

Joint entropy of continuously differentiable ultrasonic waveforms

M. S. Hughes^{a)}

Department of Medicine/Cardiology Division, Campus Box 8215, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110-1093

J. E. McCarthy

Department of Mathematics, Washington University, Campus Box 1146, One Bookings Drive, St. Louis, Missouri 63130-4899

J. N. Marsh and S. A. Wickline

Department of Medicine/Cardiology Division, Campus Box 8215, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110-1093

(Received 8 June 2012; revised 17 October 2012; accepted 19 November 2012)

This study is based on an extension of the concept of joint entropy of two random variables to continuous functions, such as backscattered ultrasound. For two continuous random variables, X and Y , the joint probability density $p(x, y)$ is ordinarily a continuous function of x and y that takes on values in a two dimensional region of the real plane. However, in the case where $X = f(t)$ and $Y = g(t)$ are both continuously differentiable functions, X and Y are concentrated exclusively on a curve, $\gamma(t) = (f(t), g(t))$, in the x, y plane. This concentration can only be represented using a mathematically “singular” object such as a (Schwartz) distribution. Its use for imaging requires a coarse-graining operation, which is described in this study. Subsequently, removal of the coarse-graining parameter is accomplished using the ergodic theorem. The resulting expression for joint entropy is applied to several data sets, showing the utility of the concept for both materials characterization and detection of targeted liquid nanoparticle ultrasonic contrast agents. In all cases, the sensitivity of these techniques matches or exceeds, sometimes by a factor of two, that demonstrated in previous studies that employed signal energy or alternate entropic quantities.

© 2013 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4770245>]

PACS number(s): 43.60.Bf [SAF]

Pages: 283–300

I. INTRODUCTION

Previous studies have demonstrated the utility of several different entropies for detection of subresolution backscattering structures in both materials characterization and medical ultrasonics in situations where other approaches fail. These include detection of unresolvable near-surface defects,^{1–5} diffuse accumulation of subresolution-sized nanoparticle-based ultrasound contrast agents,^{6–10} characterization of smooth muscle pathologies,^{11–13} and monitoring of ultrasonically induced temperature rise in tissue.¹⁴

In the first of these studies, entropy images were applied to detect defects in advanced aerospace materials.¹ In particular, detection and imaging of “near-surface” defects in graphite-epoxy composite plates was accomplished using a 5 MHz center frequency transducer.^{1–5} In this application, it is not possible to separate the front wall echo from the defect reflection. This presents a significant challenge to conventional approaches for defect detection and imaging. In fact, the case presented in the following text is nearly impossible to discern using the traditional approach. However, the joint entropy image permits

visualization of the defect, its boundary, and certain substructures contained in the defect that were not visible to conventional detection methods or even the previously published entropic images.

The bulk of our investigations have focused on medical imaging. These studies have demonstrated the utility of several different entropies for detection of subresolution structures of backscattering tissue: e.g., targeted accumulation of weakly scattering perfluorocarbon nanoparticles in tissues *in vivo*, and pathological changes in smooth muscle structure arising from muscular dystrophy. These results have been obtained in a variety of animal disease models and in patients as shown in Table I. Several features common to these studies deserve special emphasis at the outset:

- Studies have been executed using several different imaging systems (Philips HDI 5000, Philips IE33, Vevo 660, Ardent Spark) with transducer arrays covering a frequency range from 2–34 MHz.
- In all cases but one,¹¹ the entropic analysis demonstrated greater sensitivity than did signal energy to subresolution features of scattering architecture. In the exceptional case,¹¹ sensitivities were essentially equivalent.
- The most recent version of entropy published^{8,15} is amenable to real-time implementation.

^{a)}Author to whom correspondence should be addressed. Electronic mail: mshatctrain@gmail.com

TABLE I. Summary of studies based on different entropic signal receivers defined for waveform $y = f(t)$ having density function $w_f(y)$: entropy, H_f , Renyi entropy, $I_f(r)$, and its asymptotic limiting form, $I_{f,\infty}$. The joint entropy $H_{f,g}$ is frequently much more sensitive than any of these.

“Entropic” receiver	Heuristic meaning	Scattering structure	Experimental system
$H_f = - \int_{f_{min}}^{f_{max}} w_f(y) \log[w_f(y)] dy$	“Information” contained	Materials characterization,	Graphite/epoxy laminates (Ref. 5)
$H_f = - \int_0^1 \log[f'_1(t)] dt$	“Information” contained in backscattered RF $f(t)$.	Targeted nanoparticles, smooth muscle cells	MDA 435 (Ref. 16), B16 VX2, MDx mice (Ref. 11) Precancer models (Ref. 7) Humans (Ref. 12)
$I_f(r) = 1/1 - r \log \int_{f_{min}}^{f_{max}} w_f(y)^r dy$	“Weighted-information” contained in backscattered RF: $f(t)$ (note: $H_f = I_f(1)$).	Targeted nanoparticles, smooth muscle cells	MDA 435, B16 Precancer models (Ref. 17)
$I_{f,\infty} = \log\{[t_k f'(t_k) = 0] / f''(t_k) \}$	Limiting form of $I_f(r)$ as $r \rightarrow 2$. More sensitive than $H_f, I_f(r)$.	Targeted nanoparticles, smooth muscle cells	MDA 435 (Ref. 8), B16 (Ref. 10) Humans Precancer models (Ref. 15)

- (d) All results pertaining to detection of targeted nanoparticles were obtained without use of hand-drawn regions of interest. In fact, the entire analysis chain from raw data to image construction to image-based sensitivity assessment was automated and thus objective.
- (e) Image subtraction has not been required to detect changes in scattering architecture.

All of these points suggest the clinical utility of this approach, which might proceed along lines similar to that employed in Doppler imaging systems, where the conventional B-mode image is color-coded according to the blood cell velocity to present a combined B-mode/velocity image; similarly, a B-mode/entropy image could be made as well.⁹

All of these investigations required acquisition of thousands of radio frequency (RF) signals, $f(t)$. For each of these captured waveforms, viewed as a sampled version of a differentiable function, the probability density function of values was computed using the digitized values of $f(t)$. Distributions of values (or densities of distributions of values) are typically used for analysis of random variables, and all differentiable functions are, in fact, legitimate random variables. However, in the general case, it is not possible to compute the probability density function of a random variable solely from its values. For example, to determine the probability density function for noise in an amplifier, the noise source must be sampled and the distribution of sampled values subsequently fit to an assumed functional form, e.g., Gaussian or Poisson. In our case, the knowledge that the random variable $f(t)$ is differentiable permits direct computation of its probability density function without the need to assume and fit a model.^{2,5,7} While this can be a mathematically and computationally demanding task, previous applications of this approach to analysis of acoustic signals have repeatedly demonstrated its utility. The present study extends these results by developing joint entropy, which we will denote by $H_{f,g}$, to analyze the same type of data. This approach requires computation of the joint probability density of two (differentiable) functions: $y = f(t)$, representing a segment of backscattered radio frequency ultrasound, and a reference signal, $g(t)$. After carefully deriving a suitable expression for $H_{f,g}$, we show that it provides further improvements in sensitivity over those obtained in previous studies that were based on other entropies.

II. APPROACH

Our goal is to further improve the ability to detect weak scatterers in strongly scattering environments, as in the case where liquid nanoparticles are targeted to cancer molecular epitopes or in material defects in advanced aerospace materials. Although prior entropy measures have been proposed, we expect the current approach to demonstrate significant enhancement in detection in real-time applications that could be ported to existing imaging systems.

A. Background

This study is based on an extension of the concept of joint entropy of two random variables to continuous functions, such as backscattered RF, and a reference signal, such as a reflection from a weak reflector. For two continuous random variables, X, Y with joint probability density function $p(x, y)$, the joint entropy is

$$H(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) \log p(x, y) dx dy. \quad (1)$$

Ordinarily $p(x, y)$ is a continuous function of x and y that takes non-zero values in a two dimensional region of the real plane. However, in the case where $X = f(t)$ and $Y = g(t)$ are both continuously differentiable functions, the random variables X and Y are concentrated exclusively on a curve, $\gamma(t) = (f(t), g(t))$, in the x, y plane as shown in Fig. 1. Thus the densities of their values cannot be represented using conventional functions as was done in previous studies of only one differentiable random variable.^{1-5,7} Instead, this concentration can only be represented using a mathematically “singular” object like a series of Dirac Delta functions placed on $\gamma(t)$ with infinitesimal spaces in between them. Technically, such an object is not a function; in precise terms, $p(x, y)$ must be represented as a Schwartz distribution.¹⁶ Its use for imaging requires that the Schwartz distribution be coarse-grained, i.e., integrated, on a fine grid of small cells as shown in Fig. 2. The outcome of this course-graining, as shown in Appendix B, is the conversion of the probability density (which is a Schwartz distribution) into a *bona fide* probability distribution *function* such that the probability of X and Y being in a particular cell, which is the generalization of

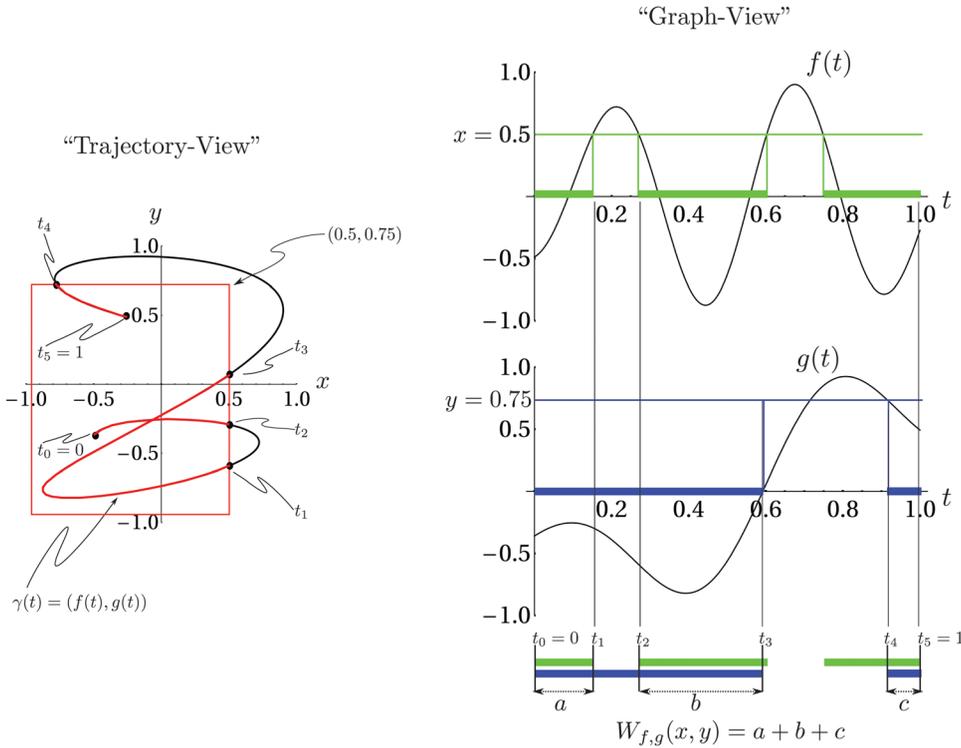


FIG. 1. (Color online) Left panel: A plot of the curve $\gamma(t)$ used to visualize the joint cumulative density function, $W(x, y)$, of values of the functions $f(t)$ and $g(t)$. The probability density is non-zero only on the curve $\gamma(t)$ because the random variables $X = f(t)$, $Y = g(t)$ are concentrated exclusively on $\gamma(t)$. On the right is a graphical calculation of $W(x, y)$ using the graphs of $f(t)$ and $g(t)$. The exact definition of $W(x, y)$ is given in Eq. (B1).

$p(x, y)$ that we seek, is given by the time that $\gamma(t)$ “spends” in the cell [Eq. (B28) in the following text]. This is what we would expect intuitively if we imagined a test particle traversing the curve, $\gamma(t)$, with velocity $v(t) = \sqrt{f'(t)^2 + g'(t)^2}$, and wanted to know the probability of finding it in a particular cell. Having this result, it is possible to extract a well-defined finite component from the coarse-grained version of Eq. (1), which we will denote by $H_{f,g}$, in terms of $f'(t)$ and $g'(t)$.

As with the entropies previously investigated, $H_{f,g}$ is applied in a moving window analysis. However, its computation requires two input waveforms. The first, $f(t)$, is exactly the same as in previous investigations. However, choice of

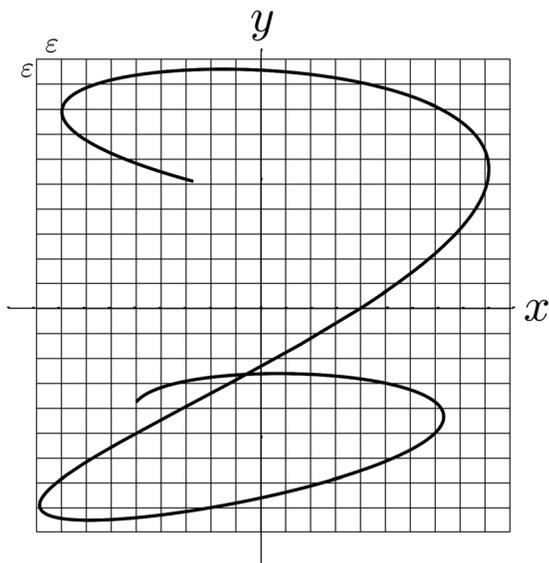


FIG. 2. Uniform partition of the xy -plane into $\epsilon \times \epsilon$ cells.

the second reference waveform, $g(t)$, constitutes an additional degree of freedom that may be exploited to increase sensitivity of the entropic analysis. Discovery of strategies for identification of the optimum reference waveform is one of the eventual goals of our research. In this study, we present the results of preliminary investigations that indicate a reflection of the imaging system’s interrogating pulse from a weak reflector, such as a water-agarose interface, constitutes a good initial guess for the optimum. To summarize our findings:

- $H_{f,g}$ analysis may be applied in the same situations as previous entropies, $H_f, I(r), I_{f,\infty}$. In fact, all of the data for this study have been analyzed previously with the use of at least one of these entropies. In this study, we repeat several of these analyses using $H_{f,g}$ images.
- $H_{f,g}$ analysis is amenable to real-time implementation, as was $I_{f,\infty}$ analysis.
- In all cases studied so far, $H_{f,g}$ analysis has proven to be more sensitive than the best previous results obtained by other entropies. For example, preliminary application of this new entropy to the MDA 435 data shown in Figs. 7 and 8 shows a roughly twofold increase in peak confidence ratios.

B. Conventions and calculation

To fix terminology, suppose that we are given two functions, $f(t)$ and $g(t)$, assumed to map $[0, 1]$ into $(-1, 1)$ (i.e., $|f_{\max}|, |f_{\min}|, |g_{\max}|, |g_{\min}|$ are all strictly less than one). We are interested in computing the joint entropy,

$$H(f, g) = \int_{-1}^1 dx \int_{-1}^1 dy w(x, y) \log[w(x, y)], \quad (2)$$

where $w(x, y)$ is their joint density, defined on the domain $\Omega = (-1, 1) \times (-1, 1)$. It turns out that for the types of

functions $f(t)$ and $g(t)$ that we will consider (intuitively thought of as “well-behaved”), $w(x, y)$ is not a real-valued function but a (Schwartz) distribution so that Eq. (2) is not well-defined.

To get around this difficulty, we employ a “coarse-grained” approximation to Eq. (2) obtained, by dividing Ω into $\varepsilon \times \varepsilon$ cells, where $\varepsilon = 1/N$ as shown in Fig. 2, and integrating $w(x, y)$ over each cell. We denote the $\varepsilon \times \varepsilon$ cells intersecting $\gamma(t)$ at initial time t_i by C_i and index them by $i = 1, \dots, M$ (M will be discussed later). For subsequent discussion, we define the distance traversed along $\gamma(t)$ when it enters the i th cell by s_i . Figure 9 illustrates our terminology. The final step in the coarse-graining process is replacement of $H(f, g)$ by the quantity $H_{f,g}$ defined by

$$H_{f,g} = \lim_{\varepsilon \rightarrow 0} [H^\Delta(f, g) - \varepsilon] \\ = \lim_{\varepsilon \rightarrow 0} \left[\sum_{j,k} p^\Delta(j, k) \log[p^\Delta(j, k)] - \varepsilon \right], \quad (3)$$

where $p^\Delta(j, k) = 0$ unless $\gamma(t)$ passes through the (j, k) th cell, in which case (as mentioned in the preceding text and as we shall show in the following text), $p^\Delta(j, k)$ is equal to the time $\gamma(t)$ spends in the (j, k) th cell.

Our ultimate goal is to obtain an expression for $H_{f,g}$ in terms of $f(t)$ and $g(t)$ because these quantities are experimentally accessible. As is carefully shown in Appendixes A–C, the following relation is true:

$$H_{f,g} \equiv -\frac{1}{2} \int_0^1 dt \frac{\min[|f'(t)|, |g'(t)|]}{\max[|f'(t)|, |g'(t)|]} \\ - \int_0^1 dt \log[\max[|f'(t)|, |g'(t)|]]. \quad (4)$$

It is a remarkable fact that derivation of Eq. (4), which uses the ergodic theorem, also requires the fact that the set of points for which $f'(t)/g'(t)$ is rational comprises a set of measure zero (in the set of real numbers). Thus the relation is essentially derived in a noncomputable setting. Nevertheless, the end result, Eq. (4), may be evaluated numerically. The full details are provided in the appendixes.

III. RESULTS

A. Noise simulation

The effect of noise on the input waveform $f(t)$ is shown in Fig. 3. To produce this plot, Gaussian distributed noise of different amplitude was added to $f(t) = e^{-10(t-0.4)^2} \sin(4\pi t)$ and $g(t) = e^{-10(t-0.4)^2} \sin(2\pi t)$ (simulations using uniformly distributed noise have also been performed yielding essentially similar results). The range of noise levels spans 20 to 120 dB. At each noise level, 1000 waveforms $f(t)$ and reference waveforms $g(t)$ were created, $H_{f,g}$ computed, and the average $\langle H_{f,g} \rangle$ and standard error computed. Each $\langle H_{f,g} \rangle$ is shown in the plot, with its standard error bar, which in all but one case is too small to see. We notice that for signal-to-noise ratios above 80 dB, it is possible to obtain three-digit agreement with the noise-free answer [which was obtained

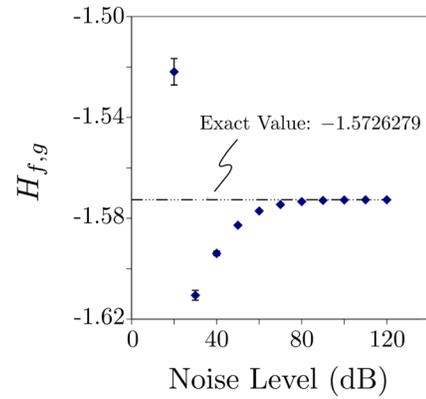


FIG. 3. (Color online) Effect noise on the computation of $H_{f,g}$ using Eq. (4). Each data point is the average of $H_{f,g}$ computed from $N = 1000$ different waveforms $g(t) = e^{-10(t-0.4)^2} \sin(2\pi t)$ to which Gaussian noise has been added. The signal-to-noise ratio simulated ranges from 20 to 120 dB. The reference trace $f(t) = e^{-10(t-0.4)^2} \sin(4\pi t)$ did not have noise added. Standard error bars are plotted with each point, most are too small to see on the scale of the plot.

using a separate MATHEMATICA implementation of Eq. (4) that was also used to double check the correctness of our C programming language implementation of the same equation]. However, for signal-to-noise ratios between 40 and 60 dB, which is the relevant range for our experimental data, agreement to at most one or two digits is achievable. This limitation shows that the numerical precision of our composite integrator is more than enough for analysis of experimental data, which always contain noise.

B. Materials characterization: Defect detection

The first application of entropy imaging was to materials characterization;¹⁷ in particular, to detection and sizing of near-surface defects in graphite/epoxy aerospace structures.^{1–5} Figure 4 shows some previously published (top row)⁵ and newer results (middle and bottom row) for imaging of a near-surface “resin-rich” defect in a thin (4 mm thick) graphite/epoxy composite. The bottom left corner shows a schematic of the specimen, which was fabricated by removing a 3.81 cm diameter circular section from the second ply prior to fabrication. The specimen was then autoclaved following standard manufacturing practice. This forced excess epoxy into the circular void and created the resin-rich region. During the fabrication procedure, additional resin penetrated the inter-laminar region bordering the circular void. The result is an oval shaped defect centered on the original circular cutout.

This specimen was scanned on a 101×101 point grid using a 5 MHz broadband transducer with the backscattered RF acquired at 8-bit digital resolution and a 100 MHz sampling rate. The width of the interrogating pulse was therefore far too wide to permit gating of front-wall reflection from that of the defect, which lay within one ply (0.25 mm) of the surface. While the top row results⁵ were originally published as part of a larger study of six different types of near-surface defects, results in the resin-rich case were less conclusive than hoped for because, as the top row shows, the best understood images failed to detect the boundaries of the defect

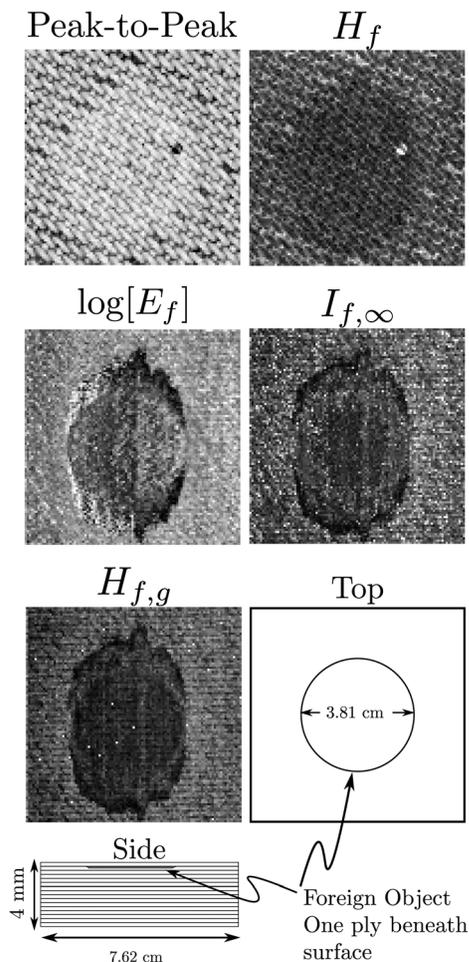


FIG. 4. Five different types of images made using RF backscattered from a graphite/epoxy plate with a near-surface defect. Top row contains previously published results (Ref. 5). These were obtained using the full 512-point digitized waveform and without optimal smoothing splines. The next two rows show the results obtained using optimal smoothing splines on a 128-point subset of the full digitized waveform. A schematic of the experimental specimen is shown in the lower left corner of the figure. The $H_{f,g}$ image shows the clearest delineation of defect boundaries.

region. While other thermodynamic receivers used in that study⁵ did delineate the defect boundaries, their physical meaning was not readily apparent.

The middle and bottom rows of Fig. 4 show that this situation has been rectified by the newer signal processing algorithms described in the preceding text. One reason for the improvement is the use of optimal smoothing splines, which are used for global noise suppression.¹⁸ Additionally, only a gated portion of the backscattered RF (128 points long for this figure) was used. Gating enables exclusion of baseline regions that contain no useful data. While $\log[E_f]$ and $I_{f,\infty}$ images show the defect with relatively good contrast, the boundary between defect and surrounding plate “fades-out” at the “3 o’clock” position. This can be significant, for instance, if the images are to be used as input for automatic feature extraction software. On the other hand, at all points on its perimeter, the $H_{f,g}$ image has greater contrast between defect boundary and surrounding “defect-free” plate (matching the clarity of the less understood images shown in earlier work (e.g., those labeled Z , C_V , E_{th} , in Fig. 6 of that

work⁵). We also observe a circular outline within the defect region. This is the diameter of the circular region removed from the second ply to produce the resin-rich region during the curing process used to fabricate the graphite/epoxy laminate. During that process, resin accumulated not only in the circular region but was also pushed out around its boundaries to create the oblong appearance of the defect. We note that the outline of the original circle is less evident in $\log[E_f]$ and $I_{f,\infty}$ images and is not visible in any of the images from the previous study.⁵ The reference $g(t)$ was a reflection from a stainless steel reflector placed at the 4-in. focal plane of the transducer. Because the experimental system used for data acquisition has been verified to be stable to at least the -40 decibel level, this reference was acquired once, prior to scanning of the entire collection of graphite/epoxy specimens. We observe that this is equivalent to the use of a reference acquired from the weak reflector that we use in biological studies. The operational criterion for choice of reflector is that it roughly matches the reflection coefficient of the scatterers in the specimen trace $f(t)$.

C. Medical imaging: Targeted tumor imaging

Previously, we reported on application of entropic quantifies H_f and $I_{f,\infty}$ for detection of subtle changes in RF backscatter induced by accumulation of targeted nanoscale contrast enhancers in tumor neovasculture^{6,8} or precancerous tissue.⁷ We have used the same raw RF from these studies to prepare $H_{f,g}$ images, which were then analyzed for evidence of nanoparticle accumulation. As the materials and methods for these studies are described in previous publications,⁸ we provide only a brief summary of data acquisition and analysis in each case and then present a comparison of the $H_{f,g}$ and $I_{f,\infty}$ analyses.

D. MDA 435 tumors implanted in athymic nude mice

Human MDA 435 cancer cells were implanted in the inguinal fat pads of 15 athymic nude mice between 19 and 20 days prior to acquisition of data. Five of these animals were injected with $\alpha_v\beta_3$ -targeted nanoparticles, five were injected with nontargeted nanoparticles, and five were injected with saline at a whole body dose of 1 ml/kg. In addition, 15 athymic nude mice not implanted with tumors were imaged in the same region following the same imaging protocol: five were injected with $\alpha_v\beta_3$ -targeted nanoparticles, five were injected with nontargeted nanoparticles, and five were injected with saline.

RF data were acquired with a research ultrasound system (Vevo 660, Visualsonics, Toronto, Canada) at 0 through 60 min in 5-min intervals after injection. The tumor was imaged with a 35-MHz center frequency single element “wobbler” probe, and the digitized RF data corresponding to single frames were stored on a hard disk for later off-line analysis. The frames consisted of 384 lines of 2048 12-bit words acquired at a sampling rate of 200 MHz using a Gage 12400 digitizer card (connected to the analog-out and sync ports of the Vevo) in a controller PC. Each frame corresponds spatially to a region 1.5 cm wide and 0.8 cm deep. Further description of materials and methods may be found

in previous publications.⁸ The reference trace $g(t)$ was obtained by digitizing a reflection of the imaging system's interrogating pulse from a water-agarose interface using the same acquisition parameters. As for the materials study, the long-term stability of the system electronics has been verified better than the 40 dB level. Consequently, one reference trace was acquired prior to beginning the study and used for all subsequent $H_{f,g}$ analysis.

A moving window analysis was performed on each waveform by moving a rectangular window (128 points long, $0.64 \mu\text{s}$) in $0.08 \mu\text{s}$ steps (16 points), resulting in 121 window positions within the output data set. A smoothing spline was fit to each window. Issues surrounding signal-to-noise (SNR) ratio, SNR estimation, and robust noise suppression for these data have been discussed in a previous publication.⁸ The fitting routine also returned an array of first and second derivatives at the locations of any critical points in the window. The arrays were used to compute $H_{f,g}$ (or $I_{f,\infty}$). This produced an image for each time point in the experiment (i.e., 0, 5, ..., 60 min).

The segment of RF within the boxcar, or moving window, is processed to produce a pixel value for an $H_{f,g}$ (or $I_{f,\infty}$) image. For each mouse used in this study, this is done

using RF data acquired at 0, 5, ..., 60 min post-injection to produce an image at each time point. For this study, in which the same portion of the anatomy was imaged at successive intervals, our goal was to identify and quantify the accumulation of targeted nanoparticles, which occurs preferentially at targeting sites. This suggests that segmenting the image into "targeted" and "non-targeted" regions will be required as part of the analysis.

One of the chief goals of our research has been to develop objective segmentation algorithms that do not require user selection (e.g., hand drawn regions of interest). Figure 5, displays the steps of such an algorithm graphically. For each mouse used in this study, a histogram of the image pixel values appearing over the entire time course (i.e., 0, 5, 10, ..., 60 min) was constructed and normalized to obtain the probability density function (PDF) of these values and then integrated to obtain the cumulative distribution function (CDF). This is shown in the top panel of the figure. Next, pixel values corresponding to "analysis-thresholds" at the lower 2%, 4%, ..., 98% of the CDF were then used to segment the images at each time point into two regions corresponding to targeted and nontargeted tissue. The figure shows the segmentation for an example analysis-threshold of

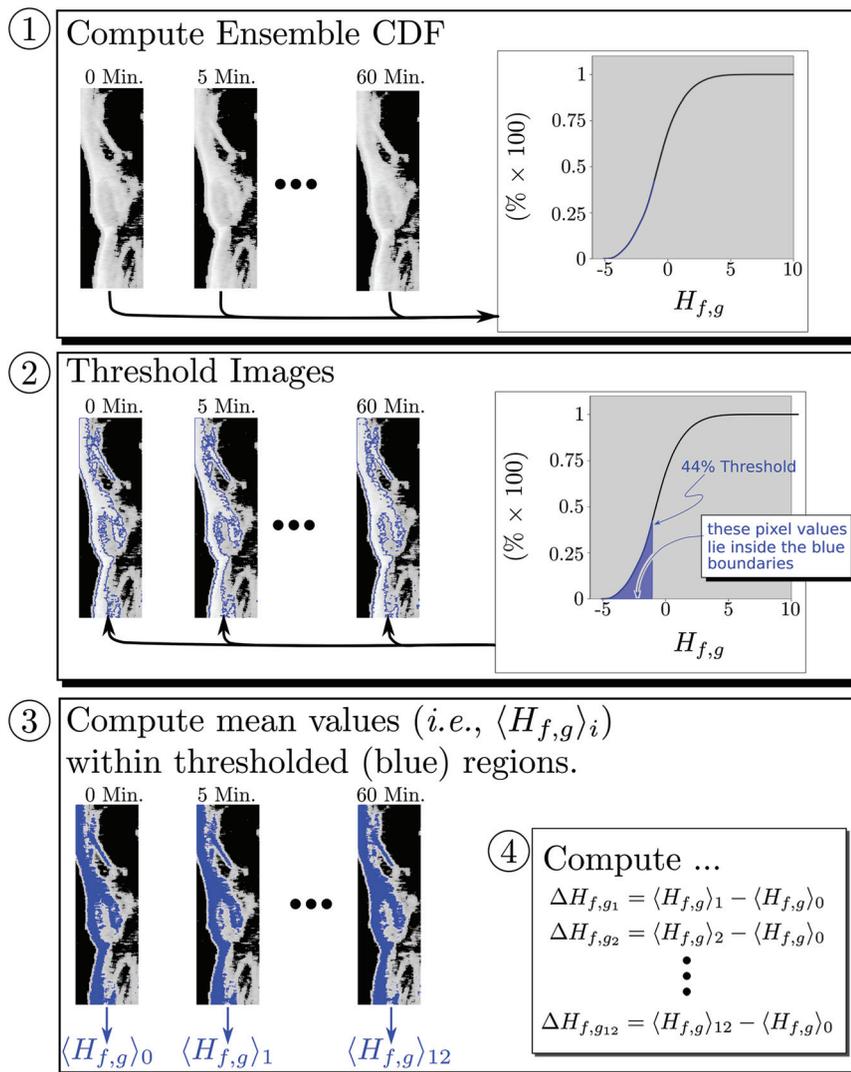


FIG. 5. Analysis steps for $H_{f,g}$ images from a single mouse. Step 1: The cumulative distribution function (CDF) of all 12 images acquired from the animal are computed. Step 2: The CDF is used to segment the image into "targeted" (inside the blue boundaries) and "non-targeted" (outside) segments according to pixel value being above or below the threshold level (44% in this case). Step 3: the mean value, $\langle H_{f,g} \rangle_i$, of $H_{f,g}$ is computed for each targeted region. Step 4: These are used to calculate $\Delta H_{f,g,i} = \langle H_{f,g} \rangle_i - \langle H_{f,g} \rangle_0$ for each time point in the study. This procedure is repeated for all mice in each group.

44%. The blue lines, shown in the second panel of the figure, indicate the boundary between targeted (inside the blue boundaries) and non-targeted (outside the blue boundaries) regions. Subsequently, the mean value of pixels in the targeted region were computed as a function of time post-injection. This is indicated in the bottom panel of the figure, in which the mean value of $H_{f,g}$ found at each time, denoted in the figure by $\langle H_{f,g} \rangle_i$, $i = 1, \dots, 12$, is shown immediately below the thresholded image. Subsequently, the change in these mean values,

$$\Delta H_{f,g} = \langle H_{f,g} \rangle_i - \langle H_{f,g} \rangle_0, \quad (5)$$

is computed ($\Delta H_{f,g_0} = 0$). The analysis diagrammed in Fig. 5 was performed for all animals in all groups and the results averaged by group. We will drop the subscript i in the remaining discussion and refer only to the $\Delta H_{f,g}$.

$\Delta H_{f,g}$, corresponding to a 98% analysis-threshold is plotted in Fig. 6, in panel number one. For the purposes of detection, it is actually the ratio of the mean value to the standard error that is significant. We have plotted an example of these ratios (for the MDA 435 implanted group injected with $\alpha_v\beta_3$ -targeted nanoparticles) in panel two of Fig. 6. To shorten subsequent discussion of our results, we will define this ratio (confidence) as¹⁹

$$c = \frac{\text{mean}}{\text{standard error}}. \quad (6)$$

As this panel shows, for an analysis-threshold of 98%, the mean value of $\Delta H_{f,g}$ is at all times at least four standard deviations or more from zero and peaks at 16 standard deviations from the mean $H_{f,g}$ at zero time. We have mapped each confidence value to a color scale shown along the x axis of the plot. Panel three of the figure shows the aggregate of all color maps resulting from analysis thresholds between 0% and 98%. These aggregated color maps, or confidence panels, may be used to quickly identify analysis thresholds at which accumulation of targeted nanoparticles is successfully detected by either $H_{f,g}$ or $I_{f,\infty}$ imaging and to quantify the sensitivity of each image type.

An inter-receiver comparison is shown in Fig. 7, which is a five-dimensional presentation having three dimensions within each confidence panel (i.e., post-injection time in the vertical direction, analysis-threshold in the horizontal direction, and confidence c in the out-of-plane or color direction), another dimension for animal group (the vertical direction within the array), and a fifth dimension being the masking level for the absolute value of confidence ratio c (the horizontal direction within the array). In the left column are $\Delta H_{f,g}$ confidence panels for all six groups used in our study as indicated in the figure caption. In the right-hand column are the corresponding $\Delta I_{f,\infty}$ confidence panels made using the same RF data and smoothing spline processing parameters.

In spite of the high dimensionality of the display, rapid comparison of $H_{f,g}$ or $I_{f,\infty}$ processing is possible, as well as

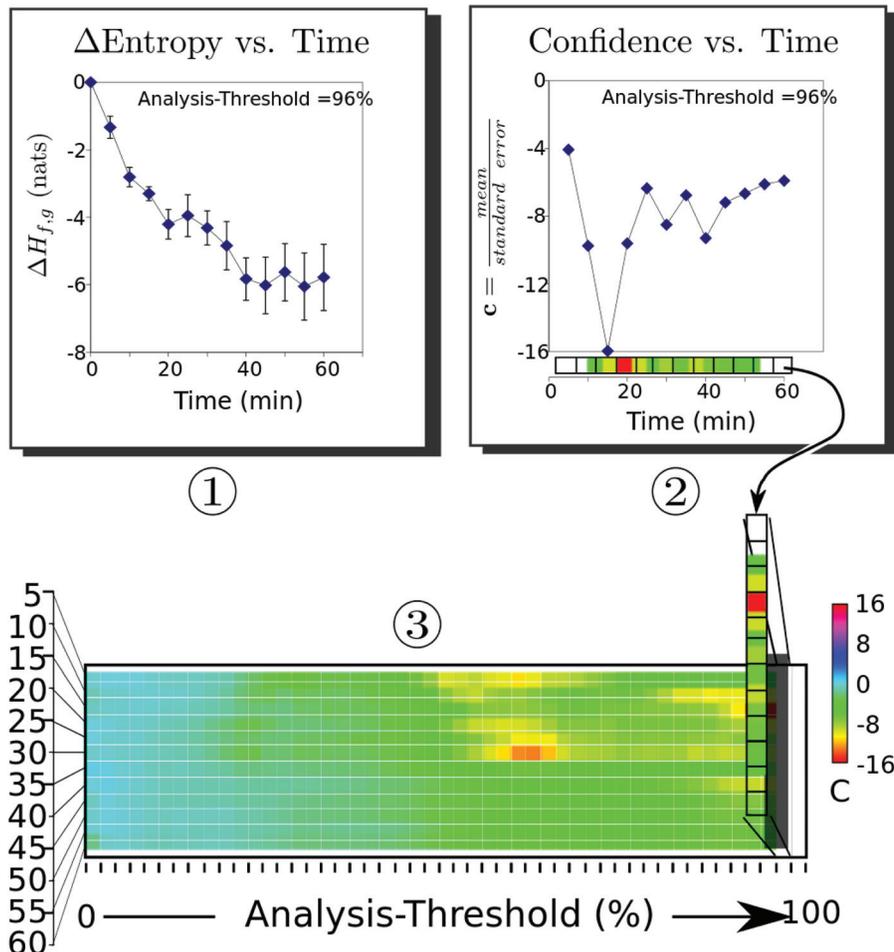


FIG. 6. Use of $\Delta H_{f,g}$ vs time plots to prepare confidence, c , panels. Step 1: An analysis threshold is set, and $H_{f,g}$ images are analyzed to produce plots of $\Delta H_{f,g}$; the time plot for the 96% analysis threshold is shown. Step 2: The confidence as a function of time is computed according to Eq. (6) and then converted to a color bar shown parallel to the x axis of the plot. Step 3: These color bars are stacked vertically left-to-right, in increasing order, to prepare a confidence panel showing the confidence obtained at all time points for all analysis thresholds between 0% and 100%. A color-to-confidence calibration bar is shown to the right of the confidence panel.

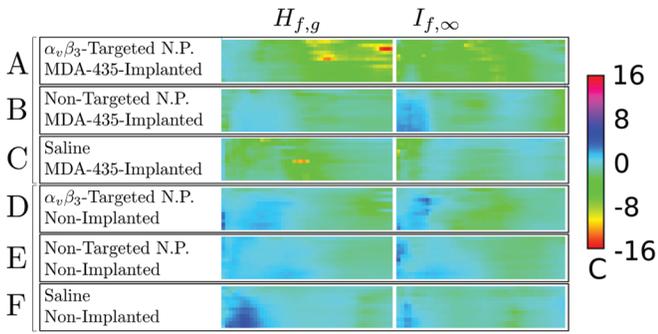


FIG. 7. Confidence, c , panels for $\Delta H_{f,g}$ (left column) and $\Delta I_{f,\infty}$ (right column) for all groups used in our study. (A) MDA 435-implanted mice injected with $\alpha_v\beta_3$ -targeted nanoparticles ($N = 5$), (B) MDA 435-implanted mice injected with non-targeted nanoparticles ($N = 5$), (C) MDA 435-implanted mice injected with saline ($N = 5$), (D)–(F), same injections into $N = 5$ tumor-free mice. The smoothing spline parameter $S = 1.0$ in all cases. A color-to-confidence calibration bar is shown to the right of the confidence panels.

inter-group comparison. For instance, in both columns of the top row, we observe “extensive” (meaning contiguous, wider than three columns over at least half the height of the confidence panel) regions, corresponding to confidence values greater than four in magnitude. The remaining confidence panels (for the control groups), with one exception, have much lower confidence value magnitudes in the same range of analysis-thresholds. Comparison of the confidence panels in the left column ($\Delta H_{f,g}$) shows that the top confidence panel, corresponding to the MDA 435-implanted group injected with $\alpha_v\beta_3$ -targeted nanoparticles, generally exhibits confidence ratios with the greatest magnitude. There is an anomalous region of high confidence ratio for the saline group C appearing at 30 min post-injection. Because no nanoparticles were injected into this group, it might be classified as a “false-positive” for the technique. Further research, aimed at understanding and eliminating this effect, e.g., by using different reference traces $g(t)$, is underway. However, we remark that this region is contained in a range of analysis thresholds that does not overlap the range of analysis thresholds for which the confidence ratio of the targeted MDA435-implanted group (row A) is large. Consequently, it possible to define a strategy for separating the groups that eliminates false positives. Moreover, for the top row, the $\Delta H_{f,g}$ images produce confidence values that are more than twice those obtained using $\Delta I_{f,\infty}$ images. Consequently, the criteria: “ $|c| \geq 4$,” with analysis threshold between 60% and 98% for $H_{f,g}$ analysis (44%–62% for $I_{f,\infty}$ analysis), becomes a means of distinguishing the targeted/tumor-implanted group from all others. We observe also that these criteria exhibit the presence of targeted nanoparticles accumulating within five minutes after injection. While the figure shows that we may distinguish the targeted tumor-implanted group from the control groups, the issue of single animal “diagnosis” remains open due to the small cohort of animals studied. The current study is a demonstration of feasibility and suggests the utility of conducting a larger experiment having a proper receiver operator curve analysis that would permit quantification of true positives, true negatives, false positives, and false negatives.

E. Transgenic K14-HPV16 mouse tumor model

The model used is the transgenic K14-HPV16 mouse that contains human papilloma virus-16 oncoproteins driven by a keratin promoter so that lesions develop in the skin, particularly in the ear. Typically the ears exhibit squamous metaplasia, a pre-cancerous condition, associated with abundant neovasculture that expresses the $\alpha_v\beta_3$ integrin. Eight transgenic mice^{20,21} were treated with 1.0 mg/kg i.v. of either $\alpha_v\beta_3$ -targeted nanoparticles ($n = 4$) or nontargeted nanoparticles ($n = 4$) and imaged dynamically for 1 h using a research ultrasound imager (Vevo 660 40 MHz probe) modified to store digitized RF waveforms acquired at 0, 15, 30, and 60 min. time points. Further details may be found in reference 8. All RF data were processed in the same manner as the MDA 435 data.

The results are shown in Fig. 8, which displays the three confidence panels obtained using $H_{f,g}$ images on the left and $I_{f,\infty}$ images on the right. Broad, contiguous, regions of large confidence ratio magnitude only occur in the right side of the panels, corresponding to analysis thresholds of 78% to 98%. Consequently, we have emphasized these portions by enclosing them with dotted lines while placing a semi-transparent layer over the complementary region (which has been displayed in this manner for completeness). Focusing attention on the 78% to 98% confidence thresholds, we see that $H_{f,g}$ analysis exhibits twice the sensitivity of $I_{f,\infty}$ analysis for the targeted group. Moreover, $H_{f,g}$ analysis, at these analysis thresholds, exhibits greater separation between the targeted group and the saline control group than does $I_{f,\infty}$ analysis. While separation between the targeted group and the non-targeted control is greater for $I_{f,\infty}$ analysis, the panels show that $H_{f,g}$ analysis still yields several standard deviations of separation between the two groups.

IV. DISCUSSION AND CONCLUSIONS

In all cases, the sensitivity afforded by analysis of $H_{f,g}$ images exceeds that obtained in our best previous analyses of the same data by roughly a factor of 2. The main difference between the previous approach and the current one is employment of a reference trace. The results presented in

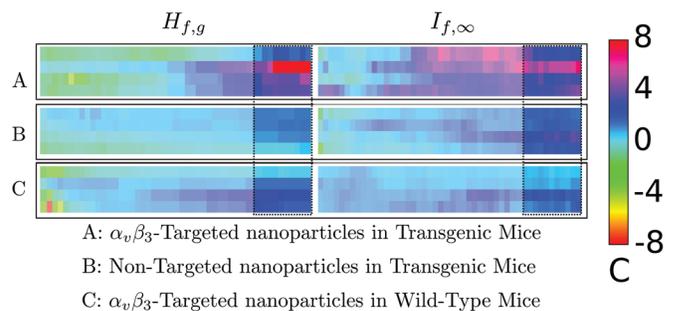


FIG. 8. Comparison $\Delta H_{f,g}$ (left) and $\Delta I_{f,\infty}$ (right) for data acquired in a pre-cancer model (Ref. 7) like that shown in Fig. 7. While analysis at all thresholds between 0% and 100% was undertaken, we wish to focus attention on the results obtained between 75% and 100% (results from lower analysis thresholds have been “dimmed” by a translucent overlay). In both cases, a broad region of confidence values above 2 is shown. While the extent of this region is larger in the $\Delta I_{f,\infty}$ confidence panel, the maximum height of the $\Delta H_{f,g}$ confidence panel is more than twice as high (16.0 vs 6.8). A color-to-confidence calibration bar is shown to the right of the confidence panels.

this study employed a reference, water-path only reflection from a weak reflector, which is the natural choice in many conventional ultrasonic analyses that deconvolve the transfer function of the experimental apparatus from the raw experimental data. In our case, the rationale for application of this reference is based on the relation between entropy, $H(A)$, joint entropy, $H(A, B)$, and conditional entropy of B given A , $H(B|A)$ for discrete random variables A , and B , which is

$$H(B, A) = H(A, B) = H(A) + H(B|A). \quad (7)$$

If A is taken as a reference variable and we compute $H(B', A)$ where B' is another random variable, then $H(B, A) - H(B', A) = H(B|A) - H(B'|A)$. Because analysis of entropy images described in the preceding text is based on differences between joint entropies, the last equation shows that they can be reinterpreted as differences between conditional entropies, i.e., heuristically at least, we are looking at the differences in information contained in backscattered RF (e.g., either B , or B') given that the tissue was interrogated with a certain reference pulse (A).

The reference trace $g(t)$ is also a “knob” that we can “turn” to adjust the sensitivity of the image to specific attributes in the reference waveform. In our heuristic interpretation, we are examining the extent of differences between the information in the reference trace and the backscatter. In other words, we are estimating how much the information content of the interrogating ultrasonic pulse has been changed by the scattering events occurring in the tissue.

In spite of the fact that our choice of reference appears natural, it is conceivable that in other circumstances, a different choice of reference waveform might be more compelling. Thus the choice of reference is to some extent arbitrary. Our response to this fact has been to post-process the resulting $H_{f,g}$ images using all analysis thresholds and to determine if there exists a rational set of criterion for differentiating the targeted from control groups. The results of the present study, as well as previous studies,^{8,15} advance one strategy for doing so. The investigation of other strategies for choice of an optimum reference given specific detection requirements is the subject of ongoing research.

ACKNOWLEDGMENTS

This study was funded by NIH EB002168, HL042950, and CO-27031 and NSF DMS 0966845. The research was carried out at the Washington University Department of Mathematics and the School of Medicine.

APPENDIX A: CALCULATION OVERVIEW

In this subsection, we derive from Eq. (2) an expression for $H_{f,g}$ in terms of the experimentally accessible quantities $f(t)$ and $g(t)$. To do this, we claim the following assertions, which will be established in subsequent sections:

- (a) **Claim i:** $p^\Delta(j, k) = \Delta t_i = t_{i+1} - t_i$ if $\gamma(t)$ enters the (j, k) th cell at time t_i and leaves it at t_{i+1} . Otherwise, $p^\Delta(j, k) = 0$. This is Eq. (B28) in the following text. Consequently, evaluation of Eq. (3) comes down to

being able to compute the limiting behavior of $\varepsilon \Delta t_i \log[\Delta t_i]$ as $\varepsilon \rightarrow 0$. We will see presently that what we really need to compute is the limiting behavior of $\varepsilon \Delta s_i \log[\Delta s_i]$.

This leads us to our second assertion:

- (b) **Claim ii:** there exists an expression for the limiting behavior of $\varepsilon \Delta s_i \log[\Delta s_i]$ in terms of $f'(t)$ and $g'(t)$, assuming that $g'(t)/f'(t)$ is irrational. This is essentially Eq. (A9) in the following text. Subsequently, we derive the expression for $H_{f,g}$ in terms of the experimentally accessible quantities $f'(t)$ and $g'(t)$. The next two subsections will justify claims i and ii. At this point, we observe from Fig. 9 that the distance traversed, Δs_i , in any member of C_i is given by

$$\Delta s_i = \Delta t_i v_i, \quad (A1)$$

where $v_i = v(t_i)$, $i = \dots, M$. Using Eq. (A1), Eq. (3) becomes the sum

$$\begin{aligned} \sum_{i=1}^M \Delta t_i \log[\Delta t_i] &= \sum_{i=1}^M \Delta t_i \log \left[\frac{\Delta s_i}{v_i} \right] \\ &= - \sum_{i=1}^M \Delta t_i \log[v_i] + \sum_{i=1}^M \Delta t_i \log[\Delta s_i]. \end{aligned} \quad (A2)$$

The first term in the sum has the limiting behavior

$$- \sum_{i=1}^M \Delta t_i \log[v_i] \rightarrow - \int_0^1 dt \log[v(t)], \quad (A3)$$

as $\varepsilon \rightarrow 0$. It remains to evaluate the sum

$$\sum_{i=1}^M \Delta t_i \log[\Delta s_i]. \quad (A4)$$

However, we do not know the coordinates of the entry (or exit) points of $\gamma(t)$ into the i th cell, and thus it appears that we lack the necessary information to

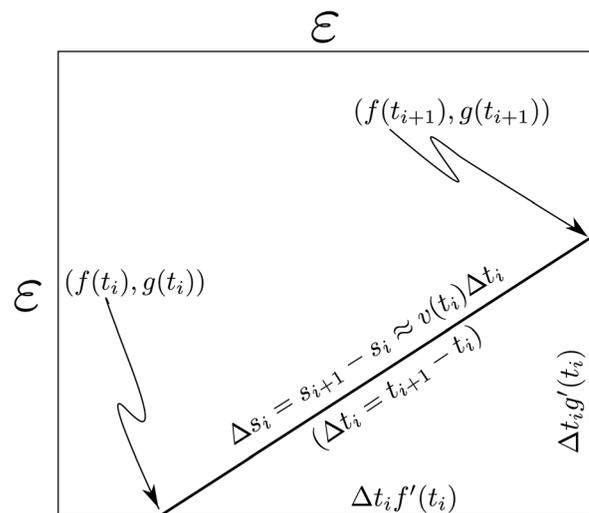


FIG. 9. A diagram of the computation of time spent by $\gamma(t)$ in an $\varepsilon \times \varepsilon$ cell in the xy plane.

calculate either Δt_i or Δs_i . Nevertheless, if we assume that the portion of the curve $\gamma(t)$ within each $\varepsilon \times \varepsilon$ cell is a straight line segment that starts at the point $(f(t_i), g(t_i))$ and has slope

$$\alpha_i \equiv g'(t_i) / f'(t_i) \quad (\text{A5})$$

so that it leaves the $\varepsilon \times \varepsilon$ cell near the point $(f(t_{i+1}), g(t_{i+1}))$, we may then continue the limiting process by sub-partitioning each cell further into $\varepsilon_s \times \varepsilon_s$ cells as shown in Fig. 10. Given this subdivision, each term in the sum, Eq. (A4), must now be replaced by the sum

$$\sum_{n=1}^{M_{S_i}} \Delta t_{i,n} \log[\Delta s_{i,n}], \quad (\text{A6})$$

where the n indexes the $\varepsilon_s \times \varepsilon_s$ cells crossed by $\gamma(t)$ as it traverses the larger $\varepsilon \times \varepsilon$ cell. Thus the sum appearing in Eq. (A4) is replaced by a double sum

$$\sum_{i=1}^M \sum_{n=1}^{M_{S_i}} \Delta t_{i,n} \log[\Delta s_{i,n}]. \quad (\text{A7})$$

We further assume that the velocity, v_i , is constant over the entire $\varepsilon \times \varepsilon$ square as we further subdivide it. Thus the inner sum in Eq. (A6) becomes

$$\begin{aligned} \sum_{n=1}^{M_{S_i}} \Delta t_{i,n} \log[\Delta s_{i,n}] &= \sum_{n=1}^{M_{S_i}} \frac{1}{v_i} \Delta s_{i,n} \log[\Delta s_{i,n}] \\ &= \frac{1}{v_i} \sum_{n=1}^{M_{S_i}} \Delta s_{i,n} \log[\Delta s_{i,n}] \\ &\equiv S_i. \end{aligned} \quad (\text{A8})$$

While it remains true that we still do not know where $\gamma(t)$ will enter each of these smaller cells, we have nevertheless made progress because if we imagine stacking all of the $\varepsilon_s \times \varepsilon_s$ cells crossed by $\gamma(t)$ on top

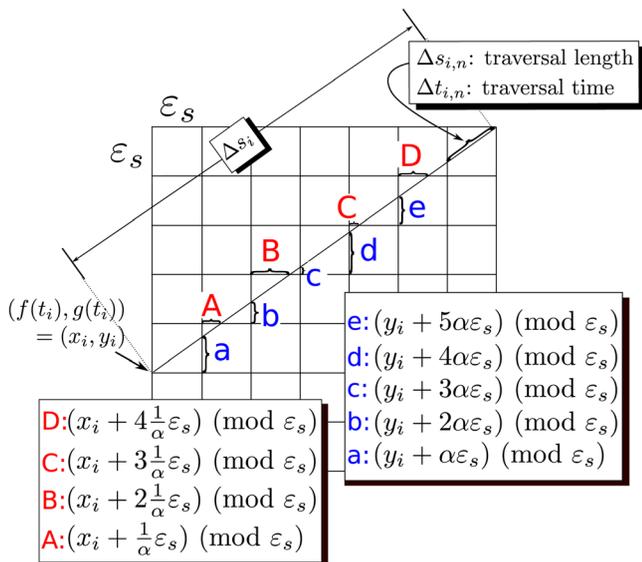


FIG. 10. (Color online) Subdivision of an $\varepsilon \times \varepsilon$ cell into finer $\varepsilon_s \times \varepsilon_s$ cells required for evaluation of Eq. (A6). We assume that ε is small enough that in the larger $\varepsilon \times \varepsilon$ cell $\gamma(t)$ is a straight line and $v(t)$ is constant. Legends indicate lengths of lettered segments.

of each other, as shown in Fig. 11, we will observe that, in almost every case (we will make this statement precise presently), the entry/exit points will be uniformly distributed around the perimeter of the smaller $\varepsilon_s \times \varepsilon_s$ cell. Given this picture, it is relatively easy to see that Δs_i may be obtained by summing the transit lengths of $\gamma(t)$ over the smaller “stacked” $\varepsilon_s \times \varepsilon_s$ cell and that in the limit where these crossings become infinite (i.e., $\varepsilon_s \rightarrow 0$), the sum may be obtained as the integral of the transit lengths, starting along the boundary of the stacked cell.

The almost every case in the preceding text occurs when the ratio $g'(t)/f'(t)$ is irrational. The reader might anticipate this fact by considering first the case where the ratio is rational (and the lengths of the sides of the $\varepsilon_s \times \varepsilon_s$ cell is rational, which is the case with our conventions), then it is easy to see that a small set of entry/exit points will be used over and over again, i.e., the trajectory cell crossings will exhibit periodic behavior. It is also relatively easy to see that the number of these entry/exit points grows as the decimal expansion of the $g'(t)/f'(t)$ grows in length. Consequently, we might anticipate that for irrational $g'(t)/f'(t)$, the number of these points is infinite. The ergodic theorem tells us that this is, in fact, true and that they are also uniformly distributed around the perimeter of the $\varepsilon_s \times \varepsilon_s$ cell. Thus the entry point into the original (larger) $\varepsilon \times \varepsilon$ cell becomes irrelevant because as the $\varepsilon_s \times \varepsilon_s$ cells are made smaller, all entry values occur with equal probability. Actually, while application of the ergodic theorem eliminates that we know where $\gamma(t)$ enters a cell, it also introduces a cell dependent scaling factor into our expression for Δs_i . Consequently, we have traded one missing piece of information for another.

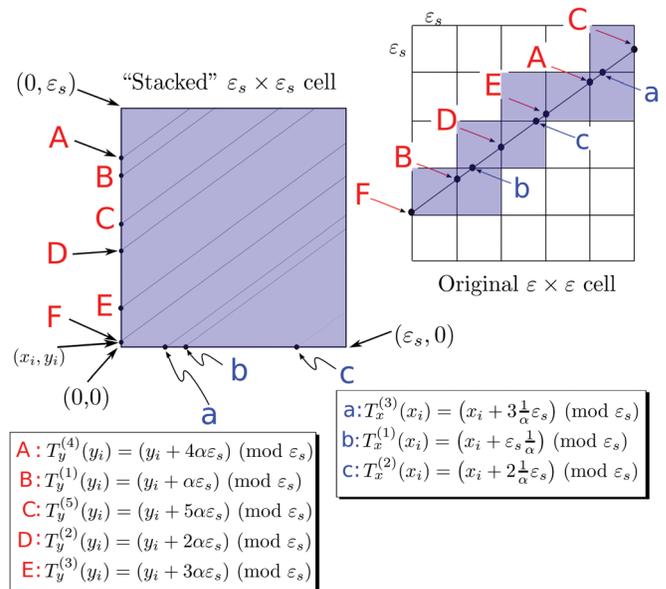


FIG. 11. (Color online) Mapping of subgrids onto the “stacked” $\varepsilon_s \times \varepsilon_s$ square. The shaded $\varepsilon_s \times \varepsilon_s$ sub-squares of the original $\varepsilon \times \varepsilon$ cell are all stacked on top of each other. The corresponding points where $\gamma(t)$ crosses boundaries of the sub-squares are lettered by capital or small letters. Legends indicate their status as iterates of the mapping $T_y(y_i)$ or $T_x(x_i)$ [Eqs. (C4) and (C6)]. For irrational $\alpha_i = g'(t_i)/f'(t_i)$, the ergodic theorem guarantees that as the number of sub-squares is increased, the density of labeled points on vertical and horizontal sides becomes uniform.

However, these scaling factors group together in the course of the calculation can be eliminated from the final expression for \mathcal{S}_i [see Eq. (C18)]. Moreover, if $f(t)$, $g(t)$ are continuously differentiable over their range, then the set of points at which $f'(t)/g'(t)$ is rational forms a set of measure zero, which may be ignored in any integral, so that it is only the cases where $g'(t)/f'(t)$ is irrational that matter. We provide the precise details in a subsequent section, the end result is:

$$\begin{aligned}
\mathcal{S}_i &= \frac{1}{v_i} \sum_{n=1}^{M_s} \Delta s_{i,n} \log[\Delta s_{i,n}] \\
&= -\frac{1}{2v_i} \Delta s_i \frac{\min[|f'(t_i)|, |g'(t_i)|]}{\max[|f'(t_i)|, |g'(t_i)|]} \\
&\quad + \frac{1}{v_i} \Delta s_i \log \left[\sqrt{1 + \left(\frac{\min[|f'(t_i)|, |g'(t_i)|]}{\max[|f'(t_i)|, |g'(t_i)|]} \right)^2} \right] \\
&\quad + \frac{1}{v_i} \Delta s_i \log[\varepsilon_s] \\
&= -\frac{1}{2} \Delta t_i \frac{\min[|f'(t_i)|, |g'(t_i)|]}{\max[|f'(t_i)|, |g'(t_i)|]} \\
&\quad + \Delta t_i \log \left[\sqrt{1 + \left(\frac{\min[|f'(t_i)|, |g'(t_i)|]}{\max[|f'(t_i)|, |g'(t_i)|]} \right)^2} \right] \\
&\quad + \Delta t_i \log[\varepsilon_s]. \tag{A9}
\end{aligned}$$

Summing these \mathcal{S}_i over i , we obtain

$$\begin{aligned}
&\sum_{i=1}^M \left(-\frac{1}{2} \Delta t_i \frac{\min[|f'(t_i)|, |g'(t_i)|]}{\max[|f'(t_i)|, |g'(t_i)|]} \right) \\
&\quad + \sum_{i=1}^M \Delta t_i \log \left[\sqrt{1 + \left(\frac{\min[|f'(t_i)|, |g'(t_i)|]}{\max[|f'(t_i)|, |g'(t_i)|]} \right)^2} \right] \\
&\quad + \sum_{i=1}^M \Delta t_i \log[\varepsilon_s]. \tag{A10}
\end{aligned}$$

The last term sums to $\log[\varepsilon_s]$ because the time required to traverse the curve is one. The remaining terms become integrals in the limit where $\varepsilon \rightarrow 0$. Collecting all terms

$$\begin{aligned}
H^{\Delta}(f, g) &\rightarrow -\int_0^1 dt \log[v(t)] - \frac{1}{2} \int_0^1 dt \frac{\min[|f'(t)|, |g'(t)|]}{\max[|f'(t)|, |g'(t)|]} \\
&\quad + \int_0^1 dt \log \left[\sqrt{1 + \left(\frac{\min[|f'(t)|, |g'(t)|]}{\max[|f'(t)|, |g'(t)|]} \right)^2} \right] \\
&\quad + \log[\varepsilon_s], \tag{A11}
\end{aligned}$$

as $\varepsilon_s \rightarrow 0$. Because the form of the singularity in this limit $\log[\varepsilon_s]$ is independent of $f(t)$ and $g(t)$, it makes sense to define

$$\begin{aligned}
H_{f,g} &\equiv -\int_0^1 dt \log[v(t)] - \frac{1}{2} \int_0^1 dt \frac{\min[|f'(t)|, |g'(t)|]}{\max[|f'(t)|, |g'(t)|]} \\
&\quad + \int_0^1 dt \log \left[\sqrt{1 + \left(\frac{\min[|f'(t)|, |g'(t)|]}{\max[|f'(t)|, |g'(t)|]} \right)^2} \right]. \tag{A12}
\end{aligned}$$

We observe that this expression is completely symmetric in $f(t)$ and $g(t)$ as expected.

Moreover, because

$$\sqrt{1 + \left(\frac{\min[|f'(t)|, |g'(t)|]}{\max[|f'(t)|, |g'(t)|]} \right)^2}$$

is equal to

$$\begin{aligned}
&= \sqrt{\left(\frac{\max[|f'(t)|, |g'(t)|]}{\max[|f'(t)|, |g'(t)|]} \right)^2 + \left(\frac{\min[|f'(t)|, |g'(t)|]}{\max[|f'(t)|, |g'(t)|]} \right)^2} \\
&= \sqrt{\frac{\min[|f'(t)|, |g'(t)|]^2 + \max[|f'(t)|, |g'(t)|]^2}{\max[|f'(t)|, |g'(t)|]^2}} \\
&= \sqrt{\frac{f'(t)^2 + g'(t)^2}{\max[|f'(t)|, |g'(t)|]^2}} \\
&= \frac{v(t)}{\sqrt{\max[|f'(t)|, |g'(t)|]^2}}, \tag{A13}
\end{aligned}$$

Eq. (A11) becomes

$$\begin{aligned}
H_{f,g} &\equiv -\frac{1}{2} \int_0^1 dt \frac{\min[|f'(t)|, |g'(t)|]}{\max[|f'(t)|, |g'(t)|]} \\
&\quad - \int_0^1 dt \log \left[\max[|f'(t)|, |g'(t)|] \right]. \tag{A14}
\end{aligned}$$

Before proceeding to the claims, we point out that in this study, we consider only the case where self-intersections of $\gamma(t)$ are isolated as shown in Figs. 1, 12, and 14. Inspection of our experimental data shows that this assumption is valid in all but a negligible number of cases, which have been “masked” as described in the following text and excluded from analysis. In the case of complete overlap, where $f(t) = g(t)$, the joint entropy $H_{f,g}$ reduces effectively to H_f entropy published previously.^{6,7} An intermediate case, of total overlap for part of $\gamma(t)$ and isolated self-intersection on other portions can be handled by breaking up the integrals we will obtain into separate portions.

APPENDIX B: JUSTIFICATION OF CLAIM I

We begin by establishing the joint density $w(x, y)$ as the (Schwartz) distribution obtained from the cumulative joint density function, $W(x, y)$, for $x = f(t)$ and $y = g(t)$. While $W(x, y)$ is a conventional function of x and y , $w(x, y)$ is not, so that a course-graining operation on $w(x, y)$ using test functions must be defined. This leads to the main result of this section, Eq. (B28). Although this permits computation of the course-grained joint-entropy, we must also verify consistency of this approach with that used to obtain previous entropies, e.g., H_f , $I_f(r)$, and $I_{f,\infty}$, which were based on the density function, $w_f(y)$ for a single function $y = f(t)$.^{6-8,12,22} This is established by showing that the marginal density functions obtained from $w(x, y)$ are just the single function densities employed previously. The same results are established for the course-grained marginal probabilities.

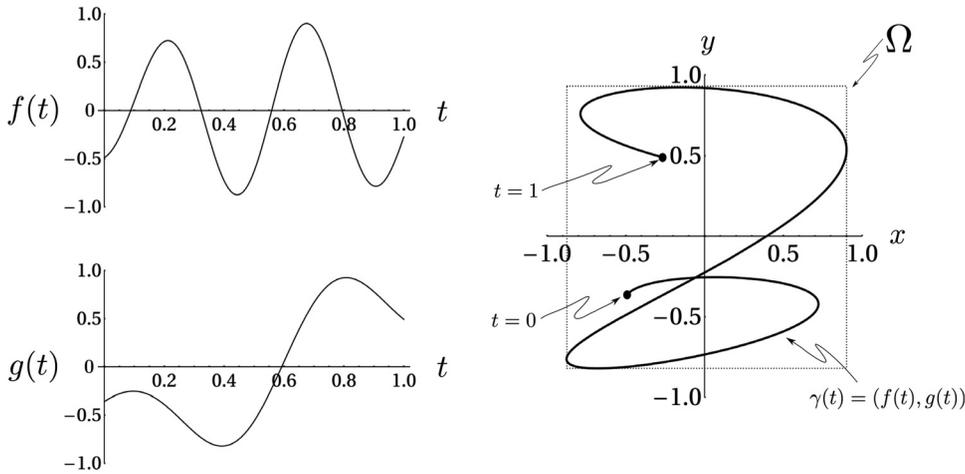


FIG. 12. Examples of two time-domain functions, $f(t) = x$, $g(t) = y$ (left side) and the path they define, $\gamma(t)$, in the x, y plane (right side). The action of the Schwartz distribution $w(x, y)$ in a two dimensional integral is to collapse it into an equivalent one dimensional integral over time.

1. Calculating $\partial W(x, y)/\partial x \partial y$ using Schwartz distributions

Typically, $w(x, y)$ is defined as the second partial derivative (with respect to x and y) of the joint cumulative density function $W(x, y)$. However, for the types of random variables, we consider (assumed to be infinitely differentiable) the joint density requires further discussion since, unlike $w_f(x)$ and $w_g(y)$,^{7,12,22} it is not a regular function (and is defined in the following text as a distribution in the sense of L. Schwartz). So we must start with the joint cumulative density function, $W(x, y)$, for $f(t) = x$, $g(t) = y$, an example of which is shown in Fig. 1. This is a regular function defined by

$$W(x, y) = |\{t \in \mathbb{R} : f(t) \leq x \text{ and } g(t) \leq y\}|, \quad (\text{B1})$$

where $|\{\cdot\}|$ denotes the measure (“length”) of the set. The right side of Fig. 1 has an example calculation of $W(0.5, 0.75)$ directly from Eq. (B1).

On the left side of the figure, an alternative (and eventually more useful) way of looking at $W(x, y)$ is shown. The trajectory defined by

$$\gamma(t) = (f(t), g(t)), \quad (\text{B2})$$

parameterized by time running from $0 \rightarrow 1$, is shown. Also shown is a red box having lower left corner at the point $(x, y) = (-1, -1)$ and its upper right corner at $(x, y) = (0.5, 0.75)$. The portions of the trajectory within this rectangular region are colored red, and the entry and exit times of the red portions are labeled to show their correspondence with times in the right-hand side of the figure. The sum of these times is $W(0.5, 0.75)$, so that $W(x, y)$ measures the time that the trajectory spends in the red box.

The figure also suggests that the derivative of $W(x, y)$ will exhibit jump discontinuities for certain values of x and y as the number of segments of $\gamma(t)$ contained in the red rectangle changes discontinuously, for instance, as the height of the red rectangle is adjusted so that its top moves from below to just above the point labeled t_0 . This result may also be anticipated by formally differentiating the integral representation of $W(x, y)$ given in Eq. (B4). Consequently, careful definition of $w(x, y)$ will require the use of Schwartz distributions.

If $H(x)$ is the Heaviside function

$$H(x) = \begin{cases} 0 & x < 0, \\ 1 & x \geq 0, \end{cases} \quad (\text{B3})$$

then

$$W(x, y) = \int_0^1 H(x - f(t))H(y - g(t)) dt. \quad (\text{B4})$$

We note that the joint density $w(x, y) = (\partial^2/\partial x \partial y)W(x, y)$ appears only in expressions like

$$I = \int_{-1}^1 dx \int_{-1}^1 dy \phi(x, y)w(x, y). \quad (\text{B5})$$

While $w(x, y)$ is not a regular “point” function, we recognize that it can be defined as a Schwartz distribution. Thus we assume that we may freely interchange orders of integration and that if $\phi(x, y) \in C_c^\infty(\Omega)$ [recall $\Omega = (-1, 1) \times (-1, -1)$], we obtain

$$\begin{aligned} I &= \int_{-1}^1 dx \int_{-1}^1 dy \phi(x, y)w(x, y) \\ &= \int_{-1}^1 dx \int_{-1}^1 dy \phi(x, y) \frac{\partial^2 W(x, y)}{\partial x \partial y} \\ &= \int_{-1}^1 dx \int_{-1}^1 dy \frac{\partial^2 \phi(x, y)}{\partial x \partial y} W(x, y), \end{aligned} \quad (\text{B6})$$

$$\begin{aligned} I &= \int_{-1}^1 dx \int_{-1}^1 dy \frac{\partial^2 \phi(x, y)}{\partial x \partial y} \int_0^1 dt H(x - f(t)) \\ &\quad \times H(y - g(t)) \end{aligned} \quad (\text{B7})$$

$$\begin{aligned} &= \int_0^1 dt \int_{-1}^1 dx \int_{-1}^1 dy \frac{\partial^2 \phi(x, y)}{\partial x \partial y} H(x - f(t)) \\ &\quad \times H(y - g(t)) \end{aligned} \quad (\text{B8})$$

$$\begin{aligned} &= \int_0^1 dt \int_{-1}^1 dx H(x - f(t)) \\ &\quad \times \left[\frac{\partial \phi(x, y)}{\partial x} H(y - g(t)) \Big|_{y=-1}^{y=1} - \int_{-1}^1 dy \frac{\partial \phi(x, y)}{\partial x} \frac{\partial}{\partial y} H(y - g(t)) \right], \end{aligned} \quad (\text{B9})$$

$$I = - \int_0^1 dt \int_{-1}^1 dx H(x - f(t)) \times \int_{-1}^1 dy \frac{\partial \phi(x, y)}{\partial x} \frac{\partial}{\partial y} H(y - g(t)), \quad (\text{B10})$$

where the boundary terms vanished because $\phi(x, y) \in C_c^\infty(\Omega)$. Using

$$\int_{-1}^1 dy F(x, y) \frac{\partial}{\partial y} H(y - z) = F(x, z), \quad (\text{B11})$$

with

$$F(x, y) = \frac{\partial \phi(x, y)}{\partial x}, \quad (\text{B12})$$

Eq. (B10) becomes

$$I = - \int_0^1 dt \int_{-1}^1 dx H(x - f(t)) \frac{\partial \phi(x, g(t))}{\partial x}. \quad (\text{B13})$$

Integrating by parts again leads to

$$I = - \int_0^1 dt \left[\phi(x, g(t)) H(x - f(t)) \Big|_{x=-1}^{x=1} - \int_{-1}^1 dx \frac{\partial}{\partial x} H(x - f(t)) \phi(x, g(t)) \right] \quad (\text{B14})$$

$$= \int_0^1 dt \int_{-1}^1 dx \frac{\partial}{\partial x} H(x - f(t)) \phi(x, g(t)), \quad (\text{B15})$$

where once again the boundary terms vanish because $\phi(x, y) \in C_c^\infty(\Omega)$. Using Eqs. (B11) and (B12), Eq. (B15) becomes

$$I = \int_0^1 dt \phi(f(t), g(t)). \quad (\text{B16})$$

Combining this result with Eq. (B5), we see that, as a Schwartz distribution, $w(x, y)$ obeys the relation

$$\int_{-1}^1 dx \int_{-1}^1 dy \phi(x, y) w(x, y) = \int_0^1 dt \phi(f(t), g(t)). \quad (\text{B17})$$

This generalizes the expression for single (differentiable) random variable density function

$$\int_0^1 \phi(y) w_f(y) dy = \int_0^1 \phi(f(t)) dt, \quad (\text{B18})$$

which we have employed in the definition of H_f as described in previous studies.^{6–8,12,22}

2. Where $w(x, y)$ “lives” in terms of $f(t)$ and $g(t)$

Equation (B16) can be visualized using the picture shown on the right-hand panel of Fig. 12. The fact that $w(x, y)$ is represented by a Schwartz distribution means that the “probability” is effectively concentrated on the curve $\gamma(t)$. Thus the current situation differs significantly from

typical applications of jointly distributed random variables, which are usually distributed continuously over two-dimensional regions of the real plane.

3. Marginal distributions

Consistency of the current approach with previous studies requires that the marginal density functions for $w(x, y)$ obtained using Eq. (B17) should be identical with the single (differential) random variable density functions of Eq. (B18). To check this, we let $\phi(x) \in C_c^\infty(\Omega)$ and compute

$$\begin{aligned} I &= \int_{f_{\min}}^{f_{\max}} dx \int_{g_{\min}}^{g_{\max}} dy \phi(x) w(x, y) \\ &= \int_0^1 dt \phi(f(t)) \\ &= \int_{f_{\min}}^{f_{\max}} dx \phi(x) w_f(x). \end{aligned} \quad (\text{B19})$$

But also

$$\begin{aligned} I &= \int_{f_{\min}}^{f_{\max}} dx \int_{g_{\min}}^{g_{\max}} dy \phi(x) w(x, y) \\ &= \int_{f_{\min}}^{f_{\max}} dx \phi(x) \int_{g_{\min}}^{g_{\max}} dy w(x, y). \end{aligned} \quad (\text{B20})$$

Equations (B19) and (B20) imply

$$w_f(x) = \int_{g_{\min}}^{g_{\max}} dy w(x, y), \quad (\text{B21})$$

and similarly

$$w_g(y) = \int_{f_{\min}}^{f_{\max}} dx w(x, y), \quad (\text{B22})$$

as required.

4. Choices of $\phi(x, y)$ for coarse-graining

The integral equations, Eqs. (B17) and (B18), defining $w(x, y)$ and $w_f(y)$ are all we need to compute the coarse-grained joint probabilities, $p^\Delta(j, k)$, and the associated coarse-grained marginal probabilities, $p^\Delta(j)$ and $p^\Delta(k)$, which we will need as intermediate quantities to obtain $H_{f, g}$.

This is accomplished by choosing an appropriate function $\phi(x, y)$ to use in Eq. (B17) [or for the marginal probabilities: choosing an appropriate $\phi(y)$ to use in Eq. (B18)]. Let $0 < \varepsilon \ll 1$ and let $0 < s \ll 1$ be a parameter used to control the sharpness of the edges of test functions $\phi(x, y)$, which will be used to effect coarse-graining of the Schwartz distribution $w(x, y)$ and the density functions $w_f(x)$, and $w_g(y)$. We partition Ω into small square cells, $\varepsilon \times \varepsilon$ in size, over which we will compute integrals of product of test functions $\phi(x, y)$ and $w(x, y)$, $w_f(x)$, or $w_g(y)$. The test function will be chosen so that as $s \rightarrow 0$, the coarse-graining test functions approach unit height square waves that sample $w(x, y)$, $w_f(x)$, or $w_g(y)$ over a small nonzero region. The aim is to simulate the operation of a digitizer that is sampling a one or

two-dimensional function. We begin by defining an infinitely differentiable smooth function that “turns on” at $x = j\varepsilon$ as x increases,

$$\phi_{L,j}(x) = \begin{cases} e^{-1/(s(x-j\varepsilon))^2}, & x > j\varepsilon; \\ 0, & x \leq j\varepsilon, \end{cases} \quad (\text{B23})$$

and a corresponding smooth function that turns off at $x = (j + 1)\varepsilon$ as x increases

$$\phi_{R,j}(x) = \begin{cases} 0, & x \geq (j + 1)\varepsilon, \\ e^{-1/(s(x-(j+1)\varepsilon))^2}, & x < (j + 1)\varepsilon. \end{cases} \quad (\text{B24})$$

Their product defines an infinitely differentiable step function,

$$\phi_j(x) = \phi_{L,j}(x)\phi_{R,j}(x), \quad (\text{B25})$$

which approaches a unit height square wave, turning on at $x = j\varepsilon$ and turning off at $x = (j + 1)\varepsilon$, as $s \rightarrow 0$. This function may be used to “sample” a function of one variable, e.g., $f(t) = x$. Similarly $\phi_k(y)$ is defined as an infinitely smooth precursor to a square wave that turns on at $y = k\varepsilon$ and turns off at $y = (k + 1)\varepsilon$. Their product may be used to define the infinitely differentiable (in x and y) function

$$\phi_{j,k}(x, y) = \phi_j(x)\phi_k(y), \quad (\text{B26})$$

which may be used to sample functions (or Schwartz distributions) of two variables at $x = j\varepsilon$, $y = k\varepsilon$. Example $\phi_j(x)$, $\phi_k(y)$, $\phi_{j,k}(x, y)$ [which are all $\in C_c^\infty(\Omega)$] are shown in Fig. 13.

We will also allow the two indices, k associated with the $f(t) = x$ variable and j associated with the $g(t) = y$ variable to run from $-\infty$ to ∞ .

5. Course-grained probabilities $p^\Delta(\mathbf{x}, \mathbf{y})$, $p^\Delta(\mathbf{x})$, and $p^\Delta(\mathbf{y})$ from $w(\mathbf{x}, \mathbf{y})$, $w_f(\mathbf{x})$, and $w_g(\mathbf{y})$

Now we have all we need to coarse-grain the densities $w(x, y)$, $w_f(x)$ and $w_g(y)$. Suppose that $\gamma(t)$ passes only once through the j, k th cell, which we will denote by $\mathcal{C}_{j,k}$. To coarse-grain $w(x, y)$ over $\mathcal{C}_{j,k}$, we compute,

$$\begin{aligned} p^\Delta(j, k) &= \lim_{s \rightarrow \infty} \int_{f_{\min}}^{f_{\max}} dx \int_{g_{\min}}^{g_{\max}} dy w(x, y) \phi_{j,k}(x, y) \\ &= \lim_{s \rightarrow \infty} \int_0^1 dt \phi_{j,k}(f(t), g(t)) \\ &= \lim_{s \rightarrow \infty} \int_{t_i}^{t_f} dt \phi_{j,k}(f(t), g(t)) \\ &= \lim_{s \rightarrow \infty} \int_{t_i}^{t_f} dt = t_f - t_i, \end{aligned} \quad (\text{B27})$$

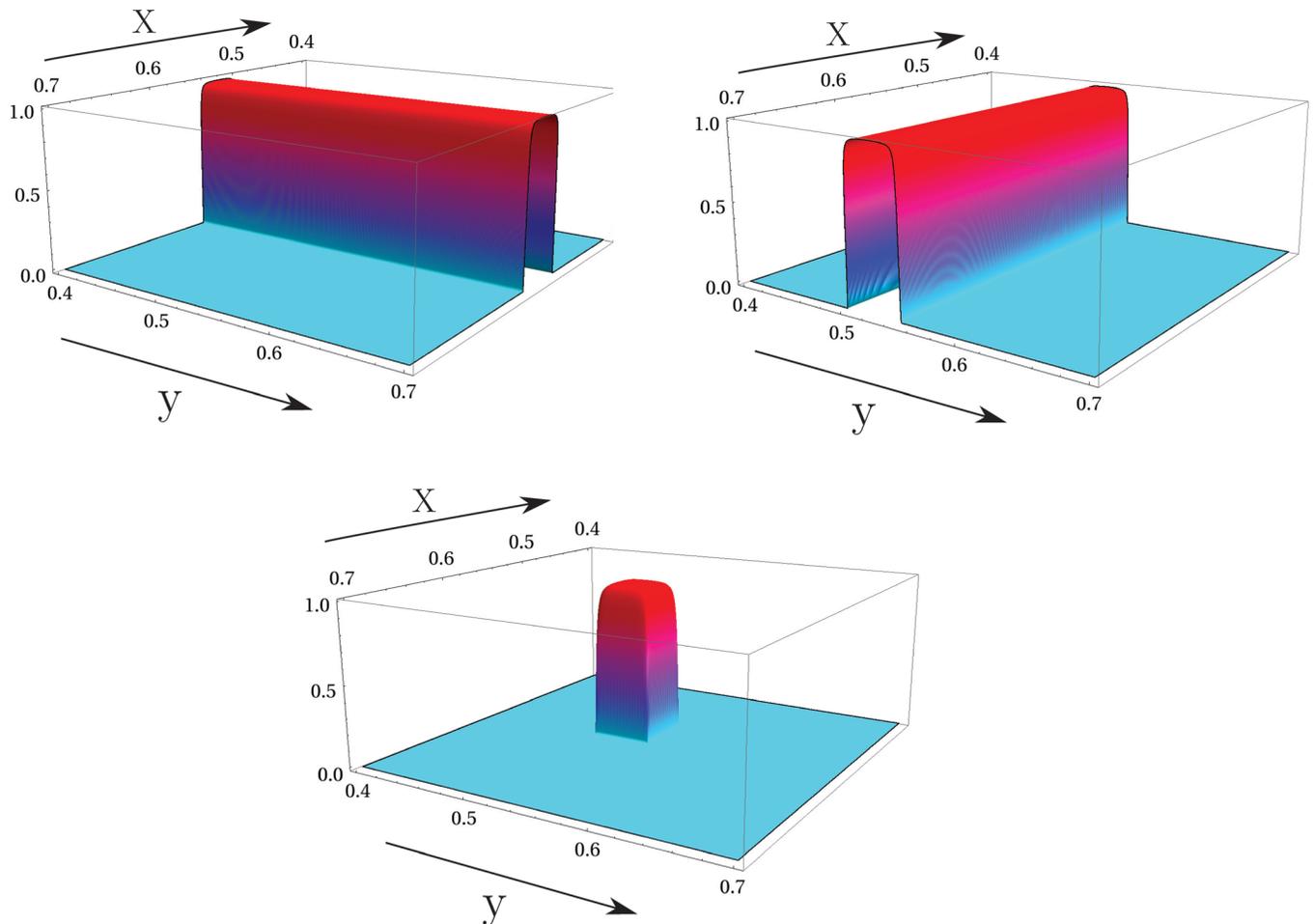


FIG. 13. (Color online) Examples of course graining functions, $\phi_j(x)$ (top left), $\phi_k(y)$ (top right), and $\phi_{j,k}(x, y)$ (bottom).

where $\phi_{j,k}(x, y)$ is defined by Eq. (B26). For cells that do not intersect $\gamma(t)$, we define $p^\Delta(j, k) = 0$ for completeness. Defining $s(t_i)$ to be the length along $\gamma(t)$ between $t = 0$ and $t = t_i$, Eq. (B27) may be rewritten as

$$p^\Delta(j, k) = t_f - t_i = \Delta t = \frac{s(t_f) - s(t_i)}{v(t_i)}, \quad (\text{B28})$$

where $v(t_i) = \sqrt{f'(t_i)^2 + g'(t_i)^2}$ is the velocity along the curve at time t_i . If more than one passage through occurs $\mathcal{C}_{j,k}$, then all traversal times t_i such that $(f(t_i), g(t_i)) \in \partial \mathcal{C}_{j,k}$ must be added, and we obtain

$$p^\Delta(j, k) = \sum_{\{t_i \in \partial \mathcal{C}_{ij}\}} \frac{s(t_f) - s(t_i)}{v(t_i)}. \quad (\text{B29})$$

6. Discrete marginal probabilities: From the standard definition

The discrete marginal probabilities are defined by

$$p^\Delta(k) = \sum_j p^\Delta(j, k). \quad (\text{B30})$$

For the course-grained marginal probabilities, we also set $p^\Delta(k) = 0$ if $\gamma(t)$ does not intersect any cell with index j ($\mathcal{C}_{j,k} \cap \gamma(t) = \emptyset, \forall k$). Similarly we define,

$$p^\Delta(j) = \sum_k p^\Delta(j, k). \quad (\text{B31})$$

7. Discrete marginal probabilities: From Eqs. (B21), (B22), and (B25)

Alternatively, we may start with Eq. (B21) and define

$$p^\Delta(j) = \int_{f_{\min}}^{f_{\max}} dx w(x, y) \phi_j(x), \quad (\text{B32})$$

where $\phi_j(x)$ is given by Eq. (B25). Similarly, using Eq. (B22),

$$p^\Delta(k) = \int_{g_{\min}}^{g_{\max}} dy w(x, y) \phi_k(y), \quad (\text{B33})$$

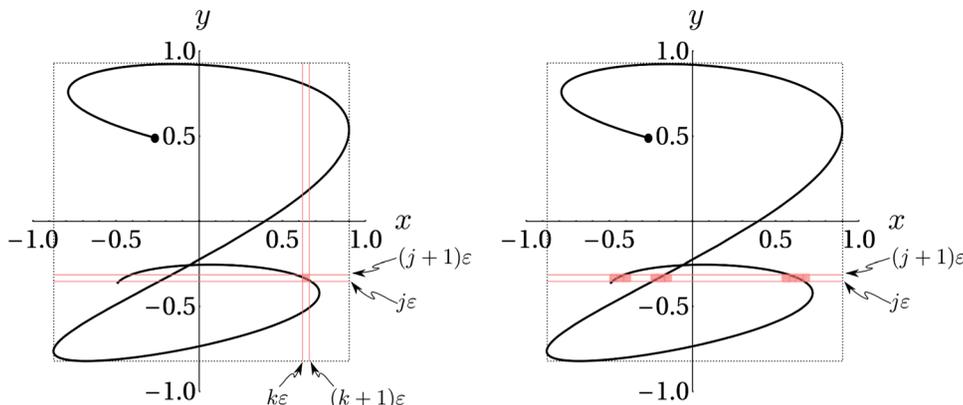


FIG. 14. (Color online) Computation of $p(j, k)$ (left). Computation of the marginal probability $p(k)$ (right).

where $\phi_k(y)$ is also given by Eq. (B25) and $p^\Delta(j), p^\Delta(k)$ are zero in the same cases described after Eqs. (B30) and (B31). Focusing on Eq. (B29), we see from Fig. 14, that the integral will reduce, in the limit where $s \rightarrow \infty$, to the sum of times spent in the cells lying between $j\epsilon$ and $(j+1)\epsilon$, so that we again obtain

$$p^\Delta(j) = \sum_k p^\Delta(j, k) \quad (\text{B34})$$

and

$$p^\Delta(k) = \sum_j p^\Delta(j, k). \quad (\text{B35})$$

We note based on previous work that as $\epsilon \rightarrow 0$ these are asymptotic to

$$p^\Delta(j) \sim \epsilon w_f(-1 + j\epsilon) = \epsilon w_f(x) \quad (\text{B36})$$

and

$$p^\Delta(k) \sim \epsilon w_g(-1 + k\epsilon) = \epsilon w_g(y). \quad (\text{B37})$$

These are just the coarse-grained versions of Eqs. (B21) and (B22) and demonstrate the consistency of computing marginals and then coarse-graining with coarse-graining and then computing marginals.

APPENDIX C: JUSTIFICATION OF CLAIM ii

While Eq. (B28) gives the probabilities needed to compute joint entropies in terms of time, Δt_i , we must express it in terms of experimentally accessible quantities, which means quantities derived from the backscattered RF, $f(t)$, and the reference signal $g(t)$. The first step in this process is Eq. (A1), which relates Δt_i to Δs_i using the velocity in the cell $v_i = \sqrt{f'(t)^2 + g'(t)^2}$. However, as discussed in the preceding text, we are still faced by an apparently unsolvable problem because we do not know where $\gamma(t)$ enters the $\epsilon \times \epsilon$ cell. Moreover, we also must eliminate the coarse-graining parameter, ϵ from the final expressions for joint entropy, which is accomplished by taking $\epsilon \rightarrow 0$. In this section, we provide the details of our approach beginning with the “stacking” process used above which may be made rigorous using the ergodic theorem²³ as we now describe. Thus we

suppose that we are given a straight line segment, ℓ , of slope $\alpha = g'(t_i)/f'(t_i)$, and length Δs_i , contained in a cell of dimensions $\varepsilon \times \varepsilon$, such as that shown in Fig. 10. We wish to subdivide the cell into $\varepsilon_s \times \varepsilon_s$ sub-cells as also shown in the figure, and compute the limiting form of

$$S_i \equiv \sum_{n=1}^{M_s} \Delta s_{i,n} \log[\Delta s_{i,n}], \quad (\text{C1})$$

where

$$\Delta s_i = \sum_{n=1}^{M_s} \Delta s_{i,n}, \quad (\text{C2})$$

and the index n keeps track of the $\varepsilon_s \times \varepsilon_s$ cells crossed as $\gamma(t)$ traverses the larger $\varepsilon \times \varepsilon$ cell. We begin, as shown in Fig. 11, by imagining that we translate all the subsegments of ℓ , to the $\varepsilon_s \times \varepsilon_s$ square, as shown in the left hand side of the figure (this is equivalent to the stacking procedure described in the preceding text). The y coordinates of those line segments that start out on the y axis are shown the figure. These are given by

$$T_y^{(r)}(y_i) = y_i + r\alpha\varepsilon_s \bmod \varepsilon_s, \quad (\text{C3})$$

where $T_y^{(r)}(y_i)$ is the r th iteration of the mapping

$$T_y(y) = y + \alpha \bmod \varepsilon_s. \quad (\text{C4})$$

Similarly the x coordinates of those line segments that start out on the x axis are given by

$$T_x^{(r)}(x_i) = x_i + r\frac{1}{\alpha}\varepsilon_s \bmod \varepsilon_s, \quad (\text{C5})$$

where $T_x^{(r)}(x_i)$ is the r th iteration of the mapping

$$T_x(x) = x + \frac{1}{\alpha} \bmod \varepsilon_s. \quad (\text{C6})$$

Thus the physical ‘‘stacking’’ procedure is implemented mathematically by iterating the maps $T_x(x)$ and $T_y(y)$. Because $g'(t_i)/f'(t_i)$ is irrational, we know that both of these mappings are ergodic.²³ Hence we have the following relations between sums and integrals involving continuous functions $\phi(x)$, $\psi(y)$ and iterates of $T_x(x)$ and $T_y(y)$,

$$\sum_{r=0}^{M_{s_x}-1} \phi\left(T_x^{(r)}(x_i)\right) \rightarrow M_{s_x} \int_0^1 dx \phi(x), \quad (\text{C7})$$

as $M_{s_x} \rightarrow \infty$ and,

$$\sum_{r=0}^{M_{s_y}-1} \psi\left(T_y^{(r)}(y_i)\right) \rightarrow M_{s_y} \int_0^1 dy \psi(y), \quad (\text{C8})$$

as $M_{s_y} \rightarrow \infty$. While it is possible to use Eqs. (C7) and (C8) to obtain Eq. (A9), we will not do so. Instead, it is simpler to refer back to Fig. 11 and observe that M_{s_x} is the number of times that the line segment $y = \alpha x$ crosses the vertical lines

$x = m\varepsilon_s$, where m is a non-negative integer, and M_{s_y} is the number of times that the line segment $y = \alpha x$ crosses the horizontal lines $x = q\varepsilon_s$, where q is a non-negative integer. Furthermore, we observe that the ergodic theorem tells us, via Eqs. (C7) and (C8), that as $s \rightarrow 0$, the crossing points become uniformly distributed along the edges of the $\varepsilon_s \times \varepsilon_s$ square in Fig. 11. This means that the random variable y shown in Fig. 15 is uniformly distributed. Thus the variable θ , which is a linear function of y , is also uniformly distributed over its range (the interval $[0, 1]$). We will use this implication of the ergodic theorem to obtain Eq. (A9).

1. Relation between M_{s_x} and M_{s_y}

Let s_h be the horizontal distance traversed and $s_v = |\alpha|s_h$ be the vertical distance traversed (so $\Delta s_i = \sqrt{1 + \alpha^2}s_h$). Then

$$M_{s_x} \in \{\lfloor s_h/\varepsilon_s \rfloor, \lceil s_h/\varepsilon_s \rceil\}, \quad (\text{C9})$$

where $\lfloor z \rfloor$ is the largest integer $\leq z$, and $\lceil z \rceil$ is the smallest integer $\geq z$. Similarly

$$M_{s_y} \in \{\lfloor s_v/\varepsilon_s \rfloor, \lceil s_v/\varepsilon_s \rceil\}. \quad (\text{C10})$$

As $s_v = |\alpha|s_h$, we get $M_{s_y} \rightarrow |\alpha|M_{s_x}$ as $\varepsilon_s \rightarrow 0$.

2. Calculation of Δs_i

We have

$$\Delta s_i = s_h \sqrt{1 + \alpha^2} = s_v \sqrt{1 + \frac{1}{\alpha^2}} \quad (\text{C11})$$

and

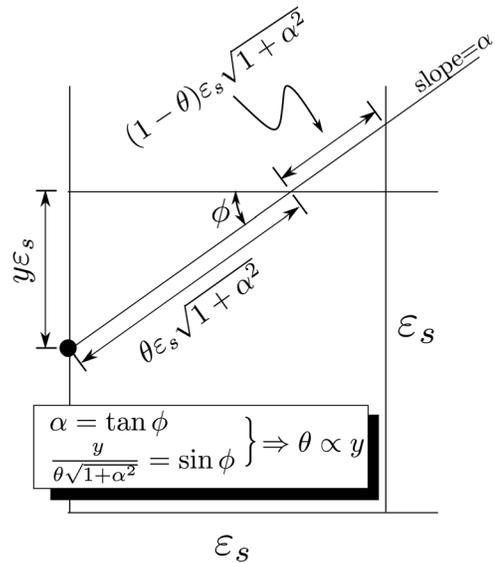


FIG. 15. The black circle is a point where $\gamma(t)$ crosses the vertical line in the $\varepsilon_s \times \varepsilon_s$ sub-grid shown in Fig. (11). Because α is irrational, the values of y are uniformly distributed along the vertical side of the $\varepsilon_s \times \varepsilon_s$ square by the ergodic theorem. This also implies that the variable θ is uniformly distributed over its range, the interval $[0, 1]$, because it is a linear function of y , as shown in the figure.

$$s_h = \varepsilon_s M_{s_x} \quad (\text{C12})$$

with an error of at most ε_s by Eq. (C9). Therefore

$$\Delta s_i \approx M_{s_x} \varepsilon_s \sqrt{1 + \alpha^2} = M_{s_x} \varepsilon', \quad (\text{C13})$$

where we have defined the symbol $\varepsilon' = \varepsilon_s \sqrt{1 + \alpha^2}$ to simplify calculations in the following text.

3. Calculation of S_i

We are now ready to calculate

$$S_i \equiv \sum_{n=1}^{M_S} \Delta s_{i,n} \log[\Delta s_{i,n}].$$

We break the calculation into two cases: $|\alpha| < 1$ and $|\alpha| > 1$.

Case $|\alpha| < 1$: We note that $M_S = M_{s_x} + M_{s_y}$ is the number of times that the trajectory, $\gamma(t)$, crosses either a vertical or a horizontal line of the $\varepsilon_s \times \varepsilon_s$ sub-grid. (Because α is irrational, there is at most one occasion where it crosses both simultaneously; we can ignore this event as $\varepsilon_s \rightarrow 0$.)

If there are two successive vertical crossings, then

$$\Delta s_{i,n} \log[\Delta s_{i,n}] = \varepsilon_s \sqrt{1 + \alpha^2} \log[\varepsilon_s \sqrt{1 + \alpha^2}]. \quad (\text{C14})$$

As $|\alpha| < 1$, if this does not happen, then there is exactly one horizontal crossing in between vertical crossings at n and $n + 2$, and

$$\begin{aligned} & \Delta s_{i,n} \log[\Delta s_{i,n}] + \Delta s_{i,n+1} \log[\Delta s_{i,n+1}] \\ &= \theta \varepsilon_s \sqrt{1 + \alpha^2} \log[\theta \varepsilon_s \sqrt{1 + \alpha^2}] \\ &+ (1 - \theta) \varepsilon_s \sqrt{1 + \alpha^2} \log[(1 - \theta) \varepsilon_s \sqrt{1 + \alpha^2}] \end{aligned} \quad (\text{C15})$$

for some $0 < \theta < 1$. By the ergodic theorem, θ will tend to a uniformly distributed random variable on $[0, 1]$ as shown in Fig. 15. So letting $\varepsilon' = \varepsilon_s \sqrt{1 + \alpha^2}$, the contribution from Eq. (C15) is, on average,

$$\int_0^1 d\theta [\theta \varepsilon' \log[\theta \varepsilon'] + (1 - \theta) \varepsilon' \log[(1 - \theta) \varepsilon']], \quad (\text{C16})$$

or

$$\varepsilon' \log \varepsilon' - \frac{\varepsilon'}{2}. \quad (\text{C17})$$

Of the M_{s_x} vertical crossings, $M_{s_y} = |\alpha| M_{s_x}$ of them will have a horizontal crossing “sandwiched” in between and contribute the quantity shown in Eq. (C17) on average; the remainder will contribute $\varepsilon' \log[\varepsilon']$. Thus we obtain

$$\begin{aligned} S_i &= |\alpha| M_{s_x} \left[\varepsilon' \log \varepsilon' - \frac{\varepsilon'}{2} \right] + (1 - |\alpha|) M_{s_x} [\varepsilon' \log \varepsilon'] \\ &= M_{s_x} \varepsilon' \left\{ |\alpha| \left[\log \varepsilon' - \frac{1}{2} \right] + (1 - |\alpha|) \log \varepsilon' \right\} \\ &= \Delta s_i \left\{ -\frac{|\alpha|}{2} + \log \varepsilon' \right\} \\ &= \Delta s_i \left\{ -\frac{|\alpha|}{2} + \log \sqrt{1 + \alpha^2} + \log \varepsilon_s \right\}. \end{aligned} \quad (\text{C18})$$

Case $|\alpha| > 1$: If $|\alpha| > 1$, we repeat the calculation with x and y “switched.” The calculation proceeds exactly as before although α must be replaced by $1/\alpha$.

Calculation summary: These results may be concisely written as

$$S_i = \Delta s_i \begin{cases} \left[-\frac{1}{2} |\alpha| + \log[\sqrt{1 + \alpha^2}] \right] + \log[\varepsilon_s] & |\alpha| \leq 1, \\ \left(-\frac{1}{2} \frac{1}{|\alpha|} + \log \left[\sqrt{1 + \frac{1}{\alpha^2}} \right] \right) + \log[\varepsilon_s] & |\alpha| \geq 1 \end{cases} \quad (\text{C19})$$

or, recalling that $\alpha = g'(t)/f'(t)$,

$$\begin{aligned} S_i &= -\frac{1}{2} \Delta s_i \frac{\min[|f'(t_k)|, |g'(t_k)|]}{\max[|f'(t_k)|, |g'(t_k)|]} \\ &+ \Delta s_i \log \left[\sqrt{1 + \left(\frac{\min[|f'(t_k)|, |g'(t_k)|]}{\max[|f'(t_k)|, |g'(t_k)|]} \right)^2} \right] \\ &+ \Delta s_i \log[\varepsilon_s], \end{aligned} \quad (\text{C20})$$

which is Eq. (A9). This completes the derivation of $H_{f,g}$.

¹M. Hughes, “Analysis of ultrasonic waveforms using Shannon entropy,” Proc.-IEEE Ultrason. Symp. **2**, 1205–1209 (1992).

²M. Hughes, “Analysis of digitized waveforms using Shannon entropy,” J. Acoust. Soc. Am. **93**, 892–906 (1993).

³M. Hughes, “NDE imaging of flaws using rapid computation of Shannon entropy,” Proc.-IEEE Ultrason. Symp. **2**, 697–700 (1993).

⁴M. Hughes, “Analysis of digitized waveforms using Shannon entropy. II. High-speed algorithms based on Green’s functions,” J. Acoust. Soc. Am. **95**, 2582–2588 (1994).

⁵M. S. Hughes, J. N. Marsh, C. S. Hall, D. Savy, M. J. Scott, J. S. Allen, E. K. Lacy, C. Carradine, G. M. Lanza, and S. A. Wickline, “Characterization of digital waveforms using thermodynamic analogs: Applications to detection of materials defects,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1555–1564 (2005).

⁶M. Hughes, J. Marsh, A. Woodson, E. Lacey, C. Carradine, G. M. Lanza, and S. A. Wickline, “Characterization of digital waveforms using thermodynamic analogs: Detection of contrast targeted tissue in MDA 435 tumors implanted in athymic nude mice,” Proc.-IEEE Ultrason. Symp. **1**, 373–376 (2005).

⁷M. S. Hughes, J. E. McCarthy, J. N. Marsh, J. M. Arbeit, R. G. Neumann, R. W. Fuhrhop, K. D. Wallace, D. R. Znidarsic, B. N. Maurizi, S. L. Baldwin, G. M. Lanza, and S. A. Wickline, “Properties of an entropy-based signal receiver with an application to ultrasonic molecular imaging,” J. Acoust. Soc. Am. **121**, 3542–3557 (2007).

⁸M. Hughes, G. M. Marsh, Lanza, S. A. Wickline, J. McCarthy, B. Wickerhauser, V. annd Maurizi, and, W. K., “Improved signal processing to detect cancer by ultrasonic molecular imaging of targeted nanoparticles,” J. Acoust. Soc. Am. **129**, 3756–3767 (2011).

⁹J. N. Marsh, J. E. McCarthy, M. Wickerhauser, J. M. Arbeit, R. W. Fuhrhop, K. D. Wallace, G. M. Lanza, S. A. Wickline, and M. S. Hughes, “Application of real-time calculation of a limiting form of the Renyi entropy for molecular imaging of tumors,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **57**, 1890–1895 (2010).

¹⁰J. N. Marsh, K. D. Wallace, G. M. Lanza, S. A. Wickline, J. E. Hughes, and M. S. McCarthy, “Application of a limiting form of the Renyi entropy for molecular imaging of tumors using a clinically relevant protocol,” Proc.-IEEE Ultrason. Symp. **1**, 53–56 (2010).

¹¹K. D. Wallace, J. Marsh, S. L. Baldwin, A. M. Connolly, K. Richard, G. M. Lanza, S. A. Wickline, and M. S. Hughes, “Sensitive ultrasonic delineation of steroid treatment in living dystrophic mice with energy-based and entropy-based radio frequency signal processing,” IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 2291–2299 (2007).

¹²M. Hughes, J. Marsh, K. Wallace, T. Donahue, A. Connolly, G. M. Lanza, and S. A. Wickline, “Sensitive ultrasonic detection of dystrophic skeletal

- muscle in patients with Duchenne's muscular dystrophy using an entropy-based signal receiver," *Ultrasound Med. Biol.* **33**, 1236–1243 (2007).
- ¹³M. Hughes, J. Marsh, K. Agyem, J. McCarthy, B. Maurizi, M. Wickerhauser, W. K. D. G. Lanza, and S. Wickline, "Use of smoothing splines for analysis of backscattered ultrasonic waveforms: Application to monitoring of steroid treatment of dystrophic mice," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **58**, 2361–2369 (2011).
- ¹⁴R. Seip, E. Ebbini, and M. O'Donnell, "Non-invasive detection of thermal effects due to highly focused ultrasonic fields," *Proc.-IEEE Ultrason. Symp.* **2**, 1229–1232 (1993).
- ¹⁵M. S. Hughes, J. E. McCarthy, M. Wickerhauser, J. N. Marsh, J. M. Arbeit, R. W. Fuhrhop, K. D. Wallace, T. Thomas, J. Smith, K. Agyem, G. M. Lanza, and S. A. Wickline, "Real-time calculation of a limiting form of the Renyi entropy applied to detection of subtle changes in scattering architecture," *J. Acoust. Soc. Am.* **126**, 2350–2358 (2009).
- ¹⁶L. Schwartz, *Mathematics for the Physical Sciences* (Dover, New York, 2008), pp. 1–368.
- ¹⁷M. Hughes, "A comparison of Shannon entropy versus signal energy for acoustic detection of artificially induced defects in Plexiglas," *J. Acoust. Soc. Am.* **91**, 2272–2275 (1992).
- ¹⁸C. H. Reinsch, "Smoothing by spline functions," *Num. Math.* **10**, 177–183 (1967).
- ¹⁹D. L. Sackett, "Why randomized controlled trials fail but needn't. II. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!)," *Can. Med. Assoc. J.* **165**, 1226–1237 (2001).
- ²⁰J. M. Arbeit, R. R. Riley, B. Huey, C. Porter, G. Kelloff, R. Lubet, J. M. Ward, and D. Pinkel, "DFMO chemoprevention of epidermal carcinogenesis in k14-hpv16 transgenic mice," *Cancer Res.* **59**, 3610–3620 (1999).
- ²¹J. M. Arbeit, K. Mnger, P. M. Howley, and D. Hanahan, "Progressive squamous epithelial neoplasia in k14-human papillomavirus type 16 transgenic mice," *J. Virol.* **68**, 4358–4368 (1994).
- ²²M. S. Hughes, J. E. McCarthy, J. N. Marsh, J. M. Arbeit, R. G. Neumann, R. W. Fuhrhop, K. D. Wallace, T. Thomas, J. Smith, K. Agyem, D. R. Znidarsic, B. N. Maurizi, S. L. Baldwin, G. M. Lanza, and S. A. Wickline, "Application of Renyi entropy for ultrasonic molecular imaging," *J. Acoust. Soc. Am.* **125**, 3141–3145 (2009).
- ²³A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems* (Cambridge University Press, Cambridge, UK, 1995), pp. 1–804.