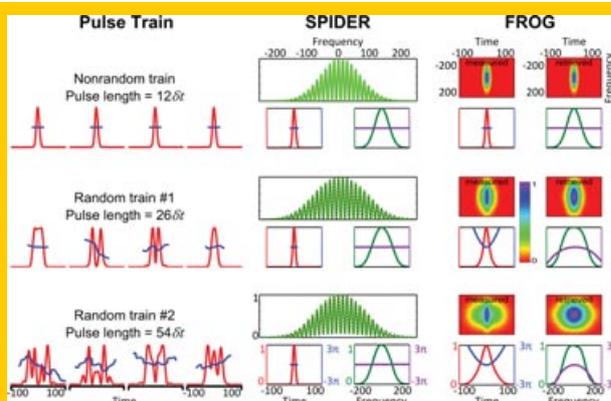


**Abstract** Multi-shot pulse-shape measurements of trains of ultrashort pulses with unstable pulse shapes are studied. Measurement techniques considered include spectral-phase interferometry for direct electric-field reconstruction (SPIDER), second harmonic generation frequency-resolved optical gating (FROG), polarization gate FROG, and cross-correlation FROG. An analytical calculation and simulations show that SPIDER cannot see unstable pulse-shape components and only measures the coherent artifact. Further, the presence of this instability cannot be distinguished from benign misalignment effects in SPIDER. FROG methods yield a better, although necessarily rough, estimate of the pulse shape and also indicate instability by exhibiting disagreement between measured and retrieved traces. Only good agreement between measured and retrieved FROG traces or 100% SPIDER fringe visibility guarantees a stable pulse train.



## Pulse-shape instabilities and their measurement

Michelle Rhodes<sup>1,\*</sup>, Günter Steinmeyer<sup>2</sup>, Justin Ratner<sup>1</sup>, and Rick Trebino<sup>1</sup>

An ideal measurement technique is not only accurate, but also robust. If the accuracy of a measurement is too easily compromised by difficult experimental circumstances, then extreme care must be taken in interpreting its results. In particular, multi-shot pulse-shape measurements of unstable pulse trains have yielded confusing or misleading results. Specifically, intensity-autocorrelation measurements of trains of differing pulses are well-known to yield a narrow spike at zero delay atop a broad background. The spike, referred to now as the “coherent artifact,” is a measure of the coherent, nonrandom, or repeatable component of the pulse train. While some have mistakenly interpreted its width as a measure of the pulse width, the broad background is the correct indicator of the actual pulse width. And the presence of a coherent artifact actually indicates instability or complexity in the pulse train. More precisely, the coherent artifact represents only the shortest repeatable substructure in the pulses and ignores any variations in the pulses.

### 1. Historical overview

The coherent artifact has caused confusion in laser pulse measurement for almost as long as pulse measurements have been conducted. In the late 1960s, significant confusion arose when researchers noticed that peaks could be observed in two-photon fluorescence (TPF) intensity-autocorrelation measurements [1] even when the lasers involved were not mode-locked [2–4]. In 1968, several authors [5,6] used previous work on light coherence [7,8] and

correlation functions [9,10] to explain the variety of autocorrelation traces that could result from measuring laser light. It was established that a peak always appears at zero delay in an autocorrelation [6,11–15] unless the light under measurement is perfectly monochromatic and has an infinite coherence length. All broadband sources yield a central peak, and the width of this peak is approximately equal to the inverse spectral width, called the coherence time [14,15]. This peak should not be assumed to be indicative of a short pulse. It is sometimes called a coherent artifact because it arises from coherence effects, but it occurs in single-shot measurements involving only the pulse intensity (e.g. autocorrelation) and should not be confused with multi-shot phenomena discussed later. In addition to a central peak, autocorrelations of complex pulses will also show a pedestal [12,16,17]. The relative strengths of the pedestal and coherence peak provide useful information about pulses. For example, broadband fluorescence and white-light sources have a coherence spike but no pedestal. Such sources usually have much longer pulse lengths, so their pedestals are difficult to separate from the usual TPF background. Imperfect mode-locking of lasers tends to yield shorter, noisy pulses whose autocorrelations have cusp-like coherence spikes on top of a noticeable broad pedestal. A fully mode-locked laser yielding transform-limited pulses has a very strong coherence peak and no pedestal. For good reviews of this discovery process, see [12] and [15].

Once clear criteria for measurements that showed true short pulses, not just coherent artifacts, had been established and experimentally verified [12,18,19], attention

<sup>1</sup> Georgia Institute of Technology, School of Physics, 837 State St NW, Atlanta, GA 30332, USA

<sup>2</sup> Max Born Institute for Nonlinear Optics and Short Pulse Spectroscopy, 12489 Berlin, Germany

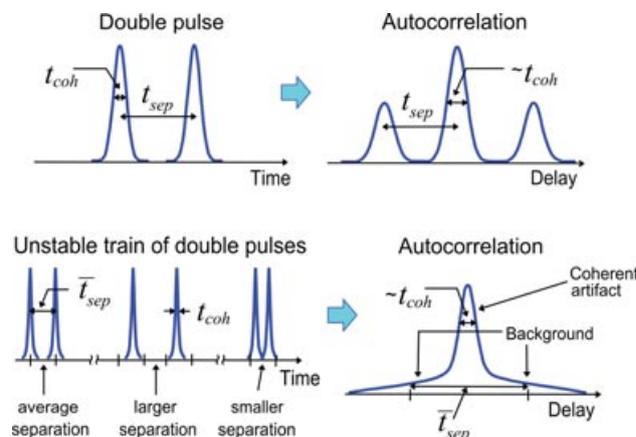
\*Corresponding author: e-mail: mrhodes3@gatech.edu

turned to the shape of the autocorrelation, especially the pedestal. The pedestal defines the temporal resolution in spectroscopic pump-probe experiments and is therefore of far greater interest than the shorter coherence time. The temporal duration of the enveloping pulse profile is consequently referred to as the pulse duration in the following, because the duration of individual sub-pulses is of little relevance for time-resolved experiments. Even though several authors noted that the contrast and shape of the autocorrelation and pedestal were not particularly sensitive to the pulse profile [17, 20–23], there were no alternative optical techniques available at the time. The problem was complicated by the fact that actual pulse lengths were typically much longer than the corresponding coherence time [6, 19]. This problem was even further complicated by the fact that most TPF measurements, with only a few exceptions [24], were in fact averages over many pulses in a pulse train. Because the intensity-autocorrelation pulse-retrieval problem had so many ambiguities, researchers could only guess at the actual pulse distortions present [20, 25, 26].

As a result, misinterpretations were common. For example, Treacy believed that he had fully compressed his pulses by compensating for positive frequency chirp [27, 28]. However, Fisher and Fleck [29] pointed out that Treacy's results (showing a short spike atop a broad background) were consistent with the output pulses of the laser each having a different random walk for their phase versus time. Compensating such pulses for positive chirp would result in random temporal structure in the intensity that varied from pulse to pulse. The multi-shot autocorrelation of such a train of pulses would be the same as that reported by Treacy, and his results therefore did not prove that the uncompensated pulses had positive chirp. This was the first identification of the multi-shot coherent artifact. The insight that a train of varying pulses could mimic a shorter-pulse, stable train in a measurement was quite surprising (see Fig. 1).

As laser technology improved and pulse trains became more stable, these types of issues became less prominent but did not disappear. In the late seventies, researchers had trouble matching autocorrelations of synchronously pumped dye lasers with a pulse shape [31–34]. The closest pulse profile was a single-sided exponential pulse; however, Van Stryland [35] argued that the autocorrelations were also consistent with a train of Gaussian pulses with a distribution of pulse widths. Birmontas *et al.* [36] extended Van Stryland's analysis to include variations in pulse energy as well as duration and showed that the autocorrelation underestimates the average pulse length of unstable trains when the fluctuations in pulse power and length are correlated. It took more than a decade to unveil the physics behind the observed unstable synchronously pumped mode-locking and to find a means for stabilization of the mode-locking process [37, 38].

The non-uniqueness of the autocorrelation function [39–41] combined with the possibility of unstable trains remained a significant problem for laser pulse measurement until the development of more powerful pulse-measurement techniques and further improvements in laser technology. Unlike autocorrelation, modern pulse measurement tech-



**Figure 1** Top: Double pulse and its background-free autocorrelation [30]. Bottom: A train of variably spaced double pulses and their multi-shot autocorrelation. The coherent artifact results from the short nonrandom coherent component of the double pulses (a single pulse), while the broader background results from the overall average pulse length (the combination of both pulses). This trace is typical of autocorrelations of nearly all trains of unstable, complex pulses.

niques do not suffer from non-uniqueness (except for trivial ambiguities) and provide not only the intensity profile, but also the phase of pulses. However, their reliability in the face of pulse trains with pulse-shape instabilities has only recently begun to be examined in detail [30, 42]. And while synchronous pumping has been widely replaced by more robust and stable passive mode-locking techniques, unstable pulse trains often still arise, especially in super-continuum compression experiments [43–45].

While instabilities besides pulse-shape fluctuations certainly exist, they are typically of little relevance for pulse measurement techniques, in particular if the techniques are completely self-referenced. Beam pointing fluctuations as well as pulse energy variations may certainly increase the noise levels in the detection and require suitable averaging techniques [46]. Carrier-envelope phase fluctuations [47] could be another possible noise source for few-cycle laser pulses. Such effects can clearly be seen when measuring cross-correlations or interferograms using two subsequent pulses from a pulsed laser source [48, 49]. It is understood, however, that carrier-envelope effects could otherwise only pose an issue for pulses whose coherent spectra span more than an octave [50], a situation far less common than the coherence spike. Moreover, all mode-locked lasers exhibit pulse timing jitter [51], which is typically characterized by RF measurement techniques [52]. However, such techniques do not have the resolution to measure ultrashort pulse shapes or pulse-shape fluctuations, making them ill-suited to many applications where intensity and phase measurements are used. All measurement techniques discussed in the following analysis employ pulse-replicas that are derived from the same oscillator or amplifier pulse. Pulse arrival time cannot be measured in this case, and variations in arrival time are unobservable. Even the cross-correlation variants of FROG and SPIDER nearly always use an

undistorted pulse from the same laser as a reference, removing the problem of synchronizing multiple laser sources. This makes these pulse measurement techniques completely immune to timing jitter noise. More importantly, none of the noise sources mentioned above cause a multi-shot coherence spike, as pulse-shape instabilities do.

When considering the possibility of pulse-shape instabilities, it becomes clear that an intensity-and-phase pulse-shape measurement has a responsibility to give some measure of the reliability of its result, which should include the stability of the pulse train. In particular, the technique should not introduce new non-trivial ambiguities resulting in confusion, for example, between a stable train of very short pulses and an unstable train of much longer ones. Given that a measurement can only give a single intensity and phase, and so cannot represent all the pulses of an unstable train, a good measurement technique should still give an approximation of a typical pulse. In addition, it is vital that there are indicators of instability. Although many modern pulse-measurement techniques have single-shot capability, including all the techniques discussed in this paper, multi-shot measurements are extremely common in practice. Consequently, multi-shot versions of pulse-measurement techniques must be evaluated to determine how they respond to unstable pulse trains.

In this paper, measurements of unstable pulse trains are simulated in order to study the performance of several self-referenced pulse measurement techniques, including spectral-phase interferometry for direct electric-field reconstruction (SPIDER)[53, 54], second-harmonic-generation (SHG) frequency-resolved optical gating (FROG)[41], and polarization-gate (PG) FROG. The reference-pulse-based technique cross-correlation FROG (XFROG) will also be examined. Our unstable train of pulses consists of a stable flat phase Gaussian pulse plus a random pulse. For this case, it is also possible to derive an analytical expression for the resulting SPIDER measurement.

## 2. Analytical calculations

In its ideal form, SPIDER involves measuring the spectrum of the sum of a pulse and a frequency-shifted and delayed replica of itself. Assuming that the pulse train has a stable, consistent component  $E(\omega)$  and a randomly varying component  $E_{rand}(\omega)$ , the expression for the ideal multi-shot SPIDER trace is:

$$S_{SPIDER} \propto |E(\omega) + E(\omega + \delta\omega) \exp(i\omega T) + E_{rand}(\omega) + E_{rand}(\omega + \delta\omega) \exp(i\omega T)|^2 \quad (1)$$

where  $T$  is the internal pulse separation. This expression contains several simplifying assumptions. Because the shear comes from stretching a pulse replica, it must be assumed that a stretched unstable pulse still has linear chirp. Further, the original pulse should still be short enough to only overlap with a small portion of the stretched pulse, or the frequency shear will not be constant for all parts of the sheared pulse. Consideration of these effects significantly

complicates interpretation of the trace, and so they will be ignored here. Multiplying out all the terms:

$$\begin{aligned} S_{SPIDER} \propto & |E(\omega)|^2 + |E(\omega + \delta\omega)|^2 + |E_{rand}(\omega)|^2 \\ & + |E_{rand}(\omega + \delta\omega)|^2 + 2\text{Re}\{E^*(\omega)E(\omega + \delta\omega) \exp(i\omega T) \\ & + E^*(\omega)E_{rand}(\omega) + E^*(\omega)E_{rand}(\omega + \delta\omega) \exp(i\omega T) \\ & + E^*(\omega + \delta\omega) \exp(i\omega T)E_{rand}(\omega) \\ & + E^*(\omega + \delta\omega)E_{rand}(\omega + \delta\omega) \\ & + E_{rand}^*(\omega)E_{rand}(\omega + \delta\omega) \exp(i\omega T)\} \end{aligned} \quad (2)$$

Even if only the zeroth-order phase of the random pulse is allowed to vary, the random field will be positive as often as it is negative. Thus, any terms that have only one factor of the random field  $E_{rand}(\omega)$  will sum to zero in the multi-shot average, yielding:

$$\begin{aligned} S_{SPIDER} \propto & |E(\omega)|^2 + |E(\omega + \delta\omega)|^2 + |E_{rand}(\omega)|^2 \\ & + |E_{rand}(\omega + \delta\omega)|^2 + 2\text{Re}\{E^*(\omega)E(\omega + \delta\omega) \exp(i\omega T) \\ & + E_{rand}^*(\omega)E_{rand}(\omega + \delta\omega) \exp(i\omega T)\} \end{aligned} \quad (3)$$

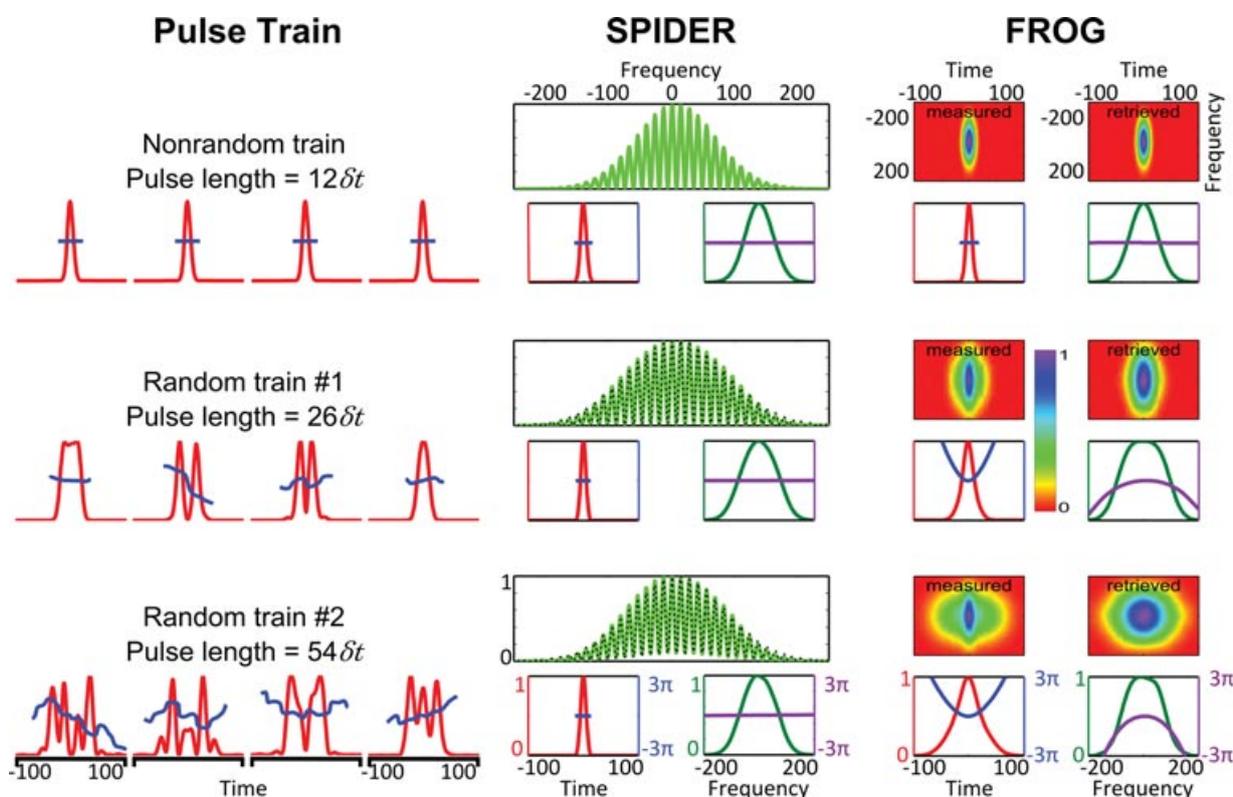
Now this expression can be written in terms of the spectra of the stable and random components,  $S(\omega)$  and  $S_{rand}(\omega)$ , and their spectral phases,  $\varphi(\omega)$  and  $\varphi_{rand}(\omega)$ , noting that the stable component has a constant spectrum and phase:

$$\begin{aligned} S_{SPIDER} \propto & S(\omega) + S(\omega + \delta\omega) + S_{rand}(\omega) \\ & + S_{rand}(\omega + \delta\omega) + 2\sqrt{S(\omega)}\sqrt{S(\omega + \delta\omega)} \\ & \cdot \cos[\varphi(\omega + \delta\omega) - \varphi(\omega) + \omega T] \\ & + 2\sqrt{S_{rand}(\omega)}\sqrt{S_{rand}(\omega + \delta\omega)} \\ & \cdot \cos[\varphi_{rand}(\omega + \delta\omega) - \varphi_{rand}(\omega) + \omega T] \end{aligned} \quad (4)$$

Finally, rewriting the expressions in the cosines in terms of the group delay as a function of frequency for each component,  $\tau(\omega) = d\varphi/d\omega$  and  $\tau_{rand}(\omega) = d\varphi_{rand}/d\omega$ :

$$\begin{aligned} S_{SPIDER} \propto & S(\omega) + S(\omega + \delta\omega) + S_{rand}(\omega) \\ & + S_{rand}(\omega + \delta\omega) \\ & + 2\sqrt{S(\omega)}\sqrt{S(\omega + \delta\omega)} \cos[\tau(\omega)\delta\omega + \omega T] \\ & + 2\sqrt{S_{rand}(\omega)}\sqrt{S_{rand}(\omega + \delta\omega)} \\ & \cdot \cos[\tau_{rand}(\omega)\delta\omega + \omega T] \end{aligned} \quad (5)$$

The net result is a sum of the spectra and sheared spectra, plus the well-known fringe term for the stable component and another for the random component. The random fringes are averaged over many shots, however. If  $\tau_{rand}(\omega)$  is constant for a given pulse, but varies from pulse to pulse, corresponding to a random component with a variable arrival time (such as a satellite pulse, as in the simple case of Fig. 1), then this term will begin to wash out. If  $\tau_{rand}(\omega)\delta\omega$  varies by as much as  $2\pi$  in this way, then the fringes from the



**Figure 2** Comparison of SPIDER and SHG FROG for a random train of pulses [30]. Red curves indicate intensity, blue phase, green spectrum, and purple spectral phase. The black dotted SPIDER traces are fits assuming flat-phase Gaussian pulses and unequal double-pulse energies. SPIDER retrieves only the nonrandom component of the pulse trains, and shows decreasing fringe visibility. SHG FROG misses the pulse structure, but retrieves longer, approximately correct pulse lengths and shows disagreement between measured and retrieved traces.

random component wash out completely. In this case, the only contribution of the random field to the SPIDER trace will be some background. Unfortunately, such a background can also be caused by misalignment and other ever-present and typically harmless effects, so, in practice, it is usually ignored. As a result, it is essentially impossible to distinguish the above harmless effects from the existence of a random component. The impact of higher-order phase variations in the random component is not immediately clear, and this will be explored in the simulations. However, the above analytical calculations suggest that SPIDER may be unable to see most random variations in multi-shot measurements. Indeed, previous work [30] shows precisely this result (see Fig. 2). This will be further investigated by the simulations.

Unfortunately, FROG does not lend itself as easily to analysis. FROG is a spectrally resolved autocorrelation, also known as a spectrogram. The general expression for a FROG measurement is:

$$I_{FROG}(\omega, \tau) = \left| \int_{-\infty}^{\infty} E(t)E_g(t - \tau) \exp(-i\omega t) dt \right|^2 \quad (6)$$

This is the Fourier transform of the signal field, created by the interaction of the pulse,  $E(t)$ , and its gate,  $E_g(t - \tau)$ , in a nonlinear medium. The resulting measurement is a

function of both frequency  $\omega$  and delay  $\tau$ . The gate function depends on the nonlinearity: it is the field itself for SHG FROG, the field intensity for PG FROG and a reference field for XFROG.

Expressing the fields in Equation (6) in terms of stable and unstable components yields no obvious cancellations or clear relationships, so no further analytical analysis will be presented here.

It should be mentioned, however, that an important feature of all versions of FROG is that the measured trace contains more points than are strictly necessary to retrieve an answer, which can help safeguard against spurious measurement effects. As a result, there are simple consistency checks that can be calculated. Summing the trace over delay produces a frequency marginal which is related to the spectrum of the pulse [41]. Summing over frequency gives a delay marginal, which is typically related to an autocorrelation. These quantities are useful in ensuring that the measurement has been performed properly and the pulse has been retrieved from the measured trace correctly. Also, simple agreement between the measured and retrieved traces is a general indicator of a high-quality measurement. Because FROG uses both time and frequency domains to determine pulse properties, in the case of instability it is likely to have discrepancies in both domains and between the measured and retrieved traces.

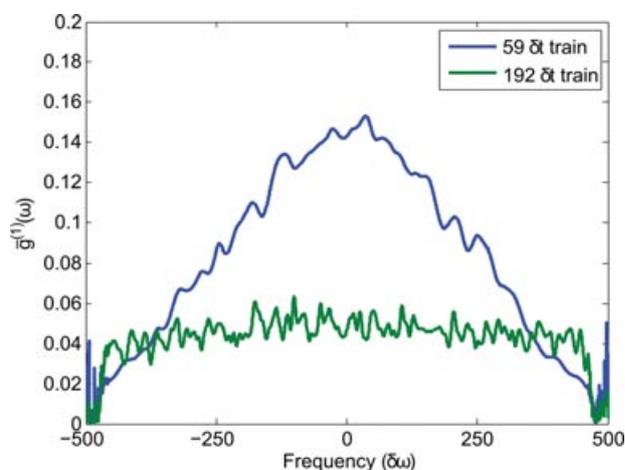
### 3. Simulation details

For the simulations, pulses were constructed consisting of a stable and a random component. The stable component is a flat phase Gaussian with temporal FWHM  $20 \delta t$ , where  $\delta t$  is the temporal sampling rate. The frequency sampling rate is  $\delta\omega = 2\pi/N\delta t$ , where  $N$  is the array size (4096). In order to avoid distorting the spectrum of the combined pulse train, the random component of each pulse began with the same spectrum as the nonrandom component. A random spectral phase was applied to the random component in the frequency domain, and a wide Gaussian envelope was applied in the time domain. Next, the amplitude of the random component was multiplied by a constant so that, on average, it had similar intensity at  $t = 0$  to the nonrandom component. Consequently, the energy contained in the random pulse is larger than in the nonrandom pulse, since the random component is much longer in time. The random and stable components were added together to create a pulse. The amplitude adjustment was verified to be appropriate afterward by calculating the average intensity profile of the train and checking that the transition from the nonrandom spike to a broad background occurred near 50% of the peak intensity. The average width of the pulses in the train was adjusted by changing the width of the Gaussian time envelope. Two pulse trains were constructed, one with an average FWHM pulse length of  $59 \delta t$  and the other with length  $192 \delta t$ , each with 5000 pulses. The average spectra of both trains are nearly identical to the spectrum of the nonrandom component. Some example pulses are included in the left columns of Figs. 4–5. The same two random pulse trains were used to compute traces for all four measurement techniques.

In order to allow comparisons between the coherence properties of these pulse trains and other works, the resulting coherence of the trains is computed from the modulus of the complex degree of first-order coherence [55]:

$$\bar{g}^{(1)}(\omega) = \left| \frac{\langle E_i^*(\omega) E_j(\omega)_{i \neq j} \rangle}{\sqrt{\langle |E_i(\omega)|^2 \rangle \langle |E_j(\omega)|^2 \rangle}} \right| \quad (7)$$

where indices  $i$  and  $j$  label numerically generated individual spectra  $E_i(\omega)$  and  $E_j(\omega)$ , respectively, and  $\omega$  is the angular frequency. Figure 3 shows the resulting coherence averaged over all pulse combinations for the pulse train with an average FWHM pulse length of  $59 \delta t$  as well as that for length  $192 \delta t$ . In both cases, maximum coherence is reached near the center frequency  $\delta\omega = 0$ , with  $\bar{g}^{(1)}$  values of about 0.14 and 0.04, respectively. For the case of  $59 \delta t$ , the coherence quickly decreases further away from the center frequency, whereas it stays nearly constant for  $192 \delta t$ . As perfect coherence is indicated by  $\bar{g}^{(1)} = 1$ , it is plain to see that noise increases with increasing FWHM pulse length in our parameterization of the problem.



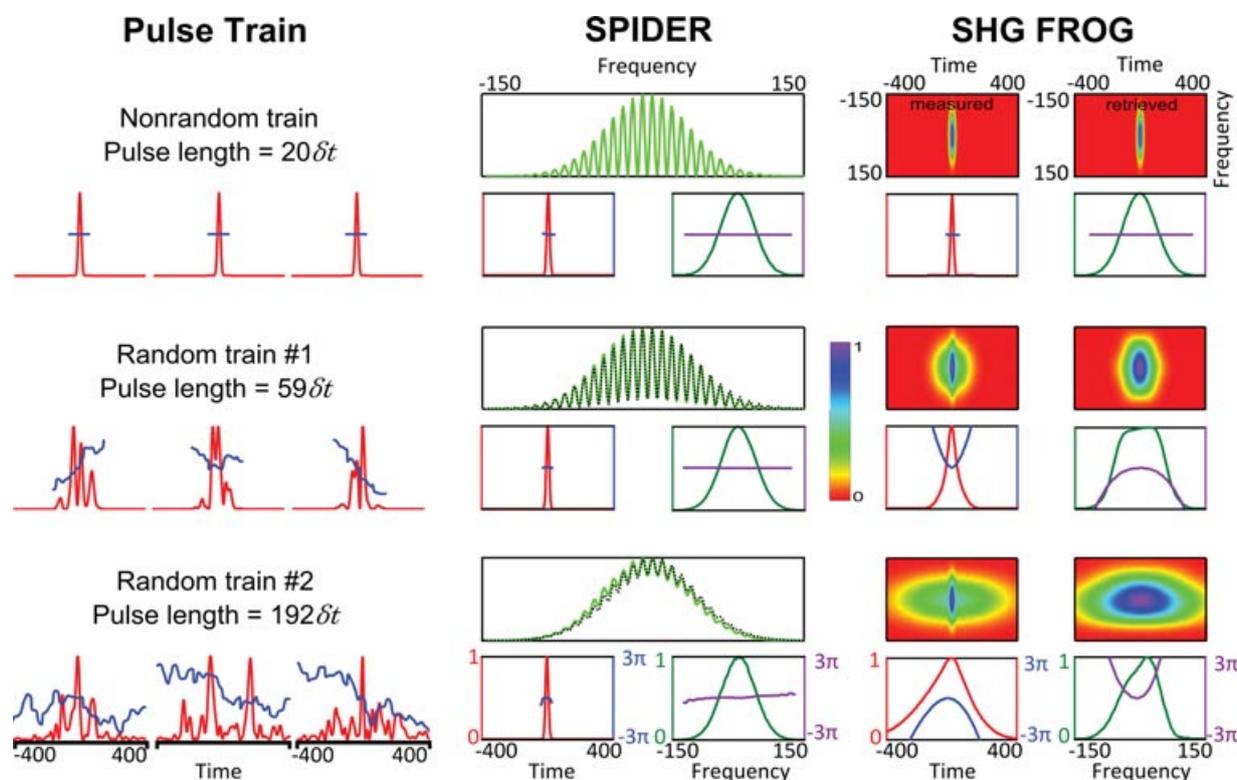
**Figure 3** Average modulus of the complex degree of first-order coherence  $\bar{g}^{(1)}(\omega)$  for the two pulse trains considered in the simulations. Perfect coherence, i.e., identical pulse shapes in the train are indicated by  $\bar{g}^{(1)} = 1$ .

## 4. Results and discussion

### 4.1. SPIDER

SPIDER traces were computed using an array size of  $N = 4096$  and pulse separation  $T$  of  $450 \delta t$ . The frequency shear used was  $9 \delta\omega$ , corresponding to 10% of the FWHM bandwidth of the trains. The trace was averaged over all 5000 pulses in a train, and then retrieved using the Takeda algorithm typically employed in SPIDER [56]. In all cases, SPIDER yields only the nonrandom component (the coherent artifact). Less than 100% fringe visibility is the indicator of the presence of instability (see Fig. 4). Consistent with the analytical calculations, train 2 has much more background than train 1. The group delay is expected to vary over a larger range from pulse to pulse in train 2, and the background should therefore be larger. For train 1, higher order phase variations appear to cancel out of the SPIDER measurement while leaving very little background. As a result, SPIDER misses a great deal of structure in the random trains.

Of course, background also occurs in SPIDER from a number of benign measurement effects. Any differences between the pulse replicas are likely to cause background, including differences in energy split, spatial overlap, spatial mode-matching, as well as any damage spots or dust present in only one arm of the experimental setup. To demonstrate that such effects can be indistinguishable from instability in the pulse train, a second SPIDER trace is fitted to the multi-shot traces (see Figs. 2 and 4), constructed by a Gaussian pulse with unequal energy in the internal double-pulse in the SPIDER setup. The pulse width of the Gaussian and the relative energy of the double pulse are allowed to vary. In all cases, the fit is in phase with the multi-shot trace, and will therefore give the same SPIDER retrieval. In summary, in SPIDER measurements, it



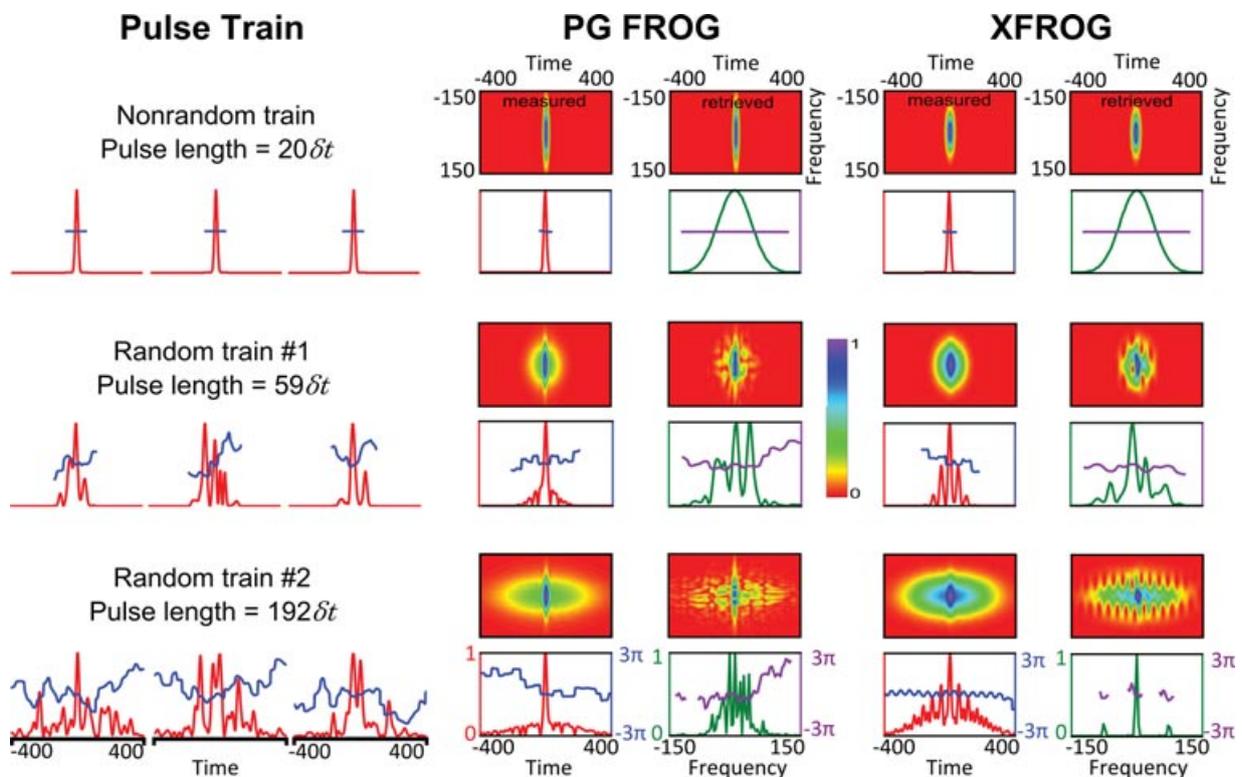
**Figure 4** Simulated multi-shot SPIDER and SHG FROG measurements of random and nonrandom trains. The black dotted SPIDER traces are fits assuming flat-phase Gaussian pulses and unequal double-pulse energies. SPIDER retrieves the nonrandom component with decreasing fringe visibility: 100%, 87%, and 20% respectively. SHG FROG exhibits an autocorrelation-like coherent artifact: the narrow blue spikes in the measured traces for the two random trains. SHG FROG fails to see the pulse structure, but the measured and retrieved traces disagree, providing at least an indication of the instability. G errors are 0.0002, 0.014, and 0.02, and  $G'$  errors are 0.45%, 19%, and 19%.

is difficult to distinguish between background due to these benign, practical effects and more serious instability. Thus very close to a 100% fringe visibility is necessary for an accurate estimate of the pulse length based on a SPIDER measurement. For example, a SPIDER fringe visibility of 87% (a better than average value) corresponds to a measured pulse length too short by a factor of approximately 3 (see Fig. 4).

#### 4.2. FROG

FROG traces were computed using a sampling rate of  $4\delta t$  and  $4\delta\omega$  to reduce the size of the traces and to be consistent with common experimental practice. The pulses were also cropped from the SPIDER array size of 4096 to 1024, yielding a FROG trace size of  $256 \times 256$ . The XFROG measurement used a flat phase Gaussian reference pulse with a temporal FWHM of  $43\delta t$ . All FROG versions were simulated in the single-shot configuration, and the traces were averaged over all pulses in a train. The resulting measured traces show a spike similar to autocorrelation's familiar coherent artifact in the center, with a broad, smooth background around it. Because no single pulse exists that

can possibly cause the measured FROG trace, the FROG code tends to stagnate and be sensitive to the initial guess. The maximum number of iterations of the algorithm was limited to 1000. This limit is more than adequate for single pulses from train 2 to converge on the first attempt. Regardless of convergence, the algorithm was run five times for each averaged measured trace, using different random noise for the initial condition each time, as suggested in Ref [57]. Because high rms error is a strong indicator of measurement problems, and because non-convergence of the algorithm in such a pathological situation is a concern, the retrieved trace with the least error was chosen from the five results. Retrieved FROG traces for the random trains show very clear differences from the measured traces. For the unstable trains, the algorithm was unable to converge on all attempts for the averaged traces, and doubling the iteration limit to 2000 had little to no effect on the resulting rms error. This strongly suggests that convergence is not possible for these traces. Both the G error and the  $G'$  error (see Appendix A) are unacceptably large, especially for noiseless theoretical traces. SHG FROG (Fig. 4) has the smallest errors, consistent with having some trivial ambiguities not present in the other FROG techniques used. It retrieves  $69\delta t$  and  $293\delta t$  as the lengths of the trains, overestimating their actual lengths of  $59\delta t$  and  $192\delta t$ .



**Figure 5** Simulated multi-shot PG FROG and XFROG measurements of random and nonrandom trains. The same trains are used as in Fig. 4, but different example pulses are shown above in order to show more of the pulses involved. Despite the autocorrelation-like coherent artifacts and smooth backgrounds in the measured traces, FROG infers pulse structure. PG FROG has  $G$  errors 0.0002, 0.016, and 0.03 and  $G'$  errors 0.39%, 27%, and 41%. XFROG has  $G$  errors 0.0001, 0.026, and 0.041 and  $G'$  errors 0.22%, 34%, and 34%.

PG FROG and XFROG (Fig. 5) retrieved temporal profiles that show structure reminiscent of the unstable pulse trains. XFROG is more sensitive than (third-order) PG FROG, and therefore is most able to see the fluctuations. Both of them underestimate the average FWHM length of the trains, however, with PG FROG retrieving  $23 \delta t$  and  $22 \delta t$ , and XFROG retrieving  $21 \delta t$  and  $120 \delta t$ . It is important to note that the retrieved spectra, traces, and temporal intensities all point to pulses more complicated than the simple Gaussian stable component, and the error indicates failed convergence.

Because the FROG algorithm has been demonstrated to be quite robust [57], significant disagreement between measured and retrieved traces should generally be attributed to instability or measurement error of some type, rather than non-convergence, as is occasionally speculated. In particular, the blotchy, structured appearance of the retrieved PG FROG and XFROG traces, contrasting with the smooth measured traces, is unlikely to result from anything but pulse-train instability. This effect has been encountered previously in multi-shot XFROG measurements of supercontinuum [42], where it revealed the true unstable nature of continuum when simple multi-shot spectral measurements did not. It appears that over-determination in the FROG trace is a significant advantage in the presence of instability.

## 5. Conclusions

In conclusion, practitioners of ultrafast optics should remain cautious about the stability of their pulses and should interpret their measurements accordingly. It should *not* be taken as given that all modern passive mode-locking techniques for bulk lasers always produce stable pulse trains. And the burden of proof that a new design yields a stable pulse train is, as always, on the researcher reporting it. Care is especially necessary when characterizing white-light supercontinuum pulses and their possible compression. Pulse break-up instabilities are common in white-light generation in microstructure fibers [42–44], filaments [58], and also mode-locked fiber lasers [59, 60].

Quite generally, SPIDER does a very good job of measuring the *nonrandom* component of the pulse train (the coherent artifact) but is not able to discern instabilities. It fails to distinguish a train of long, unstable pulses from a train of short, stable pulses. In other words, SPIDER measures *only* the coherent artifact. If claims of stability or a quantitative pulse length are to be made from a SPIDER measurement, it should have fringe visibility very close to 100%. As this is quite difficult to achieve in practice, an additional measurement of an autocorrelation or FROG trace for the same pulse train [61] could help to yield a more convincing safeguard against instability.

On the other hand, FROG measurements alone directly provide a convincing indication of stability when the measured and retrieved traces agree well. And conversely, FROG indicates instabilities via disagreement and corresponding excessively large FROG errors. In the latter case, the retrieved pulse shape is a better representation of the pulse shapes if obtained using PG FROG or XFROG than with SHG FROG. Here, once again, it is important to understand that pulse duration refers to the obtainable temporal resolution, e.g., in pump-probe experiments.

Ultimately, neither SPIDER nor FROG (nor any other method) is able to reproduce the complete picture of pulse-shape instabilities in fluctuating pulse trains. Because pulse-shape instabilities will suffice to exclude a given laser source from almost any application, researchers should use extreme caution with SPIDER measurements whose fringe visibilities are less than 100% and FROG measurements if disagreement occurs between measured and retrieved traces. In such cases, instability should likely be reported, and single-shot measurements should be considered.

**Acknowledgements.** This work was supported by the National Science Foundation, Grant #ECCS-1028825, and the Georgia Research Alliance. The authors would also like to thank Robert A. Fisher for helpful discussions.

**Conflict of Interest.** Coauthor Rick Trebino owns a company that sells FROG devices.

**Received:** 14 November 2012, **Revised:** 25 February 2013,

**Accepted:** 28 February 2013

**Published online:** 4 April 2013

**Key words:** Pulse measurement, SPIDER, FROG, instability, multi-shot, coherent artifact.

## References

- [1] J. A. Giordmaine, P. M. Rentzepis, S. L. Shapiro, and K. W. Wecht, *Appl. Phys. Lett.* **11**(7), 216–218 (1967).
- [2] M. A. Duguay, S. L. Shapiro, and P. M. Rentzepis, *Phys. Rev. Lett.* **19**(18), 1014–1016 (1967).
- [3] S. L. Shapiro, M. A. Duguay, and L. B. Kreuzer, *Appl. Phys. Lett.* **12**(2), 36–37 (1968).
- [4] M. Bass and D. Woodward, *Appl. Phys. Lett.* **12**(8), 275–277 (1968).
- [5] H. P. Weber, *Phys. Lett. A* **27**(5), 321–322 (1968).
- [6] J. R. Klauder, M. A. Duguay, J. A. Giordmaine, and S. L. Shapiro, *Appl. Phys. Lett.* **13**(5), 174–176 (1968).
- [7] R. H. Brown and R. Q. Twiss, *Nature* **177**(4497), 27–29 (1956).
- [8] R. H. Brown and R. Q. Twiss, *Proc. R. Soc. Lond. A Math. Phys. Sci.* **242**(1230), 300–324 (1957).
- [9] R. J. Glauber, *Phys. Rev.* **131**(6), 2766–2788 (1963).
- [10] L. Mandel and E. Wolf, *Rev. Mod. Phys.* **37**(2), 231–287 (1965).
- [11] H. Weber and R. Dandliker, *IEEE J. Quantum Electron.* **4**(12), 1009–1013 (1968).
- [12] M. Duguay, J. Hansen, and S. Shapiro, *IEEE J. Quantum Electron.* **6**(11), 725–743 (1970).
- [13] H. Rowe and L. Tingye, *IEEE J. Quantum Electron.* **6**(1), 49–67 (1970).
- [14] E. Ippen and C. Shank, *Techniques for measurement; Ultrashort Light Pulses*, S. Shapiro, ed. (Springer Berlin/Heidelberg, 1984), pp. 83–122.
- [15] D. J. Bradley and G. H. C. New, *Proc. IEEE* **62**(3), 313–345 (1974).
- [16] R. J. Harrach, *Phys. Lett. A* **28**(6), 393–394 (1968).
- [17] R. J. Harrach, *Appl. Phys. Lett.* **14**(5), 148–151 (1969).
- [18] S. K. Kurtz and S. L. Shapiro, *Phys. Lett. A* **28**(1), 17–18 (1968).
- [19] S. L. Shapiro and M. A. Duguay, *Phys. Lett. A* **28**(10), 698–699 (1969).
- [20] A. A. Grüter, H. P. Weber, and R. Dandliker, *Phys. Rev.* **185**(2), 629–643 (1969).
- [21] R. H. Picard and P. Schweitzer, *Phys. Rev. A* **1**(6), 1803–1818 (1970).
- [22] J. Herrmann, M. Palme, and K. E. Süsse, *Optic. Quantum Electron.* **10**(3), 195–203 (1978).
- [23] Z. Fried, *Phys. Lett. A* **33**(2), 62–63 (1970).
- [24] W. H. Glenn and M. J. Brienza, *Appl. Phys. Lett.* **10**(8), 221–224 (1967).
- [25] H. A. Pike and M. Hercher, *J. Appl. Phys.* **41**(11), 4562–4565 (1970).
- [26] R. Dandliker, A. Grütter, and H. Weber, *IEEE J. Quantum Electron.* **6**(11), 687–693 (1970).
- [27] E. B. Treacy, *Appl. Phys. Lett.* **14**(3), 112–114 (1969).
- [28] E. B. Treacy, *Phys. Lett. A* **28**(1), 34–35 (1968).
- [29] R. A. Fisher and J. J. A. Fleck, *Appl. Phys. Lett.* **15**(9), 287–290 (1969).
- [30] J. Ratner, G. Steinmeyer, T. C. Wong, R. Bartels, and R. Trebino, *Opt. Lett.* **37**(14), 2874–2876 (2012).
- [31] E. P. Ippen and C. V. Shank, *Appl. Phys. Lett.* **27**(9), 488–490 (1975).
- [32] R. K. Jain and C. P. Ausschnitt, *Opt. Lett.* **2**(5), 117–119 (1978).
- [33] C. P. Ausschnitt and R. K. Jain, *Appl. Phys. Lett.* **32**(11), 727–730 (1978).
- [34] J. P. Heritage and R. K. Jain, *Appl. Phys. Lett.* **32**(2), 101–103 (1978).
- [35] E. W. Van Stryland, *Opt. Commun.* **31**(1), 93–96 (1979).
- [36] A. Birmontas, R. Kupris, A. Piskarskas, V. Smil'gyavichyus, and A. Stabinis, *J. Quantum Electron.* **12**(6), 792–794 (1982).
- [37] J. Catherall and G. New, *IEEE J. Quantum Electron.* **22**(9), 1593–1599 (1986).
- [38] J. Q. Bi, W. Hodel, and H. P. Weber, *Opt. Commun.* **81**(6), 408–418 (1991).
- [39] E. J. Akutowicz, *Trans. Am. Math. Soc.* **83**, 234–239 (1956).
- [40] R. Trebino, E. K. Gustafson, and A. E. Siegman, *J. Opt. Soc. Am. B* **3**(10), 1295–1304 (1986).
- [41] R. Trebino, *Frequency-Resolved Optical Gating: The Measurement of Ultrashort Laser Pulses* (Kluwer Academic Publishers, 2002).
- [42] X. Gu, L. Xu, M. Kimmel, E. Zeek, P. O'Shea, A. P. Shreenath, R. Trebino, and R. S. Windeler, *Opt. Lett.* **27**(13), 1174–1176 (2002).
- [43] B. Schenkel, R. Paschotta, and U. Keller, *J. Opt. Soc. Am. B* **22**(3), 687–693 (2005).

- [44] G. Genty, S. Coen, and J. M. Dudley, *J. Opt. Soc. Am. B* **24**(8), 1771–1785 (2007).
- [45] F. Mitschke, G. Steinmeyer, and A. Schwache, *Physica. D* **96**(1–4), 251–258 (1996).
- [46] G. Stibenz and G. Steinmeyer, *Optics Express* **12**, 6319–6325 (2004).
- [47] T. M. Fortier, J. Ye, S. T. Cundiff, and R. S. Windeler, *Opt. Lett.* **27**, 445–447 (2002).
- [48] L. Xu, C. Spielmann, A. Poppe, T. Brabec, F. Krausz, and T. W. Hänsch, *Opt. Lett.* **21**, 2008–2010 (1996).
- [49] K. Osvay, M. Görbe, C. Grebing, and G. Steinmeyer, *Opt. Lett.* **32**, 3095–3097 (2007).
- [50] H. R. Telle, G. Steinmeyer, A. E. Dunlop, J. Stenger, D. H. Sutter, and U. Keller, *Appl. Phys. B* **69**, 327–332 (1999).
- [51] H. A. Haus and A. Mecozzi, *IEEE J. Quantum Electron.* **29**, 983–996 (1993).
- [52] D. von der Linde, *Appl Phys.* **39**, 201–217 (1986).
- [53] C. Iaconis and I. A. Walmsley, *Opt. Lett.* **23**(10), 792–794 (1998).
- [54] C. Iaconis and I. A. Walmsley, *IEEE J. Quantum Electron.* **35**(4), 501–509 (1999).
- [55] J. M. Dudley, G. Genty, and S. Coen, *Rev. Mod. Phys.* **78**, 1135–1184 (2006).
- [56] M. Takeda, H. Ina, and S. Kobayashi, *J. Opt. Soc. Am.* **72**(1), 156–160 (1982).
- [57] L. Xu, E. Zeek, and R. Trebino, *J. Opt. Soc. Am. B* **25**(6), A70–A80 (2008).
- [58] C. Brée, A. Demircan, and G. Steinmeyer, *Laser Phys.* **21**(7), 1313–1318 (2011).
- [59] A. Hideur, T. Chartier, M. Brunel, S. Louis, C. Özkul, and F. Sanchez, *Appl. Phys. Lett.* **79**, 3389–3391 (2001).
- [60] J. D. Kafka, T. Baer, and D. W. Hall, *Opt. Lett.* **14**, 1269–1271 (1989).
- [61] L. Gallmann, D. H. Sutter, N. Matuschek, G. Steinmeyer, and U. Keller, *Appl. Phys. B* **70**(Suppl.), S67–S75 (2000).
- [62] R. P. Scott, N. K. Fontaine, J. Cao, K. Okamoto, B. H. Kolner, J. P. Heritage, and S. J. B. Yoo, *Opt. Express* **15**(16), 9977–9988 (2007).

## 1. Appendix A: FROG errors

It has become traditional to use the so-called G error—the rms difference between the measured and retrieved traces across the entire trace—as the measure for how well the retrieved FROG trace matches the measured trace. Given a measured, normalized trace  $I_{FROG}(\omega_i, \tau_j)$  and a retrieved trace  $I_{FROG}^{(k)}(\omega_i, \tau_j)$ , the G error is given by:

$$G = \sqrt{\frac{1}{N^2} \sum_{i,j=1}^N |I_{FROG}(\omega_i, \tau_j) - \mu I_{FROG}^{(k)}(\omega_i, \tau_j)|^2} \quad (\text{A.1})$$

where  $\mu$  is chosen to minimize G. This error metric normalizes by the number of points,  $N^2$ , and is appropriate for use within the FROG algorithm, and also when additive noise is the dominant source of noise in the trace, including the trace's edges. Because most FROG traces have more multiplicative noise than additive noise and hence large regions of near-zero values in their outer regions, the value of the G error that indicates good agreement varies with the size of the trace. It has been suggested by Scott *et al.* [62] that, in most cases, a better way to normalize the error for human understanding would be to normalize by the energy in the measured trace. This error, named  $G'$ , is given by:

$$G' = \sqrt{\frac{\sum_{i,j=1}^N |I_{FROG}(\omega_i, \tau_j) - \mu I_{FROG}^{(k)}(\omega_i, \tau_j)|^2}{\sum_{i,j=1}^N |I_{FROG}(\omega_i, \tau_j)|^2}} \quad (\text{A.2})$$

Scott *et al.* found that less than 5%  $G'$  error (or a G error approximately equal to the known multiplicative noise in the measurement) was indicative of good agreement in most cases, and the simple nonrandom retrievals included in the top rows of Figs. 4 and 5 have well under 1%  $G'$  error. Both errors are used in this paper, and FROG users are encouraged to consider the more intuitive  $G'$  error.