Bayesian approach to non-Gaussian field statistics for diffusive broadband terahertz pulses

Jeremy Pearce,* Zhongping Jian, and Daniel M. Mittleman

Department of Electrical and Computer Engineering, MS 366, Rice University, Houston, Texas 77251-1892

Received June 20, 2005; revised manuscript received July 20, 2005; accepted July 20, 2005

We develop a closed-form expression for the probability distribution function for the field components of a diffusive broadband wave propagating through a random medium. We consider each spectral component to provide an individual observation of a random variable, the configurationally averaged spectral intensity. Since the intensity determines the variance of the field distribution at each frequency, this random variable serves as the Bayesian prior that determines the form of the non-Gaussian field statistics. This model agrees well with experimental results. © 2005 Optical Society of America

OCIS codes: 290.4210, 030.6600, 320.7100.

The ubiquity of multiply scattered photons has motivated the investigation of the statistical properties of electromagnetic waves that propagate through random media. Such studies have led to the development of a variety of techniques for imaging in the presence of multiple scattering.[1–6] Speckle correlation spectroscopy can be used to characterize the random medium itself.[7–9] Moreover, the statistics of the diffusive wave provide a key indicator of the onset of localization.[10,11] The vast majority of such studies have employed narrowband sources at either optical or microwave frequencies, although the case of broadband excitation has recently been of increasing interest.[12–15] Previously we described measurements using single-cycle terahertz pulses and provided a phenomenological model for understanding the departures from Gaussian statistics that can be observed in the broadband case.[12] In this Letter we develop a Bayesian formalism to predict this non-Gaussian behavior, which relies not on knowledge of the configurationally averaged spectral intensity but only on its probability distribution function.

The experimental setup has been described previously.[12] We use terahertz time-domain spectroscopy to detect the electric field emerging from a random medium at 90° to the incident wave, with subcycle temporal resolution. The medium consists of a dense collection of 0.794 mm diameter Teflon spheres, held in a 4 cm × 4 cm × 4 cm Teflon cell at a volume fraction of 0.56 ± 0.04. In these samples, the mean free path of the radiation varies dramatically within the bandwidth of the terahertz pulse, by a factor of ~70.[16] We measure the scattered electric field $E_{sc}(t)$ for numerous manifestations of the disorder and obtain the spectrum $E_{sc}(\omega)$ by Fourier transform.

We can obtain the probability distributions for both the real, $r = \mathrm{Re}[E_{sc}(\omega)]$, and the imaginary, $i = \mathrm{Im}[E_{sc}(\omega)]$, parts of the spectrum, $E_{sc}(\omega)$. As expected, the marginal distributions $P(r)$ and $P(i)$ are equivalent. From Ref. 12, they can be expressed as

$$P(a) = \frac{1}{\Delta\omega}\int_{\omega_1}^{\omega_2} \frac{1}{[\pi\langle I(\omega)\rangle]^{1/2}} \exp\left[-\frac{a^2}{\langle I(\omega)\rangle}\right] \mathrm{d}\omega, \quad (1)$$

where $a = \{r, i\}$ and where $\langle I(\omega)\rangle$ is the spectrum of the scattered field; the angle brackets indicate an average over all measured configurations of the random medium. For a narrowband incident field, the scattered field can be described as a sum of random phasors with uniformly distributed phase. From the central-limit theorem, the complex components of the scattered field are predicted to obey Gaussian statistics[17,18] with a variance that is proportional to the configurationally averaged intensity. In our earlier study we extended the monochromatic theory by introducing a frequency-dependent variance $\langle I(\omega)\rangle$ and integrating over the spectrum to obtain Eq. (1).

Although this extended model produced an accurate fit to the broadband data,[12] it is not an *a priori* result in the sense that it requires complete knowledge of the configurationally averaged intensity throughout the entire spectral range. This average, shown in Fig. 1 for our measured data, only be can estimated from the data themselves. $\langle I(\omega)\rangle$ is in general a very complicated function, depending on the details of the random medium and the power spectrum of the source. As a consequence, our original formalism lacks a closed-form expression and contains a large number of degrees of freedom.

To simplify the theory, we note that in Eq. (1) the evolution of $\langle I(\omega)\rangle$ with frequency is key to describing the probability distribution of the field. In terms of a

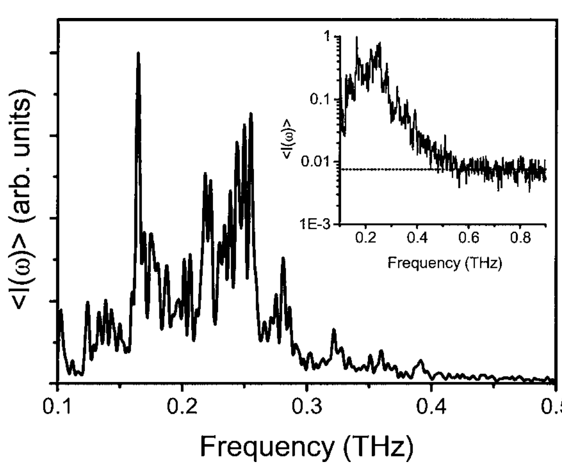

Fig. 1. Configurationally averaged intensity $\langle I(\omega)\rangle$. Inset, the same data on a log scale, normalized to unity at its peak. The horizontal line indicates the noise level, which we use to determine the value of the γ parameter in Eq. (3).

0146-9592/05/212843-3/$15.00

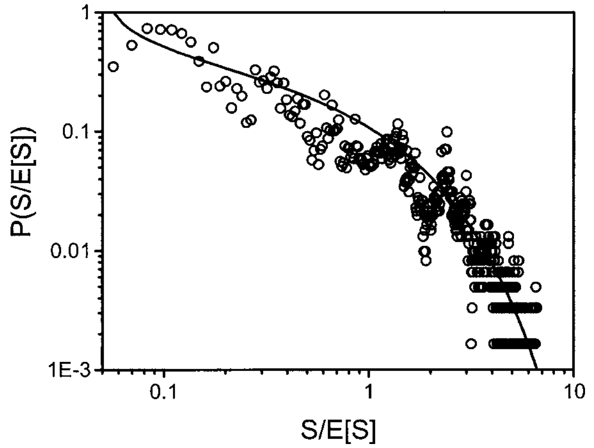

Fig. 2. Probability distribution $P(S)$ of the configurationally averaged intensity normalized to its expectation value $E[S]$, computed by taking a histogram of the estimate of $\langle I(\omega)\rangle$ shown in Fig. 1. Solid curve, a gamma distribution [Eq. (3)], with parameters $\alpha=0.689$, $\beta=1.38$, and $\gamma=0.05$, determined as described in the text.

Bayesian analysis,[19] the frequency parameter plays the role of the prior variable. As a result, this places an excessive significance on the frequency parameter as opposed to focusing on the distribution of intensities within the spectral range. It is this distribution that directly controls the weights of the Gaussians that are included in the integration in Eq. (1). We propose a new model that first considers $S=\langle I(\omega)\rangle$ as a random variable and views each frequency component as an independent observation of S. In a Bayesian framework, the marginal distributions can be formulated as

$$P(a)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{\pi S}}\exp\left(-\frac{a^2}{S}\right)P(S)\mathrm{d}S, \tag{2}$$

where $P(S)$ is the prior distribution of variable S. In this new formalism we no longer rely on knowledge of $\langle I(\omega)\rangle$ at each frequency but only on its probability distribution.

At this point it is not clear whether Eq. (2) is superior to Eq. (1). Only if there exists a closed-form expression for $P(S)$ would Eq. (2) be advantageous. Pictured in Fig. 2 is $P(S)$, the probability distribution of the configurationally averaged intensities, determined by taking a histogram of the estimate of $\langle I(\omega)\rangle$ shown in Fig. 1. We note that $P(S)$ resembles the well-known gamma distribution, which has a probability density function of the form

$$P(S)=\frac{1}{\beta\Gamma(\alpha)}\left(\frac{S-\gamma}{\beta}\right)^{\alpha-1}\exp\left(-\frac{S-\gamma}{\beta}\right), \tag{3}$$

where α, β, and γ are parameters that characterize the distribution, $\alpha>0$, $\beta>0$, $S>\gamma$, and $\Gamma(\alpha)$ is the gamma function.[20] It is possible to estimate α, β, and γ by using the data. γ is merely a shift along the horizontal axis, which we interpret as the effect of additive noise [since, in the absence of noise, $\langle I(\omega)\rangle\to 0$ for frequencies outside the source spectrum], while β acts as a scale factor. The mean and the second moment of the distribution are related to the α and β parameters according to

$$E[S]=\alpha\beta+\gamma, \tag{4}$$

$$E[S^2]=\alpha^2\beta^2+\gamma^2+\alpha\beta(\beta+2\gamma), \tag{5}$$

where $E[\ ]$ denotes the expectation of a random variable. For our situation, the expectations of S and S^2 are equivalent, respectively, to the mean values of $\langle I(\omega)\rangle$ and $\langle I(\omega)\rangle^2$ within the spectral range of the source. The solid curve in Fig. 2 is a gamma distribution with α, β, and γ estimated from $\langle I(\omega)\rangle$ and relations (4) and (5).

We note that, for a thermal light source, it is straightforward to demonstrate that the instantaneous intensity is a gamma variate.[18] However, it is difficult to draw a direct connection to the result presented here, because in our experiment the scattered wave does not satisfy the condition of stationarity, assumed for a thermal source. The use of a pulsed source necessarily defines a zero of time for the measured field and therefore ensures that it is nonstationary.[21] Indeed, we recently showed that the average spectral content of the scattered wave evolves throughout the temporal window of our measurement.[14] Nevertheless, the excellent fit to a gamma distribution shown in Fig. 2 is reminiscent of the prediction for a thermal light source.

The closed-form expression for prior distribution $P(S)$ can now be used to model the statistics of the complex parts of the electric field. We extract the complex parts of $E_{\mathrm{sc}}(\omega)$ over the 0.1–0.5 THz spectral range, where there is appreciable signal in the measured waveforms. In Fig. 3 we show the probability distributions of the real (triangles) and imaginary

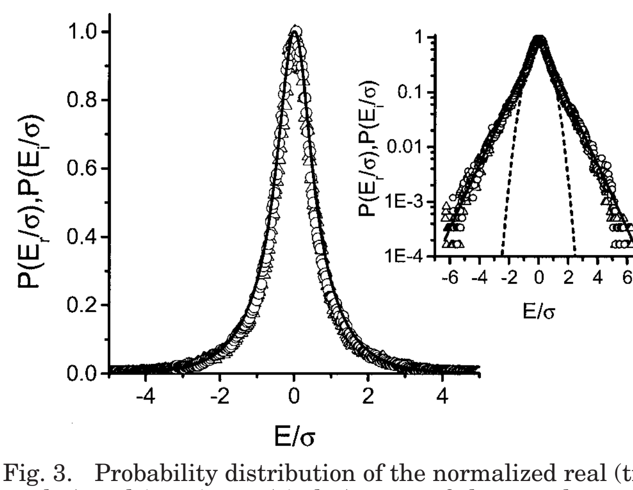

Fig. 3. Probability distribution of the normalized real (triangles) and imaginary (circles) parts of the complex scattered electric field, $P(r/\sigma)$ and $P(i/\sigma)$, normalized to their standard deviation σ. Solid curve, prediction of Eq. (2) with the gamma probability distribution function in Eq. (3) shown as the prior distribution. Inset, the same result plotted on a log scale. The predicted distribution agrees with the measurements over nearly four decades and over ±6 standard deviations in the field amplitude. The dashed curve is a best fit to a Gaussian distribution.

(circles) parts, $P(r/\sigma)$ and $P(i/\sigma)$, normalized by their standard deviation σ. As anticipated, the distribution of the field for broadband illumination does not conform to the Gaussian distribution predicted for monochromatic waves. The solid curve is computed by inserting the gamma prior distribution from Eq. (3) into the model of Eq. (2) and evaluating the integral. This model accurately describes the probability distribution of the field with the use of only three parameters: α, β, and γ. These are easily estimated from the noise level, mean, and second moment of the configurationally averaged intensity. We emphasize that the new formalism is much simpler than our original approach, which required knowledge of $\langle I(\omega)\rangle$ through the entire spectral range of the source.

In conclusion, we have developed a new model for the statistics of a random broadband field, taking note of the fact that distribution $P(a)$ is essentially a superposition of the Gaussian statistics of each individual frequency component, with a variance determined by the average intensity. We can treat this average intensity as a Bayesian prior, such that only its probability distribution function is required for a complete description of the statistics. Since this function is well described by a gamma distribution, we are able to greatly simplify the computation of the statistics. The parameters of this gamma distribution can easily be obtained from the noise and the first two moments of the configurationally averaged spectrum of the scattered field. Because of the nonstationary nature of the statistical process, the parameters α and β do not have an obvious physical meaning and in general may depend both on the spectrum of the source and on the properties of the random medium. However, the results shown in Fig. 3 suggest that the gamma distribution function, previously identified as applicable only for stationary sources, may be more generally relevant for nonstationary processes.[21]

We thank Richard Baraniuk and Hyeokho Choi for discussions. This research has been supported in part by the National Science Foundation. D. M. Mittleman's e-mail address is daniel@rice.edu.

*Present address, Shell International Exploration and Production, Inc., P.O. Box 481, Houston, Texas 77001.

References

1. M. A. O'Leary, D. A. Boas, B. Chance, and A. G. Yodh, Opt. Lett. **20**, 426 (1995).
2. D. A. Boas, L. E. Campbell, and A. G. Yodh, Phys. Rev. Lett. **75**, 1855 (1995).
3. D. A. Boas and A. G. Yodh, J. Opt. Soc. Am. A **14**, 192 (1997).
4. D. L. Lasocki, C. L. Matson, and P. J. Collins, Opt. Lett. **23**, 558 (1998).
5. J. D. McKinney, M. A. Webster, K. J. Webb, and A. M. Weiner, Opt. Lett. **25**, 4 (2000).
6. J. C. Schotland and V. A. Markel, J. Opt. Soc. Am. A **18**, 2767 (2001).
7. R. Berkovits and S. Feng, Phys. Rep. **238**, 135 (1994).
8. C. A. Thompson, K. J. Webb, and A. M. Weiner, Appl. Opt. **36**, 3726 (1997).
9. M. A. Webster, K. J. Webb, and A. M. Weiner, Phys. Rev. Lett. **88**, 033901 (2002).
10. B. A. van Tiggelen, P. Sebbah, M. Stoytchev, and A. Z. Genack, Phys. Rev. E **59**, 7166 (1999).
11. J. Rivas, R. Sprik, and A. Lagendijk, Phys. Rev. E **63**, 046613 (2001).
12. J. Pearce, Z. Jian, and D. Mittleman, Phys. Rev. Lett. **91**, 043903 (2003).
13. S. Mujumdar, K. J. Chau, and A. Y. Elezzabi, Appl. Phys. Lett. **85**, 6284 (2004).
14. J. Pearce, Z. Jian, and D. Mittleman, Opt. Lett. **29**, 2926 (2004).
15. K. J. Chau, G. D. Dice, and A. Y. Elezzabi, Phys. Rev. Lett. **94**, 173904 (2005).
16. J. Pearce and D. Mittleman, Opt. Lett. **26**, 2002 (2001).
17. A. A. Chabanov and A. Z. Genack, Phys. Rev. E **56**, R1338 (1997).
18. J. W. Goodman, *Statistical Optics* (Wiley, 2000).
19. A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis* (Chapman & Hall, 2003).
20. N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions* (Wiley, New York, 1994).
21. J. Pearce, K. Doyle, Z. Jian, J. Deibel, and D. M. Mittleman, "Non-stationary time-domain statistics of multiply scattered broadband terahertz pulses," submitted to Phys. Rev. E.