Spectrum Monitoring Network: Tradeoffs, Results, and Future Directions

Peter Mathys MATHYS@COLORADO.EDU
University of Colorado Boulder, Department of ECEE, UCB 425, Boulder, CO 80309-0425 USA

Abstract

Due to its high cost, there is limited use of professional grade test equipment in the field to monitor, characterize, and identify spectrum usage. To supplement existing high-accuracy measurements there is therefore a large interest in using lower-cost software defined radio (SDR) receivers to complement or replace the high-cost spectrum analyzers in RF (radio frequency) measurement systems. Doing this allows measurements that are (1) continuous or at least much more frequent, (2) more densely located (in frequency and location), and (3) remotely controllable, and it even allows for additional measurements not practical using current solutions such as simultaneous measurements of the same signal at multiple locations. However, there is an inherent trade-off between price and performance and there may be a complex up-front characterization procedure necessary to ensure the measurement accuracy.

1. Introduction

Radio spectrum suitable for wireless communications is a finite resource that has traditionally been allocated to users in a static long-term fashion. The demand for wireless services is expected to grow exponentially and it is widely seen as essential to enable increased and more dynamic exploitation of radio spectrum. A full characterization of the wireless environment is a monumental task that requires coordinated data collection, organization, and analysis efforts. Effective spectrum monitoring requires low cost programmable sensing hardware, secure and robust networking infrastructure, and meaningful data analysis and visualization. The conventional approach has been to use a truck, with an antenna, a professional grade spectrum analyzer, and other equipment, that is stationed in one location for some time and then moves on to the next location. The high cost of such equipment precludes its use for widespread (in terms of time, frequency and location) measurements. The emergence of low(er) cost software defined radios (SDRs) in the past decade, together with efforts to standardize spectrum characterization and occupancy sensing (SCOS) (IEEE, 2018) is an enabling technology for a much more comprehensive characterization of the radio spectrum. In this paper we discuss some of the tradeoffs between (expensive) factory calibrated professional grade spectrum anlyzers and user built and calibrated more modular approaches using (inexpensive) SDRs.

Proceedings of the 9^{th} GNU Radio Conference, Copyright 2019 by the author(s).

2. Radio Receiver Architectures

A spectrum analyzer is the same thing as a radio receiver. The main differences are the (usually) larger frequency range, the way the received signal is presented as power versus frequency, and the additional controls for gain and filtering.

2.1. Frequency Translation

An important signal processing block for communications is the frequency translation block with input $r(t)$ and output $w(t)$, consisting of a multiplier or mixer, followed by a filter, as shown in Fig. 1.

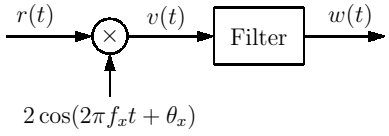


Figure 1. Mixer with Filter for Frequency Translation

Let $r(t) = s(t)\,\cos(2\pi f_c t + \theta_c)$ be a (bandlimited) radio signal. Then

$$\begin{aligned} v(t) &= 2\,s(t)\,\cos(2\pi f_c t + \theta_c)\,\cos(2\pi f_x t + \theta_x) \\ &= s(t)\,[\cos(2\pi(f_c - f_x)t + \theta_c - \theta_x) + \\ &\qquad + \cos(2\pi(f_c + f_x)t + \theta_c + \theta_x)]\,. \end{aligned} \tag{1}$$

The filter passes one of the two terms in the last equality of

eq. 1 and rejects the other and therefore $w(t)$ is either

$$w(t) = s(t)\ \cos(2\pi(f_c - f_x)t + \theta_c - \theta_x)\ , \qquad (2)$$

or

$$w(t) = s(t)\ \cos(2\pi(f_c + f_x)t + \theta_c + \theta_x)\ . \qquad (3)$$

For radio receivers the more common case is eq. 2.

2.2. Superheterodyne Receiver

The superheterodyne receiver was introduced in 1918 and its invention is generally attributed to Edwin Howard Armstrong even though the patent was later assigned to Lucien Lévy (Douglas, Nov. 1990). In the early days of radio one of the main goals besides achieving frequency selectivity was to produce enough amplification to drive an envelope detector for demodulation of AM signals with transmitted carrier (Armstrong, Feb. 1921). Modern implementations of the superheterodyne receiver generally consist of three filter stages and distribute the overall amplification between the RF frontend where a wideband low noise amplifier is used (LNA) and the intermediate frequency (IF) amplifier as shown in Fig. 2.

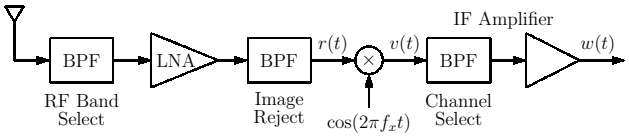


Figure 2. Superheterodyne Receiver

Assuming a received (bandlimited) radio signal $r(t) = \gamma\, s(t)\ \cos(2\pi f_c t)$ with attenuation γ and a local oscillator $\cos(2\pi f_x t)$, $v(t)$ becomes

$$v(t) = \frac{\gamma}{2}\, s(t)\, [\cos(2\pi(f_c - f_x)t) + \cos(2\pi f_c + f_x)t)]. \quad (4)$$

The third bandpass filter (BPF) in Fig. 2 passes signals with IF frequency (non-zero and typically much greater than zero)

$$f_{IF} = |f_c - f_x| \qquad (5)$$

The output $w(t)$ is then equal to

$$w(t) = \rho\, s(t)\ \cos(2\pi f_{IF} t)\ , \qquad (6)$$

where ρ is a proportionality factor. Because of the absolute value operation in eq. 5, there are two solutions for f_c in $r(t)$ for given f_{IF} and f_x

$$\begin{aligned} f_{c1} - f_x = f_{IF} \quad &\Rightarrow \quad f_{c1} = f_x + f_{IF}\ , \\ f_x - f_{c2} = f_{IF} \quad &\Rightarrow \quad f_{c2} = f_x - f_{IF}\ . \end{aligned} \qquad (7)$$

One of the two frequencies f_{c1} and f_{c2} is the desired receive frequency and the other one is called **image frequency** and needs to be rejected by the second BPF in Fig. 2 (before the mixer). Depending on the value of f_{IF}, this BPF may need to be tunable or switchable if f_x is variable. Alternatively, the bulky and expensive image reject BPF can be replaced by using electronic means and complex-valued signal operations which are more conducive to silicon integration. The principle of such image-reject receivers was introduced in 1928 by Hartley (Hartley, Apr. 1928) and in a different form by Weaver (Weaver, 1956) in 1956.

A prevailing concern in a superheterodyne receiver is image rejection. This is primarily done by the second BPF in Fig. 2 but it can be shared with the first BPF whose main function is to prevent overloading of the LNA by strong out-of-band signals. However, the second BPF is needed even if the image is rejected by the first BPF, because otherwise the image noise from the LNA will effectively double the noise figure of the receiver. Choosing a higher f_{IF} will relax the requirements on the cutoff frequency of the image reject filter, but it makes it harder to obtain the desired selectivity in the IF filter. Superheterdyne receivers with two or even three IFs can be used for more flexibility with filter selectivity. But generally, the demanding requirements on the IF and image reject filters make superheterodynes unsuitable for system on a chip (SoC) integration.

2.3. Direct Conversion Receiver

In the terminology of a superheterodyne receiver a direct conversion (also called homodyne or synchrodyne) receiver (Mashhour et al., June 1, 2001) uses an intermediate frequency of $f_{IF} = 0$, i.e., an incoming radio signal is directly mixed down to baseband. The original homodyne circuit was published by Colebrook (Colebrook, 1924) in 1924. It was an oscillating detector that oscillates at the same frequency as a received (transmitted carrier AM signal) and whose main goal was to amplify a received signal so that it could be passed on to an envelope detector, similar to the original intent of Armstrong's superheterodyne. The modern form of a direct conversion receiver (DCR) with in-phase (I) and quadrature (Q) channel outputs is shown in Fig. 3. An early version of a DCR, intended to separate overlapping sidebands of AM signals, appears in a patent by Gabrilovitch in 1936 (Gabrilovitch, 1936).

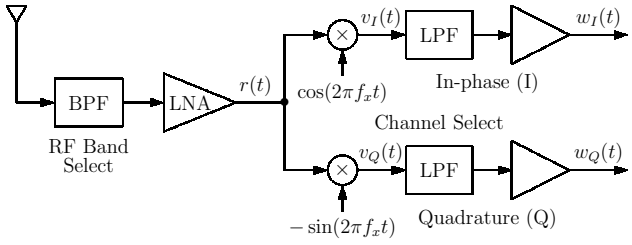


Figure 3. Direct Conversion Receiver

In essence, a DCR makes use of the frequency translation property of the Fourier transform where

$$\int_{-\infty}^{\infty} X(f+f_c)e^{j2\pi ft}df = \int_{-\infty}^{\infty} X(\nu)e^{j2\pi\nu t}d\nu\, e^{-j2\pi f_c t} = x(t)e^{-j2\pi f_c t}\,. \tag{8}$$

This leads to the entirely complex-valued baseband form of the DCR shown in Fig. 4.

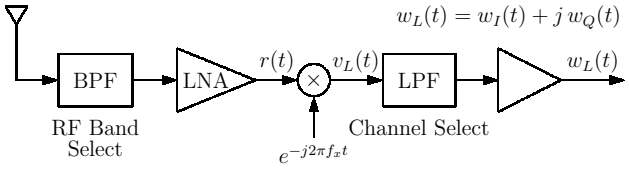


Figure 4. Complex-Valued Form of Direct Conversion Receiver

In a DCR the problem of the image frequency is eliminated since $f_{IF} = 0$. With the exception of the RF band select BPF, filtering now only occurs at baseband and can to a lage extent be done using digital signal processing (DSP). But DCRs suffer from other problems like dc offset, IQ imbalance, and flicker noise. DC components are introduced in a variety of ways, e.g., from the local oscillator coupling into the mixers' input ports, or from any stage that exhibits even order nonlinearity (since $\cos^{2n} 2\pi f_c t = [0.5(1+\cos 4\pi f_c t)]^n$). Imbalance between the I and the Q paths, including a phase shift between the two mixers that deviates from $90°$, leads to crosstalk between the I and Q channels which affects signal constellations of digital communication schemes. Low frequency noise, in particular flicker or $1/f$ noise of MOS (metal oxide semiconductor) devices, is a major baseband noise contributor. One more concern is the emission of radio signals by the DCR, going backwards through the mixer and LNA due to their non-ideal reverse isolation. Such signals are inside the receive band and are not attenuated by the RF band select filter.

2.4. Further Considerations

Both superheterodyne and DCR receivers have amplifier and mixer stages which produce non-linear responses when overdriven. IM3 products (see Section 7) affect both receiver architectures whereas IM2 products affect mostly the DCR. Phase noise of the local oscillators are another source of unwanted signal degradation. Most modern implementations of superheterodyne and DCR receivers convert the IF or baseband signal to the digital domain using sampling and quantization in an analog to digital converter (ADC). Filtering in front of the ADC is important to avoid aliasing, and often the dynamic range of such receivers is determined by the number of bits in the ADC. In a well designed system, overload of the ADC is the primary cause of non-linear receiver behavior for in-band signals.

3. Spectrum Analyzers

3.1. History

The basic function of a spectrum analyzer is to display signal power versus frequency. Early spectrum analyzers in the late 1950's were built by wiring together a radio receiver, an electronic method for sweeping the receive frequency, a bandpass filter, a power detector, and an oscilloscope to display the output. By 1960 a few companies combined these elements into expensive and bulky boxes and a few years later the first portable spectrum analyzers came out. These early *swept-tuned analyzer* instruments were entirely analog and to achieve a constant analysis bandwidth over a large tunable range it was natural to use the superheterodyne principle. With the publication of the historic paper on the computation of the fast Fourier transform (FFT) by Cooley and Tukey in 1965 (Cooley & Tukey, 1965), *FFT analyzers*, which replace sweeping by DFT/FFT computation, became possible. In 1967 the first such instrument, housed in two 6-ft racks could process a 1024 point FFT in one second (Deery, Jan. 2007).

3.2. Laboratory Grade Spectrum Analyzers

There are two broad categories of spectrum analyzers, distinguished by the methods used to obtain the spectrum of a signal, swept-tuned analyzers and FFT-based analyzers. Real-time spectrum analyzers display power versus frequency at all frequency components simultaneously. Swept-tuned spectrum analyzers measure power versus frequency sequentially as the center frequency is tuned over a selected range.

Modern swept-tuned analyzers are primarily superheterodyne receivers that are used to measure steady-state or repetitive signals because they cannot evaluate all frequencies simultaneously. They can not provide phase information, only magnitude information. Swept-tuned analyzers

often use dual or triple conversion superhets and they can cover wide frequency ranges from 10's of Hz to over 300 GHz with large dynamic range, defined as the difference between the highest and the lowest power signals that can be simultaneously measured.

Real-time spectrum analyzers measure all frequency bands in a span simultaneously. Parallel filter analyzers are used for audio applications and the vibration analysis of mechanical systems. The term "real-time" is derived from the simulation of physical systems where 100% coverage of all events must be captured. Parallel filter analyzers have large dynamic range and can measure transient and time varying signals.

FFT analyzers are the digital counterpart to parallel filter anlyzers. They sample and quantize the time-domain signal and use the FFT to convert it to the frequency domain. FFT spectrum analyzers can measure both magnitude and phase and they can easily switch between time and frequency domains. Limitations of FFT analyzers are mostly coming from the sampling and ADC operations. To avoid aliasing, a sampler requires a precise analog filter front end. The quantization of the ADC limits the dynamic range, a 14-bit ADC has 8192 amplitude levels, but if a minimum of 8 amplitude levels is needed for meaningful processing, then the dynamic range is only about $20 \log_{10} 1000 = 60$ dB (not including other limiting factors such as noise floor and intermodulation distortion).

Vector signal anlyzers (VSA) combine the advantages of superheterodyne technology with with high speed ADCs to offer fast high-resolution spectrum measurements with instantaneous bandwidths of up to several hundred MHz. VSAs are capable of characterizing complex signals such as burst, transient, or modulated signals used in communications, video, and radar systems. A VSA implements superheterodyne technology in a different fashion by replacing the analog IF by a digital IF section which incorporates digital signal processing. A significant characteristic of the VSA is that it is designed to measure and process complex-valued baseband data. The term vector in VSA comes from the ability to compute magnitude and phase from the complex-valued data.

A common characteristic of most modern spectrum analyzers is a front end with a calibrated attenuator and selective filtering that is to a large extent needed to suppress the image frequency of the superheterodyne receiver. Usually there are several filters present to switch between different bands and deal with out of band interfering signals. The price range for laboratory grade spectrum analyzers is from several 10,000 to more than 100,000 dollars.

3.3. Software-Defined Radio Spectrum Analyzers

The distinguishing feature of most SDRs is the DCR architecture and a minimum of filtering at RF frequencies. In fact, to keep the cost of SDRs down, the preferred way of implementation is in the form of a SoC. Generally SDRs are only calibrated for frequency but not for amplitude or power measurements. The higher end SDRs incorporate a FPGA (field programmable gate array) for onboard signal processing, but much of the signal processing and all of the display functions take place in a computer external to the SDR. Since SDRs convert RF signals down to complex baseband and digitize the I and Q signals, they can perform the same operations as a VSA when connected to a computer with DSP software, albeit typically with less bandwidth and less ADC resolution.

4. Performance Measurements and Results

4.1. Measurement Setup and Tools

One immediate consequence of using the lower-priced SDRs instead of professional grade spectrum analyzers is the lack of calibration and availability of detailed specifications. During the design stage of a spectrum monitoring network different SDRs need to be evaluated in terms of several performance criteria (Wepman et al., Aug. 2015) over the frequency range of interest to determine their suitability for the task. Some of the tests conducted at this stage are the displayed average noise level (DANL) test to determine the noise floor, Y-factor test to determine the noise figure, the self-generated spurious response test, the in-band signal overload test (1-dB compression point), second and third order intercept test points (IP2 and IP3) and noise power ratio test to determine intermodulation distortion (IMD), frequency and amplitude stability and accuracy test, phase noise measurement, and adjacent band signal desensitization and blocking (Meyer & Wong, Aug. 1995) test. Once a specific type of SDR is selected, each individual SDR needs to be calibrated at least for gain versus frequency, and preferably for DANL, self-generated spurious responses, and 1-dB compression points over the intended frequency range. After deployment of an SDR as part of a sensor, periodic calibration tests can be performed using calibrated noise sources such as noise diodes. A test setup that can be used in the lab for measuring SDRs is shown in Fig. 5.

Much of the measurements can be automated using suitable software scripts, but depending on the frequency range and the types of tests performed, the production test for each individual SDR can still take several hours.

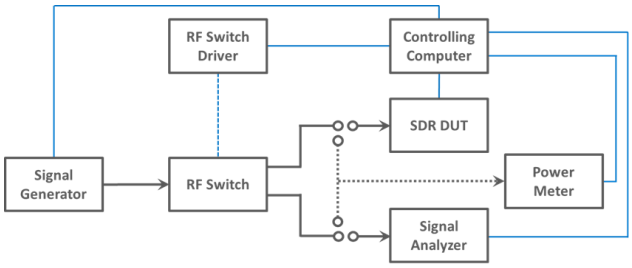


Figure 5. Test Setup for SDR Performance Measurements

4.2. Middle-Grade SDR

SDRs in this class cost about $1000. The SDR that we tested and selected for spectrum monitoring performs excellent once it has been calibrated over the frequency range of 100 MHz to 6 GHz. The maximum dynamic range is about 60 dB. The main deficiency is a wide open frontend that is very susceptible to desensitization and blocking by strong out of band signals.

4.3. Economy-Grade SDR

An SDR in this class that we tested cost about $30. It has a more narrow bandwidth (2 MHz) and a much more restricted frequency range (100 - 1700 MHz). Given the price it performs surprisingly well after calibration, but it has less longterm stability due to the lack of thermal stabilization.

5. Spectrum Characterization and Occupancy Sensing (SCOS)

The Institute of Telecommunication Sciences (ITS) in Boulder which is the research and engineering arm of NTIA (National Telecommunications and Information Administration) has released a first reference implementation of a sensor-control operating platform (ITS, 2018) proposed as part of the IEEE 802.22.3 SCOS standard (IEEE, 2018). A high-level view of the SCOS software architecture is shown in Fig. 6

The software is designed to remove some common hurdles encountered when remotely deploying sensors. It is hardware agnostic and assumes different hardware will be used depending on sensing requirements. A web-based interface gives sensor owners control over sensor tasking. At the same time, the SCOS software has also been designed with security in mind so that unauthorized users cannot access internal sensor functionality or intercept data. Data is collected using the sigMF signal metadata format (gnuradio, 2017) specification which allows for a detailed annotation of when, where, and how the collected data was taken. Currently, several sensors in Boulder and on the CU Boulder campus are active.

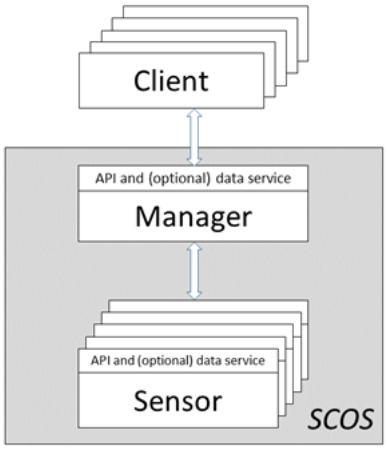


Figure 6. SCOS Architecture

6. Future Directions

Trading off sensor precision in favor of sensor density in space, time, and frequency has some interesting challenges as well as potential rewards. A dense and possibly mobile sensor network can exploit existing sources of transmission (e.g., from cell towers) to make real-time RF propagation measurements and improve the dynamic usage of RF spectrum. Machine learning can be used to learn the imperfections of the SDRs used as sensors and reinforcement learning can be used to update existing propagation models from the collected real-time data. Our next steps are to integrate geolocation and to produce cheaper and smaller sensors that can be deployed on mobile platforms such as public transportation vehicles. We are also interested in using machine learning algorithms to identify anomalies in network traffic and to locate intentional and unintentional RF intruders.

7. Appendix: Measurement Terms

7.1. Gain Calibration

One important element that is missing from most, if not all general purpose SDRs is gain calibration. RF signal strength is typically measured in dBm (power in decibels relative to 1 mW) across a resistance R matched to the source of the power (an antenna, a preamplifier, a signal generator, etc). Thus, for a power of P in watts, $P_{dBm} = 10\log_{10}(1000\,P)$ and, conversely, P is obtained from P_{dBm} as $P = (10^{P_{dBm}/10})/1000$ W. Laboratory RF signal generators are generally calibrated in P_{dBm}, but most radio receivers are fundamentally voltage amplifiers and the digital output of an SDR is a number that is propor-

tional to the voltage at the antenna input. To relate average power P or P_{dBm} to rms (root mean square) voltage V_{rms} across a matched load R for continuous wave (CW, sinusoidal) signals, the following formula is used

$$P = \frac{10^{P_{dBm}/10}}{1000} = \frac{V^2}{R} \,. \tag{9}$$

Solving for V_{rms} yields $V_{rms} = \sqrt{P\,R}$ volts, where $R = 50\,\Omega$ for most RF applications. The following table shows the power (in mW) and the RF rms voltage (in mV) for a load resistance of $R = 50\,\Omega$

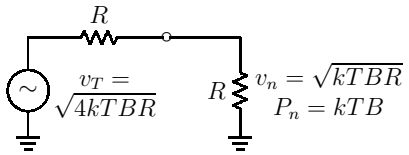
P_{dBm}	P [mW]	V_{rms} [mV]
0	1.0	223.6
−10	0.1	70.71
−20	0.01	22.36
−30	0.001	7.071
−40	0.0001	2.236
−50	0.00001	0.707

To carry out the gain calibration for a specific gain setting of the SDR, a CW signal with known power P_{dBm}, well above the noise floor but also well below the 1-dB compression point, is injected at a frequency of interest. To avoid dc offset and flicker noise problems, the SDR should be tuned to a slightly different frequency than the signal generator (e.g., 10 kHz offset). The calibration factor is then $\sqrt{2}$ **times** the V_{rms} voltage corresponding to P_{dBm} in the table above divided by the magnitude $\sqrt{V_I^2 + V_Q^2}$ of the (time averaged) rms voltages V_I and V_Q at the in-phase and quadrature (ADC) outputs of the I and Q channels of the DCR. The factor of $\sqrt{2}$ comes from the fact that the downconversion to baseband in the DCR removes half of the complex-valued Fourier transform (and thus half of the signal power). Note that this means that the power of received signals computed at baseband after calibration needs to be divided by 2 to obtain the original RF power before downconversion.

7.2. Noise, Noise Figure

Due to the movement of electrons in conductive materials, a resistor at any temperature above absolute zero will generate thermal noise. An equivalent circuit with a voltage source v_T and a noiseless resistor R is shown in Fig. 7.

The Thévenin voltage of the noise source is $\sqrt{4kTBR}$ so that the maximum noise power delivered to the matched load resistor R is $P_n = N = kTB$. The power spectral density of the noise is $N/B = kT$ in Joules=Ws. At 290 Kelvin (approximately room temperature) $N/B = 290 \times 1.38 \times 10^{-23} = 4.002 \times 10^{-21}$ which corresponds to -174 dBm (for a power P in watts, dBm are computed as $10\log_{10}(1000\,P)$). Instead of using power N, it is convenient to characterize noise sources in terms of their (equivalent) noise temperature T as



k = Boltzmann constant = $1.38 \times 10^{-23} J/K$
T = Absolute Temperature in Kelvin
B = Noise Bandwidth in Hertz
P_n = Noise Power in Watts

Figure 7. Thermal Noise Model

$$T = \frac{N}{kB} \,. \tag{10}$$

Noise factor measures the degradation of signal-to-noise ratio (SNR) in a device such as an attenuator or an amplifier. The noise factor F is defined as

$$F = \frac{S_i/N_i}{S_o/N_o} = \frac{S_i}{S_o}\,\frac{N_o}{N_i} \,, \tag{11}$$

where S_i/N_i is the SNR at the input of the device and S_o/N_o is the SNR at the output. A closely related quantity is the noise figure in dB defined as $NF = 10\log_{10} F$. Fig. 8 shows a block diagram for the derivation of the noise factor of a device under test (DUT).

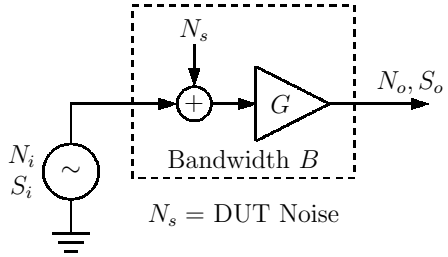


Figure 8. Model for Single Stage Noise Factor Computation

For the correct computation of F it is assumed that the device is terminated properly at the input and the output. It is also assumed that the input noise N_i and the internally generated noise N_s are uncorrelated so that the sum of the noises can be computed as $N_i + N_s$. Since $S_o = G\,S_i$ and $N_o = (N_i + N_s)\,G$ we have

$$F = \frac{1}{G}\,\frac{(N_i + N_s)\,G}{N_i} = \frac{N_i + N_s}{N_i} = 1 + \frac{T_s}{T_i}\,, \qquad (12)$$

where N_s and N_i are expressed in terms of their equivalent noise temperatures T_s and T_i. Thus,

$$T_s = (F-1)\,T_i \qquad \text{and} \qquad N_s = (F-1)\,kT_iB\,, \quad (13)$$

where T_i is usually set to 290 Kelvin. Fig. 9 shows the cascade of two devices with the same bandwidth B and gains G_1 and G_2. It is again assumed that all inputs and outputs are matched and terminated properly and that the noise sources N_i, N_{s1}, and N_{s2} are uncorrelated.

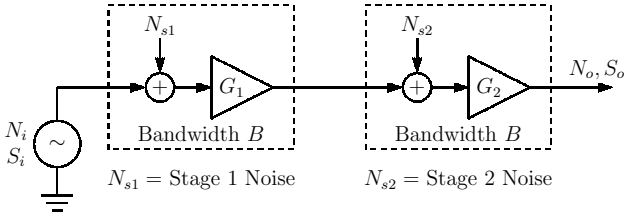


Figure 9. Noise Factor Computation for Two Stage Cascade

The overall noise factor computation proceeds as follows

$$\begin{aligned} S_o &= S_i\,G_1G_2\,, \\ N_o &= (N_i + N_{s1})\,G_1G_2 + N_{s2}\,G_2\,, \\ F &= \tfrac{1}{G_1G_2}\,\tfrac{(N_i+N_{s1})G_1G_2+N_{s2}G_2}{N_i} = \tfrac{N_i+N_{s1}}{N_i} + \tfrac{N_{s2}}{G_1N_i} \\ &= 1 + \frac{T_{s1}}{T_i} + \frac{T_{s2}}{G_1T_i}\,. \end{aligned} \qquad (14)$$

The overall equivalent system noise temperature in terms of the individual components T_{s1}, T_{s2}, and G_1 is therefore

$$T_s = (F-1)\,T_i = T_{s1} + \frac{T_{s2}}{G_1}\,, \qquad (15)$$

and the overall noise factor in terms of the noise factors of each stage is

$$F = F_1 + \frac{F_2 - 1}{G_1}\,. \qquad (16)$$

7.3. Y-Factor Calibration

The Y-factor method to measure the noise factor F (or the noise figure $NF = 10\log_{10} F$) makes use of a calibrated noise source (e.g., a noise diode) which has two distinct noise temperatures (corresponding to two different noise powers) depending on whether the noise device (ND) is turned on (i.e., powered) T_{ND}^{on} or off T_{ND}^{off}. The calibrated ND is characterized in terms of ENR (excess noise ratio) which is defined as

$$ENR = \frac{T_{ND}^{on} - T_{ND}^{off}}{T_0}\,. \qquad (17)$$

The published value of ENR is usually given in dB as $ENR_{dB} = 10\log_{10}(ENR)$. The temperature T_0 in the denominator is the reference temperature that is used for the definition of the noise factor F and usually $T_0 = 290$ K. The physical temperature of the ND is T_{ND}^{off} (in Kelvin) and, from knowing ENR and T_0, T_{ND}^{on} can be computed using eq. 17. Fig. 10 shows the Y-factor calibration setup for a single receiver (or general signal processing DUT) stage with gain G_r. It is assumed that the input and output of the DUT are properly terminated and matched, that the output of the ND and the noise N_r generated by the DUT are uncorrelated, and that the noise bandwidth of the ND is at least as large as the DUT bandwidth B.

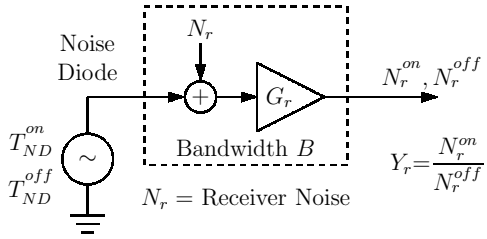


Figure 10. Y Factor Calibration Setup

The Y-factor Y_r is measured as the power ratio N_r^{on}/N_r^{off} at the output of the DUT. Mathematically

$$Y_r = \frac{(kT_{ND}^{on}B + N_r)\,G_r}{(kT_{ND}^{off}B + N_r)\,G_r} = \frac{T_{ND}^{on} + T_r}{T_{ND}^{off} + T_r}\,, \qquad (18)$$

where k is the Boltzman constant, B is the bandwidth of the DUT, and T_r is the noise temperature of N_r. Solving eq. 18 for the unknown T_r yields

$$T_r = \frac{T_{ND}^{on} - Y_rT_{ND}^{off}}{Y_r - 1} = (F_r - 1)T_0\,, \qquad (19)$$

where the last equality follows from eq. 13. With that $F_r = 1 + T_r/T_0$ and, if $T_{ND}^{off} = T_0$, then the equation simplifies to

$$F_r = \frac{ENR}{Y_r - 1}\,. \qquad (20)$$

This works well if the DUT itself can measure power (e.g., a spectrum analyzer or a SDR), but if some measurement device (MD) needs to be used for that purpose, then the noise factor F_{MD} of that device needs to be taken into account (Keysight). In this case the Y-factor Y_{tot} of the overall system is measured as N_{tot}^{on} divided by N_{tot}^{off} where N_{tot} refers to the noise of the overall system (DUT and MD). In terms of noise temperatures, this yields

$$Y_{tot} = \frac{T_{ND}^{on} + T_r + T_{MD}/G_r}{T_{ND}^{off} + T_r + T_{MD}/G_r} \,. \tag{21}$$

Solving for T_r results in

$$T_r = \frac{T_{ND}^{on} - Y_{tot} T_{ND}^{off}}{Y_{tot} - 1} - \frac{T_{MD}}{G_r} \,. \tag{22}$$

Using $T_{MD} = (F_{MD} - 1)T_0$ and assuming $T_{ND}^{off} = T_0$ the result for the noise factor F_r then becomes

$$F_r = \frac{ENR}{Y_{tot} - 1} - \frac{F_{MD} - 1}{G_r} \,. \tag{23}$$

7.4. Compression Point

A linear amplifier with power gain G and input signal power S_i outputs a signal with power $S_o = G\,S_i$ (if noise is negligible). Any deviation from this linear relationship indicates the presence of non-linear effects. The input signal level where the actual gain decreases by 1 dB from the linear gain ($S_o = 0.79\,G\,S_i$) is the 1 dB compression point (Wepman et al., Aug. 2015). Note that this may be difficult to measure in practice for digital outputs because the ADC may overload before the 1 dB compression point is reached.

7.5. Intermodulation Distortion

Most (linear) building blocks of communication systems exhibit non-linear behavior when the signal levels become too high. A simple way to model non-linearity for a memoryless system with input $x(t)$ and output $y(t)$ is the polynomial model

$$y(t) = \sum_{n=0}^{N} k_n\, x^n(t) \tag{24}$$

Intermodulation distortion (IMD) results from unwanted non-linear interactions between different frequency components of a signal $x(t)$. The simplest test signal for measuring IMD is a two-tone signal of the form $x(t) = A_1 \cos 2\pi f_1 t + A_2 \cos 2\pi f_2 t$. The $n = 2$ term in eq. 24 is then

$$\begin{aligned} x^2(t) &= (A_1 \cos 2\pi f_1 t + A_2 \cos 2\pi f_2 t)^2 \\ &= A_1^2 \cos^2 2\pi f_1 t + A_2^2 \cos^2 2\pi f_2 t + \\ &\quad + 2A_1 A_2 \cos 2\pi f_1 t \cos 2\pi f_2 t \\ &= \tfrac{A_1^2}{2}(1 + \cos 4\pi f_1 t) + \tfrac{A_2^2}{2}(1 + \cos 4\pi f_2 t) + \\ &\quad + A_1 A_2[\cos 2\pi (f_1 - f_2)t + \cos 2\pi (f_1 + f_2)t] \,. \end{aligned} \tag{25}$$

Thus, $x^2(t)$ has frequency components at dc, $2f_1$, $2f_2$, and at $f_1 \pm f_2$. The latter three are second order distortion frequency components. The term $\cos 2\pi f_1 t \cos 2\pi f_2 t$ from which $f_1 \pm f_2$ results is called a second order intermodulation (IM2) product. To quantify IM2, the ratio of an undesired IM2 output (usually at $f_1 - f_2$) over a desired output (at f_1 or f_2), when $A_1 = A_2 = A$, is defined as (Couch, 2007)

$$R_{IM2} = \frac{k_2 A^2}{k_1 A} = \frac{k_2 A}{k_1} \,. \tag{26}$$

The parameter called second order intercept point (abbreviated SOI or IP2) is then defined as the point where $R_{IM2} = 1$.

For third order non-linearities the term $x^3(t)$ for a dual tone signal is computed similar to eq. 25 to obtain the result

$$\begin{aligned} x^3(t) &= \tfrac{1}{4}\big[3A_1(A_1^2 + 2A_2^2)\cos 2\pi f_1 t + A_1^3 \cos 6\pi f_1 t + \\ &\quad + 3A_2(A_2^2 + 2A_1^2)\cos 2\pi f_2 t + A_2^3 \cos 6\pi f_2 t + \\ &\quad + 3A_1^2 A_2(\cos 2\pi(2f_1 - f_2)t + \cos 2\pi(2f_1 + f_2)t) + \\ &\quad + 3A_1 A_2^2(\cos 2\pi(2f_2 - f_1)t + \cos 2\pi(2f_2 + f_1)t)\big] \,. \end{aligned} \tag{27}$$

This has spectral components at f_1, f_2, $3f_1$, $3f_2$, $2f_1 + f_2$, $2f_2 + f_1$, $2f_1 - f_2$, and $2f_2 - f_1$ of which the last four come from third order intermodulation (IM3) products, with the last two usually being of most interest since they result in similar frequencies as f_1 and f_2 themselves. In this case the ratio of an undesired IM3 output (usually $2f_1 - f_2$ or $2f_2 - f_1$) over a desired output (f_1 or f_2) with $A_1 = A_2 = A$ is

$$R_{IM3} = \frac{k_3 3A^3/4}{k_1 A} = \frac{3k_3 A^2}{4k_1} \,, \tag{28}$$

and the third order intercept point (abbreviated TOI or IP3) is defined as the point where $R_{IM3} = 1$.

8. Acknowledgement

The author would like to thank his colleagues at ITS, in particular Todd Schumann for making all the sensor measure-

ments and Jeff Wepman, Adam Hicks, Linh Vu, Heather Ottke, and Mike Cotton as discussion partners.

References

Armstrong, Edwin H. A new system of shortwave amplification. *Proceedings of the Institute of Radio Engineers*, pp. 3 – 11, Feb. 1921.

Colebrook, F.M. Homodyne. *Wireless World and Radio Review*, (13):774, 1924.

Cooley, James W. and Tukey, John W. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301, 1965.

Couch, Leon W. *Digital and Analog Communication Systems*. Pearson Prentice Hall, 7th edition, 2007.

Deery, Joe. The 'real' history of real-time spectrum analyzers. *Sound and Vibration*, pp. 54 – 59, Jan. 2007.

Douglas, Alan. The legacies of edwin howard armstrong. *Proceedings of the Radio Club of America*, 64(3), Nov. 1990.

Gabrilovitch, J. British patent no. 504,455. *British Patent*, 1936.

gnuradio. The signal metadata format specification. https://github.com/gnuradio/SigMF, 2017. Accessed: 2019-09-08.

Hartley, Ralph V.L. Modulation system. *US Patent 1666206*, Apr. 1928.

IEEE. P802.22.3 - standard for spectrum characterization and occupancy sensing. https://standards.ieee.org/project/802_22_3.html, 2018. Accessed: 2019-09-04.

ITS. NTIA/ITS spectrum monitoring SCOS sensor reference implementation. https://github.com/NTIA/scos-sensor, 2018. Accessed: 2019-09-08.

Keysight. Noise figure measurement accuracy: The y-factor method. *Keysight Application Note*.

Mashhour, Ashkan, Domino, William, and Beamish, Norman. On the direct conversion receiver – a tutorial. https://www.microwavejournal.com/articles/3226-on-the-direct-conversion.-receiver-a-tutorial, June 1, 2001. Accessed: 2019-09-05.

Meyer, Robert G. and Wong, Alvin K. Blocking and desensitization in rf amplifiers. *IEEE Journal of Solid-State Circuits*, 30(8):944–946, Aug. 1995.

Weaver, D.K. A third method of generation and detection of single side band signals. *Proc. of IRE*, 44(12):1703 – 1705, 1956.

Wepman, Jeffery A., Bedford, Brent L., Ottke, Heather E., and Cotton, Michael G. Rf sensors for spectrum monitoring applications: Fundamentals and rf performance test plan. *NTIA Report Series*, (15-519), Aug. 2015.