KEITHLEY

WHITE PAPER

Understanding the Perils of Spectrum Analyzer Power Averaging

Steve Murray, Sr. Industry Consultant, Keithley Instruments, Inc.

Introduction

Averaging is a common technique for reducing the measurement uncertainty inherent in all measurements. Performing the same measurement a number of times and calculating the average of the measured values can often reduce the randomness of an experimental result. Many (if not most) instruments attempt to simplify the measurement process by performing averaging automatically. Rather than returning 100 noisy measurements, the instrument is responsible for taking all 100 measurements, calculating their average, and returning just the average. Averaging is so common and conceptually simple that one might assume there's little room for debate on the correct way to average. However, recent experience has demonstrated that power averaging in spectrum analyzers isn't necessarily straightforward. The following discussion explores the issues associated with power averaging in order to help readers avoid making the same mistaken assumptions the author did. The conclusions presented here are the results of an experiment that involved correlating the power measurements of two spectrum analyzers from different vendors. However, the issues discussed are generic in the sense that they apply to any spectrum analyzer power measurement with some form of post-detection averaging.

Keithley Instruments, Inc.
28775 Aurora Road
Cleveland, Ohio 44139
(440) 248-0400
Fax: (440) 248-6168
www.keithley.com

Incorrect Assumption #1: *To find the average power of a Zero-Span trace or a portion of the trace, average the RMS power.*

Averaging is so natural to engineers that it hardly seems to merit presenting the mathematical formula for calculating it. Nonetheless, to get everyone on the same page, let's refer to Eq. (1). M_{AVE} is the average of a series of individual measurements taken over N trials of an experiment, where each of those measurements is denoted as M_i:

$$M_{AVE} = \frac{1}{N}\sum_{i} M_i \tag{1}$$

In this instance, the task was to verify that instrument "A" correlated with instrument "B" to within some level of accuracy (say ±1dB). All measurements were performed in Zero-Span (ZS) mode. The fact that ZS was used is largely irrelevant to the problems with averaging; the same types of averaging issues occur in traditional frequency domain spectrum analysis. However, both vendors used the ZS technique for the measurements, in this case, Adjacent Channel Power Ratio (ACPR) measurements. This is typical of modern digital-IF analyzers, where the instrument performs multiple power measurements at varying offsets from the center frequency without re-tuning the analyzer. For those unfamiliar with ZS, it is a common spectrum analyzer technique for measuring power at a specific frequency. Put simply, ZS is a time-domain measurement that shows the variation of the signal's power envelope vs. time. In ZS mode, the analyzer is not sweeping frequency, but is instead tuned for a specific center frequency. The analyzer then measures the instantaneous detected voltage for a user-specified sweep time, and the equivalent power of this voltage "trace" is calculated and displayed vs. time. (In analog spectrum analyzers, the envelope of the signal is the output of the detector diode, while modern "digital-IF" spectrum analyzers digitize the baseband signal directly and calculate the envelope mathematically.)

Figure 1 shows a real ZS measurement of pulsed GSM signal. The blue curve represents the actual GSM pulse envelope. Note that the measurement performed here is the "Occupied RF Spectrum (ORFS) due to Modulation," which is simply an ACPR measurement. Note that the "squiggles" at the top of the burst are due to the resolution bandwidth and video bandwidth settings, both 30KHz per the GSM ORFS Mod specification. If these settings were widened, the trace would start to look much more like a rectangular pulse.

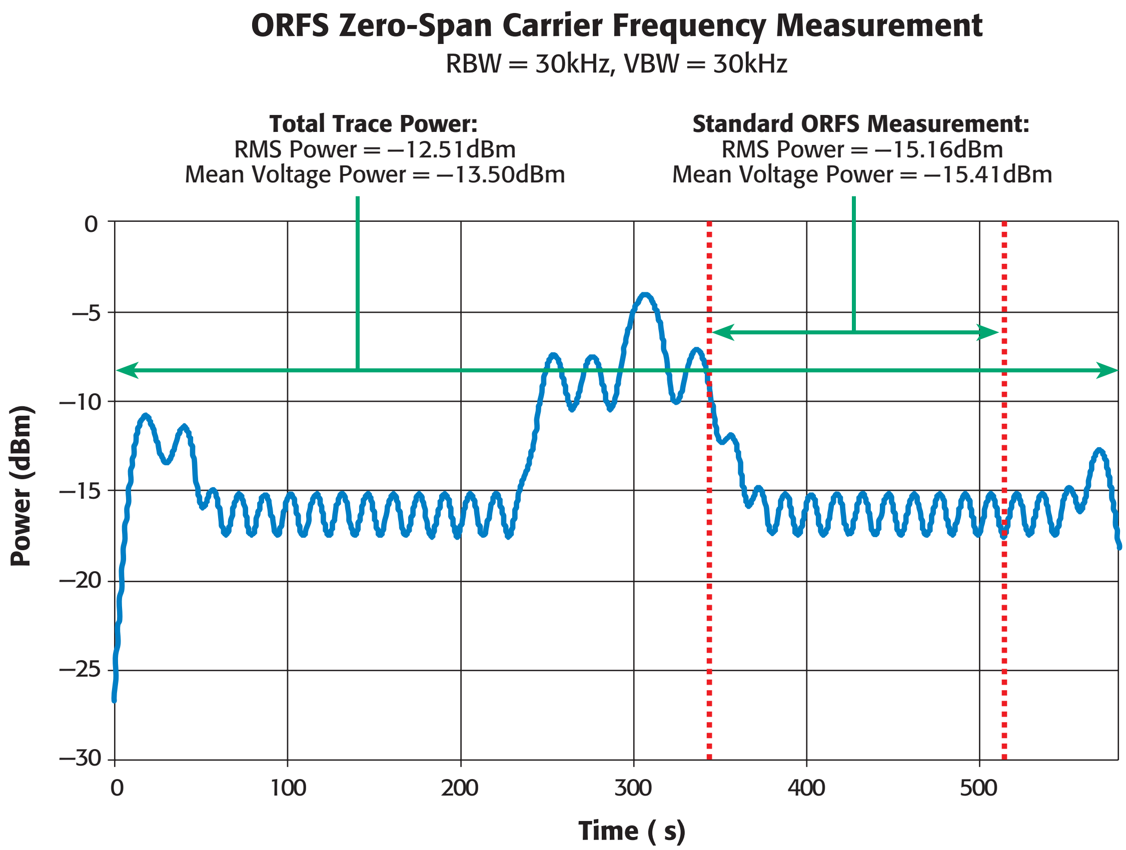

Figure 1. Zero-Span Trace

It's possible to calculate a number of useful results from the trace, such as the max peak power, min power, and average power. Finding the trace max power and min power is pretty straightforward, at least conceptually—simply have the analyzer do a max peak and min peak search on the entire trace and return the results. How can one find the average power between the dashed red lines? As a side note, the GSM ORFS Mod test requires that the average power be calculated over a limited portion of the burst—this is the region between the dashed lines.

The obvious way to calculate the average power is average across all points between the red lines. Eq. (2) accomplishes this, where N is the number of trace points between the red lines, and $P_{\text{ith point}}$ is power in the ith point.

$$P_{AVE} = \frac{1}{N}\sum_{i} P_{\text{ith point}} \tag{2}$$

Eq. (2) is completely intuitive. Moreover, it seems like the "correct" way to calculate power. Unfortunately, instrument manufacturers don't always agree. One of the instruments averaged powers as in Eq. (2), while the other instrument first converted each power point to a voltage, took the *average of all of these voltages*, then used the average voltage to calculate the average power. Eq. (3) shows the calculation.

$$P_{AVE}' = \frac{\left(\frac{1}{N}\sum_{i} V_{\text{ith point}}\right)^2}{50\Omega} \tag{3}$$

Proving that one instrument was using Eq. (2) and the other was using Eq. (3) was not a trivial exercise, because the difference between the two reported average powers wasn't that large. It was necessary to pull *multiple* traces out of both instruments and calculate the average every conceivable way until good fits were found. In the example in ***Figure 1***, the difference between the "true" average power (subsequently referred to as the RMS power) and the average *voltage* power is 0.25dB (RMS power is 0.25dB greater). Given that two different instruments were being compared, this could have been written off as a simple measurement difference (error) between the two instruments. While 0.25dB may not seem like much, when the requirement is for ~1dB of correlation (or just plain accuracy), 0.25dB becomes significant. This is particularly true at low signal levels, where the noise power becomes a significant portion of the total measured signal. Note that if the difference in powers over the whole burst is examined, the delta widens to ~1dB (again, RMS power higher than average voltage power). In this case, the difference is equal to the level of accuracy one is trying to obtain.

The average voltage power represents the "mean-square***d***" power [Eq. (3)], while the RMS power is, obviously, the "mean-square" power [Eq. (2)]. From elementary statistics, it can be shown that the mean-square minus the mean-squared is equal to the variance. What this implies, and what is probably obvious, is that the amplitude variation (amplitude variance) will directly contribute to the difference in reported powers. Finally, note that the mean-square power will *always* be greater than or equal to the mean-squared power (RMS power ≥ average voltage power).

Incorrect Assumption #2: *Average power is always calculated by averaging in Watts (linear).*

To continue with this example, assume that the average powers are themselves noisy. To remove some of the measurement noise, one may decide to apply an additional average: take multiple traces, compute each trace's average power, then average the powers across all traces (average of the averages). This is a common measurement requirement, particularly for low-level signals (in the case of the GSM ORFS Mod measurement, the standard dictates that the power results are to be averaged over 200 bursts). Eq. (4) shows the required calculation. To reiterate, each individual trace power ($P_{\text{Trace i}}$) is a *single number* calculated with Eq. (2) or Eq. (3) (either RMS power or average voltage power).

$$P_{\text{AVE}} = \frac{1}{N_{\text{Traces}}} \sum_{i} P_{\text{Trace i}} \tag{4}$$

It's reasonable to assume that the average will be computed with the $P_{\text{Trace i}}$ values in *Watts* (referred to as linear averaging). However, many analyzers offer the ability to average *logarithmically.* In this case, the "dBm"s are averaged. If, for example, given trace power averages of 1dBm and 3dBm, the linear average would be (1.25mW + 2mW) / 2 = 1.62mW = 2.11dBm. On the other hand, the log average would be (1dBm + 3dBm) / 2 = 2.0dBm. Log averaging the numbers introduces an error of 0.11dB.

In addition to the fact that averaging "dBm"s isn't really correct, there is a more subtle issue—for **repetitive** signals, linear and log averaging will produce the same result; thus, log averaging a repetitive signal introduces *no error.* Note that a repetitive signal is defined as a signal that has the same power vs. time trace for every sweep. The fact that a repetitive signal would give the same results regardless of the averaging type (linear or log) might not be intuitive, because taking the log of a number is a non-linear operation. However, it is trivial to demonstrate this fact. Starting with the equation for the log average, one has:

$$P_{\text{AVE, dBm}} = \frac{1}{N_{\text{Traces}}} \sum_{i} P_{\text{Trace i, dBm}} \tag{5}$$

The signal is repetitive, so $P_{\text{Trace i, dBm}}$ is the same value for all i. It's possible to drop the summation sign and rewrite Eq. (5) as:

$$\begin{aligned} P_{\text{AVE, dBm}} &= \left(\frac{1}{N}\right) N \cdot P_{\text{Trace, dBm}} \\ &= P_{\text{Trace, dBm}} \\ &= 10 \log\left(P_{\text{Trace, mW}}\right) \end{aligned} \tag{6}$$

So, regardless of the units used for P_{Trace} (mW or dBm), linear and log averaging will both produce the same result, provided the N values are identical. As a side note, it can also be shown that, in general, the log average will be equal to 10 times the log of the geometric mean of the linear trace powers:

$$\begin{aligned}
P_{\text{AVE, dBm}} &= \frac{1}{N_{\text{Traces}}} \sum_i P_{\text{Trace i, dBm}} \\
&= \frac{1}{N_{\text{Traces}}} \sum_i 10 \log\left(P_{\text{Trace i, mW}}\right) \\
&= \frac{1}{N_{\text{Traces}}} \left(10 \log\left(P_{\text{Trace 1, mW}}\right) + 10 \log\left(P_{\text{Trace 2, mW}}\right) + \ldots + 10 \log\left(P_{\text{Trace N, mW}}\right)\right) \\
&= \frac{1}{N_{\text{Traces}}} \left(10 \log\left(P_{\text{Trace 1, mW}} \cdot P_{\text{Trace 2, mW}} \cdot \ldots \cdot P_{\text{Trace N, mW}}\right)\right) \\
&= \frac{1}{N_{\text{Traces}}} \left(10 \log\left(\prod_i P_{\text{Trace i, mW}}\right)\right) \\
&= 10 \log\left(\left(\prod_i P_{\text{Trace i, mW}}\right)^{\frac{1}{N_{\text{Traces}}}}\right)
\end{aligned} \tag{7}$$

The fact that non-repetitive signals could produce different results is worth remembering, particularly because real-world operation conditions can differ from lab test conditions. Laboratory test signals are typically repetitive, given that they are often generated from an Arbitrary Waveform Generator (ARB). The ARB just plays back the same waveform over and over again, so it's repetitive by definition. Real-world signals are not, because they typically contain useful information that is changing in real time. Provided there isn't a large difference in the average power from trace to trace, the differences between log and linear averaging are small.

Let's now look at a real example that shows how a non-repetitive signal affects both the *per-trace* power average (either RMS or average voltage) and the power average taken across a number of traces (either linear or log). ***Figure 2*** is a plot of two traces of an EDGE burst with a constantly changing payload (pseudo-random sequence, PN15). The middle portion of the signal is repetitive; this is the GSM/EDGE Training Sequence, which is constant from burst to burst. However, the data portions on either side of the Training Sequence are changing from burst to burst. ***Table 1*** shows the results of calculating the RMS power (mean-square) and average voltage power (mean-squared) on 20 bursts, looking over the second data portion of the burst. Also, the linear and log average of all the bursts has been calculated.

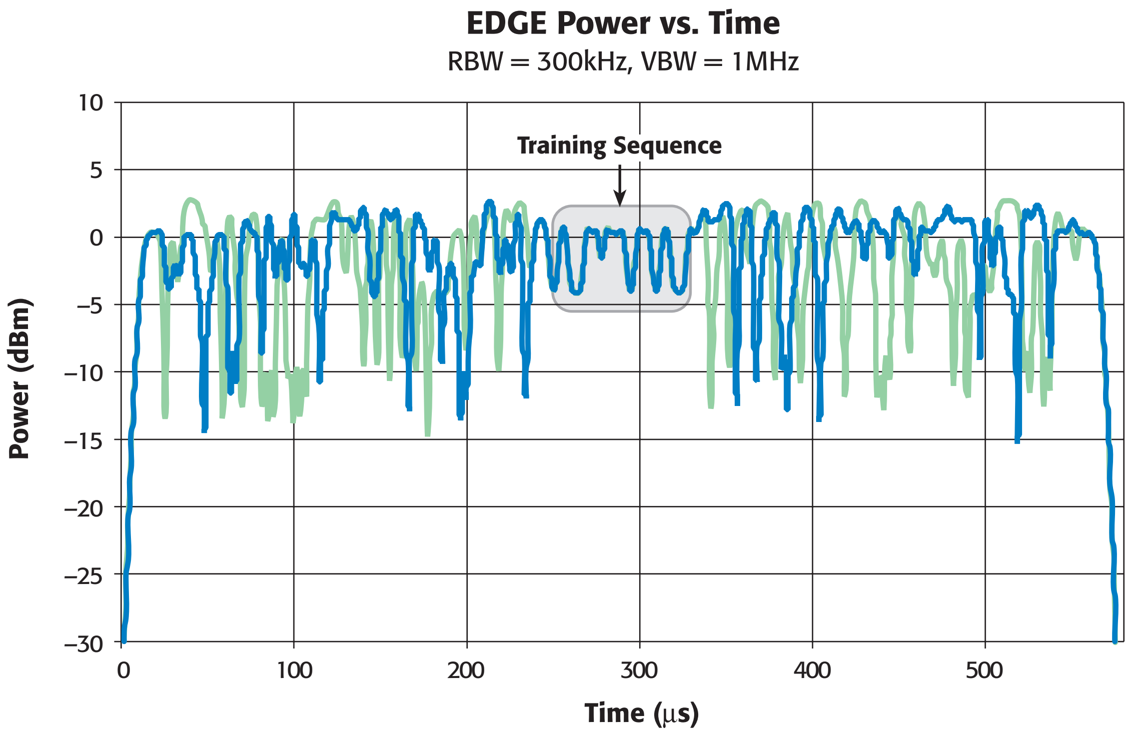

Figure 2. Non-Repetitive EDGE Signal

First, note that if one takes the "average of the average trace powers" for *either* RMS Power *or* Average Voltage Power, the difference (delta) between the linear and log average of all trace powers is very small for both RMS Power and Average Voltage Power (deltas of 0.02dB and 0.03dB respectively). That's because the average power values for all traces are reasonably close, in the sense that there isn't a large peak-to-peak swing across the averages. On the other hand, the differences between the RMS power and average voltage powers are significant if one looks at each trace *individually*—always at least 0.5dB, and approaching 0.75dB. What is more important, however, is that the difference is changing on a *trace-by-trace* basis. Again, the portion of the burst under examination is non-repetitive. The nice thing about repetitive signals is that, even though there will be a difference between RMS and voltage average power, the difference will be constant (if one were to measure power across the Training Sequence portion of the burst, this is exactly what would be found). For the current signal, however, the max to min difference is ~0.4dB. When one looks back at ***Figure 2***, the large deltas aren't very surprising. This signal in particular has ~10dB of amplitude swing, and the larger the amplitude swing is, the larger the difference between RMS and average voltage power will be. Incidentally, this is not a contrived "worst-case" signal. The

Table 1

Trace #	Trace RMS Power (dBm)	Trace RMS Power (mW)	Trace Average Voltage Power (dBm)	Trace Average Voltage Power (mW)	Trace RMS – Ave Voltage Delta (dB)
1	–1.01	0.793	–1.65	0.684	0.64
2	–0.11	0.975	–0.74	0.843	0.63
3	–0.54	0.883	–1.05	0.785	0.51
4	–0.37	0.918	–0.91	0.811	0.54
5	–1.10	0.776	–1.81	0.659	0.71
6	–0.31	0.931	–0.77	0.838	0.46
7	–0.98	0.798	–1.69	0.678	0.71
8	–0.25	0.944	–0.78	0.836	0.53
9	–0.41	0.910	–1.04	0.787	0.63
10	–0.20	1.047	–0.12	0.973	0.32
11	–1.52	0.705	–2.15	0.610	0.63
12	–1.31	0.740	–1.86	0.652	0.55
13	–0.12	0.973	–0.64	0.863	0.52
14	–0.97	0.800	–1.68	0.679	0.71
15	–1.11	0.774	–1.73	0.671	0.62
16	–1.13	0.771	–1.72	0.673	0.59
17	–0.70	0.851	–1.23	0.753	0.53
18	–0.98	0.798	–1.69	0.678	0.71
19	–0.51	0.889	–1.04	0.787	0.53
20	–0.12	0.973	–0.64	0.863	0.52
Linear Average of All RMS Trace Powers (dBm)	–0.64		Linear Average of All Average Voltage Trace Powers (dBm)	–1.21	
Log Aveerage of All RMS Trace Powes (dBm)	–0.67		Linear Average of All Average Voltage Trace Powers (dBm)	–1.25	
RMS Lin – Log Delta (dB)	**0.02**		**Average Voltage Lin – Log Delta (dB)**	**0.03**	

EDGE specification allows for even more amplitude swing, which will obviously increase the size of the difference.

Incorrect Assumption #3: *Trace averaging is always performed by calculating a single "summary" number for each trace, then averaging across these summary numbers.*

Up to this point, two forms of averaging have been discussed: single trace averaging, where all or a portion of a signal is averaged to come up with single number (RMS or average voltage power), and multiple trace averaging, where the results of performing single trace averaging on each trace are themselves averaged together (average of averages). There is another type of spectrum analyzer power averaging that should be addressed, which, for lack of a better term, can be called point-to-point averaging. Here, multiple traces are collected, and each trace point is averaged against the corresponding points in all other traces. F***igure 3*** is a graphical representation of how this works.

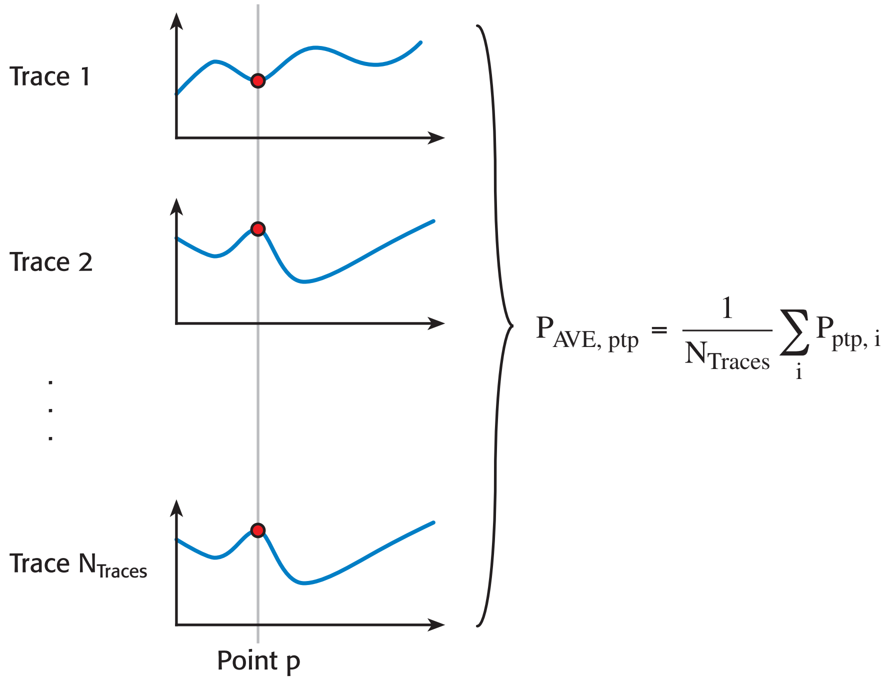

Figure 3. Point-to-Point Averaging

Again, each point is averaged with all of the points that occur at the same x value, resulting in an "average" trace. For this discussion, x will be time, but it could be frequency, and the same results will apply. As before, the points can be averaged either linearly or logarithmically. Once the averaging is complete, an additional average can be applied to the whole trace or part of it. If the waveform is repetitive, linear and log averaging will give the same average trace, because for each and every trace, a given point will have the same power. What happens when the waveform is not repetitive? ***Figure 4*** shows the average traces for both linear and log power averaging taken over 20 bursts of an EDGE signal with varying payload data. There is certainly a difference between the two traces, and it's obvious that the log averaged trace has less power than the linear averaged trace. ***Figure 5*** shows the *difference* between the two traces at every point. Note that, as expected, the Training Sequence portion of the burst shows no difference between linear and log averaging (again, the Training Sequence is repetitive from burst to burst).

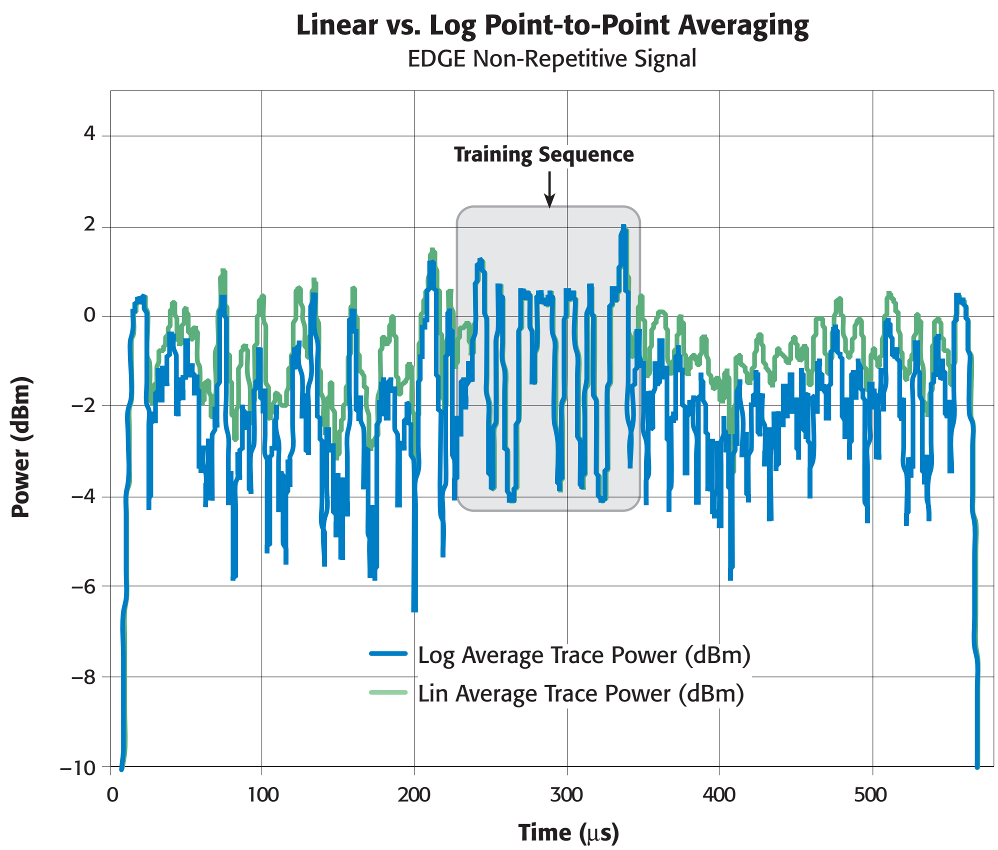

Figure 4. Linear vs. Log Point-to-point Averaging

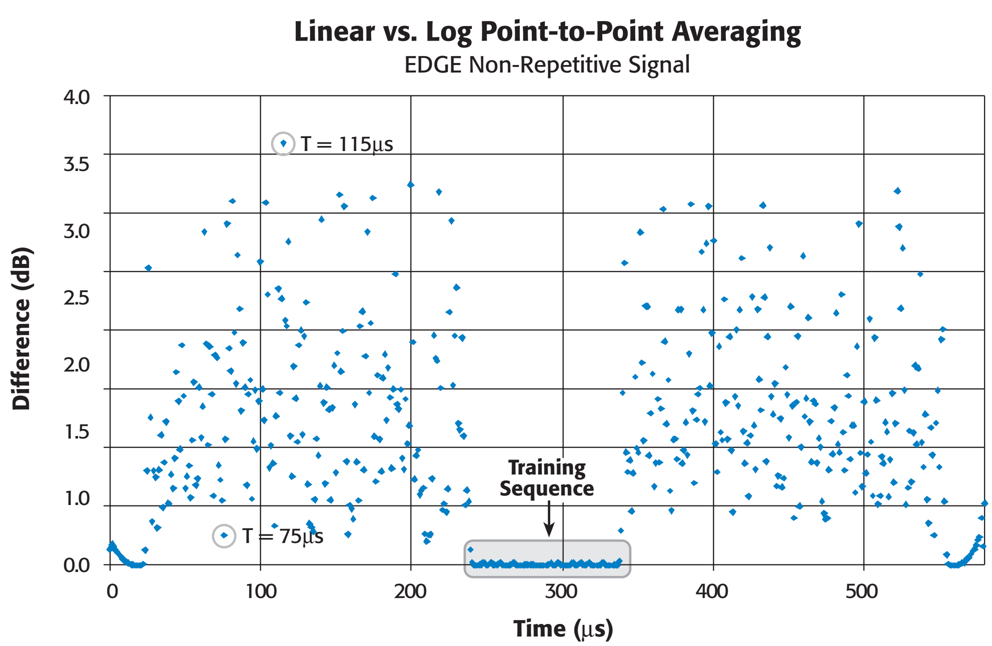

Figure 5. Linear vs. Log Delta

The difference arises from the way that log averaging exaggerates power swings. This is best illustrated by a simple example: assume one is measuring power at a specific point in time (or a specific frequency) over N bursts. The power is oscillating between two levels, for example, 0dBm and –10dBm; 50% of the power readings give 0dBm, and 50% give –10dBm. So, the peak-to-peak swing is, obviously, 10dB. What is the average power across the N bursts? Calculating the log answer is trivial: –5dBm. To calculate the linear average, one converts 0dBm and -10dBm to Watts, finds the average, and then converts this number back into dBm units. The average power in Watts is 0.55mW, or –2.6dBm. Using log averaging introduces an error of 2.4dB.

To generalize the calculation, it's known that an *x*dB change is equal to a change of $10^{(x/10)}$ in linear power. Therefore, it's possible to write the following equation, again assuming that 50% of the points are at one level M_{hi}, and the other are ΔdB down from that level:

$$\begin{aligned} P_{LinAve,\,dBm} &= 10\log\left(\frac{M_{hi}+M_{hi}\cdot 10^{-\Delta/10}}{2}\right) \\ &= 10\log\left(\frac{M_{hi}\left(1+10^{-\Delta/10}\right)}{2}\right) \\ &= 10\log\left(M_{hi}\left(\frac{1+10^{-\Delta/10}}{2}\right)\right) \\ &= 10\log\left(M_{hi}\right)+10\log\left(\frac{1+10^{-\Delta/10}}{2}\right) \end{aligned} \tag{8}$$

Note that, as Δ goes to infinity, $10\log\left(\frac{1+10^{-\Delta/10}}{2}\right)$ goes to –3dB. This means that, in the case of equal numbers of two power different levels, the resulting average linear power will be *at most* 3dB less than the higher power. It's possible to further generalize the result for an arbitrary ratio:

$$P_{LinAve,\,dBm} = 10\log\left(M_{hi}\right) + 10\log\left[r + (1-r)\cdot 10^{-\Delta/10}\right] \tag{9}$$

In Eq. (9), r is the ratio of the number of occurrences of the higher power (M_{hi}) to the total number of measurements (correspondingly, 1 – r is the ratio of the number of occurrences of the lower power to the total number of measurements). Note that when Δ goes to infinity, the resulting average power will be *at most* less than the higher power.

It's also possible to write the equation for the log average as:

$$\begin{aligned} P_{\text{LogAve, dBm}} &= 10 \log (M_{hi}) \cdot r + (1 - r)(10 \log (M_{hi}) - \Delta) \\ &= 10 \log (M_{hi}) + \Delta(r - 1) \end{aligned} \tag{10}$$

If Eq. (10) is subtracted from Eq. (9), the result is an expression for the difference between linear averaging and log averaging (this is the error introduced by log averaging):

$$P_{\text{Lin–Log, dB}} = 10 \log [r + (1 - r) \cdot 10^{-\Delta/10}] - \Delta(r - 1) \tag{11}$$

Figure 6 plots Eq. (11) vs. Δ for various values of r (Eq. 11). The plot was limited to a Δ of 20dB because this is likely to be at the upper end of common Crest Factors. (Crest Factor is often referred to as the "peak-to-average" power ratio.)

Figure 6. Maximum Log Average Error vs. Power Swing

As a sanity check, it would be helpful to look at a few points in some real data (refer back to ***Figure 5***). Here, two points in time are highlighted, one with a relatively large power difference (more than 3.5dB at T = 115µs) and the other with a much smaller difference (~0.25dB at T = 75µs). From the previous discussion, it would be reasonable to expect the corresponding power vs. time plot for those points to look considerably different; the point with the high error should show quite a bit of power swing, while the low error point should show a smaller power swing. This is, in fact, the case, as seen in ***Figure 7***. Here, the trace corresponding to the point at T = 115µs has ~15dB max amplitude swing, while the trace for the point at T = 75µs has ~5dB of swing. If one assumes that the high and low values occur equally (i.e., r = 0.5), then the trace for T = 115µs should have a maximum error of ~4.5dB, and the trace for T = 75µs should have a maximum error of ~0.5dB (see ***Figure 6*** and Eq. [11]). These values are greater than the measured 3.5dB and 0.25dB, but it's important to recall that the plot in ***Figure 6*** shows worst-case numbers (it assumes just two power levels, with equal numbers of each). One would expect the error to be smaller, since it's obvious there are more than two values (i.e., there isn't just a "high power" and a "low power").

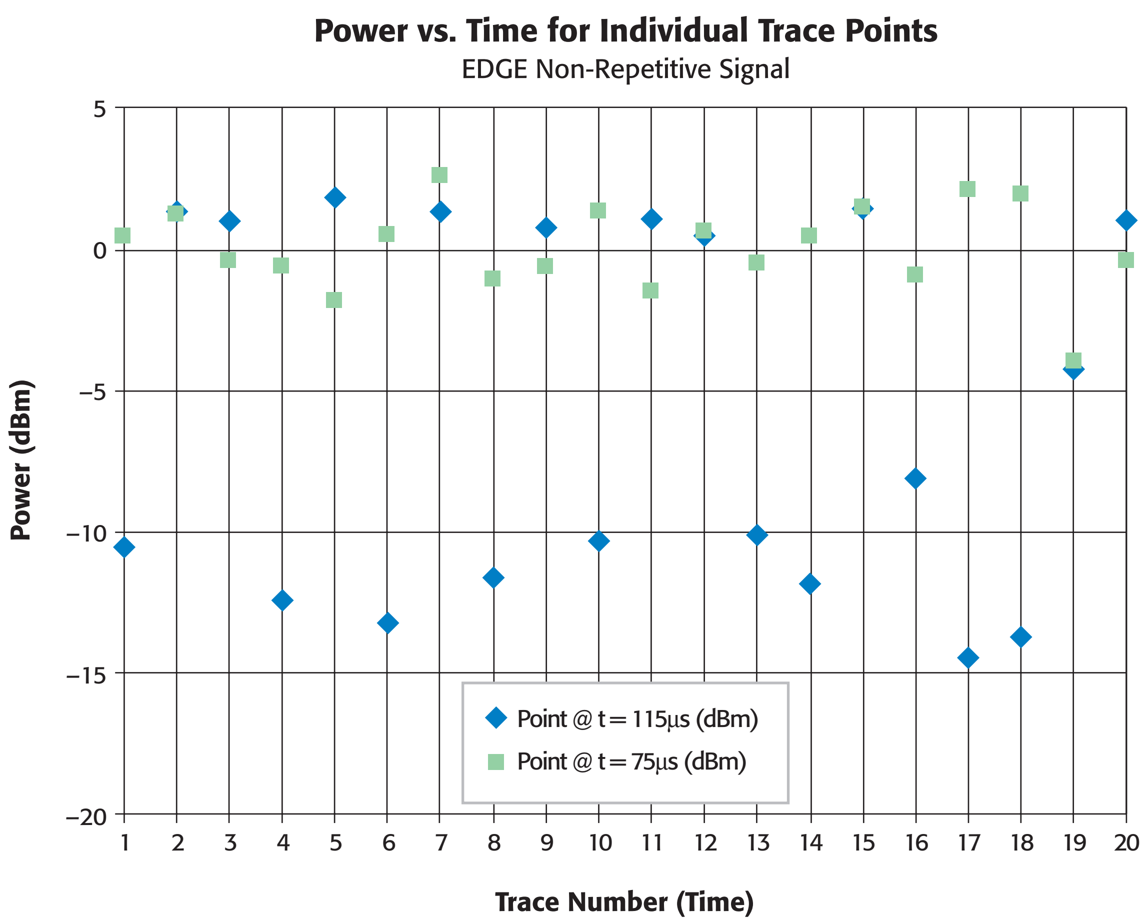

Figure 7. Power vs. Time for Individual Trace Points

Conclusion

In summary, engineers should keep in mind that spectrum analyzers don't always adhere to the "correct" way of calculating average power. Furthermore, the size of the potential errors introduced depends on the characteristics of the signal being analyzed. In particular, remember that it's important to:

- Understand the way the spectrum analyzer is calculating average power: RMS, voltage average, etc.
- Be aware that power isn't always averaged in linear units (Watts). Log averaging is also a possibility (averaging the "dBm"s).
- Repetitive signals can be misleading. The result may be either a static error (error is always the same and constant, for example, RMS vs. average voltage) or no error (linear vs. log averaging). Likewise, non-repetitive (or real-world) signals will have time-varying errors that depend on the signal swing.

As this discussion has illustrated, differences in averaging techniques can certainly lead to 1.0dB or more of error. The best way to understand how a particular spectrum analyzer calculates power averages is to pull a few traces out of the box and determine if manual calculations produce the same results as the analyzer does. While this can be a little tedious, it's well worth the effort if the application has reasonably tight accuracy requirements.

KEITHLEY

A GREATER MEASURE OF CONFIDENCE

KEITHLEY INSTRUMENTS, INC. ■ 28775 AURORA ROAD ■ CLEVELAND, OHIO 44139-1891 ■ 440-248-0400 ■ Fax: 440-248-6168 ■ 1-888-KEITHLEY ■ www.keithley.com

BELGIUM
Sint-Pieters-Leeuw
Ph: 02-363 00 40
Fax: 02-363 00 64
www.keithley.nl

CHINA
Beijing
Ph: 8610-82255010
Fax: 8610-82255018
www.keithley.com.cn

FINLAND
Espoo
Ph: 09-88171661
Fax: 09-88171662
www.keithley.com

FRANCE
Saint-Aubin
Ph: 01-64 53 20 20
Fax: 01-60-11-77-26
www.keithley.fr

GERMANY
Germering
Ph: 089-84 93 07-40
Fax: 089-84 93 07-34
www.keithley.de

UNITED KINGDOM
Theale
Ph: 0118-929 75 00
Fax: 0118-929 75 19
www.keithley.co.uk

INDIA
Bangalore
Ph: 080 22 12 80-27/28/29
Fax: 080 22 12 80 05
www.keithley.com

ITALY
Milano
Ph: 02-553842.1
Fax: 02-55384228
www.keithley.it

JAPAN
Tokyo
Ph: 81-3-5733-7555
Fax: 81-3-5733-7556
www.keithley.jp

KOREA
Seoul
Ph: 82-2-574-7778
Fax: 82-2-574-7838
www.keithley.co.kr

MALAYSIA
Kuala Lumpur
Ph: 60-3-4041-0899
Fax: 60-3-4042-0899
www.keithley.com

NETHERLANDS
Gorinchem
Ph: 0183-63 53 33
Fax: 0183-63 08 21
www.keithley.nl

SINGAPORE
Singapore
Ph: 65-6747-9077
Fax: 65-6747-2991
www.keithley.com.sg

SWEDEN
Solna
Ph: 08-50 90 46 00
Fax: 08-655 26 10
www.keithley.com

TAIWAN
Hsinchu
Ph: 886-3-572-9077
Fax: 886-3-572-9031
www.keithley.com.tw

Printed in the U.S.A.

No. 2696

07065KGW