\omega) \right|^2\) is the spectrum of the input radiation (defined ac-
\ding to the Wiener–Khintchine prescription for partially
coherent light sources), and \(\tilde{F}(\omega)\) the transfer function of the
spectral filter. The only contributions to the signal are those
from a narrow frequency range centered near the filter pass-
band at \(\omega_0\). This frequency is, in effect, a reference fre-
quency for the optical system. If the filter is sufficiently nar-
rowband, the spectral phase function of the optical system can
be truncated at the first-order term, the group delay at
\(\omega_0\). Then \(I(\tau)\) becomes

\[
I(\tau) = 2\varepsilon_{\text{in}} \left[ 1 + F(\tau - \phi^{(1)}(\omega_0)) \cos(\phi(\omega) + \phi^{(1)}(\omega_0) - \omega_0 \tau) \right],
\]

where \(F(t)\) is the inverse Fourier transform of \(\tilde{F}(\omega)\). It is easy to see that the correlation function taken with the optical
system in place is simply a delay-shifted replica of the
autocorrelation function without the system, the delay being
simply the group delay \(\phi^{(1)}(\omega_0)\), as shown in Fig. 9(b). Thus a numerical cross corre-
lation of \(I(\tau)\) measured with and without the test system yields
the group delay at the filter pass frequency with reasonable
accuracy.

It was shown by Naganuma \textit{et al.}\textsuperscript{22} that a tunable spec-
tral filter was not needed, and that without it the complete
spectral phase function could be determined. The apparatus
is shown in Fig. 10(a). The detected correlation function \(I(\tau)\)
is given by Eq. (25) with \(\tilde{F}(\omega)\) equal to a constant.

A sketch of this function when the input spectrum is
quite broad is shown in Fig. 10(b). The spectral phase func-
tion is extracted from this data in a direct way. The Fourier
transform of the data with respect to the delay \(\tau\) is taken, to
yield

\[
\tilde{I}(\Omega) = \text{FT}[I(\tau); \tau \rightarrow \Omega] = 2\delta(\Omega) + |\tilde{E}_i(\Omega)|^2 \tilde{S}(\Omega) + |\tilde{E}_i(-\Omega)|^2 \tilde{S}_0(-\Omega).
\]

\(\tilde{I}(\Omega)\) consists of a large peak near \(\Omega = 0\), and two satellite
peaks near \(\Omega = \pm \omega_0\), where \(\omega_0\) is the mean frequency of the
input light. The argument of the term representing the peak
near \(\Omega = -\omega_0\) is the spectral phase transfer function of the
optical system at \(\Omega\).

Increased accuracy can be obtained using spectral
interferometry.\textsuperscript{23} The apparatus shown in Fig. 11 illustrates a
third Michelson interferometer arrangement, in which a tun-
able spectral filter is placed before the detector. As the pass-
band of this filter is tuned, the detector sees the power in the
output arm modulate, as the spectral interferogram is mapped
out. This interferogram contains the same information as the
previous correlation functions. Returning to Eq. (25), let the
spectral filter transfer function $\tilde{F}(\omega) = T\delta(\omega - \omega_0)$ be a real function of the tunable frequency $\omega_0$. Considered now as a function of $\omega_0$, for fixed $\tau$, the measured signal is an interferogram of the form

$$\tilde{I}(\omega_0; \tau) = 2T|\tilde{E}_i(\omega_0)|^2 + 2|\tilde{E}_i(\omega_0)|T \cos(\phi(\omega_0) - \omega_0 \tau).$$  

(28)

Now, as $\omega_0$ is varied, $\tau$ can also be modified so as to keep the argument of the cosine constant. In this way the interference maximum defined by the contour can be determined. If $\tau$ is known to sufficient accuracy, then the spectral phase function can be mapped out. In fact it is only necessary to know the relative changes in $\tau$ accurately, provided the constant spectral phase term is not needed.

Rather than measuring the argument of the cosine directly, in practice the detector signal is fed back to mirror $M_2$ of the interferometer, which is moved so that the interferometer is always locked on to the peak of a fringe in the spectral interferogram. The displacement of the mirror is measured using a separate single-frequency laser (of wavelength $\lambda$) to an accuracy of better than $\lambda/50$. The displacement is equal to the group delay at the filter frequency. The increased complexity of this apparatus nevertheless leads to increased accuracy in the spectral phase function: the group delay can be mapped as a function of frequency with an accuracy of about 0.2 fs.

An arrangement that avoids scanning of any kind, either delay or spectrometer tuning, is shown in Fig. 12(a). A range of delays is provided by tilting the mirror in the reference arm of the interferometer, into which a quasi-one-dimensional (broadband) light source is collimated. At any given frequency, tilt fringes are seen at the output of the interferometer, in a direction parallel to the tilt of the reference mirror. A slit is placed at the output of the interferometer, with a grating behind it, arranged with grooves parallel to the slit direction. This causes the interferograms for each wavelength to be laterally separated at the image plane of a lens placed beyond the grating. A two-dimensional charge-coupled device camera is placed at the focal plane of the imaging lens, and the interferograms recorded. Examples of these are shown in Fig. 12(b). The horizontal axis (the direction of the grating vector) represents wavelength. The vertical axis is the tilt interferogram for each wavelength, so that this axis represents phase—a $2\pi$ shift between adjacent fringes.

The resulting two-dimensional interference pattern has fringes that follow contours of equal phase. As a result it is easy to read the spectral phase directly from the pattern—the shape of each fringe tells the form of the phase polynomial. The quadratic phase appears as parabolic fringes, cubic as $s$-shaped patterns, and so on.

Currently, perhaps the most widely used technique for the measurement of linear transfer functions is the method of Fourier transform spectral interferometry. Pioneered by Froehly, this technique has found wide specific applications in the field of ultrashort optics in the last few years; however, as for the other interferometric techniques, its application to the measurement of dispersion only depends on the spectral density of the available light source. In this method, the output of the interferometer can be frequency resolved for a fixed delay $\tau$ in a single shot using a spectrometer. The apparatus is identical to that shown in Fig. 11, except that the delay is set at a value that is longer than the correlation time of the input light. The spectrally resolved interferogram is

$$\tilde{I}(\omega_0; \tau) = 2T|\tilde{E}_i(\omega_0)|^2 + 2|\tilde{E}_i(\omega_0)|^2 T \cos(\phi(\omega_0) - \omega_0 \tau).$$  

(29)

Extraction of the spectral phase transfer function is straightforward. The Fourier transform of this interferogram with respect to the frequency is composed of three peaks. At $t=0$, one finds the Fourier transform of the noninterfering terms while at $t=\tau$ and $t=-\tau$ stand the Fourier transform of $|\tilde{E}_i(\omega_0)|^2 T \exp[i(\phi(\omega_0) - \omega_0 \tau)]$ and $|\tilde{E}_i(\omega_0)|^2 T \exp[-i(\phi(\omega_0) - \omega_0 \tau)]$. A large enough value of $\tau$ allows the filtering of one of these interferometric components, which, when Fourier transformed back to the frequency domain, gives $\phi(\omega_0) - \omega_0 \tau$ as its argument. This allows the complete measurement of the transfer function of the optical element, and therefore of its dispersion, using a single delay. It is experimentally much simpler than the previously described methods; when only orders higher than the quadratic spectral phase (or GDD) are needed, there is no need to stabilize the delay or even to know its value, as any linear term can be subtracted from the retrieved spectral phase. Because of the
C. Measurement of dispersion of a laser resonator

The intracavity dispersion of a laser is important in determining the duration of pulses that the laser can generate. The reason for this is that the coherence of the oscillating modes is established by periodic loss or phase modulation. Unless the periodicity of the modulation equals the inverse mode spacing, the number of modes that can be locked will be small. Dispersion causes a nonuniform mode spacing, thereby restricting the bandwidth of the laser and the pulse duration.

One can imagine that it is difficult to estimate the intracavity dispersion by adding the dispersion of each of the optical elements therein, since the position of the laser spatial mode in the entrance pupil of each may vary with the details of the cavity alignment. What is needed is a way to measure the dispersion of the cavity elements in the configuration that they are in when the laser is operating. Two techniques have been developed with this in mind.

The first of these uses a variant of the interferometric schemes outlined previously.31 The major modification is that the light source is placed inside the laser cavity, the output of which is directed to a balanced Michelson interferometer, as shown in Fig. 13(a). The light source is often the gain medium of the laser itself, pumped to just below threshold. The light leaving the cavity consists of periodic noise bursts, separated by the round-trip time of the cavity. Each burst has a slightly different temporal shape than the one preceding it, because it has taken one extra round trip through the dispersive cavity than the previous burst. If the arms of the Michelson are therefore adjusted to provide a delay that is equal to $m$ round trips of the laser cavity, it is possible to measure the first-order correlation between one burst and the same burst after $m$ round trips, in order to get their spectral phase difference, and therefore the dispersion of the cavity. This is done by scanning the reference arm, as in the Naganuma scheme. In Fig. 13(b), the first-order correlation of a burst with itself $S_0(\tau)$ and with the next pulse (i.e., $m = 1$) $S_1(\tau)$ are plotted. As these two functions are different, one can predict that significant dispersion has been introduced by one round trip.

A second method, illustrated in Fig. 14(a), actually measures the intracavity dispersion when the laser is operating.32 The output of the laser, when mode locked, is a train of pulses separated by a repetition period that is the cavity round-trip time at the mean wavelength of the pulse train...
spectrum. This is determined both by the physical length of the cavity and the dispersion of the intracavity elements—each adds a certain group delay to the pulse as it bounces back and forth in the resonator. Thus a measurement of the repetition rate of the laser as a function of frequency measures the change in group delay as a function of frequency, which is just the GDD [Fig. 14(b)]. The scheme devised by Knox does exactly this. The mode-locked pulse train is incident on a fast photodetector, whose output is sent to a frequency counter. The repetition frequency is recorded as a function of the tuning of the laser, and the GDD map built up directly. This method requires that the laser be tunable while remaining mode locked. This implies that the entire bandwidth of the gain medium not be mode locked while the measurement is made, since tuning is ipso facto impossible in such a circumstance.

A remarkable result of measurements of cavity dispersion using the Knox method is that the measured GDD is very close to that of the constituent bulk elements in the cavity. This is surprising for the following reason: The cavity mode spacing in frequency is not constant when the cavity contains dispersive optical elements. The superposition principle allows that when a coherent superposition of such unequally spaced modes is effected, the result will not be a steady train of identical pulses, but rather a modulated train. Such an output is not seen from a well-mode-locked laser, for good reason: The process of mode locking is both nonstationary and frequently nonlinear. The overall effect of these processes is to force the modes to be equally spaced, as was experimentally checked.\(^ {33}\) It is not obvious that such an effect will yield a local mode spacing that is so directly related to the local dispersion. Nonetheless, extensive experimental studies have shown that this appears to be the case. Therefore it would seem that it is only necessary to measure the dispersion of the intracavity elements individually in isolation, and to simply combine them to find the dispersion of the entire resonator, and this is the method used most often at present.

IV. DISPERSION IN ULTRAFAST LINEAR OPTICS
A. General considerations

In order to make effective use of short optical pulses it is necessary to be able to deliver them undistorted to an experiment. For pulses with durations of less than about 50 fs, care must be taken to ensure that there is no detrimental effect due to uncompensated dispersion between the laser and the target.

The most common form of dispersion between the laser and the experiment is that of materials. Since this is most often positive, a system with negative dispersion is needed for compensation. The most useful systems for this purpose are the double-pass grating or prism delay lines, each of which provides adjustable GDD (with associated variations of higher-order spectral phase terms). A diagram of such a setup is shown in Fig. 15. The output of a laser system which produces, say, transform-limited, 24 fs Gaussian pulses is directed toward a target experiment by an optical system which contains 4 cm of BK7 glass. Without compensation, the material dispersion in this glass would cause the pulse on target to be stretched to roughly 60 fs. In order to optimize the pulse width at the target, a prism delay line is inserted at the output of the laser system. This delay line introduces negative dispersion, which cancels the dispersion introduced by the optical system. The method of adjustment is pragmatic. A device for measuring the pulse duration is set up as close to the experimental location as possible, and the pulse duration is monitored as a function of the setting of the delay line. The prism or grating separation is adjusted until the pulse duration is minimized.

B. Dispersion in the measurement device

Several issues are important in an operation of this kind. The measuring instrument should not introduce any dispersion itself, and should be able to provide a reliable value for the pulse-quality measurement (in this example the duration) as rapidly as possible. For pulses from a mode-locked laser oscillator, which have a repetition rate of several tens of MHz, this is easy to do, especially if the pulse durations are above 50 fs. In this regime the dispersion of most standard measurement instruments is sufficiently small that the measured pulse duration is quite accurate, and the update rate of the measurement is high enough that real-time feedback is possible. In fact, it is possible to automate this procedure entirely, so that adaptive control of the dispersion is possible.

For pulses with durations of 20 fs or less, the spectral bandwidth is large enough that care must be taken to remove all dispersion in the measuring device, otherwise the resulting measurement will be inaccurate. All instruments for characterizing ultrashort optical pulses make use of two replicas of the input pulse that are mixed with each other or with another field in a nonlinear mixing process.

An example of such a technique is the intensity autocorrelator, in which two replicas of the input are mixed in a material with a second-order nonlinear susceptibility in order to generate radiation at the second-harmonic frequency. The second-harmonic power is larger when the two pulses overlap in time, due to an interference effect, compared to when they do not. Therefore the instrument maps out the intensity autocorrelation function of the input pulse as the delay between the replicas is varied.
The autocorrelator capable of extracting this function is not of the input pulse. Therefore any discrepancies between the two arms of the Michelson interferometer remain replicas of one another, since it is only in this situation that any instrumental phase can be removed after the measurement.

These methods have been applied to the characterization of ultrashort pulse shapes at the location of the experiment. A particularly interesting example of this is the measurement of a femtosecond optical pulse at the focus of a microscope used for confocal imaging applications. In this case the complete dispersion of the imaging system was quite complicated, so a direct measurement of the pulse was important in order to establish the temporal resolution of the experimental apparatus. In the measurement, which was done using the FROG technique, the sample at the focus of the microscope objective was replaced with a nonlinear crystal, so that the pulse shape at precisely the right location could be determined.

C. Optimization of a short optical pulse

Another situation in which such methods are useful is in the characterization of pulse propagation. There are both linear and nonlinear aspects to this problem, and it is impossible to discuss them in their entirety here. A single example will suffice to make the point. In telecommunications systems using time-division multiplexing of the data, the bit rate is limited by the dispersive properties of the optical fiber along which the pulses propagate. If the dispersion is too large, the pulses will lengthen in time, and eventually overlap one another, making it impossible to extract the bit pattern from the measured output power. One method to overcome the positive dispersion of common single-mode fibers at telecommunications wavelengths (in the near infrared) is to compensate it using a piece of fiber engineered (using the techniques described in Sec. II) to have negative GDD at the appropriate wavelength. The concatenation of sections of this fiber with regular fiber leads to a net zero dispersion across the entire link (of course, if the pulse intensity is large enough to give rise to nonlinear effects, then such cancellation is not possible). An experiment by Chang and Weiner demonstrates that such compensation is possible in the linear regime. A diagram of their setup is shown in Fig. 16(a). The stretched output of a 62 fs fiber laser is injected into a 2 km fiber. Without compensation, the dispersion of this fiber would cause the pulse duration to increase...
from about 250 fs to over 10 ps. By inserting a 0.5 km dispersion-compensating fiber, the output pulse width of the overall system could be reduced to about 500 fs. Figure 16(b) shows a comparison of the input and output pulses for this system. The fact that the output pulse is not completely recompressed to its original duration is indicative of the fact that the two fibers have different amounts of cubic spectral phase. Other techniques for compensating fiber dispersion have also been demonstrated.43,44

Once the pulse duration or shape at the target is known, then its shape can be altered by introducing the appropriate amount of dispersion or attenuation. Many techniques have been developed for these pulse-shaping applications.45,46 A common and practical way to do this is to make use of a Fourier-plane pulse shaper. This device uses a dispersive delay line of the sort shown in Fig. 8, set for net zero dispersion (i.e., with the two images of the grating superimposed on one another). At the mutual focal plane of the two lenses in this device, spectral components are spatially dispersed and focused.47 Spatial phase modulation in this plane then transfers into spectral phase modulation. Such an arrangement can then be considered as a programmable device inducing an arbitrary dispersion. To date, the most frequently used devices for adjusting the group delay in this manner have been a multipixel liquid-crystal array46 and a broadband acousto-optic modulator,48 but alternative solutions have also been used.49–51 Details of the operation of these devices are beyond the scope of this article. For our purposes it suffices to note that the spatial phase at any given location in the image plane of the delay line can be changed by means of a user-controlled signal. Driving this signal then allows the quas arbitary control of the induced spectral phase.

Another technique, which does not rely on a spatial modulation at the Fourier plane of a zero dispersion line, uses a collinear interaction between the optical pulse and an acoustic wave in a birefringent acousto-optic crystal.52 Proper phase matching allows the diffraction of spectral components of the input pulse polarized along one of the axes into the output pulse polarized along the other axis. As the optical indices along the two axes are different, the group delay for the diffracted frequency propagating in the crystal depends on the position where diffraction occurs. Properly driving the acoustic wave then allows one to shape the group delay in the output pulse.

The addition of a specific spectral phase to each frequency component of the light individually means that pulses of nearly arbitrary shapes can be generated. The particular spectral phase that must be applied to the input pulse is derived from the measured pulse shape (or other quantity) by defining an error signal using either local- or quasiglobal optimization algorithms.

Such schemes are most useful to the control of coherent optical processes, where not only the temporal intensity shapes of the pulse, but also its phase are important. Examples of optimization applications of this type include: compression of ultrashort pulses,53–55 maximization of two-photon absorption56 or particular channels of photodissociation,57,58 strong-response excitation of a molecular electronic–vibrational transition,59 and weak-field control of the quantum state of a Rydberg electron.60

D. Focusing of a short optical pulse

If the full variation of the field with transverse coordinates is accounted for, another important effect of dispersion is apparent. When a short pulse is focused by a lens, portions of the wave front at the edges of the lens aperture experience a different thickness of material than the portion at the center of the aperture. It is this variation in thickness that gives rise to the lenticular shape of the element and hence to its focusing power. (Although the full thickness of the material is not needed, hence the ability of Fresnel lenses to focus light, it is the most convenient way to make a lens.) A corollary to this is that the portions of the beam at the edges of the aperture also see less dispersion, so that they have a shorter group delay. This means that they arrive at the focus of the lens before the center of the beam, so that the “group front” of the beam (the surface of constant group delay) will no longer be a spherical wave centered on the focus. In other words, if one takes into account all the converging rays, the apparent duration of the pulse at this point can be dramatically larger than the effective duration of each ray.

This phenomenon was first pointed out by Bor, using simple ray tracing calculations.61 The radius-dependent part of the propagation time of a ray from a plane before any kind of focusing system to the focal point F (Fig. 17) can be calculated as

$$\Delta T(r) = -\frac{r^2}{2cf^2}\lambda \frac{df}{d\lambda},$$  \hspace{1cm} (30)$$

where $r$ is the input radius of the ray, $f$ is the focal length at the mean wavelength $\lambda$ of the short pulse, and $df/d\lambda$ is the variation of the focal length with the wavelength.

For a singlet lens made of a material of index $n(\lambda)$, one has $df/d\lambda = f[(n-1)(dn/d\lambda)]$. For example, a BK 7 lens of focal length 10 cm will induce a delay difference equal to 50 fs between the paraxial ray and a ray going through the lens at 1 cm from the axis. This is far from negligible as pulses of shorter durations are now widely used in experiments.

For an achromat optimized for the mean wavelength of the short pulse, the variation of the focal length with the wavelength is zero, so the group delay is, in this approximation, independent of the radius. This means that all pulses from any point of the aperture will arrive at the same time at the focus. Unfortunately, most doublets are optimized in the visible, and have a nonzero value of $df/d\lambda$ at 800 nm. For example, the Melles Griot doublet 01LAO123 is optimized for minimal chromatic aberrations in the visible. It yields a focal length of 100 mm and a value of $df/d\lambda$ equal to zero at 546.1 nm, but this quantity is roughly equal to $1 \text{ mm}/\mu\text{m}$ at 800 nm. The inferred delay difference on the two rays considered previously is 20 fs, i.e., far from negligible.

The group velocity dispersion also becomes a function of the radius because the rays travel through different thicknesses of a dispersive material (the lens itself). This effect,
a common layout for this class of lasers. They consist of a continuously pumped gain medium which has a very large gain bandwidth. The laser also has a mechanism for modulating the intracavity loss or phase. This mechanism may be either active (controlled by an external signal) or passive (controlled by the intracavity intensity itself). In either case, its function is to cause the phases of the cavity modes that are above threshold to be locked, that is, to have the same phase (or for adjacent modes to differ by the same constant phase). In the case of homogeneously broadened gain media, the mode-locking mechanism must also couple power between the modes.

When the modes are completely phase locked, the field inside the laser cavity consists of a single pulse, whose duration is equal to the inverse of the total mode-locked bandwidth (this is nominally equal to the bandwidth of the gain medium). The output of the laser is a train of short pulses, spaced in time at the round-trip time of the cavity. A pulse is added to the train every time the intracavity pulse is partially transmitted through the output coupler.

Dispersion plays an important role in such lasers because it causes the mode spacing to become uneven in frequency—that is the mode spacing at line center (the frequency at which the gain is maximum) is different from that away from line center. The frequencies at which the cavity supports stable Gaussian modes are given by the solutions to

\[ \phi(\omega_n) = 2n\pi, \]

where \( \phi(\omega) \) is the spectral phase accumulated by a wave of frequency \( \omega \) during a single round trip of the cavity. There are several sources of phase accumulation. That due to material propagation is simply

\[ \phi(\omega) = \frac{\omega n(\omega)L}{c}, \]

where \( L \) is the physical path length of material traversed, and \( n(\omega) \) the refractive index at frequency \( \omega \). \( L \) may also vary with frequency because of geometrical dispersion, as described in Sec. II. A second source of dispersion is due to spectrally dependent mirror reflectivities. Mirrors are typically chosen to have very flat phase transfer functions, such as is the case with metallic mirrors in the visible and near infrared, or are engineered to have very specific phase transfer functions, using sophisticated thin-film coating methods that produce dispersion via interference. Another phase offset arises from the Gouy phase shift associated with a focus.

If the modes are equally spaced in frequency, that is, the solutions \( \{\omega_n\} \) to Eq. (31) satisfy \( \omega_n - \omega_{n+1} = \Delta \omega \) for all \( n \) such that \( \omega_n \) is within the operating spectrum of the laser, then the repetition period for the laser will be \( T_{\text{rep}} = 2\pi/\Delta \omega \). If the spacing is not equal, however, it will be impossible for a periodic modulation of the cavity round-trip loss or phase to effectively lock the phases of all of the modes across the gain bandwidth.

It is clear from Eq. (31) that the uniformity of the mode spacing depends on whether \( \phi(\omega) \) is a linear function of frequency. Any deviations from linearity cause the mode spacing to become unequal. In the limits where the spectral phase proportional to \( d^2n/d\lambda^2 \) for a singlet lens, tends to modify the temporal shape of the pulse traveling on each ray. It is usually smaller than the previous effect.

These results obtained from geometrical optics have been confirmed using simulations based on a wave optical description of the problem, including the effect of spherical aberration. The manipulation of a short pulse with conventional optics is thus not an easy task, and other techniques to counteract the effects of dispersion have been studied.

The distortion of the group front of a short pulse upon propagation through a lens has been measured by Radzewicz et al. using an interferometric technique. Their method can be thought of as an extension of the method of Li et al. for measuring GDD, except now in a system where the dispersion varies within the system exit pupil. Therefore measuring the location within the pupil of fringes arising from interference between the focused and reference pulses as a function of the delay between the two, as in Fig. 12, allows one to map out the group delay across the pupil. In all cases the duration of the pulse at the target is increased, and the peak intensity is reduced over that expected from a simple calculation.

V. DISPERSION IN MODE-LOCKED LASERS

A. Introduction

The most common method of generating ultrashort optical pulses is from a cw mode-locked laser. Figure 18 shows
is a slow function of frequency, and the mode spacing is very small compared to the bandwidth of the mode-locked pulse train, the mode spacing is approximately

$$\Delta \omega = \frac{2\pi}{\delta \phi^{(1)}(\omega_n)},$$

where $\delta \phi^{(1)}(\omega_n)$ is the group delay at mode frequency $\omega_n$. As a consequence of this fact, a large component of the design of mode-locked oscillators is to find configurations of the cavity in which the dispersion is very close to linear across the widest possible range of frequencies.

The first consideration is usually to make the gain medium as short as possible, since this is the main source of positive dispersion in the laser (because of the nonresonant interaction of the laser light with the host medium of the active ions). The limit to the thinness is usually set by the doping level of active ions that can be sustained without compromising either the gain lifetime or the stimulated emission cross section through quenching of the excited state population via nonradiative transitions.

The next step is to arrange for sufficient negative dispersion, usually in a way that can be adjusted in situ. Whether one chooses the dispersion to exactly cancel that of the gain medium or not depends on the details of the mode-locking mechanism.

The shortest pulses are generated using passive mode locking. That is, it is self-action of the pulse inside the laser cavity which causes the time-dependent amplitude and phase modulation that are necessary to lock the mode phases. The most effective of such mechanisms is self-amplitude modulation. This is often implemented using a nonresonant nonlinearity (in which the phase accumulated by the pulse depends on its intensity) together with a linear mechanism that converts phase to amplitude, such as polarization rotation or spatial filtering.

**B. Kerr lens mode locking**

An important example is provided by the Kerr-lens mode-locking mechanism that is the basis of sustained mode-locked operation of cw Ti:sapphire lasers. In this technique, the amplitude of a short pulse is modulated in such a way that the intense pulses experience less loss than the weaker pulses. The passage of the pulse through the non-linear medium (in the case of the Ti:sapphire laser this is usually the sapphire host itself) causes the beam to self-focus. The pulse will pass through a subsequent aperture only if it was intense enough to be focused to a size smaller than the aperture. If not, it will experience loss. Since the pulse energy is fixed by the saturation intensity of the gain medium and the repetition rate, then shorter pulses experience lower losses. The laser thus finds stable operation with the shortest pulses that can be sustained due to other constraints, such as dispersion.

In the case of Kerr-lens mode locking, the nonlinear self-action also causes a intensity- and therefore time-dependent modulation of the pulse phase. When averaged over the pulse train, this is equivalent to the intermodal transfer of power, a necessary condition for mode locking. Nonetheless, the particular way in which the time dependence of the phase occurs (rising intensity shifts frequencies to the red) means that the dispersion must be adjusted in the right way to properly compress the pulse. Although this is not possible exactly, because the frequency shifts depend on the detail on the pulse intensity, it is generally the case that the phase shift near the peak of the pulse is a linear function of time. This requires linear negative group delay dispersion, so that the blueshifted frequencies and redshifted frequencies are resynchronized, leading to a shorter duration pulse. In some cases, it is possible to operate the laser in a solitonlike regime, in which the combination of the nonlinear temporal phase shift and quadratic spectral phase interact to form solitons.

Thus the usual situation in a Ti:sapphire laser is that the intracavity dispersion is adjusted to be somewhat negative, but still very close to linear. In order to do this, geometrical dispersion is employed to provide adjustable negative linear dispersion, and interferometric or material dispersion is used to flatten the remaining higher-order phase. Since both systems are used inside the laser, it is vital that they have extremely low loss. For this reason a two-prism delay line is used to provide the adjustable negative dispersion. The prisms are cut so that the light is incident at Brewster’s angle at the front surface, and they are used at minimum dispersion so that the incident and refracted angles are nearly the same for all wavelengths. This means that the insertion loss can be very small, only a few percent. However, it is the case that materials with low dispersion will give low angular dispersion. Therefore the separation of prisms must be larger for lower dispersive materials in order to obtain the same quadratic spectral phase. This means that the laser cavity can become unwieldy if the GDD of the gain medium is too large. Alternatively one can use highly dispersive material
for the prisms, and reduce the cavity to a more practical length. The cost of this is that the prism delay line introduces higher-order spectral phase which must also be compensated, along with higher-order material dispersion. Nonetheless, designs of prism delay lines with reduced higher-order dispersion for a given quadratic dispersion have been developed and used successfully in mode-locked lasers. As an example, Lemoff and Barty showed that careful selection of prism material allows one to compensate both second- and third-order dispersion in the laser cavity. The prism separation is chosen to balance the second-order dispersion of the laser rod. Then, as illustrated in Fig. 19 for a 1 cm Ti:sapphire rod, given a particular center wavelength, the optimal prism material is chosen to compensate the third-order dispersion of the rod.

This strategy has been a successful one for the minimization of the pulse durations from Ti:sapphire lasers. It was evident from work with mode-locked dye lasers that compensation of the quadratic spectral phase (the linear GDD) was insufficient to generate pulses with gain bandwidth limited durations. The strategy adopted in Ti:sapphire lasers has been to make the material dispersion as small as possible, by making the gain medium as short as possible. This means that the prism delay line need only compensate a small amount of quadratic dispersion, which can be done without introducing significant higher-order dispersion.

Another approach has been to design thin film coatings to provide both a high reflectivity across a broad bandwidth and a specific phase transfer function. This process is rather complicated and requires mirrors with several tens of layers. Usually there is a residual fourth-, fifth-, or sixth-order phase that leads to a small periodic modulation of the phase across the pulse spectrum. That remaining can be compensated using so-called “chirped” mirrors. Figure 20(a) illustrates the concept of a chirped mirror. They are carefully fabricated so that the effective penetration depth for each wavelength can be controlled. This allows one to tailor the dispersion introduced by each mirror. Figure 20(b) shows the standing-wave electric field patterns in a double-chirped mirror as a function of wavelength. The use of these mirrors not only leads to the simplest laser cavity with the fewest optical elements, but also allows one to produce pulses in the region of 6 fs. Measurements of the amplitude and phase of these pulses from a variety of laser systems have shown that the durations appear to be limited by the residual higher-order phase in the laser cavity (that is, spectral phase function is not constant across the pulse spectrum, which is the criterion for generating the pulse with the shortest rms duration from any given spectral amplitude).

C. Fiber lasers

The design of short-pulse fiber lasers is somewhat different. The cavity consists of several lengths of fiber connected together, with a mode locker inserted between one pair of sections. In this case, the transverse mode size in the fiber is fixed, so that Kerr-lens mode locking cannot be used. Instead, a nonlinear rotation of the laser polarization is implemented using the same physical mechanism—the Kerr effect—but now to induce a cross-phase modulation between the two polarization modes of the fiber. The polarization of the light in the fiber is therefore rotated in a way that depends on its intensity, so that a polarizing element will transmit a shorter pulse than the input pulse.

A further consideration for fiber lasers, however, is that the pulse inside the laser cavity cannot be too short, otherwise the power will be limited by uncompensable time-dependent nonlinear phase shifts, or by optical damage. Note that the mode size in the fiber is a few microns, and the fiber lengths much longer than the material lengths used in bulk lasers. This means that pulse trains with even modest average intensities can have peak intensities that will induce several radians or more of phase shift. On the other hand, the intracavity laser intensity should be comparable to the saturation intensity of the gain medium if the laser is to operate efficiently. The solution to this dilemma is to arrange the cavity dispersion so that the pulse inside the cavity is chirped, and consequently stretched in time. The duration is long enough, then the time-averaged laser intensity is large enough to saturate the gain medium without inducing nonlinear spectral phase shifts that cannot be compensated outside the laser. A short length of fiber with the opposite sign of dispersion to that used inside the laser can be made to compress the pulse outside the laser. Figure 21 shows a mode-locked fiber laser which uses this technique to produce subpicosecond pulses.

A second mode of operation is also possible in such a configuration. Whereas in the mode of operation just described amplitude modulation by nonlinear polarization rotation plays the dominant role in mode locking, it is possible to carefully balance the nonlinear phase shifts accumulated by the stretched pulse by negative dispersion (in passive sections of the fiber) in such a way as to generate solitons. The amplitude modulation is then relegated to a secondary role of stabilizing the soliton propagation against the growth of dispersive waves from amplified spontaneous emission from the gain section of the fiber laser. In these lasers, the residual higher order spectral phase limits the pulse duration, and it was found that the minimum duration is not obtained for the broadest spectrum.

Dispersive-wave fiber lasers typically generate pulses in the 100 fs region, at wavelengths of 1.5 μm, comparable to modern telecommunications fibers. On the other hand, solitonic lasers produce pulses in the picosecond regime, at similar wavelengths.

Soliton lasers can be operated without amplitude stabilization. In this case they may be unstable with respect to the growth of dispersive waves. This instability can be used to measure the dispersion of the fiber cavity while the laser is in operation. When a soliton is scattered, it may shed energy to a dispersive wave. If the scattering is periodic, as it is inside a fiber laser cavity, then the dispersive waves from each period interfere with each other. For wavelengths where the round-trip phase accumulation is 2π, then dispersive scattered waves on successive round trips interfere constructively. Frequencies for which the phase accumulation is just π will interfere destructively, so that energy scattered from the soliton on one round trip will be rescattered back into the ...
soliton on the next. For this reason, as shown in Fig. 22, the spectrum of the output pulse train consists of a central broad peak, corresponding to a train of solitons, on which is superimposed a set of peaks whose locations correspond to the frequencies that satisfy the cavity mode condition Eq. (31). These are the so-called Kelly sidebands of the soliton, and their spacing provides a map of the cavity dispersion.\textsuperscript{81}

Finally, the simple pictures of the role of dispersion in ultrafast mode-locked lasers presented here is not the most complete. It turns out to be possible to operate these lasers stably in regions where the linear dispersion of the cavity is not zero, and the sign of the linear dispersion is positive, so that solitonlike pulse formation is not operative. In these cases there is a trade-off between the nonlinear amplitude and phase modulations and the intracavity dispersion that defies a simple explanation. However, in all cases the pulses in these nonsolitonic regimes are longer than in the cases where some positive nonlinear chirp is compensated by negative group-delay dispersion, so that practically speaking, such lasers are a curiosity.

VI. DISPERSION IN ULTRAFAST AMPLIFICATION

One of the major features of ultrashort optical pulses is that they have very high peak powers even for a small energy. For example, a 10 fs pulse from a 100 MHz mode-locked oscillator with 10 mW average power, has a peak power of 10 kW. This itself is sufficient peak power to use for nonlinear optics in certain applications. It is possible, however, to generate femtosecond pulses with peak powers in the Terawatt regime in rather compact systems, and, given good spatial beam quality, to focus them to intensities in the neighborhood of $10^{19}$ W/cm$^2$ or greater. This makes possible physics in a regime where the Coulombic binding force of the valence electron in an atom is smaller than the electric field of the laser. This highly nonperturbative regime of nonlinear optics has blossomed in the past few years because of the development of high energy, ultrashort pulses from small-scale laser systems.

The key development in ultrashort pulse amplifiers is the idea of chirped pulse amplification (CPA).\textsuperscript{82,83} Illustrated schematically in Fig. 23, CPA requires proper dispersion management in order to produce ultrashort, high energy pulses.\textsuperscript{84} In this technique, an ultrashort pulse is first temporally stretched by sending it through a dispersive delay line with very large (typically positive) GDD. This procedure lowers the peak intensity so that the pulse can then be amplified by several passes through a broadband gain medium. After amplification, the final stage of CPA is to recompress the pulse in a dispersive delay line with the opposite sign.
(typically negative) GDD to that of the stretcher. This technique is used to avoid uncompensable or uncontrollable nonlinear effects in the amplifier gain medium. At lower intensities, these nonlinear effects can appear as self-phase modulation, and at higher powers, as multiphoton ionization of the gain medium itself, which can lead to avalanche breakdown of the amplifier or other optical material and thus to permanent optical damage.

An optical amplifier operates most efficiently when the input fluence is comparable to the saturation fluence, $F_s$. In Ti:sapphire the small stimulated emission cross section $\sigma$ means that $F_s (= h \omega / \sigma)$ is rather large—about 1 J/cm$^2$. If one wished to amplify directly a 10 fs duration pulse to 1 mJ energy with high efficiency, the pulse at the output of the amplifier would have a peak intensity of more than $10^{15}$ W/cm$^2$, well above the damage threshold of common optical materials and coatings. However, stretching the pulse by a factor of 10$^3$ or so will reduce the peak intensity to the point where optical damage is quite improbable. The amplifier will still operate in the saturated regime provided the duration of the stretched pulse remains much shorter than the spontaneous lifetime of the gain. For Ti:sapphire this means a duration of less than 3 $\mu$s. A larger stretch factor allows one to operate closer to saturation while avoiding uncontrollable nonlinear effects. Recent works$^{85,86}$ have shown that up to 90% of the theoretical maximum quantum efficiency for a Ti:sapphire amplifier can be achieved by using stretch factors of 10$^5$.

Under these conditions, the effects of nonlinear self-action are also negligible. A common rule of thumb in high-power laser amplifier design is that the peak nonlinear phase shift should be less than 1 rad. Although nonlinear phase shifts may be treated like any other phase as far as dispersion compensation goes,$^{87}$ they often lead to a complicated spectral phase function for the output field which is usually not completely compensable by the sort of dispersive systems one can easily construct. As a consequence, the usual design strategy is to ensure that they are as small as possible.

The most common effect that causes nonlinear phase shifts is self-phase modulation. Every optical material possesses a nonlinear refractive index, labeled $n_2$, that produces a phase accumulation on a beam of intensity, $I$, passing through it. The nonlinear phase shift $\phi_{NL}(t)$ is

$$\phi_{NL}(t) = \frac{2 \pi \omega}{c} n_2 \int_0^L dz I(z,t),$$

where $L$ is the length of the crystal, and $n_2$ for sapphire equals $2.5 \times 10^{-18}$ cm$^2$/W. With an intensity of $10^{19}$ W/cm$^2$, corresponding to the peak of the stretched pulse, $\phi_{NL}(t) \approx 0.2$ rad. It is also possible to negate the effects of small amounts of self-phase modulation by introducing nonlinear media with an $n_2$ of opposite sign to that of the amplifier media.$^{88}$

The key attributes of the stretcher and compressor used in any CPA system are:

(i) (output system) high-power and high-energy handling capability,

(ii) adjustable spectral phase to some reasonable order, over some reasonable range.

These attributes affect not only the specifics of the design, but also the physical size of the systems. The more dispersion required and the larger pulse energies to be manipulated, the larger the optical elements must be.

As the labels below the diagram in Fig. 23 indicate, the ideal situation is that the compressor have exactly the opposite second-order dispersion as that of the combination of the stretcher and amplifier. The stretcher is usually chosen to have positive dispersion so that there is no danger that the pulse be compressed in the amplifier, which itself has predominantly positive dispersion due to the host lattice. In order to achieve the required stretch factors, GDDs on the order of $10^4 - 10^6$ fs$^2$ are required.

The largest dispersions for the shortest total path length are found in angularly dispersive systems. In particular, the double-grating arrangement analyzed by Treacy in 1969,$^{89}$ and discussed in Sec. II C, provides a prototype for the compressor portion of a CPA system. The throughput of the compressor is limited by the number of interfaces that the pulse traverses. For example, since there are four reflections off the
gratings, the typical transmission of this system is about 50%–70%. An important issue with this system is determining the size of the optics, in particular the gratings. The size of the first grating depends on the output pulse peak intensity, which can damage the optic severely via multiphoton ionization. The spatial extent of the dispersed amplified pulse’s spectrum determines the size of the second grating. The size of the dispersed spectrum on the second grating is proportional to the angular dispersion of the first grating $\partial \omega / \partial \omega\omega$, the pulse spectral bandwidth $\Delta \omega$, the perpendicular distance between the gratings $D$, and the angle of diffraction from the grating $\alpha(\omega)$.

$$S \sim \frac{\partial \alpha}{\partial \omega} \Delta \omega \cos(\alpha(\omega)).$$

Using Eqs. (20) and (21), Eq. (35) can be written as

$$S \propto \frac{\phi^{(2)} \Delta \omega d \cos(\alpha)}{\lambda},$$

where $d$ is groove spacing of the first grating.

As discussed previously one of the important parameters in the design of a CPA system is the duration of the amplified stretched pulse. For large dispersions, using Eq. (7), the duration of the stretched pulse can be related to the original pulse duration by $\tau_{str} \propto d^{(2)}/\tau_{orig}$, where for a transform-limited pulse, $\tau_{orig} \propto 1/\Delta \omega$. Thus, for a given stretched pulse duration and central wavelength, the size of the second grating is proportional to the grating groove separation and the diffraction angle,

$$S \propto \tau_{str} \frac{d \cos(\alpha)}{\lambda}.$$

Note that this expression does not include the finite size of the input beam.

The corresponding stretcher, matched to all orders of dispersion (in the limit of paraxial imaging) to the Treacy compressor, is the design of Martinez. The throughput of this system is typically reduced because of the larger number of optical elements, so its positioning before the amplifier is convenient in this regard also. The presence of monochromatic aberrations in the imaging system can cause a mismatch between the cubic and quartic spectral phase of the stretcher and compressor, which limits the fidelity of pulse recompression.

However, if one includes the dispersion introduced by the amplification process, it is then not necessarily desirable for the stretcher and compressor to be completely matched to all orders of dispersion. A typical 20-pass regenerative amplifier contains a total of about 1 m of optical material, and introduces GDD of about 39 000 fs$^3$, cubic phase of 38 000 fs$^3$, and quartic phase of approximately $-22 000$ fs$^4$. A multipass amplifier, having fewer passes and less optical material, introduces about half the spectral phase of a regenerative amplifier. In order to compensate for the additional quadratic phase, one can mismatch the stretcher and compressor by simply changing the distance between the two compressor gratings [D in Eq. (19)]. Figures 24(a) and 24(b) show simulations of a 50 fs Gaussian pulse that has been amplified in a regenerative amplifier using CPA. In this example, the compressor grating separation was adjusted so that the total second-order dispersion of the entire CPA system is zero. From Fig. 24, it is clear that merely adjusting the compressor grating separation is not sufficient to achieve a short amplified pulse. This is because a change in grating separation also introduces additional third- and fourth-order spectral phase. The sign of the additional cubic is the same as that of the third-order phase introduced by the amplifier. In addition, because optical materials typically have very low fourth-order dispersion, amplifiers contribute only very small amounts of quartic phase. Therefore, the central problem is this: If one adjusts the compressor to compensate for the additional GDD of the amplifier, then excess higher order phase is introduced which severely limits the fidelity of the recompressed, amplified pulse.

In many cases, by adjusting the compressor grating incidence angle and separation, one can minimize both the second- and third-order dispersion of the CPA system. The resulting amplified pulse (for a 50 fs Gaussian input pulse) is shown in Figs. 24(c) and 24(d). Although the recompression is much improved, the amplified pulse is still not transform limited. This is due to uncompensated higher order spectral phase. For desired amplified pulse widths below 50 fs, one must also completely compensate fourth-order dispersion.

There have been many other solutions to the problem of higher order phase compensation in CPA systems. One is to use a separate optical system to compensate the higher-order terms. For example, the cubic dispersion for prism pairs is of the opposite sign to that of gratings pairs, even when both have the same sign of GDD, so that these elements together can be used to adjust both second and third-order dispersion. This idea was first applied to pulse compression in generation of 6 fs pulses from a dye laser system. This concept has been extended by Kane and Squier to modern CPA systems by using a monolithic grating and prism combination (with the gratings written on the surface of the prisms) to compensate cubic phase. In fact, a compressor of this design allows one to use a much simpler stretcher—a piece of optical fiber—and still obtain good recompression for pulses on the order of 100 fs.

A different approach to this problem is to use the stretcher to introduce the compensating higher order phase terms, rather than the compressor. This is somewhat easier because the larger number of elements in the system offers more degrees of freedom. As an example, White et al. used an air spaced doublet for the imaging lens. This allowed the lateral and longitudinal positions of the lens elements to be adjusted to accommodate various amounts of cubic and quartic phase. They therefore modify the spectral phase by deliberately introducing several monochromatic aberrations due
to decentering and defocus (such as on-axis astigmatism).

Additional methods for higher order phase compensation include placing an additional optical element between the two gratings of the compressor, which has been shown to compensate primarily cubic phase, and using gratings with spatially nonuniform groove spacing which could, in specific configurations, be used to induce purely linear group delay dispersion of positive or negative sign.

The most comprehensive approach to the design of stretchers and compressors for ultrashort pulse CPA systems is that of Lemoff and Barty. They deliberately set out to configure the grating angles, separations, and the aberrations in the stretcher in just the right way to provide exactly the correct amounts of cubic and quartic phases. The penalty is an added complexity in alignment because of the greater number of optical elements, but this technique has been used on a CPA system producing sub-20 fs, 100 TW pulses.

In previous examples, the fact that aberrations are present in the imaging system of the stretcher was used to compensate higher order dispersion. However there are advantages to aberration-free stretcher designs. It is well known that imaging systems whose surfaces have a common center of curvature have no monochromatic aberrations to all orders. An important system of this sort for these purposes is the Offner triplet stretcher. This is an all-reflective design in which the imaging system consists of a convex and a concave mirror used in an off-axis configuration with object and image planes mutually parallel and both containing the center of curvature of the mirrors, as shown in Fig. 25. The complete absence of aberrations in this design allowed Du et al. to stretch a 30 fs pulse by a factor of $10^4$ and to recompress it completely. An important practical feature of this design is that it is easy to align, and rather tolerant of slight misalignments. This design has been used to produce 30 TW, sub-30 fs pulses from a multipass amplifier CPA system.

While the fact that the Offner triplet stretcher is aberration free means that one can stretch and compress a pulse with excellent fidelity, it also means that one cannot use the aberrations of a standard stretcher to introduce fourth-order dispersion. If the amplifier of a CPA system requires a lot of material or additional passes through the gain medium (as in a regenerative amplifier), then when one adjusts the compressor to compensate for the extra quadratic and cubic phase, there is no way to compensate the additional fourth-order dispersion. A solution to this problem, introduced by Squier and Kane, is to use a compressor whose gratings have a different groove spacing than that of the Offner triplet stretcher. In fact, it works out that one can use a higher groove density grating in the compressor, leading to better diffraction efficiency, and still compensate up to fourth-order phase. This technique has been used to produce 20 mJ, 20 fs pulses at a 1 kHz repetition rate using a CPA system with a regenerative preamplifier. Because of their simplicity and flexibility, these solutions have been implemented in many commercial CPA systems.

As these systems have a limited number of parameters, it is not possible to exactly get a flat spectral phase over the complete bandwidth of the output pulse from a CPA system, which would lead to the optimal pulse in terms of duration. For example, the exact cancellation of the second- and third-order dispersion for the output pulse might leave enough fourth-order phase to deteriorate the quality of the pulse. This then calls for a global theoretical optimization of the parameters of the system (e.g., parameters of the stretcher and compressor), taking as a target a merit function for the output pulse (e.g., its duration). Also, this procedure allows one to quantify the sensitivity of the optimal set of
parameters regarding slight misalignments, which is an important practical point for the experimental implementation.

It must be noted that in all these systems, the surface quality of the optical elements plays a key role: Wherever the optical spectrum is spatially dispersed, any spatial phase feature will induce a spectral phase modulation. This effect can easily be experimentally demonstrated, e.g., when measuring the dispersion of a stretcher set in a zero dispersion position. These high order or periodiclike induced features usually cannot be compensated by the plain action of the compressor. They have been shown as detrimental to the pulse quality at the output of CPA systems.\textsuperscript{102}

Also, the use of pulse shaping devices, which are to a great extent capable of an arbitrary control of the spectral phase, on CPA systems allows one to compensate the residual spectral phase of the output pulse. These devices are placed before the amplifier in order to prevent damage to the optical modulator (for high-resolution control of the spectral phase, the spectral components need to be focused on to the modulator). This technique should be able to correct uncompensated spectral phase, either originating from theoretical mismatches in the complete CPA system or from experimental mismatches (quality of the optical elements or day-to-day misalignments).

The ability to stretch a short pulse, amplify it without detrimental nonlinear effects, and recompress the high energy pulse to a short high-power pulse with a good contrast is now routine using the optical systems presented previously. Increasing the number of parameters of one of these systems, or the number of independent systems, allows a better control of the pulse quality at the output of the CPA system. But this also usually brings an increased complexity or an overall low energy transmission. In this sense, a pulse shaper is perhaps the most radical solution, because a very large number of parameters can be independently set. Whatever the diagnostic used at the end of the system, these versatile solutions for spectral phase control, along with good quality conventional stretchers and compressors, seem to be a reasonable direction for the control of dispersion in future CPA sources.

\section{VII. Dispersion in Ultrafast Nonlinear Optics}

\subsection{A. Interaction of short pulses}

The role of dispersion in nonlinear optics takes several forms, mainly relating to over what durations and path lengths the nonlinear interaction between pulses can be maintained. Consider, for example, the frequency conversion of an intense optical pulse by its interaction with a weaker pulse of different frequency in a nonlinear medium, as shown in Fig. 26. The efficiency of the transfer of energy from the stronger to the weaker pulse depends on a large number of parameters: the length of the nonlinear medium, whether the process involves phase matching, the different group velocities of the two pulses; and the degree to which each pulse is temporally broadened due to quadratic spectral phase accumulation as it propagates through the medium.

These parameters affect the nonlinear interaction in the following ways. The most efficient interaction occurs with the longest interaction path length, so the medium itself should be as long as possible. On the other hand, the actual useable length of the medium is set by the other parameters. Assuming that phase matching can be achieved (if it is necessary), then a difference in the group velocity of the two pulses causes them to walk away from one another, so that they no longer interact.

The time taken for a pulse propagating at the velocity \(v_i\) to traverse a distance \(L_e\) is

\[ \tau_i = \frac{L_e}{v_i}; \quad i \in (1,2). \]  \hspace{1cm} (38)

The difference in traversal times is \(\tau_m\), where

\[ \tau_m = \begin{vmatrix} \frac{1}{v_1} - \frac{1}{v_2} \end{vmatrix}. \]  \hspace{1cm} (39)

If this time \(\tau_m\) is less than the input pulse duration \(\tau\), then the pulses will still overlap at the output of the medium. So the effective interaction length is then approximately

\[ L_e = \frac{\Delta v^2}{v^2 \Delta v}, \]  \hspace{1cm} (40)

where \(\Delta v\) is the difference in group velocity and \(v\) is the geometrical mean group velocity of the two waves. The actual interaction length is therefore \(L_{\text{act}} = \min(L_e, L)\), where \(L\) is the physical length of the medium. Group velocity dispersion (or GDD) affects the interaction in another way. Non-zero GDD causes the pulse to be temporally stretched. Therefore its peak intensity will decrease, and the nonlinear interaction will become much less efficient. Moreover, the presence of dispersion will cause the duration of the output pulse to be longer than that of the input, in general, for both reasons pertaining to the nonlinear interaction, as discussed here, and the linear propagation. It is therefore critical to understand the compromise between these various parameters that leads to the optimal nonlinear interaction.
The field of ultrafast nonlinear optics is too large to be effectively surveyed here, even if that were warranted by the subject matter. Therefore we shall choose a few examples to illustrate the ways in which the important parameters affect the nonlinear mixing of ultrashort optical pulses.

B. Phase-matched processes

All parametric processes require a phase-matched interaction. A simplified description of this will suffice for our purposes. The reader is referred to one of the many textbooks on nonlinear optics for a more rigorous description. Moreover we restrict ourselves to the two simplest situations, degenerate and nondegenerate three-wave mixing.

The nonlinear polarization source term generated by two waves of frequency \( \omega_1 \) and \( \omega_2 \) at position \( z \) in a medium with a \( \chi^{(2)} \) nonlinearity is (for sum frequency generation, in the weakly interacting limit)

\[
\tilde{P}(\omega,z) = \chi^{(2)}(\omega=\omega_1 + \omega_2; \omega_1, \omega_2) \tilde{E}(\omega_1; 0) \times e^{ik_{\omega_1}z} \tilde{E}(\omega_2, 0) e^{ik_{\omega_2}z},
\]

where \( \tilde{E}(\omega,0) \) is the field at the input face of the medium, \( z \) is the propagation distance, \( k(\omega) \) is the wave number of the field at frequency \( \omega \). The effectiveness of the nonlinear interaction is predicated on this term having the same spatiotemporal behavior as a field at the same frequency propagating through the medium. It is therefore common to factor out the propagation constant of the generated wave and work with the envelope of the field.

The field generated by the interaction of two short pulses is found by integrating this source polarization over the length of the nonlinear medium, and all input frequencies

\[
\tilde{E}(\omega,L) = \int d\omega_1 \chi^{(2)}(\omega; \omega_1, \omega - \omega_1) \tilde{E}(\omega_1; 0) \tilde{E}(\omega - \omega_1; 0) \times \int_0^L dz e^{i(k_{\omega_1} + k_{\omega - \omega_1} - k_{\omega})z},
\]

where we have taken into account both the conservation of energy, so that \( \omega_1 + \omega_2 = \omega \), and the propagation constant of the electric field at frequency \( \omega \). The integral over \( z \) on the right-hand side of Eq. (42) is called the phase-matching term, and contains information on the propagation parameters discussed previously.

The phase-matching term has the form

\[
\tilde{A}(L) = \int_0^L dz e^{-i\Delta k z} = L e^{i\Delta k L/2} \frac{\sin(\Delta k L/2)}{\Delta k L/2},
\]

where

\[
\Delta k(\omega, \omega_1) = k(\omega_1) + k(\omega - \omega_1) - k(\omega)
\]

is called the wave vector mismatch.

The maximum of the phase-matching function occurs when its argument is zero, so that the efficiency of the nonlinear process is highest when the wave vector mismatch is zero over the entire range of frequencies contained in the input pulse.

In order to understand the effects of dispersion on the nonlinear process, it is convenient to expand the wave vector mismatch about points near the center of the input and output spectra, say \( \omega_i \) and \( \omega_o \). Let the expansion frequency variables be \( \Delta = \omega_1 - \omega_i \) and \( \delta = \omega - \omega_o \), respectively. Then, taking into account energy conservation, the expansion of \( \Delta k \) to second order in frequency mismatches is

\[
\Delta k(\omega, \omega_1) = [k(\omega_i) + k(\omega_o - \omega_i) - k(\omega_o)]
\]

\[
+ \left[ \frac{\partial k}{\partial \omega} \right]_{\omega_i} - \left[ \frac{\partial k}{\partial \omega} \right]_{\omega_o} \delta + \left[ \frac{\partial^2 k}{\partial \omega^2} \right]_{\omega_i} - \left[ \frac{\partial^2 k}{\partial \omega^2} \right]_{\omega_o} \delta^2,
\]

where \( \partial k / \partial \omega \) is proportional to the group delay of the radiation at frequency \( \omega_o \), and \( \partial^2 k / \partial \omega^2 \) is the group delay dispersion.

It is now easy to see the role of dispersion in determining the efficiency of the nonlinear process. If the first term in square brackets on the right-hand side of Eq. (45) is set to zero, then the process is said to be phase matched for the frequencies \( \omega_i \) and \( \omega_o \). The phase mismatch is still nonzero when this is done, however, because in general the group delay and GDD mismatch terms (the second and third lines on the right-hand side of Eq. (45), respectively) are nonzero. It is clear that the process will be more efficient provided the group velocities of all three waves are the same. In the case where the group velocity of the input waves at the frequencies \( \omega_i \) and \( \omega_o - \omega_i \) are the same, the second term in the second line of Eq. (45) vanishes. This means the process will be phase matched over a broad range of frequencies.

Unfortunately, there is no such simple cancellation for the GDD terms. It is straightforward to show that the only way for the third line in Eq. (45) to vanish is for the group delay dispersions of all three frequencies to be zero. This implies that both input pulse and generated pulses do not broaden temporally as they propagate through the medium, and their peak intensity does not therefore diminish.

A common method of adjusting the wave numbers of the interacting waves is to use a birefringent crystal, along the axes of which the refractive index differs. An example of this is shown schematically in Fig. 27. Figure 27(a) indicates the symmetry axis \( c \) of a uniaxial crystal, and the direction of the wave vector of the light \( k(\omega_i) \). The axes orthogonal to the wave vector are the two different polarizations that experi-
ence the largest index difference. The polarization lying in the plane containing the vector $\mathbf{e}$ and $k(\omega_i)$ is called the extraordinary wave and the polarization orthogonal to this plane is the ordinary wave. The refractive index of the extraordinary wave can be varied by changing the angle $\sigma$ between the $c$ axis and the wave vector. The dependence of the refractive index on frequency in this configuration is shown in Fig. 27(b). The two curves represent the ordinary and extraordinary wave refractive indices, the latter curve going smoothly into the former as the angle $\sigma$ is changed. The dashed lines represent the phase-matched condition for second-harmonic generation in a type-I geometry. In this case $\omega_o=2\omega_i$, and it can be seen from Eq. (44) that this implies $k(\omega_o)=2k(\omega_i)$, or $n(\omega_o)=n(\omega_i)$.

The next step is to ensure that the group delays of the input and generated waves are equal. This can be done by careful choice of crystal parameters.\textsuperscript{103} Alternatively, the same effect can be achieved using the angular dispersion of the birefringent crystal to match the two velocities. For example, the method of Radzewicz et al.\textsuperscript{104} sets two input waves at different angles. The projection of the two wave vectors onto a vector that bisects them is adjusted by changing the angle between them. The phase-matching condition can be set by changing the crystal orientation, so that there are two degrees of freedom. These can be used to simultaneously satisfy the phase- and group-velocity matching conditions.

A second approach to the problem of broadband phase matching is to angularly multiplex the input wave so that different frequencies are incident on the nonlinear crystal at just the right angle to satisfy the phase-matching condition. Then each frequency is doubled most efficiently, and the resulting second-harmonic radiation can be reassembled into a pulse by angularly demultiplexing. The mux–demux steps are implemented using two prisms and an imaging lens arrangement.\textsuperscript{105}

A more complicated situation arises in the case of parametric downconversion, in which a pump pulse of high frequency is converted into two pulses of lower frequency, called the signal and idler waves, as shown in Fig. 28. This process is used to generate ultrashort optical pulses in the near and midinfrared region of the spectrum. In this situation, it is necessary to match all three group velocities as closely as possible. For type-I downconversion in the degenerate limit (when the signal and idler have the same frequency, exactly half that of the pump) the bandwidth of the process is often very large, since the group velocity mismatch goes almost to zero. In fact, for certain crystals pumped near 400 nm, it is possible to get a phase-matched bandwidth of several hundred nanometers,\textsuperscript{106,107} and this has been used to generate tunable pulses of less than 10 fs duration from a seeded parametric amplifier.

This process also provides enough gain that it is possible to build a femtosecond parametric oscillator. A typical arrangement for such a device in shown in Fig. 29. It consists of a parametric downconverter pumped by a train of pulses from a model-locked oscillator. The signal pulses from the downconverter are resonated in an optical cavity, arranged so that the round-trip time for the signal pulses equals the period of the pump pulses. In this way the energy of the signal pulse builds up and brings the oscillator above threshold. An intriguing feature of this synchronous pumping is that the dispersion in the cavity causes tuning of the wavelength of the oscillator output. Recall that the round-trip period of the optical cavity is set by both its physical length and the dispersion of the intracavity material. If the GDD depends on wavelength, then the wavelength of the signal (and therefore the idler, which is not resonant with the cavity) changes so that the total round-trip time $\tau_i=(2L/c)+d^{(1)}(\omega_{signal})−d^{(1)}(\omega_{pump})$ is equal to the period $T$ of the pump laser pulse train. In this way, small variations in the pump period can be accommodated, or even used to tune the OPO. In fact, this mechanism of dispersive tuning can be developed into a means to stabilize the lengths of the pump and OPO cavities, by simply ensuring that the spectrum of the OPO remains at a fixed optical frequency.

The output signal pulse duration is usually limited by the GDD at the signal wavelength. This can be compensated in the same way as in lasers, so that such oscillators can generate pulses of 40 fs or less duration in the near infrared, at a wavelength of 1.3 $\mu$m or so.\textsuperscript{108}

For generating wavelengths in the midinfrared, at wavelengths in the range from 2 to 10 microns it is common to take the signal and idler outputs of a parametric downconversion source, and to generate their difference frequency. The optimal method for this depends greatly on the material and pump wavelengths available. Nonetheless, the arguments presented previously hold in this situation also. For example, pumping a lithium niobate (LiNbO$_3$) crystal at a wavelength of 767 nm to obtain idler wavelengths in the 2–5 $\mu$m region limits the effective interaction length to 1 mm or so, leading to low output power.\textsuperscript{108} On the other hand, a silver thio gallate (AgGaS$_2$) crystal pumped at 1–3 $\mu$m has all three waves with nearly the same group velocity in the frequency range 2–12 $\mu$m, (with them exactly equal at 9.5 $\mu$m) so that a long crystal can be used to obtain short pulses with high output powers.\textsuperscript{109} Plots of the relevant group delay mismatches for this particular process are shown in Fig. 30.

Up to this point we have discussed only the regime of weak nonlinear interactions, in which the pump pulse is not depleted. In the regime of strong interaction, the situation is much more complicated, and no simple rules can be given. As an example of the modified role of dispersion in these situations, we consider the method of Umbrasas et al.,\textsuperscript{110} in the generation of second-harmonic radiation of short optical pulses, and their simultaneous compression. The experimental arrangement is shown schematically in Fig. 31(a). The input waves have orthogonal polarization and are incident on the nonlinear crystal in a collinear fashion, but delayed with respect to one another. The crystal is configured for type-II phase matching for frequency doubling. In this case, the group velocities of the $o$- and $e$-wave pulses are not the same, and they propagate through the crystal so that the delay between them at the output is exactly the negative of what it was at the input—the fundamental pulses have “walked through” each other inside the crystal, as shown in Fig. 31(b). If, however, the group velocity of the second-harmonic wavelength is close to the mean velocity of the two
input waves, and the power is sufficient to deplete the fundamental pulses during their interaction, then the generated second-harmonic pulse will be quite short. This is because energy is fed into it only from a small region where the two fundamental pulses overlap, at which point they are depleted. In this way pulses of about 0.5 ps have been generated by frequency doubling pulses of several tens of picoseconds in duration.

Another way to adjust the GDD to phase match a nonlinear interaction is to use waveguide dispersion to balance the dispersion of the material. This approach has been particularly successful in application to the generation of high harmonics of infrared pulses. Radiation at frequencies corresponding to the 150th or so harmonic of the input frequency can be generated when atoms are illuminated with an optical field whose strength is comparable to that of the Coulombic binding field of the outer-shell electron of an atom. This can be achieved by making the nonlinear susceptibility a function of the position in the nonlinear material. In bulk material, when the phase mismatch is equal to zero, the energy in the converted field grows quadratically with the length of interaction. In other cases, the energy increases for values of \( L \) in the range \( [0, L_c] \), where \( L_c = \pi/\Delta k \) is the coherence length. It reaches a maximum at \( L = L_c \) and decreases over the range \( [L_c, 2L_c] \), finally going back to zero because of reconversion from the second-harmonic field to the fundamental field. For this material, the product of the nonlinear susceptibility with the phase-matching term after a distance \( L = L_c + L' \) can be written as

\[
\chi^{(2)}(\omega; \omega_1, \omega - \omega_1) \int_0^{L_c} dz e^{i[k(\omega_1)+k(\omega_1)-k(\omega)]z} + (\chi^{(2)})_0 \int_{L_c}^L dz e^{i[k(\omega_1)+k(\omega_1)-k(\omega)]z},
\]

which simplifies to

\[
\chi^{(2)}(\omega; \omega_1, \omega - \omega_1) \int_0^{L_c} dz e^{i[k(\omega_1)+k(\omega_1)-k(\omega)]z}.
\]

\[\text{FIG. 27. Angle phase matching for second-harmonic generation in a nonlinear type I crystal. A proper choice of the phase-matching angle (a) allows the equality of the index of the fundamental along the ordinary axis and second-harmonic frequency along the extraordinary axis, leading to the cancellation of phase mismatch (b).}\]

\[\text{FIG. 28. Optical parametric generation of short pulses of different group velocities.}\]

\[\text{FIG. 29. Diagram of a synchronous pumped femtosecond optical parametric oscillator.}\]

\[\text{C. Quasi-phase-matched processes}\]

The previous techniques enhance the efficiency of the nonlinear process by using materials and configurations

\[\text{where the phase mismatch between the interacting waves is zero over a broad range of frequencies, thus maximizing the coherence length for a wide wavelength range. The technique of "quasi phase matching" relies on the engineering of the nonlinear medium itself in order to cancel the destructive interference of the interacting waves.}\]
The change in the sign of the susceptibility in the range \( L_c \), precludes the backconversion from the second-harmonic to the fundamental field. The efficiency of second-harmonic generation in this range is then the same as in the previous range \( 0, L_c \). When propagating over multiple periods of such a structure, the field then grows linearly with the propagation length, and therefore the energy grows quadratically as in the case of a perfectly phase-matched configuration. This leads to very efficient frequency mixing arrangements for sum frequency generation and parametric downconversion. An interesting aspect of quasi phase-matching is that as exact phase matching is not required, the polarization of the interacting waves can be chosen to take advantage of the highest nonlinear coefficients in the material. A complete treatment of such a structure with periodicity \( L \) shows that the phase-matching factor for the conversion efficiency has the same form as in a bulk material, but with a phase mismatch shifted by the value of the wave vector of the periodic structure \( K_m = 2 \pi m / \Lambda \), chosen close to \( \Delta k_{\text{bulk}} \):

\[
\Delta k = \Delta k_{\text{bulk}} - K_m.
\]

There is thus no specific improvement of the spectral acceptance as far as a periodic structure is concerned. However, the possibility to use aperiodic structures enhances the spectral acceptance, which is particularly useful for ultrafast optics. In this case, local sections of the structure can be engineered in order to upconvert different portions of the broad spectrum by the choice of the period. As the upconverted spectral components travel with their own group velocity, such a structure modifies the group delay in the upconverted pulse. It has been shown that these structures can have the same effect as the compressor at the end of a chirped pulse amplification system by using a chirped period in the quasi-phase-matched device, keeping in mind that the mean frequency of the compressed pulse is twice the mean frequency of the amplified pulse. Other structures have produced trains of upconverted pulses from a single fundamental pulse.

### D. Non-phase-matched processes

Nonlinear processes that do not require phase matching or are automatically phase matched are nonetheless affected by dispersion. Examples of such processes are the self- and cross-action effects induced by the nonlinear refractive index and gain processes such as stimulated Raman scattering. The effects of dispersion are quite similar to those in the phase-matched case—a mismatch of the group velocity between the interacting waves reduces the efficiency of the process, as does nonzero GDD.

For example, it has already been mentioned that the combination of positive nonlinear refractive index and negative GDD occurring in the same system can be used to generate optical solitons. These are optical pulses of particular shape, whose energy is specified by the GDD of the system. This phenomenon has been observed in waveguide geometries by many scientists, and is now the basis of high-speed...
time-division multiplexed optical telecommunications test-beds.

A critical issue in such systems is maintaining the balance of self-phase modulation and GDD even in the presence of imperfections in the fibers. If the two effects do not balance exactly, then the soliton will cease to be such, and the pulse will gradually disperse. Since all fibers have at least losses due to Rayleigh scattering in the fiber core, it might be expected that the range of a soliton communications link would be restricted. In fact, clever management of dispersion allows one to circumvent this problem. For example, an arrangement such as that shown in Fig. 33(a), where the GDD of the fiber is a monotonically decreasing function of distance, allows a soliton to be sustained even in the presence of losses. This is because the decreasing energy of the soliton as it undergoes scattering causes smaller self-phase modulation at the peak of the pulse, and thus lower GDD is required to maintain the solitonic shape. The balance between these effects is maintained as the pulse propagates by tailoring the dispersion in the correct way to match the losses. Examples of the output of the fiber (using a 1.5 ps input pulse) for uncompensated and compensated fibers are shown in Fig. 33(b). In the case of the uncompensated fiber, the increased pulse duration indicates that the pulse became dispersed rather than remaining a soliton after propagating a fraction of the distance along the fiber.

A similar argument holds for spatiotemporal solitons in bulk materials. Here there is no material-imposed waveguiding, and a spatial soliton is formed by a balance between self-focusing and diffraction. The coupling of spatial and temporal degrees of freedom via these mechanisms stabilizes the soliton, leading to a so-called optical bullet that propagates as a localized excitation in space and time. A temporal soliton requires that the dispersion and nonlinear phase shift balance one another. This is not possible in a medium with positive nonlinear refractive index and positive dispersion. In order to overcome this effect, Liu et al. used a nonresonant cascaded process to synthesize a nonlinearity of the correct sign. This process effectively concatenates an upconversion and a downconversion process in sequence, in a birefringent medium far from phase matching. The interference of the fundamental wave with the light that undergoes conversion to the second harmonic and back again causes an intensity-dependent phase shift that mimics a nonlinear refractive index. The effectiveness of this process is limited by the group velocity mismatch between the fundamental and the second harmonic. Because the material has positive dispersion, it is necessary to “precompensate” by slightly negatively chirping the pulse before entering the nonlinear medium. This is done using angular dispersion, by placing the nonlinear crystal at the focal plane of a dispersive delay line.

The interplay between dispersion and nonlinear phase shifts in the propagation of ultrashort pulses is currently a subject of intense study, and a comprehensive survey is not possible here. In many cases simple pictures have been developed only \textit{ex post facto}, and usually they provide only a partial description of the effect. Typically, a complete numerical solution of the problem is required. Nonetheless, great progress has been made in several areas, including the generation of continuum radiation in bulk and waveguided geometries, as well as four-wave mixing using third-order and cascaded second-order nonlinearities.

A third example is provided by group velocity matching for effective stimulated scattering. The method to be described was originally developed for traveling wave amplifiers with large bandwidths.

In Raman scattering, an intense pump pulse scatters light into a pulse of lower frequency, the Stokes pulse, leaving the medium in an excited vibrational, rotational, or electronic state. The process can become stimulated, and this has lead to the development of Raman amplifiers and frequency shifters for pulses from the nanosecond to the femtosecond regime. The Raman cross section is larger the closer the pump and Stokes frequencies are to resonance. But this also means that there is a difference in group delay between the two pulses. As a consequence, the effective length of the interaction region may be shorter than its physical length. Because the material in which Raman scattering occurs is usually not birefringent, it is not possible to use angular dispersion to adjust the two group velocities of pump and Stokes pulses to be the same. One method that has been adopted to overcome this problem is to configure the pump beam so it has a particular geometry, that of a “Bessel beam.”

If a planar pump beam is incident on an axicon (a conical refractive element), then the beam is deviated from its input direction in such a way to form a region along the axis of the axicon in which the intensity is very high. Moreover the region of high intensity moves along the axis with a velocity that depends on the cone angle \( \varphi \) of the focused pump light. The velocity of the illuminated region along the cone axis is approximately \( c/n \cos(\varphi) \), where \( n \) is the refractive index of the medium. This allows the possibility of matching the speed of the moving pumped region to the speed of the Stokes pulse. Even though the direction of the pump light itself is at an angle \( \varphi \) to the axis, phase matching is automatically satisfied for Stokes scattering, and therefore this angle is a free parameter that can be adjusted to match the speed of the gain region along the axis to the group velocity at the Stokes wavelength.

Raman conversion efficiencies of up to 1% for 100 fs pulses in liquids have been observed using this technique.

FIG. 32. Energy of the harmonic field in the case of perfect phase matching in a uniform material (a), nonphase matched interaction (c), and quasi-phase-matched interaction when reversing the sign of the nonlinear susceptibility every coherence length (b). The x axis is in units of coherence length for the non-phase-matched process.

\[ I(x) = I_0 + \frac{1}{a + b + c} \int_{-\infty}^{\infty} h(x') \left( a\sqrt{I_0} + b\sqrt{I_0} + c\sqrt{I_0} \right) dx' \]

\[ a = \frac{\sqrt{I_0}}{\sqrt{I_0}} \]

\[ b = \frac{\sqrt{I_0}}{\sqrt{I_0}} \]

\[ c = \frac{\sqrt{I_0}}{\sqrt{I_0}} \]

\[ x = \frac{\sqrt{I_0}}{\sqrt{I_0}} \]
Normally in liquids the competing process of four-wave mixing dominates frequency conversion, but this process requires phase matching, and the situation with Bessel beam excitation precludes such.

**VIII. DISCUSSION**

It is critical to pay attention to dispersion when dealing with ultrashort optical pulses, since it can severely compromise the focused intensity and the pulse duration of beams at a target. Even more important, it is crucial to understand how to manipulate dispersion in order to compensate for the detrimental effects of too much or too little in an optical system. This, in turn, demands a set of methods for measuring dispersion.

A reasonable rule of thumb, applicable in most cases of linear pulse propagation, and to a lesser extent in weakly nonlinear situations, is that pulses with bandwidths of less than 1 nm and durations of greater than 1 ps are not much affected by dispersion (at least in transparent media, far from absorption resonances). On the other hand, pulses with greater than 1 nm of bandwidth or durations of less than 1 ps may be affected quite dramatically by dispersion. The management of dispersion becomes increasingly more important as the brevity of the pulses decreases.

The options available for controlling dispersion depend critically on the system in which the dispersion must be managed. The trade-off is usually between a low-loss system with a small range of dispersion adjustment or a higher loss system with a much wider range of adjustment. The former strategy is nearly always adopted with mode-locked oscillator-based experiments, whereas the latter is used for amplified systems. Material dispersion in the visible is always positive (except in amplifiers, where it can be weakly negative). In combination with other elements, angular dispersion may be used to obtain controllable negative or positive group delay dispersion. Interferometric thin-film coatings can provide dispersion of arbitrary sign, but with very limited adjustability, so they are effective only as one-off devices, for particular system configurations. Nonetheless, they are very compact and have low loss, in contrast to grating-based dispersive systems.

Dispersion plays a central role in nonlinear optics. Although it is difficult to make any completely general statement about how it affects the details of the nonlinear process, it is clear that all wave mixing phenomena involving pulses with disjoint spectra will depend upon whether the pulses walk away from each other, as well as whether their peak intensity is reduced upon propagation. The subtle interplay of dispersion, diffraction, and nonlinearity will undoubtedly limit adjustability, so they are effective only as one-off settings can provide dispersion of arbitrary sign, but with very positive ways positive.

**ACKNOWLEDGMENTS**

The authors are grateful for numerous conversations with and suggestions from many persons. Of particular help in the preparation of material for this article were: Chris Barty, Eric Buckland, Irl Duling, Turan Erdogan, Chris Iaconis, Jim Kafka, Wayne Knox, Bill White, and Frank Wise. We are especially grateful to Chris Barty for helpful comments on the manuscript. The work was supported by the National Science Foundation.

---

1. E. Muybridge, *Horses and Other Animals in Motion* (Dover, New York, 1985), pp. 91.