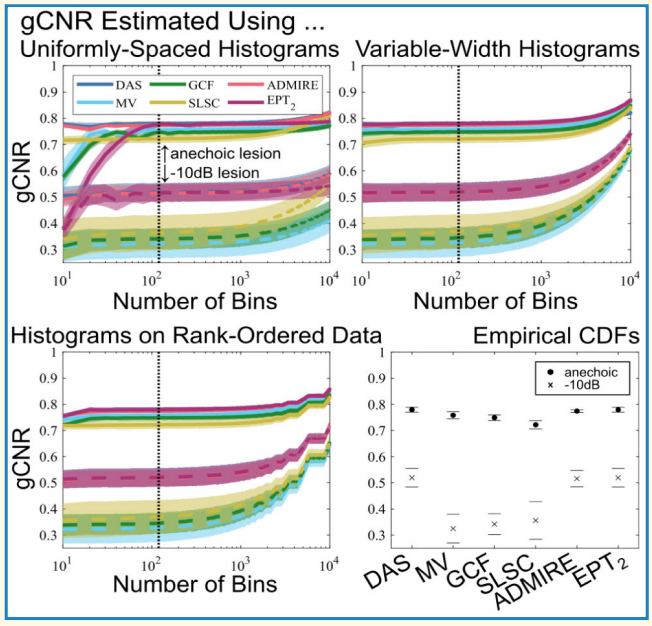
Methods for Enhancing the Robustness of the Generalized Contrast-to-Noise Ratio

Siegfried Schlunk, *Student Member, IEEE*, and Brett C. Byram, *Member, IEEE*

***Abstract*—The generalized contrast-to-noise ratio (gCNR) is a new but increasingly popular metric for measuring lesion detectability due to its use of probability distribution functions that increase robustness against transformations and dynamic range alterations. The value of these kinds of metrics has become increasingly important as it becomes clear that traditional metrics can be arbitrarily boosted with advanced beamforming or the right kinds of postprocessing. The gCNR works well for most cases; however, we will demonstrate that for some specific cases the implementation of gCNR using histograms requires careful consideration, as histograms can be poor estimates of probability density functions (PDFs) when designed improperly. This is demonstrated with simulated lesions by altering the amount of data and the number of bins used in the calculation, as well as by introducing some extreme transformations that are represented poorly by uniformly spaced histograms. In this work, the viability of a parametric gCNR implementation is tested, more robust methods for implementing histograms are considered, and a new method for estimating gCNR using empirical cumulative distribution functions (eCDFs) is shown. The most consistent methods found were to use histograms on rank-ordered data or histograms with variable bin widths, or to use eCDFs to estimate the gCNR.**



***Index Terms*— Histograms, image analysis, image quality, nonparametric statistics, probability density function (PDF), ranking, ultrasonic imaging.**

I. Introduction

BEAMFORMING methods in ultrasound imaging are most frequently judged either qualitatively by experts or quantitatively with image quality metrics. Historically, image quality metrics have been proposed as quantitative measurements that are predictive of clinical performance. Patterson [1] proposed the contrast-to-speckle ratio (CSR) as a metric that correlated well with subjective assessments of image quality. CSR is calculated as the ratio of lesion contrast to speckle noise or variance, which today we know as contrast-to-noise ratio (CNR). Smith et al. [2] proposed a similar metric, CNR_{ψ}, which can be related to the optimal signal-to-noise ratio (SNR) by incorporating the size of the lesion and an ultrasound system's spatial resolution. Using a contrast/detail phantom and choosing a threshold value for detection of each lesion, this SNR calculation was able to predict lesion detection. This suggests that CNR, which can be related to SNR, is a valuable image quality metric for beamformer comparison as an increase in CNR predicts an increase in lesion detectability.

Over time, the use of metrics such as contrast ratio and CNR has become the de facto evaluation standard, as it is faster and more convenient than having a sonographer or radiologist evaluate dozens of images whenever a beamformer is modified. As a result, adaptive beamformers are often designed with these metrics in mind. However, as these beamformers have become more common, so have the concerns over whether these beamformers are producing "real" improvements. In fact, it has specifically been shown that it is relatively easy to make a beamformer that artificially inflates CNR [3], [4]. In addition, calculating CNR on log-compressed versus linear data results in different values, making it less ideal as a go-to metric of image quality. Therefore, methods for comparing beamformers that can compensate for these factors are clear.

In investigating these issues, several new methods and metrics have been proposed by different groups. Rindal et al. [5], [6] proposed the dynamic range test (DRT), which seeks to measure the impact of dynamic range alterations on resolution,

Manuscript received 3 March 2023; accepted 16 June 2023. Date of publication 26 June 2023; date of current version 2 August 2023. This work was supported in part by NIH under Grant R01EB020040, Grant R01HL156034, and Grant S10OD023680-01. *(Corresponding author: Siegfried Schlunk.)*

The authors are with the Department of Biomedical Engineering, Vanderbilt University, Nashville, TN 37232 USA (e-mail: siegfried.g.schlunk@Vanderbilt.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/TUFFC.2023.3289157, provided by the authors.
Digital Object Identifier 10.1109/TUFFC.2023.3289157

Highlights

- **gCNR is a robust image quality metric, but this article shows that the use of histograms can result in poor estimates in some cases. Novel estimation methods are proposed, and robustness is evaluated.**
- **Novel methods such as using rank-ordered data or eCDFs are demonstrated to be more robust against unexpected distributions and require less manual effort to produce good gCNR estimates.**
- **This article hopes to be educational for researchers wishing to effectively implement or further investigate gCNR by highlighting potential risks and providing more robust estimation methods.**

contrast ratio, and CNR. Contrast ratio dynamic range (CRDR) was developed by Dei et al. [7], [8], [9] suggesting that measuring relative contrast accurately is important, rather than arbitrarily increasing contrast ratio everywhere. Histogram matching calculates the transformation between two methods using histograms, which allows a user to better discern true differences between methods [10]. And finally, and the topic of this work, generalized CNR (gCNR) is a lesion detectability metric designed to be robust against dynamic range alterations [3], [4]. All these methods were explicitly designed to allow for better comparisons between methods that may be affected by dynamic range alterations. However, as with all the image quality metrics, these new methods only measure particular facets of image quality; none can provide a complete clinical performance evaluation alone. For example, CNR and gCNR are a measure of lesion detectability, but they do not account for lesion size or spatial resolution. For a complete clinical analysis, all these factors would need to be considered.

The gCNR metric proposed by Rodriguez-Molares et al. [3], [4] estimates the probability of lesion detection by measuring the overlap between estimated probability density functions (PDFs) of a target and reference region of an image. In theory, the use of these PDFs makes gCNR resistant to dynamic range alterations. The standard implementation given is to compare the histograms of the target and reference regions and measure the overlap by the counts of the bins. Histograms are used since estimating the true distribution can be challenging for some beamformers. However, this does make the metric potentially reliant on how well the histograms represent the data. Hyun et al. [11] demonstrate this in their work, showing that gCNR when estimated with uniformly spaced histograms is reliant on the choice of the number of bins, and they suggest that even the spacing of the bins should be considered. This means that histogram design requires careful consideration on the part of the researcher to ensure gCNR is producing a good estimate.

Designing good histograms is practically a field unto itself, and methodology for optimizing and improving them has been around for nearly a century. If gCNR is implemented using histograms, these methods will need to be considered, and Hyun et al. [11] do suggest that using classic methods for choosing the number of bins or even using variable-width bins will help. However, these methods have not yet been evaluated in the context of gCNR, and an investigation into which methods will produce the most robust estimates of gCNR is required.

New in this work, we investigate beamformers and post-beamforming processors that can make histogram design difficult. We show that without careful effort, the gCNR metric can be "tricked" and even the relative ranking of some beamformers can be altered based on the choice of histogram binning. We implement and evaluate classic histogram design methods, including those suggested by Hyun et al. [11], as well as a novel way to estimate gCNR using rank-ordered data. Furthermore, we investigate a parametric implementation of fitting the data to known distributions. Finally, we propose a novel estimation method using empirical cumulative distribution functions (eCDFs) that does not rely on histograms at all. We compare these different methods across multiple simulations of varying amplitudes and sizes, and with different adaptive beamformers and post-beamforming processors.

II. Methods

MATLAB (MathWorks, Natick, MA, USA) was used for implementations of all the simulation methods, beamforming algorithms, and image quality metrics.

We used Field II [12], [13] to generate the prebeamformed channel data for all the simulations in this work. We simulated lesions of varying amplitudes (anechoic, -20, -10, and 0 dB), with six realizations for each. Lesions were simulated with different radii (1, 2, 2.5, 3, 4, and 5 mm) and were acquired with a 117-element, 0.257-mm pitch linear array transmitting at 3 MHz with a bandwidth of 60%, focused at the depth of the lesion (30 mm). Scatterers were randomly placed to achieve an average of 15 scatterers per resolution cell, which was 0.133 mm^2. In all, 128 beams were acquired spaced 0.234-mm apart, which satisfies the Nyquist theorem at the focus (expected lateral resolution is 0.517 mm). In addition, multiple-line acquisition (MLA), also referred to as parallel receive [14], was implemented to adjust the amount of data for analysis with gCNR. Except when noted, MLA was used to double ($\times 2$) the number of beams for all the cases prior to beamforming. At MLA $\times 2$, this resulted in a pixel size of 0.00226 mm^2 with approximately 59 pixels per resolution cell.

We additionally used MATLAB to generate pairs of Rayleigh distributions to measure the gCNR between them. For these distributions, $\sigma_1 = 0.5$ was constant in all the cases, and σ_2 was varied to achieve specific theoretical values of gCNR. Specifically, we chose $\sigma_2 = 0.89$, 1.27, and 2.16 to achieve approximate gCNR values of 0.4, 0.6, and 0.8, respectively, making it easy to compare the gCNR estimates to the "true" values. These sets were generated to measure the number of independent data points that might be required to

accurately estimate gCNR from a known distribution. For all the pairs, 12 realizations were generated.

A. Beamforming and Post-Beamforming Algorithms

As this work is focused on the analysis of gCNR, the discussion for the following beamformers is reserved for the original works, and we include here only specific implementation details.

1) Delay-and-Sum: Implemented with no special apodization (i.e., rectangular apodization).

2) Minimum Variance [15], [16]: Subarray lengths of $L = 0.5M$ and diagonal loading of $\epsilon = \Delta \cdot \mathrm{tr}(\hat{R})$, where $\Delta = 1/(10L)$, were applied. M is the total number of channels, and $\hat{R}$ is the estimate of the covariance matrix.

3) Generalized Coherence Factor [17]: The cutoff frequency index was set to $M_0 = 5$.

4) Short-Lag Spatial Coherence [18]: The number of lags used was 20, corresponding to approximately 17% of the active aperture.

5) Aperture Domain Model Image Reconstruction [19], [20], [21]: Parameters used were the same as in our previous work [9].

6) Envelope Power Transforms: The envelope power transform (EPT) is a post-beamforming gray-level transform (GLT) as described by Thijssen et al. [22] and used by Rindal et al. [6] in related work. We have chosen parameters specifically for beating implementations of gCNR that do not account for skewed data. We define it as a transform on the enveloped data, specifically from delay-and-sum (DAS), $|S_{\mathrm{DAS}}|$

$$\mathrm{EPT}_n(|S_{\mathrm{DAS}}|) = |S_{\mathrm{DAS}}|^n \tag{1}$$

where n is the desired power. For demonstration, we use $\mathrm{EPT}_{.5}$, EPT_2, and EPT_4 in this work, which correspond to taking the square root, the second power, and the fourth power of the enveloped data, respectively. This result is log-compressed using $20 * \log_{10}$ when displayed.

B. Contrast-to-Noise Ratio

CNR [1] is a measure of the difference in the average amplitudes of a target and background region (contrast), normalized by the speckle variance (noise). We define CNR as

$$\mathrm{CNR} = \frac{|\mu_{\mathrm{ROI}} - \mu_{\mathrm{background}}|}{\sqrt{\sigma^2_{\mathrm{ROI}} + \sigma^2_{\mathrm{background}}}} \tag{2}$$

where μ is the mean value, and σ is the standard deviation of the signal. Usually, the signal here is the enveloped, but uncompressed data; however, it is possible to calculate CNR from the log-compressed data as well. In this work, we will distinguish between the two. CNR will be the uncompressed version, and $\mathrm{CNR}_{\log}$ will be the log-compressed version.

C. Generalized CNR

An in-depth explanation of gCNR can be found from the original authors [3], [4], but we include here the important aspects of the method to understand how manipulation can occur.

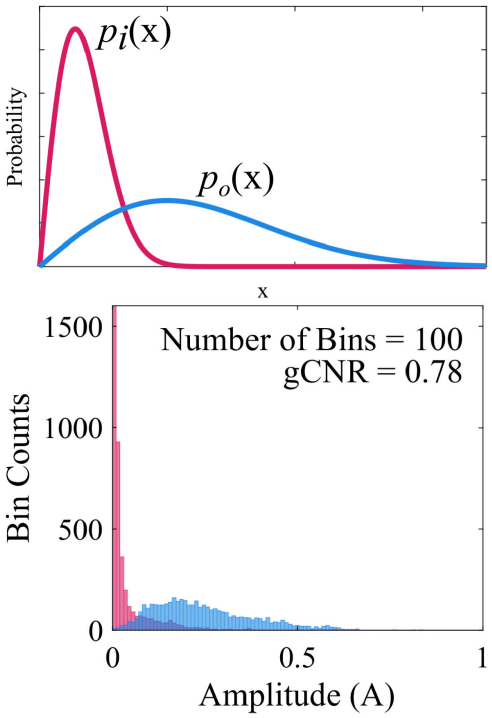


Fig. 1. (Top) Example of two probability density functions, $p_i(x)$ and $p_o(x)$, representing the lesion and reference area, respectively. (Bottom) Histograms of an anechoic lesion simulation to mimic the functions in the above plot, implemented with the 100 bins. The gCNR estimate is then based on the overlap in each bin.

gCNR is presented as a more robust version of CNR. The metric is calculated from the overlap of estimated PDFs of the target area (e.g., a lesion) and some reference area (e.g., adjacent tissue). Fig. 1 shows an example of two such functions, and solving for the overlap only requires calculating the intersection between the two curves and integrating over the respective areas of each. The actual gCNR metric is then calculated as

$$\mathrm{gCNR} = 1 - \mathrm{OVL} \tag{3}$$

where OVL is the area of overlap of the two curves. In practice, finding the actual PDF of either the lesion or background tissue can be difficult, so the original implementation estimates the PDFs using histograms, of which an example is shown in Fig. 1. For example, the original work uses $k = 100$ uniformly spaced bins, but as mentioned in the Introduction, Hyun et al. [11] later showed that careful consideration needs to be taken when choosing the number bins and potentially the spacing of the bins, suggesting popular choices such as $k = n^{1/2}$ or $k = n^{1/3}$. When using histograms, OVL can be calculated by normalizing the histograms of the target and the reference (such that the sum of each is 1), and then finding the smaller quantity between the target and the reference for each bin and summing all those together. This does rely on both the histograms having the same distribution of bins.

D. Potential Manipulations of gCNR

Histograms have been around since the seventeenth century [23], and it is thought that a properly built histogram is an

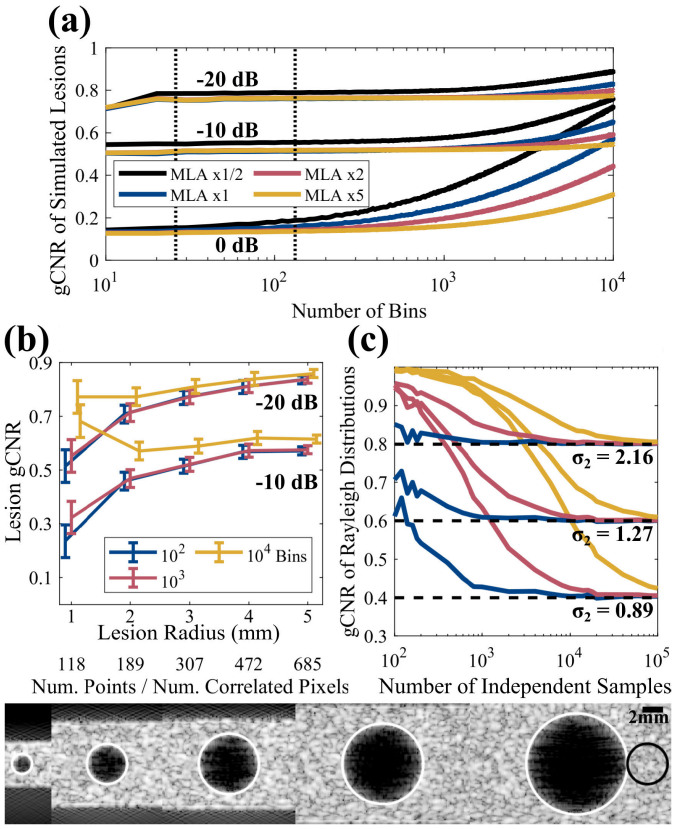


Fig. 2. (a) Plots of gCNR for a set of simulated lesions (−20, −10, 0 dB) beamformed with DAS, where the number of bins used in the histograms was varied, and additionally multiline acquisition was applied to increase or decrease the effective number of data points in the histograms. For MLA ×2, this resulted in $n = 17\,394$ total points between the target and reference regions of the image. This would result in $k = n^{1/2} = 132$ or $k = n^{1/3} = 26$ bins as often recommended (shown as the dotted lines). Note that the 0-dB case does not have a gCNR of 0, which may be the result of estimator variability or bias. (b) Plots of −20 and −10 dB simulated lesions of varying radii (1–5 mm), with gCNR estimated with 100, 1000, and 10 000 bins. Standard deviation between realizations is shown by the error bars. The lesions (bottom) show the areas used for the lesions (white) and background (black). The background was held constant to minimize its potential impact on the gCNR estimate. (c) Plots of gCNR where the target and reference groups are generated from Rayleigh distributions, where the number of independent samples generated in the distributions can be finely controlled. For all the three cases, $\sigma_1 = 0.5$, while σ_2 is varied as shown in the plot. The same binning options are included as in (b), and the dashed black line indicates the true gCNR calculated from the known PDFs of these Rayleigh distributions. Note that at the extremes, 1 bin results in 100% overlap and gCNR = 0, while an infinite number of bins results in 0% overlap and gCNR = 1.

estimator of the underlying PDF [24]. However, choosing the ideal binning of data for a histogram is an issue that has been debated for years, with Sturges [25] writing about the optimal number of bins back in 1926. Using too few bins generally results in oversmoothing of the data, while too many will likewise lead to undersmoothing [23]. Because of this, using a suboptimal number of bins can lead to misrepresentation of the data, obfuscating the PDF we are trying to estimate, and therefore interfering with our ability to accurately estimate the gCNR.

As a practical example of this, consider Fig. 2. Previously, Hyun et al. [11] observed that gCNR is dependent on the number of bins used, with the severity depending on the amplitude of the target lesion. We can extend this understanding to also consider the number of data points being considered. In Fig. 2(a), we show different amplitude lesions that have MLA applied at different scales to increase or decrease the number of data points, with gCNR estimated using uniformly spaced histograms at varying numbers of bins. We see that as we increase the number of data points with MLA for a fixed number of bins, the gCNR estimate converges on a consistent value. Alternatively, as the number of bins decreases, the sampling or the number of data points has less impact, suggesting a large ratio of data points to bins is required for a good estimate. Furthermore, as also observed by Hyun et al. [11], lower amplitude lesions are more impacted by these decisions. Obviously there is a limit to how few bins can be used (with the limit of 1 bin resulting in all the estimates being 0), but the general trend is clear. Do note, however, that MLA will only help make the histograms less reliant on the number of bins used, but it will not produce more independent data points, so it alone cannot help us to more accurately predict the true distribution of a dataset.

However, we can analyze the effect of the amount of data by simulating differently sized lesions or generating different amounts of data from known distributions. Fig. 2(b) shows a curve for lesions with radii ranging from 1 to 5 mm, while Fig. 2(c) shows when we generate a specific number of samples from known Rayleigh distributions and measure the gCNR between those sets. In these cases as well, it seems like smaller numbers of bins perform better, and larger numbers of bins cause the gCNR estimate to approach 1 (no overlap, due to every data point being in its own bin). We have also included an approximate ratio of the number of pixels for the different lesions compared with the number of pixels in a resolution cell (approximately the number of correlated pixels), to show how the two datasets might relate to each other. This suggests that even for the largest lesion here, it still falls far to the left side of the Rayleigh graph. At this stage, it is difficult to know what this means as far as an ideal minimum amount of data necessary to accurately assess gCNR.

An interesting observation from these graphs is the impact of beamforming effects on the gCNR estimates. The Rayleigh distributions, especially with the 100 bins case, suggest that the number of data points may not necessarily impact the estimate, assuming enough data that the distribution is actually represented (in fact, this will be confirmed later). However, in Fig. 2(b) with the simulated lesions we see a trend that smaller lesions have a lower gCNR, and as the size increases the estimate appears to converge. This is due to the impact of sidelobe content and the width of the main lobe introducing higher amplitude content into the target region. The shape of the main lobe produces a gradual slope of amplitude laterally between the background and the lesion. Depending on the mask we use to choose the target region, some of this content will naturally be included and will overlap partially with the distribution of the background region. The total amount of this sidelobe content will naturally vary directly with the size (radius) of the lesion, and the area of the lesion will vary directly with the square of the radius. This means that

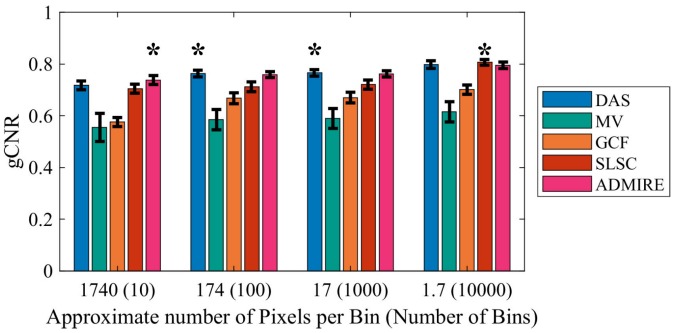


Fig. 3. Average gCNR (standard deviation shown as error bars) of −20 dB simulated lesions from the indicated beamforming methods. gCNR was calculated using different numbers of histogram bins, and the ratio of pixels to bins and the total number of bins are shown. The asterisk for each grouping indicates the highest gCNR value.

the ratio of the sidelobe content to total area will be higher with smaller cysts. This results in a larger overlap with the background region and a lower gCNR. This is also why as the lesions get larger the gCNR appears to converge, as the limit of this ratio goes to zero.

Further complicating matters, we show in Fig. 3 that varying the number of bins (or ratio of pixels/bins) used in the histograms can result in altering the rank-order of the evaluated beamformers. In this example, aperture domain model image reconstruction (ADMIRE) has the highest gCNR when estimating with 10 bins, while DAS is the best for 100 or 1000 bins. Short-lag spatial coherence (SLSC) ends up being the best if an extreme number of bins (10 000) is used. This makes it difficult for us to evaluate the relative performance of these beamformers. Note that the standard deviation shown on the error bars is from the different realizations simulated; for any single realization, we would not have error bars to judge how significant the differences are.

A second manipulation can arise even if the number of pixels and bins is fixed. Consider the EPT described before, where the enveloped data are taken to some chosen power. When the power is manipulated, the relative positioning of the data along the dynamic range can be altered. Fig. 4 shows DAS, $EPT_{.5}$, and EPT_2 with the associated histograms for a −20 dB lesion. Since all the three methods are simply transforms of each other, the gCNR should be equal in all the three cases; however, by manipulating the distribution of the data points along the data's range, the amount of overlap when using uniformly sized bins can change. This occurs due to the histograms being altered (even if the true overlap stays the same), and the binning scheme being unable to accurately represent the distributions. In the case shown here, as the skew of the data is intensified, the ability of the histogram to estimate and represent the PDF is diminished. The traditional calculation of CNR is included both on the linear and log-compressed data. Since the two transformations are based on powers, the powers cancel out such that the CNR_{log} is the same across all the three cases, while gCNR and CNR increase and decrease together. We can expand on this further in Fig. 5, where we see that normal binning rules produce poor uniformly spaced histograms and therefore poor gCNR

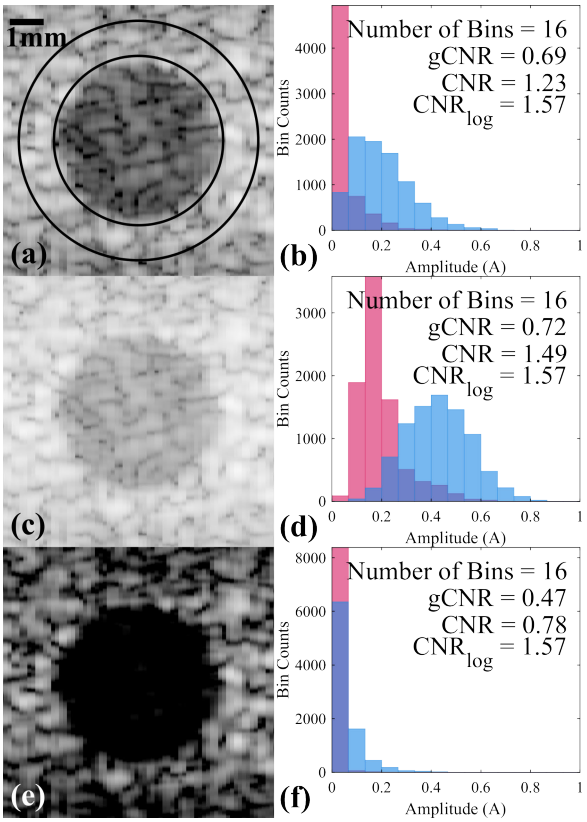


Fig. 4. Simulated −20 dB $r = 2.5$ mm lesions beamformed with (a) DAS, (c) $EPT_{.5}$, and (e) EPT_2. Corresponding histograms used to calculate gCNR are shown in (b), (d), and (f) using Sturges' formula to determine the number of bins of the histogram. In these cases where the skew of the data is high, the overlap estimate can be wrong. Traditional CNR is included for each case, both calculated on linear and log-compressed data.

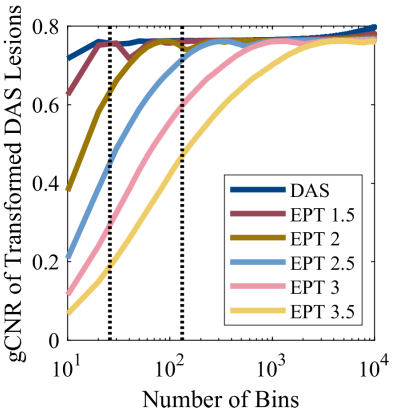


Fig. 5. Simulated −20 dB $r = 2.5$ mm lesions beamformed with DAS and transformed with the EPT with powers ranging from 1.5 to 3.5. In this case, accurate assessment of the gCNR is only consistent around $k \geq 500$, though even this is only true for powers up to 3.5 as used here. Compared with the classic binning rules, this is far more bins than might be expected purely based on the amount of data. For reference, these cases have a total number of data points $n = 17\,394$, which results in $k = n^{1/2} = 132$ or $k = n^{1/3} = 26$ bins as recommended (shown as the dotted lines), which results in different gCNR estimates for each case.

estimates for the different powers used in the EPT. Since the EPT is a GLT, the gCNR theoretically should be the same in all the cases, but the estimate does not converge until significantly larger numbers of bins are used, much higher than any binning

rule would propose. These cases show that it may not always be sufficient to simply choose an appropriate number of bins based on the amount of data.

All these observations make it clear that using uniformly spaced histograms is liable to cause issues unless careful and manual attention is paid to every case. Even then, the optimal histogram parameters may be different in separate cases, which would make it difficult to report as an author and difficult to evaluate as a reader. Instead, we will now consider different ways we can formulate histograms, such as with variable bin widths, as well as other techniques that may allow us to sidestep the issue of histograms altogether.

E. Increasing Robustness of the gCNR

To resolve the issues presented, careful consideration needs to be taken to make sure the histograms are accurately representing the data, ensuring a more robust estimation of the gCNR. We consider several possible options, detailed below.

1) Parameterized gCNR: Since estimation of gCNR relies on the calculation of the overlap between the PDFs of the lesion and reference material, a consideration can be made that estimating the distribution of the data by finding the actual PDF would be mathematically sound. Modeling different kinds of tissue in ultrasound for detection and identification purposes has been done [26], and the specific model used was often related to the properties of the backscattered echo, divided between pre-Rayleigh, Rayleigh, and post-Rayleigh (Rician) [27], [28]. Efforts to improve the modeling of ultrasound data led to the investigation of the K distribution [29], [30] and later the homodyned K distribution [31]. However, K and homodyned K distributions are significantly more computationally complex compared with previous distributions, making them relatively unappealing options. Instead, the Nakagami distribution [32] was suggested for use with ultrasound [26], [33], [34] as a much simpler alternative, while still incorporating features of both the Rayleigh and Rician models [26].

For this work, we consider the Rayleigh, Rician, and Nakagami distributions as robust, and computationally efficient options for modeling ultrasound echoes. All these distributions are relatively straightforward and require only estimation of one or two parameters. In addition, MATLAB has functions for fitting data using each of these choices, making implementation quick and efficient. Then, calculating the overlap can be done analytically or using built-in functions for the PDF and cumulative distribution function (CDF). For each distribution, we merely need to calculate the intersection using the PDF and then integrate under the relevant sections using the CDF.

2) Histogram-Based gCNR: We can also consider methods to make histogram-based gCNR estimations more robust to data range manipulations. As previously stated, the choice of the number of bins has been considered for nearly a century, and as a result there are a multitude of methods available today. Though there are too many to all be considered in this work, we will include common choices for consideration.

Sturges' formula was perhaps the first method published, and it suggests using $k = \lceil \log_2 n \rceil + 1$ [25], for n data points. Since then, the Rice rule $k = \lceil 2\sqrt[3]{n} \rceil$ has been suggested as an alternative [35], and the square root rule $k = \lceil \sqrt{n} \rceil$ is commonly used in statistics software packages [36]. Sturges' formula recommends very small numbers of bins, which would be optimal for the examples in Fig. 2, while the Rice rule produces somewhat less small estimates, while still being small enough to work well for those examples. The square root produces the largest estimates of these three, though it technically produces a better estimate for the EPT examples in Fig. 5. These rules were generally designed for certain kinds of distributions of data and to be used with bins of a fixed width, which means it is likely manipulation could still occur.

Instead, we consider variable bin widths as a method known to be more reliable with skewed data. In general statistics, variable bin widths are often avoided due to the narrow bin widths that occur at high-density locations of the data [37], and it is often specifically recommended not to use them without caution [23]. However, for the purposes of estimating the overlap for gCNR, it can adapt well to unexpected distributions and skews. If variable widths are applied, then data range manipulations should not occur, as the widths of the bins will adjust as the density of the data is altered. For the implementation presented here, we calculate the number of bins desired as either $k = \lceil \sqrt{n} \rceil$ or $k = \lceil 2n^{2/5} \rceil$ (suggested to be optimal for variable-width histograms) [38] as noted. Bin widths are designed such that each bin has an equal number of data points.

Another possible implementation we consider is to rank-order the data and estimate the gCNR from the list of rankings. This would entirely eliminate the possibility of manipulation via data range alteration, since the rank-order of the data would never change, and therefore the estimates of gCNR would be identical. Performing statistical analysis on rank-order data is hardly a new concept and can be used along side traditional metrics to gain more information about the behavior of the data [39]. In addition, rank-order data are used in many statistical tests such as those presented by Friedman [40] and Wilcoxon [41], [42]. We implement rank-ordering by sorting the entire set of data (both target and reference) and separating the ranks into the original target and reference groups. For example, consider two groups $[10, 40, 50]$ and $[20, 30, 60]$. If we rank all the data together and then separate the ranks back into their original groupings, we would get $[1, 4, 5]$ and $[2, 3, 6]$. By estimating the gCNR on these rankings rather than the original data, manipulations to the original data (such as squaring) are ignored, since the rankings will be unchanged, and thus the gCNR estimate will be unchanged. For our implementation, we use the square root rule to choose the number of bins, though the choice should be more flexible since there will be a consistent number of data points per bin, regardless of choice.

3) gCNR Using eCDFs: However, we can avoid both a parametric and histogram-based approach via the use of eCDFs. We include as supplementary material a proof showing that the gCNR can be estimated from the difference between the CDFs of the target and reference regions, without the use of any additional information. Specifically, and as visually explained in Fig. 6, gCNR can be estimated based on the amplitudes of the local maximums and minimums of the

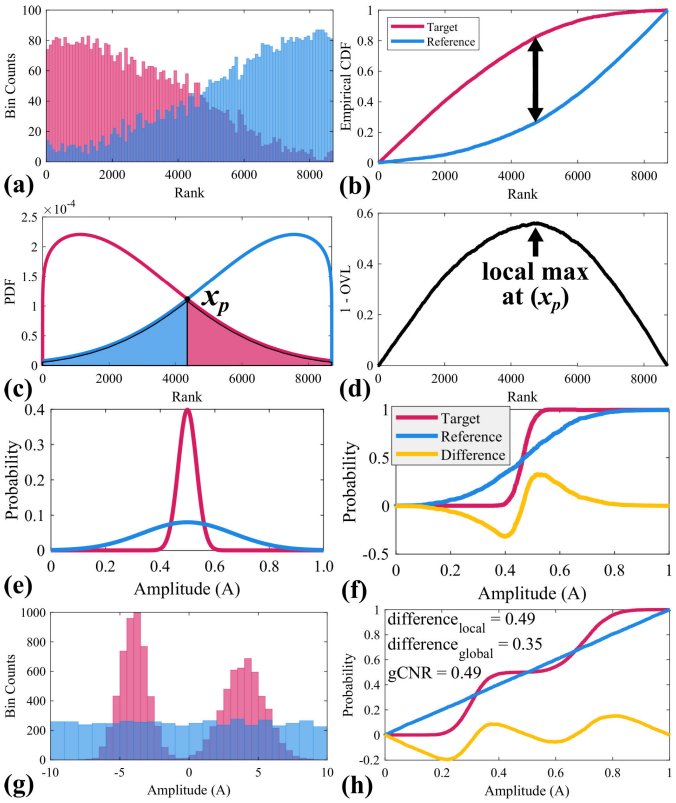


Fig. 6. Visual example of computing $\text{gCNR}_{\text{eCDF}}$ from eCDFs. (a) Histogram representation of the target and reference PDFs. (b) eCDF of the two regions. (c) Theoretical PDF of the two regions (not accurate, only for demonstration). (d) Difference function $\text{eCDF}_T - \text{eCDF}_R$. Local maximums and minimums occur at intersections of the PDFs, so the integration of the regions of the PDFs is related to the overlap, and therefore the gCNR can be calculated from the equation in (i). (e) and (f) shows an example of the process for two normal (Gaussian) distribution PDFs and the related eCDFs. We can see in this example that there are multiple intersections, which results in both local maximums and minimums. In most normal situations, there is only a single maximum and minimum, but (g) and (h) shows an example of a unique distribution which results in multiple local maximums and minimums, and calculating the difference only on the global extremes gives the incorrect gCNR.

difference function, $H = \text{eCDF}_T - \text{eCDF}_R$, and the difference between the eCDF of the target and reference regions. We prove in the supplementary material that with ideal functions (continuous, differentiable)

$$\text{gCNR}_{\text{eCDF}} = \sum \mathbb{H}_{\text{MAX}} - \sum \mathbb{H}_{\text{MIN}} \tag{4}$$

where $\mathbb{H}_{\text{MAX}}$ and $\mathbb{H}_{\text{MIN}}$ are the set of local maximums and minimums of H, respectively.

Our implementation with eCDFs uses MATLAB functions for ease of use. First, the ecdf function (©1993–2021 MathWorks, Inc.) calculates the eCDFs for both the target and reference regions. We then use some basic interpolation so that the x-axis data points are consistent across the two eCDFs. Then the H function is simply calculated from the difference in the two eCDFs, and the local maximums and minimums can be found. This then gives us the necessary information to compute the estimated gCNR as in the above equation. This is also summarized in Algorithm 1. It should be noted that

Algorithm 1 Estimating gCNR From eCDFs (MATLAB Implementation)

1 Given arrays *target* and *reference* containing the data we want to measure the overlap of
2 Calculate the empirical CDF of *target* (T) and *reference* (R)

$$[\text{eCDF}_T, x_T, \text{eCDF}_{T\_low}, \text{eCDF}_{T\_up}] = \text{ecdf}(target)$$
$$[\text{eCDF}_R, x_R, \text{eCDF}_{R\_low}, \text{eCDF}_{R\_up}] = \text{ecdf}(target)$$

which also gives the lower (low) and upper (up) confidence bounds of the estimate
3 Interpolate eCDF_T and eCDF_R to use the same x array for consistent sampling
4 Calculate the difference function, H

$$H = \text{eCDF}_{T\_\text{interp}} - \text{eCDF}_{R\_\text{interp}}$$

5 Find the set of local maximums and minimums of H, $\mathbb{H}_{\text{MAX}}$ and $\mathbb{H}_{\text{MIN}}$
6 Calculate gCNR as the difference in the sum of those sets

$$\text{gCNR}_{\text{eCDF}} = \sum \mathbb{H}_{\text{MAX}} - \sum \mathbb{H}_{\text{MIN}}$$

while theoretically the solution requires the local maximums and minimums, in practical ultrasound scenarios where lesions are being analyzed, the global maximum and minimum of the function are often sufficient. As in Fig. 6(g), technically we can formulate situations where the local values are required to estimate the correct overlap, but in this work none of our simulations or examples actually required more than the global values.

A note on implementation, the ecdf MATLAB function uses the nonparametric Kaplan–Meier estimator [43], which is used to calculate the survival function from a set of data, which in turn is the complement of the CDF. In addition, upper and lower confidence bounds can be calculated using Greenwood's formula, which is all considered in the MATLAB implementation based on prior work [44], [45]. A potential advantage of using this method is that the upper and lower confidence bounds give us a better ability to discern whether the estimated gCNR of two images or beamformers is sufficiently different to make a conclusion about ranking of the methods. By default, and as used here, the confidence is set to $\alpha = 0.05$ for a 95% confidence interval on our gCNR estimate.

III. Results

A. Parametric gCNR Results

Some examples of fitting the lesion data to known distributions for a −10 dB lesion are included in Fig. 7. For the DAS data, all the models seem relatively similar, matching the shape of the data in the histogram. However, for minimum variance (MV) and SLSC, some distributions are poor fits. The Rayleigh and Rician distributions seem poor fits for MV, while

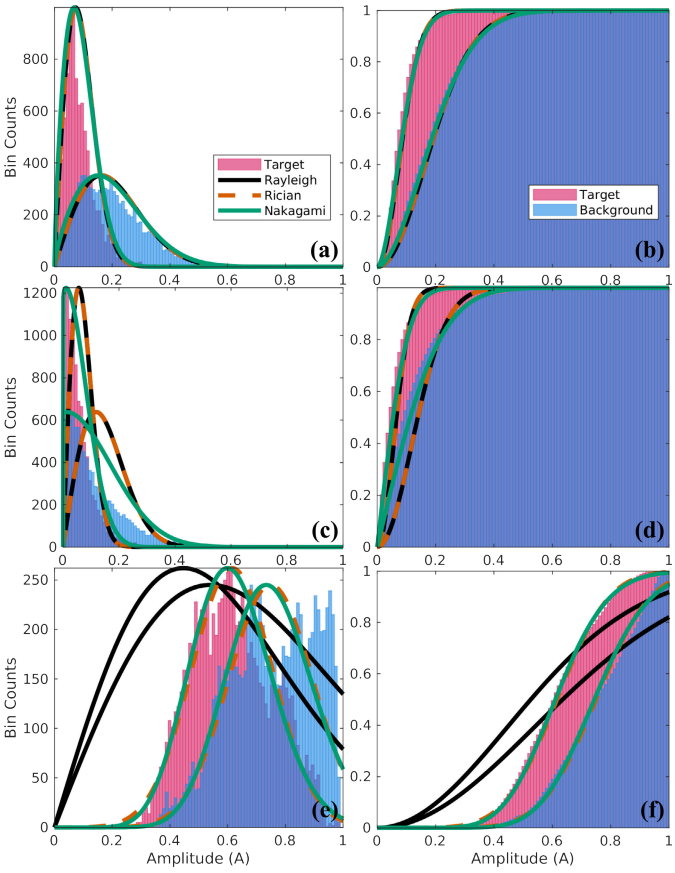


Fig. 7. Simulated −10 dB $r = 2.5$ mm lesions beamformed for (a) and (b) DAS, (c) and (d) MV, and (e) and (f) SLSC. The target and background data are shown as both a probability distribution function and cumulative distribution function (CDF), with the indicated distribution fits included.

the Rayleigh distribution does not fit SLSC at all. Generally, it seems clear that fitting the MV and SLSC data is much more challenging compared with DAS, and trying to fit the EPT data was functionally impossible (not shown) due to the amount of skew present in the data.

B. Nonparametric Results

A comparison of anechoic (high contrast) and −10 dB (low contrast) lesions is shown in Fig. 8, showing how the gCNR estimates vary by the number of bins used for each of the histogram-based methods. Uniformly spaced histograms show significant variability at extremely high and low numbers of bins with the ranking of the beamformers changing at various points. This makes it difficult to determine an optimal binning preference as the estimates change dramatically. In comparison, variable bin widths and rank-ordering show some dependence on the number of bins, but the curves are much more consistent, and importantly, the ranking of the methods appears consistent regardless of choice. By far the most consistent is the $\text{gCNR}_{\text{eCDF}}$ method. These estimates align with the variable-width and rank-ordering graphs near the $k = \sqrt{n}$ binning choice. Fig. 9 shows visual examples for one anechoic lesion case and includes the histograms or plots for the uniformly spaced, variable bin width, rank-order, and $\text{gCNR}_{\text{eCDF}}$ methods we investigated.

Fig. 10 shows the cases where we tested the amount of data available, both with differently sized lesions and by generating data from Rayleigh distributions. As discussed in Section II-D, there is a trend of gCNR increasing and converging as the lesion size increases, which is present in both uniformly spaced histograms and our robust methods. With the Rayleigh distributions, gCNR is highly variable until enough independent data points have been generated such that the two PDFs can be accurately estimated. This suggests, as we theorized, that gCNR is impacted by how much sidelobe or main lobe content is present in the lesion area and how much total independent data are available, which depends on lesion size. Then the convergence of gCNR as the lesion size increases is likely driven by both the decreasing ratio of the sidelobe content and the increasing amount of independent data in the region. Beamformers that sharpen the main lobe, like MV, may be able to improve gCNR for these smaller lesions, though in the example in Fig. 9, MV also increased the variance in the background speckle which overall resulted in more overlap between the two regions. The Rayleigh data also suggest that in some cases with ultrasound data we may not have enough independent samples to get a "true" gCNR estimate. Though these are limitations from the perspective of using gCNR to find the overlap between the "true" distributions of the target and background regions, it does not prevent the use of gCNR to generally compare the quality difference between two beamformers where the same regions are being compared.

Fig. 11 shows the post-beamforming EPT examples from Fig. 5, but now with the robust methods included. Again, we see that the variable-width, rank-ordered, and eCDF methods are consistent and predict the correct gCNR at nearly all the binning options. In this case, since the EPT data are transforms from DAS, all the curves should be identical, which mostly are for those three cases. However, uniformly spaced histograms simply cannot correctly estimate the gCNR in these cases with a reasonable choice of the number of bins.

IV. Discussion

To start, we should make it clear that in the majority of simple, homogeneous lesion cases, the uniformly spaced histogram implementation of gCNR using $k = \sqrt{n}$ bins and sufficiently large ROIs works perfectly well. After all, the only cases where it seriously struggled are the EPT lesion cases in Figs. 5 and 11, and the EPT_4 case in Fig. 9, all of which create exceptionally skewed data where we would expect a uniformly spaced histogram to struggle. What these cases show is that even a normally robust rule such as $k = \sqrt{n}$ may be insufficient if the "effective" number of data points is altered due to the skewness of the data. This makes it tricky to recommend a single blanket rule for uniformly spaced histograms.

It is easy to fall into the trap of getting lazy and picking a nice round number of bins without verifying the resulting histograms. More importantly, many researchers may be unfamiliar with gCNR and histograms such that one is unaware that the choice of the number of bins is such a potentially critical decision for some complex cases. We have demonstrated that

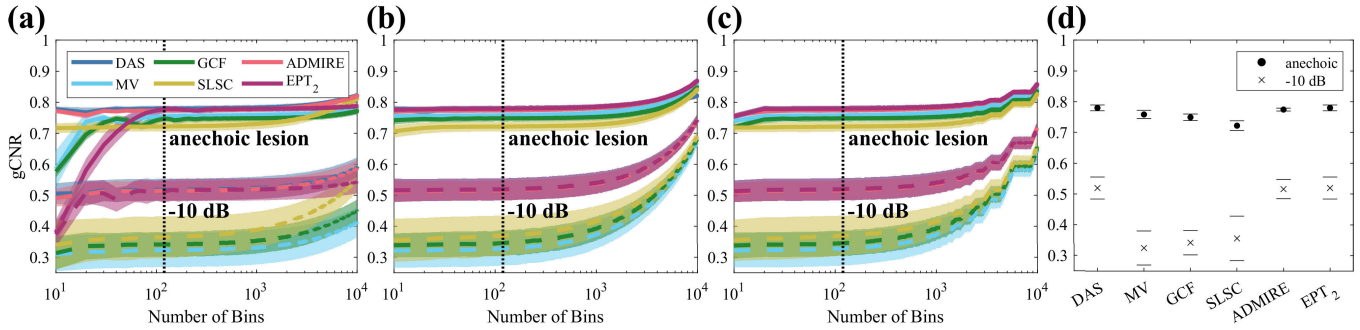


Fig. 8. gCNR for anechoic (solid) and −10 dB (dashed) $r = 2.5$ mm lesions with various beamformers estimated with (a) traditional uniform bin widths, (b) variable bin widths, and (c) rank-order uniform bin widths. The shaded regions show the standard deviation for each method. For these cases, the number of pixels used is constant ($n = 17\,394$ total for MLA $\times 2$) and the number of bins is varied. For reference, $k = \sqrt{n} = 132$ (the dotted line shown) and $k = \lceil 2n^{2/5} \rceil = 100$. (d) Plot of the average $\text{gCNR}_{\text{eCDF}}$ and standard deviation using eCDFs for that same data. Note that in (a)–(c) the DAS, ADMIRE, and EPT_2 curves overlap significantly.

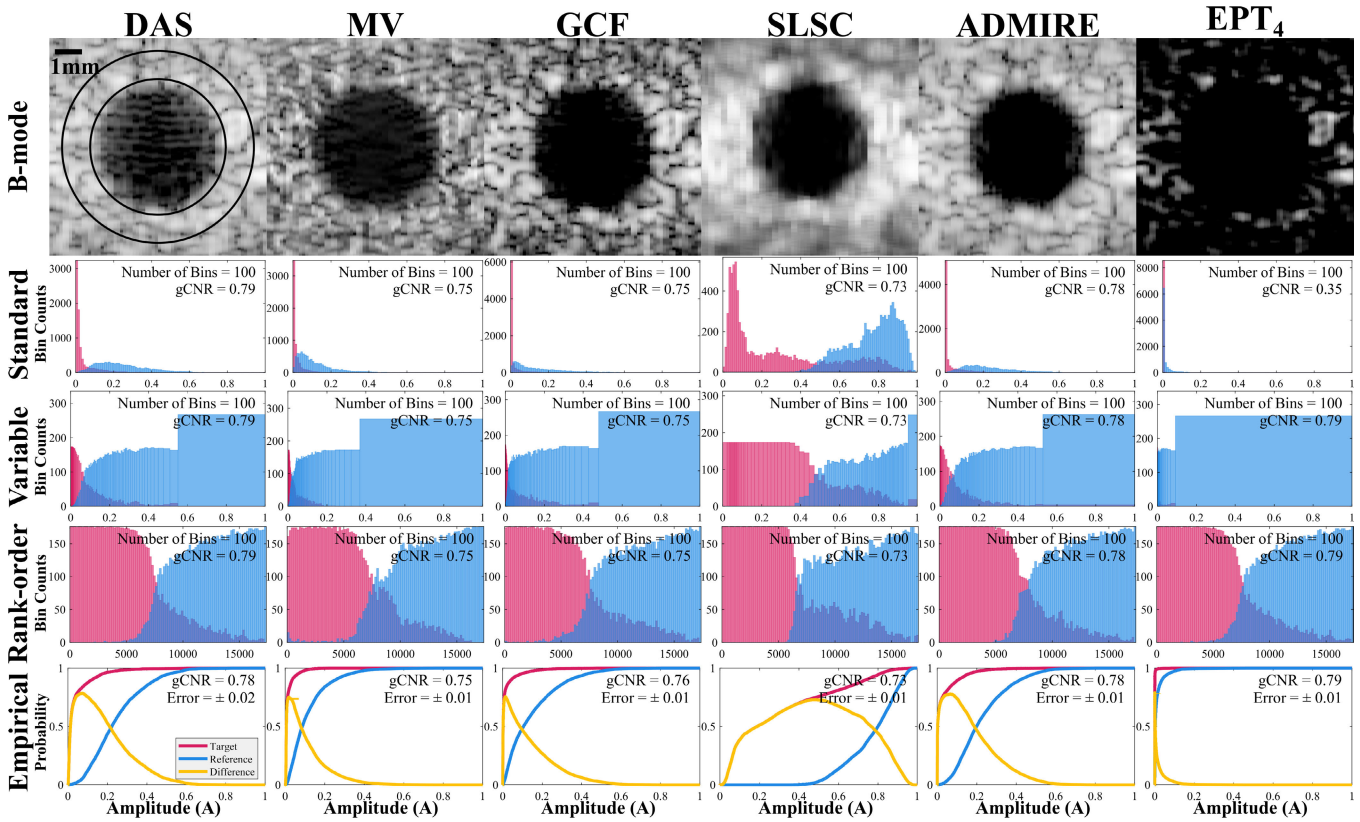


Fig. 9. B-mode images for a selection of beamforming methods for an example simulated $r = 2.5$ mm anechoic lesion. The circled regions on the DAS image indicate the lesion and background regions for gCNR calculation. Log-compressed methods are displayed with a 50-dB dynamic range. Histograms used for estimating gCNR used $k = \lceil 2n^{2/5} \rceil = 100$ for all the implementations for comparison. We additionally include the eCDF plots for these cases, with the error representing the average distance to the upper and lower bounds of the $\text{gCNR}_{\text{eCDF}}$ estimate.

there exists a possibility for methods to, willfully or not, manipulate gCNR estimates by doing something as innocent as using too many bins, which we have anecdotally witnessed. As a result, the discussion of how to increase the robustness or at least raise awareness of the implementation of gCNR is warranted.

There is a certain attractiveness to being able to model the ultrasound scatterers using parametric methods, and it is used in several applications to characterize images. Hverven et al. [46] showed us in previous work that many common beamformers can be fit with Rayleigh distributions, yet others deviate significantly. This presents the issue of determining what models are appropriate fits, and this issue is compounded by the common presence of adaptive beamformers. While DAS (raw B-mode data) has a well-understood distribution which makes fitting the data often reliable, the same is not true for many other adaptive methods. SLSC, or post-beamforming compression processes such as the EPT method, may have distributions that vary widely or fail to be fit at all in some cases. Even if a beamformer has a known distribution, a potential pitfall of a parametric method is that for a small enough sample of data the expected distribution may not be

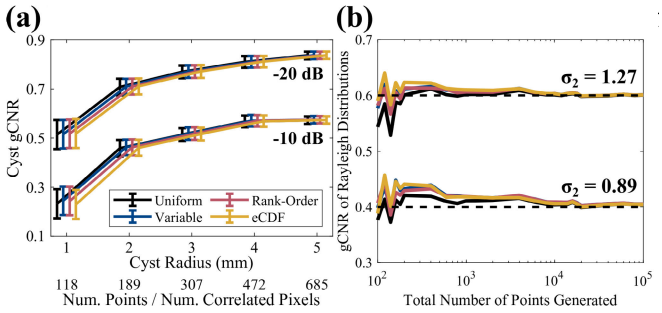


Fig. 10. Comparison of robust implementations on (a) varied lesion sizes at −20 and −10 dB and (b) Rayleigh distributions from Fig. 2. For the histogram-based methods, $k = \sqrt{n}$ bins for all the cases.

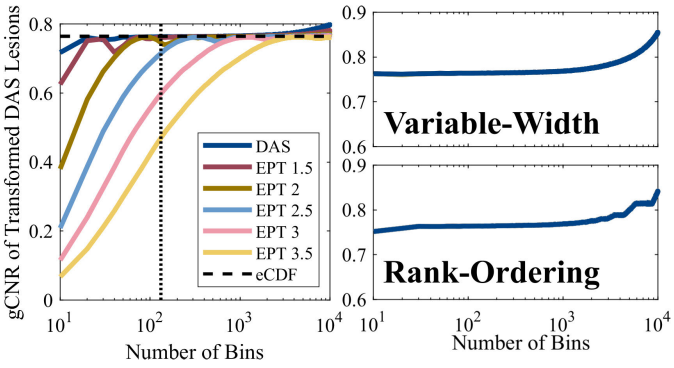


Fig. 11. Plot comparing the gCNR of (left) uniformly spaced histograms and (right) variable-width and rank-ordered histograms of the EPT data from Fig. 5. The dotted line indicates $k = \sqrt{n} = 132$, which in particular fails for the uniformly spaced histograms. The $\text{gCNR}_{\text{eCDF}}$ estimate shown on the left graph and the variable-width and rank-ordered examples show the correct estimation of gCNR across a wide range of the number of bins.

fit well anyway. Speckle is naturally correlated to some extent depending on the parameters of the acquisition, and potentially even the beamformer used, so we may need more data than expected if we pursue a parametric, or even histogram-based, approach, which may not always be possible depending on the target. This is something that would need to be considered if further investigation was desired.

For ease of use, a histogram-based implementation may still be best as it is straightforward and requires less consideration compared with parametric methods. In this case, it seems reasonable that a best-practices choice would be, at a minimum, to have some specific method for choosing the number of bins. By declaring that method and reporting the number of bins used, this gives the reader some sense of the ratio of the number of data points versus the number of bins used in the analysis. This is important, since we have shown that the number of bins and, by extension, the size of the dataset can alter not just the gCNR estimates for each method but potentially the ranking of the methods being compared. We have personally found $k = \sqrt{n}$ to be satisfactory in most cases, though this is not always sufficient.

Perhaps the most reliable histogram-based method is to use variable bin widths or rank-ordering, as it removes much of the uncertainty around choosing the number of bins and compensates for heavily skewed data. Fig. 9 shows examples for each of these methods, demonstrating how the variable widths change dramatically depending on the beamformer in question, while the rank-order histograms are identical for methods that are true transforms of each other. With histograms, these two implementations were the only two tested that were able to correctly match the results for DAS and EPT methods to within an error of 0.01 in all the circumstances. Both these implementations would seem to make gCNR estimations via histograms more resistant to extreme dynamic range alterations. It also does not prevent or restrict the use of gCNR, or increase the complexity of the interpretation of the metric, which falls in line with the goals laid out by the authors in the original work [4].

However, as we have shown, we can produce estimates of gCNR without the use of parametric methods or histograms using the eCDF. In the same Fig. 9, the empirical probability plots show that not only is the estimate relatively intuitive to calculate but also the upper and lower bounds produce a result with a small overall error. In addition, it works consistently with all the beamformers to produce nearly identical results to the robust histogram-based implementations. In relatively simple cases, like the simulated lesions here, the gCNR estimate is trivial to calculate from these plots, though even in more complex cases the full formulation we presented in (4) is sufficient. Finally, we could even apply this empirical method on the rank-ordered data to produce the rank-ordered version of this metric.

Overall, Fig. 8 suggests that using a robust rule to choose the number of bins, such as here $k = \sqrt{n}$, goes a long way to producing a robust gCNR estimate, though in cases with highly skewed data, such as in Figs. 5 and 11, this may be insufficient without further verification. However, the use of variable bin widths or rank-ordering your data removes much of the concern around heavily skewed data. Finally, we have also presented an entirely new implementation that uses eCDFs to estimate gCNR that is highly accurate in the cases shown here and eliminates much of the discussion around whether a histogram is well-formed. It is robust against skewed data and situations where there is a small amount of data available. As these robust methods are invariant to extreme dynamic range alterations, it also makes application straightforward, as it can be applied to the real data at any point in the post-beamforming process and is not even restricted to ultrasound data, just as gCNR was originally envisioned by Rodriguez-Molares et al. [3], [4].

There are outstanding questions and concerns related to gCNR that may need to be addressed in the future. First, as we saw with the differently sized lesions, there are trends that may be more or less apparent depending on how we choose the masks for our regions. Obviously, this is not an issue solely for gCNR, but rather for every metric that requires manual selection of a target and reference region. Second, Fig. 10 clearly shows that there is some minimum amount of independent data required for an accurate estimation of a PDF, and for some very small targets we may not have enough independent data to get an accurate gCNR estimate. This confirms to us that an investigation into this threshold is warranted in the future, as was also suggested by Rodriguez-

Molares et al. [4] in their work. Third, it is understood that detectability depends on lesion size and spatial resolution as Smith et al. [2] showed with their metric, CNR_ψ. gCNR only incorporates contrast and variance, similar to CNR, and as such for a complete analysis of the performance of two beamformers it may be insufficient alone. Finally, there is still a question as to how different two measurements should be to consider the performance of two image formation methods to be different. With histograms, it is reasonable to expect that there is some inherent variability in the estimation (as with any estimator), such that, for example, a gCNR of 0.77 is probably not significantly different than 0.78. eCDFs do produce upper and lower bounds (here, 95% confidence intervals), which is potentially valuable exactly for this purpose, and in Fig. 9, we can probably expect that DAS and ADMIRE are of equivalent performance, while MV and generalized coherence factor (GCF) are similar, but lower compared with the other two. This may be a valuable aspect to consider for future work.

V. Conclusion

gCNR is a powerful tool, resistant to transformations and dynamic range alterations when properly implemented. While gCNR implementations using histograms with uniformly spaced bins work properly in many circumstances, we have shown that manipulation can occur by altering the ratio of the pixels per bin or with some extreme transformation-based processes. Though a parametric approach is attractive, the inability to model an arbitrary adaptive beamformer makes it all but impossible without significantly more complex models, making a histogram-based or nonparametric solution much more appealing. In testing, making the histograms more robust is straightforward, and at a minimum a rule to determine the number of bins used, such as $k = \sqrt{n}$, should be reported. We can further improve the robustness of the histograms by either using variable bin widths (where the number of data points per bin is approximately equal) or rank-ordering the data and estimating gCNR from the lists of ranks. Alternatively, we can use eCDFs to produce an estimate that relies on neither parametric methods nor histograms, while still being robust in all the cases we considered here. All the three methods are much more resistant to transformations, and $\text{gCNR}_{\text{eCDF}}$ in particular requires no special parameter consideration, making it arguably the most "plug-and-play" of the three methods.

Acknowledgment

The authors would like to thank the staff of Advanced Computing Center for Research and Education (ACCRE), Vanderbilt University, Nashville, TN, USA, computing resource and the Vanderbilt Institute for Surgery and Engineering (VISE), Nashville, Physician-in-Residence Program.

References

[1] M. Patterson, "The improvement and quantitative assessment of B-mode images produced by an annular array/cone hybrid," *Ultrason. Imag.*, vol. 5, no. 3, pp. 195–213, Jul. 1983. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0161734683900019

[2] S. W. Smith, R. F. Wagner, J. M. Sandrik, and H. Lopez, "Low contrast detectability and contrast/detail analysis in medical ultrasound," *IEEE Trans. Sonics Ultrason.*, vol. SU-30, no. 3, pp. 164–173, May 1983.

[3] A. Rodriguez-Molares, O. M. Hoel Rindal, J. D'hooge, S. Måsøy, A. Austeng, and H. Torp, "The generalized contrast-to-noise ratio," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2018, pp. 1–4.

[4] A. Rodriguez-Molares et al., "The generalized contrast-to-noise ratio: A formal definition for lesion detectability," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 4, pp. 745–759, Apr. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8918059/

[5] O. M. H. Rindal, A. Austeng, H. Torp, S. Holm, and A. Rodriguez-Molares, "The dynamic range of adaptive beamformers," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2016, pp. 1–4.

[6] O. M. H. Rindal, A. Austeng, A. Fatemi, and A. Rodriguez-Molares, "The effect of dynamic range alterations in the estimation of contrast," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 7, pp. 1198–1208, Jul. 2019.

[7] K. Dei, A. Luchies, and B. Byram, "Contrast ratio dynamic range: A new beamformer performance metric," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

[8] S. Schlunk, K. Dei, and B. Byram, "Iterative ADMIRE for high dynamic range B-mode," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2018, pp. 1–4.

[9] S. Schlunk, K. Dei, and B. Byram, "Iterative model-based beamforming for high dynamic range applications," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 68, no. 3, pp. 482–493, Mar. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9149929/

[10] N. Bottenus, B. C. Byram, and D. Hyun, "Histogram matching for visual ultrasound image comparison," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 68, no. 5, pp. 1487–1495, May 2021.

[11] D. Hyun, G. B. Kim, N. Bottenus, and J. J. Dahl, "Ultrasound lesion detectability as a distance between probability measures," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 69, no. 2, pp. 732–743, Feb. 2022.

[12] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 39, no. 2, pp. 262–267, Mar. 1992.

[13] J. A. Jensen, "FIELD: A program for simulating ultrasound systems," in *Proc. IEEE 10th Nordic-Baltic Conf. Biomed. Imag.*, vol. 34, Mar. 1996, pp. 351–353.

[14] D. P. Shattuck, M. D. Weinshenker, S. W. Smith, and O. T. von Ramm, "Explososcan: A parallel processing technique for high speed ultrasound imaging with linear phased arrays," *J. Acoust. Soc. Amer.*, vol. 75, no. 4, pp. 1273–1282, Apr. 1984.

[15] I. K. Holfort, F. Gran, and J. A. Jensen, "Broadband minimum variance beamforming for ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, no. 2, pp. 314–325, Feb. 2009.

[16] J. F. Synnevag, A. Austeng, and S. Holm, "Adaptive beamforming applied to medical ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 54, no. 8, pp. 1606–1613, Aug. 2007.

[17] P.-C. Li and M.-L. Li, "Adaptive imaging using the generalized coherence factor," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 50, no. 2, pp. 128–141, Feb. 2003.

[18] M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl, "Short-lag spatial coherence of backscattered echoes: Imaging characteristics," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 58, no. 7, pp. 1377–1388, Jul. 2011.

[19] B. Byram, K. Dei, J. Tierney, and D. Dumont, "A model and regularization scheme for ultrasonic beamforming clutter reduction," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 62, no. 11, pp. 1913–1927, Nov. 2015.

[20] K. Dei and B. Byram, "The impact of model-based clutter suppression on cluttered, aberrated wavefronts," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 10, pp. 1450–1464, Oct. 2017.

[21] B. Byram and M. Jakovljevic, "Ultrasonic multipath and beamforming clutter reduction: A chirp model approach," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 61, no. 3, pp. 428–440, Mar. 2014.

[22] J. M. Thijssen, B. J. Oosterveld, and R. F. Wagner, "Gray level transforms and lesion detectability in echographic images," *Ultrason. Imag.*, vol. 10, no. 3, pp. 171–195, Jul. 1988, doi: 10.1177/016173468801000302.

[23] D. W. Scott, "Histogram," *WIREs Comput. Statist.*, vol. 2, no. 1, pp. 44–48, Jan. 2010.

[24] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–610, 1979.

[25] H. A. Sturges, "The choice of a class interval," *J. Amer. Stat. Assoc.*, vol. 21, no. 153, pp. 65–66, Mar. 1926.

[26] P. M. Shankar, "A general statistical model for ultrasonic backscattering from tissues," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 47, no. 3, pp. 727–736, May 2000.
[27] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, "Statistics of speckle in ultrasound B-scans," *IEEE Trans. Sonics Ultrason.*, vol. SU-30, no. 3, pp. 156–163, May 1983.
[28] R. F. Wagner, M. F. Insana, and D. G. Brown, "Statistical properties of radio-frequency and envelope-detected signals with applications to medical ultrasound," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 4, no. 5, p. 910, 1987.
[29] P. M. Shankar, "A model for ultrasonic scattering from tissues based on the k distribution," *Phys. Med. Biol.*, vol. 40, no. 10, pp. 1633–1649, Oct. 1995.
[30] R. C. Molthen, P. M. Shankar, and J. M. Reid, "Characterization of ultrasonic B-scans using non-Rayleigh statistics," *Ultrasound Med. Biol.*, vol. 21, no. 2, pp. 161–170, Jan. 1995.
[31] V. Dutt, "Ultrasound echo envelope analysis using a homodyned k distribution signal model," *Ultrason. Imag.*, vol. 16, no. 4, pp. 265–287, Oct. 1994.
[32] M. Nakagami, "The m-distribution—A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*. Amsterdam, The Netherlands: Elsevier, 1960, pp. 3–36.
[33] L. Clifford, P. Fitzgerald, and D. James, "Non-Rayleigh first-order statistics of ultrasonic backscatter from normal myocardium," *Ultrasound Med. Biol.*, vol. 19, no. 6, pp. 487–495, Jan. 1993.
[34] L. C. Gilman, "First-order statistics of pulsed-sinusoid backscatter from random media: Basic elements of an exact treatment," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 44, no. 4, pp. 798–804, Jul. 1997.
[35] D. M. Lane, "Introduction to statistics," in *Online Statistics Education: A Multimedia Course of Study*. Houston, TX, USA: Rice Univ. Press, 2007. [Online]. Available: http://onlinestatbook.com/
[36] A. C. Cameron. (2009). *EXCEL 2007: Histogram*. [Online]. Available: http://cameron.econ.ucdavis.edu/excel/excel.html
[37] D. W. Scott, *Multivariate Density Estimation* (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, Aug. 1992.
[38] *Chi-Square Goodness-of-Fit Test. NIST/SEMATECH e-Handbook of Statistical Methods*. [Online]. Available: https://itl.nist.gov/div898/handbook/prc/section2/prc211.htm. Accessed: Aug. 2022, doi: 10.18434/M32189.
[39] W. J. Conover, "The rank transformation—An easy and intuitive way to connect many nonparametric methods to their parametric counterparts for seamless teaching introductory statistics courses," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 4, no. 5, pp. 432–438, Sep. 2012.
[40] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Stat. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937.
[41] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, p. 80, Dec. 1945.
[42] F. Wilcoxon, "Some rapid approximate statistical procedures," *Ann. New York Acad. Sci.*, vol. 52, no. 6, pp. 808–814, Mar. 1950.
[43] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *J. Amer. Stat. Assoc.*, vol. 53, no. 282, pp. 457–481, Jun. 1958. [Online]. Available: https://www.jstor.org/stable/2281868?origin=crossref
[44] D. Cox and D. Oakes, *Analysis Survival Data*. London, U.K.: Chapman & Hall, Feb. 1984.
[45] J. F. Lawless, *Stat. Models Methods for Lifetime Data* (Wiley Series in Probability and Statistics). Hoboken, NJ, USA: Wiley, Nov. 2002, doi: 10.1002/9781118033005.
[46] S. M. Hverven, O. M. H. Rindal, A. Rodriguez-Molares, and A. Austeng, "The influence of speckle statistics on contrast metrics in ultrasound imaging," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

Siegfried Schlunk (Student Member, IEEE) received the B.E. degree in biomedical engineering and mathematics and the Ph.D. degree in biomedical engineering from Vanderbilt University, Nashville, TN, USA, in 2016 and 2023, respectively.

He is currently continuing research efforts at the BEAM Laboratory, Vanderbilt University, as a Postdoctoral Researcher. His research interests include developing methods for improving ultrasound image quality in cardiac and kidney applications.

Brett C. Byram (Member, IEEE) received the B.S.E. degree in biomedical engineering and mathematics from Vanderbilt University, Nashville, TN, USA, in 2004, and the Ph.D. degree in biomedical engineering from Duke University, Durham, NC, USA, in 2011.

He is currently an Associate Professor with the School of Engineering, Vanderbilt University and a Hoy Family Faculty Fellow. He has spent time at Siemens' Ultrasound Business Unit, Mountain View, CA, USA, and he spent a year with the Jorgen Jensen's Center for Fast Ultrasound, Lyngby, Denmark. He served as an Assistant Research Professor with Duke University from 2012 to 2013. He is also a part of the Vanderbilt Institute for Surgery and Engineering (VISE), Nashville, and an Affiliate with the Vanderbilt University Institute of Imaging Science (VUIIS), Nashville.