Machine Learning Techniques for Equalizing Nonlinear Distortion

A Technical Paper prepared for SCTE•ISBE by

Rob Thompson
Director, NGAN Network Architecture
Comcast
1800 Arch St., Philadelphia, PA 19103
(215) 286-7378
robert_thompson6@cable.comcast.com

Xiaohua Li
Associate Professor, Dept. of Electrical and Computer Engineering
State University of New York at Binghamton
Binghamton, NY 13902
(607) 777-6048
xli@binghamton.edu

Table of Contents

List of Figures

List of Tables

1. Introduction

Since the early 2000s, the cable television (CATV) industry has been playing its part in the Artificial Intelligence (AI) community by deploying equalization technology to enable its digital signals to survive varying frequency response conditions within its cable plants. Simon Haykin describes how the perceptron and the adaptive filter using the least mean squares (LMS) algorithm are naturally related [1]. Equalization has evolved into a powerful tool, enabling the CATV industry to achieve communication efficiencies once thought impossible -- but that story is not quite complete. The limits of equalization may extend beyond the linear frequency response, and cancel the nonlinear responses commonly associated with nodes and other active devices which use power amplifiers (PAs). Achieving nonlinear equalization requires new equalization methods, like receiver post-distorter equalization, where techniques include AI models, such as deep neural networks (DNNs). Furthermore, researchers have been advancing nonlinear distortion cancellation via other methods, including peak-to-average-power-ratio (PAPR) reduction, and digital pre-distortion (DPD). These technologies are beginning to show up in newer generation devices, where demands for radio frequency (RF) output power is high, while keeping power consumption low, like the full duplex DOCSIS (FDX) remote PHY device (RPD) nodes. DPD technologies cancel the contribution of the transmitting device only. More aggressive nonlinear distortion cancellation methods may be accomplished by advanced DNN approaches, such as incorporating input features derived from Volterra series models, which has become a popular model for nolinear distortion that can be used to describe multiple nonlinearity orders and memory. Then efficiencies across the CATV network could be considered, either by higher node RF output power, or more efficient PA architecture/bias within the node, amplifier, and/or customer premise equipment (CPE). This paper will propose how current CATV equalization systems could be enhanced to cancel severe nonlinear distortion based on some of these novel approaches to nonlinear equalization.

2. Artificial Intelligence (AI)

2.1. Historical Perspective

AI has existed for a very long time -- close to 80 years. In 1943, Warren Sturgis McCulloch and Walter Pitts published a paper titled "A Logical Calculus of Ideas Immanent in Nervous Activity," laying the foundations for artificial neural networks (ANNs) [1]. Since then, many ideas involving AI have been shared, and this community has grown appreciably. Patrick Winston, who was born in 1943 and later became a MIT professor who taught a course in AI, described it as being about algorithms, enabled by constraints, exposed by suite of representations, that support the development of models targeted at thinking, perception, and action [2]. That definition is inclusive of many things -- in fact, some very simple internet searches can yield timelines rich with AI milestones, including events such as when Deep Blue defeated World Chess champion Garry Kasparov in 1997, or a more recent milestone, on October 15th, 2019, when OpenAI enabled a robot to learn how to single-handedly solve Rubik's Cube with the support of two neural networks [3].

2.2. Common Solutions

There are many artificially intelligent solutions that we encounter every day, possibly without even realizing it. Comcast, for example, provides multiple products which incorporate AI technology. Some of these solutions include the Voice Remote, that adapts to the uniqueness of how each and every one of us speaks, and in turn assists with accessing and enjoying content on Comcast's X1 platform. The X1 recommendation engine detects patterns in the content we consume, and assists users in navigating a wealth of available content and offer recommendations. Internet-based products like xFi Advanced

Security protect our home networks against constantly evolving network threats. Interactive assistants, like xFinity Assistant, help customers by leveraging an interactive knowledge base of common solutions.

2.3. Popular Tools/Models

There are many available AI tools and frameworks, including TensorFlow, and PyTorch [4]. These systems are designed to assist with navigating the vast array of models, each with their unique set of pros and cons, when it comes to approximating the functions that couple input and output data patterns together. Some of these tools readers may have already heard of, like DNNs [5]. Others, like support vector machines (SVMs), may be less familiar. Fortunately, finding the right model fit for a particular problem has been automated via tools like automated machine learning (Auto ML), which not only selects the best function approximation model, but also assists with tuning the parameters of that model to optimize its training and generalization properties.

2.4. Biological Inspiration

The perceptron model gets its inspiration from the Pyramidal cell shown Figure 1 [1]. One of the key characteristics of this model are its synaptic weights, which it applies to each of its input signals. The input signals are summed together and applied to an activation function. The sigmoid function is a popular activation function, which limits the output response to a specific range of continuous values. An Artificial Neural Network (ANN) connects multiple perceptions together in a variety of ways, in parallel, and/or in series. In doing so, interesting behaviors begin to emerge -- the most interesting being the nonlinear adaptation of the model weights. This could present interesting opportunities for linear adaptive equalization systems used today, where nonlinear adaptation enhancements could enable systems to account for both linear and nonlinear distortion present within the communication channel to improve the performance of equalization systems overall.
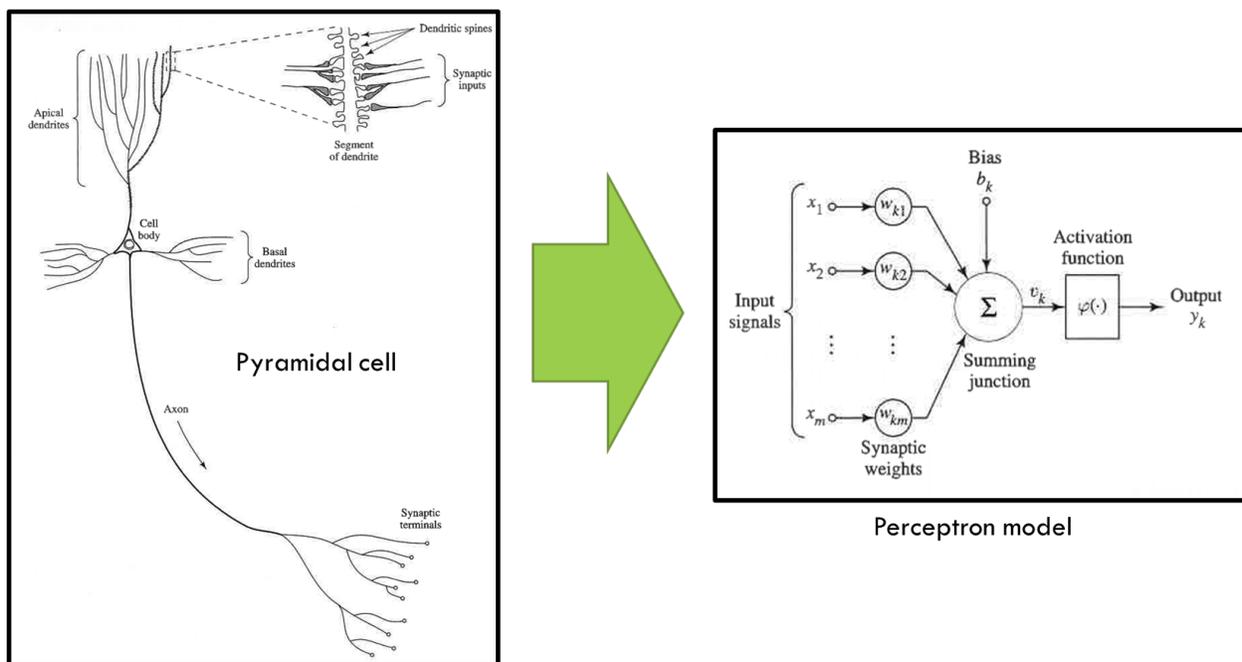

Figure 1 - Biologically Inspired Perceptron Model

2.5. DOCSIS Transmit Pre-Equalization

Since the early days of the Data Over Cable Service Interface Specifications (DOCSIS), equalization has enabled the cable modem termination systems (CMTSs) to adapt and learn the unique frequency response shared between it and each of the cable modems (CMs) to which it is connected. The CMTS shares this knowledge with the CM, via a coefficient vector or weights often times referred to as "taps", asking it to either convolve or overwrite its current set of weights, based on how quickly the frequency response was changing. The CM applies these weights to future transmissions to the CMTS, to cancel the frequency response effects of the channel, which could be micro reflections (echoes) or filter effects including group delay or amplitude roll-off [6]. This form of equalization came to be known as "transmit pre-equalization" in DOCSIS 1.1 [7].
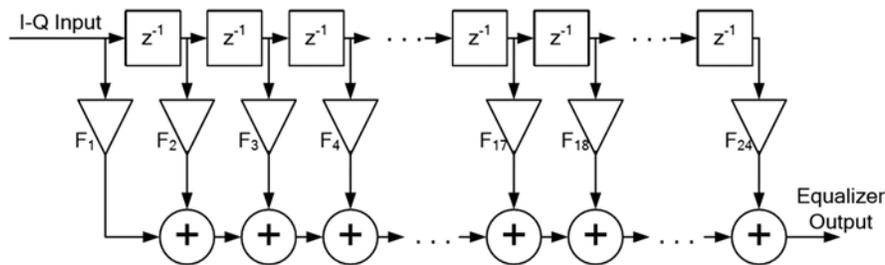
Figure 2 - DOCSIS Transmit Pre-Equalizer Structure

Comparing the DOCSIS 2.0 equalizer of Figure 2 to the perceptron of Figure 1, one cannot help but notice the similarities between the two models. What is most like the perceptron is the linear adaptive filter's weighted inputs feeding a linear combiner, and the ability to perform continuous learning – a single neuron operating in its linear mode [1]. The perceptron and an adaptive filter using LMS are naturally related [1].

3. Power Amplifier (PA) Efficiency Problem

Designer Charles Warren once gave a talk titled "How Might We: Three Words That Make Design Better" [8]. His thoughts are both refreshingly entertaining, and very helpful in organizing our ideas around innovation and establishing the following goal statement for this paper.

How might we optimize PA efficiency in our RPDs?

Improving PA efficiency in our RPDs may be beneficial in maintaining existing requirements, like RF output levels, as we introduce new capacity-enhancing technology. FDX falls into this category. Its accompanying echo cancellation (EC) technology, is necessary for facilitating bidirectional communication at the same operating frequencies [7]. EC technology additions will require compromises, especially while maintaining existing requirements for RPD RF output power, complexity, weight, power consumption, heat dissipation and cost.

At this point, the reader may be thinking "The goal statement is limited to RPDs, but why not optimize all the active components within the RF chain, including line extenders, mini-bridgers, trunk amplifiers, home drop amplifiers and even CPE front ends?" This is a thought we hope to address in paper as well, but in the spirit of following Warren's "How Might We" process through to its end, let's first consider some of the things that may stopping us from achieving our goal as stated.

First and foremost, there is a catch when it comes to increasing PA efficiency, and that is increased nonlinearity. Any savings in PA power consumption will have to be balanced with a system of mitigating any increases in nonlinearity, something we intend to explore fully in this paper. Another consideration is with respect to standards, and whether the solution requires standardization to ensure seamless interoperation across vendors that can be deployed, operated, and maintained in a consistent manner. Lastly, with introduction of any new technology, there is always consideration given to how to gracefully coexist with legacy products and services, ideally minimizing any impact to existing services.

4. Nonlinear Distortion (NLD)

Figure 3 illustrates how PAs strengthen their input signals [10]. When a PA's input power is at its lowest levels, its output power behavior is more linear than it is nonlinear, and its gain is constant. Ideally, PAs would behave linearly for all input signal levels, including high input powers, as illustrated by the dotted line. However, practical PAs generally available today cannot strengthen input signals without adding nonlinear distortion (NLD) to those input signals. As we will later see, NLD increases more rapidly than the illustrated input/output increases of the fundamental signals. Eventually, the PA reaches saturation, and its performance becomes more nonlinear than linear. At this point, the PA's output is no longer proportional to its input, and its performance is dominated by NLD. Further, a PA's linear operating region or dynamic range is a range of input powers that include a predictable mixture of impairments, including noise and NLD. At low input power, noise dominates the impairment mixture, but as input power increases, noise performance improves, while NLD worsens. The challenge for the network designer is to strike a balance between noise and NLD, so that their combined performance is within acceptable limits. Output-power-back-off (OBO) is a term used to describe this compromise, where the PA's operating point is typically several decibels (dBs) below its compression point and includes acceptable noise and NLD levels for overall system performance [11].
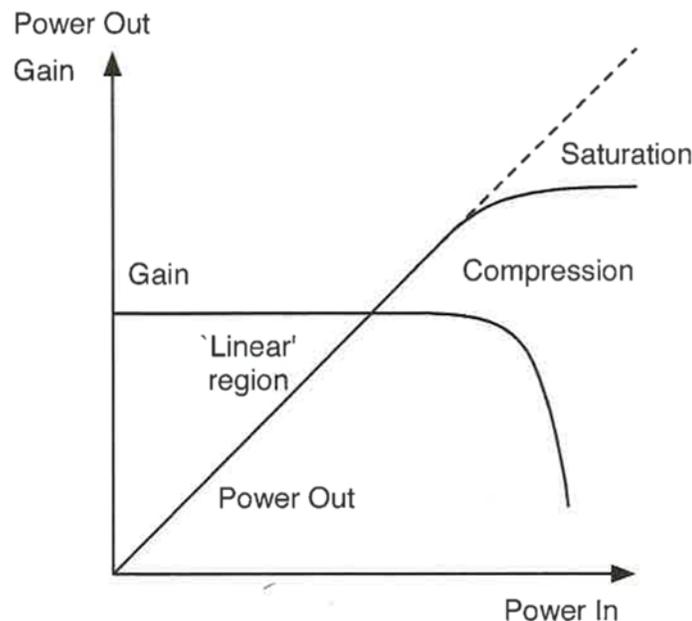
Figure 3 - Power Output and Gain Compression Characteristics of a PA

SCTE•ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 \ 2020

2020 Fall
Technical Forum
SCTE·ISBE · NCTA · CABLELABS®

4.1. Digital Signal Impact

Metrics pertaining to Figure 4 node downstream transmission path (blue) performance for noise and NLD have been included in Table 1 [9]. These specifications are measured using 6 MHz wide channels and are based on a full channel loading consisting of 194 single carrier quadrature amplitude modulation (SC-QAM) signals. Decibel-millivolts, dBmVs, are generally used, for mathematical convenience, within the CATV industry, to reference operating levels, O, for analysis of these systems [12].

Figure 4 - Sample RPD Node

SCTE•ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 2020

2020 Fall
Technical Forum
SCTE·ISBE • NCTA • CABLELABS®
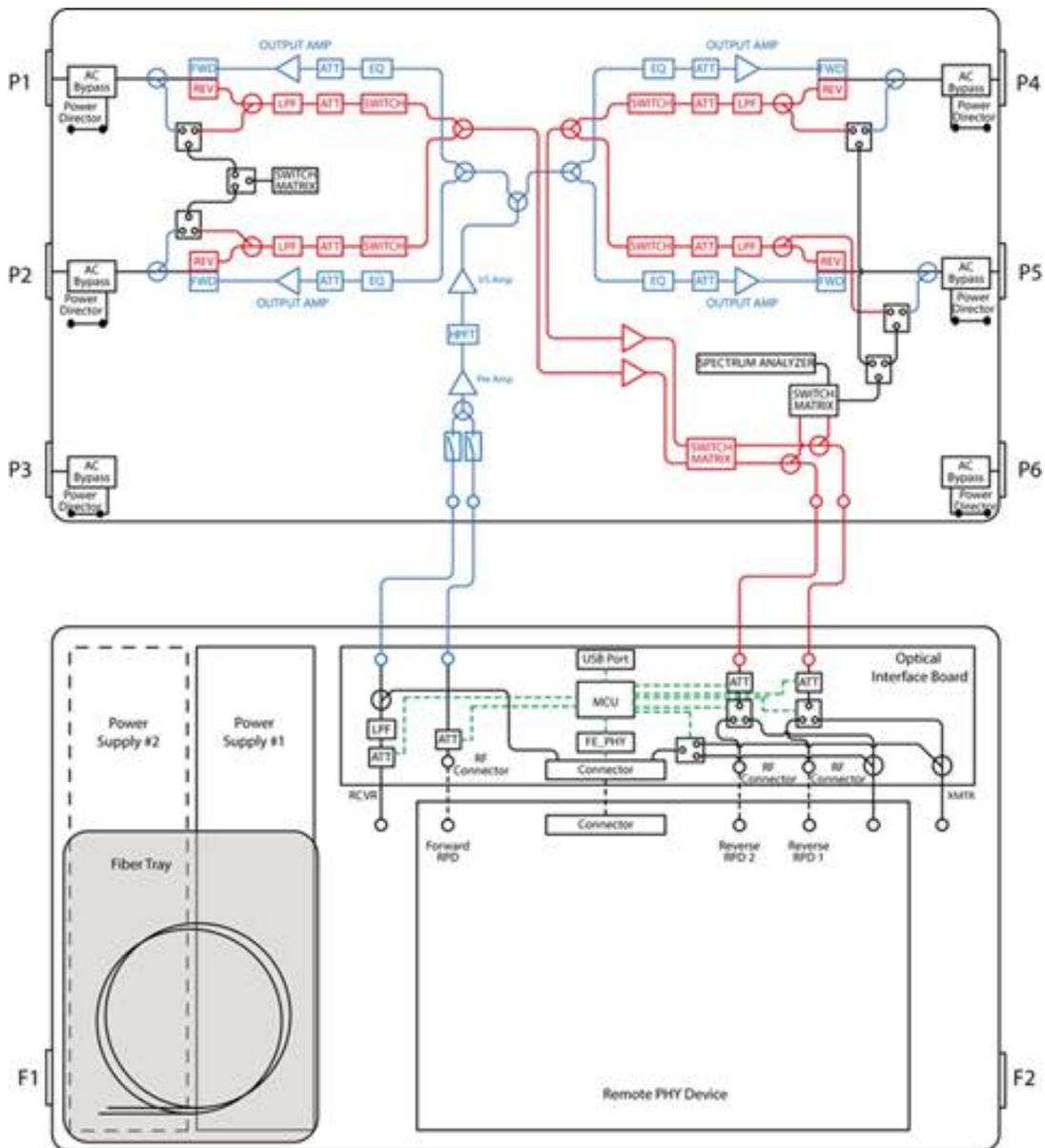
Table 1 - Sample RPD Node Datasheet Summary

Sample RPD Node Specifications	
Minimum operational gain, G	42 dB
Noise Figure, NF	15.5 dB at 54 MHz
Composite-Intermodulation-Noise, CIN_{ref}	50 dB
Reference output level, O_{ref}	42 dBmV at 54 MHz
Power	160.6 W @ 2.16 A, and 90 V AC
Weight	49.8 lbs.

Signal-to-noise ratio (SNR) describes the relative measure of signal power, S, to the noise power, N_P, which is the thermal noise or noise floor measured within the same bandwidth as the signal S, in this case 6 MHz. N_P is estimated using (1), where k is Boltzman's constant (1.374×10^{-23} $joules/°K$), T is the absolute temperature in degrees Kelvin (°K), and B is the bandwidth of the measurement in Hertz (Hz). N_P in the CATV industry is typically expressed in terms relative to 1 milli-volt (mV) across a 75 Ω impedance, therefore N_P at 62 °F is approximately -57.4 dBmV [13].

$$N_p = kTB \tag{1}$$

Composite Intermodulation Noise (CIN) is a type of NLD which results from nonlinear distortion generated from loading conditions, which include digital signals, like SC-QAM. Node contribution for SNR and CIN can be calculated using (2) and (3) respectively, for changes in its output levels O constrained over the node's dynamic range [12], [13]. In the CATV industry, CIN is typically dominated by 3rd-order NLD, which may not be the case for other communication systems, like those used in the satellite industry [14].

$$SNR = O + 57.4 - G - NF \tag{2}$$

$$CIN = CIN_{ref} - 2(O - O_{ref}) \tag{3}$$

Increasing the node's operating output power, from the originally specified O_{ref}, by 5 dB, will result in a new output level, $O = 47\ dBmV$ at 54 MHz. This new output level increase will also increase SNR to 47 dB at 54 MHz, using (2). However, CIN in (3) will decrease to 40 dB, leading to a 2 dB degradation overall of the System SNR, SNR_S, per (4).

$$SNR_S = -10log_{10}\left(10^{-\frac{SNR}{10}} + 10^{-\frac{CIN}{10}}\right) \tag{4}$$

Therefore, increasing node RF output signal levels may enable the designer to improve homes-per-node efficiency, such as with higher output level O, but will do so at the expense of increasing the node's power consumption and degrading the overall system performance criteria, SNR_S. Increasing PA RF output levels of networked devices is one of the ways in which to optimize PA efficiency. However,

SCTE·ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 2020

2020 Fall
Technical Forum
SCTE·ISBE · NCTA · CABLELABS®

higher RF output levels may drive power consumption above the network operator's acceptable threshold -- in the case of the node example, above a 160 W maximum. Power consumption threshold values are based on network operator's unique powering constraints, which may be limited by multiple factors, including network design, hardware capability and local regulatory restrictions. Regulatory restrictions here specifically involve the placement of powering hardware at specific telephone pole or pedestal locations. Increased power may also impact heat dissipation and design of the node's housing. These impacts may translate to the larger surface area and weight of the node's housing to facilitate necessary heat transfer, where weight may increase above an acceptable threshold, in this case 50 lbs. maximum. Node weight is an important factor, because technicians need to be able to lift the node in order to connect it to the cable plant. The activity could be above ground, attached to telephone poles, or below ground within a pedestal mount.
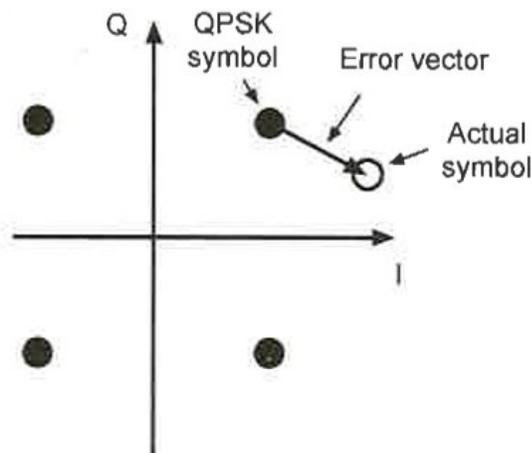
Figure 5 – Error Vector Magnitude (EVM) for a QPSK Signal

The effect of degraded performance on digital signals is illustrated in Figure 5. A quadrature phase shift keying (QPSK) signal degraded by SNR_S. will cause actual symbol reception to deviate from the ideal symbol receive point, shown as a dark circle. The resultant error vector, e_j, between the actual symbol and the ideal QPSK symbol receive point represents a measure of fidelity. Modulation error ratio (MER) in (5) measures the cluster variance in dB, that can be observed in a SC-QAM signal. It includes the effects of inter-symbol interference (ISI) spurious, phase noise, and all other degradations, where E_{av} is the average constellation energy for equally likely symbols, and N is the number of symbols averaged [7].

$$MER_{symb}(dB) = 10\log_{10}\left\{\frac{E_{av}}{\frac{1}{N}\sum_{j=1}^{N}\left|e_j\right|^2}\right\} \qquad 5$$

Poor MER can lead to symbol decision boundary crossings, translating to symbol errors. If frequent enough, these symbol errors can overwhelm forward error correction (FEC) schema, leading to packet errors and ultimately loss of network payload.

4.1. Adjacent Channel Leakage Ratio (ACLR) Measurements

Figure 6 illustrates a nonlinear response to a band-limited signal, where the output of a system represents both linear (solid line) and nonlinear components (dashed line) [10]. The term intermodulation distortion (IMD) is synonymous with CIN.
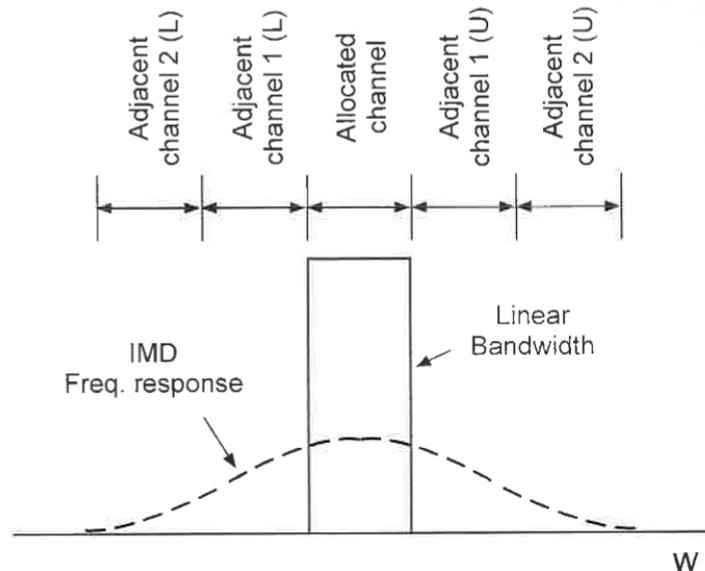
Figure 6 - Output Spectrum Comprising Allocated and Adjacent Channels

Figure 6 is an example of how CIN can accumulate under the signal(s). One in-band measurement approach of CIN involves the following steps:
1. Measuring SNR_S, in dBmV per 6 MHz
2. Turning the signal load off
3. Measuring the thermal noise contribution to SNR_S at the same frequency, N_P, also in dBmV per 6 MHz
4. Calculating the difference between these two values, via algebraic manipulation of equation 4, as the CIN contribution

This is a reasonable measurement approach in a lab environment, where turning the PA's signal load off will likely not negatively impact a customer. Figure 6 illustrates a more customer-friendly approach that involves out-of-band distortion measurements of the integrated power in the adjacent channel(s), called adjacent channel power ratio (ACPR) or adjacent channel leakage ratio (ACLR), where these measurements are expressed in decibels relative to the signal's channel power, or -dBc [10].
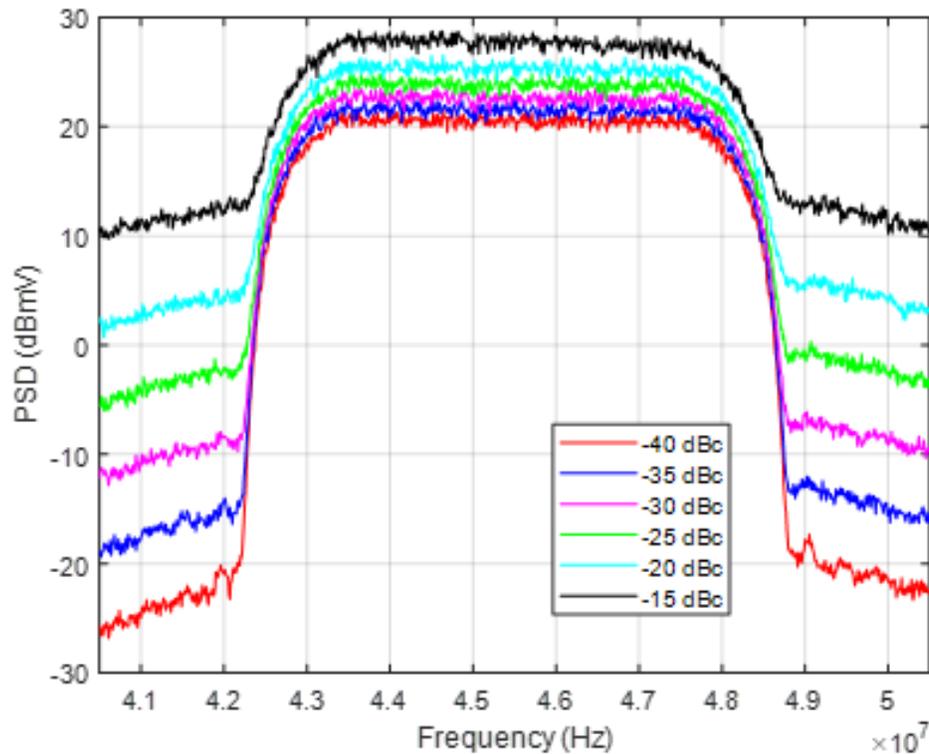
Figure 7 - Increasing Spectral Regrowth over Amplifier Linear Region

Figure 7 is composed of a series of output measurements of one PA, where each color corresponds with 5 dB changes in ACLR or spectral regrowth, resulting from incremental adjustment of the PA input power. ACLR measurements were made via Keysight's PXA signal analyzer model N9030A and vector signal analyzer (VSA) software, model 89601B. The 6.4 MHz bandwidth power delta marker measurements of both the received signal power and the spectral regrowth of the upper first adjacent channel were used to obtain the ACLR measurements.

Figure 7 data was obtained from a CATV drop amplifier, which is sometimes used within the customer's home to overcome losses associated with providing signals to multiple (1-3) client devices, primarily supporting video services, but also including voice and high-speed data services. PAs are used in many CATV network clients' front ends, used within the customer's home, like set-top boxes (STBs), cable modems (CMs), and digital terminal adapters (DTAs). The outside plant represents another group of network elements with embedded PAs, including the node previously discussed and additional amplifiers, called trunk amplifiers, line extenders and bridging amplifiers, which are used to compensate for cabling and passive losses incurred while distributing and aggregating services to and from the customer's home.

All of these cascaded PAs aggregate noise and NLD, resulting in an accumulated end-of-line (EOL) performance, which further restricts the individual contribution for any one PA in the chain to that of even higher fidelity (i.e. 50 dB per Table 1). This ensures that the customer's services meet, or ideally exceed, some minimum service level agreement (SLA). Additionally, variations in RF levels can occur for multiple reasons, like temperature changes, wind loading, and plant maintenance. Sometimes RF levels can change appreciably over short periods of time. Network designers may specify even better performance, to accommodate this performance variation in several of the components or sub systems in the network chain, a.k.a. 'margin-stacking' [10].

Similar ecosystems exist in other industries, including cellular, Wi-Fi, satellite, and the Internet of Things (IoT). For example, massive MIMO and millimeter wave transmissions use many PAs in cellular deployments and will also have similar end-to-end performance requirements [11].

4.2. PA Industry

The PA industry has a growth outlook from 2019 to 2025, with compound annual growth rate (CAGR) of 7.6%, primarily from anticipated developments in newer generation cellular communications. Some of the key providers include NXP Semiconductor (Netherlands), Broadcom Corporation (U.S.), Qorvo Inc. (U.S.), Anadigics Inc. (U.S.), RFHIC Corporation (U.S.), TekTelic Communications Inc. (U.S.), Texas Instrument (U.S.) among others [27].

Providers can supply a diverse range of PA classes suited for a variety of applications, some of which have been summarized in Table 2Table 2 - PA Classes.

Table 2 - PA Classes

Class	Description
A	Very linear, especially for smaller signal amplitudes. Used for PAs at high millimeter-wave frequencies, but considered too inefficient, < 50%, for PAs in cellular wireless communications applications
B	Power transistor is biased just at the threshold voltage, so the transistor conducts for only a half-cycle of the RF waveform. This wave generates a significant amount of even harmonic distortion, and some self-bias. Maximum theoretical efficiency rises to 78.6%, compared with a Class A.
AB	Transistor is biased slightly into the "on" condition, usually specified as a constant quiescent drain current in FETs, with the DC gate bias adjusted to provide the design standing current. The optimal load is generally very close to the Class B value, and the achievable energy efficiency can be over 70%, theoretically.
C	The Power transistor is biased below threshold, so under quiescent conditions, the transistor is switched off. This mode can be very energy efficient, but the gain is low, and the harmonic output and intermodulation distortion can be considerable.
D, E, & F	Very efficient. No overlap between the voltage and current waveforms, so this amplifier converts the DC to RF with 100% efficiency. A resonator can be used to filter out the harmonics at the load.
D	Form of switching amplifier, where the PA is driven by a pulsed waveform and the output current is switched rapidly between on and off states. Uses two transistors in a push-pull arrangement.
E	Both 2nd, and 3rd harmonics are tuned
F	2nd harmonic tuned
Inverse F	2nd harmonic tuned

The CATV industry primarily uses class A PA products because of high performance requirements, including low noise and high linearity, over a broad range of spectrum, typically between 5 and 1,218

MHz. While peak efficiencies for some of these classes may be high, they are attained only at or close to the maximum RF output power and will fall rapidly with OBO. To overcome this basic drawback, alternative PA architectures will be required.

4.1. NLD Modeling

Perhaps the simplest way to represent PA NLD is to describe what it is not. A distortionless PA would have a linear transfer characteristic, which is achieved when the waveform of the PA's output voltage precisely duplicates that of its input. Only when a PA distorts does the output signal contain additional components, at frequencies differing from the frequencies of the input signal. The nature as well as the degree of distortion is dependent not only on the shape of the transfer characteristic of the PA, but also on the loading condition and operating point (bias) [12].

At higher powers, we have seen that the output power and gain deviate significantly from the linear relationship at small signals. This is the compression region of operation, and at a sufficiently high input drive, we will get no more power out of the PA -- at this point, we are at the saturated power level. In these regions of operation, the PA is very nonlinear. This compression behavior is also known as amplitude modulation to amplitude modulation (AM-AM) conversion: by modulating (changing) the input signal amplitude, we affect or modulate the amplitude of the output signal in a nonlinear fashion [12].

Typically, frequency-domain polynomial models will be used to model the AM-AM and amplitude modulation to phase modulation (AM-PM) characteristics of the PA. In general, frequency-domain models can describe the RF frequency response phenomena quite well but are unable to accommodate the memory effects associated with long time constants -- for example, bias line reactance and charge storage [10]. A Volterra series can be thought of as a Taylor series with memory; that is, a Taylor series defines not only at the present instant in time, but includes terms at previous instants, up to some specified delay [10].

Consider the baseband discrete model of the PA $y(n) = f(x(n), x(n-1), \cdots)$, where $x(n)$ is the input signal, $y(n)$ is the output signal, and $f(\cdot)$ is some nonlinear function. The simplest nonlinear PA model is the AM-AM/AM-PM model. Let the amplitude of the input signal be $V_x = E[|x(n)|]$, where $E[\cdot]$ denotes short-term expectation or average. The output sample $y(n)$'s amplitude $V_y = E[y(n)]$ and additional phase change $\psi_y = E[\angle y(n)]$ depend on V_x in nonlinear ways as (6) and (7):

$$V_y = \frac{gV_x}{\left(1 + \frac{gV_x}{c}\right)^{\frac{1}{2\sigma}}} \tag{6}$$

$$\psi_y = \frac{\alpha V_x^p}{1 + \left(\frac{V_x}{\beta}\right)^q} \tag{7}$$

where g is the linear gain, σ is the smoothness factor, and c denotes the saturation magnitude of the PA. Typical examples of these parameters are $g = 4.65, \sigma = 0.81, c = 0.58, \alpha = 2560, \beta = 0.114, p = 2.4$, and $q = 2.3$, which are used in the PA models regulated by Institute of Electrical and Electronics Engineers (IEEE) 803.11ad task group (TG) [11].

More accurate models should take into consideration the fact that nonlinearity leads to memory effects. In this case, Volterra series (8), are typically used to model PAs [11]. A general model is shown in [11] with up to Pth order nonlinearity and up to D step memory.

$$y(n) = \sum_{d=0}^{D} \sum_{k=1}^{P} b_{kd} x(n-d) |x(n-d)|^{k-1} \tag{8}$$

It can be shown that estimation of only odd-order nonlinearity (i.e. odd k) may be necessary for limited narrowband loading conditions and specific center frequencies, because even-order nonlinearity falls outside of the passband and will be filtered out by the receiver bandpass filters [11]. To illustrate this phenomenon, we can consider some simple examples where the input signal $x(n)$ consists of a few (1-3) single frequency components only. Omitting the memory effects, if $x(n)$ is a single frequency signal, i.e., $x(n) = V_0 \cos(a_0 + \phi)$, where $a_0 = 2\pi f_0 n$. Then, using well-known trigonometric identities, the output signal can be written as

$$y(n) = k_1 V_0 \cos(a_0 + \phi + \psi_1) + \left(\frac{3}{4} k_3 V_0^3 + \frac{5}{8} k_5 V_0^5\right) \cos(a_0 + \phi + \psi_3 + \psi_5) \tag{9}$$

$$+ \frac{1}{2} k_2 V_0^2 + \frac{3}{8} k_4 V_0^4 \tag{10}$$

$$+ \left(\frac{1}{2} k_2 V_0^2 + \frac{1}{2} k_4 V_0^4\right) \cos(2a_0 + 2\phi + 2\psi_2 + 2\psi_4) \tag{11}$$

$$+ \left(\frac{1}{4} k_3 V_0^3 + \frac{5}{16} k_5 V_0^5\right) \cos(3a_0 + 3\phi + 3\psi_3 + 3\psi_5) \tag{12}$$

$$+ \frac{1}{8} k_4 V_0^4 \cos(4a_0 + 4\phi + 4\psi_4) + \cdots \tag{13}$$

$$+ \frac{1}{16} k_5 V_0^5 \cos(5a_0 + 5\phi + 5\psi_5) + \cdots \tag{14}$$

$V_0 \cos(a_0 + \phi)$ and $V_0 \sin(a_0 + \phi)$ are both sinusoidal voltages. Their waveforms are identical except for a 90° phase difference. The cosine form is used throughout this analysis because it results in simpler expressions [11]. The first line (9) is the in-band response with AM-AM/AM-PM nonlinear effects, the second line (10) is the direct current (DC) bias, and the remaining lines (11) through (14) include second through fifth order harmonics. At the receiving side, we may just have line (9) left, because all the other items will be canceled by bandpass filtering. Communication network filters, such as the root-raised cosine (RRC) filters, are typically implemented in two halves, one in the transmitter and the other in the receiver, so that overall, we get Nyquist rate sampling, and provide necessary impedance matching of the power transistor to its optimum load impedance of the network. Another critical filter function is in their use of controlling out-of-band emissions from sources including PA NLD, thus limiting the impact of distortion to within the operating band.

If $x(n)$ consists of two frequencies, i.e., $x(n) = V_1 \cos(a_1 + \phi_1) + V_2 \cos(a_2 + \phi_2)$, where $a_i = 2\pi f_i n$, then the inband response includes many more terms, such as the first order terms $k_i V_i \cos(a_i + \phi_i + \psi_i)$, the third order terms $k_3(V_i^3 + V_i V_j^2) \cos(a_i + \phi_i + \psi_i)$, the fifth order terms $k_5(V_i^5 + V_i V_j^4 + V_i^3 V_j^2) \cos(a_i + \phi_i + \psi_i)$, for $i, j \in \{1,2\}$. There are also intermodulation terms that consist of $na_i \pm ma_j$ as long as they are within the passband of the bandpass filter, such as $(V_i^2 V_j + V_i^2 V_j^3 + V_1^4 V_j) \cos(2a_i - a_j + 2\phi_i - \phi_j + 2\psi_i - \psi_j)$. For some specific loading conditions, there may be many other higher order terms with frequencies na_i, $n(a_i \pm a_j)$, or $na_i + ma_j$, that can not pass the passband filter. One of the important observations is that the contents that can pass the passband filter may consist of odd-order nonlinearity only for specific center frequency and narrowband conditions only.

If $x(n)$ consists of three or more frequencies, we can have similar observations, albeit the expressions are more complex. Let the input signal $x(n)$ be

$$x(n) = \sum_{i=1}^{3} V_i \cos(a_i), a_i = 2\pi f_i n. \tag{15}$$

The second order component includes the DC component $g_{2,0}(n)$, the sum/difference of beat components $g_{2,1}(n)$, and the second-order harmonic components $g_{2,2}(n)$. Specifically,

$$k_2 x^2(n) = g_{2,0} + g_{2,1}(n) + g_{2,2}(n) \tag{16}$$

Where

$$g_{2,0}(n) = \sum_{i=1}^{3} \frac{V_i^2}{2}, \tag{17}$$

$$g_{2,1}(n) = \sum_{i=1}^{3} \sum_{j \neq 1} V_i V_j \cos(a_i \pm a_j), \tag{18}$$

$$g_{2,2}(n) = \sum_{i=1}^{3} V_i^2 \frac{\cos(2a_i)}{2}. \tag{19}$$

The third order component includes the third-order harmonic components $g_{3,1}(n)$, the third intermodulation beat components $g_{3,2}(n)$, the triple beat components $g_{3,3}(n)$, the self-compression/expansion components $g_{3,4}(n)$, and the cross-compression/expansion components $g_{3,5}(n)$. Specifically,

$$k_3 x^3(n) = g_{3,1} + g_{3,2}(n) + g_{3,3}(n) + g_{3,4}(n) + g_{3,5}(n) \tag{20}$$

Where

SCTE•ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 2020

2020 Fall
Technical Forum
SCTE·ISBE • NCTA • CABLELABS®

$$g_{3,1}(n) = \sum_{i=1}^{3} \frac{V_i^3}{4} \cos(3a_i), \tag{21}$$

$$g_{3,2}(n) = \sum_{i=1}^{3} \sum_{j \neq 1} \frac{3V_i^2 V_j}{4} \cos(2a_i \pm a_j), \tag{22}$$

$$g_{3,3}(n) = \sum_{i=1}^{3} \sum_{j \neq 1} \sum_{k \neq 1} \frac{3V_i V_j V_k}{2} \cos(a_i \pm a_j \pm a_k), \tag{23}$$

$$g_{3,4}(n) = \sum_{i=1}^{3} \frac{3V_i^3}{4} \cos(a_i). \tag{24}$$

$$g_{3,5}(n) = \sum_{i=1}^{3} \sum_{j \neq 1} \frac{3V_i V_j^2}{2} \cos(a_i). \tag{25}$$

The simulated output containing continuous wave signals (CWs) and NLD aligns with measurement in Figure 8, given coarse approximations for nonlinear gain coefficients and odd-order memory, based on model described in (8). A Rohde and Schwarz DOCSIS Cable Load Generator (CLGD) generated the CWs. The CWs propagated through a nonlinear power amplifier. The resultant nonlinear output spectrum was measured using a Keysight vector signal analyzer, model MXA, running in spectrum analysis mode.
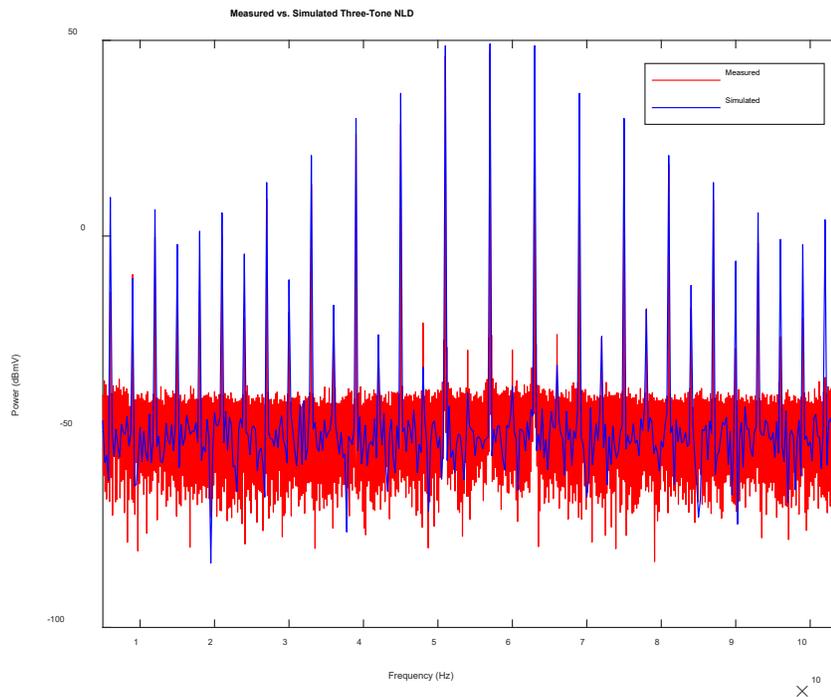
Figure 8 - 3 Tone NLD Measured vs. Simulated

5. NLD Mitigation

The focus of this paper up to this point has been on the levers associated with PA efficiencies (a) near saturation operation, and more efficient implementations via (b) PA classes, which may include (c) biasing the PA to consume less power. All these approaches result in degraded NLD, which becomes even more challenging as modern-day orthogonal frequency division multiplexing (OFDM) signals, which have a higher peak-to-average power ratio (PAPR), become more ubiquitous in communication network payloads [15]. One of the major design goals modern systems is to make the communication systems more power efficient. This needs efficient PAs, which is unfortunately more challenging since OFDM has much higher PAPR and wider bandwidth [11].

We will next explore a portfolio of current methods focused on harvesting PA efficiency and in most cases include mitigating NLD. These methods have enabled network designers to push the network operating boundaries that were previously constrained by lower amounts of NLD. These methods will include PAPR Reduction, Transmitter DPD, and Receiver Post-Distorter Equalization.

5.1. Peak-to-Average-Power-Ratio (PAPR) Reduction

Figure 9 illustrates how PAPR can interact with PA characteristics. OFDM signals can have a high degree of PAPR and push the PA operation into saturation at maximum signal amplitudes, also illustrated in Figure 9. PAPR reduction methods work to minimize the signal's PAPR, so that operation at PA saturation can be avoided, like OBO. There are multiple methods available to achieve PAPR reduction, as we will see, each with their own tradeoffs in benefit and cost.

SCTE·ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 \ 2020

2020 Fall
Technical Forum
SCTE·ISBE • NCTA • CABLELABS®
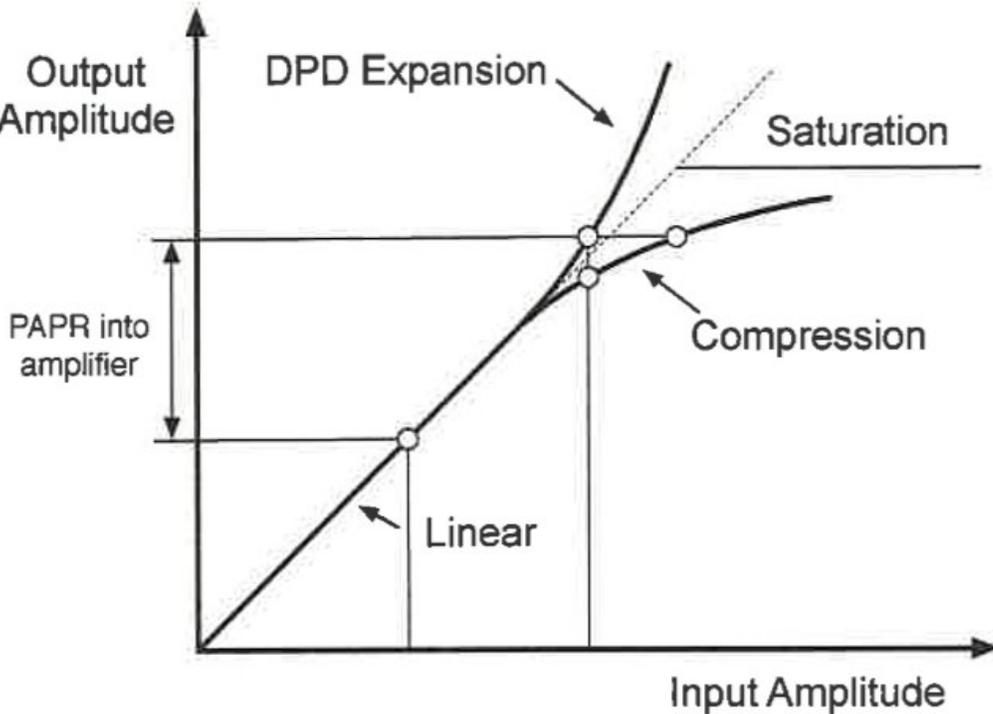
Figure 9 - PA Characteristics with PAPR and DPD

PAPR reduction was discussed during the deliberations of the DOCSIS 3.1 standard development, which introduced OFDM signaling into its portfolio of physical layer technology (PHY) [7]. The concern was that the increased PAPR of OFDM exceeded that of previous generation DOCSIS 3.0, SC-QAM on a 6 MHz bandwidth basis. However, the standards group ultimately decided against using PAPR reduction, primarily because the impact to CATV PAs would be negligible, given their broadband nature, which is an aggregate of over one hundred 6 MHz channels, up to approximately 1 GHz of bandwidth, total. Through lab measurement, these conclusions have been validated, but actual field results at scale have yet to be made [15].

High data rates have led to complex modulation schema, and ultimately higher spectral efficiency. One consequence of using spectrally efficient modulation schemes is that the dynamic range of the signal may be quite high [10].   This is generally measured in terms of the ratio of the peak signal power to the average power of the modulated signal, or PAPR [10]. While a large PAPR is not such a problem for signal transmission, it can have an impact on the efficiency of the transmitter PA [10], as illustrated in Figure 9, where the signal peaks can push the PA into saturation if the signal OBO isn't sufficiently below the PA compression point. Reducing the PAPR of the input signal using digital signal processing allows the PA to be operated at a higher efficiency, and it also reduces the dynamic range needed to represent the input signal digitally [10]. In addition, the reduced PAPR will often reduce the complexity of the linearization approach needed to compensate for PA nonlinearities [10].

The PAPR of a signal $x(n)$ is illustrated in (26), where $E[\cdot]$ denotes short-term expectation or average [10].

$$PAPR = \frac{\max(|x|^2)}{E[|x|^2]} \qquad (26)$$

(27) illustrates how PAPR is based on the statistics of the signal rather than the absolute peak, while the practical peak is the level, L, at which the signal magnitude has a 10^{-4} probability, P, of exceeding.

$$P\{\,|x|^2 > L\} = 10^{-4} \qquad (27)$$

The complementary cumulative density function (CCDF) of $|x|^2$, is a useful description of the signal statistics, often compared to that of a Gaussian waveform because the statistics of multi-carrier signals used in many of today's communications networks tend to approach that of a Gaussian [10]. Figure 10 is a Keysight PXA PAPR measurement for a 6 MHz, DOCSIS downstream SC-QAM signal whose PAPR (yellow line) is approximately 9.44 dB based on a 10 Mpt (million point) sample period [15]. Changing from DOCSIS SC-QAM to an OFDM-based PHY and limiting measurements to 6 MHz bandwidth, shows that PAPR will increase by approximately 0.96 dB, using M=256 QAM constellation levels for both signal types. However, much broader bandwidth comparisons over 192 MHz revealed that OFDM PAPR was lower than SC-QAM, by about 0.52 dB [15].
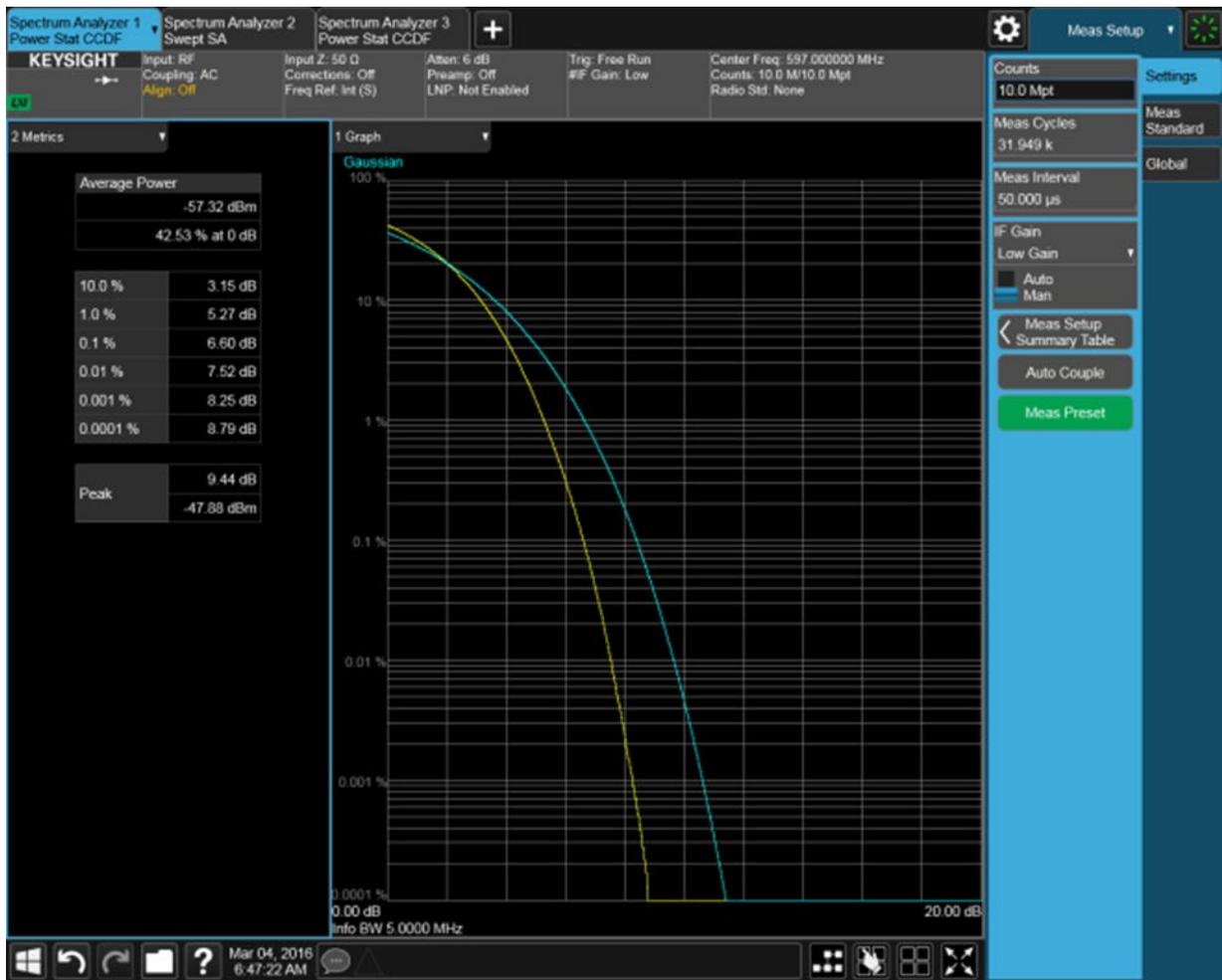
Figure 10 - DOCSIS Downstream SC-QAM PAPR Measurement

Many techniques have been developed for PAPR reduction, such as signal clipping, peak cancellation, and error waveform subtraction (noise shaping) [10]. These clip-and-filter approaches clip peaks exceeding a specified level and filter the waveform to remove out-of-band distortion [10]. Clip-and-filter and peak windowing are the easiest Crest Factor Reduction (CFR) methods to implement, making them the most likely to be used [10].

For OFDM-based formats, CFR is often achieved using redundant coding or the transmission of auxiliary information, both of which reduce data throughput of the system. Pilot tones and unmodulated subcarriers can be exploited to reduce PAPR with some special pre-coding techniques [16]. The selected mapping approach (SLM) from [16] provides good PAPR reduction performance, but may suffer from high computational complexity from using a bank of inverse fast fourier transforms (IFFTs), illustrated in Figure 11.

SCTE•ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 2020

2020 Fall
Technical Forum
SCTE·ISBE • NCTA • CABLELABS®
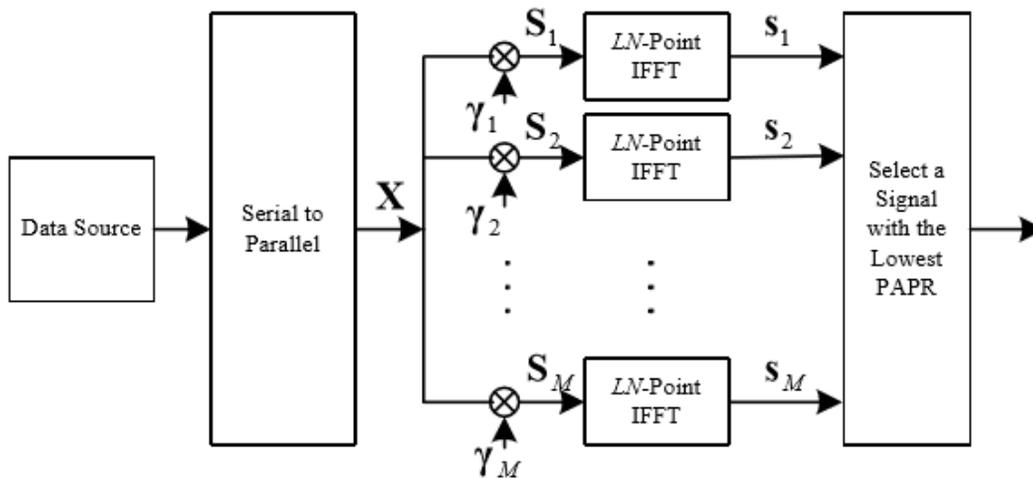
Figure 11 - Traditional Selected Mapping, PAPR Reduction via Pre-Coding

The new SLM approach from [16], shown in Figure 12, replaces IFFTs with new kinds of conversions, resulting in much lower complexity. Multiple IFFTs are replaced with transformation matrix, T_r, to produce candidate signals [16]. PAPR reduction for this approach is almost as good as traditional SLM approaches but uses much less processing overhead [16].
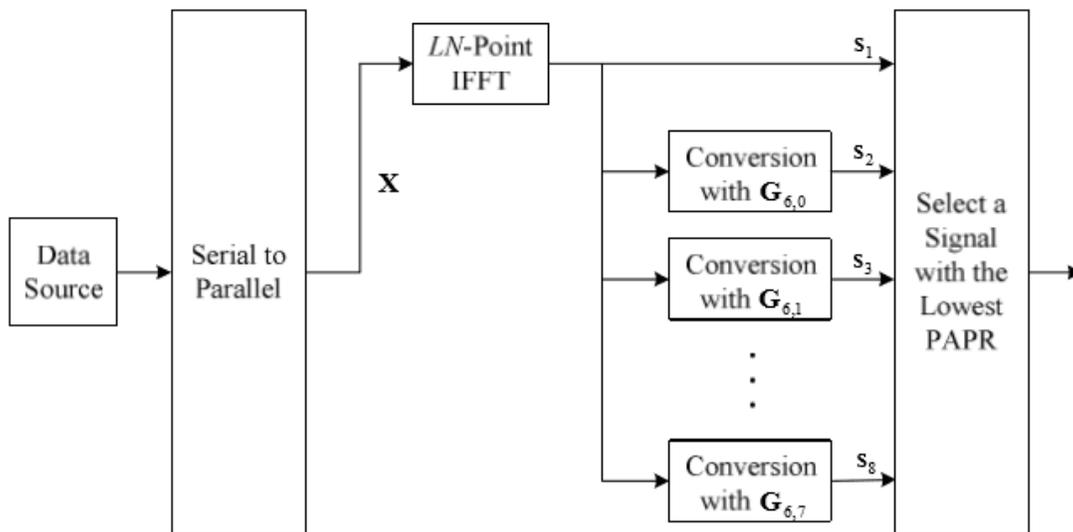
Figure 12 - Transformation Based Pre-Coding for SLM PAPR Reduction

CFR techniques are very appropriate for many wireless communications applications, such as cellular wireless handset PA applications, where Class AB operations are typically used [10]. CFR is a form of digital signal processing applied to the digital signal $x(n)$, used in combination with DPD to reduce the requirements on the RF PA within the transmit chain [10]. Although the distortionless PAPR reduction methods decrease the deleterious effect of nonlinear distortions, their effectiveness in improving the system performance is limited, since the main problem of the limited dynamic range of the PA remains unsolved [17]. Therefore, CFR often precedes DPD, per Figure 13. When CFR is used in conjunction

SCTE•ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 \ 2020

2020 Fall
Technical Forum
SCTE·ISBE · NCTA · CABLELABS®

with DPD techniques, the expansive nature of the DPD function operates on the crest factor reduced signal, so that the resulting pre-distorted signal that enters the PA does not have an excessive PAPR, and the PA can still be driven hard to operate at its highest efficiency [10]. This behavior is also illustrated in Figure 9.
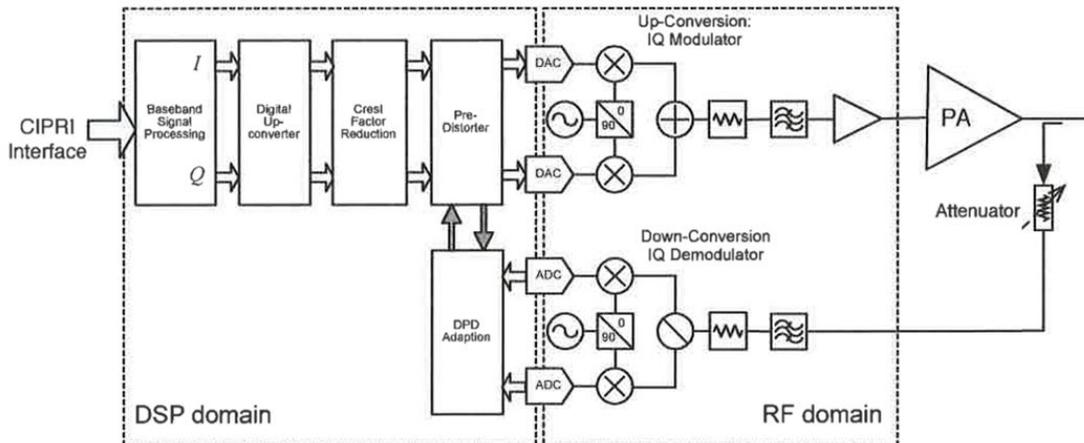
**Figure 13 - High-Level Transmitter with CFR and Adaptive DPD**

CFR assumes that reducing the peak of the signal is beneficial to the performance of the digital transmitter, essentially allowing the linearized PA to operate at a higher efficiency, while meeting the linearity requirements of the modulation format used [10]. CFR has drawbacks which degrade the system performance, either by increasing in-band degradation or reducing data throughput, but the overall system can be optimized to meet the performance specifications at higher PA efficiencies [10]. When applied properly, CFR allows the PA to operate more efficiently, thereby improving the performance of the transmitter [10].

5.1. Transmitter Digital Pre-Distortion (DPD)

One of the dominating practices in the cellular and satellite industries today is to insert a DPD circuit before the PA, which distorts the signals appropriately to compensate for the nonlinear PA response [11]. The distortion is added in such a way as to cancel the inherent nonlinearity of the PA, so that its output is a linear replica of the original input signal. DPD has been applied widely in many modern transmitters [2]. DPD can lead to the use of more efficient and cost-effective PAs [11] and is being considered for future generations of FDX RPD node hardware, where power consumption thresholds are already being encroached upon, while new capability is being added to increase its capacity as efficiently as possible.

SCTE·ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15   2020

2020 Fall
Technical Forum
SCTE·ISBE · NCTA · CABLELABS®
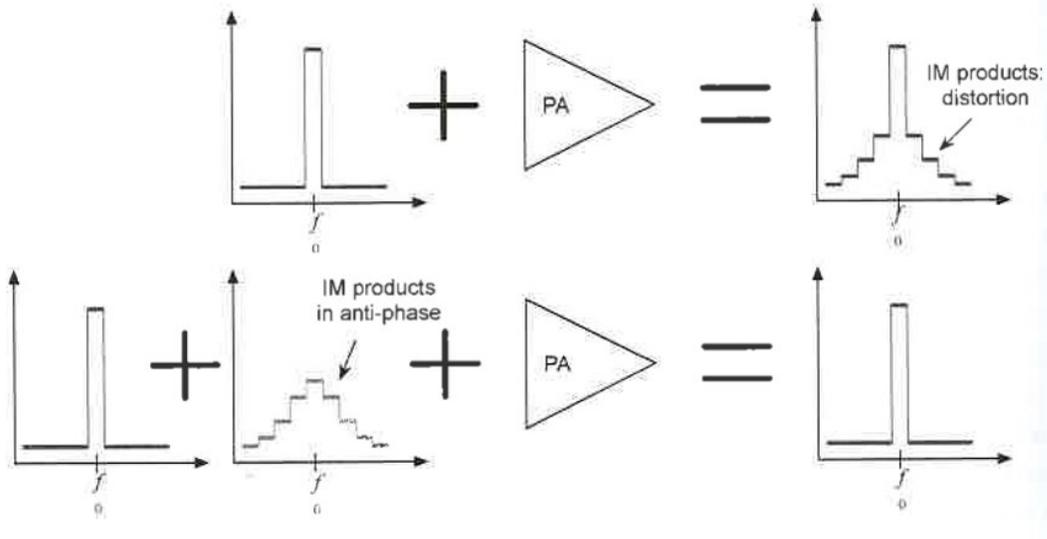
Figure 14 - Nonlinearized PA vs. Linearized PA via Digital Pre-Distortion

We will now focus on current approaches available to mitigate the effects of NLD, which are based on equalization, or creating an anti-phase version of NLD, as in the bottom part of Figure 14, which, when added to the signal from which it was derived, can negate the nonlinear effects of the PA. This essentially linearizes the PA's behavior.

In Figure 14, equalization is being performed prior to the signal's exposure to NLD, or pre-equalization. Pre-equalization of NLD is typically accomplished using DPD technology. DPD is analogous to DOCSIS Transmit Pre-Equalization, and the key difference is in the type of distortion that gets mitigated. DOCSIS Transmit Pre-Equalization mitigates linear distortion (LD), or plant echoes and other filter effects including amplitude roll-off, and group delay variation [7]. The key similarity here is that both signal processing techniques are applied as a signal bias, prior to the signal's impairment exposure, impairments being NLD for DPD, and LD for transmit pre-equalization.
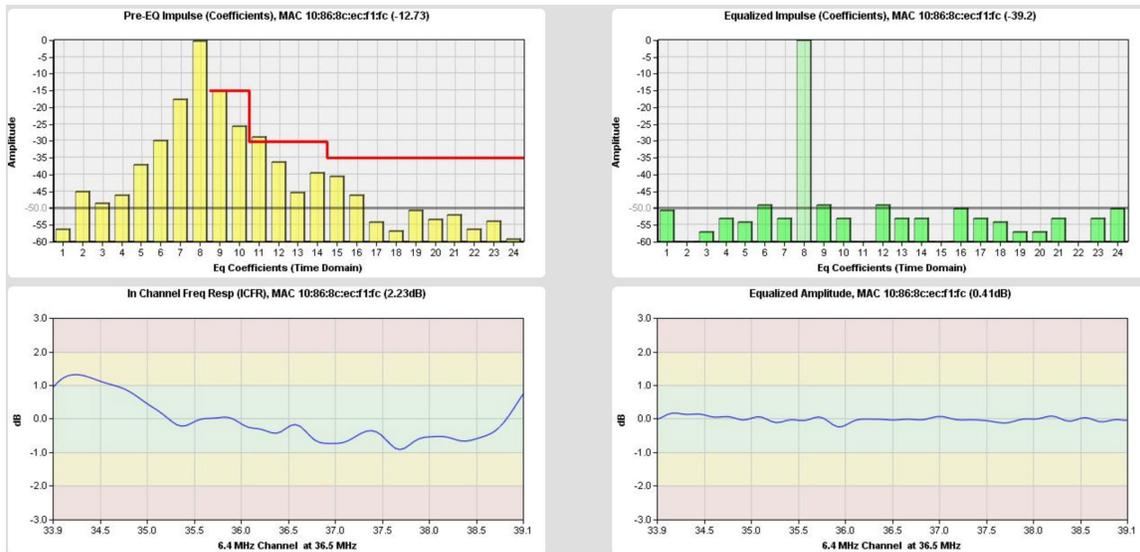
Figure 15 - CM Upstream Transmit Pre-Equalization and Post-Equalization Functions

Equalization functions can also be applied to the signal after impairment exposure, and this is known as post-distorter equalization for NLD and post-equalization for LD. In DOCSIS upstream communications, both Transmit Pre-Equalization and post-equalization functions work collaboratively to mitigate the effects of LD [7]. Figure 15 illustrates these two functions together compensating for the communication channel that exists between, say, my CM, and the CMTS that serves multiple neighboring towns, including my own. CMTSs may connect to hundreds of nodes, as discussed in earlier sections, and ultimately connecting to thousands of CMs. The CMTS's primary function is that of a router, facilitating communication between the local area network (LAN), comprised of many CMs, including mine, and the wide area network (WAN) or internet.

The collaboration between the CM and CMTS on how to equalize upstream LD is specified in DOCSIS, where transmit pre-equalization is a CM function and its DOCSIS 2.0, 24 symbol-spaced coefficients, shown in the upper left side of Figure 15 in yellow, are provided to it by the CMTS [7]. The CMTS post-equalization function is shown in the upper right side, in green. While performing its own post equalization function, the CMTS periodically sends a set of equalizer coefficients to the CM, via station-maintenance messages, with instructions to either overwrite or convolve the CM's current set of equalizer coefficients with the new coefficients.

An effect of this collaboration is to perform most of the channel equalization at the CM's transmit pre-equalization function, leaving only minor corrections in the post equalization function, at the CMTS. The equalizer's coefficients are colored yellow because of the intensity of correction, or, in other words, the variation of its amplitude frequency response, illustrated in the lower left chart, which exceeds ±1 dB peak-to-peak. Figure 15 illustrates how the CM is compensating for most of the channel's LD. Overall, LD equalization at both ends of the communication link has proven to be very robust and reliable in the CATV industry, but it does so with additional signal processing and overhead.

Adaptive pre-distortion techniques include an observation path that samples the output from the power amplifier and feeds the signal back to the pre-distorter to estimate system nonlinearity [10]. Additional components include analog-to-digital converters (ADCs) which are used to convert the observation signal back to baseband and digitize for digital domain operations. Components in this path will have better linearity than the desired performance of the transmitter to avoid introducing distortion at the PA output

arising from the observation process [10]. Input and feedback signals must be synchronized, accounting for the group delay of the PA, which tends to dominate the loop [10].

before DPD

(a)

after DPD

(b)

under-sampling

(c)

over-sampling

(d)
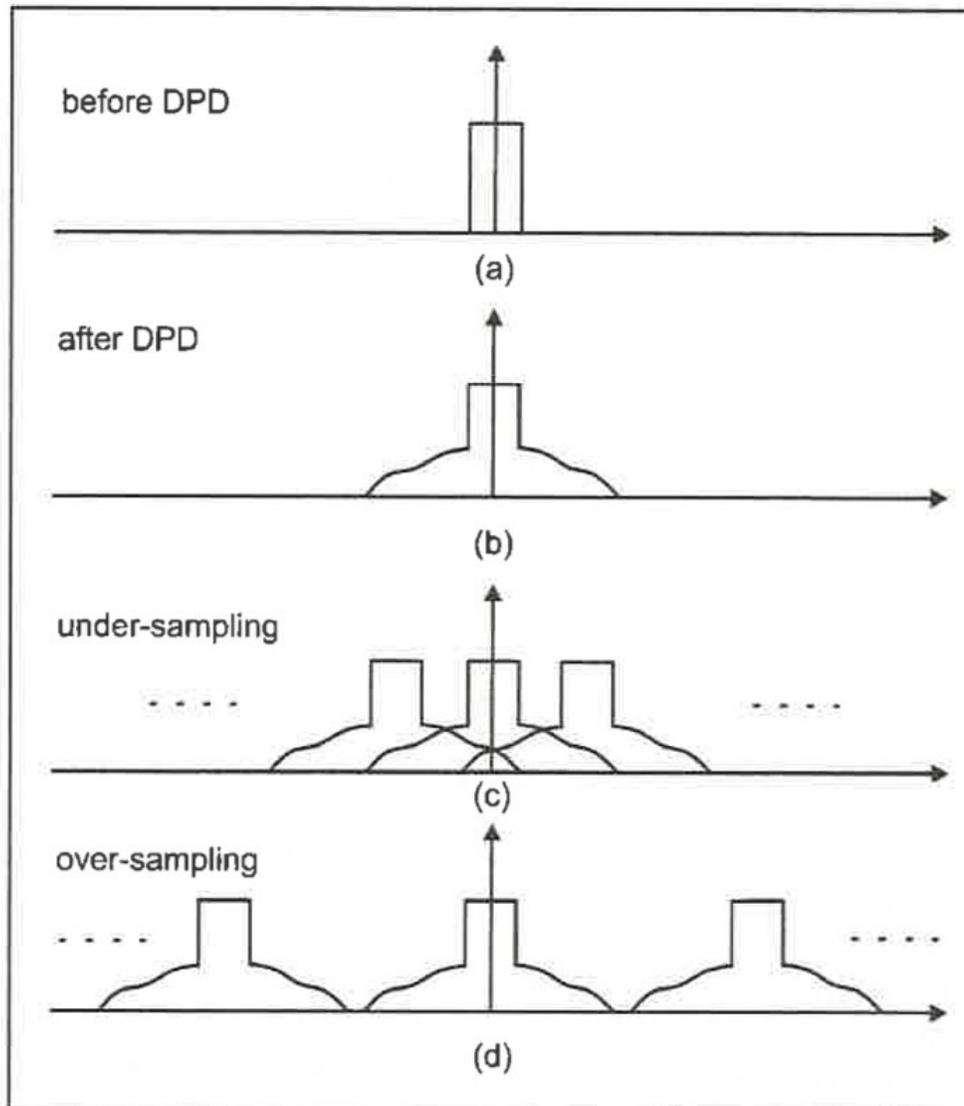
Figure 16 - Effects of Sample Rate on a Pre-Distorted Signal

Pre-distorters reconstruct the NLD in the analog domain, for input into the PA chain in the transmitter, which needs to be sampled at a higher rate to accommodate for the increased bandwidth. Therefore upsampling is required, otherwise under sampling the NLD could lead to aliasing in Figure 16 [10]. The output signal from the pre-distorter (b) has a much wider bandwidth than the input signal (a). Under-sampling of the input will lead to aliasing of the output signal (c), and the clean DPD. Typical oversampling by 5x will represent 5th order NLD [10]. Two main approaches have been adopted in the pre-distorter block:
    (1) Look-up tables (LUTs) to provide a map relating the desired pre-distorter output to the input voltage

SCTE•ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 \ 2020

2020 Fall
Technical Forum
SCTE·ISBE • NCTA • CABLELABS®

(2) Nonlinear basis functions to describe nonlinear pre-distortion function, requiring generation of the nonlinear functions and the multiplication of the basis functions by the input voltage [10].

Many PA models can describe the nonlinear pre-distortion function, including the Volterra series and Wiener model [19]. PA models will need to consider that PA characteristics do not change rapidly with time; changes in PA characteristics are often attributable to temperature drift, aging, etc., which have long time constants [19]. The cause of memory effects can be electrical or electrothermal [19]. Higher output power PA operation, such as those used in wireless base-stations, exhibit memory effects [19]. Having memory means that the output of the PA is not only a function of the current input, but also a function of past inputs and outputs [18]. Memory effects in the power amplifier limit the performance of DPD for wideband signals [18].
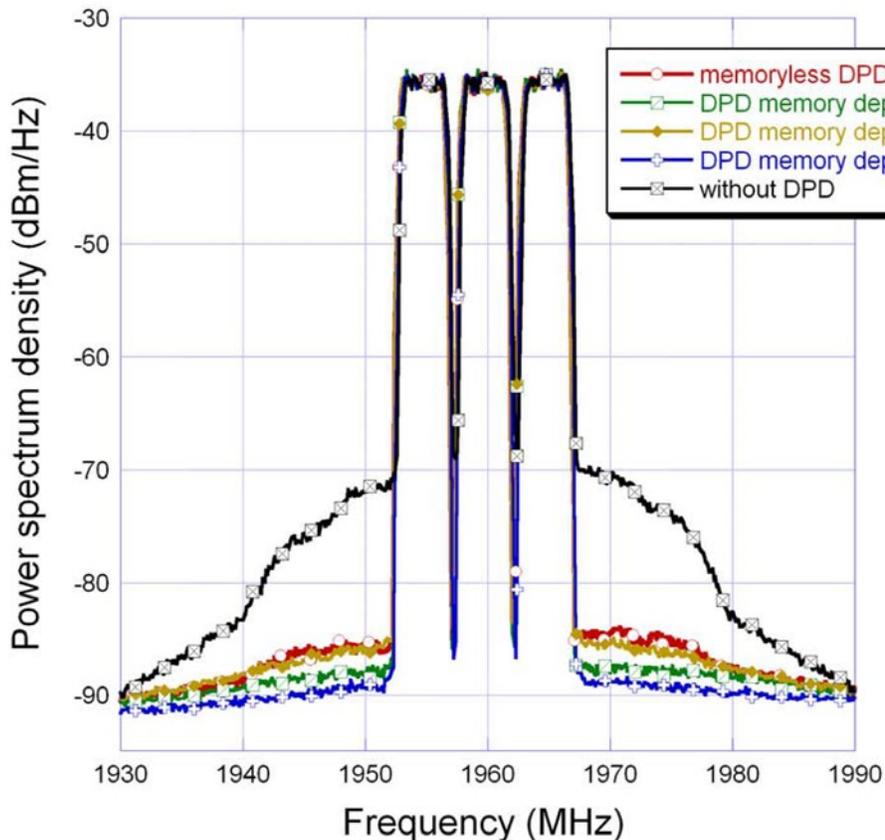
Figure 17 - Memory and Memoryless DPD Results

An issue with DPD is that it is compensating for the PA in the transmitter only, which, in most practical networks today, represents a fraction of the nonlinearity present in the communication channel. PAs are at least used in both the transmitter and receiver. Multiple PAs could be used in between the transmitter and receiver to extend the reach in many communication networks, such as CATV networks. Passive components are used to distribute signals to many receivers, and may also contain nonlinear components, such as inductors, that contribute to the channel's aggregate nonlinear NLD.

AI methods for DPD, using neural networks, illustrate the potential for improvement for NLD performance over traditional methods [20] [21]. However, incorporating additional NLD compensation

into the transmitter to account for the rest of the communication network chain, like similar closed-loop equalization systems, could lead to DPD reaching its full potential.

Implementations of DPD should be architected to compensate for stronger nonlinearity, much like the closed-loop LD equalization strategy described in the beginning of this section, for DOCSIS upstream signals can compensate for appreciably high LD. Because of bidirectional network connectivity, both the CMTS and CM can collaborate on the estimation of the total NLD present within the communication channel, and send coefficients via the downstream communication path, with instructions to either convolve or overwrite coefficients describing the estimated path NLD that will account for all the nonlinear elements within the communication network chain (transmitter, receiver, amplifiers, and nonlinear passive components). For this strategy to work, receivers must be capable of mitigating NLD. We will next review the current solutions available in receiver-based post distortion cancellation.

5.1. Receiver Post-Distorter Equalization

Another strategy is to mitigate the nonlinear PA distortions at the receivers via post-distorter equalization [22] [23] [24]. The solution presented in [25] develops a Bayesian signal detection algorithm, based on the nonlinear response of the PAs. However, this documented approach applies to a simple "AM-AM, AM-PM" nonlinear PA model only.

The authors of [22] propose a symbol-based equalizer, with nonlinear distortion cancellation for the forward link as an addition to the standard linear equalizer at the receiver, suitably adapted to incorporate specific channel functions in the forward link, including input multiplexing (IMUX) and output multiplexing (OMUX) filtering and the traveling wave tube amplifier (TWTA), using the memory polynomial model. The proposed setup is compared with current mitigation approaches, yielding significant efficiency gains [22]. The nonlinearity in the satellite channel is introduced primarily by the TWTA [22].

The objective of [22] was to compare the performance of the proposed equalizer with the state-of-the-art dynamic data pre-distortion, and to show that the best system performance is achieved when both pre-distortion at the transmitter and decision-directed equalization at the receiver are applied. In addition, the performance of a simple maximum likelihood (ML) demodulator in the detector in the cancellation loop is compared against a low density parity check (LDPC) decoder, to show the robustness of the proposed equalizer to decision errors [22]. Overall system complexity of the proposed nonlinear equalizer is argued to be less than current linear equalization approach, due to less-frequent updates [22]. As discussed earlier in the section covering DPD, similar transmitter and receiver equalization loops have proven to be robust against LD for DOCSIS upstream [7].

References [23] [24] and [25] approach receiver post NLD equalization via clustering methods, which is a subset of broader suite of AI models. [23] leverages SVMs, while [24] uses a radial basis function (RBF) network and [25] is Bayesian. ANNs have attracted researchers in the field of PA modeling, due to its successful implementation in pattern recognition, signal processing, system identification, and control [26].

6. Severe NLD Mitigation

One of the major design goals for modern systems is to make the communication systems more power efficient. This needs efficient PAs, which is unfortunately more challenging, since many modern PHY include OFDM, which has much higher PAPR and wider bandwidth [11].

Existing nonlinear PA mitigation strategies -- DPD, PAPR reduction, and receiver-based equalization, discussed in Section 5 -- may not be sufficient enough when considered individually. We can reduce PAPR to some extent only. DPD is too complex and costly for small and low-cost devices. Existing DPD and equalization techniques have moderate nonlinear distortion compensation capabilities, because they have been designed to cancel internal transmitter NLD only.

There is a larger system that must be considered, consisting of a transmitter, receiver, and nonlinear active/passive components in between them, all contributing to the overall NLD observed at the receiver. In some cases, the NLD can be quite severe, due to required higher-efficiency modes of operation. Supporting collaboration between the transmitter and receiver, like the LD cancellation systems used for DOCSIS, may be the key to achieving optimal network efficiency, while minimizing NLD. Severe nonlinearity estimation and mitigation is a requirement for the receiver, representing information that could then be shared with the transmitter through bidirectional communication. Ultimately, this lessens the burden at the receiver location and the NLD equalization system overall.

In this section, a system for cancelling severe NLD will be proposed which develops nonlinear equalizers that exploit both deep neural networks (DNNs) and Volterra series models to mitigate PA nonlinear distortions. The DNN equalizer architecture consists of multiple one-dimension convolutional layers. The input features are designed according to the Volterra series model of nonlinear PAs. This enables the DNN equalizer to mitigate nonlinear PA distortions more effectively, while avoiding over-fitting under conditions of limited training data. Experiments are conducted with real measurement data obtained from a highly nonlinear RFMD RF2317 Linear CATV Amplifier [11]. The results will demonstrate that the proposed DNN equalizer has superior performance over conventional equalization approaches, a necessary tool for more collaborative NLD mitigation strategy that could make more efficient network components potentially realizable.

6.1. Enhanced Equalization – Integrating Volterra Series and DNNs

Reference [11] proposes the use of DNNs to implement the nonlinear equalizer in the receiver, which can mitigate the nonlinear effects of the received signals, due to not only PAs but also nonlinear channels and propagations. The architecture of the DNN equalizer is shown in Figure 18.
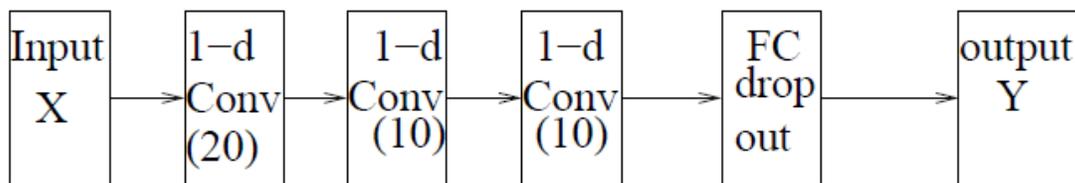
Figure 18 - Block Diagram of DNN Equalizer

Different from [28], [11] used multi-layer convolutional neural networks (CNNs). Different from conventional neural network predistorters proposed in [29], [11] used neural networks as equalizers at the receivers. Different from conventional neural network equalizers such as those proposed in [30] [31], in a DNN equalizer [11], a CNN is used and the input features in X are not only the linear delayed samples $r(n)$. Rather, Volterra series models are used to create input features.

SCTE•ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 \ 2020

2020 Fall
Technical Forum
SCTE·ISBE • NCTA • CABLELABS®
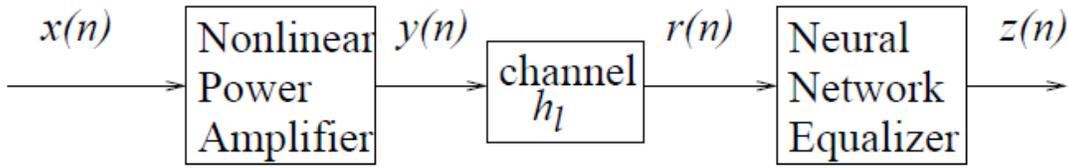
Figure 19 - System Block Diagram with Nonlinear Power Amplifier and Deep Neural Network Equalizer

To simplify presentation, according to the previous section, [11] assumes that the linear channel H has already been equalized by a linear equalizer of Figure 19, whose output signal is $r(n)$. According to Volterra series representation of nonlinear functions, the input-output response of the nonlinear equalizer can be written as

$$z(n) = \sum_{k=1}^{P} \sum_{d_1=0}^{D} \cdots \sum_{d_k=0}^{D} f_{d_1,\ldots,d_k} \prod_{i=1}^{k} r(n - d_1) \tag{28}$$

One of the major problems is that the number of coefficients $f_{d_1,\ldots,d_k}$ increases exponentially with the increase of memory length D and nonlinearity order P. There are many different ways to develop more efficient Volterra series representations with reduced numbers of coefficients. For example, in [32], the authors exploit the fact that higher-order terms do not contribute significantly to the memory effects of PAs to reduce the memory depth d when the nonlinearity order k increases. This technique can drastically reduce the total number of coefficients. In [33] [34] [35], the authors developed the dynamic deviation model to reduce the full Volterra series model (28) to the following simplified one

$$z(n) = z_s(n) + z_d(n)$$

$$= \sum_{k=1}^{P} f_{k,0} r^k(n) + \sum_{k=1}^{P} \sum_{j=1}^{k} r^{k-j}(n) \sum_{d_1=0}^{D} \cdots \sum_{d_j=d_{j-1}}^{D} f_{k,j} \prod_{i=1}^{j} r(n - d_i) \tag{29}$$

where $z_s(n)$ is the static term, and $z_d(n)$ is the dynamic term that includes all the memory effects. We can see that the total number of coefficients can be much reduced by controlling the dynamic order j which is a selectable parameter.

[11] constructs the input features of the DNN based on the model (29). Corresponding to the static term $z_s(n)$, [11] changes it to

$$\hat{z}_s(n) = \sum_{1 \leq k \leq P} f_{k,0} r(n) |r(n)|^{k-1} \tag{30}$$

The reason that (30) changes $r^k(n)$ to $r(n)|r(n)|^{k-1}$ is that only the signal frequency with the valid passband is of interest. This means that input feature vector X should include terms $r(n)|r(n)|^{k-1}$.

Similarly, corresponding to the dynamic term $z_d(n)$, we need to supply $r^{k-j}(n) \prod_{i=1}^{j} r(n - d_i)$ in the features where half of the terms $r(n)$ and $r(n - d_i)$ should be conjugated. For simplicity, in DNN equalizer used in [11], the vector X includes $r(n - q)|r(n - q)|^{k-1}$ for some q and k.

By applying Volterra series components directly as features of the input X, the DNN can develop more complex nonlinear functions with fewer hidden layers and fewer neurons. This will also make the training procedure converge much faster, with much less training data.

In Figure 18, the input X is a tensor formed by the real and imaginary parts of $r(n - q)|r(n - q)|^{k-1}$ with appropriate number of delays q and nonlinearities k. There are three one-dimension convolutional layers, each with 20 or 10 feature maps. After a drop-out layer for regularization, this is followed by a fully connected layer with 20 neurons. Finally, there is a fully connected layer to form the output tensor Y which has two dimensions. The output Y is used to construct the complex $z(n)$, where $z(n) = \hat{x}(n - d)$ for some appropriate delay d. All the convolutional layers and the first fully connected layer use the sigmoid activation function, while the output layer uses the linear activation function. [11] uses the mean square error loss function $L_{loss} = E[|x(n - d) - z(n)|^2]$, where $z(n)$ is replaced by Y and $x(n - d)$ is replaced by training data labels.

Measurement signals were obtained from an implementation of a RFMD RF2317 PA used in the cable industry, which are typically dominated by 3rd order nonlinearities. Various levels of nonlinear distortion, in terms of dBc, were generated by adjusting the PA RF input levels [11].

For the Volterra equalizer, reference [11] approximated the response of the nonlinear equalizer with delays including 8 pre- and post- main taps, and with nonlinearity including even and odd order nonlinearity up to the 5th order. To determine the values of the Volterra coefficients, we transmitted $N = 4,096$ training symbols through the PA and then collected the noisy received samples $r(n)$.

For conventional time-delay NN equalizer, [11] applied a feedforward neural network with 80-dimensional input vector X and 5 fully connected hidden layers with 20, 20, 10, 10, 10 neurons, respectively.
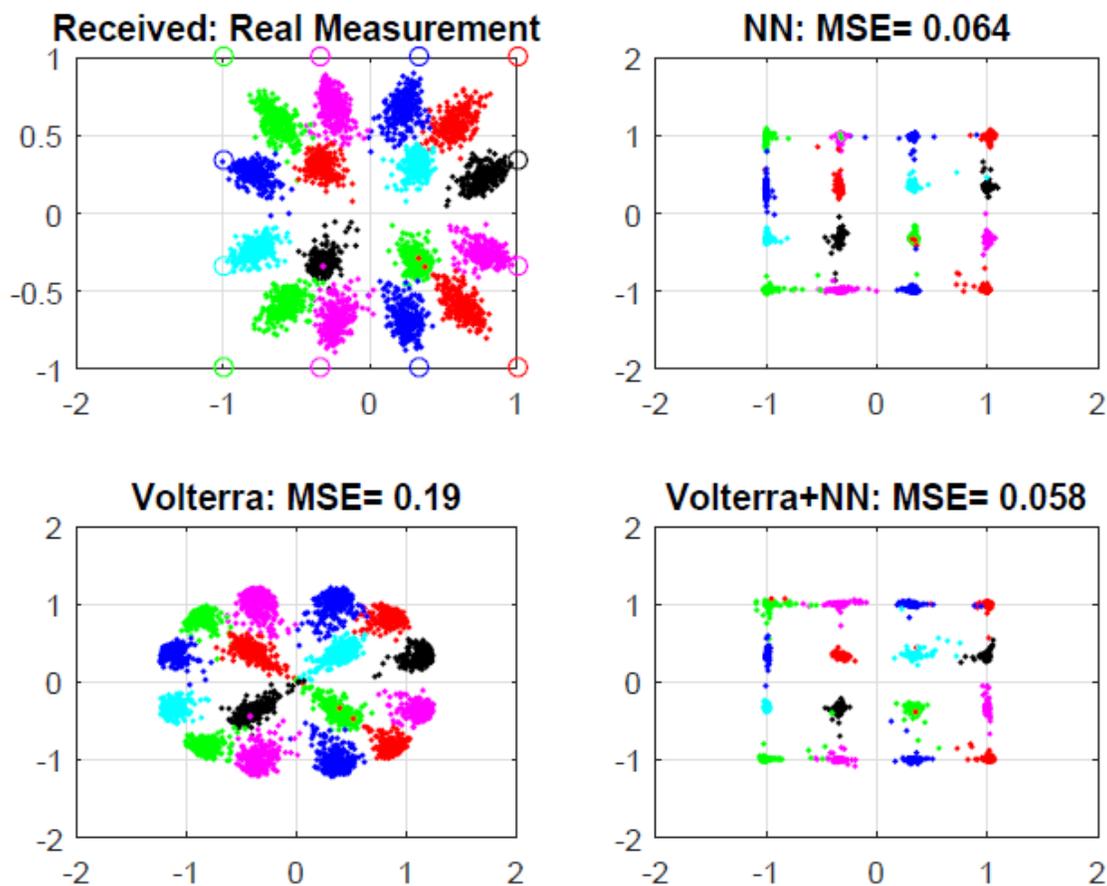
**Figure 20 - Constellation of 16-QAM Non-Equalized vs. Equalized**

Figure 20 shows the constellation of 16-QAM equalization over the real PA. The corresponding SER were 0.0067, 0.0027, 0.00025 respectively. Nonlinear filtering of both a DNN and DNN with Volterra input features show superior equalization over just Volterra filtering alone. Further, it can be seen that the proposed Volterra+NN scheme has the best performance.

SCTE•ISBE
CABLE-TEC EXPO®
VIRTUAL EXPERIENCE » OCTOBER 12-15 2020

2020 Fall
Technical Forum
SCTE·ISBE • NCTA • CABLELABS®
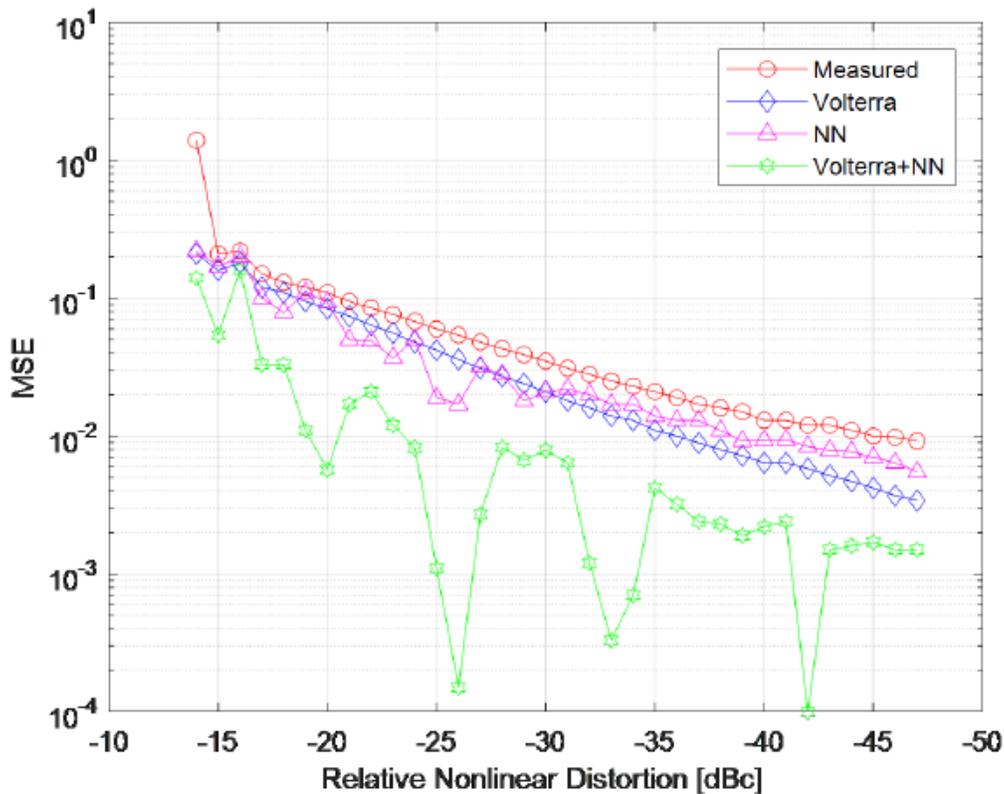
Figure 21 - Comparing Three Equalization Methods for 16-QAM under Various NLD Levels

Figure 21 provides MSE measurements for 16-QAM under various nonlinear distortion levels, dBc. For each 1 dB increase in NLD, the resultant MSE is shown for the "Measured", "Volterra", "NN", and the proposed "Volterra+NN" cases. MSE reduction diminishes appreciably as modulation order increases from QPSK to 64-QAM, but small improvements in MSE have been observed to lead to appreciable SER improvement, especially for more complex modulation orders. Unfortunately, the 4,096 symbol sample size limited measurements to a minimum measurable 0.000244 SER, which represents 1 symbol error out of 4,096 symbols.

Table 3 - Comparing MSE/SER Improvement % for the Three Equalization Methods

	Volterra		NN		Volterra+NN	
	MSE	SER	MSE	SER	MSE	SER
64-QAM	16%	26%	10%	25%	42%	44%
16-QAM	41%	2%	35%	6%	85%	28%
QPSK	57%	0%	100%	0%	100%	0%
AVERAGE	38%	9%	48%	10%	76%	24%

Table 3 summarizes equalization performance, which shows the average percent reduction/improvement in MSE and SER from NLD-impaired data for multiple modulation orders.

7. Conclusion

The enhanced capacity associated with FDX is increasing the need for higher efficiency PAs. A familiar scenario also playing out in the cellular, Wi-Fi, and satellite industries, where similar capacity-enhancing needs are increasing the needs for more efficient PAs. PA efficiencies can be realized by (a) near saturation operation, and more efficient implementations via (b) PA classes, which may include (c) biasing the PA to consume less power. All these approaches result in degraded NLD.

NLD mitigation techniques, like DPD, can lead to the use of more efficient and cost-effective PAs [11], and are being considered for future generations of FDX RPD node hardware -- where power consumption thresholds are already being encroached upon, while new capabilities are being added to increase their capacity as efficiently as possible. However, an issue with DPD is that it is compensating for the PA in the transmitter only, which, in most practical networks today, represents a fraction of the nonlinearity present in the communication channel. Incorporating additional NLD compensation into the transmitter, to account for the rest of the communication network chain, like similar closed-loop equalization systems discussed in this paper, could lead to DPD reaching its full potential.

Bidirectional network connectivity between the CMTS and CM can enable the convergence to an estimate of NLD present within the communication channel, in either direction, and send coefficients via the downstream communication paths, with instructions to either convolve or overwrite coefficients describing the estimated path NLD that will account for all the nonlinear elements within the communication network chain (transmitter, receiver, amplifiers, and nonlinear passive components).

However, for this strategy to work, receivers must be capable of mitigating severe NLD. Supporting collaboration between the transmitter and receiver, like the LD cancellation systems used for DOCSIS, may be the key to achieving optimal network efficiency, while minimizing NLD.

Thus, more aggressive NLD cancellation methods may be accomplished by advanced DNN approaches, such as incorporating input features derived from Volterra series models. Results from [11] demonstrate that the proposed DNN equalizer has superior performance over conventional equalization approaches, and is a necessary tool for more collaborative NLD mitigation strategy. Ultimately, this could make more efficient network components realizable.

Abbreviations

ACLR	adjacent channel leakage ratio
ACPR	adjacent channel power ratio
ADC	analog-to-digital converter
AI	artificial intelligence
ANN	artificial neural networks
AM-AM	amplitude modulation to amplitude modulation
AM-PM	amplitude modulation to phase modulation
Auto ML	automated machine learning
CAGR	compound annual growth rate
CATV	cable television
CCDF	complementary cumulative density function
CFR	crest factor reduction
CIN	composite-intermodulation-noise
CLGD	DOCSIS Cable Load Generator

CM	cable modem
CMTS	cable modem termination system
CNN	convolutional neural network
CPE	customer premise equipment
CW	continuous wave
dB	decibel
dBc	decibels relative to carrier power
dBmV	decibel-millivolts
DC	direct current
DNN	deep neural networks
DOCSIS	data over coax system interface specifications
DPD	digital pre-distortion
DTA	digital terminal adapter
EC	echo cancellation
EOL	end-of-line
EVM	error vector magnitude
FEC	forward error correction
FDX	full duplex DOCSIS
Hz	hertz
IEEE	Institute of Electrical and Electronics Engineers
IFFT	inverse fast fourier transform
IMD	intermodulation distortion
IMUX	input multiplex
IoT	internet of things
ISI	inter-symbol interference
LAN	local area network
LD	linear distortion
LDPC	low density parity check
LMS	least mean squares
LUT	look-up table
MER	modulation error ratio
MHz	mega-hertz
ML	maximum likelihood
Mpt	million point
mV	milli-volt
NF	noise figure
NLD	nonlinear distortion
OBO	output-power-back-off
OFDM	orthogonal frequency division multiplexing
OMUX	output multiplex
PA	power amplifier
PAPR	peak-to-average-power-ratio
PHY	physical layer
RBF	radial basis function
RF	radio frequency
RPD	remote PHY device
RRC	root-raised cosine
QPSK	quadrature phase shift keying

SC-QAM	single carrier quadrature amplitude modulation
SLA	service level agreement
SLM	selected mapping approach
SNR	signal-to-noise ratio
SNR_S	system signal-to-noise ratio
STB	set-top box
SVM	support vector machines
TG	task group
TWTA	traveling wave tube amplifier
VSA	vector signal analyzer
WAN	wide area network

Bibliography & References

[1] S. Haykin, Neural Networks: A Comprehensive Foundation. Prentice-Hall, Inc., 1999

[2] P. Winston, Artificial Intelligence, https://podcasts.apple.com/us/podcast/artificial-intelligence/id765641080, MITOPENCOURSEWARE, 2013

[3] OpenAI, Solving Rubik's Cube with a Robot Hand, https://openai.com/blog/solving-rubiks-cube/, 2019

[4] edureka!, Top 12 Artificial Intelligence Tools & Frameworks you need to know, https://www.edureka.co/blog/top-12-artificial-intelligence-tools/#scikit

[5] V. Fedak, Top 10 Most Popular AI Models, https://dzone.com/articles/top-10-most-popular-ai-models, 2018

[6] R. Thompson, C. Moore, J. Moran, R. Howald, Optimizing Upstream Throughput Using Equalization Coefficient Analysis, https://www.nctatechnicalpapers.com/Paper/2009/2009-optimizing-upstream-throughput-using-equalization-coefficient-analysis, 2009

[7] Cable Television Laboratories, Inc., Data Over Cable System Interface Specifications,

DOCSIS© 1.1, Radio Frequency Interface Specification, https://www.cablelabs.com/specifications/radio-frequency-interface-specification, September 7th, 2005

DOCSIS© 2.0, Radio Frequency Interface Specification, https://www.cablelabs.com/specifications/radio-frequency-interface-specification-2, April 22nd, 2009

DOCSIS© 3.0, Physical Layer Specification, https://specification-search.cablelabs.com/CM-SP-PHYv3.0, December 7th, 2017

DOCSIS© 3.1, Physical Layer Specification, https://specification-search.cablelabs.com/CM-SP-PHYv3.1, September 17th, 2019

DOCSIS© 4.0, Physical Layer Specification, https://www.cablelabs.com/specifications/CM-SP-PHYv4.0, April 29th, 2020

[8] Charles Warren, "How might we, three words that make design better", https://hbr.org/2012/09/the-secret-phrase-top-innovato, https://www.youtube.com/watch?v=mTpa-bJiMp4

[10] Cisco, "GS7000 1218MHz Fiber Deep Intelligent Node Data Sheet", https://www.cisco.com/c/en/us/products/collateral/video/gs7000-node/datasheet-c78-740828.html, January 9, 2019

[11] R. J. Thompson and X. Li, "Integrating Volterra Series Model and Deep Neural Networks to Equalize Nonlinear Power Amplifiers," in 2019 53rd Annual Conference on Information Sciences and Systems (CISS). IEEE, 2019, pp. 1–6.

[12] K. Simons, Technical Handbook for CATV Systems, 3rd Edition. Jerrod Publication No. 436-001-01, 1968.

[13] W. Ciciora, J. Farmer, D. Large, Modern Cable Television Technology; Voice, Video, and Data Communications. Morgan Kaufmann Publishers, Inc. 1999.

[14] S. Dimitrov, "Non-linear distortion cancellation and symbol-based equalization in satellite forward links," IEEE Trans Wireless Commun, vol. 16, no. 7, pp. 4489-4502, 2017

[15] R. Thompson, "Nonlinear Distortion Detection using DOCSIS Spectra", SCTE Conference, Philadelphia, 2016

[16] C.-L.Wangand Y.Ouyang,"Low-complexity selected mapping schemes for peak-to-average power ratio reduction in OFDM systems," IEEE Transactions on signal processing, vol. 53, no. 12, pp. 4652–4660, 2005.

[17] I. Yoffe and D. Wulich, "Predistorter for mimo system with nonlinear power amplifiers," IEEE Transactions on Communications, vol. 65, no. 8, pp. 3288–3301, 2017.

[18] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," Electronics Letters, vol. 37, no. 23, pp. 1417–1418, 2001.

[19] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," IEEE Transactions on communications, vol. 52, no. 1, pp. 159–165, 2004.

[20] M. Rawat, K. Rawat, and F. M. Ghannouchi, "Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real valued focused time-delay line neural networks," IEEE Transactions on Microwave Theory and Techniques, vol. 58, no. 1, pp. 95–104, 2010.

[21] F. Mkadem and S. Boumaiza, "Physically inspired neural network model for rf power amplifier behavioral modeling and digital predistortion," IEEE Transactions on Microwave Theory and Techniques, vol. 59, no. 4, pp. 913–923, 2011.

[22] S.Dimitrov,"Non-lineardistortioncancellationandsymbol-basedequalization in satellite forward links," IEEE Trans Wireless Communications, vol. 16, no. 7, pp. 4489–4502, 2017.

[23] D. J. Sebald and J. A. Bucklew, "Support vector machine techniques for nonlinear equalization," IEEE Transactions on Signal Processing, vol. 48, no. 11, pp. 3217–3226, 2000.

[24] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," IEEE Transactions on neural networks, vol. 4, no. 4, pp. 570–590, 1993.

[25] B. Li, C. Zhao, M. Sun, H. Zhang, Z. Zhou, and A. Nallanathan, "A Bayesian approach for nonlinear equalization and signal detection in millimeter-wave communications," IEEE Transactions on Wireless Communications, vol. 14, no. 7, pp. 3794–3809, 2015.

[26] M. Rawat, K. Rawat, and F. M. Ghannouchi, "Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real valued focused time-delay line neural networks," IEEE Transactions on Microwave Theory and Techniques, vol. 58, no. 1, pp. 95–104, 2010.

[27] PRNewswire, Small Cell Power Amplifier Market - Global Industry Analysis, Size, Share, Growth, Trends and Forecast 2017 – 2025, https://www.prnewswire.com/news-releases/small-cell-power-amplifier-market---global-industry-analysis-size-share-growth-trends-and-forecast-2017---2025-300546518.html, October 31, 2017

[28] B. Li, C. Zhao, M. Sun, H. Zhang, Z. Zhou, and A. Nallanathan, "A Bayesian approach for nonlinear equalization and signal detection in millimeter-wave communications," IEEE Transactions on Wireless Communications, vol. 14, no. 7, pp. 3794–3809, 2015.

[29] M. Rawat, K. Rawat, and F. M. Ghannouchi, "Adaptive digital predistortion of wireless power amplifiers/transmitters using dynamic real valued focused time-delay line neural networks," IEEE Transactions on Microwave Theory and Techniques, vol. 58, no. 1, pp. 95–104, 2010.

[30] D.-C. Park and T.-K. J. Jeong, "Complex-bilinear recurrent neural network for equalization of a digital satellite channel," IEEE Transactions on Neural Networks, vol. 13, no. 3, pp. 711–725, 2002.

[31] A. Uncini, L. Vecci, P. Campolucci, and F. Piazza, "Complex-valued neuralnetworkswithadaptivesplineactivationfunctionfordigital-radiolinks nonlinear equalization," IEEE Transactions on Signal Processing, vol. 47, no. 2, pp. 505–514, 1999.

[32] J. Staudinger, J.-C. Nanan, and J. Wood, "Memory fading Volterra series model for high power infrastructure amplifiers," in Radio and Wireless Symposium (RWS), 2010 IEEE. IEEE, 2010, pp. 184–187.

[34] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction based Volterra behavioral modeling of rf power amplifiers," IEEE Transactions on microwave theory and techniques, vol. 54, no. 12, pp. 4323–4332, 2006.

[35] L. Guan and A. Zhu, "Simplified dynamic deviation reduction-based Volterra model for Doherty power amplifiers," in Integrated Nonlinear Microwave and Millimeter-Wave Circuits (INMMIC), 2011 Workshop on. IEEE, 2011, pp. 1–4.