Blind Measurement of Receiver System Noise

Daniel G. Kuester, Adam Wunderlich, Duncan A. McGillivray, Dazhen Gu, and Audrey K. Puls

***Abstract*—Tightly-packaged receivers pose a challenge for noise measurements. Their only outputs are often diagnostic or benchmark information — "user data" that result from unknown processing. These include data rate test results, signal-to-noise ratio estimated by the receiver, etc. Some of these are important gauges of communication viability that may be enshrined in performance and conformance specifications. Engineers can estimate these parameters based on standards and simplified system models, but there are few means to validate against physical measurements.**

We propose here a set of measurement techniques to complement and support models of system noise. The approach is founded on a semi-parametric model of the noise response of a full-stack receiver. We probe this response experimentally by systematically perturbing signal and excess noise levels at the receiver input. The resulting technique is blind to protocol and implementation details. We introduce the design and implementation of some novel test capabilities required for these tests: a precision programmable excess noise source, and a highly directive programmable attenuator. We also introduce a regression procedure to estimate system noise (or noise figure) from the controlled input conditions and summary statistics of the user data output. We also estimate uncertainty in the measurement by combining traditional methods with a Monte Carlo method that propagates random errors through the regression.

Case studies demonstrate the measurement with consumer wireless networking and geolocation equipment. These include verification by repeatability testing and cross-comparison against Y-factor measurements.

I. INTRODUCTION

DEMAND for wireless spectrum has led to an increasing number of shared and tightly-adjacent spectrum allocations. New transmissions in these contexts, in aggregate, increase the risk of undesired noise in receivers. Yet measuring this impact has become particularly difficult, because wireless manufacturers have tended to sacrifice receive test ports to reduce size, weight, and cost. Thus, while system noise is now especially important, insights based in physical measurements are often out of reach.

The simplified receive system of Fig. 1 illustrates this conundrum in the context of a benchtop noise measurement. First, there is no analog or digitized waveform output here, because the system has bridged the radio frequency (RF) input to some other processed information domain. Further, tightly-integrated packaging prevents connections to intermediate outputs. Packaged consumer products may also require bi-directional communication to produce any output at all. It is a shortcoming of modern metrology that any one of these conditions make basic noise parameters unmeasurable.

Background and motivation: Accepted radio engineering practices includes a variety of reductive approaches to noise measurement and analysis. Designers combine various approaches that depend on practical measurement constraints

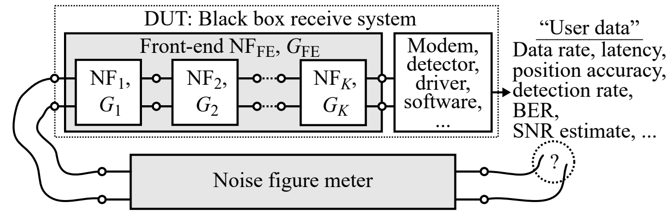


Fig. 1. The only receiver output available from packaged wireless equipment is usually heavily processed data, which is incompatible with test equipment. Cascaded analysis is also impractical because the parameters NF and gain G are unknown.

and the level of detail needed to meet a specification. A typical start is the rule of thumb that a receiver's noise figure (NF) should be only slightly greater than that of its front-end low-noise amplifier (LNA) [1, p. 495]. The Friis equation [2] adds the NF and gain of each cascaded stage in the front-end shown by Fig. 1. The model parameters are measurable with the Y-factor technique [3], [4]. More intricate models and measurements also account for wave parameters of signals and noise [5], as well as semiconductor device characteristics [6]. Unfortunately, the receiver "black box" of Fig. 1 leaves these parameters unknown and inaccessible for measurement. Systems models and third-party testers must then make guesses or assumptions to characterize the receive system as a whole.

Communication industry test standards attempt to circumvent the missing noise measurement by testing sensitivity instead [7], [8] — the input signal power threshold that provides a minimum level of system operation. The goals here are to support both performance comparison between receivers and link budget analyses. Only interoperable receivers can be compared directly this way, however, because the choice of link threshold is (i) specific to the receiver function and protocol, and (ii) depends on signal and protocol characteristics. Application to link budgeting is also limited, because the test conditions need to match the interference environment [9].

A noise measurement technique that is blind to the receiver implementation has many potential applications, such as:

i) *Characterization for link optimization:* A designer, customer, or third party could test a receiver for link budgeting or to confirm specification compliance.
ii) *Spectrum sharing and coexistence analysis:* Spectrum policy stakeholders face pressure to quantify coexistence performance between entrants and incumbents [10], [11]. These assessments require detailed signal-to-interference-plus-noise ratio (SINR) link models for data rate [12], [13] or radar detection probability [14]. Large tests have been undertaken in support of this work, including some by the authors (e.g., [15]–[20]), but have lacked the noise

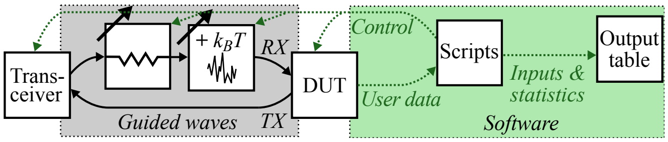


Fig. 2. The test space, which bridges the physical RF (left) and user data (right) domains.

component of SINR [21].

We elaborated the need for noise measurement data for interference testing in [21] through analysis of the coexistence test campaign in [20]. This led to our initial concept [22] for blind noise measurements.

Proposed Measurement: A test setup that accepts the "user data" outputs of Fig. 1 is shown by Fig. 2. The device on the left is a transmitter or transceiver that excites user data output from the device under test (DUT); the device can be a test instrument or even consumer wireless equipment. The measurement system attenuates this signal and adds excess thermal noise at calibrated, programmable levels. The attenuation may be directional in order to support tests of transceivers during bi-directional or full-duplex communication. We sample these inputs jointly at carefully chosen levels, and record the perturbed user data at the output. These user data, like the examples in Fig. 1, should be expected to respond as a function of input carrier-to-noise ratio (CNR) plus random variability; we verify this assumption post-hoc with the test data.

We propose a statistical regression technique to compute a measurement value from the test data. A minimum-error estimator identifies the measurand that aligns the user data as a function of CNR. To estimate measurement uncertainty, a Monte Carlo simulation repeats the regression on test data that is perturbed by (i) the variability in user data and (ii) physical models for uncertainty in signal and noise.

New work was necessary on several fronts to realize this new measurement:

i) a formal definition of "user data" as a system output,
ii) new sampling techniques to reduce the number of required samples,
iii) a new non-parametric method to estimate the measurand,
iv) a Monte Carlo simulation approach to propagate uncertainties from variability in user data through the regression into the measurand,
v) methods to assess whether the user data is CNR-dependent,
vi) a topology design and calibration method for the test system that implements the measurements, and
vii) a design for a highly-directive, programmable attenuation test system supporting live bi-directional links and a transceiver DUT.

We detail these contributions in this article, and frame their role in the measurement. As examples, we also present case studies of measurement applications to a consumer wireless local-area network (WLAN) client (Section V) and a global positioning system (GPS) L1 receiver (Section VI). Some additional discussion is also given to a WLAN access point (AP) in Appendix A.

TABLE I
RECEIVER RESPONSE PARAMETER LISTING

B	Noise integration bandwidth
C	Carrier power available to the receiver input
CNR	Carrier-to-noise ratio, in dB
E	Excess noise power injected into the receiver
f	User data response function
k_B	$1.38\ldots \times 10^{-23}$ J/K
N	Physical noise power integrated across bandwidth B
$N\|_{\text{in}}$	System noise power of a receiver or front-end
NF	Noise figure of a receive system or front-end
T	Noise temperature
T_0	Reference noise temperature (conventionally 290 K)
T_1	Minimum noise temperature of the measurement system
y	User data output from the receiver
ϵ_y	Random variable that encapsulates user data variability

II. MODEL OF RECEIVER NOISE RESPONSE

The measurand is receive system noise ($N|_{\text{in}}$) or noise figure (NF). These quantify the input-referred[1] noise performance of a DUT with an input termination at specified noise temperature. The input parameter space in experiments is CNR, which is determined by (i) incident signal power C and incident excess noise E, together and (ii) the measurand. These parameters are not new, but we review them for the present context. Finally, with these in mind, we give a simple stochastic functional model for the relationship between the input CNR and output user data.

A. Measurand Parameters

The canonical input noise level parameters in microwave networks are noise temperature T (in Kelvin), noise power N (integrated across noise bandwidth B), and NF [1]. Noise measurements are fundamentally traceable to physical temperature, and so naturally connected with T [29], but expression with N is convenient for comparison against signal power. These parameters are related through frequency-dependent power spectral density $N_0(f)$ as

$$N = \int_{f_L}^{f_H} N_0(f)\, df \approx k_B T B, \tag{1}$$

with $B = f_H - f_L$. The approximation is effectively exact except at very high frequency or cryogenic physical temperature [30]. Our convention here is to imply our use of this approximation when we use T. The frequency indicated with units implies B — for example, 0 dBm/10 MHz to suggest $B = 10^7$ Hz. The bounds for the integration in frequency (or averaging in time) need to be defined clearly because they vary by application.

B. Noise in of a Receiver Front-End

Consider the 2-port receive front-end in Fig. 1. At its input, there are signal and noise waves incident with available power C and $k_B T_1$, respectively. The available power output by the front-end has signal component C_{out} and noise component

[1] Also known as "equivalent input noise" in some acoustics and electromagnetic compatibility literature [23]–[25] and similar to the more qualitative "noise-equivalent power" in use by radio astronomy and optical detection [26]–[28].

N_{out}. The sensitivity of a receive system built with this front-end is limited by $C_{\text{out}}/N_{\text{out}}$.

When we consider the assembled receive system of Fig. 1, this output is inaccessible. This motivates the use of input-referred system noise, $N|_{\text{in}}$, which relates $C_{\text{out}}/N_{\text{out}}$ back to the front-end input as

$$\frac{C}{N|_{\text{in}}} = \frac{C_{\text{out}}}{N_{\text{out}}}. \tag{2}$$

The front-end adds new noise, so $N|_{\text{in}} > k_B T_1 B$. The NF quantifies the resulting decrease in output CNR under the condition that input temperature is equal to the reference ($T_1 \to T_0$). Thus,

$$\text{NF} = 10\log_{10}\frac{\left(\dfrac{C}{k_B T_0 B}\right)}{\left(\dfrac{C_{\text{out}}}{N_{\text{out}}}\right)} = 10\log_{10}\frac{N|_{\text{in}}}{k_B T_0 B}, \tag{3}$$

from (1) and (2), and matching [2]. The NF can therefore be understood as an alternate form of $N|_{\text{in}}$, expressed here in dB to follow modern convention. An expanded calculation for physical measurements at input temperature T_1 is given by Appendix C.

C. Equivalent Thermal Noise Power

The model (1) idealizes the added noise as additive and white. Yet, a realistic receive system is complicated by response to other factors like electromagnetic interference (EMI), state machines with hidden variables, and non-deterministic execution. We assume that these random processes, together, comprise an equivalent level of additive (but not necessarily white) noise[2].

Under this assumption, the following thought experiments are equivalent:

i) Suppose that we replace the input source and receiver electronics with noiseless copies. In this case, adding $N|_{\text{in}}$ at the receiver input reproduces the behavior of the actual receive system.
ii) Injecting noise into the receiver equal to $N|_{\text{in}}$ doubles the noise floor, reducing CNR by 3 dB.

It is therefore *equally valid* to think either in terms of the indirect noise response as (i) or the physical input noise levels in (ii).

D. Input Parameter Space

In order to probe this CNR space, we add a new degree of freedom: excess available noise, with total power E in the same band as $N|_{\text{in}}$. This noise is injected at the receiver input, and is uncorrelated with the input-referred noise of $N|_{\text{in}}$. The total CNR under these conditions is

$$\text{CNR} = 10\log_{10}\left(\frac{C}{N|_{\text{in}} + E}\right). \tag{4}$$

[2]Components of this discrepancy that vary with CNR are randomized in Monte Carlo uncertainty simulations (Section III-F) and included in the estimated uncertainty result. A verification technique like the cross-comparison in Appendix E identify some definitional and systematic errors.

We adopt the convention for this work of expressing and computing CNR in dB. The reference condition for impedance in each power quantity is available power (following the definitions of noise figure).

The experiments that follow will sweep attenuation on C and E to probe the CNR input space. These do not leave us any means to separate undesired components of the transmitter output (noise, phase noise, distortion, etc.), so we leave them as components of the signal power, C.

E. User Data and its Response Function

We adopt a simple non-parametric model to represent the transformation of input CNR to processed receiver output ("user data"). Each output sample, y, is assumed to respond as a function of CNR plus random variability,

$$y = f\left(\text{CNR}\right) + \epsilon_y, \tag{5}$$

across a range of CNR. The response function, f, is specific to type of output from the receive system. It characterizes the averaged transformation from the physical CNR domain into the user data output domain. Its argument, from (4), is in dB units[3]. The random variable ϵ_y, with unknown distribution, represents random variability in the user data. This variability encapsulates non-deterministic processes in the link and receiver, such as noisy self-estimates of CNR, impacts of noise on state transitions, and unknown impacts of memory from previous CNR conditions.

III. Measurement Method

The idea behind the experiment is to perturb a DUT at different calibrated levels of both signal and excess noise, and sample the resulting user data output. A new statistical regression technique estimates both the DUT system noise measurement value, as well as its uncertainty interval.

Each sampling point is a pair of power levels, (C, E), that produces a resulting sample of user data output, y. An experiment comprises two sets of M_y sampling points:

i) $y_{E=0}$ – excess noise disabled:
These output samples are acquired with input sampling that varies C with no excess noise. Since $N|_{\text{in}}$ is the only noise in CNR here, the trend in $y_{E=0}$ against C traces the user data response function, $f(\text{CNR})$. We therefore use these data to compute a user data response function estimate, $\hat{f}_{E=0}(\text{CNR})$.
ii) $y_{E>0}$ – excess noise enabled:
These samples result from jointly varying both C and E. The resulting user data response function estimate, $\hat{f}_{E>0}(\text{CNR})$, depends on noise as $N|_{\text{in}} + E$.

The regression that we introduce below hinges on perturbing the CNR for the $E = 0$ data differently than that of $E > 0$. We developed a new approach to selecting the input sampling points, which is detailed by Section III-B. Each $y_{E=0}$ or $y_{E>0}$ sample is an average of (and corresponding estimate

[3]This transformation follows link modeling convention, and makes the magnitude of the uncertainty sensitivity coefficient equal to one, as shown in Appendix D.

of variability in) the steady-state window of an M_s-point time series, which is described in Section III-C.

The regression process seeks the measurement value ($N|_{\text{in}}$) that transforms the CNR to align the function estimates $\hat{f}_{E=0}(\text{CNR})$ and $\hat{f}_{E>0}(\text{CNR})$. We approach this as an iterative optimization. The optimizer iterates searches trial levels of noise (N_t) for the minimum integrated residual error $|\hat{f}_{E=0}(\text{CNR})-\hat{f}_{E>0}(\text{CNR})|^2$ as

$R_{\text{min}} \leftarrow$ unset
while optimizer.not_converged() **do**
  $N_t \leftarrow$ optimizer.next()
  $\text{CNR}_t \leftarrow 10\log_{10}[C/(E+N_t)]$
  $\hat{f}_{E=0}(\text{CNR}_t) \leftarrow$ estimate on $y_{E=0}$, CNR_t *Sec. III-D*
  $\hat{f}_{E>0}(\text{CNR}_t) \leftarrow$ estimate on $y_{E>0}$, CNR_t *Sec. III-D*
  $R^2 \leftarrow \int (\hat{f}_{E=0} - \hat{f}_{E>0})^2 \, d\text{CNR}$ *Sec. III-E*
  if R_{min} is unset or $R < R_{\text{min}}$ **then**
    $N|_{\text{in}} \leftarrow N_t$
    $R_{\text{min}} \leftarrow R$
  end if
  optimizer.update (N_t, R)
end while

The optimizer is unspecified here for generality. For the examples in this paper, we use brute force optimization, presuming that the regression is computationally inexpensive. The trial CNR, CNR_t, is determined by N_t and the calibrated input levels with (4). The computation to estimate $\hat{f}_{E=0}$ and $\hat{f}_{E>0}$ applies a Gaussian kernel to the y data and CNR_t. The minimum value of the regression residual R, which characterizes the disagreement between these estimates, points to the measurement value result. Last, the Monte Carlo simulation of Section III-F gives an estimate of the measurement uncertainty by perturbing the above process with physical error distributions and user data variability.

Measurement hardware implementation is left to Section IV, and examples of application-specific details are given in the subsequent case study examples. The computations that follow are performed in linear power and attenuation (i.e., dB are only converted after all computations are complete), except CNR, which is in dB following (4).

A. *User Data Selection*

A receiver is likely to support many different user data outputs, but we select one to perform the regression. It is acceptable at this stage to simply guess that it responds as a function of CNR as defined in (5); this is validated later with the processed data (Section II-E). Suitable user data candidates could include

- benchmarking data determined by test software, such as data rate, network latency, or positioning accuracy; or,
- receiver self-diagnostic information such as self-estimated CNR, C/N_0, or bit error rate.

When possible, memory depth of user data processing in the DUT should be minimized to reduce correlation between time series samples.

B. *Input Sample Selection and Sequencing*

The selection of input power levels that comprise the input sampling pair (C, E) needs to be approached very carefully. Since the test centers on comparison between estimated user data responses ($\hat{f}_{E=0}$ and $f_{E>0}(\text{CNR})$) over CNR, the sampling points should approximate the same number and values of CNR for both $E = 0$ and $E > 0$. Close agreement in the achieved CNR for these sample points can help to reduce or eliminate biases in estimating $\hat{f}$ (and their propagation to the final $N|_{\text{in}}$). A useful secondary goal in the sampling point selection is to maximize the statistical power of the regression, in order to reduce test time or measurement uncertainty.

The sequence order of the input samples is also important, because a DUT may hold residual memory from prior input samples. This may create undesired correlation between samples of y, biasing $\hat{f}$ estimates and the resulting measurement value $N|_{\text{in}}$. To mitigate this memory effect, we randomize the sequence of input samples in the experimental acquisition.

Input samples for $E = 0$: Ideally, the vector of sampling points would exactly duplicate the input CNRs of the $E > 0$ samples. This is impossible, because the input CNR depends on the unknown $N|_{\text{in}}$. This leaves us with a bootstrapping problem, which we resolve with an initial guess. We parameterize the set of $E = 0$ sample points for this approach as follows:

- N_g: initial guess for the measurement result;
- $\min(\text{CNR}_g), \max(\text{CNR}_g)$: goal for the bounds on the achieved input CNR; and,
- M_y: the number of sampling points.

The N_g guess could come from a datasheet specification, if available, or a few dB above $k_B T_0 B$. The CNR_g bounds should be chosen with the goal of producing a strongly CNR-responsive range of $y_{E=0}$. This domain could be gauged from a datasheet or protocol specification, or exploring experimentally. We have observed the best results for M_y on the order of several tens or more.

The following algorithm generates the M_y input sample pairs $(C[k], E[k])$ from the above $E = 0$ sampling parameters:

span $\leftarrow \max(\text{CNR}_g) - \min(\text{CNR}_g)$
for $k \leftarrow 1 \ldots M_y$ **do**
  $\text{CNR}_g[k] \leftarrow \min(\text{CNR}_g) + (k/M_y) \times$ span
  $C[k] \leftarrow N_g \times 10^{\text{CNR}_g[k]/10}$
  $E[k] \leftarrow 0$
end for

The C calculation here comes from (4).

Input samples for $E > 0$: Because sampling with $E > 0$ opens a new degree of freedom, we also add another pair of constraining parameters:

- $\min(\text{ENR}_g), \max(\text{ENR}_g)$: bounds on sample values for the excess noise ratio (ENR), $10\log_{10}(E/N_g)$, (in dB).

A reasonable value for the minimum is 0 dB, so that at least half of each $E+N|_{\text{in}}$ is excess noise. The maximum should be the lesser of (i) measurement hardware output limitations and (ii) any known minimum threshold at which the DUT does not respond with CNR.

The following algorithm transforms ENR bounds and CNR into (C, E):

span $\leftarrow \max(\mathrm{ENR}_g) - \min(\mathrm{ENR}_g)$
for $k = 1$ to M_y **do**
  $\mathrm{ENR}_g[k] \leftarrow \min(\mathrm{ENR}_g) + (k/M_y) \times \mathrm{span}$
  $E[M_y + k] = N_g 10^{\mathrm{ENR}_g[k]/10}$
end for
shuffle$(E[M_y \ldots 2M_y])$
for $k = 1$ to M_y **do**
  $C[M_y + k] = (N_g + E[M_y + k])10^{\mathrm{CNR}_g[k]/10}$
end for

This procedure completes the last M_y pairs of (C, E), for $2M_y$ total samples. The E in these CNR values are spread uniformly in dB. The extra shuffling step on the non-zero elements of E decorrelates it from CNR_g, mitigating bias from any behaviors not captured by the response model in (5).

Sequencing: Some experimental errors are random variables that vary both (i) slower than a single sampling point acquisition and (ii) faster than the acquisition of $2M_y$ sampling points. Examples in C and E could include temperature drifts or time-dependent ambient noise in the laboratory. Sources of error in y could include random residual state inside the receiver from other recent samples.

We mitigate bias from these errors by randomizing the sequence of sampling points, a standard practice in experimental design. The effect is to "average out" the resulting biases [31]. Further, the slow-varying errors tend to transform into uncorrelated random errors in $y_{E=0}$ and $y_{E>0}$, which we can propagate into measurement uncertainty. We maximize this benefit by randomizing all $2M_y$ sampling points, shuffling the $E = 0$ and $E > 0$ sampling points together.

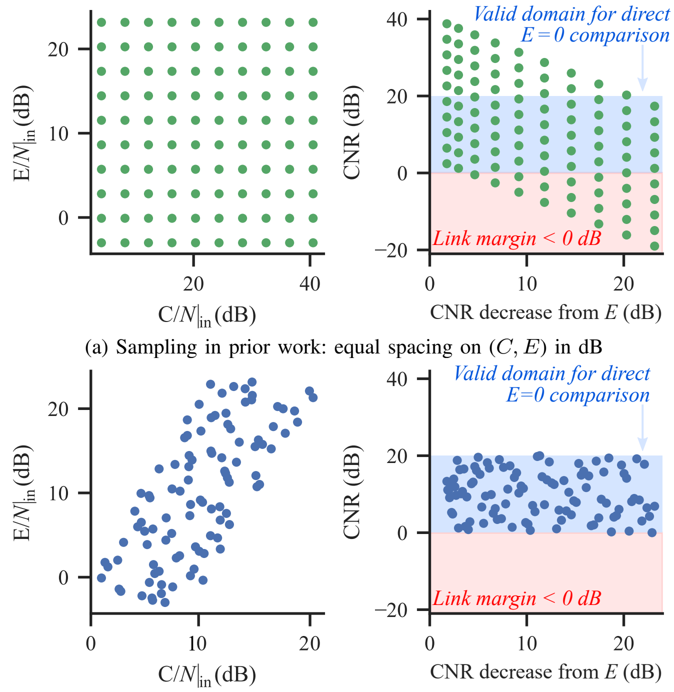
Comparison with prior efforts: Our previous experiments [22] sampled on a regular grid in (C, E) as shown by Fig. 3a. The grid edges shown are selected to ensure that sample points at 20 dB CNR are achieved even at maximum ENR. The CNR achieved here, shown on the right, is irregular. Sampling points outside the CNR range of $E = 0$, 0 dB to 20 dB, must be discarded, wasting testing time with useless data.

The new sampling method we propose to resolve this problem is illustrated by Fig. 3b. Suppose that the guess N_g is close to the measurement result, $N|_{\text{in}}$, and that the receiver requires a valid link CNR $>$ 0 dB to output user data. In this case, the CNR, based on the total noise $N|_{\text{in}} + E$, is distributed evenly across the intended range 0 dB to 20 dB. The figure on the right demonstrates that the CNR maintains a valid link matching the $E = 0$ domain (blue shaded region), and a balanced distribution in CNR across the 20 dB range of impacts from E.

C. Output Time-Series of User Data

We acquire M_s samples time series of user data at each (C, E). From each of these time series, we estimate (i) a central value y, which is the input for the estimate the user data response function $f(\mathrm{CNR})$) (Section III-D), and (ii) the user data variability, ϵ_y, for uncertainty simulations (Section III-F).

The y is computed only within the estimated steady-state span of the time series. To reject initial transients in the time-series, we apply the marginal standard error rule 5 (MSER-5) method, which was found to offer superior performance in a comparison study [32]. This algorithm locates the first sample in the time-series at which the standard deviation of batched 5-sample averages is minimized when the previous samples are deleted, and requires $M_s > 128$ [33].



(a) Sampling in prior work: equal spacing on (C, E) in dB

(b) Proposed sampling: uniform distribution on (CNR, CNR lost to noise)

Fig. 3. Examples of $E > 0$ input samples (CNR_g 0 dB to 20 dB, ENR_g -3 dB to 23 dB, $M_y = 100$) generated (a) equispaced on C and E following our prior work [22], and (b) with the proposed technique.

The steady-state window of data is the basis for the remaining statistics. The y estimate is the median average, chosen to mitigate the effect of outliers. The ϵ_y estimate is the estimated 95% confidence interval that captures the variability in this median. For time series that include strong correlations, classical estimators of confidence intervals for quantiles are unsuitable, since they are designed for independent samples. For this reason, we applied the averaged group quantile method of Heidelberger and Lewis [34] that is designed for quantile estimation from statistically dependent sequences. This method is nonparametric (it makes no distributional assumptions about the data).

The estimated confidence interval for the steady-state median captures variability only within the collected time series. It does not capture errors that are constant during the time series acquisition, such as initial DUT state at the beginning of acquisition, temperature drift, attenuation errors, etc. These factors are addressed in the uncertainty analysis of Section III-F.

D. User Data Response Function Estimate

The regression relies on comparing the two sets of user data responses that result from $E = 0$ and $E > 0$ at the same CNR. Yet, the input sample points do not produce the same set of CNR conditions for the $E = 0$ and $E > 0$ data partitions,

TABLE II
ANALYSIS PARAMETER LISTING

$\hat{f}_{E=0}$	User data response function estimate from $E = 0$ data
$\hat{f}_{E>0}$	User data response function estimate from $E > 0$ data
M_{MC}	Number of Monte Carlo simulation runs
N_t	Trial values of $N\|_{\mathrm{in}}$ selected by the optimizer
R	Regression residual error at a trial N_t
R_{min}	Regression residual error at the measurand $N\|_{\mathrm{in}}$
$y_{E=0}$	y sampled without excess noise
$y_{E>0}$	y sampled with excess noise

because $N|_{\mathrm{in}}$ is unknown. Instead, we estimate continuous-domain functions, $\hat{f}_{E=0}(\mathrm{CNR})$ and $\hat{f}_{E>0}(\mathrm{CNR})$, to support direct comparison. The same estimation process needs to be applied to each of $y_{E=0}$ and $y_{E>0}$ to minimize bias.

Since the unknown response function can take many forms, we estimate the response function by non-parametric regression. We use the locally-estimated scatterplot smoothing (LOESS) technique [?], [35], [36] here, in its original implementation [37]. It is designed to accommodate data with exactly the response function given by (5).

The principal parameter in LOESS regression, "span," sets the degree of smoothing. It is expressed as a fraction of the total span of the independent data (in this case, CNR), typically in the range 0.25 to 0.5 [38]. We have tested "span" values within this range on experimental data, but have not observed meaningful impact on the measurement value or its uncertainty; for consistency, we apply span 0.4. We scale this slightly for $E > 0$ data, to ensure that the smoothing width in dB units is the same as the $E = 0$ data. A more detailed parameter selection study may be useful in the future when a larger body of experimental data is available.

E. Noise Power Measurement Value Estimate

This is the computation that gives the measurement value, $N|_{\mathrm{in}}$. For simplicity, we use a brute-force search method with a resolution of 0.01 dB, though a study of other techniques would be worthwhile. Other algorithms could be considered in future work.

The optimization needs a cost function that quantifies the fitness of trial values of the measurand. The user data response function estimates $\hat{f}_{E=0}(\mathrm{CNR})$ and $\hat{f}_{E>0}(\mathrm{CNR})$ estimate of the same underlying f of (5). The measurement value $N|_{\mathrm{in}}$ will, ideally, align the two estimated functions as $\hat{f}_{E=0}(\mathrm{CNR}) \approx \hat{f}_{E>0}(\mathrm{CNR})$. We therefore propose that the cost function should be the residual error in this alignment, evaluated numerically as:

$$R(N_t)^2 = \frac{1}{K}\sum_{k=1}^{K} \Delta_k^2(N_t). \tag{6}$$

The sum should sample at least $K > M_y$ points. The residual at the k^{th} CNR, Δ_k, is

$$\Delta_k(N_t) = \hat{f}_{E=0}(\mathrm{CNR}_k; N_t) - \hat{f}_{E>0}(\mathrm{CNR}_k; N_t). \tag{7}$$

The semicolon notation means that the response functions need to be re-estimated from the test data for each N_t. The CNR values, in turn, are spread evenly in dB as

$$\mathrm{CNR}_k = \left(\frac{k}{K}\right)\left[\min(\mathrm{CNR}) - \max(\mathrm{CNR})\right]. \tag{8}$$

The min and max here indicate the extrema supported by both estimates $\hat{f}_{E=0}$ and $\hat{f}_{E>0}$.

To mitigate the possibility of non-convex R and to accelerate numerical evaluation, we recommend constraining the parameter search by (i) the physical lower bound $N|_{\mathrm{in}} \geq k_B T_1 B$, and (ii) the interval defined by some minimum fraction (for example, 0.5) of overlap between the CNR sampling in the $E > 0$ and $E = 0$ sampling points.

F. Measurement Uncertainty

Uncertainty estimation is mature and well-understood in total power and Y-factor noise measurements [4]. The physical sources of error in these measurements (such as impedance mismatch, connector repeatability, and detector linearity) are related to the measurand through a measurement equation that is closed-form and differentiable. This type of model suits the classical law of propagation of uncertainty [39].

In contrast, the regression process that we have defined for blind noise measurement is both nonlinear and nonparametric. We therefore require a new approach to uncertainty analysis. To make the problem tractable, we propose a hybrid method. The idea is to propagate random errors through the estimation procedure with Monte Carlo simulation. The resulting uncertainty estimate, u_{MC}, encapsulates random errors between subsequent samples of y. We then show that the classical law of propagation of uncertainty can be used to combine u_{MC} with the systematic uncertainties in C and E.

Random and definitional errors in the regression: Random errors in the regression represent variability that arises between different samples of user data. This uncertainty comprises contributions from random variability in the user data, errors introduced by the regression process, and physical errors that vary between sampling points. Definitional errors represent imperfection in aligning the user data response estimates with the CNR response model.

Propagating uncertainty from these errors is complicated by the transformations of the user data through the highly non-linear regression process and the unknown processing underlying the user data in the DUT. Purely analytical methods are not straightforward, and possibly intractable. Ideally, a data-driven approach should be nonparametric, but to our knowledge, the problem of uncertainty estimation for non-parametric regression has not been addressed in the present context[4].

We are left to use Monte Carlo simulation to estimate this uncertainty component, u_{MC}. It is computed by simulating (i) random error sources in C and E, (ii) empirical variability in the user data, and (iii) the order and sign of the disagreements between $E > 0$ function estimate relative to $E = 0$ user data ("cross-residuals"). Each Monte Carlo trial is implemented as follows:

1: perturb C and E sample points with Gaussian errors (with standard uncertainty estimated from variability as assessed in Appendix C)

[4]The statistical literature calls this regression problem "errors in variables regression," where the errors in the controlled independent variable are of the Berkson type [40].

2: perturb each user data sample y with normally-distributed errors scaled by (its 95% CI width)/(2×1.96)
3: estimate $N|_{\text{in}}$ following Section III-E
4: compute a cross-residuals vector, $\hat{f}_{E=0}(\text{CNR}) - y_{E>0}$
5: randomize the sign of each cross-residual
6: generate new perturbed $y_{E>0}$ by adding random cross-residual samples to $\hat{f}_{E=0}$
7: re-estimate $N|_{\text{in}}$ with perturbed $y_{E>0}$ (Section III-E)

Steps 1-4 perturb the result with random errors in physical inputs and user data outputs. Randomizing the structural differences in errors between $E = 0$ and $E > 0$ (the "cross-residuals") in 5-7 helps to encapsulate the extent to which the user data did not respond as a function of CNR.

The estimated 95% confidence interval (CI) bounds on $N|_{\text{in}}$ are taken from the 2.5% and 97.5% quantiles of the empirical distribution of Monte Carlo trials. We estimate the standard uncertainty from the CI as $u_{\text{MC}} \approx \text{CI length}/(2 \cdot 1.96)$. This assumes that the simulation results have Gaussian distribution, which can be verified with large M_{MC}. Variability in the estimated u_{MC} can be reduced by increasing the number of sampling points.

The resulting uncertainty estimate is unavoidably looser than one informed by a model for the process in the receiver that generates user data. One reason is that the variability in cross-residuals includes LOESS smoothing artifacts in $\hat{f}_{E=0}$, even though this smoothing should not propagate as error in $N|_{\text{in}}$ when applied to $\hat{f}_{E>0}$. Randomizing the sign of the cross-residuals helps to ensure that the error is unbiased, but broadens the uncertainty interval. As a result, this uncertainty estimate should be considered conservative.

Systematic errors in the regression: Each power level C and E was calibrated before the measurement data acquisition. The calibration is a measurement of the offset value (in dB) that corrects attenuation to physical output power at the center frequency under test. The calibration measurements themselves include errors; only some of these vary between calibration measurements. The calibration technique detailed in Appendix C applies constant offset corrections to each of C and E. Because the same offset calibration (and calibration error) applies at each input sampling point, we refer to these calibration errors as systematic errors in the regression; these are constant for all acquired data. This systematic regression error, in turn, consists of *random calibration errors* and *systematic calibration errors*.

Appendix D shows that a systematic regression error in E (in dB) produces an error that propagates to $N|_{\text{in}}$ with equal magnitude. This means that any component of uncertainty in the calibration of E propagates into the measurand with the same magnitude (both also in dB). The classical law of propagation of uncertainty therefore applies with its usual restrictions. For the uncertainty component corresponding with each error source, the sensitivity coefficient propagates to the measurand with the same value as for E [39], so analysis of these error sources applies in the classical sense, just as in [4].

Errors in the calibration of C cause an constant shift along the CNR (independent) axes in the user data. This changes the input conditions of the receiver during test. If receiver response in the new input domain is still CNR-dependent and produces

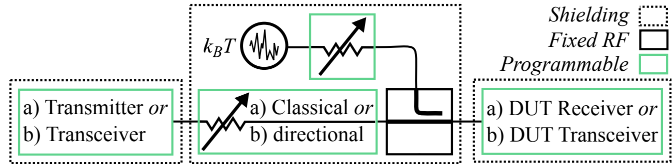


Fig. 4. Testbed topology for measurements noise figure of (a) a receiver DUTs excited by a transceiver through a classical (reciprocal) attenuator, or (b) a transceiver DUT excited by another transceiver through the directional attenuator of Fig. 5.

the same user data variability, then calibration error in C has no impact on $N|_{\text{in}}$ by the reasoning of Appendix D. Otherwise, changes in variability contribute to the random errors in the regression that are captured by Monte Carlo[5] in u_{MC}. .

Combined uncertainty: The combined standard uncertainty, u_c, is the root sum square (RSS) of uncertainties that originate from the above (i) random errors in the regression, (ii) random errors in the calibration of E, and (iii) systematic errors in the calibration of E. Following [39], the expanded uncertainty to 95% confidence is $U = 2u_c$.

This calculation presumes underlying errors are uncorrelated. This is reasonable for the calibration techniques given by Appendix C, because the dominant errors originate in measurements taken with different instruments that are calibrated against different physical standards.

G. Assessing Dependence of User Data on CNR

The minimum value of the residual (7) — achieved during optimization at the measurement value, $N|_{\text{in}}$ — gauges the extent to which user data behaves as a function of CNR. It is scaled in the arbitrary user data units, however, and we also desire a unitless relative normalization to compare the performance between different types of user data and DUTs. For this purpose, we define the following relative residual:

$$\overline{R}^2 = \frac{1}{K} \sum_{k=1}^{K} \left[\frac{\Delta_k(N|_{\text{in}})}{\hat{f}_{E=0}(\text{CNR}_k; N|_{\text{in}})} \right]^2. \tag{9}$$

The expression is evaluated at the final measurement result. The user data "responds as a function of CNR" if $\overline{R} \approx 0$. A relative residual that nears or exceeds 1, in contrast, suggests that the user data behavior exhibits some other behavior, or extremely high veriability in y. These large $\overline{R}$ tend to correspond with large measurement uncertainty.

IV. Laboratory Implementation

The basic measurement system topology is illustrated by Fig. 4. The transmitter or transceiver on the left excites the signal incident on the DUT. This signal's center frequency and bandwidth determine those of the DUT noise measurand. The new use of the directional attenuator here extends testing beyond [22] to support transceiver DUTs that share ports with

[5]Fixed shift in CNR caused by calibrations shifts the sampling range of the acquired data slightly. This change in the test domain may change (i) variability in the user data, which is captured by the random errors in the regression, or (ii) power-dependent behaviors in the receiver, as characterized by the relative residual discussed in Appendix E

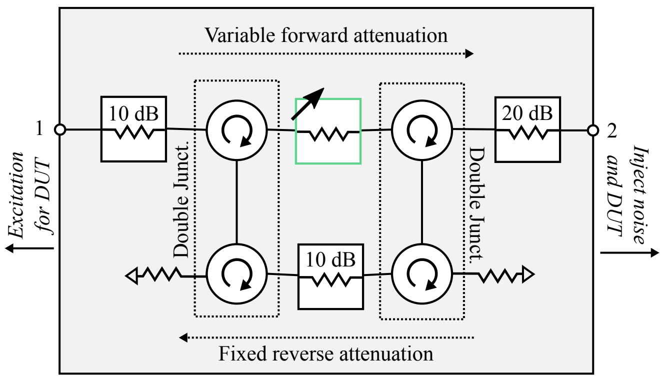
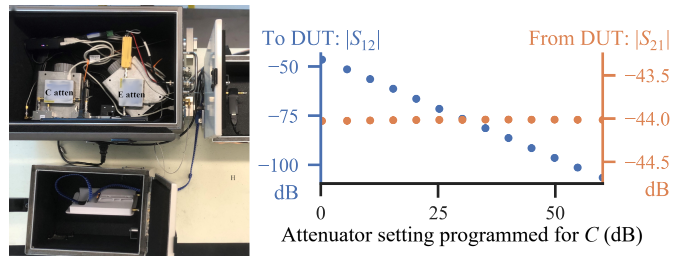


Fig. 5. Test schematic for a directional variable attenuator in tests of transceiver DUTs following Fig. 4b.



(a) (b)

Fig. 6. Directional attenuation (a) benchtop implementation, and (b) measured attenuation at 5.3 GHz. The data illustrates the design goal: flat reverse attenuation and 1 dB attenuator setting per 1 dB forward attenuation.

transmission. This ensures that only the link incident toward the DUT is tested, even for transceivers that are duplexed in any of time, frequency, or coding. We have implemented this test for receivers to 6 GHz with readily-available commercial parts. We use programmable attenuator components that span 110 dB range and 0.25 dB resolution to accommodate the wide range of tolerance for link loss in different DUTs. The calibration of the attenuator settings is detailed in Appendix C.

Programmable excess noise: Sampling points with excess noise require the test system to operate as a programmable-ENR noise source.

Amplified noise diodes are readily available to consumers and make a convenient excitation source for this application. We purchased one specified at around 57 dB ENR to 10 GHz. The minimum insertion loss between the source and the DUT in Fig. 4 is determined by the attenuator and the coupler. Our reference implementation with a 20 dB directional coupler totals 27 dB to 29 dB loss near 6 GHz, leaving a programmable ENR range on the order of 0 dB to 30 dB.

We disable noise output for $E = 0$ samples by setting the noise path attenuation to its maximum (i.e., minimum transmission) so that $E \ll k_B T_1 B$. Network analyzer measurements confirmed that this setting attenuated the output by at least 100 dB, reducing E to at least 70 dB below thermal noise. An attenuation range of 60 dB is enough to effectively disable the programmable excess noise, biasing $N|_{\text{in}} + E$ by less than 0.01 dB. We also observed no measurable difference between the attenuator and a room-temperature termination on a spectrum analyzer, confirming that the minimum excess noise power is negligibly small.

It is important to ensure that the noise level is controlled precisely, because uncertainties in the calibrated output propagate to the measurement result. We recommend calibration for the excess power level with the Dicke radiometer technique detailed in Appendix C. It is simple and yields lower uncertainty than the power measurement method in our prior work [22].

Programmable directional attenuation: Transceiver DUT testing needs programmable directional attenuation. This permits control over the signal power incident on the DUT without impact to the signal transmit from the DUT, which may be required for normal operation of the DUT.

We define directional attenuation by the following performance goals:

i) Loss in the "forward" path (waves incident into port 1, scattered from port 2 toward the DUT) is programmable. Ideally, the realized forward attenuation (in dB) is controlled exactly by attenuation in the forward attenuator (in dB).
ii) Loss in the "reverse" path (waves incident from the DUT into port 2, scattered out of port 1 toward the test system) is fixed. Ideally, this is independent of the forward attenuation.

Deviation from these ideals introduces a random error in the input sampling points, and in turn the measurement uncertainty. An automated measurement system needs at least a few tens of dB of programmable range in forward attenuation; a coarse adjustment can be made before test time with fixed attenuators. Wideband operation is desirable to reduce the number of directional attenuators that need to be implemented and calibrated. To our knowledge, it has been some time since the last published work on directional attenuation [41]. The results of that work is not suited for our purposes here, because forward attenuation tuning range was only 20 dB, achieved fractional bandwidth was about 10%, and its rectangular waveguide implementation is incompatible with most DUTs.

We developed a multi-stage coaxial directional attenuator with expanded bandwidth and attenuation range for versatile use in measurements. Its schematic is shown by Fig. 5. The forward and reverse paths are split with two stages of circulators. Forward waves propagate through the programmable attenuator, and reverse waves are attenuated by the fixed 10 dB. The remaining fixed pad attenuators help to maintain isolation between the forward and reverse paths in case of reflections at the junctions with the DUT (or its excitation). The indicated use of double-junction circulators gives a similar benefit at the junction between the constituent circulators.

A coaxial implementation for 4.4 GHz to 6 GHz is pictured in Fig. 6a. The lower bandwidth limit is the pass-band of an output filter, and the upper limit is the programmable attenuator. The main practical constraint to improving the bandwidth of this topology is the circulators, which are available commercially up to about an octave. We calibrated and characterized this directional attenuator as described in Appendix C. Across the full bandwidth and 0 dB to 60 dB attenuation settings, the forward attenuation error was 0.04 dB root mean square (RMS), and the reverse attenuation error was 0.02 dB RMS. These are the standard uncertainties in the C and E input sampling points, respectively.

Shielding: Shielding each block from the electromagnetic environment mitigates ambient noise and interference. Unshielded noise and interference adds to $N|_{\text{in}}$, biasing the measurement. We used enclosures with shielding effectiveness specified above 80 dB between hundreds of MHz to 6 GHz. This shielding provides data pass-throughs for test automation with filtered connectors.

Automation: These measurements need robust control and data acquisition from the DUT. The $2 \times M_y \times M_s$ (order of at least 10^4) time series samples need to be acquired faithfully and without crashing. This may be the greatest difficulty in the experiment, because many devices are not designed to facilitate this type of test. The automation needs to include ongoing, aggressive validation at run-time in order to confirm that the DUT is in the intended state, and command retries when appropriate.

Fortunately, the list of basic functions needed for this type of test is usually short. Many require only a subset of the following:

- `acquire`: fetch a user data time series (or part of it)
- `wait`: pause testing until the DUT is ready
- `reset`: attempt to clear the receiver state and memory
- `[dis]connect`: for testing with stateful network connections

Many DUTs act as black boxes that give little or no feedback to acknowledge proper operation. The resulting uncertainty about the state of the DUT may raise concerns about the integrity of the measurement. Luckily, the body of test data is itself useful for automation problem-solving and validation:

i) $\overline{R} \approx 0$ confirms the automation behavior by confirming that the control and outputs have produced CNR-dependent response by DUT;
ii) noisy or intermittently missing y may suggest that control over DUT state `reset`, `[dis]connect`, or `wait` are inconsistent, and;
iii) constant y suggests that `acquire` does not give the expected data.

Still, i) and ii) leave some ambiguity. Spurious outputs, or user data response that is not a function of CNR, may be a feature of the DUT that cannot be overcome externally. The unavoidable result in these cases will be noisy or spurious user data.

V. Case Study on a 5 GHz WLAN Client

Recent interest in coexistence between WLAN and LTE license-assisted access (LTE-LAA) motivated us to test WLAN equipment operating in the 5 GHz industrial, scientific, and medical (ISM) band. Physical layer modeling in this problem space typically hinges on the response of user data as a function of SINR, for example in [12]. The receive node noise performance is therefore one of the input parameters required to determine this SINR.

Our first case study here demonstrates noise measurements of a consumer WLAN client. The measurements that follow do not access the front-end output of the DUT. Use of generic data throughput tests as user data also means that no support from the manufacturer was required, because we used no special debug or diagnostic programming mode.

TABLE III
WLAN Communication Link Parameters

Communication standard	IEEE 802.11a
Center frequency	5.3 GHz (channel 60)
Channel bandwidth and B	20 MHz
AP power output setting	11 dBm
Network protocol	TCP/IP
TCP socket buffer size, M_{bytes}	8 kB

TABLE IV
WLAN Experimental Parameters

Sampling points in each of $E = 0$ and $E > 0$	M_y	41
Time series samples per y	M_s	1000
Monte Carlo simulations	M_{MC}	10^5
Test system ambient temperature	T_1	300.2 K
Sampling guess		
of the client	N_g	-95 dBm/20 MHz
of the AP	N_g	-92 dBm/20 MHz
of the LNA	N_g	-99 dBm/20 MHz
CNR goal domain	CNR_g	10 dB to 20 dB
ENR goal domain	ENR_g	0 dB to 20 dB

A case study on the AP device used to excite these tests is given in Appendix A. The raw experimental data for both tests are published in [42].

A. Equipment Under Study

The DUT is a consumer WLAN client, configured with the communication link parameters listed in Table III. The noise measurement frequency is determined by the excitation signal from the AP, 5.3 GHz.

1) WLAN Client: We purchased a consumer WLAN client with a coaxial RF connection. The test PC gave it data and power by USB. We had no access to control or diagnostic information over the DUT beyond the generic capability of the networking drivers that were installed automatically by the automation computer operating system..

2) WLAN Access Point: The WLAN access point was a packaged consumer device that also functions as a network router, manufactured by a different vendor than the client. The network connection to the automation computer was category 6 ethernet wired to a dedicated network interface, specified at 1 Gbps by its manufacturer. Control over the WLAN center frequency is in the configuration page of the AP, accessed by web browser from the automation computer. We set this before collecting any data, and changed no other settings.

3) Verification LNA: This was a commercially-available LNA with coaxial ports. We calibrated its gain and noise figure as a reference at the WLAN center frequency with a commercial Y-factor measurement, discussed in Appendix E.

B. Test Implementation

The automation computer operated the AP and client as IPV4 network interfaces. The test runs entirely on the application layer of the 802.11a protocol through the default operating system drivers. The automation computer connected the sender and receiver into a single closed network. We bound the transmission control protocol (TCP) socket connections to the corresponding send and receive interfaces to ensure that traffic passed was routed through the WLAN link. The communication between the AP and the client DUT that we used for testing here is therefore live bi-directional traffic. Most of the data traffic was WLAN uplink (AP to client), but lower layers of the 802.11 protocol here also send handshaking and other overhead through the downlink. The parameters of the experiment are listed in Table IV.

User data selection: The user data under test is median estimated data rate, tested from the AP into the DUT with TCP sockets. This "key performance indicator" is widely used and frequently tested, and applicable in general to computer networking equipment. The median statistic helps reduce the variability of the result. The time series for each y rate has $M_s = 1000$ samples, which are each estimated by sending M_{bytes} pseudo-random bytes to the DUT. The data rate estimate $M_{\text{bytes}}/\Delta t$, where Δt is the time elapsed estimated from processor clock ticks on the automation computer. This type of timing estimate is suited for this application, because it only needs to be CNR-dependent with low variability; absolute accuracy is not required.

Data acquisition: We implemented Python scripts to automate the measurement, including and the client DUT. These functions use the general-purpose WLAN drivers and TCP/IP network implementations provided by the operating system in the automation computer. We only use generic, open drivers and software libraries in order to increase the likelihood that these scripts can also support other WLAN client models and vendors.

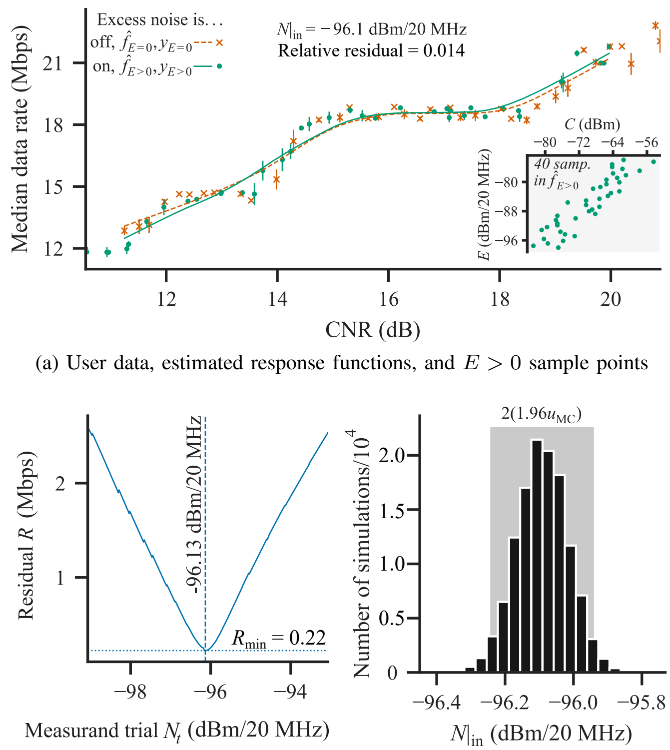
A fresh TCP/IP socket makes a new network connection for each sampling point. Nagle's algorithm [43] is disabled in these sockets, reducing the use of TCP/IP memory buffers that might span multiple measurement sample points.

The following automation loop acquired time series in each sampling point:

attempt WLAN client reconnect
if WLAN client connected **then**
  for M_s time series samples **do**
    send M_{bytes} randomized data to the DUT
    record data throughput rate
  end for
  disable traffic
  disconnect WLAN client
end if

This applies to testing either WLAN client or AP DUTs. Our automation control over the AP that interacts with the DUT was the most limited, and we were unable to implement `reset`. Each sampling point took about 10 s to 15 s, mostly spent awaiting new connections in the DUT.

Experimental parameter selection: The N_g guess for input sampling was determined by a coarse initial measurement. The guess for the LNA input noise came from the Y-factor characterization of the LNA. The goal range for CNR spans a wide range of observable changes in data rate.



(a) User data, estimated response functions, and $E > 0$ sample points

(b) $N|_{\text{in}}$ regression residual (c) Random error simulations for u_{MC}

Fig. 7. Data, regression, and simulation of uncertainty from random error in the first WLAN client measurement.

C. Results and Discussion

The measured WLAN client system noise was (-96.08 ± 0.18) dBm/20 MHz, or, equivalently, (4.84 ± 0.18) dB noise figure. The acquired data and regression analysis for this measurement are shown by Fig. 7. Further repeatability and Y-factor verification measurements are shown in Fig. 8.

Data and regression: Figure 7a shows the user data samples and resulting response functions estimates. The data rate trends upward with CNR, following changes in modulation scheme at various thresholds in CNR. The CNR values along the horizontal axis are calculated from the calibrated input signal power C, the calibrated input excess noise power E, and the measurement value $N|_{\text{in}}$. The vertical axis shows user data. The C and E sample points for $E > 0$ sampling points are inset on the lower right.

The superimposed $E = 0$ and $E > 0$ curves give some intuition for the alignment achieved by the regression. The relative residual, 1.4%, quantifies this overlap. Random errors in the underlying data seem to dominate the slight disagreement in the collected data points. Yet, each response function estimate shows clear over-smoothing relative to its underlying data. This discrepancy between the data trend and the estimated response functions increases the uncertainty due to random error in the regression; this is captured in the measurement uncertainty simulation (Section III-F).

We look for qualitative confidence in the robustness of the regression process for this data by examining its intermediate results. First, observe that the residual response with trial measurand values, given by Fig. 7b, is locally convex. We

TABLE V
NOISE MEASUREMENT UNCERTAINTY — WLAN CLIENT

Error Source	u_i	*Type*
Random & definitional errors in the regression (u_{MC})	0.07 dB	A
Excess noise attenuation at calibration	0.04 dB	A
Connector repeatability	0.035 dB	B
ENR of calibration noise diode	0.012 dB	A
DUT to test system mismatch loss	0.012 dB	A
Calibration reading variability	0.01 dB	A
Calibration reading mismatch loss	0.01 dB	A
Noise source temperature drift	0.01 dB	B
Combined: $u_c = \mathrm{RSS}(u_i) = 0.09\,\mathrm{dB}$	Expanded: $U = 2u_c = 0.18\,\mathrm{dB}$	

therefore expect that small error perturbations should still produce measurement values clustered around a central value. This is borne out by the symmetric and single-modal shape of the histogram of Monte Carlo results in Fig. 7c.

Measurement uncertainty: The uncertainty budget is listed in Table V, which gives ± 0.18 dB expanded uncertainty in $N|_{\mathrm{in}}$ to 95% confidence. The “Type A” uncertainties refer to errors modeled through statistics, while “Type B” uncertainties are estimated from datasheet information, following the standardized metrology terminology [39].

The dominant uncertainty term is the random and definitional error, u_{MC}, estimated by the Monte Carlo simulation technique from Section III. The accounting procedure according to the propagation of uncertainty method here is appropriate here, because the trials are normally distributed (Fig. 7c) [39].

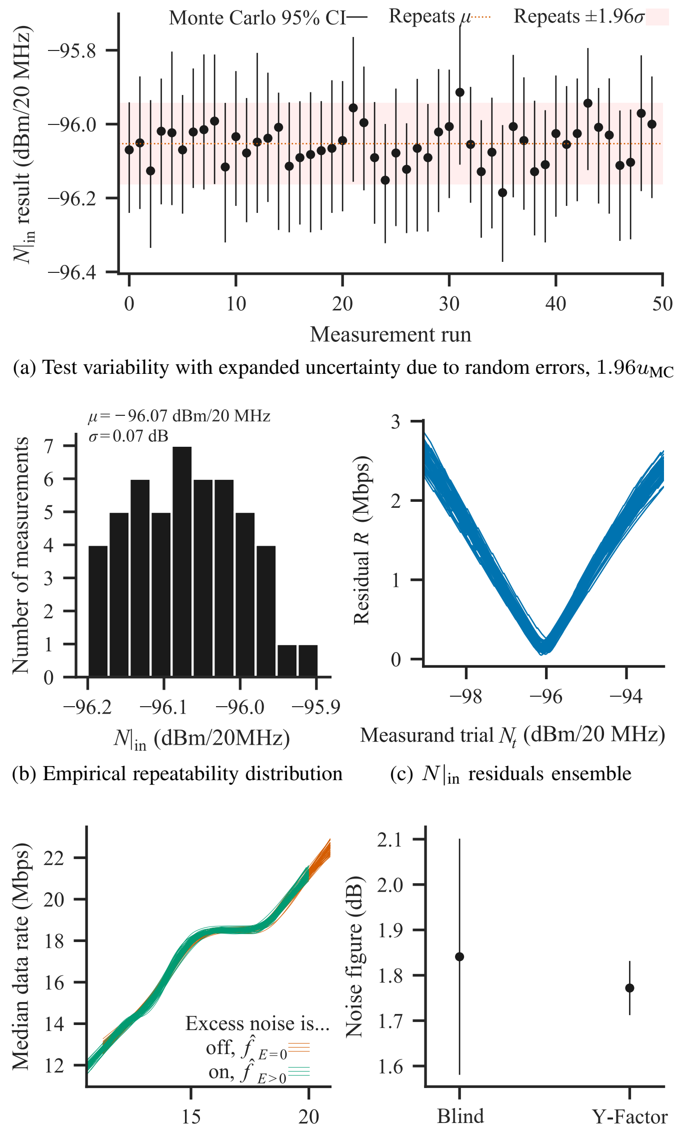
The remaining terms arise from physical errors calibrating E according Appendix C. This is the same type of analysis is detailed in, e.g., [4]. The RSS of these uncertainty components, $u_E = 0.06$ dB, is the lower bound of the standard uncertainty achievable through the acquisition and regression defined in Section III. Thus, no changes to the experimental parameters listed in Table IV can reduce the expanded uncertainty of the measurement below $2u_E = 0.12$ dB.

Each sensitivity coefficient is 1, so the i^{th} standard uncertainty, u_i, is akin to an estimated standard deviation of the error. The distribution of errors underlying each uncertainty term are assumed Gaussian.

Verification: Repeat measurements give insight into random variability within the test method. A summary of 50 test runs is shown by Fig. 8.

The repeated measurements help us validate the random variability predicted by the Monte Carlo uncertainty simulation. Each interval $N|_{\mathrm{in}} \pm 1.96 u_{\mathrm{MC}}$ is shown together with repeatability statistics in Fig. 8a. The shaded region indicates estimated 95% variability interval on repeat measurements, computed as ± 1.96 times the sample standard deviation of all measurement runs. The sample distribution of the repeat measurements is given by the histogram in Fig. 8b. Its lacks of dramatic outliers supports the assumption that u_{MC} encapsulates Gaussian-distributed errors. The empirical standard deviation in the $N|_{\mathrm{in}}$, 0.07 dB, is in this case equal to the Monte Carlo prediction, u_{MC}.

The regression residual ensemble (Fig. 8c) of these runs still shows a clear trend of curves that still point to a small cluster of minima near the measurement value. The estimated response ensemble in Fig. 8d show tight, overlapping bands



(a) Test variability with expanded uncertainty due to random errors, $1.96u_{\mathrm{MC}}$

(b) Empirical repeatability distribution

(c) $N|_{\mathrm{in}}$ residuals ensemble

(d) Response estimate ensemble

(e) Y-factor cross-validation

Fig. 8. WLAN client DUT measurement and analysis verification based on 50 repeat measurements and Y-factor measurements for cross-comparison.

for $E = 0$ (green) and $E > 0$ (orange) after alignment with the measurement value. These steps in the regression process seem robust to any imperfections in this underlying data.

To ensure proper accounting of any large systematic errors, we performed cross-comparison against Y-factor measurements. We measured a calibrated LNA cascaded with the DUT input, following Appendix E, which gave the results in Fig. 8e. The measured noise figures and corresponding uncertainty intervals on the measured noise figures overlap. Most of the uncertainty in the Y-factor measurements comes from impedance mismatch and the noise diode characterization, which are subsets of the uncertainties in the blind technique. Thus, in general, the Y-factor measurements should produce smaller uncertainties than our packaged device technique.

TABLE VI
GPS SIGNAL EXCITATION PARAMETERS

Center frequency	GPS L1, 1575.42 MHz
Satellites	11
Time	July 4, 2016 01:35:18 – 01:38:18 UTC
Location	N 31° 35.893636', 110° 16.670841' W 1352.3 m elevation
Signal codes	L1 C/A, L1C pilot, Pseudo Y, M-code
WAAS	2 augmentation signals
Satellite mask	5° elevation

VI. CASE STUDY: GPS L1 RECEIVER

Thermal noise is thought to dominate the total noise in global navigation satellite systems, so it receives close attention in receiver design. Recent concern over potential interference from cellular service proposed in adjacent bands has motivated further interest in the GPS L1 band, specifically. The application of interference test results requires ancillary noise performance data as part of a GPS link model [21]. Receivers packaged with built-in LNAs need to be assessed with a blind method like the one we proposed here.

We used noise measurements of a consumer off-the-shelf (COTS) GPS L1 receiver given in [22] with a preliminary version of the blind measurement. We reanalyze these data here with the new blind technique. We have now also released the data to the public [42].

A. Equipment Under Test

The DUT receiver was consumer equipment marketed for prototyping integration of GPS L1 capabilities. Its front-end was integrated with the rest of the electronics on a printed circuit board, with no access to physical signal outputs, so measurement was the only practical approach. The device outputs a data stream that includes an estimate of carrier-to-noise-density, C/N_0, as well as position, time, and details about the GPS satellite constellation.

B. Test Implementation

Measurements of the packaged GPS L1 receiver followed the configuration and procedures of [20]. Table VI summarizes the corresponding parameters of the GPS signal. This signal was emulated by a GPS test instrument. The measurement frequency of the blind noise measurement is determined by the excitation, 1575.42 MHz, which is also the frequency at which we calibrated attenuation and excess noise.

The original experiment differed from the techniques we developed in this paper in key ways:

- The acquisition on $E > 0$ was a hand-tuned truncated grid on (C, E) (in dB), instead of the procedure in Section III-B,
- many more samples were acquired at $E > 0$ than $E = 0$,
- the source of excess noise was a vector signal generator modulated with circular white noise,
- excess noise output E was calibrated against power and attenuation instead of calibrated noise, and
- the reference LNA for verification was characterized with a commercial noise diode based on ENR calibration data provided by the vendor.

TABLE VII
NOISE MEASUREMENT UNCERTAINTY — GPS L1 RECEIVER

Error Source	u_i	*Type*
Calibration of the spectrum analyzer reading	0.14 dB	A
Impedance mismatch	0.11 dB	B
Long term stability of the spectrum analyzer	0.10 dB	B
Random & definitional errors in the regression (u_{MC})	0.07 dB	A
Frequency response of the spectrum analyzer	0.05 dB	B
Connection repeatability	0.05 dB	A
Attenuation error at calibration	0.04 dB	A
Combined: $u_c = \mathrm{RSS}(u_i) = 0.23\,\mathrm{dB}$	Expanded: $U = 2u_c = 0.45\,\mathrm{dB}$	

The regression and uncertainty techniques of our packaged receiver measurement technique still apply to these data, despite these differences.

User data selection and time series acquisition: The output y under consideration here is the steady-state median of the DUT's self-estimate of carrier-to-noise-density, C/N_0, which was reduced to a scalar time-series by taking the median across all visible satellites at each time point. The user data increased almost monotonically across about 20 dB of input CNR.

As typical for this application, the self-estimated C/N_0 included phase noise [20, Appendix B], unlike the CNR defined by (4). This discrepancy does not impact the analysis or result, however, because the self-reported C/N_0 still responds as a *function* of CNR.

Experimental Parameters: The time-series of user data samples streamed at 20 samples/s for 180 s, producing a time series with $M_{\mathrm{s}} = 3600$ samples per sample in y. Sampling points below the illustrated range of CNR resulted in uninterpretable y outputs, so the user data response function estimate omits these data.

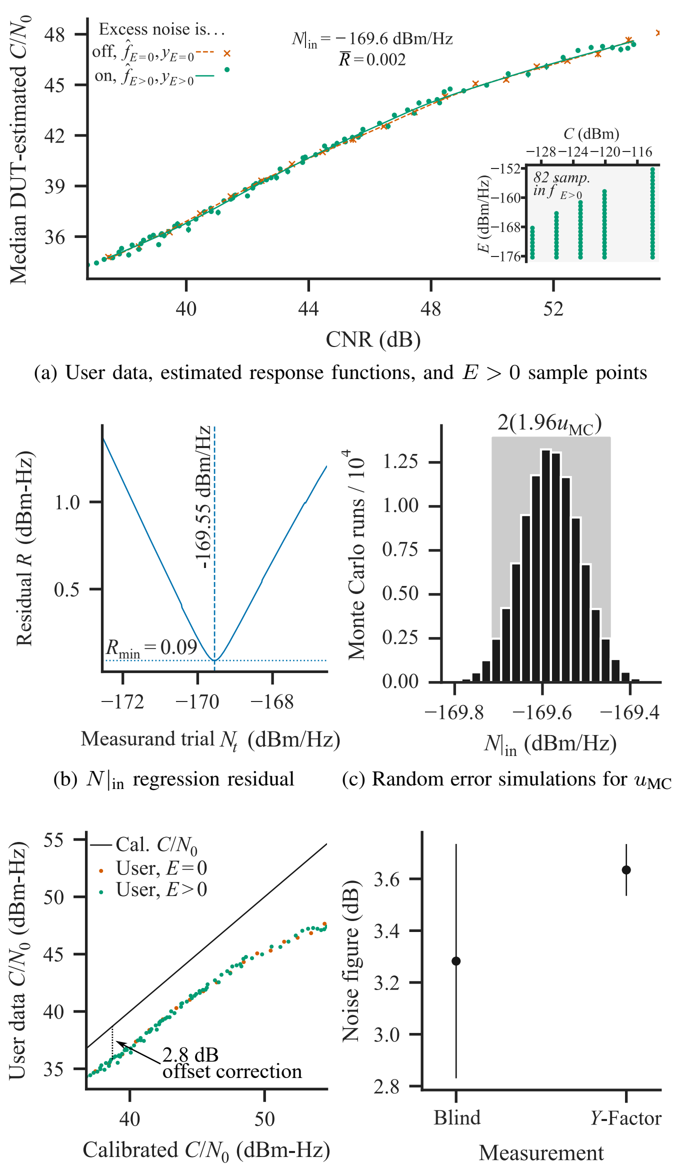
We chose bounds on C and E by manually tuning the attenuators to locate the domain that gave significant range of variation in y. Without the benefit of the new sampling techniques in Section III-B, we acquired many more sampling points in $y_{E>0}$ than $y_{E=0}$. As discussed in Section III, these extra samples did little to reduce u_{MC}. The resulting difference in the input resolution also poses an unknown risk of bias in the response function estimates. Thus, to clarify plots and reduce computation time, the reanalysis decimates the $E > 0$ data to the same number of sampling points as $E = 0$.

C. Results and discussion

The measured system noise of the GPS receiver was (-169.55 ± 0.45) dBm/Hz. The corresponding noise figure is (4.41 ± 0.45) dB.

Data and regression: Figure 9a overlays the user data and estimated response functions with and without excess noise. The CNR values shown are computed with the calibrated input levels and the measurement result. The estimated 95% confidence intervals on user data variability in each sampling point are too slim to be visible. The user data point subsets $E = 0$ and $E > 0$ also overlap too closely to separate by eye. It is not surprising that the alignment is much better than that of the WLAN data, with relative regression residual at 0.2%, confirming the assumed dependence on CNR.

Measurement uncertainty: This experiment demonstrates that the calibrations of physical noise may dominate the



(a) User data, estimated response functions, and $E > 0$ sample points

(b) $N|_{\text{in}}$ regression residual

(c) Random error simulations for u_{MC}

(d) User data calibration curves

(e) Y-factor cross-validation

Fig. 9. GPS L1 receiver DUT user data, regression, and validation, plus calibration of device-reported C/N_0.

measurement uncertainty. Uncertainty sources are listed in Table VII.

The calibration procedure here assumed flat frequency response in the spectrum analyzer. The grouped uncertainties that result, from [20, Table C.20], are dominated by this frequency response. The minimum combined standard uncertainty achievable with this test system is thus 0.22 dB. A future test could reduce this uncertainty by following the more accurate Dicke radiometer calibration of Appendix C.

The user data variability and the regression process, captured through Monte Carlo simulation in u_{MC}, is negligibly small compared to the physical calibration uncertainties. In a future measurement, this uncertainty could be reduced dramatically with the improved noise calibration technique of Appendix C.

Another application of these data is to calibrate the user data C/N_0 to a physical value at the input connector reference plane. We define the calibrated physical C/N_0 equal to the input CNR averaged over the L1 band allocation with $B = 1\,\text{Hz}$. The trend at low C/N_0 is a fixed scaling factor (losses, digitizer, manufacturing variability, etc.) relative to the calibrated C/N_0; the saturation of the user data at high C/N_0 is a common symptom of phase noise. Fig. 9d shows the calibrated and user C/N_0 data together. At the lowest C/N_0 levels, where phase noise contributions are small, the space between the curves suggests an offset correction of $+2.8\,\text{dB}$. Adding this to the user data converts to the physical value at the GPS receiver antenna connector. This number is then suitable to use in a link budget.

Verification: The Y-factor cross-comparison measurement, like the WLAN test, was a blind measurement of a calibrated LNA in cascade with the DUT. This followed the same procedure as in Appendix E.

The results of the cross-comparison against the Y-factor method are shown by Fig. 9e. The calibration of excess noise in this measurement was the same as in the bare DUT, and includes the same shortcomings. The blind measurement uncertainty here is therefore much larger than that of the Y-factor validation measurement. Still, the 95% confidence intervals overlap, validating the test method.

VII. Conclusion

We have proposed a blind method to measure the system noise in receive systems. We believe it is the first general-purpose technique for receivers and transceivers that output unknown functions of CNR.

The measurement requires an automated system to inject calibrated and programmable levels of signal attenuation and excess noise. The calibration for these levels is performed with typical laboratory measurement instruments, and could be supplied for a measurement system by an external party. These characteristics are frequency-dependent, but can otherwise be re-used to support different applications, standards, and protocols. The test execution also requires automation functions to acquire user data output; these could be implemented to support a specific DUT, or a broader industry standard. If a calibrated measurement transmitter is unavailable as a signal source for the DUT, a power sensor is also needed to characterize signal power incident on the DUT.

The case studies on consumer equipment achieved measurement uncertainties on the order of tenths of a dB. The WLAN client measurement result demonstrated that uncertainty contributions may be reasonably balanced among the traditional physical calibrations and new sources of regression uncertainty. The GPS receiver measurements showed that the regression technique can be sufficiently accurate that the physical level measurements may dominate the uncertainty of the measurement. An additional study on an WLAN AP, given in Appendix A, shows an example of the expected increase in estimated uncertainty that arises when the user data include outliers.

Further development of the technique could streamline the test execution automated selection of experimental parameters, study the regression performance with various different user data response functions, and consider further improvements to the robustness of the regression. Research could also consider support for dependence of user data on total power instead of CNR.

This basic test method opens new application opportunities. First, a figure of merit like relative residual may be a useful gauge for the effectiveness of automated control over a communication receiver. Focused attention on user data response functions may be useful on its own to develop metrology for key performance indicators. The extension of the new regression technique to over-the-air noise measurements could also be a useful benefit for integrated receive antenna systems, as in [22]. Application in the presence of interference may lead to a technique for blind measurement of receiver interference rejection, giving a more intuitive characterization of spectrum sharing impacts on incumbent receivers.

Appendix A
Case Study on a WLAN Access Point

We used the test configuration of Section V to measure the AP as the DUT. The changes needed for this test were to swap the client and the AP radio connections, send benchmark traffic to the AP instead of the client, and calibrate the signal power C from the client instead of the AP. This case illustrates the response of the regression procedure to a small number of outliers in the user data.

A. Test Implementation

The measurement followed the procedure of Section V, with an added 6 s delay to allow the DUT to adjust in each input sampling point. A shortcoming of this type of state reset approach is a lack of feedback that it has been effective at run-time. We are left to assess this with the regression residual after the test is complete.

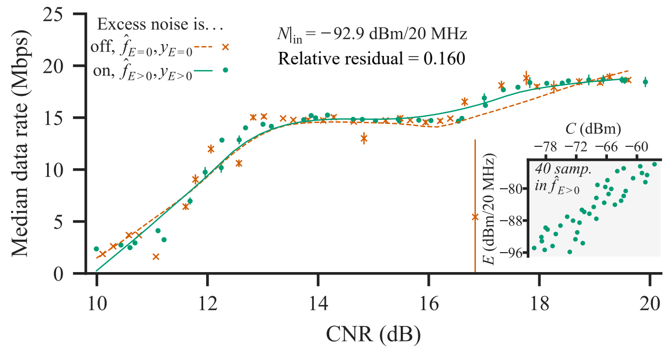
The blind measurement result for the WLAN AP was (-92.61 ± 0.32) dBm/20 MHz system noise, or, equivalently, (8.30 ± 0.32) dB noise figure. The results are presented in the same format here as for the client.

B. Results and Discussion

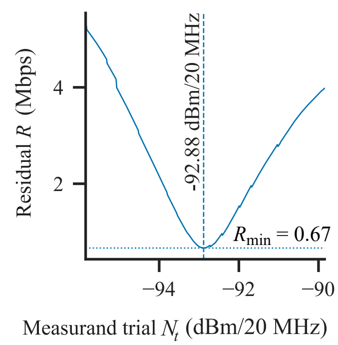
Data and regression: The AP DUT results draw attention to the impact of outliers in user data, shown in Figure 10a near CNR $=$ 17 dB for $E > 0$. This distorts the $E > 0$ response estimate. The relative residual is 16%, indicating looser alignment between $E > 0$ and $E = 0$ user data compared to the WLAN client results.

The impact of the outliers is visible in Fig. 10b as a wider spread in the minimum trough, which increases sensitivity to errors in input power. As a result, the histogram of Monte Carlo trials in Fig. 10c has greater spread than that of the WLAN client. Still, the histogram is unimodal and symmetric, so the outliers have not introduced unstable edge cases in the regression calculations.

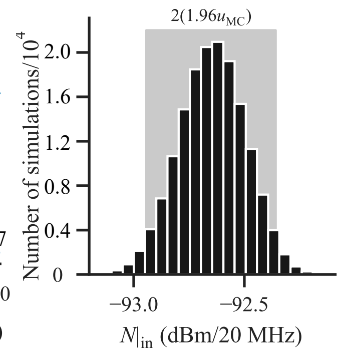
We believe the outliers in this test were caused by sending data before the link was ready. This type of synchronization problem might be corrected by increasing the wait time before acquisition or power cycling the AP between tests.



(a) $N|_{in}$ regression residual



(b) Cost function in $N|_{in}$ estimate



(c) Simulations of random errors

Fig. 10. Data, regression, and simulation of uncertainty from random error in the first WLAN AP measurement.

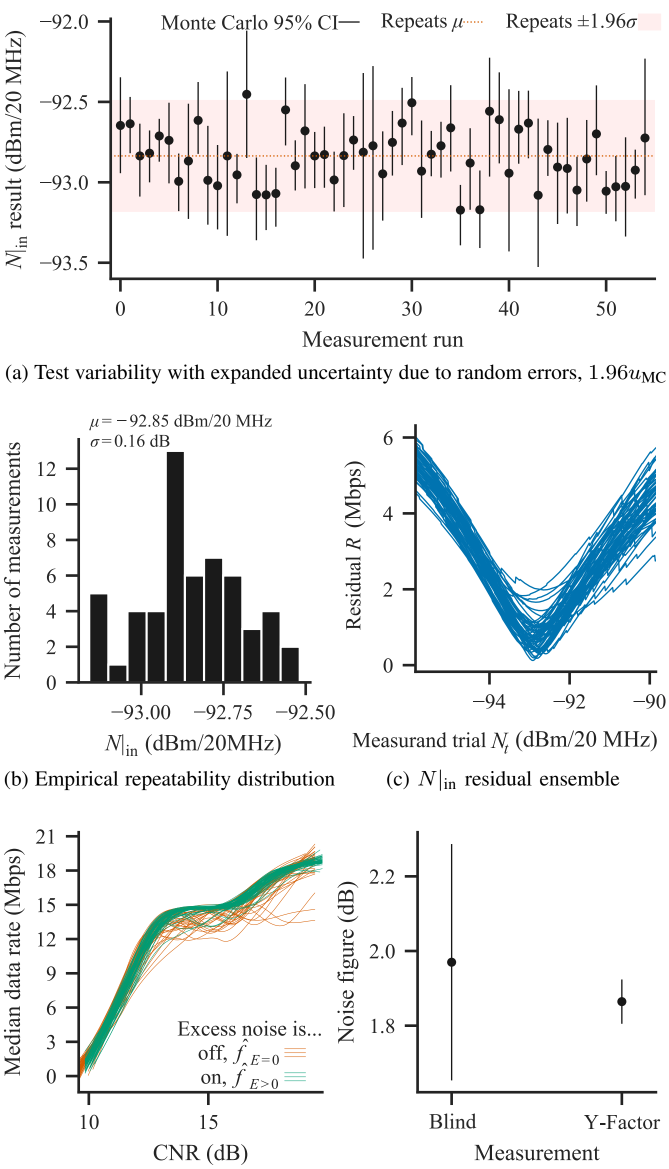
TABLE VIII
Noise Measurement Uncertainty — WLAN AP

Error Source	*Uncertainty*	*Type*
Random & definitional errors in the regression (u_{MC})	0.15 dB	A
Excess noise attenuation at calibration	0.04 dB	A
Connector repeatability	0.035 dB	B
ENR of calibration noise diode	0.012 dB	A
DUT to test system mismatch loss	0.012 dB	A
Calibration reading variability	0.01 dB	A
Calibration reading mismatch loss	0.01 dB	A
Noise source temperature drift	0.01 dB	B
Combined (RSS): $u_c = 0.16$ dB	Expanded: $U = 2u_c = 0.32$ dB	

The uncertainty budget in Table VIII breaks down the estimated expanded uncertainty, 0.32 dB. The dominant term, u_{MC}, made the uncertainty larger than that of the WLAN client. The remaining physical terms are the same as for the WLAN client, because the same calibration of E was still in use. Improving our control over the communication test might yield a total uncertainty approaching the minimum limit of 0.12 dB in expanded uncertainty.

Verification: Repeat measurements and Y-factor cross-validation present another opportunity to understand the impact of the user data outliers.

The outliers typically to appear in 1 to 2 random sampling points (similar to Fig. 10a) per measurement. They propagate into more variability in the measurement values and u_{MC} shown in Fig. 11a. The average u_{MC}, 0.15 dB, is still close to the sample standard deviation resulting from the repeat measurement runs, 0.16 dB (Fig. 11b), which indicates that



(a) Test variability with expanded uncertainty due to random errors, $1.96u_{\text{MC}}$

(b) Empirical repeatability distribution

(c) $N|_{\text{in}}$ residual ensemble

(d) Response estimate ensemble

(e) Y-factor cross-validation

Fig. 11. WLAN AP DUT measurement and analysis verification based on repeat measurements and Y-factor measurements for cross-comparison.

the Monte Carlo simulation tended to accurately characterize the measurement variability.

The ensembles in Figs. 11c,d illustrate the outlier impacts in more detail. Trends are still visible as clear bands, but some curves deviate. The trend toward lower-valued errors in the response function causes the slight shift in the minima of the residuals (and therefore the measurement result). The Monte Carlo simulations capture this effect in u_{MC} by randomizing the sign of the errors, at the expense of increased variability.

Results of Y-factor cross-comparison testing are given by Fig. 11e. The estimated 95% confidence intervals for the proposed blind noise and Y-factor methods overlap, verifying the test method and result for data rate user data on this DUT.

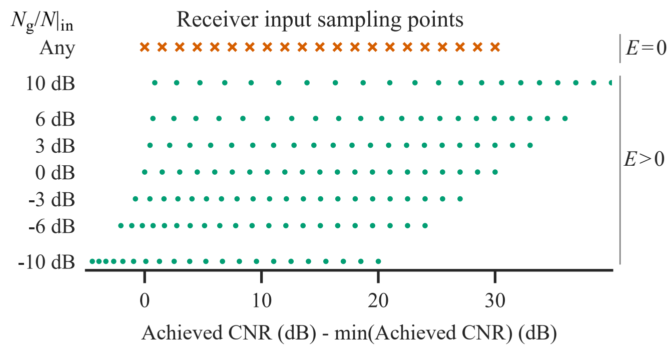


Fig. 12. Normalized location of actual input CNR samples (horizontal axis) given various levels error in the initial guess N_g (vertical axis). The $y_{E>0}$ samples (green dots) must be within the 0 dB to 30 dB range of $y_{E=0}$ (orange crosses) for use in the regression.

Appendix B
Sensitivity of Input Sampling Guesses

The accuracy needed for the initial guess for the measurement result, N_g, is not obvious. "Large" errors may place $E > 0$ sampling points outside the range of CNR supported by $E = 0$ data; these samples must be discarded during the regression. Lost test data make the regression less robust to biases in the function estimates $\hat{f}$, and increase the resulting uncertainty by reducing the statistical power of the test.

A comparison between the actual realized CNR and the goal CNR is shown by Fig. 12 for various levels of error in $N_g/N|_{\text{in}}$. The impact of these is illustrated by at various errors $N_g/N|_{\text{in}}$. Samples in $y_{E>0}$ outside the 30 dB span of $y_{E=0}$ must be discarded. The fortunate result here is that the measurement is forgiving even at the extremes $N_g/N|_{\text{in}} = \pm 10$ dB, which lead to keeping at least half of $y_{E>0}$ samples.

A measurement with significant error in N_g and large uncertainty might be improved by iterating. The new measurement takes the prior $N|_{\text{in}}$ as N_g.

Appendix C
Calibration Methods

This appendix details the calibration methods that we used for the measurement system of Section IV in the WLAN case studies (Section V and Appendix A). These general-purpose procedures also apply for other testbed systems and receiver applications.

Programmable attenuators: The programmable attenuators adjust the power levels in C and E. Attenuation level errors therefore contribute to the uncertainty in the input sample points, which propagate to uncertainty in the measurement value, $N|_{\text{in}}$.

Total attenuation through these devices (as well as the fully assembled measurement system) includes a fixed attenuation offset (in dB) plus a variable attenuation (in dB). We focus on the variable attenuation. This is the realized attenuation relative to the 0 dB attenuation setting, which varies with frequency. This relative attenuation includes deviations from the programmed attenuation setting (as large as around 2.5 dB on our devices). We leave the fixed attenuation to contribute

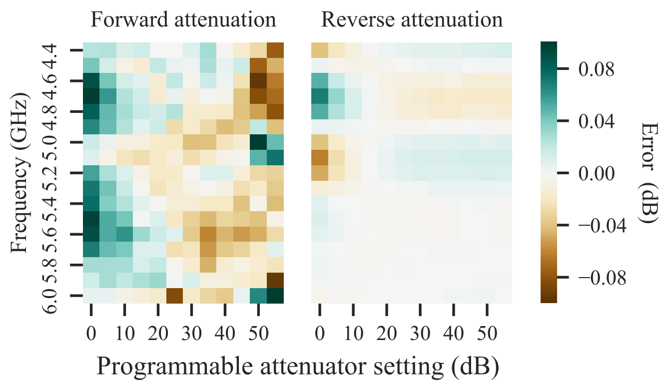


Fig. 13. Map of errors in directional signal (C) attenuation. Forward attenuation errors (0.04 dB RMSE) are relative to attenuation in the programmable attenuator; reverse attenuation errors (0.02 dB RMSE) are relative to constant attenuation.

to the total loss through the test system, which is calibrated in the next subsection.

The uncertainty in the variable attenuation characterizes the error in variable attenuation as a function of the frequency and programmed attenuation setting. This uncertainty in both C and E propagates into the measurand as random error in the input sampling point values that are used for regression. This uncertainty is an input to the Monte Carlo simulation that is used to perturb the input power. Uncertainty in attenuation of E also contributes to systematic error in the regression in the next section.

The calibration measurement is a 2-port S-parameter characterization of each attenuator. They need to be disconnected from the measurement system for this test, because the added loss in the test system is difficult to measure at higher attenuation settings. The measurement sweep covers all supported attenuation settings in 100 MHz frequency steps, and power and resolution bandwidth configured to achieve $|S_{21}|$ noise floor around -120 dB. Our network analyzer, chosen to maximize dynamic range, achieved this requirement at 1 Hz resolution bandwidth, and 8 dBm power output. We calibrate the attenuators to settings as high as 90 dB. We also test with added vector averaging to ensure that the 110 dB attenuation setting is accurate to within 1 dB, ensuring its effectiveness as an “off” switch.

We record the measured relative attenuation with the attenuation setting as a lookup table for use during test.

S-parameter measurements that verify the assembled test system are shown by Fig. 13. The residual attenuation errors result from mismatch between the attenuator blocks and the rest of the test system. We incorporate them into Monte Carlo simulations as random uncertainties.

Offset correction for excess noise power: The excess noise output calibration plane is the interface between the fully-assembled test system and the DUT. The calibration frequency should be the center frequency of the excitation signal (i.e., the signal with power level C). For a traceable noise reference, we use a connectorized noise diode calibrated against NIST primary standards, following [29]. The noise diode used to perform the calibration should ideally be specified with ENR of at least 15 dB (for strong detection on the spectrum analyzer), but less than the maximum testbed output power, $\max E$.

The calibration procedure for the excess noise output E follows that of a Dicke radiometer [44]:

1: connect the reference noise source to a spectrum analyzer and record the noise level P_{ref} that is integrated across the measurement band, $[f_L, f_H]$, in linear units
2: substitute the measurement system output in place of the reference noise source at the DUT reference plane
3: adjust the calibrated excess noise attenuation level until the spectrum analyzer reading matches that of the reference noise source; record it as P_{sys}

The residual imbalance, $10\log_{10}(P_{\text{ref}}/P_{\text{sys}})$, should be within 1 attenuation step. The maximum excess noise power is the calibration offset,

$$\begin{aligned} E_{\max}\,(\text{dBm}/B) =& 10\log_{10}(kT_0) + \text{ENR}\,(\text{dB}) \\ &+ \text{balance attenuation}\,(\text{dB}) \\ &+ \text{residual imbalance}\,(\text{dB}). \end{aligned} \tag{10}$$

The offset correction to determine the calibrated excess noise during operation of the testbed is then

$$E\,(\text{dBm}/B) = E_{\max}\,(\text{dBm}/B) - \text{Atten. in } E\,(\text{dB}), \tag{11}$$

with the calibrated attenuation value described in the previous section.

The correction applied by (10) and (11) makes the combined uncertainty in E (and therefore $N|_{\text{in}}$) dependent on errors in the reference noise diode calibration, attenuator calibration, and spectrum analyzer measurement. These need to be considered in the total uncertainty budget.

Offset correction for signal power: The signal power measurement helps to ensure the intended CNR test conditions at the receiver input. The concern here is the impact of random errors in the attenuation levels on C. We calibrated the offset in C by measuring signal power at the (non-DUT) transceiver output with a coupler and power sensor, and subtracting loss to the DUT.

Correction to measurement system physical temperature: We need to correct the noise figure from the physical noise temperature of the measurement system, T_1, to the reference temperature, T_0. Consider the the measured input noise powers characterized at T_1 as $N|_{\text{in}}^{T=T_1}$, and the idealized output produced with no input noise as $N|_{\text{in}}^{T=0}$. The additive contributions of input noise are

$$\begin{aligned} N|_{\text{in}}^{T=T_1} &= N|_{\text{in}}^{T=0} + kT_1B, \text{ and} \\ N|_{\text{in}} &= N|_{\text{in}}^{T=0} + kT_0B. \end{aligned}$$

Solving for $N|_{\text{in}}$ with (3) cancels $N|_{\text{in}}^{T=0}$, giving

$$\text{NF} = 10\log_{10}\left(\frac{N|_{\text{in}}^{T=T_1}}{kT_0B} + \frac{T_0 - T_1}{T_0}\right), \tag{12}$$

the corrected noise figure.

Appendix D
Measurand Sensitivity to Excess Noise Errors

The guide to the expression of uncertainty in measurement (GUM) [39] standardizes error propagation techniques that estimate the standard uncertainty of the measurand as a weighted RSS of many constituent uncertainty sources. The *sensitivity coefficient*, a component of each weighting coefficient, scales the uncertainty term by its impact on the measurand. The scaling term applied to the uncertainty term is the square of the sensitivity.

The sensitivity coefficient that we focus on here tracks errors in E through to the measurand. We demonstrate here that the sensitivity coefficient for E is -1 when the measurand is in dB.

Consider systematic errors that shift all calibrated values of E by a fixed offset, $10\log_{10}(a)$ dB, in the user data response,

$$y = f\left(10\log_{10}\left[\frac{C}{aE + N|_{\text{in}}}\right]\right) + n. \tag{13}$$

Now define a response function with offset (in dB), $g(\text{CNR}) = f(10\log_{10} a + \text{CNR})$. The user data response with the same error as (13) in terms of g is

$$y = g\left(10\log_{10}\left[\frac{C}{E + N|_{\text{in}}/a}\right]\right) + n. \tag{14}$$

Executing the experimental procedure in Section III with g (in place of f) shifts the CNR of the input sampling points by $-10\log_{10}(a)$ dB. The constant offset in dB in the argument of g cancels in the optimization by equation (7). Hence, the sensitivity coefficient on excess noise is -1, because an error in E of $10\log_{10}(a)$ dB produces the error in $N|_{\text{in}}$ of $-10\log_{10}(a)$ dB. The square of this coefficient is therefore 1.

Appendix E
Y-Factor Cross-Comparison Method

Verification to address systematic errors requires comparison against a traceable noise figure measurement. Our approach here is to compare two measurements of a reference LNA in cascade with the DUT input: (i) our proposed blind measurement and (ii) cascaded noise figure calculation based on calibrated two-port Y-factor measurement of the reference LNA.

The LNA gain and noise figure were measured on a commercial noise-figure meter with the Y-factor method. The noise figure of this cascaded system, NF_{CASC}, is approximately equal to the LNA Y-factor measurement result. More precisely, by [2],

$$\text{NF}_{\text{CASC}} = 10\log_{10}\left(F_{\text{LNA}} + \frac{F_{\text{DUT}} - 1}{G_{\text{LNA}}}.\right). \tag{15}$$

F_{DUT} and F_{LNA} are the DUT and LNA noise factors (noise figure in linear units), respectively; G_{LNA} is the available power gain of the LNA. This verification depends on the DUT noise figure measurand that is under verification ($10\log_{10} F_{\text{DUT}}$), but with sufficient LNA gain, only very weakly. The calculated noise figure becomes effectively a Y-factor measurement result that is nearly independent of our proposed blind measurement. The calibration values of these parameters are listed for each DUT in this paper by Table IX.

TABLE IX
Y-Factor Cross-Comparison Parameters

WLAN Client	$F_{\text{LNA}} = 1.72 \pm 0.06$ dB
	$G_{\text{LNA}} = 20.60 \pm 0.13$ dB
	$F_{\text{DUT}} = 4.84 \pm 0.22$ dB
GPS L1 Receiver	$F_{\text{LNA}} = 3.60 \pm 0.10$ dB
	$G_{\text{LNA}} = 19.90 \pm 0.13$ dB
	$F_{\text{DUT}} = 4.41 \pm 0.45$ dB
WLAN AP	$F_{\text{LNA}} = 1.72 \pm 0.06$ dB
	$G_{\text{LNA}} = 20.60 \pm 0.13$ dB
	$F_{\text{DUT}} = 8.30 \pm 0.34$ dB

The use of the LNA in measurements of a transceiver DUT *attenuates* the link in the reverse direction (from the DUT into the measurement system). This is acceptable as long as there is sufficient link margin.

We estimate cascaded uncertainty on NF_{CASC} by Monte Carlo analysis. Errors in F_{LNA}, F_{DUT}, and G_{LNA} are taken to follow a Gaussian distribution (truncated to positive values) with standard uncertainty equal to half of the stated uncertainty (expanded at 95% confidence). Each of these are treated as uncorrelated random variables, which we use to perturb (15) over 10^6 Monte Carlo trials. The 2.5% and 97.5% quantiles of the empirical distribution for these trials yield an approximate 95% uncertainty interval for NF_{CASC}.

Acknowledgments

This paper was greatly improved by suggestions and moral support from Michael Frey (with NIST); Billy Riddle, Bill Young, and Sheryl Genco (formerly with NIST); and our many helpful and diligent reviewers.

References

[1] D. M. Pozar, *Microwave Engineering*, 3rd ed. Hoboken, NJ: Wiley & Sons, 2005.

[2] H. Friis, “Noise Figures of Radio Receivers,” *Proc. IRE*, vol. 32, no. 7, pp. 419–422, Jul. 1944.

[3] H. C. Torrey and C. A. Whitmer, *Volume 15 of MIT Radiation Laboratory Series: Crystal Rectifiers*. New York: McGraw-Hill Book Company, 1948.

[4] M. G. Arthur, “The Measurement of Noise Performance Factors: A Metrology Guide,” U.S. National Bureau of Standards, Boulder, CO, Tech. Rep. 142, June 1974.

[5] S. Wedge and D. Rutledge, “Wave techniques for noise modeling and measurement,” *IEEE Trans. Microw. Theory Tech.*, vol. 40, no. 11, pp. 2004–2012, 1992.

[6] G. Gonzales, *Microwave Transistor Amplifiers*, 2nd ed. Pearson, 1996.

[7] “3GPP Technical Specification 36.101 Release 15 version 15.3.0,” 3GPP, Tech. Rep., Oct. 2018.

[8] “Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Amendment 4: Enhancements for Very High Throughput for Operation in Bands below 6 GHz,” IEEE, Tech. Rep. Std 802.11ac-2013, 2013.

[9] G. Koepke, W. Young, J. Ladbury, and J. Coder, “Complexities of Testing Interference and Coexistence of Wireless Systems in Critical Infrastructure,” NIST, Tech. Rep. 1885, July 2015.

[10] *ANSI C63.27-2017: American National Standard for Evaluation of Wireless Coexistence.* IEEE, 2017.

[11] “A Quick Introduction to Risk-Informed Interference Assessment,” FCC Technical Advisory Council, Tech. Rep., Apr. 2015.

[12] A. M. Voicu, L. Simic, and M. Petrova, “Coexistence of pico- and femto-cellular LTE-unlicensed with legacy indoor Wi-Fi deployments,” in *2015 IEEE International Conference on Communication Workshop, ICCW 2015*, 2015, pp. 2294–2300.

[13] Y. Ma, W. Young, E. Anderson, and J. Coder, "Probability of Coexistence of LTE-LAA and WLAN Systems Based on Delay Constraints," in *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, July 2018, pp. 1–9.

[14] A. Munari, N. Grosheva, L. Simić, and P. Mähönen, "Performance of radar and communication networks coexisting in shared spectrum bands," *arXiv preprint arXiv:1902.01359*, 2019.

[15] N. Golmie, N. Chevrollier, and O. Rebala, "Bluetooth and WLAN coexistence: challenges and solutions," *IEEE Wireless Communications*, vol. 10, no. 6, pp. 22–29, 2003.

[16] D. Yang, Y. Xu, and M. Gidlund, "Wireless coexistence between IEEE 802.11-and IEEE 802.15.4-based networks: A survey," *International Journal of Distributed Sensor Networks*, vol. 7, no. 1, p. 912152, 2011.

[17] S. Seidman and N. LaSorte, "An experimental method for evaluating wireless coexistence of a bluetooth medical device," *IEEE Electromagnetic Compatibility Magazine*, vol. 3, no. 3, pp. 49–54, 2014.

[18] W.-B. Yang, M. Souryal, and D. Griffith, "LTE uplink performance with interference from in-band device-to-device (D2D) communications," in *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, 2015, pp. 669–674.

[19] J. H. Reed, A. W. Clegg *et al.*, "On the co-existence of TD-LTE and radar over 3.5 GHz band: An experimental study," *IEEE Wireless Communications Letters*, vol. 5, no. 4, pp. 368–371, 2016.

[20] W. F. Young *et al.*, "LTE Impacts on GPS Final Report," NIST, Boulder, CO, Tech. Rep. 1952, Feb. 2017.

[21] D. G. Kuester, D. A. McGillivray, A. Wunderlich, and W. F. Young, "Interference tests at room temperature applied to deployed low-noise receivers," NIST, Gaithersburg, MD, Tech. Rep. 1971, Oct. 2017.

[22] D. G. Kuester, D. A. McGillivray *et al.*, "Black box measurement of system noise in integrated wireless receivers," in *IEEE-APS Conf. Antennas Propag. Wirel. Commun.* IEEE, Sep. 2017, pp. 209–212.

[23] "EMC Testing and Measurement Techniques - Radiated, Radio-Frequency, Electromagnetic Field Immunity Test," IEC, Geneva, Tech. Rep. 61000-4-3, 2010.

[24] "Specification of Hearing Aid Characteristics," ANSI, Melville, NY, Tech. Rep. 53.22-2003, 2003.

[25] G. L. Pearson, "Fluctuation Noise in Vacuum Tubes*," *Bell Syst. Tech. J.*, vol. 13, no. 4, pp. 634–653, Oct. 1934.

[26] R. C. Jones, "A Method of Describing the Detectivity of Photoconductive Cells," *Cit. Rev. Sci. Instruments*, vol. 24, 1953.

[27] G. G. Kretschmar and L. E. Schilberg, "Preparation and Photoconductive Properties of Cadmium Telluride Films," *J. Appl. Phys.*, vol. 28, no. 8, p. 865, 1957.

[28] W. W. Lindemann and R. K. Mueller, "Grain Boundary Photodetector," *J. Appl. Phys.*, vol. 29, p. 1770, 1958.

[29] W. C. Daywitt, "A Coaxial Noise Standard for the 1 GHz to 12.4 GHz Frequency Range," National Bureau of Standards, Tech. Rep. 1074, 1984.

[30] J. Randa, S. Member, E. Gerecht, D. Gu, S. Member, and R. L. Billinger, "Precision Measurement Method for Cryogenic Amplifier Noise Temperatures Below 5 K," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 3, 2006.

[31] D. C. Montgomery, *Design and Analysis of Experiments*, 8th ed. New York: John Wiley & Sons, 2012.

[32] K. Hoad, S. Robinson, and R. Davies, "Automating warm-up length estimation," *J. Oper. Res. Soc.*, vol. 61, no. 9, pp. 1389–1403, 2010.

[33] K. Hoad and S. Robinson, "Implementing MSER-5 in commercial simulation software and its wider implications," in *Proc. 2011 IEEE Winter Simul. Conf.*, 2011, pp. 495–503.

[34] P. Heidelberger and P. A. W. Lewis, "Quantile Estimation in Dependent Sequences," *Oper. Res.*, vol. 32, no. 1, pp. 185–209, 1984.

[35] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American statistical association*, vol. 74, no. 368, pp. 829–836, 1979.

[36] P. H. Kvam and B. Vidakovic, *Nonparametric Statistics with Applications to Science and Engineering*. New York: John Wiley & Sons, 2007.

[37] W. S. Cleveland, E. Grosse, and M.-J. Shyu, "Locally-weighted regression," https://www.netlib.org/a/dloess, 1992.

[38] "LOESS (aka LOWESS)," 2012. [Online]. Available: https://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm

[39] "Guide to the expression of uncertainty in measurement," Bureau International des Poids et Mesures, Sèvres, France, Tech. Rep. 100, Sept 2008.

[40] J. Berkson, "Are there two regressions?" *Journal of the American Statistical Association*, vol. 45, no. 250, pp. 164–180, June 1950.

[41] B. R. Gandhi, S. K. Agarwal, and S. Sundaram, "Directional Attenuators," *IETE J. Res.*, vol. 24, no. 5, pp. 210–212, May 1978.

[42] D. G. Kuester, A. Wunderlich, D. A. McGillivray, D. Gu, and A. K. Puls, "Noise Measurements of Packaged Wireless Transceivers (Public Data Resource)." [Online]. Available: https://datapub.nist.gov/od/id/mds2-2121

[43] J. Nagle, "Congestion Control in IP/TCP Internetworks," Internet Engineering Task Force, Tech. Rep., Jan 1984.

[44] R. H. Dicke, "The Measurement of Thermal Radiation at Microwave Frequencies," *Review of Scientific Instruments*, vol. 17, pp. 268–275, July 1946.

Daniel G. Kuester (S'11-M'12-SM'16) received the Ph.D. degree in electrical engineering from the University of Colorado in Boulder in 2012.

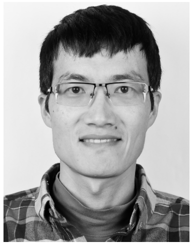
He is a Project Leader with the Communications Technology Laboratory, National Institute of Standards and Technology (NIST), Boulder, CO, USA. In the past, he worked as an RF Engineer, designing and integrating phased array systems and passive wireless sensors with FIRST RF Corporation (2013-2015) and Phase IV Engineering (2012-2013). His graduate work with advisor Zoya Popović (2007-2012) focused on ultra-lightweight wireless energy harvesting, in conjunction with research on passive backscatter communication at NIST.

His research focus is on robust metrology for spectrum and noise, spanning the full stacks of communications and radar systems. Dr. Kuester's work has been awarded the U.S. Department of Commerce Gold Medal (2017), Most Innovative Use of RFID by *RFID Journal* (2015), and Best Paper at the IEEE Conf. on Wireless Power Transfer (2013).

Adam Wunderlich received the B.S. degree in electrical engineering and the M.S. degree in theoretical and applied mechanics from the University of Illinois at Urbana–Champaign in 1999 and 2002, respectively, the M.S. degree in mathematics from Oregon State University in 2006, and the Ph.D. degree in electrical and computer engineering from the University of Utah in 2009. He is the data science lead for the National Advanced Spectrum and Communications Test Network (NASCTN), which is hosted by the Communications Technology Laboratory, National Institute of Standards and Technology, Boulder, CO, USA. In this role, Dr. Wunderlich leads experimental design and data analysis for test and measurement projects performed by NASCTN, and additionally carries out research in applied statistics and machine learning. He has co-authored over 50 technical publications and is a recipient of the U.S. Department of Commerce Gold Medal.

Duncan A. McGillivray was born in Muenster, Germany. He received his B.Sc. degree in physics from the State University of New York at Binghamton University, Vestal, NY, USA in 2008, and Ph.D. degree in electrical and computer engineering from the University of Virginia, Charlottesville VA, USA in 2015.

In 2016, he joined the Communications Technology Laboratory at the National Institute of Standards and Technology (NIST) as an Electronics Engineer in the Spectrum Sharing Metrology group. He led the group in an acting capacity between 2018 and 2019. As of 2019, Dr. McGillivray joined the National Advanced Spectrum and Communications Test Network (NASCTN) at NIST as Chief Engineer. He leads technical efforts of NASCTN projects and contributes to research activities in testing spectrum sharing, wireless coexistence, and interference susceptibility between telecommunications systems and DoD assets. Dr. McGillivray was a recipient of the Department of Commerce Gold Medal.



Dazhen Gu (S'01-M'08-SM'10) received the Ph.D. degree in electrical engineering from University of Massachusetts, Amherst, in 2007.

He has been with the RF Technology Division, National Institute of Standards and Technology, Boulder, CO, since November 2003. During the first three and a half years, he did his doctoral research in development of terahertz imaging components and systems. From 2007 to 2009, he was with the Microwave Measurement Services project, where he was involved in microwave metrology, in particular thermal noise measurements and instrumentation. From 2009 to 2015, he took a position in the microwave remote-sensing project, in which a microwave brightness-temperature standard was successfully demonstrated. From 2015 to 2018, he was in charge of the microwave power project and developed the NIST power traceability with correlated uncertainties for 5G communication researches. Since March 2018, he has been with the Shared-Spectrum Metrology Group, where he is involved in technical developments in the Spectrum Sensing and Noise Project.

Audrey K. Puls has a B.S. degree in electrical and computer engineering from the University of Colorado Boulder (2016) and a M. Eng. degree in electrical engineering from the University of Colorado at Boulder (2019).

She worked at the National Institute of Standards and Technology from 2013-2019 developing test methods for RF-based electronic safety equipment and collecting radiation pattern measurements using a reverberation chamber. In 2019 she joined Ball Aerospace, where she is presently a RF/Microwave Engineer. She has published several technical papers including *Antenna Radiation Pattern Measurements Using a Reverberation Chamber, Waveforms for Inference Testing of Emergency Responder Safety Devices*, and *Development of Laboratory Test Methods for RF-Based Electronic Safety Equipment: Guide to the NFPA 1982 Standard*.