Maximum-likelihood parameter estimation in terahertz time-domain spectroscopy

LALEH MOHTASHEMI, PAUL WESTLUND, DEREK G. SAHOTA, GRAHAM B. LEA, IAN BUSHFIELD, PAYAM MOUSAVI, [ID] AND J. STEVEN DODGE[*] [ID]

Department of Physics, Simon Fraser University, V5A 1S6, Canada
[*]*jsdodge@sfu.ca*

Abstract: We present a maximum-likelihood method for parameter estimation in terahertz time-domain spectroscopy. We derive the likelihood function for a parameterized frequency response function, given a pair of time-domain waveforms with known time-dependent noise amplitudes. The method provides parameter estimates that are superior to other commonly used methods and provides a reliable measure of the goodness of fit. We also develop a simple noise model that is parameterized by three dominant sources and derive the likelihood function for their amplitudes in terms of a set of repeated waveform measurements. We demonstrate the method with applications to material characterization.

1. Introduction

At the heart of most applications of terahertz time-domain spectroscopy (THz-TDS) is a mathematical procedure that relates raw THz-TDS waveform measurements to parameters of scientific and technological interest [1–3]. Typically this analysis is formulated in the frequency domain, since it provides the most natural description of any linear, time-invariant system of interest. But since a THz-TDS waveform describes the electric field as a function of *time*, it must be Fourier-transformed for use in any frequency-domain analysis. The Fourier transform scrambles the THz-TDS measurement noise, which is more naturally represented in the time domain, and produces artifacts that can degrade the overall measurement quality and yield misleading results [4–13]. Furthermore, the standard approaches to parameter estimation in THz-TDS involve the ratio of two noisy waveforms, which further distorts the noise spectrum and can yield noise distributions that are far from Gaussian. Other approaches have emerged that improve on the standard procedures [14–16], but so far all of the existing approaches to THz-TDS analysis lack clear, model-independent statistical criteria for establishing parameter confidence intervals or for deciding whether a given physical model describes the data well or not. Here, we introduce a general framework to remedy this [17]. We describe a maximum-likelihood approach to THz-TDS analysis in which both the signal and the noise are treated explicitly in the time domain, together with a frequency-domain constraint between the input and the output signal. We show that this approach produces superior results to existing analysis methods.

2. Signal and noise in THz-TDS

A Monte Carlo simulation of an elementary THz-TDS analysis illustrates the basic problems that our method solves; it also allows us to develop notational conventions that we will use to describe our main results. Figure 1(a) shows an ideal, noiseless THz-TDS waveform $\mu(t)$ normalized to its peak value μ_p [18], together with a time-domain Gaussian noise amplitude $\sigma_\mu(t)$ given by

$$\sigma_\mu^2(t) = \sigma_\alpha{}^2 + [\sigma_\beta \mu(t)]^2 + [\sigma_\tau \dot{\mu}(t)]^2 \tag{1}$$

with amplitudes $\sigma_\alpha = 10^{-4}\mu_p$, $\sigma_\beta = 10^{-2}$, and $\sigma_\tau = 1$ fs. Physically, the additive noise term σ_α is produced by the detection electronics (in units of μ_p); the multiplicative noise term $\sigma_\beta|\mu(t)|$ is produced by laser power fluctuations, which modulate the signal strength; and the jitter term $\sigma_\tau|\dot{\mu}(t)| \approx |\mu(t + \sigma_\tau) - \mu(t)|$ is produced by fluctuations in the optical path lengths that control the delay between the terahertz pulse arrival time and the receiver sampling time. In this example the overall noise is dominated by $\sigma_\beta|\mu(t)|$, which produces the two peaks near $t = 2.75$ ps and $t = 3.50$ ps. The contribution from $\sigma_\tau|\dot{\mu}(t)|$ is less evident, except at $t = 3.10$ ps, where $\mu(t)$ crosses zero with a nonzero slope. The contribution from σ_α dominates at times when both $\mu(t)$ and $\dot{\mu}(t)$ are small, although in the figure it is barely discernible from zero. Such structured, signal-dependent, time-domain noise is common in THz-TDS waveform measurements, and leads to well-known ambiguities in defining the signal-to-noise ratio and dynamic range for them [5,6,8].
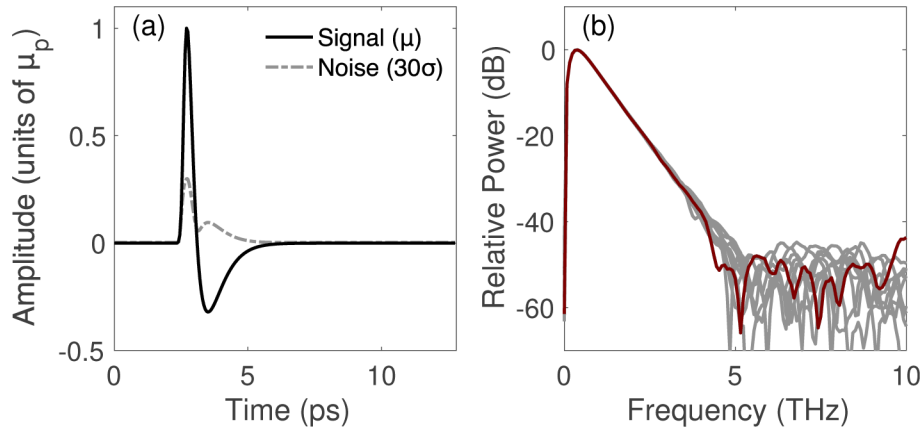

Fig. 1. (a) Simulated time-domain signal (black solid line) and noise amplitude (gray dashed line). (b) Power spectral density (relative to peak) obtained from ten simulated time-domain measurements, with one shown in red and the rest shown in gray.

We simulate a single THz-TDS waveform measurement by computing

$$x(t_k) = \mu(t_k) + \sigma_\mu(t_k)\epsilon(t_k) \tag{2}$$

at equally spaced times $t_k = kT$, $k \in \{0, 1, \ldots, N-1\}$, with $N = 256$ and $T = 50$ fs, where each $\epsilon(t_k)$ is an independent random variable with a standard normal distribution. This sequence has the discrete Fourier transform (DFT)

$$X(\omega_l) = \tilde{x}(\omega_l) = \tilde{\mu}(\omega_l) + [\tilde{\sigma}_\mu \circledast \tilde{\epsilon}](\omega_l) \tag{3}$$

at the discrete frequencies $\omega_l = 2\pi l/(NT)$, $l \in \{0, 1, \ldots, N-1\}$, where $\tilde{x}(\omega_l)$, $\tilde{\sigma}_\mu(\omega_l)$, and $\tilde{\epsilon}(\omega_l)$ denote the DFTs of $x(t_k)$, $\sigma_\mu(t_k)$, and $\epsilon(t_k)$, respectively, and the $\circledast$ symbol denotes circular discrete convolution. Figure 1(b) shows the power spectral density $S_{xx}(\omega_l) = (T/N)|X(\omega_l)|^2$ at the unique frequencies given by $l \leq \text{floor}(N/2) = \lfloor N/2 \rfloor$ for ten such simulations, where each spectrum is normalized to its peak value. The signal decreases exponentially with frequency, falling by a bit more than 40 dB from its peak power before reaching the noise floor near 5 THz. The red-highlighted spectrum in Fig. 1(b) shows that while the noise is relatively flat between 5 THz and 10 THz, it exhibits oscillatory fluctuations with a period of about 0.75 THz. These also appear in all of the other spectra in Fig. 1(b), and arise because the convolution $[\tilde{\sigma}_\mu \circledast \tilde{\epsilon}](\omega_l)$ smooths the noise over a frequency scale given by the inverse width of $\sigma_\mu(t)$. The resulting correlations blur the distinction between the signal and the noise, and their presence

implies that the true uncertainty in $X(\omega_l)$ is significantly smaller than the noise floor suggested by $S_{xx}(\omega_l)$, which neglects the information that noise at one frequency provides about the noise at neighbouring frequencies.

3.   Transfer function estimation in THz-TDS

To measure the response of a system requires measurements of two THz-TDS waveforms: an input, $\mu(t)$, and an output, $\psi(t) = [h*\mu](t)$, where $h(t)$ is the impulse response of the system and the $*$ symbol denotes continuous-time convolution. Fourier transforming the ideal relationship $\psi(t) = [h*\mu](t)$ converts the time-domain convolution to a frequency-domain multiplication, $\tilde{\psi}_c(\omega) = H(\omega)\tilde{\mu}_c(\omega)$, where $\tilde{\mu}_c(\omega)$ and $\tilde{\psi}_c(\omega)$ denote the continuous-time Fourier transforms of $\mu(t)$ and $\psi(t)$, respectively, and $H(\omega) = \tilde{h}_c(\omega)$ is the continuous-time transfer function of the system. Proceeding as we did for a single waveform, we simulate input and output waveform measurements $x(t_k)$ and $y(t_k)$, respectively, by computing $\mu(t)$ and $\psi(t)$ at the discrete times t_k and adding time-domain noise $\sigma_\mu(t_k)\epsilon(t_k)$ and $\sigma_\psi(t_k)\delta(t_k)$, where each $\epsilon(t_k)$ and $\delta(t_k)$ is an independent random variable with a standard normal distribution. After applying the discrete Fourier transform to obtain $X(\omega_l) = \tilde{x}(\omega_l)$ and $Y(\omega_l) = \tilde{y}(\omega_l)$, we can determine the *empirical transfer function estimate* (ETFE) [19],

$$\hat{H}_{\mathrm{ETFE}}(\omega_l) = \frac{Y(\omega_l)}{X(\omega_l)} = \frac{\tilde{\psi}(\omega_l) + [\tilde{\sigma}_\psi \circledast \tilde{\epsilon}](\omega_l)}{\tilde{\mu}(\omega_l) + [\tilde{\sigma}_\mu \circledast \tilde{\delta}](\omega_l)},\tag{4}$$

where $\tilde{\mu}(\omega_l)$ and $\tilde{\psi}(\omega_l)$ are the DFTs of the ideal input and output signals, respectively, that we would obtain in the absence of noise, and $[\tilde{\sigma}_\mu \circledast \tilde{\epsilon}](\omega_l)$ and $[\tilde{\sigma}_\psi \circledast \tilde{\delta}](\omega_l)$ are the corresponding frequency-domain noise terms. The ETFE is a common starting point for THz-TDS analysis—it is easy to compute from the raw data, and it provides an estimate of $H(\omega)$ for any linear, time-invariant system. Frequently, though, the focus of interest is not $H(\omega)$ itself but a relatively small number of parameters that specify $H(\omega)$ through a physical model, such as the thickness and refractive index of a material. In this case, fitting the model directly to $\hat{H}_{\mathrm{ETFE}}(\omega_l)$ can yield ambiguous results, because $\hat{H}_{\mathrm{ETFE}}(\omega_l)$ is biased and has noise that is both correlated and non-Gaussian.

To illustrate these deficiencies, Fig. 2 shows 250 simulations of $\hat{H}_{\mathrm{ETFE}}(\omega_l)$ with nominally identical input and output pulses, corresponding to $H(\omega) = 1$. The red-highlighted curves show correlations similar to those shown in Fig. 1, but that extend to frequencies where the signal is much stronger. The average over all simulations reveals the bias in $\hat{H}_{\mathrm{ETFE}}(\omega_l)$: in Fig. 2(a), $\mathrm{Re}\{\hat{H}_{\mathrm{ETFE}}\}$ falls from the correct value of 1 to the biased value of 0 as the frequency increases above 5 THz, where the signal reaches the noise floor in Fig. 1(b). Also, the noise distribution for $\hat{H}_{\mathrm{ETFE}}(\omega_l)$ develops wide, non-Gaussian tails at frequencies where the signal is small, because noise fluctuation that cause $|X(\omega_l)| \to 0$ become more likely, and these in turn cause $\hat{H}_{\mathrm{ETFE}}(\omega_l)$ to diverge [20]. If the noise is uncorrelated, such large fluctuations are only expected when the signal is small. But in the more typical case that the noise is correlated, they can influence frequencies well within the primary signal bandwidth, as indicated by the red-highlighted curves in Figs. 2(c) and 2(d).

Figure 3 shows how these problems distort standard measures of fit quality. It displays the results of weighted least-squares fits to the ETFE simulations in Fig. 2 with the model

$$H_1(\omega; \boldsymbol{\theta}) = \theta_1 \exp(i\omega\theta_2),\tag{5}$$

which rescales the input by an amplitude θ_1 and shifts it by a delay θ_2 when we adopt the $\exp(-i\omega t)$ convention for harmonic time dependence. For each fit, the estimated parameter vector
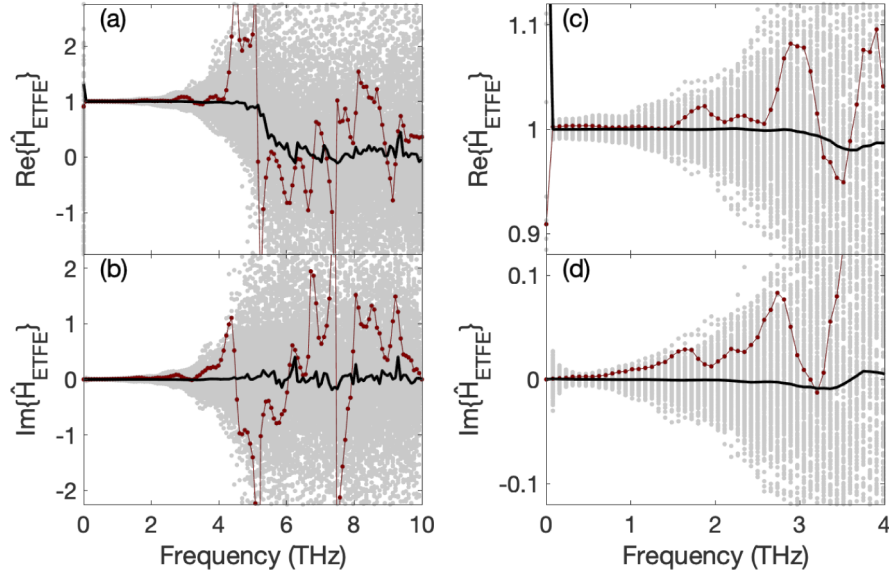
Fig. 2. Gray dots show the real (a,c) and imaginary parts (b,d) of the empirical transfer function estimate $\hat{H}_{\text{ETFE}}$, Eq. (4), for 250 pairs of simulated time-domain measurements of the waveform shown in Fig. 1(a). One estimate is highlighted in red, with the dots connected by a thin red line. The solid black line shows the average over all simulations. Panels (a,b) show the full bandwidth and (c,d) show the same data over the primary signal bandwidth.

$\hat{\boldsymbol{\theta}}$ is chosen to minimize

$$Q_{\text{ETFE}}(\boldsymbol{\theta}) = \sum_{l=0}^{N-1} \frac{\left|\hat{H}_{\text{ETFE}}(\omega_l) - H_1(\omega_l; \boldsymbol{\theta})\right|^2}{\sigma_{\omega_l}^2}, \tag{6}$$

where each $\sigma_{\omega_l}^2 = (\text{Var}\{\text{Re}[\hat{H}_{\text{ETFE}}(\omega_l)]\} + \text{Var}\{\text{Im}[\hat{H}_{\text{ETFE}}(\omega_l)]\})$ is determined from the Monte Carlo samples. These fits return accurate estimates for the model parameters—over all simulated samples, we obtain $\hat{\theta}_1 = 1.000 \pm 0.005$ and $\hat{\theta}_2 = (0.0 \pm 0.8)$ fs, which are consistent with the true values, $\theta_1 = 1$ and $\theta_2 = 0$, used in the simulation. But the normalized fit residuals, given by

$$r_{\text{ETFE}}(\omega_l; \hat{\boldsymbol{\theta}}) = \frac{\hat{H}_{\text{ETFE}}(\omega_l) - H_1(\omega_l; \hat{\boldsymbol{\theta}})}{\sigma_{\omega_l}}, \tag{7}$$

show strong deviations from the standard normal distribution, exhibit structure that could easily be mistaken for real physical features, and yield a goodness of fit (GOF) statistic $S_{\text{ETFE}} = Q_{\text{ETFE}}(\hat{\boldsymbol{\theta}})$ that looks nothing like the χ^2 distribution that we would normally use to evaluate the fit quality. Also, the uncertainty estimates, $\hat{\sigma}_{\theta_1} = 0.001$ and $\hat{\sigma}_{\theta_2} = 0.2$ fs, obtained by the usual method of inverting the curvature matrix of $Q_{\text{ETFE}}(\boldsymbol{\theta})$ around the minimum of each fit, significantly underestimate the values that are actually observed over the set of Monte Carlo samples, $\sigma_{\theta_1} = 0.005$ and $\sigma_{\theta_2} = 0.8$ fs. In short, while an ETFE fit with Eq. (6) may provide accurate parameter estimates when the underlying model is known in advance, it offers poor discrimination between good models and bad ones, and it yields parameter uncertainty estimates that are unrealistically precise.
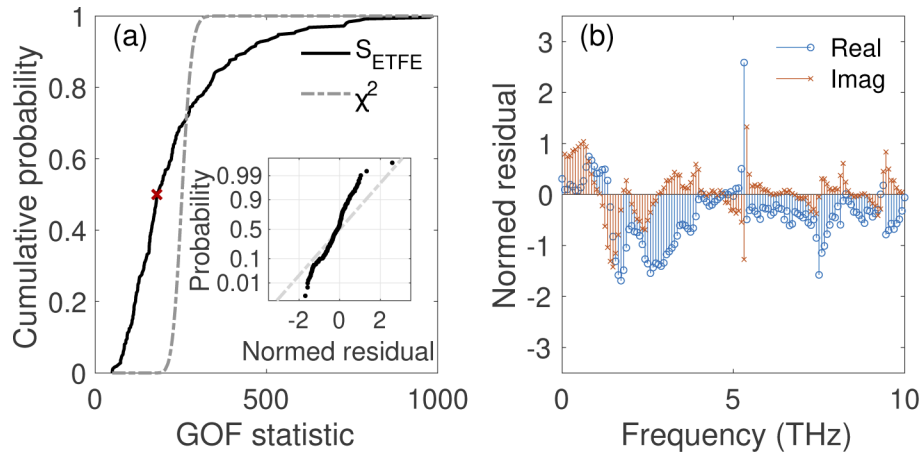
Fig. 3. Measures of fit quality for ETFE fits with Eq. (5), obtained by minimizing Eq. (6). (a) Cumulative distribution of the GOF statistic, S_{ETFE}, for the Monte Carlo simulations shown in Fig. 2. The χ^2 distribution for the same number of degrees of freedom ($\nu = 254$) is shown for comparison. The red $\times$ marks a fit with $S_{\text{ETFE}} \approx 180$, which is just above the median value. The normalized residuals for this fit, $r_{\text{ETFE}}(\omega_l; \hat{\boldsymbol{\theta}})$, are shown in (b) as a function of frequency, and in the inset to (a) with a normal probability plot (black dots, which represent both the real and the imaginary parts of r_{ETFE}). The gray dash-dotted line in the inset represents the standard normal distribution.

An alternative approach is to use a least-squares (LS) procedure that minimizes the sum of the squared differences between the output signal and the transformed input signal [14–16],

$$Q_{\text{LS}}(\boldsymbol{\theta}) = \sum_{l=0}^{N-1} |Y(\omega_l) - H(\omega_l; \boldsymbol{\theta}) X(\omega_l)|^2 = \sum_{k=0}^{N-1} \{y(t_k) - [h(\boldsymbol{\theta}) \circledast x](t_k)\}^2 . \qquad (8)$$

Here, $h(t_k; \boldsymbol{\theta})$ is the model impulse response, equal to the inverse DFT of $H(\omega_l; \boldsymbol{\theta})$, and $[h(\boldsymbol{\theta}) \circledast x](t_k)$ represents the convolution of impulse responses with the input vector at the time t_k. The equivalence between the frequency-domain sum and the time-domain sum is assured by Parseval's theorem. The LS procedure avoids the division by $X(\omega_l)$ that distorts the noise distribution of the ETFE in the small-signal limit, and the time-domain residuals $r_{\text{LS}}(t_k; \hat{\boldsymbol{\theta}}) = y(t_k) - [h(\boldsymbol{\theta}) \circledast x](t_k)$ are a sensitive indicator of fit quality. But the uniform weighting in Eq. (8) implicitly assumes that this noise is constant in time, which as we have noted is frequently not the case.

In principle, we should be able to account for time-dependent noise by assigning appropriate weights to the sum, but this is not as simple as it might seem. The problem is that Eq. (8) treats the input and output noise asymmetrically, which we can see more clearly if we express it explicitly in terms of the time-domain signal and noise sequences:

$$Q_{\text{LS}}(\boldsymbol{\theta}) = \sum_{k=0}^{N-1} \left[\psi(t_k) - [h(\boldsymbol{\theta}) \circledast \mu](t_k) + \sigma_\psi(t_k) - [h(\boldsymbol{\theta}) \circledast (\sigma_\mu \epsilon)](t_k) \right]^2 . \qquad (9)$$

This asymmetry results from the implicit assumption that all noise is restricted to the output variable in an LS fit, so that the term $[h(\boldsymbol{\theta}) \circledast (\sigma_\mu \epsilon)](t_k)$ can be ignored. When input noise is present—as it always is in THz-TDS—the optimal fit weights will depend not just on the measurement noise sequences but also on $h(t_k; \boldsymbol{\theta})$. Any modification of Eq. (8) with constant weights will generically bias the parameter estimates toward values that minimize the input noise

contribution, and will cause the GOF statistic, $S_{\mathrm{LS}} = Q_{\mathrm{LS}}(\hat{\boldsymbol{\theta}})$, to have a distribution that also depends on $h(t_k; \boldsymbol{\theta})$. As we discuss below, these problems can all be solved with a fit procedure derived from the maximum-likelihood principle.

4. Maximum-likelihood estimation of a parameterized transfer function model

The likelihood function is equivalent to the probability density for obtaining the measured data under the assumptions of a given model, and forms the theoretical basis for the standard least-squares fit procedure. To define it in the THz-TDS context, we express the measured input and output signals in vector notation as $\boldsymbol{x} = [x(t_0), x(t_1), \ldots, x(t_{N-1})]^{\mathsf{T}}$ and $\boldsymbol{y} = [y(t_0), y(t_1), \ldots, y(t_{N-1})]^{\mathsf{T}}$, respectively, and assume that they represent noisy measurements of bandlimited but otherwise unknown ideal signal vectors $\boldsymbol{\mu} = [\mu(t_0), \mu(t_1), \ldots, \mu(t_{N-1})]^{\mathsf{T}}$ and $\boldsymbol{\psi} = [\psi(t_0), \psi(t_1), \ldots, \psi(t_{N-1})]^{\mathsf{T}}$ with noise covariance matrices $\boldsymbol{\Sigma_x}$ and $\boldsymbol{\Sigma_y}$, respectively. We further assume that $\boldsymbol{\mu}$ and $\boldsymbol{\psi}$ satisfy the linear constraint $\boldsymbol{\psi} = \boldsymbol{h}(\boldsymbol{\theta})\boldsymbol{\mu}$, where $\boldsymbol{h}$ is a known matrix function of an unknown p-dimensional parameter vector $\boldsymbol{\theta}$. The likelihood function is then a product of two multivariate Gaussian distributions,

$$L(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\psi}; \boldsymbol{x}, \boldsymbol{y}) = \frac{\exp[-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathsf{T}} \boldsymbol{\Sigma_x}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})]}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma_x})}} \frac{\exp[-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\psi})^{\mathsf{T}} \boldsymbol{\Sigma_y}^{-1}(\boldsymbol{y} - \boldsymbol{\psi})]}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma_y})}}, \quad (10)$$

where the dependence on $\boldsymbol{\theta}$ on the right side of the expression is implicit in the constraint between $\boldsymbol{\mu}$ and $\boldsymbol{\psi}$. The signal vectors $\boldsymbol{\mu}$ and $\boldsymbol{\psi}$ appear as arguments in the likelihood function but are otherwise unimportant—the statistical literature aptly describes these as *nuisance parameters*, which we must eliminate to estimate $\boldsymbol{\theta}$, the parameters of interest [21].

We can use discrete-time Fourier analysis to obtain explicit expressions for $\boldsymbol{h}(\boldsymbol{\theta})$, $\boldsymbol{\Sigma_x}$, and $\boldsymbol{\Sigma_y}$. The time-domain transfer matrix $\boldsymbol{h}(\boldsymbol{\theta})$ is

$$h_{jk}(\boldsymbol{\theta}) = \frac{1}{N} \sum_{l=-N/2+1}^{N/2-1} H(\omega_l; \boldsymbol{\theta}) \exp[-2\pi i(j-k)l/N] + \frac{1}{N} \mathrm{Re}[H(\omega_{N/2}; \boldsymbol{\theta})] \cos[\pi(j-k)], \quad (11)$$

where ω_l now extends to negative values of l, and N is assumed even; for odd N, the sum runs from $l = -(N-1)/2$ to $(N-1)/2$ and the anomalous term at the Nyquist frequency is absent. Similarly, we can define the discrete-time matrix representation of the time-derivative operator, $\boldsymbol{D}$, as

$$D_{jk} = \frac{1}{N} \sum_{l=-\lfloor(N-1)/2\rfloor}^{\lfloor(N-1)/2\rfloor} -i\omega_l \exp[-2\pi i(j-k)l/N], \quad (12)$$

by recognizing that the transfer function for the continuous time-derivative operator is $H(\omega) = -i\omega$. Note that Eq. (12) is valid for all values of N because $\mathrm{Re}[-i\omega_{N/2}] = 0$, so the anomalous term in Eq. (11) is zero. Equation (12) allows us to express the noise variance function $\sigma_\mu^2(t)$ in Eq. (1) as a matrix function in discrete time,

$$[\boldsymbol{V}(\boldsymbol{\mu})]_{jk} = \delta_{jk} \left[\sigma_\alpha{}^2 + \sigma_\beta{}^2 \mu_k^2 + \sigma_\tau{}^2 (\boldsymbol{D}\boldsymbol{\mu})_k^2 \right]. \quad (13)$$

The covariance matrices for $\boldsymbol{x}$ and $\boldsymbol{y}$ are then

$$\boldsymbol{\Sigma_x} = \boldsymbol{V}(\boldsymbol{\mu}), \quad \boldsymbol{\Sigma_y} = \boldsymbol{V}(\boldsymbol{\psi}). \quad (14)$$

The maximum-likelihood (ML) estimate for the model parameters is given by the vectors $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{\psi}}$, and $\hat{\boldsymbol{\theta}}$ that maximize L subject to the constraint $\boldsymbol{\psi} = \boldsymbol{h}(\boldsymbol{\theta})\boldsymbol{\mu}$. Alternatively, we can minimize the

negative-log likelihood,

$$
\begin{aligned}
-\ln L(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\psi}; \boldsymbol{x}, \boldsymbol{y}) = {} & N \ln(2\pi) + \frac{1}{2} \ln\{\det[V(\boldsymbol{\mu})]\} + \frac{1}{2} \ln\{\det[V(\boldsymbol{\psi})]\} \\
& + \frac{1}{2} \left\{ (\boldsymbol{x} - \boldsymbol{\mu})^{\mathsf{T}} [V(\boldsymbol{\mu})]^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{y} - \boldsymbol{\psi})^{\mathsf{T}} [V(\boldsymbol{\psi})]^{-1} (\boldsymbol{y} - \boldsymbol{\psi}) \right\},
\end{aligned}
\tag{15}
$$

also subject to the constraint, where the last term now has the familiar least-squares form. The dependence of the covariance matrices on the signal vectors prevents us from using a standard least-squares optimization algorithm to minimize Eq. (15), but to a very good approximation we can substitute $V(\boldsymbol{\mu}) \approx V(\boldsymbol{x})$ and $V(\boldsymbol{\psi}) \approx V(\boldsymbol{y})$ to obtain

$$
\begin{aligned}
-\ln L(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\psi}; \boldsymbol{x}, \boldsymbol{y}) \approx {} & N \ln(2\pi) + \frac{1}{2} \ln\{\det[V(\boldsymbol{x})]\} + \frac{1}{2} \ln\{\det[V(\boldsymbol{y})]\} \\
& + \frac{1}{2} \left\{ (\boldsymbol{x} - \boldsymbol{\mu})^{\mathsf{T}} [V(\boldsymbol{x})]^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{y} - \boldsymbol{\psi})^{\mathsf{T}} [V(\boldsymbol{y})]^{-1} (\boldsymbol{y} - \boldsymbol{\psi}) \right\}.
\end{aligned}
\tag{16}
$$

The first three terms are now all constant, so we can focus on minimizing the last term, which we multiply by two to obtain a cost function in what is known as a total least-squares form [22],

$$
\tilde{Q}_{\mathrm{TLS}}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\psi}) = (\boldsymbol{x} - \boldsymbol{\mu})^{\mathsf{T}} [V(\boldsymbol{x})]^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) + (\boldsymbol{y} - \boldsymbol{\psi})^{\mathsf{T}} [V(\boldsymbol{y})]^{-1} (\boldsymbol{y} - \boldsymbol{\psi}),
\tag{17}
$$

which is also subject to the constraint $\boldsymbol{\psi} = \boldsymbol{h}(\boldsymbol{\theta})\boldsymbol{\mu}$. Note that the total least-squares cost function treats the input and output variables symmetrically, unlike the conventional least-squares cost function in Eq. (8). Geometrically, it can be interpreted as the sum of the squared distances between each measurement point (x_k, y_k) and its associated model point (μ_k, ψ_k), using the metric tensors $V(\boldsymbol{x})$ and $V(\boldsymbol{y})$, respectively, for the input and output variables. As discussed in Sec. 3, the LS cost function in Eq. (8) focuses entirely on the deviations in the output variable.

Introducing an N-dimensional vector of Lagrange parameters $\boldsymbol{\lambda}$ to implement the constraint, we can define the modified cost function,

$$
\tilde{\tilde{Q}}_{\mathrm{TLS}}(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\lambda}) = \boldsymbol{\lambda}^{\mathsf{T}} [\boldsymbol{\psi} - \boldsymbol{h}(\boldsymbol{\theta})\boldsymbol{\mu}] + \tilde{Q}_{\mathrm{TLS}}(\boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\theta}),
\tag{18}
$$

which we may now minimize with respect to $\boldsymbol{\mu}$, $\boldsymbol{\psi}$, $\boldsymbol{\theta}$, and $\boldsymbol{\lambda}$. The parameters of interest are $\boldsymbol{\theta}$; minimizing with respect to the remaining parameters gives the equations

$$
\left. \frac{\partial \tilde{\tilde{Q}}_{\mathrm{TLS}}}{\partial \lambda_m} \right|_{\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}}, \boldsymbol{\theta}} = \hat{\psi}_m - \sum_{l=1}^{N} h_{ml}(\boldsymbol{\theta}) \hat{\mu}_l = 0,
\tag{19}
$$

$$
\left. \frac{\partial \tilde{\tilde{Q}}_{\mathrm{TLS}}}{\partial \psi_m} \right|_{\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}}, \boldsymbol{\theta}} = \hat{\lambda}_m - 2 \sum_{l=1}^{N} \left\{ [V(\boldsymbol{y})]^{-1} \right\}_{ml} (y_l - \hat{\psi}_l) = 0,
\tag{20}
$$

$$
\left. \frac{\partial \tilde{\tilde{Q}}_{\mathrm{TLS}}}{\partial \mu_m} \right|_{\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\psi}}, \hat{\boldsymbol{\lambda}}, \boldsymbol{\theta}} = -\sum_{l=1}^{N} \hat{\lambda}_l h_{lm}(\boldsymbol{\theta}) - 2 \sum_{l=1}^{N} \left\{ [V(\boldsymbol{x})]^{-1} \right\}_{ml} (x_l - \hat{\mu}_l) = 0,
\tag{21}
$$

which have solutions

$$
\hat{\boldsymbol{\psi}}_{\boldsymbol{\theta}} = \boldsymbol{h}(\boldsymbol{\theta}) \hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}},
\tag{22}
$$

$$
\hat{\boldsymbol{\lambda}}_{\boldsymbol{\theta}} = 2 [V(\boldsymbol{y})]^{-1} (\boldsymbol{y} - \hat{\boldsymbol{\psi}}_{\boldsymbol{\theta}}),
\tag{23}
$$

$$
\hat{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = \left\{ \mathbf{1} + V(\boldsymbol{x}) [\boldsymbol{h}(\boldsymbol{\theta})]^{\mathsf{T}} [V(\boldsymbol{y})]^{-1} \boldsymbol{h}(\boldsymbol{\theta}) \right\}^{-1} \left\{ \boldsymbol{x} + V(\boldsymbol{x}) [\boldsymbol{h}(\boldsymbol{\theta})]^{\mathsf{T}} [V(\boldsymbol{y})]^{-1} \boldsymbol{y} \right\}.
\tag{24}
$$

Evaluating $\tilde{Q}_{\text{TLS}}$ at $\mu = \hat{\mu}_{\theta}$, $\psi = \hat{\psi}_{\theta}$, and $\lambda = \hat{\lambda}_{\theta}$ yields a simplified cost function that involves only the parameter vector θ and the data vectors x and y,

$$Q_{\text{TLS}}(\theta) = (x - \hat{\mu}_{\theta})^{\mathsf{T}} [V(x)]^{-1}(x - \hat{\mu}_{\theta}) + (y - \hat{\psi}_{\theta})^{\mathsf{T}} [V(y)]^{-1}(y - \hat{\psi}_{\theta}). \tag{25}$$

We can simplify these expressions further by defining

$$U(x, \theta) = h(\theta)V(x) [h(\theta)]^{\mathsf{T}}. \tag{26}$$

Substituting Eq. (26) in Eqs. (24) and (22) reveals that $\hat{\psi}_{\theta}$ is just the weighted average of y and $h(\theta)x$,

$$\hat{\psi}_{\theta} = \left\{ [V(y)]^{-1} + [U(x, \theta)]^{-1} \right\}^{-1} \left\{ [V(y)]^{-1}y + [U(x, \theta)]^{-1}[h(\theta)x] \right\}, \tag{27}$$

where the matrices $U(x, \theta)$ and $V(y)$ are approximately equal to the true (but unknown) covariance matrices $\text{Cov}[h(\theta)x] = U(\mu, \theta)$ and $\text{Cov}(y) = V(\psi)$, respectively. We emphasize here that although $h(\theta)\mu = \psi$, the covariance matrices $\text{Cov}[h(\theta)x]$ and $\text{Cov}(y)$ are generally different, since $h(\theta)$ transforms the noise in x as well as the signal. For example, if $h(A) = AI, A \neq 1$, and the noise is purely additive, we have $\text{Cov}(y) = \sigma_{\alpha}^2 I \neq \text{Cov}[h(\theta)x] = A^2 \sigma_{\alpha}^2 I$.

Substituting Eq. (26) in Eq. (25) yields

$$Q_{\text{TLS}}(\theta) = [y - h(\theta)x]^{\mathsf{T}} [V(y) + U(x, \theta)]^{-1} [y - h(\theta)x]. \tag{28}$$

Like the simpler LS cost function in Eq. (8), the TLS cost function in Eq. (28) is expressed in terms of the deviations $y - h(\theta)x$, but is now properly normalized with respect to the associated covariance matrix, $V(y) + U(x, \theta)$. Introducing the matrix square root, $A^{1/2} = X \leftrightarrow A = X^2$, we may define the normalized residual vector as

$$r_{\text{TLS}}(\theta) = [V(y) + U(x, \theta)]^{-1/2} [y - h(\theta)x], \tag{29}$$

which allows us to express the TLS cost function in the compact form,

$$Q_{\text{TLS}}(\theta) = [r_{\text{TLS}}(\theta)]^{\mathsf{T}} r_{\text{TLS}}(\theta). \tag{30}$$

The TLS estimate for the parameter vector, $\hat{\theta}$, may now be determined by minimizing $Q_{\text{TLS}}(\theta)$ using a standard least-squares optimization algorithm.

Figure 4 shows fit results for the same model and data shown in Fig. 3, but obtained by minimizing $Q_{\text{TLS}}(\theta)$ in Eq. (30) instead of $Q_{\text{ETFE}}(\theta)$ in Eq. (6). The statistical properties obtained with $Q_{\text{TLS}}(\theta)$ are clearly superior to those with $Q_{\text{ETFE}}(\theta)$. The GOF statistic, $S_{\text{TLS}} = Q_{\text{TLS}}(\hat{\theta})$, approximates a χ^2 distribution with $\nu = N - p$ degrees of freedom, as expected. The normalized residual vector $r_{\text{TLS}}(\hat{\theta})$, shown for one of the fits, exhibits a standard normal distribution with no discernible correlations among neighboring time points. The distribution for S_{TLS} is more than 7 times narrower than the distribution for S_{ETFE}, making it that much more sensitive to a poor fit. Similarly, the lack of structure in $r_{\text{TLS}}(\hat{\theta})$ makes it more useful for residual analysis than $r_{\text{ETFE}}(\omega_l; \hat{\theta})$, since structure associated with poor fits may be discerned more readily. Finally, unlike the ETFE fits, the TLS fits yield estimates for the parameter uncertainties that agree with the values observed over the set of Monte Carlo samples, $\sigma_{\theta_1} = 0.002$ and $\sigma_{\theta_2} = 0.4$ fs, which are both about a factor of 2 smaller than the values observed in the ETFE parameter estimates. In summary, when compared with the standard ETFE fit procedure, the TLS fit procedure offers better discrimination between good models and bad ones, more precise values for the parameters, and more accurate estimates of the parameter uncertainties.
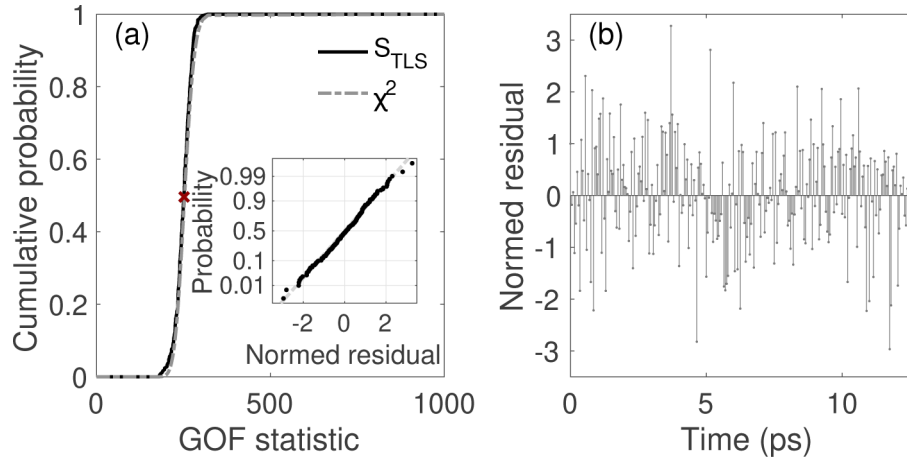
Fig. 4. Measures of fit quality for TLS fits with Eq. (5), obtained by minimizing Eq. (30); compare with Fig. 3. (a) Cumulative distribution of the GOF statistic, S_{TLS}, for the Monte Carlo simulations shown in Fig. 2. The χ^2 distribution for the same number of degrees of freedom ($\nu = 254$) is shown for comparison. The red × marks a fit with $S_{\text{TLS}} \approx 252$, which is just above the median value. The normalized residuals for this fit, $\boldsymbol{r}_{\text{TLS}}(\hat{\boldsymbol{\theta}})$, are shown in (b) as a function of time, and in the inset to (a) with a normal probability plot (black dots). The gray dash-dotted line in the inset represents the standard normal distribution.

5. Maximum-likelihood estimation of the noise model

We have assumed so far that the noise amplitudes $\sigma_\alpha, \sigma_\beta$, and σ_τ in Eq. (13) are known, but in general they must also be determined experimentally. One way to do this is to measure the noise at three different points on the THz waveform: as we saw in Fig. 1, the σ_β term dominates near the peak, the σ_τ term dominates where the signal crosses zero with a large slope, and the σ_α term dominates where both the signal and its slope are small. Another approach is to determine the variance as a function of time from a set of repeated waveform measurements [5], then obtain the noise parameters from a fit with Eq. (1). But both of these methods are susceptible to systematic errors from drift, which produces excess signal variability over the timescale of multiple measurements [8,9,23]. In this section, we describe a likelihood method for estimating the noise parameters that accounts for this drift.

We consider repeated measurements of an ideal primary waveform, $\mu(t)$, which has an amplitude and a delay that drift on a timescale longer than the waveform acquisition time. We can then associate each measurement with an ideal secondary waveform,

$$\zeta(t; A_l, \eta_l) = A_l \mu(t - \eta_l), \tag{31}$$

where A_l is the relative amplitude, η_l is the delay, and $l \in \{0, 1, \ldots, M - 1\}$. We also set $A_0 = 1$ and $\eta_0 = 0$, so that $\zeta(t; A_0, \eta_0) = \mu(t)$. As before, we sample these waveforms at the nominal times $t_k = kT, k \in \{0, 1, \ldots, N - 1\}$ to obtain the ideal primary waveform vector $\boldsymbol{\mu} = [\mu_0, \mu_1, \ldots, \mu_{N-1}]^{\mathsf{T}}$ and M measurement vectors $\boldsymbol{x}_l = [x_l(t_0), x_l(t_1), \ldots, x_l(t_{N-1})]^{\mathsf{T}}$, which we can arrange in an $N \times M$ matrix $\boldsymbol{X} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N-1}]$. We also write the amplitudes and delays in vector form, $\boldsymbol{A} = [A_0, A_1, \ldots, A_{M-1}]^{\mathsf{T}}$ and $\boldsymbol{\eta} = [\eta_0, \eta_1, \ldots, \eta_{M-1}]^{\mathsf{T}}$, respectively.

With these definitions, we can express the sampled secondary waveforms in vector form,

$$\boldsymbol{\zeta}_l = \boldsymbol{\zeta}(\boldsymbol{\mu}, A_l, \eta_l) = A_l \boldsymbol{S}(\eta_l) \boldsymbol{\mu}, \tag{32}$$

where the matrix $S(\eta_l)$ is defined to impart a delay by η_l. Using Eq. (11) with the transfer function $H(\omega; \eta) = \exp(i\omega\eta)$, we can write this matrix explicitly as

$$S_{jk}(\eta) = \frac{1}{N} \sum_{l=-N/2+1}^{N/2-1} \exp(i\omega_l\eta) \exp[-2\pi i(j-k)l/N] + \frac{1}{N} \cos(\omega_{N/2}\eta) \cos[\pi(j-k)] \qquad (33)$$

for even N, with changes for odd N as described for Eq. (11). Following Eqs. (13) and (14), and arranging the secondary waveform vectors in an $N \times M$ matrix $\mathbf{Z} = [\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_{M-1}]$, we also have

$$[\mathrm{Cov}(\boldsymbol{x}_l)]_{jk} = [\boldsymbol{V}(\boldsymbol{\zeta}_l; \sigma_\alpha, \sigma_\beta, \sigma_\tau)]_{jk} = \delta_{jk} \left[\sigma_\alpha{}^2 + \sigma_\beta{}^2 Z_{kl}^2 + \sigma_\tau{}^2(\boldsymbol{DZ})_{kl}^2 \right], \qquad (34)$$

where $\boldsymbol{D}$ is defined in Eq. (12), $\boldsymbol{\zeta}_l$ and $\boldsymbol{Z}$ depend implicitly on $\boldsymbol{A}$ and $\boldsymbol{\eta}$, and the dependence of $\boldsymbol{V}$ on the noise amplitudes is now expressed explicitly.

The likelihood function for observing $\boldsymbol{X}$ under these assumptions is

$$L(\sigma_\alpha, \sigma_\beta, \sigma_\tau, \boldsymbol{\mu}, \boldsymbol{A}, \boldsymbol{\eta}; \boldsymbol{X}) = \prod_{l=0}^{M-1} \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{x}_l - \boldsymbol{\zeta}_l)^\mathsf{T} [\boldsymbol{V}(\boldsymbol{\zeta}_l; \sigma_\alpha, \sigma_\beta, \sigma_\tau)]^{-1}(\boldsymbol{x}_l - \boldsymbol{\zeta}_l)\right\}}{\sqrt{(2\pi)^N \det[\boldsymbol{V}(\boldsymbol{\zeta}_l; \sigma_\alpha, \sigma_\beta, \sigma_\tau)]}}, \qquad (35)$$

in which the noise amplitudes σ_α, σ_β and σ_τ are the parameters of interest and the signal vectors $\boldsymbol{\mu}, \boldsymbol{A}$ and $\boldsymbol{\eta}$ are nuisance parameters. As with Eq. (10), it is more convenient computationally to work with the negative-log likelihood,

$$-\ln[L(\sigma_\alpha, \sigma_\beta, \sigma_\tau, \boldsymbol{\mu}, \boldsymbol{A}, \boldsymbol{\eta}; \boldsymbol{X})] = \frac{MN}{2} \ln(2\pi) + \frac{1}{2} \sum_{l=0}^{M-1} \ln\left\{\det[\boldsymbol{V}(\boldsymbol{\zeta}_l; \sigma_\alpha, \sigma_\beta, \sigma_\tau)]\right\}$$
$$+ \frac{1}{2} \sum_{l=0}^{M-1} \left\{(\boldsymbol{x}_l - \boldsymbol{\zeta}_l)^\mathsf{T} [\boldsymbol{V}(\boldsymbol{\zeta}_l; \sigma_\alpha, \sigma_\beta, \sigma_\tau)]^{-1}(\boldsymbol{x}_l - \boldsymbol{\zeta}_l)\right\}. \qquad (36)$$

Ignoring the constant first term, multiplying the remaining terms by 2, and expressing the matrix elements explicitly, we can define the ML cost function,

$$Q_{\mathrm{ML}}(\sigma_\alpha, \sigma_\beta, \sigma_\tau, \boldsymbol{\mu}, \boldsymbol{A}, \boldsymbol{\eta}; \boldsymbol{X}) = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} \left\{\ln\left[\sigma_\alpha{}^2 + \sigma_\beta{}^2 Z_{kl}^2 + \sigma_\tau{}^2(\boldsymbol{DZ})_{kl}^2\right] \right.$$
$$\left. + \frac{(X_{kl} - Z_{kl})^2}{\sigma_\alpha{}^2 + \sigma_\beta{}^2 Z_{kl}^2 + \sigma_\tau{}^2(\boldsymbol{DZ})_{kl}^2}\right\}, \qquad (37)$$

which we minimize to obtain ML estimates for all of the free parameters in the model.

The resulting estimates for the noise parameters (σ_α, σ_β, and σ_τ) are biased below their true values, which we can attribute to the presence of the nuisance parameters ($\boldsymbol{\mu}$, $\boldsymbol{A}$, and $\boldsymbol{\eta}$) [21]. For example, in 1000 Monte Carlo simulations of $M = 10$ repeated measurements using the waveforms described in Sec. 2, the ratios of the ML estimates to their true values are $\hat{\sigma}_\alpha/\sigma_\alpha = 0.95 \pm 0.02$, $\hat{\sigma}_\beta/\sigma_\beta = 0.94 \pm 0.03$, and $\hat{\sigma}_\tau/\sigma_\tau = 0.89 \pm 0.09$, all significantly below unity. Increasing the number of observations to $M = 50$ reduces the bias to $\hat{\sigma}_\alpha/\sigma_\alpha = 0.990 \pm 0.007$, $\hat{\sigma}_\beta/\sigma_\beta = 0.98 \pm 0.01$, and $\hat{\sigma}_\tau/\sigma_\tau = 0.93 \pm 0.04$, but some bias remains in $\hat{\sigma}_\beta$ and $\hat{\sigma}_\tau$ even in the limit $M \to \infty$. This residual bias arises because the number of elements in $\boldsymbol{A}$ and $\boldsymbol{\eta}$ both grow with the number of observations, which allows us to account for drift but dilutes some of the information that the data provide about the noise [21,24]. In principle, we can resolve this problem by integrating out all of the nuisance parameters in Eq. (35) to obtain a marginal likelihood that depends only on σ_α, σ_β, and σ_τ [21], but this involves computationally expensive integrations for a relatively small benefit. As we discuss below, we can remove much of the bias by simply rescaling the noise parameters by a suitable correction factor.

To determine the bias correction factor it is helpful to consider a simplified example in which there is no drift and only additive noise, so that $A_l = 1$ and $\eta_l = 0$ for all l and $\sigma_\beta = \sigma_\tau = 0$. The ML cost function then reduces to

$$\tilde{Q}_{\mathrm{ML}}(\sigma_\alpha, \boldsymbol{\mu}; \boldsymbol{X}) = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} \left[\ln {\sigma_\alpha}^2 + \frac{(X_{kl} - \mu_k)^2}{{\sigma_\alpha}^2} \right], \tag{38}$$

which is minimized by

$$\hat{\mu}_k = \frac{1}{N} \sum_{l=0}^{M-1} X_{kl} = \bar{x}_k, \qquad \hat{\sigma}_\alpha^2 = \frac{1}{N} \sum_{k=0}^{N-1} s_M^2(t_k) = \overline{s_M^2}, \tag{39}$$

where

$$s_M^2(t_k) = \frac{1}{M} \sum_{l=0}^{M-1} (X_{kl} - \bar{x}_k)^2. \tag{40}$$

This result for $\hat{\sigma}_\alpha^2$ is biased because it is the waveform average of $s_M^2(t_k)$, which in turn is just the biased sample variance of the measurements at t_k. To remove this bias we can apply Bessel's correction, which replaces the factor of $1/M$ with $1/(M-1)$ in Eq. (40). Alternatively, we can multiply $\hat{\sigma}_\alpha^2$ by $M/(M-1)$ in Eq. (39). Returning to Eq. (37) and following similar reasoning, we can justify rescaling all of the ML noise parameter estimates by the same factor,

$$\hat{\sigma}_\alpha^* = \hat{\sigma}_\alpha \sqrt{\frac{M}{M-1}}, \qquad \hat{\sigma}_\beta^* = \hat{\sigma}_\beta \sqrt{\frac{M}{M-1}}, \qquad \hat{\sigma}_\tau^* = \hat{\sigma}_\tau \sqrt{\frac{M}{M-1}}. \tag{41}$$

With these corrections, the Monte Carlo simulations with $M = 10$ yield $\hat{\sigma}_\alpha^*/\sigma_\alpha = 1.00 \pm 0.02$, $\hat{\sigma}_\beta^*/\sigma_\beta = 0.99 \pm 0.03$, and $\hat{\sigma}_\tau^*/\sigma_\tau = 0.94 \pm 0.09$, which are all within the statistical uncertainty of the true values. The simulations with $M = 50$ yield lower statistical uncertainty, but with the same average values: $\hat{\sigma}_\alpha^*/\sigma_\alpha = 1.000 \pm 0.007$, $\hat{\sigma}_\beta^*/\sigma_\beta = 0.99 \pm 0.01$, and $\hat{\sigma}_\tau^*/\sigma_\tau = 0.94 \pm 0.04$. For all practical purposes, the remaining bias in $\hat{\sigma}_\beta^*$ and $\hat{\sigma}_\tau^*$ is negligible.

6. Experimental verification

In this section we present experimental results that verify our analysis. Figure 5(a) shows the raw data that we use to specify the noise model, $\boldsymbol{X}$, which comprises $M = 50$ waveforms, each with $N = 256$ points sampled every $T = 50$ fs. With this data as input, we minimize Eq. (37) to obtain the ML estimates, $\hat{\sigma}_\alpha$, $\hat{\sigma}_\beta$, $\hat{\sigma}_\tau$, $\hat{\boldsymbol{\mu}}$, $\hat{\boldsymbol{A}}$, and $\hat{\boldsymbol{\eta}}$ for all of the free parameters in the model. The resulting time-dependent noise amplitude, corrected for bias using Eq. (41), is

$$\hat{\sigma}_\mu^*(t_k) = \sqrt{V_{kk}(\hat{\boldsymbol{\mu}}; \hat{\sigma}_\alpha^*, \hat{\sigma}_\beta^*, \hat{\sigma}_\tau^*)}. \tag{42}$$

In Fig. 5(b) we compare this model to the observed time-dependent noise, which we estimate by using $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{\eta}}$ to correct for the drift,

$$\tilde{\boldsymbol{x}}_l = \frac{1}{\hat{A}_l} \boldsymbol{S}(-\hat{\eta}_l) \boldsymbol{x}_l, \tag{43}$$

then compute the unbiased standard deviation at each time point over the set $\{\tilde{\boldsymbol{x}}_l\}$. The model quantitatively describes most features of the measurements, with small deviations near some of the peaks. As a further consistency check, Fig. 5(c) shows the normalized residuals for one of the waveforms,

$$\boldsymbol{r}_{\mathrm{ML},l} = \left\{ \boldsymbol{V} \left[\boldsymbol{\zeta}(\hat{\boldsymbol{\mu}}, \hat{A}_l, \hat{\eta}_l); \hat{\sigma}_\alpha^*, \hat{\sigma}_\beta^*, \hat{\sigma}_\tau^* \right] \right\}^{-1/2} \left[\boldsymbol{x}_l - \boldsymbol{\zeta}(\hat{\boldsymbol{\mu}}, \hat{A}_l, \hat{\eta}_l) \right], \tag{44}$$

which demonstrates that they approximately follow a standard normal distribution, with small but statistically significant correlations among neighboring points.
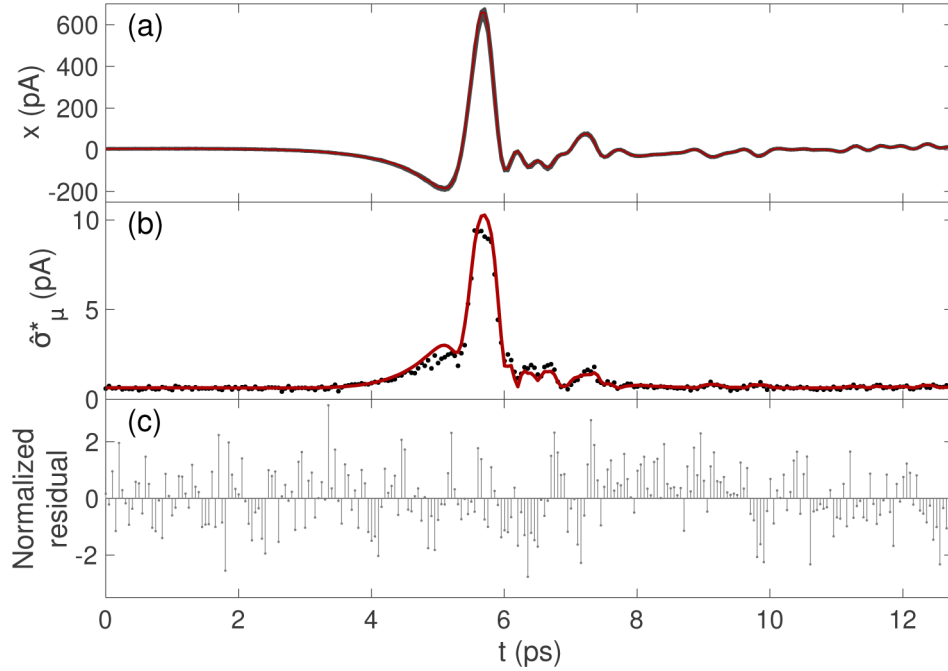
Fig. 5. Noise-model fits with laboratory measurements. (a) Set of 50 nominally identical waveform measurements that compose the data matrix X. (b) Time-dependent noise amplitude obtained from a fit with Eq. (37) to X. The solid line shows the ideal noise model, with the bias-corrected ML estimates $\hat{\sigma}_\alpha^* = (0.623 \pm 0.005)$ pA, $\hat{\sigma}_\beta^* = (1.55 \pm 0.03)\%$, and $\hat{\sigma}_\tau^* = (1.8 \pm 0.1)$ fs. Points show the standard deviation of the measurements after correcting for drift, as described in the text. (c) Normalized residuals for one of the waveform measurements shown in (a).

Figure 6 shows fits to two sequential waveforms from this set. A fit with Eq. (5) in the time domain, obtained by minimizing Q_{TLS} in Eq. (28), yields $\hat{\theta}_1^{\text{TLS}} = 1.019 \pm 0.003$ for the amplitude and $\hat{\theta}_2^{\text{TLS}} = (-2.8 \pm 0.5)$ fs for the delay, with the normalized residuals shown in Fig. 6(b). A fit with the same model in the frequency domain, obtained by minimizing Q_{ETFE} in Eq. (6), yields $\hat{\theta}_1^{\text{ETFE}} = 1.020 \pm 0.004$ and $\hat{\theta}_2^{\text{ETFE}} = (-4.4 \pm 0.6)$ fs, with the normalized residuals shown in Fig. 6(d). Both fit methods reveal significant differences from the ideal values of $\theta_1 = 1$ and $\theta_2 = 0$, reflecting the measurement drift discussed in Sec. 5. As we found for the residuals of the noise-model fit in Fig. 5(c), the residuals of the transfer-function fit in Fig. 6(b) show small but statistically significant correlations. But as we also found with the idealized Monte Carlo simulations, the residuals of the ETFE fit in the frequency domain are much more structured than the residuals of the TLS fit to the same data in the time domain.

To illustrate the analysis problem that this raises, in Fig. 6(c) we compare an ETFE fit with Eq. (5) to an ETFE fit with a more flexible transfer-function model,

$$H_2(\omega; \boldsymbol{\theta}) = (\theta_1 + \theta_3\omega^2)\exp[i\omega(\theta_2 + \theta_4\omega^2)]. \tag{45}$$

Since the two input waveforms are nominally identical, we know that the downturns in $\text{Re}[\hat{H}_{\text{ETFE}}(\omega)]$ and $\text{Im}[\hat{H}_{\text{ETFE}}(\omega)]$ with increaing frequency near 2 THz are spurious. But if we did not know this in advance, we might conclude from Fig. 6(c) that Eq. (45) describes the measurements better than Eq. (5). We would also be able to support this conclusion by comparing the GOF statistic for the fit with Eq. (5), $S_{\text{ETFE}}^{(1)} \approx 223$, with $\nu = 254$ degrees of freedom, to that
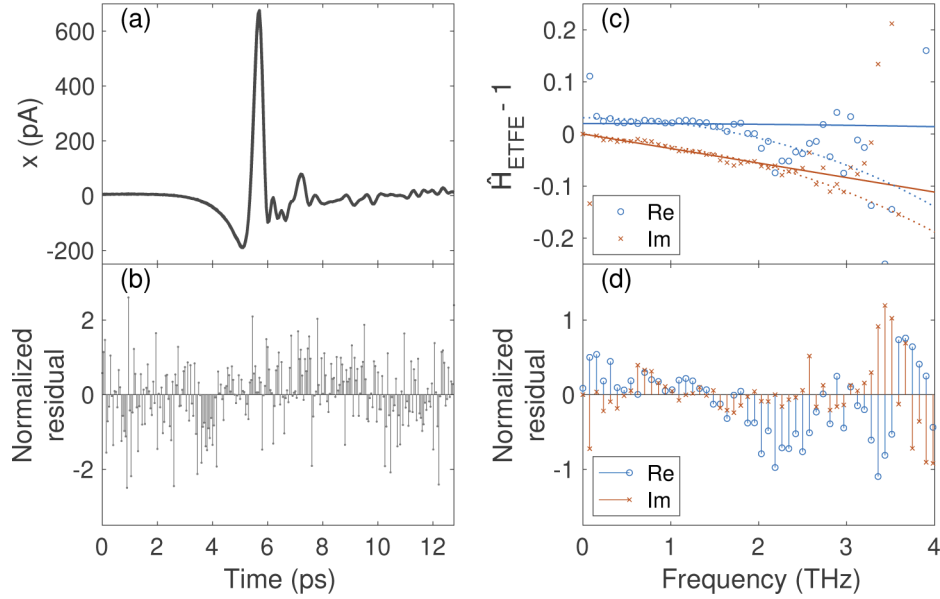
Fig. 6. Transfer-function fits with laboratory measurements. (a) Two sequential waveform measurements, taken from the set shown in Fig. 5. (b) Normalized time-domain residuals, $r_{\text{TLS}}(\hat{\boldsymbol{\theta}}^{\text{TLS}})$, for the TLS fit with Eq. (5). (c) Real and imaginary parts of $\hat{H}_{\text{ETFE}}(\omega_l) - 1$ (dots), fitted with Eq. (5) (solid lines) and Eq. (45) (dotted lines) by minimizing Q_{ETFE} in Eq. (6). (d) Normalized frequency-domain residuals, $r_{\text{ETFE}}(\omega_l; \hat{\boldsymbol{\theta}}^{\text{ETFE}})$, for the fit with Eq. (5).

obtained with Eq. (45), $S_{\text{ETFE}}^{(2)} \approx 191$, with $\nu = 252$ degrees of freedom. By adding only two additional fit parameters, we reduce S_{ETFE} by 33, which erroneously suggests that the added complexity of Eq. (45) captures a real physical effect. The TLS method is more robust against such overfitting: the GOF statistic with Eq. (5) is $S_{\text{TLS}}^{(1)} \approx 198$, while with Eq. (45), $S_{\text{TLS}}^{(2)} \approx 196$. In this case, adding two free parameters reduces S_{TLS} by only two, so Occam's razor favors the simpler model, Eq. (5). More formally, to select from a set of transfer-function models $H_k(\omega; \boldsymbol{\theta}_k)$, $k = 1, 2, \ldots, N_{\text{model}}$, each with p_k free parameters, we can minimize a modified cost function based on the Akaike information criterion [21],

$$Q_{\text{AIC}}(k) = S_{\text{TLS}}^{(k)} + 2p_k, \tag{46}$$

where $S_{\text{TLS}}^{(k)}$ is the TLS GOF statistic for the model $H_k(\omega; \boldsymbol{\theta}_k)$. As we discussed in Sec. 4, this ability to discriminate among competing models is one of the major advantages of the TLS method.

If we divide the 50 experimental waveforms into 25 sequential pairs and fit each pair with Eq. (5), the results are qualitatively consistent with the Monte Carlo simulations and support the conclusion that S_{TLS} offers the better absolute measure of fit quality. The distribution of S_{ETFE} over the experiments has a mean $\bar{S}_{\text{ETFE}} \approx 235$ and standard deviation $\sigma(S_{\text{ETFE}}) \approx 113$, while the distribution for S_{TLS} has a mean $\bar{S}_{\text{TLS}} \approx 246$ and standard deviation $\sigma(S_{\text{TLS}}) \approx 39$. For the simulated distributions shown in Fig. 3(a) and Fig. 4(a), we obtain $\bar{S}_{\text{ETFE}} \approx 235$, $\sigma(S_{\text{ETFE}}) \approx 160$, $\bar{S}_{\text{TLS}} \approx 250$, and $\sigma(S_{\text{TLS}}) \approx 22$. As we discussed at the end of Sec. 4, a narrower GOF distribution provides better sensitivity to fit quality. And while the experimental distribution for S_{TLS} is nearly twice as broad it is in the simulations, it is still nearly a factor of 3 narrower than the

experimental distribution for S_{ETFE}. Despite the quantitative differences between the simulations and the experiment, the TLS method consistently shows better performance than the ETFE.

7. Conclusion

In summary, we have developed a maximum-likelihood approach to THz-TDS analysis and demonstrated that it has numerous advantages over existing methods. Starting from a few simple assumptions, we derived a method to determine the transfer function relationship between two THz-TDS measurements, together with a method to specify a parameterized noise model from a set of repeated measurements. In each case, we also derived expressions for fit residuals in the time domain that are properly normalized by the expected noise. Through a combination of Monte Carlo simulations and experimental measurements, we verified that these tools yield results that are accurate, precise, responsive to fit quality, and generally superior to the results of fits to the ETFE.

We focused on simple examples to emphasize the logical structure of the method, but we can readily apply the same approach to more complex problems. For example, we have successfully used the framework presented here to measure a weak frequency dependence in the Drude scattering rate of a metal, which is predicted by Fermi liquid theory; we have also used it to measure small variations in the THz-frequency refractive index of silicon with temperature [17]. In both of these applications we found that maximum-likelihood analysis in the time domain provided better performance than a least-squares analysis based on the ETFE in the frequency domain. We expect similar improvements in other applications, and provide MATLAB functions in Code Repository 1 (Ref. [25]) for others to try.

Disclosures. The authors declare no conflicts of interest.

References

1. L. Duvillaret, F. Garet, and J.-L. Coutaz, "Highly precise determination of optical constants and sample thickness in terahertz time-domain spectroscopy," Appl. Opt. **38**(2), 409 (1999).
2. T. D. Dorney, R. G. Baraniuk, and D. M. Mittleman, "Material parameter estimation with terahertz time-domain spectroscopy," J. Opt. Soc. Am. A **18**(7), 1562–1571 (2001).
3. M. Naftaly ed., *Terahertz Metrology* (Artech House, Boston, 2015).
4. W. Withayachumnankul, B. M. Fischer, H. Lin, and D. Abbott, "Uncertainty in terahertz time-domain spectroscopy measurement," J. Opt. Soc. Am. B **25**(6), 1059 (2008).
5. M. Naftaly and R. Dudley, "Methodologies for determining the dynamic ranges and signal-to-noise ratios of terahertz time-domain spectrometers," Opt. Lett. **34**(8), 1213–1215 (2009).
6. M. Naftaly, "Metrology issues and solutions in THz time-domain spectroscopy: Noise, errors, calibration," IEEE Sens. J. **13**(1), 8–17 (2013).
7. M. Naftaly, R. G. Clarke, D. A. Humphreys, and N. M. Ridler, "Metrology state-of-the-art and challenges in broadband phase-sensitive terahertz measurements," Proc. IEEE **105**(6), 1151–1165 (2017).
8. M. Skorobogatiy, J. Sadasivan, and H. Guerboukha, "Statistical Models for Averaging of the Pump–Probe Traces: Example of Denoising in Terahertz Time-Domain Spectroscopy," IEEE Trans. Terahertz Sci. Technol. **8**(3), 287–298 (2018).
9. H. W. Hubers, M. F. Kimmitt, N. Hiromoto, and E. Brundermann, "Terahertz spectroscopy: System and sensitivity considerations," IEEE Trans. Terahertz Sci. Technol. **1**(1), 321–331 (2011).
10. M. Krüger, S. Funkner, E. Bründermann, and M. Havenith, "Uncertainty and ambiguity in terahertz parameter extraction and data analysis," J. Infrared, Millimeter, Terahertz Waves **32**(5), 699–715 (2011).
11. D. Jahn, S. Lippert, M. Bisi, L. Oberto, J. C. Balzer, and M. Koch, "On the influence of delay line uncertainty in THz time-domain spectroscopy," J. Infrared, Millimeter, Terahertz Waves **37**(6), 605–613 (2016).
12. A. Rehn, D. Jahn, J. C. Balzer, and M. Koch, "Periodic sampling errors in terahertz time-domain measurements," Opt. Express **25**(6), 6712–6724 (2017).
13. W. Withayachumnankul and M. Naftaly, "Fundamentals of measurement in terahertz time-domain spectroscopy," J. Infrared, Millimeter, Terahertz Waves **35**(8), 610–637 (2014).
14. J. L. M. van Mechelen, A. B. Kuzmenko, and H. Merbold, "Stratified dispersive model for material characterization using terahertz time-domain spectroscopy," Opt. Lett. **39**(13), 3853 (2014).

15. S. Krimi, J. Klier, J. Jonuscheit, G. von Freymann, R. Urbansky, and R. Beigang, "Highly accurate thickness measurement of multi-layered automotive paints using terahertz technology," Appl. Phys. Lett. **109**(2), 021105 (2016).

16. R. Peretti, S. Mitryukovskiy, K. Froberger, M. A. Mebarki, S. Eliet, M. Vanwolleghem, and J. Lampin, "THz-TDS time-trace analysis for the extraction of material and metamaterial parameters," IEEE Trans. Terahertz Sci. Technol. **9**(2), 136–149 (2019).

17. L. Mohtashemi, "Test of Fermi liquid theory with terahertz conductivity measurements of MnSi," Ph.D. thesis, Simon Fraser University (2020).

18. D. Grischkowsky and N. Katzenellenbogen, "Femtosecond pulses of terahertz radiation: physics and applications," in *OSA Proceedings on Picosecond Electronics and Optoelectronics*, vol. 9 T. C. L. G. Sollner and J. Shah, eds. (Optical Society of America, Washington, DC, 1991), p. 9–14.

19. L. Ljung, *System Identification: Theory for the User*, Prentice-Hall Information and System Sciences Series (Prentice Hall, Upper Saddle River, N.J., 1999), 2nd ed.

20. P. Guillaume, I. Kollar, and R. Pintelon, "Statistical analysis of nonparametric transfer function estimates," IEEE Trans. Instrum. Meas. **45**(2), 594–600 (1996).

21. Y. Pawitan, *In All Likelihood: Statistical Modelling and Inference Using Likelihood* (Oxford University Press, Oxford, 2001).

22. S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem*, Frontiers in Applied Mathematics (Society for Industrial and Applied Mathematics, Philadelphia, 1991).

23. W. Withayachumnankul, B. M. Fischer, and D. Abbott, "Material thickness optimization for transmission-mode terahertz time-domain spectroscopy," Opt. Express **16**(10), 7382–7396 (2008).

24. J. Neyman and E. L. Scott, "Consistent Estimates Based on Partially Consistent Observations," Econometrica **16**(1), 1–32 (1948).

25. J. S. Dodge, L. Mohtashemi, P. Westlund, D. G. Sahota, G. B. Lea, I. Bushfeld, and P. Mousavi, "MATLAB functions for maximum-likelihood parameter estimation in terahertz time-domain spectroscopy," Zenodo (2020). https://zenodo.org/record/4326594.